Title : “Regression Models Course project”

author: “Chamathka Thilakarathna”

Date : “17/07/2020”

This is the course project report to the Final project of Regression Model Course.

This report contains some Exploratory analysis and Regression analysis to mtcars dataset.

```
library(ggplot2)
library(dplyr)
```

```
##
## Attaching package: ’dplyr’
```

```
## The following objects are masked from ’package:stats’:
##
##     filter, lag
```

```
## The following objects are masked from ’package:base’:
##
##     intersect, setdiff, setequal, union
```

```
data(mtcars)
str(mtcars)
```

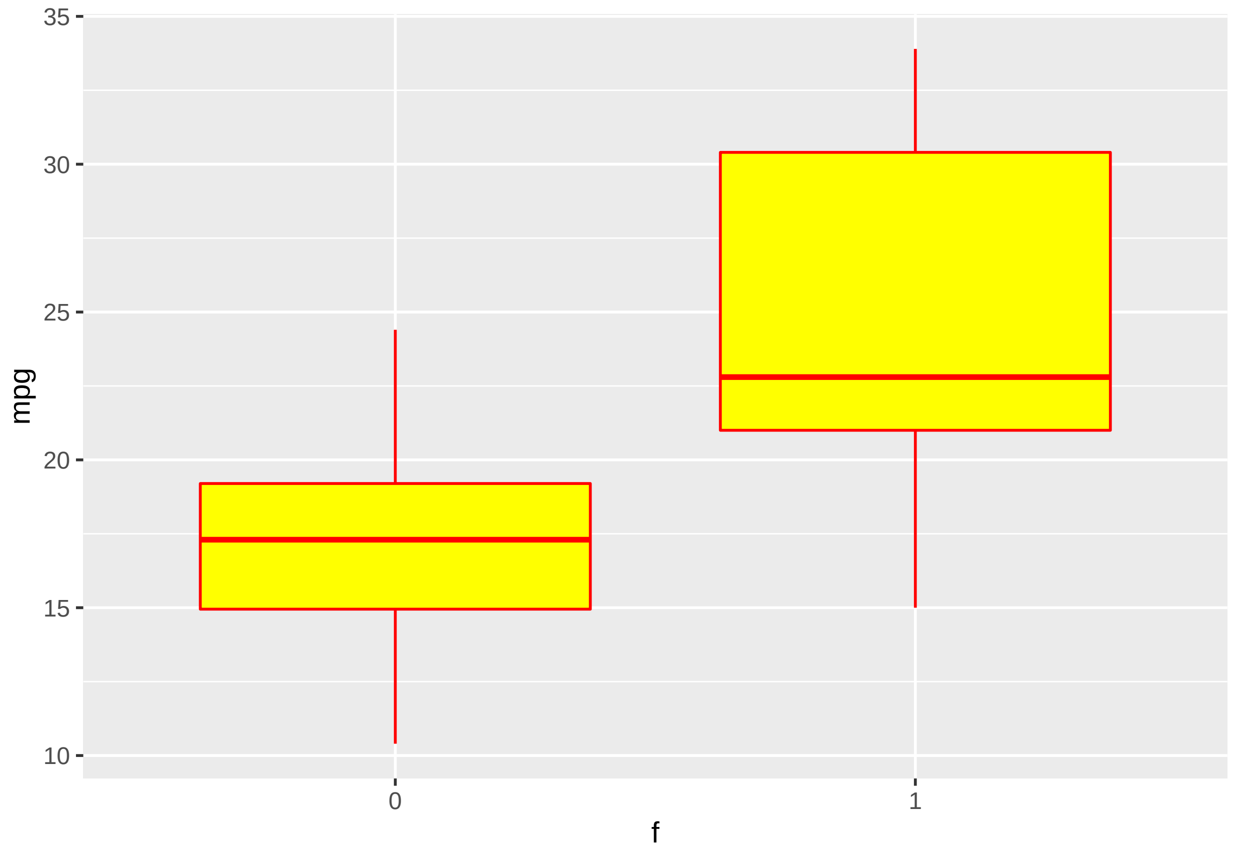
```
## ’data.frame’:    32 obs. of  11 variables:
##  $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
##  $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
##  $ disp: num  160 160 108 258 360 ...
##  $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
##  $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 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 ...
```

```
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
```

Exploratory Data Analysis

```
f=factor(mtcars$am)
ggplot(mtcars,aes(x=f,y=mpg))+geom_boxplot(colour="red",fill="yellow")
```

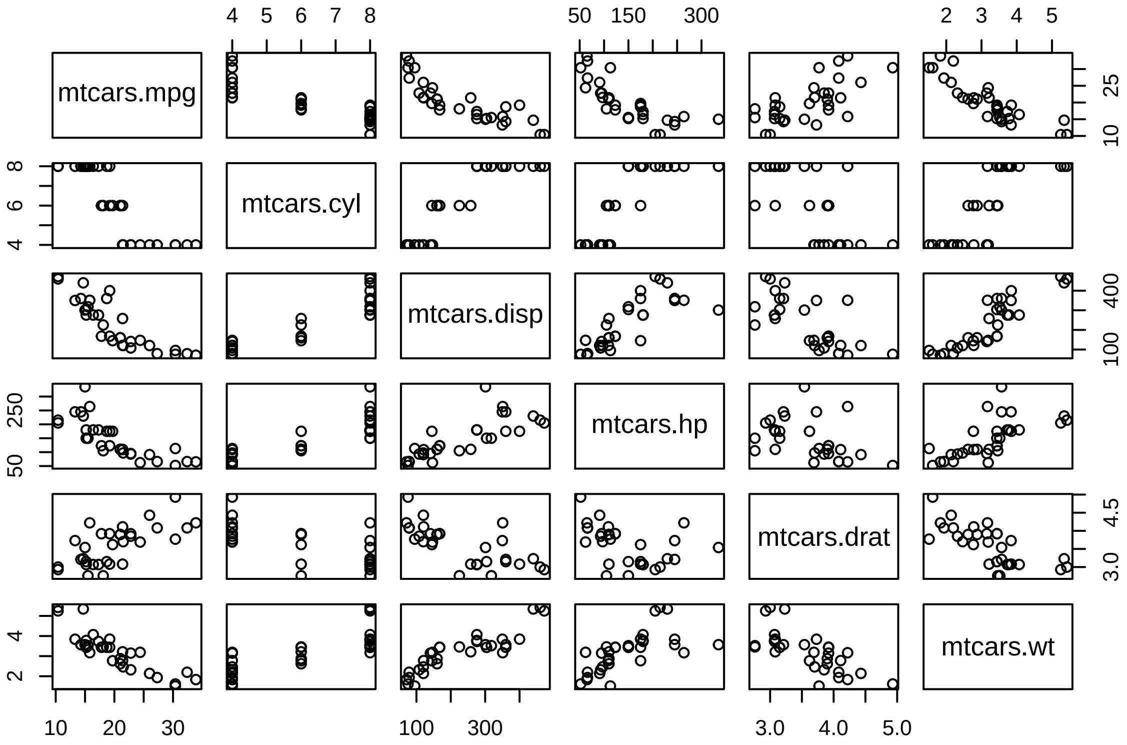


```
###Mean,median and all the quantiles of miles per gallon are higher in manual vehicles.
```

According to the above bloxplot we can see that mean,median and the all the quantile values of manual are higher than the automatic.Therefore miles fer gallons in manual vehicles are high.

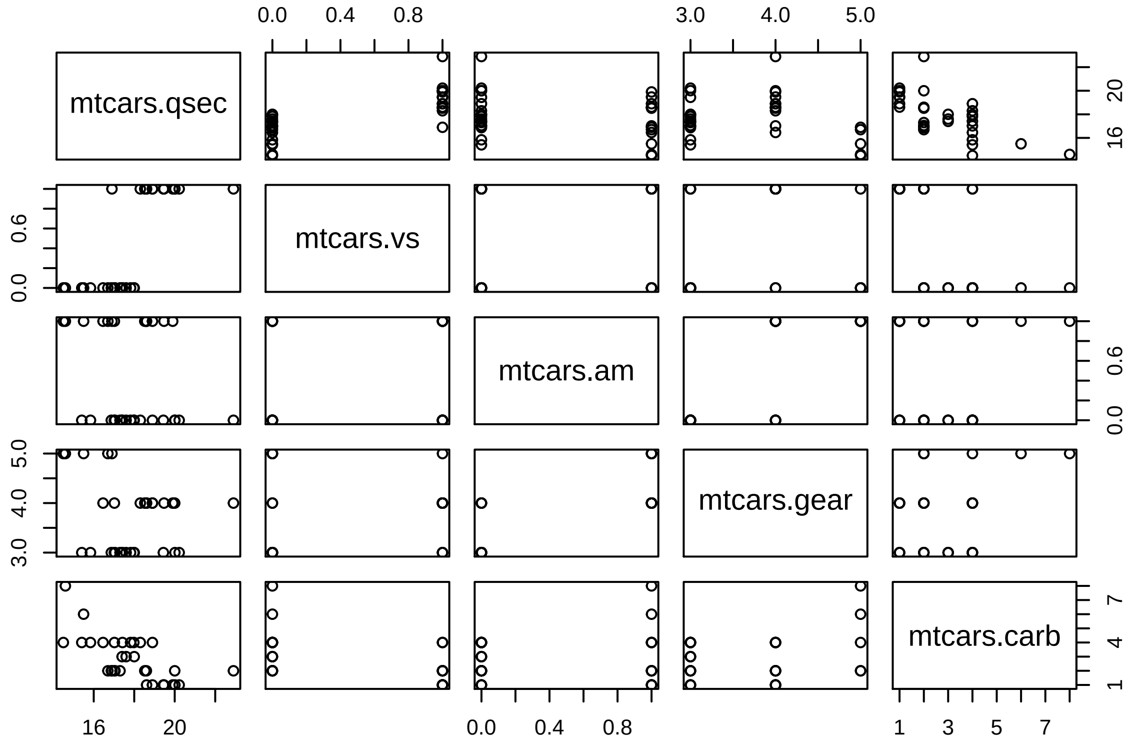
Multiple Regression Analysis.

```
data(mtcars);
data1=data.frame(mtcars$mpg,mtcars$cyl,mtcars$disp,mtcars$hp,mtcars$drat,mtcars$wt)
pairs(data1)
```



Above figure shows us fairwise correlation between variables mpg,cyl,disp, hp,drat,wt.we can see that theese variables are positively or negatively correlated approximately.

```
data2=data.frame(mtcars$qsec,mtcars$vs,mtcars$am,mtcars$gear,mtcars$carb)
pairs(data2)
```



Fairwise correlation between qsec,vs,am,gear,carb.Not strickly correlated as above variables

```
fit_manual=lm(mpg[mtcars$am==1]~disp[mtcars$am==1],data=mtcars)
summary(fit_manual)$coef
```

```
##                          Estimate Std. Error  t value     Pr(>|t|)
## (Intercept)           32.86613705 1.95033212 16.85156 3.328083e-09
## disp[mtcars$am == 1] -0.05903842 0.01173523 -5.03087 3.834293e-04
```

```
fit_automatic=lm(mpg[mtcars$am==0]~disp[mtcars$am==0],data=mtcars)
summary(fit_automatic)$coef
```

```
##                         Estimate  Std. Error  t value    Pr(>|t|)
## (Intercept)           25.1570641 1.592922405 15.79303 1.36335e-11
## disp[mtcars$am == 0] -0.0275836 0.005145991 -5.36021 5.19427e-05
```

***Above regression models coefficients show that Intercept and slope is higher in manual.Thus decrement of miles per gallon per 1 increment of Displacement is higher in manual.

```
mtcars=mutate(mtcars,amn=1*(am==1))
fit<-lm(mpg~cyl*factor(amn),data=mtcars)
summary(fit)$coef
```

```
##                     Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)      30.873529  3.1882316  9.683591 1.947808e-10
## cyl              -1.975735  0.4485295 -4.404917 1.407342e-04
## factor(amn)1     10.175407  4.3045523  2.363871 2.525769e-02
## cyl:factor(amn)1 -1.305116  0.7070400 -1.845887 7.550690e-02
```

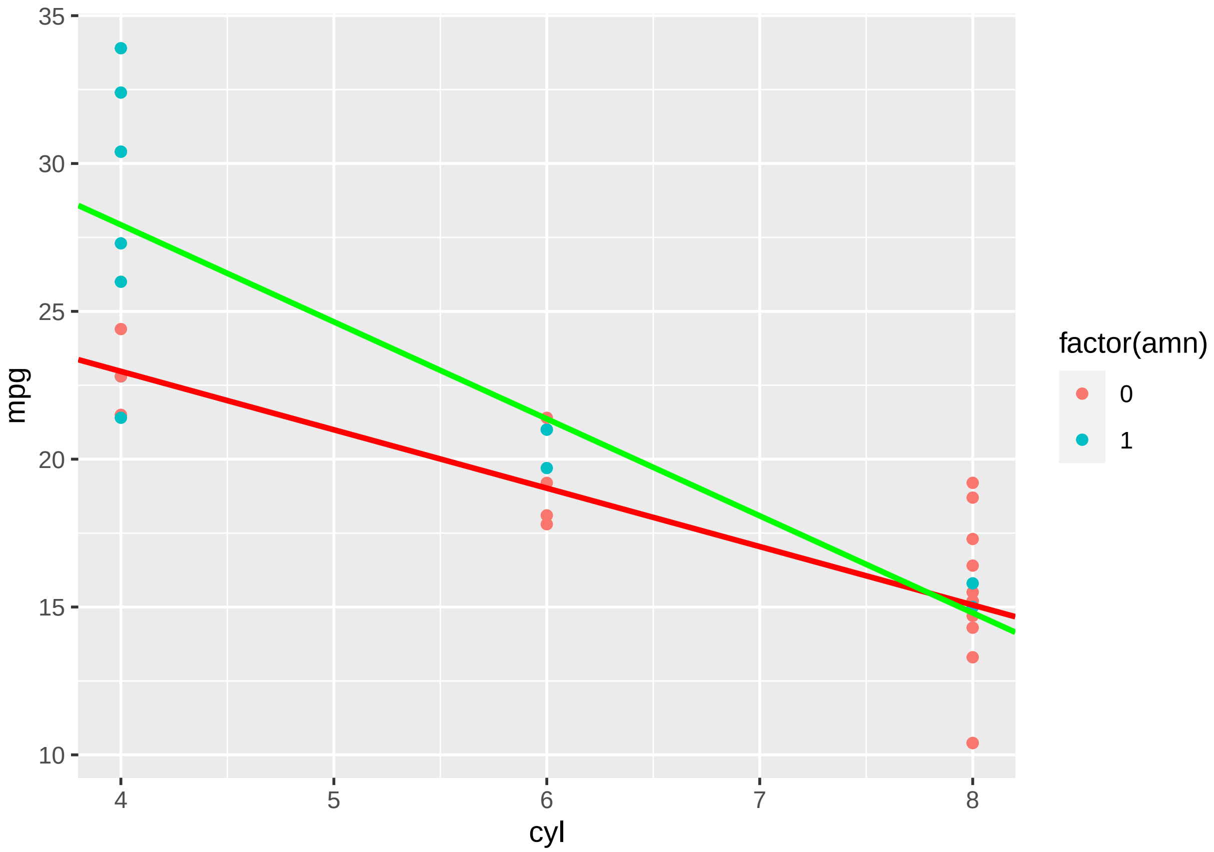
```
g=ggplot(mtcars,aes(x=cyl,y=mpg,colour=factor(amn)))+geom_point()
g=g+geom_abline(intercept = coef(fit)[1],slope=coef(fit)[2],
                method="lm",size=1,colour="red",na.rm = T)
```

```
## Warning: Ignoring unknown parameters: method
```

```
g=g+geom_abline(intercept = coef(fit)[1]+coef(fit)[3],
                slope = coef(fit)[2]+coef(fit)[4],method="lm",size=1,
                colour="green",na.rm = T)
```

```
## Warning: Ignoring unknown parameters: method
```

```
g
```



According to the graph the slope(decreasing) of the miles per galoon(mpg) for manual is greater than to the automatic vehicles corresponding to the number of cylinders. Therefore we can conclude that automatic transmission better for mpg.

Model Selection.

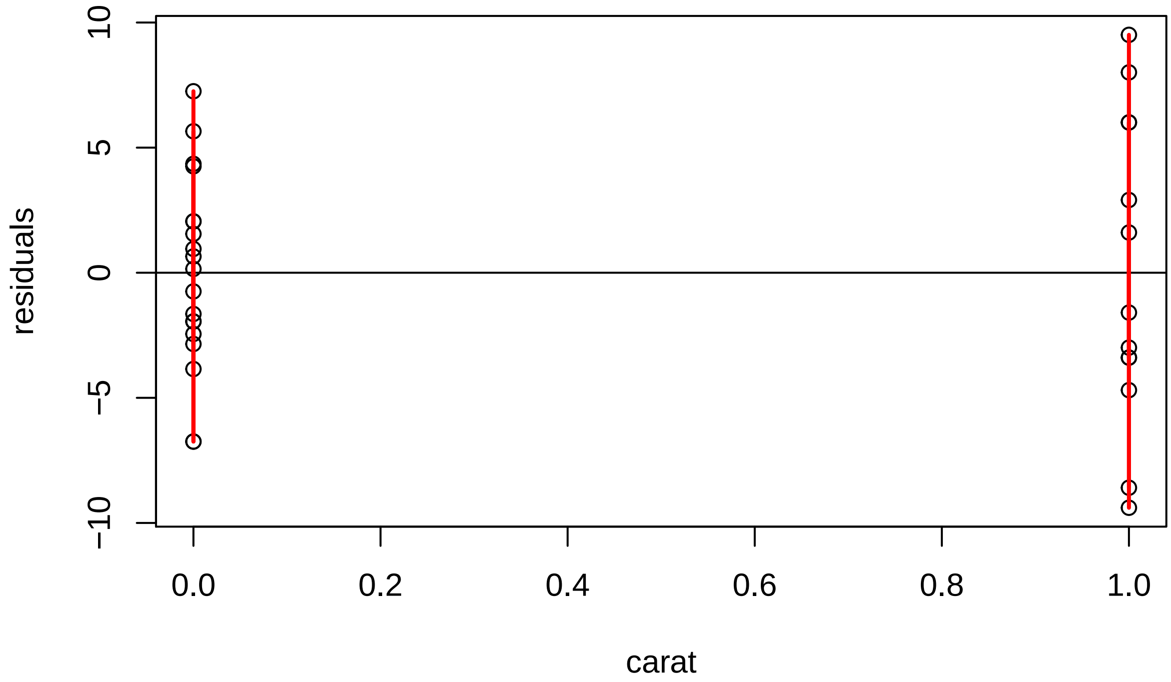
```
fit<-lm(mpg~.,data=mtcars)
fit1<-lm(mpg~cyl,data=mtcars)
fit3<-update(fit,mpg~cyl+disp+hp)
fit5<-update(fit,mpg~cyl+disp+hp+drat+wt)
fit7<-update(fit,mpg~cyl+disp+hp+drat+wt+qsec+vs)
fit9<-update(fit,mpg~cyl+disp+hp+drat+wt+qsec+vs+gear+carb)
anova(fit1,fit3,fit5,fit7,fit9)
```

```
## Analysis of Variance Table
##
## Model 1: mpg ~ cyl
## Model 2: mpg ~ cyl + disp + hp
## Model 3: mpg ~ cyl + disp + hp + drat + wt
## Model 4: mpg ~ cyl + disp + hp + drat + wt + qsec + vs
## Model 5: mpg ~ cyl + disp + hp + drat + wt + qsec + vs + gear + carb
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1     30 308.33
## 2     28 261.37  2    46.965 3.2689 0.057153 .
## 3     26 167.43  2    93.943 6.5387 0.005907 **
## 4     24 163.35  2     4.079 0.2839 0.755549
## 5     22 158.04  2     5.306 0.3693 0.695418
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

According to the anova table of with respect to the above models Rss values are lower in model3,mode4,model5.comparing the F statistics model 4 is better.

Residual Plot.

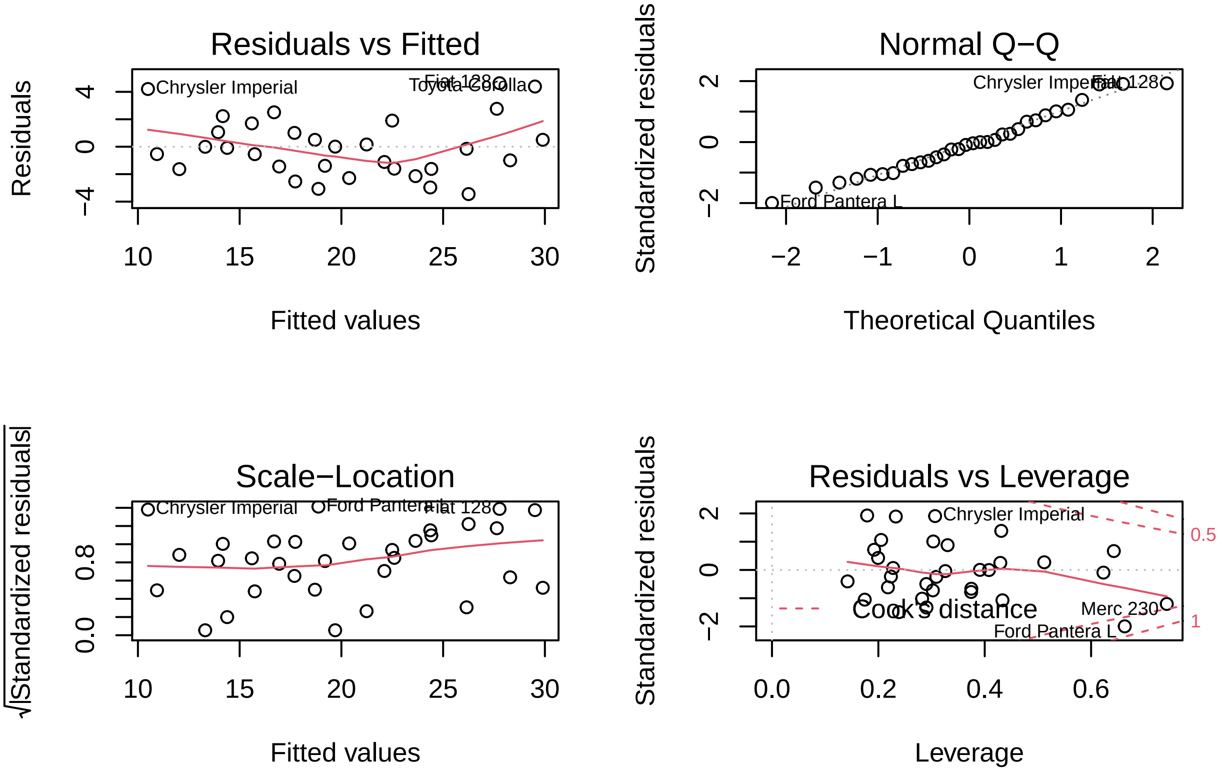
```
y=mtcars$mpg ; x=mtcars$am; n=length(y)
fit=lm(y~x)
e=resid(fit)
plot(x,e,xlab = "carat",ylab = "residuals")
abline(h=0,lwd=1)
for(i in 1:n)
  lines(c(x[i],x[i]),c(e[i],0),col="red",lwd=2)
```



Residual Plot for miles per gallon and am.variation of residuals in automatic is greater than to manual.

Diagonastic Approach.

```
data(mtcars); par(mfrow=c(2,2))
fit<-lm(mpg~.,data = mtcars); plot(fit)
```



***This figure shows some diagonastic test for this model.

Conclusion.

when we compare to the manual and mpg Transmission according to the above figures and model data we can conclude that manual is better than the automatic.