Summary of How to Graph Badly or What NOT to Do

Nick Blackbourn

September 20, 2016

While graphing can be a great way to display and analyze data, it can also be a source of confusion and frustration for the average reader. One of the flaws that gets introduced in this article is the term known as "Chartjunk". This can be defined as outlandish decoration or artistic detail that adds no statistical value for the viewer when trying to draw conclusions about a dataset. Some of the main techniques associated with this term include: "Moire" shading, Pseudo 3-D, and Artificial Color. When a graphic designer engages in "Moire" shading, he or she is using shading patterns that are overwhelming to the viewer and draw attention to the exterior parts of the graph. Pseudo 3-D is a concept showing that how a graph is designed can make all the difference in the observations made about data (pie chart, ex.). Similar to the first two readings in the course, color can be defined as the make or break element to a graph and should only be employed when there aren't other options to clearly display the data. We see this with the car example at the end of the article. In order to clearly see the driver's controls inside the vehicle, the diagram needs to have the proper shading which will allow the viewer to only see this part of the car.

Throughout this piece, the author discusses the importance of finding meaning in data by including more information in less graph displays (bar chart, pie chart, line graph, etc.). The term used midway through the article to describe a documents extraneous amount of graphs is what can be described as graphical "carpet bombing". In order to avoid this issue, people need to know how to condense a data set and assess the crucial points, be looking for ways to combine similar graphs into one, and avoid adding any features that will take away from the data (high data density).

Another key point about graphs is how labeling can affect our observations. This is especially apparent when looking at graphs that have been combined into a single one with multiple datasets. Without any words or numbers attached in the appropriate places, the reader will have a difficult time answering questions. This will especially help with graphs that have multiple dashed and solid lines as well as help the reader get information about the "context" of what's in front of him or her. These labels can be harmful when there aren't enough of them, they haven't been finished, or the font used is illegible.

In addition, one other concept that comes up in this reading is called "Data Hiding". This occurs when there are so many extraneous details included that the main details of a graph are invisible. This can be observed with the dark black grid in Figure 1.7 and the author's lack of data in Figure 1.8. "Data Hiding" can also occur when there are two separate quantities with completely different magnitudes graphed on the same scale. By doing this, it does not allow us to visualize a difference on that same graph.

`Summary of The Gospel According to Tufte`

After analyzing the piece by Tufte, he makes it very clear which points are the most crucial when putting together graphs for datsets. These include: maximize the data-ink ratio, above all else show the data, emphasize the data, erase non-data and redundant data-ink, and revise and edit.In looking at Tufte's principles, emphasizing and displaying the data can be combined into one for several reasons. When asked to show the data on a graph, not only is your intended purpose or meaning of your diagram crucial, but also how the data curves are depicted. Tufte talks about the graphs each having a similar labeling style or format to make inferences from the different data points. One way to do this is by expanding the date lines with thicker or more pronounced ink that sticks out to the reader.

One element of erasing non-data ink is the presence of grid lines when plotting graphs. For the most part, grid lines should be avoided on diagrams with the exception of nonograms (graphical calculator), three-dimensional figures, and special cases where it's required to draw a conclusion. If they absolutely need to be used, they should be lightly drawn to prevent them from impacting our conclusions. Another way to condense a graph is to use a half frame as opposed to a full frame. This only utilizes the horizontal and vertical axes and still provides the reader with the same data in a more condensed form. In looking at bar graphs, one can eliminate the bars and utilize line segments to simplify the data and reduce the amount of shading required.

For redundant data ink, even though many people are able to use the generic concept of symmetry to visualize a full object from only one half shown, Tufte emphasizes the fact that showing the image entirely is the best way to visualize the graph and remember it in the long-term. While it is an important feature to include the whole graph, one should always be looking for ways to simplify and eliminate non-graphical data that takes away from it's intended message.

An important principle that seems to be common in looking at graphs and data is "high data density". This is known as the process of combining several different data groups (200 thermal conductivity measurments in Figure 2.12) into one single graph. While this seems like a crazy course of action, the idea behind Tufte's theory is to spot themes or goals that can be taken away from the graph's purpose/thesis. This can be found by looking for clear labeling, shading, and organization of the different data elements. This can be compared to the Shrink Principle in that by simplifying data plots and graphs into one single figure, we are able to make conclusions based on one simplified page instead of multiple scattered pages.

When looking at graphs, there are many different ways that one can view them as tools for analyzing data. For example, graphs can also serve as table-graphics in how they are displayed (left to right). In Table 2.2, even though the numbers are positively skewed to the right, we are able to draw conclusions based on the labels included on the horizontal axis. While tables can serve as graphs, they are also known to be a better depiction of data when there are too many groups or columns present in a dataset. This requires us to eliminate the graph entirely and use a more structured and organized table to make inferences about the data.

In Tufte's writing, he states that there are several ways to make data more understandable to the average viewer. One of the ways he explains this is with his principle known as "small multiples"/animations on a page. This is defined as a method of reducing the amount of graphs one sees and reduce multiple graphs into a mere one. For many datasets, space can be seen as a substitute for time on a graph. Tufte called this the Collapsing Dimension/Escaping Flatland principle (Japanese beetles). In addition, this article emphasizes the importance of a wide aspect ratio as opposed to a heightened one because this allows for more labels, words, and numbers to be included without confusing the reader.

Throughout our analysis of graphs, color continually comes up as an essential element in our ability to differentiate between data and non-data/exterior elements. Unfortunately, color can be used the wrong way and certain color patterns can hinder us from making meaningful observations about a graph. This can occur when many different colors of the spectrum are put together in one designated area. One proposed solution is to place contours on top of a pseudocolor map to create separation between color patterns.