## Regression Analysis For MotorTrend

The dataframe mtcars contains 32 observations on 11 variabels like miles/gallon(MPG), number of cylinders etc.

Our main focus in the study is how the Transmission type( automatic or manual) affects the miles per gallon. define a relationship between mileage and transmission type.

### Loading and the Data

data("mtcars")  
mtcars <- mtcars%>%  
 mutate(am = as.factor (am))  
levels(mtcars$am)<- c("Automatic","Manual")  
  
summary(mtcars)

## mpg cyl disp hp   
## Min. :10.40 Min. :4.000 Min. : 71.1 Min. : 52.0   
## 1st Qu.:15.43 1st Qu.:4.000 1st Qu.:120.8 1st Qu.: 96.5   
## Median :19.20 Median :6.000 Median :196.3 Median :123.0   
## Mean :20.09 Mean :6.188 Mean :230.7 Mean :146.7   
## 3rd Qu.:22.80 3rd Qu.:8.000 3rd Qu.:326.0 3rd Qu.:180.0   
## Max. :33.90 Max. :8.000 Max. :472.0 Max. :335.0   
## drat wt qsec vs   
## Min. :2.760 Min. :1.513 Min. :14.50 Min. :0.0000   
## 1st Qu.:3.080 1st Qu.:2.581 1st Qu.:16.89 1st Qu.:0.0000   
## Median :3.695 Median :3.325 Median :17.71 Median :0.0000   
## Mean :3.597 Mean :3.217 Mean :17.85 Mean :0.4375   
## 3rd Qu.:3.920 3rd Qu.:3.610 3rd Qu.:18.90 3rd Qu.:1.0000   
## Max. :4.930 Max. :5.424 Max. :22.90 Max. :1.0000   
## am gear carb   
## Automatic:19 Min. :3.000 Min. :1.000   
## Manual :13 1st Qu.:3.000 1st Qu.:2.000   
## Median :4.000 Median :2.000   
## Mean :3.688 Mean :2.812   
## 3rd Qu.:4.000 3rd Qu.:4.000   
## Max. :5.000 Max. :8.000

### Exploratory Data Anlaysis

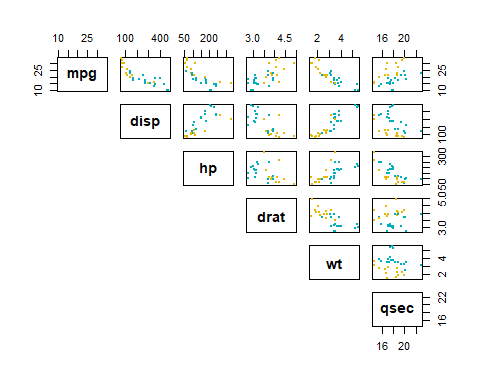
The Displacement ,Mileage, HorsePower, axle ratio, quator mile time, weight are all the continous variables.

And other varibles are categorical

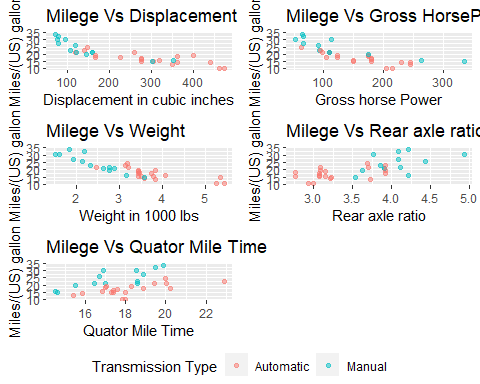
And our only intrest is to find relationship between Transmission type and Mileage.

we will analyse the continous variables sactter plot with mileage

mtcars.con <- mtcars[c("mpg","disp","hp","drat","wt", "qsec")]  
my\_cols <- c("#00AFBB", "#E7B800", "#FC4E07")   
pairs(mtcars.con ,pch = 19, cex =0.5, col = my\_cols[mtcars$am], lower.panel = NULL, font.labels = 2, cex.labels = 1.3)



g <- ggplot(data = mtcars, aes(x = disp, y = mpg, color = am))  
g <- g + geom\_point( alpha = 0.5)  
g <- g + labs(x = "Displacement in cubic inches", y = "Miles/(US) gallon", title = "Milege Vs Displacement", color = "Transmission Type")  
  
g1 <- ggplot(data = mtcars, aes(x = hp, y = mpg, color = am))  
g1 <- g1 + geom\_point( alpha = 0.5)  
g1 <- g1 + labs(x = "Gross horse Power", y = "Miles/(US) gallon", title = "Milege Vs Gross HorsePower", color = "Type")  
  
g2 <- ggplot(data = mtcars, aes(x = wt, y = mpg, color = am))  
g2 <- g2 + geom\_point( alpha = 0.5)  
g2 <- g2 + labs(x = "Weight in 1000 lbs", y = "Miles/(US) gallon", title = "Milege Vs Weight",color = "Type")  
  
g3 <- ggplot(data = mtcars, aes(x = drat, y = mpg ,color = am))  
g3 <- g3 + geom\_point( alpha = 0.5 )  
g3 <- g3 + labs(x = "Rear axle ratio", y = "Miles/(US) gallon", title = "Milege Vs Rear axle ratio", color = "Type")  
  
  
g4 <- ggplot(data = mtcars, aes(x = qsec, y = mpg ,color = am))  
g4 <- g4 + geom\_point( alpha = 0.5 )  
g4 <- g4 + labs(x = "Quator Mile Time", y = "Miles/(US) gallon", title = "Milege Vs Quator Mile Time", color = "Type")  
  
ggarrange(g,g1,g2, g3,g4, ncol = 2, nrow = 3,   
 common.legend = TRUE, legend = "bottom")



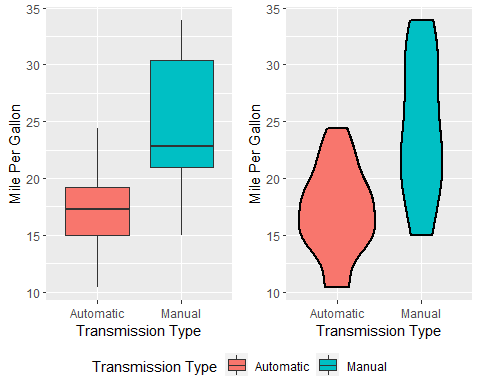
corr\_disp<- cor(mtcars$disp , mtcars$mpg)  
corr\_hp<- cor(mtcars$hp , mtcars$mpg)  
corr\_wt<- cor(mtcars$wt , mtcars$mpg)  
corr\_drat <-cor(mtcars$drat , mtcars$mpg)  
corr\_qsec<- cor(mtcars$qsec , mtcars$mpg)

\*\* The Correlation values of the different relationship \*\*

* The Plot shows a negative Trend with correlation values of **-0.848** between Displacement and Mileage.
* The Plot shows a negative Trend with correlation values of **-0.776** between HosrePower and Mileage.
* The Plot shows a negative Trend with correlation values of **-0.868** between weight and Mileage.
* The Plot shows a postive Trend with correlation values of **0.681** between rear axle ratio and Mileage.
* The Plot shows a postive Trend with correlation values of **0.419** between quator mile time and Mileage.

The Dependency of MPG value on Transmission Type is explained by the Bar and Violin Plots.

l<- labs(x = "Transmission Type", y = "Mile Per Gallon", fill = "Transmission Type")  
box <- ggplot(data = mtcars, aes(am , mpg, fill = am))  
box\_plot <- box+geom\_boxplot()+l  
violin <- box+geom\_violin(color = "black", size = 1)+l  
  
ggarrange(box\_plot, violin, ncol = 2, common.legend = TRUE, legend = "bottom")



The Box plot reveals that there is a huge differnce in mean mpg for the automatic and manual Transmission

Since, Our question of analysis is relationship between the Mileage with respect to transmission. And Displacement and Weight Shows high correlation with Mileage.

The Regression analysis of Mileage as outcome and Weight,Mileage and Type as Predictors.

### Model of Regression

First to test the Transmission Type is really a categorical value to determine the MPG.

t.test(mtcars$mpg~mtcars$am,conf.level=0.95)

##   
## Welch Two Sample t-test  
##   
## data: mtcars$mpg by mtcars$am  
## t = -3.7671, df = 18.332, p-value = 0.001374  
## alternative hypothesis: true difference in means is not equal to 0  
## 95 percent confidence interval:  
## -11.280194 -3.209684  
## sample estimates:  
## mean in group Automatic mean in group Manual   
## 17.14737 24.39231

The T-test rejects the null Hypothesis, the difference between Transmission on MPG is 0.

mdl <- lm(mpg~disp+wt+am , data = mtcars)  
coef\_mdl <- coef(mdl)  
rsquare\_val <- summary(mdl)$adj.r.squared

* The adjusted R square value is **0.757583**

|  |  |
| --- | --- |
| Feature | coeffcient value |
| Intercept | 34.6759109 |
| displacement | -0.0178049 |
| Weight | -3.2790439 |
| manual transmission | 0.1777241 |

### Model Selection

We can step method to as R to choose the best model itself

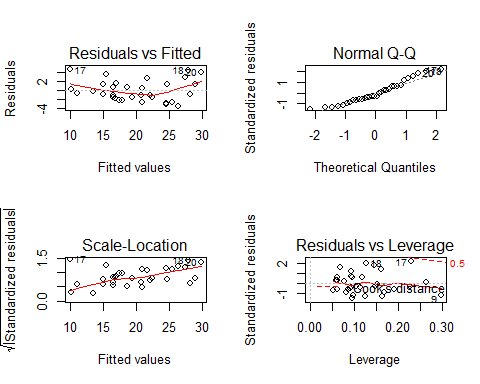
bestmodel = step(lm(mpg~. , data = mtcars), trace = 0)  
coef\_bdl <- coef(bestmodel)  
rsquare\_bval <- summary(bestmodel)$adj.r.squared  
vif\_model <- vif(bestmodel)

* The BestModel that fits perfectly for MPG as outcome is with predictors Weight,Quator Mile time and Transmission Type.
* The adjusted R square value for best model is **0.8335561**

|  |  |  |
| --- | --- | --- |
| Feature | coeffcient value | VIF |
| Weight | -3.9165037 | 2.4829515 |
| Quator mile Time | 1.225886 | 1.3643391 |
| manual transmission | 2.9358372 | 2.5414372 |

The Residual Plots for the Fitted values and inputs

par(mfrow = c(2,2))  
plot(bestmodel)



### Conclusion

Based on the previous analysis, we can say that on average manual transmission is better than automatic transmission by 2.9 mpg but also transmission type is not the only factor accounting for MPG, weight, and acceleration (1/4 mile time) also needs to be considered.