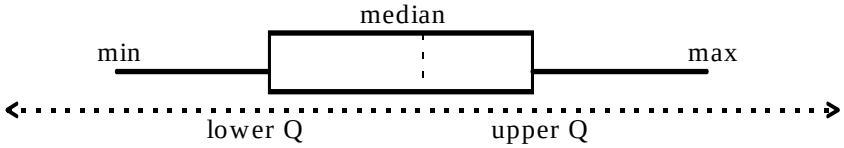
Objectives:

- To understand the meaning of the phrase ***statistical tendency***
- To learn a formal rule-of-thumb for determining whether an observation is an ***outlier***
- To discover ***modified boxplots*** as displays conveying more information than ordinary boxplots
- To acquire extensive experience with using graphical, numerical, and verbal means of comparing and contrasting distributions from two or more groups
- To use the computer as an important tool for performing such analyses

Statistical tendency Though each individual observation may not agree, you look for a tendency - or in other words for most observations what is happening. This is the statistical tendency. *e.g.* Men tend to be taller than women, This is not the case for every man and every women but there is a tendency.

Five-number summary is made up of the minimum and maximum observations along with the lower, middle (median) and upper quartiles.

boxplot uses the *five-number summary* to create a visual display of a distribution. To create a *boxplot* first draw a number line with a proper scale to illustrate the observations being displayed. Draw a *box* with one end at the lower quartile and the other end at the upper quartile. draw a dotted line in the box at the median. Draw "whiskers" from each end of the box to the minimum and the maximum.



Modified boxplots for normal data with no outliers the standard boxplot is useful, but whenever an outlier is present that outlier takes the whisker well beyond where the non-"screwed-up" data is. *i.e.* if we were to do a boxplot of income, The whisker would have to go all the way out to Ross Perots income - this does not help us distinguish the typical upper income.

Outliers: To overcome this problem we find the value of 1.5**IQR* let the whisker go out to this value if the data goes out this far. Any data beyond this value will be denoted by a "*" in the boxplot.