**Problem**

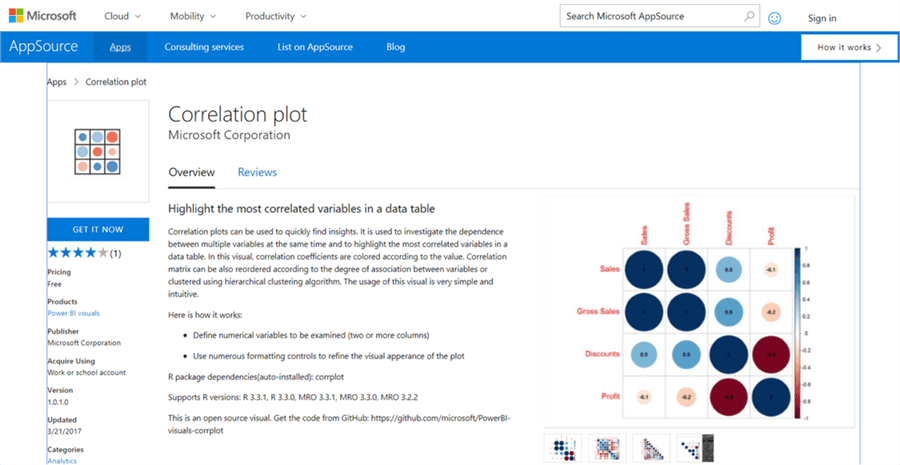
Correlation Analysis is a fundamental method of exploratory data analysis to find a relationship between different attributes in a dataset. Statistically, correlation can be quantified by means of a correlation co-efficient, typically referred as Pearson’s co-efficient which is always in the range of -1 to +1. A value of -1 indicates a total negative relationship and +1 indicates a total positive relationship. Any number closer to zero represents very low or no relationship at all. There is a statistical calculation involved to find this co-efficient and using this you can identify the correlation between two attributes with numerical data. If interested, you can read more about it [here](https://en.wikipedia.org/wiki/Pearson_correlation_coefficient). It can be a very statistically intensive process if the task is to identify correlation between many numeric variables. Correlation plots can be used to quickly calculate the correlation coefficients without dealing with a lot of statistics, effectively helping to identify correlations in a dataset.

**Solution**

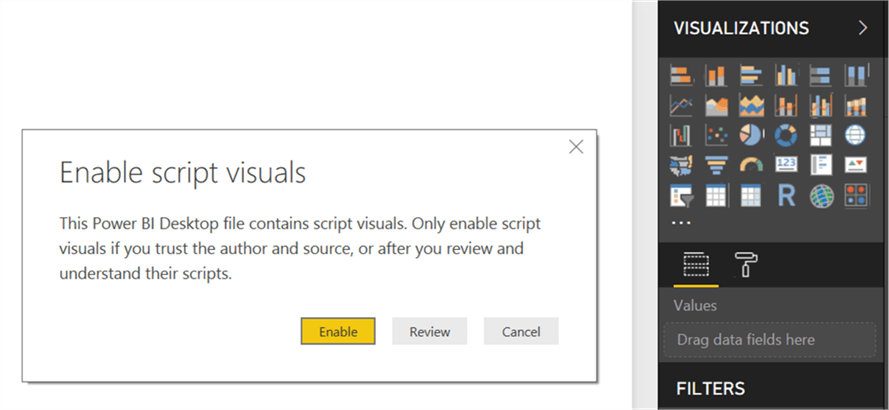
Power BI provides correlation plot visualization in the Power BI Visuals Gallery to create Correlation Plots for correlation analysis.

In this tip we will create a correlation plot in Power BI Desktop using a sample dataset of car performance. It is assumed that Power BI Desktop is already installed on your development machine. So please follow the steps as mentioned below.

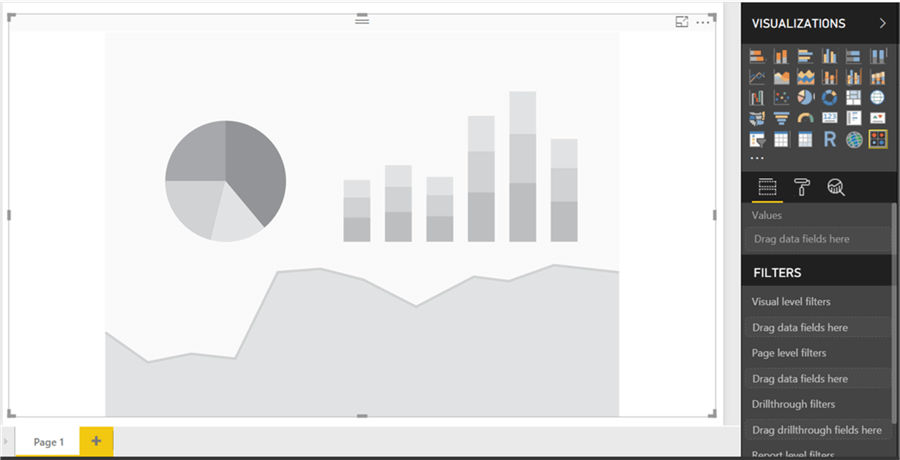
1) The first step is to download the correlation plot from [here](https://appsource.microsoft.com/en-us/product/power-bi-visuals/WA104380814?src=office&tab=Overview), as it is not available by default in Power BI Desktop. This visualization makes using of the R corrplot package. The same plot can be generated using the R Script visualization and some code. Instead this visualization eliminates the need for coding and provides parameters to configure the visualization.



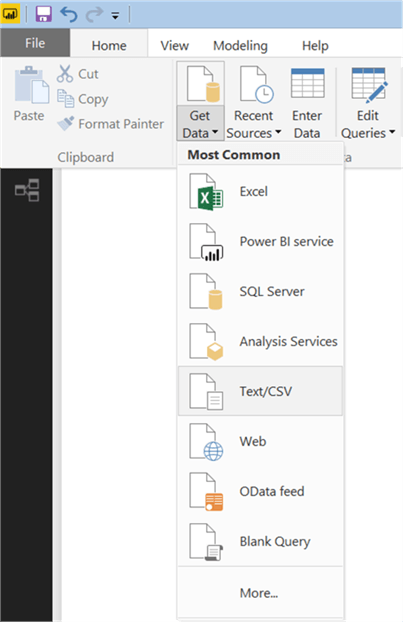
2) After downloading the file, open Power BI Desktop. You can click on the ellipsis in the visualization tab and select the “Import from file” menu option. This will open a dialog box to select the visualization package file, to add the visualization in Power BI. Select the downloaded file and it should add the correlation plot to Power BI Desktop visualizations gallery. Click on the correlation plot icon in the gallery and you should be prompted with a warning as shown below to enable script visuals. Click on **Enable** which will make the correlation plot operational.



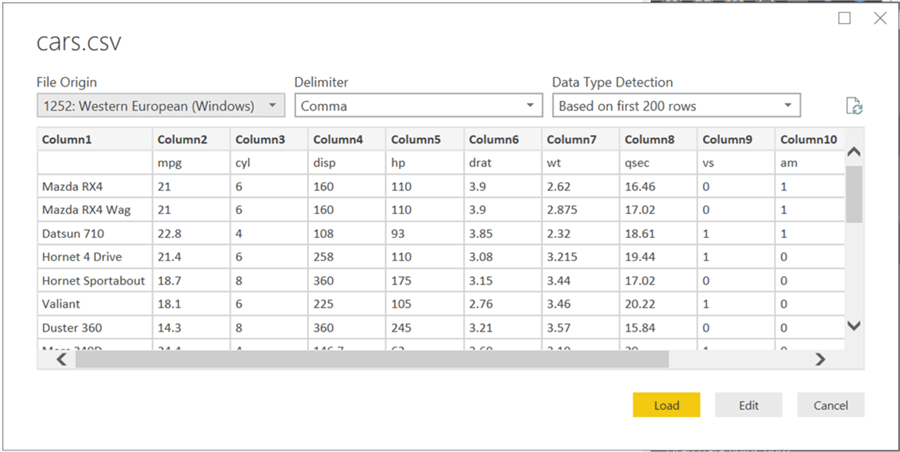
3) After the correlation plot is added to the report layout, enlarge it to occupy the entire available area on the report. After you have done this, the interface should look as shown below.



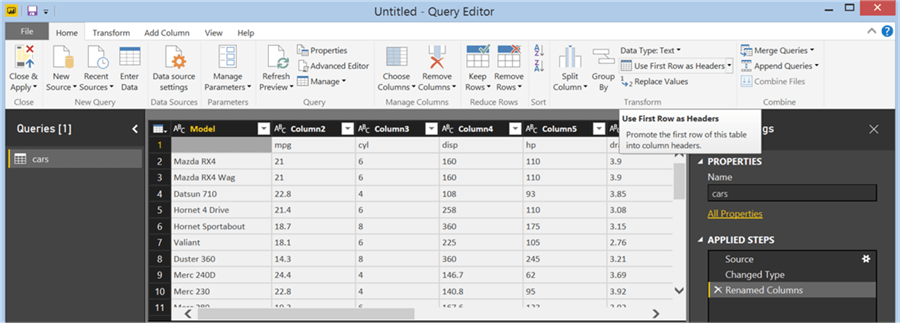
4) Now that we have the visualization, it is time to populate it with some data on which correlation analysis can be performed. We need a dataset with many numerical attributes. You can download a sample dataset from [here](https://www.mssqltips.com/tipimages2/5197_cars.csv) that contains data on car performance with metrics like miles per gallon, horsepower, transmission, acceleration, cylinder, displacement, weight, gears, etc. Click on the Get Data menu and select CSV since we have the data in a csv file format.



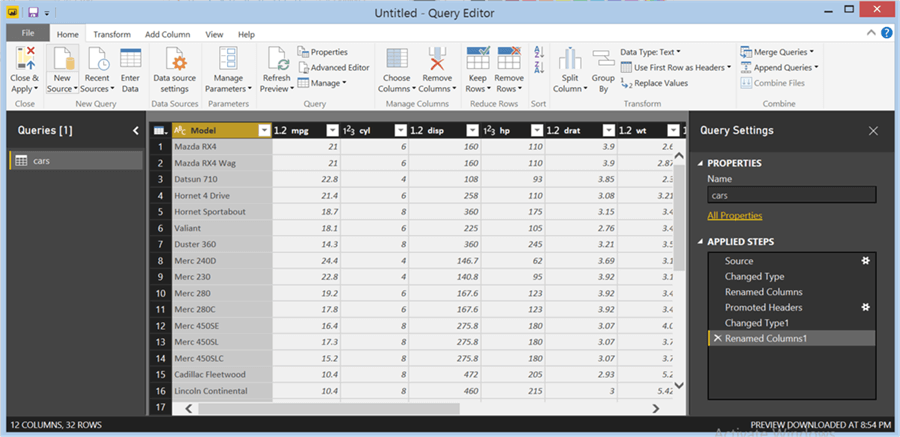
5) This will open a dialog box to select the file. Navigate to the downloaded file and select it. This will read a few records from the file and show a data preview as shown below.



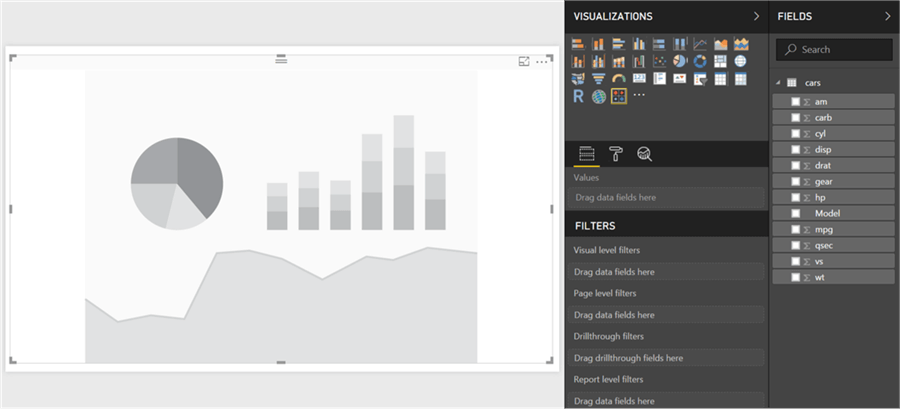
6) As you can see, the column headers are in the first row. So, click on the edit button to indicate this before importing the dataset. Click on the “Use First Row as Headers” so that we get the column names properly. Also, you can rename the Car Names column and name it Model.



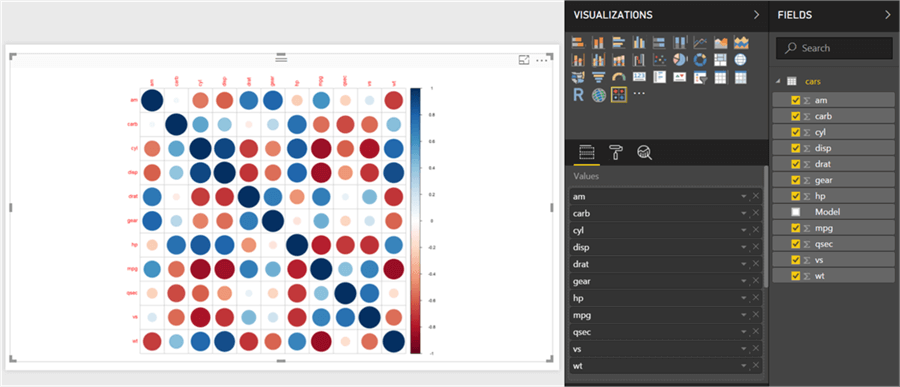
7) After you apply the setting, the column names should look as shown below. Click on the Close and Apply button to complete the import process.



8) After the data is imported in the Power BI Desktop, the model should look as shown below.



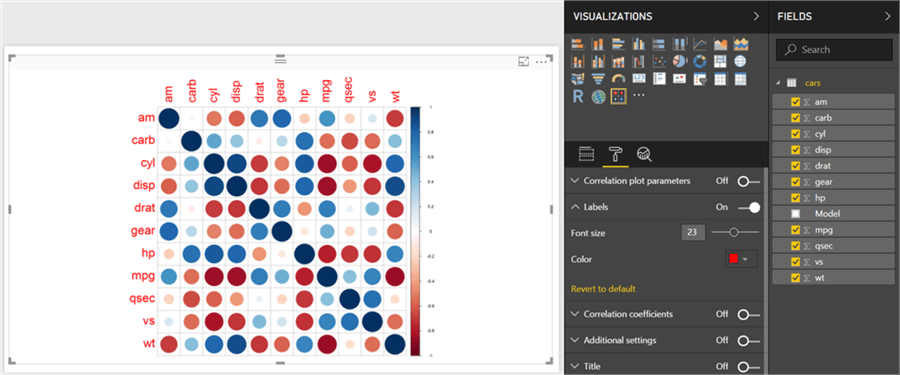
9) Now it is time to select the fields and add them to the visualization. Click on the visualization in the report layout, and add all the fields from the model except the model field which is a categorical / textual field. The visualization would look as shown below.



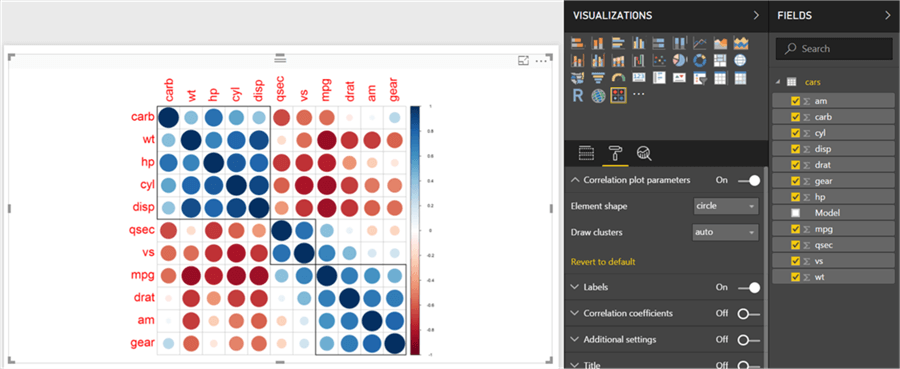
10) The below points should be considered while reading this plot:

* The dark blue circles in a diagonal line from top left to bottom right shows correlation of an attribute with itself, which is always the strongest or 1. So this should not be read as correlation, but just as a separator line.
* The more the circle has a dark blue color, it signifies stronger positive correlation. The darker the red color, it signifies a negative correlation. Lighter or white colors signifies weak or no correlation.
* The scale can be used to estimate the correlation coefficient value.

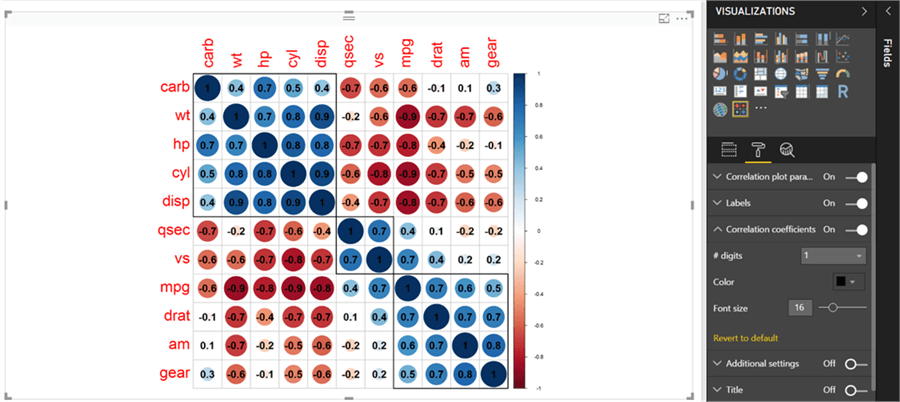
11) We need to make a few modifications in this plot to make it visually analyzable. Click on the Format option, in the Labels section and increase the font size, so that the field labels are clearly visible as shown below. As you can see, weight (**wt**) has a strong positive correlation with displacement (**disp**) and miles per gallon (**mpg**) has a strong negative correlation with weight (**wt**). The data is shown in a matrix format and there are many positive and negative correlation spreads in the plot.



12) It would be easier to analyze correlation if attributes with the same type of correlation are clustered together. To do so, select the correlation plot parameters and set the “Draw clusters” property to “Auto”. This will cluster and reorganize the attributes as shown below.



13) The strength of the correlation is still shown by the depth of the color. It would be easier to analyze the data if it is shown by a number indicating this strength – i.e. correlation coefficient. To do so, switch On the **Correlation Coefficients** section and increase the font size, so that you can see the coefficient clearly. Using the values as a reference, you can easily find out the strongest and weakest correlation in the entire dataset.



14) There are other sections for formatting the data, but those are mostly related to cosmetic aspects of the plot like title, background, transparency, title, etc. You can try to modify those settings and make the plot more suitable to the theme of the report.

With Power BI, without digging into any coding or complex statistical calculations, one can derive correlation analysis from the data by using the correlation plot in Power BI Desktop.