IS 362 – Week 8 Assignment

Your task in this week's assignment is to load a dataset, perform some minor cleanup and transformation tasks, then use exploratory data analysis to learn about the distribution of variables and the relationship between variables.
Here is a link to the "`Auto MPG`" dataset in the UC Irvine data repository:
https://archive.ics.uci.edu/ml/datasets/Auto+MPG
.

1. Load the data in the (yellow highlighted) Data Folder in the file `auto-mpg.data` into a pandas DataFrame.
2. Review the data and the provided data set description.
3. Use the attribute information provided in `auto-mpg.names` to name your pandas DataFrame columns.
4. Some of the horsepower data is marked as missing with a '?'. You need to replace this '?' with a pandas-appropriate missing value, then convert the column to numeric.
5. The origin column values are 1, 2, and 3. Convert these values respectively to 'USA', 'Asia', and 'Europe'
6. Create a bar chart that shows the distribution for cylinders.
7. Create a scatterplot that shows the relationship between horsepower and weight.
8. Ask, then answer (using matplotlib and seaborn graphics capabilities) an interesting question about the distribution of one of the variables, or the relationship between two (or more!) of the variables in the "Auto MPG" dataset.

The figure to the right shows an example of an advanced plot that can be created from the auto-mpg dataset using pandas and seaborn.