LV13008

“Visualizing “big data”

Ronald Saltinski, Ph.D.
National University

With remarkable increases in computing power, the graphic visualization of “big data” to generate complex and imaginative digital displays of large quantities of information, numerical and non-numerical, can only be describe as nothing less than “staggering.” Such information visualizations offer not only an artistic quality for displaying information but an analytical ability to see clearly and intuitively opportunities that otherwise would remain hidden and unavailable for further investigation.

Introduction

"Big data" starts small with a digital "ping," the kilobyte (KB). A typical page of text and/or numerical data is about four (4) kilobytes of information. Over 2.5 quintillion bytes of data posted every 24-hours. This data is" big data," This data comes from everywhere: sensors gathering scientific and medical data, posts to social media sites, digital pictures and videos, commercial business transactions, global stock market trading, and cell phone texting, tweets, and GPS signals to name a few.

Over the past fifty years worldwide enterprises whether commercial entities or scientific institutions have moved from an ancient analog world to a fully digitized society. The analog world lasted five thousand years recording information in ways that began with scratches on stones to circular scratches on latex records. The digital world will (arguably) last forever with each "ping" or "bit" of information recorded electronically in a variety of storage media. Aside from Doomsday scenarios it remains that the present digital world will provide past, present, and especially future insights into information never before imagined.

Big data is more than simply a matter of size; it is an a growing issue of "complexity." When the varying elements of a tsunami of information instantly interact in a veracity of different ways the whole takes on a life of its own: "it adapts and evolves in response to changing conditions. It can be prone to sudden and seemingly unpredictable changes, market crashes or climate collapse for examples. One or more trends can reinforce other trends in a positive feedback loop until things swiftly spiral out of control and cross a tipping point, beyond which conditions dramatically and radically change" (West 14).

Mining "Big Data"

Analysts are mining "big data" for solutions that range from scientific secrets to the Universe to commercial secrets making commerce more competitive and profitable. Some of the most encouraging and rewarding "mining" tools are very powerful mathematical systems that combine computer simulations with advanced statistical analysis. A suite of statistical tools, "Maximal Information-Based Nonparametric Exploration," (MINE) have been created by The Broad Institute of MIT and Harvard University. "MINE" can "mine" and "tunnel" through huge masses of data to "tease" out hidden relationships. The Broad Institute states that MINE is "essentially an "experiment" in a new way of doing science," generating new ideas and connections that were ordinarily not readily apparent. At the core of MINE is the capacity to generate visualizations resulting from intense statistical analyses.

One of the most accessible and powerful web sites, *Many Eyes*, is freely supported by the corporate computer giant IBM. *Many Eyes* is interactive and serves to "explore different visual representations of large amounts of data and share it with others to help them collectively make better sense of the information." *Many Eyes* is a dynamic system for information visualization, allowing users can upload their own datasets, and/or

work with existing datasets on the site. Text options include a word tree, phrase net, tag cloud, and word clouds. Other options include bar chart, histogram, bubble chart, scatter plots, pie chart, matrix chart, tree maps, standard maps, and tools for tracking changes over time. This very complete set of tools is quite easy to use. Registration is required to upload and save creations. The web site can be explored using already uploaded data without registration.

Another open access web site providing a set of software tools for data visualizations is the *Prefuse Visualization Toolkit*. *Prefuse* "provides optimized data structures for tables, graphs, and trees, a host of layout and visual encoding techniques, and support for animation, dynamic queries, integrated search, and database connectivity." The web site contains an extensive gallery of visualizations and a demonstration video including a user's manual.

Perhaps one of the most comprehensive and informative discussions of visualizing "big data" is the YouTube presentation by Noah Iliinsky (2012). An impressive framework is laid out for how to determine what question one is actually trying to answer, what data is needed (and what data isn't needed) in order to answer that question, and the steps to take through effective visualization to convey the bottom line. Noah emphasizes selecting crucial visual encodings to relate or map to digital data generating graphical qualities inherent in position, size, shape, and color.

"A Tour Through the Visualization Zoo"

Digital "big data" comes down to quantifying that data no matter what the origins of that data. The bigger the "big data" the more the need "to leverage the understanding of that data via the human visual capability that is acutely tuned to identifying patterns, trends, and especially outliers (extreme limits of statistical analysis). "Well-designed visual representations can replace cognitive calculations with simple perceptual inferences and improve comprehension, memory, and decision-making." Visualization of "big data" makes data more relevant for decision-makers and potentially more accessible and appealing to engage more diverse audiences for novel and unique exploration and analysis.

In the journal article "A Tour Through the Visualization Zoo," an exotic tour is created of the more sophisticated and unusual techniques for visualizing "big data." The journal article and illustrations are non-technical and meant to serve as a "taste" of the intriguing possibilities for visual encoding and the potential unexplored and undiscovered "species" of visualizations. The authors conclude with "All visualizations share a common "DNA" - set of mappings between data properties and visual attributes such as position, size, shape, and color – and customized species of visualizations might always be constructed by varying these encodings."

Conclusion

Today, in the "think tank" centers of the world associated with governments, corporations, known and unknown entities, powerful computers are scanning, collecting, and analyzing global digital traffic. There are no raucous printers generating reams of paper with meaningless text and numerical data associated with the computer "hunt." There are only arrays of large flat monitor screens displaying realms of visualized information ranging from bizarre to eerie. The book published in 1999, *Information Visualization: Using Vision to Think* foresaw that in those images lie as yet undiscovered stories and realms of unimagined meaning.

Work Cited

Card, Stuart K., Mackinlay, Jock D., and Shneiderman, Ben. 1999. *Information Visualization: Using Vision to Think*. San Francisco, CA: Morgan Kaufmann Publishers.

Many Eyes
http://www-958.ibm.com/software/data/cognos/manyeyes/

MINE: Maximal Information-Based Nonparametric Exploration
http://www.exploredata.net/

Iliinsky, Noah. "Designing Data Visualizations"
http://www.youtube.com/watch?v=R-oiKt7bUU8

Heer, Jeffrey, Bostock, Michael, and Ogievetsky, Vadim. "A Tour Through the Visualization Zoo." *Communications of the ACM* (Association of Computing Machinery). 53.6 (June 2010): Print.
http://hci.stanford.edu/jheer/files/zoo/

West, Geoffrey. "Wisdom in Numbers." *Scientific American*. 305.5 (May, 2013): Print.