Regression

Mayar

2/10/2022

Below is a summmary of our models. We showed that automatic transmission cars have lower mpg compared with manual transmission cars. The mpg is largely determined by the interplay between weight, acceleration and transmission. Given the above analysis, the original question (automatic transmission vs manual transmission) is not really answered, and should be considered in the context of weight and acceleration speed.

0. Preprocessing

```
data(mtcars)
names(mtcars)
```

```
##  [1] "mpg"  "cyl"  "disp" "hp"   "drat" "wt"   "qsec" "vs"   "am"   "gear"
## [11] "carb"
```

```
summary(cars)
```

```
##      speed           dist       
##  Min.   : 4.0   Min.   :  2.00  
##  1st Qu.:12.0   1st Qu.: 26.00  
##  Median :15.0   Median : 36.00  
##  Mean   :15.4   Mean   : 42.98  
##  3rd Qu.:19.0   3rd Qu.: 56.00  
##  Max.   :25.0   Max.   :120.00
```

1. Analysis

```
cor(mtcars$mpg,mtcars[,-1])
```

```
##            cyl       disp         hp      drat         wt     qsec        vs
## [1,] -0.852162 -0.8475514 -0.7761684 0.6811719 -0.8676594 0.418684 0.6640389
##             am      gear       carb
## [1,] 0.5998324 0.4802848 -0.5509251
```

2. Automatic or manual transmission?

```
mtcars$am <- as.factor(mtcars$am)
levels(mtcars$am) <-c("Automatic", "Manual")

t.test(mtcars$mpg~mtcars$am,conf.level=0.95)
```

```
##
##  Welch Two Sample t-test
##
## data:  mtcars$mpg by mtcars$am
## t = -3.7671, df = 18.332, p-value = 0.001374
## alternative hypothesis: true difference in means between group Automatic and group Manual is not equa
## 95 percent confidence interval:
##  -11.280194  -3.209684
## sample estimates:
## mean in group Automatic    mean in group Manual
##                17.14737                24.39231
```

The p-value is 0.001374, we may reject the null hypothesis and conclude, that automatic transmission cars have lower mpg compared with manual transmission cars - but this assumption is based on all other characteristics of automatic transmission cars and manual transmission cars are same (e.g: both have same weight distribution). This needs to be further explored in a multiple linear regression analysis.

3. Quantifying mpg difference

```
stepmodel = step(lm(data = mtcars, mpg ~ .),trace=0,steps=10000)
summary(stepmodel)
```

```
##
## Call:
## lm(formula = mpg ~ wt + qsec + am, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.4811 -1.5555 -0.7257  1.4110  4.6610
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.6178     6.9596   1.382 0.177915
## wt           -3.9165     0.7112  -5.507 6.95e-06 ***
## qsec          1.2259     0.2887   4.247 0.000216 ***
## amManual      2.9358     1.4109   2.081 0.046716 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.459 on 28 degrees of freedom
## Multiple R-squared:  0.8497, Adjusted R-squared:  0.8336
## F-statistic: 52.75 on 3 and 28 DF,  p-value: 1.21e-11
```

At this point we have a model, which includes 3 variables

wt qsec am This model has a 0.85 of total variance. To further optimize the model, we can examine mpg ~ wt + qsec correlation with am.

```
model <- lm(mpg~ factor(am):wt + factor(am):qsec,data=mtcars)
summary(model)
```

```
##
```

```
## Call:
## lm(formula = mpg ~ factor(am):wt + factor(am):qsec, data = mtcars)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.9361 -1.4017 -0.1551  1.2695  3.8862
##
## Coefficients:
##                          Estimate Std. Error t value Pr(>|t|)
## (Intercept)               13.9692     5.7756   2.419  0.02259 *
## factor(am)Automatic:wt    -3.1759     0.6362  -4.992 3.11e-05 ***
## factor(am)Manual:wt       -6.0992     0.9685  -6.297 9.70e-07 ***
## factor(am)Automatic:qsec   0.8338     0.2602   3.205  0.00346 **
## factor(am)Manual:qsec      1.4464     0.2692   5.373 1.12e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.097 on 27 degrees of freedom
## Multiple R-squared:  0.8946, Adjusted R-squared:  0.879
## F-statistic: 57.28 on 4 and 27 DF,  p-value: 8.424e-13
```

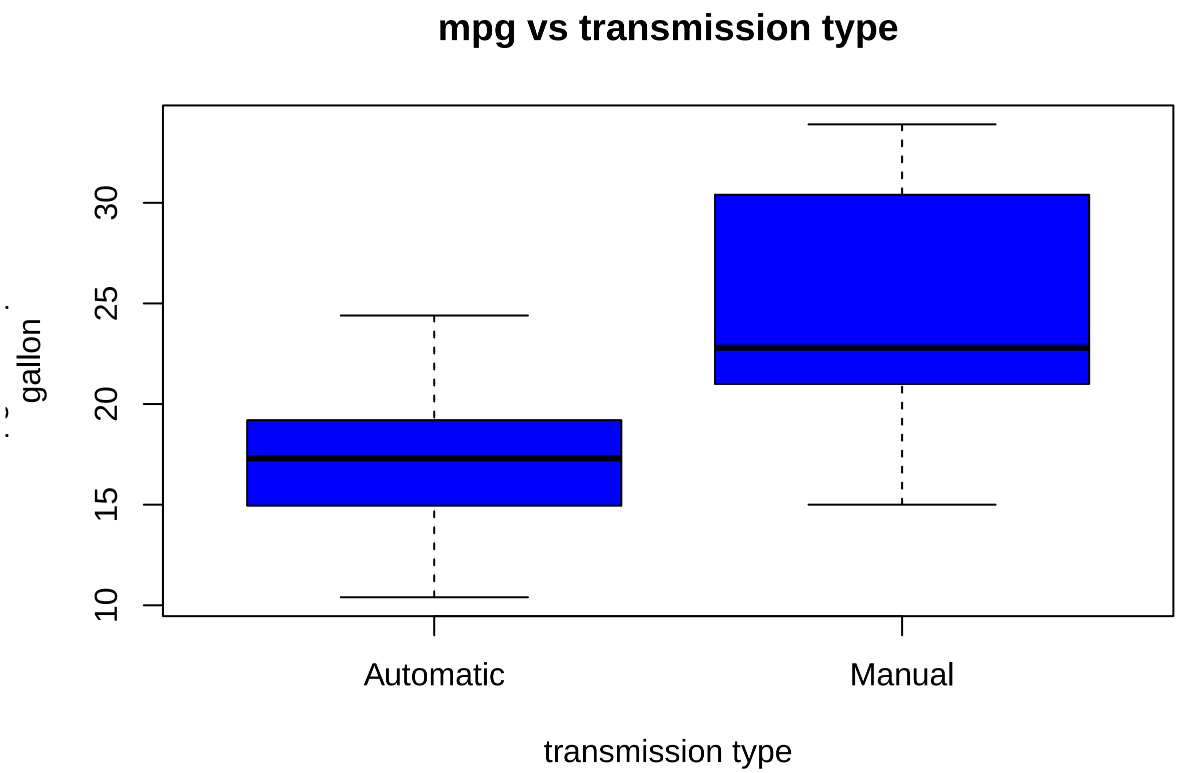
4. Summary Interpreting the results, we can see this model has a 89.5% total variance with an adjusted variance of 0.879. By adding the coefficients, we have the following conclusions:

when the weight increased by 1000 lbs, the mpg decreased by -3.176 for automatic transmission cars, and -6.09 for manual transmission cars so with increasing car weight we should choose manual transmission cars when the acceleration speed dropped, and 1/4 mile time increased (by 1 sec), the mpg factor increased by 0.834 miles for automatic transmission cars, and 1.446 miles for manual transmission cars so with lower acceleration speed, but same weight, manual transmission cars are better for mpg

Main conclusion The mpg is largely determined by the interplay between weight, acceleration and transmission. Given the above analysis, the original question (automatic transmission vs manual transmission) is not really answered, and should be considered in the context of weight and acceleration speed.

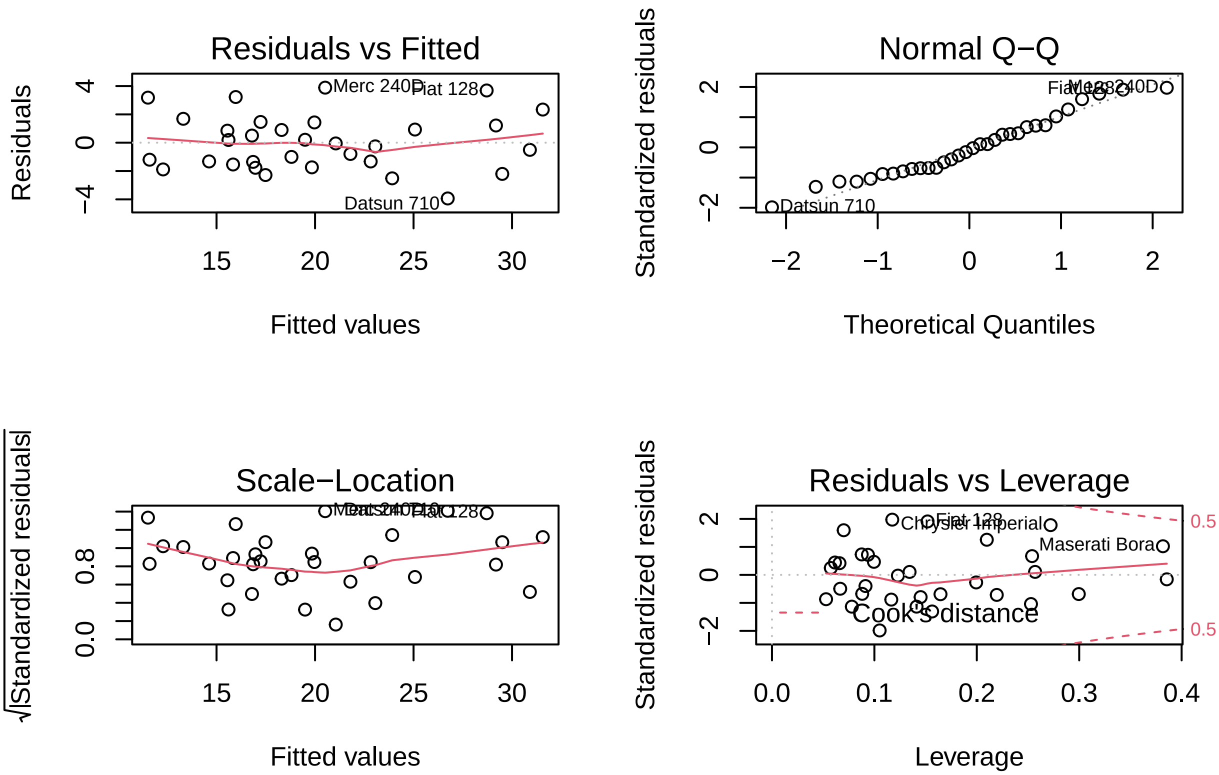
Appendix Appendix 1. Boxplot of mpg vs transmission type

```
boxplot(mtcars$mpg ~ mtcars$am, data = mtcars, outpch = 19, ylab="mpg:miles per
gallon",xlab="transmission type",main="mpg vs transmission type", col="blue")
```



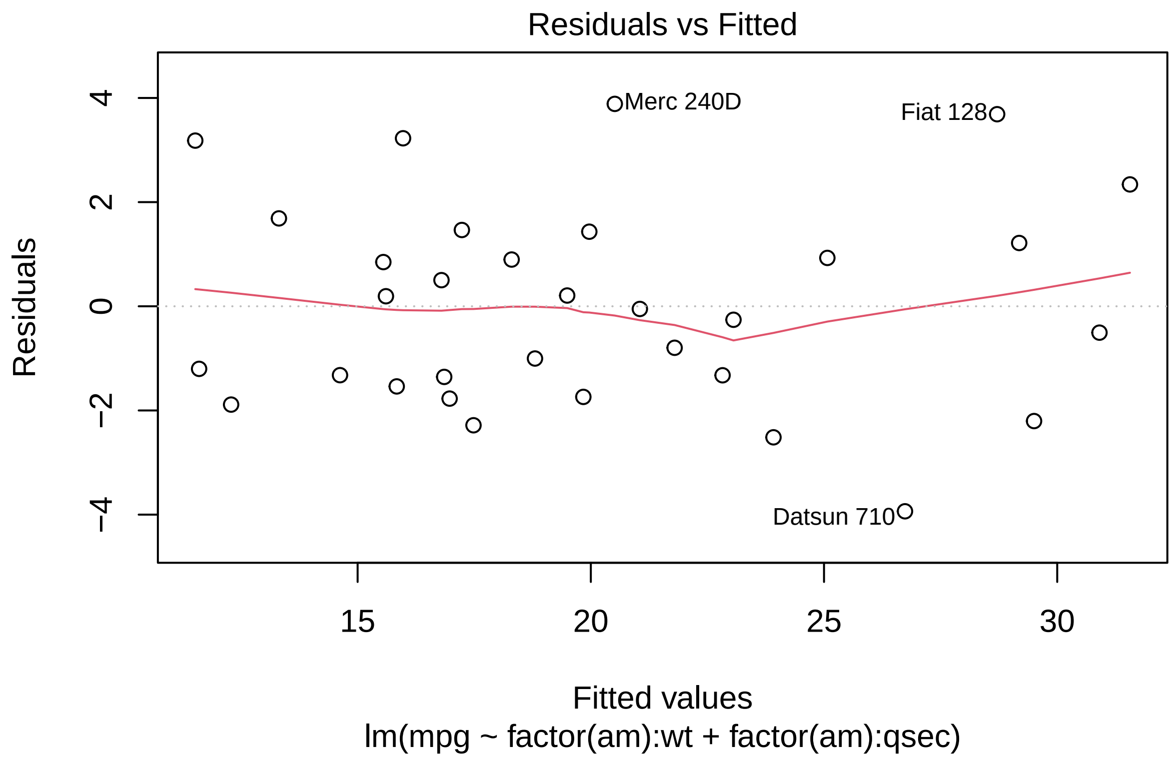
Appendix 2. Residual check and diagnostics plot

```
par(mfrow=c(2,2))
plot(model)
```

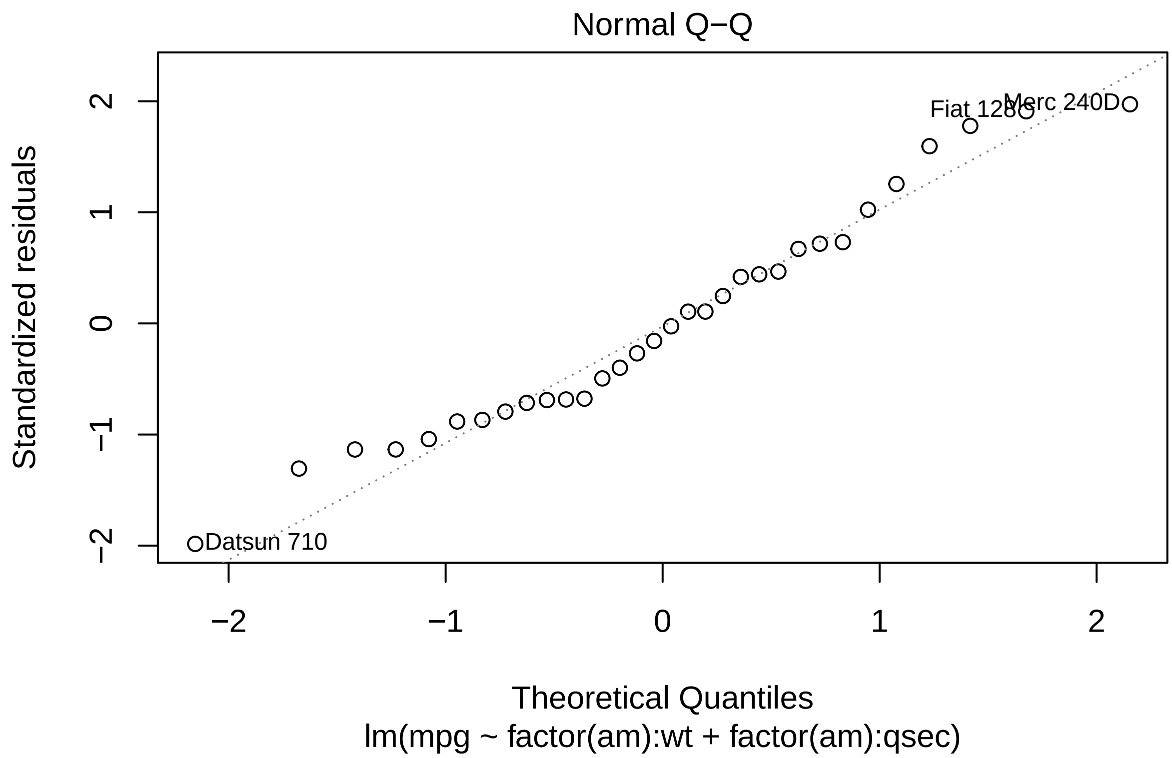


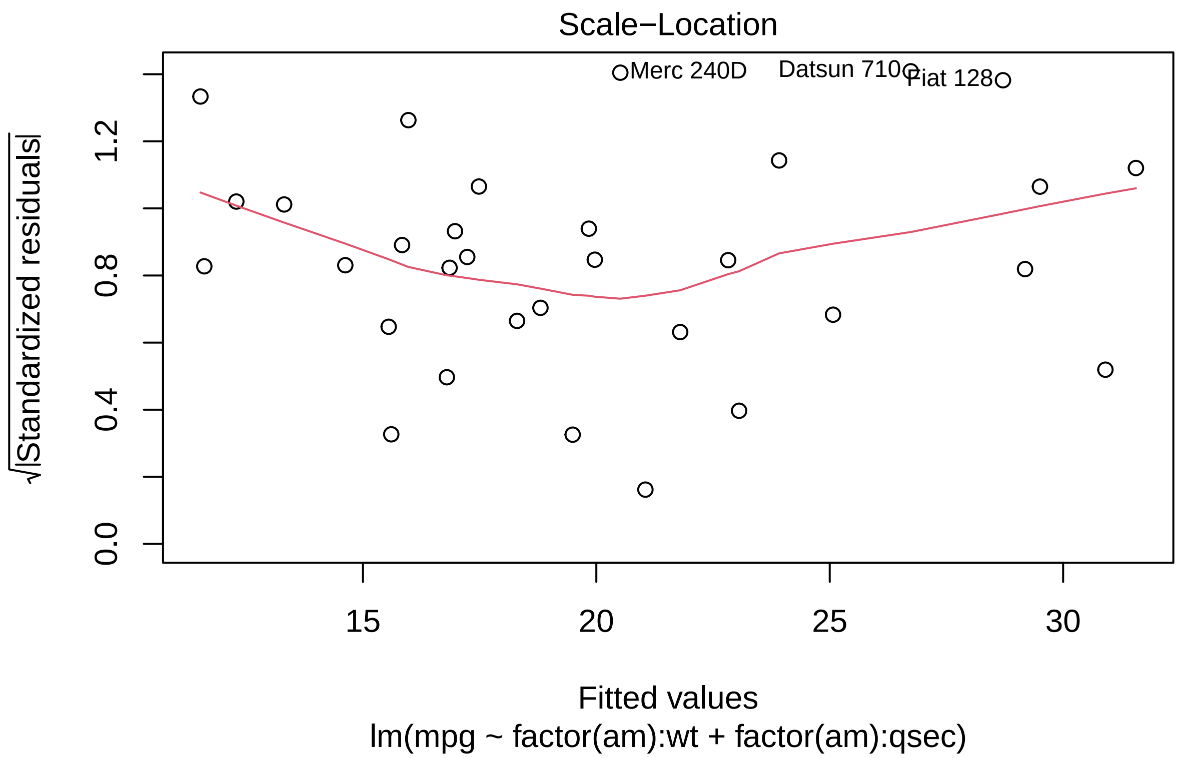
Appendix 3. Further plots

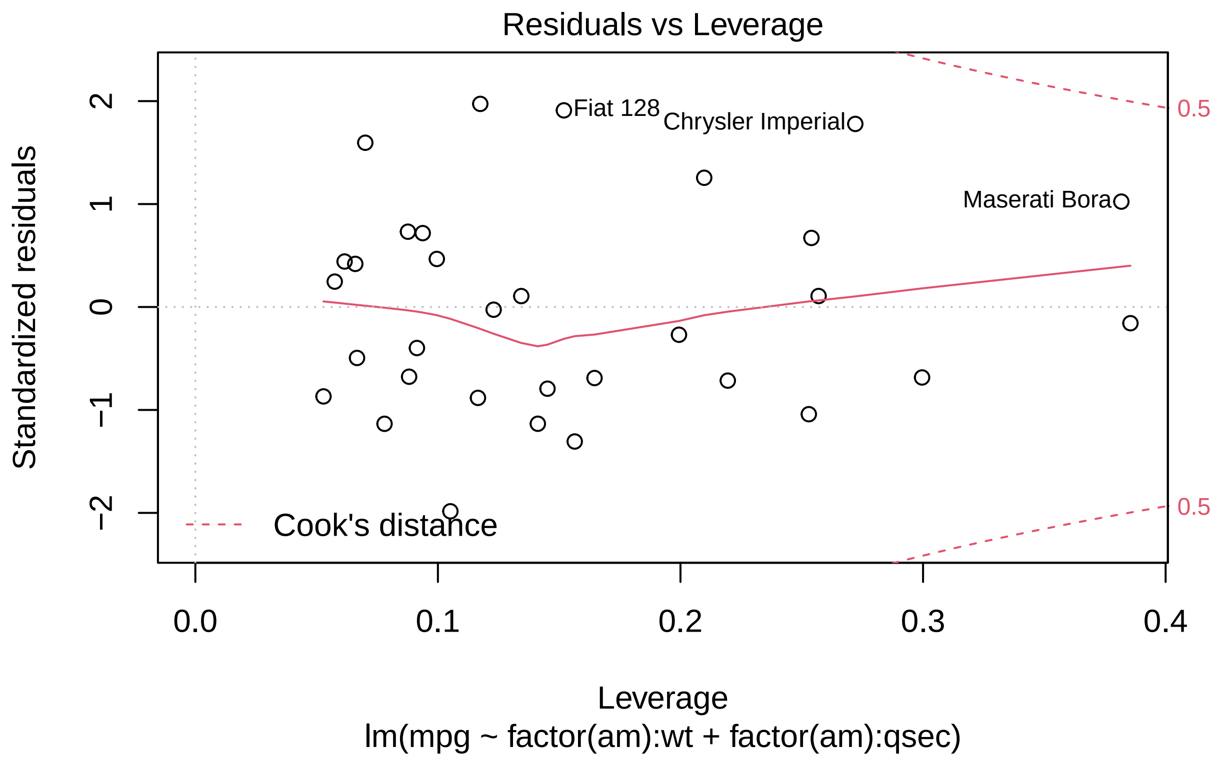
```
plot(model)
```



Residuals vs Fitted
Merc 240D
Fiat 128
Datsun 710
4
2
0
−2
−4
Residuals
15
20
25
30
Fitted values
lm(mpg ~ factor(am):wt + factor(am):qsec)

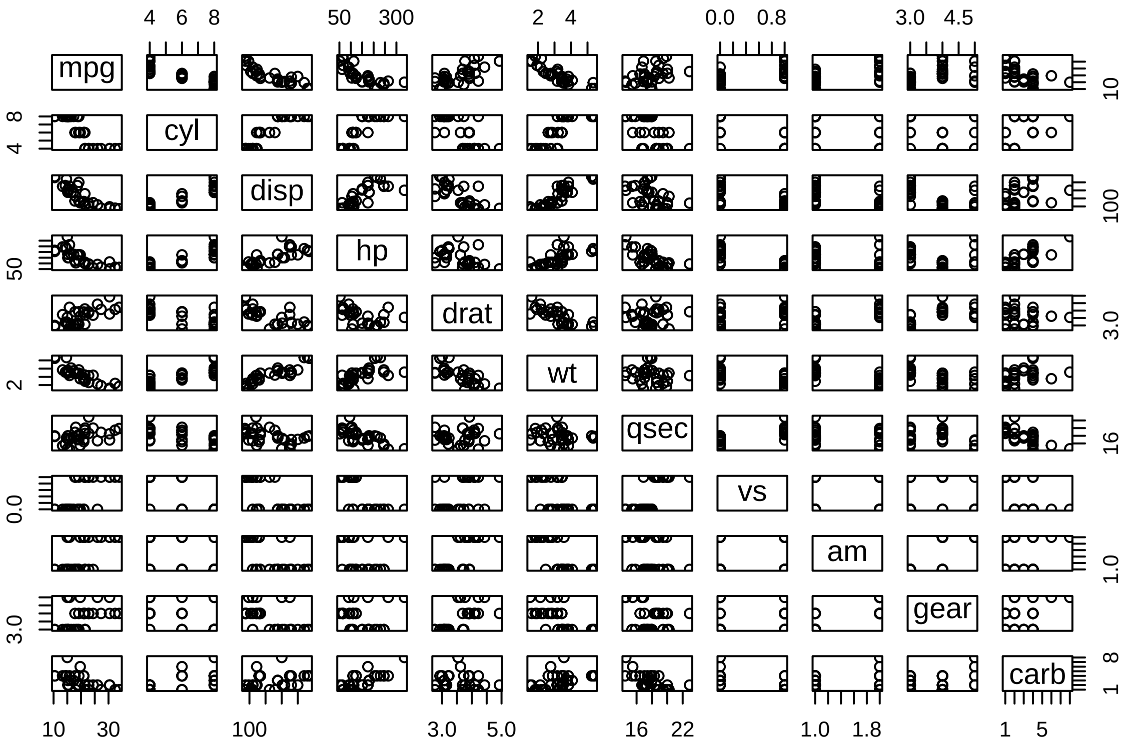

Normal Q−Q
Standardized residuals
Theoretical Quantiles
lm(mpg ~ factor(am):wt + factor(am):qsec)
Datsun 710
Fiat 128
Merc 240D
−2
−1
0
1
2


Scale−Location
Merc 240D
Datsun 710
Fiat 128
1.2
0.8
0.4
0.0
√|Standardized residuals|
15
20
25
30
Fitted values
lm(mpg ~ factor(am):wt + factor(am):qsec)



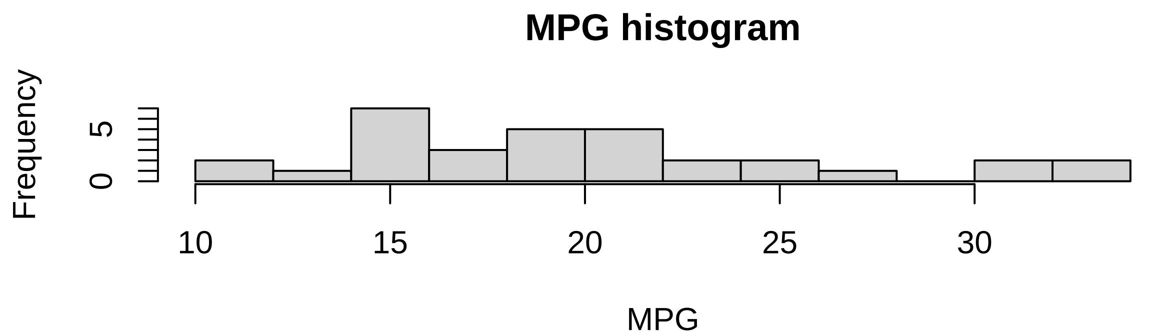
Appendix 4. Scatterplots

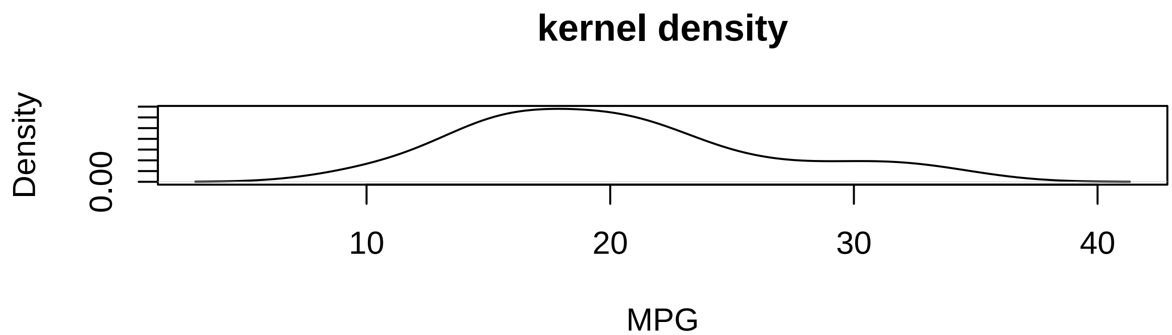
```
pairs(mpg ~ ., data = mtcars)
```



Appendix 5. Density and histogram

```
par(mfrow=c(2,1))
hist(mtcars$mpg, breaks=10, xlab="MPG", main="MPG histogram")
plot(density(mtcars$mpg), main="kernel density", xlab="MPG")
```


MPG histogram
Frequency
0
5
10
15
20
25
30
MPG


kernel density
Density
0.00
10
20
30
40
MPG