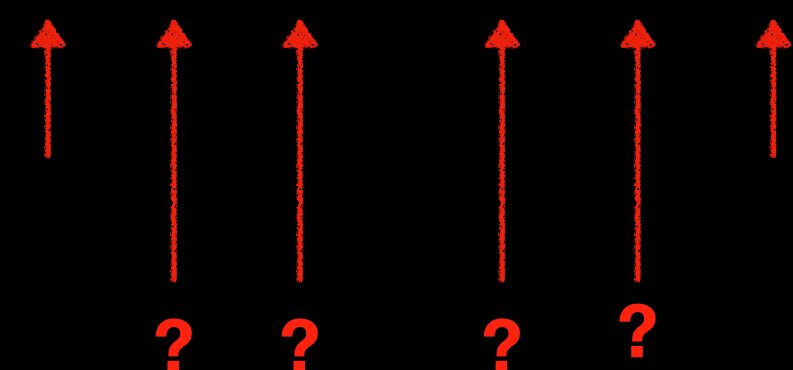


```
summary(after)
  Min. 1st Qu. Median  Mean 3rd Qu.  Max.
 39.00 41.25 44.50  44.21 46.75 50.00
```



A diagram consisting of six red arrows pointing upwards. The first arrow points from a red question mark to the 'Min.' value (39.00). The second arrow points from a red question mark to the '1st Qu.' value (41.25). The third arrow points from a red question mark to the 'Median' value (44.50). The fourth arrow points from a red question mark to the 'Mean' value (44.21). The fifth arrow points from a red question mark to the '3rd Qu.' value (46.75). The sixth arrow points from a red question mark to the 'Max.' value (50.00).

Some of these may look familiar but some might be new -- the mean, median and quartiles might be new.

Summary Statistic Definitions!

Mean (Sample) = sum of all data values divided by number of data points

$$\text{Mean} = \frac{\text{Sum of all values}}{\text{Number of values}}$$

$$\text{Symbolically, } \bar{x} = \frac{\sum x}{n}$$

where \bar{x} (read as 'x bar') is the mean of the set of x values,
 $\sum x$ is the sum of all the x values, and
 n is the number of x values.

(note - only works with “numerical” data types... more about data types later)

Median = if we order the data from smallest to largest, this is the observation in the middle (splits the data in 2 halves)

First/Third Quartiles = where 25% of the data falls below/above

Standard Deviation = this is the square root of the variance, where the variance is roughly the average distance of data values from the mean

$$\text{Standard Deviation (sample)} = \sqrt{\frac{\sum_{i=1}^n (x_i - \text{mean})^2}{n-1}}$$

Mean and median give a sense of the “typical value” in your dataset.

Summary Statistic Definitions!

Mean (Sample) = sum of all data values divided by number of data points

$$\text{Mean} = \frac{\text{Sum of all values}}{\text{Number of values}}$$

Symbolically, $\bar{x} = \frac{\sum x}{n}$

where \bar{x} (read as 'x bar') is the mean of the set of x values,
 $\sum x$ is the sum of all the x values, and
 n is the number of x values.

(note - only works with “numerical” data types... more about data types later)

Median = if we order the data from smallest to largest, this is the observation in the middle (splits the data in 2 halves)

First/Third Quartiles = where 25% of the data falls below/above

Standard Deviation = this is the square root of the variance, where the variance is roughly the average distance of data values from the mean

$$\text{Standard Deviation (sample)} = \sqrt{\frac{\sum_{i=1}^n (x_i - \text{mean})^2}{n-1}}$$

Standard deviation and quartiles give measures of the variability in the data.