The Elements of Graphing Data, Chapters 1-2
Summary by Nick Siler and Regina Carns, 11/19/2009

Adam C. and Brian S. led this week's seminar, which covered the first two chapters of William Cleveland's book *The Elements of Graphing Data*. The seminar was divided into three parts: 1) Presentation of the main points of Chapter Two, 2) Discussion of the controversial points in Chapter Two, and 3) Critical evaluation of examples of graphs.

Part 1: Presentation of the four main points in Chapter Two

1. Make the data stand out. Avoid superfluity.
    a. In some cases, having a graph at all may be superfluous. Make your figures count.
    b. Use visually prominent elements to emphasize your data and to draw attention to the point of your figure. (However, as discussed later in class, making elements *too* visually prominent can make graphs harder to read.)
2. Do not clutter the data region.
    a. Avoid "chart junk", which includes anything in your graph that does not convey information to the reader.
        i. Steve W. cited the example of university logos that often appear at the bottom of Powerpoint slides at conferences.
        ii. As Adam C. pointed out, Edward Tufte likened chart junk to the Big Duck store on Long Island, which was built in the shape of a duck. In his book *The Visual Display of Quantitative Information*, Tufte writes, "When a graphic is taken over by decorative forms or computer debris; when the data measures and structures become Design Elements; when the overall design purveys Graphical Style rather than quantitative information, then that graphic may be called a duck in honor of the duck-form store, 'Big Duck'."
        iii. Rob W. noted that 3-d histograms and other 3-d graphics qualify as chart junk since they are often more difficult to understand than simple 2-d graphs. Brian S. pointed out that most default chart types in Excel and other Microsoft products are so cluttered as to be unusable.
    b. In the interest of reducing clutter, Cleveland recommends always placing the legend outside the graph. Many people objected to this, however, noting that the location of the legend should depend on the data presented in the graph and the formatting guidelines of the journal in which it is published.
3. Strive for clarity.
    a. Adam C. noted that "every figure should convey a clear point." In other words, it should not require great effort on the part of the reader to understand the data presented in the graph or the conclusion the graph is meant to support.
    b. Brian S. recommended making the points on a scatter plot sufficiently prominent to be easily noticed by the reader.

c. Adam C. said that, in his experience, “What people remember from a paper is often what they see, not what they read”. He therefore suggested spending as much time developing your figures as writing your paper.
d. A question arose as to whether it is necessary to include a best-fit curve as part of a scatter plot. Al R. argued that best-fit curves should be included because they reveal the pattern of residuals, which is often important. However, others argued that a simple scatter plot is sometimes sufficient, depending on what the author wants the plot to convey.

4. Graphing data should be an iterative, experimental process.

Part 2: Discussion of the controversial points of Chapter Two

1. Offsetting the zero point on one or both axes.
   a. In cases where the data include points that fall on the x- or y-axis, Cleveland recommends offsetting the zero point so that the points at zero do not fall directly on top of the axis line.
   b. Adam C. found this problematic, since it might imply the possibility of negative data points on graphs where none was possible. The class, however, felt that this practice would generally improve clarity.

2. Using outward-pointing tick marks.
   a. According to Cleveland, tick marks should point outward, in keeping with the general principle of clutter reduction.
   b. However, Adam C. argued that they should usually point inward, since tick marks that point outward tend to draw the reader’s attention away from the data. After some discussion, a vote revealed that most of the class agreed with Adam.
   c. Two vocal dissenters were Ed W. and Steve W., who both testified to only using outward-pointing tick marks. Ed noted that inward-pointing tick marks often draw too much attention to data points that lie close to them, and correspondingly too little attention to data points that lie farther away.

3. Placing data points within a box.
   a. Adam C. wondered whether having a box around the data was necessary at all, or if it simply qualified as clutter. The class generally felt that a box was helpful.
   b. Steve W. pointed out that a box indicates where to look for data.
   c. Sandra P. also argued that when quantitative data are presented, the right and top borders should have tick marks as well as the x- and y-axes. When only the axes have tick marks, it is often difficult for the reader to estimate the numerical values of the data.

Part 3: Critical evaluation of examples of graphs

For the final portion of the seminar, the class was divided into five groups, each of which evaluated 3 different graphs that had been published in scientific journals. All of these graphs had at least one major flaw that obscured the content and/or significance of the data presented. The groups discussed the graphs independently for about ten minutes. The groups then came together to discuss

each graph, highlighting its errors and discussing ways it could be improved. Suggestions for improvement included:

1. Choose symbol sizes and line thicknesses to draw attention to the point of the graph while maintaining readability. Data-point symbols that are too large, or lines that are too thick, will clutter the figure and obscure the data. Symbols and lines that are too small or thin, on the other hand, will be difficult to read, especially when reduced.
2. Use data resolution appropriate to your conclusions. Smoothed data may be easier to read. Tick marks should be spaced appropriately for the data resolution.
3. Consider carefully how much data should appear on one graph. Many graphs show too much data, obscuring their overall point. Others use multiple graphs to display data that would be presented more clearly on a single graph.
4. Use a visualization appropriate to your data; not all data should be displayed on a scatter plot. When the independent variable is naturally divided into discrete units (such as a graph of attendance figures for an annual conference), histograms will usually present the data more clearly. Tables may even be used instead of graphs when there is little data to present.
5. Take care when using effects such as stippling, hatching, or other visually noisy patterns. They can make figures hard to read and unpleasant to look at, and they may cause problems if they are reduced in size.
6. Make sure each element of the graph is clearly identified. Use informative axis labels. Include a legend if necessary.