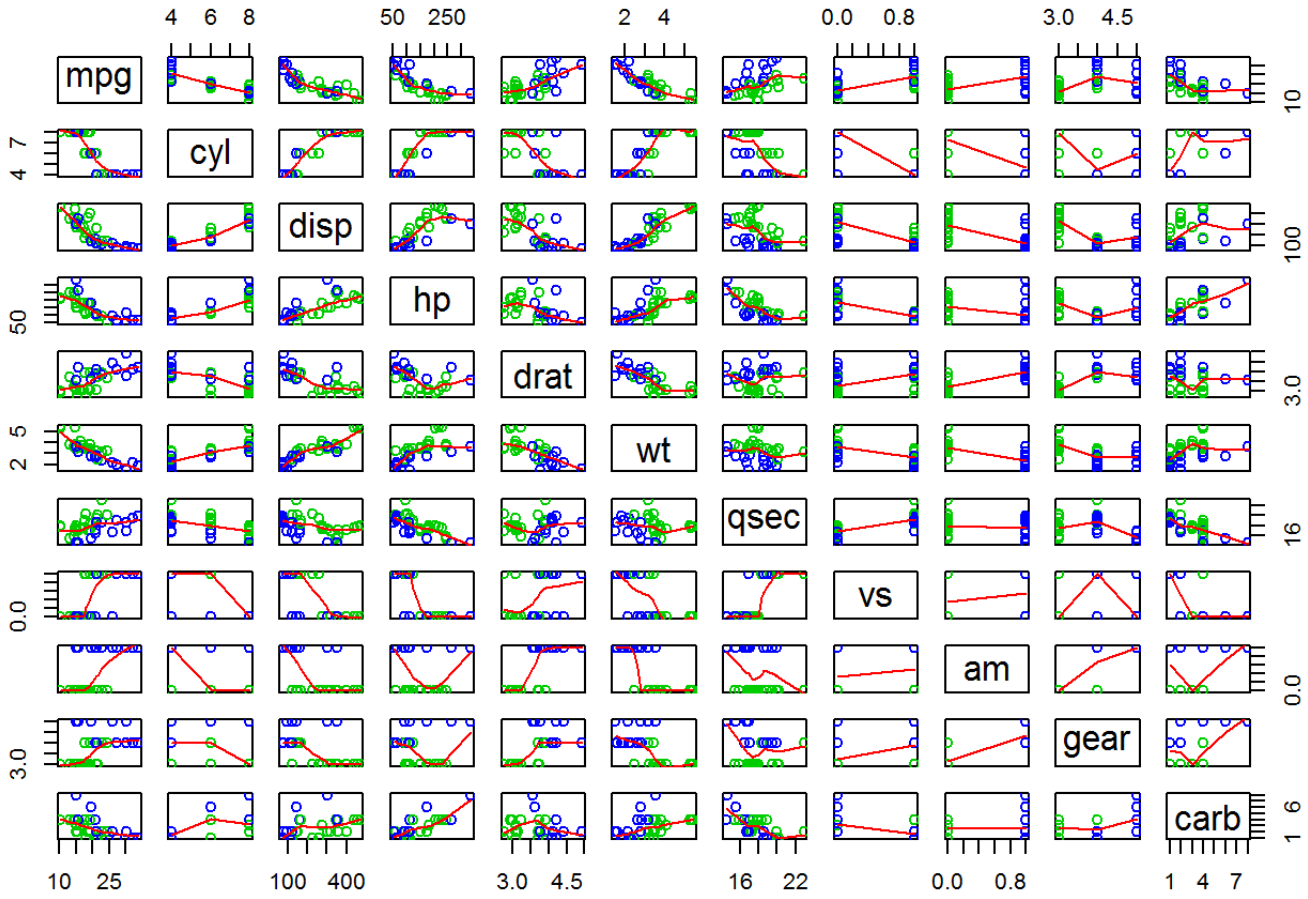
Compare MPG between automatic and manual transmission cars

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 the below plot 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` :

```
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)
```

Variable Selection

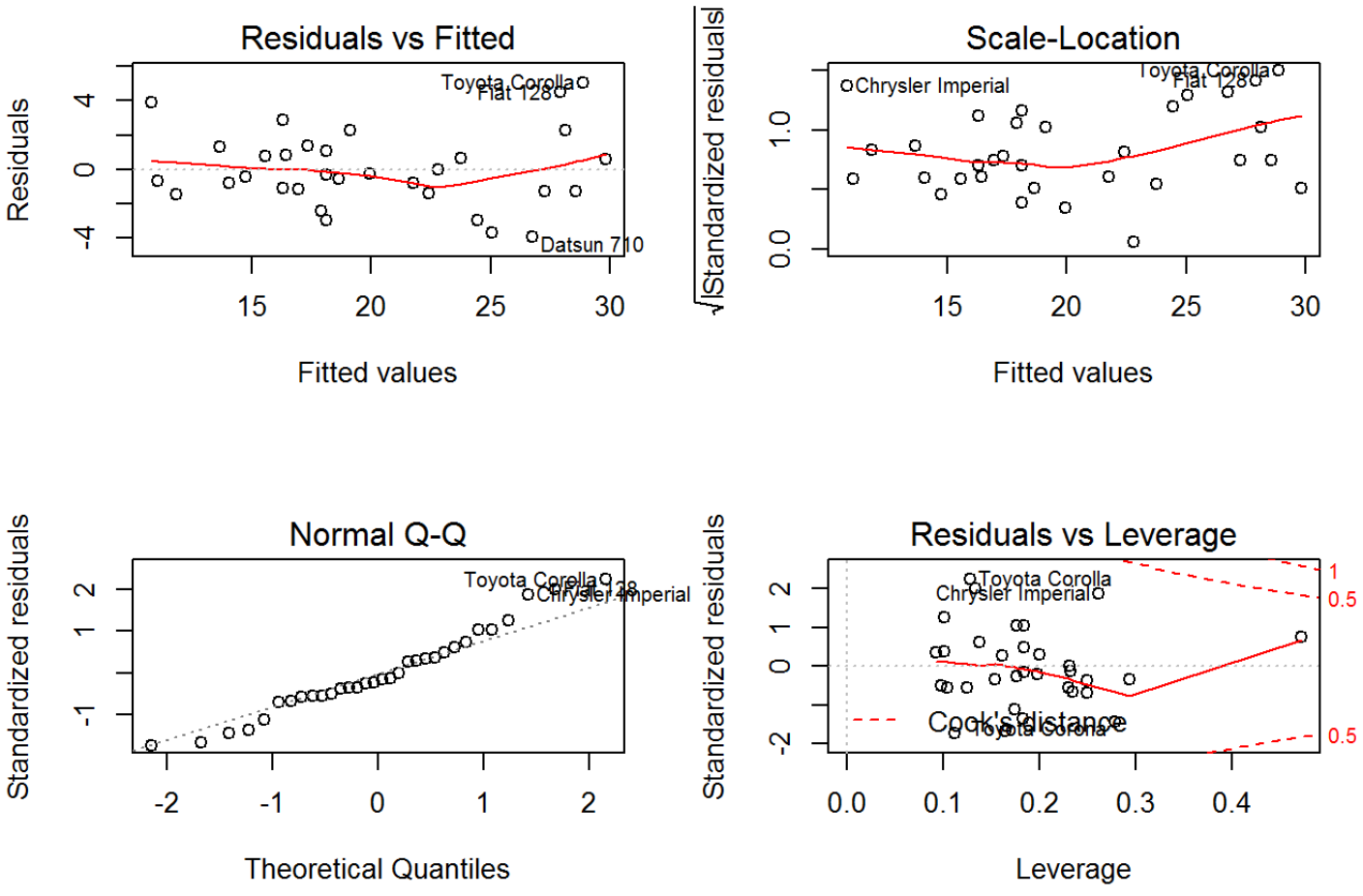
The fittest model was decided by the smallest AIC computed by the stepwise algorithm. The smallest AIC was acquried through `both` directions. Three predictors `wt`, `qsec`, and `am` were selected in the fittest model.

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

```
##              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

Below plots 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.



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 rownames(mtcars)
[c(17, 18, 20)]** from the dataset.
...the average MPG for automatic car was **r round( as.numeric(unlist(t.test(mpg ~ a
m, var.equal = TRUE,mtcars1))[6]) , digits=2)**...
```