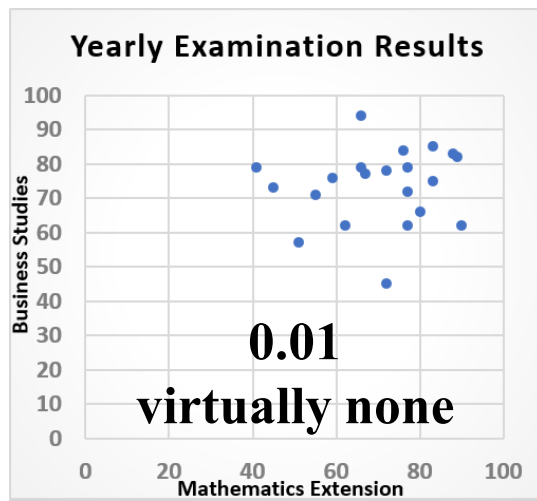# *Scatterplots & Correlation*
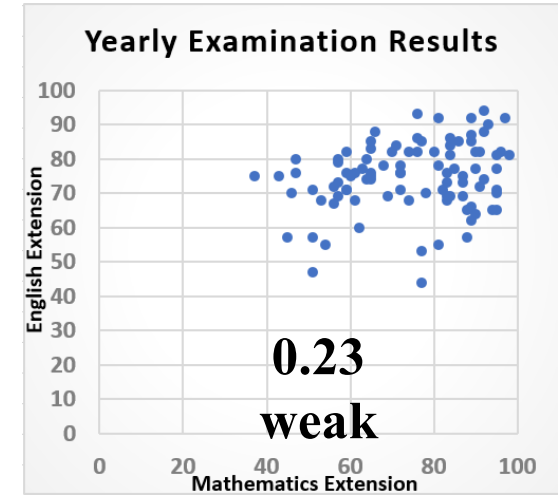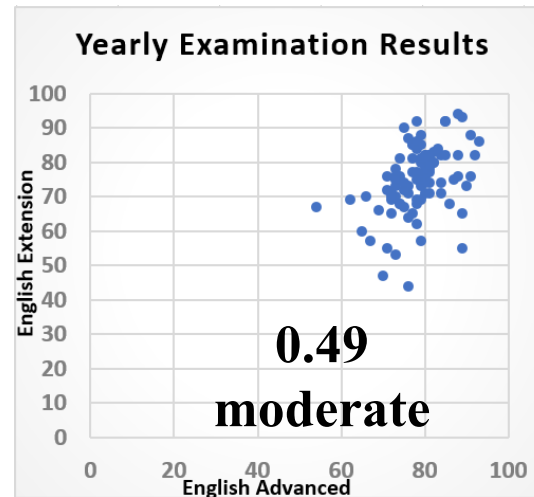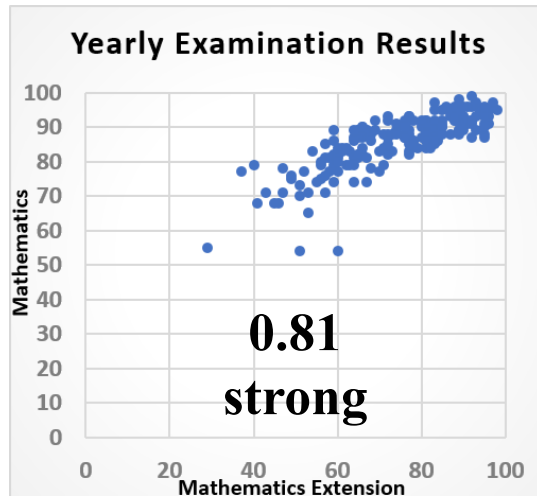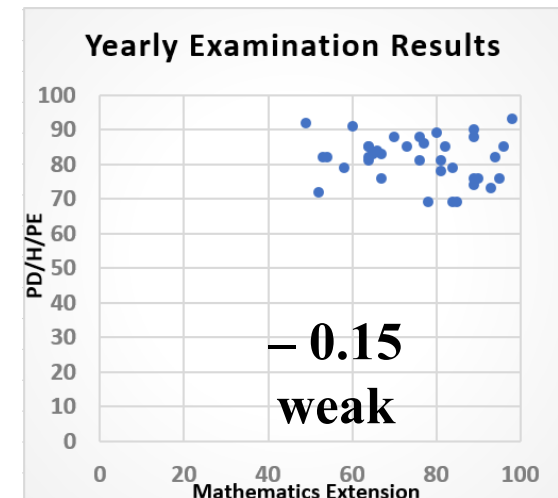
Bivariate data measures two variables on the same dataset to see if a relationship (**correlation**) exists between the data

Scatterplots are used to graph bivariate data

Yearly Examination Results

0.81 strong

Mathematics vs. Mathematics Extension

Yearly Examination Results

0.49 moderate

English Extension vs. English Advanced

Yearly Examination Results

0.23 weak

English Extension vs. Mathematics Extension

Yearly Examination Results

0.01 virtually none

Business Studies vs. Mathematics Extension

Correlation is the strength of the scatter

$$-1 \leq \text{correlation} \leq 1$$

Yearly Examination Results

− 0.15 weak

PD/H/PE vs. Mathematics Extension
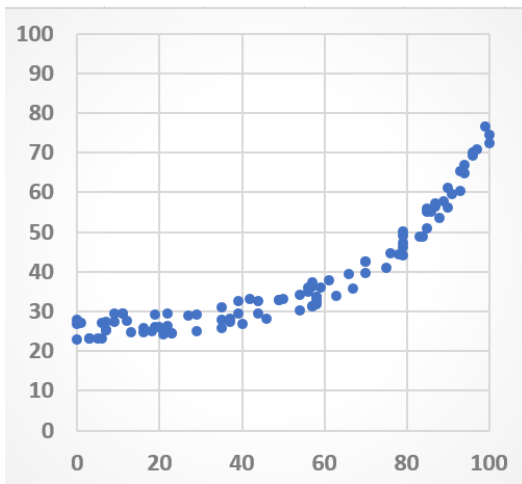
As with any graph the independent variable should be placed on the horizontal axis and the dependent variable on the vertical axis.

Not all bivariate datasets have a clear independent and dependent variable, in which case the choice of axes does not matter.
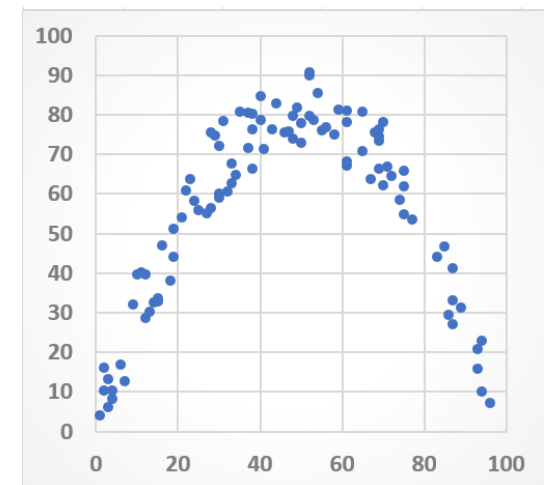
We will be concerned with **linear correlation**, however it is possible to have a non-linear correlation.

strong
exponential
correlation
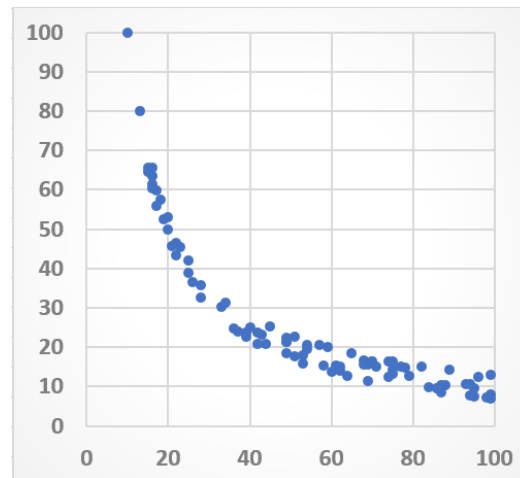
strong
inverse
correlation

strong
quadratic
correlation

Linear Correlation

A single line that represents the general pattern of the data is called a **line of best fit**

It is important to identify any multiple points in the scatter graph, using larger circles, other shapes or placing the number of occurrences inside the point.

If the correlation can be seen clearly, a line of best fit can be drawn by eye.
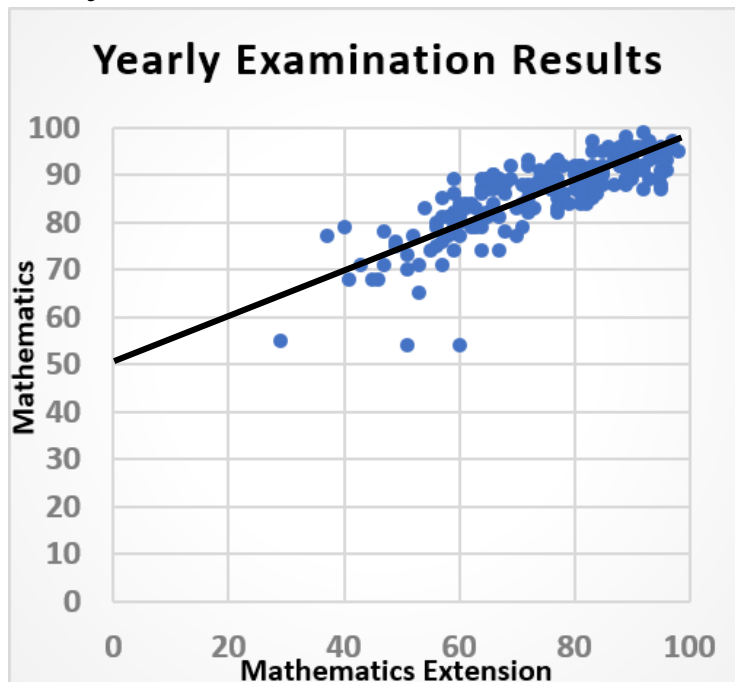
Assuming the data is reliable, predictions can be made by;

interpolation; predicting results within the range of data e.g. a student scoring 40% in Extension should score 70% in Advanced

extrapolation; predicting results outside the range of data e.g. a student scoring 20% in Extension should score 60% in Advanced

Correlation Vs Causation

Just because we have a correlation doesn't mean it is true!

e.g. (i) During summer, people get hayfever. There is also a lot of pollen in the air.

There is a strong **correlation** between presence of pollen and hayfever

Therefore, pollen **causes** hayfever

This is a sensible correlation and there have been studies to prove it.

(ii) During summer, the sale of icecream increases and so does death by drowning

There is a strong **correlation** between icecream sales and death by drowning

Therefore, icecream **causes** drowning

This is NOT a sensible correlation they are just both things that happen more in summer

correlation does not mean causation.

Exercise 15D; 1, bde, 2, 3bd, 4, 5, 6, 7, 8