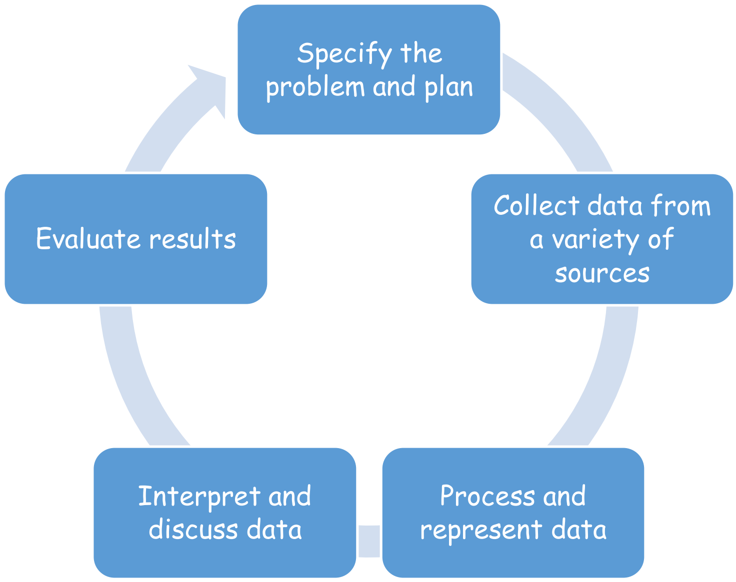
Guide for Statistics – MAC Calculation policy

It is important that graphs and diagrams are drawn on the appropriate paper:

- bar charts and line graphs on squared or graph paper.
- pie charts on plain paper.

Any such work needs to be embedded in the **data handling cycle.**



If learners understand this cycle, then they will see how the work they are doing at home is a part of something bigger, something that will give them the chance to answer questions that they are interested in. Hence the starting point is not 'Let's gather some data' but 'Have we got a problem we want to investigate?

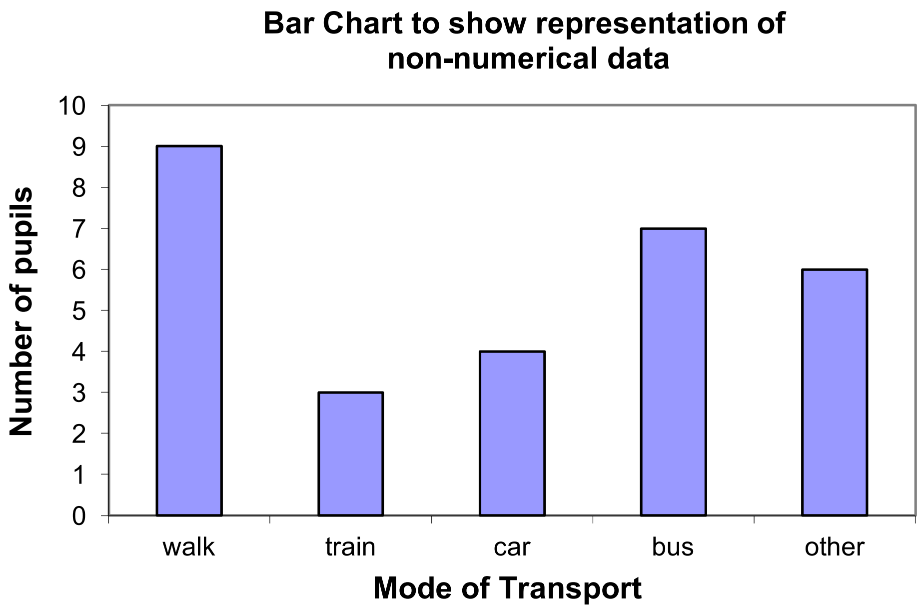
Bar Charts

These are the diagrams most frequently used in areas of the curriculum other than mathematics. The way in which the graph is drawn depends on the type of data to be processed.

Graphs should be drawn with **gaps between the bars** if the data categories are not numerical (colours, makes of car, names of pop star, etc). There should also be gaps if the data is numeric but can only take a particular value – DISCRETE DATA (shoe size, KS3 level, etc). In cases where there are gaps in the graph the horizontal axis will be labelled beneath the columns.

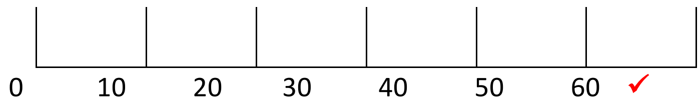
The labels on the vertical axis should be on the lines.

eg.

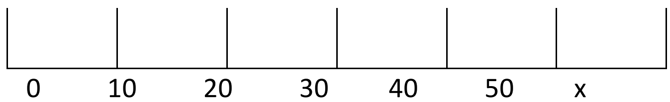


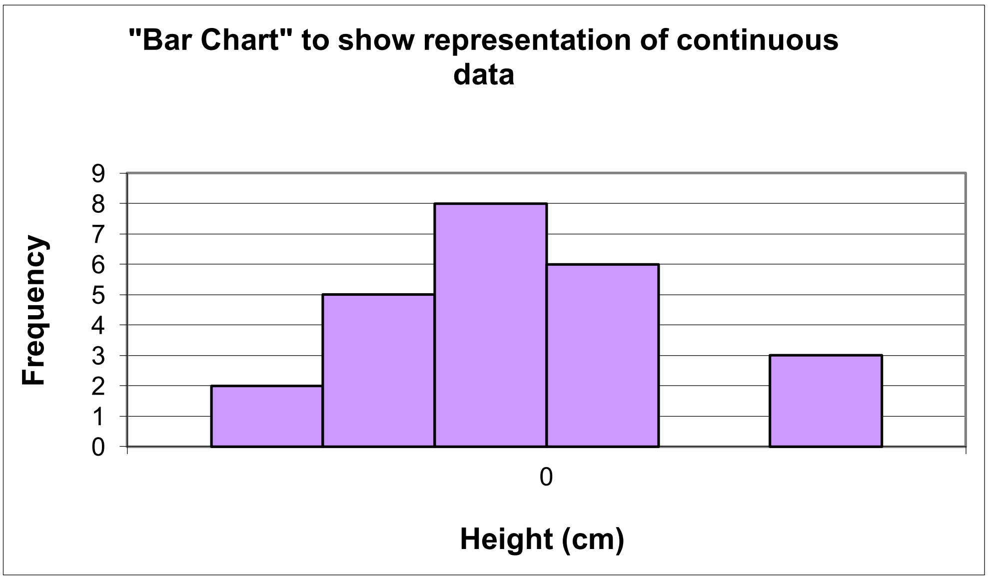
Where the data are CONTINUOUS, eg. lengths, the horizontal scale should be like the scale used for a graph on which points are plotted.

eg



NOT

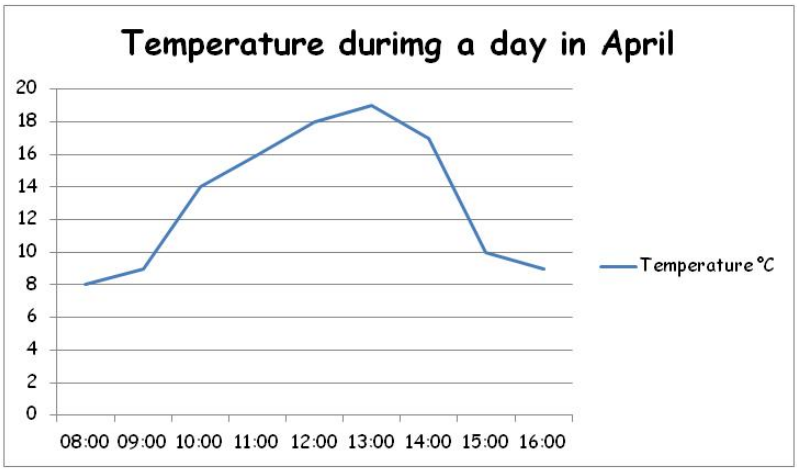




Line Graphs

Line graphs should only be used with data in which the order in which the categories are written is significant.

Points are joined if the graph shows a trend or when the data values between the plotted points make sense to be included. For example the measure of a patient's temperature at regular intervals shows a pattern but not a definitive value.



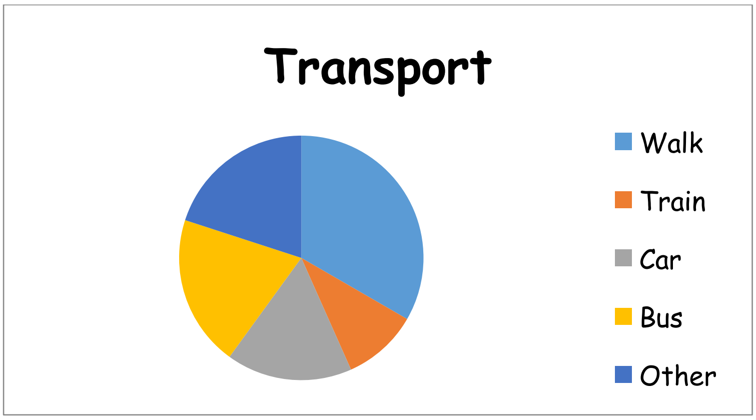
Pie Charts

Pie charts should be used to show how the data is split up between the different categories. The area of the whole circle represents the total number of items.

The way in which pupils should be expected to work out angles for a pie chart will depend on the complexity of the question. If the numbers involved are simple it will be possible to calculate simple fractions of 360°.

eg. The following table shows the results of a survey of 30 pupils travelling to school. Show this information on a pie chart.

Mode of Transport	Frequency	Fraction	Angle
Walk	10	$\frac{1}{3}$	120º
Train	3	$\frac{1}{10}$	36º
Car	5	$\frac{1}{6}$	60º
Bus	6	$\frac{1}{5}$	72º
Other	6	$\frac{1}{5}$	72º
Total	**30**	**1**	**360º**



However, with more difficult numbers which do not readily convert to a simple fraction pupils should first work out the share of 360° to be allocated to **one** item and then multiply this by its frequency.

eg. 180 pupils were asked their favourite core subject.

Each pupils has 360 ÷ 180 = 2° of the pie chart.

Subject	Number of pupils	Pie Chart Angle
English	63	63 x 2 = 126º
Mathematics	75	75 x 2 = 150º
Science	42	42 x 2 = 84º
Total	180	360º

If the data is in percentage form each item will be represented by 3.6° on the pie. To calculate the angle pupils will need to multiply the frequency by 3.6.

eg. 43% will be represented by 43 X 3.6 = 154.8°
≈ <u>155°</u>

Any calculations of angles should be rounded to the nearest degree only at the **final stage of the calculation.** If the number of items to be shown is 47 each item will need:

$360 \div 47 = 7.659574468°$

This complete number should be used when multiplying by the frequency and then rounded to the nearest degree.

Care needs to be taken when using a pair of **compasses**. Students should hold the pivot (not the arms) when drawing a circle to ensure precision. The pencil must be level with the point of the compass.

Ensure when using a **protractor** that students measure from 0º, not 180º (compare to a ruler – you wouldn't measure a line starting from 30cm!)

Using Data

Range

The range of a set of data is the difference between the highest and the lowest data values.

eg. If in an examination the highest mark is 80% and the lowest
mark is 45%, the range is 35% because 80% - 45% = 35%

The range is always a **single number** , so it is **NOT** 45% - 80%

Averages

Three different averages are commonly used:

- **Mean –** is calculated by adding up all the values and dividing by the number of values**.**
- **Median –** is the middle value when a set of values has been arranged in order.
- **Mode -** is the most common value. It is sometimes called the **modal group.**

eg. for the following values: **3, 2, 5, 8, 4, 3, 6, 3, 2,**

Mean = $\frac{3 + 2 + 5 + 8 + 4 + 3 + 6 + 3 + 2}{9} = \frac{36}{9} = 4$

Median – is 3 because 3 is in the middle when the values are put in order.

2, 2, 3, 3, 3, 4, (5), 6, 8

Mode - is 3 because 3 is the value which occurs most often.

Averages from a frequency table

It is often convenient to put data into frequency tables.

- The mode can still be identified as the value with the highest frequency.
- The median can be identified by locating the $(\frac{n+1}{2})^{th}$ value in the frequency table, and which category it falls into.
- The mean can be found using the formula $\frac{\Sigma fx}{\Sigma f}$, where f is the frequency, x is the variable and Σ means "the sum of".

Eg.

Number of goals (x)	Frequency (f)	Goals x frequency (fx)
0	8	0
1	15	15
2	12	24
3	7	21
4	3	12
5	1	5
Total (Σ)	46	77

Mean = $\frac{\Sigma fx}{\Sigma f} = \frac{77}{46}$ = 1.67 goals

Mode = 1 goal (highest frequency)

Median = The total frequency is 46 so the median will be the 23.5^{th} value, that is halfway between the 23^{rd} and 24^{th} value.

There were 8 games with no goals scored.

There were 15 + 8 = 23 games with 0 or 1 goals scored.

This means that the 23^{rd} value is 1, the 24^{th} value is 2, so the median is 1.5 goals.

Averages from a grouped frequency table

Similar rules apply for continuous data and grouped frequency tables, although our results will be less accurate. We can only find the modal class and the median class rather than an accurate mode and median, and we can only calculate an estimate for the mean. As the variable (x) is now a group, it is necessary to use the middle value of each class interval.

Eg.

Speed (s mph)	Frequency (f)	Class width (x)	Speed x frequency (fx)
20 ≤ s < 25	4	22.5	90
25 ≤ s < 30	10	27.5	275
30 ≤ s < 35	12	32.5	390
35 ≤ s < 40	315	37.5	562.5
40 ≤ s < 45	9	42.5	382.5
Total (Σ)	50		1700

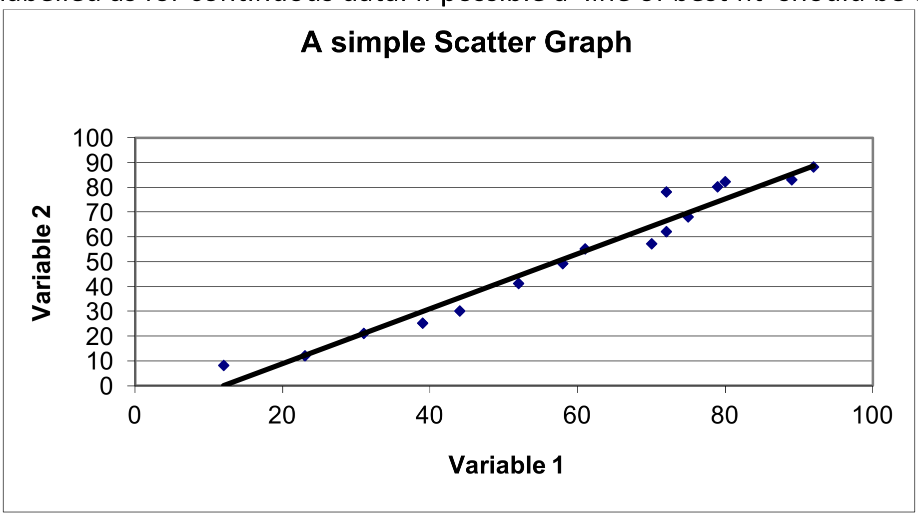
Estimate for the mean = $\frac{\Sigma fx}{\Sigma f} = \frac{1700}{50}$ = 34 mph

The modal class is 35 ≤ s < 40.

The median falls in the class 30 ≤ s < 35.

Scattergraphs

These are used to compare two sets of numerical data. The two values are plotted on two axes labelled as for continuous data. If possible a 'line of best fit' should be drawn.



The degree of correlation between the two sets of data is determined by the proximity of the points to the 'line of best fit'

The above graph shows a positive correlation between the two variables. However you need to ensure that there is a reasonable connection between the two, e.g. ice cream sales and temperature. Plotting use of mobile phones against cost of houses will give two increasing sets of data but are they connected?

Negative correlation depicts one variable increasing as the other decreases, no correlation comes from a random distribution of points. See diagrams below.