Topic	*Mathematics Advanced Stage 6 Statistical Analysis*
	MA-S2 Descriptive Statistics and Bivariate Data Analysis
	S2.1: Data (grouped and ungrouped) and summary statistics

LESSON 2 OBJECTIVE	compare the suitability of different methods of data presentation in real-world contexts (ACMEM048)

Syllabus Outcomes

- chooses and uses appropriate technology effectively in a range of contexts, models and applies critical thinking to recognise appropriate times for such use MA12-9

Links with Previous Learning

- calculate measures of central tendency (mean, median, mode)
- identify univariate and bivariate data
- calculate cumulative frequencies of data groups
- Construct and compare a range of data displays, including stem-and-leaf plots and dot plots (ACMSP170)
  - calculate the percentage of the whole represented by different categories in a divided bar graph or sector graph (Problem Solving)
  - compare the strengths and weaknesses of different forms of data display (Reasoning)
  - identify and explain which graph types are suitable for the type of data being considered, e.g. sector graphs and divided bar graphs are suitable for categorical data, but not for numerical data (Communicating, Reasoning)
  - draw conclusions from data displayed in a graph, e.g. 'The graph shows that the majority of Year 8 students who play a musical instrument play a string instrument' (Communicating, Reasoning)

Key Language Words / Concepts

- Univariate data, bivariate data
- mean, median
- minimum, maximum
- quartile, interquartile range (IQR)
- cumulative frequency
- Pareto chart, box plot/box-and-whisker chart, parallel box-and-whisker chart, two-way frequency table

Summary

Recapping the pre-class Flipped Learning exercise, students consider ways they might use the different measures of central tendency of a univariate data group – mean, median, and mode – in this case, a set of test scores. They recap what a central measure might lack, then review a richer view of the data group, including minimum, maximum, and quartiles, resulting in the five-number summary.

From these examples, they will create a **box-and-whisker plot** using the GeoGebra online site and discuss what the plot shows about the data group.
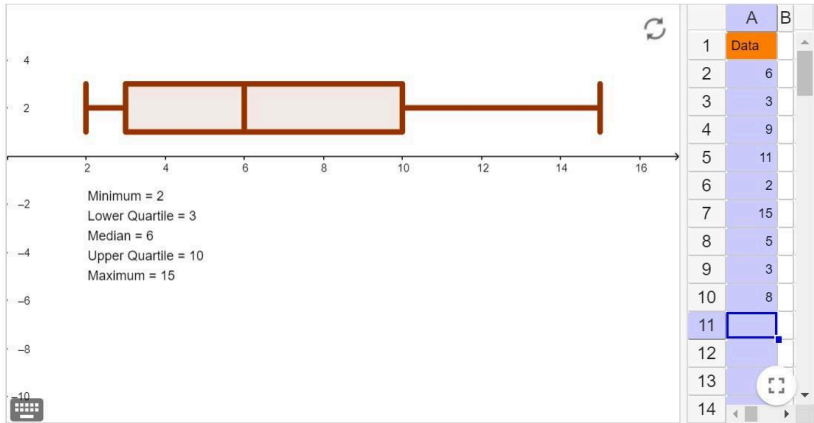
From there, they will consider the group, instead of as test scores, as counts of work units completed, e.g. computers sold, or problem sets completed. Discussing what different questions might be of interest, they will be introduced to the **cumulative frequency diagram**, showing how it illustrates totals above or below threshold marks. The relationship between box-and-whisker and cumulative frequency diagrams can be shown. They return to GeoGebra to a page where they can adjust settings and see the effects on the cumulative chart and the associated box-and-whisker plot.
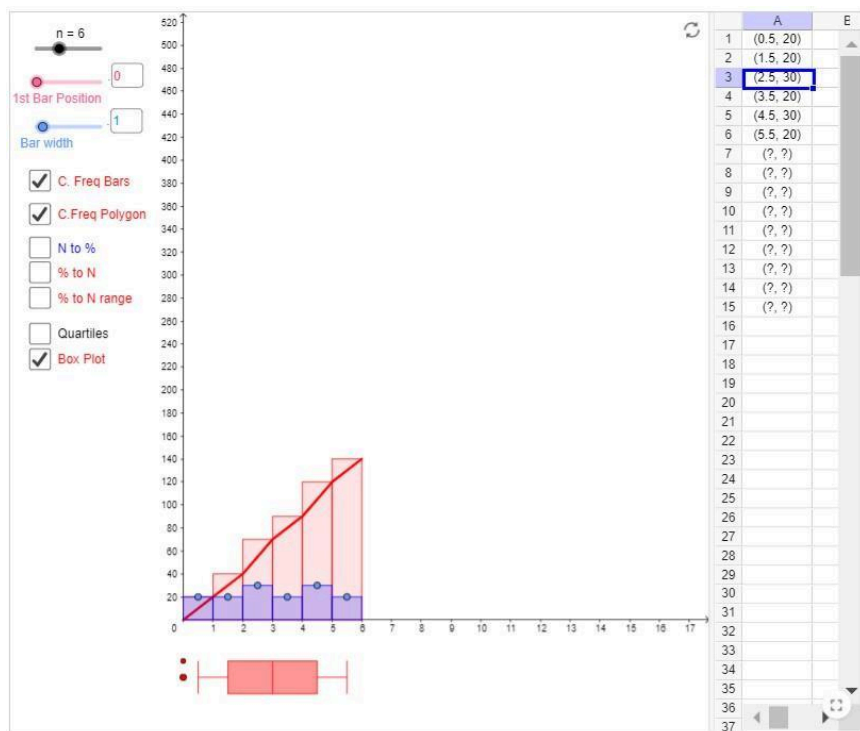
They then look at **parallel box-and-whisker plots** of two different data groups, reflecting on similarities and differences between the groups. What can they tell about the two groups relative to each other?

At this stage, the students reconsider the first group, in its original context as test scores, but with an additional set of data attached: for each score, the number of hours of sleep the student had the night before the test. This change opens the discussion to **bivariate data**, and construction of **two-way frequency tables**.

First, students discuss how to group the data to analyse; the most obvious approach would be to define ranges of hours slept, but if grouping by test score is proposed, consider that too and follow it through. (It might raise the excellent question of whether the data collection presumes the result that more sleep improves test scores, and a discussion on that point would be most worthwhile.) Converting the totals to relative frequencies is next; students discuss how they might calculate them, and what the results would tell.

Finally, the students will be given the test scores again, but in detail: the test comprised three scores for different solution stages, and the scores will be examined by stage (Setup, Computation, and Result). The new information suggests a different organisation; after discussing what might be meaningful to know from these results, the class will explore and compose **Pareto charts** of the data.

Sequence of Teaching / Learning Experiences	Time	Resources / Organisation
Introduction (Link to previous understanding): Ask students thoughts on why we would use charts. As makers, as readers. Review mean/median/mode as concepts of "middle". Revisit minimum/maximum, quartiles; finish with five-number summary.	3 mins	Flipped Learning resource
Univariate Data: Box-and-Whisker Plots Each student will create a **box-and-whisker plot** using the GeoGebra site and the below data from Example 1 of the Flipped Learning Text. Identify the parts of the plot. (Note box-and-whisker plots can be presented horizontally OR vertically.) *Keep this tab open, or save a screen shot of the plot!* 	7 mins	Box-and-Whisker generator (pirsquared, 2013) : https://www.GeoGebra.org/m/gQkXhAvd Data points are entered into the applet under the Data column.
Cumulative Frequency Diagrams Use the data from Example 3 in the Flipped Learning resource and consider it as units of work completed (e.g. computers sold, or pizzas delivered within guaranteed time). What questions about this data might be different from the test score set? Explain the **cumulative frequency diagram**, invite students to define appropriate intervals, and count the frequency for each. In GeoGebra, build the diagram (or sketch on graph paper. Use the controls to show the box-and-whisker plot. Play with the settings if time allows, viewing effects.	7 mins	Cumulative Frequency diagram generator (Dabbs, 2014) : https://www.GeoGebra.org/m/uNkxDWP5 Data entered as follows: Slide button "n = …" at top = number of intervals Frequency within each interval is controlled by the blue buttons atop each bar in the graph. (NOTE: graph scale is not changeable; in illustration here,

The Score table:

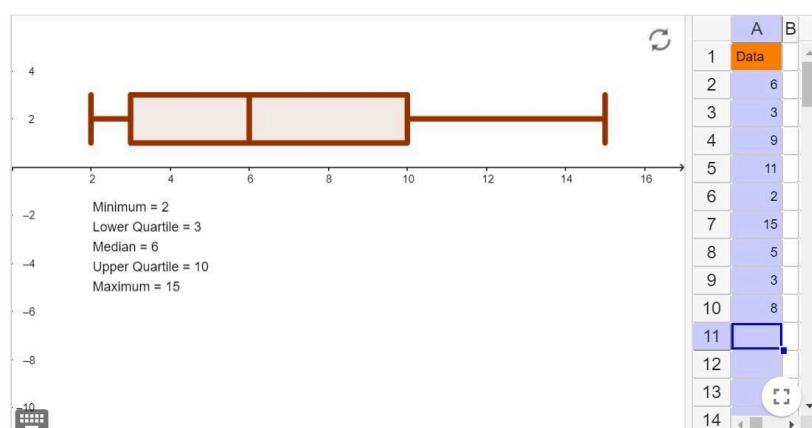| Score	6	3	9	11	2	15	5	3	8

Bars, polygon (line graph), quartiles, and box-and-whisker display are controlled by the tick boxes.

Parallel Box-and-Whisker Plots

Opening a *new tab* of GeoGebra (or referring to the saved screenshot), create a box-and-whisker plot of the below data from Example 2 in the Flipped Learning text. What similarities and differences can be observed? (For example, IQR position relative to min and max; similar median and mean; IQR size; extreme max in Example 1)

Score	4	1	12	10	7	5	9	6

Example 1: Mean = 6.9; Median = 6

Example 2: Mean = 6.75; Median = 6.5

6 mins

Box-and-whisker generator (pirsquared, 2013) : https://www.GeoGebra.org/m/gQkXhAvd
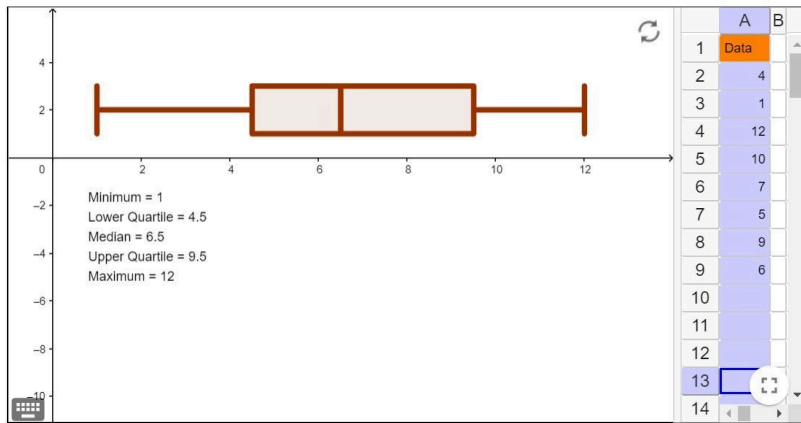
Bivariate Data: Two-Way Frequency Tables Using the data below from the Flipped Learning Text, discuss grouping into intervals – on which variable? Why?	6 mins	Two-way table demonstration page (Walch Education, 2014) : https://www.GeoGebra.org/m/wbkAsEz5 The slide button allows switching between stages of the demonstration table.

Score	6	3	9	11	2	15	5	3	8
Hrs Sleep	7	4	7.5	6	6.5	7.5	6	4	5.5

The GeoGebra page gives students a step-by-step example of a two-way table showing joint frequencies. Completing the two-way table for the Sleep Hours data may be given as a further exercise, perhaps with row or column frequency calculations.

Pareto Charts Using the below data from Example 1, by Stage, work with the students to organise a Pareto chart: 1) Total by each stage; 2) Arrange the stages in descending order of totals; 3) Draw the bar chart	6 mins	Whiteboard or screen share

Total Score	6	3	9	11	2	15	5	3	8
Setup	2	1	4	4	2	5	1	1	3
Comput.	2	2	3	4	0	5	3	1	3
Result	2	0	2	3	0	5	1	1	2

Ask students: What do you think this chart can tell us? Are there other situations where this style of chart might be more helpful? (e.g. errors in process, such as causes of late delivery: mistyped order, parts out of stock, truck delayed)

NOTE: In this case, each stage is valued at the same number of points. A particularly astute student might ask.

Conclusion (Review of new understanding)	5 mins	
Do charts make things clearer? Do these charts have clearly different strengths – and what are they? Can you think of situations when it is better *not* to use a chart, but to leave the data in its original form?		

Assessment of Learning

Ask students what they think would be interesting to know from a data set, and what decisions or actions it might drive. Comparing two charts, ask what the differences might tell them. Encourage them to use the charting tools to play, and see how changes in data points affect the charts' shape.

Evaluation

[*This Self-Evaluation should be written after you present the lesson.*

Post items you might focus on.

What factors impacted on the success of the lesson or limited learning of students?

Consider your actions, students' actions, resources, prior learning etc.

No more than 1 A4 page.]

References

Dabbs, M. (2014). *Cumulative Frequency Properties.* GeoGebra. https://www.GeoGebra.org/m/uNkxDWP5

pirsquared. (2013). *Box-and-Whisker Plot Generator.* GeoGebra. https://www.GeoGebra.org/m/gQkXhAvd

Walch Education. (2014). *UCSS Math I 4.2.1 Example 3.* GeoGebra. https://www.GeoGebra.org/m/wbkAsEz5