Ejercicio 9 LAB2

Laura Sudupe

12/10/2020

A partir de unos datos bioclínicos o biosanitarios que escojáis y que importéis a R, explicad sus variables (un mínimo de 8) y:

He elegido un data set de 1988 que contiene el mismo tipo de informción que el data set del ejercicio 8 pero de año y población disntinto. https://www.kaggle.com/johnsmith88/heart-disease-dataset?select=heart.csv

```
health_data <- read.csv("/Users/lsudu/code/rlab/LAB2/heart_ejercicio9.csv")
health_data$sex <- factor(health_data$sex, levels=c(0,1),
                          labels=c("men", "women"))
health_data$thal <- factor(health_data$thal, levels=c(1,2,3),
                          labels=c("normal", "fixed defect",
                                   "reversable defect"))
head(health_data)
```

```
##   age   sex cp trestbps chol fbs restecg thalach exang oldpeak slope ca
## 1  52 women  0      125  212   0       1     168     0     1.0     2  2
## 2  53 women  0      140  203   1       0     155     1     3.1     0  0
## 3  70 women  0      145  174   0       1     125     1     2.6     0  0
## 4  61 women  0      148  203   0       1     161     0     0.0     2  1
## 5  62   men  0      138  294   1       1     106     0     1.9     1  3
## 6  58   men  0      100  248   0       0     122     0     1.0     1  0
##                thal target
## 1 reversable defect      0
## 2 reversable defect      0
## 3 reversable defect      0
## 4 reversable defect      0
## 5      fixed defect      0
## 6      fixed defect      1
```

Tenemos 13 variables age sex chest pain type (4 values) resting blood pressure serum cholestoral in mg/dl fasting blood sugar > 120 mg/dl resting electrocardiographic results (values 0,1,2) maximum heart rate achieved exercise induced angina oldpeak = ST depression induced by exercise relative to rest the slope of the peak exercise ST segment number of major vessels (0-3) colored by flourosopy thal: 0 = normal; 1 = fixed defect; 2 = reversable defect

- (a) Realizad un resumen estadístico completo del dataset y explicad los resultados.

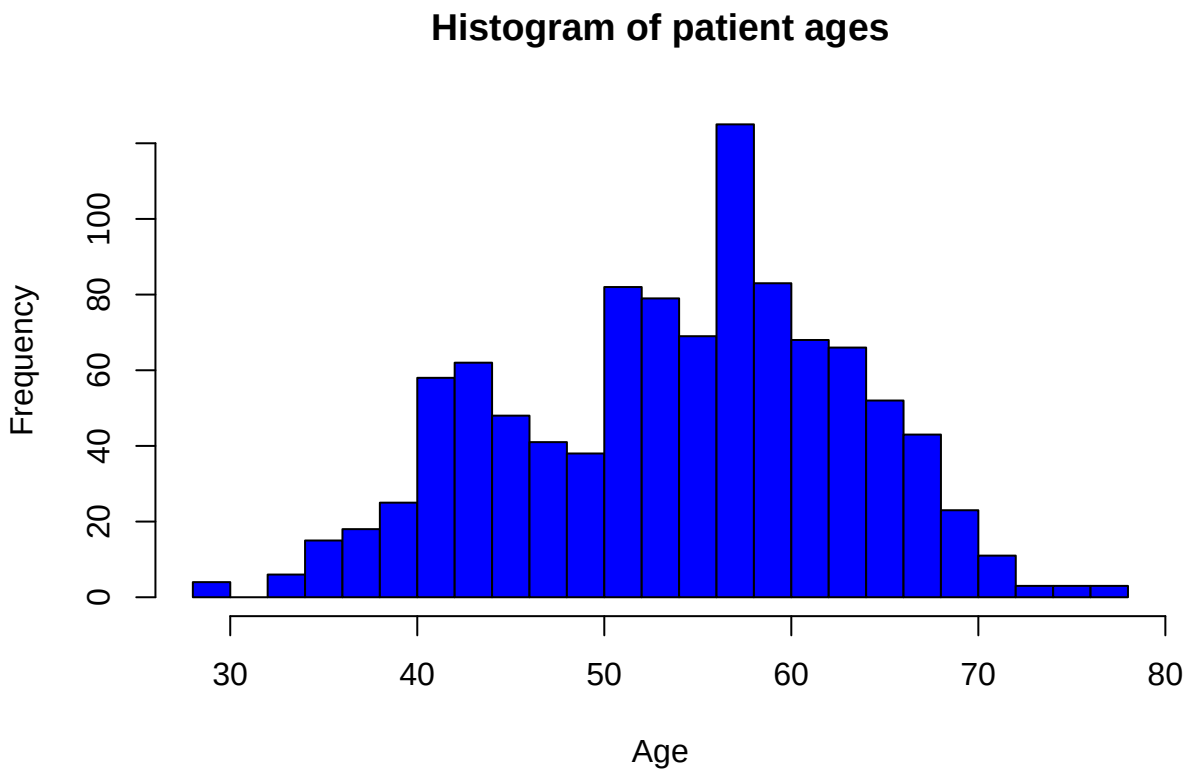
```
summary(health_data)
```

```
##       age           sex            cp            trestbps          chol
##  Min.   :29.00   men  :312   Min.   :0.0000   Min.   : 94.0   Min.   :126
```

```
##  1st Qu.:48.00   women:713   1st Qu.:0.0000   1st Qu.:120.0   1st Qu.:211
##  Median :56.00               Median :1.0000   Median :130.0   Median :240
##  Mean   :54.43               Mean   :0.9424   Mean   :131.6   Mean   :246
##  3rd Qu.:61.00               3rd Qu.:2.0000   3rd Qu.:140.0   3rd Qu.:275
##  Max.   :77.00               Max.   :3.0000   Max.   :200.0   Max.   :564
##       fbs            restecg          thalach          exang
##  Min.   :0.0000   Min.   :0.0000   Min.   : 71.0   Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:132.0   1st Qu.:0.0000
##  Median :0.0000   Median :1.0000   Median :152.0   Median :0.0000
##  Mean   :0.1493   Mean   :0.5298   Mean   :149.1   Mean   :0.3366
##  3rd Qu.:0.0000   3rd Qu.:1.0000   3rd Qu.:166.0   3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :2.0000   Max.   :202.0   Max.   :1.0000
##     oldpeak          slope             ca                         thal
##  Min.   :0.000   Min.   :0.000   Min.   :0.0000   normal           : 64
##  1st Qu.:0.000   1st Qu.:1.000   1st Qu.:0.0000   fixed defect     :544
##  Median :0.800   Median :1.000   Median :0.0000   reversable defect:410
##  Mean   :1.072   Mean   :1.385   Mean   :0.7541   NA's             :  7
##  3rd Qu.:1.800   3rd Qu.:2.000   3rd Qu.:1.0000
##  Max.   :6.200   Max.   :2.000   Max.   :4.0000
##      target
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :1.0000
##  Mean   :0.5132
##  3rd Qu.:1.0000
##  Max.   :1.0000
```

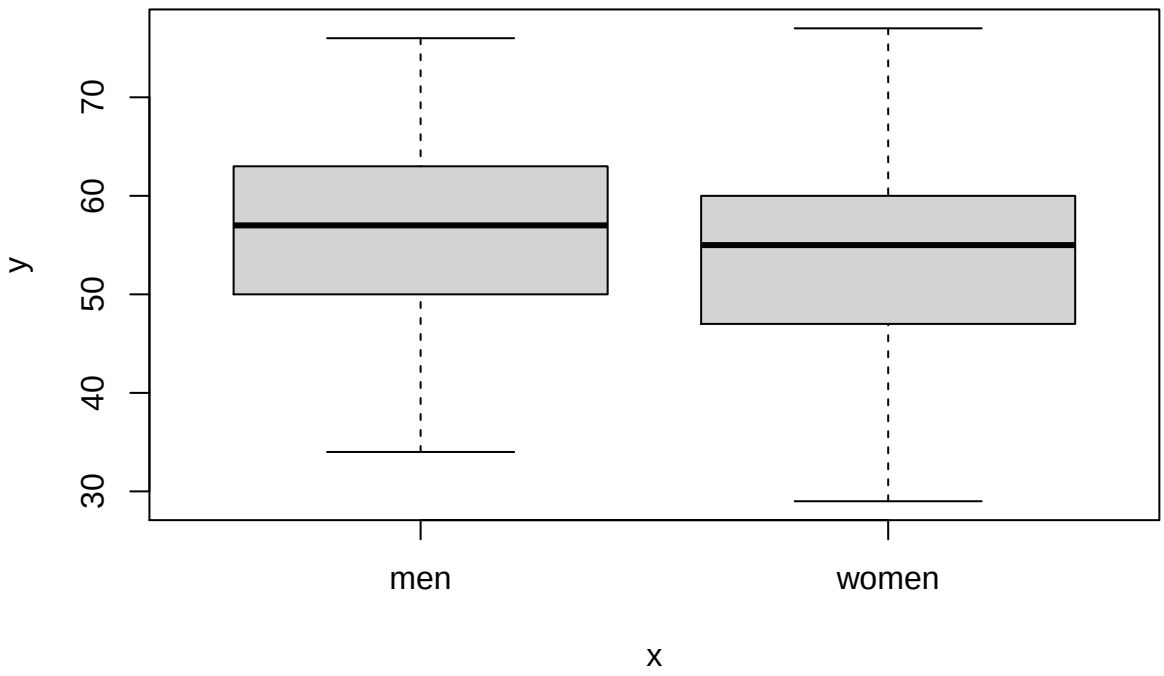
- (b) Realizad 5 gráficos con las variables, explicad su significado y guardadlos como imágenes (jpeg o bmp).

```
hist(health_data$age, main = "Histogram of patient ages",
     breaks=20,col="blue", xlab="Age")
```



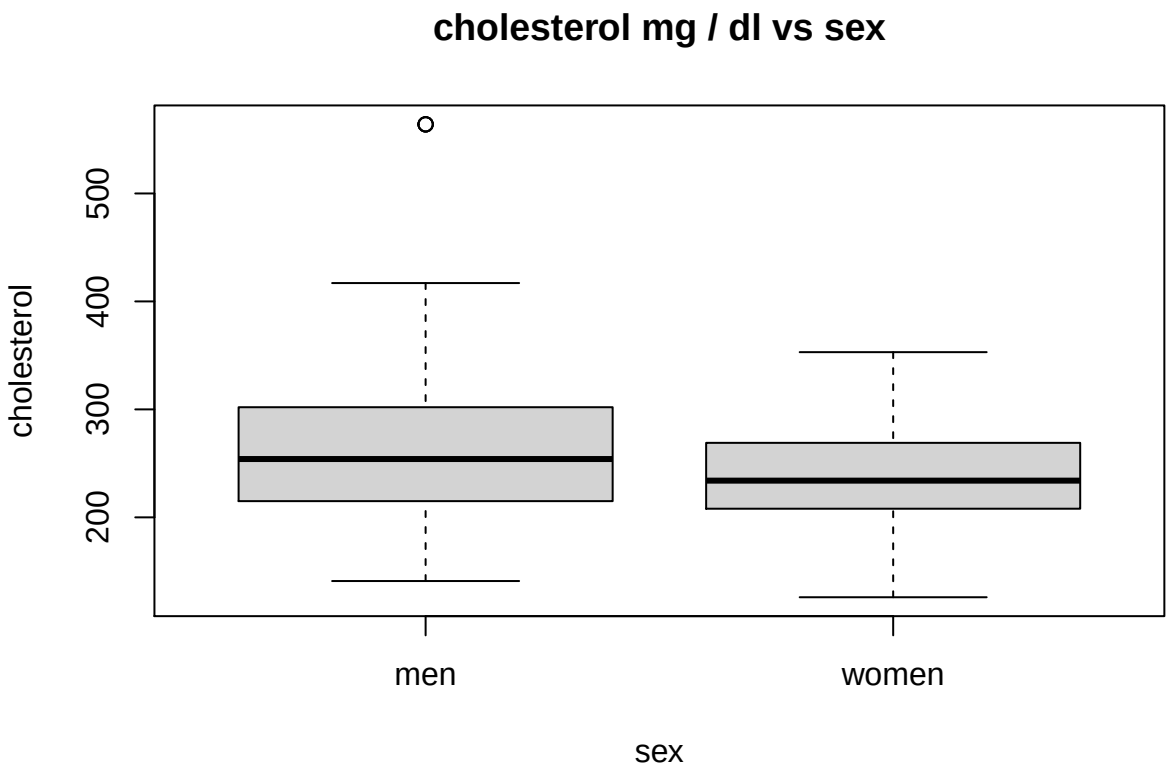

```
#Vemos que la edad que mas se repite en las muestras esta entre 50-60 años
```

```
plot(health_data$sex , health_data$age)
```

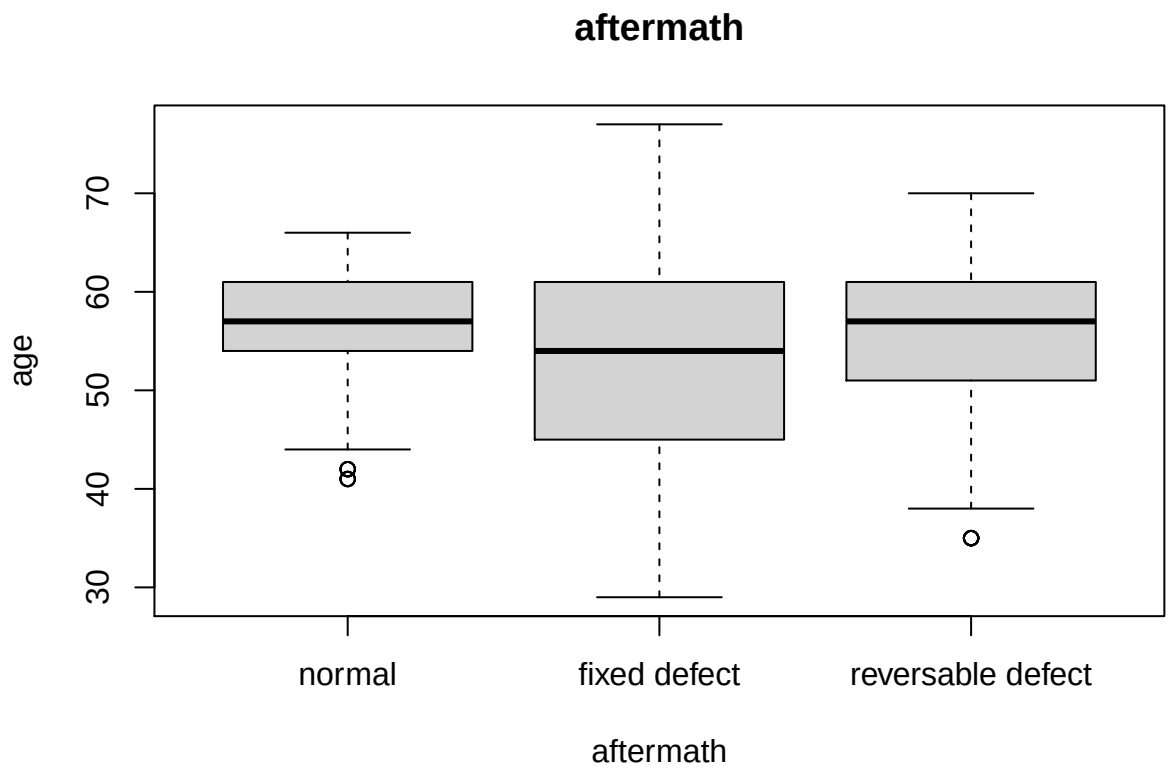



```
plot(health_data$age, health_data$thalach,main = "Heart rate vs age",
     xlab="age", ylab="max heart rate")
```


```
plot(health_data$sex, health_data$chol, main="cholesterol mg / dl vs sex",
     xlab="sex", ylab="cholesterol")
```




```
plot(health_data$thal, health_data$age, main="aftermath",
     xlab="aftermath", ylab="age")
```



- (c) Generad una regresión lineal entre 2 de sus variables paso a paso y comentad los resultados obtenidos.

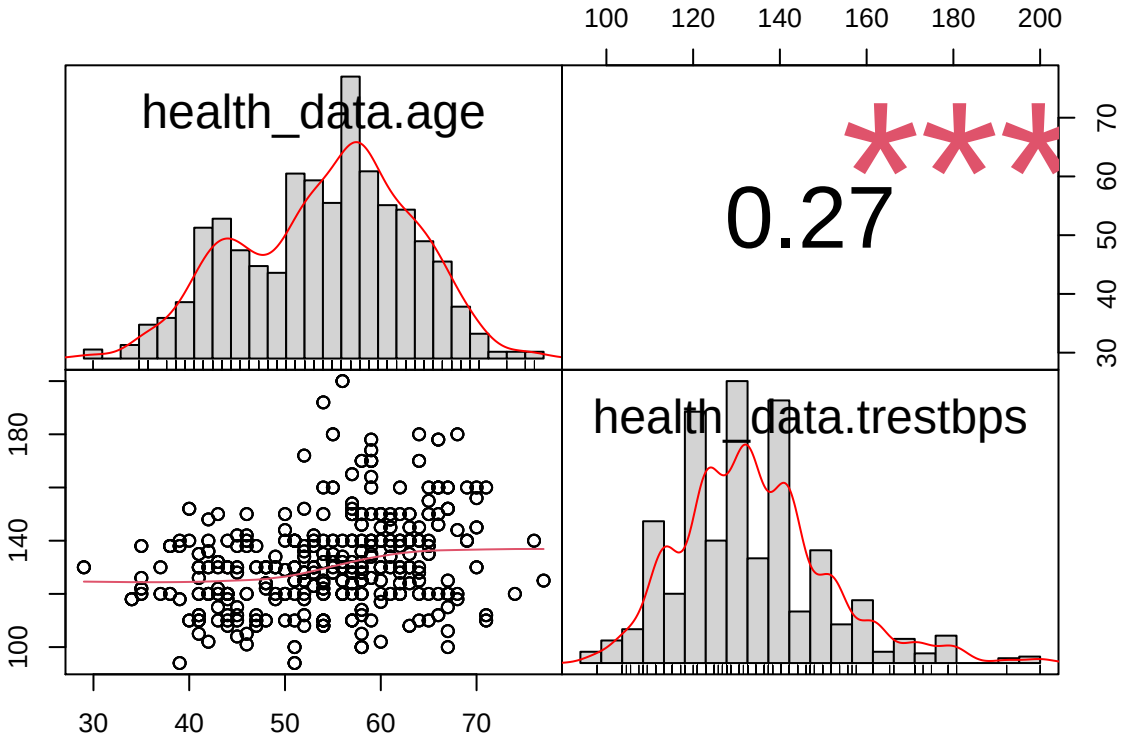
```
cor(health_data$age, health_data$trestbps)
```

```
## [1] 0.2711214
```

```
library(PerformanceAnalytics)
```

```
## Loading required package: xts

## Loading required package: zoo

##
## Attaching package: 'zoo'

## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

##
## Attaching package: 'PerformanceAnalytics'

## The following object is masked from 'package:graphics':
##
##     legend
```

```
data <- data.frame(health_data$age, health_data$trestbps)
chart.Correlation(data)
```



Vemos que no hay mucha relacion entre la edad y la pression de la sangre en descanso. Pero si hay una correlación

```
model <- lm(formula= health_data$age~health_data$trestbps, data=health_data)
summary(model)
```

```
##
## Call:
## lm(formula = health_data$age ~ health_data$trestbps, data = health_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -25.2078  -6.2078   0.7922   6.3197  23.4943
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)           35.95327    2.06943  17.373   <2e-16 ***
## health_data$trestbps   0.14042    0.01559   9.009   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 8.737 on 1023 degrees of freedom
## Multiple R-squared:  0.07351,    Adjusted R-squared:  0.0726
```

```
## F-statistic: 81.16 on 1 and 1023 DF,  p-value: < 2.2e-16
```

```r
plot(health_data$age, health_data$trestbps, xlab="age",
     ylab="resting blood pressure")
abline(model)
```