Year 10 · Statistics: Picturing Data

Name: _____ Date: _____

This sheet helps you choose the right graph, pick sensible scales, and read patterns from data. You'll also draw a line of best fit to make predictions, a skill you'll use lots next year!

Warm-up

1. Name the best visualisation to show how many students at Avondale College travel to school by bus, car, bike, or walking (a category variable with counts).

2. A dot plot of shoe sizes has these dots: size 7 (2 dots), size 8 (5 dots), size 9 (3 dots), size 10 (1 dot). How many people are in the data set, and what is the most common (modal) shoe size?

3. For a scatter plot investigating whether hours studied predicts a test score, which variable goes on the x-axis and which on the y-axis?

4. A data set of daily temperatures ranges from 6°C to 27°C. Suggest a sensible scale for the vertical axis of a graph.

Practice

5. A café in Avondale records the number of flat whites sold each day for a week: Mon 42, Tue 38, Wed 55, Thu 47, Fri 63, Sat 71, Sun 29. Which visualisation best shows the pattern across the week, and what does it show?

6. Twenty Year 10 students recorded their reaction times (seconds) in an experiment. Explain why a histogram, not a bar graph, is the right choice, and describe how the bars differ.

7. Here are the maths test marks (out of 20) for two classes. Class A: 12, 14, 15, 15, 16, 18, 19. Class B: 8, 11, 13, 15, 17, 18, 20. Find the median of each class and state which class performed better on average.

8. A survey asks Year 10 students whether they own a bike (Yes/No) and whether they are Junior or Senior. Results: Junior Yes 18, Junior No 12, Senior Yes 9, Senior No 21. Draw a two-way table and find the proportion of Juniors who own a bike.

9. A student measures the height (cm) of a bean plant each week: Week 0: 2, Week 1: 5, Week 2: 9, Week 3: 14, Week 4: 20. On a scatter plot she draws a straight line of best fit that passes close to (0, 2) and (4, 20). Use this line to predict the height at Week 5.

10. A shop owner has this data linking the daily high temperature (°C) to ice creams sold: (18, 22), (20, 30), (23, 41), (25, 52), (28, 66). Describe the relationship you would see on a scatter plot and explain how a line of best fit helps.

Challenge

11. A teacher records study hours (x) and test score out of 100 (y) for six students: (1, 45), (2, 52), (3, 61), (4, 68), (5, 74), (6, 82). Draw a line of best fit, estimate its equation, and use it to predict the score for a student who studies 7 hours. Comment on how reliable this prediction is.

12. An investigation collects each Year 10 student's arm span (cm) and height (cm). Explain (a) which visualisation is appropriate and why, (b) how to choose axis scales if arm spans range 140–185 cm and heights range 145–190 cm, and (c) what a line of best fit would tell you.

Answer key (for parents and tutors)

1. Name the best visualisation to show how many students at Avondale College travel to school by bus, car, bike, or walking (a category variable with counts).
   A bar graph. The four travel methods are categories, and we want to compare their frequencies. Bars (with gaps between them) let you compare category counts at a glance. A frequency table would also record the counts, but a bar graph best SHOWS the comparison.

2. A dot plot of shoe sizes has these dots: size 7 (2 dots), size 8 (5 dots), size 9 (3 dots), size 10 (1 dot). How many people are in the data set, and what is the most common (modal) shoe size?
   Total people = 2 + 5 + 3 + 1 = 11. The modal shoe size is 8, because it has the most dots (5), which is the highest frequency.

3. For a scatter plot investigating whether hours studied predicts a test score, which variable goes on the x-axis and which on the y-axis?
   Hours studied is the explanatory (predictive/independent) variable, so it goes on the x-axis. Test score is the response (dependent) variable, so it goes on the y-axis. The predictor always goes on x.

4. A data set of daily temperatures ranges from 6°C to 27°C. Suggest a sensible scale for the vertical axis of a graph.
   Start the axis at 0°C and go up to 30°C, marking every 5°C (0, 5, 10, 15, 20, 25, 30). This covers the full range (6°C to 27°C) with even, easy-to-read intervals and clear units (°C).

5. A café in Avondale records the number of flat whites sold each day for a week: Mon 42, Tue 38, Wed 55, Thu 47, Fri 63, Sat 71, Sun 29. Which visualisation best shows the pattern across the week, and what does it show?
   A time-series graph (points joined in day order) is best because the data is measured over time. Plot day on the x-axis and flat whites sold on the y-axis (scale 0 to 80, marking every 10). The pattern shows sales generally rising through the week, peaking on Saturday (71), then dropping sharply on Sunday (29).

6. Twenty Year 10 students recorded their reaction times (seconds) in an experiment. Explain why a histogram, not a bar graph, is the right choice, and describe how the bars differ.
   Reaction time is a numerical (continuous) variable, so the values are grouped into intervals (e.g. 0.20–0.29 s, 0.30–0.39 s, …). A histogram is used for grouped numerical data and its bars TOUCH (no gaps), showing a continuous number line. A bar graph is only for categories and has gaps between bars.

7. Here are the maths test marks (out of 20) for two classes. Class A: 12, 14, 15, 15, 16, 18, 19. Class B: 8, 11, 13, 15, 17, 18, 20. Find the median of each class and state which class performed better on average.
   Each class has 7 values, so the median is the 4th value when ordered. Class A ordered: 12, 14, 15, 15, 16, 18, 19 → median = 15. Class B ordered: 8, 11, 13, 15, 17, 18, 20 → median = 15. Both medians are 15, so the classes performed the same 'on average' by the median. (Class B is more spread out: range 20−8 = 12 vs Class A's 19−12 = 7.)

8. A survey asks Year 10 students whether they own a bike (Yes/No) and whether they are Junior or Senior. Results: Junior Yes 18, Junior No 12, Senior Yes 9, Senior No 21. Draw a two-way table and find the proportion of Juniors who own a bike.
   Two-way table: Junior: Yes 18, No 12, Total 30 Senior: Yes 9, No 21, Total 30 Totals: Yes 27, No 33, Grand total 60. Proportion of Juniors who own a bike = 18 ÷ 30 = 0.6, which is 60%.

9. A student measures the height (cm) of a bean plant each week: Week 0: 2, Week 1: 5, Week 2: 9, Week 3: 14, Week 4: 20. On a scatter plot she draws a straight line of best fit that passes close to (0, 2) and (4, 20). Use this line to predict the height at Week 5.
   Find the slope of the line through (0, 2) and (4, 20): slope = (20 − 2) ÷ (4 − 0) = 18 ÷ 4 = 4.5 cm per week. The line's equation is height = 2 + 4.5 × week. At Week 5: height = 2 + 4.5 × 5 = 2 + 22.5 = 24.5 cm. So predict about 24–25 cm. (Note: real growth may curve, so this is an estimate.)

10. A shop owner has this data linking the daily high temperature (°C) to ice creams sold: (18, 22), (20, 30), (23, 41), (25, 52), (28, 66). Describe the relationship you would see on a scatter plot and explain how a line of best fit helps.
    Plot temperature on the x-axis (explanatory) and ice creams sold on the y-axis (response). The points rise from lower-left to upper-right, showing a strong positive relationship: as temperature increases, ice cream sales

increase. An eyeballed line of best fit drawn through the middle of the points lets the owner predict sales for temperatures not measured, for example, reading up from 22°C gives a prediction of roughly 36–38 ice creams.

11. A teacher records study hours (x) and test score out of 100 (y) for six students: (1, 45), (2, 52), (3, 61), (4, 68), (5, 74), (6, 82). Draw a line of best fit, estimate its equation, and use it to predict the score for a student who studies 7 hours. Comment on how reliable this prediction is.
Choose two points the line passes near, e.g. (1, 45) and (6, 82). Slope = (82 − 45) ÷ (6 − 1) = 37 ÷ 5 = 7.4 marks per hour. Using point (1, 45): y = 45 + 7.4(x − 1) = 7.4× + 37.6. Check at x = 6: 7.4×6 + 37.6 = 44.4 + 37.6 = 82 ✓. Predict for x = 7: y = 7.4 × 7 + 37.6 = 51.8 + 37.6 = 89.4, so about 89 marks. Reliability: 7 hours is just beyond the data range (1–6 hours), so this is extrapolation and less reliable, scores also can't exceed 100, and the real relationship may flatten off.

12. An investigation collects each Year 10 student's arm span (cm) and height (cm). Explain (a) which visualisation is appropriate and why, (b) how to choose axis scales if arm spans range 140–185 cm and heights range 145–190 cm, and (c) what a line of best fit would tell you.
(a) A scatter plot is appropriate because both variables are numerical and we are investigating a RELATIONSHIP between paired data (each student gives one point). (b) Don't start at 0 (the values cluster far from zero). For the x-axis (arm span) start at 140 cm and go to 185 cm marking every 5 cm; for the y-axis (height) start at 145 cm and go to 190 cm marking every 5 cm. Even 5 cm intervals spread the points out clearly and use the graph space well. (c) A line of best fit summarises the trend, for arm span vs height we'd expect a strong positive relationship (taller people tend to have wider arm spans, roughly equal to their height), so the line would rise steadily and let us predict a likely height for a given arm span.