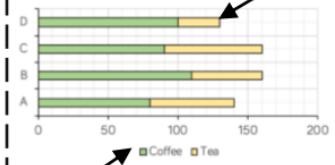
Bar and line charts

Compare the bars green compared to yellow. The size of each bar is the frequency
Overall total easily comparable

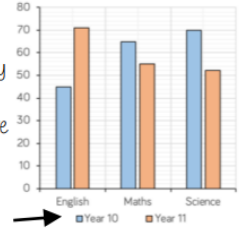
Composite bar charts



Categories clearly indicated

Dual bar charts

Bars are compared side by side
Easier to compare subgroups



Categories clearly indicated

Linear Correlation R

Positive Correlation

As one variable increases so does the other variable

Negative Correlation

As one variable increases the other variable decreases

No Correlation

There is no relationship between the two variables

Speed, Distance, Time

minutes ÷ 60 hours × 60

Before calculations – make sure you are working in the same units as the speed

Learn or learn how to rearrange the formular for speed, distance and time

$$\text{time} = \frac{\text{distance}}{\text{speed}}$$

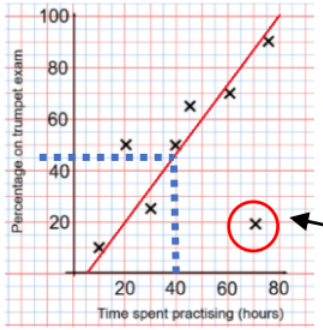
$$\text{distance} = \text{speed} \times \text{time}$$

Substitute in the variables given

Using a line of best fit R

Interpolation is using the line of best fit to estimate values inside our data point

e.g. 40 hours revising predicts a percentage of 45



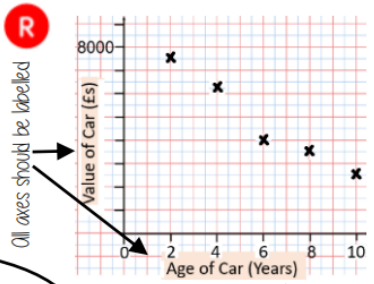
Extrapolation is where we use our line of best fit to predict information outside of our data
This is not always useful – in this example you cannot score more that 100%. So revising for longer can not be estimated

This point is an "**outlier**"
It is an outlier because it doesn't fit this model and stands apart from the data

Draw and interpret a scatter graph R

Age of Car (Years)	2	4	6	8	10
Value of Car (£s)	7500	6250	4000	3500	2500

- This data may not be given in size order
- The data forms information pairs for the scatter graph
- Not all data has a relationship



All axes should be labelled

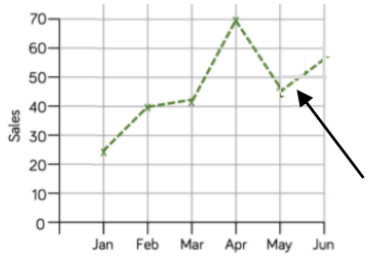
"This scatter graph show as the age of a car increases the value decreases"

The link between the data can be explained verbally

The axis should fit all the values on and be equally spread out

Time-Series R

This time-series graph shows the total number of car sales in £1000 over time

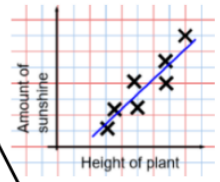


Look for general trends in the data. Some data shows a clear increase or a clear decrease over time

Readings in-between points are estimates (on the dotted lines). You can use them to make assumptions

The line of best fit R

The Line of best fit is used to make estimates about the information in your scatter graph



Things to know:

- The line of best fit **DOES NOT** need to go through the origin (The point the axes cross)
- There should be approximately the same number of points above and below the line (It may not go through any points)
- The line extends across the whole graph

It is only an estimate because the line is designed to be an average representation of the data

It is always a **straight line.**

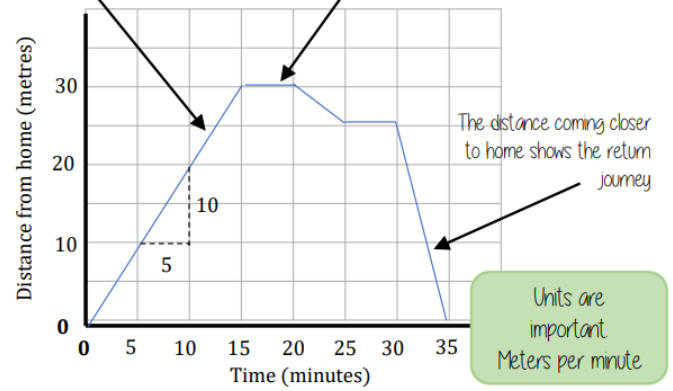
Distance – Time graphs

Gradient = speed

The steeper a gradient the faster the speed

$\frac{10}{5}$ = 2 metres per min

Horizontal lines represent staying still



The distance coming closer to home shows the return journey

Units are important
Meters per minute

Southmoor Academy
ASPIRE · ACHIEVE · ENJOY

Histograms

A histogram is similar to a bar chart, but where a bar chart is used for categorical or discrete data, we use a histogram for continuous data e.g. heights, weights, time etc.

Key features:

- There are no gaps between bars and bars may be different widths
- The horizontal scale is linear and not grouped
- The vertical axis is labelled **frequency density**
- The **frequency** is represented by the **area of each bar** rather than the height of each bar

e.g. Draw a histogram of the following data

Length of time	Frequency
$0 \le t < 10$	60
$0 \le t < 15$	40
$15 \le t < 20$	75
$20 \le t < 50$	150

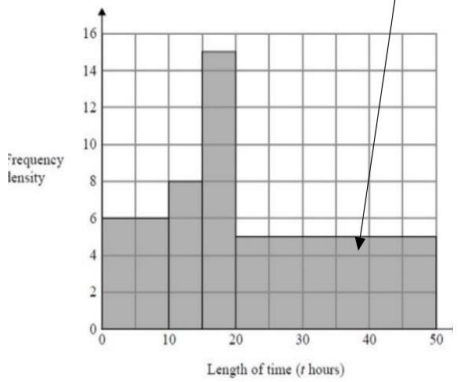
First we need to calculate the frequency density

$$\text{Frequency density} = \frac{\text{Frequency}}{\text{Class width}}$$

Length of time	Frequency	Frequency density
$0 \le t < 10$	60	$60 \div 10 = 6$
$10 \le t < 15$	40	$40 \div 5 = 8$
$15 \le t < 20$	75	$75 \div 5 = 15$
$20 \le t < 50$	150	$150 \div 30 = 5$

Class width is the difference between the 2 bounds so this one is $50 - 20 = 30$

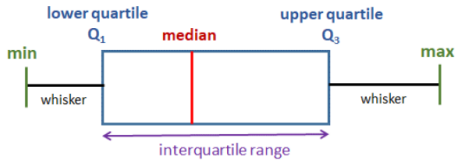
The area of this bar is 150 which is the frequency of the group $20 \le t < 50$



Boxplots (or box and whisker diagrams)

A boxplot is used as a visual representation of the spread of data. It shows the smallest value, largest value, lower quartile, upper quartile and median. The actual box represents the spread of the middle 50% of the data which is known as the **interquartile range**. The first 25% of the data is the first whisker and the final 25% of the data is the second.

Interquartile range = upper quartile – lower quartile.



Cumulative Frequency graphs

A cumulative frequency table shows a running total of the frequencies. A cumulative frequency diagram or graph, is drawn by **plotting** the **cumulative frequency** against the **upper boundary** of the class interval and then joined together.

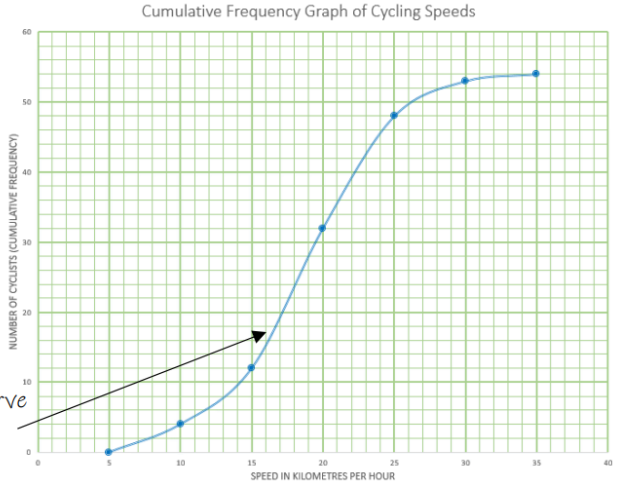
e.g. Plot a cumulative frequency diagram of the following data

Speed of cyclists	Frequency	Cumulative frequency
$5 \le t < 10$	4	4
$10 \le t < 15$	8	$4 + 8 = 12$
$15 \le t < 20$	20	$12 + 20 = 32$
$20 \le t < 25$	16	$32 + 16 = 48$
$25 \le t < 30$	5	$48 + 5 = 53$
$30 \le t < 35$	1	$53 + 1 = 54$

Start by calculating the cumulative frequency

Plot the upper bound of the groups against cumulative frequency so (10, 4), (15, 12) and so on.

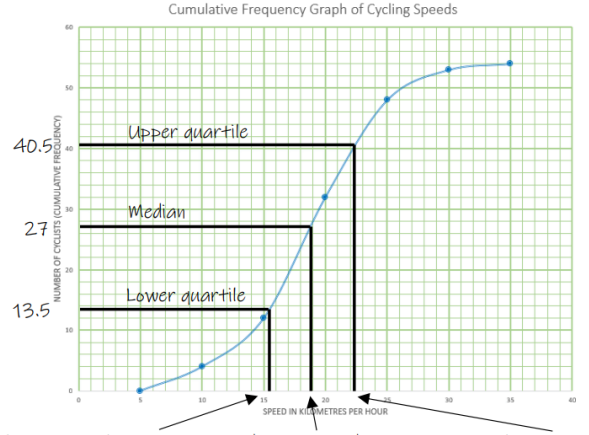
Join as a smooth curve or as straight lines between each point



A cumulative frequency graph can be used to estimate the lower quartile, median and upper quartile of grouped data.

Find one quarter of the total cumulative frequency, (in this case $54 \div 4 = 13.5$), one half of the total cumulative frequency ($54 \div 2 = 27$) and three quarters of the total cumulative frequency ($13.5 \times 3 = 40.5$).

Draw a line across from each of these points until they hit the curve, then go down to the horizontal scale and read off.



The lower quartile is 15.5, the median is 19 and the upper quartile is 24.5