

This text represents the submission for the degree of Doctor of Philosophy at the Royal College of Art. This copy has been supplied for the purpose of research for private study, on the understanding that it is copyright material, and that no quotation from the thesis may be published without proper acknowledgement.

Visualising Cultural Data

Exploring Digital Collections
Through Timeline Visualisations

Florian Kräutli

A thesis submitted in partial fulfilment of the requirements of the Royal College of Art for the degree of Doctor of Philosophy
Examined by Professor Sue Walker, University of Reading, and Professor Paolo Ciuccarelli, Politecnico di Milano, on 20 April 2016 at the Royal College of Art in London.

This PhD is funded by EPSRC CASE award EP/J502169/1 and System Simulation Ltd.

Abstract

This thesis explores the ability of data visualisation to enable knowledge discovery in digital collections. Its emphasis lies on time-based visualisations, such as timelines.

Although timelines are among the earliest examples of graphical renderings of data, they are often used merely as devices for linear storytelling and not as tools for visual analysis. Investigating this type of visualisation reveals the particular challenges of digital timelines for scholarly research. In addition, the intersection between the key issues of time-wise visualisation and digital collections acts as a focal point. Departing from authored temporal descriptions in collections data, the research examines how curatorial decisions influence collections data and how these decisions may be made manifest in timeline visualisations.

The thesis contributes a new understanding of the knowledge embedded in digital collections and provides practical and conceptual means for making this knowledge accessible and usable.

The case is made that digital collections are not simply representations of physical archives. Digital collections record not only what is known about the content of an archive. Collections data contains traces of institutional decisions and curatorial biases, as well as data related to administrative procedures. Such 'hidden data' - information that has not been explicitly recorded, but is nevertheless present in the dataset - is crucial for drawing informed conclusions from digitised cultural collections and can be exposed through appropriately designed visualisation tools.

The research takes a practice-led and collaborative approach, working closely with cultural institutions and their curators. Functional prototypes address issues of visualising large cultural datasets and the representation of uncertain and multiple temporal descriptions that are typically found in digital collections.

The prototypes act as means towards an improved understanding of and a critical engagement with the time-wise visualisation of collections data. Two example implementations put the design principles that have emerged into practice and demonstrate how such tools may assist in knowledge discovery in cultural collections.

Calls for new visualisation tools that are suitable for the purposes of humanities research are widespread in the scholarly community. However, the present thesis shows that gaining new insights into digital collections does not only require technological advancement, but also an epistemological shift in working with digital collections. This shift is expressed in the kind of questions that curators have started seeking to answer through visualisation. Digitisation requires and affords new ways of interrogating collections that depart from putting the collected artefact and its creator at the centre of humanistic enquiry. Instead, digital collections need to be seen as artefacts themselves. Recognising this leads curators to address self-reflective research questions that seek to study the history of an institution and the influence that individuals have had on the holdings of a collection; questions that so far escaped their areas of research.

Contents

	Abstract	4
	Acknowledgements	16
	Author's Declaration	18
1	Introduction	22
	Definitions	24
	Timeline · Digital Collection · Curator	
	Research Questions	26
	Thesis Structure	27
	Methodology	31
	Research Through Design	
	Framework	36
	Technical Methods	36
	Users, Experts, Collaborators	37
2	Foundations	41
	Digital Humanities	42
	Origins · Distant Reading · Criticism	
	Visualisation	50
	History of (Time-Wise) Visualisation · Visualisation for Sense-Making	
	Collections	62
	Collections As Data	
	Time	70
	A Framework for Sense-Making · Time, History and Chronology	
	Curators' Perspectives	76
	Objects and Time in Digital Collections	
	Discussion	84
	Data Visualisation for the Humanities · Digital Collections As Objects of Research · Digital Historiography	

3	Digital Timeline (Tools)	92
	Defining Analytical Timelines	94
	Evaluation Criteria for Timeline Visualisations	
	Static Timelines	100
	Dynamic Timelines	106
	Exploratory Timelines	112
	Open Timelines	117
	Discussion	123
	Findings Led by Criteria · Findings Led by Spectrum	
	Focus Issues	125
	Scale · Layout · Temporal Descriptions	
4	Prototypes	130
	Prototype 1 – Metaballs	132
	Prototype 2 – Temporal Projections	137
	Prototype 3 – Drawing Uncertainties	141
	Prototype 4 – First Encounters With Tate	145
	Prototype 5 – ChronoZoom	151
	Prototype 6 – Tate Artists	155
	User Testing	
	Prototype 7 – Britten Force	161
	Prototype 8 - Temporal Jittering/Parallel Timelines	165
	Emerging Design Principles	169
	Timelines for Distant Reading · Embedding Uncertainties · Multiple Temporalities	
5	Paradigmatic Prototypes	174
	TT	175
	Walkthrough · Architecture · Discussion – Temporal Jittering Layout	
	Britten’s Poets / Temporal Perspectives	187
	Walkthrough · Architecture · Discussion – Temporal Perspectives	

6	Evaluation	200
	Mediated Insights – Timeline Tools	201
	Cataloguing Biases · Missing Dates · Dating Strategies · The Structure of a Collection	
	Mediated Insights – Perspectives	214
	Explicit and Implicit Relationships · Distant Reading · Close Reading · Different Emphasis Through Different Timescales	
	Expert Insights	225
	Design Decisions and Collaborative Design · Critical Thinking, Interpretation and Explanation · Enjoyment · Seeing Everything · Critical Chronographics · The Merits of Errors · Revealing Biases and Interpretations · Meta-Questions · Privacy · Coda	
	Discussion	238
7	Conclusion	242
	Contributions	243
	Process and Limitation	244
	Cross-Contamination	245
	Future Work	247
	Beyond Newtonian Time · Beyond Collections Data	
	Dissemination and Impact	249

	Appendix	250
A	Timeline Criteria	251
B	Datasets	263
C	List of Digital Timelines	270
D	Documentation – TT	301
E	Documentation – Temporal Perspectives	307
	Bibliography	310
	Glossary	339

List of Figures

- Figure 2.1 Roberto Busa (1913-2011). p. 43
- Figure 2.2 Misinterpretation of two-digit dates in JavaScript. p. 48
- Figure 2.3 A line chart by William Playfair (1786). p. 50
- Figure 2.4 One of the first known bar charts by William Playfair. p. 52
- Figure 2.5 Priestley's *Chart of Biography* (1764). p. 53
- Figure 2.6 Barbeau-Dubourg's 16.5m long *Carte Chronographique*. p. 54
- Figure 2.7 De La Bruyère's *Mappe-Monde Historique* (1750a). p. 55
- Figure 2.8 Priestley's notation of uncertainty in dates. p. 56
- Figure 2.9 A map of time by Girolamo Andrea Martignoni. p. 57
- Figure 2.10 Nigel Holmes' diagrams of US House and Senate expenditures. p. 58
- Figure 2.11 The Smithsonian National Museum of Natural History bird collection. p. 64
- Figure 2.12 One of more than 230,000 user curated collections on the Rijksmuseum's website. p. 65
- Figure 2.13 Europeana. p. 69
- Figure 2.14 *Lifestreams*. p. 71
- Figure 2.15 Decade Summary of the Australian Prints + Printmaking online collection. p. 87
- Figure 2.16 Artscope. p. 89
- Figure 3.1 What on Earth? Wallbook Timeline of Big History. p. 94
- Figure 3.2 Criteria tree for studying visual timelines. p. 99
- Figure 3.3 Timeline visualisation plotting twenty-one historic empires. p. 101
- Figure 3.4 Modern Gantt chart. p. 102
- Figure 3.5 Edward Lee's interactive timeline of empires. p. 103

- Figure 3.6 Dotted line in Accurat’s timeline. p. 104
- Figure 3.7 Events in the life of Nelson Mandela. p. 106
- Figure 3.8 The biography of Steve Jobs. p. 107
- Figure 3.9 The C.I.A.’s secret interrogation programme. p. 107
- Figure 3.10 Timeline by the UK Guardian. p. 108
- Figure 3.11 Emma Willard’s *Temple of Time* (1846). p. 109
- Figure 3.12 *Kindred Britain*. p. 112
- Figure 3.13 Timeline layout of *Kindred Britain*. p. 113
- Figure 3.14 Periscopic’s visualisation *U.S. Gun Deaths*. p. 116
- Figure 3.15 Continuum. p. 117
- Figure 3.16 TimeFlow. p. 118
- Figure 3.17 Simile Timeline. p. 119
- Figure 3.18 Palladio. p. 120
- Figure 3.19 Palladio’s Timespan filter. p. 121
- Figure 3.20 Palladio introduced two time-based views “Time” and “Duration”. p. 122
- Figure 4.1 Grouping of contiguous event using a Metaball algorithm. p. 133
- Figure 4.2 A prototype application, based on the Geffrye Museum’s collection. p. 134
- Figure 4.3 Overview timeline in TimeFlow. p. 135
- Figure 4.4 Coiling up a timeline. p. 137
- Figure 4.5 Concept prototype allows a timeline to be folded. p. 137
- Figure 4.6 A two-dimensional model of time in a circular projection. p. 138
- Figure 4.7 A tree of possible histories. p. 139
- Figure 4.8 Spaces of uncertain time frames. p. 139
- Figure 4.9 Events with varying certainties. p. 141
- Figure 4.10 Standard error bars. p. 142
- Figure 4.11 Tate’s collection. p. 146
- Figure 4.12 Works by J.M.W. Turner account for more than half of the Tate’s collection. p. 147

- Figure 4.13 Prototype visualisation of the ChronoZoom dataset. p. 152
- Figure 4.14 Sketch of the tree-based timeline layout. p. 153
- Figure 4.15 Lifespans of the pioneering timeline designers. p. 154
- Figure 4.16 Reusable timeline tool that features semantic zooming. p. 156
- Figure 4.17 Date brackets in Britten-Pears' dataset. p. 157
- Figure 4.18 Benjamin Britten's works visualised as dots. p. 161
- Figure 4.19 The force-directed layout algorithm generates regular clusters. p. 162
- Figure 4.20 Prototype that visualises Britten's works using the Temporal Jittering layout. p. 165
- Figure 4.21 Britten's pieces are visualised by date of composition (top) and date of first performance (bottom). p. 166
- Figure 5.1 Tate's complete artwork dataset as represented by the TT prototype. p. 175
- Figure 5.2 Zooming in on the aggregate shape. p. 176
- Figure 5.3 Clusters of works that share a commonality. p. 177
- Figure 5.4 A customisable panel with details about a record. p. 178
- Figure 5.5 Eric Gill in the Tate dataset. p. 178
- Figure 5.6 Credit line reveals a cluster. p. 179
- Figure 5.7 Panel display. p. 181
- Figure 5.8 Temporal Jittering algorithm. p. 182
- Figure 5.9 Three variations of a histogram. p. 183
- Figure 5.10 Time-based histograms assign events to one of several predefined timeframes. p. 185
- Figure 5.11 Temporal Jittering algorithm allows records to spread out across their permitted timeframe. p. 185
- Figure 5.12 Histogram-based timeline of the Australian Dress Register. p. 186
- Figure 5.13 Initial view of Britten's Poets. p. 187

- Figure 5.14 Blake, de la Mare and Auden. p. 189
- Figure 5.15 Scaling each timeline independently (bottom). p. 190
- Figure 5.16 *The future, as foretold by the past.* p. 192
- Figure 5.17 *Art movie time machine.* p. 193
- Figure 5.18 SemTime. p. 194
- Figure 5.19 Positioning authors. p. 197
- Figure 5.20 Side-effect of positioning events by their temporal midpoint. p. 198
- Figure 6.1 Artwork dataset of the Tate. p. 201
- Figure 6.2 Zooming in on a cluster between 1800–1850. p. 202
- Figure 6.3 Every page in J.M.W. Turner's sketchbooks. p. 203
- Figure 6.4 Another suspicious cluster remains after removing Turner from the diagram. p. 204
- Figure 6.5 TimeLine Curator. p. 205
- Figure 6.6 Artworks in the Tate dataset, visualised by dates of acquisition. p. 206
- Figure 6.7 Oxford's Beazley Archive. p. 207
- Figure 6.8 Cooper-Hewitt's object records. p. 208
- Figure 6.9 Records with precise dates (exact years) appear in vertical spikes. p. 208
- Figure 6.10 Records that correspond to the 1930s. p. 209
- Figure 6.11 Curators at Cooper-Hewitt refrained from deciding between two relevant dates. p. 210
- Figure 6.12 Object records in Cooper-Hewitt's digital collection by their associated decade. p. 211
- Figure 6.13 Dataset of the Geffrye Museum. p. 212
- Figure 6.14 Large cluster of missing images. p. 212
- Figure 6.15 Spiky cluster at the right end of the diagram represents recent items in Geffrye's collection. p. 213
- Figure 6.16 Temporal Perspectives view of a part of the V&A dataset. p. 216
- Figure 6.17 MoDA. p. 216

- Figure 6.18 Individual items relate to other aspects of the collection. p. 217
- Figure 6.19 Publication on Lewis Jones. p. 219
- Figure 6.20 The objects MoDA featured on their blog. p. 219
- Figure 6.21 A university art history course, which apparently focuses on two distinct timeframes. p. 219
- Figure 6.22 Selecting a poet. p. 220
- Figure 6.23 Individual authors share a relationships through Britten's works. p. 220
- Figure 6.24 *Voices for Today*. p. 221
- Figure 6.25 A time axis that can be dynamically adjusted. p. 223
- Figure 6.26 A prototype iteration that uses fisheye distortion. p. 223
- Figure 6.27 Britten's Poets implements three possible time scales. p. 224
- Figure 6.28 Tryout visualisation of Johnston-related items. p. 228
- Figure 6.29 Oxford's Beazley Archive plots database records on a map. p. 230

List of Tables

- Table 2.1 Date descriptions in the V&A collections data.
Dots (•) represent numbers from 0-9. p. 77
- Table 3.1 A schema for classifying digital timelines according
to their interactive features. p. 96

Acknowledgements

I wish to express my sincerest gratitude to Professor Stephen Boyd Davis, whose commitment and dedication made this research possible. Thanks for supporting my PhD in the best possible way as a crucial and critical advisor and collaborator.

I thank my external supervisor Dr. Mike Stapleton at System Simulation for his invaluable advice and encouraging support throughout the course of my PhD.

Professor Sue Walker and Professor Paolo Ciuccarelli have generously shared their experience and expertise by acting as my external examiners. I am very thankful for the diligence with which they examined my work, their critical appraisal and valuable input.

I am grateful for the generous funding provided by the EPSRC in collaboration with the Knowledge Transfer Network and System Simulation Ltd.

I thank everyone at System Simulation for their support and guidance. Thanks to Dr. Mike Stapleton, Dr. Michael Selway, and the entire team: Alec, Fulvio, Gavin, George, John, Mali, Mark, Matt, Matthew, Mia, Nich, Olly, Richard, Sacha and Steve.

I had the honour to meet and collaborate with many inspiring people at cultural institutions across the UK. They have been extremely generous in sharing with me their time, expertise and enthusiasm. I can't thank them enough for the contribution they have made to my research. I specifically thank Dr. Lucy Walker of the Britten-Pears Foundation, who has been an exceptional collaborator early-on and throughout the time of my PhD. I truly appreciate the help of all the curators and thank them for making this research possible.

I was fortunate to get the opportunity to bring my research topic to a variety of places. During an internship at Microsoft Research I explored the potential of time-wise visualisation for personal data. I would like to thank Professor Richard Harper for making this possible. I thank the organisers for inviting me to *Beautiful Data*, a Getty funded Summer School at Harvard University, and all the participants for their inspiration and continuing collaborations.

I'm grateful for my colleagues at the Royal College of Art who helped me maintain sanity during the sometimes peculiar journey of a PhD. I'm thankful for all the friends who supplied me with advice,

company, food and shelter: Delfina, Helga, Veronica, Cho, Mia, Sabina, Matt and Aindreas. I thank my fellow Chronographers Olivia and Sam and wish them all the best for their own PhD journeys.

I wish to thank Dr. Jonathan Antonio Edelman and Dr. John Stevens who have been extremely helpful in the lead up to my PhD viva.

A special thanks goes to Ayham Ghraawi, for invaluable advice on the layout and typography of this thesis, to Professor Michael Hohl, for acting as an RCA research outpost in Berlin, and to Professor Mark Bishop, for making me consider doing a PhD in the first place.

I could not have achieved any of this if it were not for my truly amazing wife Katharina. Thank you for your never-ending support, the confidence you give me, your love and friendship.

Author's Declaration

During the period of registered study in which this thesis was prepared the author has not been registered for any other academic award or qualification. The material included in this thesis has not been submitted wholly or in part for any academic award or qualification other than that for which it is now submitted.

Florian Kräutli
11 February 2016

Publications

I have previously published outcomes of the presented research in the following publications and conference contributions:

Boyd Davis, S. & Kräutli, F., 2015. The Idea and Image of Historical Time: Interactions between Design and Digital Humanities. *Visible Language*, 49(3), pp.101-119. Available at: <http://visiblelanguagejournal.com/issue/172/article/1202> [Accessed 5 February 2016].

Kräutli, F. & Boyd Davis, S., 2015. Revealing Cultural Collections Over Time. In R. Borgo & C. Turkay, eds., *EG UK Computer Graphics & Visual Computing*. London: The Eurographics Association.

Boyd Davis, S. & Kräutli, F., 2014. Scholarly chronographics: can a timeline be useful in historiography? *European Social Science and History Conference*. Vienna, 23-26 April 2014. Amsterdam: International Institute of Social History.

Kräutli, F., 2013. Challenges for time as digital data. *Herrenhausen Conference: (Digital) Humanities Revisited*. Hanover, 5-7 December 2013. Hanover: Volkswagen Stiftung.

Boyd Davis, S. & Kräutli, F., 2013. Time in Perspective: a visual approach to models of time. *CHI 2013 Workshop Changing Perspectives on Time*. Paris, 27 April - 2 May 2013. New York: ACM.

Kräutli, F. & Boyd Davis, S., 2013. Known unknowns: representing uncertainty in historical time. In K. Ng, J. P. Bowen, & S. McDaid, eds., *Proceedings of EVA London*. London: BCS Learning and Development Ltd, pp. 61-68.

Outcomes of my research have appeared in the following publications by other authors:

Boyd Davis, S., 2016. Time Machines. In J. Pilcher, ed., *Culture, technology and the image*. Bristol: Intellect. In Press.

Boyd Davis, S., 2016. "To see at one glance all the centuries that have passed" - early visualisations of historical time. In A. Black, O. Lund, & S. Walker, eds., *Gower handbook of information design*. Farnham: Ashgate. In Press.

Mitchell, L. ed., 2014. *Beautiful Data: A field Guide for Exploring Open Collections*. Cambridge, MA: metaLab (at) Harvard University. Available at: http://beautifuldata.metalab.harvard.edu/wp-content/uploads/2014/09/BD_Field_Guide%20small.pdf [Accessed 14 January 2016].

1

“There are circumstances where the best or only way to shed light on a proposition, a principle, a material, a process or a function is to attempt to construct something.” (Archer 1995, p.11)

1 Introduction

For the past three years I have been researching the use of interactive visualisations of cultural data, with a particular emphasis on mapping data against time. Visual timelines are among the earliest examples of graphical renderings of data, yet today they are often seen merely as a medium for linear storytelling rather than as tools for visual analysis – a purpose that is studied and advanced in the present thesis.

Digital catalogues of collections form the starting point of my research. Cataloguing an archive, library or collection is done for a variety of reasons. On the most pragmatic level, a catalogue helps to keep track of where items in a collection may be physically located. In addition, it can hold information of what is considered to be known about an item, such as its origin, its age, its value, its dimensions, etc. All these descriptions make up the metadata: the data about the ‘actual’ data in the form of an item in a collection or archive, or a book in a library.¹

The increasing digitisation of collection catalogues has enabled them to contain far more (meta) data than their paper and card based equivalents.² Digital catalogues have become digital collections in their own right and valuable resources for research, a development that cultural institutions are beginning to acknowledge.³

The present thesis is motivated by these developments and the inherent question of what kind of knowledge digital collections are able to contain and how new insights may be enabled by digital methods. It is based on two core propositions: the expectation that interactive data visualisations are able to serve as tools for knowledge discovery and the notion that the model of time provides a suitable frame of reference for sense-making regardless of the content and structure of the datasets that are to be studied.

My research has involved engaging with the topic of timeline visualisations for digital collections on a practical, theoretical and social level: by writing code, designing functional prototypes and acquiring and refining the skills necessary for doing so; by immersing myself in the discourse around visualisation and the implications of digital tools on humanities scholarship; and by collaborating with curators and archivists from different institutions, negotiating the use of their

- 1 I will refrain from drawing a clear distinction between "data" and "metadata" in the continuation of this thesis as my research is, in a sense, only concerned with the latter.
- 2 The current SPECTRUM standard for collections management specifies more than 500 attributes to describe a single item in a museum collection (Dawson & Hillhouse 2011).
- 3 "Collection metadata is both a key asset representing centuries of person years of investment and a potential enabler for current operations and future developments. However, despite its strategic importance for numerous stakeholders, the potential value [...] remains under exploited." (The British Library 2015)

datasets and getting their insights and feedback on the way I visualised their data.

The research contributes practical approaches as well as a conceptual understanding of how we can design and use digital visualisation tools in the shape of visual timelines to extract new knowledge from existing digital collections, and presents new findings on the nature of such knowledge that encompasses their content and histories.

I position my research within the framework of the Digital Humanities, a relatively young area⁴ that seeks to pursue and understand research at the intersection of digital technology, design and humanities. My research contributes to the ongoing challenge of developing and (re)defining digital methods for knowledge production in the humanities.

4 Digital Humanities has borne this name since 2001 (Kirschenbaum 2012).

My methodology is grounded in a practice based approach of iterative prototyping and critical reflection, where new knowledge emerges through the process of making and collaborating: what Frayling (1993) describes as research *through* art and design. It is informed by the challenges I have identified in the process, as well as by studying the relevant literature, participating in workshops and conferences, and conversations and collaborations with expert users.

While initially I considered the potential beneficiaries of visualisations of digital collections to be the general public, my focus soon shifted to people within the institutions owning the collections: their curators, archivists and researchers.

Despite the fact that they are experts in the subject matter, digitisation exposes them to new challenges. Visualisation, it turned out, enables these users to gain new perspectives on their own digital collections. In addition, their expertise allowed me to verify the efficacy of my visualisations - see if patterns that are visible in the visualisation align with their knowledge of the subject matter -, and find out what kind of discoveries they allow beyond what is already known by collaborating in the design process. These collaborations produced valuable insights in the research areas of the scholars and institutions I worked with, as well as on the potential of interdisciplinary research in the Digital Humanities.

My research contributes a thorough understanding of the kinds of knowledge that is embedded in digital collections with regards to the cultural artefacts they represent and, most importantly, in relation to the embedded histories and biases of collections data that will become evident through the course of my research. I offer novel approaches to the visualisation of temporal data and suggest new ways for timeline visualisations to be used to access and analyse collections data with regards to their identified potential as resources for knowledge within and beyond traditional humanities scholarship.

Definitions

A glossary is included on page 339 to offer explanations and working definitions of some of the technical terms that appear in this thesis. Here I offer a brief description of essential and recurring concepts.

Timeline

A timeline is, in this context, a specific kind of data visualisation which visually represents data organised by a model of time – be it by calendrical dates, sequence, cycles or any other temporal structure. Many types of datasets may be visualised on a timeline, such as categorical and numerical, or structured and unstructured data. A timeline is different from a time series, which plots a regular series of measurements over time. In contrast to other diagram formats such as bar charts or histograms, timelines do not need to summarise events – although they can – and often represent events graphically as individual instances.⁵

⁵ What counts as an ‘instance’ often depends on the dataset and the interpretation of the creator of a visualisation.

Although the term ‘timeline’ may also be used to describe simple text-based lists of chronologically ordered events or the display of content on a social media website, I will refer to timelines mainly as time-centric visualisations and will sometimes use the term ‘chronographic’ to make the distinction between this and other concepts evident.

Digital Collection

I have worked with digital collections obtained from institutions I collaborated with, which are listed in appendix B on page 263. In addition, I obtained datasets from institutions that let the public download a copy of their collections data, such as the Museum of Modern Art and the Cooper Hewitt Design Museum. What constitutes a collection more broadly and a digital collection in particular will be discussed in more detail in chapter 2 (page 62).

Curator

I will use the term ‘curator’ to refer to a diverse group of experts with whom I collaborated. The title is not necessarily applicable to all of them, nor does it cover their area of work entirely. While some of the people do carry ‘curator’ in their professional roles and are responsible for putting together public exhibitions, others have different areas of responsibilities. The commonality of the people I collaborated with is that their profession involves accessing and curating cultural datasets; in the original Latin sense of the word ‘taking care of’.

Their comments and remarks, which I have transcribed from audio recordings of our meetings, are included in this thesis and distinguished as C1 to C12. For reasons of confidentiality and privacy, I have not included the complete transcripts of our conversations. Nevertheless, I have taken great care that the original context of the extracted quotes is retained.

Research Questions

The central question I seek to answer through my research is: What kind of knowledge can we gain from visually analysing digital collections through timeline visualisations?

I aim to contribute to our understanding of how digital tools shape and enable humanities knowledge production, through their application as well as during their development. What kind of knowledge is embedded in digital collections, which timeline visualisations could provide access to? What can we learn by examining collections data in visual timelines?

In order to address my main research question I need to answer a range of sub-questions:

What is collections data? How does it relate to collections?

What are the implications of digitisation in this context and in humanities research more broadly?

What are the particular challenges around time based visualisations of large cultural collections? How can we approach them?

What questions do those that use cultural collections want to ask? How do visualisations benefit them?

Thesis Structure

The present chapter continues with a description of the methodology and framework of my PhD, followed by an introduction of the individuals involved.

The rest of the thesis is structured as follows:

Chapter 2: Foundations

I contextualise my research through a discussion of its field and the central concepts of my PhD. I offer a brief account of the history and promises of the Digital Humanities, as well as the criticism it faces.

A historic perspective on visualisation emphasises the important role that timelines have played early-on in the development of data visualisation formats, both for the purpose of communicating and analysing data. Time acts as a unified framework for structuring and making sense of data - an ideal model for historic research on cultural collections. Time however poses its own challenges specifically with regards to digitally stored date descriptions that form part of collections data. Collections and the consequences of their digitisation are discussed.

I present current views on the prevalent understanding of collections, including my own findings on the representation of time and objects in collections data that I have gathered by collaborating with museum experts. These conversations illustrate how collections increasingly cease to be understood as neutral sources of evidence; curators begin to show awareness for the subjectivities in collection practices.

Embedded interpretations, I argue, must be considered as a fundamental aspect of digital collections in order for visualisation tools to not be misleading: a recurring critique of software for humanities research, but also one that diverts the responsibility to the tool. What is also necessary is that humanities researchers become more informed concerning the kind of questions that digital tools can and cannot answer.

Chapter 3: Digital Timeline (Tools)

In this chapter I review existing projects and software tools that implement timeline visualisations for analysing data. I employ a schema for classifying visualisation tools according to their supported level of interactivity.

To ensure a rigorous analysis of this broad class of visualisations, I draft a set of criteria according to which I examine the presented projects. The criteria act as a kind of checklist in order to prevent me

from taking familiar graphical representations for granted and instead question their underlying assumptions.

In my review I look closer at established concepts and identify potential shortcomings with regards to the time-wise visualisation of digital collections. Current timeline visualisations tend to focus on the representation of individual events or sequences, making them unable to visualise large cultural datasets in an insightful manner.

In addition, events are generally treated as occupying a singular and clearly defined point in time - an expectation which cultural data is unable to fulfil. Historic events often have a multiplicity of associated dates and these temporal descriptions are defined with varying degrees of certainty.

Lastly, the readability of a timeline visualisation depends to a great extent on the graphical layout - nevertheless digital timelines often employ pragmatic layout algorithms that tend to hide patterns that may be present in a dataset. These focus issues form the starting point of my practice-led explorations.

Chapter 4: Prototypes

The fourth chapter discusses the progression of my research through iterative prototyping and evaluation. I discuss eight prototypes (P1-P8) with a focus on the insights they produced in order to enable the reader to retrace the progression and refinement of the outlined issues.

Beginning with innovations in the time-wise representation of cultural data based on mathematical concepts and model implementations, my process soon takes a turn closer to experimenting with existing datasets; the messiness of real-world datasets and the intricacy of humanities research cannot be simplified, but need to be addressed in all their complexity.

The issues of timeline layouts and the representation of large datasets soon converge and I discover how both can be tackled by developing a timeline format that produces a representation of a complete dataset through an emergent behaviour operating on every individual record - which is different from existing approaches that either represent each event separately or summarise several events based on pre-defined thresholds. The issue of multiple and uncertain temporal descriptions in cultural data leads me to a conceptual shift in seeing them not as additional qualifiers, but as a defining property of cultural data which facilitates new perspectives and multiple viewpoints on an individual digital collection.

Based on the defined focus issues, I formulate principles for the design of timeline visualisations for cultural data; these relate to distant reading, embedded uncertainties and multiple temporalities.

Chapter 5: Paradigmatic Prototypes

The proposed design principles derived from my prototype-led explorations manifest in two example implementations. This chapter describes their architecture, discusses their conceptual background and relevant aspects of their design, and puts them in context by comparing them with previous work in the field.

Timeline Tools is a reusable visualisation library that facilitates exploration and visual analysis of arbitrary cultural datasets. It is specifically aimed at large datasets and has successfully been tested with collections that contain several hundred thousand records.

Temporal Perspectives implements an algorithm that restructures cultural datasets according to multiple temporal dimensions and produces a timeline layout that enables researchers to study the temporal relationships within a cultural dataset. It is put into practice in a prototype that extracts and visualises the relationships of authors whose writings were used in compositions by Benjamin Britten.

Chapter 6: Evaluation

Finally, I return to my main research question: What kind of knowledge can we gain from visually analysing digital collections through timeline visualisations?

This is answered, in one part, through an account of the specific insights that surfaced through the application of the presented visualisation tools. In addition to offering an impression of a collection's content, collections *data* - visualised according to the developed principles - reveals embedded characteristics of a collection: how its structure has developed, how embedded biases in cataloguing determine the shape of a collection and how the use and public understanding of a collection has changed over time.

The second part of the answer is given through the voices of collaborating curators. Visualisation tools, paired with an improved knowledge of the implications of digital methods on their research practices enables them to ask new questions. These show a departure from 'traditional' research interests that concern aspects of cultural artefacts and the biographies of their creators - questions which would not necessarily require digital means to be answered. Instead, curators became interested in 'meta-questions' that explore the history of their institution as represented by data and how they and their predecessors shape and reshape the knowledge contained in collections.

Chapter 7: Conclusion

The thesis concludes with a summary of my contributions to knowledge and puts the findings in perspective by pointing out the limitations of my research and follow-up questions. I reflect on the implications of an interdisciplinary research project for my PhD and Digital Humanities scholarship more broadly. Finally, I highlight some of the immediate outcomes that have so far resulted from the presented research.

Methodology

I address my research questions through a combined methodology of theoretical and practical research, whereby my main focus lies on the practice-based elements: the knowledge that I have acquired through my research and which is made explicit in this thesis is derived from a method of enquiry that depends on the design of digital prototypes. The order in which the methods are presented below is not a chronology of how the research was conducted; the different steps intertwine and informed each other, but are presented here, for clarity, as a sequential model.

Beginning with a **literature review** I embed my questions and subsequent findings in the ongoing discourse of the field. Through a thorough study of current and historic research, and a review of existing timeline visualisation projects I am able to identify gaps that my work should address which I formulate as ‘focus issues’ that will later be developed and explored.

By conducting **interviews** with curators, archivists and other representatives of cultural institutions I gather insights on the nature of digital collections and their embedded assumptions, specifically with regards to time and temporal descriptions. My findings contribute to a better understanding of the significance of collections data and will be vital for my further practical enquiries. Conversations with curators and archivists continued throughout my research process. These produced valuable insights leading to a deeper understanding of the role of digital data and visualisations for cultural institutions in general, and most importantly served as evaluation method for my prototypes.

By making **prototypes** of timeline visualisations I am able to instantiate and test my ideas and hypotheses, and critically reflect on their implications. *Making* constitutes my main research method: the creation of an artefact, the interaction with it and its evaluation leads to discoveries, insights and new questions. Prototypes tackle one or several of the focus issues that I have identified and help to refine them through ongoing testing and evaluation. This practice-based element is the main driving force of my research and I discuss its purpose below, along with my reasons for addressing my research questions through a design-led approach.

Evaluation of the prototypes occurs through several ways. First of all through critical reflection on the process of making as well as while interacting with the prototypes. In instances where I have published work in progress on my blog or through conferences, peer critique serves as an additional mode of evaluation. Finally, I evaluate and test the prototypes together with their future users. Over time, my interac-

tion with curators and archivists developed from consultation on matters of data access and collection specifications towards collaboration on the design and development of visualisations. The resulting discussions were critical for the evaluation of my prototypes.

Evaluating the prototypes leads to insights on several levels. On a local scale, I can identify the potential and drawbacks of individual prototypes. Furthermore, evaluating allows me to reflect on what kind of practical insights may be scalable and transferable. Most importantly, the evaluation process enables me to develop an understanding of the kind of knowledge that is contained in digital collections and how it appears in time-wise visualisations. The knowledge gained through evaluation is made explicit in writing and feeds into later iterations of prototypes. By refining and re-defining the goals and challenges, evaluation defines the approach and criteria for the next iteration of prototypes.

Research Through Design

My practical research method is based on iterative design of functional visualisation prototypes for digital cultural collections. Prototyping acts as a method to generate knowledge by making and reflecting on the creation process as well as through interaction with and evaluation of a created prototype.

I work with publicly available datasets as well as data I have obtained directly from the institutions I collaborated with. Because my prototypes are based on existing cultural datasets, they reflect and expose the challenges of time-wise visualisation in real-world applications.

A core element of my process is the constant evaluation of the created artefacts in the form of critical reflection and ongoing dialogues with museum curators and archivists, who are both experts and the future users of my visualisation tools. This is a form of *Critical Making* (Ratto 2011) which emphasises iterative and collaborative methods and uses the working process itself as the locus of evaluation, rather than employing a separately designed user study.⁶

The starting point of every prototype iteration is a problem statement. It defines the specific focus issue the prototype should address and the proposed solution that is implemented by the prototype. Through the course of the research these are refined and reformulated. Design researchers have to regard the design brief as part of the problem, as was pointed out by Archer (1968). I present my point of departure – my own design brief as it were – and hypothesis at the beginning of the discussion of each prototype iteration, and specify the datasets the visualisation is based on or has been tested with.⁷

6 I conducted one controlled user study, but returned to this mode of reflective and ongoing evaluation as the results of the study did not cover the full scope of my research questions. The study will be discussed later on (see page 156).

7 My goal is to arrive at approaches that generalise to a variety of cultural collections and datasets that I may not be able to anticipate. I therefore make sure my visualisations can – in principle and in practice – function with a variety of datasets.

The creation of a prototype and the subsequent interaction with it constitute the research by uncovering new insights and questions that emerge during development and evaluation of the prototype, as well as by providing evidence-based support or disproof for a hypothesis. Hereby, the practice-based elements are not disconnected from the theoretical foundations. Prototyping and making both inform and are informed by my theoretical enquiries. Making cannot be separated from the other methods of a practice based research project as it necessarily causes one to reconsider and reformulate issues, questions and problems, which may only have been revealed by the design process itself and newly discovered issues require a further engagement using all available methods.

Why Do I Conduct Research Through Design?

Design based research methods lend themselves to address a particular type of problems: the kind that Rittel and Webber called “wicked problems”:

The problems that scientists and engineers have usually focused upon are mostly “tame” or “benign” ones. [...] the mission is clear. It is clear, in turn, whether or not the problems have been solved. Wicked problems, in contrast, have neither of these clarifying traits [...] (1973, p.160)

Rittel and Webber refer to problems that can not be solved through analysis; the process of enumerating all possible solutions in order to pick the best one. Wicked problems exhibit incomplete and often unrecognisable characteristics, making them nearly impossible to solve completely. Designers address such problems not by analysing - the scientific method - but by making.⁸

8 “The scientific method is a pattern of problem-solving behaviour employed in finding out the nature of what exists, whereas the design method is a pattern of behaviour employed in inventing things of value which do not yet exist. Science is analytic; design is constructive.” (Gregory 1966, p.6)

There are numerous factors that may determine what a visualisation may reveal in cultural data, even without the additional and potentially arbitrary requirements mandated by a specific dataset and the individual expectations of a researcher. The problem at hand is a wicked problem, one where trying to make is more likely to lead to satisfactory results, or any results at all, than attempting to first comprehend the problem in its entirety. As Archer writes:

There are circumstances where the best or only way to shed light on a proposition, a principle, a material, a process or a function is to attempt to construct something (1995, p.11).

Attempting a solution in practice is, of course, no guarantee for success. In fact, it may even reveal more and previously unidentified problems. In practice-based research however, this does not constitute

a failure. We can use these instances where a system breaks due to the complexity of the problem as a way of getting insights. Attempting to create a solution rather than drafting a complete analysis enables one to problematise what may not initially be identified as a problem, thus creating knowledge and uncovering new opportunities for research.

Practice-based research is not without its drawbacks. By making and acting on the real world, a design researcher influences and is influenced by the problem at hand (Schön 1983; Dorst 2003); he cannot simply position himself as a 'neutral' observer. Practice-based research is always to a certain degree subjective. It is therefore essential to compensate for a lack in objectivity by following a rigorous and transparent process and to strive to make the plans and decisions, as well as one's beliefs, encounters and experiences that may have influenced them, explicit in writing or using another appropriate method of documentation.

For the same reasons, practice-based research may not be completely reproducible. A person undertaking the same research, following the same questions based on the same materials, may nevertheless produce different artefacts and arrive at different conclusions. Again a thorough documentation is therefore necessary in order for a research process to be – if not exactly reproducible – at least comprehensible. The research outcome will then nevertheless be deducible and can be subjected to review and peer critique.

⁹ A more comprehensive definition of design comes from HA Simon, who sees design as the act of "devis[ing] courses of action aimed at changing existing situations into preferred ones." (1996, p.111). However, this could be understood as if "preferred" situations exist and can be universally agreed upon. Instead, I argue, that a preferred situation can only be defined according to a specific set of criteria as well as subjective preference. Furthermore, what constitutes "preferred" may change during a design process; even the "existing" state could turn out to be the "preferred" state.

When Does Design Constitute Research?

Design is the discipline which addresses problems through an act of making.⁹ The problems may be specific, such as the design of an exhibition in a particular space. They can be reproducible, such as the design for a mass-manufactured product. We speak of graphic or communication design, when the design problem applies to the optimal arrangement of type, shapes and images in a two dimensional space, or of vehicle or transportation design when the problem has to do with people or goods overcoming distances.

Designing for these problems is informed by research, involves doing research and results in new knowledge for the people participating in the design process. Can we therefore argue that doing design inherently results in doing research?

Design is interdependent on research, yet it is not equivalent. Where design seeks an acceptable solution for a particular problem within the given circumstances, research aims to arrive at communicable and generalisable 'truths':

Design is a way to ask questions. Design Research, when it occurs through the practice of design itself, is a way to ask larger questions beyond the limited scope of a particular design problem. (Zimmerman 2003, p.176)

In order to determine when design constitutes research, we have to answer:

[...] was the practitioner activity an enquiry whose goal was knowledge? Was it systematically conducted? Were the data explicit? Was the record of the conduct of the activity 'transparent' [...] ? Were the data and the outcome validated in appropriate ways? (Archer 1995, p.10)

If, and only if, the intention of doing design is to arrive at communicable knowledge and if the design process is undertaken transparently and rigorously, can we speak of research through design.

Framework

I undertake this research project in the framework of an industrial CASE studentship funded by the EPSRC in partnership with System Simulation, a central London software engineering company with decades of experience in database applications for the heritage sector. Their clients include a wide range of holders of cultural collections, such as the V&A, the British Museum and the Wellcome Trust. Part of the PhD included an industrial placement at their offices and supervision by Mike Stapleton. This provided me with access to their staff's expertise and enabled me to collaborate with many of their clients. At the Royal College of Art I am enrolled in the Innovation Design Engineering department and supervised by Professor Stephen Boyd Davis. This constellation creates a unique triad for an interdisciplinary PhD in the Digital Humanities; covering the areas of humanities, design and computing equally.

Technical Methods

To develop functional prototype visualisations I made use of standard web-development tools. Languages such as JavaScript, HTML and CSS were used for the public-facing interfaces, server-based languages such as PHP for communicating with APIs of online collections. Most visualisations are based on D3.js, a widely used JavaScript library geared towards web-based interactive data visualisations.

Users, Experts, Collaborators

Hereby a brief introduction of the individuals involved in my research: first, there is the staff of System Simulation, a group of highly skilled software engineers. To them I turned to learn about the technical aspects of cultural datasets; the inner workings of the databases and the conceptual and computational strategies for dealing with the particularities of collections data and processes. I developed many of my visualisations inside their offices, which often made them some of the first to test and give feedback on prototype visualisations. Such informal tests on a small number of users are highly effective for identifying most of the common usability problems - researchers find that testing with more than four to five subjects does not lead to substantially better insights (Virzi 1990; Nielsen & Landauer 1993; Nielsen 2000).

Then, there are the scholars and professionals working at museums, archives and other cultural organisations who I met over the course of my research. Seeking institutions willing to share their datasets for my visualisation projects, I was lucky to meet early on with Dr Lucy Walker, director of Learning at the Britten-Pears foundation. Having herself experimented with Microsoft Excel diagrams of the dataset she just finalised, she already had an idea of the potential of visualisations for cultural data and the possible questions they might be able to answer.

Once I had my own prototype visualisations to offer, I could better communicate the purpose and potential of my research to curators unfamiliar with the topic and it became easier to encourage other scholars and museum professionals to share their data and expertise. Individuals with whom I collaborated include scholars, curators and technical staff from (in alphabetical order) Britten-Pears foundation (Aldeburgh), Courtauld Institute of Art (London), Geffrye Museum (London), ICA Philadelphia, King's College (London), National Archives (Kew), National Library of Wales (Aberystwyth), London Transport Museum, Museum of Domestic Design and Architecture (London), Oxford Beazley Archives, Science Museum (London) and Tate (London). In addition, participating in conferences, workshops and a summer school on the topic of digital museum collections and visualisation allowed me to meet and discuss with representatives of a range of other institutions and gather feedback on my prototypes on an informal basis.

Initially I approached institutions mainly seeking material to experiment with. Collections data and APIs were still few and far between when I began my research. Only recently have museums started to be more generous with their data.¹⁰ Meeting with curators in

¹⁰ Thanks to pioneering examples such as Cooper-Hewitt and the Rijksmuseum who successfully demonstrated the value of making collections data publicly available.

person, however, also gave me the opportunity to learn more about the history of their institution and the nature of their collection.

Last but not least, I need to introduce the envisaged users of visualisation tools for cultural collections. But who are they? Existing work in visualising museum collections largely explores their benefits for the general public. Hinrichs et al. (2008) discuss the use of interactive information visualisation within the museum exhibition environment and consequently by museum visitors: a “diverse audience [...] ranged from toddlers to elderly people” (ibid., p.1185). Hinrichs is a co-author of continuing studies on visualisations of digital collections in museums (Rogers et al. 2014) and libraries (Thudt et al. 2012), which evaluate interfaces with actual library users or simulated groups of typical museum visitors. Whitelaw (2009) who continuously studies and works on visual interfaces for cultural collections since 2009, assumes that “a user who is unfamiliar with the collection’s scope, contents, or structure” (ibid.) would benefit most from visualisations of archival collections; the main target group are “visitors unfamiliar with a collection” (Whitelaw 2012).

Users As Experts

In my own research I look at the topic from the opposite direction: regarding informed users, who are interested in scholarly analysis as the primary beneficiaries of visualisation tools of cultural collections, while assuming that also casual users will be able to take advantage of them. While visualisations can enable even uninformed viewers to observe patterns and make discoveries that, without the aid of diagrams, would require a great amount of desk research and expertise, it does not follow that visualisations of cultural collections have nothing to offer to those who are already familiar with the datasets. Existing work appears to suggest that the common search interfaces are sufficient for the normal tasks of expert users, while my interest is precisely in finding out what alternative perspectives on cultural datasets in the shape of time-centred visualisation tools may be able to offer even for advanced users.

Focussing on scholarly use still constitutes a large target group. Museum visitors span a wide variety of individuals, but so do researchers who might be looking into a specific archive for very different reasons. Oxford’s Beazley Archive describes their main users as a “small, but loyal number of researchers” (C7) who turn to the collection as a substantial resource for studying ancient pottery. More and more scholars, however, are looking at this archive to learn more about the personal life of its founder Sir John Beazley and may arrive at the data with relatively little knowledge, or immediate interest, in classic

archeology. It therefore makes sense, as Whitelaw does, to not presuppose that users of an archive are necessarily knowledgeable about its field, scope and content. However, in my own experience, we should also not underestimate the potential of visualisation to offer new insights even to expert users; those who have in-depth knowledge of a collection and are very able to also use conventional interfaces effectively.

Involving this type of user in my research allows me to identify the requirements and abilities of visual interfaces to reveal 'genuine' new knowledge: insights which an expert equipped with detailed understanding of an archive and its subject domain has not previously gained, discoveries that deepen existing knowledge, and new research questions which have been stimulated by using and examining a collection through a visualisation tool. In addition, collaborating with experts allows me to verify the discoveries which I and my users have made; to see if what we may read from a diagram or from interacting with a visualisation is effectively telling of the dataset itself and not just an artefact that resulted from the visualisation process. Expert knowledge provides a baseline for evaluating the honesty and 'truthfulness' of a visualisation tool.

The type of users of my visualisation tools include therefore the very scholars, curators and archivists with whom I collaborate on the data-gathering, development and evaluation of the visualisation tools. This requires them to wear more than one hat; specialist and collaborator, and test subject.

When meeting with these individuals, I steered our conversations so that the different roles - expert and test subject - of my counterpart stay, as far as possible, separated. First, we would talk about their subject domain, exchange views on the nature of digital collections and the potential of visualisations - these conversations contributed insights that informed the requirements of my visualisation tools and my understanding of humanities research more broadly. Then, I would prompt them with my visualisation prototypes and gather feedback and questions, while we jointly explore the datasets - this phase allowed me to collect evidence, to make and verify discoveries and to further evaluate my practical work. Both modes of conversation arrive at generalisable knowledge: questions and insights that are relevant within and outside the scope of a particular scholar or institution. The individuals with whom I collaborate are therefore not clients - I am not creating prototypes to meet their specific requirements - rather they are representative samples¹¹ of the type of scholarly user my work seeks to address.

11 Representative in their role as interested scholars and researchers, not in the statistical sense as a representative sample of all researchers.

2

“20 years as a curator, I was always forced to be certain about things I wasn’t certain about.” (C3)

2 Foundations

Coming from a background in design and computer science, I initially saw my research at the intersection of these two disciplines: the field of human-computer interaction (HCI). HCI studies the relationship between digital technology and its users in order to make computer-based systems more efficient, more ergonomic and more rewarding to use. Soon it became apparent, however, that my research questions require a deeper engagement with the subject matter of archives and collections, their digital instantiations and the epistemological role of digital tools and visualisation.

I found myself in the Digital Humanities, a field that describes the growing area of research in the humanities with and about digital technology, and builds on pioneering work that dates back to the 1940s (McCarty 1998). Digital Humanities is positioned at the intersection of humanities, computer science and - increasingly - design.

I begin this chapter with a brief history of Digital Humanities along with a review of the prevalent criticisms it faces and the challenges it creates, to which my research responds. I then continue with a discussion of the fundamental concepts of my thesis: digital collections, time and visualisations of data.

My review of collections consists of a look at their definitions according to the literature and the consequences and opportunities created at their coalescence with digital technology. Time is examined as an information-structuring model and we will see how representations of time have played an important role in the history of data visualisation. Next to a historic perspective, I will look at the specific claims of visualisation as tools for knowledge discovery.

Finally, I complement the literature review with original research on the manifestations of time in digital collections and a brief study on the understanding of digital collections among the collaborating curators. The chapter provides the foundation of my research, offering an overview of existing theory and an appraisal of the gaps in current knowledge and practices that my research aims to fill.

Digital Humanities

Humanities scholarship has adopted computational methods from the quantitative sciences to follow new research questions within its own field, but has also realised that available digital tools do not necessarily do justice to the nature of humanistic enquiry. In fact it seems necessary to consider the role of digital technology in humanistic knowledge production as a whole, to reconsider established concepts of data, interface and visualisation in a humanities context and subsequently to build new tools and methods for digital research in the humanities.

The Digital Humanities looks at the wider implications of the computational turn in humanities research. These include the introduction of computational methods to traditional humanities research, but also the contribution of humanities methods to computer science, the establishment of non-traditional forms of knowledge production and publishing as well as new synergies and collaborations between individuals and institutions. In this section I will outline the history and criticisms of Digital Humanities in order to provide the framework in which my research is positioned.

Digital Humanities has fundamentally changed the work of many humanities scholars. Albeit, not primarily through the introduction of digital technology per se - any avid researcher will be well versed in the use of computers - but through a shift in humanities research from *writing* to *making*. As Lunenfeld et al. state:

The advent of Digital Humanities implies a reinterpretation of the humanities as a generative enterprise: one in which students and faculty alike are making things [...] (2012, p.10)

The things that are being made by digital humanists - whether those things are visuals, software, objects or platforms - not only constitute their research output, but new modes of enquiry for the field of humanities.

Design contributes to the Digital Humanities in two ways: most visibly in the form of interaction design by creating visualisation tools, employing best-practices of interface design and developing ways to graphically represent humanities data. More importantly however on a more fundamental level, through the application of practice based research methods to address questions and challenges in Digital Humanities scholarship. Design, as the discipline of making things, is able to contribute its knowledge and methods of practice based enquiries to the humanities, where digital methods have introduced scholars to new modes of artefact-based knowledge production.

Origins

Kirschenbaum (2012) identifies an event in 2001 as the birth date of the term ‘Digital Humanities’. Discussing the title of a reader which would be published three years later, John Unsworth suggested “A Companion to Digital Humanities” (Schreibman et al. 2004) as a reaction to the publisher’s proposition “Companion to Digitized Humanities” (Kirschenbaum 2012, p.3). In changing “digitized” to “digital” Unsworth wanted to shift the emphasis from mere digitisation of resources towards a wider scope of the area. Digital Humanities aims to divert the attention from the field of computing to the field of humanities. While the humanities have seen a number of pioneering projects using early digital technology, these endeavours have largely been considered as humanities research adopting methods from the computer sciences, a convergence that goes by the name of ‘humanities computing’.

Figure 2.1 – Roberto Busa (1913-2011), a pioneer of Humanities Computing, pictured in front of his Index Thomisticus
image: Wikimedia (CC0)



Roberto Busa (Figure 2.1) - an Italian Jesuit priest - is generally regarded as the father of humanities computing (McCarty 1998; Boonstra et al. 2004; Hockey 2004). As part of his doctoral research on the concept of ‘presence’ in the writings of Thomas Aquinas, Busa manually produced an index of 10,000 sentences by Aquinas containing the preposition ‘in’ (Busa 1980). He realised that his lexicographical enquiry could also serve as a basis for other scholars and started imagining an ‘Index Thomisticus’, which would contain a concordance of all the words of Thomas Aquinas. With the help of IBM and thirty years of work he completed an index containing 11 million lemmatised words, all stored on punchcards and ready for further computational analysis.

Busa laid the foundation for a humanities scholarship that not only uses digital methods to contribute new knowledge, but at the same time makes these methods available for others. In fact, the endeavours

in adopting computing technology in humanities research that followed throughout the 1960s (Hockey 2004) were primarily concerned with building databases and converting text-based material into digital formats (Lunenfeld et al. 2012) in order to make these resources accessible and usable with digital research methods.

Around the same time, the term Quantitative History was popularised to describe the growing adoption of statistical methods from the social sciences for research in history and, with them, the concepts of *data* and *datasets* :

Quantitative history [...] required a new set of skills and techniques for historians. Most importantly, they had to incorporate the concept of the data set and data matrix into their practice. (Anderson 2007, p.246f)

While historians had incorporated quantitative evidence in their studies, the greater availability of computational processing in the form of mainframe computing¹ offered new ways of looking at historical evidence. By aggregating a large number of events and using statistical analysis historians were able to look at long-term patterns and developments in events which, on their own, would be insignificant.

1 In the 1960s, mainframe computers – centralised systems that filled entire rooms – were the first digital data processing systems widely available for corporate use.

Distant Reading

The study of large-scale patterns across humanistic sources comprises the practice of “distant reading”. Franco Moretti introduced this term in an article proposing new ways to study world literature (Moretti 2000) – it has subsequently also been used to describe research practices in the wider scope of the field of humanities and is a continuation of earlier endeavours in the area of Quantitative History.

Distant reading is the process of using computational methods to answer narrowly defined questions across a large body of works. In distant reading, Moretti writes,

distance [...] is a condition of knowledge: it allows you to focus on units that are much smaller or much larger than the text: devices, themes, tropes—or genres and systems. (2000, p.57)

Digital technology enables the processing of much larger bodies of works through distant reading, in opposition to close reading; studying a small sample of works or even a single work in detail. Close reading expects a more thorough engagement with limited source material – but focussing on a small number of works is not necessarily done by choice. Moretti’s argument builds around the dichotomy between claiming to be studying World literature, when in fact what is being studied is only ever a tiny sample of existing literature due to limits

in human interpretive capacities. Merely by selecting which works to study, a researcher inherently discards an enormous body of work without ever having looked at it.

In contrast, distant reading allows an individual researcher to consider an archive in its entirety:

By confronting scale, “distant reading” [...] – or, in our case, “distant listening” – reveals structures, patterns, and trends that are not discernable when the focus remains on just a handful of close readings of individual texts. [...] Distant listening facilitates whole corpus analysis and, potentially, the democratization of knowledge. Instead of privileging “human listening” (in which we necessarily have to limit ourselves to a tiny canon of works, probably a few hundred), distant listening is performed by a computer and can easily “listen to” thousands, if not millions, of works. (Presner 2015, p.23)

Digital technology enables distant reading but in some respects the digital also demands it. Human researchers can make an informed guess as to what will be studied and what not. The fundamental principle of an algorithm by which a computer has to operate, is a set of step-by-step instructions that require each step to be executable independent of the entire process; therefore an algorithm cannot consider a set of sources as a whole, but needs to examine each one individually and often exhaustively.²

2 Modern optimised search algorithms do not need to make an exhaustive search, which is impossible when the search space is very large. If the entire internet would be examined, a Google search would not return a result within seconds. Therefore the search space is indexed and search algorithms operate on a representation of a dataset.

Moretti did not expect his idea “to be particularly popular” (Moretti 2000) and opposition soon followed (Trumpener 2009; Culler 2010; Marche 2012). Humanistic knowledge production should not be systematised, Trumpener argues:

We are, first and foremost, highly trained readers [...] the unsystematic nature of our discipline is actually its salvation. (2009, p.171)

Endeavours in systematising humanistic enquiry are however not a recent phenomenon – see Boyd Davis (2016a) for an account of eighteenth century attitudes to mechanical knowledge production. The criticisms around close reading also echoed earlier concerns about the advent of Quantitative History. As Anderson writes

‘traditional’ historians expressed doubts about the new methods, challenging them as reductionist, brittle and not pertinent to the main goal of the historical narrative (Anderson 2007, p.257)

With the main modes of accessing collections turning digital and a growing use in born-digital collections – such as archives of Tweets – historians are increasingly forced to use distant reading:

As historians, we will need to reorient our approach to studying the past so that it does not involve reading every one of those thirty-one million lines of text. (Sternfeld 2014)

However, distant reading is only useful if it also reveals novel insights. Distant reading entails not only processing large amounts of data to ask humanities questions, but to do so in a useful manner:

The sheer size and scope of today's digital sources demand a level of methodological rigor that we are not yet accustomed to applying. (Sternfeld 2014)

3 Available at <https://books.google.com/ngrams> (accessed 12.01.2016)

4 An n-gram, in this context, is a contiguous sequence of n words, which can also be a single word (1-gram).

Google's NGram Viewer,³ for example, is an interface for distant reading that comes close to studying 'all' of World Literature in the way that Moretti envisaged. It is based on a corpus of literary sources printed between 1500 and 2008 and contains, for every year, an indexed count of n-grams.⁴ It enables users to trace – through graphical line plots over a horizontal dimension of time – the appearance of certain words and word combinations in the literature. However, the tool allows few insights into the cause of an apparent change in the use of certain words. NGram Viewer is a tool for distant reading, but largely fails in enabling humanistic enquiry into the history of the sources it represents. Sternberg continues to argue that

big data visualizations [...] wipe away any remnant of historical causality (2014)

a statement which certainly applies to Google NGram's line charts whose basis remains largely inaccessible. However, I argue, the existence of simplistic examples of implementations does not rule out the usefulness of distant reading, especially when it is combined with and in support of traditional methods of close reading. As Gibbs writes:

Any robust digital research methodology must allow the scholar to move easily between distant and close reading, between the bird's eye view and the ground level of the texts themselves. [...] Historical trends – or anomalies – might be revealed by data, but they need to be investigated in detail in order to avoid conclusions that rest on superficial evidence. (2011, p.76)

Distant reading must not replace, but can accompany close reading and if implemented in the manner that Gibbs describes, it can enable novel insights into digital collections whose size and structure requires the consideration and development of methods for distant reading.

Criticism

Many historians are critical of the methods inherited from the quantitative sciences and fear a discrediting of traditional humanities research.

In the context of the Digital Humanities it is the, arguably blind (Swierenga 1974; J. Drucker 2011b), application of digital methods which is most susceptible to criticism. Historians especially faced the limitations of off-the-shelf database software, which since their early days suffered from a range of challenges related to dates and the modelling of fuzzy data that immediately became apparent in historical research:

The most obvious limitation that historians encounter [...] is the date function, since most programs cannot interpret or query nineteenth-century or earlier dates and a nest of problems in date functionality accompany any attempt to use historical dating in these programs. (Thomas 2004, p.60).

5 The widely used programming language JavaScript still misinterprets 1st century dates; the Date() function automatically converts two-digit years to the twentieth century as visible in Figure 2.2.

Current software packages might have mostly done away with the problem of pre-twentieth century dates.⁵ Still, software is often seen as unfit for scholarly research (Borgman 2009). The problem of representing historic dates is not solved by simply extending the available timeframe. Dating historical events is a complex endeavour which may include conflicting evidence, missing data or varying degrees of certainty. Furthermore, the problem of digitally representing complex, fuzzy and uncertain information applies to all kind of data in the humanities. In order for digital tools to be usable for humanistic research, they have to be able to computationally model such data and, in addition, offer bespoke interfaces to query datasets and answer the kind of questions humanities researchers want to ask. According to Borgman (2009) it is the responsibility of humanities scholars to lead the development of new tools in line with the challenges and requirements of humanistic enquiry.

With the advances in graphical user interfaces and improved usability, it has become even easier for humanists to adopt existing software originally designed for engineers, statisticians or business analysts for their own purposes. However, Drucker argues that these sophisticated interfaces tend to hide software's underlying assumptions, which stem from outside the humanities discipline:

Such graphical tools are a kind of intellectual Trojan horse. (2011b, §1)

The limitations of available software tools for humanities research have been pointed out since the 1980s (Winchester 1980), but early ef-

forts in developing database tools specifically for humanities computing (Thaller 1987) found little adoption. Therefore, many of the available software tools and datasets still suffer from limitations related to fuzzy, uncertain or incomplete data and the handling of time and temporal information.

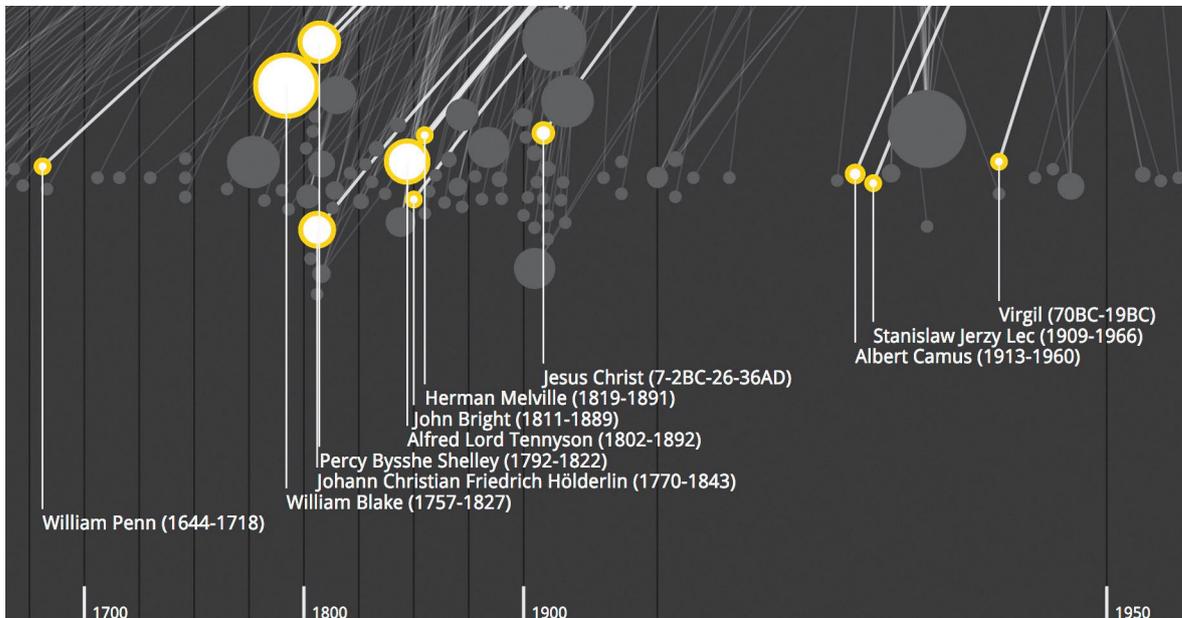


Figure 2.2 – Misinterpretation of two-digit dates in JavaScript means that – according to this prototype timeline – Jesus Christ and Virgil appear as contemporaries of Hölderlin and Camus.

In addition to the challenges posed by software implementations, the Digital Humanities have inherited many of the criticisms already raised about Quantitative History with regards to the conceptual shift to a numeric, mechanistic knowledge production that is feared to replace traditional, interpretative humanistic scholarship.

Carl Bridenbaugh, then president of the American Historical Association warned historians against “worship[ing] at the shrine of that Bitch-goddess, QUANTIFICATION” (Bridenbaugh 1963). The worry, shared by many historians, was that quantitative methods “may dehumanize history because of the emphasis on collectivities instead of individuals” (Swierenga 1974, p.1064). Few, on the other hand, were as dismissive to quantitative methods in history as Arthur Schlesinger, Jr., who announced that “almost all important questions are important precisely because they are *not* susceptible to quantitative answers” (Kousser 1980, p.434).

Bridenbaugh’s “Bitch-goddess” has become somewhat of a dictum to illustrate the aversion of historians and other humanists towards numerical methods (Bogue 1983; Thomas 2004; Anderson 2007), although many historians were not completely opposed to these quantitative, and later digital, methods. Including Bridenbaugh himself, whose main concern was a disconnection of future historians from their craft and their subjects of study. His worry was not directed at

quantitative methods per se, but at the danger of them being seen as a replacement, rather than an addition to qualitative methods.

Similar concerns are still raised about the Digital Humanities and digital collections in particular. It is suspected that digital methods may be seen as a replacement, rather than an addition to traditional humanities scholarship. Schüler-Springorum observes a decline in student's readiness to engage deeply with an archive on location as more and more collections are being digitised and made available online: "a specific intellectual space will get lost by providing online access to historical documents" (2015).

Museums have indeed been concerned that they will make themselves and their physical presences obsolete by providing improved online access to their collections (Lejeune 2009). So far museums have seen their visitor numbers increasing after collections were made available online (Goldman & Wadman 2002; Thomas & Carey 2005; Thomas & Crossman 2006), but the studies allow for little insights about the change in numbers based on the purpose of the visits. It might well be that the numbers of casual visitors have gone up, while digital collections may have reduced the need for scholars to visit a collection in person. Oxford's Beazley Archive, for example, experienced fewer visits after the complete catalogue was put online (C7). The Museum of Domestic Design and Architecture (MoDA) at Middlesex University, in contrast, observed an increase in usage of collection artefacts that previously saw little use, as a result of their collection being digitally accessible. A perceived competition between physical and digital collections stems from them being often seen as equivalent, rather than complementary. The digital is feared to replace the physical. To shed light on this apparent concurrence, I will take a closer look at how collections are generally perceived and defined in a later section.

In the case of existing digital collections, the growing calls for more sophisticated methods of modelling humanities data may have come too late; these datasets necessarily have been created with and suffer from the limitations of available software. The question is: is it possible - to some degree - to reverse-engineer the complexity of the 'original' humanistic data, the "thick description" (Geertz 1973) that is likely to be only insufficiently represented in a collections dataset?

Visualisation

Digital Humanities is a remarkably visual form of research. In contrast to traditional humanities research, visual materials act not merely as illustrative companions to written output, but visuals constitute the actual research output. Lunenfeld et. al see this development rooted in societal changes:

Digital Humanities necessarily partakes in and contributes to the “screen culture” of the 21st century [...]. What this means is that the visual becomes ever more fundamental to the Digital Humanities, in ways that complement, enhance, and sometimes are in tension with the textual. (2012, p.12)

6 ‘Mapping’ in the mathematical, not in the cartographic sense. See also the glossary entry on page 341.

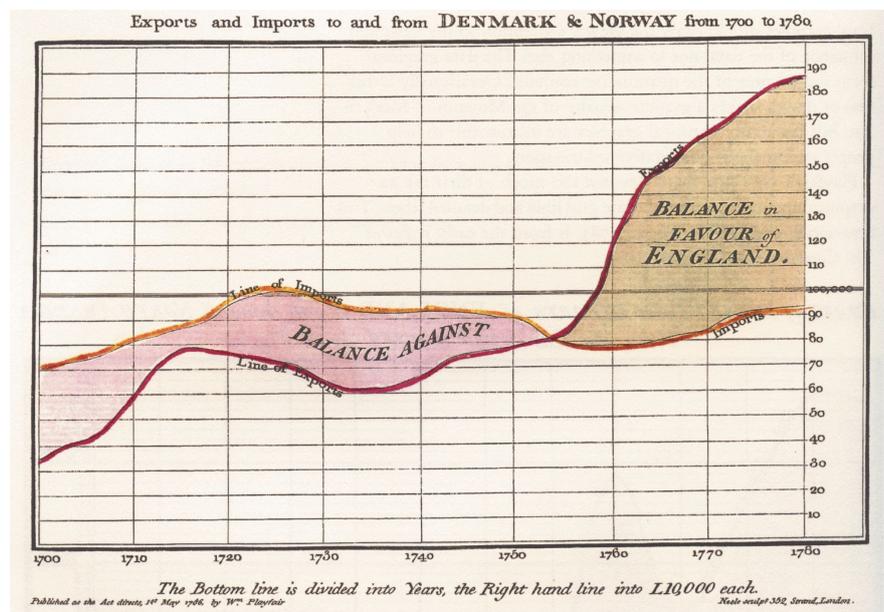
7 Bertin (1967/2010) offers a comprehensive, though somewhat idiosyncratic list of possible mappings.

Figure 2.3 – A line chart by William Playfair (1786) plotting exports and imports between England and Denmark and Norway.

image: Wikimedia Commons

However, the proliferation of screens is one that affects all aspects of academic, professional and social life. The visual enters the Digital Humanities, in my view, through the adoption of concepts of data and datasets and consequently the practice of data visualisation.

In this thesis I will focus on the method of analysing data through visualisation. Visualisation refers to the use of graphical elements to encode data: data visualisation. Essentially it describes a mapping⁶ from the input space of data to a graphical output space: ordinal, quantitative or categorical data attributes control the shape, size, colour, area and other visual properties of graphical elements.⁷



The development and study of visualisations that serve the purposes of data analysis, knowledge discovery and communication generally falls in the field of Information Visualisation (Gershon & Eick 1997; Card et al. 1999; Bederson & Shneiderman 2003; Ware 2004; Kerren et al. 2008). Data Visualisation is often seen as an equivalent term (Card

8 Friendly explains that “the term information visualization is generally applied to the visual representation of large-scale collections of non-numerical information” (2009) and that “we focus on the slightly narrower domain of data visualization, the science of visual representation of ‘data’, defined as information which has been abstracted in some schematic form” (ibid.).

9 Gershon and Eick distinguish scientific visualisation “which focuses on data” (1997) and information visualisation that “focuses on information, which is often abstract” (ibid.).

10 Infographics do not enjoy the best reputation within the information visualisation community. Examples of infographics that are (arguably) badly designed, depart from best-practice approaches or misrepresent data are widespread, which leads many to draw one-sided conclusions about infographics in general.

et al. 1999; Bederson & Shneiderman 2003; Kosara 2013; Bailey & Pregill 2014) and sometimes treated as a sub-discipline of Information Visualisation (Friendly 2009).⁸ Occasionally, Information Visualisation and Data visualisation are described as separate, but related fields (Gershon & Eick 1997).⁹

When data visualisation is treated as a specialisation within information visualisation, it is generally paired with scientific visualisation (Friendly 2009) – for example 3D visualisations of natural phenomena – and Infographics (Kosara 2010; Gelman & Unwin 2013) – explanatory visualisations accompanied by illustrations. Infographics are often contrasted with data visualisations and the benefits of each, as well as the value of differentiating at all, provide ample food for theorisation (Few 2011; Cairo 2012; Kosara 2013; Wickham 2013). For some (Gelman & Unwin 2013), all data visualisations that depart from traditional statistical charts count as infographics, along with their negative connotations.¹⁰ For others (Kosara 2010; Wickham 2013) – including myself – the difference between data visualisations and infographics lies not in their adherence to written or unwritten rules of information visualisations, but in their ability to generalise and – consequently – to be generated.

Visualization is (largely) automatic, infographics are hand-crafted. Neither are objective, and both require hand-tuning and understanding to get right. (Kosara 2010)

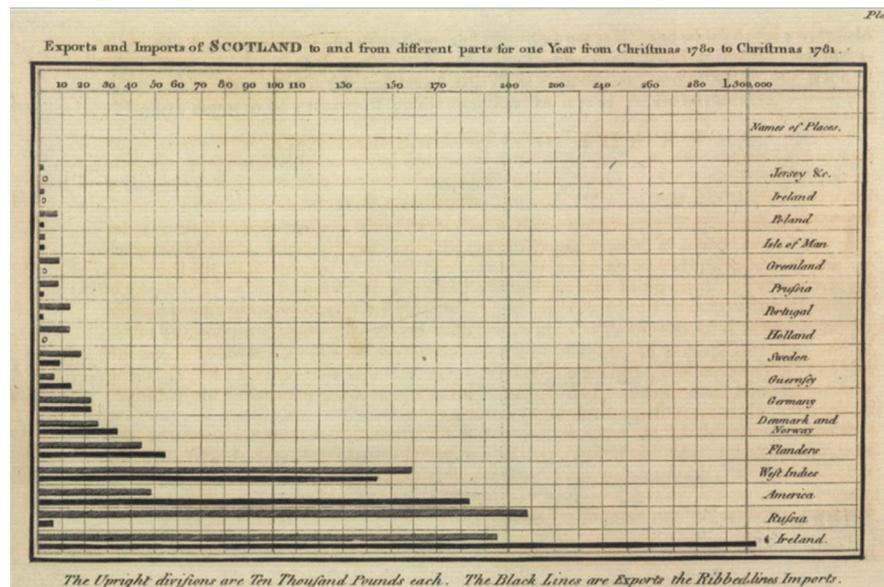
The automatic view on visualisation envisages digitally generated diagrams, which are also the focus of this thesis. However, the next section will introduce historic examples of hand-crafted visualisations, which have, to an extent, been produced mechanically, an aspect that also constitutes their main innovation.

History of (Time-Wise) Visualisation

Graphical representations of - what we now would call - data are scattered throughout history, too numerous to pay them the attention here they deserve.¹¹ By the end of the 18th century, data visualisation - in the strict sense of generalisable and mechanically achievable charts - became mainstream through the publishing of William Playfair's (1759-1823) *Commercial and Political Atlas* in 1786 (Playfair 2005). Geographical atlases were already available at that time and had been since the 16th century, but Playfair's book was not mapping land, it was mapping economic data.

11 For a concise history of data visualisation see Friendly (2006).

Figure 2.4 – Exports and imports of Scotland plotted in one of the first known bar charts by William Playfair.
image: Wikimedia Commons



To do so, Playfair invented the line chart (Figure 2.3) as well as the bar chart (Figure 2.4), although he credits a fellow Englishman for providing the inspiration for the latter (Wainer 2014). The bar chart actually originated from a compromise - Playfair lacked sufficient data to draw a continuous line (Wainer 2005). Line charts were also not completely novel; Christiaan Huygens' (1629-1695) mortality chart is regarded as the earliest example (Wainer 2005). Huygens' graph however depicts a mathematical function. Playfair's graphs on the other hand are graphical mappings from empirical data and hence 'true' data visualisations.

His book did not initially find a publisher in England, so Playfair tried his luck in France. When he sent a copy to Louis XVI in 1781, the king had never seen such diagrams, yet Playfair notes

he at once understood the charts and was highly pleased. He said they spoke all languages and were very clear and easily understood. (Wainer 2014)

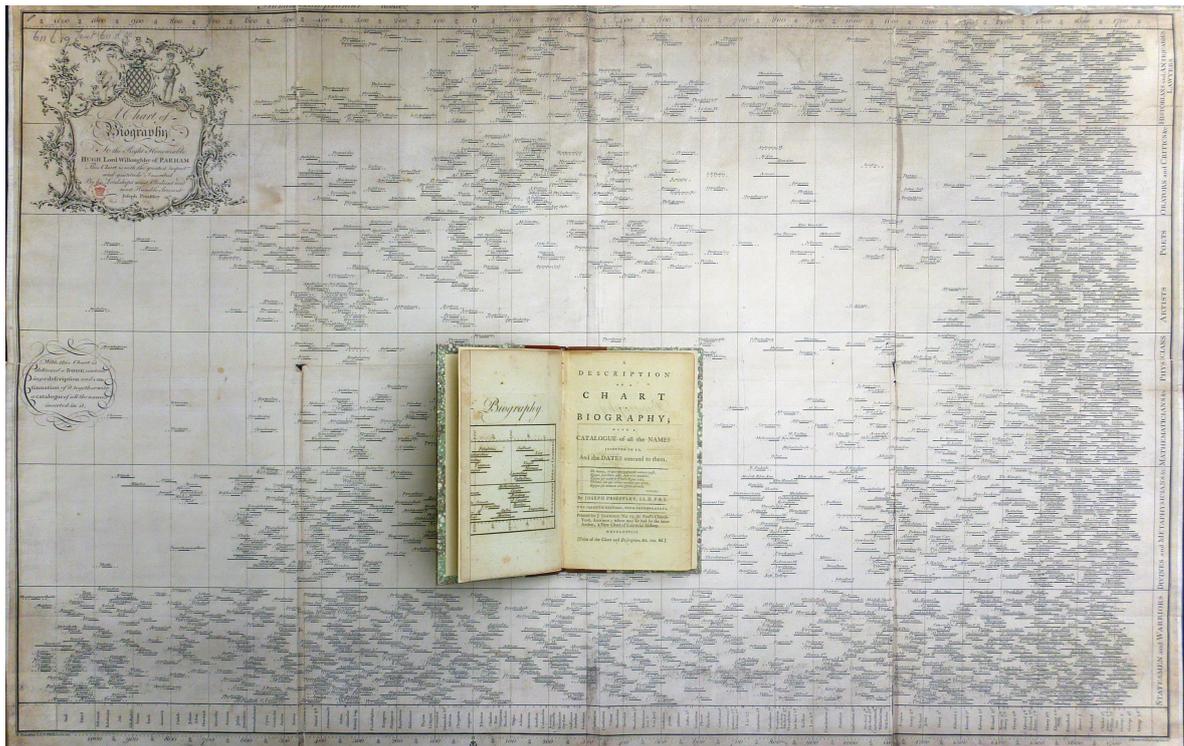


Figure 2.5 – Priestley's *Chart of Biography* (1764) maps the lives of 2,000 individuals on an arithmetic timeline.

image: Stephen Boyd Davis/
British Library

- 12 See Rosenberg (2007) and Boyd Davis (2011) for a more thorough discussion of Priestley's Chart.

Others soon shared the king's insight. Playfair's Atlas was published, became known and his charts continue to be widely adapted.

The Englishman who was Playfair's inspiration for the bar chart is Joseph Priestley (1733-1804), remembered more for his discovery of oxygen than for his pioneering work on visualisations. Playfair was referring to Priestley's *Chart of Biography* (1764, Figure 2.5), one of the first visual timelines published in 1764 and graphing the lives of about two thousand important individuals.¹² Priestley drew lines of different lengths, representing the individual lifetimes of those he mapped, an innovation at that time (Boyd Davis 2011) which Playfair adopted by expressing quantities as lengths of bars.

The history of data visualisation is, to a great extent, also a history of time-wise visualisations. Playfair's charts visualised economical data on a (horizontal) dimension of time and drawn diagrams that visualise the movement of celestial bodies through space and time were used as early as the 10th century (Friendly 2006). Rosenberg and Grafton (2010) offer a comprehensive and richly illustrated history of timeline visualisations; here I will only focus on some key examples to which I will return later on in this thesis.

I have already mentioned Priestley's *Chart of Biography* as Playfair's inspiration. At the same time, it is one of the first true arithmetic timelines that graphically maps data according to a linear time axis, which runs horizontally and is divided equally into centuries. The principle of using equal units of space to represent equal number of years was introduced only eleven years prior to Priestley's chart.

Figure 2.6 – Barbeau-Dubourg’s 16.5m long *Carte Chronographique* mounted in a device that lets viewers scroll ‘through time’ by turning the handles.

image: Princeton University Library



Jacques Barbeau-Dubourg (1709-1779) produced a *Carte Chronographique* (Figure 2.6) of the “main events of every century, since the Creation of the world up until the present”¹³ (Barbeau-Dubourg 1753), a timeframe that covered 16.5 metres of paper, each year a quarter of a centimetre wide.¹⁴ On it the names of notable individuals as well as summaries of important events are positioned horizontally according to their time and grouped vertically by their country of origin. A category with unassignable or general events is positioned at the bottom of the chart and labelled “memorable events” (“événements mémorables”) (Barbeau-Dubourg 1753).

Barbeau-Dubourg’s and Priestley’s visualisations represent a significant step towards abstract visual representations of data. Earlier chronologies adhered to a text-based tabular layout while other examples of early timelines, such as the work of Girolamo Martignoni (died ca. 1743, see Boyd Davis 2016b), borrowed their visual language from geographic maps (Figure 2.9). A piece by Jean-Louis Barbeau de la Bruyère (1710-1781) appears like a cross-over between a tabular layout and a geographic map. Through his *Mappe-Monde Historique* (Barbeau de la Bruyère 1750a, Figure 2.7) he intended to offer a complete view of the history and geography of the known world.¹⁵ The *Mappe-Monde* borrows from a tabular layout, with geographical divisions marked at the head of the chart, forming a column for each country,¹⁶ with empires that extend across several columns set in colour.

The ability of Priestley and Barbeau-Dubourg to graphically map data according to a coordinate system required a conceptual shift in the understanding of space and data that was offered by Descartes (1596-1650) through his proposition that anything that can be expressed in number can be represented graphically (1641/1996).

13 “des principaux événements de chaque Siècle, depuis la Création du Monde jusqu’à présent” [sic] (Barbeau-Dubourg 1753)

14 See Furguson (1991) and Schmidt-Burkhardt (2011) for more information on Barbeau-Dubourg’s *Carte Chronographique*.

15 See Boyd Davis (2015a) for a more thorough discussion of this chart as well as de la Bruyère’s own *Explication* (1750a).

16 “Ces Divisions forment ainsi pour chaque Pays une Colonne” (Barbeau de la Bruyère 1750a)

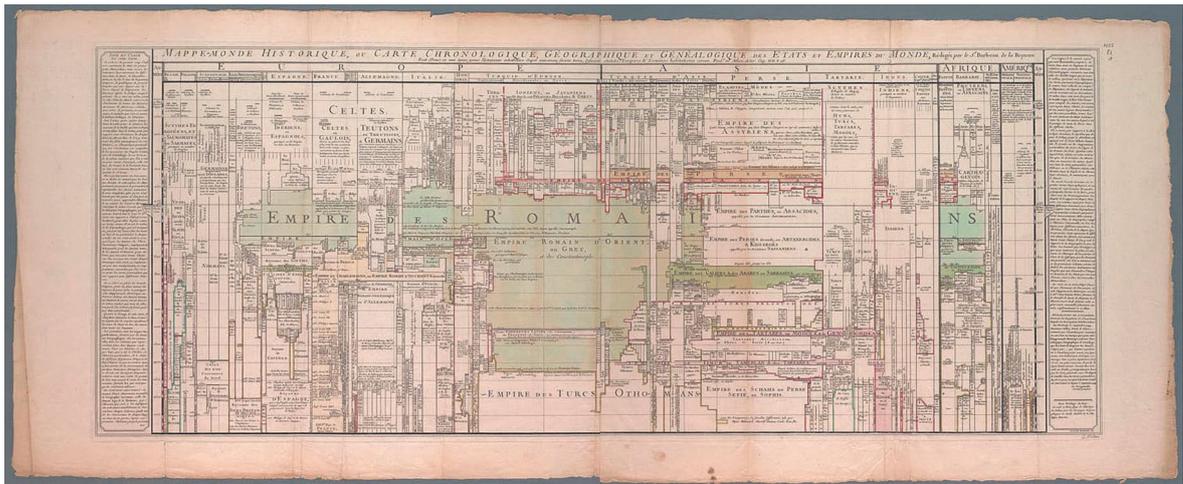


Figure 2.7 – De La Bruyère's *Mappe-Monde Historique* (1750a).

image: Koninklijke
Bibliotheek NL

Another fundamental innovation for the development of arithmetical timelines was the advancement of a numerical model of time famously promoted by Isaac Newton (1642 – 1726/7). Newton considered time to be an absolute, uniform frame of reference where events could be ‘located’ independent of other events or external perceivers. Time, according to Newton, is “absolute, true, and mathematical” (Newton, 1687), a fundamental quantity like length or mass, which can be measured and expressed in a manner that may be universally agreed upon.

While Priestley was not the first to draft a true arithmetic timeline, he was original in graphically representing duration as a line:

THUS the abstract idea of TIME, though it be not the object of any of our senses, and no image can properly be made of it, yet because it has a relation to quantity, and we can say a greater or less space of time, it admits of a natural and easy representation in our minds by the idea of a measurable space, and particularly that of a LINE; which, like time, may be extended in length, without giving any idea of breadth or thickness. (Priestley 1764, p.6)

It is worth noting, however, that the understanding of time as a quantity along with the affordance of quantities to be represented as space that Priestley describes, are a result of the conceptual shift in thinking advanced by Newton and Descartes; the “natural and easy” representation required a significant intellectual effort and decades of familiarisation.

In the literature, arguments for the merits of particular graphical representation that rest on the assumption of one being more “natural” than the other are widespread; we will see more examples later on and the claimed ‘naturalness’ is rarely free of previously established and unconsciously acquired paradigms.

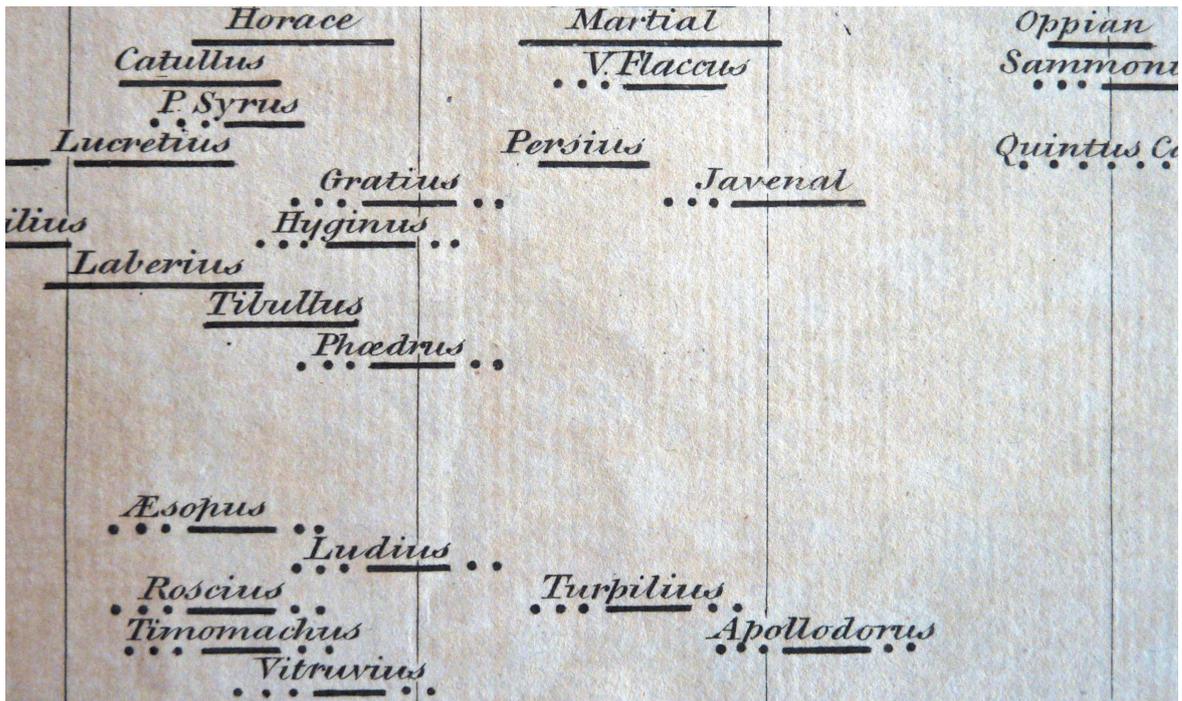


Figure 2.8 – Priestley's notation of uncertainty in dates expressed as variable numbers of dots.

image: Stephen Boyd Davis, taken at Chetham's Library Manchester, 6 July 2009

Priestley's decision to map duration to length is one such innovation that stood the test of time; his graphic taxonomy is embedded in many modern visual timelines. While the effectiveness of the line was widely recognised, its problems that Priestley identified as well have found far less consideration later on. A line has a clear beginning and ending; historical dates however often do not. Having spotted the inconsistency, Priestley offered a solution (Figure 2.8):

In this case the compiler must content himself with placing his line as near as he can conjecture from history where his true place was, leaving marks to express the uncertainty there is attending it. The method I have used in this chart is to express certainty by a full line, and what is uncertain by dots or a broken line [...] (1764, p.11)

Priestley, Barbeau-Dubourg and de la Bruyère each complemented their visualisations with detailed descriptions of the rationale for their design decisions, knowledge they acquired while making their charts as well as the insights they gained from inspecting them (Barbeau de la Bruyère 1750a; Barbeau-Dubourg 1753; Priestley 1764). One could see these pioneering works as a form of Critical Making (Ratto 2011), as their makers not only had to design new graphical formats, they had to develop a new visual rhetoric and, most importantly, explain and reflect on their ideas, processes, and rationales.

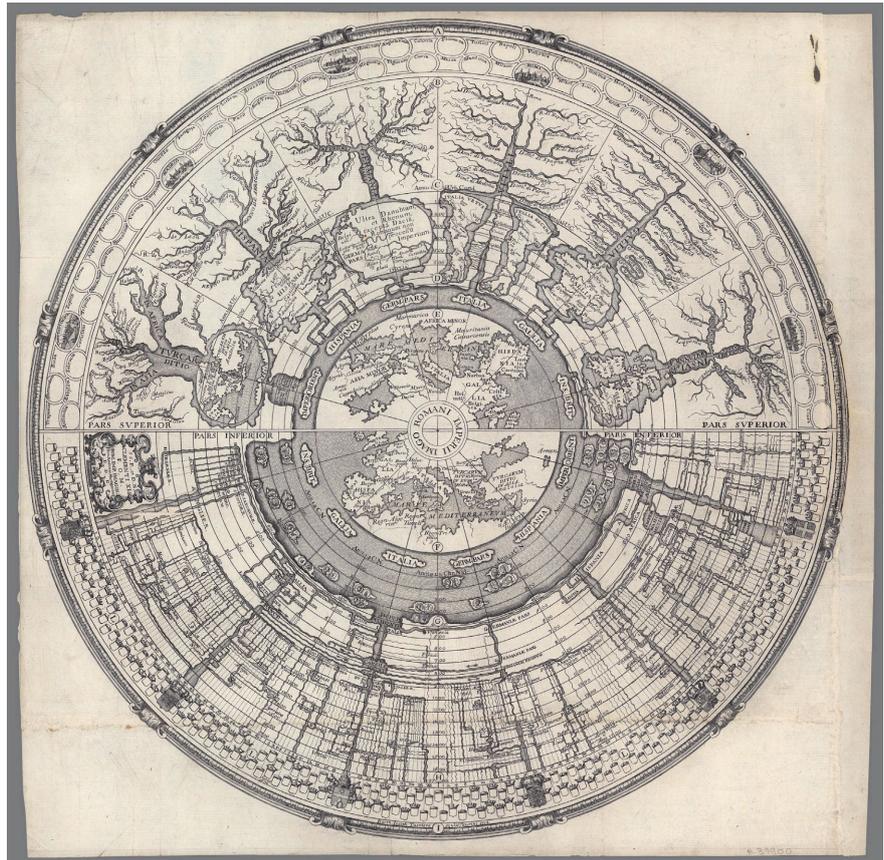
Today, it is rare for designers to have to defend and justify their decisions in relation to visual representations of time, which is maybe why timelines are generally regarded as simple, even as “a bit of banal

medium” (Behrendt 2011). This brief look at their history however demonstrates that they can be, and have been, much more than that; effective tools for visual analysis, or in the words of Joseph Priestley:

What words would do but very imperfectly, and in a long time, this method effects in the compleatest [sic] manner possible, and almost at a single glance [...] (1764, p.9)

Figure 2.9 – A map of time by Girolamo Andrea Martignoni that borrows cartographic visual language.

image: David Rumsey Historical Map Collection

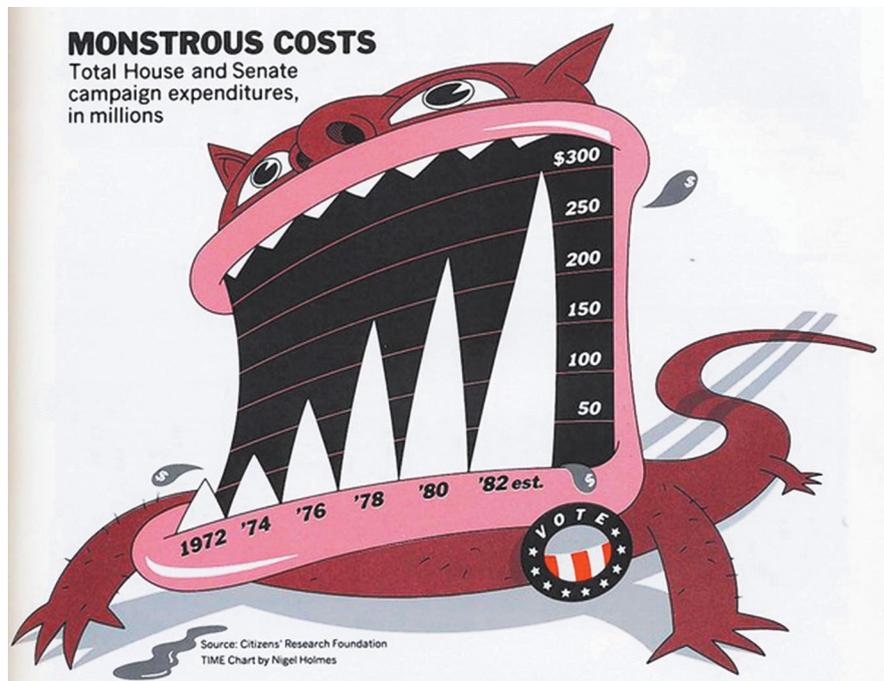


Visualisation for Sense-Making

Priestley’s quote that concludes the previous section leads me to address the question of what purposes visualisation serves. In the booklet that accompanies his chart, he dedicates his work “to the youth [...]; showing them what names will most frequently attract their attention, and how they stand related in point of time to one another” (Priestley 1764, p.5), underlining its educational qualities. Yet he also states that only through the use of visualisation, he was able to “see [...] the relation of these lives to one another in any period” (Priestley 1764, p.10), suggesting that he drafted the visualisation in order to understand for himself. These observations relate to DeFanti et al. (1989) who describe visualisations as either “a tool for communication and teaching” or “a tool for discovery and understanding”. Similar discrim-

inations appear throughout the visualisation literature. My emphasis is on the second category of analytical visualisation, although there is often a wide area of overlap.

Figure 2.10 – Nigel Holmes' diagrams of US House and Senate expenditures exhibits a large proportion of so-called chartjunk that is theorised to impede people's ability to comprehend the diagram (Tufte 2001). Nevertheless, experimental evidence found this chart to perform better than a version stripped off its "visual embellishment" (Bateman et al. 2010).
image: Nigel Holmes



Few (2013) separates visualisations for the purpose of sense-making from visualisations for communication, while Kosara (2007) identifies "very technical, analysis-oriented work on the one, and artistic pieces on the other hand". Tufte (2001) even coined the term "chartjunk" to describe elements of a visualisation which, according to him, may be justified as artistic additions, but are in his view purely decorative and unnecessary, even counterproductive for the readability of a visualisation (see Figure 2.10). Tufte's radical dismissal of rhetorical elements in visualisations is debated however and experimental investigations found that so-called chartjunk may in fact be helpful for the reading of a visualisation (Craft & Cairns 2005; Inbar et al. 2007; Bateman et al. 2010; Hullman et al. 2011; Moere et al. 2012; Borkin et al. 2013).

While there is largely an agreement that visualisations can enable understanding, how they might do so is still not fully understood. A reoccurring claim is that by visualising data, instead of for example displaying it in a table, one can take advantage of the human visual system's abilities to recognise patterns immediately (McCormick et al. 1987; Gershon & Eick 1997; Van Dam et al. 2000; Ware 2004). In situations where large amounts of measurements are visualised in scatter plots, this explanation sounds reasonable. But there is more to visualisation than just seeing patterns. A different process is at play when, for example, two numbers are represented as lines of different lengths or quantities are encoded by colour ranges.

We may take for granted that visuals can enable understanding - we say “I see” to mean “I understand” - and to suggest that visualisation may be beneficial for our understanding of large datasets and complex issues is not a daring statement. During the first half of the twentieth century the credo was different. Arnheim, in his seminal work on visual thinking, quotes Karl Bühler’s (1879-1963) writing - “In principle, any subject can be thought and meant completely and distinctly without any help of imagery” (Arnheim 1969, p.100) - and the American psychologist Robert Woodworth (1869-1962), who goes as far as stating “the more effective the thinking process is at any moment, the more likely is imageless thought to be detected” (Arnheim 1969, p.100). In this context, it is no surprise that Arnheim sees his own theory - to suggest that the process of thinking is very close to seeing - as “an obvious contradiction. How can there be intelligence in perception?” (1969, p.13).

Bertin argues in a functional way for the advantages of graphics over other forms of representation (1967/2010). He contrasts the visual system in the form of graphics to the auditory. The auditory also includes written transcriptions of language, music or mathematic as they are, in his view, merely ways of capturing what is essentially auditory. Graphical notation uses the two dimensional spatial plane while auditory systems operate in one dimensional time. This argument has to be used cautiously however, as an increase in dimensionality does not necessarily make a visualisation more understandable. In fact the usefulness of visualisation in a lot of scientific contexts stems from reducing high dimensional datasets in two- or three-dimensional projections.

Bertin argues for the efficiency of visualisations:

If, in order to obtain a correct and complete answer to a given question, all other things being equal, one construction requires a shorter period of perception than another construction, we can say that it is more efficient for this question (1967/2010, p.9)

In his line of reason, a tabular representation, for example, would allow the same insights as a plot of the data, just with differences in speed. Is visualisation then a short-cut, but not a unique path to understanding? Larkin and Simon (1987) offer evidential support for this hypothesis. Farquhar and Farquhar on the other hand are convinced that some conclusions can only be drawn from a graphical representation:

The graphical method has considerable superiority for the exposition of statistical facts over the tabular. [...] the popular mind is as

incapable of drawing any useful lessons from [the tabular method] as of extracting sunbeams from cucumbers (1891, p.55).

For Don Norman, “the power of the unaided mind is highly overrated” (1993). He introduces the concept of “cognitive artefacts” (Norman 1991) that enhance human abilities: like a megaphone amplifies voice, pen and paper can ‘amplify’ thought. Visualisations, too, are a type of cognitive artefact.

Norman gives the example of a visual mapping of values to the sizes of graphical elements as a “superior form of representation” (1991) to arabic numerals. According to Norman, the superiority of the visualisation, in this case, stems from the ‘natural’ relationship between the “perceptual representation” (ibid.) of the visualisation to the numerical values:

The “naturalness” of a mapping is related to the directness of the mapping, where directness can be measured by the complexity of the relationship between representation and value (Norman 1991, p.28)

The problem with this line of reasoning is that it rests on the intuition of what one perceives as ‘natural’, even though Norman tries to circumvent this issue by suggesting that one is able to objectively measure how closely related a value and its representation is. Even if we could, according to Scaife and Rogers

we cannot simply assume a privileged relationship between a graphical representation of a system [...] and someone’s understanding or ability to reason about it. (1996, p.201)

Scaife and Rogers explored and tested how diagrams permit “computational offloading” (1996, p.188); similar to Norman’s concept of cognitive artefacts, visualisations afford “external cognition” (Scaife & Rogers 1996) by relieving users from having to picture information mentally and instead allow them to focus on studying information by interacting with a visual artefact. While Scaife and Rogers, following a thorough review of existing literature, conclude that little is known still on why visualisations work, they do observe that interacting with a visualisation - be it through manipulating a digital representation or through pen and paper - benefits users’ understanding significantly.

The, admittedly, unspectacular conclusion is that despite the wide range of possible theories and speculations, we do not know for certain why and how visualisations benefit understanding. Largely, I argue, because we also know very little still about the process of understanding itself.

The functionalist view of the mind argues that cognition is independent of the world and in principle realisable by a computer (Block 1980; Fodor 1987; Putnam 1988). This was the dominant philosophy of mind throughout the 1980s, when digital data visualisation was on the rise and its ability to assist in understanding was first theorised (Card et al. 1999).

Functionalist explanations of how visualisations enable understanding focus on the idea of the mind as a “pattern-seeking machine” (Popova 2013) that operates on internal representations. Therefore, they have to argue for a qualitative difference in the internal processing of visualisations as opposed to other representations of data, for example by “exploiting people’s natural strengths in rapid visual pattern recognition” (Gershon & Eick 1997). As I have argued, claiming an advantage based on ‘natural’ human abilities can be problematic.

More recently, an ‘embodied’ theory of mind is emerging which argues that consciousness arises from interacting with the world (Varela et al. 1992; O’Regan & Noë 2001; Noë 2004). Understanding stems not from operating on internal representations, but from acting and interacting with the world.¹⁷

¹⁷ This involves both manipulating the world as well as ‘passive’ observation. What leads to understanding is not only the ‘actual’ interaction, but the knowledge of how things in the world change when being acted on.

Both the concepts of “external cognition” (Scaife & Rogers 1996) and “cognitive artefacts” (Norman 1991), and the improved understanding that arises from interacting with and manipulating a diagram that Scaife and Rogers observed, gives support to the embodied theory of mind and provides an explanation of how visualisations might support sense-making: not by offering universal functional advantage over other forms of representations, but by serving as tools that can be manipulated, both manually as well as by observation, and aid and extend our inherent and acquired means for cognition.

Collections

When I refer to collections within this thesis, I envisage the holdings of museums, archives, libraries, galleries, universities and cultural institutions more broadly, as well as the specific collections of the institutions I worked with. To define a collection is however far from straightforward, even with this pragmatic concept, which leaves the definition of a collection to the institution that owns it; the internal criteria of what comprises an institution's collection may vary considerably over time:

This system [...] [contains] everything that's catalogued, but not necessarily in the main collection. So for example this [artefact] is not considered - at the moment - a part of the collection. (C10)

Archives and collections are diverse and dynamic - despite the common image of them being a static conservation of the past - and as a result there is not a unanimous definition of collections in the literature.

Lee (2005) studies the concept of library collections from the perspective of its users. Most of the findings translate to collections in general; given that libraries increasingly contain and catalogue a variety of items and media, it ceases to make sense to hold up a strict discrimination between libraries on one hand and collections and archives on the other. Digitisation further blurs the boundaries, as Curall et al. point out: "the digital world in some senses equates to a library in that items enjoy an independence that is analogous to a book" (2005).¹⁸

¹⁸ Lynch on the other hand argues that "digital collections and digital libraries aren't the same thing" (2002), which illustrates the complexity of the issue and the disputes that surround it.

Lee finds that "as an entity, the library collection seemed to be extremely vague in the users' minds" (Lee 2005, p.71). Some participants of the study equated it with certain physical items housed in a building (see Figure 2.11). Especially in the example of libraries and museums, the strong association between an institution's collection and its physical building may strengthen this view - even though many museums and libraries store their collections in various off-site locations. And of course, the multiplicity of types of items in a collection - be they objects, files, references or concepts - may require various storage 'locations': physical, virtual, juridical, etc.

Despite users' apparent difficulty to formulate what a collection is, most of them could easily describe what they expect from it. Key factors include selectivity - the fact that a collection is curated - but also flexibility - that a collection can be shaped to one's needs. Finally, a collection is required to be "readily available":

if something were not readily available, it could not be considered part of the collection (Lee 2005, p.72)

Currall et al. (2005) assembled a comprehensive review of institutional definitions of collections. For example, the description by JISC IE that states “a collection is any discrete aggregation of one or more items of content, but will often take the form of a database of one kind or another”:¹⁹ Slightly more useful is the definition in the Canadian Rules for Archival Description, despite its restriction on one specific type of collected item:

¹⁹ see <http://www.ukoln.ac.uk/distributed-systems/jisc-ie/arch/collection-description/> (accessed 27.12.2015)

An artificial accumulation of documents of any provenance brought together on the basis of some common characteristic, e.g. way of acquisition, subject, language, medium, type of document, name of collector, which may be treated for descriptive purposes as a unit under a common title.²⁰

²⁰ see https://www.collectionscanada.gc.ca/archivianet/020123/0201230103_e.html (accessed 11.12.2015)

The view of the authors is that collections can not be unambiguously defined as they are always a product of the people that maintain, produce and govern them:

Our position is simply this: it is human beings, with their language and intentions, who determine the categories and, thus, the collections into which things are placed. (Currall et al. 2005, p.135)

Currall et al. criticise that the people responsible for the collections are often oblivious of this:

Although curators are aware that objects within their collection or collections have been selected for inclusion, it rarely occurs to them that the process of selection is both dynamic and constructed. (Currall et al. 2005, p.139)

My own research however could not fully confirm this observation. Many curators and archivists I talked to were conscious of the subjective role they and their institution are playing in shaping a collection. However, often it was suspicious patterns in my visualisations - visual evidence of the influence of individuals on the collection - that steered our conversation to address the subjective biases they and their institution impose on a collection. These findings will be discussed later on.



Figure 2.11 – The Smithsonian National Museum of Natural History bird collection along with some of the human beings that determine the categorisation of each bird.

image: Chip Clark (CC BY-NC-SA 2.0)

Considerations of a curator’s influence on collections have not found wide adoption in specialised scholarly literature, ignoring the power that individuals may have in shaping a collection’s preservation criteria. Hedstrom observes that

debates about appraisal occur along a continuum ranging from a Jenkinsonian ^[21] approach, which takes as its point of departure assumptions about the neutrality and impartiality of records and the objectivity of the archivist, to a more socio-technical approach. (2002, p.34)

21 Sir Hilary Jenkinson (1882-1961), Deputy Keeper of the Public Record Office, defended the view that archive records are “impartial” (Jenkinson 1922), a neutral and objective representation of the past.

22 “Tangibility”, in the definitions that use the term, is used in a strict sense and refers to sets of physical objects (Lee 2000).

Many classical views on collections define it with regards to “tangibility”, “ownership” and its “user community” (Lee 2000). We can safely discard “tangibility” from a definition of a collection that is compatible with the digital age.²² “Ownership” does not hold up either, with institutions such as the MoMA “acquiring” the @-sign (Antonelli 2015); a symbolic entity that is part of the public realm and hence cannot be owned, but apparently nevertheless can be included in a museum collection. A more common example is items that are on loan and often enter a digital cataloguing system along with an institution’s own items. It is not unusual that these records remain in the digital collection even after a physical artefact has been returned.

Figure 2.12 – One of more than 230,000 user curated collections on the Rijksmuseum’s website, exploring the topic of bubbles and balloons.

image: screenshot of rijksmuseum.nl, collection by Yvonne Klop



door engeltjes&bengeltjes by angels&beagles



Het koffiehuis Quincampoix te...
Anonymous



Het koffiehuis Quincampoix te...
Anonymous



Kinderen spelend met een...
Jan Luyken

What remains from these definitions is the “user community”. I argue that the user community includes both internal and external scholars, researchers and individuals and that they both exert power on what a collection is. Currently the main power still lies in the

hands of the curators and the ones governing preservation criteria. With increased digital access to collections data however, the power structures are shifting. Since the 2013 redesign of the Rijksmuseum website, users can create their own collections from what is available in the digital catalogue (Figure 2.12). So far, more than 230,000 custom collections have been curated, according to taxonomies that often are far remote from scholarly art history: Bubbles, Breast Feeding, Pornography, etc.²³

23 See all of them at <https://www.rijksmuseum.nl/en/rijksstudio> (accessed 06.01.2016)

In conclusion, a collection, it appears, is primarily defined through the people that use and maintain it. Recent definitions and discussions display an increased awareness of the influence of individuals on the shape of archives and collections, and depart from supposedly objective and neutral standpoints.

24 "New records and images are being added every week" (The British Museum 2013) to the database of the British Museum, which is described as follows: "The database is an inventory of the Museum's collection and aims to record what is known about it." (ibid.)

Collections As Data

Libraries, archives and museums have been among the early adopters of digital technology for cataloguing the contents of their collections (Chenhall 1975). GRIPHOS (Heller 2013), one of the earliest digital cataloguing systems was released in 1967 and some museums use it even today (Williams 2010).

Catalogues were primarily drafted as a way of making inventory, a way of answering: what do we actually have? A non-trivial question, not only for holders of large collections such as the British Museum, whose catalogue currently contains more than two million items and is estimated to grow to more than six million by the time it will be completed.²⁴

25 This distinction implies that a digital collection represents an archive consisting of physical items. Most collections I worked with, in fact, do collect objects – or photographic representations thereof. A collection could also consist exclusively of digital items. It would not however change the present argument. Rhizome, for example, who keep a collection of digital artworks, still maintain a database catalogue, while the works are stored separately (often as virtual machines, when their code relies on discontinued hardware).

A catalogue functions like an index: an efficient way of searching for a specific item within a collection by imposing a certain structure. Digital catalogues have been adopted because they allow for searching through an entire archive almost instantly and without practical limits in its size. Notably, also without limits in the size of the catalogue itself. While physical index cards can only hold a restricted set of data, such as, for example, the name of an object and its location in a storage, their digital counterparts can store nearly unlimited amounts of (meta) data: the date it was made or acquired, the name of its creator or inventor, its material or colour – in short – anything which the authors of a catalogue consider meaningful to record about anything which might be in a collection.

Thanks to this rich metadata, the digital catalogue becomes more than just a list of things, it becomes a resource in its own right. It turns into a digital collection, which is not a mere reproduction of the 'original' analogue collection. Neither is it imperatively, by any standard, better or inferior to a physical archive, but a resource that can be used both in conjunction as well as independent from it.²⁵

The digital catalogue not only makes working with a collection more efficient; it enables completely new ways of interacting with an archive. I suggest to look at digital collections primarily as ‘data’; digitally stored information which can be accessed, shared, manipulated, enriched, reinterpreted and – most importantly for my research – visualised.

26 Finding aids in archives contain written descriptions of the overall content of a collection, but also specific data on the provenance of individual items, keywords, etc.

I define digital collections as digital versions of indexing catalogues and finding aids²⁶ and not as digital substitutes of physical collections. This is on one hand to avoid confusion about the terminology around digital collections and the process of digitising a collection. Digitising can mean making digital copies of cultural artefacts, through photographic reproduction or 3D scanning. Often it simply refers to converting index cards into a machine-readable format. Looking at digital collections primarily as catalogues, that might or might not be enriched with digital reproductions, serves as a lowest common denominator to ensure that methods and findings can be applied to a wide range of digital collections.

Europeana (Figure 2.13), the European aggregator of digital collections, is more strict in its definition:

Europeana focuses on giving access to the digital version of physical objects held in institutions rather than just abstract digital information about these objects. Therefore such catalogue descriptions are not considered as digital objects in their own right within the context of Europeana. This definition includes digitised catalogue cards as well, as they function as finding aids and not as objects. (Europeana 2014)

This, however, results in the platform itself exercising a bias on the represented cultural artefacts, excluding collections that do not comprise physical items or that do not lend themselves to digital reproduction. For example, many 20th century artefacts may not be photographically reproduced due to copyright restrictions, leaving data from that period underrepresented (Gomez & Keller 2015).

27 Some might find this transition easier to make if not the cultural artefact, but the index card is ‘reduced’ to data: a cultural artefact is unique – data, in contrast, is generalisable and reproducible.

The second reason for wanting to differentiate between a collection and the data *about* a collection is to avoid a competition between digital and non-digital collections, and the respective methods for studying them; many historians feared that traditional methods would waste away with the advent of Quantitative History.

Using digital methods for historic research requires scholars to rethink archives as data.²⁷ Data, however, is an ambiguous term (Machlup 1983). When it is understood in a realist sense as given facts, it is unfit for use in the humanities, argues Drucker:

To begin, the concept of data as a given has to be rethought through a humanistic lens and characterized as *capta*, taken and constructed (2011b)

Drucker argues that the concepts and tools – such as *data* and *data visualisation* – that have been borrowed from the natural and social sciences “carry with them assumptions of knowledge as observer-independent and certain” (ibid.). However, data is also in the ‘hard’ sciences not generally seen as a neutral piece of evidence (Buckland 1991; Scheiner 2004; Borgman 2009). Borgman writes:

In our research on science and technology researchers in the environmental sciences, we found differing views of data on concepts as basic as temperature. (2009, sec.28)

A biologist, for example, regards temperature not as given, but as a piece of data that has been acquired by selected means, under specific circumstances in a certain environment. Scientists generate their own data, making them aware of its constructedness and the individual circumstances under which the data has been acquired. In contrast:

The humanities and arts are the least likely of the disciplines to generate their own data [...] Humanities scholars rely most heavily on records (Borgman 2009, sec.33)

Scholars who use existing digital collections operate blindly – the conditions in which the data has been created are often not recorded:

Archivists [...] make little effort to leave clues about the basis for their appraisal decisions or the contexts in which they are made. (Hedstrom 2002, p.37)

Hedstrom suggested that the digital turn in collections could be an opportunity to improve this situation:

New interfaces could serve as gateways to structured information about appraisal and selection. To build such interfaces, however, archivists would have to share their insights [...], and, most importantly, reveal their uncertainties about, and discomfort with, the choices that confront them. (2002, p.37)

So far, however, the proposed changes in how archivists record data have not taken place: the reasons for selection, subjective decisions and beliefs, which are increasingly accepted as forming part of collections, are generally not made explicit.

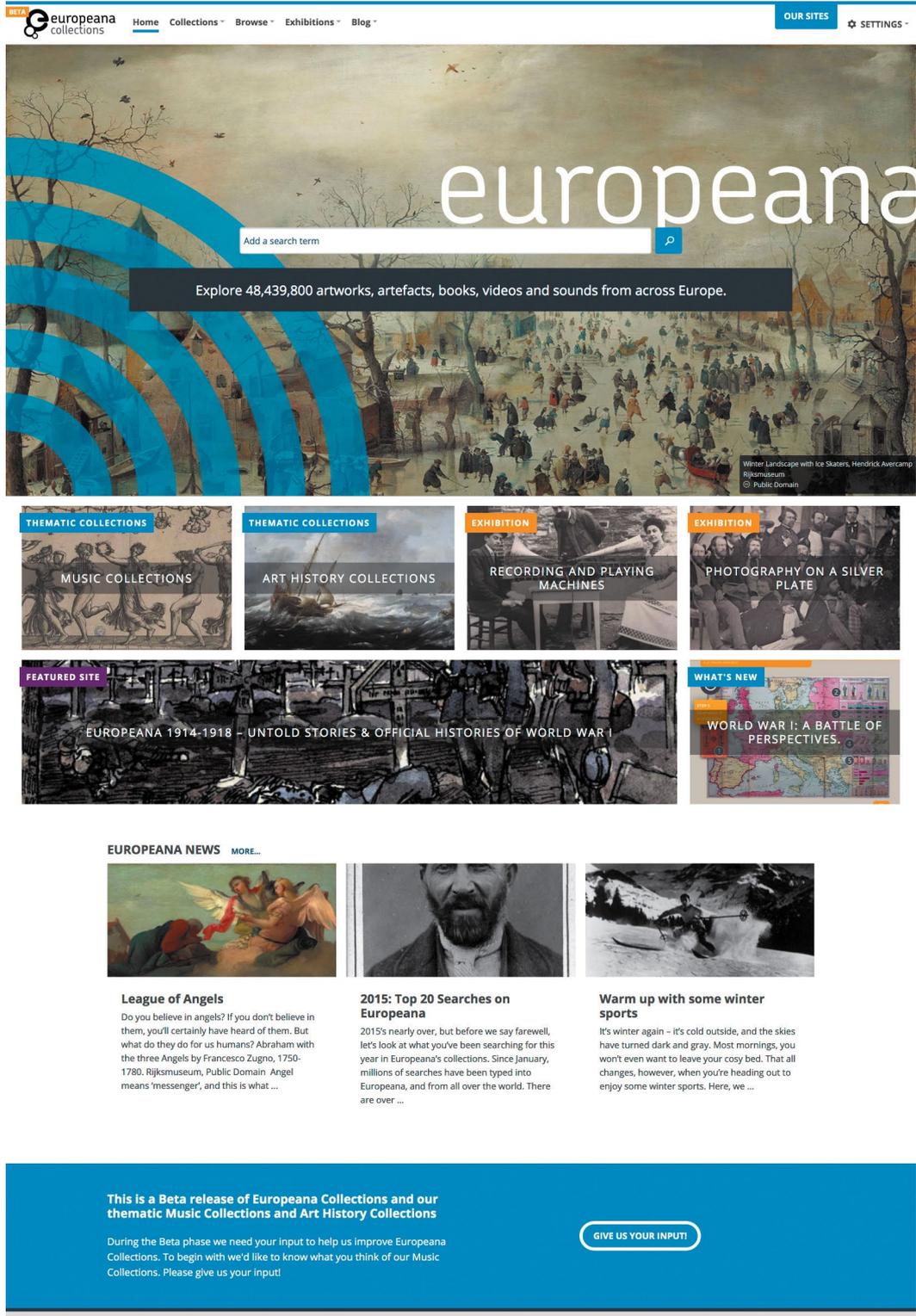


Figure 2.13 – Europeanana provides access to more than 48 million digitised collection artefacts via its online search interface.

image: screenshot europeanana.eu (accessed 06.01.2016)

Time – A Framework for Sense-Making

We have seen how time and the conceptual shift in its understanding as a quantity in the Newtonian sense has played an important role in the invention of data visualisation. Of course, time can be understood in a variety of ways and such numeric and apparently objective models of time have famously been contested.

Bergson (1950) discusses time in relation to consciousness. He distances experienced (concrete duration) from mathematical time (abstract time), the latter seen by Bachelard (1963) as a sequence of discontinuous, countable instants. Bachelard, as a philosopher of science, favoured a quantified model of time, for only what can be expressed in numbers would, in his view, count as scientific. By contrast, Bergsonian duration is “a qualitative multiplicity, with no likeness to number” (Bergson 1950, p.226). His duration is unique and extends continuously from past to present.

‘Scientific’ time is no longer the simple uniform progression from past, to present, to future that non-scientists sometimes like to suggest. Einstein introduced a kind of subjectivity with the theory of relativity, and time’s very existence is repeatedly questioned, including in the ‘hard sciences’ such as physics (Barbour 1999). For Gödel, too (Weinert 2013), time is unreal, a conclusion that has been reached by thinkers such as Spinoza, Kant, and Hegel (McTaggart 1993), and many others throughout history.

In the field of information technology, a number of innovations are introducing interpretive and subjective (Drucker & Nowvieskie 2003), complex and uncertain (Kräutli & Boyd Davis 2013; Meeks & Grossner 2014), and social (Martin 2010) models of time.

Nevertheless, Newtonian time is still the prevalent underpinning model in computing; and, if we keep in mind that it is just one of many, it has considerable merits for analysing data through visualisation by providing a unified frame of reference that can be easily mapped on to the numerical space of a digital screen.

Time-wise visualisations, I argue, can enable users to gain new knowledge from digital collections. Knowledge, according to the European Committee for Standardization is defined as “the combination of data and information, to which is added expert opinion, skills and experience [...]” (CEN 2004). Embedded in this statement is the DIKW pyramid. The letters stand for Data, Information, Knowledge and Wisdom. In a hierarchical order, one follows from the other, although not automatically. Knowledge and Wisdom form the transcendent top half of the pyramid, while Data and Information form the broad basis of all of knowledge.²⁸ The DIKW pyramid is generally attributed to Ackoff

28 The DIKW pyramid is (rightly) criticised for being a simplistic model (Weinberger 2010). In this context it is useful for highlighting that arriving at knowledge based on data is a process that proceeds through several stages. Masud et al. (2010) have employed the DIKW principle to highlight the value of visualisation as a process – in contrast to seeing visualisation merely as an output.

29 Two verses in T.S Eliots “The Rock” (1934) describe the transformations from knowledge to information in reverse order:
Where is the wisdom we have lost in knowledge?
Where is the knowledge we have lost in information?

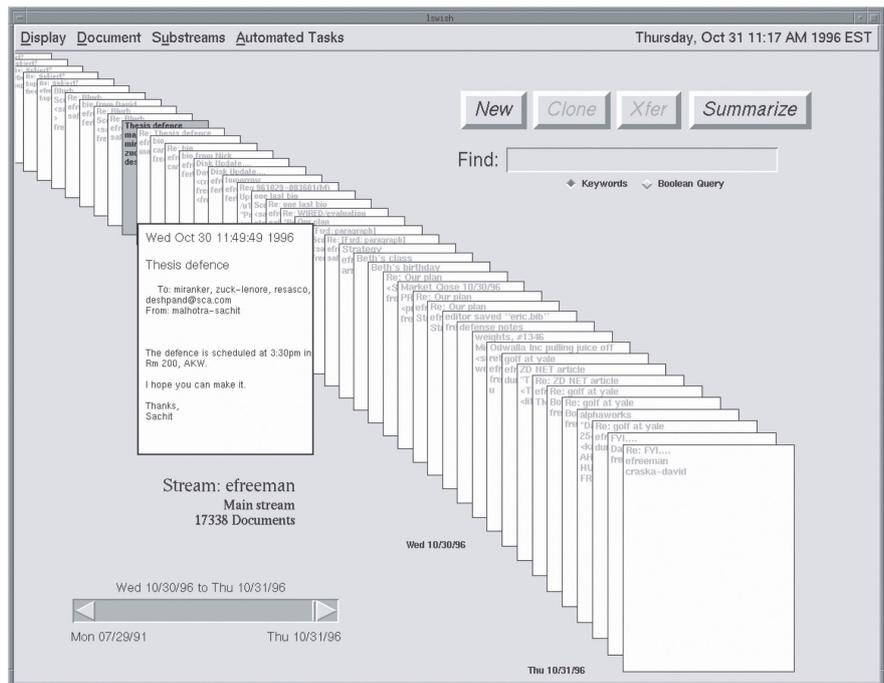
30 Although Wurman talks about structuring information rather than data, he continues to argue that by reorganising information in these ways, additional information emerges. (ibid, p.70) Again, this is an example of information acting as data and the difficulty of discriminating between the two.

Figure 2.14 – *Lifestreams* is an information retrieval system that arranges files in a strict chronology image: Eric Freeman (1997) used with permission

(2010), but apparently has its origin in the writings of T.S. Eliot.²⁹

It is not always trivial to distinguish data from information and there are competing views of what discriminates the two. Rowley (2007) defines information as “organized or structured data”. A slightly problematic view when applied to digital data. Digital data is inherently structured; it must always be stored in a data structure. Is digital data therefore always information? Not if we accept that information is not universally defined, but dependent on a subject. This is the view that is taken here – a common view but by far not the only one generally held (Zins 2007). Digitally structured data may be information for a computer, but it needs to be restructured and represented in order to become information for a user.

What should be the criterion by which data, specifically cultural data, may be organised and finally turned to knowledge? Wurman (1990) writes that there is only a limited number of criteria for organising information: location, alphabet, time, category and hierarchy.³⁰



31 The numerical concept of time can also appear combined with other criteria. For example, data can be organised in years, months and weeks – a hierarchical form of time – or in time periods, which is a categorical type of time.

I argue that only time can be universally applied to digital collections. Alphabet, category and hierarchy are dependent on individual decisions of naming and classification and the necessary data may often be absent. Location could be a universally informative criteria, but again, it is often absent. This leaves us with time.³¹

As with location, the date an object has been created or a piece of text written may not always be known, but nevertheless temporal elements are hardly completely absent in a digital collection. Time is recorded not only explicitly as attributes of individual items or artefacts in a collection, but also in the form of time stamps: when it entered

the archive, when it was recorded in the collection, when it has been accessed, moved, disposed, etc. Even when no explicit temporal information is recorded, time is implied in the sequence in which elements have been catalogued.³²

32 Digital computers, despite the advent of parallel processing techniques, still operate ultimately in sequence, according to the theoretical model of computing as described by Turing (1938).

Arguably, time might simply be the most obvious way to order data. Time – as measured by the running of a clock – structures and coordinates most aspects of our lives, from daily routines, to appointments and yearly holidays. It is the mnemonic by which we remember pivotal events such as 9/11 or 1789 and it is the yardstick by which we slice the past into manageable portions, be it the 1960s or the Renaissance. Time has become so ubiquitous as an organising principle that we tend to take it for granted.

Time has silently become one of the main modes of organising and viewing digital data. A study on email triage found that 89% of all users sort their emails by time (Neustaedter et al. 2005). Social networks sites suggest a chronological ordering, notably Facebook with changing the name of the “Facebook Wall” to “Timeline” in 2011 (Cellan-Jones 2011).³³ Twitter’s attempts at deviating from displaying content in a strict chronology regularly caused an outrage among its user community.³⁴

33 The ordering is however not strictly chronological; ‘popular’ items or subsidised content appear early in the timeline, regardless of the time they have originally been published.

What motivated this shift towards a time-centric mode of digital information structuring? Malone (1983) observed a general struggle of users dealing with complex folder structures already in the age of paper based documents. By adopting the file and folder metaphor in desktop computing, the problems of these organising principles have been inherited as well (Lansdale 1988). Moreover, the rise in the amount of documents that people have to deal with due to the almost zero cost involved in sharing them has led users to become increasingly unable to properly file them into folders (Boardman et al. 2003).

34 see <https://storify.com/fkraeutli/twitter-timeline-out-of-order> (accessed 21.01.2016)

Freeman et al. (1996) have similarly argued that hierarchical folder structures quickly become both too complex and obsolete. They instead propose a time based model: *Lifestreams* (Freeman & Fertig 1995; Freeman & Gelernter 1996, Figure 2.14) is a visual file retrieval system that arranges all personal files and documents in a growing visual timeline. It starts with a person’s birth certificate and extends into the future, including to-dos and documents a user will be needing at some later point in time. They chose time as the structuring element, because in their view

time is a natural guide to experience; it is the attribute that comes closest to a universal skeleton-key for stored experience (Freeman & Gelernter 1996, p.2).

Time has since successfully been used as ordering dimension in many digital information management and retrieval tasks, such as general

desktop computing (Malone 1983; Freeman & Fertig 1995; Rekimoto 1999; Llorens et al. 2011), search tasks (Alonso et al. 2007; Alonso et al. 2010) and in applications targeting specific types of electronic files such as emails (Yiua et al. 1997; Kiritchenko et al. 2004; Ringel et al. 2003) or digital photographs (Platt et al. 2003; Huynh et al. 2005).

Hierarchical ordering of data is increasingly abandoned in favour of time-based sorting as more and more personal files are stored on servers in the proverbial cloud. Names and folders – alphabet and hierarchy – as the obvious modes of organising files are replaced by time based sorting. Photos cease to be files and turn into streams when they are stored on the servers of companies like Apple and Flickr.³⁵

35 In OS X 10.10 and iOS 8 Apple launched iCloud Drive and reintroduced classic file and folder structures for storing arbitrary documents. iCloud documents that are stored and accessed within apps, such as photos, retained their chronological ordering.

The novelty of cloud storage is not merely that files are stored on a server, but that data can be copied, edited and subsequently synchronised across a multitude of devices. Keeping track of the latest version of a piece of data is a non-trivial challenge. After all, a document might have been edited on a device without internet connection and only uploaded later, but the same document could have been changed on a different device in the meantime. Tracking changes over time is necessary in order to resolve potential conflicts and to only retain the most current version of a piece of data.

It is unclear whether the shift towards time-centric structuring of information was informed by conscious design decisions, such as the ones proposed by Freeman et al. (1995; 1996), or if the focus on time from an engineering perspective has spilled over to the front-end. Time might arguably not be a “skeleton-key”, but it is an organising model that has proven to be very effective across a large number of knowledge domains.

Time, History and Chronology

One view of collections I left out in my earlier discussion is the dictionary definition, which equates archives to accumulations of historical documents and artefacts.³⁶ This historical perspective makes studying them along a model of time an obvious choice, especially if ‘history’ itself is considered according to the dictionary as “the past considered as a whole” (OED Online 2015b). However, history cannot simply be equated with the past. According to Walter Benjamin (1940/1991), history is located in the present (“Jetztzeit”). History is constructed and subjective – a view shared by many historians and theorists, dating back at least to Hegel (1770-1831) who argues that

36 The Oxford Dictionary of English defines ‘archive’ as a “collection of historical documents or records providing information about a place, institution, or group of people” or as “the place where historical documents or records are kept” (OED Online 2015a).

even the ordinary, the “impartial” historiographer, who believes and professes that he maintains a simply receptive attitude; surrendering himself only to the data supplied him – is by no means

passive as regards the exercise of his thinking powers. He brings his categories with him, and sees the phenomena presented to his mental vision, exclusively through these media. (2001, p.24)

When we study collections over time we are employing *chronology* in order to derive meaning. John Locke (1632-1704) paired chronology, together with geography, as the essential prerequisites for history:

Without geography and chronology [...] history will be very ill retained, and very little useful. (1693)

Locke emphasises the importance of geography and chronology for structuring events which otherwise would just be a “jumble of matters of fact” (1693).

In its original greek meaning, ‘historia’ is concerned with retelling of what one has learned through investigation, not the least by chronology. In that sense, history is concerned with communicating knowledge, while the task of chronology is to turn the jumble of matters that is data into structured information, to be interpreted by a historian in order to derive knowledge.

Time, in the view of some authors of the 18th century, may serve as the unbiased framework by which events are presented “in a truer Light than regular Histories” (Pointer 1714). While the claim for a possible truth value of time has to be taken with a grain of salt, Pointer continues to argue that when events are presented in a chronological manner, it enables the individual on his own to make sense and draw new knowledge - to “make their own Inferences from simple Matters of Fact” (Pointer 1714).

Where history is the - necessarily subjective - retelling of events, chronology provides rigour: a kind of ‘honesty’ in studying and representing events. Events are studied according to a schema that is traceable, communicable and transparent: the model of calendrical time.³⁷

Pre-calendrical descriptions of events depended on correlations with temporal landmarks. Roman historians would chart events according to a list of past consuls, while Athenian used their own reigns (Feeney 2007, p.9f) - a cumbersome endeavour and converting between different cultures’ models of timekeeping meant identifying where they align and differ. The primary motivation for creating a unified calendrical model was, according to Feeney (2007, p.18), not practical or logistical considerations, but the writing of history, for which chronologies that were compatible across cultural differences in timekeeping were crucial.

The Julian calendar provided a unified model according to which events can be ‘positioned’ - a model which allows their users to easily find out whether two events have happened before or after each other,

³⁷ Numeric, calendrical time is only one of many models of time that historians may use in studying past events. Fernand Braudel (1902-1985), for example, employs three: geological time (*la longue durée*), social time and individual time (Hughes-Warrington 2008).

or simultaneously. Graphical renderings of events, such as Priestley's *Chart of Biography* that followed 1700 years later, made such comparisons even more accessible:

You see at one glance, without the help of Arithmetic, or even of words, and in the most clear and perfect manner possible, the relation of these lives to one another in any period of the whole course of them. (Priestley 1764, p.10)

The ease by which events can be located in chronologies can however introduce also a kind of dishonesty, when it leads to historic data being represented in a truer light than is possible. The date of past events can never be truly known, even if they are described according to a rigorous framework such as the calendrical model of time-keeping. A rigorous calendar is also a rigid model of time and no guarantee for 'truth'. Calendrical dates do not rule out the presence of doubts, contradictions and inconsistencies.³⁸

³⁸ Dershowitz and Reingold (2008) list 30 calendars, some of which still in use, that are historically significant and not necessarily compatible.

History, in the historiographer's definition, is therefore not a neutral representation of the past, but, as far as possible, an honest *retelling* of past events. Collections serve, to the researcher of history, as a source of evidence for past events. Chronology - structuring events through a model of time - equips a researcher with an 'honest' framework for making sense of collections: a transparent, rigorous, but not necessarily 'true' model for studying history.

Chronology and a unified model of time equipped historians with methods for being rigorous and transparent - in short - honest about what they know and how they know it. Digital research methods need to follow similar ambitions, which in the case of timeline visualisation tools means being rigorous and transparent about the visual representation of sets of data.

Curators' Perspectives – Objects and Time in Digital Collections

Through my conversations and interviews with curators and other representatives of cultural institutions, and by examining their collections datasets, I was able to gain a better understanding of the prevalent views and issues around digital collections. Specifically, I will present insights on the handling of dates; how temporal data is stored in cultural datasets and the possible interpretations that digitally stored dates may entail.

Most digital collections store dates as a pair of values denoting an earliest and latest possible date. Additionally, the date is generally stored as free text: it is this representation that the curators work with and is exposed on a website when the collection is accessible online. Numeric dates are, in day-to-day use, primarily relevant for searching records, as the textual dates are not machine readable.³⁹ Data visualisation depends on them for the same reason: generating graphics digitally relies on the underlying data to be machine readable.

³⁹ They are machine readable as text, but – without additional processing steps – not amenable to machine interpretation as dates.

The numeric date pairs typically bracket the date of production of an artefact. Other dates, if present, are usually only recorded as a single date. The date brackets therefore do not normally relate to a durational time period, but a possible time frame for a historic event. Collections data thus holds a measure of confidence in the numeric dates, which is derived from the written date descriptions.

Often, there is a significant discrepancy between the free text that the curator enters manually and the numeric date pairs that lie 'behind' them and are sometimes generated automatically. The numeric values for the pairs of dates are typically stored as years, even in cases where more precise information would be available in the written date. In other cases, where the precision of the known date is less than a year, the numeric dates are set as a precisely defined range of years. In the Cooper Hewitt objects database, for example, "mid-20th century" becomes 1940-1958, "possibly ca. 1960" is stored as 1955-1965 and "1946 or later" is quantised to 1946-1989.⁴⁰ Date-formatting and processing protocols add their own interpretations on the representation of the data, typically implying greater precision than was available to the person who originally entered the information.

⁴⁰ The dates refer to a piece of textile, a necklace and a pitcher. See <http://cprhw.tt/o/2CqE2/>, <http://cprhw.tt/o/2DFYi/> and <http://cprhw.tt/o/2DuSr/> (accessed 10.08.2015)

count	format	count	format
897	•	1	probably •• june •••• - • july ••••
298	•••• - ••••	1	possibly ••••
278	ca. ••••	1	middle of ••th century
105	••th century	1	mid sixteenth century
97	ca. •••• - ••••	1	mid - ••••s
57	ca. ••••	1	late ••th century - early ••th century
51	•s	1	late ••••s to late ••••s
26	late ••th century	1	july •••• ?
23	early ••th century	1	july ••••
17	••••'s	1	first half ••th century
14	••••s - ••••s	1	eighteenth century
12	about ••••	1	early •rd century
10	late ••••s	1	circa ••••
10	••th century - ••th century	1	ca. •th century
8	•••• - ••	1	ca. ••••'s
7	mid ••th century	1	ca. •••• to ca. ••••
7	late ••th - mid ••th century	1	ca. •••• to ••••
7	••• - •••	1	ca. •••• - - ••
7	••/••/••••	1	ca •••• - ••••
6	late ••th century - ••th century	1	ca ••••
6	ca. •••• - ••	1	c. •••• - ••
5	•••• to ••••	1	c. ••••
5	•	1	•th century to •th century
4	late ••th - early ••th century	1	•st quarter ••th century
4	ca. •••• - ••••	1	•st century - •nd century
4	ca. •••• - ca. ••••	1	•s
4	•th century to ••th century	1	•rd quarter ••th century
4	••••'s - ••••'s	1	•nd half of ••th century
3	ca. •••• - ••	1	••th march ••••
3	after ••••	1	•••• or ••••
3	•th century	1	•••• - ••••s
2	ca. ••••s	1	••• - ••••
2	[••••]	1	••/•••• - ••/••••
2	•nd century - •th century	1	••/••••
2	••th century to ••th century	1	•• july ••••
1	spring/summer ••••	1	• november, possibly ••••

Table 2.1 – Date descriptions in the V&A collections data. Dots (•) represent numbers from 0-9.

Collections datasets exhibit a range of textual date descriptions. Table 2.1 on page 77 lists, in order of frequency of occurrence, the format of date descriptions in a representative sample of 200 records from the V&A dataset. A large number of them include expressions of uncertainties and doubts. There is, however, no indication of the source or reason of these uncertainties. A user who is unfamiliar with the history of an individual item will be oblivious if the uncertainties originate from a lack of information, conflicting records, unreadable or ambiguous sources - in short: the reasons for a lack of precision in the temporal information remains undisclosed in the dataset. It is important to remember that the same applies also to the dates without such qualifiers; one cannot inherently assume that these dates are certain. Certainty - as well as uncertainty - is merely a matter of sources and interpretation, none of which is captured in either of these digital date descriptions.

We can identify the vocabulary used to express uncertainties, doubts and timespans: “ca.”, “probably”, “possibly”, “after”, “late”, etc. A fairly consistent set of terms, probably because the software used to catalogue the datasets tries to automatically convert these descriptions to numeric dates and displays an error, or requires manual user input, whenever it cannot recognise an expression. Due to them partly being controlled by the expectation of the software, this list of terms might reveal more about the inner workings of the recognition algorithm than about the curator's language for dating objects. Britten-Pears Foundation's thematic catalogue of works of Benjamin Britten contains textual date descriptions that have been manually converted to machine readable dates. Nevertheless, the descriptions follow similar patterns and expressions of confidence - “circa”, “pre”, “post” - which could mean that the terminology remains relatively fixed even when the manual conversion would allow more freedom.

A curator I talked to about the dates I encountered in collections databases suggests that these date descriptions are influenced by conventions even when they are not mechanically enforced:

20 years as a curator, I was always forced to be certain about things I wasn't certain about. Think about how you express both uncertainty - not trueness - and which date you pick - however certain you are. Is it the start of construction, end of construction, sale, purchase, destruction, whatever. (C3)

When an object is exhibited in a museum, it usually is described by just a single date, the year of production:

There are two concealments that museums have practiced. One is, the concealment of uncertainties. The second is the concealment of the long lives of things. [...] historians have only been interested in invention; the moment of genesis. [...] we try to conceal everything since 1829 to the present by saying Stephenson's Rocket is 1829. (C3)

Conceptually one might think of dates in all different kind of ways - "There's a language that I use to express temporality, which is a fuzzy language and it's a subjective language, as you made me realise." (C3). Conventions of the field, however, led to a standardisation of the terminology of describing dates. Digital technology enforces its own standards on dates through the processes of quantification I described. These effects are added on top of the ones date descriptions have already undergone before they entered a digital system.

The use of digital databases to represent knowledge is often associated with a loss of information resulting from the structural requirements posed by the digital system, a recurring critique of the computational turn in the humanities:

In cutting up the world in this manner, information about the world necessarily has to be discarded in order to store a representation within the computer. (Berry 2011, p.2)

At least in the realm of cultural collections, where structures and conventions have dictated how information is stored even before the advent of computers, digitisation also partly had the opposite effect. Cataloguing software abolishes the physical space restrictions of index cards and allows more freedom in the descriptions and choice of fields. These richer forms of data storage, however, introduce their own problems to collection holders, as one curator recalls:

That was a quite difficult transition from when you did paper records to computer records. We used to have a glossary of terms, a thesaurus. You made your objects fit a prescribed framework of terminology. There was physically a book which gave you the thesaurus of terms. Now with more free flowing cataloguing, someone could call a bowl a 'bowl', or a 'basin'. [...] It was sort of easier to extract the information, because under 'ewer' it would say 'see basin'. (C11)

The challenges resulting from the relative freedom granted by digital cataloguing systems were echoed by a number of curators. As users are not forced to adhere to specific standards, datasets can easily become messy:

[...] we have a large number of different people cataloguing all the time. Because you can catalogue away in free text [...] it can cause you real issues. (C1)

Collections datasets often become complex and hard to interpret, because the system may allow a great amount of liberty in choosing how and where to store data -

You'll find some objects with description fields. Others without, but [descriptions] in another field. (C2)

- but also because the data structures have been changed and expanded as new items entered a collection that were not originally anticipated when a system was put in place:

We have customised our collections management system very much. [...] It has given us flexibility, but there's a flip-side to that [...] it's quite difficult to standardise it. (C2)

Besides making it often challenging for curators to retrieve data, collection holders pay for increased customisation and flexibility when data from different collections has to be combined. A process which is quite common; except for a small minority of relatively recent catalogues, all datasets I have worked with contain items that have previously been maintained in (or constituted) separate collections, often as part of different departments within museums.

In the old days we had a 2D collection and a 3D collection and they had quite different views about cataloguing [...] We've been trying to bring the two together. (C2)

It is during these processes of merging datasets that data has to be reinterpreted or may be discarded when old data finds no place within new databases - "We call it 'squashing'. It's not a technical term but it's very, very difficult" (C5). Without background knowledge of a collection's history, one may not always know how trustworthy certain pieces of information are. Especially with regards to data that may not be considered the most essential aspect of a collected item - often temporal information:

They treated this whole batch [of posters] as if it was 'an archive' [...] The individual posters within that [batch] - and of course all subsequent posters no matter what date they are - have just been given inventory numbers based on the date they were acquired, even if, of course, they could be much earlier. (C8)

Other factors, such as changes in the administrative structure of a collection also have a great influence on the accuracy and credibility of dates:

We have a lot of objects, supposed to be acquired in 1876. They have been acquired before but that's when we started a new inventory system. [...] We know as curators that when you get [this date], you don't believe it. (C3)

Dates are particularly problematic to interpret in collections that occupy a large timeframe, both in terms of the items they contain as well as with regards to the period of their own existence. In younger and smaller collections that might only focus on the works of an individual artist, and therefore are able to apply more consistent dating strategies, knowledge about the origins of dates is more easily maintained and curators may know which dates they can trust...

There's a lot that could be done with the period of 1928-38 because he kept diaries then. He didn't keep them before 1928, so any date that's on there before 1928 will be because it's written on it. (C6)

...or conversely which dates have to be treated with caution:

[The artist] was really bad at dating. He just painted something and sold it, so stuff has gone into lots of different collections and come back to us at the Library, but we've got very limited information about dating. (C5)

These factors are, of course, embedded also in the data of larger collections, but it is less likely that knowledge about the origins of dates - and most other aspects of the dataset - has been acquired about the complete collection, passed on and maintained throughout changes in governance and staff.

Archives and museum collections are seen as the neutral, material evidence of our cultural past. Lynch describes digital collections as "uninterpreted databases of raw cultural heritage material" (2002). The UK's National Museum Director's Council states the purpose of museums in their manifesto as follows:

We are a mirror to our own times and illuminate developments in our culture and society. (2004)

What museums are generally not eager to admit in public, yet are aware internally, is that they are a distorting mirror, reflecting culture and society through various filters and biases. Beginning with the choice of material by which the past is represented...

In a museum, your core-thing is the object [...] I suppose as curators we work on the premise that the object is what is fascinating. (C11)

... and continuing with the various factors that determine which objects are preserved, which can be physical -

Maybe people were very selective in the past sometimes [...] when storage and conservation were big issues. (C11)

- administrative -

[...] the shape of the collection is determined by the administrative structure and preservation criteria; what the museum deemed important enough [...] (C8)

- strategic -

Sometimes you just have to take something and hope that with passage of time it will become significant (C1)

- or environmental:

The content of the database is a non-random selection of the painted pottery that survived from ancient Athens [...] (C7)

In my review of the literature I have already encountered a growing recognition that collections and collections data are not neutral, but subject to various influences. Through my conversations with curators and encounters with their datasets I have been able to get a better idea of the specific kinds of interpretations that are exerted by the individuals, institutions, conventions and the digital systems respectively.

Digital databases and the people that use them add various layers of interpretation to a cultural collection. There can be no “uninterpreted” (Lynch 2002) databases as much as there can be no uninterpreted collections. Digital data structures add their own distortions to collections datasets. Computational representations of time, which, in the computer science tradition is treated as a numerical space, are particularly affected by this, as we have seen.

Making sense of cultural data over time therefore entails making sense of the distortions that a collection has experienced throughout its history. The potential personal, institutional and organisational biases of collections, and consequently, collections data needs to be taken into account when visualising it on a timeline.

Humanities researchers are used to having to be critical about their sources and the need for criticality applies to digital datasets as well. Several scholars I have talked to observe that digital sources are often not approached with a healthy level of suspicion, which could well

be a symptom of digital interfaces not allowing or promoting critical enquiry.

Ideally, visualisation tools can enable this criticality and help to expose the various layers of human and machine-made interpretations a dataset has undergone. It might be the most important difference when designing visualisations for scholarly, rather than for casual use; not to simplify, not to make a dataset appear more perfect than it is, but on the contrary, to respect and emphasise its imperfections and to steer the attention to inconsistencies and possible sources of knowledge.

Discussion

If we regard a collection as a dynamic entity that is shaped and re-shaped by the people that govern and use it, digitisation has not fundamentally redefined the concept of a collection. Digitisation has, however, changed how collections may be used and studied, and added – to the traditional methods of humanities research – those methods that are universally applicable to digital data: data mining, visual analysis, automated processing, visualisation, etc. Digital tools and methods are, however, not easily transferable from the quantitative sciences, from which they originate, as is evident in the difficulties of curators to digitally model notions of time in databases as well as other kinds of humanities data. Digital Humanities methods require established paradigms for data analysis, including visualisation, to be reconsidered. Researchers need to develop an understanding of the kind of insights digital tools can enable. Collections data offers an ideal testbed for simultaneously exploring – by studying and developing appropriate visualisation methods – the requirements of humanities research tools as well as the knowledge currently hidden in digital collections.

Data Visualisation for the Humanities

Data visualisation is one of the essential research methods for the Digital Humanities, especially when the research is – as it is most of the time – concerned with large datasets. However, established paradigms for visualisations in the sciences do not necessarily translate to humanities data. Lunenfeld et. al write:

Currently, visualization in the humanities uses techniques drawn largely from the social sciences, business applications, and the natural sciences, all of which require self-conscious criticality in their adoption. Such visual displays, including graphs and charts, may present themselves as objective or even unmediated views of reality, rather than as rhetorical constructs. (2012, p.42)

Jessop on the other hand argues that digital visualisation methods are not in any way “revolutionary” (2008) or “lacking in rigorous scholarly value” (ibid.). Rather they are a continuation of established academic practice. A similar line of argument is pursued by Unsworth (2000), who lists a range of “scholarly primitives”; common discrete activities that humanities scholars need to be able to perform when doing research, regardless whether these are carried out with digital or analogue tools.

Drucker argues that the effectiveness of visualisation has caused humanities researchers to lose the necessary criticism for rigorous scholarly research:

The sheer power of the graphical display of “information visualization” [...] seems to have produced a momentary blindness among practitioners who would never tolerate such literal assumptions in textual work. (2011b, sec.5)

Drucker complements her article with suggestions of less reductionist graphical displays which try to convey some of the subjective, emotional and uncertain components of the presented data. These are toy examples based on artificial data and it is unclear how they would behave with and scale to real world datasets. They do however convey a sense of the necessity to scrutinise visual representations and to question the implied precision, honesty and trustworthiness of visual diagrams.

The reason for an apparent lack of rigour in digital humanities research stems not only from inappropriate tools and visualisation methods, it might also lie in researcher’s unfamiliarity with visualisation tools and a lack of criticism towards them. The ability of graphs and diagrams to mislead - whether intentionally or out of lack of acuity - and the risk of compromising on the ‘truthfulness’ when representing data visually is a known problem in the scientific community, and even in popular literature (Huff 2010).

On one hand, there is a clear need for visualisation tools that are able to fit the characteristics of humanities data and research - on the other hand it is also necessary that humanities researchers become aware of the mechanisms of digital data visualisation and how they in turn shape humanities knowledge production.

My research aims to address both of these issues; by developing new tools, and by developing them in collaboration with curators and archivists.

It is crucial that researchers gain an understanding of the epistemological consequences of knowledge production from digital sources by digital means and that we learn to understand what kind of insights we can expect to gain. Even if more appropriate tools were available to humanities researchers, without a better understanding of their working principles they remain “intellectual Trojan horse[s]” (J. Drucker 2011b).

Digital Collections As Objects of Research

I argued that we should study digital collections independently from the collections they are based on; by regarding them primarily as catalogues and, thus, data.

However, I am hesitant to refer to digital collections as ‘raw data’. The term is generally used to describe data that requires to be processed – for example through visualisation – in order to deliver useful insights. Clean data, on the other hand, is organised and pre-processed so that it can be read also in its ‘original’ form.⁴¹ I suggest that it is necessary and valuable to consider any kind of dataset, whether well organised or disorderly, as open to reinterpretation.

41 i.e. as a text document or spreadsheet, not in the ‘original’ form of bits on a digital storage device.

Differentiating ‘raw data’ from data in the context of digital collections is, I argue, neither useful nor possible. Lynch, who defends the view of collections being uninterpreted, questions at the same time how “interpretation-neutral” a collection can be (2002); because, for example, “interpretation creeps into the descriptive metadata” (Lynch 2002).⁴² In his view, interpretation-less digital collections would nevertheless be the ideal to strive for and he suggests that biases in collections could be neutralised by drawing from different sources.

42 We will later see the various stages throughout the lifespan of a collection where interpretation ‘creeps in’.

What I discovered through my research – and will describe later on – is that embedded “interpretations” in digital collections must not present an obstacle to making sense of it. They form part of any collection and hence should not be suppressed; doing so would mean consciously disregarding one of their fundamental properties. Furthermore, when the methods for accessing the datasets acknowledge the presence of embedded biases, they can enable valuable insights.

Existing work in visualisation-based interfaces for cultural collections largely treats them as digital substitutes of physical archives and focusses on making their content more accessible to specialists and the public – with an emphasis on the latter. These include for example the SFMOMA ArtScope (2007, Figure 2.16), which allows users to explore more than 6,000 artworks in a tile-based zoomable visualisation⁴³ or the map based interface of the Natural Science Museum of Barcelona.⁴⁴

43 I have discussed SFMOMA ArtScope in more detail elsewhere (Kräutli 2012)

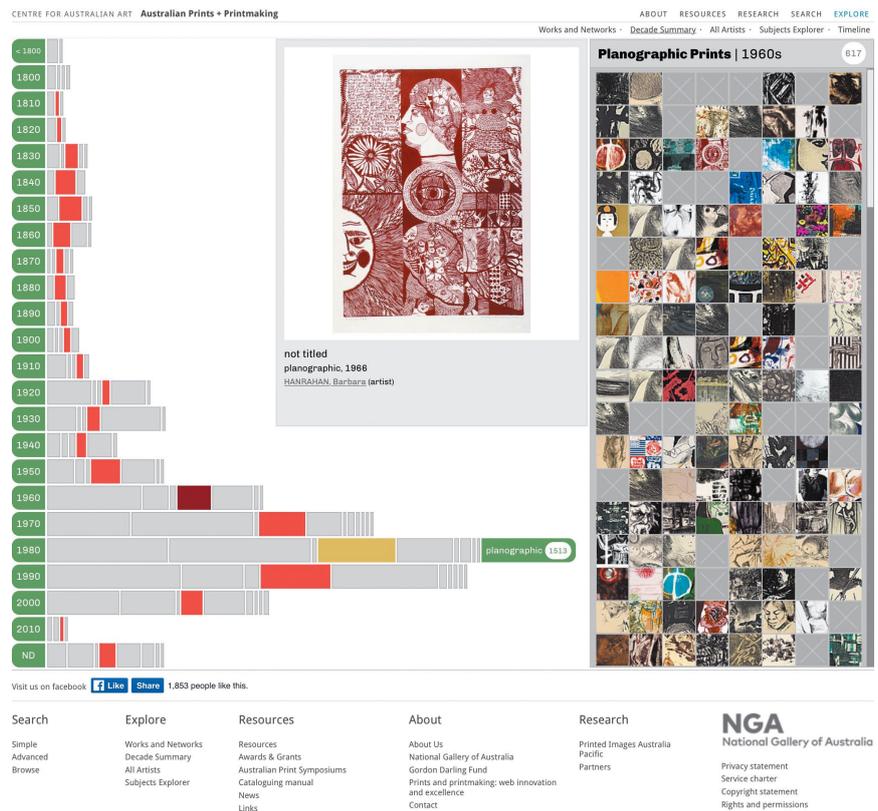
Several interfaces for cultural collections have been developed by Whitelaw (2015), who advocates a “generous” approach of enabling access to digital collections:

44 <http://mapa.bioexplora.cat/> (accessed 12.12.2015)

Generous interfaces offer rich, browsable views; provide evocative samples of primary content; and support an understanding of context and relationships. (Whitelaw 2012)

Figure 2.15 – The Decade Summary of the Australian Prints + Printmaking online collection uses a vertical temporal histogram, organised by print categories, as the primary mode of navigating the collection.

image: screenshot prints-andprintmaking.gov.au



45 <http://printsandprintmaking.gov.au/explore/> (accessed 15.08.2015)

An example has been developed by Ennis Butler (2013) – Whitelaw’s PhD student – based on the Centre for Australian Arts Print collection (Figure 2.15).⁴⁵ Through five different visual interfaces, online visitors can explore the collection based on the relationship between works, keywords, their associated individuals as well as two time-based views: a “decade summary” of the entire collection and a timeline view, which plots the works of an individual artist along a vertical timeline. Different access points and rich contextual cross-references allow users to immerse themselves in the collection. Users are permitted to get lost in a collection instead of – the normal task of curators and what interfaces are usually designed for – looking for something specific.

Generous interfaces build on Bates’ (1989) *berrypicking* technique and Dörk’s (2012) concept of the *information flâneur*: models that see search and information seeking not only as a goal-directed task, but as an explorative and – not the least – enjoyable activity.

The interfaces Whitelaw and his collaborators have built so far demonstrate the richness of digital collections, which is often hidden behind ‘ungenerous’ search forms. The interfaces are aimed at the general public and are not specifically intended for scholarly use, although “they may prompt such analyses both by scholars and (importantly) wider communities of interest.” (Whitelaw 2015, sec.38).

Whitelaw gives an example of what can be derived from the summative diagram in the decade view of Ennis Butler's prototype:

The resulting graph is informative in itself, showing the chronological shape of the collection and the relative distribution of different print types. The boom in stencil printing in the 1970s and 80s, for example, is clear. (2015, sec.26)

A scholarly tool would need to support a deeper engagement with the occurrence of such patterns. Is it really representative of a general increase in the usage of this printing technique? Is the entire collection a representative sample of prints produced during that time or is the pattern telling of the taste of an individual curator? Are these all unique prints or maybe several reproductions of the same original stencil? Are the dates reliable⁴⁶ or could there be errors or uncertainties? Could there be biases in the structure of the catalogue or the database that skew the digital data in this category?

Visual analytic tools for cultural collections should have the ability to answer such questions and be able to convey the reliability of the conclusions that can be drawn from observations.

⁴⁶ In digital data, items that carry the year 1970 are particularly suspicious as the date could in fact be unknown. 1970 marks the beginning of the Unix time period. Dates that are missing are sometimes interpreted as 0, or in the Gregorian calendar, 1.1.1970. Many witnessed the possible implication of this misinterpretation when Facebook congratulated them to 46 years of friendship on January 1 2016 (Dockterman 2016) – these users befriended each other before Facebook started to store the dates of their friendships, resulting in many dates stored as 0.

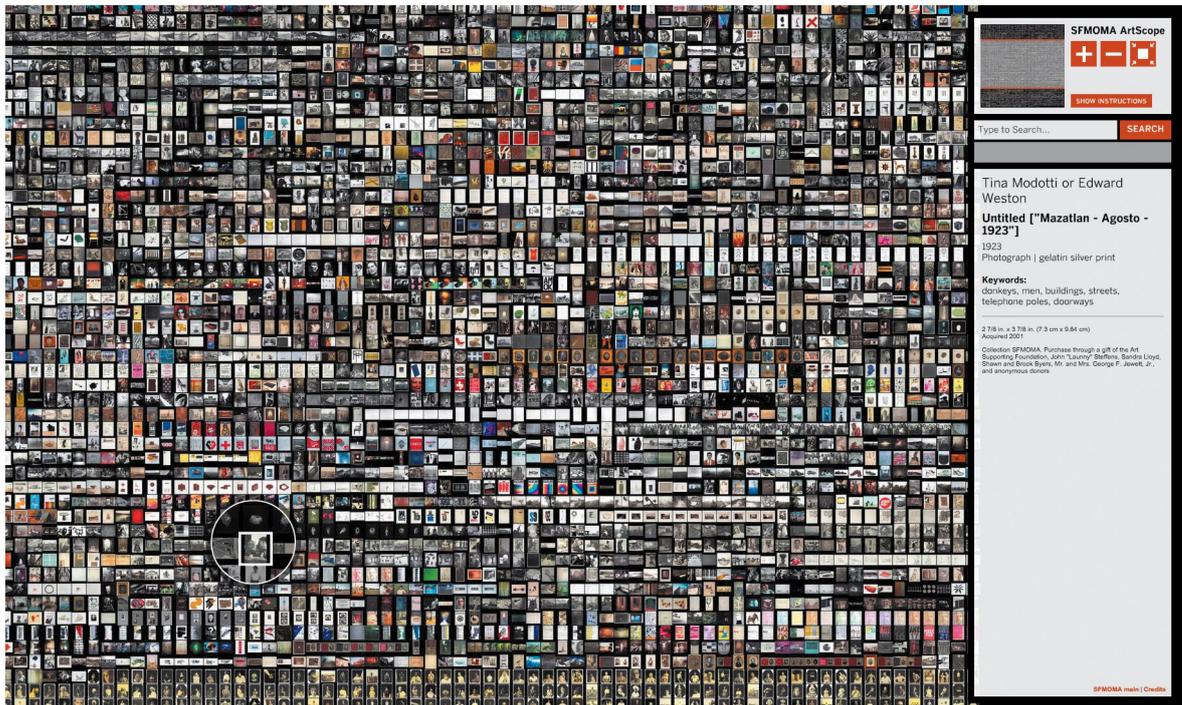


Figure 2.16 – Artscope is a zoomable online interface for accessing and searching a part of the SFMOMA art collection. image: screenshot of www.sfmoma.org/experience/artscope (accessed 20.08.2014)

Digital Historiography

Traditionally, there has been a clear separation between a researcher and a collection; between the archivist – the maintainer of the collection – and the historian – the user of a collection. Collections provide the ‘neutral’ evidence for the historian, which the archivist has gathered and structured. Archives are organised by basic principles such as “*respectus des fonds*” (Duchein 1983) which ensures that the original order of the artefacts and documents in an archive is maintained and suggests that they represent a neutral conservation of the past. The task of a historian is then to make sense of this material through tracing their historic narrative.

We have seen, however, that collections are increasingly conceived as neutral accumulations of evidence, but instead as dynamic entities that are defined and redefined by subjective individuals. This has consequences for historians and anyone doing research with archival materials.

Historians separate their materials into ‘primary’ and ‘secondary’ sources: ‘original’ pieces of evidence as those contained by archives and other people’s interpretations of such evidence. If we accept, however, that historical evidence as contained in collections is already subjected to individual and institutional influences and appraisal, we

consequently have to accept that a researcher working with a collection is in fact not working with primary sources in the strictest sense.

It is not surprising that this is an unacceptable situation for some and one of the reasons why many scholars advocate the use of 'traditional' over digital methods; studying the 'actual' archival material over working with digitised artefacts. The competition emerges most visibly in digital collections that, in addition to cataloguing data contain high-resolution scans of historic documents. When high-resolution photographs allow researchers to examine the quality of a brush stroke or the texture of a paper more closely than they would be allowed and able to with the 'real' physical artefact, the implied hierarchy between the digital and the physical object becomes hard to defend.

However, this is not an argument for the digitised artefact to be regarded as equivalent to the original. It is an argument for the opposite. When the digital, an obviously remediated, interpreted and subjective 'secondary' source, is as useful for humanistic research as the 'primary', it becomes hard to deny that also the physical artefact in a collection is to a certain degree a secondary source.

The digital, in the context of collections, not only requires new methods, but also a new understanding of historiography: of what we are able to derive from historic evidence and on what constitutes that evidence. This is a development that has been identified by Kramer:

So what, then, does it mean within the digital domain to address historiography when it is understood to be the collection of secondary sources and ongoing debates about a historical topic? It would mean, perhaps, rethinking the relationship between primary and secondary sources in new ways, not just going to the supposedly pure sources, fetishized as they are in the field of history. (2014)

Digital collections emphasise the need to be critical about archival sources - that primary sources in an archive carry a certain degree of secondary interpretation. Digital collections require historians to reconsider assumptions about primary sources in collections and researchers need to develop new methods that are honest about the degree of pre-interpretation of sources in curated collections. The way in which digital structures and representations emphasise the constructedness of a collection and, consequently, the data about the collection requires us to face the implications of researching history through remediated sources.

3

“Until analytical tools and services are more sophisticated, robust, transparent, and easy to use [...], it will be difficult to attract a broad base of interest within the humanities community.” (Borgman 2009, §5)

3 Digital Timeline (Tools)

By embedding my research in the wider field of the Digital Humanities and studying its foundations with respect to collections data and time-wise visualisation I have identified a number of challenges for my research to address. I will now investigate the status quo of digital timelines in order to identify possible shortcomings in existing implementations. I will explore the abilities of digital timelines to qualify as analytical tools through a review of existing projects. My aim is to explore and make explicit what separates a naïve timeline from a timeline that can fulfil the needs of scholarly research and enable visual analysis of a wide range of digital datasets. In doing so I will identify shortcomings and opportunities of existing digital timelines for visual analysis of cultural collections. These will result in the formulation of focus issues - concerning the time-wise visualisation of large datasets, the development of timeline layouts and the incorporation of multiple temporal descriptions - that will guide my practical exploration of the research questions through a series of prototype visualisations.

I will review a selection of timelines drawn from examples I have gathered throughout the duration of my studies,¹ outline their characteristics and potential problems and how they translate to timeline visualisations more broadly. I will initiate a deeper discourse around this class of diagrams in the way that is already taking place with other formats such as network visualisations (Fruchterman & Reingold 1991; Golbeck & Mutton 2006; Martin et al. 2011; Krzywinski et al. 2012) and even pie charts (Brinton 1919; Eells 1926; Croxton & Stryker 1927; Cleveland & McGill 1984; Spence 2005).

I begin by introducing a spectrum to systematise timeline implementations for scholarly analysis. The spectrum is presented as an alternative to DeFanti et al.'s (1989) discrimination between visualisations for either *communication* or *discovery* (analysis). While classifications are helpful for making comparisons through providing a framework for discussion, I found that this particular distinction is not very useful in the case of visual timelines. Priestley's *Chart of Biography*

¹ A list of timeline tools is included in appendix C on page 270.

(1764), for example, can be seen as both communicative and analytical and the same applies to most of the timelines I will introduce in this chapter.

I therefore employ a spectrum that distinguishes timelines by progressing in four stages from *static* to *open*. It does so through a set of thresholds that describe interactive and technical features that signify an advancement in a timeline's use as an analytical tool. Open timelines, the ultimate stage, contain interactive features to analyse, filter, select and compare data, and are open in the sense that they are able to work with different kinds of datasets and are themselves open to interpretation and critique. The tools I set out to develop, will consequently be open timelines.²

² In accordance with established paradigms for visual analysis (Tukey 1977) and digital tools for humanities research (Unsworth 2000).

My review of existing tools is led by a set of criteria. Based on an initial review of requirements I devised and refined these criteria over the course of my research and revised them according to new findings that resulted from my practical and theoretical investigations. The criteria help me to direct my attention to the specifics of how data is visually represented over time. Timelines are a familiar diagram format and consequently a lot of their visual rhetoric is often taken for granted. By examining timelines through a 'lens' of criteria, I am able to look beyond certain preconceptions and reconsider present assumptions and design decisions. These criteria are included in appendix A on page 251 and form the method by which I study the examples in order to ensure that my enquiry follows a transparent and rigorous process and covers all the aspects I set out to examine. The criteria therefore mainly serve the execution of the present investigation and are not intended to be used outside of this thesis - although they could be appropriated and improved by others.

Defining Analytical Timelines

At first I looked at timelines as either visualisations for “discovery” and “communication”. Communicative timelines may be illustrations to a narrative that has been constructed based on historical evidence: a way to communicate time-based information in a graphical manner in order to educate others. Most visual timelines typically found in museums are of this kind - engaging communicative devices, but often too simplified to serve as scholarly research tools and, some argue, even too simplified for communicating the complexity of historical narratives (Lubar 2013). The datasets that they represent are often limited and heavily curated, and the arrangement and visual representation of events follows aesthetic and practical constraints.

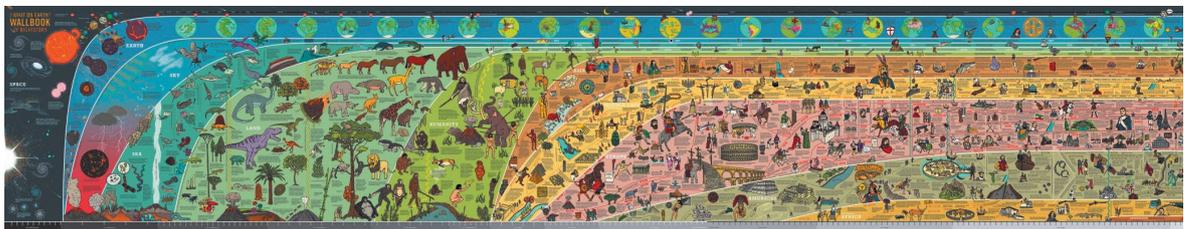


Figure 3.1 – The What on Earth? Wallbook Timeline of Big History is a 2m long visual timeline. It includes more than 1,000 hand-drawn illustrations that cover natural history as well as the history of human civilisation.

image: Christopher Lloyd, What on Earth Publishing (used with permission)

Such timelines may be of significant educational value, but one would not be inclined to draw any reliable conclusions from visualisations that are evidently not drafted for the purpose of scholarly research. For example, we would not expect Christopher Lloyds *Wallbook of Big History* (2011, Figure 3.1), a “comprehensive, visual voyage” through all of history, to - in fact - be comprehensive. Nor would we expect to be able to precisely read the period of time different species lived on our planet from the graphical space that they occupy on the time axis; it is not an arithmetic mapping of events for the purpose of analysis. The timeline is not intended to represent all of history, but to give an overview of important events in an accessible and engaging manner. The purpose of this timeline is not to discover new knowledge in the field of history, but to communicate insights from the field to a public.

A timeline for the purpose of communication is the product of a *transformer*, a term coined by the designer and theorist of educational graphics Marie Neurath:

It is the responsibility of the ‘transformer’ to understand the data, to get all necessary information from the expert, to decide what is worth transmitting to the public, how to make it understandable, how to link it with general knowledge or with information already given in other charts. In this sense, the transformer is the trustee of the public. (Neurath & Kinross 2009)

Communicative illustrative timelines may allow for novel insights, but their primary goal is not to process data, but to convey pre-processed information.

We may not always be able to clearly distinguish communicative from analytical timelines, especially when it comes to their digital instantiations. When a graphical timeline is generated programmatically, the rendering of events has to be made explicit which results in digital timelines often, but not always, being more strict and therefore transparent in the visual mapping of events to (numerical) time. This side-effect of digital timelines facilitates their use as analytical devices, but it does not automatically follow merely from them being digital.

A helpful classification I employed in the beginning of my research - when I collected different examples of digital timelines - was the distinction between timeline *projects* and timeline *tools*. A project represents a purpose-built timeline visualisation around a specific dataset, while a tool constitutes a dedicated software or library for visually representing arbitrary datasets in time.

In many cases we can deduce that purpose built timeline projects are communicative visualisations as they are based on a single selected and curated dataset and often include elements of storytelling. Timeline tools, on the other hand, accept external datasets and may be used as analytical devices, whether or not this use was originally intended. However, there are a range of exceptions to this rule, such as dedicated timeline implementations that later have been adapted to accept different kinds of datasets, or timeline tools that are, in fact, authoring environments designed to produce simple narrative timelines.

I therefore propose a four-step description for classifying digital timeline visualisations: *static*, *dynamic*, *exploratory* and *open* timelines (Table 3.1). We can look at it like a continuous spectrum, as a visualisation might develop from a static to a dynamic timeline, and further. As a progressive spectrum, each threshold presupposes the conditions of the previous classification(s).

	Static	Dynamic	Exploratory	Open
Interactive features may include	–	<ul style="list-style-type: none"> - switching between different views - more detail on individual records - scripted interaction - navigating parts of the display separately 	<ul style="list-style-type: none"> - manipulate graphic representation beyond story-boarded interaction - more detail on several records - filtering, searching 	<ul style="list-style-type: none"> - ability to manipulate any aspect of the display through UI or programmatically - import custom datasets

Table 3.1 – A schema for classifying digital timelines according to their interactive features.

Static timelines are digital timelines, which function equally well on screen as they could when printed on a paper. A user might be able to pan and zoom the display in order to look at aspects of the visualisation in detail, but static timelines do not offer any additional interactive features, by which I mean the ability to change the presented view or aspect of its rendering.

Dynamic timelines offer basic functionality for a viewer to manipulate the display. They could be similar to slide shows or scroll based websites where interactivity is present, but mostly confined to a limited and pre-scripted storyboard – the author has anticipated most of the possible manipulations. Additional information about aspects of the visualisation may be provided, for example, by highlighting or selecting graphical elements.

Exploratory timelines allow for a deeper level of interactivity and manipulation by, for example, not only offering different views on a display, but giving a user more freedom in affecting the rendering of individual events, or offering methods for filtering or searching a given display. The timeline affords ways for interrogation beyond visual inspection.

Open timelines give a user not only more freedom in the way the data is represented, but also in the kind of datasets they visualise. This could be facilitated, in the case of dedicated software, through an appropriate user interface or, in the case of software libraries, through a well-designed API or sufficient documentation of the source code.

For arriving at digital timelines that allow visual analysis and insights, I regard open timelines as the desirable format. This does not automatically mean that a static timeline is necessarily inferior in this regard – a well-designed static visualisation may be more insightful than a thoughtlessly designed one that is technically superior. However, open timelines allow for verifiable insights which may be reproduced and reused by others, which I – and others (Borgman 2009) – see as a main requirement of tools for scholarly analysis.

Evaluation Criteria for Timeline Visualisations

- 3 Based on their presence in scholarly literature, social media and discussion platforms for visualisations.

In order to be able to evaluate different kinds of digital timelines, despite their dissimilarities and the many possible implementations, purposes and styles, I have drafted a set of criteria which I have used to examine and compare the variety of digital timelines I have collected, of which I will present a selection. I selected the projects primarily based on their acceptance³ and diversity. My aim is to be comprehensive within my abilities - I can only safely evaluate a tool to which I was granted access. As is the nature of digital artefacts, some timelines I have studied may in the meantime have been updated or changed or - at worst - disappeared. I therefore captured screencasts of the discussed projects and included the URLs to the videos in appendix C on page 270.

The criteria I propose and apply are not meant to be a means for assigning grades. There are no correct answers and a visualisation which can respond to several criteria is not automatically better than one which just answers a few. My approach relates to Twyman's schema for the study of graphic language "as a device for directing our thinking and not as an end in itself" (1979, p.202). The criteria I employ are intended as a checklist to help me focus on issues that might get overlooked:

All this is made necessary because our training and experience, whether primarily verbal, numerical, or visual, tends to predispose us towards particular approaches to graphic communication. (Twyman 1979, p.202)

The criteria are organised thematically as a tree (Figure 3.2). Upper nodes represent wider themes that depend on several factors and might require individual judgment based on personal requirements. Therefore, they are further broken down such that the criteria at the leaf nodes may be assessed, as far as possible, without subjective appraisal. We move from larger questions such as the overall representation of records, to finer grained aspects such as the graphical vocabulary used to render individual items - what kind of shapes and colours are present and what do they signify?

- 4 It is included in appendix A on page 251.

The complete tree contains 112 criteria, which I will not discuss in detail here.⁴ I will merely describe the five central topics which are below the root node:

"Data acquisition and curation" looks at a visualisation based on the data level. What kind of datasets are allowed or included? What kind of expectations does a dataset need to meet?

"Representation of dataset" evaluates the visual rendering of the complete dataset. How is the layout assembled? What kind of temporal

models are used? What factors influence the readability of the entire dataset?

“Representation of records” focusses on the visual representation of individual items. What kind of graphics are used? How are temporal elements mapped onto visual aspects? What kind of temporal descriptions are supported, such as events, periods, uncertain or discontinued times, etc.? How consistent is the mapping of time to graphical space? Are there exceptions?

“Interaction” forms the largest branch and looks at aspects that facilitate exploration and sense-making in the visualisation. What modalities and techniques are used to navigate the visualisation? Can the visualisation be searched or filtered? What aspects of the display can be manipulated?

“Technology/platform” is the most pragmatic branch to evaluate and looks at the technical implementation of the digital timeline. Is it a standalone tool or a web-based application? Does it rely on standard libraries or is it coded largely from scratch?

I will comment on the specific insights that these criteria facilitated at the end of this chapter.

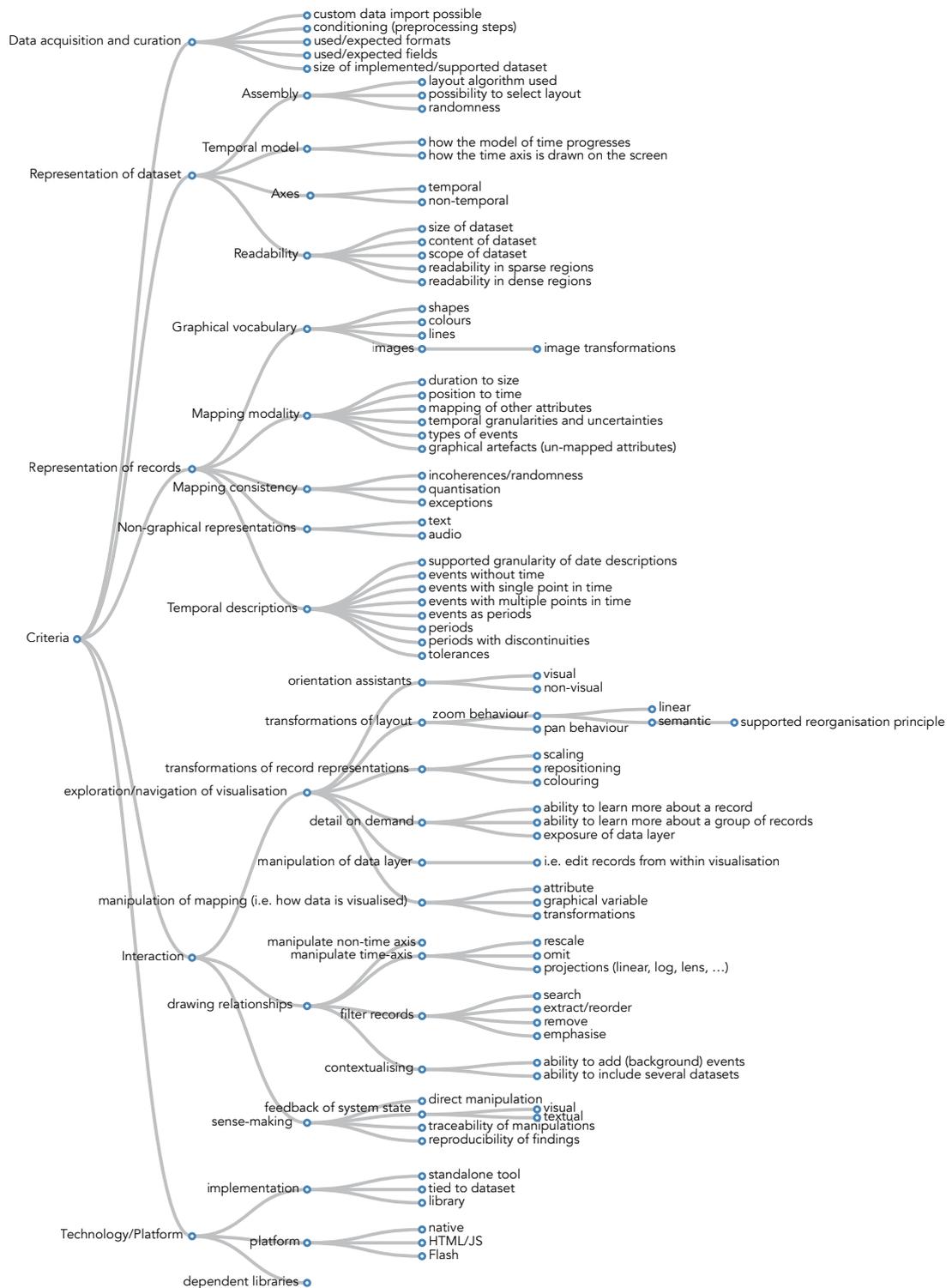


Figure 3.2 – A criteria tree for studying visual timelines, containing 112 items organised in 5 major categories.

Static Timelines

Static timelines include all digital chronographics that offer no additional level of interactivity beyond that provided by their medium. When a digital artefact is interactive, we expect it to change appropriately as a result of an input that a user executes through an input device, such as a mouse, a touch screen or a keyboard. We might implicitly understand the concept of interactivity, which nevertheless leaves a lot of room for interpretation and differences in definitions (Smuts 2009). Manovich argues that the term is meaningless in digital media because, as he states, “once an object is represented in a computer, it automatically becomes interactive” (2001). Here, I use the term simply to separate static timelines that offer no interactivity on their own – except, what Manovich refers to, through the browser or other software that is required to view them – from other digital timelines.

The timeline I will focus on most here has in fact been published in print, in addition to being accessible as a digital image. I nevertheless treat it as an example of a digital timeline as it has been created with digital means and exhibits some of the problems that are typical for digitally generated timelines. It is a piece by Accurat, an Italian design studio specialised in data driven media design. Their portfolio includes a variety of timeline based works, of which I want to take a closer look at a particular visualisation that features a reoccurring topic throughout the history of visual timelines: the rise and fall of empires. Like Barbeau de la Bruyere’s *Mappe-Monde* (1750b), and other historic examples such as Friedrich Strass’s *Strom der Zeiten* (1804) or John Spark’s *Histomap* (1931), Accurat’s *The empire strikes back* (Beltramin et al. 2012, reproduced in Figure 3.3) plots the reigning period of a selection of the world’s greatest empires. I have chosen this particular visualisation as a representative example of a static digital timeline as it employs a prevalent – if not the de facto standard – timeline layout.

Each empire is drawn as a rectangular bar, positioned horizontally on a linear time axis according to its year of rise and extending width-wise according to its duration. Typically in this timeline format, each bar would be of equal height, but in this particular example, the height encodes an additional dimension of the dataset. Each bar is scaled vertically in proportion to the maximum geographic expansion of the empire it represents. A yellow vertical line indicates the date of maximum expansion. The vertical order of the empires follow their date of fall, putting them in almost sequential order. A detailed legend on the left informs the viewer of the data that is represented graphically and gives an impression of the richness of the underlying dataset.

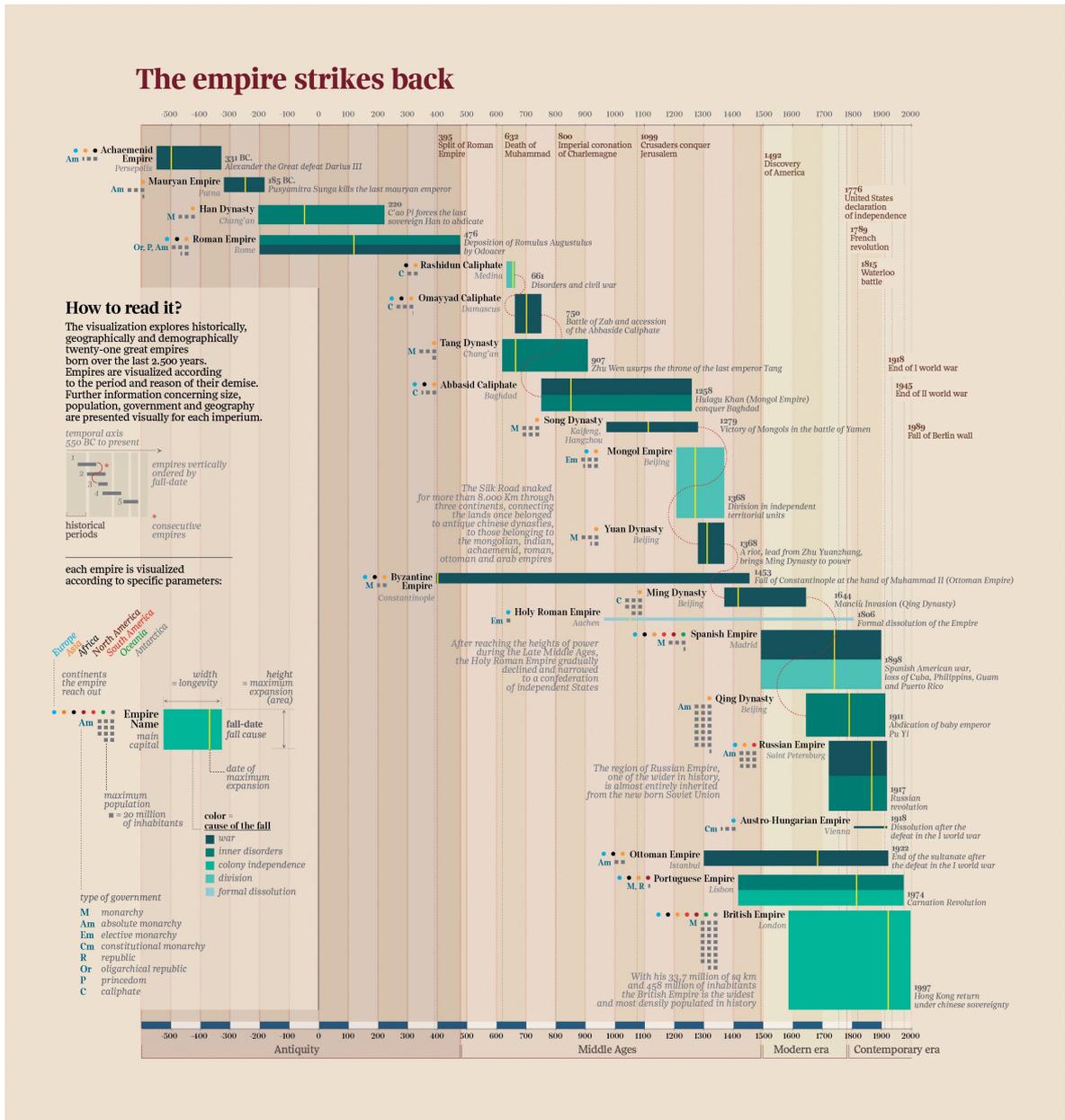


Figure 3.3 – A timeline visualisation plotting twenty-one historic empires over the course of 2,500 years. image: accurat (reproduced with the author's permission)

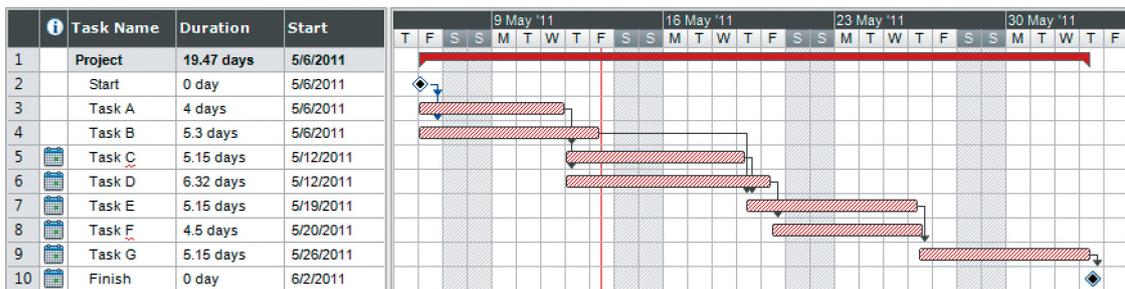


Figure 3.4 – A modern Gantt chart displayed in a project planning software.

image: Wikimedia commons (CC BY-SA 3.0)

Before I look at the details of what this visualisation can tell us and what could be improved, I want to highlight an aspect in its layout that applies to a lot of similar timelines.

This particular layout has its roots in Karol Adamiecki's *harmonogram* (Marsh 1974), better known today as Gantt chart (Figure 3.4), after Henry Gantt (1919) who popularised the use of this chart in the early 1900s. Gantt charts are a common format for planning and visualising the scheduling of tasks using a tabular format. Columns typically represent units of time, such as days or weeks, while rows represent individual tasks, ordered vertically in sequence of their planned completion. Shaded cells indicate the time and duration of a particular task. In a Gantt chart time proceeds horizontally and the vertical axis contains categorical information - the name of each task or group of tasks. In Accurat's timeline, the vertical ordering is not used to show an independent dimension of the dataset. Both the horizontal axis as well as the vertical ordering of the chart represent temporal information: calendrical date of rise on the horizontal and sequence of fall in relation to the other empires on the vertical.

The strong correlation of these two temporal axes results in the apparent downwards sloping of the entire layout: a clear pattern, but one that is a result of the chosen layout and not a pattern that is present in the dataset. This effect is recognisable in many mechanically generated timelines. Automatically generating a timeline's visual layout is not a trivial task; having each record occupy a separate row, like in a Gantt chart, is a possible solution that can be implemented relatively easily.

Although the authors have clearly thought about their chosen layout - otherwise they would not have explained it in the legend - the decision to adopt a Gantt-like arrangement might have been more of a pragmatic than an optimal choice. While the height of the empires corresponds to aspects of the underlying dataset, their vertical positioning carries no meaning in this regard, but solely depends on the number of chosen empires that precedes them. When looking at the representation of the entire dataset, the overall shape the individual bars form does not reveal any insightful patterns.

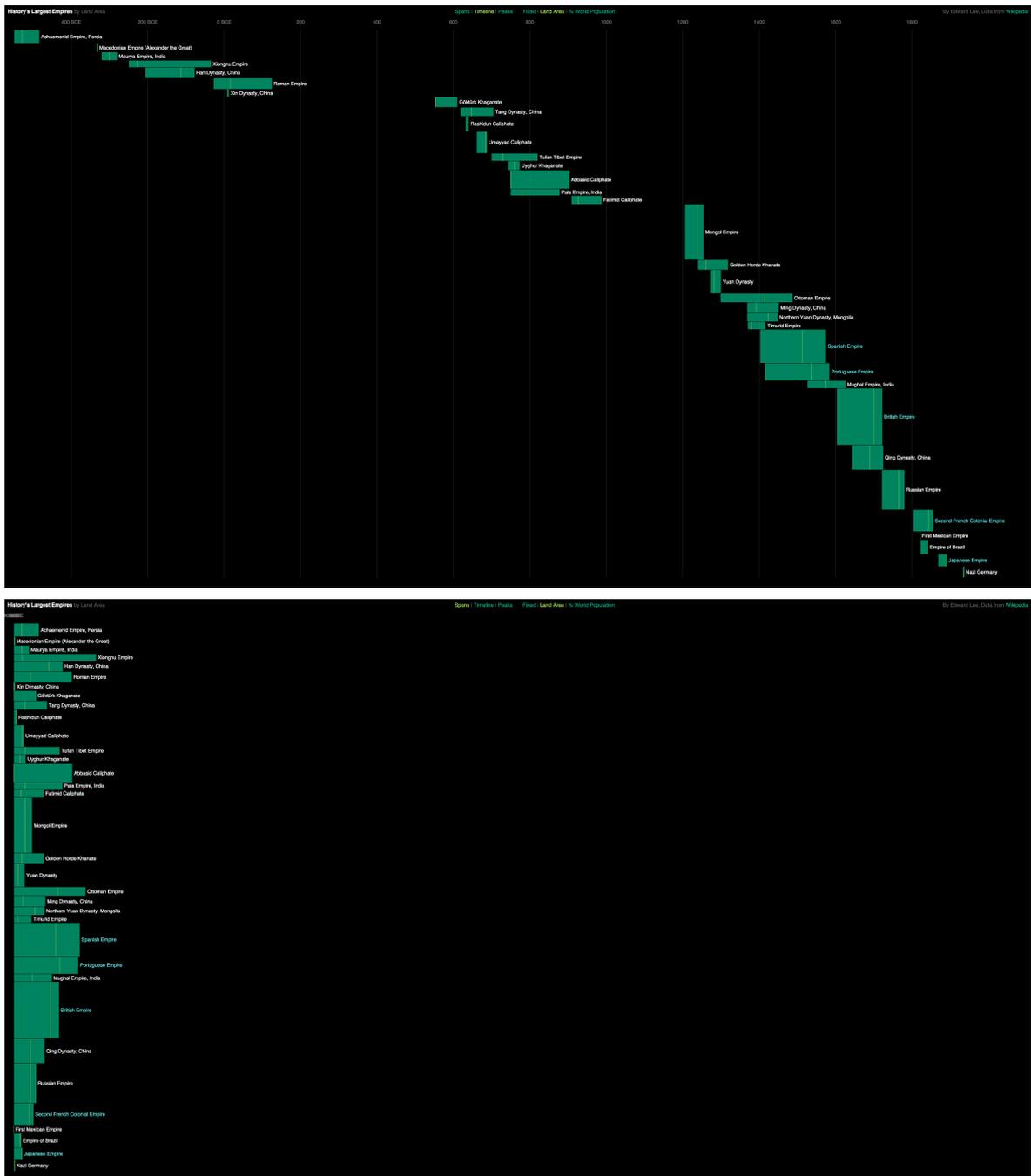


Figure 3.5 – In Edward Lee's interactive timeline of empires the horizontal bars can be aligned by their start date, allowing users to make absolute and relative comparisons of time periods and durations.

image: screenshots of <http://edwardclementlee.com/vis/empires/> (accessed 20.01.2016)

- 5 Based on the publication dates and personal conversation with both authors (14.08.2015 and 23.08.2015).



Figure 3.6 – A dotted line in Accurat's timeline indicates consecutive empires.

image: accurat (used with permission)

- 6 "liaisons fréquentes, ou plus grands traits historiques communs" (Barbeau de la Bruyère 1750a, p.19)
- 7 "cette Chronologie uniforme, avant & après J.C. est préférable à la manière de compter de la Création du Monde, qui varie selon les différents Auteurs" (Barbeau de la Bruyère 1750a, p.6)

A digital predecessor of Accurat's visualisation can be identified in Edward Lee's *History's Largest Empires* (2011, Figure 3.5). This timeline bears a striking resemblance to Accurat's visualisation and is very likely to have informed it.⁵ Lee's timeline offers the ability to change the horizontal positioning of the bars, aligning them all by their beginnings or according to their date of maximum expansion. Here we are able, for example, to compare the relative duration of each empire, while still retaining their temporal order encoded in the vertical arrangement.

In Accurat's timeline, we can look at individual empires and in some cases, through an S-shaped dotted line, follow consecutive empires (Figure 3.6). Analysing overall patterns or judging the temporal distance of one empire to another, however, is not something the visualisation affords easily. The vertical arrangement of the empires could have profited from more deliberation. Using the vertical position for geographical categorisation might have been an obvious choice, making it easier to see how empires that occupied similar territories progressed through time rather than causing empires that are close in geographical space, but distant in time to be spread out across the chart.

We can see how Barbeau de la Bruyère has identified, and indeed addressed this problem in his *Mappe-Monde* (1750b - Figure 2.7 on page 55). Bruyère is aware of the consequences of representing geographical space on a one-dimensional axis. He therefore selected neighbouring countries based on their shared borders as well as their relationship and shared history (Barbeau de la Bruyère 1750a, p.19)⁶ and made use of colour shading to indicate the coherence of empires that are graphically separated.

Time is indicated as vertical axis on the left and right edge, with horizontal lines spanning across the chart. The time axis is non-uniform, using a condensed time scale for the distant past, which makes the horizontal separation slightly misleading - equal distances do not represent equal time periods across the chart. I will revisit the subject of uniform and non-uniform scalings of time axes later on. A double line marks the beginning of the Christian Era, the uniform reference point for counting time which Bruyère notes is "preferable to the method of counting from the Creation of the World" (Barbeau de la Bruyère 1750a, p.6).⁷

As Boyd Davis (2015a), points out we are essentially faced with a coordinate space: the vertical and horizontal position as well as the width and height of the plotted empires carry significance. In contrast to the two digital timelines discussed above, we can also see how the

extent of individual empires changes over time and how the same geographical regions have been governed by different rulers.

8 “[...] chaque Pays a été habité depuis les premières Peuplades faites après le Déluge [...]” (Barbeau de la Bruyère 1750a, p.37)

Every point in this chart therefore carries significance which, as Bruyère points out, logically follows from the fact that every represented country has been inhabited at all times.⁸ This is a level of consistency in the mapping of data to graphical space that we may not always find in timeline visualisations. Nevertheless, it is something we should strive for in digital timelines or at least – as Bruyère does as well – reflect on the implications and limitations of a chosen layout.

Bruyère created his chart, of course, without the help of digital tools. Nevertheless its rectified, diagrammatic appearance in principle lends itself to be generated mechanically and differs from similar charts produced around the same time that made use of more rhetorical graphical devices – for example rivers to represent the history and convergence of different empires (Rosenberg & Grafton 2010).

Accurat’s *Empire* timeline, as well as other examples from their portfolio which I will discuss later, are some of the most thoughtfully designed timelines that have been made recently – the great care for detail is evident in the richness of the represented data and the meticulous explanation of the visual parameters provided in the legend. The problems I have outlined are mainly a result of the algorithmic production of the timeline layout. Diagrams, whether constructed manually or mechanically, may never be able to escape the danger of patterns being misunderstood. However, if the representation can be manipulated, such as in Lee’s (2011) example, it might be possible to allow viewers to question a pattern and examine different ways of mapping a dataset along time.

Dynamic Timelines

Dynamic timelines can be characterised as static timelines with the addition of certain interactive features that, ideally, benefit their use as scholarly research devices. The online editions of the New York Times (NYT) and The Guardian feature a number of digital timelines that I consider to be *dynamic* based on the features that they offer to navigate the visualisation. Online journalism offers a rich source of dynamic timelines, although there has recently been a decline in favour of more slide-show based formats - possibly due to the unsuitability of most implementations for mobile devices. My selection here does not include any scholarly examples - these will follow soon - mostly because interactive timelines for research purposes tend to include more advanced interactive features. However, the argument here is that even evidently simple modes of interactivity - such as presenting detail information in a tooltip - can support the use of visual timelines as research tools.

The Life and Legacy of Nelson Mandela: 1918-2013

Nelson Mandela's quest for freedom in South Africa's system of white rule took him from the court of tribal royalty to the liberation underground to a prison cell to the presidency.



Figure 3.7 – A horizontal, linear timeline accompanies an article of the New York Times. It plots important events in the life of Nelson Mandela and simultaneously serves to navigate the article.

image: screenshot of <http://www.nytimes.com/interactive/2013/12/05/world/af-rica/Mandela-Timeline.html> (accessed 17.08.2015)

A variety of dynamic timelines have been developed by the team at the New York Times (Belopotosky et al. 2011; Delvisicio et al. 2013; Andrews & Parlapiano 2014; Clark & Bilefsky 2015 - Figures 3.7-3.9). They each seem to have been intended as reusable visualisations - their code is modular - nevertheless most of them only ever appear once and only a few of them have been used to visualise a variety of data-sets. This diversity of timelines appears to be primarily a consequence of progressing web technology: more recent implementations make use of the latest technology available at their time as well as more image and video based data sources. The visual rhetoric of the different timelines as well as their technical capabilities on the other hand have not changed significantly.

Events are represented as dots (Delvisicio et al. 2013; Andrews & Parlapiano 2014) or dashes (Belopotosky et al. 2011) and arranged from left to right on a horizontal time axis, forming a one-dimensional timeline visualisation. Buttons enable a step-by-step navigation, emphasising the chronological sequence of the events. This sequential navigation can also be helpful in more 'busy' periods, where events tend to overlap. In one of the earlier timelines (Belopotosky et al. 2011, Figure 3.8) it is possible to adjust the scaling of the time axis via zoom buttons which makes it easier to examine busy period - a feature which is not present anymore in later examples of NYT-timelines.

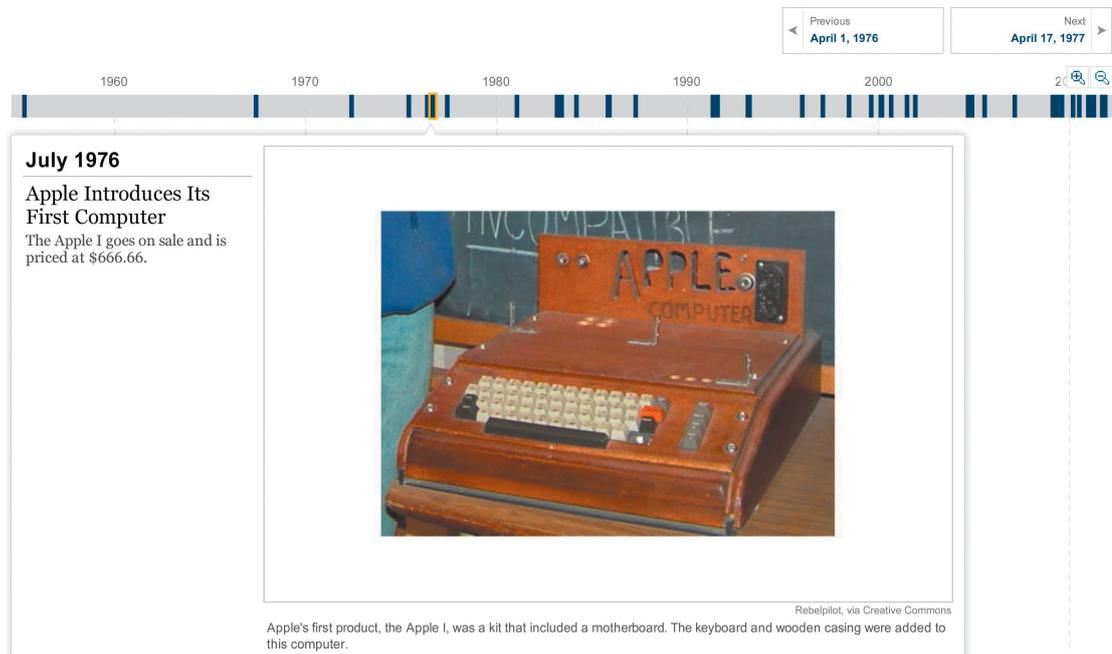


Figure 3.8 – A Flash-based timeline by the New York Times is used to narrate the biography of Steve Jobs. Zoom buttons in the upper right corner allow the time axis to be rescaled.

image: screenshot of nytimes.com/interactive/2011/10/05/business/20111005jobs-life-timeline.html (accessed 17.08.2015)

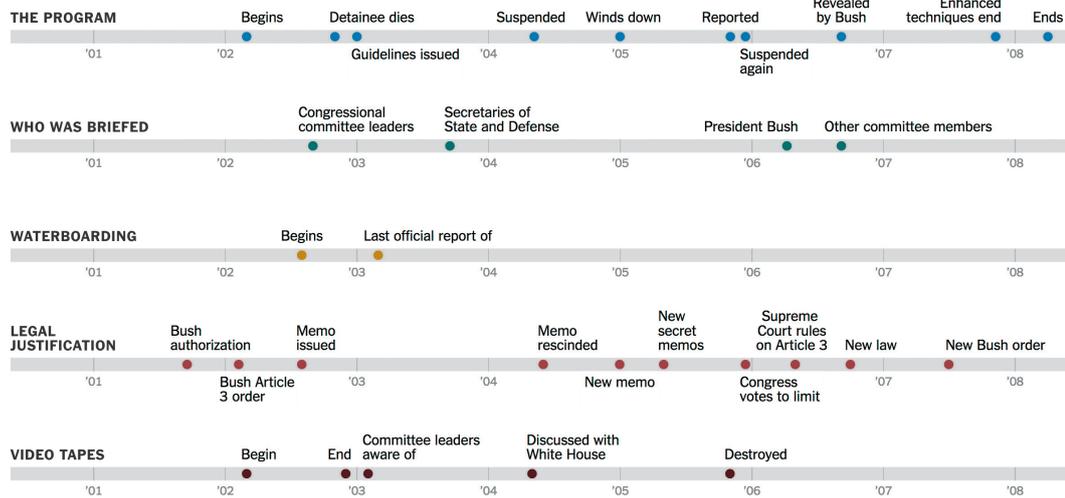


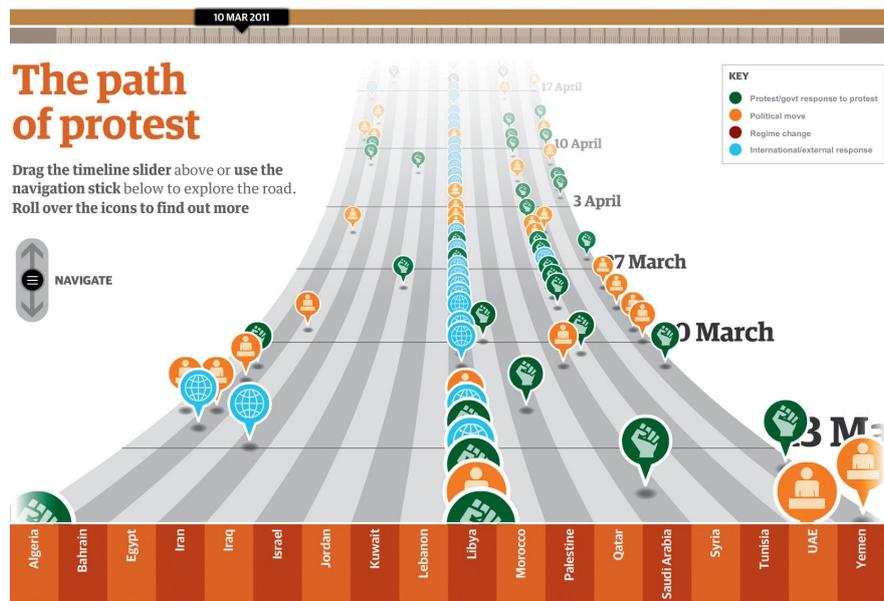
Figure 3.9 – Events related to the C.I.A.'s secret interrogation programme are categorised and visualised on separate timelines. A description of the event appears by hovering over a dot.

image: screenshot of nytimes.com/interactive/2014/12/09/world/timeline-of-cias-secret-interrogation-program.html (accessed 17.08.2015)

The user is able to interact with the event representations by using the mouse to hover over or click them. These limited, even seemingly trivial ways of interacting with the display offer a significant advantage over static visualisations by allowing access to secondary information on demand and - more crucially for the use as analytical tools - the original source of the visually represented data.

In one example, *A History of the C.I.A.'s Secret Interrogation Program* (Andrews & Parlapiano 2014, Figure 3.9), the NYT departs from the one-dimensional timeline format. The vertical dimension is used to denote the category of different events related to the uncovering of the contentious interrogation methods practiced by the CIA. This enables a viewer to gain a better impression of the complexity and the interrelatedness of events, something which a one-dimensional timeline and indeed an accompanying text or chronology tends to hide in favour of a sequential narrative.

Figure 3.10 – A reusable, Flash-based timeline by the UK Guardian employs a first-person perspective on events. image: screenshot of theguardian.com/world/interactive/2011/mar/22/middle-east-protest-interactive-timeline (accessed 18.08.2015)



Slightly odd appears the decision to use an individual time axis for every category despite all of them using the same time scale. The multiple timelines, each with its own axis, seems to suggest that either separate courses of events are visualised using different timescales, or that the timeline in fact flows continuously from one line to the next.

Inspecting the generated code of the timeline reveals that each category is drawn independently on a separate HTML element. It is quite likely that the timeline was originally intended to depict a single one-dimensional and linear narrative, but was 'misused' by a creative editor in order to visualise a more complex story.

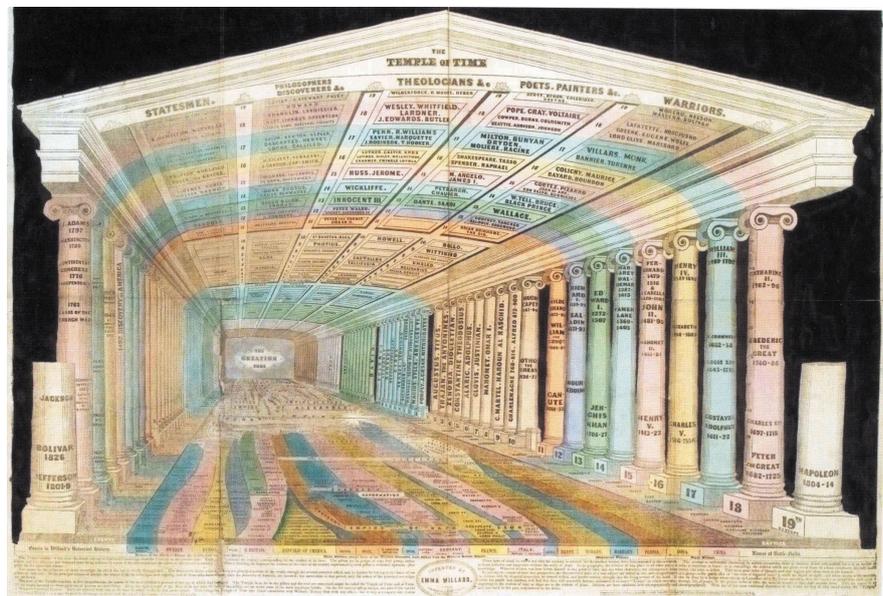
The UK Guardian developed a dynamic timeline that supports mapping more complex courses of events. They have used it to visualise a variety of topics, such as the Middle East protests (Blight et al.

2012, Figure 3.10), the UK riots (Blight et al. 2011) or the Eurozone crisis (Mead & Blight 2014). Running on Adobe Flash, a technology not available on Apple mobile devices, its use declined with growing mobile access to the Guardian website.

Events, corresponding to newspaper articles are assigned different categories and represented by colour coded icons. All icons are arranged on individual bands according to their geographic region. Unlike the columns in Bruyère's *Mappe-Monde*, the bands are sorted not by geographical distance, but alphabetically - presumably a pragmatic decision to allow the arrangement to be more easily generated automatically. Together, the bands form a timeline which extends in a perspectival projection from the virtual viewpoint of the user into the screen, curving upwards as it proceeds into the future.

Figure 3.11 – Emma Willard's *Temple of Time* (1846) is a timeline of history that employs a three-dimensional perspective. Each column of the temple marks a century, with the viewer looking towards the past.

image: American Antiquarian Society



The perspectival curvature in the timeline is quite untypical, yet effective in several ways. First of all it allows a user to better see the distribution of events in the future, which would otherwise be hidden behind each other would the timeline simply continue straight. Secondly, a kind of subjectivity is introduced by the use of the first-person perspective, locating the viewer 'within' the course of events and, perhaps, communicating that one is not looking at a neutral representation of history, but at one particular viewpoint thereof - see also Boyd Davis (2009).

Such spatial representations, with time moving from the present into distant space can be traced back to Emma Willard's *Temple of Time* published in 1846 (Figure 3.11). Kullberg and Mitchell (1995) as well as Korralo et al. (2012) have studied the effects of three-dimensional renderings of digital timelines and both argue for an increase in information recall as a result of the 3D rendering. In the wider discussion of

best-practices in data visualisation however, the use of ‘unnecessary’ or faux 3D - extra dimensions that are not used to show additional data - is generally discouraged (Shneiderman 2003; Few 2007) and experimental studies found its effect on readability to be adverse (Zacks et al. 1998; Cockburn & McKenzie 2002; Hicks et al. 2003), or to be dependent on the implementation and task at hand on (Levy et al. 1996; Wiss et al. 1998; Ridsen et al. 2000; Smallman et al. 2001).

In the Guardian timeline we are able to navigate through time by dragging a horizontal slider at the top of the visualisation or by manipulating a prominently placed handle which causes the visualisation to move forward or backward through time, like a train on tracks. Hovering over an icon highlights the corresponding band and presents a preview of the newspaper article, which can be accessed with a click of the mouse.

With respect to the qualities as a tool for visual analysis there are a few things worth highlighting in the Guardian timelines. The upwards curving of the bands allows us to make out patterns in the dataset, albeit only over a few weeks into the future. The past moves out of view as we navigate the visualisation and an overview of the entire dataset is not available.

In contrast to the NYT timelines, all records are representative of news articles, hence the provenance of the records is transparent and can be traced back to their source. However, we do not know whether the dataset is complete, if in fact all relevant articles are represented or if the dataset has been curated. We also need to trust the authors on the classification of the nature and geographic region of the events, which might be biased both subjectively as well as through the visualisation, which does not accommodate for events occupying several categories or locations.

An important observation which is not unique to this implementation, is that we are not looking at a “timeline of Middle East protests”, but at a timeline of *reports about* the Middle East protests. What is represented are not the protests themselves, but data about the protests - a seemingly obvious distinction, but one that is often overlooked.⁹ Even Priestley, usually remarkably conscious of the relationship between data and display falls into this trap, making no distinction between lack of data and lack of events in one part of his interpretation:

The thin and void places in the chart are, in fact, not less instructive than the most crowded, in giving us an idea of the great interruptions of science, and the intervals at which it has flourished. (Priestley 1764, p.24)

⁹ The “attractive” appearance of a graphic in contrast to, for example, a tabular chronology is often held responsible for this omission; or in the words of Drucker: “the sheer power of the graphical display” (2011b, sec.5)

We have seen how already limited functionality can be helpful for analysing datasets through visualisation by providing detail on demand, maintaining provenance to the underlying data, and offering different views on a dataset by allowing users to manipulate the domain and scaling of the time axis.

There are a few ways of manipulating the display that one might expect, but which are not supported by the Guardian timeline. It is not possible to select a geographic region and have all the corresponding events highlighted, nor is it possible to highlight or filter the data based on a category. Labels and legends act purely as output devices and do not, as Ahlberg and Shneiderman (1994) recommend, double as user interface elements.

This lack of ways for manipulating the display make this visualisation fall short of a truly exploratory timeline, although one can easily imagine such functionality being added. Part of the purpose of discussing examples along this spectrum is precisely to identify how we can push timeline visualisations further.

Exploratory Timelines

Exploratory timelines allow for a greater level of engagement with a visualisation than dynamic timelines, by providing additional interfaces for manipulating the data and its visual representation. The term refers to Tukey's (1977) proposition for exploratory data analysis (EDA). EDA outlines a working method with datasets which is not primarily aimed at confirming predefined hypotheses. Instead, datasets should

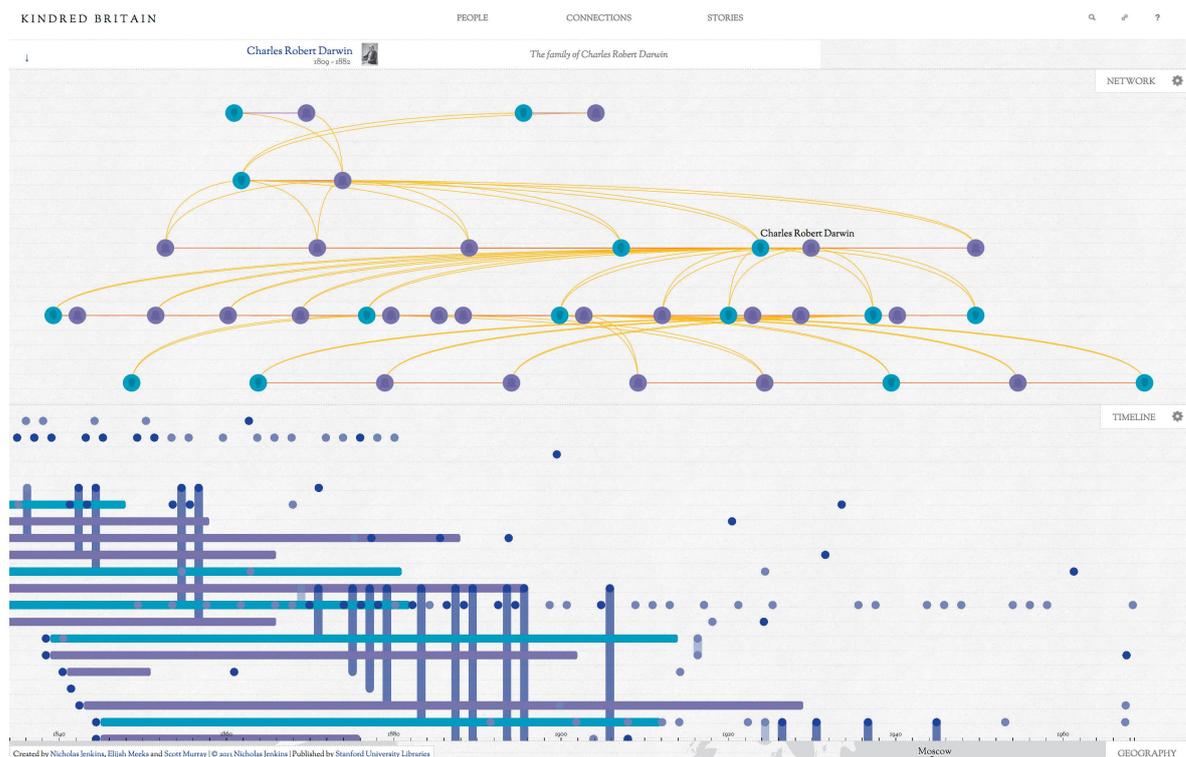


Figure 3.12 – Selecting an individual in *Kindred Britain* visualises its family tree both as a network and in a timeline.

image: screenshot of <http://kindred.stanford.edu/> (accessed 20.08.2015)

be approached with an open mind and hypotheses formulated based on observations of the datasets, not prior knowledge. Tukey describes and introduces a number of graphical techniques, suggesting that visualisation may be an ideal method for EDA. Pousman et al. see exploratory analysis as an essential prerequisite for visual analysis:

Analytic insights come from exploratory analysis, extrapolation, and consist in the large or small eureka moments where a body of data comes into focus for a user. (2007)

While my selection of dynamic timelines originates from the field of journalism, we increasingly see explorative timelines appearing as part of, or even constituting scholarly publications.

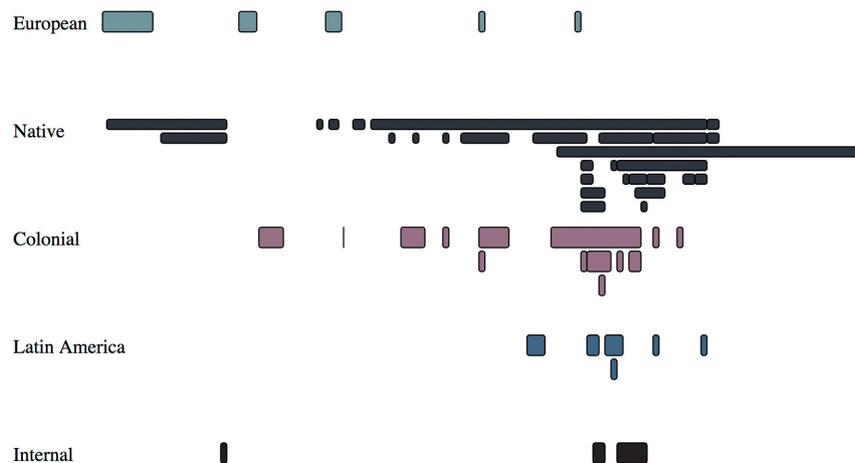
Kindred Britain (Jenkins et al. 2013, Figure 3.12) has been developed by the Stanford University Libraries and visualises a detailed dataset of nearly “30,000 individuals - many of them iconic figures in British culture” (Jenkins et al. 2013). The timeline visualisation forms one of

three visual layouts through which the dataset can be examined - the other two are based on a network visualisation and a geographic map.

Time is plotted linearly from left to right and records are represented as dots, signifying individual events, and bands for visualising the period of individual's lifetimes. Zooming out of the timeline hides the dots, leaving only the bars; an interaction paradigm known as 'semantic zooming' which adjusts the graphical representation based on the zoom level, in this case, by removing less relevant data.

Figure 3.13 – The timeline layout of *Kindred Britain* makes use of a space-saving layout to position individual bars and is available as a reusable D3 layout.

image: screenshot of <http://bl.ocks.org/emeeks/3184af35f4937d878ac0> (accessed 14.01.2016)



The vertical arrangement follows largely the temporal sequence of events. Bars are moved to the top where gaps would otherwise appear, resulting in a more condensed layout, but also in unrelated events appearing on the same row. Similar layouts are often used in both analogue and digital timelines in order to arrive at a more space-saving display than a Gantt-like timeline. Elijah Meeks later released the timeline component of *Kindred Britain* as a standalone D3 layout¹⁰ (Figure 3.13) and it quickly attracted the attention - and prompt rejection - of fellow developer Ethan Jewett, who works on the Palladio project.¹¹ His critique was that the algorithm reacts with drastic changes in the layout even when the data only changes minimally, making him go back to “a sort of ‘step’ layout” - a Gantt-like chart.¹²

Despite my reservations about the Gantt-layout, we see how a pragmatic space-saving layout may be even more problematic for scholarly analysis. In a Gantt-like layout the vertical position of an item is determined by the ones that precede it; a behaviour that is transparent and predictable. Space-saving algorithms are influenced by numerous factors, such as the temporal distribution of all events in the dataset, the graphical space available and the inner workings of the layout algorithm: factors that are unpredictable and opaque, not exactly the hallmarks of a scholarly research tool.

10 See the glossary entry on page 339 for more information on D3.

11 I will discuss Palladio in the next section.

12 The conversation took place on Twitter and is archived here: <https://storify.com/fkraeutli/d3-layout-timeline/> (accessed 08.02.2016)

In the Kindred Britain visualisation, only 74 events are visible at any one time, allowing only a tiny fraction of the entire dataset to be examined on the timeline. It is very likely that a low number of records is visualised partly for performance reasons, partly because the focus of the visualisation is in the comparison of only a few individuals. Why we are presented exactly with this number of records and on which basis they are chosen remains undisclosed.

If we do select one or two individuals for comparison, the timeline gains vertical bars that denote family relations between individuals through marriage or birth. We are in fact presented with a hybrid visualisation that represents both temporal and hierarchical network data. Unlike the network view that connects events through thin lines, the timeline uses bars of the same width as the events to indicate relationships. This graphical equivalence between events and connections is rather unusual when compared to similar examples that visualise hierarchical network data on a single time axis (Card et al. 2006; André et al. 2007; McClure 2014; Müller et al. 2015).

The exploratory timeline allows for insights which were not all anticipated by the designer of the visualisation:

Kindred Britain is also not about supplying definitive answers.

Instead, on a broad foundation of historical fact, it deploys modern computational methods to suggest possibilities, metaphors, ideas about Britain, about its subjects and its culture. (Jenkins 2013)

Exploratory features do not rule out the presence of narrative elements. Kindred Britain contains suggestions for comparisons, as well as a number of essays. Periscopic's timeline visualisation U.S. Gun Deaths (Periscopic 2013, Figure 3.14), a different example of an exploratory timeline, begins with a text-based narration explaining how an individual person killed by a gun in the US is represented as an arc on the timeline, slowly building up the entire visualisation of all gun deaths in a particular year. Detailed filtering options then enable a viewer to dissect the dataset and make comparisons. This is in line with an observation made by Boonstra et al.:

The difference between explorative visual data analysis and visual presentation is small, however. Similar tools can be used in both ways: if a special technique allows researchers to explore and interpret their data well, it will serve as a means to present their data efficiently as well. (2004, p.73)

At first sight, this visualisation uses a familiar timeline format, with events being plotted on a horizontal time axis running from left to right, except for the arcs that extend above it. Looking at the tem-

poral description of individual records, we notice that each record is mapped using three temporal descriptions: the year of birth, the year of death, and the hypothetical year of death if the individual had died of natural causes. To make comparisons easier, the data is normalised and arranged by age.

Unfortunately, the arc rendering of the events and the over-plotting caused by the relatively large number of records visualised (ca. 10,000) turns out to be uninformative. As an alternative view to the timeline display, the data may also be visualised as a set of two histograms representing the distribution of the actual and potential lifespans of the victims. In this quantitative view we can often spot patterns which the arc-layout tends to hide. For example, white southwestern victims account for 14% of the deaths, which corresponds to the graphical proportion of the histogram above the timeline. In the timeline view however, the top set of arcs only looks slightly smaller than the bottom one. We also fail to spot temporal patterns in the timeline visualisation that are apparent in the histogram, such as the extent and ages of peaks. Representing a large number of items individually on a timeline in a meaningful way is a challenge which we will see recur.

Exploratory timelines that are built around specifically curated datasets allow for the inclusion of rich background information, stories and customised interactive features. Kindred Britain makes full use of this by turning to visualisation as a way to convey the complexity of the data. It does represent a humanities research output - the 'traditional' scholarly work has been done by Nicholas Jenkins, accumulating the genealogical data on the represented individuals. But instead of capturing the knowledge he gained in a paper or a book, he collaborated with developers Elijah Meeks, Karl Grossner and Scott Murray to create an interactive visualisation that allows others to make sense of the data.

Both examples show how an exploratory timeline may enable different perspectives on a dataset and allow various narratives to unfold. They also show how a single dataset allows for various insights, merely by manipulating it using a custom interface. This is maybe one of the most powerful abilities of visualisation - to facilitate interpretation - but it also makes evident the need for transparency in the visualisation process.

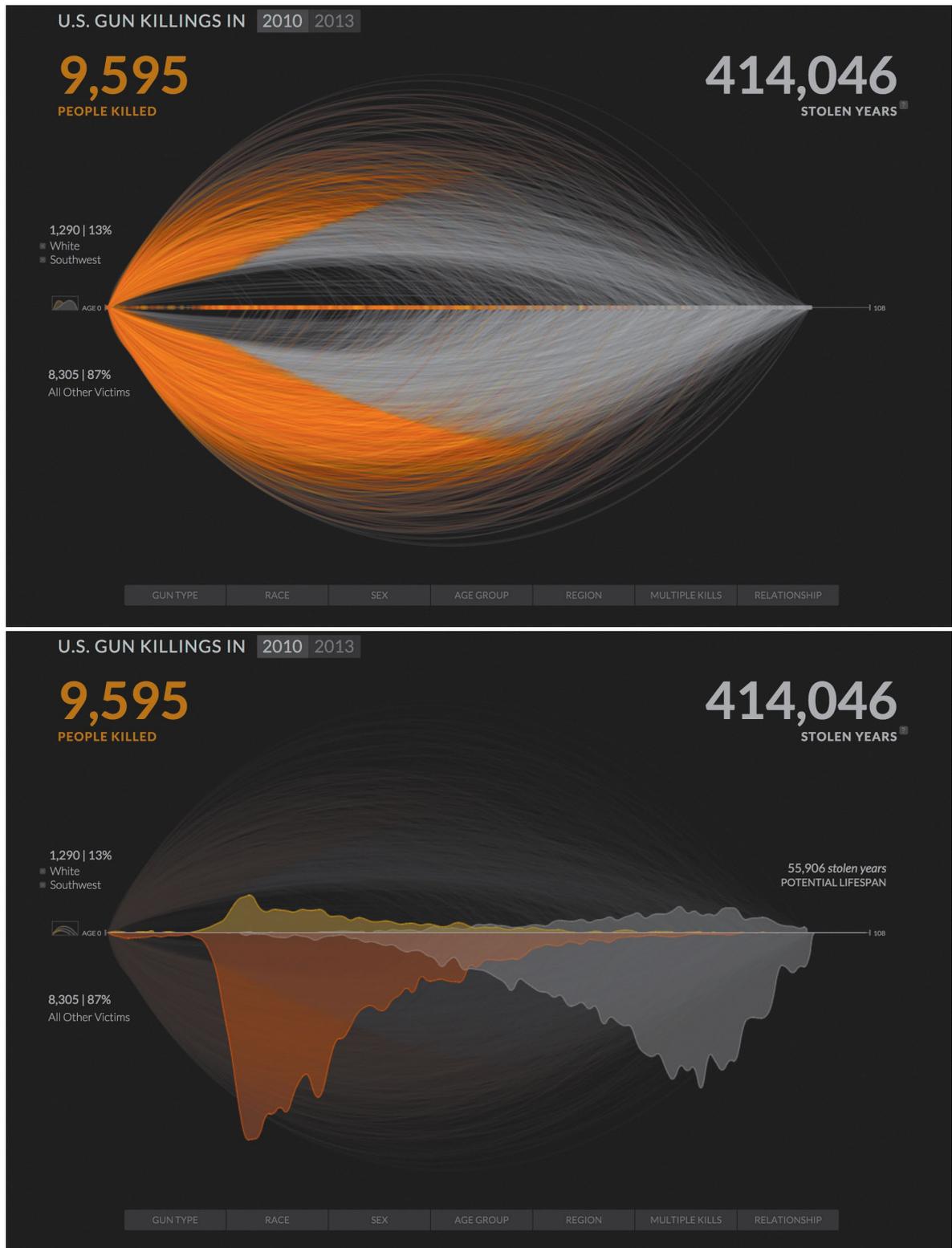


Figure 3.14 – Periscope’s visualisation U.S. Gun Deaths draws each killed individual as an arc. Summative patterns become hard to discriminate, which is probably why the visualisation includes an additional histogram view.

image: screenshots of <http://guns.periscope.com/> (accessed 14.01.2016)

Open Timelines

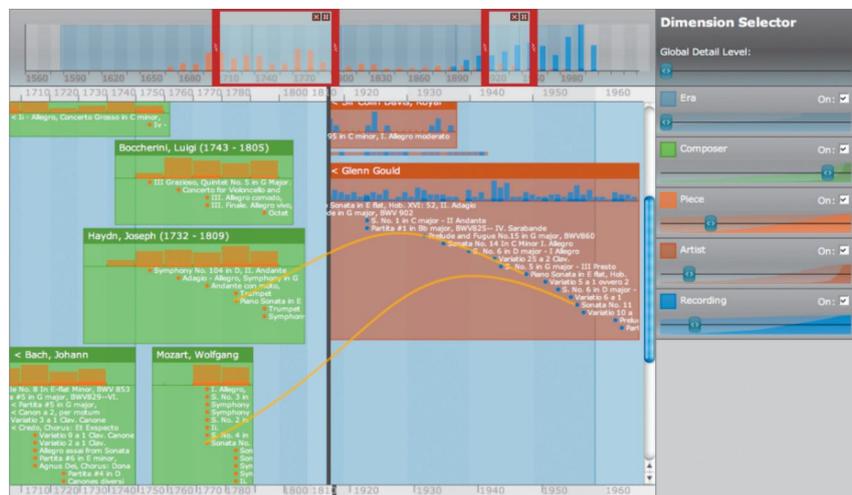
Open timelines should allow for custom datasets to be visualised and explored so that the visualisation can truly serve as a research tool. Such software is crucial for scholarly research in Digital Humanities as only when both the tools and the examined datasets are open, can we produce reliable and verifiable insights. As Borgman writes:

Until analytical tools and services are more sophisticated, robust, transparent, and easy to use [...], it will be difficult to attract a broad base of interest within the humanities community. (2009 §5)

Open timelines, on the spectrum that I employ here, satisfy all prerequisites of static, dynamic and exploratory timelines, with the most important addition of added transparency and reusability.

Figure 3.15 – Continuum visualises a single dataset in a split-view timeline, facilitating comparisons over arbitrary timeframes.

image: André, Wilson, Russell, et al. 2007



Once again, my selection of projects here is not exhaustive; there are timeline tools I will not discuss that would qualify as open, notably vis.js or Timeline, both open source timelines that support various modes of browsing and searching, but are clearly designed for presenting, rather than analysing data.¹³ The timelines I do discuss here all have been designed for the purpose of doing research.

13 See the list of digital timelines in appendix C on page 270.

Continuum (André, Wilson & schraefel 2007; André, Wilson, Russell, et al. 2007, Figure 3.15) is an exceptional timeline tool, not only with regards to the thinking and features it implements, but also in the context of the present selection of tools. Unlike the other examples I discuss, Continuum is not and has never been released. Strictly speaking, it therefore should not be included here, but it would be a gross omission not to discuss it and I will refer to this particular timeline tool later on when I present my own prototype implementations.

Continuum was developed as a response to the lack of sophisticated and reusable timeline visualisations for data analysis and is conceived as a timeline “for hierarchies, relationships and scale” (André, Wilson,

Russell, et al. 2007) - responding specifically to the challenge of visualising large datasets. This is facilitated through a diagram that includes summative histogram views of data within a conventional timeline format. Relationships between individual events can be identified through lines that connect matching entities. This is possible even over a large timeframe as the visualisation implements two separate views of a single timeline, omitting an arbitrary timeframe in between.

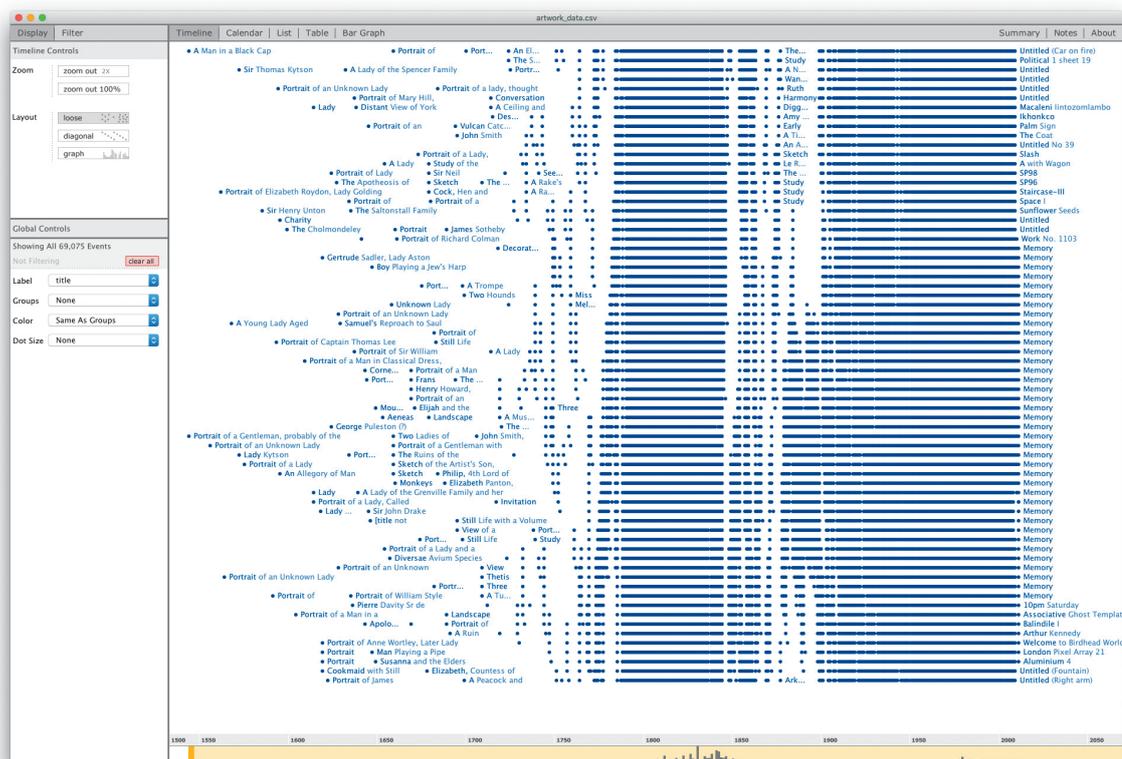


Figure 3.16 – Tate’s artwork data visualised in TimeFlow, a reusable timeline tool intended for journalistic use.

A comparative study between Continuum and the SIMILE timeline, which I discuss below, verified the claimed advantage as a visualisation tool to analyse large cultural datasets (André, Wilson & schraefel 2007) – a movie database containing 11,000 records. Despite plans of releasing the tool into the public domain, development of the tool appears to have ceased and it has never been made available.

This makes TimeFlow (Viégas et al. 2010, Figure 3.16) one of the few publicly available visualisation tools which is dedicated purely to visual analysis of arbitrary data on a timeline. TimeFlow was developed as a “visualisation tool for analysing temporal data” aimed at data journalists. Other data visualisation suites and libraries¹⁴ put a stronger emphasis on non-temporal and quantitative data and most tools dedicated to temporal data tend towards a narrative, rather than analytic use.

14 For example IBM’s ManyEyes (IBM 2007) or Tableau (n.d.).

The open source Java application accepts datasets in the tabular CSV format via an import screen that automatically assigns one of five types to each field (Number, Date/Time, Text, List or URL). After import, the data is visualised along a horizontal time axis using a “loose” layout where “the vertical location of data points is flexible” (Flowing-Media 2010b). Other supported layouts include a “diagonal” arrangement, which uses sequence of events to determine vertical positioning – starting back at the top when records would fall out of view at the

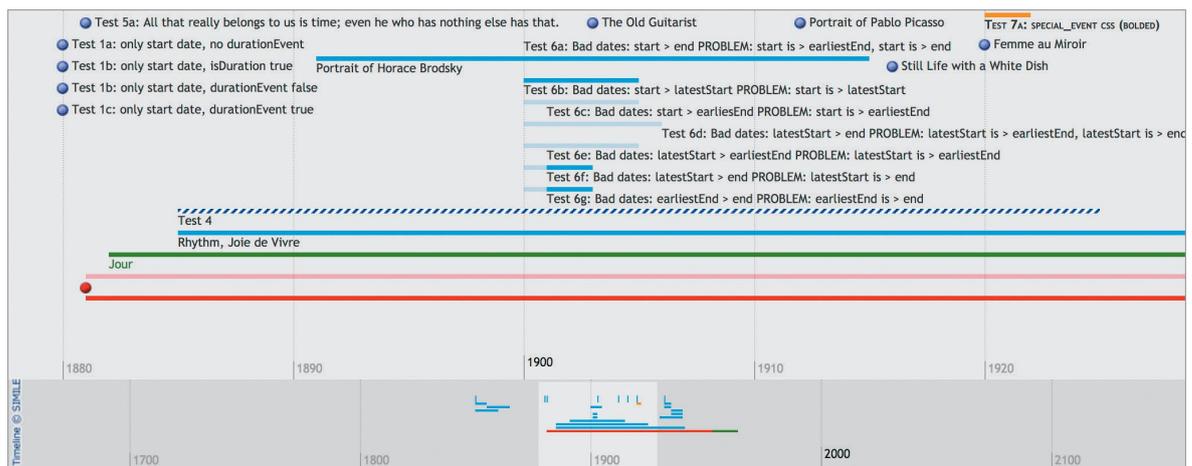


Figure 3.17 – An example of the Simile Timeline, demonstrating all possible event descriptions and representations.

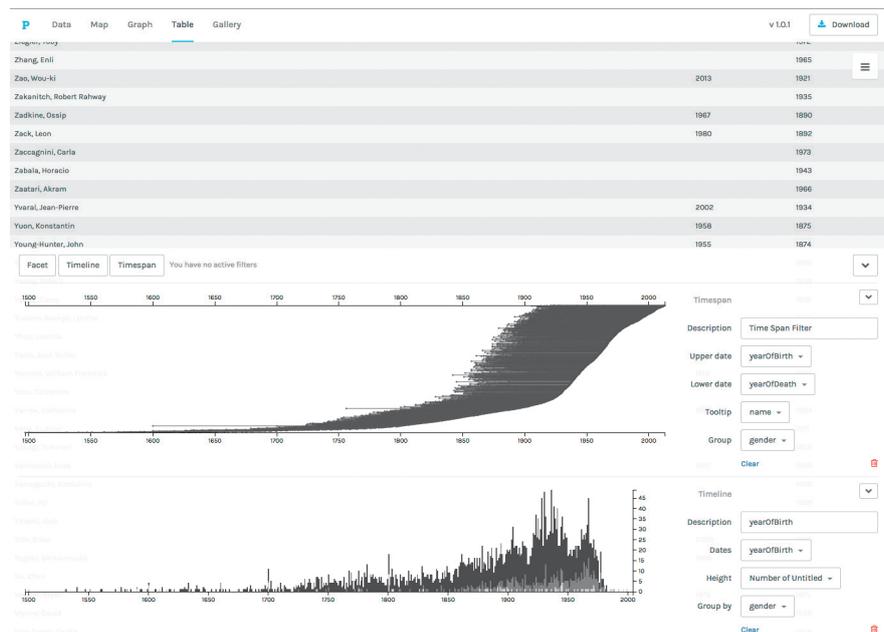
image: screenshot of http://www.simile-widgets.org/timeline/examples/test_example/test.html (accessed 17.10.2014)

bottom – and “graph”, essentially an aggregated histogram view. It is possible to group events on horizontal bands and assign colours according to a chosen (text or date) field, as well as map a numeric value to the size of the dot used to represent individual events.

There is no forced limitation on the size of the datasets that may be visualised, but the authors state that a considerable decrease in performance is to be expected with datasets of more than 20,000 rows (FlowingMedia 2010a). While performance did suffer a bit, I had no trouble importing a dataset of almost 70,000 records. Nevertheless, the timeline layouts did no longer prove to be informative when more than a few hundred records were in view. As the display becomes cluttered it is very difficult to distinguish individual events and overall patterns that might be present in the dataset are no longer visible.

SIMILE Timeline (Huynh 2009, Figure 3.17) is arguably one of the timeline tools with the largest user base, partly also because of its integration in other scholarly projects, such as Neatline (Nowviskie et al. 2012; Nowviskie et al. 2013). SIMILE was a project that ran at the MIT from 2003 until 2011 and aimed to develop reusable interactive tools for the presentation and exploration of digital assets. The tools have since been released and are partly developed further as SIMILE Widgets, a collection of open source web based data visualisation tools. SIMILE Timeline’s basic layout is a horizontally scrolling timeline

Figure 3.18 – Palladio allows users to explore arbitrary datasets – here, artist data from the Tate. Time-based visualisations form part of two of the three available filters. image: screenshot of <http://palladio.designhumanities.org/> (accessed 20.08.2015)



populated either by reading an XML file or by dynamically loading events in the JSON format. This architecture allows the timeline to be manually curated, as well as potentially being generated through an existing dataset.

The designers have dedicated a remarkable level of consideration to the kind of temporal descriptions the timeline is able to visualise. In contrast to other timelines, which only support singular events and periods, it is possible to include elements of uncertainty by specifying a latest start or earliest end date to events. The appearance of individual events can be further customised and additional information about events can be revealed by clicking on them. In its later versions the tool also allows basic filtering and highlighting based on entered search terms.

Another unique feature is the ability to inflect the time axis by defining periods with a different temporal granularity than the rest of the visualisation. “Hotzones” (SIMILE 2009) are intended to locally zoom in on busy time periods, where events that closely follow each other can otherwise not be discriminated. These distortions have to be predefined and cannot be changed interactively by a user of the timeline. Like most timelines, SIMILE Timeline is primarily intended as a means for telling an authored story, rather than exploring unstructured and potentially large datasets.

The last project I’d like to discuss within this section is Palladio (Figure 3.18) by Stanford’s Humanities+Design Lab (2014), an application dedicated to visual analysis of datasets for humanities researchers. Palladio’s origins lie in a research project on the Republic of Letters (Chang et al. 2009), which included an exploratory visualisa-

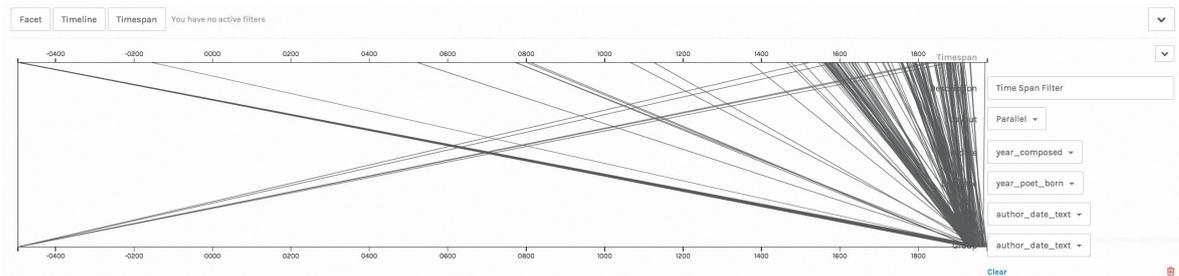


Figure 3.19 – Palladio’s Timespan filter can be coerced to show temporal relationships. However, the fixed scale of the time axis causes problems with unevenly distributed data.

image: screenshot of <http://palladio.designhumanities.org/> (accessed 28.01.2016)

tion tool. The working principles and the insights gained through the development of this tool have been repurposed and made reusable.

Palladio is not primarily a timeline visualisation tool – its focus is clearly on network and geographical visualisations. In fact, two dedicated timeline views (titled ‘Time’ and ‘Duration’ – see Figure 3.20) that appeared in a late beta version (Stanford Humanities + Design Lab 2015a) were removed at the release of the final version only a few months later (Stanford Humanities + Design Lab 2015b) – for undisclosed reasons.

Temporal visualisations have remained in the tool not in the shape of dedicated views, but as interfaces for filtering datasets. The “Timeline” filter generates an aggregated histogram view, which can be grouped by any of the available fields. A conventional timeline visualisation with events represented individually as horizontal lines hides behind the “Timespan” filter. Clicking and dragging anywhere inside these filter views creates a brush, which allows the datasets to be filtered dynamically based on the chosen timeframe.¹⁵

Both temporal filters employ timeline representations established by Priestley,¹⁶ but in earlier releases of the software we can see how the creators experimented with different ways of visualising temporal events. The timespan filter offers an alternative layout formed of two parallel time axes with events plotted as connections from their beginning dates at the top to their end dates on the bottom axis. This layout was, for a short time, also available as a dedicated view to analyse the dataset.

A similar layout has been used, for example by Thudt et al. (2012) to explore patterns between the publishing dates of books and their fictional time periods. In Palladio, however, the upper and lower time axis are fixed and synchronised – the time axes can not be scaled or shifted – which makes comparisons across larger timeframes difficult.

It might indicate a bias which originates from Palladio’s ancestor project that was built around a dataset concerned with individual’s lifespans over a common timeframe. Figure 3.19 is a screenshot of the Timespan filter in Palladio set up with one of the datasets I will visualise later on – the works of Benjamin Britten. The dataset contains

15 A ‘brush’ is an interaction paradigm where users select a subset of values on a numerical axis by defining an upper and lower bound – for example beginning and end dates. The selection then may be dragged along the axis, resulting in different portions of the dataset gradually appearing and disappearing – this process is also called ‘brushing’.

16 Events represented as lines and arranged on a horizontal arithmetic time axis.

works with their year of composition as well as associated writers and their birth year. Visualising both dates in Palladio's TimeSpan filter scales the entire diagram to the larger timeframe of the writers, losing all detail in the smaller timeframe that the works occupy.

This raises the question of how far a timeline visualisation can be truly reusable as it is impossible to anticipate every possible use and dataset. It is a question we may not be able to answer, as every implementation exposes new limitations. However, by opening up visualisation tools and making them generalisable, as the creators of Palladio have done, we can not only get closer to such tools, but also learn more about the biases we might unknowingly impose by expecting a certain kind of datasets and by asking certain questions.

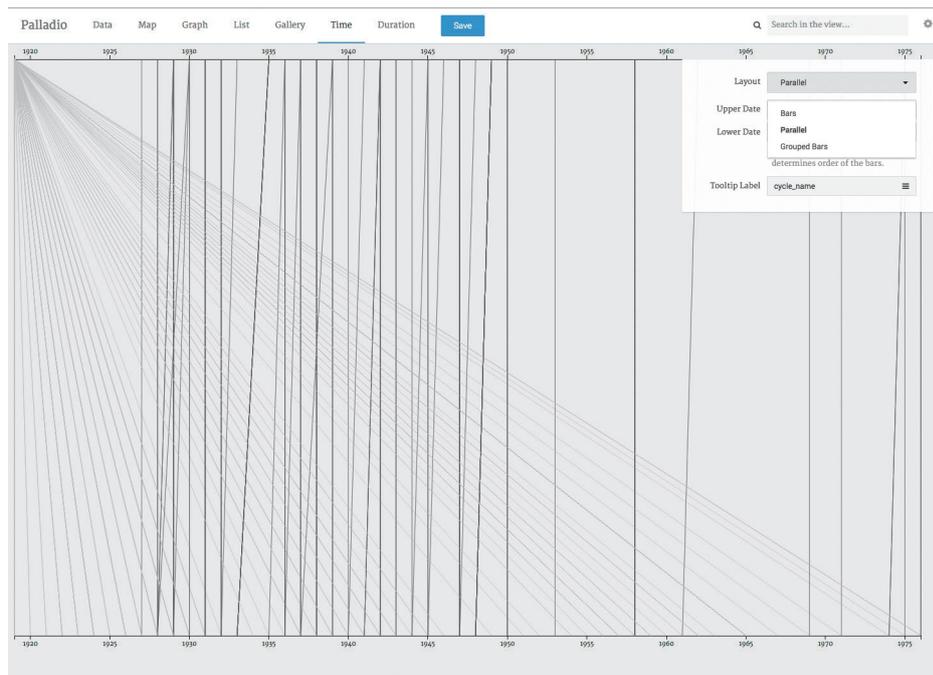


Figure 3.20 – In version 0.9 Palladio introduced two time-based views “Time” (pictured here) and “Duration”, which have been removed again from the final version.

image: screenshot of <http://palladio.designhumanities.org/> (accessed 28.01.2015)

Discussion

Before I summarise the main findings of this project review, I briefly summarise how the different branches of my criteria tree have mediated my appraisal of the discussed examples and reflect on the implications of my employed spectrum on the definition of a timeline as a visual analytics tool.

Findings Led by Criteria

The “Data Acquisition and Curation” branch caused me to focus on the ‘expectations’ that a visualisation format may have towards the dataset it visualises. Visualisations that accept custom datasets often only support a very limited set of fields, while examples that are tied to a single dataset suggest that there are many more attributes that might be useful to visualise on a timeline – especially with regards to temporal descriptions as I will outline below. How might a timeline tool be able to make use of all temporal dimensions of a dataset?

The criteria under “Representation of Dataset” and “Representation of Records” delivered the most useful observations, resulting in the focus issues described below. Distinguishing between how records are represented individually and how they form an overall representation of the dataset, revealed how many timeline visualisations exhibit problems with the latter. Even when great attention is given to the graphical rendering of individual records, as for example in SIMILE, patterns across several records are often unreadable.

Looking at the examples in terms of “Interaction” has revealed issues that are more relevant for individual implementations than for the class of timelines as a whole. Most timelines are able to provide “details on demand” (Shneiderman 1996). Only few, however, offered the ability to operate on several records, indicating again a lack of focus on the totality of visualised events.

In terms of their “Technology and Platform”, most digital visualisation represented the prevalent state of the art at the time of their creation. Currently, web-based applications or libraries have replaced earlier Java-based standalone software. A development which is beneficial for scholarly research, when it leads to software that is based on open standards and/or open source practices.

Findings Led by Spectrum

I have already indicated that I do not mean to suggest that the fitness of a digital timeline tool for scholarly research and visual analysis automatically follows from the presence of certain interactive features. Still, I would like to highlight some of the technical aspects I have

used to distinguish between static, dynamic, explorative and open visualisations on my spectrum which, after reviewing the examples, I think do make a difference.

Static timelines may be created without any coding experience, but that does not make them necessarily an easier format, as it means having to include all data in one display and deciding on one way of representing a dataset. Dynamic timelines on the other hand may offer different views on a dataset, each highlighting certain aspects. Besides enabling different insights, alternative views may reveal and emphasise that some patterns that are visible in a visualisation may be a side-effect of a particular layout, scaling or graphical representation.

Even limited interactive features may significantly improve the value of a timeline as a research tool by offering secondary information on demand and by providing links between the graphical representation of data and its original source. Ensuring that the underlying data remains accessible is a key issue in data visualisation (Kasik et al. 2009) as well as in the wider context of Digital Humanities (Flanders & Muñoz 2011). The timelines I have presented as *dynamic* support this kind of “details on demand” (Shneiderman 1996) and source linkage.

Individual records may be interrogated as such, but what dynamic timelines fail to offer is a way of asking questions about the dataset in its entirety or about larger groups of records. In exploratory timelines these features include the option to colour records based on specified field values - as in *Kindred Britain* (Jenkins et al. 2013) - or the ability to select and compare data that fits certain criteria against the remaining dataset - as in *U.S. Gun deaths* (Periscopic 2013).

Open timelines offer in terms of interactive features not much more than exploratory. Arguably, they may often even be less sophisticated as they cannot take advantage of certain fields that may or may not be present in any particular dataset. They do however feature mechanisms and interface elements required for importing data and assigning fields to graphical dimensions of the visualisation. These make a crucial difference, especially with regards to temporal data. In an open timeline we have to tell the visualisation which field contains the data that governs the position of its graphical representation on the time axis. This leaves us with the option to use a different field than the authors of a tool might have anticipated; for example, to visualise objects by acquisition or cataloguing date, rather than production date. *Palladio* requires two dates to be selected, expecting that these will represent beginning and end dates of a single event, yet we can also choose dates with no apparent relationship and use the visualisation to examine if we can discover one. When a visualisation functions as a tool, it can not only be used, it may also be productively misused.

Focus Issues

This review of timeline visualisations has offered a snapshot of the diversity of implementations available and their various (interactive) features, which may support a use for visual analytics. It has also uncovered a range of embedded assumptions in the design of timelines and how events should be represented on them, along with the potentially counterproductive effects of adopting particular graphic layouts due to their ease of computational implementation or space-saving properties.

The main challenge that emerged from my previous studies of the literature is the difficulty of accounting - within timeline visualisation tools - for the uncertainties, biases and subjectivities that are embedded within digital collections as a whole and in date descriptions specifically. Through my examination of implementations of digital timelines and visual analytic tools, I have further refined my analysis and identified practical challenges which will guide my prototype-led explorations. These issues concern the analysis of large scale datasets, the algorithmic generation of appropriate timeline layouts, and the modelling and representation of temporal descriptions in cultural datasets.

Scale

Perhaps the most far-reaching consequence of digital technology in the humanities context is an unprecedented increase in the size of collections that historians, archivists and archeologists are able to process:

When digital technologies allow for the storage and analysis of millions of books, billions of tweets, and hundreds of billions of interactions, the ways in which we can query and comprehend the cultural record explodes. [...] we will have to design and employ new tools to thoughtfully and meaningfully sift through, analyze, visualize, map, and evaluate the deluge of data and cultural material that the digital age has unleashed [...] (Lunenfeld et al. 2012, p.38)

Challenges around visualisations of large datasets are a recurring topic in the data visualisation literature (Lamping & Rao 1996; Fua et al. 1999; Keim 2001; Shneiderman 2008). None of these however focus on visualising large datasets on timelines. In the context of data visualisation time is often treated as “just another dimension” (Kosara et al. 2004), neglecting the differences between time and typical scalar quantities, but also missing out on the opportunities that time might offer for visualisation.

So far, the issue of large datasets in timelines has mainly been addressed in Continuum (André, Wilson, Russell, et al. 2007; André, Wilson & schraefel 2007), whose approach includes a histogram based overview timeline and limiting the number of events in view in the main timeline. Other timelines have implemented both strategies, albeit separately, and I will revisit and build on these approaches in my own prototypes.

Large datasets are particularly challenging for timelines. In my own definition I have pointed out that timelines, in contrast to other visualisation formats, do not need to make quantitative summaries of data in order to visually represent it. Events can be studied individually and being able to examine the temporal position, succession and relationship of single events is one of the main benefits of graphical timelines.

However maintaining individual representations of events creates problems when large numbers have to be visualised in a small amount of graphical space: a common problem of the timeline visualisations that I have reviewed here is their inability to handle large datasets elegantly, resulting in cluttered and unreadable displays.

The challenge is therefore to either develop a suitable layout that maintains readability of individual events or to find a suitable way of summarising events on a timeline - which could be a histogram or a novel diagram format.

Layout

The issue of large datasets is intrinsically linked to the challenges around timeline layouts. In the digital realm, timelines are often understood literally as lines. Time is not only treated as a one-dimensional mathematical space but also represented as such graphically, as is visible in the timelines by the NYT as well as the work by Periscopic. This circumvents a problem I have hinted at in my discussion of Accurat's *Empires* timeline: the significance of the orthogonal graphical dimension.

Bruyère (1750a) underlines how every point in his chart carries significance due to the respective mapping of geographical and temporal dimensions to the horizontal and vertical axes of the graphical space. The Guardian timeline similarly uses the non-temporal axis to indicate the geographical region of an event.¹⁷ In these examples, the position of a record map on to a field value. This, however is not always the case in graphical timelines that visualise records on more than one graphical dimension - timeline layouts often are *non-deterministic*.

A scatter plot is a *deterministic* visualisation layout - the position of every graphical representation of a record forms an arithmetic rela-

¹⁷ In contrast to Bruyère's chart, the order of regions in the Guardian timeline bears no relationship to their geographic locations.

18 In the digital context random effects will almost always be 'pseudo-random'; there can be no true randomness generated by a computer. As John von Neumann famously wrote: "Any one who considers arithmetical methods of producing random digits is, of course, in a state of sin" (1951). He did however not argue against pseudo-randomness per se – it is essential for many applications and in the context of visualisation it is used, for example, in some network layout algorithms. Neumann merely warned against confusing pseudo-randomness with true randomness.

tionship with the data being visualised. In non-deterministic layouts, there can be many factors that influence the position of a mark, such as the content and structure of the dataset, the size of the visualisation, the chosen layout algorithms as well as randomness.¹⁸

For network visualisations, a number of different and clearly defined deterministic and non-deterministic layouts exist and their effects and merits are widely discussed (Fruchterman & Reingold 1991; Golbeck & Mutton 2006; Martin et al. 2011; Krzywinski et al. 2012). As Krzywinski et al. write:

Network layouts are difficult to predict and interpret because their creation is in part, driven by an aesthetic heuristic that can influence how specific structures are rendered. (2012, p.2)

In addition, the same layout can produce significantly different visualisations even when the dataset has only been changed slightly:

Most algorithms are sensitive to small changes in the network and create perceptually nonuniform layouts, where differences vary out of proportion to changes in the network. (Krzywinski et al. 2012, p.2)

Although timeline visualisations face very similar challenges, these are generally not discussed and the issue of timeline layouts are not addressed outside very specific implementations.

The mechanisms used to generate the visual layout of a timeline affect not only the position of individual records, but the overall visual appearance and our likelihood of spotting patterns and connections among items that appear close together. When the arrangement of items on the non-temporal axis is arbitrary or undisclosed, the relevance of the position and distance between items on that axis becomes meaningless and overall patterns may be misinterpreted.

Priestley exploited this flexibility in arrangements that the timeline affords by "placing those persons the nearest together who had the most connections, and whom I thought it would be most amusing to compare together" (Priestley 1764, p.19). He could rely on his individual judgement to do so, which digital timelines that are generated through software are unable to do. We therefore need to develop timeline layouts that can be generated automatically and are readable, comprehensible, and robust against changes in the dataset or display size.

Temporal Descriptions

At first sight it appears as if we are mainly faced with two kinds of temporal descriptions in digital timelines: single events and time periods. Single events are defined by a unique date and tend to be graphi-

cally represented as dots (Huynh 2009; Viégas et al. 2010; Delvisicio et al. 2013; Jenkins et al. 2013; Andrews & Parlapiano 2014), dashes (Belopotosky et al. 2011), or icons (Blight et al. 2012). Periods on the other hand consist of two dates, beginning and end dates and are rendered as bands (Huynh 2009; Lee 2011; Beltramin et al. 2012; Jenkins et al. 2013; Stanford Humanities + Design Lab 2014).

Only a few examples make use of different event formats. US Gun Deaths (Periscopic 2013) maps three dates per event - birth, actual death and hypothetical death. Accurat's (Beltramin et al. 2012) and Lee's (2011) timeline visualise empires as periods with beginnings and ends, as well as an additional third date representing the date of maximum expansion. SIMILE timeline even supports four dates per event (period), equivalent to the quadruple proposed by Kauppinen et al. (2010) for describing imprecise temporal information in cultural heritage data: earliest start, latest start, earliest end and latest end. SIMILE however only uses these description for the graphical rendering; they remain inaccessible when interacting with the visualisation. In Lee's interactive timeline we are, to a certain extent, able to pick the date that determines the spatial arrangement, allowing for comparisons in sequence, total duration and durations before and after peak date.

The presented examples employ multiple date descriptions either to express uncertainty or to provide secondary information. I am interested in exploring both types of date multiplicities. Digital timelines predominantly treat events as singular occurrences in time, ignoring the uncertainties and multiplicities in dating that might be associated with real events. Born-digital data reinforces this practice by carrying a unique timestamp marking its creation. Historically, temporal diagrams have not only been used to analyse and understand the 'actual' chronology of events, but also to examine and compare the variety of possible accounts of chronology - for example Lenglet du Fresnoy's *Tables Chronologiques* discussed in Boyd Davis (2015b).

Cultural data such as digital collections, and any other authored datasets, may have a multitude of known and unknown relevant dates which would lend themselves to be explored through visualisation. The question is, how can a timeline make use of different temporal descriptions? What kind of insights could be enabled by timeline visualisations that do so?

4

“By giving people access to the collection, you give them the ability to ask questions of the museums’ work that we inside these museums might never think to ask.”

(Rodley 2014)

4 Prototypes

Building on the theoretical and methodological groundwork established in the previous chapters, I here present how I respond to the specified focus issues through making and evaluating prototype visualisations. They lead me to refine the challenges and finally derive a set of principles for the design of timeline visualisation tools for visual analysis of digital collections.

The prototypes were evaluated primarily through critical reflection and discussion with expert users. In cases where I published initial results online, comments from other scholars in the form of peer critique further helped me in evaluating my prototypes.

Prototype 6 (P6) is evaluated through a controlled user study - see page 155. Initially I expected this to be the preferred method to ultimately evaluate my research output. However I found that this method did not perform as expected in answering my research questions. The discussion of P6 includes a summary of the user testing procedure and my reasons for refraining from further controlled user studies.

The practical work I discuss here developed alongside the theoretical enquiries that I presented earlier. Making, in the form of sketches on paper as well as with code, studies of different forms of visuals, interaction paradigms and algorithms up to the design of fully fledged visualisation tools took place throughout the course of my research.

In retrospect, I can identify three stages in my process. My very first experiments were driven by a 'scientific' enquiry into specified research questions through prototypes that modelled and addressed isolated problems. From there I soon progressed to a more data-driven approach, working closely with existing cultural collections. Finally, collaborating with experts in museums, archives and libraries, enabled me to gain a better understanding of how timeline visualisations are able to address the experts' specific research questions along with how they translate to knowledge production in the context of digital collections more broadly.

Here I will discuss eight prototype iterations that each address some or all of the focus issues by implementing candidate solutions and build on each other by following leads that arose in previous prototype iterations. Each presentation of a prototype is followed by a

discussion of the insights that it enabled and an appraisal of the successful elements that are developed further, as well as their drawbacks and failures that are to be tackled or avoided.

Focus Issue	
Scale	Visualising large cultural datasets on a timeline
Layout	A timeline-format that generates a comprehensible and honest representation of a dataset
Temporal Descriptions	Making use of alternative temporal descriptions in collections data

Prototype 1 – Metaballs

Problem Statement	
Issue	- Scale
Proposed solution	- Horizontal (temporal) clustering through Metaball metaphor - Vertical distribution by dominant category
Dataset	- Generated - Geffrye
Evaluation	- Casual user testing - Critical reflection

Process

My awareness of the challenges posed by timeline visualisations for large datasets initially led me to focus on clustering algorithms. Clustering techniques are used in data analysis to identify groups of records that share a similarity based on set parameters. When applied to visualisation, clustering may be used to summarise several records with one graphical representation in order to minimise the number of elements drawn on the screen, hence reducing complexity and improving readability in visualisations of large datasets. Clustering algorithms have been developed for timelines of photo collections (Huynh et al. 2005) and documents more broadly (Alonso et al. 2009). Some scholars (Goldstein & Roth 1994; Chuah 1998; Fredrikson et al. 1999; Tang & Shneiderman 2001) prefer to speak of ‘aggregation’ in order to discriminate between the visual effect of records accidentally clustering together due to their proximity in a visualisation and the computational method of intentionally summarising groups of data into a single representation. The term aggregation however is similarly ambiguous in the data visualisation context due to its relation to data retrieval, but I will use both terms to refer to the deliberate visual summarising of data representations.

I developed a time-based clustering method built around the principle of two-dimensional Metaballs (Blinn 1982). Metaballs are a mathematical technique for producing circular shapes with droplet-like behaviour; entities that are in close proximity of each other merge into a larger shapes, which makes it a useful concept for visual clustering and has already been applied to a variety of datasets (Rilling &

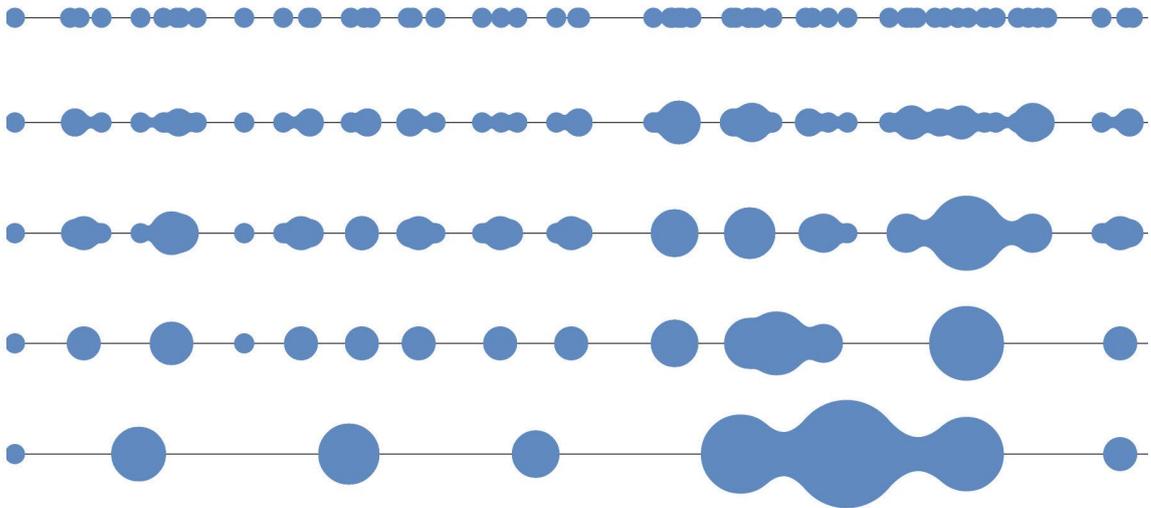


Figure 4.1 – Grouping of contiguous event using a Metaball algorithm on random data.

Mudur 2002; Müller et al. 2007). Admittedly, my decision of adopting this principle was not a purely rational choice. Rather it was heavily influenced by an example script I used to get acquainted with the JavaScript drawing framework Paper.js (Lehni & Puckey 2011); such accidental influences are perhaps typical of a research project that follows an applied, rather than an idealised model.

After a few iterations of tryout visualisations (Figure 4.1), the concept resulted in a prototype application which allowed users to explore the collection of the Geffrye Museum¹ based on given search terms (Figure 4.2). Two clustering methods are at play in this prototype. The Metaball clustering summarises records that are close in time by production date on a horizontal time axis according to a threshold, which the user may adjust via a slider. A second clustering method organises the presented records according to their dominant categories and arranges them by category vertically, starting with the category that contains the most items.

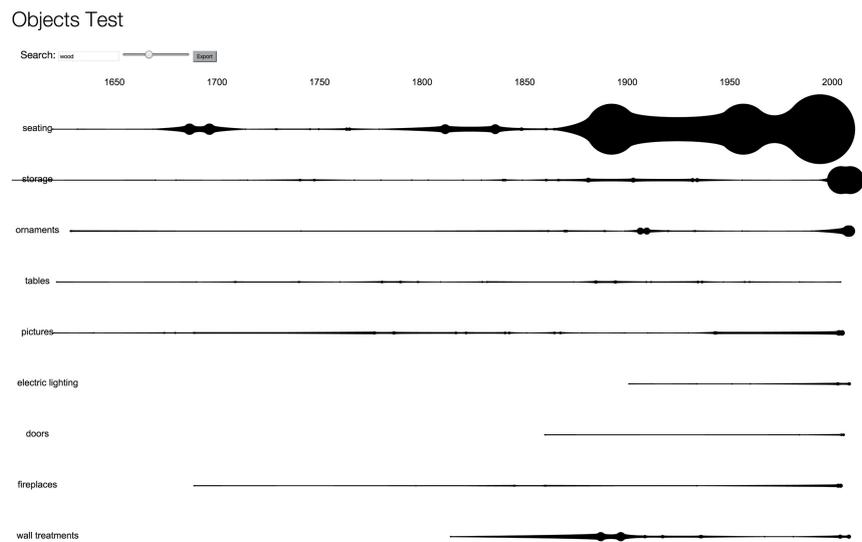
I abandoned the Metaball concept after this prototype because the resulting visualisation did not have the desired effect when applied to this particular cultural collection. In my model tryouts I used randomly generated data with a relatively even distribution throughout time. The Geffrye collection - and most other collections I have worked with so far - display an uneven temporal distribution.² Geffrye’s dataset causes my prototype to overly exaggerate the data in more recent time periods, while earlier items are hardly visible.

The same would probably happen with any dataset of historic events and it is the consequence of adhering to a strictly linear time scale. The left-most part of Barbeau-Dubourg’s 16.5 metre long chronography are almost empty, with the right end of it crammed full of data (Boyd Davis 2016b).

¹ See the complete list of datasets I used in appendix B on page 263.

² With regards to the production dates of their items and their acquisition dates.

Figure 4.2 – A prototype application, based on the Geffrye Museum’s collection, combining time-based Metaball grouping and content-based clustering.



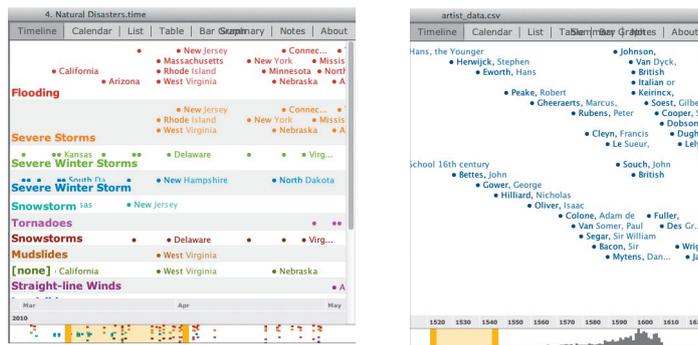
Insights

Dynamic Aggregations

The metaphor of drops of liquid combining and separating made the process of clustering transparent and comprehensible, besides producing a pleasing visual effect - according to casual user trials. In addition, the process worked seamlessly, allowing me to move between representing records individually and as parts of progressively larger clusters. Shneiderman (2008) calls these two ways of looking at a dataset ‘atomic’ views and ‘aggregations’. He specifically recommends including both as separate views in visualisation tools for large datasets in order to be able to examine datasets as a whole in the aggregated view and smaller selections in atomic views. TimeFlow (Viégas et al. 2010) applies this technique in a timeline by using a histogram in the overview timeline when datasets exceed a certain size (Figure 4.3). Continuum (André, Wilson, Russell, et al. 2007) follows a similar approach, while the widely used SIMILE Timeline (Huynh 2009) does implement two separate views for overview and detail, but uses atomic representations for both.

What I found intriguing in my prototype is the ability to use the same visual paradigm for both atomic and aggregated views, in a single display, including the ability to seamlessly transition between the two. The ‘liquid’ metaphor could aid in understanding how the aggregated view is constructed. Histograms on the other hand may be less readable for unaccustomed users. Such a proposition would however have to be verified.

Figure 4.3 – The overview timeline in TimeFlow automatically switches from an atomic (left) to an aggregated (right) representation when datasets exceed a certain size.



Automated Curating

My prototype automatically assigns a category for every record. In this specific case, it is chosen from the website keywords curators have assigned to individual records. A single record may have several keywords, which is why the prototype implementation gives preference to the one that is most frequently present in all the records currently displayed. Searching for ‘wood’ results in most records being assigned the ‘seating’ category, causing it to occupy the top row in the visualisation, followed by the remaining categories in decreasing order of frequency. This enables a user to get an impression of the thematic distribution of a selection along with the possibility to spot patterns. We can see how ‘seating’ is present across the entire timespan of the collection, while ‘electric lighting’ evidently only comes into place towards the end of the 19th century. This, or a similar form of ‘automated curation’ could potentially be very useful.

On the other hand, it could also put a damaging emphasis on dominant themes, causing records in less popular categories to disappear from sight. The ‘long tail’ of a collection may inadvertently get suppressed. One also needs to scrutinise the basis of automated curation. Website keywords, as I have used here, may not be the most unbiased of all possible criteria. A curator may have a lot of time and dedication drafting keywords for one item, while others may be neglected.

Presenting a pre-conditioned display may or may not be desirable for the user. It could appear patronising; users may not be interested in seating at all, for example, yet it gets imposed on them. On the other hand, such a display could equally be engaging, offering a more accessible entry point for users less familiar with the collection. There are arguments (Whitelaw 2012) as well as evidence (Gorgels 2013; Brenner 2015) in favour of the latter view. These authors however look primarily at engaging external visitors of online collections and not at scholarly uses, which require a more ‘neutral’ display.

Continuing

How can a timeline visualise heterogenous data without over-emphasising dominant patterns?

What visual paradigms allow a fluid transition between atomic and aggregate views within a single display?

Prototype 2 – Temporal Projections

Problem Statement	
Issue	- Temporal descriptions
Proposed solution	- Alternative projections of time - Transforming the (time)line
Dataset	- Generated
Evaluation	- Critical reflection

Process

Looking for alternative temporal models and descriptions that a timeline could represent I was prompted by Zerubavel (2004), whose writings on sociocultural concepts of time range from illustrated enumerations of temporal visions and narratives that would lend themselves to visualisation - cyclical, tree-shaped, staccato, legato, etc. - to accounts of historical discontinuities introduced through discrepancies between perceived, narrated and historical time, and time as measured by a calendar. A digital timeline could be used to explore such conflicts and counterintuitive correlations. There exists a body of work directed at visualising patient histories (Plaisant et al. 1996; Park & Choi 2012) or for criminal investigations (Plaisant et al. 1996), which includes the development of alternative visual and computational models of time to account for the complexities and ambiguities of narrated events.

My initial prototypes follow the footsteps of these examples in conceiving novel temporal models to computationally represent diverse accounts of historical time. By considering alternative concepts - non-Newtonian, subjective, relational, branching etc. - my aim was to develop new models that allow timeline tools to exploit and exhibit the complications of time, and, as a result, be more rigorous and honest about the temporal data they represent. I produced a number of diagrams which explored different visual and computational representations of time as well as ways for navigating and manipulating such temporal spaces.

These include a timeline that could be coiled up, in order to examine cyclical patterns within a linear timeline (Figure 4.4), or one that can be folded (Figure 4.5). I experimented with the possibility of a two dimensional model of time (Figure 4.6) as well as branching time (Figure 4.7). Mapping ambiguous events is enabled by a diagram format where years are projected downwards, creating a singularity at the bottom edge where the time of events is unknown (Figure 4.8).

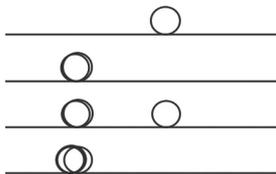


Figure 4.4 – A concept prototype exploring the interaction paradigm of coiling up a timeline.

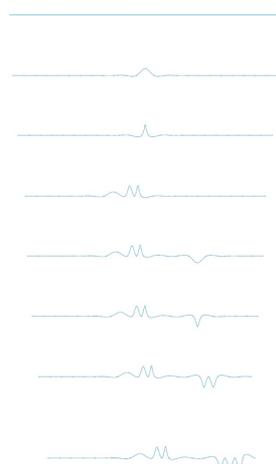


Figure 4.5 – This concept prototype allows a timeline to be folded at arbitrary times, in order to compress certain timespans and bring others closer together.

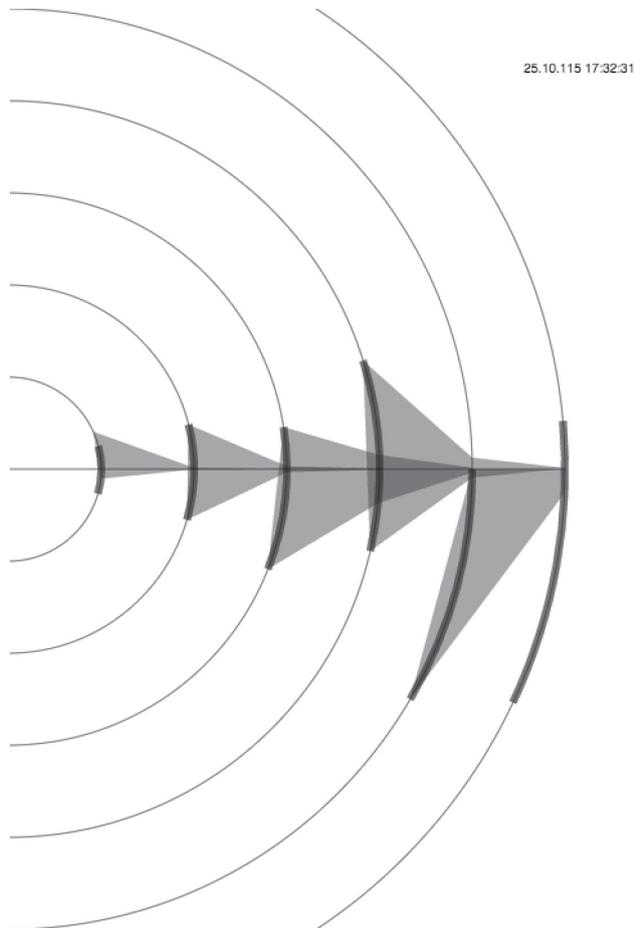


Figure 4.6 – This interactive prototype tests the idea of a two-dimensional model of time in a circular projection. The innermost ring represents seconds, while outer rings encode coarser granularities of time.

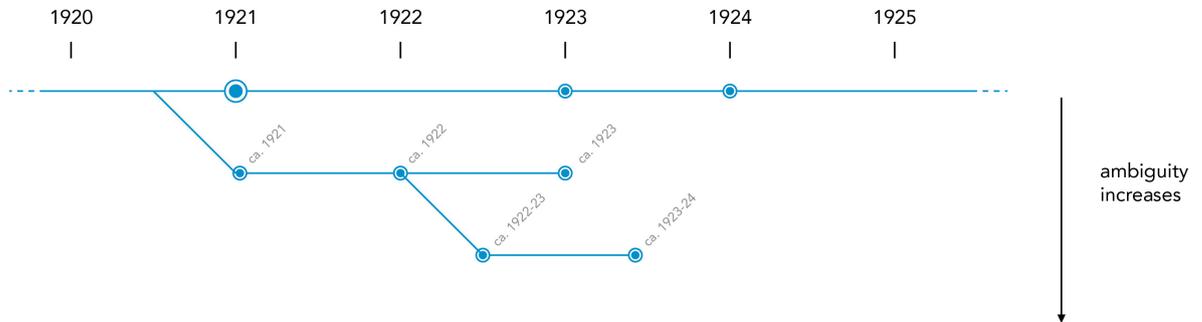


Figure 4.7 – A tree of possible histories, grown out of ambiguities of time.

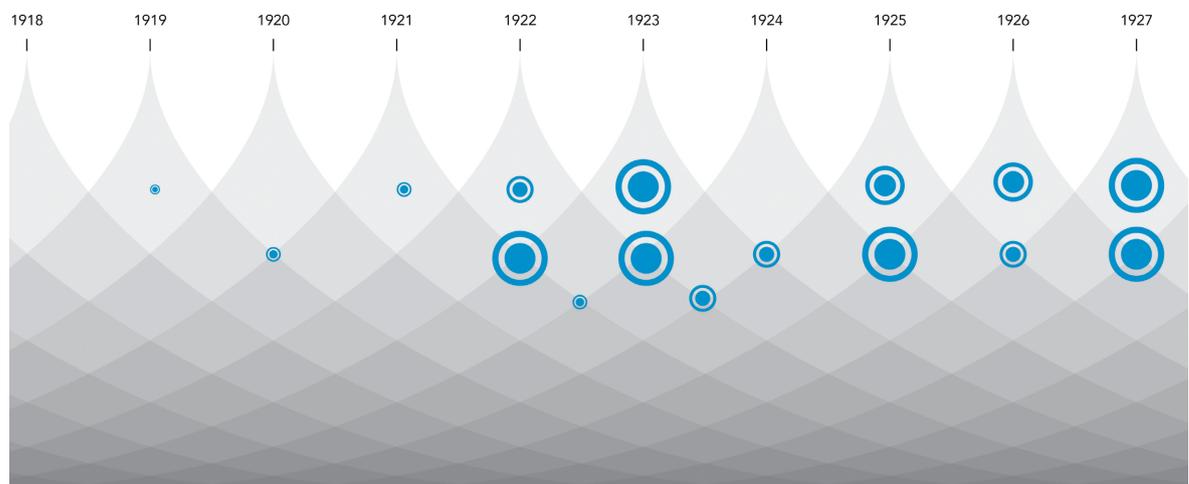


Figure 4.8 – Years project into neighbouring years and produce overlapping spaces of uncertain time frames.

Insights

I initially distanced myself consciously from specific datasets in order to explore the topic more freely. However my endeavours towards unconventional and possibly more human notions of time translated poorly to the application in real digital collections. Existing datasets are incompatible with my proposed alternative temporal approaches. Relations, explicit levels of confidence, source identifications, alternatives - information that might have been available to a curator at the time of dating a cultural artefact and necessary qualifiers for supporting richer models of time - are all conflated into numeric date specifications.

It was crucial therefore, to move my prototyping closer to existing cultural collections and to look at the opportunities for alternative temporal representations that are commonly supported by digital collections. An obvious step, perhaps, but also one that is easier said than done. Technically there were no obstacles in getting access to cultural datasets through my collaboration with System Simulation. Getting permission to use those datasets required the consent of their rights holders, which first had to be negotiated.

Continuing

What possibilities for alternative models of time are cultural datasets able to afford?

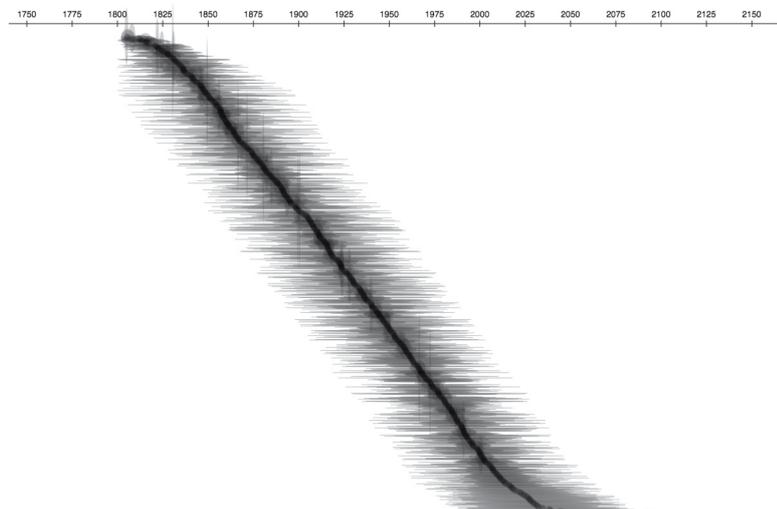
Prototype 3 – Drawing Uncertainties

Problem Statement	
Issue	- Temporal descriptions
Proposed solution	- Probability Density Functions
Dataset	- Generated - Geffrye data
Evaluation	- User conversations - Critical reflection

Process

After my previous prototypes I moved my endeavours out of the ‘lab’ and closer to the datasets at hand. With regards to my ambitions to allow timelines to use different temporal descriptions, I decided to take a closer look at what we might be able to achieve with the date pairs that are used in digital collections for specifying imprecise temporal information.

Figure 4.9 – A timeline visualisation of a randomly generated dataset, consisting of events with varying certainties.



3 A date field, for example, may only accept a value for a year, even when more precise information would be available.

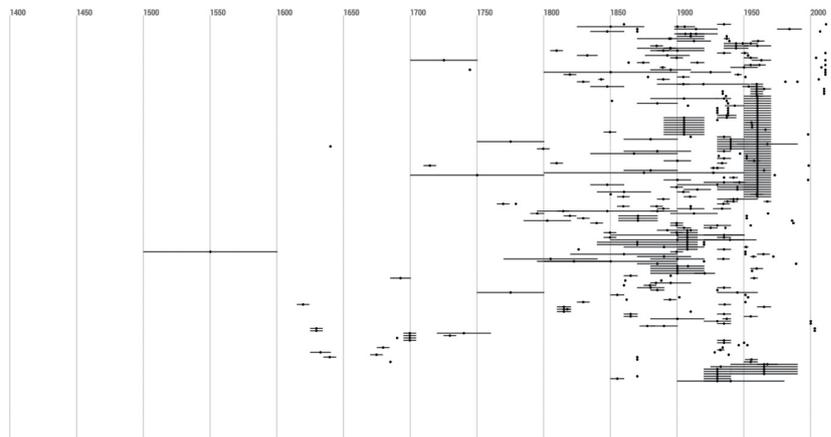
Digitally stored descriptions of time imply a certain level of precision through the need to specify discrete times or time frames and by having to match the granularity expected by the database.³ Date brackets suggest that an event has happened precisely within these boundaries, with equal probability. However unlikely this scenario is, when digital timelines map date boundaries onto precise bars and rectangles, this is what the graphics imply. To represent such data more honestly, their purpose in defining a possible timeframe for an event needs to be taken seriously. A date bracket could signify an event that

has happened at some unknown precise date within the date bracket, it could represent an event that is inherently imprecise or it could express a curator’s confidence in a date description.

Based on these observations I devised a method to computationally model confidence levels from the date descriptions that are available in cultural collections. I have discussed their rationale and working method elsewhere (Kräutli & Boyd Davis 2013) and will only briefly summarise them here, focussing instead on why I later departed from this approach. The idea is based on translating date brackets to probability density functions. These functions are able to calculate a varying level of confidence for any period of time within and outside a specified date bracket, allowing visualisations to render more sophisticated event representations than uniform blocks.⁴ The confidence is calculated by the integral of the function over the given period, which means we always need to look at a period of time rather than a point in time. The probability of an event having taken place at a single point in time is always zero, which reflects our intuition of what we are at all able to know about the date of historic events.

4 Boyd Davis et al. (2013) have explored various glyphs for representing temporal uncertainty.

Figure 4.10 – Already minimal indication of confidence – standard error bars – can make a timeline visualisation crowded.



Insights

I developed prototype implementations of my proposed approach, first based on model data (Figure 4.9) and later using the Geffrye dataset (Figure 4.10). I experimented with different visual renderings of the probabilistic events, but did not arrive at a fully convincing solution. I indicated confidence levels by changing the way events are graphically represented; adding elements such as lines to mark confidence intervals, or using shading or opacity to indicate that events are uncertain. This graphical rhetoric has already been proposed by Priestley (1764) and reiterated by Drucker:

Edges might be permeable, lines dotted and broken, dots and points vary in size and scale or degree of ambiguity of placement, and so on. (2011b, sec.20)

On the level of an individual event, these strategies worked well and could be interpreted as an indication of uncertainty. In the timeline visualisation as a whole, they however tended to introduce a bias towards events that have a wider date bracket which, as a result, also occupy a larger area on the time axis.

What is visible as well are quantisation effects of the dating strategies. Date brackets in the Geffrye dataset are not of arbitrary width, but are defined in regular intervals of often 10 or 25 years – a pattern which could be informative about dating strategies in an institution.

Perhaps the most crucial observation from attempting to model and render uncertainties in this manner is the danger of implying precision in the imprecision; suggesting that we can derive from the available data exactly how uncertain a specific event is. This might be possible with sensor data, where the imprecision of a reading can be measured; where the unknowns are known. Dates in cultural collections are not meter readings but human crafted inputs whose basis are, as far as the digital data is concerned, unknown:

You can talk about time duration with error bars [...] or the number of decimal places of significant figures. That's the standard uncertainty expression for figures. But epochs, dates, are not durations. That's why we have the problem we have, there's no way to express it. (C3)

The problem of temporal uncertainty, and how it is represented in a visualisation, will need further consideration.

Continuing

How to represent temporal uncertainty in an entire dataset?

How to represent uncertainty, without implied certainty?

Prototype 4 – First Encounters With Tate

Problem Statement	
Issue	- Scale - Layout
Proposed solution	- Summarising records - Vertical distribution dependent on data density
Dataset	- Tate Artworks and Artists dataset
Evaluation	- Peer critique - Critical reflection

Process

An opportunity to delve into the challenge of visualising large datasets on a timeline presented itself through the Tate and its digital catalogue, which contains data on almost 70,000 items. In an effort towards increasing the accessibility and use of their digital collection, Tate published an export of the metadata of their entire collection of works of art and an additional dataset with tombstone information of 3,500 associated artists on GitHub, a platform normally used by programmers as a collaborative code repository. Tate followed the footsteps of Cooper Hewitt⁵ in choosing this fairly unusual publishing method as the “most time- and cost-effective method of releasing data to [the] public for now”.⁶

5 Cooper-Hewitt published their dataset on GitHub in 2012 and was followed by Tate in 2013 and MoMA in 2015, among other institutions.

6 According to a Twitter status by the (now former) developer at Tate Richard Barrett-Small: <https://twitter.com/richbs/status/396206787277643776> (accessed 27.12.2015)

Being able to download this dataset influenced my perception of a digital collection. Even though I had access to exact copies of archival databases, they still made it difficult to consider a dataset as a single entity, as a whole. In a database, information is scattered across a number of tables and in order to retrieve it, one has to formulate a precise query. Many cultural institutions mirror this paradigm of interacting with a collection online by providing search forms and APIs that return only a limited number of records. On GitHub, it is possible to download an entire digital collection as a single CSV file.

In principle, one could create a single file export from any database, but this is not something the database paradigm affords - in the Gibsonian (1977) sense. Databases afford partial access, while downloading a file entails that all data is contained within that file. A study conducted by Harper et al. (2013) highlights how users see files as something they can own and manipulate, giving them a sense of control and completeness - both qualities that are useful also for analysing data.

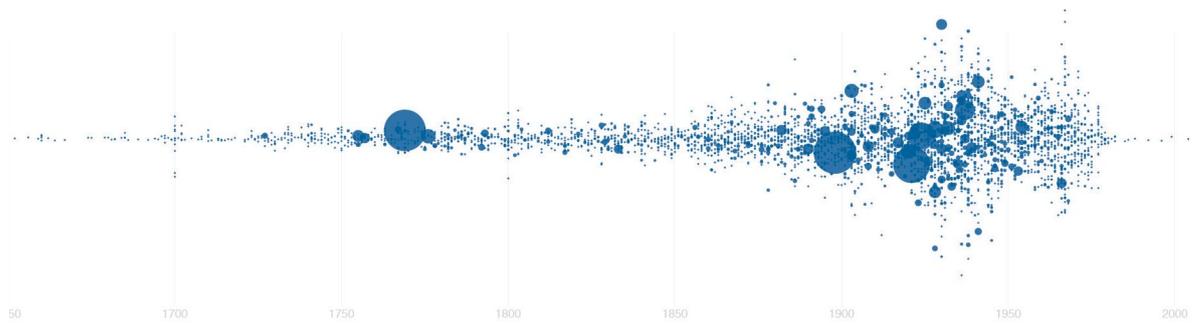


Figure 4.11 – All but one artist in Tate’s collection are plotted as bubbles and arranged by dates of birth on a horizontal timeline. The size of each bubble corresponds to the artist’s number of works in the collection.

Tate’s degree of popularity quickly gave rise to a range of projects. Researchers built visualisations to look at the different sizes of artworks (Davenport 2013), or the distribution and correlations of assigned topics (Drass 2013), or to map the gender gap in Tate’s art purchases over time (Szpak 2013).

My own interest lied in getting a sense of the scope and temporal distribution of the collection as a whole, following my endeavours in representing large datasets on a timeline. The prototypes and sketches, which I documented online (Kräutli 2013), were some of the very first anyone ever did with the dataset – a few weeks after Tate published it – therefore one of my visualisations (Figure 4.11) quickly became a poster child for illustrating the potential of visualising the Tate, and collections data more broadly. It has been included in a number of articles on this topic (Simon 2013; Bailey & Pregill 2014; Rodley 2014; Winesmith & Carey 2014). The only problem is, the diagram is misleading.

Bailey and Pregill describe it as

an example [...] in which the collection data of the Tate Museum was used to parse the distribution of artworks by the birthdate of artists, with the size of each bubble representing the number of works by that artist in the collection (the vertical position is for visual clarity and has no statistical meaning). (2014)

What they and many other authors who wrote about the visualisation do not point out is that the visualised collection actually misses its most prominent artist: J.M.W. Turner. For their omission I have but myself to blame as I did not sufficiently indicate Turner’s absence in the visualisation’s caption – a reader needed to follow the reasoning in my writing in order to be aware of it.

My decision to remove Turner from the diagram was partly for the sake of its visual appearance. When I envisaged the diagram that I hoped would result from mapping the data to bubbles in the described manner, I imagined it as pictured above. Instead, what I was presented with was a much more monotonous representation: a single large blob embedded in a plane of similarly-sized dots (Figure 4.12).

The large circle, positioned around 1800, testifies to the overshadowing dominance of Turner in Tate’s collection, accounting for about 40,000 works, or more than half of the entire art collection. This observation by itself was a surprising insight the visualisation provided and which I later studied more thoroughly. Tate’s second most prominent artist in the collection, George Jones, is present with ‘only’ 1,046 works, or 0.015% of the collection.

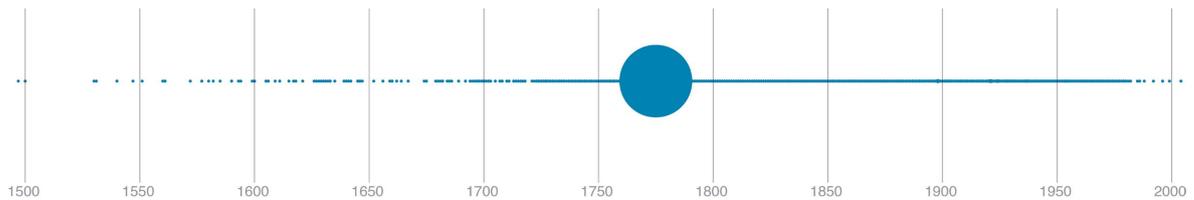


Figure 4.12 – Works by J.M.W. Turner account for more than half of the Tate’s collection, overshadowing all other artists in this early version of the visualisation.

As my visualisation uses a linear scale to map number of works to circle size, the unequal distribution of artists makes it impossible to spot any patterns among the non-Turners; all their dots are only a few pixels in size. Removing the singular extreme case of Turner allows for a more differentiated scaling and gives the diagram more depth, enabling me to see what I wanted to see; the scope and temporal distribution, although no longer of the complete collection. “In a sense he was blocking your view”, a curator later commented, “By taking away the Tate’s chief glory, you got a better idea of their collection” (C11).

Insights

My own experience with working with the Tate dataset and the responses that the work received confirm my hypothesis, that timeline visualisations can be an insightful tool for visualising and analysing collections data. Bailey and Pregill comment on my diagram:

This simple visualization would be impossible through interpretation of individual collection records themselves, but open data combined with an interested researcher has allowed for the building of a visualization that concisely provides a high level of detail about characteristics of the museum’s collection that would be impossible without the encapsulating power of this chart. (2014)

They conclude that

Visualizing information from the perspective of collection analysis allows curators and collection managers to gain novel insights into the nature of their holdings – insights that facilitate and enhance supporting research uses and imagining new methods of exhibition. (2014)

Simon (2013) takes a more critical stance in discussing Davenport's (2013) and my own visual interpretation of the Tate dataset. While Simon is excited to see "what surprising things can be visualized and learned from the collections data", she also points out that "the data is sufficiently flawed and idiosyncratic to yield conclusions of questionable value" (Simon 2013). In particular, she proposes to use collections data to explore more "compelling" (ibid.) questions than Davenport's size comparisons: issues of gender or race of artists, or which works of art may have been loaned out to whom.

Information on loans constitutes data that most institutions are unwilling or unable to make available to the public, leaving one to speculate what might be hiding in this kind of data. By my own experience of negotiating the use of collections datasets, the main reasons for not sharing (parts of) a dataset are concerns regarding copyright and the security of the physical collection. Loans data, for example, might equip a malicious individual with detailed knowledge about the financial value and physical location of an object. Another reason is the more obscure feeling of insecurity about what one might give away by sharing such data, not knowing what exactly it contains and what it could tell a researcher equipped with the right tools. An unfortunate and vicious cycle, when releasing data to researchers could dismantle such insecurities.

The other kind of questions Simon proposes may however be answered by the level of detail present in the Tate dataset and in most other available digital collections, given suitable tools that allow all available fields to be used. I would be cautious however to favour - in a visualisation tool - one type of question over seemingly less compelling ones, as even the most benign starting point may lead to surprising insights.

Simon rightly observes that "the data you have is not always the data you want, but you often don't know that until you start monkeying with it" (Simon 2013) and when one does so, questions might get answered one did not previously think. Ed Rodley, Associate Director of Integrated Media at Peabody Essex Museum, observes:

By giving people access to the collection, you give them the ability to ask questions of the museums' work that we inside these museums might never think to ask. Florian Kräutli's visualization of Tate's collecting history was not only a new way to look at Tate, but also uncovered flaws in how Tate's collection data were organized [...] that a collections manager or information architect could overlook or rationalize. Putting it out in public both surfaced the issue and provided impetus for addressing it. (2014)

In conclusion, producing these visualisations of the Tate dataset at this stage in my research resulted in a number of insights relevant for the progression of my research questions.

Accounting for Inconsistencies in Collections

Erasing Turner's works from the visualisation, as drastic as this intervention might seem from an art historical perspective, was essential for being able to study the distribution of artists in the collection without this extreme case. Filtering is often a necessary, if not indispensable step in exploratory data analysis; progressively excavating new patterns and relationships in a dataset by intentional omission or separation. Visualisation tools need to support such filtering options.

However, my reasons for omitting Turner were not purely for analytical purposes. I removed his works also in order for my diagram to be visually more lively. I tackled the 'problem' by cleaning up the diagram, a 'satisficing' (Simon 1997) solution for this individual case, but none that would satisfy my research goals.

Instead, I have to reflect on how the diagram could be redesigned in order to be able to accommodate such inconsistencies likely to be found in other cultural datasets as well - and likely to produce valuable insights when visualised. Of course, it is not harmful when such diagrams take on a visually pleasing appearance and it is impossible for subjective design choices - in terms of selecting colours, fonts and shapes - to be completely absent in a design-led research process. However, visual appearances should not be the main driver.

Communicating Completeness

The version of the visualisation that was publicised is a partial view of Tate's collection, but one is led to believe - when it is presented detached from its history - that one is actually looking at a representation of the complete collection.⁷ This observation is insightful in two respects. First, it is important to be instructive about what a visualisation displays, both in terms of the data as well as the mapping of the data.

Second, viewers' expectation that a diagram is a complete representation needs to be taken into account when designing a visualisation. The challenge for my research here is not only to accommodate for large datasets in my timeline visualisations, but to account for a sense of knowing the presence of the complete collection; designing not only for scale, but completeness. This issue was later also evoked by a user study and even for the early chronographers, being able to represent the "complete" history was a recurring ambition (Boyd Davis 2015a).

7 When this work is discussed it is referred to as a visualisation of the Tate collection, not the Tate collection *without Turner*.

Continuing

How can a large dataset be visualised on a timeline as a ‘complete entity’?

How can biases and inconsistencies not only be revealed, but explored in detail?

Prototype 5 – ChronoZoom

Problem Statement	
Issue	- Scale - Layout
Proposed solution	- Hierarchical Time Tree (zoomable and collapsible)
Dataset	- ChronoZoom
Evaluation	- Peer critique - Critical reflection

Process

ChronoZoom Hierarchical TimeTree (Figure 4.13) was my winning contribution to the Visualizing Time challenge (Pappas 2014), organised by Visualizing.org and Microsoft Research. The contest invited participants to “use the ChronoZoom APIs and create novel solutions to visualize the entire ChronoZoom dataset” (visualizing.org 2013). ChronoZoom is an interactive visual timeline that enables users to explore Big History, the history from the Big Bang up to our present time.

Strictly speaking, the ChronoZoom dataset does not exactly qualify as a digital collection. When I began the ChronoZoom project I saw it as an aside to my ‘actual’ research endeavours. I mainly chose to participate in the contest because the description stated unsolved problems in timeline visualisations very similar to the ones I identified as well: challenges that are only rarely discussed and addressed in visualisation research and practice.⁸

⁸ Of course, another good reason to participate was the prospect of winning a \$2,000 cash prize.

The original ChronoZoom is both a timeline interface and a dataset that contains all the events and histories that make up the timeline – the dataset can also be accessed separately from the visual interface. As the dataset grew in size, it became difficult to explore using the current interface. How to design a novel timeline interface that enables large datasets to be visualised is a challenge the developers of ChronoZoom faced as well, as they outline in their contest description:

There are currently about 7,000 content items in the default Big History dataset and we anticipate many more. It is crucial for users to be able to get a clear overview as they begin to browse this data. [...] [Contest] entries should aim to make timelines and related content more clear at a high level. (visualizing.org 2013)

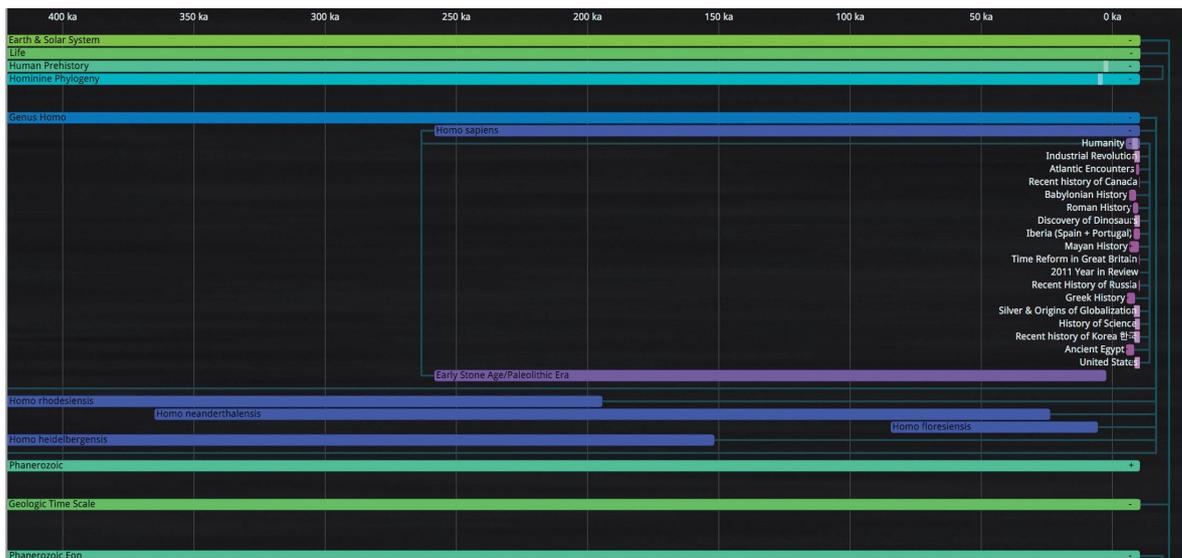


Figure 4.13 – My prototype visualisation of the ChronoZoom dataset displays recent events in the larger context of Big History.

Their brief specifically asks for clarity at “high level” in alignment with my own observations that visual timelines often fall short on providing an overview of a complete dataset.

ChronoZoom’s content items are defined as a set of nested timelines all contained by the root, the *Cosmos* timeline. Within the *Cosmos* timeline we can find the history of the *Earth & Solar System* and in there the history of *Life* down to finer grained histories even of the Microsoft Corporation. In the original ChronoZoom interface, these nested timelines are drawn, not as bars, but as rectangles, within rectangles, within rectangles – forcing each graphical timeline to be tall enough to accommodate all of its children. This produced the difficulties the ChronoZoom developers faced as the dataset grew larger:

When viewing “humanity” for example, it should be easy to see quickly the names of a reasonable number of visible timeline titles. When viewing the “Industrial Revolution”, the timeline is unnaturally very tall due to the large number of parallel items. (visualizing.org 2013)

To solve this problem I developed a timeline layout that is a hybrid between a common Gantt-like timeline, with events represented as horizontal bars, and a collapsible folder tree: timelines are drawn, not within, but below each other with dependencies indicated through connected lines and additionally through colour coding (Figure 4.14). I have elsewhere provided a more thorough description of the project (Kräutli 2014b) and its technical details (Kräutli 2014a).



Figure 4.14 – A sketch of the tree-based timeline layout. Dependencies are indicated by lines and colour shading.

Insights

Abstraction

My immediate insight from this project was on the relationship between the structure of the dataset and its visual representation. Manovich (2011, p.13) argues for the benefits of “direct visualisation”: the reorganisation of data “into a new visual representation that preserves its original form” (Manovich 2011, p.13). I am highly critical of this view, first of all because digital data does not have any inherent visual appearance or “form” to begin with - any data visualisation is necessarily an act of interpretation. Secondly, for the reason that my prototype offers a better⁹ overview of the dataset than the original ChronoZoom: precisely by *not* graphically mimicking the data structure of nested timelines and instead adopting an abstracted visual representation of implied hierarchy. An increased level of abstraction, even when it means moving away from ‘reality’, can allow for better sense-making and understanding; a simplified topographic map may be a more useful orientation device than a detailed aerial photograph.

⁹ According to the contest’s judges:
 “Florian Kräutli’s ChronoZoom Hierarchical Time Tree, [...] best presents a novel solution for visualizing the entire ChronoZoom dataset. It has an approachable and intuitive interface that allows the user to easily dive in and discover new topics. It nicely also has very detailed panels, almost like an open book, for each selected topic or event.”
 see <http://visualizing.org/stories/visualizing-time-winners-0>
 (accessed 15.05.2014)

Reusability

The second insight concerns the reusability of (parts of) visualisations. I was writing a section of a paper on the history of visual timelines and the pioneering contributions by Jacques Barbeau-Dubourg (1709-1779), Joseph Priestley (1733-1804), William Playfair (1759-1823) and Adam Ferguson (1723-1816) when I wondered how their lifespans may have overlapped and whether or not they have been aware of each others’ work. As Priestley (1764, p.10) argued in defence of his chart, such relationships are hard to calculate mentally, but obvious when inspected on a visual diagram. I had no suitable diagram at hand, so I began to draft one in MS Excel when it dawned on me that, provided with a suitable dataset of the lifespans of these individuals, my ChronoZoom visualisation would be able to give me the answer quickly - and make me aware of just how far Playfair, who is often regarded as the father of modern charts, was predated by other proponents of arithmetic data visualisations (Figure 4.15).



Figure 4.15 – I adapted my ChronoZoom prototype to study the lifespans of the pioneering timeline designers, which subsequently led to the development of a reusable visualisation tool.

The architecture of my ChronoZoom visualisation allowed individual components to be decoupled and repurposed. I defined and reviewed open timeline visualisations that accept different types of datasets and (in the case of SIMILE) can be embedded as plugins into larger projects. Based on my experience with ChronoZoom I started considering visualisations not as self-contained tools, but as a set of interlinked components. I directed my focus on the reusability of my proposed visualisation tools, both complete as well as in parts, in order to not only make the tools more adaptable, but most importantly more transparent. University of Virginia’s Neatline (Nowviskie et al. 2012) has similarly diverted its focus. Described at the time of launch as “a tool for the creation of interlinked timelines and maps” (UoV Library Scholars’ Lab 2009) it is since its second version (McClure 2013) presented as “a lightweight, extensible framework for creating interactive editions of visual materials” (UoV Library Scholars’ Lab n.d.). The SIMILE timeline, originally a central element, is now just one of several optional plugins that can be included. This shift does not follow a change in the technical architecture - Neatline has been designed as a plugin on top of the Omeka CMS¹⁰ since the beginning - rather it is a change in the perception of a digital visualisation suite as a reusable scholarly tool.

¹⁰ A Content Management System, or CMS, takes care of organising and delivering website content. Omeka is a specialised CMS for use by museums and archives and includes an interface to common digital collections.

Continuing

Is a modular visualisation tool more appropriate for scholarly research than standalone software?

Prototype 6 – Tate Artists

Problem Statement	
Issue	- Scale
Proposed solution	- Sampling based semantic zooming
Dataset	- Tate - Britten-Pears - London Transport Museum
Evaluation	- Controlled user testing - User conversations - Critical reflection

Process

I repurposed the timeline view and bars layout of my ChronoZoom prototype and turned it into a modular timeline tool. I tested the visualisation with data from the Tate, the Britten-Pears Foundation and the London Transport Museum, and developed an extended prototype based on the Tate artists data that includes additional interface elements to filter the dataset (Figure 4.16).

In order to represent up to 70,000 records in a Gantt-like timeline, I implemented a type of semantic zooming based on a sampling approach.¹¹ Semantic zooming is an established concept, for example, in digital maps, where it is used to draw only the essential features of a landscape at high zoom levels and reveal more details dynamically as the user zooms in. Perlin and Fox (1993) invented both the term and its first application. Semantic zooming has been used in timeline visualisations for quantitative data (Brodbeck & Girardin 2003; Aigner et al. 2012; Hoffmann et al. 2012) as well as for navigating photographic data on a graphical timeline (Huynh et al. 2005). In a digital map, the hierarchy of geographic features controls what is visible at which zoom level: country, district, city or street-level information. In less structured datasets, such as cultural collections, a suitable mechanism needs to be established first. My prototype requires a measure of ‘importance’ to determine the rendering of individual items at different zoom levels; only the most ‘important’ records are represented at higher zoom levels, with less ‘important’ ones either drawn as a thin line or hidden completely. The importance of artists in the Tate datasets is measured by their number of works in the collection, but any other measure could work equally.¹² The design of this prototype is more thoroughly discussed elsewhere (Boyd Davis & Kräutli 2014).

11 This approach effectively means that the number of items drawn on screen is reduced at higher zoom levels. An alternative would be to reduce detail, but keep item numbers intact.

12 For the London Transport Museum dataset I defined a measure of ‘uniqueness’ calculated based on the proportion of certain item types within the entire dataset. Any relationship that can be expressed as a quantitative measure can in principle be used to control the semantic zoom.



Figure 4.16 – The artists in the Tate dataset are visualised in a reusable timeline tool that features semantic zooming and a number of filtering interfaces for users to explore the data.

User Testing

I conducted a controlled user testing on this prototype, following best-practice UX evaluation methods after Nielsen (1994). Subsequently I abandoned this route and instead followed the evaluation method outlined in my methodology. Here I want to give a brief account of the initial testing procedure and discuss the reasons for my departure from evaluating through controlled user studies.

Candidates were recruited from the student body of the Royal College of Art via the college’s online notice board, ensuring that the participants had at least basic knowledge of information technology. The call informed interested students that they would be interacting with a software interface based on the Tate collection and that some interest in art and art history was desirable. Seven candidates were initially scheduled but due to unfortunate circumstances only five were able to attend on the day.¹³

The participants were led to a quiet room where they were asked to sit in front of a laptop computer running the visualisation and

13 The date of the testing coincided with a Tube strike.

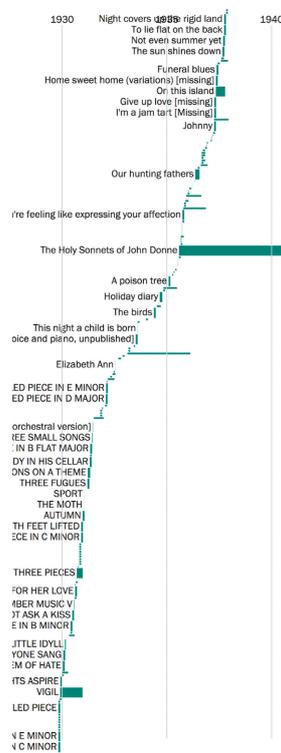


Figure 4.17 – The date brackets in Britten-Pears’ dataset encode – if known – the time it took Britten to compose a piece. In a Gantt-like timeline layout, works with known or unusually long durations of composition therefore stand out.

received four sheets of paper with instructions and small exercises to familiarise themselves with how to interact with the prototype. I then verbally read out the test procedure which consisted of them executing a set of twelve tasks using the timeline visualisation and filling in a short questionnaire after completing (or giving up on) a task. Participants were also asked to describe to me, while they were working on the task, what they were doing and why, following Lewis’ “thinking-aloud” protocol (Lewis 1982; Nielsen 1994; Lewis 2006). At the end of the session I captured their general impression of the prototype in non-structured interviews. During the testing I took notes and all sessions were video recorded and transcribed.

Results

In general, participants reacted positively to the interface and were eager and motivated to delve into the visualisation and follow their own leads beyond my predefined tasks for exploring the dataset using the timeline tool. Being presented with a lot of data at once made people curious. They quickly started to, for example, explore art movements and find corresponding artists for movements they were not aware existed, or test their own knowledge about the artists present in the dataset.

While the experience of testing and the feedback I received were encouraging and confirmed the ability of digital timeline visualisation tools to assist in knowledge discovery in digital collections, most of the insights and problems the sessions revealed could not be translated beyond this specific prototype.

The strength of this and similar studies lies in exposing flaws in the specific implementation of a visual interface; usability problems related to interface elements, insecurities about the signification of certain labels or the use of filters and selectors. My goal is however not to evaluate the effectiveness of a particular implementation of a visualisation in achieving a certain tasks, but to gain a broader understanding of the benefits and challenges of timeline visualisation tools in enabling and assisting in knowledge discovery in cultural collections. I encountered a contradiction when trying to evaluate the ability of a tool to facilitate an essentially creative act through predefined tasks and a controlled testing scenario; or in the words of Shneiderman et al. (2006), “[...] specifying tasks is somehow at odds with the goals of supporting innovation or discovery” (Shneiderman & Plaisant 2006).

By interacting with the visualisation, the participants were able to make discoveries that were novel to them and formulate research questions that they then tried to answer using the tool. This observa-

tion is however insufficient evidence for drawing an informed conclusion on the ability of my visualisation tools to enable sense-making, as the users' discoveries were lacking a baseline. A widespread method for establishing a baseline is to use a comparative testing scenario, with a control group exploring the same dataset through a standard interface and comparing the kind of insights they were able to make with the ones that were facilitated by the tool to be evaluated. Such A/B testing is a common practice for evaluating visual search interfaces. However, van Hoek et al. (2014) question the reliability of these tests and identify the need for a common reference system, while Shneiderman et al. (2006) express a general doubt on the effectiveness of these, and similar, evaluation methods: "Controlled experiments of specific features seem too narrow as do gross comparisons of one tool versus another" (Shneiderman & Plaisant 2006). The baseline I use instead for evaluating the heuristic quality of my visualisation prototypes is the knowledge that experts have acquired during years of their professional and academic career. If this knowledge can be augmented through a visualisation tool - even ever so slightly - this will equip me with more reliable evidence than testing within a narrow framework.

In sum, what I was missing from this controlled user study - and what made me choose a different mode of evaluation - was a lack of transferable insights beyond specific usability problems, a difficulty of verifying discoveries that the visualisation facilitated and the impossibility of discriminating them from superficial or pre-existing knowledge because of the lack of a suitable baseline to 'measure' the ability of my visualisations to provoke new insights.

Insights

Confirmed Issues With Gantt-Like Timelines

In several respects the Gantt-like layout proved to be unsuitable for visualising cultural data. Bars encode durations and, with the exception of the artist lifespans, the temporal data in the cultural datasets refers not typically to time-periods but, as I showed earlier, to estimates of a possible timeframe for an event to have taken place. Collections data is thus largely event-based, although not exclusively. Britten-Pears' collection sometimes includes works that have a known period of composition - when a work comprises several pieces and each has been individually dated or when Britten referred to having spent time on a composition over several days in his diary. Solely based on the numeric dates in the database however, we are unable to decide whether the dates refer to a (documented) period of composition or to an estimated timeframe (Figure 4.17).

Mapping timeframes to lengths of bars produces a bias towards events with wider confidence in dates; larger timeframes visually stand out in the diagram. I have already observed this in P3 and the user study confirmed the effect. Indeed many users were drawn to identifying the oldest artist in the Tate collection, or to find out why certain artists seemingly reached an unrealistically high age – those ‘artists’ typically turned out to be umbrella categories for works of art whose creator was, in fact, unknown. However, in most datasets the signification of a duration is not as evident as in the case of a person’s lifespan, therefore a layout that emphasises this aspect may not be the ideal choice for visualising cultural data.

Earlier, when discussing the *Empires* timeline of Accurat (Beltramin et al. 2012) and Lee (2011), I pointed out the pattern that results from the correlation of the horizontal time axis and the vertical sequential ordering in Gantt-like timeline layouts. This effect occurs in my own prototype as well and I was able to observe the consequences for the viewer through my user study.

Users did not realise that the vertical ordering of the events corresponded to the birth-dates of the artists, even when they were consciously reflecting on it. Instead, users began to speculate on the significance of the vertical axis and sought to make sense of the overall pattern, giving up quickly as they could not come to a satisfying conclusion. This demonstrated not only the problem inherent in the layout which I criticised, but also that users try to interpret the visualisation both in respect to its totality as well as individual items, supporting the need for timelines to be effective on the level of individual events as well as in totality.

Biases Through Sampling-Based Semantic Zoom

Due to the sampling that was required by my semantic zooming mechanism, any conclusion one could derive from this overall pattern would not only be distorted by the layout algorithm, but also by the applied selection criteria of the semantic zooming algorithm. The S-shaped pattern that emerges in the overall representation of the records could signify an uneven temporal distribution of artists in the Tate collection, with a cluster of artists born around 1800 and a second one around 1950, or, it could show an uneven distribution of *significant* artists – according to a quantitative measure of significance. Due to the constraints of the chosen semantic zooming method, which only ever shows a fraction of the data, one is not able to draw any reliable insights about the entire dataset. However, users often did assume they were presented with the complete dataset when interacting with the tool. The layout did not sufficiently communicate the extent of the

dataset and whether or not all that is available is, in fact, in view. Some users stated that they expected the visualisation to be complete whenever it did not entirely fill the available space.

A prevalent assumption that a visualisation displays a complete picture of a dataset appeared already when the lack of Turner in my earlier Tate diagram (P4, Figure 4.11 on page 146) went unnoticed. Through this user testing I was once more made aware of the importance of not only making an entire dataset available, but also of the challenge of finding an appropriate representation that indicates the boundaries of a collection. In the physical world the constraints of an information source are often obvious through the medium. A physical archive is terminated by the spaces it is contained in. The thickness and weight of a book lets us estimate the size of its content; as we flip through pages, we get a sense of our 'location' within it. In the digital realm we need to design visualisations so that users get a similar sense of orientation.¹⁴ It is not sufficient to use external indicators. My prototype did state the number of items currently visible along with the total in writing, yet this information was ignored, with no exception, by all users.

¹⁴ The downsides of the lack of "physical concreteness" (Conklin 1987) of digital media have been observed early on: "It will be difficult to provide on a computer screen the equivalent effectiveness of the more subtle tactile and visual cues such as size, color, texture, absolute and relative position, weight and heft, etc. that paper documents offer." (ibid.)

Continuing

How can we design a semantic zoom that retains completeness?

What layout is more suitable to represent event-based data than a Gantt-like chart?

Prototype 7 – Britten Force

Problem Statement	
Issue	<ul style="list-style-type: none"> - Temporal descriptions - Scale - Layout
Proposed solution	<ul style="list-style-type: none"> - Time-aware force-directed layout
Dataset	<ul style="list-style-type: none"> - Britten-Pears
Evaluation	<ul style="list-style-type: none"> - User conversations - Critical reflection

Process

“Britten Force” is motivated by my enquiries into representing uncertainties in dating events and is based on the digital archive of the Britten-Pears foundation, a dataset that offered, from all the collections

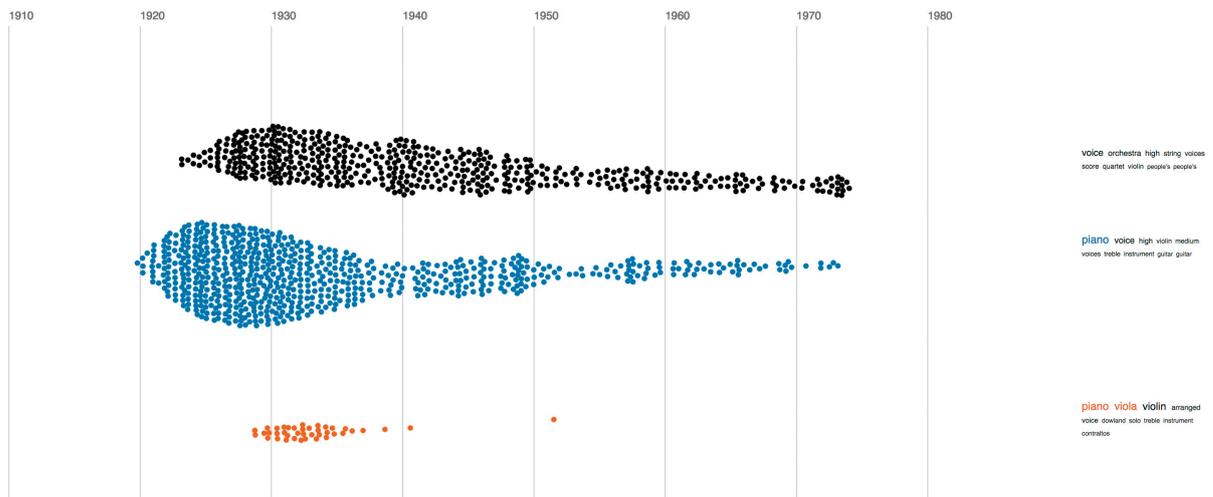


Figure 4.18 – Benjamin Britten’s works are visualised as dots that arrange themselves based on the (un)certainly of their temporal descriptions. A tag cloud allows users to split the dataset based on dominant terms in the works’ subtitles.

available to me, the most diverse source for different levels of (un)certainities in the temporal descriptions of the catalogued works. Reflecting on how I could even attempt to do justice to the variety of different temporal granularities and to the richness of the descriptions, I asked myself what would happen if I disregard most of the classifications, cataloguing and categorisation and instead just consider this significant cultural collection as, quite literally, a bunch of stuff.

In this visualisation every work of Benjamin Britten is graphically represented as a dot. Each dot’s position is determined by a simulated force; it is drawn by a gravitational pull to its destined location on a horizontal time axis. The more certain a dot’s temporal description,



Figure 4.19 – The force-directed layout algorithm generates regular clusters where several works share the same date description.

the stronger the force which is exerted on it. As individual dots are also repelled from each other, the layout arranges itself and ultimately settles in a flock-like composition – uncertain events will have settled with a position somewhere in the vicinity of their respective dates, while more narrowly dated events will stick tightly to their ‘proper’ temporal location (Figure 4.19). The technique is borrowed from a force-directed graph visualisation method (Fruchterman & Reingold 1991) that is also responsible for the fluid motion in which the dots align themselves – a visual effect that curators found very appealing.

Users can dissect the dataset by clicking and dragging one of the words that appear at the right of the flock in a tag cloud. They are based on word occurrences in the works’ subtitles, ordered and sized by frequency of appearance. I used the subtitles as they effectively described a work’s instrumentation and contained values for all the pieces. Additionally, this method of machine-interpreting free text is robust and reusable for other cultural datasets.¹⁵

15 Tag clouds or word clouds of this sort have faced criticism (Harris 2011) as simplified and potentially misleading text visualisations. My proposed use here relates to analysing free-text database fields, which are typically short and semi-structured pieces of text, making the shortcomings of tag clouds less relevant.

Dragging one of the words splits the group of records in two based on whether they match the chosen criteria. For example, “piano” appears as the dominant category in the Britten dataset and dragging the term downwards creates two new collections, one with and one without piano pieces. To the curators’ surprise the piano works comprised a large subset of the collection:

He is not known at all as being a piano composer, and there it is.
(C6)

What the visualisation revealed was not so much representative of Britten’s oeuvre, as of the decision to classify most of his many childhood works as piano pieces.

We can continue dragging selections based on instruments out of the flocks (Figure 4.18). “Viola” now appears as a term besides the piano pieces and separating it reveals a cluster of pieces around 1935 and a single piece, an outlier, composed around 1950.¹⁶

16 The piece “Lachrymae – Reflections on a song of John Dowland, for viola and piano, o. 48” composed 16 May 1950 and revised in June 1970. See <http://www.brittenproject.org/works/BTC1032> (accessed 14.12.2015)

Insights

The prototype was well received by the curators of the Britten-Pears archive who described it as “very expressive” (C6) and immediately began to suggest queries and how it could be enriched by using it to explore different aspects of their dataset. The tactile metaphor of start-

ing with a heap of things and then being able to sift through it, while getting visual feedback of the query result, proved to be very effective.

Due to the layout algorithm and the coarse granularity of the time axis the visualisation could only give an approximate indication of when certain works were composed. However, the curators raised no concerns on the factual imprecision of the events' date representation. They used the visualisation to examine overall temporal patterns rather than looking at precise date of composition of individual works.

The time-aware force-directed layout successfully addresses a number of challenges that arose in my research. Through the arrangement one can clearly see the extent of the collection and comprehend it as a single entity. Patterns that occur through variations in the overall shape are representative of temporal patterns in the dataset and not merely side-effects of the layout algorithm.

I have previously questioned the effectiveness of communicating uncertainties in visualisations by additional graphical elements. In the design of this timeline layout, uncertainty is treated not as an additional and therefore optional qualifier, but as a fundamental property of temporal data. Uncertainty is taken into account as a basis of the layout - a radical departure from Gantt-like layouts that expect the timing of events to be known and introduce an additional graphical vocabulary for the exceptional case of uncertain dates. In collections data, however, uncertain dates are the rule, not the exception.

Conceptually, this is an important insight, technically however the implementation of the layout did not scale well to datasets larger than a few hundred items due to the complexity of the physics simulation. By optimising the algorithm and changing the architecture¹⁷ a performance improvement would be conceivable, however the layout suffered from more than just technical difficulties. Unsurprisingly, the curators found it difficult to interrogate records that were moving around constantly. I realised that I had to find a way to retain accessibility to individual records, even in collections that span hundreds of thousands of them.

¹⁷ i.e. switching from SVG to canvas or WebGL, to take advantage of hardware acceleration in the browser.

Continuing

How can this type of layout scale to larger datasets?

How to use uncertainty as the basis of a visualisation and not merely as an optional qualifier?

Prototype 8 - Temporal Jittering/Parallel Timelines

Problem Statement	
Issue	<ul style="list-style-type: none"> - Temporal descriptions - Scale - Layout
Proposed solution	<ul style="list-style-type: none"> - Temporal Jittering Layout - Time-based reorganising
Dataset	- Britten-Pears
Evaluation	<ul style="list-style-type: none"> - User conversations - Critical reflection

Process

The starting point is the previously discussed implementation of a time-aware force-directed layout based on the Britten-Pears archive

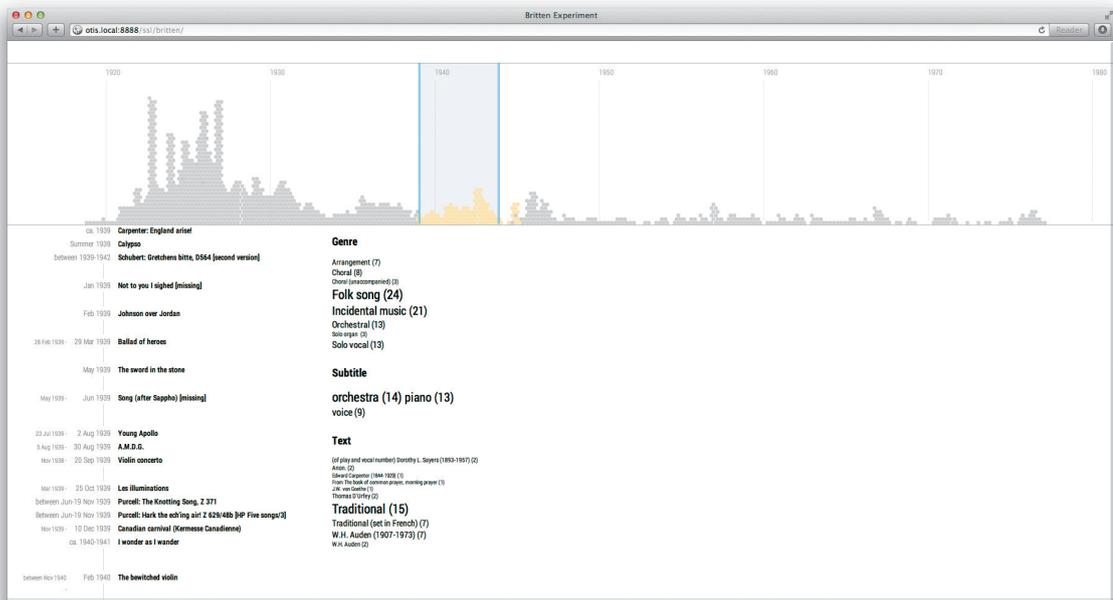


Figure 4.20 – A prototype that visualises Britten’s works using the Temporal Jittering layout, in addition to a vertical non-linear timeline and a number of tag based filters.

dataset (P7). The premise of temporal descriptions as inherently uncertain, the tactile metaphor of the heap of ‘stuff’ and the atomic representation of individual records proved to be effective. These are characteristics that should be retained.

Major downsides in the previous prototype turned out to be the fact that the approach scales badly to larger datasets and that the final

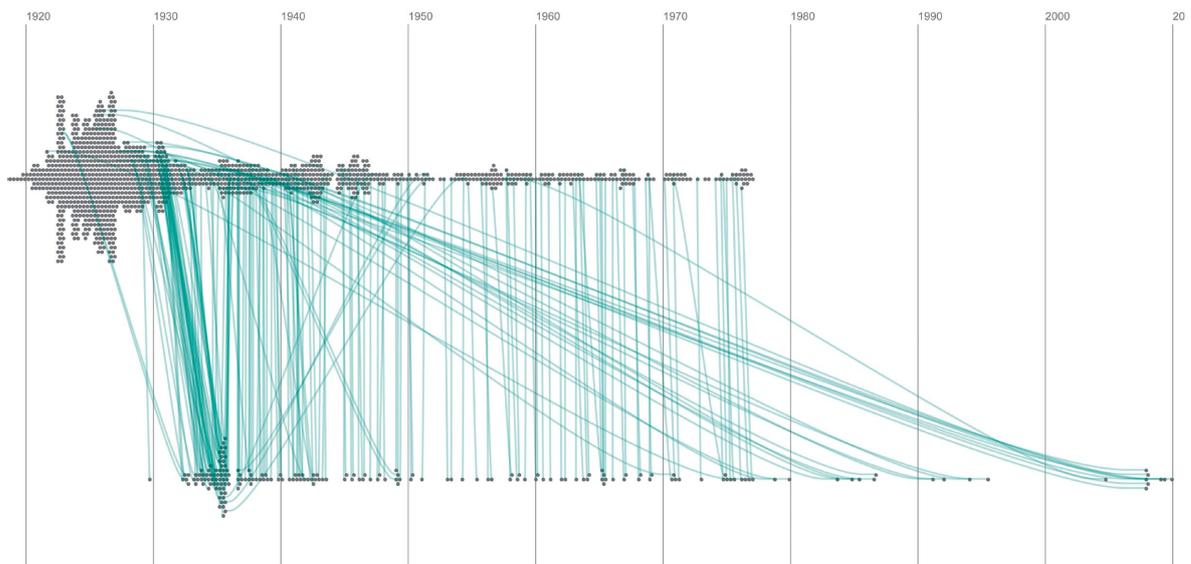


Figure 4.21 – Britten’s pieces are visualised by date of composition (top) and date of first performance (bottom). Connections between individual pieces allow for insights into the public interest in Britten’s works and reveal errors in the dataset.

layout is non-deterministic; the position of any single item depends on the attributes of all other items. I have already pointed out the adverse effects on readability of non-deterministic layouts in the context of the space-saving layout implemented in *Kindred Britain* (see page 113).

To counteract these properties, I developed an algorithm that arranges the dots in a reproducible, deterministic manner, making use of the uncertainties embedded in the temporal descriptions. The Temporal Jittering (TJ) layout represents each record as a same-sized dot, positioned horizontally according to its allowed timeframe – as defined by the date brackets – and stacked vertically in order of accession.¹⁸ This produces a compact and aggregated overview of an entire collection. Essentially, it exploits the width of the timeframe allowed by the date brackets to accommodate a more densely packed display than would otherwise be possible.

I produced two initial prototype implementations using the layout: an application that combines the layout with a vertical list based timeline and filters to explore and dissect the dataset (Figure 4.20), and a parallel timeline that looks at the relationship between the composition date of works and their date of first performance (Figure 4.21).

Following my focus issue on temporal descriptions I looked at what kind of dates are available in the Britten-Pears archive and what this visualisation format might be able to reveal about them. The parallel timeline visualises Britten’s works twice: the layout in the top half is generated according to their date of composition, the bottom half according to the date of first performance.¹⁹ I use a symmetrical version of the layout here – for reasons I will outline later. As the layout allows each work to be represented individually, it is possible to connect matching works by a thin line.

18 The name *Temporal Jittering* borrows from an established technical term that describes a clutter reduction technique in data visualisation (Ellis & Dix 2007). *Jittering* pragmatically displaces graphical elements, positioning them off their destined location, to prevent records from overlapping.

19 The date of first performance is not available for all the works, which is why the lower timeline contains fewer data.

Insights

We can spot a number of lines that are going the ‘wrong way’. Apparently these works have been performed before they were written - as it turns out, a result of textual dates misinterpreted by the cataloguing software. A vertical pattern of lines indicates that most of the works have been performed shortly after having been composed. Oblique lines heading from the performances into the past appear shortly before, and long after Britten stopped composing - and indeed living. These are first performances of juvenilia taking place. By the point in time at which they appear, we can study how people must have begun to understand Britten as a whole and started being interested in his very early works.

This can speak about the telling of biography. You revealed in this diagram that propensity for looking for the origins. (C3)

This prototype is not the first one where I experiment with using other temporal descriptions than the date of creation to visualise the dataset. However, it is the first one that does so successfully. Primarily because the layout allows not only an entire collection to be represented in a compact form, but also several ‘versions’ of a collection according to different temporal attributes. The collection itself - reorganised according to different temporalities - provides its own context.

When discussing this visualisation with the curators at Britten-Pears they were immediately able to spot the outliers:

How were some performed before they were written? (C6)

The also began to question and interpret the diagram:

I’m curious about this thick bunch of lines here. What’s that telling us? (C6)

Going back to the Temporal Jittering layout - this proved to be successful in tackling the issues that my research brought up. The strict rule-based positioning of the dots eliminates randomness; the diagram is built up gradually, placing each dot at the lowest available vertical position, while minimising the horizontal distance to the midpoint of the allowed time period. This relates to issues observed earlier about the meaning of the non-temporal axis; the vertical order matters as it is a consequence of the sequence in which records have been accessioned - a feature whose benefits become more apparent when I apply it to larger datasets later on.

In contrast to Gantt-like layouts that produce a sloping pattern in the overall representation of the records as a side-effect of the correlation between the two axes, the patterns produced by this layout are

informative about the dataset and - specifically - the temporal uncertainties

Both the layout algorithm and the implementation proved to be very effective in addressing the focus issues as well as by raising positive reactions and evoking insights and questions among my collaborators.

Continuing

How can the layout and the insights gained from the implementations be generalised?

What kind of insights are enabled by incorporating temporal multiplicities?

Emerging Design Principles

A development is taking place with regards to the focus issues. While I initially looked at them as separate problems, they have started to converge. Tackling issues of scale and layout is, in the last prototypes, done effectively by incorporating uncertainty as a fundamental principle in the timeline visualisation. The compactness of the layout in turn is achieved by making use of the uncertain temporal descriptions embedded in a dataset.

Based on the initial focus issues and the insights that emerged from the development and evaluation of the prototypes I propose three design principles that address specific challenges related to the time-wise visualisation of digital collections. These include approaches to examining large datasets through timelines that support distant reading, and new approaches to model and represent historic time by taking advantage of embedded uncertainties and temporal multiplicities.

Timelines for Distant Reading

Existing timelines and timeline layouts can be magnificent tools for locating an event in time or tracing the history of a course of events. Current timelines excel at close reading of histories, but they often fall short when trying to make sense of an entire dataset.

I showed how timeline tools collapse when faced with large datasets. It is not necessarily the scale of a dataset, but its temporal distribution that causes a lot of records to occupy a small timeframe, or even identical times. Unlike regular time series, digital collections and many other ‘manually’ produced datasets usually exhibit uneven temporal distributions.

By tackling the issues of layouts for visualising large datasets on timelines, I have paved the way for timelines that support a distant reading of digital collections. The Temporal Jittering layout results in overall patterns that are informative about the visualised dataset and readable in both dense and sparse temporal regions.

The difficulty of visualising large datasets over time is to achieve this not by imposing severe biases on how the data is represented. Combining events that are close in time and organising them by category (P1), or controlling a semantic zoom by hiding ‘unimportant’ records (P4) emphasises the dominant themes in a dataset and suppresses its long tail of records that do not match the specified measure of importance - an undesirable outcome. Preventing clutter by reducing the number of displayed items means exerting a bias on the rep-

resented dataset and violating the basic principle of distant reading: being able to study ‘everything’ without discarding data unquestioned.

Semantic zooming could be a viable solution for cluttered timelines and indeed has been applied in the past, but it needs to be able to function without reducing the number of represented records. Kindred Britain’s timeline adjusts the detail of drawn events when zooming out – but the maximum of records visible is capped at 74 records, while the dataset is said to contain 30,000 records. Other examples of digital timelines with semantic zooming all work by reducing the number of events represented.²⁰ Having to sample a dataset in order for it to be represented however compromises on the honesty of the visualisation and can mean that observed overall patterns are meaningless because not all records in a dataset contribute to it.

²⁰ These include Histro-
pedia Timeline, Time-
glider, Wellcome Time-
line (see appendix C on
page 270)

In my discussion of P1 I have pointed out how existing timeline visualisations approach the issue of distant reading by implementing two separate views – an atomic representation of events and an aggregate view of the entire dataset in a histogram. In a histogram we however lose the quality of being able to work with individual events. Ideally, I concluded, aggregate and atomic views could both be implemented using the same layout and representation. The Temporal Jittering layout I developed in P8 does have the potential to manage this because its overall shape is determined, can therefore be drawn like a histogram and is equally informative, while its composition consists of individual units that can also be drawn separately.²¹ The result is a timeline that supports distant reading by implementing an aggregate view that retains the completeness of a dataset, communicates its extent and exposes large-scale patterns that are informative, rather than side-effects of an algorithmic layout.

²¹ This concept will become
clearer when I discuss its
implementation later on.

Embedding Uncertainties

Closing in on the issue of making use of temporal descriptions I focused on the uncertainty embedded in numeric date brackets, to see how temporal uncertainty could be represented on a timeline (P3, P7, P8). Graphically indicating uncertainty as error bars – a standard approach in statistical diagrams – works well on the level of individual events. SIMILE implements a variation of this by indicating uncertain parts of events through a lighter shading. However, we are still left with the issue of clutter, to which graphical error bars might even contribute.

Other concerns emerged on the danger of suggesting – through a precise graphical rendering – certainty within the uncertainty. On the other hand, it is insightful that curators raised no concerns about the imprecision of the temporal positioning in the Britten Force layout

(P7). In this particular case it might be because this is a comparatively small collection. Curators typically know when most of the works were composed. In a larger dataset this might be more problematic. However, this example also shows how a timeline of cultural data can bring new knowledge to experts who are very familiar with every single piece in a collection. What the visualisation was able to tell them was something about overall patterns or connections and ask questions about a larger number of records.

Is it possible to enable a distant reading of a dataset, along with its temporal uncertainties? In P7 and P8 I have deviated from the standard practice of including uncertainties at the end of a visualisation ‘pipeline’ - by adding graphical elements or changing the representation of an event - to the very beginning of it - by making it the driving force of a timeline layout. Temporal uncertainty is then a fundamental component of the timeline visualisation, as much as it is a fundamental part of dating historic events. Timelines that incorporate temporal uncertainty at their very basis are able to respond to and be transparent about inconsistencies in temporal data that are typical for digital collections.

Multiple Temporalities

In P1-P7 I retained the established convention of museums to date cultural artefacts by their “moment of genesis” (C3). In P8 I deviated from this by arranging Britten’s works also by their dates of first performances. This allowed me and curators to study the history of the public popularity of Britten’s works in addition to their own history of composition. New insights in digital collections could be facilitated by making use of alternative temporal attributes as a structuring principle and subsequently also looking at a collection from different perspectives: rather than seeing the Britten-Pears collection as an archive of compositions, it could be an archive of first performances or - as I show later - an archive of poets.

The issue of multiple temporalities in cultural collections is possibly an under-explored one because in archiving practices the central unit is the object and its ‘date of birth’ is the dominant temporal point.

Chronology, the study of the proper temporal location of events and the foundation of modern timelines has enabled structured and comparative sense-making along a uniform model of mathematical time, but also reinforced a singular model of time, which treats events as occupying a single position in time. As I have shown, in digital collections, it is not unusual to record several temporal attributes of an item, such as date of acquisition or relocation, and an item may have related

events - exhibitions, persons, other items - which each have their own associated dates.

Digital visualisations of entire datasets have so far made little use of including multiple temporal perspectives for individual events and established formats for visual timelines are unable to represent rich relationships through multiple granularities of time. Visualising for multiple temporalities - in the way I have demonstrated in P8 and will develop further - means decreasing the authority of a single date associated with an event and instead rendering multiple temporal perspectives of individual events, which as a result draws a richer representation of the temporal relationships within an entire dataset.

5

“Very much Minority Report” (C4)

5 Paradigmatic Prototypes

In this chapter I will present two example implementations that apply the proposed design principles in timeline visualisations for digital collections. They appropriate successful strategies and develop them further.

The concepts of distant reading for large collections, representation of uncertainties and multiple temporalities - with a focus on the first two - are exemplified in a timeline visualisation that offers a reusable interface to explore and analyse various cultural datasets. The tool is unremarkably named Timeline Tool (TT).

The concept of multiple temporalities is further developed in a singular prototype based on the relationships between compositions by Benjamin Britten and the lives of the authors whose writing he used in his vocal works. A core aspect of this prototype is the newly developed Temporal Perspectives layout, aimed at visually analysing temporal relationships. Independent of the Britten's Poets example, it serves as a reusable timeline layout for establishing and visually tracing connections in arbitrary datasets.

The chapter includes brief descriptions of the prototypes' software architecture in order to familiarise the reader with their individual components and how they can be adapted to visualise arbitrary datasets. A technical documentation of the prototypes is included in Appendices D and E (page 301 and page 307).

Following a presentation of the working principles of the prototypes and how they can be used to visually analyse different kinds of cultural collections, I will highlight some of the specific insights they are able to facilitate.

In the next chapter, I will discuss findings that these visualisation evoked in my collaboration with curators, archivists and scholars and present their views on how digital timelines facilitate a new understanding of their cultural collections. Together, these discussions will answer my main research question: what kind of knowledge can we gain from visually analysing digital collections through timeline visualisations?

TT

- 1 I implemented the tool according to standard JavaScript design patterns, specifically the Module pattern (Osmani 2012). This required me to pick a name when I began the project, which effectively started out with the ChronoZoom visualisation (P5). I therefore chose a generic name, not knowing how exactly the tool will develop and planned to pick a more suitable name later – however in the end the name stuck.
- 2 A screencast of the TT prototype in use with the Tate dataset is available online at: https://www.youtube.com/watch?v=5WqyJX_-VJU (accessed 08.02.2016)

TT is a timeline visualisation library for cultural collections. The acronym stands for **Timeline Tool**.¹ Following my insights on the merits of reusable visualisation components I devised this prototype as an extendable library for analytic timeline visualisations.

First, the Tate digital collection is imported as an example and I will demonstrate how the visualisation can be used to explore and analyse a dataset. Specific insights that the visualisation enabled in this and other digital collections will follow in the next chapter.

I conclude this section with a discussion of the Temporal Jittering layout – the distinctive feature of the TT prototype – explaining some of the design decisions and evaluate it against alternative diagram formats for representing cultural collections on a timeline.

Walkthrough

Tate’s digital collection, which encompasses almost 70,000 works of art, is summarised in Figure 5.1 using the Temporal Jittering layout.² A semantic zoom functionality aggregates the view by displaying it as a contiguous shape, revealing the dots that comprise it only at deeper zoom levels.

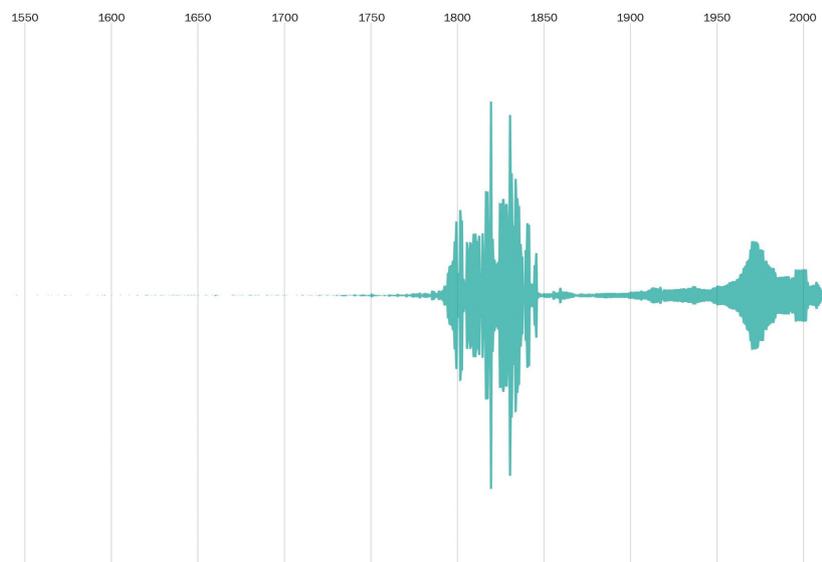


Figure 5.1 – the Tate’s complete artwork dataset as represented by the TT prototype.

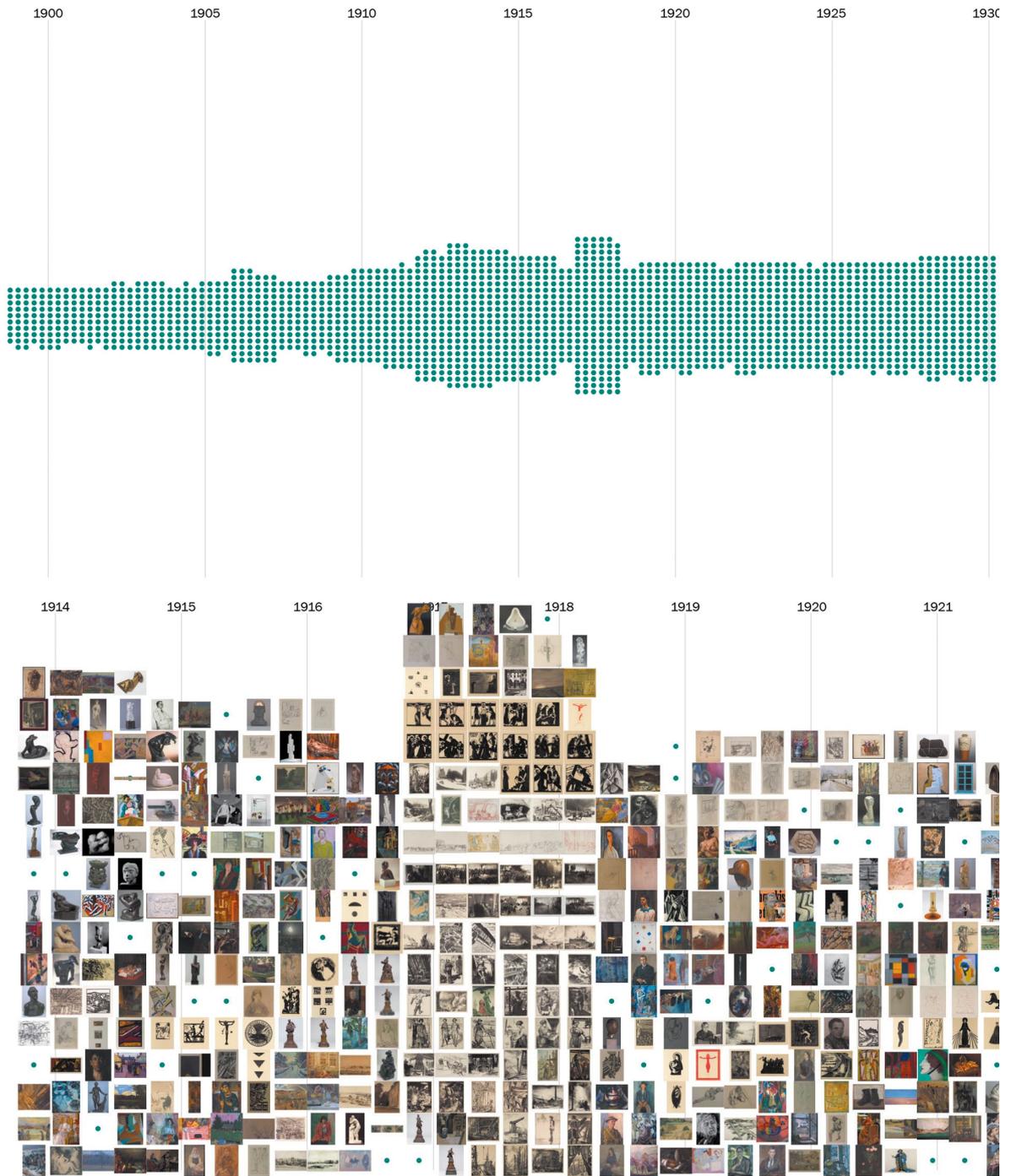
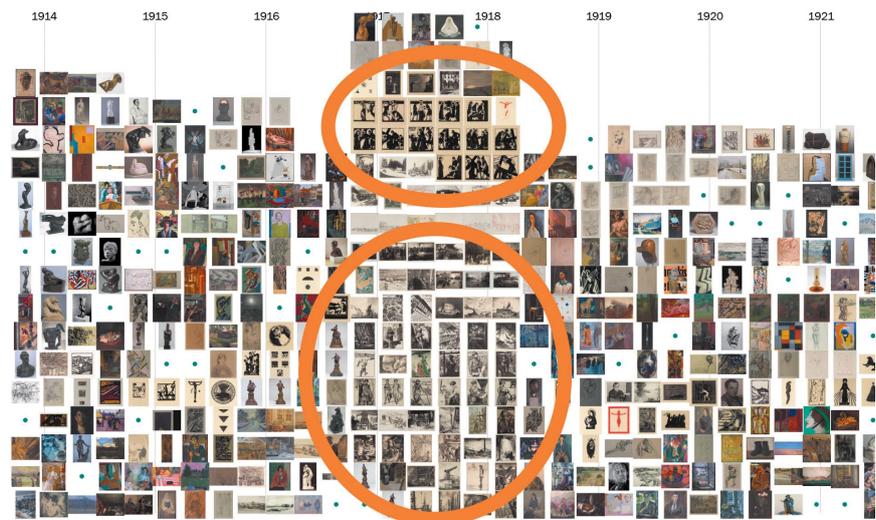


Figure 5.2 – Zooming in on the aggregate shape reveals the dots that compose the overall representation and – further in – the thumbnails associated with each record.

We can see the temporal extent and distribution of the collection. Early works date back as far as the 16th century and individual items are evenly scattered until numbers begin to increase slowly in the 18th century. A spiky cluster of items is visible between 1800 and 1850; a large number of works from this time period appears to be present in the collection. Works from later periods until about 1950 are sparser again and a second, smoother cluster towards the end of the time period indicates a second focus in the collection on contemporary art works.

Zooming in reveals additional detail in the shape of dots that compose the diagram (Figure 5.2). Each dot stands for an individual record - datasets are not arithmetically summarised, as would be the case in a histogram, but only visually aggregated when the number of dots in view would exceed the number of elements that a modern web browser can display without causing noticeable performance issues. The predefined threshold is at 2,000 elements, which can still be displayed smoothly by a standard PC or iPad.

Figure 5.3 – Similar-looking thumbnails reveal clusters of works that share a commonality. For example, they originate from the same artist (top cluster) or they could have formed part of the same acquisition (bottom cluster).

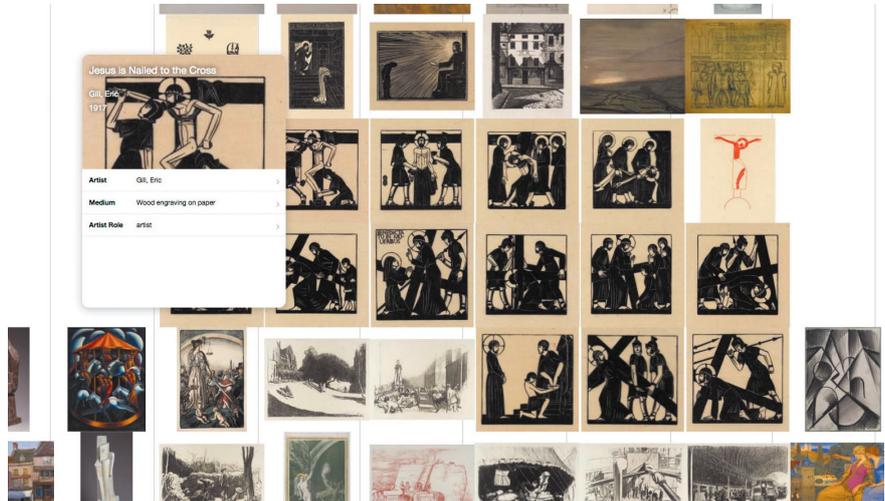


As we zoom in further, dots are replaced by a representative thumbnail image, wherever one is available. This feature - described by a curator as “very much Minority Report” (C4) - is optional, making the visualisation tool usable also for digital collections that lack images, be it because of incomplete digitisation, copyright limitations or because the items in the collection do not lend themselves to be photographed (e.g. audio collections). In this particular case, most records do contain images and panning through the diagram gives a visual impression of the works of art of different time periods in the Tate collection.

All works are arranged by production date and in order of accession, which sometimes results in clusters of similar images being formed where several related works have been accessioned immediately after each other. In Figure 5.3 we can identify a cluster of wood engravings

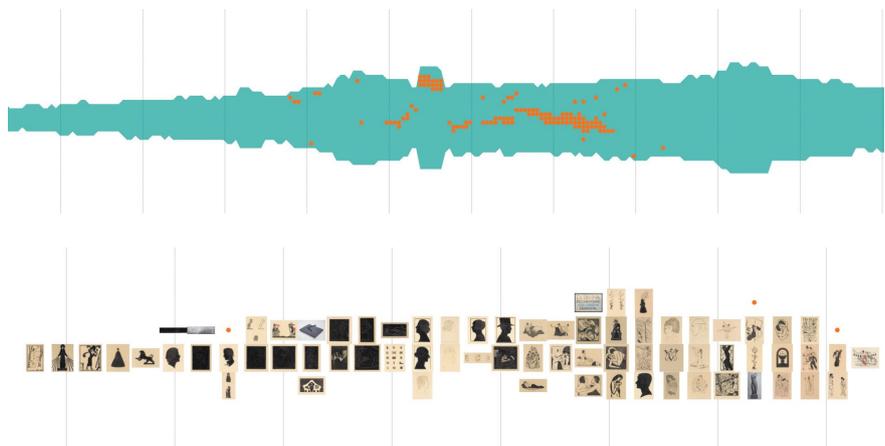
and, below it, a second cluster of similar looking lithographs. The upper cluster is a series of engravings by the artist Eric Gill. By clicking on one of the dots or images, a panel lets us inspect the details of the record (Figure 5.4).

Figure 5.4 – Clicking a dot reveals a customisable panel with details about a record. Each attribute can then be used to filter the dataset.



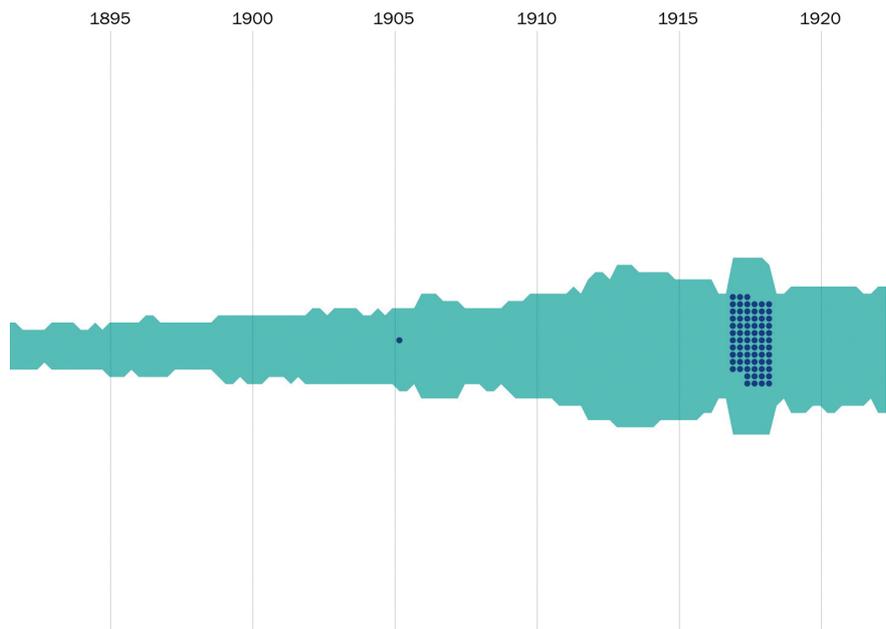
Using the same panel, we can also look for other occurrences of Eric Gill in the collection by highlighting matching records in colour; Figure 5.5 shows a section of the collection with works attributed to Eric Gill coloured in orange. Some of the now coloured dots form contiguous shapes within the collection. These are representative of items that either form a series of works or works that have been accessioned, and probably acquired, as a whole. We can also choose to only examine this particular artist's work by duplicating the corresponding records into a new sub-collection or split and filter the collection based on any other criteria. This behaviour is adopted from P7, where dragging tags split the datasets into sub-collections that can be compared and dissected further.

Figure 5.5 – Each occurrence of Eric Gill in the Tate dataset is highlighted (top) and separated from the remaining dataset (bottom).



The second cluster in Figure 5.3 contains works by several artists, which we find out through probing them by clicking and inspecting the records' details that are displayed in the panel. What is noticeable

Figure 5.6 – Highlighting works by credit line reveals a cluster of items that share the same date of production, along with one outlier.



in these examples is that they all share the same credit line: “Presented by the Ministry of Information 1918” and are all dated either “1917” or “c. 1917”. Highlighting the records by matching the credit line exposes the cluster we already saw, but also one outlier (Figure 5.6): an item that has been dated 1905.³ While the other prints in this selection display scenes of working life, this one is a depiction of Somerset House in London, which may or may not have something to do with the dating of this particular print. We can surmise that also the other prints have in fact been produced at earlier dates, but that they were assigned the date of acquisition as a possible date of production.

3 A print attributed to Sir Moorhead Bone, see <http://www.tate.org.uk/art/artworks/bone-somerset-house-p03007> (accessed 02.01.2016)

Whatever may be the history behind these discoveries – and it would certainly be necessary to turn to other sources and methods to fully understand them – we see how the visualisation tool lets us examine the content of the collection and at the same time discover insights into its history and collecting and acquisition practice. I will review the discoveries the visualisation enabled users to make and the questions that arose in the next chapter – here they serve merely as an example of how the visualisation tool functions when in use.

Architecture

TT is conceived as a JavaScript-based library for visualising cultural datasets along a temporal dimension. It is therefore not a standalone piece of software, but rather a set of extendable components with which a timeline visualisation tool can be created for arbitrary datasets. The public-facing interface can be accessed in a standard web browser, enabling a variety of users to use the tool to visually analyse a given collection.

Some coding knowledge is required in order to set up the visualisation suite to work with a new dataset. This approach allows for a greater level of flexibility in how the visualisation connects to datasets and database fields and how individual fields control aspects of the visualisation - an essential feature for visualising cultural collections, as there is no standard definition of fields and file formats. Even if institutions use the same cataloguing software, the format in which the data is made available through an API or file export can still be arbitrarily defined. The tool has been successfully tested in conjunction with live and static data sources and will be made available open source for public use and further development.

TT is designed as a modular, extendable platform consisting of three types of components:

The **Timeline** component acts as a container for the visualisation and provides a canvas with a horizontal time axis and the ability to zoom and pan the view.⁴

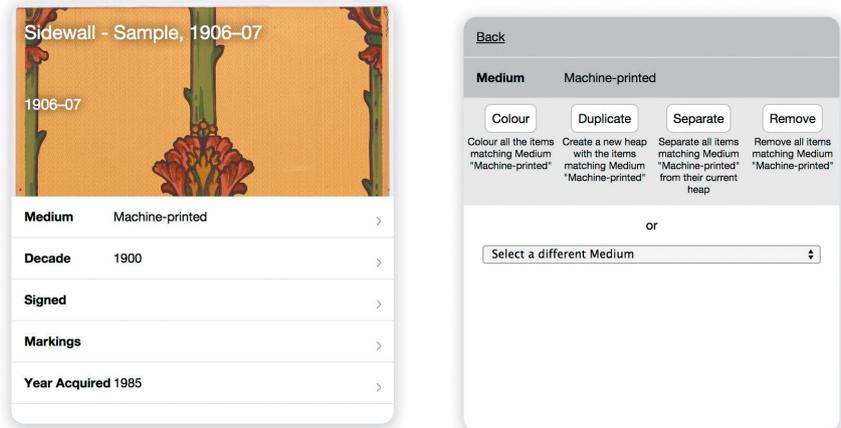
4 The Timeline component has been adopted from P5 (ChronoZoom, see page 151).

Layouts can be attached to the Timeline container and connect to a dataset. At the minimum, a dataset must contain records with a unique identification and one or more date brackets that represent, for example, an item's production date, or any other temporal aspect of the data item. Layouts produce the diagrammatic visual representation of the dataset. Several layouts may be used within a single timeline container to explore different collections simultaneously or to visualise a single dataset according to various temporal aspects. TT offers two layouts; **Bars**, a Gantt-like chart which represents durational periods as rectangular bars⁵ and **Heap**, my standard layout specifically designed for visually representing cultural collections that implements the Temporal Jittering format.

5 The Bars layout has been adopted from P6 (see page 155).

UIs attach to Layout components and offer additional levels of interactivity. The **Panel** component is intended for browsing as well as analysing the dataset (Figure 5.7). One alternative UI component has been developed by System Simulation for including the visualisation on touch screen devices for visitors in museum environments. It enables records and their associated images to be viewed in a slideshow and provides more screen space for descriptive text. As my focus is on

Figure 5.7 – The panel displays a record's details and offers different ways to manipulate and filter the dataset.



analytical use, I will only discuss the former. A Panel displays details of a record on demand and allows a user to filter and manipulate the dataset. Some operations result in a change in the visual representation of the layout – e.g. colouring items that match a specific criterion – while others produce new layouts – e.g. duplicating the visualisation or splitting the dataset in two. A Panel takes two parameters: a record description and a set of fields. The second parameter describes which fields of the dataset are exposed in the panel's details and are available for manipulating the dataset. Often it will be an identical mapping between the database record and the field value.⁶ The architecture also allows more complex transformations, such as splitting the content of a field, generating new fields from existing ones (e.g. a 'decade' field based on date information) or populating a field based on a server request.⁷

- 6 For example, a field may be named "object_title" and could be exposed in the panel simply as "Title".
- 7 This enables the visualisation to communicate directly with an online API, which is crucial for it to work with online collections that are not (yet) available as downloadable files.

Discussion – Temporal Jittering Layout

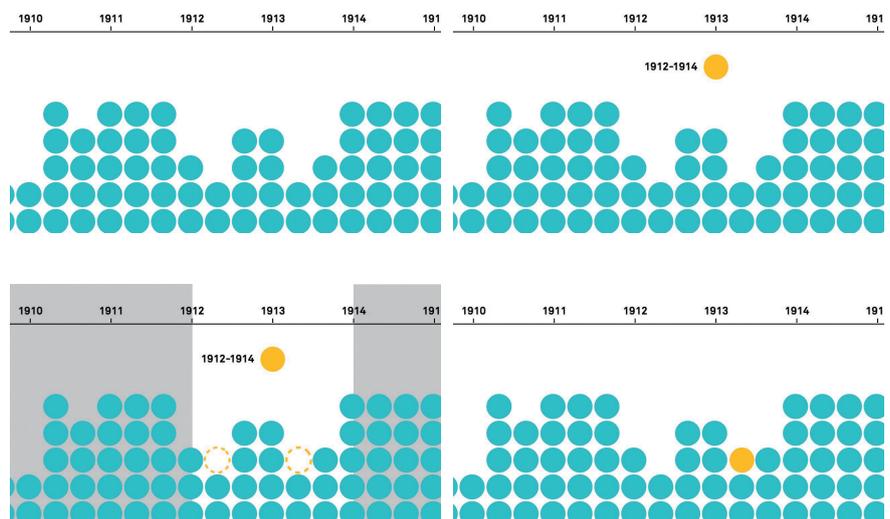
Central to TT is the diagrammatic representation of the datasets, my timeline layout for visualising heterogenous datasets along a temporal dimension. The Temporal Jittering layout is designed specifically around the challenges of visualising cultural data with regards to the uniqueness of individual records, the uncertainty of temporal descriptions and the scalability from a few dozens to hundred thousands of records.

Principle

Temporal Jittering constructs a diagram of a digital collection by exploiting the liberty in temporal positioning granted by the records' date brackets. The strict rule-based positioning of the dots eliminates randomness; the diagram is built up gradually, placing each dot at the lowest available vertical position, while minimising the horizontal distance to the midpoint of the allowed time period (Figure 5.8).

Internally, a table keeps track of the available positions, forming a grid that is overlaid on top of the visualisation and filled bottom up, similar to a game of Connect Four. Once the entire dataset has been processed, the columns of the grid are centred vertically, resulting in a shape similar to the flock in the earlier Britten prototype or, depending on the granularity of date descriptions, a more jagged representation similar to a sound-wave representation of an audio recording.

Figure 5.8 – The Temporal Jittering algorithm determines the position of each record by picking the lowest available slot that is closest to the midpoint of the record's allowed timeframe.

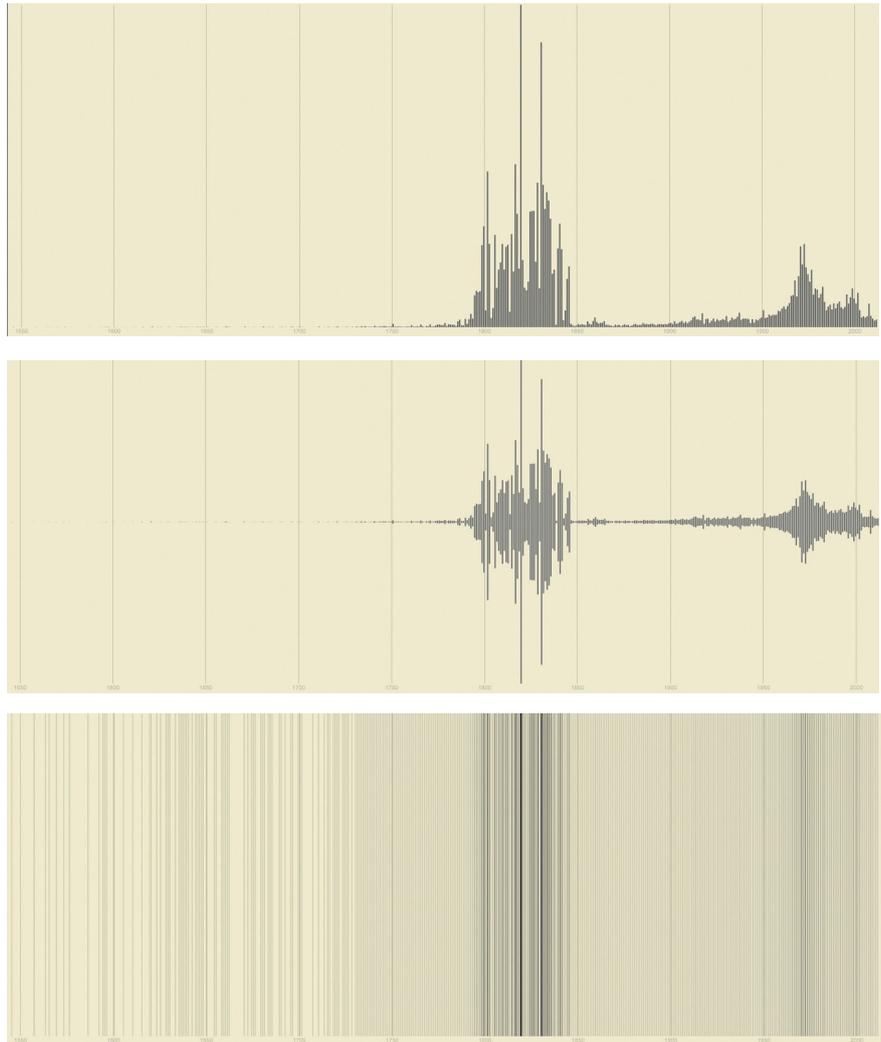


Why Symmetry

The symmetrical layout of the diagram around a horizontal axis is caused by the vertical centring of each individual column. I have debated the rationale and potential benefits and disadvantages of the symmetrical layout with collaborators and myself.

Before the development of the Temporal Jittering algorithm, I have experimented with alternative designs for traditional charts in order to make them look less like statistical graphics. My ambition was to defamiliarise users from these formats that are strongly associated with numerical data in order to allow them to be reinterpreted in the context of cultural data. I produced a series of alternative histograms based on the Tate artworks datasets. Figure 5.9 depicts a standard histogram along with two variations; one with vertically centred bars and one where the height of the bars is constant and instead their values are encoded by their shading.

Figure 5.9 – Three variations of a histogram: conventional (top), symmetrical (middle) and encoded by transparency (bottom). The unconventional formats maintain readability in sparse regions of the diagram.



All versions are able to communicate the ‘denser’ time periods, but what I found striking is how the non-standard versions give a much better idea of the distribution of items in the sparser regions than the original histogram. This observation was one reason to adopt the symmetrical distribution in TT, as it meant that records in areas with little data stand out more and the dramatic differences in quantities – which are typical for collections data – are halved, making the entire diagram more readable.

8 This is why the symmetry already appears in the parallel timeline example in P8.

Another reason for the symmetrical layout is that the visualisation is not meant to show just one of these arrangements, but several in parallel;⁸ be they different datasets, one that has been split into parts or the same datasets ordered according to different temporal attributes. Having an arrangement with a central baseline makes it easier to distribute the different collections vertically and emphasises the lack of a common y-Axis. While the vertical position of the records does carry meaning, it is not possible to read it off a numerical axis. What matters is the vertical order in relation to the other records in the collection, not their absolute position.

The symmetrical layout is not without its downsides however. Some users tried to make comparisons between the records above and below the baseline. They were therefore misled into believing that there are two dimensions visualised, when in fact the symmetry presents redundant information. Symmetry thus creates visual noise by displaying a single quantity in two places. Several users compared the shape of the diagram to a sound-wave. Some appreciated the similarity, as both formats visualise data over a horizontal time dimension, while others found it confusing - expecting to be presented with audio data.

There are valid arguments for both using a symmetrical and the more standard layout of building upwards from a bottom baseline. TT therefore supports both, although I favour the symmetrical variant for the outlined reasons and the largely positive user feedback.

Temporal Jittering versus Histogram

The diagram is comparable, in both use and visual appearance, to a histogram. Histograms typically represent the distribution of numerical data but, as we have seen, they are often used in conjunction with a temporal horizontal axis to visualise the distribution of items, events or records in a given timeframe. Just like in a histogram, we are able to study the distribution of records over time by examining the aggregated view of the Temporal Jittering diagram. However, the construction method of the diagram minimises visual artefacts that occur from the quantisation of data necessary when constructing a histogram. This results in a more compact and honest overview of cultural datasets than a histogram could offer. Additionally, the resulting shape can give useful insights about the way items in the collection are dated.

To build a temporal histogram, a timeframe is sliced into predefined same-sized time-periods - the bins of the histogram - and events are assigned to their respective time-slice. In doing so, we inherently must assume that events can be unambiguously assigned; a condition

Figure 5.10 – Time-based histograms assign events to one of several predefined timeframes.

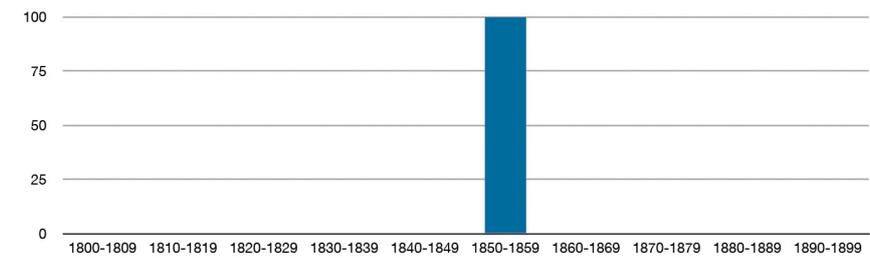
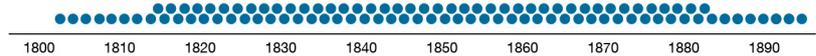


Figure 5.11 – Temporal Jittering algorithm allows records to spread out across their permitted timeframe.



that cultural data usually cannot fulfil, as the date brackets may overlap with the borders of more than one bin.

A possible solution would be to assign records to a bin according to the midpoint of their date brackets. However, this results in a skewed representation of a collection's temporal scope. For example, items dated as 19th century would end up being represented as a single bar exactly in the middle between 1800 and 1899 (Figure 5.10).

The skewing effects of histograms do not occur only in toy examples. Figure 5.12 depicts screenshots from the online collection of the Australian Dress Register, which features an interactive timeline visualisation that could be described as a “stream graph” (Byron et al. 2008) histogram; essentially a stacked histogram organised vertically by garment types and horizontally by decade. Selecting the fragment that represents Skirts made between 1890-1899 brings up a total of 9 records, the second of which is “Ellen Sharam’s Olive Green Skirt” made between 1878 and 1900. Skirts from 1900-1909 include 25 records, however, we are again presented with the same record that was already part of the previous selection. Because its production date overlaps with the bin boundaries of the histogram, it has been counted and represented several times, once for every decade its date range spans. To make things worse, the bin size is adjustable via zoom buttons – in principle a useful feature. However, in this case it means that the smaller the bin size, the bigger the collection appears to be, making it impossible to get any reliable insights on the size, scope and temporal distribution of a collection.

In the diagram format that I propose, records are not summarised in bins, but are individually processed as units. The hypothetical example of 19th century items would, in my diagram, result in a horizontal row of dots growing symmetrically outwards from 1850 (Figure 5.11). Instead of a single peak on this year, the aggregate view would be more representative of a collection's temporal distribution with

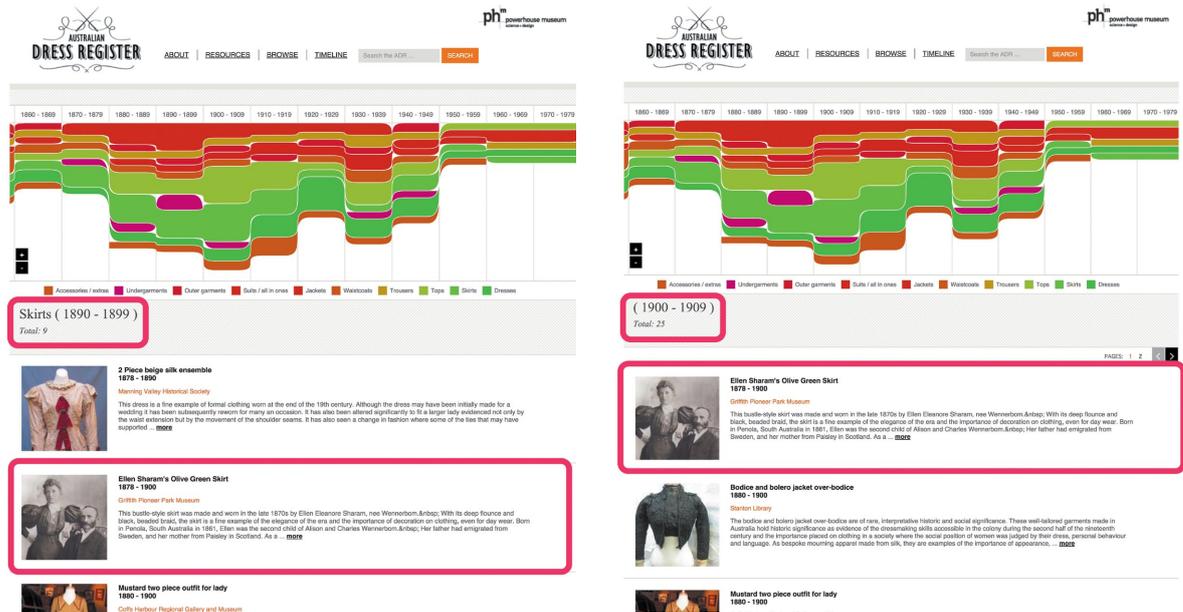


Figure 5.12 – The histogram-based timeline of the Australian Dress Register online collection visualises individual records multiple times. image: Screenshots of www.australiandressregister.org (accessed 12.01.2016)

regards to the cataloguers uncertainty in dating - rather than a representation of a system's inability to model it. Similarly multiple counting of items, apparent in the Australian Dress Register timeline, is avoided as records are only represented once at a single point in time.

The Temporal Jittering layout is more suitable for visualising cultural data over time than a histogram as it prevents many of the skewing effects that arise from the quantitative nature of histograms. It borrows a histogram's ability to display the overall temporal distribution of data, but does so in a way that is more representative of how records in collections tend to be dated.

Britten's Poets / Temporal Perspectives

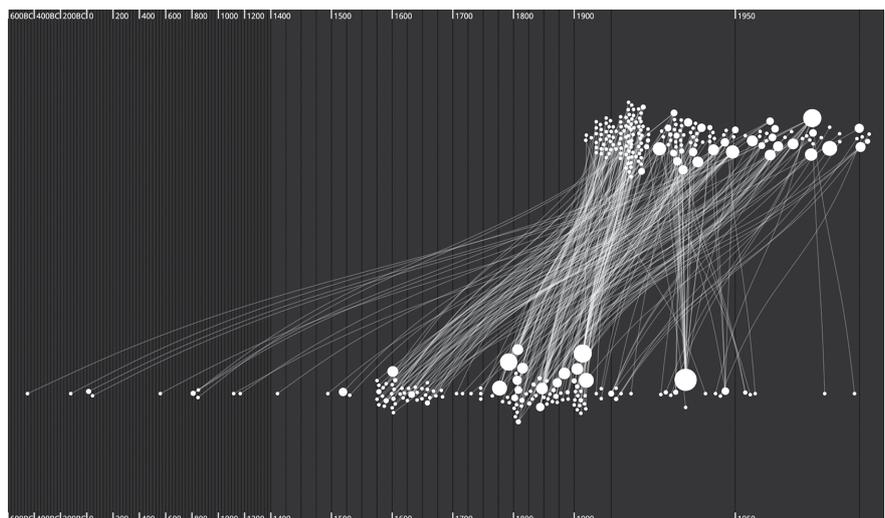
Britten's Poets looks at the temporal relationships between authors, whose texts the composer Benjamin Britten set to music, and the individual works where these texts appear. The visualisation is conceived as a standalone prototype developed in collaboration with the Britten-Pears Foundation, who sought to answer questions such as: at what times did the authors that Britten set to music live? Are there patterns or preferences appearing in Britten's use of poets from different time-periods? Which authors did he only set once, or only during a specific period in his working life? Are there authors Britten was interested in throughout his life? The interactive version of the prototype is published on the foundation's website.⁹ In addition, the curators and I produced a narrated screencast of the visualisation in use as part of an exhibition.¹⁰

9 See <http://brittenweb.hosting.ssl.co.uk/static/interactives/poets/> (accessed 08.02.2016)

10 <http://www.brittenpears.org/resources/song-visualisation> (accessed 08.02.2016)

Britten's Poets (Figure 5.13) is an advancement of the double parallel timeline format that appeared in P8. I have developed the layout further into a reusable timeline format. The Temporal Perspectives layout uses two or more timeline constructs for comparing different temporal aspects of a single collection. In this chapter, I present the implementation of the layout by example of the Britten's Poets visualisation and compare it to similar formats. The next chapter presents insights that the layout is able to reveal in cultural collections more broadly, based on examples of collections by the V&A and MoDA.

Figure 5.13 – The initial view of Britten's Poets giving an impression of the temporal distribution of Britten's compositions and the authors whose works he set to music.



Walkthrough

Britten's Poets is a double timeline that visualises the temporal relations between Britten's songs and the lifetimes of the poets whose writings Britten set to music. Britten's individual songs and song cycles are represented by dots along a timeline in the upper half. A

curved line links each of them to the author of the text, who is represented as a dot in the lower half. All are arranged along a horizontal time axis that covers a timeframe of roughly 2500 years: from the age of Sophocles around 400 BC to Britten's own lifetime (1913-1976) until the present. In the upper timeline the size of a dot represents the number of poems within a song cycle. In the lower timeline the size indicates the number of poems Britten set by a particular author.

A cluster of poets around 1800 is noticeable, showing Britten's fondness for poets of that era. The early eighteenth century is much sparser by comparison, while the late sixteenth and early seventeenth are seen to comprise Britten's favourite period. Selecting a poet in the lower timeline highlights the corresponding works in the upper timeline, and vice versa. We can spot authors who appear throughout Britten's lifetime - such as Alfred Lord Tennyson - and others like Walter de la Mare who only make an appearance during Britten's youth (Figure 5.14). Clicking on a dot 'pins' the current selection and assigns it a unique colour, allowing users to make comparisons and look for correlations across authors and works.

In the default view the time axis is the same for both the upper and lower timeline: a non-linear scale with three different time-scales. The distant past is compressed in order to allow some of the earliest writers Britten set to be displayed within the limitations of the screen size, while Britten's own time period is expanded to prevent the individual works from being positioned too close to each other. Alternatively, a linear timescale may be selected and both axes may be scaled independently to the extent of their respective datasets - the works timeline then ranges linearly from 1915 to 1980, the poets from 600 BC to 2015 (Figure 5.15).

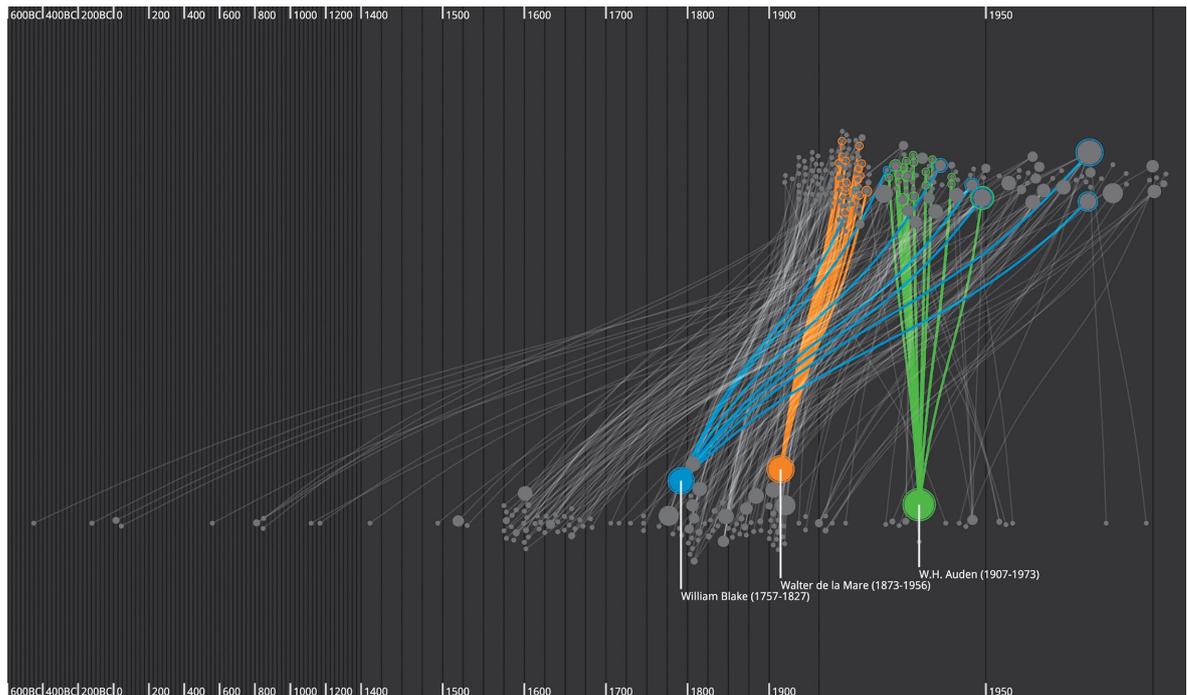


Figure 5.14 – Blake, de la Mare and Auden are three of the most prominent authors in Britten's Oeuvre. Blake (blue) appears throughout Britten's life. De la Mare (orange) is used exclusively in his childhood pieces and Britten's use of Auden's works (green) coincides with their short period of friendship.

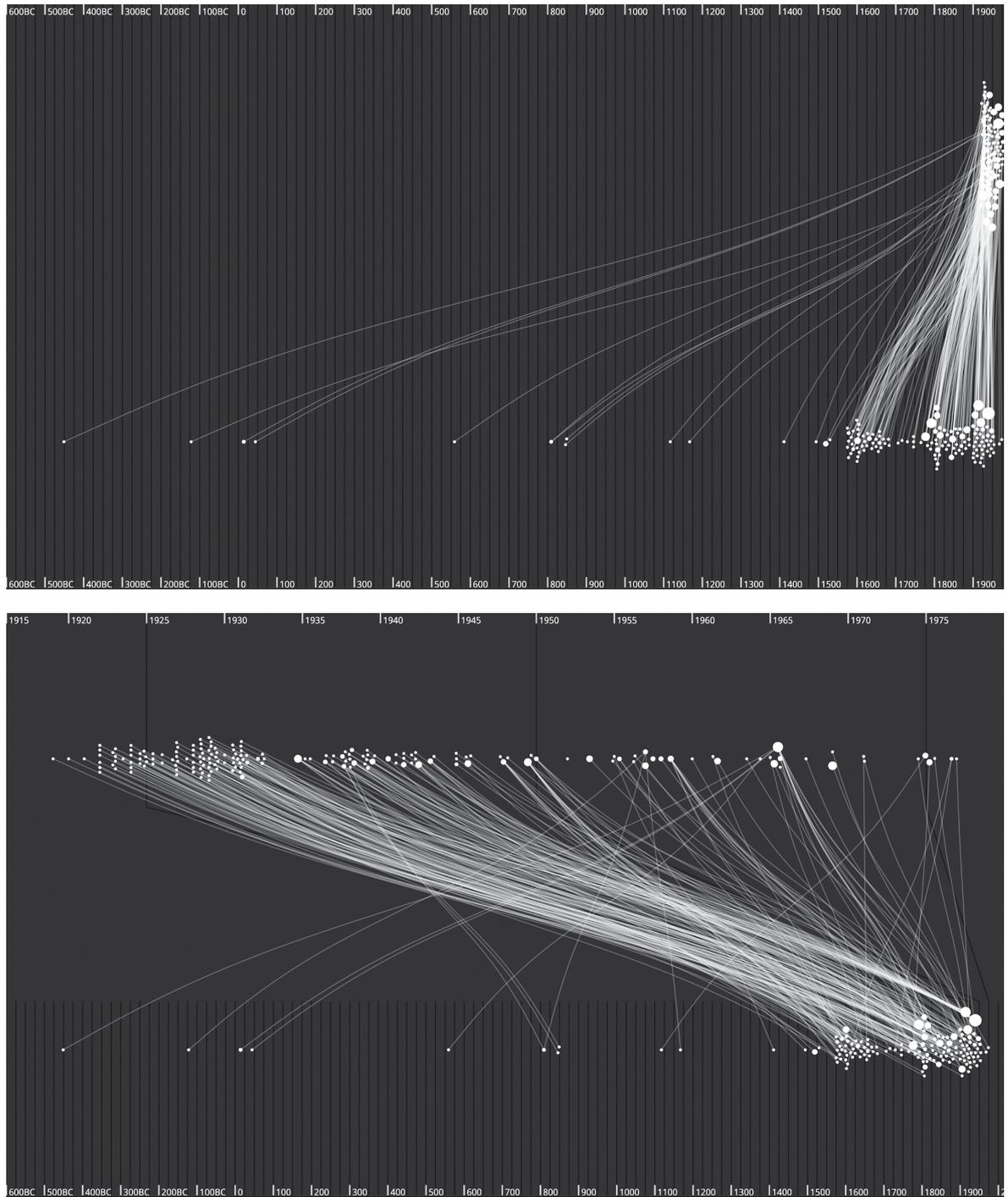


Figure 5.15 – A linear timescale (top) emphasises the time distances between the authors, while scaling each timeline independently (bottom) provides a more fine-grained view of the temporal distributions in each perspective.

Architecture

The individual parts that comprise the Britten's Poets prototype are designed as reusable D3 components. Components are the individual units that implement the functionality of the D3 visualisation library and are organised according to their responsibilities. Scale components assume the task of converting input data to a specified numeric output range, usually pixel values that control the size or position of a graphical mark. Scales also drive the visual representation of axes. For Britten's poets I extended the standard linear timescale to support periods with different time-resolutions, which can be dynamically adjusted, added or removed.¹¹

The core functionality is however provided by my Temporal Perspectives (TP) layout component.¹² A TP layout is initialised by supplying it with a flat dataset. Flat in this context means, that hierarchical data is converted into a tabular format and one-to-many or many-to-many relationships appear, if necessary, as separate individual records.¹³

One or several Perspectives are then added to the layout; these control how the dataset is reorganised and define the (temporal) parameters for generating the different timeline views. Britten's Poets contains two Perspectives that visualise the dataset from the viewpoint of work cycles and authors respectively. In another prototype I used the MoDA dataset to plot the relationships between individual objects in the collection and their usage in exhibitions, books and loans - any field can serve as the defining category and the resulting subset of data may be organised according to any available temporal dimension. The layout takes care of computing the relationship, which then may be plotted using lines or other visual encodings.

11 Comparable to HotZones in Simile Timeline, but not tied to a single time axis. Simile Timeline also lacks the ability to add, remove or change HotZones at runtime.

12 In D3, layout components take care of reorganising input data in order to prepare it to be visualised in standard or non-standard diagram formats. A histogram layout, for example, organises data into bins and computes the width and count (i.e. height) of the bins. D3 layouts often assume a certain visual translation; they however do not enforce or define a diagram's visual appearance.

13 For example a single artist that is associated with three art movements will be translated to three records, one for every movement.

Discussion – Temporal Perspectives

Related Work

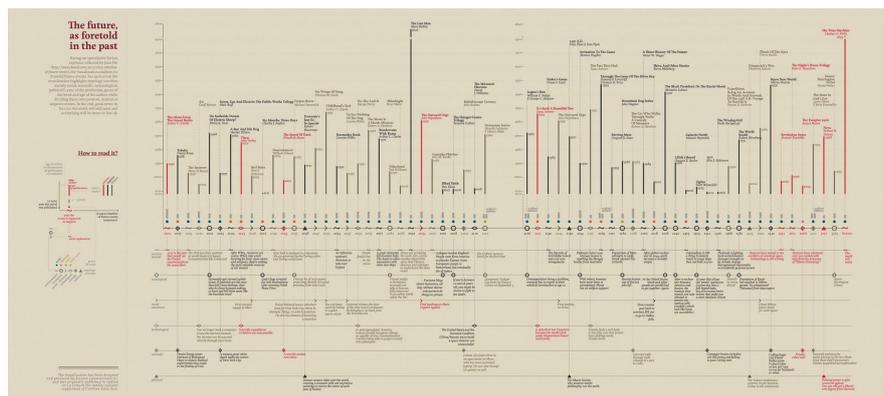
The dataset of Britten's authors forms a typical example of the multitude of temporalities that are often present in cultural datasets. Graphical timelines emphasise a single temporal dimension of events by plotting data on one continuous time axis. However, this has not always been the case. As Rosenberg and Grafton point out, the timeline "in its modern form, with a single axis [...] is a relatively recent invention" (2010, p.14). Often, the purpose of pre-modern visualisations of time was precisely to understand, compare and reconcile different accounts of histories.¹⁴ The tabular Chronicles by the Roman historian Eusebius (CE 260/265 – 339/340) trace, in parallel columns, the chronologies of Jewish, Christian and pagan histories. A recurring motivation for graphical timelines that use several time axes was to make sense of the conflicting accounts of biblical history as told in the gospels (e.g Barr 1938; Priestley 1780). Current visual timelines only rarely include more than one time axis for exploring and comparing separate chronologies. While singular timelines have become a standard visualisation method, there is no established format for visualising multiple chronologies.

A likely choice of format for visualising two temporal dimensions is to borrow from scatter plots. Scatter plots visualise data in a planar space along two axes. By letting the x- and y-axes of a diagram encode separate temporal dimensions, it is possible to study events according to two temporalities.

¹⁴ See also page 74 for my discussion of calendrical time as a model for reconciling different accounts of historical events.

Figure 5.16 – *The future, as foretold by the past* is a printed timeline by Accurat mapping the publishing date of science fiction novels against their fictional time.

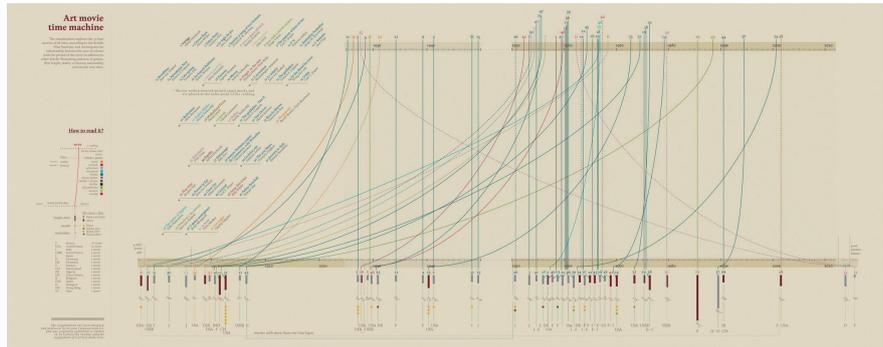
image: accurat, used with permission.



The Italian design studio Accurat has employed this concept in their much publicised diagram *The future, as foretold by the past* (accurat 2014, Figure 5.16), a timeline mapping the years of fictional events as predicted by novels against the works' year of publication. The horizontal, fictional time-axis however is non-linear, which brings this visualisation closer to a bar chart than a true scatter plot.

Figure 5.17 – In *Art movie time machine* two parallel timelines are used to map the year of release of 52 movies to their fictional time periods.

image: accurat, used with permission.



Accurat's portfolio also features examples of designs that use a format of two horizontal time axes. Here the relationship between two synchronous, parallel timelines is indicated – as in my own diagram – by connecting lines. “Art movie time machine” (accurat 2012, Figure 5.17) plots the relationships between the release dates of movies and their fictional time period. Lines curve left or right – depending on whether a movie is set in the past or future relative to its release date – or descend straight vertically, when a film is set in its present time.

In terms of digital timelines, the literature offers two examples that use multiple timelines to enable visual analytics of temporal relationship of events across different timeframes:¹⁵ *SemTime* (Jensen 2003) and *Continuum* (André, Wilson, Russell, et al. 2007; André, Wilson & schraefel 2007). Both are intended as reusable visualisation tools, potentially accepting different sorts of datasets. However, both prototypes have so far not been released – my account of their functionality is therefore based only on the information available in the corresponding publications.

SemTime allows users to explore data using sets of vertically stacked timelines which each can be navigated and scaled independently (Figure 5.18). Relationships are indicated by arrows that connect events within and across timelines. *Continuum* visualises connections in a similar way, but only allows two simultaneous timelines positioned side-by-side. In discussing their own and previous work the authors make a compelling case for the value of multiple timelines to show relationships across time. Both examples demonstrate new research questions that such designs could answer. Nevertheless, despite the evident potential, the concept has hardly been adopted or developed further.

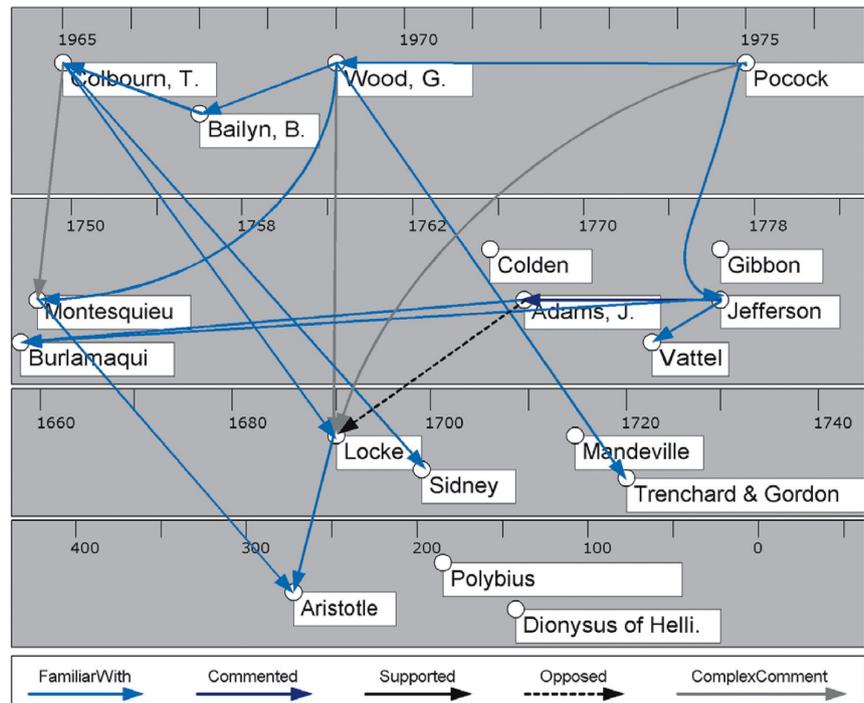
The proposed scenarios for the use of *SemTime* – and indeed most visualisations that focus on temporal relationships – are in tracing and retelling a course of events: identifying and studying individual relationships between incidents. Jensen gives the example of wanting to study the Watergate scandal: “one might wish to answer questions such as ‘what did the President know, and when did he know

15 There are other digital timeline tools that focus on visualising temporal relationships (Cousins & Kahn 1991, Plaisant et al. 1996, Chittaro & Combi 2001, Miksch 2004, Card et al. 2006, Burch et al. 2008). The discussed projects, however, are unique in their use of multiple, independent time-axes.

it?" (Jensen 2003). The focus is on close reading of relationships. Even Continuum, which is specifically aimed at identifying large-scale patterns in datasets, nevertheless displays relationships only in the detail view, not in the aggregated overview. My design instead allows for a distant reading of relationships, in addition to enabling users to study individual events up close. I will discuss the implications and possible insights that this enables as part of the next chapter.

Figure 5.18 – In SemTime, multiple views on a timeline can be stacked and scaled independently in order to trace the relationships of individual events.

image: Jensen (2003)



Non-Uniform Time Scales

Linear time scales always map equal periods of time to equal units of space, while non-linear time scales may be inflected. A remarkably large number of digital timelines use a linear time scale (e.g. Plaisant et al. 1996; Yiua et al. 1997; Havre et al. 2002; Card et al. 2006; Chang et al. 2009; Alonso et al. 2010; Baur et al. 2010; Fouse et al. 2011; Hoffmann et al. 2012; Nowviskie et al. 2012; Straup Cope 2013) - making it the de-facto standard temporal and graphical model.

Non-linear time scales often appear in digital timelines as a consequence of graphical metaphors of space, such as in the case of the Guardian timeline or similar three-dimensional timeline projects (Kullberg & Mitchell 1995; Korallo et al. 2012). MacKinlay et al. (1991) first proposed the use of perspectival mappings as a way to use screen-space more efficiently when visualising linear data. When used with temporal data, their Perspective Wall resembles a conventional interface that has been folded twice, with the past and future extending 'into' the screen space while the present 'wall' faces the user. The visualisation employs distortions to provide focus and context in datasets

within a single view: the wall is in focus, while the data on the walls provides the context in a condensed display.

Another popular technique uses simulated fisheye distortions (Furnas 1986), a metaphor which has been applied to temporal data, for example in Bederson et al. (2004). Again, the assumption is “that fisheye views are appropriate when users need to see details of some specific items in the context of a large information space” (Bederson et al. 2004).

Providing focus and context is the recurring motivation for using distortions and non-linear scalings. I argue that, in the context of timeline visualisations, there is more than this to be gained by considering alternative projections. As Boyd Davis et al. write:

So often is historiography obliged to deal with dramatically uneven densities of data, especially over extended historical periods, that it is impossible to deny the value of various non-linear [time] scales. These include not only logarithmic scales but pragmatic adjustments of scale based on the density of data at that point. (2013, p.254)

Common to Accurat's “The future, as foretold by the past” (2014) and my own work with Britten's Poets is the challenge of comparing two vastly different timescales. In the case of science fiction novels the publishing dates range from early 19th century to 2012, while their fictional dates range up to the year 800,000. The lifetimes of the poets Britten set cover a timeframe of roughly 2,500 years, while he composed his works during a period of 57 years.

Accurat addressed the issue of the large fictional timeframe by using a sequential timescale for the horizontal axis, placing each item an equal amount of space from the previous one regardless of the timeframe that separates them. The vertical time axis for the publishing dates is almost linear, with a break in order to account for one outlier that has been published in 1826, while the other novels all were published after 1950.

In both axes the decision to refrain from using a true linear timescale is justified by the limitations imposed through the printed medium. However, as a result we are unable to comprehend the vast timeframe that the visualisation actually occupies. Neither is it possible to visually discriminate any patterns in the fictional timeframe in order to find out which time periods were most popular among science fiction writers and how much of an outlier “The Time Machine” by Herbert G. Wells, with its narrated time dating to the year 802701, actually is.

The merits of a linear timescale have already been stressed by both Barbeau Dubourg and Priestley (Barbeau-Dubourg 1753; Priestley 1764).

- 16 "Les Cartes représentatives des temps peuvent & doivent être toutes construites sur le même point, qui exprime constamment des années; & comme cette uniformité dispense d'y joindre des échelles de rapport, on n'a pas besoin de les étudier le compas à la main." – "Maps of time can and must be constructed with the same scale, expressing a consistent number of years; and because this uniformity dispenses the need for an accompanying scale reference, one does not have to study them with a pair of compasses in hand." (Barbeau-Dubourg 1753, p.7)

- 17 Colin Graham (1931-2006) and Yevgeniy Yevtushenko (born 1932)

- 18 This method is used by Priestley where precise dates were unavailable. Priestley used the time "when it is said that a writer *flourished at*" (Priestley 1764, p.11) to position an individual on a timeline.

Dubourg argues that maps of time – in contrast to geographic maps – can and must be constructed with a uniform scale and that this approach eliminates the need for any visual or external indication of scale.¹⁶ Priestley similarly argues in favour of uniform scales when he criticises the use of a non-linear timescale in a "Chart of History imported from France" (1764, p.8), an English version of Bruyère's *Mappe-Monde* (1750b). Bruyère uses a non-linear timescale to accommodate the variations in data density in different time periods. According to Priestley a false impression is made by this non-uniform scale which, lacking sufficient visual notice, the mind of a viewer is not capable of correcting.

Both linear and non-linear timescales bring with them their own advantages and disadvantages. In digital interactive timelines however, we are luckily able to implement both. Linear timescales are most expressive in communicating the extent of historic time periods and in accurately representing the relative timeframes between events. When graphing temporal relationships within cultural datasets that often comprise large timeframes with unevenly distributed events the value of a non-linear timescale becomes obvious.

Positioning

In the example of Britten's Poets, the creation of a work is compared with a poet's lifetime. As the layout demands that records are assigned a point in time, I had to decide what that date should be: the date of birth, death, or a different point in time. At first I did not give it too much thought as it seemed logical to me to position all poets by their birth year. As a result it appeared as if all but two writers¹⁷ predate Benjamin Britten, whose life is indirectly represented by his works. In fact, however, many of the writers Britten set were his contemporaries, yet this method of positioning authors fails to communicate that (Figure 5.19).

Again it is important to remember the inequality of the comparison made in this parallel timeline – lifetimes of authors to uses of their works. We can safely assume that the authors were not born as writers, but that it has taken them a certain amount of years to acquire the necessary skills and gain recognition as poets. One could position the authors according to the peak moment in their career,¹⁸ but defining it would need additional data and indeed subjective judgment. Moving all authors forward by a set amount of years to give a more realistic image of their productive age might be a viable solution,

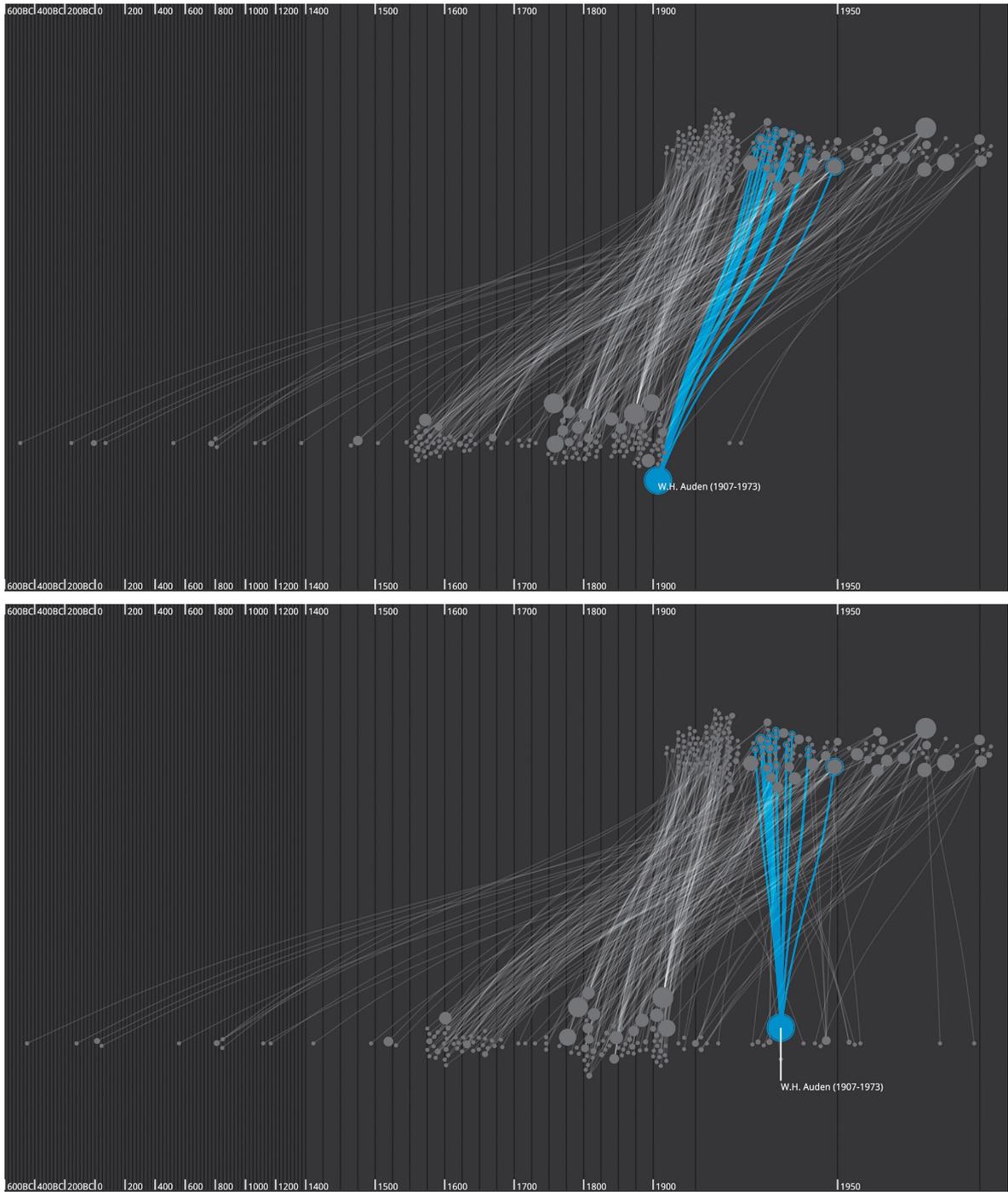


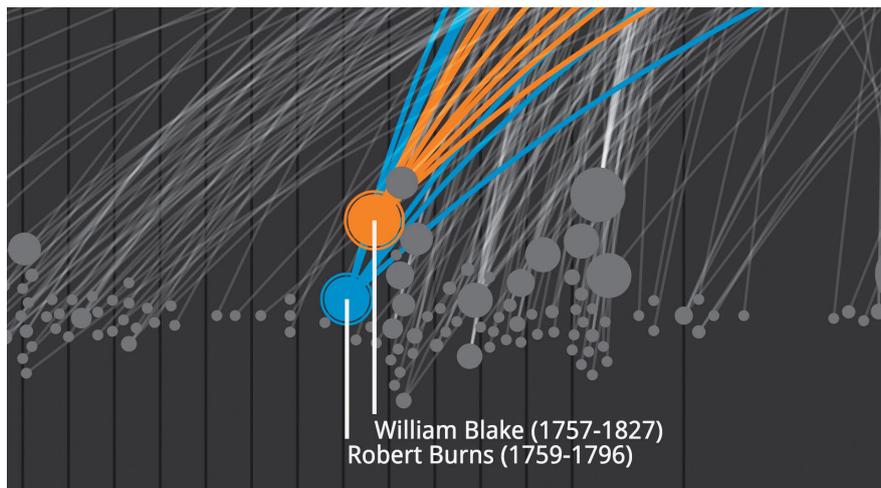
Figure 5.19 – Positioning authors by year of birth (top) moves most writers outside the timeframe of Britten’s working life – even Auden, who was a contemporary and friend. Positioning them by midpoint (bottom) communicates their shared timeframe.

but this could result in poets that died early to appear after their year of death.

Instead, the approach I have taken is to position the authors by the midpoints of their lifetimes. Now, Britten's contemporaries are recognisable as such. W.H. Auden, for example, was a close friend and collaborator of Britten after they met in 1935, and until an "irrevocable quarrel" (Hensher 2009) divided them. This relationship is visible in the way the connecting lines ascend relatively straight up from Auden to a compact group of Britten's works.

The curators, upon interrogating the visualisation immediately spotted an apparent error. Robert Burns is located before William Blake, although Burns was actually born two years later (Figure 5.20). It turned out to be an anomaly resulting from the rationale behind the positioning of the poets. Burns died at the age of 37, while Blake lived almost twice as long. We decided, however, to not 'fix' this occurrence, but rather that positioning the dots consistently by their temporal midpoints is the better option.

Figure 5.20 – A side-effect of positioning events by their temporal midpoint: Blake was born two years before Burns, yet appears later in the timeline. This is because Blake also lived twice as long – the position encodes the author's lifespan, not only the date of birth.



6

“How is knowledge being generated? How is new knowledge being generated with the same stuff? You’re generating new knowledge by visualising it, but then that is visualising the fact that new knowledge is being generated.” (C12)

6 Evaluation

In this penultimate chapter, the developed visualisation tools are put into practice and I demonstrate what kind of discoveries they enable in digital collections in collaboration with expert users. First, I focus on specific insights that we are able to gain through the two paradigmatic visualisation tools I presented in the previous chapter. The following discussion will provide empirical evidence for their contribution to knowledge discovery in existing digital collections. Next, I will assess the kind of insights the visualisation offered to the expert users by offering an account of their perceived utility and the discoveries they are able to make that would be invisible to a casual user. These are partly ‘pragmatic’ insights that arise from studying digital collections through the proposed visualisation tools. However, to a greater extent they represent conceptual insights stemming from an increased awareness of the consequences of the digital turn in their research that results from our co-investigations and from partaking in the design process.

Mediated Insights – Timeline Tools

TT is unique in its ability to not only represent the content of digital collections, but also the biases, patterns and strategies that hide behind the cataloguing data. The presented insights tie in with previously discussed issues around embedded interpretations in collections, the difficulty of modelling dates digitally and assigning dates to cultural artefacts in the first place, and changing collecting practices in the history of institutions.

Cataloguing Biases

In my review of the literature on collections and through my own conversations with professionals I came across the issue of biases in collections; collections are not a neutral representation of the past, they have been shaped by various personal, institutional, administrative and technical influences. Collections data, the cataloguing information, does not record what “is known about it” (The British Museum 2013), but what individuals with their own interests and biases know and have recorded, or were able to record in cataloguing structures and conventions, which themselves produce a bias within the recorded data. Suitable visualisation tools, I suggested, could be able to provide further insights and evidence of these biases.

In developing P6 I encountered the disproportionate presence of works by J.M.W. Turner in Tate’s catalogue. Through analysing the collection in my last timeline tool I will demonstrate how visual analysis can shed a better light on this, and similar, biases.

Figure 6.1 – The artwork dataset of the Tate exhibits a large cluster after 1800 when examined within TT.

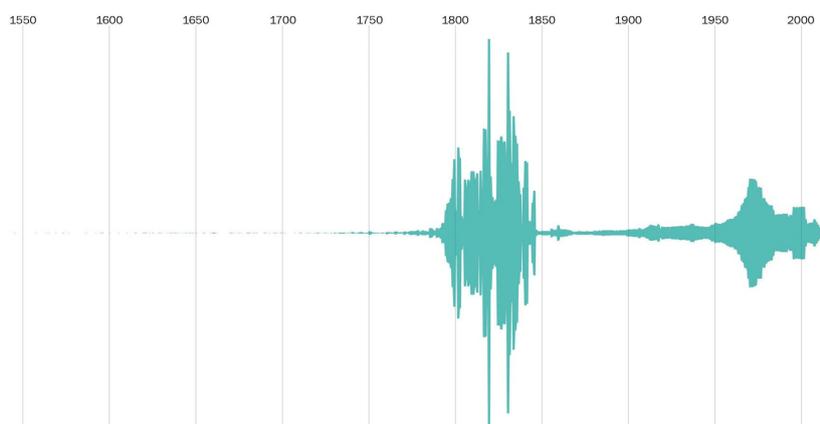


Figure 6.1 shows the overview after loading the Tate dataset into the tool, organised by the works’ assigned date of production. The visualisation automatically adjusts the zoom level so that the time period in view matches the temporal extent of the collection. I pointed out the cluster at 1800-1850 previously; I will now examine it by zooming in closer (Figure 6.2).

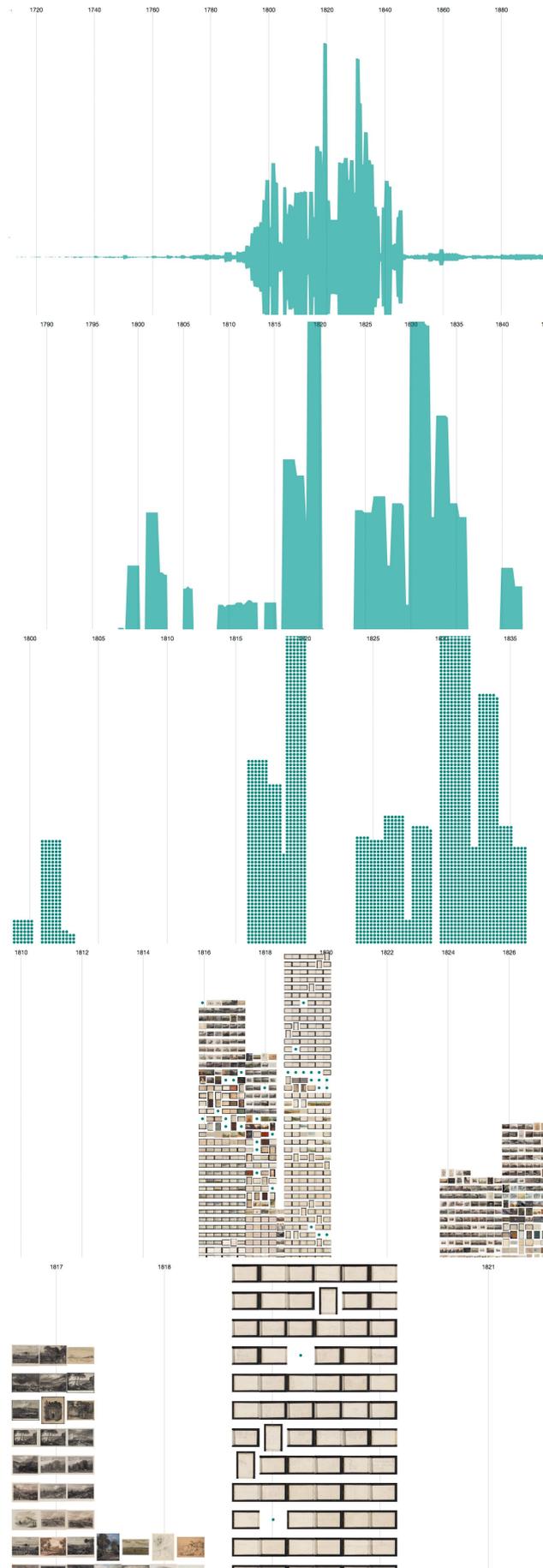
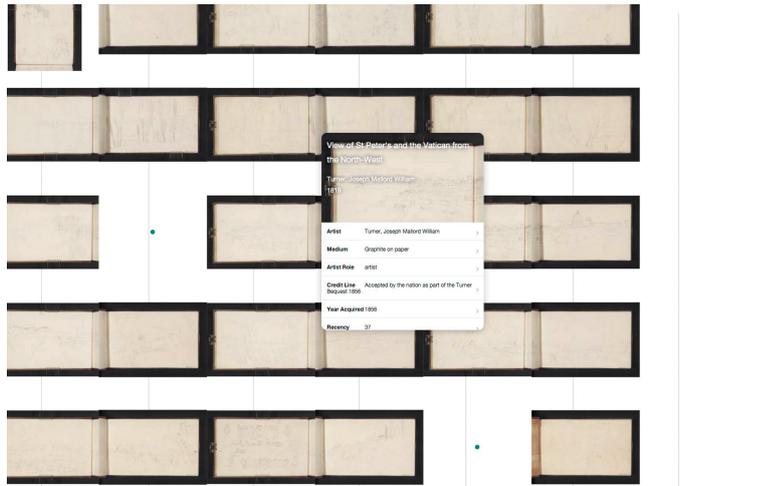


Figure 6.2 – Zooming in on a cluster between 1800–1850 reveals how it is composed of suspiciously uniform thumbnail images.

Figure 6.3 – Every page in J.M.W. Turner's sketchbooks is catalogued as an individual work.



It turns out that the cluster is mostly composed of works by J.M.W. Turner – an anomaly that we already recognised. It appears, however, that most of these works are actually individual pages of his sketchbooks (Figure 6.3). His overarching dominance in the Tate collection is not only a result of his prolific output of paintings, but of the decision to catalogue every single page of his sketchbooks as a work of art. As far as the digital catalogue is concerned, a page in Turner's sketchbook is on a par with other 'finished' artworks in the collection. This observation has been confirmed in a comment on the early findings I published online (Kräutli 2013) and explained by Tate's developer Richard Barrett-Small:

The Tate holds the Turner Bequest on behalf of the nation, which comprises a large number of Turner sketchbooks. Each page of these sketchbooks is classified as an individual artwork on paper, which makes up the lion's share of this rather singular collection.¹

¹ see <https://github.com/tategallery/collection/issues/16> (accessed 12.01.2016)

George Oates, museum technologist and board member of the British Library Labs, picked up on this occurrence in her keynote at the 2014 Museum and the Web conference, naming it the "Turner Problem" – the fact that the skewing of data that appears in the visualisation is not an error in the design that should be fixed, but representative of cataloguing conventions:

What Florian didn't know was that every page in all of Turner's sketchbooks has been catalogued. Instead of being an anomaly, it was a true representation of Tate's decision to elevate the Turner materials in their catalog. (Oates 2014)

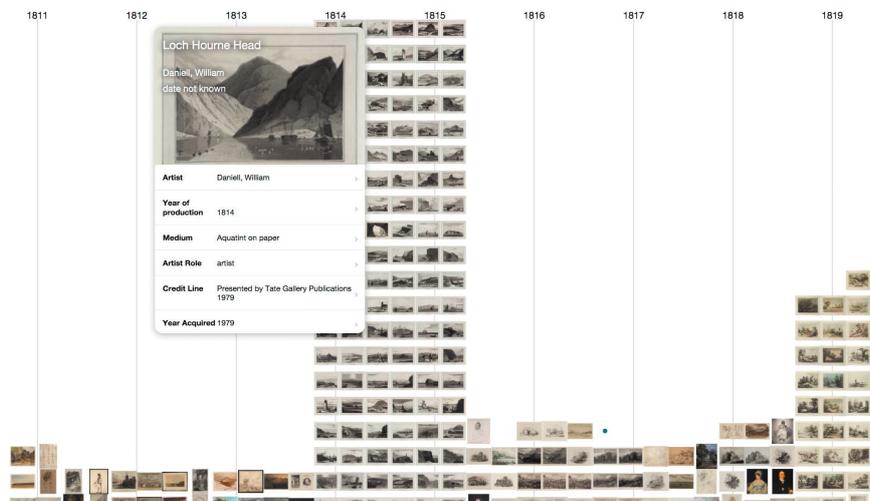
Oates also comments on the ability of the visualisation to provide evidence for collection's biases:

Every catalog is different, no data is perfect, every cataloger’s decisions are different, every institution is different. I like that these differences are becoming more apparent. (2014)

To take a closer look at the “Turner Problem” we can separate all the Turner-records from the non-Turners. We can now either examine the Turner-related items separately, or continue with the remaining items in the Tate collection. These show a more even temporal distribution, although a suspicious peak is still present around 1815. As we zoom in, it becomes apparent that it represents a set of prints by the artist William Daniell (Figure 6.4). Their date of production is, in fact, unknown; the text-based date field states “date not known”, while the numeric dates have apparently been set to 1814. It’s unclear why, though we further learn that they have previously been part of the Tate Publications collection and have been included in their art collection in 1979.

Highlighting records that match the credit line reveals other works that were part of the same acquisition batch. These records are all associated to William Daniell, revealed by highlighting the corresponding records in a different colour, and they are also the only works by this artist that the Tate owns.

Figure 6.4 – Another suspicious cluster remains after removing Turner from the diagram. These records all relate to a single acquisition and have probably all given the same production date, although they might have been produced at different times.



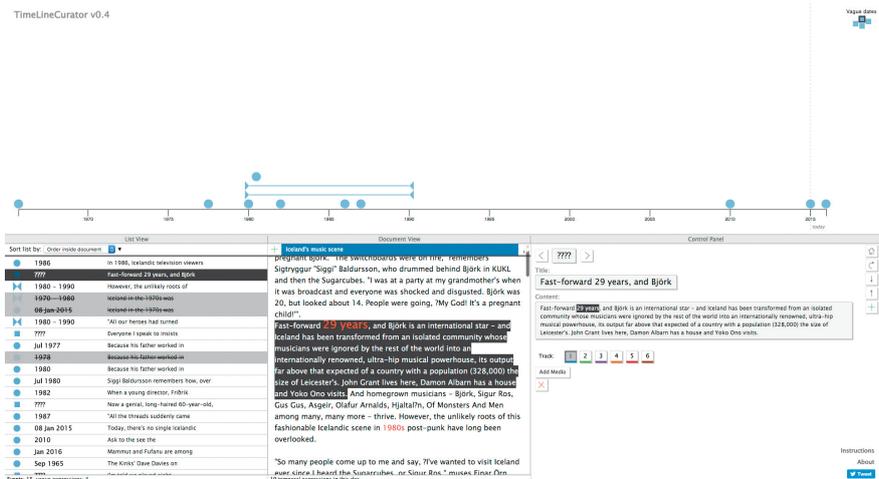
Missing Dates

Daniell’s works have no known production date according to their textual dates, yet they appear in the timeline because their numeric dates have been assigned. Other records that have no numeric descriptions of their production date are not displayed in this view. In Tate’s example this currently corresponds to 5,392 records, or 0.08% of the dataset.

Timelines display events by positioning a graphical mark according to their temporal location, which inherently brings up the question if – and if yes, how – events with missing dates should be displayed. Existing examples of digital timelines employ different strategies.

Figure 6.5 – TimeLine Curator extracts temporal information from unstructured text and generates a visual timeline. Events with vague dates appear as rectangles in the top right corner.

image: screenshot of <http://tl-generator.herokuapp.com/> (accessed 24.02.2015)



- 2 The tool generates timelines based on TimelineJS (see page 297).
- 3 My account is based on personal communication with the authors. The cited paper does not describe how unknown dates are dealt with.

TimeLineCurator (Fulda et al. 2016), a tool for generating timelines semi-automatically from unstructured text, displays “vague dates” as squares in the top right corner of the view (Figure 6.5). When the timeline is exported to its final format,² vague dates are automatically removed. HiT (Boyd Davis et al. 2013), a prototype visualisation of part of MoDA’s database places all events with missing dates at the left edge of the screen.³ Most digital timelines simply omit events that have no associated dates.

Arguably, not displaying undated events might be a sensible approach: users looking at data on a timeline do so through ‘time-tinted’ glasses – non-temporal data will be invisible to them. Mauri et al. (2013) therefore propose to implement multiple “views” when visualising cultural data – lists, maps, timelines, etc. – and allowing users to switch between them.

Visualisations are not neutral representations; their format and focus dictates what is visible. However, users might not be sufficiently aware that their view is incomplete, assuming instead that a visualisation shows all available data, as my evaluation of P4 and P6 indicated.

Early iterations of my visualisation prototype falsely interpreted missing dates as zero point in the twentieth century – 1.1.1900 – or Unix time – 1.1.1970. Anomalies that these interpretations caused to appear in the visualisation were spotted quickly by the curators. However – lacking the background knowledge in computational date interpretations – they were taken at face value as occurrences in the dataset, and not a result of missing dates.

In my discussion of time as a framework for sense-making (see page 70) I argued that temporal data is never truly absent in digital datasets. When production dates are missing in cultural records, there are other temporal aspects that may be used to visualise a dataset on a timeline. Exploiting these constitutes my proposed design principle to visualise multiple temporalities and it is also a viable strategy for

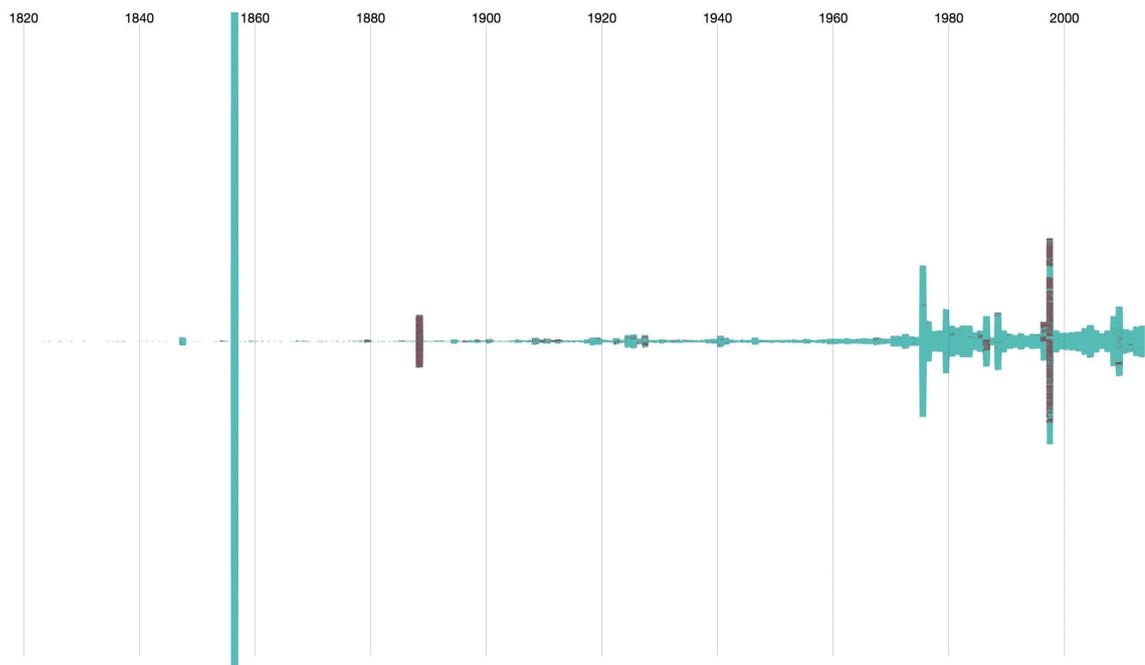


Figure 6.6 – The artworks in the Tate dataset, visualised by dates of acquisition. Highlighting items with no production date reveals records that were missing in the earlier views that visualised Tate’s data by year of production.

dealing with missing dates – which in this context denote dates that are merely missing along one temporal dimension, while others may still be present.

Figure 6.6 visualises the Tate collection from the perspective of acquisition dates. These are set for all but 45 records.⁴ Visualising by acquisition reveals an institution’s history. For the Tate, it begins in 1820, although it has only been an independent institution since 1955 (The National Gallery n.d.). Originally it was a part of The National Gallery that grew out of a donation of the industrialist Henry Tate and led to the opening of a separate gallery for British art. When J.M.W. Turner died he bequeathed his work to the National Gallery, visible in the diagram by a vertical strip in 1856.

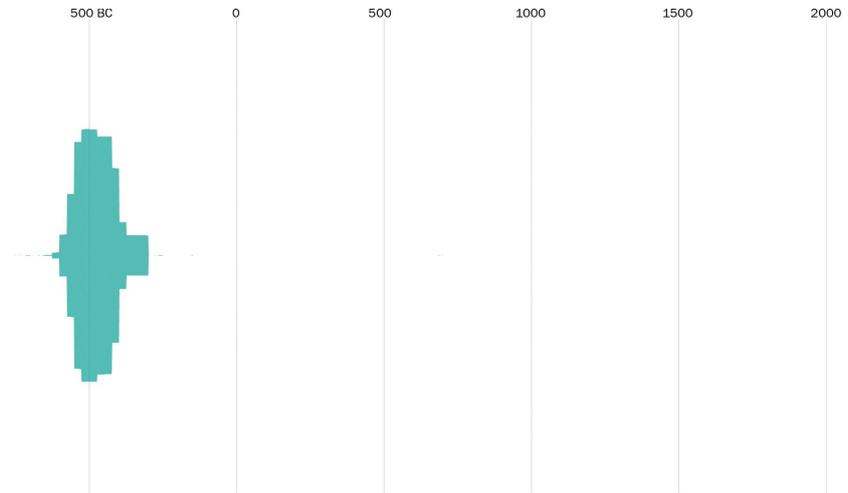
⁴ 42 of these 45 records contain production dates, which leaves three records that appear in neither of these views. These records could be revealed by exhausting other temporal attributes, such as the timestamp a record has been updated – a value that is defined for all records – or using date information that is assigned to a record indirectly – via associated artists, movements, accession batches, etc.

TT allows any dataset field to be used as filter, including temporal data. In Figure 6.6 items with missing production dates are highlighted; these are the records that were absent in the previous view. Block-shaped clusters form where a batch of undated works have been acquired. Positioning these works by acquisition dates provides an upper bound of their possible date of production and neighbouring records provide context. Undated works appear among other records with which they share an informative temporal commonality. When they are positioned ‘outside’ the main timeline diagram, as some examples do, the only shared temporal relationship of these separately visualised clusters is their lack of production dates, which does not allow for much insight about their history.

This discussion of missing dates puts a caveat on my earlier observation of the importance of visualising datasets completely; what counts as a complete dataset often depends on the viewpoint, and the

specific viewpoint determines what the dataset consists of. We are not only looking at data of the artworks in the Tate collection, we are looking at datasets of art production and art acquisition. Time creates its very own viewpoint; seeing everything necessitates looking from several perspectives of time.

Figure 6.7 – More than 100,000 records make up this diagram that visualises vases in Oxford’s Beazley Archive by their estimated date of production.



Dating Strategies

Timelines for distant reading should enable insights about overall patterns and occurrences in a dataset. Histogram-based overview timelines allow users to inspect the temporal distribution of a collection. As explained earlier, the TJ layout I developed minimises artefacts that result from the incompatibility between the way cultural items are dated and the assumptions embedded in the construction of a histogram by utilising the tolerance permitted by the date brackets. The layout algorithm operates on the level of individual records, but it produces an emergent pattern that can reveal new insights about the strategies that have been used when dating items in a collection.

Figure 6.8 and Figure 6.9 depict the objects database of the Cooper Hewitt Smithsonian Design Museum collection organised by date of production. In comparison with the Tate’s collection that served as an example before it’s noticeable that the diagram is much more evenly distributed; likely a result of a conscious decision to offer a representative view of the history of industrial design. Spikes, appearing at regular yearly intervals throughout the diagram indicate sets of objects that carry a precise year as a date. Between the years, a continuous band stretches horizontally throughout the visualisation with only minor variation in thickness, suggesting that the objects that appear here have been dated with a much coarser granularity – five years or more, visible by the lack of sudden changes. One anomaly stands out from this pattern: a thicker portion in the late 1930s turns out to originate from a set of items dated 1928-29.

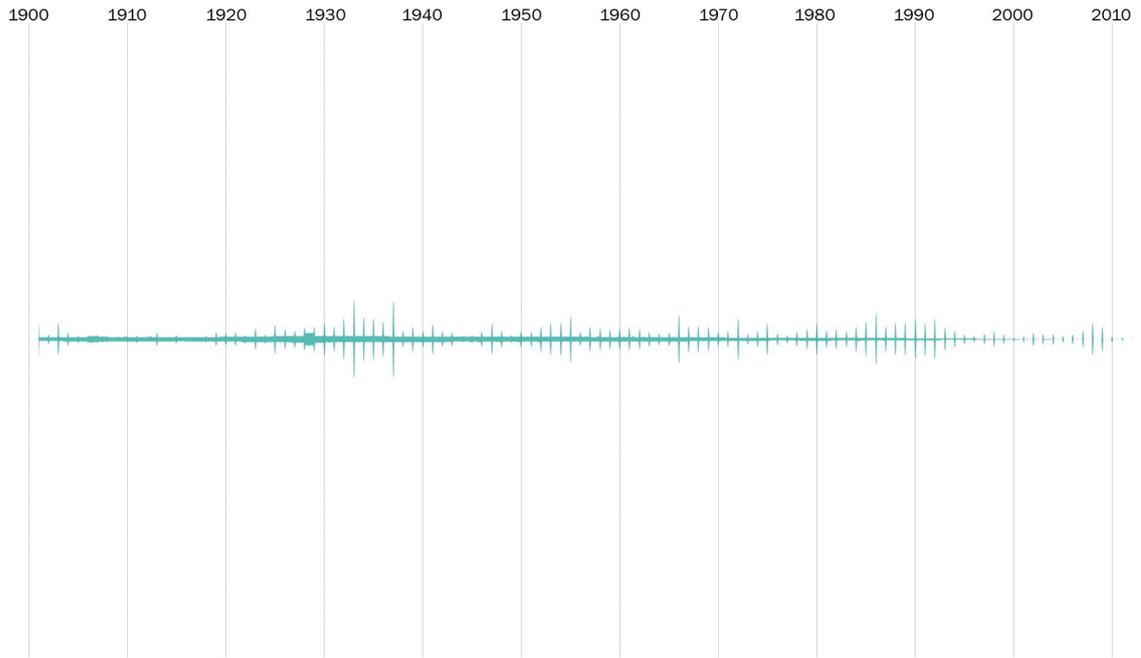


Figure 6.8 – Cooper-Hewitt's object records are visualised by production date.

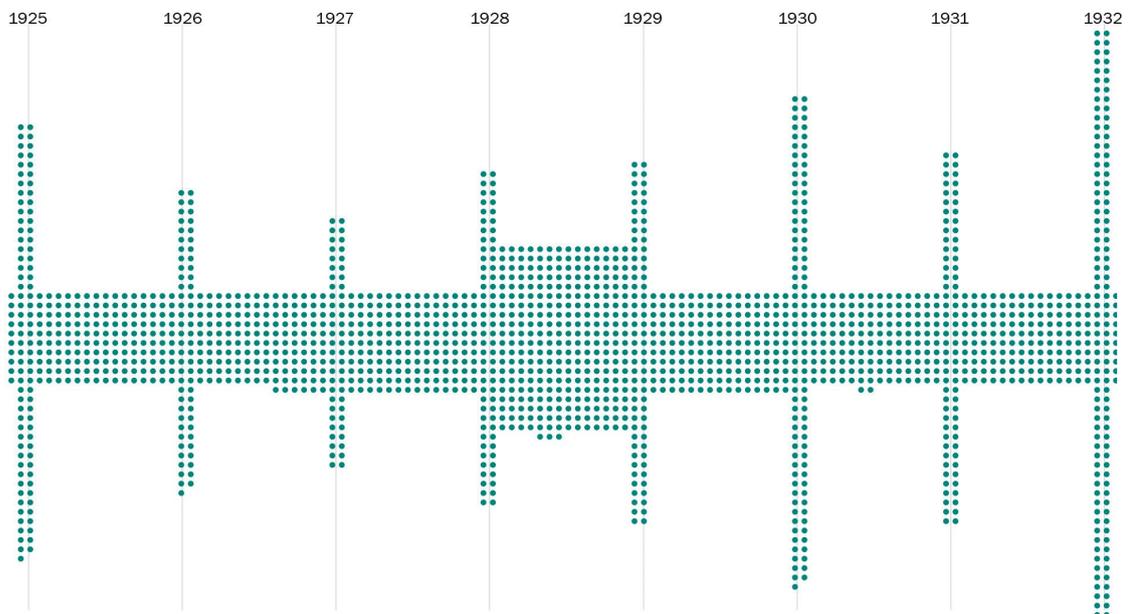
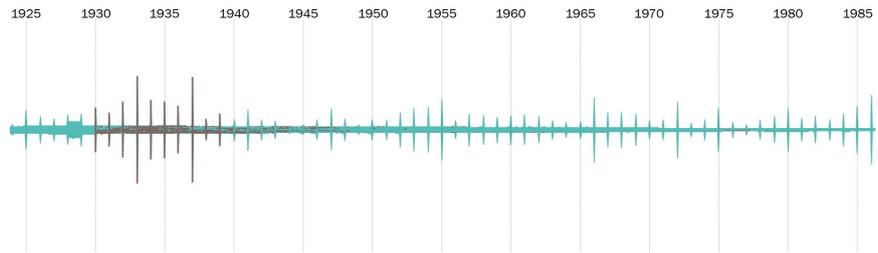


Figure 6.9 – Records with precise dates (exact years) appear in vertical spikes, while uncertain items spread out between them.

Figure 6.10 – Records that correspond to the 1930s are highlighted in Cooper-Hewitt’s object dataset. The assigned dates of these items reach as far as the 1980s.



The lens-shaped distribution representing Oxford’s Beazley archive (Figure 6.7) displays a staggered, step-like distribution. This indicates that the dates have been assigned in regular intervals of 50 years each. Indeed a numeric analysis of the data reveals that 93% of the records in the collection have a date bracket that is precisely half a century wide. Almost 7% are dated within a century, leaving less than 1% of individually dated items. Incidentally, seven records have a date bracket of negative 50 years, likely to be result of confusing CE and BCE dates. This regularity in the dating of the records that partly originates from Sir John Beazley himself has not previously been recognised.

The detailed breakdown of the dating intervals required a numeric analysis to be performed ‘outside’ the timeline – the visualisation ‘only’ provided the initial stimulus that there might be a pattern that would be interesting to study. However, the visualisation made the pattern immediately apparent and, crucially, without a researcher having to explicitly look for it. Using the visualisation one could then decide to follow up on the insights from the numeric analysis and, for example, discover if there is anything distinct about the 1% of items that do not follow the dating pattern.

Besides the usual format of a production date stored as textual date and date bracket, objects in the Cooper Hewitt database carry an additional ‘decade’ field. It is not unlikely that such a field contains redundant data effectively derived from the numeric date field in order to make it easier to navigate the dataset by decade. However in this particular case it appears to serve a different purpose. Highlighting, for example, records that match the decade 1930 indeed dyes most of the entries between 1930 and 1940 (Figure 6.10) – although not all of them, and neither is the selection fully confined by these dates, extending instead as far as the 1980s. How come these are labelled as 1930s?

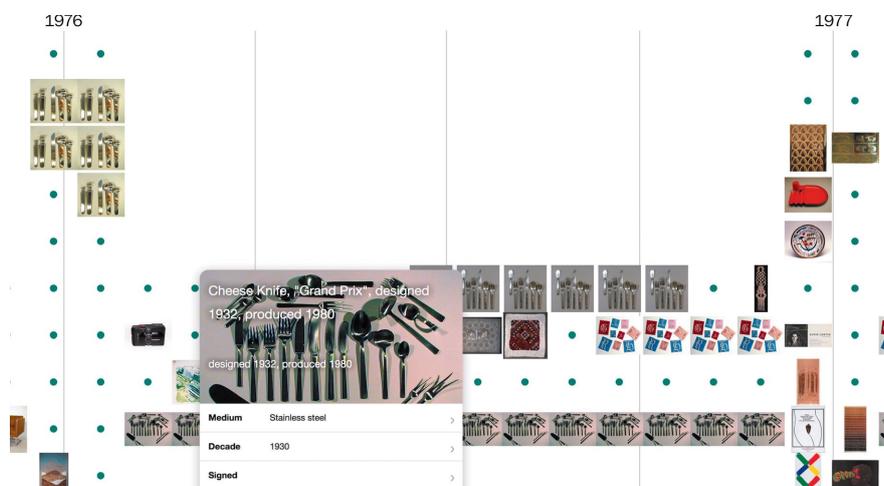
Probing the records by pulling up their details provides an explanation. A horizontal row of dots in the 1970s refers to a set of cutlery. Each record carries the same image and represents an individual piece of cutlery; dessert spoon, cheese knife, salad fork, etc. Another occurrence of the “Turner problem”: a set of matching cutlery is treated as a

- 5 Internally, cataloguing software is able to treat records such as these both as single items as well as sets of items. However, such relationships are usually not maintained when a digital collection is published online.

Figure 6.11 – Curators at Cooper-Hewitt refrained from deciding between two relevant dates (date of design and date of production) for this set of cutlery. Instead, they assigned the timespan 1932–1980, a decision which the visualisation revealed. Later, the textual date description was updated, giving preference to the date of production.

collection of individual objects.⁵ Yet, we are still interested in learning why a 1930s item appears so late in the timeline.

The textual date description offers a clue: “designed 1932, produced 1980” (Figure 6.11). Numerically, the date is defined as 1932-1980; the width of this date bracket caused the records in the diagram to spread out between those dates. The ambiguous date of this item is an example of the kind of problem in relation to dating objects a curator (C3) alluded to. Conventions expect the specification of a single date, forcing a curator to make a compromise when faced with many relevant dates; such as in this case, date of design and date of production. The visualisation is able to respond to this dilemma and offers a glimpse of



a curator’s reasoning behind the dating of objects. Incidentally, it appears that the dispute between these two dates has been settled in the meantime. When the dataset was released in November 2013, the date on this item read “designed 1932, produced 1980”. An update on 12 June 2014 changed the description to “1980”, while leaving the numeric dates at 1932-1980.⁶

- 6 Following an edit on 12 June 2014, the item’s date is updated. The edit history is accessible here: <https://github.com/cooperhewitt/collection/blame/master/objects/187/066/33/18706633.json> (accessed 12.01.2016)

- 7 Visible by an increase in yearly spikes and a disappearance of the stretched-out band in the decade shapes.

I will refrain from exploring all the histories behind the records that appear seemingly outside their assigned decade; we now have a sense of how this would be done and what kind of insights it could reveal. Coming back to the concept of distant reading in relation to the strategies for dating cultural objects: this can be done here, for example, by separating the dataset according to the decade field, as visible in Figure 6.12. We can see how the precision with which items are dated increases in more recent decades;⁷ a trend that is to be expected. However, there is a difference in how far the decades stretch into the future; an indication of how ‘popular’ designs of these periods were, measured by how long they were still in production. Both the 1900s and the 1920s extend as far as five decades outside of their scope, while others typically do not reach out more than two.

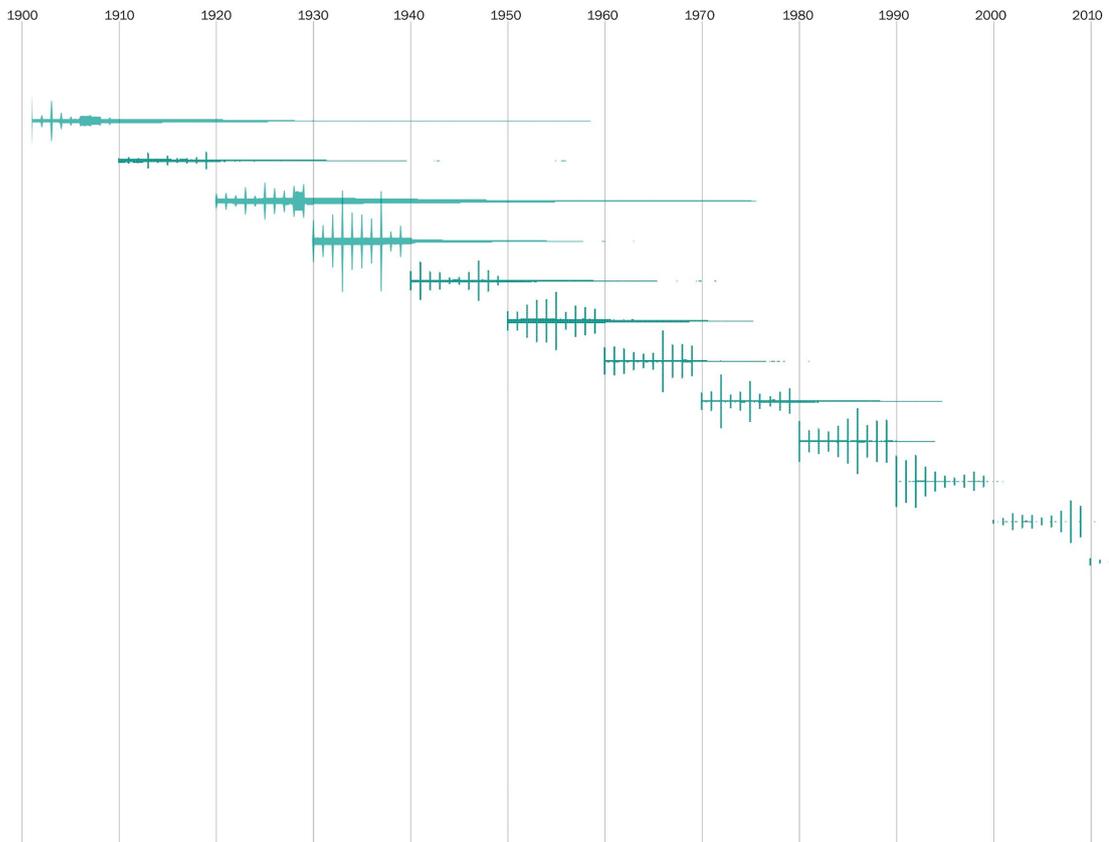
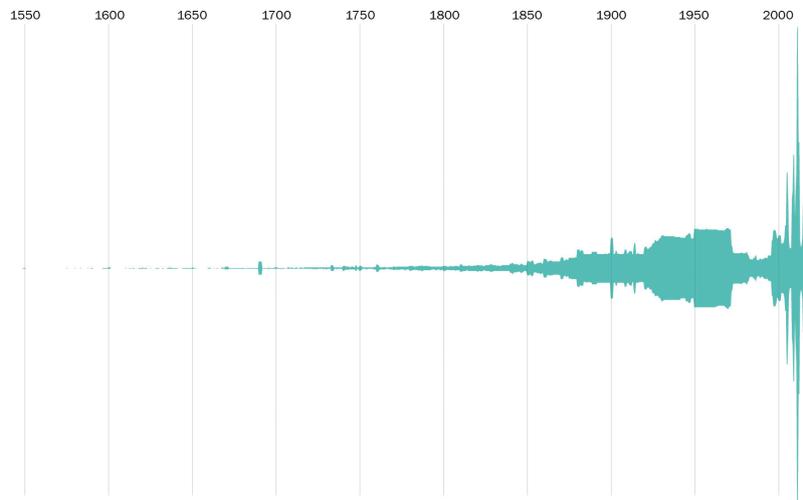


Figure 6.12 – Separating the object records in Cooper-Hewitt’s digital collection by their associated decade shows for how long items of a particular time period remained in production.

The Structure of a Collection

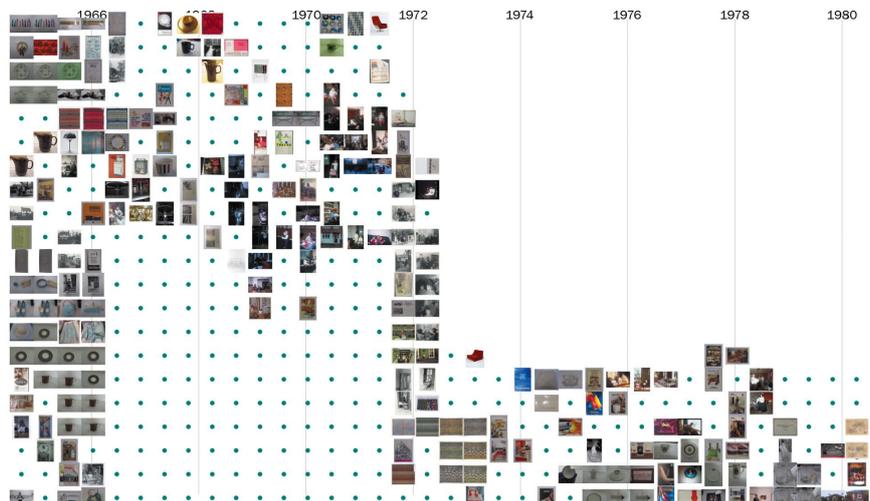
Through the visualisation it is possible to see the temporal scope, size, distribution and composition of a collection. The latter often requires a bit of probing. Figure 6.13 visualises the Geffrye dataset by date of

Figure 6.13 – The dataset of the Geffrye Museum reflects changes in the focus of its collection. These are made visible in this diagram by the distinct variations in its shape.



production. From the beginnings in the 16th century until the 19th century the shape of the diagram widens continuously. At the right end towards the present time, the shape and structure undergoes a series of abrupt changes. After a steady growth around 1925, a block appears in 1950, which then quickly narrows and morphs into a spiked cluster that remains until the present time.

Figure 6.14 – A large cluster of missing images is responsible for an abrupt change in Geffrye's collection. These could indicate an acquisition that is not yet catalogued completely.



Spikes, as we learned, refer to items that have been precisely dated, while continuous patches contain records with ambiguous dates. When a collection contains images, it is often immediately informative to examine them by zooming in. Around 1970, the images display objects, photographs, drawings – a heterogenous composition of items before and after the sudden change in size. However, a set of patches with missing images is visible, which is responsible for the pattern

we saw (Figure 6.14). They all appear to be photographs, uncatalogued and dated as 1950-1970 - likely the traces of a large set of items that has been acquired, but never completely catalogued.

The spiked cluster at the end of the diagram does represent a change in the composition of the collection. While earlier items relate to objects or drawings, these records represent photographs of homes, largely part of their Documenting Homes collection.⁸ Most of them carry a timestamp and are therefore dated with greater precision of up to a day. We can also see how present-day physical objects are only rarely collected anymore - instead the more recent items are almost exclusively made up of photographs and magazines. This shift in the collection's ontology represents a change from collecting to documenting - traceable in Figure 6.15 which highlights the records of the Documenting Homes project in blue.

- 8 Geffrey's Documenting Homes collection is an archive that captures twentieth and twenty-first century living environments. See <http://www.geffrye-museum.org.uk/collections/explore-our-collections/advanced-search/?groupId=PRO24406> (accessed 21.01.2016)

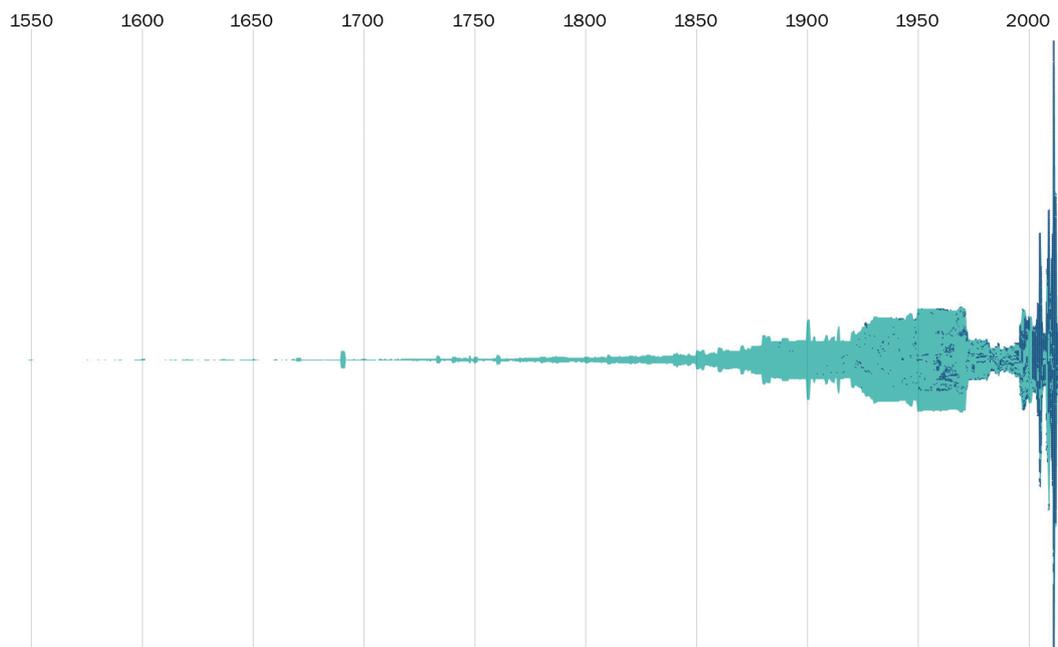


Figure 6.15 – A spiky cluster at the right end of the diagram represents recent items in Geffrye's collection that correspond mostly to their Documenting Homes project.

Mediated Insights – Perspectives

The interest of the curators of the Britten-Pears foundation in the temporal relationship of authors and compositions formed the basis on which I developed a visualisation format that emphasises the multiple temporalities of cultural artefacts and collections data, and makes them available to visual analysis. While previous work on visualising temporal relationships focusses on tracing the history of individual events, the Temporal Perspectives layout enables patterns to be studied both up close as well as across a complete dataset; close and distant reading of temporal relationships each produce their own insights. Furthermore, the layout encourages the basic unit of a cultural dataset, i.e. the object, to be reconsidered, multiplying the possible viewpoints on a dataset and its potential as a resource for new knowledge.

Explicit and Implicit Relationships

SemTime (Jensen 2003) relies on the relationships between events being encoded in the dataset: the type of connection is indicated in the visualisation, which means it must have previously been defined. Continuum (André, Wilson, Russell, et al. 2007) similarly requires links to other entities being defined in the data type as separate nodes. Both visualisation tools therefore rely on explicitly described relationships.

None of this is unusual, in fact Linked Data (Bizer et al. 2009) formats promote the idea that datasets should be defined exclusively as sets of relationships - in contrast to seeing records as disconnected, individual entities, like a row in a table. Linked Open Data is especially suitable for storing and retrieving cultural data as it does away with many problems related to matching data to a predefined structure (europeana labs n.d.).

Many cultural datasets have grown from database formats that are built around more 'traditional' data structures; they are therefore not typically available as Linked Data. At the centre is the entity that usually describes an object, along with its different attributes; material, location, maker, etc. In Linked Data terms this would translate to a set of tuples: object "made from" material, object "originates from" location, object "produced by" maker, etc. Linked Data therefore provides explicit information on the type of link and allows for finer grained descriptions. For example a location could be specified as "found in", "made in", "used in" etc.

Although tabular formats often lack explicit link descriptions, they can nevertheless be treated as sets of relationships by reconsidering what constitutes the basic unit of a cultural dataset. Britten Poets

maps the relationship between Britten's works and the lives of individual authors essentially by putting a different attribute at the centre of the data record than would be expected; instead of comprising a collection of works by Benjamin Britten, the lower part of the visualisation represents a collection of authors, whose commonality is that they all have been set to music by this particular composer.

Relationships do not need to be explicitly expressed in a dataset to be visualised in a timeline, instead we also can reconsider what constitutes an event. By visualising different temporalities of a record, different events are being mapped and their relationships can be studied. This is the mechanism behind the Temporal Perspectives layout; a conceptual, rather than just a technical innovation.

Distant Reading

Figure 6.16 is a visualisation of a part of the V&A database as a collection of types of objects, of artists, of places and a collection of collections. Each shift of perspective results in a different temporal distribution and the connections between them plots the relationships. Their order matters for the distant reading of connections as the links are only drawn from one row to the next. This particular arrangement displays the temporal extent and distribution of object types in terms of contributing artists, how these artists relate to places and whether they have lived there at similar times, and the diversity of collections in terms of the geographic origins of the containing objects. We see, for example, the temporal focus and spread of individual collections with regards to the items they contain. Two outliers appear around 1700; the Sculpture Collection and the South and South East Asian Collection both contain objects distributed evenly across the entire timeframe, while other collections typically cover a smaller time-period or exhibit more asymmetrical distributions.

In Figure 6.17 the entire database of MoDA is visualised as categories, objects and procedures. The term 'procedure' is used by the cataloguing software to denote exhibitions, publications or loans. Objects and categories are both arranged by production dates and we can see, by the width of the clusters of connections, what timeframe objects in a certain category typically occupy. Procedures are arranged by their date of creation and occupy a shorter, more recent timeframe. By examining the size and distribution of the dots in the procedures row, we are able to identify patterns in the usage of the objects in the collection. A large number of objects has been exhibited as part of four major and one smaller exhibition throughout the 1980s.

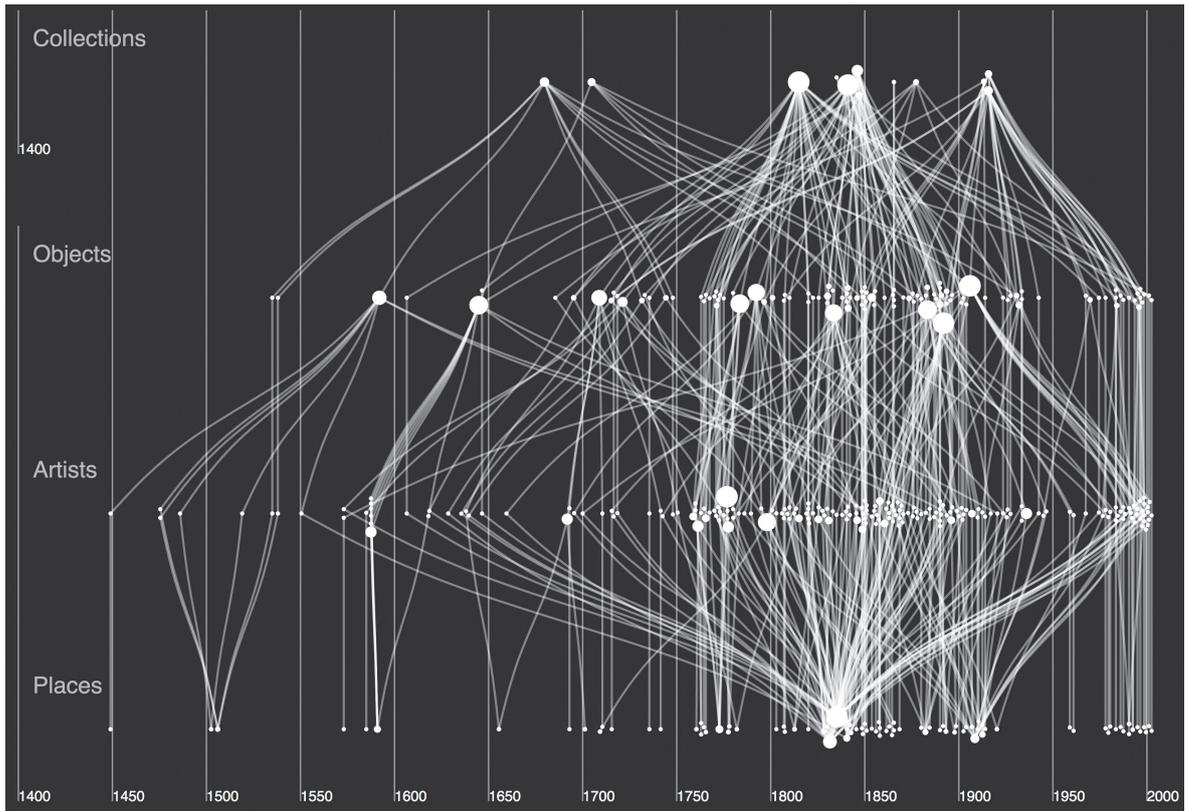


Figure 6.16 – A Temporal Perspectives view of a part of the V&A dataset. We are looking at the dataset as sets of collections, objects, artists and places.

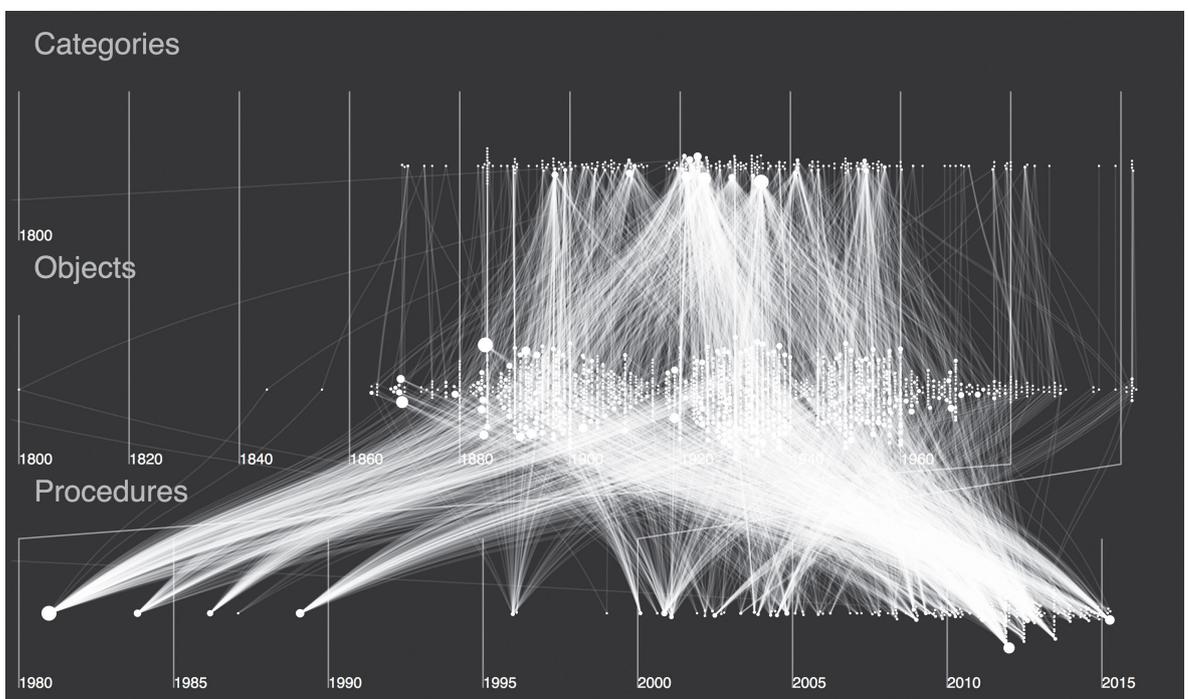
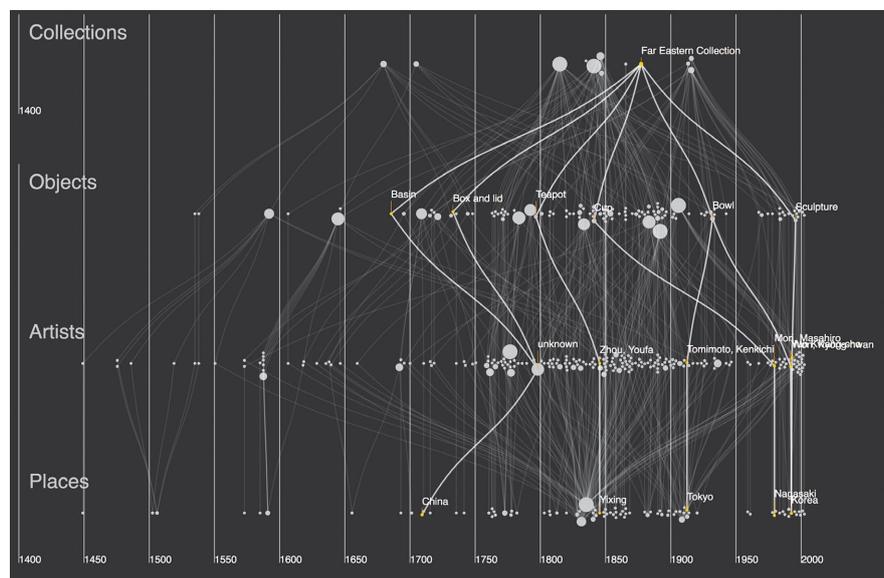


Figure 6.17 – This dataset by MoDA contains object records along with associated “procedures”: exhibitions, publications, lectures, etc. In this Temporal Perspective view, we can examine patterns in usage of their collection.

In the 1990s, there is only little documented use of the collection. The 2000s display a steady growth in usage of objects as parts of publications, theses and exhibitions. Again some larger dots represent major exhibitions, but none of them as expansive as the first documented show in the 1980s.

I have already discussed some of the patterns visible in the Britten dataset; his favourite time periods in terms of writers and the temporal ‘spread’ of writers in Britten’s life that the visualisation reveals. Similarly, the vertical ‘rain’ of compositions and immediate first performances, followed by a sudden backtrack to Britten’s childhood pieces in P8 (Figure 4.21 on page 166) is an example of how the visualisation affords distant reading of relationships. These patterns are only apparent when they are visualised in context of relationships in the entire dataset; anomalies, such as lines reaching diagonally to Britten’s juvenilia stand out against the ‘normal’ case of vertical lines. The timeline format allows for both close and distant reading of datasets, following my described design principles of distant reading and multiple temporalities, and offers a way of looking at temporal patterns in datasets that is not available in current visualisation tools.

Figure 6.18 – By highlighting dots in this view of the V&A dataset, we can study how individual items relate to other aspects of the collection.



Close Reading

By highlighting a dot, it is possible to study how a ‘virtual record’ connects to others and how it occupies different timeframes according to the perspective from which it is seen. For example, we can examine the V&A’s Far Eastern Collection, highlighted in the top perspective in Figure 6.18. The selection propagates downwards and connects related items. Hereby it is possible to identify if relationships are reciprocal. For example, “Basin” in the objects timeline is highlighted in yellow,

which denotes a reciprocal link - all represented basins are part of the Far Eastern Collection. The link to “Teapot” on the other hand is labelled, but the dot remains grey - teapots can also be found in other collections.

We can study the MoDA dataset in the same way. Figures 6.19–6.21 depict the visualisation in use. One of the three larger items appearing in the recent history of the procedures row are highlighted in each screenshot. Each results in a different distributions of objects that are associated with a particular procedure. In Figure 6.19 the procedure “Robert Jones’ Biography of Lewis Jones” is highlighted and results in a condensed ray of lines aiming at a narrow portion of the objects timeline, dated around 1920-1950. This corresponds to the time Lewis Jones (1894-1953) has worked for The Silver Studio⁹ - with one outlier, probably an item with an incorrect production date.¹⁰ The objects in Figure 6.20 are distributed across the entire collection. These correspond to “Objects Featured on the Blog”, a selection likely intended to offer a balanced overview of the collection. Lastly, in Figure 6.21 the highlighted objects are associated with an art history course taught at the university. Here we can clearly identify two temporal clusters of objects, denoting the temporal focus the course has chosen, along with which time periods and objects are omitted.

Visualising procedures and museum objects in this way enables an institution to study which parts of their collection gets used in what types of contexts. Such insights are of great value for museums, not least financially, as they often need to show funders if and how their collections are being used and what impact certain investments have had on the usage of archives.¹¹ On a less fiscal, but no less valuable level, curators are able to experience how their holdings lead to new insights:

How is knowledge being generated? How is new knowledge being generated with the same stuff? You’re generating new knowledge by visualising it, but then that is visualising the fact that new knowledge is being generated. (C12)

In the example of Britten’s Poets, studying the relationships between authors and works enables curators to discover new connections and examine patterns they were already aware of in more detail. I discussed Alfred Lord Tennyson’s appearance in Britten’s oeuvre with the curators ahead of creating the visualisation. His writing appears several times throughout Britten’s life, set as individual works or as part of a work cycle in combination with other authors. By highlighting Tennyson’s dot in the bottom perspective - as visible in Figure 6.22 - we can see precisely how his writings are distributed in Brit-

9 “After it closed, the contents of the Silver Studio - artwork, record books, photographs, correspondence and so on - were given to Hornsey College of Art, which became part of what is now Middlesex University. This material is the focal collection of the Museum of Domestic Design & Architecture (MoDA).” (MoDA 2008)

10 According to the collection’s curator.

11 For example, a museum may justify to devoting funds to the conservation of artefacts when they are able to show that there is (public) interest in them.

ten’s works and - by the size of the dots - which works contain several author’s writings. Subsequently selecting these works we can examine which poets appear together with Tennyson and how they are distributed through time (Figure 6.23).

Figure 6.19 – A publication on Lewis Jones – highlighted in the bottom perspective – features objects from MoDA’s collection. Their production dates correspond to the time Jones was working at the Silver Studio.

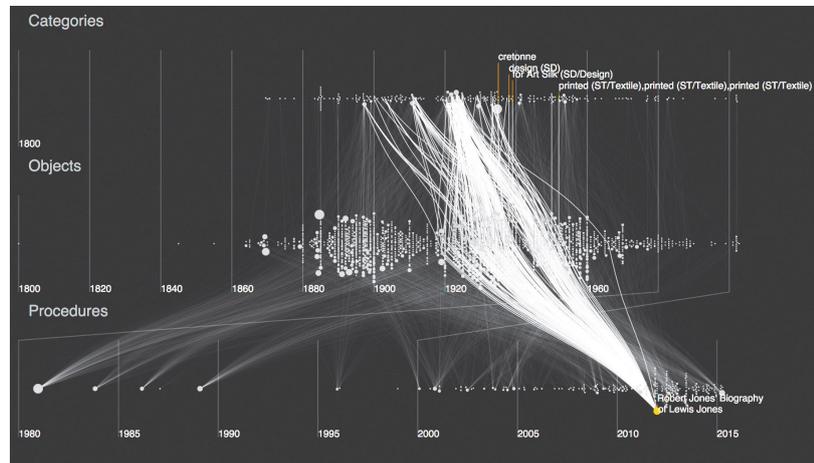


Figure 6.20 – An internal sub-collection keeps track of the objects MoDA featured on their blog. These are distributed evenly across the entire timeframe the objects in the collection occupy.

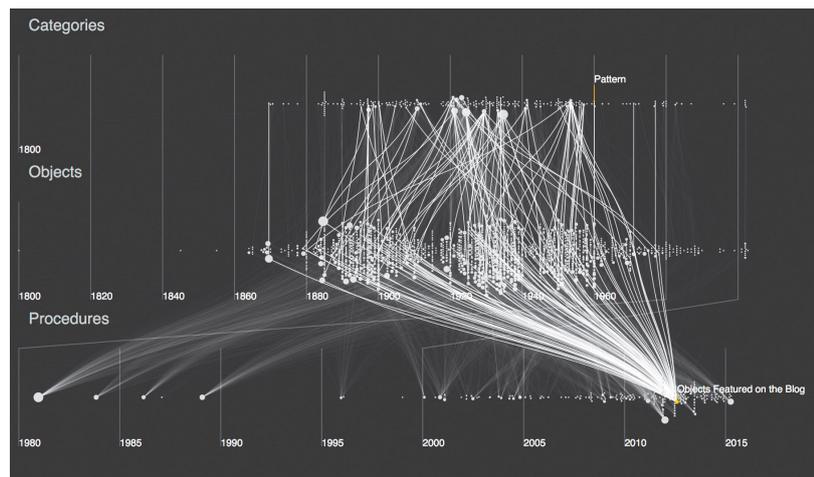
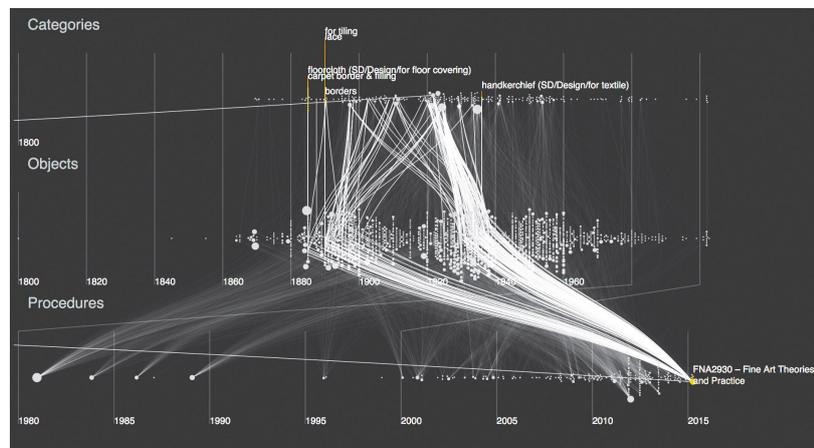


Figure 6.21 – The highlighted procedure corresponds to a university art history course, which apparently focuses on two distinct timeframes in MoDA’s objects collection.



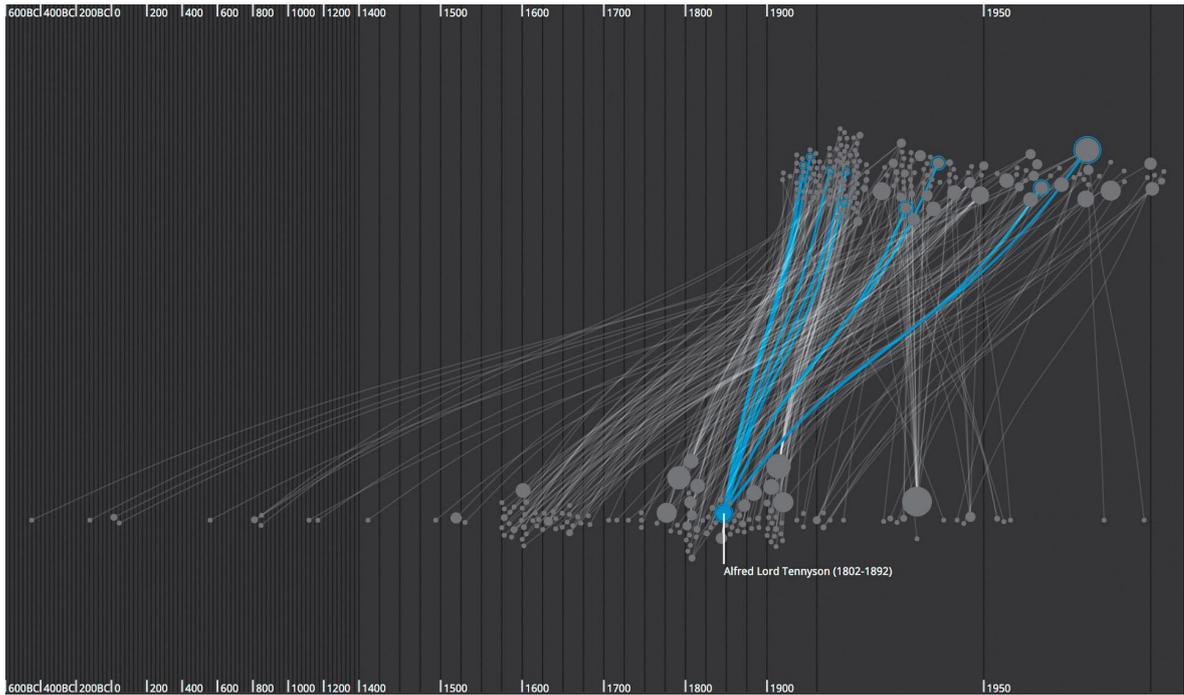


Figure 6.22 – Selecting a poet highlights the corresponding works in Britten's oeuvre, allowing researchers to study Britten's interest in specific writers throughout his working life.

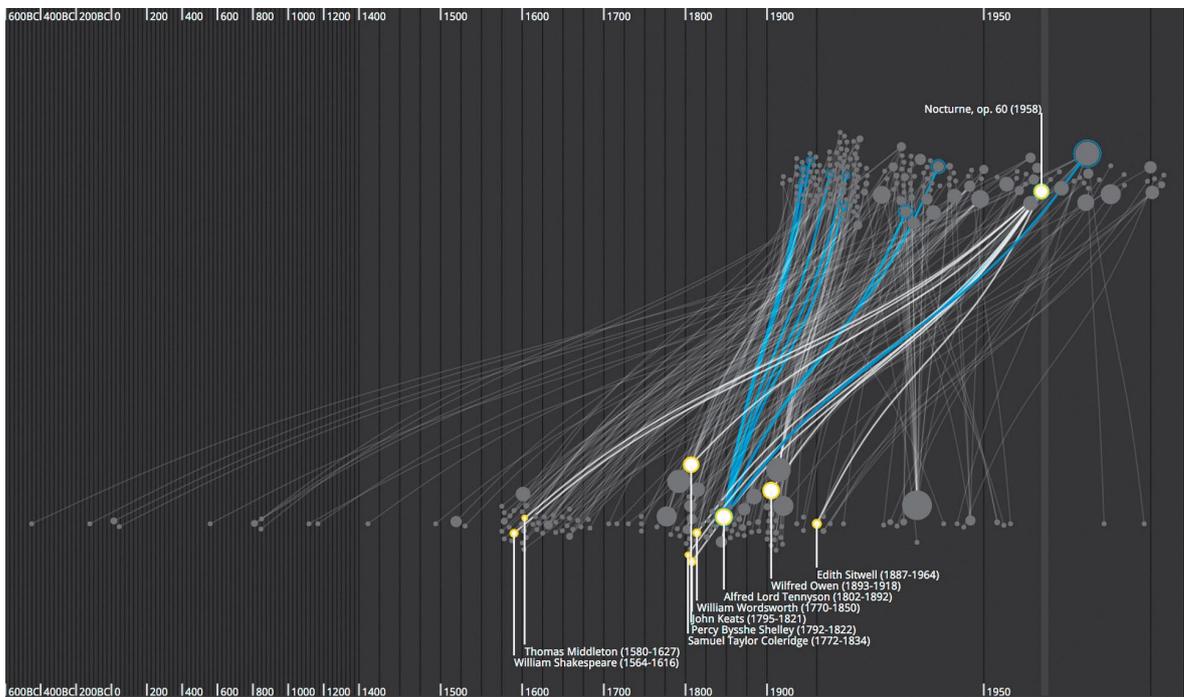


Figure 6.23 – It is possible to study how individual authors share a relationships through Britten's works and examine which writers appear together in a work cycle.

One piece stands out when we look at its temporal distribution in terms of poets. *Voices for Today* is a singular example, spanning the widest timeframe of all of Britten's work (Figure 6.24). This piece includes both the earliest poet Britten set – Sophocles (496–406 BCE) – as well as the latest – the Russian poet Yevgeniy Yevtushenko (born 1932). A novel insight for the curators, and one for which they could provide an explanation. *Voices for Today* was a commission for the twentieth anniversary of the United Nations and the poets Britten set were suggested by his advisors. It is quite possible that this is the reason why poets from Britten's favourite time period are joined by 'outsiders' and that his advisors were keen to include a selection of poets spanning a wide frame of history and geography.

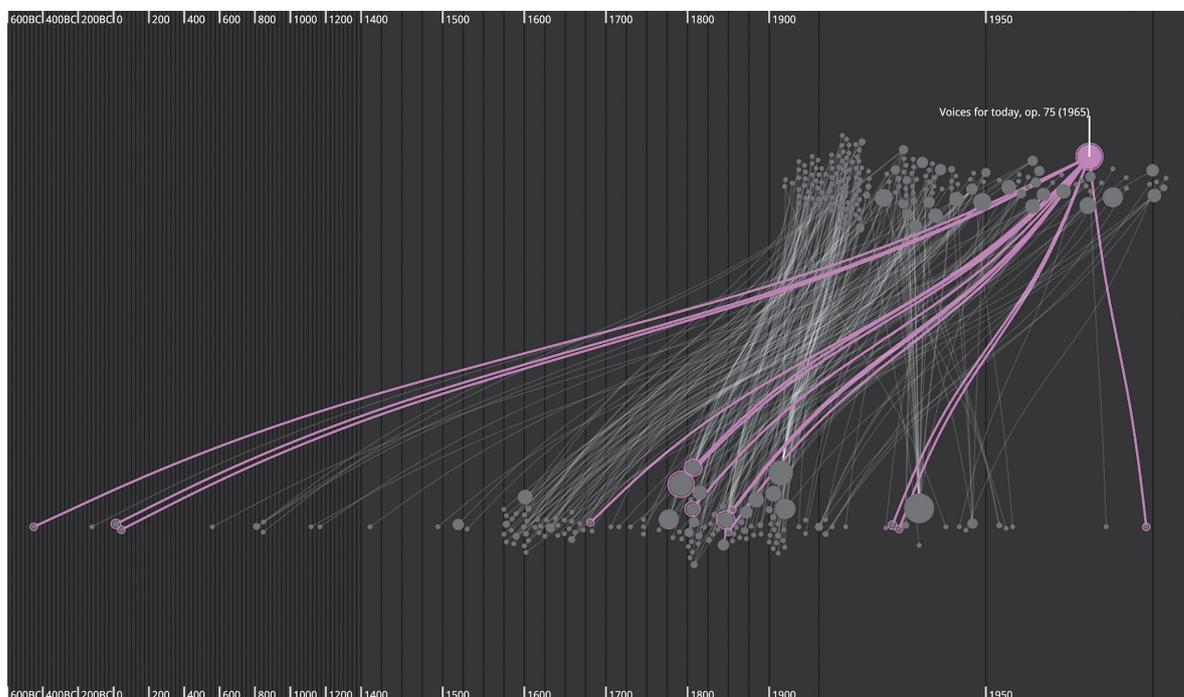


Figure 6.24 – *Voices for Today* stands out in the timeline due to the wide timeframe it occupies. Apparently, this is not a coincidence – the writings have been selected by Britten's advisors.

Different Emphasis Through Different Timescales

In earlier iterations of the Britten's Poets prototype I experimented with time axes whose scale can be manipulated in various ways in order to find the 'ideal' combination of timescales for the parallel timeline format.

Figure 6.25 is a screenshot of an earlier version with two parallel time axes. On the upper timeline the poets are represented in a unit histogram and linked to the individual works, which are represented on the lower axes. Both axes initially use a linear scale. By clicking and dragging anywhere on the axis, users can set break points and stretch or compress certain timeframes.

The prototype pictured in Figure 6.26 arranges the data vertically in two columns; on the left are Britten's works and on the right the corresponding authors. In this example the axes are scaled according to a Fisheye transformation (Furnas 1999): a dynamic lens which lets users expand and compress timeframes by moving the mouse vertically over either column.

Like Priestley I expected the non-linear scaling of the axes to be confusing to the viewer. However it did not seem to pose any problems for the individuals I tested the prototypes with. It seems that because the user was in control of the transformations, the scaling of the axes became itself a useful tool for filtering the data and for understanding the relationships between events across time.

An important insight is how a user's ability to explore and understand a dataset through a timeline visualisation is not diminished when a timeline does not use a linear time axis. In fact it became evident how the scaling of a time axis, including the familiar linear scale, is an arbitrary choice - one of many possibilities - and plays an important role in determining what aspects and patterns in a dataset get emphasised.

The non-linear scale in my Britten's Poets visualisation - which by default was the same for both work cycles and authors - clearly reveals Britten's preferred literary periods and the patterns of compositions in his own working life. By switching to a linear scale, detail in the composition's temporal distribution is lost, but a more accurate impression of the poet's distribution in time is gained. The temporal distance between the 18th century and the ancient Greek and Chinese writers is evident.

Last, but not least, scaling each timeline independently from the other reveals patterns that were hidden in the previous views. Three connections from works to authors stand out as their angle deviates from the majority of lines. Around 1935 Britten uses, for the first time in his working life, the writings of authors that date back by more than a millennium: notably the writings of the Chinese poet Po-Chü-i (772-846), who appears again later in Britten's life (Figure 6.27).

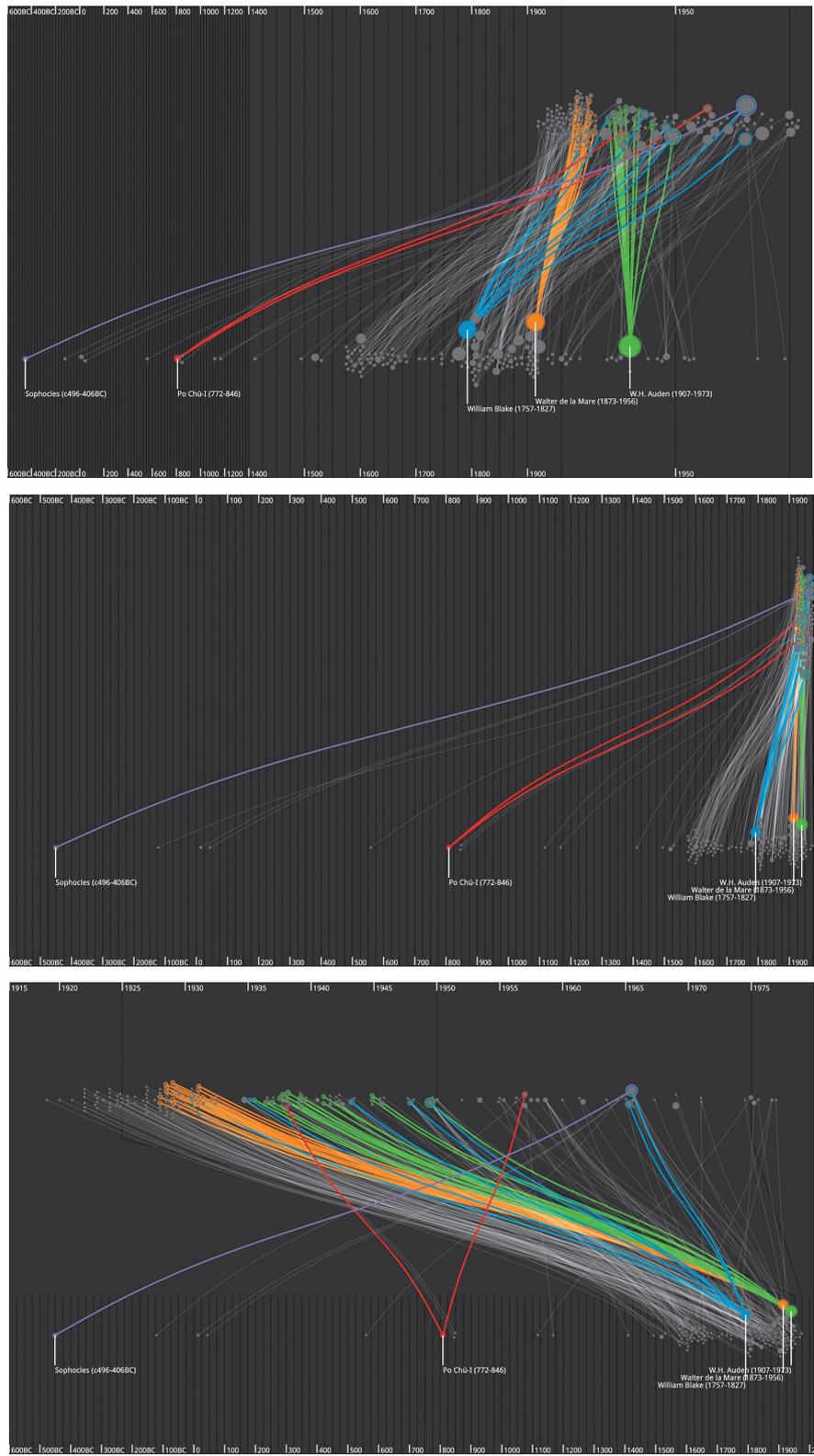


Figure 6.27 – The final version of Britten's Poets implements three possible time scales. Each emphasises a different aspect of the temporal relationships.

Expert Insights

In addition to the functional aspects of the visualisation prototypes that facilitate the discovery of new knowledge with regards to the issues mentioned earlier, the following section presents the insights that curators were able to gain from interacting with the visualisation prototypes and from participating in the design process. These range from practical uses with regards to the identification and interpretation of ‘suspicious’ data, to affective qualities and the ability of visualisations to encode and release a curator’s tacit knowledge.

Design Decisions and Collaborative Design

When I discussed the prototypes with curators and expert users, our conversations revolved not only around what the visualisations reveal – how they could be interpreted – but also around the design of the tools: how the visualisations could be improved, what works and what does not. Hardly did the comments touch on the prototypes’ appearances, their ‘aesthetic’ qualities. Comments were aimed at allowing tools to answer specific questions or enabling certain modes of enquiry: suggestions for the design of the prototype were intrinsically linked to insights, or possible insights, that they enable. I will discuss these below. Notably, sometimes it was also just the idea of visualising data without actually visualising it that triggered new strands of thoughts. After having gained experience in the possibilities of visualisation and having seen examples of visualised collections, curators started to speculate, which artefacts might appear how and what that might tell them. We can see this as evidence of the different ways that a design process produces new knowledge; it can be generated by using timeline visualisation tools, but also by reflecting on how visualisations should be designed.

When comments on the prototypes’ appearances were made, they were largely positive. Curators found the prototypes to be “pretty”, “beautiful”, even “tasty”. One scholar stated that the TT prototype looks unfinished.¹² This sparked a discussion among those present on whether or not this is necessarily a problem. One participant argued that the fact that the visualisation does not look overly polished invites conversation and criticism and suggests that the tool itself is open to be adapted – key properties for a scholarly research method.

¹² This point was raised, not by a curator, but by one of three professors of design who were present at a seminar where I demonstrated my work.

Critical Thinking, Interpretation and Explanation

Conversations revolved around how to represent the characteristics of collections data I discussed, specifically with respects to temporal uncertainty. Incorporating uncertainty in the Temporal Jittering layout

in the way I have done corresponds to how curators intuitively deal with uncertain dates. Similarly, positioning single dots by average date in the Temporal Perspectives prototypes, rather than representing beginning and end dates, was appreciated.

That's sort of intuitive [...] It's actually the years in which they were largely working and obviously that's the crucial point. We don't really care, in that sense, when they were born. Because they didn't start writing poetry the moment they were born. (C6)

By making uncertainty in dates an essential part of the visualisation, curators increasingly learned to be critical about the dataset; in respect to digital dates as well as other aspects of the data. This is a transition I noticed in comparison to when I began my research and, lacking at that time my own visualisation tools, would often produce timeline visualisations using existing software that had been designed for numerical data analysis. When interpreting these generated graphs, curators were quick in drawing conclusions from patterns that later turned out to be inconsistencies in the visualisation format, but were regarded as being representative of the collection. These false conclusions could be debunked when the same datasets were visualised in the TT prototype that employs dating uncertainty and represents records individually, not as summarised charts. While this showed me that my own prototypes indeed presented an improvement over existing solutions when it comes to visualising cultural data, it also made me aware how quickly users will try to make sense of what they see.

The term “intuitive” was used repeatedly by the curators to describe certain aspects of the visualisation tools, however sometimes only after we discussed the rationale behind them. Non-standard visual representations often required explanation:

It seems to me that you also need a sub-strata of textual information. (C11)

Without explanation one might not understand, or even consciously think about, the meaning of some graphical representations. An example of this is the vertical positioning of the dots in the TJ layout:

The height is a bit artificial, isn't it? That does not really mean anything. (C4)

After I explained that the vertical ordering corresponds to the records' sequence in the dataset and should match how the archive was originally catalogued, patterns were becoming evident:

Oh, ok and that's probably right, because that's from one particular publication that someone must have been going through. So if you look at those, these must quite possibly be from the same publication. There are definitely groups. (C4)

Questioning and discussing the thinking behind a visualisation can be essential for drawing reliable conclusions from it. Researchers and the public have largely grown accustomed to the idea that data does not speak for itself (Behn 2009). The same, however, is true for data visualisation. Wherever diagrams depart from established paradigms, an explanation becomes necessary:¹³ Having to explain a visualisation does however not mean admitting defeat; believing that a design has failed in being self-explanatory or intuitive.¹⁴ To explain provides an opportunity to reflect on design choices and gain insights.

13 In the *Description* (1764) that accompanies Priestley's chart, he devotes several pages to explaining the rationale behind representing time as a line.

14 Companies such as Apple promote the idea of user interfaces having to be "intuitive" (Stern 2014). Steve Jobs himself expressed that "the main thing in our design is that we have to make things intuitively obvious" (Isaacson 2012). In the academic HCI literature however, the term is less convincing because "many claims of intuitiveness, when examined, fail." (Raskin 1994)

Enjoyment

Matters of affect and pleasure when using the visualisation tools have also appeared as an important factor. That working with visual timelines is more enjoyable than studying text-based chronologies is a claim that has already been made in the 18th century (Boyd Davis & Kräutli 2014). Features of the digital medium enhanced the visualisation's affective qualities through animation and motion. Zooming in on the Tate timeline, for example, and making the images appear often elicited from spectators a sound of delight. I did not expect this when I demonstrated the prototype for the first time at a conference. Later I could predict the reaction and provoked the dramatic effect of the images appearing by zooming in on the right portion with the right speed. Including images, where available, is however more than just a visual effect:

I think the other really distinctive thing about this is that ability to see the thumbnails when you go in. You could have a graph that's plotted from a spreadsheet that is generated with a statistical tool [...] I'm sure we have users that do that kind of thing for their research. But, being able to play around with it, to see, to get a sense of what kinds of objects fall in particular periods is much more intuitive and much more informative - immediately informative. (C7)

Seeing Everything

One of the most impressive effects on curators, when they looked at their dataset through the TT visualisation, was a sense of seeing everything. This did not apply to early visualisation prototypes and sketches I made using off-the-shelf software (Figure 6.28). These often required the datasets to be split up into categories, for example, in order for the visualisations to be comprehensive.

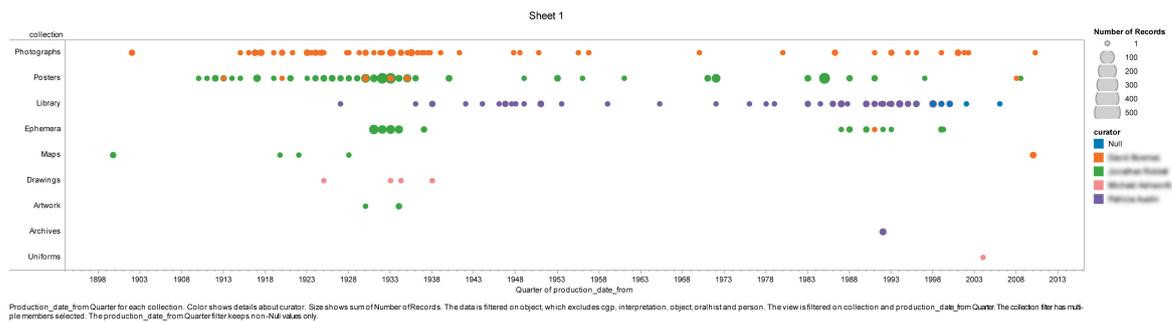


Figure 6.28 – A tryout visualisation of Johnston-related items in the London Transport Museum collection.

image: chart created in Tableau Desktop

By using the interactive TT visualisation and by being able to look at the images of the records, curators got a sense of seeing the complete collection. Although not all the dots and images are visible in the overview, it appears that the knowledge of being able to immediately query the visualisation in detail supported the notion of seeing everything, without actually seeing all of the images at once. This observation gives support to the theories of visualisations acting as “cognitive artefacts” (Norman 1991) that support “external cognition” (Scaife & Rogers 1996), as I discussed in chapter 2.

What I find really appealing [is this] ability to see the entire database in a sense. You literally see it all, even aggregated in that shape. And then you can zoom in on it. (C7)

Selecting and filtering could then be done within the visualisation:

The process of trying to find patterns is a human one. You provide a tool that enables a human being to play around and see what’s significant. (C7)

On a practical level, the ability to filter data through the visualisation interface was widely identified as one of the most useful features:

I don’t think we can filter down to that level of detail. That’s just really useful and it’s very clear what’s actually happening. (C6)

This behaviour, and how the visualisation affords it, is in line with Shneiderman’s visualisation mantra “overview first, details on demand” (Shneiderman 1996); the principle of starting out with an all-encompassing view and gradually exposing details.

Somewhat ironic is that the above statement on the “ability to see the entire database” was made while we were sitting inside the room that contained the entire archive; stored in folders, arranged on bookshelves that covered almost every wall of the room. By turning one’s head one could see the entire physical collection, yet it seemed more comprehensible and immediately accessible when the data was visualised on the screen. This corresponds to an earlier observation by Lee (2005) that a collection must be “readily available” – a digital interface

is able to facilitate that when it can visualise a collection on a single screen:

It's quite different from going page to page, or even a continuous scrolling type interface. (C4)

A digital collection may be examined as a whole in a way that, for various reasons, is not always possible with a physical archive or museum collections:

Often objects are missing for a while. Often you get some examples of what is in a particular travelling exhibition, but you don't get the whole range. That would be a very useful way of seeing things. (C11)

Critical Chronographics

Curators were, in general, content with the format of the timeline and using time as the framework for structuring the data:

Chronology wouldn't have to be the organising principle, but it's the most intuitive one. You could also have a blob organised to some other principle [...] but chronology is kind of obvious. (C7)

Analytic timelines also fill a gap, as existing tools that curators use to organise and clean their data often fall short when working with temporal data:

You have to sort an Excel spreadsheet in one way and chronology is the least easy to do. I'm sorting by store or by inventory number, and then I try to tell a story... [Dates are difficult] because the date fields are inconsistent and in different areas. (C8)

A chronological ordering matches with curators' interest in narratives and their work on telling stories with and about cultural artefacts. The other visualisation format that often came up as a suggestion is a geographic map. Many curators shared the desire to examine data according to its place of origin and some archives already produced their own, often Google Maps based visualisations of the 'locations' of objects.

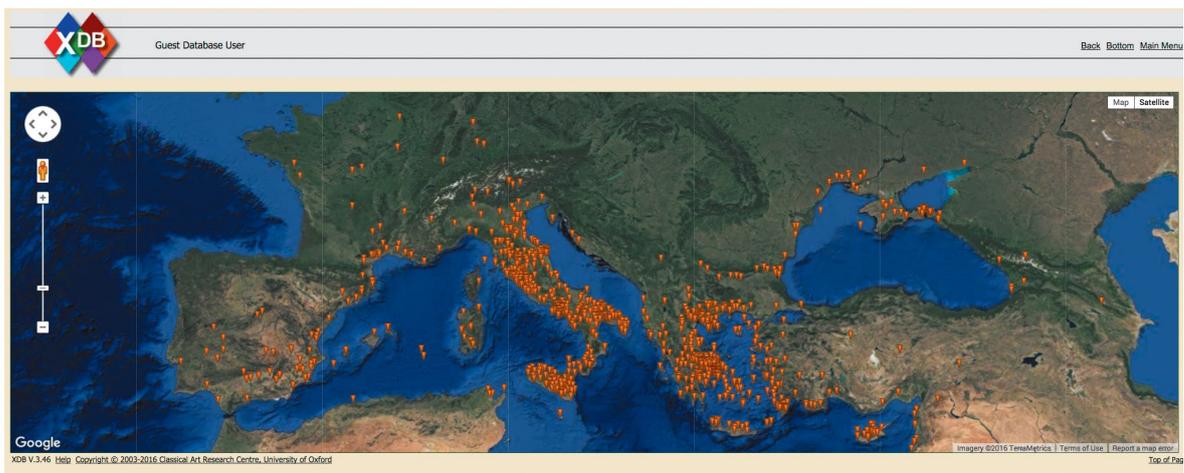


Figure 6.29 – Oxford's Beazley Archive plots database records on a map. However, each vase may have several relevant locations, only one of which is visualised.

image: screenshot <http://www.beazley.ox.ac.uk/xdb/ASP/mapXDB.asp> (accessed 09.12.2015)

Oxford Beazley Archive's online presence shows how the pots in their collections originate from all over Europe (Figure 6.29). Curators remind us, however, that the narrative behind this mapping is more complex than it appears:

Most of the ones we have on the database were made in Athens, but there's only a minority that was actually found there. Most of them come from Italy [...] The Etruscans, the ancient people of northern Italy, liked acquiring these Athenian vases and they put them in their tombs. And the kinds of tombs they built tended to protect the vases. (C7)

Geographic mapping revealed inconsistencies in location data; the fact that artefacts can hardly ever be associated with a single place and that a place may be relevant for a variety of reasons. As we have seen, the same applies to time and mapping events on a timeline.

Curators have typically not previously visualised their datasets in this way; arranging all events automatically on a visual timeline. Seeing their data mapped in time and participating in the design process of digital timeline visualisations caused them to make similar observations about the origins of dates and their accuracy:

These signs were around in the thirties, but we don't know when in the thirties. We're looking at photos to identify when they were around. There's nothing actually recorded [...]. I guess because of when the collection came together, which is quite a lot after these things were removed from their sites, it's difficult to know exactly their story; they are a bit vague. (C10)

In collaboration with the curators we were able to explore the cause of suspicious patterns in the datasets and identify their roots. For the London Transport Museum's upcoming centenary of Edward Johnston – the typeface designer who shaped the identity of London's public

transport - they asked me to analyse all the Johnston-related items in their database. In 1913 he was commissioned to design a lettering for London underground; the relation to Johnston of items that appear on the timeline before this date is therefore questionable. To me, the earliest item appearing on the timeline in 1868, an image of Westminster station with pre-Johnston typeface, looked like an error in the data. However, as the record represented not the photograph, but the actual station - dated by its date of opening - the issue was a bit more complex. To complicate things even further, a curator remarked that the actual photograph must have also been taken later than the timeline suggests: it depicted the station after its remodelling by architect Frank Pick in 1923, but still before the lettering was changed to Johnston's typeface.

This particular record represented the entire history of Westminster station, conflated in a single date. When represented on a timeline, it was evident that more must be hidden behind it. According to the record description, an early version of a sign designed by Johnston was found on the site in 1920 - which is why the record carries a relation to Johnston, albeit more than 50 years later than it appears on a timeline.

The Merits of Errors

Before I was handed over datasets of collections, curators often expressed concerns about them being imperfect and not organised the way they should be. When I presented them visualisations of other datasets I did earlier and the errors that emerged, they feared and admitted that their data will probably show similar mistakes, gaps and inconsistencies. Once they saw their data visualised, and indeed spotted numerous errors, they were however not that concerned about them anymore, because they recognised that the errors are, as a consequence of being visible, also fixable:

This is brilliant. This is extremely useful for proof reading isn't it.
(C6)

Seeing errors or missing images produced a call to action:

Just in itself, the visual representation of a sea of dots with very few thumbnails being utterly practical from a curatorial point of view. This immediately says, here is an issue we need to tackle. (C8)

Missing images are not only an issue for the quality of the data, they can introduce a bias in the usage of a collection:

The importance is that artefacts with images in their object record get used and produced for people more often than artefacts that don't have an image in their object record. (C8)

What became apparent is that there are hardly any binaries between clean data, and data that is messy, incomplete or erroneous. Curators often strived to present me with clean data. One of the main reasons I was sometimes not allowed to get my hands on datasets was because they were just in the process of cleaning them up. Once they were finished I would be able to work with them. Needless to say I hardly ever received the datasets; the process of cleaning up is never-ending.

A dissatisfaction with the condition of their current datasets is often keeping institutions from publishing them, along with a fear of presenting data that is 'wrong'. Data, however, can only be wrong if it is treated as facts - which arguably it often is, also in the context of cultural collections. A shelf number stored in a database should tell a user where a physical item is located. If the item is not where it is supposed to be, that piece of data is wrong. When looking at a cultural dataset on its own, however, the truth value of individual fields matters less; it may just be one dimension of it.

The observed outlier in the form of a record of Westminster station is wrong in some respects and correct in others. Supposed errors appearing in the visualisation cause curators to look closer and discover the hidden narrative, and reflecting on how this could be represented better in the data that drives the visualisation. Errors gave a sense of confidence to the curators. The fact that the visualisation is not perfect, that what the visualisation shows is not perfect, is not seen as problematic, but as honest:

This could be quite pure because what you're coming up with, in a way it's coming out of information that hasn't gone through someone's selection process to what is to be shown. (C11)

Revealing Biases and Interpretations

Representing cultural data honestly brings up all kinds of inconsistencies, ambiguities and irregularities that, in a visual timeline aimed at informing museum visitors, would normally be cleaned up or swept under the carpet. As I have argued, digital collections can not be "uninterpreted databases of raw cultural heritage material" (Lynch 2002). The interpretations that are necessarily present can however be revealed and made evident through appropriately designed timelines like the ones I presented:

Really fascinating what you've been doing with that in the way you present it by curator. You see where trends in collecting have taken place, the themes different curators were having in the development of the collection. (C1)

The traces individuals have left on a cultural collection can become tangible, one curator speculates:

It could be showing where an out-of-control curator has gone mad and acquired a lot of material. (C11)

The influence of their predecessors on the datasets is something which is very much relevant for the day-to-day practice of curators working with digital collections.

A lot of what we know about our collection is because the history has passed. Down the line we know [...] why we didn't catalogue a lot of those things because that curator was not interested in that. You could imagine in about fifty years time when that history is gone, that this [visualisation tool] is really interesting. (C1)

The same curator also suggested the use of the visualisation as an educational tool for new curators. New staff coming in will be unfamiliar with the collection and will lack the background knowledge required to properly interpret collections data. This applies however to any user outside an institution who wants to make use of a collection:

I urge you to continue this great work. Most of my life has just been a pain, trying to find things in archives and databases, where you have to spend a week trying to understand how they thought when they catalogued things. (C3)

With improved technical accessibility of online collections it is very likely that more users will struggle with identifying embedded interpretations of cultural data. Curators develop a tacit knowledge of how to interpret their own collections when working with them - knowledge that can be approximated, to an extent, with the presented visualisation tools:

What you're doing is, you're encoding some of that internal tacit craft - historical knowledge - you're blackboxing it into a series of tools that anyone can understand. (C3)

Meta-Questions

As my prototypes for visualising cultural data progressed from initially naïve timelines to the more sophisticated tools I arrived at in the end, they were more and more able to reveal biases that originated

from subjective preferences or the institutional history. Simultaneously, the kinds of questions that curators wanted to be able to answer with visualisation tools developed from low level to high level insights.

By 'low level' I do not refer to simplistic questions or mundane findings, but insights that stay close to the physical material the data relates to: the artefacts that are collected. Low level questions regard the data as closely related or even equivalent to the material. These questions do not necessarily require digital means in order for them to be answered. Higher level questions, or 'meta-questions', ask about the data, its history, and the circumstances that led to its creation. High level questions are removed from the artefacts and closer to the data. This also means taking advantage of the capabilities of digital technology and developing a greater awareness of its limitations.

Low level questions that were posed in the beginning concerned mainly aspects of the content of the collection:

I am interested in the genres [...] [visualising] all of his work and showing what type they were, how long they took to write. Having that represented visually would be wonderful. (C6)

Curators' questions related to ongoing discourses among the experts in their particular field, perhaps related to common beliefs and assumptions about the items in the collection or their creators, for which a visualisation might provide evidence, or maybe disproof.

Quite often curators stated that it would be necessary to gather more data before being able to answer the questions they really thought interesting. It appeared almost as if they either believed their archival data was insufficient - that the information it contains is not useful enough - or that they thought the possible insights would be trivial and not be able to exceed what they already knew about their collection.¹⁵ This relates to Simon's (2013) observation I mentioned earlier on the inability of available collections data to offer meaningful insights (see page 148). As experts, curators might have expected to know what the data entails and therefore the possibility of answering new questions through their datasets was often made dependent on more data to be gathered.

Spotting biases in collections data, both in their own as well as in other collections, caused the questions of the curators to become more self-reflective. While initially they were interested in finding out something about the data and the artefacts it represents, they were increasingly wondering what the data could tell them about their institution. Not the artefact and its creator - the artist - but the data and

15 This underlines the importance of a distinction between *collections* and *collections data*. Curators know a lot about the former, but less about the latter.

its creators – themselves, the institution and previous staff – entered the spotlight.

Engaging with cultural data instead of cultural artefacts reflects a shift in the way a humanist works. Instead of studying physical artefacts and finding out how their creator and the people around them must have lived and thought, curators now have to study their own traces and retrace the thinking of those before them.

Archaeologists working on a digging site will try to minimise the traces that they leave themselves. With digital data, researchers are currently much less careful, or less able to be careful about the consequences of their doing. Archaeological practices have perfected methods for minimising the influence of researchers on the excavated material. Digital collections have only been around for about fifty years and knowledge of the possible damaging effects of human ‘traces’ in digital data, as well as how to prevent them, still needs to be developed and passed on. Until then, visualisation tools that not only represent collections data, but also its history and biases, will be essential for the interpretation of cultural data.

To give an example of higher level, or meta-question, and how these progress; a reoccurring task curators wanted to perform is making comparisons:¹⁶

Could you do, for example, a comparison of the pots from the [museum collection] against all the pots in the archive. (C7)

This is a research question that already shows an awareness of the kind of questions that can be answered with digital data by a timeline visualisation. It makes use of an attribute in the database that holds the physical location of an artefact not in order to gain insights about the content of a collection, but to find out something about the thinking of the people that assemble a selection of an archive in order to exhibit it in public. The resulting insight is that the temporal distribution of the items that are on display is much more even, in comparison to the totality of items available.

I think they deliberately tried to have a representative range of material. (C7)

Earlier I mentioned London Transport Museum and their Johnston centenary. After our first encounter, I revisited them a year later and after a new curator has taken over the responsibility of the project. In the meantime, their project has progressed and the departing curator has spent a lot of time on cleaning the data, annotating it and tagging items that bear a relationship to Johnston. The task that the current curator suggested we could perform with the visualisation, was to

¹⁶ According to Unsworth, comparing is one of the “scholarly primitives” (2000); the basic tasks that humanists do, along with discovering, annotating, referring, sampling, illustrating, and representing.

import the latest version of the collections dataset along with the ones I had from before and to visualise the differences, to see how the collection data has changed in the meantime and which artefacts have gained or lost a relationship with Johnston as a result of their project.

The notion of visualising collections data on a timeline to learn more about the history of an institution came up several times before:

You'll see a range of clusters where we are planning a major exhibition. [...] We might acquire 2,000 objects in the space of two years, where normally we acquire one a year. [...] You don't need to know that we opened a new shipping gallery in 1963, you may just simply look at the data and it will tell you. (C3)

However, these insights concern the history of the physical collection as told by the collections data, while the suggestion of comparing changes to a catalogue aims to examine the history of collections data itself. The digital catalogue gains in relevance as an independent source of knowledge.

As more and more digital collections are released on GitHub, it will become feasible to study the history of collections data in a larger scale. Originally intended as a code repository, GitHub does not overwrite anything when new data is added to it: the complete revision history is maintained and remains accessible. I already encountered traces of editing activity in the Cooper Hewitt dataset when I updated my local copy to the latest version (see page 210).

Privacy

At this stage, I encounter increasing concerns over privacy issues with the availability of digital collections. When I began my research, the institutions I collaborated with readily shared their entire datasets. Concerns emerged when it became evident that collections data does not only hold data on artefacts, but also traces of individuals. Access and updates to a collection are logged with timestamps and user IDs. I suggested making use of multiple temporal aspects when visualising collections data; visualising data along those timestamps, however, results in being able to look at a detailed activity profile of individual curators, to see their preferred working hours, maybe when someone skipped work, etc.

An increasing awareness of issues of privacy is closely related to the formulation of higher level questions; when the interests revolve around cultural artefacts and their creators - who often have died long ago - privacy is less of an issue. However, when questions relate to collection activity and the decisions of individuals, these matters need to be taken into account when developing visualisation tools.

Coda

When I began my research project and developing visualisation tools, I expected the tools to lead to new knowledge of the kind that I now characterise as ‘low level’ insights. The original proposal for my PhD stated that “for example, a member of the public interested in a museum object would be able to see it in the context of all its neighbours in time”. While the TT visualisation, as we have seen, is indeed able to display images of objects in the context of other objects that share a temporal relation, there is an extended context of every record in a cultural dataset pointing towards the history of the collecting institution, the cataloguing individuals, biases and data revisions, which properly designed timeline tools are able to reveal as well.

Crucial to enabling these insights are not only the final tools, but the process and knowledge of creating them and the collaborative efforts of curators as experts, scholars and interested researchers, and myself. Participating in the design process made curators more aware of the characteristics of digital cultural data and the potential for visualising it. Collaborating and involving them in the entire process made me and them more informed and critical about tools and all aspects of the datasets.

My initial expectation was that visualisation tools will broaden curators’ understanding of their collection; these have been met, but in addition they extended their knowledge and understanding of the nature of digital data and the type of knowledge it is able to contain at high and low levels; close to the artefact and close to the data.

Discussion

The knowledge that timeline visualisations are able to produce from digital collections appears in different shapes. Being able to explore a collection through the visualisation enables a user to get an overview and see a collection (almost) in its entirety. The sense of ‘seeing everything’ is facilitated not only by having all the records graphically represented - individually or aggregated - but by allowing users immediate access to individual records.

Having an overview is useful for both casual and expert users. Even curators who are very familiar with a collection do not necessarily grasp it in its entirety. The abstraction of the diagrammatic representation produces a ‘map’ of the collection, both in the mathematical sense of mapping input data to a graphical output space, as well as in the common understanding of a cartography of a digital collection.

Visualisation can support common curatorial tasks such as selecting items for an exhibition, identifying errors - or supposed errors - and making corrections. Curators appreciated the ability to begin with everything and then filter down to the level of detail that they required, instead of their usual way of working with collections data; start empty and, through successive and targeted queries, build up the selection that they envisaged. This makes the presented timeline visualisation an analytic device in the true sense of the word; by allowing users to decompose a collection. Being able to filter in this way also facilitates understanding. Interactively manipulating a selection and seeing the effects visualised on the screen provides a cognitive advantage over mental problem solving.¹⁷

¹⁷ See also Kirsh and Maglio (1994) who observed that expert Tetris players find the optimal position and rotation of a brick not by planning, but by trying out different options through actively rotating and moving elements.

When used to support ‘normal’ curatorial tasks, the visualisation acts as a cognitive artefact. Curators commented that the visualisation encodes some of their tacit craft, the background knowledge that they require for studying their collections. Through visualisation, expert curators are spared from having to think consciously about the history of a dataset. Instead biases and inconsistencies are, to a certain extent, visible on the diagram. For curators who are new, or users outside an institution accessing a collection, appropriate visualisation tools allow them, too, to get an overview of a collection, but also to be able to see and consider the embedded interpretations and biases. This enables outside users to draw more informed conclusions from the dataset they examine.

When the visualisations are used to answer questions about the artefacts within a collection, it often helped to confirm existing knowledge. For example, the curators of the Britten-Pears foundation already knew who Britten’s favourite poets were and how they are distributed

in time. When insights are not surprising, they are nevertheless useful evidence. Patterns that are visible in the visualisation are immediately informative and the knowledge they confirm previously required extensive desk research. As a scholarly research tool, the comprehensible confirmation of existing knowledge or hypotheses is as important as the possibility of discovering ‘genuinely’ new knowledge. In addition, the visualisation allowed curators to study known patterns in more detail. We can precisely observe the distribution of Walter de la Mare poems in Britten’s oeuvre and we are able to not only see the extremes, but all shades in between.

When it comes to the area of temporal descriptions with regards to their multiplicities and uncertainties, the visualisation has offered previously unknown findings also on the level of a collection’s artefacts. By comparing the relationship between dates of composition and dates of first performance, we can see how Britten usually delivered his works shortly before they were first performed. By seeing the moment where first performances of childhood pieces started to take place, we can tell when people started to get interested in the oeuvre of Britten in its entirety.

This insight brings us towards the finding of meta-narratives, further away from the artefacts; possibly the most important aspect that these visualisations afford and a diversion from the established use of collections visualisation and museum timelines that focus on the artefacts or the biographies of their creators. When we look at a collection with regards to the embedded temporal uncertainties and, in the example of the Beazley archive or the Cooper-Hewitt collection, spot patterns and irregularities in the way the objects are dated, we essentially look at the traces of curators. We interpret how they have thought when they catalogued the artefacts, not at the artefacts themselves.

When I speculated that collections data is a resource in its own right besides the ‘physical’ collection I mainly envisaged that the sheer amount of data might lead to new insights. Instead, it appears that it is mostly the quality of data, how the data has been ‘crafted’, that bears important insights. In my research, I was able to discover new knowledge even with only a few fields per record. What is crucial, however, is to have some data about all the records – in contrast to having all data about some records. This constitutes a shift in humanities research which traditionally is concerned with close reading of a few artefacts rather than distant reading of a large set of items. Seeing everything is essential for distant reading as well. Context is everything; seeing data in its own context of all other records is what enables one to spot patterns.

The next step on the abstraction ladder of questions that collections data is able to answer is the study of data revisions; an area of research a curator has suggested and which I already touched on when I encountered data edits in the Cooper Hewitt dataset. Revision histories of collections data will be an important resource in the future, when more and more collections are published on platforms, such as GitHub, that store multiple versions of a dataset.

The presented insights appear as immediate outcomes of the prototypes and their application on museum data. It is important to consider however, the role that making the prototypes played and how the collaborative process of making led to new insights. By working together with the curators and understanding how they work, their actions and traces came into focus of my research more than they might have had I examined the datasets without their participation. Similarly, taking part in the design process and learning about visualisation - the functional principles of data visualisation and what kind of insights it enables - allowed curators to develop an understanding on the kind of questions they could answer through visualisation. While the curators generally exhibited a great awareness of the potential biases in their collection - in contrast to earlier studies of curators' views (Currall et al. 2005) - they did not anticipate that something which is part of their intuition could be visualised and studied.

7

“Curators and archivists are given the means to confront and reflect on the collecting history and cataloguing practices of their institutions.” (page 243)

7 Conclusion

Digital collections are increasingly made available online as a result of a growing recognition of their potential as resources for research. However, their role in knowledge production is still not fully understood. Often they are regarded as digital reproductions of the archives they represent; losing the physicality of the collection, but gaining the merits of digital accessibility.

I have argued for digital collections to be more closely related to catalogues of archives than to the archives themselves, and claimed that the kind of data these digital catalogues are able to contain leads to them becoming independent resources for knowledge production. As many theorists in the Digital Humanities observe (Winchester 1980; Hedstrom 2002; Thomas 2004; Borgman 2009; J. Drucker 2011a), we currently lack suitable tools to make sense of such humanities data, a problem my research addresses immediately through the development of suitable visualisation tools and, crucially, through making the thinking and knowledge behind their design explicit.

Visual timelines, my argument continues, can be effective and versatile scholarly research tools for making sense of cultural datasets. This claim is backed up by the successful history of this visualisation format of communicating and revealing new knowledge in the field of humanities. However, modern digital timelines do not face up to the challenges posed by visual analytics more broadly and cultural datasets specifically, as my review of existing digital timelines and the literature on visualisation in the Digital Humanities demonstrates. This thesis has explored these challenges with regards to the issues around visualising and interpreting large datasets, the development of suitable layout algorithms and the challenges around temporal uncertainties and multiplicities that are embedded in digital collections.

Contributions

By following a practice-based approach my research contributes novel insights that are relevant for real-world datasets and the kind of research questions that they are able to answer. Through collaborating closely with experts and users of digital collections, my work evoked a shift in the understanding of the epistemic role of digital collections; recognising and appreciating their value as resources that contain not only documented knowledge about cultural artefacts, but at the same time encapsulate a collection's history along with the embedded biases, ambiguities and subjectivities.

In contrast to the findings of earlier studies, my research demonstrated that curators are very much aware of the embedded interpretations of collections and know that they need to take them into account when working with cultural datasets. However, so far this has been largely anecdotal knowledge that could not be verified, or intuitive, tacit knowledge that was difficult to pass on. The visualisation tools I developed and the paradigms that they encapsulate enable curators and other users, for the first time, to see and interrogate the history of their collections.

By collaborating early on in a design process and jointly uncovering the complexities of humanities data and time-wise visualisation I was able to use potential biases and inconsistencies as a point of departure for critical discourse and provide the means for exploring and tackling them. These form an important contribution of this thesis.

Working closely with the 'owners' of the source material, I have been able to specify the key issues around visualisation of cultural data and formulate new paradigms to address the unique challenges of timeline visualisations and the issues emerging from partial, incomplete, contingent - in other words, real - data. Museums and archives are aware of the fact that their collections are biased, but instead of having to surrender to this reality, curators and archivists are given the means to confront and reflect on the collecting history and cataloguing practices of their institutions.

Process and Limitation

The practice-based research method was crucial in enabling my insights and those of the people I worked with. When I originally proposed the project I expected that the prototypes themselves would be the final outcome and the main contribution of the PhD. However all the prototypes I developed over the course of the research were also, and most crucially, a means to an end; a method to problematise and both discover and address the emergent problems.

When faced with the problem of representing uncertainty I, at first, followed a conventional approach and conceptualised ways of visually representing it in a standard timeline diagram. Only by testing the conceived solution with complete datasets - something that is hardly done in similar research projects¹- I became aware that this might not be the most viable solution. Existing work on alternative visual interfaces tends to present paradigmatic implementations based on a subset of data, often claiming that scaling a solution to larger datasets is merely a technical problem.² However, my research showed that it is more complex than that, that the real world is a world of “wicked problems”. My design-based research method was fundamental, not merely for verifying solution concepts, but to shape and reshape the problem.

The downside of practice-based research is its high level of subjectivity. Another researcher would have built other prototypes, would have drawn different conclusions from them and would maybe have focused on other challenges. Through diversifying sources and opinions I aimed to ease the subjective element in my research; through working with a variety of datasets and collaborating with curators of different institutions. Of course, these individuals all have their own subjectivities. For example, they all work for institutions in the UK and possibly have similar educational and professional backgrounds.

Collections data can, of course, not only be visualised on a timeline. Many curators asked for geographic maps and I have outlined that the timeline itself executes a bias towards the visualised data and the created knowledge; a timeline looks at data through time-tinted glasses. There are many possibilities for visualising data and each is able to produce new insights from collections data. The angle I chose to visualise datasets not only influences the potential knowledge, but also the course of the research process, as this thesis demonstrated; if I had not picked digital timelines - as a comparatively uncharted area of data visualisation - I would not have focused on the temporal element of collections data and worked with the aspects I did.

1 *Bohemian Bookshelf*, for example, an alternative visual interface for exploring digital libraries “included a collection of 250 books to ensure fluid real-time interaction” (Thudt et al. 2012). These were selected from the Open Library (openlibrary.org), a source that offers millions of books. SFMOMA’s ArtScope, another visualisation of a collection, features only around 20% of the data.

2 “While the performance of our prototype can easily be improved by applying more potent implementation strategies, some of the visualizations have to be adjusted to allow for larger data sets. [...] Common visualization techniques such as edge bundling could be used” (Thudt et al. 2012)

Cross-Contamination

For an interdisciplinary research project, it is not unusual for everyone involved in the research to fulfil different roles and I too needed to be aware of what position I am occupying in different parts of the project: not only a researcher, but also a designer and developer, and increasingly, an art-historian, musicologist, digital humanist. This progression of my own position into the field of humanities is not only a passive one - complementing my own knowledge and skills by collaborating with other researchers. Rather I argue that the cross-contamination between disciplines grows deeper, leading to participants making significant contributions outside their 'own' field.

Digital Humanities projects are characterised by being collaborative (Unsworth 2003; Hayles 2012, p.34). The adoption of computer science methods into humanities is an obvious cross-fertilisation of disciplines in this field. Recent discussions have moved towards more fine-grained elaborations of the epistemological consequences of digital technology in humanities disciplines, but the Digital Humanities is still fundamentally equated with the adoption of digital methods; a view that resonates in repeated questions if - or indeed demands that - humanists need to learn to code. Ramsay (2011) is a prominent defendant of this view, a stance that has been widely discussed and criticised (Gold 2012); Posner (2012) rightly flags up the intrinsic gender bias in calls for the necessity of practical computer skills.

Based on my research and my encounters with representatives of various disciplines, I found that the exchange between computer sciences and humanities are far more profound than the adoption of programming skills by humanities researchers. After all, coding is merely a skill learned and applied by professionals in various disciplines, and a skill that is not even essential for computer science itself. The science of computing is independent of the specific implementation.³

However, I do not mean to diminish what can be achieved through skilful programming. I do want to emphasise what we can do without. Intellectual progress is made through collaborations between the humanities and computer sciences in both disciplines by the exchange and awareness of problems and the proposal - not necessarily the practical implementation - of solutions. Talking about uncertainty and confidence in temporal expressions and seeing effects of database cataloguing visualised, prompted curators to make suggestions on the computational modelling of their data and engage with issues around data structures. In doing so they have not only adopted concepts from the computer sciences but made a valid contribution to the development of the field.

³ I have met various computer scientists who do not, and often cannot, write code.

In contrast to the intersections between humanities and computing, the opportunities and implications of collaborations between designers and humanists have only recently entered academic debate through scholars who remarked on the fundamental importance of ‘making’ in Digital Humanities (Presner et al. 2009) and the need to critically reflect on established paradigms for visual representation (J. Drucker 2011a; J. Drucker 2011b). Nevertheless, designers are rarely involved either in the conceptualisation or in the execution of Digital Humanities projects (Burdick 2009), despite growing evidence of the benefits of a deep relationship between designers and humanities scholars as collaborators and co-researchers (Caviglia 2013; Pellegrini et al. 2013; Uboldi et al. 2013).

As a designer, or as any ‘outsider’ collaborating on a research project, one should not underestimate the contributions that one can make also to the knowledge domain of other disciplines. There are discoveries I was able to make - by using methods that are available to me as a designer and developer to study cultural datasets in relation to time - that were novel to the owners and keepers of these datasets. Collaborating on Digital Humanities projects does not only mean adopting the skills and methods of each others’ discipline, but recognising the contributions that each field is able to make within the other fields in terms of creating new knowledge through their specific viewpoints and philosophies.

Future Work

I have centred my research around the specified focus issues I discussed. Elsewhere (Kräutli & Boyd Davis 2013) I identified additional challenges of time-wise visualisations in the humanities that the current research left untouched. I have addressed the ones I found most crucial, possibly creating more challenges in the process, of which some will be tackled by continuing PhD projects that were informed by the presented work.

Beyond Newtonian Time

With regards to time and how it is represented I have remained close to the computational, the Newtonian concept of time. Consequently I largely ignored alternative models of time along with the opportunities they might create for tackling some of the challenges I identified. I employed Newtonian time as an established system for structuring information, for its ease of use in a digital context and because it is the format which collections data largely use.⁴ Nevertheless, there is a lot of potential to develop the digital concept of time further, also within the confined realm of digital collections, for example by considering time not as an absolute coordinate space, but as a relational concept. The patterns in the form of contiguous assemblies that I identified in many collections and were often telling of batch acquisitions constitute events that happened at the ‘same time’, but nevertheless were recorded, as a computer demands it, in sequential order. We can see, therefore, that other traces of temporal relationships are embedded in collections data even though time is explicitly specified in Newtonian terms.

4 Linked Open Data formats might, in the future, lead to alternative and more complex definitions of time in collections data.

Beyond Collections Data

My findings and the paradigms I developed are not uniquely applicable to digital collections. Appropriately designed timeline visualisation tools can lend themselves to visually analyse other datasets, who might exhibit some of the same qualifiers. As part of an internship at Microsoft Research, I applied some of my thinking to personal data - email, calendar data, photographs. Born-digital data turns out to be as affected by subjective biases as manually created datasets. Examining the temporal aspects of personal data and making use of embedded temporal multiplicities leads to new insights. As in the case of collections data, these insights are most telling of the people that created it.⁵

5 Although the data is ‘born-digital’, it ultimately remains a human created artefact.

For example, events on calendars are displayed by the date they take place, but calendars also store when an entry has been added or edited. Email software tracks when an email has been received, but

also when it has been answered or forwarded. Photographs 'know' when they have been shot, but also when they have been downloaded from the camera or looked at. These temporal multiplicities allow for useful insights. In my findings, the timespan between the scheduling of an event and its date is strongly correlated with the amount of preparation that is needed ahead of it - 'hidden' data that could be used to remind a user ahead of time of the event. As data is increasingly stored in the cloud, time plays an ever more important role in digital data and the potential knowledge embedded in it that may become evident through visualisation multiplies.

Dissemination and Impact

The two prototypes that contain the insights from my enquiries are also available as open source libraries and for use by the public and researchers in and outside the humanities. So far they have been adopted by the Geffrye museum within an exhibition setting to visualise the objects that are on display. Visitors can explore their temporal distribution and the visualisation provides context by displaying their position in the entire archive, communicating curatorial decisions of selecting representative artefacts.

At the Royal College of Art, two PhD projects - by Olivia Vane and Sam Cottrell - look closer the concept of narration in and through visual timelines and at elements of uncertainty in temporal visualisations. Outcomes of the present research have informed and been used to secure these studentships.

My work in relation to the history of collections data have led to a commission by the ICA Philadelphia and Visual AIDS: an interactive timeline that visualises the revision history of the Wikipedia article on HIV/AIDS. For Microsoft Research I explored the temporal dimension of email communication through a set of visualisation-based research tools. My prototype interface of the ChronoZoom platform was awarded the Grand Prize of the Visualizing Time challenge.

My research had an immediate effect on the work of many of the institutions I collaborated with. The Britten-Pears Foundation applied conceptual and practical results in their own research and teaching activities. The London Transport Museum was able to gain new insights relevant for their ongoing research into the legacy of Edward Johnston. At the Courtauld Institute, my work has led to the development and funding of a new visualisation-based research project.

I have been invited to a Summer School on exploring digital collections at Harvard University's metaLab, where outcomes of my research have contributed to a handbook on exploring collections data (Mitchell 2014).

Based on my research output I have been awarded grants by the European Science Foundation, the Getty Foundation and the Volkswagen Stiftung for contributing to specialist conferences and workshops. As more and more cultural institutions are realising the potential of their digital collections, my work has often been used as evidence for the knowledge they contain and how visualisation can reveal it (Simon 2013; Bailey & Pregill 2014; Oates 2014; Rodley 2014; Winesmith & Carey 2014).

Appendix

A Timeline Criteria

1 Data Acquisition and Curation	252
2 Representation of Dataset	253
3 Representation of Record	255
4 Interaction	258
5 Technology/Platform	262

1 Data Acquisition and Curation

How does data enter the system? What requirements does a dataset need to meet? How does a dataset have to be processed in order for it to be visualised?

1.1 Custom Data Import Possible

Does the tool support custom data import, either through coding or within the UI?

1.2 Conditioning (Preprocessing Steps)

Does the tool support or require conditioning of the imported data (e.g. specify data type, granularity, etc.) as part of the import process?

1.3 Used/Expected Format

What (standard) formats are used for the represented data (e.g. CSV, XML, JSON)?

1.4 Used/Expected Fields

What (kind of) fields are supported and/or required (e.g. date, text, tags)?

1.5 Size of Implemented/Supported Dataset

What is the total/maximum number of records supported?

2 Representation of Dataset

How is the entire dataset graphically represented?

2.1 Assembly

What kind of diagram format is used to represent the dataset visually?
Which principles does it follow?

2.1.1 Layout Algorithm Used

What is the automated procedure by which the timeline layout is constructed?

2.1.2 Possibility to Select Layout

Are there alternative timeline layouts?

2.1.3 Randomness

Is there an element of (pseudo-)randomness or is the layout deterministic?

2.2 Temporal Model

How is time digitally and visually present?

2.2.1 How the Model of Time Progresses

What is the mathematical model of time? (linear, non-linear, logarithmic, distorted, continuous, with omissions, transformations...)

2.2.2 How the Time Axis Is Drawn on the Screen

What is the graphical model of time? (straight, curved, circular, ...)

2.3 Axes

If present, what is the significance of graphical axes?

2.3.1 Temporal

How is/are the temporal axis/axes represented?

2.3.2 Non-Temporal

How is/are the non-temporal axis/axes represented?

2.4 Readability

How readable is the graphical representation of the dataset?

2.4.1 Size of Dataset

Does the size of the represented dataset correspond to the total size of the dataset?

2.4.2 Content of Dataset

Is the content of the dataset readable in the representation?

2.4.3 Scope of Dataset

Is the structure/composition of the dataset comprehensible?

2.4.4 Readability in Sparse Regions

Are patterns readable in regions with little data?

2.4.5 Readability in Dense Regions

Are patterns readable in regions with a lot of data?

3 Representation of Record

How are individual records graphically represented?

3.1 Graphical Vocabulary

Which graphical elements are used to visually represent records?

3.1.1 Shapes

3.1.2 Colours

3.1.3 Lines

3.1.4 Images

3.1.4.1 Image Transformations

Are images processed before they are represented? (crop, scale, filter, ...)

3.2 Mapping Modality

How do attributes of a record control their visual representation?

3.2.1 Duration to Size

Is the temporal duration of an attribute represented as an element of size?

3.2.2 Position to Time

Is a temporal attribute of a record translated to a position of a graphical representation?

3.2.3 Mapping of Other Attributes

How are other attributes of a record graphically represented?

3.2.4 Temporal Granularities and Uncertainties

To what level of (im)precision are temporal attributes graphically represented?

3.2.5 Types of Events

Are there different representations for e.g. single events, periods, background events,...

3.2.6 Graphical Artefacts (Un-Mapped Attributes)

Which visual features are not driven by data?

3.3 Mapping Consistency

How consistent and comprehensible is the visual mapping of data attributes?

3.3.1 Incoherences/Randomness

Are there random effects in the graphical mapping of data?

3.3.2 Quantisation

What level of quantisation do individual attributes undergo?

3.3.3 Exceptions

Are there exceptions in the visual mapping? (e.g. records that have been positioned manually?)

3.4 Non-Graphical Representations

Are there elements of the dataset which are represented through other means than graphics?

3.4.1 Text

3.4.2 Audio

3.5 Temporal Descriptions

What kind of temporal descriptions in a record does the visualisation expect and support?

3.5.1 Supported Granularity of Date Descriptions

Is there a limitation towards the maximum or minimum precision of temporal descriptions

3.5.2 Events Without Time

Are events with missing dates represented?

3.5.3 Events With Single Point in Time

Does the visualisation support events with a single temporal attribute?

3.5.4 Events With Multiple Points in Time

Does the visualisation support events with multiple temporal attributes?

3.5.5 Events As Periods

Does the visualisation support representing events that are non-durational, but nevertheless have a time period associated with them?

3.5.6 Periods

Does the visualisation support durational events?

3.5.7 Periods With Discontinuities

Does the visualisation support durational events with breaks?

3.5.8 Tolerances

Does the visualisation support imprecisions in temporal description?

4 Interaction

How does a user interact with the visualisation?

4.1 Exploration / Navigation of Visualisation

What interaction paradigms are implemented that support navigation and exploration of the visualisation?

4.1.1 Orientation Assistants

Are there features that offer additional information on the user's position in the visualisation?

4.1.1.1 Visual

4.1.1.2 Non-Visual

4.1.2 Transformations of Layout

Can the assembly of the layout be manipulated?

4.1.2.1 Zoom Behaviour

Can the layout be zoomed? What zooming mechanism is used?

4.1.2.1.1 Linear

4.1.2.1.1 Semantic

4.1.2.1.1.1 Supported Reorganisation Principle

What is the driving mechanism of the semantic zoom? (e.g. summarise events, sample events, reduce detail, ...)

4.1.2.2 Pan Behaviour

Can the layout be panned?

4.1.3 Transformations of Record Representations

Can the representation of individual records be manipulated?

4.1.3.1 Scaling

Can records be rescaled?

4.1.3.2 Repositioning

Can records be repositioned?

4.1.3.3 Colouring

Can records be coloured?

4.1.4 Detail on Demand

Is it possible to attain more details about a graphically represented record?

4.1.4.1 Ability to Learn More About a Record

Is it possible to get more details about an individual record?

4.1.4.2 Ability to Learn More About a Group of Records

Is it possible to get common details about several records?

4.1.4.3 Exposure of Data Layer

Is the dataset that drives the visualisation accessible?

4.1.5 Manipulation of Data Layer

Can the dataset be changed from within the visualisation? I.e. can individual records be edited?

4.1.6 Manipulation of Mapping

Can the representation of individual records be manipulated?

4.1.6.1 Attribute

Is it possible to select which attribute controls a graphical variable?

4.1.6.2 Graphical Variable

Is it possible to select the graphical variable an attribute controls?

4.1.6.3 Transformations

Is it possible to change the mapping modality? (i.e. applying thresholds, different scalings, etc.)

4.2 Drawing Relationships

What interaction paradigms does the visualisation support for drawing relationships?

4.2.1 Manipulate Non-Time Axis

Can the non-time axis be scaled or reassigned?

4.2.2 Manipulate Time-Axis

Can the time axis be manipulated?

4.2.2.1 Rescale

Is it possible to adjust the scaling of the time axis?

4.2.2.2 Omit

Is it possible to suppress certain time frames?

4.2.2.3 Projections

Is it possible to select the mapping modality of the time axis? (linear, log, lens, ...)

4.2.3 Filter Records

Is it possible to manipulate the records?

4.2.3.1 Search

Does the visualisation support searching for records?

4.2.3.2 Extract/reorder

Can records be reorganised or looked at separately?

4.2.3.3 Remove

Can records be removed?

4.2.3.4 Emphasise

Can records be selected or highlighted?

4.2.4 Contextualising

Does the visualisation allow putting the data into (temporal) context?

4.2.4.1 Ability to Add (Background) Events

Is it possible to add custom events?

4.2.4.2 Ability to Include Several Datasets

Is it possible to examine several datasets at once?

4.3 Sense-Making

What interaction paradigms support understanding and discovery?

4.3.1 Direct Manipulation

Is every output also an input?

4.3.2 Feedback of System State

Is the user kept informed of internal processes?

4.3.2.1 Visual

4.3.2.2. Textual

4.3.3 Traceability of Manipulations

Is it possible to retrace custom manipulations? Is there an 'undo' function?

4.3.4 Reproducibility of Findings

Do the same manipulations always lead to the same outcome?

5 Technology/Platform

On what system does the visualisation run? What are the expectations towards it?

5.1 Implementation

How is it implemented? Is it a standalone software or a modular system?

5.1.1 Standalone Tool

5.1.2 Tied to Dataset

5.1.3 Library

5.2 Platform

Is the visualisation platform independent or does it rely on proprietary formats?

5.2.1 Native

OS X, Windows, Unix, Linux, etc.

5.2.2 HTML/JS

Standard web-technologies

5.2.3 Flash, Etc.

Proprietary web-technologies

5.3 Dependent Libraries

e.g. jQuery, d3, ...

B Datasets

The Beazley Archive	264
Britten-Pears Foundation	264
ChronoZoom	265
Cooper Hewitt, Smithsonian Design Museum	266
Geffrye Museum	266
London Transport Museum	267
Museum of Domestic Design and Architecture (MoDA)	267
Museum of Modern Art (MoMA)	268
Penn Museum	268
Tate	269
Victoria and Albert Museum	269

The Beazley Archive

The Beazley Archive forms part of Oxford University's Classical Art Research Centre. It maintains the world's largest collection of photographs depicting and describing ancient decorated pottery. An online search interface provides researchers access to the collection's dataset and enables them to export search results in an XML format. For my research I obtained a full export of the database.

<http://www.beazley.ox.ac.uk>

Publicly Available

<http://www.beazley.ox.ac.uk/xdb/ASP/default.asp> (accessed 09.02.2016)

Format

XML

Size

107,391 records

Britten-Pears Foundation

The Britten-Pears Foundation is based in Aldeburgh, UK, and maintains a collection of works by Benjamin Britten, along with other relevant material such as manuscripts, documents and personal artefacts related to Britten and his partner Peter Pears.

<http://www.brittenpears.org>

Britten Thematic Catalogue

The Britten Thematic Catalogue is a database detailing all of Britten's works, including his early childhood pieces, along with manuscript sources and audio incipits. It is available via an online interface, however for the purpose of this research I have obtained a copy of the database in the MySQL format directly from the Britten-Pears foundation.

Publicly Available

<http://www.brittenproject.org/> (accessed 13.01.2016)

Format

MySQL/XML

Size

1,238 works, 192 first performances

Britten's Poets

For an exhibition on the writers and poets that appear in Britten's works, the curators assembled a dataset that contains tombstone information of authors along with the compositions they appear in. I received this dataset as an Excel sheet and converted it into a MySQL database, as some data was expressed using cell formatting in Excel and had to be translated to a relational database format in order to be modelled computationally.

Publicly Available

no

Format

Excel, converted to MySQL

Size

481 works, 62 work cycles, 156 authors

ChronoZoom

ChronoZoom is an interactive visual timeline for studying Big History. It started out as a project by Walter Alvarez and Roland Saekow at the University of California, Berkeley, and has been developed further in collaboration with Microsoft Research. ChronoZoom is both a visual interface and a community-created dataset of historic events, organised in nested timelines. The dataset is accessible separately from the interface through an API.

<http://www.chronozoom.com/>

Publicly Available

<http://www.chronozoom.com/api/gettimelines> (accessed 25.01.2016)

Format

JSON

Size

367 timelines, ca. 7000 events

Cooper Hewitt, Smithsonian Design Museum

The Cooper Hewitt, Smithsonian Design Museum in New York collects historical and contemporary design. Their collection comprises more than 210,000 objects. Their collections dataset is available on the code-sharing platform GitHub and includes data on individual objects, as well as people and periods. For my research I worked with the objects dataset.

<https://www.cooperhewitt.org>

Publicly Available

<https://github.com/cooperhewitt/collection> (accessed 13.01.2016)

Format

JSON

Size

125,364 objects

Geffrye Museum

The Geffrye Museum of the Home in London documents and explores the history of English homes. Their collection includes objects that partly comprise “period rooms” - representative interiors for a given time period - as well as photographic documentation of past and present living environments. Their collection is accessible through their website. An API is available, but not currently advertised to the public.

<http://www.geffrye-museum.org.uk/>

Publicly Available

web-based: <http://www.geffrye-museum.org.uk/collections/>
(accessed 13.01.2016)

API (beta): <http://mailgate.geffrye-museum.org.uk:7452/api/rest/beta/objects/> (accessed 13.01.2016)

Format

JSON

Size

27,055 records

London Transport Museum

The London Transport Museum maintains a collection of more than 450,000 objects, including posters, photographs, vehicles and many items that relate to the type designer Edward Johnston. I collaborated with museum curators working on the centenary commemorations for Edward Johnston and explored a collection of items relevant for their research.

<http://www.ltmuseum.co.uk>

Publicly Available

only web-based: <http://www.ltmcollection.org> (accessed 13.01.2016)

Format

CSV (export)

Size

2,807 records

Museum of Domestic Design and Architecture (MoDA)

MoDA forms part of Middlesex University, London, and houses a large collection of textiles and items related to the Silver Studio, an influential UK design company, along with further collections related to domestic design. For research purposes, the curators gave me access to an export of their collections data, which includes records on objects along with records that describe usage of those objects as parts of exhibitions and publications.

<http://www.moda.mdx.ac.uk/>

Publicly Available

only objects data, through web-based interface: <http://www.moda.mdx.ac.uk/Collections> (accessed 09.02.2016)

custom export not available

Format

CSV (export)

Size

3,579 object records, 265 procedure records

Museum of Modern Art (MoMA)

The Museum of Modern Art in New York collects modern and contemporary art. Its collection spans around 200,000 works. Their collections data is available on GitHub. In contrast to the online collection on their website, the GitHub data also includes records that contain incomplete information, marked as “not Curator Approved”.

<http://www.moma.org/>

Publicly Available

<https://github.com/MuseumofModernArt/collection>
(accessed 14.01.2016)

Format

CSV, JSON

Size

126,119 records

Penn Museum

The University of Pennsylvania Museum of Archaeology and Anthropology is the largest university museum in the United States and provides researchers in anthropology access to about one million objects. A large part of their digital collection is available for download on their website in a variety of formats.

<http://www.penn.museum/>

Publicly Available

<http://www.penn.museum/collections/data.php> (accessed 14.01.2016)

Format

CSV, JSON, XML

Size

365,057 object records (representing more than 835,000 objects)

Tate

The Tate maintains the United Kingdom's national collection of British and international art, which it exhibits in four museums: Tate Britain and Tate Modern, both in London, Tate Liverpool and Tate St Ives. Their collections data is available on GitHub and covers almost 70,000 artworks, some of which are jointly owned by the Tate and the National Galleries of Scotland. In addition, the dataset contains tombstone information on associated artists.

<http://www.tate.org.uk/>

Publicly Available

<https://github.com/tategallery/collection> (accessed 09.02.2016)

Format

CSV, JSON

Size

69,202 artwork records, 3,534 artist records

Victoria and Albert Museum

The Victoria and Albert Museum, or V&A, is a major museum of art and design and maintains a collection of roughly 4.5 million objects. More than one million of them are available through an online search interface. An API is available, but currently (14.01.2016) not fully functional.

<http://www.vam.ac.uk/>

Publicly Available

online: <http://collections.vam.ac.uk/>

API: <http://www.vam.ac.uk/api/json/museumobject/>

Format

JSON

Size

1,167,880 (online), 2,000 (API)

c List of Digital Timelines

Aeon Timeline	272
BEEDOCS Timeline 3D	273
CATT Lab Timeline	274
CATT Lab Time Spiral	275
ChronoZoom	276
Citeology	277
Continuum	278
Cooper Hewitt Event Horizons	279
GE Annual Reports	280
Guardian Timeline: The path of protest	281
History Flow	282
History's Largest Empires	283
Histropedia	284
Kindred Britain	285
Neatline	286
NYT Timelines	287
Palladio	289
Scaled in Miles	290
SIMILE Timeline	291
The Evolution of Innovation	292
Tiki-Toki Timeline	293
TimeFlow	294
Timeglider	295

List of Digital Timelines

271

Timeline	296
TimelineJS	297
U.S. Gun Deaths	298
vis.js Timeline	299
Wellcome Timeline	300

Aeon Timeline

Developer

Scribble Code

Website

www.scribblecode.com

Type

Standalone (Native OS X)

License

Commercial

Dataset

custom

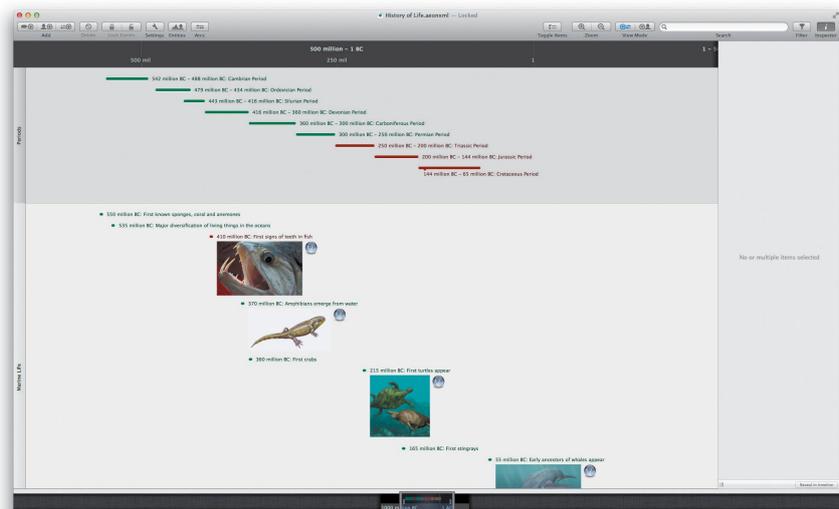
Release History

19.06.2012 1.0.5 (first public 1.x version)

31.10.2008 Beta Release 0.1

Screencast

<https://www.youtube.com/watch?v=tWosFJKJk8o>



CATT Lab Timeline

Developer

CATT Lab University Of Maryland

Website

<http://www.cattlab.umd.edu/?portfolio=timeline>

Type

Browser based (Flash)

License

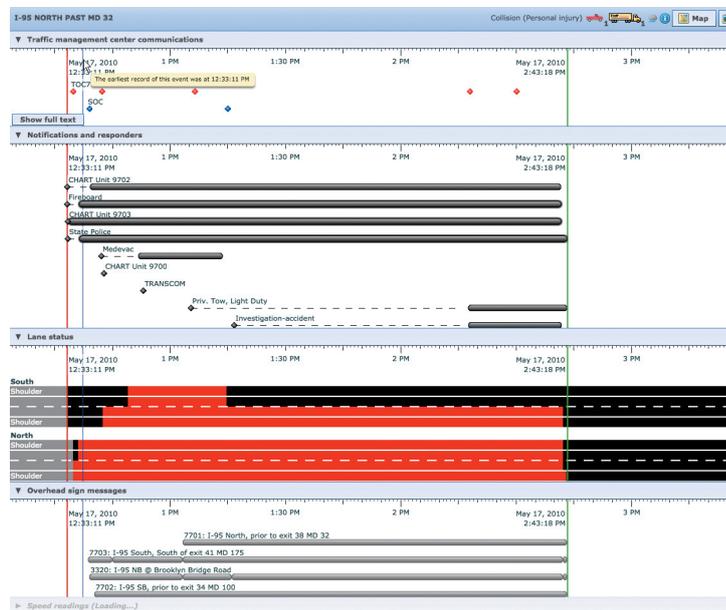
Internal

Dataset

Traffic events (intended for custom datasets)

Release History

2007? (based on data in timeline)



CATT Lab Time Spiral

Developer

CATT Lab University Of Maryland

Website

<http://www.cattlab.umd.edu/?portfolio=timeline>

Type

Browser based (Flash)

License

Internal

Dataset

Traffic events (intended for custom datasets)

Release History

2008? (based on data in timeline)

Screencast

<https://www.youtube.com/watch?v=NTP9p7-DeuY>



ChronoZoom

Developer

Walter Alvarez and Roland Saekov (UC Berkeley), Microsoft Research,
Moscow State University

Website

www.chronozoom.com

Type

Browser based

License

Open Source

Dataset

Events of World History (customisable)

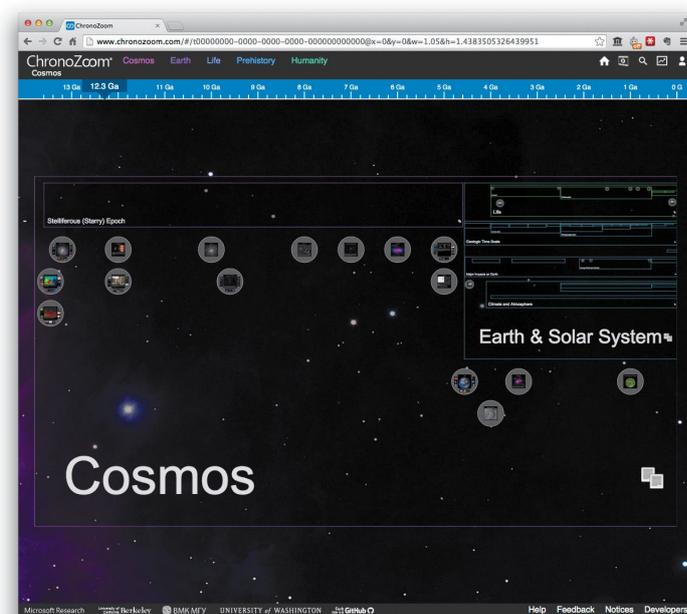
Release History

14.3.2012 (2.0, complete rewrite in HTML5 and made open source)

29.4.2010 (1.0)

Screencast

<https://www.youtube.com/watch?v=OBGi8ZKPof0>



Citeology

Developer

Autodesk Research

Website

<http://www.autodeskresearch.com/projects/citeology>

Type

Java based visualisation

License

Internal

Dataset

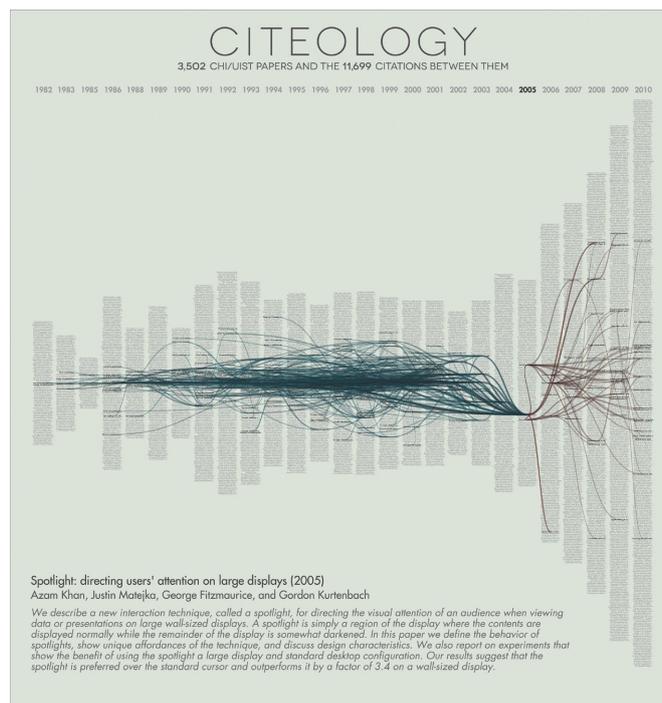
CHI Papers

Release History

2010

Screencast

<https://www.youtube.com/watch?v=aEwS47G3rIi>



Continuum

Developer

André, P., Wilson, M. L. and schraefel, m. c.

Website

<http://research.mspace.fm/projects/continuum>

Type

Standalone tool

License

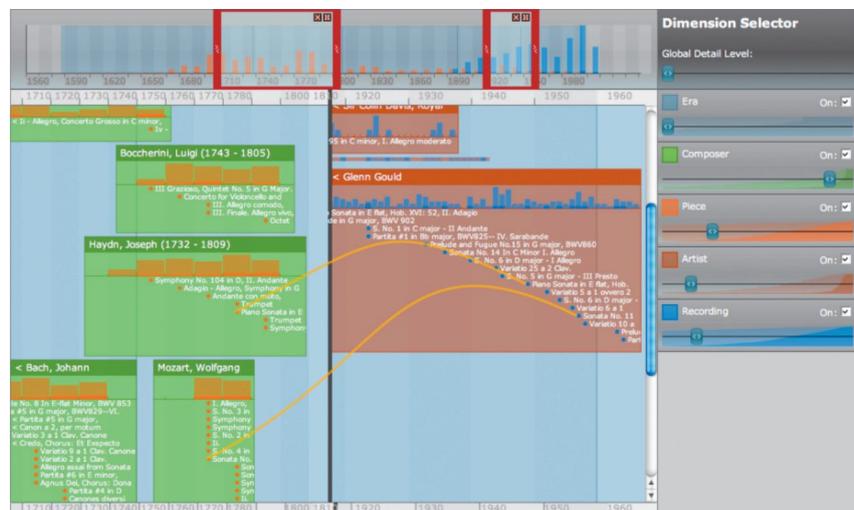
Internal

Dataset

Compositions (intended for custom datasets)

Release History

2007 (paper published, unreleased)



Cooper Hewitt Event Horizons

Developer

Cooper Hewitt (Aron Straup Cope)

Website

<https://github.com/cooperhewitt/d3-timeline-event-horizon>

Type

Browser based (Library Module)

License

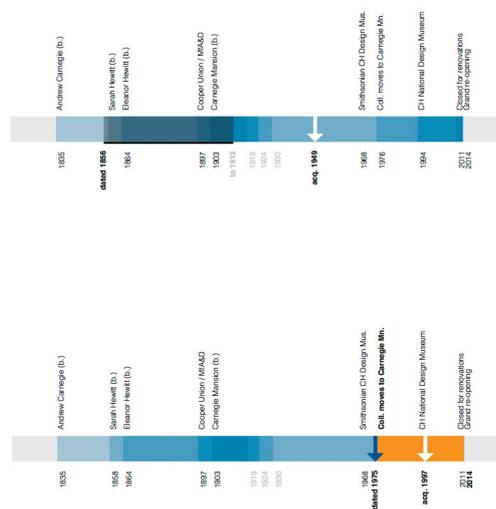
Open Source

Release History

September 2013 (<http://labs.cooperhewitt.org/2013/a-timeline-of-event-horizons>)

Screencast

<https://www.youtube.com/watch?v=t3souqxnVPM>



GE Annual Reports

Developer

Fathom

Website

<http://www.ge.com/visualization/annual> (down currently, 18.11.2014)

<http://fathom.info/latest/2237>

Type

Browser based visualisation, iOS app

License

Internal

Dataset

6000 pages; 120 years' worth of GE annual reports, spanning the years 1892-2011

Release History

20.3.2012



Guardian Timeline: The Path of Protest

Developer

Guardian

Website

<http://www.theguardian.com/world/interactive/2011/mar/22/middle-east-protest-interactive-timeline>

Type

Browser based visualisation (Based on Away3D/Flash)

License

Internal

Dataset

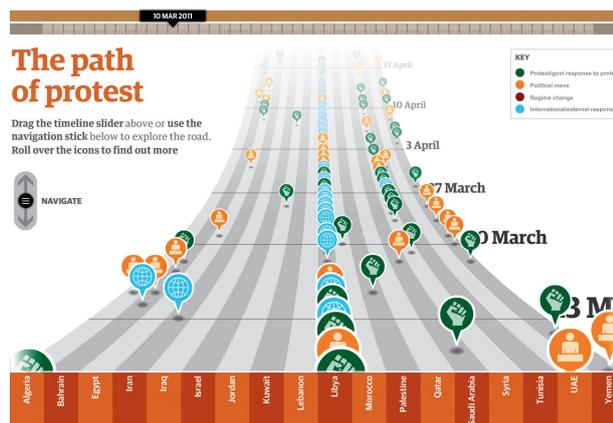
Curated dataset of Guardian news reports

Release History

2011

Screencast

<https://www.youtube.com/watch?v=B1YoGystZLE>



History Flow

Developer

Fernanda Viegas & Martin Wattenberg (at IBM)

Website

<http://www.bewitched.com/historyflow.html>

<http://fernandaviegas.com/wikipedia.html>

Type

Browser based visualisation

License

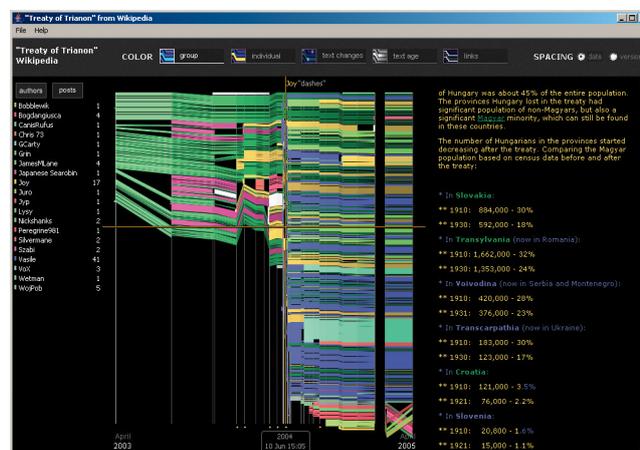
Internal

Dataset

custom

Release History

2003 (not publicly available)



Histropedia

Developer

Navino Evans

Sean McBirnie

Website

<http://histropedia.com/>

Type

Browser based

License

Creative Commons (CC-BY-SA)

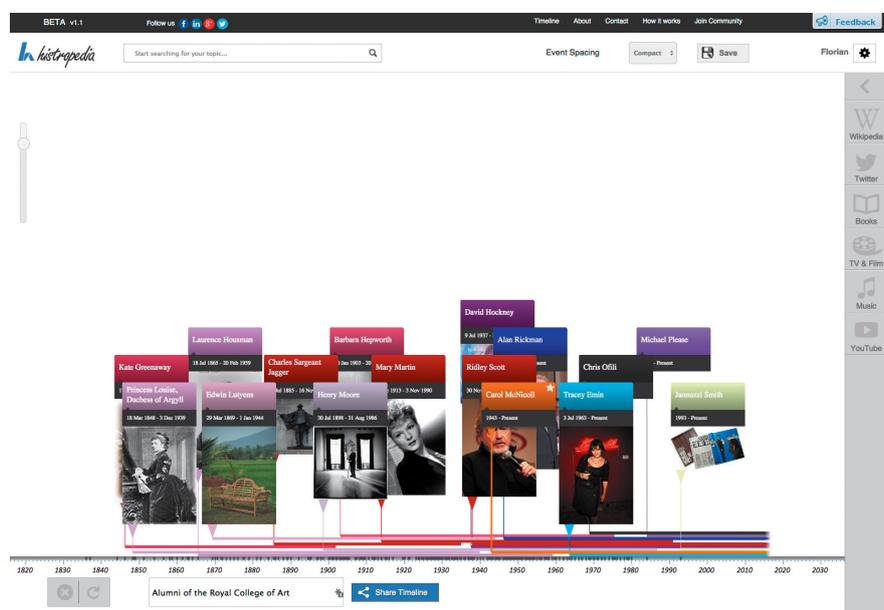
Release History

29.07.2015 (v. 0.8)

19.02.2013 (first blog post)

Screencast

<https://www.youtube.com/watch?v=y8KjoQHHh1M>



Kindred Britain

Developer

Stanford University Libraries

Website

<http://kindred.stanford.edu/>

Type

Browser based visualisation

License

Internal

Dataset

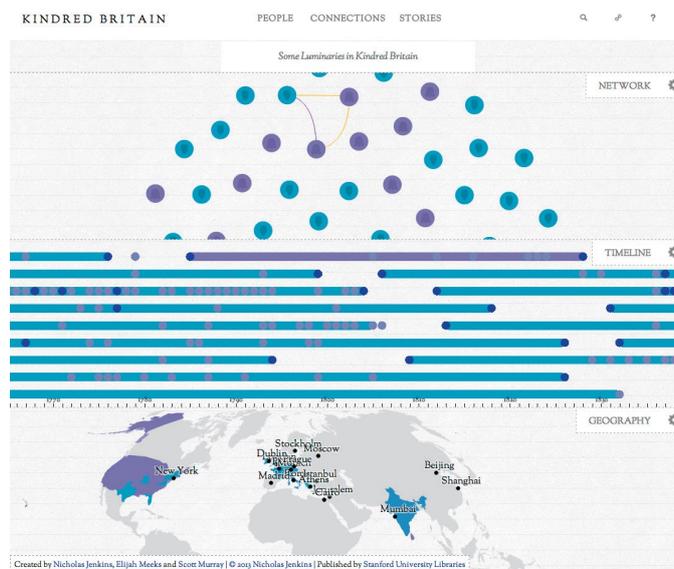
Dataset of 30,000 individuals; iconic figures in British culture

Release History

2013 (copyright notice)

Screencast

<https://www.youtube.com/watch?v=3LQI6czR1Co>



Neatline

Developer

Scholars' Lab (UVA Library et al.)

Website

www.neatline.org

Type

Browser based / Library

License

Open Source

Dataset

custom

Release History

28.7.2014 (2.3)

9.7.2013 (2.0)

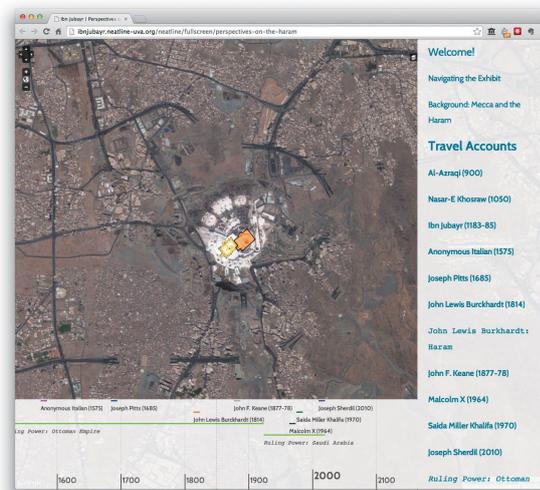
2.7.2012 (1.0)

Initially released as an online tool and a package for custom hosting.
Online tool discontinued since v. 2.0

Screencasts

https://www.youtube.com/watch?v=u_vy6DXtM9A

https://www.youtube.com/watch?v=ggiG_orF9hs



NYT Timelines

Developer

The New York Times

Website

A History of the C.I.A.'s Secret Interrogation Program, <http://www.nytimes.com/interactive/2014/12/09/world/timeline-of-cias-secret-interrogation-program.html> (accessed 17.08.2015)

Steven P. Jobs: His Life, His Companies, His Products. <http://www.nytimes.com/interactive/2011/10/05/business/20111005jobs-life-timeline.html> (accessed 17.08.2015)

Timeline of Nelson Mandela: 1918-2013, <http://www.nytimes.com/interactive/2013/12/05/world/africa/Mandela-Timeline.html> (accessed 17.08.2015)

Type

Browser based visualisations

License

Internal

Dataset

Custom datasets of news events

Release History

several versions released in 2011, 2013 and 2014

Screenshots

Previous April 1, 1976 Next April 17, 1977

July 1976
Apple Introduces Its First Computer
 The Apple I goes on sale and is priced at \$666.66.

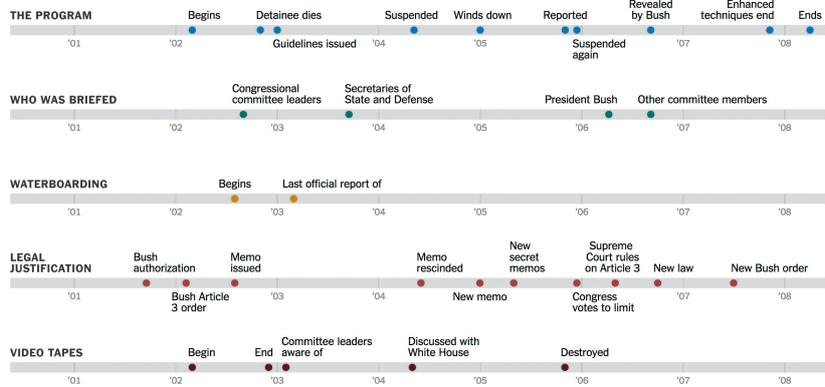
Rebelplot, via Creative Commons
 Apple's first product, the Apple I, was a kit that included a motherboard. The keyboard and wooden casing were added to this computer.

The Life and Legacy of Nelson Mandela: 1918-2013

Nelson Mandela's quest for freedom in South Africa's system of white rule took him from the court of tribal royalty to the liberation underground to a prison cell to the presidency.

Previous Next

July 18, 1918 June 1, 2004 2013



Palladio

Developer

Stanford Design + Humanities

Website

<http://palladio.designhumanities.org/>

Type

Browser based visualisation tool

License

internal

Dataset

custom

Release History

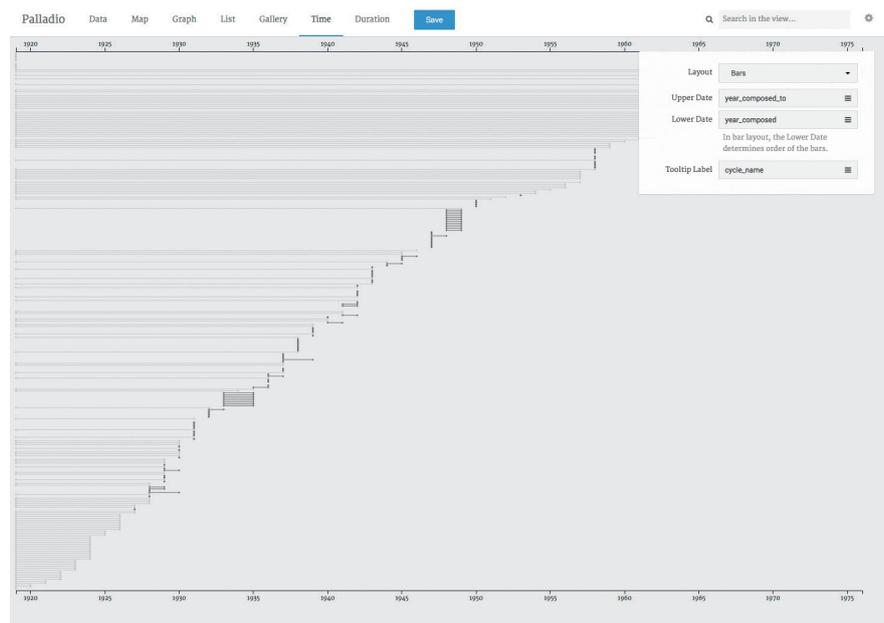
23.09.2014 (0.7.0)

16.04.2014 (release)

Screencasts

<https://www.youtube.com/watch?v=zbh9jz8oEb8>

<https://www.youtube.com/watch?v=yXZtzimhmlYi>



Scaled in Miles

Developer

Fathom

Website

<http://fathom.info/miles/>

Type

Browser based visualisation

License

Internal

Dataset

Contributors to Miles Davis recordings

Release History

2014

Screencast

<https://www.youtube.com/watch?v=obUEKEVj7SI>



The Evolution of Innovation

Developer

Popular Science

Website

<http://www.popsci.com/content/best-whats-new-graphic>

Type

Browser based visualisation

License

Internal

Dataset

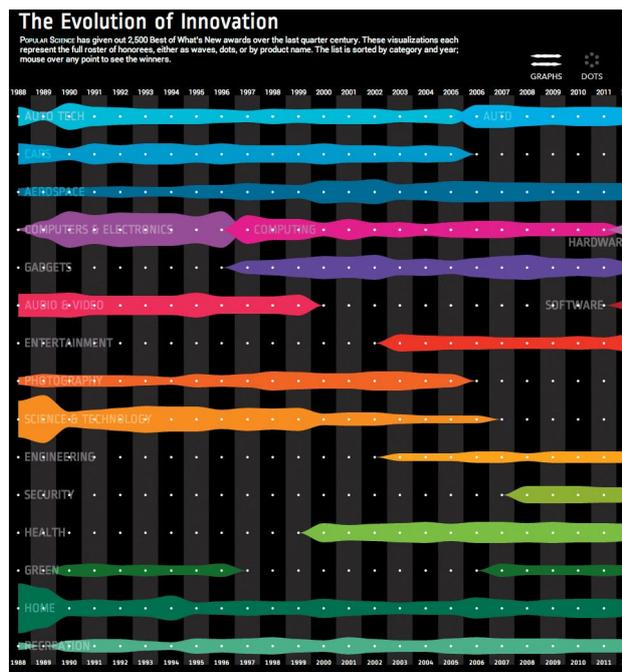
Popular Science 2,500 Best of What's New awards over the last quarter century

Release History

2011

Screencast

<https://www.youtube.com/watch?v=NR-qpeXunu8>



Tiki-Toki Timeline

Developer

Webalon Ltd.

Website

www.tiki-toki.com

Type

Browser based

License

Commercial

Dataset

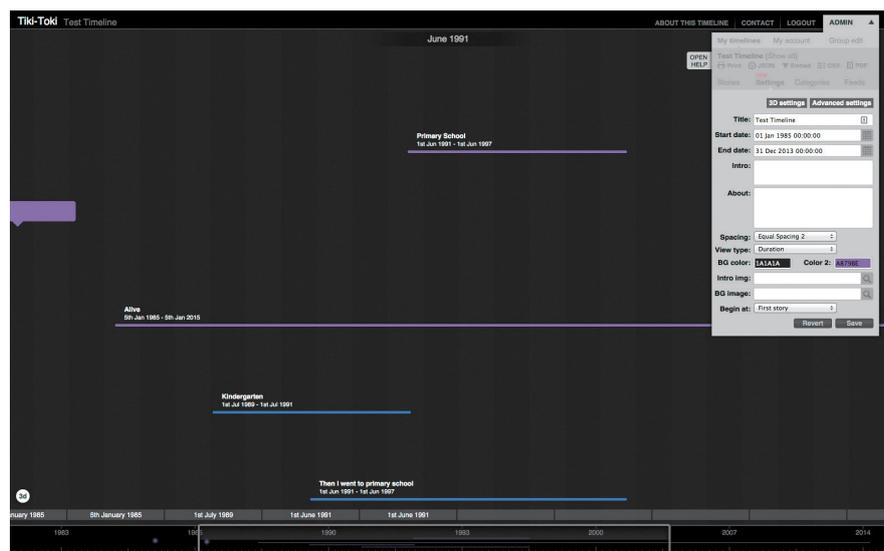
custom

Release History

2011 (first appearance on Google.com)

Screencast

https://www.youtube.com/watch?v=9hbHj_oM1W4



TimeFlow

Developer

Flowing Media (Fernanda Viegas, Martin Wattenberg)

Website

www.flowingmedia.com/timeflow.html

Type

Standalone (Java)

License

Open Source(?)

Dataset

custom

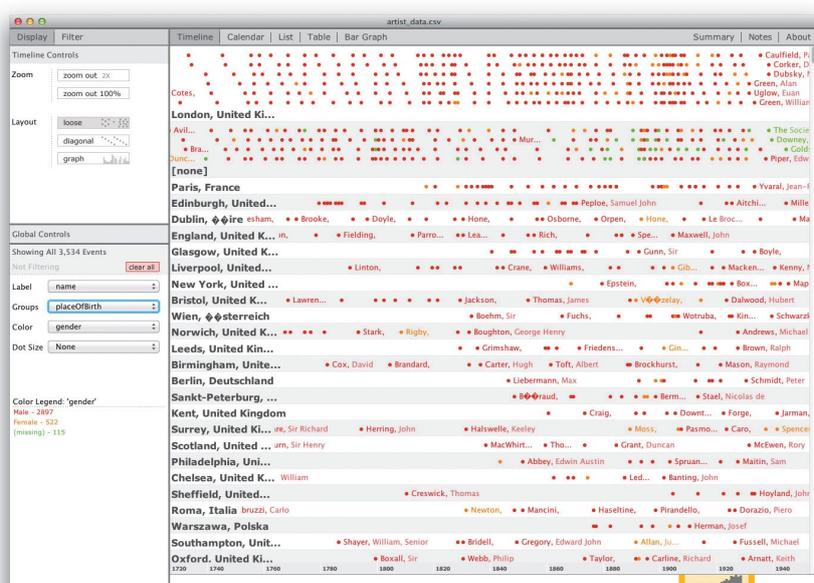
Release History

ca. 2010 (Alpha)

Screencast

<https://www.youtube.com/watch?v=39dCyH-A5ik>

https://www.youtube.com/watch?v=O_LWfb_wnwE



Timeglider

Developer

Timeglider (Memograph LLC)

Website

<http://timeglider.com/>

Type

Browser based (Widget)

License

Free, Limited Commercial, and OEM/SaaS Integration

Dataset

custom

Release History

October 2010 (MIT License, Flash based)

Later (2012?) move to JS based solution

Screencast

<https://www.youtube.com/watch?v=KPI4iOBQSYo>



Timeline

Developer

Rickard Lindberg

Website

thetimelineproj.sourceforge.net/

Type

Standalone (Windows, Linux)

License

Open Source (GNU)

Dataset

custom

Release History

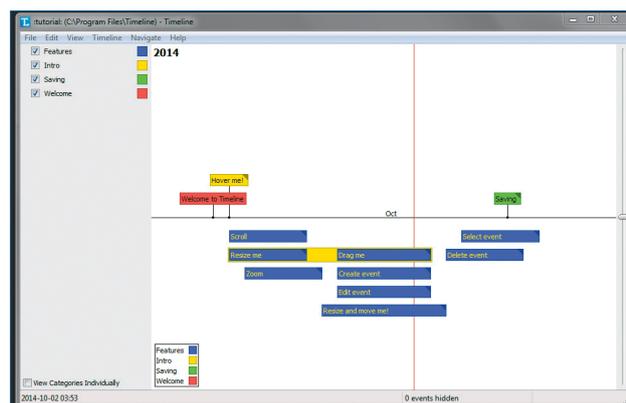
30.6.2014 (1.3)

30.9.2013 (1.0)

11.4.2009 (0.1.0 first usable version)

Screencast

https://www.youtube.com/watch?v=_KenYSuSoJA



TimelineJS

Developer

knightlab (Northwestern University)

Website

timeline.knightlab.com/

Type

Browser based (library)

License

Open Source (MPL 2.0)

Dataset

custom

Release History

Probably 2012 (first git commit)

Screencast

<https://www.youtube.com/watch?v=2XT326FJew8>

MARCH 14, 1948

EDITH CLARKE: BUCKING THE TREND

There is no demand for women engineers, as such, as there are for women doctors; but there's always a demand for anyone who can do a good piece of work.

—Edith Clarke, first female professor of electrical engineering in the United States

Timeline JS

U.S. Gun Deaths

Developer

Periscopic.com

Website

<http://guns.periscopic.com/?year=2013>

Type

Browser based visualisation

License

Internal

Dataset

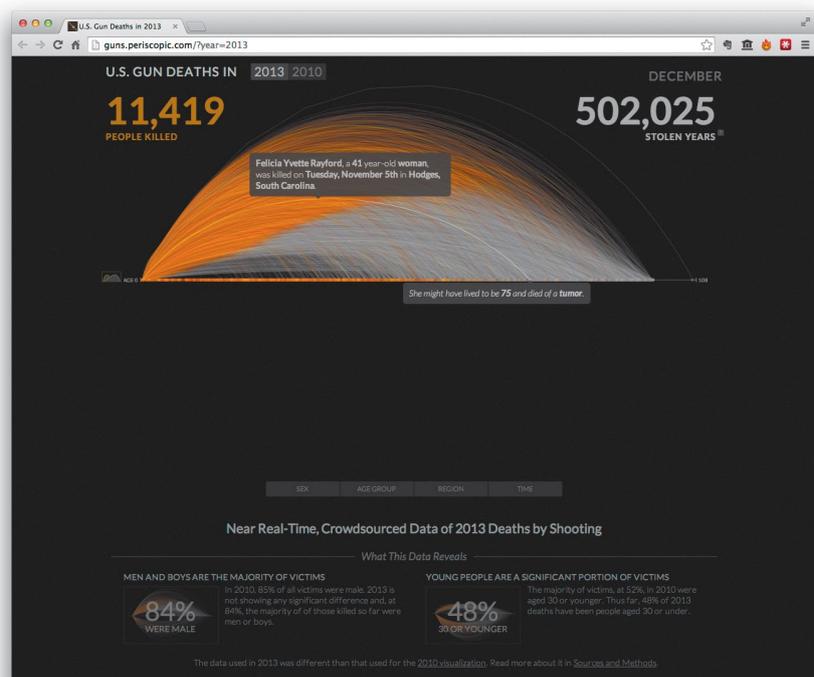
US gun deaths in 2010 and 2013 from FBI's Uniform Crime Reports (2010) and @GunDeaths (2013)

Release History

2013

Screencast

https://www.youtube.com/watch?v=07nLL_Fg_N8



vis.js Timeline

Developer

Almende BV

Website

<http://visjs.org/>

Type

Browser based (Library Module)

License

Open Source (Apache 2.0)

Dataset

custom

Release History

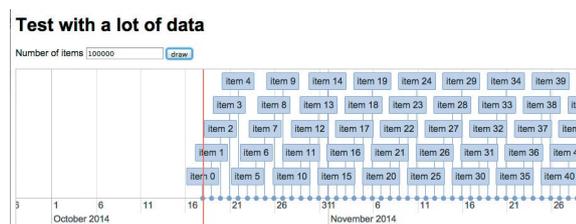
17.10.2014 (latest commit, ongoing)

2013 (first commit of vis.js Timeline)

2012 (first commit of predecessor CHAP Links Timeline)

Screencast

<https://www.youtube.com/watch?v=K4DhxuEUh-o>



Wellcome Timeline

Developer

Wellcome Trust

Website

<http://wellcomelibrary.org/using-the-library/subject-guides/genetics/makers-of-modern-genetics/genetics-timeline/#>

<https://github.com/wellcomelibrary/timeline>

Type

Browser based visualisation

License

MIT License

Dataset

at release: curated dataset of events related to the history of genetics
since 20.01.2015: Custom datasets possible

Release History

? (last updated 9.10.2014)

20.01.2015: 10.2.2013 (released on github)

Screencast

<https://www.youtube.com/watch?v=FVbzFIYFFYU>

wellcome library Contact Us JOIN LOGIN

THE HISTORY OF GENETICS: FROM DARWIN TO THE 21ST CENTURY: 1859 - 2012

1859 CHARLES DARWIN PUBLISHES 'ON THE ORIGIN OF...'

1866 GREGOR MENDEL PUBLISHES HIS THEORY OF...

1869 FRANCIS GALTON PUBLISHES 'HEREDITARY GENIUS'

1913 CARLOS BLAKEAR APPOINTED GENERAL SECRETARY...

1918 THE GENETICISTS' MANIFESTO

1943 HANS JONHESBERG PUBLISHES 'THE GENETICS OF THE...'

1953 FIRST HUMAN CHROMOSOMES SEQUENCED

2001 PRESENT DAY

2002 GRANT NEANDERTHAL GENOME PUBLISHED

1900 2000

International HapMap project launched

2002

An international project to map variation in 200-400 individuals from Nigeria, the USA, China and Japan was launched in October. The goal of the HapMap project was to extend the understanding of variation between individuals. It used the fact that single-letter variants in the DNA code, called single nucleotide polymorphisms, or SNPs, tend to group together in blocks known as haplotypes. Consequently, just a few SNP tags were needed to identify a haplotype. The HapMap documented variation in haplotype patterns among the different populations studied, a first step towards identifying genetic variation that was associated with disease.

More about the International HapMap project

D Documentation – TT

Timeline

Using the Timeline

Creating a New Timeline Container

Below is an example of how to create a basic timeline container. You need to specify the domain of the timeline to create an initial view. These values will be updated once a new layout containing data is added to the timeline.

```
timeline = TT.timeline()  
  .domain( [ new Date(1800, 0, 1), new Date() ] );
```

To render the timeline you'll need to assign it to a timeline container. In the example below, there is already a SVG element with the ID "timeline" present in the DOM.

```
d3.select( "svg#timeline" )  
  .attr( "width", $j(window).width() )  
  .attr( "height", $j(window).height() )  
  .call( timeline );
```

Adding Children

Layouts are added to the timeline in the following manner:

```
heap = TT.layout.heap().data( dataset );  
timeline.add( heap );
```

Reference

`timeline.add(TT.layout().*)`

Attach a layout to the timeline.

```
heap = TT.layout.heap().data( dataset );
timeline.add( heap );
```

`timeline.apply(:SVGElement)`

Creates the necessary DOM elements for the timeline container when called on an SVG element

```
timeline.apply( d3.select( "svg#timeline" ) );
```

`timeline.from(Date())`

Sets the lower bound of the timeline or returns it if called without an argument.

`timeline.domain([from:Date(), to:Date()])`

Sets or returns the domain of the timeline if called without arguments.

```
var from = new Date( 1900, 0, 1 );
var to = new Date ( 2000, 0, 1);
timeline.domain( [ from, to ] );
```

`timeline.height(:int)`

Returns the height of the timeline if called without arguments. Otherwise sets the height to the specified value and calls `.refresh()`

`timeline.refresh()`

Redraws the timeline and updates scales and axes.

`timeline.to(Date())`

Sets the upper bound of the timeline or returns it if called without an argument.

`timeline.view()`

Returns the view object of the timeline which contains the domain as well as the size in pixels. Usually gets called by the children of a timeline.

`timeline.width(:int)`

Returns the width of the timeline if called without arguments. Otherwise sets the width to the specified value and calls `.refresh()`

`timeline.x()`

Returns the horizontal scale used for zooming and panning the the timeline. Usually gets called by the children of a timeline.

`timeline.y()`

Returns the vertical scale used for zooming and panning the the timeline. Usually gets called by the children of a timeline.

TT.layout.heap()

Using the Heap Layout

Create a new heap layout by initialising it with a dataset and adding it to an existing timeline

```
heap = TT.layout.heap().data( dataset );
timeline.add( heap );
```

Reference

heap.data(Dataset:[])

If called with no arguments, returns the dataset. Otherwise replaces current dataset with new data and adjusts the domain of the parent timeline if necessary.

heap.identifier()

Returns a string which uniquely identifies the current heap and is used to generate id attributes for dependant DOM elements.

heap.initialise()

Is called by the timeline when the layout is added as a child.

heap.scales(Scales:[])

Sets or returns the scales object

heap.styles.events.(Name, Value)

Set the CSS values for certain attributes of the event element.

diameter: the diameter of the circles

```
heap.styles.events( "diameter", 4 ) // Sets diameter of
heap circles to 4
```

heap.images.(Name, Value)

Set the CSS values for certain attributes of the image element.

factor: the size of the images in relation to the zoom factor

```
heap.styles.images( "factor", 2 ) // Sets the size of the
images to 2 times the zoom factor
```

heap.threshold.(Name, Value)

Sets the threshold values which control the display of the heap

display: the maximum number of events which are displayed individually. Below this number, the heap is rendered as an outline

images: the zoom factor at which the images are loaded for the events

```
heap.threshold( "display", 1000) // Sets maximum amount  
of events displayed to 1000
```

TT.UI.panel()

The UI functions control the interaction with the layouts.

Usage:

```
ui = TT.ui.panel().heap( heap ).fields( fields )
    .initialise();
```

Reference

ui.fields([])

Accepts an array of objects, which describe the name of the fields in the panel as well as how they are accessed in the dataset. A basic field object looks like this:

```
[
  title: "Type",
  accessor: function(d) {
    return d.type;
  }
]
```

It is also possible to execute more complex operations in the accessor functions:

```
[
  title: "Age (years)",
  accessor: function(d) {
    return new Date( d.to.valueOf() -
      d.from.valueOf() ).getFullYear();
  }
]
```

ui.heap(:TT.layout.heap())

Attach the UI element to a heap.

ui.initialise()

Initialises the UI element after it has received a layout and an array of fields. It will then query the data from the layout and extract the unique values for every field.

E Documentation

– Temporal Perspectives

Temporalities()

Usage

The Multiple Temporalities layout is instantiated as follows:

```
var layout = new Temporalities();
```

A layout needs to be initialised with a flattened dataset:

```
layout.data ( dataset );
```

Dataset is expected to consist of a one dimensional array of JSON objects.

Perspectives can be added to a layout through the add() method:

```
var perspective = layout.add();
```

A number of methods are used to initialise a perspective. Below is an example that uses an attribute object_category as a viewpoint to create new events. They use object_category as their titles and will be positioned on the timeline by their attribute object_date. The methods scale(), width() and radius() control the mapping of the newly created events on the screen.

```
layout.caption( "Categories" )
  .nest( function( d ) {
    return d.object_category;
  } )
  .date( function( d ) {
    return d.object_date;
```

```

    } )

    .title(
      function( d ) {
        return d.object_category;
      } )
    .scale( d3.time.scale()
      .domain( [ new Date( 1800, 0, 1 ), new Date( 2016, 0,
1 ) ] ) )
    .range( [ 0, 1024 ] ) )
    .width( 1024 )
    .radius( [ 2, 10 ] );

```

Several perspectives can be added to a Multiple Temporalities layout in this manner. Finally, the data for the layout is generated by calling the `build()` method:

```
layout.build();
```

Reference

`Temporalities.add():Temporalities.set`

Initialises a new perspective (set) to be added to a Multiple Temporalities layout.

`Temporalities.build():[]`

Assembles the data for the Multiple Temporalities layout.

`Temporalities.set.caption(Accessor:())`

Defines the accessor that generates the caption for a Multiple Temporalities perspective.

`Temporalities.data(Dataset:[])`

If called with no argument, returns the dataset. Otherwise replaces current dataset with new data.

`Temporalities.set.date(Accessor:())`

Defines the accessor that returns the date that is used to position a generated event on the time axis.

`Temporalities.set.nest(Accessor:())`

Defines the accessor that is used to restructure a given dataset.

Temporalities.set.radius([MinRadius:int, MaxRadius:int])

Defines a maximum and minimum value for the radii of the generated events.

Temporalities.set.scale(Scale:d3.time.scale())

Defines the time scale of a perspective.

Temporalities.set.width(Width:int)

Sets or returns the width of a perspective

Bibliography

- accurat**, 2012. *Art Movie Time Machine (English)*. Available at: <http://visual.ly/art-movie-time-machine-english> [Accessed 7 November 2015].
- accurat**, 2014. *The Future, as Foretold in the Past*. Available at: <http://visual.ly/future-foretold-past> [Accessed 7 November 2015].
- Ackoff, R.**, 2010. From Data to Wisdom. *Journal of Applied Systems Analysis*, 16(1), pp. 3–9.
- Ahlberg, C. & Shneiderman, B.**, 1994. Visual information seeking: tight coupling of dynamic query filters with starfield displays. In B. Adelson, S. Dumais & J. Olson, eds., *Proceedings of the SIGCHI Conference on Human factors in Computing Systems*. Boston, MA: ACM, pp. 313–317.
- Aigner, W., Rind, A. & Hoffmann, S.**, 2012. Comparative Evaluation of an Interactive Time-Series Visualization that Combines Quantitative Data with Qualitative Abstractions. *Computer Graphics Forum*, 31(3pt2), pp. 995–1004.
- Alonso, O., Berberich, K., Bedathur, S. & Weikum, G.**, 2010. Time-based exploration of news archives. In *Proceedings of the Fourth Workshop on Human-Computer Interaction and Information Retrieval*. New Brunswick, 22 August 2010, pp. 12–15. Available at: <http://research.microsoft.com/en-us/um/people/ryenw/hcir2010/docs/H CIR-2010Proceedings.pdf> [Accessed 11 December 2012].
- Alonso, O., Gertz, M. & Baeza-Yates, R.**, 2009. Clustering and exploring search results using timeline constructions. In *Proceedings of the 18th ACM conference on Information and knowledge management*. New York: ACM, pp. 97–106.
- Alonso, O., Gertz, M. & Baeza-Yates, R.**, 2007. Search results using timeline visualizations. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*. New York: ACM, pp. 908–908.
- Anderson, M.**, 2007. Quantitative History. In W. Outwaite & S. Turner, eds., *The Sage Handbook of Social Science Methodology*. London: Sage.

- Andrews, W. & Parlapiano, A.**, 2014. *A History of the C.I.A.s Secret Interrogation Program*. The New York Times. New York: The New York Times Company. Available at: <http://www.nytimes.com/interactive/2014/12/09/world/timeline-of-cias-secret-interrogation-program.html> [Accessed 17 August 2015].
- André, P., Wilson, M.L. & schraefel, m c**, 2007. *Continuum: A Task-Based Evaluation of Temporal Visualisations*. Southampton: University of Southampton. Available at: <http://eprints.soton.ac.uk/264548/> [Accessed 6 February 2016]
- André, P., Wilson, M., Russell, A., Smith, D.A., Owens, A. & schraefel, m c**, 2007. Continuum: designing timelines for hierarchies, relationships and scale. In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology*. New York: ACM, pp.101-110.
- Antonelli, P.**, 2015. *@ at MoMA*. Inside/Out. New York: Museum of Modern Art. Available at: http://www.moma.org/explore/inside_out/2010/03/22/at-moma [Accessed 22 April 2015].
- Archer, L.B.**, 1995. The Nature of Research. *Co-design, interdisciplinary journal of design*, 2(11), pp.6-13.
- Archer, L.B.**, 1968. *The Structure Of Design Processes*. PhD thesis. London: Royal College of Art.
- Arnheim, R.**, 1969. *Visual Thinking*. Berkeley: University of California Press.
- Bachelard, G.**, 1963. *La dialectique de la durée*. Paris: Les Presses universitaires de France.
- Bailey, J. & Pregill, L.**, 2014. Speak to the Eyes: The History and Practice of Information Visualization. *Art Documentation*, 33(2), pp. 168-191. Available at: <http://www.jeffersonbailey.com/speak-to-the-eyes-the-history-and-practice-of-information-visualization/> [Accessed 19 September 2015].
- Barbeau de la Bruyère, J.L.**, 1750a. *Explication Générale de la Mappemonde Historique*. Paris: Hérissant. Bibliothèque municipale, Dijon. Fonds Ancien 12990 CGA.
- Barbeau de la Bruyère, J.L.**, 1750b. Mappemonde historique ou carte chronologique, géographique et généalogique des états et empires du monde. Paris: Ph. Buache. Available at: <http://gallica.bnf.fr/ark:/12148/btv1b72002870> [Accessed 5 February 2016]
- Barbeau-Dubourg, J.**, 1753. *Chronographie, ou Description des tems*. Paris: Bibliothèque nationale de France. Available at: <http://gallica.bnf.fr/ark:/12148/bpt6k1314025.r=Chronographie%2C+ou+Description+des+tems.langEN> [Accessed 23 March 2014].

- Barbour, J.**, 1999. *The End of Time: The Next Revolution in Our Understanding of the Universe*. London: Weidenfeld & Nicholson.
- Barr, A.**, 1938. *A diagram of synoptic relationships*. Edinburgh: T. & T. Clark
- Bateman, S. , Mandryk, R. L., Gutwin, C., Genest, A., McDine, D. & Brooks, C.**, 2010. Useful junk?: the effects of visual embellishment on comprehension and memorability of charts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York: ACM, pp. 2573-2582.
- Bates, M. J.**, 1989. The design of browsing and berrypicking techniques for the online search interface. *Online review*, 13(5), pp. 407-424.
- Baur, D., Seiffert, F., Sedlmair, M. & Boring, S.**, 2010. The Streams of Our Lives: Visualizing Listening Histories in Context. *IEEE Transactions on Visualization and Computer Graphics*, 16(6), pp.1119-1128.
- Bederson, B.B. & Shneiderman, B.**, 2003. *The Craft of Information Visualization*. Burlington: Morgan Kaufmann.
- Bederson, B.B., Clamage, A., Czerwinski, M. P. & Robertson, G. G.**, 2004. DateLens: A fisheye calendar interface for PDAs. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 11(1), pp. 90-119.
- Behn, R.**, 2009. The data don't speak for themselves. *Bob Behn's Management Report*, 6(7).
- Behrendt, A.**, 2011. Review of Cartographies of Time: A History of the Timeline. *Essays In History*. Available at: <http://www.essaysinhistory.com/review/2011/37> [Accessed 8 July 2015].
- Belopotosky, D., Kawakami, L. & Brustein, J.**, 2011. *Steven P. Jobs: His Life, His Companies, His Products*. The New York Times. New York: The New York Times Company. Available at: <http://www.nytimes.com/interactive/2011/10/05/business/20111005jobs-life-timeline.html> [Accessed 17 August 2015].
- Beltramin, P. & accurat**, 2012. *Gli imperi colpiscono ancora*. Available at: <http://visual.ly/empires-strike-again> [Accessed 14 August 2015].
- Benjamin, W.**, 1991. *Gesammelte Werke*. H. Schweppenhäuser & R. Tiedeman, eds. Frankfurt: Suhrkamp.
- Bergson, H.**, 1950. *Time and Free will: An Essay on the Immediate Data of Consciousness*. London: George Allen & Unwin Ltd.
- Berry, D.M.**, 2011. The computational turn: Thinking about the digital humanities. *Culture Machine*, 12(0), p.2.
- Bertin, J.**, 2010. *Semiology of Graphics: Diagrams, Networks, Maps*. Redlands: ESRI Press.

- Bizer, C., Heath, T. & Berners-Lee, T.**, 2009. Linked data: the story so far. *International Journal on Semantic Web and Information Systems*, 5(3), pp.1-22.
- Blight, G., Burn-Murdoch, J., Ball, J. & McCormick, M.**, 2011. *England riots: an interactive timeline*. The Guardian. London: Guardian News and Media ltd. Available at: <http://www.theguardian.com/uk/interactive/2011/sep/05/england-riots-timeline-interactive> [Accessed 18 August 2015].
- Blight, G., Pulham, S. & Torpey, P.**, 2012. *Arab spring: an interactive timeline of Middle East protests*. The Guardian. London: Guardian News and Media ltd. Available at: <http://www.theguardian.com/world/interactive/2011/mar/22/middle-east-protest-interactive-timeline> [Accessed 18 August 2015].
- Blinn, J.F.**, 1982. A generalization of algebraic surface drawing. *ACM Transactions on Graphics (TOG)*, 1(3), pp. 235-256
- Block, N.**, 1980. What is functionalism? In N. Block, ed., *Readings in Philosophy of Psychology*. Cambridge, MA: Harvard University Press.
- Boardman, R., Spence, R. & Sasse, M.A.**, 2003. Too many hierarchies? The daily struggle for control of the workspace. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York: ACM, pp. 616-620.
- Bogue, A.G.**, 1983. *Clio & the Bitch Goddess: Quantification in American Political History*. Beverly Hills: SAGE Publications.
- Boonstra, O., Breure, L. & Doorn, P.**, 2004. Past, present and future of historical information science. *Historical Social Research/Historische Sozialforschung*, 29(2), pp. 4-132.
- Borgman, C.L.**, 2009. The Digital Future is Now: A Call to Action for the Humanities. *Digital Humanities Quarterly*, 3(4).
- Borkin, M.A., Vo, A.A., Bylinskii, Z., Isola, P., Sunkavalli, S., Oliva, A. & Pfister, H.**, 2013. What Makes a Visualization Memorable? *IEEE Transactions on Visualization and Computer Graphics*, 19(12), pp.2306-2315.
- Boyd Davis, S.**, 2015a. Beholder of All Ages: The History of the World in a French Mappemonde. *Textimage: revue d'étude du dialogue text-image*, 7(1). Available at: http://revue-textimage.com/11_illustration_sci-ence/boyd-davis1.html [Accessed 2 July 2015].

- Boyd Davis, S.**, 2011. Joseph Priestley: The Man Who Drew Time. *Postings from the Priestley House*, 49. Available at: http://www.joseph-priestleyhouse.org/wp-content/uploads/The_Man_Who_Drew_Time.pdf [Accessed 17 September 2012].
- Boyd Davis, S.**, 2009. Mapping the unseen: making sense of the subjective image. In C. Nold, ed., *Emotional Cartography: Technologies of the Self*. London: Emotional Cartography. Available at: <http://www.emotionalcartography.net/EmotionalCartographyLow.pdf> [Accessed 4 February 2016].
- Boyd Davis, S.**, 2015b. May Not Duration Be Represented as Distinctly as Space? Geography and the Visualization of Time in the Early Eighteenth Century. In D. Beck, ed., *Knowing nature in early modern europe*. London: Pickering & Chatto.
- Boyd Davis, S.**, 2016a. Time Machines. In J. Pilcher, ed., *Culture, technology and the image*. Bristol: Intellect. In Press.
- Boyd Davis, S.**, 2016b. "To see at one glance all the centuries that have passed" - early visualisations of historical time. In A. Black, O. Lund, & S. Walker, eds., *Gower handbook of information design*. Farnham: Ashgate. In Press.
- Boyd Davis, S. & Kräutli, F.**, 2014. Scholarly chronographics: can a timeline be useful in historiography? *European Social Science and History Conference*. Vienna, 23-26 April. Amsterdam: International Institute of Social History
- Boyd Davis, S. & Kräutli, F.**, 2015. The Idea and Image of Historical Time: Interactions between Design and Digital Humanities. *Visible Language*, 49(3), pp.101-119. Available at: <http://visiblelanguagejournal.com/issue/172/article/1202> [Accessed 5 February 2016].
- Boyd Davis, S., Bevan, E. & Kudikov, A.**, 2013. Just in Time: Defining Historical Chronographics. In P. Bowen, S. Keene & K. Ng, eds., *Electronic Visualisation in Arts and Culture*. London: Springer, pp. 243-257.
- Brenner, S.**, 2015. Reconsidering searching and browsing on the Cooper Hewitts Collections website 2015. *MW2015: Museums and the Web*. Available at: <http://mw2015.museumsandtheweb.com/paper/reconsidering-searching-and-browsing-on-the-cooper-hewitts-collections-website/> [Accessed 29 March 2015].
- Bridenbaugh, C.**, 1963. The Great Mutation. *American Historical Review*, 68(2), pp.315-331.

- Brinton, W.C.**, 1919. *Graphic methods for presenting facts*. New York: The Engineering Magazine Company. Available at: <https://archive.org/download/graphicmethodsfoobrnrch/graphicmethodsfoobrnrch.pdf> [Accessed 18 January 2016].
- Brodbeck, D. & Girardin, L.**, 2003. Trend Analysis in Large Timeseries of High-Throughput Screening Data Using a Distortion-Oriented Lens with Semantic Zooming. *IEEE Symposium on Information Visualization*. Seattle, 19-21 October. New York: IEEE
- Buckland, M.K.**, 1991. Information as thing. *Journal of the American Society for Information Science*, 42(5), p.351.
- Burch, M., Beck, F. & Diehl, S.**, 2008. Timeline trees: visualizing sequences of transactions in information hierarchies. In *Proceedings of the Working Conference on Advanced Visual Interfaces*. New York: ACM, pp. 75-82.
- Burdick, A.**, 2009. Design without designers. *Keynote for a Conference on the Future of Art and Design Education in the 21st century*. Brighton: University of Brighton. Available at: http://anneburdick.com/Design-wo-Designers/Burdick_Design_wo_Designers_sm.pdf [Accessed 29 September 2015].
- Busa, R.**, 1980. The annals of humanities computing: The index Thomisticus. *Computers and the Humanities*, 14(2), pp.83-90.
- Byron, L. & Wattenberg, M.**, 2008. Stacked Graphs - Geometry & Aesthetics. *IEEE Transactions on Visualization and Computer Graphics*, 14(6).
- Cairo, A.**, 2012. *Reclaiming the word "infographics"*. The Functional Art. Available at: <http://www.thefunctionalart.com/2012/12/claiming-word-infographics-back.html> [Accessed 13 August 2015].
- Card, S.K., Mackinlay, J.D. & Shneiderman, B., eds.**, 1999. *Readings in Information Visualization: Using Vision to Think*. San Francisco: Morgan Kaufmann Publishers.
- Card, S.K., Suh, B., Pendleton, B., Heer, J. & Bodnar, J.W.**, 2006. Time-Tree: Exploring Time Changing Hierarchies. In *IEEE Symposium On Visual Analytics Science And Technology (VAST)*. New York: IEEE, pp. 3-10.
- Caviglia, G.**, 2013. *The Design of Heuristic Practices*. PhD thesis. Milan: Politecnico di Milano Dipartimento di Design.
- Cellan-Jones, R.**, 2011. *Facebook focuses on media sharing and adds timeline*. BBC News. London: BBC. Available at: <http://www.bbc.co.uk/news/technology-15028920> [Accessed 9 December 2013].

- CEN**, 2004. *European Guide to good Practice in Knowledge Management Part 1: Knowledge Management Framework*. Brussels: European Committee for Standardization. Available at: <ftp://cenftp1.cenorm.be/PUBLIC/CWAs/e-Europe/KM/CWA14924-01-2004-Mar.pdf> [Accessed 28 October 2013].
- Chang, D., Ge, Y., Song, S. & Coleman, N.**, 2009. *Visualizing the Republic of Letters*. Stanford Humanities Center. Stanford, CA: Stanford University. Available at http://web.stanford.edu/group/toolingup/rplviz/papers/Vis_RofL_2009 [Accessed 6 February 2016].
- Chenhall, R.G.**, 1975. *Museum Collections and Today's Computers*. Nashville, TN: American Association for State & Local History
- Chittaro, L. & Combi, C.**, 2001. Representation of temporal intervals and relations: Information visualization aspects and their evaluation. In *Proceedings of the Eighth International Symposium on Temporal Representation and Reasoning*. New York: IEEE, pp.13-20.
- Chuah, M.C.**, 1998. Dynamic aggregation with circular visual designs. In *Proceedings of IEEE symposium on Information Visualization*. New York: IEEE, pp. 35-43.
- Clark, N. & Bilefsky, D.**, 2015. *Germanwings Crash in French Alps*. The New York Times. New York: The New York Times Company. Available at: <http://www.nytimes.com/2015/03/25/world/europe/germanwings-crash.html> [Accessed 17 August 2015].
- Cleveland, W.S. & McGill, R.**, 1984. Graphical perception: Theory, experimentation, and application to the development of graphical methods. *Journal of the American Statistical Association*, 79(387), pp.531-554.
- Cockburn, A. & McKenzie, B.**, 2002. Evaluating the effectiveness of spatial memory in 2D and 3D physical and virtual environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York: ACM, pp. 203-210.
- Cousins, S.B. & Kahn, M.G.**, 1991. The visual display of temporal information. *Artificial intelligence in medicine*, 3(6), pp.341-357.
- Conklin, J.**, 1987. A survey of hypertext. In *MCC Technical Report #STP-356-86*. Available at: http://iicm.tugraz.at/thesis/cguetl_diss/literatur/Kapitel04/References/Conklin1987/Conklin87.html [Accessed 4 February 2016].

- Craft, B. & Cairns, P.**, 2005. Beyond Guidelines: What Can We Learn from the Visual Information Seeking Mantra? In *Proceedings of the Ninth International Conference on Information Visualisation*. New York: IEEE, pp. 110-118
- Croxtton, F.E. & Stryker, R.E.**, 1927. Bar Charts versus Circle Diagrams. *Journal of the American Statistical Association*, 22(160), pp.473-482.
- Culler, J.**, 2010. The closeness of close reading. *ADE Bulletin*, 149, pp.20-25. Available at: <https://ade.mla.org/bulletin/article/ade.149.20> [Accessed 6 February 2016].
- Currall, J., Moss, M. & Stuart, S.**, 2005. What is a collection. *Archivaria*, 58, Fall Issue, pp.131-146.
- Davenport, J.R.A.**, 2013. *The Dimensions of Art*. If We Assume. Available at: <http://www.ifweassume.com/2013/11/the-dimensions-of-art.html> [Accessed 19 September 2015].
- Dawson, A. & Hillhouse, S. eds.**, 2011. *SPECTRUM 4.0*. London: Collections Trust.
- DeFanti, T.A., Brown, M.D. & McCormick, B.H.**, 1989. Visualization: expanding scientific and engineering research opportunities. *Computer*, 22(8), pp.12-16.
- Delvisicio, J., Fisher, R., Allert, M. & Sinha, S.**, 2013. *Timeline of Nelson Mandela: 1918-2013*. The New York Times. New York: The New York Times Company. Available at: <http://www.nytimes.com/interactive/2013/12/05/world/africa/Mandela-Timeline.html> [Accessed 17 August 2015].
- Dershowitz, N. & Reingold, E.M.**, 2008. *Calendrical Calculations*. Cambridge, MA: Cambridge University Press.
- Descartes, R.**, 1996. *Descartes: Meditations on First Philosophy*. Cambridge, MA: Cambridge University Press.
- Dockterman, E.**, 2016. *This Is Why Facebook Thinks You Have 46-Year Friendships*. TIME. New York: Time Inc. Available at: <http://time.com/4165241/facebook-friends-glitch-46-years/> [Accessed 5 January 2016].
- Dorst, K.**, 2003. The problem of design problems. *Expertise in Design: Design Thinking Research Symposium*. Sydney, 17-19 November. Available at: <http://research.it.uts.edu.au/creative/design/papers/23DorstDTRS6.pdf> [Accessed 6 February 2016].

- Dörk, M.**, 2012. *Visualization for Search: Exploring Complex and Dynamic Information Spaces*. PhD thesis. Calgary: University of Calgary. Available at: <http://mariandoerk.de/phd/dissertation.pdf> [Accessed 2 May 2013].
- Drass, E.**, 2013. *Exploring The Tate Collection*. Shardcore. Available at: <http://www.shardcore.org/shardpress/2013/11/06/tate-data-explorer/> [Accessed 19 September 2015].
- Drucker, J.**, 2011a. *Graphesis: Visual knowledge production and representation*. Available at: http://www.johannadrucker.com/pdf/graphesis_2011.pdf [Accessed 25 November 2013].
- Drucker, J.**, 2011b. Humanities approaches to graphical display. *Digital Humanities Quarterly*, 5(1). Available at: http://www.johannadrucker.com/pdf/hum_app.pdf [Accessed 22 November 2013].
- Drucker, J. and Nowviskie, B.P.**, 2003. Temporal Modelling: Conceptualization and Visualization of Temporal Relations for Humanities Scholarship. Available at: <http://www2.iath.virginia.edu/time/reports/infodesign.doc> [Accessed 8 October 2012].
- Duchain, M.**, 1983. Theoretical Principles and Practical Problems of Respect des fonds in Archival Science. *Archivaria*, 1(16).
- Eells, W.C.**, 1926. The Relative Merits of Circles and Bars for Representing Component Parts. *Journal of the American Statistical Association*, 21(154), pp.119-132.
- Ellis, G.P. & Dix, A.J.**, 2007. A Taxonomy of Clutter Reduction for Information Visualisation. *IEEE Transactions on Visualization and Computer Graphics*, 13(6), pp.1216-1223.
- Ennis Butler, B.**, 2013. Visual Exploration of Australian Prints and Printmaking. *MW2013: Museums and the Web 2013*. Available at: <http://mw2013.museumsandtheweb.com/paper/visual-exploration-of-australian-prints-and-printmaking/> [Accessed 14 March 2013].
- Europeana**, 2014. *Provider FAQs*. Den Haag: Europeana. Available at: <http://pro.europeana.eu/page/provider-faqs> [Accessed 21 November 2015].
- europeana labs**, *Europeana Linked Open Data*. Den Haag: Europeana. Available at: <http://labs.europeana.eu/api/linked-open-data-introduction> [Accessed 5 November 2015].
- Farquhar, A.B. & Farquhar, H.**, 1891. *Economic and Industrial Delusions*. New York & London: G. P. Putnam's Sons.
- Feeney, D.**, 2007. *Caesar's Calendar: Ancient Time and the Beginnings of History*. Berkeley: University of California Press.

- Few, S.**, 2011. *Are Infovis and Statistical Graphics Really All That Different?* El Dorado Hills: Perceptual Edge. Available at: http://www.perceptualedge.com/files/are_infovis_and_statistical_graphics_really_all_that_different.pdf [Accessed 13 August 2015].
- Few, S.**, 2013. Data Visualization for Human Perception. In M. Soegaard & R. F. Dam, eds., *The encyclopedia of human computer interaction, 2nd ed.* Aarhus: Interaction Design Foundation. Available at: <https://www.interaction-design.org/literature/book/the-encyclopedia-of-human-computer-interaction-2nd-ed/data-visualization-for-human-perception> [Accessed 6 February 2016].
- Few, S.**, 2007. *Save the pies for dessert.* Visual Business Intelligence Newsletter. El Dorado Hills: Perceptual Edge. Available at: <https://courses.washington.edu/info424/2007/readings/Save%20the%20Pies%20for%20Dessert.pdf> [Accessed 18 August 2015].
- Flanders, J. & Muñoz, T.**, 2011. *An introduction to humanities data curation.* DH Curation Guide. Available at: <http://guide.dhcuration.org/intro/> [Accessed 26 June 2015].
- Flowing Media**, 2010a. *Data.* Available at: <https://github.com/Flowing-Media/TimeFlow> [Accessed 20 August 2015].
- Flowing Media**, 2010b. *Timeline View.* Available at: <https://github.com/FlowingMedia/TimeFlow> [Accessed 20 August 2015].
- Fodor, J.**, 1987. *Psychosemantics: The problem of meaning in the philosophy of mind.* Cambridge, MA: MIT Press.
- Fouse, A., Weibel, N., Hutchins, E. & Hollan, J.D.**, 2011. ChronoViz: a system for supporting navigation of time-coded data. In *CHI Extended Abstracts on Human Factors in Computing Systems*. New York: ACM, pp. 299–304.
- Frayling, C.**, 1993. Research in art and design. *Royal College of Art Research Papers*, 1(1), pp.1–5. Available at: http://researchonline.rca.ac.uk/384/3/frayling_research_in_art_and_design_1993.pdf [Accessed 17 September 2012].
- Fredrikson, A., North, C., Plaisant, C. and Shneiderman, B.**, 1999. Temporal, geographical and categorical aggregations viewed through coordinated displays: a case study with highway incident data. In *Proceedings of the Workshop on New Paradigms in Information Visualization and Manipulation in Conjunction with the Eighth ACM International Conference on Information and Knowledge Management*. New York: ACM, pp. 26–34.

- Freeman, E. & Fertig, S.**, 1995. Lifestreams: Organizing your electronic life. In *Proceedings of the AAAI Fall Symposium on AI Applications in Knowledge Navigation and Retrieval*. Palo Alto: The Association for the Advancement of Artificial Intelligence, pp.38-44.
- Freeman, E. & Gelernter, D.**, 1996. Lifestreams: a storage model for personal data. *ACM SIGMOD Record*, 25(1), pp. 80-86.
- Friendly, M.**, 2006. A brief history of data visualization. In C. Chen, W. Härdle, & A. Unwin, eds., *Handbook of Data Visualization*. Berlin/Heidelberg: Springer.
- Friendly, M.**, 2009. *Milestones in the history of thematic cartography, statistical graphics, and data visualization*. Available at: <http://euclid.psych.yorku.ca/SCS/Gallery/milestone/milestone.pdf> [Accessed 30 March 2015].
- Fruchterman, T. & Reingold, E.M.**, 1991. Graph drawing by force-directed placement. *Software-Practice and Experience*, 21(11), pp.1129-1164.
- Fua, Y.-H., Ward, M.O. & Rundensteiner, E.A.**, 1999. Hierarchical parallel coordinates for exploration of large datasets. In *Proceedings of Visualization '99*. New York: IEEE, pp.43-50.
- Fulda, J., Brehmer, M. & Munzner, T.**, 2016. TimeLineCurator: Interactive Authoring of Visual Timelines from Unstructured Text. *IEEE Transactions on Visualization and Computer Graphics*, 22(1), pp.300-309.
- Ferguson, S.**, 1991. *The 1753 Carte Chronographique of Jacques Barbeu-Dubourg*. Princeton, NJ: Princeton University Library Chronicle. Available at: http://www.princeton.edu/~ferguson/PULC_1991_duBourg.pdf [Accessed 4 April 2013].
- Furnas, G.W.**, 1986. Generalized fisheye views. *ACM SIGCHI Bulletin - Special issue: CHI '86 Conference Proceedings*, 17(4), pp.16-23.
- Furnas, G.W.**, 1999. The FISHEYE view: a new look at structured files. In S.K Card, J.D. Mackinlay & B. Shneiderman, eds., *Readings in Information Visualization: Using Vision to Think*. San Francisco: Morgan Kaufmann Publishers.
- Gantt, H.L.**, 1919. *Work, wages, and profits* 2nd ed., New York: The Engineering Magazine Co. Available at: <https://archive.org/details/work-wagesprofitsoogant> [Accessed 14 August 2015].
- Geertz, C.**, 1973. Thick Description: Toward an Interpretive Theory of Culture. In C. Geertz, ed., *The Interpretation of Cultures Selected Essays*. New York: Basic Books, pp. 3-30.

- Gelman, A. & Unwin, A.**, 2013. Infovis and statistical graphics: different goals, different looks. *Journal of Computational and Graphical Statistics*, 22(1), pp.2-28.
- Gershon, N. & Eick, S.G.**, 1997. Information Visualization. *IEEE Computer Graphics and Applications*, 17(4), pp.29-31.
- Gibbs, F.W. & Cohen, D.J.**, 2011. A Conversation with Data: Prospecting Victorian Words and Ideas. *Victorian Studies*, 54(1), pp.69-77.
- Gibson, J.J.**, 1977. The Theory of Affordances. In R. Shaw & J. Bransford, eds., *Perceiving, Acting, and Knowing: Toward an Ecological Psychology*. Hillsdale: Lawrence Erlbaum.
- Golbeck, J. & Mutton, P.**, 2006. Spring-Embedded Graphs for Semantic Visualization. In V. Geroimenko & C. Chen, eds., *Visualizing the Semantic Web*. London: Springer, pp.172-182.
- Gold, M.K., ed.**, 2012. *Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press.
- Goldman, K.H. & Wadman, M.**, 2002. "There's something happening here, what it is ain't exactly clear". In *Proceedings of Museums and The Web*. Toronto: Archives & Museum Informatics. Available at: <http://www.museumsandtheweb.com/mw2002/papers/haleyGoldman/haleygoldman.html> [Accessed 1 June 2015].
- Goldstein, J. & Roth, S.F.**, 1994. Using aggregation and dynamic queries for exploring large data sets. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York: ACM, pp.23-29.
- Gomez, P.U. & Keller, P.**, 2015. *The missing decades: the 20th century black hole in Europeana*. Den Haag: Europeana. Available at: <http://pro.europeana.eu/blogpost/the-missing-decades-the-20th-century-black-hole-in-europeana> [Accessed 21 November 2015].
- Gorgels, P.**, 2013. Rijksstudio: Make Your Own Masterpiece! *MW2013: Museums and the Web 2013*. Portland, OR, 17-20 April. Silver Spring: Museums and the Web LLC. Available at: <http://museumsandtheweb.com/paper/rijksstudio-makeyour-own-masterpiece> [Accessed 24 May 2013].
- Gregory, S.A.**, 1966. Design and the Design Method. In S.A. Gregory, ed., *The Design Method*. New York: Springer.
- Harper, R., Lindley, S., Thereska, E., Banks, R., Gosset, P., Smyth, G., Odom, W. & Whitworth, E.**, 2013. What is a File? In *Proceedings of the 2013 Conference on Computer Supported Cooperative Work*. New York: ACM, pp.1125-1136.

- Harris, J.**, 2011. *Word clouds considered harmful*. NiemanLab. Cambridge, MA: Nieman Foundation at Harvard University. Available at: <http://www.niemanlab.org/2011/10/word-clouds-considered-harmful/> [Accessed 19 October 2015].
- Havre, S., Hetzler, E., Whitney, P. & Nowell, L.**, 2002. Themeriver: Visualizing thematic changes in large document collections. *IEEE Transactions on Visualization and Computer Graphics*, 8(1), pp.9-20.
- Hayles, N.K.**, 2012. *How We Think: Digital Media and Contemporary Technogenesis*. Chicago, IL: University of Chicago Press.
- Hedstrom, M.**, 2002. Archives, Memory, and Interfaces with the Past. *Archival Science*, 2(1-2), pp.21-43.
- Hegel, G.W.F.**, 2001. *The Philosophy of History*. Kitchener: Batoche Books.
- Heller, J.**, 2013. GRIPHOS (*General Retrieval and Information Processor for Humanities Oriented Studies*). TAPoR 3. Available at: <http://tapor-test.artsrn.ualberta.ca/tools/237> [Accessed 15 September 2015].
- Hensher, P.**, 2009. *Benjamin Britten and WH Auden*. The Guardian. London: Guardian News and Media ltd. Available at: <http://www.theguardian.com/culture/2009/nov/07/britten-auden-philip-hensher> [Accessed 4 November 2015].
- Hicks, M., Nicholls, S., O'Malley, C. & Anderson, B.**, 2003. Comparison of 2D & 3D Representations for Visualising Telecommunication Usage. *Behaviour & Information Technology*, 22(3), pp.185-201.
- Hinrichs, U., Schmidt, H. & Carpendale, S.**, 2008. EMDialog: Bringing Information Visualization into the Museum. *IEEE Transactions on Visualization and Computer Graphics*, 14(6), pp.1181-1188.
- Hockey, S.**, 2004. The History of Humanities Computing. In S. Schreibman, R. Siemens & J. Unsworth, eds., *A companion to digital humanities*. Oxford: Blackwell, pp. 3-29.
- Hoek, W. van & Mayr, P.**, 2014. *Is Evaluating Visual Search Interfaces in Digital Libraries Still an Issue?* ArXiv e-prints. Ithaca, NY: Cornell University Library. Available at: <http://arxiv.org/abs/1408.5001> [Accessed 6 February 2016].
- Hoffmann, S., Aigner, W., Rind, A., Miksch, S. & Turic, T.**, 2012. *Visualization Technique for Time-Oriented Data with Semantic Zoom*. Vienna: Vienna University of Technology Faculty of Informatics, Institute of Software Technology & Interactive Systems. Available at: <http://www.cvast.tuwien.ac.at/projects/semtimezoom> [Accessed 20 March 2014].

- Huff, D.**, 2010. *How to Lie with Statistics*. New York: W.W. Norton & Company.
- Hughes-Warrington, M.**, 2008. *Fifty Key Thinkers on History*. London & New York: Routledge.
- Hullman, J., Adar, E. & Shah, P.**, 2011. Benefitting infovis with visual difficulties. *IEEE Transactions on Visualization and Computer Graphics*, 17(12), pp.2213-2222.
- Huynh, D.F.**, 2009. *SIMILE Widgets | Timeline*. SIMILE Widgets. Cambridge MA: Massachusetts Institute of Technology. Available at: <http://simile-widgets.org/timeline/> [Accessed 17 January 2013].
- Huynh, D.F., Drucker, M. S., Baudisch, P. & Wong, C.**, 2005. Time quilt: scaling up zoomable photo browsers for large, unstructured photo collections. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems*. New York: ACM, pp. 1937-1940.
- IBM**, 2007. *Many Eyes*. New York: IBM. Available at: <http://www-01.ibm.com/software/analytics/many-eyes/> [Accessed 20 August 2015].
- Inbar, O., Tractinsky, N. & Meyer, J.**, 2007. Minimalism in information visualization: attitudes towards maximizing the data-ink ratio. In *Proceedings of the 14th European conference on Cognitive ergonomics: invent! explore!*. New York: ACM.
- Isaacson, W.**, 2012. *How Steve Jobs' Love of Simplicity Fueled A Design Revolution*. Smithsonian Magazine. Washington, DC: Smithsonian Institution. Available at: <http://www.smithsonianmag.com/arts-culture/how-steve-jobs-love-of-simplicity-fueled-a-design-revolution-23868877/> [Accessed 9 December 2015].
- Jenkins, N.**, 2013. *Originating Kindred Britain*. Kindred Britain. Stanford: Stanford University. Available at: <http://kindred.stanford.edu/notes.html?section=originating> [Accessed 24 November 2015].
- Jenkins, N., Meeks, E. & Murray, S.**, 2013. *Kindred Britain*. Stanford: Stanford University. Available at: <http://kindred.stanford.edu/> [Accessed 19 August 2015].
- Jenkinson, H.**, 1922. *A Manual of Archive Administration Including the Problems of War Archive Making*. Oxford: The Clarendon Press. Available at: <https://archive.org/details/manualofarchiveaoojenkuoft> [Accessed 6 February 2016].
- Jensen, M.**, 2003. *Visualizing complex semantic timelines*. NewsBlip Technical Report NBTR2003-001. Seattle, WA: NewsBlip. Available at <http://www.newsblip.com/tr/nbtr2003001.pdf> [Accessed 8 October 2012].

- Jessop, M.**, 2008. Digital visualization as a scholarly activity. *Literary and Linguistic Computing*, 23(3), pp.281-293.
- Kasik, D.J., Ebert, D., Lebanon, G., Park, H. & Pottenger, W.M.**, 2009. Data transformations and representations for computation and visualization. *Information Visualization*, 8(4), pp.275-285.
- Kauppinen, T., Mantegari, G. & Paakkari, P.**, 2010. Determining relevance of imprecise temporal intervals for cultural heritage information retrieval. *International Journal of Human-Computer Studies*, 68(9), pp.549-560.
- Keim, D.A.**, 2001. Visual exploration of large data sets. *Communications of the ACM*, 44(8), pp.38-44.
- Kerren, A., Stasko, J.T., Fekete, J.-D. & North, C. eds.**, 2008. *Information Visualization: Human-Centered Issues and Perspectives*. Berlin/Heidelberg: Springer.
- Kiritchenko, S., Matwin, S. & Abu-Hakima, S.**, 2004. Email classification with temporal features. In M. Kłopotek, S. Wierzchoń, & K. Trojanowski, eds., *Intelligent information processing and web mining*. Berlin/Heidelberg: Springer, pp. 523-533.
- Kirschenbaum, M.G.**, 2012. What Is Digital humanities and What's It Doing in English Departments? In M. K. Gold, ed. *Debates in the Digital Humanities*. Minneapolis: University of Minnesota Press.
- Kirsh, D. & Maglio, P.**, 1994. On Distinguishing Epistemic from Pragmatic Action. *Cognitive science*, 18(4), pp.513-549.
- Korallo, L., Foreman, N., Boyd Davis, S., Moar, M. & Coulson, M.**, 2012. Can multiple spatial virtual timelines convey the relatedness of chronological knowledge across parallel domains? *Computers & Education*, 58(2), pp.856-862.
- Kosara, R.**, 2013. InfoVis is so much more: a comment on Gelman and Unwin and an invitation to consider the opportunities. *Journal of Computational and Graphical Statistics*, 22(1), pp.29-32.
- Kosara, R.**, 2010. *The Difference Between Infographics and Visualization*. Eagereyes. Available at: <https://eagereyes.org/blog/2010/the-difference-between-infographics-and-visualization> [Accessed 13 August 2015].
- Kosara, R.**, 2007. Visualization Criticism - The Missing Link Between Information Visualization and Art. In *11th International Conference on Information Visualization*. New York: IEEE, pp. 631-636.

- Kosara, R., Bendix, F. & Hauser, H.**, 2004. Timehistograms for large, time-dependent data. In *Proceedings of the Sixth Joint Eurographics - IEEE TCVG conference on Visualization*. Aire-la-Ville: The Eurographics Association, pp. 45-54.
- Kousser, J.M.**, 1980. Quantitative social-scientific history. In M. Kammen, ed., *The Past Before Us: Contemporary Historical Writing in the United States*. Ithaca, NY: Cornell University Press, pp. 433-456.
- Kramer, M.J.**, 2014. Going Meta on Metadata. *Journal of Digital Humanities*, 3(2). Available at: <http://journalofdigitalhumanities.org/3-2/going-meta-on-metadata/> [Accessed 5 January 2016].
- Kräutli, F.**, 2014a. *Behind the Scenes: ChronoZoom*. Available at: <http://research.kraeutli.com/index.php/2014/01/behind-the-scenes-chronozoom/> [Accessed 10 October 2015].
- Kräutli, F.**, 2012. *SFMOMA ArtScope - Exploring Digital Collections*. Available at: <http://research.kraeutli.com/index.php/2012/12/sfmoma-artscope/> [Accessed 20 November 2015].
- Kräutli, F.**, 2013. *The Tate Collection on GitHub*. Available at: <http://research.kraeutli.com/index.php/2013/11/the-tate-collection-on-github/> [Accessed 19 September 2015].
- Kräutli, F.**, 2014b. *Tree of Time: ChronoZoom*. Available at: <http://research.kraeutli.com/index.php/2014/01/tree-of-time-chronozoom/> [Accessed 10 October 2015].
- Kräutli, F. & Boyd Davis, S.**, 2013. Known unknowns: representing uncertainty in historical time. In K. Ng, J. P. Bowen, & S. McDaid, eds., *Proceedings of EVA London*. London: BCS Learning and Development Ltd, pp. 61-68.
- Krzywinski, M., Birol, I. & Jones, S.**, 2012. Hive plots - rational approach to visualizing networks. *Briefings in bioinformatics*, 13(5).
- Kullberg, R.L. & Mitchell, W.J.**, 1995. *Dynamic Timelines-Visualizing Historical Information in Three Dimensions*. PhD thesis. Cambridge, MA: Massachusetts Institute of Technology
- Lamping, J. & Rao, R.**, 1996. The hyperbolic browser: A focus + context technique for visualizing large hierarchies. *Journal of visual languages and computing*, 7(1), pp.33-55.
- Lansdale, M.W.**, 1988. The psychology of personal information management. *Applied ergonomics*, 19(1), pp.55-66.
- Larkin, J.H. & Simon, H.A.**, 1987. Why a diagram is (sometimes) worth ten thousand words. *Cognitive science*, 11(1), pp.65-100.

- Lee, E.**, 2011. *History's Largest Empires*. Available at: <http://visualising.org/visualizations/historys-largest-empires> [Accessed 14 August 2015].
- Lee, H.-L.**, 2005. The concept of collection from the user's perspective. *Library Quarterly*, 75(1).
- Lee, H.-L.**, 2000. What is a collection? *Journal of the American Society for Information Science*, 51(12), pp.1106-1113.
- Lehni, J. & Puckey, J.**, 2011. *Paper.js* Available at: <http://paperjs.org/about/> [Accessed 10 May 2013].
- Lejeune, B.**, 2009. The Effects of Online Catalogues in London and other Museums: A Study of an Alternative Way of Access. *Papers from the Institute of Archaeology*, 18(S1), pp.79-97.
- Levy, E., Zacks, J., Tversky, B. & Schiano, D.**, 1996. Gratuitous graphics? Putting preferences in perspective. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York: ACM, pp. 42-49.
- Lewis, C.**, 1982. *Using the thinking-aloud method in cognitive interface design*. Report RC 9265. New York: IBM Research.
- Lewis, J.R.**, 2006. Usability testing. In G. Salvendy, ed., *Handbook of Human Factors and Ergonomics*. New York: John Wiley & Sons, pp. 1275-1316.
- Llorens, H., Saquete, E., Navarro, B. & Gaizauskas, R.**, 2011. Time-surfer: time-based graphical access to document content. In Clough, P., Foley, C., Gurrin, C., Jones, G. J. F., Kraaij, W., Lee, H. & Mudoch, V., eds., *Advances in Information Retrieval*. Berlin/Heidelberg: Springer, pp. 767-771.
- Lloyd, C. & Forshaw, A.**, 2011. *The What on Earth? Wallbook of Big History*, Tonbridge, UK: What On Earth Foundation.
- Locke, J.**, 1693. *Some Thoughts Concerning Education*, London: Churchill.
- Lubar, S.**, 2013. Timelines in Exhibitions. *Curator: The Museum Journal*, 56(2).
- Lunenfeld, P., Burdick, A., Drucker, J., Presner, T. & Schnapp, J.**, 2012. *Digital Humanities*, Cambridge MA: MIT Press. Available at: http://mitpress.mit.edu/sites/default/files/titles/content/9780262018470_Open_Access_Edition.pdf [Accessed 26 November 2012].
- Lynch, C.**, 2002. Digital collections, digital libraries & the digitization of cultural heritage information. *Microform & imaging review*, 31(4), pp.131-145.

- Machlup, F.**, 1983. Semantic quirks in studies of information. In F. Machlup & U. Mansfield, eds., *The study of information interdisciplinary messages*. New York: John Wiley & Sons, pp. 641-671.
- Mackinlay, J.D., Robertson, G.G. & Card, S.K.**, 1991. The perspective wall: detail and context smoothly integrated. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York: ACM, pp. 173-176.
- Malone, T.W.**, 1983. How do people organize their desks?: Implications for the design of office information systems. *ACM Transactions on Information Systems (TOIS)*, 1(1), pp.99-112.
- Manovich, L.**, 2001. *The Language of New Media*. Cambridge, MA: MIT Press.
- Manovich, L.**, 2011. What is visualisation? *Visual Studies*, 26(1), pp.36-49.
- Marche, S.**, 2012. *Literature is not Data: Against Digital Humanities*. Los Angeles: Los Angeles Review of Books. Available at: <https://lareviewofbooks.org/essay/literature-is-not-data-against-digital-humanities> [Accessed 8 January 2016].
- Marsh, E.R.**, 1974. The Harmonogram of Karol Adamiecki. *Academy of Management Proceedings*, 18(2), pp.358-364.
- Martin, R.E.**, 2010. *Kairoscope: coordinating time socially*. PhD thesis. Cambridge, MA: Massachusetts Institute of Technology.
- Martin, S., Brown, W.M., Klavans, R. & Boyack, K.W.**, 2011. OpenOrd: an open-source toolbox for large graph layout. In P. C. Wong et al., eds., *Proceedings of SPIE - The International Society for Optical Engineering*. Bellingham: International Society for Optics & Photonics.
- Masud, L., Valsecchi, F., Ciuccarelli, P., Ricci, D. & Caviglia, G.**, 2010. From Data to Knowledge - Visualizations as Transformation Processes within the Data-Information-Knowledge Continuum. In *14th International Conference in Information visualisation*. New York: IEEE, pp. 445-449.
- Mauri, M., Pini, A., Ciminieri, D. & Ciuccarelli, P.**, 2013. Weaving data, slicing views: a design approach to creating visual access for digital archival collections. In *Proceedings of the Biannual Conference of the Italian Chapter of SIGCHI*. New York: ACM, Article No. 21.
- McCarty, W.**, 1998. *What is humanities computing? Toward a definition of the field*. Centre for Computing in the Humanities. London: King's College. Available at: <http://www.mccarty.org.uk/essays/McCarty,%20What%20is%20humanities%20computing.pdf> [Accessed 8 April 2015].

- McClure, D.**, 2013. *Announcing Neatline 2.0.0! A stable, production-ready release*. Available at: <http://scholarslab.org/announcements/neatline-2-0/> [Accessed 12 October 2015].
- McClure, D.**, 2014. *Visualizing 27 years, 12 million words of the Humanist list*. Available at: <http://dclure.org/essays/visualizing-the-humanist/> [Accessed 2 January 2015].
- McCormick, B., DeFanti, T.A. & Brown, M.D.**, 1987. Visualization in Scientific Computing. *IEEE Computer Graphics and Applications*, 7(10), pp.69–69.
- McTaggart, J.M.E.**, 1993. The Unreality of Time. In R. Le Poidevin & M. MacBeath, eds., *The philosophy of time*. New York: Oxford University Press.
- Mead, N. & Blight, G.**, 2014. *Eurozone crisis timeline: crisis, recession and growth*. The Guardian. London: Guardian News and Media Ltd. Available at: <http://www.theguardian.com/business/interactive/2012/oct/17/eurozone-crisis-interactive-timeline-three-years> [Accessed 18 August 2015].
- Meeks, E. & Grossner, K.**, 2014. *Topotime*. Stanford: Stanford University, Libraries. Available at: <http://dh.stanford.edu/topotime/index.html> [Accessed 2 May 2014].
- Miksch, S.**, 2004. Connecting time-oriented data and information to a coherent interactive visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York: ACM, pp. 105–112.
- Mitchell, L. ed.**, 2014. *Beautiful Data: A field Guide for Exploring Open Collections*. Cambridge, MA: metaLab (at) Harvard University. Available at: http://beautifuldata.metabolab.harvard.edu/wp-content/uploads/2014/09/BD_Field_Guide%20small.pdf [Accessed 14 January 2016].
- MoDA**, 2008. *The Silver Studio (1880-1963) A Design Practice At Work*. London: Museum of Domestic Art and Design. Available at: <http://www.moda.mdx.ac.uk/download?id=18663> [Accessed 16 December 2015].
- Moere, A.V., Tomitsch, M., Wimmer, C., Christoph, B. & Grechenig, T.**, 2012. Evaluating the Effect of Style in Information Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 18(12), pp.2739–2748.
- Moretti, F.**, 2000. Conjectures on World Literature. *New Left Review*, (1), pp. 54–68.

- Müller, B., Albrecht, K. & Dörk, M.**, 2015. Culturegraphy, Visualizing Cultural Network Dynamics. In R. Malina, M. Schich & I. Meirelles, eds., *Arts, humanities, and complex networks*. Cambridge, MA: MIT Press.
- Müller, C., Grottel, S. & Ertl, T.**, 2007. Image-Space GPU Metaballs for Time-Dependent Particle Data Sets. In H.P. Seidel, H.P.A. Lensch, B. Rosenhahn, P. Slusallek & J. Weickert, eds., *Proceedings of the Annual Workshop: Vision, modelling and visualization 2007*. Berlin: Akademische Verlagsgesellschaft, pp. 31-40.
- National Museum Directors' Council**, 2004. *A Manifesto for Museums*. London. Available at: http://www.nationalmuseums.org.uk/media/documents/publications/manifesto_for_museums.pdf [Accessed 5 October 2015].
- Neumann, J. von**, 1951. Various techniques used in connection with random digits. *Journal of Research of the National Bureau of Standards, Applied Mathematics Series*, 3, pp. 36-38.
- Neurath, M. & Kinross, R.**, 2009. *The transformer: principles of making Isotype charts*. London: Hyphen Press.
- Neustaedter, C., Brush, A.J. & Smith, M.A.**, 2005. Beyond from and received: exploring the dynamics of email triage. In *CHI Extended Abstracts on Human Factors in Computing Systems*. New York: ACM, pp. 1977-1980.
- Nielsen, J.**, 1994. *Usability Engineering*. San Francisco: Morgan Kaufmann.
- Nielsen, J.**, 2000. *Why You Only Need to Test with 5 Users*. Warm Springs: Nielsen Norman Group. Available at: <https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/> [Accessed 6 February 2016]
- Nielsen, J. & Landauer, T.K.**, 1993. A mathematical model of the finding of usability problems. In *CHI Proceedings of the INTERACT and CHI Conference on Human Factors in Computing Systems*. New York: ACM, pp. 206-213.
- Noë, A.**, 2004. *Action In Perception*. Cambridge, MA: MIT Press.
- Norman, D.A.**, 1991. Cognitive Artifacts. In J.M. Carroll, ed., *Designing Interaction*. Cambridge, MA: Cambridge University Press.
- Norman, D.A.**, 1993. *The things that make us smart: defending human attributes in the age of the machine*. Boston, MA: Addison-Wesley Publishing Company

- Nowviskie, B.P., McClure, D., Graham, W. & Soroka, A.**, 2013. Geo-Temporal Interpretation of Archival Collections with Neatline. *Literary and Linguistic Computing*, 28(4), pp.692-699.
- Nowviskie, B.P., McClure, D., Graham, W., Rochester, E. & Boggs, J.**, 2012. *About Neatline*. Scholar's Lab. Charlottesville: University of Virginia Library. Available at: <http://neatline.org/about/> [Accessed 5 March 2013].
- Oates, G.**, 2014. Keynote. *UK Museums and The Web 2014*. London, 7 November 2014. London: Museums Computer Group. Available at: <https://goodformandspectacle.wordpress.com/2014/11/10/uk-museums-and-the-web-keynote/> [Accessed 19 November 2014].
- OED Online**, 2015a. *Archive*. Oxford Dictionary of English. Oxford: Oxford University Press. Available at: <http://www.oxforddictionaries.com/definition/english/archive> [Accessed 11 January 2016].
- OED Online**, 2015b. *History*. Oxford Dictionary of English. Oxford: Oxford University Press. Available at: <http://www.oxforddictionaries.com/definition/english/history> [Accessed 11 January 2016].
- Osmani, A.**, 2012. *Learning JavaScript Design Patterns*. Sebastopol: O'Reilly Media.
- O'Regan, J.K. & Noë, A.**, 2001. A sensorimotor account of vision and visual consciousness. *Behavioral and brain sciences*, 24(5), pp.939-972.
- Pappas, A.**, 2014. *Visualizing Time Winners*. Visualizing.org. Available at: <http://visualizing.org/stories/visualizing-time-winners-o> [Accessed 15 May 2014].
- Park, H. & Choi, J.**, 2012. V-model: a new innovative model to chronologically visualize narrative clinical texts. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York: ACM, pp. 453-462.
- Pellegrini, V., Caviglia, G. & Ciuccarelli, P.**, 2013. Minerva: an Information Visualization Tool to Support Philosophical Historiography. In N. Ceccarelli, ed., *Proceedings of 2CO Communicating Complexity*. Rome: Edizioni Nuova Cultura. Available at: http://www.2coconference.org/docs/2CO_book_light_bn.pdf [Accessed 6 February 2016].
- Periscope**, 2013. *United States Gun Deaths*. Portland, OR: Periscope. Available at: <http://guns.periscope.com> [Accessed 19 August 2015].
- Perlin, K. & Fox, D.**, 1993. Pad: an alternative approach to the computer interface. In *Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques*. New York: ACM, pp. 57-64

- Plaisant, C., Milash, B. & Rose, A.**, 1996. LifeLines: visualizing personal histories. In *Conference Companion on Human Factors in Computing Systems*. New York: ACM, pp. 392-393.
- Platt, J.C., Czerwinski, M. & Field, B.A.**, 2003. PhotoTOC: automatic clustering for browsing personal photographs. In *Proceedings of the 2003 Joint Conference of the Fourth International Conference on Information, Communications and Signal Processing*. New York: IEEE, pp. 6-10.
- Playfair, W.**, 2005. *Playfair's Commercial and Political Atlas and Statistical Breviary*. New York: Cambridge University Press.
- Pointer, J.**, 1714. *A Chronological History of England*. Oxford: Oxford University Collection. Available at: <https://archive.org/details/achronologicalhoopooingooog> [Accessed 12 December 2013].
- Popova, M.**, 2013. *Stephen Jay Gould, the Greatest Science Essayist of All Time, on Evolution and Storytelling*. Brain Pickings. Available at: <https://www.brainpickings.org/2013/09/10/stephen-jay-gould-2000-interview/> [Accessed 19 November 2015].
- Posner, M.**, 2012. *Some Things to Think About Before You Exhort Everyone to Code*. Available at: <http://miriamposner.com/blog/some-things-to-think-about-before-you-exhort-everyone-to-code/> [Accessed 29 September 2015].
- Pousman, Z., Stasko, J.T. & Mateas, M.**, 2007. Casual Information Visualization: Depictions of Data in Everyday Life. *IEEE Transactions on Visualization and Computer Graphics*, 13(6), pp.1145-1152.
- Presner, T.**, 2015. The Ethics of the Algorithm: Close and Distant Listening to the Shoah Foundation Visual History Archive. In C. Fogu, W. Kansteiner & T. Presner, eds., *History Unlimited: Probing the Ethics of Holocaust Culture*. Cambridge: Harvard University Press. In Press.
- Presner, T., Schnapp, J. & Lunenfeld, P.**, 2009. *The digital humanities manifesto 2.0*. Los Angeles: UCLA Mellon Seminar in Digital Humanities. Available at: <http://manifesto.humanities.ucla.edu/2009/05/29/the-digital-humanities-manifesto-20/> [Accessed 6 February 2016]
- Priestley, J.**, 1764. *A Description of a Chart of Biography*, Warrington: Eyre. British Library: 611.d.30, 3.
- Priestley, J.**, 1780. *A Harmony of the Evangelists in English*. Hampshire: Gale ECCO
- Putnam, H.**, 1988. *Representation and reality*. Cambridge, MA: MIT Press
- Ramsay, S.**, 2011. *Who's In and Who's Out*. Available at: <http://stephen-ramsay.us/text/2011/01/08/whos-in-and-whos-out/> [Accessed 16 April 2015].

- Raskin, J.**, 1994. Intuitive equals familiar. *Communications of the ACM*, 37(9), p.17.
- Ratto, M.**, 2011. Critical making: Conceptual and material studies in technology and social life. *The Information Society*, 27(4), pp. 252-260.
- Rekimoto, J.**, 1999. Time-machine computing: a time-centric approach for the information environment. In *Proceedings of the 20th Annual ACM Symposium on User Interface Software and Technology*. New York: ACM, pp. 45-54.
- Rilling, J. & Mudur, S.P.**, 2002. On the use of metaballs to visually map source code structures and analysis results onto 3D space. In *Proceedings of the Ninth Working Conference on Reverse Engineering*. New York: IEEE, pp. 299-308.
- Ringel, M., Cutrell, E., Dumais, S. & Horvitz, E.**, 2003. Milestones in time: The value of landmarks in retrieving information from personal stores. In G.W.M. Rauterberg, M. Menozzi, & J. Wesson, eds., *Proceedings of INTERACT 2003*. Lansdale: IOS Press, pp. 184-191.
- Risden, K., Czerwinski, M.P. & Munzner, T.**, 2000. An initial examination of ease of use for 2D and 3D information visualizations of web content. *International Journal of Human-Computer Studies*, 53(5), pp. 695-714.
- Rittel, H.W.J. & Webber, M.M.**, 1973. Dilemmas in a general theory of planning. *Policy Sciences*, 4, pp. 155-169.
- Rodley, E.**, 2014. The Virtues of Promiscuity, or, Why giving it away is the future. In E. Rodley, R. Stein, & S. Cairns, eds., *CODE / WORDS technology and theory in the museum*. Boston & Edinburgh: Museums-ETC.
- Rogers, K., Hinrichs, U. & Quigley, A.**, 2014. It Doesn't Compare to Being There: In-Situ vs. Remote Exploration of Museum Collections. *The Search is Over: Exploring Cultural Collections with Visualization, Workshop Held in Conjunction with the Digital Libraries Conference DL 2014*. London, 11-12 September. London: City University London.
- Rosenberg, D.**, 2007. Joseph Priestley and the Graphic Invention of Modern Time. *Studies in Eighteenth Century Culture*, 1(1), pp. 55-103.
- Rosenberg, D. & Grafton, A.**, 2010. *Cartographies of Time*. New York: Princeton Architectural Press.
- Rowley, J.**, 2007. The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of Information Science*, 33(2), pp. 163-180.

- Scaife, M. & Rogers, Y.**, 1996. External cognition: how do graphical representations work? *International Journal of Human-Computer Studies*, 45(2), pp. 185-213.
- Scheiner, S.M.**, 2004. Experiments, Observations, and Other Kinds of Evidence. In M. L. Taper & S. R. Lele, eds., *The nature of scientific evidence*. Chicago, IL: University of Chicago Press, pp. 51-72.
- Schmidt-Burkhardt, A.**, 2011. *Maciunas' Learning Machines*. Berlin/Heidelberg: Springer.
- Schön, D.A.**, 1983. *The Reflective Practitioner*. New York: Basic Books.
- Schreibman, S, Siemens, R & Unsworth, J**, 2004. *A Companion to Digital Humanities*. Oxford: Blackwell Publishing.
- Schüler-Springorum, S.**, 2015. Keynote speech. *Presentation of the European Holocaust Research infrastructure*. Berlin, 26 March 2015. Amsterdam: European Holocaust Research Infrastructure.
- SFMOMA**, 2007. *SFMOMA ArtScope*. Available at: <http://www.sfmoma.org/projects/artscope/index.html> [Accessed 1 March 2013].
- Shneiderman, B.**, 2008. Extreme visualization: squeezing a billion records into a million pixels. In *Proceedings of the ACM SIGMOD International Conference on Management of Data*. New York: ACM, pp. 3-12.
- Shneiderman, B.**, 1996. The eyes have it: A task by data type taxonomy for information visualizations. In *Proceedings of the IEEE Symposium on Visual Languages*. New York: IEEE, pp. 336-343.
- Shneiderman, B.**, 2003. Why not make interfaces better than 3D reality? *IEEE Computer Graphics and Applications*, 23(6), pp. 12-15.
- Shneiderman, B. & Plaisant, C.**, 2006. Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. In *Proceedings of the 2006 AVI Workshop on Beyond Time and Errors: Novel Evaluation Methods for Information Visualization*. New York: ACM.
- SIMILE**, 2009. *Timeline CreatingHotzones*. SIMILE Widgets. Cambridge MA: Massachusetts Institute of Technology. Available at: http://simile-widgets.org/wiki/Timeline_CreatingHotzones [Accessed 20 August 2015].
- Simon, H.A.**, 1997. *Administrative Behavior* 4th ed. New York: The Free Press.
- Simon, H.A.**, 1996. *The Sciences of the Artificial*. Cambridge, MA: MIT Press.

- Simon, N.**, 2013. *Visualizing the Tate's Collection: What Open Data Makes Possible*. Museum 2.0. Available at: <http://museumtwo.blogspot.co.uk/2013/11/visualizing-tates-collection-what-open.html> [Accessed 27 November 2015].
- Smallman, H.S., John, M., Oonk, H.M. & Cowen, M.B.**, 2001. Information availability in 2D and 3D displays. *IEEE Computer Graphics and Applications*, 21(5), pp. 51-57.
- Smuts, A.**, 2009. What Is Interactivity? *The Journal of Aesthetic Education*, 43(4), pp. 53-73.
- Sparks, J.B.**, 1931. *The Histomap*. Chicago, IL: Rand McNally & Company.
- Spence, I.**, 2005. No Humble Pie: The Origins and Usage of a Statistical Chart. *Journal of Educational and Behavioral Statistics*, 30(4), pp. 353-368.
- Stanford Humanities + Design Lab**, 2014. *Palladio*. Stanford Center for Spatial and Textual Analysis. Stanford: Stanford University. Available at: <http://hdlab.stanford.edu/projects/palladio/> [Accessed 20 August 2015].
- Stanford Humanities + Design Lab**, 2015a. *Palladio Release Notes: 0.9.0*. Stanford Center for Spatial and Textual Analysis. Stanford: Stanford University. Available at: <http://hdlab.stanford.edu/lab-notebook/palladio/2015/01/27/palladio-release-0-9-0-release-notes/> [Accessed 20 August 2015].
- Stanford Humanities + Design Lab**, 2015b. *The Palladio 1.0 Release*. Stanford Center for Spatial and Textual Analysis. Stanford: Stanford University. Available at: <http://hdlab.stanford.edu/lab-notebook/palladio/2015/06/12/palladio-release-1-0-release-notes/> [Accessed 20 August 2015].
- Stern, M.**, 2014. *Designing Intuitive User Experiences*. WWDC 2014. San Francisco: Apple. Available at: <https://developer.apple.com/videos/play/wwdc2014-211/> [Accessed 9 December 2015].
- Sternfeld, J.**, 2014. Historical Understanding in the Quantum Age. *Journal of Digital Humanities*, 3(2). Available at: <http://journalofdigitalhumanities.org/3-2/historical-understanding-in-the-quantum-age/> [Accessed 5 January 2016].
- Strass, F.**, 1804. *Strom der Zeiten*. Vienna. Available at: <http://mapy.mzk.cz/en/mzko3/001/065/842/2610355751/> [Accessed 6 February 2016].
- Straup Cope, A.**, 2013. *A Timeline of Event Horizons*. Cooper Hewitt Labs. New York: Smithsonian Cooper Hewitt. Available at: <http://labs.cooperhewitt.org/2013/a-timeline-of-event-horizons/> [Accessed 20 August 2015].

- Swierenga, R.P.**, 1974. Computers and American History: The Impact of the “New” Generation. *The Journal of American History*, 60(4), p. 1045.
- Szpak, M.**, 2013. *Tate Britain Collection*. Available at: <http://zenlan.com/tate/rickshaw.html> [Accessed 19 September 2015].
- Tableau**, Tableau. Available at: <http://www.tableau.com/> [Accessed 20 August 2015].
- Tang, L. & Shneiderman, B.**, 2001. Dynamic Aggregation to Support Pattern Discovery: A Case Study with Web Logs. In *Discovery Science: Volume 2226 of the series Lecture Notes in Computer Science*. Berlin/Heidelberg: Springer, pp. 464-469.
- Thaller, M.**, 1987. Methods and Techniques of Historical Computation. In P. Denley & D. Hopkin, eds., *History and Computing*. Manchester: Manchester University Press.
- The British Library**, 2015. *Unlocking The Value - The British Librarys Collection Metadata Strategy 2015-2018*. London: The British Library. Available at: <http://www.bl.uk/bibliographic/pdfs/british-library-collection-metadata-strategy-2015-2018.pdf> [Accessed 28 November 2015].
- The British Museum**, 2013. *About the Collection database online*. London: The British Museum. Available at: http://www.britishmuseum.org/research/collection_online/about_the_database.aspx [Accessed 17 November 2015].
- The National Gallery**, *The new Tate Gallery*. London: The National Gallery. Available at: http://www.nationalgallery.org.uk/paintings/history/collection-history/collection-history/*/viewPage/4 [Accessed 3 December 2015].
- Thomas, H. & Crossman, S.**, 2006. *Museums, Galleries and Digitisation*. Edinburgh: Scottish Museum Council
- Thomas, W.A. & Carey, S.**, 2005. Actual/virtual visits: What are the links. In *Museums and the Web 2005*. Toronto: Archives & Museum Informatics.
- Thomas, W.G.**, 2004. Computing and the Historical Imagination. In S. Schreibman, R. Siemens & J. Unsworth, eds., *A Companion to Digital Humanities*. Oxford: Blackwell
- Thudt, A., Hinrichs, U. & Carpendale, S.**, 2012. The bohemian bookshelf: supporting serendipitous book discoveries through information visualization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York: ACM, pp. 1461-1470.

- Trumpener, K.**, 2009. Critical Response I. Paratext and Genre System: A Response to Franco Moretti. *Critical inquiry*, 36(1), pp. 159-171.
- Tufte, E.R.**, 2001. *The visual display of quantitative information*. Cheshire, CT: Graphics Press.
- Tukey, J.W.**, 1977. *Exploratory Data Analysis*. Boston, MA: Addison-Wesley Publishing Company.
- Turing, A.M.**, 1938. On computable numbers, with an application to the Entscheidungsproblem. A correction. *Proceedings of the London Mathematical Society*, 2(1), p. 544.
- Twyman, M.**, 1979. A Schema for the Study of Graphic Language. In P.A. Kolers, M.E. Wrolstad & H. Bouma, eds., *Processing of Visible Language*. Boston, MA: Springer US, pp. 117-150.
- Uboldi, G., Caviglia, G., Coleman, N., Heymann, S., Mantegari, G. & Ciuccarelli, P.**, 2013. Knot: an interface for the study of social networks in the humanities. In *Proceedings of the Biannual Conference of the Italian Chapter of SIGCHI*. New York: ACM, Article No. 15.
- Unsworth, J.**, 2000. Scholarly primitives: What methods do humanities researchers have in common, and how might our tools reflect this. In *Symposium on Humanities Computing: Formal Methods, Experimental Practice*. London, May 13. London: King's College.
- Unsworth, J.**, 2003. The Humanist: 'Dances with Wolves' or 'Bowls Alone'. In *Scholarly Tribes and Tribulations: How Tradition and Technology Are Driving Disciplinary Change*. Washington D.C.: Association of Research Libraries
- UoV Library Scholars' Lab**, 2009. *Neatline: Geospatial & Temporal Interpretation of Archival Collections*. Scholar's Lab. Charlottesville: University of Virginia Library. Available at: <http://neatline.org/> [Accessed 12 October 2015].
- UoV Library Scholars' Lab**, *Plugins*. Scholar's Lab. Charlottesville: University of Virginia Library. Available at: <http://neatline.org/plugins/> [Accessed 12 October 2015].
- Van Dam, A., Forsberg, A.S., Laidlaw, D.H., LaViola, J.J. Jr & Simpson, R.M.**, 2000. Immersive VR for scientific visualization: A progress report. *IEEE Computer Graphics and Applications*, 20(6), pp. 26-52.
- Varela, F.J., Thompson, E. & Rosch, E.**, 1992. *The Embodied Mind: Cognitive Science and Human Experience*. Cambridge, MA: MIT Press.
- Viégas, F.B., Wattenberg, M. & Cohen, S.**, 2010. *TimeFlow*. hint.fm. Available at: <http://hint.fm/projects/timeflow/> [Accessed 14 May 2014].

- Virzi, R.A.**, 1990. Streamlining the Design Process: Running Fewer Subjects. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, 34(4), pp. 291-294.
- visualizing.org**, 2013. *Visualizing Time Challenge*. Available at: <http://visualising.org/contests/chronozoom> [Accessed 10 October 2015].
- Wainer, H.**, 2005. *Graphic Discovery: A Trout in the Milk and Other Visual Adventures*. New York: Princeton University Press.
- Wainer, H.**, 2014. William Playfair. In *StatProb: The Encyclopedia Sponsored by Statistics and Probability Societies*. Boston MA: Springer US. Available at: <http://statprob.com/encyclopedia/WilliamPLAYFAIR.html> [Accessed 22 January 2014]
- Ware, C.**, 2004. *Information Visualization: Perception for Design*. Philadelphia: Elsevier.
- Weinberger, D.**, 2010. *The Problem with the Data-Information-Knowledge-Wisdom Hierarchy*, Harvard Business Review Blog Network. Available at: <https://hbr.org/2010/02/data-is-to-info-as-info-is-not> [Accessed 19 Nov. 2015].
- Weinert, F.**, 2013. *The March of Time*. Berlin/Heidelberg: Springer.
- Whitelaw, M.**, 2015. Generous Interfaces for Digital Cultural Collections. *Digital Humanities Quarterly*, 9(1). Available at: <http://www.digitalhumanities.org/dhq/vol/9/1/000205/000205.html> [Accessed 11 August 2014].
- Whitelaw, M.**, 2012. Towards generous interfaces for archival collections. *Comma*, 2012(2), pp.123-132.
- Whitelaw, M.**, 2009. Visualising Archival Collections: The Visible Archive Project. *Archives & Manuscripts*, 37, pp.22-40. Available at: <http://mtchl.net/papers/VisualisingArchivalCollections.pdf> [Accessed 21 June 2015].
- Wickham, H.**, 2013. Graphical criticism: some historical notes. *Journal of Computational and Graphical Statistics*, 22(1), pp. 38-44.
- Williams, D.**, 2010. A Brief History of Museum Computerization. In R. Parry, ed., *Museums in a Digital Age*. New York: Routledge, pp. 15-21.
- Winchester, I.**, 1980. Priorities for Record Linkage: A Theoretical and Practical Checklist. In J. M. Clubb & E. K. Scheuch, eds., *Historical Social Research: The Use of Historical and Process-Produced Data*. Stuttgart: Klett-Cota.

- Winesmith, K. & Carey, A.**, 2014. *Why Build an API for a Museum Collection?* San Francisco: SFMOMA. Available at: http://www.sfmoma.org/about/research_projects/lab/why_build_an_api [Accessed 19 September 2015].
- Wiss, U., Carr, D. & Jonsson, H.**, 1998. Evaluating three-dimensional information visualization designs: a case study of three designs. In *Proceedings of the IEEE Conference on Information Visualization*. New York: IEEE, pp. 137-144.
- Wurman, R.S.**, 1990. *Information Anxiety*. New York: Bantam Books.
- Yiua, K.S., Baeckerb, R., Silverb, N. & Longb, B.**, 1997. A Time-Based Interface for Electronic Mail and Task Management. *Advances in Human Factors Ergonomics*, 21, pp.19-22.
- Zacks, J., Levy, E., Tversky, B. & Schiano, D.J.**, 1998. Reading bar graphs: Effects of extraneous depth cues and graphical context. *Journal of Experimental Psychology: Applied*, 4(2), pp. 119-138.
- Zerubavel, E.**, 2004. *Time Maps: Collective Memory and the Social Shape of the Past*. Chicago, IL: University of Chicago Press.
- Zimmerman, E.**, 2003. Play as research: The iterative design process. In B. Laurel, ed., *Design Research: Methods and Perspectives*. Cambridge, MA: MIT Press.
- Zins, C.**, 2007. Conceptual approaches for defining data, information, and knowledge. *Journal of the American Society for Information Science and Technology*, 58(4), pp.479-493.

Glossary

API

An application program interface (API) enables different kinds of programmes as well as individual components of a software application to communicate with each other. In the context of digital collections, an API enables machine-readable access to online collections data, as opposed to browser-based interfaces which are intended for human users.

CSV

CSV is a file format to store tabular data in plain text and stands for “comma-separated values”. There is no official standard, but it is customary for the first line of the text file to consist of the table headers, separated by commas, followed by the table rows on individual lines. CSV is a popular format for sharing datasets that can be represented as a spreadsheet, as it avoids compatibility issues that may arise from using proprietary formats such as Microsoft Excel files.

D3

D3 (or “Data-driven Documents”) is a JavaScript library geared towards web-based data visualisation. It operates by binding datasets to (visual) elements in a browser and offers routines for manipulating and updating a browser view based on data. Some ‘layouts’ that automatically prepare data for visualisation in standard diagram formats are included, as well as components that generate recurring elements of diagrams, such as axes. Most of the library’s functionalities are however not directed at creating specific types of diagrams, focussing instead on basic building blocks and interaction paradigms that allow the design of arbitrary and completely novel visualisation formats.

Database

A database is a structured collection of data. In the context of information technology, digital data is organised according to the database model. Relational models are among the most prevalent and rely on tables to structure information. Several tables can be used to describe relationships between records in order to store data that cannot easily be modelled in a single table. This could include fields that may contain several values (e.g. the material of an object) or connections between different kind of entities (e.g. an object and its creator).

Field

A database record consists of one or more fields along with their corresponding values. The field identifies the type of stored data along with its format; a title field may expect text of not more than a few hundred characters while a date field will expect temporal information.

Histogram

A diagram format which represents the distribution of data according to a specific dimension. To construct a histogram, the data is sorted into “bins” - ranges of possible values according to a chosen interval - and the number of items per bin is counted. Each bin is then graphically represented as a rectangle, whose height corresponds to its number of items and whose widths equal the chosen interval. Histograms are often used to visualise time-dependent data, using units of time as binning intervals (hours, days, months etc.). Choosing an appropriate bin size is crucial as the observable patterns may differ significantly depending on the chosen value.

JSON

JavaScript Object Notation (JSON) is a standard format for sharing data in a text-format, which is readable for both humans and machines. It is based on two structures: objects, which are a collection of pairs of names and values, and arrays, an ordered list of structures. In contrast to CSV files, JSON is able to store datasets which can not easily be represented in a spreadsheet, such as datasets which consist of records with several field values as well as hierarchical data. Although

JSON is based on JavaScript, the format itself is language independent.

Example:

```
{
  "accessionNumber": "0123456",
  "contributors": [
    {
      "name": "Doe, Joanne",
      "role": "artist",
      "gender": "female"
    }
  ],
  "date": "1794",
  "title": "Untitled"
}
```

Mapping

In the context of this thesis ‘mapping’ refers - unless specified - not to the mapping of land, but to the arithmetical mapping of input data to a (graphical) output space. A mapping function produces a (numeric) output for every acceptable input. The mapping can be, but often is not, reversible. For example, a perspectival mapping function translates data from a three-dimensional input space to a two-dimensional output space, losing some of the original data in the process.

Record

A database record can be understood as a single data item, like a single row in a spreadsheet. In the context of this thesis, a record from a digital collection usually represents the catalogue descriptions of a single item within the collection. What constitutes an item depends on the conventions of the database owner.

XML

Extensible Markup Language or XML is an open standard for defining structured documents. It is often used in online APIs to transmit data in a format that is both human and machine-readable. In the recent past, XML is increasingly replaced by JSON.