Graphs

A graph is a visual representation of your data, and you want your graph to be as clear as possible to the reader for interpretation. First, **you have to decide** whether to use a scatter plot in order to draw a "best fit" line through data points, a bar graph, or some other representation with appropriate units. Use a line graph if your data are continuous (e.g., the appearance of product over time in an enzyme reaction). If your data are discontinuous (e.g., the amount of water consumption in different high schools), use a bar graph. Your teacher might have other suggestions.

A graph must have a title that informs the reader about the experiment. Labeling a graph as simply "Graph Number Four" doesn't tell the reader anything about the experiment, or the results. In comparison, the title "The Effect of Different Concentrations of Auxin on Root Growth" tells the reader exactly what was being measured. Make sure each line or bar on your graph is easily identifiable by the reader.

Axes must be clearly labeled with units:

• The x-axis shows the independent variable. Time is an example of an independent variable. Other possibilities for an independent variable might be light intensity, or the concentration of a hormone or nutrient.

• The y-axis denotes the dependent variable, or what is being affected by the condition (independent variable) shown on the x-axis.

• Intervals must be uniform. For example, if one square on the x-axis equals five minutes, each interval must be the same and not change to ten minutes or one minute. If there is a break in the graph, such as a time course over which little happens for an extended period, note this with a break in the axis and a corresponding break in the data line.

• For clarity, you do not have to label each interval. You can label every five or ten intervals, or whatever is appropriate.

• Label the x-axis and y-axis so that a reader can easily see the information.

More than one condition of an experiment may be shown on a graph using different lines. For example, you can compare the appearance of a product in an enzyme reaction at different temperatures on the same graph. In this case, each line must be clearly differentiated from the others — by a label, a different style, or color indicated by a key.

These techniques provide an easy way to compare the results of your experiments. Be clear as to whether your data start at the origin (0,0) or not. Do not extend your line to the origin if your data do not start there. In addition, do not extend your line beyond your last data point (extrapolation) unless you clearly indicate by a dashed line (or some other demarcation) that this is your prediction about what may happen.

Descriptive Statistics

You've Got Numbers. Now What?

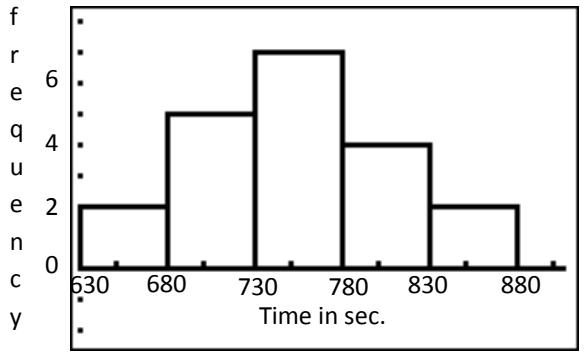
Floating disk lab. Data listed in seconds.

Control	809	785	697	766	737	736	821	856	862	724
	690	766	809	756	635	724	744	732	676	708
Experimental	834	756	735	877	779	831	829	888	929	853
	797	912	790	1004	805	872	1007	630	746	877

Graphical Displays – first you need a picture to see the distribution

Histograms – Excellent for large data sets

- Label both horizontal and vertical axes
  (vertical axis is usually the frequency but may be cumulative)
- Bin widths should be the same
- Comparative data should use the same scale



Stemplot – Works for small data sets. Too cumbersome for large data sets.

- Use only one digit for the "leaves"
- Stem may be more than one digit – rounding may be necessary
- Include a key

6	4
*	8 9 9
7	1 2 2 3 4 4 4
*	6 7 7 9
8	1 1 2
*	6 6

6 | 4 = 640 seconds

Data rounded to nearest 10 seconds

Back –to-back stemplot – used for comparison

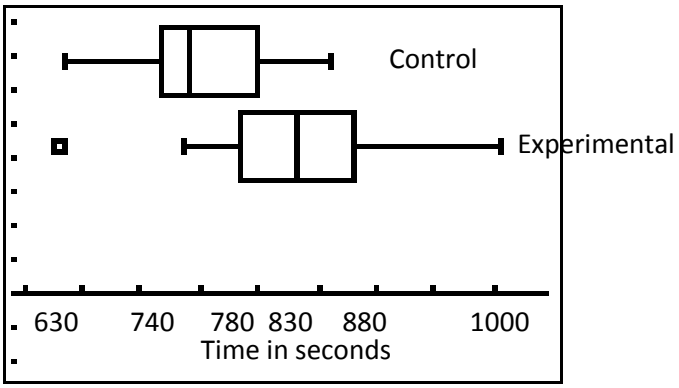
Experimental		Control
3	6	4
	*	8 9 9
4	7	1 2 2 3 4 4 4
9 8 6 5	*	6 7 7 9
3 3 3 1 0	8	1 1 2
9 8 8 7 5	*	6 6
3 1	9	
	*	
1 0	10	

6 | 4 = 640 seconds

All data rounded to nearest 10 seconds

Boxplots – appropriate for large data sets

- Label the data, especially if more than one boxplot used
- Include scale
- May show outliers
- Cannot be used to show normality – use the term symmetric



Summary Statistics – When describing a distribution, CUSS and BS.

- C – Center : mean or median
- U – Unusual features such as outliers, clusters, or gaps in the data
- S - Shape: approximately normal, skewed right or left, bimodal
- S – Spread: range, standard deviation, IQR(interquartile range)

BS = Be specific

Center

mean: simply the average of the data

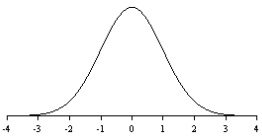
median: the center of the data when ordered from smallest to largest

The median is not appropriate when the data are skewed or large gaps occur

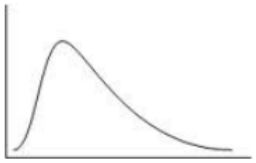
Unusual features

Outliers are typically more than 1.5(IQR) above quartile 3 or 1.5(IQR) below quartile 1

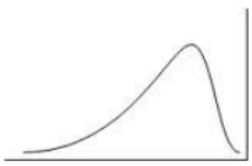
Shape



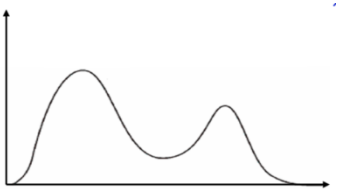
normal



skewed right



skewed left



bimodal

Spread

Range: simply the maximum minus the minimum

Standard deviation: basically the average distance from the mean

IQR: $Q_3 - Q_1$ This is the length of the box in a boxplot.

5-number summary

minimum, Q_1 , median , Q_3 , maximum