Regression Models Project Executive Summary

The goal of this study is to determine the difference in miles per gallon in automatic versus manual transmission vehicles. To estimate this, we will use linear regression models. The overall conclusion is that we cannot definitevely say whether the type of transmission affects MPG.

Preprocesing and Data Exploration

```
data(mtcars)
dim(mtcars)

## [1] 32 11

sum(is.na(mtcars))

## [1] 0
```

The mtcars dataset has 32 observations and 11 variables. There is no missing data. Looking at the summary statistics (see appendix), 41% of the observations are vehicles with automatic transmissions. the average MPG is approximately 20 MPG. Further, we observe that automatic vehicles get more MPG (see appendix box plot). Finally, with mean > median MPG shows a right skew.

Modeling, Model #1

```
model <- lm(mpg ~ am, data = mtcars)
summary(model)

##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.13e-15 ***
## am             7.245      1.764   4.106 0.000285 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

The first model built is a simple linear regression, with AM as the independent variable. The result indicates that automatic vehicles get 7.245 more MPG than manual transmission vehicles.
#Modeling, Model #2

```
model2 <- lm(log(mpg) ~ am + wt + cyl + hp, data = mtcars)
summary(model2)

##
## Call:
## lm(formula = log(mpg) ~ am + wt + cyl + hp, data = mtcars)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.14658 -0.07827 -0.03107  0.07064  0.24430
##
## Coefficients:
##               Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.8375665  0.1374861  27.912  < 2e-16 ***
## am           0.0266911  0.0638171   0.418 0.679077
## wt          -0.1645318  0.0407323  -4.039 0.000399 ***
## cyl         -0.0307380  0.0258070  -1.191 0.243991
## hp          -0.0011682  0.0006043  -1.933 0.063755 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.1111 on 27 degrees of freedom
## Multiple R-squared:  0.8787, Adjusted R-squared:  0.8608
## F-statistic: 48.91 on 4 and 27 DF,  p-value: 5.498e-12

vif(model2)

##       am       wt      cyl       hp
## 2.546159 3.988305 5.333685 4.310029
```

The second model uses am, wt, and hp as the independent variables and the dependent variable uses a natrual log to help address the right skew. Variable selection is based on the variables with the strongest covariance with MPG (see appendix). I chose to omit many to avoid multicollinearity. This model has an adj. r-squared of .8267 and shows that am is insignificant, and that only the wt variable is significant with a -16.4% change with each 1000 pound increase. The VIF doesn't indicate that multicollinearity is a concern. Finally, the fitted vs. residual plot (see appendix) shows a little evidence of heteroskedasticity.

Conclusion

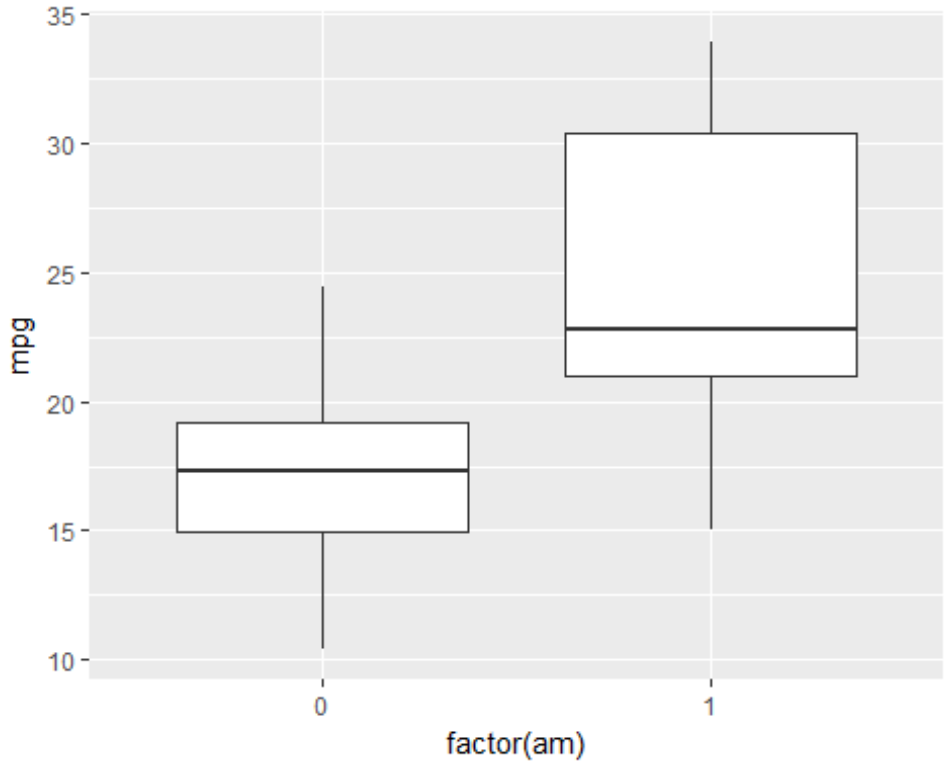
After fitting two models, this study did not observe a statistically significant difference in MPG for automatic and manual transmissions. As such, one cannot quantify the difference. The best one could estimate is ~ +7 MPG for automatics, not controlling for other variables.

Appendix

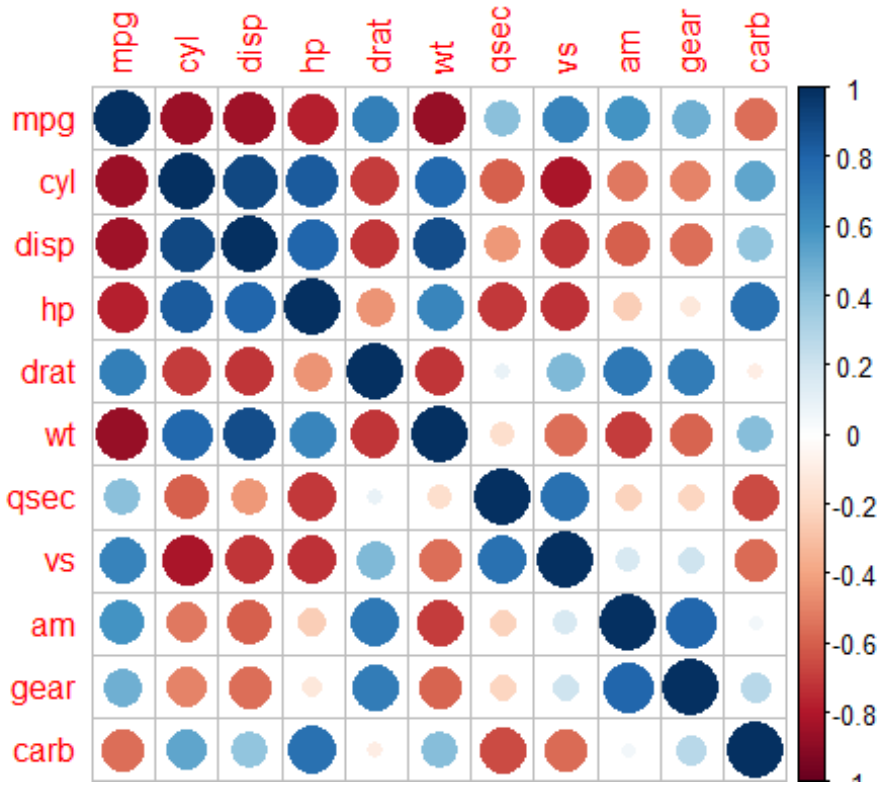
```r
#Summary Stats
summary(mtcars)
```

```
##       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      
##  Min.   :0.0000   Min.   :3.000   Min.   :1.000  
##  1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000  
##  Median :0.0000   Median :4.000   Median :2.000  
##  Mean   :0.4062   Mean   :3.688   Mean   :2.812  
##  3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000  
##  Max.   :1.0000   Max.   :5.000   Max.   :8.000
```

```r
#Box Plot
ggplot(mtcars, aes(x = factor(am), y = mpg)) + geom_boxplot()
```




```
#Correlations
corrplot(cor(mtcars))
```



```r
#Model 2 Plots
par(mfrow = c(2,2))
plot(model2)
```