ISL2 Ch2.3 HW1 Submission

2024-09-26

```
knitr::opts_chunk$set(echo = TRUE) # sets default behavior to include code in the HTML file, please kee

rm(list = ls()) # clears the environment of any existing objects

suppressPackageStartupMessages({ # reduces annoying printing
  library(janitor) # has helpful cleaning functions
  library(knitr) # knitting functions (transforming RMarkdown file to HTML)
  library(lubridate) # date parsing functions
  library(scales) # library for formatting ggplot axes
  library(tidyverse) # workhorse dplyr package
})

options(dplyr.summarise.inform = FALSE) # turns off an annoying dplyr behavior

set.seed(487) # setting the seed makes random operations reproducible
```

2.3.1 Basic Commands

```
x <- c(1,3,2,5) # creates a vector of 1,3,2,5
x               # calls and display vector x
```

```
## [1] 1 3 2 5
```

```
x <-c(1,6,2)
y = c(1,4,3)

length(x)       # display length of x
```

```
## [1] 3
```

```
length(y)
```

```
## [1] 3
```

```
x+y             # calls and display of the vector addition of x and y
```

```
## [1]  2 10  5
```

```r
ls()           # display the list of all objects
```

```
## [1] "x" "y"
```

```r
rm(x,y)       # remove x and y objects

rm(list = ls()) # remove all objects in the list


x <- matrix(c(1,2,3,4),2,2) # creates a matrix x with 2 col and 2 row of [1,2],[3,4] (filled by column

x <- matrix(c(1,2,3,4),2,2, byrow = TRUE) # creates a matrix x with 2 col and 2 row of [1,3],[2,4] fill

sqrt(x) # takes the square root of each element of the vector/matrix
```

```
##          [,1]     [,2]
## [1,] 1.000000 1.414214
## [2,] 1.732051 2.000000
```

```r
x^2 # squares each element of the vector/matrix
```

```
##      [,1] [,2]
## [1,]    1    4
## [2,]    9   16
```

```r
x <- rnorm(50) # generates a vector of random normal variables with n = 50
y<- x+ rnorm(50, mean = 50, sd = .1) # add x vector to the another random normal vector with standard d
cor(x,y) # calculate the correlation between x and y
```

```
## [1] 0.9953846
```

```r
set.seed(3) # generate a reproducible random set of numbers with code 3 (number itself is arbitrary) fo
y <- rnorm(100)

mean(y) # display mean of y
```

```
## [1] 0.01103557
```

```r
var(y) # display variance of y
```

```
## [1] 0.7328675
```

```r
sqrt(var(y)) #display square root of variance of y
```

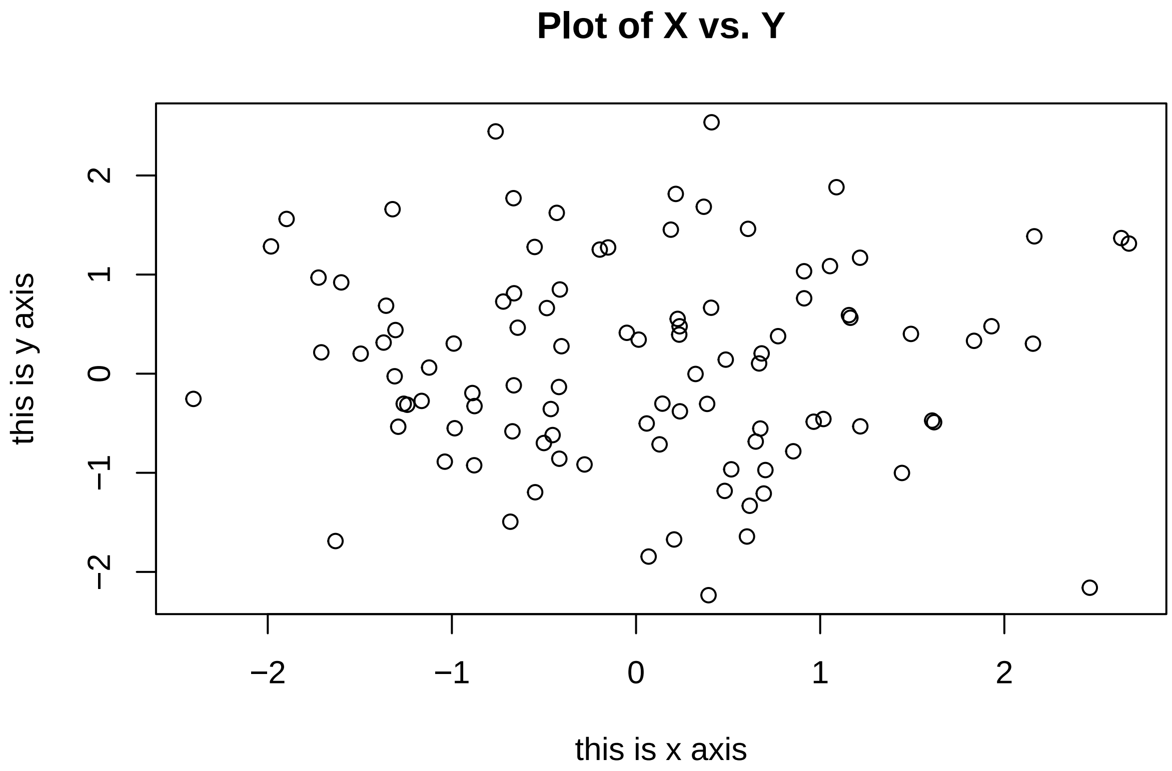
```
## [1] 0.8560768
```

```
sd(y) #display standard deviation of y
```

```
## [1] 0.8560768
```

2.3.2 Graphics

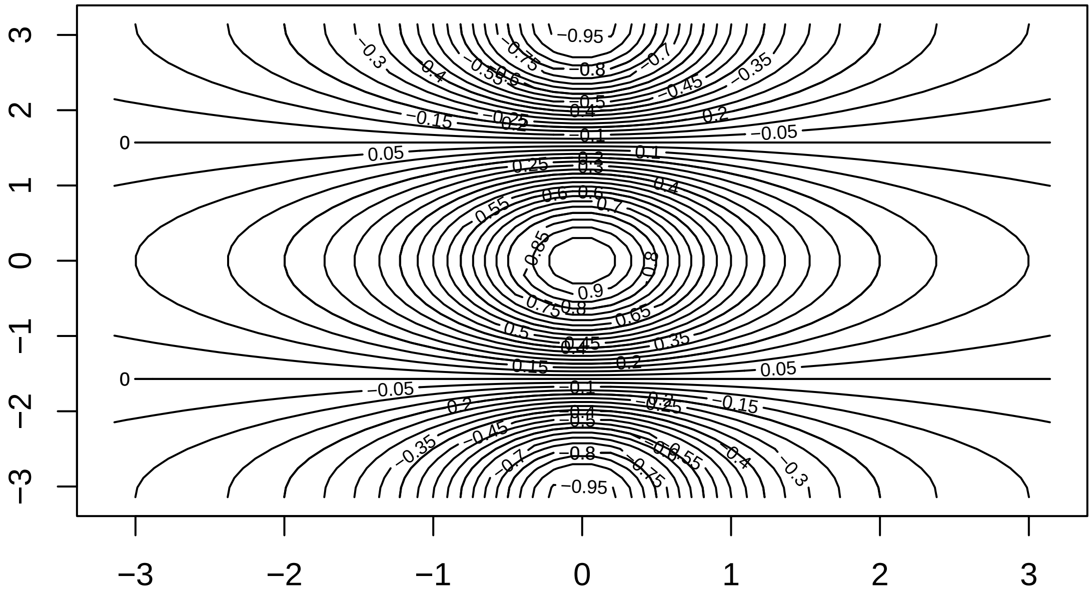
```
x <- rnorm(100)
y <- rnorm(100)

plot(x,y, xlab = 'this is x axis',
     ylab = 'this is y axis',
     main = 'Plot of X vs. Y') # creates a plot based on normal random x and y vectors with x, y labels
```



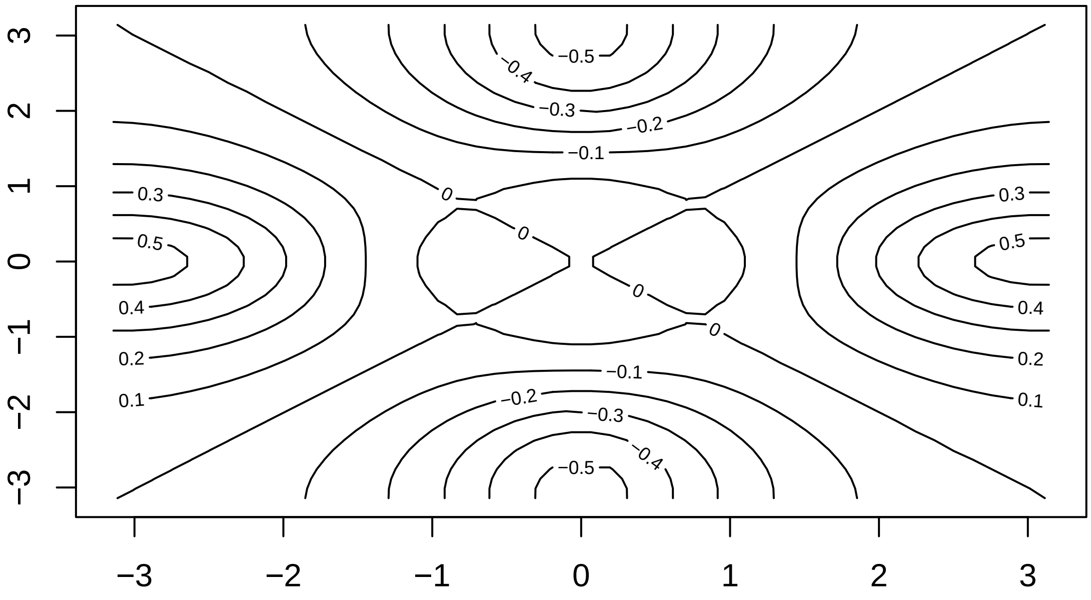
```
pdf('Figure.pdf') # creates pdf (for jpeg, use jpeg('figure name'))
plot(x,y,col = 'green')
dev.off() # indicates a finish operation, which in our case is the plotting
```

```
## pdf
##   2
```

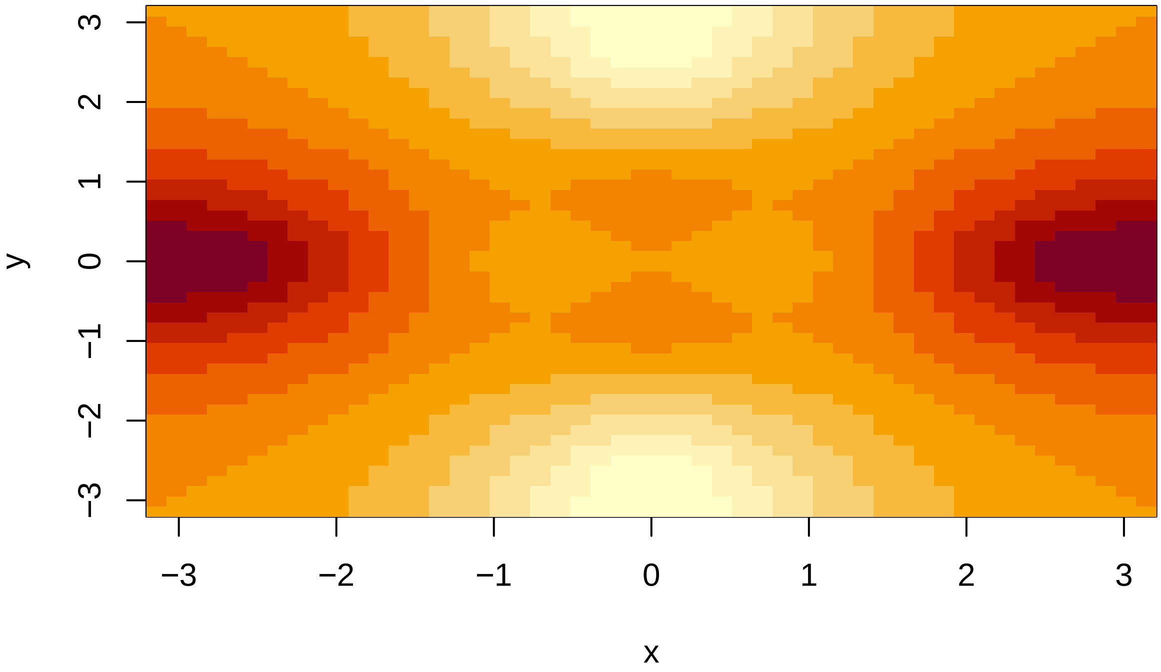
```
x <- seq(1,10) # creates a sequence of integer number from 1 to 10 (better alternative: seq(1,10, lengt
x <- 1: 10
x<- seq(-pi, pi, length = 50)

y<-x
f <- outer(x,y, function(x,y) cos(y)/(1+x^2)) # outer products of the x and y arrays operated with cos(
contour(x,y,f)
contour(x,y,f, nlevels = 45, add = T) # creates a contour plot of f on x,y axis with 45 levels of conto
```



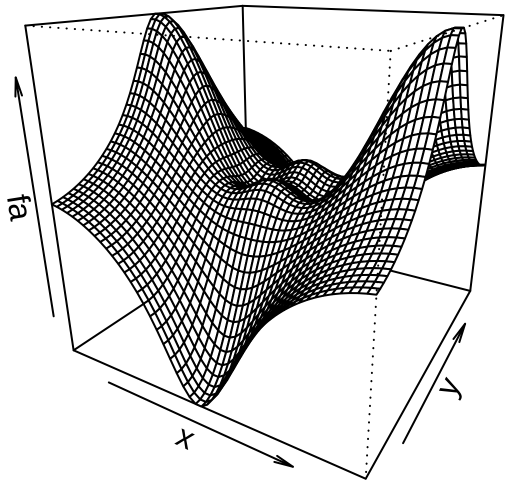
```
fa <- (f - t(f)) /2 # set a function products of f subtract to its transposed version and divide the fu
contour(x,y,fa, nlevel = 15)
```



```r
image(x,y,fa) # creates a color-coded plot (heatmap) dependent on z value to its color
```



```
persp(x,y,fa, theta = 30, phi = 20) # creates a 3-d plot with theta and phi to augment to angle of the
```



2.3.3 Indexing Data

```
A <- matrix(1:16,4,4)
A
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    1    5    9   13
## [2,]    2    6   10   14
## [3,]    3    7   11   15
## [4,]    4    8   12   16
```

```
A[2,3] #access the element on the second row, third column
```

```
## [1] 10
```

```
A[c(1,3), c(2,4)] # access and create a matrix based on the row 2 (limited by 1,3) and column 3 (limiet
```

```
##      [,1] [,2]
## [1,]    5   13
## [2,]    7   15
```

```r
A[1:3, 2:4] #access and create matrix with row 1-3 and 2-4 in column elements
```

```
##      [,1] [,2] [,3]
## [1,]    5    9   13
## [2,]    6   10   14
## [3,]    7   11   15
```

```r
A[1:2, ] #access and create matrix with row 1-2 and all the column elements
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    1    5    9   13
## [2,]    2    6   10   14
```

```r
A[, 1:2] #access and create matrix with column 1-2 and all the row elements
```

```
##      [,1] [,2]
## [1,]    1    5
## [2,]    2    6
## [3,]    3    7
## [4,]    4    8
```

```r
A[-c(1,3), ] # access and create matrix with row 1, 3 excluded and of all column
```

```
##      [,1] [,2] [,3] [,4]
## [1,]    2    6   10   14
## [2,]    4    8   12   16
```

```r
dim(A) # returns the dimension information of the matrix
```

```
## [1] 4 4
```

2.3.4 Loading Data

```r
Auto <- read.table("Auto.data") # read Auto text file as table
head(Auto) # Provide the first five entry of the data, View(Auto) # View Auto data in tabular format
```

```
##     V1        V2           V3         V4     V5           V6   V7     V8
## 1  mpg cylinders displacement horsepower weight acceleration year origin
## 2 18.0         8        307.0      130.0  3504.         12.0   70      1
## 3 15.0         8        350.0      165.0  3693.         11.5   70      1
## 4 18.0         8        318.0      150.0  3436.         11.0   70      1
## 5 16.0         8        304.0      150.0  3433.         12.0   70      1
## 6 17.0         8        302.0      140.0  3449.         10.5   70      1
##                          V9
## 1                      name
## 2 chevrolet chevelle malibu
## 3         buick skylark 320
## 4        plymouth satellite
## 5             amc rebel sst
## 6               ford torino
```

```r
Auto <- read.table("Auto.data", header = T, na.strings = "?", stringsAsFactors = T) # read Auto text fi
head(Auto)
```

```
##   mpg cylinders displacement horsepower weight acceleration year origin
## 1  18         8          307        130   3504         12.0   70      1
## 2  15         8          350        165   3693         11.5   70      1
## 3  18         8          318        150   3436         11.0   70      1
## 4  16         8          304        150   3433         12.0   70      1
## 5  17         8          302        140   3449         10.5   70      1
## 6  15         8          429        198   4341         10.0   70      1
##                        name
## 1 chevrolet chevelle malibu
## 2         buick skylark 320
## 3        plymouth satellite
## 4             amc rebel sst
## 5               ford torino
## 6          ford galaxie 500
```

```r
Auto <- read.csv("Auto.csv", na.strings = "?", stringsAsFactors = T) # read Auto csv files as table wit
head(Auto)
```

```
##   mpg cylinders displacement horsepower weight acceleration year origin
## 1  18         8          307        130   3504         12.0   70      1
## 2  15         8          350        165   3693         11.5   70      1
## 3  18         8          318        150   3436         11.0   70      1
## 4  16         8          304        150   3433         12.0   70      1
## 5  17         8          302        140   3449         10.5   70      1
## 6  15         8          429        198   4341         10.0   70      1
##                        name
## 1 chevrolet chevelle malibu
## 2         buick skylark 320
## 3        plymouth satellite
## 4             amc rebel sst
## 5               ford torino
## 6          ford galaxie 500
```

```r
dim(Auto)
```

```
## [1] 397   9
```

```r
Auto[1:4, ]
```

```
##   mpg cylinders displacement horsepower weight acceleration year origin
## 1  18         8          307        130   3504         12.0   70      1
## 2  15         8          350        165   3693         11.5   70      1
## 3  18         8          318        150   3436         11.0   70      1
## 4  16         8          304        150   3433         12.0   70      1
##                        name
## 1 chevrolet chevelle malibu
## 2         buick skylark 320
## 3        plymouth satellite
## 4             amc rebel sst
```

```
Auto <-na.omit(Auto) # omits the na data/ missing data
dim(Auto)
```

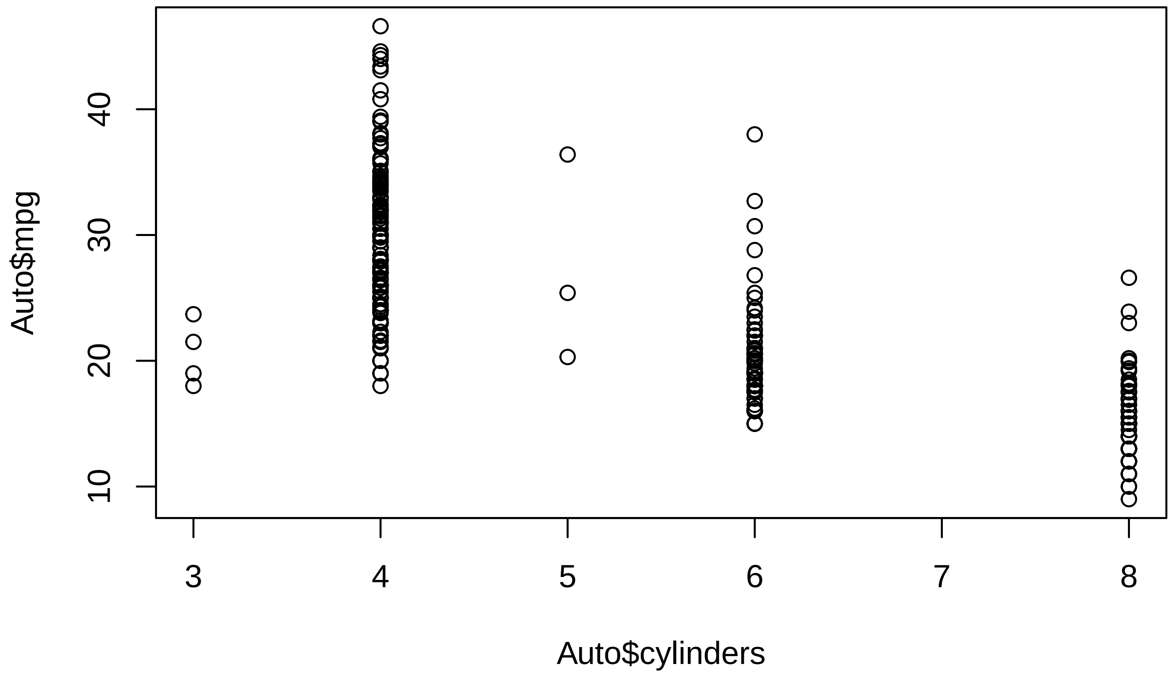
```
## [1] 392   9
```

```
names(Auto) # checks variable names
```

```
## [1] "mpg"          "cylinders"    "displacement" "horsepower"   "weight"
## [6] "acceleration" "year"         "origin"       "name"
```

2.3.5 Additional Graphical and Numerical Summaries

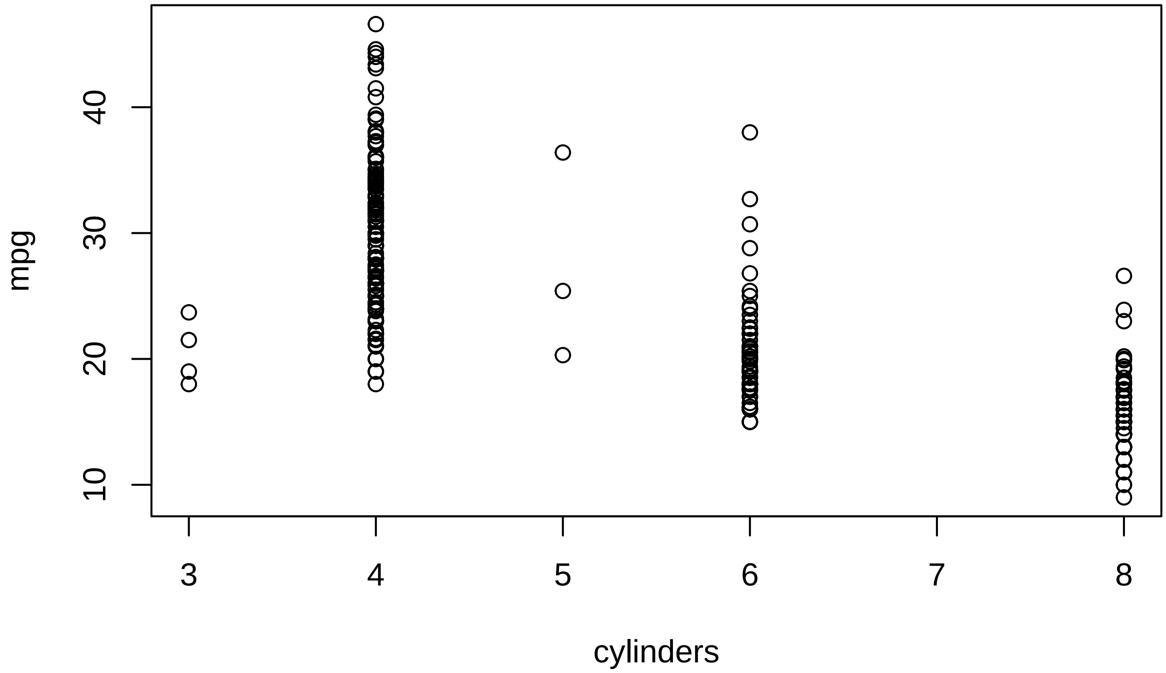
```
plot(Auto$cylinders , Auto$mpg) # plot cylinders on mpg from Auto
```



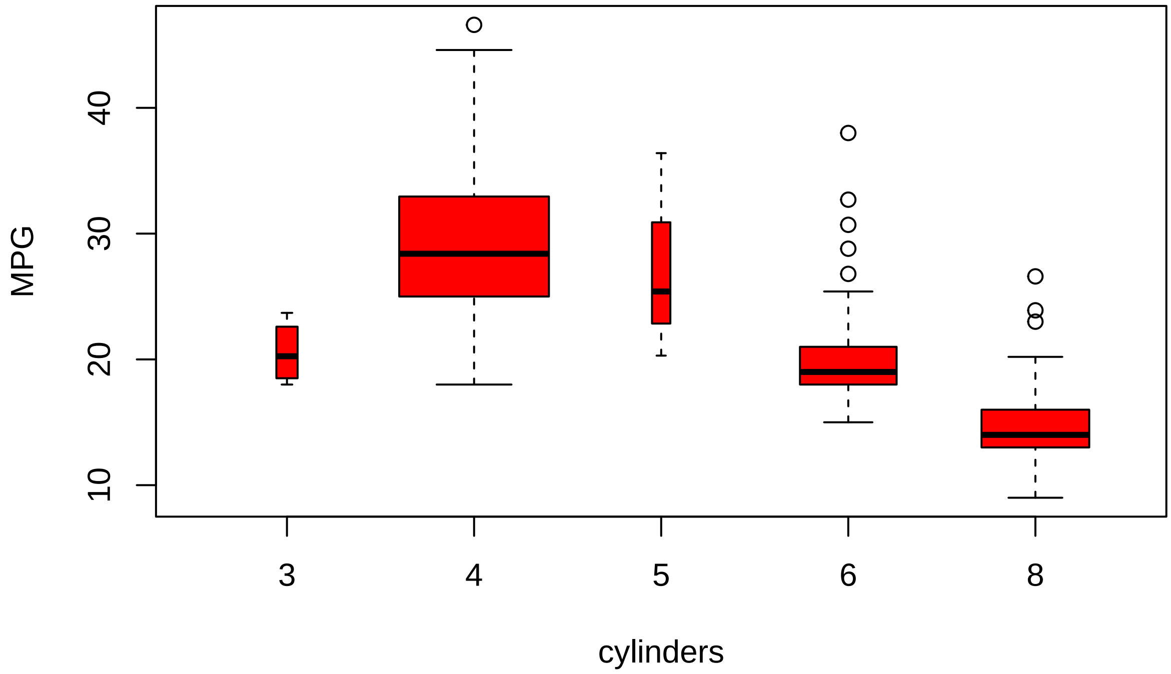
```
attach(Auto) # call the Auto data
```

```
## The following object is masked from package:ggplot2:
##
##     mpg

## The following object is masked from package:lubridate:
##
##     origin
```

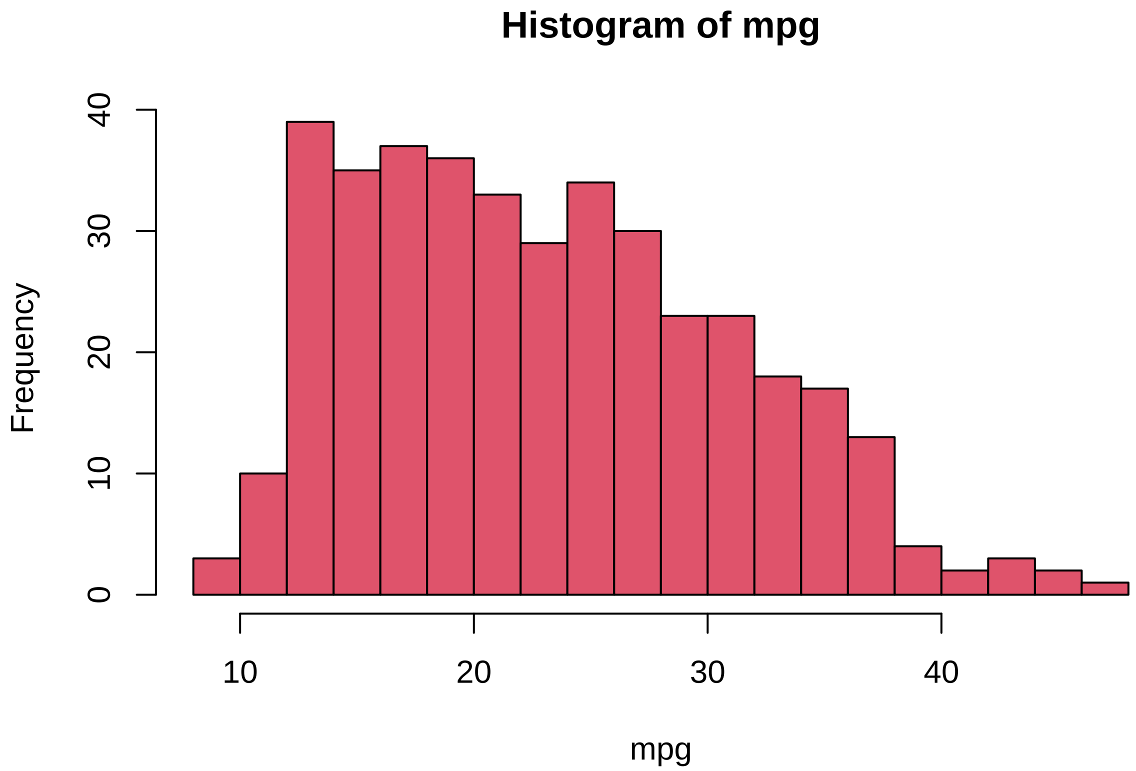
```r
plot(cylinders , mpg) # plot cylinders on mpg
```



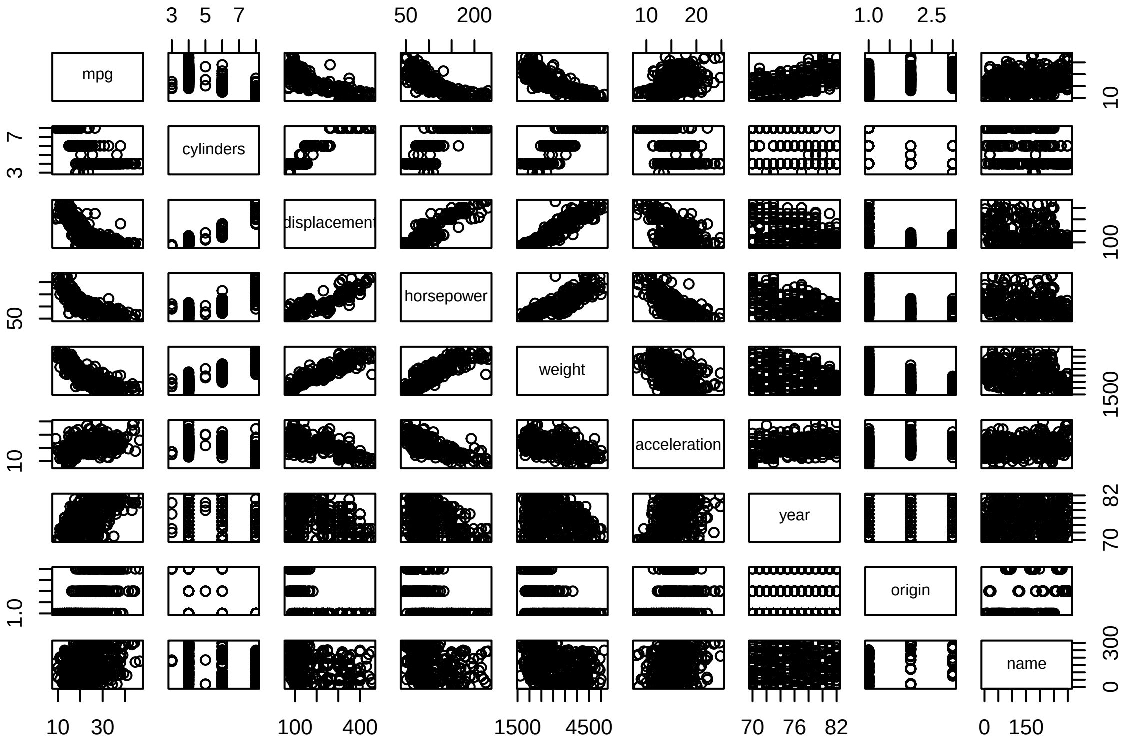
```r
cylinders <- as.factor(cylinders) # store cylinder numerical value as qualitative value

plot(cylinders, mpg, col = "red", varwidth = T, xlab = "cylinders", ylab = "MPG") # box plot produced w
```



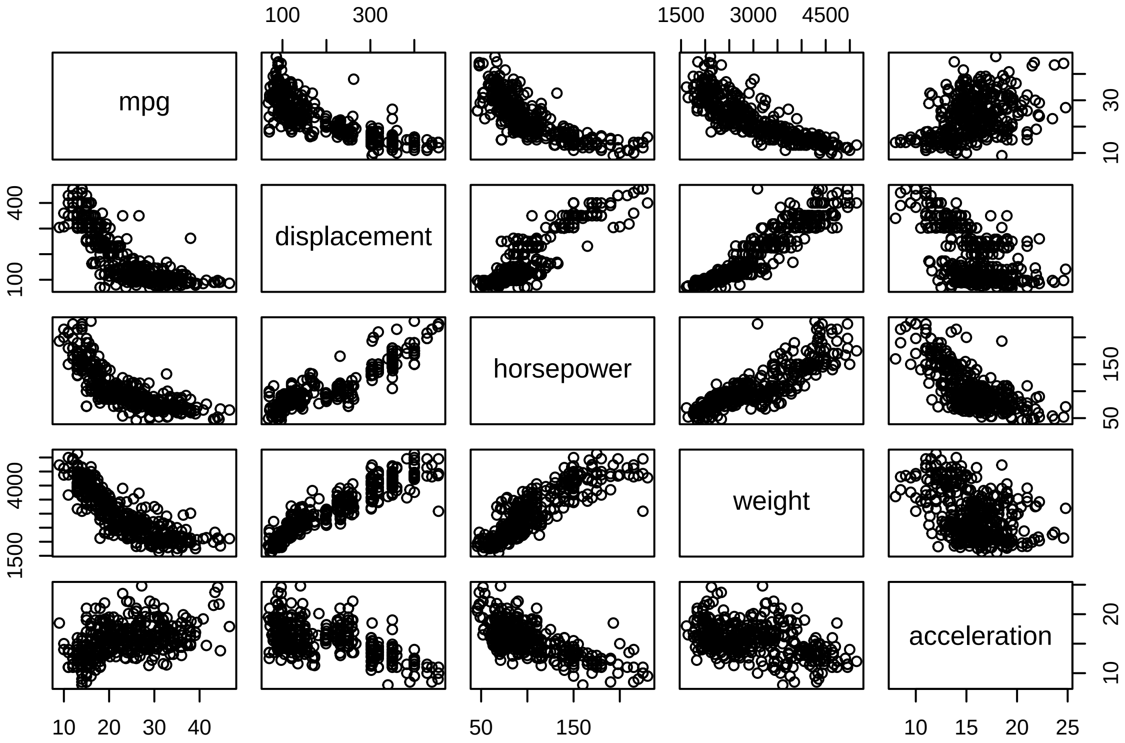
```
hist(mpg, col = 2, breaks = 15) # plot histogram based on mpg with col = 2 (equiv. to red)
```



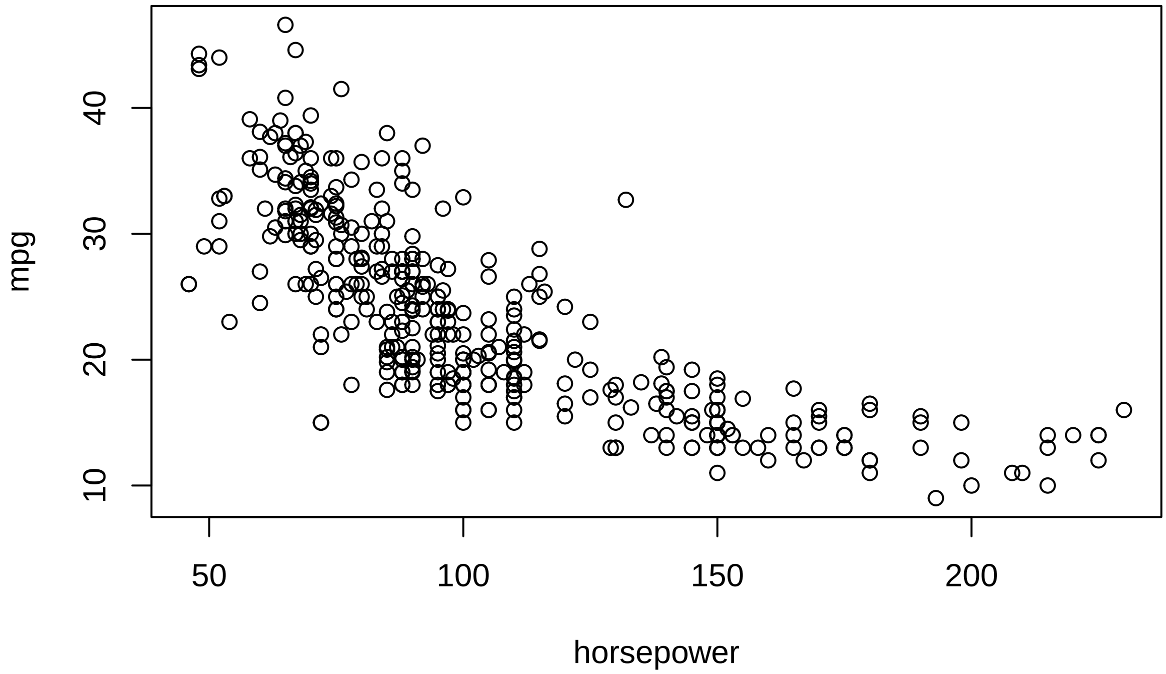
```r
pairs(Auto) # plot scatter plots based on all numerical columns in Auto
```



```r
pairs(~mpg + displacement + horsepower + weight + acceleration, data = Auto
) # plot scatter plots based on specified numerical columns in Auto
```



```
plot(horsepower , mpg)
identify(horsepower, mpg, name) # when plot of horsepower on mpg, name label will be displayed
```



```
## integer(0)
```

```
summary(Auto) # illustrate out the descriptive statistics of Auto
```

```
##       mpg          cylinders      displacement     horsepower        weight
##  Min.   : 9.00   Min.   :3.000   Min.   : 68.0   Min.   : 46.0   Min.   :1613
##  1st Qu.:17.00   1st Qu.:4.000   1st Qu.:105.0   1st Qu.: 75.0   1st Qu.:2225
##  Median :22.75   Median :4.000   Median :151.0   Median : 93.5   Median :2804
##  Mean   :23.45   Mean   :5.472   Mean   :194.4   Mean   :104.5   Mean   :2978
##  3rd Qu.:29.00   3rd Qu.:8.000   3rd Qu.:275.8   3rd Qu.:126.0   3rd Qu.:3615
##  Max.   :46.60   Max.   :8.000   Max.   :455.0   Max.   :230.0   Max.   :5140
##
##   acceleration        year           origin                      name
##  Min.   : 8.00   Min.   :70.00   Min.   :1.000   amc matador       :  5
##  1st Qu.:13.78   1st Qu.:73.00   1st Qu.:1.000   ford pinto        :  5
##  Median :15.50   Median :76.00   Median :1.000   toyota corolla    :  5
##  Mean   :15.54   Mean   :75.98   Mean   :1.577   amc gremlin       :  4
##  3rd Qu.:17.02   3rd Qu.:79.00   3rd Qu.:2.000   amc hornet        :  4
##  Max.   :24.80   Max.   :82.00   Max.   :3.000   chevrolet chevette:  4
##                                                  (Other)           :365
```

```
summary(mpg)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    9.00   17.00   22.75   23.45   29.00   46.60
```