Making Better Graphs

If you allow it, most commercial spreadsheet software will generate lousy graphs. You should learn how to modify the various parameters of your graphs to produce better results. Remember that the reason you are creating a graph is to provide information to the reader, and to clarify statements that you make in the text. Just as you edit your writing to make it clear and concise, you should tailor each graph based on the information it contains and the message you want to convey.

A Bad Graph



Figure 1: A graph of data that was collected by hooking up a meter to the circuit without the muffler belts showing measured voltage as we changed the current. Note that the yellow data clearly fits a horizontal line.

A particularly bad graph is shown in Figure 1. This graph shows two sets of data, represented by the yellow and light blue data points. This is the first problem: relying on color to distinguish between different sets of data is a mistake. Your documents will invariably be scanned, faxed, and copied before your boss's boss's boss sees them and colored objects will all become the same ambiguous shade of gray. Lighter colors, like yellow and light blue, will barely be seen at all. Similarly, lightly colored text within the graph may disappear.

The labels for the Y axis are placed next to the origin line near the horizontal center of the graph, which is the default setting for the software used to create it. This can confuse the reader about which set of labels corresponds to which axis title, and the labels may be obscured by the data points themselves.

Make sure that the limits for each axis are chosen to eliminate unnecessary white space in

your graph. In this example, the limits for the logarithmic Y axis start two decades (powers of ten) too low and extend one decade two high.

The number formats for the tic labels on both axes were poorly selected. The labels on the X axis have three digits to the right of the decimal point even though two of those digits are always zero. This clutters the chart and fools the reader into thinking that somewhere on the graph there is a label that might really need all of those digits. The labels for the Y axis are also poor. They have unnecessary digits and the choice of the scientific 'E' format is hard to read.

The title for the X axis should indicate the units used for the data. The title for the Y axis has specified units, but for a logarithmic axis you should always use the base units... volts in this case.

When plotting measured data (as opposed to a mathematical function) be very careful about "connecting the dots". In Figure 1 the top (yellow) set of data points has been fit to a trend line that assumes a linear relationship between current and voltage, but the true relationship happens to be a sinusoid. Don't add a trend line unless you have reason to believe that the measured phenomenon actually behaves according to the function selected for the trend, and the fact that the data points appear to line up is not enough evidence. The bottom (blue) set of data points has been connected with a spline curve. This is a completely arbitrary function to associate with these measurements and it can be misleading. In general, you shouldn't need to add lines between data points unless there are several sets of measurements in one graph and it helps the reader keep them sorted out, and in that case just use a straight line segment between each pair of points. In any case, your text should always explain why a particular line was added to the graph.

Finally, the caption for this graph is far too long. The caption should be just long enough to clarify the difference between the various graphs that may be presented in the same document. Details about how the measurements were collected or conclusions about their meaning should be provided in the narrative text.

A Better Graph

The same data values are plotted again in Figure 2. No color is used so this graph should be just as readable after it is faxed. The two sets of data can be distinguished by the different symbol used for the data points: a diamond for one and a square for the other. A legend in the upper right corner clarifies the difference, and the same information should be provided in the narrative text. An alternative to using a graph legend is to add text annotations next to each set of data points, but the particular spreadsheet software used here does not have that option.

A trend line was added to the bottom set of data points, but I happened to know that these measurements should follow an exponential function so that is what was used for the trend line. Once again, the narrative text should explain why the trend line was added. This software doesn't have an option to use a sine function for the trend line, so the upper data set is presented without a trend line.

The labels for the axis tic marks have been moved to the edges of the graph area and their number format was adjusted to make them more readable. The font size for these labels was also increased so that the size of the text in the graph would be roughly the same as the text in the rest of the document, even after the graph was resized to fit the available space. The label



Figure 2: Circuit response without muffler belts

for the X axis now indicates that the current values are in milliamperes.

The caption has been shortened considerably, knowing that the text explicitly refers to the graph and thoroughly discusses the methods used to obtain the measurements. Your narrative text is also the best place to discuss any conclusions you draw from the graph. Note that if the graph was taken from another source then the caption, as well as the text, should provide a citation of the original source.