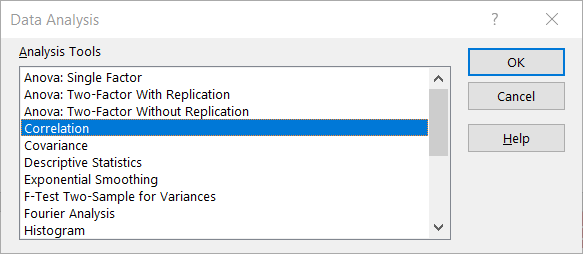
Pearson correlation: demo notes

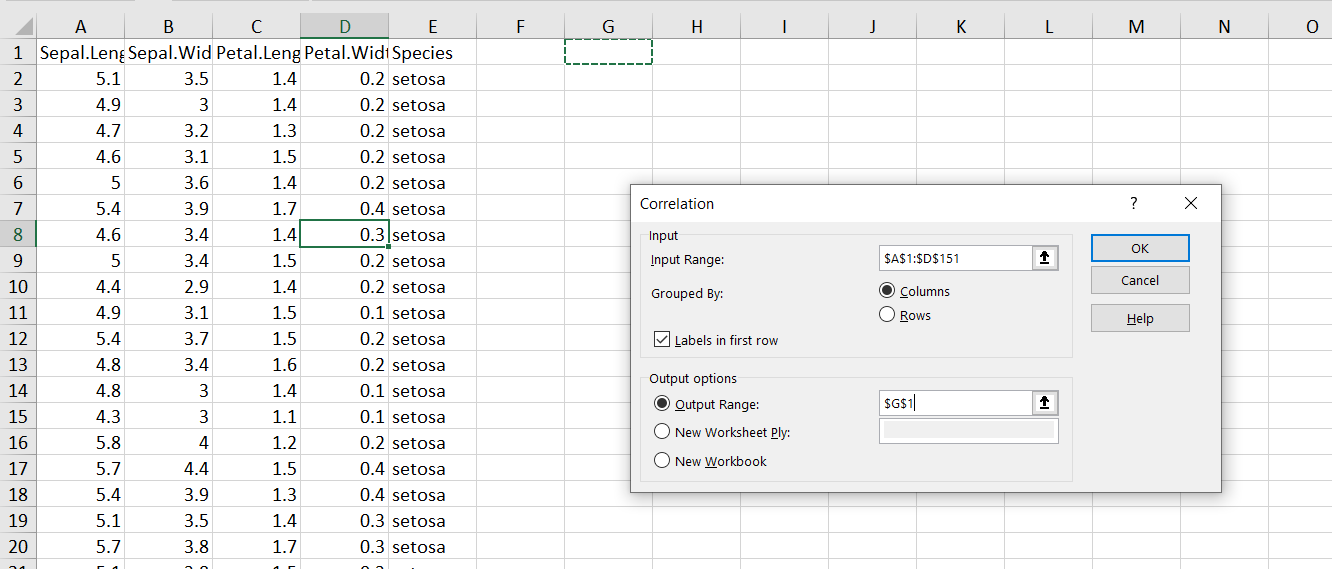
**Pearson correlations**

File: iris-corr.xlsx

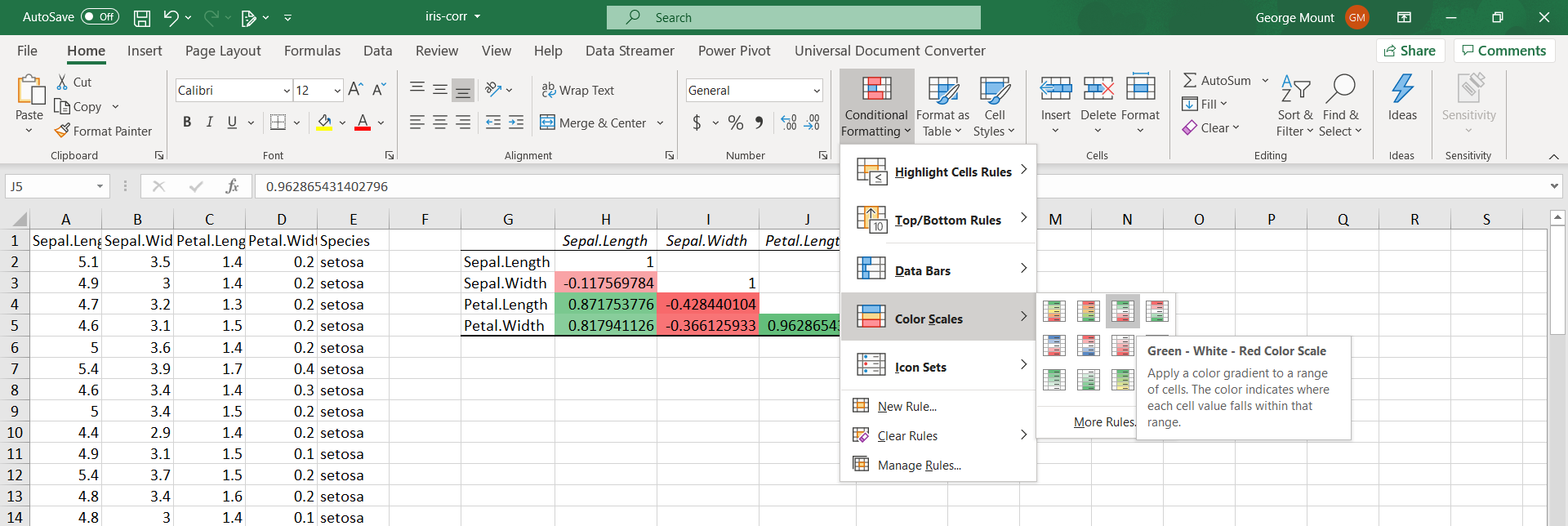
1. To insert a correlation matrix, go to the Analysis ToolPak and select Correlation.



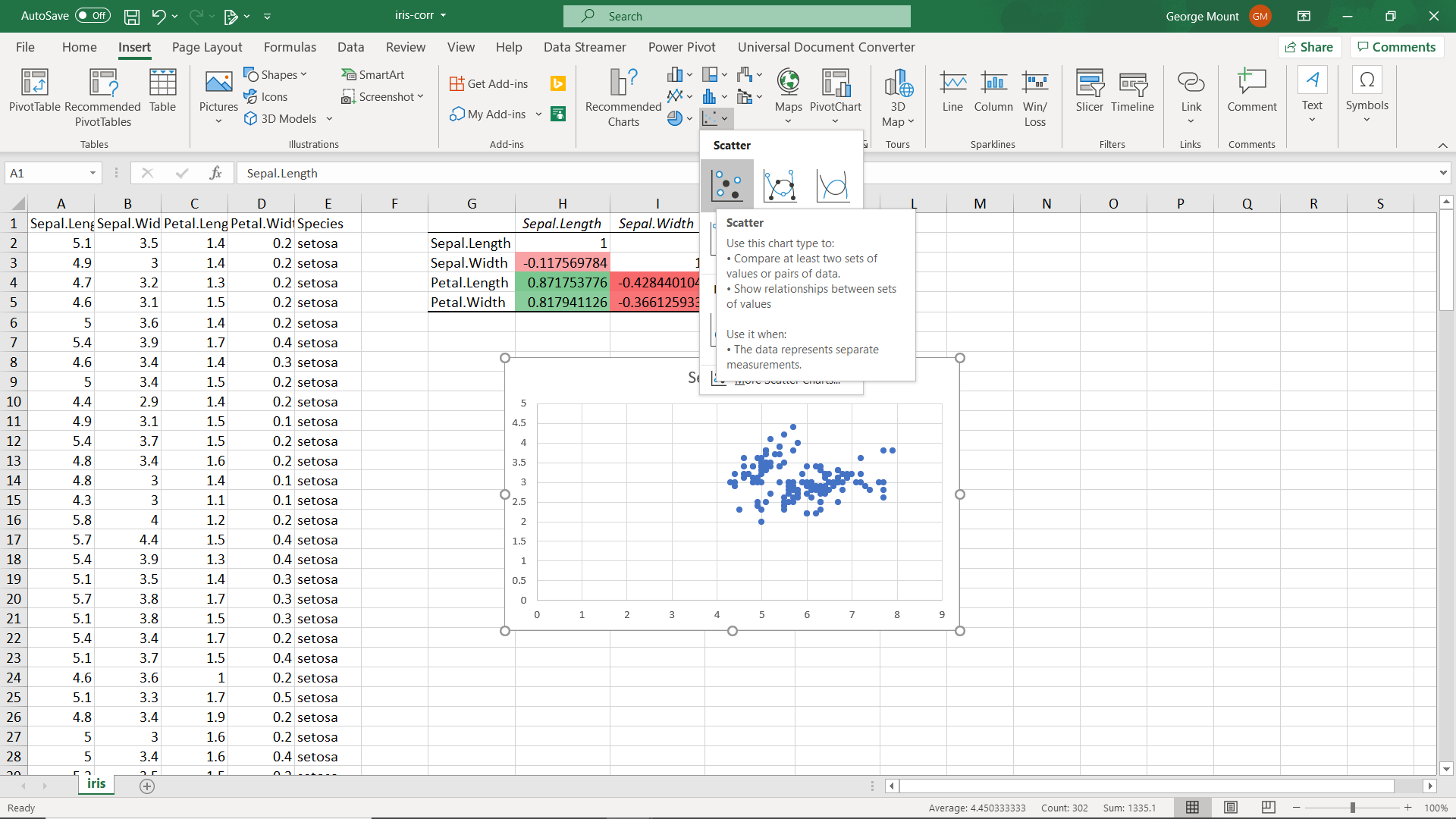
1. The input range will be all the *numeric* fields. String fields cannot be included in the correlation analysis.



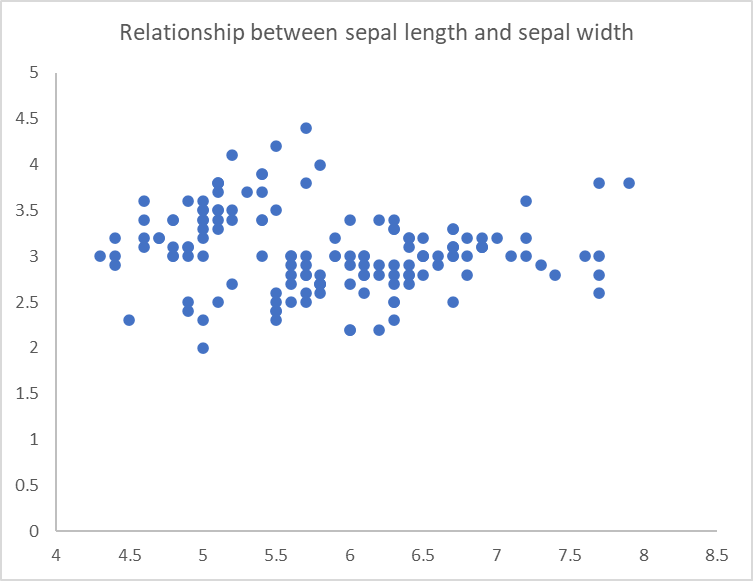
1. For ease of interpretation, select all the correlation values and select Home > Conditional Formatting > Color Scales > Green – White – Red Color Scales.



1. To plot the relationship between two variables (in this case, sepal length and sepal width), highlight the data and select Insert > Scatter.



1. To clean up this scatter chart, set the X-axis to 4 and remove the gridlines. It’s also a good idea to label this chart more clearly.

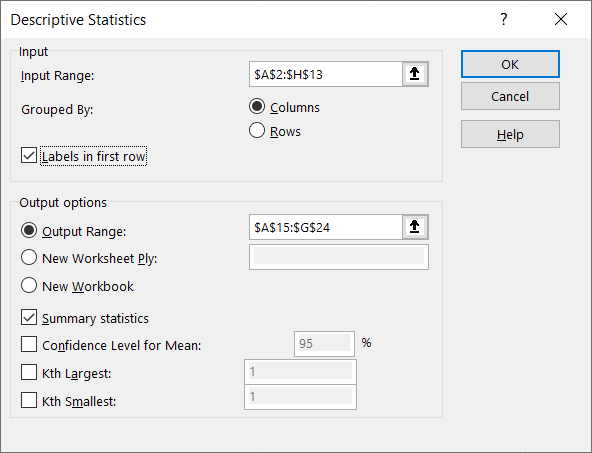


1. What does the scatterplot of sepal length and petal length look like?
2. By custom, you want to put the *independent* variable on the X-axis, and the *dependent* on the Y-axis.

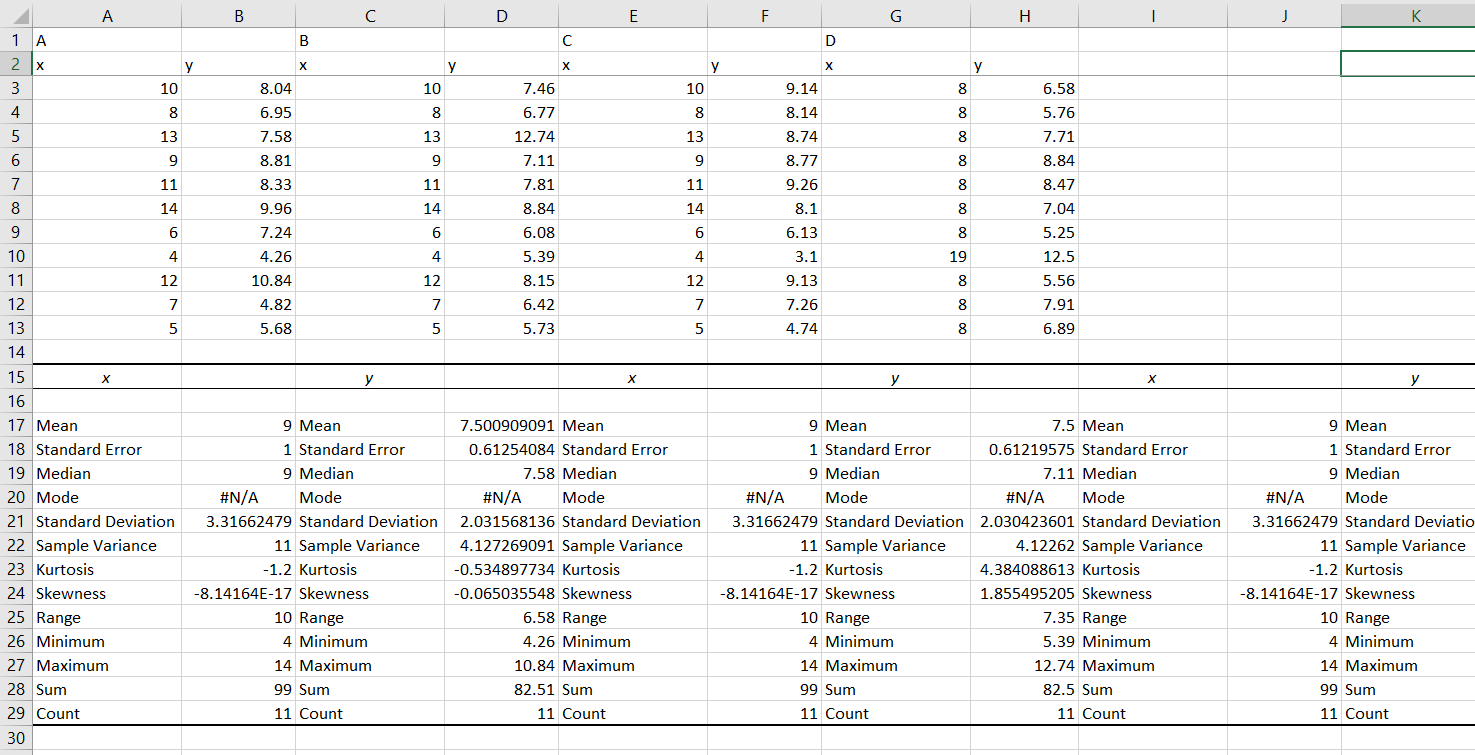
**Careful with correlations!**

File: anscombe.xlsx

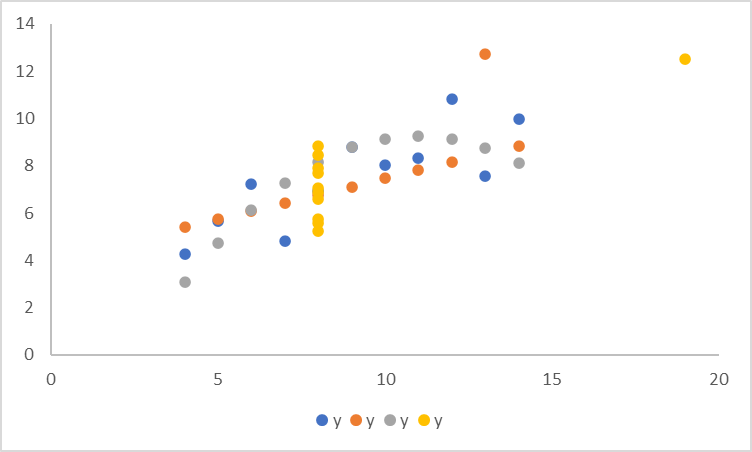
1. Perform descriptive statistics on all variables by selecting Descriptive Statistics from the ToolPak. Make sure you select “Summary statistics” from the Output Options.



1. From the results of these descriptive statistics, each X-Y pair is *very* similar.



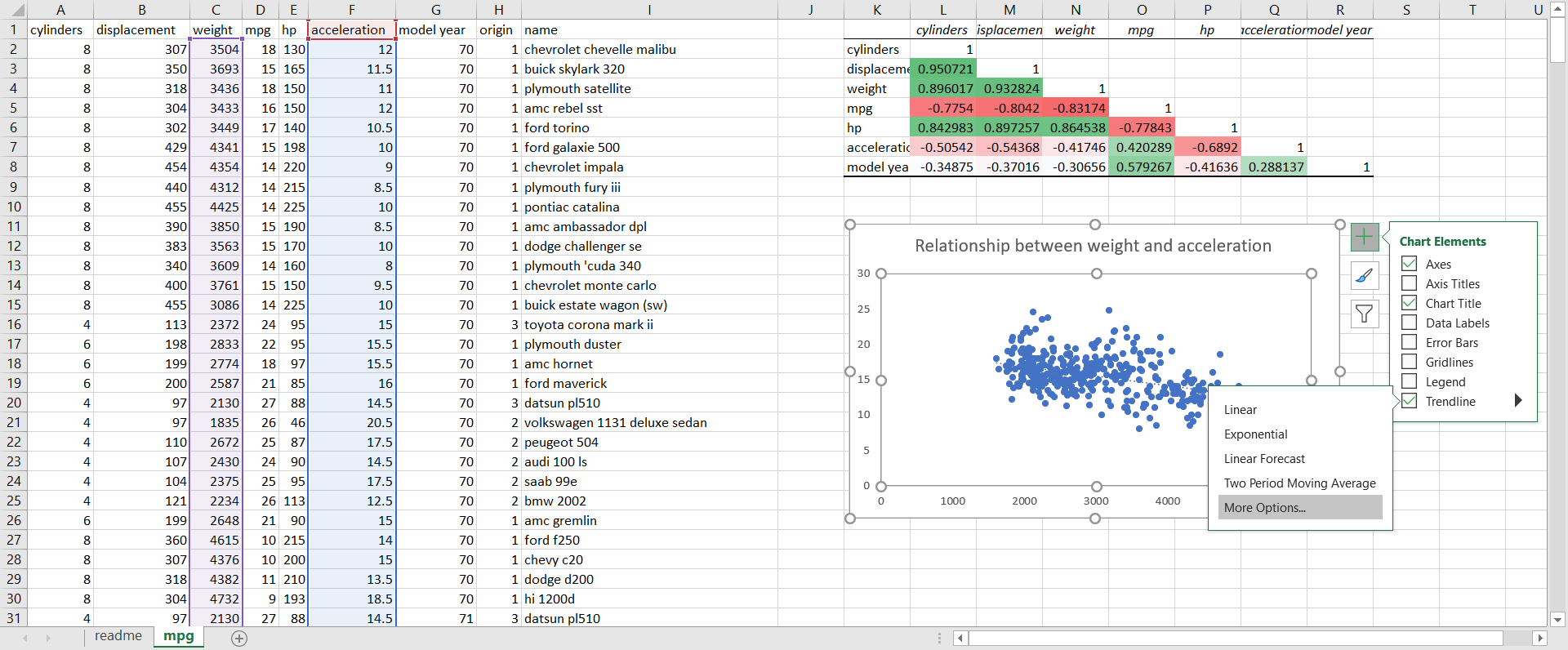
1. What happens if we graph this data using a scatter plot? Highlight the range and select Insert > Recommended Charts > Scatter.
   1. Only from the visualization can we easily see that these datasets are quite different!



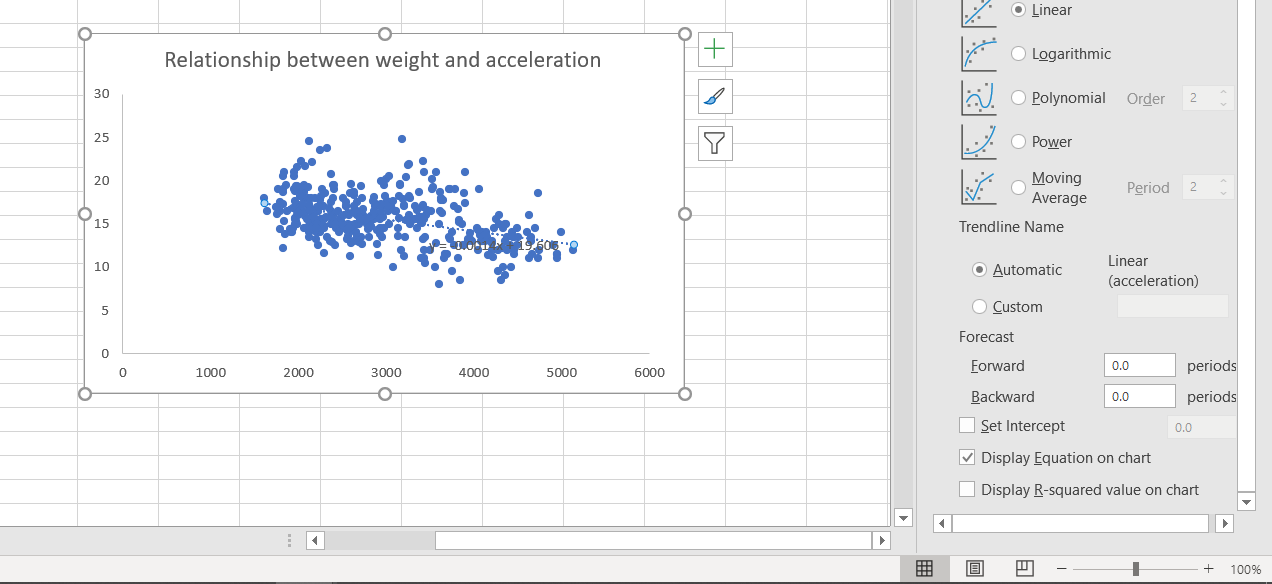
Up and running with linear regression: demo notes

File: mpg-regression.xlsx

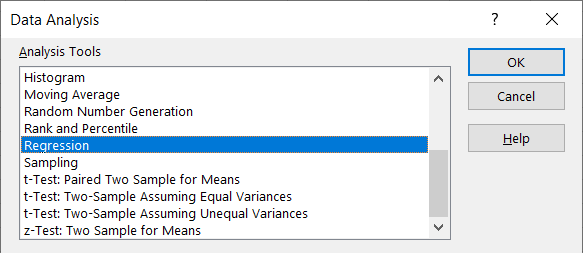
1. The easiest way to find the regression coefficients in Excel is to add a fitted line to a scatter plot:

-

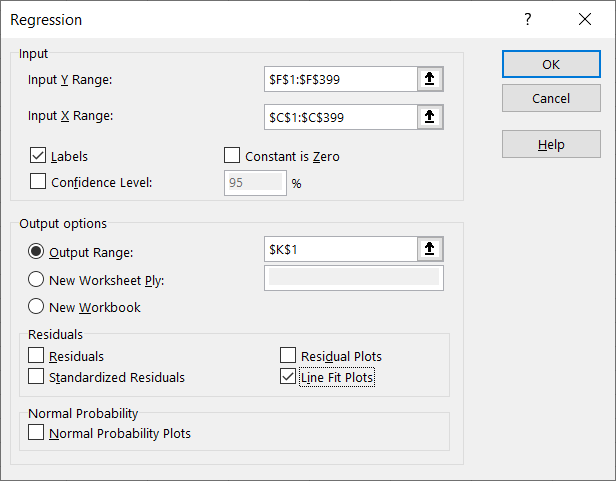
* 1. If we click on More Options, we can even insert the regression equation on the chart. Pretty cool, but we can get additional important information by running a full regression model via the ToolPak.



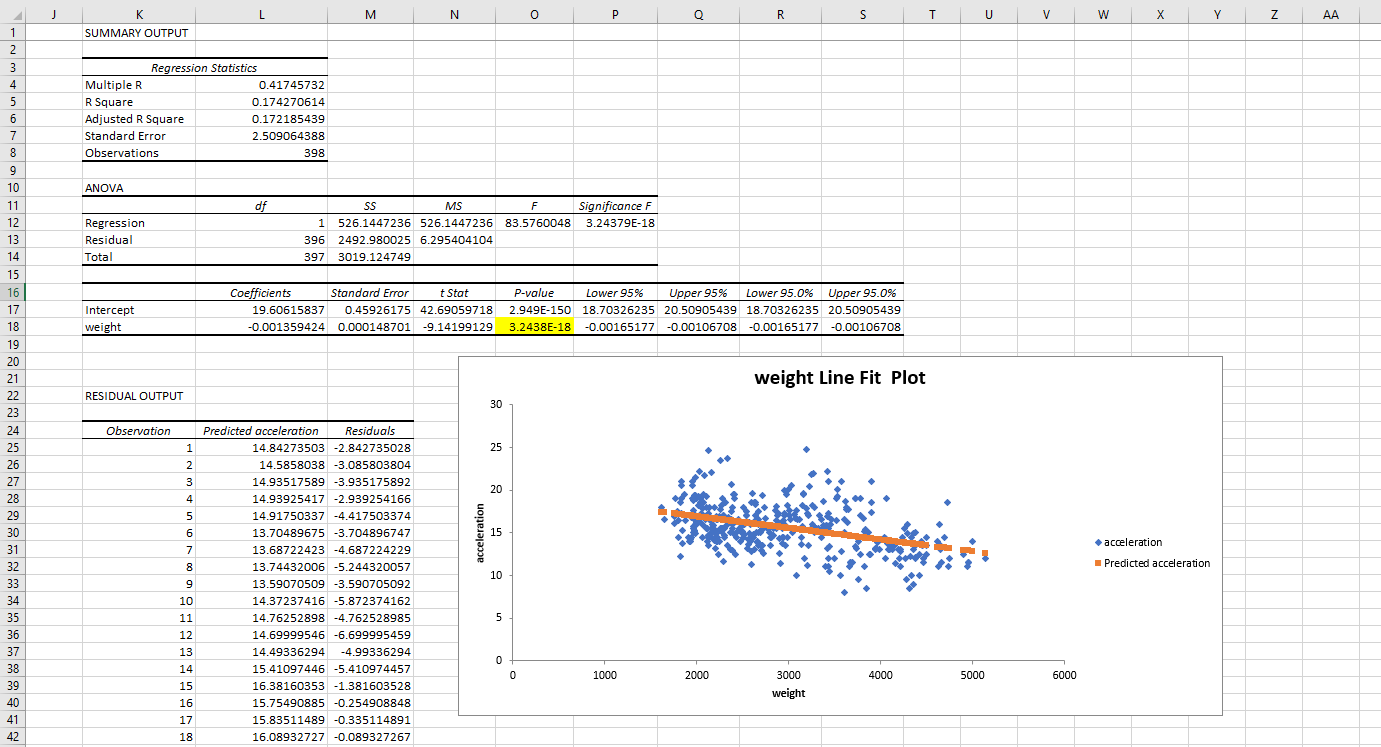
1. This won’t give you the statistical significance along with some other key metrics. For that we will select the ToolPak.



* 1. Fill it out. Turn line fit on.

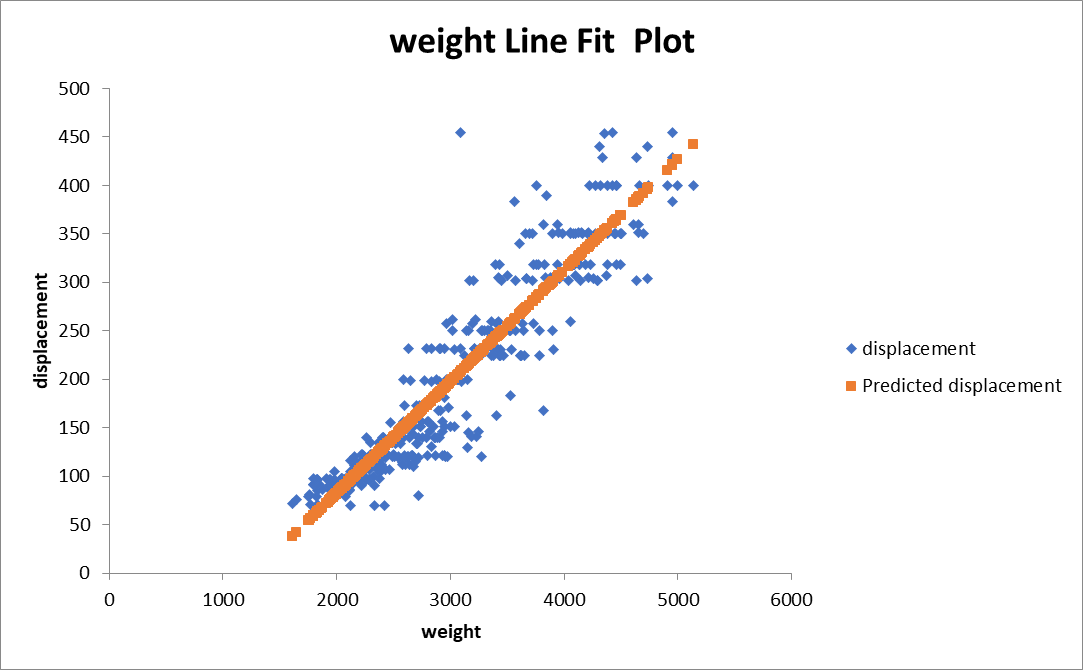


* 1. From here we can see that the slope of the fitted line is significantly different from zero. By turning on the line fit plot, we can also see our predicted values against the actual values.



File: mpg-regression-drill-solutions.xlsx

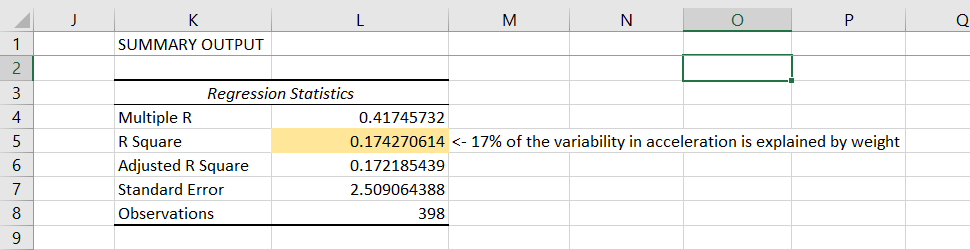
1. Keep in mind that linear regression assumes no influential cases. From visual inspection, there appears to be at least one influential case in this analysis.
   1. Spotting and adjusting for influential cases will be addressed in the next course in this sequence.



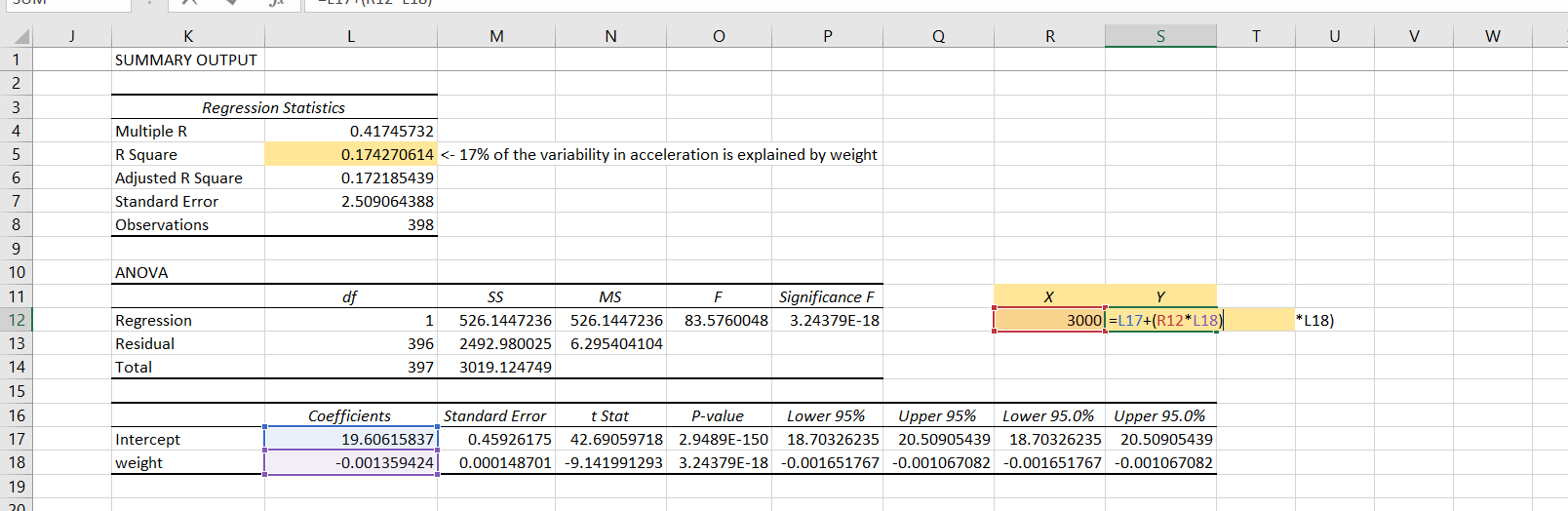
File: mpg-regression-diagnostics.xlsx

To increase our understanding and use of the model, let’s do two more things with our model.

1. R-square: This tells us what percent of our variability in Y is explained by X. We can get the r-square from the results of the ToolPak.
   1. In this case, 17% of variability in acceleration is explained by weight.



1. Making point predictions: Using the intercept and slope of our regression line, we can predict a value for Y given a value for X.
   1. This is done by adding our intercept to the product of the slope and our value for X.



* 1. In this example, we predict that a car weighing 3,000 pounds will accelerate from 0-60 in 15.5 seconds.