Year 8 · Statistics: Picture the Data

Name: _____ Date: _____

This sheet helps you choose the best graph for different kinds of data, read what graphs are telling you, and spot the odd data points we call outliers. Have a ruler handy for the graphs!

Warm-up

1. Name the type of graph that is best for showing how a single number changes over time (for example, daily temperature in Avondale over one week).

2. A class recorded their favourite lunch: 8 chose pies, 5 chose sushi, 4 chose sandwiches, 3 chose salad. Is this data categorical or numerical?

3. Look at this dot plot of goals scored by a netball team in 6 games: 2, 3, 3, 4, 4, 12. Which value looks like an outlier?

4. True or false: an outlier pulls the mean towards itself but has little effect on the median. Explain in one sentence.

Practice

5. Mele surveyed 20 students at Vertex Academy about how they travel to the centre: Walk 7, Bus 6, Car 5, Bike 2. Describe how you would draw a bar graph for this, including what each axis shows.

6. A shop tracked ice-cream sales (in dollars) each month from January to June: $900, $760, $540, $420, $300, $250. What graph should you use, and what trend does it show?

7. Year 8 students were asked their favourite sport, split by boys and girls. Boys: Rugby 10, Football 6, Netball 2. Girls: Rugby 4, Football 5, Netball 9. What kind of graph best compares the two groups, and why?

8. Here are the heights (cm) of 9 plants in a school garden: 22, 24, 25, 25, 26, 27, 28, 29, 61. (a) Find the median. (b) Find the mean. (c) Explain what the outlier does to each.

9. Aroha made a time-series graph of her weekly savings and noticed one week jumped from about $15 up to $140, then went back to normal. Give one sensible reason this is NOT an error, and say whether she should delete it.

10. A café owner has this data and wants ONE graph: the number of customers each hour from 8am to 12pm (8am: 12, 9am: 20, 10am: 35, 11am: 18, 12pm: 40). What graph type fits best, and what does the pattern suggest?

Challenge

11. A class tracked recycling collected (kg) over 5 weeks for two rooms. Room A: 8, 9, 10, 11, 12. Room B: 2, 3, 4, 5, 46. (a) Which room has an outlier and what is it? (b) Compare the medians of the two rooms. (c) Which room truly recycles more consistently, and why can the mean be misleading here?

12. A student wants to show, in ONE graph, how many students in each of Years 7, 8 and 9 answered 'Yes' or 'No' to using the school bus. Data, Year 7: Yes 30, No 10; Year 8: Yes 20, No 20; Year 9: Yes 15, No 25. (a) Name a suitable graph and explain your choice. (b) Describe what a stacked bar graph would show that a clustered one might not. (c) State one clear pattern in the data.

Answer key (for parents and tutors)

1. Name the type of graph that is best for showing how a single number changes over time (for example, daily temperature in Avondale over one week).
   A time-series graph. It plots time along the bottom (horizontal) axis and the numerical variable up the side, joining the points to show the trend.

2. A class recorded their favourite lunch: 8 chose pies, 5 chose sushi, 4 chose sandwiches, 3 chose salad. Is this data categorical or numerical?
   Categorical. The values are named groups (pies, sushi, sandwiches, salad), not measurements or counts of a quantity on each person.

3. Look at this dot plot of goals scored by a netball team in 6 games: 2, 3, 3, 4, 4, 12. Which value looks like an outlier?
   12 is the outlier. Five of the values sit between 2 and 4, but 12 is far larger than all of them.

4. True or false: an outlier pulls the mean towards itself but has little effect on the median. Explain in one sentence.
   True. The mean uses the actual size of every value, so a very large or small value drags it across, while the median is just the middle-ordered value and barely moves.

5. Mele surveyed 20 students at Vertex Academy about how they travel to the centre: Walk 7, Bus 6, Car 5, Bike 2. Describe how you would draw a bar graph for this, including what each axis shows.
   Put the four travel types (Walk, Bus, Car, Bike) along the horizontal axis, spaced apart with gaps between the bars because the data is categorical. The vertical axis shows frequency (number of students) from 0 to at least 7, marked in 1s. Draw a bar of height 7 for Walk, 6 for Bus, 5 for Car and 2 for Bike. Give the graph a title such as 'How 20 Vertex students travel to the centre'.

6. A shop tracked ice-cream sales (in dollars) each month from January to June: $900, $760, $540, $420, $300, $250. What graph should you use, and what trend does it show?
   Use a time-series graph, because the data is a numerical variable measured over time. Plot months Jan–Jun along the horizontal axis and sales ($) up the vertical axis, then join the points. The trend is a steady decrease in sales from summer into winter, dropping from $900 in January to $250 in June.

7. Year 8 students were asked their favourite sport, split by boys and girls. Boys: Rugby 10, Football 6, Netball 2. Girls: Rugby 4, Football 5, Netball 9. What kind of graph best compares the two groups, and why?
   A clustered (side-by-side) bar graph or a stacked bar graph, because there are two categorical variables: sport and gender. A clustered bar graph works well here: for each sport, draw a pair of bars (one for boys, one for girls) next to each other, using a key/legend. This lets you compare boys and girls directly for each sport, for example, more boys chose rugby while more girls chose netball.

8. Here are the heights (cm) of 9 plants in a school garden: 22, 24, 25, 25, 26, 27, 28, 29, 61. (a) Find the median. (b) Find the mean. (c) Explain what the outlier does to each.
   (a) Order the data (already ordered). With 9 values the median is the 5th value = 26 cm. (b) Mean = (22+24+25+25+26+27+28+29+61) ÷ 9 = 267 ÷ 9 = 29.67 cm (2 d.p.). (c) The outlier is 61 cm. It pulls the mean up to about 29.7 cm, higher than 8 of the 9 plants, so the mean is misleading. The median stays at 26 cm, close to most plants, so the median describes this data better.

9. Aroha made a time-series graph of her weekly savings and noticed one week jumped from about $15 up to $140, then went back to normal. Give one sensible reason this is NOT an error, and say whether she should delete it.
   A one-off event could explain it, for example she received a birthday gift or a pay-out that week. This is a genuine rare event, not a mistake, so she should keep it in her data, but she can note it so anyone reading the graph understands the spike.

10. A café owner has this data and wants ONE graph: the number of customers each hour from 8am to 12pm (8am: 12, 9am: 20, 10am: 35, 11am: 18, 12pm: 40). What graph type fits best, and what does the pattern suggest?
    A time-series graph, since it is a numerical count recorded across time (each hour). Plot the hours on the horizontal axis and customers on the vertical axis, then join the points. The pattern shows a busy mid-morning peak at 10am (35 customers) and a lunchtime rise to 40 at 12pm, with a quieter dip at 11am (18). The owner could add extra staff around 10am and midday.

11. A class tracked recycling collected (kg) over 5 weeks for two rooms. Room A: 8, 9, 10, 11, 12. Room B: 2, 3, 4, 5, 46. (a) Which room has an outlier and what is it? (b) Compare the medians of the two rooms. (c) Which room truly recycles more consistently, and why can the mean be misleading here?
(a) Room B has an outlier: 46 kg, far larger than its other values (2–5 kg). (b) Room A median = 3rd value = 10 kg. Room B median = 3rd value = 4 kg. So Room A's median is much higher. (c) Room A recycles more consistently, its values are all close together (8–12 kg). Room B's mean = (2+3+4+5+46) ÷ 5 = 60 ÷ 5 = 12 kg, which looks equal to Room A, but that mean is inflated by the single 46 kg week. The median (4 kg) shows Room B usually recycles very little, so the mean is misleading.

12. A student wants to show, in ONE graph, how many students in each of Years 7, 8 and 9 answered 'Yes' or 'No' to using the school bus. Data, Year 7: Yes 30, No 10; Year 8: Yes 20, No 20; Year 9: Yes 15, No 25. (a) Name a suitable graph and explain your choice. (b) Describe what a stacked bar graph would show that a clustered one might not. (c) State one clear pattern in the data.
(a) A clustered or stacked bar graph is suitable because there are two categorical variables: year level and bus use (Yes/No). Put Year 7, 8, 9 on the horizontal axis with a key for Yes/No. (b) A stacked bar graph stacks Yes on top of No so each bar's total height shows the class size (40 students each), making it easy to compare proportions within each year, for example the Yes portion shrinks as year level rises. A clustered graph makes it easier to compare exact Yes and No counts side by side but does not show the total as clearly. (c) Clear pattern: bus use decreases with age, 'Yes' falls from 30 (Year 7) to 20 (Year 8) to 15 (Year 9), while 'No' rises the opposite way.