# MPG Analysis of Transmission Type in Motor Trend Data

## Executive Summary

For this project, we were asked "You work for Motor Trend, a magazine about the automobile industry. Looking at a data set of a collection of cars, they are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome)." With focus on the effect of the transmission type for MPG. We will show that the data suggests that manual transmissions are better, however once we include the appropriate variables in our model, it cannot be shown with high probability.

## Load Data

Perform an initial data load while setting the appropriate columns as factors

```
data(mtcars)
mtcars$am <- factor(mtcars$am, labels=c('Automatic', 'Manual'))
mtcars$cyl <- factor(mtcars$cyl)
mtcars$vs <- factor(mtcars$vs)
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
```

## Regression Model

I used model nesting and variance inflation to determine the right mode to use.

### Model Selection

After review of the variables, and given the question, we tested a model with just the transmission variable vs. a model that includes variables with low variable inflation... and tried nesting model testing to sow the correct model.

This version nests and starts to add appropraite varaibles. Based on the anova output, fit4 is the appropriate model.

```
fit1 <- lm(mpg ~ am, data = mtcars)
fit2 <- lm(mpg ~ am + wt, data = mtcars)
fit3 <- lm(mpg ~ cyl + am + wt, data = mtcars)
fit4 <- lm(mpg ~ cyl + hp + wt + am, data = mtcars)
fit5 <- lm(mpg ~ cyl + hp + wt + am + drat + gear, data = mtcars)
fit6 <- lm(mpg ~ cyl + hp + wt + am + drat + gear + carb, data = mtcars)
anova(fit1, fit2, fit3, fit4, fit5, fit6)$"Pr(>F)"
```

```
## [1]         NA 4.891e-07 8.683e-03 5.561e-02 9.802e-01 8.919e-01
```

We then compare the Adjusted R-squared which clearly shows the model with the best coverage is fit4, and also clearly shows just taking into account transmission type as a variable would be a mistake.

```r
c(summary(fit1)$adj.r.squared, summary(fit2)$adj.r.squared, summary(fit3)$adj.r.squared, summary(fit4)$a
```

```
## [1] 0.3385 0.7358 0.8134 0.8401 0.8209 0.7900
```

## Analysis/Interpretation

The output of the fit4 model suggests that mpg decreases as you add either cylinders (the intercept is 4 cylinders) weight, or horsepower.

```r
fit4$coeff
```

```
## (Intercept)        cyl6        cyl8          hp          wt    amManual
##    33.70832    -3.03134    -2.16368    -0.03211    -2.49683     1.80921
```

To the two core questions:

### "Is an automatic or manual transmission better for MPG"

If you look at the fit1 model that only includes transmission type, you would conclude that manual transmissions are better. The fit4 model also suggests this, however it does so with poor probability. So, you have to conclude that with the right coefficients taken into account, we cannot say that a manual transmission has a high probability of having better mpg given the data.

### "Quantify the MPG difference between automatic and manual transmissions"

According to a model only taking into account transmission, the answer is:

```r
coef(fit1)[2]
```

```
## amManual
##    7.245
```

However, we know that is not the right model. The appropriate model, fit4, I've selected shows it to be a much smaller increase with a low probablity.

```r
#Estimate of manual transmission
fit4summary <- summary(fit4)$coeff
fit4summary[6,1]
```

```
## [1] 1.809
```

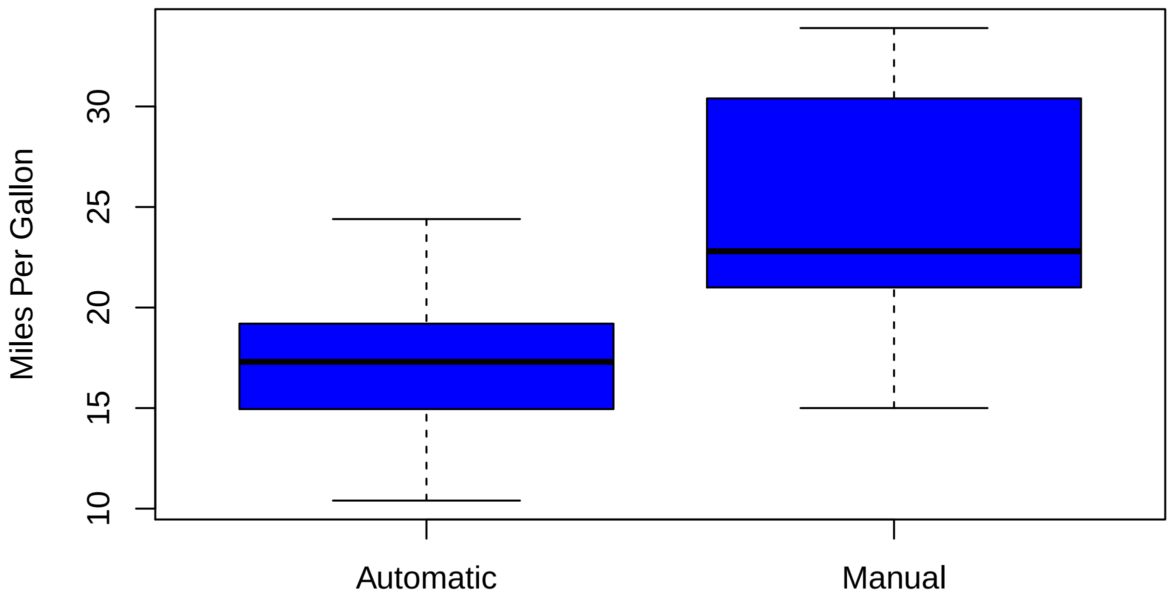
```r
#Probability of manual transmission
fit4summary[6,4]
```

```
## [1] 0.2065
```

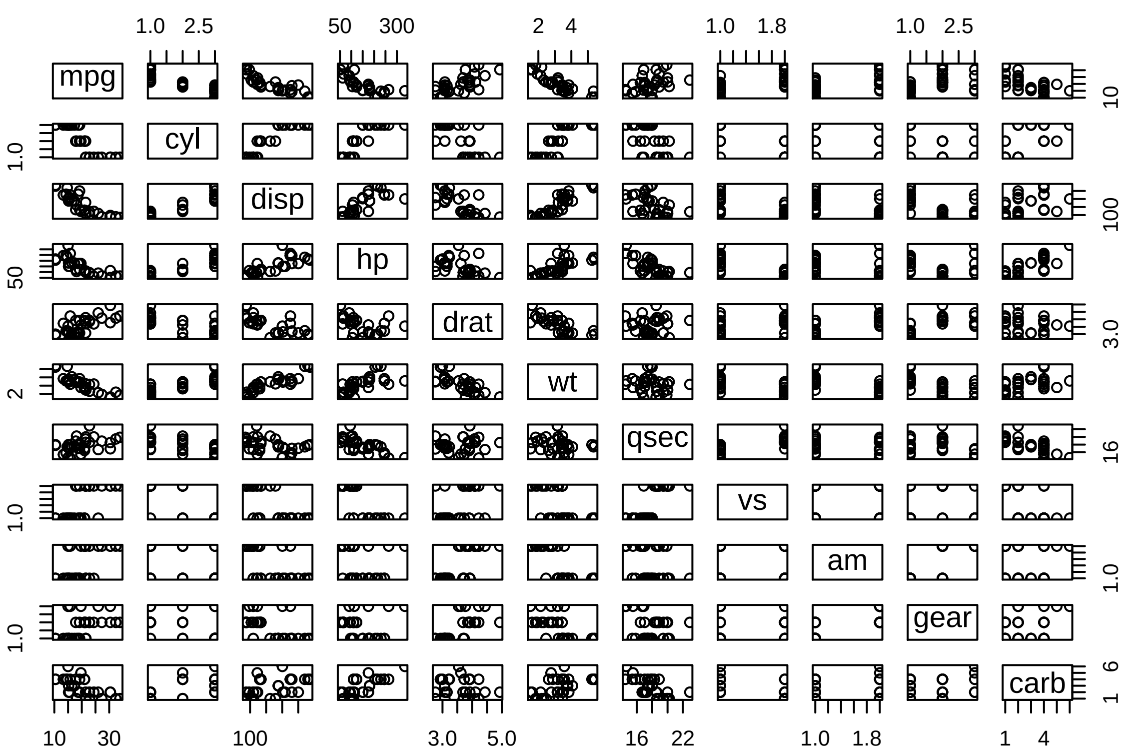
# Appendix

## Data Exploration

### Boxplot of MPG by transmission Type




### Pairs of mtcars data



## Residual Plot