
Speak to the Eyes: The History and Practice of Information Visualization

Author(s): Jefferson Bailey and Lily Pregill

Source: *Art Documentation: Journal of the Art Libraries Society of North America*, Vol. 33, No. 2 (September 2014), pp. 168-191

Published by: The University of Chicago Press on behalf of the Art Libraries Society of North America

Stable URL: <https://www.jstor.org/stable/10.1086/678525>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <https://about.jstor.org/terms>



The University of Chicago Press and Art Libraries Society of North America are collaborating with JSTOR to digitize, preserve and extend access to *Art Documentation: Journal of the Art Libraries Society of North America*

JSTOR

Speak to the Eyes:

The History and Practice of Information Visualization

Jefferson Bailey, *Internet Archive*

Lily Pregill, *New York Art Resources Consortium*

Abstract—Information visualization techniques are being used increasingly by scholars, museum curators, and collection managers to analyze cultural heritage data sets in novel and dynamic ways. Shifting palettes, spatial density, and other material aspects of works can now be examined digitally to provide new insights into creativity, form, genre, and change. Cultural heritage professionals are also beginning to use visualizations and computational tools to expand the availability and explorability of their collections. This article locates the current field of information visualization within its historical context, demonstrating the shift in aesthetic practice within the field from the mid-eighteenth century to the present day. A number of current projects are presented to illustrate how information visualization is mediating formal humanities research and the study and management of collections.

[This article is based on papers presented at “*The Visual Language of Data: Reshaping Humanities Research*” session at the ARLIS/NA Conference held in Pasadena, CA, in April 2013.]

“... the best way to capture the imagination is to speak to the eyes.”—William Playfair¹

INTRODUCTION

Applying digital visualization techniques to cultural heritage data sets is celebrated as a new and innovative research methodology. However, mapping data to visual representations has been used for centuries to reveal patterns, to communicate complex ideas, and to tell stories—even the story of art. Perhaps the most

Jefferson Bailey is partner specialist and program manager at the Internet Archive, San Francisco, California; jeffersonbailey@gmail.com. Lily Pregill is a consultant working with the New York Art Resources Consortium and The Museum of Modern Art; lily@technelib.com.

1. William Playfair, *Elemens de statistique* (Paris: s.n., 1802), XX, <http://books.google.com/books?vid=BCUL1094754581>.

Art Documentation: Journal of the Art Libraries Society of North America, vol. 33 (fall 2014)
0730-7187/2014/3302-0002 \$10.00. Copyright 2014 by the Art Libraries Society of North America. All rights reserved.

famous (and most debated) art historical example of information visualization is the chart created by Alfred H. Barr Jr., founding director of The Museum of Modern Art, illustrating the development of modern art from 1890 to 1935 (Figure 1).² Used as the dust jacket image for the 1936 exhibition catalog *Cubism and Abstract Art*,³ the diagram deconstructs art movements by mapping genres, cities, and artist names onto a grid of time and into categories of organic or geometric abstract art. Barr's powerful chart illustrates and communicates the evolution of modern art using a number of visualization principles: proximity (positioning related movements close to each other), color (highlighting external influences), size (increasing type size for historical relevance), and connectedness and directionality (indicating paths of influence).⁴

This map of modern art is only one of many examples from a long and rich history of information visualization that can be traced back to the use of tree diagrams as a visual classification system in illustrated manuscripts and deployed as a medieval mnemonic technique.⁵ While the field has ancient roots, development of modern practices began in the mid-eighteenth century with the advent of statistical graphics and continues on into the present day using advanced computational techniques. Twenty-first-century humanities scholars find themselves in the midst of a visualization renaissance of sorts with information analysis and visualization literacy recognized as fundamental skills in the academy.⁶ While fluency with digital tools and techniques is essential, knowledge of the history of the field of information visualization, principles of graphical representation, analog antecedents such as Barr's mapping of modern art, and the contemporary theoretical context in which visualization is taking place, are all equally important in framing current practice. As professor Michael Friendly aptly states: "There certainly *have* been many new things in the world of visualization; but unless you know its history, everything might seem novel."⁷

The technology-driven discipline that is recognized today as information visualization emerged in the 1980s alongside the development of computer graphics programs, with the first annual IEEE Conference on Visualization held in 1990.⁸ Edward

2. Jennifer Tobias, reader services librarian at MoMA, maintains a bibliography on Barr's chart at https://www.zotero.org/groups/moma_charts/items. Barr's visual language continues to resonate: see the "Graffiti and Street Art" diagram (also referred to as the "Feral Diagram") by Daniel Feral. The image first appeared in the March 29, 2011, press release for *Pantheon: A History of Art from the Streets of New York City*, an exhibit at the Donnell Library Center of the New York Public Library, https://wayback.archive-it.org/4472/20140701210222/https://dl.dropboxusercontent.com/u/12523908/PANTHEON%20Press%20Release/PANTHEON-PR4_X_032911.pdf.

3. *Cubism and Abstract Art* (New York: Museum of Modern Art, 1936).

4. For a complete analysis of Barr's chart see Edward Tufte, *Beautiful Evidence* (Cheshire, CT: Graphics Press, 2006), 65–68.

5. For a comprehensive history of the use of tree metaphors as a visualization technique, see Manuel Lima, *The Book of Trees: Visualizing Branches of Knowledge* (New York: Princeton Architectural Press, 2014).

6. Maureen Stone, "Information Visualization: Challenge for the Humanities," in *Working Together or Apart: Promoting the Next Generation of Digital Scholarship: Report of a Workshop Cosponsored by the Council on Library and Information Resources and the National Endowment for the Humanities*, CLIR Publication, no. 145 (Washington, DC: Council on Library and Information Resources, 2009), 43–56.

7. Michael Friendly, "Milestones in the History of Thematic Cartography, Statistical Graphics, and Data Visualization," August 2, 2009, <https://web.archive.org/web/20140709231808/http://www.math.yorku.ca/SCS/Gallery/milestone/milestone.pdf>, as quoted in Lima, *The Book of Trees*, 9.

8. Arie Kaufman, ed., *VIS90 Visualization '90, San Francisco, CA, October 23–26, 1990* (Los Alamitos, CA: IEEE Computer Society Press, 1990).

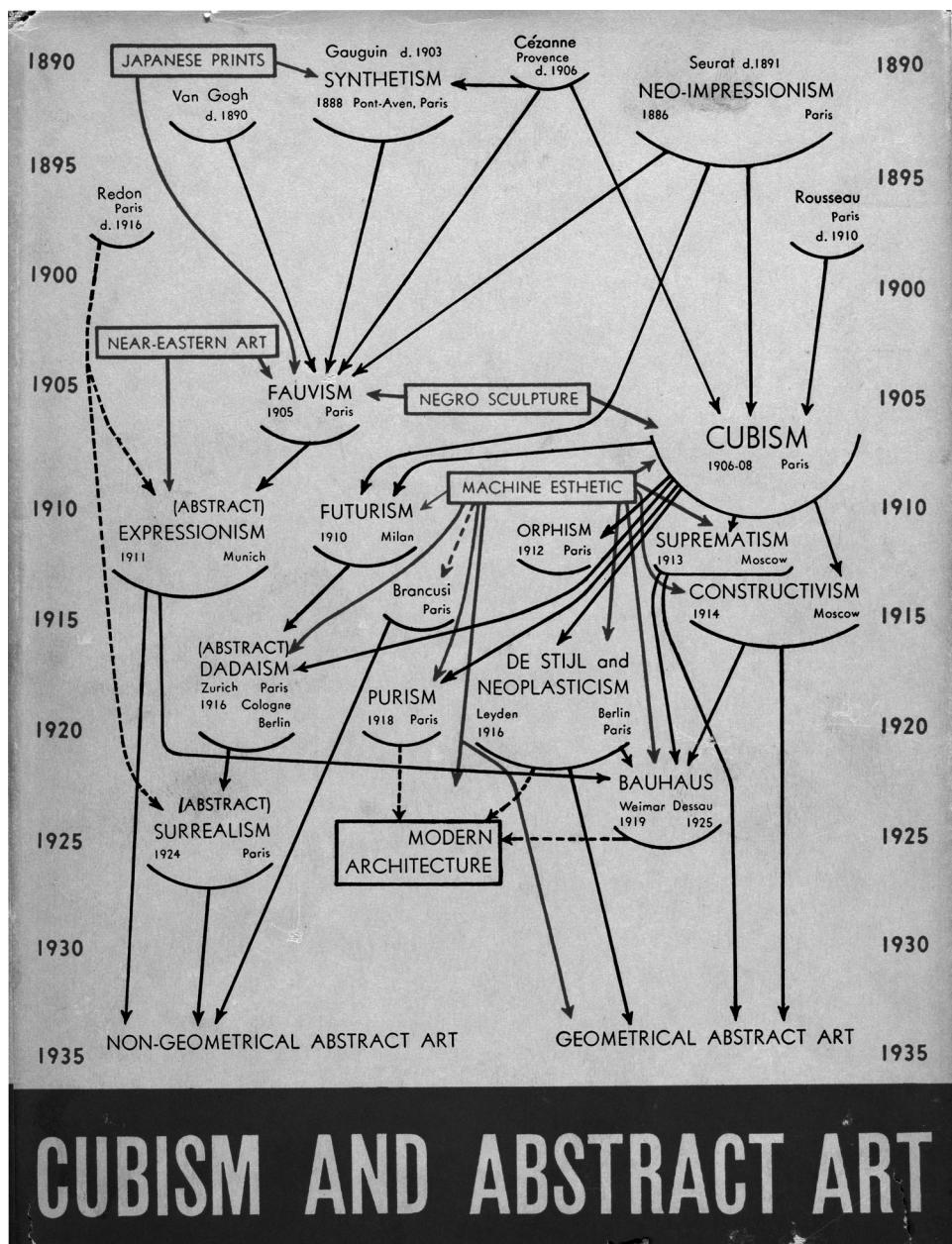


Figure 1. Dust jacket with chart prepared by Alfred H. Barr Jr., of the exhibition catalog, *Cubism and Abstract Art*, by Alfred H. Barr Jr., 1936. Offset, printed in color; 10 1/8 x 7 3/4" (25.7 x 19.7 cm). The Museum of Modern Art Library, New York. Digital Image © The Museum of Modern Art/Licensed by SCALA / Art Resource, NY. Used with permission. Please see the online edition of *Art Documentation* for a color version of this image.

Tufte, professor emeritus of political science, computer science, and statistics at Yale University, published *The Visual Display of Quantitative Information* in 1983, a groundbreaking work that predated the formal discipline. Called “the da Vinci of

data,”⁹ Tufte has become one of the most influential scholars in the field of information design and visualization. His work examines the history of statistical graphics and provides a practical theory for visual displays of information. Any historical treatment of this subject owes much to Tufte’s work. The goal of this article is to provide choice examples from that history, focusing on the ongoing evolution of aesthetic practice within the field. This historical background provides context for the subsequent examination of contemporary digital visualization use cases from the digital humanities, art and technology communities, and the information sciences. This arrangement connects the past with the present and points to future directions for visualization-based research in the humanities.

WILLIAM PLAYFAIR: LINE GRAPHS, BAR CHARTS, AND PIE GRAPHS

Graphical methods for displaying data grew out of the rise of economic and social statistics in the mid-eighteenth and early nineteenth centuries. Many individuals contributed to the growth of the field during this time, including John Snow (1813–1858) and Florence Nightingale (1820–1910), both champions of the use of graphics to advance public health. However, William Playfair (1759–1823), a Scottish engineer and political scientist, is credited as being the earliest pioneer of statistical graphics, first publishing on the topic in 1786. Playfair invented three standard visualization techniques that are still ubiquitous today: line graphs, bar charts, and pie graphs. A plate from *The Statistical Breviary* (Figure 2) illustrates his innovation. Playfair maps the area of each country (circle), the country’s population in millions (left line), and taxes collected in millions of pounds sterling (right line). The sloping line connecting the population and taxes is used to illustrate which is higher. At a glance, this chart clearly shows Britain and Ireland as being most burdened by taxes. Tufte cites this chart as being notable on three accounts: 1) the graphic is an early example charting multivariate data; 2) it uses area to show quantity; and 3) it is the first known use of the pie chart, as seen in the proportions of the Ottoman (“Turkish”) Empire.¹⁰ Playfair developed graphical forms for plotting data to demonstrate evidence and communicate economic statistics in an effective way by “speaking to the eyes.”

CHARLES JOSEPH MINARD: VISUAL STORYTELLING

The premiere example of visual storytelling was produced by the French engineer Charles Joseph Minard (1781–1870). In his *Carte Figurative*, Minard uses casualty data to illustrate the fate of Napoleon’s Grand Army in Russia throughout the campaign of 1812 (Figure 3). The viewer traces the troop size by following the tan line from the army’s starting point at 422,000 men strong at the left through a series of losses over time and space as the soldiers march towards Moscow. The black line indicates their retreat and continued rank reduction. A dramatic loss of 22,000 lives over the Berezina

9. Deborah Shapely, “The da Vinci of Data,” *New York Times*, March 30, 1998, <https://web.archive.org/web/20140630232219/http://www.nytimes.com/1998/03/30/business/the-da-vinci-of-data.html?pagewanted=2&src=pm>.

10. Edward Tufte, *The Visual Display of Quantitative Information* (Cheshire, CT: Graphics Press, 1983), 44.

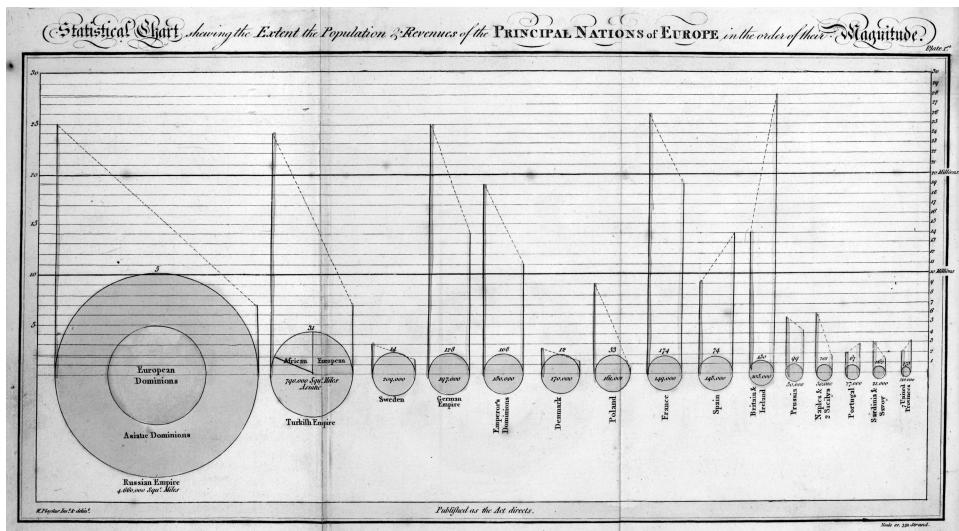


Figure 2. "Statistical Chart Showing the Extent of the Population and Revenues of the Principal Nations of Europe in the Order of their Magnitude," Plate 1, from *The Statistical Breviary*, 1801. The Ottoman Empire pie chart is the second from the left. Courtesy of The Thomas Fisher Rare Book Library, University of Toronto. Please see the online edition of *Art Documentation* for a color version of this image.

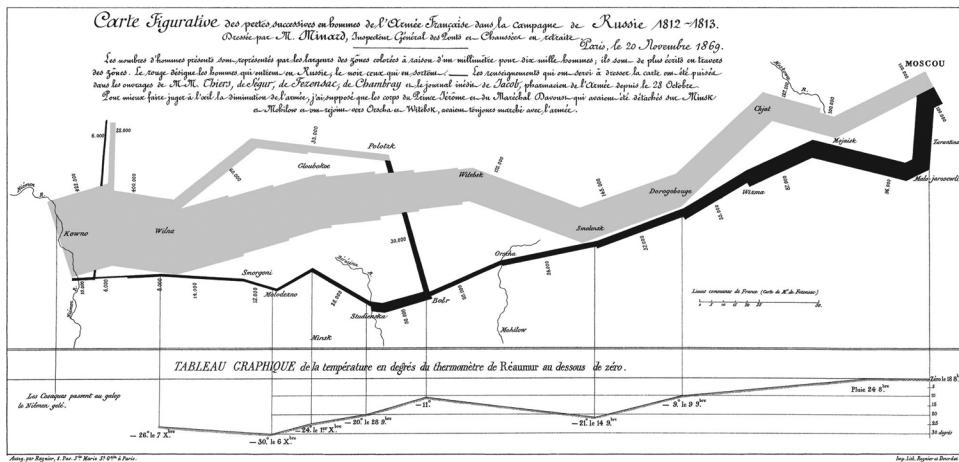


Figure 3. Charles Joseph Minard, *Tableaux Graphiques et Cartes Figuratives de M. Minard, 1845-1869*. 1869. Public domain. Source: Wikimedia Commons, via Internet Archive, <https://web.archive.org/web/20140301150547/> <https://commons.wikimedia.org/wiki/File:Minard.png>. Please see the online edition of *Art Documentation* for a color version of this image.

River is strikingly apparent as the dark line narrows almost by half. Along the bottom of the graphic, the temperature plummets to thirty degrees below zero as the army made its way back to Kovno with only 10,000 men.

This graphic depicts six different dimensions: latitude, longitude, direction of movement, time, temperature, and the size of the Grand Army. Minard's achievement here is expertly illustrating a complex, multivariate time-space story in a

compelling and concise display. Tufte has famously endorsed it as the “best statistical graphic ever drawn.”¹¹ Historian E.J. Marey, a contemporary of Minard, praised the map as “seeming to defy the pen of the historian by its brutal eloquence.”¹²

OTTO NEURATH: ISOTYPE

In the 1920s Otto Neurath (1882–1945), an Austrian mathematician, sociologist, and statistician, introduced a new visual language to information graphics. Neurath was a founding member of the Vienna Circle, a group of philosophers interested in the democratization of knowledge and a renewal of the Enlightenment spirit in the sciences.¹³ For Neurath, developing a common symbolic language was essential to this goal. Neurath believed that statistical information and abstract thought could be presented with clarity as a universal language of pictograms that all people, regardless of their cultural or educational background, could easily understand.¹⁴ As founder and director of the Gesellschafts- und Wirtschaftsmuseum in Vienna, Neurath had both the raw data and public audience to explore his visual pedagogical theories. It was at the museum that he developed his pictorial language called ISOTYPE, an acronym for International System of Typographic Picture Education, which uses simple icons to represent data.

The ISOTYPE technique was very influential and used internationally into the 1940s. A graphic included in The Museum of Modern Art’s 1939/1940 annual report (Figure 4) serves as an example of this new pictorial language. Here it is used to classify and quantify the museum’s acquisitions during that fiscal year. The Pictograph Corporation, whose founder Rudolph Modley studied under Neurath, created the charts for this report.

During his tenure at the museum, Neurath honed his chart making into a collaborative process that he called transformation. The team involved in creating charts included the director, transformers, artists, and technicians. The transformer played the key bridge role, working with experts to understand the data and making decisions about how best to convey it to the public. Scholars from across a variety of disciplines, including statistics, industrial management, and art history, were often consulted for their advice. The transformer would distill the data and develop a blueprint for the artist to develop into a finished layout.¹⁵ This collaborative spirit in which multiple areas of expertise are brought together to inform data-driven work continues today in the digital humanities centers and collaborative research projects that pair technologists, information professionals, and scholars in project-based work facilitating access and analysis of humanities materials.

Neurath’s work represented the emergence of a new visual style for representing data. Rather than using abstractions such as points on a graph or lines in a chart,

¹¹. Ibid., 40.

¹². Étienne-Jules Marey, *La méthode graphique dans les sciences expérimentales et principalement en physiologie et en médecine* (Paris: G. Masson, 1878), 73, <http://vlp.mpiwg-berlin.mpg.de/library/data/lit3585>.

¹³. Amelie Rorty, ed., *Philosophers on Education: New Historical Perspectives* (New York: Routledge, 1998), 418–21.

¹⁴. Otto Neurath, “Visual Education: A New Language,” *Survey Graphic* 26, no. 1 (1937): 25.

¹⁵. Marie Neurath and Robin Kinross, *The Transformer: Principles of Making Isotype Charts* (London: Hyphen Press, 2009), 77–78.

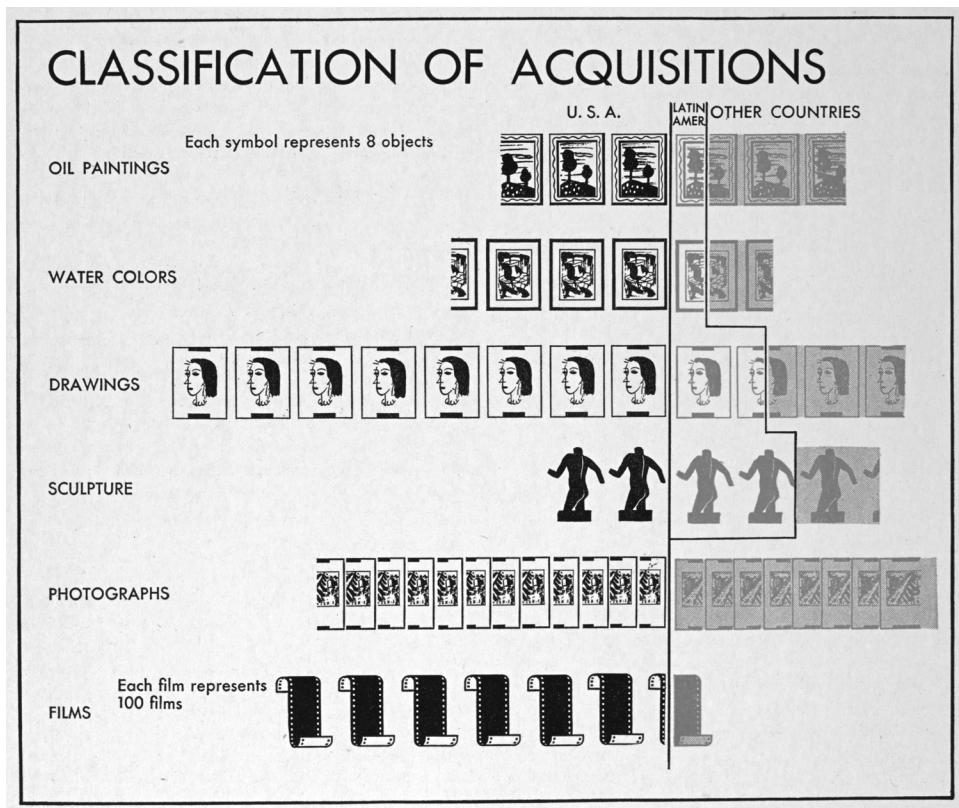


Figure 4. An example of icons used in statistical graphics. "Classification of Acquisitions," page 35 from "The Year's Work, Annual Report to The Board of Trustees and the Corporation Members of The Museum of Modern Art, for the Year June 30, 1939–July 1, 1940." 1940. Reports and Pamphlets, 9.2. The Museum of Modern Art Archives. Digital Image © The Museum of Modern Art/Licensed by SCALA / Art Resource, NY. Used with permission. Please see the online edition of *Art Documentation* for a color version of this image.

ISOTYPE transforms data into universal symbols. His charts are friendly, simple, and clear; they were designed for mass consumption.

NIGEL HOLMES: INFOGRAPHICS

Referring to Neurath as "probably the biggest single influence on my work and thinking," Nigel Holmes introduced the infographic to popular culture during the late 1970s.¹⁶ While earlier examples of visualizations can be found in the pages of *Fortune* and scattered among other publications, Holmes's charts had a significant impact on the field. Holmes worked at *Time* magazine from 1978 to 1994, creating what he called "explanation graphics" to accompany the reporting. Using humor and emotion in his illustrative charts, Holmes presented data in memorable and visually arresting ways to capture the attention of the reader (Figure 5).

¹⁶ Steven Heller, *Nigel Holmes: On Information Design* (New York: Jorge Pinto Books, 2006), 12.

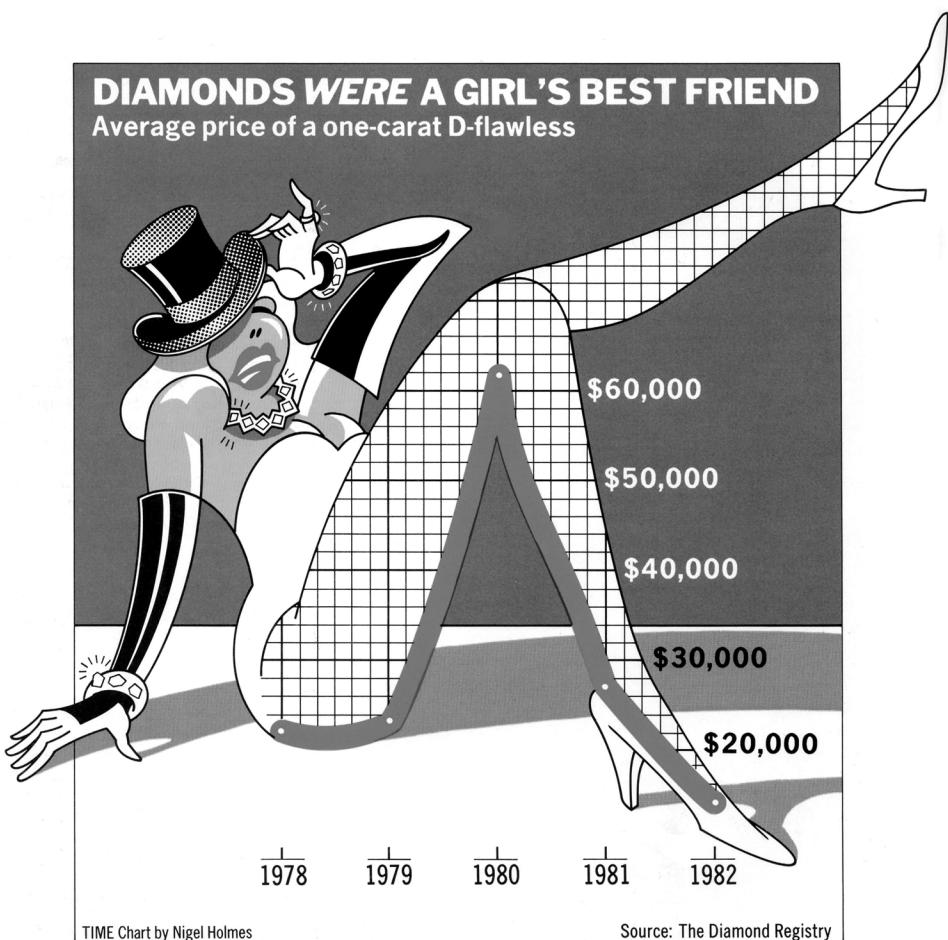


Figure 5. “Diamonds Were a Girl’s Best Friend.” Chart by Nigel Holmes in “A Gem That Lost Its Luster,” *Time* 120, no. 9 (August 30, 1982): 66. Used with permission. Please see the online edition of *Art Documentation* for a color version of this image.

Tufte famously points to this “unsavory exhibit” as “chartjunk.”¹⁷ Chartjunk is a term he coined to characterize graphics embellished with unnecessary decoration that detracts from the data. Graphical efficiency, one of Tufte’s primary principles, is presented in his data-ink ratio. This ratio calculates the proportion of ink used in a graphic to display data information in relation to the total amount of ink used. Successful designs are those that keep the ratio as high as possible with all superfluous graphical elements omitted.¹⁸ Beyond breaking this cardinal rule, Tufte asserts that chartjunk designers show “contempt both for information and for the audience.”¹⁹

¹⁷ Edward Tufte, *Envisioning Information* (Cheshire, CT: Graphics Press, 1990), 34. The publication reproduces the chart but does not credit Holmes or cite the source of the image.

¹⁸ Tufte, *Visual Display*, 91–121.

¹⁹ Tufte, *Envisioning Information*, 34.

Holmes disagrees: "A good approach to information graphics includes an appeal to the reader, immediately followed by a true account of the story . . . I want to make room for enjoyment, delight, aesthetic appreciation and wit, and a friendly 'you can understand this' approach."²⁰

The debate as to whether chartjunk aids or hinders understanding in information graphics continues. A 2010 study measuring the interpretation and recall of Holmes-style charts and plain visualizations questions the minimalist approach to chart design.²¹ The researchers found that chartjunk had no impact on reading comprehension of the information presented in the chart. Furthermore, the study found that recall of the embellished charts was much better than that of the plain charts following a two-to-three-week gap. In essence, chartjunk makes charts more memorable. This evidence suggests that visual embellishments may benefit the reader, and an austere style may not be the best approach to chart design for all publications or contexts.

BEN SHNEIDERMAN: TREEMAPS

Technological advances in the 1990s shifted the field from paper into the digital realm, sparking the development of powerful information visualization methods. As information overload quickly followed the birth of the Information Age, researchers began inventing dynamic graphic displays to explore information collections. This visual approach leverages the eye's natural abilities to quickly scan, recognize, distinguish, and recall images. The approach also draws upon earlier methods developed for presenting static information, using positioning, color, and directional techniques to convey information. Innovations during this period include dynamic query sliders, fish-eye views, hyperbolic trees, perspective walls, and treemaps.

HistoryWired: A Few of Our Favorite Things, a 2001 virtual exhibition from the Smithsonian Institution's National Museum of American History, illustrates the treemap technique (Figure 6). Treemaps were invented by Ben Shneiderman, a computer scientist and human-computer interaction researcher at the University of Maryland. He also developed the "Visualization Information-Seeking Mantra," namely, "overview first, zoom and filter, then details on demand."²² This mantra can be seen in practice in *HistoryWired* where the map provides an overview of 450 museum objects. The objects are divided into regions by broad categories, and object details can be viewed by clicking on individual squares. Direct manipulation, immediate feedback, linked displays, and dynamic queries become key design principles for this new language of digital information visualization.

LEV MANOVICH: DIRECT VISUALIZATION

Advances in processing bandwidth, computer graphics, software applications, and digital visualization techniques during the late twentieth century have led to exciting

20. Heller, *Nigel Holmes*, 78.

21. Scott Bateman, Regan Mandryk, Carl Gutwin, Aaron Genes, David McDine, and Christopher Brooks, "Useful Junk? The Effects of Visual Embellishment on Comprehension and Memorability of Charts," in *Proceedings of ACM Conference on Human Factors in Computing Systems (CHI 2010)* (New York: ACM, 2010), 2573–82.

22. Ben Shneiderman, "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualization," in *Proceedings: IEEE Symposium on Visual Languages: September 3–6, 1996, Boulder, Colorado* (Los Alamitos, CA: IEEE Computer Society Press, 1996), 336.

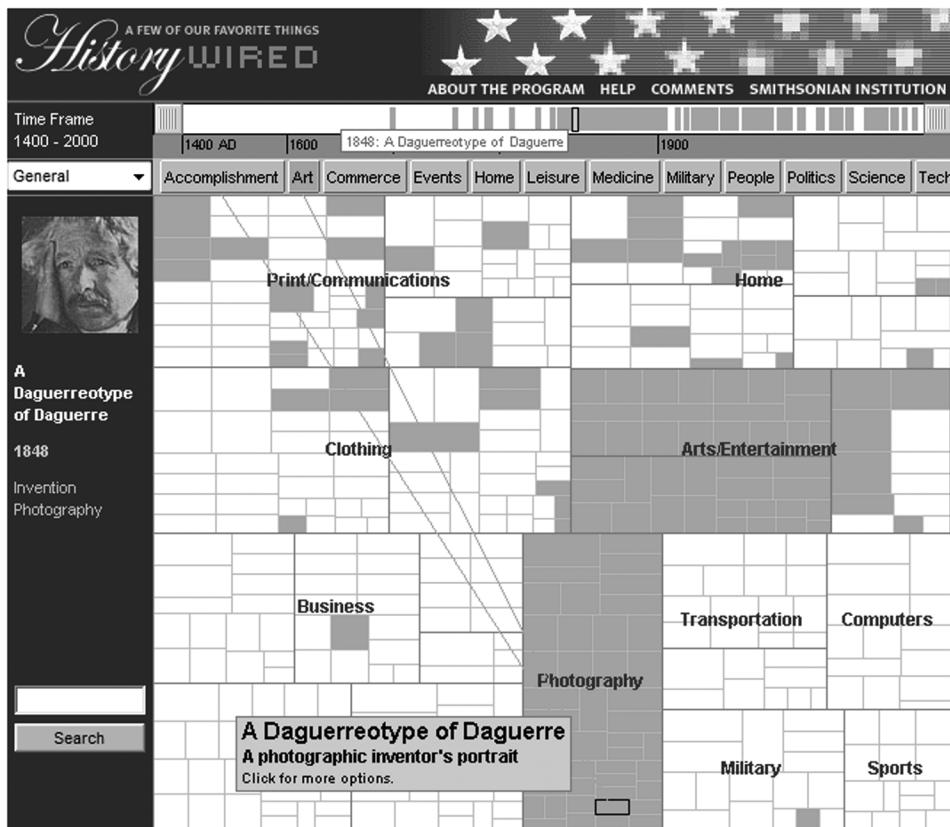


Figure 6. Screenshot from HistoryWIRED, via Internet Archive, <https://web.archive.org/web/20140326035717/http://historywired.si.edu>. Please see the online edition of *Art Documentation* for a color version of this image.

approaches to visualizing cultural heritage data. Lev Manovich, professor in the Computer Science Department at the Graduate Center, City University of New York, promotes a new paradigm of analyzing and visualizing cultural data using computational methods called cultural analytics.²³ His work uses the big data approach—data mining, statistical data analysis, simulation, and information visualization—applied to cultural data sets, including images and video. Manovich uses what he calls the direct visualization method, which breaks with the traditional visual language of information visualization where graphical primitives (points, lines, rectangles, circles, and icons) are used to indicate objects and show relations between them. Direct visualization does not substitute visual symbols for data objects, but builds visualizations out of the original form, though often in miniature or reduced form, which is aggregated in large numbers on a display.²⁴

23. Lev Manovich, "Cultural Analytics: Analysis and Visualization of Large Cultural Data Sets," September 30, 2007, https://web.archive.org/web/20130718043831/http://www.manovich.net/cultural_analytics.pdf, and "Cultural Analytics: Visualizing Cultural Patterns in the Era of 'More Media,'" 2009, https://web.archive.org/web/20140630234908/http://manovich.net/content/o4-projects/o6o-cultural-analytics-visualizing-cultural-patterns/6o_article_2009.pdf.

24. Lev Manovich, "What Is Visualisation?," *Visual Studies* 26, no. 1 (March 15, 2011): 36–49.

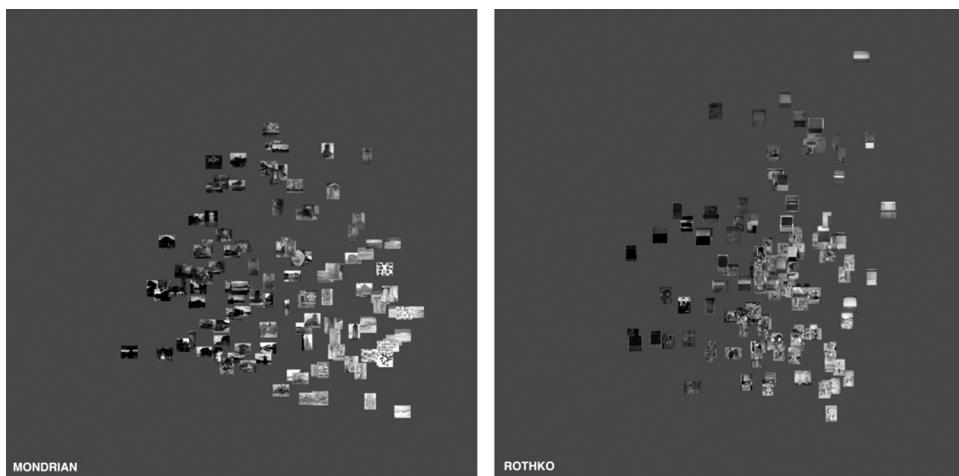


Figure 7. Paintings by Mondrian and Rothko are mapped on image plots. From “Mondrian vs Rothko: footprints and evolution in style space,” <https://web.archive.org/web/20140630022324/http://lab.softwarestudies.com/2011/06/mondrian-vs-rothko-footprints-and.html>. Please see the online edition of *Art Documentation* for a color version of this image.

This technique is shown in Figure 7, which presents 128 paintings by Piet Mondrian and 123 paintings by Mark Rothko as image plots. Using ImagePlot software, miniature images of the paintings themselves are mapped according to their brightness (x-axis) and saturation (y-axis).²⁵ The visualization shows the artists’ palettes, indicating the range of brightness and hue in each of their work. Showing the original data object, in this case the image of the painting rather than a representational surrogate, allows patterns to emerge in their original context that could not be represented in mapping the raw data itself.

Discovering patterns across large data sets in the humanities is also known as a mode of inquiry called distant reading.²⁶ As seen here, the direct visualization approach supports both close reading, selecting an individual image to examine, and distant reading, where one can view a whole set of paintings at once. Manovich questions the usefulness of visualization reductionism in analyzing cultural heritage data: “We throw away 99 percent of what is specific about each object to represent only 1 percent—in the hope of revealing patterns across this 1 percent of objects’ characteristics.”²⁷ This emphasis on both distinct elements and context hearkens back to Rudolph Arnheim’s observations in *Visual Thinking*. He notes that “to lift something out of its context means to neglect an important aspect of its nature.”²⁸

Consideration of context becomes more meaningful when examining contemporary theories and practices of information visualization, especially in the art history

25. ImagePlot is a freely available software tool to visualize image and video collections, <https://web.archive.org/web/20140628114640/http://lab.softwarestudies.com/p/imageplot.html>.

26. For an introduction to distant reading, see Franco Moretti, *Graphs, Maps, Trees: Abstract Models for Literary History* (London: Verso, 2007).

27. Manovich, “What Is Visualisation?,” 38.

28. Rudolf Arnheim, *Visual Thinking* (Berkeley: University of California Press, 1969), 71.

and cultural heritage areas of collection management and fostering and supporting scholarship. Information visualization is now offering researchers methods of exploring visual data sets both by object and by aggregate, and in the process reshaping the visual language of data. It is also providing curators, librarians, and archivists new ways to enhance metadata, expand contextual description, and foster new uses of materials under their stewardship.

In describing the many ways that data visualization can facilitate new ways of analyzing both individual works and aggregated collections, representative examples are perhaps best organized within certain conceptual categories. Three different means by which contemporary information visualization use cases have supported new kinds of inquiry and understanding are pattern analysis, narrative modeling, and collection analysis. This is a broad ontology—new methods of visualization will emerge as new technologies themselves emerge. No single ontology can capture the dynamism and creativity of humanistic analysis even within the realm of information visualization. However, this loose categorization provides an entry point into thinking of ways that visualizing artistic and historical data can spur new methods of scholarship, curatorial practice, collection management, and aesthetic and creative interpretation. The categories can help delineate the shifting, overlapping ways that data visualization can be utilized.

PATTERN ANALYSIS

Exposing outliers, trends, and abrupt changes are all traditional functions of pattern analysis as a visualization strategy. What marks contemporary information visualizations as different from those of previous eras, however, is the scale of data these visualizations can capture and the novel types of patterns they can display. Network diagrams are a method of visualizing relations and interactions between persons or entities. Barr's diagram *Cubism and Abstract Art*, discussed earlier, attempted to use lines to chart the evolution and mutation of artistic genres. This was less a representation of a network, with its assumption of transactional exchange, than it was an abstraction of influence. Network diagrams, instead, have the ability to visually represent actual occurrences, be they citations, purchases, or other modes of interdependence. The Getty Provenance Index has extracted information on art buyers, sellers, auctioneers, and others to analyze the patterns of exchange in the European art market in the early nineteenth century. Using 230,000 records spanning 1802–1820, this visualization accounts not just for individuals and transactions, but also place, type of transaction, and total market activity (Figure 8).

The volume of individual data points contained in this single visualization (zoomable for detailed viewing) goes far beyond simple statistics. In revealing patterns at this scale, it begins to take on a certain autonomy from the data that powers it—that distance in Moretti's formulation of distant reading becomes more pronounced, more profound. Clusters of activity illustrate relationships within the data that simple numbers and statistics might not expose. Patterns reveal larger relationships. For collection managers, these patterns can expose other qualities. Network diagrams and other visualizations of aggregate data can help identify outliers, inconsistencies, or

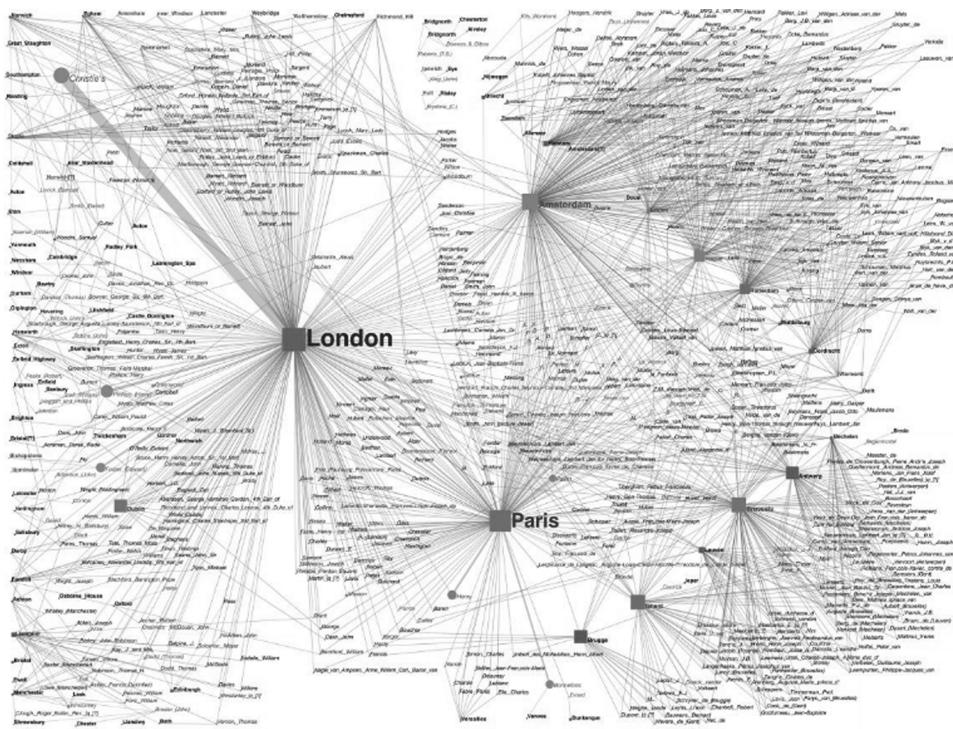


Figure 8. Network diagram of agents connecting the British, Belgian, Dutch, and French auction markets from 1801–1820 using 230,000 records from the Getty Provenance Index databases. © J. Paul Getty Trust and Maximilian Schich. Visualization can be seen at <https://web.archive.org/web/20131105182412/http://www.getty.edu/research/tools/provenance/zoomify/index.html>. Please see the online edition of *Art Documentation* for a color version of this image.

inaccuracies within data that can be indicative of errors in cataloging, uncontrolled taxonomies, or other metadata issues requiring correction.

Though not a traditional data visualization, the project What Makes Paris Look Like Paris? expands the concept of interpreting visual elements to include images as both the subject of scrutiny and as the product of illustration. The project analyzed tens of millions of individual images scraped from Google Street View of Paris and ran algorithms to determine which visual features of the built environment occurred most commonly in the city and which architectural details most evoked the character of the urban landscape (Figure 9).²⁹ This example calls into question the predefined, though not quantifiable, presumptions a user or researcher may have about the defining visual feature that identifies a certain city. While identifying common characteristics has long been the purview of academic analysis, the computational processing power of projects like this have the ability to identify characteristics that may not appear through individual, subjective scrutiny. Specific architectural elements—doorway arches, balcony railing patterns, window shutter styles—are

29. Carl Doersch, Saurabh Singh, Abhinav Gupta, Josef Sivic, and Alexei A. Efros, "What Makes Paris Look like Paris?" <https://web.archive.org/web/20140118140608/http://graphics.cs.cmu.edu/projects/whatMakesParis/>.



Figure 9. Images from What Makes Paris Look Like Paris? <http://youtu.be/s5-30NKSwo8>. Please see the online edition of *Art Documentation* for a color version of this image.

identified as commonalities and come to be more representative of visual identity than what may be more apparent to the classically trained eye.

As the project's creators say, the "look and feel" of a city is defined "largely on a set of stylistic elements, the visual minutiae of daily urban life." The understanding of what makes Paris look like Paris is possible because of this visualization (aggregated instantiations of representative features), but it also relies upon the particular idioms of visualization itself (in this case, street view imagery, which is itself a visualization)—innumerable tiny single images stitched together to give the appearance of a single view of a boulevard or building. Visualization, in this example, is both generative and deductive. The visual features of a place, of created space, assume their own grammar and, through redundancy distilled from a massive amount of image data and juxtaposition, evoke a novel pattern and meaning latent in a set of architectural images.

In a similar project, researcher John Resig used computational image similarity analysis to examine digital images of anonymous (i.e., art lacking attribution) Italian art in the Frick Art Reference Library's Photoarchive (Figure 10). As Resig notes, "Image similarity analysis is an exciting computer vision technique for matching photos whose image content is substantially or completely similar."³⁰ In applying visual analysis tools to a corpus of images and their associated meta-

³⁰ John Resig, "Using Computer Vision to Increase the Research Potential of Photo Archives," <https://web.archive.org/web/20140701125645/http://ejohn.org/research/computer-vision-photo-archives/>.



Figure 10. “All of these binned matches were then passed on to researchers at the Frick Photoarchive for further analysis and record correction.” - John Resig. <https://web.archive.org/web/20140701125645/http://ejohn.org/research/computer-vision-photo-archives/>. Please see the online edition of *Art Documentation* for a color version of this image.

data, Resig was able to discover relationships between works and identify metadata conflicts at a scale that would be impossible to achieve through manual review. His project also generated a visual interface of similar works that subject experts and researchers could then review for correction or annotation. In this instance, visualization, and its underlying algorithms, enables curators and collection managers to review potential matches and update and augment descriptive information. Visual analysis brings computational tools and a preliminary level of automation to metadata enhancement.

NARRATIVE MODELING

“Graphical features organize a field of visual information, but the activity of reading follows other tendencies. These depend on embodied and situated knowledge, cultural conditions and training, the whole gamut of individually inflected and socially conditioned skills and attitudes.”³¹ This tension between the static representation of a

³¹ Johanna Drucker, “Humanities Approaches to Interface Theory,” *Culture Machine* 12 (2011): 6.

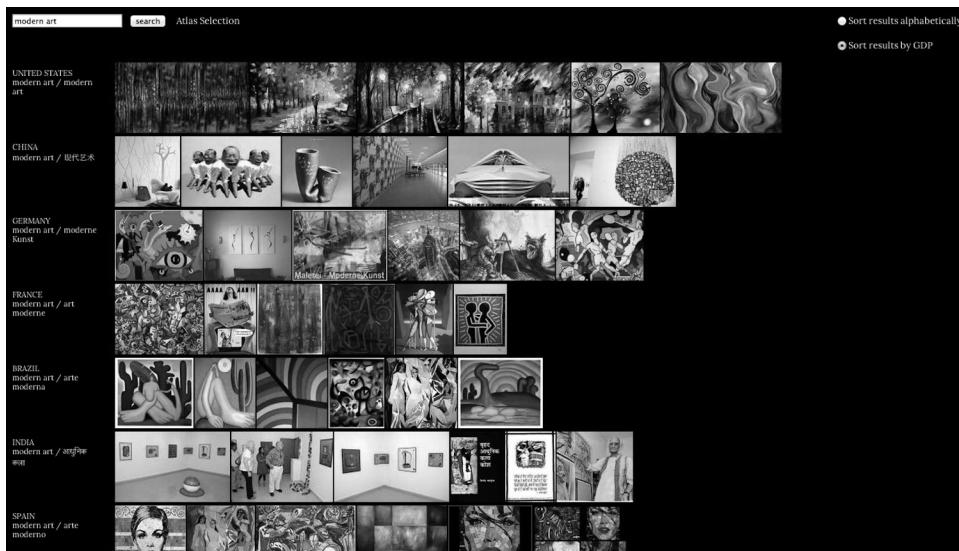


Figure 11. Image Atlas results on searching “modern art” and sorted by GDP. <http://www.imageatlas.org/>. Please see the online edition of *Art Documentation* for a color version of this image.

chart or interface and the interpretive and conditional act of reading, articulated in this quotation from Johanna Drucker, underpins the interplay between information visualizations and narrative ambition. Data is visualized, charted, and parsed in order to frame an assertion and to provide an interface into an avowal, yet in many ways the data itself remains static. This stagnancy of the data impedes the narrative movement necessary for a visualized argument. Contemporary visualizations, however, are able to make use of the interactive nature of the web to allow greater user manipulation of visualizations in a manner that supports the widening understanding needed to fuel interpretive insight.

Image Atlas is an example of visualizing information in that it is both content-dependent and also equivocal towards its narrative model. Image Atlas is a visualization tool created by Aaron Swartz, a programmer and open data advocate, and Taryn Simon, an artist, as part of Rhizome’s Seven on Seven program that matches technologists and artists to develop a specific project within a single day. The website (Figure 11) allows the user to enter a search term, and it then pulls the top Google Image results from different country domains, which can be sorted by Gross Domestic Product (GDP) or alphabetically.³² Here content itself drives the visualization, suggesting how aggregation can expose unexpected epistemologies and unique visual syntax, in this case by country. It manages to recontextualize visual materials through the lens of nationality and economic power and reveal comic or astonishing comparisons through juxtaposition. It also evokes the mystery of the algorithm that drives the tool and reminds the user that any argument, verbal or visual, remains idiosyncratic.

Like the ImagePlot visualizations produced by Lev Manovich, Image Atlas relies on

32. Aaron Swartz and Taryn Simon, “Image Atlas,” <http://www.imageatlas.org/>.

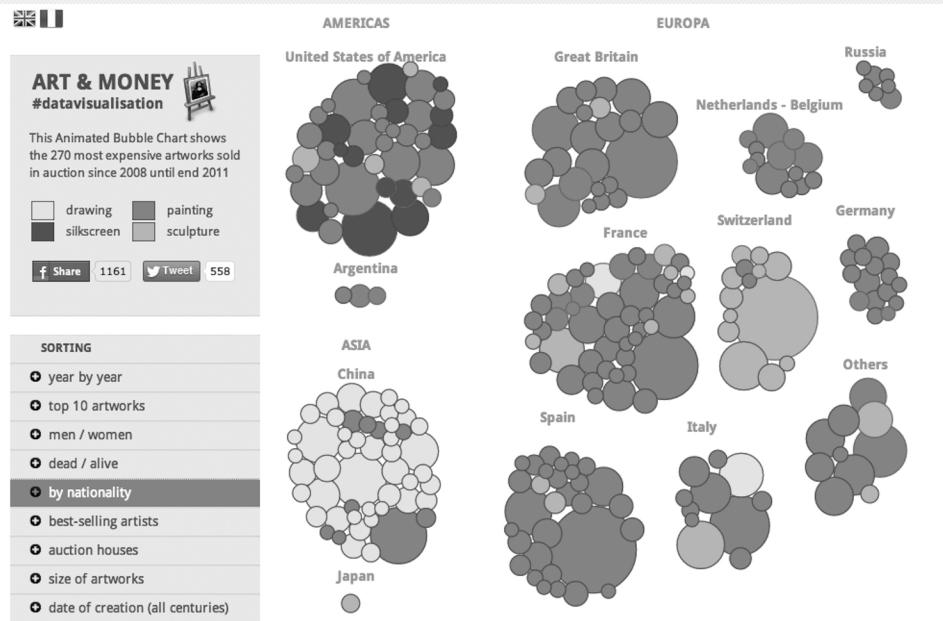
Art & Money
 By: Jean Abbiateci


Figure 12. The Art & Money visualization interface showing the most expensive artworks sold between 2008 and 2011 and sorted by artist nationality. <http://visualizing.org/full-screen/41161>. Please see the online edition of *Art Documentation* for a color version of this image.

the collocation of large sets of individual objects. Image Atlas, though, also slyly demonstrates the potential for the flat neutrality of visualization. Though clearly a mediated, produced interface (with a tacit political argument and clear artistic intent), its content is returned via a search engine whose PageRank algorithm remains unknown. Are the similarities and contrast of Image Atlas's results quirks of culture? Are they illustrative of national taste in aesthetics? Are they simple algorithmic sorting of search popularity? Information visualization, of course, presumes that these are all equally valid propositions, and that all are potential avenues of analysis and meaning. In this case, the visualization inverts the logic of traditional research and analysis, instead suggesting that cultural identity is reflected and refracted by the similarity and contrast of images related to its search habits. These narratives emerge from both the immediacy of the images and the obscurity of the algorithm.

The Art & Money website by Jean Abbiateci is an example of a visualization that provides multiple avenues of entry to analyze a data set of artworks sold between 2001 and 2008 (Figure 12). By providing multiple methods of sorting and color-coding, along with interactivity, animation, and rollover thumbnail images, the visualization accomplishes a number of goals. Much like traditional static visualizations, it reveals trends and patterns, outliers and commonalities, assigning importance by size and using colors and arrangement to categorize. But by being dynamic and interactive, it allows the same underlying set of data (artwork sales) to be parsed in different, interconnected visualizations, all appearing within the same interface. Selecting an alter-

nate sorting method allows the user to reform the same set of visualized objects (in this case, color-coded bubbles) into new arrangements. Toggling between different methods of arrangement begins to allow one to iteratively explore the data in a way that is layered, progressive, and akin to discursive exposition.³³

Geographic visualizations provide their own narrative impetus to cultural and artistic change as they track the movement of trends across time and place. Projects at Stanford and Harvard's metaLAB serve as storytelling examples using cartographic narratives of growth and transmission. Stanford's "Journalism's Voyage West" tracks the spread of newspapers across the United States between 1690 and 2011 by visualizing data from the Library of Congress Chronicling America Collection. Harvard uses this technique to document the spread of printing across Europe in the fifteenth century by visualizing the location of printed works by year and in a separate project to map the influence of Adam Smith's *Wealth of Nations* by visualizing printed editions by country.³⁴ In all these cases, geographic movement through time is charted across maps and national borders to signify cultural, social, or economic influence or adoption. The narrative is one of connection and sway, showing how the impact of objects, collections, or technologies can be understood as a process, a movement, or a flow and not just as a set of discontinuous data points.

COLLECTION ANALYSIS

Though the previous examples have been situated conceptually, their utility to art historians, curators, and content stewards is clear. Visualization can reveal patterns and uncover narratives that have the potential to enhance how collections are managed, accessed, and used. The term collection analysis is used here quite broadly, referring both to the contents of a curated collection as well as a large, less-mediated aggregation of items or objects sharing a similar trait, such as creator, owner, exhibition, or other technical or descriptive metadata. Information visualization, beyond illuminating shared characteristics for further scrutiny or exploration, has also come to serve an administrative or technical function, allowing curators and collection managers to gain intellectual or physical control of a collection. This function is, however, not merely an administrative one, but one that offers its own revelations about provenance, collecting trends, and otherwise unseen characteristics of institutional practice.

The Samuel H. Kress History and Conservation Database Project is an example that offers a bridge between the concepts of collection and pattern analysis. In this case, a custodial use of visualizing collection data facilitated a better understanding of the collection's origins, qualities, and donation patterns. Metadata documenting the 3,600 paintings, sculptures, medals, and decorative art collected by the Samuel H. Kress Foundation between 1927 and 1958 was exported from the database system, refined, and then imported into the online visualization tool Viewshare. The multiple

³³ Jean Abbateci, Art & Money, 2012. <http://visualizing.org/full-screen/41r61>.

³⁴ Rural West Initiative, Stanford University. Data Visualization at http://www.stanford.edu/group/ruralwest/cgi-bin/drupal/visualizations/us_newspapers; Van Ballegg and Kasia Cieplak-Mayr, "The Explosion of 15th Century Printing: A Data Visualization," *The Atlantic*, December 5, 2012, <https://web.archive.org/web/20130313143039/http://www.theatlantic.com/video/archive/2012/12/the-explosion-of-15th-century-printing-a-data-visualization/265902/>.

visualization interfaces revealed trends that, for a researcher, could take days or weeks of documenting and analyzing hundreds of individual records. But in this example, the data can be visualized immediately in a way that elicits information immediately confirmable through individual analysis. The method of inquiry is, in some ways, reversed. Instead of the characteristics of exemplars informing analysis, the trends within an aggregation can serve as wayfinders for further exploration.³⁵

In Figure 13, manipulating histogram sliders for date range changes the scatterplot visualization. Here the x-axis represents date and the y-axis indicates the purchase price of each individual artwork. Comparing these two images (a process that can be done automatically using the online interface) shows us the dramatically different costs of artworks purchased in two different time periods. As a comparison of the images shows, Kress family purchasing patterns turned towards far more expensive works during the second half of their timespan of collecting. In Figure 14, from the same data set, the pie charts quantify the number of works purchased from a specific dealer. Here, limiting the results by specific date spans shows that one dealer dominated the first eighteen years of collecting. It can also be easily seen that the same dealer played a far smaller role in the final thirteen years of collecting.³⁶

While these trends are made apparent through dynamic visualization, they give away none of the reasoning behind, or origins of, this shift. The scatterplot or pie chart cannot reveal, as happens to be the story in these specific examples, that Kress family acquisition tastes changed due to the death of one of the key family members overseeing purchases. The visualization does not show that near-exclusive reliance on one specific Italian art dealer waned at a certain period due to the complications of World War II. But visualizing the provenance information of this collection leads to a suggestion, a curiosity, that may not emerge from study of the collection data itself, and these qualities of intrigue have value not just to the researcher but to the collection stewards supporting that research and maintaining this and other related data sets. The other essential feature of these visualizations is their dynamic and interactive nature. This is not a static visualization that codifies a statistic or an argument, but one that responds to faceting, sliders, text search, and other user inputs. They are visualizations that can be manipulated according to the interest and analysis of the user and the affordances of the underlying data itself.

While information visualizations are often thought of as being generated externally from the data they represent, they can also be an elemental part of online collections themselves. The Cooper-Hewitt Labs, part of the Smithsonian Cooper-Hewitt National Design Museum, has experimented with automatically generated timeline visualizations that can be “turned on” so that they appear on each item’s webpage (Figure 15). As the site notes, “the timeline’s goal is to visualize an individual object’s history relative to the velocity of major events that define the larger collection. . . . To continue to develop a visual language to represent the richness and the

35. The online visualization can be found at <http://viewshare.org/views/GalleryArchives/samuel-h-kress-collection-paintings-with-purchase/>.

36. Lauren Algee, Jefferson Bailey, and Trevor Owens, “Viewshare and the Kress Collection: Creating, Sharing, and Rapidly Prototyping Visual Interfaces to Cultural Heritage Collection Data,” *D-Lib Magazine* 18, no. 11 (November/December 2012), <http://www.dlib.org/dlib/november12/algee/11algee.html>.

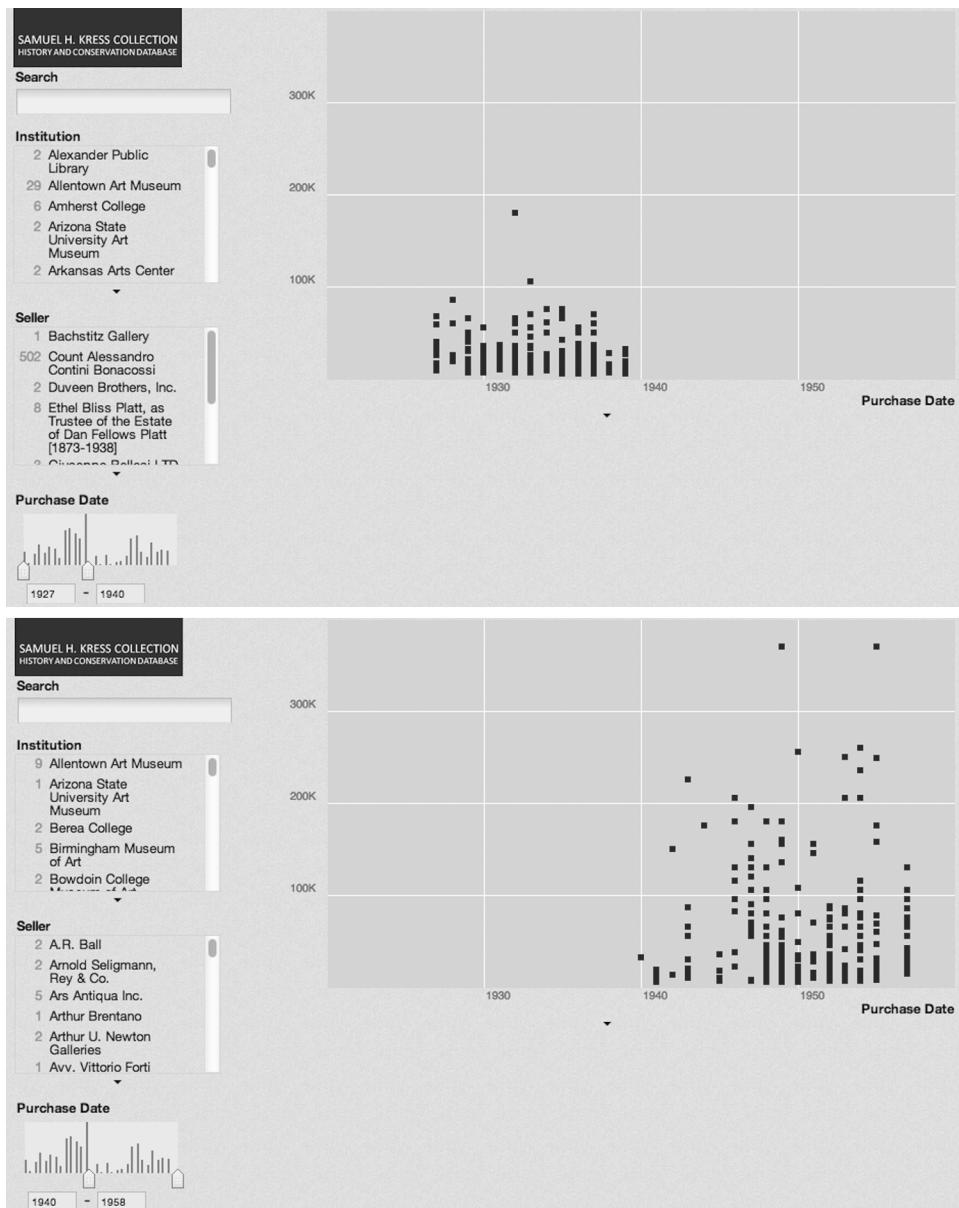


Figure 13. Screenshot from Kress Collection in Viewshare. <http://viewshare.org/views/GalleryArchives/samuel-h-kress-collection-paintings-with-purchase/>. Please see the online edition of *Art Documentation* for a color version of this image.

complexity of our collection. To create views that allows a person to understand the outline of a history and invite further investigation.”³⁷

In this instance, the visualization serves to place the individual item within the

³⁷ Aaron Straup Cope, “A Timeline of Event Horizons,” Cooper-Hewitt Labs blog, September 25, 2013, <http://labs.cooperhewitt.org/2013/a-timeline-of-event-horizons/>.

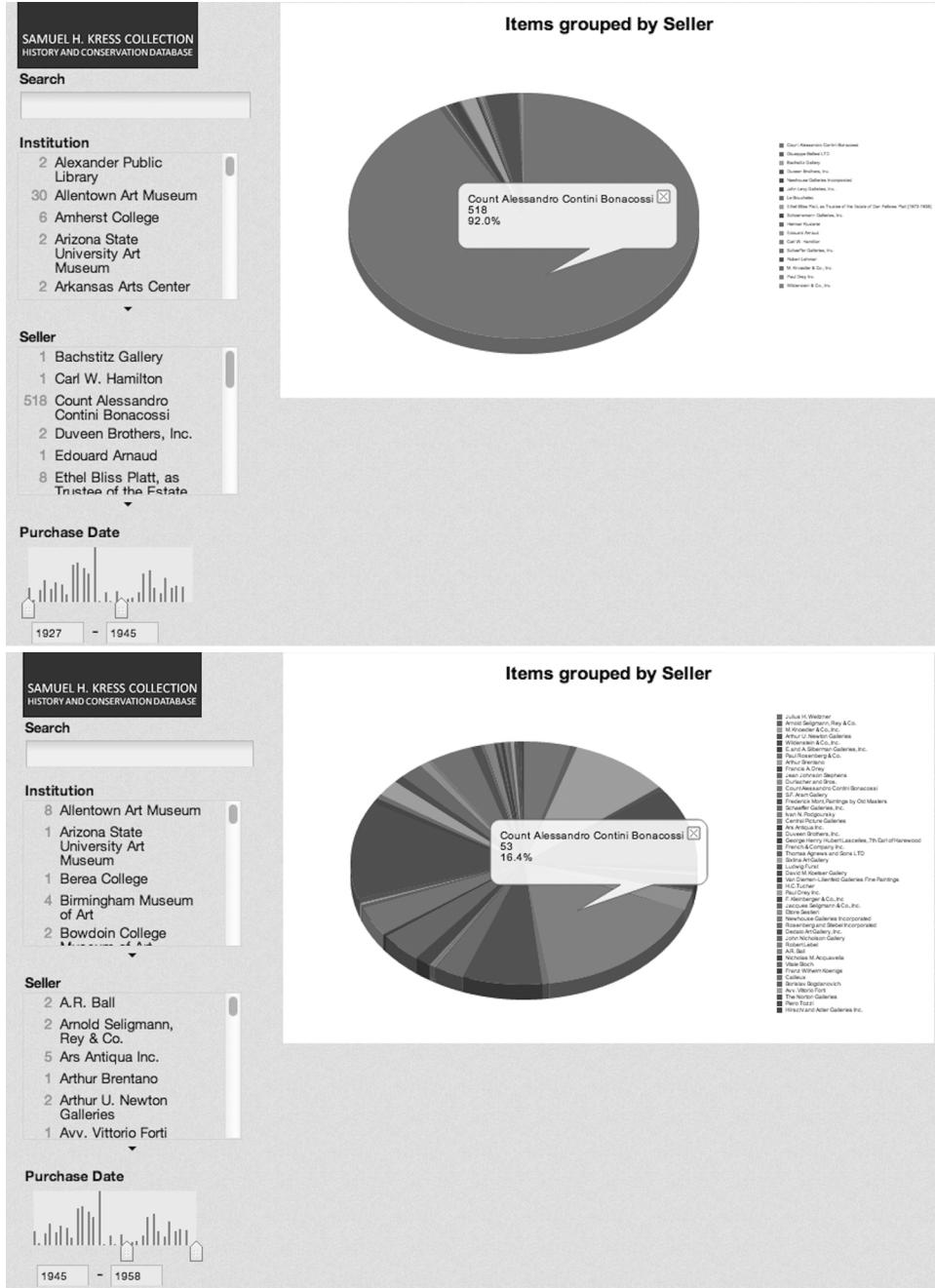


Figure 14. Screenshot from Kress Collection in Viewshare. <http://viewshare.org/views/GalleryArchives/samuel-h-kress-collection-paintings-with-purchase/>. Please see the online edition of *Art Documentation* for a color version of this image.

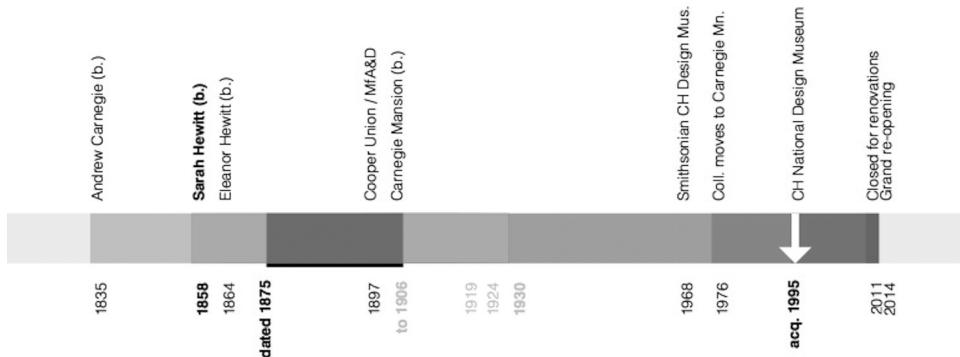


Figure 15. Item-specific timeline visualization from the Smithsonian Cooper-Hewitt Museum collection website. <http://labs.cooperhewitt.org/2013/a-timeline-of-event-horizons/>. A version of this interactive visualization can be turned on to appear on every item page. Please see the online edition of *Art Documentation* for a color version of this image.

context of both the larger collection and the evolution of the institution itself. Information such as the date of creation or acquisition, while valuable, lacks the ability to manifest the larger historical and circumstantial provenance. The timeline visualization notes key moments in the museum's formation and growth, identifies the date of the item's acquisition by the museum (the white arrow), and highlights in specific color (in this case the yellow span) the assumed creation date of the object (if the exact date is known, it is represented by an arrow of a different color). In this case, the visualization serves to provide relevance, to situate an individual item within a larger institutional context, and to allow a user to better understand collecting practices and the historical contingency of creation and acquisition.

Collection analysis is, of course, the province of researchers as well as curators, collection managers, and museum technologists. The advent of open data in the museum world, primarily collection data extracted from systems like The Museum System (TMS), has allowed individuals outside the institution to create their own visualizations. An example is Figure 16, 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). 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.³⁸

Visualizing information from the perspective of collection analysis allows curators and collection managers to gain novel insights into the nature of their holdings—

³⁸ Florian Kräutli, "The Tate Collection on GitHub," YYYY-MM-DD blog, November 2, 2013, <http://research.kraeutli.com/index.php/2013/11/the-tate-collection-on-github/>. Florian Kräutli's blog post offers an insightful analysis of the challenges and successes of working with museum open data collections.

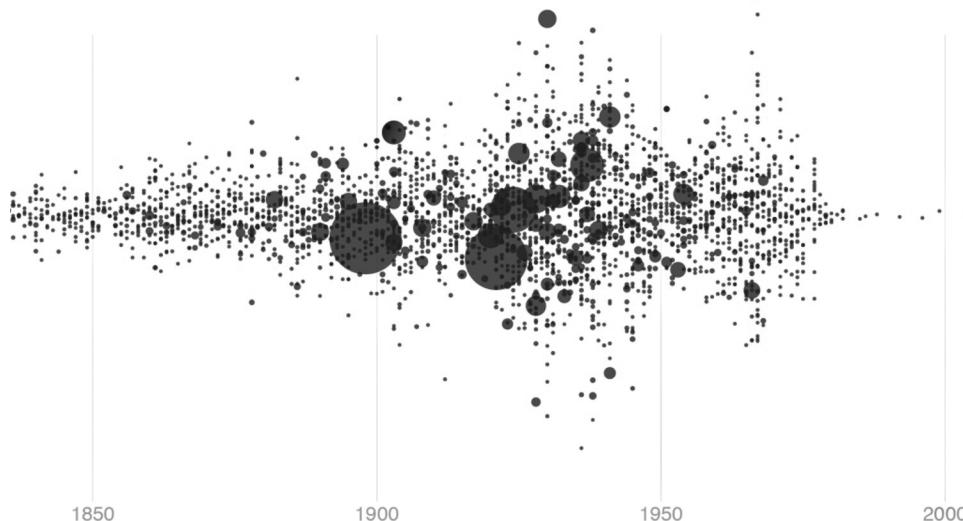


Figure 16. Distribution of artworks in the Tate collection by birthdate of artists. Visualization by Florian Kräutli. <http://research.kraeutli.com/index.php/2013/11/the-tate-collection-on-github/>. Please see the online edition of *Art Documentation* for a color version of this image.

insights that facilitate and enhance supporting research uses and imagining new methods of exhibition. In addition, visualizations can enhance contextual detail where it is most valuable and hardest to summarize—on the item’s web page. Embedding dynamic visualizations on object web pages allows curators and stewards to broaden the focus of an object’s history to account for institutional changes and to elucidate more meaningful historical connections.

CURRENT CONTEXTS

One of the exciting aspects of data visualization for art librarians, museum archivists, and collection managers is that it breathes new life into the descriptive functions these roles often entail or support. While some examples given here, such as the ImagePlot or What Makes Paris Look Like Paris?, rely on machine-extracted or generated technical information of digital objects, much of the underlying data that enables visualization (and the meaning that users derive from those data-driven visualizations) has its origin in metadata created by professionals: librarians, archivists, curators, and others. Catalog records, TMS notes fields, curatorial annotations, registrar records, contextual descriptions and abstracts—the data that enables visualization is the data that the art library and museum community is especially talented at and uniquely trained for creating. With information visualization, that ability to describe, to assign meaning and detail, serves not just to help users find items in a catalog, or better understand the contents of a book or the origins of a sculpture, but also allows for the creation of visualization and data-dependent interfaces that can enable new modes of research and inquiry across all of a collection, encompassing entire genres and *oeuvres*. Information visualization is far from being an ancillary discipline divergent from the services and skills of collection managers and catalogers, but

is instead a domain that can reenergize, amplify, and expand the talents of curatorial and informational cultural heritage professionals.

Finally, while this article extolls the value of data visualization for librarians, archivists, and collection managers, many of the examples presented herein were not the work of that community but the work of software developers, digital humanists, or just talented enthusiasts and museum supporters. Their work was largely dependent upon the availability of museum data in open formats and delivered on publicly accessible platforms. The open data movement has recently shown signs of accelerating, and it bears remembering that much of the cutting-edge work of data visualization with art and museum data has happened because of an increased willingness of museums to offer their collection data and images to data-inclined researchers in unrestricted, reusable ways. This article's authors encourage art libraries and museums to adopt, continue, or expand their commitment to open data policies and practices to further support interdisciplinary collaboration and exchange necessary to cultivate new and emerging types of research, patron recontextualization, and reuse of collections and data.

CONCLUSION

Visualizing data serves many roles: clarification, argumentation, identification and abstraction, and consolidation and fragmentation. By examining the historical evolution of data visualization and linking it to contemporary uses in scholarly and computational analysis of art historical and cultural contexts, this article elucidates how visualizing data can support new modes of understanding, serve an administrative function, and play a unique role in mediating how users explore and understand groups of materials. Visualizing information is not just a communicative act but also an interpretive one, a process, both “generative and iterative, capable of producing new knowledge through the aesthetic provocation.”³⁹ Visualizations can interrogate, explicate, and reveal. In this sense, visualization contains the potential for an almost insurgent empowerment. When Moretti speaks of distant reading and the power of aggregation, he notes how visualization “reveals form as a diagram of forces; or, perhaps, even, as ‘nothing but force.’”⁴⁰ He untangles the intertwined reduction and abstraction that visualization can offer. It can serve as both wayfinder to a specific object, moment, or formal feature, but it can also allow one to discern trends and patterns across very large sets of heterogeneous data. But most evocatively, Moretti places visualization in the context of a broader radicalism, one that questions preconceptions, that usurps entrenched narratives and staid formulations. Information visualization is not just about discovering new meanings but can also be about denouncing the assumed verities, exposing the forces behind the forms, and becoming a more expansive vocabulary, a lexicon that revels, connects, and unmasks.

ACKNOWLEDGMENTS

The authors would like to thank Jennifer Tobias, Christian Huemer, Nigel Holmes, Judy Dyki, The Museum of Modern Art Archives, and The Thomas Fisher Rare Book Library, University of Toronto.

³⁹. Johanna Drucker, “Graphesis: Visual Knowledge Production and Representation,” *Poetess Archive Journal* 2, no. 1 (December 10, 2010): 33.

⁴⁰. Moretti, *Graphs, Maps, Trees*, 64.