Appendix A

Presenting Your Data -Tips for Making Tables and Graphs

You've just done an important experiment and need to figure out what the data have to say, as well as present them in a way that communicates effectively to others. This is where tables and graphs come in. When data are properly organized and presented, the story they tell is easier to see and patterns that support or refute the hypothesis become clearer. You want your graphs to present the maximum amount of information, while being easy to interpret.

Data can be presented in many different ways. Which way you choose depends largely on the type and complexity of the data. But all tables and graphs have to contain enough information to be interpretable without looking for an explanation in the accompanying text. This means that they should have a title and labels that identify treatments and the units of the data being reported.

INDEPENDENT AND DEPENDENT VARIABLES

A hypothesis identifies one or more factors that influence the process or pattern that you are studying. These factors are called **variables**. The factor hypothesized to determine change in an experiment is the **independent** (or explanatory) **variable.** This is usually, but not always, the factor you manipulate or vary among treatments in an experiment. The factor that responds to changes in the independent variable is the **dependent** (or response) **variable**—so-called because its value depends on changes in the independent variable. For example, suppose you think that a certain compound influences plant growth, and set up several treatments in which plants are grown in identical conditions except for the amount of this substance that is present. In this case, the abundance of the substance is the independent variable; some measure of plant growth is the dependent variable.

Many experiments include more than one dependent variable but only one independent variable. For example, you might hypothesize that the substance influences growth rate (gain in height/day), aboveground biomass (the weight of the shoot system at a specified point in time), and number of seeds produced. All three are dependent variables. But other factors that might influence growth such as light and water availability are best held constant, to be evaluated in separate experiments. Light and water are other independent variables.

Experiments that evaluate the effects of more than one independent variable (and the interactions between them) are possible and often extremely interesting, but the experimental designs and statistical analysis required to draw valid conclusions are beyond the scope of Bio180.

TABLES

Tables present data in chart form. The variables are arrayed in columns, usually with the independent variable located to the left and the dependent variables to the right. A table is a good choice if:

- The independent and/or dependent variables occur in categories that are hard to graph (e.g. lists of species, locations, characteristics, or other qualitative data);
- Exact numbers or only a few data points need to be shown; or
- Data for multiple dependent variables are presented.

Tables are numbered consecutively in a report and each should have a descriptive title. They may also have a legend or caption that conveys information about sample size or how the data were acquired. Footnotes sometimes clarify or add information that is difficult to include in headings. Units must be shown in column headings, wherever relevant. For example:

Table 1. Characteristics of Four Species of British Poppies Sampled in 1978
100 individuals of each species were sampled.

Species of *Papaver*	Distribution	Mean No. of Flowers per Plant	Mean No. of Seeds per Flower	Mean % Germination of Seeds
P. argemone	Uncommon*	6.8	314	63
P. hybridum	Uncommon*	7.3	230	91
P. dubium	Common[1]	6.8	2,008	42
P. rhoeas	Common[1]	12.5	1,360	64

*occurs in less than 10% of the meadows examined
[1]occurs in more than 60% of the meadows examined

GRAPHS

A picture really *is* worth a thousand words. Often, displaying data in graphical form is much more compelling than presenting the same data in a table. In Bio180 labs, you will use scatter plots, bar graphs, and perhaps other kinds of graphs.

Scatter Plots

Scatter plots are best when:

- Both independent and dependent variables are continuous—meaning that they can take many values within a range of values;
- More than two or three data points need to be plotted;
- Only a few data sets are involved—usually no more than 3-4 lines on the same graph—so the graph remains uncluttered and easy to read;
- You want to emphasize trends in the data.

Follow these steps to produce good graphs:

1. By convention, **the independent variable *always* goes on the horizontal axis and the dependent variable is *always* plotted on the vertical axis**. In a graph of growth rate versus temperature, for example, temperature goes on the horizontal axis and growth rate goes on the vertical axis. This is because an organism's growth rate changes with (depends on) temperature, and because the temperature in an environment is not influenced by how fast organisms are growing.

   Time can be an independent or dependent variable, depending on circumstances. Usually you want to measure something that changes as time passes, such as the height of a bean plant measured daily for a month. In this case, time is the independent variable and goes on the horizontal axis. Why? In this experiment, time goes by independently and does not depend on the height of the beans you're measuring. Occasionally, though, time is a dependent variable and belongs on the vertical axis. For example, time is a dependent variable when the amount of time taken to accomplish something is being measured as a response to change in some other variable. Think about how you'd graph the time needed to finish a marathon when individuals in different treatment groups had received different types of liquids during the race. In this case, you are hypothesizing that time depends on the liquids ingested.

   Sometimes the relationship between the variables you are plotting isn't dependent or independent—they simply vary together, meaning that they are **correlated**. If you are plotting leaf width against leaf length, you could put either length or width on the horizontal axis since there is no reason to think that changes in width cause changes in length, or vice versa.

2. Organize your data. If you are making a graph by hand, putting the data in low-to-high or some other order can help you recognize the minimum and maximum values and decide the proper scale to use.
3. If you are plotting by hand, determine the minimum and maximum values to be displayed (the **range** of data points). Then decide on the origin. For many data sets, the point where the *x*- and *y*-axes cross will be the 0,0 point. This is not appropriate, though, if it results in a lot of empty space. For example, if the variable on the *x*-axis ranges from 500-900, it might be better to have the *x*-axis start at 450 than at 0. Set the maximum value shown on each axis so that it is only slightly larger than the maximum value in your data set.
4. If you are plotting by hand, divide each axis into evenly spaced intervals, marked with ticks, to span the range of the data. Number enough tick marks so a viewer can read them easily and interpret the data points, but don't clutter or overcrowd the axis. For example, if data values range from 0-200, you might put tick marks at intervals of every 10 units, but only number ticks at 50 unit intervals. Label each axis, identifying the variable and units involved.
5. Plot the data points. Use different symbols or colors, and provide a key, if more than one set of data is shown on the same graph. Avoid putting too much on one graph—generally no more than 3-4 lines per graph.
6. Should you connect the data points? Lines that connect data points can be added to clarify the relationship between variables, especially if more than one set of data is being plotted on the same graph. In many cases, though, you will want to add a line of best-fit—a line that averages out random variability. A statistical method called linear regression provides a way of finding the best-fit line. A regression line helps identify the general trend, if one is evident, and may also allow you to predict the value of Y for a given X within the range of the existing data, but for which you have no data point.
7. Finish the graph by giving it a title, an informative caption that summarizes the conclusions that can be drawn from the graph or that clarifies the conditions of the experiment, and a key that identifies any symbols or other added information. A graph should be capable of standing alone, without the need to go to the text to figure out what it says. Figures (graphs and other diagrams) are usually numbered separately from Tables in the same report.

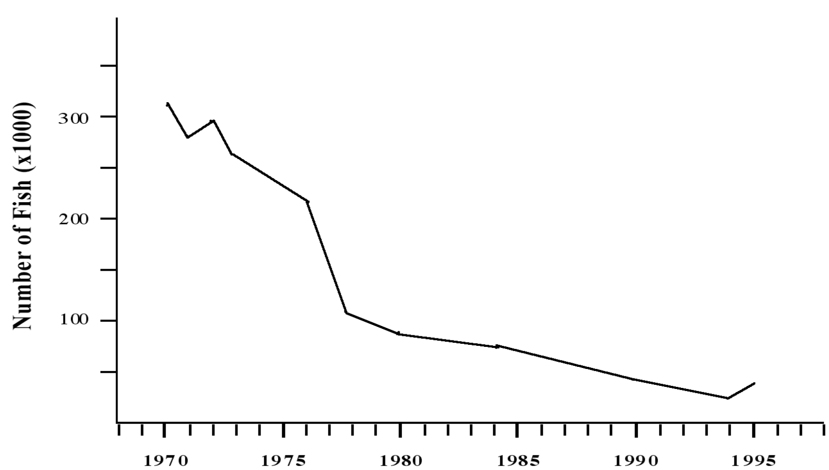

Fig. 2: Estimated Population of Bluefin Tuna, Age 8 and over, in the West Atlantic, 1970-1995

Note that scatter plots have to be interpreted carefully. The main issue is that you can't assume that changes in X cause changes in Y, even if the data show a clear relationship between them. Stated another way, a **correlation** between two variables does not always indicate a cause-effect relationship. The variable on the *x*-axis may indeed cause changes in the variable on the *y*-axis, as hypothesized, but it could also be true that the changes observed in both variables are related to changes in some other factor. For example, suppose you decided to plot the average number of cups of coffee consumed in Seattle each year from 1950 to 2009 versus the average number of cars per year passing over the I-5 University bridge each year over the same time period. There will almost certainly be an impressive increase in both over this time interval. There is a strong correlation between the variables, but no cause-effect relationship. The real cause of the increase in both variables is the increase in population.

Bar Graphs

Bar graphs are used when:

- The independent variable is discrete, meaning that the variable being plotted occurs in separate categories. Sometimes these categories are non-numerical (like colors or species or geographic locations); sometimes they may represent different treatment groups (e.g. males vs. females, or low vs. high food availability); sometimes they represent a range of numerical types grouped together (e.g. plants 0-10 mm tall vs. plants 11-20 mm tall). In each case, the categories are each treated as single data points. Note that the dependent variable in a bar graph must be continuous.
- A visual, rather than tabular, comparison of categories is wanted, with emphasis on comparing data points more than overall trends.

  Bar graphs are often used to plot means (averages) from different treatments. In general, bar graphs follow the same rules as scatter plots, except that values are represented by vertical bars rather points. Each graph must have a title, axes and units must be labeled, independent variables must be shown on the *x*-axis, and so on.