Regression Models Final Project

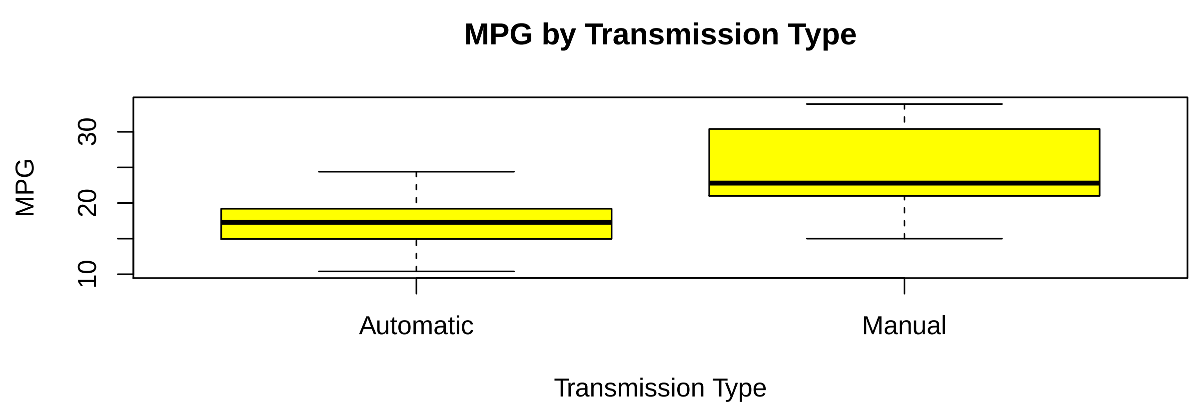
Executive Summary

This report investigates the relationship between automobile transmission type and fuel efficiency, leveraging the mtcars data set. The data were extracted from the 1974 Motor Trend US magazine, and comprise fuel consumption and 10 aspects of automobile design and performance for 32 automobiles (1973-74 models).

The data indicate that, within the sample set of vehicles, those with a manual transmission achieve higher miles per gallon (MPG) than those with an automatic transmission by a factor of 1.8. However, this ratio does not account for other contributing factors, including relative performance and weight of vehicles. Within the sample set, vehicles with automatic transmissions tend to have more powerful, heavier engines, which will inherently reduce fuel efficiency due to engineering for power (versus efficiency) and the increase in vehicle weight being propelled.

Exploratory Data Analysis

```r
library( datasets ); data( "mtcars" )
mtcars$am <- factor( mtcars$am, labels = c( "Automatic", "Manual" ) )
mtcars$cyl <- as.factor( mtcars$cyl )
mtcars$vs <- as.factor( mtcars$vs )
mtcars$gear <- as.factor( mtcars$gear )
mtcars$carb <- as.factor( mtcars$carb )
str( mtcars )
boxplot(mpg ~ as.factor(am), data = mtcars, col="yellow", xlab = "Transmission Type",
        ylab = "MPG", main="MPG by Transmission Type")
```



Hypothesis Testing

An initial visual interpretation indicates that vehicles with manual transmissions have better MPG performance than vehicles with automatic transmissions. This can be confirmed with a t-test:

```r
t.test(mpg~am,mtcars,paired=FALSE,var.equal=FALSE)$p.value
```

```
## [1] 0.001373638
```

The resulting p-value of 0.0014 is less than 0.05, which confirms the hypothesis.

Predictive Model Fitting

```r
simplified <- lm( mpg ~ am, data = mtcars )
summary( simplified, digits = 3 )$adj.r.squared
```

```
## [1] 0.3384589
```

The adjusted R^2 value illustrates that the simplified model accounts for 34% of variance.

```r
extended <- step( lm( mpg ~ . , mtcars ), trace = 0 )
summary( extended )$adj.r.squared
```

```
## [1] 0.8400875
```

The adjusted R^2 value illustrates that the extended model accounts for 84% of variance.

```r
anova( simplified, extended )
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ cyl + hp + wt + am
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     26 151.03  4    569.87 24.527 1.688e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Due to the low F value within the variance analysis, it is preferable to use the extended model for the remainder of the exercise.

Conclusion

```r
summary(extended)$coeff[c("(Intercept)", "amManual"),]
```

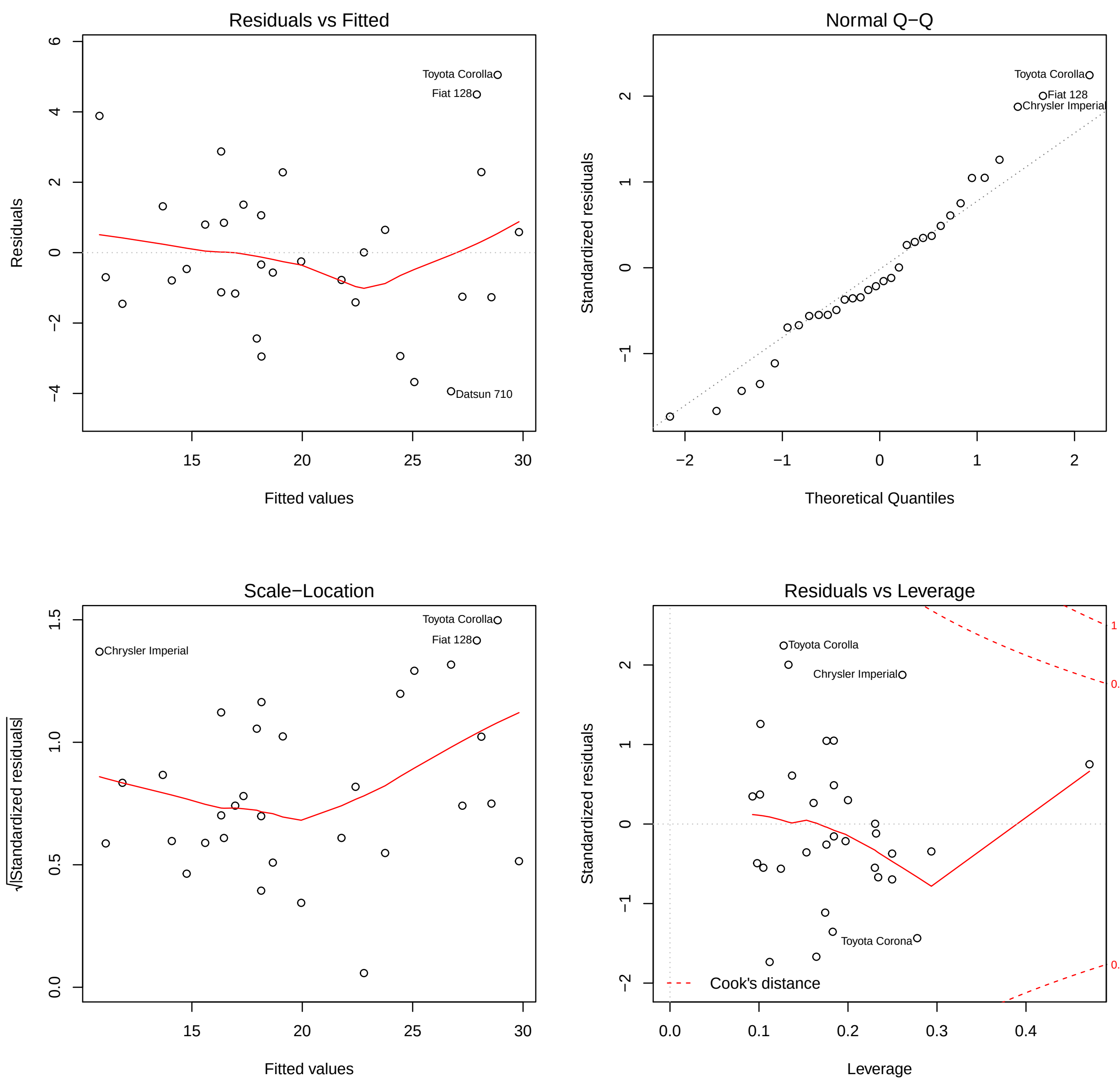
```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 33.708324   2.604886 12.940421 7.733392e-13
## amManual     1.809211   1.396305  1.295714 2.064597e-01
```

Based upon the analysis of this data set, we can conclude that manual transmissions operate at a higher MPG than automatic by a factor 1.8092114 mpg.

Appendix

Comparison of Residuals and Fitted Values

```
par( mfrow = c( 2, 2 ) )
plot( extended )
```



Residuals versus Fitted and Scale-Location plots do not indicate a pattern. A Normal Q-Q plot plot suggests that Residuals follow an approximate Normal distributions. Residuals versus Leverage plot does not incidate any statistically-notable outliers.

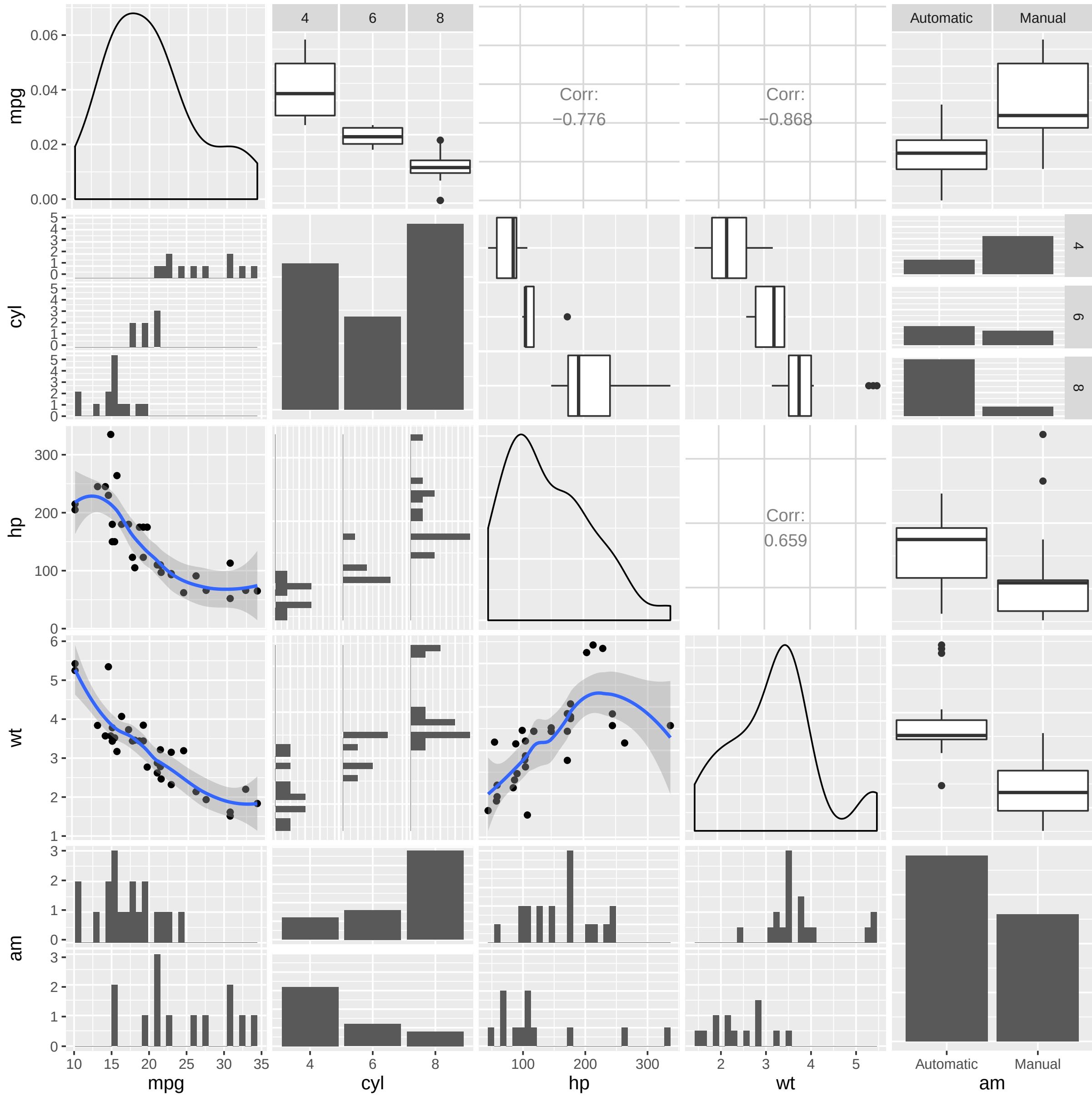
Pairwise Plot Matrix

```
require(GGally); require(ggplot2)
```

```
## Warning: package 'GGally' was built under R version 3.1.3
```

```
## Warning: package 'ggplot2' was built under R version 3.1.3
```

```
g <- ggpairs( mtcars[, c( "mpg", "cyl", "hp", "wt", "am") ],
        lower = list(continuous = function( data, mapping, method="loess", ...){
                p <- ggplot( data = data, mapping = mapping ) +
                        geom_point() +
                        geom_smooth( method = method, ... )
                p
        } ) )
g
```


mpg
cyl
hp
wt
am
4
6
8
Automatic
Manual
Corr:
−0.776
Corr:
−0.868
Corr:
0.659