The Use of Data Visualizations for Analyzing Social Trends

Dr. Richard Schaefer

Professor of Mass Communication

Department of Communication and Journalism

University of New Mexico

Correspondence:

505-917-9909

schaefer@unm.edu

The Use of Data Visualizations for Analyzing Social Trends

Abstract

This essay posits that now is a particularly propitious time for the development and use of data visualization as a means for communicating abstract baseline information about society's complex institutions, organizations, and social structures. It reviews recent developments in off-the-shelf visualization software and describes supporting literature and tutorials. Finally, it presents some of the ethical dilemmas and constraints confronting visualization producers.

For generations, communication theorists, like Fisher (1985), have pointed to the importance of storytelling as a technique humans use to inform themselves and transmit culture. The long trajectory of humans' social evolution arcs from telling tales around the fire and primitive cave drawings through the development of writing, printing presses, photography, and film—all, at least partly, in the service of communicating complex understandings about ourselves, our historical moments, and the transmission of our culture.

Understandings gleaned from these communications have produced organizational and technological feats of ever-greater complexity and scope. In a blink of biological time, our species has migrated out of Africa to inhabit every continent and developed tribal structures that grew into ethnically homogenous civilizations unified by complex visual symbols and communication. In turn, these more homogenous civilizations have been subsumed within increasingly heterogeneous secular formations characterized by overlapping civic and economic

hierarchies that, in turn, depend upon even more complex communications. Thus, many individuals today identify themselves as members of families, communities, political organizations, nation-states, and, finally, constituents of complex globalized civic and economic structures, such as the United Nations and WTO—structures that present their own unique sets of informational and communication challenges.

As one moves to ever greater complexity and abstraction, from the individual through the familial to the tribal, and then toward ever more complex and elaborated social understandings of globalized entities, simple anecdotal storytelling is no longer sufficient for communicating the overarching patterns and "stories" available in aggregated social and economic data—patterns that are essential for understanding the efficacy and trajectories of so much of our civic, national, and global activities.

This essay posits that now is a particularly propitious time for the development and use of data visualization as a means for communicating abstract baseline information about our complex institutions, organizations, and social structures. It also offers some advice on doing so efficiently and ethically.

Visualizing Abstract Concepts

Visualizations, or what some call infographics, provide a communicative alternative to the more prevalent and traditional anecdote-based narratives and storytelling. For thousands of years mass communicators have relied on the timeworn technique "to put a face on it" in order to report on a problem or achievement. This method of illustrating the human dimensions of an issue by depicting the plight of a single individual still dominates informative communications today, as theorists have posited, and even recently tested, that such individual stories are the best vehicles for piquing audience members' emotional interests, identifications, and social

action (Amit & Greene, 2012; Baird, Scheffer & Wilson, 2011; Oliver, Dillard, Bae & Tamul, 2012; and Slovic, 2007).

But, for centuries, maps and informational graphics of various types have existed side-by-side with anecdote-based word, photo, film and video stories. Even before the widespread diffusion of computers, databases, and digital communication technologies, communication theorists described the best practices for visualizing abstract information and concepts—a practice that had preoccupied map makers for centuries.

Beniger and Robyn (1978) described the history of the budding field they called "quantitative graphics" at a time when it was just beginning to draw on existing databases and mainframe computer programs written specifically for what the theorists of the 1960s and 1970s called "scientific visualizations." Extending the foundational work of Bertin (1967, 1977, 1983) in France and Tufte (1983, 1990) in the United States, a growing body of theoretical best practices had begun to evolve for map makers and digital data analysts in the 1980s and 1990s, when desktop computers and spreadsheet and data analysis programs were just beginning to be mass marketed to visual designers. Thus, visualization and the art of digital visual design morphed from a skill practiced primarily by computer programmers to a communicative art practiced by journalists, researchers, and analysts in particular industries who wanted to communicate abstract concepts from their own fields of study to wider audiences.

The art of visualization is highly intuitive, but it is also conventional and tied to human perceptual capabilities. In the 1980s Cleveland and McGill (1985) provided a taxonomy of the types of graphical forms that best communicate certain dimensions of variable relationships most easily and accurately. Using evaluative testing, they specified the infographic forms that best communicate various types of data relationships to human's visual perceptual system. In the process they guided and encouraged early infographic makers to use specific types of

charts for particular relationships in order to communicate the relationship without introducing the types of visual illusions popularized by op-artists—illusions that would be considered distortions of the data. In recent years, Groeger (2016) explored that subject more creatively and accessibly for a journalistic audience. That tradition of making graphic depictions more accessible to information communicators has been taken to new heights by Krum (2014) and Cairo (2012, 2016).

More recently, new graphic communicators have been able to increasingly rely on publicly available data sets and inexpensive commercial and open-source spreadsheet and visualization software to produce graphics that communicate baseline understandings— understandings that provide a portrait of key social and economic indicators, including the trajectory and progress of those indicators across years and decades. This broad ability to illustrate the trajectory of everyday concepts enables the new visual storytellers to generate graphics that provide surveillance of complex social phenomena and relationships, often based on publicly available data. Thus, it can help provide a long sought fact-based platform for public opinion formation that democratic theorists, such as Lippmann (1922), envisioned a century earlier. This ability of communicators to provide surveillance of the daily routines and progress of key institutions is a far cry from the type of more sensational journalistic and informational reporting that focuses on individuals' exceptional feats, personal scandals, or improbable activities.

Surveillance and Infographics

At a time when all sorts of digital technologies have made it possible for institutions and governing bodies to collect and share unprecedented amounts of baseline data on nearly every measurable aspect of daily life, the ability to visually represent the trends and patterns evident in

that data has never been more important. The value of this type of routine surveillance of available data is far greater than most citizens realize. Its cost largely depends on the cost of analyzing and envisioning the data and the availability and utility of the aggregated data to the public and particular information elites.

In recent years, the Bloomberg terminal provides what is perhaps the most expensive example of the value of such visualized information capabilities. A typical Bloomberg terminal is a Windows-based information and visualization system that provides access to real-time financial data and near-instantaneously produced graphic charts that can be displayed on hardware consisting of two to six monitors (The terminal, n.d.). Besides offering trading platforms, data, and news for investors and investing professionals in the "fintech" industry, Bloomberg L.P. has penetrated the market for various types of business news users and corporate executives. Access to a single Bloomberg terminal runs $24,000 a year, and in 2013 the company's leasing rights had grown to 315,000 subscribers and 57 percent of the quickly-developing fintech information market (Nath, 2015).

Legacy news media have also enhanced their infographic capabilities in recent years. Riordan (2016) documented the birth and end of the New York Times R&D Lab, which was largely responsible for the development of visualizations and other media innovations within the organizational structure of "The Gray Lady." Though the R&D Lab closed in 2016, infographics and visual interactive features have become an integral part of the both the print and digital editions of the news organization.

Similarly, the other great U.S. national newspaper, the Wall Street Journal, recently reorganized its print and online graphics departments into a single entity and began experimenting with its Daily Shot visualization-driven digital storytelling (Roush, 2016; Thompson, 2016). Thus, America's two great national newspaper organizations appear to be

investing in the future of infographics even as they and other major newspaper organizations, including Gannet and the Los Angeles Times, take steps elsewhere to cut their legacy staffs and expenses (Barron, 2016; Ember, 2016).

Still, one does not have to purchase a Bloomberg terminal subscription or even subscribe to the digital versions of the New York Times or Wall Street Journal to depend upon infographics to follow the "surveillance" function of the media—a function that Lasswell (1948) articulated as the media's ability to inform the public about key happenings in the world in a way that makes events more understandable and less confusing. Although the organization of human society and social activity is more complex than ever, information workers' ability to obtain baseline data and present it visually has never been more accessible.

Bringing Numbers and Data to Life Visually

The term "big data" has become something of a cliché within academia and business. But it refers to an aspect of contemporary life that is unassailable, as the digital traces of all sorts of everyday activities can be recorded—that is, recorded digitally and quantitatively—and then stored in a database that researchers can retrieve and explore for patterns, trends, and associations.

The "big" part of big data refers to data sets that are so large that ordinary desktop computers have trouble processing their information, which, of course, is a moving target because computing power expands regularly along the lines described by Moore's Law. In addition to describing a database by its volume, we also refer to the nature, or structure, of the data. Structured data is the more traditional and familiar type of numeric data in which the stored numbers have specific definitions and interpretations—something like a definitional roadmap of the value as a measurement. In short, the Xs and Os that represent numbers or

text also refer to something clearly defined (i.e., money charged with a credit card) that already has a meaning and structure.

In contrast, unstructured databases have no clear key with regard to what they describe or measure, so the analyst must devise an algorithm to make sense of the data before any analysis can take place, and an unstructured data set may have various meanings depending on the algorithm that is used to render it "structured." For example, video recordings of the sky over a playground could be subsequently analyzed for brightness of the sun or the moon, artificial light at night, the flight patterns of planes, or of birds, or of insects, or an analysis of the clouds and their moisture content. Each analysis would require different algorithms to structure and interpret the otherwise unspecific data.

Despite the fact that the "Internet of things" is producing heaps of structured and unstructured data, this analysis deals mainly with the deliberately structured and often shared data produced primarily by institutions. Such highly structured and well-defined data produced by institutions, particularly government institutions, is routinely being made available to the public for examination and analysis as a legitimating social value. Many public institutions are sharing the data in an effort to be transparent with their constituents. These institutions are making the digital information available in commonly usable forms, such as databases in the format of the ubiquitous Microsoft Excel spreadsheet program, or in the CSV (comma-separated values) format that programs such as Excel, SAS, SPSS and nearly every other type of commercial spreadsheet and data analysis tool can read and re-write for manipulation and analysis. The availability of this type of highly structured and conventionally formatted data is driving the contemporary trend toward visualization.

Desktop computing spreadsheets, such as Microsoft Excel and Apple's Numbers, can not only create and perform routine mathematical manipulations on the variable values, but the

software can also produce various types of "charts" or graphic representations based on one or two of the variable dimensions. The same is true for some of the most commonly distributed statistical analysis programs, such as SAS, IBM's SPSS, and Statistical Solutions' BMDP.

All these programs produce highly conventional and simple images, such as pie charts, bar charts, line graphs, histograms, and scatter plots, that make up the common lexicon of quantitative visual representation. Some, such as the most current versions of Excel and Google Charts, can interpret GPS coordinates and translate location names into map coordinates. This enables the programs to produce either static or interactive maps with representations of crimes, documented weather events, or other types of variables associated with geographic location markers. And with each year, the data analysis and visualization capabilities of these relatively common spreadsheet and data analysis programs increase, as media producers and business and research-oriented users demand more exploration and visual capabilities from the latest software. This demand for more capability encourages the software programmers to include new analysis and visual representational algorithms in their new releases.

In recent years, another set of software programs has been created specifically to produce more elegant visualizations. Many are either free to the public or provide public versions with slightly reduced capabilities as teasers for the more complete commercial program. Tableau's commercial version is particularly popular with journalists and other media producers, and the free version—Tableau Public—produces many of the same visually stunning graphics as the commercial version. Google's Fusion makes maps and other types of visualizations to complement Google Charts, the company's more traditional data graphing software. And each year visual designers develop new programs, often under grants requiring the programmers to place the software in the public domain. Therefore, a simple search for

visualization tools on the Web yields a wealth of new programs, such as Vennage, Canva Infographic Maker, and Visualize, that go beyond simple charting to produce elegant, professional-looking, and pictorially complex infographics.

Other programs are particularly adept at making animated or motion graphics. Flare is probably the most popular in this category for visualizations that run with Adobe Flash, and Dipity specializes in helping users create interactive timelines. Video producers who are familiar with the keying and the motion effects available in the off-the-shelf Adobe Premiere, Avid Media Composer, and Apple Final Cut Pro commercial video editing software can add motion and progressive layering to imported charts. All three of the abovementioned video editing programs can import layered stills from static visualization programs and add or subtract layers, provided the stills are imported as TIFF files rather than JPEGs. These capabilities enable video producers to build elaborate animated visualizations for broadcast or streaming.

Almost everyone who has ever tried to learn a software program in a hurry knows that even the simplest programs can be frustrating to the uninitiated, and visualization software is no exception. Most infographic programs have their own tutorials, and the instructional software training program Lynda.com has tutorials for nearly all the programs listed above. Lynda.com also offers a series of best practices tutorials on data visualization design, with the latest, produced by Bill Shander, titled "Data Visualization Storytelling Essentials" (Shander, 2016). In recent years the Poynter Institute (Cairo, 2017) has also begun offering instruction in basic infographic production that has been targeted specifically toward journalists with little or no prior visualization training.

Among the many how-to manuals, Krum's (2013) primer on effective visualization and design is an easily accessible "quick read" and, like the work of Tufte, is visually stimulating. Visualization designers Segel and Heer (2010) have also placed a manuscript in the public

domain that describes the most common types of interactive infographics. Like most primers on visualization, it suggests that a clear storyline is needed to select and translate compelling data into viable narratives.

Indeed, when the author of the present essay teaches visualization to journalism students, researchers, and other information workers, the author requires the aspiring visualization designers to start with a hypothesis that can be either proven or disproven by careful examination of a data set. This requirement was inspired by a chapter from a UNESCO primer for investigative journalists that was written by Hunter, Sengers, and Thordsen (2011). Using a hypothesis ensures that the time spent finding the data and creating the visualization, which is often tedious and relatively expensive, will result in the production of an effective infographic story. Developing a hypothesis also forces students to look at a single compelling question, which prevents "visual data snooping" and thereby reduces the odds that the students will commit what statisticians call a "type one error" by examining so many visualizations that designers increase their chances of stumbling across a completely anomalous, and therefore misleading, data set.

Ethical Issues

The human visual system is a powerful organizer and judge of information, but as anyone who has ever been tricked by optical illusions knows, visual imagery can be deceiving. Groeger's (2016) short article for Propublica illustrates some of the most common optical illusions that have plagued infographics designers and consumers. For a more thorough analysis of various aspects of visualization, Huang's (2014) graduate-level handbook contains chapters on some of the more theoretical, evaluative, and ethical issues arising from the uses of visualizations, including a section on the cognitive utility of infographics. And Schaefer and

Jácquez (2015) have produced a pilot study that found animated video infographics to be better transmitters of abstract information than scene- and person-oriented video.

Although a statistical background is not typically required for visualization designers, having some understanding of statistical principles can be helpful. For example, it is imperative that visual designers understand the difference between nominal, ordinal, and ratio variables, as most of the visualization programs distinguish among the types of imagery that can be produced with specific types of data. Unfortunately, many of the visualization programs use less conventional or idiosyncratic terminology for the different types of variables and relationships. Along these lines, Tableau uses the fairly logical terms "discreet" and "continuous" to describe nominal and ratio variables. But it also uses the term "measures" for ratio variables and "dimension" for nominal variables. In this sense, visualization designers need to be ready to decode each program's less conventional terminology.

And for those who live outside the United States, most of the visualization software has inadequate linguistic translations or no translated materials at all, even for such common languages as Spanish and Chinese. Hopefully, as programs become more popular and globalization progresses, better translations will be developed.

Finally, most of the interfaces and software described above were designed primarily to handle associations of two variables. But the author has found it particularly helpful to create two-dimensional visualizations that place more than one variable and their associated scales on either the X or Y axis of a chart, thereby illustrating the correlations between more than two variables over time. Most of the programs designed for visualization cannot do this within a single graph, so other techniques, such as video keying or image layering, are needed to create these more complex graphics. Again, users can only hope that this limitation will be overcome through future software releases.

It has already been noted that visualized data sets have the ability to communicate complex abstract concepts while arousing less emotion than the more traditional "put a face on it" anecdotal narratives. In a time when the efficacy of public institutions is being questioned and baseline understandings of complex social trends are desperately needed to determine and communicate the overarching effects of complex social phenomena, visualization holds tremendous potential to play an important role in this vital communication task.

References

Amit, E., & Greene, J. D. (2012). You see, the ends don't justify the means: Visual imagery and moral judgment. *Psychological Science: OnlineFirst, 20*(10), 1–8. Advance online publication. doi: 10.1177/0956797611434965

Baird, A. D., Scheffer, I. E., & Wilson, S. J. (2011). Mirror neuron system involvement in empathy: A critical look at the evidence. *Social Neuroscience, 6*(4), 327–335. doi: 10.1080/17470919.2010.547085

Barron, J. (2016, October 19). A. G. Sulzberger: Leading change at The New York Times as journalism evolves. *The New York Times*. Retrieved from https://www.nytimes.com/2016/10/20/business/media/ag-sulzberger-leading-change-at-the-new-york-times-as-journalism-evolves.html?action=click&contentCollection=Media&module=RelatedCoverage&region=Marginalia&pgtype=article

Beniger, J. R., & Robyn, D. L. (1978). Quantitative graphics in statistics: A brief history. *American Statistician, 32,*1–11.

Bertin, J. (1967). *Semiology of Graphics: Diagrams, Networks, Maps (Semiologie Graphique: Les Diagrammes, les Reseaux, les Cartes)*. Paris: Gauthier-Villars.

Bertin, J. (1977). *Graphics and Graphic Information Processing*. Berlin: De Gruyter.

Bertin, J. (1983). *Semiology of graphics* (W. Berg, Trans.). Madison, WI: University of Wisconsin Press.

Cairo, A. (2012). *The functional art: An introduction to information graphics and visualization.* Berkeley, CA: New Riders.

Cairo, A. (2016). *The truthful art: Data, charts, and maps for communication.* Berkeley, CA: New Riders.

Cairo, A. (Producer) (2017, Feb. 16). *Data visualization for journalists: Stripping the dread from the graphics.* [Web course from Poynter Institute]. Retrieved from https://www.poynter.org/newsu/courses/data-visualization-journalists-stripping-dread-graphics

Cleveland, W. S., & McGill, R. (1985, Aug. 30). Graphical perception and graphical methods for analyzing scientific data. *Science,* 229:4716, 828-833. Available through JSTOR: https://web.cs.dal.ca/~sbrooks/csci4166-6406/seminars/readings/Cleveland_GraphicalPerception_Science85.pdf

Ember, S. (2016, October 27). More wretched news for newspapers as advertising woes drives anxiety. *The New York Times*. Retrieved from https://www.nytimes.com/2016/10/28/business/media/buyouts-wall-street-journal-bad-news-for-newspapers.html?_r=0

Fisher, W. R. (1985). The narrative paradigm: In the beginning. *Journal of Communication, 35,* 74–89.

Groeger, L. (2016, April 20). How information graphics reveal your brain's blind spots. *Propublica.* Retrieved from https://www.propublica.org/article/how-information-graphics-reveal-your-brains-blind-spots

Huang, W. (2014). *Handbook of human centric visualization.* New York: Springer.

Hunter, M. L., Sengers, L., & Thordsen, P. (2011). Using hypotheses: The core of investigative

method. In M. L. Hunter (Ed.), *Story-based inquiry: A manual for investigative journalists*

(pp. 13–26). Paris: UNESCO. Retrieved from

http://unesdoc.unesco.org/images/0019/001930/193078e.pdf

Krum, R. (2013). *Cool infographics: Effective communication with data visualization and design.*

Indianapolis, IN: Wiley.

Lasswell, H. D. (1948). The structure and function of communication in society. In W.

Schramm (Ed.), *Mass Communications* (2nd Ed. 1960, pp. 117–129). Urbana, IL:

University of Illinois Press.

Lippmann, W. (1922). *Public opinion.* New York: Harcourt, Brace.

Nath, T. (2015, May 28). Financial news comparison: Bloomberg vs. Reuters (BAC, GOOG).

Investopedia. Retrieved from

http://www.investopedia.com/articles/investing/052815/financial-news-comparison-

bloomberg-vs-reuters.asp

Oliver, M. B., Dillard J. P., Bae, K., & Tamul, D. J. (2012). The effect of narrative news format on

empathy for stigmatized groups. *Journalism and Mass Communication Quarterly*, *89*(2),

205–224. doi: 10.1177/1077699012439020

Riordan, D. (2016, May 6). Eulogy for The New York Times R&D Lab. *Medium.* Retrieved from

https://medium.com/@riordan/eulogy-for-the-new-york-times-r-d-lab-

c8d6301450ab#.29reowcag

Roush, C. (2016, Sept. 20). WSJ makes changes to Money & Investing team. *Talking Biz News.*

Retrieved from http://talkingbiznews.com/1/wsj-makes-changes-to-money-investing-

team/

Schaefer, R. J., & Jácquez, N. (2015, August). Evaluating issue-oriented video journalism

    techniques. A competitive paper delivered to the Electronic News Division at the annual

    convention of the Association for Education in Journalism and Mass Communication,

    San Francisco.

Segel, E., & Heer, J (2010). *Narrative visualization: Telling stories with data.* Retrieved from

    http://vis.stanford.edu/files/2010-Narrative-InfoVis.pdf

Shander, B.(Producer) (2016, Aug. 31). *Data visualization storytelling essentials.* [Video file

    series in Lynda.com]. Retrieved from https://www.lynda.com/Excel-tutorials/Data-

    Visualization-Storytelling-Essentials/435230-2.html

Slovic, P. (2007). If I look at the mass I will never act: Psychic numbing and genocide.

    Judgment and Decision Making, 2(2), 79–95. Retrieved from

    http://journal.sjdm.org/7303a/jdm7303a.htm

The terminal (n.d.). Retrieved Feb. 24, 2018, from Bloomberg Professional Services website,

    https://www.bloomberg.com/professional/solution/bloomberg-terminal-learn-

    more/?utm_medium=dotcom&utm_campaign=Dotcom&utm_source=Website&utm_cont

    ent=bcom-article&bbgsum=dg-ws-prof-bcom-a1

Thompson, S. A. (2016, August 25). How we rebuilt the Wall Street Journal's graphics team.

    Source. Retrieved from https://source.opennews.org/en-US/articles/wsj-graphics-team/

Tufte, E. R. (1983). *The visual display of quantitative information.* Cheshire, CT: Graphics

    Press.

Tufte, E. R. (1990). *Envisioning information.* Cheshire, CT: Graphics Press