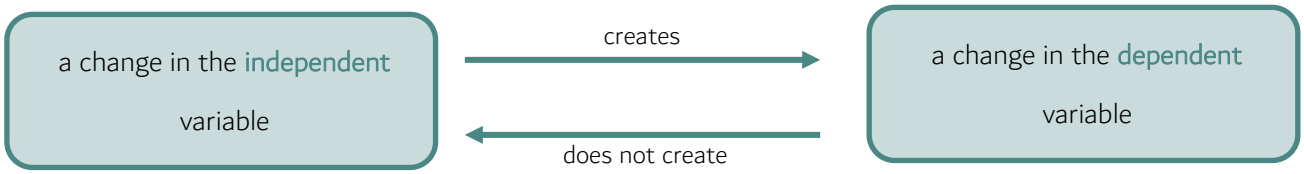
Understanding correlation

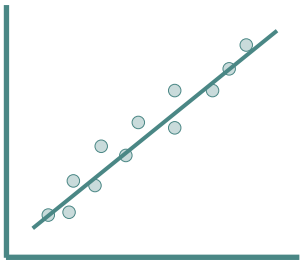
A correlation is a statistical relationship between two variables. One of these variables is referred to as the independent variable (plotted against the x axis) and the other as the dependent variable (plotted against the y axis).



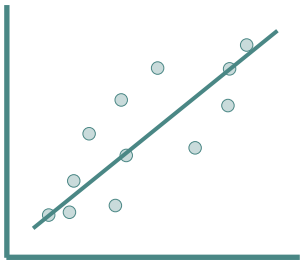
In most geographical investigations the correlations found are linear. This means that as one variable increases or decreases, the second variable will increase or decrease, and a line is observed when the data is plotted in a scattergraph.

Where an increase in one variable results in an increase in the second variable, we refer to the correlation as positive. Where an increase in one variable results in a decrease in the second variable, we refer to the correlation as negative.

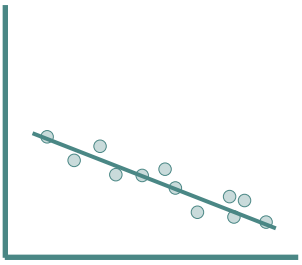
In a scattergraph a line of best fit can be drawn where it passes through as many of the data plots as possible. It is extremely unlikely that all the data plots will fall on the line exactly - in this case an equal number of plots should sit either side of the line. The closer the data plots are to the line of best fit, the stronger the correlation between the variables. The angle of the line or the position of the line within the set of axes is no measure of the strength of the correlation.



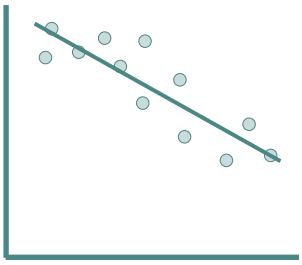
Strong positive correlation



Weak positive correlation

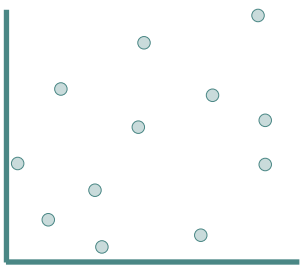


Strong negative correlation

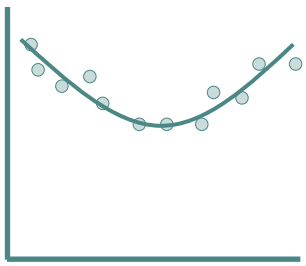


Weak negative correlation

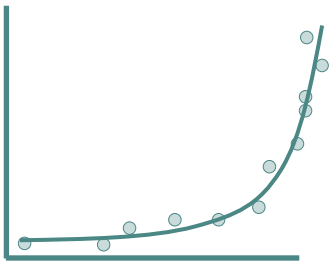
Where it is not possible to draw a line of best fit, we refer to the variables as having no correlation. However, it may also be the case that the data is displaying a non-linear correlation between the variables such as a curvilinear correlation (where the data increases and then subsequently decreases with certain values) or an exponential correlation (where the change in one variable does not occur at a uniform rate with the change in the second variable).



No correlation



Curvilinear correlation



Exponential correlation

Understanding any correlation between two variables helps geographers to better understand the impact geographical changes might have.

Correlations also allow geographers to estimate, to within a good degree of accuracy, the value of unknown data plots based on the line of best fit. For example, one could extrapolate data beyond the range of the known sample to say what data might be found when one variable is doubled or tripled. Equally one could interpolate data. This means one could estimate the value of data plots found in between known data plots within the sample.

