An anomaly is a piece of data, collected by researchers that does not fit with the general pattern of results shown in the sample. In geography this definition is taken further to include the idea that anomalous data is that which can not be explained by any known geographical phenomenon. This means that data might be an outlier (does not fit the normal pattern of results) but would not be considered anomalous because other understandable and known geographical variables may have caused that particular piece of data to stand out.
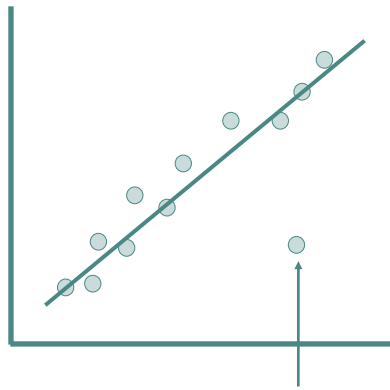
This point is sitting far
outside the line of best fit.
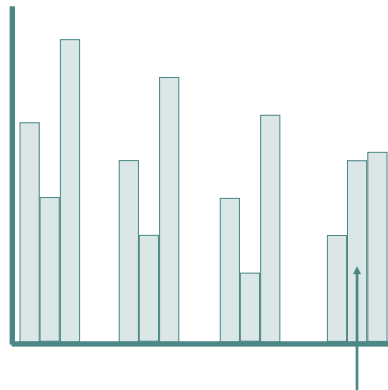
This column appears to be far
taller than one might expect.

Outliers are often identified when the data is presented in a graphical form. For example, plotting data on a scattergraph or on a map for example will clearly show any single data value which is not like the others or which lies further away from something else than one might have predicted.

Explaining outliers

Once an outlier has been identified, there are a number of questions one might ask as a way of trying to see whether the data is a true outlier or is in fact anomalous. It is important for geographical researchers to think through these questions carefully. Too often outliers are dismissed from further investigation because it is quickly assumed that human error has caused the data to be different. In fact, the outlier might prove to be the most interesting piece of data within the set as it may include the influence of variables that the researcher had not initially considered.

- Has the data been accurately transcribed into the graphical data presentation technique? Could the data recording sheet have been read in error when drawing the graph or map?

- Was the data recording equipment used accurately? Were multiple readings taken? (followed by a calculated mean?)

- Was all the data collected and recorded in the same manner? Was the same equipment used in the same way, by the same person in each occasion?

- Was the data all collected at the same time? And on the same day? Could weather conditions or seasons have had an impact on the data and caused an outlier to be recorded?

- Was a second reading taken when it was found that there was an outlier? Did this turn up the same result the second time round?

- What other geographical variables (that weren't measured and recorded) could have influenced the data? Why might these variables have only influenced this data point and not the others?

Answering these questions should allow the researcher sufficient means to find a possible reason for any outlying data. Only once all of the above have been considered might the geographer call the data truly anomalous.

The next decision is whether to include any anomalous quantitative data in any statistical test that the research demands. Much of that decision rests on just how different the anomalous value is from the overall pattern - the smaller the difference, the less likely it will be to greatly influence the overall result of any statistical test. It is equally important to consider whether it will be ethically and morally right to exclude some data from the overall sample.