Foundation Mathematics Units 1 & 2

Chapter 6 Summarising and comparing data

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

Theory

Most real questions about data are **comparisons**. Is Year 11 or Year 12 fitter? Which of two bus routes is more reliable? Did a new roster make service times more consistent? To answer them we put two (or more) data sets side by side and look for differences. This section is about **why** and **how** we compare data sets fairly; the following sections (6D–6F) show how to measure the differences exactly.

Make the comparison fair.

Before comparing, check that you are comparing *like with like*:

- **same variable** — measure the same thing in each group (both in minutes, both counts of the same event);
- **same units** — convert first if one set is in metres and the other in centimetres;
- **comparable groups** — if the groups are different sizes, compare **proportions, percentages or rates** rather than raw totals. A shop open twice as long can serve more customers *in total* even if it is quieter *per hour*.

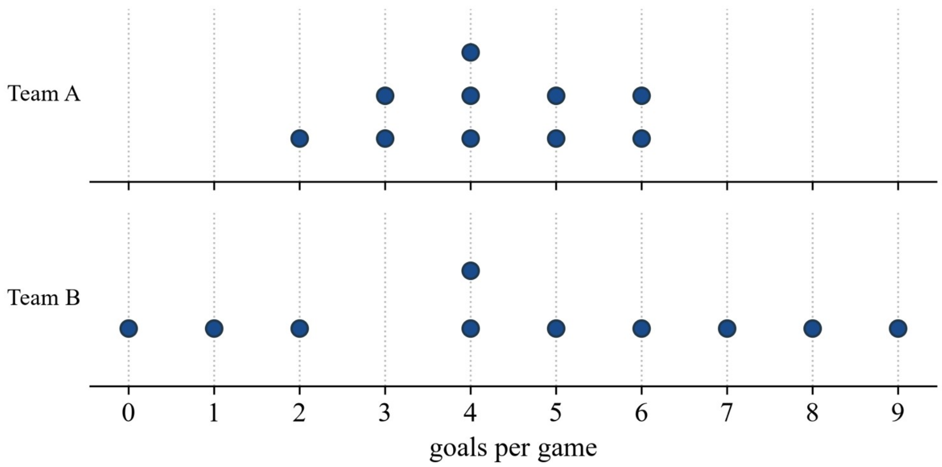
Comparing raw totals from groups of different sizes is the most common way a comparison is made misleading.

Displays that compare.

Some displays are built to show two data sets together:

- **Parallel (side-by-side) dot plots** — one dot plot per group, drawn on the *same* horizontal scale so you can line the groups up.
- **Back-to-back stem-and-leaf plots** — the two groups share a single central “stem”, with the leaves of one group written to the left and the other to the right.
- **Side-by-side (grouped) column graphs** — for categorical data, the two groups’ bars are drawn next to each other in each category.
- **Parallel boxplots** — a compact five-number picture of each group on a shared scale (developed in 6F).

The parallel dot plot below compares the goals scored per game by two teams over ten games. Both teams are centred near the same place, but Team B’s dots are far more spread out.



What to look for.

When you compare two numerical data sets, describe three things:

- **Centre** — is one group typically higher? Compare the **median** (or the mean, or the mode).
- **Spread** — is one group more variable? Compare the **range** (and, from 6F, the interquartile range).
- **Shape** — is the data symmetric or skewed? Are there **clusters, gaps or unusual values (outliers)**?

Then write the comparison in words, always mentioning the **context** and the **units** — for example, “Team A typically scored about the same as Team B (both medians near 4 goals), but Team A was far more consistent (range 4 goals versus 9 goals).”

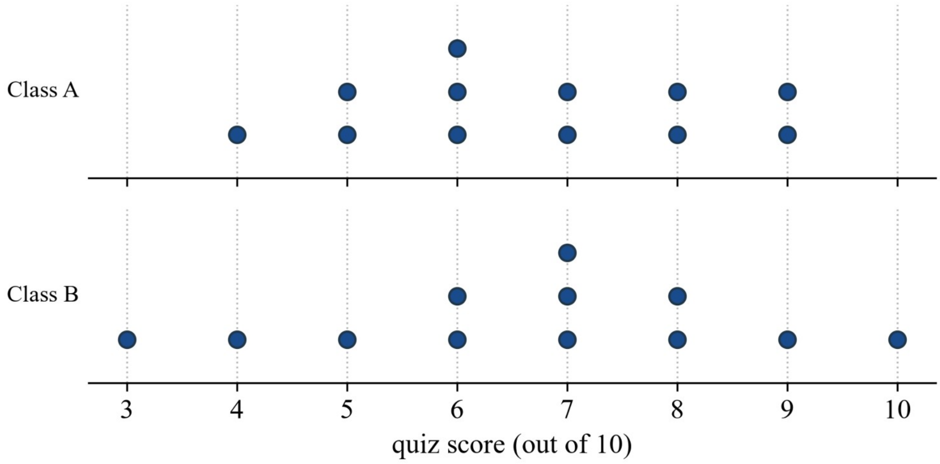
Watch for outliers.

A single unusual value can make one group *look* far more spread out. The **range** is very sensitive to it. When comparing spread, it is worth asking what the range would be *without* the unusual value, so the comparison reflects the **typical** members of each group.

Worked examples

Example 1 — scaffolded

Two Year 11 classes sat the same numeracy quiz, marked out of 10. Their results are shown in the parallel dot plot below. Compare the two classes.



Working	Comment
Class A scores: $4, 5, 5, 6, 6, 6, 7, 7, 8, 8, 9, 9$ ($n = 12$). Median = $\dfrac{6+7}{2} = 6.5$.	For $n = 12$ (even), average the 6th and 7th values.
Class B scores: $3, 4, 5, 6, 6, 7, 7, 7, 8, 8, 9, 10$ ($n = 12$). Median = $\dfrac{7+7}{2} = 7$.	Class B’s centre is slightly higher.
Range: Class A = $9 - 4 = 5$; Class B = $10 - 3 = 7$.	Class B is more spread out.
Shape: Class A peaks at 6 (its mode); Class B peaks at 7 and stretches wider.	Read the tallest stack for the mode.

∴ Class B scored slightly higher on average (median 7 versus 6.5) but was a little more spread out (range 7 versus 5); Class A’s results were more tightly bunched around 6.

Example 2 — scaffolded

The speeds (words per minute) achieved on a typing test by a group of students who completed a touch-typing course and a group who did not are shown in the back-to-back stem-and-leaf plot. Compare the two groups.

No course (leaf)	Stem	Course (leaf)
8 6 4 2	2	
8 5 3 1 0	3	4 6 9
2 0	4	1 3 5 7
	5	0 2 5 8

Key: $3\,/\,1 = 31$ *wpm. The no-course leaves are read leftward from the stem.*

Working	Comment
No course: $22, 24, 26, 28, 30, 31, 33, 35, 38, 40, 42$ ($n = 11$). Median $= 31$ wpm.	The median is the 6th value.
Course: $34, 36, 39, 41, 43, 45, 47, 50, 52, 55, 58$ ($n = 11$). Median $= 45$ wpm.	Also the 6th value.
Range: no course $= 42 - 22 = 20$; course $= 58 - 34 = 24$ wpm.	Similar spread.
The course group's leaves sit higher on the stem.	Higher stems = faster typing speeds.

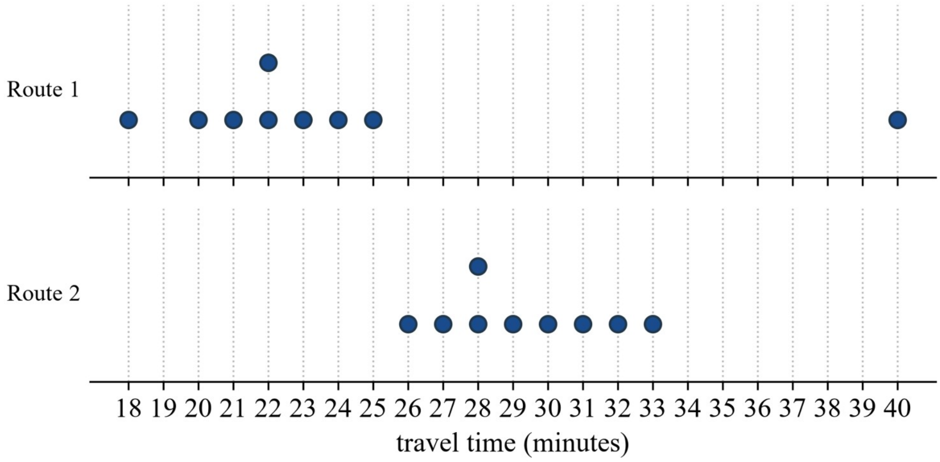
∴ the students who completed the course typed clearly faster (median 45 versus 31 wpm), while the spread of the two groups is about the same (ranges 20 and 24 wpm).

Example 3 — unscaffolded

A commuter timed the morning trip (in minutes) on two bus routes over nine weekdays:

$$\text{Route 1}: 18, 20, 21, 22, 22, 23, 24, 25, 40\,;\ \text{Route 2}: 26, 27, 28, 28, 29, 30, 31, 32, 33.$$

Compare the routes, and decide which is more reliable.



Both sets have $n = 9$, so each median is the 5th value: Route 1 has median 22 min and Route 2 has median 29 min, so Route 1 is usually **faster**. The ranges are $40 - 18 = 22$ min for Route 1 and $33 - 26 = 7$ min for Route 2, which makes Route 1 look far more variable. But Route 1's spread is inflated by a single unusual value, the 40-minute trip: ignoring it, Route 1 runs from 18 to 25 min, a range of 7 min — exactly the same spread as Route 2. So on a typical day Route 1 is both quicker and just as consistent, but it occasionally suffers a long delay.

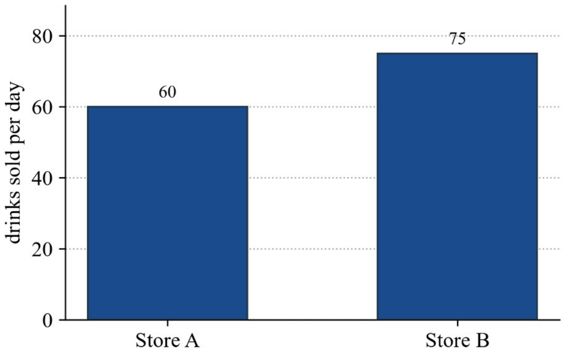
∴ Route 1 is typically faster (median 22 versus 29 min) and, apart from one 40-minute delay, just as consistent as Route 2; Route 2 is slower but never had a bad day in this sample.

Example 4 — unscaffolded

A manager says, "Store A sold 480 energy drinks and Store B only 300, so Store A is the better seller." Store A's figure is over 8 trading days and Store B's is over 4 trading days. Explain why the manager's comparison is unfair, and make a fair one.

Comparing the raw totals is unfair because the two stores were open for different numbers of days — Store A had twice as long to make its sales. A fair comparison uses the **rate**, sales *per day*:

$$\text{Store A} = \frac{480}{8} = 60\,\text{drinks/day}\,, \text{Store B} = \frac{300}{4} = 75\,\text{drinks/day}\,.$$



Per trading day, Store B actually sells **more** energy drinks than Store A.

∴ although Store A sold more in total, Store B is the stronger seller at 75 drinks per day versus 60 — the raw totals were misleading because the trading periods differed.

Practice questions

Scaffolded

Q1. The daily maximum temperatures (°C) over one week in two towns were

$$\text{Town X:}\,24,26,27,28,30,31,33;\ \text{Town Y:}\,20,22,25,29,31,34,36.$$

(a) Find the median maximum temperature of each town. (b) Find the range of each town. (c) Which town is warmer on average, and which has the more variable weather?

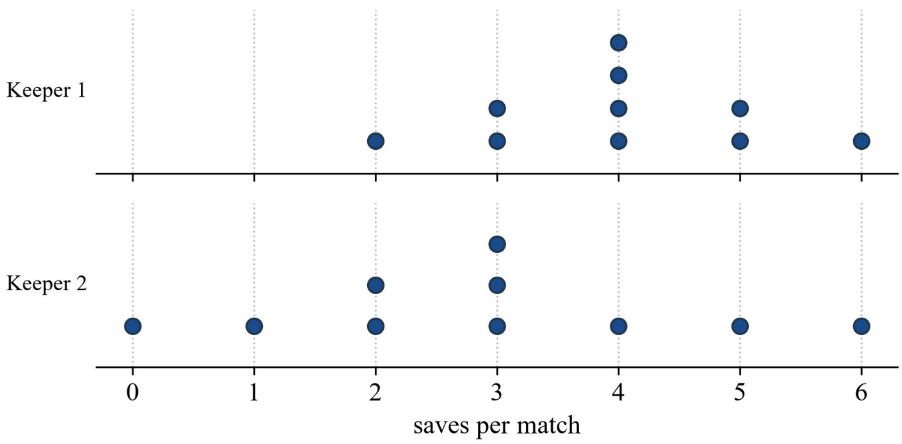
Q2. Two classes' test scores (out of 50) are shown in the back-to-back stem plot.

Class P (leaf)	Stem	Class Q (leaf)
8 5 2	2	
8 5 3 1	3	0 2 5 8 9
5 1 0	4	1 2 4 6 8

Key: $2\,/\,2 = 22$ *marks.*

(a) How many students are in each class?
(b) Find the median score of each class.
(c) State which class performed better and by roughly how much (compare the medians).

Q3. The parallel dot plot shows the number of saves made by two goalkeepers over ten matches.



(a) Write down the mode for each keeper.
(b) Find the range for each keeper.

(c) Compare the keepers in terms of centre and spread.

Q4. In School A, 60 of the 200 students surveyed walk to school. In School B, 45 of the 120 students surveyed walk to school. (a) Explain why comparing the raw numbers (60 versus 45) is unfair. (b) Find the percentage of students who walk in each school. (c) State which school has the greater proportion of students who walk.

Q5. Two baristas' service times (seconds) for eight orders were

$$\text{Barista 1:} 40, 42, 45, 48, 50, 52, 55, 60; \text{Barista 2:} 30, 35, 44, 48, 52, 58, 66, 75.$$

(a) Find the median service time of each barista. (b) Find the range of each barista. (c) The two medians are almost equal. Use the ranges to say which barista is more consistent.

Q6. For each comparison below, name a suitable display and give a reason. (a) The favourite type of movie (action, comedy, drama, …) chosen by a Year 10 group and a Year 12 group. (b) The heights (cm) of the students in two different classes. (c) The monthly rainfall (mm) of two towns across a whole year.

Unscaffolded

Q7. Two orchards recorded the number of apples on nine sample trees:

$$\text{Orchard A:} 30, 34, 36, 38, 40, 42, 44, 46, 60; \text{Orchard B:} 44, 45, 46, 47, 48, 49, 50, 51, 52.$$

Compare the orchards using the median and the range, and comment on any unusual value.

Q8. Worker A assembled 84 items in a 6-hour shift; Worker B assembled 75 items in a 5-hour shift. Decide who is the faster worker, justifying your answer with a fair comparison.

Q9. The back-to-back stem plot shows the weekly hours worked by a group of part-time and a group of full-time staff.

Part-time (leaf)	Stem	Full-time (leaf)
8 5 2	1	
8 5 4 2 0	2	
	3	2 5 6 8 8
	4	0 2 5

Key: $1 / 2 = 12$ *hours.*

Compare the two groups' weekly hours using the median and the range.

Q10. Two groups played a number game; their scores were

$$\text{Group A:} 1, 2, 2, 3, 3, 3, 3, 4, 4, 5; \text{Group B:} 1, 1, 1, 2, 2, 3, 4, 5, 6, 7.$$

(a) Find the median of each group. (b) Describe and compare the **shape** of each group (symmetric, or skewed, with any clusters).

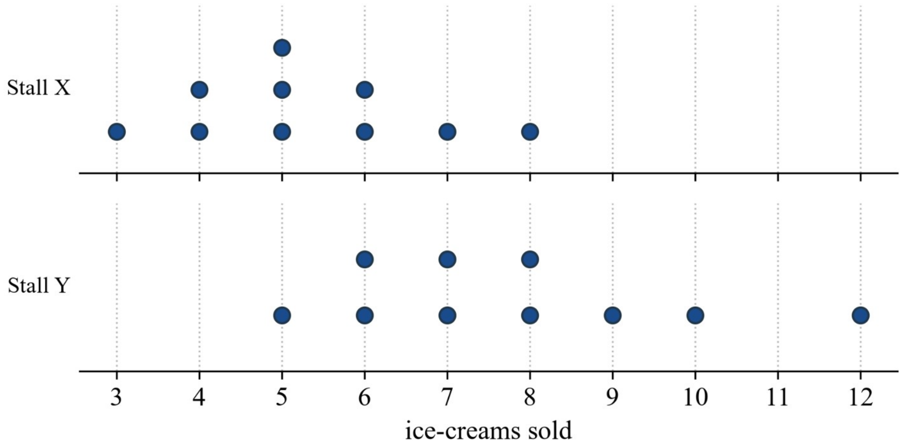
Q11. In a season, Team A scored a goal in 12 of its 20 games and Team B scored in 15 of its 22 games. Using a fair comparison, decide which team scored in the greater proportion of its games.

Q12. The annual salaries (in thousands of dollars) of the seven staff in two small departments were

$$\text{Dept A:} 55, 58, 60, 62, 64, 66, 120; \text{Dept B:} 68, 70, 72, 74, 76, 78, 80.$$

(a) Find the median and the mean of each department. (b) Explain why the **median** is a fairer measure of a typical salary in Dept A than the mean. (c) Using the medians, state which department pays more typically.

Q13. The parallel dot plot shows the number of ice-creams sold each day by two stalls over ten days.



(a) Find the median daily sales for each stall.
(b) Find the range for each stall.
(c) Write a sentence comparing the two stalls in terms of centre and spread.

Q14. A student wants to compare the number of pets owned by families in two suburbs. Explain which single display would let a reader compare the two suburbs most easily, and say what to put on each axis.

Q15. Town A recycled 4200 kg of glass from a population of 3000 people; Town B recycled 3500 kg from a population of 2000 people. Decide which town recycles more effectively, using a fair comparison.

Q16. The finishing times (minutes) of the first ten runners in a fun run in two successive years were

$$2024 : 210, 215, 220, 225, 230, 235, 240, 245, 250, 300; 2025 : 205, 210, 212, 215, 218, 220, 222, 225, 228, 230.$$

Compare the two years using the median and the range, and comment on the effect of any unusual value.

Q17. An online post claims, “Café A has 90 five-star reviews and Café B only 60, so Café A is better.” Café A has 200 reviews in total and Café B has 100. Explain the flaw in the claim and make a fair comparison.

Q18. Two bank branches recorded customer wait times (minutes) for eleven customers:

$$\text{Branch A}: 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 25; \text{Branch B}: 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15.$$

(a) Find the median wait time of each branch. (b) Find the range of each branch, and then the range of Branch A **without** its unusual value. (c) Using your results, state which branch serves customers faster and whether they are really more variable.

Answers

Q1. (a) Town X 28 °C, Town Y 29 °C. (b) Town X 9 °C, Town Y 16 °C. (c) Town Y is slightly warmer on average and much more variable. **Q2.** (a) 10 each. (b) Class P 34, Class Q 40. (c) Class Q performed better by about 6 marks (medians 40 versus 34). **Q3.** (a) Keeper 1 mode 4; Keeper 2 mode 3. (b) Keeper 1 range 4; Keeper 2 range 6. (c) Keeper 1 makes more saves typically (median 4 versus 3) and is more consistent (range 4 versus 6). **Q4.** (a) The schools surveyed different numbers of students, so the totals are not comparable. (b) School A 30%, School B 37.5%. (c) School B. **Q5.** (a) Barista 1 49 s, Barista 2 50 s. (b) Barista 1 20 s, Barista 2 45 s. (c) Barista 1 is more consistent (much smaller range). **Q6.** (a) A side-by-side (grouped) column graph — the data is categorical. (b) Parallel dot plots or a back-to-back stem plot — numerical data, similar sizes. (c) A line graph with two lines (one per town) — numerical data over time. **Q7.** Orchard A median 40, range 30; Orchard B median 48, range 8. Orchard B has more apples per tree and is far more consistent; Orchard A has an unusually high tree (60) that inflates its range. **Q8.** Worker A $=14$ items/h; Worker B $=15$ items/h. Worker B is faster once the different shift lengths are taken into account. **Q9.** Part-time median 21 h, range 16 h; full-time median 38 h, range 13 h. Full-time staff work far more hours and slightly more consistently. **Q10.** (a) Group A 3, Group B 2.5. (b) Group A is roughly symmetric, clustered around 3; Group B is skewed to the right (bunched at 1–2 with a tail up to 7). **Q11.** Team A 60%, Team B $\approx 68.2\%$; Team B scored in the greater proportion of games. **Q12.** (a) Dept A median 62, mean ≈ 69.3; Dept B median 74, mean 74. (b) The $\$120$k salary is an outlier that pulls Dept A's mean well above a typical salary, so the median is fairer. (c) Dept B pays more typically (median 74 versus 62). **Q13.** (a) Stall X 5, Stall Y 7.5. (b) Stall X 5, Stall Y 7. (c) Stall Y sells more per day (median 7.5 versus 5) and is a little more spread out (range 7 versus 5). **Q14.** Parallel dot plots (or a back-to-back stem plot): number of pets on the horizontal axis, one lane per suburb, so the two suburbs share a scale. **Q15.** Town A $=1.4$ kg/person; Town B $=1.75$ kg/person. Per person, Town B recycles more, even though Town A recycled more in total. **Q16.** 2024 median 232.5 min, range 90 min; 2025 median 219 min, range 25 min. Runners were faster and far more consistent in 2025; 2024's range is inflated by a slow 300-minute finisher.

Q17. The claim ignores the total number of reviews. Café A: $\frac{90}{200}=45\%$ five-star; Café B: $\frac{60}{100}=60\%$ five-star. As a proportion, Café B rates better. **Q18.** (a) Branch A 7 min, Branch B 10 min. (b) Branch A range $25-2=23$ min; without the 25, $11-2=9$ min; Branch B range 10 min. (c) Branch A serves customers faster (median 7 versus 10) and, apart from one 25-minute wait, is no more variable than Branch B (typical range 9 versus 10 min).

Fully-worked solutions

Q1. Each town has $n=7$, so the median is the 4th value. (a) Town X: $24, 26, 27, 28, 30, 31, 33 \rightarrow$ median 28 °C. Town Y: $20, 22, 25, 29, 31, 34, 36 \rightarrow$ median 29 °C. (b) Range: Town X $=33-24=9$ °C; Town Y $=36-20=16$ °C. (c) The medians (28 and 29 °C) are close, so Town Y is only slightly warmer on average; but its range (16 °C) is much larger than Town X's (9 °C), so Town Y has the more variable weather.

Q2. (a) Class P has leaves $3+4+3=10$ students; Class Q has $5+5=10$ students. (b) Class P: $22, 25, 28, 31, 33, 35, 38, 40, 41, 45$, median $=\frac{33+35}{2}=34$. Class Q: $30, 32, 35, 38, 39, 41, 42, 44, 46, 48$, median $=\frac{39+41}{2}=40$. (c) Class Q's median (40) is 6 marks above Class P's (34), so Class Q performed better by about 6 marks.

Q3. (a) The tallest stack is at 4 for Keeper 1 and at 3 for Keeper 2, so the modes are 4 and 3 saves. (b) Keeper 1: $2, 3, 3, 4, 4, 4, 4, 5, 5, 6 \rightarrow$ range $=6-2=4$. Keeper 2: $0, 1, 2, 2, 3, 3, 3, 4, 5, 6 \rightarrow$ range $=6-0=6$. (c) Keeper 1 (median 4) typically makes more saves than Keeper 2 (median 3) and is more consistent (range 4 versus 6).

Q4. (a) School A surveyed 200 students and School B only 120, so a larger raw count for A does not mean a larger *share* — the group sizes differ. (b) School A: $\frac{60}{200}\times 100\%=30\%$. School B: $\frac{45}{120}\times 100\%=37.5\%$. (c) $37.5\%>30\%$, so School B has the greater proportion of students who walk.

Q5. (a) Both have $n=8$, so the median averages the 4th and 5th values. Barista 1: $\frac{48+50}{2}=49$ s; Barista 2: $\frac{48+52}{2}=50$ s. (b) Barista 1 range $=60-40=20$ s; Barista 2 range $=75-30=45$ s. (c) The medians are almost the same (49 and 50 s), but Barista 1's range is far smaller (20 versus 45 s), so Barista 1 is more consistent.

Q6. (a) A **side-by-side (grouped) column graph**: movie type is categorical, and grouped columns let each category be compared across the two year levels. (b) **Parallel dot plots** or a **back-to-back stem-and-leaf plot**: heights are numerical and the classes are similar in size, so a shared scale compares them directly. (c) A **line graph with two lines**: rainfall is numerical measured over the twelve months, and two lines on the same axes show the towns' patterns over time.

Q7. Both have $n=9$, median = 5th value. Orchard A: $30, 34, 36, 38, 40, 42, 44, 46, 60$ → median 40, range $60-30=30$. Orchard B: $44, 45, 46, 47, 48, 49, 50, 51, 52$ → median 48, range $52-44=8$. Orchard B typically has more apples per tree (median 48 versus 40) and is far more consistent. Orchard A has one unusually productive tree (60 apples) that stretches its range; without it, Orchard A runs 30–46, a range of 16 — still more variable than B.

Q8. Compare the rates. Worker A: $\frac{84}{6}=14$ items per hour. Worker B: $\frac{75}{5}=15$ items per hour. Although Worker A assembled more in total, Worker B assembles more each hour, so **Worker B is the faster worker**.

Q9. Part-time: $12, 15, 18, 20, 22, 24, 25, 28$ ($n=8$), median $=\frac{20+22}{2}=21$ h, range $=28-12=16$ h. Full-time: $32, 35, 36, 38, 38, 40, 42, 45$ ($n=8$), median $=\frac{38+38}{2}=38$ h, range $=45-32=13$ h. Full-time staff work far more hours per week (median 38 versus 21) and are slightly more consistent (range 13 versus 16 h).

Q10. (a) Both have $n=10$. Group A: $1, 2, 2, 3, 3, 3, 3, 4, 4, 5$ → median $=\frac{3+3}{2}=3$. Group B: $1, 1, 1, 2, 2, 3, 4, 5, 6, 7$ → median $=\frac{2+3}{2}=2.5$. (b) Group A is roughly **symmetric**, clustered around its centre of 3 with values falling away evenly on each side. Group B is **skewed to the right**: most scores bunch at 1 and 2, with a tail of larger values stretching out to 7.

Q11. Compare proportions. Team A: $\frac{12}{20}=0.60=60\%$. Team B: $\frac{15}{22}=0.6818\ldots\approx 68.2\%$. Since $68.2\%>60\%$, **Team B** scored in the greater proportion of its games, even though it played more games.

Q12. (a) Dept A: $55, 58, 60, 62, 64, 66, 120$ ($n=7$), median $=62$; mean $=\frac{485}{7}\approx 69.3$ (thousand dollars). Dept B: $68, 70, 72, 74, 76, 78, 80$, median $=74$; mean $=\frac{518}{7}=74$. (b) In Dept A the single very high salary of $\$120$k is an **outlier**; it drags the mean up to about $\$69.3$k, well above what most staff earn. The median ($\$62$k) is not affected by that one value, so it better represents a typical salary. (c) Comparing medians, Dept B ($\$74$k) pays more typically than Dept A ($\$62$k).

Q13. (a) Both have $n=10$. Stall X: $3, 4, 4, 5, 5, 5, 6, 6, 7, 8$ → median $=\frac{5+5}{2}=5$. Stall Y: $5, 6, 6, 7, 7, 8, 8, 9, 10, 12$ → median $=\frac{7+8}{2}=7.5$. (b) Stall X range $=8-3=5$; Stall Y range $=12-5=7$. (c) Stall Y sells more ice-creams per day (median 7.5 versus 5) and is a little more spread out (range 7 versus 5).

Q14. Use **parallel dot plots** (or a back-to-back stem-and-leaf plot). Put the **number of pets** on the horizontal axis with a single shared scale, and give each suburb its own lane (or side of the stem). Because both suburbs share the same axis, a reader can line them up and compare centre and spread at a glance.

Q15. Compare the amount recycled **per person**. Town A: $\frac{4200}{3000} = 1.4$ kg/person. Town B: $\frac{3500}{2000} = 1.75$ kg/person. Town B recycles more per person (1.75 versus 1.4 kg), so it recycles more effectively, even though Town A recycled more in total.

Q16. Both have $n = 10$, so the median averages the 5th and 6th values. 2024: $\ldots 230, 235 \ldots$ → median $= \frac{230 + 235}{2} = 232.5$ min, range $= 300 - 210 = 90$ min. 2025: $\ldots 218, 220 \ldots$ → median $= \frac{218 + 220}{2} = 219$ min, range $= 230 - 205 = 25$ min. Runners in 2025 were faster (median 219 versus 232.5 min) and far more consistent (range 25 versus 90 min). The 2024 range is greatly inflated by one slow finisher (300 min); ignoring that runner, 2024 runs 210–250, a range of 40 min — still more spread than 2025.

Q17. The claim compares only the raw counts of five-star reviews and ignores that the two cafés have different total numbers of reviews. As a proportion of all reviews, Café A has $\frac{90}{200} = 45\,\%$ five-star and Café B has $\frac{60}{100} = 60\,\%$ five-star. Café B has the higher proportion of five-star reviews, so the original claim is misleading.

Q18. Both have $n = 11$, so the median is the 6th value. (a) Branch A: $2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 25$ → median 7 min. Branch B: $5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15$ → median 10 min. (b) Branch A range $= 25 - 2 = 23$ min; without the 25-minute wait the largest value is 11, so the range is $11 - 2 = 9$ min. Branch B range $= 15 - 5 = 10$ min. (c) Branch A serves customers faster on average (median 7 versus 10 min). Its full range (23 min) makes it look far more variable, but that is due to a single 25-minute wait; the typical spread (9 min) is about the same as Branch B's (10 min). So Branch A is faster and not really more variable, apart from one unusually long wait.

Theory

Data becomes most useful when we **compare** one group with another: two teams, two suburbs, a "before" group and an "after" group. A fair comparison never rests on a single number — it looks at three things.

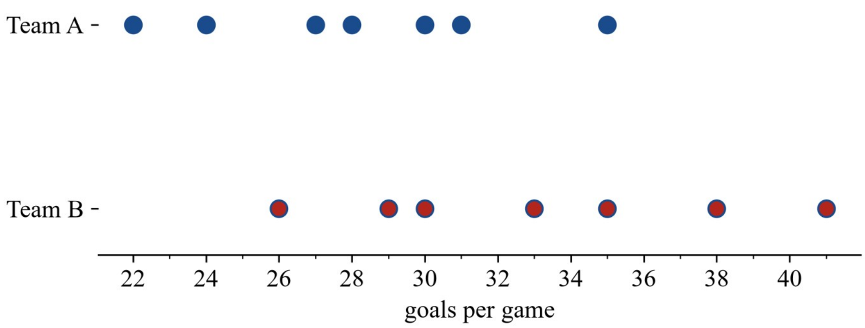
The three things we compare.

- **Centre** — where the data "sits". Use the **median** (the middle value) or the **mean** (the average). The median position in an ordered list of n values is the $\frac{n+1}{2}$th value.
- **Spread** — how stretched out the data is. Use the **range** $(\text{maximum} - \text{minimum})$. A smaller spread means the data is more **consistent**. (Section 6F develops a second measure of spread, the interquartile range.)
- **Shape** — is the data roughly **symmetric**, or **skewed** (bunched at one end with a tail)? Are there **clusters**, **gaps**, or unusually high/low values (**outliers**)?

Median or mean? When a data set has an outlier or is skewed, the **median** describes its centre better, because it is barely affected by one extreme value. The mean can be dragged a long way by a single outlier — and so can the range, which uses only the two most extreme values.

Displays that line two groups up. Two displays are especially useful here:

- **Parallel dot plots** — two dot plots drawn one above the other on the *same* scale.
- **Back-to-back stem-and-leaf plots** — one shared stem down the middle, one group's leaves to the left and the other's to the right.



These parallel dot plots compare goals scored by two teams. Reading them: Team B's dots sit further right (a higher median, 33 vs 28 goals) and stretch a little further (slightly more spread). Notice how quickly the centre, spread and shape can be read from a well-drawn display.

Writing a comparison. A complete comparison mentions **centre and spread, in context**, using comparative words:

> *"Team B typically scored more goals than Team A (median* 33 *compared with* 28*), and Team B's scores were slightly more spread out (range* 15 *compared with* 13*)."*

Worked examples

Example 1 — scaffolded

Two netball teams recorded the goals they scored in seven games. Team A: $24, 31, 28, 35, 22, 30, 27$. Team B: $33, 29, 41, 26, 38, 30, 35$. Find the median and range of each team, and compare the teams.

Working	Comment
Team A ordered: $22, 24, 27, 28, 30, 31, 35$. Median $= 28$.	$n = 7$, so the median is the $\frac{7+1}{2} = 4$th value.
Team A range $= 35 - 22 = 13$.	Maximum − minimum.
Team B ordered: $26, 29, 30, 33, 35, 38, 41$. Median $= 33$.	The 4th value again.
Team B range $= 41 - 26 = 15$.	
Team B typically scored more (median 33 vs 28) with slightly more spread (range 15 vs 13).	Compare **centre** and **spread**, in context.

Example 2 — scaffolded

The back-to-back stem-and-leaf plot shows the number of eggs collected each morning over ten days from two chicken coops. Compare the two coops.

Coop A (leaf)	Stem	Coop B (leaf)
9 8	1	4 7 9
6 5 4 3 2 1 1 0	2	2 3 5 7 8
	3	0 1

Key: $2 / 1 = 21$ eggs.

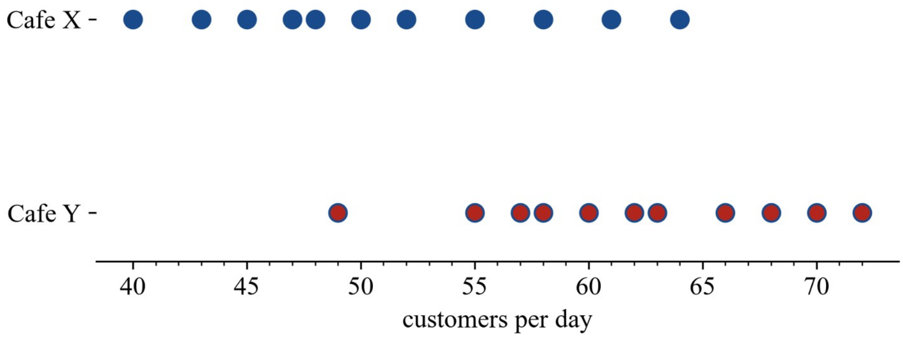
Working	Comment
Coop A ordered: $18, 19, 20, 21, 21, 22, 23, 24, 25, 26$. Median $= \frac{21+22}{2} = 21.5$ eggs.	$n = 10$: average the 5th and 6th values.
Coop B ordered: $14, 17, 19, 22, 23, 25, 27, 28, 30, 31$. Median $= \frac{23+25}{2} = 24$ eggs.	The 5th and 6th values again.
Coop A range $= 26 - 18 = 8$; Coop B range $= 31 - 14 = 17$.	Maximum − minimum.
Coop B typically laid more eggs (median 24 vs 21.5) but varied more (range 17 vs 8); Coop A was more consistent.	Higher centre **but** larger spread.

Example 3 — unscaffolded

Over 11 days a café owner counted the customers at two branches. Café X: $40, 52, 47, 61, 55, 43, 58, 50, 64, 48, 45$. Café Y: $55, 62, 58, 70, 49, 66, 60, 72, 57, 63, 68$. Construct parallel dot plots and compare the branches.

Ordering each set ($n = 11$, so the median is the 6th value):

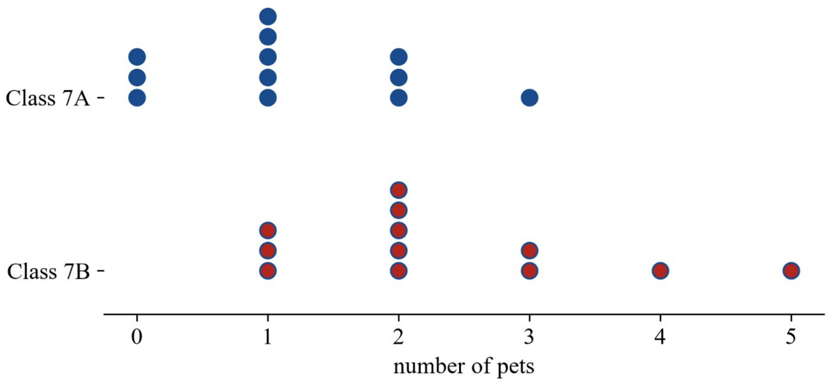
Café X: $40, 43, 45, 47, 48, 50, 52, 55, 58, 61, 64$ — median 50, range $= 64 - 40 = 24$. Café Y: $49, 55, 57, 58, 60, 62, 63, 66, 68, 70, 72$ — median 62, range $= 72 - 49 = 23$.



∴ Café Y is busier than Café X (median 62 vs 50 customers), while the two branches vary by a similar amount from day to day (range 23 vs 24).

Example 4 — unscaffolded

Two Year 7 classes were asked how many pets they own. Class 7A: $0, 1, 1, 2, 0, 3, 1, 2, 1, 0, 2, 1$. Class 7B: $1, 2, 2, 3, 1, 2, 4, 2, 3, 1, 2, 5$. Display the data as parallel dot plots and compare the classes.



Class 7A ordered: $0, 0, 0, 1, 1, 1, 1, 1, 2, 2, 2, 3$ — median $= \frac{1+1}{2} = 1$ pet, range $3 - 0 = 3$. Class 7B ordered: $1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 4, 5$ — median $= \frac{2+2}{2} = 2$ pets, range $5 - 1 = 4$.

The 7A dots cluster at 0 and 1; the 7B dots sit higher and trail off to the right with a couple of large values (4 and 5), a slight right skew.

∴ Class 7B students tend to own more pets (median 2 vs 1) and their counts are a little more spread out (range 4 vs 3), with a small right tail.

Practice questions

Scaffolded

Q1. The waiting times (minutes) at a bus stop were recorded one morning and one afternoon. Morning: $4, 7, 5, 9, 6, 8, 5$. Afternoon: $8, 6, 11, 7, 9, 10, 7$. (a) Find the median waiting time for each. (b) Find the range for each. (c) Write one sentence comparing the two periods.

Q2. Two classes sat the same test (out of 20). Class P: $12, 15, 14, 18, 11$. Class Q: $16, 13, 19, 15, 17$. (a) Find the mean mark of each class. (b) Find the range of each class. (c) Which class performed better overall? Justify your answer.

Q3. The table summarises the points scored by two basketball teams across their last 15 games.

	Min	Median	Max
Team X	52	61	78
Team Y	55	64	70

(a) Find the range of each team's scores.
(b) Compare the centres of the two teams' scores.
(c) Which team's scores are more consistent? Justify your answer.

Q4. The back-to-back stem plot shows the goals scored per season by two players.

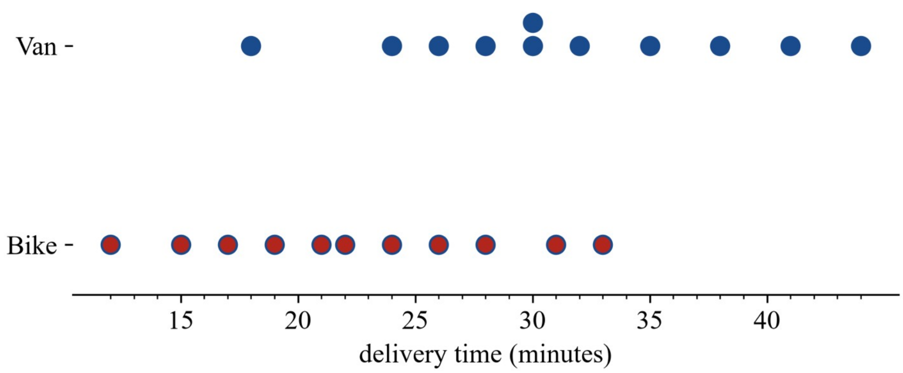
Player L (leaf)	Stem	Player R (leaf)
	0	8
8 5 2	1	4 6 9
7 4 3 1	2	2 5 8 9
3 1	3	5

Key: $1\,/\,2 = 12$ goals.

(a) Find the median for each player.
(b) Find the range for each player.
(c) Which player is more consistent? Explain.

Q5. The weekly rents (dollars) of seven properties in two suburbs are listed. Suburb M: $380, 400, 410, 420, 430, 440, 900$. Suburb N: $450, 460, 470, 480, 490, 500, 510$. (a) Find the median rent of each suburb. (b) Find the mean rent of each suburb. (c) For Suburb M, is the median or the mean a better measure of the "typical" rent? Explain.

Q6. The parallel dot plots show the delivery times (minutes) of a van and a bike courier over 11 deliveries each.



(a) Write down the median delivery time for each courier.
(b) Find the range for each courier.
(c) Which courier is faster and more consistent? Refer to your figures.

Unscaffolded

Q7. Daily sales (items) at two stores over a week were: Store A: $21, 25, 28, 30, 34, 26, 29$. Store B: $33, 27, 40, 31, 38, 29, 35$. Compare the two stores' sales in terms of centre (median) and spread (range), in context.

Q8. The back-to-back stem plot shows the number of points scored in a general-knowledge quiz by a Year 11 and a Year 12 class.

Year 11 (leaf)	Stem	Year 12 (leaf)
9 5	4	
8 7 5 4 2	5	8
3 1 0	6	1 3 4 6 7 9
	7	0 2 5

Key: $4 / 5 = 45$ points.

Find the median and range of each class, then write a comparison of the two year levels.

Q9. At a swimming club's time trials, the 50 m freestyle times (seconds) of two swimmers over a season are summarised below.

	Min	Median	Max
Swimmer A	33.9	35.8	39.6
Swimmer B	35.0	36.2	37.4

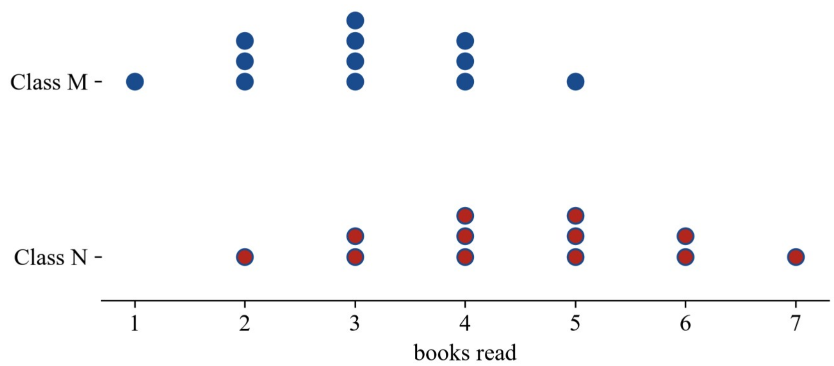
Compare the two swimmers' times, referring to centre and spread. Which swimmer would the coach choose for a relay team if she values a *reliable* race time?

Q10. A charity recorded the value (dollars) of gifts received from seven donors in a country town and seven donors in the city. Town: $15, 18, 20, 22, 24, 26, 95$. City: $40, 44, 47, 50, 53, 56, 60$. (a) Find the median and the mean gift for each group. (b) Explain why the mean is misleading for the Town data. (c) Using the medians, compare the typical gift in the two locations.

Q11. The comparison in Question 6 used 11 deliveries for each courier and drew both dot plots on the same scale. Explain why this makes the comparison fair, and state one thing the dot plots show that the two medians and ranges alone would not.

Q12. The monthly rainfall (mm) for a year was recorded in two towns. Alba: $22, 35, 48, 60, 71, 80, 78, 66, 54, 40, 29, 24$. Bica: $40, 44, 52, 63, 70, 74, 72, 68, 60, 55, 48, 43$. (a) Find the median and range of each town's rainfall. (b) Which town has the more variable rainfall across the year? Support your answer with the ranges. (c) Comment on what your comparison suggests about the two climates.

Q13. The parallel dot plots show the number of books read in a month by two classes.



Compare the two classes, referring to centre, spread and the shape of each distribution.

Q14. A factory checks the fill (grams) of jars from two machines. Machine 1: $248, 252, 250, 249, 251, 250, 247, 253$. Machine 2: $240, 262, 245, 258, 238, 265, 250, 242$. (a) Show that both machines have the same mean fill. (b) Find the range of each machine. (c) Which machine is more reliable? Explain why the mean alone does not tell the whole story.

Q15. Eight runners recorded their time (seconds) for a sprint before and after a six-week training program. Before: $42, 38, 45, 50, 47, 41, 44, 39$. After: $55, 60, 52, 63, 58, 54, 61, 57$. (a) Find the median before and after. (b) By how much did the median change? (c) Did the *spread* of times change much? A student claims “the training made everyone faster.” Evaluate this claim using the data. (Note: a lower time is faster.)

Q16. Two garden plots each grew 10 tomatoes. Their masses (grams) were Plot A: $71, 72, 74, 75, 76, 76, 77, 78, 79, 105$. Plot B: $60, 68, 75, 80, 83, 85, 88, 93, 100, 108$. (a) Find the median mass and the range of each plot. (b) Describe the shape of each distribution, referring to clusters, gaps and outliers. (c) Compare the two plots in context.

Q17. A newspaper compares two hospitals’ waiting times by reporting only the *mean* wait at each. Give **two** reasons why comparing only the means could give a misleading picture, and name one extra statistic and one display that would make the comparison fairer.

Q18. Ten batteries of two brands were tested for life (hours). Brand P: $88, 92, 85, 90, 79, 94, 87, 91, 83, 89$. Brand Q: $95, 82, 99, 90, 102, 88, 97, 93, 105, 86$. (a) Find the median life of each brand. (b) Find the range of each brand. (c) A shopper wants the longest-lasting battery; a hospital wants the most *predictable* battery life. Recommend a brand for each, justifying both choices from your statistics.

Answers

Q1. (a) Morning 6, afternoon 8 min. (b) both range 5 min. (c) Afternoon waits were typically longer (median 8 vs 6) with the same spread. **Q2.** (a) P 14, Q 16. (b) P 7, Q 6. (c) Class Q — higher mean (16 vs 14) and slightly smaller spread. **Q3.** (a) X 26, Y 15. (b) Team Y's centre is slightly higher (median 64 vs 61). (c) Team Y — its scores are much less spread out (range 15 vs 26). **Q4.** (a) L 23, R 22. (b) L 21, R 27. (c) Player L — similar median but a smaller range (21 vs 27), so more consistent. **Q5.** (a) M $\$420$, N $\$480$. (b) M $\$482.86$, N $\$480$. (c) Median — the $\$900$ outlier inflates M's mean; the median $\$420$ better reflects the typical rent. **Q6.** (a) Van 30, Bike 22 min. (b) Van 26, Bike 21 min. (c) The bike — lower median (22 vs 30) and smaller range (21 vs 26), so faster and more consistent. **Q7.** Median A 28, B 33; range both 13. Store B sells more typically (median 33 vs 28) with the same overall spread; A's values cluster in the middle while B's are evenly spaced. **Q8.** Year 11: median 56, range 18. Year 12: median 66.5, range 17. Year 12 scored higher (median 66.5 vs 56) with a similar spread. **Q9.** Ranges: A 5.7 s, B 2.4 s; medians 35.8 vs 36.2 s. Swimmer A is faster on average but far less consistent — for a reliable race time, choose Swimmer B. **Q10.** (a) Town median $\$22$, mean $\$31.43$; City median $\$50$, mean $\$50$. (b) The $\$95$ outlier pulls the Town mean well above most gifts. (c) City gifts are typically higher (median $\$50$ vs $\$22$). **Q11.** Same variable, same number of deliveries and the same scale make it fair; the dot plots show every individual time (clusters, gaps, repeated values, shape), which the medians and ranges alone do not. **Q12.** (a) Alba median 51, range 58; Bica median 57.5, range 34. (b) Alba (range 58 vs 34). (c) Bica has steadier rainfall; Alba is more strongly seasonal. **Q13.** Class M median 3, range 4; Class N median 4.5, range 5. Class N reads more books (higher median) with a similar spread; both roughly symmetric, N shifted higher. **Q14.** (a) Both means $= 250$ g. (b) Machine 1 range 6; Machine 2 range 27. (c) Machine 1 — far smaller spread, so more reliable; equal means hide very different consistency. **Q15.** (a) Before 43, After 57.5 s. (b) Increased by 14.5 s. (c) Spread barely changed (range 12 vs 11). Times **rose**, and since lower is faster, runners got *slower*, not faster — the claim is wrong. **Q16.** (a) A median 76 g, range 34 g; B median 84 g, range 48 g. (b) A clusters at 71–79 g with a gap and a high outlier (105 g); B is roughly symmetric with no gaps or outliers. (c) B's tomatoes are typically heavier (higher median) and more variable (larger range); A's range is inflated by its outlier. **Q17.** e.g. the mean hides spread/consistency and is distorted by outliers; add the median (or the range), and show parallel dot plots. **Q18.** (a) P 88.5, Q 94 h. (b) P range 15; Q range 23. (c) Longest-lasting: Brand Q (higher median). Most predictable: Brand P (smaller range).

Fully-worked solutions

Q1. Morning ordered $4, 5, 5, 6, 7, 8, 9$: median $= 6$ (4th value), range $= 9 - 4 = 5$. Afternoon ordered $6, 7, 7, 8, 9, 10, 11$: median $= 8$, range $= 11 - 6 = 5$. Afternoon waits are typically two minutes longer, with the same range.

Q2. Class P mean $= \dfrac{12+15+14+18+11}{5} = \dfrac{70}{5} = 14$; range $= 18 - 11 = 7$. Class Q mean $= \dfrac{16+13+19+15+17}{5} = \dfrac{80}{5} = 16$; range $= 19 - 13 = 6$. Class Q performed better: a higher mean and a slightly smaller spread.

Q3. (a) $\text{range}_X = 78 - 52 = 26$; $\text{range}_Y = 70 - 55 = 15$. (b) Team Y's median (64) is slightly higher than Team X's (61), so Team Y's centre is a little higher. (c) Team Y is more consistent: its scores stay within a 15-point band, while Team X's swing over a 26-point band.

Q4. Reading the stems (L leaves read right-to-left): Player L $= 12, 15, 18, 21, 23, 24, 27, 31, 33$; Player R $= 8, 14, 16, 19, 22, 25, 28, 29, 35$. (a) $n = 9$, median is the 5th value: L $= 23$, R $= 22$. (b) $\text{range}_L = 33 - 12 = 21$; $\text{range}_R = 35 - 8 = 27$. (c) Player L: the medians are almost equal, but L's range is smaller (21 vs 27), so L's scoring is more consistent.

Q5. (a) M ordered ends at 900; median (4th of 7) $= \$420$. N median $= \$480$. (b) M mean $= \dfrac{380+400+410+420+430+440+900}{7} = \dfrac{3380}{7} = \$482.86$; N mean $= \dfrac{3360}{7} = \$480$. (c) The single very high rent ($\$900$) is an outlier that drags M's mean up to $\$482.86$, above six of the seven rents. The median $\$420$ sits among the typical values, so it is the better measure.

Q6. Reading the dot plots — Van: $18, 24, 26, 28, 30, 30, 32, 35, 38, 41, 44$; Bike: $12, 15, 17, 19, 21, 22, 24, 26, 28, 31, 33$ ($n = 11$ each). (a) With 11 values the median is the 6th: Van 30 min, Bike 22 min. (b) $\text{range}_{\text{Van}} = 44 - 18 = 26$ min; $\text{range}_{\text{Bike}} = 33 - 12 = 21$ min. (c) The bike is faster — its median (22 min) is 8 minutes below the van's (30 min) — and more consistent, since its range is smaller (21 vs 26 min).

Q7. Store A ordered $21, 25, 26, 28, 29, 30, 34$: median 28, range $34 - 21 = 13$. Store B ordered $27, 29, 31, 33, 35, 38, 40$: median 33, range $40 - 27 = 13$. Store B typically sells more (median 33 vs 28). Both have the same range (13), so their overall spread is similar, but their shapes differ: A's values cluster around 25–30, while B's are more evenly spread out.

Q8. Year 11 $= 45, 49, 52, 54, 55, 57, 58, 60, 61, 63$: median $= \frac{55+57}{2} = 56$, range $= 63 - 45 = 18$. Year 12 $= 58, 61, 63, 64, 66, 67, 69, 70, 72, 75$: median $= \frac{66+67}{2} = 66.5$, range $= 75 - 58 = 17$. Year 12 scored higher across the board (median 66.5 vs 56, and its lowest score 58 almost matches Year 11's median), with a similar spread of scores.

Q9. Ranges: Swimmer A $= 39.6 - 33.9 = 5.7$ s; Swimmer B $= 37.4 - 35.0 = 2.4$ s. Swimmer A is faster on average (median 35.8 vs 36.2 s), but their times swing through almost 6 s, while Swimmer B's stay within 2.4 s. For a relay team that values reliability, choose Swimmer B — only slightly slower on average, but far more predictable.

Q10. (a) Town ordered $15, 18, 20, 22, 24, 26, 95$: median $\$22$, mean $= \frac{220}{7} = \$31.43$. City ordered $40, 44, 47, 50, 53, 56, 60$: median $\$50$, mean $= \frac{350}{7} = \$50$. (b) The $\$95$ gift is an outlier; it lifts the Town mean to $\$31.43$, higher than six of the seven gifts, so the mean overstates the typical gift. (c) Using medians, city donors typically gave more than twice as much ($\$50$ vs $\$22$).

Q11. Both couriers were measured on the same variable (delivery time, in minutes) over the same number of deliveries (11), and the two dot plots share one scale, so the comparison is like with like and the plots can be read directly against each other. The dot plots show every individual time, so clusters, gaps, repeated values and the overall shape are visible; the two medians and ranges compress each data set to two numbers and hide all of that.

Q12. (a) Alba ordered $22, 24, 29, 35, 40, 48, 54, 60, 66, 71, 78, 80$: median $= \frac{48+54}{2} = 51$, range $= 80 - 22 = 58$. Bica ordered $40, 43, 44, 48, 52, 55, 60, 63, 68, 70, 72, 74$: median $= \frac{55+60}{2} = 57.5$, range $= 74 - 40 = 34$. (b) Alba's rainfall is more variable — its range (58 mm) is much larger than Bica's (34 mm). (c) Bica's rainfall stays in a narrower band all year (a steadier, milder climate), while Alba swings between dry and very wet months (a strongly seasonal climate).

Q13. Reading the dot plots — Class M $= 1, 2, 2, 2, 3, 3, 3, 3, 4, 4, 4, 5$: median $= \frac{3+3}{2} = 3$, range $= 5 - 1 = 4$. Class N $= 2, 3, 3, 4, 4, 4, 5, 5, 5, 6, 6, 7$: median $= \frac{4+5}{2} = 4.5$, range $= 7 - 2 = 5$. Class N reads more books (median 4.5 vs 3); the spreads are alike (range 5 vs 4). Both distributions are roughly symmetric and single-peaked, with N's whole pattern shifted about $1\,1/2$ books higher.

Q14. (a) Machine 1 sum $= 248 + 252 + 250 + 249 + 251 + 250 + 247 + 253 = 2000$, mean $= 2000/8 = 250$ g. Machine 2 sum $= 240 + 262 + 245 + 258 + 238 + 265 + 250 + 242 = 2000$, mean $= 250$ g. (b) Machine 1 ordered $247, 248, 249, 250, 250, 251, 252, 253$: range $= 253 - 247 = 6$ g. Machine 2 ordered $238, 240, 242, 245, 250, 258, 262, 265$: range $= 265 - 238 = 27$ g. (c) Machine 1 is far more reliable: although both average 250 g, Machine 1's fills stay within 6 g while Machine 2's spread over 27 g. Equal means can hide very different consistency, so spread must also be reported.

Q15. (a) Before ordered $38, 39, 41, 42, 44, 45, 47, 50$: median $= \frac{42+44}{2} = 43$ s. After ordered $52, 54, 55, 57, 58, 60, 61, 63$: median $= \frac{57+58}{2} = 57.5$ s. (b) The median rose by $57.5 - 43 = 14.5$ s. (c) Spread hardly changed: the ranges are close ($50 - 38 = 12$ s before, $63 - 52 = 11$ s after). Because times *increased* and a lower time is faster, the runners actually became **slower** after the program, so the claim that "training made everyone faster" is not supported — the data show the opposite.

Q16. (a) Both have $n = 10$, so the median is the mean of the 5th and 6th values. Plot A: median $= \frac{76+76}{2} = 76$ g, range $= 105 - 71 = 34$ g. Plot B: median $= \frac{83+85}{2} = 84$ g, range $= 108 - 60 = 48$ g. (b) Plot A clusters at 71–79 g, then has a gap and a single high outlier at 105 g. Plot B is roughly symmetric about its centre, with no gaps or outliers. (c) Plot B grew typically heavier tomatoes (median 84 vs 76 g) and was more variable (range 48 vs 34 g), although Plot A's range is inflated by its 105 g outlier — without it, Plot A's masses run only 71–79 g.

Q17. Two reasons: (1) the mean says nothing about **spread**, so a hospital with steady 20-minute waits and one with waits swinging from 5 to 60 minutes could report the same mean; (2) the mean is **distorted by outliers**, so a few very long waits can inflate it above the typical experience. A fairer comparison would add the **median** (or the **range**) and show **parallel dot plots** so both centre and spread are visible.

Q18. (a) Brand P ordered $79, 83, 85, 87, 88, 89, 90, 91, 92, 94$: median $= \frac{88+89}{2} = 88.5$ h. Brand Q ordered $82, 86, 88, 90, 93, 95, 97, 99, 102, 105$: median $= \frac{93+95}{2} = 94$ h. (b) P: range $= 94 - 79 = 15$ h. Q: range $= 105 - 82 = 23$ h. (c) The shopper who wants the longest life should choose **Brand Q** — its median (94 h) is higher and it reaches 105 h. The hospital wanting predictable life should choose **Brand P** — its life is far more tightly grouped (range 15 vs 23 h), so it is less likely to fail unexpectedly early.

Theory

When we collect data we often want a single value that best represents the whole set — a **typical** or **average** value. A number that describes the middle of a data set is called a **measure of centre**. There are three in common use: the **mean**, the **median** and the **mode**. Each finds the "middle" in a different way, and part of working with data is choosing which one describes a set most fairly.

The mean.

The **mean** (the everyday "average") is found by adding all the values and dividing by how many there are:

$$\bar{x} = \frac{\text{sum of all data values}}{\text{number of data values}}.$$

The symbol $\bar{x}$ is read "x-bar". The mean uses **every** value, so it is a good summary when the data has no unusually large or small values.

The median.

The **median** is the **middle value** when the data is written in order from smallest to largest. To find its position in a set of n values, use

$$\text{median position} = \frac{n+1}{2}.$$

- If n is **odd**, this gives a whole number and the median is that single middle value.
- If n is **even**, the position falls halfway between two values, and the median is the **average of the two middle values**.

Because the median depends only on the middle of the ordered list, one unusually large or small value barely moves it.

The mode.

The **mode** is the value that occurs **most often**. A set can have **one** mode, **more than one** mode (if two or more values tie for the highest frequency), or **no mode** at all (if every value occurs the same number of times). The mode is the only average that can be used for **categorical** data such as sizes, colours or brands.

Centre from a frequency table.

When data is organised in a **frequency table**, we do not need to write every value out. For the values x with frequencies f:

$$\bar{x} = \frac{\sum(f \times x)}{\sum f} = \frac{\text{sum of (value} \times \text{frequency)}}{\text{total frequency}}.$$

The **mode** is the value with the **highest frequency**. The **median** is found by using its position $\frac{n+1}{2}$ (where $n = \sum f$) and reading down a **cumulative frequency** (running total) to see which value that position lands on.

The effect of an outlier.

An **outlier** is a value that is much larger or much smaller than the rest. An outlier **pulls the mean** towards it, but has **almost no effect on the median**. For example, in a small business where most staff earn a modest wage but the owner earns a great deal, the mean wage can be higher than nearly everyone actually earns, while the median wage still describes a typical worker. When a data set has an outlier or is lop-sided (skewed), the **median is usually the fairer measure of centre**.

Choosing a measure of centre.

Measure	Uses	Affected by an outlier?	Best when
Mean	every value	Yes	the data is fairly even, with no outliers
Median	the middle value(s)	No (resistant)	the data has an outlier or is skewed
Mode	the most common value	No	the data is categorical, or you want the most common value

Always state the **units** with a measure of centre — for example "the median wait was 6 minutes".

Worked examples

Example 1 — scaffolded

A café recorded the number of takeaway coffees sold in the first hour on 7 mornings:

$$18, 22, 19, 25, 22, 20, 22.$$

Find the mean, the median and the mode.

Working	Comment
$\bar{x} = \frac{18+22+19+25+22+20+22}{7} = \frac{148}{7} \approx 21.1$ coffees.	Mean = sum ÷ number of values ($n = 7$).
In order: $18, 19, 20, 22, 22, 22, 25$.	Always order the data before finding the median.
Median position $= \frac{7+1}{2} = 4$, so the median is the 4th value: 22 coffees.	Odd n: the median is the single middle value.
22 occurs three times, more than any other value, so the mode is 22 coffees.	Mode = most frequent value.

Example 2 — scaffolded

A class of 30 students was surveyed on the number of pets they own. The results are in the frequency table. Find the mean, the median and the mode.

Pets (x)	Frequency (f)	$f \times x$	Cumulative frequency
0	6	0	6
1	11	11	17
2	8	16	25
3	3	9	28
4	2	8	30

Working	Comment
$\sum f = 30$ and $\sum (f \times x) = 0+11+16+9+8 = 44$.	Total frequency and the sum of value × frequency.
$\bar{x} = \frac{44}{30} \approx 1.47$ pets.	Mean $= \frac{\sum (f \times x)}{\sum f}$.
Median position $= \frac{30+1}{2} = 15.5$, so average the 15th	Even total: the median sits between two values.

Working	Comment
and 16th values.	
Cumulative frequency reaches 6 at “0” and 17 at “1”, so the 15th and 16th values are both 1: median $=1$ pet.	Read down the running total to locate positions 15 and 16.
The highest frequency is 11 (at 1 pet), so the mode is 1 pet.	Mode = value with the greatest frequency.

Example 3 — unscaffolded

The ages of the 10 workers in a small team are

$$22, 24, 27, 29, 31, 34, 38, 41, 45, 49.$$

Find the mean, the median and the mode.

The data is already in order, with $n=10$ (even). The mean is

$$\bar{x}=\frac{22+24+27+29+31+34+38+41+45+49}{10}=\frac{340}{10}=34\text{ years}.$$

The median position is $\frac{10+1}{2}=5.5$, so the median is the average of the 5th and 6th values:

$$\text{median}=\frac{31+34}{2}=32.5\text{ years}.$$

Every age occurs exactly once, so there is **no mode**.

∴ the mean age is 34 years, the median age is 32.5 years, and there is no mode.

Example 4 — unscaffolded

A business has 7 employees. Their annual salaries, in thousands of dollars, are

$$52, 55, 58, 60, 62, 65, 180.$$

Find the mean and the median salary, and state which is the fairer measure of a typical salary.

There are $n=7$ values, already in order. The mean is

$$\bar{x}=\frac{52+55+58+60+62+65+180}{7}=\frac{532}{7}=76\left(\text{thousand dollars}\right).$$

The median is the 4th value ($7+1/2=4$): median $=60$ (thousand dollars). The one large salary of $\$180\,000$ is an **outlier**; it pulls the mean up to $\$76\,000$, which is higher than what six of the seven employees actually earn. The median of $\$60\,000$ is not affected by that one salary.

∴ the mean is $\$76\,000$ and the median is $\$60\,000$; the **median** is the fairer measure of a typical salary because it is not distorted by the outlier.

Practice questions

Scaffolded

Q1. The maximum daily temperatures (°C) recorded over 9 days were $24, 27, 22, 30, 25, 27, 23, 27, 26$. (a) Find the mean, correct to one decimal place. (b) Find the median. (c) Find the mode.

Q2. The number of goals scored by a soccer team in each of 20 games is shown below.

Goals (x)	Frequency (f)
0	3
1	5
2	6
3	4
4	2

(a) Find the mean number of goals.
(b) Find the median number of goals.
(c) Find the mode.

Q3. Eight students scored these marks out of 25 on a quiz: $12, 15, 18, 15, 20, 15, 11, 14$. (a) Find the mean. (b) Find the median. (c) Find the mode.

Q4. Six houses in a street sold for the following prices, in thousands of dollars: $420, 445, 460, 480, 505, 1200$. (a) Find the mean sale price. (b) Find the median sale price. (c) State which measure better represents a typical sale price, and why.

Q5. A shoe shop recorded the sizes sold during one hour: $7, 8, 8, 9, 9, 9, 10, 10, 11$. (a) Find the mode. (b) Explain why the mode is the most useful average for the shop when it reorders stock. (c) Find the median size.

Q6. Five students sat a test. The mean of their five marks is 68. Four of the marks are $60, 65, 70$ and 75. (a) Find the total of all five marks. (b) Find the fifth mark. (c) Find the median of all five marks.

Unscaffolded

Q7. The delivery times (minutes) for 11 online orders were $25, 28, 30, 28, 35, 40, 28, 32, 45, 30, 38$. Find the mean (to one decimal place), the median and the mode.

Q8. A survey counted the number of people in each of 40 cars passing a school.

People (x)	Frequency (f)
1	18
2	12
3	6
4	3
5	1

Find the mean (to three decimal places), the median and the mode.

Q9. A café recorded the number of customers served in each of 8 half-hour slots: $15, 18, 20, 21, 22, 24, 25, 75$. (a) Find the mean and the median. (b) The busy slot of 75 customers is removed. Find the mean and the median of the remaining seven values. (c) State which measure — the mean or the median — was changed more by removing that value, and explain why.

Q10. The annual salaries of the 7 staff at a start-up, in thousands of dollars, are $48, 52, 55, 58, 60, 62, 240$. (a) Find the mean and the median salary. (b) The company advertises “a typical salary”. State which measure it should quote to fairly describe a typical worker, and why.

Q11. The mean of six numbers is 15. Five of the numbers are $12, 14, 16, 18$ and 11. Find the sixth number.

Q12. Over 8 days a shop sold a mean of 24 jackets per day. On the 9th day it sold 42 jackets. Find the mean number of jackets sold per day over all 9 days.

Q13. The frequency table shows the number of siblings of the students in a class. One frequency is missing.

Siblings (x)	Frequency (f)
0	4
1	7
2	?
3	1

The mean number of siblings is 1.3. Find the missing frequency.

Q14. A clothing store sold shirts in the following sizes during a sale: size S (12 shirts), size M (20), size L (15) and size XL (8). (a) State the modal size. (b) Explain why the mean is not an appropriate average for this data.

Q15. The numbers of bicycles hired from a shop on each of 14 days were $8, 10, 11, 11, 12, 13, 15, 15, 15, 16, 18, 20, 22, 30$. Find the mean (to one decimal place), the median and the mode.

Q16. A set of five numbers, written in order, is $8, 11, x, 17, 21$. The median of the set is 14. (a) Find the value of x. (b) Find the mean of the five numbers.

Q17. The ages of the 9 members of a lawn-bowls social group are $18, 19, 19, 20, 21, 22, 24, 55, 60$. (a) Find the mean, the median and the mode. (b) The group wants to advertise a “typical member’s age” to attract new members. State which measure best describes the group, and explain why.

Q18. A survey of 25 people recorded how many nights each week they exercise.

Nights (x)	Frequency (f)
0	2
1	3
2	5
3	7
4	4
5	3
6	1

Find the mean (to two decimal places), the median and the mode, and write one sentence interpreting the median in this context.

Answers

Q1. (a) 25.7 °C. (b) 26 °C. (c) 27 °C. **Q2.** (a) 1.85 goals. (b) 2 goals. (c) 2 goals. **Q3.** (a) 15 marks. (b) 15 marks. (c) 15 marks. **Q4.** (a) 585 (thousand dollars). (b) 470 (thousand dollars). (c) The median; the $\$1\,200\,000$ sale is an outlier that inflates the mean. **Q5.** (a) 9. (b) size 9 is sold most often, so it should be restocked most. (c) 9. **Q6.** (a) 340. (b) 70. (c) 70. **Q7.** mean 32.6 min, median 30 min, mode 28 min. **Q8.** mean 1.925 people, median 2 people, mode 1 person. **Q9.** (a) mean 27.5, median 21.5. (b) mean ≈ 20.7, median 21. (c) the mean changed far more (from 27.5 to 20.7); the outlier 75 pulls the mean but barely moves the median. **Q10.** (a) mean ≈ 82.1, median 58 (thousand dollars). (b) the median ($\$58\,000$); the $\$240\,000$ salary is an outlier that makes the mean unrepresentative. **Q11.** 19. **Q12.** 26 jackets. **Q13.** 8. **Q14.** (a) size M. (b) shirt sizes are categorical labels, so adding them and dividing has no meaning. **Q15.** mean 15.4, median 15, mode 15. **Q16.** (a) $x = 14$. (b) 14.2. **Q17.** (a) mean ≈ 28.7, median 21, mode 19. (b) the median (or mode); most members are around 19–21, and the two older members pull the mean up. **Q18.** mean 2.84 nights, median 3 nights, mode 3 nights; half the people surveyed exercise on 3 or more nights each week.

Fully-worked solutions

Q1. In order: $22, 23, 24, 25, 26, 27, 27, 27, 30$, with $n = 9$. (a) $\bar{x} = \frac{24+27+22+30+25+27+23+27+26}{9} = \frac{231}{9} \approx 25.7$ °C. (b) Median position $\frac{9+1}{2} = 5$, so the median is the 5th value, 26 °C. (c) 27 occurs three times, more than any other value, so the mode is 27 °C.

Q2. Total frequency $\sum f = 3+5+6+4+2 = 20$. (a) $\sum(f \times x) = 0(3)+1(5)+2(6)+3(4)+4(2) = 0+5+12+12+8 = 37$, so $\bar{x} = \frac{37}{20} = 1.85$ goals. (b) Median position $\frac{20+1}{2} = 10.5$, the average of the 10th and 11th values. The cumulative frequency is 3 (at 0), 8 (at 1), 14 (at 2), so the 10th and 11th values are both 2: median $= 2$ goals. (c) The highest frequency is 6, at 2 goals, so the mode is 2 goals.

Q3. In order: $11, 12, 14, 15, 15, 15, 18, 20$, with $n = 8$. (a) $\bar{x} = \frac{12+15+18+15+20+15+11+14}{8} = \frac{120}{8} = 15$ marks. (b) Median position $\frac{8+1}{2} = 4.5$, so the median is $\frac{15+15}{2} = 15$ marks. (c) 15 occurs three times, so the mode is 15 marks.

Q4. The prices (thousands of dollars) are $420, 445, 460, 480, 505, 1200$, with $n = 6$. (a) $\bar{x} = \frac{420+445+460+480+505+1200}{6} = \frac{3510}{6} = 585$. (b) Median position $\frac{6+1}{2} = 3.5$, so the median is $\frac{460+480}{2} = 470$. (c) The 1200 (that is, $\$1\,200\,000$) is an outlier that pulls the mean up to 585, well above five of the six prices. The median of 470 better represents a typical sale.

Q5. The sizes in order are $7, 8, 8, 9, 9, 9, 10, 10, 11$, with $n = 9$. (a) Size 9 occurs three times, more than any other, so the mode is 9. (b) When reordering, the shop wants the size it sells most often so it does not run out; that is exactly the mode, size 9. (c) Median position $\frac{9+1}{2} = 5$, so the median is the 5th value, 9.

Q6. (a) The mean of five marks is 68, so the total is $68 \times 5 = 340$. (b) The four known marks total $60+65+70+75 = 270$, so the fifth mark is $340 - 270 = 70$. (c) The five marks in order are $60, 65, 70, 70, 75$; the median is the 3rd value, 70.

Q7. In order: $25, 28, 28, 28, 30, 30, 32, 35, 38, 40, 45$, with $n = 11$. The mean is $\frac{25+28+30+28+35+40+28+32+45+30+38}{11} = \frac{359}{11} \approx 32.6$ min. Median position $\frac{11+1}{2} = 6$, so the median is the 6th value, 30 min. The value 28 occurs three times, so the mode is 28 min.

Q8. Total frequency $\sum f = 18+12+6+3+1=40$. $\sum(f\times x)=1(18)+2(12)+3(6)+4(3)+5(1)=18+24+18+12+5=77$, so $\bar{x}=\frac{77}{40}=1.925$ people. Median position $\frac{40+1}{2}=20.5$, the average of the 20th and 21st values. The cumulative frequency is 18 (at 1) then 30 (at 2), so both the 20th and 21st values are 2: median $=2$ people. The highest frequency is 18, at 1 person, so the mode is 1 person.

Q9. (a) With all eight values $15,18,20,21,22,24,25,75$ ($n=8$): $\bar{x}=\frac{220}{8}=27.5$ customers, and the median is $\frac{21+22}{2}=21.5$ customers. (b) Removing 75 leaves $15,18,20,21,22,24,25$ ($n=7$): $\bar{x}=\frac{145}{7}\approx 20.7$ customers, and the median is the 4th value, 21 customers. (c) The mean fell sharply, from 27.5 to about 20.7, while the median hardly moved (21.5 to 21). The value 75 is an outlier: it pulls the mean strongly but has little effect on the median.

Q10. The salaries (thousands of dollars) are $48,52,55,58,60,62,240$, with $n=7$. (a) $\bar{x}=\frac{48+52+55+58+60+62+240}{7}=\frac{575}{7}\approx 82.1$; median position $\frac{7+1}{2}=4$, so the median is the 4th value, 58. (b) The $\$240\,000$ salary is an outlier that lifts the mean to about $\$82\,000$, higher than six of the seven staff earn. The company should quote the **median**, $\$58\,000$, to describe a typical worker fairly.

Q11. A mean of 15 over six numbers means the total is $15\times 6=90$. The five known numbers total $12+14+16+18+11=71$, so the sixth number is $90-71=19$.

Q12. Over 8 days at a mean of 24, the total sold is $24\times 8=192$. Adding the 9th day: $192+42=234$ over 9 days, so the new mean is $\frac{234}{9}=26$ jackets per day.

Q13. Let the missing frequency be x. The total frequency is $4+7+x+1=12+x$, and $\sum(f\times x\text{-value})=0(4)+1(7)+2(x)+3(1)=10+2x$. The mean is 1.3, so

$$\frac{10+2x}{12+x}=1.3.$$

Then $10+2x=1.3(12+x)=15.6+1.3x$, giving $0.7x=5.6$, so $x=8$. The missing frequency is 8.

Q14. (a) The greatest frequency is 20, for size M, so the modal size is M. (b) The "values" S, M, L and XL are **categories**, not numbers. There is no meaningful way to add them and divide, so a mean cannot be calculated; the mode is the appropriate average.

Q15. The data $8,10,11,11,12,13,15,15,15,16,18,20,22,30$ has $n=14$. The mean is $\frac{216}{14}\approx 15.4$ bicycles. Median position $\frac{14+1}{2}=7.5$, so the median is $\frac{15+15}{2}=15$ bicycles. The value 15 occurs three times, so the mode is 15 bicycles.

Q16. (a) With five numbers in order, the median is the 3rd value, which is x. Since the median is 14, $x=14$. (b) The five numbers are $8,11,14,17,21$, so $\bar{x}=\frac{8+11+14+17+21}{5}=\frac{71}{5}=14.2$.

Q17. The ages $18,19,19,20,21,22,24,55,60$ have $n=9$. (a) $\bar{x}=\frac{18+19+19+20+21+22+24+55+60}{9}=\frac{258}{9}\approx 28.7$ years; median position $\frac{9+1}{2}=5$, so the median is the 5th value, 21 years; 19 occurs twice, so the mode is 19 years. (b) Seven of the nine members are aged between 18 and 24 ; the two members aged 55 and 60 pull the mean up to about 28.7 years, which overstates a typical age. The median (21) or mode (19) better describes the mostly-young group.

Q18. Total frequency $\sum f = 2+3+5+7+4+3+1 = 25$.
$\sum(f \times x) = 0(2)+1(3)+2(5)+3(7)+4(4)+5(3)+6(1) = 0+3+10+21+16+15+6 = 71$, so $\overline{x} = \frac{71}{25} = 2.84$ nights. Median position $\frac{25+1}{2} = 13$, the 13th value. The cumulative frequency is $2, 5, 10, 17, \ldots$, so the 13th value falls in "3": median $= 3$ nights. The highest frequency is 7, at 3 nights, so the mode is 3 nights. Interpretation: half of the people surveyed exercise on 3 or more nights each week.

Chapter 6 Summarising and comparing data — 6F Measures of spread (range, quartiles)

Theory

A **measure of centre** — the mean, median or mode — tells you where the middle of a data set sits. But two data sets can share the same centre and still look completely different. Suppose two cafés each sell a median of 30 coffees an hour: one is steady all day, the other swings between quiet mornings and frantic lunch rushes. To describe *how spread out* the values are, we use a **measure of spread**. In Units 1 & 2 you use two: the **range** and the **interquartile range (IQR)**.

The range.

The **range** is the simplest measure of spread:

$$\text{range} = \text{maximum value} - \text{minimum value}.$$

It is quick to find, but it uses only the two most extreme values, so a single unusually large or small value (an **outlier**) can make it misleading.

Quartiles and the median-of-halves method.

The **quartiles** divide an ordered data set into four equal parts. The **median** is the middle value — the second quartile. The **lower quartile** Q_1 is the median of the lower half of the data, and the **upper quartile** Q_3 is the median of the upper half.

To find the quartiles, order the data and use the **median-of-halves method**:

1. Order the data from smallest to largest and find the **median**.
2. If there is an **odd** number of values, leave the middle value out: the lower half is everything below it, the upper half everything above it.
3. If there is an **even** number of values, split the data into two equal halves.
4. Q_1 is the median of the lower half; Q_3 is the median of the upper half.

The interquartile range.

The **interquartile range** is the spread of the middle 50 % of the data:

$$\text{IQR} = Q_3 - Q_1.$$

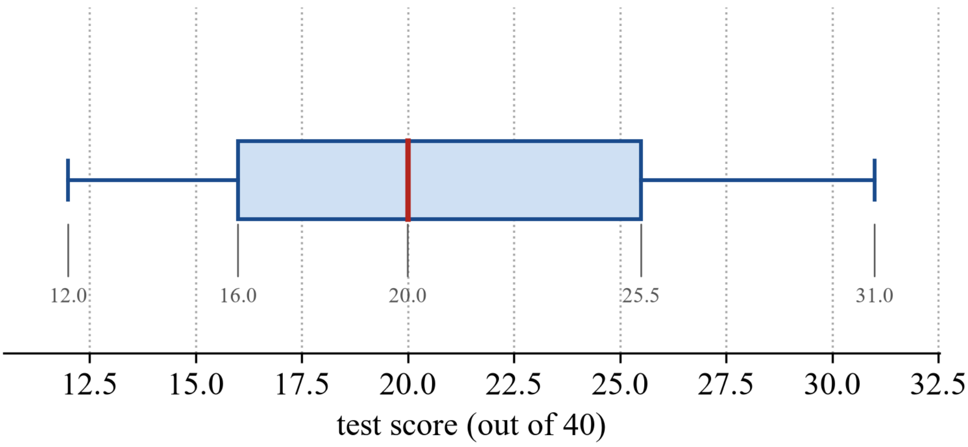
Because the IQR ignores the smallest 25 % and the largest 25 % of the values, one extreme value barely changes it — so the IQR is a far more reliable measure of spread than the range when a data set contains an outlier.

The five-number summary and the boxplot.

The five values

$$\text{minimum}, Q_1, \text{median}, Q_3, \text{maximum}$$

are the **five-number summary**. They are displayed with a **boxplot** (box-and-whisker plot): a box from Q_1 to Q_3 with the median marked inside it, and a whisker reaching out to the minimum and to the maximum. The length of the box is the IQR, and the distance from one whisker tip to the other is the range.



This boxplot summarises nine test scores. You can read the five-number summary straight off it: minimum 12, $Q_1 = 16$, median 20, $Q_3 = 25.5$ and maximum 31, giving a range of 19 and an IQR of 9.5.

Outliers.

A value that is clearly separated from the bulk of the data — well above or well below all the other values — is called an **outlier**. In Units 1 & 2 you identify outliers informally: look for a value that plainly does not belong with the rest.

Extension — beyond Units 1 & 2. *The formal rule below is beyond the Foundation Mathematics Units 1 & 2 study design — spread at this level stops at the range, the quartiles and the IQR. The rule belongs to General Mathematics, and is included here as a preview because later questions in this book use it.* The rule treats a value as an **outlier** if it lies more than $1.5 \times \text{IQR}$ below Q_1 or above Q_3 — that is, outside the **fences**

$$Q_1 - 1.5 \times \text{IQR} \text{ and } Q_3 + 1.5 \times \text{IQR}\,.$$

On a boxplot an outlier is shown as a separate point, with the whisker stopping at the most extreme value that is *not* an outlier.

Worked examples

Example 1 — scaffolded

Over nine days a rain gauge recorded, in millimetres: $8, 3, 12, 5, 3, 9, 15, 6, 11$. Find the range, the median, the quartiles Q_1 and Q_3, and the interquartile range.

Working	Comment
Order the data: $3, 3, 5, 6, 8, 9, 11, 12, 15$.	Always order first; there are $n = 9$ values.
Range $= 15 - 3 = 12$ mm.	Maximum − minimum.
Median $= 8$ mm (the 5th value).	The middle value of nine.
Lower half: $3, 3, 5, 6$; so $Q_1 = \dfrac{3+5}{2} = 4$ mm.	Odd n: leave the median out. Median of four is the mean of the middle two.
Upper half: $9, 11, 12, 15$; so $Q_3 = \dfrac{11+12}{2} = 11.5$ mm.	Median of the upper half.
$\text{IQR} = Q_3 - Q_1 = 11.5 - 4 = 7.5$ mm.	Spread of the middle 50 %.

Example 2 — scaffolded

Ten parcels have masses (kg): $2.1, 3.5, 1.8, 4.2, 2.9, 3.1, 5.0, 2.4, 3.8, 4.6$. Write the five-number summary and find the IQR.

Working	Comment
Order the data:	$n = 10$ (even).

Working	Comment
$1.8, 2.1, 2.4, 2.9, 3.1, 3.5, 3.8, 4.2, 4.6, 5.0$.	
Median $= \dfrac{3.1+3.5}{2} = 3.3$ kg.	Mean of the 5th and 6th values.
Lower half: $1.8, 2.1, 2.4, 2.9, 3.1$; $Q_1 = 2.4$ kg.	Even n: split into two halves of five. Median of five is the middle value.
Upper half: $3.5, 3.8, 4.2, 4.6, 5.0$; $Q_3 = 4.2$ kg.	Median of the upper half.
Five-number summary: $1.8, 2.4, 3.3, 4.2, 5.0$.	min, Q_1, median, Q_3, max.
$\text{IQR} = 4.2 - 2.4 = 1.8$ kg.	$Q_3 - Q_1$.

Example 3 — unscaffolded

A casual worker's earnings over eleven weeks, in dollars, were $180, 195, 210, 205, 190, 200, 185, 215, 220, 198, 640$ (the last week included a large one-off bonus). Find the range and the IQR, draw a boxplot, and state which measure better describes the typical spread of the weekly pay.

Ordered: $180, 185, 190, 195, 198, 200, 205, 210, 215, 220, 640$ ($n = 11$).

Median $= 200$ (the 6th value).

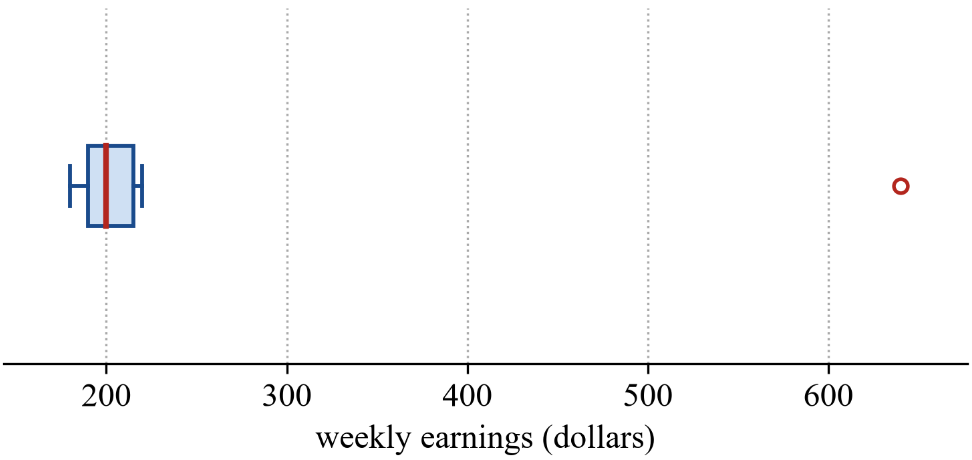
Lower half $180, 185, 190, 195, 198$, so $Q_1 = 190$.

Upper half $205, 210, 215, 220, 640$, so $Q_3 = 215$.

Range $= 640 - 180 = 460$.

$\text{IQR} = 215 - 190 = 25$.

(Extension) Upper fence $= Q_3 + 1.5 \times \text{IQR} = 215 + 1.5 \times 25 = 252.5$, and $\$640$ lies beyond it, so $\$640$ is an outlier (whisker stops at $\$220$).



The range of $\$460$ is inflated by the single bonus week, while the IQR of $\$25$ describes the pay in a normal week.

$\therefore$ the IQR better describes the typical spread, because it is not distorted by the outlier.

Example 4 — unscaffolded

Eight commuters recorded their morning travel time in minutes: $24, 31, 28, 45, 33, 29, 38, 26$. Write the five-number summary, find the range and the IQR, and draw a boxplot.

Ordered: $24, 26, 28, 29, 31, 33, 38, 45$ ($n = 8$).

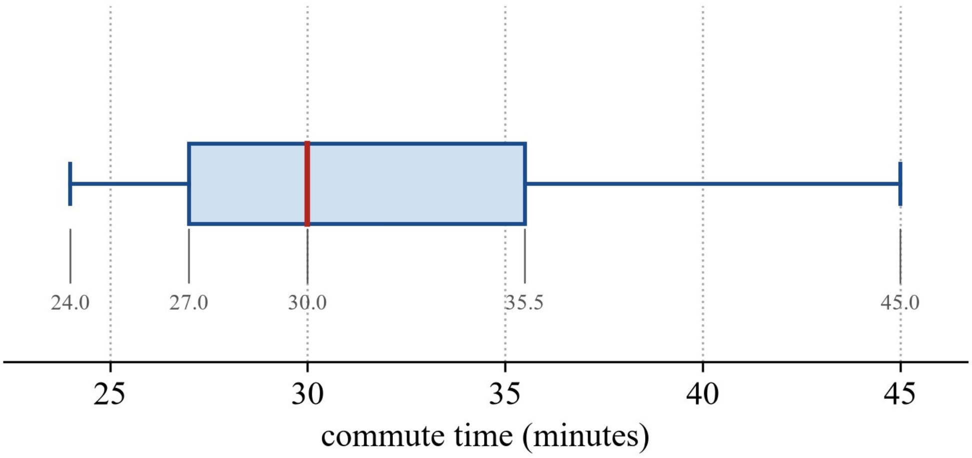
Median $= \dfrac{29+31}{2} = 30$.

Lower half $24, 26, 28, 29$, so $Q_1 = \dfrac{26+28}{2} = 27$.

Upper half $31, 33, 38, 45$, so $Q_3 = \dfrac{33+38}{2} = 35.5$.

Five-number summary: $24, 27, 30, 35.5, 45$.

Range $= 45 - 24 = 21$, and $\text{IQR} = 35.5 - 27 = 8.5$.



$\therefore$ the middle half of commuters travel within an 8.5-minute band, while the full range of travel times is 21 minutes.

Practice questions

Scaffolded

Q1. The ages of seven junior players are $14, 15, 15, 16, 17, 18, 20$. (a) Find the range. (b) Find the median. (c) Find Q_1 and Q_3.

Q2. Eight items cost, in dollars: $3.20, 4.50, 2.80, 5.10, 3.90, 4.20, 2.50, 4.80$. (a) Write the data in order. (b) Write the five-number summary. (c) Find the IQR.

Q3. Nine daily maximum temperatures (°C) were $22, 19, 25, 18, 24, 21, 27, 20, 23$. (a) Find the range. (b) Find the median. (c) Find the IQR.

Q4. A student's daily step count (in thousands) over ten days was $6, 8, 5, 9, 7, 11, 4, 10, 8, 12$. (a) Find the median. (b) Find Q_1 and Q_3. (c) Write the five-number summary.

Q5. In nine netball games a team scored $2, 3, 3, 4, 5, 5, 6, 7, 20$ goals. (a) Write the five-number summary. (b) Find the IQR. (c) *(Extension)* Using the upper fence $Q_3 + 1.5 \times \text{IQR}$, decide whether 20 is an outlier.

Q6. Eight employees earn, in thousands of dollars, $48, 52, 55, 58, 60, 62, 65, 120$. (a) Find the range. (b) Find the IQR. (c) State which measure better describes the typical spread of these salaries, and why.

Unscaffolded

Q7. Over eleven days a town's rainfall (mm) was $0, 2, 3, 5, 5, 6, 8, 9, 10, 12, 15$. Write the five-number summary, and find the range and the IQR.

Q8. Twelve customers waited, in minutes: $22, 25, 25, 28, 30, 31, 33, 34, 36, 38, 40, 45$. Find the median, Q_1, Q_3 and the IQR.

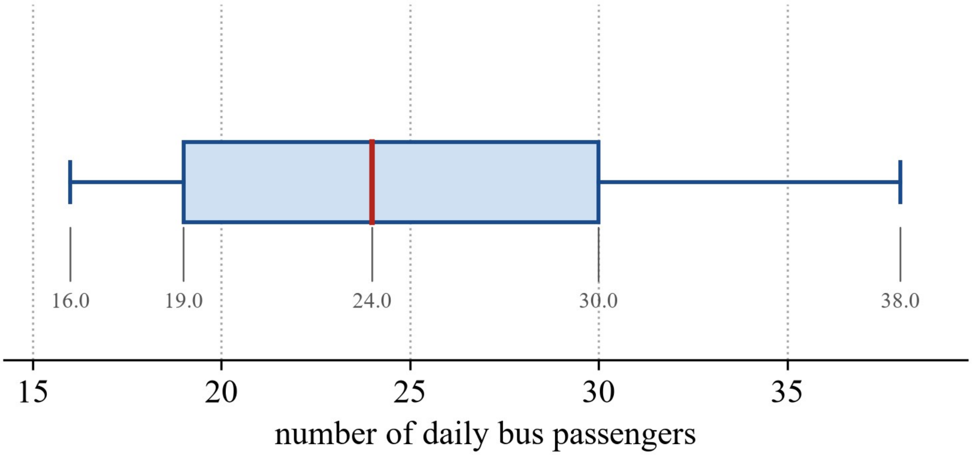
Q9. Two machines fill bags of flour. The masses (grams above the labelled weight) of seven bags from each are Machine A: $46, 48, 51, 53, 55, 57, 60$; Machine B: $40, 49, 52, 54, 56, 58, 66$. (a) Find the range and the IQR for each machine. (b) The two machines have the same IQR but different ranges. Explain what this tells you about how the two machines compare.

Q10. A striker's goals in nine matches were $1,2,2,3,3,4,4,5,14$. (a) Find Q_1, Q_3 and the IQR. (b) *(Extension)* Show that 14 is an outlier using the upper fence.

Q11. Eleven students have heights (cm) $150,152,155,158,160,162,165,168,170,172,180$. Write the five-number summary and find the range and the IQR.

Q12. Twelve exam marks (out of 80) were $35,41,44,48,52,55,58,60,63,67,70,78$. Find the median, the quartiles and the IQR.

Q13. The boxplot below shows the number of passengers on a morning bus route over ten days.



(a) Write the five-number summary.
(b) Find the range.
(c) Find the IQR.

Q14. A shop's nine daily online orders had values (dollars) $12,14,15,15,16,18,19,20,60$. (a) Find the median, Q_1 and Q_3. (b) Find the range and the IQR. (c) *(Extension)* Determine whether $\$60$ is an outlier.

Q15. Ten months of rainfall (mm) were $3,4,4,6,7,9,10,12,13,16$. Write the five-number summary and find the IQR.

Q16. Two towns record their midday temperature (°C) on seven days: Town X: $27,28,29,30,31,32,33$; Town Y: $18,22,26,30,34,38,42$. (a) Show that the two towns have the same median. (b) Find the range and the IQR of each town. (c) Which town has the more consistent temperature? Justify your answer using the IQR.

Q17. Thirteen delivery times (minutes) were $30,32,33,35,36,38,39,40,41,42,44,45,90$. (a) Find the median, Q_1, Q_3 and the IQR. (b) *(Extension)* Using the fences, identify any outlier.

Q18. Eight rainfall readings (mm) were $5.2,6.1,6.8,7.0,7.5,8.2,8.9,9.4$. Write the five-number summary, and find the range and the IQR.

Answers

Q1. (a) 6. (b) 16. (c) $Q_1 = 15$, $Q_3 = 18$. **Q2.** (a) $2.50, 2.80, 3.20, 3.90, 4.20, 4.50, 4.80, 5.10$. (b) $2.50, 3.00, 4.05, 4.65, 5.10$. (c) IQR $= 1.65$. **Q3.** (a) 9. (b) 22. (c) IQR $= 5$. **Q4.** (a) 8. (b) $Q_1 = 6$, $Q_3 = 10$. (c) $4, 6, 8, 10, 12$. **Q5.** (a) $2, 3, 5, 6.5, 20$. (b) IQR $= 3.5$. (c) *(Extension)* upper fence $= 11.75$; $20 > 11.75$, so 20 is an outlier. **Q6.** (a) range $= 72$. (b) IQR $= 10$. (c) the IQR, because the salary of $\$120\,000$ is an outlier that inflates the range. **Q7.** $0, 3, 6, 10, 15$; range $= 15$, IQR $= 7$. **Q8.** median $= 32$, $Q_1 = 26.5$, $Q_3 = 37$, IQR $= 10.5$. **Q9.** (a) A: range 14, IQR 9; B: range 26, IQR 9. (b) the middle 50% of both machines is equally spread (same IQR), but Machine B has more extreme individual bags (larger range). **Q10.** (a) $Q_1 = 2$, $Q_3 = 4.5$, IQR $= 2.5$. (b) *(Extension)* upper fence $= 4.5 + 1.5 \times 2.5 = 8.25$; $14 > 8.25$, so 14 is an outlier. **Q11.** $150, 155, 162, 170, 180$; range $= 30$, IQR $= 15$. **Q12.** median $= 56.5$, $Q_1 = 46$, $Q_3 = 65$, IQR $= 19$. **Q13.** (a) $16, 19, 24, 30, 38$. (b) range $= 22$. (c) IQR $= 11$. **Q14.** (a) median $= 16$, $Q_1 = 14.5$, $Q_3 = 19.5$. (b) range $= 48$, IQR $= 5$. (c) *(Extension)* upper fence $= 27$; $\$60 > \$27$, so it is an outlier. **Q15.** $3, 4, 8, 12, 16$; IQR $= 8$. **Q16.** (a) both have median 30°C. (b) X: range 6, IQR 4; Y: range 24, IQR 16. (c) Town X, because its IQR is much smaller (4 vs 16), so its temperatures are far more consistent. **Q17.** (a) median $= 39$, $Q_1 = 34$, $Q_3 = 43$, IQR $= 9$. (b) *(Extension)* upper fence $= 43 + 1.5 \times 9 = 56.5$; $90 > 56.5$, so 90 is an outlier. **Q18.** $5.2, 6.45, 7.25, 8.55, 9.4$; range $= 4.2$, IQR $= 2.1$.

Fully-worked solutions

Q1. Ordered: $14, 15, 15, 16, 17, 18, 20$ ($n = 7$). (a) Range $= 20 - 14 = 6$. (b) Median $= 16$ (the 4th value). (c) Lower half $14, 15, 15$, so $Q_1 = 15$; upper half $17, 18, 20$, so $Q_3 = 18$.

Q2. (a) In order: $2.50, 2.80, 3.20, 3.90, 4.20, 4.50, 4.80, 5.10$ ($n = 8$). (b) Median $= \dfrac{3.90 + 4.20}{2} = 4.05$. Lower half $2.50, 2.80, 3.20, 3.90$, so $Q_1 = \dfrac{2.80 + 3.20}{2} = 3.00$; upper half $4.20, 4.50, 4.80, 5.10$, so $Q_3 = \dfrac{4.50 + 4.80}{2} = 4.65$. Five-number summary $2.50, 3.00, 4.05, 4.65, 5.10$. (c) IQR $= 4.65 - 3.00 = 1.65$.

Q3. Ordered: $18, 19, 20, 21, 22, 23, 24, 25, 27$ ($n = 9$). (a) Range $= 27 - 18 = 9$. (b) Median $= 22$ (the 5th value). (c) Lower half $18, 19, 20, 21$, so $Q_1 = 19.5$; upper half $23, 24, 25, 27$, so $Q_3 = 24.5$. IQR $= 24.5 - 19.5 = 5$.

Q4. Ordered: $4, 5, 6, 7, 8, 8, 9, 10, 11, 12$ ($n = 10$). (a) Median $= \dfrac{8 + 8}{2} = 8$. (b) Lower half $4, 5, 6, 7, 8$, so $Q_1 = 6$; upper half $8, 9, 10, 11, 12$, so $Q_3 = 10$. (c) Five-number summary $4, 6, 8, 10, 12$.

Q5. Ordered: $2, 3, 3, 4, 5, 5, 6, 7, 20$ ($n = 9$). (a) Median $= 5$ (the 5th value). Lower half $2, 3, 3, 4$, so $Q_1 = 3$; upper half $5, 6, 7, 20$, so $Q_3 = \dfrac{6 + 7}{2} = 6.5$. Five-number summary $2, 3, 5, 6.5, 20$. (b) IQR $= 6.5 - 3 = 3.5$. (c) *(Extension)* Upper fence $= 6.5 + 1.5 \times 3.5 = 11.75$. Since $20 > 11.75$, the value 20 is an outlier.

Q6. Ordered: $48, 52, 55, 58, 60, 62, 65, 120$ ($n = 8$). (a) Range $= 120 - 48 = 72$. (b) Median $= \dfrac{58 + 60}{2} = 59$. Lower half $48, 52, 55, 58$, so $Q_1 = 53.5$; upper half $60, 62, 65, 120$, so $Q_3 = 63.5$. IQR $= 63.5 - 53.5 = 10$. (c) The IQR better describes the typical spread: the salary of $\$120\,000$ is far above the others and inflates the range, whereas the IQR reflects the middle 50% of earners.

Q7. Data already ordered ($n = 11$). Median $= 6$ (the 6th value). Lower half $0, 2, 3, 5, 5$, so $Q_1 = 3$; upper half $8, 9, 10, 12, 15$, so $Q_3 = 10$. Five-number summary $0, 3, 6, 10, 15$. Range $= 15 - 0 = 15$; IQR $= 10 - 3 = 7$.

Q8. Data ordered ($n = 12$). Median $= \dfrac{31 + 33}{2} = 32$. Lower half $22, 25, 25, 28, 30, 31$, so $Q_1 = \dfrac{25 + 28}{2} = 26.5$; upper half $33, 34, 36, 38, 40, 45$, so $Q_3 = \dfrac{36 + 38}{2} = 37$. IQR $= 37 - 26.5 = 10.5$.

Q9. (a) Machine A ($n=7$): median 53, $Q_1=48$, $Q_3=57$; range $=60-46=14$, IQR $=57-48=9$. Machine B ($n=7$): median 54, $Q_1=49$, $Q_3=58$; range $=66-40=26$, IQR $=58-49=9$. (b) The two machines fill the middle 50 % of bags with the same consistency (IQR $=9$ for both), but Machine B produces a few bags that are much lighter or heavier than Machine A's (its range is 26 versus 14), so B is less reliable at the extremes.

Q10. Ordered: $1,2,2,3,3,4,4,5,14$ ($n=9$). (a) Median $=3$. Lower half $1,2,2,3$, so $Q_1=2$; upper half $4,4,5,14$, so $Q_3=4.5$. IQR $=4.5-2=2.5$. (b) *(Extension)* Upper fence $=Q_3+1.5\times\text{IQR}=4.5+1.5\times2.5=8.25$. Since $14>8.25$, the value 14 is an outlier.

Q11. Data ordered ($n=11$). Median $=162$ (the 6th value). Lower half $150,152,155,158,160$, so $Q_1=155$; upper half $165,168,170,172,180$, so $Q_3=170$. Five-number summary $150,155,162,170,180$. Range $=180-150=30$; IQR $=170-155=15$.

Q12. Data ordered ($n=12$). Median $=\dfrac{55+58}{2}=56.5$. Lower half $35,41,44,48,52,55$, so $Q_1=\dfrac{44+48}{2}=46$; upper half $58,60,63,67,70,78$, so $Q_3=\dfrac{63+67}{2}=65$. IQR $=65-46=19$.

Q13. Reading the boxplot: the left whisker is at 16, the box runs from 19 to 30 with the median at 24, and the right whisker is at 38. (a) Five-number summary $16,19,24,30,38$. (b) Range $=38-16=22$. (c) IQR $=30-19=11$.

Q14. Ordered: $12,14,15,15,16,18,19,20,60$ ($n=9$). (a) Median $=16$ (the 5th value). Lower half $12,14,15,15$, so $Q_1=\dfrac{14+15}{2}=14.5$; upper half $18,19,20,60$, so $Q_3=\dfrac{19+20}{2}=19.5$. (b) Range $=60-12=48$; IQR $=19.5-14.5=5$. (c) *(Extension)* Upper fence $=19.5+1.5\times5=27$. Since $\$60>\$27$, the order of $\$60$ is an outlier.

Q15. Ordered ($n=10$). Median $=\dfrac{7+9}{2}=8$. Lower half $3,4,4,6,7$, so $Q_1=4$; upper half $9,10,12,13,16$, so $Q_3=12$. Five-number summary $3,4,8,12,16$. IQR $=12-4=8$.

Q16. (a) Town X ($n=7$): median $=30$ (the 4th value). Town Y ($n=7$): median $=30$ (the 4th value). Both medians are 30°C. (b) Town X: lower half $27,28,29$, so $Q_1=28$; upper half $31,32,33$, so $Q_3=32$; range $=33-27=6$, IQR $=32-28=4$. Town Y: lower half $18,22,26$, so $Q_1=22$; upper half $34,38,42$, so $Q_3=38$; range $=42-18=24$, IQR $=38-22=16$. (c) Town X is more consistent: although both towns have a median of 30°C, Town X's IQR is only 4°C compared with Town Y's 16°C, so Town X's temperatures are clustered much more tightly around the middle.

Q17. Ordered ($n=13$). (a) Median $=39$ (the 7th value). Lower half $30,32,33,35,36,38$, so $Q_1=\dfrac{33+35}{2}=34$; upper half $40,41,42,44,45,90$, so $Q_3=\dfrac{42+44}{2}=43$. IQR $=43-34=9$. (b) *(Extension)* Upper fence $=43+1.5\times9=56.5$; lower fence $=34-1.5\times9=20.5$. Since $90>56.5$, the delivery time of 90 minutes is an outlier (no value is below 20.5).

Q18. Data ordered ($n=8$). Median $=\dfrac{7.0+7.5}{2}=7.25$. Lower half $5.2,6.1,6.8,7.0$, so $Q_1=\dfrac{6.1+6.8}{2}=6.45$; upper half $7.5,8.2,8.9,9.4$, so $Q_3=\dfrac{8.2+8.9}{2}=8.55$. Five-number summary $5.2,6.45,7.25,8.55,9.4$. Range $=9.4-5.2=4.2$; IQR $=8.55-6.45=2.1$.

Topic 2 Test A — Multiple-choice

Data analysis, probability and statistics (Chapters 5–6). Approved materials: one bound reference that may be annotated; one scientific calculator. Section A: 15 multiple-choice questions (15 marks). Time: about 25 minutes. Unless otherwise indicated, diagrams are not drawn to scale.

Section A — Multiple-choice questions

Choose the response that is correct for the question.

Question 1

The colour of a car in a car park is an example of

A. categorical (nominal) data

B. categorical (ordinal) data

C. discrete numerical data

D. continuous numerical data

Question 2

The exact time, in seconds, that a runner takes to finish a race is an example of

A. categorical (nominal) data

B. categorical (ordinal) data

C. discrete numerical data

D. continuous numerical data

Question 3

A school is divided into its year levels and a number of students is chosen at random from each year level. This is an example of

A. convenience sampling

B. a census

C. systematic sampling

D. stratified sampling

Question 4

A television program asks viewers to phone in to vote on a question. The main problem with the resulting sample is that

A. the sample is stratified

B. it is actually a census

C. the sample is self-selected and not representative

D. the sample is too large

Question 5

Every student in a school is asked which language they speak at home. This method of collecting data is

A. a sample

B. a systematic sample

C. a stratified sample

D. a census

Question 6

In a pie chart, a category that makes up 30% of the total is shown by a sector with an angle of

A. 30°

B. 108°

C. 90°

D. 120°

Question 7

The most appropriate graph to compare the number of students in a class who chose each of five different sports is a

A. dot plot

B. line graph

C. column graph

D. stem-and-leaf plot

Question 8

The table shows the number of goals a team scored in 20 matches.

Goals	0	1	2	3	4
Frequency	3	7	5	3	2

The modal number of goals is

A. 0

B. 1

C. 2

D. 7

Question 9

The mean of $6, 8, 11, 15$ is

A. 10

B. 11

C. 40

D. 8

Question 10

The median of $5, 9, 11, 14, 16, 18, 23$ is

A. 16

B. 11

C. 14

D. 15

Question 11

The number of days it rained each month was $12, 19, 7, 25, 14, 9$. The range is

A. 16

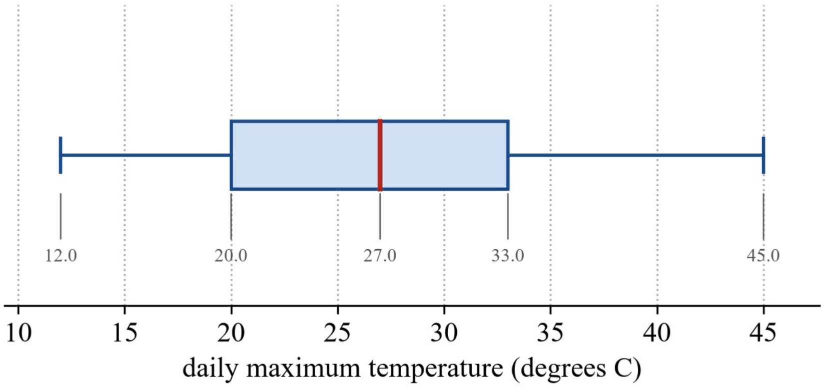
B. 18

C. 25

D. 7

Question 12

The boxplot below shows the daily maximum temperatures (°C) recorded at a town over a month.



The median daily maximum temperature is

A. 27

B. 33

C. 45

D. 20

Question 13

For the ordered data set $4, 6, 9, 12, 15, 18, 20, 24$, the lower quartile Q_1 is

A. 7.5

B. 9

C. 12

D. 6

Question 14

A set of house prices contains one very expensive house. Compared with the median, the mean will be

A. zero

B. higher than the median

C. lower than the median

D. the same as the median

Question 15

In a survey of 40 people, 18 said they preferred tea to coffee. The percentage who preferred tea is

A. 18%

B. 22%

C. 45%

D. 72%

End of Section A

Topic 2 Test A — Solutions

Question	Answer	Working
1	**A**	A colour is a label with no order, so it is categorical (nominal) data.
2	**D**	A time that can take any value on a scale (measured) is continuous numerical data.
3	**D**	Selecting from within each year-level group is stratified sampling.
4	**C**	Viewers choose whether to phone in, so the sample is self-selected and unrepresentative.
5	**D**	Every member of the population is surveyed, so it is a census.
6	**B**	$0.30 \times 360^\circ = 108^\circ$.
7	**C**	A column graph compares the counts in separate categories.
8	**B**	The highest frequency is 7, for 1 goal, so the mode is 1.
9	**A**	$\frac{6+8+11+15}{4} = \frac{40}{4} = 10$.
10	**C**	With $n = 7$, the median is the 4th value, 14.
11	**B**	$25 - 7 = 18$.
12	**A**	The median is the line inside the box, at 27.
13	**A**	The lower half is $4, 6, 9, 12$; its median is $\frac{6+9}{2} = 7.5$.
14	**B**	One very high value pulls the mean up but not the median.
15	**C**	$\frac{18}{40} \times 100\% = 45\%$.

Topic 2 Test B — Short answer

Data analysis, probability and statistics (Chapters 5–6). Approved materials: one bound reference; one scientific calculator. Section B: 6 questions (30 marks). Time: about 40 minutes. In questions where more than one mark is available, appropriate working must be shown. Only round when instructed.

Section B

Question 1 (5 marks)

The masses (in kilograms) of 11 parcels posted by a small business one morning were

$$2, 3, 5, 6, 8, 9, 11, 12, 14, 16, 35.$$

a. Find the mean mass. 1 mark

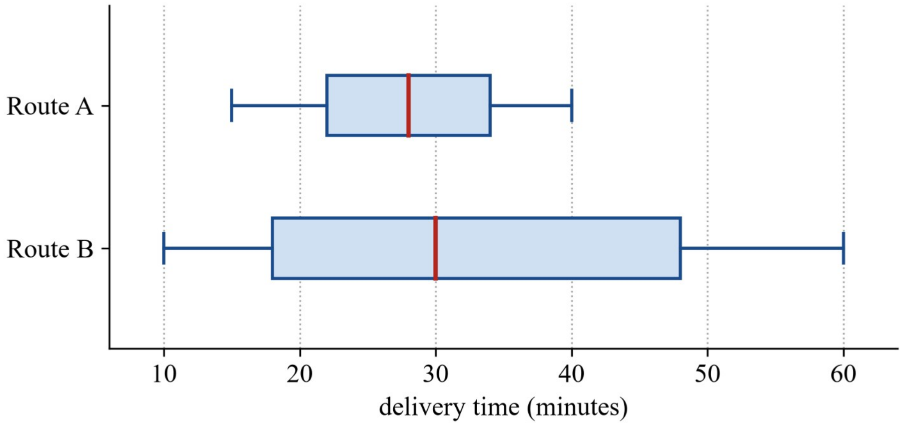
b. Find the median, the lower quartile Q_1 and the upper quartile Q_3. 2 marks

c. Find the interquartile range. 1 mark

d. Show that the 35 kg parcel is an outlier. 1 mark

Question 2 (5 marks)

The delivery times (in minutes) for two courier routes are shown in the parallel boxplots below.



a. Write down the median delivery time of each route. 2 marks

b. Find the interquartile range of each route. 2 marks

c. The two routes have a similar median. State which route was more consistent, and give a reason. 1 mark

Question 3 (5 marks)

A survey of 50 students recorded the number of siblings each student has.

Number of siblings	0	1	2	3	4
Frequency	8	18	14	7	3

a. Complete a relative frequency column for this table, giving each value as a percentage. 2 marks

b. What percentage of students have 2 or more siblings? 1 mark

c. State the most appropriate type of graph to display this data, and write down the modal number of siblings. 2 marks

Question 4 (5 marks)

A local council wants to know the opinions of its 5000 residents about a new park. It surveys 60 people at a shopping centre on a Tuesday morning.

a. Identify the population and the sample in this survey. 2 marks

b. Name the sampling method used, and give one reason the sample may be biased. 2 marks

c. Describe one way the council could obtain a more representative sample. 1 mark

Question 5 (5 marks)

A survey of 100 workers recorded whether they own a bicycle and whether they ride a bicycle to work. Some entries are missing.

	Rides to work	Does not ride to work	Total
Owns a bicycle	28		45
Does not own a bicycle			
Total	40		100

a. Copy and complete the two-way table. 2 marks

b. How many of the workers own a bicycle? 1 mark

c. Of the workers who ride a bicycle to work, what percentage own a bicycle? 2 marks

Question 6 (5 marks)

A company's yearly profit is shown in a column graph on social media. The vertical axis starts at $\$4.0$ million (not at $\$0$). The Year 1 column reaches $\$4.2$ million and the Year 2 column reaches $\$4.5$ million.

a. Find the actual increase in profit from Year 1 to Year 2, and express it as a percentage of the Year 1 profit, correct to one decimal place. 2 marks

b. On the graph, the Year 2 column appears more than twice the height of the Year 1 column. Explain why the graph gives a misleading impression of the change in profit. 2 marks

c. State one change that would make the graph a fair representation of the data. 1 mark

End of Section B

Topic 2 Test B — Solutions

Question 1.

a. $\bar{x} = \dfrac{2+3+5+6+8+9+11+12+14+16+35}{11} = \dfrac{121}{11} = 11$ kg.

b. With $n = 11$, the median is the 6th value, 9 kg. The lower half $2, 3, 5, 6, 8$ gives $Q_1 = 5$ kg; the upper half $11, 12, 14, 16, 35$ gives $Q_3 = 14$ kg.

c. $\text{IQR} = Q_3 - Q_1 = 14 - 5 = 9$ kg.

d. The upper fence is $Q_3 + 1.5 \times \text{IQR} = 14 + 1.5 \times 9 = 14 + 13.5 = 27.5$ kg. Since $35 > 27.5$, the 35 kg parcel lies beyond the upper fence and is therefore an outlier.

Question 2.

a. Reading the median line inside each box: Route A median $= 28$ minutes; Route B median $= 30$ minutes.

b. For Route A, $Q_1 = 22$ and $Q_3 = 34$, so $\text{IQR} = 34 - 22 = 12$ minutes. For Route B, $Q_1 = 18$ and $Q_3 = 48$, so $\text{IQR} = 48 - 18 = 30$ minutes.

c. Route A was more consistent: its box and whiskers are much narrower (IQR 12 versus 30 minutes), so its delivery times are less spread out even though the medians are similar.

Question 3.

a. Each relative frequency is $\dfrac{\text{frequency}}{50} \times 100\%$:

Number of siblings	0	1	2	3	4
Relative frequency	16%	36%	28%	14%	6%

b. Students with 2 or more siblings: $14 + 7 + 3 = 24$, which is $\dfrac{24}{50} \times 100\% = 48\%$.

c. The data is discrete numerical counts in separate categories, so a **column (bar) graph** is most appropriate. The modal number of siblings is 1 (the highest frequency, 18).

Question 4.

a. The population is all 5000 residents of the council area; the sample is the 60 people who were surveyed at the shopping centre.

b. The method is convenience sampling. It may be biased because the people at a shopping centre on a Tuesday morning are not representative of all residents (for example, people in full-time work are less likely to be there), so their opinions may differ from the population as a whole.

c. The council could take a simple random sample from its full list of residents (or survey people across a range of days, times and locations), so that every resident has an equal chance of being chosen.

Question 5.

a. Completing the table using the row and column totals:

	Rides to work	Does not ride to work	Total
Owns a bicycle	28	17	45

	Rides to work	Does not ride to work	Total
Does not own a bicycle	12	43	55
Total	40	60	100

(Owns-not $= 45 - 28 = 17$; rides-no-bike $= 40 - 28 = 12$; no-bike total $= 100 - 45 = 55$; no-bike-not $= 55 - 12 = 43$; not-ride total $= 100 - 40 = 60$.)

b. The number who own a bicycle is the "Owns a bicycle" row total, 45.

c. Of the 40 workers who ride to work, 28 own a bicycle: $\frac{28}{40} \times 100\,\% = 70\,\%$.

Question 6.

a. The actual increase is $\$4.5 - \$4.2 = \$0.3$ million. As a percentage of the Year 1 profit, $\frac{0.3}{4.2} \times 100\,\% = 7.1\,\%$ (to 1 d.p.).

b. Because the vertical axis starts at $\$4.0$ million instead of $\$0$, only the tops of the columns are shown. The visible heights are 0.2 against 0.5, a ratio of 2.5 to 1, so Year 2 looks more than twice Year 1 even though the profit actually rose by only about $7\,\%$. The truncated axis exaggerates the change.

c. Start the vertical axis at $\$0$ so that the column heights are proportional to the profits.