Martin Grandjean

Data Visualization for History

Abstract: It is logical that the generalization of digital approaches in history is leading to a democratization of the graphic representation of the data produced by these processes. Rather than presenting long series of examples, this very cursory chapter seeks to fuel reflection on our uses: why do we visualize historical data? Is it for illustrative purposes, to "show" our historical object and make it understandable to a large audience? Or is it, on the contrary, because the raw data is unintelligible to us, and visualization is therefore a heuristic tool intended for their exploration? The central point of my argument is based on a typology of sources and uses, a double entry table which is intended as a kind of decision-making aid for those seeking to make their data speak in the right way to the right audience.

Keywords: data visualization, digital history, digital humanities, infographics, visual analytics

Introduction

The widespread use of visual representations in historical science reflects the public's passion for historical questions in general: as this subject arouses a strong and lasting interest from society and as such a rich iconographic tradition has made history intelligible to so many people for so long, the integration of graphics into historical narratives is now common.[1] Presently, this practice is mainly descriptive, taking advantage of relatively universal visual codes to summarize an object or a historical phenomenon, "simplifying" it to make it more understandable. Of course, visualization is not a prerogative of history. Far from it. It is primarily a cross-curricular skill inspired by the hard sciences and statistical methods, spreading into all areas of research. It is this hybridization that drives a large community of historians, researchers in social sciences, economics, literary, or artistic studies, journalists, and graphic designers to produce visual representations that go far beyond the mere presentation of results toward more analytic and exploratory approaches. In this essay, we will focus on issues related to visualization in history practice, without disregarding the works of some references in neighboring areas,

1 Johanna Drucker, 2015. "Graphical Approaches to the Digital Humanities," in *A New Companion to Digital Humanities*, ed. Susan Schreibman, Ray Siemens and John Unsworth (John Wiley & Sons, Ltd., 2008), 238–50; Martyn Jesson, "Digital Visualization as a Scholarly Activity," *Literary and Linguistic Computing* 23 (3): 281–293.

https://doi.org/10.1515/9783110430295-024

such as those of Tukey, Tufte, and others.[2] Process automation, corpus massification,[3] "distant reading,"[4] interactivity, and data sharing have emerged as methods for sharing research and involving audiences. Nonetheless, the evolution from an illustrative visualization to visualization that integrates the research process, or allows readers or exhibition visitors to navigate the data for themselves, raises many questions.

Rather than provide an inventory of projects and methods, this paper offers a more conceptual reflection around the production of these graphic representations, from the author's ideas to the reader's comprehension. While data visualization is a prevalent practice in historical science and its ancillary disciplines, at least in its simplest forms, the recourse to these sometime fascinating visual objects demands a critical discussion and typology of uses.

A History of Visualizing Data

It is important to remember that visualization is only one step among others when processing a dataset. Visualizations simplify and, thus, are not always able to express the richness of the object they describe. If this visual product naturally makes it possible to "see" the historical data, or at least one of its facets, a graphic representation may not be the perfect end point of a demonstration or an exhibition.

But this visual simplification, which is a very ancient practice, is of course beneficial to information dissemination. For example, representing territory in the form of a map augmented with markers indicating elements of physical (such as mountain ranges in rock paintings) or human geography is a form of visualization that preceded formal writing. Next came symbolic and political cartography,[5] the objectification of the "frontier", which made territory maps and cadastral plans performative visualizations. Celestial cartography, which is suspected to be even older than its earthly counterpart, is also a perfect example of simplifying information for

2 See John W. Tukey, *Exploratory Data Analysis* (Reading: Pearson, 1977); Edward R. Tufte, *The Visual Display of Quantitative Information* (Cheshire: Graphics Press, 1983); Amy Maxmen, "Three Minutes with Hans Rosling Will Change Your Mind about the World," *Nature* no. 540 (2016): 330–33; Michael Bostock, Vadim Ogievetsky, and Jeffrey Heer, "D3: Data-Driven Documents," *IEEE Trans. Visualization & Comp. Graphics* (2011).

3 Shawn Graham, Ian Milligan and Scott Weingart, *Exploring Big Historical Data, The Historian's Macroscope* (London: Imperial College Press, 2015); Andreas Fickers, "Towards a New Digital Historicism? Doing History in the Age of Abundance," *VIEW Journal of European Television History and Culture* (2012): 19–26.

4 Franco Moretti, *Distant Reading* (Verso Books, 2013).

5 Christian Jacob, *L'empire des cartes. Approche théorique de la cartographie à travers l'histoire*. (Paris: Albin Michel, 1992).

practical purposes: even if, since the Antiquity, certain astronomical atlases sought to be as complete as possible, most were simplified maps of the brightest stars and major constellations for sailors and travelers. In the register of highly codified visual representations, we also find medieval family trees. On one or more sides of the cenotaph of a recumbent, for example, the representation in bas-relief of the coats of arms were organized according to marriages and relatives. This allowed the viewers to glance over the genealogy of the monarch, understand the alliances formed, and reconstruct the recent history of power without needing to know how to read. This is an early example of a "public" history – a history that was available for all to see.

More recently, during the enlightenment, engineer William Playfair is often credited with the first graphical representations of statistical data.[6] Inspired by the timelines of Priestley who, in 1765, visualized the lifespan of two thousand personalities along an axis of time spanning almost three millennia,[7] Playfair offered many time series documenting the British foreign trade balance and can be therefore considered as the inventor of the histogram (the bar graph). In 1869, Charles Joseph Minard made the "first" emblematic historical visualization that grew from a statistical framework. That graphic, *Carte figurative des pertes successives en hommes de l'armée française dans la campagne de Russie* (1812–1813), was described by Robinson[8] and notably popularized by Tufte (Fig. 1).[9] This map simplified the constantly decreasing number of soldiers in the Napoleonic army into a graphic that very effectively demonstrated Napoleon's ill-fated march; it still inspires many graphic designers today. At the beginning of the twentieth century the use of graphs that combined statistical data, conceptual plans, and geographical maps grew dramatically,[10] although the practice is far from being generalized in the humanities. From the rise of statistical atlases and their fascinating thematic maps developed in the second half of the nineteenth

6 William Playfair, *Commercial and Political Atlas: Representing, by Copper-Plate Charts, the Progress of the Commerce, Revenues, Expenditure, and Debts of England, during the Whole of the Eighteenth Century* (London (1786), recently reissued by Howard Wainer and Ian Spence, eds. 2005); *The Commercial and Political Atlas and Statistical Breviary* (New York: Cambridge University Press, cited by Michael Friendly, 2007); "A Brief History of Data Visualization," in *Handbook of Computational Statistics: Data Visualization*, edited by C. Chen, W. Härdle and A. Unwin (Heidelberg: Springer), 1–34.

7 Joseph Priestley, *A Chart of Biography* (London: British Library, 1765), 611.I.19.

8 Arthur H. Robinson, "The Thematic Maps of Charles Joseph Minard," *Imago Mundi* 21, no. 1 (1967): 95–108.

9 Tufte (1983).

10 Charles van den Heuvel, "Building Society, Constructing Knowledge, Weaving the Web: Otlet's Visualizations of a Global Information Society and His Concept of Universal Civilization," in *European Modernism and the Information Society*, edited by W. Boyd Rayward (London: Ashgate, 2008), 127–153.

Fig. 1: Visitors in front of a giant reproduction of Minard's 1869 map at the Mundaneum, in Mons (Belgium). CC-BY-SA Martin Grandjean 2021.

century to "graphic semiology",[11] the modes of representation have evolved and stabilized. In this field, their use remains nevertheless relatively descriptive and heuristic data visualizations are still rare.

Typologies of Visualization: Sources and Uses

Having data that can be visualized does not mean that it needs to be visualized, or at least not in any way for any audience. Thus, we propose a typology for visualization adapted to (public) history. It unfolds along to two main axes (see Fig. 2): a typology of historical sources that distinguishes representations that are drawings based on information aggregation from those that are based on quantitative datasets; and a typology of use that differentiates the representations intended to illustrate or describe a situation in a simple way to make it immediately understandable, those produced to make data accessible through an interactive interface, and those likely to be the analyst's tool to generate new knowledge in a research process.

Axis 1: Which Sources? Infographics and Data Visualizations

The relationship between a researcher and his sources is a fundamental distinguishing characteristic of historical science. Thus, we expect a high degree of

11 Jacques Bertin, *Sémiologie Graphique. Les Diagrammes. Les Réseaux. Les Cartes* (Paris: Mouton et Gauthier-Villars, 1967).

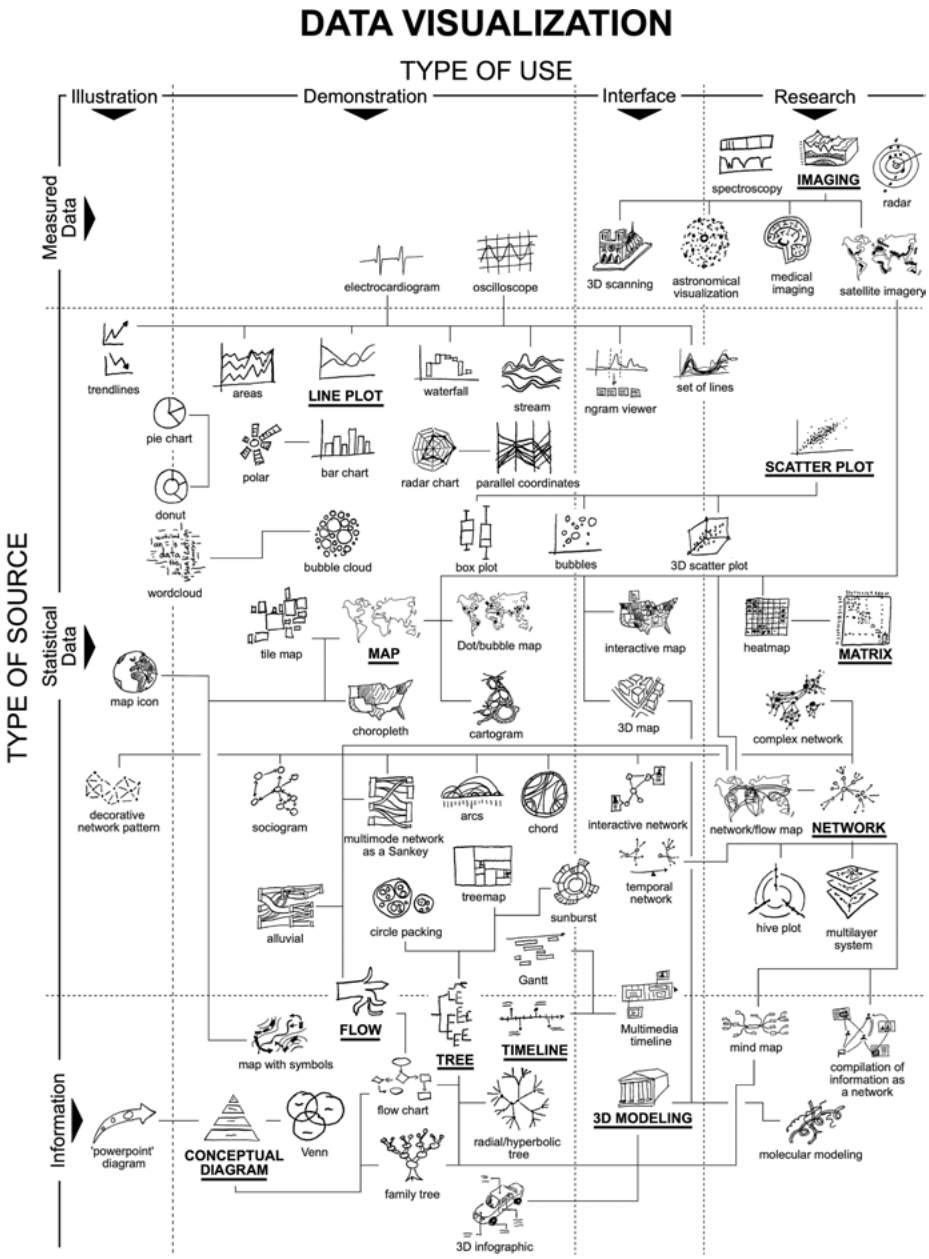

Fig. 2: Typology of data visualization along two axes: the type of data sources (vertical) and the type of use (horizontal). This table is intended to fuel reflection, to support the realization of a visualization. It cannot contain all the scenarios but offers 70 examples of frequent graphic forms, organized in families, to serve as reference points. CC-BY-SA Martin Grandjean 2021.

precision from historical visualization as regards the type of data that lies beneath the graphical representation. The nuance between an "infographic" and "data visualization" can be difficult to see, as the terms are often used interchangeably, especially now that the use of such visual aids has increased, especially in digital public history practices.

But the difference is important: making a representation based on a compilation of information (and on short historical narratives) is an act that involves a graphic and manual layout, whereas the visualization of a dataset is an operation that can be automatically done by software. And so, it is according to the type of sources used, and their serial processing, that one distinguishes "infographics" from "data visualizations", even though they sometimes may be visually similar. For example, the historical literature is very fond of maps – which may be created from quite different sources. A map of global movements of populations, battlefields, or railway networks may be an object that is "drawn" by its author on the basis of the information it graphs (Fig. 2 "map with symbols"). By contrast, a map with markers of industrial production, population density, or the position of monuments may not be a drawing at all, but the product of a formalized procedure based on a list of geographical coordinates and values (Fig. 2 "dot/bubble map").

In the field of information graphics, we therefore find all representations that are not based on numerical data. Thus, a diagram that clarifies the hierarchy or the succession of several elements (Fig. 2 "flow chart") and a dendrogram describing the structure of an institution or a family (Fig. 2 "family tree") are "drawings" because they do not imply a formal data table. Conversely, a curve tracing the evolution of a value over time (Fig. 2 "line plot"), a set of points in a two-dimensional space (Fig. 2 "scatter plot"), or a histogram of a value that evolves over time (Fig. 2 "bar chart") are pure statistical representations. This is a category in which we also find much more complex forms, especially when it represents relational data (Fig. 2 "matrix" and "network"), like social relations or the circulation of goods, people, or documents.[12] This typological axis might suggest that information design is a less noble art than data visualization because it produces simpler images or those in which technical connotation is less pronounced. However, it must be remembered that these two types of graphic representations are complementary rather than hierarchical. They simply show sources with different characteristics.

12 A more precise classification of historical networks is proposed in Martin Grandjean, "Analisi e visualizzazioni delle reti in storia. L'esempio della cooperazione intellettuale della Societa delle Nazioni," *Memoria e Ricerca* 55, no. 2 (2017): 371–393.

Axis 2: What Uses and Which Audiences? Demonstration, Interface, and Research

Is a graphic created for the purpose of synthesizing a historical object intended for an audience? Or is it made in order to understand a massive research object and to crunch data for new knowledge – in which case it will probably only be useful to the historian? This typology of uses between "demonstration" and "research" can be traced back to Tukey[13] and applied to history.[14] They also can be refined, using an intermediate stage that relates particularly to public history. Indeed, visualizations can now serve as an "interface" to explore data and access more information or for the audience to add some data by themselves, in a collaborative process. These three uses are implicitly classified from the simplest to the most complex. First, the "demonstration" visualization is usually straightforward so that the reader can immediately understand, such as a diagram that classifies elements into sets (Fig. 2 "Venn") or a flow diagram for simple relational data (Fig. 2 "alluvial"). For its part, the "interface" visualization can be more complex since the public can interact with it and it is not limited to a unique view, like a timeline, which allows the public to choose the temporality and to display additional information (Fig. 2 "multimedia timeline"), or frequency curves of words, which makes it possible to return to the indexed text (Fig. 2 "ngram viewer"). More and more frequently, these visualizations are turned into dashboards that accumulate indicators, allowing the public to navigate between the screens. Finally, the "research" visualization is sometimes illegible for someone other than its author – as the scholar (often alone) has gone through the whole development of processing and analysis, such as in very dense graph analyses (Fig. 2 "complex network"). Heuristic drawings are quite similar as they represent the organization of concepts or research ideas (Fig. 2 "mind map"), where only the author can understand the intrinsic logic. It should be noted that these three different uses depend on the evolution of technology. For example, the data visualization remained very descriptive or demonstrative until the development of the computation power of the personal computer, making the visualizations more affordable as research tools. And it has been the emergence of web technology, mobile applications, and touch screens that today favor the exploitation of visualization interfaces in online publications, databases, or museums.

13 Tukey (1977).

14 Martin Grandjean, "Introduction à la visualisation de données: L'analyse de réseau en histoire," *Histoire et Informatique* no. 18 (2015): 107–126.

Maintaining a Critical Discussion on Visual Representations

Visualization provides a unique and increasingly important avenue through which to convey historical knowledge. However, sometimes the intended effect is completely spoiled by misinformed or faulty practices. Thus, it remains critical that the graphic representation is not dissociated from the data; we also must provide critical and demanding discourse to support the data modeling process.[15]

The aesthetic and heuristic aspects of visualization often unintentionally reinforce the impression that these graphical representations are self-sufficient and that they can become a substitute for traditional scholarly methods. To visualize is to objectify, to "totalize" a historical object, sometimes giving the impression that complex subjects can be grasped at once. Although this may serve the general public, it is an open door for the positivist temptation to reduce this object to its empirically measurable part. Yet, visualization should not replace the fundamental material of the historian: the figures, the organization charts, and the timelines are hiding the people who make history. This reminder is all the more important because the innovative and extremely specific aspect of certain technologies forces those who use them to isolate themselves in scientific communities that sometimes cut them off from the public. This leads those historians, designers, or cultural institutions to make visualizations that deprive their audience of important critical information and necessary context. Moreover, and this is also a point that needs to be discussed in any research project or public history work, the tools themselves are rarely designed for historical analyses and narratives. Using them in other ways is a rich and fascinating task but exploiting them without discerning that they need to be adapted to the contexts of arguments and audience can lead to irrelevant conclusions.

The recent developments of "digital history",[16] which have been accompanied by a democratization of access to visualization tools, only accentuates such issues. While it is obviously valuable that software resources are not monopolized by a caste of specialists, audiences sometimes lack the analytic or technical skills to use such resources and understand their added value. Network analysis is a classic case of complex and extremely powerful tools that sometimes are used naïvely for projects that wish to benefit from the impressive and complicated expression of deep entanglements evident in such representations. But, for unsophisticated audiences, these graphs may be inexplicable. Adding complexity in this manner may

15 Frederick W. Gibbs, 2016. "New Forms of History: Critiquing Data and Its Representations," *The American Historian*, http://tah.oah.org/february-2016/new-forms-of-history-critiquing-data-and-its-representations/.

16 Serge Noiret and Frédéric Clavert, eds., *L'histoire contemporaine à l'ère numèrique – Contemporary History in the Digital Age* (Bern: Peter Lang, 2013).

ultimately obscure more than it reveals and create an artificial distance with the public. Likewise, it is also common for museums and the media to use visual objects that have the characteristics of a visualization but that do not explain or illustrate the subjects they are meant to elucidate. Instead, they fill empty spaces to impress the audience with a debauchery of aestheticism.

Conclusion: Historians' Responsibilities to Audiences

The main pitfall facing the producers of historical data visualization is probably that they are not always able to predict the reception of their audience and design a product that matches their needs. Although it is now common to encounter graphic representations of statistical data in mass media publications, the use of such objects is still relatively underdeveloped in scientific publications in history or in their popularization to the general public. It is often limited to very simple and intuitive forms: histograms, curves, or geographical maps. Consequently, using a visualization resulting from complex data processing, such as a multiple correspondence analysis or a network graph, whose codes are unfamiliar to the majority of people, often produces mixed reactions. Some do not understand or refuse these visualizations as a means of proof. Others may be so fascinated by the object, its attractive aesthetic, and the impression of completeness, that they accept the interpretation without questioning the modeling choices. Problems also arise when interactive data visualizations are proposed to the public. In practice, many museum visitors confronted with an interactive audiovisual interface or Internet users who browse a heritage database on a website are lost after the first click because the interface was not designed with the user in mind. In some cases, such visualizations may offer so much browsing freedom that the visitor is immediately disoriented or overwhelmed.

In the end, all these reactions are the product of the same cause, a form of "visual illiteracy", an inability to read these graphs and understand the visualization issues that underlie them. However, it is not a question of ridding ourselves of our responsibility by blaming audiences. On the contrary, it is crucial to take into account the fact that they are not always educated to read such objects. Audiences, therefore, must be accompanied – often metaphorically through text or user design – with the historians' guidance when decrypting these graphic representations. For public historians, whose work is so much a part of historical dialogues with diverse audiences, this is especially true. And as with great powers come great responsibilities, we have the mission to produce visualizations that are up to the rigor of historical sciences, respecting their data on the one hand and the public of readers on the other.

Bibliography

Cairo, A. *The Functional Art: An Introduction to Information Graphics and Visualization*. Indianapolis: New Riders, 2012.

Drucker, J. *Visualization and Interpretation. Humanistic Approaches to Display*, Cambridge: MIT Press, 2020.

Graham, S., I. Milligan, and S. Weingart. *Exploring Big Historical Data, The Historian's Macroscope*. London: Imperial College Press, 2015.

Kee, K., and T. Compeau, eds. *Seeing the Past with Computers: Experiments with Augmented Reality and Computer Vision for History*. Ann Arbor: University of Michigan Press, 2019.

Rosenberg, D., and A. Grafton. *Cartographies of Time: A History of the Timeline*. New York: Princeton Architectural Press, 2010.