Mathematics Level 6

Chapter 7 — Statistics

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

Comparing data sets: mode, range and shape

> *interpret and compare data sets for ordinal and nominal categorical, discrete and continuous numerical variables using comparative displays or visualisations and digital tools; compare distributions in terms of mode, range and shape* — VC2M6ST01

Theory

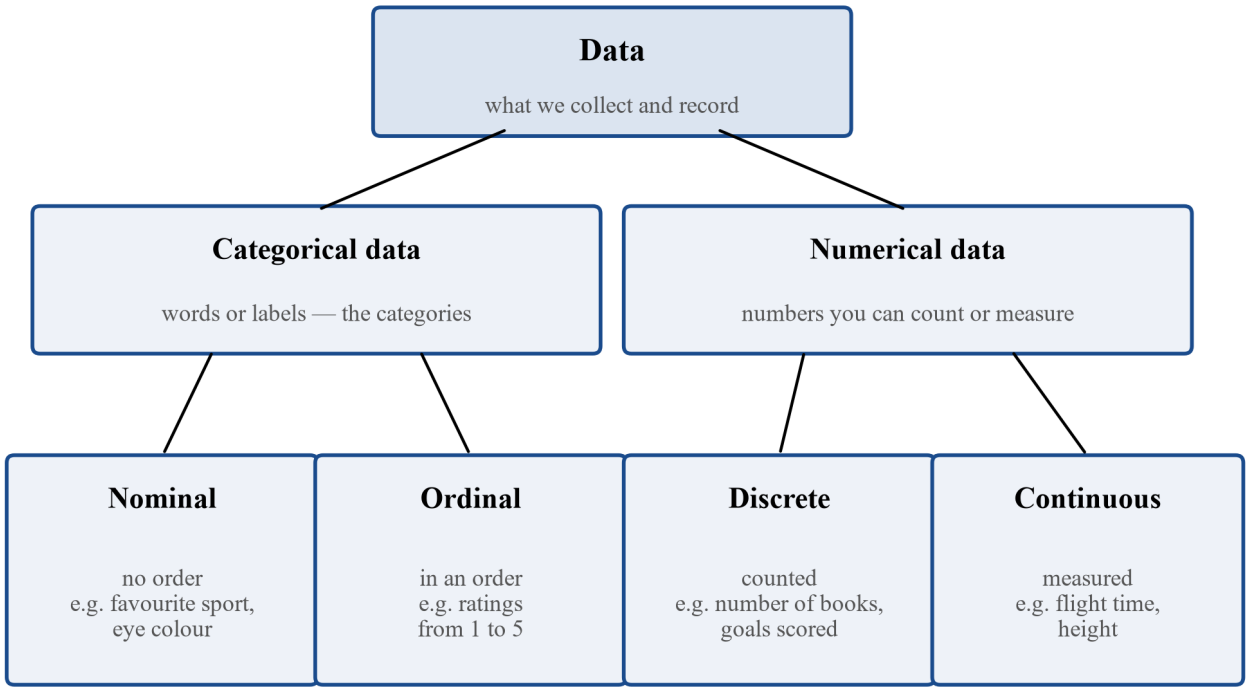
Why we compare data

Two classes run a reading challenge. Two paper-plane designs go head to head. Two towns share a weather report. In each case the question is the same: *which one, and how can we tell?* One number is almost never enough. A list of numbers on its own does not tell the story — we need to look at the whole set and describe what it is like.

In this subtopic you learn three tools for comparing data sets. The **mode** tells you the most frequent result. The **range** tells you how far the numbers spread out. The **shape** tells you what the whole distribution looks like — where it peaks, whether it leans to one side, or whether it is flat. Together, they let you compare any two data sets and say *why* they are different.

The four types of variables

Data comes in types, and the type decides what you can do with it. The thing you ask about or measure for each person or object is called a **variable**. There are four types, and the names matter: you will use them in this subtopic, in 7C, and again in Year 7.



Each type suits a different display — that is what 7A teaches.

A concept map with Data at the top, splitting into categorical data, meaning words or labels, and numerical data, meaning numbers you can count or measure; categorical splits into nominal, with no order, for example favourite sport and eye colour, and ordinal, in an order, for example ratings from 1 to 5; numerical splits into discrete, counted, for example number of books and goals scored, and continuous, measured, for example flight time and height

Categorical data is words or labels. It splits into two types.

Type	What it is	Examples
Nominal	labels with no order — the names just name things	favourite sport, eye colour, type of pet

Type	What it is	Examples
Ordinal	labels you *can* put in order	a rating from 1 to 5, a medal (gold, silver, bronze), a size (small, medium, large)

Numerical data is numbers. It also splits into two types.

Type	What it is	Examples
Discrete	numbers you **count** — they jump in steps, often whole numbers	number of books read, goals scored, people in a car
Continuous	numbers you **measure** — any value in a range is possible	time of a paper-plane flight, height, temperature

A continuous measurement can take any value: between 2.0 s and 2.1 s there is 2.05 s, and between those there is another value, and so on. When we record continuous data we usually round it — to the nearest 0.1 s, or the nearest cm — so it can be written down and graphed. A quick test helps you sort the four types: *do I count it or measure it?* (numerical) or *do I just name it?* (categorical), and if I name it, *can I put the names in order?*

Displays to suit the data

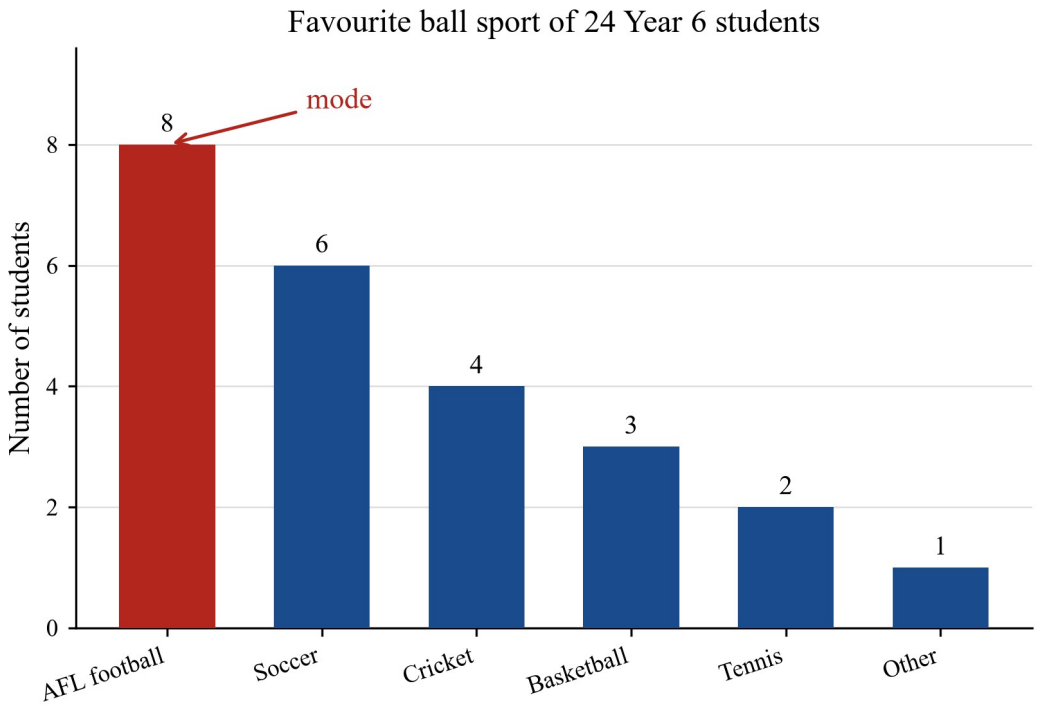
Pick a display that suits the variable type, and the story is easy to read.

- **Nominal** data suits a **column graph** — one column for each label, and the height shows how many.
- **Ordinal** data suits a **dot plot** or a bar chart, with the labels in order along the bottom (`VC2M6ST01.E03`).
- **Discrete** numerical data suits a dot plot or a column graph — one dot or column for each value.
- **Continuous** numerical data suits a dot plot once the values are rounded and recorded (each dot is one measurement), or a column graph of grouped values.

To compare **two** data sets, draw them on the same scale, side by side. Two common comparative displays are the **side-by-side column graph** (`VC2M6ST01.E02`) — two columns for each category, one for each set — and two dot plots stacked one above the other on the same axis. Comparative displays only work when both sets share the same scale, so the bars and dots can be read against each other.

Mode: the most frequent result

The **mode** of a data set is the value or label that occurs **most often**. On a column graph or a dot plot it is the tallest column or the tallest stack of dots. The mode works for every type of data — nominal, ordinal, discrete and continuous — because it only ever asks *which one turned up the most?*



A column graph of the favourite ball sport of 24 Year 6 students: AFL football 8, soccer 6, cricket 4, basketball 3, tennis 2 and other 1, with the tallest column for AFL football marked as the mode

In the survey above, 24 students named their favourite ball sport. The tallest column is AFL football with 8 students, so the mode is **AFL football**.

A data set can have **two modes**: if two values are joint most frequent, both are modes. And if every value occurs equally often — a perfectly flat data set — there is **no mode** at all.

Range: highest − lowest

The **range** measures how spread out a numerical data set is. Find it by subtracting the lowest value from the highest value (`VC2M6ST01.E01`):

$$\text{range} = \text{highest value} - \text{lowest value}$$

For the data set 2, 4, 4, 5, 9:

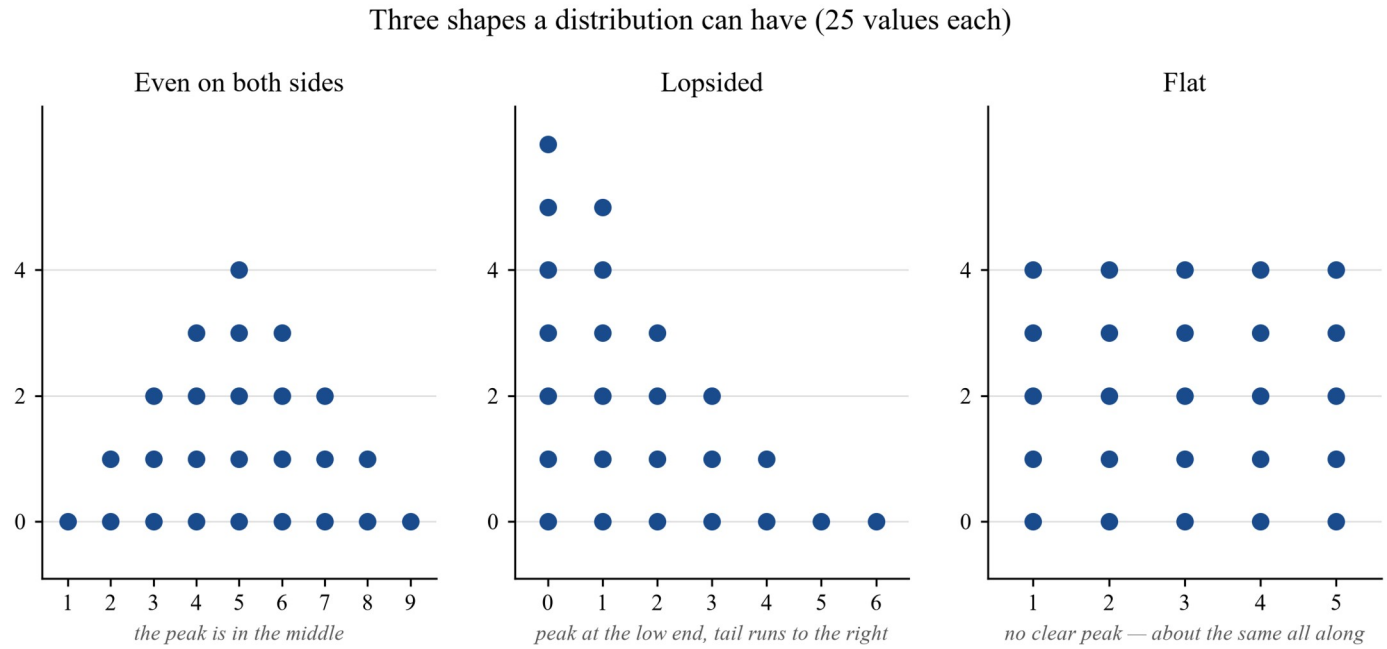
$$\text{range} = 9 - 2 = 7$$

A **wide range** means the values are more spread out. A **small range** means they sit close together. Two data sets can have the same mode but very different ranges — the mode says where the data piles up, and the range says how far it stretches.

The range is only for **numerical** data. You cannot find the range of eye colours — words have no highest and lowest, and you cannot subtract blue from brown. The range also uses only the two end values, so one very big or very small value can change it a lot. You will see this in Q16.

Shape: what the distribution looks like

The **shape** of a distribution is what the whole display looks like when you stand back from the individual numbers. Three shapes to know:



Three dot plots of 25 values each: one even on both sides with the peak in the middle, one lopsided with the peak at the low end and a tail running to the right, and one flat with no clear peak

- **Even on both sides** — the peak sits in the middle and the two halves are roughly the same.
- **Lopsided** — the peak sits near one end, and a long tail of low stacks runs away to the other end.
- **Flat** — no clear peak; the stacks are about the same height all along.

Words that help you describe shape: the **peak** is where the tallest stack sits, a **tail** is the thin line of small stacks that runs away from the peak, and a **cluster** is a bunch of values sitting close together.

Watch out for

- **Do not find the range of categorical data.** Labels have no numbers in them. For a nominal data set you can report the mode — never a range.
- **The mode is about how often, not how big.** The mode of 0, 0, 0, 1, 2 is 0 — most families in that set have *no* pets at all.
- **One extreme value moves the range.** The range looks only at the two end values, so one very hot day or one very long throw stretches it.
- **Two modes and no mode are both possible.** Joint most frequent values give two modes; a flat data set has no mode.

Comparing two groups side by side

Often the interesting question is not *what does this data say?* but *do two groups give different answers?* Data can be grouped by year level, by gender, by age group, or by another variable, and each group gets its own columns in one display (`VC2M6ST01.E04`). Then the mode, range and shape of each group can be compared directly — and sometimes the groups tell different stories. In Worked example 6 you will see an excursion vote where Years 3–4 and Years 5–6 choose in different ways.

Digital tools do the drawing

Digital tools are part of this subtopic (`VC2M6ST01`). A **spreadsheet** is where a data set is set out: one column for the values, one column for each group's counts, a heading row, and one row for every value — even when a count is 0, so the two sets stay lined up. **Graphing software** then draws the display from the entered numbers in seconds, and you can change the display type without re-entering anything. Technology is also how you **access** data sets: the weather data in Worked example 7 comes from the Bureau of Meteorology, which publishes its climate statistics online.

	A	B	C
1	**Books read**	**Year 5 class**	**Year 6 class**
2	2	1	0
3	3	2	0
4	4	4	1
5	5	3	2
6	6	2	4
7	7	1	3
8	8	1	2
9	9	1	1
10	10	0	2

Every value gets a row — a 0 keeps the two classes lined up.

A spreadsheet layout ready for graphing: column A lists the numbers of books read from 2 to 10, column B holds the Year 5 class counts and column C the Year 6 class counts, with a 0 kept in place so the two classes line up

Words of this subtopic

Term	Meaning here
data	the values collected in a survey, count or measurement
variable	the thing you ask about or measure for each person or object
categorical data	data that is words or labels
nominal	categorical labels with no order
ordinal	categorical labels that can be put in order
numerical data	data that is numbers
discrete	numerical data you count
continuous	numerical data you measure
distribution	a data set together with how its values are spread across the display
display	a graph, chart or plot that shows data
comparative display	a display that shows two or more data sets on the same scale
mode	the value or label that occurs most often
range	highest value − lowest value, for numerical data
shape	what the whole distribution looks like — even, lopsided or flat
spread	how far out the values reach; a big spread means a big range
frequency	how often a value occurs
dot plot	a display that stacks one dot above each value
side-by-side column graph	a column graph with one column per set for each category, so two sets can be compared

Worked examples

Example 1 — with help

Classify each variable as **nominal**, **ordinal**, **discrete** or **continuous**.

1. Favourite sandwich filling
2. Rating the canteen lunch from 1 to 5
3. Number of books read in Term 2
4. Time to walk to school
5. Eye colour
6. Height of each student in the class

Variable	Type	Why
Favourite sandwich filling	nominal (categorical)	a label, and the fillings have no order
Rating from 1 to 5	ordinal (categorical)	labels, but 1 to 5 puts them in order
Number of books read	discrete (numerical)	a number you count — 4 books, never 4.5
Time to walk to school	continuous (numerical)	a number you measure — any value is possible
Eye colour	nominal (categorical)	a label with no order
Height	continuous (numerical)	measured, and can take any value in a range

The quick test: count or measure → numerical; name → categorical; and if it names, can it be ordered?

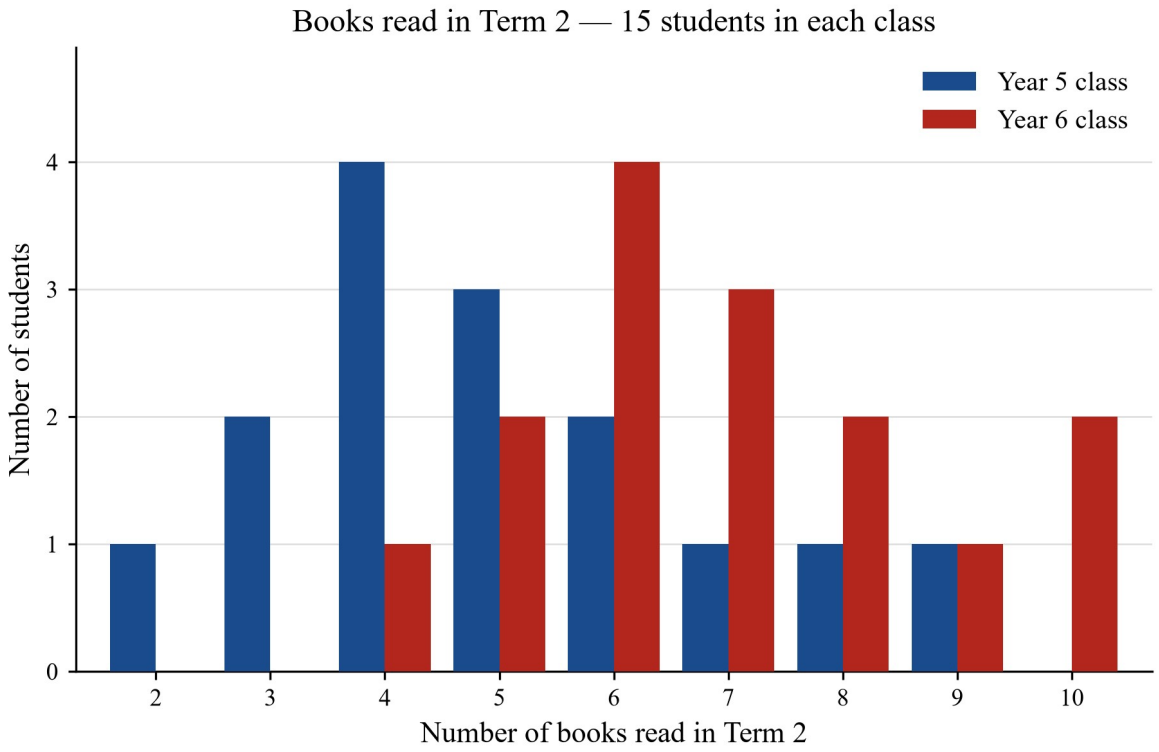
Example 2 — with help

Read the favourite-sport column graph above. (a) How many students were surveyed? (b) What is the mode? (c) Can this data set have a range?

Working	Comment
Add the column heights: $8+6+4+3+2+1=24$.	Every student is counted in exactly one column.
The tallest column is AFL football at 8.	The mode is the most frequent label.
Mode: **AFL football**.	Eight students chose it — more than any other sport.
The data is nominal — the answers are labels, not numbers.	You cannot order sports by size or subtract one from another.
No — a range makes no sense here.	Range is only for numerical data.

Example 3 — with help

Two classes of 15 students each recorded how many books they read in Term 2. The spreadsheet and the side-by-side column graph below show the same data.



A side-by-side column graph of books read in Term 2 for 15 students in each class: the Year 5 class peaks at 4 books and the Year 6 class peaks at 6 books

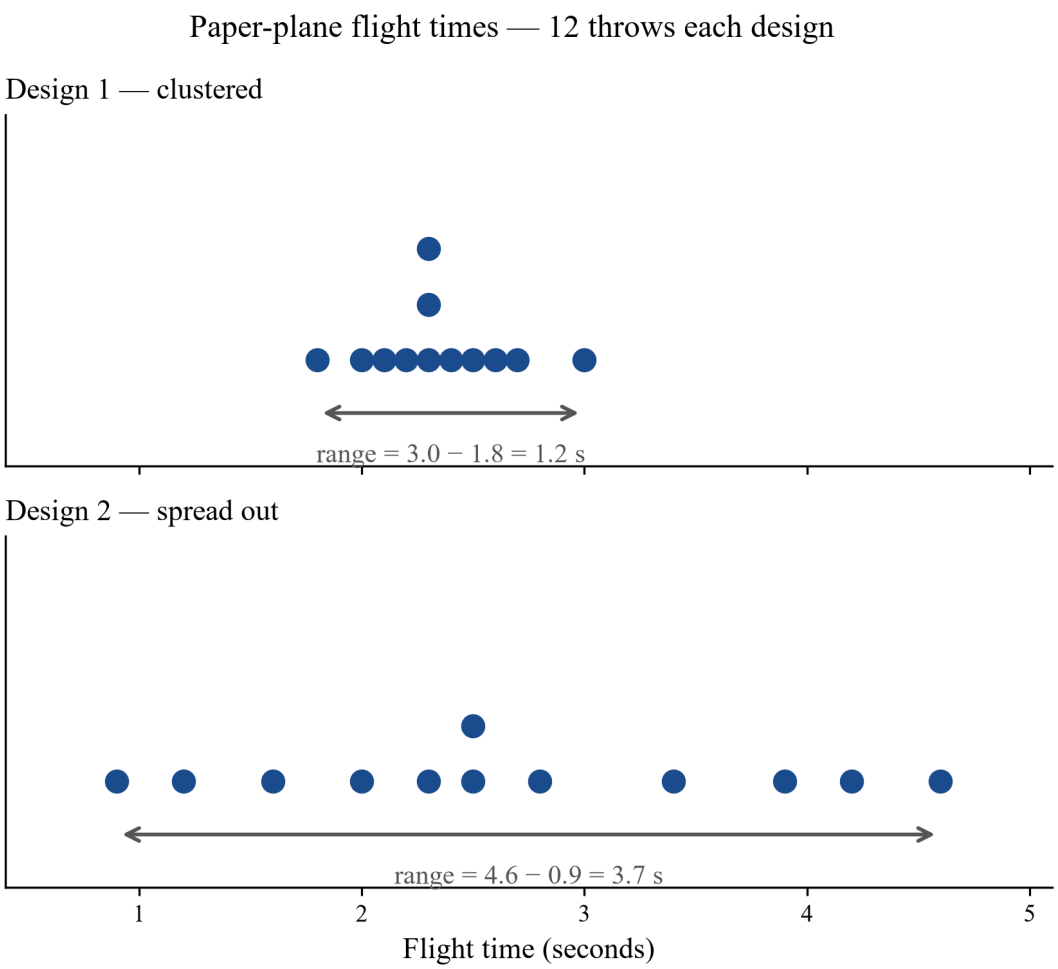
Find the mode and range for each class, then compare the two distributions.

Working	Comment
Year 5: the tallest column is at 4 books. Mode = **4**.	Read the mode straight off the graph.
Year 5: highest value with students is 9, lowest is 2. $9 - 2 = 7$. Range = **7**.	Range is highest − lowest (`VC2M6ST01.E01`).
Year 6: the tallest column is at 6 books. Mode = **6**.	
Year 6: highest is 10, lowest is 4. $10 - 4 = 6$. Range = **6**.	
Compare: the Year 6 class read more books in the middle (mode 6 against 4), but the spreads are about the same (ranges 7 and 6).	Same question, two groups, one comparative display.

In the spreadsheet, each value gets a row, and a 0 keeps the classes lined up where no student read that many books.

Example 4 — with help

Two paper-plane designs were each thrown 12 times. The flight times (in seconds, to one decimal place) are shown as side-by-side dot plots.



Two dot plots of paper-plane flight times: Design 1 is clustered around 2.3 seconds with a range of 1.2 seconds, and Design 2 is spread out with a range of 3.7 seconds

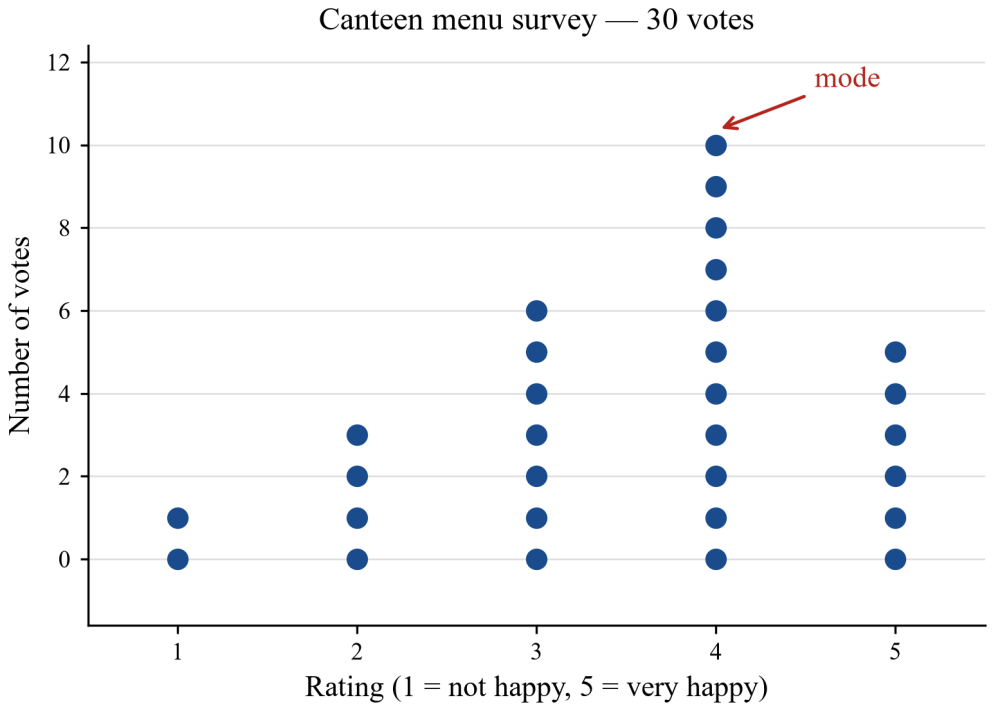
(a) Find the range for each design. (b) Which design is more predictable? (c) Find the mode for each design.

Working	Comment
Design 1: highest 3.0 s, lowest 1.8 s. $3.0 - 1.8 = 1.2$. Range = **1.2 s**.	Read the ends of the bracket under the dot plot.
Design 2: highest 4.6 s, lowest 0.9 s. $4.6 - 0.9 = 3.7$. Range = **3.7 s**.	
Design 1's dots sit close together; Design 2's dots stretch across the axis.	Design 1 has the smaller spread.
Design 1 is more predictable — its throws nearly always land near 2.3 s.	Small range → results you can count on.
Design 1 mode: **2.3 s** (three throws). Design 2 mode: **2.5 s** (two throws).	The tallest stack on each dot plot.

Design 2 did produce the single longest flight (4.6 s) — but you could not count on it.

Example 5 — on your own

The school canteen surveyed its new menu. Thirty students each voted a rating from 1 (not happy) to 5 (very happy). The results are in the dot plot.



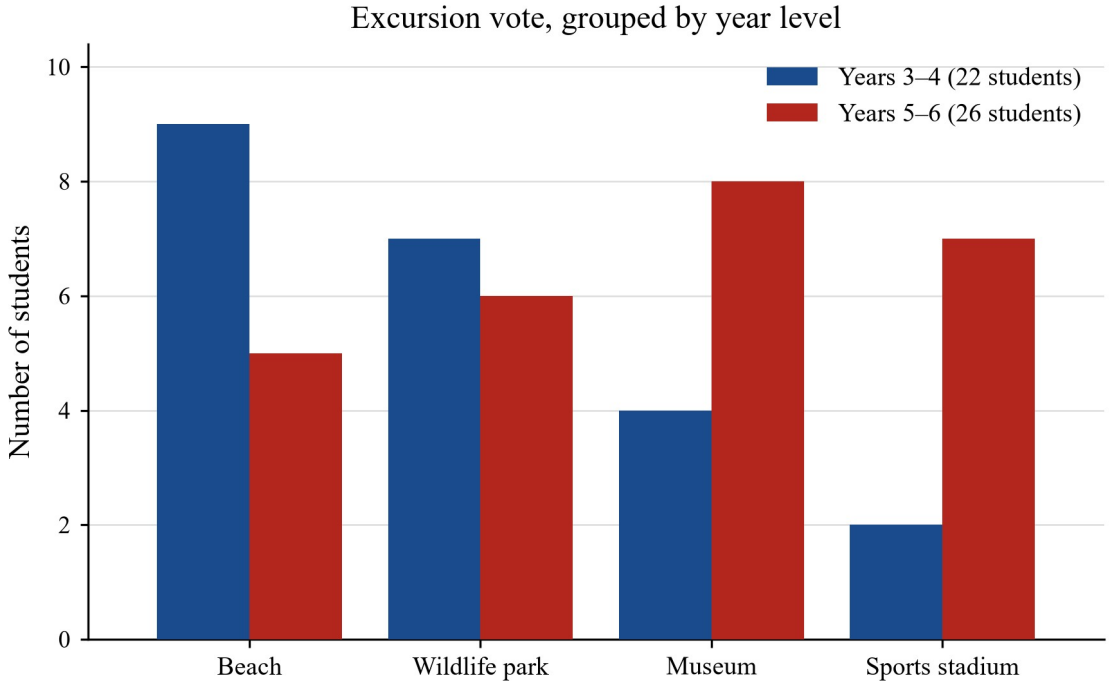
A dot plot of 30 canteen ratings from 1 to 5: the stack at rating 4 is the tallest with 11 votes and is marked as the mode

(a) What is the mode of the ratings?
(b) Describe the shape of the distribution.
(c) Why is a dot plot a good display for this data?

The mode is **4** — the tallest stack, with 11 votes. The shape is **lopsided towards the high ratings**: the peak sits at 4, most votes are 3 or higher, and only a few students voted 1 or 2, so the tail runs down the low end. A dot plot suits this data because the ratings are **ordinal** — the labels run in order from 1 to 5, so one dot per vote stacked above each rating shows both the order and the shape at a glance.

Example 6 — on your own

A primary school asked its students to vote for the end-of-year excursion. The results were grouped by year level.



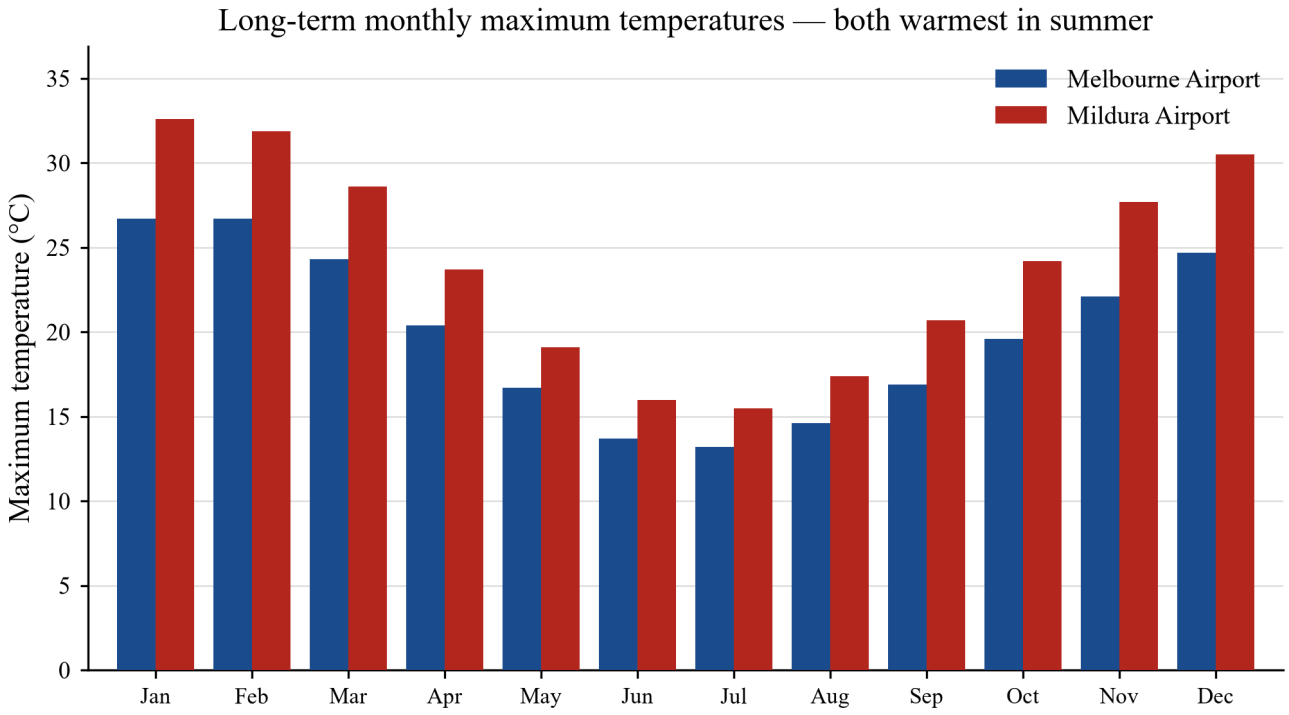
A side-by-side column graph of the excursion vote grouped by year level: Years 3–4 prefer the beach with 9 votes and Years 5–6 prefer the museum with 8 votes

(a) What is the mode for Years 3–4? (b) What is the mode for Years 5–6? (c) The overall vote, with both groups combined, is beach 14, wildlife park 13, museum 12, sports stadium 9. What is the overall mode — and why might the school still think twice before booking it?

The mode for Years 3–4 is the **beach** (9 votes), but the mode for Years 5–6 is the **museum** (8 votes). Combined, the mode is the **beach** (14 votes) — yet the museum came first with the older students. Grouping the data by year level (`VC2M6ST01.E04`) shows what the combined total hides: the two groups chose in different ways. A school that books only the overall mode is booking against half its students' first choice.

Example 7 — on your own

The Bureau of Meteorology publishes long-term climate statistics for Victorian towns. The graph shows the long-term monthly maximum temperatures for two of them.



A side-by-side column graph of long-term monthly maximum temperatures for Melbourne Airport and Mildura Airport, with both towns warmest in summer

Source: Bureau of Meteorology, Climate statistics for Australian locations, Melbourne Airport (site 086282) and Mildura Airport (site 076031), long-term averages. Retrieved 20 August 2026.

Month	January	February	March	April	May	June	July	August	September	October	November	December
Melbourne Airport (°C)	26.7	26.7	24.3	20.4	16.7	13.7	13.2	14.6	16.9	19.6	22.1	24.7
Mildura Airport (°C)	32.6	31.9	28.6	23.7	19.1	16.0	15.5	17.4	20.7	24.2	27.7	30.5

(a) In which month are both towns warmest? (b) Find the range of each town's temperatures. (c) Which town's temperatures change more across the year? (d) Describe the shape of each town's year.

(b) **January** — summer in Australia, so the temperatures peak at the start of the year (and the year ends warm too). (b) Melbourne Airport: $26.7 - 13.2 = 13.5$, range **13.5 °C**. Mildura Airport: $32.6 - 15.5 = 17.1$, range **17.1 °C**. (c) **Mildura Airport** — its range is the wider, $17.1\,°C$ against $13.5\,°C$. Mildura runs much hotter in summer (32.6 °C in January) while its winters stay close to Melbourne's. (d) Both towns have a **hill shape**:

the temperatures climb to a summer peak in January and fall away through winter and climb again. Mildura's hill is the taller one, because its summers are so much hotter.

Example 8 — on your own (digital-tool path)

Two Year 6 classes voted on the colour of their new sports shirts. The results are:

Colour	Class 6A	Class 6B
Red	8	5
Blue	6	9
Green	4	7
Gold	7	4

Use a **spreadsheet and graphing software** (or any graphing tool your school uses) to draw a side-by-side column graph, then read off each class's mode.

Enter the colours in one column, with the two classes' votes in the next two columns and a heading row on top, just like the books-read table in the Theory. Select all three columns including the headings, choose a column graph, and give the graph a title and axis labels. The graph shows two columns at each colour — one for each class — so the classes can be compared colour by colour. Reading the tallest column for each class: Class 6A's mode is **red** (8 votes) and Class 6B's mode is **blue** (9 votes). The two classes would choose different colours — a good reason to keep the data grouped instead of adding the classes together. (Check: $8+6+4+7=25$ and $5+9+7+4=25$, so every student in both classes is counted.)

Practice questions

*The questions marked **digital tools** need access to a spreadsheet or graphing software. The rest can be done with pencil and paper.*

With help

Q1. Classify each variable as **nominal**, **ordinal**, **discrete** or **continuous**. (a) the number of brothers and sisters a student has (b) a student's favourite movie (c) a rating of the school camp from 1 to 5 (d) the time it takes to walk to school

Q2. Which of these data sets suits a dot plot best? A the colours of 20 cars in the staff car park B the heights of 25 students, measured to the nearest cm C the names of the days of the week D the brands of phones that students own

Q3. A school soccer team scored these goals over 8 matches: 1, 2, 2, 3, 2, 4, 2, 3. (a) Find the mode. (b) Find the range.

Q4. Over 8 nights, a student recorded the minutes spent reading before bed: 20, 30, 30, 25, 20, 35, 20, 30. (a) Find the mode, or say if there is more than one. (b) What does your answer say about the student's reading habits?

Q5. Six students jumped these distances in the long jump: 3.2 m, 3.5 m, 3.1 m, 3.8 m, 3.4 m, 3.6 m. (a) Find the range. (b) Is this data discrete or continuous? How do you know?

Q6. Twenty students were asked their eye colour: brown 9, blue 6, green 3, hazel 2. (a) Find the mode. (b) Explain why this data set cannot have a range.

Q7. Three dot plots are described in words. Match each description to a shape — **even on both sides**, **lopsided** or **flat**. (a) The stacks are about the same height all along, with no clear peak. (b) The tallest stack is in the middle, and the two halves look roughly the same. (c) Most of the dots pile up at the low values, and a thin tail runs to the right.

Q8. Look at the spreadsheet in the Theory, where the books-read data is set up for graphing. (a) What does column A record? (b) The Year 5 column holds a 0 in the row for 10 books. Why is that row kept in the table? (c) To draw the side-by-side column graph, which columns do you select?

Q9. Read the favourite-sport column graph in the Theory. (a) How many students were surveyed? (b) What is the mode? (c) How many students chose soccer or cricket?

Q10. Read the canteen dot plot in Worked example 5. (a) How many votes were cast altogether? (b) What is the mode? (c) How many students voted 4 or 5?

On your own

Q11. Classify each variable, and give the reason for each answer — *counted*, *measured*, *labels with no order*, or *labels in order*. (a) the temperature at 9 am each day for a week (b) the number of text messages sent in one day (c) the brand of a student's school bag (d) how much a student agrees with a class rule: *agree a lot*, *agree a little*, *disagree*

Q12. Twelve students borrowed these numbers of library books in a month: 0, 1, 1, 2, 2, 2, 2, 3, 3, 4, 5, 5. (a) Find the mode. (b) Find the range. (c) Describe the shape of the distribution.

Q13. Spot the errors. (a) Marco lists his favourite colours: blue, green, blue, red, blue. He writes "range = red − blue". What is wrong with this? (b) Sofia says the range of the data set 4, 4, 4, 4 is 0. Is she right? Explain why.

Q14. Two data sets are below. Set A: 1, 2, 2, 3, 1, 4 Set B: 1, 2, 3, 4, 1, 2, 3, 4 (a) Find the mode of Set A, or say if there is more than one. (b) Set B has no mode. Explain why, and name the shape it would make in a dot plot.

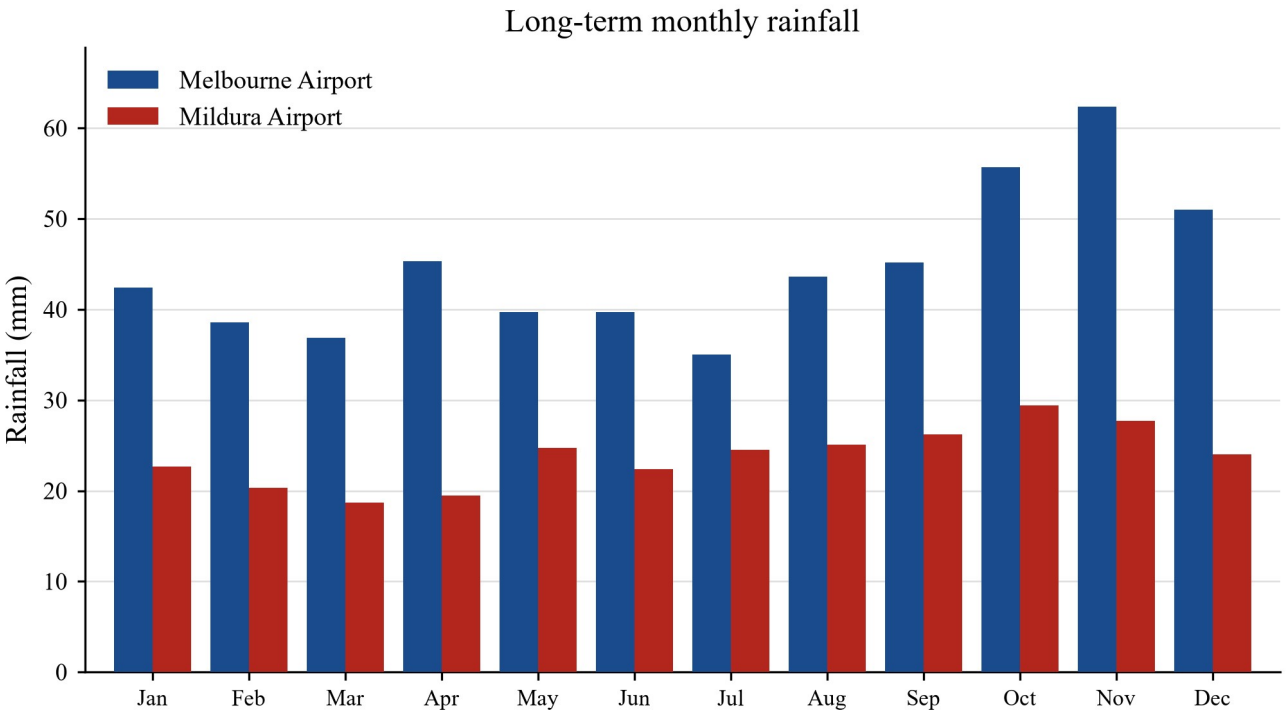
Q15. A school asked how students travel to school and grouped the answers by year level.

Travel	Years 3–4	Years 5–6
Walk	12	5
Bus	6	14
Car	9	4
Bike	3	7

(a) Find the mode for Years 3–4.
(b) Find the mode for Years 5–6.
(c) Find the mode when both groups are combined. (Add the columns first.)
(d) What does grouping tell you that the combined mode hides?

Q16. A school weather station recorded the maximum temperature each day for one week: 18.2 °C, 19.6 °C, 21.0 °C, 24.5 °C, 26.3 °C, 27.1 °C, 28.4 °C. (a) Find the range of the week. (b) The next day was a hot one: 38.9 °C. Find the new range of all 8 days. (c) What does this say about the range as a measure of spread?

Q17. The graph in this question shows long-term monthly rainfall for two Victorian towns. The values are printed below it.



A side-by-side column graph of long-term monthly rainfall for Melbourne Airport and Mildura Airport

Source: Bureau of Meteorology, Climate statistics for Australian locations, Melbourne Airport (site 086282) and Mildura Airport (site 076031), long-term averages. Retrieved 20 August 2026.

Month	January	February	March	April	May	June	July	August	September	October	November	December
Melbourne Airport (mm)	42.4	38.6	36.9	45.3	39.7	39.7	35.0	43.6	45.2	55.7	62.4	51.0
Mildura Airport (mm)	22.7	20.3	18.7	19.5	24.7	22.4	24.5	25.1	26.2	29.4	27.7	24.0

(a) Which town has the wettest month, and which month is it?

(b) Find the range of each town's rainfall.

(c) Which town's rainfall changes more from month to month?

(d) Describe the shape of Melbourne's rainfall across the year.

Q18. Look again at the temperature graph and table in Worked example 7. (a) In which month are both towns warmest? Why is that month warm in Victoria? (b) Find the range for each town. (c) Which town's temperatures change more across the year? How do you know? (d) Describe the shape of each town's year in one sentence.

Q19. For each variable, choose the display that suits it best — a **column graph**, a **dot plot**, or a **side-by-side column graph** — and say why. (a) the favourite subject of 30 students (b) ratings of the school fete from 1 to 5, collected from 40 students (c) the number of wet days in each month, for two towns at once

Q20. Look again at the paper-plane dot plots in Worked example 4. (a) Write down the range for each design. (b) Which design is more predictable? How does the range show it? (c) Your team gets one throw in a longest-flight contest. Which design do you choose, and why?

Q21. Draw two dot plots of your own, each using 9 values. (a) Make one distribution even on both sides. (b) Make one distribution lopsided. (c) Label each plot with its shape.

Q22. Two classes each asked 15 students how many minutes they spent on homework one night. Class A: mode 30 minutes, range 40 minutes. Class B: mode 30 minutes, range 10 minutes. (a) What is the same about the two classes? (b) What is different? (c) Which class's homework times are more spread out?

Q23. Three data sets: Set P: 5, 5, 5, 5, 5 Set Q: 3, 5, 5, 5, 7 Set R: 1, 2, 5, 8, 9 (a) Find the range of each set. (b) Order the sets from least spread to most spread. (c) All three sets have the mode 5. What does this tell you about what the mode does *not* measure?

Q24. (digital tools) Two classes each surveyed 20 students about the number of pets at home. The results are:

Number of pets	0	1	2	3	4	5
Class 6W	2	6	5	4	3	0
Class 6K	4	5	6	3	1	1

(a) Set this data out in a spreadsheet: one column for the number of pets, one column for each class, and a heading row.

(b) Use the spreadsheet or graphing software to draw a side-by-side column graph. Write down the steps you took, in order. No product names or menu paths are needed.

(c) From your graph, read off the mode and the range for each class.

Q25. Sort these categorical variables into **nominal** and **ordinal**. eye colour · medal won (gold, silver, bronze) · favourite subject · level of a video game (easy, medium, hard) · type of pet · size of a drink (small, medium, large)

Q26. Sort these numerical variables into **discrete** and **continuous**. number of goals in a season · time to ride to school · number of eggs in a carton · length of the classroom · mass of a pumpkin · number of people in a car

Q27. Look again at the excursion vote in Worked example 6. (a) Write down the mode for each year-level group. (b) Find the overall total for each option, and name the overall mode. (c) Give one reason the school might not simply book the overall mode.

Q28. In subtopic 7C you will plan and run your own statistical investigation. Get ready now. (a) Write a survey question that collects **ordinal** data. (b) Write a survey question that collects **nominal** data. (c) Write a survey question that collects **discrete numerical** data. (d) Choose one of your questions and name the display you would use for its results. Say why it suits the variable type.

Q29. Two teams each took 9 shots at a target and scored these points. Team A: 6, 7, 7, 8, 8, 8, 9, 9, 10 Team B: 2, 4, 6, 8, 8, 8, 10, 12, 14 (a) Find the mode and the range for each team. (b) Describe the shape of each team's distribution. (c) Write two sentences comparing the teams, using mode, range and shape.

Answers

Q1. (a) discrete; (b) nominal; (c) ordinal; (d) continuous. **Q2.** B. **Q3.** (a) 2; (b) 3. **Q4.** (a) two modes, 20 and 30; (b) the reading most often lands at 20 or 30 minutes — there is no single most common time. **Q5.** (a) 0.7 m; (b) continuous — jump lengths are measured, and any value in a range is possible. **Q6.** (a) brown; (b) the data is nominal — the answers are labels with no numbers, so there is no highest or lowest to subtract. **Q7.** (a) flat; (b) even on both sides; (c) lopsided. **Q8.** (a) the number of books read in Term 2; (b) so the two classes stay lined up — the 0 shows that no Year 5 student read 10 books; (c) all three columns, headings included. **Q9.** (a) 24; (b) AFL football; (c) 10. **Q10.** (a) 30; (b) 4; (c) 17. **Q11.** (a) continuous — measured; (b) discrete — counted; (c) nominal — labels with no order; (d) ordinal — labels in order. **Q12.** (a) 2; (b) 5; (c) lopsided — the peak is at 2 and a tail runs to the higher values. **Q13.** (a) the data is nominal, so it has no range — colours cannot be subtracted; (b) yes: $4 - 4 = 0$; all the values are the same, so there is no spread at all. **Q14.** (a) two modes, 1 and 2; (b) every value occurs the same number of times, so no value is most frequent; the dot plot is flat. **Q15.** (a) walk; (b) bus; (c) bus (20 against 17, 13 and 10); (d) walking is the mode for Years 3–4 but the bus is the mode for Years 5–6 — one combined mode hides that. **Q16.** (a) 10.2 °C; (b) 20.7 °C; (c) one very hot day can stretch the range a long way, because the range only looks at the two end values. **Q17.** (a) Melbourne Airport, November (62.4 mm); (b) Melbourne Airport 27.4 mm, Mildura Airport 10.7 mm; (c) Melbourne Airport — its range is the wider; (d) nearly flat all year, with a gentle rise to a spring peak around November. **Q18.** (a) January — January is in the middle of summer in Australia; (b) Melbourne Airport 13.5 °C, Mildura Airport 17.1 °C; (c) Mildura Airport — its range is 17.1 °C against 13.5 °C; (d) each town climbs to a summer peak in January and falls away again — a hill shape, taller for Mildura. **Q19.** (a) column graph — the subjects are nominal labels; (b) dot plot — the ratings are ordinal and run in order; (c) side-by-side column graph — two groups compared month by month. **Q20.** (a) Design 1: 1.2 s; Design 2: 3.7 s; (b) Design 1 — the smaller range means its throws sit close together; (c) any reasoned choice, for example Design 2, because it has produced the longest single flight (4.6 s), or Design 1, because its result can be counted on. **Q21.** any correct pair of dot plots, 9 values each — for example 1, 2, 2, 3, 3, 3, 4, 4, 5 for the even one and 0, 0, 0, 1, 1, 2, 3, 4, 5 for the lopsided one; check the shape after plotting. **Q22.** (a) the mode — both 30 minutes; (b) the range — 40 against 10; (c) Class A. **Q23.** (a) P: 0, Q: 4, R: 8; (b) P, Q, R; (c) the spread — the mode only says where the data piles up, not how far it stretches. **Q24.** (a) layout as described; (b) steps in order, no product names; (c) Class 6W: mode 1, range 4; Class 6K: mode 2, range 5. **Q25.** nominal: eye colour, favourite subject, type of pet; ordinal: medal won, level of a video game, size of a drink. **Q26.** discrete: number of goals, number of eggs, number of people in a car; continuous: time to ride to school, length of the classroom, mass of a pumpkin. **Q27.** (a) Years 3–4: beach; Years 5–6: museum; (b) beach 14, wildlife park 13, museum 12, sports stadium 9 — overall mode: beach; (c) for example, the older students' first choice was the museum, so the beach would be booked against their wishes. **Q28.** examples will vary — see the worked solutions for the checks each answer must pass. **Q29.** (a) Team A: mode 8, range 4; Team B: mode 8, range 12; (b) both are even on both sides with the peak at 8 — Team B is the wider hill; (c) example given in the worked solutions.

All the working

Q1. (a) You **count** brothers and sisters — a whole number of them — so the variable is **discrete** numerical. (b) A favourite movie is a label, and movie titles have no order, so it is **nominal** categorical. (c) Ratings from 1 to 5 are labels you can put in order, so the variable is **ordinal** categorical. (d) Time is measured — any value is possible between one time and the next — so it is **continuous** numerical.

Q2. A dot plot stacks one dot per value, so it needs numerical data that runs in order. Heights measured to the nearest cm are exactly that. The colours of cars, the days of the week and phone brands are all labels, not numbers. $\therefore$ **B**.

Q3. (a) Count how often each score appears: 1 once, 2 four times, 3 twice, 4 once. The score that appears most often is 2. $\therefore$ Mode = **2**. (b) Highest score 4, lowest score 1: $4 - 1 = 3$. $\therefore$ Range = **3**.

Q4. (a) Sort the data: 20, 20, 20, 25, 30, 30, 30, 35. Both 20 and 30 appear three times, and no value appears more. $\therefore$ There are **two modes, 20 and 30**. (b) There is no single most common reading time — the nights pile up at two different lengths, 20 minutes and 30 minutes.

Q5. (a) Highest jump 3.8 m, lowest jump 3.1 m: $3.8 - 3.1 = 0.7$. $\therefore$ Range = **0.7 m**. (b) **Continuous** — a jump length is measured, and a jump could land at any value, like 3.45 m, even though these results are recorded to one decimal place.

Q6. (a) Brown was chosen by 9 students, more than any other colour. ∴ Mode = **brown**. (b) Eye colour is nominal data: the answers are labels, not numbers. There is no highest label or lowest label, and subtracting one colour from another means nothing, so the range is not defined for this data set.

Q7. (a) **Flat** — stacks of about the same height, no peak. (b) **Even on both sides** — the peak in the middle with matching halves. (c) **Lopsided** — the pile at the low end with a tail running right.

Q8. (a) Column A records the **number of books read in Term 2** — one row for each value from 2 to 10. (b) The row is kept because the 0 is information: no Year 5 student read 10 books. If the row were dropped, the two classes would no longer line up value by value, and the side-by-side graph would compare the wrong rows. (c) Select **all three columns** — the values column and both classes' counts — **with the heading row**, so the graph can label each series.

Q9. (a) $8+6+4+3+2+1=24$ students. (b) The tallest column is AFL football at 8. ∴ Mode = **AFL football**. (c) Soccer 6 and cricket 4: $6+4=10$ students.

Q10. (a) Add the votes at each rating: $2+4+7+11+6=30$. (b) The tallest stack is at rating 4 (11 votes). ∴ Mode = **4**. (c) Rating 4 got 11 votes and rating 5 got 6: $11+6=17$ students.

Q11. (a) Temperature is **measured** — it can take any value in a range — so it is **continuous** numerical. (b) Text messages are **counted** in whole numbers, so the variable is **discrete** numerical. (c) Bag brands are **labels with no order**, so the variable is **nominal** categorical. (d) The answers are **labels in order** — agree a lot, agree a little, disagree — so the variable is **ordinal** categorical.

Q12. (a) Count the values: 0 once, 1 twice, 2 four times, 3 twice, 4 once, 5 twice. The value that appears most often is 2. ∴ Mode = **2**. (b) Highest 5, lowest 0: $5-0=5$. ∴ Range = **5**. (c) The peak sits at 2, near the low end, and the stacks fall away towards 5 — the distribution is **lopsided**, with its tail running to the higher values.

Q13. (a) Favourite colour is **nominal** data — labels with no order and no numbers in them. There is no highest colour or lowest colour, and "red − blue" is not a calculation anyone can do. The range only exists for numerical data. (b) Sofia is **right**. The highest value is 4 and the lowest value is 4, so $4-4=0$. A range of 0 says every value in the set is the same — there is no spread at all.

Q14. (a) In Set A, 1 appears twice and 2 appears twice, while 3 and 4 appear once each. ∴ Set A has **two modes, 1 and 2**. (b) In Set B, every value — 1, 2, 3 and 4 — appears exactly twice. No value appears more often than the others, so there is **no mode**. In a dot plot the stacks would all be the same height: a **flat** shape.

Q15. (a) For Years 3–4 the biggest count is 12, for walking. ∴ Mode = **walk**. (b) For Years 5–6 the biggest count is 14, for the bus. ∴ Mode = **bus**. (c) Combine the groups: walk $12+5=17$, bus $6+14=20$, car $9+4=13$, bike $3+7=10$. ∴ Combined mode = **bus**. (d) Grouping shows that the two year levels travel in different ways — walking is the mode for Years 3–4, while the bus is the mode for Years 5–6. The combined mode "bus" hides the walking majority in Years 3–4.

Q16. (a) Highest 28.4 °C, lowest 18.2 °C: $28.4-18.2=10.2$. ∴ Range = **10.2 °C**. (b) The new highest is 38.9 °C and the lowest is still 18.2 °C: $38.9-18.2=20.7$. ∴ New range = **20.7 °C**. (c) The range doubled because of **one** very hot day. The range looks only at the two end values, so a single extreme value can stretch it a long way.

Q17. (a) Melbourne Airport's wettest month is **November** at 62.4 mm. (Mildura's wettest month is November too, but only 29.4 mm.) (b) Melbourne Airport: highest 62.4, lowest 35.0, so $62.4-35.0=27.4$. Range = **27.4 mm**. Mildura Airport: highest 29.4, lowest 18.7, so $29.4-18.7=10.7$. Range = **10.7 mm**. (c) **Melbourne Airport** — its range of 27.4 mm is much wider than Mildura's 10.7 mm, so its rainfall changes more from month to month. (d) Melbourne's rainfall is **nearly flat** across the year — every month is somewhere between 35.0 and 62.4 mm — with a gentle rise to a spring peak around November.

Q18. (a) **January**. January sits in the middle of summer in Australia, so both towns peak then. (b) Melbourne Airport: $26.7-13.2=13.5$. Range = **13.5 °C**. Mildura Airport: $32.6-15.5=17.1$. Range = **17.1 °C**. (c) **Mildura Airport**, because its range is wider: 17.1 °C against 13.5 °C. Mildura's summers are far hotter while its winters stay close to Melbourne's. (d) Each town's year is a **hill shape** — temperatures climb to the January peak and fall away again — and Mildura's hill is the taller one.

Q19. (a) **Column graph** — the subjects are nominal labels with no order, one column per subject. (b) **Dot plot** — the ratings are ordinal, so the labels run in order from 1 to 5 and one dot per vote shows the shape. (c) **Side-by-side column graph** — two towns are being compared month by month, so each month needs one column per town on the same scale.

Q20. (a) Design 1: $3.0 - 1.8 = 1.2$. Range = **1.2 s**. Design 2: $4.6 - 0.9 = 3.7$. Range = **3.7 s**. (b) **Design 1**. Its range is the smaller one, which means its throws sit close together — nearly all of them land around 2.3 s. (c) Any reasoned choice. For example: **Design 2**, because it has already produced the longest single flight of 4.6 s, and a contest only needs one long throw. Or **Design 1**, because its small range means the team can count on a flight near 2.3 s rather than risk a short one.

Q21. Any correct pair of dot plots. Check each one by standing back from it. For the even one, the peak should sit in the middle with roughly matching halves — for example the values 1, 2, 2, 3, 3, 3, 4, 4, 5 give stacks of 1, 2, 3, 2 and 1 dots, matching on both sides. For the lopsided one, pile most values at one end — for example 0, 0, 0, 1, 1, 2, 3, 4, 5 — so the peak sits at 0 and a tail runs to the right. Label each plot with its shape.

Q22. (a) The **mode** is the same: both classes have 30 minutes as the most common homework time. (b) The **range** is different: 40 minutes for Class A against 10 minutes for Class B. (c) **Class A** — the wider range means its times stretch further out from the mode.

Q23. (a) Set P: $5 - 5 = 0$. Set Q: $7 - 3 = 4$. Set R: $9 - 1 = 8$. (b) From least spread to most spread: **P, Q, R** (ranges 0, 4, 8). (c) The mode does not measure **spread**. All three sets pile up at 5, but Set P has no spread at all while Set R stretches across 8 — the mode cannot see any of that.

Q24. (a) One correct layout: a heading row reading *Number of pets*, *Class 6W*, *Class 6K*; then one row for each value 0 to 5, with each class's count beside it. Every value gets a row, so the 0s stay in place and the classes line up. (b) Steps, in order (wording will vary): enter the headings and the data as in part (a); select all three columns including the headings; choose a column graph from the spreadsheet or graphing software; give the graph a title, such as "Number of pets at home", and axis labels, such as "Number of pets" and "Number of students". Check that two columns appear at each pet value, one per class. (c) Class 6W: the tallest stack is at 1 pet (6 students) → mode **1**; highest value 4, lowest 0, so range $4 - 0 = 4$ → range **4**. Class 6K: the tallest stack is at 2 pets (6 students) → mode **2**; highest value 5, lowest 0, so range $5 - 0 = 5$ → range **5**.

Q25. Nominal (labels with no order): eye colour, favourite subject, type of pet. **Ordinal** (labels in order): medal won (gold, silver, bronze — the order matters), level of a video game (easy, medium, hard), size of a drink (small, medium, large). The test: can the labels be put in a sensible order?

Q26. Discrete (counted): number of goals in a season, number of eggs in a carton, number of people in a car. **Continuous** (measured): time to ride to school, length of the classroom, mass of a pumpkin. The test: count or measure?

Q27. (a) Years 3–4: **beach** (9 votes). Years 5–6: **museum** (8 votes). (b) Add the groups: beach $9 + 5 = 14$, wildlife park $7 + 6 = 13$, museum $4 + 8 = 12$, sports stadium $2 + 7 = 9$. ∴ Overall mode = **beach**. (c) The older students voted for the museum as their mode. Booking the beach would go against the first choice of the whole upper group — a reason to look at the grouped results, or to ask the question another way, before deciding.

Q28. Examples will vary; each answer is checked against its type. For example: (a) Ordinal: "Rate the new playground from 1 to 5." The answers are labels in order. (b) Nominal: "What is your favourite lunch to buy at the canteen?" The answers are labels with no order. (c) Discrete numerical: "How many times did you visit the library this term?" The answers are counted whole numbers. (d) A dot plot suits the ordinal question, because the ratings run in order; a column graph suits the nominal question, because the labels are names; a dot plot or column graph suits the discrete question, because the answers are counted values.

Q29. (a) Team A: 8 appears three times, more than any other score → mode **8**; highest 10, lowest 6, so $10 - 6 = 4$ → range **4**. Team B: 8 appears three times, more than any other score → mode **8**; highest 14, lowest 2, so $14 - 2 = 12$ → range **12**. (b) Both distributions are **even on both sides**, with the peak at 8 and the values falling away on each side of it. Team B's hill is the wider one — it stretches from 2 to 14, while Team A stays between 6 and 10. (c) Example: "Both teams scored 8 points most often, and both distributions are even on both sides, so their shapes match. But

Team B's scores are far more spread out — its range is 12 against Team A's 4 — so Team B had brilliant shots and poor ones, while Team A's shots stayed close together."

What comes next

Next in Chapter 7, subtopic **7B** turns these skills around and points them at the media: you will learn to spot the tricks that make displays mislead — broken axes, scales that do not start at 0, and claims the numbers do not support (VC2M6ST02). Then in **7C** you run the whole show yourself: pose and sharpen a question, collect the data, and report your findings with the comparative displays and measures from this subtopic (VC2M6ST03).

In Year 7, your comparison toolkit grows again. You will meet two new numbers that describe the middle of a data set (VC2M7ST01) and a new plot that breaks each value into its digits (VC2M7ST02). The mode, range and shape you have learned here carry straight into that work — and into every science investigation that needs a fair graph.

Statistics in the media

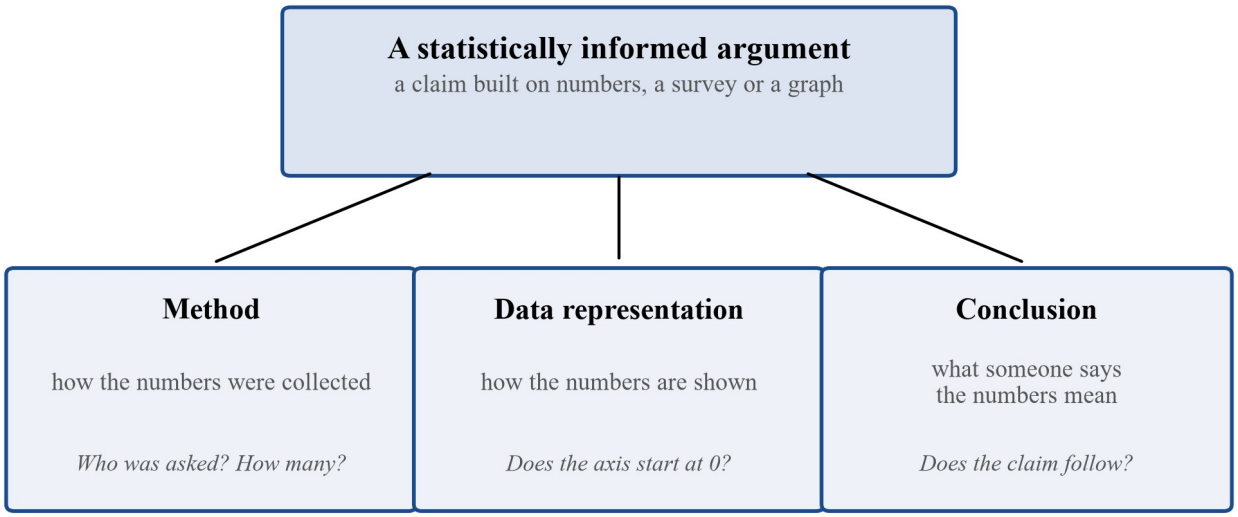
identify statistically informed arguments presented in traditional and digital media; discuss and critique methods, data representations and conclusions — VC2M6ST02

Theory

The news is full of numbers. A headline says crime has doubled. A television graph says one team is far ahead. A social media post says 9 out of 10 people agree. Each of these is a **statistically informed argument** — a claim that is built on numbers, a survey or a graph. You meet them in **traditional media** (newspapers, television and radio) and in **digital media** (websites and social media). In this subtopic you learn to spot these claims and to test them.

Three things to check. Every statistically informed argument has three parts, and each one can go wrong.

Part	What it is	Questions to ask
Method	how the numbers were collected	Who was asked or counted? How many? From which date to which date? Was it a fair way to collect the numbers?
Data representation	how the numbers are shown — a graph, chart, table or picture	Does the axis start at 0? Are the jumps on the scale equal? Is the picture drawn to scale? Do the pie chart slices add to 100%?
Conclusion	what someone says the numbers mean	Does the claim follow from the numbers, or does it say more than the numbers show?



Each part can go wrong — a critique checks all three.

A map of the three parts of a statistically informed argument: method, how the numbers were collected; data representation, how the numbers are shown; and conclusion, what someone says the numbers mean, each with a question to ask

Ask about the message. Every graph and headline is made by someone, for a reason (`VC2M6ST02.E01`). Before you believe a display, ask: *what message does the maker want me to take away?* A shop wants you to buy. A team

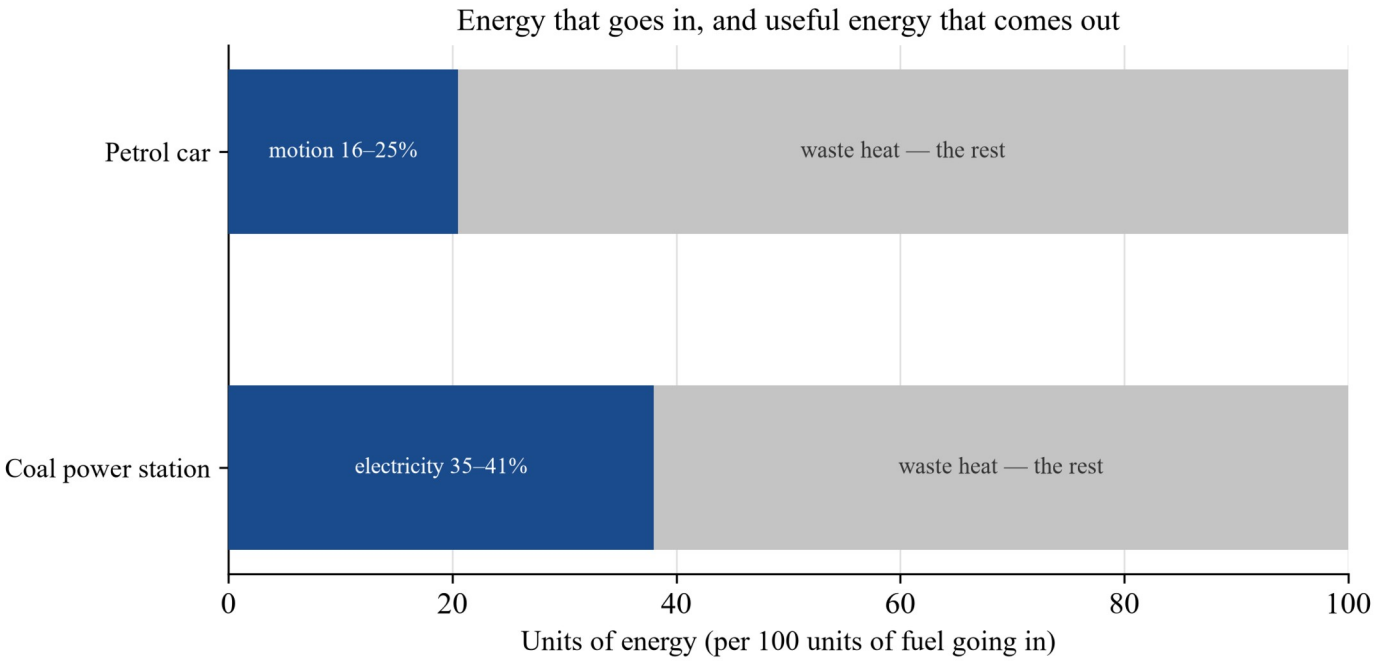
wants your support. Sometimes the message is honest and sometimes it is not — either way, you read the display better when you know what it is trying to do.

Five faults to look for. Most misleading displays use one of five tricks. Learn the names, and you can say exactly what is wrong.

Fault	What happens	The quick check
Broken axis	part of the scale is cut away, marked with a zigzag break, so a small gap looks big	Is there a break in the axis? Which numbers are missing between the last number before the break and the first number after it?
Non-linear scale	the numbers on the axis go up in jumps of different sizes, so some bars grow too fast or too slowly	Are all the jumps on the scale the same size?
Graphic not drawn to scale	a 3D effect or a picture makes one part look too big — the picture changes but the numbers do not	Would the picture tell the same story if it were flat and plain?
Data about the wrong population	numbers about one group are used to make a claim about a different group; the **population** is the whole group the claim is about	Who was actually counted or asked? Is that the same group the claim talks about?
Pie chart whose whole is not the population	the whole pie stands for only part of the group the claim is about, or the slices do not add to 100%	What does the whole pie stand for? Do the slices add to 100%?

The pie chart test. A pie chart shows the parts of one whole. If the slices cover the whole population, they must add to 100%. If they add to less, something is left out. If they add to more, something has been counted twice — or the numbers are simply wrong. Adding the slices is a ten-second test, and it catches a surprising number of bad charts.

Reading someone else's numbers. Often the numbers in a news story were collected by someone else — a report, a survey company or a government. Numbers like this are called **secondary data** (VC2M6ST02.E02). Before you use them, ask: who collected these numbers, how, and when — and do they measure what the claim says they measure? The measure chosen can change the message. Reports about how non-renewable resources such as coal, oil and gas are distributed and used around the world are a good example: count the fuel that goes *in*, and one story appears; count the useful energy that comes *out*, and a very different story appears. You will meet this in Worked example 3.



Count the fuel that goes in, or the useful energy that comes out — the story changes.

Two bars, each standing for 100 units of energy going in: a coal power station turns 35 to 41 units into electricity and the rest leaves as waste heat; a petrol car turns 16 to 25 units into motion and the rest leaves as waste heat

A critique is not an insult. A graph can mislead on purpose, or by accident. Your job in a critique is not to call anyone names. It is to **name the fault**, say what it does to the story, and say what a fair version would look like. Fact-check groups do this every day, testing claims made by every side of politics.

Words for this subtopic, with what they mean here:

Term	Meaning here
statistically informed argument	a claim built on numbers, a survey or a graph
traditional media	newspapers, television and radio
digital media	websites and social media
method	how the numbers were collected — who was asked, how many, over what time
data representation	the way numbers are shown — a graph, chart, table or picture
conclusion	what someone says the numbers mean
population	the whole group a claim is about
scale	the numbers marked along the side or bottom of a graph
broken axis	an axis with a break in it that leaves part of the scale out
secondary data	numbers collected by someone else, which you read about in a report or the media

In subtopic 7C you turn these habits around and use them on your own work: you plan an investigation, collect the data, and present it so that no one can point to any of these five faults in your display. The same skill keeps going after school — senior mathematics has a whole topic built on spotting distorted data, and every science subject needs it when a claim has to be tested.

Worked examples

Example 1 — with help

On 3 April 2024, Peter Dutton, the leader of the party not in government, said that personal income tax had risen by a "record 23 per cent" under the government of the day. RMIT ABC Fact Check tested the claim and published its result on 11 April 2024. Use the three-part check — method, representation, conclusion — to see what the fact check found.

Source: ABC News, 'Fact check: Have personal income taxes risen by a "record" 23 per cent under Anthony Albanese?', 11 April 2024. Retrieved 19 August 2026.

Working	Comment
The 23% compares tax collected in March 2022 with tax collected in December 2023.	First, find the method: which numbers were used, and from when to when?
The federal election was in May 2022, and a new government took office. So March and April 2022 belong to the *previous* government.	The start date of the comparison is the heart of the problem.
Measured from June 2022, after the new government took office, the rise was 20%, not 23%.	A fairer start date gives a smaller rise.
Now check the word "record". Two earlier governments saw tax rise even more in their first 18 months: 107.2% and 41.7%.	The conclusion is checked against other cases, not taken on trust.
23% is less than 41.7% and far less than 107.2%, so the rise was not a record.	The claim says more than the numbers support.

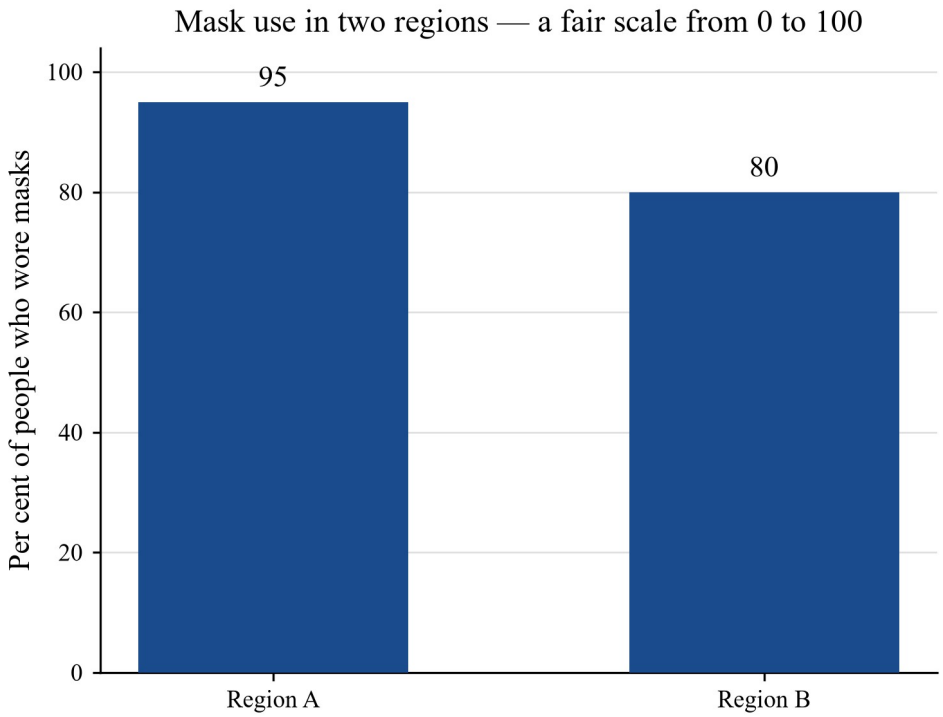
The fact check judged the claim overblown: the start date was chosen in a way that helped the claim, and "record" was wrong. Notice that nothing was made up — the critique only used the numbers and the dates.

Example 2 — with help

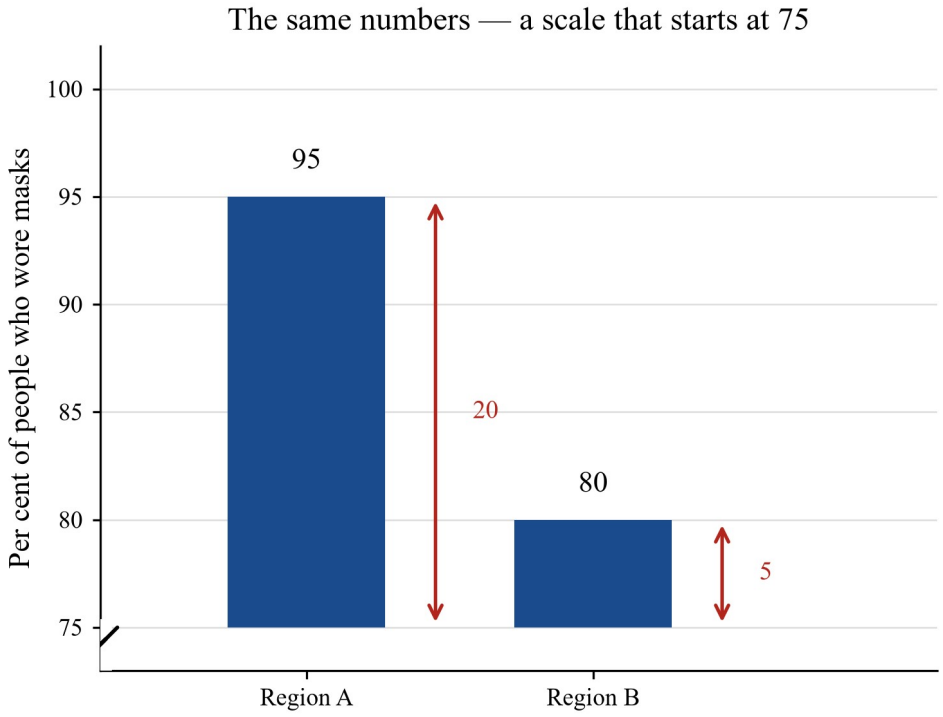
In December 2020, a television station called N1 showed a graph about how many people wore masks in different parts of Croatia. The Conversation reported that the graph's scale started at 75% instead of 0%, which made small differences look big. Here is a smaller example of the same fault. In Region A, 95 out of every 100 people wore masks. In Region B, 80 out of every 100 did.

Source: The Conversation, 'What children can teach governments about making graphs', 19 December 2020. Retrieved 19 August 2026.

Working	Comment
On a fair scale from 0 to 100, the two bars stand at 95 and 80.	Both bars are long. The regions look close, because they are.
The gap between them is $95 - 80 = 15$ points.	The real difference, found by subtraction.
On a scale starting at 75, the bars only show what sticks out above 75: $95 - 75 = 20$ and $80 - 75 = 5$.	This works like a broken axis — everything below 75 is left out.
Now one bar looks much taller than the other, even though the gap is still just 15 points.	Starting the scale high stretches small gaps until they look like big ones.



A bar graph of mask use on a fair scale from 0 to 100: the bar for Region A stands at 95 and the bar for Region B at 80, so the two regions look close



The same two bars on a scale that starts at 75: the bar for Region A rises 20 units above 75 and the bar for Region B rises only 5, so the gap of 15 points looks big

The numbers never changed — only the axis did. That is the giveaway: when a scale does not start at 0, check what the picture does to the gaps before you believe it.

Example 3 — on your own

An article in The Conversation checks a kind of chart you see a lot. The chart measures *primary energy* — all the coal, oil and gas that goes into the system. It shows that more than 90% of Australia's raw energy supply still comes from fossil fuels, and it makes the switch to clean energy look almost impossible. The author points out that this measure counts energy that is never used. A coal-fired power station turns only 35–41% of the energy in its coal into electricity; the rest leaves as waste heat. A petrol car turns only 16–25% of its fuel's energy into motion.

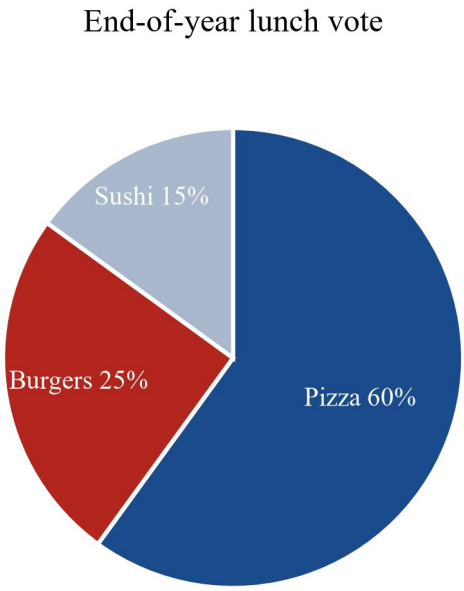
Source: The Conversation, ‘This widely used chart makes the clean energy switch seem much harder than it actually is’, 18 November 2025. Retrieved 19 August 2026.

(a) What message does the primary-energy chart send?
(b) Name one reason the fossil-fuel part of the chart looks too big.
(c) What would a fairer chart measure instead?

The chart sends the message that switching away from fossil fuels is almost too hard to do. The fossil-fuel part looks too big because the chart counts all the wasted heat — the coal’s energy that never becomes electricity, and the petrol’s energy that never moves the car. A fairer chart would measure *useful energy*: the energy that actually heats homes, moves vehicles and runs devices. The message of that chart is very different, because electric systems waste far less. This is the `VC2M6ST02.E02` idea in action: with secondary data about non-renewable resources, *the measure chosen can change the story*, so always ask which measure was used and why.

Example 4 — on your own

A school newsletter prints a pie chart with the headline: “Students at our school want pizza for the end-of-year lunch!” The slices are pizza 60%, burgers 25% and sushi 15%. But the survey only asked the 30 students in one Year 6 class.



A pie chart with three slices: pizza 60 per cent, burgers 25 per cent and sushi 15 per cent, with the whole pie standing for the 30 students in one Year 6 class

(a) Do the slices add to 100%? Is the pie chart arithmetic the problem here?
(b) Name the fault in the headline.
(c) Rewrite the headline so it is fair.

$60\% + 25\% + 15\% = 100\%$, so the pie chart arithmetic is fine. The fault is the population: the headline makes a claim about the whole school, but the data only covers one class of 30 students. The whole pie stands for that class, not for the school. A fair headline is: “Students in our Year 6 class want pizza for the end-of-year lunch.” The numbers are real — the claim just reached further than they do.

Practice questions

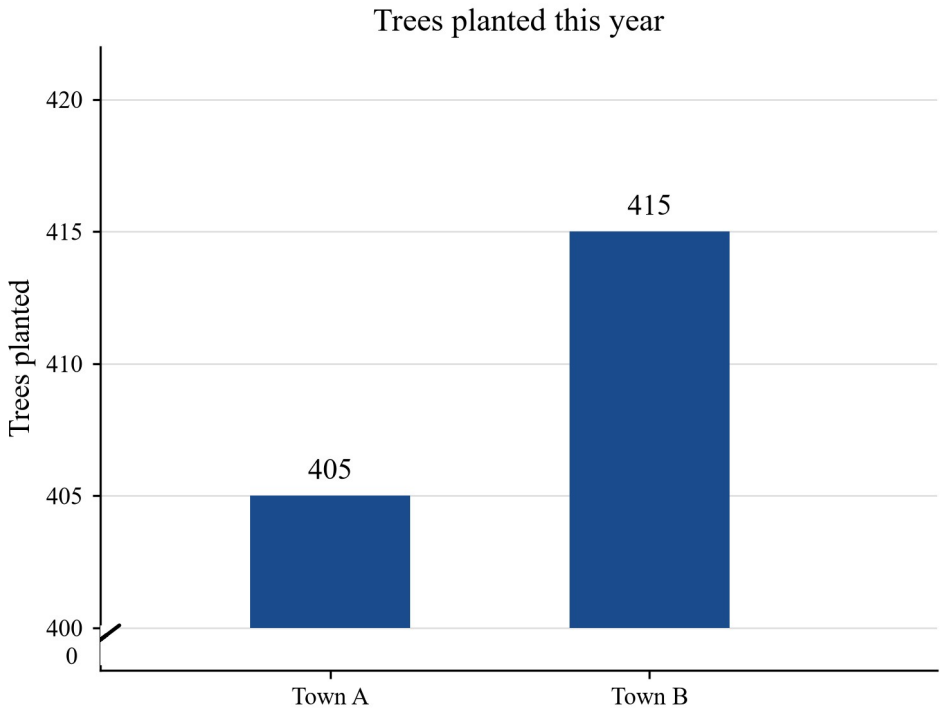
With help

Q1. Three students are critiquing the same news claim. Match each student to the part of the argument they are checking — **method**, **data representation** or **conclusion**. (a) Priya: "The survey only asked 40 people." (b) Jack: "The vertical axis starts at 90, so small gaps look big." (c) Aroha: "The numbers show prices rose a little — not doubled, like the headline says."

Q2. Match each question to the part it checks — **method**, **data representation** or **conclusion**. (a) Who was asked, and how many? (b) Does the axis start at 0, and are the jumps equal? (c) Does the claim follow from the numbers?

Q3. Four schools collected 40 kg, 45 kg, 48 kg and 50 kg of recycling. A graph of the results starts its vertical axis at 38 kg instead of 0 kg. (a) On the printed graph, the bar for 50 kg rises 12 units above the start of the axis, and the bar for 40 kg rises only 2. Why does this make the 50 kg school look like a giant next to the others? (b) What number should the axis start at, so the bars tell the truth?

Q4. Two towns report how many trees they planted this year: Town A planted 405, and Town B planted 415. A newspaper graph shows a vertical axis labelled 0, then a zigzag break, then 400, 405, 410, 415, 420.



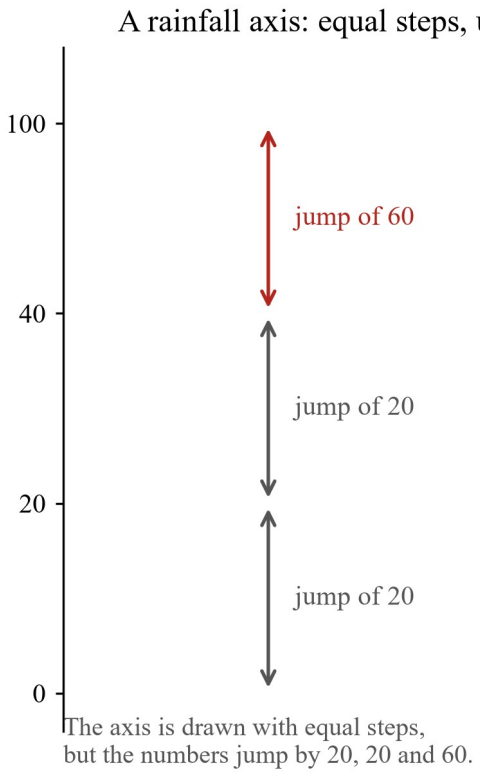
A bar graph with a broken axis: the vertical axis shows 0, then a zigzag break, then 400, 405, 410, 415 and 420; Town A's bar stands at 405 and Town B's bar at 415, but Town B's bar looks much taller

(a) Which numbers does the break leave out?

(b) How many more trees did Town B plant than Town A?

(c) On the graph, Town B's bar looks much taller. Why is that misleading?

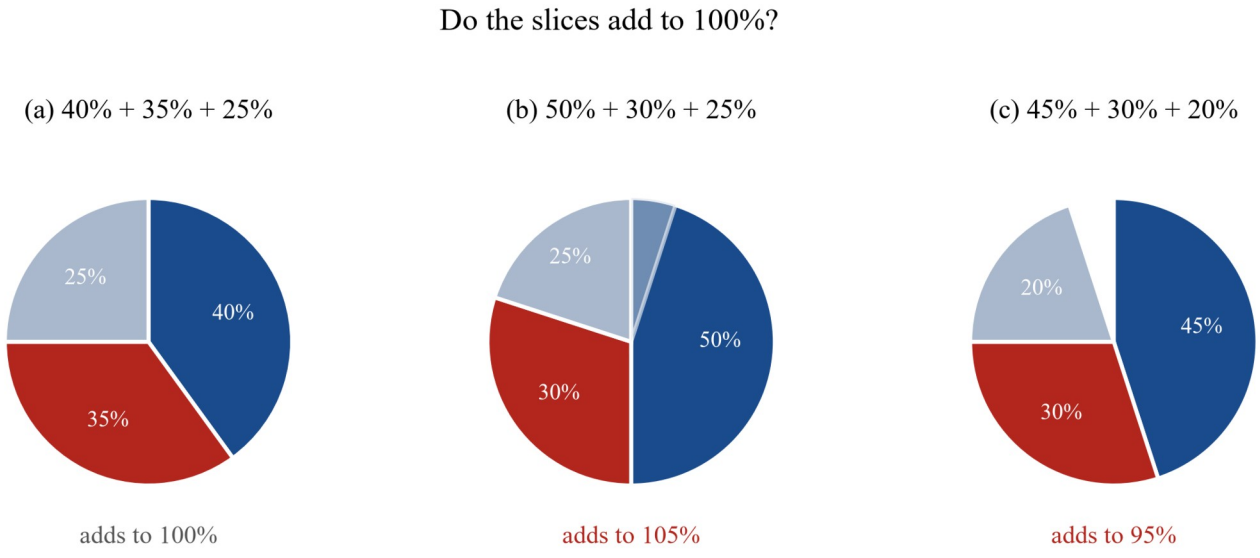
Q5. A weather website prints a bar graph of rainfall for three months. The vertical axis is labelled 0, 20, 40, 100.



A vertical axis drawn with three equal steps labelled 0, 20, 40 and 100, so the jumps on the numbers are 20, 20 and 60

(a) What are the jumps on this scale?
(b) What is wrong with them?
(c) Relabel the axis so the jumps are fair.

Q6. Three pie charts have these slices. Work out each total, and say which chart *could* be a fair picture of a whole population.



Three pie charts: the first has slices of 40, 35 and 25 per cent that close exactly; the second has slices of 50, 30 and 25 per cent adding to 105 per cent, so the last slice runs past the start of the first; and the third has slices of 45, 30 and 20 per cent adding to 95 per cent, leaving a gap

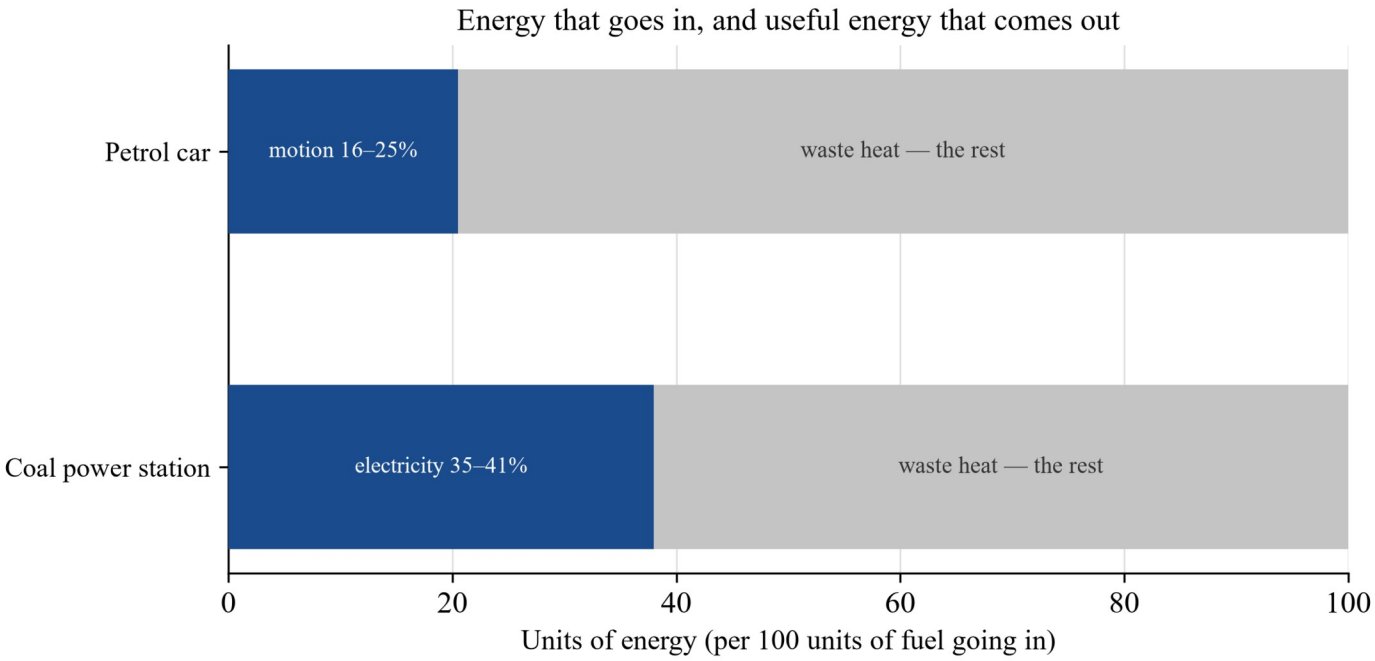
(a) 40%, 35%, 25%
(b) 50%, 30%, 25%
(c) 45%, 30%, 20%

On your own

Q7. The fact check in Worked example 1 found that two earlier governments had tax rises of 107.2% and 41.7% in their first 18 months — both more than 23%. (a) Put these three numbers in order from smallest to largest: 23%, 41.7%, 107.2%. (b) In one sentence, say why the word "record" was wrong. (c) In one sentence, say why the start date of a comparison matters.

Q8. Read the facts from Worked example 3 again.

Source: The Conversation, 'This widely used chart makes the clean energy switch seem much harder than it actually is', 18 November 2025. Retrieved 19 August 2026.



Count the fuel that goes in, or the useful energy that comes out — the story changes.

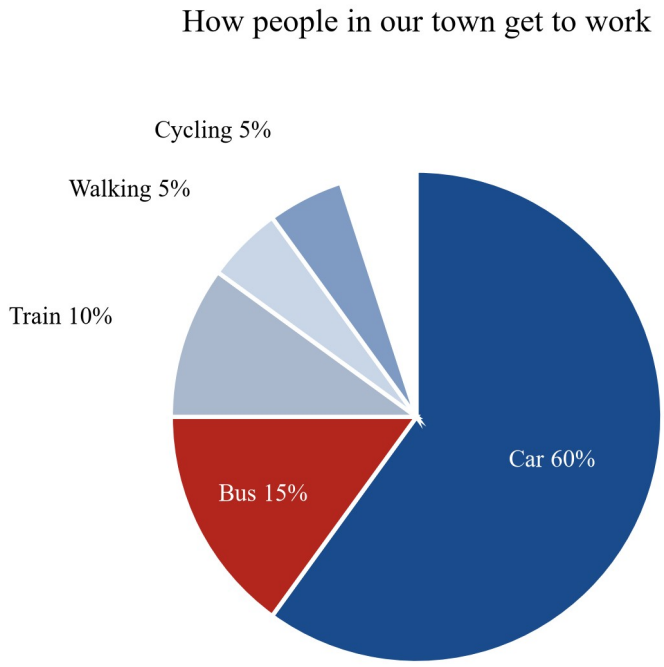
Two bars, each standing for 100 units of energy going in: a coal power station turns 35 to 41 units into electricity and the rest leaves as waste heat; a petrol car turns 16 to 25 units into motion and the rest leaves as waste heat

(a) More than 90% of Australia's raw energy supply comes from fossil fuels. What message does a chart of raw energy send about switching to clean energy?
(b) A coal-fired power station turns only 35–41% of the coal's energy into electricity. What happens to the rest?
(c) Why does counting waste heat make the fossil-fuel part of the chart look too big?
(d) What measure does the article say would tell a fairer story?

Q9. The website of an online game runs a survey. 500 players vote, and 90% say gaming should be a school subject. The website posts the headline: "Kids want gaming in school!" (a) Who was actually asked in this survey? (b) Give two reasons this survey cannot tell us what *all* Australian kids want. (c) Rewrite the headline so it is fair.

Q10. A graph shows two towns' library loans: Town A borrowed 100 books, Town B borrowed 120. The vertical axis starts at 95. (a) How many books above the start of the axis does each bar rise? (b) Explain why Town B's bar looks like a landslide when it is not. (c) Describe a fair axis for this graph.

Q11. A poster in the library shows a pie chart titled "How people in our town get to work": car 60%, bus 15%, train 10%, walking 5%, cycling 5%.



A pie chart of how people in our town get to work: car 60 per cent, bus 15 per cent, train 10 per cent, walking 5 per cent and cycling 5 per cent, with a gap where the last 5 per cent would sit

(a) Add the slices. What is wrong with this pie chart?

(b) What could the missing part of the pie be?

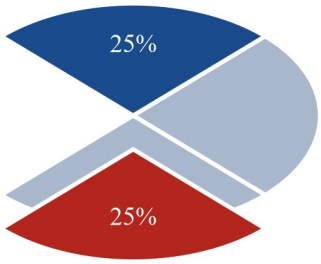
(c) The poster also says “Everyone in our town drives to work.” Give two reasons this conclusion does not follow from the chart.

Q12. An article in The Conversation says that 3D pie charts can make equal slices look unequal, and that a plain bar chart is usually a better choice.

Source: The Conversation, ‘Here’s why you should (almost) never use a pie chart for your data’, 8 January 2024. Retrieved 19 August 2026.

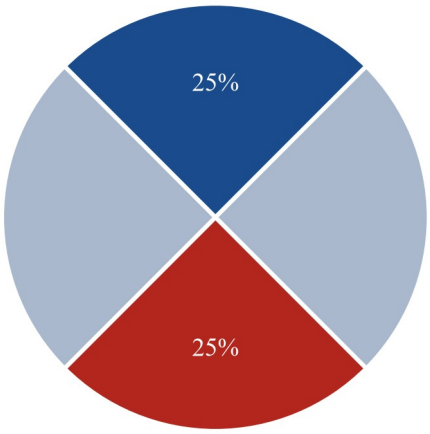
Equal slices of 25% can look unequal on a tipped 3D pie

A tipped 3D pie chart



the front slice takes up more of the picture

A flat pie chart



all four slices are the same size

Two pie charts with four slices of 25 per cent each: on the left, a tipped 3D pie chart where the front slice takes up more of the picture than the back slice, and on the right, a flat pie chart where all four slices look the same size

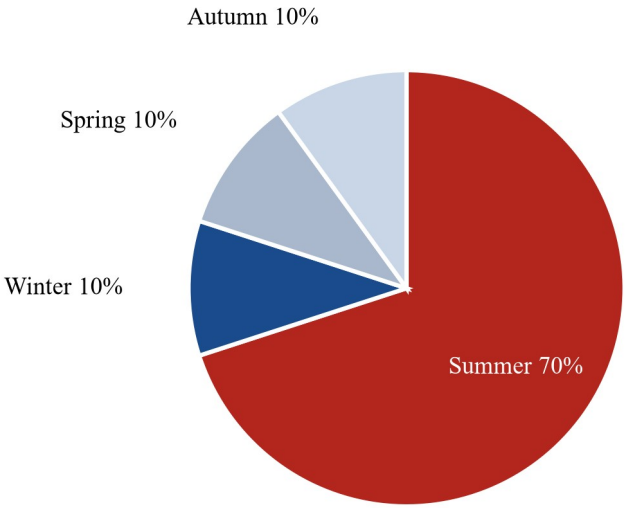
(a) A 3D pie chart has two slices of 25% each. One slice is tipped towards the viewer, and the other is tipped away. Why might they *look* different in size even though they are equal?

(b) Name the fault from the fault table that this is an example of.

(c) What plainer display does the article say usually works better?

Q13. A social media post says: "Everyone agrees — summer is the best season!" Underneath is a pie chart of votes from the poster's own followers: summer 70%, winter 10%, spring 10%, autumn 10%.

Favourite season — votes from the poster's own followers



The whole pie is the poster's followers — not everyone.

A pie chart of votes from the poster's own followers: summer 70 per cent, winter 10 per cent, spring 10 per cent and autumn 10 per cent, so the whole pie is one small group

(a) Do the slices add to 100%? Is the pie chart arithmetic the problem here?

(b) Name the fault in the claim “Everyone agrees”.

(c) Rewrite the claim so it is fair.

Q14. The Conversation reported that a Press Association line graph of coronavirus cases and deaths around the world had no usable scale on either axis — and that it drew cases and deaths on the same numbers, even though cases ran into the millions while deaths were far fewer.

Source: The Conversation, ‘What children can teach governments about making graphs’, 19 December 2020. Retrieved 19 August 2026.

(a) Why can’t you trust a graph with no scale on its axes?

(b) What happens to the deaths line when it shares a scale with the millions of cases?

(c) Write two questions you would ask before believing this graph.

Answers

Q1. (a) method; (b) data representation; (c) conclusion. **Q2.** (a) method; (b) data representation; (c) conclusion. **Q3.** (a) everything below 38 kg is left out, so the small gaps between the schools are stretched; (b) 0 kg. **Q4.** (a) the numbers from 0 up to 400; (b) 10; (c) the gap of 10 trees is shown on a stretched-out piece of axis, so it looks like a big lead. **Q5.** (a) 20, 20 and 60; (b) the jumps are not all the same size, so the bars do not grow at the same rate as the numbers; (c) for example 0, 20, 40, 60, 80, 100. **Q6.** (a) 100% — this one could be a fair picture of a whole; (b) 105% — something counted twice or wrong; (c) 95% — something left out. **Q7.** (a) 23%, 41.7%, 107.2%; (b) two earlier governments had rises of 107.2% and 41.7%, both more than 23%, so 23% is not the biggest ever; (c) choosing an earlier or later start can make the same rise look more or less than it really is. **Q8.** (a) that switching is almost too hard to do; (b) it leaves as waste heat; (c) because heat that is never used gets counted as if it were; (d) useful energy — the energy that actually heats homes, moves vehicles and runs devices. **Q9.** (a) 500 players of that one game; (b) only players of one game were asked, and only the ones who chose to vote; (c) example: "Most players in our survey want gaming in school." **Q10.** (a) Town A rises 5, Town B rises 25; (b) the axis starts at 95, so the real gap of 20 books is stretched until it looks huge; (c) start the axis at 0 and use equal jumps. **Q11.** (a) 95% — the slices do not add to 100%, so part of the population is missing; (b) any sensible answer, for example people who work from home, or ride a bike and a train; (c) 40% of the chart uses other ways to get to work, and 60% is not everyone. **Q12.** (a) the slice tipped towards the viewer takes up more of the picture, so it seems to stand for more than it really does; (b) graphic not drawn to scale; (c) a bar chart. **Q13.** (a) yes, 100% — the arithmetic is fine; (b) the data only covers the poster's followers, not everyone; (c) example: "Most of my followers say summer is the best season." **Q14.** (a) with no scale you cannot read any numbers from the graph, so you can only guess what it shows; (b) it is squashed down near the bottom and looks almost flat; (c) any two sensible questions, for example: what numbers are on the axes, and where did the data come from?

Worked solutions

Q1. Priya asks *how* the numbers were collected (only 40 people were surveyed), so she is checking the **method**. Jack asks how the numbers are *shown* (the axis starts at 90), so he is checking the **data representation**. Aroha asks whether the headline follows from the numbers, so she is checking the **conclusion**.

Q2. (a) "Who was asked, and how many?" is about collecting the numbers — **method**. (b) The axis and its jumps are about how the numbers are drawn — **data representation**. (c) "Does the claim follow?" is about what the numbers are said to mean — **conclusion**.

Q3. (a) The axis starts at 38 kg, so everything below 38 kg is left out. The bar heights above the start are $40 - 38 = 2$, $45 - 38 = 7$, $48 - 38 = 10$ and $50 - 38 = 12$. The biggest school's bar looks six times as tall as the smallest school's, even though 50 kg is only a little more than 40 kg. The small gaps between the schools are stretched until they look big. (b) The axis should start at **0 kg**, so each bar's height matches its number.

Q4. (a) The break jumps from 0 to 400, so it leaves out all the numbers in between — everything from just above 0 up to 400. (b) $415 - 405 = 10$ trees. (c) Town B planted only 10 more trees than Town A, but the graph draws that gap on a stretched-out piece of axis running from 400 to 420, where every small change looks tall. The bars look like a big lead when the lead is small. A fair graph would draw the whole axis from 0 with equal jumps, or would say clearly that the axis has been shortened.

Q5. (a) From 0 to 20 is a jump of 20. From 20 to 40 is a jump of 20. From 40 to 100 is a jump of 60. The jumps are 20, 20 and 60. (b) The jumps are not all the same size. A bar that grows by 60 mm of rain only moves up one step on the axis, while a bar that grows by 20 mm also moves up one step — so the picture hides how much more rain the last month really had. This is a **non-linear scale**. (c) Use equal jumps that cover the data, for example 0, 20, 40, 60, 80, 100.

Q6. (a) $40\% + 35\% + 25\% = 100\%$. This chart **could** be a fair picture of a whole population — the slices fill exactly one whole. (b) $50\% + 30\% + 25\% = 105\%$. The slices are more than the whole, so something is wrong — a group was counted twice, or a number is simply incorrect. (c) $45\% + 30\% + 20\% = 95\%$. The slices leave 5% of the whole out, so the pie chart does not show the full population. It needs one more slice, or a label saying what is missing.

Q7. (a) From smallest to largest: **23%, 41.7%, 107.2%**. (b) The word "record" means the biggest ever — but two earlier governments had rises of 107.2% and 41.7% in their first 18 months, so 23% is not a record. (c) The start date decides where the rise is measured from. Starting earlier or later can make the same rise look more or less than it really is — which is exactly why the fact check tested the claim with a fairer start date.

Q8. (a) The raw-energy chart makes it look as if almost everything still runs on fossil fuels, so switching away from them looks almost too hard to do. (b) The rest of the coal's energy — more than half of it — leaves the power station as **waste heat**. (c) Because the heat that is never used by anyone still gets counted. That makes the fossil-fuel part of the chart look too big — much of it is heat that never does any work. (d) **Useful energy** — the energy that really heats homes, moves vehicles and runs devices. Electric systems waste far less of it, so a useful-energy chart tells a fairer story about the switch.

Q9. (a) The survey only asked the 500 people who play that one game and chose to vote. (b) First, the survey was not asking Australian kids in general — it was only asking players of one particular game, who are likely to love gaming already. Second, only the people who chose to vote are counted; kids who did not visit the website get no say at all. So the population the survey covered is not the population the headline claims. (c) Any fair version, for example: "Most players in our survey say gaming should be a school subject."

Q10. (a) The axis starts at 95, so Town A's bar rises $100 - 95 = 5$ units, and Town B's bar rises $120 - 95 = 25$ units. (b) Town B only borrowed 20 more books than Town A, but because the axis starts at 95, the bars rise 5 units against 25 units. The real gap is stretched until Town B looks like a landslide. This is the same fault as a scale that starts too high: everything below 95 is left out. (c) A fair axis starts at **0** and goes up in equal jumps, for example 0, 30, 60, 90, 120. Then each bar's height matches its number.

Q11. (a) $60\% + 15\% + 10\% + 5\% + 5\% = 95\%$. The slices do not add to 100%, so the pie chart is missing a part of the population. Something is left out — a pie chart of a whole population must add to exactly 100%. (b) Any sensible answer: for example, people who work from home, people who ride a motorbike, or people who use more than one way to get to work. (c) First, 40% of the pie is *not* car travel — bus, train, walking and cycling add to 40%, so most people in the chart do not drive alone. Second, 60% is not "everyone": the word everyone means the whole population, and the chart itself shows four other ways to travel.

Q12. (a) In a 3D pie chart, a slice tipped towards the viewer takes up more of the picture than a slice of the same size tipped away. Your eye reads the size on the screen, not the number, so equal slices can look unequal. (b) **Graphic not drawn to scale** — the picture changes the size without the numbers changing. (c) A plain **bar chart**, where every bar stands on the same flat baseline, so the comparison is fair.

Q13. (a) $70\% + 10\% + 10\% + 10\% = 100\%$. The arithmetic is fine — the problem is somewhere else. (b) The data only covers the poster's own followers. "Everyone" means all people, but the pie's whole population is one small group that probably already agrees with the poster. This is **data about the wrong population**. (c) Any fair version, for example: "Most of my followers say summer is the best season."

Q14. (a) The scale is what lets you turn the picture into numbers. With no usable scale, you cannot read any value from the graph at all — you can only guess whether a line is going up or down, and you have no way to check the story the graph seems to tell. (b) Cases ran into the millions. On a scale big enough to hold them, the deaths line is squashed down near the bottom and looks almost flat — even if deaths were rising, the shared scale hides it. (c) Any two sensible questions, for example: *What numbers are marked on the axes? · Where did the data come from, and who collected it? · Why are cases and deaths drawn on the same scale? · What are the dates on the bottom axis?*

Planning and conducting a statistical investigation

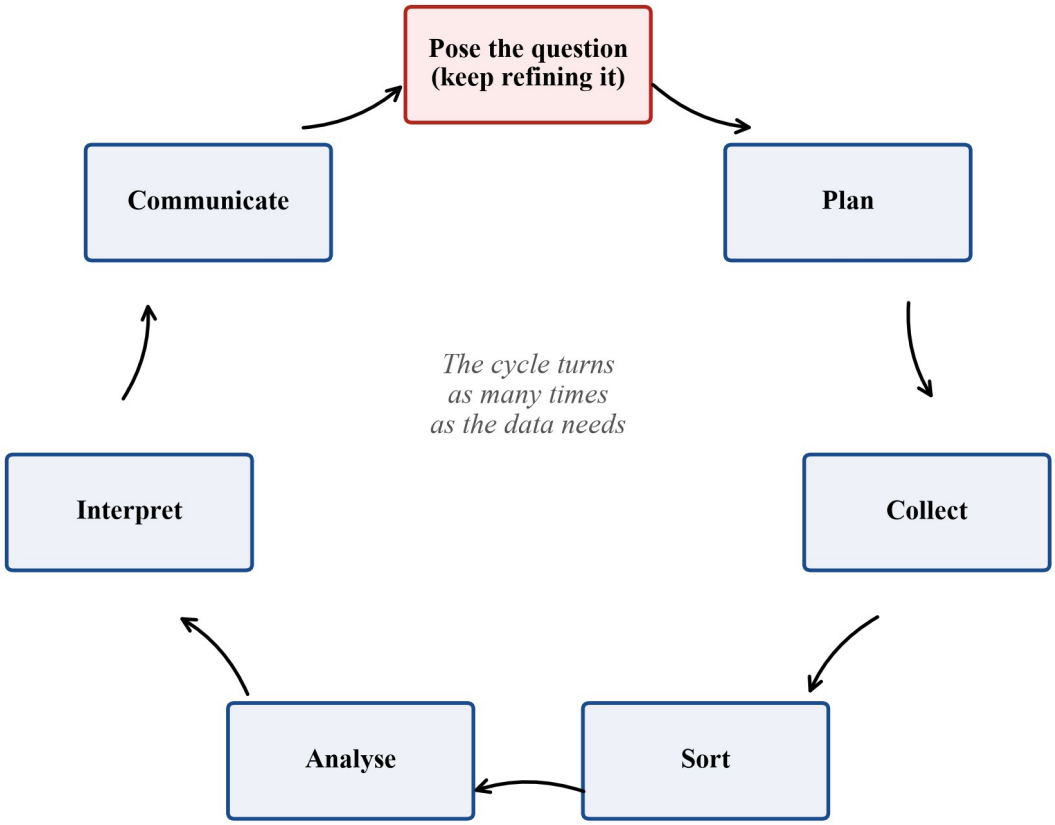
> *plan and conduct statistical investigations by posing and refining questions to collect categorical or numerical data by observation or survey, or identifying a problem and collecting relevant data; analyse and interpret the data and communicate findings within the context of the investigation* — VC2M6ST03

*In 7A you learned to compare data sets, and in 7B you learned to question them. In 7C the data is yours: you ask the question, you collect the numbers, and you report what they say — in your own words, about your own group. That whole journey has a name: a **statistical investigation**.*

Theory

The investigation cycle

A statistical investigation is not a straight line — it is a **cycle**, and going around it twice is part of the method, not a sign of failure.



A finding can send you back refining the question — that is not a failure, it is the cycle working.

A cycle of seven boxes joined clockwise by arrows: Pose the question (outlined in red, with a note to keep refining it), Plan, Collect, Sort, Analyse, Interpret, Communicate, with the note that the cycle turns as many times as the data needs

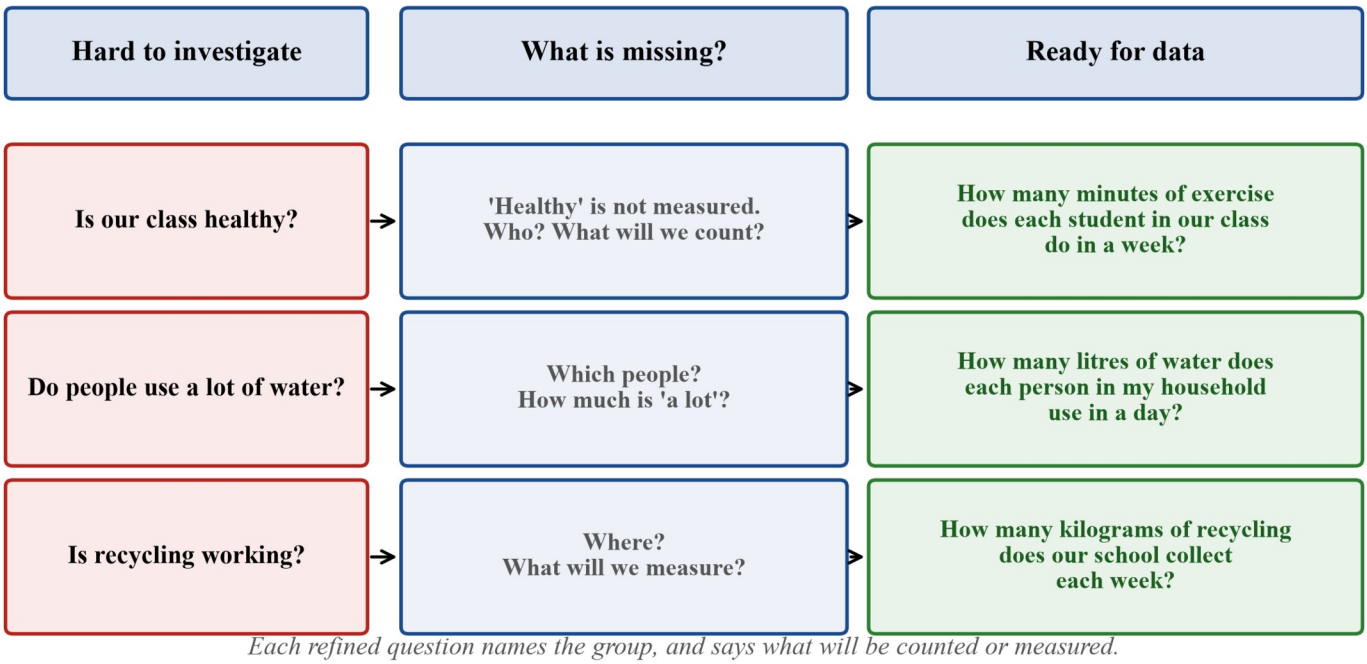
- **Pose the question** — and keep **refining** it until data can actually answer it (the red box, because everything follows from it).
- **Plan** — choose observation or survey, decide who or what gives the data, and how many.
- **Collect** — gather the data the same way every time.
- **Sort** — tidy the data into a table or spreadsheet so nothing is lost.
- **Analyse** — draw the display that suits the data type, and find the mode, the range and the shape.
- **Interpret** — say what those numbers mean about your question.

- **Communicate** — report the finding in the context of the investigation: about the group you asked or measured, in plain words.

And sometimes the last step feeds the first: a finding can send you back **refining the question**, and the cycle turns again.

Posing and refining the question

A question is ready for data when it passes two tests: it **names the group**, and it **says what will be counted or measured**. A question that fails either test cannot be investigated — yet.



Each refined question names the group, and says what will be counted or measured.

Three rows, each turning a vague question into one that is ready for data: "Is our class healthy?" becomes "How many minutes of exercise does each student in our class do in a week?"; "Do people use a lot of water?" becomes "How many litres of water does each person in my household use in a day?"; "Is recycling working?" becomes "How many kilograms of recycling does our school collect each week?"

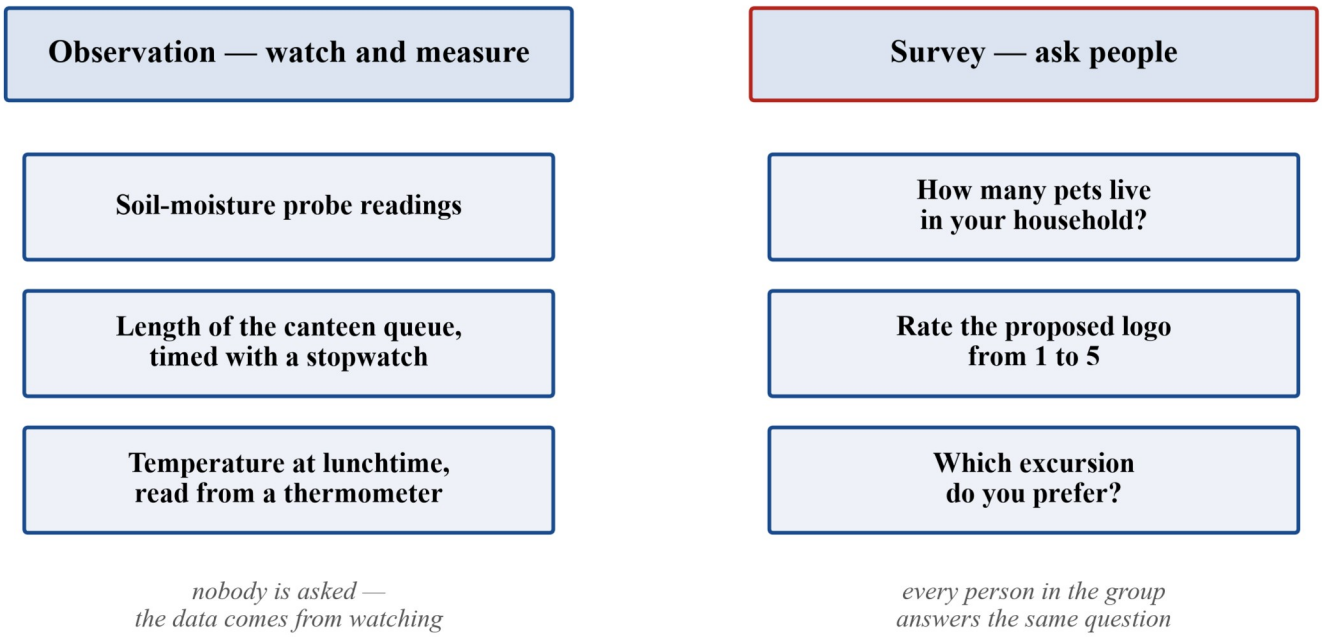
Look at what each pass of refining does. *Who?* is answered — our class, my household, our school. *What will we count or measure?* is answered — minutes, litres, kilograms. Words like "healthy", "a lot" and "working" are opinions until they are turned into something counted or measured. Refining is not changing your question into a different one; it is sharpening the question you already had until data can answer it.

The ready-question checklist

Test	Ask yourself
Names the group	Who or what will give the data — my class, my household, our school?
Says what is counted or measured	What number or label will each answer give — minutes, litres, a rating?
One question, asked the same way	Will everyone in the group answer exactly this question?

Observation or survey

There are two main ways to collect data, and choosing between them is the first planning decision.



Two panels. Observation - watch and measure: soil-moisture probe readings, the length of the canteen queue timed with a stopwatch, the temperature at lunchtime read from a thermometer; nobody is asked, the data comes from watching. Survey - ask people: how many pets live in your household, rate the proposed logo from 1 to 5, which excursion do you prefer; every person in the group answers the same question

In an **observation** nobody is asked anything — you watch, time, weigh or measure. In a **survey** you ask every person in the group the same question. The test is simple: *does someone have to tell you, or can you see it for yourself?* How students feel about a logo must be asked; how long the canteen queue takes can be timed.

The content description names a third way in: sometimes you start from **a problem that is already there** — the canteen queue is always long, the garden bed dries out every summer — and you collect the data that bears on it. The method is still observation or survey; the problem simply hands you the question to keep refining.

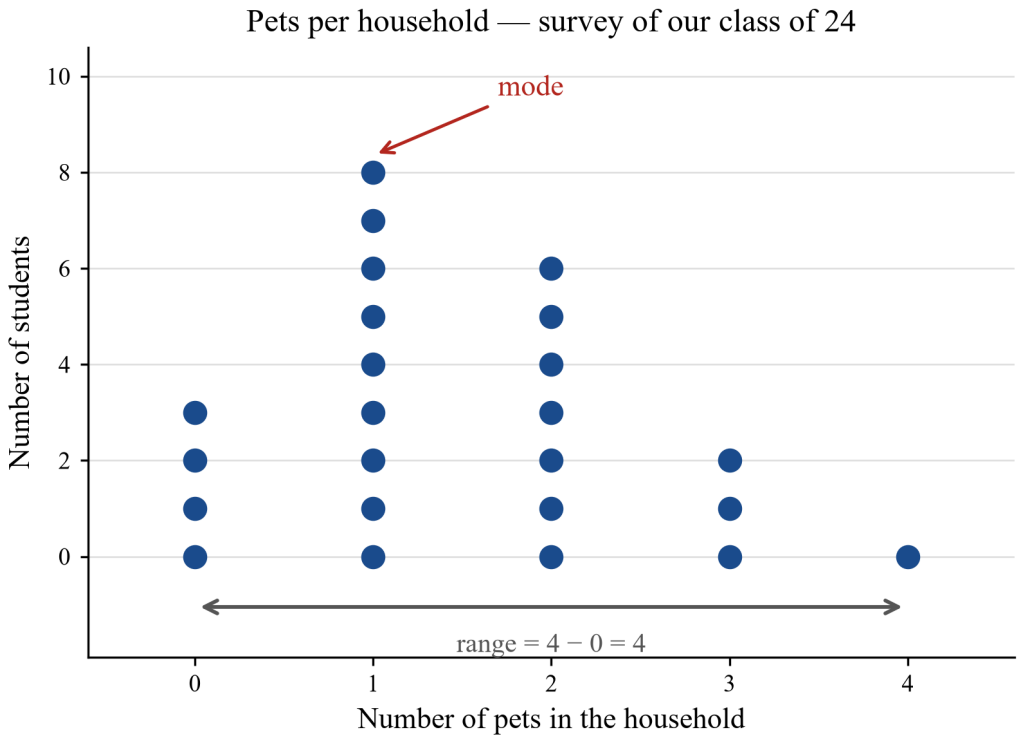
The data type decides the display

Before you collect, know what type of data your question will produce — the four types from 7A, recalled here because they decide your display and your measures.

Type	What it is	The display that suits it	What to report
Nominal	labels with no order	column graph	the mode — never a range
Ordinal	labels in order, like a 1-5 rating	dot plot with the labels in order	the mode and the shape
Discrete numerical	counted, in steps	dot plot or column graph	mode, range and shape
Continuous numerical	measured, any value in a range	dot plot of rounded values, or a line chart over time	mode, range and shape

To compare **two groups**, put both sets on the same scale — a side-by-side column graph, or two dot plots on one axis — exactly as in 7A.

Your analysis is the 7A toolkit, used on your own data: the **mode** (the most frequent value or label), the **range** (highest − lowest, for numerical data only) and the **shape** (even on both sides, lopsided, or flat). Here is a class survey analysed that way.



A dot plot of pets per household for a class of 24 students: 4 stacks of dots at 0, 9 at 1, 7 at 2, 3 at 3 and 1 at 4, with the stack at 1 marked as the mode and a bracket underneath marking range = 4 − 0 = 4

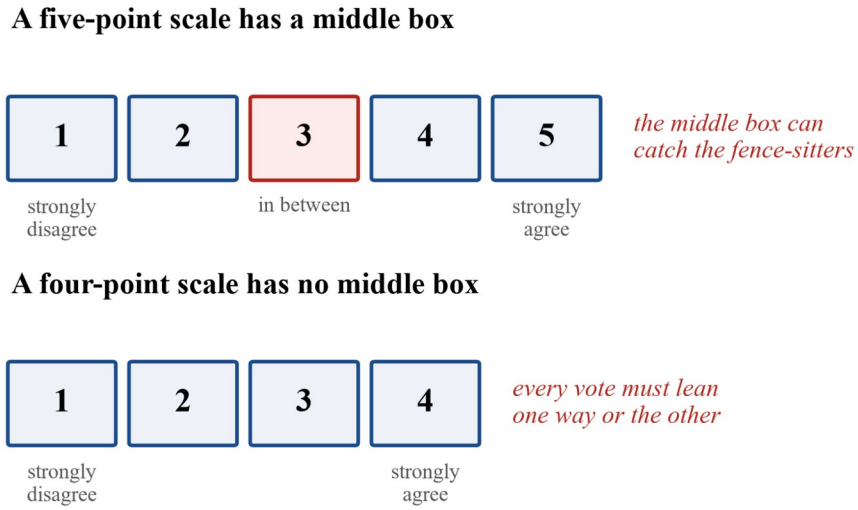
Twenty-four students answered "How many pets live in your household?" The mode is **1 pet** (the tallest stack, 9 students), the range is $4 - 0 = 4$ pets, and the shape is **lopsided** — the peak sits at the low end with a tail running to the right. Three sentences, and the whole survey is said.

Designing a fair survey

A survey is only as good as its question. Two habits keep it fair.

Ask everyone, the same way. Every person in the group answers the same question with the same choices, and the question must not push an answer. "Our canteen is great — don't you agree?" is not a question, it is a nudge. "How do you rate the canteen food, from 1 to 5?" lets the data speak.

Choose your scale on purpose. When answers are ratings, the scale itself is a design decision (`VC2M6ST03.E05`, `.E06`).



A five-point scale, boxes 1 to 5, with the middle box 3 outlined in red and labelled "in between": the middle box can catch the fence-sitters. Below it, a four-point scale, boxes 1 to 4, with no middle box: every vote must lean one way or the other

A **five-point scale** has a middle box — *not for and not against* — so nobody is forced to lean. A **four-point scale** has no middle: every vote must land on one side or the other. One scale is not more correct than the other; they measure different things. The five-point scale shows where people sit, including the undecided middle. The four-point scale shows which way the group leans once the undecided are asked to choose. Picking one without thinking about the other is a planning mistake you can now name.

Recording data in a spreadsheet

As you collect, the data goes somewhere tidy — and a **spreadsheet** is the natural home (`VC2M6ST03.E02`). One column holds each part of the record, one row holds each data point, and a heading row names the columns.

The one rule that makes a spreadsheet work: **the format of a cell has to fit the data**. A date is formatted as a date, a measurement as a number, a note as text. A number typed into a text cell is just a picture of a number — the spreadsheet cannot add it, subtract it or graph it.

	A	B	C	D
1	**Date**	**Meter reading (kL)**	**Use (L)**	**Note**
2	12/05/2026	412.920	620	laundry day
3	13/05/2026	413.500	580	rain — no garden watering
4	14/05/2026	414.160	660	car washed
5	15/05/2026	414.800	640	
6	16/05/2026	415.500	700	extra showers
7	17/05/2026	416.120	620	
8	18/05/2026	416.780	660	garden watered

format: date *format: numerical* *format: numerical* *format: text*

A reading typed as text could not be subtracted or graphed — the format has to fit the data.

A spreadsheet holding a week-long water log: column A Date, formatted as date; column B Meter reading in kilolitres, formatted numerical; column C Use in litres, formatted numerical; column D Note, formatted text; with the warning that a reading typed as text could not be subtracted or graphed

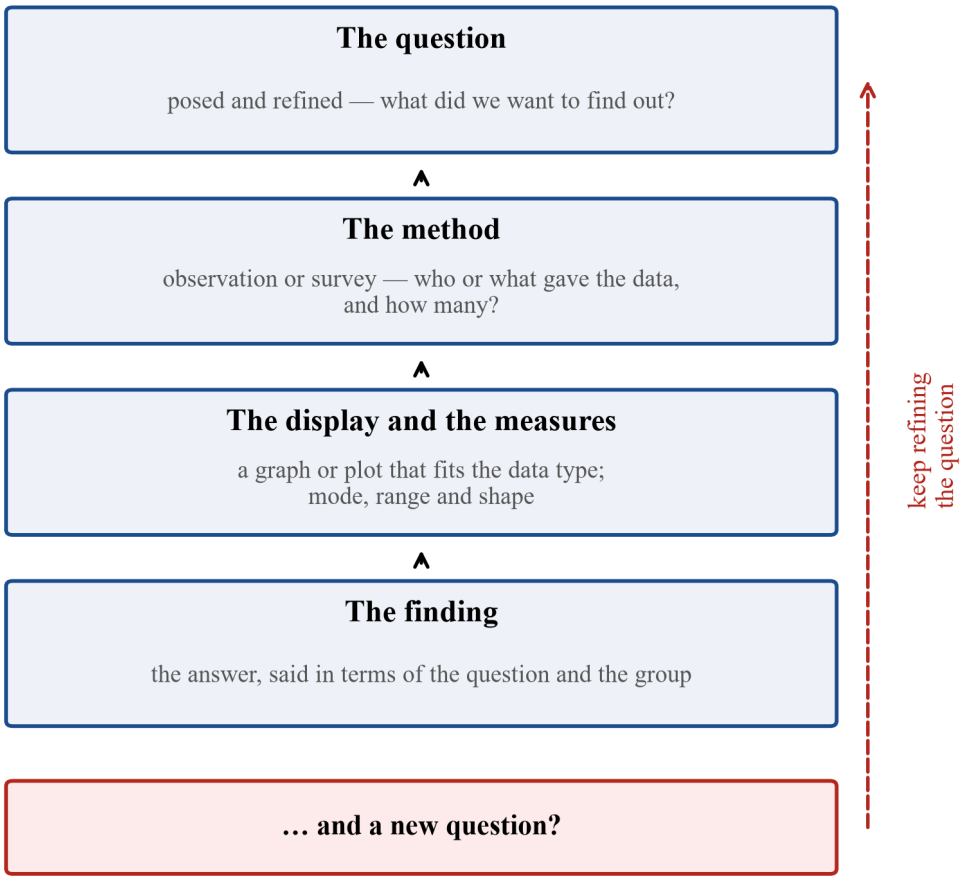
Set the formats before you type the data, and check a calculation on the first two rows before you trust the rest. Worked example 4 walks through this with the water log.

Analyse, interpret, communicate

The last three stages of the cycle are where the investigation earns its name.

- **Analyse** — build the display that suits the data type, then read off the mode, the range and the shape.
- **Interpret** — turn those three facts into meaning: *the peak at 1 pet means most households have one pet; the tail says a few households have many.*
- **Communicate** — say the finding **within the context of the investigation**: about the group you named in your question, in words a reader who was never there can follow.

A findings statement has four parts, and often a fifth.



The parts of a findings statement stacked with arrows: The question (posed and refined - what did we want to find out?), The method (observation or survey - who or what gave the data, and how many?), The display and the measures (a graph or plot that fits the data type; mode, range and shape), The finding (the answer, said in terms of the question and the group), then a red box "… and a new question?" with a dashed red arrow running back up the side labelled keep refining the question

The fifth part is the honest one: a good finding usually spots the next question. That is the cycle turning, and it is a sign the investigation worked.

Words for this subtopic

Term	Meaning here
statistical investigation	posing and refining a question, collecting and sorting data, analysing and interpreting it, and communicating the finding
pose	to put forward a question that data could answer
refining	sharpening a question until it names the group and says what will be counted or measured
observation	collecting data by watching, timing, weighing or measuring — nobody is asked
survey	collecting data by asking every person in the group the same question
scale	the ordered set of answer boxes in a rating question, like 1 to 5
five-point scale	a rating scale with a middle box, so a voter can sit neutral
four-point scale	a rating scale with no middle box, so every vote leans one way

Term	Meaning here
cell format	a spreadsheet's setting for a cell — date, numerical or text — which decides what can be done with what is typed in it
mode	the value or label that occurs most often (7A)
range	highest value − lowest value, for numerical data (7A)
shape	what the whole display looks like — even on both sides, lopsided or flat (7A)
finding	the answer to the investigated question, said in terms of the question and the group

Worked examples

Example 1 — with help

Three questions are not ready for data yet. For each, say what is missing and write a refined question that names the group and says what will be counted or measured.

1. "Are games bad for you?"
2. "Is our canteen any good?"
3. "Does recycling do anything?"

Vague question	What is missing?	Refined question
Are games bad for you?	Who? And "bad" is not measured.	How many minutes of screen time does each student in our class have on a school night?
Is our canteen any good?	Who? And "good" to whom?	How do the students in our school rate the canteen food, from 1 to 5?
Does recycling do anything?	Where? And what will we measure?	How many kilograms of recycling does our school collect each week?

Each refined question passes both checklist tests: the group is named (our class, our school), and the answer is a count or a measure (minutes, a rating, kilograms). Notice that each refined question keeps the spirit of the original — refining sharpens, it does not swap.

Example 2 — with help

Now plan the method and the display for each refined question from Example 1.

Working	Comment
Screen time: nobody can *see* someone else's screen time, so it must be asked — a **survey**.	Does someone have to tell you, or can you see it for yourself?
Minutes are counted in whole numbers — **discrete** numerical data.	Counted → discrete.
Display: a **dot plot**, one dot per student's minutes.	Discrete numerical data suits a dot plot.
Canteen rating: feelings must be asked — a **survey**, on a **five-point scale**.	The scale is chosen on purpose: the middle box is allowed.
Ratings 1-5 are **ordinal** — labels in order.	Ordinal data suits a dot plot with the labels in order.
Recycling: the bins can be weighed — an **observation**.	Nobody is asked; the scale tells you.
Mass is measured — **continuous** numerical data, one value per week.	Measured → continuous.
Display: a **column graph**, one column per week.	One measured value per week suits columns over time.

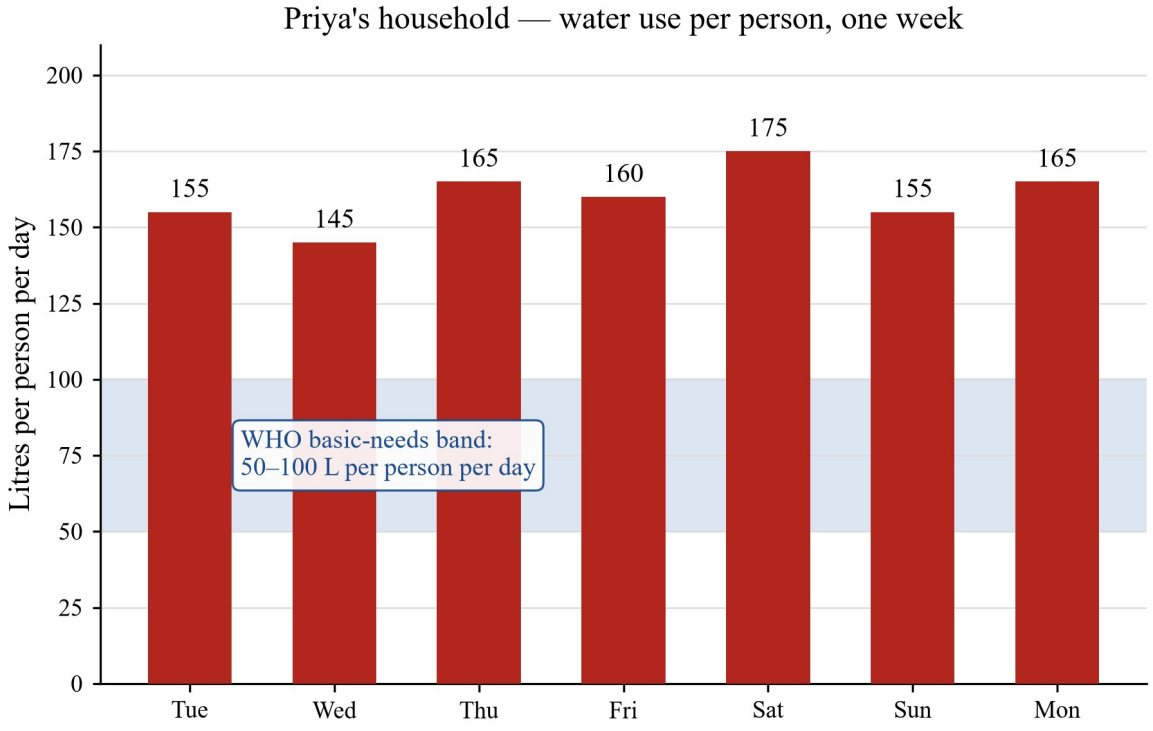
The chain is always the same: *question → method → data type → display*. Get the first three right and the display chooses itself.

Example 3 — with help

Priya's household of 4 people investigated the question: *How many litres of water does each person in my household use in a day?* (**observation** — the water meter is read, nobody is asked). Priya read the meter at 8 am each morning for a week. The use each day is the difference between two readings; 1 kL = 1 000 L.

Day	Meter reading at 8 am (kL)	Use that day (L)	Note
Monday 11 May	412.300	—	start reading
Tuesday 12 May	412.920	620	laundry day
Wednesday 13 May	413.500	580	rain — no garden watering
Thursday 14 May	414.160	660	car washed
Friday 15 May	414.800	640	
Saturday 16 May	415.500	700	extra showers
Sunday 17 May	416.120	620	
Monday 18 May	416.780	660	garden watered

(a) Check two of the daily uses by subtraction. (b) Share the week's total out equally to find the litres per person per day. (c) Read the column graph below and compare the household with the basic-needs band. (d) Write a finding in the context of the investigation.



A column graph of Priya's household water use per person per day: Tuesday 155, Wednesday 145, Thursday 165, Friday 160, Saturday 175, Sunday 155, Monday 165 litres, with a shaded band from 50 to 100 litres labelled the WHO basic-needs band

Source: World Health Organization, 'Domestic Water Quantity, Service Level and Health' (WHO/SDE/WSH/03.02, G Howard and J Bartram), 2003. Retrieved 20 August 2026.

Working	Comment
Tuesday: $412.920 - 412.300 = 0.620$ kL $= 620$ L. Wednesday: $413.500 - 412.920 = 0.580$ kL $= 580$ L.	Each day's use is the difference of two readings — this is what the spreadsheet in WE4 does.

Working	Comment
Week total: $620 + 580 + 660 + 640 + 700 + 620 + 660 = 4\,480$ L.	Every day counted once.
Shared out equally across 4 people and 7 days: $4\,480 \div 4 \div 7 = 160$ L per person per day.	Sharing out equally is division — it is not one of 7A's measures, and it is not needed again until Year 7.
The graph: every day sits **above** the 50–100 L basic-needs band; the highest day is Saturday (175 L), the lowest is Wednesday (145 L); range $175 - 145 = 30$ L.	Read the display, then the measures.
Finding: *In my household of 4, water use shared out equally was 160 L per person per day — above the 50–100 L basic-needs band on every day of the week, with the most used on Saturday, the day of the extra showers.*	The finding names the group, the measure and the comparison.

The notes column is doing quiet work: Saturday's peak lines up with "extra showers", and the lowest day is the rainy one when the garden was not watered. That is **interpretation** — the numbers and the context read together.

Example 4 — with help (digital-tool path)

Priya entered her water log in a spreadsheet, one column per part of the record. If your school permits a spreadsheet, set this up yourself; the pencil path below shows exactly what you would see.

	A	B	C	D
1	**Date**	**Meter reading (kL)**	**Use (L)**	**Note**
2	12/05/2026	412.920	620	laundry day
3	13/05/2026	413.500	580	rain — no garden watering
4	14/05/2026	414.160	660	car washed
5	15/05/2026	414.800	640	
6	16/05/2026	415.500	700	extra showers
7	17/05/2026	416.120	620	
8	18/05/2026	416.780	660	garden watered
	format: date	*format: numerical*	*format: numerical*	*format: text*

A reading typed as text could not be subtracted or graphed — the format has to fit the data.

The water log in a spreadsheet: column A Date formatted as date, column B Meter reading (kL) formatted numerical, column C Use (L) formatted numerical, column D Note formatted text, with the warning that a reading typed as text could not be subtracted or graphed

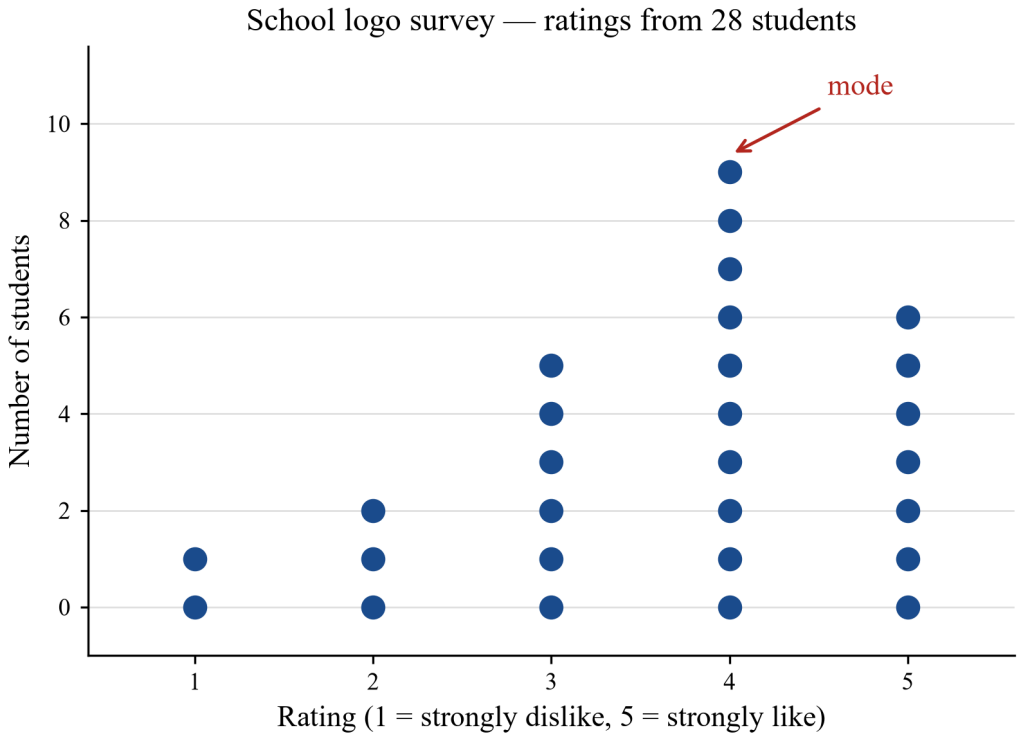
Working	Comment
Column A is formatted **date**, so 12/05/2026 is a real date the spreadsheet can order.	A date typed as general text may not sort into order.
Columns B and C are formatted **numerical**, so readings can be subtracted: B3 − B2 gives Tuesday's use.	Only a numerical cell can take part in a calculation.
Column D is formatted **text** on purpose — "laundry day" is a note, not a number.	Text is the right format for words.

Working	Comment
Now the trap: a reading typed as *text* — say "414.800" entered in a text cell — looks fine on screen but cannot be subtracted or graphed.	The format has to fit the data. A picture of a number is not a number.
Check: compute B3 − B2 by hand (0.580 kL = 580 L) and compare with the spreadsheet's answer before trusting the rest.	Check the first calculation, then trust the column.

Setting the formats **before** typing the data is the planning habit; the hand-check is the checking habit. Together they are the spreadsheet version of "collect the data the same way every time".

Example 5 — on your own

The student council showed 28 students a proposed new school logo and asked them to rate it from 1 (strongly dislike) to 5 (strongly like) — a five-point scale (`VC2M6ST03.E05`). The results are in the dot plot.



A dot plot of 28 logo ratings from 1 to 5: two dots at 1, three at 2, six at 3, ten at 4 and seven at 5, with the stack at 4 marked as the mode

(a) What type of variable is a rating from 1 to 5, and why?

(b) How many students were surveyed?

(c) What is the mode?

(d) Describe the shape, and write one sentence the council could use as its finding.

The variable is **ordinal** — the ratings are labels, but 1 to 5 puts them in order. Adding the stacks, $2+3+6+10+7=28$ students. The mode is **4** — the tallest stack, 10 students. The shape is **lopsided towards the high ratings**: the peak sits at 4, most dots are at 3 or higher, and a thin tail runs down to 1. A fair finding for the council: *Of the 28 students we asked, most rated the proposed logo 4 or 5, and the most common rating was 4 — the logo is liked.* Note what the finding does *not* say: it does not say the whole school likes it, because only 28 students were asked.

Example 6 — on your own

A class of 12 students trialled two scales on the same statement: *"Our school should start a Year 6 buddy program."* First they voted on a five-point scale, then on a four-point scale (`VC2M6ST03.E06`).

Rating	1	2	3	4	5
Five-point scale	1	1	6	2	2
Four-point scale	1	4	5	2	—

(a) On the five-point scale, what is the mode — and what does it hide?

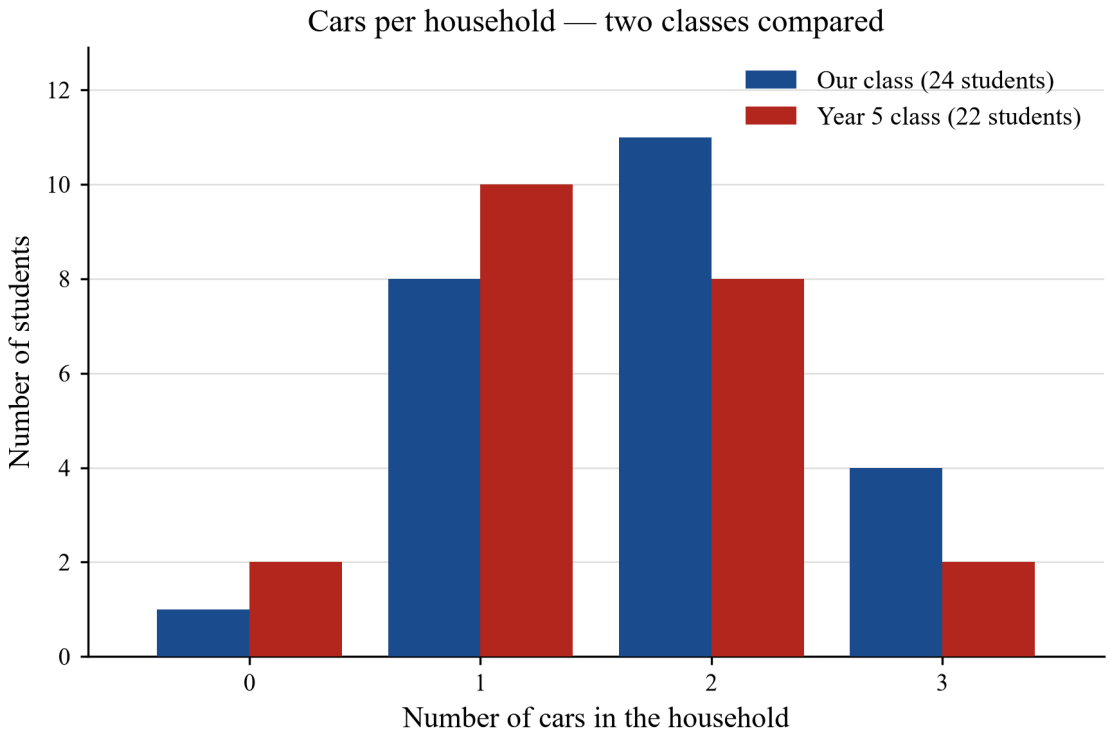
(b) On the four-point scale, count the votes on each side: against (ratings 1 and 2) and for (ratings 3 and 4).

(c) The six "in between" voters from the five-point scale had to choose a side on the four-point scale. Which way did they break, and what does that tell the principal?

(d) The mode on the five-point scale is **3 — the middle box** (6 students). It hides the lean: with the undecided sitting in the middle, the vote looks like a shrug. (b) Four-point scale: against $1+4=5$, for $5+2=7$. (c) The original leaners were 2 against (1+1) and 4 for (2+2). Once the six middle-box students had to choose, the for side grew from 4 to 7 and the against side from 2 to 5 — the middle broke **3 to 3**, and the lean that was hiding under the middle box was **for** the program, 7 to 5. A finding for the principal: *When a middle box was offered, most students sat in it; when it was taken away, the class leaned for the buddy program.* The same 12 students, two different pictures — that is why the scale is a design decision.

Example 7 — on your own

Two classes ran the same survey question: *How many cars are in your household?* Our class (24 students) and a Year 5 class (22 students) answered (`VC2M6ST03.E04`). The side-by-side column graph shows both sets on one scale.



A side-by-side column graph of cars per household for two classes: our class 1, 8, 11, 4 for 0 to 3 cars, and the Year 5 class 2, 10, 8, 2, with our class peaking at 2 cars and the Year 5 class at 1 car

(a) Find the mode for each class.

(b) Find the range for each class.

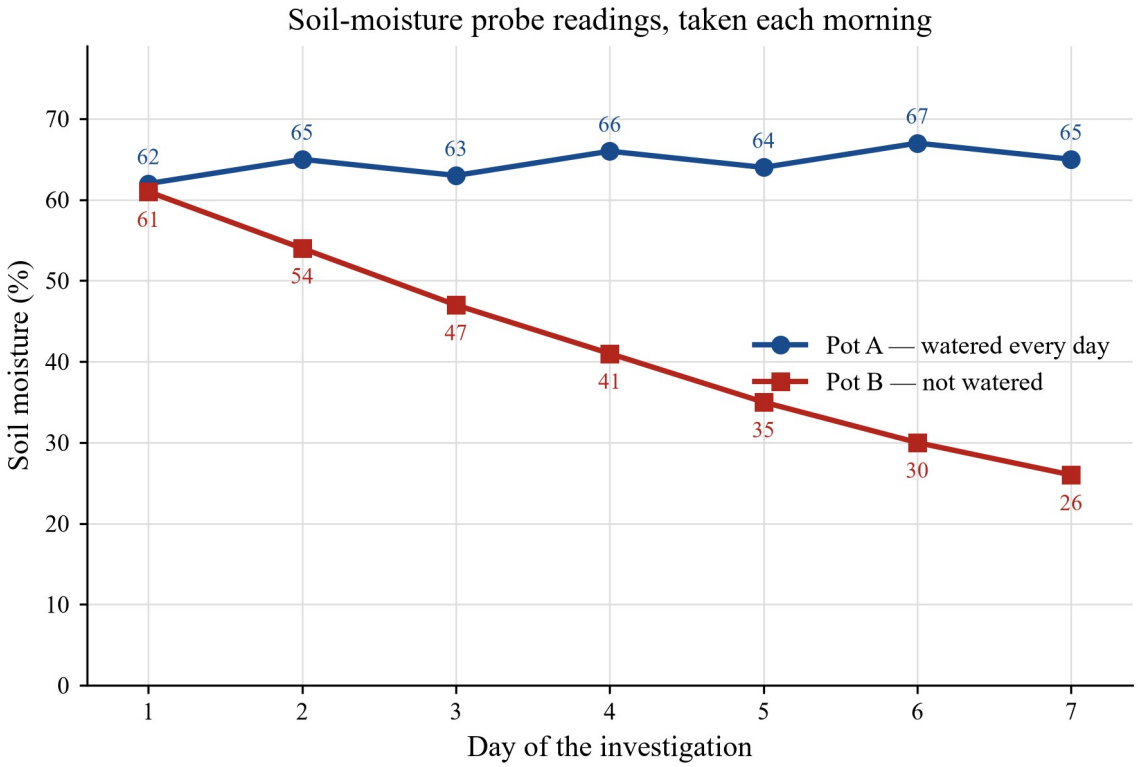
(c) Compare the two distributions in two sentences, using mode and shape.

(d) Our class: the tallest column is 2 cars (11 students) — mode **2**. Year 5: the tallest column is 1 car (10 students) — mode **1**. (b) Both classes: households ran from 0 cars to 3 cars, so the range is $3-0=3$ for **both**. (c) Example comparison: *Both classes have the same range of 3 cars and a lopsided shape peaking near the low end, but our class peaks at 2 cars while the Year 5 class peaks at 1 — most of our households have one more*

car than most Year 5 households. Same spread, different peak: that is exactly what comparing mode, range and shape is for.

Example 8 — on your own

A science class set up two identical pots of soil. Pot A was watered every day; Pot B was not watered at all. A **soil-moisture probe** (`VC2M6ST03.E01`) logged the moisture each morning, and its digital chart drew the lines below. This is **observation** — nobody is asked anything; the probe measures the same way every day.



A line chart of soil moisture over 7 days: Pot A, watered every day, stays between 62% and 67%; Pot B, not watered, keeps falling from 61% to 26%

(a) The investigation question was: *Does watering keep the soil moisture up over a week?* Which line is which pot, and how can you tell?

(b) Find the range of the readings for each pot.

(c) Describe what each line does over the week, in context.

(d) Write a finding for the question, and name one way the investigation could be refined if the cycle turns again.

(e) The flat blue line is **Pot A** (watered) — it stays high all week; the falling red line is **Pot B**. (b) Pot A: $67 - 62 = 5$ percentage points — a range of **5**. Pot B: $61 - 26 = 35$ — a range of **35**. (c) Pot A's moisture stays around the mid-60s all week, while Pot B's keeps falling, losing about 5 points a day after day 1. (d) Example finding: *Over the 7 days, the watered pot held its moisture (range 5) while the unwatered pot kept drying out (range 35) — watering keeps the soil moisture up.* A change for the next turn of the cycle: run it for two weeks, or add more pots, to check that the pattern holds (`VC2S6I02` — a fair test asks for more than one pot per treatment).

Practice questions

*Digital tools are permitted, not required, in this subtopic: where a question is marked **digital tools**, a spreadsheet or graphing software may help, and a pencil-and-paper path always works too.*

With help

Q1. Put the seven stages of the investigation cycle in order, starting from the red box in the cycle diagram: analyse · collect · communicate · interpret · plan · pose the question · sort.

Q2. Which questions are ready for data? For each one that is not, say what is missing. (a) How many brothers and sisters does each student in our class have? (b) Is sport good? (c) How many hours of sleep did each student in our class get last night?

Q3. Which question is collected by **observation**, not survey? A How many pets live in your household? B How long is the canteen queue at 12:30 each day this week? C Which excursion do you prefer — beach or museum? D How do you rate the new logo, from 1 to 5?

Q4. Name the variable type for each planned question: **nominal**, **ordinal**, **discrete** or **continuous**. (a) Rate the school fete from 1 to 5. (b) How many cars are in your household? (c) How long does a puddle on the oval take to dry? (d) What is your favourite fruit?

Q5. Match each data type to the display that suits it best: **column graph**, **dot plot with the labels in order**, **side-by-side column graph**. (a) the favourite colour of 30 students (b) ratings of the library from 1 to 5 (c) the number of books read in Term 2, for two classes at once

Q6. Read the pets dot plot in the Theory. (a) How many students were surveyed? (b) What is the mode? (c) Find the range. (d) Describe the shape.

Q7. Read the water-log spreadsheet in the Theory. (a) Which column is formatted as text, and which is formatted as date? (b) What subtraction gives Tuesday's use, and what is it in litres? (c) A meter reading typed into a text cell looks fine on screen. Why is it still a problem?

Q8. Look at the two scales in the Theory. (a) Which scale has a middle box, and what does that box let a voter do? (b) Which scale makes every vote lean one way or the other? (c) A council wants to know whether the group leans for or against a plan. Which scale shows the lean better, and why?

Q9. Make this question ready for data, with help: "Do people recycle?" Fill in the two blanks — *Who is the group?* and *What will be counted or measured?* — then write the refined question.

Q10. Read the logo dot plot in Worked example 5. (a) What is the mode? (b) How many students rated the logo 4 or 5? (c) Name the shape. (d) Why is a dot plot a good display for this data?

On your own

Q11. Write a refined, ready-for-data version of each question. (a) "Is our school green?" (b) "Do kids sleep enough?"

Q12. Choose **observation** or **survey** for each question, and give the reason. (a) The temperature at the school gate at 8:30 am each day this week. (b) The favourite subject of each student in our class. (c) How long each student takes to eat lunch. (d) How students feel about the new uniform.

Q13. Name the variable type, with the reason — *counted*, *measured*, *labels with no order*, or *labels in order*. (a) The number of library books each student borrowed this term. (b) The mass of each student's school bag. (c) A rating of the school camp from 1 to 5. (d) The type of pet in each household.

Q14. A household of 5 people reads its water meter: 128.400 kL at 8 am on Monday, 129.050 kL at 8 am on Tuesday. (a) How many litres did the household use that day? (b) Share that out equally across the 5 people. How many litres per person? (c) The basic-needs band is 50–100 L per person per day. Where does this day sit?

Q15. Spot the faults. Tom wants to know how his class feels about the canteen. He asks the four friends at his table, "Our canteen is great, right?", and writes down their answers. (a) Name two things that make this survey unfair. (b) Rewrite the question and the method so the survey is fair.

Q16. The council is asking students whether school should start half an hour later. It can use a five-point scale or a four-point scale. Choose one and give the reason for your choice.

Q17. Read the cars side-by-side column graph in Worked example 7. (a) Write down the mode for each class. (b) Write down the range for each class. (c) What do the two classes have the same, and what is different?

Q18. Read the soil-moisture chart in Worked example 8. (a) Which line is Pot B, and how do you know? (b) Find the range of Pot B's readings. (c) What does the probe do that guessing the moisture by eye could not?

Q19. The parts of a findings statement are shuffled below. Put them in the order they appear in the report diagram, and say which part is allowed to be missing. the finding · the method · the question · a new question · the display and the measures

Q20. Our class surveyed how students travel to school: walk 11, car 8, bus 3, bike 2. Write a one-sentence finding in the context of the investigation.

Q21. For each plan, choose a display and say why it suits the data type. If your school permits it, you may draw the display with a spreadsheet or graphing software (**digital tools**); a neat hand-drawn display counts just as well. (a) Ratings of the library from 1 to 5, from 30 students. (b) The favourite season of two different classes. (c) The number of goals our soccer team scored in each game this season.

Q22. A survey produced these six answers: 3, 3, 3, 3, 3, 10. (a) Find the mode. (b) Find the range. (c) Which of the two numbers tells you about the 10? What does this show the mode cannot do?

Maths in real life

Q23. A household of 3 people used 2 730 L of water in one week. (a) Share the total out equally to find the litres per person per day. (b) The basic-needs band is 50–100 L per person per day. How does this household compare? (c) Name one change the household could make, and what you would expect the next week's data to show.

Source: World Health Organization, 'Domestic Water Quantity, Service Level and Health' (WHO/SDE/WSH/03.02, G Howard and J Bartram), 2003. Retrieved 20 August 2026.

Q24. A class investigated the canteen queue by **observation**: one student timed the queue with a stopwatch at 12:30 each day for a week. The waiting times were 4, 6, 6, 9 and 12 minutes. (a) Find the mode and the range. (b) Why was this question collected by observation rather than survey? (c) Write a one-sentence finding for the canteen staff.

Q25. A (made-up) phone company asked 200 customers to rate the service from 1 to 5 on a five-point scale: 20 voted 1, 30 voted 2, 80 voted 3, 50 voted 4 and 20 voted 5. (a) What is the mode? (b) The company reads the mode as "customers are happy". Give one reason the middle box makes that reading shaky. (c) Suggest a follow-up question, and the scale you would use for it.

Choosing your strategy

Q26. Choose **observation** or **survey** for each investigation, and say what will be counted or measured. (a) Does the amount of water a plant gets change how much it grows in a month? (b) Which day of the week do students like best for sport? (c) Does the crossing queue move faster before 8:40 am or after 8:40 am?

Q27. Two students refined the same starting question, "How much water do people use?". Candidate A: "Do people use too much water?" Candidate B: "How many litres of water does each person in my household use in a day?" Choose the better refined question and say exactly which checklist test the other one fails.

Q28. Plan a whole investigation for this question: *How many hours of sleep did each student in our class get last night?* Write down, in order: the method; the variable type; the display you would draw; the two measures you would report; and a skeleton finding sentence with blanks for the results.

Q29. A class ran the soil investigation from Worked example 8 for only **two** days, and the two pots read almost the same. One student says the investigation failed. Another says the cycle just needs to turn again. (a) Which student is right, and why? (b) Name two things you could change on the next turn of the cycle.

Answers

Q1. Pose the question (keep refining it) → plan → collect → sort → analyse → interpret → communicate. **Q2.** (a) ready; (b) not ready — no group is named and "good" is not counted or measured; (c) ready. **Q3.** B. **Q4.** (a) ordinal; (b) discrete; (c) continuous; (d) nominal. **Q5.** (a) column graph; (b) dot plot with the labels in order; (c) side-by-side column graph. **Q6.** (a) 24; (b) 1 pet; (c) 4 pets; (d) lopsided — peak at the low end, tail to the right. **Q7.** (a) column D is text, column A is date; (b) $412.920 - 412.300 = 0.620$ kL = 620 L; (c) a text cell holds a picture of a number — it cannot be subtracted or graphed. **Q8.** (a) the five-point scale; it lets a voter sit neutral — not for and not against; (b) the four-point scale; (c) the four-point scale — with no middle box, every vote lands on one side, so the lean shows. **Q9.** example: group = our school; counted = kilograms of recycling per week; "How many kilograms of recycling does our school collect each week?" **Q10.** (a) 4; (b) 17; (c) lopsided towards the high ratings; (d) the ratings are ordinal, so one dot per vote stacked above each rating shows the order and the shape at once. **Q11.** examples: (a) "How many kilograms of recycling does our school collect each week?" (b) "How many hours of sleep did each student in our class get last night?" **Q12.** (a) observation — it is measured, nobody is asked; (b) survey — a favourite must be asked; (c) observation — it can be timed; (d) survey — feelings must be asked. **Q13.** (a) discrete — counted; (b) continuous — measured; (c) ordinal — labels in order; (d) nominal — labels with no order. **Q14.** (a) 650 L; (b) 130 L per person; (c) above the band. **Q15.** (a) he asked only his own table, not the whole class, and his question pushes the answer "yes"; (b) example: ask every student in the class, "How do you rate the canteen food, from 1 to 5?" **Q16.** any reasoned choice — four-point: the council wants the lean, and no middle box forces every vote to a side; or five-point: the council should see how many students are still undecided. **Q17.** (a) our class 2 cars, Year 5 class 1 car; (b) 3 cars for both; (c) same range and same lopsided shape, different mode. **Q18.** (a) the falling red line — the unwatered pot dries out; (b) 35 percentage points; (c) it measures the same thing the same way every day, giving numbers that can be compared. **Q19.** question → method → display and the measures → finding (→ a new question); the new question is the part that may be missing. **Q20.** example: *In our class of 24, walking is the most common way to school, with 11 students.* **Q21.** (a) dot plot with the labels in order — ordinal; (b) side-by-side column graph — two groups of nominal labels on one scale; (c) dot plot or column graph — discrete counts. **Q22.** (a) 3; (b) 7; (c) the range — the mode cannot see spread or one out-of-place value. **Q23.** (a) 130 L per person per day; (b) above the band on an even share; (c) example: shorter showers — the next week's bars should sit lower. **Q24.** (a) mode 6 minutes, range 8 minutes; (b) the length of a queue is measured, not asked; (c) example: *The queue most often takes about 6 minutes, but it can reach 12 minutes on a Friday.* **Q25.** (a) 3; (b) 3 is the neutral box — the mode may be a shrug, not happiness; (c) example: the same statement on a four-point scale, to see which way the middle breaks. **Q26.** (a) observation — height growth is measured in cm; (b) survey — a preference must be asked; (c) observation — queue speed is timed. **Q27.** B; A fails the "what will be counted or measured?" test — "too much" is an opinion with no number attached. **Q28.** method: survey of the whole class; variable type: discrete numerical (hours counted); display: dot plot; measures: mode and range; skeleton: *In our class, the most common amount of sleep last night was __ hours, and the answers ran from __ to __ hours.* **Q29.** (a) the second student — two days is too short to see a difference, and refining the plan is the cycle working; (b) any two of: run it for longer, add more pots, water one pot less instead of not at all, take readings twice a day.

All the working

Q1. The cycle runs clockwise from the red box: **pose the question (and keep refining it)**, then **plan** the method, **collect** the data, **sort** it into a table or spreadsheet, **analyse** with a display and the measures, **interpret** what they mean, and **communicate** the finding — with a dashed line back to the start whenever a finding sends you back refining the question.

Q2. (a) Ready: the group is named (each student in our class) and the answer is a count. (b) Not ready: no group is named, and "good" is an opinion — it needs a count or a measure, such as a rating. (c) Ready: the group is named and the answer is a count of hours.

Q3. The queue length is timed with a stopwatch — nobody is asked anything — so it is observation. Pets, excursion preference and logo ratings all have to be asked. ∴ **B**.

Q4. (a) Ratings 1-5 are labels in order — **ordinal**. (b) Cars are counted in whole numbers — **discrete**. (c) Drying time is measured, and any value is possible — **continuous**. (d) Fruits are labels with no order — **nominal**.

Q5. (a) Favourite colour is nominal — one column per label: **column graph**. (b) Ratings are ordinal — **dot plot with the labels in order**. (c) Two groups of counts on one scale — **side-by-side column graph**.

Q6. (a) $4+9+7+3+1=24$ students. (b) The tallest stack is at 1 pet (9 students) — mode **1 pet**. (c) Highest 4, lowest 0: $4-0=4$ — range **4 pets**. (d) The peak sits at the low end with a tail running right — **lopsided**.

Q7. (a) Column **D (Note)** is text; column **A (Date)** is date. (b) $412.920-412.300=0.620$ kL, and $0.620\times1\,000=620$ L. (c) Because a text cell holds a picture of a number: the spreadsheet cannot subtract it, add it or graph it, so the log quietly breaks at that row.

Q8. (a) The **five-point** scale; its middle box lets a voter sit neutral — not for and not against. (b) The **four-point** scale — with no middle, every vote must lean. (c) The **four-point** scale: the lean is exactly what the council wants, and removing the middle box forces every vote onto one side so the lean can be counted.

Q9. The group might be *our school*, and the thing counted might be *kilograms of recycling per week*. Refined: "How many kilograms of recycling does our school collect each week?" Any refined question that names a group and a count or measure passes.

Q10. (a) The tallest stack is at 4 (10 students) — mode **4**. (b) $10+7=17$ students. (c) **Lopsided towards the high ratings** — peak at 4, tail down to 1. (d) The ratings are ordinal: the labels run in order, so one dot per vote stacked above each rating shows the order and the shape at a glance.

Q11. (a) Example: "How many kilograms of recycling does our school collect each week?" — the vague word "green" becomes a measured amount. (b) Example: "How many hours of sleep did each student in our class get last night?" — "kids" becomes our class, and "enough" becomes a counted number of hours that can be compared with a recommendation later.

Q12. (a) **Observation** — the temperature is read from a thermometer; nobody is asked. (b) **Survey** — a favourite is in people's heads; it must be asked. (c) **Observation** — lunch time can be timed with a stopwatch. (d) **Survey** — feelings about the uniform must be asked.

Q13. (a) **Discrete** — books are counted. (b) **Continuous** — mass is measured, and any value in a range is possible. (c) **Ordinal** — the ratings are labels, and 1 to 5 puts them in order. (d) **Nominal** — types of pet are labels with no order.

Q14. (a) $129.050-128.400=0.650$ kL $=650$ L. (b) $650\div5=130$ L per person. (c) 130 L sits **above** the 50–100 L band.

Q15. (a) First, the group is wrong: four friends at one table are not the class, so the data is about the wrong population. Second, the question pushes the answer — "Our canteen is great, right?" is a nudge, not a question. (b) Example: ask **every** student in the class the same neutral question, such as "How do you rate the canteen food, from 1 to 5?"

Q16. Four-point, with this reason: the council wants to know which way the group leans, and with no middle box every vote must land for or against, so the lean can be counted (the Worked example 6 pattern). Five-point, with this reason: the council should know how many students are still undecided, and only a middle box shows that. Either choice passes when the reason matches the scale.

Q17. (a) Our class: mode **2 cars** (11 students). Year 5: mode **1 car** (10 students). (b) Both: $3-0=3$ cars. (c) The same: the range (3 cars) and the lopsided shape. Different: the mode — our class peaks one car higher.

Q18. (a) The **falling red line**: the unwatered pot loses moisture day by day, from 61% down to 26%. (b) $61-26=35$ — range **35** percentage points. (c) The probe measures the same thing the same way every morning, so the two pots are compared with real numbers — guessing by eye could not give a range to compare.

Q19. The order is: **the question → the method → the display and the measures → the finding**, with **a new question** last if the cycle turns. The new question is the part that may be missing — a finished investigation does not have to spawn another.

Q20. Check the total: $11+8+3+2=24$ students. Example finding: *In our class of 24, the most common way to school is walking (11 students), while only 2 students ride.* The finding names the group and answers the question in context.

Q21. (a) **Dot plot with the labels in order** — ratings 1-5 are ordinal, so the order along the axis shows the shape. (b) **Side-by-side column graph** — the seasons are nominal labels, and two groups on one scale can be compared label by label. (c) **Dot plot (or column graph)** — goals per game are discrete counts, one dot per game.

Q22. (a) 3 appears five times — mode **3**. (b) $10 - 3 = 7$ — range **7**. (c) The **range** sees the 10: it stretches from 3 all the way to 10. The mode cannot — it only says which value piles up, not how far out an odd value sits.

Q23. (a) $2\,730 \div 3 = 910$ L per person per week; $910 \div 7 = 130$ L per person per day. (b) 130 L is **above** the 50–100 L basic-needs band. (c) Example: fix the dripping tap and take shorter showers — then the next week's per-person figure should sit lower, closer to the band.

Q24. (a) 6 appears twice, more than any other time — mode **6 minutes**; $12 - 4 = 8$ — range **8 minutes**. (b) The length of a queue is something you measure with a stopwatch; asking people would only guess it. (c) Example: *The 12:30 queue most often takes about 6 minutes, but it stretched to 12 minutes on the busiest day — the range of 8 minutes says the queue is unpredictable.*

Q25. (a) 3, with 80 votes — mode **3**. (b) On a five-point scale, 3 is the **neutral** box. A pile of neutral votes is not happiness — the mode may be a shrug, exactly the trap in Worked example 6. (c) Example: ask the same customers to agree or disagree with "The service is good" on a **four-point scale**, so the 80 neutrals have to break one way or the other.

Q26. (a) **Observation** — growth is measured: the height of each plant in cm, once a week. (b) **Survey** — the favourite day must be asked of every student in the group. (c) **Observation** — the queue is timed with a stopwatch before and after 8:40 am.

Q27. Candidate B is better. Candidate A fails the second checklist test: "too much" is an opinion, so the question never says what will be counted or measured. Candidate B names the group (my household) and the measure (litres per person per day).

Q28. Method: **survey** — ask every student in the class the same question. Variable type: **discrete numerical** — hours are counted. Display: **dot plot**, one dot per student above each hour value. Measures: **mode and range** (and the shape in one word). Skeleton finding: *In our class, the most common amount of sleep last night was __ hours, and the answers ran from __ to __ hours — a range of __ hours.*

Q29. (a) The **second student**. Two days is too short a window for the unwatered pot to dry out — the plan, not the idea, needs refining. That is the cycle turning, not a failure. (b) Any two of: run the investigation for a week or more; add more pots so each treatment has repeats; take readings twice a day; water Pot A less often instead of every day, to compare two watering levels.

What comes next

Chapter 7 ends here — but the toolkit you now carry (pose and keep refining, collect, sort, analyse, interpret, communicate) is the one the Chapter 7 task assesses. Next, Chapter 8 turns to probability: **8A** puts chance on a number line from 0 to 1 with fractions, decimals and percentages (VC2M6P01), and **8B** runs simulations to watch what more trials do to the results (VC2M6P02) — and a simulation is itself a tiny investigation: plan it, run it, display it, report it.

In Year 7 the investigation cycle returns with more tools: a fuller investigation standard (VC2M7ST03), two new numbers that describe the middle of a data set (VC2M7ST01), and a new display for numerical data (VC2M7ST02). Everything you learned to do in 7C — including the habit of refining the question when the data surprises you — carries straight into that work, and into every science investigation that begins with a fair question.

Test

Chapter test T7 — Statistics

Mathematics Level 6 · Chapter 7 · 7A Comparing data sets: mode, range and shape · 7B Statistics in the media · 7C Planning and conducting a statistical investigation

Assesses:

CD	Subtopic
VC2M6ST01	7A Comparing data sets: mode, range and shape
VC2M6ST02	7B Statistics in the media
VC2M6ST03	7C Planning and conducting a statistical investigation

This is a classroom check, not an exam. Answer all questions, and show your working where a question asks for it. Only Question 3 needs the digital tool; the rest can be done with pencil and paper.

Marks	15 (Section A: 5 · Section B: 5 · Section C: 5)
Guidance time	20 minutes
Equipment	Spreadsheet or graphing tool access; data sets supplied

> This is a school-designed paper. It is not a VCAA instrument. Achievement in the Victorian Curriculum is reported against the Level 6 achievement standard, not against a mark on this paper.

Section A — Comparing data sets: mode, range and shape (5 marks)

Q1. A class survey asks every student two questions. Name the type of variable each question collects — nominal, ordinal, discrete or continuous.

a. "Rate the new school garden from 1 to 5." (1 mark)
b. "How many times did you ride a bike last week?" (1 mark)

Q2. The same class asks its 30 students, "What is your favourite fruit?" The answers are labels with no order. Which display suits this data best?

A. A dot plot B. A column graph C. A side-by-side column graph D. A line chart

(1 mark)

Q3. Two classes each asked 10 students, "How many minutes did you read last night?" The results are supplied below.

Minutes	0	10	20	30	40
Class 6E	1	3	4	2	0
Class 6R	2	5	1	1	2

a. Use a spreadsheet or graphing tool to draw a side-by-side column graph of the two data sets on the same scale. (1 mark)
b. From your graph, write down the mode for each class. (1 mark)

Section B — Statistics in the media (5 marks)

Questions 4 to 6 use made-up school media items. No real school or organisation is behind them.

Q4. The school newsletter prints a pie chart with the headline "What our students buy at the canteen". The slices are: pies 35%, sausage rolls 30%, sandwiches 25%, fruit 15%.

a. Add the slices. What is wrong with this pie chart? (1 mark)
b. What does the total tell you about the numbers behind the chart? (1 mark)

Q5. A sports-house poster shows a bar graph of house points for the week: Blue House 84 points, Red House 88 points. The vertical axis is labelled 0, then a zigzag break, then 80, 82, 84, 86, 88 — so Red House's bar looks twice as tall as Blue House's.

a. How many more points did Red House score than Blue House? Show your working. (1 mark)
b. Name the fault in the poster's graph, and say what it does to the story the poster sends. (1 mark)

Q6. A student posts on the school's social media page: "Students want a longer lunch break!" The claim is based on a survey of 12 students, all of them friends of the student who posted. The population the claim talks about is all the students at the school. Give one reason the conclusion does not follow from the survey. (1 mark)

Section C — Planning and conducting a statistical investigation (5 marks)

Q7. Two students refine the question *"Do students care about the environment?"*

Candidate A: *"Do students care too much about the environment?"* Candidate B: *"How many students in our class walk or ride to school each week?"*

Which candidate is the better refined question, and which test in the ready-question checklist does the other candidate fail? (1 mark)

Q8. A class investigates the question: *"How long does it take each student in our class to walk from the school gate to the classroom?"*

a. Is this data collected by observation or by survey? Give the reason. (1 mark)
b. Is the walking time nominal, ordinal, discrete or continuous? (1 mark)

Q9. The student council asks 20 students, "How many times did you borrow from the school library this term?" The results are supplied below.

Number of borrowings	0	1	2	3	4	5
Students	2	3	6	5	3	1

a. Find the mode and the range of the data. (1 mark)
b. Write a one-sentence finding for the student council, in the context of the survey. Include the shape of the data in your sentence. (1 mark)

End of test — total 15 marks.

Solutions

Answer key

Q	Answer	Marks
1a	Ordinal	1
1b	Discrete	1
2	B — a column graph	1
3a	A side-by-side column graph, both classes on the same scale	1
3b	Class 6E: 20 minutes; Class 6R: 10 minutes	1
4a	35% + 30% + 25% + 15% = 105% — the slices are more than the whole	1
4b	Something was counted twice, or a number is wrong	1
5a	4 points	1
5b	Broken axis — the small gap is stretched until it looks big	1
6	Any valid reason, for example only 12 friends were asked	1
7	Candidate B; Candidate A fails the "counted or measured" test	1
8a	Observation — the time is measured, nobody is asked	1
8b	Continuous	1
9a	Mode 2; range 5	1
9b	Any in-context finding that includes the shape (lopsided)	1
	Total	**15**

Worked solutions

Q1. Ask what the answers look like. a. The ratings 1 to 5 are labels, but 1 to 5 puts them in order, so the variable is ordinal (categorical). b. The number of bike rides is counted in whole numbers — 3 rides, never 3.5 — so the variable is discrete (numerical). ∴ a ordinal; b discrete.

Q2. Favourite fruit is nominal data — labels with no order. A column graph gives one column for each label, and the tallest column shows the mode, so it is the display that suits nominal data. A dot plot needs values that run in order, a side-by-side column graph compares two groups at once, and a line chart shows change over time. ∴ **B, a column graph.**

Q3. a. Enter the minutes in one column, with the two classes' counts in the next two columns and a heading row on top. Draw a column graph from all three columns. The correct display has the minutes 0, 10, 20, 30, 40 along the bottom axis, with two columns at each value — one for Class 6E and one for Class 6R — and both classes drawn against the same vertical scale, so the two sets can be read against each other. That comparative display earns the mark. b. Read the tallest column for each class: Class 6E's tallest column is at 20 minutes (4 students), so the mode is **20 minutes**; Class 6R's tallest column is at 10 minutes (5 students), so the mode is **10 minutes**. (The modes can also be read straight from the supplied table — the graph mark is for the comparative display itself.)

Q4. a. $35\% + 30\% + 25\% + 15\% = 105\%$. The slices are more than the whole: a pie chart that covers a whole population must add to exactly 100%, so something is wrong with this chart. b. A total over 100% means something was counted twice — or one of the percentages is simply wrong. The numbers behind the chart cannot all be correct as printed. ∴ a 105%, more than the whole; b counted twice or wrong.

Q5. a. $88 - 84 = 4$. Red House scored **4 points** more than Blue House. b. The fault is a **broken axis**: the axis jumps from 0 to 80 and leaves out every number in between. The real gap of 4 points is drawn on the stretched piece of axis from 80 to 88, where every small change looks tall — so Red House's bar looks twice as tall as Blue House's when the lead is only 4 points. A fair graph would run the whole axis from 0 with equal jumps. ∴ a 4 points; b broken axis — a small gap is stretched until it looks big.

Q6. The survey only asked 12 students, and every one of them is a friend of the student who posted. Friends are likely to agree with the poster, and 12 students is not the whole school — so the data is about the wrong population, and it cannot support a claim about all students. Any answer that names the gap between who was asked (12 friends) and who the claim is about (all students at the school) earns the mark. ∴ example as given.

Q7. Candidate B names the group (our class) and says what will be counted (the students who walk or ride to school), so it is ready for data. Candidate A keeps the vague group "students", and "care too much" is an opinion, not a count or a measure — it fails the checklist test *says what will be counted or measured*. ∴ Candidate B; Candidate A fails the "counted or measured" test.

Q8. a. **Observation** — the walking time is measured with a stopwatch, the same way for every student, and nobody is asked a question. b. Time is measured, and any value in a range is possible between one time and the next, so the variable is **continuous** (numerical). ∴ a observation, because the time is measured; b continuous.

Q9. a. The tallest count is 6 students, at 2 borrowings, so the mode is **2**. The highest value is 5 and the lowest is 0, so the range is $5 - 0 = 5$ — range **5**. b. Example: *Of the 20 students we surveyed, the most common number of library borrowings this term was 2, and the shape is lopsided — most students borrowed a few times, with a tail of students who borrowed 4 or 5 times.* The mark needs a sentence that is about the group surveyed, uses the data, and includes the shape: the peak sits at 2, near the low end, with the stacks falling away towards 5, so the distribution is lopsided. ∴ a mode 2, range 5; b example as given.