CARS Analysis

This analysis is done to explore a dataset of collection of cars. The main aim is to explore the relationship between a set of variables and miles per galllon. The two questions which needs to be answered are:-

1- Is an automatic or manual transmission better for MPG. 2- Quantify the MPG difference between automatic and manual transmissions.

Executive summary

During this analysis it was found that the transmission type(automatic/ manual) has sugnificant effect on the miles per gallong (mpg) of the vehicle. In addition to transmission, two other variables, weight and qsec has impact on the mpg as well. Keeping other variables constant, Manual transmission is approx. `9.62` mpg more than automatic transmission.

Analysis

```
library(ggplot2)
```

load the data.

```
data(mtcars)
attach(mtcars)
```

```
## The following object is masked from package:ggplot2:
##
##     mpg
```

convert automatic variable to logical type

```
mtcars$am  <- as.logical(mtcars$am)
```

summarise automatic and maual cars in the dataset.

```
summary(mtcars$am)
```

```
##    Mode   FALSE    TRUE    NA's
## logical      19      13       0
```

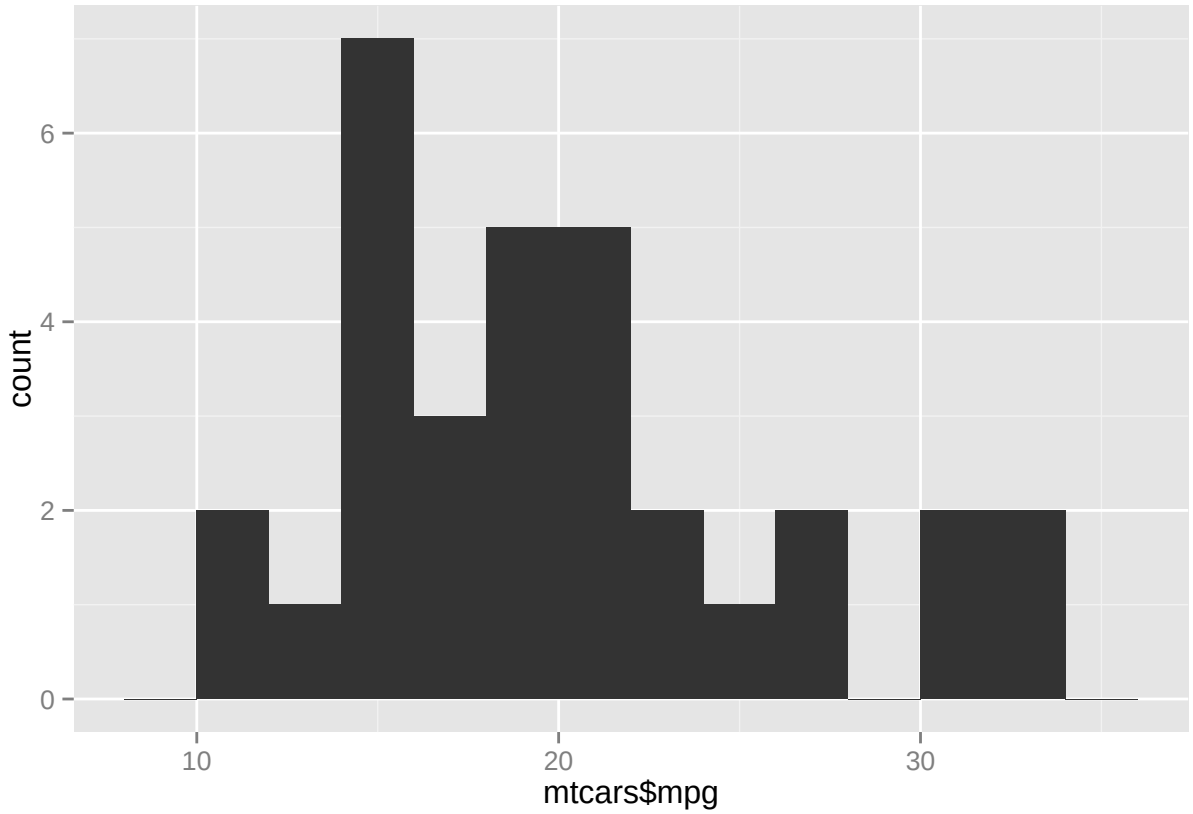
There is no missing value.

check the summary of mpg variable.

```
summary(mtcars$mpg)
```

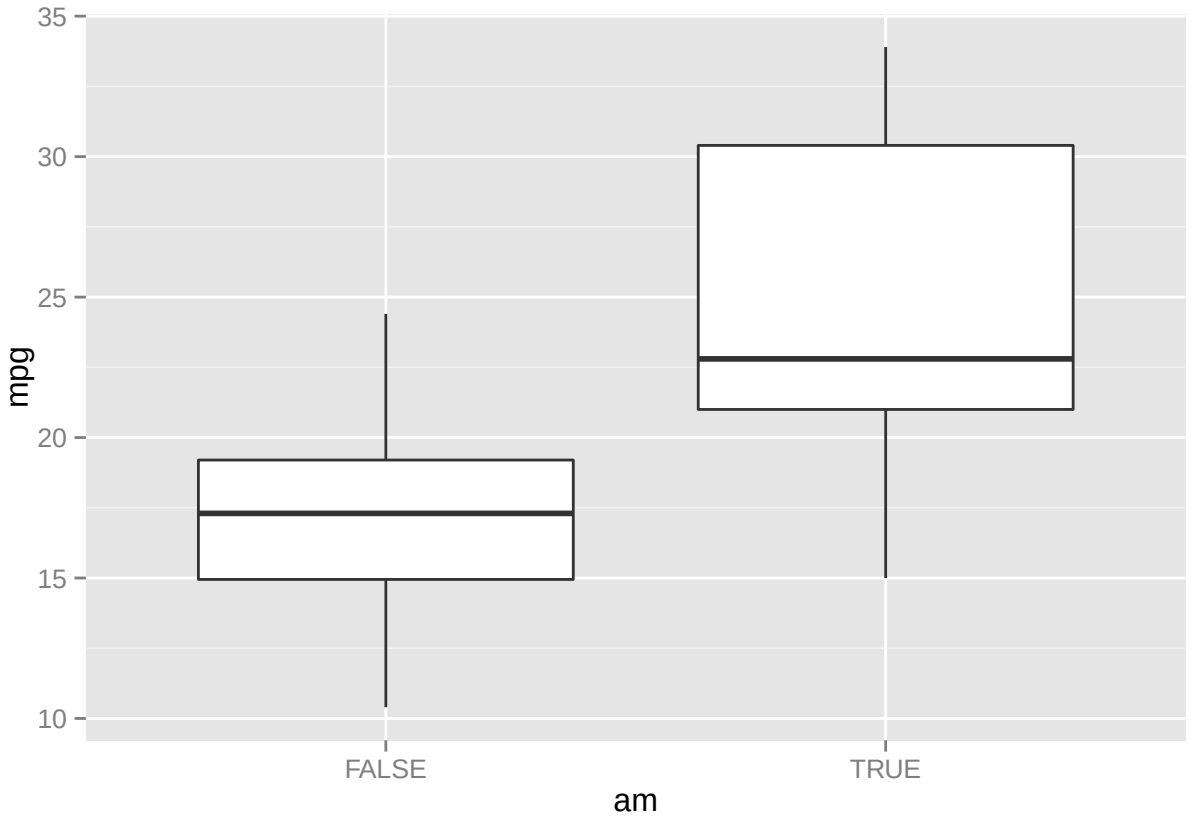
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    10.4    15.4    19.2    20.1    22.8    33.9
```

```
qplot(mtcars$mpg,binwidth=2)
```



plot boxplot of mpg against the automatic/ manual transmission

```
p  <- ggplot(mtcars, aes(am, mpg))
p + geom_boxplot()
```



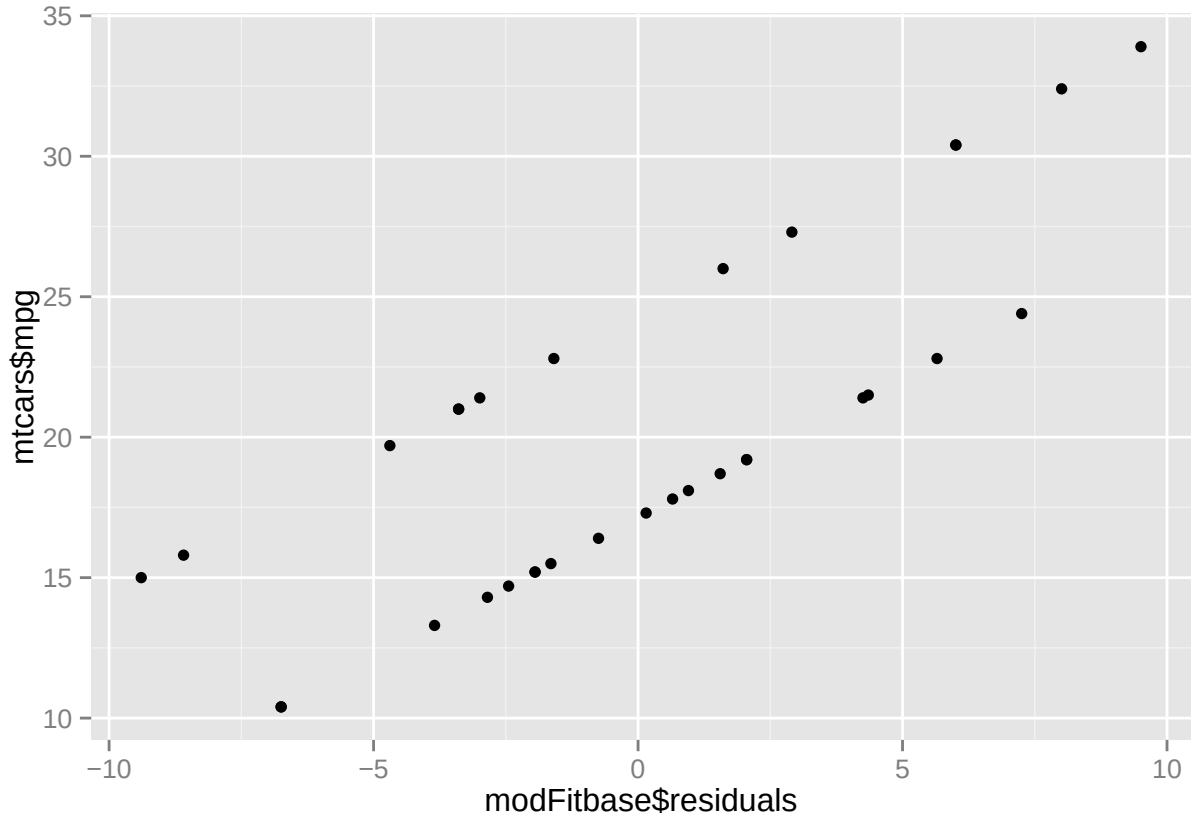
generate the base model with linear regression between mpg and automatic transmission

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

```
##
## Call:
## lm(formula = mpg ~ am, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -9.392 -3.092 -0.297  3.244  9.508
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)    17.15       1.12   15.25  1.1e-15 ***
## amTRUE          7.24       1.76    4.11  0.00029 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.9 on 30 degrees of freedom
## Multiple R-squared:  0.36,   Adjusted R-squared:  0.338
## F-statistic: 16.9 on 1 and 30 DF,  p-value: 0.000285
```

plot a graph between residual and mpg to see if these two are correlated.

```
qplot(modFitbase$residuals, mtcars$mpg)
```



The correlation seems to be significant. So Find the correlation between residual and output mpg

```
cor(modFitbase$residuals, mtcars$mpg)
```

```
## [1] 0.8001
```

This correlation is significant. we are missing some other important variables in our model.

```
modFitall  <- lm(mpg ~ ., data= mtcars)
summary(modFitall)
```

```
##
## Call:
## lm(formula = mpg ~ ., data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##  -3.45  -1.60  -0.12   1.22   4.63
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  12.3034    18.7179    0.66    0.518
## cyl          -0.1114     1.0450   -0.11    0.916
## disp          0.0133     0.0179    0.75    0.463
```

```
## hp           -0.0215     0.0218   -0.99    0.335
## drat          0.7871     1.6354    0.48    0.635
## wt           -3.7153     1.8944   -1.96    0.063 .
## qsec          0.8210     0.7308    1.12    0.274
## vs            0.3178     2.1045    0.15    0.881
## amTRUE        2.5202     2.0567    1.23    0.234
## gear          0.6554     1.4933    0.44    0.665
## carb         -0.1994     0.8288   -0.24    0.812
## ---
## 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:  0.807
## F-statistic: 13.9 on 10 and 21 DF,  p-value: 3.79e-07
```

By looking at `Pr(>|t|)` value for each variables, we see that `wt`, `am` and `qsec` are three significant variables.

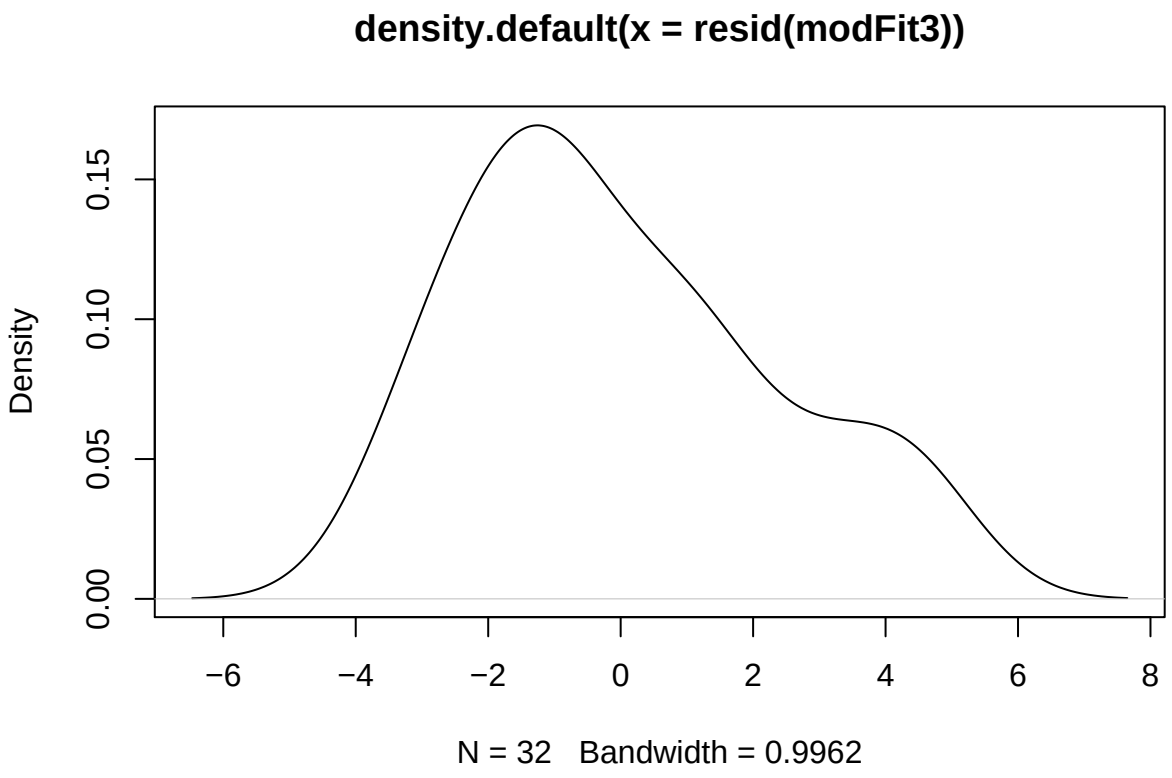
```
modFit3  <- lm(mpg ~ wt+qsec+am-1, mtcars)
summary(modFit3)
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am - 1, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.481 -1.556 -0.726  1.411  4.661
##
## Coefficients:
##         Estimate Std. Error t value Pr(>|t|)
## wt        -3.917      0.711   -5.51    7e-06 ***
## qsec       1.226      0.289    4.25  0.00022 ***
## amFALSE    9.618      6.960    1.38  0.17792
## amTRUE    12.554      6.057    2.07  0.04754 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.46 on 28 degrees of freedom
## Multiple R-squared:  0.988,  Adjusted R-squared:  0.986
## F-statistic:  574 on 4 and 28 DF,  p-value: <2e-16
```

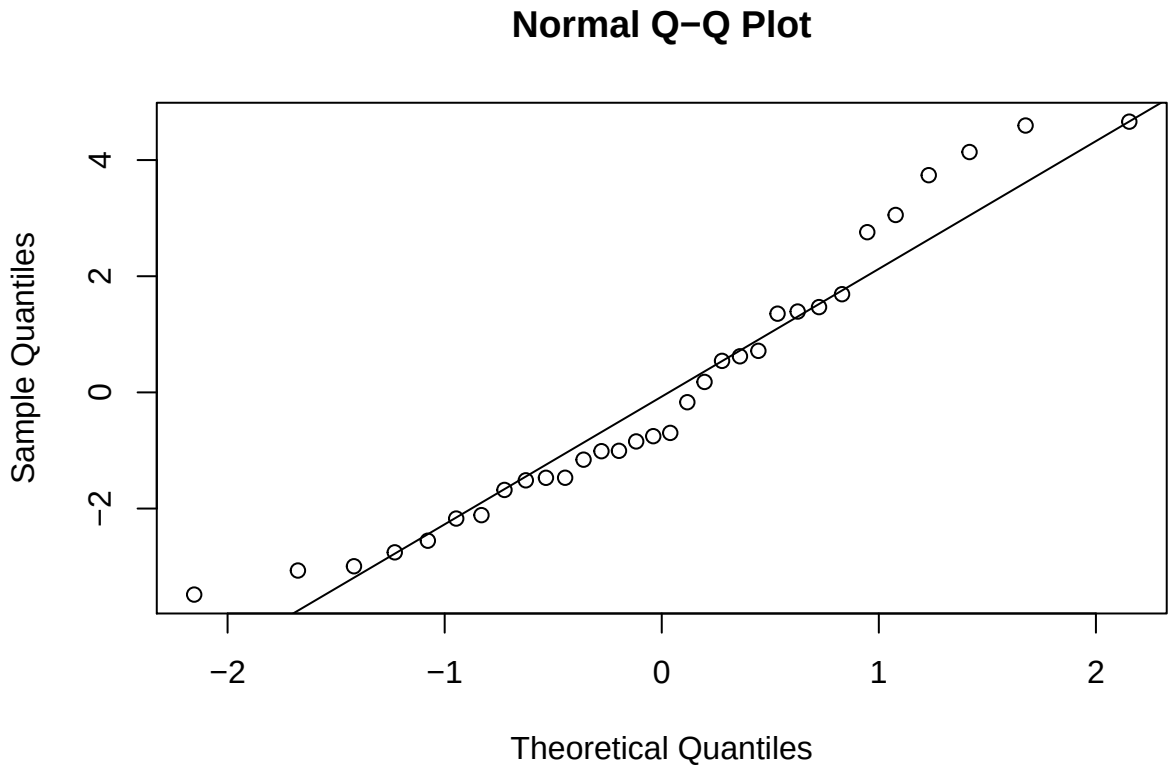
We can conclude that `am`(transmission type) has significant influence on `mpg` but 'wt'(weight) and `qsec`(1/4 mile time) also influence `mpg`.

Below are the residual plots to see the fit of data.

```
plot(density(resid(modFit3 ))) #A density plot
```




```
qqnorm(resid(modFit3 )) # A quantile normal plot - good for checking normality
qqline(resid(modFit3 ))
```



- Is an automatic or manual transmission better for MPG Manual transmission is better for MPG

- Quantify the MPG difference between automatic and manual transmissions" Having the manual transmissions would be 9.6178 mpg more effeciant than those having automatic transmissions on MPG, if the cars having same condition on all other features, particularly, the weight and quarter mile time.