EG UK Computer Graphics & Visual Computing (2015)
Rita Borgo, Cagatay Turkay (Editors)

# Revealing Cultural Collections Over Time

F. Kräutli[1] and S. Boyd Davis[1]

[1]Royal College of Art, London

**Abstract**

*We present a novel timeline visualisation tool designed for visual analysis of digital cultural collections. We evaluated our prototype in use with a number of different collections, such as the digital catalogue of the art collection of the Tate and Cooper Hewitt's object collection. Through our research, we explore the particular challenges for visualisation of digital collections data and, in close collaboration with museum professionals, identify the potential of visual analytics of cultural data in a humanities context.*

## 1. Introduction

The Digital Humanities (DH) — a field that has emerged to describe the growing synergies between humanities scholarship and computer science [SSU04] — have inherited a range of technical and conceptual approaches from the quantitative sciences. Visual analytics features prominently in DH projects [LBD*12, SSU04] and in related approaches to sense-making from digital data eg. [WCR*11], but the nature of the data that is processed in the humanities as well as the kind of knowledge that may be discovered differ substantially from the sciences.

With advances in graphical user interfaces and improved usability, it became practical for humanists to adopt existing visual analysis software originally designed for engineers, statisticians or business analysts for their own purposes. However, it is argued that these sophisticated interfaces tend to hide software's underlying assumptions, which stem from outside the humanities discipline [SSU04, Bor09, Dru14, Lee00]; "such graphical tools are a kind of intellectual Trojan horse" [Dru11].

The limitations of available software tools for humanities research have been pointed out since the 1980s [Win80], but early efforts in developing database tools specifically for humanities computing [Tha87] then found little acceptance. Humanities data, which often may be fuzzy, incomplete, uncertain or disputed, remains difficult to model using existing database software.

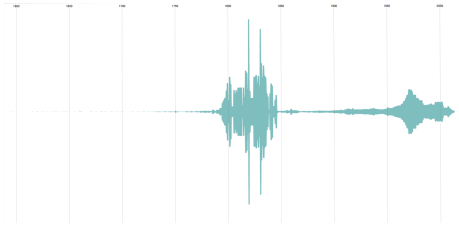

**Figure 1:** *The works of art in the Tate collection visualised according to their production date. Two distinct clusters are visible.*

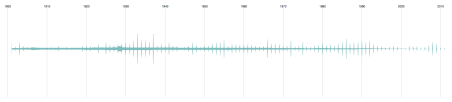

**Figure 2:** *Cooper Hewitt's object collection displays an even distribution when visualised by production date. The vertical spikes reveal quantisation caused by the granularities of the date descriptions.*

## 2. The nature of collections data

Libraries, archives and museums have nevertheless adopted database software early on in order to maintain their collections in digital catalogues. Over the years these catalogues have been enriched with meta data, turning these digital catalogues into valuable resources for research in their own right.

Our primary aim is to reveal new knowledge by developing visual analytic tools to explore these diverse large

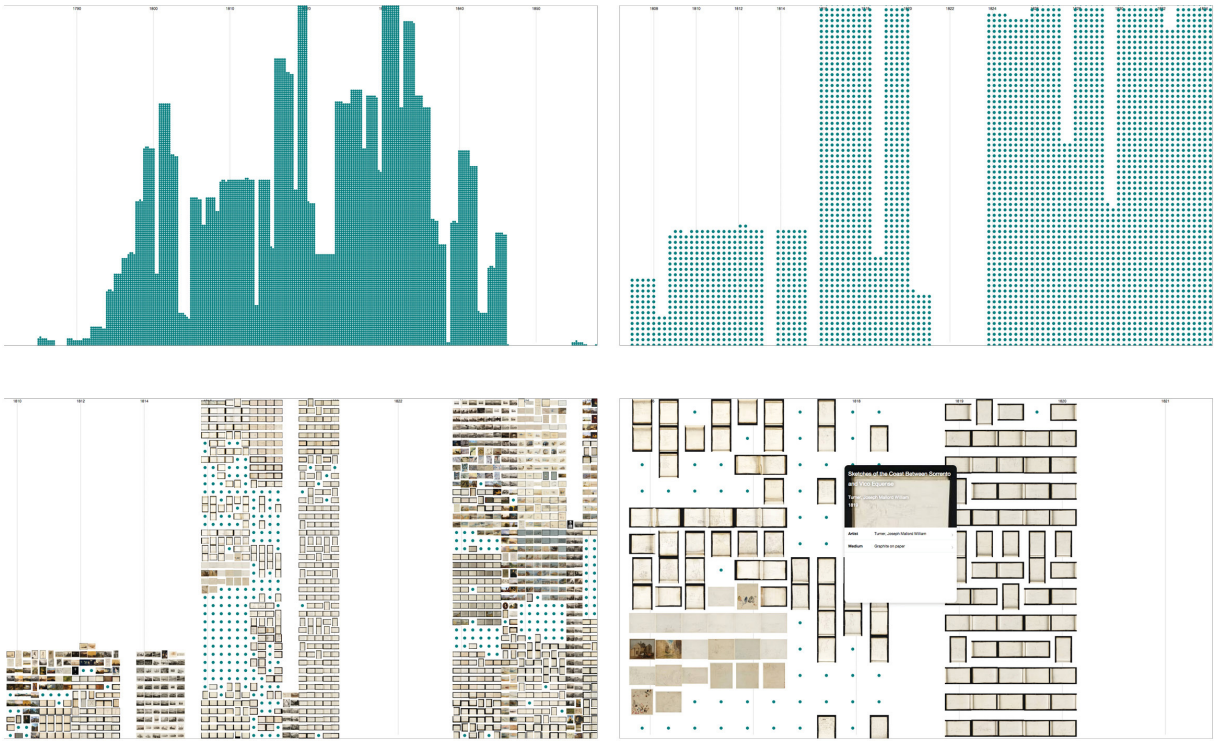

**Figure 3:** *Zooming in on the earlier cluster in the Tate collection reveals a bias in cataloguing: pages from Turner's sketchbooks have been recorded individually.*

datasets in rich interactive timelines and to understand what visual analytics could contribute to culture and heritage institutions.

Collections data pose specific challenges with regards to time-wise visualisations. Each record represents an individually authored set of data; in contrast to a time series, where each data point represents a controlled and comparable set of measurements. Items in collections may be inconsistently catalogued by a number of different curators. Fields may expect a certain input format, which skews the information that a cataloguer meant to enter. Cataloguing practices vary across and within institutions. Although there have been efforts to create unified standards for digitally storing cultural data [DH11], museums already have difficulties adhering to their internal standards. "Museum data (and to a lesser extent library and archive data) is non-standard, largely unstructured and dependent on complex relations." [Poo15]

## 3. Visualisation design

In our research we have explored how to address these challenges through functional visualisation prototypes and — in collaboration with curators and archivists — evaluated the benefits of visual analysis for digital collections. We will present a short overview of the tool we developed and the insights it is able to facilitate.

Following Shneiderman's mantra "overview first, detail on demand" [Shn96], we developed a visualisation format which allows users to explore their entire digital collection on a timeline. The merits of being able to inspect a "complete" dataset on a visual timeline have been pointed out since the 18th century [BD15] and so have the difficulties of graphically mapping large datasets onto a linear timescale [Pri64]: the visualisation becomes cluttered and, at the worst, unreadable. A range of clutter reduction strategies are available to address this problem [ED07], some of which are appropriate for timeline visualisations. For our tool, we developed a custom algorithm which makes use of the particularities of cultural datasets to generate a readable layout also for large datasets.

We represent each record as a disk of the same size, positioned on the horizontal axis according to its production date — or any other appropriate date — and arranged vertically by order of accession. In order to reduce clutter and position the disks in a space-efficient manner, we capitalised on the differences in granularity and uncertainty the dates have been specified in. Records dating from the same timeframe would therefore not immediately be placed on top of

each other, but first distributed horizontally within the allowed timeframe – (on our representation and exploitation of uncertainty see [KBD13] and [BDK15]). The resulting visualisation is visually similar to a histogram and may be read as an overall distribution of records over time, but it circumvents the need to summarise records and allows individual items to be examined. This is supported through a zoom and pan behaviour, allowing the visualisation to be navigated like a digital map.

## 4. Evaluation and Discussion

We have used the tool to analyse a range of digital collections, collaborating with archivists and curators in order to gather feedback and find out what kind of insights our tool might enable. Our methods are thus grounded in Design Research, using iterative design of functioning visualisation prototypes with ongoing evaluation through reflection and dialogue with the curators and archivists. We assume, as recognised early in the years of Design Research, that the questions, issues and problems to be interrogated and presented are reformulated during the design and development process [Arc68].

Generally, our humanities colleague's first impression is a sense of the position and distribution of the collection over time: the extent of the timeframe the collection covers and the distribution of items within that timeframe. The Tate collection produces two clusters of works between the years 1800 and 1850 and between 1950 and 2000 (see figure 1), while Cooper Hewitt's object collection is more evenly distributed (see figure 2).

The scale of a collection, the sheer number of items, becomes apparent by zooming in, gradually rendering the individual disks representing the records and, where appropriate, their associated images. Focusing on the large cluster in the Tate visualisation causes a remarkably uniform set of images to appear, which turn out to be individually catalogued pages from J. M. W. Turner's sketch books (see figure 3). We found similar biases and inconsistencies in cataloguing in all of the collections we examined. In some cases, the archivists already suspected them and the visualisation could serve as evidence or give further insights into the nature and extent of the bias. Visual analytics of cultural collections reveals as much about past curatorial and archival practices as it does about the items in the collection.

Working closely with the archivists and curators we identified a number of potential use cases for these visual analytic tools. One curator commented on the usefulness of such visualisations for new staff to learn about the content and history of a collection. Several commented on the ability of these visualisations in helping to find errors in the data and identifying records that lack essential information. Besides the immediate insights the visualisations allowed, they triggered new research questions as well as practical suggestions for the development of other visualisation formats. More specialised visualisations have now been developed that pursue curators specific questions, and in the process also reveal unsuspected patterns [BDK15].

## 5. Conclusion

By developing a visualisation tool specifically tailored to the peculiarities of collections data, and not the least by working closely with archivists and museum curators, we were able to gather new insights into the potential of visual analytics for cultural data. Besides offering a new perspective on the contents of digital collections, visualisation allows for a closer examination of the history of a collection and the practices and decisions of the individuals responsible for it.

## References


[Arc68] ARCHER B. L.: *The Structure of Design Processes*. PhD thesis, Royal College of Art, 1968. 3

[BD15] BOYD DAVIS S.: "To see at one glance all the centuries that have passed" - early visualisations of historical time. In *Gower Handbook of Information Design*, Black A., Lund O., Walker S., (Eds.). Farnham, UK, 2015. 2

[BDK15] BOYD DAVIS S., KRÄUTLI F.: Image and idea: Interactions between Design and Digital Humanities. *Visible Language special issue 'Critical Making: Design and the Digital Humanities'. In preparation (abstract accepted)* (2015). 3

[Bor09] BORGMAN C. L.: The Digital Future is Now: A Call to Action for the Humanities. *Digital Humanities Quarterly 3* (2009), 1–20. 1

[DH11] DAWSON A., HILLHOUSE S. (Eds.): *SPECTRUM 4.0*. Collections Trust, 2011. 2

[Dru11] DRUCKER J.: Humanities approaches to graphical display. *Digital Humanities Quarterly 5*, 1 (2011). 1

[Dru14] DRUCKER J.: Graphesis. Harvard University Press, 2014. 1

[ED07] ELLIS G. P., DIX A. J.: A Taxonomy of Clutter Reduction for Information Visualisation. *IEEE Trans. Vis. Comput. Graph. () 13*, 6 (2007), 1216–1223. 2

[KBD13] KRÄUTLI F., BOYD DAVIS S.: Known unknowns: representing uncertainty in historical time. In *EVA London* (2013), pp. 1–8. 3

[LBD*12] LUNENFELD P., BURDICK A., DRUCKER J., PRESNER T., SCHNAPP J.: *Digital_Humanities*. MIT Press, 2012. 1

[Lee00] LEE H. L.: What is a collection? *Journal of the American Society for Information Science 51*, 12 (Jan. 2000), 1106–1113. 1

[Poo15] POOLE N.: Aggregation & the Culture Grid, June 2015. 2

[Pri64] PRIESTLEY J.: *A Description of a Chart of Biography*. Warrington: Eyre. British Library: 611.d.30, 3., 1764. 2

[Shn96] SHNEIDERMAN B.: The eyes have it: A task by data type taxonomy for information visualizations. *Visual Languages, 1996. Proceedings., IEEE Symposium on* (1996), 336–343. 2

[SSU04] SCHREIBMAN, S, SIEMENS, R, UNSWORTH, J: *A companion to digital humanities*. Blackwell Publishing, 2004. 1

[Tha87] THALLER M.: Methods and Techniques for Historical Computation. In *History and computing*, Denley P., Hopkin D., (Eds.). Manchester University Press, 1987. 1

[WCR*11] WONG B., CHOUDHURY S., ROONEY C., CHEN R., XU K.: Invisque: Technology and methodologies for interactive information visualization and analytics in large library collections. In *Research and Advanced Technology for Digital Libraries*, Gradmann S., Borri F., Meghini C., Schuldt H., (Eds.), vol. 6966 of *Lecture Notes in Computer Science*. Springer Berlin Heidelberg, 2011, pp. 227–235. 1

[Win80] WINCHESTER I.: Priorities for record linkage: A theoretical and practical checklist. In *Historical social research the use of historical and process-produced data*, Clubb J. M., Scheuch E. K., (Eds.). Stuttgart, 1980. 1