

Science 9–10

Chapter 4 — Processing, modelling and analysing

This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.

Attribution: Classbasics Pty Ltd, vwx.au

Aboriginal and Torres Strait Islander content has been excluded as accuracy cannot be guaranteed, and it is better to be respectful than cover the entire curriculum inaccurately.

Significant effort has been made to ensure this content is legal. Please send any areas of concern to admin@vwx.au with appropriate document/legal references so errors can be verified and changes be made.

Contents

Section	Page
Organising and summarising data: tables, graph choice and descriptive statistics	2
Models, symbols, formulas and mathematical relationships as representations	17
Patterns, trends, relationships and anomalies — including correlation versus causation	33

Organising and summarising data: tables, graph choice and descriptive statistics

Chapter 4 — Processing, modelling and analysing (Science, Levels 9 and 10)

Subtopic 4A · VC2 Science — Processing, modelling and analysing

CD code: VC2S10I04 (split delivery †; this subtopic is the accountable owner) **Content description**

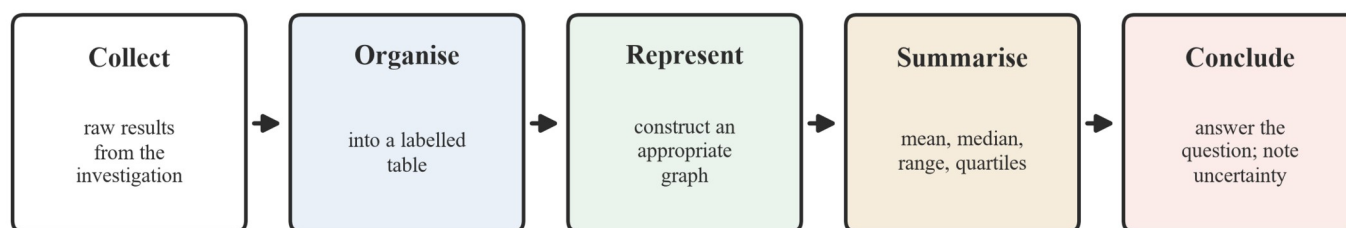
(verbatim): “data and information can be organised, processed and summarised by selecting and constructing representations including tables, graphs, descriptive statistics, models, symbols, formulas and mathematical relationships” **This subtopic’s limb (TOC §4 clause split):** tables, graph choice, descriptive statistics. Partner 4B (not an owner) delivers models, symbols, formulas and mathematical relationships.

Elaboration ids drawn on: VC2S10I04.E01 · VC2S10I04.E02 · VC2S10I04.E04 **Achievement standard (extract served):** “They select and construct a range of appropriate representations to organise, process and summarise data and information.” **Maths interlock (D11):** VC2M10ST01 — quartiles, interquartile range, box plots, histograms and dot plots; used here for summarising spread.

All data in this subtopic are simulated data for practice, and each dataset is labelled where it appears. No real-world measurement is attributed to any person or place.

Theory

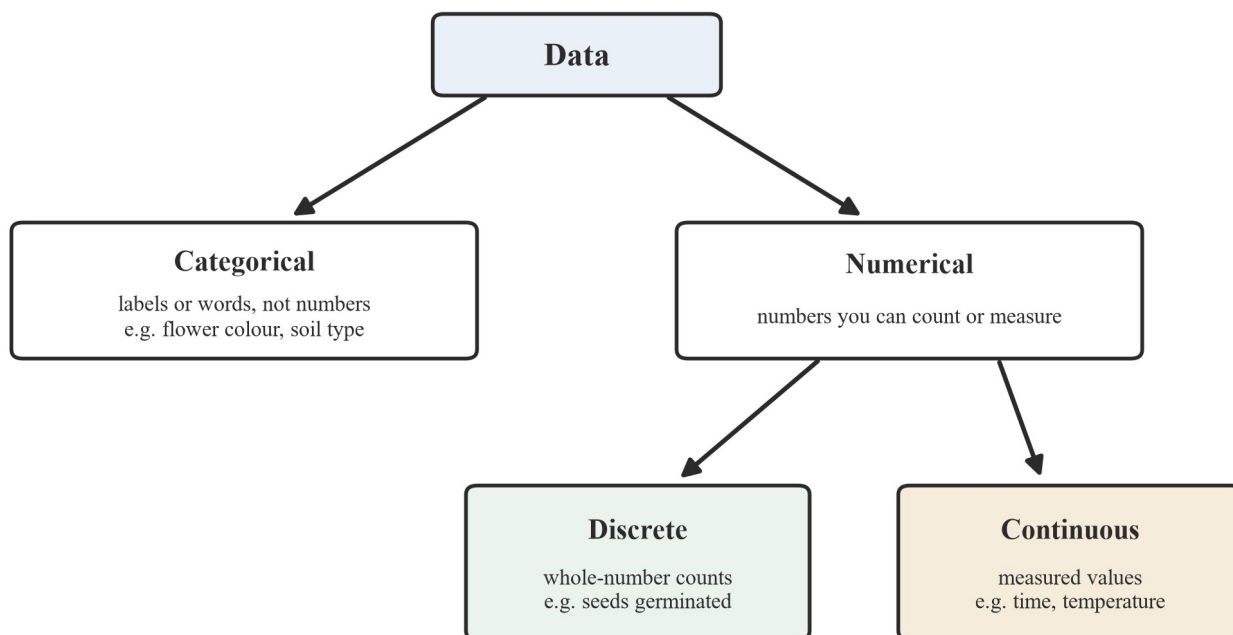
The data-processing pipeline. Every investigation produces raw results — a page of numbers that, on its own, tells you very little. To turn results into evidence, scientists move through the same five steps every time.



Flow diagram of the five steps: collect the raw results, organise them into a labelled table, represent them with an appropriate graph, summarise them with statistics such as the mean, median, range and quartiles, and conclude by answering the question with a note of uncertainty.

Each step makes the data easier to read. The table stops you losing values, the graph shows the shape of the data at a glance, and the summary numbers let you compare two investigations in a single line. This subtopic works through the middle three steps: organise, represent, summarise.

Types of data. The graph you may choose depends on the type of data you collected. Data is either **categorical** (labels or words, not numbers) or **numerical** (numbers you count or measure), and numerical data is either **discrete** (whole-number counts) or **continuous** (measured values that can sit anywhere on a scale).



Classification tree: data splits into categorical (labels or words, not numbers, for example flower colour and soil type) and numerical (numbers you can count or measure); numerical splits into discrete (whole-number counts, for example seeds germinated) and continuous (measured values, for example time and temperature).

Organising data into tables. A table is the first home for raw results. A well-built table puts the **independent variable** (the one you change) in the first column and the **dependent variable** (the one you measure) in the columns that follow; writes each unit once, in the heading, rather than in every cell; keeps one column per repeated trial; and ends with a mean column.

Colour of light	Trial 1 (bubbles/min)	Trial 2 (bubbles/min)	Trial 3 (bubbles/min)	Mean (bubbles/min)
Blue	34	38	36	36
Green	12	10	11	11
Red	41	43	39	41
White	29	31	30	30

Annotations for Table 1:

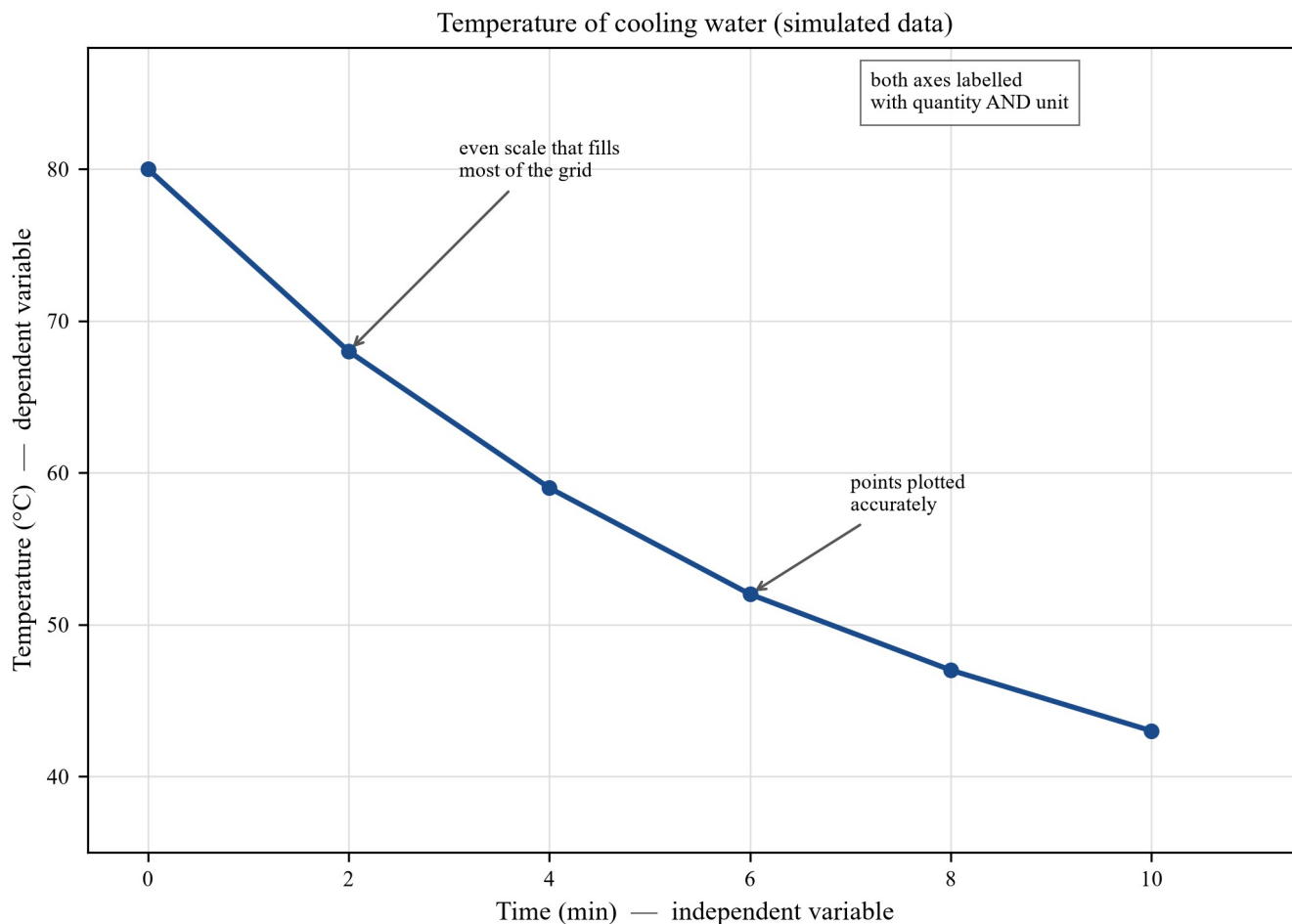
- independent variable (first column):** Points to the 'Colour of light' column.
- unit written once, in the heading:** Points to the '(bubbles/min)' unit in the Trial 1 heading.
- repeated trials:** Points to the Trial 1, Trial 2, and Trial 3 columns.
- mean column:** Points to the 'Mean (bubbles/min)' column.
- dependent variable (the measured values):** Points to the numerical values in the data rows.

Table 1. Oxygen bubbles produced by pondweed under different colours of light (simulated data).

Model results table, Table 1: oxygen bubbles from pondweed under blue, green, red and white light, with three trial columns and a mean column; callouts mark the independent variable in the first column, the unit written once in the heading, the repeated trials, the mean column, and the dependent variable as the measured values.

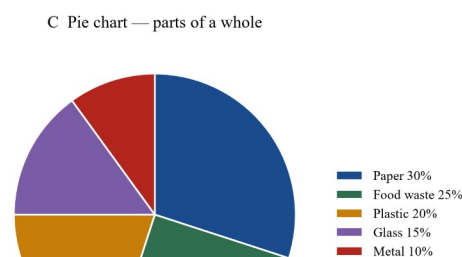
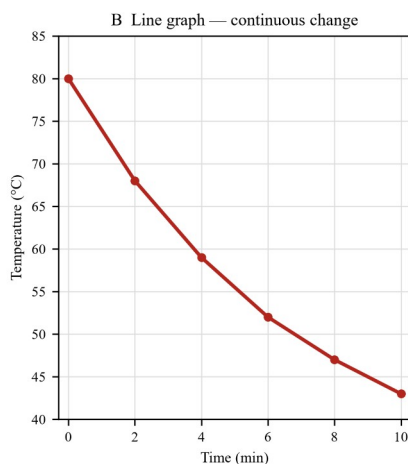
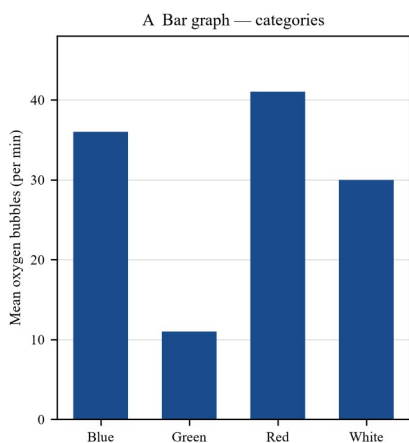
In Table 1 each colour is tested three times. Repeats matter: the blue trials give 34, 38 and 36, and the mean of 36 smooths out the small differences between trials. If one trial lands far from the others, you see it at once and can check what happened, instead of discovering the problem after the data has been summarised.

Setting up axes. When you construct a graph, the independent variable goes on the horizontal (x) axis and the dependent variable on the vertical (y) axis. Every axis is labelled with the quantity **and** its unit, the scale is even and fills most of the grid, and the points are plotted accurately.



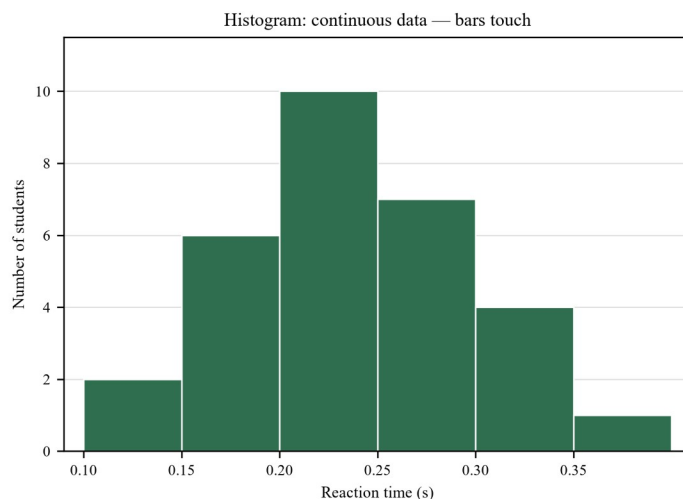
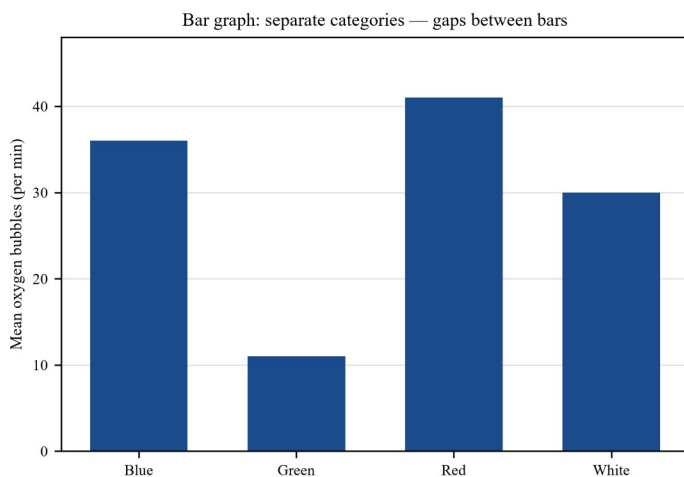
Line graph of the temperature of cooling water against time, with notes that both axes are labelled with quantity and unit, the even scale fills most of the grid, and the points are plotted accurately.

Three common graph types. Three graphs cover most school investigations. **Bar graphs** compare categories. **Line graphs** show how the dependent variable changes as a continuous variable — usually time — changes. **Pie charts** show how one whole splits into parts.



Three panels. A: bar graph of mean oxygen bubbles per minute for blue, green, red and white light. B: line graph of the temperature of cooling water against time. C: pie chart of household waste, paper 30%, food waste 25%, plastic 20%, glass 15%, metal 10%.

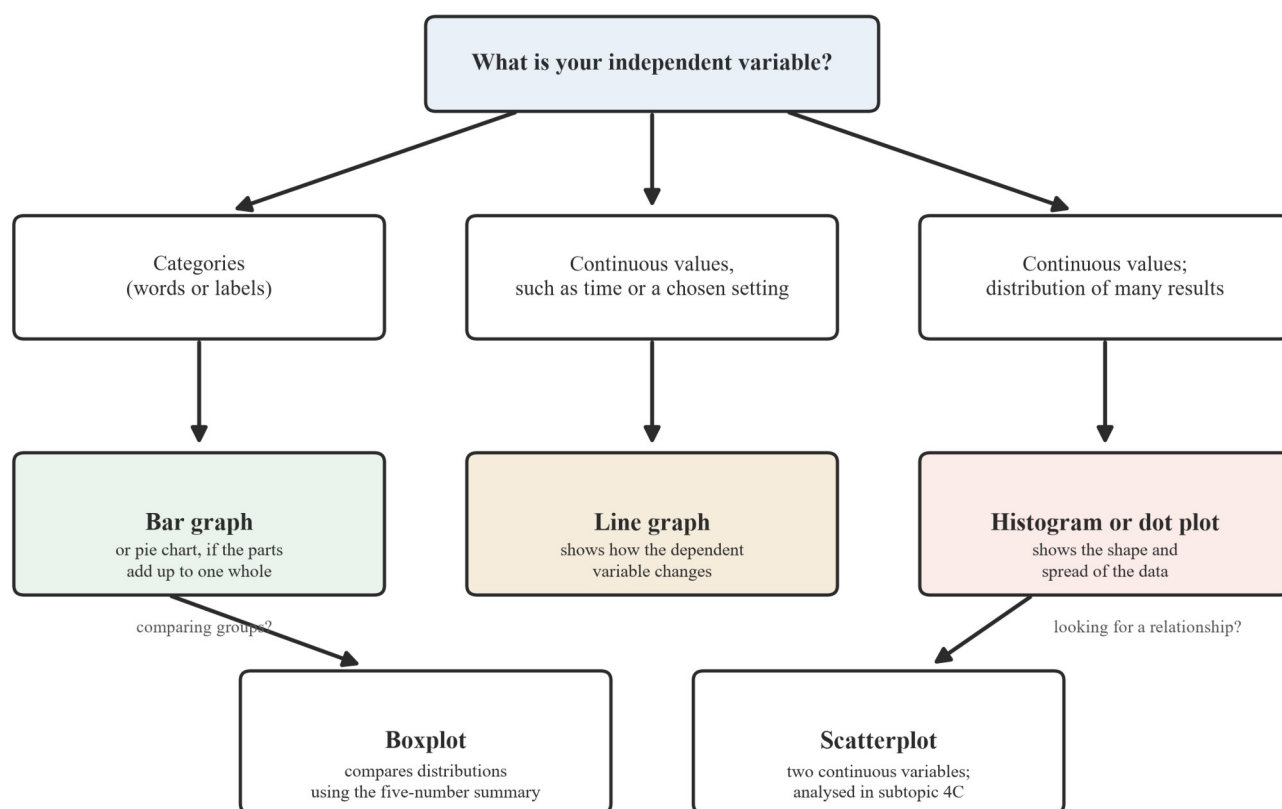
Bar graphs and histograms. A **histogram** looks like a bar graph but is a different tool. Its data is continuous and grouped into equal-width bins, so the bars touch; a bar graph compares separate categories, so the bars keep a gap.



Two panels side by side: a bar graph of separate categories with gaps between the bars, and a histogram of the reaction times of 30 students with touching bars.

In the histogram, each bar counts the students whose reaction time falls in one bin, such as 0.20–0.24 s. The bins are neighbouring slices of one continuous scale with no gap between them, which is why the bars touch.

Choosing a graph. When you are unsure which graph to construct, ask what your independent variable is.



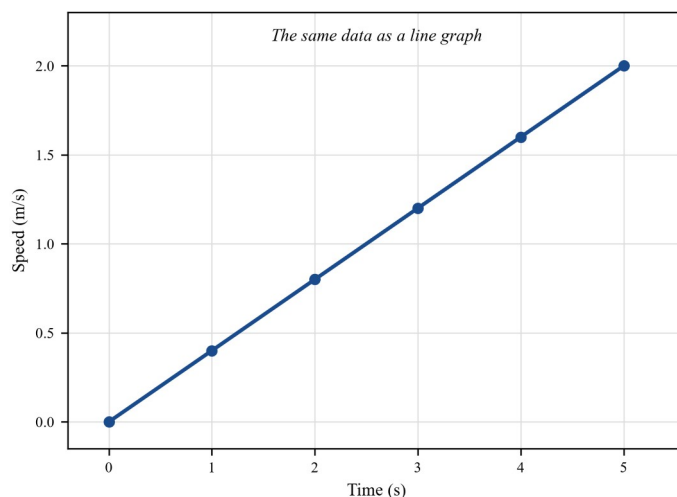
Decision flowchart starting from the independent variable: categories lead to a bar graph, or a pie chart if the parts add up to one whole; continuous values such as time or a chosen setting lead to a line graph; the distribution of many results leads to a histogram or dot plot; comparing groups suggests a boxplot; two continuous variables suggest a scatterplot, analysed in subtopic 4C.

A **scatterplot** is the graph for two continuous variables when you are looking for a relationship rather than a change over time. Constructing and analysing scatterplots, including lines of good fit, belongs to subtopic 4C; the flowchart here simply tells you when a scatterplot would be the right choice.

Speed and acceleration data (elaboration .E04). Dataloggers make it easy to collect speed and acceleration data, and the same data can be shown as a table or as a line graph.

Time (s)	Speed (m/s)	Acceleration (m/s ²)
0	0.0	—
1	0.4	0.4
2	0.8	0.4
3	1.2	0.4
4	1.6	0.4
5	2.0	0.4

The same data as a table

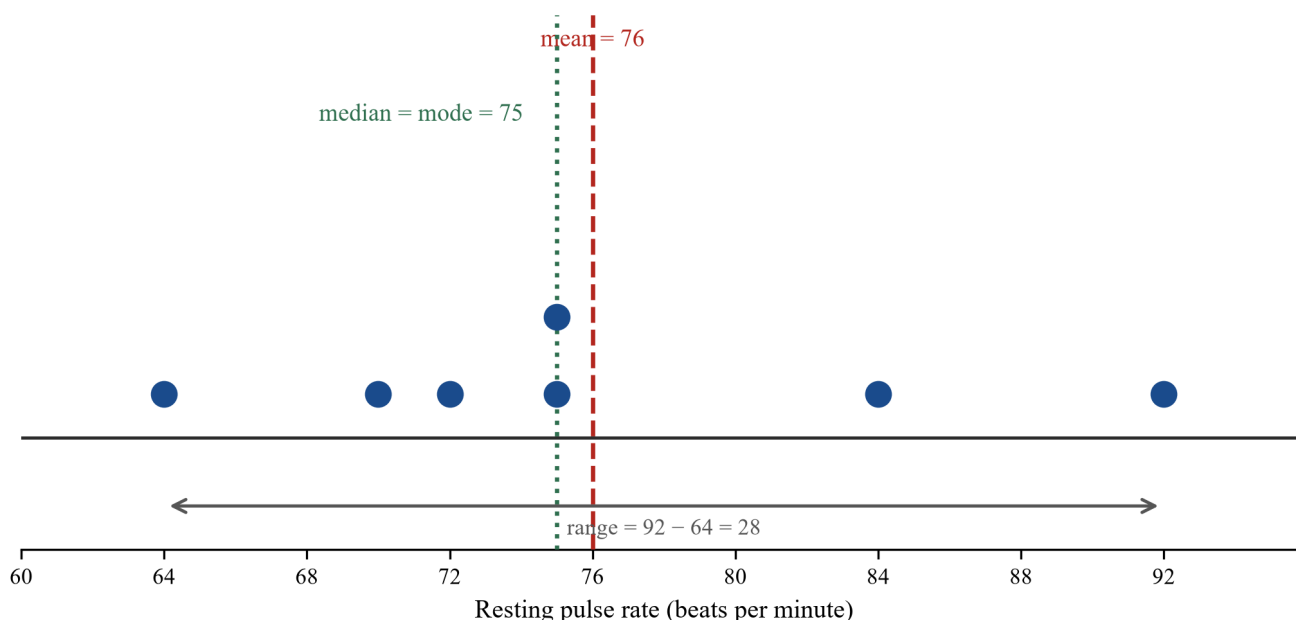


A table of time, speed and acceleration for a trolley beside the same data drawn as a line graph of speed against time; the straight line shows the speed growing by the same amount each second.

In this run the datalogger records the trolley's speed growing by 0.4 m/s every second, and it reports the same acceleration, 0.4 m/s², at every reading. Both representations hold the same information, and comparing them shows why the graph earns its place: in the table you must check the numbers row by row to see the steady growth, while in the graph the steady growth is a straight line you can see at a glance.

Descriptive statistics: centre and spread. Summary numbers come in two families. Measures of **centre** — the mean, the median and the mode — describe a typical value. Measures of **spread** — the range, and later the interquartile range — describe how much the values disagree with each other.

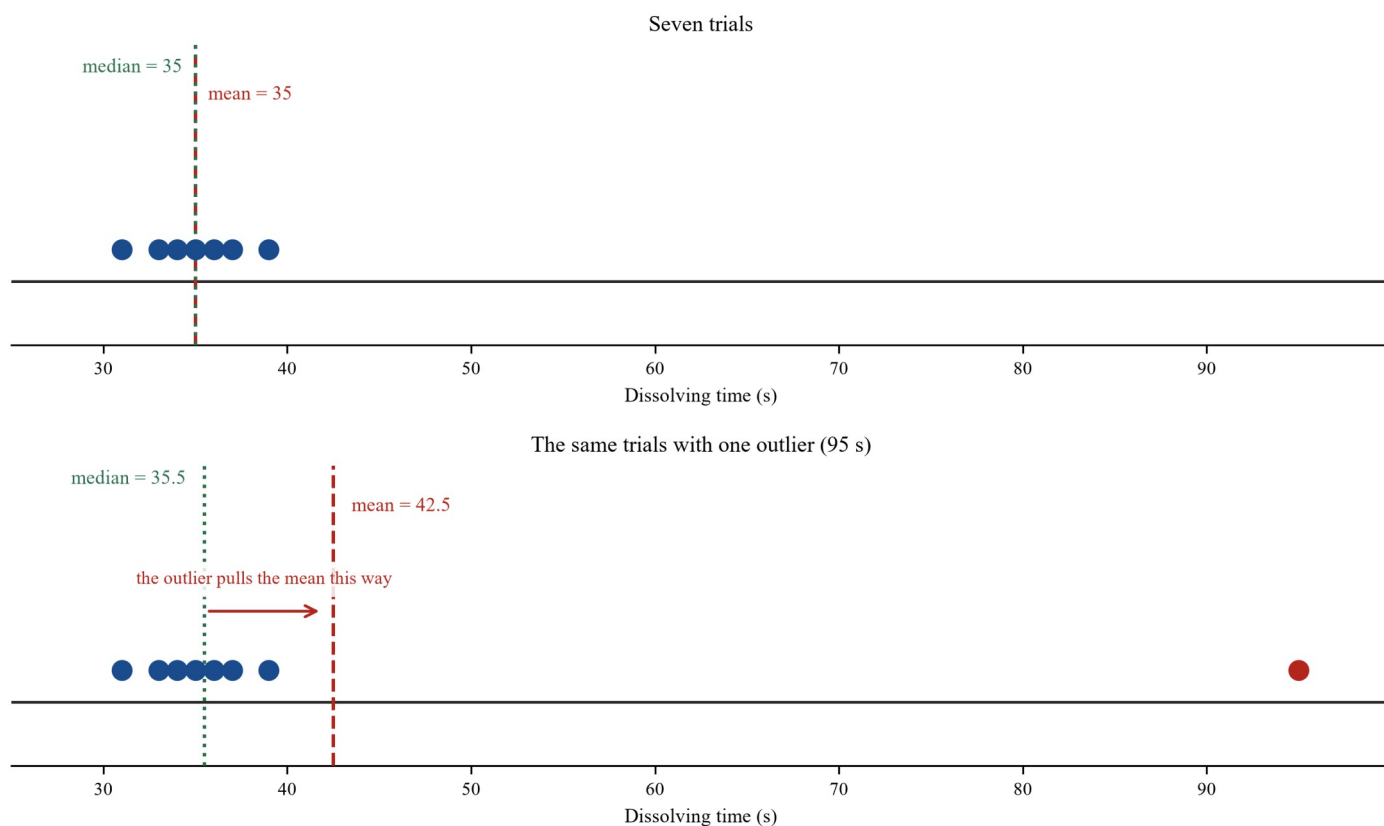
- The **mean** is the sum of the values divided by how many there are.
- The **median** is the middle value once the data is ordered.
- The **mode** is the value that appears most often.
- The **range** is the largest value minus the smallest value.



Dot plot of seven resting pulse rates in beats per minute, with the mean marked at 76, the median and mode marked at 75, and a double-headed arrow showing the range, 92 minus 64, equals 28.

A word about “average”. In everyday speech, *average* stands in for any typical value — “the average family”, “on average”. In science we are more careful. In this book, **average** means the **mean**, and the median and the mode are always named on their own. If you meet “average” in another book or in the news, check from the context which measure is meant before you calculate anything.

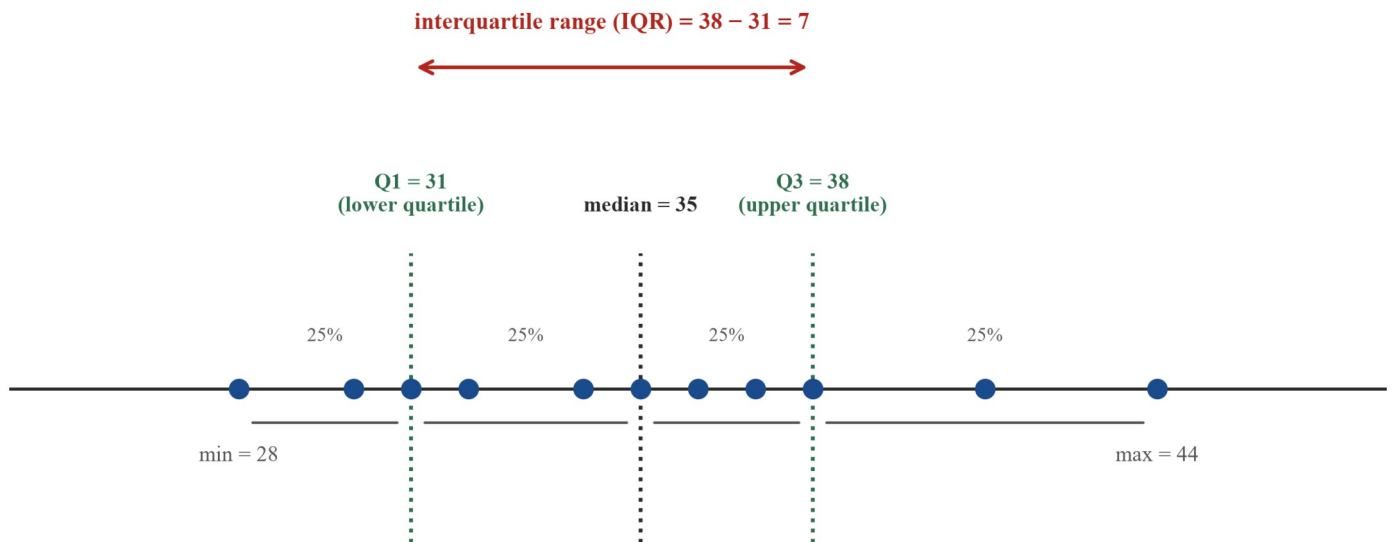
Outliers (elaboration .E02). An **outlier** is a value that sits far away from the rest of the data — on a graph you see it as a large gap between one point and the cluster of the others, and that large gap is always worth a second look.



Two dot plots of antacid dissolving times in seconds. Top: seven trials with the mean and the median both at 35. Bottom: the same trials with one outlier of 95 s, which pulls the mean to 42.5 while the median moves only to 35.5; an arrow shows the outlier pulling the mean.

The outlier changes the mean far more than the median. With seven trials the mean and the median are both 35 s; add one trial of 95 s and the mean jumps to 42.5 s while the median creeps to 35.5 s. An outlier may be a mistake — a misread scale, a swapped unit — or a genuine but unusual result, so the right move is to check the raw record. If it is a mistake, correct or discard it and say so; if it is genuine, keep it, but acknowledge its effect by reporting the median as well as the mean (“the mean is 42.5 s, but one unusual trial pulls it, so the median of 35.5 s is the honest summary”). Simply deleting an awkward value is never acceptable.

Quartiles and the interquartile range. The median splits ordered data into two halves. The **lower quartile (Q1)** is the median of the lower half and the **upper quartile (Q3)** is the median of the upper half; when there is an odd number of values, the median itself is left out of both halves. The **interquartile range (IQR)** is $Q3 - Q1$: the width of the middle 50% of the data. Unlike the range, the IQR is not distorted by one wild value, because the outer values never enter the calculation.

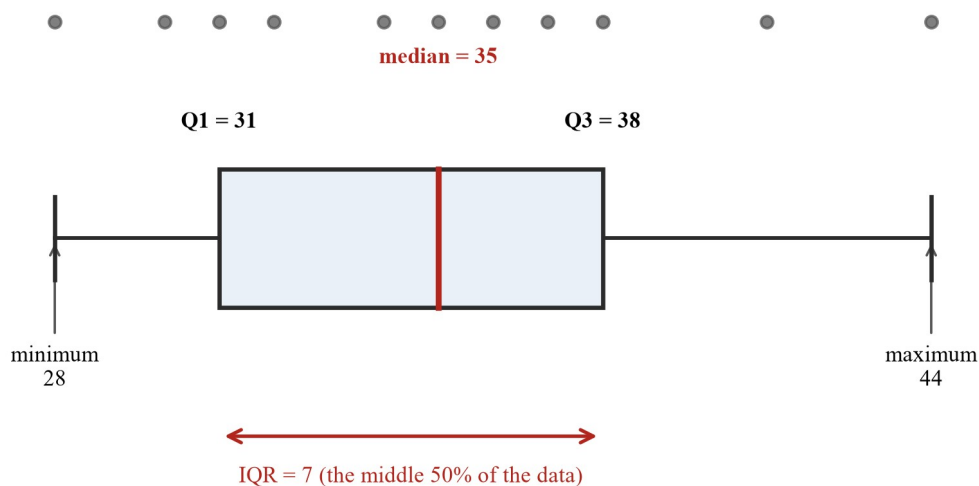


Dot plot of eleven ordered values split into quarters, marking minimum 28, $Q1 = 31$, median = 35, $Q3 = 38$ and maximum 44, each quarter labelled 25%, with a bracket showing the interquartile range, 38 minus 31, equals 7.

The IQR also gives a standard rule for **flagging** outliers: a value is flagged if it lies more than $1.5 \times \text{IQR}$ below $Q1$ or above $Q3$. For the data above, $1.5 \times 7 = 10.5$, so the limits sit at $31 - 10.5 = 20.5$ and $38 + 10.5 = 48.5$; every value sits inside, so nothing is flagged.

Box plots. The five numbers — minimum, $Q1$, median, $Q3$, maximum — form the **five-number summary**, and a **box plot** draws exactly those five numbers as a picture.

the data the boxplot summarises



Box plot built from the five-number summary 28, 31, 35, 38, 44, with the data points shown above it, labels for $Q1 = 31$, median = 35 and $Q3 = 38$, and a bracket marking the $\text{IQR} = 7$, the middle 50% of the data.

The box holds the middle 50% of the data, the line inside the box is the median, and the whiskers reach to the minimum and the maximum. Box plots shine when you compare two or more groups side by side — you will use them that way in Question 12.

Spreadsheets (elaboration .E01). Spreadsheet software carries out the calculations for you, which is how most real data is processed. If your data sits in cells B2 to B8, then `=AVERAGE (B2 : B8)` returns the mean, `=MEDIAN (B2 : B8)` the median, and `=MAX (B2 : B8) - MIN (B2 : B8)` the range. The quartiles come from `=QUARTILE (B2 : B8, 1)` ($Q1$) and `=QUARTILE (B2 : B8, 3)` ($Q3$).

	A	B	C
1	Trial	Volume of gas (cm³)	
2	1	33	
3	2	31	
4	3	36	
5	4	32	
6	5	37	
7	6	34	
8	7	35	
9			
10	Mean	=AVERAGE(B2:B8)	34
	Median	=MEDIAN(B2:B8)	34

What each formula returns

=AVERAGE(B2:B8)

→ mean = 34 cm³

=MEDIAN(B2:B8)

→ median = 34 cm³

=MAX(B2:B8)–MIN(B2:B8)

→ range = 37 – 31 = 6 cm³

=QUARTILE(B2:B8,3)–QUARTILE(B2:B8,1)

→ IQR = 36 – 32 = 4 cm³

Seven trials of a gas-collection measurement (simulated data)

Spreadsheet grid holding seven trials of a gas-collection measurement in column B, with formulas alongside:
 =AVERAGE(B2:B8) returns mean 34, =MEDIAN(B2:B8) returns median 34, =MAX–MIN returns range 6, and the difference of two =QUARTILE calls returns IQR 4.

A spreadsheet is only as trustworthy as the numbers typed into it — mistype 37 as 73 and the mean will be wrong, and the software will not complain. That is why hand calculation on small datasets, like the ones in this subtopic, is still worth practising: it tells you the size of answer to expect, so a mistyped cell stands out.

Activities

Activity 1 — Sort the variables. Classify each variable as categorical, discrete numerical or continuous numerical, and name the graph type from the flowchart you would start with: (a) the species of bird at a feeder; (b) the number of seeds that germinate in each pot; (c) the time for a sugar cube to dissolve; (d) the temperature of the water in each of five beakers; (e) the brand of fertiliser used; (f) the number of legs on the insects in a pitfall trap.

Activity 2 — Build the table. A class measures how far a toy car rolls on grass, concrete and carpet, testing each surface three times. The raw results, in the order they were written down, are: carpet 42 cm, grass 88 cm, concrete 150 cm, grass 92 cm, carpet 38 cm, concrete 156 cm, grass 90 cm, concrete 144 cm, carpet 40 cm. Draw a table that organises these results the way Table 1 does — independent variable first, unit in the heading, one column per trial, and a mean column — then fill it in and calculate the three means.

Activity 3 — Let the spreadsheet do the arithmetic (elaboration .E01). Enter the seven gas volumes from the spreadsheet figure (33, 31, 36, 32, 37, 34, 35 cm³) into a spreadsheet column. In the cells beside the data, enter formulas for the mean, the median, the range and the IQR, and check that you get 34, 34, 6 and 4 cm³. Then change the 37 to 73, watch what happens to the mean and the median, and write one sentence explaining why one of them moved more. (Then change it back.)

Worked examples

All data here are simulated data for practice. Summary values are exact unless shown with ≈.

Example 1 — scaffolded

A student drops a tennis ball from 1 m and measures the height of the first bounce, in cm, on nine tries: 38, 35, 39, 38, 41, 36, 38, 37, 40. Calculate the mean, median, mode and range.

Working

Ordered: 35, 36, 37, 38, 38, 38, 39, 40, 41

Comment

Ordering first makes the median, mode and range easy

Working	Comment
Mean = $(35 + 36 + \dots + 41) \div 9 = 342 \div 9 = 38$ cm	to read. The mean uses every value.
Median = 5th value = 38 cm	With nine values, the 5th is the middle one.
Mode = 38 cm (appears three times)	38 appears more often than any other value.
Range = $41 - 35 = 6$ cm	Largest minus smallest.

All four summaries agree that a bounce of about 38 cm is typical, and the small range of 6 cm tells you the trials were consistent.

Example 2 — scaffolded

A student records the temperature of hot water as it cools, every 2 min for 10 min. The results are 80, 68, 59, 52, 47 and 43 °C. Choose a graph type and set up the axes.

Working	Comment
Independent variable: time, which is continuous	Time is measured, not labelled, so the categories branch of the flowchart does not apply.
Choose a line graph	A continuous independent variable changing over time is the line-graph branch.
x-axis: Time (min), from 0 to 10	The independent variable goes on the horizontal axis, labelled with quantity and unit.
y-axis: Temperature (°C), starting at 40 rather than 0	The dependent variable goes on the vertical axis; starting above zero keeps the points spread across the grid.
A title or caption that names the run	“Temperature of cooling water” tells the reader what was measured.

Constructed this way, the graph is the cooling curve shown in the axes figure above.

Example 3 — unscaffolded

A student measures a classmate’s reaction time with a dropped ruler, eleven times, in seconds: 0.27, 0.24, 0.29, 0.21, 0.30, 0.28, 0.52, 0.26, 0.31, 0.25, 0.34. Find the five-number summary and the IQR, and use the $1.5 \times \text{IQR}$ rule to decide whether any value should be flagged as an outlier.

Order the data: 0.21, 0.24, 0.25, 0.26, 0.27, 0.28, 0.29, 0.30, 0.31, 0.34, 0.52.

The median is the 6th value, 0.28 s. The lower half is 0.21, 0.24, 0.25, 0.26, 0.27, so $Q1 = 0.25$ s; the upper half is 0.29, 0.30, 0.31, 0.34, 0.52, so $Q3 = 0.31$ s. The five-number summary is 0.21, 0.25, 0.28, 0.31, 0.52 s, and $\text{IQR} = 0.31 - 0.25 = 0.06$ s.

The limits sit at $Q1 - 1.5 \times \text{IQR} = 0.25 - 0.09 = 0.16$ s and $Q3 + 1.5 \times \text{IQR} = 0.31 + 0.09 = 0.40$ s. The value 0.52 s lies beyond the upper limit, so it is flagged as an outlier; every other value sits inside. Notice the effect on the centre: the mean of all eleven values is about 0.30 s, above the median of 0.28 s, because the outlier pulls the mean upward.

Example 4 — unscaffolded

A class counts the birds at a school feeder at lunchtime on eight days: 4, 5, 5, 6, 5, 4, 6, 26. On the last day a passing flock briefly landed. Calculate the mean and the median, decide which better represents a typical day, and check your decision by recalculating the mean without the 26.

With all eight values, $\text{mean} = 61 \div 8 \approx 7.6$ birds, and the ordered data 4, 4, 5, 5, 5, 6, 6, 26 gives $\text{median} = (5 + 5) \div 2 = 5$ birds. On seven of the eight days the count was 4–6 birds, so a typical day is about 5 birds — the median. The mean of 7.6 is higher than every count except the one day that produced it, so it represents no day at all.

Without the 26, mean = $35 \div 7 = 5.0$ birds, equal to the median. Removing the outlier does not move the median — it was 5 either way — but it pulls the mean back to where the data actually sits, exactly the behaviour shown in the outlier figure.

Practice questions

Scaffolded questions

Question 1 Classify each variable as categorical, discrete numerical or continuous numerical. (a) the colour of a flower (b) the number of bubbles counted in one minute (c) the mass of a seed, in grams (d) the type of soil in each pot

Question 2 Look at the pondweed results table (Table 1) in the Theory. (a) Which variable is the independent variable? (b) Which variable is the dependent variable? (c) Why were three trials run for each colour? (d) The blue trials are 34, 38 and 36 bubbles per minute. Confirm the mean shown in the table.

Question 3 Ten sunflower seeds are weighed, in grams: 0.42, 0.51, 0.47, 0.55, 0.49, 0.44, 0.52, 0.48, 0.46, 0.50. Calculate (a) the mean, (b) the median, (c) the range.

Question 4 Use the graph-choice flowchart to pick a graph for each investigation. (a) a student records her pulse rate every minute while she exercises for ten minutes (b) a class surveys the favourite fruit of each student (c) a technician records the heights of 200 Year 10 students, grouped into 5 cm bins (d) a report shows how Australia's electricity generation splits between coal, gas, wind and solar

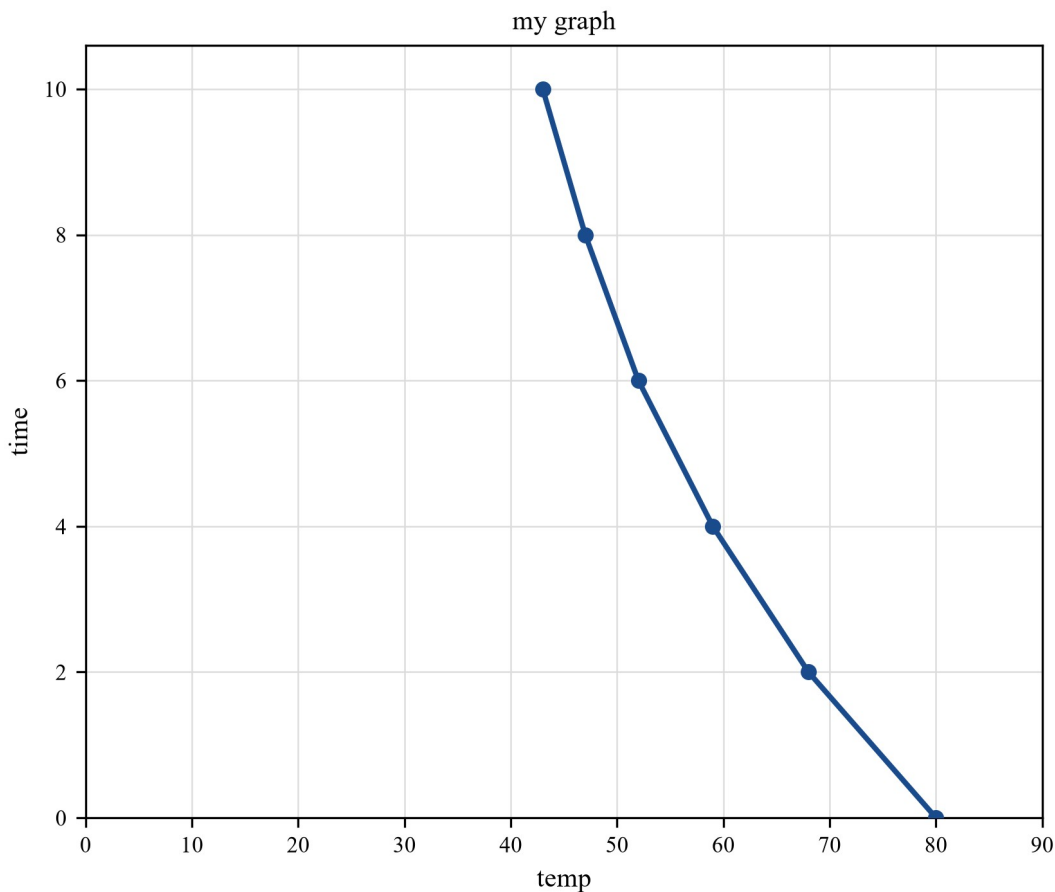
Question 5 Twelve gum leaves are measured, in mm: 41, 43, 44, 46, 47, 49, 51, 52, 54, 56, 58, 61. Find (a) the median, (b) Q1 and Q3, (c) the IQR, (d) whether the $1.5 \times \text{IQR}$ rule flags any value.

Question 6 A spreadsheet holds seven measurements in cells B2 to B8. Write the formula that returns (a) the mean, (b) the median, (c) the range, (d) the upper quartile.

Unscaffolded questions

Question 7 A ball rolls down a ramp and its speed at the bottom is measured, in cm/s, for three ramp heights. The trials give: 10 cm — 42, 45, 45; 20 cm — 78, 81, 81; 30 cm — 112, 114, 116. (a) Calculate the mean speed for each height. (b) Which graph type from the flowchart suits these data, and why? (c) Describe the pattern the data shows.

Question 8 The graph below is a student's first attempt at drawing the cooling-water data from the Theory. List four things that must be fixed before the graph is acceptable.



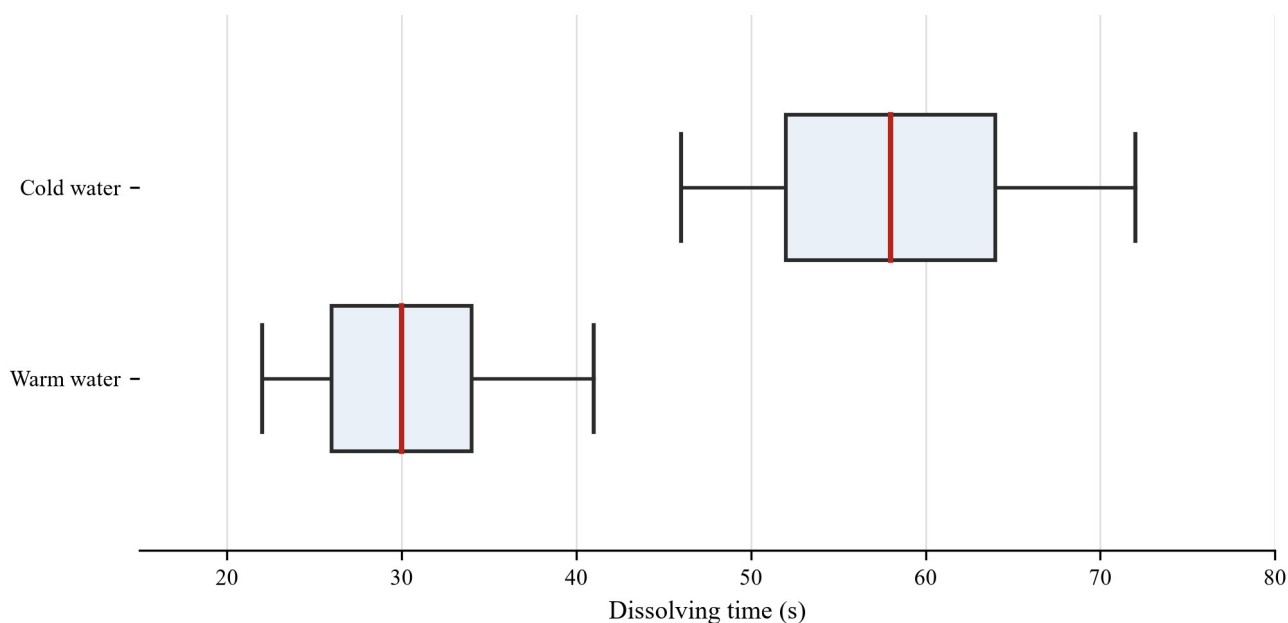
A flawed graph of the cooling-water data: the axes are swapped so time sits on the vertical axis, the axis labels are shortened and carry no units, the title reads only “my graph”, and a poor choice of scale squeezes the points into one corner of the grid.

Question 9 The times, in seconds, for ten students to solve a maze are: 24, 31, 27, 25, 38, 26, 30, 27, 29, 27. Calculate the mean, median, mode and range, then use the $1.5 \times \text{IQR}$ rule to check whether the 38 s should be flagged as an outlier.

Question 10 A pool is tested each morning for a week, and the pH readings are 6.8, 7.0, 7.1, 6.9, 7.2, 7.0 — then, on the seventh day, 2.1. The student who wrote down 2.1 finds that the probe had not been rinsed, so the reading is a mistake. (a) Calculate the mean and the median of all seven readings. (b) Calculate the mean of the six good readings. (c) What does the difference tell you about reporting the mean when a bad value slips in?

Question 11 The overnight minimum temperatures, in $^{\circ}\text{C}$, recorded over thirteen days at a school weather station are: 12, 13, 14, 14, 15, 16, 17, 18, 19, 19, 21, 22, 24. Write down the five-number summary and the IQR, and state whether the $1.5 \times \text{IQR}$ rule flags any value.

Question 12 The box plots below summarise the time, in seconds, for an antacid tablet to dissolve in cold water and in warm water.



Two box plots on one scale of dissolving time in seconds: cold water, minimum 46, $Q1$ 52, median 58, $Q3$ 64, maximum 72; warm water, minimum 22, $Q1$ 26, median 30, $Q3$ 34, maximum 41.

- Write down the median dissolving time for each.
- Calculate the IQR for each.
- In your own words: which water dissolves the tablet faster, and which set of results is more spread out in its middle 50%?

Question 13 A datalogger records the speed of a trolley, in m/s, each second from 0 s to 5 s: 0.0, 0.4, 0.8, ?, ?, 2.0. Two values were lost. (a) Use the pattern in the data to fill in the missing speeds at 3 s and 4 s. (b) Why is a line graph the right choice for these data?

Question 14 The histogram in the bar-graph-versus-histogram figure counts the reaction times of 30 students. (a) How many students fall in the busiest bin, and which bin is it? (b) Show that the bin counts add to 30. (c) Explain in one sentence why the bars of a histogram touch.

Question 15 The pie chart in the three-graphs figure shows how a household's waste splits by type. If the household threw out 2 000 kg of waste in a year, use the chart to calculate the mass of (a) paper, (b) metal.

Question 16 A gardener counts the number of ladybeetles on each of nine rose bushes: 15, 11, 17, 13, 15, 18, 14, 15, 17. Calculate the mean, median, mode and range.

Answers

Question 1 (a) categorical; (b) discrete numerical; (c) continuous numerical; (d) categorical.

Question 2 (a) the colour of light; (b) the rate of oxygen bubbles (bubbles per minute); (c) repeats check consistency and expose any trial that went wrong; (d) $(34 + 38 + 36) \div 3 = 108 \div 3 = 36$ bubbles per minute, as shown.

Question 3 (a) $0.484 \text{ g} \approx 0.48 \text{ g}$; (b) 0.485 g ; (c) 0.13 g .

Question 4 (a) line graph; (b) bar graph (a pie chart is also acceptable); (c) histogram; (d) pie chart.

Question 5 (a) 50 mm; (b) $Q1 = 45 \text{ mm}$, $Q3 = 55 \text{ mm}$; (c) 10 mm; (d) the limits sit at 30 mm and 70 mm, so nothing is flagged.

Question 6 (a) $=\text{AVERAGE}(B2:B8)$; (b) $=\text{MEDIAN}(B2:B8)$; (c) $=\text{MAX}(B2:B8)-\text{MIN}(B2:B8)$; (d) $=\text{QUARTILE}(B2:B8,3)$.

Question 7 (a) 44, 80 and 114 cm/s; (b) a line graph — the ramp height is a continuous value the investigator chose; (c) the higher the ramp, the greater the speed at the bottom.

Question 8 The axes are swapped (time belongs on the horizontal axis); the axis labels are shortened and carry no units; the title “my graph” says nothing about the data; the scale squeezes the points into one corner instead of filling most of the grid.

Question 9 Mean 28.4 s, median 27 s, mode 27 s, range 14 s; $Q1 = 26 \text{ s}$, $Q3 = 30 \text{ s}$, $IQR = 4 \text{ s}$, limits 20 s and 36 s — the 38 s lies beyond the upper limit and is flagged.

Question 10 (a) mean 6.3, median 7.0; (b) 7.0; (c) one bad value dragged the mean a full pH unit off the truth while the median did not move.

Question 11 12, 14, 17, 20, 24 °C; $IQR 6 \text{ °C}$; limits 5 °C and 29 °C — nothing is flagged.

Question 12 (a) cold 58 s, warm 30 s; (b) cold $IQR 12 \text{ s}$, warm $IQR 8 \text{ s}$; (c) warm water dissolves the tablet faster, and the cold-water results are more spread out in their middle 50%.

Question 13 (a) 1.2 m/s and 1.6 m/s; (b) time is continuous and the speed changes with it — the line-graph branch of the flowchart.

Question 14 (a) 10 students, in the 0.20–0.24 s bin; (b) $2 + 6 + 10 + 7 + 4 + 1 = 30$; (c) the bins are neighbouring slices of one continuous scale, so there is no gap between them.

Question 15 (a) 30% of 2 000 kg = 600 kg; (b) 10% of 2 000 kg = 200 kg.

Question 16 Mean 15, median 15, mode 15, range 7 ladybeetles.

Fully-worked solutions

Question 1 (a) Flower colour is a label, not a number: categorical. (b) Bubbles are counted in whole numbers: discrete numerical. (c) Mass is measured and can sit anywhere on a scale: continuous numerical. (d) Soil type is a label: categorical.

Question 2 (a) The colour of light is chosen by the investigator, so it is the independent variable and sits in the first column. (b) The rate of bubbles is what is measured in response: the dependent variable. (c) Three trials let you see whether the results are consistent; a trial that went wrong shows up as a value far from the other two and can be checked or repeated. (d) $(34 + 38 + 36) \div 3 = 108 \div 3 = 36$ bubbles per minute, matching the table.

Question 3 $\text{Sum} = 0.42 + 0.51 + 0.47 + 0.55 + 0.49 + 0.44 + 0.52 + 0.48 + 0.46 + 0.50 = 4.84 \text{ g}$. (a) $\text{Mean} = 4.84 \div 10 = 0.484 \text{ g} \approx 0.48 \text{ g}$. (b) Ordered: 0.42, 0.44, 0.46, 0.47, 0.48, 0.49, 0.50, 0.51, 0.52, 0.55; the two middle values are 0.48 and 0.49, so $\text{median} = (0.48 + 0.49) \div 2 = 0.485 \text{ g}$. (c) $\text{Range} = 0.55 - 0.42 = 0.13 \text{ g}$.

Question 4 (a) Pulse rate against time: a continuous variable measured over time → line graph. (b) Favourite fruit is categories → bar graph; a pie chart is also acceptable, because the parts add up to the whole class. (c) Heights are

continuous and the question asks about the distribution of many results → histogram. (d) Parts of one whole → pie chart.

Question 5 The data is already ordered. (a) With twelve values the median is the average of the 6th and 7th: $(49 + 51) \div 2 = 50$ mm. (b) The lower half is 41, 43, 44, 46, 47, 49, so $Q1 = (44 + 46) \div 2 = 45$ mm; the upper half is 51, 52, 54, 56, 58, 61, so $Q3 = (54 + 56) \div 2 = 55$ mm. (c) $IQR = 55 - 45 = 10$ mm. (d) $1.5 \times IQR = 15$ mm, so the limits sit at $45 - 15 = 30$ mm and $55 + 15 = 70$ mm; the smallest leaf (41 mm) and the largest (61 mm) both sit inside, so nothing is flagged.

Question 6 (a) =AVERAGE(B2:B8). (b) =MEDIAN(B2:B8). (c) =MAX(B2:B8)-MIN(B2:B8). (d) =QUARTILE(B2:B8,3) — the 3 asks for the upper quartile; a 1 would give the lower quartile.

Question 7 (a) 10 cm: $(42 + 45 + 45) \div 3 = 132 \div 3 = 44$ cm/s. 20 cm: $(78 + 81 + 81) \div 3 = 240 \div 3 = 80$ cm/s. 30 cm: $(112 + 114 + 116) \div 3 = 342 \div 3 = 114$ cm/s. (b) The ramp height is a continuous value the investigator chose, so the line-graph branch applies; a bar graph would hide that the heights sit on one scale. (c) As the ramp height increases, the speed at the bottom increases; from 10 cm to 30 cm the mean speed more than doubles.

Question 8 Four fixes: (1) the axes are swapped — time, the independent variable, belongs on the horizontal axis and temperature on the vertical axis; (2) the axis labels are shortened and carry no units — they must read, for example, “Time (min)” and “Temperature (°C)”; (3) the title “my graph” says nothing — a title or caption must name what was measured; (4) the scale wastes most of the grid and squeezes the points into one corner — the scale should let the data fill most of the grid.

Question 9 Sum = $24 + 31 + 27 + 25 + 38 + 26 + 30 + 27 + 29 + 27 = 284$ s; mean = $284 \div 10 = 28.4$ s. Ordered: 24, 25, 26, 27, 27, 27, 29, 30, 31, 38; median = $(27 + 27) \div 2 = 27$ s; mode = 27 s (three times); range = $38 - 24 = 14$ s. The lower half is 24, 25, 26, 27, 27, so $Q1 = 26$ s; the upper half is 27, 29, 30, 31, 38, so $Q3 = 30$ s; $IQR = 4$ s. $1.5 \times 4 = 6$, so the limits sit at $26 - 6 = 20$ s and $30 + 6 = 36$ s. The 38 s lies beyond the upper limit and is flagged as an outlier — worth checking whether that student misread the stopwatch, but not to be deleted without a reason.

Question 10 (a) Sum of all seven = 44.1; mean = $44.1 \div 7 = 6.3$. Ordered: 2.1, 6.8, 6.9, 7.0, 7.0, 7.1, 7.2; median = 7.0. (b) Sum of the six good readings = 42.0; mean = $42.0 \div 6 = 7.0$. (c) The bad reading dragged the mean from 7.0 all the way to 6.3 — a full pH unit — while the median stayed at 7.0 either way. When one suspect value sits in a small dataset, the mean cannot be trusted; report the median, and find out what went wrong before deciding whether the value stays.

Question 11 The data is already ordered. Median = 7th value = 17 °C. The lower half is 12, 13, 14, 14, 15, 16, so $Q1 = (14 + 14) \div 2 = 14$ °C; the upper half is 18, 19, 19, 21, 22, 24, so $Q3 = (19 + 21) \div 2 = 20$ °C. Five-number summary: 12, 14, 17, 20, 24 °C; $IQR = 6$ °C. The limits sit at $14 - 9 = 5$ °C and $20 + 9 = 29$ °C; the minimum 12 and the maximum 24 sit inside, so nothing is flagged.

Question 12 (a) Cold water median 58 s; warm water median 30 s. (b) Cold: $IQR = 64 - 52 = 12$ s. Warm: $IQR = 34 - 26 = 8$ s. (c) The warm-water box sits well to the left of the cold-water box, so tablets dissolve faster in warm water — the median time is about half. The cold-water box is wider (IQR 12 s against 8 s), so the middle 50% of the cold-water results is more spread out.

Question 13 (a) The speed grows by 0.4 m/s each second, so at 3 s the speed is $0.8 + 0.4 = 1.2$ m/s and at 4 s it is $1.2 + 0.4 = 1.6$ m/s (and $1.6 + 0.4 = 2.0$ m/s at 5 s, matching the table). (b) Time is a continuous variable and the speed changes with it, so the line-graph branch of the flowchart applies; the graph also makes the steady growth visible as a straight line, as the speed figure shows.

Question 14 (a) The tallest bar holds 10 students, in the 0.20–0.24 s bin. (b) $2 + 6 + 10 + 7 + 4 + 1 = 30$ students. (c) The bins are neighbouring slices of one continuous scale, so there is no gap between them and the bars touch.

Question 15 (a) Paper is 30% of the whole: $0.30 \times 2\,000$ kg = 600 kg. (b) Metal is 10%: $0.10 \times 2\,000$ kg = 200 kg.

Question 16 Sum = $15 + 11 + 17 + 13 + 15 + 18 + 14 + 15 + 17 = 135$; mean = $135 \div 9 = 15$ ladybeetles. Ordered: 11, 13, 14, 15, 15, 15, 17, 17, 18; median = 5th value = 15; mode = 15 (three times); range = $18 - 11 = 7$ ladybeetles.

Where this leads: constructing and analysing scatterplots, trends and lines of good fit belongs to subtopic 4C (with VC2M10ST02); models, symbols, formulas and mathematical relationships as representations belong to the partner subtopic 4B; motion graphs and the gradient as a physical quantity belong to 15C. Box plots and quartiles return in VCE studies, mentioned here as reference only (TOC D9).

Models, symbols, formulas and mathematical relationships as representations

Chapter 4 — Processing, modelling and analysing (Science, Levels 9 and 10)

Subtopic 4B · VC2 Science — Processing, modelling and analysing

CD code: VC2S10I04 (split delivery †; this subtopic is the partner, not the owner) **Content description (verbatim):** “data and information can be organised, processed and summarised by selecting and constructing representations including tables, graphs, descriptive statistics, models, symbols, formulas and mathematical relationships” **This subtopic’s limb (TOC §4 clause split):** models, symbols, formulas and mathematical relationships. Owner 4A delivers tables, graph choice and descriptive statistics. **Elaboration ids drawn on:** VC2S10I04.E03 · VC2S10I04.E05 · VC2S10I04.E06 · VC2S10I04.E07 (VC2S10I04.E01, .E02 and .E04 are drawn on by Subtopic 4A.) **Achievement standard (extract served):** “They select and construct a range of appropriate representations to organise, process and summarise data and information.” **Maths interlock (D11):** VC2M9M05 — direct proportion, rates, ratio and scale.

Most data in this subtopic are simulated data for practice, and each dataset is labelled where it appears. Two values are real and carry source lines where they are used: the Earth–Moon dimensions from NASA’s Moon Fact Sheet (retrieved 2026-08-20), and the carbon-storage rule of thumb from the IPCC 2006 Guidelines, Volume 4 (retrieved 2026-08-20). No real-world measurement is attributed to any person or place.

Theory

One data set, many representations. A walker moving at a steady pace covers 80 m every minute. That single fact can be recorded as a table of values, drawn as a line graph, compressed into a formula, or said out loud as a sentence. Each version is a **representation** — a way of standing in for the real situation so you can think about it, share it or calculate with it.

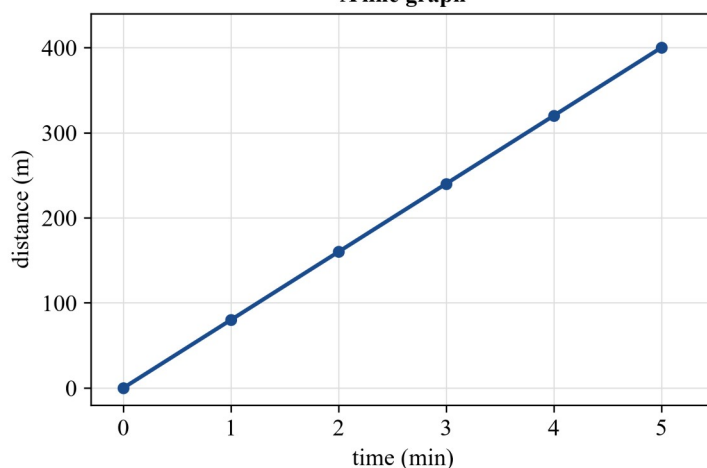
The same data set, four different representations
(simulated data for practice: a walker covers 80 m each minute)

A table

t (min)	0	1	2	3	4	5
d (m)	0	80	160	240	320	400

exact values, one at a time

A line graph



A formula

$$d = 80t$$

d in metres, t in minutes

every value in one line — predicts any input

A sentence

“The distance grows by 80 m every minute.”

a pattern you can say out loud

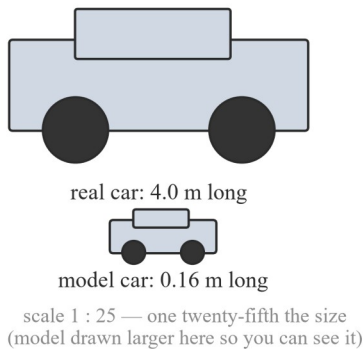
The same data set shown four ways: a table of time and distance values, a line graph of distance against time, the formula $d = 80t$, and the sentence “The distance grows by 80 m every minute”.

Subtopic 4A covers the first family of representations — tables, graphs and descriptive statistics. This subtopic covers the rest of the content description: **models, symbols, formulas and mathematical relationships**. None of these is “better” than the others. Each is a tool, and the skill is choosing the tool that fits the job: a table gives exact values one at a time, a graph shows the shape of the data at a glance, a formula predicts any value in one line, and a sentence states the pattern in words.

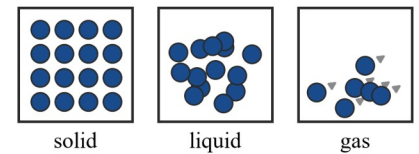
What a model is. A **model** is a simplified stand-in for something real, built to describe, explain or predict it. Scientists reach for models whenever the real thing is too big, too small, too fast, too slow, too far away or too dangerous to work with directly. Four kinds of model do most of the work in science.

Four kinds of model

Physical (scale) model



Conceptual (picture) model



the particle model — matter drawn as tiny particles

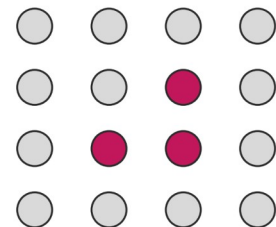
Mathematical model

$$v = \frac{d}{t}$$

speed = distance ÷ time

a formula that stands in for the real motion

Simulation (computer) model



a program steps a situation forward in time
(three dots carrying an infection spread to others)

Four kinds of model: a physical scale model of a car, a conceptual particle model of solids, liquids and gases, the mathematical model $v = d \div t$, and a computer simulation of an infection spreading through a group.

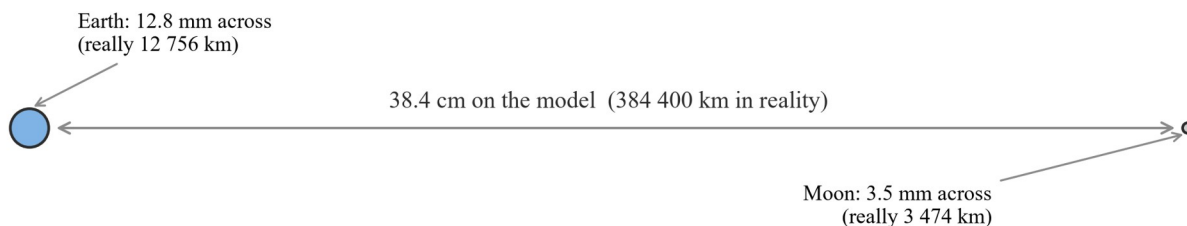
- A **physical model** is a smaller or larger copy you can hold, such as a 1 : 25 model car or a ball-and-stick molecule.
- A **conceptual model** is a picture in words or diagrams, such as the **particle model** of matter, which draws solids, liquids and gases as tiny particles arranged differently.
- A **mathematical model** is a formula that stands in for a real situation, such as $v = d \div t$ standing in for any steady motion.
- A **simulation** is a computer program that steps a situation forward in time, round by round, such as a model of an infection spreading through a group.

Every model is simpler than the thing it stands for — that is the point, and also the limit. A model car has no engine; the particle model draws particles as balls even though atoms are not little balls; a simulation follows the rules it was given and no others. When you use a model, you should always be able to say what it leaves out.

Scale models and scale factors. A **scale model** shrinks (or enlarges) every length of the real object by the same factor. The scale is written as a ratio, model size : real size. A scale of **1 : 25** means every length on the model is one twenty-fifth of the real length: a real car 4.0 m long becomes $4.0 \div 25 = 0.16$ m, or 16 cm, on the model.

The same idea works at far larger scales. At a scale of **1 : 1 000 000 000** — one to a billion — every length is one billionth of the real one, so the Earth–Moon system shrinks to something that fits on a desk.

A 1 : 1 000 000 000 scale model of the Earth–Moon system, sizes and distance to the same scale
(sizes from NASA's Moon Fact Sheet, retrieved 2026-08-20)



A 1 to 1 000 000 000 scale model of the Earth–Moon system: Earth drawn 12.8 mm across, the Moon 3.5 mm across, and the two 38.4 cm apart, sizes and distance to the same scale.

The sizes used in that figure are real: Earth's diameter is 12 756 km, the Moon's diameter is 3 474 km, and the mean Earth–Moon distance is 384 400 km. (Source: NASA, "Moon Fact Sheet", <https://nssdc.gsfc.nasa.gov/planetary/factsheet/moonfact.html>, retrieved 2026-08-20.) The calculation divides each real length by the scale factor, then converts to a unit you can hold a ruler against: $12\,756\text{ km} \div 1\,000\,000\,000 = 0.000\,012\,756\text{ km} = 12.8\text{ mm}$; $3\,474\text{ km}$ becomes 3.5 mm ; and $384\,400\text{ km}$ becomes 38.4 cm . The surprise is the message: the distance between the two bodies is about 30 times Earth's diameter, so a true-to-scale picture is mostly empty space.

Symbols carry exact, agreed meanings. A **symbol** is a mark that stands for something else, and science runs on them. Three families do different jobs.

Symbols carry exact, agreed meanings

Chemical formulas	Unit symbols	Quantities and arrows
H_2O subscript 2 = two hydrogen atoms no subscript = one oxygen atom CO_2 one carbon, two oxygens CH_4 one carbon, four hydrogens H = hydrogen, O = oxygen, C = carbon — agreed worldwide	m metre kg kilogram s second °C degrees Celsius N newton m/s metres per second J joule the same symbols in every country (SI)	d distance t time v speed m mass $A + B \rightarrow C + D$ the arrow means "react to form": reactants on the left, products on the right say what each letter stands for the first time you use it

Three columns of symbols: chemical formulas such as H_2O , CO_2 and CH_4 with their subscripts explained; unit symbols m , kg , s , $^{\circ}C$, N , m/s and J ; and quantity symbols d , t , v and m beside the reaction arrow $A + B \rightarrow C + D$.

- Chemical formulas** name substances. Each element has a one- or two-letter symbol agreed worldwide — H for hydrogen, O for oxygen, C for carbon — and a small lowered number, the **subscript**, counts the atoms of the element just before it. In H_2O the subscript 2 says there are two hydrogen atoms; an element with no subscript has one atom, so H_2O has one oxygen atom. CO_2 is one carbon and two oxygens; CH_4 is one carbon and four hydrogens.
- Unit symbols** are the same in every country that uses the SI system: m for metre, kg for kilogram, s for second, $^{\circ}C$ for degrees Celsius, N for newton, m/s for metres per second, J for joule.

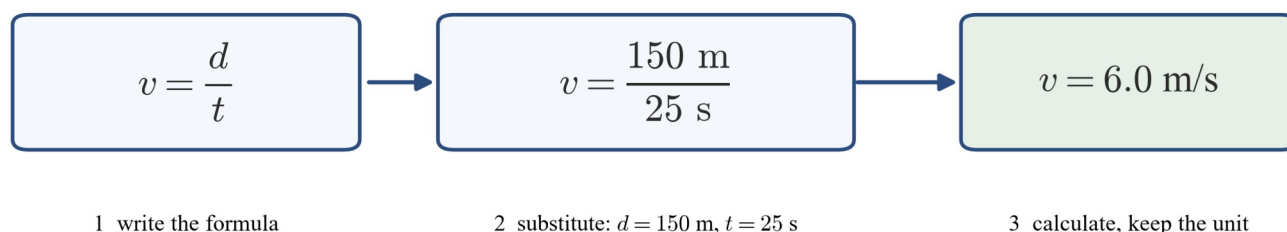
- **Quantity symbols** are letters that stand for measurable amounts: d for distance, t for time, v for speed, m for mass. Say what each letter stands for the first time you use it. In chemistry, the arrow $\rightarrow$ means “react to form”: the substances on the left are the **reactants**, and the substances on the right are the **products**.

Formulas and substitution. A **formula** is a compact mathematical relationship between quantities. The speed formula, for example, is

$$v = \frac{d}{t}$$

where v is speed in m/s, d is distance in metres and t is time in seconds. To **substitute** means to replace each letter with its number and unit, then calculate.

Substituting into a formula — replace each letter with its number and unit

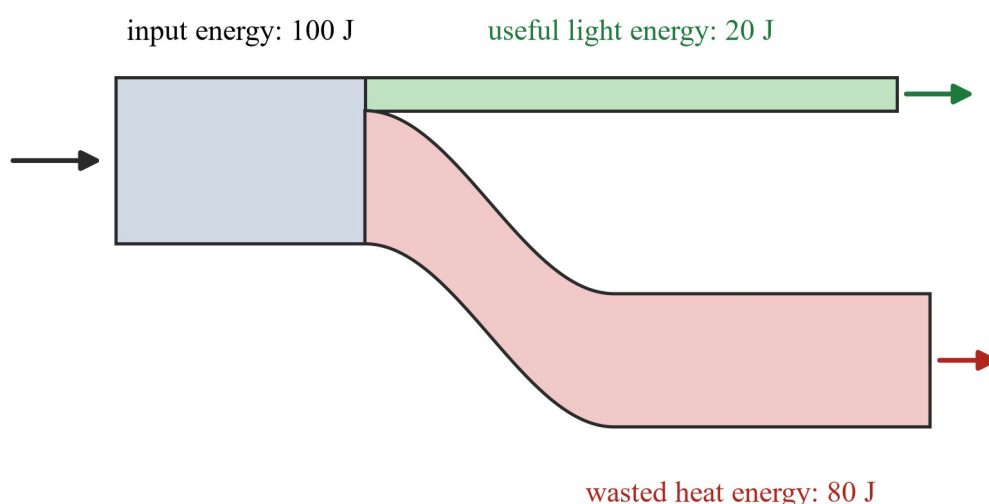


Substituting into $v = d \div t$ in three steps: write the formula, substitute $d = 150 \text{ m}$ and $t = 25 \text{ s}$, then calculate $v = 6.0 \text{ m/s}$ and keep the unit.

The three steps are always the same: write the formula, substitute the numbers with their units, then calculate and keep the unit in the answer. In this subtopic you substitute into formulas that are already set up for the letter you want; rearranging a formula to make a different letter the subject comes later, in the topics that need it.

Energy-flow diagrams show ratios (elaboration VC2S10I04.E03). When energy is transferred, some of it ends up where it is wanted and the rest spreads into the surroundings, usually as heat. An **energy-flow diagram** (also called a **Sankey diagram**) draws this split with branches whose widths are exactly proportional to the amounts of energy.

An energy flow (Sankey) diagram



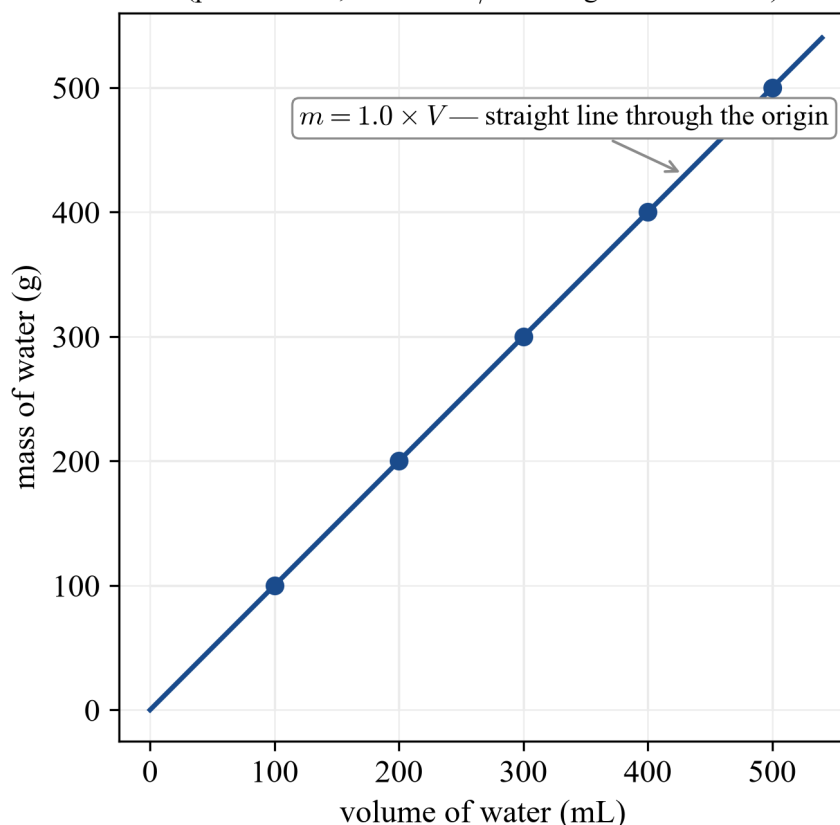
Widths are drawn to scale. $20 : 80$ simplifies to $1 : 4$ — for every 1 J of useful light, 4 J spread out as heat. (Practice example; efficiency is calculated in subtopic 14C.)

An energy-flow (Sankey) diagram with 100 J of input energy splitting into 20 J of useful light energy and 80 J of wasted heat energy, branch widths drawn to scale with the ratio 1 : 4 marked.

Reading the diagram is ratio work. The input is 100 J; the useful branch is 20 J and the wasted branch is 80 J. Energy is conserved, so the branches always add back up to the input: $20\text{ J} + 80\text{ J} = 100\text{ J}$. The ratio of useful to wasted energy is $20 : 80$, which simplifies to **1 : 4** — for every 1 J of useful light, 4 J spread out as heat. Because the widths are drawn to scale, a branch that carries four times the energy is four times as wide, and you can read ratios straight off the picture. Calculating the percentage of the input that ends up useful — the efficiency — belongs to Subtopic 14C; here the task is the ratio and the to-scale drawing.

Direct proportion. Two quantities are in **direct proportion** when one is always the same multiple of the other — when their ratio never changes. The mass of water is a clean example: measure the mass of different volumes of water and the ratio $\text{mass} \div \text{volume}$ is always 1.0 g/mL.

Direct proportion: mass of water vs volume
(practice data; the ratio $m/V = 1.0 \text{ g/mL}$ is constant)



A line graph of the mass of water against its volume, the points 100 g at 100 mL through 500 g at 500 mL falling on a straight line through the origin, labelled $m = 1.0 \times V$.

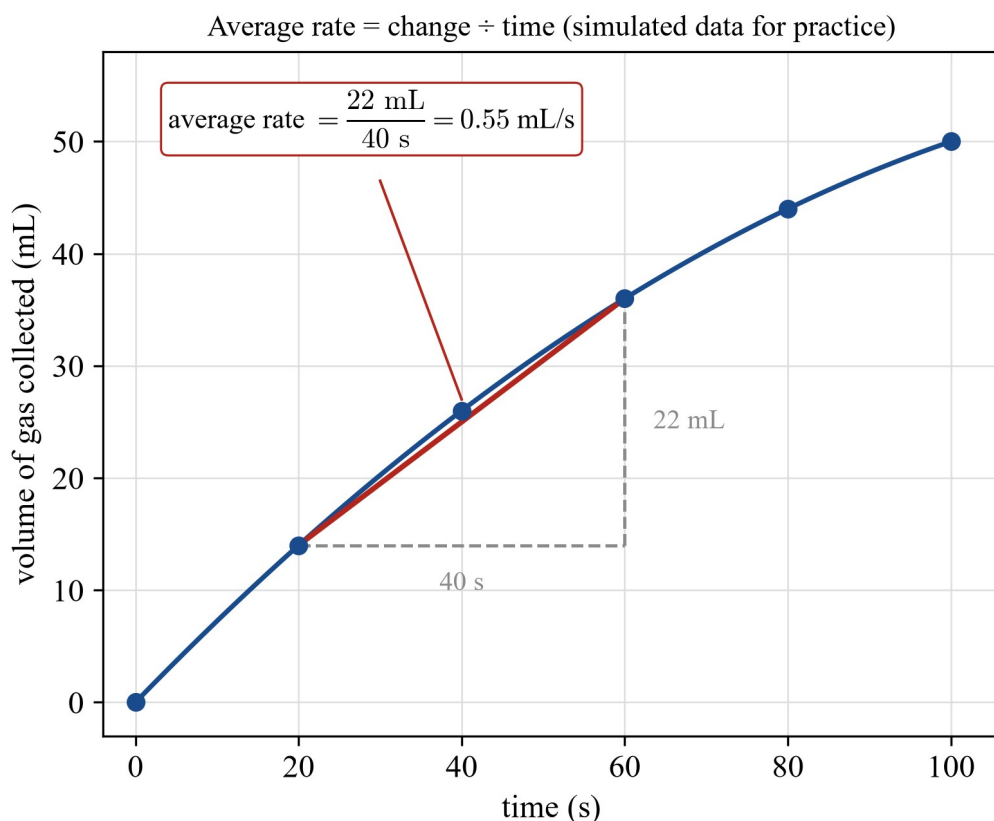
(Practice data; the ratio $m/V = 1.0 \text{ g/mL}$ is constant.)

Two tests mark direct proportion, and both are visible in the figure:

- the **ratio is constant** — here $m \div V = 1.0 \text{ g/mL}$ for every point;
- the **graph is a straight line through the origin** — zero volume means zero mass.

When two quantities are directly proportional you can write a formula of the form $m = 1.0 \times V$: one quantity equals a fixed number times the other. The fixed number is the constant ratio. It also means doubling one quantity doubles the other, tripling one triples the other, and so on. This is the Maths interlock VC2M9M05 at work.

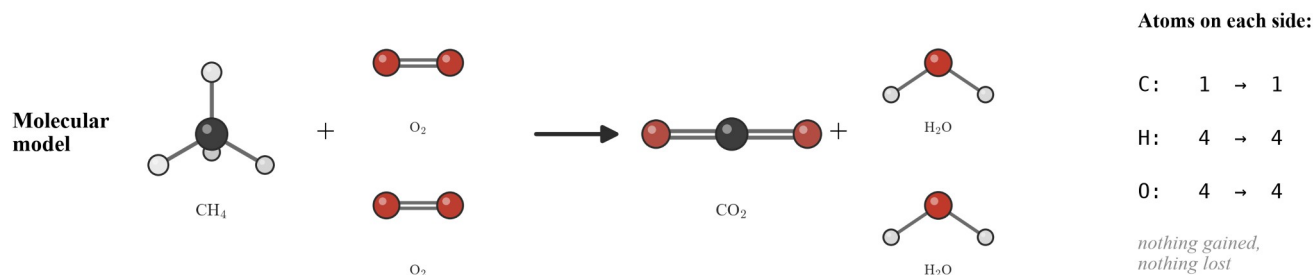
Rates. A **rate** measures how quickly one quantity changes, and it is also a ratio: change $\div$ time. Speed is the rate of change of distance; the average speed of the walker above is 80 m per minute. When the change is not steady, you can still find the **average rate** over any chosen stretch.



A curve of gas volume collected against time with the points 0, 14, 26, 36, 44 and 50 mL at 0, 20, 40, 60, 80 and 100 s; the chord between 20 s and 60 s marks a change of 22 mL in 40 s, an average rate of 0.55 mL/s.

(Simulated data for practice.) The reaction in the figure collects gas quickly at first and then more slowly. Over the stretch from 20 s to 60 s the volume rises by $36 - 14 = 22 \text{ mL}$ in $60 - 20 = 40 \text{ s}$, so the average rate there is $22 \div 40 = 0.55 \text{ mL/s}$. On the graph, an average rate is the slope of the straight line joining the two ends of the stretch — and the slope is smaller for later stretches, which is how the graph shows the reaction slowing down. The unit of a rate is always one unit divided by the other: mL/s, m/s, g/min.

One reaction, three representations (elaboration VC2S10I04 . E07). When methane burns, it reacts with oxygen to form carbon dioxide and water. The same reaction can be written three ways: as a molecular model, as a word equation, and as a symbol equation.



Word equation methane + oxygen → carbon dioxide + water

Balanced symbol equation $CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O$

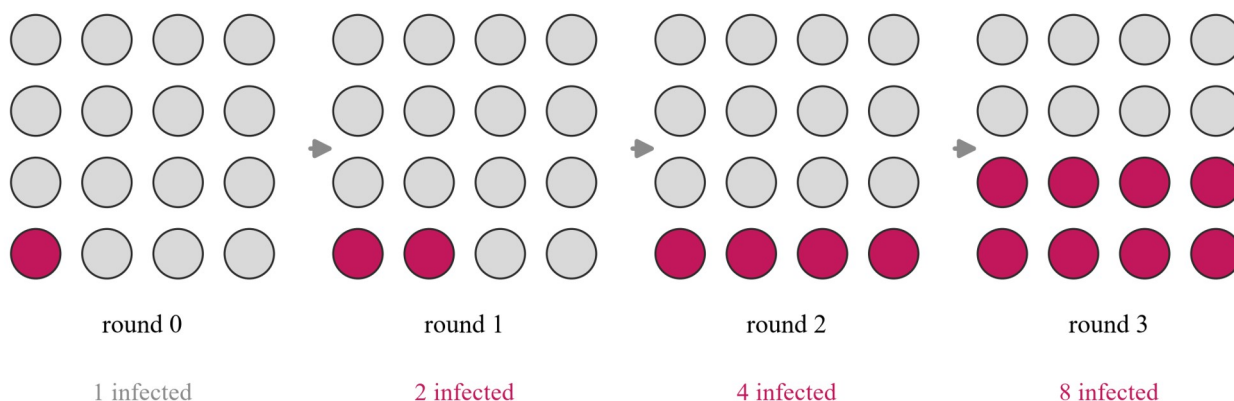
One reaction shown three ways: molecular models of methane and oxygen rearranging into carbon dioxide and water with an atom tally C 1 → 1, H 4 → 4, O 4 → 4; the word equation methane + oxygen → carbon dioxide + water; and the balanced symbol equation $CH_4 + 2O_2 \rightarrow CO_2 + 2H_2O$.

Count the atoms on each side and nothing has been gained or lost: one carbon, four hydrogens and four oxygens go in, and one carbon, four hydrogens and four oxygens come out. That is the **law of conservation of mass** — in a chemical reaction, matter is neither created nor destroyed, only rearranged. The atom tally is what the word “balanced” in a balanced symbol equation means.

Each representation shows something different. The molecular model shows atoms being rearranged — bonds broken and new ones formed — which is the picture behind the reaction. The word equation names the substances in plain language, so it is the easiest to read. The symbol equation is the most compact, and the numbers in front of the formulas show the ratios the molecules react in. In this subtopic you read and compare equations that are already balanced; balancing equations yourself is taught in Chapter 11.

Simulations and algorithms (elaborations VC2S10I04 .E05 and .E06). A **simulation** steps a situation forward in time by applying the same rule again and again. The figure below simulates a disease spreading through a group of 16 people. Its rule is kept simple on purpose: each round, every infected person passes the infection to exactly one new person.

A simulation of disease spread (simulated data for practice): each round, every infected person passes the infection to exactly one new person — 1, 2, 4, 8 infected out of 16

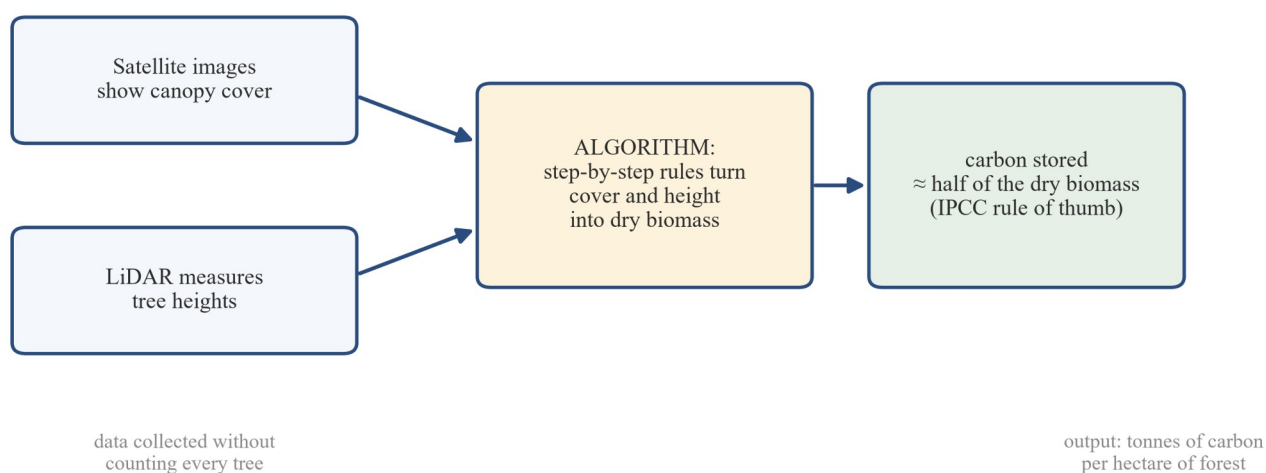


A simulation of disease spread over four rounds: the number of infected people out of 16 doubles each round, from 1 to 2 to 4 to 8.

(Simulated data for practice.) The numbers double each round — 1, 2, 4, 8 — so a small start grows fast. Real diseases do not follow a rule this tidy: people recover, some are already immune, and each infected person meets a different number of others. The value of the simulation is that it makes one idea visible — that steady doubling grows quickly — and its limits are part of the lesson too.

An **algorithm** is a step-by-step set of rules for finishing a task, and algorithms are how simulations and calculations are actually carried out. Scientists use them at real scale: to estimate how much carbon a forest stores, without counting every tree. Satellite images show how much ground is covered by tree canopy, and LiDAR (light detection and ranging) measurements show how tall the trees are. An algorithm turns cover and height into an estimate of the dry biomass — the mass of the living plant material — and the carbon stored follows from a simple ratio: about half of dry biomass is carbon. (Source: IPCC, “2006 IPCC Guidelines for National Greenhouse Gas Inventories, Volume 4: Agriculture, Forestry and Other Land Use”, <https://www.ipcc-nggip.iges.or.jp/public/2006gl/vol4.html>, retrieved 2026-08-20.)

An algorithm turns remote-sensing data into a carbon-storage estimate



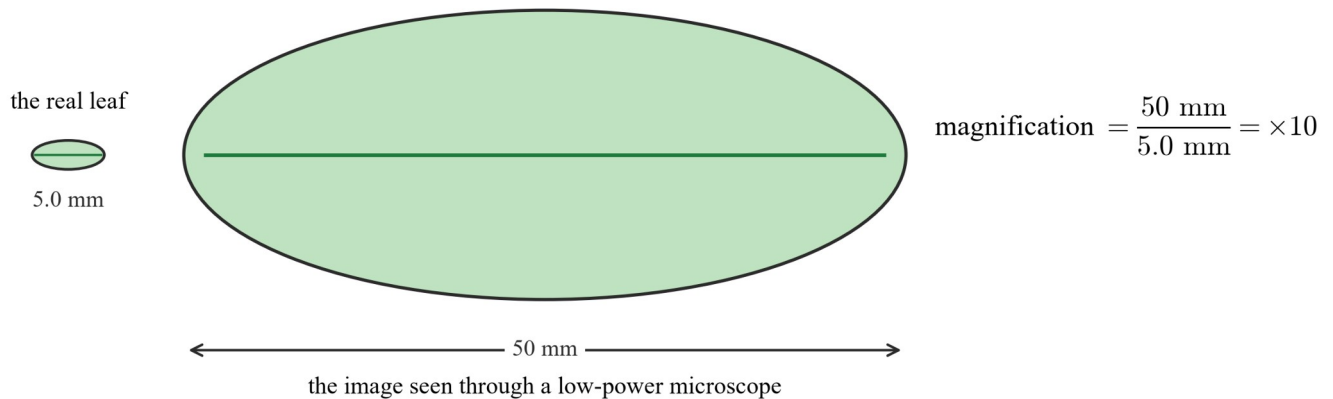
A flow diagram: satellite images of canopy cover and LiDAR tree-height measurements feed into an algorithm that converts them to dry biomass, and carbon stored is about half of the dry biomass, giving tonnes of carbon per hectare of forest.

The same shape runs through every algorithm: data in, step-by-step rules, result out. The output — tonnes of carbon per hectare — is a number a government can report and compare, and it was built by applying a ratio, the same mathematics as the rest of this subtopic.

Magnification. Images seen through a microscope or a magnifying glass are bigger than the real object. The **magnification** tells you how many times bigger, and it is another ratio:

$$\text{magnification} = \frac{\text{image size}}{\text{actual size}}$$

$$\text{Magnification} = \text{image size} \div \text{actual size}$$



A leaf 5.0 mm long beside its microscope image 50 mm long, both drawn to the same millimetre scale, with $\text{magnification} = 50 \text{ mm} \div 5.0 \text{ mm} = \times 10$.

A magnification of $\times 10$ means the image is 10 times the actual size. The two sizes must be in the same unit before you divide, and the magnification itself has no unit — it is a pure ratio. The same formula works backward: if you know the image size and the magnification, dividing gives the actual size. That is how the scale bar on a microscope photograph lets you recover the true size of what you are looking at.

Choosing a representation. Every representation in this subtopic is a way of saying the same thing in a form that suits the job. Use a table when exact values matter, a graph when the shape of the change matters, a formula when you need to predict, a sentence when you need to explain, a model when the real thing cannot be handled, and symbols when you need exact, agreed shorthand. The content description lists all of them for a reason: a scientist who can move between representations can see a problem from several sides at once.

Activities

Activity 1 — Modelling disease spread with chemistry (elaboration VC2S10I04 . E05). Each student gets a small cup holding clear sodium carbonate (washing soda) solution. Without the students watching, one cup also receives a few drops of phenolphthalein solution; that student is the one “infected” person. Phenolphthalein turns pink in sodium carbonate solution, so the pink colour marks the infection. In each round, every student pours a small amount of their cup into one other student’s cup and receives one back, then the class counts the pink cups. Record the count after each round and compare it with the doubling model 1, 2, 4, 8 from the simulation figure. You will usually see fast early growth that slows before everyone is pink — because pouring into a cup that is already pink creates no new infection, which the pure doubling model ignores. (Safety: wear safety glasses; sodium carbonate solution irritates eyes and skin, so wash splashes with water at once; the phenolphthalein solution is usually made up in alcohol, so keep it away from flames; nothing in this activity may be consumed; wash hands afterwards.)

Activity 2 — A true-to-scale Earth–Moon model. Using the 1 : 1 000 000 000 scale from the Theory, first calculate the model sizes yourself: divide Earth’s diameter (12 756 km), the Moon’s diameter (3 474 km) and the Earth–Moon distance (384 400 km) by one billion, and convert each answer to millimetres or centimetres. Then build the model: mark a dot 12.8 mm across and a dot 3.5 mm across, and place them 38.4 cm apart on a strip of paper. Compare your strip with any picture of the Earth and Moon you have seen in books or online, and write two sentences on what a true-scale picture shows that the usual pictures hide.

Activity 3 — Constant-ratio hunt. Three relationships are given below as small tables. For each one, calculate the ratio (second quantity $\div$ first quantity) for every row, decide whether the ratio is constant, and state whether the two quantities are in direct proportion. For the ones that are, write a formula and predict the next row. (a) Number of identical apples n and total mass m : $n = 2, 4, 6, 8$ with $m = 300, 600, 900, 1\,200$ g. (b) Side length s of a square and its area A : $s = 1, 2, 3, 4$ cm with $A = 1, 4, 9, 16$ cm². (c) Time t and distance d for the walker from the opening figure: $t = 1, 2, 3, 4$ min with $d = 80, 160, 240, 320$ m. Then choose one proportional case, plot its points, and check they fall on a straight line through the origin.

Worked examples

All data here are simulated data for practice unless a source is given. Values are exact unless shown with $\approx$.

Example 1 — scaffolded

A small electric motor takes in 250 J of electrical energy. 50 J ends up as useful movement and the rest spreads out as heat. (a) How much energy is wasted? (b) Write the ratio useful : wasted in its simplest form. (c) You draw an energy-flow diagram whose input band is 10 cm wide. How wide should the useful branch and the wasted branch be?

Working	Comment
(a) Wasted = $250 - 50 = 200$ J	Energy is conserved: input = useful + wasted.
(b) Useful : wasted = $50 : 200 = 1 : 4$	Divide both sides by 50.
(c) Useful branch = $(50 \div 250) \times 10 = 2.0$ cm	Branch widths are proportional to the energies they carry.
Wasted branch = $(200 \div 250) \times 10 = 8.0$ cm	Check: $2.0 + 8.0 = 10$ cm, the full input band.

The 1 : 4 ratio says the same thing as the branch widths: for every 1 J of useful movement, 4 J spread out as heat.

Example 2 — scaffolded

A museum is building a 1 : 25 scale model of a car that is 4.0 m long. How long is the model, in centimetres?

Working	Comment
Scale 1 : 25 means model = real $\div$ 25	Every length shrinks by the same factor.
$4.0 \div 25 = 0.16$ m	Divide the real length by the scale factor.
0.16 m = $0.16 \times 100 = 16$ cm	Convert metres to centimetres to give a usable answer.

Example 3 — unscaffolded

Identical steel nails are weighed in groups: 5 nails have mass 11 g, 10 nails 22 g, 15 nails 33 g and 20 nails 44 g. Show that the mass is directly proportional to the number of nails, write a formula connecting mass m (in g) and number of nails n , and use the formula to predict the mass of 40 nails.

The test for direct proportion is a constant ratio. Here $m \div n$ is $11 \div 5 = 2.2$, $22 \div 10 = 2.2$, $33 \div 15 = 2.2$ and $44 \div 20 = 2.2$ — every row gives 2.2 g per nail, so the mass is directly proportional to the number of nails, and the constant ratio is 2.2 g per nail. The formula is $m = 2.2 \times n$. For 40 nails: $m = 2.2 \times 40 = 88$ g. Plotted, the points would fall on a straight line through the origin, which is the second mark of direct proportion — and 40 nails is twice 20 nails, so the mass is twice 44 g, the same answer by doubling.

Example 4 — unscaffolded

A photograph taken through a microscope shows a pollen grain at a magnification of $\times 300$. In the photograph the grain is 36 mm across. What is the actual diameter of the grain?

Magnification = image size $\div$ actual size, so actual size = image size $\div$ magnification. Actual diameter = $36 \div 300 = 0.12$ mm. The answer is a small number because a microscope image is larger than the object, so the division goes the other way from what feels natural — always check that the actual size comes out smaller than the image.

Practice questions

Scaffolded questions

Question 1 Count the atoms of each element in each chemical formula. (a) H_2O (b) CO_2 (c) CH_4 (d) O_2

Question 2 A lamp takes in 200 J of electrical energy each second; 50 J becomes useful light and the rest spreads out as heat. (a) How much energy is wasted each second? (b) Write the ratio useful : wasted in its simplest form. (c) You draw an energy-flow diagram whose input band is 12 cm wide. How wide should the useful and wasted branches be?

Question 3 A bushwalking map has a scale of 1 : 25 000. Two campsites are 6 cm apart on the map. How far apart are they in reality? Give your answer in metres and in kilometres.

Question 4 A sample of an unknown metal has mass 54 g and volume 20 mL. (a) Calculate the ratio mass $\div$ volume for the sample. (b) The density of aluminium is 2.7 g/mL. Is the sample consistent with being aluminium? (c) If mass and volume are in direct proportion for one metal, what would the graph of mass against volume look like?

Question 5 A conveyor belt moves at a steady 1.5 m/s, so the distance d (in m) it carries a package in t seconds is $d = 1.5 \times t$. (a) Find d when $t = 8$ s. (b) Find d when $t = 20$ s. (c) Is distance directly proportional to time here? Give the reason that decides it.

Question 6 The particle model draws matter as tiny particles. (a) Describe the arrangement of the particles in a solid, a liquid and a gas, as the model shows them. (b) State one thing the particle model does not show.

Un scaffolded questions

Question 7 A swimmer crosses a 50 m pool in 40 s at a steady pace. Calculate the swimmer's speed.

Question 8 A photograph taken through a microscope shows a human cell at a magnification of $\times 400$. In the photograph the cell is 24 mm wide. What is the actual width of the cell?

Question 9 A garden centre sells a 1 : 40 scale model of a tree that is 12 m tall in reality. How tall is the model, in centimetres?

Question 10 Two quantities x and y are measured together, giving the pairs $x = 1, 2, 3, 4$ with $y = 3, 6, 12, 24$. A student claims y is directly proportional to x . Test the claim by calculating the ratio $y \div x$ for each pair, and state your conclusion.

Question 11 An electric heater takes in 120 W of electrical power; 30 W warms the room usefully and 90 W is wasted. (a) Write the ratio useful : wasted in its simplest form. (b) You draw an energy-flow diagram whose input band is 16 cm wide. How wide should the two branches be? (c) Check that energy is conserved in your diagram.

Question 12 The rate figure in the Theory shows gas collected in a reaction: 0, 14, 26, 36, 44, 50 mL at times 0, 20, 40, 60, 80, 100 s. (a) Calculate the average rate of gas production from 0 s to 40 s. (b) Calculate the average rate from 60 s to 100 s. (c) Compare the two rates and state what they say about the reaction.

Question 13 Look at the three-representations figure of burning methane in the Theory. (a) Name the reactants and the products. (b) Copy and complete an atom tally: carbon atoms on each side, hydrogen atoms on each side, oxygen atoms on each side. (c) Which of the three representations best shows that atoms are rearranged rather than destroyed? Give a reason. (d) State the law of conservation of mass in one sentence.

Question 14 The disease-spread simulation starts with 1 infected person out of 16 and doubles each round: 1, 2, 4, 8. (a) If the doubling continued, how many people would be infected at round 4 and at round 5? (b) Round 5 needs more people than the group of 16 in the figure. Explain what this says about applying the model to a small group. (c) Give two reasons a real disease would not keep doubling forever.

Question 15 A remote-sensing algorithm estimates that a forest patch has a dry biomass of 220 t per hectare. (a) Use the carbon rule of thumb from the Theory to estimate the carbon stored per hectare. (b) The forest patch covers 5.0 hectares. Estimate the total carbon stored.

Question 16 For each task below, choose the representation from this subtopic — table, graph, formula, sentence, model or symbols — that best fits it, and give one reason. (a) Record the exact masses of eight soil samples. (b)

Predict the distance the walker from the Theory covers after 7.5 min. (c) Show how the volume of gas changes over the whole 100 s of the rate figure at a glance. (d) Explain to a younger student what happens when methane burns. (e) Study the arrangement of atoms in a molecule of methane.

Answers

Question 1 (a) 2 hydrogen, 1 oxygen; (b) 1 carbon, 2 oxygen; (c) 1 carbon, 4 hydrogen; (d) 2 oxygen. **Question 2** (a) 150 J; (b) 1 : 3; (c) useful 3.0 cm, wasted 9.0 cm. **Question 3** 150 000 cm = 1 500 m = 1.5 km. **Question 4** (a) 2.7 g/mL; (b) yes — the ratio matches aluminium's density; (c) a straight line through the origin. **Question 5** (a) 12 m; (b) 30 m; (c) yes — the ratio $d \div t$ is the constant 1.5. **Question 6** (a) solid: particles packed in a regular arrangement; liquid: particles close together but irregular and able to move; gas: particles far apart and moving; (b) any one, for example the forces between particles, or that particles are not solid balls. **Question 7** 1.25 m/s. **Question 8** 0.06 mm. **Question 9** 30 cm. **Question 10** Ratios 3, 3, 4, 6 — not constant, so not directly proportional. **Question 11** (a) 1 : 3; (b) useful 4.0 cm, wasted 12.0 cm; (c) $4.0 + 12.0 = 16$ cm, the input band. **Question 12** (a) 0.65 mL/s; (b) 0.35 mL/s; (c) the rate falls, so the reaction slows down. **Question 13** (a) reactants methane and oxygen; products carbon dioxide and water; (b) C $1 \rightarrow 1$, H $4 \rightarrow 4$, O $4 \rightarrow 4$; (c) the molecular model — it shows the same atoms regrouped into new molecules; (d) in a chemical reaction, matter is neither created nor destroyed. **Question 14** (a) 16 at round 4, 32 at round 5; (b) a group of 16 runs out of new people before round 5, so pure doubling cannot hold in a small group; (c) any two, for example people recover or become immune, and infected people meet fewer new people to infect. **Question 15** (a) ≈ 110 t of carbon per hectare; (b) ≈ 550 t. **Question 16** (a) table — exact values one at a time; (b) formula — $d = 80 \times 7.5$ gives any value in one line; (c) graph — the shape of the change at a glance; (d) sentence — plain language names the substances; (e) model — the molecule is too small to see directly.

Fully-worked solutions

Question 1 The subscript after an element symbol counts the atoms of that element; no subscript means one atom. (a) H_2O : 2 hydrogen, 1 oxygen. (b) CO_2 : 1 carbon, 2 oxygen. (c) CH_4 : 1 carbon, 4 hydrogen. (d) O_2 : 2 oxygen.

Question 2 (a) Energy is conserved, so wasted = $200 - 50 = 150$ J. (b) Useful : wasted = $50 : 150$; divide both sides by 50 to get 1 : 3. (c) Branch lengths are proportional to the energies, so the branch widths are: useful = $(50 \div 200) \times 12 = 3.0$ cm, wasted = $(150 \div 200) \times 12 = 9.0$ cm. Check: $3.0 + 9.0 = 12$ cm, the input band.

Question 3 A scale of 1 : 25 000 means 1 cm on the map is 25 000 cm in reality. $6 \times 25\,000 = 150\,000$ cm. Converting: $150\,000 \text{ cm} = 1\,500 \text{ m}$ (divide by 100) = 1.5 km (divide by 1 000 again).

Question 4 (a) mass $\div$ volume = $54 \div 20 = 2.7$ g/mL. (b) The ratio equals the density of aluminium, 2.7 g/mL, so the sample is consistent with being aluminium (a single measurement like this is evidence, not proof — other metals would need to be ruled out). (c) Direct proportion gives a constant ratio, so the graph of mass against volume is a straight line through the origin.

Question 5 (a) $d = 1.5 \times 8 = 12$ m. (b) $d = 1.5 \times 20 = 30$ m. (c) Yes. The ratio $d \div t$ is 1.5 for every value of t — a constant ratio — which is the test for direct proportion; the graph would be a straight line through the origin.

Question 6 (a) In the model, a solid's particles are packed close together in a regular arrangement; a liquid's particles are still close but irregularly arranged and able to move around one another; a gas's particles are far apart and move freely. (b) The model does not show the forces between particles (and it draws particles as balls, which atoms are not). One clear statement of either point is enough.

Question 7 Speed = distance $\div$ time = $50 \div 40 = 1.25$ m/s.

Question 8 Actual size = image size $\div$ magnification = $24 \div 400 = 0.06$ mm. The actual cell is smaller than its image, as it must be at a magnification greater than $\times 1$.

Question 9 Model = real $\div 40 = 12 \div 40 = 0.30$ m. Converting, 0.30 m = 30 cm.

Question 10 The ratios $y \div x$ are $3 \div 1 = 3$, $6 \div 2 = 3$, $12 \div 3 = 4$ and $24 \div 4 = 6$. For direct proportion the ratio must be the same for every pair; here it changes from 3 to 4 to 6, so the claim is false. (The y -values are doubling when x doubles only up to the second pair — after that they grow faster than direct proportion.)

Question 11 (a) Useful : wasted = $30 : 90 = 1 : 3$. (b) Branch widths are proportional to the powers: useful = $(30 \div 120) \times 16 = 4.0$ cm; wasted = $(90 \div 120) \times 16 = 12.0$ cm. (c) The two branches add to $4.0 + 12.0 = 16$ cm, exactly the input band, so the diagram conserves energy: $30 \text{ W} + 90 \text{ W} = 120 \text{ W}$.

Question 12 Average rate = change in volume $\div$ time taken. (a) From 0 s to 40 s the volume rises $26 - 0 = 26$ mL in 40 s: $26 \div 40 = 0.65$ mL/s. (b) From 60 s to 100 s the volume rises $50 - 36 = 14$ mL in 40 s: $14 \div 40 = 0.35$ mL/s. (c) The later rate is about half the earlier one, so the reaction is slowing down as time passes.

Question 13 (a) The reactants, on the left of the arrow, are methane and oxygen; the products, on the right, are carbon dioxide and water. (b) Carbon: $1 \rightarrow 1$; hydrogen: $4 \rightarrow 4$; oxygen: $4 \rightarrow 4$ (two O_2 molecules give four oxygen atoms, and CO_2 plus two H_2O contain four). (c) The molecular model — it shows the same atoms before and after, regrouped into different molecules, so you can see that nothing disappeared. The word and symbol equations say the same thing but you have to count the atoms yourself. (d) In a chemical reaction, matter is neither created nor destroyed — the total mass of the products equals the total mass of the reactants.

Question 14 (a) Doubling continues: round 4 gives 16 infected, round 5 gives 32. (b) The group in the figure has only 16 people, so by round 4 everyone is infected and round 5 cannot happen at all — the model needs a bigger population before the doubling pattern is even possible. (c) Any two of: infected people recover or die and stop spreading it; some people are immune; an infected person runs out of uninfected people nearby to infect; real contact is not one new person per round.

Question 15 (a) Carbon stored $\approx$ half of the dry biomass: $220 \div 2 \approx 110$ t of carbon per hectare. (b) Total $\approx 110 \times 5.0 \approx 550$ t. Both answers keep $\approx$ because the “half” rule is itself an approximation.

Question 16 (a) Table — the task wants exact values recorded one at a time. (b) Formula — $d = 80 \times t$ predicts any time, and 7.5 min sits between the table’s whole minutes: $d = 80 \times 7.5 = 600$ m. (c) Graph — the whole shape of the change is visible at a glance, including the slowing down. (d) Sentence — plain language names the substances for someone with no symbols yet; the word equation of the reaction is exactly this sentence form. (e) Model — the molecule is far too small to see, so a ball-and-stick model stands in for it. Every choice earns its reason; other pairings are acceptable if the reason fits.

Where this leads: balancing symbol equations is taught in Chapter 11 (Subtopic 11B) — here they are read and compared, not balanced; the law of conservation of mass is treated fully in Subtopic 11A. Efficiency percentages and the full analysis of energy-flow diagrams belong to Subtopic 14C. Rearranging formulas (VC2M10A05) arrives in Subtopic 15B with Newton’s second law; this subtopic substitutes into formulas that are already set up. Patterns, trends and relationships in data belong to the neighbouring Subtopic 4C. Mole-ratio calculations are VCE-only and are not taught here (TOC D9).

Patterns, trends, relationships and anomalies — including correlation versus causation

Chapter 4 — Processing, modelling and analysing (Science, Levels 9 and 10)

Subtopic 4C · VC2 Science Inquiry Skills — Processing, modelling and analysing

CD code: VC2S10I05 (delivered whole by this subtopic) **Content description (word for word):**

“information and processed data can be analysed and compared to identify and explain qualitative and quantitative patterns, trends, relationships and anomalies” **Elaboration ids drawn on:** VC2S10I05.E01 · VC2S10I05.E02 · VC2S10I05.E03 · VC2S10I05.E04 · VC2S10I05.E05 · VC2S10I05.E06

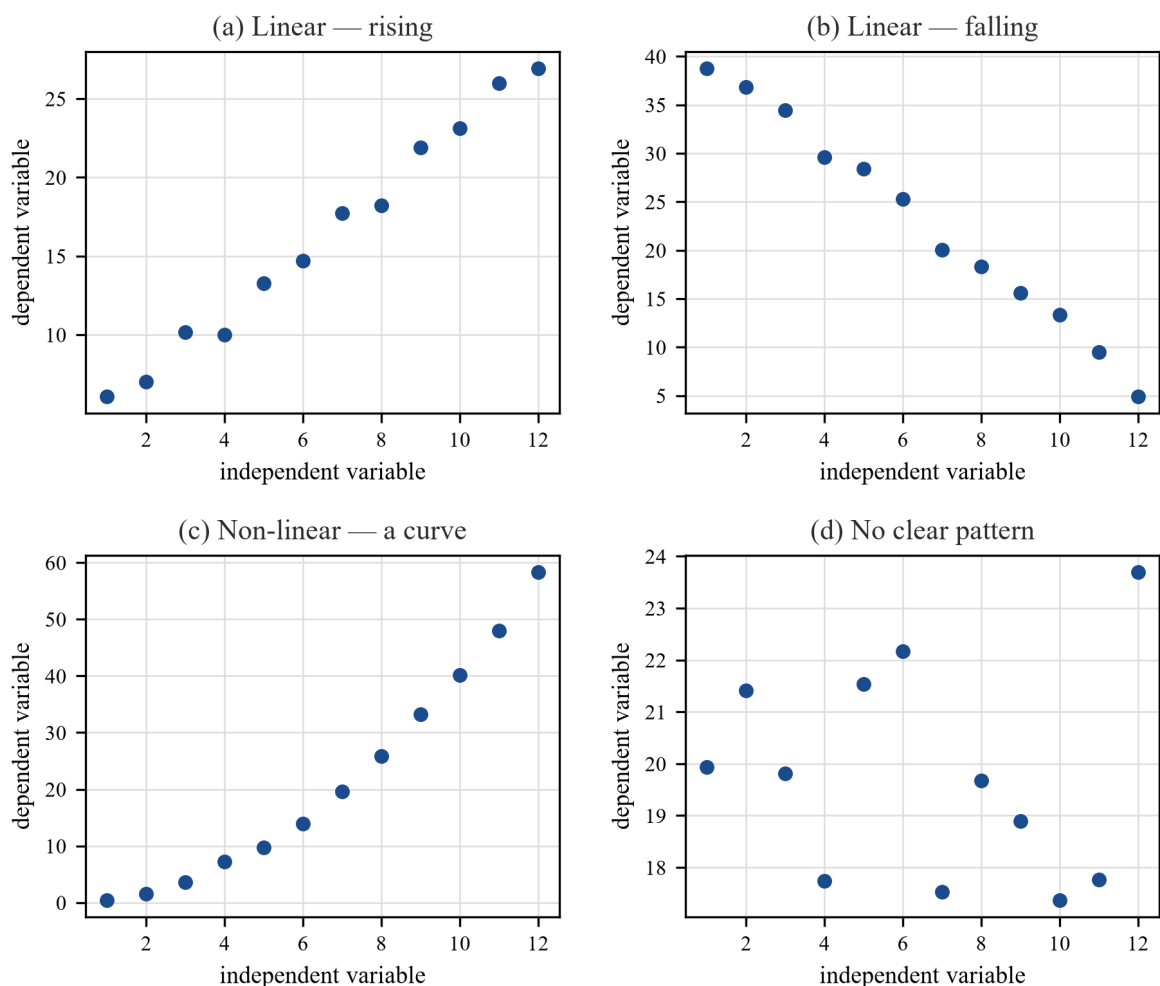
Achievement standard (extract served): “They analyse and compare a variety of data and information to identify and explain qualitative and quantitative patterns, trends, relationships, assumptions and anomalies.” **Maths interlock (D11):** VC2M10ST02 — construct scatterplots and consider a line of good fit; comment on the association between the two numerical variables in terms of strength, direction and

linearity. This subtopic uses that language to describe scientific data.

Theory

Analysis — what the word means here. By the time you reach this subtopic, your data has already been organised and processed (4A) and displayed in a suitable way (4B). **Analysis** is the next step: looking at the data and asking *what is it telling me?* You are looking for four things — **patterns** (repeated or regular behaviour in the data), **trends** (the overall direction the data moves in), **relationships** (how one variable changes when another changes) and **anomalies** (results that do not fit the pattern). This subtopic builds a method for finding and explaining each of them.

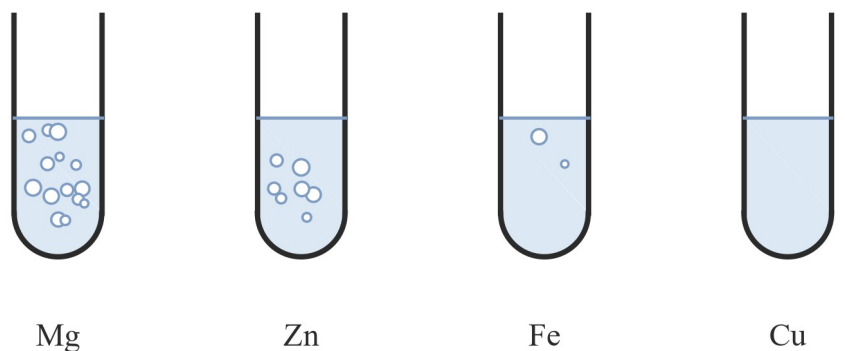
Qualitative and quantitative patterns. Data and patterns come in two kinds. **Quantitative** data is numerical — masses, times, temperatures, counts — so a quantitative pattern is one you can state with numbers, like “the mass halves every 10 minutes”. **Qualitative** data is descriptive — colours, textures, what you saw, heard or smelled — so a qualitative pattern is one you state in words.



Four scatterplots of simulated data: (a) points rising along a straight line, (b) points falling along a straight line, (c) points following an upward-bending curve, (d) points scattered with no clear pattern

The four panels above show the kinds of quantitative pattern you will meet: a **linear** (straight-line) pattern rising, a linear pattern falling, a **non-linear** pattern (a curve), and no clear pattern at all.

equal-sized pieces of metal in equal volumes of dilute hydrochloric acid



bubbles fast

bubbles steadily

very few bubbles

no bubbles

Four test tubes of dilute hydrochloric acid (a weak-acid solution) with equal-sized pieces of metal: magnesium bubbles fast, zinc bubbles steadily, iron shows very few bubbles, copper shows no bubbles

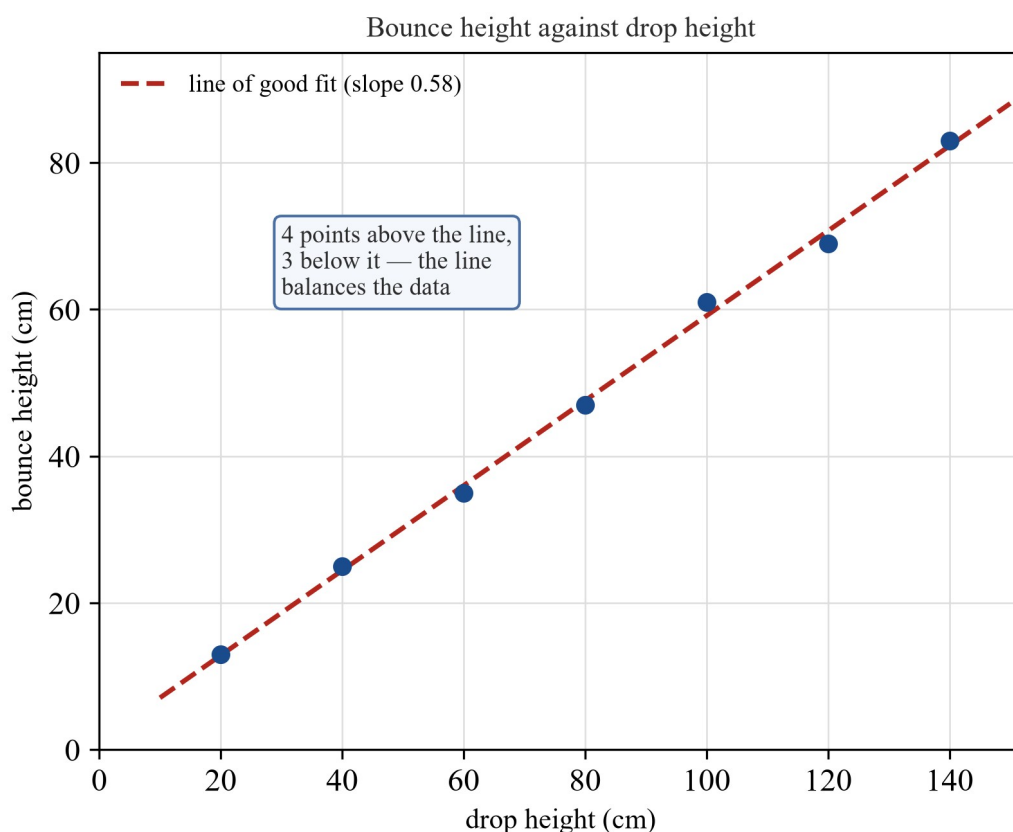
The test tubes above show a qualitative pattern. Equal-sized pieces of four metals are placed in equal volumes of dilute hydrochloric acid, and the observations are recorded in words: magnesium bubbles fast, zinc bubbles steadily,

iron shows very few bubbles, and copper shows none. There are no numbers anywhere, yet there is a clear pattern: **the metals can be put in an order, from most to least vigorous reaction**. A pattern does not need numbers to be real.

Describing a quantitative pattern: direction, shape, strength. When both variables are numerical, the graph of the paired measurements is called a **scatterplot** — one point for each pair of values. To describe the pattern in a scatterplot, state three things (this is exactly the language of VC2M10ST02):

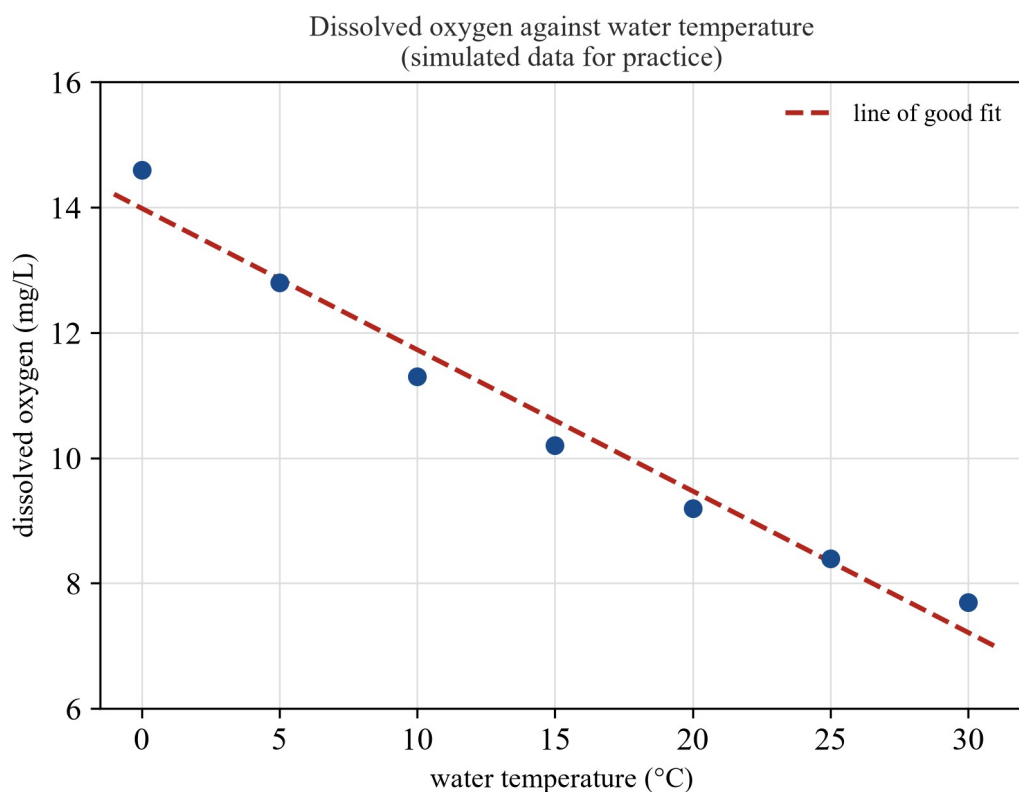
- **Direction** — is the pattern rising (as one variable increases, so does the other) or falling (as one increases, the other decreases)?
- **Shape** — is the pattern linear (the points sit close to a straight line) or non-linear (they follow a curve)?
- **Strength** — do the points sit tightly along the line or curve (a strong pattern) or are they widely scattered (a weak pattern)?

Where a pattern is roughly linear, a **line of good fit** is drawn: a straight line placed through the points so that it balances them, with about as many points above the line as below it.



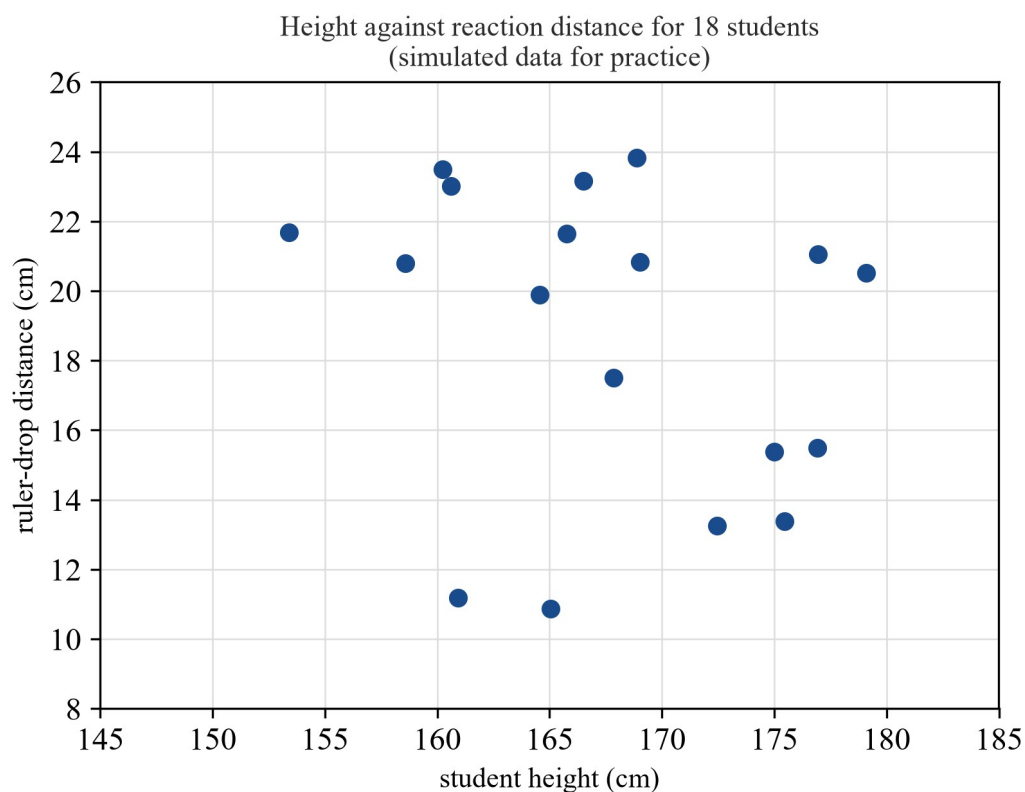
Scatterplot of bounce height against drop height for a dropped ball: seven points lying close to a dashed line of good fit with slope 0.58, four points above the line and three below it

In the bounce-height scatterplot above, the pattern is **strong** (the points sit very close to the line), **rising**, and **linear**. The line of good fit balances the data — four points above it, three below.



Scatterplot of dissolved oxygen against water temperature: points falling smoothly from about 14.6 mg/L at 0 degrees C to about 7.7 mg/L at 30 degrees C, with a falling line of good fit

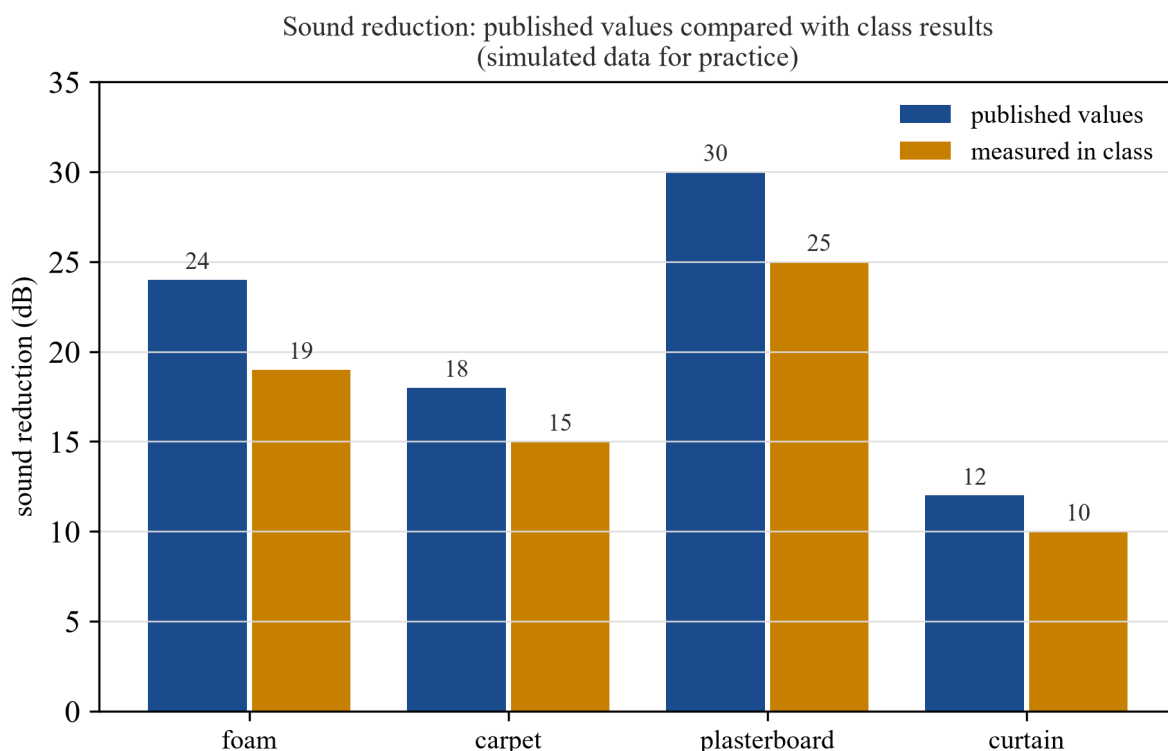
The dissolved-oxygen graph shows a **strong, falling** pattern: the warmer the water, the less oxygen stays dissolved in it. The pattern is close to linear, though it curves a little — the falls are slightly smaller at the hot end. Noticing that kind of detail (“close to linear, but curving”) is part of describing shape honestly.



Scatterplot of student height against ruler-drop reaction distance for 18 students: points scattered all over the graph with no pattern

And the height-and-reaction graph shows **no clear pattern** at all: the points are scattered, and knowing a student's height tells you nothing about their reaction distance. "There is no relationship" is a perfectly good — and important — conclusion.

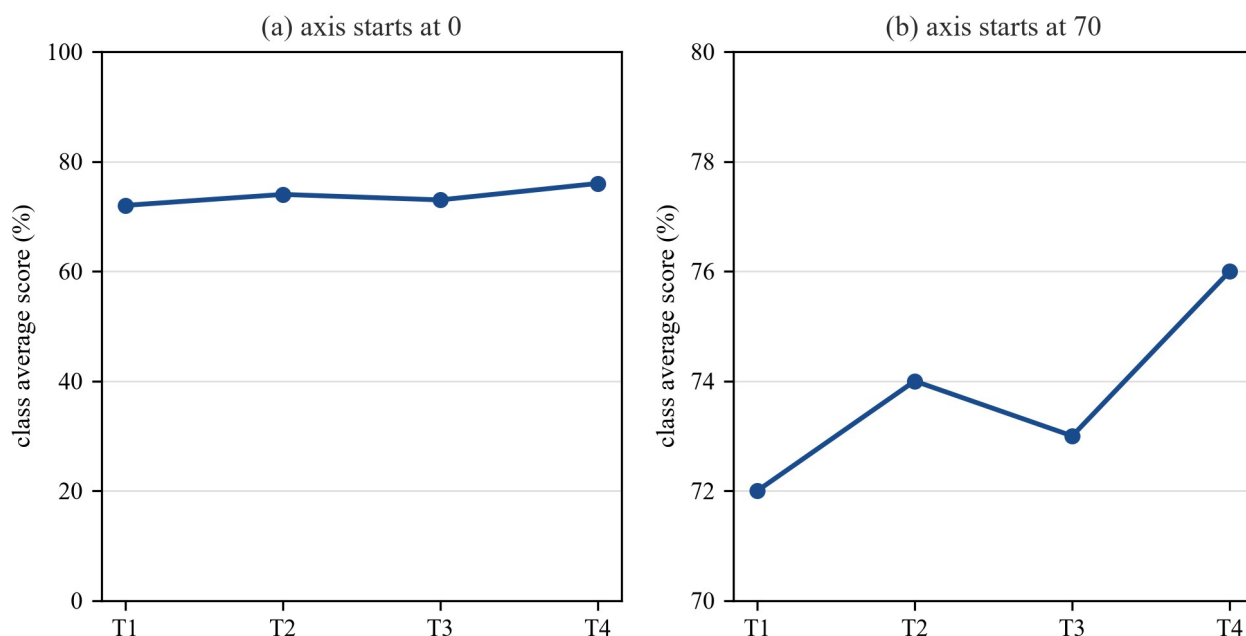
Comparing two sets of data (elaboration VC2S10I05.E01). Analysis often means comparing one dataset with another — for example, comparing your class results with published values from a manufacturer or a handbook.



Bar chart comparing published sound-reduction values with class measurements for four materials: foam 24 vs 19 dB, carpet 18 vs 15 dB, plasterboard 30 vs 25 dB, curtain 12 vs 10 dB (simulated data for practice)

In the sound-reduction chart above, a class tested how well four materials block sound and compared the results with published values. The analysis has three parts. First, **the ranking is the same in both datasets**: plasterboard reduces the most sound, then foam, then carpet, then curtain. Second, **every measured value is lower than the published value**, by 2–5 dB. Third, **explain the pattern**: published values are usually higher than experimentally determined levels because they are measured in ideal laboratory conditions — sealed edges, exact thicknesses, perfect equipment — while a classroom setup lets some sound leak around the material. Same pattern, smaller numbers: that is the relationship to explain.

The same data can give different impressions (elaboration VC2S10I05.E02). How data is organised and displayed changes what a reader sees in it — sometimes so much that two people draw different conclusions from the *same numbers*.

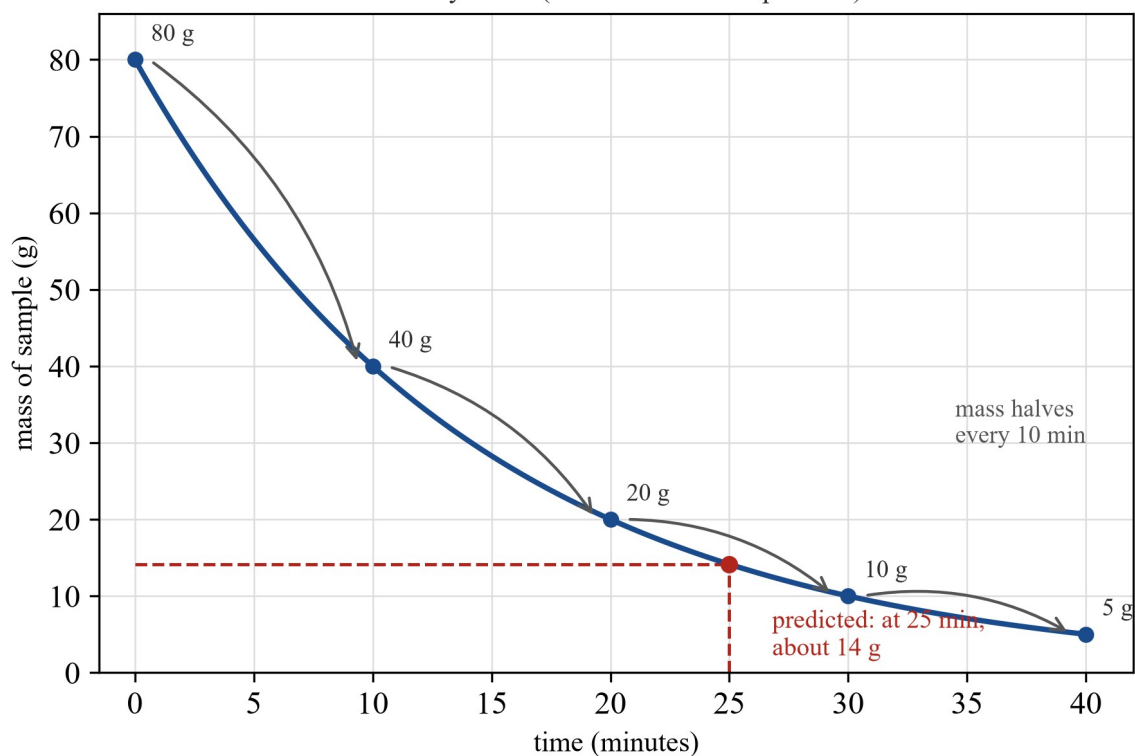


Two line graphs of the same four class average scores, T1 to T4 (72, 74, 73, 76 percent): panel (a) with the axis starting at 0 shows an almost flat line; panel (b) with the axis starting at 70 shows a dramatic rise and fall

Both panels above plot exactly the same four class average scores. In panel (a), with the axis starting at 0, the scores look almost flat — “nothing much changed”. In panel (b), with the axis starting at 70, the same changes look dramatic — “scores rose and fell sharply”. Neither panel is lying, yet each gives a different impression, and a reader could honestly quote either one. The lesson for analysis is two-sided: when you **read** a graph, always check where the axis starts and how the scale runs before you believe the impression; and when you **draw** a graph, choose the display on purpose, knowing it shapes what your reader sees. The same idea applies to organising data: grouping results by term, by topic, or by student can each bring out a different pattern in the same numbers.

Using a pattern to make predictions (elaboration VC2S10I05 .E03). A well-established pattern lets you predict values you have not measured. Radioactive decay is the classic case: a radioactive sample loses mass in a regular pattern described by its **half-life**, the time taken for the mass to halve.

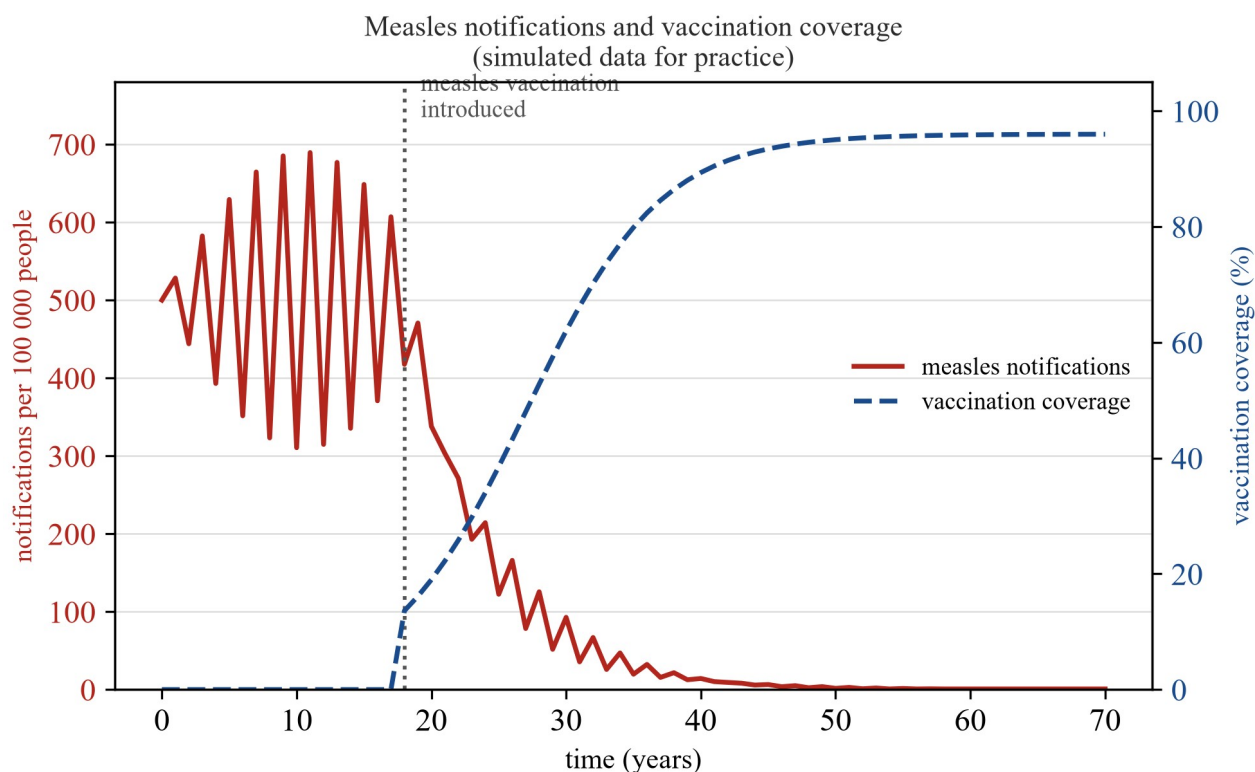
A decay curve (simulated data for practice)



Decay curve: mass of a radioactive sample falling from 80 g to 5 g over 40 minutes, halving every 10 minutes, with a prediction marked at 25 minutes of about 14 g (simulated data for practice)

In the decay curve above, the mass halves every 10 minutes: 80 g $\rightarrow$ 40 g $\rightarrow$ 20 g $\rightarrow$ 10 g $\rightarrow$ 5 g. The pattern is non-linear — a falling curve that flattens out, because the same *fraction* (one half) is lost each 10 minutes, not the same *amount*. Once the pattern is established, the curve can be used to predict: at 25 minutes the mass is about 14 g, between the 20 g at 20 minutes and the 10 g at 30 minutes, and extending the pattern one more half-life to 50 minutes predicts about 2.5 g. A prediction inside the measured range is fairly safe; a prediction far beyond the last measurement assumes the pattern continues, and that assumption should be stated.

Trends in health data (elaboration VC2S10I05.E04). Vaccination records show some of the clearest trends in public data.



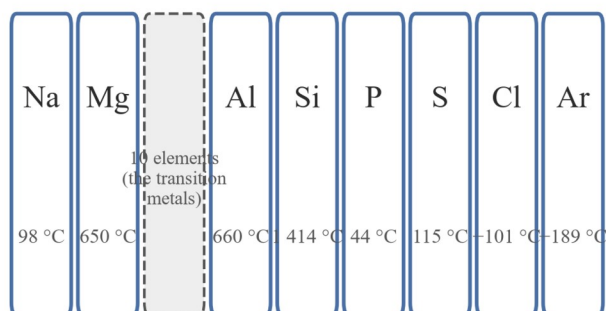
Line graph over 70 simulated years: yearly measles notifications oscillating around 500 per year until a vaccination programme is introduced at year 18, after which notifications collapse towards zero while vaccination coverage rises to about 96 percent (simulated illustrative data)

In the (simulated) record above, yearly measles notifications oscillate around 500 per year — the oscillation is itself a pattern, the familiar two-to-three-year rise and fall of outbreaks — and then, after the vaccination programme is introduced at year 18, notifications collapse towards zero while coverage climbs to about 96%. The analysis names the trend (high and oscillating, then collapsing), the relationship (as coverage rises, notifications fall), and the explanation: vaccination gives **immunity**, so the disease can no longer find enough people to infect. The same story appears in real records for measles, polio, tetanus, chickenpox and whooping cough, and in recent data for influenza, HPV and COVID-19 — which is why public-health organisations track vaccination coverage and disease notifications together.

One pattern, two representations (elaboration VC2S10I05.E05). The same data can be displayed in different ways, and each way has merits and limitations.

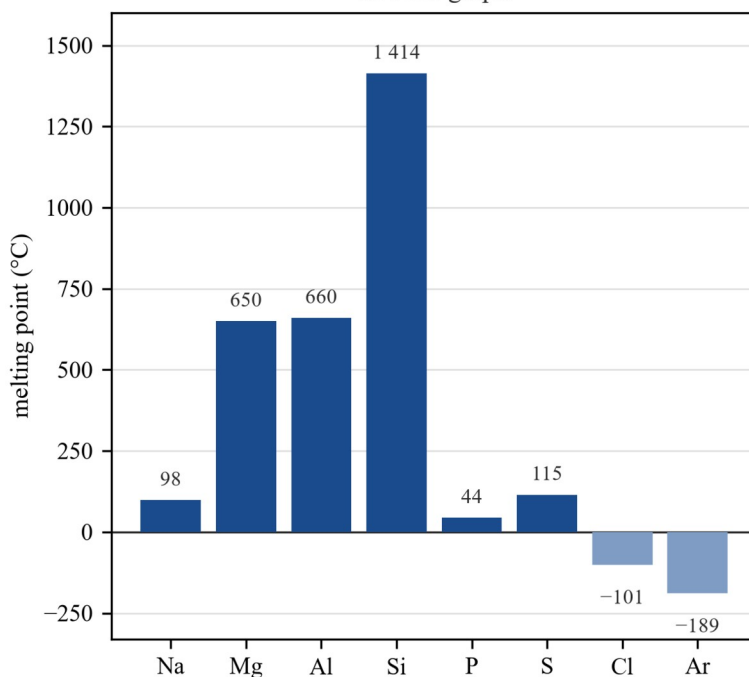
One pattern, two representations — melting points of the period-3 elements
(rounded reference values for classroom use)

as part of the periodic table layout



the pattern is hidden in the layout

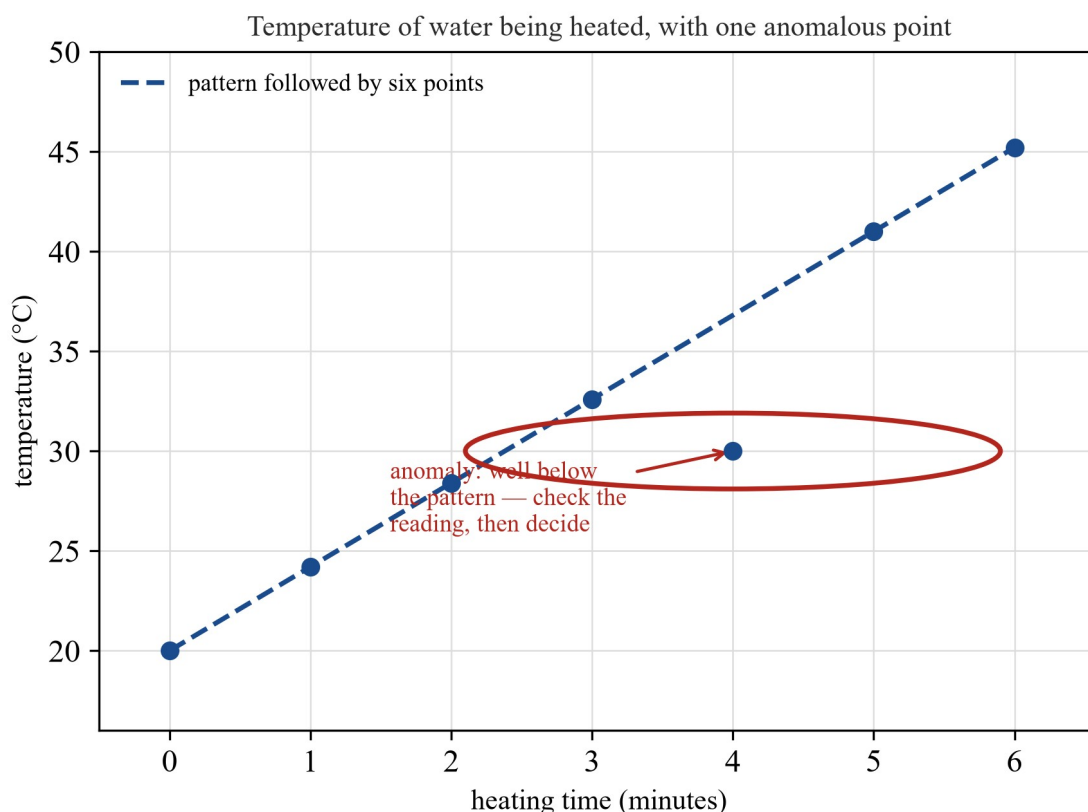
as a bar graph



Melting points of the period-3 elements shown two ways: left, as element tiles in the periodic-table layout (Na 98, Mg 650, Al 660, Si 1 414, P 44, S 115, Cl minus 101, Ar minus 189 degrees C) where the pattern is hidden; right, as a bar graph where the rise to silicon and fall after is obvious

The melting points of the period-3 elements (sodium to argon, with the ten transition elements sitting in the gap) are shown above in two representations, using rounded reference values. In the periodic-table layout, the numbers are right next to each element's symbol — the merit is that position and element family are visible — but the melting-point pattern is hidden in the layout. As a bar graph, the pattern jumps out: melting point **rises from sodium to a peak at silicon (1 414 °C), then falls sharply**, with chlorine and argon melting below 0 °C (they are gases at room temperature). The graph's limitation is the reverse: it shows the trend beautifully but says nothing about where each element sits in the table. A good analyst chooses the representation that shows the pattern the question is asking about — and can state what the other representation would have shown.

Anomalies. An **anomaly** is a result that does not fit the pattern shown by the other data. Anomalies are easy to miss if you only look at numbers in a table, and obvious once the data is graphed — one more reason to display data well.

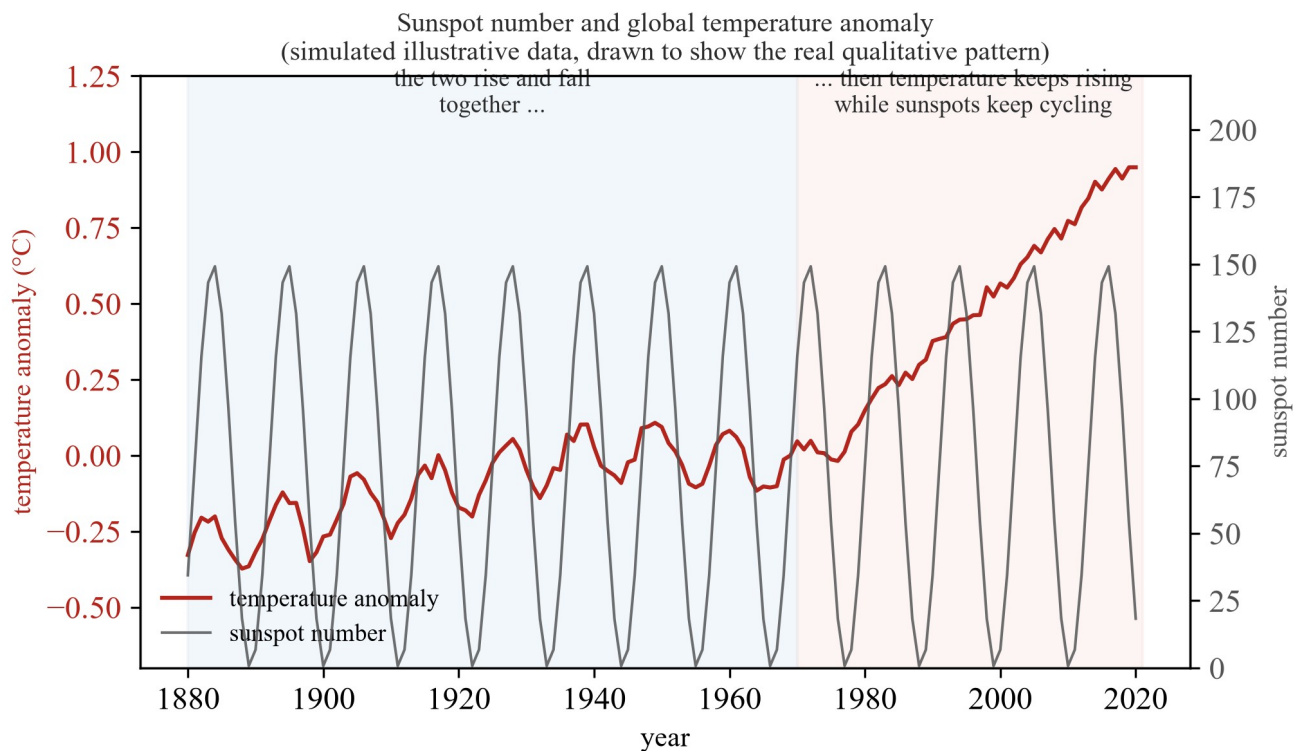


Graph of the temperature of water being heated, rising 4.2 degrees C per minute from 20 degrees C, with six points on the pattern line and one anomalous point at 4 minutes (30 degrees C instead of about 37 degrees C) circled in red

In the heating data above, six of the seven points follow a clean linear pattern (about 4.2 °C per minute), but the point at 4 minutes sits well below the line. The method for an anomaly is: **check, repeat, then decide**. First check for a mistake — a misread thermometer, a timer pressed late, the burner turned down. If you can, repeat that one measurement. Only then decide: if the repeat gives a value on the pattern, the original was an error and can be set aside (with a note saying so); if the repeat confirms the odd value, then it is not an error at all but a discovery — the pattern is telling you something new. An anomaly is never just deleted in silence.

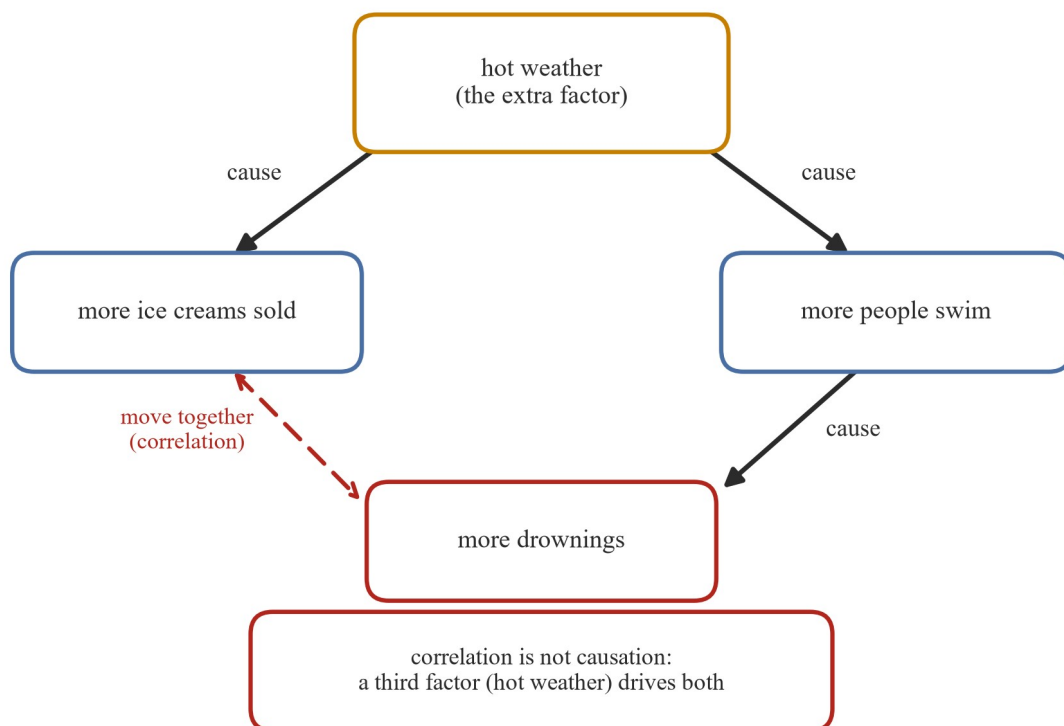
Correlation and causation (elaboration VC2S10I05 .E06). Two variables show a **correlation** when they move together — when one changes, the other tends to change in a predictable way. A relationship is **causal** when one variable actually produces the change in the other. Every causal relationship shows correlation, but a correlation on its own does not prove causation, and confusing the two is the most common error in data analysis.

A causal case: bacterial infection and disease. For most of history, stomach ulcers were blamed on stress and diet, because stress and ulcer onset correlated. In 1984, Barry Marshall and Robin Warren reported curved bacteria in the stomachs of patients with gastritis (an inflammation of the stomach lining) and peptic ulcers (Marshall & Warren, “Unidentified curved bacilli in the stomach of patients with gastritis and peptic ulceration”, *The Lancet*, 1984). Correlation alone would not have settled the question; what established **causation** was the extra evidence — patients whose bacteria were killed with antibiotics recovered, and swallowing the bacteria produced the same inflammation. The bacteria *caused* the disease: changing one variable changed the other.



Two-axis line graph from 1880 to 2020 (simulated illustrative data, drawn to show the real qualitative pattern): sunspot number cycling every 11 years and temperature anomaly rising and falling with it until about 1970, after which temperature keeps rising while sunspots keep cycling unchanged

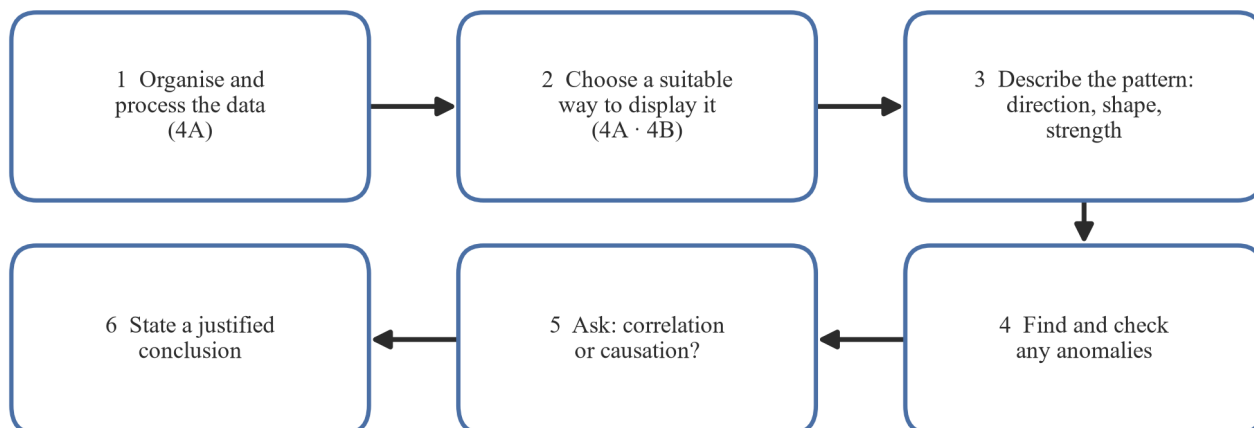
A correlation-only case: sunspot activity and climate change. Sunspots rise and fall on an 11-year cycle, and for much of the record above the two lines rise and fall together — a correlation. But from about 1970 the lines **diverge**: temperature keeps climbing while the sunspot cycle continues exactly as before. If sunspot activity were the *cause* of the recent warming, temperature would still be tracking the cycle. The correlation breaks exactly where a causal relationship would have to hold, so sunspot activity cannot be the main cause of recent climate change.



Flow diagram: hot weather (the extra factor) causes both more ice creams sold and more people swim; more people swim causes more drownings; a dashed double arrow between more ice creams sold and more drownings is labelled move together (correlation); caption reads correlation is not causation, a third factor (hot weather) drives both

A third kind of case: a correlation produced by a **third factor**. Ice-cream sales and drownings rise and fall together each year — a strong correlation — but eating ice cream does not cause drowning. A third factor, hot weather, causes *both*: in hot weather more ice creams are sold and more people swim, and more swimming leads to more drownings. When you see a correlation, always ask: is this causal, is it the other way around, or is a third factor driving both?

Pulling it together — the analysis method. Every part of this subtopic joins into one method, shown in the flow diagram below.



Flow diagram of the six-step analysis method: 1 organise and process the data (4A), 2 choose a suitable way to display it (4A and 4B), 3 describe the pattern (direction, shape, strength), 4 find and check any anomalies, 5 ask correlation or causation, 6 state a justified conclusion

A strong analysis answer in this subtopic does the following, in order:

- **Organise and display** the data suitably (4A, 4B) — and check the axes and scale before trusting an impression.
- **Describe the pattern** in direction, shape and strength, quoting actual values from the data (“the mass halved every 10 minutes”, not just “it went down”).
- **Compare** datasets where there are two — your results against published values, one group against another — and state the relationship between them.
- **Find and check anomalies**, and say what you did about them.
- **Ask correlation or causation** before claiming one thing caused another.
- **State a conclusion** that answers the question asked, justified by the data you have just quoted.

Activities

Activity 1 — Bounce height and drop height. Work in pairs with a tennis ball, a metre ruler and a flat floor. Drop the ball from 20 cm and measure the bounce height; repeat three times and average. Repeat for drop heights of 40, 60, 80, 100, 120 and 140 cm. Plot your averages on a scatterplot (drop height on the horizontal axis), draw a line of good fit, and describe the pattern in direction, shape and strength. Use your line to predict the bounce from a 90 cm drop, then test the prediction. Does the prediction sit on, above or below your measured value?

Activity 2 — A qualitative pattern in metals. *Teacher demonstration.* Your teacher adds equal-sized pieces of magnesium, zinc, iron and copper to equal volumes of dilute hydrochloric acid. Record what you observe for each metal in words (bubbling, warmth, change in the metal). Put the metals in order from most to least vigorous, describe the qualitative pattern in one sentence, and then suggest one quantitative measurement (for example, the volume of gas collected in one minute) that would turn your qualitative pattern into a quantitative one.

Activity 3 — Published versus measured sound reduction. Place a phone playing a steady sound inside a box, and measure the sound level outside the box with a sound-meter app. Wrap the box in a layer of foam, then carpet, then plasterboard offcut, then curtain fabric, recording the sound level each time. Work out the reduction each material

gives (no-wrap level minus wrapped level), and compare your ranking and values with the published-style chart in this subtopic. Is your ranking the same? Are your values higher or lower than the published ones, and what conditions in a laboratory would explain the difference?

Worked examples

Datasets in this section are simulated data for practice unless labelled as reference values.

Example 1 — scaffolded

A class measured the sound reduction (in dB) of four materials and compared the results with published values:

Material	Published (dB)	Measured (dB)
Foam	24	19
Carpet	18	15
Plasterboard	30	25
Curtain	12	10

- Describe the pattern in the published values.
- Calculate the difference (published – measured) for each material, and describe the relationship between the two datasets.
- Suggest a reason for the relationship you found in (b).

Working	Comment
(a) Published ranking: plasterboard 30 > foam 24 > carpet 18 > curtain 12	Read the pattern in one dataset first.
(b) Differences: $24 - 19 = 5$; $18 - 15 = 3$; $30 - 25 = 5$; $12 - 10 = 2$ dB	One difference per material.
The ranking is identical in both datasets, and every measured value is 2–5 dB below the published value	State the relationship between the datasets, not just the numbers.
(c) Published values come from ideal laboratory conditions (sealed edges, exact thicknesses); a classroom setup leaks sound around the material	Explain the relationship — analysis is “identify <i>and</i> explain”.

The class data follows the published pattern but sits 2–5 dB below it, because classroom conditions are not the sealed laboratory conditions of the published tests.

Example 2 — scaffolded

A sample of a radioactive material is measured every 10 minutes:

Time (min)	0	10	20	30	40
Mass (g)	80	40	20	10	5

- Describe the pattern in the data.
- Use the pattern to predict the mass at 25 minutes.
- Predict the mass at 50 minutes, and state the assumption your prediction makes.

Working	Comment
(a) $80 \rightarrow 40 \rightarrow 20 \rightarrow 10 \rightarrow 5$: the mass halves every 10 minutes	A pattern stated with numbers, not “it gets smaller”.
The graph is a falling curve that flattens — non-linear	Halving is a constant fraction, not a constant amount.
(b) At 20 min the mass is 20 g; at 30 min, 10 g. The smooth curve at 25 min gives about 14 g	A prediction between measured values reads from the pattern.
(c) 40 min = 5 g, so 50 min = $5 \div 2 = 2.5$ g	One more half-life.

Working	Comment
Assumption: the halving pattern keeps going after the last measurement	State the assumption behind predicting beyond the data.

The pattern is exponential decay with a half-life of 10 minutes; it predicts about 14 g at 25 minutes and 2.5 g at 50 minutes, assuming the pattern continues.

Example 3 — unscaffolded

The dissolved-oxygen data of the theory graph are repeated here.

Water temperature (°C)	0	5	10	15	20	25	30
Dissolved oxygen (mg/L)	14.6	12.8	11.3	10.2	9.2	8.4	7.7

Describe the association between the two variables in terms of strength, direction and linearity, and estimate the dissolved oxygen at 10 °C and at 12.5 °C.

The direction is **falling**: as water temperature increases, dissolved oxygen decreases. The strength is **strong**: the points sit tightly along a smooth line with no outliers. The shape is **close to linear but slightly curved** — the falls between readings shrink from 1.8 mg/L (0 → 5 °C) to 0.7 mg/L (25 → 30 °C), so a straight line would only be an approximation. Reading from the data, the dissolved oxygen at 10 °C is **11.3 mg/L**, and at 12.5 °C, halfway between the 11.3 at 10 °C and the 10.2 at 15 °C, it is about **10.8 mg/L**.

Example 4 — unscaffolded

For each pair of variables, decide whether the relationship is best described as causal or as correlation only, and justify each decision. (a) Vaccination coverage and measles notifications. (b) Sunspot number and global temperature, 1970–2020. (c) Ice-cream sales and drownings.

- (a) **Causal.** Coverage and notifications move in opposite directions, but the case rests on the mechanism and the test: vaccination produces immunity, so the disease cannot spread, and when coverage falls (as it has in some places and times) outbreaks return. Changing one variable changes the other — the signature of causation.
- (b) **Correlation only.** Over 1970–2020 the sunspot number keeps cycling with the same 11-year period and the same peak size, while temperature rises steadily. A cause must be followed by its effect; here the supposed effect has stopped following, so the relationship cannot be causal.
- (c) **Correlation only, via a third factor.** The two rise and fall together season after season, but neither produces the other: hot weather causes both more ice-cream sales and more swimming, and more swimming leads to more drownings. The correlation is real; the causation runs through the third factor.

Practice questions

Scaffolded questions

Question 1 Look at the four-panel figure of pattern types in the theory. (a) Which panel shows a linear, rising pattern? (b) Which panel shows a non-linear pattern? How can you tell it is not linear? (c) Which panel shows no clear pattern, and what would you conclude about the relationship between the two variables in that panel?

Question 2 Look at the test-tube figure of four metals in dilute hydrochloric acid. (a) Describe the qualitative pattern across the four metals. (b) Is this pattern qualitative or quantitative? Give a reason. (c) Suggest one quantitative measurement that would turn the pattern into a quantitative one.

Question 3 Use the sound-reduction bar chart (published versus measured). (a) Which material reduced the most sound in both datasets? (b) Calculate the difference between the published and measured values for foam and for plasterboard. (c) State the relationship between the published and measured datasets.

Question 4 Use the two-panel figure of the same class average scores (72, 74, 73, 76%). (a) Calculate the change in the average score from T1 to T4. (b) Which panel makes the change look larger? (c) What two things should a careful reader check about the axes before trusting the impression a line graph gives?

Question 5 Use the decay-curve figure. (a) Read the mass of the sample at 20 minutes. (b) State the pattern in the data in one sentence. (c) Use the pattern to find the mass at 30 minutes.

Question 6 Use the scatterplot of student height against ruler-drop distance. (a) Describe the pattern shown by the points. (b) What does the pattern say about the relationship between height and reaction distance? (c) Should a line of good fit be drawn on this graph? Give a reason.

Unscaffolded questions

Question 7 Use the bounce-height scatterplot and its line of good fit, $y = 0.58x + 1.3$ (x = drop height in cm, y = bounce height in cm). Describe the association in terms of strength, direction and linearity, and use the line to predict the bounce height from a 90 cm drop.

Question 8 Use the dissolved-oxygen scatterplot. Describe the association in terms of strength, direction and linearity, and estimate the dissolved oxygen concentration at 20 °C.

Question 9 A class heats water and records its temperature each minute: 20.0, 24.2, 28.4, 32.6, 30.0, 41.0 and 45.2 °C at minutes 0–6. (a) Identify the anomaly, and state the value the pattern of the other points suggests at that time. (b) Give two possible causes of the anomaly. (c) State what the class should do before deciding to set the anomalous value aside.

Question 10 Use the vaccination figure (notifications and coverage over 70 simulated years, vaccine introduced at year 18). (a) Describe the trend in notifications before year 18 and after year 30. (b) Describe the relationship between coverage and notifications. (c) Is this relationship causal or correlation only? Justify your answer with the mechanism.

Question 11 Use the two representations of the period-3 melting points. (a) Describe the pattern in melting point from sodium to argon. (b) Which representation shows the pattern more clearly? Justify your choice. (c) State one merit of the periodic-table layout that the bar graph does not have.

Question 12 Use the sunspot-and-temperature figure. (a) Describe the relationship between sunspot number and temperature anomaly over 1880–1970. (b) Describe what happens to the relationship after 1970. (c) Explain why the pattern after 1970 shows that sunspot activity is not the main cause of recent warming.

Question 13 Ice-cream sales and drownings show a strong correlation over the year. (a) Explain why this correlation is not a causal relationship. (b) Identify the third factor that drives both variables. (c) Give your own example of two variables that correlate because a third factor drives both.

Question 14 Two students look at the same quiz averages (T1 72%, T2 74%, T3 73%, T4 76%). Student A, looking at the panel with the axis from 0 to 100, says “the scores are basically flat — nothing changed”. Student B, looking at the panel with the axis from 70 to 80, says “the scores jumped 4 points by T4 — a real improvement”. Explain how the same data supports both impressions, and say which display you would choose to report the data, with a reason.

Question 15 A basketball is dropped and its bounce height measured: from 20 cm it bounces 15 cm, from 60 cm it bounces 45 cm, from 100 cm it bounces 75 cm, and from 140 cm it bounces 105 cm. The tennis ball of the theory scatterplot follows $y = 0.58x + 1.3$. (a) Describe the pattern in the basketball data. (b) Which ball returns the greater fraction of its drop height? Justify by comparing the two patterns. (c) Predict the basketball’s bounce height from a 90 cm drop.

Question 16 Classify each relationship as causal or correlation only, with a one-line justification. (a) Bacterial infection and disease (such as the bacteria Marshall and Warren found in ulcer patients). (b) Sunspot activity and climate change. (c) Vaccination coverage and measles notifications. (d) Ice-cream sales and drownings.

Answers

Question 1 (a) panel (a); (b) panel (c) — the points follow a bending curve, not a straight line; (c) panel (d); no relationship between the variables. **Question 2** (a) vigour of bubbling falls from Mg to Cu: fast, steady, very few, none; (b) qualitative — recorded in words, no numbers; (c) e.g. volume of gas collected in one minute. **Question 3** (a) plasterboard; (b) foam 5 dB, plasterboard 5 dB; (c) same ranking, measured values all lower (by 2–5 dB). **Question 4** (a) $76 - 72 = 4$ points; (b) panel (b); (c) where the axis starts, and the size of each scale step. **Question 5** (a) 20 g; (b) the mass halves every 10 minutes; (c) 10 g. **Question 6** (a) no clear pattern — scattered points; (b) no relationship; height does not predict reaction distance; (c) no — a line of good fit needs a roughly linear pattern to balance. **Question 7** Strong, rising, linear; $0.58 \times 90 + 1.3 \approx 53$ cm. **Question 8** Strong, falling, close to linear but slightly curved; about 9.2 mg/L. **Question 9** (a) the 30.0°C at minute 4; the pattern suggests about 36.8°C ($\approx 37^\circ\text{C}$); (b) e.g. misread thermometer; timer pressed late (or burner turned down); (c) check the reading and repeat that measurement. **Question 10** (a) before: oscillating around 500 per year; after year 30: near zero; (b) as coverage rises, notifications fall; (c) causal — vaccination gives immunity, so the disease cannot spread. **Question 11** (a) rises from Na to a peak at Si ($1\,414^\circ\text{C}$), then falls sharply; Cl and Ar melt below 0°C ; (b) the bar graph — the rise-and-fall is visible at a glance; (c) it shows each element's position and family in the table. **Question 12** (a) the two rise and fall together (correlated); (b) temperature keeps rising while the sunspot cycle continues unchanged — the lines diverge; (c) a cause must be followed by its effect; temperature has stopped following the cycle. **Question 13** (a) neither variable produces the other; (b) hot weather; (c) any valid third-factor pair, e.g. sunscreen sales and sunburn (sunny days drive both). **Question 14** Same four numbers; panel (a)'s 0–100 scale makes 4 points look tiny, panel (b)'s 70–80 scale makes them look large — the impression comes from the scale, not the data; choice of display with a justified reason. **Question 15** (a) linear, rising; bounce = $0.75 \times$ drop; (b) the basketball — its line is steeper (0.75 vs 0.58); (c) $0.75 \times 90 = 67.5 \approx 68$ cm. **Question 16** (a) causal — killing the bacteria cures the disease; (b) correlation only — the lines diverge after 1970; (c) causal — immunity blocks spread; (d) correlation only — hot weather drives both.

Fully-worked solutions

Question 1 (a) Panel (a) shows points rising along a straight line: linear and rising. (b) Panel (c). The points follow a curve that bends upwards — the rises get bigger (about 1, 2, 4, 7 ...) instead of staying the same size, so the pattern is non-linear. (c) Panel (d). The points are scattered with no direction or shape, so the conclusion is that there is no relationship between the two variables.

Question 2 (a) Magnesium bubbles fast, zinc bubbles steadily, iron shows very few bubbles and copper shows none: the vigour of the reaction falls in the order $\text{Mg} > \text{Zn} > \text{Fe} > \text{Cu}$. (b) Qualitative. The observations are recorded in words (how the bubbling looks), not as numbers. (c) For example, collect the gas for one minute and measure its volume, or time how long each metal takes to collect 10 mL of gas. Either turns the word pattern into a number pattern.

Question 3 (a) Plasterboard (30 dB published, 25 dB measured) — the tallest bar in both datasets. (b) Foam: $24 - 19 = 5$ dB. Plasterboard: $30 - 25 = 5$ dB. (c) The two datasets have the same ranking (plasterboard > foam > carpet > curtain), and every measured value is lower than the published value, by 2–5 dB.

Question 4 (a) $76 - 72 = 4$ percentage points. (b) Panel (b), the one whose axis starts at 70. (c) Where the axis starts (0 or some other value), and the size of each step on the scale — the two things that decide how big a change looks.

Question 5 (a) 20 g. (b) The mass halves every 10 minutes ($80 \rightarrow 40 \rightarrow 20 \rightarrow 10 \rightarrow 5$ g) — a falling, non-linear pattern. (c) One more halving: $20 \div 2 = 10$ g.

Question 6 (a) The points are scattered across the graph with no direction or shape — no clear pattern. (b) There is no relationship between a student's height and their ruler-drop distance; height cannot be used to predict reaction distance. (c) No. A line of good fit balances points around a roughly linear pattern; with no pattern there is nothing for the line to fit, and drawing one would invent a relationship the data does not show.

Question 7 Strength: the points sit very close to the line, so the association is strong. Direction: rising — the higher the drop, the higher the bounce. Linearity: linear — the points follow a straight line. Prediction: $y = 0.58 \times 90 + 1.3 = 52.2 + 1.3 = 53.5$, so the bounce height is about **53 cm**.

Question 8 Direction: falling — as the water warms, less oxygen stays dissolved. Strength: strong — the points sit tightly along a smooth falling line with no outliers. Linearity: close to linear but slightly curved, flattening at the hot end. At 20 °C the graph and table give about **9.2 mg/L**.

Question 9 (a) The other six points rise steadily by 4.2 °C per minute (20.0, 24.2, 28.4, 32.6, then 41.0, 45.2), so the pattern at minute 4 should be $32.6 + 4.2 = 36.8$ °C (≈ 37 °C). The recorded 30.0 °C at minute 4 is the anomaly. (b) Any two sensible causes: the thermometer was misread; the timer was pressed late so the temperature was recorded a minute early; the burner was briefly turned down; the thermometer touched the cooler beaker wall. (c) Check the reading against the pattern, and repeat that one measurement. Only if the repeat sits on the pattern should the 30.0 °C value be set aside — and the note of it should stay in the record.

Question 10 (a) Before year 18, notifications oscillate around 500 per year (roughly 300–700), rising and falling every two to three years. After year 30, notifications sit near zero and stay there. (b) As vaccination coverage rises towards 96%, notifications fall towards zero — a strong falling relationship. (c) Causal. Vaccination gives people immunity, so the disease cannot find enough people to infect and outbreaks stop; the fall begins exactly when the programme starts, and the same story repeats for other diseases. The mechanism plus the timing is what lifts this above correlation.

Question 11 (a) Melting point rises from sodium (98 °C) to a peak at silicon (1 414 °C), then falls sharply: phosphorus 44 °C, sulfur 115 °C, then chlorine –101 °C and argon –189 °C, both below 0 °C. (b) The bar graph. The rise-to-silicon-then-fall shape is visible at a glance, and the negative values stand out below the zero line; in the table layout the same numbers are buried in the row of symbols. (c) The layout shows where each element sits in the periodic table — its position and its family — which the bar graph throws away.

Question 12 (a) Over 1880–1970 the two lines rise and fall together: when sunspot number is high, the temperature anomaly tends to be high. The variables are correlated. (b) After 1970 the relationship breaks: temperature climbs steadily while the sunspot cycle continues with the same 11-year period and the same peak size. The lines diverge. (c) In a causal relationship the effect must keep following the cause. Since 1970 the temperature has stopped following the sunspot cycle, so sunspot activity cannot be the main cause of the recent warming — the correlation before 1970 was not causation.

Question 13 (a) For causation, changing one variable would have to change the other, and neither does: selling more ice cream does not put people in the water, and fewer drownings would not sell fewer ice creams. The two simply move together. (b) Hot weather: in hot weather more ice creams are sold and more people swim, and more swimmers leads to more drownings. One third factor causes both variables. (c) Any pair with a shared driver, for example: sunscreen sales and sunburn (sunny days drive both), or woollen-hat sales and hot-drink sales (cold days drive both).

Question 14 Both students are reading the same four numbers (72, 74, 73, 76). The change from T1 to T4 is 4 points either way. In panel (a) the axis runs 0–100, so 4 points is a small fraction of the axis and the line looks flat; in panel (b) the axis runs 70–80, so the same 4 points fills most of the axis and looks like a dramatic climb. The different impressions come from the choice of scale, not from the data — this is elaboration VC2S10I05.E02 in action. To report the data I would choose the 0–100 scale if I wanted to show the scores as a level (the class is consistently around 70–76%), or the 70–80 scale if I wanted to show the small term-to-term changes honestly and prominently — stating the axis choice so the reader is not misled either way.

Question 15 (a) The bounce height is always 0.75 of the drop height (15/20, 45/60, 75/100, 105/140), so the pattern is linear and rising: $\text{bounce} = 0.75 \times \text{drop}$. (b) The basketball. Its line (slope 0.75) is steeper than the tennis ball's line of good fit (slope 0.58), so the basketball returns about 75% of its drop height while the tennis ball returns about 58%. (c) $0.75 \times 90 = 67.5$, so about **68 cm**.

Question 16 (a) Causal. The correlation was confirmed as causation by the extra evidence: killing the bacteria cured the disease, and swallowing the bacteria produced the same stomach inflammation — changing one variable changed the other. (b) Correlation only. After 1970 the temperature keeps rising while the sunspot cycle runs unchanged, so the supposed effect no longer follows the supposed cause. (c) Causal. Vaccination produces immunity, which stops the disease spreading; when coverage falls, outbreaks return — changing the cause changes the effect. (d) Correlation only. Hot weather is a third factor that increases both ice-cream sales and swimming (and therefore drownings); neither variable causes the other.

Where this leads: the skills of this subtopic — describing patterns in direction, shape and strength, comparing datasets, checking anomalies and separating correlation from causation — are re-used by every knowledge chapter in this book that graphs data (for example 12D’s environmental trends and 15C’s motion graphs, where gradient becomes a physical quantity), and they are exactly what VCE Science investigations assess in all five studies (Biology, Chemistry, Physics, Psychology, Environmental Science) — reference-only for this book (TOC D12). No VCE content is pre-taught here (TOC D9).