Nicholas Russell

Data Visualization

September 21, 2016

Assignment #2

**How to Graph Badly or What Not to Do**

An effective learning strategy for making successful graphs is to learn how to make poor, uninformative graphs. This is because skill in visualization depends on not making mistakes. This article looks at the main mistakes of graphs and charts, specifically chart junk, Wainer’s rules for bad graphs, high data density, data-hiding, inconsistent visual metaphor, context free-data, area instead of length as a visual metaphor, label woes, emphasizing the unimportant and unnecessary graphic novelty.

Chart junk is defined as features that add nothing to the information content of a graph. These examples can be found in newspapers and non-technical magazines, because of professional artists who are interested in visual beauty rather than getting a clear message across. Examples include unnecessary fonts that may be illegible and attention grabbing shading that brings focus to unimportant features. Pseudo 3-D graphs, like the one example of the Maize production, make certain portions of the graph seem bigger than others. Artificial color is an expensive tool that doesn’t convey much information. One must only use color when it is necessary. Over interpretation is found when a result is desperately wanted to be found, like the example of the African bees, where the other made a fake correlation in a random cloud of data. The last example of Chart junk is actually too many similar graphs which loses a reader. If many graphs are needed then the author should vary the graph type, highlight the key figures in the caption and the graph itself, triage the analysis and combine closely-related graphs into a single multi-panel graph.

Wainer develops rules for bad graphs. The rules are as follows:

1. Show as little data as possible [minimize the data density]

2. Hide what data you do show [minimize the ratio of data/ink]

3. Show the data inaccurately [ignore the visual metaphor and randomize the connection between graphical elements and the numbers]

4. Use length as the visual metaphor when the area of two-dimensional icons is what is actually perceived

5. Graph data out of context [sparse captions and vague text]

6. Obfuscation #1: Change scales in mid-axis

7. Obfuscation #2: Emphasize the trivial [ignore the important]

8. Obfuscation #3: Jiggle the baseline [use different axis ranges for two graphs which will be printed side-by-side and need to be compared]

9. Obfuscation #4: Alabama first! [Order the data by some criterion, such as alphabetical order, which is irrelevant to all of the interesting patterns in the data]

10. Obfuscation #5: Label: (a) illegibly (b) incompletely (c) incorrectly (d) ambiguously

11. Obfuscation #6: More is murkier: (a) more decimal places and (b) more dimensions

12. If it has been done well in the past, think of a new way to do it [New graph types are sometimes needed, but they require a lot of concentration from the reader, and should be used sparingly in good graphics] in the next few sections, we will illustrate some of these principles.

Low data density is only justified when emphasizing very important information because it displays little information or illustrating a concept but low density data graphs tend to display little information which is why high density data graphs are more desirable. Data hiding is a bad strategy that involves adding elements that distract from the actual data points. One common mistake is graphing the similar quantities on the same scale which can hide data. Inconsistent Visual Metaphor is common mistake for a relationship between a specific graphical elements and data. A graph is a failure is the data is context free, like if the text and caption fail to give the reader enough information so that the graph is comprehensible or if the curves fail to make comparisons. In other words a good graph should show enough curves with enough information.

As humans we have trouble accurately representing area and volume in graphs. That is why using length better for the reader. With the barrels example what may be seen as twice as big is actually three times as big. Illegible labels are surprising problem that stem from too small type size, poor placement and too few labels. Graphs can be over labeled that distract from the curve or too few labels that hinder clarity. Emphasizing the unimportant comes from overemphasizing data that removes focus from the main point. In figure 1.16 the solid black color jumps off the page and we immediately emphasize based off of color rather than numerical value. Unnecessary the Nightingale rose may be an effective for certain researchers but the general population has to take more time to understand the graph.

**The Gospel According to Tufte**

Edward Tufte covers certain principles and truths in accordance with proper graphing. This author takes note of his theories and evaluates them. The topics he covers in relation to graphing are data ink, data density, multifunctioning graphical elements, small multiples or animations on a page, elements of a graph work with one another, layering, separation and rubrication, word labels, escaping flatland, supplementary material, aesthetics of aspect ratio, the ineffectiveness of color, parallelism and friendly graphic.

In regards to data ink, he Tufte advises to show the data by describing what the data is about. Then he suggests emphasizing the data by “maximizing the data-ink ration”. In other words use thicker ink for the data curves and deemphasize the axis lines, tic marks and labels. Tufte is, in most cases, against using grids because of its distracting nature and is especially against non-data ink like the frame of graphs and standard bar chart. Tufte is for symmetry like using 5/3 of the globe but he is a political scientist so he is not familiar with the world map as much. When all else fails and a graph does not seem to be the answer, Tufte asserts to use a table instead. Despite that these are graphs and not papers, these charts and graphs still need revision and editing to make sure there are no mistakes.

Tufte is a big fan of high data density but the key themes and goals of the figure need determine the design and the not the other way around. High data density is a virtue but is not more important than clear labels, clear theme and other essential features that made a graph perfect. Sometimes a high density graph may be hard to read. A better strategy may be two graphs or a black and white graph. Multifunctioning graphs are very effective that can convey more than two pieces of information. In regards to animations the symbolic equation is “small multiple” graph – one animation with morphing between frames. When labeling graphs lines, the label should be above the line but close enough to realize that they are related. One should use full words and numbers when labeling so the reader does not have to go back and forth. When using a timeline, it is best to use a long a spaced out timeline to see the differences in the graph.

Other aspects included in the authors account include inset graphs meaning to show the transformation between two graphs and the external aspect ratio of a graph which is Re= width on page/height on page. Color is again brought up as an expensive asset that can be used wrongly and needs to be used carefully in a way that is not distracting. Parallelism is effective when a plot is parallel in layout as well. Lastly the author creates a friendly data chart with friendly titling one column and unfriendly titling the other. One of my favorite examples of friendly is the graphic attracts viewers and provokes curiosity because it makes sense to me. An example of unfriendly is using a “design insensitive to color-deficient viewers; red and green used for essential contrasts.