Graphing Style

Introduction

The purpose of using graphs to present data is to help you and the reader understand what the data are trying to say. We will spend a lot of time in this class thinking about how to graph data sensibly. This handout covers some of the basic points.

Be sure to read the paper by Loftus, “A picture is worth a thousand p values: on the irrelevance of hypothesis testing in the microcomputer age.”

Identify variables to be used in plots, and assemble data points

Generally, you will have one dependent variable per graph (the data set), organized by values of one or more independent variables (the experimental conditions). The first independent variable is used for the ordinate (x-axis), and others will be represented as different plots on the graph. The dependent variable is used for the abscissa (y-axis).

Variables can be discrete (separate values such as male/female, New Jersey or New York, Happy/Sad or large/small) or continuous (values which represent a point along a continuum such as stimulus size, drug dosage or score on a 12-point happiness scale). Continuous variables are often plotted with connecting lines or adjacent bars, while discrete variables shouldn’t be joined since the space between them doesn’t have any meaning (halfway between male and female?). Data from different observers can be represented on the same graph as another independent variable, but this can lead to cluttered graphs and should be avoided. Generally you should only combine observers onto a single graph if you need to show how similar or different their results are, or if there is only one other independent variable being represented on the graph.

Choose appropriate type of plot

This depends on the type of data, how many points there are and how they are organized, and what the graph is seeking to show (differences between groups or similarity of data to a model etc.). Choices include scatterplots, data points connected by lines (most common), histograms and many more. For plots which represent means rather than individual data points, it is customary to include error bars to give the reader an idea of the variability of the sample.

What axis scale to use?

To do this, first find the largest and smallest values. As a rule of thumb, use min and max values that allow a bit of “breathing space” above the highest point and below the lowest point. Think about what limits make sense, however. Decide whether it is sensible to include zero in the scale. Including zero will enable readers to compare the relative sizes of different values of the dependent variable, but may result in the data points getting squashed together and mask effects in the experiment. It’s your call. You will want to add about 5 “ticks”, labeled with values of the variable. Use numbers with as few significant digits as possible (10, 20, 30 rather than 12, 24, 36 and 1.05, 1.10, 1.15 rather than 1.025, 1.075, 1.125). Sometimes the variable values themselves will dictate the values to be labeled. (If set size in your experiment has values of 8, 16, 32, then these are the only labels you need to include.)

Note that if your graph includes a model, or set of reasonable predictions or assumptions, these should be included when figuring out the ranges. Also, be sure to figure out the extent of any error bars ahead of time.

Often you will have several different graphs that plot similar data. For example, you might run an experiment using two different subjects and graph the data for each. An important principle is to use the same axis scaling across graphs that show comparable data. This allows easy comparisons to be made. To do this, you will often need to change the axis scale from the default produced by the graphing program.

How should I label the axes?

Appropriate level of description

The goal is to create a graph that is clear without requiring the reader to refer to the text. Of course, this is open to interpretation—one person might find an axis label clear without much explanation, while someone else may need several paragraphs of explanation that are best left to the body of the paper. Aim for a label that provides an accurate description of what that axis represents, but without using too many words.

Examples of labels with too little information: "Means", "Response Value"

Example of label with too much information: "Average motion threshold of a field of stationary dots after adapting to a field of moving dots (mm/sec)"

About right: "Average motion threshold (mm/sec)"

Units of measurement:

It's quite important to include the units of measurement in the labels you give the axes. For instance, if the X-axis shows the different line lengths you presented to your subjects, and the units shown on that axis are measured in pixels, make a label that says something like: "Line Length (pixels)", rather than just "Line Length".

Font, size, style

You may use your own judgment about which Font type to use. The main rule of thumb here is to use the same font, size and style for both axes.

Plot the data

Add the data points. There are many ways to do this. Favorites include a small "x", or a point surrounded by a circle, triangle, square etc. Sometimes it is appropriate to use a filled-in symbol (e.g. square), especially when it represents the mean value of (scatter) plotted data. If you are plotting lines, join up the dots with straight, dashed or dotted lines.

For a histogram or bar graph, be sure to make the bars of equal width, and only allow them to touch if they represent a continuously measured variable. If there are multiple bars representing multiple independent variables, it may be appropriate to add a pattern or color to the bars to help distinguish them.

Add error bars, if you are using them, and possibly a symbol to represent significance levels.

Make a legend

Each graph which contains more than one type of symbol or line needs to have a legend illustrating what each of these stands for. Legends should be clearly labeled as such, and should be concise. Place the legend outside of the area bounded by the axes.

If you use multiple graphs to represent the data from different observers, in which the plots are otherwise identical and share a legend, it is customary to identify the graph with a 2-letter code for the subject in one of the top corners, away from the data and line plots. Alternatively, multiple plots from one observer can be grouped as sub-plots of a single figure, with the observer identified by code name in the caption.

Write a caption

Every figure (graphs are identified as Fig. 1, Fig. 2, etc.) must have a figure caption. The caption should include the figure number, followed by a clear and concise description of what is being plotted (what are the variables), and which symbols/line types are used to represent them. Do not attempt to describe the data, but identify whether points represent raw data, means etc. Also describe the meaning of any error bars.

The standard form of a caption describes a graph as "Figure (n). (graph type) of (dependent variable, y-axis) as a function of (independent variable, x-axis, other independent variable, ...) for (number of observers / observer ID). Data are plotted as means of (N) observations. (independent variable value 1) are represented by (symbol type 1, line type etc.). Predictions of the (model type?) are represented as (line type). Error bars represent ±1 (SE or SD)."

When submitting fully formatted papers for publication, figure captions are included on a separate sheet at the end. For our purposes, you can write them on the same page as the graph, just below the graph.

Color

Be careful in the use of color in your graphs. Color can make things easier to discern when colors are carefully chosen. On the other hand, graphs are often printed on black-and-white laser printers (e.g., before we read and grade them!), so it is generally prudent to make graphs entirely with black symbols, lines, etc., and use symbol types and line styles to differentiate different conditions.

Common Graph Errors and Codes

In grading lab reports, there are certain comments that we find ourselves writing out over and over again. To save our time and to alert you to the common mistakes, here is a list of error codes that will be used in making comments on graphs. The order is not indicative of the relative importance. Nor is this list exhaustive.

Code 1: The error bars run off the edge of the graph. This is an error. The solution is to re-scale the axes. Click directly on the axis and a dialog box will come up that lets you choose the axis limits.

Code 2: The fonts on the axis labels don't match. This is ugly. See “Changing Axis Labels” below to see how to change the fonts and font sizes.

Code 3: Use solid/dotted lines to distinguish different conditions graphed. This is more a suggestion than an error. How strongly you should heed it depends on how close the lines are on the graph. In addition to changing the symbols, you may also change the lines that connect them. The same dialog box where you change the symbols has additional options for these attributes. Another way to distinguish line types is to use a colored felt tip pen and draw over the individual lines afterwards.

Code 4: The heading is moderately cryptic or contains an abbreviation that might not be obvious to all readers. This may or may not be an error, depending on what you said in your text to accompany the graph. But if there is space on the graph, it is usually nicer to spell things out.

Code 5: It is useful to indicate somewhere on the graph or in the text what the error bars denote. Typically you will be using the standard deviation or the standard error of the mean.

Code 6: The heading is very cryptic. This is an error. Use meaningful headings and labels.

Code 7: Provide units for numerical values when appropriate.