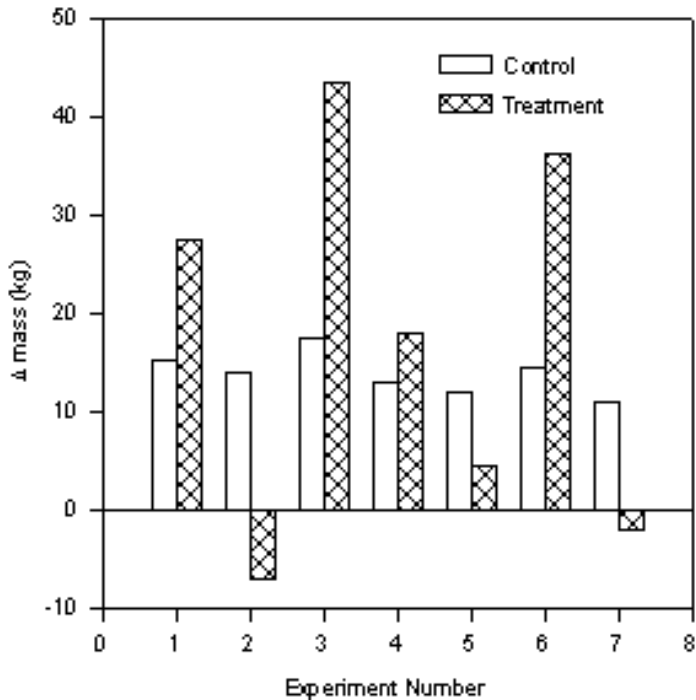


## Graph Types in Biology

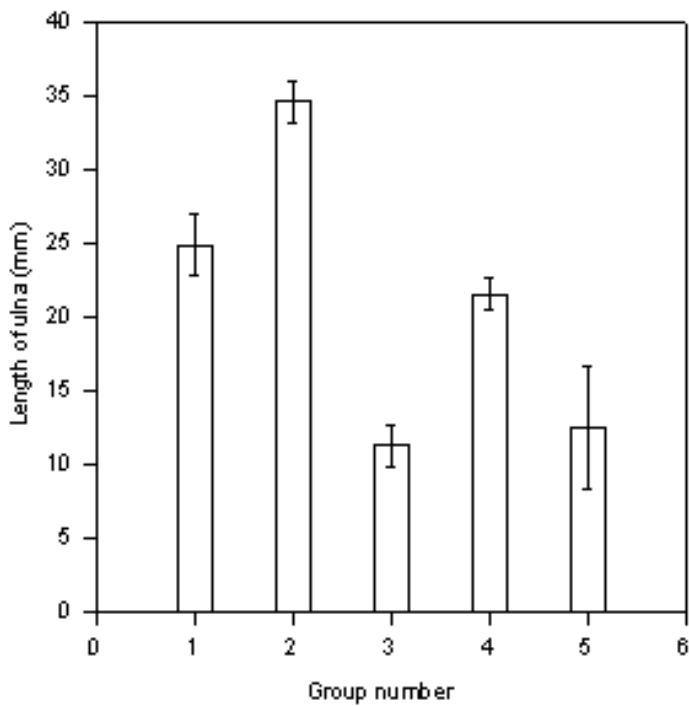
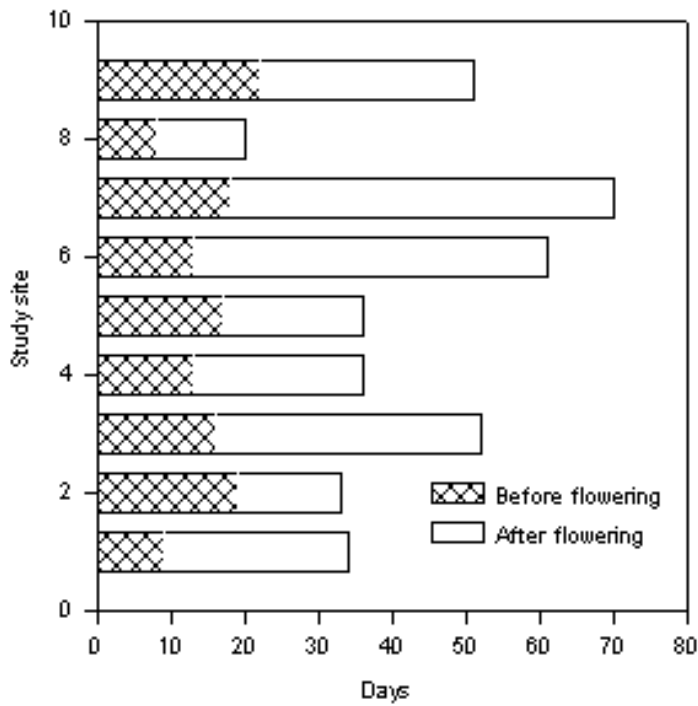
Despite the variety of graphs used in business and the popular press, there are only a few basic styles used in biology, and generally straightforward criteria for which to use in each situation. The object of graphing is to depict numeric data visually, so it is important to avoid visual elements that do not add to seeing the data, and to choose a graph design that visually shows the comparisons you intend to make. Following are examples of five basic styles of two-dimensional graphs, along with a brief discussion of 3-dimensional graphs.

**Bar graphs:** These are best used to show numeric data that represent discrete items or experiments. Bars imply that there are no intermediate values (contrast with lines below), and in many (but not all) cases the order of the bars along the X-axis will be arbitrary.

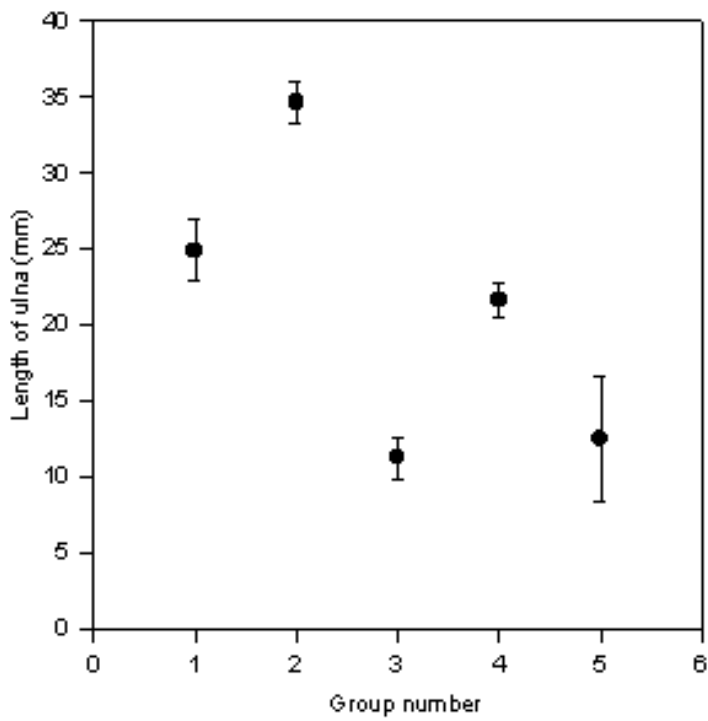


**Side-by-side** - Bar graphs can contain but a single series of data, but when they contain more than one, the additional series can be arranged in two ways. In a side-by-side graph, the bars are exactly that. This allows the series to be visually compared on an item-by-item basis

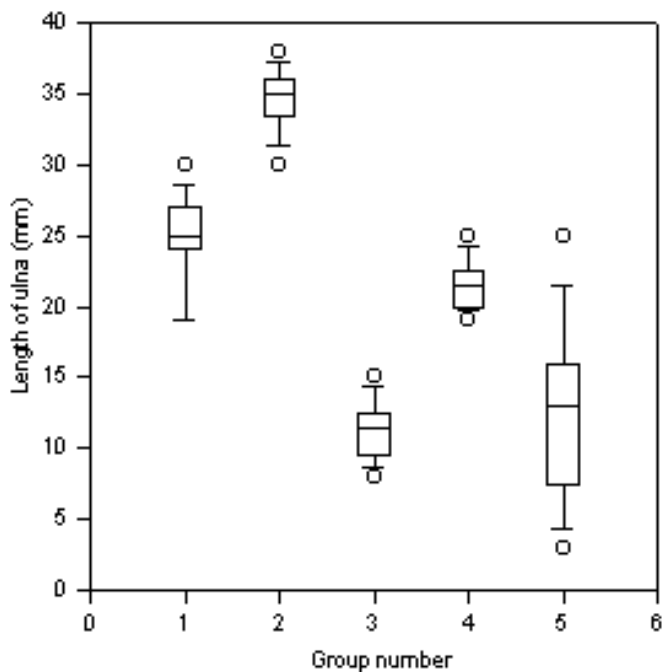
**Stacked** - Sometimes the numeric values for an item accumulate between series, and the important visual comparison is between items rather than series. In this case, a stacked bar graph is more appropriate. In this example, the bars are oriented horizontally, because the flow of time is often represented horizontally, and the X-axis is now the dependent variable. As a general rule, horizontal bars should only be used if there is a reason to do so.



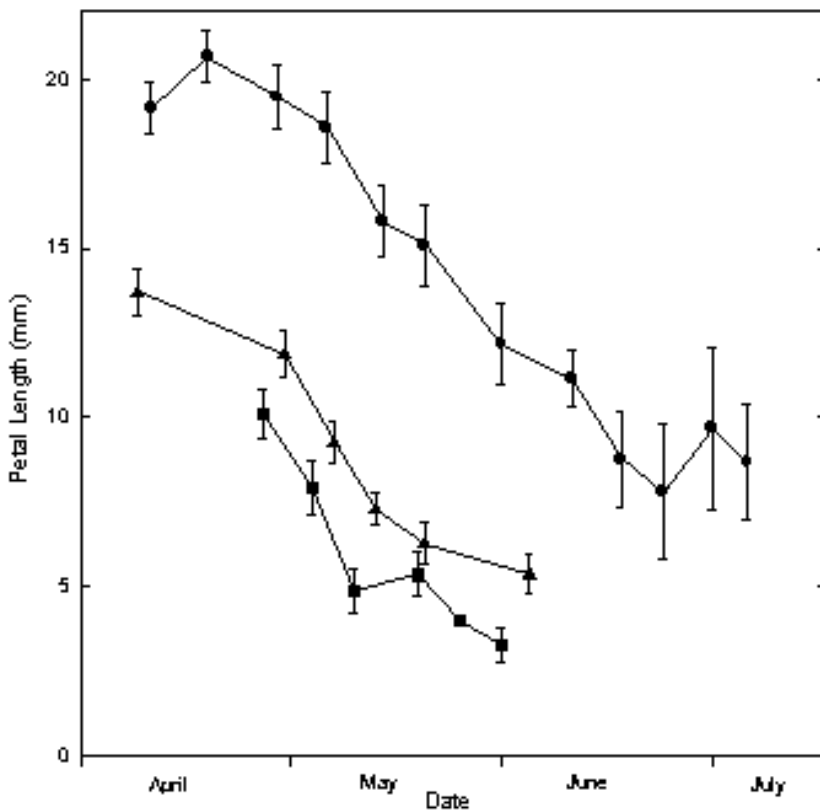
**Error bars** - When the numeric value of a bar is a mean, it is often important to show variability. A common way of doing this is with error bars: lines extending above and below the top of the bar to show some aspect of variability, such as the standard deviation, the standard error of the mean, or the 95% confidence level of the mean. The error bars can extend up away from the top of the bar only, or both above and below (in that case the bar should have no fill).



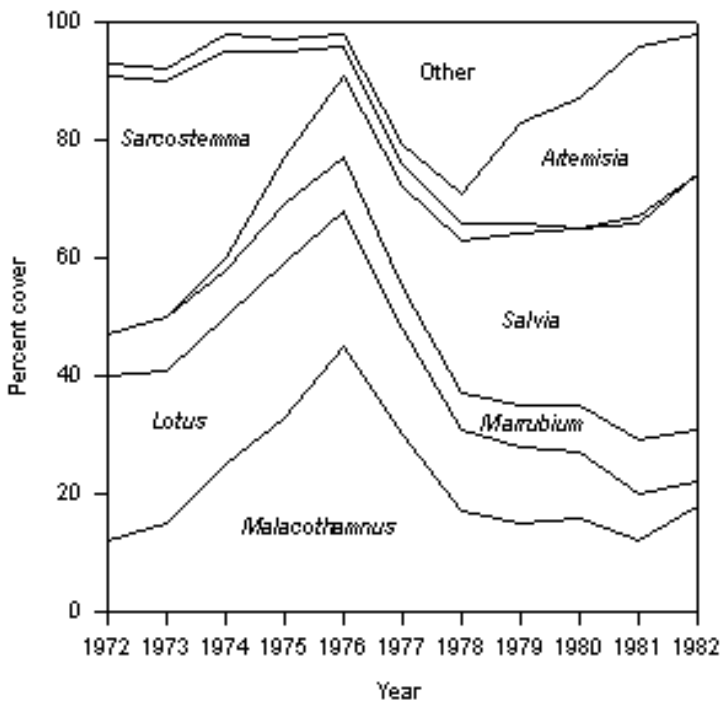
**Floating error bars** - The same graph can be constructed without the bars: the error bars remain, but the mean is now represented by a symbol. The choice between this and the graph above is not straightforward, and different disciplines characteristically use one or the other. The example above visually stresses comparison of the means over comparison of the variation; this example stresses comparison of the variation, and de-emphasized comparison of the means.



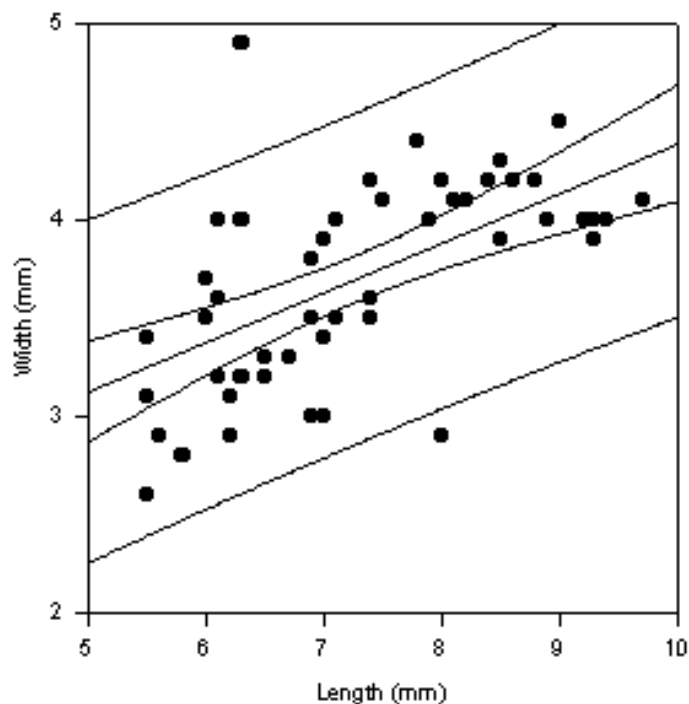
**Box plots** - Sometimes it is useful to show a visual representation of variability in data without resorting to parametric measures of variation. A box plot depicts the median, rather than the mean (although many graph programs substitute the mean), and the quartiles (the 25% of the data above the median and the 25% below the median). This example adds thin lines including 90% of the data, and those individual data points that are outliers. Note that these are not confidence intervals; they are measures of the actual data.



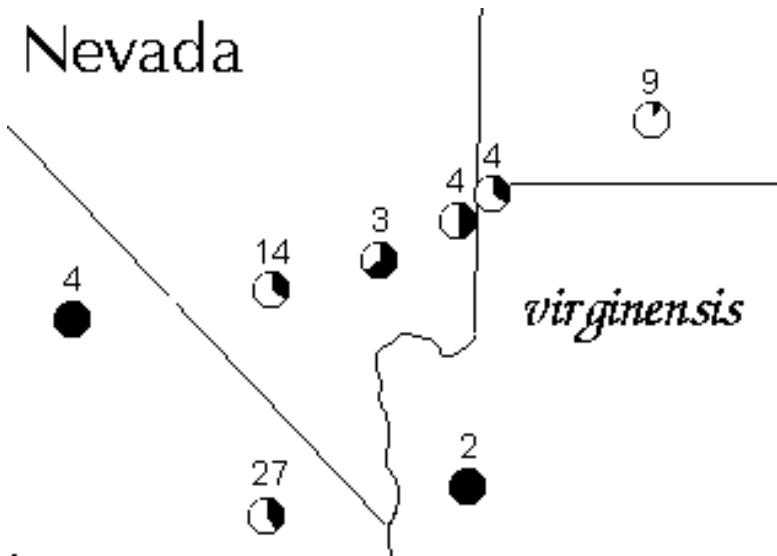
**Line graph:** Line graphs best represent data that are samples from continuous phenomena. The visual implication of the line is that intermediate points exist, but were not sampled. Values taken over time or through space fit this criterion, as do observations at different dosages (assuming that the dosage could be varied continuously). The order of the data along the X-axis is of course not arbitrary with a line graph. In this example, there are error bars for the individual samples. The samples are also connected by straight lines; they could also be connected by *spline curves*, which would give a smoother appearance, but which are no better predictors of intermediate values.



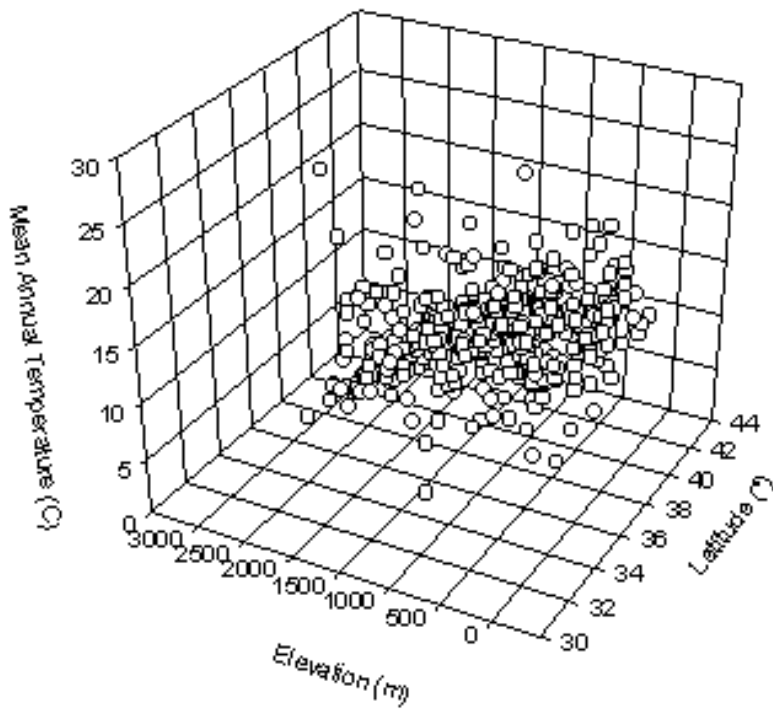
**Area graph:** An area graph is a stacked line graph, somewhat like a stacked bar graph, in that the values determine the height of a line above the line below it, rather than the origin. Area graphs are especially well-suited to show changes in percentages or cumulative values over time, but they have few other uses in biology.



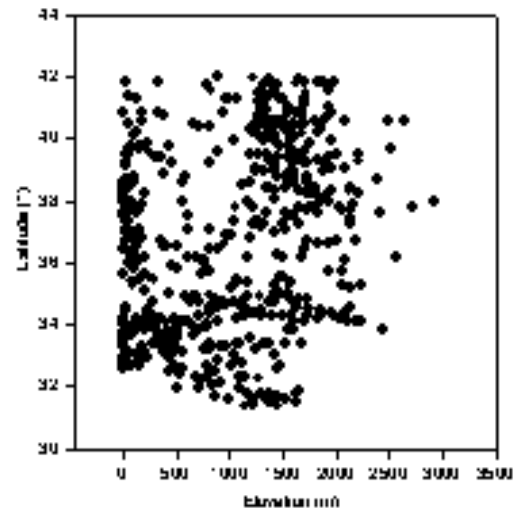
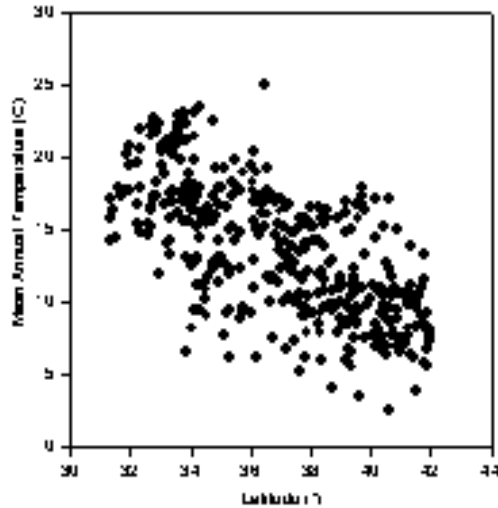
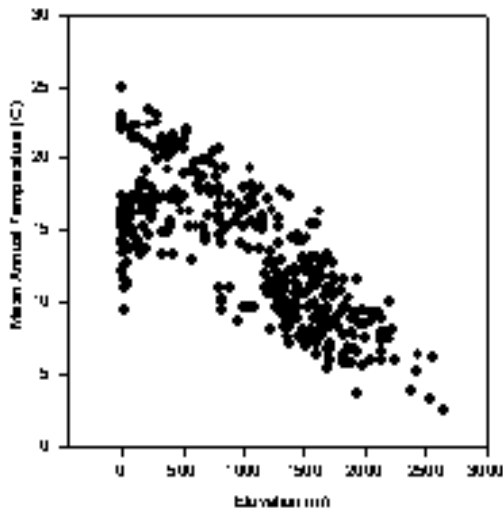
**Scatterplot:** In bar, line, and area graphs, the values for the independent variable are truly independent and are generally chosen in advance. A scatterplot instead shows the relationship between two measured variables as a scatter of individual points, each representing an item with its position determined along the X and Y axes by its values for the two variables. In terms of the way computer programs construct them, scatterplots and line graphs are very similar, but in their application and interpretation they are fundamentally different. The points of a scatterplot are never connected by lines, but instead a regression line is often plotted, showing how one measurement varies in relation to the other. (The example includes lines showing the 95% confidence intervals of the regression line and of the sample). If each point is a mean, error bars can be included in both the X and Y directions (forming a cross).



**Pie graph:** These are of limited usefulness in biology. They effectively show percentages, but a single pie graph can often be replaced by a table, and a group of them can be better compared as stacked bars. One appropriate use is to show percentages plotted on a map.



Three-dimensional graphs: These are of course two-dimensional projections of three-dimensional graphs. In this example, the top graph is a scatterplot of elevation, latitude, and mean annual temperature for the southwestern United States. The bottom three graphs are the three variables plotted pairwise in 2-D scatterplots. Notice that it is difficult to visualize the relationships among the variables in the 3-D plot (no, rotating won't help, I tried that). In the 2-D plots, the strong relationship between elevation and temperature, the weak one between latitude and temperature, and the (not unexpected) lack of relationship between elevation and latitude are all easy to see. Often, 2-D graphs are better at depicting relationships among multiple variables than 3-D graphs.



2-D graphs