regression model project

Executive Summary

In this report, we will examine the mtcars data set and explore how miles per gallon (MPG) is affected by different variables. In particularly, we will answer the following two questions: (1) Is an automatic or manual transmission better for MPG, and (2) Quantify the MPG difference between automatic and manual transmissions.

From our analysis we can show that manual transmission has an MPG 2.8 greater than an automatic transmission.

```
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --
```

```
## v ggplot2 3.3.2     v purrr   0.3.4
## v tibble  3.0.4     v dplyr   1.0.2
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.0
```

```
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
data("mtcars")
head(mtcars)
```

```
##                    mpg cyl disp  hp drat    wt  qsec vs am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  0  1    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  0  1    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  1  1    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  1  0    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  0  0    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  1  0    3    1
```

Motor Trend Car Road Tests

Description

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 (1973–74 models).

A data frame with 32 observations on 11 (numeric) variables.

[, 1] mpg Miles/(US) gallon [, 2] cyl Number of cylinders [, 3] disp Displacement (cu.in.) [, 4] hp Gross horsepower [, 5] drat Rear axle ratio [, 6] wt Weight (1000 lbs) [, 7] qsec 1/4 mile time [, 8] vs Engine (0 = V-shaped, 1 = straight) [, 9] am Transmission (0 = automatic, 1 = manual) [,10] gear Number of forward gears [,11] carb Number of carburetors

EDA

```
## possibly more meaningful, e.g., for summary() or bivariate plots:
mtcars2 <- within(mtcars, {
   vs <- factor(vs, labels = c("V", "S"))
   am <- factor(am, labels = c("automatic", "manual"))
   cyl  <- ordered(cyl)
   gear <- ordered(gear)
   carb <- ordered(carb)
})
summary(mtcars2)
```

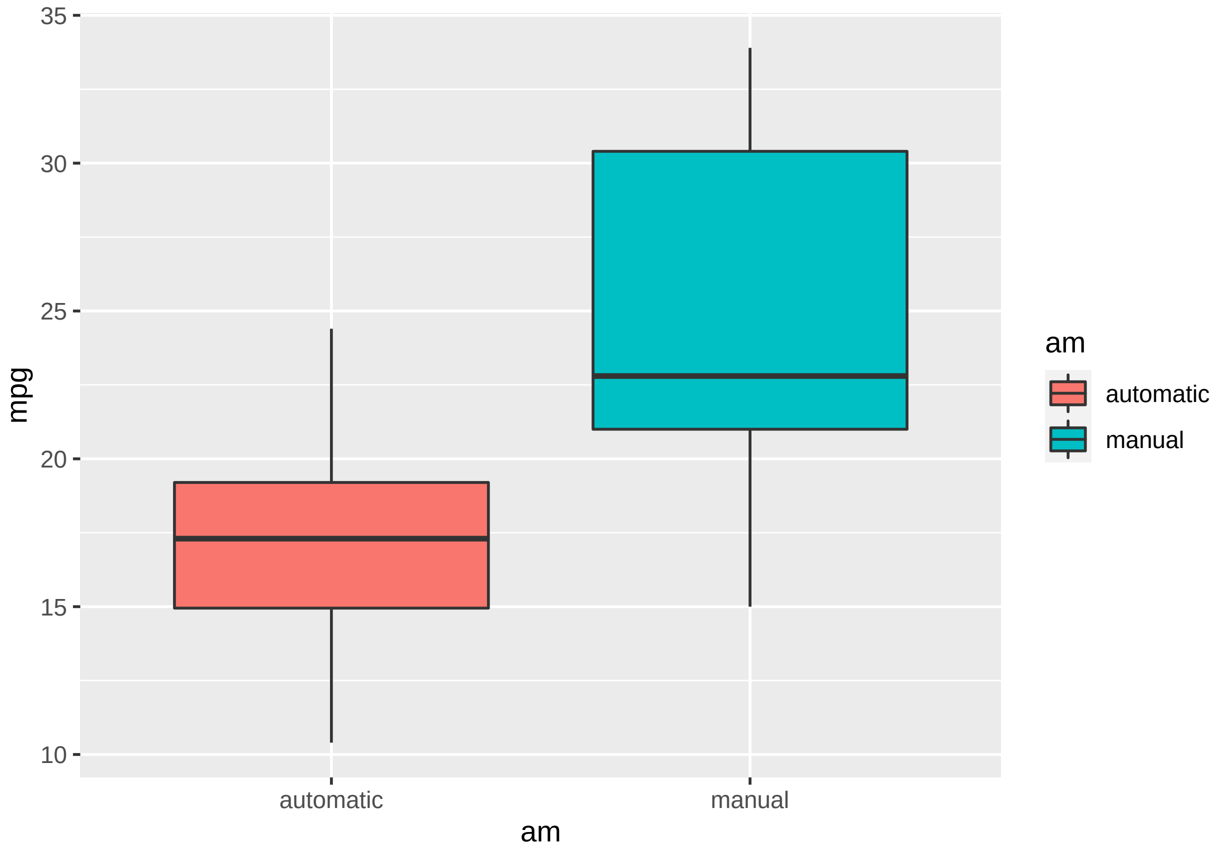
```
##       mpg        cyl         disp             hp             drat      
##  Min.   :10.40   4:11   Min.   : 71.1   Min.   : 52.0   Min.   :2.760  
##  1st Qu.:15.43   6: 7   1st Qu.:120.8   1st Qu.: 96.5   1st Qu.:3.080  
##  Median :19.20   8:14   Median :196.3   Median :123.0   Median :3.695  
##  Mean   :20.09          Mean   :230.7   Mean   :146.7   Mean   :3.597  
##  3rd Qu.:22.80          3rd Qu.:326.0   3rd Qu.:180.0   3rd Qu.:3.920  
##  Max.   :33.90          Max.   :472.0   Max.   :335.0   Max.   :4.930  
##        wt             qsec       vs             am     gear   carb  
##  Min.   :1.513   Min.   :14.50   V:18   automatic:19   3:15   1: 7  
##  1st Qu.:2.581   1st Qu.:16.89   S:14   manual   :13   4:12   2:10  
##  Median :3.325   Median :17.71                         5: 5   3: 3  
##  Mean   :3.217   Mean   :17.85                                4:10  
##  3rd Qu.:3.610   3rd Qu.:18.90                                6: 1  
##  Max.   :5.424   Max.   :22.90                                8: 1  
```

```
head(mtcars2)
```

```
##                    mpg cyl disp  hp drat    wt  qsec vs        am gear carb
## Mazda RX4         21.0   6  160 110 3.90 2.620 16.46  V    manual    4    4
## Mazda RX4 Wag     21.0   6  160 110 3.90 2.875 17.02  V    manual    4    4
## Datsun 710        22.8   4  108  93 3.85 2.320 18.61  S    manual    4    1
## Hornet 4 Drive    21.4   6  258 110 3.08 3.215 19.44  S automatic    3    1
## Hornet Sportabout 18.7   8  360 175 3.15 3.440 17.02  V automatic    3    2
## Valiant           18.1   6  225 105 2.76 3.460 20.22  S automatic    3    1
```

"Is an automatic or manual transmission better for MPG"

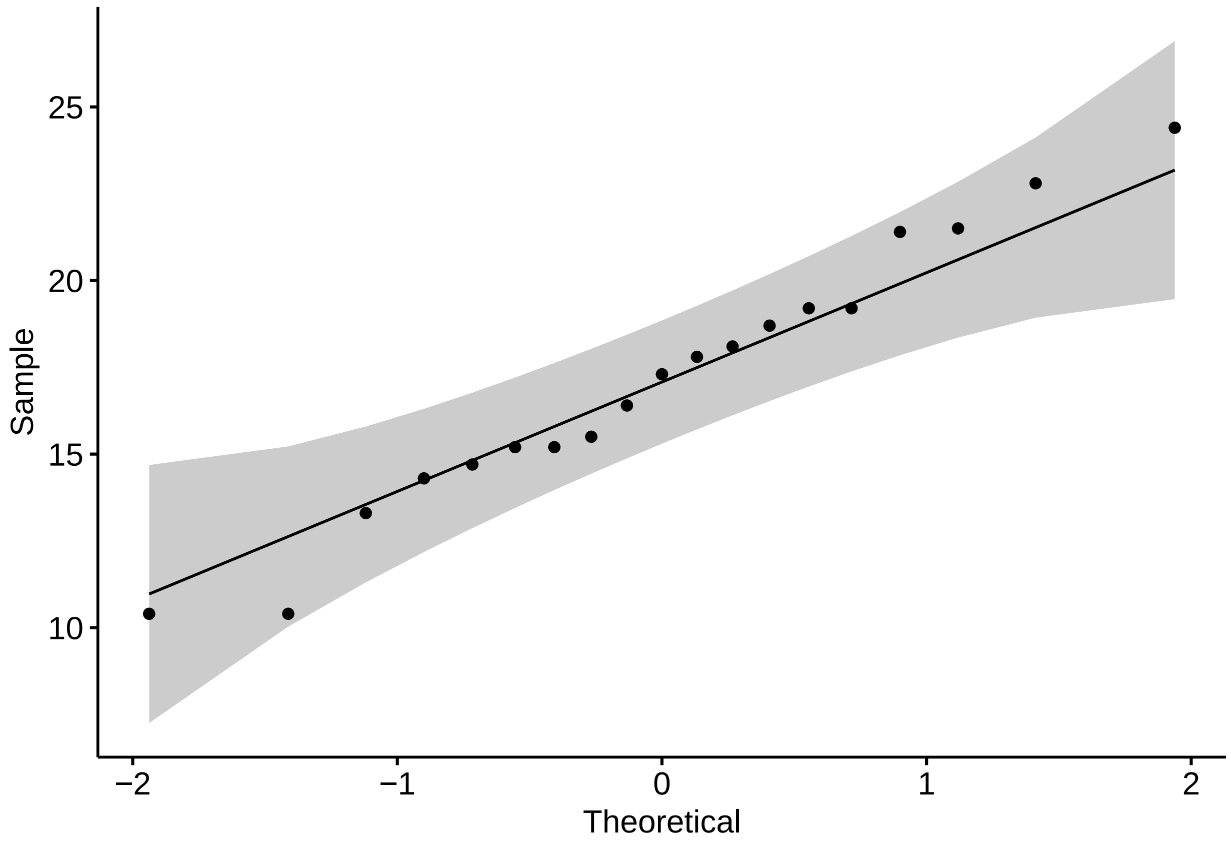
```
mtcars2 %>%
  ggplot(aes(am,mpg,fill=am))+
  geom_boxplot()
```



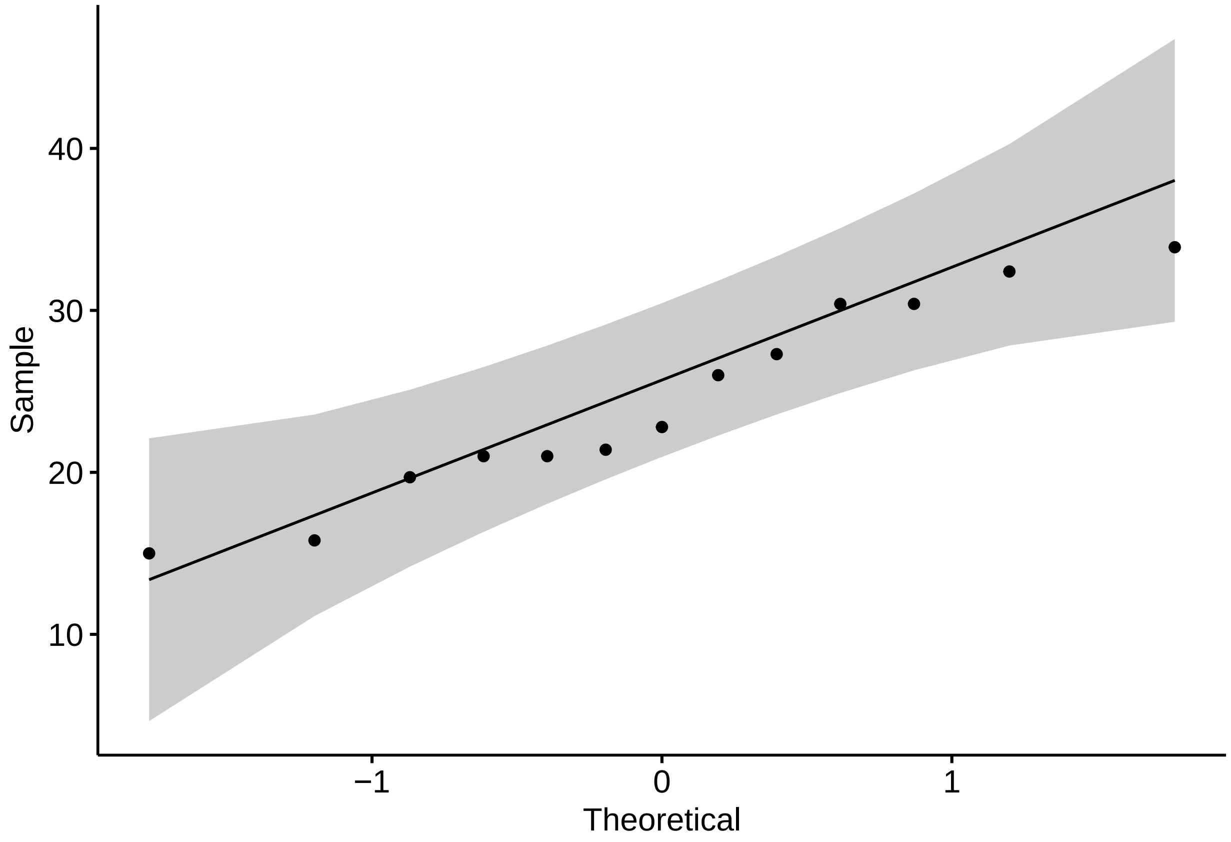
```
mpg_auto <- mtcars2$mpg[mtcars2$am == "automatic"]
mpg_manual <- mtcars2$mpg[mtcars2$am == "manual"]
```

Check assumtion: Assumtion : Are the data from each of the 2 groups follow a normal distribution?

```
library(ggpubr)
ggqqplot(mpg_auto)
```



```
ggqqplot(mpg_manual)
```



```r
shapiro.test(mpg_auto)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  mpg_auto
## W = 0.97677, p-value = 0.8987
```

```r
shapiro.test(mpg_manual)
```

```
##
##  Shapiro-Wilk normality test
##
## data:  mpg_manual
## W = 0.9458, p-value = 0.5363
```

p value > 0.05 so we cannot reject hypothesis that all data is normal distribution

Assumption Do the two populations have the same variances?

```r
res.ftest <- var.test(mpg_auto,mpg_manual)
res.ftest
```

```
##
```

```
##  F test to compare two variances
##
## data:  mpg_auto and mpg_manual
## F = 0.38656, num df = 18, denom df = 12, p-value = 0.06691
## alternative hypothesis: true ratio of variances is not equal to 1
## 95 percent confidence interval:
##  0.1243721 1.0703429
## sample estimates:
## ratio of variances
##          0.3865615
```

The p-value of F-test is greater than the significance level alpha = 0.05. In conclusion, there is no significant difference between the variances of the two sets of data. Therefore, we can use the classic t-test witch assume equality of the two variances.

```
t.test(mpg_auto, mpg_manual, alternative = "less", paired = FALSE, var.equal = TRUE, conf.level = 0.95)
```

```
##
##  Two Sample t-test
##
## data:  mpg_auto and mpg_manual
## t = -4.1061, df = 30, p-value = 0.0001425
## alternative hypothesis: true difference in means is less than 0
## 95 percent confidence interval:
##       -Inf -4.250255
## sample estimates:
## mean of x mean of y
##  17.14737  24.39231
```

As the p-value is lower than 0.05, then, we reject the null hypothesis. it means that mpg automatic less than mpg manual

"Quantify the MPG difference between automatic and manual transmissions"

```
simple_model <- lm(mpg ~ am, data = mtcars)
summary(simple_model)
```

```
##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -9.3923 -3.0923 -0.2974  3.2439  9.5077
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   17.147      1.125  15.247 1.13e-15 ***
## am             7.245      1.764   4.106 0.000285 ***
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## Residual standard error: 4.902 on 30 degrees of freedom
## Multiple R-squared:  0.3598, Adjusted R-squared:  0.3385
## F-statistic: 16.86 on 1 and 30 DF,  p-value: 0.000285
```

The R2 value is 0.36 thus telling us this model only explains us 36% of the variance. As a result, we need to build a multivariate linear regression.

```
complex_model <- lm(mpg ~.,data=mtcars)
anova(simple_model,complex_model)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ cyl + disp + hp + drat + wt + qsec + vs + am + gear + carb
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
## 1     30 720.90
## 2     21 147.49  9     573.4 9.0711 1.779e-05 ***
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
```

p-value < 0.05 so we can claim the complex_model model is significantly better than our simple_model model.

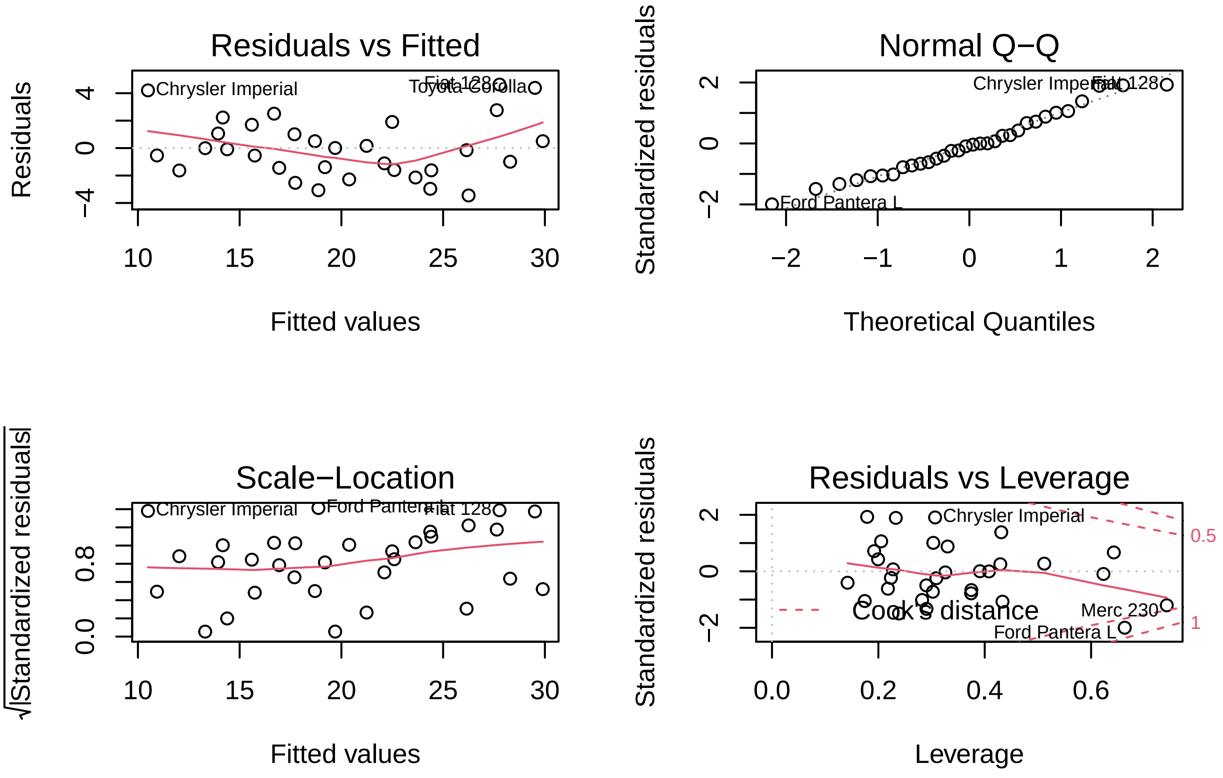
```
summary(complex_model)
```

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4506 -1.6044 -0.1196  1.2193  4.6271
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 12.30337   18.71788   0.657   0.5181
## cyl         -0.11144    1.04502  -0.107   0.9161
## disp         0.01334    0.01786   0.747   0.4635
## hp          -0.02148    0.02177  -0.987   0.3350
## drat         0.78711    1.63537   0.481   0.6353
## wt          -3.71530    1.89441  -1.961   0.0633 .
## qsec         0.82104    0.73084   1.123   0.2739
## vs           0.31776    2.10451   0.151   0.8814
## am           2.52023    2.05665   1.225   0.2340
## gear         0.65541    1.49326   0.439   0.6652
## carb        -0.19942    0.82875  -0.241   0.8122
## ---
## Signif. codes:  0 ’***’ 0.001 ’**’ 0.01 ’*’ 0.05 ’.’ 0.1 ’ ’ 1
##
## Residual standard error: 2.65 on 21 degrees of freedom
## Multiple R-squared:  0.869,  Adjusted R-squared:  0.8066
## F-statistic: 13.93 on 10 and 21 DF,  p-value: 3.793e-07
```

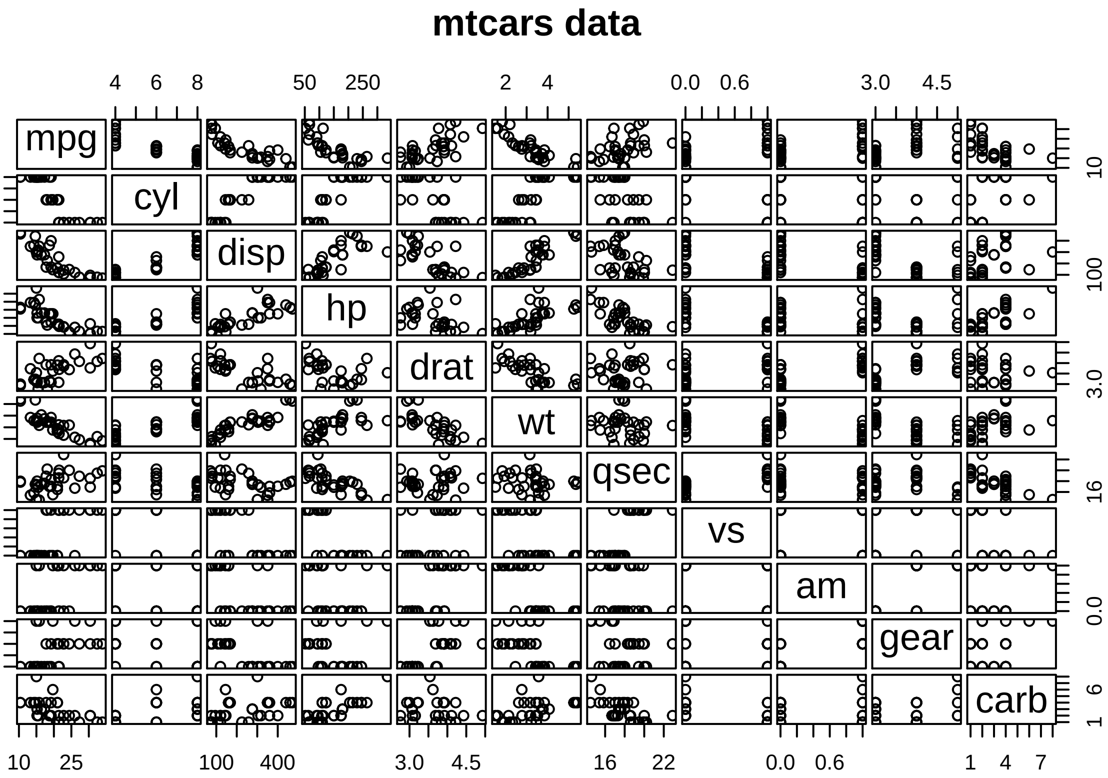
The model explains 86.9% of the variance and as a result, Thus, we can say the difference between automatic and manual transmissions is 2.81 MPG. ## Appendix

- Check residuals

```r
par(mfrow = c(2,2))
plot(complex_model)
```



```r
require(graphics)
pairs(mtcars, main = "mtcars data", gap = 1/4)
```


mtcars data
mpg
cyl
disp
hp
drat
wt
qsec
vs
am
gear
carb
4
6
8
50
250
2
4
0.0
0.6
3.0
4.5
10
25
100
400
3.0
4.5
16
22
0.0
0.6
1
4
7
7
4
50
5
2
0.0
3.0
10
100
3.0
16
0.0
6
1