

Quantifying Culture: The Value of Visualization inside (and outside) Libraries, Museums, and the Academy

Chris Alen Sula
Pratt Institute, School of Information & Library Science
144 W 14th Street, 6th Floor
New York, NY 10016
csula@pratt.edu

Maps, diagrams, illustrations, and other visual materials have long been part of cultural institutions, as well as the academic disciplines of the arts, sciences, and humanities. In the past several years, these visual materials have been increasingly centred on quantitative data, with sensors, geotags, social networks, and “big data” now occupying the forefronts of research and public engagement. With this use of quantitative data comes the need for more sophisticated and adequate visual representations, particularly through the field of information visualization (i.e., infovis). In this paper, I explore five ways in which infovis can enrich the visual culture of libraries, museums, and the academy: (1) digital, interactive visualizations can take advantage of linked data to provide participants with richer, contextualized experiences; (2) high-volume, longitudinal datasets can be seen from a macroscopic perspective, in which patterns, processes, and systems-level phenomena all become visible; (3) the cognitive science foundations of infovis help produce designs that extend working memory and amplify cognition, allowing many viewers to grasp large, complex data for the first time; (4) the empirical foundations of quantitative data collection help to reduce biases in representing events; and (5) this empirical validity helps to produce visualizations that are more ethical in the sense that they are more inclusive of various groups and disinterested on the whole—the victors can still write history, but only insofar as they can measure it (and cannot avoid *all* measurements of it).

*Quantitative data. Information visualization. Cultural heritage. Digital humanities.
Linked data. Macroscopic, Representation. Ethics.*

1. INTRODUCTION

Maps, diagrams, illustrations, and other visual materials have long been part of cultural institutions. In fact, much of the work of cultural heritage in the past decade has gone into digitizing, preserving, and liberating these materials for use by the public. So too have academic disciplines relied on visual materials to document, defend, and disseminate their work. One need only think of the vast influence of weather maps for bringing meteorology into public awareness, or the value of photographs, illustrations, and moving images for creating a more public history of labor movements, human rights struggles, and political resistances—especially ones that challenge established and official narratives of power.

In the past several years, these visual materials have been increasingly centered on *quantitative data*, with sensors, geotags, social networks, and “big data” now occupying the forefronts of research

and public engagement, in tandem with the massive digitization of print documents. Birth certificates and census records are, in one sense, textual documents. Their characters may be decoded and their contents, transformed into large datasets of demographic information; even handwritten records may now be automatically minded for data. But at the same time, these records are also visual materials, whose structural layouts and aesthetic conventions allow for such datamining to take place. There is a *field* for “given name” or “race” or more administrative metadata, such as record number or preparer. So-called “born digital” documents are even richer in quantitative information. Many photos, tweets, and posts now carry embedded geospatial data, and the platforms that host them capture relationships between people and groups, forming large-scale social networks, the documentation of which is unprecedented in human history.

With this use of quantitative data comes the need for more sophisticated and adequate visual representations. In many of these cases, one simply *cannot* process such high-volume longitudinal data in a textual form. Attempting to do so would exhaust cognitive resources of working memory and attention long before detailed relationships and patterns could be extracted. Think of the difficulty of understanding even one major artist's trajectory as understood through a collection of a few hundred works. The art *community* may spend years, even decades attempting to construct such a narrative, let alone situate that narrative within the context of other artists and thousands of their works. And all this for only one medium! There are countless writings, sounds, and other information that could be brought to bear on the problem, many of them in digital form through "linked data" or the semantic web. Though the metadata standards and technologies for implementing them may still be years away, such possibilities are on the horizon, and there are already more specialized topics for which curated, linked data exists. The point of this thought experiment is not to overwhelm the tasks of interpretation, knowledge creation, and aesthetic enjoyment; it is simply to bring home the magnitude of information that exists in digital form at present—and the inadequacy of many existing interfaces for capturing that complexity.

The field of information visualization (hereafter, "infovis") can make a helpful intervention here. Infovis, broadly defined, sits at the center of cognitive science, computer visualization, and data analysis. According to Ware (2004), the term 'visualization' is understood as "a graphical representation of data or concepts. It attempts to harness the powers of human cognition for the task of comprehending information, especially large-scale data. Infovis allows viewers to browse through such datasets, noting top-level patterns and trends and often drilling down into more detailed information. Lin (1997) identifies five conditions in which browsing techniques are especially useful:

- when there is a good underlying structure so that items close to one another can be inferred to be similar,
- when users are unfamiliar with a collection's contents,
- when users have limited understanding of how a system is organized and prefer a less cognitively loaded method of exploration,
- when users have difficulty verbalizing the underlying information need, and
- when information is easier to recognize than describe.

All of these have wide application to the visual materials found in cultural institutions and

academic settings, and the last two conditions in particular are especially relevant for quantitative data. The remainder of this paper discusses five ways in which infovis can enrich the visual culture of libraries, museums, and the academy: adding context through linked data, providing a macroscopic perspective. extending working memory and amplifying cognition. reducing biases, and facilitating more ethical and inclusive representations. Each section further introduces infovis (with reference to the added value under discussion) and provides an example in which infovis is capable of enriching cultural information.

2. LINKED DATA

While the major uses of infovis are analytic, Pousman, Stasko and Mateas (2007) identify "casual" uses of information visualization, which include ambient infovis, social infovis, and artistic infovis. These uses engage with a wide spectrum of users including novices; are "repeatable (over weeks and months), or contemplative (a long moment at an art gallery);" and are personally important or relevant to the use. Still, casual uses of infovis deliver insight about information, albeit non-analytic insight.

Casual infovis may be especially well suited for creating interactive displays that draw in linked data to provide participants with richer, contextualized experiences. For example, one might read a digital edition of a Modern novel in an interface that shows images of the natural wildlife of author's time, local agriculture, etc. As a more quantitative example, the New York Public Library's Direct Me NYC 1940 (<http://directme.nypl.org>) combines digitized images of 1940 New York City telephone directories with newly released 1940 Census records, historical maps, and archived headlines from The New York Times. The site also allows patrons to convert residential street addresses of 1940s New Yorkers into precise census enumeration districts (the smallest organizational unit of the census), which enables them to lookup an individual's census record. Visitors can also contribute stories about particular addresses.

3. MACROSCOPY

In 1979, Joël de Rosnay announced the coming of a the microscope, "a symbolic instrument made of a number of methods and techniques borrowed from very different disciplines...It is not used to make things larger or smaller but to observe what is at once too great, too slow, and too complex for our eyes." *The Macroscope* examined topics of energy and survival, information and society, and time and evolution—systems-level phenomena that elude casual perception and exhibit various

interaction effects between their constituent elements. Macroscopy, however, has a much wider scope, and infovis functions as a macroscopic instrument for examining high-volume, longitudinal datasets and making visible patterns, processes, and systems-level phenomena. Additional reflections on macroscopy in the humanities are found in Sula (2012).

Two examples of infovis macroscopy applied to cultural materials are Mapping the Republic of Letters: Exploring Correspondence and Intellectual Community in the Early Modern Period (1500–1800), located at Stanford University (<http://republicofletters.stanford.edu>), and this author's Phylo project (<http://phylo.info>), co-founded with David Morrow. Both projects are based in primary source archival documents (letters, dissertations, faculty records) and leverage mapping and network analysis techniques to trace interactions across intellectual networks. The Republic of Letters currently focuses on over 2,000 correspondents who formed a communication network across Europe, Asia, Africa, and the Americas. Phylo combines various data sources, user-submitted information, and visual analytics to advance the study of the discipline of philosophy. It traces the flow of ideas across time by documenting the people, places, and institutions associated with philosophy, and currently contains information on over 17,000 philosophers, mainly from twentieth-century North America.

A key, macroscopic component to both of these projects is their focus on high-level trends, structures, and patterns, rather than the individuals that compose and exist within those larger elements. Such visualization is no substitute for detailed analysis but rather an alternative method for understanding the phenomena at hand. The thousands of individuals and hundreds of thousands of connections between them (perhaps even millions) could not be apprehended in textual form, yet visualization renders them quite saliently at a glance.

4. EXTENDING MEMORY AND AMPLIFYING COGNITION

As previously noted, infovis aids human understanding of complex datasets that cannot always be comprehended in textual form. This function is facilitated by the cognitive science foundations of infovis, which attempt to harness quick perceptual systems for the purpose of processing information. Card, Mackinlay and Shneiderman (1999) even *define* 'visualization' as "the use of computer-supported, interactive visual representations of data to amplify cognition." In discussing this definition, they list a number of

ways in which visualizations can accomplish this, including increasing memory and processing resources available, reducing search for information, enhancing the recognition of patterns, enabling perceptual inference operations, using perceptual attention mechanisms for monitoring, and encoding information in a manipulable medium. According to Larkin and Simon (1987), many of these benefits are achieved by substituting rapid perceptual inferences for more difficult logical ones. This substitution is made possible by preventive processing, low-level tasks in the human visual system that occur in less than 200–250 milliseconds from the time an observer is exposed to a visual stimulus. Healey (2009) summarizes a range of psychology experiments that have used preattentive visual features to perform the following preattentive tasks:

- *target detection*: users rapidly and accurately detect the presence or absence of a "target" element with a unique visual feature within a field of distractor elements,
- *boundary detection*: users rapidly and accurately detect a texture boundary between two groups of elements, where all of the elements in each group have a common visual property,
- *region tracking*: users track one or more elements with a unique visual feature as they move in time and space, and
- *counting and estimation*: users count or estimate the number of elements with a unique visual feature.

Though all of these are relevant for identifying general trends, patterns, outliers, and differences in visual displays, they are particularly well-suited for quantitative information, which relies on differences in magnitude between values.

An added bonus of preattentive processing in visualizations is the ability of many viewers to grasp large, complex datasets for the first time. As Plaisant (2004) observes, visualizations "may be able to formulate and answer questions they didn't anticipate having before looking at the visualization." This characterization is reflected in Franco Moretti's *Graphs, Maps, Trees: Abstract Models for a Literary History* (2005). As opposed to the close readings of a single text that typify literary scholarship, Moretti employs a "distance reading" method: "instead of concrete, individual works, a trio of artificial constructs—graphs, maps, trees—[is used] in which the reality of the text undergoes a process of deliberate reduction and abstraction...fewer elements, hence a sharper sense of overall interconnection. Shapes, relations, structures. Forms. Models." In particular, Moretti's graph of the rise of the novel in Britain and Japan (18th Century), Italy and Spain (19th Century), and Nigeria (20th Century) provokes new questions for

him about the development of the genre and the underlying forces of industrialization that account for this robust trend. “[M]ost radically,” he notes, “we see them [visualizations] *falsify* existing theoretical explanations, and ask for a theory.”

5. BIAS REDUCTION

The problem of bias has long been discussed in reference to acts of collection and curating, especially where cultural materials are concerned. Whether this problem is as well recognized with quantitative data is less certain. However neutral or objective a dataset or collection purports to be, there may be certain residual biases in measurement design, modeling techniques, or background assumptions about the phenomena observed. Cathy Davidson (2008) puts the point more strongly in saying, “Data transform theory; theory, stated or assumed, transforms data into interpretation. As any student of Foucault would insist, data collection is really data selection. Which archives should we preserve? Choices based on a complex ideational architecture of canonical, institutional, and personal preferences are constantly being made.”

Infovis, rather than ameliorating this problem, brings it to the fore in visual representations. If large portions of continuous data are missing or a significant number of outliers present, such omissions or deviations will be visible in faithful representations. As Huff (1954) pointed out long ago, it is always possible to lie with statistics, and so, too, is it possible to lie with the datasets that form the basis of visualizations. In this respect, a more robust “ethics of visualization” is needed. At present, it is worth noting that the shift toward quantitative data provides a level of empirical verifiability that is not found in many non-quantitative forms of visualization. This shift provides any wronged parties a framework in which to question inferences and conclusions, seek redress, and present counter-narratives to restore balance—not unlike cases of human rights violations, though of significantly less magnitude.

In addition to the injunction, “When counting, count everything (relevant),” an ethics of visualization would also likely include the rule, “Visualize uncertainties as well as certainties.” Clues to facilitating this latter task, are found in a study by Skeels, *et al.* (2010), which examined uncertainty in 18 different subject domains (humanities is largely absent) and developed five cross-domain categories for understanding uncertainty: measurement, completeness, inference, credibility, and disagreement. Though visual techniques may not be able to address all types or degrees of uncertainty, they can represent many of them more

fully than statistical measures (esp. measures of central tendency), helping to reduce the impression that findings are determinate or at least more certain than they are.

6. MORE ETHICAL AND INCLUSIVE REPRESENTATIONS

The empirical basis of infovis data may be able to do more than simply reduce harm through bias, error, and false completeness; it may also be able to produce visualizations that are more ethical in the sense that they are more inclusive of various groups, especially those that go unrepresented or underrepresented in many cases.

A prominent example of this use of visualization is Invisible Australians (<http://invisibleaustralians.org>), which documents Indigenous Australians and thousands of non-Europeans, including Chinese, Japanese, Indians, Afghans, Syrians and Malays, who faced discriminatory laws and policies for not being white. The site draws together government records documenting these people, and attempts to “link together their lives.” While the site currently focuses on more qualitative aspects of these individuals, the quantitative infovis possibilities abound, from frequency charts and line graphs of their history, to geospatial mapping and network graphs making visible their activities and connections. Another example is the Transborder Immigrant Tool, a digital art project by Micha Cardenas and Jason Najarro at the University of California San Diego, which uses cracked Nextel cell phones to track immigrant’s geolocations across the Mexico/U.S. Border. In addition to allowing would-be illegal immigrants access to map information, the application’s creators hope it will “add an intelligent agent algorithm that would parse out the best routes and trails on that day and hour for immigrants to cross this vertiginous landscape as safely as possible.”

7. CONCLUSION

Though academics and cultural institutions are faced with a deluge of digital objects information, the process of presenting such materials is greatly facilitated by information visualization, particular in the case of quantitative data. Infovis holds vast potential for providing context, insight, and perspective of large-scale datasets, and the empirical foundations of such datasets support visualizations that reduce bias and represent individuals, groups, and events more fully. Though significant work remains in developing and preserving such visualizations, the interdisciplinary foundations of the field provide robust ground for the task of quantifying—and visualizing—culture.

8. REFERENCES

- Card, S., Mackinlay, J. & Shneiderman, B. (1999) *Readings in Information Visualization: Using Vision to Think*. Morgan Kaufmann, San Francisco.
- Davidson, C. (2008) Humanities 2.0: Promise, perils, predictions. *PMLA* 123(3), 707–717.
- de Rosnay, J. (1979) *The Macroscope*. Harper & Row, New York.
- Healey, C. G. (2009). Perception in visualization. <http://www.csc.ncsu.edu/faculty/healey/PP/> (3 April 2012).
- Huff, D. (1954) *How to Lie with Statistics*. Norton, New York.
- Larkin, J. H., and Simon, H. A. (1987) Why a diagram is (sometimes) worth 10,000 words. *Cognitive Science* 11, 65–100.
- Lin, X. (1997) Map displays for information retrieval. *Journal of the American Society for Information Science* 48(1), 40–54.
- Moretti, F. (2005) *Graphs, Maps, Trees: Abstract Models for a Literary History*. Verso, Brooklyn, New York.
- Plaisant, C. (2004) The challenge of information visualization evaluation. *Proceedings of Conference on Advanced Visual Interfaces AVI'04*. IEEE, 109–116.
- 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), 1145–52.
- Skeels, M., Bongshin, L., Smith, G. & Robertson, G. G. (2010) Revealing uncertainty for information visualization. *Information Visualization* 9(1), 70–81.
- Sula, C. A. (2012) Philosophy through the macroscope: History, technologies, and representations of the field. *Journal of Interactive Technology and Pedagogy*, 1.
- Ware, C. (2004) *Information Visualization: Perception for Design*, 1st ed. Morgan Kaufmann, New York.