Regression Week 4 Project: Motor Trends : Automatic or Manual transmission?

You work for Motor Trend, a magazine about the automobile industry. Looking at a data set of a collection of cars, they are interested in exploring the relationship between a set of variables and miles per gallon (MPG) (outcome). They are particularly interested in the following two questions:

"Is an automatic or manual transmission better for MPG" "Quantify the MPG difference between automatic and manual transmissions"

Peer Grading

The criteria that your classmates will use to evaluate and grade your work are shown below. Each criteria is binary: (1 point = criteria met acceptably; 0 points = criteria not met acceptably) Criteria

Did the student interpret the coefficients correctly? Did the student do some exploratory data analyses? Did the student fit multiple models and detail their strategy for model selection? Did the student answer the questions of interest or detail why the question(s) is (are) not answerable? Did the student do a residual plot and some diagnostics? Did the student quantify the uncertainty in their conclusions and/or perform an inference correctly? Was the report brief (about 2 pages long) for the main body of the report and no longer than 5 with supporting appendix of figures? Did the report include an executive summary? Was the report done in Rmd (knitr)?

Exploring the Dataset

```
library(ggplot2)
library(dplyr)
data(mtcars)
```

```
# Checking for the dimension of the dataset
dim(mtcars)
```

```
## [1] 32 11
```

```
# Checking for columns names
names(mtcars)
```

```
##  [1] "mpg"  "cyl"  "disp" "hp"   "drat" "wt"   "qsec" "vs"   "am"   "gear"
## [11] "carb"
```

```
# viewing the first six rows of the dataset
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
```

```
# checking the structure of the dataset
str(mtcars)
```

```
## 'data.frame':    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
##  $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
##  $ qsec: num  16.5 17 18.6 19.4 17 ...
##  $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
##  $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
##  $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
##  $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

```
# checking the summary of the dataset
summary(mtcars)
```

```
##       mpg             cyl             disp             hp       
##  Min.   :10.40   Min.   :4.000   Min.   : 71.1   Min.   : 52.0  
##  1st Qu.:15.43   1st Qu.:4.000   1st Qu.:120.8   1st Qu.: 96.5  
##  Median :19.20   Median :6.000   Median :196.3   Median :123.0  
##  Mean   :20.09   Mean   :6.188   Mean   :230.7   Mean   :146.7  
##  3rd Qu.:22.80   3rd Qu.:8.000   3rd Qu.:326.0   3rd Qu.:180.0  
##  Max.   :33.90   Max.   :8.000   Max.   :472.0   Max.   :335.0  
##       drat             wt             qsec             vs        
##  Min.   :2.760   Min.   :1.513   Min.   :14.50   Min.   :0.0000  
##  1st Qu.:3.080   1st Qu.:2.581   1st Qu.:16.89   1st Qu.:0.0000  
##  Median :3.695   Median :3.325   Median :17.71   Median :0.0000  
##  Mean   :3.597   Mean   :3.217   Mean   :17.85   Mean   :0.4375  
##  3rd Qu.:3.920   3rd Qu.:3.610   3rd Qu.:18.90   3rd Qu.:1.0000  
##  Max.   :4.930   Max.   :5.424   Max.   :22.90   Max.   :1.0000  
##        am              gear            carb      
##  Min.   :0.0000   Min.   :3.000   Min.   :1.000  
##  1st Qu.:0.0000   1st Qu.:3.000   1st Qu.:2.000  
##  Median :0.0000   Median :4.000   Median :2.000  
##  Mean   :0.4062   Mean   :3.688   Mean   :2.812  
##  3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000  
##  Max.   :1.0000   Max.   :5.000   Max.   :8.000
```

```
# checking the correlation between mpg and the other variable
cor(mtcars$mpg, mtcars[,-1])
```

```
##             cyl       disp         hp      drat         wt     qsec        vs
## [1,] -0.852162 -0.8475514 -0.7761684 0.6811719 -0.8676594 0.418684 0.6640389
##             am      gear       carb
## [1,] 0.5998324 0.4802848 -0.5509251
```

The result of the correlation provides a negative correlation to mpg and the variables are: *cyl, disp, hp, wt, carb*

```
# Looking at the t.test at the confident level of 95 and the p value
mtcarsTestResults <- t.test(mpg ~ am, data = mtcars, conf.level = 0.95)
mtcarsTestResults$p.value
```

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

If the null hypothesis is presented as that a car with a manual transmission will have more mpg than a car with an automatic transmission. However, the p value been 0.001374 may reject the null hypothesis. More over, as the cars do have different characteristics a deeper analysis must be pursue.

```
# Checking the estimate of t.test results
mtcarsTestResults$estimate
```

```
## mean in group 0 mean in group 1
##        17.14737        24.39231
```

The result from the estimate shows that the manual transmission has an advantage of 7.24494 extra miles per gallon

```


# Checking the regression values of the univariable mtcarsfitUv
mtcarsfitUv <- lm(mpg ~ am, data = mtcars)
summary(mtcarsfitUv)
```

Regression Analysis

```
##
## 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
```

```r
# Rounding the univariable mtcarsfitUv to two digits
round(summary(mtcarsfitUv)$r.square*100, digits=2)
```

```
## [1] 35.98
```

The result of the regression can only explain that 35.98% of the mpg variation

```r
# Checking the regression values of the multivariable mtcarsfitMv
mtcarsfitMv <- lm(mpg ~ ., data = mtcars)
mtcarsfitMv
```

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Coefficients:
## (Intercept)          cyl         disp           hp         drat           wt
##    12.30337     -0.11144      0.01334     -0.02148      0.78711     -3.71530
##        qsec           vs           am         gear         carb
##     0.82104      0.31776      2.52023      0.65541     -0.19942
```

```r
summary(mtcarsfitMv)
```

```
##
## 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
```

```r
# Rounding the multivariable mtcarsfitMv to two digits
round(summary(mtcarsfitMv)$r.square*100, digits=2)
```

```
## [1] 86.9
```

Summary We use the mtcars dataset to determine which type of transmission would offer better mileage. The t-test between the automatic and manual transmission shows us that the maul transmission does have a 7.24494 extra mileage per gallon. However, the regression model on a multivariable shows that the difference between automatic and manual transmission is much less

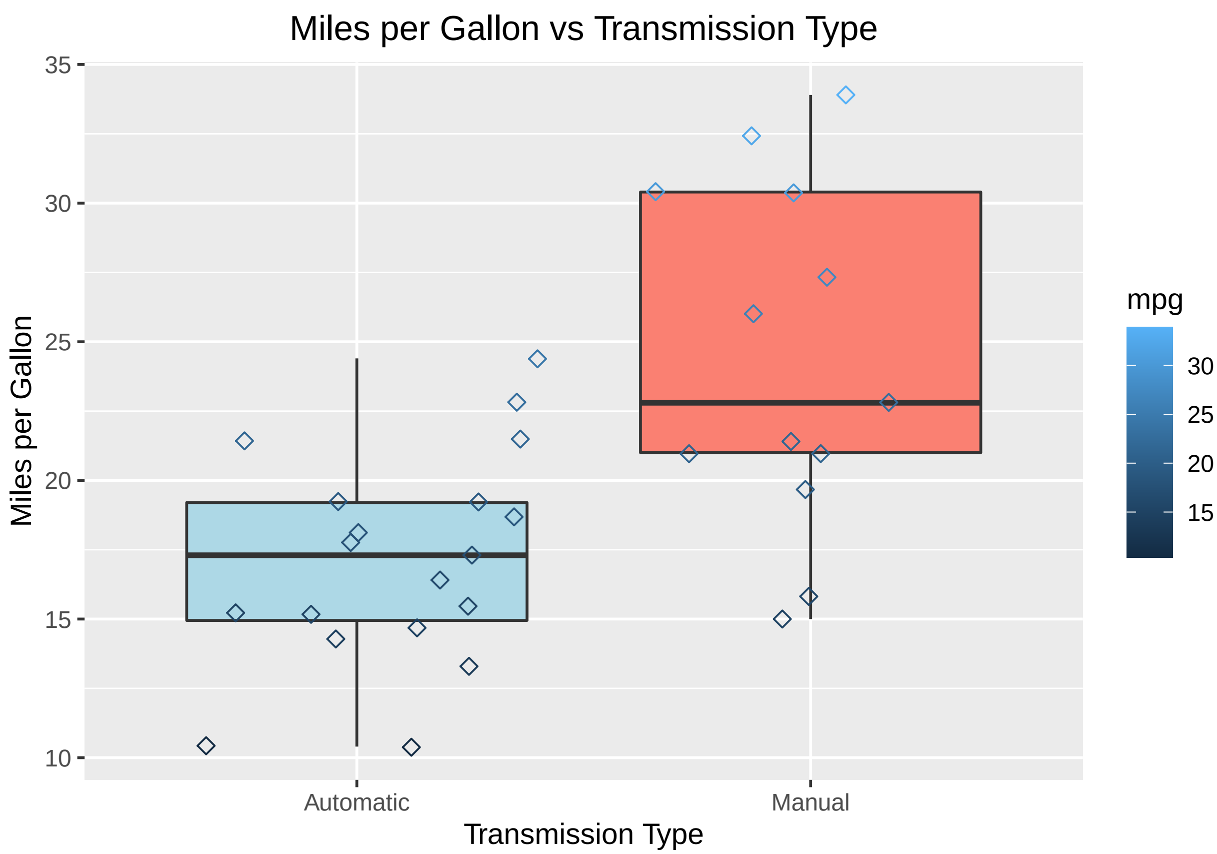
Conclusion Although the manual transmission would have a small advantage than an automatic transmission; there are other variable like cyl, wt, hp which have more more influence when determine

Residuals Analysis and Diagnostics

- The residual vs fitted plots shows not consistent pattern
- The Normal Q-Q plot indicates that the residuals are normal
- The scale-location plot confirms the constant variance assumption
- The Residuals vs leverage confirms that no outliers are present

Appendix

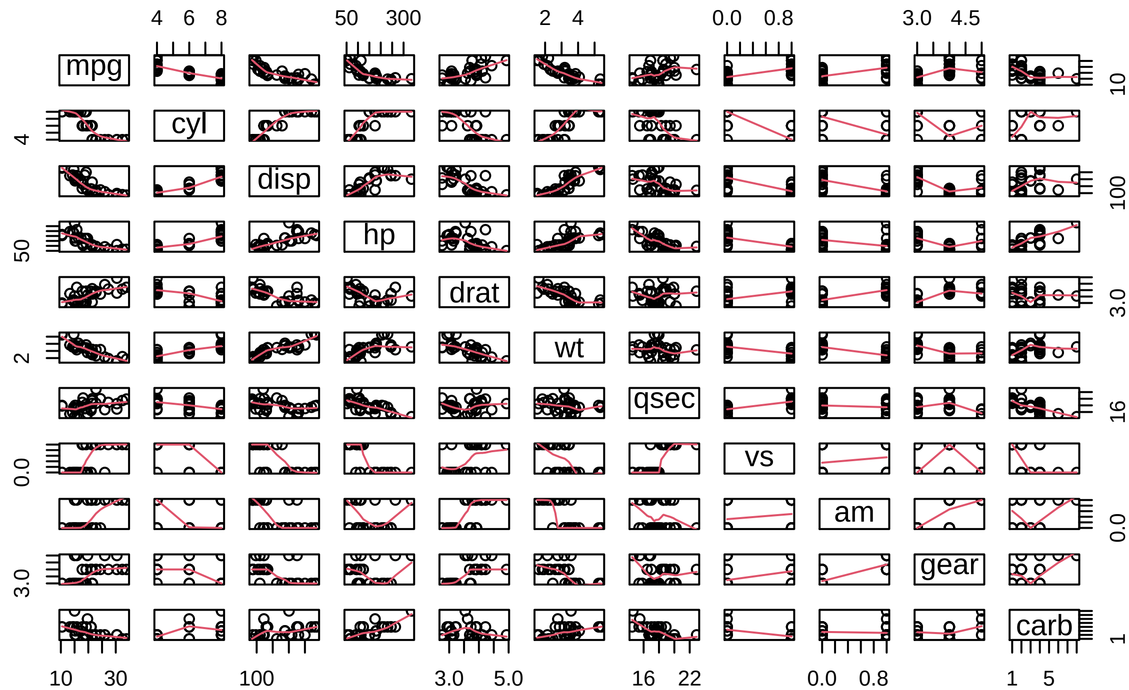
```r
# Creating a new variable as a factor and changing the levels
mtcars$am2 <- factor(mtcars$am)
levels(mtcars$am2) <-c("Automatic", "Manual")

# Using ggplot2 to create a boxplot to look at mpg and transmission type distribution
g <- ggplot(mtcars,aes(x=am2, y=mpg)) +
  geom_boxplot(aes(group=am2), fill=c("lightblue","salmon")) +
  labs(x="Transmission Type") +
  ylab("Miles per Gallon") +
  ggtitle("Miles per Gallon vs Transmission Type") +
  theme(plot.title = element_text(hjust = 0.5)) +
  geom_jitter(cex = 1.8, shape = 5, aes(colour = mpg))
g
```



1.

```
# pairs all the variables without the newly create variable am2
pairs((mtcars %>% select(-am2)), panel = panel.smooth, main = "Pair Graph of mtcars")
```

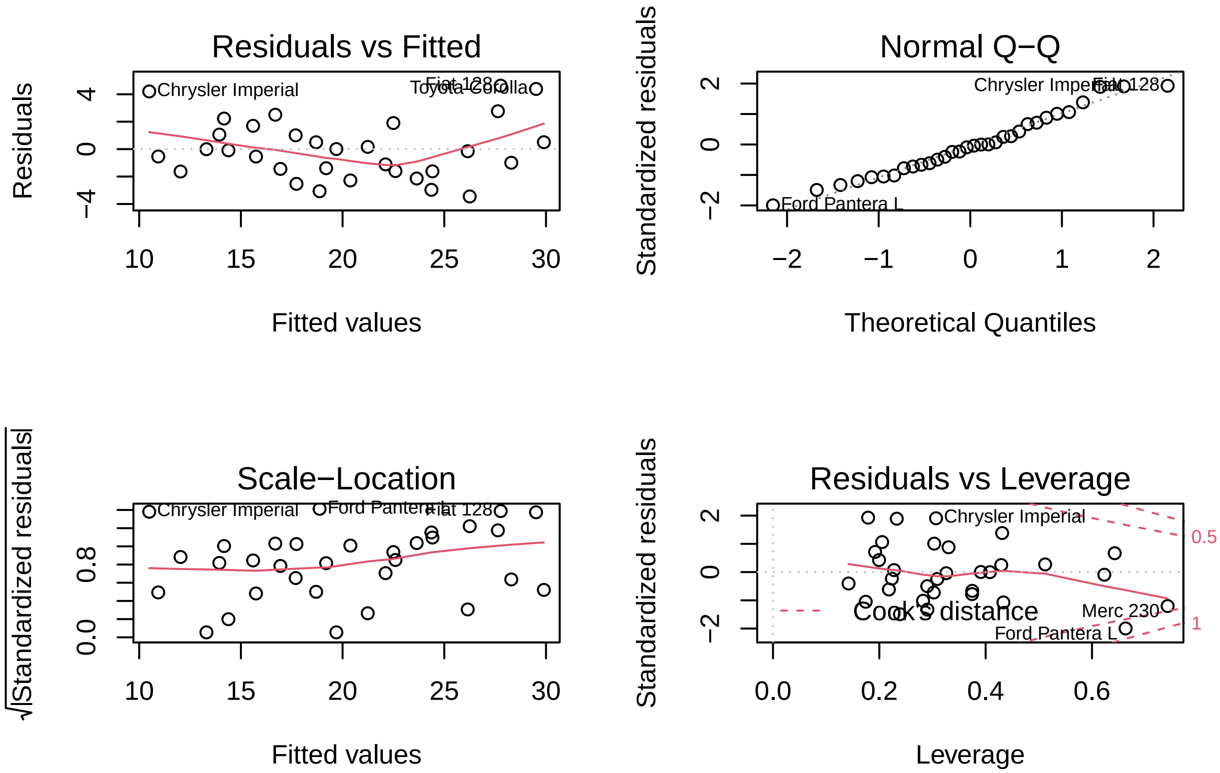
Pair Graph of mtcars



2.

The Pair graph shows that several variables are high correlated with mpg.

```
# Multivariable
par(mfrow=c(2, 2))
plot(mtcarsfitMv)
```


Residuals vs Fitted
Residuals
Fitted values
Chrysler Imperial
Fiat 128
Toyota Corolla
Normal Q−Q
Standardized residuals
Theoretical Quantiles
Ford Pantera L
Scale−Location
√|Standardized residuals|
Residuals vs Leverage
Leverage
Cook's distance
Merc 230
0.5
1

3.