Using Data
with Confidence

Centre for Education Statistics and Evaluation

Contents

Introduction

Welcome to Using Data with Confidence, a suite of resources that will assist you to read, understand and interpret data.

The resources have predominantly been developed for school executives (for example Principals, Deputy Principals, Assistant Principals and Head Teachers); however it is also a valuable resource for anyone needing to understand the basics.

Each section within this manual is made up of multiple units. It is recommended that you work through each unit within the manual from start to finish, as each unit leads onto the next.

Each unit within sections 1-4 commence with an overview, followed by information boxes:

Please read

Please read boxes are displayed at the beginning of each section and provide information regarding content.

Example

The example boxes provide information regarding calculations and provide you with an example as well as an Excel formula.

Try it

Try it # boxes provide you with an exercise based on the current unit. Answers to these exercises can be found at the back of the manual.

When do I…?

When do I…? boxes list suggestions as to when/why the unit may be relevant to you.

Beware!

Beware! boxes advise you of any considerations or warnings that may be relevant.

FYI

The FYI only boxes provide additional and useful information on a unit.

Supporting resources

There are a number of supporting resources available, to be used in conjunction with this manual:

The Educator Calculator (TEC)

TEC is a multipurpose calculator comprising of 7 tabs – **TEC 1** to **TEC 6** and **Boxplots comparisons**. Simply enter all the marks from tests or assessments into the TEC tabs and they will automatically calculate:

- number of values/ sum of values
- the mean, median and modes
- minimum and maximum
- variance
- standard deviation
- confidence interval amount and scale
- quartiles and interquartile range
- percentiles
- boxplots

The **Boxplot** tab displays **all** boxplots, enabling you to compare up to six sets of marks.

A TEC tab
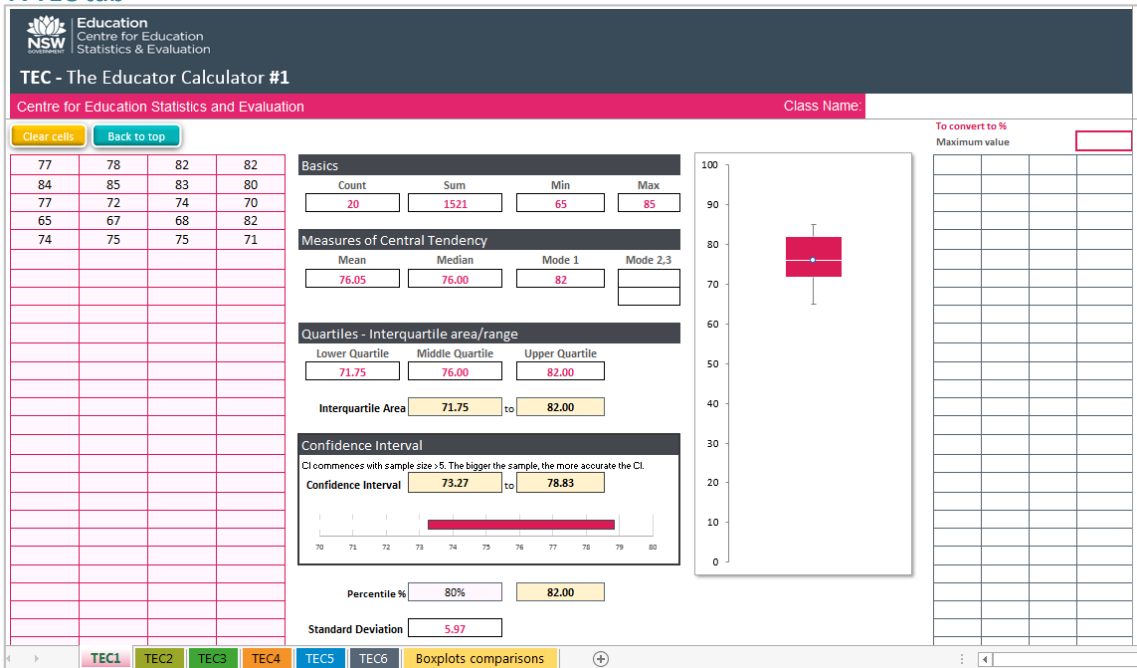
Boxplot tab
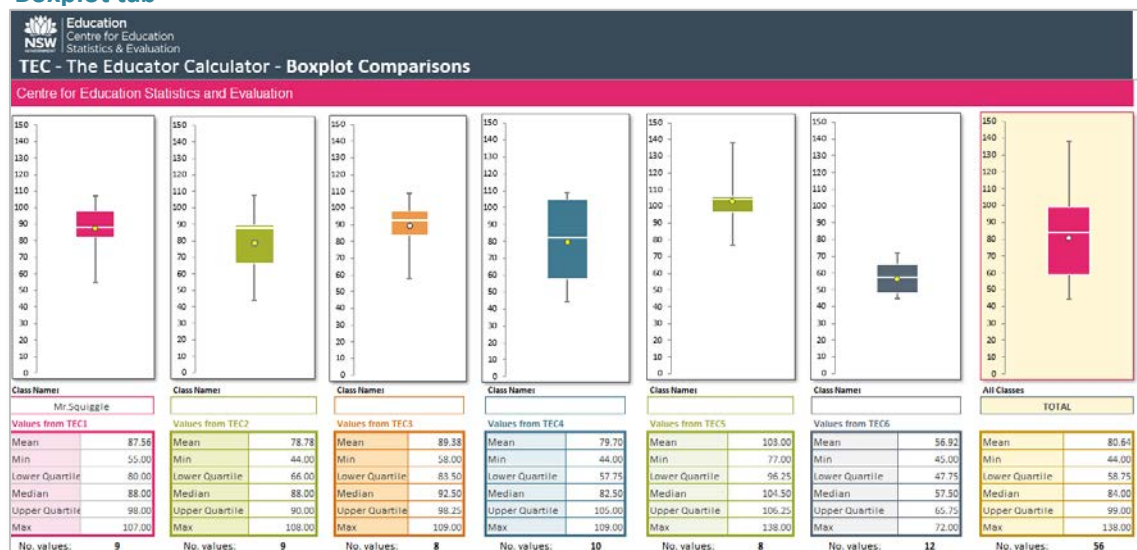
Using Data with Confidence Excel Workbook

The Excel Workbook contains **9 tabs**. Each tab corresponds with units within this manual, and is where you enter Excel formulas.

Using Data with Confidence Workbook

This booklet is used during face-to-face training and is useful for making notes on each unit within the workshop.

Using Data with Confidence Overview

Learning outcomes

At the completion of all exercises within these resources, you will be able to:

- identify and describe basic data fundamentals and use correct terminology when interpreting information
- demonstrate an understanding and use of data literacy concepts such as value added and confidence intervals
- accurately read and interpret the information within a range of charts
- report on and apply the findings from information, using the correct terminology.

Target audience

The target audience for these resources is predominantly school executives, but they may also be relevant to teachers.

The Australian Professional Standards for Teachers comprise of four professional career stages:

1. Graduate
2. Proficient
3. Highly Accomplished
4. Lead.

For each of these career stages, focus areas provide further illustration of teaching knowledge, practice and professional engagement. Understanding and interpreting data play a significant role.

Relevant teaching standards

Focus	5.3 Make consistent and comparable judgements	5.4 Interpret student data
Graduate	5.3.1 Demonstrate understanding of assessment moderation and its application to support consistent and comparable judgements of student learning.	5.4.1 Demonstrate the capacity to interpret student assessment data to evaluate student learning and modify teaching practice.
Proficient	5.3.2 Understand and participate in assessment moderation activities to support consistent and comparable judgements of student learning.	5.4.2 Use student assessment data to analyse and evaluate student understanding of subject/content, identifying interventions and modifying teaching practice.
Highly Accomplished	5.3.3 Organise assessment moderation activities that support consistent and comparable judgements of student learning.	5.4.3 Work with colleagues to use data from internal and external student assessments for evaluating learning and teaching, identifying interventions and modifying teaching practice.
Lead	5.3.4: Lead and evaluate moderation activities that ensure consistent and comparable judgements of student learning to meet curriculum, school or system requirements.	5.4.4: Coordinate student performance and program evaluation using internal and external student assessment data to improve teaching practice.

School Excellence Framework

In the Teaching Domain section within the School Excellence Framework, data skills and use are acknowledged as being an essential skill in:

THEMES	DELIVERING	SUSTAINING AND GROWING	EXCELLING
Data literacy	Teachers access and engage in professional learning that builds skills in the analysis, interpretation and use of student progress and achievement data.	The school promotes and demonstrates professional learning in data concepts, analysis and use of student assessment data and related tools. Teachers use data effectively to evaluate student understanding of lesson content.	All teachers have a sound understanding of student assessment and data concepts (e.g. causality, bias). They analyse, interpret and extrapolate data and they collaboratively use this to inform planning, identify interventions and modify teaching practice.
Data analysis	The leadership team regularly uses student progress and achievement data to inform key decisions such as resourcing and implementation of new programs or initiatives.	The leadership team comprehensively analyses student progress and achievement data for insights into student learning and discusses results with the whole staff. All teachers contribute to gathering and analysing data.	The learning goals for students are informed by analysis of internal and external student progress and achievement data. Progress towards goals is monitored through collection of quality, valid and reliable data. Reporting on school performance is based on valid and reliable data and analysis.
Data use in teaching	Teachers review student assessment data and compare results from external assessments (e.g. NAPLAN, ICAS, PAT, HSC) with internal measures to build consistent and comparable judgement of student learning.	Assessments are developed/sourced and used regularly across stages/year levels/ subject areas or the whole school to help promote consistent and comparable judgement of student learning, monitor student learning progress, and identify skill gaps for improvement and areas for extension.	Teachers clearly understand, develop and apply a full range of assessment strategies – assessment for learning, assessment as learning and assessment of learning – in determining teaching directions, monitoring and assessing student progress and achievement, and reflecting on teaching effectiveness.
Data use in planning	Clear and accurate analysis of student progress and achievement data informs planning that is shared with the school community in the Annual Report.	There is a coordinated effort by school staff to engage the school community to reflect on student progress and achievement data and develop plans and strategies for improvement.	School staff collaborate with the school community to use student progress and achievement data to identify strategic priorities, and develop and implement plans for continuous improvement. Strategies implemented reflect research on best practice and include ongoing monitoring of success.

For more information about the School Excellence Framework V2 visit the [department's website.](#)

Why do we need to understand data?

We use and report on data in everyday life without necessarily realising it, for example when we ask: how much? How often? How little? If taken to the next level, the data can not only be reported on, but it can also be turned into information from which evidence based decisions can be made. But, data is useless without the skills to understand and interpret it!

In a school environment, when data has been effectively interpreted, the findings can assist with:

- the allocation of resources
- the implementation of support schemes for students
- increasing professional development for staff members
- improving student performance.

The difference between data, analysis and statistics

Raw data are the direct result of gathering observations and measurements. Data analysis is the process of synthesising and summarising data and turning it into meaningful information. Statistics are the results of the analysis of the data and are usually collated in tables, charts and reports. Statistics provide us with clear, objective, numerical data on important characteristics or performances.
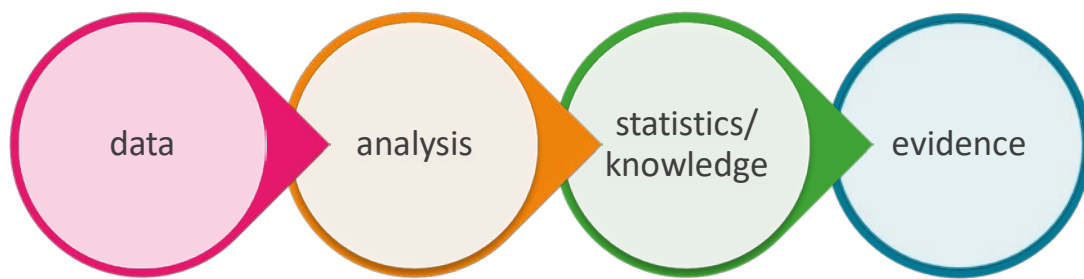
Evidence based decision-making

The evidence based decision-making process includes gathering data and processing the data into meaningful information and statistics, the interpretation of which builds knowledge and provides the basis for making informed decisions.

In order to make these well informed decisions about statistics, it is necessary to be equipped with the skills and knowledge to be able to read, understand, interpret and communicate statistical information. Therefore a level of statistical literacy is required.

Data fundamentals and terminology

Quantitative data

Quantitative data is any information that can be reduced to a set of numbers, from which you can create averages, counts, percentages, differences or totals. The data represented within this manual is predominantly quantitative data. For more information, please refer to the Evidence Guide for School Excellence – Quantitative Data.

Example:

Favourite colour	Totals
Green	5
Red	7
Sky Blue	15
Orange	9
Lime	22
Magenta	8
Cyan	11

Favourite colours
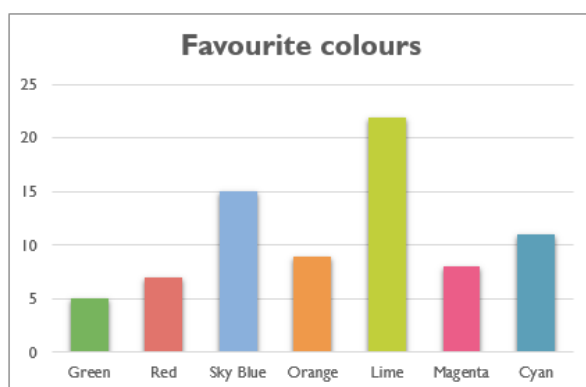
Qualitative data

Qualitative data tends to be a collection of thoughts, observations, feelings, opinions and/or experiences. It can be useful to answer questions about what, how and why rather than how many or how much. Sometimes, qualitative data can be turned into quantitative data by assigning numbers to categories, such as a ranking.

Qualitative data is not covered in great detail in this material. For more information, please refer to the Evidence Guide for School Excellence – Qualitative Data.

What is/are data?

Data are measurements or observations that are collected as a source of information. There are many different types of data, and ways to represent it (such as charts and tables).

Data unit
A data unit is a single entity (such as a student, staff member or school). It may also be referred to as a data record, case or a row.

Data item (variable)
A data item is a characteristic of a data unit which can be measured or classified (such as gender or exam scores). It is also known as a variable, because the characteristics may vary (see quantitative data, page 9).

Observation
An observation is a data item that has been recorded about a data unit. Observations may be numeric or non-numeric, such as 23 or female.

Dataset
A dataset is a collection of all observations.

Dataset   Data items (variables)

Name	Year	Gender	NAPLAN Average
James	7	M	732
Nina	9	F	586
Eleanor	7	F	612
Joe	5	M	517
Melanie	7	M	716

Data unit   Observations

Aggregate/granular

Aggregate information or charts contain high-level information, summarised for groups of data units. They do not tend to reveal personally identifiable information, and may instead provide an overview of examination results, ages, value-added information or statistical information relating to communities.

Granular information or charts contain information of the individual level (for example students or staff) and often may include personally identifiable information such as names.

Please note: Privacy laws are in place to protect information about individuals. If you are unsure whether you can report on or share information, check first!

Example

Aggregate
This is an aggregate chart taken from Scout. It displays the percentage of the population with diabetes, ages 16 years and over, within a local health area. It does not provide any personally identifiable information.
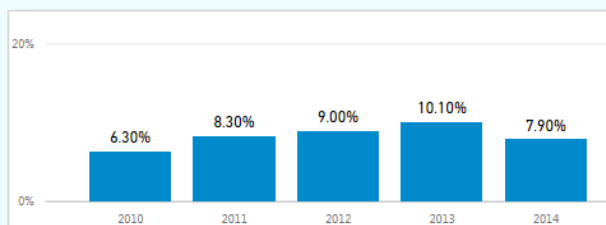
Granular
This is a granular chart from Scout. It displays professional development hours by staff members.
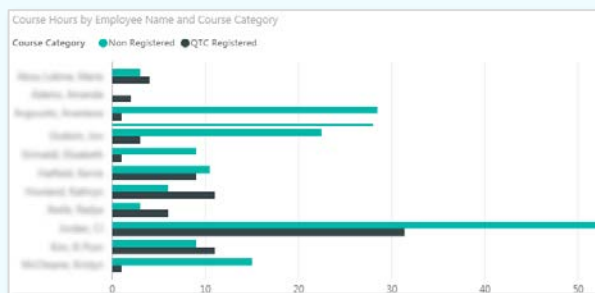
Dot points within this chart represent the average NAPLAN and scaled HSC scores for individual students. Student names are visible when the mouse hovers over the dots. Is this a granular or aggregate chart?
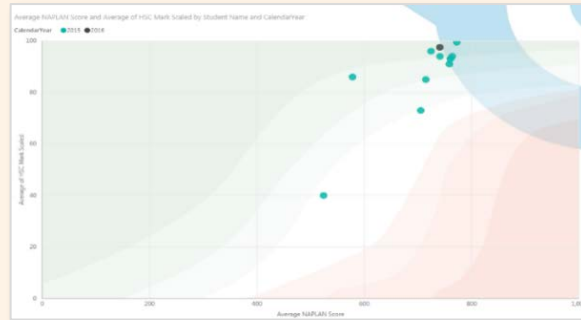
Answers to 'Try its'

…use aggregate or granular charts?

- Granular information/ charts can be used to display information regarding individuals, such as total staff development hours, student grades or which subjects teachers are accredited to teach.

- Aggregate information/ charts are used for high-level information, such as community data, school enrolment numbers or percentage of students achieving at or above national minimum standards.

Calculating percentages

A percentage (%) is a way to express a number as part of a total (for example, 60% of students achieved band 5). To calculate a percentage, the formula must contain a numerator and a denominator.

Numerator and denominator

The numerator is the top number in a fraction, and the denominator is the bottom number.

Example

To calculate the percentage of the pink triangles (3) within the total number of shapes (18), divide the numerator (3 pink triangles) by the denominator (18 total shapes) then multiply by 100.
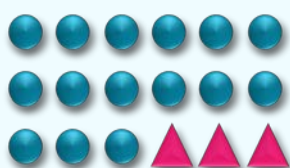
$$\frac{\text{Numerator}}{\text{Denominator}} = \frac{3}{18} \times 100 = 16.67\%$$

Excel formula: **=(3/18)*100**

Try it #2

What is the percentage of lime diamonds within these shapes? Hint: There are 8 lime diamonds, 28 shapes.
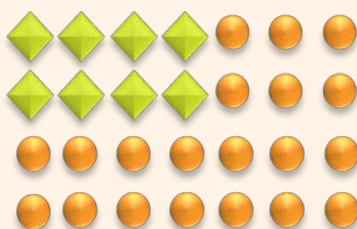
Answers to 'Try its'

When do I...?

...calculate percentages?

- % of students in each grade bands
- % of students above or below standard
- % of students with English as an additional language or dialect (EAL/D) background
- % of school funds spent on professional learning.

Beware!

Beware of interpreting differences in percentages, especially changes in percentages over time when the denominator is small (<20), as the sample size may be too small to give an accurate representation of the data.

Measures of central tendency - mean, median and mode

A measure of central tendency is a central value within a group of numbers. Several statistics can be used to represent the central values; the most common are mean, median and mode.

Mean (average)

The mean is the most common measure of central tendency and the one that can be mathematically calculated. It is the average of a group of numbers.

Example

The mean is calculated by summing up all the individual values and dividing by the total number of values. It adds together all the numbers within a range and divides by the quantity of numbers.

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2		10	14	17	21	24	26	28	29	29		22
3												

22

Excel formula in cell L2   =Average(B2:J2)

Try it #3

In the Excel workbook, using the **mean** tab:
In cell **J2**, enter the formula to calculate the mean.

	A	B	C	D	E	F	G	H	I	J
1										
2		12	6	16	42	22	67	87		
3										

Answer _____   **Answers** to 'Try its'

When do I...?

...use a mean?

- to report average school performance in internal or external assessments
- to compare previous year or state averages
- to compare results against another class or to a previous assessment task
- to establish the performance of students as a group in a class assessment task
- to examine trends over time in external testing, such as NAPLAN.

Beware!

- The mean includes every value in the distribution and is therefore influenced by outliers.
- The mean in isolation does not account for the distribution or variance of student performance.

Median

The median number is the middle value of a list of numbers. It divides the distribution in halves; half of the numbers are above the median and half are below it. The median is also known as the 50th percentile.

The calculation for median is as follows:

Odd quantity of numbers: There are 9 numbers in this example. The number in the centre, with 4 numbers at either side is the 5th number. The median is 14.

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2		13	13	13	13	14	14	16	18	21		14

14

Even quantity of numbers: Find the middle two numbers, add them together, then divide them by 2. There are 8 numbers in this group. The middle 2 numbers are 22 and 24. 22 and 24 added together equals 46 divided by 2 is 23.

The median is 23.

	A	B	C	D	E	F	G	H	I	J	K
1											
2		12	18	19	22	24	26	28	32		23
3											

(22+24) ÷ 2 = 23

Excel formula in cell **K2** =Median(B2:I2)

In the Excel workbook, using the **median** tab**:**

In cell **J2**, enter the formula to calculate the median.

	A	B	C	D	E	F	G	H	I	J
1										
2		88	37	45	92	22	12	17		
3										

Answer: _____

Answers to 'Try its'

...use median?

- In a class test, if there are some very high or very low performers, a median will better represent the class average. The median is less affected by outliers than the mean, and is usually the preferred measure of central tendency when the distribution is skewed.
- It can be used to understand the point at which 50% of students are above and below a percentile.
- The median is an element of a boxplot.

Mode, bimodal or multimodal

The mode is the value that appears most frequently within a set of numbers. If two or more sets of numbers appear with the same frequency within a set, this is called bimodal or multimodal.

Example

Mode: In the following set of numbers, **13** appears 4 times, more times than any other number, and is therefore the mode.

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2		13	13	13	13	14	14	16	18	21		13

Bimodal: If two numbers appear the same amount of times, and more frequently than any other number, this is referred to as bimodal.

Multimodal: If multiple numbers appear the same amount of times, and more frequently than any other number, this is referred to as multimodal.

Excel formula for single mode: =Mode(B2:J2)

Try it #5

In the Excel workbook, using the **mode** tab:

In cell **Q2**, enter the formula to calculate the mode.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1																	
2		11	14	11	18	14	18	13	14	13	16	13	16	16	14		

Answer: _____

Answers to 'Try its'

When do I…?

…use mode, bimodal or multimodal?

- If your mode represents a large proportion of cases (for example, 12 times in 20), this would indicate that the assessment does not discriminate well between student achievement of learning outcomes.
- If the mode is apparent at either end of the assessment scale, this would suggest lack of differentiation between the students. The assessment may be too easy or too difficult.

Outliers

Outliers are extreme values that are notably different from the rest of the data. It is important to detect outliers because they can significantly alter the results of the data analysis. The mean is more sensitive to the existence of outliers than the median or mode.

12 students who participated in the regional 200m under 15's running finals had their times recorded. There is an outlier amongst the figures. One student (Trevor) completed the race in 82 seconds, all other recorded times were less than 30 seconds.

Example

Times: 22 22 24 25 25 **26** 26 **27 27 27** 28 **82**

outlier

median (26)   mode (27)   mean (30.08)
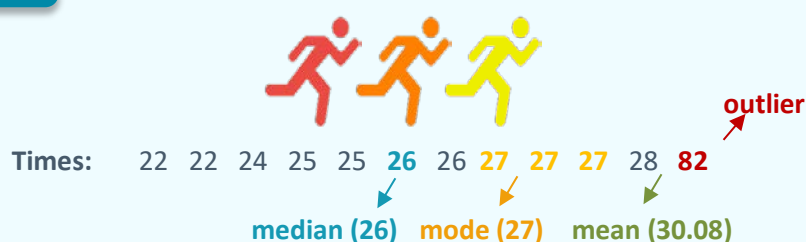
- **Median**
  The median is a good representation of the results. The outlier has had no effect on the median figure. The median figure is **26 seconds.**

- **Mode**
  The mode is a good representation of the results and has not been affected by the outlier.
  The mode is **27 seconds.**

- **Mean**
  The mean is affected by the outlier and may not be a suitable representation when reporting on this information. The mean is **30.08 seconds.**

All of the students' times have been plotted on the image below.

Mode = 27 secs

Trevor's time = 82 secs

0 5 10 15 20 25 30 35 40 45 50 55 60 65 70 75 80 85

Median = 26 secs

Mean = 30.08 secs
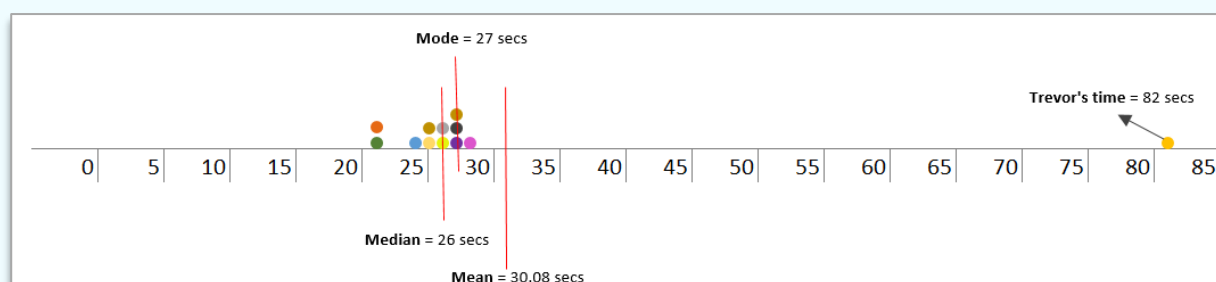
12 students were asked to guess the number of marbles within a glass jar. The actual number of marbles within the jar was 114.

Their responses are shown at below. There are 2 obvious outliers amongst the guesses.

12 82 90 102 110 118 118 118 125 132 140 **630**

outlier   **outlier**

1.   What are the mean, median and mode of the numbers shown?

2.   Which measure/s of central tendency could be a good representation of all the guesses?

Answer 1:_____

Answer 2:_____

Answers to 'Try its'

As a rule, outliers that are accurate should be kept within the dataset and not removed. You should always clarify whether the outlier is an accurate representation or if it is an error, such as a typing error. If it is an error, endeavour to find the correct information and change it.

You should mention any outliers when reporting on statistics. Only in extreme cases should the outliers be removed.

Measures of dispersion/distribution

Quartiles

Quartiles are the cuts that divide a group of numbers into 4 quarters of equal size.

- **lower quartile:** the 25th percentile (the cut at 25% of the set of values)
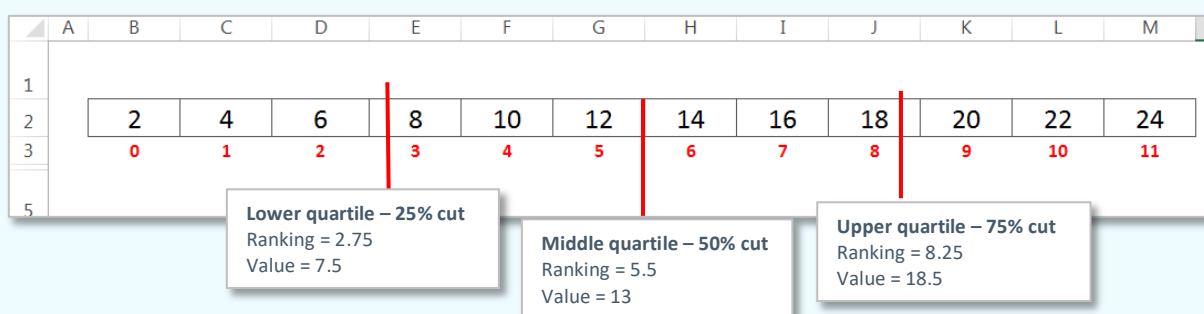- **middle quartile** (median): the 50th percentile (the cut at 50% of the set of values)
- **upper quartile:** the 75th percentile (the cut at 75% of the set of values)

Example

There are **3** quartiles; they are the points at which values are cut into four quarters.

When calculating quartiles, a formula called **(N-1) basis** is used. **N** is the count of values, **-1** takes into account the fact that the ranking starts at zero.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1													
2		2	4	6	8	10	12	14	16	18	20	22	24
3		0	1	2	3	4	5	6	7	8	9	10	11
5													

Lower quartile – 25% cut
Ranking = 2.75
Value = 7.5

Middle quartile – 50% cut
Ranking = 5.5
Value = 13

Upper quartile – 75% cut
Ranking = 8.25
Value = 18.5

In the image above, there are **12** values ranging from **2** to **24**. Each value is in ascending order and has been given a ranking, commencing at **0** (red values).

Calculating lower quartile (quartile 1)
The lower quartile makes the cut at the 25% mark of the rankings, to split first and second quarters.

The total count of values **(N)** is **12,** the **N-1 basis** formula means we need to deduct **1** from **12, = 11.**

The cut point for the lower quartile is **25%**. Therefore **25%** x **11 = 2.75.**

The value that aligns with the ranking point of **2.75** is **7.5**. This is the lower quartile value.

Excel formula for **lower quartile**: =Quartile(B2:M2,1)

Calculating middle quartile (quartile 2)
The middle quartile is also the median, so you could use the =Median() function, otherwise the formula without Excel would be 50% x (N-1) which equates to 50% x 11 = 5.5. The value crossing the 5.5 ranking mark is **13.**

Excel formula for **middle quartile**: =Quartile(B2:M2,2) OR =Median(B2:M2)

Calculating Quartile 3 (Upper Quartile)
The upper quartile formula without excel would be 75% x (N-1) which equates to 75% x 11 = **8.25**. The value crossing the **8.25** ranking mark is **18.5**.

Excel formula for **upper quartile**: =Quartile(B2:M2,3)

In the Excel workbook, open the **quartiles** tab:

In cells D5, D6 and D7, enter the formula to calculate the quartiles.

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2		4	7	9	11	13	17	19	22	24	27	33
3		0	1	2	3	4	5	6	7	8	9	10
4												
5												
6												
7		Lower Quartile										
8		Middle Quartile										
9		Upper Quartile										

Answer:_____

Answers to 'Try its'

When do I…?

…use quartiles?

Quartiles are:

- included in boxplot charts
- often used to group students by socioeconomic status, for example using the Index of Community Socio-Educational Advantage (ICESEA) or Family Occupation and Education Index (FOEI).
- displayed within SMART Student Growth for NAPLAN charts.

Interquartile range

The interquartile range represents the range from the lower quartile (25th percentile) to the upper Quartile (75th percentile). This is the middle 50% of a set of data.

Example

The formulas for quartiles have been used to find the interquartile range. The area highlighted in red is the middle 50% of values.

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2		4	7	9	11	13	17	19	22	24	27	33
3		0	1	2	3	4	5	6	7	8	9	10
4												
5				Lower Quartile = 10					Upper Quartile = 23			
6												
7		Lower Quartile		10.0								
8		Middle Quartile		17.0								
9		Upper Quartile		23.0								

Try it #8

In the Excel workbook, use the **IQR** tab:

In cells D5, D6 and D7 enter the formulas to calculate the quartiles and highlight the interquartile range on the image below.

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2		3	6	7	11	13	15	16	17	21	23	26
3		0	1	2	3	4	5	6	7	8	9	10
4												
5		Lower Quartile										
6		Middle Quartile										
7		Upper Quartile										

Answers to 'Try its'

When do I…?

…use an interquartile range?
- to give a more realistic representation of spread of performance than minimum or maximum (as they are not affected by outliers).
- within boxplots
- to determine where the middle 50% of students perform in an assessment task.

Percentiles

A percentile is the value below which a percentage of data falls, for example, 90% of values fall below the 90th percentile. The values that are above the 90th percentile are the top 10% of values.

As with quartiles, the **N-1 basis** formula should be used. Each value is given a ranking from lowest to highest, commencing at 0. **N** is the count of values, **-1** indicates the commencement of ranking.

Example

	A	B	C	D	E	F	G	H	I	J	K	L	M
1													
2													
3		34	42	51	56	65	69	71	78	84	85	88	92
4		0	1	2	3	4	5	6	7	8	9	10	11

80th Percentile
Ranking = 8.88
Value = 84.80

Calculating 80th percentile

In the image above, there are **12** values ranging from **34** to **92**. Each value has been given a ranking, commencing at **0** (red values).

The total count of values is **12**. The calculation *without* using Excel would be: 80**%** x **(N-1)** which equates to 80% x 11 = **8.88.**

Find the **8.88** mark for the ranking, and the value returned is **84.8.** This is the 80th percentile.

Excel formula for **80th Percentile**: =Percentile(B2:M2,80%)

Try it #9

In the Excel workbook, use the **percentile** tab**:**

In cell **D6** add the formula to calculate the 75th percentile.

	A	B	C	D	E	F	G	H	I	J	K
1											
2											
3		23	27	33	35	39	42	47	49	55	61
4		0	1	2	3	4	5	6	7	8	9
5											
6		Percentile									

Answer: _____

Answers to 'Try its'

When do I...?

...use percentiles?

- To ascertain which students are below or above a certain percentage.
- can be used to map student performance.
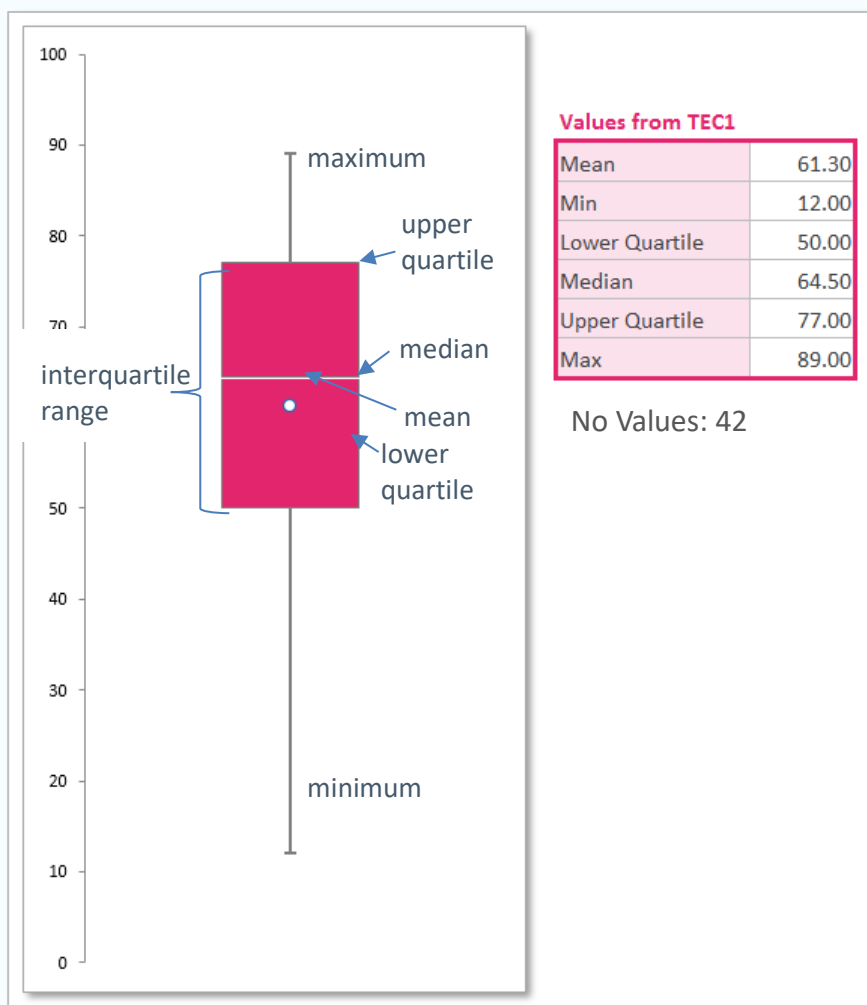
Boxplots (box and whisker plots)

Boxplots provide a very clear representation of elements covered in this manual. They display the mean, median, quartiles, interquartile range, minimum and maximum.

There are different types of boxplots; the image below is just one example. Sometimes boxplots may use the 10th and 90th percentile instead of minimum and maximum.

For further information about boxplots, please refer to boxplot charts.

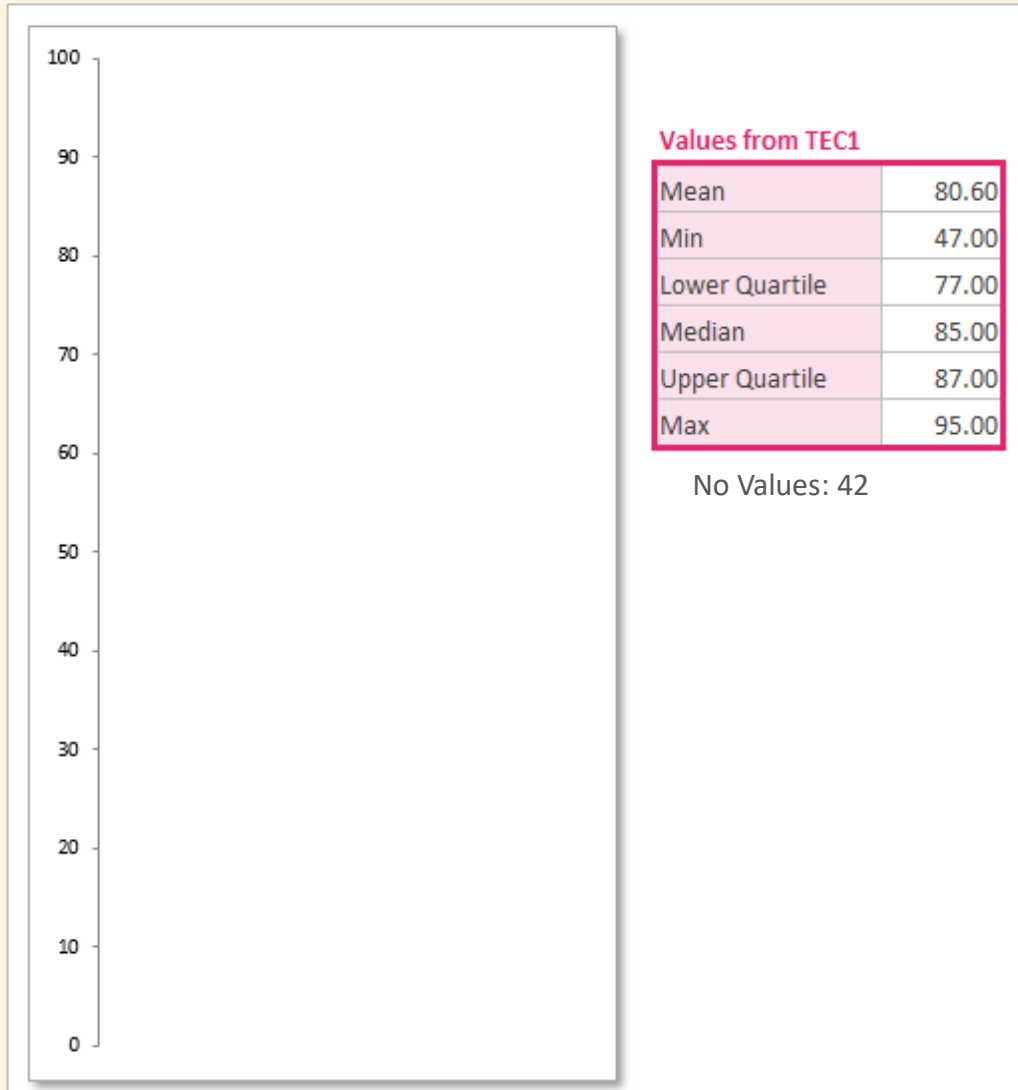**Example**

Below, values calculated from TEC1 have been tabled on the right, and represented in a boxplot on the left.

Values from TEC1

Mean	61.30
Min	12.00
Lower Quartile	50.00
Median	64.50
Upper Quartile	77.00
Max	89.00

No Values: 42

Draw a boxplot based on the values. What information can we ascertain from the boxplot?

100	
90	
80	
70	
60	
50	
40	
30	
20	
10	
0	

Values from TEC1

Mean	80.60
Min	47.00
Lower Quartile	77.00
Median	85.00
Upper Quartile	87.00
Max	95.00

No Values: 42

Answers to 'Try its'

…use boxplots?

- Boxplot charts are used within Scout to represent HSC results for a range of key learning areas. Refer to boxplot charts to see an example of boxplots within an HSC report. The report incorporates all of the information we have covered in this section.

- Boxplots are a really useful way of displaying information on student achievement, as they show many calculations at a glance.
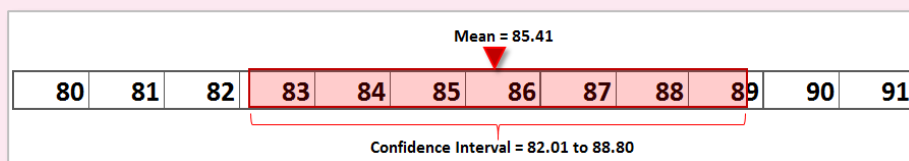
Confidence intervals

A confidence interval is the range of values surrounding an estimate or statistic within which we are fairly confident the true value lies. It describes the uncertainty associated with knowing what the actual mean is, especially when taking into account many variables, such as:

- teacher's judgement on the day of marking
- students not performing well that day due to illness
- students performing particularly well that day, above their usual standard
- broken air-conditioning unit, causing uncomfortable conditions
- maintenance workers making a lot of noise outside the hall in which the exam was taking place, causing disruption.

Important: when calculating the average for a group of students, the average (mean) may be 85.41 but the confidence interval indicates the range in which the average might vary if we re-tested the students several times over a period of time, or the range in which it might vary from year to year with different groups of students. In the following example, the confidence interval is between 82.01 and 88.80.

T-distribution figure
TEC calculates the confidence interval using a t-distribution figure of 95% (the most commonly used percentage). Using this, we can say we are 95% confident that the true average falls between 82.01 and 88.80.

Another way of representing the confidence interval is using a bell curve:
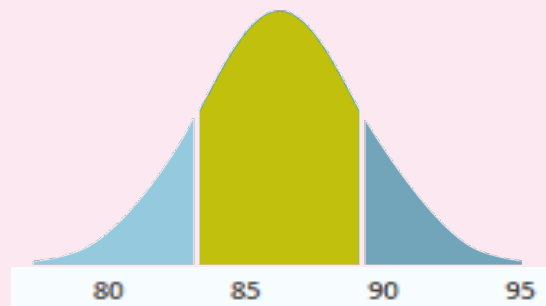
A confidence interval is calculated using multiple steps, including (but not limited to) the calculation of:

- variance
- standard deviation
- square root of number of values
- standard error.

The calculations are shown in Appendix C at the end of this manual.

Enter the following information into TEC1 tab.

78	88	79	82
86	85	84	89
90	86	87	72
86	84	65	77
95	94	92	96
92	92		

What is the:

- mean, median and mode?
- confidence interval range?
- Interquartile range?
- standard deviation?

Using the boxplot tab, what does the boxplot tell you about the range of marks?

_____
_____
_____

Answers to 'Try its'

When do I...?

...use confidence intervals?

- collation of quantitative survey data of specific groups of parents, students or teachers within your school
- collective value added growth of students from one standardised task to another. This will allow you to measure impact of practice across a number of years.
- value added chart within Scout estimates of a school's contribution to student growth, reported with confidence intervals:
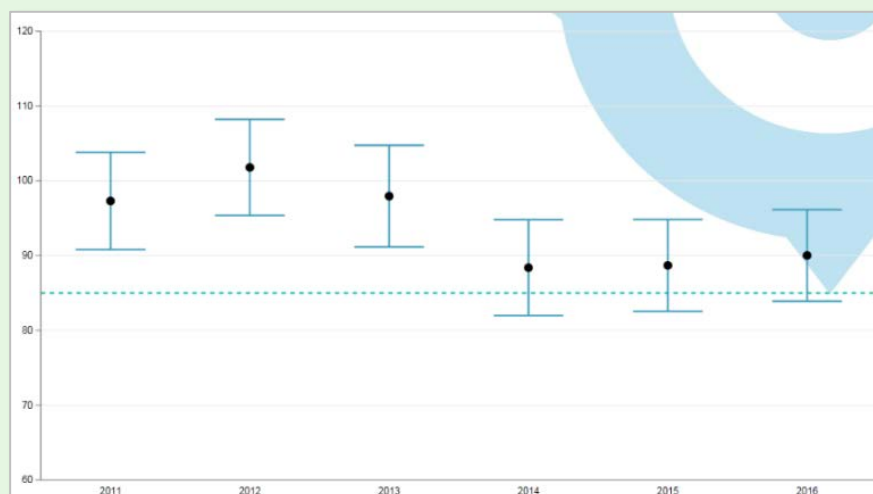

See **Value added** charts for more information.

Standard deviation

Standard deviation is a measure surrounding the mean in which the true average is likely to sit. It is a component within the calculation process for confidence intervals, but can also be used as a stand-alone measure. It does not have the same level of accuracy or certainty as that of a confidence interval.

A low standard deviation indicates that the range is close to the estimated average. A high standard deviation indicates a wider spread. For example, if an average is 35 and the standard deviation is 7, the range of the standard deviation would be split evenly across the average – from **31.5** to **38.5**

| 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 |

Standard deviation

Average
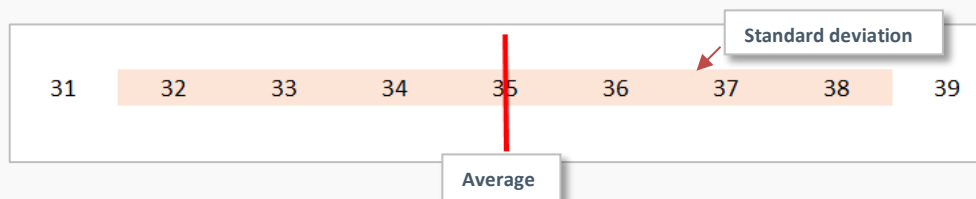
Exploring different types of charts

The following section provides an overview of different types of charts. Each chart provides information about when it could be used, how to interpret it, how to report on and plan with it and how it could be misinterpreted.

The section commences with basic charts fundamentals, followed by a range of charts that have been selected from Scout, Tell Them From Me and SMART. They are:

- bar charts
    - overlapped
    - grouped
    - stacked
- value added
- boxplots
- line charts
- radar charts
- bubble charts
- student growth charts
- sankey diagram
- pie charts.

Basic chart fundamentals

A chart is also called a graph and is a graphical representation of data, in which the data may be represented by bars, lines, slices or bubbles. Different charts are used to display different types of data.

The majority of charts, other than pie charts, have the same common elements. Many can be oriented vertically or horizontally.

X-axis
The x-axis is usually the horizontal scale, running from left to right.

Y-axis
The y-axis is usually the vertical scale.

Gridlines
Charts contain gridlines which display the scale intervals, which assist with being able to read the units.

Dot points
In line graphs, data is usually represented on charts as data points, often connected by lines, bars or columns.   The data points represent the values on the y-axis for the given category for the x-axis.

Lines
The lines in this chart connect the dot points to allow us to clearly see if the information is going up or down.

Trend line
The trend line is an optional element and shows the general trend of the direction the data points seem to be heading. It is useful for describing the relationship between the two variables represented on the graph and for predicting future information.

Data label
Data labels display the value of dot points or bars.

Line chart
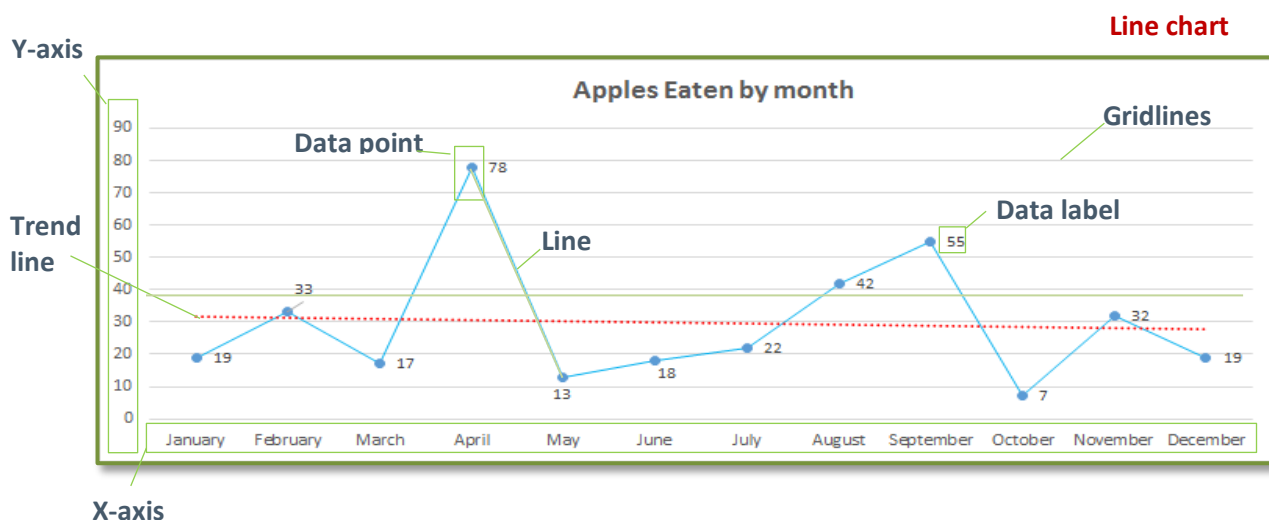
Bar charts (also column charts)
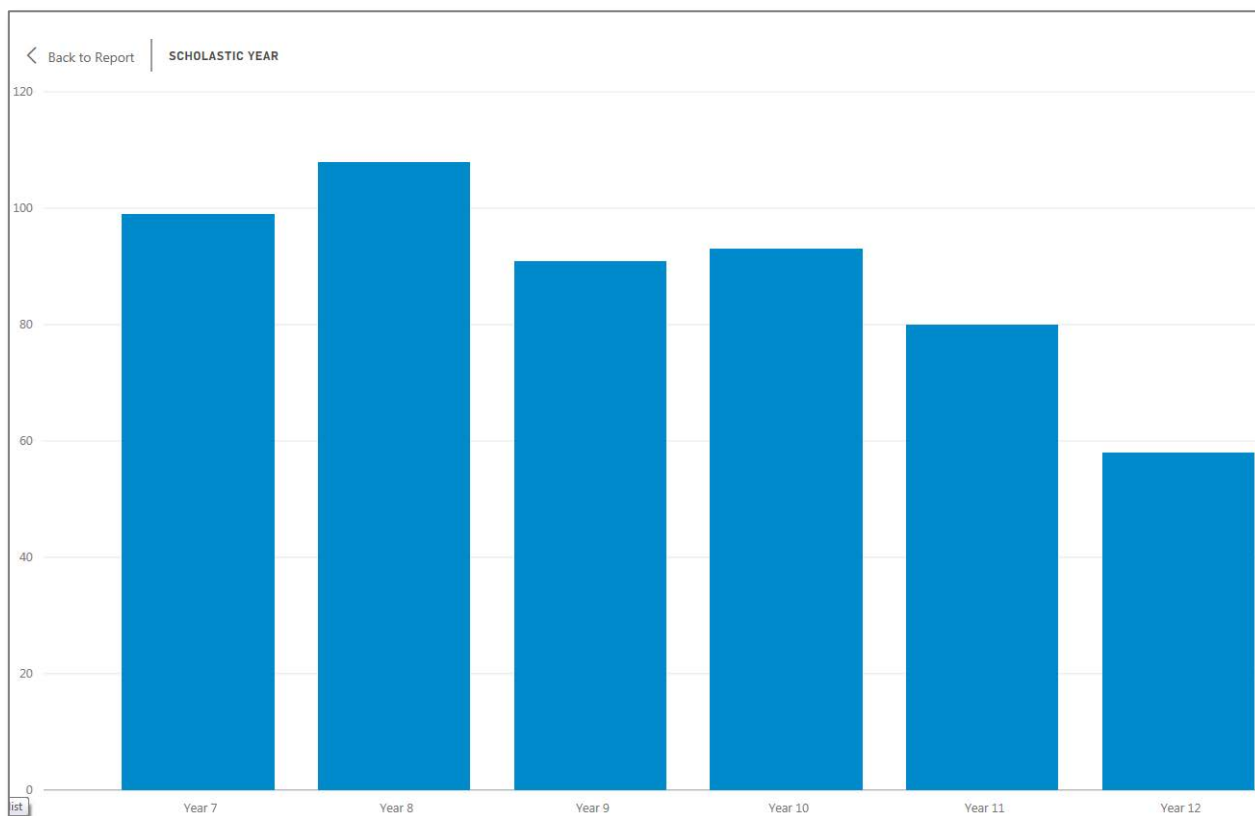

Scout – Current enrolments chart.

Name of charts: bar charts	
When would it be used?	Bar charts display a series of numbers / data as a set of vertical or horizontal bars. They can also be used to represent trends over time.
How to interpret the charts	Bar charts can reflect data changes over a period of time or illustrate comparisons amongst items. The heights of the bars correspond to the value shown on the y-axis.
How to report on and plan with it	Bar charts provide a clear representation of information in a simple format. The higher the bar, the greater the value.
How can it be misinterpreted?	The bars are not labelled with values in the example above and therefore difficult to report on. You may consider adding data labels to bars.
	In a vertical bar chart, you have less space for category labels to display along the x-axis, in which case horizontal bar charts may be more suitable.
	The interpretation of differences between categories or trends over time depends on the scale on the y-axis. Beware interpreting apparently large differences in bar heights when the intervals on the y-axis are small.

Overlapped bar charts
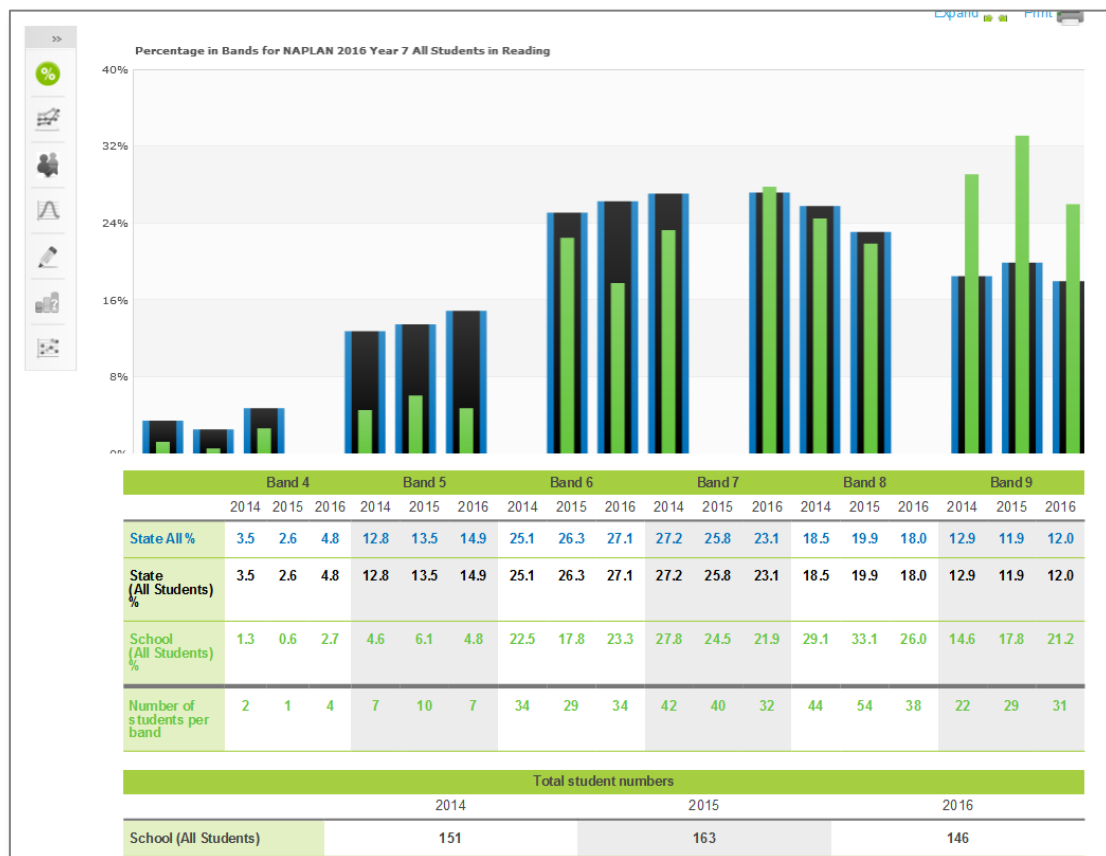
SMART - percentage in bands for NAPLAN.

| Name of charts: overlapped column charts	
When would it be used?	This chart could be used by teachers to:
	• analyse and compare student achievement in NAPLAN over time
	• compare student results with other students across the state and within NSW DoE schools.
How to interpret the charts	In this chart the x-axis shows the NAPLAN bands and the y-axis shows the percentage of students achieving each NAPLAN band.
	• The green column represents the percentage of all students within a school in Year 7 achieving in each of the NAPLAN bands for reading.
	• The blue column represents the percentage of all students in the state achieving each NAPLAN band.
	• The black column represents the percentage of NSW DoE students achieving the band.
	• The charts can be filtered by gender, aboriginality, language background other than English (LBOTE) and English as an additional language or dialect (EAL\D) category to identify achievements of each group of students.
How to report on and plan with it	This report is very useful for mapping trends in student achievement over time and comparing achievement of students across department schools and the state. This information can contribute to the evaluation of existing teaching programs and future planning strategies for all students within the school.

Other types of bar charts

Additional data can be added by using grouped, overlapped or stacked bar charts.

Grouped bar charts
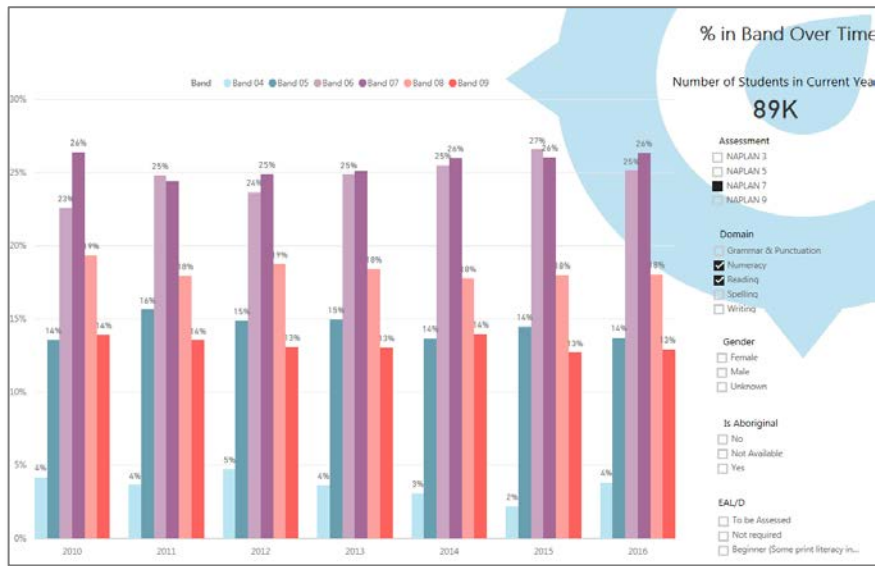
Scout – NAPLAN percentage in bands over time

A grouped bar chart is a good way to show information about sub-groups of the main category.

Different bar colours represent each sub-group, each displayed in the legend on the top of the chart.

In this example, each subgroup represents NAPLAN Bands. It is quite clear to see that there has been a consistent percentage of students within each band across the years.

Stacked bar charts
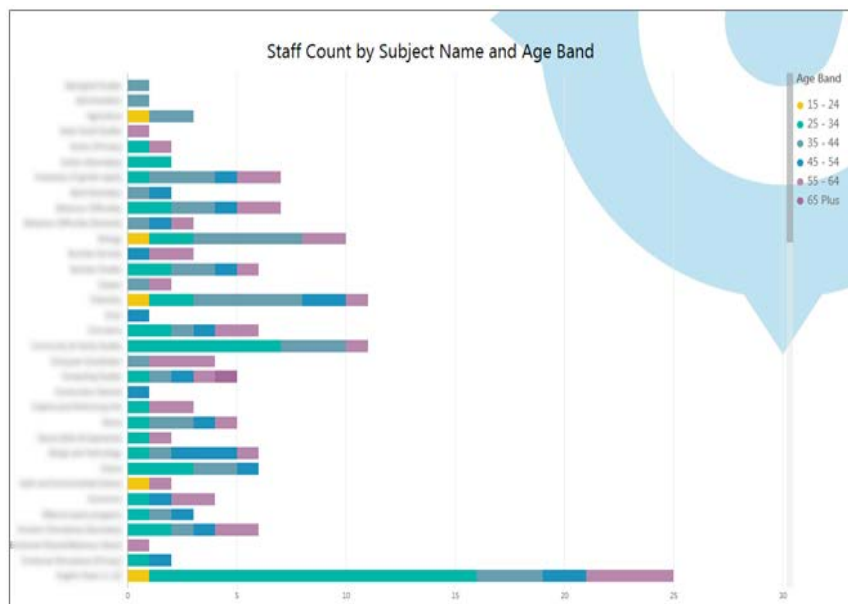
Scout – staff by subject and age

Stacked bar charts place values on top of each other, showing sub-groups on the same bar. Different colours represent different sub-groups, in this instance age bands of staff members accredited to teach each subject.

This type of chart is useful for showing the total size of groups, broken down by sub-groups.

Value added

Scout – value added

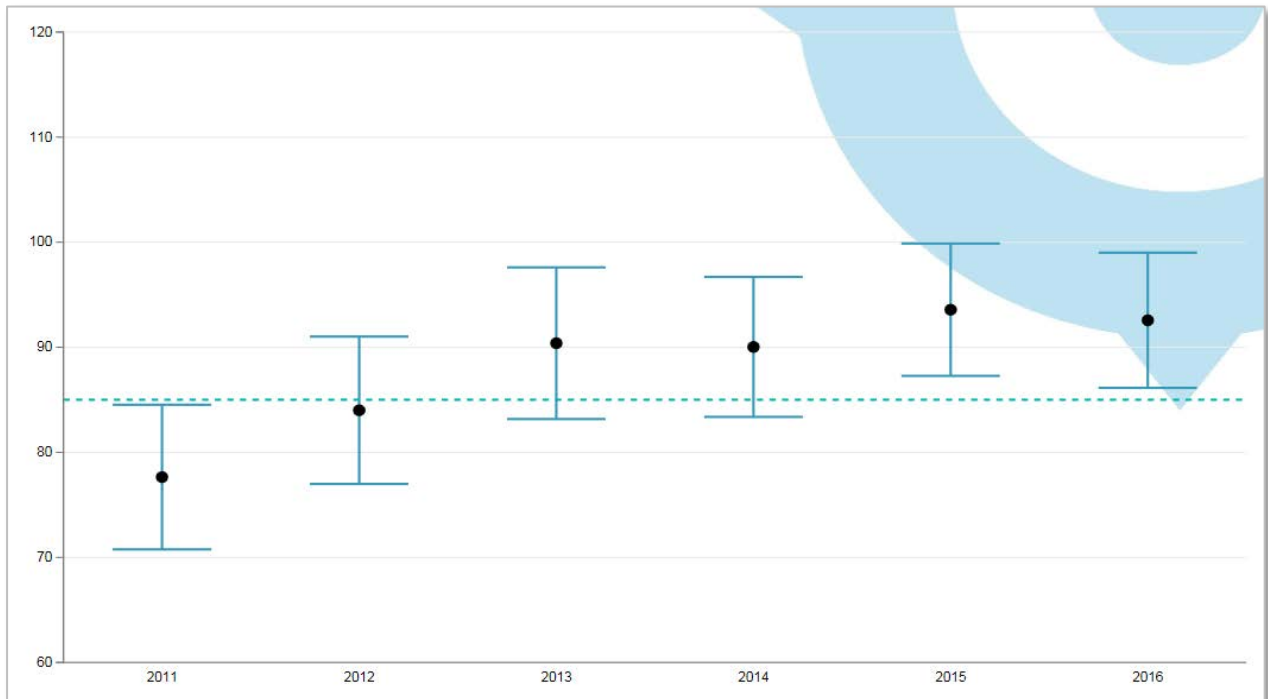| Name of chart: value added	
When would it be used?	The value added report displays information about learning growth of students in schools, between assessment years. For example, Years 3 to 5 or Years 9 to 12.
How to interpret the chart	This value added example shows a school progressing from a status of being 'significantly below average' (whereby the confidence interval falls wholly below the line representing the average school) to 'average' (the confidence interval crosses the line representing the average school) to 'significantly above average' (confidence interval sits above the line representing the average school).
	The dot points are the estimate (mean) of the NAPLAN results. The confidence intervals (or error bars) represent the range in which we are confident the true average lies.
	The dashed line across the chart represents the value added by the average school.
How to report on and plan with it	Principals can use this report to assess value add for their school. It is important to consider the value added for the school in conjunction with other measures of student performance and growth, such as information from SMART, internal measures within a school and school context.
	Value added compares trends year on year. In the example above, you could conclude that from 2011 to 2016 there has been an improvement large enough to justify a claim of statistical significance. If the chart was trending negatively however, it would identify an area of need for improvement which would require planning.

Boxplots

Please also refer to boxplots (box and whisker plots).
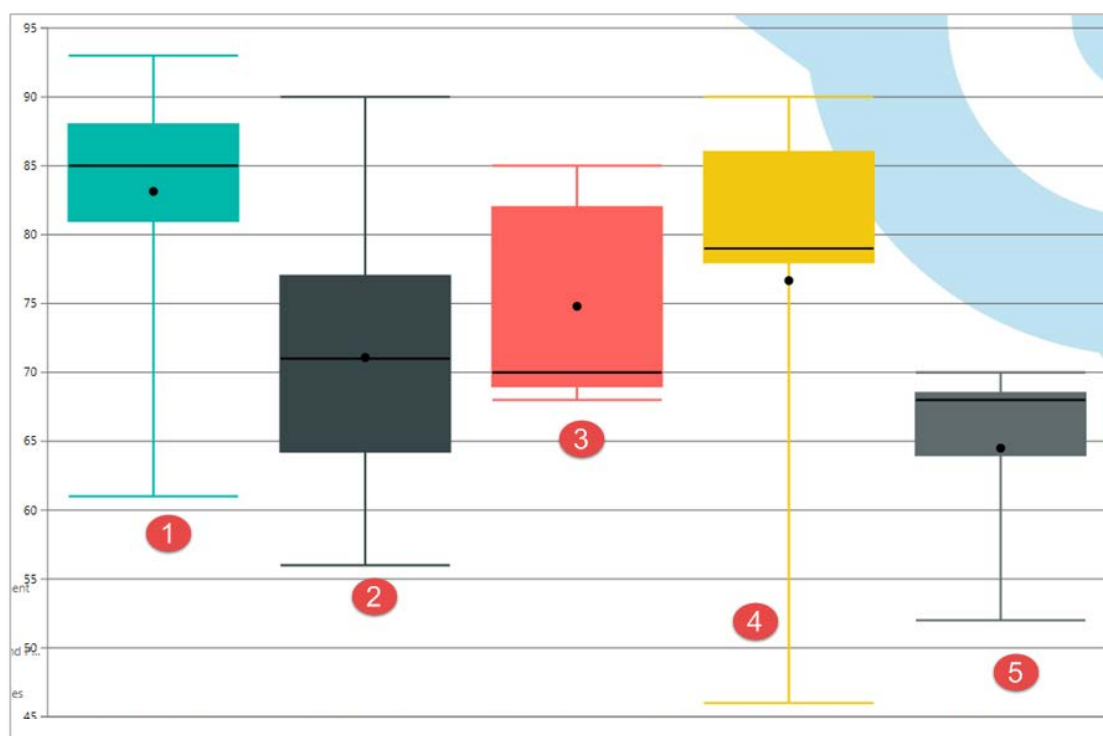
Scout – HSC results

Name of chart: boxplot	
When would it be used?	A boxplot is a convenient way of graphically displaying the quartiles, interquartile range, mean, median, maximum and minimum. This chart can be used to show differences between courses within a key learning area.
How to interpret the chart	A boxplot chart has five main components. The mean is represented by the black dot on each boxplot. The black line is the median and represents the point where 50% of scores fall below and 50% above. The coloured block on each boxplot represents the interquartile range from 25th percentile up to the 75th percentile.
How to report on and plan with it	In the above example the marks are out of 100. Group number 1 has the highest mark of 93 but also, the top 75% of students achieved more than 80 (top 2 NAPLAN bands). The mean being lower than the median suggests that outliers in the lower mark range have had an impact on the mean. This is supported by the fact that the distribution of marks in the lower 25% is a larger range than the highest 75%.
	The middle 50% of students represented in the box number 2 have a relatively large range of marks from 64 to 77. With the mean and median being the same mark of 71 suggests that the distribution of marks in this course are quite even. The lowest 25% of students have a much smaller range of marks that those found in group 1.
	The range of marks in dataset number 3 is quite small from 68 to 85. With the median of 70, 50% of students have achieved in the range between 68 and 70. This represents a large concentration of students within this mark range. The higher mean suggests another large concentration of students in 80 to 85 mark range.
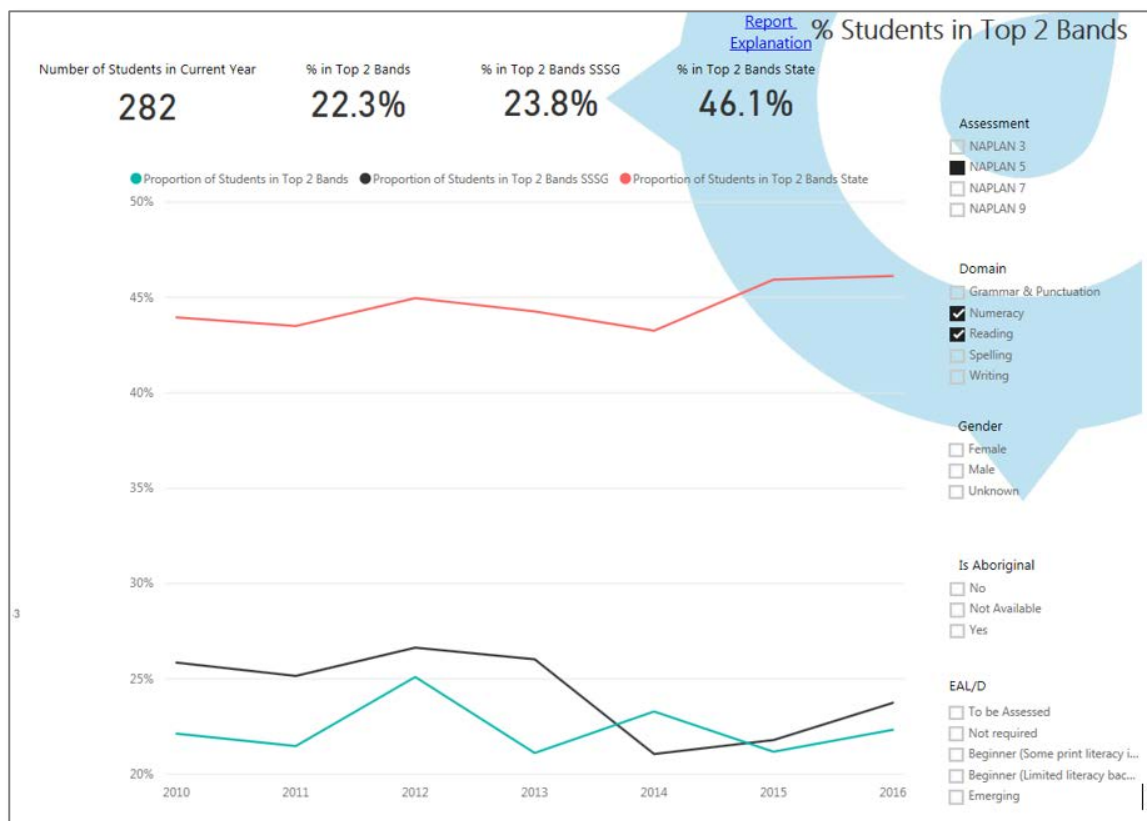
Line charts

Scout – percentage of students in top 2 bands in NAPLAN

Name of chart: line chart	
When would it be used?	A line chart is often used to visualise a trend in data over intervals of time. It is usually drawn chronologically. The above example displays 3 lines, each representing a different series as shown in the legend.
How to interpret the charts	This report provides a high level view of a school's percentage in the top two NAPLAN bands, compared to the percentage for a Statistically Similar School Group (SSSG) and state.
How to report on and plan with it	This report can be very useful to gain an overview of changes in performance, as it highlights differences in results over time. The data in this report can be used to support several School Evidence Framework elements, especially 'Student Performance Measures, Assessment and Reporting' and 'Curriculum and Learning'.
How can it be misinterpreted	Beware of small changes and small numbers! Measures of student performance can be volatile, especially when the number of students represented is small. A positive or negative change for one year to the next may reflect volatility rather than a real change in student performance. To claim a positive or negative trend, you should consider changes over at least 3 years.

Radar charts
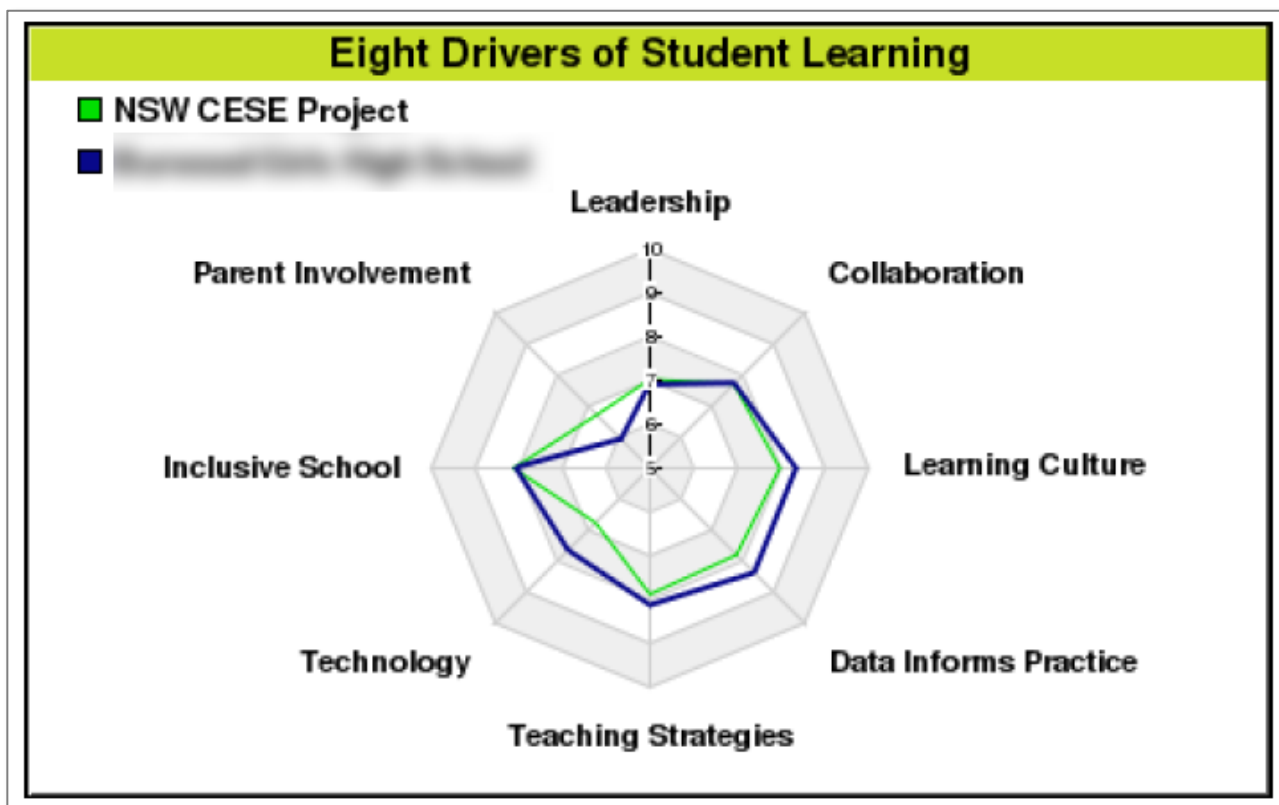
Tell Them From Me - eight drivers of student learning

Name of charts: radar charts	
When would it be used?	A radar chart is a graphical method of displaying data in a two-dimensional chart with three or more variables, represented on axes starting from the same point.
	It can be useful for displaying:
	• staff / school performance data
	• class comparisons.
How to interpret the charts	Results for variables that appear close to the centre indicate a lower effect size compared to results further away from the centre.
How to report on and plan with it	When one result, such as the blue line (school) is compared to another, such as the green line (CESE Project), you can make comparisons of the data. For example, in terms of the Tell Them From Me eight drivers of student learning, this school is higher in Technology compared to the CESE project but lower than the CESE project in parent involvement
How can it be misinterpreted?	Radar charts may not be easy to interpret on first glance. The chart should have clear values for the scale and each of the data points for each category.

Bubble charts
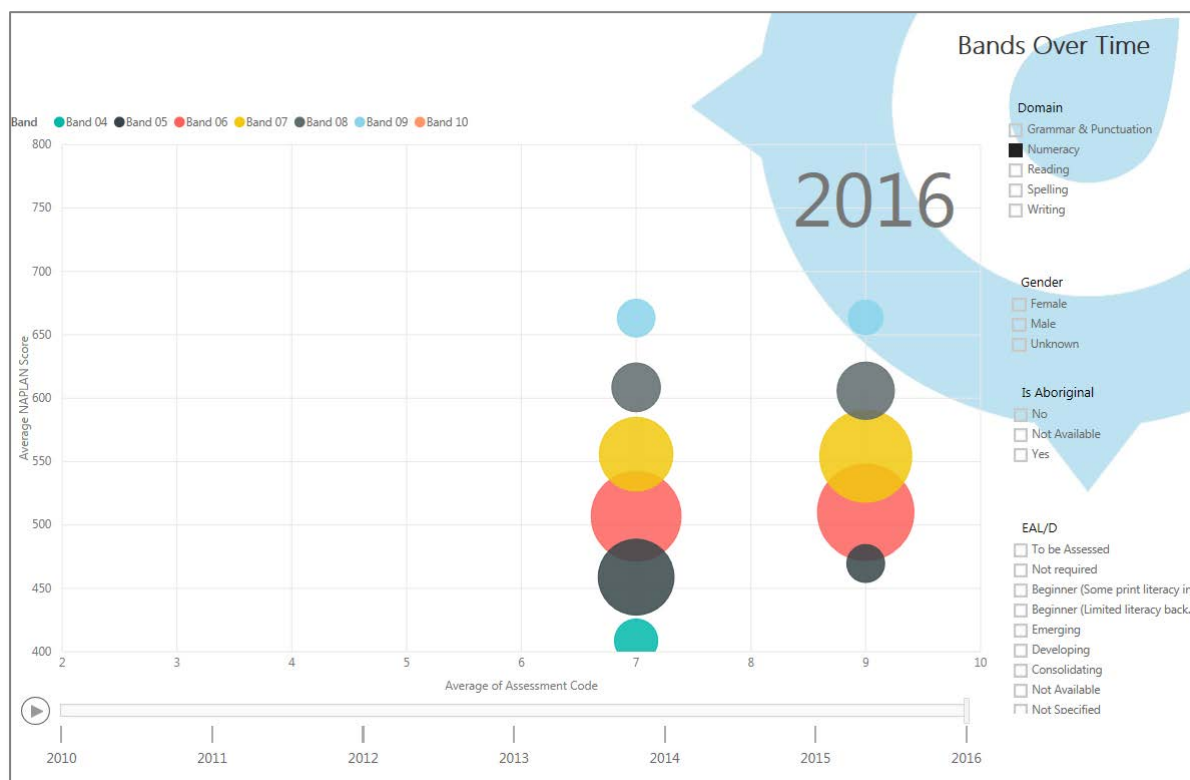
Scout - NAPLAN bands over time

Name of charts: bubble charts	
When would it be used?	A bubble chart is a variation of a scatter graph. Each bubble can display three different measures at the same time, such as average score, quantity of students within a band and NAPLAN year.
How to interpret the charts	The x-axis displays the NAPLAN bands – 3, 5, 7 or 9. This example provides information for Years 7 and 9 NAPLAN Numeracy average score. The red bubble represents band 6 and comprises the greatest number of students in both Year 7 and Year 9. From this chart we can conclude that the majority of Year 9 students achieved a band 6 or band 7 in 2016. The y-axis displays the average NAPLAN score for each student in each band at that school in the selected year.
How to report on and plan with it	The bands over time report can be used to show changes in the distribution of NAPLAN student band results over time. This can be very useful to gain an overview of changes in performance as it highlights any differences in results over time.
How can it be misinterpreted?	Depending on the scaling used to represent the relative size of each bubble, this can change the way the differences between data points are perceived. Data labels or a hovering function could be added to eliminate the possibility of having to guess the values.

Student growth chart

Student Growth for NAPLAN 2016 Year 7 All Students in Reading (Number of students: 131)

Score Year 5 : 528
Score Year 7 : 608
Growth : 80
Expected Growth : 30.3
State Average (All Students) : 38.1
NSW DoE Average (All Students) : 39.2
School Average : 40.1

Students in order of Year 5 achievement

■ 25th Percentile   ■ 75th Percentile

☑ Show Bands
☐ Show Unmatched Students

Average Scaled Score Growth	
Growth in	Scaled score
State (All Students)	38.1
NSW DoE (All Students)	39.2
School (All Students)	40.1

Percentile Ranges	
Percentile	% of students
Less than 25th	17.6%
25th to 75th	44.3%
75th and above	38.2%

Expected Growth	
Growth	% of students
Less than expected growth	35.1%
Greater than or equal to expected growth	64.9%

SMART chart – student growth for NAPLAN

| Name of chart: student growth chart	
When would it be used?	This chart provides data on student growth by showing each student's previous and current scores, connected by a vertical line. The length of the line represents the change or growth students have shown over time. As it is a granular report, individual student growth can be examined more closely.
How to interpret the chart	The y-axis displays the NAPLAN scores; the x-axis represents each student ordered by their previous score.
	The up arrows represent growth. Orange arrows show above expected growth, blue up arrows show below expected growth
	All down arrows are blue and show negative growth (that is, a student scored less than their previous results).
	Information pertaining to each student can be displayed by hovering the mouse over each arrow.
	The red line shows the 25th percentile, the black line the 75th percentile, and therefore the interquartile range is the space between the two.
How to report on and plan with it	It is easy to see areas of improvement or concern that require the development of strategies to support students.

This chart contains a table that provides a clear and concise overview of all the information contained within it.

| **How can it be misinterpreted?** | For the expected growth chart in SMART, the 25th and 75th percentiles are rendered as an approximation. Students appearing on the rendered 75th percentile line may not have achieved growth equivalent to the actual 75th percentile of growth for that prior score. |

Comparing growth between students with different prior scores is not recommended, as growth needs to be interpreted relative to the starting score. For example, growth of 50 score points is well below expected for students with low starting scores, but is well above expected for students with high starting scores.

Sankey diagram
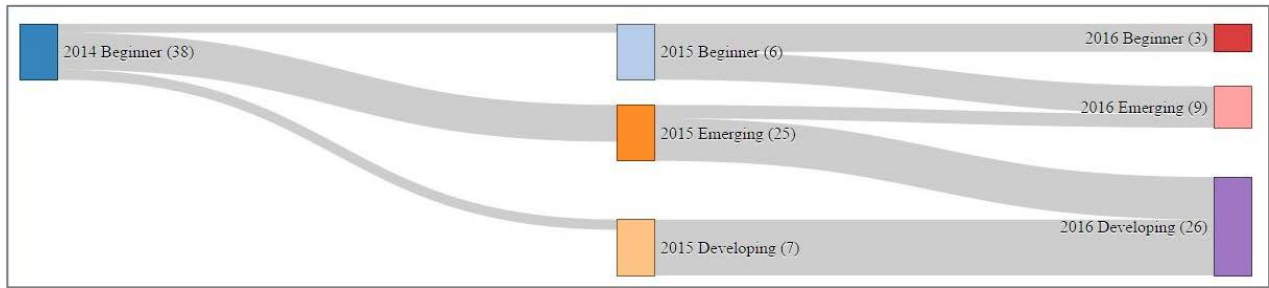
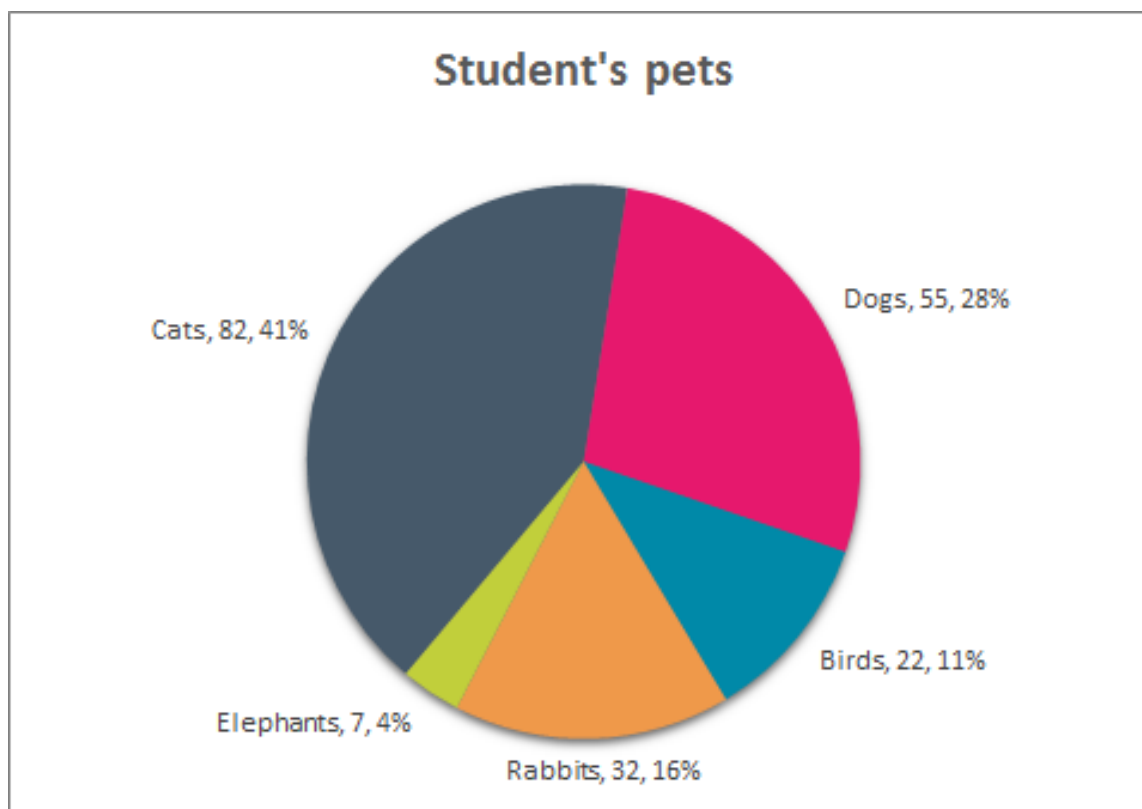| Name of chart: sankey diagram	
When would it be used?	Sankey diagrams are a specific type of flow diagram, in which the width of the arms are proportional to the size of the flow quantity.
How to interpret the charts	This sankey diagram represents English as an Additional Language or Dialect (EAL/D) students classified as level 'beginner' in 2014 and their levels in 2015 and 2016.
How to report on and plan with it	The example above has 3 arms leaving the 2014 beginner point; the total number of students is equal to 38. The widest arm progresses to the 2015 'emerging' point and the other two arms are of equal size and linked to 2015 'beginner' and 2015 'developing' points.
	From this chart we can conclude that 26 students of the original 38 have progressed from beginner stage in 2014 through to developing stage in 2016.

Pie charts

Student's pets

Cats, 82, 41%
Dogs, 55, 28%
Birds, 22, 11%
Rabbits, 32, 16%
Elephants, 7, 4%

| Name of charts: pie charts	
When would it be used?	A pie chart is a circular statistical graphic, divided into slices to illustrate numerical proportion. The angle of each slice is proportional to the quantity it represents. It is named for its resemblance to a pie that has been sliced. It would be used to show percentages or proportional data.
How to interpret the chart	In this example, the largest slice of pie is the grey slice; representing 41% of students with pets have a cat. The smallest piece of pie indicates that 4% of students with pets have an elephant.
How to report on and plan with it	It is recommended that no more than 6 categories are included in the pie chart, as it is more difficult for the eye to distinguish between the relative sizes of the different sectors and can become difficult to interpret without an accompanying table.
How can it be misinterpreted?	It is important that you add data labels to each slice of the pie, to ensure the percentages are not guessed or misinterpreted.

Scenarios

Scenario 1 - Boxplot

What do the boxplots suggest about the performance of students in the two courses?

_____

_____

_____

_____

_____
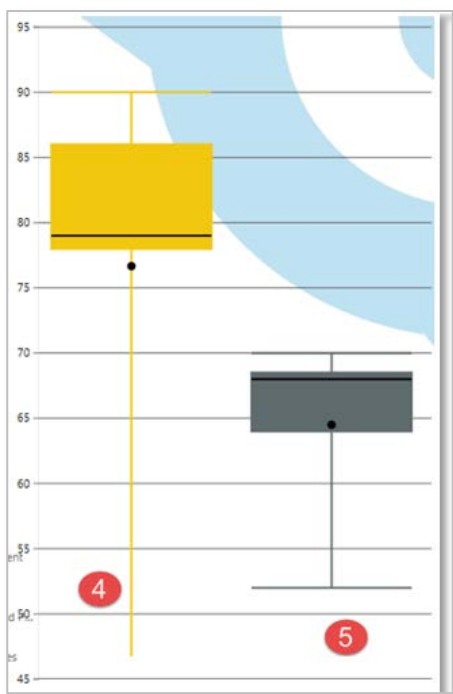
Scenario 2 - charts, student growth

Which type of graph or chart would you use to visually display the reading growth of your students over the course of their education? Justify your choice.

_____

_____

_____

_____

_____

_____

Scenario 3 - distribution

You have collated the results from a formative assessment task. You would like to better understand the distribution of student performance so that you can more effectively plan future learning activities. Outline how you could use calculation methods, charts and/or graphs of student performance to understand this?

_____

_____

_____

_____

_____

_____

_____

_____

_____

_____

Answers to 'Try its'

1. The charts display dot points which represent individual students, so the chart is granular. The ability to view students' names means that it contains personal data.
2. 28.57%
3. 36
4. 37
5. Mode – there are 4 x number 14
6. Mean = 148, median = 118, mode = 118. The median or mode would be the best representation.
7. Lower quartile = 10, middle quartile = 17, upper quartiles = 23
8. Lower quartile = 9 middle quartile = 15 upper quartiles = 19

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2		3	6	7	11	13	15	16	17	21	23	26
3		0	1	2	3	4	5	6	7	8	9	10
4												
5		Lower Quartile		9.0								
6		Middle Quartile		15.0								
7		Upper Quartile		19.0								

9. 48.5
10. The top 75% of students achieved more than 77. The mean is lower than the median which shows that there is an outlier/s in the bottom marks (47). This is supported by the fact that the distribution of marks in the lower 25% is much larger than the top 75%.

Values from TEC1

Mean	80.60
Min	47.00
Lower Quartile	77.00
Median	85.00
Upper Quartile	87.00
Max	95.00
No. values:	**10**

11. **Mean** = 85.41, **median** = 86, **mode** = 86
    Confidence interval range = 82.02 to 88.80
    Interquartile range / area = 82.50 to 91.50
    Standard deviation = 7.66

Scenario answers

Scenario 1

This scenario is a continuation of the boxplot charts section.

Boxplot #4 - the range of marks in this course is quite high. The mean has been brought down by outliers in the lowest 25% of students. The mean and median suggests a large concentration of students in the 75 to 80 range. The top 25% of students are only separated by 4 marks.

Boxplot #5 - This course is the lowest performing course in this set. 50% of students fall within the mark range of 68 and 70. The mean has also been brought down by outliers in the lowest 25% of students.

Scenario 2

A number of different graphs could be used in this circumstance, some more effective than others for communicating the information.

Some graphs may not represent the information effectively, such as the student growth chart which only represents growth from one NAPLAN test to the next and is unable to effectively show the growth of students over the course of their education.

A line graph *could* be used for this scenario; however the limitation of making meaningful sense of multiple lines for individual students can be difficult to comprehend. It may be easier to represent students on their own personalised graph such as a line graph.

Scenario 3

There are a number of key calculations and graphs which would be required in this circumstance.

Calculating the mean would provide you with a measure of the general performance of the group. This might be more or less significant depending on the number and extent of outliers in the distribution of marks.

The median is an additional calculation which will help you understand the exact point at which 50% of students are above or below. This may be more helpful than the mean depending on outliers.

Calculating the standard deviation would be useful for understanding the spread of student performance within the class. The larger the standard deviation, the larger the spread of student performance within the class.

These calculations are also important if you are looking at graphically representing the performance of your students. A line graph which plots the performance of students from lowest to highest would help to visualise the general performance of students. A pie chart or column graph would be helpful to visualise mark or grade distribution of the class. This would help ascertain the number of students who are performing to a particular standard. To further enhance this understanding, a boxplot could also be utilised to understand the spread and grouping of students.

Glossary of terms

A

Aggregate

Aggregate information or charts contain high-level information summarised for groups of data units. They do not tend to reveal personally identifiable information, and may instead provide an overview of examination results, ages, value-added information or statistical information relating to communities.

B

Bar charts (column)

A graph containing horizontal or vertical bars that represent variables.

Boxplot (box and whisker plots)

A type of chart that may contain information on quartiles, mean, median, minimum and maximum.

Bubble charts

A bubble chart plots data in the form of bubbles, the size of which reflects a 3rd dimension of the data (for example, volume of students within a NAPLAN band).

C

Column charts

See bar charts.

Confidence intervals

A range of values in which it is estimated the true value lies.

D

Data

Measurements or observations that are collected as a source of information.

Data item (variable)

A characteristic of a data unit which can be measured or classified (such as gender or exam scores). It is also known as a variable, because the characteristics may vary.

Data unit

A single entity (such as a student, staff member or school). It may also be referred to as a data record, case or a row in a spreadsheet.

Dataset

A collection of all observations.

Denominator

The number on the bottom of a fraction (for example, the 2 in ½). It is important to define what the denominator is when calculating percentages.

Dispersion / distribution

Dispersion refers to the spread or distribution of values for a numerical data item (for example, the distribution of exam scores across a class) or the distribution of data units across categories (for example, percentage of students across NAPLAN bands).

E

Error

A generic term that describes a variety of ways in which data can vary from its true value. Different types of error include sampling error and measurement error.

Estimate

A summary statistic calculated from the data collected (for example, school mean for Year 3 reading scores).

Evidence based decision-making

The process of analysing information from which to make effective decisions.

G

Granular

Granular data/charts contain information on individuals which may include identifiable information.

I

Index

The index (indexed number) is the numeric rank of a percentile.

Interquartile range

The interquartile range is calculated as the difference between the upper quartiles and lower quartile and describes the range of the middle 50% of values when ordered from lowest to highest.

L

Line charts

A line chart or line graph is a type of chart which displays information as a series of data points called 'markers' connected by straight lines.

M

Measures of central tendency

The measures of central tendency are statistical ways to represent the centre of a distribution of values. They are:

- **mean**
  The mean (also known as the arithmetic average) is the sum of the values of each observation in a dataset, divided by the number of observations.
- **median**
  The median is the middle value in distribution when the values are arranged in ascending or descending order. The median is the 50th percentile.
- **mode**
  The mode is a value that appears most often in a dataset.

N

Numerator
The numerator is the number on the top of a fraction (for example, the 1 in ½). It is important to define what the numerator is when calculating percentages.

Numerical data
Numerical data are measured or identified on a numeric scale. The primary methods of reporting this data are measures of central tendency (mean and median) and dispersion (standard deviation and interquartile range).

O

Outliers
Outliers are extreme or atypical data value/s that are notably different from the rest of the data.

Observation
An observation is a measurement or classification made for a data item (variable) that has been recorded about a data unit (individual).

P

Pie charts
A pie chart is a circular statistical graph, divided into slices to illustrate numerical proportions.

Percentages
A percentage (%) is a number or ratio expressed as a fraction of 100.

Percentiles
A percentile is the value below (or above) which a percentage of data falls (for example, the median is the 50th percentile).

Q

Qualitative data
Qualitative data can be obtained from open-ended questions, discussions or observations. It can include diary accounts, unstructured survey questions and classroom observations.

Quantitative data
Quantitative data is any measurement or classification that can be summarised numerically. Numerical summaries include averages, counts, percentages or differences.

Quartiles
Quartiles are the values that divide a list of ordered values into four groups of equal size.

R

Radar charts
A radar chart is a graphical representation of data in the form of a web-type chart. Each spoke represents a different data item (variable).

S

Sankey diagram

Sankey diagrams are a specific type of flow diagram, in which the width of the arms are proportional to the size of the flow quantity.

Scatter plot

A scatter plot is a type of chart which plots data points based on their values against two numeric variables.

Standard deviation

Standard deviation measures the spread or dispersion of the data and is an indicator of the average deviation from the mean. The standard deviation is the square root of the variance.

Standard error

The standard error is an estimate of the error around a mean or other statistical estimate. It is calculated from the standard deviation of the estimate and the number of data units (records) on which the estimate is based.

V

Value added

Value added reflects the contribution over and above other explanatory variables.

Variance

Variance measures the spread or dispersion of the data and is an indicator of the average deviation from the mean. The variance is the square of the standard deviation.

Appendix A - key concepts

FOEI

The Family Occupation and Education Index (FOEI) is a school-level index of educational disadvantage related to socio-economic background. The information is sourced from the parental education and occupation information collected from student enrolment forms and recorded in ERN (Enrolment Registration Number system).

FOEI uses a statistical regression model to produce a weighted combination of school-level parental education and occupation variables based on the extent to which each variable uniquely predicts average school performance.

Visit the CESE website for more information on FOEI.

RAM

The Resource Allocation Model (RAM) was developed to ensure a fair, efficient and transparent allocation of the education budget for every school and recognises that students and school communities have different needs and require different levels of support.

Visit the department's website for more information on RAM.

ICSEA

The Index of Community Socio-Educational Advantage (ICSEA) is a measure that enables meaningful comparisons to be made across schools. It measures key factors that correlate with educational outcomes.

ICSEA uses Australian Bureau of Statistics (ABS) and school data to create an index that best predicts performance on NAPLAN tests. The variables that make up ICSEA include socio-economic characteristics of the small areas where students live (in this case ABS census collection district of about 200 houses), as well as whether a school is in a regional or remote area, and the proportion of Aboriginal students enrolled at the school.

FCI

The Facility Condition Index (FCI) is used to calculate the rating of a school, based on the amount of repair work needed. The calculation for the FCI is the value of the school divided by the estimated cost of repair.

EAL/D

English as an Additional Language or Dialect (EAL/D) students are those whose first language or dialect is not English and they require additional support to assist them to develop proficiency in English.

ERN

Enrolment and Registration Numbers are provided to individual students once they register with a school. The students should retain the same ERN throughout their school years.

Scout

Scout is a suite of business analytics tools, developed to provide school and corporate staff with information about what is working and what can be improved in their school or business area. Scout provides access to data from multiple sources across the department as well as external sources, enabling:

- analysis and support planning
- reporting
- data-driven, informed decision-making

With more data than ever before, Scout is an invaluable resource for school and corporate staff.

For more information, please contact: **business.intelligence@det.nsw.edu.au**

SSSG

Statistically Similar School Groups (SSSG) is calculated in the same way as ICSEA. SSSG for primary, secondary and selective schools include 20 schools above and 20 below, a total of +40.

For central schools, 20 primary and 20 secondary below and 20 primary and 20 secondary above (+80 schools in total).

Appendix B - calculations for confidence intervals

- **Step 1:** Calculate the sum of all values.
  504

- **Step 2:** Count the number of values.
  7

- **Step 3:** Calculate the mean (average) of the total values.
  504 ÷ 7 = 72

- **Step 4:** Calculate the deviation of the individual values from the mean (subtract the mean from each value).
  Minus **72** (Mean) from each of the 7 values

- **Step 5:** Square each individual deviation value (multiply each deviation value by itself and add all values together).
  1820

- **Step 6:** Calculate the mean of the squared deviation to get the variance (divide total of deviation values [1820] by count of values [7]).
  1820 ÷ 7 = 260 = variance

- **Step 7:** Calculate the deviation of the individual values from the mean (subtract the mean from each value).
  Minus **72** (mean) from each of the 7 new values

- **Step 8:** Calculate square root of the variance to get the standard deviation.
  √260 = 16.12 = standard deviation

- **Step 9:** Calculate add and subtract standard deviation (16.12) from the mean (72) to get the range.
  Standard deviation range = 55.88 to 88.12

- **Step 10:** Calculate the standard error (calculate the square root of the count of values)
  √7 = 2.646

- **Step 11:** Divide the standard deviation(16.13) by the standard error (2.646)
  16.13 ÷ 2.646 = 6.095

- **Step 12:** Calculate the t-distribution probability figure (95%) for 7 values and multiply by step 11.
  6.095 x 2.365 = 14.41

Numbers	95%
1	12.71
2	4.303
3	3.182
4	2.776
5	2.571
6	2.447
7	2.365
8	2.306

Confidence interval = 72 (mean) **minus** 14.41 (57.6) and **plus** 14.41 (86.4)

Mean = 72

| 55 | 60 | 65 | 70 | 75 | 80 | 85 | 90 |

Confidence Interval = 57.6 to 86.4