Statistics in R

Sadaf Zuhra

2024-10-01

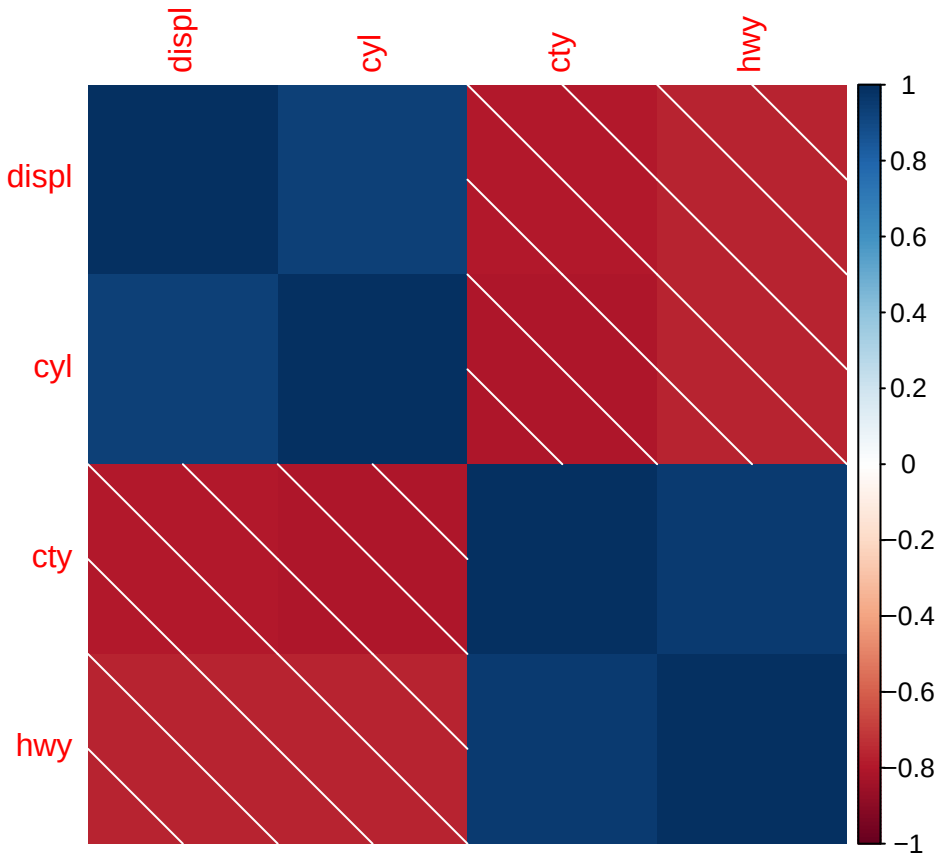
relationship between Data Variables

Correlation and Correlation Matrix

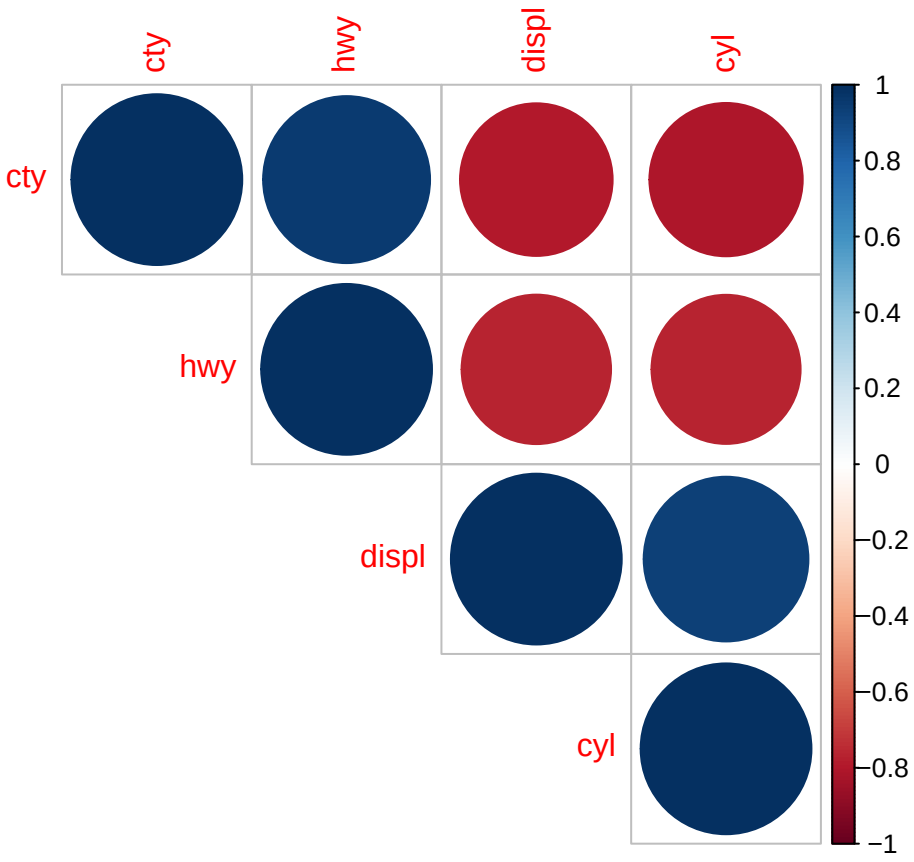
```
library(ggplot2)
library(tidyverse)
library(corrplot)

mpg <- mpg


corr<- mpg %>%
   select(., displ, cyl, cty, hwy) %>%
    cor()
```

```
library(corrplot)

corrplot(corr, method = "number", type = "full")
```

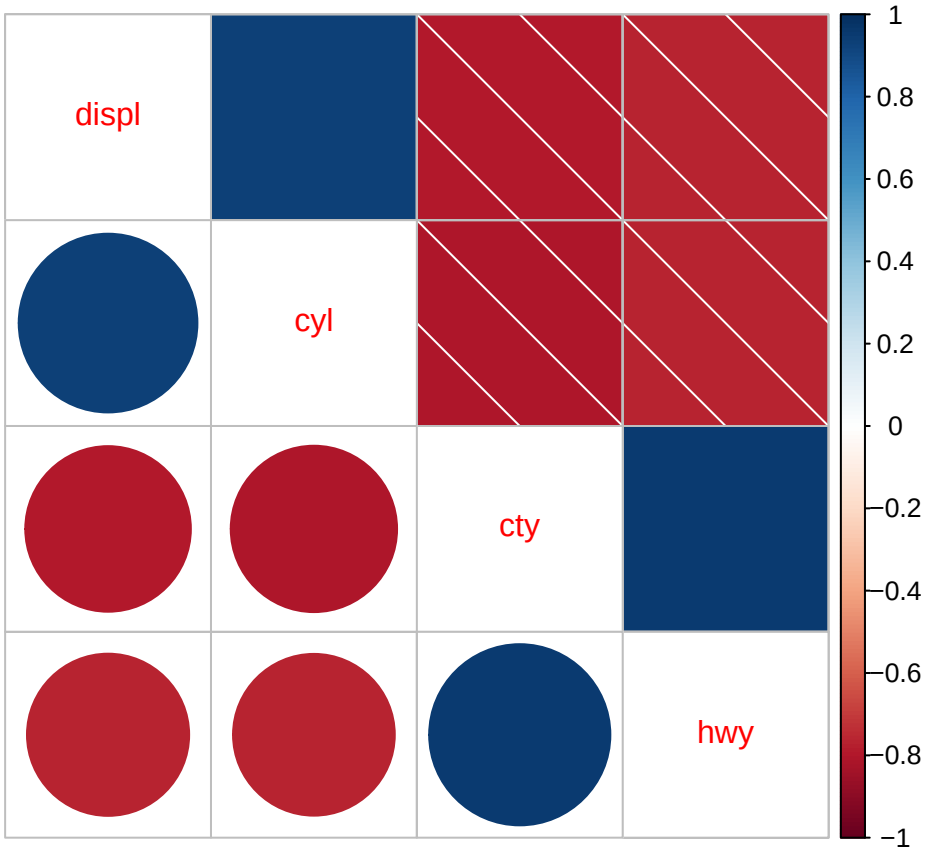

```
corrplot(corr, method = "shade", type = "full")
```




```
corrplot(corr, method = "circle", type = "upper", order = "hclust", addrect = TRUE )
```




```
corrplot.mixed(corr, lower = "circle", upper = "shade")
```

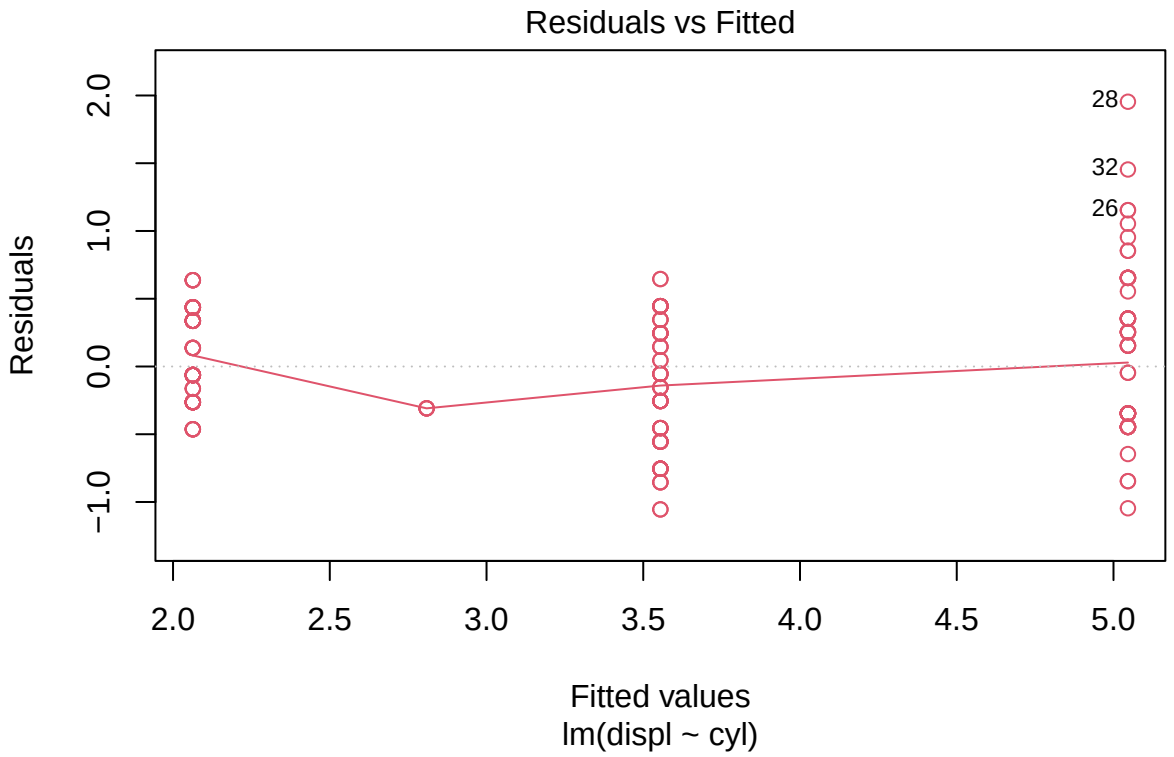


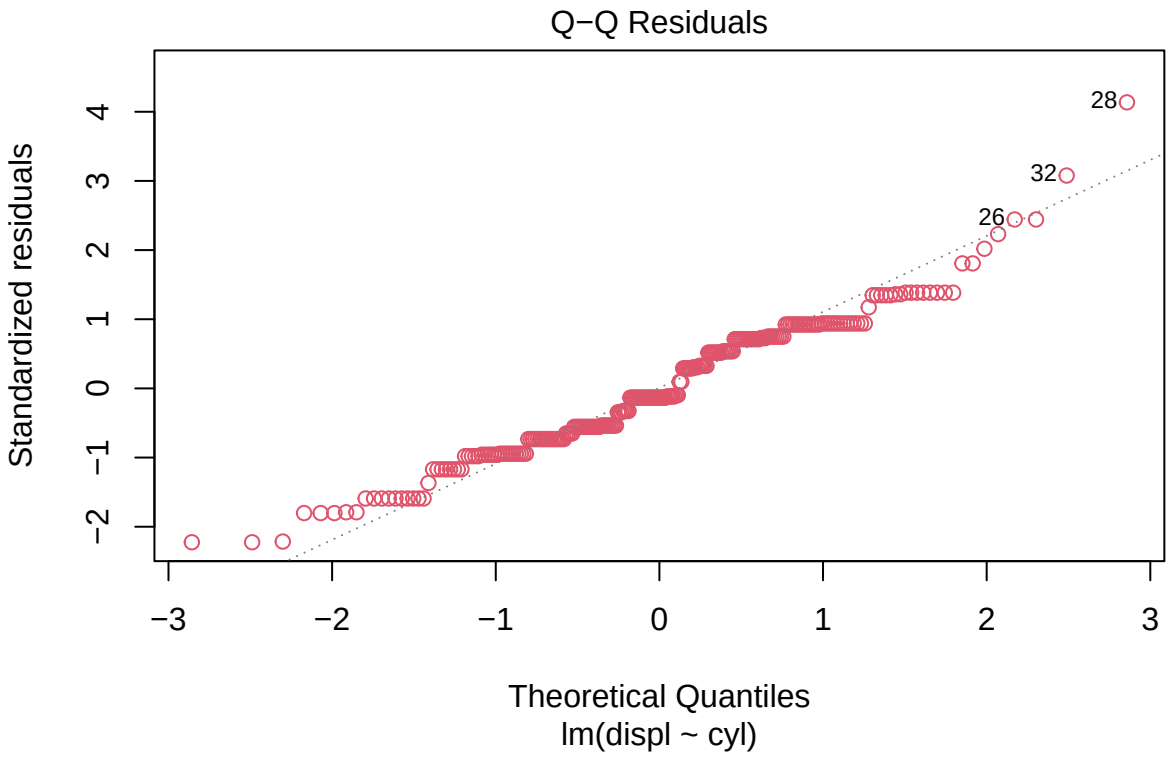
regression

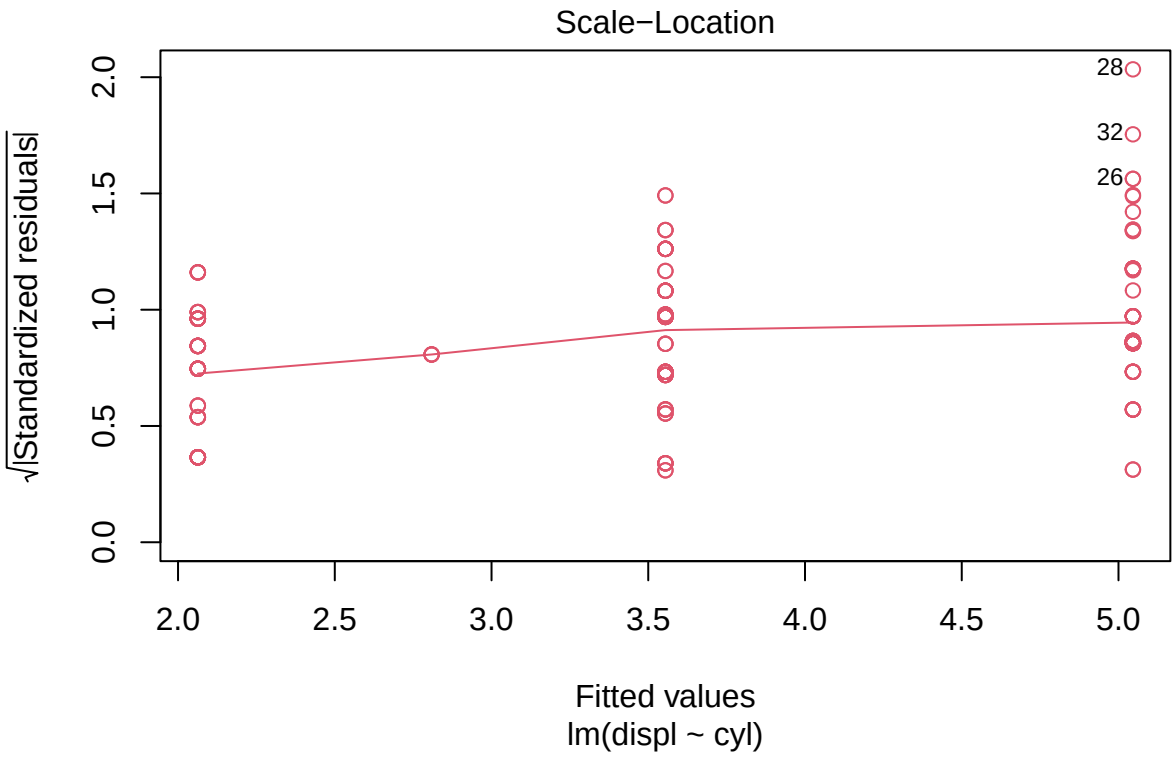
```
mod <- mpg %>%  lm( formula = displ ~ cyl)

summary(mod)
```

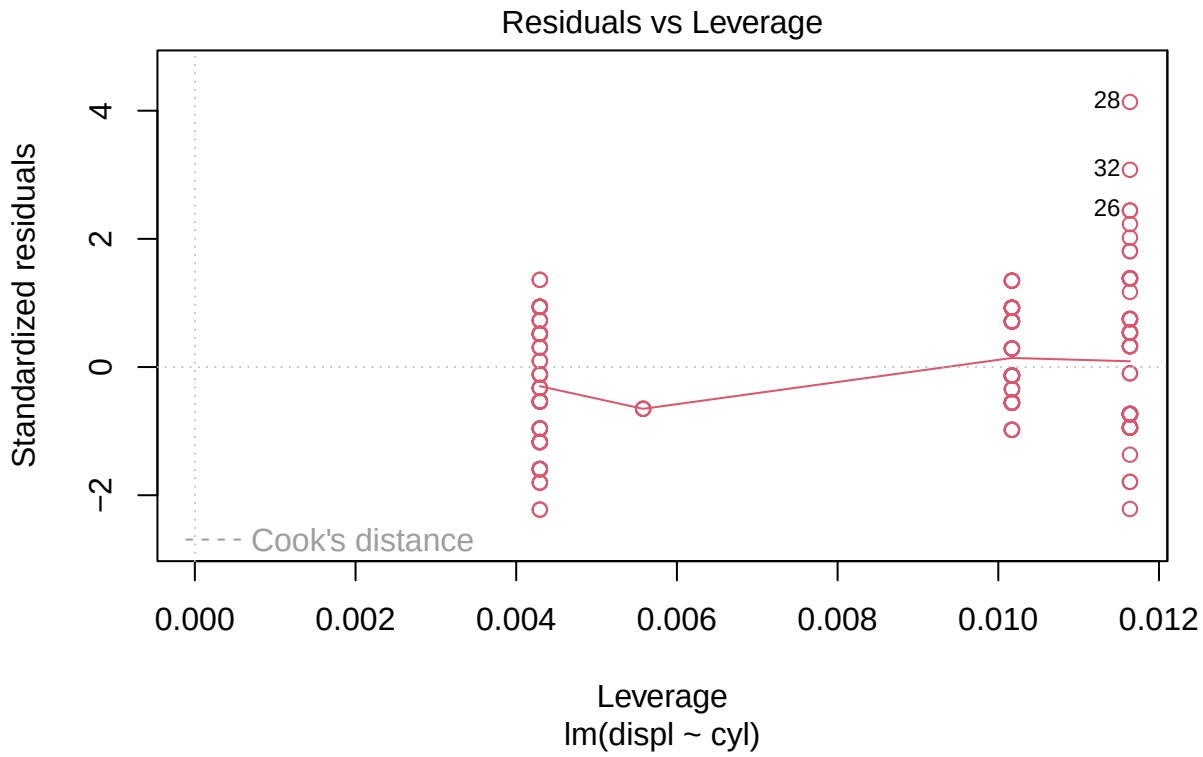
```
##
## Call:
## lm(formula = displ ~ cyl, data = .)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.05466 -0.34617 -0.06314  0.35383  1.95383
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.91989    0.11791  -7.801 2.07e-13 ***
## cyl          0.74576    0.01932  38.609  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4751 on 232 degrees of freedom
## Multiple R-squared:  0.8653, Adjusted R-squared:  0.8647
## F-statistic:  1491 on 1 and 232 DF,  p-value: < 2.2e-16
```

```
plot(mod, col=2, lwd=1)
```



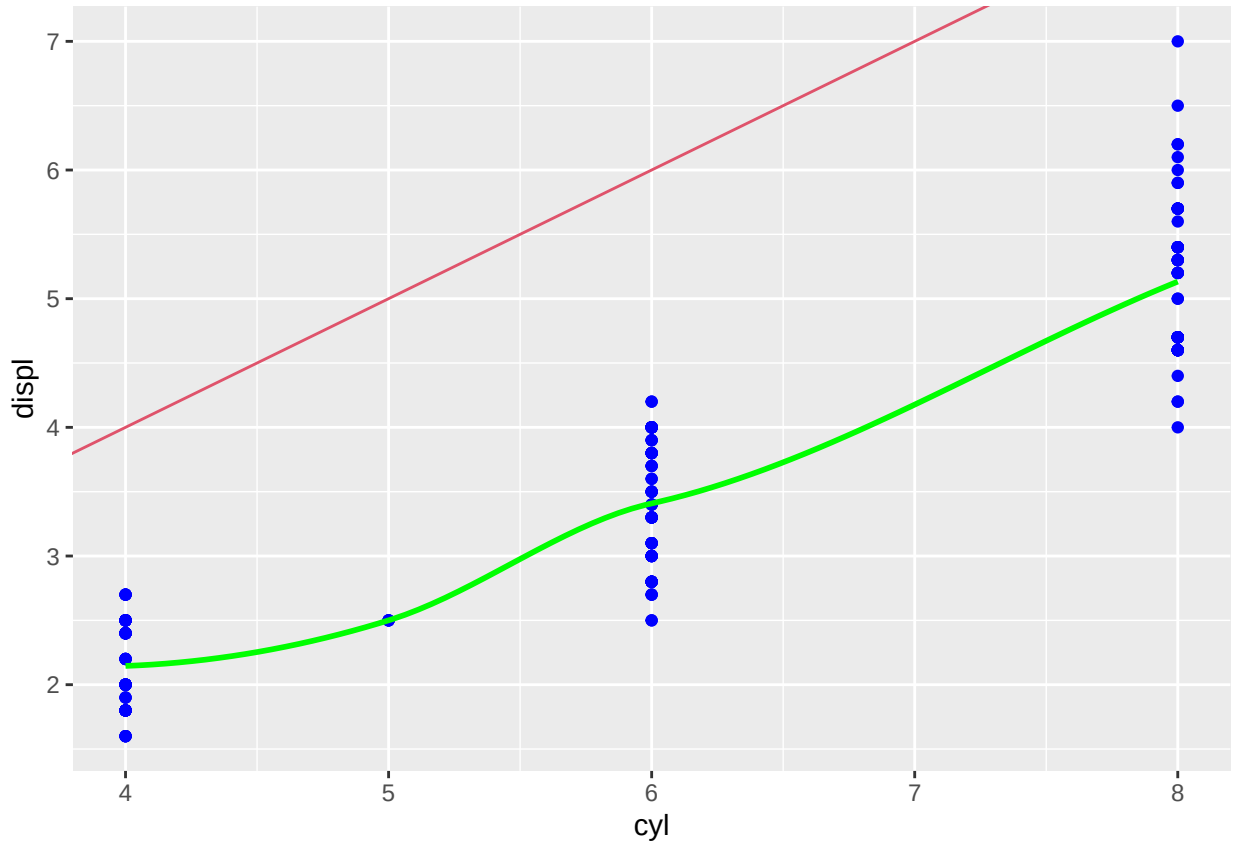

Q−Q Residuals
Standardized residuals
−2
−1
0
1
2
3
4
28
32
26
−3
−2
−1
0
1
2
3
Theoretical Quantiles
lm(displ ~ cyl)


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
28
32
26
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Fitted values
lm(displ ~ cyl)

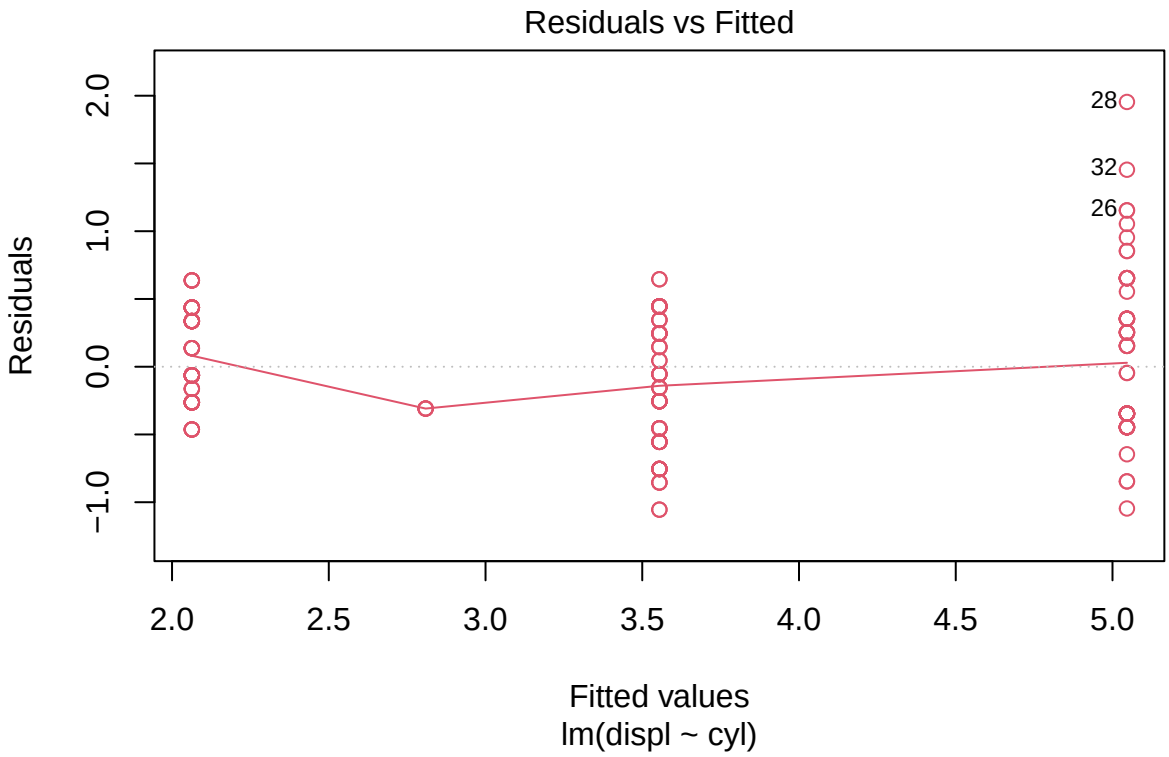


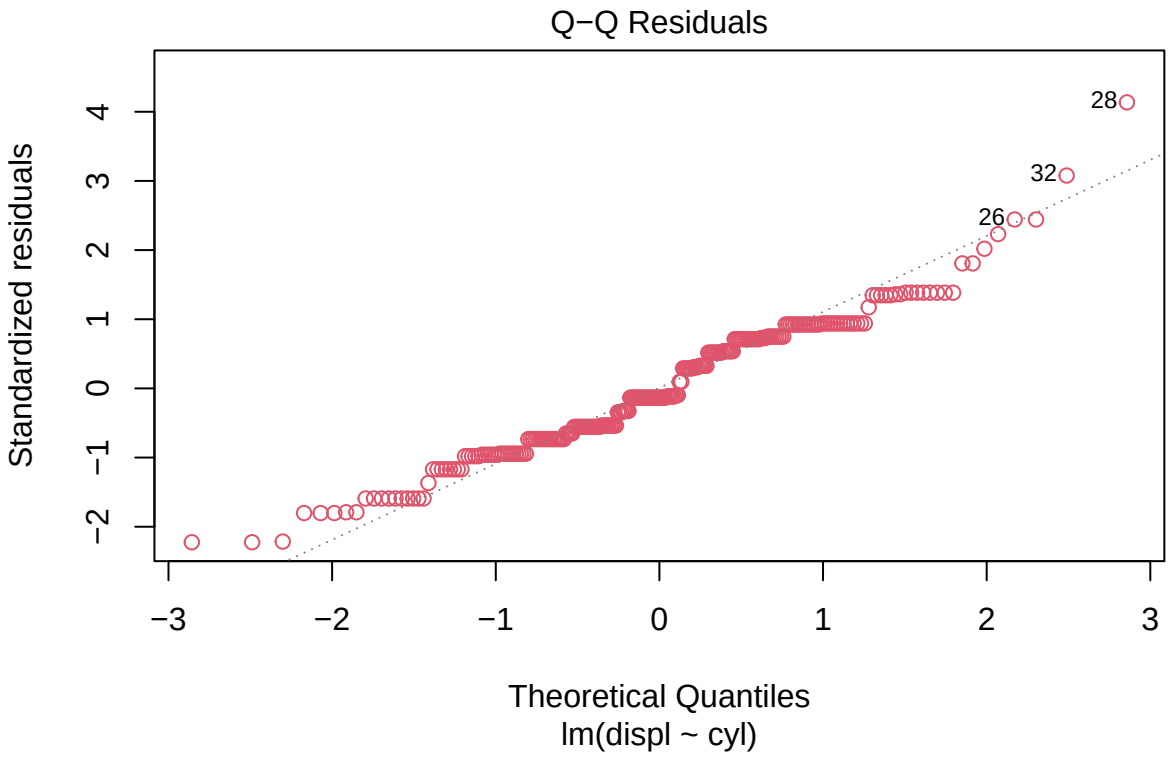

```
par(mfrow = c(2,2))
```

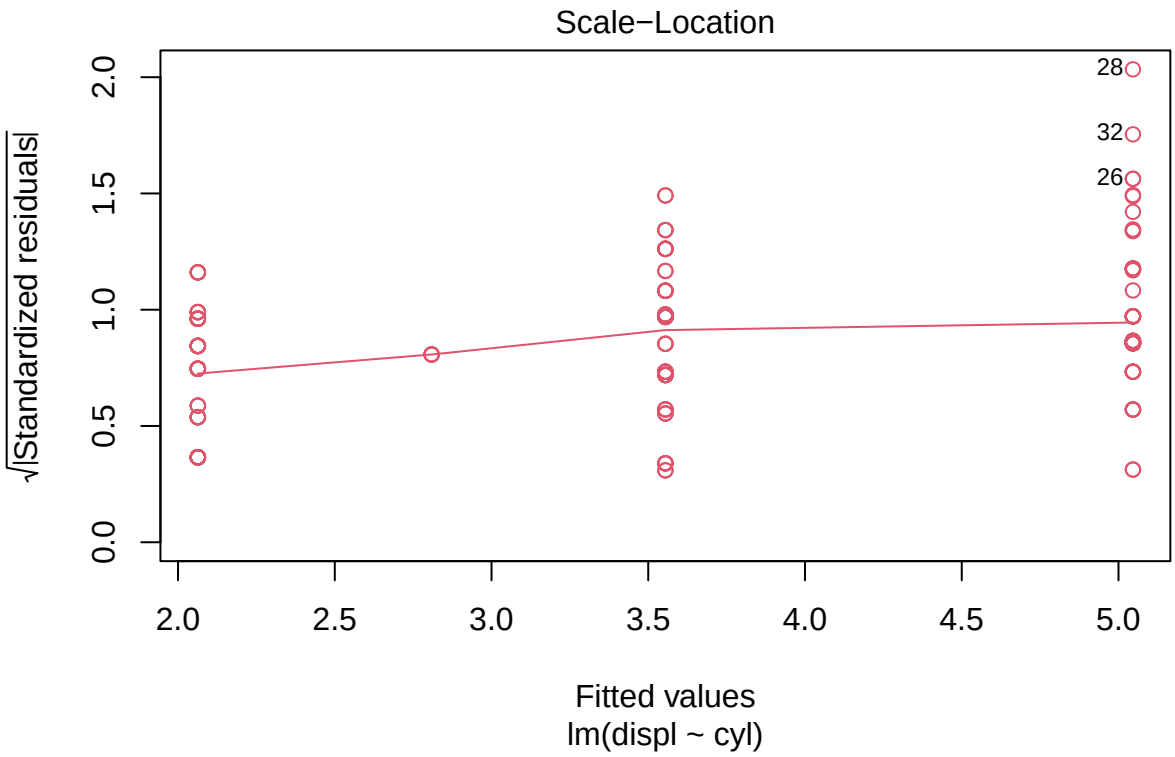
```
ggplot(data = mpg, aes(x=cyl, y=displ))+geom_abline(col=2)+

  geom_point(col="blue")+geom_smooth(se=FALSE, col="green")
```

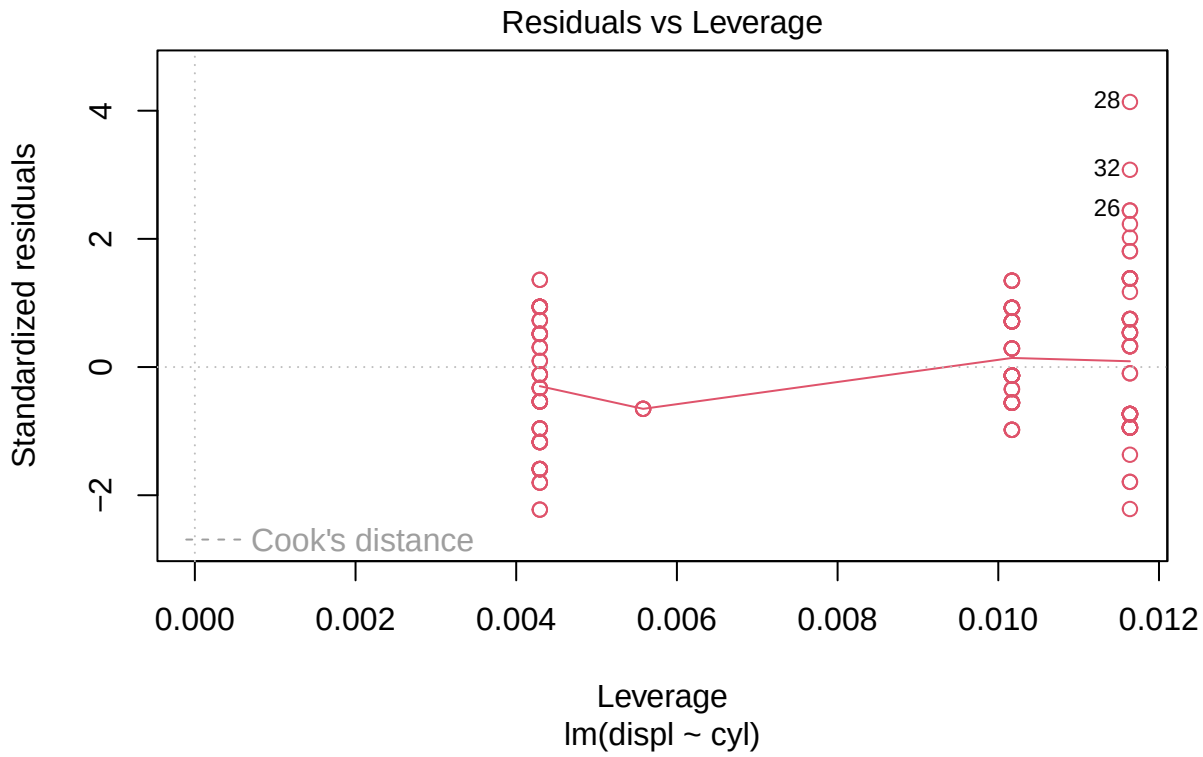



```
plot(mod, col=2, lwd=1)
```

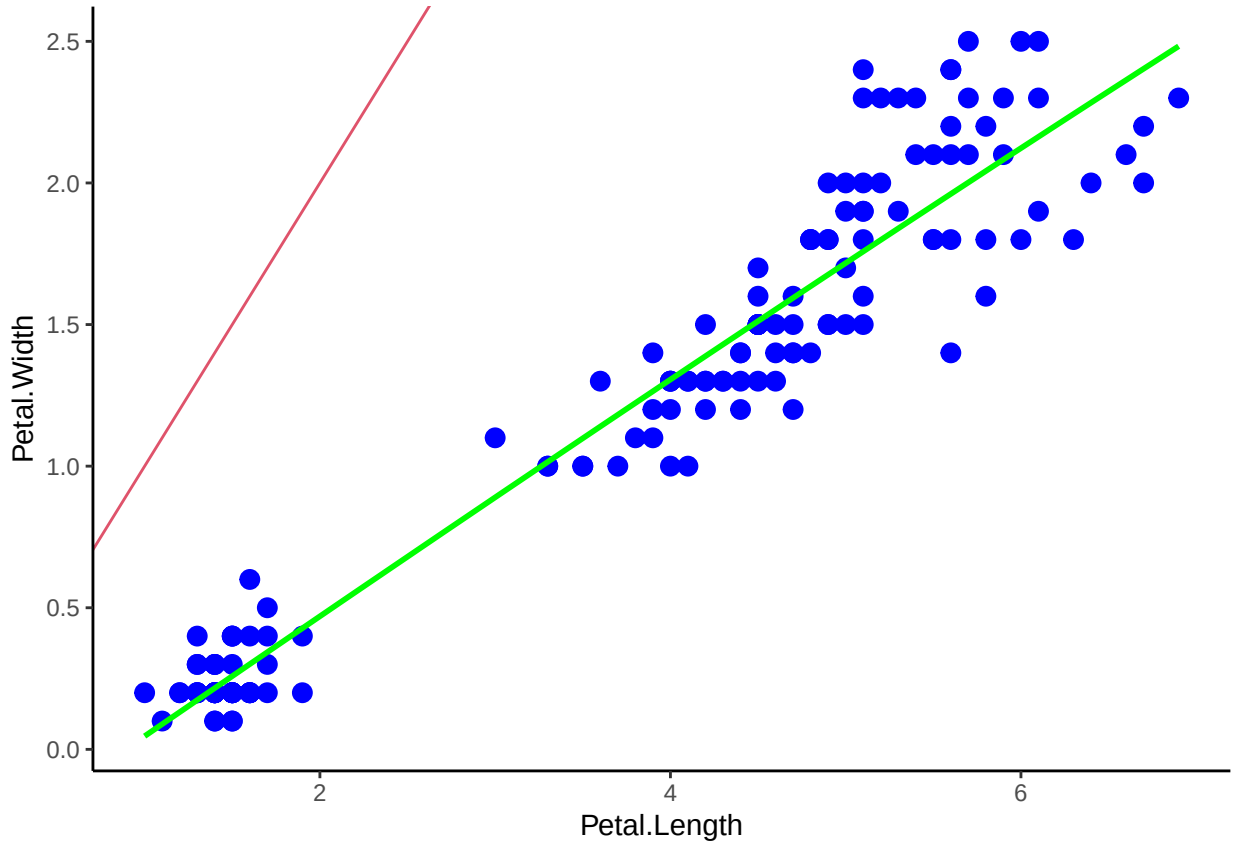

Residuals vs Fitted
Residuals
2.0
1.0
0.0
−1.0
28
32
26
2.0
2.5
3.0
3.5
4.0
4.5
5.0
Fitted values
lm(displ ~ cyl)


Q−Q Residuals
Standardized residuals
−2
−1
0
1
2
3
4
28
32
26
−3
−2
−1
0
1
2
3
Theoretical Quantiles
lm(displ ~ cyl)


Scale−Location
√|Standardized residuals|
0.0
0.5
1.0
1.5
2.0
2.0
2.5
3.0
3.5
4.0
4.5
5.0
28
32
26
Fitted values
lm(displ ~ cyl)




```
ggplot(data=iris, aes(x=Petal.Length, y=Petal.Width)) +

  geom_abline(col=2) + geom_point(size=3, col="blue") +

  geom_smooth(se=FALSE, col="green", span=10) + theme_classic()
```



comparison between Data Variables

T-test in R

One sample T-test

```
# mpg <-mpg

t.test(mpg$hwy , mu =23.4 )
```

```
##
##  One Sample t-test
##
## data:  mpg$hwy
## t = 0.1032, df = 233, p-value = 0.9179
## alternative hypothesis: true mean is not equal to 23.4
## 95 percent confidence interval:
##  22.67324 24.20710
## sample estimates:
## mean of x
##  23.44017
```

Two sample T-test

```
t.test(mpg$cty , mpg$hwy)
```

unpaired/ independent T-test

```
##
##  Welch Two Sample t-test
##
## data:  mpg$cty and mpg$hwy
## t = -13.755, df = 421.79, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -7.521683 -5.640710
## sample estimates:
## mean of x mean of y
##  16.85897  23.44017
```

```
pre_treatment <- rnorm(n= 2000 ,mean= 160 ,sd = 15)
post_treatment <- rnorm(n=2000 , mean = 120 , sd = 15)
```

```
t.test(pre_treatment ,post_treatment , paired = TRUE)
```

paired T-test

```
##
##  Paired t-test
##
## data:  pre_treatment and post_treatment
## t = 84.668, df = 1999, p-value < 2.2e-16
## alternative hypothesis: true mean difference is not equal to 0
## 95 percent confidence interval:
##  38.95700 40.80451
## sample estimates:
## mean difference
##        39.88076
```

use of if-else command

```
if( is.numeric(mpg$hwy))

{
  print("highway is a numeric column")
}
```

```
## [1] "highway is a numeric column"
```

```
if( is.numeric(mpg$manufacturer))

{
  print("highway is a numeric column")
}else  {
  print("highway is not a numeric column ")
}
```

```
## [1] "highway is not a numeric column "
```