

: *What is best practice for data visualization?*

by Dan Power

Editor, DSSResources.COM

Visualizing data means that we can see relationships if they exist. A visualization is a short cut to understanding the underlying pattern in the data. The danger is that the pattern is spurious and that we fail to test for the significance of the relationship. So what are "best practices" for visualizing a data set? Should we start with descriptive information on the data set? Should we have a prior hypothesis that guides our analysis? Should we try a number of techniques and tools and to see if we find a relationship?

The answers to the above questions are complex and the lay reader might easily be led astray by the technical arguments that some experts would make. In general, best practices are conservative. We seek to insure that a relationship really exists and that it is meaningful and useful. Be cautious in data analysis and display and use a supposition or hypothesis to guide a limited number of statistical tests for significant relationships. A hypothesis is a tentative explanation that can be tested. The dangers of undirected or unplanned data analysis are real and it is easy for analysts to fall into the trap of testing so many relationships that by chance they find a spurious or false relationship.

We want to begin with a hypothesis about a relationship and a data set that will allow us to test the relationship. The data set needs to be sufficiently large and derived in a manner that does not have bias so that the relationship is from the data and is not an artifact of how the data was gathered. We begin by doing a descriptive analysis of a relevant data set. Then we test a limited number of hypotheses to reduce the chances we will find a false or chance relationship. The scientific approach provides safe guards to insure the repeatability of the results and the truthfulness that we demand. Repeatable results reoccur predictably.

Data visualization can convey incorrect information as well as show meaningful relationships. It is important that everyone associated with creating and interpreting a visualization exercises caution in choosing the visualization tools and in interpreting the results. Visualization should tell an accurate story. The presence of a powerful visualization can actually hinder our understanding if we are misled to think that something is true when in fact it is false.

So how do we begin? We begin with a question we want to answer and a hypothesized answer. For example, we want to know who are our best customers. We hypothesize that married women with children are our best customers. The hypothesis helps define data we need to analyze. So we need to obtain purchase and demographic data and then find tools that will let us test the relationship.

: *What is best practice for data visualization?*

The visualization alone is not enough, we must use statistical tests to ensure that the relationship shown in the visualization is meaningful. We need to correct for the bias that would lead us to incorrectly conclude that a relationship exists when in fact it does not exist. We don't want a false positive -- a type I error -- any more that we want a type II error -- a false negative.

According to Vitaly Friedman (2008), the "main goal of data visualization is to communicate information clearly and effectively through graphical means." Data analysis insures we are effectively examining the data. We study and summarize data with the intent to find useful information and develop conclusions. We need effective data analysis and not just effective data visualization. Showing a false relationship is not helpful. We should **NOT** try a number of techniques and tools and see if we find a relationship.

Edward Tufte (1983; 2001) wrote a classic on advanced data visualization. He cautions against decorative and non-informative content added to charts -- "chartjunk", the lie factor, the data-ink ratio, and the data density of a graphic. One can lie with visualizations by picking deceptive scales or selecting data. He argues for using all the relevant data and presenting it accurately and in a visually attractive manner.

More recently Ryan Bell suggests we should begin by understanding the problem domain; get sound data; show the data and show comparisons; incorporate visual design principles; allow for quick visual comparisons; add extra levels of information and preserve the high-level summary data; add axes or coding patterns; and add a network metaphor to show complex connections. The basic design principles are: 1) align and position elements; 2) create clear contrasts and a visual hierarchy; 3) create visual unity with repetition of design elements across representations; and 4) use proximity and grouping of design elements with white space (cf., Tchakirides, 2011).

As Bell notes "Business analysts, IT staff and knowledge workers will need more skills designing, building and using fluid, interactive, dynamic visualizations." A good starting point is the theory and practice writings of Edward Tufte (www.edwardtufte.com). Finding and sharing the meaning in our data is our goal. Better visualization tools increase our ability to understand and persuade and our ability to deceive both others and ourselves.

References

Bell, R. "Eight Principles of Data Visualization," Information Management, August 17, 2012 at URL http://www.information-management.com/news/Eight-Principles-of-Data-Visualization-10023032-1.html?ET=informationmgmt:e3469:2078848a:&st=email&utm_source=editorial&utm_medium=email&ut

: *What is best practice for data visualization?*

m_campaign=IM_Da

Data visualization, from Wikipedia, the free encyclopedia at URL http://en.wikipedia.org/wiki/Data_visualization

Friedman, V., "Data Visualization and Infographics," Smashing Magazine, January 14th, 2008 at URL <http://www.smashingmagazine.com/2008/01/14/monday-inspiration-data-visualization-and-infographics/>.

Tchakirides, D., "Principles of Visual Design," Slideshow, January 21, 2011 at URL <http://www.slideshare.net/dianetch/principles-of-visual-design-6647916#btnNext> .

Tufte, E., *The Visual Display of Quantitative Information*. Cheshire, CT: Graphics Press, 1983; 2001 (2nd edition).

Tufte, E., *Envisioning Information*. Cheshire, CT: Graphics Press, 2001.

Whitson, N., "Best Practices for Data Visualizations," Slideshow, October 3, 2011 at URL <http://www.slideshare.net/visually/best-practices-for-data-visualizations-9527840#btnNext> .

Author: Daniel Power

Last update: 2012-12-23 04:38