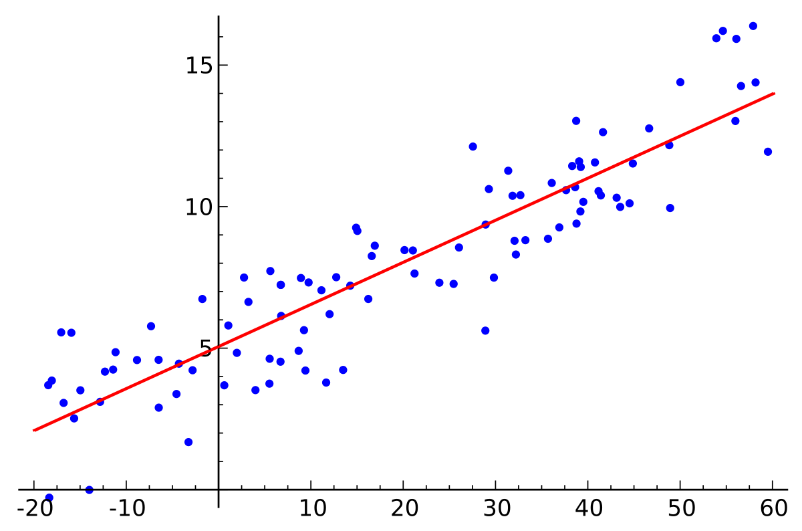
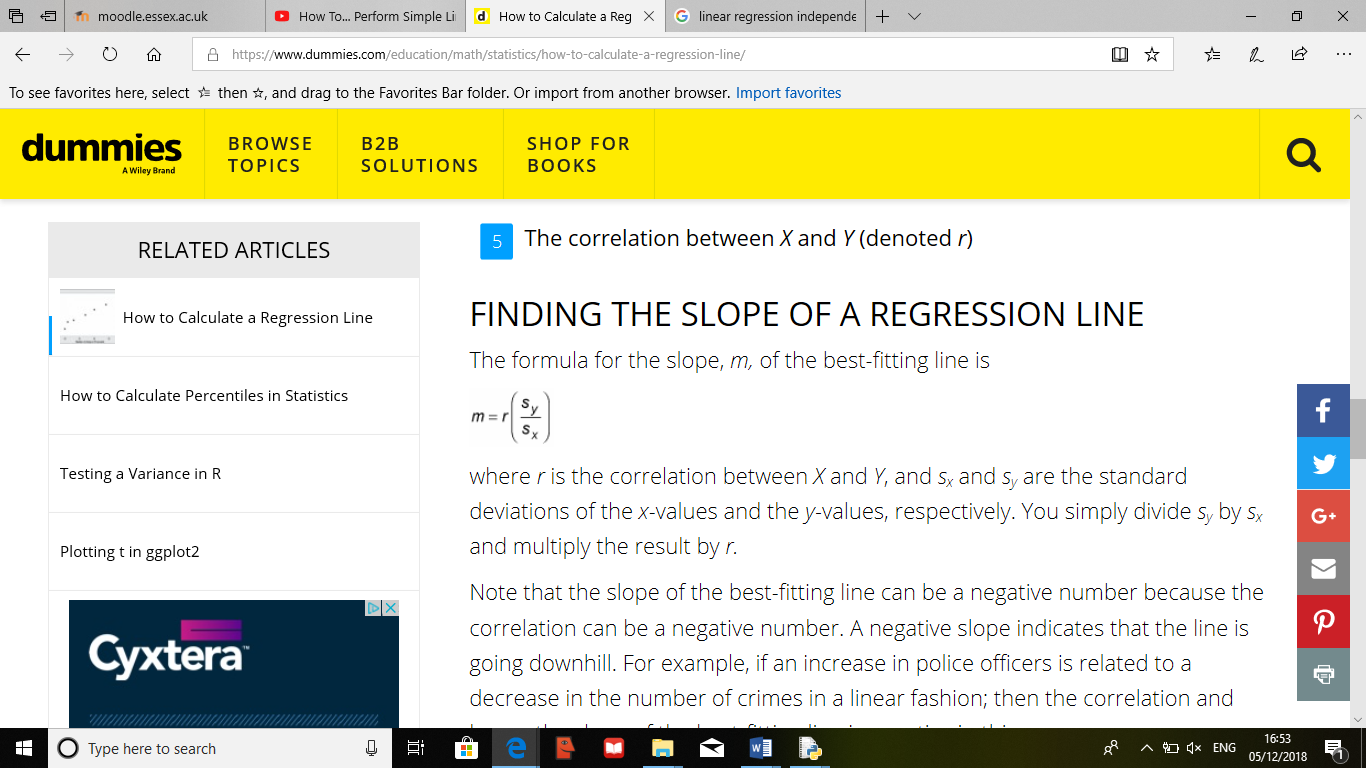
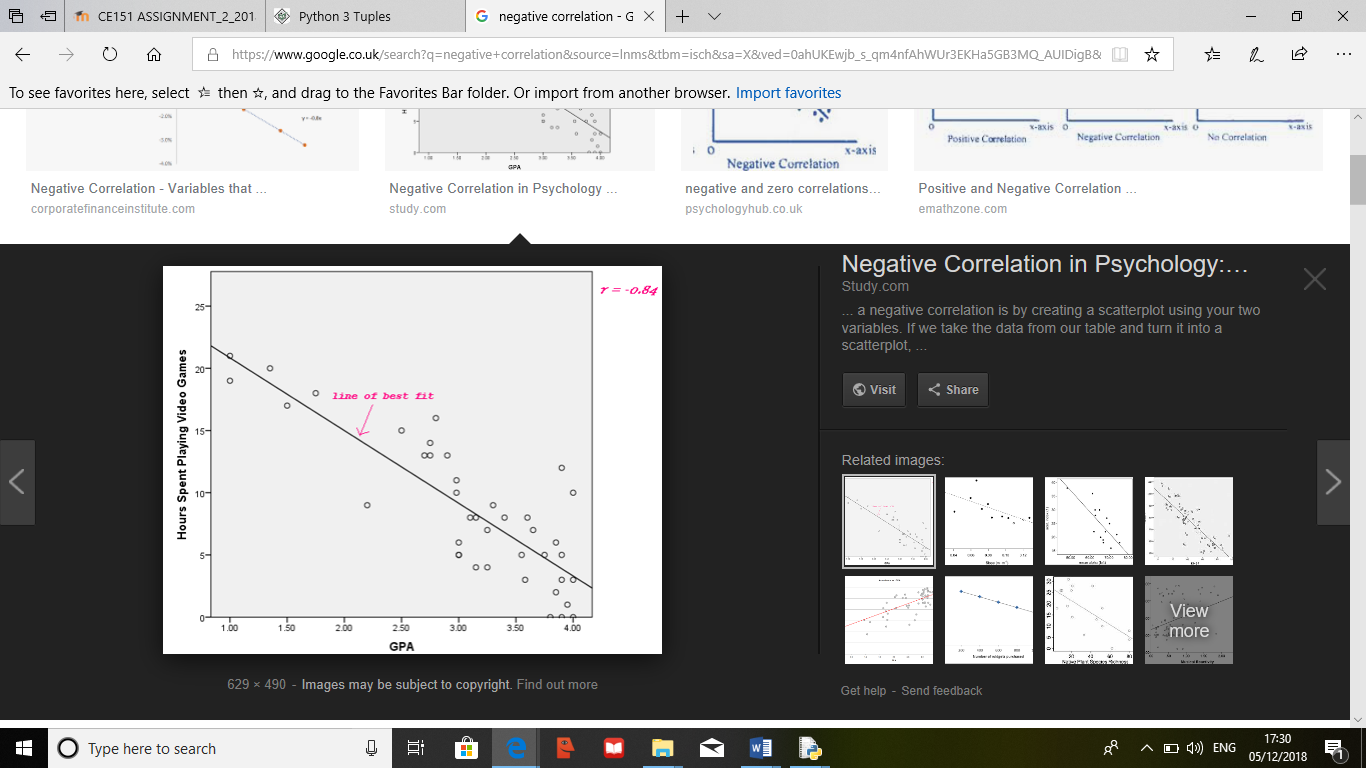
Linear regression models

* The main idea of regression is for two things, 1) does a variables do a good job predict an outcome, 2) which variables are significant predictor of the outcome variable
* [](https://www.google.co.uk/url?sa=i&rct=j&q=&esrc=s&source=images&cd=&cad=rja&uact=8&ved=2ahUKEwiBx7iN9_beAhXNJFAKHQwTA8IQjRx6BAgBEAU&url=https://en.wikipedia.org/wiki/Regression_analysis&psig=AOvVaw3gld3roQ6DCK4RG8by5_xA&ust=1543489098118621)Regression is defined by the formula, y = c + b\*x, where y = estimated dependent variable score, c = constant, b = regression coefficient, and x = score on the independent variable.
* First, the regression might be used to identify the strength of the effect that the independent variable(s) have on a dependent variable. The independent variable goes on the x axis and the dependent variable on the y axis.
* To work out linear regression we must first work out the standard deviation of x and y. These are denoted as Sx and Sy. We must also work out the mean of x and y. The correlation between x and y is r.
* The linear regression formula is y = a + bx and the y intercept formula is a = y – bx
* So we substitute the mean of y, the mean of x and the gradient into the y intercept formula. This helps us work out ‘a’.
* Next we can plug these values into a linear regression formula and we’ll have our regression line.



-The correlation denoted by ‘r’ can be negative or positive.

The diagram to the right is a negative correlation, we notice this by the direction of the line of best fit