Data Representation and Interpretation 9

ACACIA TUTORING

Free and always will be!

Focus: A set of questions and solutions for Year 9 students focused on 'Data Representation and Interpretation' under the "Statistics and Probability" strand, tailored to the Australian Curriculum:

1. Understanding Time Series Data:

a) Define what time series data is.

Solution:
Time series data is a sequence of data points collected or recorded at regular time intervals. It's used to track changes over time in areas like economics, weather forecasting, sales, etc.

b) What are the main components of time series data?

Solution:
Trend: Long-term movement in data over time (upward, downward, or steady).
Seasonality: Regular patterns that repeat over known, fixed periods (e.g., yearly, monthly).
Cyclical Variations: Fluctuations occurring at irregular intervals, often related to business or economic cycles.
Irregular/Random Variations: Unpredictable or short-term fluctuations due to random events.

2. Identifying Trends:

a) Look at the following data representing monthly units sold for a shop over the 2024 period:

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Units Sold	100	110	120	130	140	150	160	170	180	190	200	210

Is there an evident trend?

Solution:
Yes, there is an upward trend as the sales increase each month, suggesting growth over time.
Graph data to visually observe trend. *(See below)*

b) How can you describe the trend in question 2b mathematically?

Solution:
The trend can be described as linear since sales increase by approximately 10 units each month. The equation could be:

Sales $= 100 + 10 \times (Month \ - 1\ )$, where January is considered. $Month\ 1$.

c) Look at the following data representing monthly units sold for a shop over the 2025 period:

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Units Sold	100	105	120	129	140	142	165	170	180	112	200	210

Is there an evident trend?

Solution:
Yes, there is an upward trend as the sales increase each month, suggesting growth over time. However, during October there was a considerable drop in sales.
Graph data to visually observe trend. *(See below)*

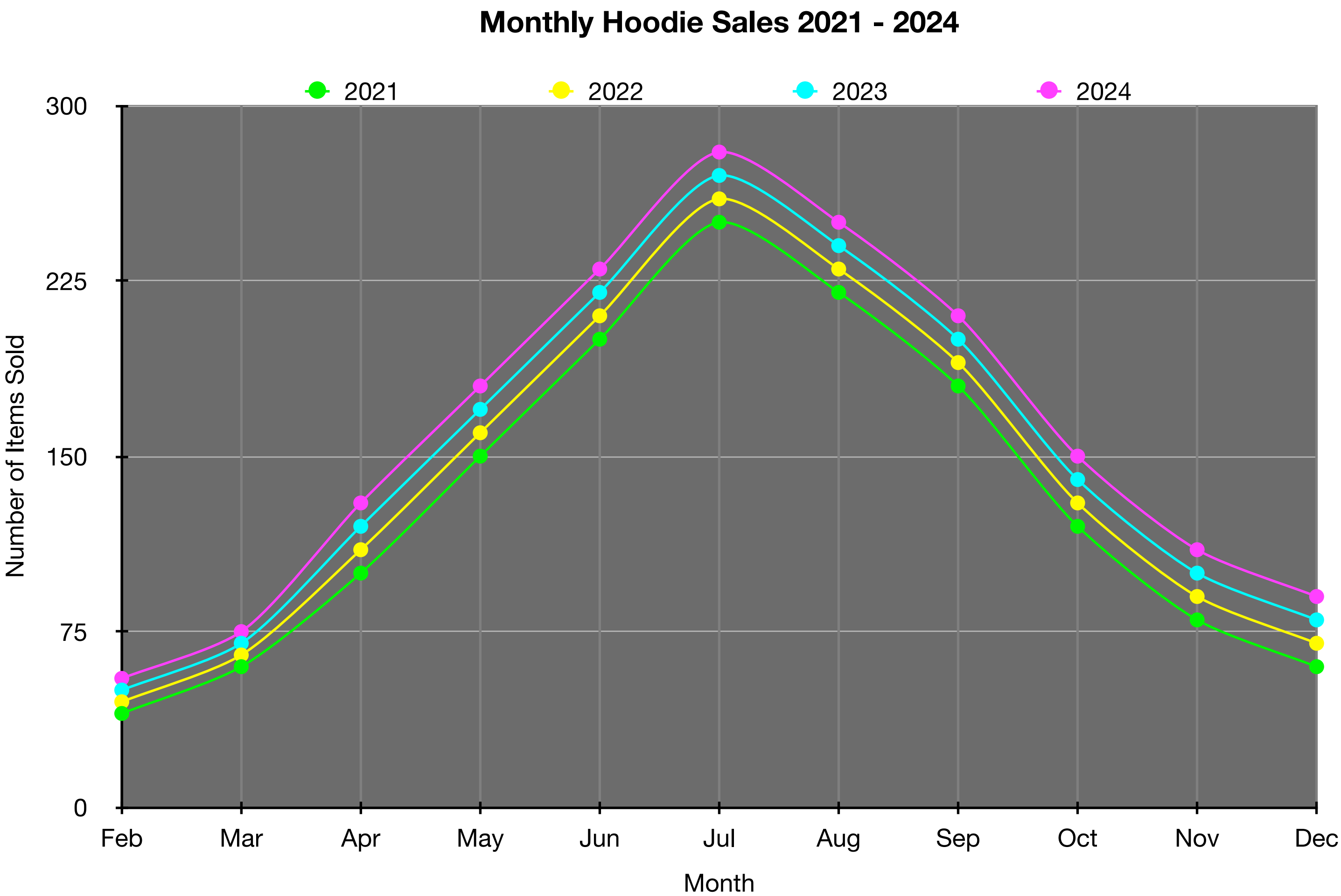
3. Analysing Seasonality:

Given the following sales data for hoodies over four years:

Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
2021	50	40	60	100	150	200	250	220	180	120	80	60
2022	55	45	65	110	160	210	260	230	190	130	90	70
2023	60	50	70	120	170	220	270	240	200	140	100	80
2024	65	55	75	130	180	230	280	250	210	150	110	90

Identify any seasonal patterns.

Solution:
There's a clear seasonal pattern where sales peak in July and August due to summer, and trough in winter months like January and February. The pattern repeats annually.



4. Smoothing Data:

a) Explain what data smoothing is and why it's done.

Solution:
Data smoothing involves removing noise or short-term fluctuations to reveal underlying patterns or trends more clearly. It's done to make long-term forecasting easier and to understand the true trend of the data.

b) Apply a $3-month$ moving average to the sales data from question 2c. (Units sold during the 2025 period). What does this tell you about the trend?

Solution:

$$\text{Mean} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{\text{Sum of scores}}{\text{Number of Scores}}$$

Jan-Feb-Mar:

$$\rightarrow \frac{(100+105+120)}{3} \approx 108.33\,.$$

Feb-Mar-Apr:

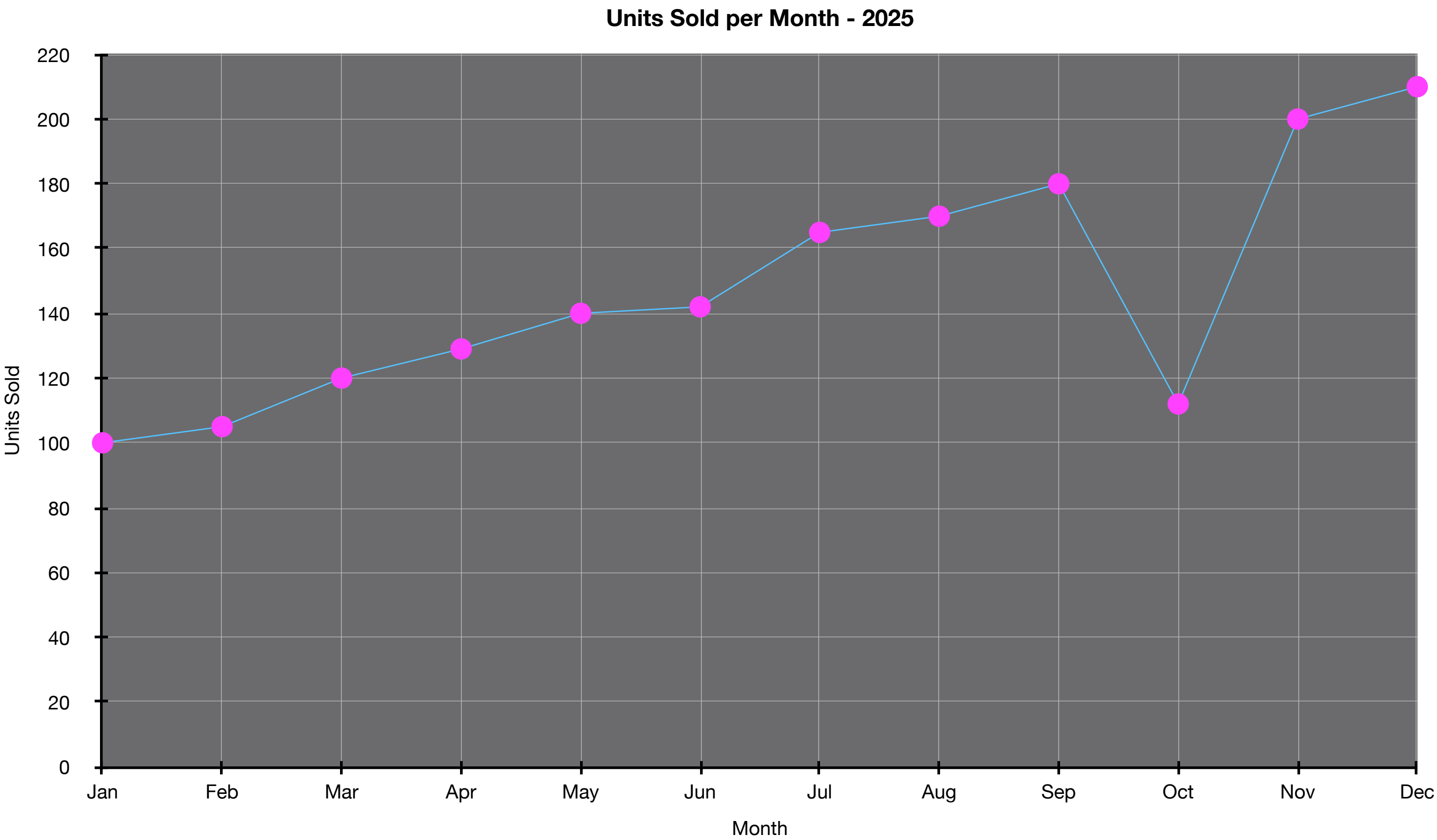
$$\rightarrow \frac{(105+120+129)}{3} = 118\,.$$

... (Continue for each 3-month period.)

This moving average shows a smoother upward trend, reducing the impact of monthly fluctuations.

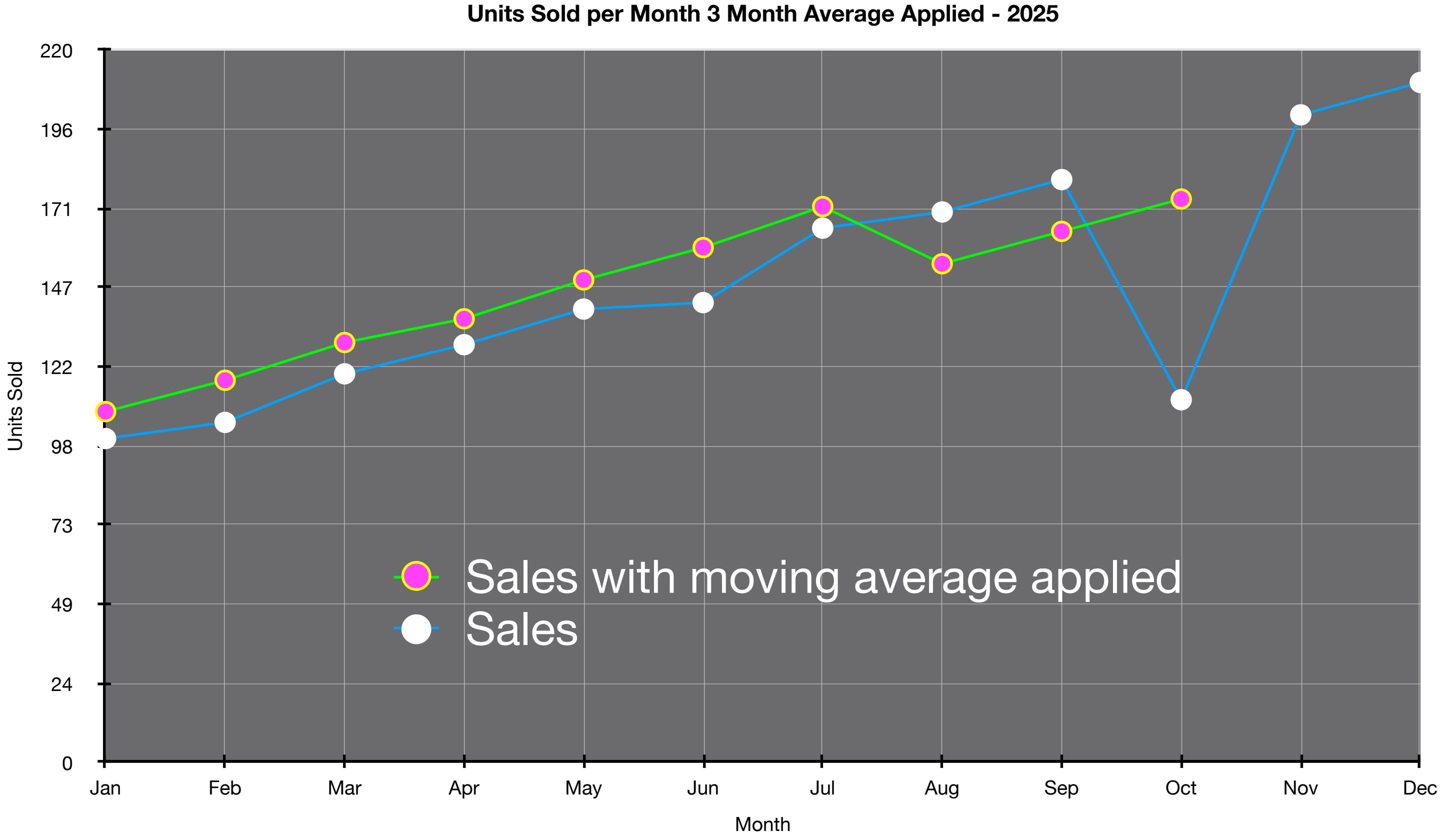
Data *before* **moving average applied:**

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Units Sold	100	105	120	129	140	142	165	170	180	112	200	210



Data *after* moving average applied (including data before average applied):

Month	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
Units Sold	100	105	120	129	140	142	165	170	180	112	200	210
Moving Average	108.33	118.00	129.67	137.00	149.00	159.00	171.67	154.00	164.00	174.00	-	-



5. Forecasting:

a) Using the trend from question 2b, predict the sales for January next year (2025).

Solution:
Sales $= 100 + 10 \times (Month \ - 1)$, where January is considered. $Month\ 1$.

If we continue the linear trend, January of the next year would be Month 13 (January 2025 :)
Sales $= 100 + 10 \times (13 - 1)$
$= 220\, units$.

b) How might you adjust this forecast (prediction), considering seasonality?

Solution:
Given the data shows no clear seasonality for January in 2024, we might use the trend forecast to predict units sold in 2025, but if there were historical patterns, we'd adjust based on those.

For example in question 4b, the 2025 data shows a drop in sales during October compared to the trend, therefore you would lower the prediction for 2026 accordingly.

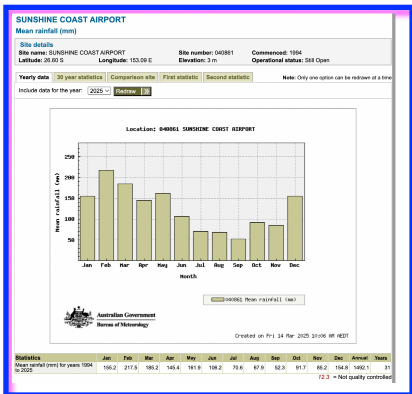
6. Practical Application:

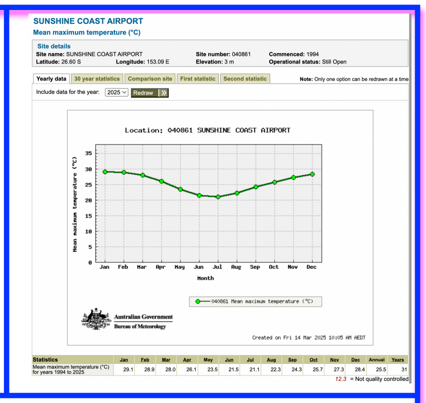
a) A local weather station records daily temperatures. How would you use time series data analysis to predict future temperature trends?

Solution:

Trend Analysis: Look for long-term increases or decreases in average temperature.
Seasonal Analysis: Identify recurring patterns in temperature related to seasons.
Smoothing: Use techniques like moving averages to see past daily fluctuations.
Forecasting: Use identified trends and seasonality to predict future temperatures, perhaps adjusting for known climatic events or anomalies.

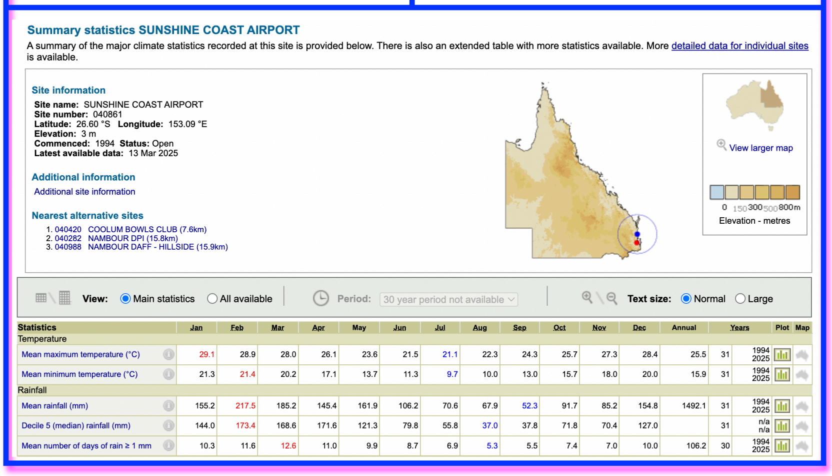
Fig. 1 →





← Fig. 2

Fig. 3 →



Statistics	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Annual	Years		Plot	Map
Temperature																	
Mean maximum temperature (°C)	29.1	28.9	28.0	26.1	23.6	21.5	21.1	22.3	24.3	25.7	27.3	28.4	25.5	31	1994 2025		
Mean minimum temperature (°C)	21.3	21.4	20.2	17.1	13.7	11.3	9.7	10.0	13.0	15.7	18.0	20.0	15.9	31	1994 2025		
Rainfall																	
Mean rainfall (mm)	155.2	217.5	185.2	145.4	161.9	106.2	70.6	67.9	52.3	91.7	85.2	154.8	1492.1	31	1994 2025		
Decile 5 (median) rainfall (mm)	144.0	173.4	168.6	171.6	121.3	79.8	55.8	37.0	37.8	71.8	70.4	127.0		31	n/a n/a		
Mean number of days of rain ≥ 1 mm	10.3	11.6	12.6	11.0	9.9	8.7	6.9	5.3	5.5	7.4	7.0	10.0	106.2	30	1994 2025		

Australian Bureau of Meteorology, Data and Statistics, accessed 14 March 2025, from:
Fig. 1 - http://www.bom.gov.au/ - Sunshine Coast Airport Rainfall
Fig. 2 - http://www.bom.gov.au/ - Sunshine Coast Airport Temperature Data
Fig. 3 - http://www.bom.gov.au/ - Sunshine Coast Airport Climate Averages

7. Challenges with Time Series Data:

Discuss one challenge when dealing with time series data.

Solution:
One challenge is dealing with outliers or anomalies which can skew the data, making it difficult to identify true trends or patterns. These could be due to unexpected events like natural disasters, economic shocks, or errors in data collection.

8. Understanding Box Plots:

a) What is a box plot, and what does it show?

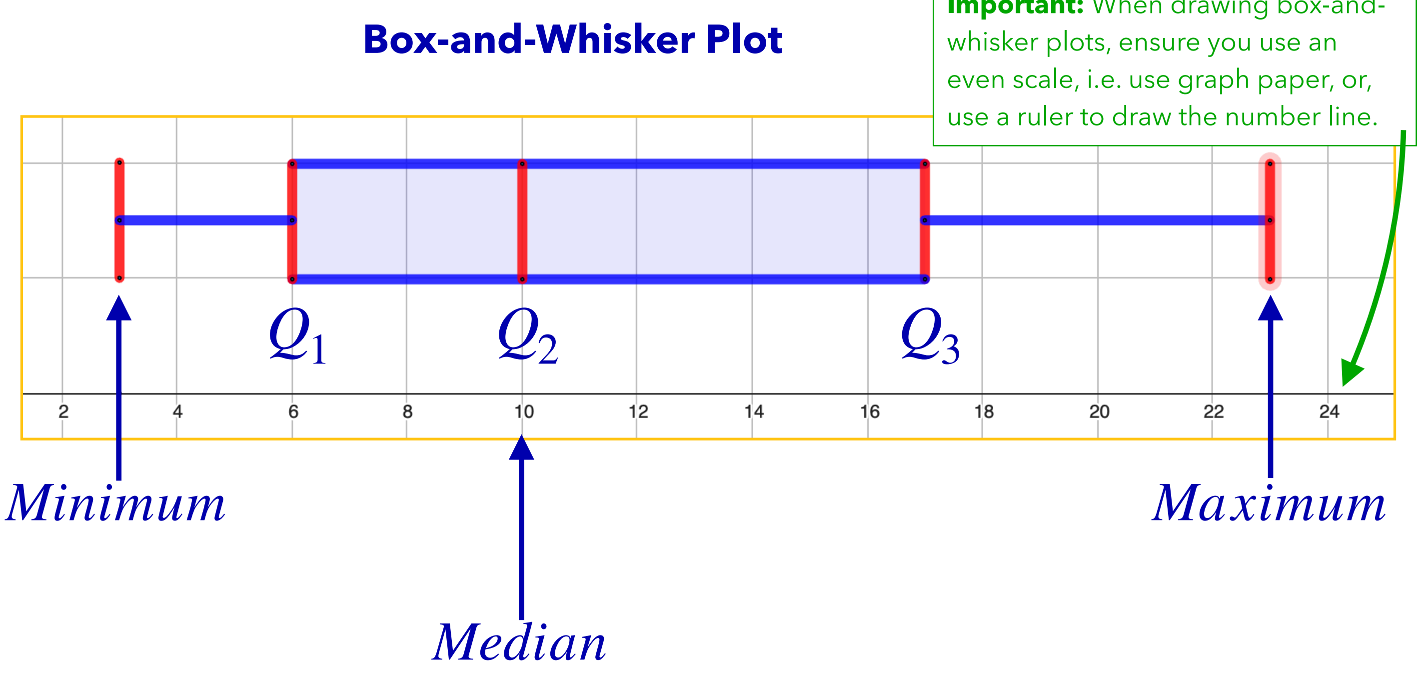
Solution:
A box plot, or box-and-whisker plot, is a graphical representation of numerical data through their quartiles. It shows the median, quartiles (Q1, Q3), and the range of the data including potential outliers. It displays the distribution's central tendency, dispersion, and skewness.

b) Identify the five-number summary used in a box plot.

Solution:

Minimum Value:	The smallest data point excluding outliers.
First Quartile (Q_1) :	The median of the lower half of the data.
Median (Q_2) :	The middle value when data is ordered (median of whole data set).
Third Quartile (Q_3) :	The median of the upper half of the data.
Maximum Value:	The largest data point excluding outliers.

Important: When drawing box-and-whisker plots, ensure you use an even scale, i.e. use graph paper, or, use a ruler to draw the number line.



9. Constructing a Box Plot:

a) Given the following dataset: $\{ 5, 7, 8, 12, 15, 18, 20, 22, 25, 28, 30 \}$ **. Construct a box plot.**

Solution:

Ordered Data: $5, 7, 8, 12, 15, 18, 20, 22, 25, 28, 30$

Minimum: 5

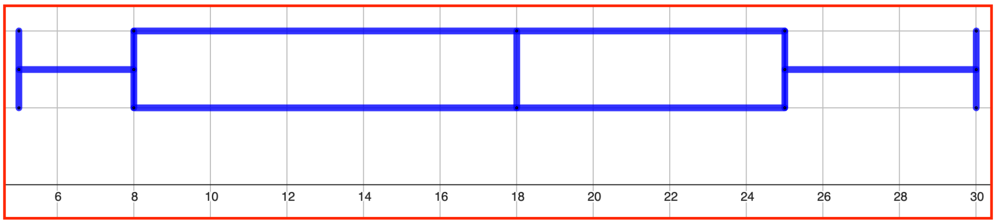
Q_1 **:** Median of the lower half ($5, 7, 8, 12, 15$) is 8

Median (Q_2 **):** 18 (middle number)

Q_3 **:** Median of the upper half ($20, 22, 25, 28, 30$) is 25

Maximum: 30

The box plot would have:
A line at 5 (minimum). A line at 30 (maximum). A line at 18 (median).
Box from 8 (Q_1) to 25 (Q_3)
Whiskers extending from 5 to 8 and from 25 to 30 .



b) What would you do if you had to check for outliers before drawing the box plot?

Solution:

Use the Interquartile Range (IQR) method:

$$
\begin{aligned}
IQR &= Q3 - Q1 \\
&= 25 - 8 \\
&= 17 .
\end{aligned}
$$

Any data point,

Below : ($Q_1 - 1.5 \times IQR$) or,

Above : ($Q_3 + 1.5 \times IQR$) is considered an outlier.

Lower bound: $8 - 1.5 \times 17$

$= -17.5$. (no data points / outliers here)

$\rightarrow$ There are no data points below -17.5 in the dataset.

Upper bound: $25 + 1.5 \times 17$

$= 50.5$. (no data points / outliers here either)

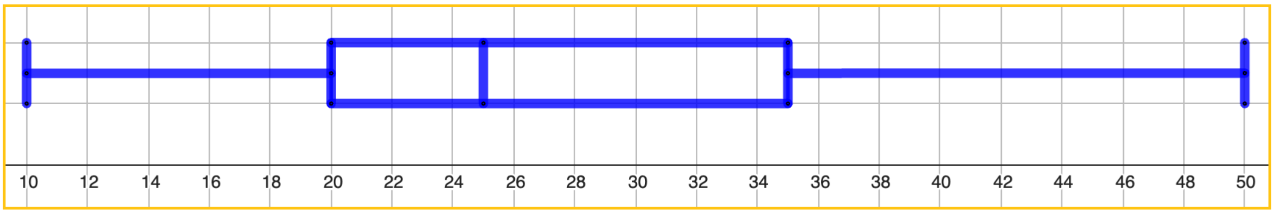
$\rightarrow$ There are no data points above 50.5 in the dataset.

Dataset: $\{ 5, 7, 8, 12, 15, 18, 20, 22, 25, 28, 30 \}$

10. Interpreting Box Plots:

a) Look at the following box-and-whisker plot and interpret it:

| Min: 10 – | Q1: 20 | Median: 25 Q3: 35 | –Max: 50 |



Solution:

The data spans from 10 to 50 .

The median is 25 , indicating that half the data points are below and half above this value.

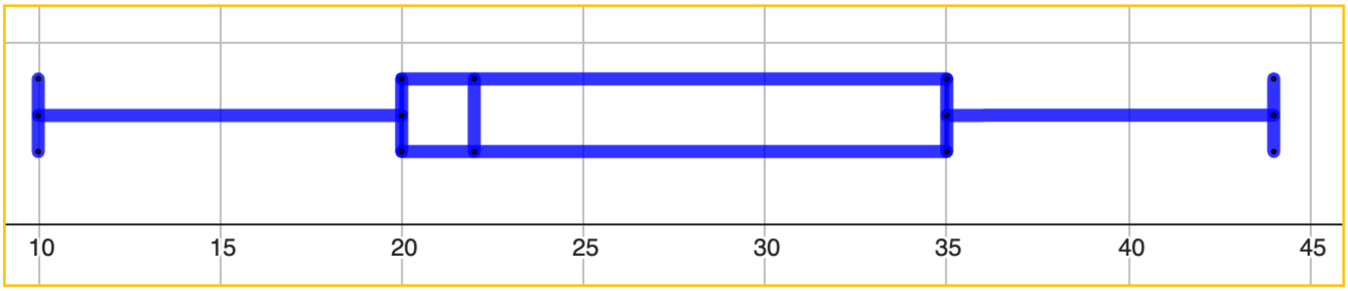
The spread from Q_1 to Q_3 (20 to 35) indicates where the middle $50\,\%$ of the data lies, showing the interquartile range is 15 . There's skewness or spread in the data, with more variability above the median than below.

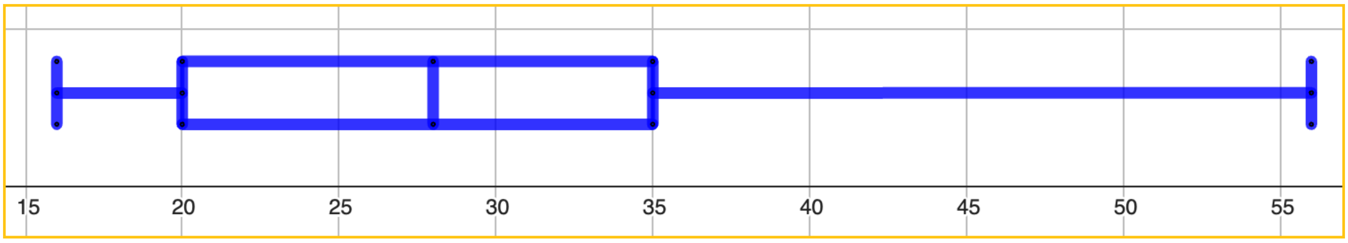
b) How can you tell if a dataset is skewed from its box plot?

Solution:

If the median is not centred within the box (closer to one end), it suggests skewness (*left diagram below).*

If one whisker is notably longer than the other (*right diagram below)*, it indicates skewness in the direction of the longer whisker.





11. Comparing Distributions:

a) Two classes took the same test. Class A's box plot has a median of 70, Q_1 at 60, and Q_3 at 80. Class B's has a median of 75, Q_1 at 70, and Q_3 at 85. Compare the performance of the two classes.

Solution:

Class A:

Median: 70 (centre of performance),

IQR: $80 - 60$

$= 20$ (spread of middle $50\,\%$).

Class B:

Median: 75 (higher centre),

IQR: $85 - 70$

$= 15$ (less spread in the middle $50\,\%$).

Comparison:

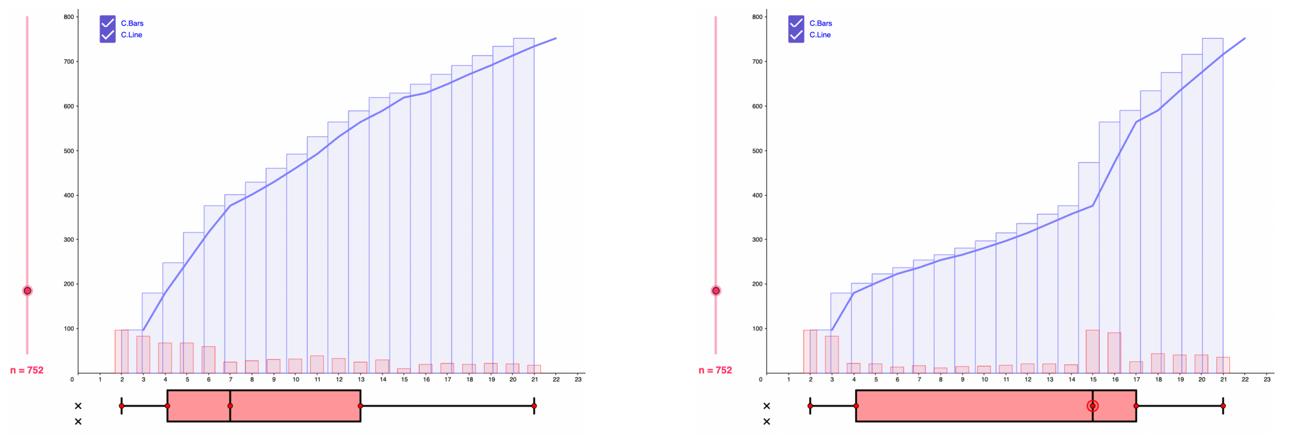
Class B generally performed better (higher median),
and had less variability in their middle scores,
suggesting more consistent performance,
while Class A had a wider spread of scores.

12. Practical Application:

a) A teacher wants to use box plots to compare test scores between two classes. What advantages does a box plot offer over a bar graph in this context?

Solution:

Box plots show more detailed statistical information like median, quartiles, and outliers, offering insights into the distribution and variability of the data. They are particularly useful for comparing distributions, highlighting differences in central tendency, spread, and skewness, which can be less clear with bar graphs.

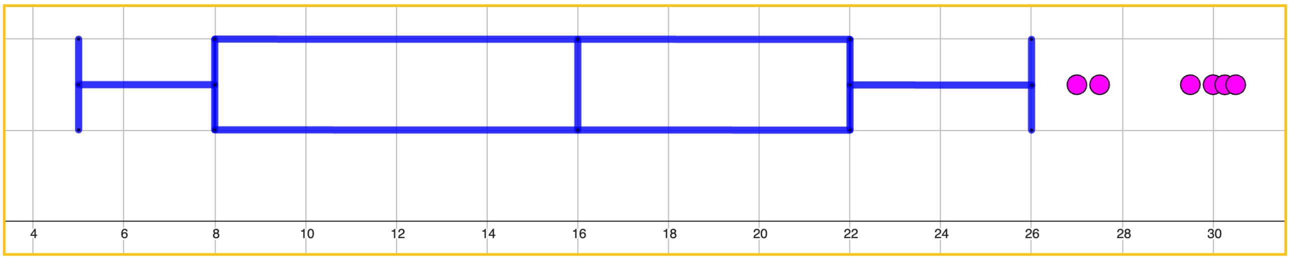


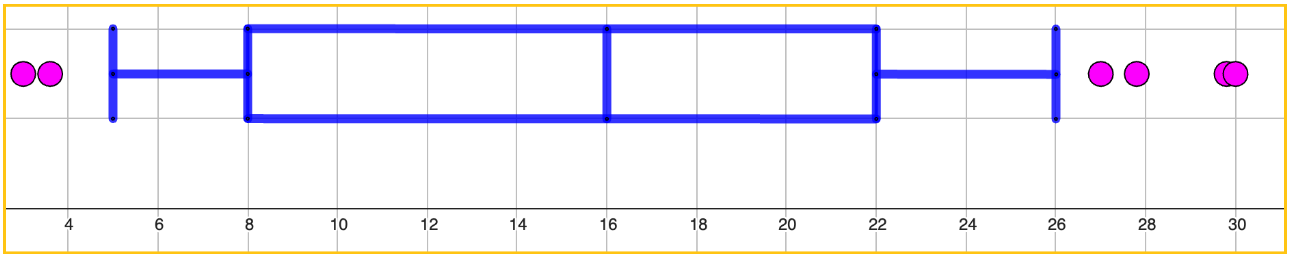
GeoGebra Interactive Box Plot Generator, accessed 15 March 2025 from: https://www.geogebra.org/

13. Box Plot Variations:

a) Explain what a modified box plot is and how it differs from a standard box plot.

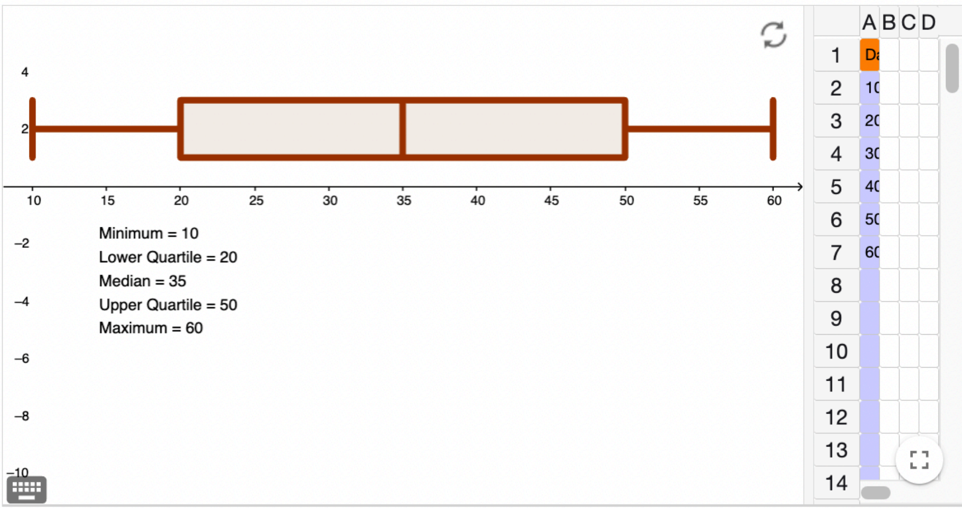
Solution:
A modified box plot includes outliers as individual points beyond the whiskers, which are set at the most extreme data points within $\pm 1.5 \times IQR$ from Q_1 or Q_3. This allows for a clearer view of data distribution including outliers, which in a standard box plot might be represented by extending whiskers to the minimum and maximum values without distinguishing outliers.





GeoGebra Interactive, accessed 15 March 2025 from: https://www.geogebra.org/m/gQkXhAvd

Box and Whisker Plot Generator



Additional Notes for Teachers:

Learning Outcomes: Students should understand how to analyse time series data, recognise trends, seasonality, and use smoothing techniques to make predictions. Students should be able to construct, interpret, and compare data using box plots, understanding how they represent the distribution of data.
Teaching Strategies:
Use real or simulated data sets from various fields like economics, climate, or health for analysis.
Implement hands-on activities where students graph time series data to visually identify patterns.
Discuss the implications of trends and forecasts in real-world scenarios.
Use real or simulated datasets for students to practice creating box plots.
Discuss how box plots can be used in real-life scenarios like comparing sports performances or test scores.
Employ software or online tools for interactive box plot construction and analysis.
Assessment: Evaluate through tasks involving data interpretation, trend analysis, smoothing, and forecasting based on given or collected data.
Assess through tasks where students must construct box plots from given data, interpret plots, and draw conclusions about data distribution.
Resources: Spreadsheets for data manipulation, graphing tools for visualisation, or online weather or stock market data for practical examples. Graph paper for manual plotting, statistical software for visualisation, or datasets from various contexts for analysis.

This question set aligns with the Australian Curriculum for Year 9, focusing on the proficiencies of understanding, fluency, problem-solving, and reasoning in statistics and probability, specifically in the context of Data Representation and Interpretation.