A **response variable** measures or records an outcome of a study.

An **explanatory variable** explains changes in the response variable.

In a **scatterplot** one axis is used to represent each of the variables, and the data are plotted as points on the graph.

Typically, the **explanatory** or **independent variable** is plotted on the *x* axis and the **response** or **dependent variable** is plotted on the *y* axis.

Positive association: High values of one variable tend to occur together with high values of the other variable.

Negative association: High values of one variable tend to occur together with low values of the other variable.

The **strength** of the relationship between the two variables can be seen by how much variation, or **scatter**, there is around the main form.

With a strong relationship, you can get a pretty good

estimate of y if you know x.

With a weak relationship, for any *x* you might get a wide range of *y* values.

In a scatterplot, outliers are points that fall outside of the overall pattern of the relationship.

The correlation coefficient is a measure of the direction and strength of a relationship. It is calculated using the mean and the standard deviation of both the *x* and *y* variables.

$$\mathbf{r} = \frac{1}{\mathbf{n}-1} \sum_{i=1}^{n} \left(\frac{x_i - \overline{x}}{s_x} \right) \left(\frac{y_i - \overline{y}}{s_y} \right)$$