Assignment 2: Summary of *How to Graph Badly,* and *The Gospel According to Tufte.*

David Boose

September 14, 2015

The purpose of a graph is to convey a large amount of information in a very small space. Graphics take advantage of the fact that humans can take in and process visual information much more rapidly than we can through any of the other sensory systems. While we obviously read words using our eyes as well, the cognitive process of making meaning from text is much slower than that of discerning patterns and relationships among images. Certain similarities and differences among visual elements can be detected literally within milliseconds, with no conscious mental processing. And the brain can gain meaningful, detailed understanding of an image in far less time than it would take to describe that understanding in words. So graphic representations can be very information-dense.

But to be effective, graphical representations of information must be carefully designed. Because the initial interpretation of images happens so quickly and without conscious thought, a carelessly created graph may lead the reader to "see" a very different message than that intended by the designer. Similarly, poor graph design can obscure patterns that the author wishes to make visible. John Boyd, a faculty member in engineering at the University of Michigan, presents some basic recommendations for designing effective graphs that work with, and not against, the natural processes by which we make meaning from images. In the first two chapters of his class notes, Boyd first teaches by (bad) example, with information on "How to Graph Badly". He then lays out and comments on some more formal recommendations from Edward Tufte, an influential leader in the theory and practice of data visualization. These recommendations can be organized and applied to the three primary components of any graph: the pattern, the variables, and the context.

Pattern

Graphs are most effective at showing relationships among two or more variables–that is, patterns in the change of one variable with respect to another. Graphs are generally not as useful when the focus is on specific values of a variable; in that case, a table is a better choice. Rather, the goal is to depict the direction, rate, and relative magnitude of changes in the variable, features that are much more difficult to glean from a tabular listing of values. Similarly, a graph is not necessary if the relationship between two variables is constant. In that case, there is nothing added to the understanding that cannot be adequately described in a few short words. There are, of course, exceptions. Boyd cites a couple of examples where the actual data values are used as elements of the graph, so that both the pattern and the values can be discerned. But even in these cases, the central feature of the graphic–the thing the eye picks up first–is the pattern.

In this usage of "pattern", I don't necessarily mean something regular or repeating. Rather, I use it to mean a visually detectable relationship among graph elements. A random distribution of points in a plane can depict a meaningful relationship between two variables that the eye would quickly discern, and I would include that in my definition of "pattern." To be effective, a graph should be designed to emphasize, and not hide, the patterns of interest. At the same time, it should not inadvertently create visual impressions of patterns that either don't exist or are unimportant. In *How to Graph Badly*, Boyd illustrates how adding a "3-d" design to a pie chart creates a misleading comparison of the size of the wedges in the circle. He also points to such errors as overinterpretation and "graphical carpet bombing" as examples of nonexistent or unimportant patterns. He emphasizes the need to "triage" your analysis, deciding which patterns are most important and therefore deserve graphical representation, and which do not.

That's not to say that a graph can't contain a great deal of informaton and present it clearly and effectively. But doing so requires careful planning. First, the elements used to depict the variables should be those that the visual system processes most accurately. We interpret length information much more accurately than width, area, or volume, for example. Second, the representation of those variables should be consistent with the pattern being depicted. Colors, shading, or differences in line width or form should reflect important differences–identifying different values of a nominal variable, for example. They should not be used when they attract attention to unimportant differences. Similarly, data should be ordered by meaningful criteria, not irrelevant ones. For example, when graphing poverty rates among the various U.S. states, it would make more sense to order the data from highest to lowest, rather than alphabetically by the name of the state ("Alabama first!"). In particularly data-dense graphs, it may be beneficial to omit those elements that don't contribute significantly to the important pattern being depicted. Boyd uses two examples–a graph of 200 measurements of the conductivity of copper at different temperatures, and a map of the densities of galaxies in the visible universe–to show how removing some of the data can actually make the pattern of interest clearer. Finally, colors should be used carefully. Bright, contrasting colors convey the message of large contrasts in values, which may be what is desired. But including too many such colors creaes a visually confusing graph; if numerous gradual changes are to be depicted, a more muted color scheme is likely to be more effective.

Variables

A pattern only has meaning if one understands the variables that are in relation to one another. So while the overall pattern may be the first thing the mind perceives, the graph must also clearly identify and give the characteristics of the variables involved. This includes: the variable names; whether they are continuous, ordinal, or nominal; and for categorical variables, the names of the categories. Labels need to be designed so that the graphical elements that represent different variables can be easily and quickly identified, without repeated reference to a key or a caption. Thus Boyd recommends using words instead of letters as labels, placing labels as close to the features they identify as possible, and avoiding distracting elements such as boxes around the labels. If labels can't be placed close to what they identify without creating clutter, then lines may be used to connect them, as long as those lines are distinct from any lines connecting data values. For increased clarity, such lines can be made a different color, to distinguish label lines from data lines.

Another key task is deciding how many variables to include in a single graph. As the number of variables goes up, the mind's ability to perceive important patterns among them goes down. Sometimes, variables can be omitted entirely without loss of important information. Boyd illustrates this with two examples of figures depicting sunspot patterns. In one, the longitudinal dimension of sunspot location is ignored, since the important pattern relates to latidude and time. In the other, time and space are collapsed into a single dimension of the figure. Because sunspots move across the face of the sun, time and location are essentially redundant measures of the same thing, so one variable can represent both. Another strategy for reducing the dimensionality of a figure is to plot several similar figures together. Boyd discusses this strategy under three different headings: "The Shrink Principle" (section 2.2.1), "Small Multiples" (section 2.4), and "Parallelism" (section 2.12). When several small graphs are placed together, they essentially become elements in a larger graphic. The ability of the viewer to detect patterns among those elements will depend on both the differences in content among the panels, and the similarity of the "background" components (e.g., axes). If the individual panels dislplay lines that vary in length, position, or slope, for example, while the framework surrounding each panel is constant, the eye will detect that quickly. If the panels depict shapes that differ in area or smaller surface details, or if the frames vary in their size or orientation, those differences will be harder to see.

Context

Finally, the graph must provide the context within which this relationship among variables is to be interpreted. This includes the range of values over which the variables are changing, as well as any other important information, such as location, time of day or year, directional perspective, etc. Equally important is the theoretical context. Some of this may be considered as part of the "pattern" that is embodied in the graph. For example, plotting observed data from an experiment is much more informative if the results that would be expected from theory are plotted along with them. Some of this information will be given in a caption, or in the text describing the figure, but the less a reader needs to move back and forth between image and text, the more effective the visualization will be.

At the same time, more is not always better. A key principle here is that the graph should, in Tufte's words, "maximize the data-ink ratio." That is, the graph should include the fewest and least intrusive elements possible while still providing sufficient context. For example, Tufte argues strongly against grid-lines and full frames around graphs. Values along axes provide important information on the ranges of the variables involved, but reducing the number of tick marks and labels to the minimum eliminates distraction and improves clarity. He even goes so far as to argue that half of the data can be eliminated in a symmetrical or anti-symmetrical graph, because our brains can fill in the missing components based on what is there.

But this "erasing" can go too far. Boyd points out that while our brains are certainly *capable of* filling in the other half of a symmetrical figure, we are not neessarily *used to* doing that. At a minimum, the additional cognitive processing involved in completing the figure takes time, slowing the rate at which meaning is being made and reducing the effectiveness of the figure. Worse, the viewer may simply miss the fact that the figure is symmetrical, thereby overlooking a significant feature of the graph. A similar effort at reduction that can hide important results is plotting data of very different magnitudes on the same axis. In this case, adding a second axis–or even a second graph–with a different range of values will allow a clearer comparsion and convey more information.

Simplify, simplify, simplify

The ability of the human visual system to take in and process so much information so quickly is a double-edged sword for graphical designers. On the one hand, an effective graphic can tell an important and detailed story in a very compact space. On the other hand, our eyes will take in all that a graphic has to offer, and will make sense of it according to specific rules and processes. If the graphic is not designed with those rules in mind, the impression the viewer creates may not be the one the designer intended. So graphs must emphasize those features the designer feels are important, and remove all elements that are not essential to telling that particular story. Boyd describes those features that distract from the clarity of the graph as "chartjunk," and he includes in that category such things as hard-to-read fonts, unnecessary or misleading coloring or shading, visual effects such as 3-d, and superfluous text. As with any good piece of communication, a graph must be carefully revised and edited for maximum clarity. Having others take a look at draft versions is also an excellent idea, because the patterns we think are so obvious may not be as clear to a naive viewer. It has been said that writers shouldn't write to be understood; rather, they should write so that they can't be misunderstood. Designers of data visualizations should follow the same injunction, and should carefully craft their images so they take maximum advantage of the powerful information processing capabilities of the human eye and brain.