

VCE VM Numeracy Units 3 & 4

Chapter 8 — Data

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

Attribution: Classbasics Pty Ltd, vwx.au

Distribution note: this document may be distributed, and whole pages may be removed from it, provided only WHOLE pages are removed (never part of a page). This allows teachers to remove tests, exams and subtopics they do not want to issue.

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

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

Contents

Section	Page
8A Collecting and categorising data	2
8B Tables and graphs: axes and scales	16
8C Measures of centre: mean, median and mode	29
8D Spread, range and outliers	41
8E Interpreting data and drawing conclusions	53
Chapter 08 — Test A (multiple choice)	69
Test A — Answers	72
Test A — Fully-worked solutions	73
Chapter 08 — Test B (written response)	75
Test B — Answers	80
Test B — Fully-worked solutions	81

Chapter 8 Data — 8A Collecting and categorising data

Theory

Data is everywhere: on receipts, in footy stats, on workplace rosters, in health apps, in the news. Before any of it can be graphed or analysed, someone has to **collect** it, **sort** it, and **organise** it. That is this section. Section 8B turns organised data into tables and graphs, and sections 8C–8E measure and interpret it.

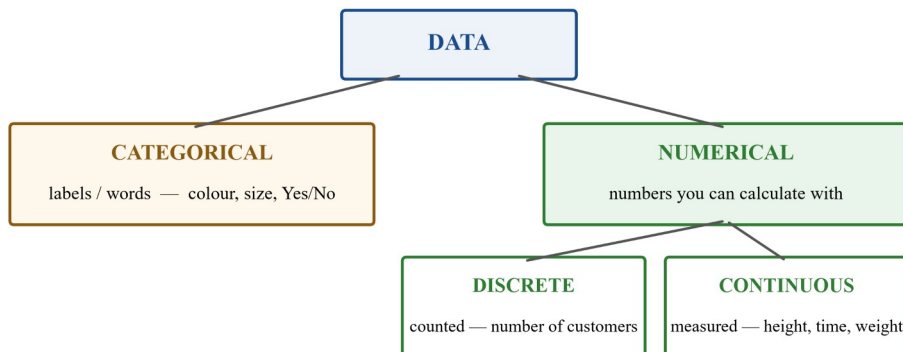
What is data?

Data is facts and numbers collected to answer a question. A single number on its own means little; a set of numbers, sorted and organised, starts to tell you something. A **data set** is a collection of data gathered for one purpose — every order a café takes in a day, every height measured in a class, every goal kicked in a season.

Two kinds of data.

Every piece of data is either **categorical** or **numerical**. Sorting data into these kinds is the first step of **categorisation** — putting data into groups (categories) that share something in common.

- **Categorical data** sorts things into groups using words or labels: eye colour, type of transport, “Yes” or “No”, sizes like small, medium and large. The labels may have an order (large is bigger than small), but they are still labels, not numbers you can calculate with.
- **Numerical data** is numbers you can do maths with. It comes in two kinds:
 - **Discrete data** is **counted** in whole units: the number of customers in a shift, the number of texts you send, the number of goals kicked. You cannot have 2.5 customers.
 - **Continuous data** is **measured**, and it can take any value in a range: height, weight, time, temperature, the volume of milk in a jug. A time can be 32 minutes, 32.5 minutes, or 32.51 minutes.



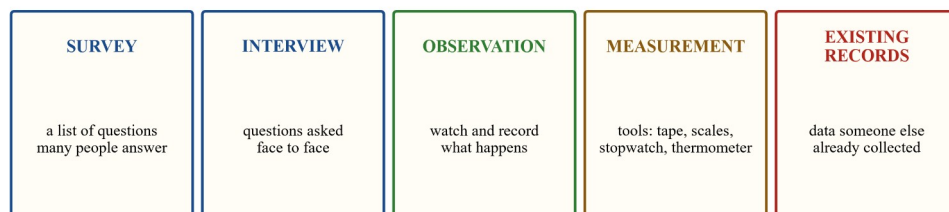
Data item	Kind	Why
Favourite pizza topping	categorical	a label (pepperoni, cheese, ...)
Number of pizzas sold on Friday	numerical, discrete	counted in whole pizzas
Weight of a pizza, in grams	numerical, continuous	measured — any value is possible
Delivery: pickup or delivered	categorical	two labels only

Where data comes from: primary and secondary.

- **Primary data** is data **you collect yourself**, fresh, for your own question — by survey, by observation, or by measuring.
- **Secondary data** is data **someone else already collected** — a table on a government statistics website, a company’s sales records, numbers quoted in a news report.

	Good because	Watch out because
Primary	it is exactly what you need, and you know how it was collected	it takes time and effort to collect
Secondary	it is quick and cheap, and can cover huge groups (a whole country)	it may be out of date, or collected for a different purpose

Tools for collecting data.



five common ways to collect data

- A **survey** is a list of questions that many people answer — on paper, online, or through a phone app.
- An **interview** is a survey you ask one person face to face.
- **Observation** means watching and recording what happens: standing at a corner and counting the cars that pass.
- **Measurement** uses tools — a tape measure, scales, a stopwatch, a thermometer (§1E) — to record numbers directly.
- **Existing records** are secondary data already sitting somewhere: rosters, receipts, a shop's sales system.

Technology helps. A spreadsheet or a phone app can store responses and sort them in seconds (§1F). But nothing in this section needs one — pen, paper, tally marks and a basic calculator do the whole job.

Asking good survey questions.

A survey is only as good as its questions.

- Ask about **one thing at a time**. “Do you exercise and eat well?” is really two questions.
- Use **neutral** wording. A **leading question** pushes people toward the answer the asker wants: “You agree our canteen is the best, don't you?” tells people what to say.
- **Closed questions** offer a list of options to tick (bus, train, car, walk, bike). They are quick to answer and easy to tally. **Open questions** let people write anything. They can tell you more, but the answers are harder to organise.
- Options must **cover every possible answer** (add “Other” if needed) and must **not overlap**, so nobody wonders which box to tick.

Who do you ask? The sample.

You almost never ask everybody. The people you actually ask or measure are called the **sample**. A good sample is **big enough** to trust and **fair** — it includes all the sorts of people you want to say something about. A sample is **biased** when it unfairly over-represents one group: asking only footy players what sport the school should play next term leaves out everyone else, and the data leans one way.

Tallying and frequency.

Raw answers arrive as a messy list. **Collating** means bringing the list together into counts. The classic tool is the **tally mark**: one stroke per answer, and every fifth stroke drawn diagonally across the bundle of four, so you can count bundles at a glance.

Favourite drink — 20 people asked		
Drink	Tally	Frequency
Latte	 	8
Tea	 	5
Juice		4
Water		3
Total		20

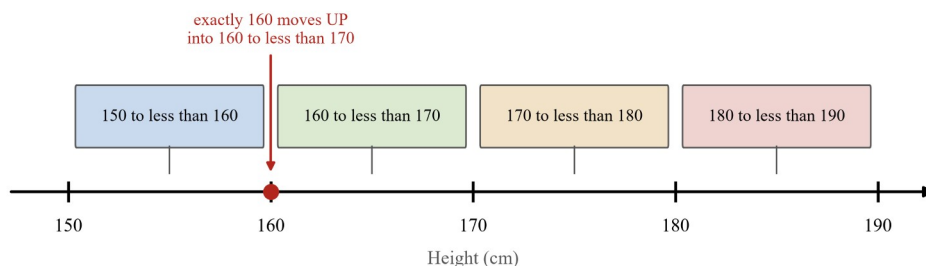
every 5th stroke slants across the bundle

The **frequency** of a category is how many times it occurs. A **frequency table** lists each category with its frequency. Always include a **Total** row: it must equal the number of people (or items) surveyed. That total is a built-in check — if it does not match, a tally got lost.

In the model above, 20 people were asked their favourite drink: latte 8, tea 5, juice 4, water 3. Check: $8 + 5 + 4 + 3 = 20$ ✓. Lattes were chosen by 8 of the 20 people, which is $\frac{8}{20} \times 100 = 40\%$ of them.

Grouping numerical data into intervals.

A list of 24 measured times is hard to read. Grouping it into **intervals** — ranges of equal width — turns the list into a short table. Write each interval so it cannot trap a boundary value: “20 **to less than** 30” means 20 is in, but 30 is not. A value exactly on a boundary always goes into the **higher** interval.



Worked model — 10 heights, in cm, grouped into intervals of 10 cm:

Height (cm)	Interval it lands in
151, 158	150 to less than 160 (frequency 2)
160, 163, 167	160 to less than 170 (frequency 3)
170, 172, 175, 178	170 to less than 180 (frequency 4)
181	180 to less than 190 (frequency 1)

Notice 160 does **not** go into “150 to less than 160” — it sits exactly on the boundary, so it moves up. Check: $2 + 3 + 4 + 1 = 10$ ✓. Grouping makes the list easy to read, but it loses the individual values — you can see that 4 heights fall in “170 to less than 180”, but not which exact heights they were.

Organising: one-way and two-way tables.

A frequency table for one question is a **one-way table**. When each item carries **two** categorical labels — a drink *and* a cup size — a **two-way table** puts one label along the rows and the other along the columns, and counts land in the cells where a row and a column meet. Row totals and column totals are added at the edges, and the grand total can be checked two ways: down the rows, and across the columns. Both must give the same number.

The collection process, step by step.

Collecting data is a process, like the problem-solving cycle (§1A):

1. **Plan.** What question do you want to answer? Write it down.
2. **Categorise.** What kind of data will answer it — categorical, discrete or continuous?
3. **Choose the tool.** Survey, interview, observation, measurement, or existing records?
4. **Collect.** Gather the responses or take the measurements.
5. **Record.** Write everything in the same unit and the same precision.
6. **Collate.** Tally the raw list into counts.
7. **Organise.** Build the frequency table (grouped, if the data is continuous).
8. **Check.** Totals must match; recount anything doubtful; look for impossible values.

Checking data quality.

Data goes wrong in predictable ways. Watch for:

- **keying errors** — a mistyped number, like an age of 214 on a gym list. Ask: could this value happen in real life?
- **missing values** — a blank cell or “n/a”. Decide before you start whether to leave it out or chase it up, and count only the responses you actually have.
- **out-of-date secondary data** — last decade’s prices will not plan this year’s budget.
- **biased samples and leading questions** — the collection itself leans one way.

Precision and rounding.

Record measurements at one stated precision — to the nearest cm, to the nearest minute, to the nearest cent — and keep that precision all the way through (§2C). A value recorded as 14 cm to the nearest cm really means “at least 13.5 cm and under 14.5 cm”. When a question asks you to round an answer, it says how far: “correct to the nearest cent”, “to 1 decimal place”. Carry full precision through the working and round only at that stated point.

About the data in this section. The surveys, shops, clubs, tables and numbers here are **made up** (not real): pretend businesses and pretend numbers, built for this book so you can practise the maths. Nothing here is real data about a real business, person or place.

Worked examples

Example 1 — scaffolded

A café manager records six pieces of information for every order:

1. the type of drink (long black, latte, tea, smoothie)
2. the cup size (small, medium, large)
3. the number of people in the group
4. the time taken to make the order, in seconds
5. the volume of milk used, in mL
6. whether the order is takeaway or dine-in

Classify each item as categorical, or numerical and discrete, or numerical and continuous.

Working	Comment
Item 1, type of drink: the answers are labels (latte, tea, ...), so this is categorical .	Words or labels → categorical.
Item 2, cup size: small, medium, large are labels with an order, still categorical .	Ordered labels are still labels, not numbers.
Item 3, number of people: you count people in whole numbers (2, 3, 4, ...), so this is numerical, discrete .	Counted → discrete.
Item 4, time in seconds: time is measured and can be any value in a range (47.3 s is possible), so numerical ,	Measured → continuous.

continuous.

Item 5, volume of milk in mL: volume is also **measured**, so **numerical, continuous**.

Item 6, takeaway or dine-in: only two labels, so **categorical**.

Same test: could the value fall between two whole numbers?

Yes/No-style answers are categorical.

Example 2 — scaffolded

Thirty students were asked: “How do you get to school most days?” The raw responses are listed below.

School survey				
"How do you get to school most days?"				
30 responses — raw list				
bus	walk	car	bus	train
bus	walk	bike	car	bus
walk	car	train	bus	walk
car	bus	bike	train	walk
bus	car	walk	bus	bus
walk	car	bus	bus	bus

Made-up data: the school, the 30 students and every response are made up for this section.

- (a) Collate the responses into a frequency table. (b) Use the Total row to check your table. (c) Find the percentage of the 30 students who get to school by bus.

Choose the categories first: bus, walk, car, train, bike.

Go through the 30 responses one at a time, putting a tally stroke against each category.

Counts: bus 12, walk 7, car 6, train 3, bike 2.

Frequency table: bus 12, walk 7, car 6, train 3, bike 2;
Total 30.

Check: $12 + 7 + 6 + 3 + 2 = 30$, which equals the number of students asked ✓.

Percentage by bus: $\frac{12}{30} \times 100 = 40\%$.

Categories must cover every answer in the list.

One pass, one stroke per response — cross off each word as you go so nothing is counted twice.

In tally bundles: bus is two bundles of five plus 2.

The table is the collated data, organised.

The Total row is the built-in error check.

Part $\div$ whole $\times 100$ (§2F). No rounding needed — the answer is exact.

Example 3 — unscaffolded

A bakery checks its ovens by weighing 20 loaves, to the nearest gram. The weights are shown below, with an empty grouped table.

Bakery oven check				
20 loaf weights, in grams (nearest g)				
756	768	771	785	792
803	811	818	826	834
842	847	855	861	866
874	888	903	917	926

Group these weights	
Weight, w (g)	Frequency
750 to less than 800	?
800 to less than 850	?
850 to less than 900	?
900 to less than 950	?
Total	20

Made-up data: the bakery, the 20 loaves and every weight are made up for this section.

- (a) Copy and complete the grouped frequency table. (b) How many loaves weigh less than 850 g? (c) A loaf weighs exactly 850 g. Which interval does it belong in?

Step 1. The intervals run from 750 g in steps of 50 g, and each interval includes its bottom value but not its top value: “750 to less than 800”, and so on.

Step 2. Sort each weight into its interval and tally. The completed table:

Weight, w (g)	Frequency
$750 \leq w < 800$	5
$800 \leq w < 850$	7
$850 \leq w < 900$	5
$900 \leq w < 950$	3
Total	20

Step 3. Check: $5 + 7 + 5 + 3 = 20$, which equals the number of loaves weighed ✓. Loaves under 850 g are the ones in the first two intervals: $5 + 7 = 12$.

Step 4. A loaf of exactly 850 g sits on a boundary. The interval “800 to less than 850” excludes 850; the interval “850 to less than 900” includes it. Boundary values always move up.

∴ the frequencies are 5, 7, 5, 3 (total 20 ✓), 12 loaves weigh less than 850 g, and a loaf of exactly 850 g belongs in the interval $850 \leq w < 900$.

Example 4 — unscaffolded

Mia needs data for an assignment. Her first attempt is the card below.

First attempt

"You study hard, don't you?"

How many hours a day do you study?"

Asked: my 8 friends

8 responses

Second attempt

"How many hours did you study last week?"

Tick one: 0–2 3–5 6–8 9 or more

Asked: 50 students from across the year levels

Hours studied	Frequency
0–2	6
3–5	17
6–8	21
9 or more	?
Total	50

Made-up data: Mia, her school and every number here are made up for this section.

- (a) Give two reasons why Mia’s first question collects poor data. (b) Mia rewrites the question neutrally — “How many hours did you study last week?” with options 0–2, 3–5, 6–8, 9 or more — and asks 50 students from

across the year levels. Her frequency table has one number missing. Find it. (c) Find the percentage of the 50 students who studied 6 hours or more.

Step 1 (a). The question has flaws. “You study hard, don’t you?” is **leading** — it tells people the answer the asker expects. And she asked only 8 of her own friends: the sample is too **small** and **biased** (her friends study the way she does, and nobody else is represented).

Step 2 (b). The frequencies must add to the total of 50 students:

Hours studied	Frequency
0–2	6
3–5	17
6–8	21
9 or more	?
Total	50

Missing frequency = $50 - (6 + 17 + 21) = 50 - 44 = 6$.

Step 3. Check: $6 + 17 + 21 + 6 = 50$ ✓. Students who studied 6 hours or more are the last two rows: $21 + 6 = 27$.

Step 4 (c). Percentage = $\frac{27}{50} \times 100 = 54\%$ — exact, so no rounding is needed.

∴ the first question is leading and the sample of 8 friends is small and biased; the missing frequency is 6; and 54 % of the 50 students studied 6 hours or more.

Practice questions

Scaffolded

Q1. Classify each data item as categorical, numerical discrete, or numerical continuous. (a) The colour of each of 15 cars in a car park. (b) The number of customers entering a shop in each hour of a day. (c) The weight of each parcel at a post office, in kg.

Q2. Say whether each is primary or secondary data. (a) A gardener measures the heights of 20 tomato seedlings in her own garden. (b) A student downloads a table of last year’s rainfall from a government weather website. (c) A road-safety group counts the cars that stop at an intersection over 10 minutes.

Q3. Choose the best collection tool for each job — survey, observation, measurement with an instrument, or existing records — and give a reason. (a) Find the most popular lunch option at the canteen. (b) Find how long a swimmer takes to swim 50 m freestyle. (c) Find how many cyclists use a bike path between 8 am and 9 am on a Tuesday.

Q4. The canteen took 25 food orders in one morning: pie, sausage roll, sandwich, salad or fruit. The orders were: pie 9, sausage roll 6, sandwich 5, salad 3, fruit 2. (a) Complete a frequency table of the orders, and add a Total row. (b) Use the Total row to check your table. (c) Find the percentage of orders that were pies, and the percentage that were salad or fruit. Give each percentage to the nearest whole percentage.

Q5. Each survey question below has a fault. Name the fault and rewrite the question. (a) “Don’t you agree that the new gym is fantastic?” (b) “Do you play sport and eat breakfast?” (c) “How many hours of TV do you watch per week? Tick one: 0–5 · 5–10 · 10–20” (Hint for (c): think about someone who watches exactly 5 hours, and someone who watches 25.)

Q6. A junior football club groups its players by age using the intervals “10 to less than 12”, “12 to less than 14”, “14 to less than 16” and “16 to less than 18”. (a) Which interval does a player who is exactly 12 belong in? (b) Which interval does a player aged 15 years 6 months belong in? (c) The ages of 10 players are:

10, 11, 12, 13, 13, 14, 15, 16, 17, 17. Find the frequency of each interval, and check that your frequencies add to 10.

Unscaffolded

Q7. A shop's sales spreadsheet has one column for each item below.

Sales spreadsheet — one row per item sold (extract)

Item name	Size	Number sold	Weight (kg)	Customer rating (1–5)	Delivery time (days)
T-shirt	M	6	0.30	4	2.5
Hoodie	L	3	0.75	5	3.0
Cap	S	11	0.12	3	1.5
Sock pack	M	8	0.20	4	2.0

the shop recorded every sale itself, from its own register

Made-up data: the shop and every entry in the sheet are made up for this section.

- Which columns hold categorical data?
- Of the numerical columns, which are discrete and which are continuous? Give a reason for each.
- The shop recorded these sales itself, from its own register. Is this primary or secondary data?

Q8. A gym asked each of its 45 members to name the equipment they use most. The results are partly tallied below.

Equipment	Frequency
Treadmill	14
Bike	11
Weights	9
Rower	?
Other	4
Total	45

Made-up data: the gym, its members and every number in the table are made up for this section.

- Find the missing frequency for the rower.
- Find the percentage of members who use the treadmill most, correct to 1 decimal place.
- Find the percentage who use the bike or “other” most, correct to 1 decimal place.

Q9. A council wants opinions on a proposed skate park. Its first survey question is: “Do you agree that a skate park is a waste of council money?” Only 8 people answer, and 6 of them are members of the local skate club. (a) Name the fault in the way the question is worded. (b) Find the percentage of the people who answered who were skate club members. Give your answer to the nearest whole percentage. (c) Rewrite the question neutrally, and suggest one way to get a fairer sample.

Q10. A community garden records how many volunteers turn up for each of 20 working shifts:

2, 3, 3, 4, 4, 4, 5, 5, 5, 5, 6, 6, 6, 6, 7, 7, 7, 8, 8, 9

Made-up data: the garden and every number in the list are made up for this section.

- Organise the list into a frequency table with rows for 2, 3, 4, 5, 6, 7, 8 and 9 volunteers.
- How many shifts had 6 or more volunteers?
- Check that your frequencies add to 20, then find the percentage of shifts with 5 or fewer volunteers, to the nearest whole percentage.

Q11. A delivery company times 24 of its deliveries, to the nearest minute. The times, in minutes, are:

22, 24, 26, 27, 29, 30, 31, 33, 35, 36, 38, 39, 40, 42, 44, 45, 46, 48, 51, 53, 55, 58, 61, 64

Made-up data: the company and every time in the list are made up for this section.

- (a) Complete this grouped frequency table using intervals of 10 minutes: “20 to less than 30”, “30 to less than 40”, “40 to less than 50”, “50 to less than 60”, “60 to less than 70”.
- (b) How many deliveries took less than 40 minutes?
- (c) Find the percentage of deliveries that took 30 minutes or more, correct to 1 decimal place.
- (d) A delivery took exactly 50 minutes. Which interval does it belong in?

Q12. A council report written in 2026 uses bicycle-count data that was collected in 2019. (a) How many years old is the data? (b) Give two risks of using secondary data like this. (c) Give one situation where secondary data is still the best choice.

Q13. A school has 1200 students and wants to know which canteen meals to add next term. Plan A: ask the first 10 students through the front gate on Monday morning. Plan B: ask 80 students chosen from every year level. (a) Find the percentage of the school Plan B asks, correct to 1 decimal place. (b) Which plan is more likely to give fair data? Give a reason. (c) Name the problem with Plan A’s sample.

Q14. A clothes shop counts its t-shirt stock. Every shirt costs \$24.95.

Riverside Threads — t-shirt stocktake	
Price: \$24.95 per shirt	
Size	Count
S	18
M	?
L	21
XL	14
Total	80

Made-up data: the shop, the stock counts and the price are made up for this section.

- (a) Find the missing count for size M.
- (b) Which size is the most common in stock?
- (c) Find the value of the size-M stock, correct to the nearest cent.

Q15. A nursery worker measures 16 seedlings with a ruler and records each height to the nearest cm. (a) One seedling is recorded as 14 cm. Between which two values must its true height lie? (b) The recorded heights, in cm, are: 11, 12, 13, 14, 14, 15, 16, 17, 18, 19, 19, 20, 21, 22, 23, 24. Group them into the intervals “10 to less than 15”, “15 to less than 20” and “20 to less than 25”, and find each frequency. (c) Find the percentage of seedlings under 20 cm tall, correct to 1 decimal place.

Made-up data: the nursery, the seedlings and every measurement are made up for this section.

Q16. A café tallies one day’s drink orders by drink and cup size.

Café drink orders — one day

Drink	Small	Medium	Large	Total
Latte	14	26	10	?
Cappuccino	8	18	4	?
Tea	6	9	5	?
Total	?	?	?	?

Made-up data: the café and every number in the table are made up for this section.

- Copy the table and fill in every “?” — the three row totals, the three column totals and the grand total.
- Check the grand total two ways: down the row totals, and across the column totals.
- Find the percentage of the day’s orders that were large, to the nearest whole percentage, and name the most-ordered size.

Q17. A gym’s member spreadsheet has an Age column with 12 entries:

19, 22, 214, 27, 29, 31, (blank), 34, 38, 41, 47, 55

The office later confirms the 214 was a keying error for 24, and the member with the blank cell is 33.

Made-up data: the gym, its members and every number in the list are made up for this section.

- Even before the office confirms it, why is 214 obviously wrong?
- After both corrections, how many of the 12 members are under 30?
- Find the percentage of the 12 members under 30, correct to 1 decimal place.

Q18. A sports club asks all its members: “Which night suits training best — Tuesday, Wednesday, Thursday or Friday?” The 36 responses are listed on the card below.

Club survey — training night					
"Which night suits training best?"					
36 responses — raw list					
Tue	Wed	Wed	Wed	Tue	Wed
Wed	Wed	Tue	Wed	Thu	Fri
Tue	Wed	Thu	Fri	Tue	Wed
Thu	Fri	Tue	Wed	Thu	Fri
Tue	Wed	Thu	Fri	Tue	Wed
Thu	Fri	Tue	Wed	Thu	Fri

Made-up data: the club, its members and every response are made up for this section.

- Collate the responses into a frequency table, and check that the Total row equals 36.
- Which night wins, and what percentage of members chose it? Give the percentage correct to 1 decimal place.
- The organiser claims: “More than 60 % of members chose Tuesday or Wednesday.” Use your table to decide whether the claim is correct, showing your working to 1 decimal place.

Answers

Q1. (a) categorical. (b) numerical, discrete. (c) numerical, continuous. **Q2.** (a) primary. (b) secondary. (c) primary. **Q3.** (a) survey — opinions live in people’s heads. (b) measurement (stopwatch) — time is measured. (c) observation — count what happens, nobody needs to be asked. **Q4.** (a) pie 9, sausage roll 6, sandwich 5, salad 3, fruit 2; Total 25. (b) $9 + 6 + 5 + 3 + 2 = 25$ ✓. (c) pies 36 %; salad or fruit 20 % (both exact — no rounding needed). **Q5.** (a) leading — “Do you think the new gym is good? Tick: Yes / No / Not sure”. (b) two questions in one — split into “Do you play sport?” and “Do you eat breakfast?”. (c) options overlap at 5 and 10, and nothing covers more than 20 hours — use $0-4 \cdot 5-9 \cdot 10-19 \cdot 20$ or more. **Q6.** (a) “12 to less than 14”. (b) “14 to less than 16”. (c) 2, 3, 2, 3; total 10 ✓. **Q7.** (a) Item name; Size. (b) discrete: Number sold, Customer rating (counted); continuous: Weight, Delivery time (measured). (c) primary. **Q8.** (a) 7. (b) 31.1 % to 1 d.p. (c) 33.3 % to 1 d.p. **Q9.** (a) leading question. (b) 75 % (exact). (c) e.g. “Do you support or oppose a skate park? Support / Oppose / Not sure”; ask a larger group from across the whole council area, not just one interested club. **Q10.** (a) 2:1, 3:2, 4:3, 5:4, 6:4, 7:3, 8:2, 9:1. (b) 10 shifts. (c) total 20 ✓; 50 % (exact). **Q11.** (a) 5, 7, 6, 4, 2; total 24 ✓. (b) 12. (c) 79.2 % to 1 d.p. (d) “50 to less than 60”. **Q12.** (a) 7 years. (b) any two of: out of date; collected for a different purpose; cannot check how it was collected; may not cover the area needed. (c) e.g. national statistics — too big for one person to collect fresh. **Q13.** (a) 6.7 % to 1 d.p. (b) Plan B — bigger, and spread across every year level, so it is fairer. (c) it is biased (one gate, one morning; misses whole groups). **Q14.** (a) 27. (b) size M. (c) \$673.65 (exact to the cent). **Q15.** (a) at least 13.5 cm and under 14.5 cm. (b) 5, 6, 5; total 16 ✓. (c) 68.8 % to 1 d.p. **Q16.** (a) rows 50, 30, 20; columns 28, 53, 19; grand total 100. (b) $50 + 30 + 20 = 100$ and $28 + 53 + 19 = 100$ ✓. (c) 19 % (exact); medium. **Q17.** (a) no gym member is 214 years old — it is outside any sensible range. (b) 5. (c) 41.7 % to 1 d.p. **Q18.** (a) Tue 9, Wed 13, Thu 7, Fri 7; total 36 ✓. (b) Wednesday, 36.1 % to 1 d.p. (c) $\text{Tue} + \text{Wed} = 22$; $\frac{22}{36} \times 100 = 61.1$ % to 1 d.p. — the claim is correct.

Fully-worked solutions

Q1. (a) Colours are labels (white, silver, red, ...), so **categorical**. (b) Customers are counted in whole numbers, so **numerical, discrete** — there is no such thing as half a customer walking in. (c) Weight is measured and can take any value in a range (2.35 kg is possible), so **numerical, continuous**. Check: labels → categorical; counted → discrete; measured → continuous.

Q2. (a) The gardener collects the heights herself — **primary**. (b) The rainfall table was collected by someone else, earlier, for their own purpose — **secondary**. (c) The group does the counting itself — **primary**.

Q3. (a) **Survey** — you cannot observe a favourite lunch; you have to ask people. (b) **Measurement with an instrument** — a stopwatch measures the time directly, to a stated precision. (c) **Observation** — stand at the path and tally each cyclist; no question needs asking, and the record is of what actually happened.

Q4. (a) Tally the 25 orders:

Order	Frequency
Pie	9
Sausage roll	6
Sandwich	5
Salad	3
Fruit	2
Total	25

(b) Check: $9 + 6 + 5 + 3 + 2 = 25$ ✓ — matches the number of orders taken.

(c) Pies: $\frac{9}{25} \times 100 = 36$ % (exact). Salad or fruit: $3 + 2 = 5$ orders, and $\frac{5}{25} \times 100 = 20$ % (exact). To the nearest whole percentage: 36 % and 20 %.

Q5. (a) Fault: **leading** — the wording tells people the answer expected. Rewrite: “What do you think of the new gym? Very good / Good / Not good / Not sure”. (b) Fault: **two questions in one** — a person who plays sport but skips breakfast cannot answer. Rewrite: ask “Do you play sport? Yes / No” and, separately, “Do you eat breakfast? Yes / No”. (c) Fault: the options **overlap** (exactly 5 hours fits two options, and so does exactly 10) and they do not **cover** every answer (someone watching 25 hours has nowhere to tick). Rewrite: “0–4 · 5–9 · 10–19 · 20 or more” (whole hours per week).

Q6. (a) Exactly 12 sits on the boundary between “10 to less than 12” and “12 to less than 14”. The first interval excludes 12, so the player belongs in “12 to less than 14” — boundary values move up. (b) 15 years 6 months = 15.5 years, and $14 \leq 15.5 < 16$, so the interval is “14 to less than 16”. (c) Sort and tally the 10 ages:

Age interval	Ages	Frequency
10 to less than 12	10, 11	2
12 to less than 14	12, 13, 13	3
14 to less than 16	14, 15	2
16 to less than 18	16, 17, 17	3
Total		10

Check: $2 + 3 + 2 + 3 = 10$ ✓.

Q7. (a) **Item name** and **Size** hold labels (a product’s name; S, M, L), so they are categorical. (b) **Number sold** is counted in whole units → discrete. **Customer rating (1–5)** is a counted score from a fixed list of whole numbers → discrete. **Weight (kg)** and **Delivery time (days)** are measured and can take any value in a range → continuous. (c) The shop gathered the sales itself, fresh, from its own register → **primary data**.

Q8. (a) The frequencies must add to 45: $\text{Rower} = 45 - (14 + 11 + 9 + 4) = 45 - 38 = 7$. Check: $14 + 11 + 9 + 7 + 4 = 45$

✓. (b) Treadmill: $\frac{14}{45} \times 100 = 31.111\dots$, which is 31.1 % correct to 1 decimal place. (c) Bike or other: $11 + 4 = 15$

members, and $\frac{15}{45} \times 100 = 33.333\dots$, which is 33.3 % correct to 1 decimal place. Check: $31.1\% + 33.3\%$ is well under 100 %, as expected, because the weights and rower members are not counted in either ✓.

Q9. (a) The question is **leading** — “a waste of council money” pushes people toward one answer before they have thought. (b) $\frac{6}{8} \times 100 = 75\%$ — exactly 75 % of the people who answered were skate club members (to the nearest whole percentage: 75 %). (c) Neutral rewrite: “Do you support or oppose building a skate park in the council area? Support / Oppose / Not sure”. Fairer sample: ask many more people from across the whole council area — for example, a survey at several locations and times — not just the one club with the strongest interest.

Q10. (a) Tally each value in the list of 20:

Volunteers	Frequency
2	1
3	2
4	3
5	4
6	4
7	3
8	2
9	1
Total	20

(b) Shifts with 6 or more: $4 + 3 + 2 + 1 = 10$ shifts.

(c) Check: $1 + 2 + 3 + 4 + 4 + 3 + 2 + 1 = 20$ ✓. Shifts with 5 or fewer: $1 + 2 + 3 + 4 = 10$, so the percentage is $\frac{10}{20} \times 100 = 50\%$ — exactly 50% to the nearest whole percentage. Check: (b) and (c) split the 20 shifts into $10 + 10$ ✓.

Q11. (a) Sort the 24 times into intervals of 10 minutes (bottom value included, top value excluded):

Time, t (min)	Times in the list	Frequency
$20 \leq t < 30$	22, 24, 26, 27, 29	5
$30 \leq t < 40$	30, 31, 33, 35, 36, 38, 39	7
$40 \leq t < 50$	40, 42, 44, 45, 46, 48	6
$50 \leq t < 60$	51, 53, 55, 58	4
$60 \leq t < 70$	61, 64	2
Total		24

Check: $5 + 7 + 6 + 4 + 2 = 24$ ✓. (b) Less than 40 minutes: the first two intervals, $5 + 7 = 12$ deliveries. (c) 30 minutes or more: $7 + 6 + 4 + 2 = 19$ deliveries. $\frac{19}{24} \times 100 = 79.1666\dots$, which is 79.2% correct to 1 decimal place. (d) Exactly 50 minutes sits on a boundary, so it moves up: the interval “50 to less than 60” (the interval “40 to less than 50” excludes 50).

Q12. (a) $2026 - 2019 = 7$ — the data is **7 years old**. (b) Any two of: it may be **out of date** — cycling habits change in 7 years; it was collected for **someone else’s purpose**, so it may not measure what the report needs; the council **cannot check how it was collected**, so its quality is taken on trust; it may not cover the area the report is about. (c) When fresh collection is impossible or silly — for example, national population or weather statistics cover millions of records that no council could collect itself. Secondary data is quick, cheap, and sometimes the only option.

Q13. (a) $\frac{80}{1200} \times 100 = 6.666\dots$, which is 6.7% correct to 1 decimal place. (b) **Plan B** — it asks more students, and it reaches every year level, so no group of the school is left out. A fair sample represents the whole group you want to talk about. (c) Plan A’s sample is **biased**: the first 10 through one gate on one morning are whoever happens to arrive that way (perhaps only bus riders or only one year level), and the other 1190 students have no voice.

Q14. (a) The counts must add to the total of 80 shirts: $M = 80 - (18 + 21 + 14) = 80 - 53 = 27$. Check: $18 + 27 + 21 + 14 = 80$ ✓. (b) 27 is the largest count, so **size M** is the most common in stock. (c) Value $= 27 \times \$24.95 = \$673.65$ — exactly **\$673.65**, correct to the nearest cent. Check: estimate $27 \times \$25 = \$675$, just above \$673.65 ✓ (the price is 5 cents under \$25, and $27 \times 5 \text{ c} = \$1.35$).

Q15. (a) Recorded to the nearest cm, 14 cm means the true height is at least 13.5 cm and under 14.5 cm. (b) Sort the 16 heights (15 sits on a boundary, so it moves up into “15 to less than 20”):

Height, h (cm)	Frequency
$10 \leq h < 15$	5
$15 \leq h < 20$	6
$20 \leq h < 25$	5
Total	16

Check: $5 + 6 + 5 = 16$ ✓. (c) Under 20 cm: $5 + 6 = 11$ seedlings. $\frac{11}{16} \times 100 = 68.75$, which is 68.8 % correct to 1 decimal place.

Q16. (a) Row totals: latte $14 + 26 + 10 = 50$; cappuccino $8 + 18 + 4 = 30$; tea $6 + 9 + 5 = 20$. Column totals: small $14 + 8 + 6 = 28$; medium $26 + 18 + 9 = 53$; large $10 + 4 + 5 = 19$. Grand total: $50 + 30 + 20 = 100$. (b) Other direction: $28 + 53 + 19 = 100$ — the same grand total ✓, so the table hangs together. (c) Large: $\frac{19}{100} \times 100 = 19\%$ — exactly 19 % to the nearest whole percentage. The most-ordered size is **medium** (53 orders, the largest column total).

Q17. (a) 214 is impossible as an age — no gym member is 214 years old. A value this far outside any sensible range is a classic keying error, and checking values against what is possible in real life is part of the process (step 8). (b) Replace 214 with 24 and the blank with 33. The corrected ages are: 19, 22, 24, 27, 29, 31, 33, 34, 38, 41, 47, 55. Under 30: 19, 22, 24, 27, 29 — that is 5 **members**. (c) $\frac{5}{12} \times 100 = 41.666\dots$, which is 41.7 % correct to 1 decimal place. Check: before the fix, the “under 30” count would have been 4 — the error had hidden a member in the wrong place, which is why cleaning comes before counting.

Q18. (a) Tally the 36 responses:

Night	Frequency
Tuesday	9
Wednesday	13
Thursday	7
Friday	7
Total	36

Check: $9 + 13 + 7 + 7 = 36$ ✓. (b) Wednesday has the largest frequency (13), so **Wednesday** wins. Percentage: $\frac{13}{36} \times 100 = 36.111\dots$, which is 36.1 % correct to 1 decimal place. (c) Tuesday or Wednesday: $9 + 13 = 22$ members. $\frac{22}{36} \times 100 = 61.111\dots = 61.1\%$ correct to 1 decimal place. $61.1\% > 60\%$, so the organiser’s claim is **correct**.

Theory

In section 8A you sorted data into categories and counted what fell in each one. This section is about showing those counts to somebody else. There are two big families of display: **tables** and **graphs**. A table holds the exact numbers so anyone can look up a value. A graph turns the same numbers into a picture, so the shape of the data — which category is biggest, where things rise and fall — can be seen at a glance. Both are only useful if they are labelled properly, so this section also covers the parts of a graph, and how to choose and read a scale.

About the data in this section. Every business, person, price, plan and workplace in this section is **made up** (not real). They are built for this book so you can practise reading and drawing tables and graphs. Nothing here is real data about a real person, business or place.

Tables. A **table** sets numbers out in rows and columns under headings, so you can look up an exact value quickly. The count for a category is called the **frequency** of that category. Here is one week of bottle counting at a community shop:

Day	Bottles collected
Mon	45
Tue	60
Wed	35
Thu	70
Fri	85
Sat	95
Total	390

Made-up data: the shop and its bottle counts are made up for this section.

When data is measured (like time or mass), the values are usually gathered into **groups** (also called **intervals**) before they are counted. This is a **grouped frequency table**:

Waiting time (minutes)	Number of customers
0–9	6
10–19	14
20–29	11
30–39	6
40–49	3
Total	40

Every wait is tallied in exactly one group. A wait that falls between two group labels is rounded to the nearest minute first — a wait of 9.5 minutes rounds to 10 minutes, so it is tallied in the 10–19 group.

Made-up data: the service counter and its waiting times are made up for this section.

A **two-way table** holds two categories at once. This one counts gym classes by type and by day:

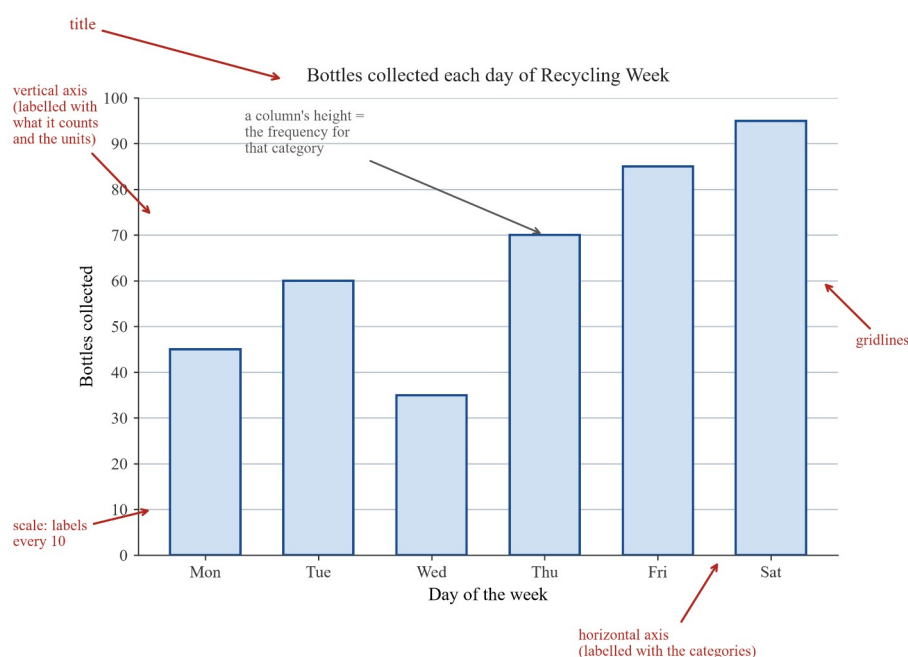
Class	Tue	Thu	Sat
Yoga	12	18	9
Spin	15	14	20
Boxing	8	10	11

Read any cell by finding its row and its column: the Spin row and the Thu column meet at 14, so 14 spin classes ran on Thursday. Totals can be added as an extra column and an extra row, and each one is just the sum of one row or one column — the Spin row totals $15 + 14 + 20 = 49$ classes.

Made-up data: the gym and its class counts are made up for this section.

Every table needs four things to be readable: a **title** saying what is in it, **row and column headings** saying what each line holds, the **units** in the headings where there are units (minutes, kg, dollars), and the **numbers** themselves lined up under their headings.

Parts of a graph. A graph is the same data drawn as a picture. Every graph in this book has the same parts, shown in red on this column chart:



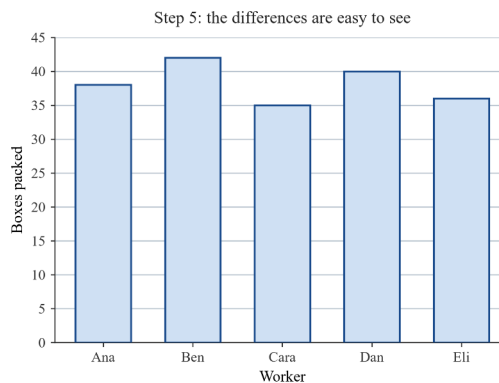
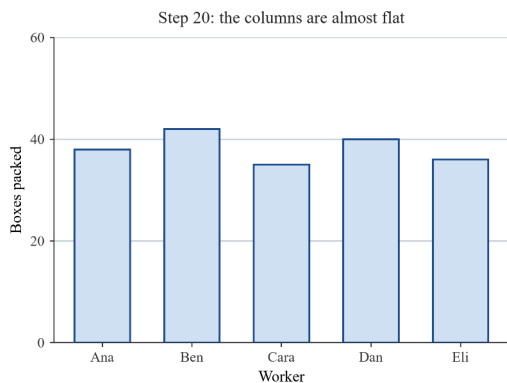
Made-up data: the bottle counts in this graph are made up for this section.

- The **title** says what the graph shows.
- The **horizontal axis** (along the bottom) carries the categories.
- The **vertical axis** (up the side) carries the counts, with the units in its label.
- The **scale** is the set of labels on an axis. On this vertical axis the labels climb by the same amount every square, which is what makes the graph readable.
- The **gridlines** run from the labels across the graph so heights can be read.

Choosing a scale. The scale you choose decides whether a graph is easy to read or not. On squared paper:

1. Find the largest value you must plot.
2. Count the squares you have on the axis you are drawing.
3. Pick a step from the friendly list — 1, 2, 5, 10, 20, 50, 100 — so that the largest value fits in the squares you have and most of your values land on gridlines.
4. Label the top square with the multiple of your step that sits just above (or exactly at) the largest value, then label the squares below it in even steps.

Worked model: the largest value is 47 and the paper has 10 squares on the axis. Step 1 would need 47 squares — far too many. Step 5 reaches a top label of 50 in exactly 10 squares, and 47 sits just below 50, so step 5 fits and step 5 it is. The same data drawn two ways shows why the step matters:



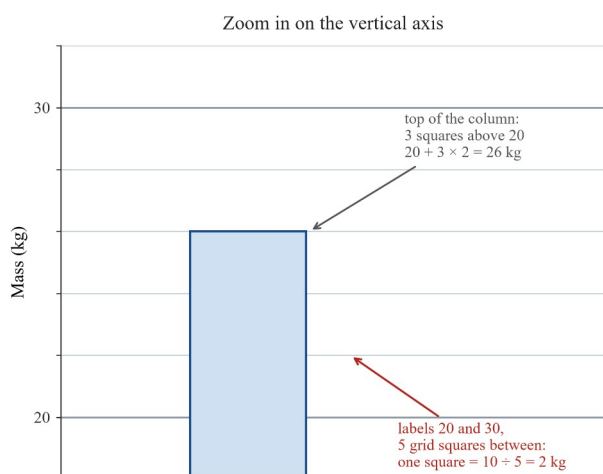
Made-up data: the box counts in these two graphs are made up for this section.

With step 20 the tallest column is only about 2 squares high and the differences between people are almost invisible. With step 5 the top label is 45, reached in 9 squares, and every difference is easy to see. One rule never bends: on a column chart, a bar chart or a histogram, the vertical scale **starts at 0**. Columns are compared by height, and heights only tell the truth when they are measured from zero.

Reading a scale. Sometimes you meet a graph whose grid squares are not labelled one-by-one. Then the worth of one square is a small division:

$$\text{one square} = \frac{\text{gap between two neighbour labels}}{\text{squares between them}}$$

On this axis the labels 20 and 30 sit 5 squares apart:



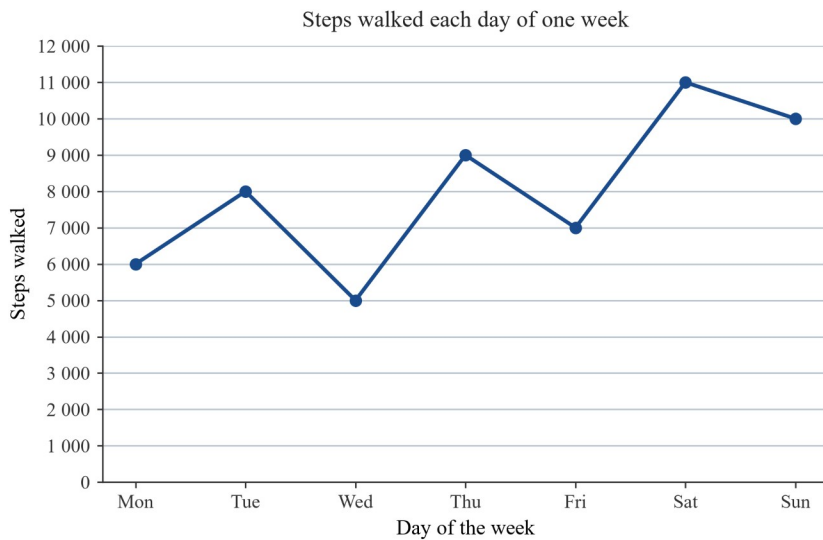
Made-up data: the column in this picture is made up for this section.

So one square is worth $(30 - 20) \div 5 = 2$. The column top sits 3 squares above the 20 label: $20 + 3 \times 2 = 26$.

Commonly used graphs. These are the displays you will meet most often in this book.

Column and bar charts. A **column chart** stands one column on each category, like the bottle graph above; a **bar chart** is the same picture lying down, with the bars running left to right. Use either to compare categories.

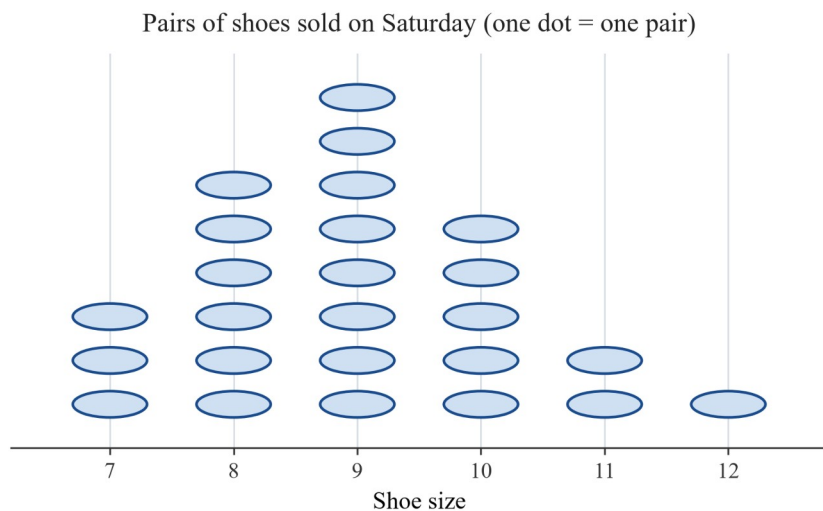
Line graph. A **line graph** plots a dot at each (category, value) pair and joins neighbouring dots with straight lines. Use it when the order of the categories matters — most often time — because the rises and falls of the line show change:



Made-up data: the step counts in this graph are made up for this section.

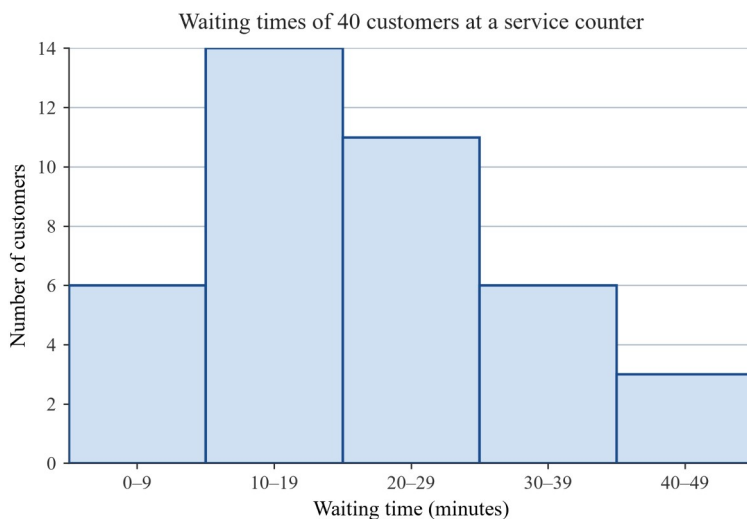
Each dot sits at a day and a step count — Tuesday's dot is at 8 000 steps — and the line between two dots shows how the week was moving.

Dot plot. A **dot plot** stacks one dot per data value above each value on a number line. It keeps every single value on show, so it suits small data sets:



Made-up data: the shoe sales in this graph are made up for this section.

Histogram. A **histogram** looks like a column chart with the columns touching, because the data is measured and grouped: each column is one group of a grouped frequency table, in order, with no gaps between groups:



Made-up data: the waiting times in this graph are made up for this section.

The histogram above is the waiting-time table from earlier, drawn as a picture — same groups, same frequencies, columns touching.

Pie chart. A **pie chart** cuts a circle into **sectors** (slices), one per category, so each sector's angle shows that category's share of the whole. The angle of a sector is

$$\text{sector angle} = \frac{\text{frequency}}{\text{total}} \times 360^\circ$$

and its percentage share is

$$\text{percentage} = \frac{\text{frequency}}{\text{total}} \times 100\%$$

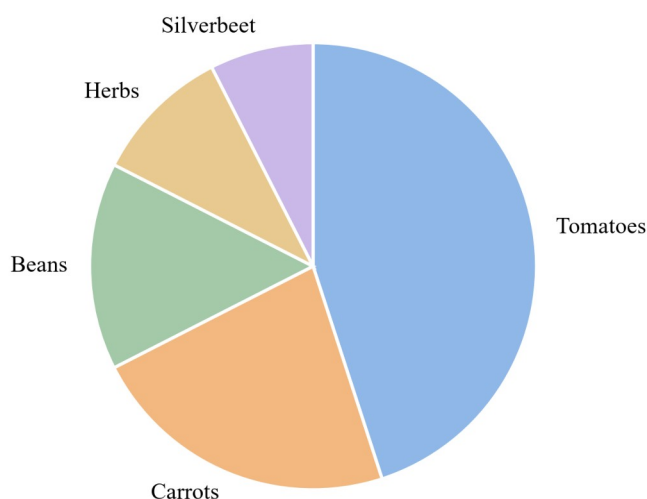
One week's harvest at a community garden weighed 40 kg in total:

Vegetable	Mass (kg)
Tomatoes	18
Carrots	9
Beans	6
Herbs	4
Silverbeet	3
Total	40

Made-up data: the garden harvest and its masses are made up for this section.

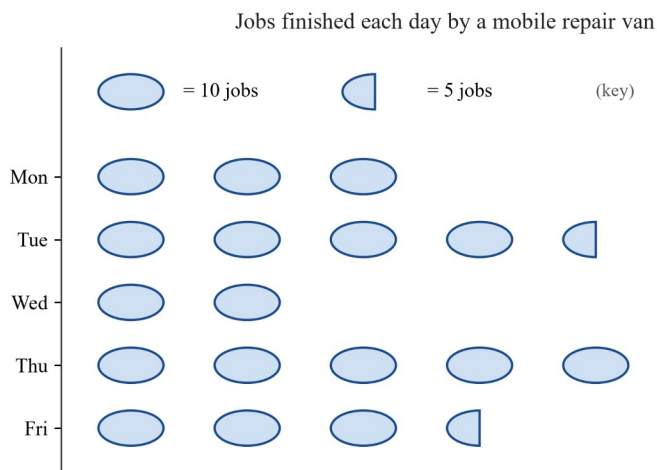
The tomatoes sector angle is $\frac{18}{40} \times 360^\circ = 162^\circ$ — nearly half the circle, which fits tomatoes being nearly half the weight. Drawing every sector at its angle gives:

Vegetables harvested at a community garden (40 kg in total)



Made-up data: the harvest shares in this pie chart are made up for this section.

Pictograph. A **pictograph** uses picture symbols instead of columns, with a **key** saying what one whole symbol is worth; a half symbol is worth half of that:



Made-up data: the repair-van job counts in this pictograph are made up for this section.

The key says one circle is 10 jobs, so a half circle is 5 jobs. Thursday's row shows 5 whole circles: $5 \times 10 = 50$ jobs.

Choosing the right display. The best display is the one that answers your question most clearly:

If you want to ...	use ...
look up exact values	a table
compare categories	a column or bar chart
show change over time	a line graph
keep every value of a small data set on show	a dot plot
show measured data grouped into intervals	a histogram
show shares of a whole	a pie chart
give a quick picture with symbols	a pictograph
hold two categories at once	a two-way table

Tools that help. A spreadsheet or a graphing calculator will draw every display in this section for you, and a spreadsheet will even pick a scale — but nothing here needs one. Every graph in this section can be read with a ruler and drawn by hand on squared paper, and drawing them by hand is how the scale rules above become yours. If you do use technology, check its axes afterwards: the same rules (title, labels, units, scale starting at 0) still apply.

Worked examples

Example 1 — scaffolded

A café counts its hot drinks for one day: flat white 45, latte 40, cappuccino 30, tea 25, hot chocolate 10. The manager wants a column chart on squared paper with 10 squares on the vertical axis. Choose the scale, then find how many squares tall the latte column is.

Made-up data: the café and its drink counts are made up for this section.

Working	Comment
Largest value to plot: 45.	The scale must cover this one.
Try step 5: top label 45, reached in $45 \div 5 = 9$ squares — it fits the 10 squares with one to spare.	Friendly steps are 1, 2, 5, 10, ...
Every count (45, 40, 30, 25, 10) is a multiple of 5, so every column ends exactly on a gridline.	A good scale puts column tops on gridlines.
Step 10 would fail: 45 is not a multiple of 10, so the tallest column would end between gridlines.	Test the step against your own values.

Working

Latte column: $40 \div 5 = 8$ squares tall.

The scale is step 5 from 0 to 45 on the 10-square paper, and the latte column is 8 squares tall.

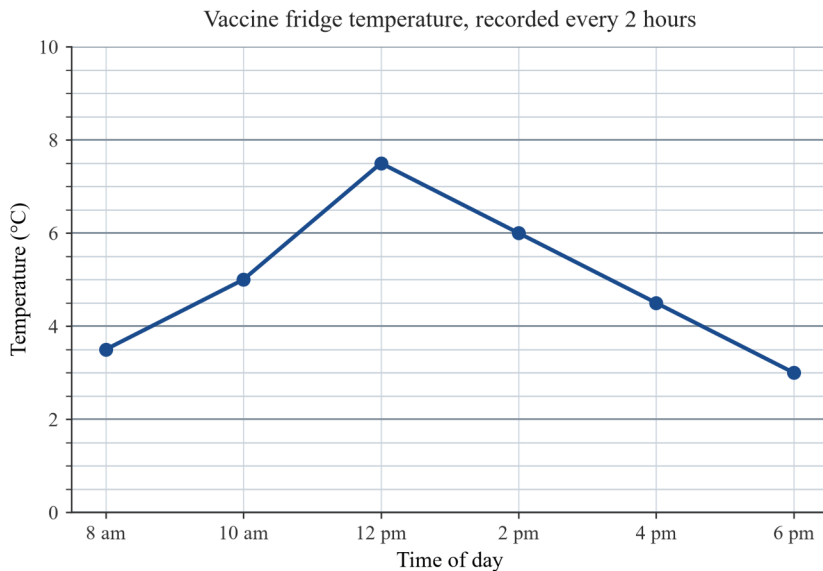
Comment

One square is worth 5 drinks.

Answer both parts, in words.

Example 2 — scaffolded

A clinic's vaccine fridge records its temperature every 2 hours through one day. The line graph below has labels every 2°C on the vertical axis, with 4 grid squares between each pair of labels. Find what one grid square is worth, read the noon temperature, and decide whether the fridge stayed inside its safe band of 2°C to 8°C all day.



Made-up data: the fridge and every temperature on the graph are made up for this section.

Working

One square: $2 \div 4 = 0.5$, so each grid square is worth 0.5°C .

Noon: the dot sits 3 squares above the 6 label:
 $6 + 3 \times 0.5 = 7.5^{\circ}\text{C}$.

Highest reading 7.5°C (noon); lowest reading 3°C (6 pm).

Both extremes sit inside the safe band 2°C to 8°C , so every reading does.

One square is worth 0.5°C , noon read 7.5°C , and the fridge stayed inside the safe band all day.

Comment

Square worth = label gap $\div$ squares between.

Count squares up from the nearest label below.

Scan every dot before deciding.

With the extremes inside, all points between them are too.

Answer all three parts, in words.

Example 3 — unscaffolded

A plumber lists April's jobs by type: repairs 16, installations 10, maintenance 8, quote visits 6. Draw a pie chart of this data — find each sector angle, check the angles, and label the sectors.

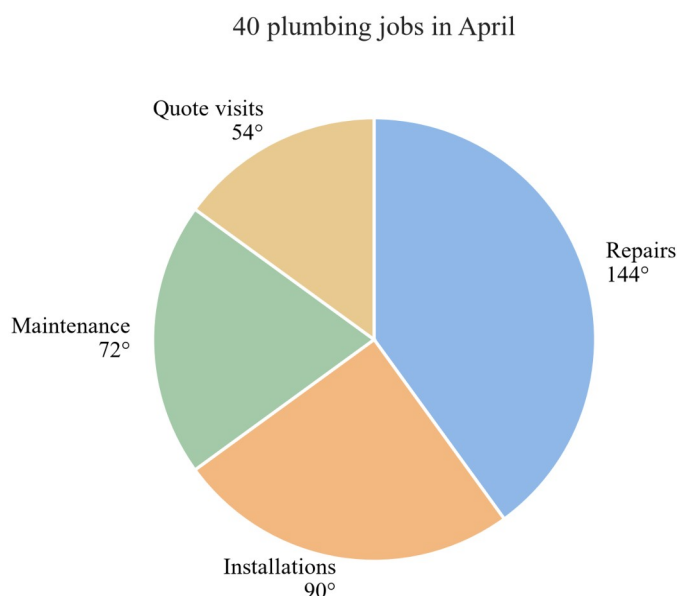
Made-up data: the plumber's jobs and their counts are made up for this section.

The total is $16 + 10 + 8 + 6 = 40$ jobs. Each angle is the frequency over the total, times 360° :

Job type	Frequency	Sector angle
Repairs	16	$\frac{16}{40} \times 360^{\circ} = 144^{\circ}$
Installations	10	$\frac{10}{40} \times 360^{\circ} = 90^{\circ}$

Job type	Frequency	Sector angle
Maintenance	8	$\frac{8}{40} \times 360^\circ = 72^\circ$
Quote visits	6	$\frac{6}{40} \times 360^\circ = 54^\circ$

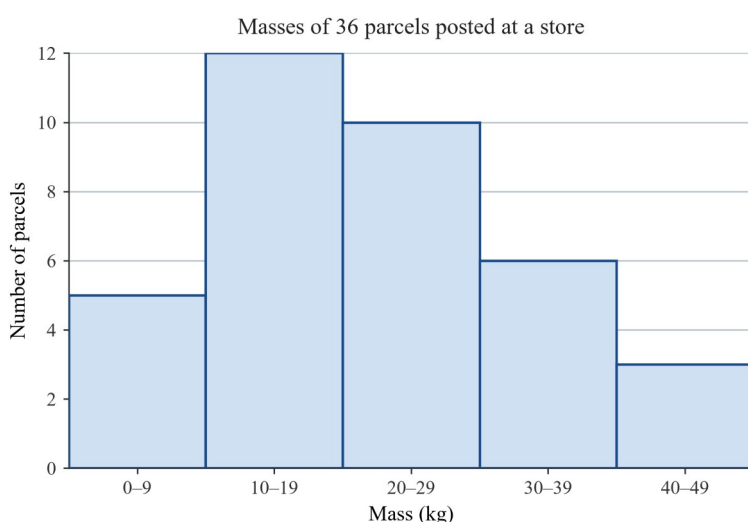
Check: $144^\circ + 90^\circ + 72^\circ + 54^\circ = 360^\circ$ — a full circle, so no angle is out. Drawing the sectors in order from the top gives:



$\therefore$ the sector angles are 144° , 90° , 72° and 54° , and they add to a full circle.

Example 4 — unscaffolded

A store weighs every parcel posted in one day and groups the masses: 0–9 kg holds 5 parcels, 10–19 kg holds 12, 20–29 kg holds 10, 30–39 kg holds 6 and 40–49 kg holds 3.



Made-up data: the store and its parcel masses are made up for this section.

- Which display suits this data, and why?
- Choose a step and a top label for the vertical scale.
- How many parcels weighed 20 kg or more?

Worked answers:

- (a) A histogram is the right display because mass is measured and the data is already grouped into intervals in order: a histogram gives one touching column per group, in order, so the shape of the masses shows clearly. A column chart would be wrong here — its gaps would suggest the groups are unrelated categories.
- (b) The tallest column is 12. Step 2 reaches a top label of 12 in $12 \div 2 = 6$ squares and puts the even frequencies (12, 10, 6) on gridlines. Step 1 would need 12 squares of labels and crowd the page; step 5 would put most column tops (12, 6, 3) between gridlines. So: step 2 with top label 12, as drawn above.
- (c) 20 kg or more means the 20–29, 30–39 and 40–49 groups: $10 + 6 + 3 = 19$ parcels.

∴ a histogram suits the grouped masses, the vertical scale is step 2 with top label 12, and 19 parcels weighed 20 kg or more.

Practice questions

Scaffolded

Q1. A bakery records one day's loaves by type: white 86, wholemeal 57, sourdough 34 and rye. The day's total was 200 loaves. (a) Find the frequency for rye. (b) Write the white loaves as a percentage of the day's total. (c) Name the display you would use to compare the four types, and say why.

Made-up data: the bakery and its loaf counts are made up for this section.

Q2. Use the bottle column chart in the Theory section (the graph with the red part labels). (a) How many bottles were collected on Friday? (b) How many more bottles were collected on Saturday than on Wednesday? (c) What is one grid square on the vertical axis worth?

Q3. A market stall counts its customers each hour: 12, 28, 35, 8, 22, 40. You will draw a column chart on squared paper with 8 squares on the vertical axis. (a) Write down the largest value you must plot. (b) Choose a step for the vertical scale and say why it fits the paper. (c) List the labels your step puts on the axis, from 0 to the top.

Made-up data: the stall and its customer counts are made up for this section.

Q4. Use the shoe-size dot plot in the Theory section. (a) How many pairs of size 9 were sold? (b) How many pairs were sold in total? (c) Which shoe size sold the most?

Q5. Use the steps line graph in the Theory section. (a) How many steps were walked on Wednesday? (b) How many more steps were walked on Saturday than on Sunday? (c) On which day were the most steps walked?

Q6. A youth club asks its 60 members which activity they join most:

Activity	Sport	Music	Art	Gaming
Members	24	15	9	12

- (a) Write the sport share as a percentage of all members.
- (b) Find the sector angle for music in a pie chart of this data.
- (c) Find the sector angle for gaming.

Made-up data: the youth club and its member counts are made up for this section.

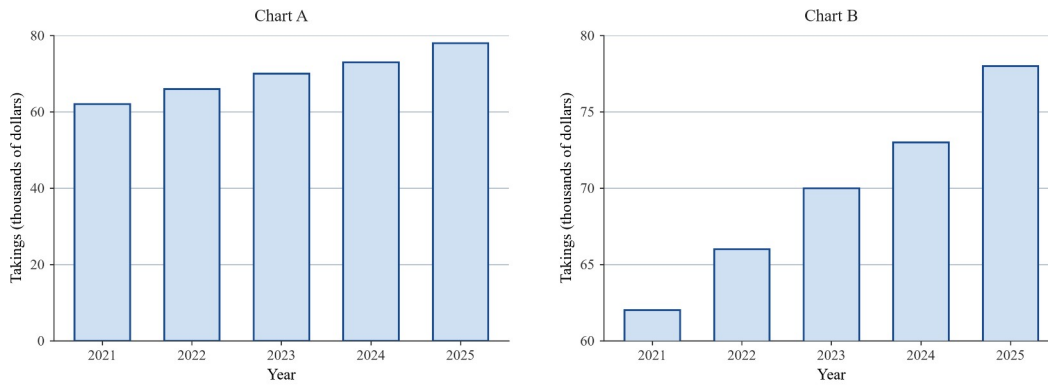
Un scaffolded

Q7. A library counts one week's loans by type: junior fiction 60, adult fiction 45, non-fiction 30, magazines 25, DVDs 15. You will draw a column chart on squared paper with 12 squares on the vertical axis. (a) Say which quantity goes on each axis, and why the vertical scale must start at 0. (b) Choose a step and a top label for the vertical scale, and say why your step suits these values. (c) How many squares tall is the adult fiction column on your scale?

Made-up data: the library and its loan counts are made up for this section.

Q8. Use the waiting-time histogram in the Theory section. (a) Which group holds the most customers, and how many? (b) How many customers waited 20 minutes or longer? (c) Explain why the columns in this graph touch.

Q9. The two column charts below show the same made-up takings of a small shop from 2021 to 2025: \$62 000, \$66 000, \$70 000, \$73 000 and \$78 000.



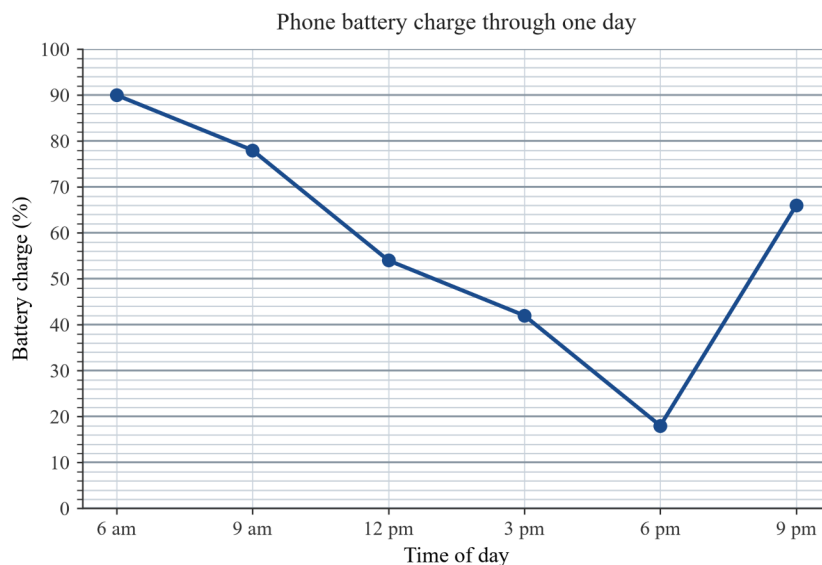
Made-up data: the shop and its takings are made up for this section.

- Which chart makes the rise in takings look bigger?
- Explain how that chart does it.
- Which chart is the fairer picture, and which rule does it follow?

Q10. A fisher counts the fish caught on each of 12 trips: 3, 5, 2, 5, 5, 7, 3, 4, 5, 3, 6, 5. (a) Make a frequency table for these counts. (b) Describe the horizontal axis of a dot plot for this data: which values get a stack of dots? (c) What was the most common catch?

Made-up data: the fisher and the catches are made up for this section.

Q11. The line graph below shows a phone's battery charge through one day. The vertical axis has labels every 10, with 5 grid squares between each pair of labels.



Made-up data: the phone and its battery readings are made up for this section.

- What is one grid square on the vertical axis worth?
- What was the charge at midday?
- The phone was put on charge at 6 pm. By how much did the charge rise between 6 pm and 9 pm?

Q12. Use the harvest pie chart in the Theory section and its table, repeated here:

Vegetable	Tomatoes	Carrots	Beans	Herbs	Silverbeet	Total
Mass (kg)	18	9	6	4	3	40

- What percentage of the harvest was tomatoes?
- Show that the silverbeet sector angle is 27° .
- Find the beans sector angle.

Q13. A takeaway shop counts one weekend's orders by size and by day:

Size	Fri	Sat	Sun
Small	18	25	22
Medium	30	42	35
Large	12	20	16

- How many medium orders were there on Saturday?
- How many large orders were there across the whole weekend?
- How many orders were there in total? Check your total two ways — by rows, then by columns.

Made-up data: the takeaway shop and its orders are made up for this section.

Q14. A TAFE open day counts interest by trade: electrical 14, plumbing 11, carpentry 9, painting 6. (a) Name the display you would draw and say why it suits this data. (b) On squared paper with 8 squares on the vertical axis, choose a step and a top label, and say why a smaller step and a bigger step both fail. (c) List four labelled features your finished graph must carry.

Made-up data: the open day and its counts are made up for this section.

Q15. Use the repair-van pictograph in the Theory section. (a) How many jobs were done on Tuesday? (b) How many jobs were done on Wednesday? (c) How many jobs were done across the whole week?

Q16. Choose the best display for each job, and say why in one sentence. (a) A nurse records a patient's temperature once an hour through a shift and wants to show how it moved. (b) A club splits a \$2000 event budget into rent \$800, food \$500, fuel \$450 and other costs \$250, and wants to show each share of the whole. (c) A coach keeps every sprint time from one race and wants every value on show.

Made-up data: the event budget and its splits are made up for this section.

Q17. A cup of tea cools on a bench while its temperature is graphed. The vertical axis carries the labels 0, 20, 40, 60, 80 (in $^\circ\text{C}$), with 4 grid squares between each pair of labels. (a) What is one grid square worth? (b) A dot sits 3 squares above the 60 label. What temperature is it? (c) How many squares above 0 does 50°C sit?

Made-up data: the tea and its temperatures are made up for this section.

Q18. A canteen counts icy poles sold on each of the first 8 days of term: 12, 18, 25, 30, 28, 20, 14, 10. (a) Choose a step and a top label for the vertical axis of a column chart, and say why. (b) Which day sold the most, and how many? (c) Describe the shape of the week in one sentence.

Made-up data: the canteen and its sales are made up for this section.

Answers

Q1. (a) 23 loaves. (b) 43 %. (c) a column chart — four categories compared by size. **Q2.** (a) 85 bottles. (b) 60 more bottles. (c) 10 bottles. **Q3.** (a) 40. (b) step 5 — 40 sits exactly at the top of the 8th square, filling the paper. (c) 0, 5, 10, 15, 20, 25, 30, 35, 40. **Q4.** (a) 8 pairs. (b) 25 pairs. (c) size 9. **Q5.** (a) 5 000 steps. (b) 1 000 more steps. (c) Saturday. **Q6.** (a) 40 %. (b) 90° . (c) 72° . **Q7.** (a) loan type on the horizontal axis, number of loans on the vertical; columns are compared by height, so the scale starts at 0. (b) step 5, top label 60 — every value is a multiple of 5 and 60 fills the 12 squares. (c) 9 squares. **Q8.** (a) the 10–19 group, with 14 customers. (b) 20 customers. (c) the data is measured and grouped into intervals that run on into each other, so the columns touch. **Q9.** (a) Chart B. (b) its vertical axis starts at 60 000 instead of 0, so the same rise fills most of the picture. (c) Chart A — a column chart's vertical scale starts at 0. **Q10.** (a) 2: 1, 3: 3, 4: 1, 5: 5, 6: 1, 7: 1. (b) one stack at each of 2, 3, 4, 5, 6, 7. (c) 5 fish. **Q11.** (a) 2 %. (b) 54 %. (c) 48 % (from 18 up to 66). **Q12.** (a) 45 %. (b) shown — 27° . (c) 54° . **Q13.** (a) 42 orders. (b) 48 orders. (c) 220 orders. **Q14.** (a) a column chart — categories compared by size. (b) step 2, top label 14 (7 squares); step 1 needs 14 squares and step 5 needs a top label of 15. (c) title; horizontal-axis label; vertical-axis label with units; an even, labelled scale starting at 0. **Q15.** (a) 45 jobs. (b) 20 jobs. (c) 180 jobs. **Q16.** (a) line graph — the change across the shift is the point. (b) pie chart — the four costs are shares of one whole. (c) dot plot — a small data set with every value on show. **Q17.** (a) 5°C . (b) 75°C . (c) 10 squares. **Q18.** (a) step 5, top label 30 (6 squares) — roomy, and the round counts land on gridlines. (b) day 4, with 30 poles. (c) sales rise to day 4, then fall away.

Fully-worked solutions

Q1. (a) Rye is what is left after the other three types: $200 - 86 - 57 - 34 = 23$ loaves. (b) $\frac{86}{200} \times 100 = 43$, so 43 %.

(c) Four categories compared by size: a column chart, one column per type.

Q2. (a) Up from Friday to the column top, then across: half a square above the 80 label, so 85 bottles. (b) Saturday 95, Wednesday 35: $95 - 35 = 60$ more bottles. (c) The labels climb by 10 with one square between them: $10 \div 1 = 10$ bottles a square.

Q3. (a) 40. (b) Step 5: $40 \div 5 = 8$, so 40 sits exactly at the top of the 8th square and the paper is filled, not wasted. Step 2 would need 20 squares — far more than 8 — and step 10 would stop at the 4th square, leaving the top half of the paper empty. (c) 0, 5, 10, 15, 20, 25, 30, 35, 40.

Q4. (a) Count the dots above 9: 8 pairs. (b) Add the stacks: $3 + 6 + 8 + 5 + 2 + 1 = 25$ pairs. (c) The tallest stack stands above size 9.

Q5. (a) Up from Wednesday to the dot, then across: 5 000 steps. (b) Saturday 11 000, Sunday 10 000: $11\,000 - 10\,000 = 1\,000$ more steps. (c) The highest dot is Saturday's.

Q6. (a) $\frac{24}{60} \times 100 = 40$, so 40 %. (b) $\frac{15}{60} \times 360^\circ = 90^\circ$. (c) $\frac{12}{60} \times 360^\circ = 72^\circ$. (Check: art's angle is $\frac{9}{60} \times 360^\circ = 54^\circ$, and the four angles add to 360° .)

Q7. (a) Loan type (the categories) on the horizontal axis; number of loans (the frequency) on the vertical axis, with the units in its label. Columns are compared by height and heights are only honestly measured from zero, so the vertical scale starts at 0. (b) Step 5 with top label 60: every value (60, 45, 30, 25, 15) is a multiple of 5, so every column ends on a gridline, and $60 \div 5 = 12$ squares fills the 12-square axis exactly. Step 10 would leave 45, 25 and 15 between gridlines. (c) $45 \div 5 = 9$ squares.

Q8. (a) The tallest column is the 10–19 group, at 14 customers. (b) The groups from 20 up: $11 + 6 + 3 = 20$ customers. (c) Waiting time is measured, and the groups are intervals that run on into each other with no gaps between them, so the columns touch — the touching is what says “one continuous measured scale”.

Q9. (a) Chart B. (b) Chart B's vertical axis starts at 60 000, not 0. The takings rise by the same \$16 000 on both charts, but on Chart B that rise fills most of the picture, so it reads as a jump instead of a steady climb. (c) Chart A is the fairer picture: its vertical scale starts at 0, so column heights tell the truth about size.

Q10. (a) Tally each value:

Fish per trip	2	3	4	5	6	7
Trips	1	3	1	5	1	1

The frequencies add to 12, the number of trips. (b) One stack of dots at each value that appears: 2, 3, 4, 5, 6 and 7 on the horizontal axis, with one dot per trip stacked above its value. (c) The tallest stack stands above 5, so 5 fish was the most common catch.

Q11. (a) Labels every 10 with 5 squares between: $10 \div 5 = 2$, so one square is worth 2%. (b) Up from 12 pm to the dot, then across: 2 squares above the 50 label, $50 + 2 \times 2 = 54$, so 54%. (c) 6 pm reads 18 and 9 pm reads 66: $66 - 18 = 48$, a rise of 48% while the phone charged.

Q12. (a) $\frac{18}{40} \times 100 = 45$, so 45%. (b) One step per line:

$$\text{angle} = \frac{3}{40} \times 360^\circ$$

$$\text{angle} = 27^\circ$$

which is the 27° required. (c) $\frac{6}{40} \times 360^\circ = 54^\circ$.

Q13. (a) Medium row, Saturday column: 42 orders. (b) The large row: $12 + 20 + 16 = 48$ orders. (c) By rows: small $18 + 25 + 22 = 65$, medium $30 + 42 + 35 = 107$, large 48; then $65 + 107 + 48 = 220$. By columns: Fri $18 + 30 + 12 = 60$, Sat $25 + 42 + 20 = 87$, Sun $22 + 35 + 16 = 73$; then $60 + 87 + 73 = 220$. Both ways agree: 220 orders.

Q14. (a) A column chart: four categories (the trades) compared by size, with no time order and no whole to share. (b) Step 2 with top label 14: $14 \div 2 = 7$ of the 8 squares used, one to spare. Step 1 would need 14 squares — nearly twice the paper — and step 5 would need a top label of 15 and would leave 11, 9 and 6 between gridlines. (c) The four labelled features: a title; a horizontal-axis label naming the categories; a vertical-axis label naming what is counted, with the units; and an even, labelled scale starting at 0.

Q15. (a) Tuesday shows 4 whole circles and a half: $4 \times 10 + 5 = 45$ jobs. (b) Wednesday shows 2 whole circles: $2 \times 10 = 20$ jobs. (c) Add the days: $30 + 45 + 20 + 50 + 35 = 180$ jobs.

Q16. (a) A line graph: the readings run in time order and the rise and fall across the shift is the point. (b) A pie chart: the four costs are shares of one whole — and they do add to it: $800 + 500 + 450 + 250 = 2000$. (c) A dot plot: a small data set in which every single value earns its own dot.

Q17. (a) $(20 - 0) \div 4 = 5$, so one square is worth 5°C . (b) $60 + 3 \times 5 = 75^\circ\text{C}$. (c) $50 \div 5 = 10$ squares above 0.

Q18. (a) The biggest day is 30, so step 5 with top label 30: $30 \div 5 = 6$ squares — roomy enough to compare the days, and the round counts (25, 30, 20, 10) land on gridlines. (b) The tallest column is day 4, at 30 poles. (c) Sales climb to day 4 and then fall away — a rise-then-fall shape.

Chapter 8 Data — 8C Measures of centre: mean, median and mode

Theory

A data set can be long. A week of sales figures, a season of scores, a list of wages — too many numbers to hold in your head at once. So we squeeze the set down to one number that stands for it. A **measure of centre** is a single number that sits near the middle of a data set and describes the whole set in one value. This section teaches the three common measures of centre — the **mean**, the **median** and the **mode** — and, just as importantly, how to choose the right one for the data in front of you.

About the data in this section. Every business, team, person and number in this section is **made up** (not real). They are built for this book so you can practise finding and choosing measures of centre. Nothing here is real data about a real person, business or place.

The mean.

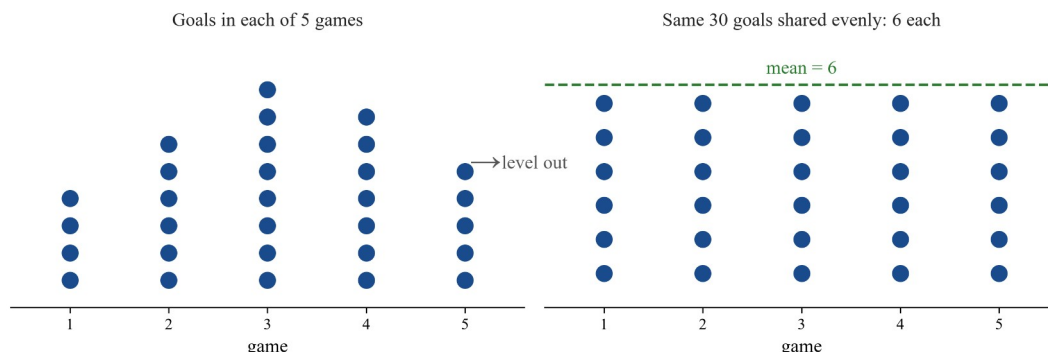
The **mean** is the measure of centre most people call “the average”. Add up all the values, then divide by how many values there are:

$$\text{mean} = \frac{\text{sum of all values}}{\text{number of values}}$$

In formulas the mean is often written $\bar{x}$, said “x-bar”.

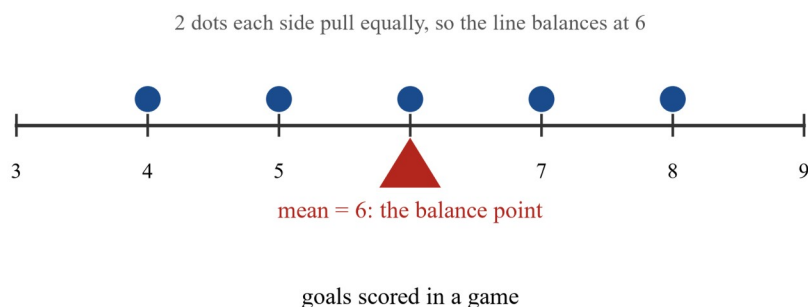
A team scores 4, 6, 8, 7 and 5 goals in five games. The sum is $4 + 6 + 8 + 7 + 5 = 30$ goals, and there are 5 values, so the mean is $30 \div 5 = 6$ goals a game.

One way to picture the mean is to **level out** the data: take the same 30 goals and share them evenly across the five games.



Made-up data: the team and every score in this picture are made up for this section.

Every tower ends at 6, so 6 goals a game is the levelled-out score — the mean. The same picture works as a balance: put one dot on a number line for each value, and the mean is the point where the line balances.



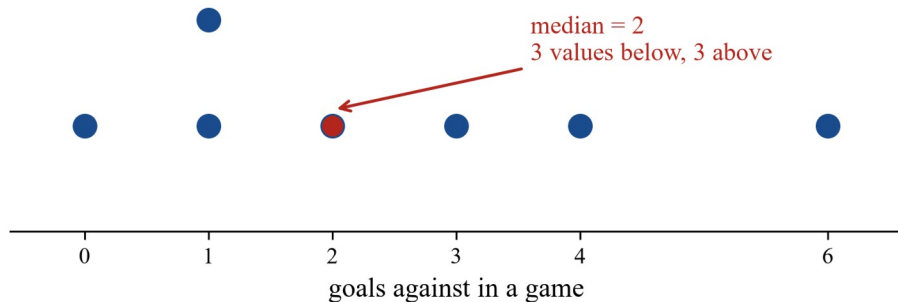
Made-up data: the scores in this picture are made up for this section.

The 4 and 5 on the left pull down just as hard as the 7 and 8 on the right pull up — 2 dots each side pull equally — so the line balances at 6, the mean. The dot on 6 itself sits over the fulcrum and pulls neither way.

The median.

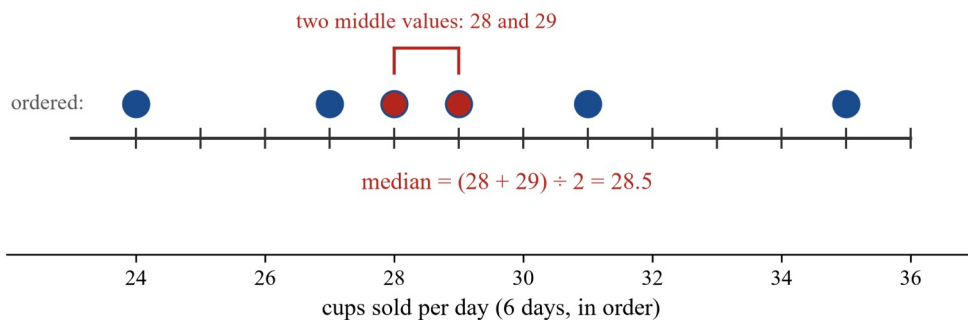
The **median** is the middle value once the data is written in order, from smallest to largest. With an odd number of values, one value sits exactly in the middle. A team concedes 0, 1, 1, 2, 3, 4 and 6 goals in seven games — already in order — and the middle value is 2:

7 games: the middle value is the median



Made-up data: the team and every score in this picture are made up for this section.

Three values sit below the 2 and three above it, which is exactly what “middle” means. With an even number of values, two values share the middle, and the median is halfway between them: add the two middle values and divide by 2. A café sells 24, 31, 28, 35, 27 and 29 cups of coffee over six days. In order the data is 24, 27, 28, 29, 31, 35, so the two middle values are 28 and 29:



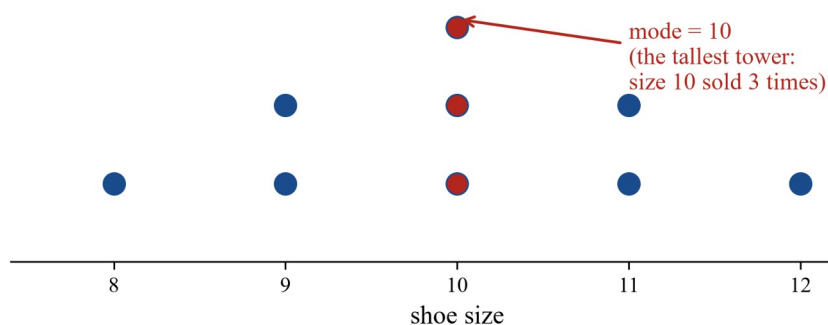
Made-up data: the café and its daily cup counts are made up for this section.

The median is $(28 + 29) \div 2 = 28.5$ cups. A median can land between two values like this; it does not have to be a value from the data set.

The mode.

The **mode** is the value that appears most often. No adding and no ordering — just count how many times each value turns up. A shoe shop records the sizes it sells in one day: 8, 9, 9, 10, 10, 10, 11, 11, 12.

Shoe sizes sold in one day

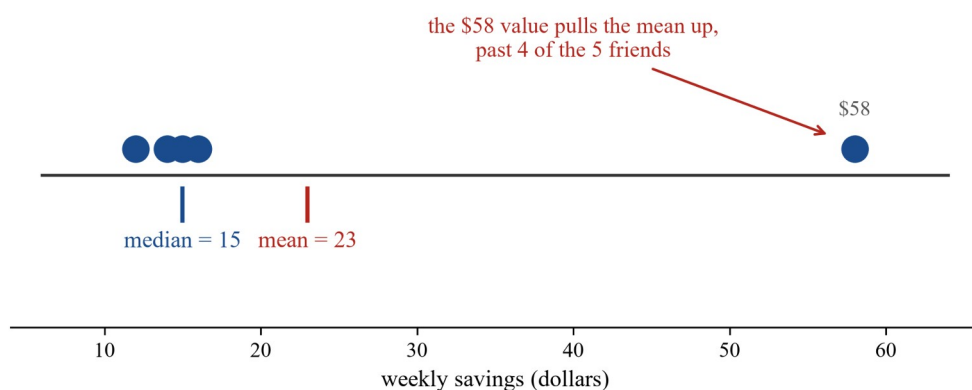


Made-up data: the shop and its sales in this picture are made up for this section.

Size 10 sold 3 times, more than any other size, so the mode is 10. (In this particular set the mean and the median both come to 10 as well — sometimes all three measures agree.) Some sets have no mode at all, when every value appears just once, and some have two modes, when two values tie for most often.

When the mean and the median tell different stories.

The mean uses every value in its sum, so one far-out value can drag it a long way. Five friends save \$12, \$15, \$14, \$16 and \$58 a week:



Made-up data: the five friends and their savings are made up for this section.

The mean is $(12 + 15 + 14 + 16 + 58) \div 5 = 115 \div 5 = \$23$, while the median — the middle of 12, 14, 15, 16, 58 — is \$15. Four of the five friends save **less** than the mean of \$23; the one big saving of \$58 pulls the mean up past almost the whole group. The median stays with the cluster, so here it is the better picture of a typical friend's savings. Values that sit far away from the rest of the data, like the \$58 here, change the mean much more than they change the median. Section 8D gives those far-out values a name and looks at how spread out a data set is.

Data in a frequency table.

When values repeat, it is shorter to list each value once with its **frequency** — the number of times it appears — than to write the whole list out. A street counts the cars in each household:

Cars, x	0	1	2	3	4
Households, f	2	7	6	4	2
$f \times x$	0	7	12	12	8

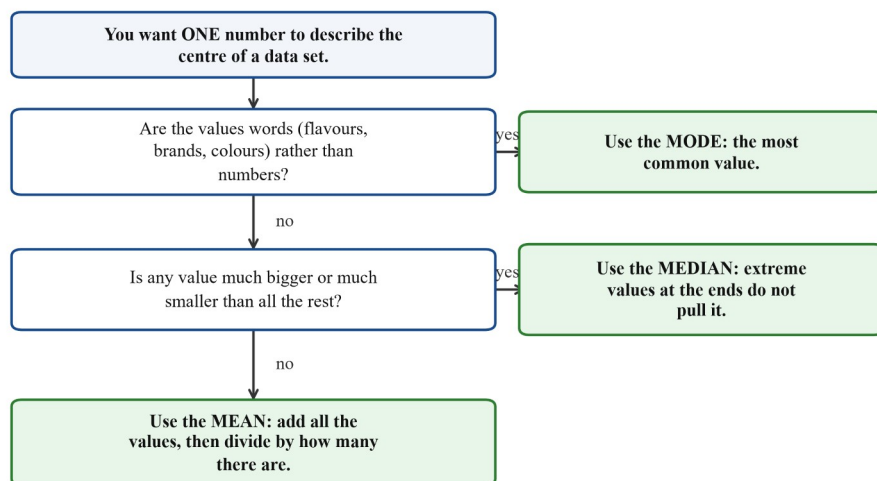
Made-up data: the street and its households are made up for this section.

The frequencies add to $2 + 7 + 6 + 4 + 2 = 21$ households. To add up the cars without writing 21 numbers out, multiply each value by its frequency (the $f \times x$ row) and add that row: $0 + 7 + 12 + 12 + 8 = 39$ cars. So the mean is $39 \div 21 = 1.857 \dots$, which is 1.9 cars to one decimal place.

The median is the middle of the 21 values in order, the 11th value. Count up through the frequencies: the 1st and 2nd values are 0, the 3rd to 9th are 1, and the 10th to 15th are 2 — so the 11th value is 2 cars. The mode is the value with the biggest frequency: 1 car, with 7 households.

Choosing the right measure.

All three measures are correct arithmetic; the skill is picking the one that tells the truth about your data.



Whichever measure you choose, say which one you used — the three measures give different numbers and different stories.

Made-up data: this chart describes the methods in this section, not real data.

Word data — flavours, brands, colours — has no sum and no order, so the mode is the only measure of centre that works. Numbers with a far-out value at one end are best described by the median, which the far-out value cannot pull. Numbers with nothing far-out are best described by the mean, because the mean is the only measure that uses every value. Whichever you choose, say which one you used: the three measures give different numbers and different stories.

Rounding and checking.

When an answer needs rounding, the question tells you how far — to one decimal place, to the nearest cent, to the nearest whole number. Keep the full value in your working and round only at the end, then say in your answer how far you rounded. Afterwards, sanity-check the number: a measure of centre should sit somewhere among the values, not off past the ends of the data. If your “centre” is bigger than every value in the set, something has gone wrong.

Tools that help.

A calculator or a spreadsheet will add up a long list and divide in seconds, and at work you will often let it. But the tool cannot tell you which measure suits the data, and it will happily give you a mean of a list of shoe colours if you ask. Choosing the measure and checking the answer against the data stay your job.

Worked examples

Example 1 — scaffolded

A bakery counts the loaves of bread it sells each hour over a nine-hour market day: 42, 55, 48, 55, 60, 38, 55, 47 and 50 loaves. Find the mean, the median and the mode of this data.

Working	Comment
Sum: $42 + 55 + 48 + 55 + 60 + 38 + 55 + 47 + 50 = 450$ loaves.	Add every value before anything else.
Mean: $450 \div 9 = 50$ loaves.	Nine hours were counted, so divide by 9.
In order: 38, 42, 47, 48, 50, 55, 55, 55, 60.	Always order the data before finding the median.
Median: the 5th of the 9 values is 50 loaves.	With 9 values, 4 sit below the middle and 4 above.
Mode: 55 loaves, which appears 3 times; every other value appears once.	The mode is the most common value.
Mean 50, median 50, mode 55 loaves per hour.	Answer everything that was asked, with units.

Example 2 — scaffolded

A print shop counts the online orders it gets in an hour, over 13 hours:

Orders in one hour, x	7	8	9	10	11
Hours, f	2	3	4	2	2
$f \times x$	14	24	36	20	22

Made-up data: the print shop and its order counts are made up for this section.

The same data is drawn below. Find the mean (to one decimal place), the median and the mode.

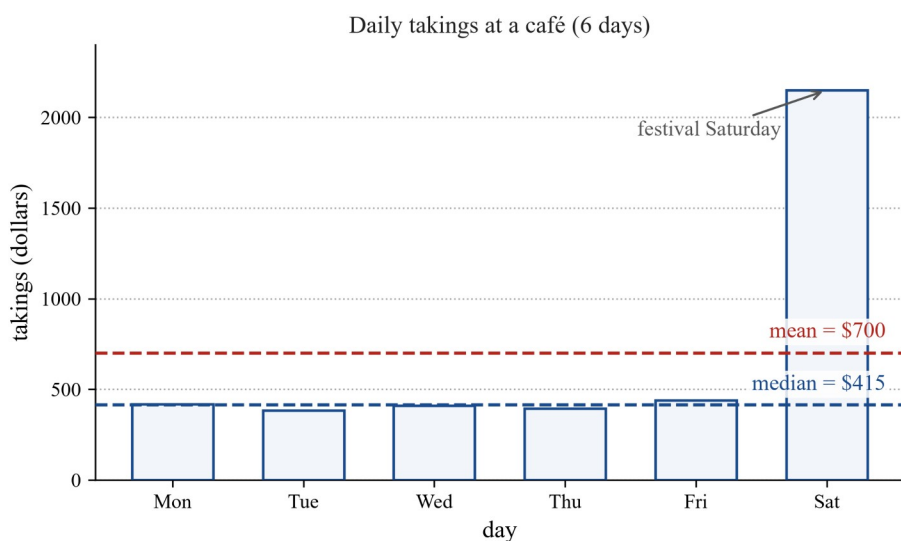


Made-up data: every number in this picture is made up for this section.

Working	Comment
Total hours: $2 + 3 + 4 + 2 + 2 = 13$.	Add the frequency row to count the values.
Total orders: $14 + 24 + 36 + 20 + 22 = 116$.	Add the $f \times x$ row instead of writing out 13 numbers.
Mean: $116 \div 13 = 8.923\dots$, so 8.9 orders to one decimal place.	Keep the full value in working; round once, at the end.
Median: the 7th of 13 values. Counting up, values 1–2 are 7, 3–5 are 8, 6–9 are 9 — so the 7th is 9 orders.	The middle of an odd number of values is value number $(13 + 1) \div 2 = 7$.
Mode: 9 orders, with frequency 4 — the tallest bar on the graph.	The biggest frequency marks the mode.
Mean 8.9 (one decimal place), median 9, mode 9 orders per hour.	State the rounding you used.

Example 3 — unscaffolded

A café's daily takings over six days are \$ 420, \$ 385, \$ 410, \$ 395, \$ 440 and \$ 2 150 — the last day was a festival Saturday.



Made-up data: the café and its takings are made up for this section.

- (a) Find the mean and the median of the takings.
- (b) Which of the two measures better describes a typical day at this café? Give a reason using the numbers.
- (c) The sum is $420 + 385 + 410 + 395 + 440 + 2\,150 = \$4\,200$, so the mean is $4\,200 \div 6 = \$700$. In order the takings are 385, 395, 410, 420, 440, 2 150; the two middle values are 410 and 420, so the median is $(410 + 420) \div 2 = \$415$.
- (d) The median. Five of the six days took less than the mean of \$ 700 — the festival Saturday drags the mean up past every ordinary day. The median of \$ 415 sits in among the ordinary days, so it is the honest picture of a typical day.

$\therefore$ the mean is \$ 700 and the median is \$ 415, and the median better describes a typical day.

Example 4 — unscaffolded

A basketballer scores 14, 22, 18, 26 and 15 points in her first five games. She wants a mean of exactly 20 points per game over her first six games. How many points does she need in game six?

The five games so far add to $14 + 22 + 18 + 26 + 15 = 95$ points. A mean of 20 over 6 games needs a total of $20 \times 6 = 120$ points. Game six must supply the gap: $120 - 95 = 25$ points. Check: $(95 + 25) \div 6 = 120 \div 6 = 20$ ✓ — the mean comes out at exactly 20.

$\therefore$ she needs 25 points in game six.

Practice questions

Scaffolded

Q1. Find the mean, the median and the mode of the data set 7, 3, 9, 3, 8. (a) Find the mean. (b) Find the median. (c) Find the mode.

Q2. Find the measures of centre of the data set 15, 9, 17, 11, 8, 13. (a) Find the mean, to one decimal place. (b) Find the median. (c) Find the mode, or explain why there is none.

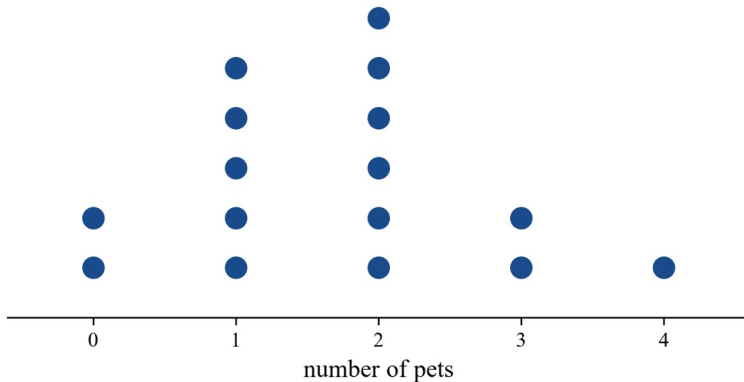
Q3. Students in a class count the pets living in their house:

Pets, x	0	1	2	3	4
Students, f	2	5	6	2	1

Made-up data: the class and its pet counts are made up for this section.

The same data is drawn below.

Pets living in each student's house (16 students)



Made-up data: every number in this picture is made up for this section.

- How many students were surveyed?
- Write down the mode.
- Find the mean number of pets, to one decimal place.

Q4. A canteen records which pies it sells one lunchtime:

Flavour	Meat	Chicken	Veggie	Cheese
Pies sold	38	22	12	18

Made-up data: the canteen and its pie sales are made up for this section.

- How many pies were sold altogether?
- Write down the mode.
- Explain why the mean and the median cannot be found for this data, and why the mode still works.

Q5. A team's goals per match over a season are recorded in this table:

Goals, x	4	5	6	7	8
Matches, f	2	1	4	2	1

Made-up data: the team and its scores are made up for this section.

- Find the mean number of goals per match.
- Find the median number of goals per match.
- Write down the mode.

Q6. The maximum temperatures ($^{\circ}\text{C}$) in one town over eight days are 21, 18, 24, 19, 26, 22, 17, 24. (a) Find the median temperature. (b) Find the mean temperature, to one decimal place. (c) Write down the mode.

Unscaffolded

Q7. The mean of five numbers is 16. Four of the numbers are 12, 18, 14 and 17. Find the fifth number.

Q8. Seven workers at a small firm earn \$ 620, \$ 650, \$ 620, \$ 700, \$ 680, \$ 620 and \$ 1 450 a week. (a) Find the mean weekly wage, to the nearest cent. (b) Find the median and the mode of the wages. (c) Which measure — mean, median or mode — best describes a typical wage at this firm? Give a reason using the numbers.

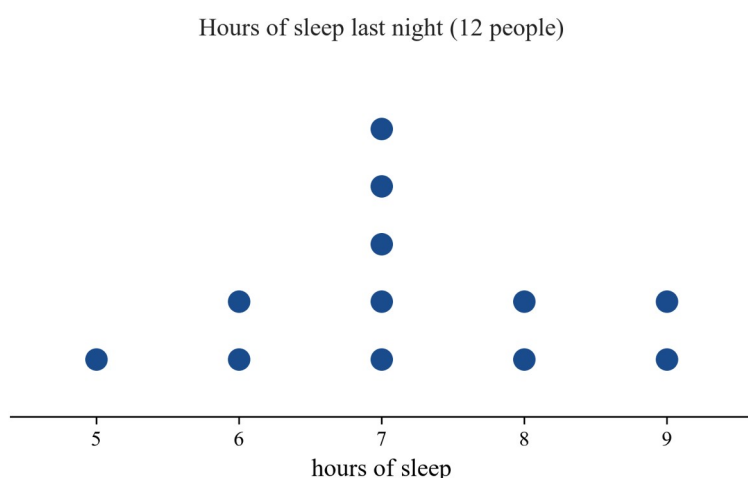
Q9. The mean of eight numbers is 12. A ninth number, 21, is added to the set. Find the new mean.

Q10. People in a building record how many hours they slept last night:

Hours of sleep, x	5	6	7	8	9
People, f	1	2	5	2	2

Made-up data: the building and its sleep hours are made up for this section.

The same data is drawn below.



Made-up data: every number in this picture is made up for this section.

- Find the mean hours of sleep, to one decimal place.
- Find the median.
- Write down the mode.

Q11. A driver times two routes to work, eight trips each, in minutes. Route A: 22, 25, 24, 31, 23, 25, 24, 26. Route B: 18, 35, 20, 19, 34, 21, 20, 33. (a) Find the mean time for each route. (b) Find the median time for each route. (c) The two means are the same, yet the routes are not alike. Use the medians to explain what the mean hides about Route B.

Q12. A household's six electricity bills are \$ 87.40, \$ 92.15, \$ 84.60, \$ 96.30, \$ 88.75 and \$ 91.45. (a) Find the mean bill, to the nearest cent. (b) Find the median bill. (c) Write down the mode, or explain why there is none.

Q13. A nurse records a patient's resting heart rate (beats per minute) on nine days: 62, 68, 65, 71, 64, 66, 63, 67, 64. (a) Find the mean, to the nearest whole beat. (b) Find the median. (c) Find the mode, then say whether all three of your measures sit inside the 60–70 beats-per-minute band.

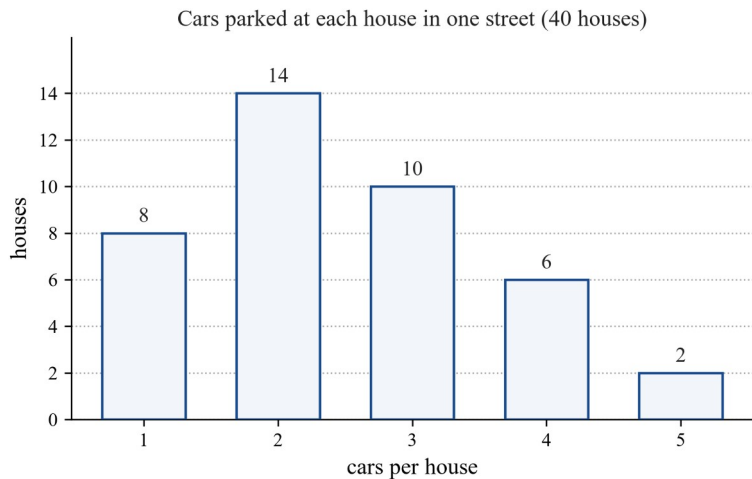
Q14. A courier's delivery times (minutes) over eight deliveries are 34, 41, 38, 35, 92, 40, 36, 39. One entry, the 92, turns out to be a mis-scan with no valid reading behind it, so it is deleted from the record entirely. (a) Find the mean (to one decimal place) and the median of the eight recorded times. (b) Find the mean (to one decimal place) and the median of the seven times left after the 92 is deleted. (c) How much did the mean fall, to one decimal place? Compare this with how much the median moved, and say in one sentence why the two measures behaved so differently.

Q15. Houses in a street are surveyed for the number of cars parked at them:

Cars, x	1	2	3	4	5
Houses, f	8	14	10	6	2

Made-up data: the street and its houses are made up for this section.

The same data is drawn below.



Made-up data: every number in this picture is made up for this section.

- Find the mean number of cars per house.
- Find the median.
- Write down the mode.

Q16. Rainfall (mm) over ten days at a farm is 0, 0, 0, 4, 2, 0, 1, 8, 3, 0. (a) Find the mean daily rainfall. (b) Find the median. (c) Write down the mode, then say in one sentence which measure — mean, median or mode — best describes a typical day's rain at this farm, and why.

Q17. The weekly hours worked by 20 casual staff at a venue are recorded:

Hours, x	8	10	12	15	20
Staff, f	3	5	6	4	2

Made-up data: the venue and its roster are made up for this section.

- Find the mean hours worked per week.
- Find the median.
- Write down the mode.

Q18. Mia spends \$4.50, \$6.20, \$3.80 and \$5.50 on lunches from Monday to Thursday. She wants her mean lunch cost over the five weekdays to be at most \$5.00. (a) Find the total she spent from Monday to Thursday. (b) Find the most she can spend on Friday so that the five-day mean is \$5.00 or less. (c) Show by calculation that spending one cent more than your answer to (b) puts the five-day mean above \$5.00.

Answers

Q1. (a) 6. (b) 7. (c) 3. **Q2.** (a) 12.2 (to one decimal place). (b) 12. (c) no mode — every value appears exactly once. **Q3.** (a) 16 students. (b) 2 pets. (c) 1.7 pets (to one decimal place). **Q4.** (a) 90 pies. (b) meat. (c) flavours are words, not numbers — there is nothing to add and no order — so only the mode works. **Q5.** (a) 5.9 goals. (b) 6 goals. (c) 6 goals. **Q6.** (a) 21.5 °C. (b) 21.4 °C (to one decimal place). (c) 24 °C. **Q7.** 19. **Q8.** (a) \$762.86 (to the nearest cent). (b) median \$650, mode \$620. (c) the median — the one \$1450 wage pulls the mean up above 6 of the 7 wages. **Q9.** 13. **Q10.** (a) 7.2 hours (to one decimal place). (b) 7 hours. (c) 7 hours. **Q11.** (a) both 25 minutes. (b) A 24.5 minutes, B 20.5 minutes. (c) B's median of 20.5 says half of B's trips take 20 minutes or less; the mean of 25 hides the three slow trips of 33–35 minutes that drag it up, so B is usually quicker than its mean suggests but sometimes much slower. **Q12.** (a) \$90.11 (to the nearest cent). (b) \$90.10. (c) no mode — all six bills are different. **Q13.** (a) 66 beats per minute (to the nearest whole beat). (b) 65. (c) 64; yes — 66, 65 and 64 all sit inside the 60–70 band. **Q14.** (a) mean 44.4 minutes (to one decimal place), median 38.5 minutes. (b) mean 37.6 minutes (to one decimal place), median 38 minutes. (c) the mean fell 6.8 minutes (to one decimal place) while the median moved only 0.5 — one far-out time drags the mean a long way and barely moves the middle. **Q15.** (a) 2.5 cars. (b) 2 cars. (c) 2 cars. **Q16.** (a) 1.8 mm. (b) 0.5 mm. (c) 0 mm; the mode (or the median) — five of the ten days were dry, and the two wet days pull the mean up above the rainfall of six of the ten days. **Q17.** (a) 12.3 hours. (b) 12 hours. (c) 12 hours. **Q18.** (a) \$20.00. (b) \$5.00. (c) at \$5.01 the mean is $25.01 \div 5 = \$5.002$, which is above \$5.00.

Fully-worked solutions

Q1. (a) The sum is $7 + 3 + 9 + 3 + 8 = 30$ over 5 values, so the mean is $30 \div 5 = 6$. (b) In order: 3, 3, 7, 8, 9; the middle of the five values is 7. (c) The value 3 appears twice and every other value once, so the mode is 3.

Q2. (a) The sum is $15 + 9 + 17 + 11 + 8 + 13 = 73$ over 6 values, so the mean is $73 \div 6 = 12.166\dots$, which is 12.2 to one decimal place. (b) In order: 8, 9, 11, 13, 15, 17; the two middle values are 11 and 13, so the median is $(11 + 13) \div 2 = 12$. (c) Every value appears exactly once, so no value is more common than the rest: this set has no mode.

Q3. (a) Add the frequency row: $2 + 5 + 6 + 2 + 1 = 16$ students. (b) The biggest frequency is 6, against 2 pets, so the mode is 2 pets — the tallest tower on the dot plot. (c) Add an $f \times x$ row: $0 \times 2 = 0$, $1 \times 5 = 5$, $2 \times 6 = 12$, $3 \times 2 = 6$, $4 \times 1 = 4$, which adds to $0 + 5 + 12 + 6 + 4 = 27$ pets. The mean is $27 \div 16 = 1.6875$, which is 1.7 pets to one decimal place.

Q4. (a) $38 + 22 + 12 + 18 = 90$ pies. (b) Meat sold 38 times, more than any other flavour, so the mode is meat. (c) The mean needs a sum, and flavours are words — there is nothing to add. The median needs the values in order from smallest to largest, and words have no order. The mode only needs counting, and counting words works fine, so the mode is the only measure of centre this data can have.

Q5. (a) The $f \times x$ row is 8, 5, 24, 14, 8, adding to 59 goals; the frequencies add to 10 matches, so the mean is $59 \div 10 = 5.9$ goals. (b) With 10 values the median is halfway between the 5th and 6th. Counting up: values 1–2 are 4, value 3 is 5, values 4–7 are 6 — so the 5th and 6th values are both 6, and the median is $(6 + 6) \div 2 = 6$ goals. (c) The biggest frequency is 4, against 6 goals, so the mode is 6 goals.

Q6. (a) In order: 17, 18, 19, 21, 22, 24, 24, 26. The two middle values are 21 and 22, so the median is $(21 + 22) \div 2 = 21.5$ °C. (b) The sum is $21 + 18 + 24 + 19 + 26 + 22 + 17 + 24 = 171$ over 8 days, so the mean is $171 \div 8 = 21.375$, which is 21.4 °C to one decimal place. (c) The value 24 appears twice and every other value once, so the mode is 24 °C.

Q7. A mean of 16 over 5 numbers means the five add to $16 \times 5 = 80$. The four known numbers add to $12 + 18 + 14 + 17 = 61$, so the fifth is $80 - 61 = 19$. Check: $(61 + 19) \div 5 = 80 \div 5 = 16$ ✓.

Q8. (a) The sum is $620 + 650 + 620 + 700 + 680 + 620 + 1450 = \$5340$ over 7 wages, so the mean is $5340 \div 7 = 762.857142\dots$, which is \$762.86 to the nearest cent. (b) In order: 620, 620, 620, 650, 680, 700, 1450; the middle of the seven values is the 4th, \$650, so the median is \$650. The wage \$620 appears three times and

every other wage once, so the mode is \$ 620. (c) The median. Six of the seven wages sit below the mean of \$ 762.86 — the single \$ 1 450 wage drags the mean up past nearly everyone — while the median of \$ 650 sits in among the ordinary wages and describes what a typical worker here actually earns.

Q9. A mean of 12 over 8 numbers means they add to $12 \times 8 = 96$. Adding 21 gives a sum of $96 + 21 = 117$ over 9 numbers, so the new mean is $117 \div 9 = 13$.

Q10. (a) The $f \times x$ row is 5, 12, 35, 16, 18, adding to 86 hours; the frequencies add to 12 people, so the mean is $86 \div 12 = 7.166\dots$, which is 7.2 hours to one decimal place. (b) With 12 values the median is halfway between the 6th and 7th. Counting up: value 1 is 5, values 2–3 are 6, values 4–8 are 7 — so the 6th and 7th values are both 7, and the median is 7 hours. (c) The biggest frequency is 5, against 7 hours, so the mode is 7 hours.

Q11. (a) Route A: $22 + 25 + 24 + 31 + 23 + 25 + 24 + 26 = 200$ over 8 trips, mean $200 \div 8 = 25$ minutes. Route B: $18 + 35 + 20 + 19 + 34 + 21 + 20 + 33 = 200$ over 8 trips, mean $200 \div 8 = 25$ minutes. (b) Route A in order: 22, 23, 24, 24, 25, 25, 26, 31; the middle two are 24 and 25, so the median is $(24 + 25) \div 2 = 24.5$ minutes. Route B in order: 18, 19, 20, 20, 21, 33, 34, 35; the middle two are 20 and 21, so the median is $(20 + 21) \div 2 = 20.5$ minutes. (c) Route A's median (24.5) sits close to its mean (25): seven of A's eight trips fall between 22 and 26 minutes. Route B's median (20.5) is well below its mean (25): half of B's trips take 20 minutes or less, around 20 minutes, and the mean only reaches 25 because three slow trips of 33, 34 and 35 minutes drag it up. The same mean of 25 therefore hides that B is usually the quicker route but occasionally much slower.

Q12. (a) The sum is $87.40 + 92.15 + 84.60 + 96.30 + 88.75 + 91.45 = \$ 540.65$ over 6 bills, so the mean is $540.65 \div 6 = 90.1083\dots$, which is \$ 90.11 to the nearest cent. (b) In order: 84.60, 87.40, 88.75, 91.45, 92.15, 96.30; the two middle values are 88.75 and 91.45, so the median is $(88.75 + 91.45) \div 2 = 180.20 \div 2 = \$ 90.10$. (c) All six bills are different amounts, so no value appears more often than the rest: this set has no mode.

Q13. (a) The sum is $62 + 68 + 65 + 71 + 64 + 66 + 63 + 67 + 64 = 590$ over 9 readings, so the mean is $590 \div 9 = 65.555\dots$, which is 66 beats per minute to the nearest whole beat. (b) In order: 62, 63, 64, 64, 65, 66, 67, 68, 71; the middle of the nine values is the 5th, 65. (c) The value 64 appears twice and every other value once, so the mode is 64. All three measures — 66, 65 and 64 — sit inside the 60–70 band, so they agree that this patient's resting rate is in the 60s.

Q14. (a) The sum of the eight times is $34 + 41 + 38 + 35 + 92 + 40 + 36 + 39 = 355$, so the mean is $355 \div 8 = 44.375$, which is 44.4 minutes to one decimal place. In order: 34, 35, 36, 38, 39, 40, 41, 92; the middle two are 38 and 39, so the median is $(38 + 39) \div 2 = 38.5$ minutes. (b) Without the 92 the sum is $34 + 41 + 38 + 35 + 40 + 36 + 39 = 263$, so the mean is $263 \div 7 = 37.571\dots$, which is 37.6 minutes to one decimal place. In order: 34, 35, 36, 38, 39, 40, 41; the middle of the seven values is the 4th, 38 minutes. (c) The mean fell $44.375 - 37.571\dots = 6.803\dots$, which is 6.8 minutes to one decimal place, while the median moved only $38.5 - 38 = 0.5$ minutes. The mean adds every value into its sum, so one far-out time drags it a long way; the median only cares about the middle of the ordered list, so deleting the top value shifts it by half a step at most.

Q15. (a) The $f \times x$ row is 8, 28, 30, 24, 10, adding to 100 cars; the frequencies add to 40 houses, so the mean is $100 \div 40 = 2.5$ cars. (b) With 40 values the median is halfway between the 20th and 21st. Counting up: values 1–8 are 1, values 9–22 are 2 — so the 20th and 21st values are both 2, and the median is 2 cars. (c) The biggest frequency is 14, against 2 cars, so the mode is 2 cars.

Q16. (a) The sum is $0 + 0 + 0 + 4 + 2 + 0 + 1 + 8 + 3 + 0 = 18$ mm over 10 days, so the mean is $18 \div 10 = 1.8$ mm. (b) In order: 0, 0, 0, 0, 0, 1, 2, 3, 4, 8; the two middle values are 0 and 1, so the median is $(0 + 1) \div 2 = 0.5$ mm. (c) The value 0 appears five times, more than any other, so the mode is 0 mm. The mode (or the median) tells the typical day best: five of the ten days had no rain at all, and six of the ten days had less than the mean of 1.8 mm, because the 4 mm and 8 mm days pull it up.

Q17. (a) The $f \times x$ row is 24, 50, 72, 60, 40, adding to 246 hours; the frequencies add to 20 staff, so the mean is $246 \div 20 = 12.3$ hours. (b) With 20 values the median is halfway between the 10th and 11th. Counting up: values 1–3

are 8, 4–8 are 10, 9–14 are 12 — so the 10th and 11th values are both 12, and the median is 12 hours. (c) The biggest frequency is 6, against 12 hours, so the mode is 12 hours.

Q18. (a) $4.50 + 6.20 + 3.80 + 5.50 = \$20.00$. (b) A five-day mean of \$5.00 needs a five-day total of $5 \times 5.00 = \$25.00$, so Friday can be at most $25.00 - 20.00 = \$5.00$. (c) At \$5.01 the total is $20.00 + 5.01 = \$25.01$ and the mean is $25.01 \div 5 = \$5.002$, which is above \$5.00 — so one cent more than \$5.00 breaks the goal, and \$5.00 is exactly the most she can spend.

Theory

Section 8C squeezed a data set down to one number for its middle — the mean, the median or the mode. But a middle is not the whole story. Two lists can share the same mean and still feel completely different: one steady and predictable, the other all over the place. A **measure of spread** is a single number that says how scattered a data set is — how far apart the values sit from each other. This section teaches the range, the simplest measure of spread, and the **outlier**: the odd value far away from the rest that can bend every other measure out of shape.

About the data in this section. Every business, team, person and number in this section is **made up** (not real). They are built for this book so you can practise measuring spread and spotting outliers. Nothing here is real data about a real person, business or place.

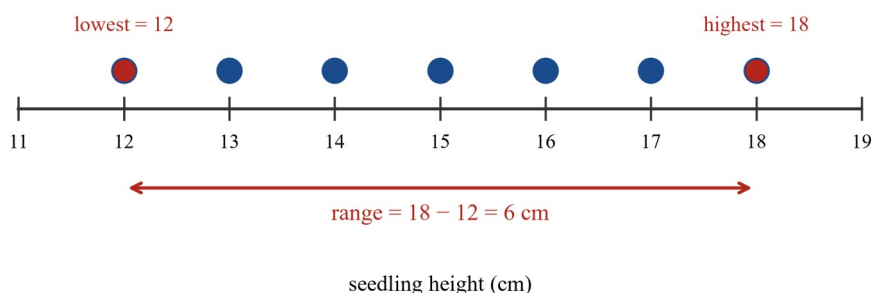
Spread and the range.

The **spread** of a data set is how much the values vary — whether they sit close together or stretch out over a wide band of numbers. The simplest measure of spread is the **range**: the gap between the highest value and the lowest value.

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

A garden centre measures seven seedlings: 12, 15, 14, 18, 13, 16 and 17 cm tall. The lowest is 12 cm and the highest is 18 cm, so the range is $18 - 12 = 6$ cm.

Seven seedlings: the spread from lowest to highest



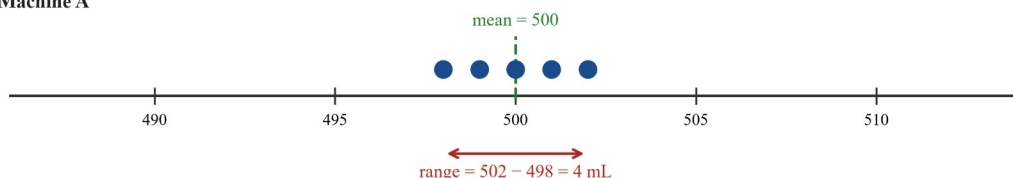
Made-up data: the seedlings and every height in this picture are made up for this section.

Ordering the data first — 12, 13, 14, 15, 16, 17, 18 — makes the lowest and highest easy to see, and the range is then one subtraction. Keep the units: a range of heights is a number of centimetres, a range of times is a number of minutes. The range uses only two values out of the whole set, the two at the ends, which makes it quick to find — and, as you will see below, easy to wreck with one odd value.

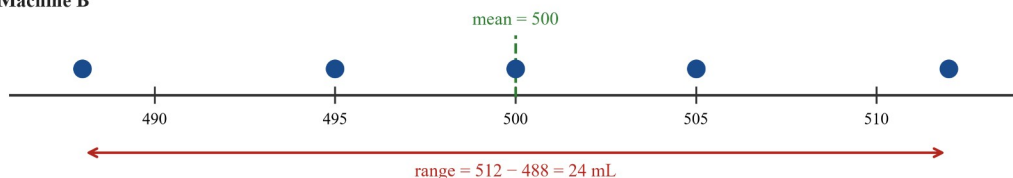
Same centre, different spread.

Two machines fill bottles meant to hold 500 mL. Five bottles from each machine are measured. Machine A gives 498, 501, 500, 499 and 502 mL. Machine B gives 488, 512, 500, 495 and 505 mL. Both sets add to 2 500 mL over 5 bottles, so both have a mean of exactly 500 mL — and in both sets the middle value is 500 mL too. Same mean, same median. Are the machines the same? No:

Machine A



Machine B



mL in the bottle (target 500 mL)

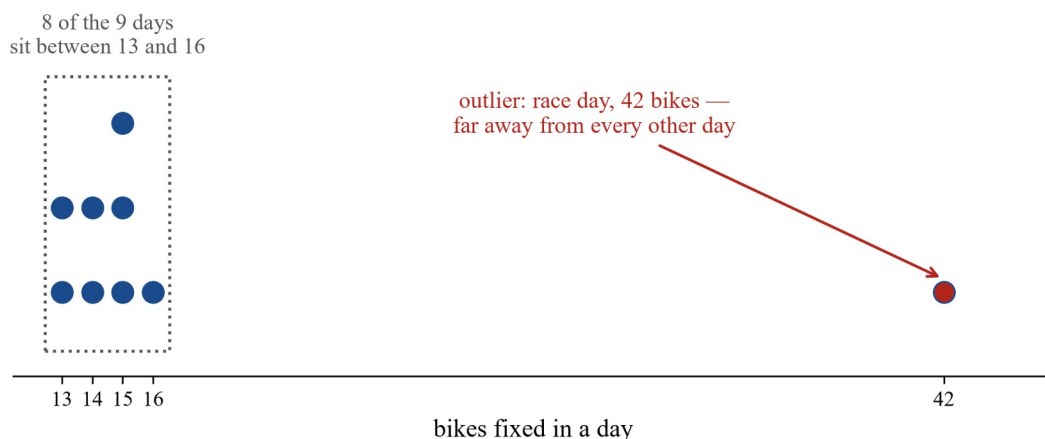
Made-up data: the machines and every volume in this picture are made up for this section.

Machine A's fills sit within 4 mL of each other (range $502 - 498 = 4$ mL), while Machine B's stretch over 24 mL (range $512 - 488 = 24$ mL). Machine A is **consistent** — it gives much the same answer every time. Machine B is not, and a customer could get a badly short bottle from it. The mean alone hid all of that. This is why a data set is best described by a centre **and** a spread together: the centre says where the data sits, the spread says how much you can rely on it.

Outliers.

An **outlier** is a value that sits far away from all the other values in a data set. There is no formula to find one: you order the data, or draw it, and look for a value with a big empty gap between it and everything else. A bike repair stand counts the bikes it fixes each day over nine days: 14, 15, 13, 16, 15, 14, 42, 13 and 15.

Bikes fixed each day at a repair stand (9 days)



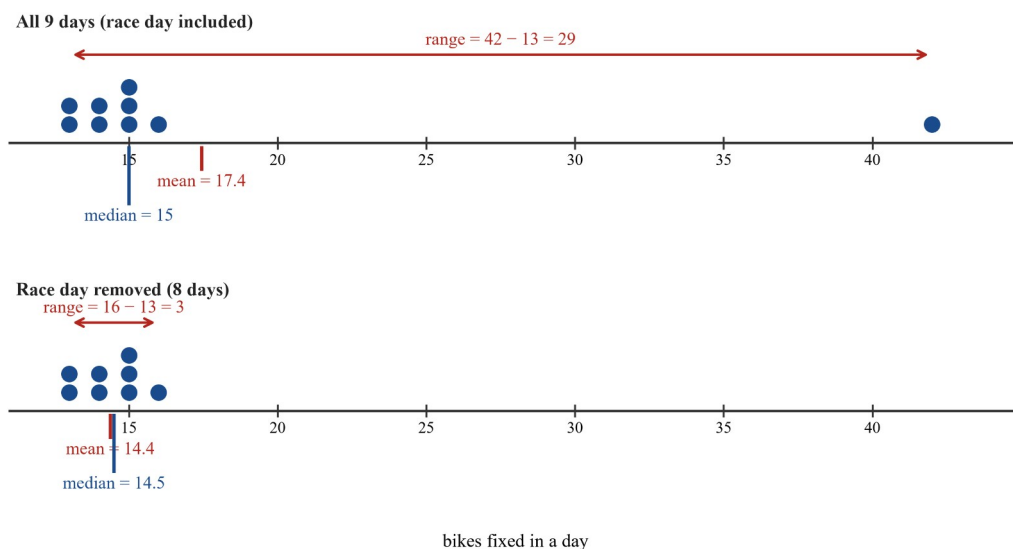
Made-up data: the repair stand and every count in this picture are made up for this section.

Eight of the nine days sit between 13 and 16 bikes. The 42 sits on its own with a huge gap — the next value below it is only 16 — because that day was a fun-run race day and half the town brought in punctures. The 42 is an outlier.

What an outlier does to the measures.

One far-out value does not treat every measure the same way. Compare the bike data with and without the race day:

One outlier: the mean and range move, the median barely does



Made-up data: the counts in this picture are made up for this section.

With all nine days the mean is $157 \div 9 = 17.444\dots$, so 17.4 bikes to one decimal place, and 8 of the 9 days fixed **fewer** bikes than that mean — the one race day dragged it up past nearly every ordinary day. Take the race day out and the mean falls to $115 \div 8 = 14.375$, so 14.4 bikes: a drop of about 3.1 bikes from one value. The median barely notices: it moves from 15 to 14.5, half a step. And the range collapses from $42 - 13 = 29$ to $16 - 13 = 3$, because the range is **decided** by the extreme values — an outlier does not just stretch the range, it *is* the range.

Why outliers happen, and what to do with one.

An outlier usually has one of two causes. It can be **real but unusual** — a race day, a festival Saturday, a heatwave — in which case it is a true part of the record and you keep it, but you report the median as well as the mean so the unusual day does not masquerade as a typical one. Or it can be a **mistake** — a keying error, a mis-read scale, a mis-scan — in which case you correct it if you can find the true value, or remove it if you cannot, and you say in your working exactly what you did and why. What you never do is delete a value quietly: a reader must be able to see which numbers were used.

Which measure tells you what.

Measure	What it uses	What one extreme value does to it
mean	every value, added into the sum	gets pulled a long way toward the extreme value
median	only the middle of the ordered list	barely moves — half a step at most
mode	only how often values repeat	usually does not change at all
range	only the highest and lowest values	is decided by it — the range stretches to reach it

So report a centre AND a spread together — and look for extreme values before you trust either.

Made-up data: this chart describes the methods in this section, not real data.

The mean uses every value, so it is the measure most worth having when nothing is far-out — and the one most worth doubting when something is. The median and the mode shrug off extreme values. The range is quick and useful, but it hangs entirely on the two end values, so check those two first: if one of them is an outlier, say so, and give the range with and without it.

Rounding and checking.

When an answer needs rounding, the question tells you how far — to one decimal place, to the nearest cent, to the nearest whole number. Keep the full value in your working and round only at the end, then say in your answer how far you rounded. Afterwards, sanity-check: a range can never be negative, and it is 0 only when every value in the set is the same. A centre should sit between the lowest and highest values; if your “centre” lies outside them, something has gone wrong.

Tools that help.

A calculator or a spreadsheet will find the highest and lowest of a long list in seconds — a spreadsheet has a built-in function for each — so at work you will often let the tool do the subtracting. But no tool can tell you *why* the 42 is there, whether it is a race day or a typing slip, or whether the mean or the median is the honest number for your report. Those judgements stay yours.

Worked examples

Example 1 — scaffolded

A market stall records the litres of fresh juice it sells each day over eight days: 22, 35, 28, 41, 30, 26, 38 and 32 litres.

- (a) Find the range of the juice sales.
- (b) Find the mean.
- (c) Find the median.

Working	Comment
In order: 22, 26, 28, 30, 32, 35, 38, 41.	Order first — the ends and the middle both show up.
Lowest = 22 L, highest = 41 L.	The range needs only these two values.
Range: $41 - 22 = 19$ L.	Highest minus lowest; keep the unit.
Sum: $22 + 35 + 28 + 41 + 30 + 26 + 38 + 32 = 252$ L.	Add every value, then divide by how many there are.
Mean: $252 \div 8 = 31.5$ L.	
Eight values, so the median is halfway between the 4th and 5th: $(30 + 32) \div 2 = 31$ L.	The middle two of an even-sized ordered list.
Range 19 L, mean 31.5 L, median 31 L.	Answer every part asked, with units.

Example 2 — scaffolded

A fish-and-chip shop counts the customers it serves each hour over 12 recorded hours:

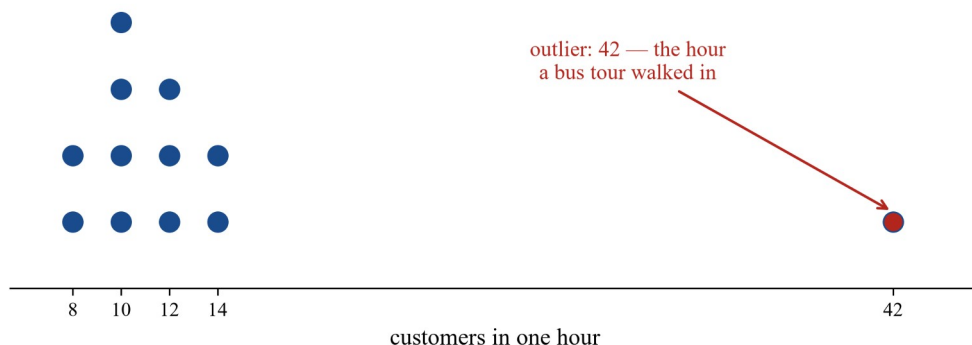
Customers
in one hour,

x	8	10	12	14	42
Hours, f	2	4	3	2	1
$f \times x$	16	40	36	28	42

Made-up data: the shop and its customer counts are made up for this section.

The same data is drawn below. The hour with 42 customers was the hour a bus tour arrived.

Customers per hour at a fish-and-chip shop (12 hours)



Made-up data: every number in this picture is made up for this section.

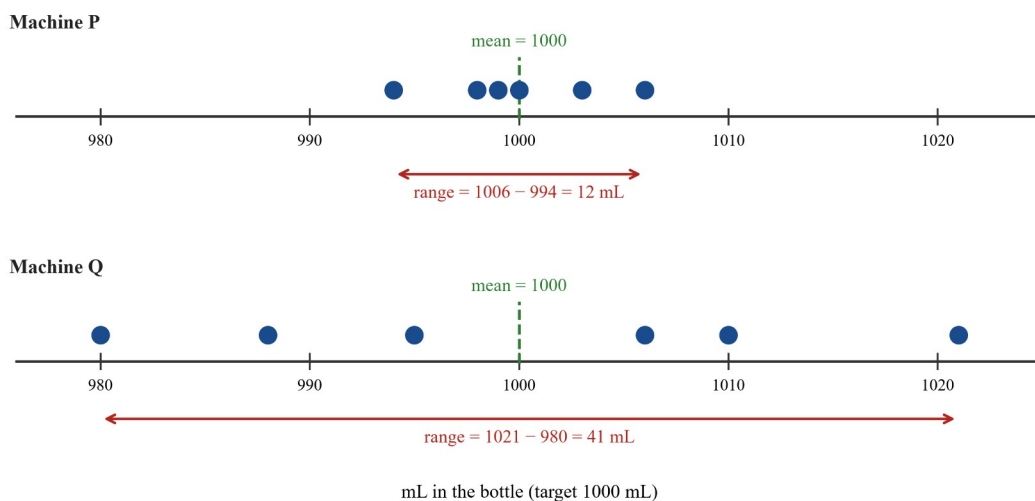
- Find the range of customers per hour.
- Find the mean and the median.
- Identify the outlier, then find the range and the mean (to one decimal place) of the data without it. What does that tell you about the range?

Working	Comment
Hours: $2 + 4 + 3 + 2 + 1 = 12$.	Add the frequency row to count the values.
Lowest = 8, highest = 42. Range: $42 - 8 = 34$ customers.	The two end values give the range at once.
Customers: $16 + 40 + 36 + 28 + 42 = 162$. Mean: $162 \div 12 = 13.5$ customers.	Add the $f \times x$ row instead of writing out 12 numbers.
Median: the 6th and 7th of the 12 values. Counting up, values 1–2 are 8, 3–6 are 10, 7–9 are 12 — so $(10 + 12) \div 2 = 11$ customers.	Half of 12 is 6: the middle is between the 6th and 7th values.
Outlier: 42 — the next value below it is only 14, so it sits far from the rest.	A big empty gap in the ordered data marks an outlier.
Without the 42: 11 hours, $162 - 42 = 120$ customers, mean $120 \div 11 = 10.909\dots$, so 10.9 customers to one decimal place. Range: $14 - 8 = 6$ customers.	Removing one value shrank the range from 34 to 6.
The range is decided by the ends: one unusual hour made it 34 when 11 of the 12 hours spread over only 6.	The mean fell a little; the range collapsed.

Example 3 — unscaffolded

A quality check measures six bottles from each of two filling machines, in mL. Machine P: 994, 1003, 998, 1006, 999, 1000. Machine Q: 980, 1021, 995, 1010, 988, 1006.

Two filling machines: equal means, unequal ranges



Made-up data: the machines and every volume in this picture are made up for this section.

- Find the mean and the range for each machine.
- Which machine is more consistent? Give a reason using your numbers.
- A buyer says, “Both machines average 1000 mL, so it makes no difference which one we buy.” Use your numbers to explain why the buyer is wrong.
- Machine P adds to $994 + 1003 + 998 + 1006 + 999 + 1000 = 6\,000$ mL over 6 bottles, so its mean is $6\,000 \div 6 = 1000$ mL; its range is $1006 - 994 = 12$ mL. Machine Q adds to $980 + 1021 + 995 + 1010 + 988 + 1006 = 6\,000$ mL over 6 bottles, so its mean is also $6\,000 \div 6 = 1000$ mL; its range is $1021 - 980 = 41$ mL.
- Machine P. The means are equal, so the centre cannot separate them, but P’s range of 12 mL is much smaller than Q’s 41 mL: P’s bottles stay close to 1000 mL every time, while Q’s wander as much as 20 mL below and 21 mL above it.
- The mean only says where the middle of each machine’s fills sits. P never misses 1000 mL by more than 6 mL, while Q can put out a 980 mL bottle or a 1021 mL bottle and still average 1000 mL. The ranges show the buyer that Q is the riskier machine, so the choice does make a difference.

$\therefore$ both machines have a mean of 1000 mL, but Machine P (range 12 mL) is more consistent than Machine Q (range 41 mL), and the means alone cannot show that.

Example 4 — unscaffolded

A clinic records the waiting times (in minutes) of eight patients: 18, 22, 15, 25, 20, 17, 21 and 162. The last entry turns out to be a keying error: the receptionist typed 162 when the true wait was 16 minutes.

- Find the mean, the median and the range of the data **as recorded**.
- Correct the 162 to 16 and find the mean, the median and the range again.
- Which measure moved the most, and which moved the least? Explain why, using what each measure adds into its working.
- As recorded the sum is $18 + 22 + 15 + 25 + 20 + 17 + 21 + 162 = 300$ minutes over 8 patients, so the mean is $300 \div 8 = 37.5$ minutes. In order the times are 15, 17, 18, 20, 21, 22, 25, 162; the middle two are 20 and 21, so the median is $(20 + 21) \div 2 = 20.5$ minutes. The range is $162 - 15 = 147$ minutes.

- (e) Corrected, the sum is $18 + 22 + 15 + 25 + 20 + 17 + 21 + 16 = 154$ minutes, so the mean is $154 \div 8 = 19.25$ minutes. In order the times are 15, 16, 17, 18, 20, 21, 22, 25; the middle two are 18 and 20, so the median is $(18 + 20) \div 2 = 19$ minutes. The range is $25 - 15 = 10$ minutes.
- (f) The range moved the most: $147 - 10 = 137$ minutes, because the wrong value *was* the highest value, and the range uses nothing but the two ends. The mean moved a long way too, $37.5 - 19.25 = 18.25$ minutes, because every value goes into the sum, so one huge value drags the total. The median moved least, $20.5 - 19 = 1.5$ minutes, because it uses only the middle of the ordered list: replacing the top value with 16 shifted the middle by one place at most.

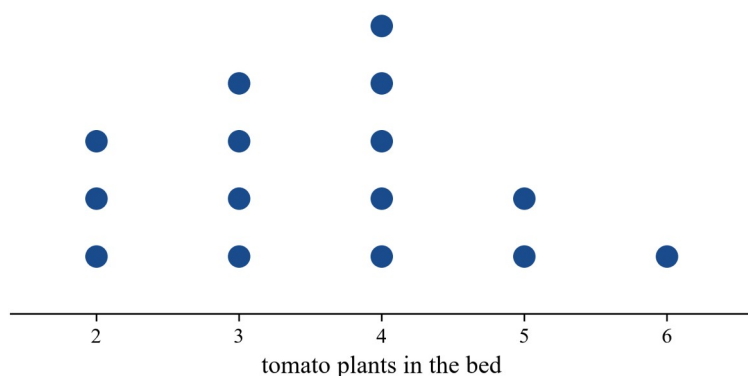
$\therefore$ as recorded the mean is 37.5 minutes, the median 20.5 minutes and the range 147 minutes; corrected they are 19.25 minutes, 19 minutes and 10 minutes — one keying error wrecked the range and dragged the mean, while the median barely moved.

Practice questions

Scaffolded

- Q1.** Find the highest value, the lowest value and the range of the data set 4, 9, 2, 11, 7. (a) Write the data in order, then state the lowest and highest values. (b) Find the range. (c) Find the median.
- Q2.** A golfer counts the putts taken in each of six rounds: 31, 35, 29, 33, 36 and 30 putts. (a) Write the data in order. (b) Find the range of the putts. (c) Find the mean number of putts, to one decimal place.
- Q3.** Garden beds on a shared plot are counted for tomato plants:

Tomato plants per garden bed (15 beds)



Made-up data: the beds and every count in this picture are made up for this section.

- (a) How many garden beds were counted?
 (b) Find the range of the plant counts.
 (c) Write down the mode.

- Q4.** Players in a football match are counted for goals kicked:

Goals, x	0	1	2	3	7
Players, f	4	6	5	2	1

Made-up data: the players and their goal counts are made up for this section.

- (a) How many players were counted?
 (b) Find the range of the goal counts.
 (c) Is the 7 an outlier? Give a reason using the gap between it and the next value below it.

Q5. The minimum temperatures ($^{\circ}\text{C}$) in one town over six nights are $-2, 3, 0, 5, -1$ and 4 . (a) Write down the highest and the lowest temperature. (b) Find the range of the temperatures. (c) Find the mean temperature.

Q6. A bus is late by these minutes over eight mornings: $0, 2, 1, 3, 2, 0, 4, 2$. (a) Find the range of the lateness. (b) Find the median lateness. (c) Find the mean lateness, to one decimal place.

Unscaffolded

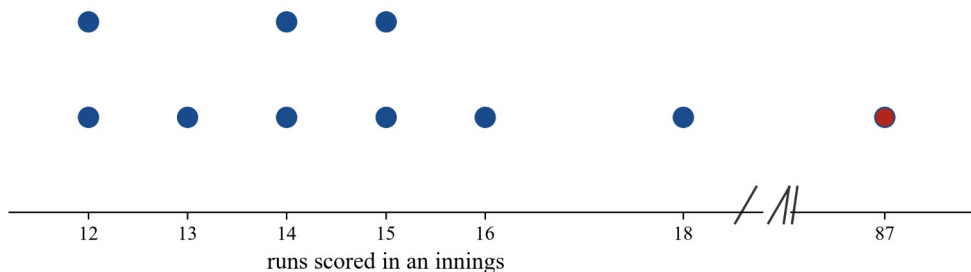
Q7. A runner records the kilometres run each day over ten days: $8, 10, 9, 12, 11, 10, 45, 9, 12$ and 10 km. The 45 km day was a trail race. (a) Find the mean (to one decimal place) and the median of the ten distances. (b) Identify the outlier and give a reason using the gap in the ordered data. (c) Find the range with the race day included and the range without it, and find the mean without it (to one decimal place). What do the three numbers tell you about the range?

Q8. A list of five parcel weights (in kg) is keyed into a spreadsheet as $24, 31, 93, 28, 35$. The 93 is a keying error: the true weight was 39 kg. (a) Find the range of the weights as keyed, and the range once the 93 is corrected to 39 . (b) Find the mean as keyed and the mean once corrected. (c) One measure of centre did not change at all. Show which one by finding it both ways, and explain why it survived the error.

Q9. Six batteries of each of two brands are tested for life in hours. Brand X: $44, 46, 45, 47, 43, 45$. Brand Y: $30, 58, 45, 38, 52, 47$. (a) Find the mean life and the range for each brand. (b) A torch user wants a battery that behaves much the same every time. Which brand should they choose? Give a reason using your numbers. (c) A shop sign says “Brand Y: same average life as Brand X, so the same value.” Use the ranges to say what the sign leaves out.

Q10. A batter’s scores over ten innings are drawn below.

One batter's scores over 10 innings



Made-up data: the batter and every score in this picture are made up for this section.

- Find the mean score (to one decimal place), the median and the range.
- Identify the outlier, then find the mean (to one decimal place), the median and the range of the nine innings without it.
- Which measure — the mean of all ten innings or the median — better describes a typical innings for this batter? Give a reason using your numbers from (a) and (b).

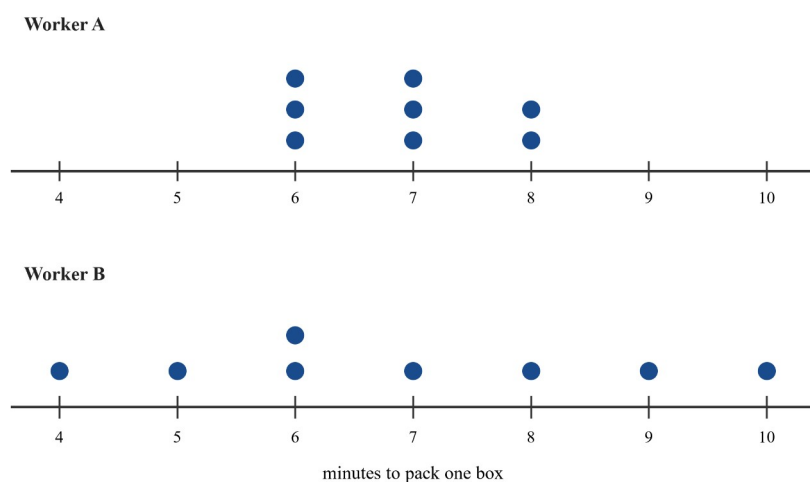
Q11. A smoothie bar records the flavours sold on Saturday: mango 31 , berry 24 , banana 18 , pineapple 12 . (a) How many smoothies were sold altogether? (b) Write down the mode. (c) Explain why a mean and a range cannot be found for this data, even though the counts are numbers.

Q12. A data set is $6, 9, 12$ and one unknown value k . The range of the set is 9 . (a) Show by calculation that $k = 15$ gives a range of 9 . (b) Find the only other possible value of k , and show by calculation that it too gives a range of 9 . (c) For each of your two values of k , write down the lowest and highest value of the set.

Q13. Over nine games a team’s scores have a lowest score of 18 and a highest score of 54 . (a) Find the range of the nine scores. (b) The team scores 61 in its tenth game. Find the new range, and give one tenth score that would have left the range at the answer to (a). (c) Find the two tenth scores that would make the new range exactly 50 .

Q14. Two workers are timed packing boxes, eight boxes each, in minutes.

Packing times: 8 boxes each for two workers



Made-up data: the workers and every time in this picture are made up for this section.

- Find the mean time (to one decimal place) and the range for each worker.
- Which worker is more consistent? Give a reason using your numbers.
- A manager says, “They both average the same, so either one can run the packing line.” Use the ranges to explain the risk of choosing Worker B.

Q15. A gym records how long members stay, in minutes, over 16 visits:

Minutes, x	20	30	40	50	110
Visits, f	3	6	4	2	1

Made-up data: the gym and its visit times are made up for this section.

- How many visits were recorded?
- Find the mean visit time (to one decimal place) and the median.
- Find the range, say whether the 110 is an outlier, and find the range of the data without it.

Q16. The maximum temperatures ($^{\circ}\text{C}$) in one coastal town over ten days are 18, 26, 22, 34, 19, 25, 21, 30, 12 and 23. (a) Find the mean and the range of the temperatures. (b) A tourist reads the mean and says, “It is about 23°C every day, so I only need summer clothes.” Use the range, and the number of days below 20°C , to say what is wrong with that claim. (c) How many of the ten days were within 5°C of the mean — that is, between 18°C and 28°C inclusive?

Q17. A farm stand’s takings over seven days are \$240, \$265, \$250, \$235, \$270, \$255 and \$890. The \$890 day was the regional field day, when the whole district came through. (a) Find the mean taking, correct to the nearest cent, and the median taking. (b) Is the \$890 an outlier? Should it be deleted from the record? Give a reason. (c) Find the range of the seven days and the range of the six days without the field day. Which single measure would you quote to describe a typical day, and why?

Q18. A rideshare driver’s fares one week are \$12, \$15, \$14, \$18, \$13, \$16, \$15, \$14 and \$108. The \$108 fare was an airport run. (a) Find the mean fare, the median fare and the range. (b) The driver advertises “my average fare is \$25”. How many of the nine fares were below \$25? Use that count to say whether the advertisement describes a typical fare. (c) Find the mean (to one decimal place) and the range of the eight fares without the airport run, and say which two numbers you would quote to describe a typical week of fares.

Answers

Q1. (a) in order 2, 4, 7, 9, 11; lowest 2, highest 11. (b) 9. (c) 7. **Q2.** (a) 29, 30, 31, 33, 35, 36. (b) 7 putts. (c) 32.3 putts (to one decimal place). **Q3.** (a) 15 beds. (b) 4 plants. (c) 4 plants. **Q4.** (a) 18 players. (b) 7 goals. (c) yes — the next value below 7 is only 3, so one player sits 4 goals clear of everyone else. **Q5.** (a) highest 5 °C, lowest -2 °C. (b) 7 °C. (c) 1.5 °C. **Q6.** (a) 4 minutes. (b) 2 minutes. (c) 1.8 minutes (to one decimal place). **Q7.** (a) mean 13.6 km, median 10 km. (b) the 45 — the next value below it is 12, a gap of 33 km. (c) range 37 km with the race day, 4 km without; mean without 10.1 km (to one decimal place) — one day stretched the range from 4 to 37. **Q8.** (a) 69 kg as keyed, 15 kg corrected. (b) 42.2 kg as keyed, 31.4 kg corrected. (c) the median: 31 kg both ways, because it uses only the middle of the ordered list and the error sat at the top. **Q9.** (a) X: mean 45 h, range 4 h; Y: mean 45 h, range 28 h. (b) Brand X — same mean, far smaller range. (c) the sign hides that a Brand Y battery can die at 30 h or run to 58 h: same average, much less predictable. **Q10.** (a) mean 21.6 runs (to one decimal place), median 14.5 runs, range 75 runs. (b) the 87; without it mean 14.3 runs (to one decimal place), median 14 runs, range 6 runs. (c) the median — the one 87 drags the mean above 9 of the 10 scores. **Q11.** (a) 85 smoothies. (b) mango. (c) flavours are words: there is nothing to add for a mean and no highest-minus-lowest for a range. **Q12.** (a) with $k=15$ the set is 6, 9, 12, 15: range $15-6=9$. (b) $k=3$: range $12-3=9$. (c) $k=15$: lowest 6, highest 15; $k=3$: lowest 3, highest 12. **Q13.** (a) 36. (b) 43; any tenth score from 18 to 54 (for example 40) leaves the range at 36. (c) 68 or 4. **Q14.** (a) A: mean 6.9 minutes (to one decimal place), range 2 minutes; B: mean 6.9 minutes, range 6 minutes. (b) Worker A — same mean, range 2 against 6. (c) B's range of 6 means a box can take 10 minutes when the average says 7: the line will stall on B's slow boxes. **Q15.** (a) 16 visits. (b) mean 38.1 minutes (to one decimal place), median 30 minutes. (c) range 90 minutes; yes, 110 is an outlier (next below is 50); range without it 30 minutes. **Q16.** (a) mean 23 °C, range 22 °C. (b) the range shows days from 12 °C to 34 °C, and 3 of the 10 days were below 20 °C, so the mean hides real cold days. (c) 7 days. **Q17.** (a) mean \$343.57 (to the nearest cent), median \$255. (b) yes, an outlier — but a real event, so keep it and quote the median for a typical day. (c) \$655 over seven days, \$35 over six; the median, because the field day drags the mean above 6 of the 7 days. **Q18.** (a) mean \$25, median \$15, range \$96. (b) 8 of the 9 — the airport run drags the mean above nearly every fare, so “average \$25” is not typical. (c) mean \$14.6 (to one decimal place), range \$6; quote the median \$15 and the range \$6 (or \$96 with the airport run, clearly labelled).

Fully-worked solutions

Q1. (a) In order: 2, 4, 7, 9, 11; the lowest value is 2 and the highest is 11. (b) The range is $11-2=9$. (c) With five values the middle one is the 3rd, so the median is 7.

Q2. (a) In order: 29, 30, 31, 33, 35, 36. (b) The range is $36-29=7$ putts. (c) The sum is $31+35+29+33+36+30=194$ putts over 6 rounds, so the mean is $194\div6=32.333\dots$, which is 32.3 putts to one decimal place.

Q3. (a) Read the dot columns: $3+4+5+2+1=15$ beds. (b) The lowest count is 2 plants and the highest is 6 plants, so the range is $6-2=4$ plants. (c) The tallest column is at 4 plants (5 beds), so the mode is 4 plants.

Q4. (a) Add the frequency row: $4+6+5+2+1=18$ players. (b) The lowest count is 0 goals and the highest is 7, so the range is $7-0=7$ goals. (c) Yes. In order the counts run 0, 0, 0, 0, 1, ..., 3, 3, 7: the next value below 7 is 3, so one player sits 4 goals clear of every other player — a big gap, which is what marks an outlier.

Q5. (a) The highest temperature is 5 °C and the lowest is -2 °C. (b) The range is $5-(-2)=7$ °C — mind the double negative: subtracting a negative adds. (c) The sum is $-2+3+0+5+(-1)+4=9$ over 6 nights, so the mean is $9\div6=1.5$ °C.

Q6. (a) In order: 0, 0, 1, 2, 2, 2, 3, 4; the range is $4-0=4$ minutes. (b) With eight values the median is halfway between the 4th and 5th, both 2, so the median is 2 minutes. (c) The sum is $0+2+1+3+2+0+4+2=14$ minutes over 8 mornings, so the mean is $14\div8=1.75$, which is 1.8 minutes to one decimal place.

Q7. (a) The sum is $8+10+9+12+11+10+45+9+12+10=136$ km over 10 days, so the mean is $136\div10=13.6$ km. In order: 8, 9, 9, 10, 10, 10, 11, 12, 12, 45; the middle two are 10 and 10, so the median is 10 km. (b) The outlier is the 45 km race day: the next value below it is 12, a gap of 33 km while every other gap is at most 1 km. (c)

With the race day the range is $45 - 8 = 37$ km. Without it the nine distances run from 8 to 12 km, so the range is $12 - 8 = 4$ km, and the mean is $(136 - 45) \div 9 = 91 \div 9 = 10.111\dots$, so 10.1 km to one decimal place. One unusual day stretched the range from 4 km to 37 km, which is what “the range is decided by its ends” means; the mean moved too (13.6 to 10.1), while the median stayed at 10 km either way.

Q8. (a) As keyed, the highest is 93 and the lowest 24, so the range is $93 - 24 = 69$ kg. Corrected, the highest is 39 and the lowest 24, so the range is $39 - 24 = 15$ kg. (b) As keyed the sum is $24 + 31 + 93 + 28 + 35 = 211$ kg over 5 parcels, mean $211 \div 5 = 42.2$ kg. Corrected the sum is $24 + 31 + 39 + 28 + 35 = 157$ kg, mean $157 \div 5 = 31.4$ kg. (c) The median. As keyed, in order: 24, 28, 31, 35, 93 — median 31 kg. Corrected, in order: 24, 28, 31, 35, 39 — median 31 kg. The error sat at the top of the ordered list, and the median uses only the middle value, so the middle never changed; the mean, which adds every value, jumped by 10.8 kg.

Q9. (a) Brand X: sum $44 + 46 + 45 + 47 + 43 + 45 = 270$ h over 6 batteries, mean $270 \div 6 = 45$ h; range $47 - 43 = 4$ h. Brand Y: sum $30 + 58 + 45 + 38 + 52 + 47 = 270$ h over 6 batteries, mean $270 \div 6 = 45$ h; range $58 - 30 = 28$ h. (b) Brand X. The means are equal, but X’s range of 4 h says every battery landed within 2 h of 45, while Y’s range of 28 h says a battery can die at 30 h — for a torch that must behave the same every time, X is the safer buy. (c) The sign is arithmetically true and still misleading: the same average of 45 h hides that Y’s batteries are spread over 28 h against X’s 4 h, so a Y buyer gets far less certainty for the same price.

Q10. (a) The sum is $12 + 15 + 14 + 18 + 13 + 16 + 14 + 87 + 15 + 12 = 216$ runs over 10 innings, mean $216 \div 10 = 21.6$ runs. In order: 12, 12, 13, 14, 14, 15, 15, 16, 18, 87; the middle two are 14 and 15, so the median is $(14 + 15) \div 2 = 14.5$ runs. The range is $87 - 12 = 75$ runs. (b) The outlier is the 87: the next score below it is 18, a gap of 69 runs when every other gap is at most 2. Without it the nine scores sum to $216 - 87 = 129$, mean $129 \div 9 = 14.333\dots$, so 14.3 runs to one decimal place; in order 12, 12, 13, 14, 14, 15, 15, 16, 18 the 5th of nine is 14, so the median is 14 runs; the range is $18 - 12 = 6$ runs. (c) The median. With the 87 in, the mean of 21.6 is above 9 of the 10 scores — one innings dragged it past almost the whole record — while the median of 14.5 (and the no-outlier mean of 14.3) sits in among the ordinary innings. The mean only becomes an honest summary of this record once that one big score is allowed for.

Q11. (a) $31 + 24 + 18 + 12 = 85$ smoothies. (b) Mango, with 31 sales — more than any other flavour. (c) The mean needs a sum of the values, and the values here are words: mango plus berry is nothing at all. The range needs a highest minus a lowest, and words have no size to subtract. Only the counts can be added — which is why the mode, a count of repeats, is the one measure of centre that works, and no measure of spread works at all.

Q12. (a) With $k = 15$ the set is 6, 9, 12, 15. The highest value is 15 and the lowest is 6. $15 - 6 = 9$, which is the range we had to show. (b) Try k as a new lowest instead, so 12 stays the highest. The lowest must then be $12 - 9 = 3$, so $k = 3$, and the set is 3, 6, 9, 12. $12 - 3 = 9$, which is again the required range. (No other value works: any k strictly between 3 and 15 leaves the range smaller than 9, and any k outside them makes it bigger than 9.) (c) For $k = 15$: lowest 6, highest 15. For $k = 3$: lowest 3, highest 12.

Q13. (a) The range is $54 - 18 = 36$. (b) A tenth score of 61 becomes the new highest, so the range is $61 - 18 = 43$. Any tenth score from 18 to 54 leaves the highest at 54 and the lowest at 18: for example a score of 40 keeps the range at $54 - 18 = 36$. (c) A range of 50 needs the ends 50 apart. If the tenth score is a new highest: $k - 18 = 50$, so $k = 68$. If it is a new lowest: $54 - k = 50$, so $k = 4$. The two scores are 68 and 4.

Q14. (a) Worker A: sum $6 + 7 + 6 + 8 + 7 + 6 + 7 + 8 = 55$ minutes over 8 boxes, mean $55 \div 8 = 6.875$, so 6.9 minutes to one decimal place; range $8 - 6 = 2$ minutes. Worker B: sum $4 + 9 + 7 + 6 + 10 + 5 + 8 + 6 = 55$ minutes over 8 boxes, mean $55 \div 8 = 6.875$, so 6.9 minutes to one decimal place; range $10 - 4 = 6$ minutes. (b) Worker A. The means are the same to one decimal place, but A’s range of 2 minutes says every box took 6, 7 or 8 minutes, while B’s range of 6 minutes runs from 4 to 10. (c) The manager’s average is true and still not the whole picture: on B’s line a box can take 10 minutes — more than twice B’s quickest 4 — so trucks and customers queue on B’s slow boxes even though the weekly average looks fine. The range is the number that shows that risk; A’s small range is the safer staffing choice.

Q15. (a) Add the frequency row: $3 + 6 + 4 + 2 + 1 = 16$ visits. (b) The $f \times x$ row is 60, 180, 160, 100, 110, adding to 610 minutes; the mean is $610 \div 16 = 38.125$, so 38.1 minutes to one decimal place. With 16 values the median is

halfway between the 8th and 9th: counting up, visits 1–3 are 20 and 4–9 are 30, so both middle values are 30 and the median is 30 minutes. (c) The range is $110 - 20 = 90$ minutes. Yes, the 110 is an outlier: the next value below it is 50, a gap of 60 minutes when every other gap is 10. Without it the range is $50 - 20 = 30$ minutes — one visit tripled the range, exactly what a measure built only on its two ends does.

Q16. (a) The sum is $18 + 26 + 22 + 34 + 19 + 25 + 21 + 30 + 12 + 23 = 230$ over 10 days, so the mean is $230 \div 10 = 23$ °C. The highest is 34 °C and the lowest 12 °C, so the range is $34 - 12 = 22$ °C. (b) The claim reads the mean as if every day matched it. The range of 22 °C says the days ran from a chilly 12 °C to a hot 34 °C, and 3 of the 10 days (18, 19 and 12 °C) were below 20 °C — jumper weather, not shorts weather. The mean describes the middle of the ten days, not each day. (c) Within 5 °C of 23 means between 18 and 28 inclusive: 18, 26, 22, 19, 25, 21, 23 — that is 7 of the 10 days. The other three (34, 30, 12 °C) are the days the tourist would feel.

Q17. (a) The sum is $240 + 265 + 250 + 235 + 270 + 255 + 890 = \$2\,405$ over 7 days, so the mean is $2\,405 \div 7 = 343.571428\dots$, which is \$343.57 correct to the nearest cent. In order: 235, 240, 250, 255, 265, 270, 890; the middle of seven is the 4th, so the median is \$255. (b) Yes, \$890 is an outlier — the next taking below it is \$270, a gap of \$620 when every other gap is \$10 or less — but it is a **real** event, the field day, not a mistake, so it stays in the record. The honest report keeps it and quotes the median of \$255 for a typical day, because 6 of the 7 days took less than the mean of \$343.57. (c) Over all seven days the range is $890 - 235 = \$655$; over the six ordinary days it is $270 - 235 = \$35$. The single best measure for a typical day is the median, \$255: the field day drags the mean above nearly every day and stretches the range to \$655, while the median sits in among the ordinary takings.

Q18. (a) The sum is $12 + 15 + 14 + 18 + 13 + 16 + 15 + 14 + 108 = \$225$ over 9 fares, so the mean is $225 \div 9 = \$25$. In order: 12, 13, 14, 14, 15, 15, 16, 18, 108; the middle of nine is the 5th, so the median is \$15. The range is $108 - 12 = \$96$. (b) The fares below \$25 are 12, 13, 14, 14, 15, 15, 16, 18 — that is 8 of the 9. So the advertised “average \$25” is true arithmetic and a poor description: eight of nine customers paid well under it, and only the single airport run lifts the mean that high. (c) Without the airport run the eight fares sum to $225 - 108 = \$117$, mean $117 \div 8 = 14.625$, so \$14.6 to one decimal place, and the range is $18 - 12 = \$6$. To describe a typical week I would quote the median, \$15, together with the range, \$6 for ordinary fares (or \$96 once airport runs are counted, clearly labelled) — the median for the typical fare and the range for how much fares can jump.

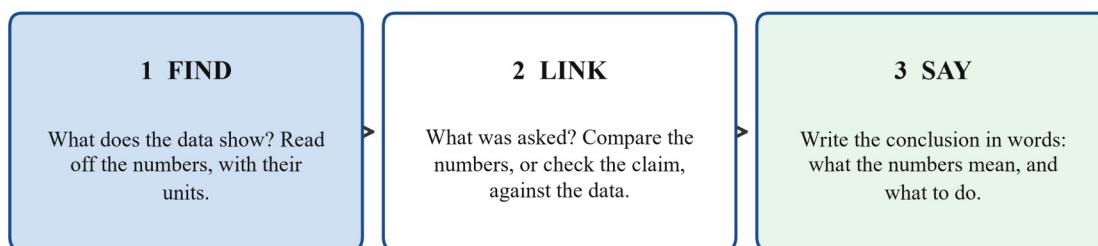
Theory

Sections 8A to 8D gave you ways to *show* data — tables, graphs, pictographs — and ways to squeeze a data set down to single numbers: the mean, the median, the mode, the range. This section is about what you do next. Someone hands you a graph, a table, or a claim like “most of our customers are happy”, and asks: what does it all mean? **Interpreting data** means reading what a display or a summary number actually says — describing the patterns, the variation and the trends in it — and then **drawing a conclusion**: a sentence, backed by a number, that answers the question asked.

About the data in this section. Every business, club, team, person and number in this section is **made up** (not real). They are built for this book so you can practise reading data and drawing conclusions from it. Nothing here is real data about a real person, business or place.

Three steps: find, link, say.

Nearly every conclusion in this section comes from the same three moves. **Find** what the data shows: read off the numbers you need, with their units. **Link** them to what was asked: compare two numbers, check a claim against a count, watch how a line moves. **Say** the conclusion in a plain sentence that answers the question.



A conclusion without a number is an opinion; a number without a sentence is not an answer.

Made-up data: this chart describes the steps in this section, not real data.

A conclusion without a number is an opinion. A number without a sentence is not an answer — it is just arithmetic. The three steps keep both halves together.

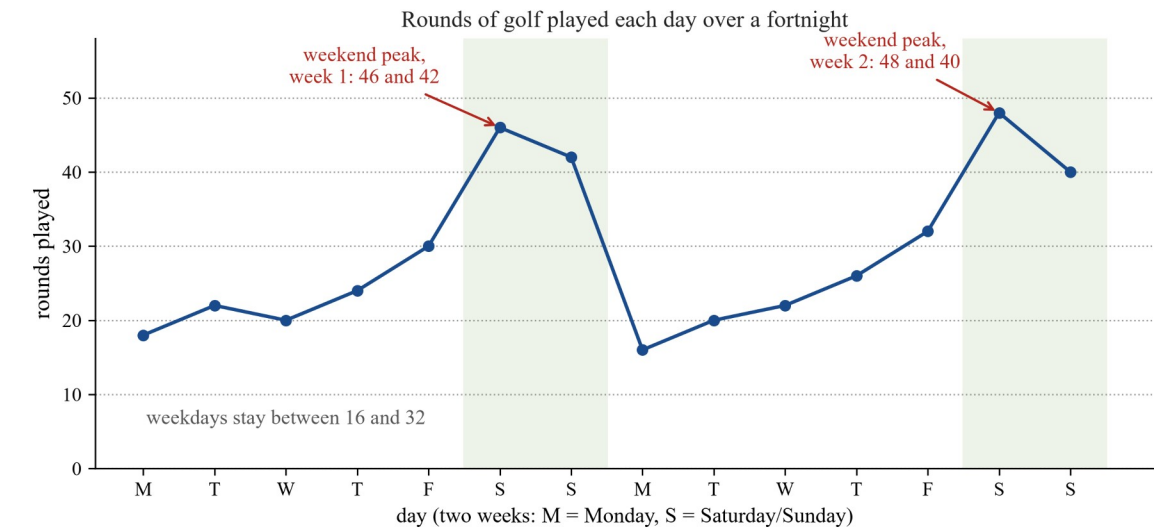
Patterns in a display.

A **pattern** is a shape that repeats, or holds right across the data, so you can describe it in one sentence. A golf course counts the rounds of golf played each day over a fortnight:

Day	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Week 1 rounds	18	22	20	24	30	46	42
Week 2 rounds	16	20	22	26	32	48	40

Made-up data: the golf course and its round counts are made up for this section.

The same data drawn as a line graph:



Made-up data: every number in this picture is made up for this section.

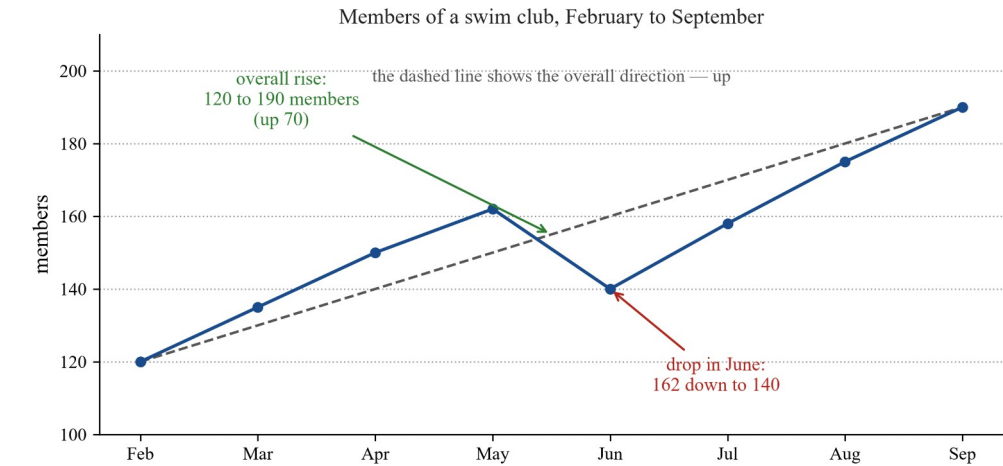
Read the shape first, then the numbers. The line climbs through each week, jumps on Saturday and Sunday, and falls back on Monday — and it does this in *both* weeks, which is what makes it a pattern rather than a coincidence. Weekdays sit between 18 and 30 rounds in week 1 and between 16 and 32 in week 2; weekends sit between 40 and 48. Even the quietest weekend day (40 rounds) beats the busiest weekday (32 rounds). The conclusion writes itself: the course’s busy days are Saturdays and Sundays, so weekend staff and tee times matter most.

Trends.

A **trend** is the overall direction of data that changes over time: up, down, or flat. A swim club records its membership at the start of each month from February to September:

Month	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep
Members	120	135	150	162	140	158	175	190

Made-up data: the swim club and its membership counts are made up for this section.



Made-up data: every number in this picture is made up for this section.

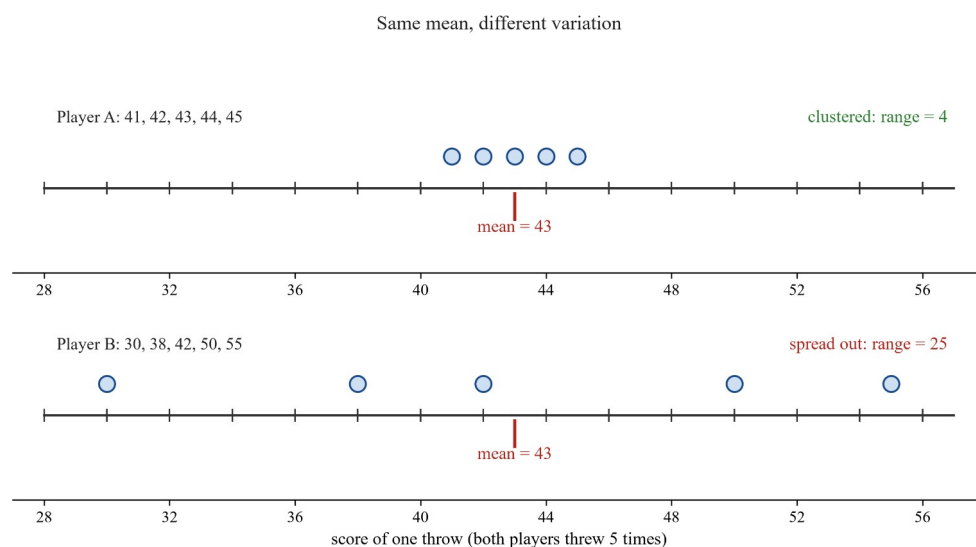
The trend is up: 120 members in February, 190 in September — an overall rise of 70 members. But “up” is not the whole story. The line dips in June, from 162 down to 140 members, a fall of 22, before climbing again. A good interpretation names both the overall trend *and* the dips and jumps around it. September is the highest month and February the lowest.

A trend describes the data you have, not the data you do not. It is fair to extend a steady pattern one step and call the result an **estimate** — Question 10 asks you to — as long as you label it an estimate rather than a fact. And data shows

what happened, not why: the June dip might be fees, weather, or a pool closure across the road, and the numbers alone cannot tell you which. A trend is a reason to go and ask, not proof of a cause.

Variation.

Variation is how spread out the data is — whether the values cluster tightly together or scatter widely. Two darts players throw five times each:



Made-up data: the two players and their scores are made up for this section.

Both players average 43 a throw, so the mean cannot separate them. The variation can. Player A's throws sit between 41 and 45 — a range of 4 — while Player B's run from 30 to 55, a range of 25. (The **range**, the biggest value minus the smallest, is the simple measure of spread from Section 8D.) Same mean, very different variation: A is steady, B is hit-or-miss. If your team needs a reliable 40-plus every turn, A is the pick — and “same mean” would have hidden that entirely.

When a summary number misleads.

A single summary number can hide the shape of the data behind it. The mean is the easiest to fool, because every value goes into its sum: one far-out value — Section 8D gives these the name **outliers** — drags the mean away from the rest of the set, while the median stays with the cluster. Example 4 puts a real-style job ad on trial over exactly this. Before you quote a mean, glance at the display: if one column towers over the others, say which measure you are using, and why.

Checking a claim.

Plenty of conclusions arrive pre-written: an advertisement, a report, a friend saying “most people prefer...”. Treat any claim as a conclusion someone else drew, and test it against the data yourself. Words like “most”, “average” and “more than half” each have a precise number hiding inside them — “most” means more than half, and “average” usually means the mean — so first work out what number the claim needs, then compute that number from the data and compare.

A conclusion of your own is worth writing when it ticks all three of these:

- it quotes at least one number from the data, with its units;
- it answers the question that was actually asked, not a nearby one;
- it says how far it was rounded, if it was rounded at all.

On rounding: keep the full value in your working, round once at the end, and repeat the rounding in your answer sentence — “to the nearest whole order”, “to one decimal place” — so a reader knows exactly what your number is. Questions 10, 15, 16 and 17 each ask for a stated rounding.

Worked examples

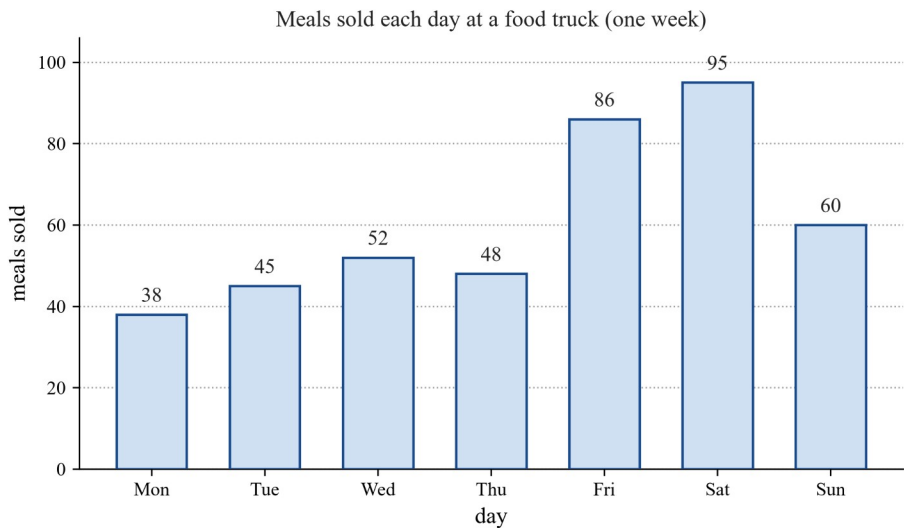
Example 1 — scaffolded

A food truck records the meals it sells each day over one week:

Day	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Meals sold	38	45	52	48	86	95	60

Made-up data: the food truck and its meal counts are made up for this section.

The same data drawn as a column chart:



Made-up data: every number in this picture is made up for this section.

Read the display and describe the week: which block of days sells most, and by how much?

Working	Comment
Read the seven column heights: 38, 45, 52, 48, 86, 95, 60 meals.	Check the chart against the table before using it.
Mon–Thu: $38 + 45 + 52 + 48 = 183$ meals.	Group the four weekdays first.
Fri–Sun: $86 + 95 + 60 = 241$ meals.	Then the three-day weekend block.
Difference: $241 - 183 = 58$ meals.	The question asks “by how much”, so subtract.
Busiest day: Saturday, 95 meals; quietest day: Monday, 38 meals.	Name the ends of the pattern as well.
Check: $183 + 241 = 424$ meals, which is all seven days added together.	A quick total catches reading slips.
Conclusion: the three days Fri–Sun sell 58 more meals than the four days Mon–Thu, so the truck’s busy block is the weekend.	A conclusion is a sentence about what the numbers mean.

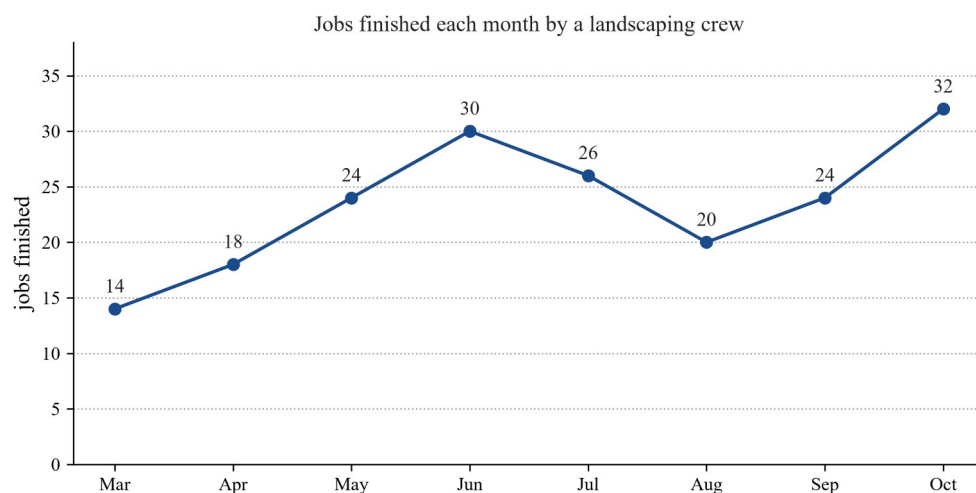
Example 2 — scaffolded

A landscaping crew counts the jobs it finishes each month from March to October:

Month	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct
Jobs finished	14	18	24	30	26	20	24	32

Made-up data: the landscaping crew and its job counts are made up for this section.

The same data drawn as a line graph:



Made-up data: every number in this picture is made up for this section.

Describe the crew's eight months: the total, the overall change, and the biggest month-to-month drop.

Working	Comment
Total: $14 + 18 + 24 + 30 + 26 + 20 + 24 + 32 = 188$ jobs.	Add all eight months.
Overall change: $32 - 14 = 18$ more jobs in October than in March.	Last value minus first value gives the net change.
Month-to-month changes: $+4, +6, +6, -4, -6, +4, +8$.	Subtract each month from the month before it.
Biggest drop: 6 jobs, from July (26) to August (20).	The largest negative change is the biggest drop.
Shape: climbs March to June (14 up to 30), dips June to August (30 down to 20), then climbs to 32 in October.	Describe the shape in words, with the numbers attached.
Conclusion: the crew finished 188 jobs and ended the period busier than it started — up 18 jobs — despite a mid-period drop of 6 jobs between July and August.	Tie the total, the trend and the dip into one sentence.

Example 3 — unscaffolded

A cricket team compares two batters over six innings each. Batter A scores 32, 41, 38, 45, 36 and 48 runs. Batter B scores 12, 78, 5, 88, 20 and 61 runs.

- Find the mean and the range of each batter's scores.
- The team needs one batter who will reliably score around 40 runs. Which batter would you pick? Justify your choice with the numbers from part (a).
- Batter A: the scores add to $32 + 41 + 38 + 45 + 36 + 48 = 240$ runs over 6 innings, so the mean is $240 \div 6 = 40$ runs. The range is $48 - 32 = 16$ runs. Batter B: the scores add to $12 + 78 + 5 + 88 + 20 + 61 = 264$ runs, so the mean is $264 \div 6 = 44$ runs. The range is $88 - 5 = 83$ runs.
- Batter A. B's mean is the higher one (44 against 40), but it is built from two big scores (78 and 88) and two very small ones (5 and 12): B's range of 83 runs says the scores scatter across almost the whole scale, so "44 on average" tells you little about any single innings. A's range of 16 runs says every innings landed between 32 and 48 — near the 40 the team needs, every time.

$\therefore$ Batter A: mean 40, range 16; Batter B: mean 44, range 83; pick Batter A for reliable scoring around 40 runs.

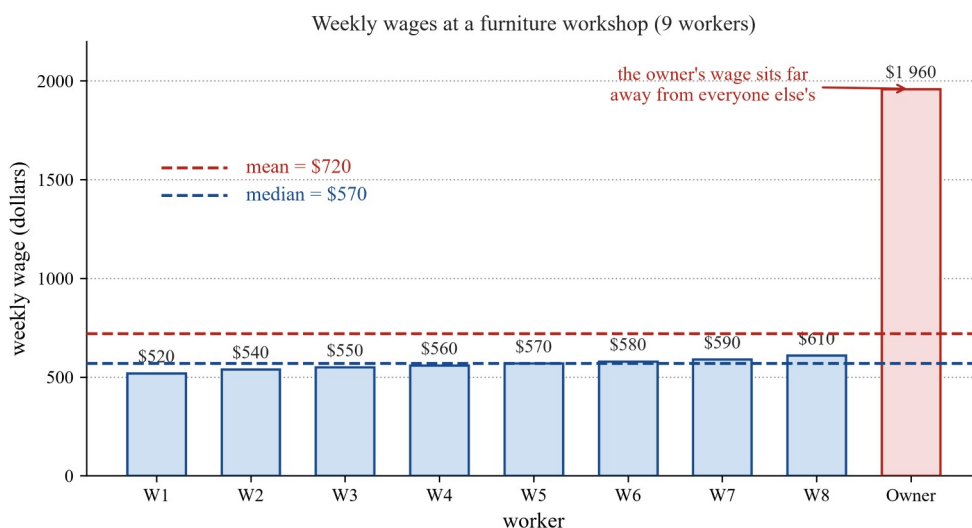
Example 4 — unscaffolded

A furniture workshop's job advertisement reads: "Great team! Average weekly wage \$720." The nine people at the workshop earn:

Worker	W1	W2	W3	W4	W5	W6	W7	W8	Owner
Weekly wage	\$ 520	\$ 540	\$ 550	\$ 560	\$ 570	\$ 580	\$ 590	\$ 610	\$ 1 960

Made-up data: the workshop and its wages are made up for this section.

The same data drawn as a column chart, with the mean and median marked:



Made-up data: every number in this picture is made up for this section.

- Show that the mean weekly wage is exactly \$ 720.
- Find the median weekly wage.
- Is the advertisement a fair picture of what a worker at this workshop earns? Answer with the numbers.
- The wages add to $520 + 540 + 550 + 560 + 570 + 580 + 590 + 610 + 1\,960 = \$6\,480$ over 9 workers, and $6\,480 \div 9 = \$720$ exactly. The advertisement's number is the mean, and the mean is correct.
- With 9 values the median is the 5th value in order. The table is already in order from smallest to largest, so the median is the 5th wage, \$ 570.
- Not really. Eight of the nine workers earn less than \$ 720 — every wage except the owner's \$ 1 960 sits between \$ 520 and \$ 610. The owner's far-out wage (an outlier, in Section 8D's language) drags the mean \$ 150 above the median, so "average \$ 720" is true arithmetic but a poor picture of the workshop. A new worker should expect something near the median of \$ 570, not the mean of \$ 720.

$\therefore$ the mean is \$ 720 and the median is \$ 570; the advertisement quotes the mean correctly, but the median shows a typical worker earns \$ 150 less than the advertisement suggests.

Practice questions

Scaffolded

Q1. A pool counts its admissions each day over one week:

Day	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Admissions	145	160	155	170	210	320	295

Made-up data: the pool and its admission counts are made up for this section.

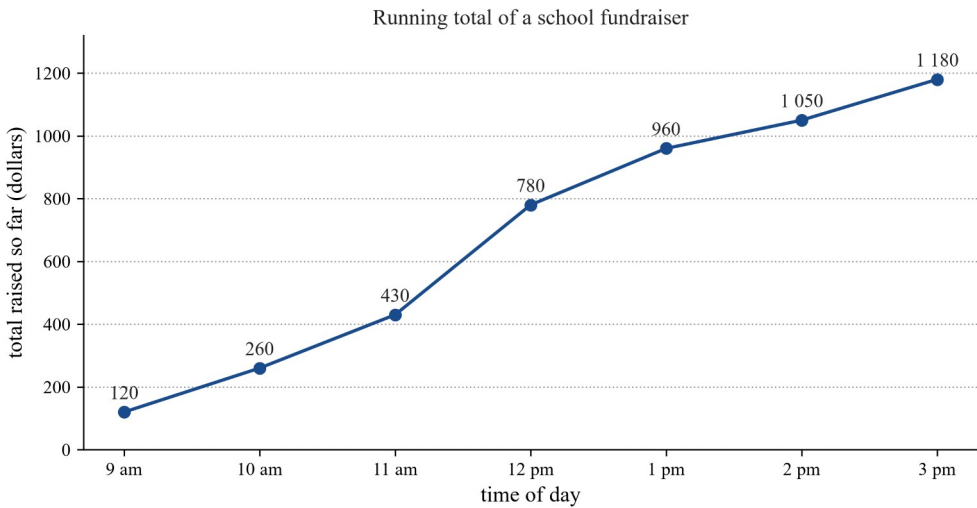
- (a) How many admissions were there on Friday?
- (b) How many more admissions were there on Saturday than on Wednesday?
- (c) Find the total admissions for the week.

Q2. A school fundraiser records its running total — the money raised so far — each hour from 9 am to 3 pm:

Time	9 am	10 am	11 am	12 pm	1 pm	2 pm	3 pm
Total raised so far	\$ 120	\$ 260	\$ 430	\$ 780	\$ 960	\$ 1 050	\$ 1 180

Made-up data: the fundraiser and its totals are made up for this section.

The same data drawn as a line graph:

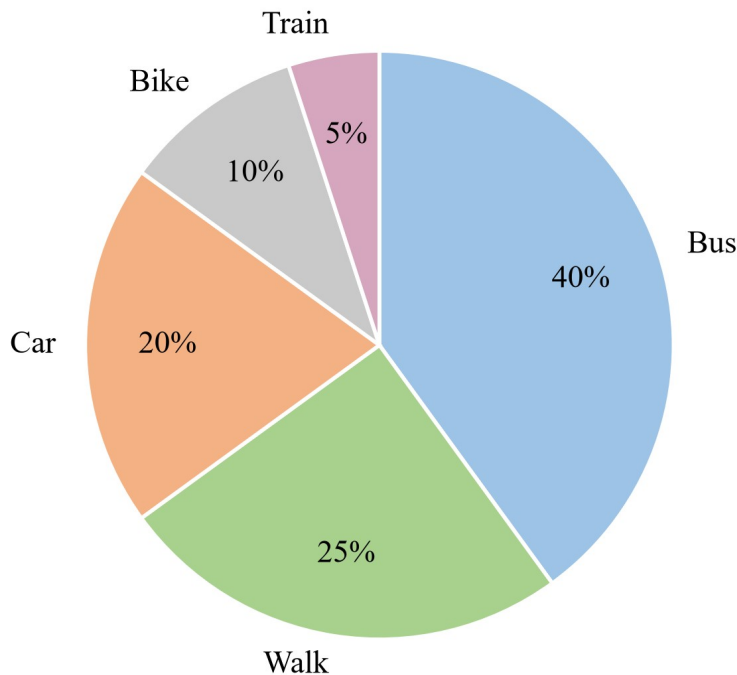


Made-up data: every number in this picture is made up for this section.

- (a) How much had been raised by 12 pm (noon)?
- (b) Between which two hours was the rise in the total the biggest, and how big was it?
- (c) How much had been raised altogether by 3 pm?

Q3. A school surveys how its students travel to school, as a share of all students:

How students travel to school (share of all students)



Made-up data: the shares in this picture are made up for this section.

- Write down the most common way of travelling, and its share.
- What share of students walk or ride a bike?
- A student writes: “More than half the students travel by bus.” Is this true or false? Give the numbers that settle it.

Q4. A school fete stall records the beanies it sells in one winter week:

Size	S	M	L	XL
Beanies sold	6	14	10	4

Made-up data: the stall and its sales are made up for this section.

- How many beanies were sold altogether?
- Write down the mode of the sizes sold.
- The stall orders more beanies for next winter. Which size should it order most of? Give a reason using the numbers.

Q5. A hardware store sums up one week of sales with four numbers: a mean of 46 sales a day, a median of 44, a mode of 40, and a range of 25. The quietest day had 32 sales.

- Find the number of sales on the busiest day.
- Which of the three measures of centre — mean, median or mode — sits highest?
- What does your answer to (b) tell you about the week’s busy days? Answer in a sentence.

Q6. A music school counts its students by level and by lesson day:

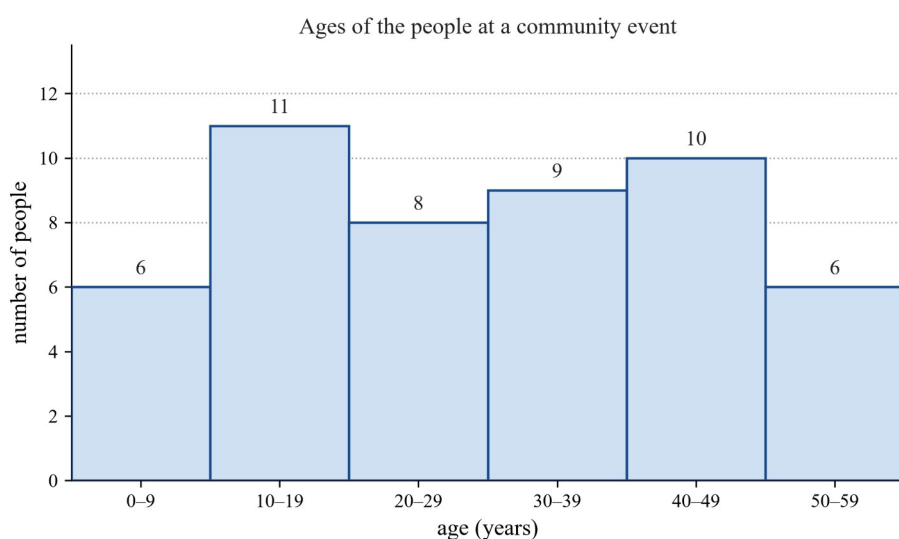
Level	Mon	Wed	Fri
Beginner	14	9	11
Intermediate	8	12	10
Advanced	4	6	9

Made-up data: the music school and its student counts are made up for this section.

- How many beginner students have lessons on Wednesday?
- Which lesson day is the busiest overall? Show the totals that decide it.
- Check the table by adding two ways: down the level totals and across the day totals. What should both additions give?

Un scaffolded

Q7. The ages of the people at a community event are drawn as a histogram:



Made-up data: every number in this picture is made up for this section.

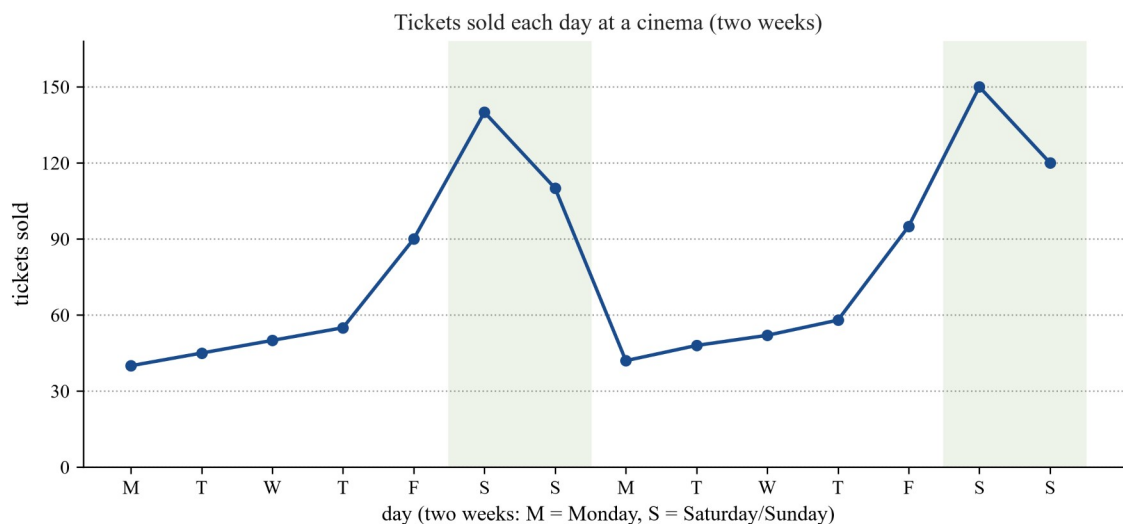
- How many people were at the event?
- Write down the modal age group (the tallest column) and how many people are in it.
- The organiser says: “Most of our people are teenagers (aged 10–19).” Do you agree? Justify your answer with the numbers.

Q8. A cinema counts the tickets it sells each day over two weeks:

Day	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Week 1 tickets	40	45	50	55	90	140	110
Week 2 tickets	42	48	52	58	95	150	120

Made-up data: the cinema and its ticket counts are made up for this section.

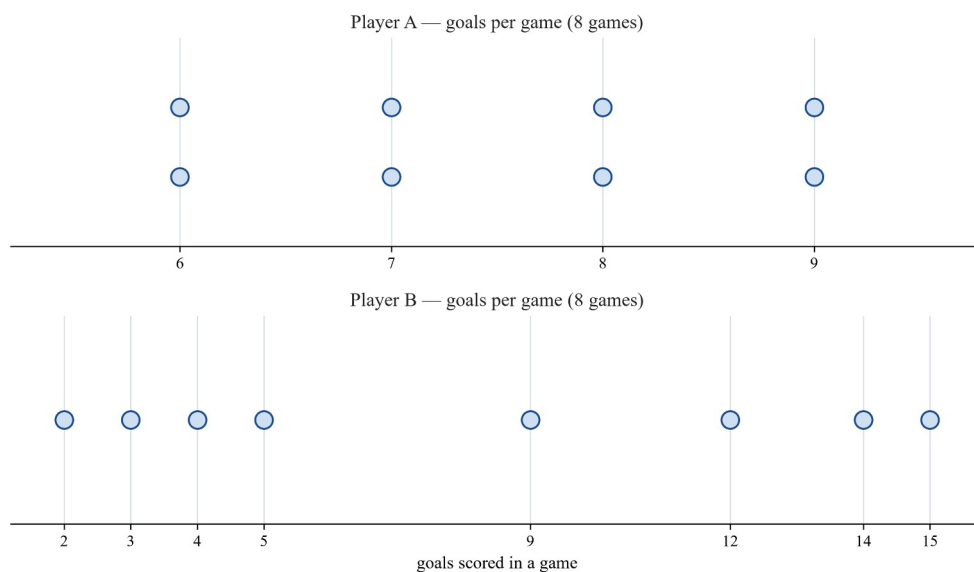
The same data drawn as a line graph:



Made-up data: every number in this picture is made up for this section.

- Find the total tickets for week 1 and for week 2, and the rise from week 1 to week 2.
- How many tickets were sold over the two weeks altogether?
- Describe the pattern the two weeks share, and write down the busiest single day.

Q9. Two netball players' goals per game over eight games are drawn as dot plots:



Made-up data: the players and their scores are made up for this section.

- Find the mean and the range of each player's goals.
- Find the median of each player's goals.
- The coach must pick one of the two for a close final and wants a dependable scorer. Which player should she pick? Justify your choice with the numbers.

Q10. An online store counts the orders it ships each month over six months: 42, 48, 55, 61, 68 and 74 orders.

- (a) Find the rise in orders from each month to the next, and describe what you notice.
- (b) Find the total rise from the first month to the last, and the average rise per month.
- (c) Estimate the next month's orders if the pattern continues, to the nearest whole order. Say in one sentence why your answer is an estimate and not a fact.

Q11. A market stall records how each dollar it takes is spent, by category:

Category	Fruit	Vegetables	Eggs	Honey	Flowers
Share of money taken	35 %	25 %	15 %	10 %	15 %

Made-up data: the stall and its shares are made up for this section.

- (a) What share of the money is spent on fruit and vegetables together? Is that more than half?
- (b) The stall's sign says: "9 out of every 10 dollars spent here are spent on food (fruit, vegetables, eggs and honey)." Check the claim against the table.
- (c) Which single category brings in the most money?

Q12. A hockey team's goals per game over ten games are 0, 1, 1, 2, 2, 2, 3, 3, 4 and 12 goals.

- (a) Find the mean, the median, the mode and the range of the goals.
- (b) In how many of the ten games did the team score 2 goals or fewer?
- (c) Which measure of centre best describes a typical game for this team? Give a reason using the numbers.

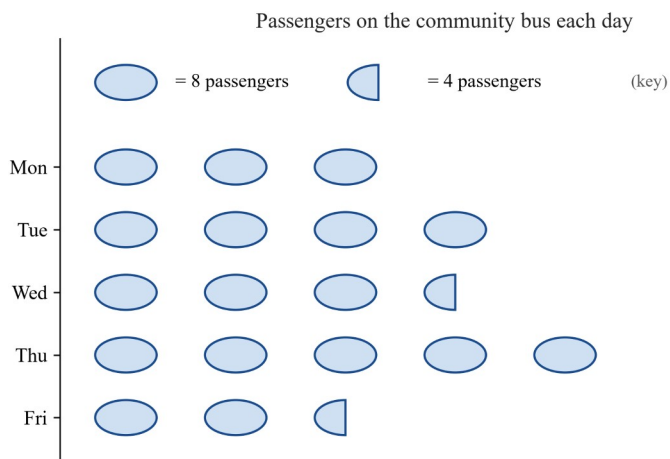
Q13. A council surveys 200 residents about which one project they want funded:

	Pool upgrade	Skate park	Community garden
Under 30	30	45	15
30 and over	50	20	40

Made-up data: the council survey and its counts are made up for this section.

- (a) Find the total votes for each project, and check that the three totals add to 200.
- (b) Write down the most-chosen project overall, and the most-chosen project among residents under 30.
- (c) The council's report says: "Most residents chose the pool upgrade." Check the claim against the table.

Q14. A community bus line draws its weekday passengers as a pictograph:



Made-up data: every symbol in this picture is made up for this section.

- How many passengers travelled on Thursday?
- How many passengers travelled on Wednesday?
- Find the total passengers for the five days.

Q15. A maker tests 60 batteries and records how long each one lasts:

Battery life (hours)	40–44	45–49	50–54	55–59	60–64	65–69
Batteries	4	9	18	16	9	4

Made-up data: the batteries and their lifetimes are made up for this section.

- Check the total number of batteries tested.
- The maker’s box says: “Most of our batteries last 55 hours or more.” Check the claim against the table.
- What percentage of the batteries lasted 50 hours or more? Give your answer to the nearest whole percent.

Q16. A lift’s sign reads: maximum 8 people or 600 kg. Eight people step in, weighing 68, 72, 85, 91, 64, 78, 80 and 70 kg.

- Find the total weight of the group, and their mean weight.
- Does the group break the 600 kg limit? Show the comparison that decides it.
- The heaviest person takes the stairs instead. Find the new mean weight of the remaining seven people, to one decimal place, and state whether the lift is then within the 600 kg limit.

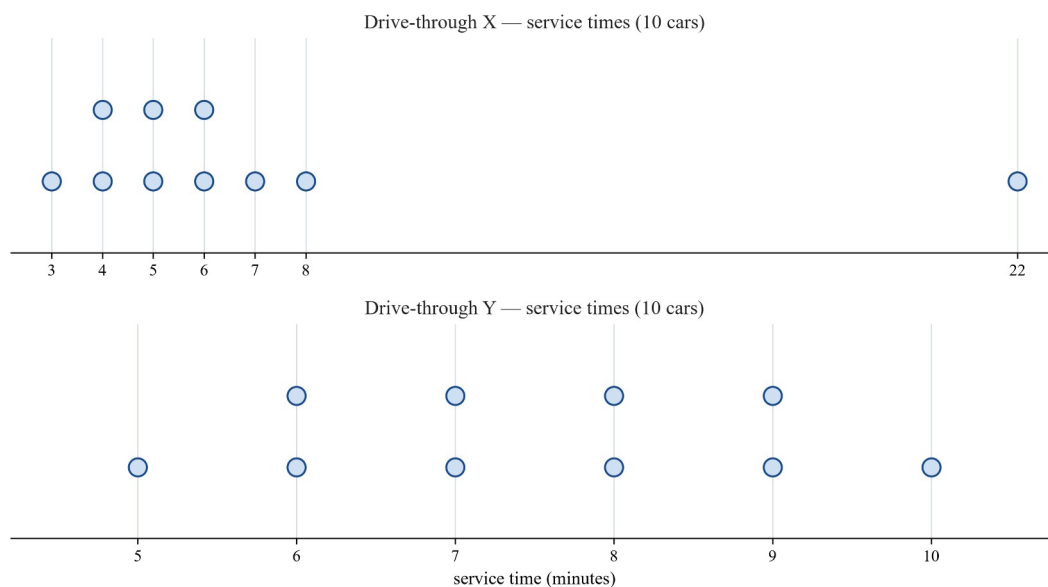
Q17. A household records the gas it uses each month from April to November:

Month	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov
Gas used (MJ)	240	320	410	460	430	350	280	220

Made-up data: the household and its gas readings are made up for this section.

- Find the total gas used over the eight months.
- Find the mean gas used per month, to the nearest 10 MJ.
- The household makes two claims. Claim 1: “Our gas use rose every month from April to July.” Claim 2: “The winter months June, July and August account for more than half of our gas over the eight months.” Check each claim against the table.

Q18. The service times of ten cars at each of two drive-throughs are drawn as dot plots:



Made-up data: the drive-throughs and their times are made up for this section.

- Find the mean, the median and the range of each drive-through's service times.
- Explain in a sentence why drive-through X's mean and median sit so far apart.
- Your lunch break allows ten minutes for the queue. Which drive-through do you choose? Justify your choice with the numbers.

Answers

Q1. (a) 210 admissions. (b) 165 more admissions. (c) 1 455 admissions. **Q2.** (a) \$ 780. (b) 11 am to 12 pm, a rise of \$ 350. (c) \$ 1 180. **Q3.** (a) bus, 40 %. (b) 35 %. (c) false — 40 % is less than the 50 % that “more than half” needs. **Q4.** (a) 34 beanies. (b) M. (c) M — it sold 14 times, more than any other size, so it is the size most likely to run out first. **Q5.** (a) 57 sales. (b) the mean, 46. (c) the mean sits above the median and the mode, so a few very busy days pull it up while most days sit nearer 40–44 sales. **Q6.** (a) 9 students. (b) Friday, with 30 students against 26 and 27. (c) 83 students, both ways. **Q7.** (a) 50 people. (b) 10–19 years, 11 people. (c) no — 11 of the 50 people are teenagers, exactly 22 %; the other 39 are not, so “most” is wrong. **Q8.** (a) week 1: 530; week 2: 565; a rise of 35 tickets. (b) 1 095 tickets. (c) both weeks climb steadily to a weekend peak and fall back on Monday, with week 2 above week 1 on every day; the busiest day is the second Saturday, 150 tickets. **Q9.** (a) A: mean 7.5, range 3; B: mean 8, range 13. (b) A: 7.5; B: 7. (c) Player A — the means are close, but A’s range of 3 says she scores near 7 or 8 every game, while B’s range of 13 hides games of 2 and 3 goals behind two big ones. **Q10.** (a) rises of 6, 7, 6, 7, 6 orders — a steady climb of about 6 or 7 orders a month. (b) total rise 32 orders; average rise 6.4 orders a month. (c) 80 orders (to the nearest whole order). **Q11.** (a) 60 %, yes — that is more than half. (b) false — food is 85 % of the money, which is 8.5 out of every 10 dollars, not 9. (c) fruit, 35 %. **Q12.** (a) mean 3, median 2, mode 2, range 12 goals. (b) 6 of the 10 games. (c) the median (or the mode), 2 goals — the single 12-goal game pulls the mean of 3 above six of the ten games. **Q13.** (a) pool upgrade 80, skate park 65, community garden 55; grand total 200. (b) overall: the pool upgrade, 80 votes; under 30: the skate park, 45 votes — exactly half of the 90 under-30s. (c) false — 80 of 200 is 40 %, and 120 residents chose something else. **Q14.** (a) 40 passengers. (b) 28 passengers. (c) 144 passengers. **Q15.** (a) 60 batteries. (b) false — 29 of the 60 batteries lasted 55 hours or more, and “most” needs more than half, so 31 or more. (c) 78 % (to the nearest whole percent). **Q16.** (a) 608 kg; mean 76 kg. (b) yes — 608 kg is over the 600 kg limit, by 8 kg. (c) 73.9 kg (to one decimal place); yes — the new total of 517 kg is within the limit. **Q17.** (a) 2 710 MJ. (b) 340 MJ (to the nearest 10 MJ). (c) Claim 1 true: $240 < 320 < 410 < 460$. Claim 2 false: June–August is 1 300 MJ of the 2 710 MJ, and half would be 1 355 MJ, so it falls short by 55 MJ. **Q18.** (a) X: mean 7, median 5.5, range 19 minutes; Y: mean 7.5, median 7.5, range 5 minutes. (b) the single 22-minute car drags X’s mean up while the median stays with the other nine cars, all served in 8 minutes or less. (c) Drive-through Y — every one of Y’s ten cars was served within 10 minutes, while X’s mean of 7 hides a 22-minute wait.

Fully-worked solutions

Q1. (a) Read Friday’s cell in the table: 210 admissions. (b) Saturday had 320 admissions and Wednesday 155, so Saturday had $320 - 155 = 165$ more admissions. (c) Add the week: $145 + 160 + 155 + 170 + 210 + 320 + 295 = 1\,455$ admissions.

Q2. (a) Read the 12 pm point on the line: \$ 780 raised so far. (b) The rises hour by hour are $260 - 120 = \$ 140$, $430 - 260 = \$ 170$, $780 - 430 = \$ 350$, $960 - 780 = \$ 180$, $1\,050 - 960 = \$ 90$ and $1\,180 - 1\,050 = \$ 130$. The biggest rise is \$ 350, between 11 am and 12 pm — the steepest part of the line. (c) Read the 3 pm point: \$ 1 180 raised altogether.

Q3. (a) The biggest sector is the bus, at 40 % of students. (b) Walk is 25 % and bike is 10 %, so walk or bike is $25 \% + 10 \% = 35 \%$ of students. (c) False. “More than half” means more than 50 %, and the bus sector is 40 % — below half. The other 60 % of students travel some other way.

Q4. (a) $6 + 14 + 10 + 4 = 34$ beanies. (b) Size M sold 14 times, more than any other size, so the mode is M. (c) Size M. It sold 14 of the 34 beanies — more than S and XL together ($6 + 4 = 10$) — so M is the size most likely to run out first next winter and the one to order most of.

Q5. (a) The range is the busiest day minus the quietest day, so the busiest day is $32 + 25 = 57$ sales. (b) The mean, at 46 — above the median of 44 and the mode of 40. (c) A mean pulled above the median and the mode is the signature of a few very big values: a couple of very busy days (up near the 57) drag the mean up, while most days of the week sit nearer the mode of 40 and the median of 44 sales.

Q6. (a) Read the Beginner row, Wednesday column: 9 students. (b) Add down each day: Monday $14 + 8 + 4 = 26$, Wednesday $9 + 12 + 6 = 27$, Friday $11 + 10 + 9 = 30$. Friday is the busiest day, with 30 students. (c) Add across each

level row: Beginner $14 + 9 + 11 = 34$, Intermediate $8 + 12 + 10 = 30$, Advanced $4 + 6 + 9 = 19$; those level totals add to $34 + 30 + 19 = 83$. The day totals from (b) add to $26 + 27 + 30 = 83$ as well. Both additions count every student exactly once, so both must give the same grand total — 83 students — and they do.

Q7. (a) Add the column heights: $6 + 11 + 8 + 9 + 10 + 6 = 50$ people. (b) The tallest column is the 10–19 group, with 11 people. (c) No. Teenagers are the 10–19 group, 11 of the 50 people, and 11 out of 50 is exactly 22% — well under half. The other 39 people are not teenagers, so “most of our people are teenagers” does not match the data.

Q8. (a) Week 1: $40 + 45 + 50 + 55 + 90 + 140 + 110 = 530$ tickets. Week 2: $42 + 48 + 52 + 58 + 95 + 150 + 120 = 565$ tickets. The rise is $565 - 530 = 35$ tickets. (b) Together: $530 + 565 = 1095$ tickets. (c) Both weeks show the same shape: a slow climb Monday to Friday, a jump to a weekend peak, and a fall back on Monday — and week 2 sits above week 1 on every single day. The busiest day is the second Saturday, with 150 tickets.

Q9. (a) Player A: the goals add to $6 + 8 + 7 + 9 + 7 + 8 + 6 + 9 = 60$ over 8 games, so the mean is $60 \div 8 = 7.5$ goals; the range is $9 - 6 = 3$ goals. Player B: the goals add to $2 + 14 + 5 + 12 + 3 + 15 + 4 + 9 = 64$, so the mean is $64 \div 8 = 8$ goals; the range is $15 - 2 = 13$ goals. (b) Player A in order: 6, 6, 7, 7, 8, 8, 9, 9; the two middle values are 7 and 8, so the median is $(7 + 8) \div 2 = 7.5$ goals. Player B in order: 2, 3, 4, 5, 9, 12, 14, 15; the two middle values are 5 and 9, so the median is $(5 + 9) \div 2 = 7$ goals. (c) Player A. The means are close (7.5 against 8), but A’s range of 3 goals says every game landed between 6 and 9 — a score the team can plan around. B’s range of 13 goals says her mean of 8 is built on two big games (14 and 15) papering over games of 2 and 3; in a close final, that variation is the risk.

Q10. (a) The rises are $48 - 42 = 6$, $55 - 48 = 7$, $61 - 55 = 6$, $68 - 61 = 7$ and $74 - 68 = 6$ orders: a steady climb of 6 or 7 orders every month, with no dips and no jumps. (b) The total rise is $74 - 42 = 32$ orders over 5 month-to-month steps, so the average rise is $32 \div 5 = 6.4$ orders a month. (c) Adding the average rise to the last month gives $74 + 6.4 = 80.4$, which is 80 orders to the nearest whole order. It is an estimate because it extends the pattern one month past the data — the store might get 78 or 85 — and the pattern itself might not hold.

Q11. (a) Fruit and vegetables together are $35\% + 25\% = 60\%$ of the money, and 60% is more than half (50%), so yes. (b) Food is fruit + vegetables + eggs + honey = $35\% + 25\% + 15\% + 10\% = 85\%$ of the money. 85% is 8.5 out of every 10 dollars, not 9 out of 10 (which would need 90%), so the sign overstates it — the claim is false. (c) Fruit, at 35% — the largest single share.

Q12. (a) The sum is $0 + 1 + 1 + 2 + 2 + 2 + 3 + 3 + 4 + 12 = 30$ goals over 10 games, so the mean is $30 \div 10 = 3$ goals. The data is already in order; the two middle values are both 2, so the median is 2 goals. The value 2 appears three times, more than any other, so the mode is 2 goals. The range is $12 - 0 = 12$ goals. (b) Counting the values 2 or less: 0, 1, 1, 2, 2, 2 — that is 6 of the 10 games. (c) The median (or the mode), 2 goals. Six of the ten games scored 2 or fewer, and the single 12-goal game drags the mean of 3 above six of the ten games; the median stays with the cluster of ordinary scores, so it is the honest picture of a typical game.

Q13. (a) Pool upgrade: $30 + 50 = 80$ votes. Skate park: $45 + 20 = 65$ votes. Community garden: $15 + 40 = 55$ votes. Check: $80 + 65 + 55 = 200$ residents ✓, which matches the survey size. (b) Overall the most-chosen project is the pool upgrade, with 80 votes. Among residents under 30 the most-chosen is the skate park, with 45 votes — exactly half of the 90 under-30 residents. The two age groups want different things, which is exactly what a two-way table can show and a single total cannot. (c) False. “Most residents” needs more than half of the 200, so more than 100 votes; the pool upgrade got 80, which is 80 out of 200 = 40%. The other 120 residents chose the skate park or the garden, so the pool upgrade was the most popular single project but not what most residents chose.

Q14. (a) Thursday shows 5 whole circles: $5 \times 8 = 40$ passengers. (b) Wednesday shows 3 whole circles and a half circle: $3.5 \times 8 = 28$ passengers. (c) Day by day: Monday $3 \times 8 = 24$, Tuesday $4 \times 8 = 32$, Wednesday 28, Thursday 40, Friday $2.5 \times 8 = 20$. The total is $24 + 32 + 28 + 40 + 20 = 144$ passengers. (Check: $3 + 4 + 3.5 + 5 + 2.5 = 18$ symbols, and $18 \times 8 = 144$ ✓.)

Q15. (a) Add the frequency row: $4 + 9 + 18 + 16 + 9 + 4 = 60$ batteries ✓. (b) Batteries lasting 55 hours or more are $16 + 9 + 4 = 29$. “Most” means more than half, and half of 60 is 30, so the claim needs 31 or more batteries; 29 is not enough, so the claim is false — in fact slightly fewer than half lasted 55 hours or more. (c) Batteries lasting 50 hours

or more are $18 + 16 + 9 + 4 = 47$. As a percentage, $47 \div 60 = 0.78333\dots$, which is $78.333\dots\%$, so 78% to the nearest whole percent.

Q16. (a) The weights add to $68 + 72 + 85 + 91 + 64 + 78 + 80 + 70 = 608$ kg, and the mean is $608 \div 8 = 76$ kg. (b) Yes: the limit is a total of 600 kg and the group totals 608 kg, which is 8 kg over. (Equivalently, $600 \div 8 = 75$ kg a head is the most the limit allows on average, and the group's mean of 76 kg is above it.) (c) Without the 91 kg person the total is $608 - 91 = 517$ kg over 7 people, so the new mean is $517 \div 7 = 73.857\dots$, which is 73.9 kg to one decimal place. The new total of 517 kg is under 600 kg, so the lift is then within the limit (and 7 people is within the 8-person limit too).

Q17. (a) The total is $240 + 320 + 410 + 460 + 430 + 350 + 280 + 220 = 2710$ MJ. (b) The mean is $2710 \div 8 = 338.75$ MJ, which is 340 MJ to the nearest 10 MJ. (c) Claim 1: April to July the readings go 240, 320, 410, 460 — each one bigger than the last — so the claim is true. Claim 2: June, July and August add to $410 + 460 + 430 = 1300$ MJ. Half of the eight-month total is $2710 \div 2 = 1355$ MJ, and 1300 MJ is 55 MJ short of that, so the winter months are *less* than half — the claim is false, though only just.

Q18. (a) Drive-through X: the times add to $3 + 4 + 4 + 5 + 5 + 6 + 6 + 7 + 8 + 22 = 70$ minutes over 10 cars, so the mean is $70 \div 10 = 7$ minutes. In order the two middle values are 5 and 6, so the median is $(5 + 6) \div 2 = 5.5$ minutes. The range is $22 - 3 = 19$ minutes. Drive-through Y: the times add to $5 + 6 + 6 + 7 + 7 + 8 + 8 + 9 + 9 + 10 = 75$ minutes, so the mean is $75 \div 10 = 7.5$ minutes. In order the two middle values are 7 and 8, so the median is $(7 + 8) \div 2 = 7.5$ minutes. The range is $10 - 5 = 5$ minutes. (b) Nine of X's ten cars were served in 8 minutes or less, so the median sits at 5.5; the single 22-minute car goes into the mean's sum in full and drags the mean up to 7, well above the middle of the cluster. (c) Drive-through Y. Every one of Y's ten cars — all of them — was served within 10 minutes, and Y's range of 5 minutes says its times stay bunched between 5 and 10. X's mean of 7 minutes looks quicker, but it hides a 22-minute wait: one car in ten missed a ten-minute break by a wide margin, and that is the risk a ten-minute break cannot afford.

Chapter 08 — Test A (multiple choice)

Chapter 08 (Data: collecting and categorising data; tables and graphs; mean, median and mode; range and outliers; interpreting data). Materials: one scientific calculator. Test A: 12 multiple-choice questions (12 marks). Approximate time: 20 minutes. Choose the response that is correct for each question.

About the data in this test. The shops, clubs, teams and numbers in this test are **made up**: pretend situations with pretend numbers, written for this book so that you can practise the mathematics. None of the data is real data about a real business, person or place.

Question 1

A florist records these details for every order it delivers: the type of flowers, the delivery distance in kilometres, the number of stems in the order, and the payment method (cash or card). Which detail is numerical continuous data?

- A. the number of stems in the order
- B. the type of flowers
- C. the delivery distance in kilometres
- D. the payment method (cash or card)

Question 2

A council surveys residents about the town's new library. What is the main problem with the survey question 'You love our new library, don't you?'

- A. It pushes people towards a particular answer.
- B. It is too long to read quickly.
- C. It cannot be answered with yes or no.
- D. It gives people too many choices.

Question 3

A bakery asked each of 40 customers which single item they had bought that day. Of these customers, 16 bought bread, 10 bought cake, 8 bought pies and 6 bought biscuits. What percentage of the 40 customers bought biscuits?

- A. 6 %
- B. 35 %
- C. 40 %
- D. 15 %

Question 4

A farmer wants to show how the height of a seedling changes over 8 weeks of growth. Which display is the best choice?

- A. a two-way table
- B. a line graph
- C. a pie chart
- D. a dot plot

Question 5

A pie chart shows the 60 books borrowed from a library in one week. Of these books, 15 were novels. The angle of the sector for novels is

- A. 25°
- B. 15°
- C. 270°
- D. 90°

Question 6

On a column chart of apples sold each day, the vertical scale has labels 0 and 20 with 4 grid squares between them. The top of Monday's column sits 3 grid squares above the 20 label. The number of apples sold on Monday is

- A. 23
- B. 35
- C. 26
- D. 60

Question 7

On a column chart, the vertical scale must start at 0. Why is this?

- A. Columns are compared by their heights, and heights are only honest when they are measured from zero.
- B. It makes the top of every column land exactly on a gridline.
- C. It means the chart needs fewer grid squares.
- D. It means the step size must always be 10.

Question 8

A netball team scored 11, 14, 9, 16, 12 and 10 goals over six games. The mean number of goals scored per game is

- A. 72
- B. 7.2
- C. 11.5
- D. 12

Question 9

Six workers at a small business each earn between \$580 and \$710 per week. The manager earns \$1920 per week. Which measure best describes the 'typical' weekly wage at this business?

- A. the mean, because it uses every wage
- B. the median, because it is not pulled around by the manager's far-out wage
- C. the range, because it shows how spread out the wages are
- D. the mode, because the mode is always the best measure of centre

Question 10

The temperatures ($^{\circ}\text{C}$) recorded at 6 am at a ski lodge over five days were -4 , 1 , 0 , 6 and -2 . The range of these temperatures is

- A. 2°C
- B. -10°C
- C. 13°C
- D. 10°C

Question 11

A museum counted the visitors to one small gallery on each of 9 days. The lowest daily count was 12 visitors, and the highest daily count apart from the exhibition day was 16 visitors. On the ninth day a special exhibition brought in 45 visitors. What is the range of all 9 counts, and the range when the 45 is set aside as an outlier?

- A. 33 and 4
- B. 4 and 33
- C. 33 and 16
- D. 45 and 16

Question 12

A grocery store recorded the main item bought by each of 80 customers: 30 bought fruit, 25 bought vegetables, 15 bought bread and 10 bought other items. The store claims, 'Most of our customers buy fruit.' Which of the following best judges this claim?

- A. The claim is true, because fruit was the most popular item.
- B. The claim is true, because more customers bought fruit than bread.
- C. The claim is false, because 30 out of 80 is less than half, even though fruit was the most popular item.
- D. The claim is false, because 30 out of 80 is more than half of the customers.

End of Test A

Test A — Answers

Question	1	2	3	4	5	6	7	8	9	10	11	12
Answer	C	A	D	B	D	B	A	D	B	D	A	C

Test A — Fully-worked solutions

Question	Answer	Working
1	C	The number of stems is counted in whole numbers (discrete), and the type of flowers and the payment method are words or labels (categorical). The delivery distance in kilometres is measured and can take any value in between, so it is numerical continuous.
2	A	‘You love our new library, don’t you?’ is a leading question: the tag ‘don’t you?’ pushes people towards the answer the surveyor wants, so it is not fair.
3	D	$\frac{6}{40} \times 100 = 15\%$ of the customers bought biscuits.
4	B	A line graph tracks a measured value (the height) as it changes over time, so it best shows how the seedling’s height changes over the 8 weeks.
5	D	Sector angle = $\frac{15}{60} \times 360^\circ = 90^\circ$.
6	B	One grid square is worth $20 \div 4 = 5$. The top of the column is $20 + 3 \times 5 = 35$ apples.
7	A	Columns are compared by their heights, and a height is only honest when it is measured from zero. A scale that starts higher up makes differences look bigger than they are.
8	D	Sum $= 11 + 14 + 9 + 16 + 12 + 10 = 72$; mean $= 72 \div 6 = 12$ goals.
9	B	The manager’s \$1 920 is a far-out value that drags the mean up above every worker’s wage. The median sits among the six workers’ wages and is not pulled around, so it best shows the ‘typical’ wage.
10	D	In order: $-4, -2, 0, 1, 6$. Range = highest – lowest $= 6 - (-4) = 6 + 4 = 10^\circ\text{C}$.
11	A	With all 9 counts: range $= 45 - 12 = 33$. With the 45 set aside as an outlier: range $= 16 - 12 = 4$.
12	C	‘Most’ means more than half. Half of 80 is 40, and only 30 customers

bought fruit, so the claim is false — even though fruit was the most popular single item.

Chapter 08 — Test B (written response)

Chapter 08 (Data). Materials: one scientific calculator. Test B: 8 written-response questions (40 marks). Approximate time: 40 minutes. Answer all questions in the spaces provided. Where a numerical answer is required, round only when the question instructs you to, and use the precision the question asks for. Where more than one mark is available, show your working.

About the data in this test. The clubs, farms, teams and numbers in this test are **made up**: pretend situations with pretend numbers, written for this book so that you can practise the mathematics. None of the data is real data about a real business, person or place.

Question 1 (5 marks)

A swimming club keeps three columns of information for every training session it runs:

- the name of the session — for example, squad training, learn-to-swim or water aerobics;
- the water temperature during the session, in °C;
- the number of swimmers who attend.

Made-up data: the swimming club, its sessions and its records are made up for this test.

a. Classify each of the club's three columns as categorical, numerical discrete or numerical continuous. 3 marks

b. The club recorded all of this information itself at the pool, session by session. Is this primary or secondary data? Give a reason for your answer. 2 marks

Question 2 (5 marks)

Fifteen students recorded their travel time to school one morning. Their times, in minutes and already in order, were

3, 6, 8, 11, 12, 14, 16, 18, 21, 23, 25, 27, 31, 33, 36

The students grouped the times into this frequency table.

Travel time, t (minutes)	Frequency
$0 \leq t < 10$	3
$10 \leq t < 20$	?
$20 \leq t < 30$	?
$30 \leq t < 40$	3
Total	15

Made-up data: the students and their travel times are made up for this test.

a. Find the two missing frequencies. 2 marks

b. How many students travelled for less than 20 minutes? 1 mark

c. One student travelled for exactly 20 minutes. Into which interval does this time go? 1 mark

d. Using your answer to **b**, find the percentage of students who travelled for less than 20 minutes. Give your answer to the nearest whole percent. 1 mark

Question 3 (5 marks)

A sports club asked each of its 36 members to name their favourite training activity. The results are shown in the table.

Favourite activity	Number of members
Gym	15
Swimming	9
Tennis	8
Cycling	4

The club displays the results in a pie chart.

Made-up data: the sports club and its members are made up for this test.

a. Explain why a pie chart is a good display for these data. 1 mark

b. Show that the angle of the gym sector is 150° . 2 marks

c. Find the angle of the cycling sector. 1 mark

d. Find the percentage of members whose favourite activity is swimming. 1 mark

Question 4 (5 marks)

An egg farm counted the eggs collected on each day of one week.

Day	1	2	3	4	5	6	7
Eggs collected	25	40	30	45	20	35	15

The farm draws a column chart of the data on grid paper. The vertical scale starts at 0 and uses a step of 5.

Made-up data: the egg farm and its daily counts are made up for this test.

a. On which day did the farm collect the most eggs, and how many eggs were collected on that day? 1 mark

b. What is the top label on the vertical scale? Explain why a step of 5 is a better choice than a step of 10 for these data. 2 marks

c. How many grid squares tall is the column for day 2? 1 mark

d. Find the total number of eggs collected over the week. 1 mark

Question 5 (5 marks)

A bookshop kept track of the number of bookmarks sold in each hour of a sale. The results are summarised in this frequency table.

Bookmarks

sold in one

hour, x

6

7

8

9

Number of

2

3

5

2

hours, f

Made-up data: the bookshop and its bookmark sales are made up for this test.

a. Find the number of hours the shop kept track of. 1 mark

b. Find the mean number of bookmarks sold per hour. Give your answer correct to 1 decimal place. 2 marks

c. Find the median number of bookmarks sold per hour. 1 mark

d. Find the mode. 1 mark

Question 6 (5 marks)

A netball shooter scored 18, 22, 16, 24 and 15 goals in her first five matches of the season. She wants a mean of exactly 21 goals per match over her first six matches.

Made-up data: the netball shooter and her scores are made up for this test.

- a.** Find the total number of goals scored over the first five matches. 1 mark

- b.** Find the total number of goals needed over the six matches for a mean of 21 goals per match. 1 mark

- c.** How many goals must she score in the sixth match? 1 mark

- d.** Check your answer to **c** by calculating the mean of all six scores. 2 marks

Question 7 (5 marks)

A community pool counted the admissions on each day of one week.

210, 240, 225, 260, 230, 220, 815

The day with 815 admissions was the day of a free family fun day at the pool.

Made-up data: the pool and its admission counts are made up for this test.

- a.** Find the range of the daily admissions. 1 mark

- b.** Find the mean and the median of the daily admissions. Give the mean correct to the nearest whole number. 2 marks

- c.** Explain why 815 is an outlier for these data. 1 mark

- d.** Would you quote the mean or the median to describe a typical day at the pool? Give a reason for your choice. 1 mark

Question 8 (5 marks)

A library counted the picture books borrowed in each of five months.

Month	1	2	3	4	5
Picture books borrowed	34	39	45	50	56

Made-up data: the library and its borrowing counts are made up for this test.

- a.** Find the rise from one month to the next for each pair of months, and describe the trend in the data. 2 marks

- b.** Find the average monthly rise. 1 mark

- c.** Use the average rise to estimate the number of picture books borrowed in month 6. Give your answer to the nearest whole book, and explain why your answer is an estimate. 1 mark

- d.** The library's newsletter claims that the number of picture books borrowed 'more than doubled' from month 1 to month 5. Is this claim true or false? Justify your answer with a calculation. 1 mark

End of Test B

Test B — Answers

Question 1. **a.** Session name: categorical; water temperature: numerical continuous; number of swimmers: numerical discrete. **b.** Primary — the club collected the information itself, first-hand.

Question 2. **a.** 5 and 4. **b.** 8 students. **c.** $20 \leq t < 30$ (the '20 to less than 30' interval). **d.** 53 %.

Question 3. **a.** It shows each activity as a share of the whole club. **b.** $\frac{15}{36} \times 360^\circ = 150^\circ$. **c.** 40° . **d.** 25 %.

Question 4. **a.** Day 4, with 45 eggs. **b.** Top label 45; with a step of 5 every count lands on a gridline, while with a step of 10 the 45 would not. **c.** 8 grid squares. **d.** 210 eggs.

Question 5. **a.** 12 hours. **b.** 7.6 bookmarks per hour (to 1 decimal place). **c.** 8 bookmarks. **d.** 8 bookmarks.

Question 6. **a.** 95 goals. **b.** 126 goals. **c.** 31 goals. **d.** Mean = $(95 + 31) \div 6 = 126 \div 6 = 21$ goals per match.

Question 7. **a.** 605 admissions. **b.** Mean 314 admissions (nearest whole number); median 230 admissions. **c.** The gap from 260 to 815 is 555, while every other gap is at most 20. **d.** The median — the mean is dragged above six of the seven days by the fun-day count.

Question 8. **a.** Rises of 5, 6, 5, 6; a steady climb. **b.** 5.5 books per month. **c.** About 62 books (nearest whole book); it assumes the trend continues. **d.** False — $2 \times 34 = 68$, and $56 < 68$.

Test B — Fully-worked solutions

Question 1.

a. The name of the session is a word or label, so it is **categorical**. The water temperature is measured and can take any value in between, so it is **numerical continuous**. The number of swimmers is counted in whole numbers, so it is **numerical discrete**.

b. **Primary** data. The club collected the information itself, first-hand, at the pool — it did not reuse data collected by someone else.

Question 2.

a. For $10 \leq t < 20$ the times are 11, 12, 14, 16 and 18, so the frequency is 5. For $20 \leq t < 30$ the times are 21, 23, 25 and 27, so the frequency is 4. Check: $3 + 5 + 4 + 3 = 15$. ✓

b. Less than 20 minutes covers the first two intervals: $3 + 5 = 8$ students.

c. A boundary value moves up into the higher interval, so exactly 20 minutes goes into $20 \leq t < 30$ — the ‘20 to less than 30’ interval.

d. $\frac{8}{15} \times 100 = 53.33\ldots\%$, which is 53 % to the nearest whole percent.

Question 3.

a. A pie chart shows each activity as a sector of one circle, so it displays each activity as a share of the whole club.

b. The gym sector covers 15 of the 36 members.

$$\text{angle} = \frac{15}{36} \times 360^\circ$$

$$\text{angle} = 150^\circ$$

So the gym sector has an angle of 150° , as required.

c. Cycling: $\frac{4}{36} \times 360^\circ = 40^\circ$.

d. Swimming: $\frac{9}{36} \times 100 = 25\%$. This is exact, so no rounding is needed.

Question 4.

a. The largest count is 45, collected on day 4.

b. With a step of 5 the scale runs 0, 5, 10, 15, ..., 45, so the top label is 45 — exactly 9 grid squares. Every count in the table is a multiple of 5, so with a step of 5 every column top lands exactly on a gridline and the chart can be drawn and read accurately. With a step of 10 the top label would be 50 (only 5 grid squares), and the day-4 column (45) would finish halfway between two gridlines.

c. Day 2 = 40 eggs, and $40 \div 5 = 8$, so the column is 8 grid squares tall.

d. $25 + 40 + 30 + 45 + 20 + 35 + 15 = 210$ eggs.

Question 5.

- a. The frequencies add to $2 + 3 + 5 + 2 = 12$, so the shop kept track of 12 hours.
- b. Add an $f \times x$ row: $2 \times 6 = 12$, $3 \times 7 = 21$, $5 \times 8 = 40$ and $2 \times 9 = 18$. The $f \times x$ row adds to $12 + 21 + 40 + 18 = 91$ bookmarks.

$$\text{mean} = 91 \div 12 = 7.5833\dots$$

which is 7.6 bookmarks per hour, correct to 1 decimal place.

c. 12 is even, so two values share the middle and the median is halfway between them: the 6th and 7th values in order. Count up through the frequencies: the 1st and 2nd values are 6, the 3rd to 5th are 7, and the 6th to 10th are 8. The 6th and 7th values are both 8, so the median is $(8 + 8) \div 2 = 8$ bookmarks.

d. The mode is the value with the biggest frequency: 8 bookmarks, with 5 hours.

Question 6.

- a. $18 + 22 + 16 + 24 + 15 = 95$ goals.
- b. A mean of 21 over 6 matches needs a total of $21 \times 6 = 126$ goals.
- c. The sixth match must supply the gap: $126 - 95 = 31$ goals.
- d. Check with 31 goals in the sixth match: the total is $95 + 31 = 126$ goals, and the mean is $126 \div 6 = 21$ goals per match — exactly the mean she wanted.

Question 7.

- a. In order: 210, 220, 225, 230, 240, 260, 815. Range = $815 - 210 = 605$ admissions.
- b. The sum is $210 + 240 + 225 + 260 + 230 + 220 + 815 = 2200$, so

$$\text{mean} = \frac{2200}{7} = 314.28\dots$$

which is 314 admissions, correct to the nearest whole number. There are 7 values, so the median is the 4th value in order: 230 admissions.

c. In the ordered list, the gap from 260 up to 815 is 555, while every other gap is at most 20. That big empty gap marks 815 as an outlier: it sits a long way from all of the other values.

d. Quote the **median**, 230 admissions. The fun day drags the mean up to 314, which is above the count on six of the seven days, so the median — which the far-out value cannot pull — gives the better picture of a typical day.

Question 8.

a. The rises are $39 - 34 = 5$, $45 - 39 = 6$, $50 - 45 = 5$ and $56 - 50 = 6$. The number of picture books borrowed climbs steadily, by 5 or 6 books each month.

b. Average rise = $\frac{5 + 6 + 5 + 6}{4} = \frac{22}{4} = 5.5$ books per month.

c. The estimate for month 6 is $56 + 5.5 = 61.5$, which is about 62 picture books to the nearest whole book. It is an estimate because the trend is assumed to continue in the same way — the actual month-6 count has not been recorded.

d. Double month 1: $2 \times 34 = 68$. Month 5 is 56, and $56 < 68$, so the count did not quite double — the claim is **false**.