MotorTrend - MPG vs. Transmission

Executive Summary

In this analysis, data from 1974 Motor Trend US Magazine comprising of fuel consumption and 10 aspects of automible design & performance for 32 automobiles (1973-74 models) is used to answer the following questions:

1. Is an automatic or manual transmission better for MPG?
2. Quantify the MPG difference between automatic and manual transmissions

Data Model

Description of the data model is found @ https://stat.ethz.ch/R-manual/R-devel/library/datasets/html/mtcars.html.

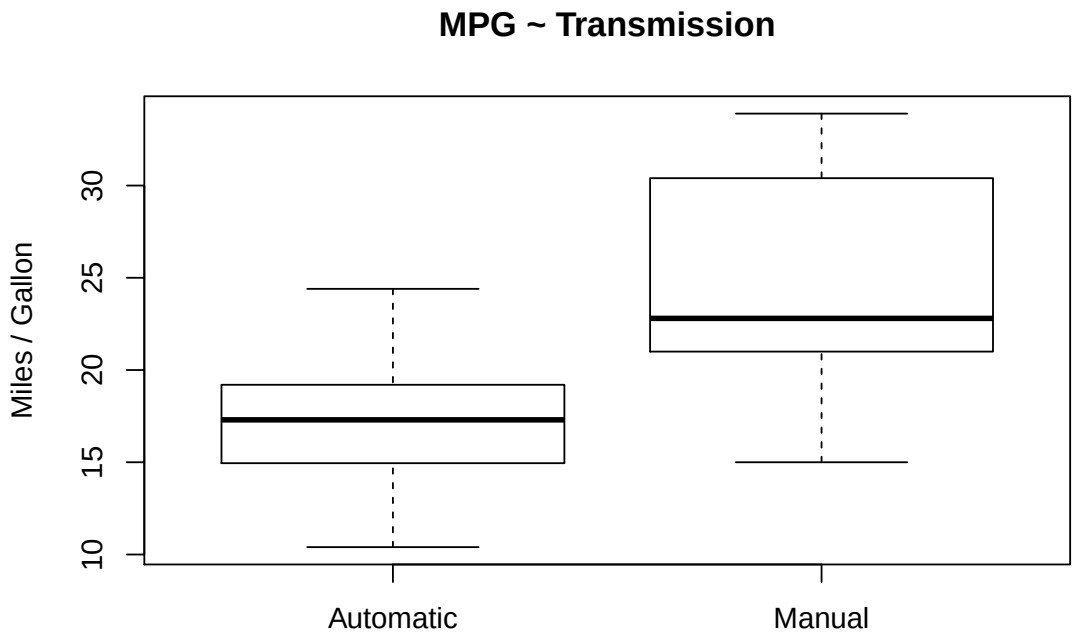
Loading the data model, reviewing the structure, and setting it up for exploratory analysis.

```
library(datasets)
data(mtcars)
mtcars$am <- factor(mtcars$am, labels = c("Automatic", "Manual"))
```

Exploratory Analysis - Single predictor regression

Lets check visually to see if a correlation exists between mpg and transmission.

```
boxplot(mpg ~ am, data = mtcars, ylab = "Miles / Gallon", main = "MPG ~ Transmission")
```



A linear regression model with one predictor, Transmission, can provide a quick answer to the second question:

```
fit <- lm(mpg ~ am, data = mtcars)
summary(fit)$coefficients
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 17.147368   1.124603 15.247492 1.133983e-15
## amManual     7.244939   1.764422  4.106127 2.850207e-04
```

1. The p-value is 0.000285, so the null hyppothesis is rejected (i.e. there is statistical difference in MPG for cars with Manual vs. Automatic transmissions).
2. Manual transmission is better than Automatic transmission by 7.245 (am Estimate)

However the regression model covers only 35.98% of the variance. So we need to continue the regression analysis to include more variables to provide a definite conclusion.

Exploratory Analysis - Multi predictor regression

Lets use the step() method to find the best fit.

```
fit.best <- step(lm(mpg ~ ., data = mtcars), trace = 0)
summary(fit.best)
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.6178     6.9596   1.382 0.177915
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## amManual      2.9358     1.4109   2.081 0.046716 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

The 3 predictors (wt = Weight, qsec = , am = 1/4 mile time, Transmission) contributes to 84.97% of variance. As the p-value is 0.046716, we can safely reject the null hypothesis. However the impact on MPG is now decreased to 2.9358 (amMT estimated coefficient).

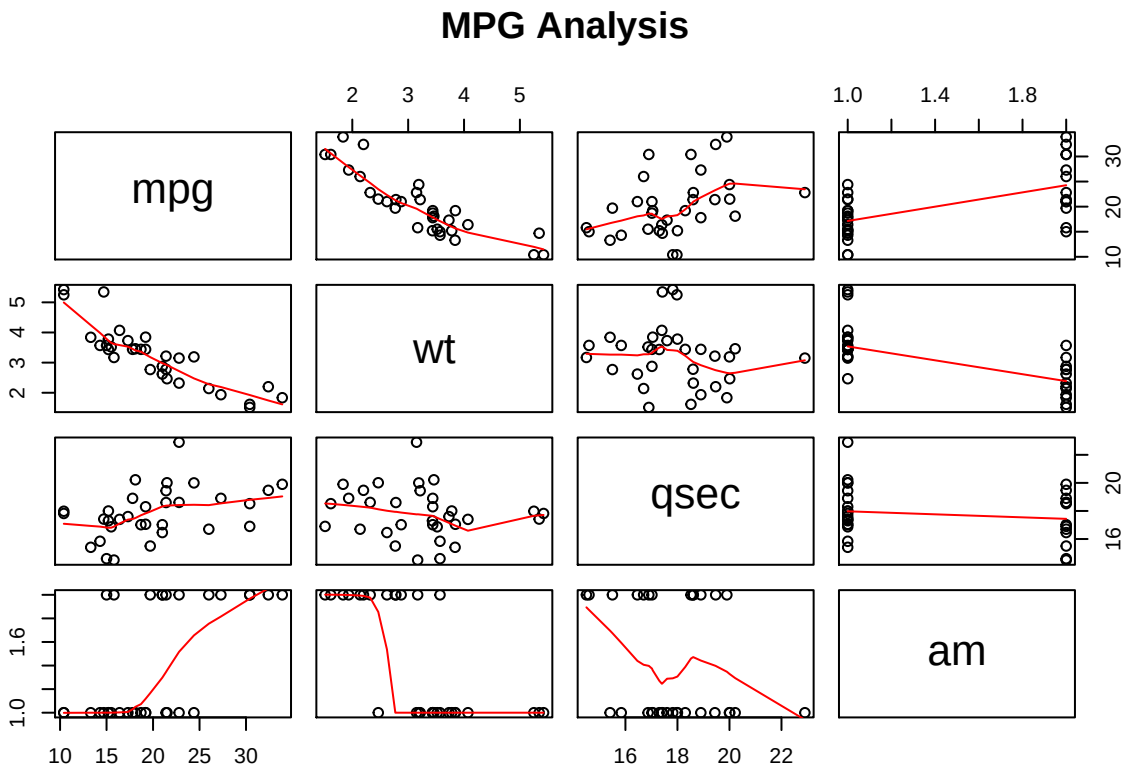
Conclusion

Manual transmission is better for MPG. However, its signficance is reduced (lowers) when we consider the other predictors (such as weight), so requires a larger set of data (including driving conditions, road type) to make a definite conclusion.

Appendix

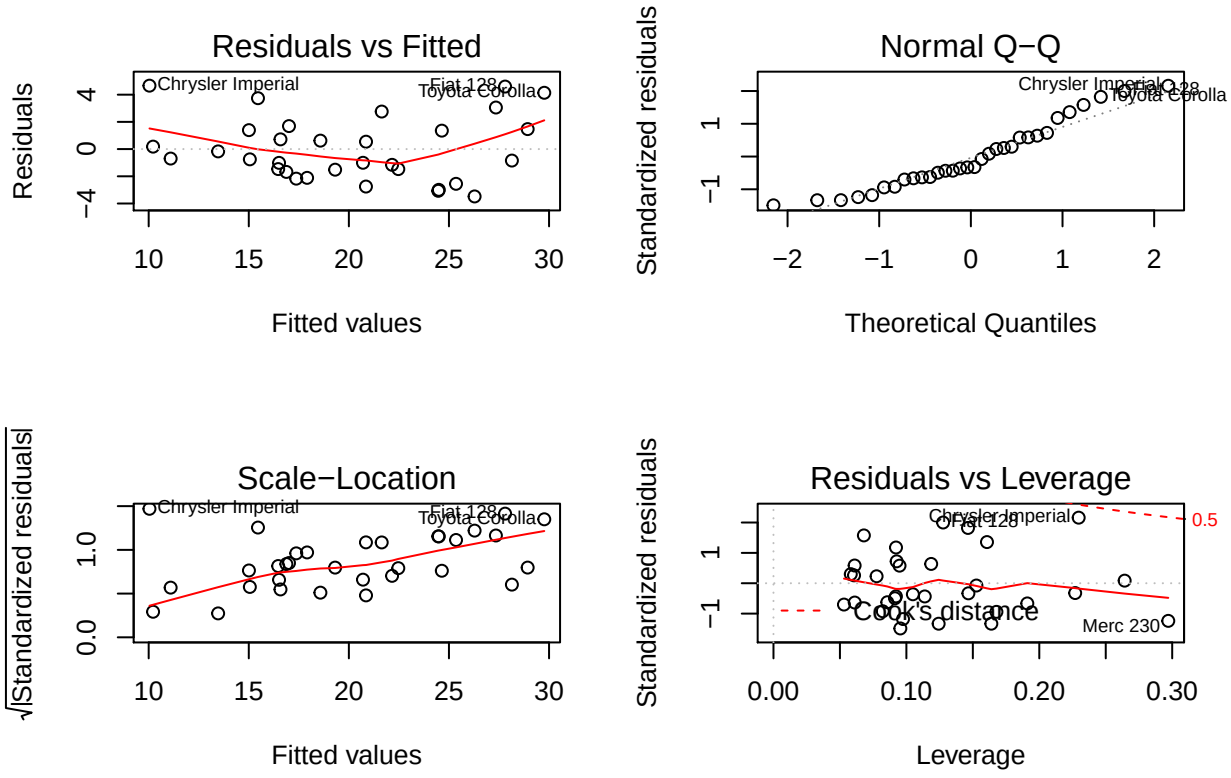
Appendix 1 - Relation between significant variables

```
pairs(mtcars[, c("mpg", "wt", "qsec", "am")], panel = panel.smooth, main = "MPG Analysis")
```



Appendix 2 - Residuals for best fit

```
par(mfrow = c(2, 2))
plot(fit.best)
```



Obervations - The higher MPG values have higher residuals. Atleast 3 cars with high residual values, indicating a poor fit for these cars. It will be worth exploring further if these high influence and / or leverage bais the fit.