STAT 4510 HW1

Drew Dahlquist

1/21/2022

Problem 1

2.3.1 Basic Commands

```
x = c(1,2,3,5)
x
```

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

```
x = c(1,6,2)
x
```

```
## [1] 1 6 2
```

```
y = c(1,4,3)
```

```
length(x)
```

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

```
length(y)
```

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

```
x+y
```

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

```
ls()
```

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

```
rm(x,y)
ls()
```

```
## character(0)
```

```
rm(list=ls())
```

```
?matrix
```

```
x=matrix(data=c(1,2,3,4), nrow=2, ncol=2)
x
```

```
##      [,1] [,2]
## [1,]    1    3
## [2,]    2    4
```

```
matrix(c(1,2,3,4),2,2,byrow=TRUE)
```

```
##      [,1] [,2]
```

```
## [1,]    1    2
## [2,]    3    4
```

```
sqrt(x)
```

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

```
x^2
```

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

```
x=rnorm(50)
y=x+rnorm(50,mean=50,sd=.1)
cor(x,y)
```

```
## [1] 0.9944895
```

```
set.seed(1303)
rnorm(50)
```

```
##  [1] -1.1439763145  1.3421293656  2.1853904757  0.5363925179  0.0631929665
##  [6]  0.5022344825 -0.0004167247  0.5658198405 -0.5725226890 -1.1102250073
## [11] -0.0486871234 -0.6956562176  0.8289174803  0.2066528551 -0.2356745091
## [16] -0.5563104914 -0.3647543571  0.8623550343 -0.6307715354  0.3136021252
## [21] -0.9314953177  0.8238676185  0.5233707021  0.7069214120  0.4202043256
## [26] -0.2690521547 -1.5103172999 -0.6902124766 -0.1434719524 -1.0135274099
## [31]  1.5732737361  0.0127465055  0.8726470499  0.4220661905 -0.0188157917
## [36]  2.6157489689 -0.6931401748 -0.2663217810 -0.7206364412  1.3677342065
## [41]  0.2640073322  0.6321868074 -1.3306509858  0.0268888182  1.0406363208
## [46]  1.3120237985 -0.0300020767 -0.2500257125  0.0234144857  1.6598706557
```

```
set.seed(3)
y=rnorm(100)
mean(y)
```

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

```
var(y)
```

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

```
sqrt(var(y))
```

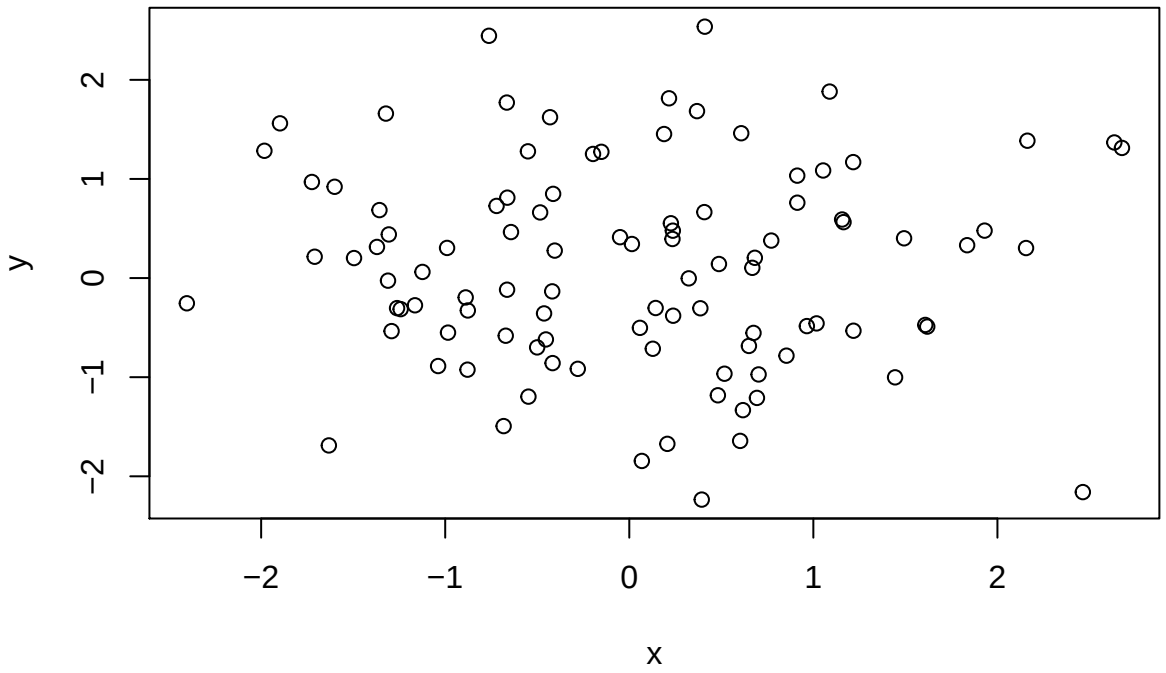
```
## [1] 0.8560768
```

```
sd(y)
```

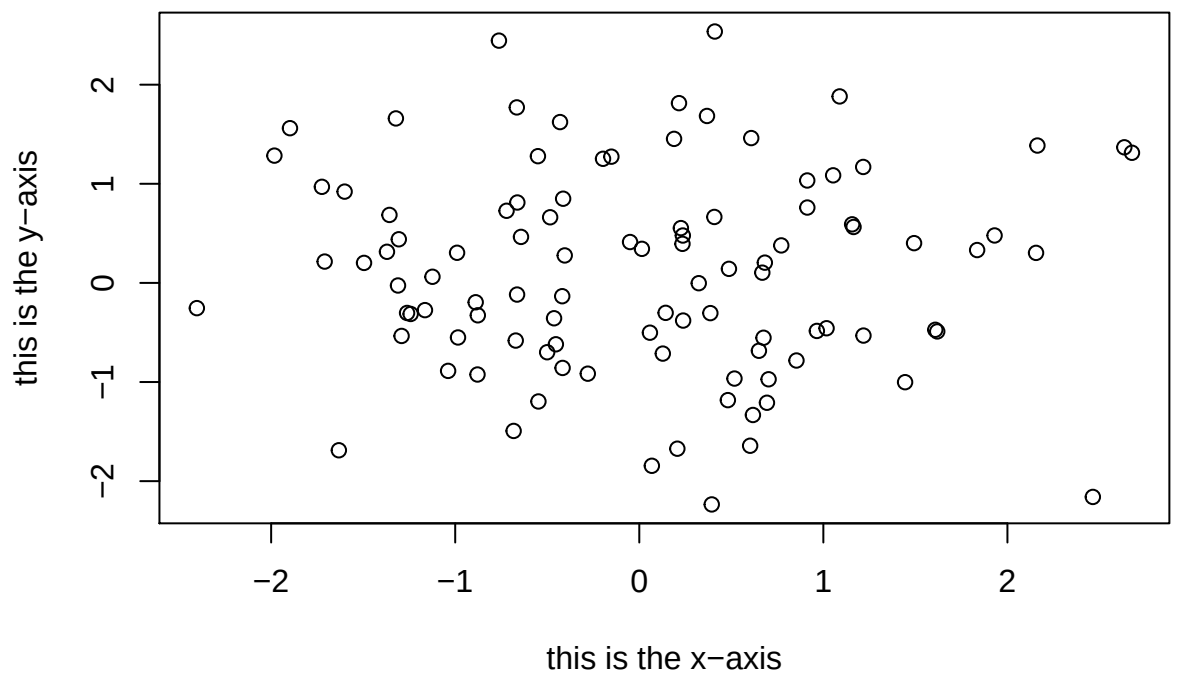
```
## [1] 0.8560768
```

2.3.2 Graphics

```
x=rnorm(100)
y=rnorm(100)
plot(x,y)
```




```
plot(x,y,xlab="this is the x-axis",ylab="this is the y-axis", main="Plot of X vs Y")
```




```
pdf("Figure.pdf")
plot(x,y,,col="green")
dev.off()
```

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

```
x=seq(1,10)
x
```

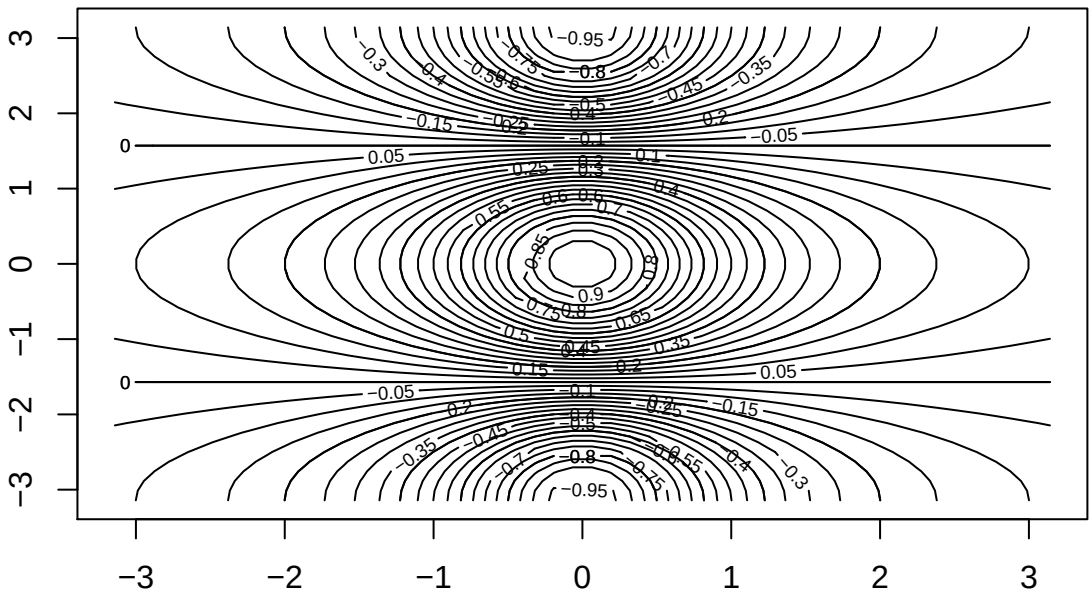
```
##  [1]  1  2  3  4  5  6  7  8  9 10
```

```
x=1:10
x
```

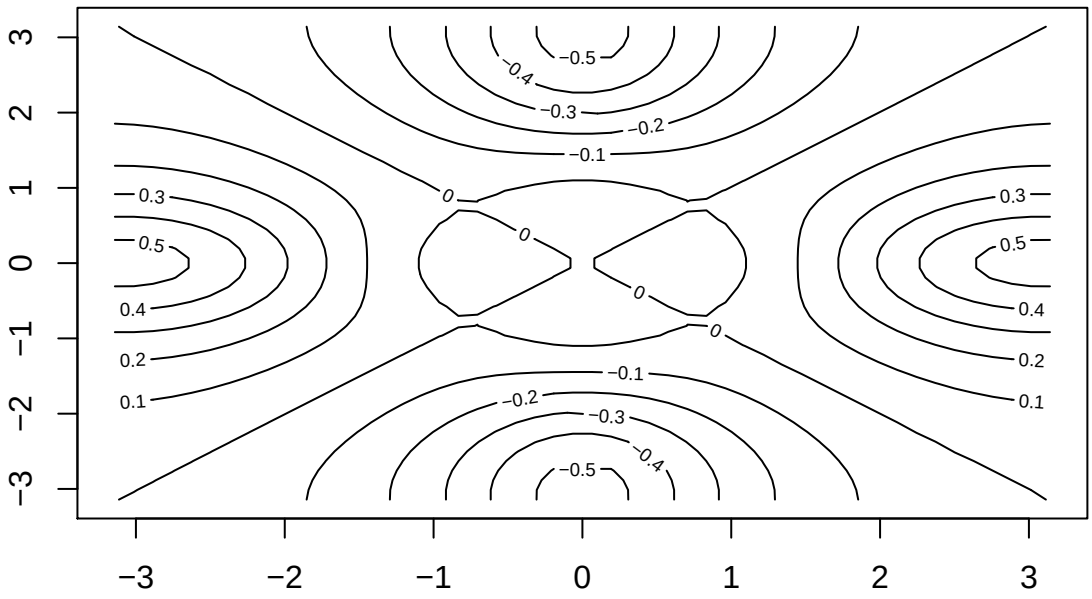
```
##  [1]  1  2  3  4  5  6  7  8  9 10
```

```
x=seq(-pi,pi,length=50)
```

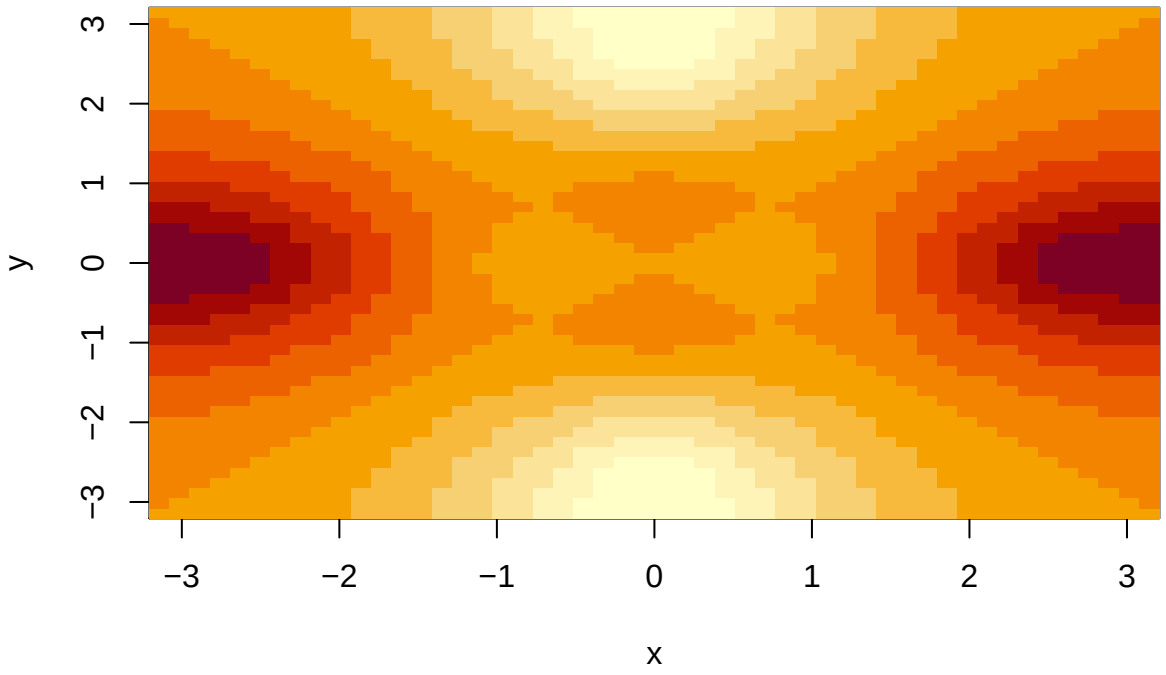
```
y=x
f=outer(x,y,function(x,y)cos(y)/(1+x^2))
contour(x,y,f)
contour(x,y,f,nlevels=45,add=T)
```



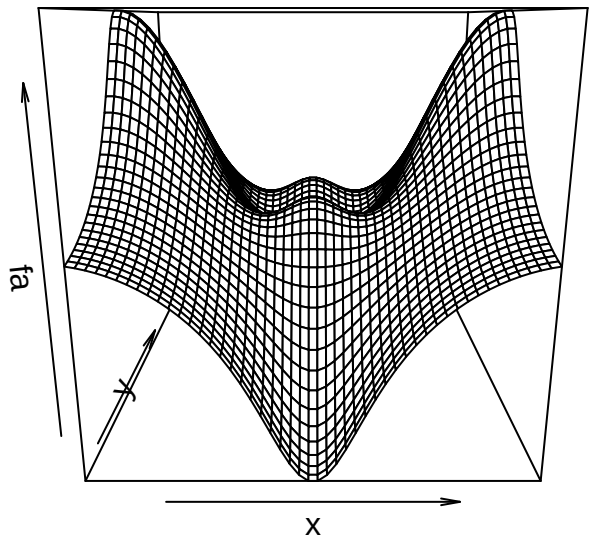

```
fa=(f-t(f))/2
contour(x,y,fa,nlevels=15)
```



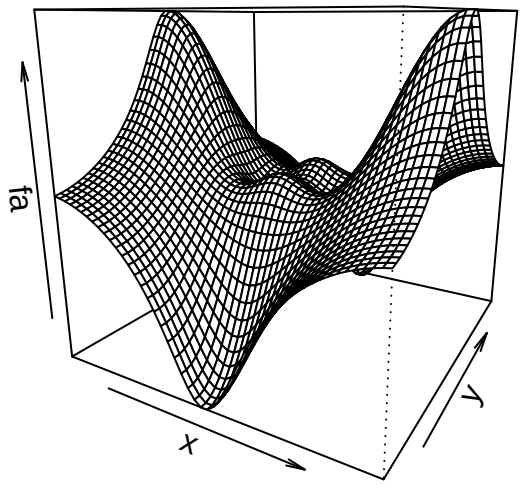

```
image(x,y,fa)
```



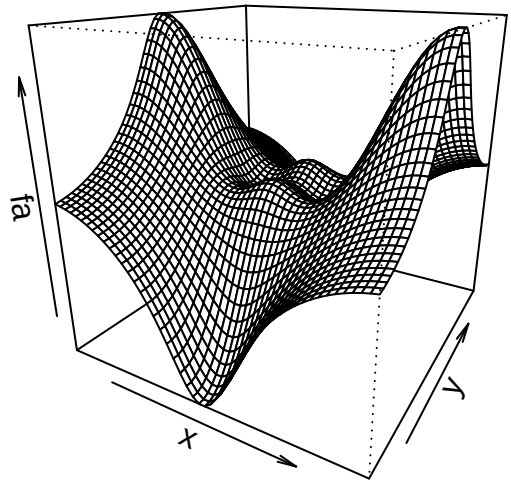

```
persp(x,y,fa)
```



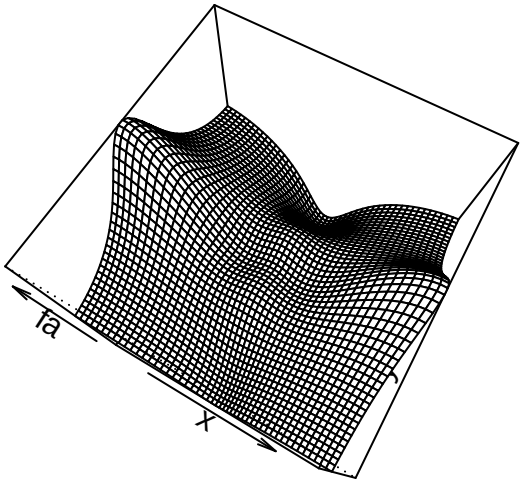

```
persp(x,y,fa,theta=30)
```



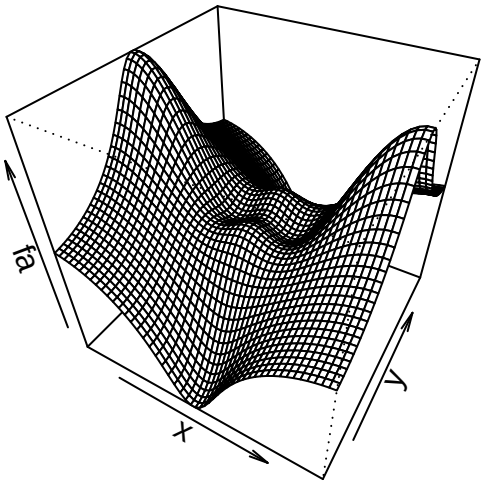

```
persp(x,y,fa,theta=30,phi=20)
```




```
persp(x,y,fa,theta=30,phi=70)
```




```
persp(x,y,fa,theta=30,phi=40)
```



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]
```

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

```
A[c(1,3),c(2,4)]
```

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

```
A[1:3,2:4]
```

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

```
A[1:2,]
```

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

```
A[,1:2]
```

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

```
A[1,]
```

```
## [1]  1  5  9 13
```

```
A[-c(1,3),]
```

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

```
A[-c(1,3),-c(1,3,4)]
```

```
## [1] 6 8
```

```
dim(A)
```

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

2.3.4 Loading Data

```
Auto=read.table("Auto.data")
fix(Auto)
```

```
Auto=read.table("Auto.data",header=T,na.strings="?")
fix(Auto)
```

```
Auto=read.csv("Auto.csv",header=T,na.strings="?")
fix(Auto)
dim(Auto)
```

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

```
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)
dim(Auto)
```

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

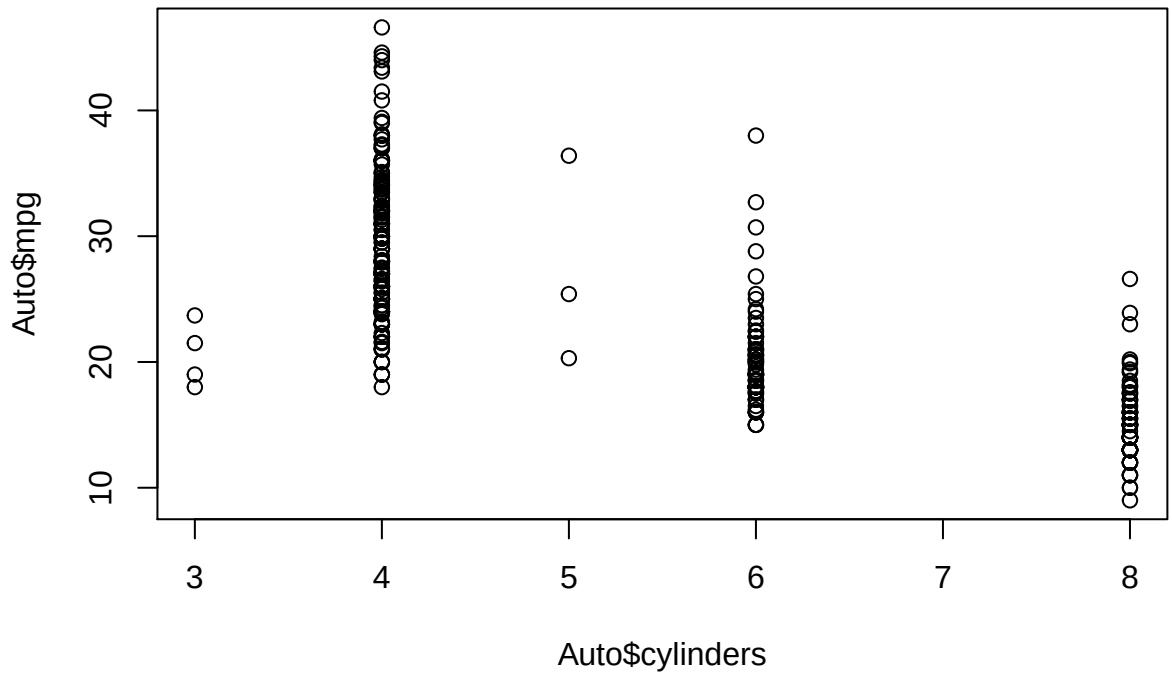
```
names(Auto)
```

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

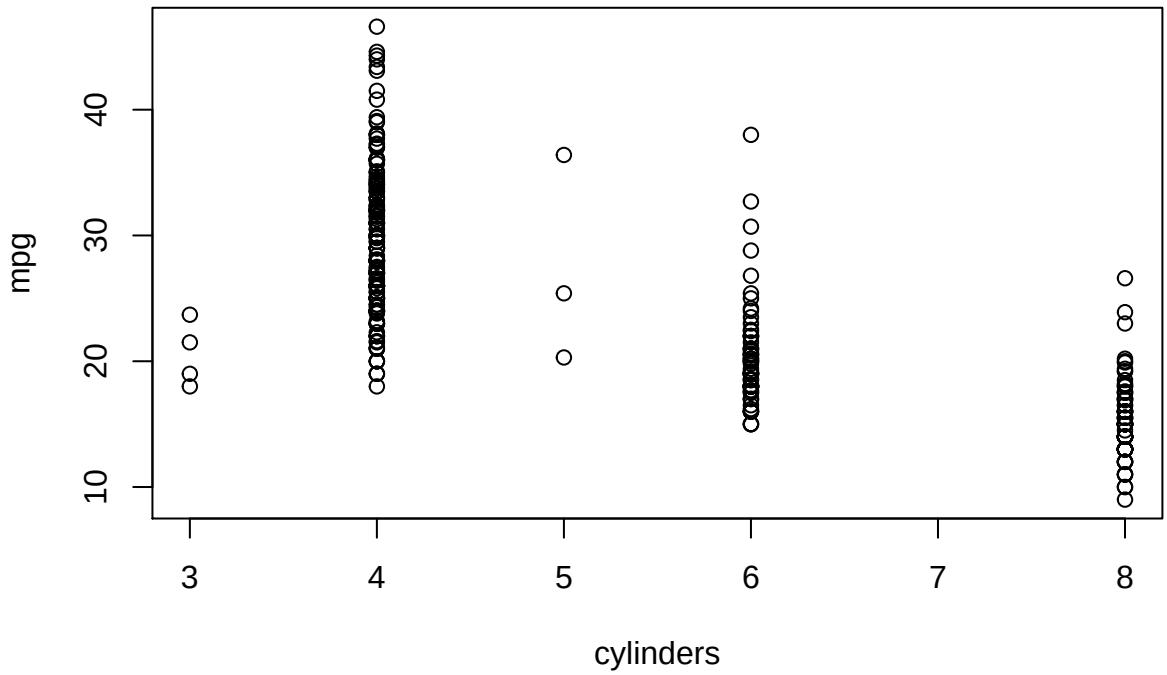
2.3.5 Additional Graphical and Numerical Summaries

```
plot(cylinders, mpg)
```

```
plot(Auto$cylinders, Auto$mpg)
```

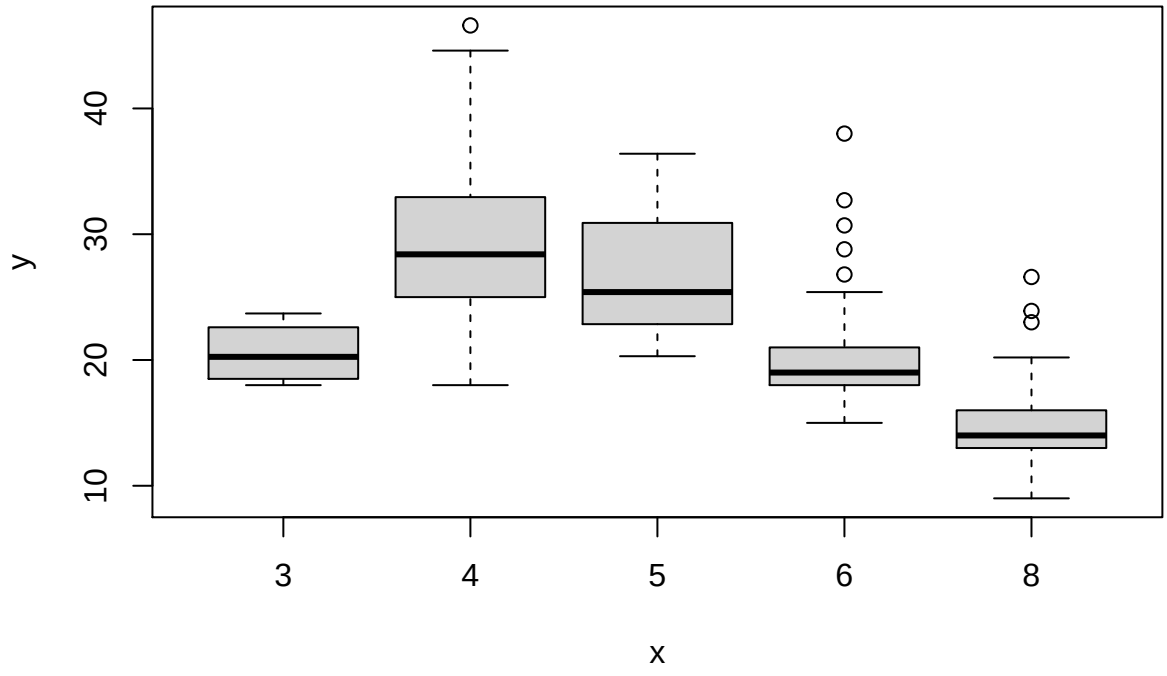



```
attach(Auto)
plot(cylinders, mpg)
```

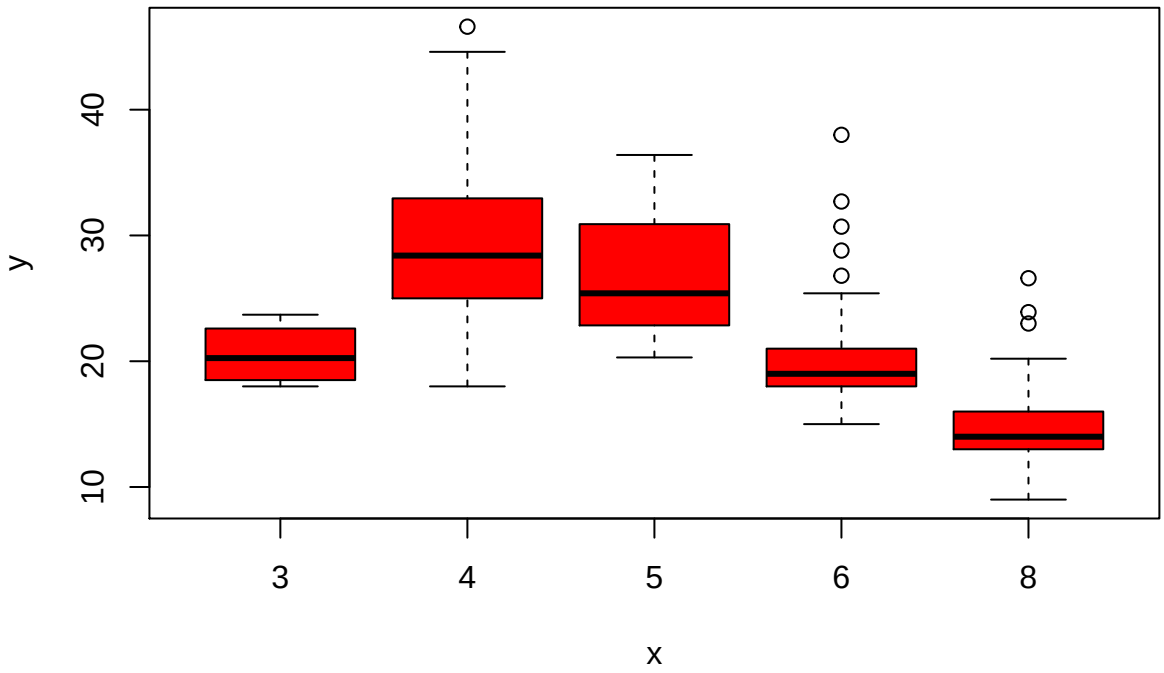



```
cylinders=as.factor(cylinders)
```

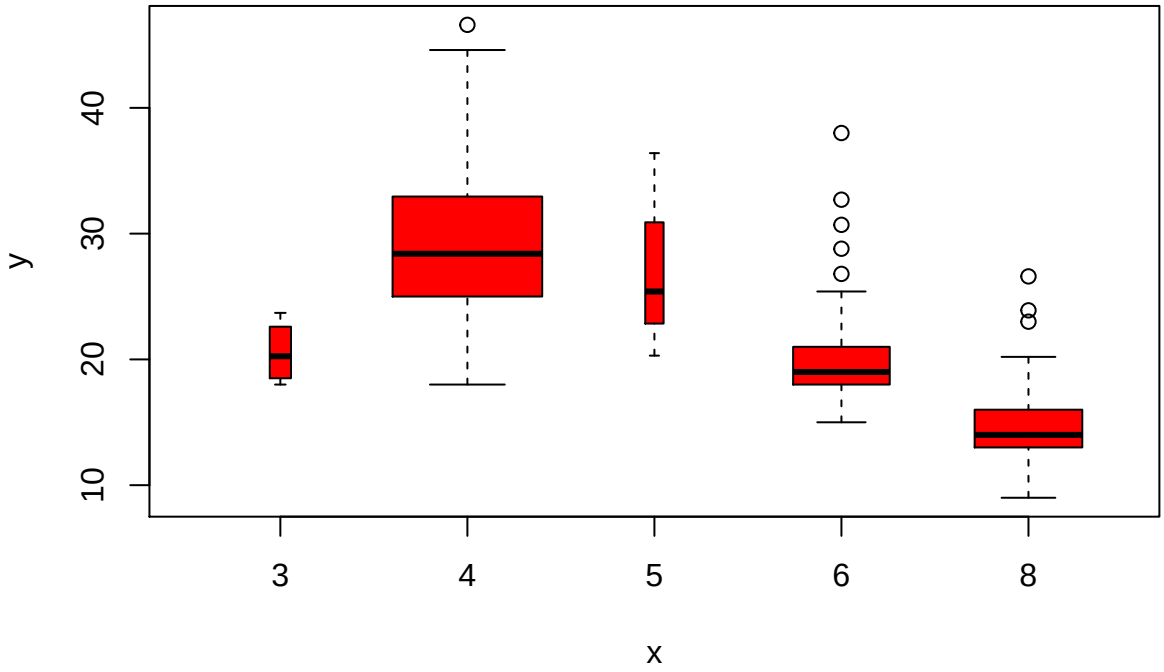
```
plot(cylinders, mpg)
```



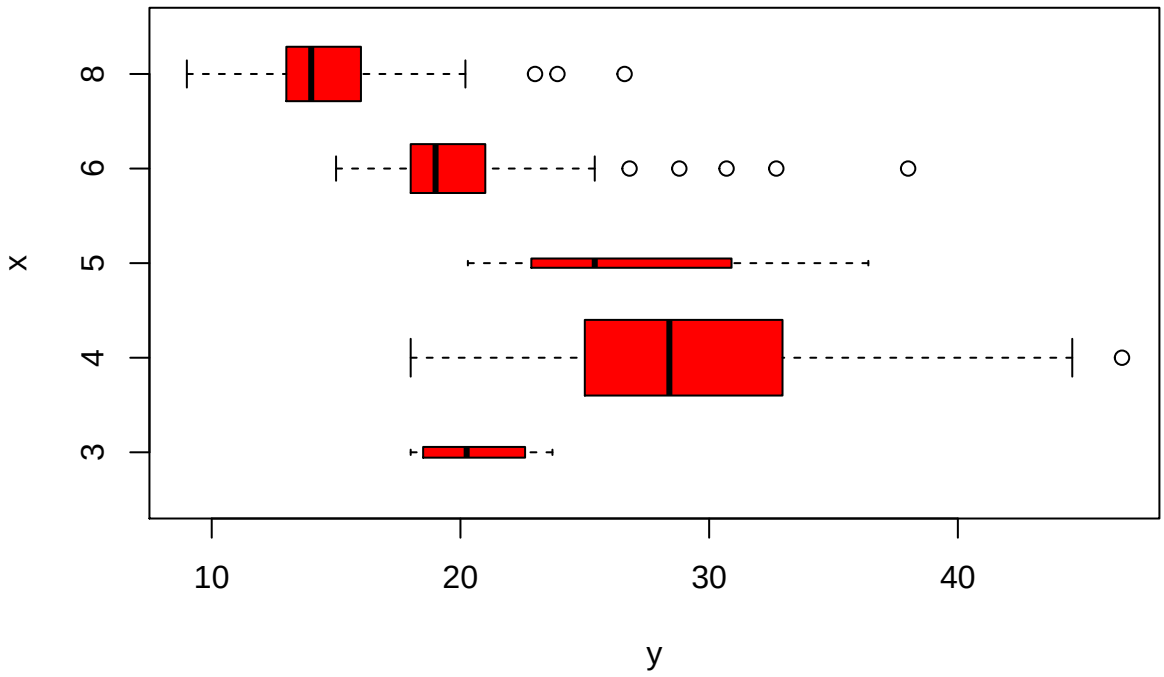

```
plot(cylinders, mpg, col="red")
```



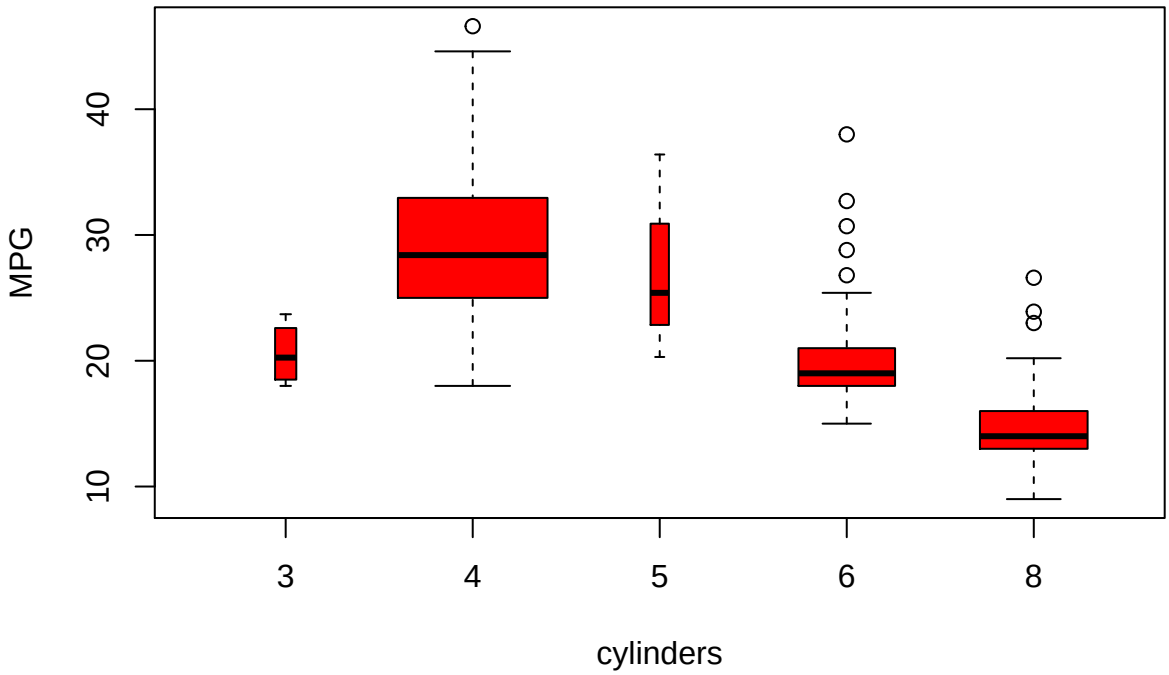

```
plot(cylinders, mpg, col="red", varwidth=T)
```



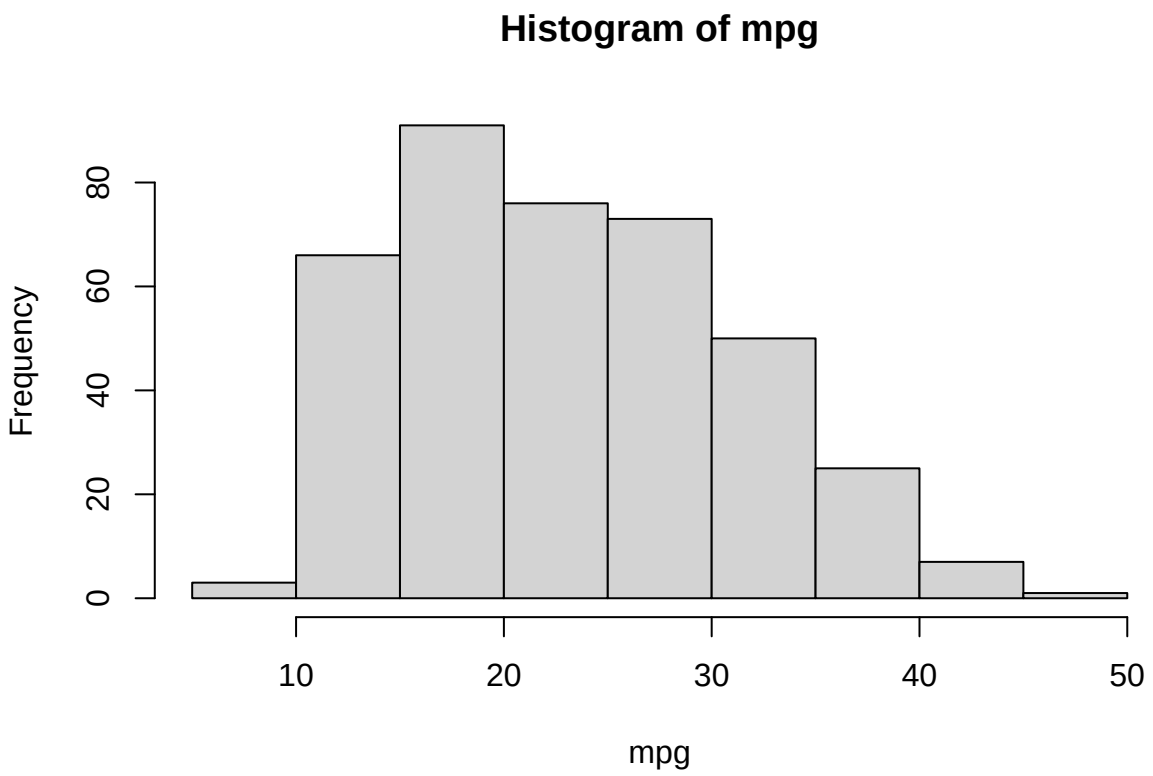

```
plot(cylinders, mpg, col="red", varwidth=T, horizontal=T)
```



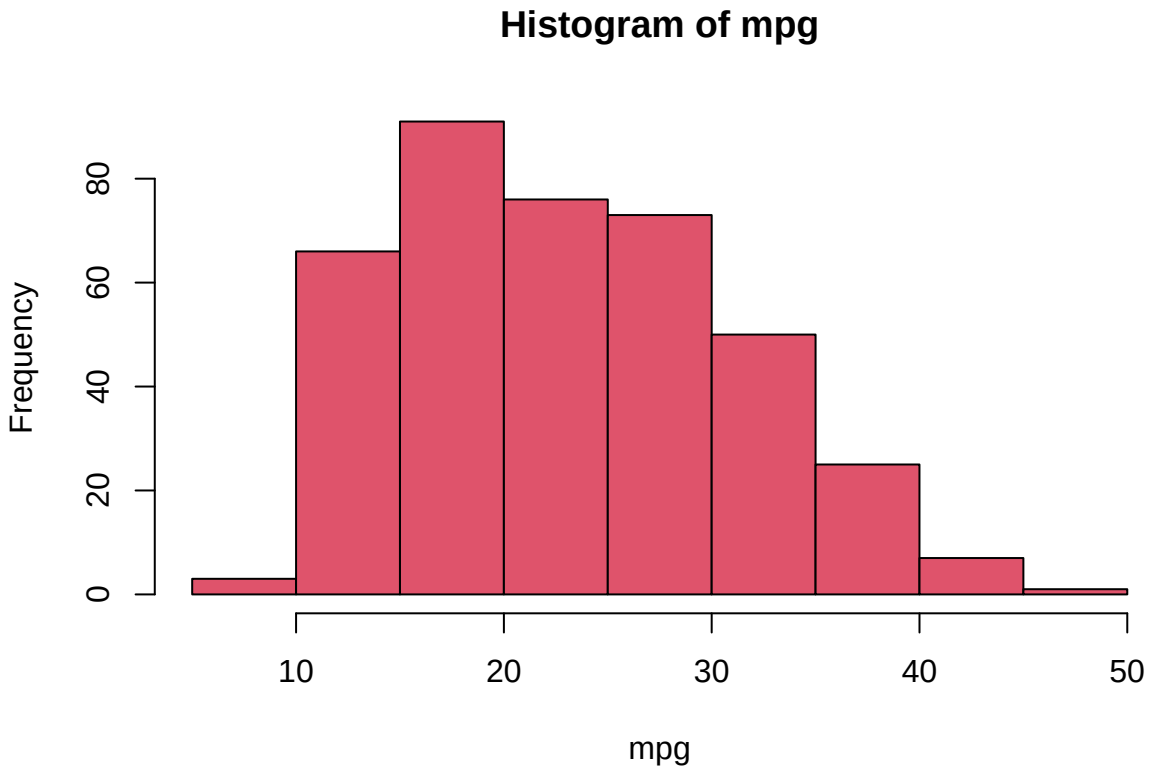

```
plot(cylinders, mpg, col="red", varwidth=T, xlab="cylinders", ylab="MPG")
```



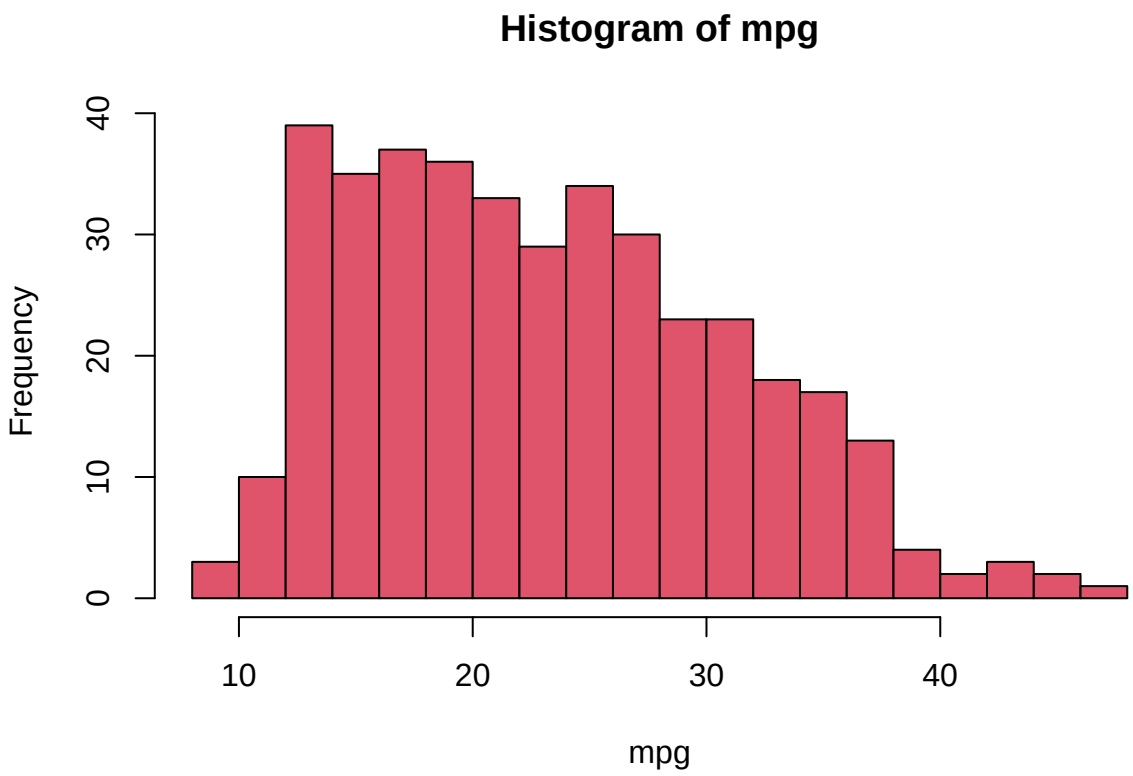

```
hist(mpg)
```




```
hist(mpg, col=2)
```

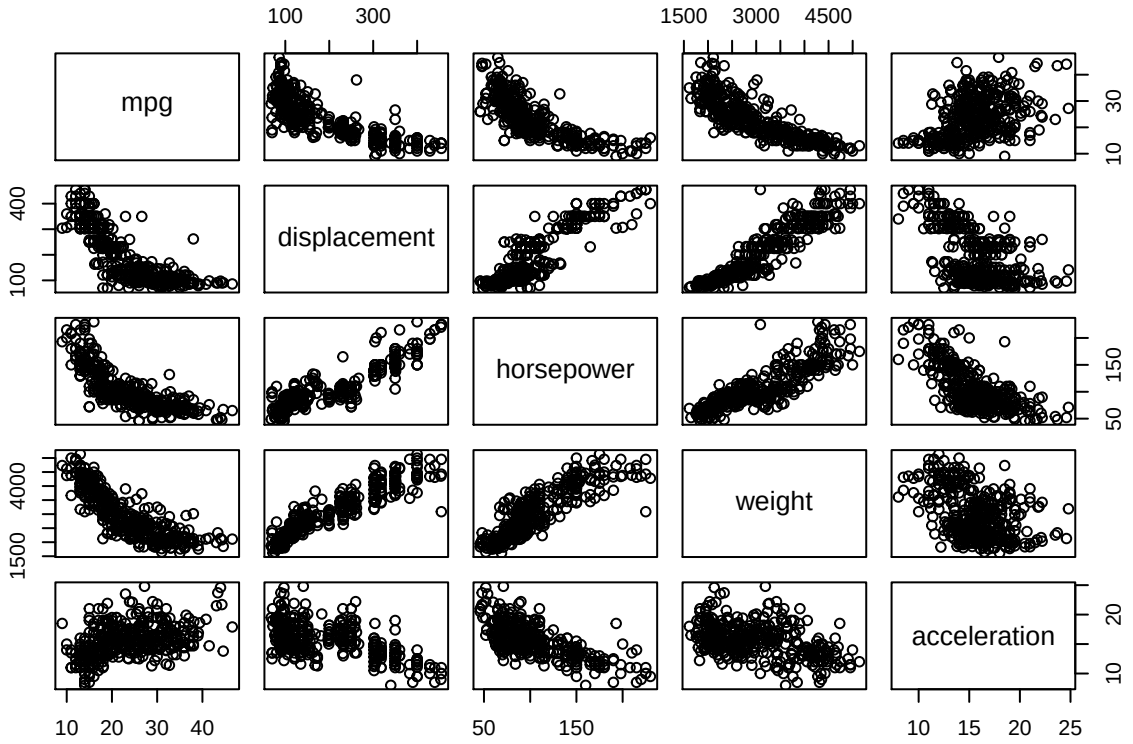



```
hist(mpg, col=2, breaks=15)
```

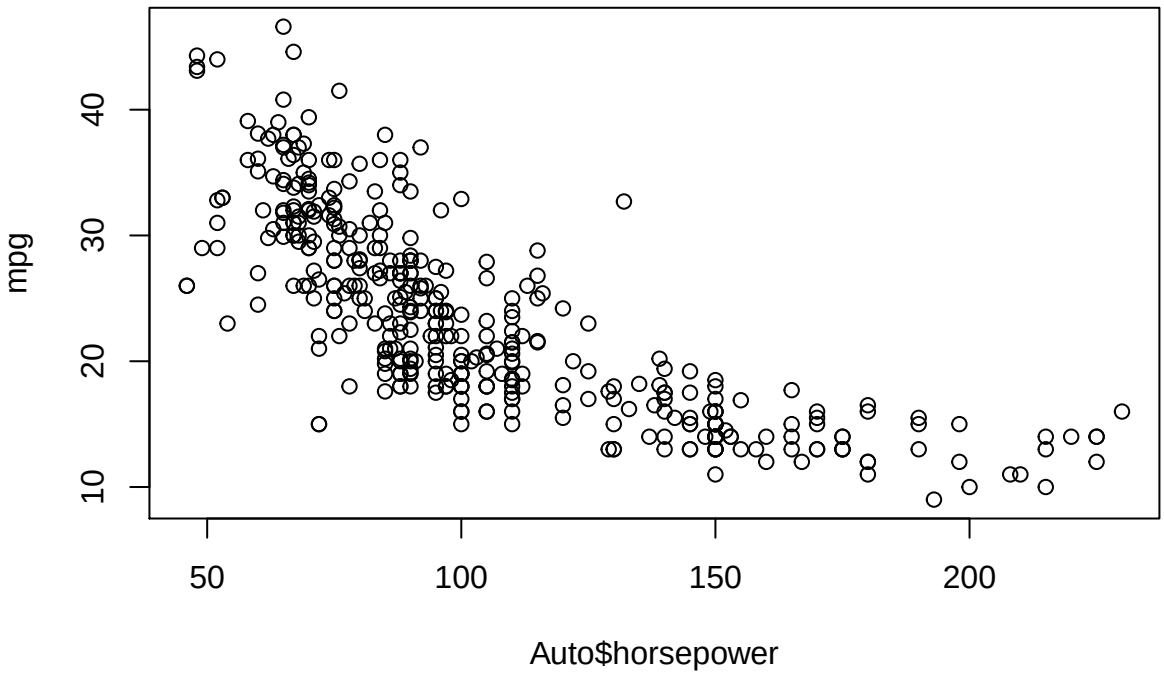



```
pairs(Auto)
```

```
pairs(~ mpg + displacement + horsepower + weight + acceleration, Auto)
```




```
plot(Auto$horsepower, mpg)
identify(horsepower, mpg, name)
```




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

```
summary(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   Length:392
##  1st Qu.:13.78   1st Qu.:73.00   1st Qu.:1.000   Class :character
##  Median :15.50   Median :76.00   Median :1.000   Mode  :character
##  Mean   :15.54   Mean   :75.98   Mean   :1.577
##  3rd Qu.:17.02   3rd Qu.:79.00   3rd Qu.:2.000
##  Max.   :24.80   Max.   :82.00   Max.   :3.000
```

```
summary(mpg)
```

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

Problem 2

a)

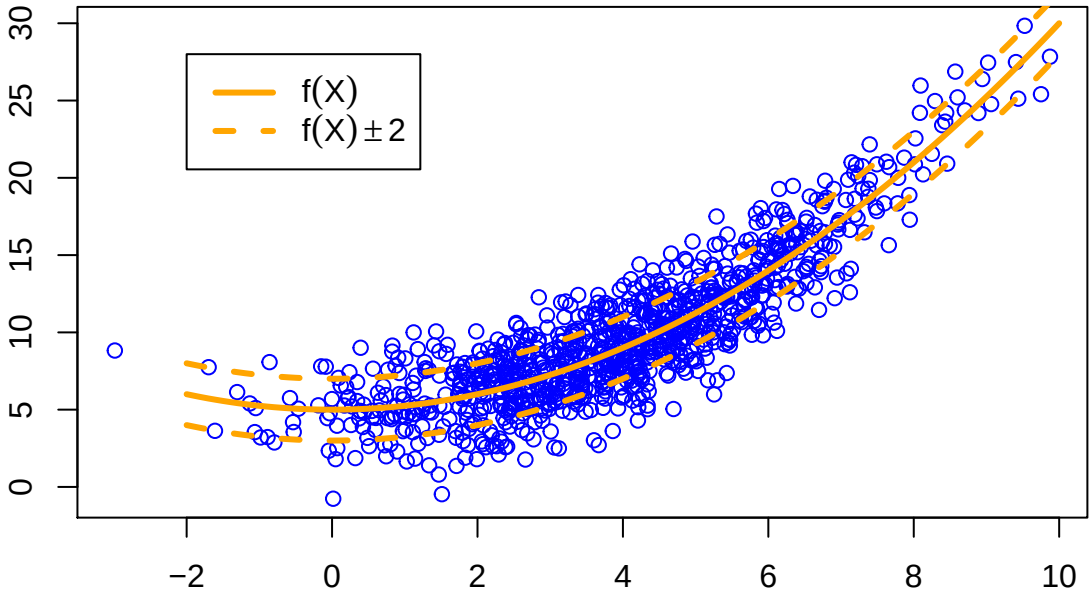
```
?rnorm
```

The `rnorm` function generates a normal distribution with given mean and sd. You may also give `rnorm` a vector of quantiles or probabilities, a number of observations to make. It also allows for given probabilites to be given as log(p) or as P[X < x].

b)

The reducible error for the prediction is 1.245.

c)



After changing the sd of the error terms to 2, the sampled data points become less distributed around the true underlying function.

Problem 3

a)

```
install.packages("ISLR")
```

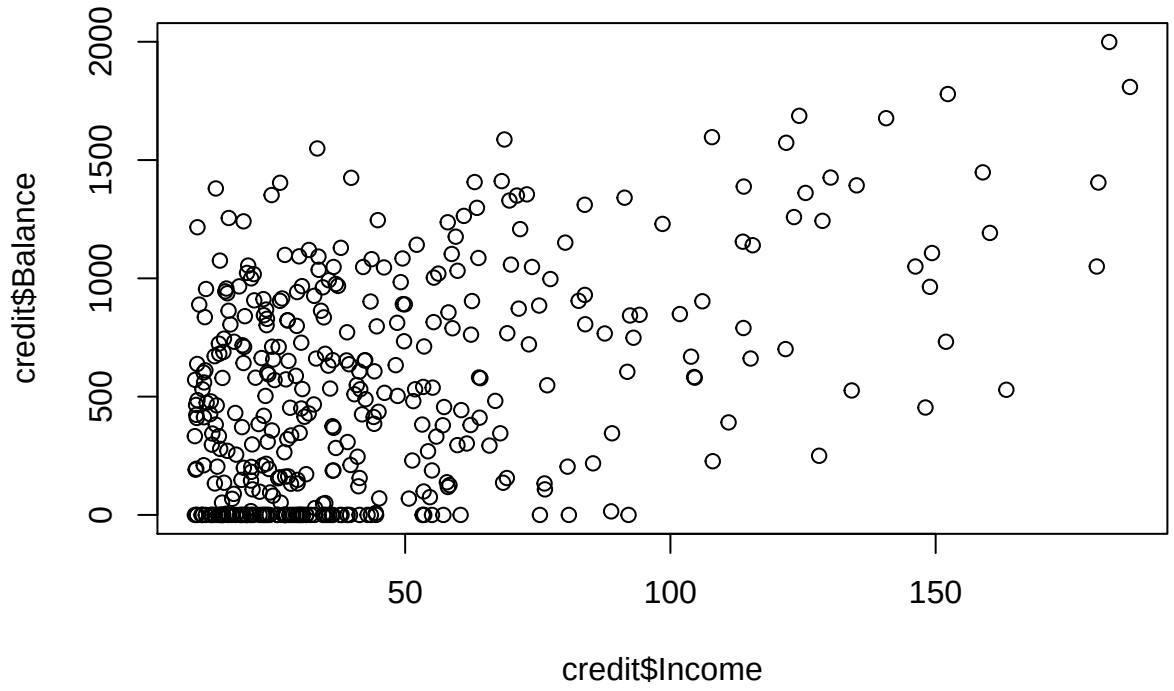
```
library(ISLR)

?Credit
```

b)

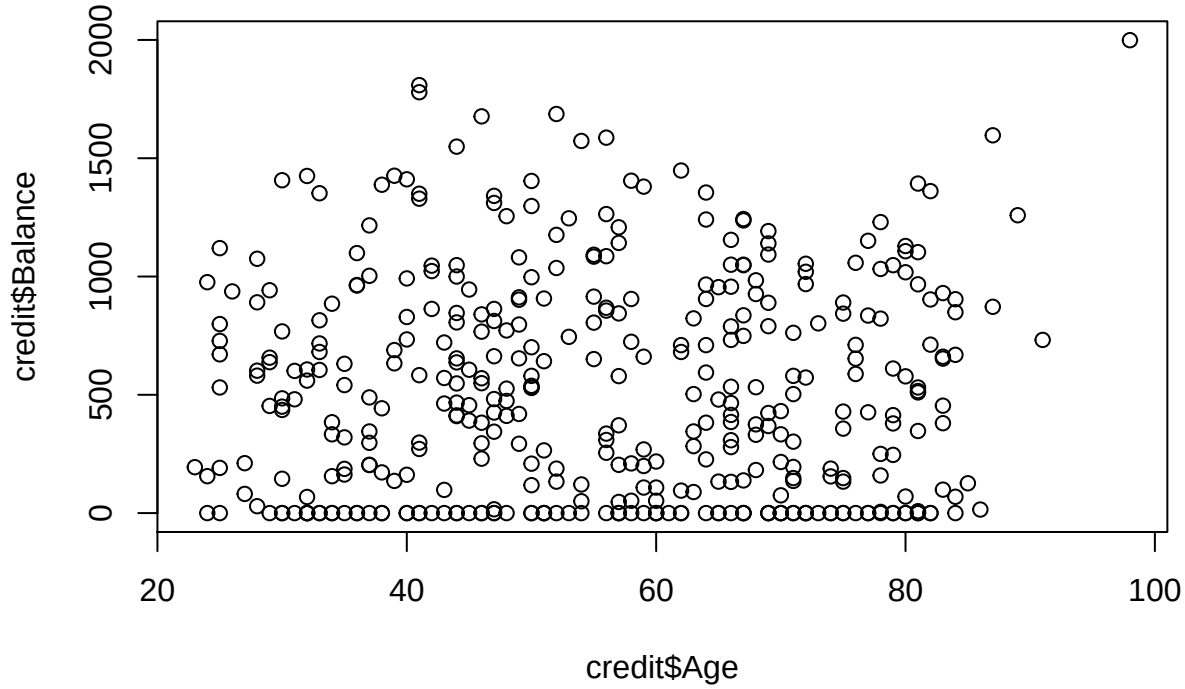
```
credit = read.csv("Credit.csv")
```

c)

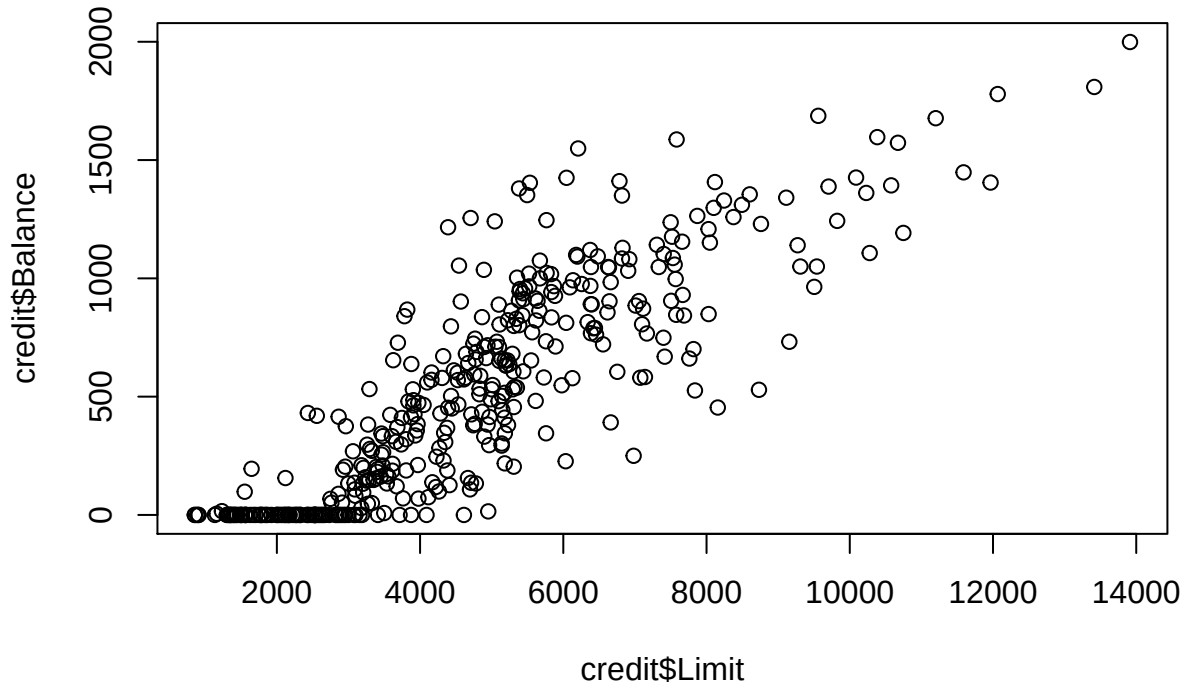
```
plot(credit$Income, credit$Balance)
```




```
plot(credit$Age, credit$Balance)
```




```
plot(credit$Limit, credit$Balance)
```



d)

```
summary(credit)
```

```
##      Income           Limit           Rating          Cards      
##  Min.   : 10.35   Min.   :  855   Min.   : 93.0   Min.   :1.000  
##  1st Qu.: 21.01   1st Qu.: 3088   1st Qu.:247.2   1st Qu.:2.000  
##  Median : 33.12   Median : 4622   Median :344.0   Median :3.000  
##  Mean   : 45.22   Mean   : 4736   Mean   :354.9   Mean   :2.958  
##  3rd Qu.: 57.47   3rd Qu.: 5873   3rd Qu.:437.2   3rd Qu.:4.000  
##  Max.   :186.63   Max.   :13913   Max.   :982.0   Max.   :9.000  
##       Age          Education         Own              Student         
##  Min.   :23.00   Min.   : 5.00   Length:400         Length:400        
##  1st Qu.:41.75   1st Qu.:11.00   Class :character   Class :character  
##  Median :56.00   Median :14.00   Mode  :character   Mode  :character  
##  Mean   :55.67   Mean   :13.45                                        
##  3rd Qu.:70.00   3rd Qu.:16.00                                        
##  Max.   :98.00   Max.   :20.00                                        
##    Married             Region             Balance       
##  Length:400         Length:400         Min.   :   0.00  
##  Class :character   Class :character   1st Qu.:  68.75  
##  Mode  :character   Mode  :character   Median : 459.50  
##                                        Mean   : 520.01  
##                                        3rd Qu.: 863.00  
##                                        Max.   :1999.00
```

```
credit$Own = as.factor(credit$Own)
credit$Student = as.factor(credit$Student)
credit$Married = as.factor(credit$Married)
credit$Region = as.factor(credit$Region)
```

e)

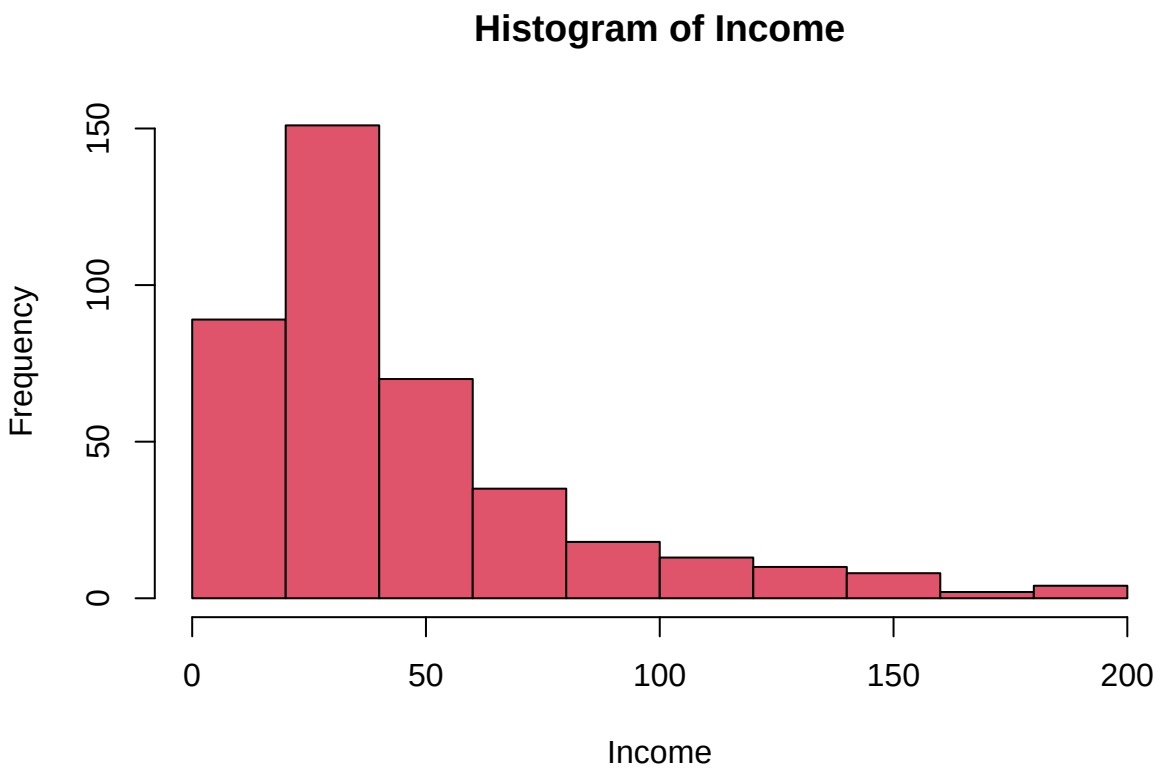
```
summary(credit)
```

```
##      Income           Limit           Rating          Cards
##  Min.   : 10.35   Min.   :  855   Min.   : 93.0   Min.   :1.000
##  1st Qu.: 21.01   1st Qu.: 3088   1st Qu.:247.2   1st Qu.:2.000
##  Median : 33.12   Median : 4622   Median :344.0   Median :3.000
##  Mean   : 45.22   Mean   : 4736   Mean   :354.9   Mean   :2.958
##  3rd Qu.: 57.47   3rd Qu.: 5873   3rd Qu.:437.2   3rd Qu.:4.000
##  Max.   :186.63   Max.   :13913   Max.   :982.0   Max.   :9.000
##       Age          Education      Own      Student   Married     Region
##  Min.   :23.00   Min.   : 5.00   No :193   No :360   No :155   East : 99
##  1st Qu.:41.75   1st Qu.:11.00   Yes:207   Yes: 40   Yes:245   South:199
##  Median :56.00   Median :14.00                                 West :102
##  Mean   :55.67   Mean   :13.45
##  3rd Qu.:70.00   3rd Qu.:16.00
##  Max.   :98.00   Max.   :20.00
##     Balance
##  Min.   :   0.00
##  1st Qu.:  68.75
##  Median : 459.50
##  Mean   : 520.01
##  3rd Qu.: 863.00
##  Max.   :1999.00
```

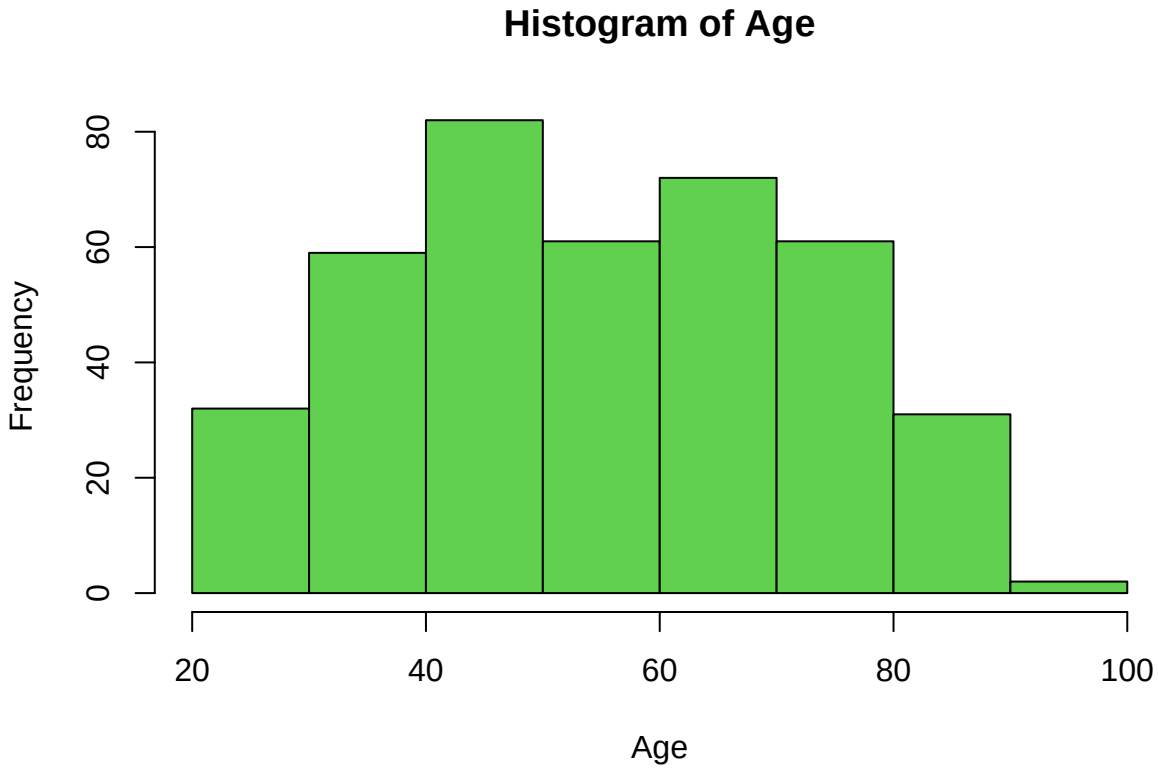
The factor variables now the amount of observations for each category for the variable.

f)

```
hist(credit$Income, main="Histogram of Income", xlab="Income", ylab="Frequency", col="2")
```

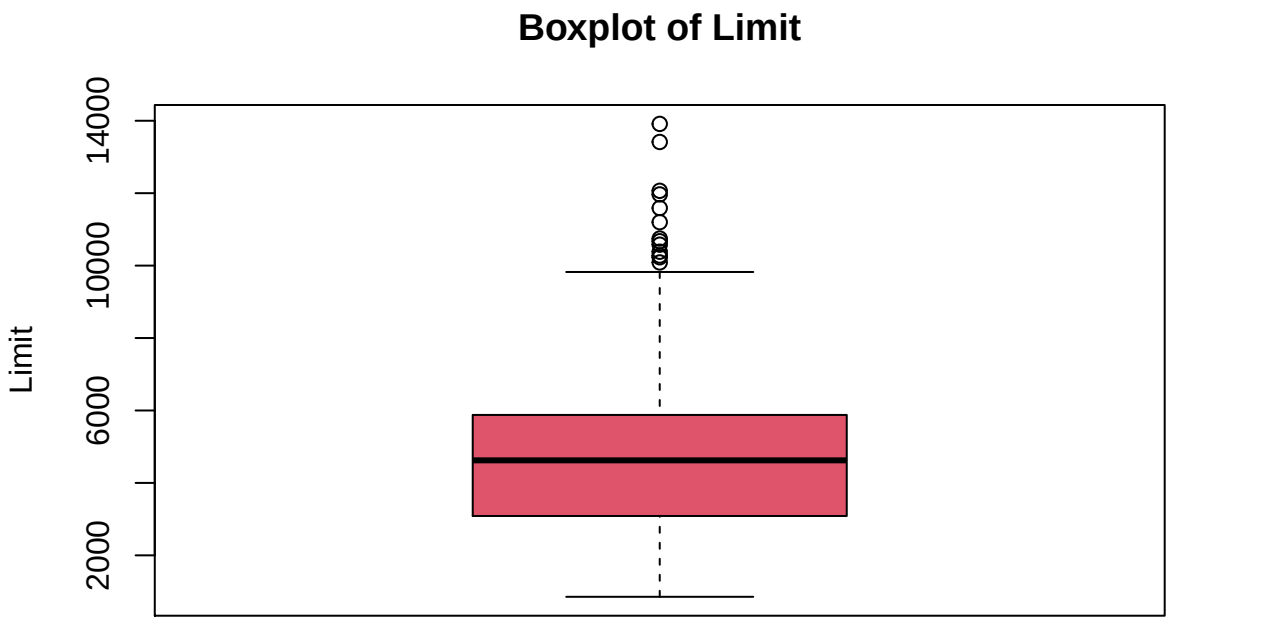



```r
hist(credit$Age, main="Histogram of Age", xlab="Age", ylab="Frequency", col="3")
```



g)

```
boxplot(credit$Limit, main="Boxplot of Limit", ylab="Limit", col="2")
```




```
boxplot(credit$Balance, main="Boxplot of Balance", ylab="Balance", col="3")
```