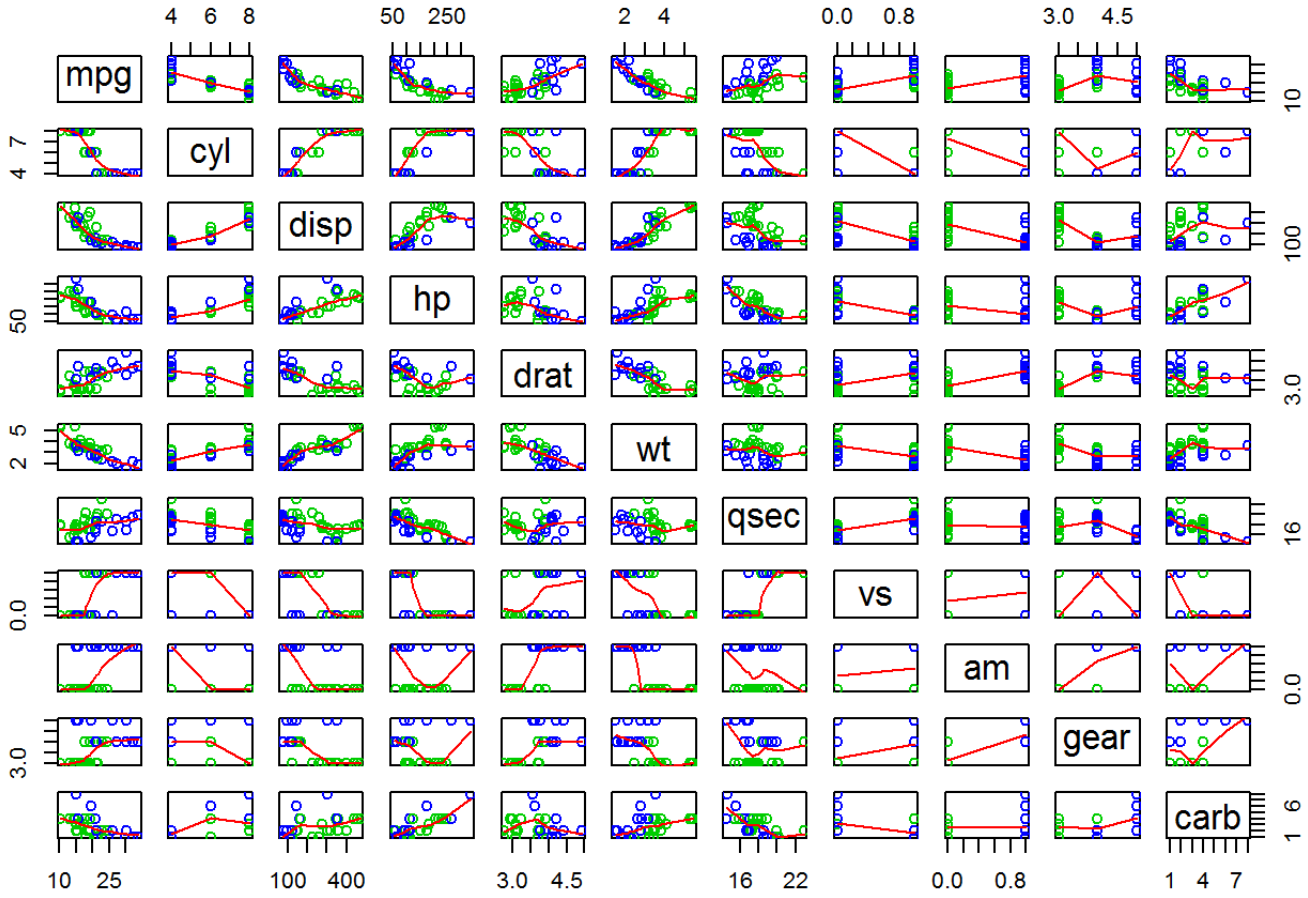
Compare MPG between automatic and manual transmission cars

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, five variables are categorical variables.



The original `mtcars` dataset was duplicated to the set `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

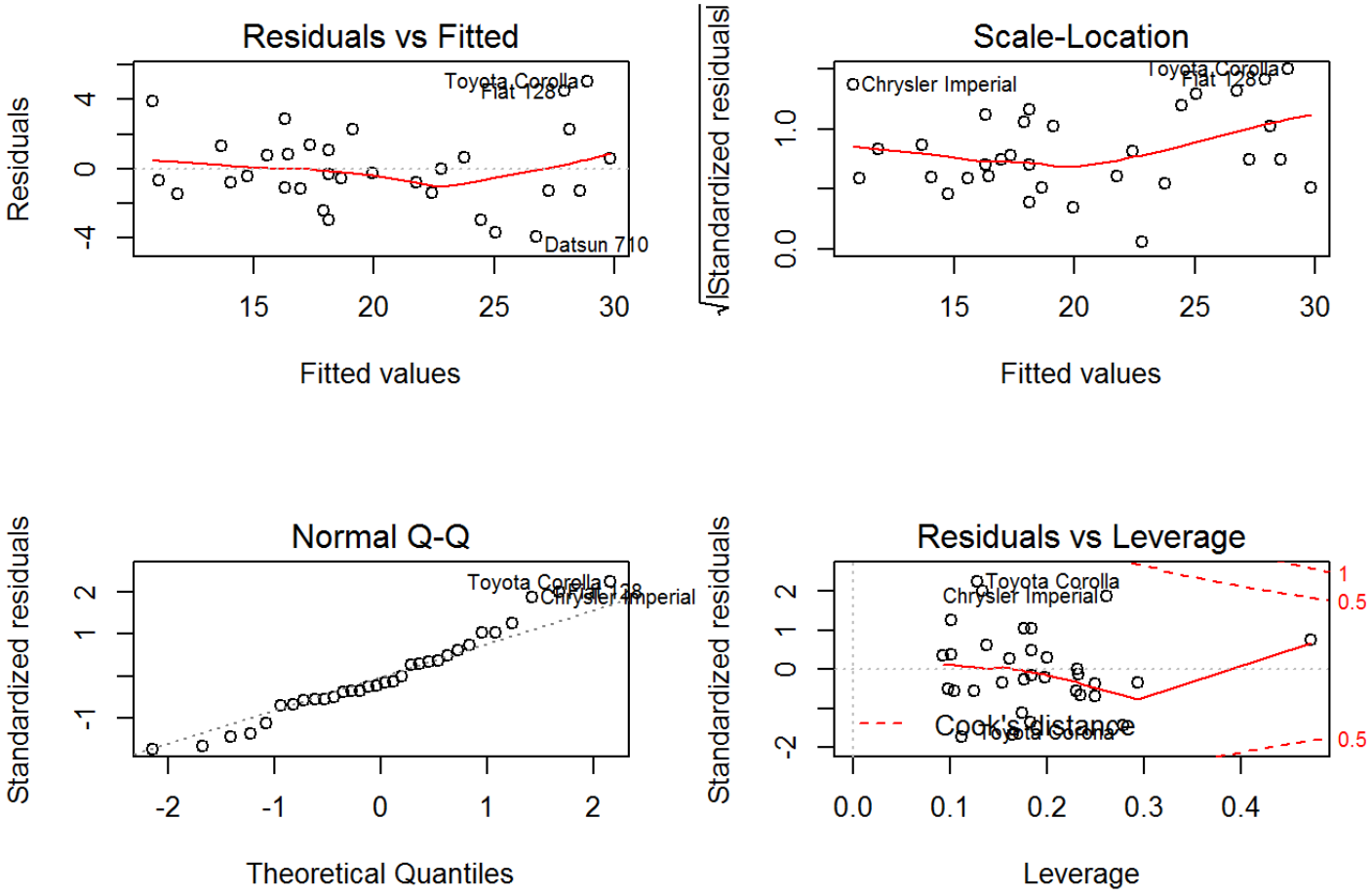
The fittest model was decided by the smallest AIC computed by the stepwise algorithm. The smallest AIC was acquried through `both` directions. Three predictors `wt`, `qsec`, and `am` were selected in the fittest model.

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

```
##                 Estimate Std. Error   t value     Pr(>|t|)
## (Intercept) 33.70832390 2.60488618 12.940421 7.733392e-13
## cyl6        -3.03134449 1.40728351 -2.154040 4.068272e-02
## cyl8        -2.16367532 2.28425172 -0.947214 3.522509e-01
## hp          -0.03210943 0.01369257 -2.345025 2.693461e-02
## wt          -2.49682942 0.88558779 -2.819404 9.081408e-03
## am1          1.80921138 1.39630450  1.295714 2.064597e-01
```

Diagnostics of Residuals

Below plots illustrated that the residuals of most observations match the normalization assumption in addition to Chrysler Imperial, Fiat 128, Toyota Corolla. The plot comparing resesuals and leverages indicate the three observations contributed the least influence to the model. Therefore, I indicated Chrysler Imperial, Fiat 128, Toyota Corolla as the outliers of `mtcars` dataset.



To decrease the uncertainty of our conclusion, I compared the MPG betlen automatic and manual transmission in use of t test. Both T tests with and without outliers shold 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**.
Without 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**.

Appendix

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