Compare MPG between automatic and manual transmission cars

In this report I demonstrated the process to isolate the fittest model for the `mtcars` dataset. Through the diagnosis of this model, I picked up three outliers **Chrysler Imperial, Fiat 128, Toyota Corolla** from the dataset. The comparison of transimission types `am` on the miles per gallon `mpg` was evaluated from the reservation of outliers to the removal of outliers. At the initiation of this process, I called the package `MASS`.

Exploratory Data Analyses

Firstly I explored the correlations for each pair of variables in `mtcars` dataset. As Figure 1 (see Appendix) presented, `cyl`, `vs`, `am`, `gear`, and `carb` are categorical variables.

The original `mtcars` dataset was duplicated to new dataset `mtcars1` and the five categorical variables were transformed into `factor`. The model included all variables was stored in `all`. The coefficient of every variable were listed below:

```
mtcars1 <- mtcars
mtcars1$cyl <- as.factor(mtcars1$cyl)
mtcars1$vs <- as.factor(mtcars1$vs)
mtcars1$am <- as.factor(mtcars1$am)
mtcars1$gear <- as.factor(mtcars1$gear)
mtcars1$carb <- as.factor(mtcars1$carb)
all <- lm(mpg ~ . , data=mtcars1)
summary(all)$coef
```

```
##                Estimate  Std. Error     t value   Pr(>|t|)
## (Intercept) 23.87913244 20.06582026  1.19004018 0.25252548
## cyl6        -2.64869528  3.04089041 -0.87102622 0.39746642
## cyl8        -0.33616298  7.15953951 -0.04695316 0.96317000
## disp         0.03554632  0.03189920  1.11433290 0.28267339
## hp          -0.07050683  0.03942556 -1.78835344 0.09393155
## drat         1.18283018  2.48348458  0.47627845 0.64073922
## wt          -4.52977584  2.53874584 -1.78425732 0.09461859
## qsec         0.36784482  0.93539569  0.39325050 0.69966720
## vs1          1.93085054  2.87125777  0.67247551 0.51150791
## am1          1.21211570  3.21354514  0.37718957 0.71131573
## gear4        1.11435494  3.79951726  0.29328856 0.77332027
## gear5        2.52839599  3.73635801  0.67670068 0.50889747
## carb2       -0.97935432  2.31797446 -0.42250436 0.67865093
## carb3        2.99963875  4.29354611  0.69863900 0.49546781
## carb4        1.09142288  4.44961992  0.24528452 0.80956031
## carb6        4.47756921  6.38406242  0.70136677 0.49381268
## carb8        7.25041126  8.36056638  0.86721532 0.39948495
```

Variable Selection

The fittest model was decided by the smallest AIC computed by the stepwise algorithm. The smallest

AIC was acquried through `both` directions. Three predictors `wt`, `qsec`, and `am` were selected in the fittest model.

```
fit <- stepAIC(all, direction="both")
summary(fit)$coef
```

```
##              Estimate Std. Error   t value     Pr(>|t|)
## (Intercept)  9.617781  6.9595930  1.381946 1.779152e-01
## wt          -3.916504  0.7112016 -5.506882 6.952711e-06
## qsec         1.225886  0.2886696  4.246676 2.161737e-04
## am1          2.935837  1.4109045  2.080819 4.671551e-02
```

Diagnostics of Residuals

Figure 2(see Appendix) illustrated that the residuals of most observations obey the normalization assumption exceot **Chrysler Imperial, Fiat 128, Toyota Corolla**. The plot compared resesuals and leverages indicates that the three observations contributed the least influence to the fittest model. Therefore, I indicated **Chrysler Imperial, Fiat 128, Toyota Corolla** as the outliers of `mtcars` dataset.

Conclusion

To decrease the uncertainty of my conclusion, I compared the MPG between automatic and manual transmission in use of t test. Both t tests, with and without outliers, showed the significant result that the mannual cars perfomred better on MPG than the automatic cars. I set sigificant level at `.05`.
With the reservation of outliers, the average MPG for automatic car was **17.15** and the average MPG for mannual car was **24.39**. The statistical value was t(**30**) = **-4.11**, p value **< .05**. That difference of MPG **7.24** supports the better MPG for the automatic transmission.
After excluded the outliers, the average MPG for automatic car was **17.28** and the average MPG for mannual car was **22.8**. The statistical value was t(**27**) = **-3.25**, p value **< .05**. That difference of MPG **5.52** supports the better MPG for the automatic transmission.

Note

Bold words are produced by the codes embeded in line. Below column shows two examples I used in this report.

```
Through the diagnosis of this model, I picked up three outliers **r rowname
s(mtcars)[c(17, 18, 20)]** from the dataset.
...the average MPG for automatic car was **r round( as.numeric(unlist(t.tes
t(mpg ~ am, var.equal = TRUE,mtcars1))[6]) , digits=2)**...
```

Appendix

Figure 1. Explore the relation of variables

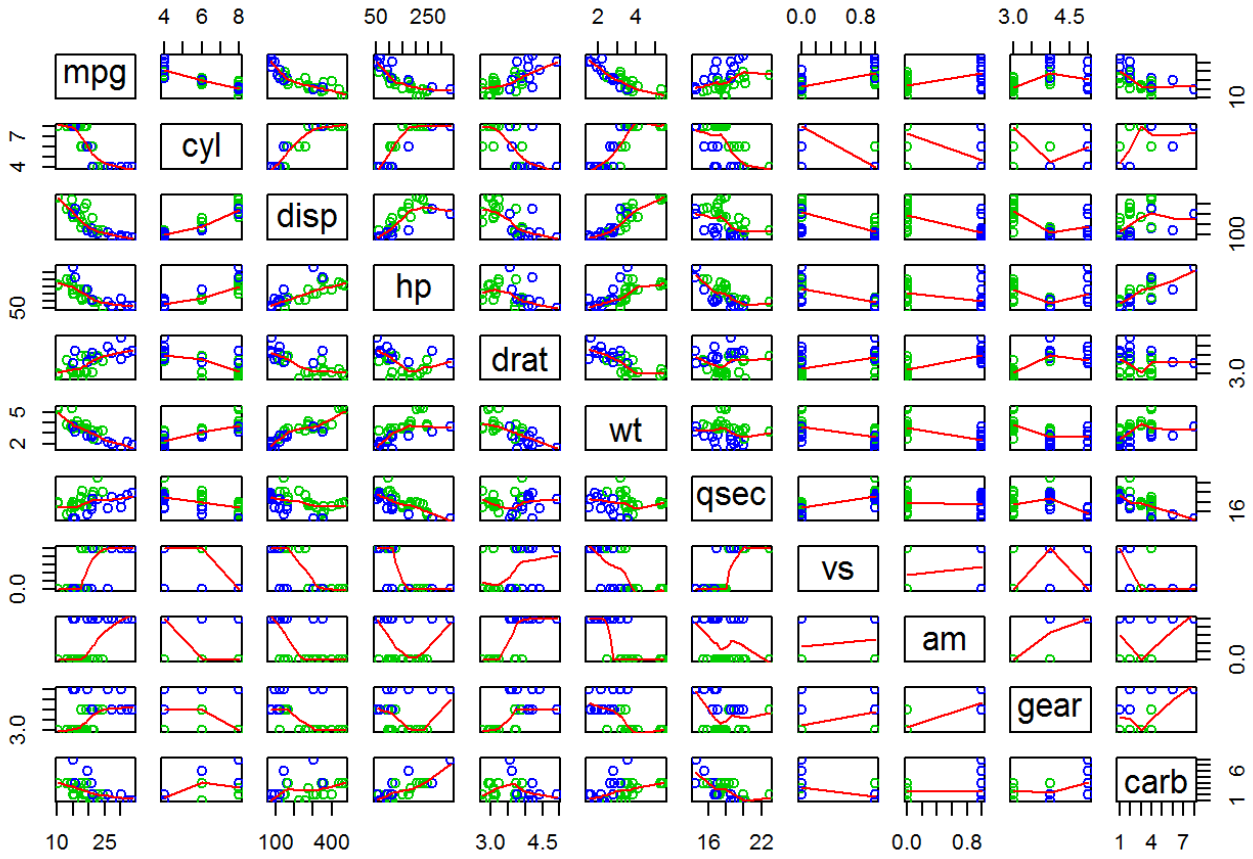

Figure 2. Plots for evaluate outliers