Scatter graphs

A scatter graph shows the relationship between two variables or 'factors'.

Scatter graphs are used to give a visual impression of the direction or strength of a relationship between two variables. They are one of the most commonly used types of graphs because geographical enquiries are often concerned with the relationship between paired variables. For example, air temperature and altitude, slope angle and soil depth, natural population increases and life expectancy, and building age with distance from the CBD.

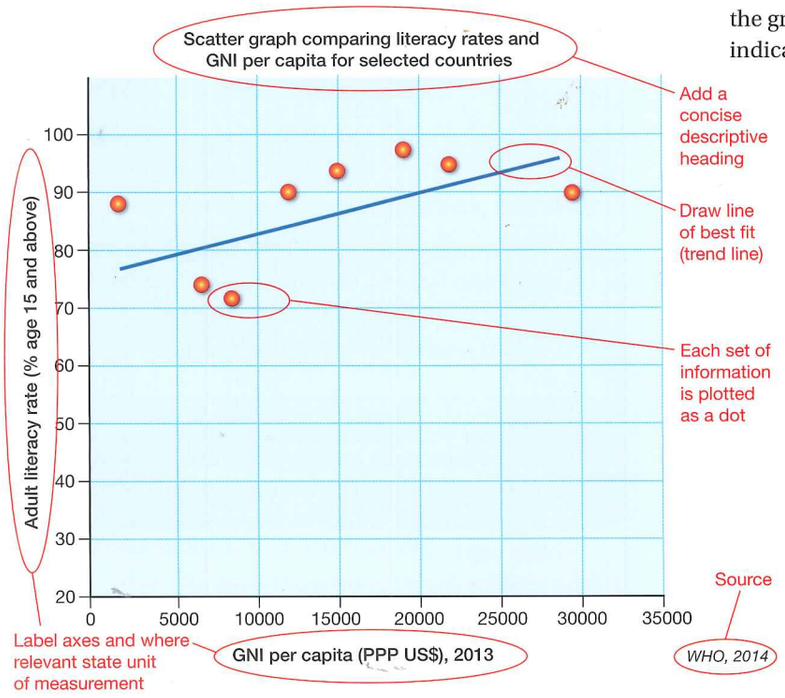
Table 9.34 Adult literacy rates and GNI per capita for selected countries, 2013

Country	Adult literacy rate (% at age 15 and above), 2013	GNI per capita (PPP US$), 2013
Portugal	95	22900
Chile	98	19100
United Arab Emirates	90	29900
Mexico	94	15000
Brazil	90	12100
Egypt	74	6600
Algeria	72	7500
Kenya	87	1800

Source: World Health Organization, 2014

The scatter graph in Figure 9.35 shows the relationship between the two variables: adult literacy rate and **GNI** per capita (**PPP** US $).

Figure 9.35 Scatter graph of data shown in Table 9.34



Drawing a scatter graph

1. Draw a horizontal axis and a vertical axis on a piece of graph paper. Allow at least half an A4 page to complete your graph. Use a pencil in case you make a mistake.
2. Label the axes by deciding which variable is causing the change in the other. The simplest way to decide is to look for the 'cause-and-effect' relationship. The horizontal axis shows the cause (independent variable) and the vertical axis shows the effect (the dependent variable). You decide this, based on what you wish to show. For example, in Figure 9.35:
   - Does GNI per capita affect adult literacy rate—i.e. are you trying to show that wealthier countries can afford more schooling, which therefore increases the literacy rate?

   or
   - Does adult literacy rate affect GNI per capita—i.e. are you trying to show that more literate countries tend to have better skills for employment, which therefore increases GNI?

   In this case, we want to explore how far GNI per capita influences (causes) the adult literacy rate (the effect).
3. Develop a scale for each of the two axes. Do this by looking at the range of data that has to be graphed.
4. Plot the data by placing a single point at the meeting point (coordinate) of the two values. The more points (paired data) there are, the more reliable the graph. Do not join up these points.
5. Give the graph a title.
6. Draw in a line of best fit (see Figure 9.35) by using a ruler and pencil. The straight line should be as close as possible to all the points plotted with a balanced scatter of points above and below the line. The line does not need to pass through the point of origin (where the two axes meet).
7. Determine the correlation (relationship) between the two variables by examining the pattern of points plotted on the graph. The closeness of the points to the line of best fit indicates the strength of the correlation or 'relationship'.

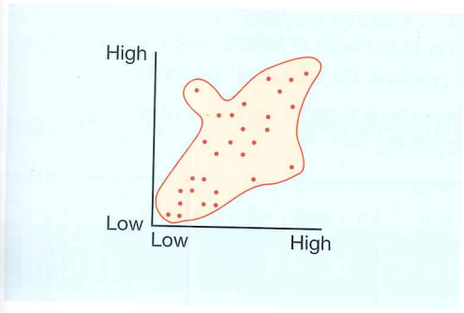
Glossary

GNI gross national income (the average personal income in a country), which has recently replaced GDP as a measure of economic development and wealth. GNI is more accurate than GDP because it takes PPP into account and therefore assesses what income is really worth.

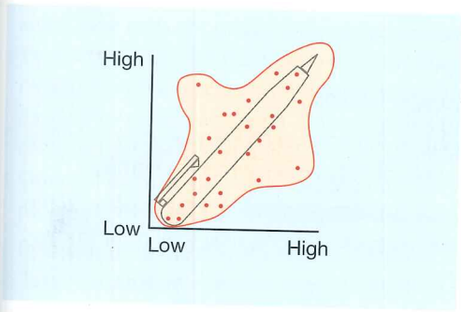
PPP purchasing power parity, which is a way of expressing a country's GNI that takes into account the local price of a number of standard goods and services. For instance, Uganda's GDP per capita is about A$400, but because goods and services there are three to four times cheaper than in Australia, its GNI is recalculated as A$1450 per capita.

Identifying the line of best fit on a scatter graph

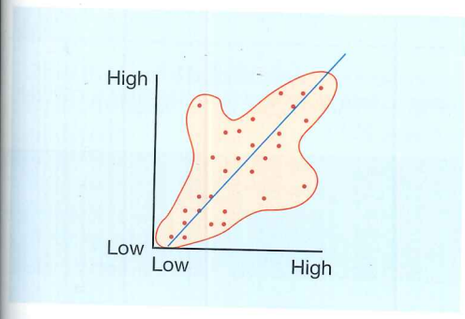
1 Draw a pencil line around the plotted points.



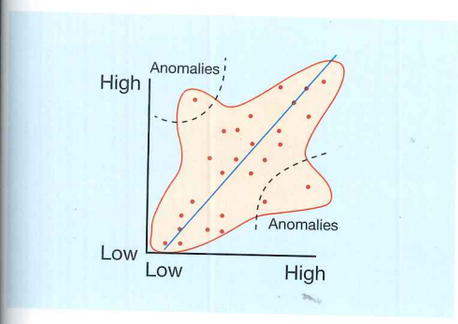
2 Use a pen or pencil to help visualise the line of best fit. Place it in the position that best reflects the direction of the shape you have drawn (the longest diagonal).



3 Remove the pen or pencil and draw a line in its place. Try to ensure that approximately half the points lie above and below the line, and that most are relatively close.



Those (few) points furthest away from the line are identifiable as the anomalies to the pattern identified by the line.



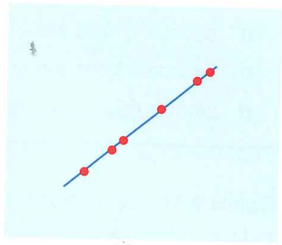
Correlation patterns

The scatter graphs in Figures 9.36 to 9.39 and their accompanying explanations indicate what some correlation patterns could mean.

Perfect correlation

- All points lie on the line of best fit.
- The correlation may be positive (as illustrated in the example) or negative.

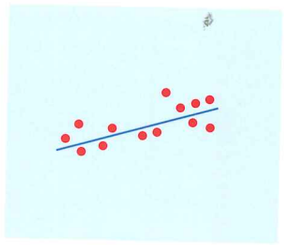
Figure 9.36



Positive correlation

- The points plotted on the graph run from the bottom left to the top right, and are close to the line of best fit.
- As one of the variables increases in value, so does the other.

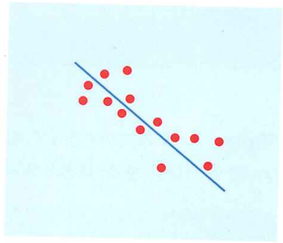
Figure 9.37



Negative correlation

- The points plotted on the graph run from the top left to the bottom right.
- As one variable increases in value, the other decreases.

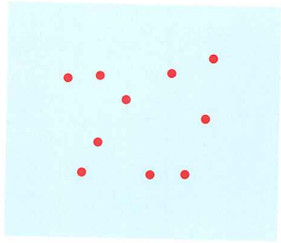
Figure 9.38



No correlation

- The points are scattered at random with no clear trend.
- There is no correlation (relationship) between the variables.

Figure 9.39



The existence of a strong positive correlation does not prove that there is an interaction. In other words, it is not always the case that there is a cause-and-effect relationship between two variables even if a positive correlation exists.

Individual points that stand out because they do not follow the pattern are called isolates or anomalies.

GeoSkills **9.19**

On which axis (horizontal or vertical) would each of the following sets of variables be placed? Refer to page 120 to help you.

a air temperature and altitude

b access to safe water and infant mortality rate

c sediment load and stream velocity

d plant nursery sales and number of months of drought

Figure 9.40 Scatter graph comparing maximum daily temperature and sales of refrigerated drinks

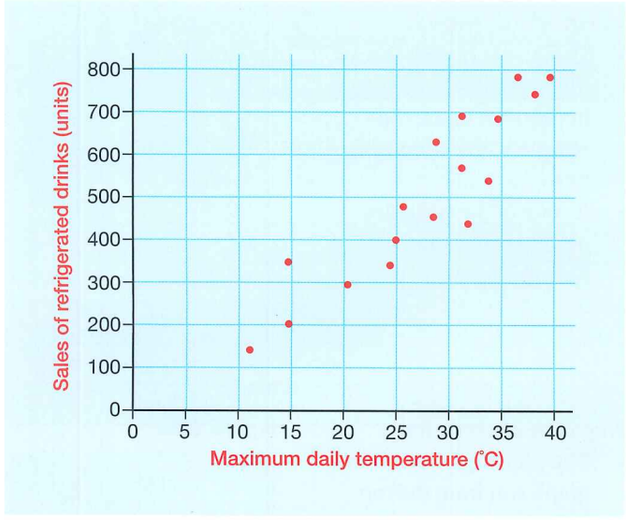
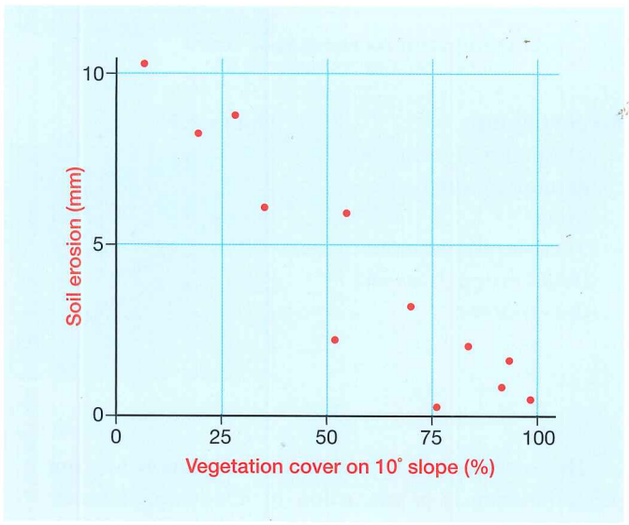


Figure 9.41 Scatter graph comparing percentage vegetation cover on 10-degree slope and depth of soil erosion



GeoSkills **9.20**

Examine the scatter graphs in Figures 9.40 and 9.41. For each graph, respond to the following questions.

1. What type of correlation (relationship) is shown—perfect, positive, negative, no correlation?
2. Which variable is independent and which is dependent?

GeoInquiry **9.21**

Refer to Table 9.42, which shows the percentage of the population of different countries with access to safe water and the mortality rate for children under five years of age.

1. Construct a scatter graph using the data from the table.
2. Draw a line of best fit.
3. Comment on the correlation (relationship) between the two sets of data as shown on the graph.
4. Identify any isolates (anomalies) that appear on the graph and suggest reasons why they differ from the trend.
5. Do you believe there is a relationship between access to safe water and child mortality? Justify your response.
6. Discuss: what other factors might contribute to child mortality?

Table 9.42 Access to safe water and mortality rate for children under five years of age, 2013

Country	Access to safe water (%)	Mortality rate for children under five years of age (per 1000)
Bangladesh	80	33
Bolivia	85	39
Brazil	91	20
Chad	34	96
Egypt	98	29
India	88	44
Indonesia	80	32
Mali	56	58
New Zealand	99	4
Philippines	91	23
Uganda	67	57

Source: World Health Organization, 2013