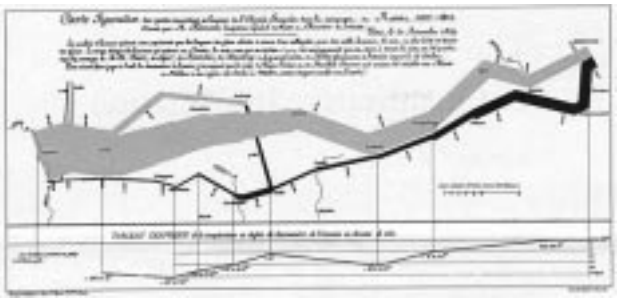
Beautiful Data: Digital Tools that Make Data Look Sexy

By John Sutton Lutz

The vast amount of data: numerical, textual, cartographic, and visual now available is a fabulous resource for scholars but it brings some new problems. How can we present our analyses of hundreds or thousands of pages of text, or millions of transactions, or networks of people in easy to understand, engaging, even beautiful, formats?

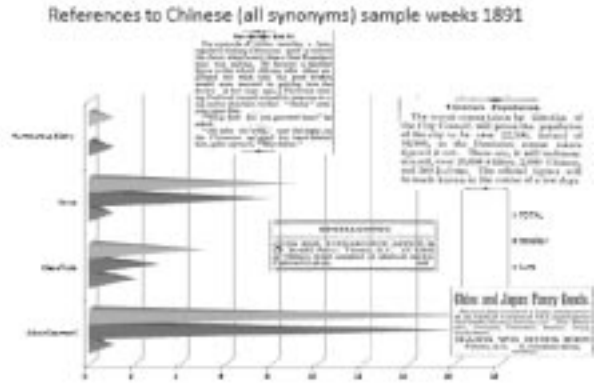
Like so many of the challenges of the digital revolution this is not an entirely new question. Quantitative historians and demographers have long attempted this with simple statistical measurements like chi square, and indices of dissimilarity, which are undeniably elegant, but for most readers, not engaging or intuitive. In the late 18th century William Playfare invented the pie-chart, bar chart and time-series graphs which can convey complex information. One of the earliest, and still considered one of the most elegant examples of visualization tools, is Charles Minard's 1869 pictorial representation of Napoleon's march on and retreat from Moscow in 1812, which conveys several layers of complexity in an easy to grasp representation, with the size of the line proportional to the surviving troops and related to the dropping temperature and distance.



Minard's Map of Napolean's March on Moscow with the larger line indicating his advance and the narrower one in the tableau indicating his retreat.

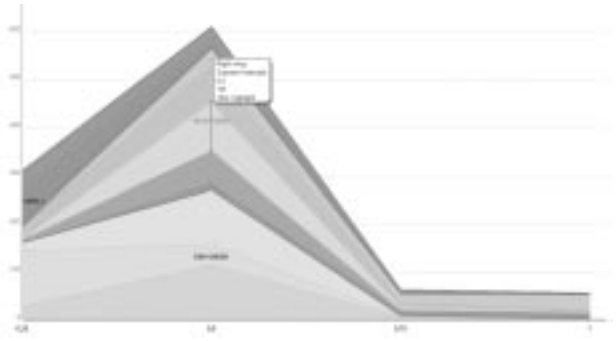
What is revolutionary about our current environment is the vast store of knowledge now searchable and analyzable digitally, and what is needed, and becoming available, are new tools of exploration, analyses and presentation. Fluency in these visualization tools needs to expand in synchronization with research fluency. The creation of new modes and methods of visualizing data is a scholarly activity in its own right. The challenge, as Maureen Stone says is "to do so in a way that balances complexity with conciseness, and accuracy with essence, that speaks authoritatively, yet inspires exploration and personal insight" (2009, 23)

Most of us have some fairly powerful tools on our computer this moment in our Microsoft Excel spreadsheets or its Open Office equivalent "Calc". With or without the addition of visual clues the push of a few buttons can create a range of interesting presentations.



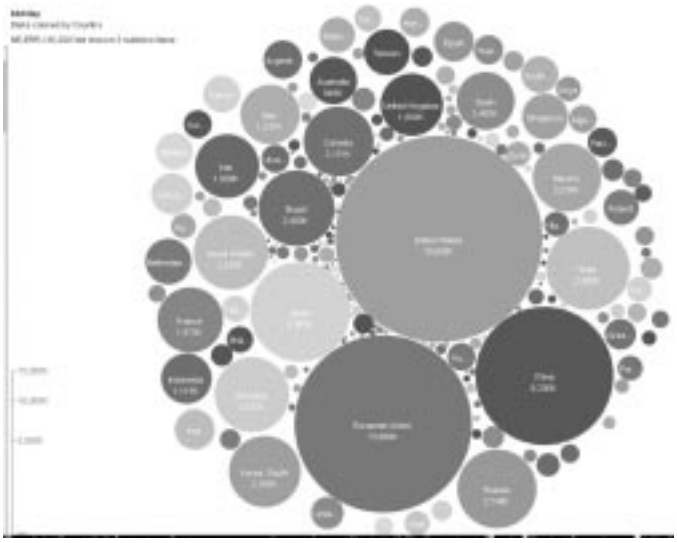
Frequency of Mention of Chinese in British Colonist Newspaper Generated By MS Excel with Sample Articles imported as Images

For more complex data there are now some web based tools that allow users to create much more complex visualizations. One such website is Many Eyes http://www-958.ibm.com/software/data/cognos/manyeyes/ created by IBM. On this website you can use upload your own data to create Stacked Charts, Bubble Charts, Scattergrams, to name only a few options. These are better viewed in colour and on the web they become interactive so mousing over a slice reveals the data used to create it.



Stacked Chart, Finnish Parties Support for NATO

Interactive stack charts are demonstrated in a short video on the Sense.Us website: http://vis.berkeley.edu/papers/sense.us/video/



Total World Wide Consumption of Oil in Barrels per Day by Country, March 1, 2011

Bubble charts can legibly show data values that differ by a ratio of 100,000, and can display hundreds of individual values at once. You can think of them as performing a visual square-root transform on the data set.

Perhaps the most familiar of the tools for analyzing text is the Word cloud tool available at Many Eyes but most famously at Wordle http://www.wordle.net/create. Even simpler than uploading a data set, all the user needs to do is paste in text into the box and click and the result is visible on the cover of this issue of the *Bulletin*, showing the frequency of word use by size of word. The user can adjust colours, contrast, orientation and create a textual piece of art that summarizes a key aspect of a text, beautifully. Tag clouds do the same thing and are common on trendy websites.

At "Many Eyes" you can paste in a text and generate a word tree which is a visual tool for unstructured text, such as a book, article, speech or poem. It lets you pick a word or phrase and shows you all the different contexts in which the word or phrase appears. The contexts are arranged in a tree-like branching structure to reveal recurrent themes and phrases. If you are interested in how different words are linked, the "Phrase Net" shows the strength of word linkages.

For scholars interested in how different people are linked, or want to show how networks function, including social networks, the "Network Diagram" tool will create a visual network from a table with columns of data: ie Person in one column, and Person they know in another.

For inspiration and for the more adventurous check out Flare http://flare.prefuse.org/ and Prefuse which are interactive visualization tools for the web using Adobe Flash. These require the help of a friendly programmer but the code is free and available for the willing.

Flare Graph showing Dependent Relationships

At Prefuse http://www. prefuse.org/ you can explore document content relative to the central idea.

Understanding the flow of people or objects can be made easier through a tool like Flow Map. (Below)

Flow Map Chart Showing Migration to California

Like many kinds of digital scholarship, the algorithms that transform numbers into graphs and words in clouds, operate under the hood. As a profession, to take full advantage of the power of the computer we need to develop peer review processes that certify certain data transformations or at least present their strengths and weaknesses. We also need to become more visually literate and data-critical so we can be in a better position to analyze meaning from the striking image. Data can be beautiful and we all know the seductive power of beauty.

References: Stone, Maureen. "Information Visualization: Challenge for the Humanities." *Working Together or Apart: Promoting the Next Generation of Digital Scholarship*. Washington: Council on Library and Information Resources, 2009. 43-56 < http://www.clir.org/pubs/reports/ pub145/pub145.pdf>

This data is so hot it's smoking – animation from Flare.