Ejercicios LAB2 1-7

Laura Sudupe

12/10/2020

Ejercicio 1.En este ejercicio construimos las variables en nuestro script, recodificamos variables no numéricas y buscamos resúmenes estadísticos y tablas de frecuencias de nuestras variables y datos:

```
id=c(100,110,120,130,140,150,160,170,180,190)
edad=c(18,19,NA,18,24,27,22,25,22,15)
sexo=c(2,1,2,2,1,2,1,1,2,1)
peso=c(65,58,56,61,84,99,50,64,87,87)
altura=c(161,170,174,165,150,171,181,170,184,190)
niv_col=c(1,2,3,1,3,2,3,1,2,3)

tabla_datos = data.frame(id, edad, sexo, peso, altura, niv_col)

sexo = factor(sexo, levels = c(1, 2), labels = c("Hombre", "Mujer"))
niv_col = factor(niv_col, levels = c(1, 2, 3),
                 labels = c("Colesterol alto", "COlesterol normal",
                            "Colesterol bajo"))
tabla_datos_2 = data.frame(id, edad, sexo, peso, altura, niv_col)

edad = summary(edad) #datos estadisticos
edad_five_num = fivenum(edad) #5 numeros caracteristicos
edad_mean = mean(edad)
edad_var = var(edad)
peso_altura_cor = cor(peso, altura) #correlacion entre dos variables
edad_desviacion_estandar = sd(edad)
```

Ejercicio 2.Se quiere estudiar la posible asociación entre el hecho de que una gestante fume durante el embarazo y que el bebé presente bajo peso al nacer. Se realiza un estudio de seguimiento de 2.000 embarazadas.

```
Si = c(43, 105)
No = c(207, 1645)
tabla = data.frame(Si, No)
rownames(tabla) <- c("Fumadora", "No fumadora")
tabla
```

```
##              Si   No
## Fumadora     43  207
## No fumadora 105 1645
```

```
#Test Chi-cuadrado para la asociacion entre las dos variables.
Chi_test <- chisq.test(tabla)
Chi_test
```

```
##
##  Pearson's Chi-squared test with Yates' continuity correction
##
## data:  tabla
## X-squared = 38.427, df = 1, p-value = 5.685e-10
```

```
#La conclusión es que hay una asociación significativa entre las variables
#(Chi2=38.42, gl = 1, p-value<0.05).
```

Ejercicio 3. Escogemos dos conjuntos de datos incorporados en los paquetes Datasets y MASS de R. Los data frame que queremos son Orange. (https://stat.ethz.ch/R-manual/R- atched/library/datasets/html/Orange.html) e Iris del que ya hemos hablado anteriormente.

```
install.packages('datasets')
```

```
## Warning: package 'datasets' is in use and will not be installed
```

```
#cargamos los paquetes
library(datasets)
data(package='datasets') #ambos dataframes estan en esta libreria, ORANGE e IRIS
data('Orange') #vemos la data
head(Orange)
```

```
##   Tree  age circumference
## 1    1  118            30
## 2    1  484            58
## 3    1  664            87
## 4    1 1004           115
## 5    1 1231           120
## 6    1 1372           142
```

```
data('iris')
head(iris)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

```
#. (a) Buscar un resumen de las variables de cada dataset.
summary(Orange)
```

```
##  Tree       age         circumference
##  3:7   Min.   : 118.0   Min.   : 30.0
##  1:7   1st Qu.: 484.0   1st Qu.: 65.5
##  5:7   Median :1004.0   Median :115.0
##  2:7   Mean   : 922.1   Mean   :115.9
##  4:7   3rd Qu.:1372.0   3rd Qu.:161.5
##        Max.   :1582.0   Max.   :214.0
```

```r
summary(iris)
```

```
##   Sepal.Length    Sepal.Width     Petal.Length    Petal.Width
##  Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100
##  1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300
##  Median :5.800   Median :3.000   Median :4.350   Median :1.300
##  Mean   :5.843   Mean   :3.057   Mean   :3.758   Mean   :1.199
##  3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
##  Max.   :7.900   Max.   :4.400   Max.   :6.900   Max.   :2.500
##        Species
##  setosa    :50
##  versicolor:50
##  virginica :50
##
##
##
```

```r
summary_iris <- summary(iris)
summary_iris
```

```
##   Sepal.Length    Sepal.Width     Petal.Length    Petal.Width
##  Min.   :4.300   Min.   :2.000   Min.   :1.000   Min.   :0.100
##  1st Qu.:5.100   1st Qu.:2.800   1st Qu.:1.600   1st Qu.:0.300
##  Median :5.800   Median :3.000   Median :4.350   Median :1.300
##  Mean   :5.843   Mean   :3.057   Mean   :3.758   Mean   :1.199
##  3rd Qu.:6.400   3rd Qu.:3.300   3rd Qu.:5.100   3rd Qu.:1.800
##  Max.   :7.900   Max.   :4.400   Max.   :6.900   Max.   :2.500
##        Species
##  setosa    :50
##  versicolor:50
##  virginica :50
##
##
##
```

```r
#. (b) Generar una tabla de frecuencias absolutas y una de relativas con
#cualquier variable del dataset.
#frecuencia absoluta
table(Orange$age)
```

```
##
##  118  484  664 1004 1231 1372 1582
##    5    5    5    5    5    5    5
```

```r
table(iris$Sepal.Length)
```

```
##
## 4.3 4.4 4.5 4.6 4.7 4.8 4.9   5 5.1 5.2 5.3 5.4 5.5 5.6 5.7 5.8 5.9   6 6.1 6.2
##   1   3   1   4   2   5   6  10   9   4   1   6   7   6   8   7   3   6   6   4
## 6.3 6.4 6.5 6.6 6.7 6.8 6.9   7 7.1 7.2 7.3 7.4 7.6 7.7 7.9
##   9   7   5   2   8   3   4   1   1   3   1   1   1   4   1
```

```r
#frecuencia relativa
prop.table(Orange$age)
```

```
##  [1] 0.003656081 0.014996127 0.020573199 0.031107668 0.038140976 0.042509682
##  [7] 0.049016266 0.003656081 0.014996127 0.020573199 0.031107668 0.038140976
## [13] 0.042509682 0.049016266 0.003656081 0.014996127 0.020573199 0.031107668
## [19] 0.038140976 0.042509682 0.049016266 0.003656081 0.014996127 0.020573199
## [25] 0.031107668 0.038140976 0.042509682 0.049016266 0.003656081 0.014996127
## [31] 0.020573199 0.031107668 0.038140976 0.042509682 0.049016266
```

```r
prop.table(iris$Sepal.Length)
```

```
##   [1] 0.005818597 0.005590416 0.005362236 0.005248146 0.005704507 0.006160867
##   [7] 0.005248146 0.005704507 0.005019966 0.005590416 0.006160867 0.005476326
##  [13] 0.005476326 0.004905876 0.006617228 0.006503137 0.006160867 0.005818597
##  [19] 0.006503137 0.005818597 0.006160867 0.005818597 0.005248146 0.005818597
##  [25] 0.005476326 0.005704507 0.005704507 0.005932687 0.005932687 0.005362236
##  [31] 0.005476326 0.006160867 0.005932687 0.006274957 0.005590416 0.005704507
##  [37] 0.006274957 0.005590416 0.005019966 0.005818597 0.005704507 0.005134056
##  [43] 0.005019966 0.005704507 0.005818597 0.005476326 0.005818597 0.005248146
##  [49] 0.006046777 0.005704507 0.007986309 0.007301768 0.007872219 0.006274957
##  [55] 0.007415859 0.006503137 0.007187678 0.005590416 0.007529949 0.005932687
##  [61] 0.005704507 0.006731318 0.006845408 0.006959498 0.006389047 0.007644039
##  [67] 0.006389047 0.006617228 0.007073588 0.006389047 0.006731318 0.006959498
##  [73] 0.007187678 0.006959498 0.007301768 0.007529949 0.007758129 0.007644039
##  [79] 0.006845408 0.006503137 0.006274957 0.006274957 0.006617228 0.006845408
##  [85] 0.006160867 0.006845408 0.007644039 0.007187678 0.006389047 0.006274957
##  [91] 0.006274957 0.006959498 0.006617228 0.005704507 0.006389047 0.006503137
##  [97] 0.006503137 0.007073588 0.005818597 0.006503137 0.007187678 0.006617228
## [103] 0.008100399 0.007187678 0.007415859 0.008670850 0.005590416 0.008328580
## [109] 0.007644039 0.008214489 0.007415859 0.007301768 0.007758129 0.006503137
## [115] 0.006617228 0.007301768 0.007415859 0.008784940 0.008784940 0.006845408
## [121] 0.007872219 0.006389047 0.008784940 0.007187678 0.007644039 0.008214489
## [127] 0.007073588 0.006959498 0.007301768 0.008214489 0.008442670 0.009013120
## [133] 0.007301768 0.007187678 0.006959498 0.008784940 0.007187678 0.007301768
## [139] 0.006845408 0.007872219 0.007644039 0.007872219 0.006617228 0.007758129
## [145] 0.007644039 0.007644039 0.007187678 0.007415859 0.007073588 0.006731318
```

```r
#. (c) Generar una tabla de frecuencias absolutas con cada una de las variables
#del conjunto de datos Orange. ¿Todas las tablas generadas tienen sentido
#para vosotros?
table(Orange$Tree)
```

```
##
## 3 1 5 2 4
## 7 7 7 7 7
```

```r
table(Orange$age)
```

```
##
##  118  484  664 1004 1231 1372 1582
##    5    5    5    5    5    5    5
```

```r
table(Orange$circumference)
```

```
##
##  30  32  33  49  51  58  62  69  75  81  87 108 111 112 115 120 125 139 140 142
##   3   1   1   1   1   1   1   1   1   1   1   1   1   1   2   1   1   1   1   2
## 145 156 167 172 174 177 179 203 209 214
##   1   1   1   1   1   1   1   2   1   1
```

```r
#Con la frecuencia absoluta de three no obtenemos ningun dato concluyente
```

Ejercicio 4.En este ejercicio realizaremos paso a paso la creación de gráficos y modificación de sus opciones para que veáis cómo se pueden genera

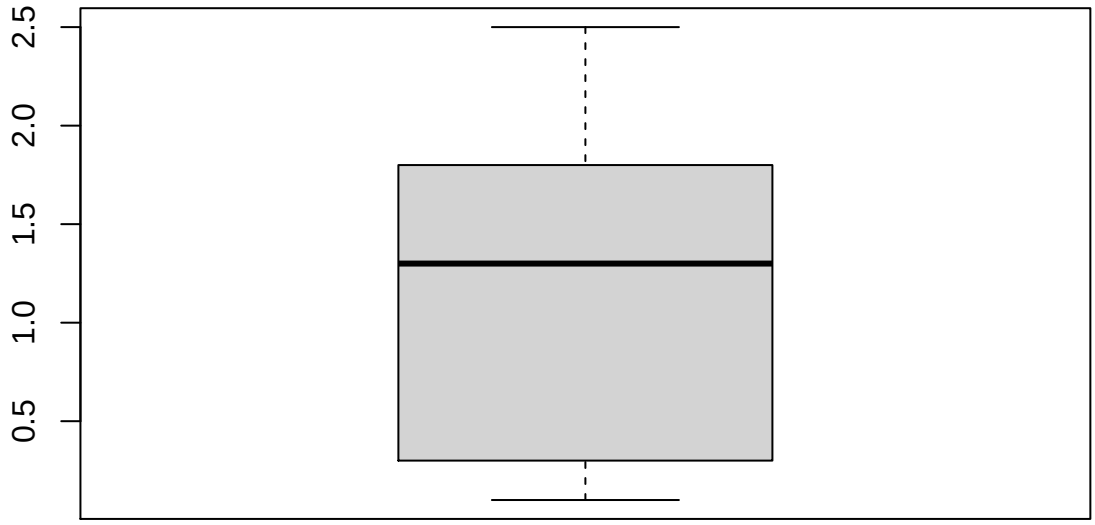
```r
library(datasets)
data('iris')
head(iris)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

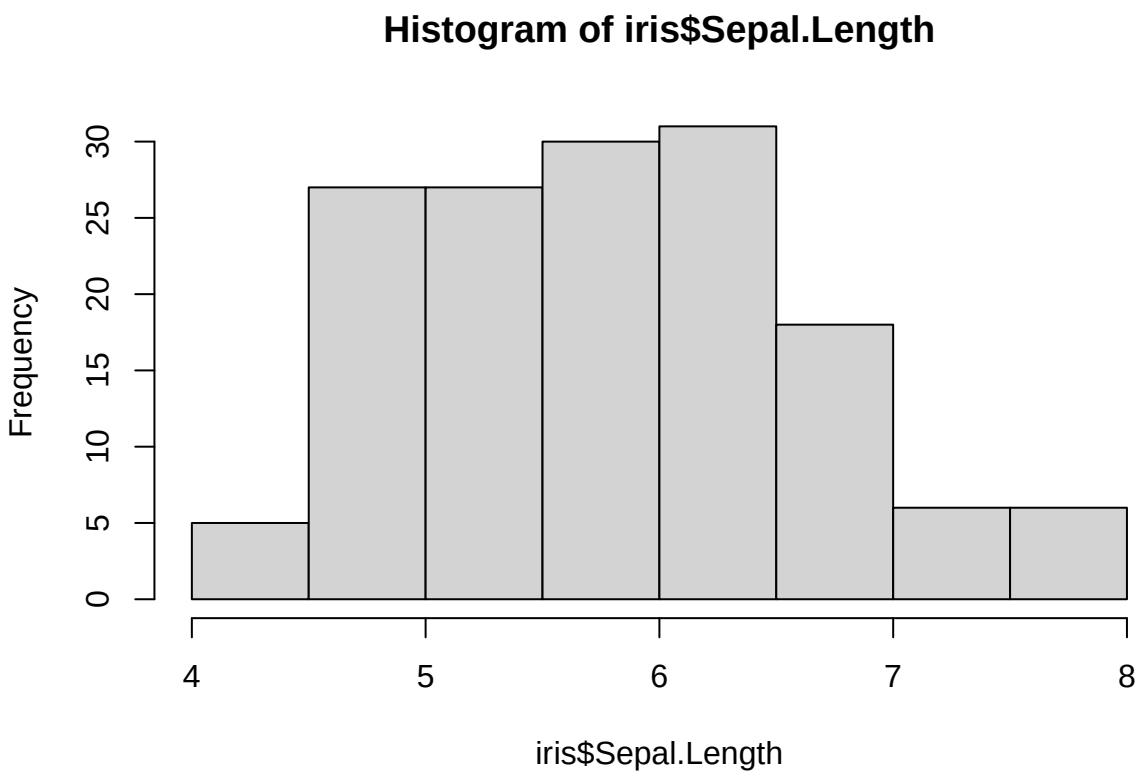
```r
stem(iris$Sepal.Length) #grafico de tallo y hojas
```

```
##
##   The decimal point is 1 digit(s) to the left of the |
##
##   42 | 0
##   44 | 0000
##   46 | 000000
##   48 | 00000000000
##   50 | 00000000000000000000
##   52 | 00000
##   54 | 0000000000000
##   56 | 00000000000000
##   58 | 0000000000
##   60 | 000000000000
##   62 | 0000000000000
##   64 | 000000000000
##   66 | 0000000000
##   68 | 0000000
##   70 | 00
##   72 | 0000
##   74 | 0
##   76 | 00000
##   78 | 0
```

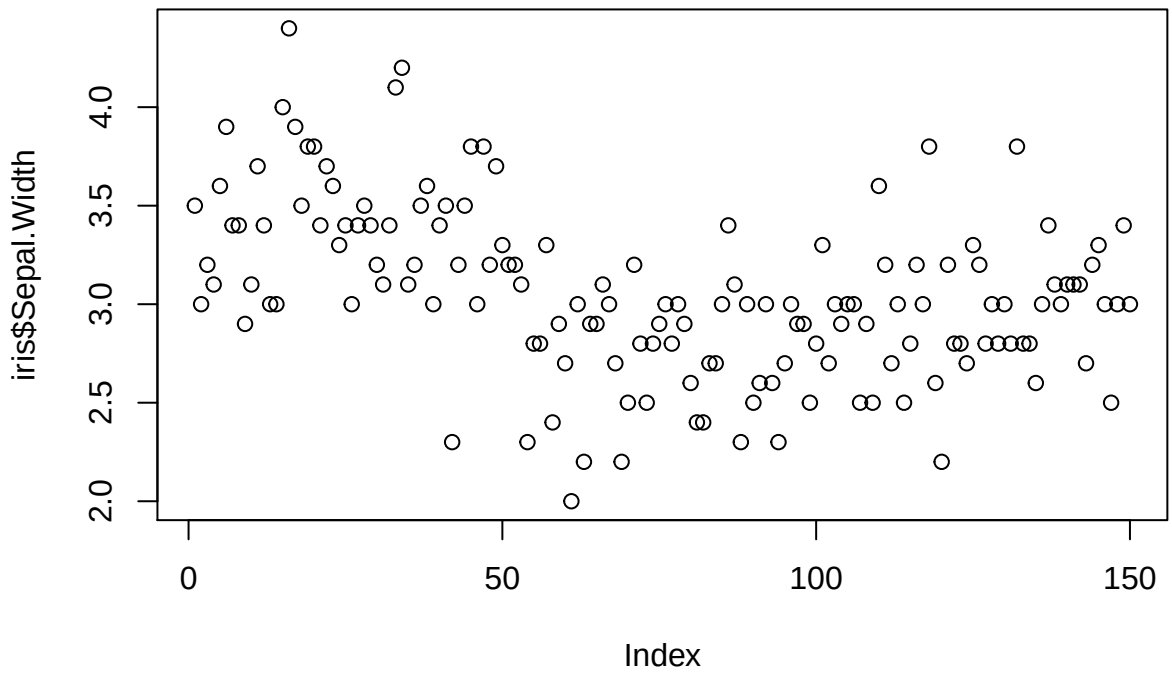
```r
boxplot(iris$Petal.Width) #grafico de cajas
```




```r
hist(iris$Sepal.Length) #grafico de histogramas
```



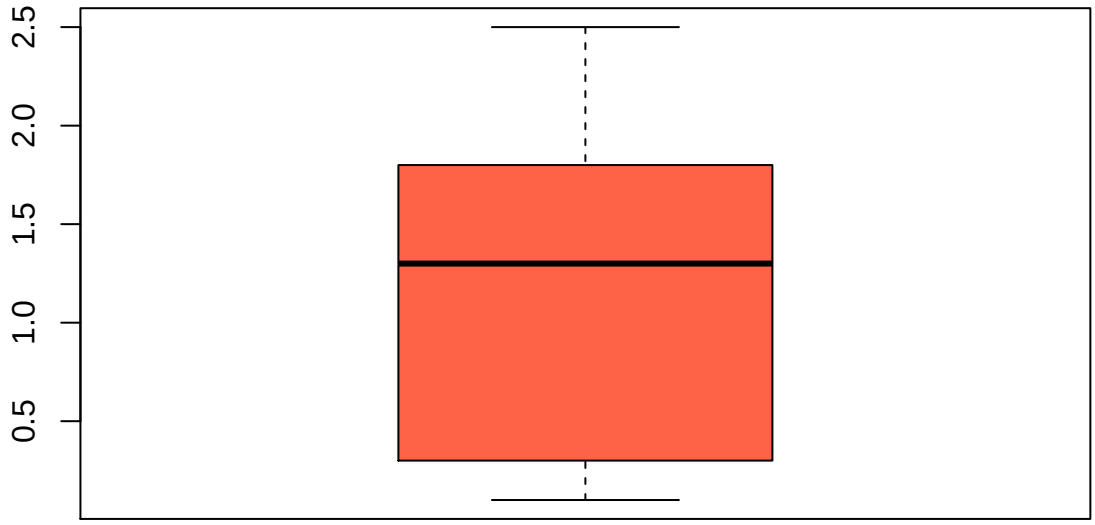

```r
plot(iris$Sepal.Width) #grafico de puntos
```



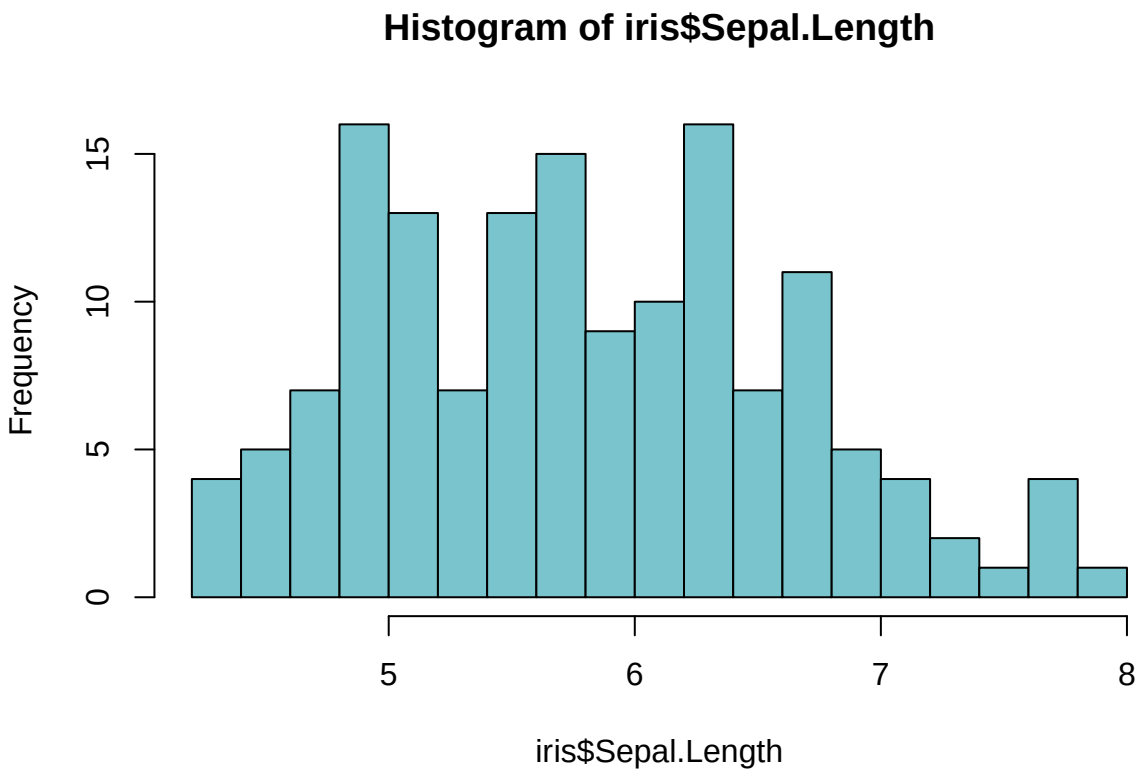

```
#modificaciones varias
stem(iris$Sepal.Length, scale =1, width =80)
```

```
##
##   The decimal point is 1 digit(s) to the left of the |
##
##   42 | 0
##   44 | 0000
##   46 | 000000
##   48 | 00000000000
##   50 | 00000000000000000000
##   52 | 00000
##   54 | 0000000000000
##   56 | 00000000000000
##   58 | 0000000000
##   60 | 000000000000
##   62 | 0000000000000
##   64 | 000000000000
##   66 | 0000000000
##   68 | 0000000
##   70 | 00
##   72 | 0000
##   74 | 0
##   76 | 00000
##   78 | 0
```

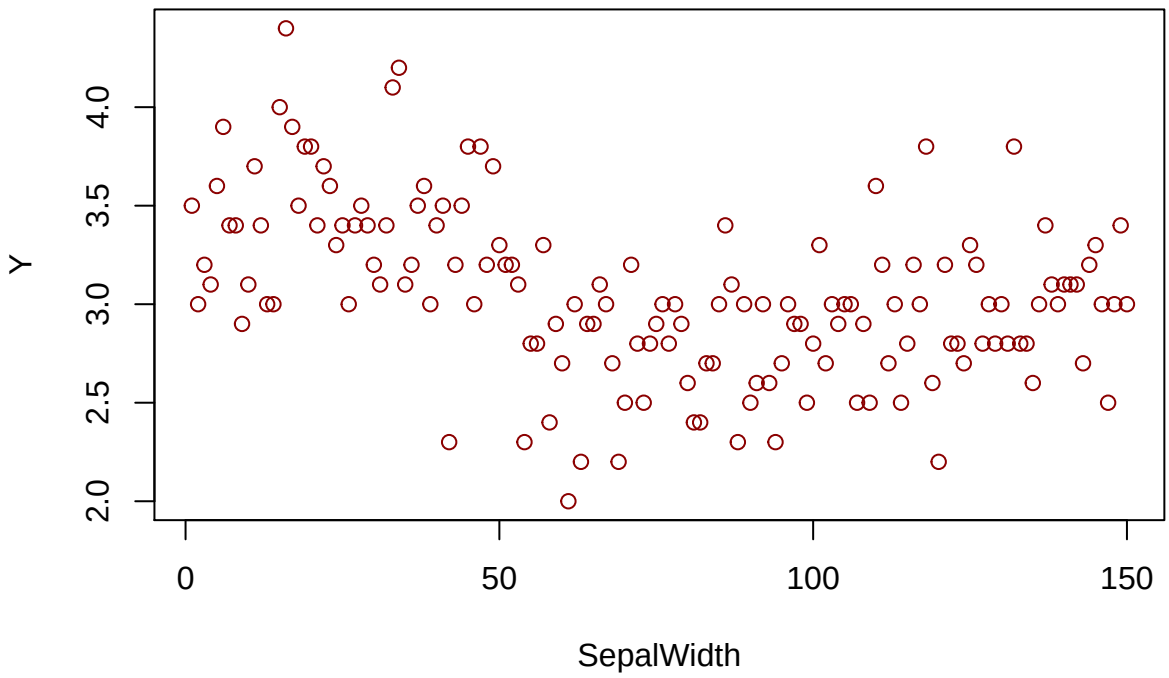
```r
boxplot(iris$Petal.Width, col='tomato', title='Diagrama de caja')
```




```r
hist(iris$Sepal.Length, breaks=20, col='cadetblue3')
```




```
plot(iris$Sepal.Width, col='dark red', xlab='SepalWidth', ylab='Y')
```

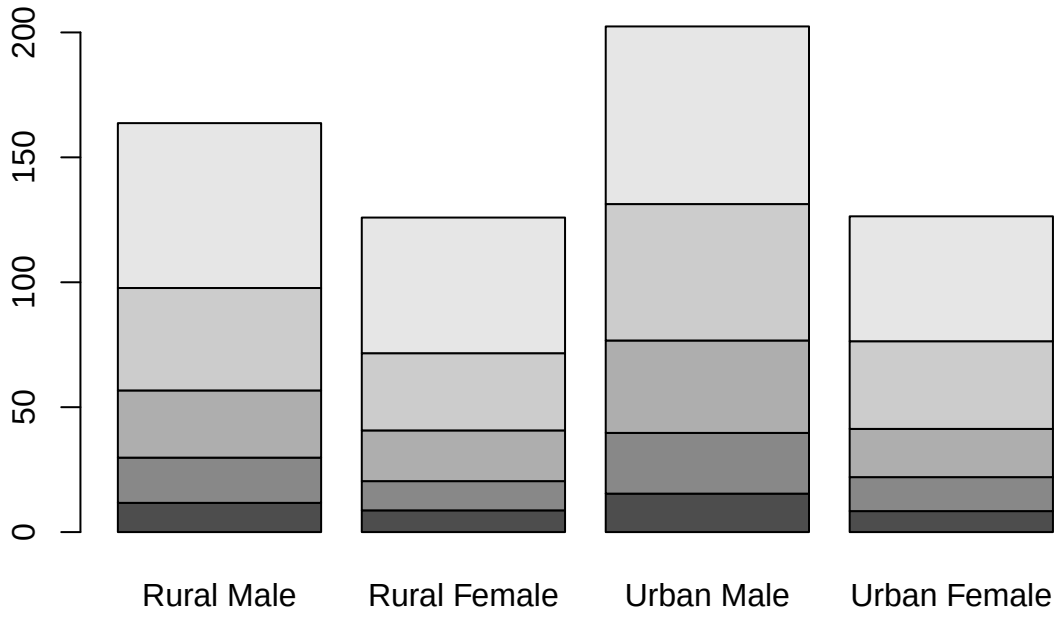


Ejercicio 5.

```
#1.a) Incorporad el dataset VaDeaths de R y generad un gráfico barplot.
library(datasets)
data("VADeaths")
head(VADeaths)
```

```
##       Rural Male Rural Female Urban Male Urban Female
## 50-54       11.7          8.7       15.4          8.4
## 55-59       18.1         11.7       24.3         13.6
## 60-64       26.9         20.3       37.0         19.3
## 65-69       41.0         30.9       54.6         35.1
## 70-74       66.0         54.3       71.1         50.0
```

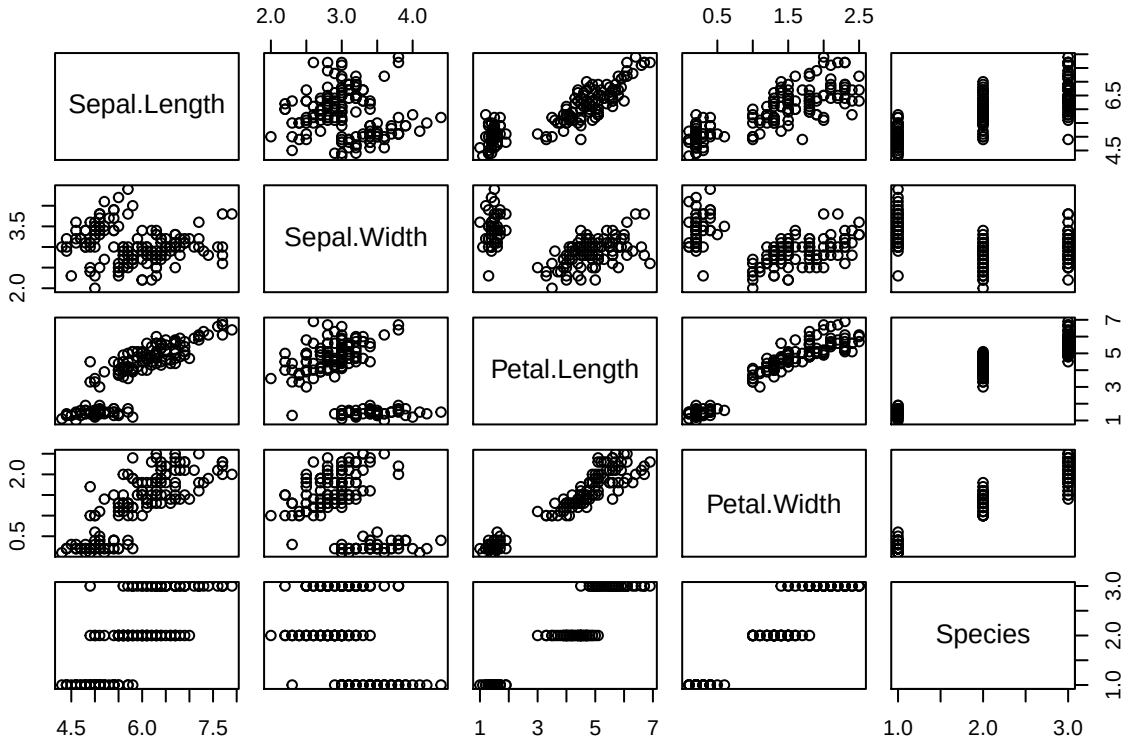
```
barplot(VADeaths)
```



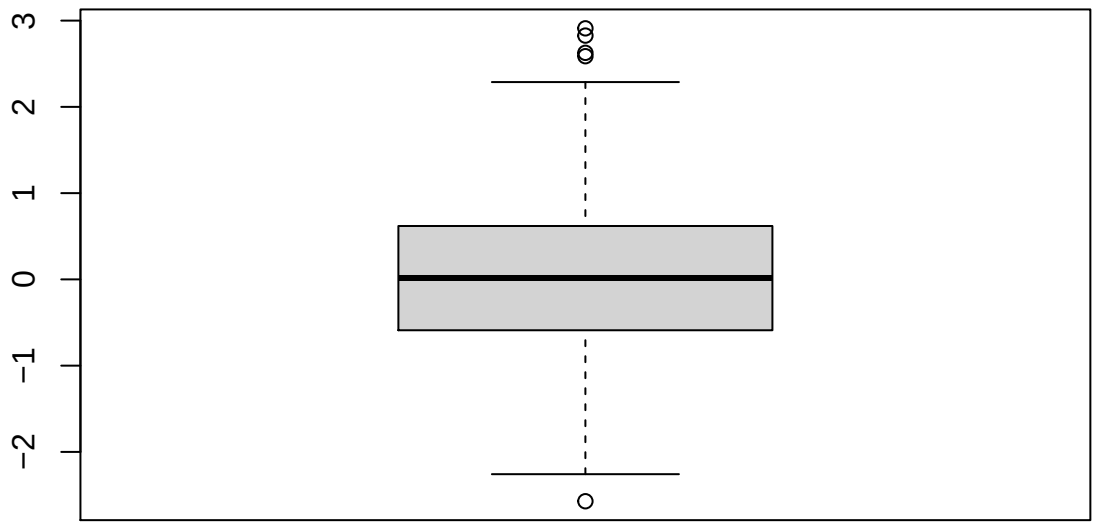

```
#1.b) Usad la función pairs sobre el conjunto de datos iris.
data("iris")
head(iris)
```

```
##   Sepal.Length Sepal.Width Petal.Length Petal.Width Species
## 1          5.1         3.5          1.4         0.2  setosa
## 2          4.9         3.0          1.4         0.2  setosa
## 3          4.7         3.2          1.3         0.2  setosa
## 4          4.6         3.1          1.5         0.2  setosa
## 5          5.0         3.6          1.4         0.2  setosa
## 6          5.4         3.9          1.7         0.4  setosa
```

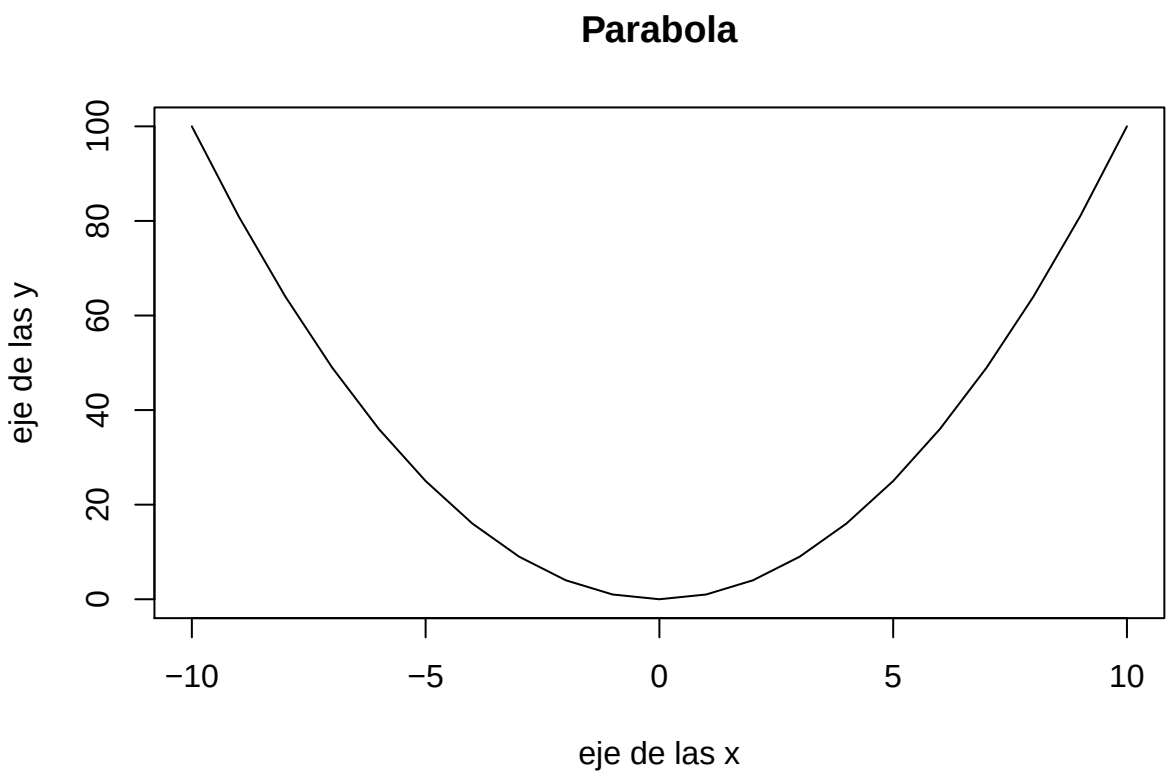
```
pairs(iris)
```



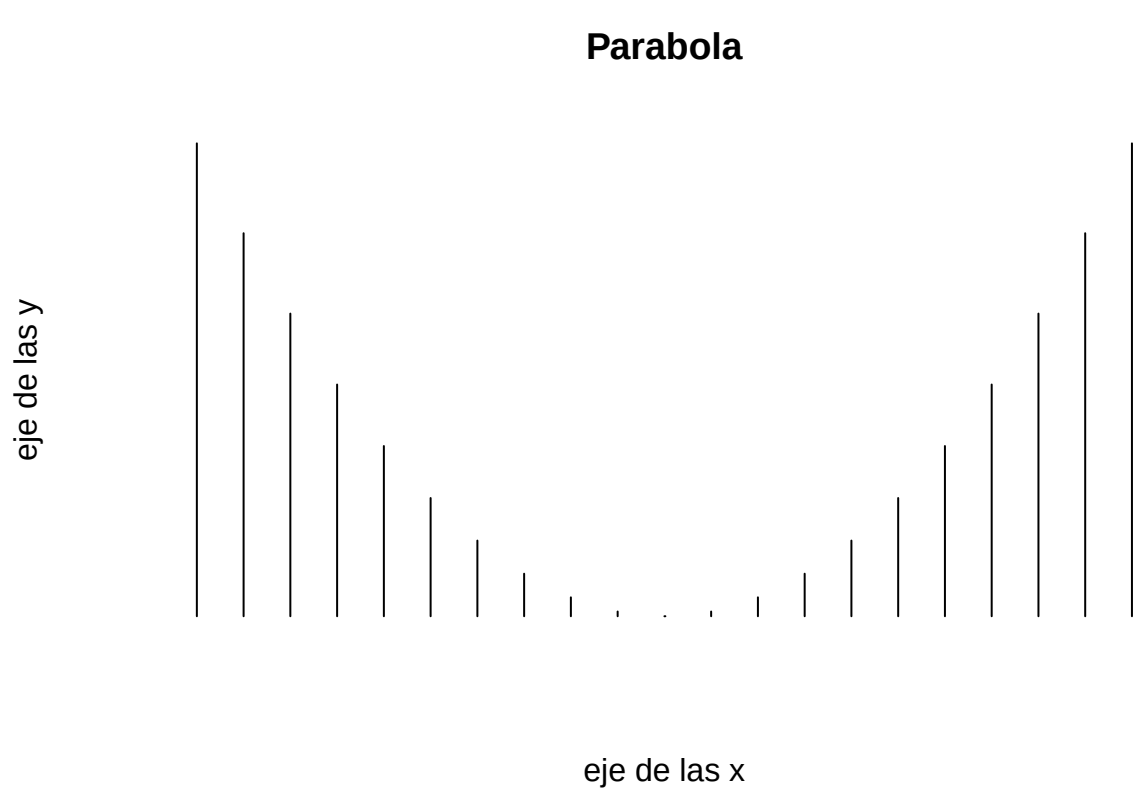

```
#1.c) Generad unos datos inventados y cread un boxplot con ellos.
x = rnorm(300)
boxplot(x)
```



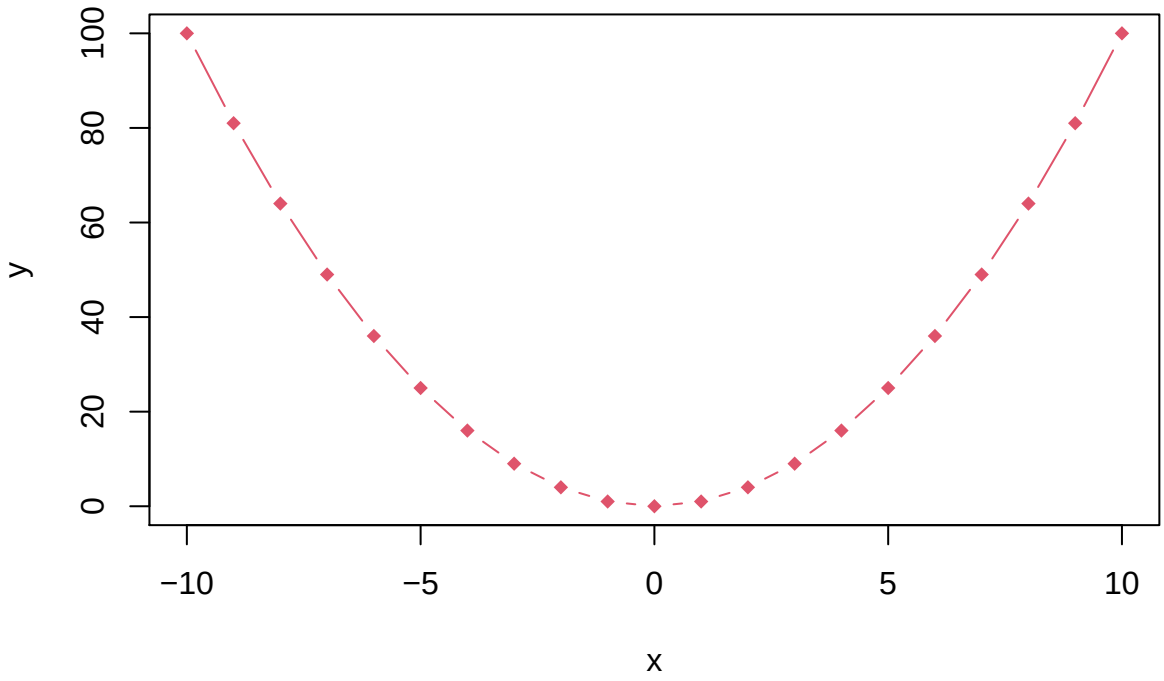

```
#1.d) Dibujad una parábola y=x^2 con valores que van de x entre -10 y 10.
x = seq(-10, 10)
y = x^2
plot(x, y, type='l', xlab='eje de las x', ylab='eje de las y', main='Parabola')
```




```
plot(x, y, type='h', xlab='eje de las x', ylab='eje de las y', main='Parabola',
     axes=F)
```




```
plot(x, y, pch=18, col=2, type='b')
```



Ejercicio 6.

```
#2.a)Cargad (o instalad primero y luego cargad) el paquete UsingR. El conjunto
#de datos brightness (incorporado en este paquete) contiene información sobre el
#brillo de 963 estrellas.Representad estos datos mediante un histograma y un
#gráfico de densidad superpuesto. Combinad los dos gráficos.
install.packages("UsingR",repos = "http://cran.us.r-project.org")
```

```
## Installing package into 'C:/Users/lsudu/Documents/R/win-library/4.0'
## (as 'lib' is unspecified)
```

```
## package 'UsingR' successfully unpacked and MD5 sums checked
##
## The downloaded binary packages are in
##  C:\Users\lsudu\AppData\Local\Temp\Rtmp4wwo59\downloaded_packages
```

```
#install.packages('UsingR')
library(UsingR)
```

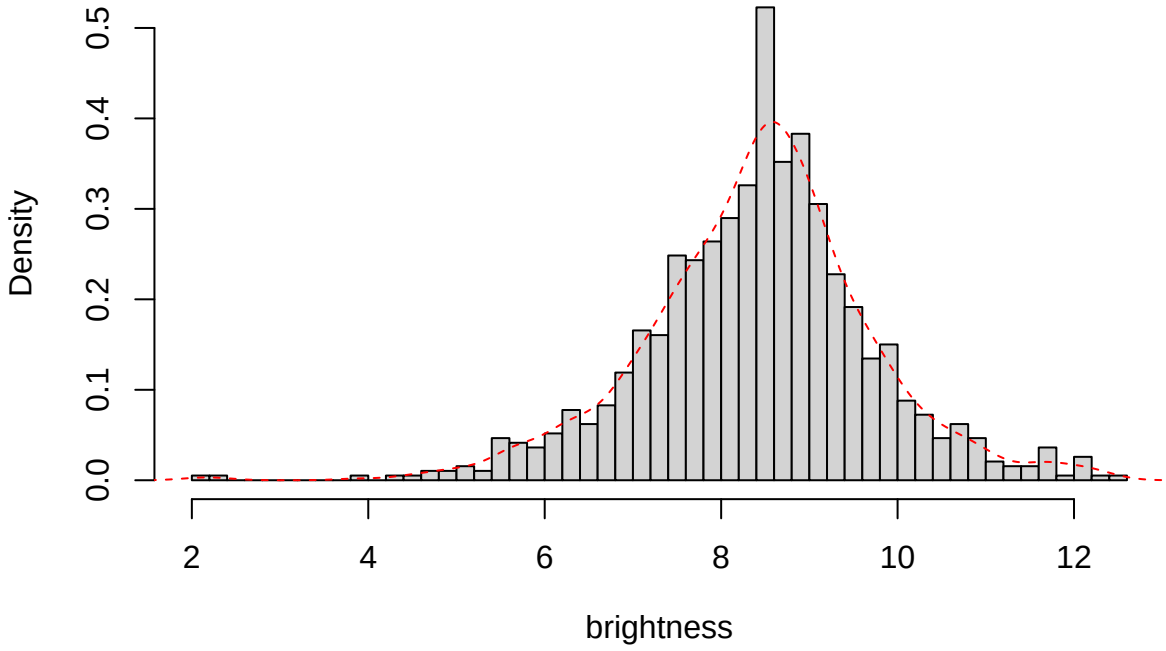
```
## Loading required package: MASS
```

```
## Loading required package: HistData
```

```
## Loading required package: Hmisc
```

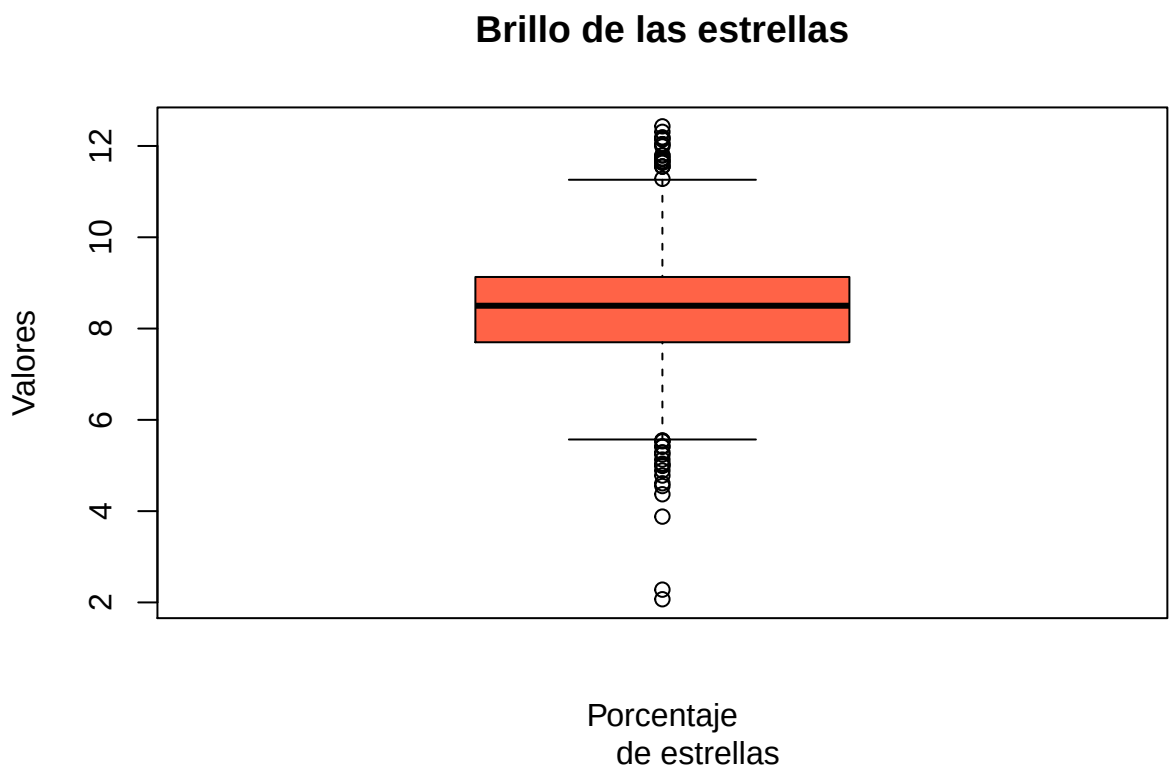
```
## Loading required package: lattice

## Loading required package: survival

## Loading required package: Formula

## Loading required package: ggplot2

##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':
##
##     format.pval, units

##
## Attaching package: 'UsingR'

## The following object is masked from 'package:survival':
##
##     cancer
```

```
data(package='UsingR')
data("brightness")
Bright <- data.frame(brightness)
hist(Bright, freq = F, main = 'Histograma', xlab='Brillo de las estrellas',
     ylab= 'Numero de estrellas')
lines(density(Bright$brightness),lty=2, col='red')
```



n:966 m:0

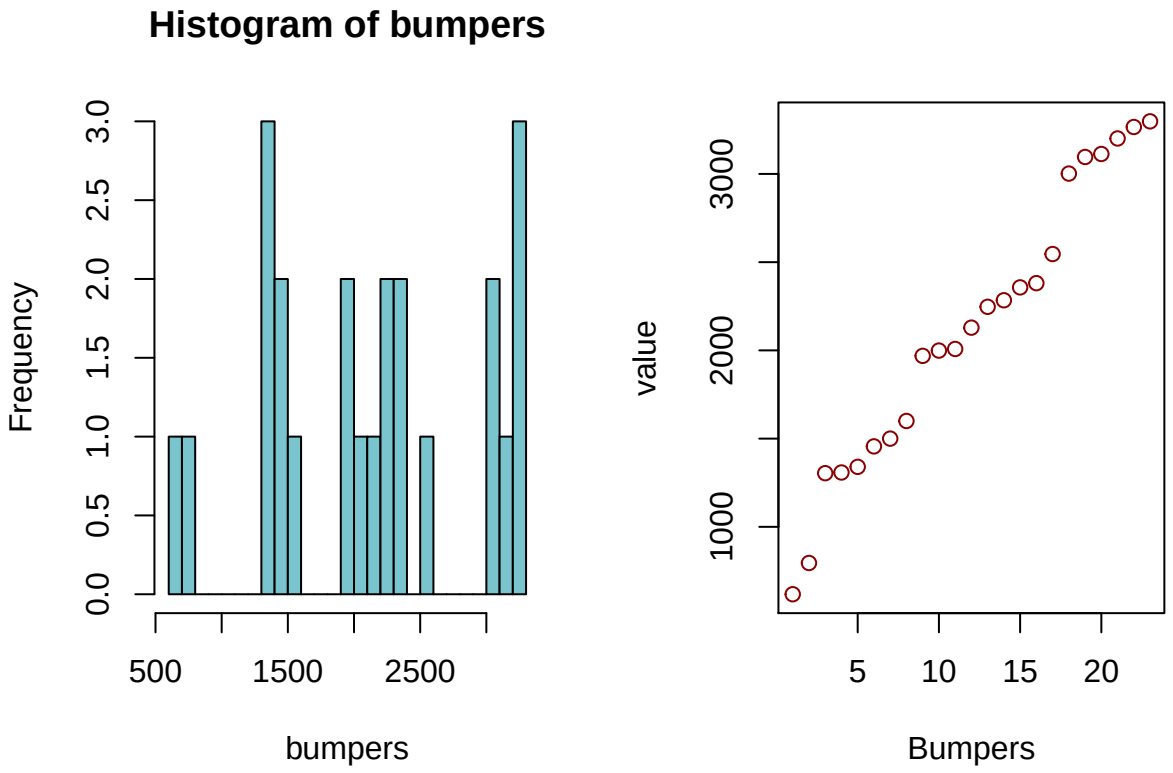
```
#2.b)Representad gráficamente estos datos mediante un diagrama de caja (boxplot)

boxplot(brightness, data=Bright, main='Brillo de las estrellas',xlab='Porcentaje
        de estrellas', ylab='Valores',col='tomato')
```



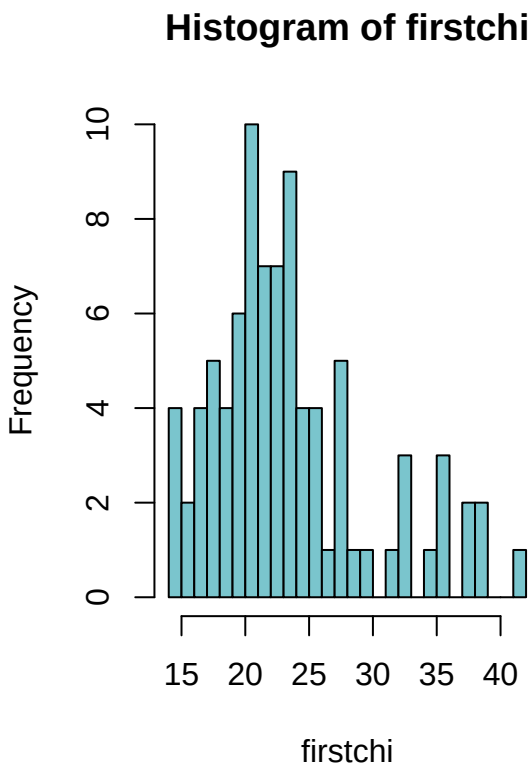

```
#2.c)¿Diríais que los datos presentan «outliers»?
#Si, hay varios datos alejados
```

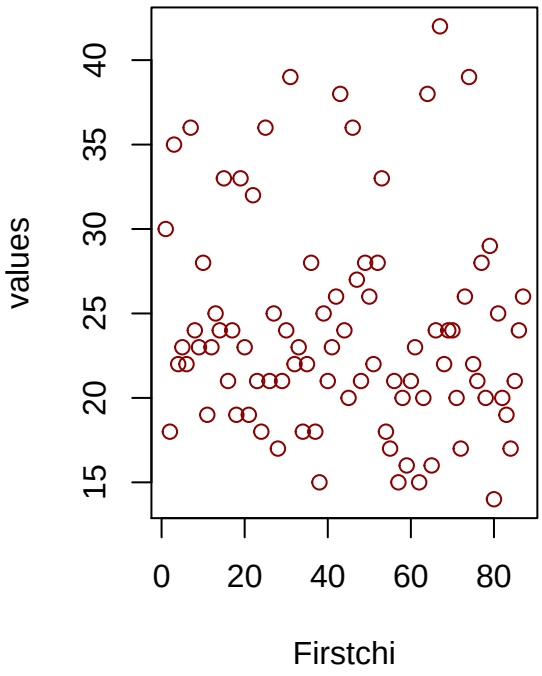
Ejercicio 7.Representad gráficamente los datos contenidos en los conjuntos de datos (incorporados en UsingR) bumpers, firstchi, math con un histograma y/o un boxplot usando ggplot2.

```
library(UsingR)
a <- data.frame(bumpers)
b <- data.frame(firstchi)
c <- data.frame(math)

par(mfrow=c(1,2))
hist(bumpers, breaks=30, col='cadetblue3')
plot(bumpers, col='dark red', xlab='Bumpers', ylab='value')
```

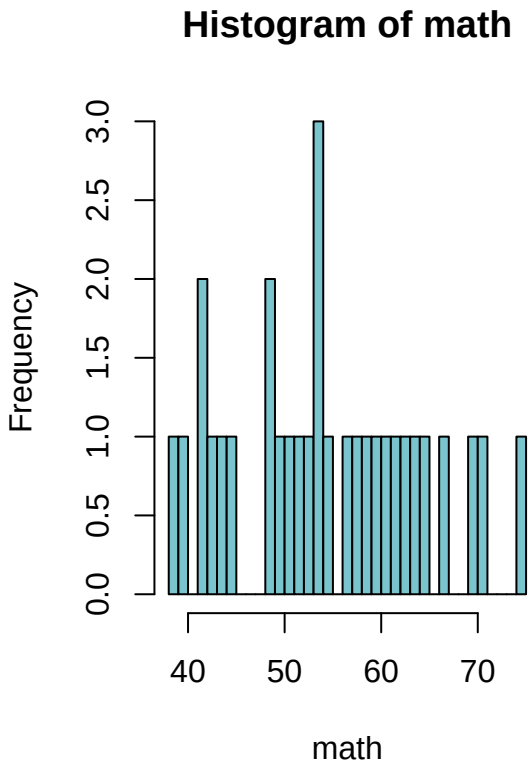


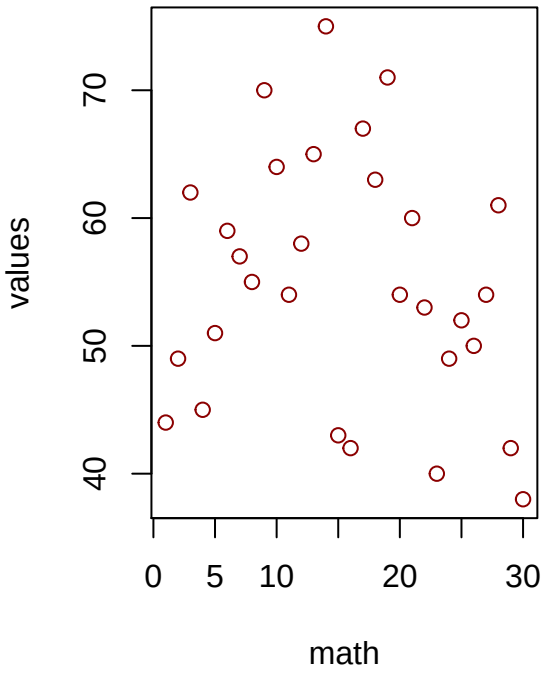

```
par(mfrow=c(1,2))
hist(firstchi, breaks=30, col='cadetblue3')
plot(firstchi, col='dark red', xlab='Firstchi', ylab='values')
```






```
par(mfrow=c(1,2))
hist(math, breaks=30, col = 'cadetblue3')
plot(math, col='dark red', xlab='math', ylab='values')
```


Histogram of math
Frequency
0.0
0.5
1.0
1.5
2.0
2.5
3.0
40
50
60
70
math


values
40
50
60
70
0
5
10
20
30
math