If you're planning on analysing your data using a statistical test, you first need to come up with a **null hypothesis** — this is a special type of hypothesis that states there is no significant difference (or correlation) between the things you're investigating. You then collect data to try to disprove the null hypothesis before analysing it statistically. There's more on null hypotheses on pages 257 and 393.

Student's t-test

You can use the Student's t-test when you have two sets of data that you want to compare. It tests whether there is a significant difference in the means of the two data sets. The value obtained is compared to a critical value, which helps you decide how likely it is that the results or 'differences in the means' were due to chance. If the value obtained from the t-test is greater than the critical value at a probability (**P value**) of 5% or less (≤ 0.05), then you can be 95% confident that the difference is significant and not due to chance. This is called a **95% confidence limit** — which is good enough for most biologists to reject the null hypothesis.

Chi-squared test

You can use the chi-squared test when you have categorical (grouped) data and you want to know whether your observed results are statistically different from your expected results. You compare your result to a critical value — if it's larger than the critical value at $P = 0.05$, you can be 95% certain the difference is significant. There's more on chi-squared on pages 393-395.

Correlation coefficient

A correlation coefficient allows you to work out the degree to which two sets of data are correlated (see page 15 for more on correlation). It is given as a value between 1 and –1. A value of 1 indicates a strong positive correlation, 0 means there is no correlation and –1 is a strong negative correlation. You can then compare your result to a critical value to find out whether or not the correlation is significant. The **Spearman's rank correlation coefficient** is an example of a correlation coefficient. See pages 256-257 for how to use it.

Exam Tip
In the exams, you could be asked which statistical test you'd use to analyse some data and to explain why you'd use it, as well as to interpret the test result.

Exam Tip
When you're talking about the results of a statistical test and using the 95% confidence limit, make sure you refer to the probability as less than 0.05 or 5%, not 0.05%.

Tip: If the result of your statistical test is greater than the critical value at a P value of less than 2% (< 0.02), or even 1%, you can be even more confident that the difference is significant.

Tip: Make sure you're familiar with the symbols > (greater than), >> (much greater than), < (less than) and << (much less than).

4. Presenting Data

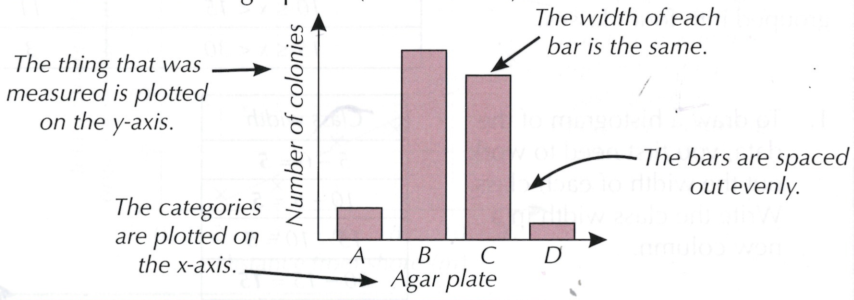
Presenting your data can make it easier for you to understand your results and spot any trends. You need to choose the best way to present your data.

Qualitative and discrete data

Qualitative data is non-numerical data, e.g. blood group. **Discrete** data is numerical data that can only take certain values in a range, e.g. number of patients. You can use **bar charts** or **pie charts** to present these types of data.

Example

The bar chart below shows the number of bacterial colonies present on a series of agar plates (labelled A-D).



Tip: Qualitative data can also be called categorical data — all the data can be sorted into categories and values between categories don't exist.

Exam Tip
If you are asked to draw a graph or chart in your exam, don't forget to label the axes (including the quantity and units), choose a sensible scale and make sure that it covers at least half the graph paper.

Tip: The graph on the right is a line graph. Line graphs look a bit like scattergrams (see next page), but the points on line graphs are joined together.

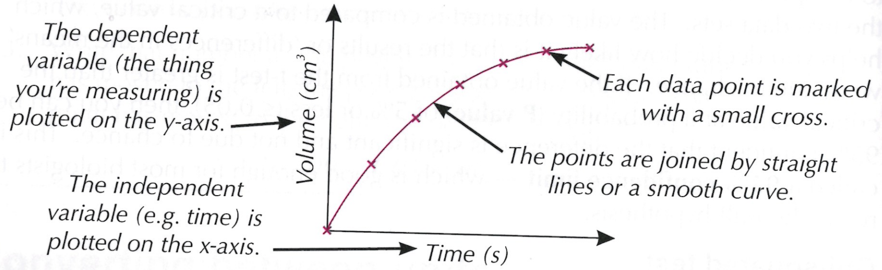
Continuous data

Continuous data is data that can take any value in a range, e.g. height or weight. You can use **line graphs** or **histograms** to present this type of data.

Line graphs

Line graphs often show how a variable changes over time. The data on both axes is continuous.

Example



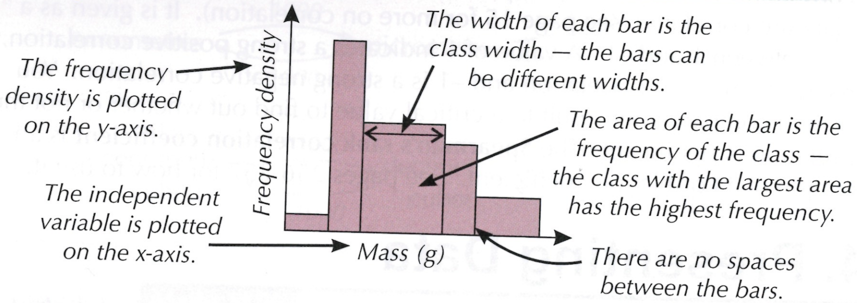
Tip: Don't be fooled by the height of the bars in a histogram — the tallest bar doesn't always belong to the class with the greatest frequency.

Tip: The data for a histogram is split into groups called classes, rather than categories. The class width is the range of the class.

Histograms

Histograms are a useful way of displaying frequency data when the independent variable is continuous. They may look like bar charts, but it's the area of the bars that represents the frequency (rather than the height). The height of each bar is called the **frequency density**.

Example



You calculate the frequency density using this formula:

$$\boxed{\text{frequency density} = \text{frequency} \div \text{class width}}$$

Tip: The continuous data here has been split into classes. $0 \leq x < 5$ means the data in the class is more than or equal to 0 and less than 5.

Tip: If all the class widths are the same, you can just plot the frequency on the y-axis.

Example — **Maths Skills**

The table on the right shows the results of a study into variation in pea plant height. The heights of the plants were grouped into four classes.

Height of pea plant (cm)	*Frequency*
$0 \leq x < 5$	5
$5 \leq x < 10$	14
$10 \leq x < 15$	11
$15 \leq x < 30$	3

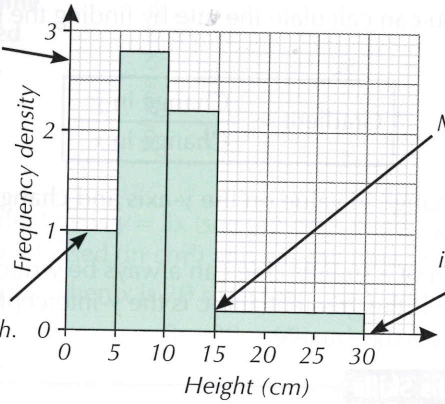
1. To draw a histogram of the data, you first need to work out the width of each class. Write the class width in a new column.

Class width
$5 - 0 = \mathbf{5}$
$10 - 5 = \mathbf{5}$
$15 - 10 = \mathbf{5}$
$30 - 15 = \mathbf{15}$

2. Use the formula on the previous page to calculate the frequency density for each class and write it in another column.
3. Work out a suitable scale for each axis, then plot the histogram. It should look like this:

Frequency density
$5 \div 5 = \mathbf{1}$
$14 \div 5 = \mathbf{2.8}$
$11 \div 5 = \mathbf{2.2}$
$3 \div 15 = \mathbf{0.2}$

Frequency density goes on the y-axis. (So the height of the bars is the frequency density.)



The width of each bar is the class width.

Make sure the bars are touching.

Height (the independent variable) goes on the x-axis.

Tip: You might have to round the frequency density — if so, choose a sensible number of decimal places that will be possible to plot using your graph's scale.

Tip: You can calculate the frequency of a class from a histogram by rearranging the formula: frequency = frequency density × class width.

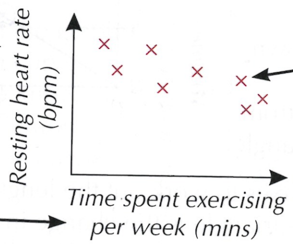
Tip: The width of the whole histogram shows the range of results (how spread out they are).

Scattergrams

When you want to show how two variables are related (or correlated, see page 15) you can use a **scattergram**. Both variables must be numbers.

Example

The y-axis shows one variable...



...and the x-axis shows the other variable.

Each data point is marked with a cross, but they're never joined up.

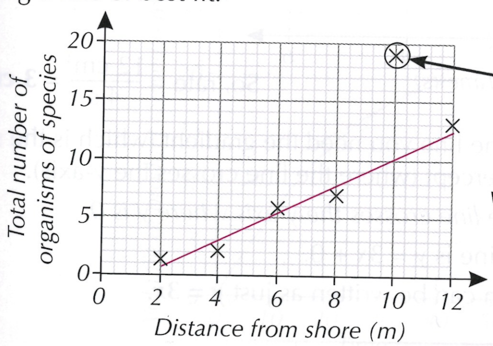
Tip: Scattergrams can also be called scatter graphs, or scatter diagrams.

Tip: Data that's made up of numbers is called quantitative data.

You can draw a **line** (or curve) **of best fit** on a scattergram to help show the trend in your results. To do so, draw the line through or as near to as many points as possible, ignoring any anomalous results.

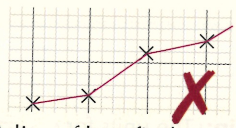
Example

The number of organisms of one species on a rocky beach was recorded at different distances from the shore. The graph below shows the results, including a line of best fit.



Anomalous results don't fit in with the rest of the data — just ignore them when you're drawing a line of best fit.

Tip: You should never join the points together on a scattergram.



A line of best fit shows a trend rather than changes between data points.

Tip: A trend shown by a scattergram is called a correlation. There's more about correlation and what it means on page 15.