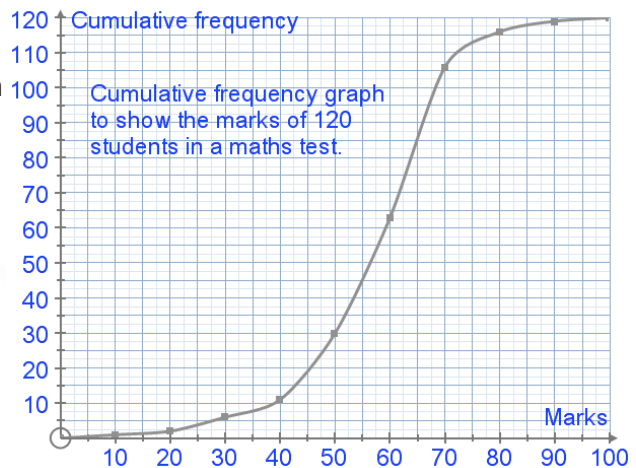


## Statistical diagrams

### Starter

1. **(Review of GCSE material)**  
 Consider the cumulative frequency for the number of marks gained by a year group in a maths exam.
- Estimate the median.
  - Estimate the interquartile range.
  - Calculate the percentage of students who scored between 45 and 75 marks



### Notes

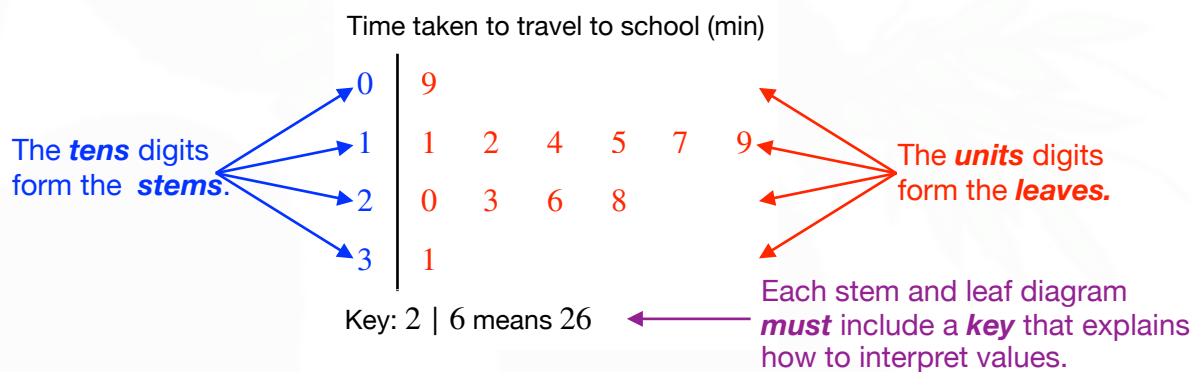
#### Stem and leaf diagrams

A **stem and leaf diagram** is a great way to express a **small amount** of **raw data** clearly when each data value must be included.

The data values for the time to travel to school of a group of students is below...

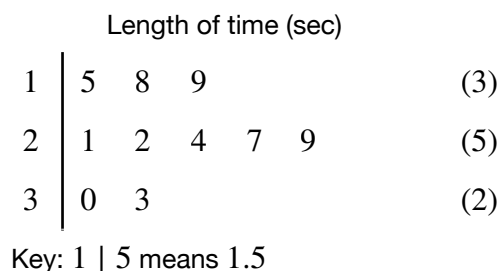
12, 26, 9, 17, 14, 28, 11, 31, 23, 19, 15, 20

...written in a stem and leaf diagram would look like this:



Since the values are put in ascending order in each row, it is called an **ordered stem and leaf diagram**. When asked to create a stem and leaf diagram, assume it must be an ordered stem and leaf diagram.

A number in brackets is **sometimes** written at the end of the each row — this indicates how many data values there are in each row, which is useful when there are lots of values.



- E.g. 1** (a) The weights of a group of students were recorded as:  
55, 67, 43, 46, 51, 81, 64, 52, 74, 57, 63, 68, 57, 77  
Express this data points in a stem and leaf diagram.
- (b) State the value of the mode for the data.
- (c) Use your diagram to state the median for the data.

**Back-to-back stem and leaf diagrams**

Two sets of data values can be compared using a **back-to-back** stem and leaf diagram.

Below shows the heights of a sample of students from Year 8 and Year 9.

Year 8					Year 9			
7	6	3	0	15	9	9		
8	5	4	4	16	1	3	8	
		7	5	17	0	3	4	7
			0	18	2	5	6	8
				19	2			

Key: 15 | 9 means 159

**N.B.** The values on the left-hand side of the diagram are in descending order i.e. the digits closest to the centre are the smallest.

**E.g. 2** Answer these questions on the back-to-back stem and leaf diagram above.

- (a) State the height of the largest student in Year 9.
- (b) State the height of the smallest student in Year 8
- (c) Comment on how the table would look different if samples of Year 12 and Year 13 students were taken. Think of two ways it could be different. Give a reason for your answer.

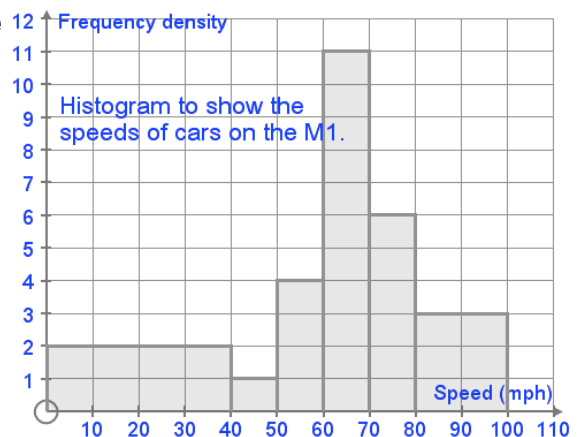
**Histograms**

**Area of bar = frequency of the group**

The formula for the vertical scale, **frequency density** is:  $\text{Frequency density} = \frac{\text{Frequency}}{\text{Class width}}$

**E.g. 3** Consider the histogram to the right showing the speeds of cars on the M1.

- (a) Calculate the number of cars that were travelling between 50 and 60 mph.
- (b) Calculate the percentage of cars that were breaking the speed limit of 70 mph.
- (c) Calculate the percentage of cars that were travelling less than 50 mph.
- (d) Calculate an estimate for the mean of the data.



### Cumulative frequency graphs

When data is presented as table of grouped values (e.g.  $10 \leq h < 20$ ) a **cumulative frequency graph** can be drawn to **estimate** the median and quartiles. In the AS or A2 examination, it is unlikely you'll be asked to draw a cumulative frequency curve due to time constraints. The method is provided here for completeness.

Success Criteria — drawing a cumulative frequency curve:

1. Add an extra column to the table and find the cumulative frequency.
2. Draw the axes — **cumulative frequency** is always on the **vertical axis**. Label your axes, make sure there is a suitable title on the axes.
3. The line must start from the horizontal axis. The **left-hand value** of the **first interval** is the **first number** on the **horizontal axis**.
4. Plot the points at the **end of each interval**.
5. Draw **one smooth curve** through **all the points**.
6. Give your graph a **title** "Cumulative frequency graph to show...".

**N.B.** The **vertical axis** always starts at **zero**  
The **horizontal axis** can start at the **left-hand value** of the **first interval in the table**.  
The interquartile range considers the **middle 50%** of the data.

$$\text{Median, } Q_2 = \frac{1}{2}n \text{ th value}$$

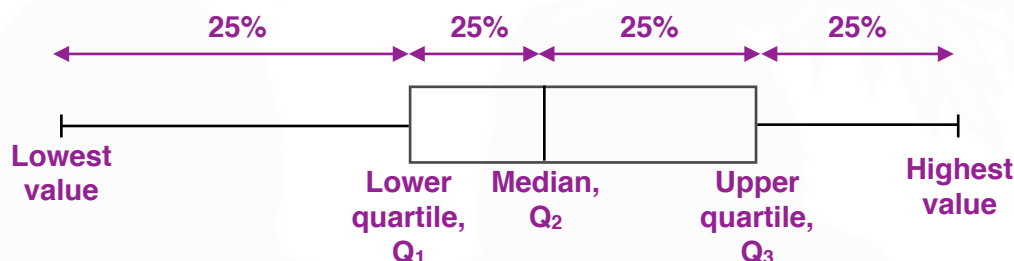
$$\text{Lower quartile, } Q_1 = \frac{1}{4}n \text{ th value}$$

$$\text{Upper quartile, } Q_3 = \frac{3}{4}n \text{ th value}$$

$$\text{Interquartile range, IQR} = \text{Upper quartile} - \text{Lower quartile} = Q_3 - Q_1$$

### Box and whisker plots

A box plot is a simple diagram that displays key information.



A box plot should be drawn accurately and the end points of the lines and boxes must have the numbers written on them. You might find it useful to draw a scale under your diagram.

**N.B.** If the data is grouped (i.e.  $10 \leq h < 20$ ), the lowest and highest values are not known. In which case, the highest and lowest **possible** values are used.

Box plots are often used to **compare data** and may appear in large data set questions.

When comparing data we need to make two comments

1. State which group has the highest **central tendency** — i.e. the highest **median** (or mean).
2. State which group has the least **spread** — so use the **interquartile range** (or range if the interquartile ranges are equal).

**Video:** [Stem and leaf diagrams](#)  
**Video:** [Box and whisker plots](#)

Video: [Histograms](#)  
Video: [Histograms - finding the height of a bar](#)  
Video: [Median, quartiles & IQR for discrete data](#)  
Video: [Cumulative frequency curves](#)  
Video: [How to interpret class intervals](#)

Exam questions: [Box and whisker plots](#)  
Exam questions: [Stem and leaf & box and whisker plots](#)  
Exam questions: [Histograms](#)  
Exam questions: [Histograms - finding the height of a bar](#)  
Exam questions: [Median, quartiles & IQR for discrete data](#)  
Exam questions: [Estimating the median from a histogram](#)

[Solutions to Starter and E.g.s](#)

### Exercise

p329 16A Qu 1A, 2i, 3-10, (11 red)

### Summary

#### **Stem and leaf diagrams**

A **stem and leaf diagram** is a great way to express a **small amount** of **raw data** clearly when each data value must be included.

#### **Histograms**

**Area of bar = frequency of the group**

Vertical scale, **frequency density** is:  $\text{Frequency density} = \frac{\text{Frequency}}{\text{Class width}}$

#### **Cumulative frequency**

The interquartile range considers the **middle 50%** of the data.

$$\text{Median, } Q_2 = \frac{1}{2}n \text{ th value}$$

$$\text{Lower quartile, } Q_1 = \frac{1}{4}n \text{ th value}$$

$$\text{Upper quartile, } Q_3 = \frac{3}{4}n \text{ th value}$$

$$\text{Interquartile range, IQR} = \text{Upper quartile} - \text{Lower quartile} = Q_3 - Q_1$$

#### **Box and whisker plots**

A box plot is a simple diagram that displays key information.

