In reporting the results of scientific investigations, it is often useful to present data in the form of a graph. When graphing data in the natural sciences, by convention the independent variable is placed on the X (horizontal) axis, and the dependent variable is placed on the Y (vertical) axis. Since there are many different types of graphs, choosing the proper graph for your data is essential. The type of data you collect will generally determine the type of graph you will use.

Data types: data collected in a scientific investigation will be in one or more of the following classes:

Numeric data. Numeric data involves measurements that are quantitative, and where the numbers also describe a mathematical relationship to each other. Snout vent length (SVL) is an example of numeric data. A snake that has a SVL of 40 cm is twice as long as a snake with a SVL of 20 cm.

Numeric data can be either **continuous** or **discrete**. Continuous data is any data which can take an infinite number of values over a finite interval. Length measurements are an example of continuous data. Discrete data is data that can take a finite number of values over a finite interval. An example of discrete data would be the number of peas in a pod (there cannot be 2.5 peas in a pod).

Ranked data. Ranked data is data that is ordered, but there is no mathematical relationship implied by the data. Birth order is an example of ranked data (the second born is not twice the first born, hence the numbers do not have a mathematical relationship to each other except to rank the variable).

Attributes. Attributes are discrete character states that have no mathematical relationship. A common attribute data variable used is alive or dead. Other examples could include presence or absence, infected or not infected, etc.

Graph types:

Scientific data is often presented using one of the following graphs:

Scatter plot   Pie chart
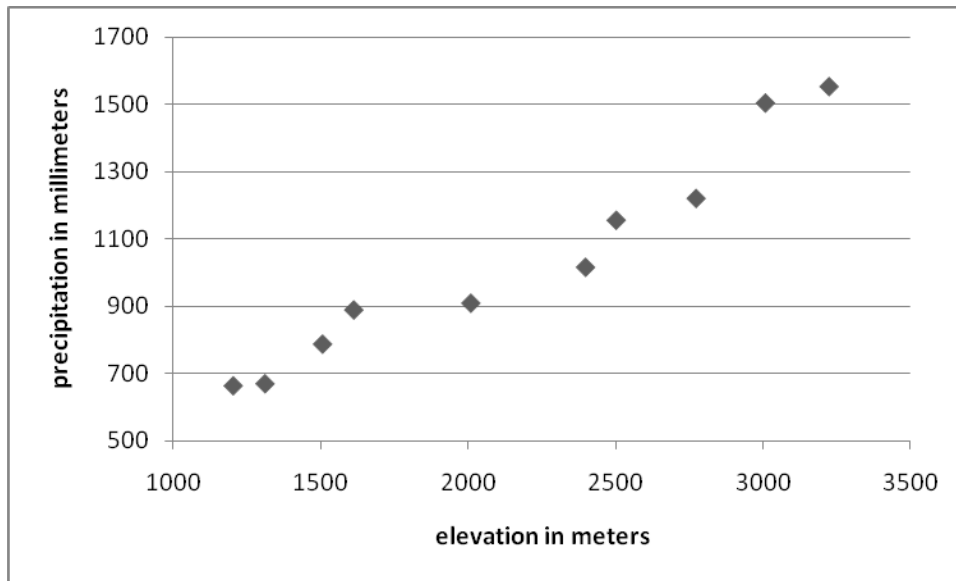
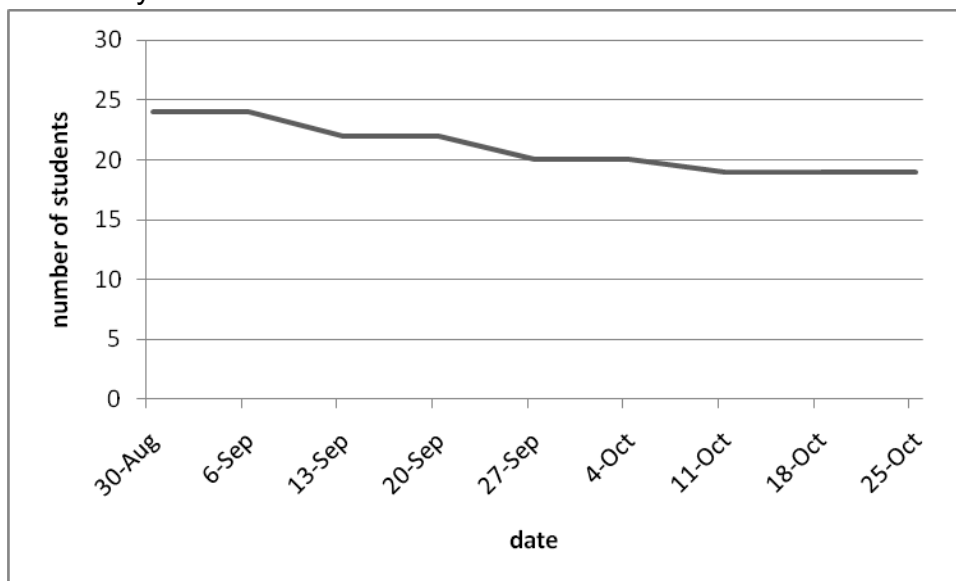Line graph   Histogram

Bar graph

Scatter plot. A scatter plot is used to graph data where both the independent and dependent data are numeric. An example of a scatter plot would be examining the relationship between elevation and yearly precipitation.
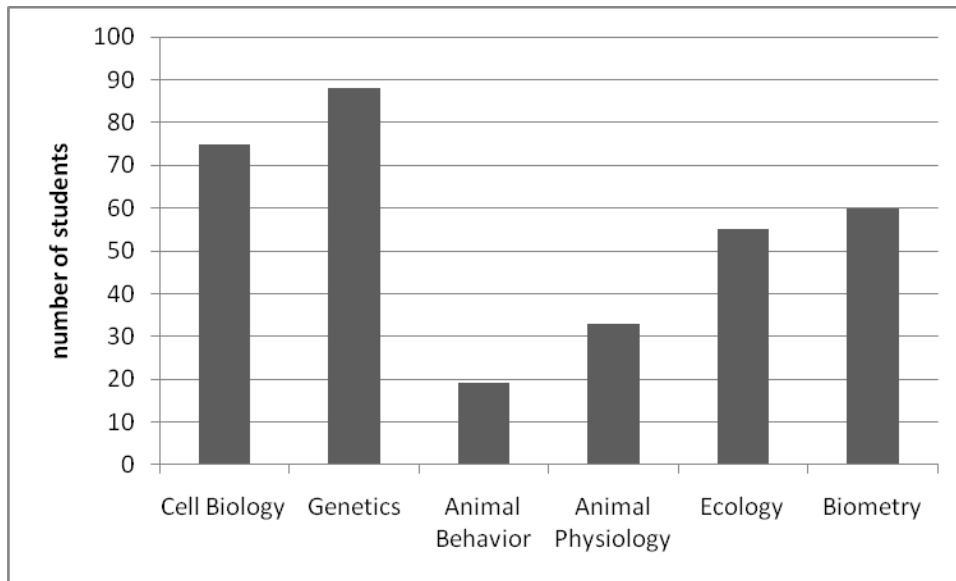

Notice that the axes do not start at zero. One should choose the axis values to maximize the resolution of the data.
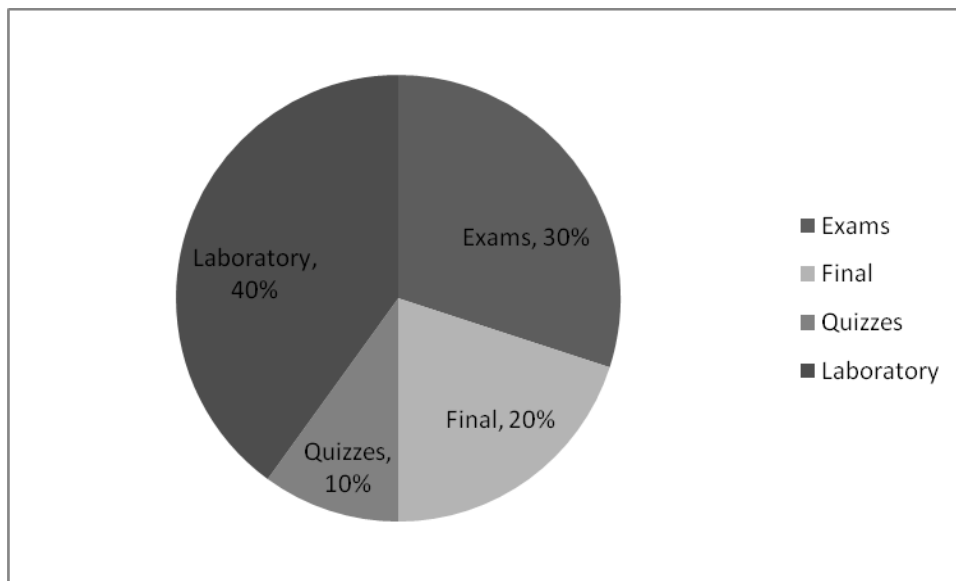
Line graph. A line graph is used when both variables are numeric data and the author is determining if there is a trend in the data. A common use of the line graph is to describe changes in the dependent variable over time. An example of this would be how many students are enrolled in this lab over the course of the semester.
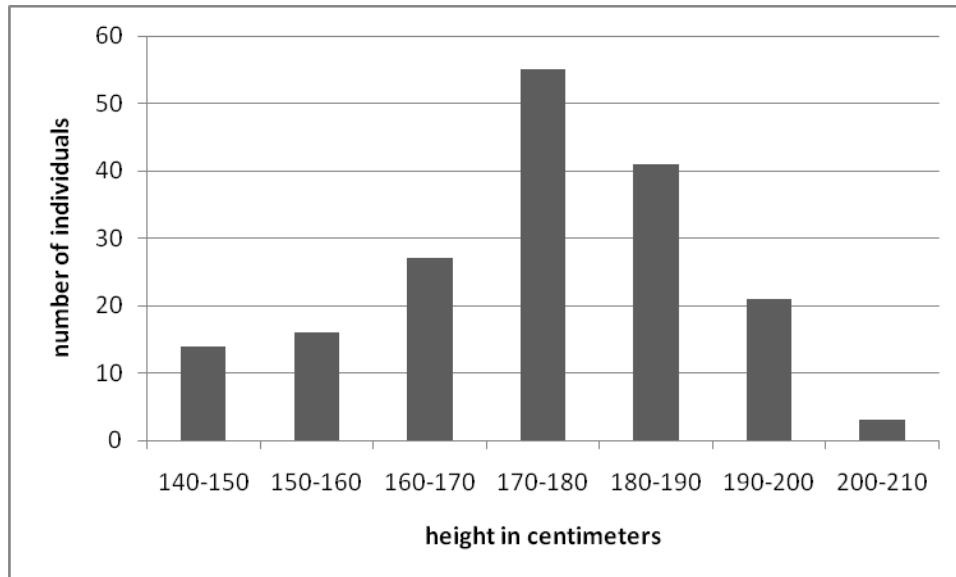
Bar graph. A bar graph can be used when the independent variable is discrete, ranked or an attribute. An example of a bar graph would be enrollment comparisons of the different science courses offered at ASU West.

Pie chart. A pie chart is used to show percentage data or parts of a whole. An example of a pie chart could be a description of how your grade is calculated in a college course:

Histogram. A histogram is used to group continuous data into discrete categories. An example of this would be human height. If you graph each height individually in a scatter plot, the results would be difficult to interpret. By grouping the data into intervals, the data can be graphed in a manner that facilitates interpretation.

To make a histogram, one must decide how many groups, or intervals, one wants to sort the data into. In general, twelve to twenty groups should be used, although this is also dependent on the sample size of the data that was collected. The size of each group should be the same, such as the example above where each group spans 10 cm.

Measuring Variability of Data: the Standard Error and Standard Error Bars.

Often in presenting data, the researcher will need to convey the amount of variability in a sample. One way this can be done is to put a standard error bar on the graph. Standard error bars are appropriate when you are graphing an average as opposed to graphing individual data points. The standard error is a measure of how much variability the mean of a sample would have if repeated samples were taken. The standard error is calculated by dividing the sample standard deviation by the square root of the sample size. The standard deviation is found by taking the positive square root of the variance. The variance is caluculated by first subtracting each data point from the mean, squaring this difference and then suming it for all of the data points. Finally, this sum is divided by the number of samples that were used minus one.

In this laboratory all of these statistics can be calculated in Excel. The following section gives the formuals to calculate the statistics of variation for reference.

The formula used for the variance (S^2) is:

$$S^2 = \frac{\sum (x_i - \text{mean})^2}{(N - 1)}$$

Where

    N = number of data points in the sample]

    X_i = the ith data point

$\sum$ indicates that you must sum the results of this formula for each data point (i = 1 to N).

To find the standard deviation (S), simply take the positive square root of the variance:

$$S = \sqrt{S^2}$$

And finally to calculate the standard error (S_x), simply divide the standard deviation by the square root of your number of samples:
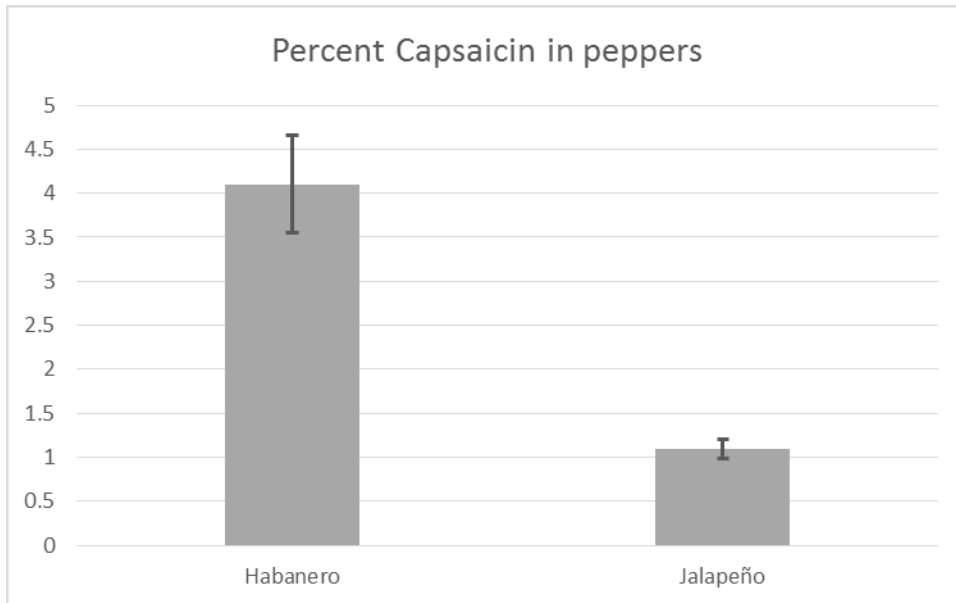
$$S_x = S/(\sqrt{N})$$

Calculation of the standard error can be done in Excel, and the values used to determine the length of the standard error bars. After the standard error for each average is calculated, draw the bar graph in Excel. Then select the data and click on the design menu under chart tools at the top of the screen. Under the design menu, click on the button to add chart element. Under the pull down menu you should see error bars. Under this menu select custom values and use the standard errors you calculated. Please note that checking off the box on the menu for standard error will NOT graph the correct error bar. You can easily determine this by looking at the size of the two error bars-if they are the same size, it is (almost always) incorrect.

An example of a bar graph with standard error bars can be made from the following data:

Variety:	
Habanero	Jalapeño
3.24	1.10
2.43	0.880
5.45	1.01
4.41	1.50
4.98	0.993

Table 1. % Capsaicin by dry weight in five Habanero and Jalapeño peppers.

Your graph should appear similar to the following:

Percent Capsaicin in peppers

For the following problems, choose the appropriate graph type and graph the data. Where appropriate, put standard error bars on your graphs.

1. A research botanist is trying to determine when in the year that most Mexican poppies (*Eschscholzia glyptosperma*) bloom. The scientist has recorded how many flowers are open on a study plot each week for six weeks.

Date:	# of flowers
1 March	6
8 March	11
15 March	25
22 March	21
29 March	15
5 April	6

What can you infer from the graph of the data about the time of year this flower blooms?

2. An animal physiologist is trying to analyze the relative contribution to diet different food sources make for burrowing owls (*Athene cunicularia*). The scientist has collected the following data:

Food source	Percent of diet by weight
Mammals	45%
Reptiles	15%
Invertebrates	40%

What can you infer from the graph about the diet of burrowing owls?

3. A research ichthyologist is attempting to study the relationship between size and reproductive rates in Apache trout (*Oncorhynchus apache*). The scientist has weighed ten fish, and recorded the number of eggs they produced.

Weight (grams)	Number of eggs
235	112
225	104
334	201
354	199
407	254
505	214
487	306
612	411
774	398
801	409

What can you infer from the graph about the relationship between the weight of a trout and the number of eggs it produces?

4. A research limnologist is comparing primary production in four different lakes. The scientist has collected the following data:

Lake	Sample number	Primary productivity (grams carbon fixed per square meter per day)
Roosevelt	1	0.56
	2	0.42
	3	0.11
Pleasant	1	0.65
	2	0.66
	3	0.54
Canyon	1	0.32
	2	0.31
	3	0.22
Powell	1	0.68
	2	0.71
	3	0.66

What can you infer from the graph about the productivity of these lakes?