ASSIGNMENT 1: MOTOR TREND CAR ROAD TESTS

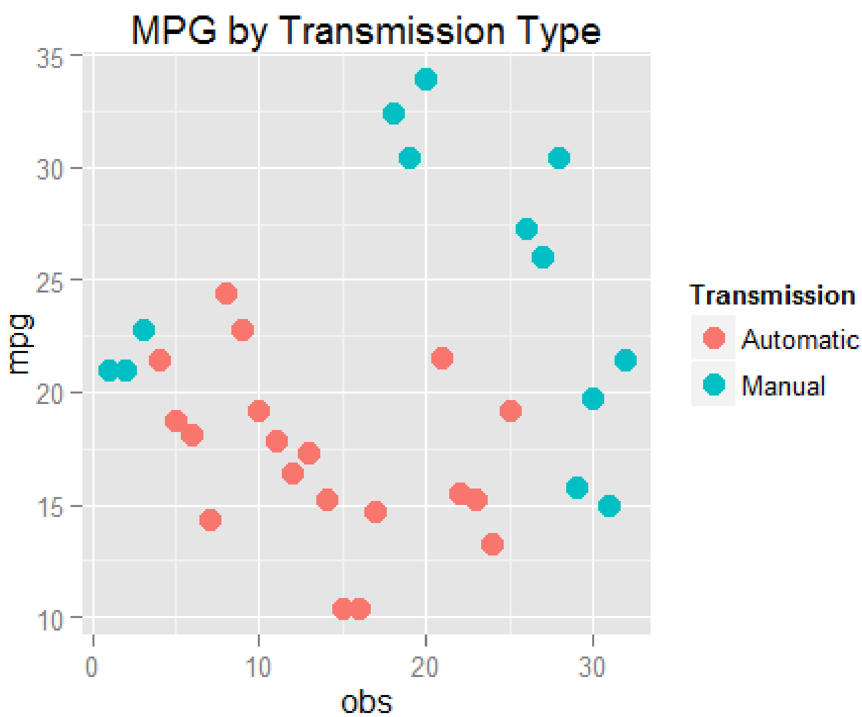
```
library(ggplot2)
```

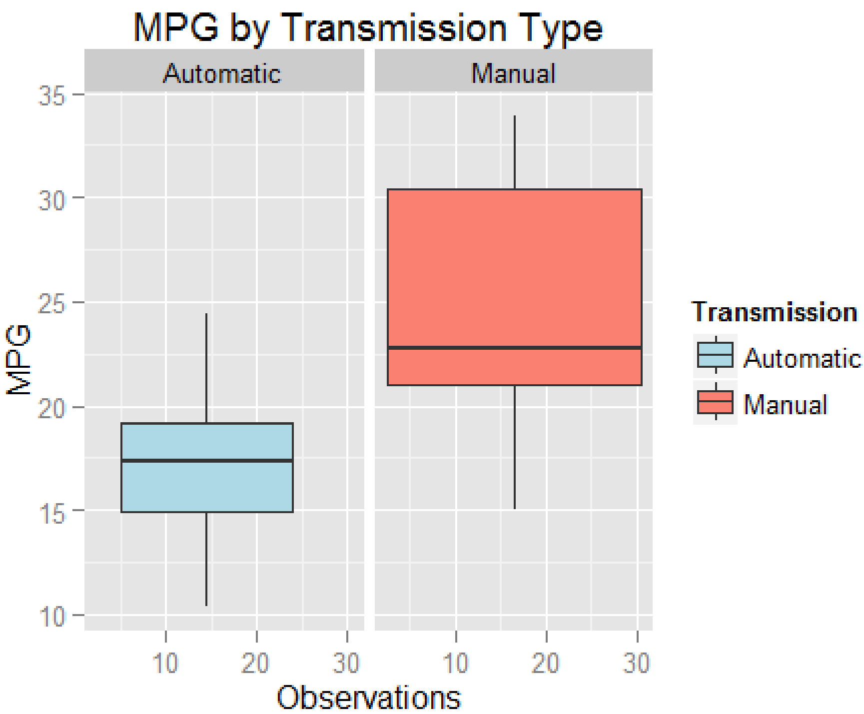
EXECUTIVE SUMMARY

The data was extracted from the 1974 Motor Trend US magazine, and comprises fuel consumption and 10 aspects of automobile design and performance for 32 automobiles. My analysis has shown that weight and number of cylinders are the best predictors of the car's MPG. The transmission type was important in the univariate regression, indicating a positive relationship between the two variables, however, it falls out as insignificant when weight is added into the equation.

DATA ANALYSIS

First, let's plot vehicles' MPGs by transmission type and inspect the data visually.





Let's run a T-test to compare mean MPGs by transmission type and infer whether they differ.

```
ttest <- t.test(mpg~am, mtcars, paired=FALSE, var.equal=FALSE, conf.level=0.95)
```

The T-test indicates that we can reject the null hypothesis that the means are the same at a 95% confidence level. The mean MPG for cars with automatic transmission is lower than that for cars with manual transmission.

Let's explore linear relationships between the MPG and other variables. We will start with a simple regression of MPG on transmission type:

The regression output indicates that transmission type has a significant impact on the MPG. The average MPG for the automatic transmission type is 17.15. However, for the cars with manual transmission, another 7.24 are added to the average MPG. The R-Squared for this regression is 0.36 meaning that 36% of the variance in MPG is explained with the transmission type.

Now, let's try a more complex model:

```
eq2 <- lm(mpg~am+wt, data=mtcars)
```

When we add the weight variable, the coefficient sign on the transmission type changes and the variable becomes insignificant. It is possible that there exists a relationship between transmission type and weight, which makes transmission type redudnant in the MPG equation.

```
eq3 <- lm(wt~am, data=mtcars)
```

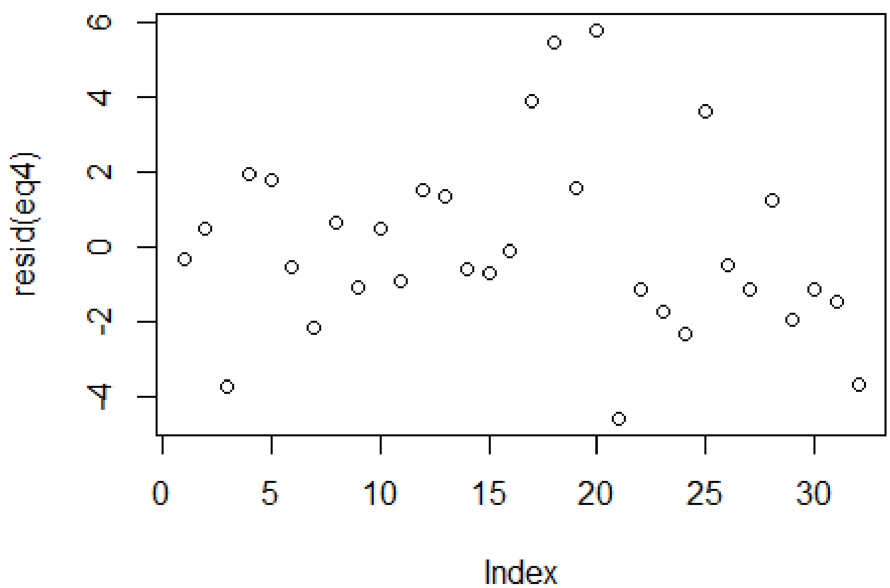
The regression of weight on transmission type does indicate a significant negative relationship between the two variables.

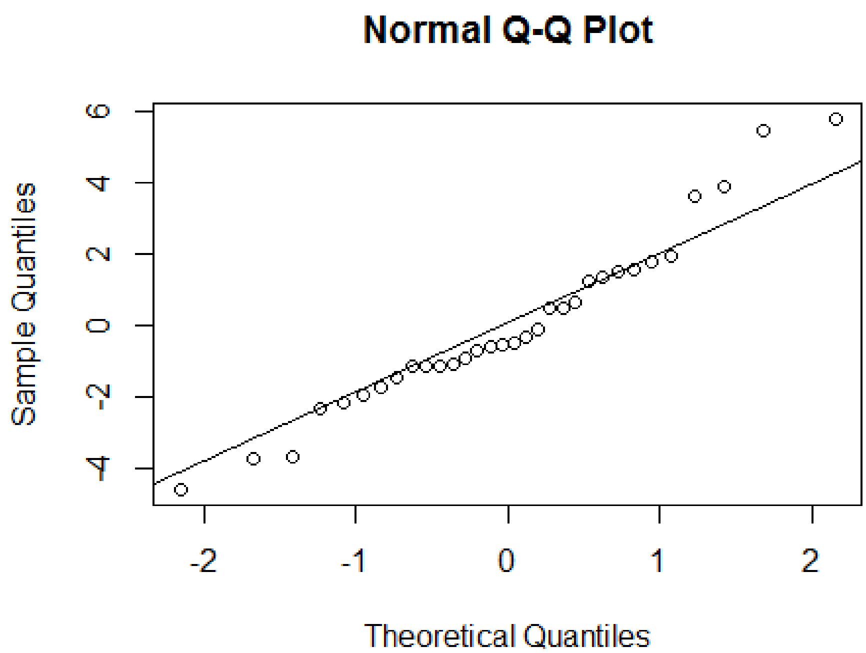
Let's drop the transmission type and regress MPG on weight and number of cylinders, where cylinders is a dummary variable.

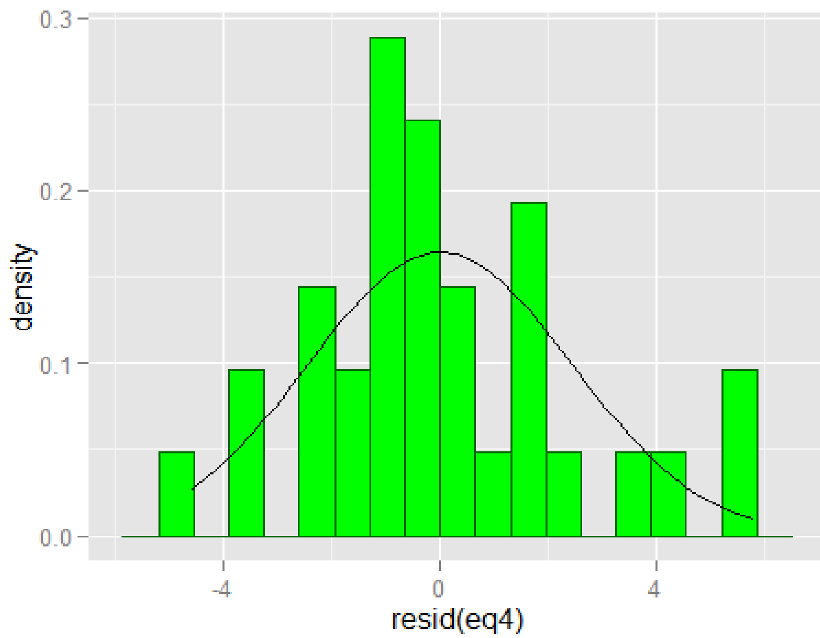
```
mtcars$cyl <- as.factor(mtcars$cyl)
eq4 <- lm(mpg~wt+cyl, data=mtcars)
```

The regression results demonstrate that the number of cylinders is a significant variable. Relative to the base number of 4 cylinders, a change to 6 cylinders reduces average MPG by 4.25 miles, and a change to 8 cylinders reduces average MPG by 6.07 miles. The residual standard error in the regression is 2.56, which is lower than that what we observed before. The R-Squared is 0.84, which implies that 84% of variance in the dependent variable is explained by the regression. Considering that none of the other models I ran performed better, I am going to choose this model as my final model.

Now, let's look at the residuals:







The Q-Q plot and the histogram of the residuals show that the residuals are clustered around the middle of the normal distribution. We do also observe more significant divergence from the normal quantile line at both ends of the graph.

CONCLUSION

Based on my analysis, car weight and the number if cylinders are the better predictors of the MPG. When regressing MPG on transmission type alone, we find a significant relationship between the two variables. However, when weight is added to the regression, transmission type falls out as no longer significant. Given that the transmission type and the weight also have a statistically significant relationship, I assume that a lot of the variation in MPG explained by the transmission type is also explained by the weight.