Year 7 · Statistics: Picturing Data

Name: _____ Date: _____

Hi! This sheet helps you choose the right graph for different kinds of data, read graphs carefully, and spot outliers. Take your time and label everything neatly.

Warm-up

1. Name the type of graph best used to show categorical data such as favourite school lunch (pie, sandwich, sushi, wrap).

2. A time-series graph is used to show data collected over what?

3. True or false: an outlier is a data point that is very much bigger or smaller than most of the others.

4. Which average is pulled towards an outlier: the mean or the median?

Practice

5. Aroha asked her class their favourite fruit. Apple = 8, Feijoa = 5, Mandarin = 6, Kiwifruit = 3. What type of graph should she use, and what should she put on each axis?

6. A shop in Avondale records its ice-cream sales each month over summer: Nov = 120, Dec = 180, Jan = 210, Feb = 160. Which graph shows the trend over time, and describe the trend.

7. A graph is supposed to show data. Write down two things a good data visualisation should let a viewer work out straight away.

8. Here are the number of pets owned by 9 students: 1, 2, 1, 0, 2, 1, 3, 2, 9. Which value is an outlier, and does it need to be an error?

9. A teacher wants to compare boys' and girls' choices of favourite sport in one graph. Which two graph types could she use, and what does each show best?

10. The times (in minutes) students took to run 2 km were: 11, 12, 11, 13, 12, 25. Find the mean and the median, then explain which better describes a typical time.

Challenge

11. A café tracks flat-white sales for two workers over a week: Mon–Fri. Wiremu: 20, 22, 21, 23, 24. Sina: 18, 40, 19, 20, 21. Sina's total looks high. Explain using an outlier, and say which graph would best let a manager compare the two workers day by day.

12. A class made 30 test scores. When they drew a dot plot, all dots were between 60 and 80 except one dot at 15. The mean was 68 and the median was 70. Explain why the mean is lower than the median, and describe how removing the 15 would change each.

Answer key (for parents and tutors)

1. Name the type of graph best used to show categorical data such as favourite school lunch (pie, sandwich, sushi, wrap).
   A bar graph (or a dot plot). Categorical data groups things into named categories, so a bar graph with one bar per category is a good choice.

2. A time-series graph is used to show data collected over what?
   Over time. A time-series graph plots a variable against time (for example, temperature each day, or sales each month) so you can see trends going up or down.

3. True or false: an outlier is a data point that is very much bigger or smaller than most of the others.
   True. An outlier sits well away from the main cluster of data, much higher or much lower than the rest.

4. Which average is pulled towards an outlier: the mean or the median?
   The mean. Because the mean uses the size of every value, a very large or very small outlier drags it towards itself. The median (middle value) is barely affected.

5. Aroha asked her class their favourite fruit. Apple = 8, Feijoa = 5, Mandarin = 6, Kiwifruit = 3. What type of graph should she use, and what should she put on each axis?
   Favourite fruit is categorical data, so a bar graph (or dot plot) is best. The horizontal axis shows the fruit categories (Apple, Feijoa, Mandarin, Kiwifruit) and the vertical axis shows the frequency (number of students), going 0, 2, 4, 6, 8. Each bar height matches its count.

6. A shop in Avondale records its ice-cream sales each month over summer: Nov = 120, Dec = 180, Jan = 210, Feb = 160. Which graph shows the trend over time, and describe the trend.
   This is time-series data (values collected over time), so use a time-series graph with months on the horizontal axis and sales on the vertical axis, joining the points with lines. The trend rises from November to a peak in January (120 → 180 → 210), then falls in February to 160, sales climb through early summer and start dropping as summer ends.

7. A graph is supposed to show data. Write down two things a good data visualisation should let a viewer work out straight away.
   Any two of: (1) what the variable(s) is/are (what is being measured or grouped); (2) who the data was collected from (the group of people or things); (3) the units for numerical variables; (4) the frequency, proportions, patterns or trends. A clear title and labelled axes usually provide these.

8. Here are the number of pets owned by 9 students: 1, 2, 1, 0, 2, 1, 3, 2, 9. Which value is an outlier, and does it need to be an error?
   9 is the outlier, most students have 0 to 3 pets, but one student has 9, which is far above the others. It does not have to be an error; that student may genuinely own 9 pets (for example, a family that fosters animals). Outliers can be real, rare events.

9. A teacher wants to compare boys' and girls' choices of favourite sport in one graph. Which two graph types could she use, and what does each show best?
   She could use a clustered (grouped) bar graph or a stacked bar graph, because this is paired categorical data (sport AND gender). A clustered bar graph places the boys' and girls' bars side by side for each sport, making it easy to compare the two groups directly. A stacked bar graph puts one on top of the other in a single bar per sport, making the total for each sport easy to see.

10. The times (in minutes) students took to run 2 km were: 11, 12, 11, 13, 12, 25. Find the mean and the median, then explain which better describes a typical time.
    Mean = (11 + 12 + 11 + 13 + 12 + 25) ÷ 6 = 84 ÷ 6 = 14 minutes. Median: order them 11, 11, 12, 12, 13, 25; middle two are 12 and 12, so median = 12 minutes. The 25 is an outlier that pulls the mean up to 14. The median (12) better describes a typical time because it is not affected by the one very slow runner.

11. A café tracks flat-white sales for two workers over a week: Mon–Fri. Wiremu: 20, 22, 21, 23, 24. Sina: 18, 40, 19, 20, 21. Sina's total looks high. Explain using an outlier, and say which graph would best let a manager compare the two workers day by day.
    Wiremu's total = 20+22+21+23+24 = 110. Sina's total = 18+40+19+20+21 = 118. Sina looks higher only because of

Tuesday's 40, which is an outlier, her other days (18–21) are actually a bit lower than Wiremu's. Sina's mean = 118 ÷ 5 = 23.6, but her median = 20 (order: 18,19,20,21,40), so the median shows her typical day is around 20. To compare the two workers each day, a clustered bar graph works best: for each day, place Wiremu's and Sina's bars side by side, making the Tuesday spike easy to spot.

12. A class made 30 test scores. When they drew a dot plot, all dots were between 60 and 80 except one dot at 15. The mean was 68 and the median was 70. Explain why the mean is lower than the median, and describe how removing the 15 would change each.
The score of 15 is a low outlier far below the main cluster (60–80). Because the mean uses every value's size, this very low score drags the mean down to 68, while the median (middle value) stays at 70 in the main group, so the mean ends up lower than the median. If the 15 is removed: the mean would rise (it is no longer being pulled down, moving closer to 70), while the median would change only slightly, since taking away one end value barely shifts the middle of the ordered data. This shows the median is more resistant to outliers than the mean.