mtcars Analysis

##Overview

The data from mtcars will be looked at and analyzed to determine the relationship between the miles per gallon and the transmission of the car.

##Exploratory Analysis

The data is first loaded and converted to factors

```
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

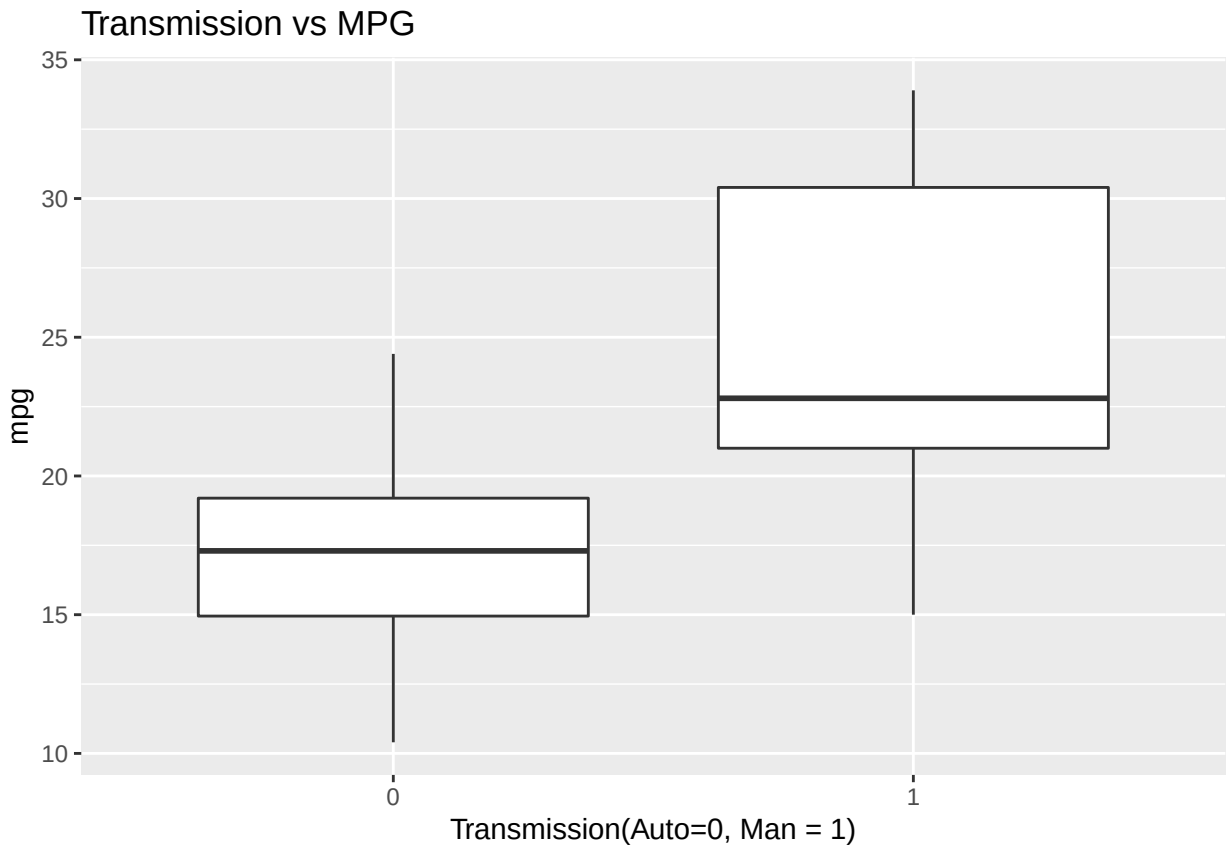
```
require(gridExtra)
```

```
## Loading required package: gridExtra
```

```
## Warning: package 'gridExtra' was built under R version 3.6.3
```

```
data(mtcars)
mtcars$am = as.factor(mtcars$am)
mtcars$cyl = as.factor(mtcars$cyl)
mtcars$gear = as.factor(mtcars$gear)
mtcars$car = as.factor(mtcars$carb)
mtcars$vs = as.factor(mtcars$vs)
auto = subset(mtcars, mtcars$am == 0)
man = subset(mtcars, mtcars$am == 1)
```

A quick plot is setup in order to see how the data is distributed amonst transmission and mpg



As can be seen in the figure above, manual cars appear to have a higher maximum mpg and minimum mpg. The range in mpg widely varies amongst manual cars compared to automatic.

```
t.test(man$mpg, auto$mpg)
```

```
##
##  Welch Two Sample t-test
##
## data:  man$mpg and auto$mpg
## t = 3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##   3.209684 11.280194
## sample estimates:
## mean of x mean of y
##  24.39231  17.14737
```

In addition, as can be seen in the two item T test of both the manual and automatic transmissions, there is a 95% confidence interval of 3.21-11.28 showing that there is likely to be some difference between manual and automatic transmissions. The p value is low enough to reject the notion that there is no difference between mpg amongst transmissions and that manual is better for mpg than automatic. In addition, the difference in the means can be easily seen with manual have an average mpg of 24.39 and automatic have an anverage of 17.1

##Regression modeling

```r
model1 = lm(mpg ~ ., data=mtcars)
summary(model1)
```

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.5087 -1.3584 -0.0948  0.7745  4.6251
##
## Coefficients: (1 not defined because of singularities)
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 22.84336   19.99148   1.143   0.2711
## cyl6        -2.64870    3.04089  -0.871   0.3975
## cyl8        -0.33616    7.15954  -0.047   0.9632
## disp         0.03555    0.03190   1.114   0.2827
## hp          -0.07051    0.03943  -1.788   0.0939 .
## drat         1.18283    2.48348   0.476   0.6407
## wt          -4.52978    2.53875  -1.784   0.0946 .
## qsec         0.36784    0.93540   0.393   0.6997
## vs1          1.93085    2.87126   0.672   0.5115
## am1          1.21212    3.21355   0.377   0.7113
## gear4        1.11435    3.79952   0.293   0.7733
## gear5        2.52840    3.73636   0.677   0.5089
## carb         1.03577    1.19437   0.867   0.3995
## car2        -2.01513    2.20142  -0.915   0.3745
## car3         0.92809    3.40346   0.273   0.7888
## car4        -2.01590    2.89087  -0.697   0.4963
## car6        -0.70130    4.25377  -0.165   0.8713
## car8              NA         NA      NA       NA
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.833 on 15 degrees of freedom
## Multiple R-squared:  0.8931, Adjusted R-squared:  0.779
## F-statistic:  7.83 on 16 and 15 DF,  p-value: 0.000124
```

A preliminary linear model on the mtcars dataset blindly checking the effect other variables would have on the mpg shows that mpg is not highly dependent on any of the independent variables. However, the coffecient of am is about 2.52023 which means that it has a greater weight in determining the mpg, as the value goes higher(car is manual).

```r
data(mtcars)
sort(cor(mtcars)[1,])
```

```
##         wt        cyl       disp         hp       carb       qsec       gear
## -0.8676594 -0.8521620 -0.8475514 -0.7761684 -0.5509251  0.4186840  0.4802848
##         am         vs       drat        mpg
##  0.5998324  0.6640389  0.6811719  1.0000000
```

The correlation of the data is checked to see which variables are related to mpg. wt, cyl, disp, and hp all appear to be highly correlated to mpg. disp and cyl appear correlated to one another and would be as higher

cylinders in a engine would be capable of greater displacement. wt and hp would generally be typical logical guesses to indicators of mpg.

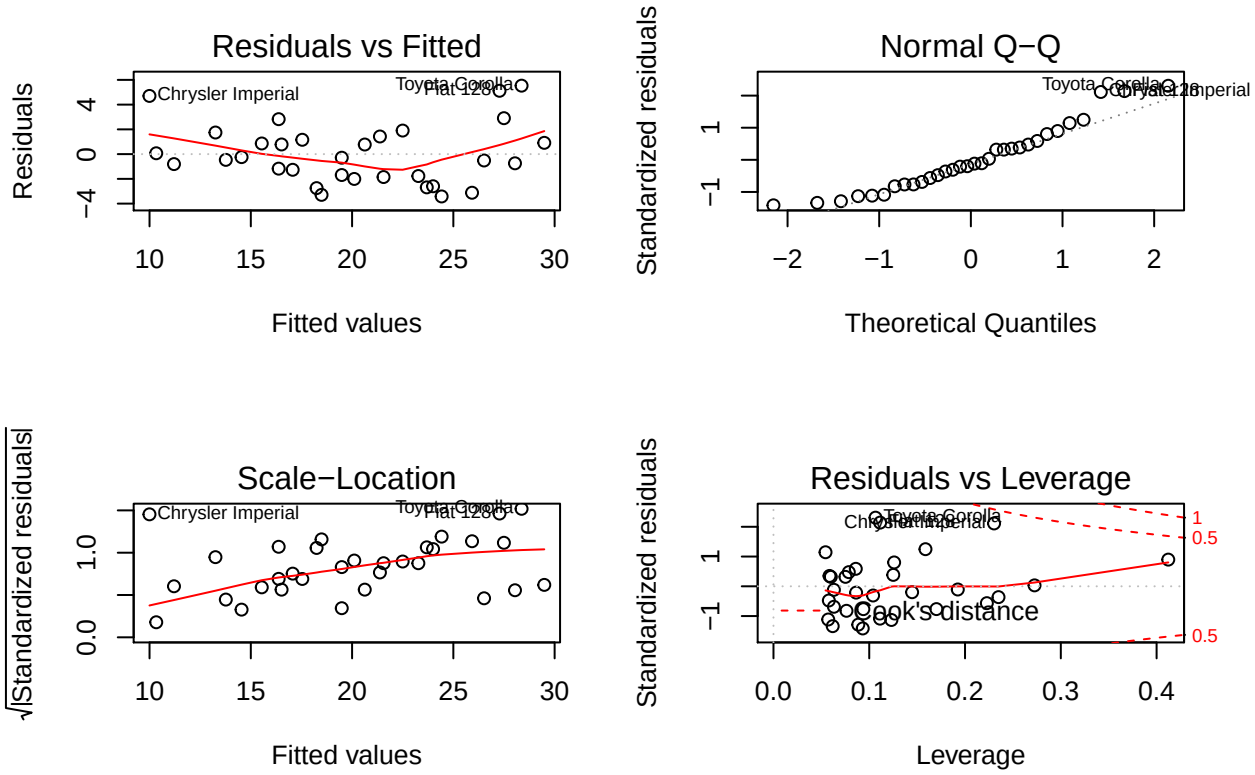
```
model2 = lm(mpg ~ am + wt + hp, data=mtcars)
summary(model2)
```

```
##
## Call:
## lm(formula = mpg ~ am + wt + hp, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4221 -1.7924 -0.3788  1.2249  5.5317
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 34.002875   2.642659  12.867 2.82e-13 ***
## am           2.083710   1.376420   1.514 0.141268
## wt          -2.878575   0.904971  -3.181 0.003574 **
## hp          -0.037479   0.009605  -3.902 0.000546 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.538 on 28 degrees of freedom
## Multiple R-squared:  0.8399, Adjusted R-squared:  0.8227
## F-statistic: 48.96 on 3 and 28 DF,  p-value: 2.908e-11
```

```
anova(model1, model2)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb +
##     car
## Model 2: mpg ~ am + wt + hp
##   Res.Df    RSS  Df Sum of Sq      F Pr(>F)
## 1     15 120.40
## 2     28 180.29 -13   -59.888 0.5739 0.8394
```

Model3 is not signficantly better than the first model however, it's variables in determining the mpg are individually much more significant to determing the mpg.



In addition, the residuals appear normally distributed on the Q-Q graph, and there are no obvious patterns on the residuals vs fitted on the distribution of residuals.

##Summary

The transmission of the car certainly plays a role in the mpg of the car. A 95% T test of the transmission of the cars to mpg shows that they are not the same and is significantly significant. In addition, a model for mpg using the transmission, weight and horsepower for the dependent variables can compensate for about 84% of the variation. It should be noted that it is an extremely small sample size and using data from 1974 and is in no way represenative of cars of today.