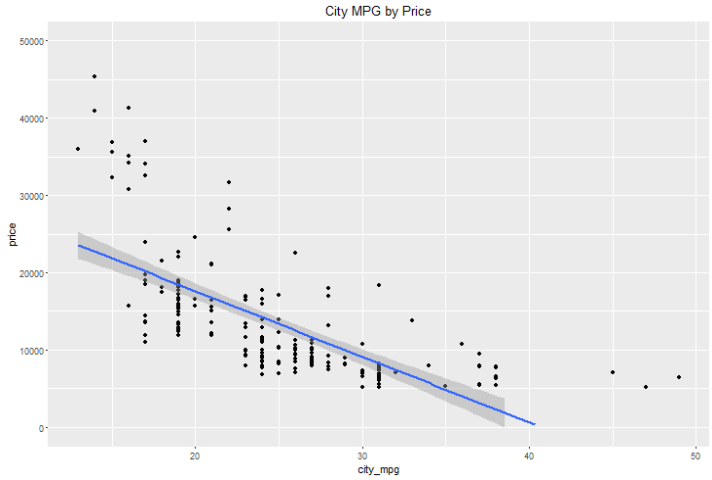
Method for Data Analysis – Assignment 1 – Takeaways

Josh Jensen

1. Investigating MPG

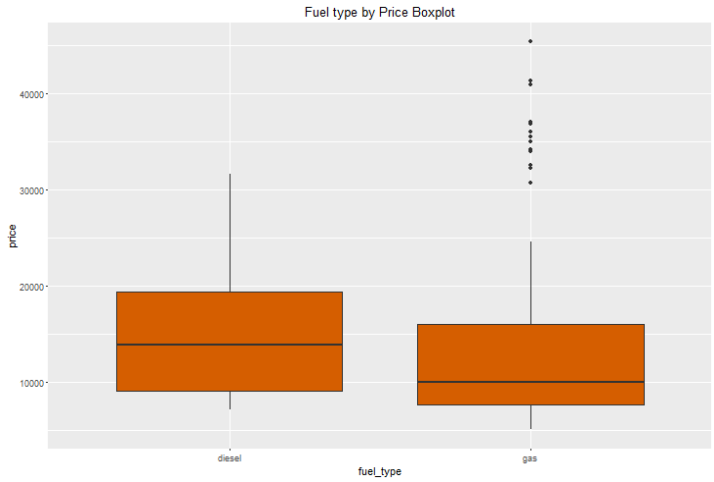
The price of automobile is dependent on MPG measures (city & highway). Looking at the scatter plots below, we observe that MPG measures are negatively correlated price. MPG is likely related to fuel type, so let's investigate this next.

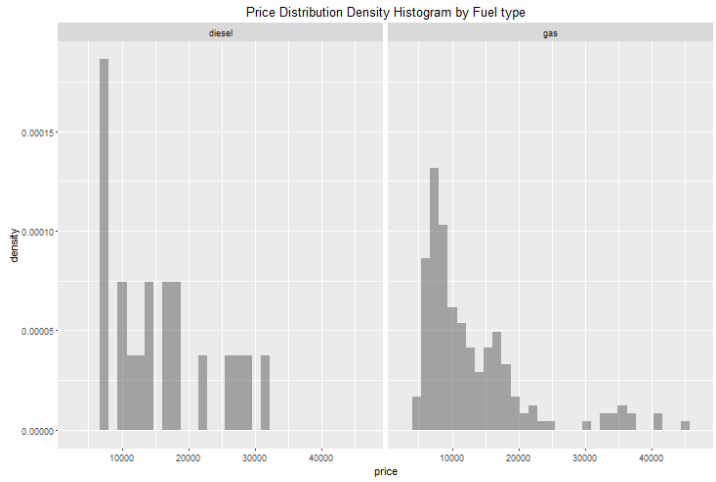


2. Investigating Fuel Type

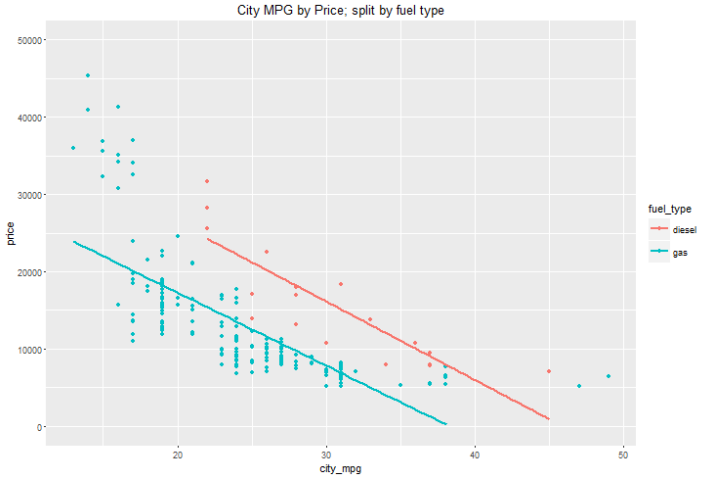
The price of automobiles is dependent on fuel type (gas vs. diesel). In calculating the mean price of each car in the sample by the respective fuel type, we observe that diesel car prices are on average 23% higher than gas cars. However, the standard deviation is very high, so this difference could be due to noise. In addition, diesel only has 20 observations, so small n is a factor here. Looking at the boxplot & histogram below we see that (1) the higher average is not due to a diesel outlier, (2) generally diesel vehicles have a slightly higher distribution of prices, and (3) gas vehicles have a small cluster of expensive vehicles.

```
  fuel_type mean_price sd_price
     (fctr)      (dbl)    (dbl)
1    diesel   15838.15 7759.844
2       gas   12916.41 7934.923
```





However, also bringing in the city MPG variable we see that in every case observed, diesel vehicles have a higher price than a gas vehicle of the same mileage (see scatter below). Next lets look at the cluster of expensive vehicles.



3. Investigating the cluster of expensive vehicles (> $30K)

The price of automobiles is dependent on engine size. We first create a flag to split expensive autos, with a price greater than $30K, from other autos. Using this flag we are able to summarize the mean of each column by this flag. Looking at these means, we are able to observe that the mean engine size of the expensive autos are almost double that of the non-expensive autos. Looking at the box plot below, we can see that the distribution of engine size varies dramatically between expensive and non-expensive autos. Further, in the scatter plot below we see that (1) there is a stark contrast in how expensive & non-expensive cars are grouped, and (2) engine size is positively correlated with price.