Assignment 1: Motor Trend report

C. Deramond

March, 2015

Executive summary

In this document we explored car performance in terms of miles per gallon (mpg). By methods of regression, we conclude that manual transmission will give better performance in a boost of app 1.8 MPG, ceteris paribus.

Alas the question are answered by the model in a pretty straightforward way. In the following sections you may review the procedure as well as the validations of the model.

Introduction

In this report we'll study the relationship between MPG and several features of vehicles. By the end of this study we hope to have answered these two questions:

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

All the data used in this report comes from `mtcars` dataset available on R datasets package.

Analysis

Exploratory

We'll start the process by discarding all variables which have a lower than 0.55 correlation with our output `mpg`. This process lefts us with 8 variables.

A quick scatter matrix shows many interesting relationships that we should address:

- `cyl, vs, am` are discrete variable variables. We take out `vs` as it is of low interest. And turn the rest into factor
- the correlation within the predictors is quite high, so risk of multicollinearity is there.

If you see Graph 02, the risk for muticolinearity (i.e. our predictors are not independent among each other) is high.

Regression

As starting point, let's consider the suggested model by stepwise: **mpg ~ cyl + hp + wt + am**

As we are including 2 factors, interaction will be considered see interaction plot. A second look at the ANOVA table, tells that interaction is not significant:

```
## Analysis of Variance Table
##
## Response: mpg
##                      Df Sum Sq Mean Sq F value    Pr(>F)
## am                    1 405.15  405.15 67.0195 2.094e-08 ***
## cyl                   2 456.40  228.20 37.7487 3.880e-08 ***
## hp                    1  67.30   67.30 11.1321  0.002755 **
## wt                    1  46.17   46.17  7.6379  0.010801 *
## interaction(cyl, am)  2   5.94    2.97  0.4912  0.617886
## Residuals            24 145.09    6.05
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

We rerun model without interaction, thus suggesting using manual transmission has a gain of 1.81 in mpg, which makes sense. Time to evaluate our model.

```
##
## Call:
## lm(formula = mpg ~ am + cyl + hp + wt, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9387 -1.2560 -0.4013  1.1253  5.0513
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 33.70832    2.60489  12.940 7.73e-13 ***
## amManual     1.80921    1.39630   1.296  0.20646
## cyl6        -3.03134    1.40728  -2.154  0.04068 *
## cyl8        -2.16368    2.28425  -0.947  0.35225
## hp          -0.03211    0.01369  -2.345  0.02693 *
## wt          -2.49683    0.88559  -2.819  0.00908 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.41 on 26 degrees of freedom
## Multiple R-squared:  0.8659, Adjusted R-squared:  0.8401
## F-statistic: 33.57 on 5 and 26 DF,  p-value: 1.506e-10
```

Model evaluation

Nested ANOVA

as we have merely 4 predictor variables, a nested ANOVA, suffices to prove our variables are indeed adding explicative value. As it shows below,

```
## Analysis of Variance Table
##
## Model 1: mpg ~ am
## Model 2: mpg ~ am + cyl
## Model 3: mpg ~ am + cyl + hp
## Model 4: mpg ~ am + cyl + hp + wt
##   Res.Df    RSS Df Sum of Sq      F    Pr(>F)
```

```
## 1     30 720.90
## 2     28 264.50  2    456.40 39.286 1.388e-08 ***
## 3     27 197.20  1     67.30 11.585  0.002164 **
## 4     26 151.03  1     46.17  7.949  0.009081 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model diagnostics & validation

Graphically we can examine our normality and homoskedacity assumptions see graphs 4. Additionally a validation of components vs residuals can be seen on graphs 5

Results

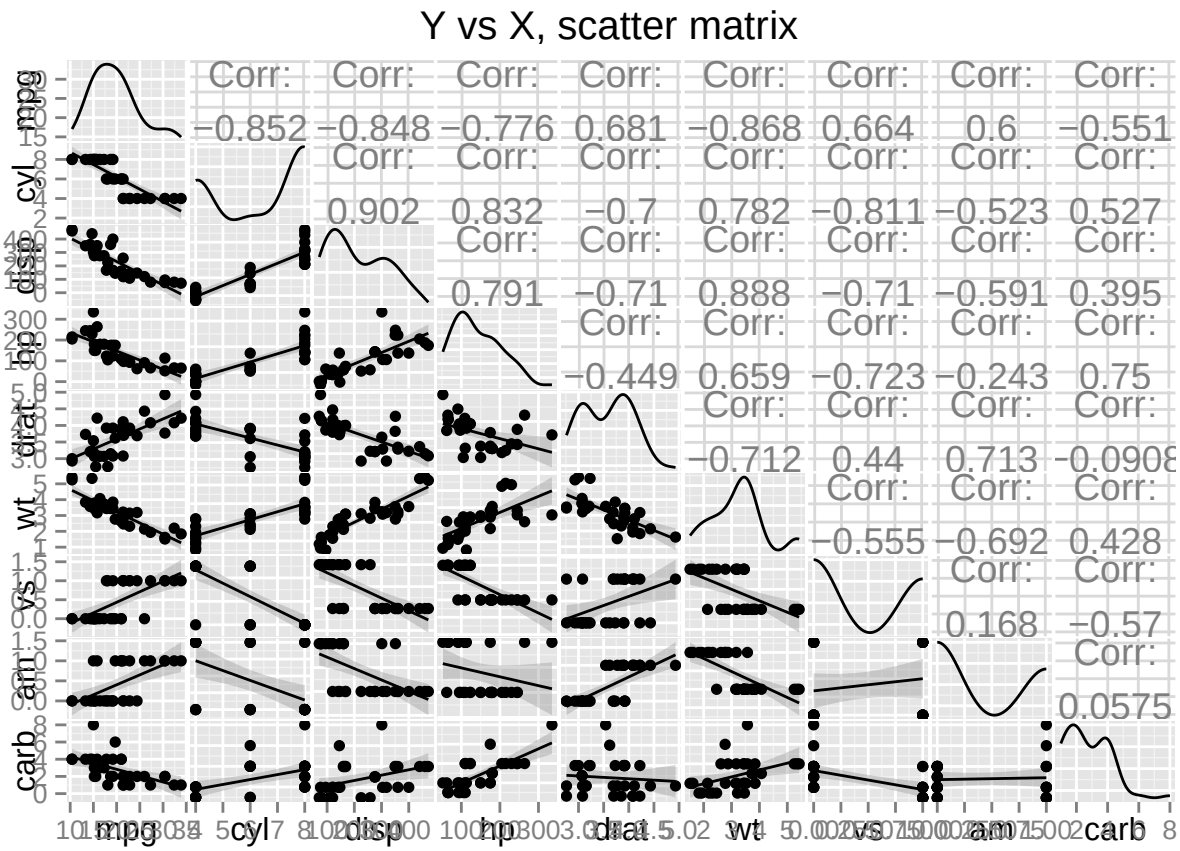
The model as presented by its coefficients is self explanatory:

```
## (Intercept)    amManual        cyl6        cyl8          hp          wt
## 33.70832390  1.80921138 -3.03134449 -2.16367532 -0.03210943 -2.49682942
```

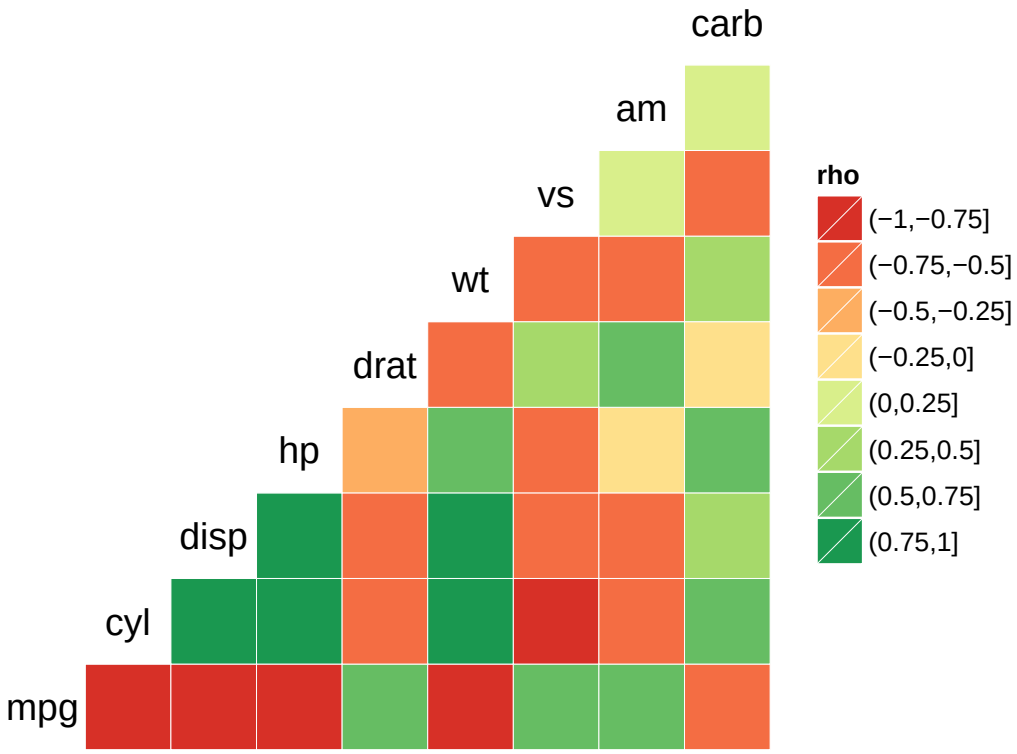
Manual transmission cars will put out better mileage app 1.81, ceteris paribus.

Appendix

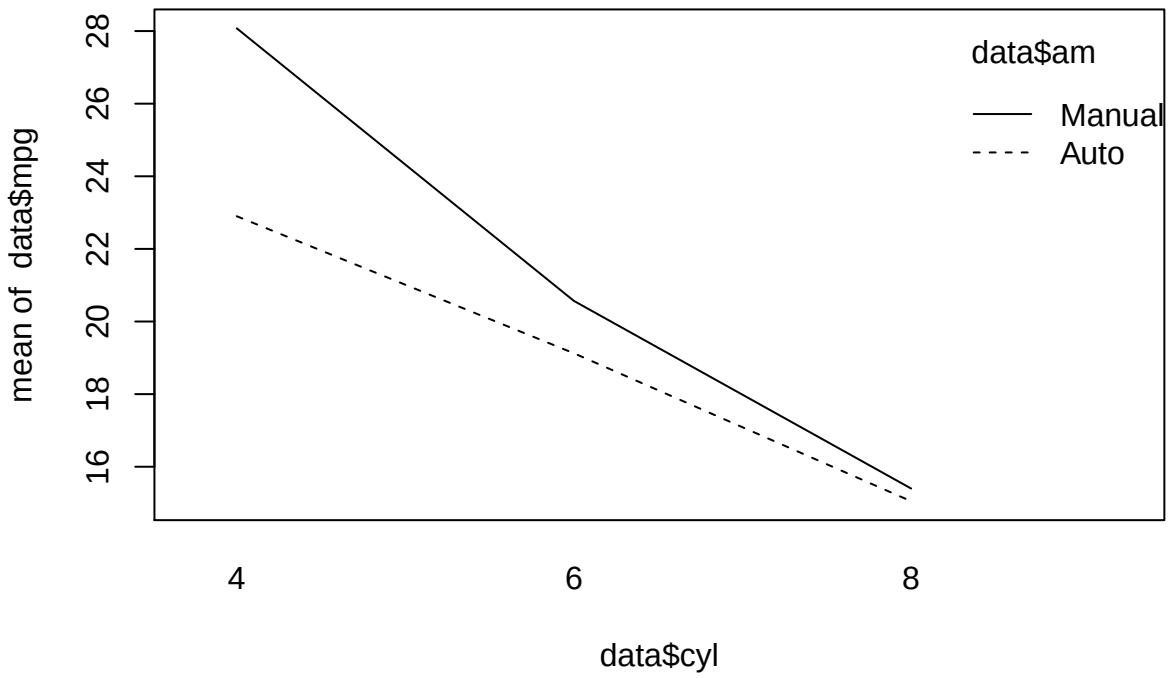
Graph 1: Scatter matrix, all variables



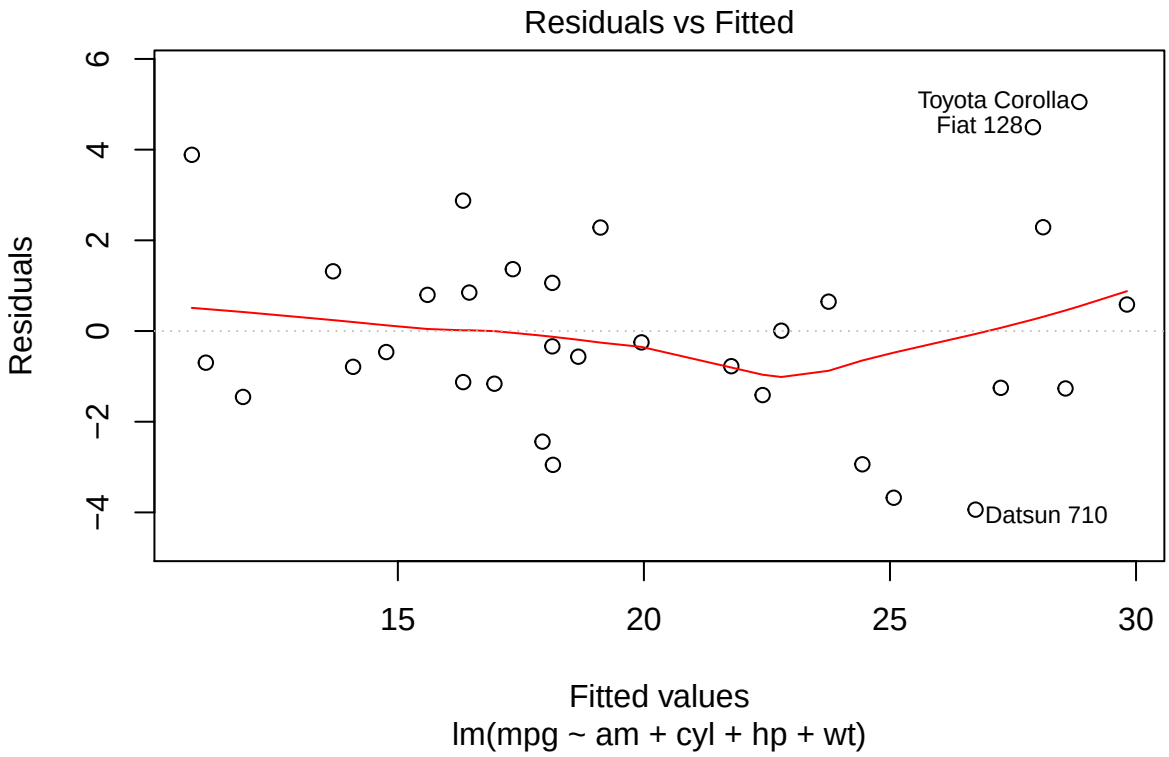
Graph 2: Correlation matrix

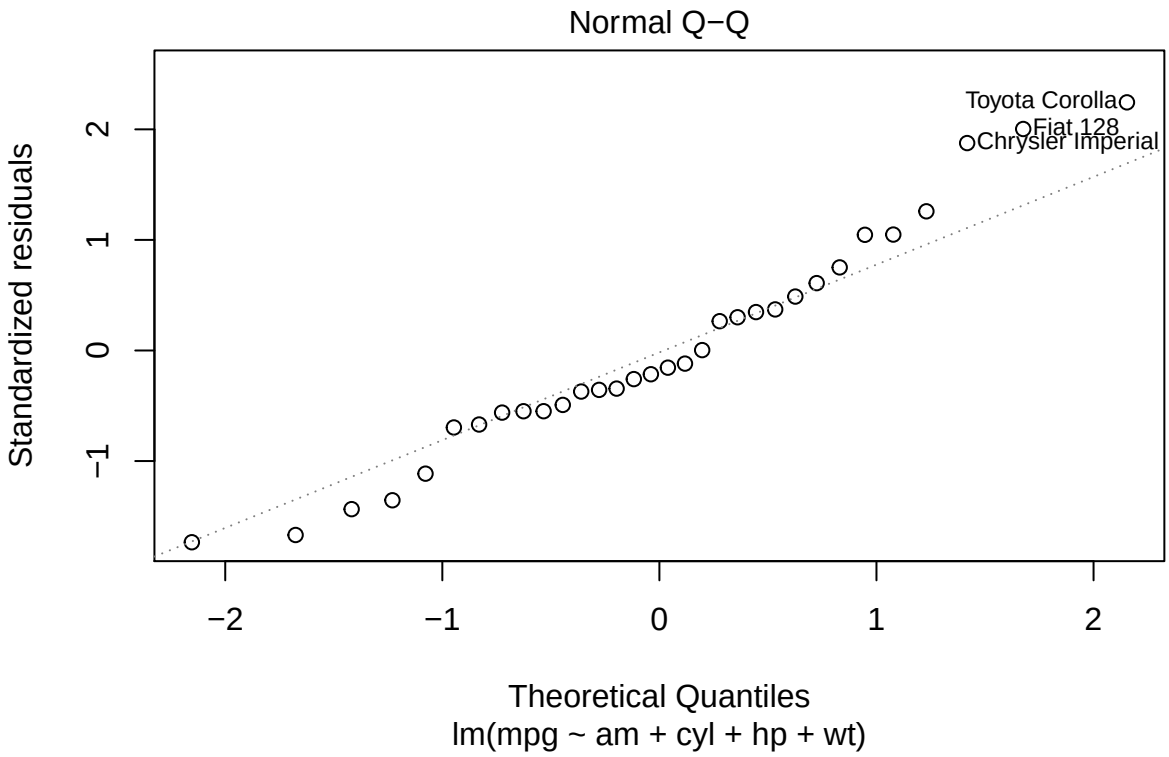


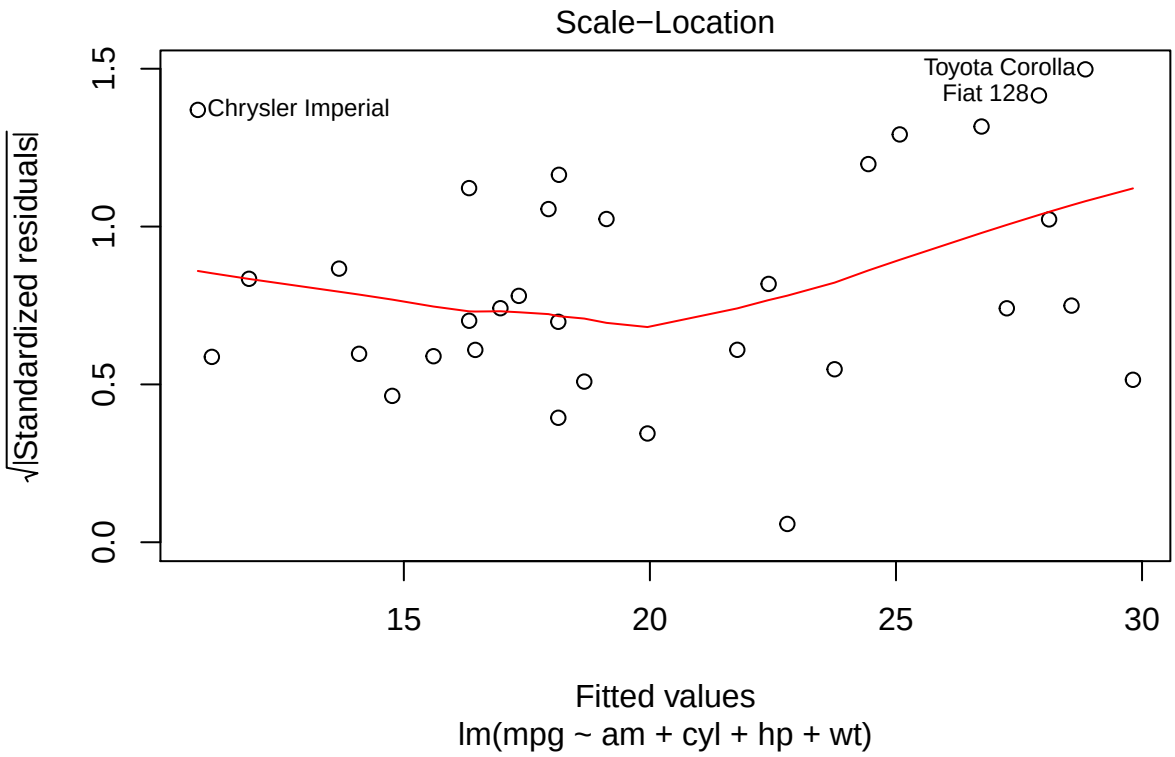
Graph 3: Interaction plot am vs cyl

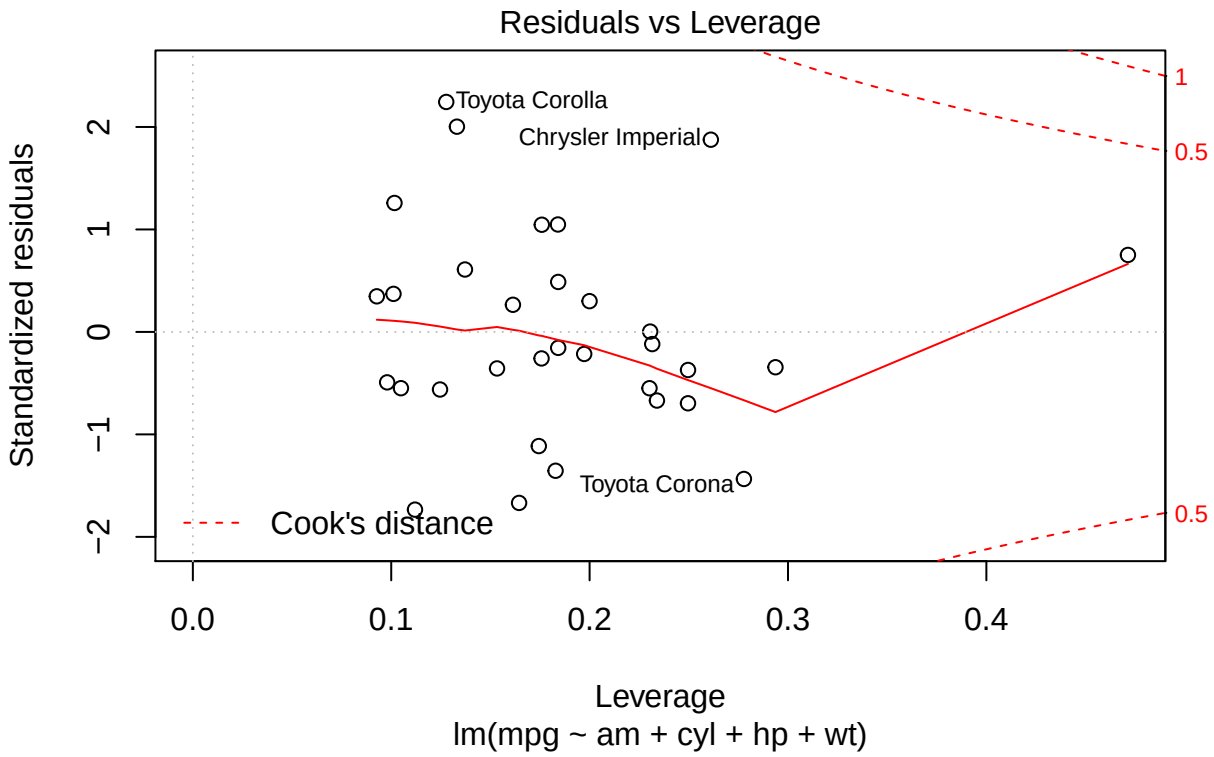


Graph 4: Diagnostics plots for model

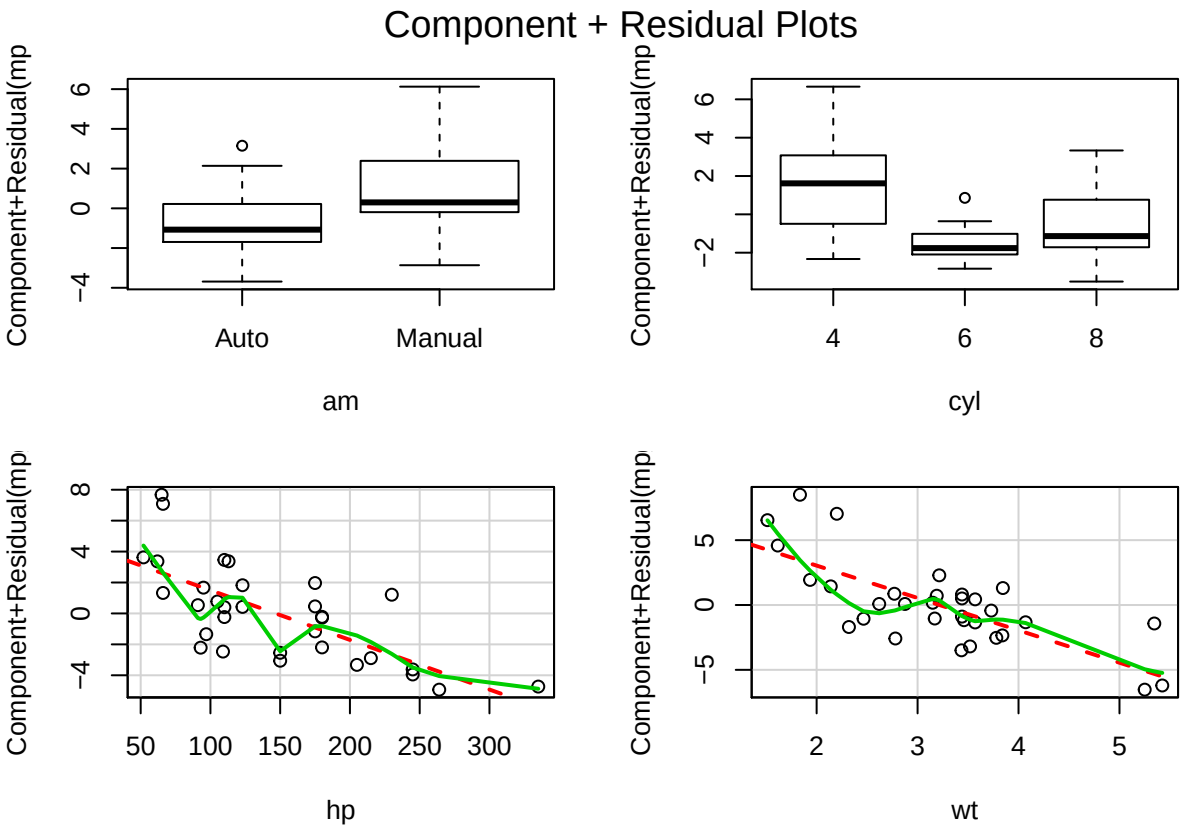

Residuals vs Fitted
Residuals
Fitted values
lm(mpg ~ am + cyl + hp + wt)
Toyota Corolla
Fiat 128
Datsun 710
6
4
2
0
−2
−4
15
20
25
30


Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(mpg ~ am + cyl + hp + wt)
Toyota Corolla
Fiat 128
Chrysler Imperial
−2
−1
0
1
2


Scale−Location
Chrysler Imperial
Toyota Corolla
Fiat 128
√|Standardized residuals|
0.0
0.5
1.0
1.5
15
20
25
30
Fitted values
lm(mpg ~ am + cyl + hp + wt)



Graph 5: Components and Residual plots


Component + Residual Plots
Component+Residual(mp
−4
0
2
4
6
Auto
Manual
am
Component+Residual(mp
−2
2
4
6
4
6
8
cyl
Component+Residual(mp
−4
0
4
8
50
100
150
200
250
300
hp
Component+Residual(mp
−5
0
5
2
3
4
5
wt