Christine Zlotek

Dr. Patil

September 21, 2016

Data Visualization

**How to Graph Badly or What not to do:**

The author starts the article with a statement that graphing good visualizations depends on trying to not make mistakes. Chartjunk is often referred to as the features that artist add to visualizations that do not enhance the information content. Within this term chartjunk, there are factors such as fonts, shading, third dimension, and artificial color which can leave the reader misdirected or over interpreting the data. There is also the practice of “carpet bombing” which illustrates that authors, professors, or students who drop numerous graphs (bombs) in one single paper or place. This practice is very problematic because it is essential for one to know the purpose of the graphs, why they are important, and not just bombard them with incomprehensible information. Wainer has twelve rules for bad graphs, in which some will be discusses with remedies.

One principal is that graphs should not have little amounts of information but display as much high density data as possible, if presented in a clear manner. Another component of poor graphs is having a ridiculous amount of added elements or noise that the content is not clear. The components mentioned above (fount, shading, texture, ect.) need to utilized in favor of presenting the data, not hindering it. The scaling of graphs is important if the data wants to be comparable or consistent, having even a small increase on an axis can change how the data is interpreted. Another flaw of bad graphs is if they are unable to give the reader enough information through the text, caption and labels, thus making it unable for the reader to understand the data. Even though captions and labels may seem small or even an irrelevant factor, they degrade graphs if incomplete or done poorly. There are numerous factors that make for bad graphs, but for each mistake there is a remedy and the goal is to make it clear and understandable for the reader.

**The Gospel According to Tufte:**

The idea of “data-ink” is defined as, “the non-erasable core of a graphic”. To better understand this definition there are five maxims to elaborate his data; (1) above all else show the data (2) maximize the data-ink ratio (3) erase non-data-ink (4) erase redundant data-ink (5) revise and edit. The first is the most important because it is similar to an essay, there needs to be a sufficient topic sentence or purpose of the graph. The second maxim has to do with the data-ink ratio which focuses on the portion of the graph that can be erased without the loss of valuable data information. This second theory wants to eliminate the noise and keep only the relevant information. The third concept is erase non-data-ink which Tufe relates to eliminating gridlines, half framing graphs, and using a minimalist bar chart. The fourth maxim, erase redundant data-ink, focuses on when to use symmetric or antisymmetric graphs as well as when using a table is more appropriate than a graph. Lastly, is the revise and edit where graphs can have multiple drafts that do not change the data but the clarity for readers.

As discussed in the last article, data density (having adequate amount of data in a graph) is crucial. Tufte calculates data density by the number of entries in a data array divided by the area of the data graphic. In data dense graphs the goal for the reader is to be able to identify themes or goals in the graph, this is easily accomplished using the shrink principal. Tufte also endorsed “small multiples” which is where there is a sequence of images that all have a slight change that makes it appear it is a continuous motion. The noise elements in data graphs are discussed further, primarily the emphasis on word labels over letter labels, how separation and laying are more advanced with technology, and the degree to which dimensional or flat graphs are useful. The aspect ratio says that width of the page should be divided by the height of the page because humans naturally have better vision horizontally than vertically. Likely, when looking at graphs, color plays a large role in the ease of reading. Color needs to emphasis the correct elements in a graph, by sticking color with natural associations (do not make the ocean brown in a graphic because society associates it with blue) and tint can be useful to ensure the labels are eligible. Parallelism can be useful because the message is easier to grasp because of the axis and format are consistent and only the data varies.