Applying vision science to improve data visualization

How do we extract meaning from a vast stream of data? For security analysts, a critical part of doing this is seeing the data in visual form. If done right, such visualizations can allow our visual intelligence to spot trends, outliers, and other patterns in a set of data, patterns that might not otherwise be detected. In essence, such visualizations enable the operator to "use vision to think" (Card, Mackinlay, & Shneiderman, 1999). Combining this with more disciplined techniques in critical thinking has led to the emergence of the field of "visual analytics", where the goal is to create an analytics system that can noticeably amplify the intelligence of its user (Thomas & Cook, 2005).

But what is the best way to do this? Is there such a thing as a best visualization? If so, is there a way to find it? Considerable effort has been expended on this issue over the years. Many of the improvements in visualization systems have involved extensive use of vision science. A common approach has been to use what is known about human vision (sensitivity to color, say, or the limits of visual attention) to create a design that takes maximal advantage of the strengths of human vision while avoiding its weaknesses (see e.g., Horowitz & Rensink, 2016; Ware, 2012).

Another, deeper approach has also been developing recently. Here, the idea is to take a simplified version of a visualization design and investigate why it works. For example, how do people see correlation in a scatterplot? (Rensink, 2016). Careful investigation of this has shown that correlation perception can be done in a fraction of a second, although it has a pronounced bias (observers see less correlation than there is). What underlies this appears to be the ability to perceive the entropy of the dots in the scatterplot dot cloud. This explained many aspects of this correlation perception, and made several predictions. And in addition to uncovering a new kind of visual intelligence in humans, it also found new, simple ways to evaluate the effect of scatterplot design parameters such as color or size of dots. And finally, as a side-effect of the experimental design, it even uncovered a more compact—and potentially better—way of displaying data graphically.

Such work is just starting. But there is considerable potential in these approaches. Looked at more generally, a "science of visualization" could be a solid foundation for the design of analytics systems that will scale well as size and time constraints become increasing severe (Rensink, 2014).

References

Card, S.K., Mackinlay, J.D., & Shneiderman, B. Information visualization. (1999). In Card, S.K., Mackinlay, J.D., & Shneiderman, B. (Eds.) *Readings in information visualization: Using vision to think*, chapter 1. San Francisco: Morgan Kaufman.

Horowitz, T.S., & Rensink, R.A. (2016). Extended vision for oncology: A perceptual science perspective on data visualization and medical imaging. In B.W Hesse, D.K. Ahern, E. Beckjord (Eds). *Oncology informatics: Using health information* *technology to improve processes and outcomes in cancer care*. (pp. 287-303) Elsevier.

Rensink, R.A. (2014). On the prospects for a science of visualization. In W. Huang (Ed.), *Handbook of human centric visualization*. New York: Springer. pp. 147–175.

Rensink, R.A. (2016). The nature of correlation perception in scatterplots. *Psychonomic Bulletin & Review*. doi: 10.3758/s13423-016-1174-7.

Thomas, J.J., & Cook, K.A. (2005). *Illuminating the path: The research and development agenda for visual analytics*. National Visualization and Analytics Center.

Ware, C. (2012) *Information Visualization: Perception for Design*. (3nd edition) New York: Morgan Kaufmann.