A Study on Vehicles' Gas Mileage via Linear Regression Models

Wenhe 'Wayne' Ye

September 25, 2015

Overview

Data Processing

First we need to call a few useful R packages to facilitate our analysis and load data *mtcars* into work space.

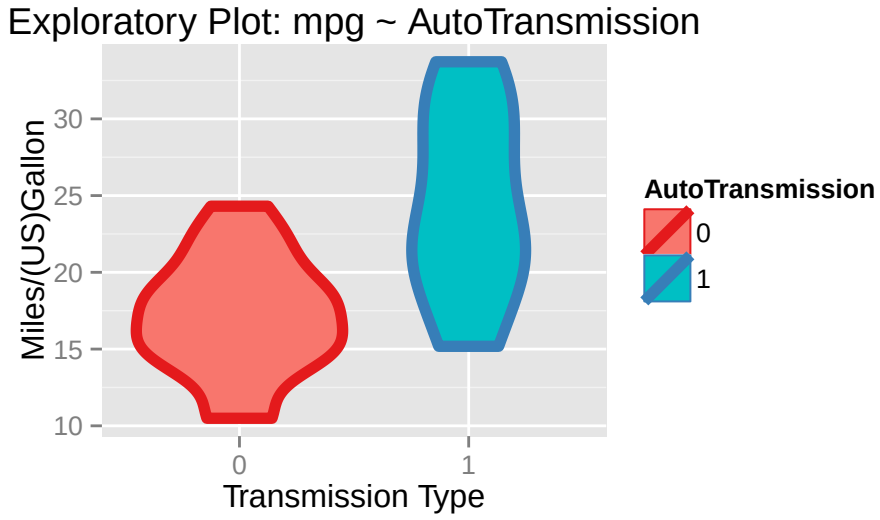
Clean up the raw data and convert some varibles into factors.

```
mtcars2<-mutate(mtcars,mpg,disp,wt,hp,Cylinder=as.factor(cyl),AutoTransmission=as.factor(am))
mtcars2<-select(mtcars2,mpg,disp,wt,hp,Cylinder,AutoTransmission)
```

Exploratory Data Analysis

Since we want to explore the relationship between the mpg and whether the cars are manual or auto transmission. We make a violin plot between mpg and factor of different transmission types to see the over all relationship.

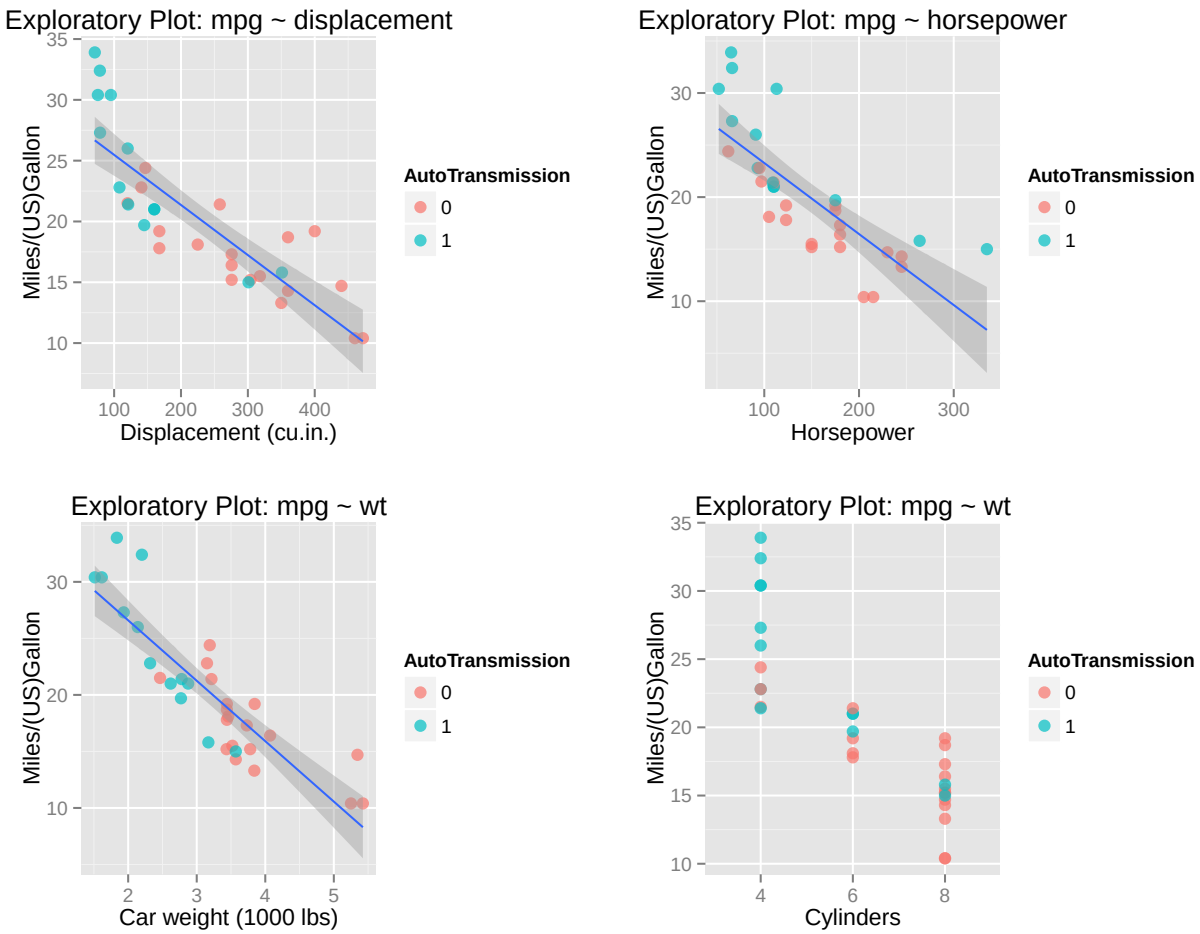
```
(g1<-ggplot(data=mtcars2,aes(x=AutoTransmission,y=mpg,col=AutoTransmission))+
        geom_violin(aes(fill=AutoTransmission),size=2)
+scale_color_brewer(palette="Set1")
+ggtitle("Exploratory Plot: mpg ~ AutoTransmission")
+labs(x="Transmission Type",y="Miles/(US)Gallon"))
```



From the plot we see manual transmission cars have a higher gas mileage over the automatic transmission. However, there might be other confounding variables need to be taken into account. The common sense

tells us that most high-end vehicles, probably those gas guzzlers have automatic transmission rather than manual transmission. The later one is more likely to be found on some compact vehicles especially on the basic editions. We select a few other variables as candidates to see their correlation with the mpg data. We picked displacement (disp), horsepower (hp), weight(wt) and number of cylinders(Cylinder) as confounding variables. It is worth noting here, we transform the cylinder number into factors rather than a continuous varible in the following study.

```
g2<-ggplot(data=mtcars2,aes(x=disp,y=mpg))+geom_point(size=3,alpha=0.7,aes(col=AutoTransmission))+
        geom_smooth(method="lm")+ggtitle("Exploratory Plot: mpg ~ displacement")+
        labs(x="Displacement (cu.in.)",y="Miles/(US)Gallon")
g3<-ggplot(data=mtcars2,aes(x=hp,y=mpg))+geom_point(size=3,alpha=0.7,aes(col=AutoTransmission))+
        geom_smooth(method="lm")+ggtitle("Exploratory Plot: mpg ~ horsepower")+
        labs(x="Horsepower",y="Miles/(US)Gallon")
g4<-ggplot(data=mtcars2,aes(x=wt,y=mpg))+geom_point(size=3,alpha=0.7,aes(col=AutoTransmission))+
        geom_smooth(method="lm")+ggtitle("Exploratory Plot: mpg ~ wt")+
        labs(x="Car weight (1000 lbs)",y="Miles/(US)Gallon")
g5<-ggplot(data=mtcars2,aes(x=Cylinder,y=mpg))+geom_point(size=3,alpha=0.7,aes(col=AutoTransmission))+
        ggtitle("Exploratory Plot: mpg ~ wt")+
        labs(x="Cylinders",y="Miles/(US)Gallon")
grid.arrange(g2,g3,g4,g5,ncol=2)
```



All the candidates showed some suspicious correlation with mpg. In order to quantify the difference between an auto transmission car and a manual transmission car, we need to carefully select the model to make our

estimation.

Regression Modeling

Conclusion