Coursera Regression Models project

Executive summary

I'm trying to explore the relationship between a set of variables and miles per gallon (MPG) (outcome). Particulary I'm going to discuss the questions below:

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

Data processing

```
data(mtcars)
mtcars$cyl <- factor(mtcars$cyl)
mtcars$am <- factor(mtcars$am)
mtcars$gear <- factor(mtcars$gear)
mtcars$carb <- factor(mtcars$carb)
mtcars$vs <- factor(mtcars$vs)
```

Model building

To compare automatic transmission vs. manual we need to keep all other variable equal, so create model, which have all variables.

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

Many of variables add almost no information, lets exclude them. Also plot mpg versus weight and horsepower, as they have biggest significance (Fig 1)

Second model

```
fit2 <- lm(mpg~wt+hp+am, data = mtcars)
summary(fit2)$coefficients
```

```
##             Estimate Std. Error t value  Pr(>|t|)
## (Intercept) 34.00288   2.642659  12.867 2.824e-13
## wt          -2.87858   0.904971  -3.181 3.574e-03
## hp          -0.03748   0.009605  -3.902 5.464e-04
## am1          2.08371   1.376420   1.514 1.413e-01
```

On Fig 2 lets analize residuals of this model and regression variables. It's known to be that automatic transmission is heavier. Lets see if we can improve the model considering intaraction between weight and transmission type. ### Model upgrade

```r
fit3 <- update(fit2, mpg~ wt + hp + am + wt * am )
anova(fit3)
```

```
## Analysis of Variance Table
##
## Response: mpg
##           Df Sum Sq Mean Sq F value  Pr(>F)
## wt         1    848     848  155.87   1e-12 ***
## hp         1     83      83   15.31 0.00056 ***
## am         1     15      15    2.71 0.11111
## wt:am      1     33      33    6.15 0.01968 *
## Residuals 27    147       5
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
summary(fit3)
```

```
##
## Call:
## lm(formula = mpg ~ wt + hp + am + wt:am, data = mtcars)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -3.064 -1.332 -0.935  1.218  5.082
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  30.9473     2.7234   11.36  8.5e-12 ***
## wt           -2.5156     0.8445   -2.98   0.0061 **
## hp           -0.0270     0.0098   -2.75   0.0105 *
## am1          11.5548     4.0233    2.87   0.0078 **
## wt:am1       -3.5779     1.4428   -2.48   0.0197 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.33 on 27 degrees of freedom
## Multiple R-squared:  0.87,   Adjusted R-squared:  0.85
## F-statistic:   45 on 4 and 27 DF,  p-value: 1.45e-11
```

As we can see interraction between weight and transmission in significant.

Conclusions

- Cars with manual transmission have better for MPG (beta before am is positive)
- Manual transmission gives 11.58 higher MPG. As we can see there is different relationship between weight and MPG for automatic and manual transmission. As average car weight with manual transmission is 2.411, average increase in MPG is 2.95 = 11.58-2.411*3.5779

Appendix

```
par(mfrow = c(1,2))
plot(mtcars$wt, mtcars$mpg , col = mtcars$am,pch=19)
plot(mtcars$hp, mtcars$mpg , col = mtcars$am,pch=19)
```

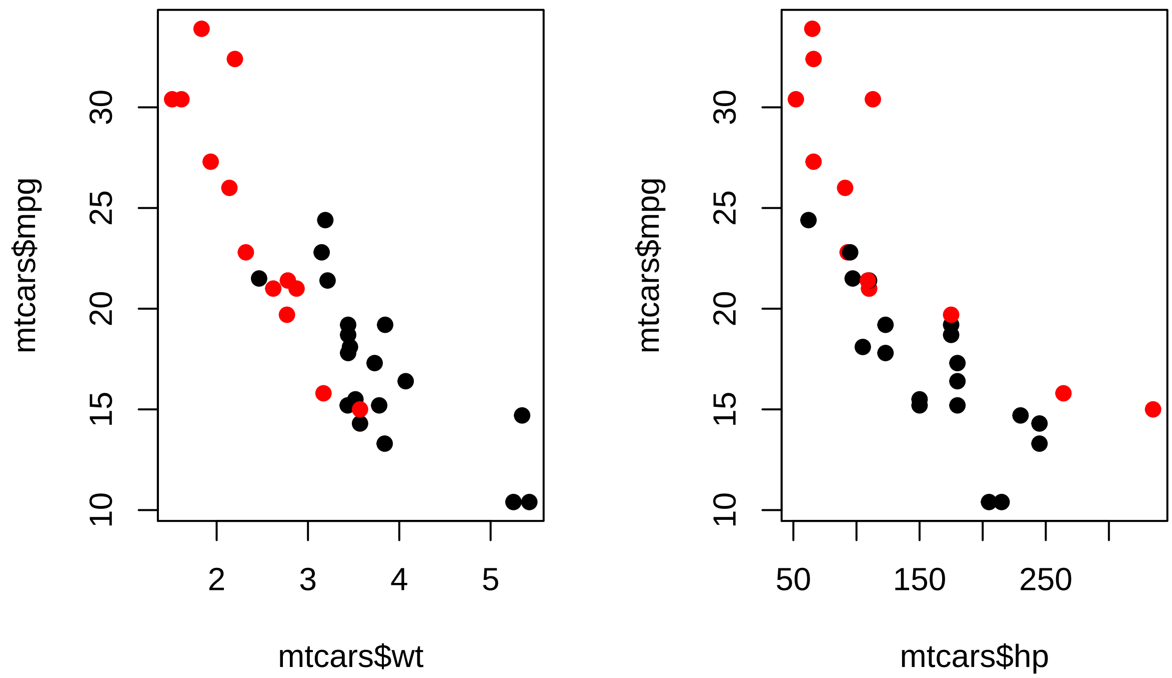
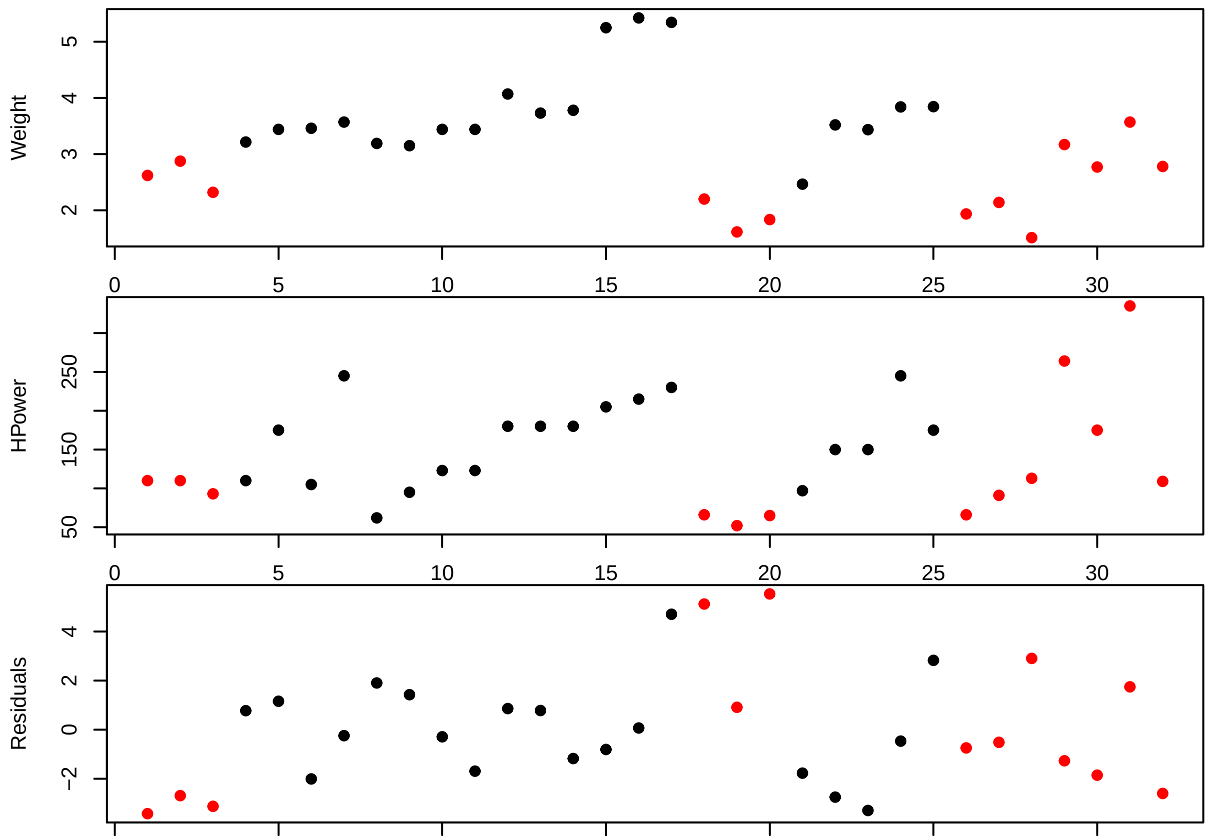
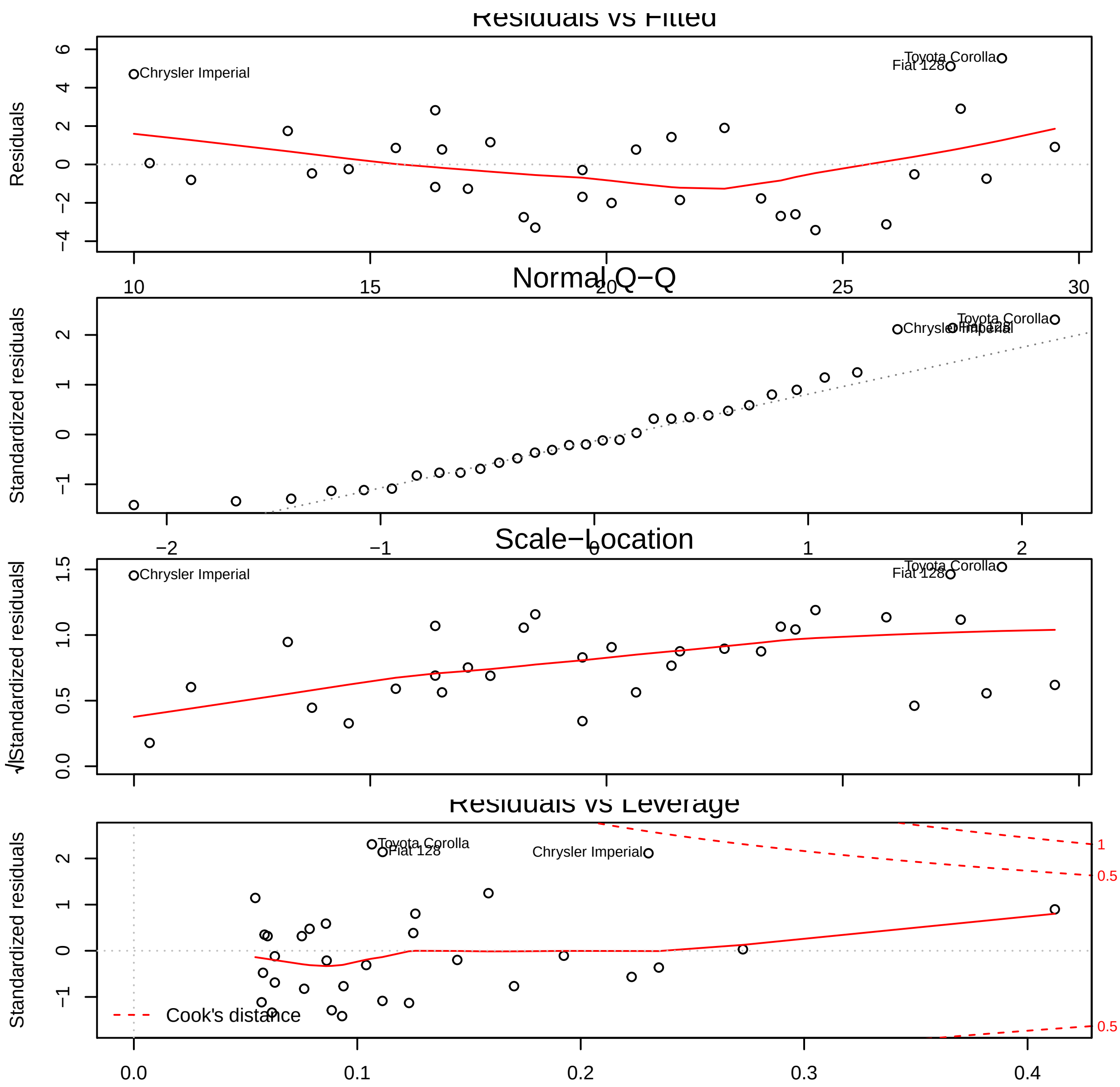


Fig. 1

```r
par(mfrow = c(3,1), mar = c(1,4,1,2))
plot(mtcars$wt,  col = mtcars$am,pch=19, ylab='Weight')
plot(mtcars$hp,  col = mtcars$am,pch=19, ylab='HPower')
# plot residuals
plot(fit2$residuals, col = mtcars$am, pch =19, ylab='Residuals')
```




```r
plot(fit2)
```


Residuals vs Fitted
Residuals
Fitted values
Chrysler Imperial
Toyota Corolla
Fiat 128
Normal Q−Q
Standardized residuals
Chrysler Imperial
Toyota Corolla
Fiat 128
Scale−Location
√|Standardized residuals|
Chrysler Imperial
Toyota Corolla
Fiat 128
Residuals vs Leverage
Standardized residuals
Toyota Corolla
Fiat 128
Chrysler Imperial
Cook's distance