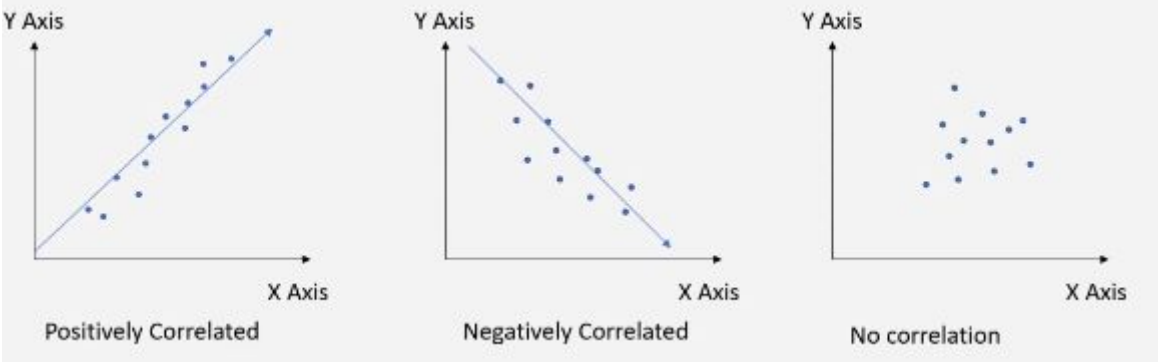
LESSON HANDOUT

Correlation

Correlation indicates the direction and strength which two variables are related to one another or otherwise explained, the extent to which variable fluctuate together. While correlation does not equal causation, it's a useful precursor to causation and can indicate areas for further analysis.

Strength and direction

Correlation provides the strength and direction of a relationship. Here is what we expect a scatter plot to look like with respect to positively, negatively and non-correlated variables:



- Positive correlation means an increase in one variable relates to a proportional increase in another variable.
- Negative correlation means an increase in one variable relates to a proportional decrease in another variable.
- No correlation shows no relationship between variables.

Correlation coefficient

The Pearson's correlation of coefficient is a measure of correlation which indicates the extent and direction of a relationship (negative or positive). The below table can be used to interpret the correlation coefficient:

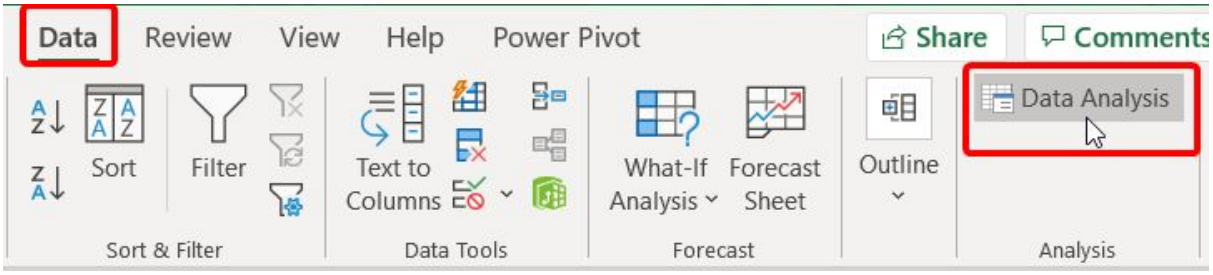
-0.80 to -1	-0.60 to -0.80	-0.40 to -0.60	-0.40 to 0.40	0.40 to 0.60	0.60 to 0.80	0.80 to 1
Strong negative correlation	Moderate negative correlation	Weak negative correlation	No correlation	Weak positive correlation	Moderate positive correlation	Strong positive Correlation

Correlation in Excel

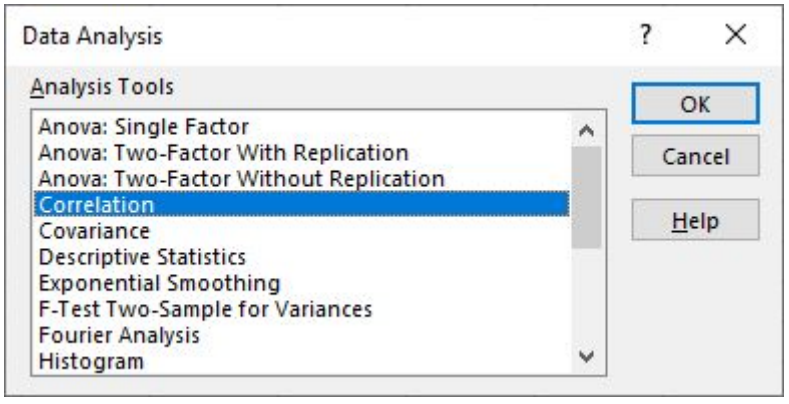
We are able to easily calculate correlation in Excel using the CORREL formula and display a grid of how variables correlate to one another in a correlation plot.

COREL Formula. =CORREL(array1, array2)

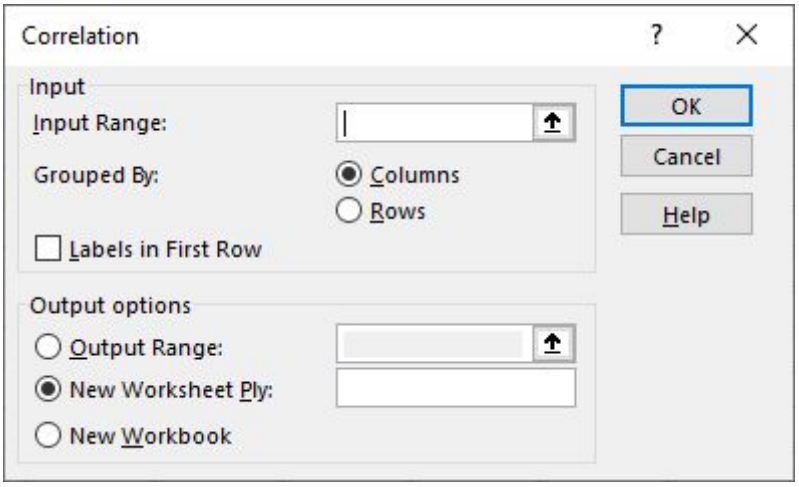
Correlation plot. To create a correlation plot - in the ribbon, go to the Data tab and within 'Analysis' section click on 'Data Analysis'.



- Select 'Correlation' from the list and click OK

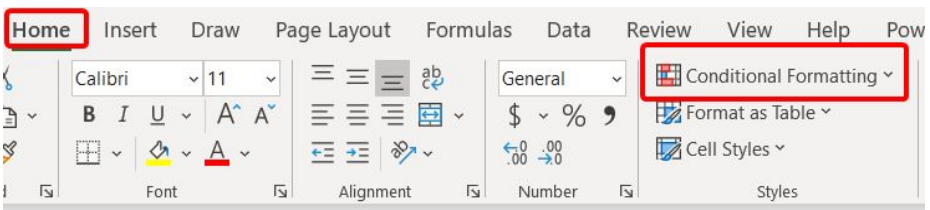


- Select your Input Range (for example $B:$F for columns B through to F)
- Tick 'Labels in First Row', then
- Choose where you want your correlation plot to go

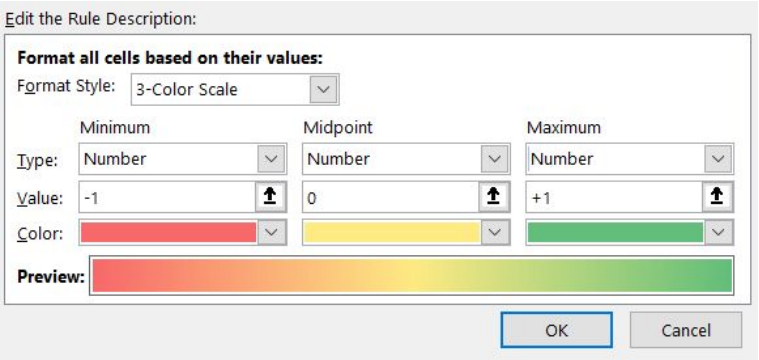


Conditional Formatting. To make your correlation plot easier to understand, you can use conditional formatting.

- Select the values within your correlation plot
- In the Home tab, under Styles, click on Conditional Formatting > New Rule



- Choose the Format Style ‘3-Color Scale’ with the types as Numbers, with the Minimum Value of -1, Midpoint 0, and Maximum +1.



- Your Correlation Plot should look like this, making it easier to relate to the Pearson’s correlation of coefficient.

	Age	*Height*	*Weight*	*Games Played*	*PTS*
Age	1.00				
Height	0.03	1.00			
Weight	0.09	0.84	1.00		
Games Played	0.14	0.07	0.08	1.00	
PTS	0.06	-0.02	0.03	0.73	1.00