PEC1

Laura Sudupe

13/10/2020

EJERCICIO 1. Resuelve las siguientes cuestiones, mostrando las instrucciones de código utilizadas así como el resultado de la ejecución de dicho código. 1.1. Importa los datos del fichero Health_heart.csv correspondientes a la probabilidad de sufrir un ataque al corazón. Guarda estos datos en un data frame llamado health_heart y muestra los primeros y últimos registros del conjunto de datos.

```
health_heart <- read.csv("/Users/lsudu/code/rlab/PEC1/heart.csv")
head(health_heart)
```

```
##   ï..age sex cp trestbps chol fbs restecg thalach exang oldpeak slope ca thal
## 1     63   1  3      145  233   1       0     150     0     2.3     0  0    1
## 2     37   1  2      130  250   0       1     187     0     3.5     0  0    2
## 3     41   0  1      130  204   0       0     172     0     1.4     2  0    2
## 4     56   1  1      120  236   0       1     178     0     0.8     2  0    2
## 5     57   0  0      120  354   0       1     163     1     0.6     2  0    2
## 6     57   1  0      140  192   0       1     148     0     0.4     1  0    1
##   target
## 1      1
## 2      1
## 3      1
## 4      1
## 5      1
## 6      1
```

```
tail(health_heart)
```

```
##     ï..age sex cp trestbps chol fbs restecg thalach exang oldpeak slope ca thal
## 298     59   1  0      164  176   1       0      90     0     1.0     1  2    1
## 299     57   0  0      140  241   0       1     123     1     0.2     1  0    3
## 300     45   1  3      110  264   0       1     132     0     1.2     1  0    3
## 301     68   1  0      144  193   1       1     141     0     3.4     1  2    3
## 302     57   1  0      130  131   0       1     115     1     1.2     1  1    3
## 303     57   0  1      130  236   0       0     174     0     0.0     1  1    2
##     target
## 298      0
## 299      0
## 300      0
## 301      0
## 302      0
## 303      0
```

1.2. A partir del data frame definido, health_heart, muestra algunas características como: a) Nombre de las variables que forman el conjunto de datos.

```r
names(health_heart)
```

```
##  [1] "ï..age"   "sex"      "cp"       "trestbps" "chol"     "fbs"
##  [7] "restecg"  "thalach"  "exang"    "oldpeak"  "slope"    "ca"
## [13] "thal"     "target"
```

b) Estructura del conjunto de datos

```r
str(health_heart)
```

```
## 'data.frame':    303 obs. of  14 variables:
##  $ ï..age  : int  63 37 41 56 57 57 56 44 52 57 ...
##  $ sex     : int  1 1 0 1 0 1 0 1 1 1 ...
##  $ cp      : int  3 2 1 1 0 0 1 1 2 2 ...
##  $ trestbps: int  145 130 130 120 120 140 140 120 172 150 ...
##  $ chol    : int  233 250 204 236 354 192 294 263 199 168 ...
##  $ fbs     : int  1 0 0 0 0 0 0 0 1 0 ...
##  $ restecg : int  0 1 0 1 1 1 0 1 1 1 ...
##  $ thalach : int  150 187 172 178 163 148 153 173 162 174 ...
##  $ exang   : int  0 0 0 0 1 0 0 0 0 0 ...
##  $ oldpeak : num  2.3 3.5 1.4 0.8 0.6 0.4 1.3 0 0.5 1.6 ...
##  $ slope   : int  0 0 2 2 2 1 1 2 2 2 ...
##  $ ca      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ thal    : int  1 2 2 2 2 1 2 3 3 2 ...
##  $ target  : int  1 1 1 1 1 1 1 1 1 1 ...
```

c) Tamaño de la muestra y número de variables

```r
dim(health_heart)
```

```
## [1] 303  14
```

d) ¿Existen valores nulos en el conjunto de datos?

```r
table(is.null(health_heart))
```

```
##
## FALSE
##     1
```

e) ¿Existen datos perdidos (missing values) en la tabla?

```r
table(is.na(health_heart))
```

```
##
## FALSE
##  4242
```

EJERCICIO 2. Definid un data frame, dataHealthHeart_w, y otro data frame, dataHealthHeart_m, que corresponderán a los conjuntos de datos de las mujeres y de los hombres, respectivamente. Posteriormente, exportad ambos data frame a dos ficheros (HealthHeart_w,HealthHeart_m) de tipo csv.

```
#health_heart <- read.csv("/Users/lsudu/code/rlab/LAB2/PEC1/heart.csv")
health_heart$sex <- factor(health_heart$sex, levels=c("0","1"),
                       labels=c("male","female"))
#Hice el ejercicio pensando que 0 se le asignaba al hombre y 1 a la mujer

dataHealthHeart_w <- data.frame(subset(health_heart,
                                      health_heart$sex == "female"))
dataHealthHeart_m <- data.frame(subset(health_heart,
                                health_heart$sex == "male"))

write.csv(dataHealthHeart_w, file=
            "/Users/lsudu/code/rlab/PEC1/HealthHeart_w.csv")

write.csv(dataHealthHeart_m, file =
            "/Users/lsudu/code/rlab/PEC1/HealthHeart_m.csv")
```

EJERCICIO 3. Considerando el data frame health_heart, resolver las siguientes cuestiones: a) ¿Cuál es la media de edad, que definiremos como age_mean, de los pacientes de la muestra de datos?

```
#health_heart <- read.csv("/Users/lsudu/code/rlab/LAB2/PEC1/heart.csv")

#primero voy a renombrar la columna 'ï..age'
#install.packages("tidyverse")
library(plyr)
health_heart <- rename(health_heart, c("ï..age" = "age" ))

age_mean <- mean(health_heart$age)
```

b) Definid una variable, age_max_trestbps, que guarde la edad del paciente que registra el máximo valor de presión arterial.

```
age_max_trestbps <- health_heart$age[max(health_heart$trestbps)]
```

c) Mostrad los/as pacientes cuya probabilidad de tener un ataque al corazón (target=0(menor probabilidad), target=1 (mayor probabilidad)) considerando aquellos/as pacientes con presión arterial (trestbps) superior a la media.

```
high_heart_attack_prob <- health_heart$target[health_heart$trestbps >
                                                  mean(health_heart$trestbps)]
high_heart_attack_prob
```

```
##   [1] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1
##  [38] 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 0 0 0 0 0 0 0 0
##  [75] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
## [112] 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
```

d) Definid un data frame, heartAttack_chance_0, que contenga los/as pacientes cuya probabilidad de sufrir un ataque al corazón sea baja (target=0). A partir de este data frame, definid un vector que contenga los valores máximos de las variables edad "age", colesterol "chol" y pulsaciones "thalach".

```r
heartAttack_chance_0 <- subset(health_heart, health_heart$target == "0")
heartAttack_chance_0
```

```
##     age    sex cp trestbps chol fbs restecg thalach exang oldpeak slope ca thal
## 166  67 female  0      160  286   0       0     108     1     1.5     1  3    2
## 167  67 female  0      120  229   0       0     129     1     2.6     1  2    3
## 168  62   male  0      140  268   0       0     160     0     3.6     0  2    2
## 169  63 female  0      130  254   0       0     147     0     1.4     1  1    3
## 170  53 female  0      140  203   1       0     155     1     3.1     0  0    3
## 171  56 female  2      130  256   1       0     142     1     0.6     1  1    1
## 172  48 female  1      110  229   0       1     168     0     1.0     0  0    3
## 173  58 female  1      120  284   0       0     160     0     1.8     1  0    2
## 174  58 female  2      132  224   0       0     173     0     3.2     2  2    3
## 175  60 female  0      130  206   0       0     132     1     2.4     1  2    3
## 176  40 female  0      110  167   0       0     114     1     2.0     1  0    3
## 177  60 female  0      117  230   1       1     160     1     1.4     2  2    3
## 178  64 female  2      140  335   0       1     158     0     0.0     2  0    2
## 179  43 female  0      120  177   0       0     120     1     2.5     1  0    3
## 180  57 female  0      150  276   0       0     112     1     0.6     1  1    1
## 181  55 female  0      132  353   0       1     132     1     1.2     1  1    3
## 182  65   male  0      150  225   0       0     114     0     1.0     1  3    3
## 183  61   male  0      130  330   0       0     169     0     0.0     2  0    2
## 184  58 female  2      112  230   0       0     165     0     2.5     1  1    3
## 185  50 female  0      150  243   0       0     128     0     2.6     1  0    3
## 186  44 female  0      112  290   0       0     153     0     0.0     2  1    2
## 187  60 female  0      130  253   0       1     144     1     1.4     2  1    3
## 188  54 female  0      124  266   0       0     109     1     2.2     1  1    3
## 189  50 female  2      140  233   0       1     163     0     0.6     1  1    3
## 190  41 female  0      110  172   0       0     158     0     0.0     2  0    3
## 191  51   male  0      130  305   0       1     142     1     1.2     1  0    3
## 192  58 female  0      128  216   0       0     131     1     2.2     1  3    3
## 193  54 female  0      120  188   0       1     113     0     1.4     1  1    3
## 194  60 female  0      145  282   0       0     142     1     2.8     1  2    3
## 195  60 female  2      140  185   0       0     155     0     3.0     1  0    2
## 196  59 female  0      170  326   0       0     140     1     3.4     0  0    3
## 197  46 female  2      150  231   0       1     147     0     3.6     1  0    2
## 198  67 female  0      125  254   1       1     163     0     0.2     1  2    3
## 199  62 female  0      120  267   0       1      99     1     1.8     1  2    3
## 200  65 female  0      110  248   0       0     158     0     0.6     2  2    1
## 201  44 female  0      110  197   0       0     177     0     0.0     2  1    2
## 202  60 female  0      125  258   0       0     141     1     2.8     1  1    3
## 203  58 female  0      150  270   0       0     111     1     0.8     2  0    3
## 204  68 female  2      180  274   1       0     150     1     1.6     1  0    3
## 205  62   male  0      160  164   0       0     145     0     6.2     0  3    3
## 206  52 female  0      128  255   0       1     161     1     0.0     2  1    3
## 207  59 female  0      110  239   0       0     142     1     1.2     1  1    3
## 208  60   male  0      150  258   0       0     157     0     2.6     1  2    3
## 209  49 female  2      120  188   0       1     139     0     2.0     1  3    3
## 210  59 female  0      140  177   0       1     162     1     0.0     2  1    3
## 211  57 female  2      128  229   0       0     150     0     0.4     1  1    3
## 212  61 female  0      120  260   0       1     140     1     3.6     1  1    3
## 213  39 female  0      118  219   0       1     140     0     1.2     1  0    3
## 214  61   male  0      145  307   0       0     146     1     1.0     1  0    3
```

```
## 215  56 female  0       125  249   1       0      144      1      1.2      1  1    2
## 216  43   male  0       132  341   1       0      136      1      3.0      1  0    3
## 217  62   male  2       130  263   0       1       97      0      1.2      1  1    3
## 218  63 female  0       130  330   1       0      132      1      1.8      2  3    3
## 219  65 female  0       135  254   0       0      127      0      2.8      1  1    3
## 220  48 female  0       130  256   1       0      150      1      0.0      2  2    3
## 221  63   male  0       150  407   0       0      154      0      4.0      1  3    3
## 222  55 female  0       140  217   0       1      111      1      5.6      0  0    3
## 223  65 female  3       138  282   1       0      174      0      1.4      1  1    2
## 224  56   male  0       200  288   1       0      133      1      4.0      0  2    3
## 225  54 female  0       110  239   0       1      126      1      2.8      1  1    3
## 226  70 female  0       145  174   0       1      125      1      2.6      0  0    3
## 227  62 female  1       120  281   0       0      103      0      1.4      1  1    3
## 228  35 female  0       120  198   0       1      130      1      1.6      1  0    3
## 229  59 female  3       170  288   0       0      159      0      0.2      1  0    3
## 230  64 female  2       125  309   0       1      131      1      1.8      1  0    3
## 231  47 female  2       108  243   0       1      152      0      0.0      2  0    2
## 232  57 female  0       165  289   1       0      124      0      1.0      1  3    3
## 233  55 female  0       160  289   0       0      145      1      0.8      1  1    3
## 234  64 female  0       120  246   0       0       96      1      2.2      0  1    2
## 235  70 female  0       130  322   0       0      109      0      2.4      1  3    2
## 236  51 female  0       140  299   0       1      173      1      1.6      2  0    3
## 237  58 female  0       125  300   0       0      171      0      0.0      2  2    3
## 238  60 female  0       140  293   0       0      170      0      1.2      1  2    3
## 239  77 female  0       125  304   0       0      162      1      0.0      2  3    2
## 240  35 female  0       126  282   0       0      156      1      0.0      2  0    3
## 241  70 female  2       160  269   0       1      112      1      2.9      1  1    3
## 242  59   male  0       174  249   0       1      143      1      0.0      1  0    2
## 243  64 female  0       145  212   0       0      132      0      2.0      1  2    1
## 244  57 female  0       152  274   0       1       88      1      1.2      1  1    3
## 245  56 female  0       132  184   0       0      105      1      2.1      1  1    1
## 246  48 female  0       124  274   0       0      166      0      0.5      1  0    3
## 247  56   male  0       134  409   0       0      150      1      1.9      1  2    3
## 248  66 female  1       160  246   0       1      120      1      0.0      1  3    1
## 249  54 female  1       192  283   0       0      195      0      0.0      2  1    3
## 250  69 female  2       140  254   0       0      146      0      2.0      1  3    3
## 251  51 female  0       140  298   0       1      122      1      4.2      1  3    3
## 252  43 female  0       132  247   1       0      143      1      0.1      1  4    3
## 253  62   male  0       138  294   1       1      106      0      1.9      1  3    2
## 254  67 female  0       100  299   0       0      125      1      0.9      1  2    2
## 255  59 female  3       160  273   0       0      125      0      0.0      2  0    2
## 256  45 female  0       142  309   0       0      147      1      0.0      1  3    3
## 257  58 female  0       128  259   0       0      130      1      3.0      1  2    3
## 258  50 female  0       144  200   0       0      126      1      0.9      1  0    3
## 259  62   male  0       150  244   0       1      154      1      1.4      1  0    2
## 260  38 female  3       120  231   0       1      182      1      3.8      1  0    3
## 261  66   male  0       178  228   1       1      165      1      1.0      1  2    3
## 262  52 female  0       112  230   0       1      160      0      0.0      2  1    2
## 263  53 female  0       123  282   0       1       95      1      2.0      1  2    3
## 264  63   male  0       108  269   0       1      169      1      1.8      1  2    2
## 265  54 female  0       110  206   0       0      108      1      0.0      1  1    2
## 266  66 female  0       112  212   0       0      132      1      0.1      2  1    2
## 267  55   male  0       180  327   0       2      117      1      3.4      1  0    2
## 268  49 female  2       118  149   0       0      126      0      0.8      2  3    2
```

```
## 269  54 female  0       122  286   0        0      116      1      3.2      1  2    2
## 270  56 female  0       130  283   1        0      103      1      1.6      0  0    3
## 271  46 female  0       120  249   0        0      144      0      0.8      2  0    3
## 272  61 female  3       134  234   0        1      145      0      2.6      1  2    2
## 273  67 female  0       120  237   0        1       71      0      1.0      1  0    2
## 274  58 female  0       100  234   0        1      156      0      0.1      2  1    3
## 275  47 female  0       110  275   0        0      118      1      1.0      1  1    2
## 276  52 female  0       125  212   0        1      168      0      1.0      2  2    3
## 277  58 female  0       146  218   0        1      105      0      2.0      1  1    3
## 278  57 female  1       124  261   0        1      141      0      0.3      2  0    3
## 279  58   male  1       136  319   1        0      152      0      0.0      2  2    2
## 280  61 female  0       138  166   0        0      125      1      3.6      1  1    2
## 281  42 female  0       136  315   0        1      125      1      1.8      1  0    1
## 282  52 female  0       128  204   1        1      156      1      1.0      1  0    0
## 283  59 female  2       126  218   1        1      134      0      2.2      1  1    1
## 284  40 female  0       152  223   0        1      181      0      0.0      2  0    3
## 285  61 female  0       140  207   0        0      138      1      1.9      2  1    3
## 286  46 female  0       140  311   0        1      120      1      1.8      1  2    3
## 287  59 female  3       134  204   0        1      162      0      0.8      2  2    2
## 288  57 female  1       154  232   0        0      164      0      0.0      2  1    2
## 289  57 female  0       110  335   0        1      143      1      3.0      1  1    3
## 290  55   male  0       128  205   0        2      130      1      2.0      1  1    3
## 291  61 female  0       148  203   0        1      161      0      0.0      2  1    3
## 292  58 female  0       114  318   0        2      140      0      4.4      0  3    1
## 293  58   male  0       170  225   1        0      146      1      2.8      1  2    1
## 294  67 female  2       152  212   0        0      150      0      0.8      1  0    3
## 295  44 female  0       120  169   0        1      144      1      2.8      0  0    1
## 296  63 female  0       140  187   0        0      144      1      4.0      2  2    3
## 297  63   male  0       124  197   0        1      136      1      0.0      1  0    2
## 298  59 female  0       164  176   1        0       90      0      1.0      1  2    1
## 299  57   male  0       140  241   0        1      123      1      0.2      1  0    3
## 300  45 female  3       110  264   0        1      132      0      1.2      1  0    3
## 301  68 female  0       144  193   1        1      141      0      3.4      1  2    3
## 302  57 female  0       130  131   0        1      115      1      1.2      1  1    3
## 303  57   male  1       130  236   0        0      174      0      0.0      1  1    2
##     target
## 166      0
## 167      0
## 168      0
## 169      0
## 170      0
## 171      0
## 172      0
## 173      0
## 174      0
## 175      0
## 176      0
## 177      0
## 178      0
## 179      0
## 180      0
## 181      0
## 182      0
## 183      0
```

```
## 184           0
## 185           0
## 186           0
## 187           0
## 188           0
## 189           0
## 190           0
## 191           0
## 192           0
## 193           0
## 194           0
## 195           0
## 196           0
## 197           0
## 198           0
## 199           0
## 200           0
## 201           0
## 202           0
## 203           0
## 204           0
## 205           0
## 206           0
## 207           0
## 208           0
## 209           0
## 210           0
## 211           0
## 212           0
## 213           0
## 214           0
## 215           0
## 216           0
## 217           0
## 218           0
## 219           0
## 220           0
## 221           0
## 222           0
## 223           0
## 224           0
## 225           0
## 226           0
## 227           0
## 228           0
## 229           0
## 230           0
## 231           0
## 232           0
## 233           0
## 234           0
## 235           0
## 236           0
## 237           0
```

```
## 238           0
## 239           0
## 240           0
## 241           0
## 242           0
## 243           0
## 244           0
## 245           0
## 246           0
## 247           0
## 248           0
## 249           0
## 250           0
## 251           0
## 252           0
## 253           0
## 254           0
## 255           0
## 256           0
## 257           0
## 258           0
## 259           0
## 260           0
## 261           0
## 262           0
## 263           0
## 264           0
## 265           0
## 266           0
## 267           0
## 268           0
## 269           0
## 270           0
## 271           0
## 272           0
## 273           0
## 274           0
## 275           0
## 276           0
## 277           0
## 278           0
## 279           0
## 280           0
## 281           0
## 282           0
## 283           0
## 284           0
## 285           0
## 286           0
## 287           0
## 288           0
## 289           0
## 290           0
## 291           0
```

```
## 292      0
## 293      0
## 294      0
## 295      0
## 296      0
## 297      0
## 298      0
## 299      0
## 300      0
## 301      0
## 302      0
## 303      0
```

```
vector <- c(max(heartAttack_chance_0$age), max(heartAttack_chance_0$chol),
            max(heartAttack_chance_0$thalach))
vector
```

```
## [1]  77 409 195
```

e) Definid una matriz que muestre las columnas "age" y "sex".

```
matrix_health <- matrix(c(health_heart$age, health_heart$sex), ncol=2)
head(matrix_health)
```

```
##      [,1] [,2]
## [1,]   63    2
## [2,]   37    2
## [3,]   41    1
## [4,]   56    2
## [5,]   57    1
## [6,]   57    2
```

```
matrix_hearAttack <- matrix(c(heartAttack_chance_0$age,
                              heartAttack_chance_0$sex), ncol=2)
head(matrix_hearAttack)
```

```
##      [,1] [,2]
## [1,]   67    2
## [2,]   67    2
## [3,]   62    1
## [4,]   63    2
## [5,]   53    2
## [6,]   56    2
```

EJERCICIO 4. a) Realizad un resumen estadístico de health_heart que muestre los parámetros básicos más importantes.

```
health_heart_num <- read.csv("/Users/lsudu/code/rlab/PEC1/heart.csv")
head(health_heart) #observamos que columnas tenemos
```

```
##   age    sex cp trestbps chol fbs restecg thalach exang oldpeak slope ca thal
## 1  63 female  3      145  233   1       0     150     0     2.3     0  0    1
## 2  37 female  2      130  250   0       1     187     0     3.5     0  0    2
## 3  41   male  1      130  204   0       0     172     0     1.4     2  0    2
## 4  56 female  1      120  236   0       1     178     0     0.8     2  0    2
## 5  57   male  0      120  354   0       1     163     1     0.6     2  0    2
## 6  57 female  0      140  192   0       1     148     0     0.4     1  0    1
##   target
## 1      1
## 2      1
## 3      1
## 4      1
## 5      1
## 6      1
```

```
summary(health_heart) #nos da un vistazo de la descripcion estadistica
```

```
##       age             sex            cp           trestbps          chol      
##  Min.   :29.00   male  : 96   Min.   :0.000   Min.   : 94.0   Min.   :126.0  
##  1st Qu.:47.50   female:207   1st Qu.:0.000   1st Qu.:120.0   1st Qu.:211.0  
##  Median :55.00                Median :1.000   Median :130.0   Median :240.0  
##  Mean   :54.37                Mean   :0.967   Mean   :131.6   Mean   :246.3  
##  3rd Qu.:61.00                3rd Qu.:2.000   3rd Qu.:140.0   3rd Qu.:274.5  
##  Max.   :77.00                Max.   :3.000   Max.   :200.0   Max.   :564.0  
##       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.:133.5   1st Qu.:0.0000  
##  Median :0.0000   Median :1.0000   Median :153.0   Median :0.0000  
##  Mean   :0.1485   Mean   :0.5281   Mean   :149.6   Mean   :0.3267  
##  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.   :0.00   Min.   :0.000   Min.   :0.0000   Min.   :0.000  
##  1st Qu.:0.00   1st Qu.:1.000   1st Qu.:0.0000   1st Qu.:2.000  
##  Median :0.80   Median :1.000   Median :0.0000   Median :2.000  
##  Mean   :1.04   Mean   :1.399   Mean   :0.7294   Mean   :2.314  
##  3rd Qu.:1.60   3rd Qu.:2.000   3rd Qu.:1.0000   3rd Qu.:3.000  
##  Max.   :6.20   Max.   :2.000   Max.   :4.0000   Max.   :3.000  
##      target      
##  Min.   :0.0000  
##  1st Qu.:0.0000  
##  Median :1.0000  
##  Mean   :0.5446  
##  3rd Qu.:1.0000  
##  Max.   :1.0000
```

```
cor(health_heart_num) #vemos que dependencia tienen las variables entre si. Baja
```

```
##                ï..age         sex          cp    trestbps         chol
## ï..age     1.00000000 -0.09844660 -0.06865302  0.27935091  0.213677957
## sex       -0.09844660  1.00000000 -0.04935288 -0.05676882 -0.197912174
## cp        -0.06865302 -0.04935288  1.00000000  0.04760776 -0.076904391
## trestbps   0.27935091 -0.05676882  0.04760776  1.00000000  0.123174207
```

```
## chol      0.21367796 -0.19791217 -0.07690439  0.12317421  1.000000000
## fbs       0.12130765  0.04503179  0.09444403  0.17753054  0.013293602
## restecg  -0.11621090 -0.05819627  0.04442059 -0.11410279 -0.151040078
## thalach  -0.39852194 -0.04401991  0.29576212 -0.04669773 -0.009939839
## exang     0.09680083  0.14166381 -0.39428027  0.06761612  0.067022783
## oldpeak   0.21001257  0.09609288 -0.14923016  0.19321647  0.053951920
## slope    -0.16881424 -0.03071057  0.11971659 -0.12147458 -0.004037770
## ca        0.27632624  0.11826141 -0.18105303  0.10138899  0.070510925
## thal      0.06800138  0.21004110 -0.16173557  0.06220989  0.098802993
## target   -0.22543872 -0.28093658  0.43379826 -0.14493113 -0.085239105
##                  fbs     restecg      thalach       exang      oldpeak
## ï..age    0.121307648 -0.11621090 -0.398521938  0.09680083  0.210012567
## sex       0.045031789 -0.05819627 -0.044019908  0.14166381  0.096092877
## cp        0.094444035  0.04442059  0.295762125 -0.39428027 -0.149230158
## trestbps  0.177530542 -0.11410279 -0.046697728  0.06761612  0.193216472
## chol      0.013293602 -0.15104008 -0.009939839  0.06702278  0.053951920
## fbs       1.000000000 -0.08418905 -0.008567107  0.02566515  0.005747223
## restecg  -0.084189054  1.00000000  0.044123444 -0.07073286 -0.058770226
## thalach  -0.008567107  0.04412344  1.000000000 -0.37881209 -0.344186948
## exang     0.025665147 -0.07073286 -0.378812094  1.00000000  0.288222808
## oldpeak   0.005747223 -0.05877023 -0.344186948  0.28822281  1.000000000
## slope    -0.059894178  0.09304482  0.386784410 -0.25774837 -0.577536817
## ca        0.137979327 -0.07204243 -0.213176928  0.11573938  0.222682322
## thal     -0.032019339 -0.01198140 -0.096439132  0.20675379  0.210244126
## target   -0.028045760  0.13722950  0.421740934 -0.43675708 -0.430696002
##               slope          ca        thal      target
## ï..age   -0.16881424  0.27632624  0.06800138 -0.22543872
## sex      -0.03071057  0.11826141  0.21004110 -0.28093658
## cp        0.11971659 -0.18105303 -0.16173557  0.43379826
## trestbps -0.12147458  0.10138899  0.06220989 -0.14493113
## chol     -0.00403777  0.07051093  0.09880299 -0.08523911
## fbs      -0.05989418  0.13797933 -0.03201934 -0.02804576
## restecg   0.09304482 -0.07204243 -0.01198140  0.13722950
## thalach   0.38678441 -0.21317693 -0.09643913  0.42174093
## exang    -0.25774837  0.11573938  0.20675379 -0.43675708
## oldpeak  -0.57753682  0.22268232  0.21024413 -0.43069600
## slope     1.00000000 -0.08015521 -0.10476379  0.34587708
## ca       -0.08015521  1.00000000  0.15183213 -0.39172399
## thal     -0.10476379  0.15183213  1.00000000 -0.34402927
## target    0.34587708 -0.39172399 -0.34402927  1.00000000
```

```r
sapply(health_heart_num, mean) #echamos un vistazo a la media de todas las variables.
```

```
##      ï..age         sex          cp    trestbps        chol         fbs
##  54.3663366   0.6831683   0.9669967 131.6237624 246.2640264   0.1485149
##     restecg     thalach       exang     oldpeak       slope          ca
##   0.5280528 149.6468647   0.3267327   1.0396040   1.3993399   0.7293729
##        thal      target
##   2.3135314   0.5445545
```

b) Definid un vector con la probabilidad de sufrir un ataque (target).

Etiquetad la variable "target" con "Menor" si el valor es 0 y "Mayor" si el valor es 1.

```r
health_heart$target <- factor(health_heart$target, levels=c(0,1),
                              labels=c("Menor", "Mayor"))
prob_attack <- c(health_heart$target)
```

c) Definid un vector con los niveles de colesterol (chol) de los/as pacientes, después ordenad dicho vector, calculad la media, la varianza y desviación estándar.

```r
chol_patient <- c(health_heart$chol)
chol_patient <- sort(chol_patient)
chol_patient_mean <- mean(chol_patient)
chol_patient_var <- var(chol_patient)
chol_patient_sd <- sd(chol_patient)
```

d) Mostrad las tablas de frecuencias relativas y absolutas de los vectores de la edad (age) y la probabilidad (target). Posteriormente representad una tabla de frecuencias relativas cruzadas, de manera que podamos visualizar cuántos/as pacientes según edad, tiene una probabilidad mayor o menor.

```r
age_patient <- c(health_heart$age)
target_patient <- c(health_heart$target)

prop.table(table(age_patient, target_patient))
```

```
##            target_patient
## age_patient          1          2
##          29 0.00000000 0.00330033
##          34 0.00000000 0.00660066
##          35 0.00660066 0.00660066
##          37 0.00000000 0.00660066
##          38 0.00330033 0.00660066
##          39 0.00330033 0.00990099
##          40 0.00660066 0.00330033
##          41 0.00330033 0.02970297
##          42 0.00330033 0.02310231
##          43 0.00990099 0.01650165
##          44 0.00990099 0.02640264
##          45 0.00660066 0.01980198
##          46 0.00990099 0.01320132
##          47 0.00660066 0.00990099
##          48 0.00990099 0.01320132
##          49 0.00660066 0.00990099
##          50 0.00990099 0.01320132
##          51 0.00990099 0.02970297
##          52 0.01320132 0.02970297
##          53 0.00660066 0.01980198
##          54 0.01980198 0.03300330
##          55 0.01650165 0.00990099
##          56 0.01980198 0.01650165
##          57 0.03300330 0.02310231
##          58 0.03960396 0.02310231
##          59 0.02970297 0.01650165
##          60 0.02640264 0.00990099
```

```
##          61 0.02310231 0.00330033
##          62 0.02310231 0.01320132
##          63 0.01980198 0.00990099
##          64 0.01320132 0.01980198
##          65 0.01320132 0.01320132
##          66 0.00990099 0.01320132
##          67 0.01980198 0.00990099
##          68 0.00660066 0.00660066
##          69 0.00330033 0.00660066
##          70 0.00990099 0.00330033
##          71 0.00000000 0.00990099
##          74 0.00000000 0.00330033
##          76 0.00000000 0.00330033
##          77 0.00330033 0.00000000
```

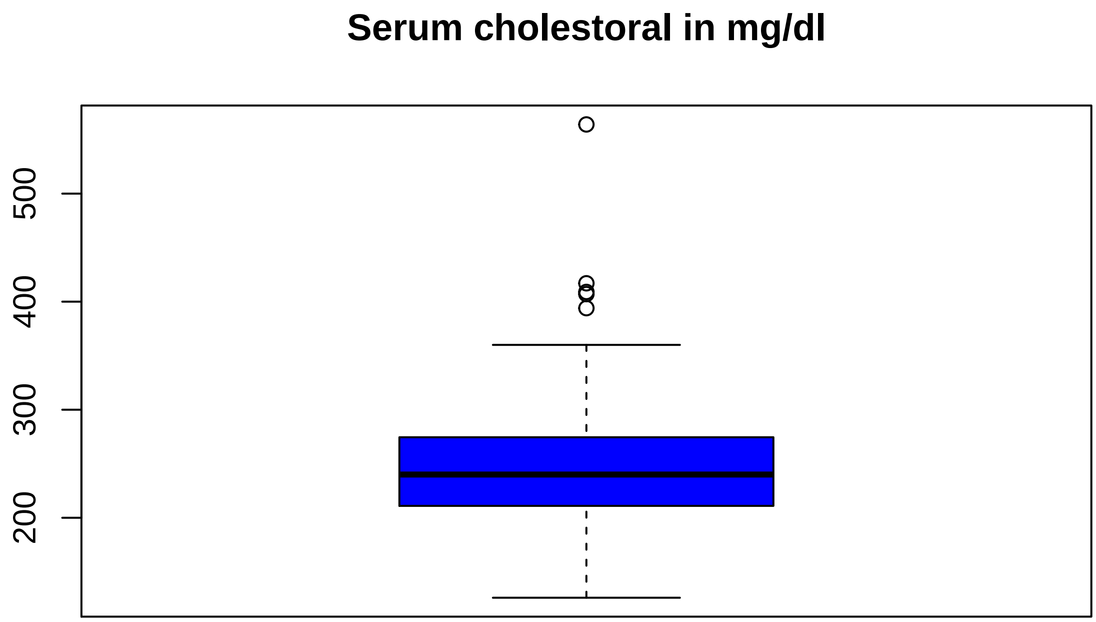
EJERCICIO 5. Realizad los siguientes gráficos: a) Diagrama de tallo y hojas de la variable "age".

```
stem(health_heart$age, scale=1, width=80)
```

```
##
##   The decimal point is at the |
##
##   28 | 0
##   30 |
##   32 |
##   34 | 000000
##   36 | 00
##   38 | 0000000
##   40 | 0000000000000
##   42 | 0000000000000000
##   44 | 0000000000000000000
##   46 | 000000000000
##   48 | 000000000000
##   50 | 0000000000000000000
##   52 | 000000000000000000000
##   54 | 000000000000000000000000
##   56 | 0000000000000000000000000000
##   58 | 000000000000000000000000000000000
##   60 | 0000000000000000000
##   62 | 00000000000000000000
##   64 | 000000000000000000
##   66 | 0000000000000000
##   68 | 0000000
##   70 | 0000000
##   72 |
##   74 | 0
##   76 | 00
```

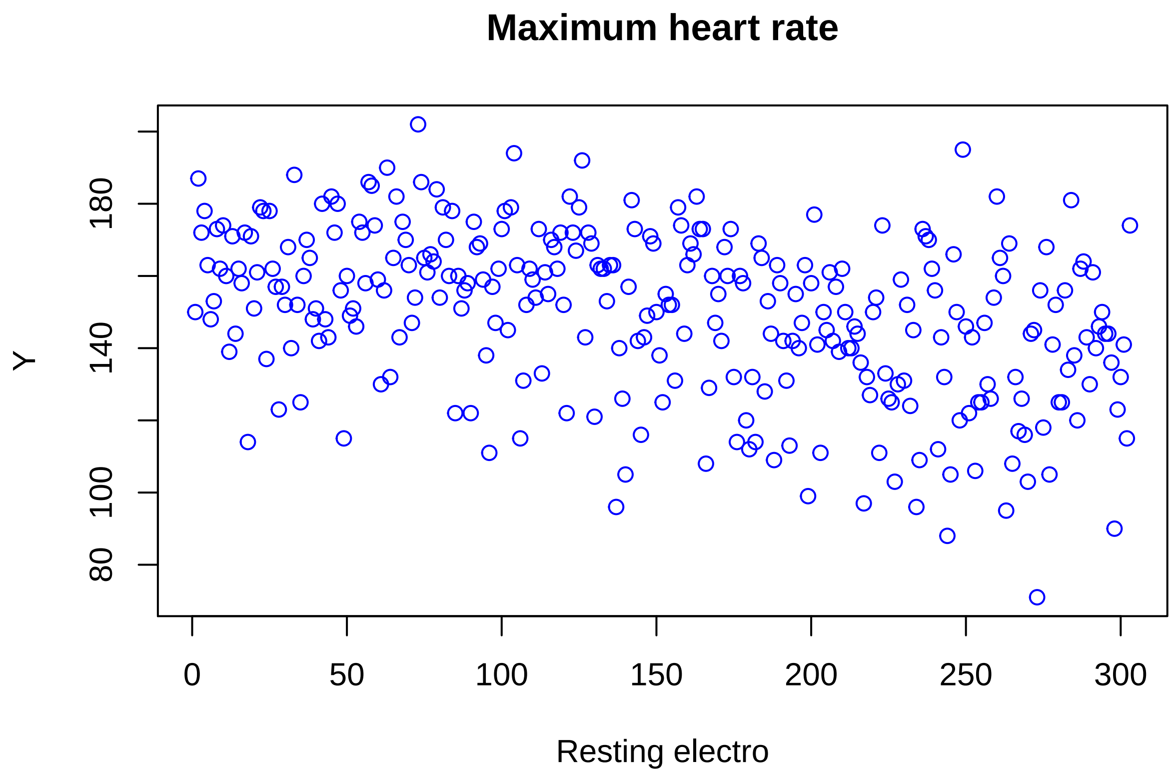
b) Diagrama de cajas y bigotes de la variable "chol".

```
boxplot(health_heart$chol,main = "Serum cholestoral in mg/dl",  col="blue")
```



d) Diagrama de puntos de la variable “thalach” (frecuencia cardíaca máxima alcanzada).

```
plot(health_heart$thalach, main= "Maximum heart rate",
     col="blue", xlab="Resting electro", ylab ="Y")
```



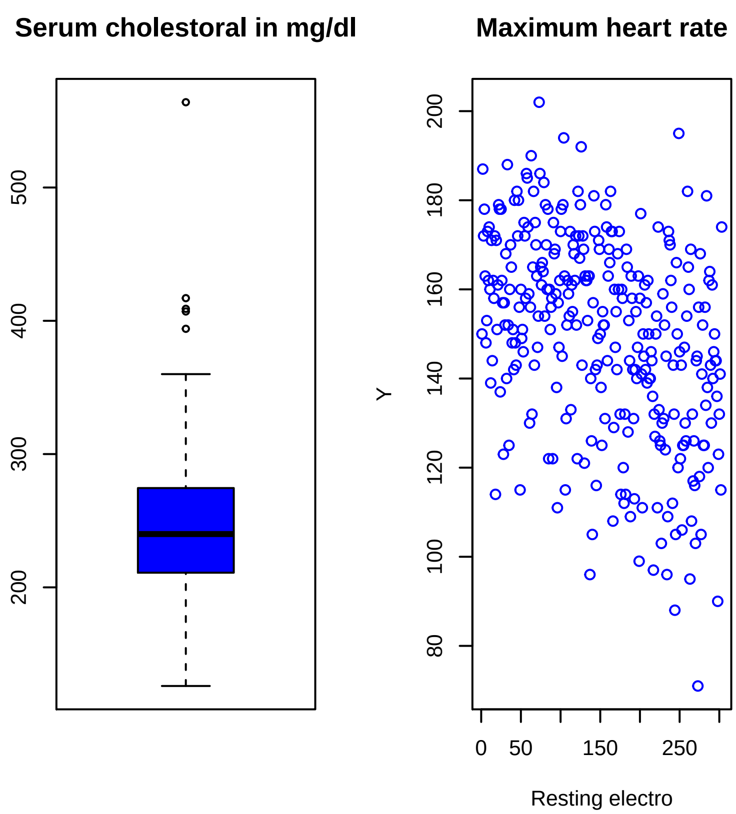
e) Combinad los gráficos anteriores en una representación gráfica común. Podéis utilizar la función layout() para ajustar la distribución de los gráficos si fuera necesario.

```
par(mfrow=c(1,3))
boxplot(health_heart$chol,main = "Serum cholestoral in mg/dl", col="blue")
stem(health_heart$age, scale=1, width=80)
```

```
##
##   The decimal point is at the |
##
##   28 | 0
##   30 |
##   32 |
##   34 | 000000
##   36 | 00
##   38 | 0000000
##   40 | 0000000000000
##   42 | 0000000000000000
##   44 | 0000000000000000000
##   46 | 000000000000
##   48 | 000000000000
##   50 | 0000000000000000000
##   52 | 000000000000000000000
##   54 | 000000000000000000000000
##   56 | 0000000000000000000000000000
```

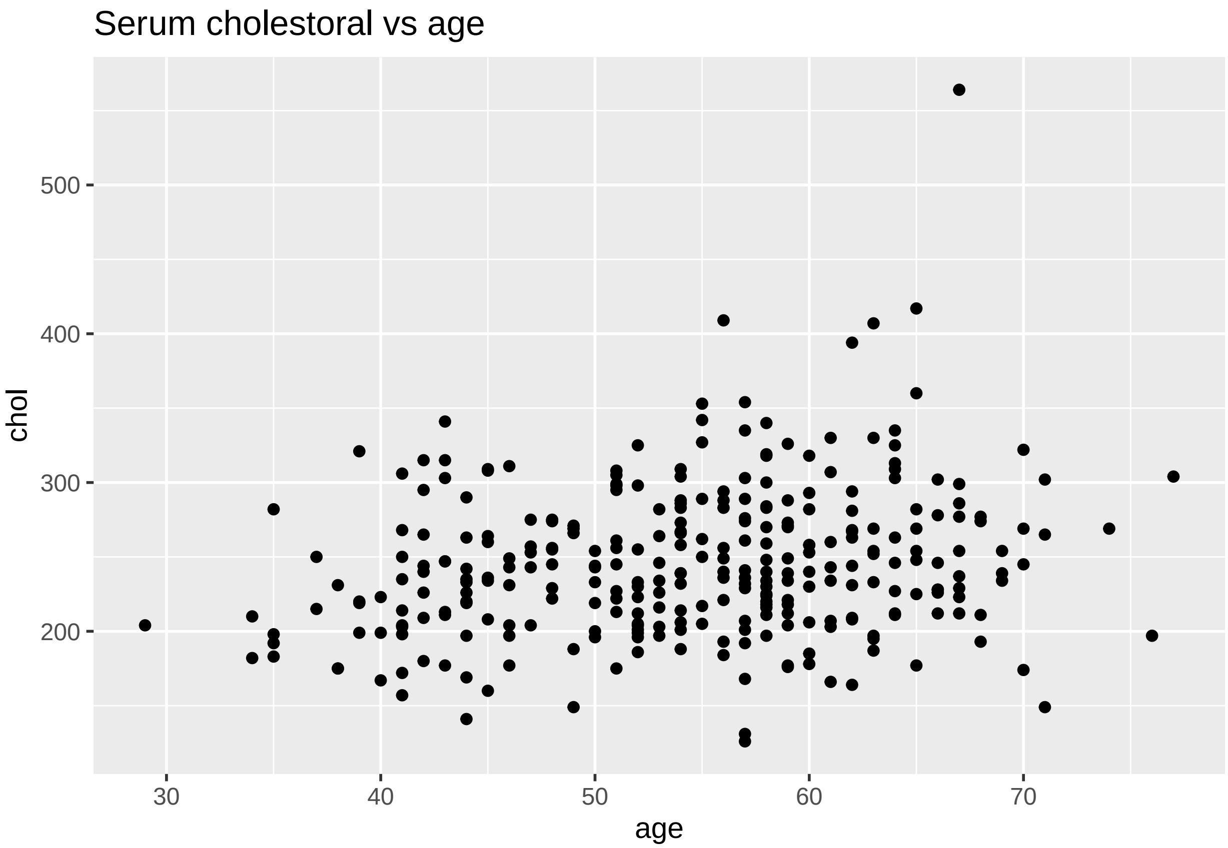
```
##    58 | 00000000000000000000000000000000
##    60 | 0000000000000000000
##    62 | 00000000000000000000
##    64 | 000000000000000000
##    66 | 0000000000000000
##    68 | 0000000
##    70 | 0000000
##    72 |
##    74 | 0
##    76 | 00
```

```
plot(health_heart$thalach,main= "Maximum heart rate",
     col="blue", xlab="Resting electro", ylab ="Y")
```



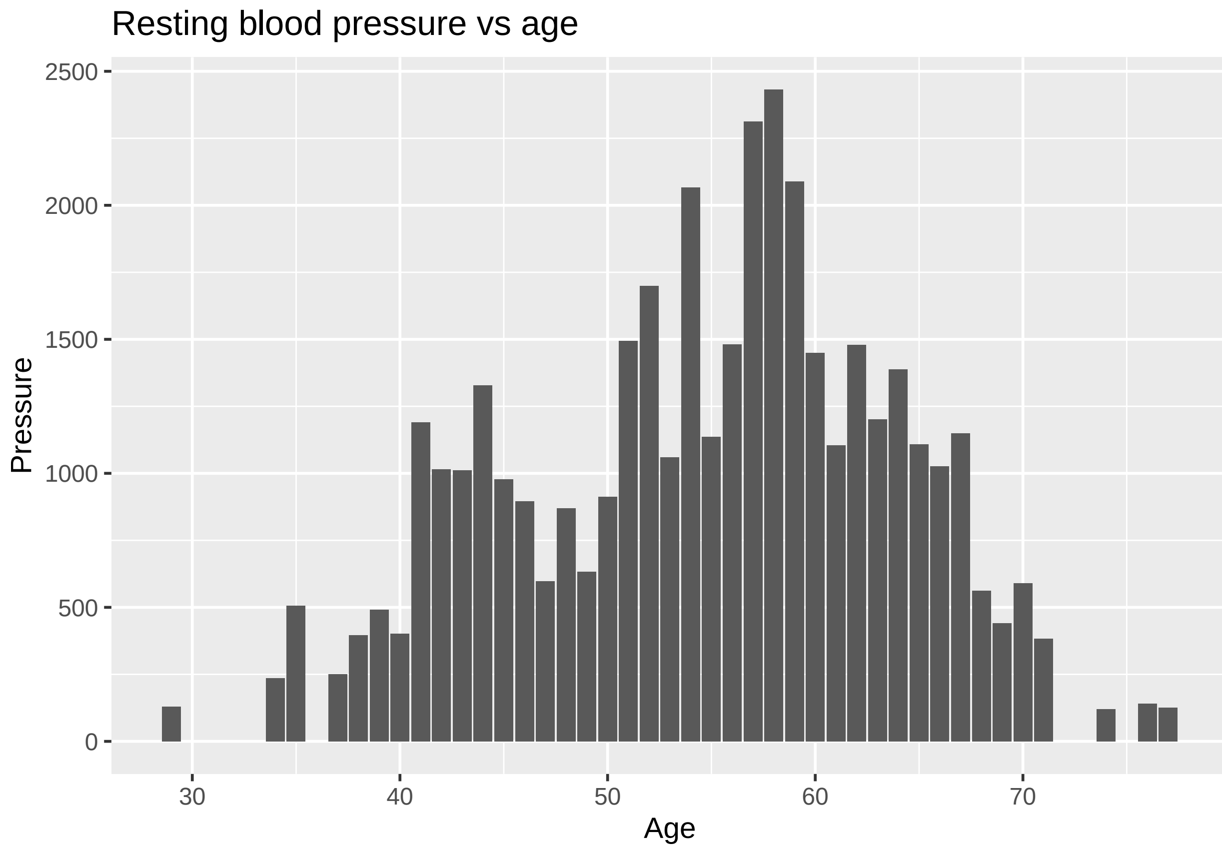
EJERCICIO 6. A partir del data frame "health_heart" se pide resolver las siguientes cuestiones. 6.1. Utilizando el paquete de gráficos ggplot2, se pide: a) Realizar un gráfico de tipo qplot de la variable "age" y la variable "chol".

```
library(ggplot2)
qplot(main = "Serum cholestoral vs age",
      x=age, y=chol, data=health_heart, geom="point")
```



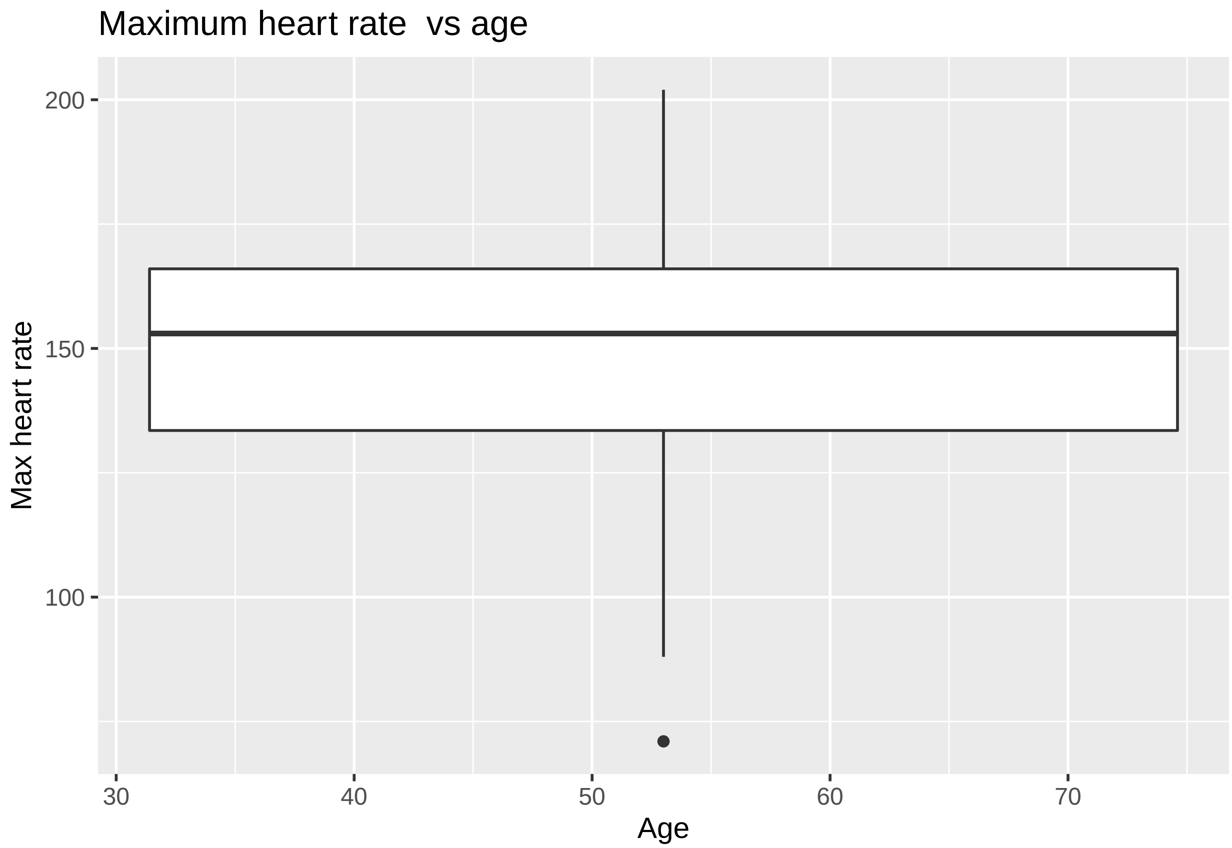
b) Realizar un gráfico de barras para las variables "age" y "trestbps".

```
ggplot(data = health_heart,  aes(x=age, y= trestbps)) +
  labs(title= "Resting blood pressure vs age", x="Age", y="Pressure") +
  geom_col()
```



6.2. Realizad un breve estudio de regresión y correlación lineal, resolviendo las siguientes cuestiones: a) Realizad un diagrama de cajas sobre las variables "age" y "thalach". ¿Qué se puede extraer de este gráfico?

```
ggplot(data = health_heart,  aes(x=age, y= thalach, group=1)) +
  labs(title= "Maximum heart rate  vs age", x="Age", y="Max heart rate") +
  geom_boxplot()
```



El grueso de nuestros valores “heart rate” se encuentran entre 125 y 175. Tiene una dispersión equilibrada. Vemos un valor extremo tambien en la parte baja de la grafica.

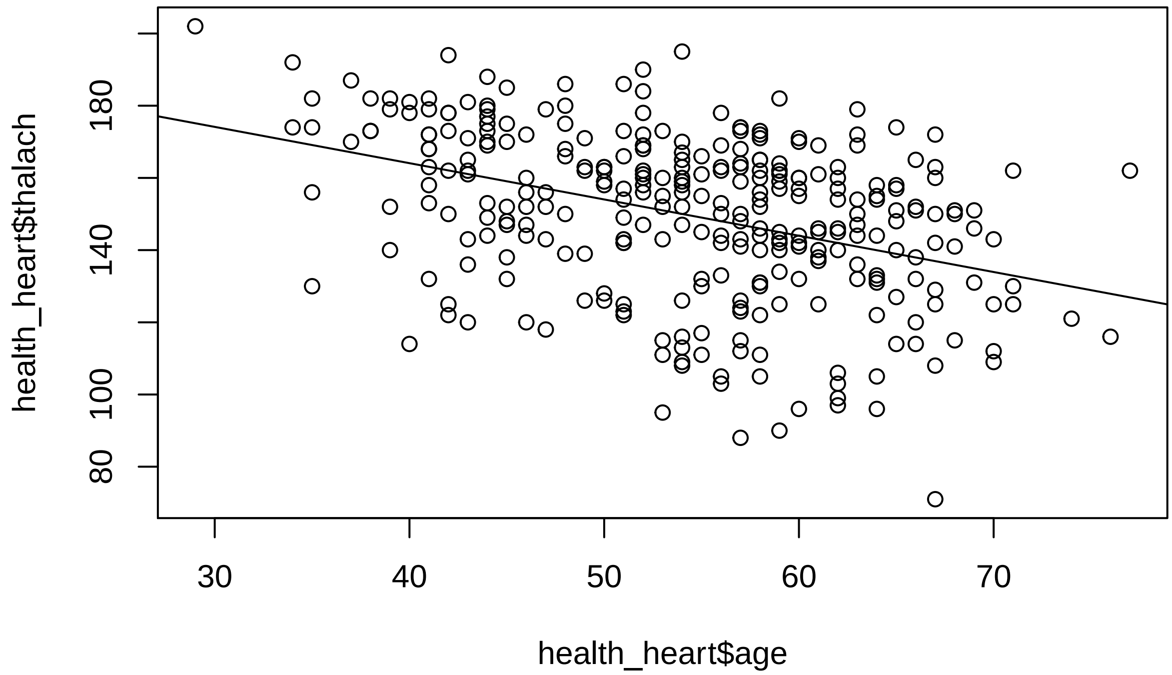
b) Realizad un modelo de regresión para las dos variables anteriores y un diagrama de puntos que ajuste al modelo anterior. ¿Qué conclusiones pueden obtenerse?

```
modelo <- lm(health_heart$thalach~health_heart$age, data=health_heart)
summary(modelo)
```

```
##
## Call:
## lm(formula = health_heart$thalach ~ health_heart$age, data = health_heart)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -65.949 -11.954   3.975  15.921  44.985
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      204.2892     7.3485  27.800  < 2e-16 ***
## health_heart$age  -1.0051     0.1333  -7.539 5.63e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21.04 on 301 degrees of freedom
## Multiple R-squared:  0.1588, Adjusted R-squared:  0.156
```

```
## F-statistic: 56.83 on 1 and 301 DF,  p-value: 5.628e-13
```

```r
plot(health_heart$age, health_heart$thalach)
abline(modelo)
```




```r
cor.test(health_heart$age, health_heart$thalach,
                               method="pearson")
```

```
##
##  Pearson's product-moment correlation
##
## data:  health_heart$age and health_heart$thalach
## t = -7.5386, df = 301, p-value = 5.628e-13
## alternative hypothesis: true correlation is not equal to 0
## 95 percent confidence interval:
##  -0.4892312 -0.2992831
## sample estimates:
##        cor
## -0.3985219
```

Podemos ver que si hay una tendencia, pero hay una correlación baja entre las variables

c) Realizad la matriz de correlación del conjunto de datos health_heart. ¿Qué se puede afirmar?

```r
age <- c(health_heart$age)
thalach <- c(health_heart$thalach)
df <- data.frame(age, thalach)
cor(df)
```

```
##                age    thalach
## age      1.0000000 -0.3985219
## thalach -0.3985219  1.0000000
```

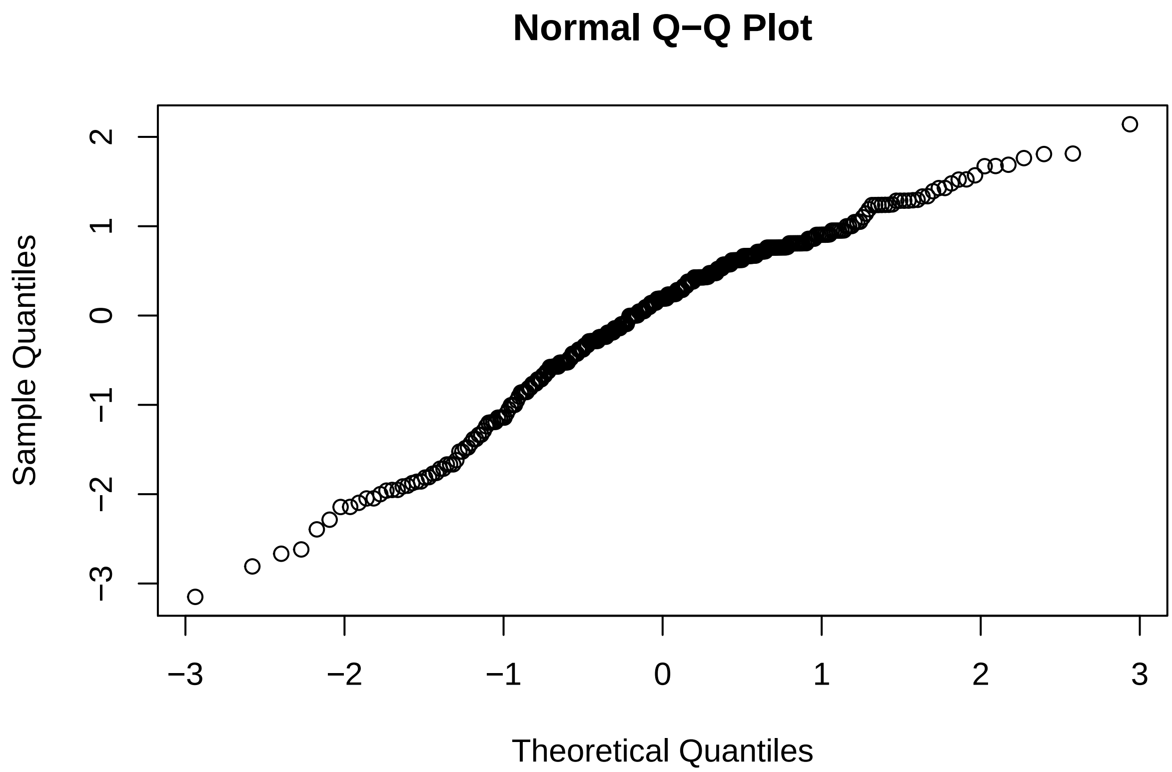
Reafirmamos en la poca dependencia de una variable respecto a la otra.

d) Calculad los residuos del modelo ajustado anterior y realizad un gráfico de normalidad.

```r
summary(modelo)
```

```
##
## Call:
## lm(formula = health_heart$thalach ~ health_heart$age, data = health_heart)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -65.949 -11.954   3.975  15.921  44.985
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      204.2892     7.3485  27.800  < 2e-16 ***
## health_heart$age  -1.0051     0.1333  -7.539 5.63e-13 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 21.04 on 301 degrees of freedom
## Multiple R-squared:  0.1588, Adjusted R-squared:  0.156
## F-statistic: 56.83 on 1 and 301 DF,  p-value: 5.628e-13
```

```r
residuos <- rstandard(modelo)
qqnorm(residuos)
```



e) De los apartados anteriores, ¿qué se puede concluir? No existe correlación directa entre las variables seleccionadas.

EJERCICIO 7. A partir del conjunto de datos "women" del paquete MASS, guardad en un vector la altura (height) de las mujeres cuyo peso (weight) es superior a 131lb e inferior a 150lb.

```
library(MASS) #cargamos el paquete
data("women")
table <- data.frame(women)

height_women <- c()#Creo el vector vacio
j <- 1  #Creo la instancia j que señala al valor 1 que me dara la posición
        #equivalente de la [i] en el vector

F_height <- function(){   #creo la función
  for(i in 1:length(women$height))  #hago el bucle
    if (women$weight[i] > 131)      #creo las condiciones
    if (women$weight[i] < 150) {
      height_women[j] <- women$height[i]  #Voy añadiendo los valores al vector
      j <- j + 1 #Asi voy añadiendo en la posición equivalente
    }
  return (height_women)
}
heigth <- F_height()
heigth
```

```
## [1] 64 65 66 67 68
```

Calculad la media del vector y las posiciones asociadas (del conjunto de datos "women") a los valores incluidos en el nuevo vector creado.

```r
mean(heigth)
```

```
## [1] 66
```

```r
match(heigth, women$height) #para encontrar las posiciones asociadas
```

```
## [1]  7  8  9 10 11
```

EJERCICIO 8. A partir del conjunto de datos "BOD" de la libreria RSQLite, realizad un breve estudio utilizando la sintaxis de SQL con R. En concreto, se pide:

a) Una lectura de las 4 primeras líneas del conjunto de datos.

```r
library(RSQLite)
```

```
## Warning: package 'RSQLite' was built under R version 4.0.3
```

```r
library(sqldf)#cargamos la libreria
```

```
## Warning: package 'sqldf' was built under R version 4.0.3

## Loading required package: gsubfn

## Warning: package 'gsubfn' was built under R version 4.0.3

## Loading required package: proto

## Warning: package 'proto' was built under R version 4.0.3
```

```r
db <- "RSQLite"::datasetsDb()  #accedemos a las bases de datos de uno de los
#ficheros de RSQLite
dbReadTable(db, "BOD")#accedemos a BOD y lo guardamos en db
```

```
##   Time demand
## 1    1    8.3
## 2    2   10.3
## 3    3   19.0
## 4    4   16.0
## 5    5   15.6
## 6    7   19.8
```

```r
sqldf("SELECT * FROM BOD LIMIT 4")
```

```
##   Time demand
## 1    1    8.3
## 2    2   10.3
## 3    3   19.0
## 4    4   16.0
```

b) La información de la tabla ordenada de forma descendente de acuerdo a la variable “Demand”.

```r
sqldf("SELECT * FROM BOD ORDER BY demand DESC")
```

```
##   Time demand
## 1    7   19.8
## 2    3   19.0
## 3    4   16.0
## 4    5   15.6
## 5    2   10.3
## 6    1    8.3
```

c) Los datos de la tabla cuyos valores de la variable “Demand” se encuentran entre 12 y 25.

```r
sqldf("SELECT * FROM BOD where demand > 12 AND demand < 25")
```

```
##   Time demand
## 1    3   19.0
## 2    4   16.0
## 3    5   15.6
## 4    7   19.8
```