**Exploratory Analysis of MPG Data Set**

Suppose it’s your first time to purchase a car. You are really excited but don’t know how to choose a car. Don’t worry, if you look at this report, you’ll probably know how to make your own decision.

The data set is called ‘Auto MPG Data Set‘, there are 406 instances in this data set with each instance is a description of a car. Totally there are nine attributes for each instance, the Car Name, MPG, number of Cylinders, the amount of Displacement, Horsepower, Weight, Acceleration, Model type, and the Origin. For a car, what you are concerned may be whether it is a good car or not. Suppose you have a limited budget, well maybe a good car would be a car that could run fast and fuel-efficient, which means the car has a high value of Acceleration and MPG.

For the next step, we will explore our most interested attributes, MPG and, maybe, Acceleration. The following Fig.1 shows the plots of correlations across 7 attributes. By visual observation, we could see that there exist positive correlations between attribute MPG and attributes Model and Acceleration, negative correlations between attribute MPG and attributes Weight, Horsepower, Displacement and Cylinders.

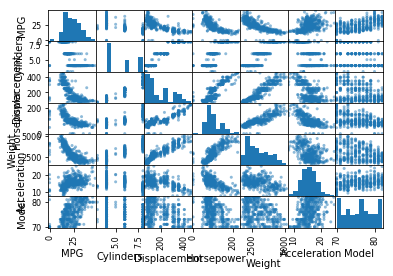


Figure 1 Correlation Plots of Attributes

From Fig.1 we can also know that Horsepower is in negative correlation with MPG, which is contrary to the common sense. However, the main factor here is the weight. According to physical laws, the larger the weight, the more power you need to drive up as well as to maintain the speed. That’s why MPG is low even the value of Horsepower is large.

Now let’s put several attributes together to see what are the relationships between MPG and the other 3 attributes including ‘Weight’, ‘Cylinders’ and ‘Horsepower’ together.

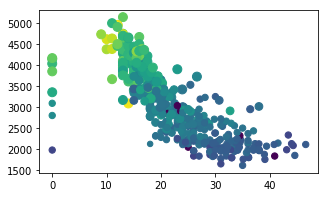


Figure 2 Relationships between MPG and Weight, Cylinders and Horsepower

Fig.2 provide a great visualization for these four attributes. The x axis is for MPG, y axis is for Weight. The size of nodes is for Cylinders, the larger size means the larger value of Cylinders. The color is for Horsepower, the darker the color, the smaller Horsepower. These results are in accordance with what we’ve already figured out from Fig.1.

So far, we are clear with the relationships among these attributes, the next and the most important question is, which car shall I choose? Or more specifically, which model of car shall I choose/which origin of car shall I choose/which brand of car shall I choose?

From Fig.1 we could see that there is a positive correlation relation between MPG and Model. The more detailed statistics could be seen in Fig.3. From the boxplot we could see the highest and the lowest value as well as the median, Q1, Q3 quartile values of different Models. The trend is as the model number increases, the MPG value increases. Observing this, when you decide to buy a car, maybe you need to consider a larger-numbered model.

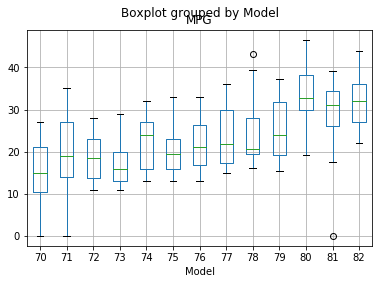
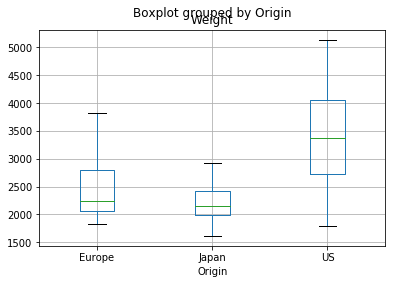
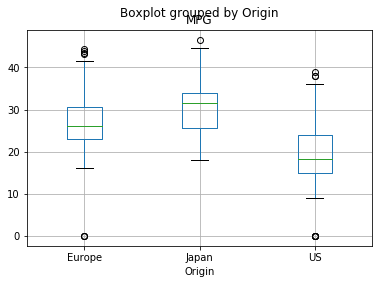
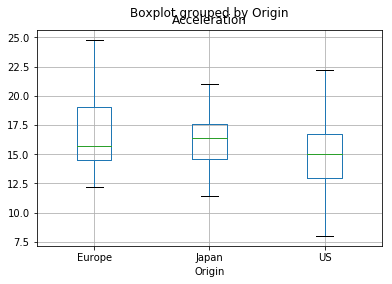
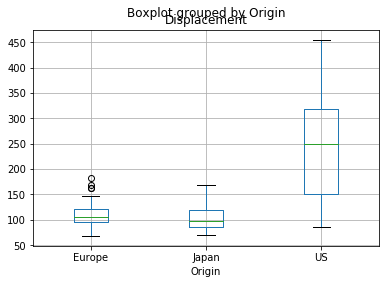


Figure 3 Boxplot grouped by Model

Besides the attribute Model, there are two other attributes that by now we haven’t talked about yet, which are Car and Origin. Putting the attribute Car aside, the following are the boxplot of Origin and some other attributes.





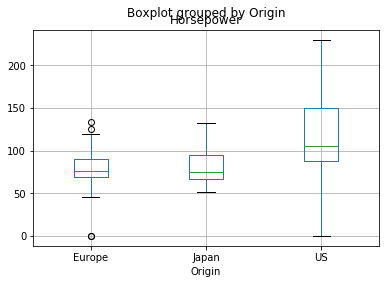


Figure 4 Boxplots of MPG and Other 5 Attributes

An obvious conclusion is that you are suggested to purchase a Japanese car. There are at lease three reasons. First is that over all Origins, Japanese cars have the largest median, high and low MPG value. Second is, comparing the other cars, Japanese cars have the lowest weight, thus need less fuel to drive and at the same time ‘Displacement’ value is also the lowest although European cars is almost the same ‘Displacement’ statistics as Japanese cars. Last but not the least is that although the Horsepower of Japanese cars are not very large, they still get a very impressive acceleration performance, comparing with cars from the other origins.

For the above reasons, I suggest that you need to consider Japanese cars if you are the first time to purchase a car. As for car brands, choose whatever you like.