

BOX AND WHISKER PLOTS**What:**

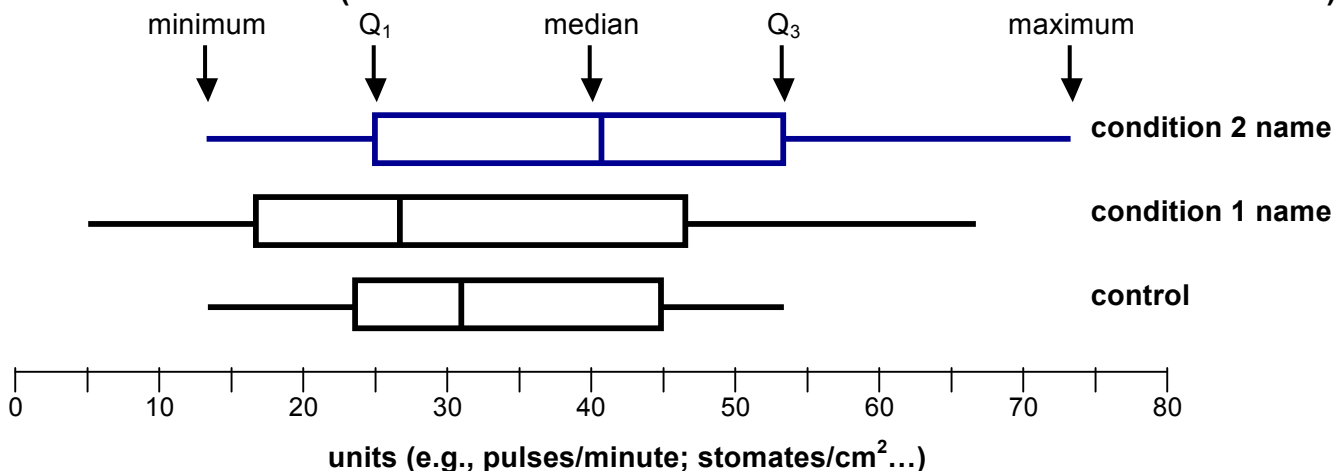
The box plot, also known as box-and-whiskers, is a relatively new data display created by John Tukey of Princeton (he also created the stem-and-leaf plot). A box plot is a scaled display in which the data are divided into four equal parts and are used to show the spread of the data relative to the median and the quartiles.

Why:

Looking at the mean and standard deviation is one method of measuring the center and spread of a data set. However, these statistics can be more difficult to calculate and interpret. Using the median to measure the center and the quartiles to show spread is a simpler, more easily visualized method. Also the median and quartiles are not as easily affected by extreme values, as are the mean and standard deviation.

How:

1. Rank the raw data in numerical order from least to greatest (you should have a minimum of 15 data points per condition for a valid box plot).
2. Divide the data into two equal halves at the **median** (the middle value). If the median is a data point, do not include it into either group. (If you have 15 data points the median would be the 8th value.)
3. Find the median of the lower half. This value is the first or lower quartile (Q_1). This is equivalent to the 25th percentile. (If you have 15 data points the first quartile would be the 4th value.)
4. Find the median of the higher half. This value is the third or upper quartile (Q_3). It is equivalent to the 75th percentile. (If you have 15 data points the median would be the 12th value.)
5. Use a number line scaled in appropriate equal intervals.
6. Draw short vertical lines at the values for the lower and upper quartile. Complete the "box" with horizontal lines joining the vertical lines. This represents 50% of the data. Mark the median as a vertical line inside the box. The "whiskers" are drawn as horizontal segments extended from the middle of the sides of the box to the minimum and maximum values.

DESCRIPTIVE TITLE (SHOULD INCLUDE KEY VARIABLES AND NAME OF EXPERIMENTAL ORGANISM)

Special uses:

Parallel box plots(as shown above) are a useful way of comparing groups of data. For example, compare the data from different lab groups or different trials of an experiment. Or compare the data collected from an experiment conducted at different times when a variable is changed, such as time of the day.

Interquartile range (IQR):

The interquartile range is the difference between the upper and lower quartile. ($Q_3 - Q_1$) This gives the spread of the middle 50% of the data. A rule of thumb is that 0.75 times the IQR is a good approximation of the standard deviation of the data set.

Inner quartiles:

Comparisons of the locations of the inner quartiles (in addition to comparisons of the medians) **are the most valuable statistical comparison for you to make**. If the inner quartiles (the “boxes”) overlap by a full quartile, this indicates that there is not a *significant* statistical difference in the data sets.

Outlier:

An outlier is an extreme value that is not typical of the data. These are values lying more than 1.5 times the interquartile range away from the nearer quartile. Outliers should be represented by open circles (or asterisks) and the whiskers should be drawn to the next closer value within the acceptable range.

Caution: It is important that students understand that a smaller side of the “box” or a shorter “whisker” does not mean that there are fewer values represented by that the values are grouped closer together, in other words, they have a smaller spread.

Communication of summary values: ⇒You should include the following table of values with your box plot.

DESCRIPTIVE TITLE TELLING WHAT YOU ARE COMPARING:

	Control	Condition 1	Condition 2
1st Quartile			
Median			
3rd Quartile			
Max / Min Values			
Inner Quartile Range ($Q_3 - Q_1$)			
Whisker Limit** = 1.5 (IQR)			
Max. Limit** = $Q_1 - \text{Whisker limit}$			
Min. Limit** = $Q_3 + \text{Whisker limit}$			
Outliers** (Values above and below limits)			

The starred (**) items in the table are for more advanced analysis of the data.