

For the following section, the ages of 13 students in a class are used:

15, 14, 14, 15, 15, 15, 15, 16, 15, 14, 15, 15, 15

Mean

The *mean* is the *average* of all the data. There is a special symbol for the mean of a set of numbers – it’s an ‘x’ with a horizontal line above it: $\bar{x}$.

$$\bar{x} = \frac{\sum x}{n}$$
 where $\sum x$ = sum of all numbers

For the class:

$$\bar{x} = \frac{\sum x}{n}$$
$$\bar{x} = \frac{15+14+14+15+15+15+15+16+15+14+15+15+15}{13}$$
$$\bar{x} = 14.85$$

Median

The median value is the *middle* value in the list once you’ve ordered them:

14, 14, 14, 15, 15, 15, **15**, 15, 15, 15, 15, 15, 16

15 is the median value. If you have an even number of values, you take the average of the middle two numbers. So if there were only 12 ages:

14, 14, 14, 15, 15, **15**, **15**, 15, 15, 15, 15, 16

The median is the average of the two middle values 15 and 15. So the median is 15.

Mode

The mode is the most commonly occurring number in the entire list. Can be more than one mode.

Even More Statistics

Create a frequency table for the original 13 ages:

Student age	Number of students with that age
14	3
15	9
16	1

The mode is 15.

Mean Versus Median

The mean generally seems to take into account *all* the values in a data set, which means that one *outlier* (an outlier is a value that is significantly different to most of the other values) can significantly change the mean.

The median is much better at effectively *ignoring* outliers.

Because of the median’s ability to ignore outlying values, it is often regarded as a more *robust* measure, because it is focused around the middle values and ignores extreme values on either side.

Median and mean are used a lot in real estate.

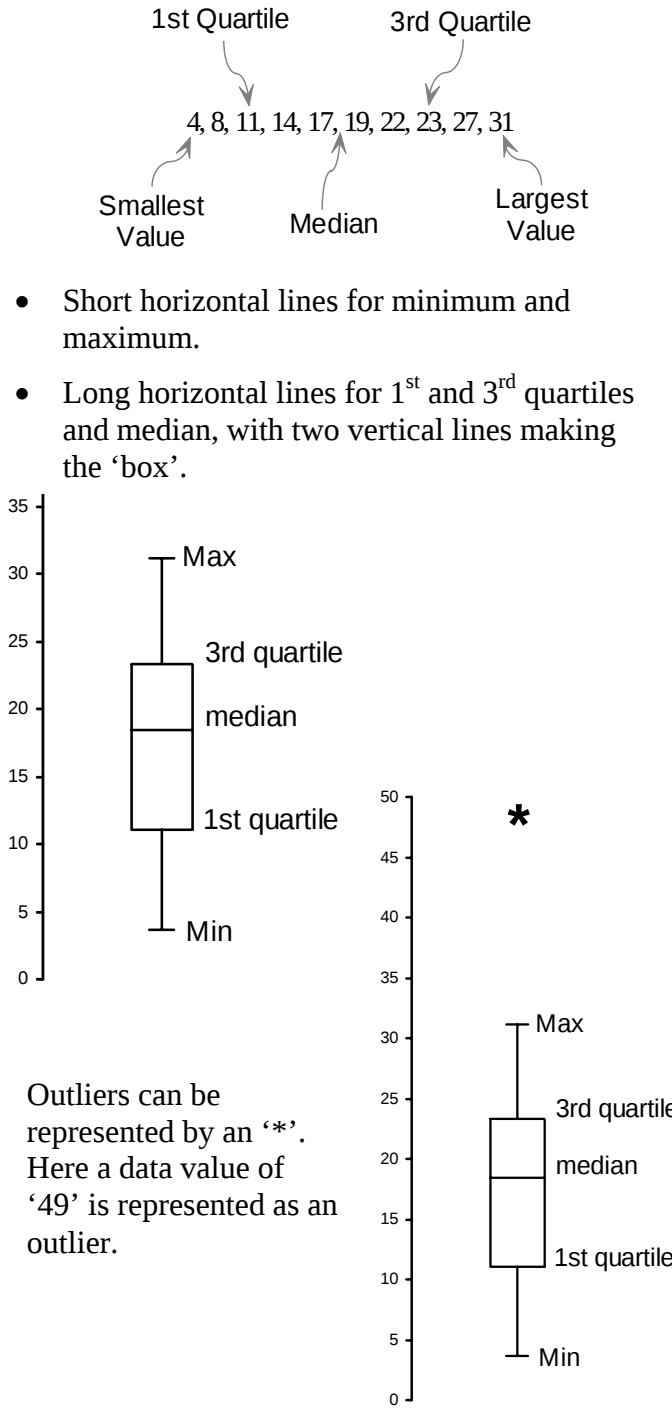
Box and Whisker Plots

Start with data: 11, 23, 14, 27, 8, 4, 31, 22, 17, 19

Order it: 4, 8, 11, 14, 17, 19, 22, 23, 27, 31

The **median** is the middle value - the average of 17 and 19 – 18.

The **1st quartile** value is the median value *out of the lower half of the data*, the **3rd quartile** value is the median value *out of the upper half of the data*:



- Short horizontal lines for minimum and maximum.
- Long horizontal lines for 1st and 3rd quartiles and median, with two vertical lines making the ‘box’.

Outliers can be represented by an ‘*’. Here a data value of ‘49’ is represented as an outlier.