knn_algorithm

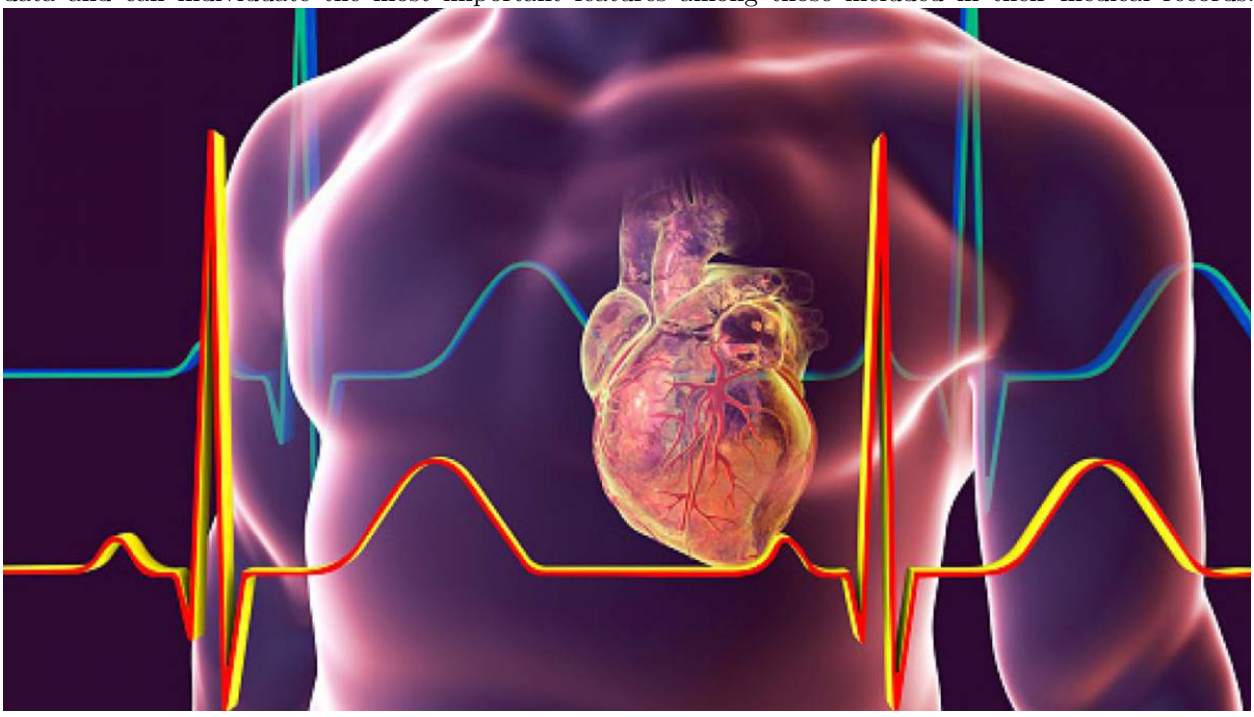
Manuel Herrera Lara y Anahí Berumen Murillo

3/11/2020

Perform a basic data analysis describing the dataset, summary statistics, data distribution, etc.

The data domain

With this dataset we try to predict survival of patients with heart failure applying the algorithm knn. Cardiovascular diseases kill approximately 17 million people globally every year, and they mainly exhibit as myocardial infarctions and heart failures. Heart failure (HF) occurs when the heart cannot pump enough blood to meet the needs of the body. Available electronic medical records of patients quantify symptoms, body features, and clinical laboratory test values, which can be used to perform biostatistics analysis aimed at highlighting patterns and correlations otherwise undetectable by medical doctors. Machine learning, in particular, with classification algorithms like knn can predict patients survival from their data and can individuate the most important features among those included in their medical records.



Cardiovascular diseases are disorders of the heart and blood vessels, including coronary heart disease (heart attacks), cerebrovascular diseases (strokes), heart failure (HF), and other types of pathology. In particular, heart failure occurs when the heart is unable to pump enough blood to the body, and it is usually caused by diabetes, high blood pressure, or other heart conditions or diseases.

Algorithm KNN

- **KNN application** – Classify data based on similarity to its neighbors
  - simple and effective
  - has a quick training phase
  - KNN uses information from the nearest neighbors against the new observations that have just arrived.
  - k. Number of neighbors.
  - It is based on the calculation of distances, that is, it calculates the distances between a point and its neighbors.
  - There are different metrics for measuring distances between objects, the most common being the Euclidean distance.

How the data was recollected, limitations of the study, disadvantages, etc.

To diagnose heart failure, your doctor will carefully review your medical history and symptoms, and do a physical exam. Your doctor can also check for risk factors, such as high blood pressure, coronary artery disease, or diabetes. This dataset contains the medical records of 299 patients who had heart failure, collected during their follow-up period, where each patient profile has 13 clinical features. The current version of the dataset was elaborated by Davide Chicco (Krembil Research Institute, Toronto, Canada) and donated to the University of California Irvine Machine Learning Repository.

Description of the variables of the dataset

Feature	Explanation	Measurement	Range
Age	Age of the patient	Years	[40,..., 95]
Anaemia	Decrease of red blood cells or hemoglobin	Boolean	0, 1
High blood pressure	If a patient has hypertension	Boolean	0, 1
Creatinine phosphokinase	Level of the CPK enzyme in the blood	mcg/L	[23,..., 7861]
(CPK)			
Diabetes	If the patient has diabetes	Boolean	0, 1
Ejection fraction	Percentage of blood leaving	Percentage	[14,..., 80]
	the heart at each contraction		
Sex	Woman or man	Binary	0, 1
Platelets	Platelets in the blood	kiloplatelets/mL	[25.01,..., 850.00]
Serum creatinine	Level of creatinine in the blood	mg/dL	[0.50,..., 9.40]
Serum sodium	Level of sodium in the blood	mEq/L	[114,..., 148]
Smoking	If the patient smokes	Boolean	0, 1
Time	Follow-up period	Days	[4,...,285]
(target) death event	If the patient died during the follow-up period	Boolean	0, 1

» (dataset reading)

```
knitr::opts_chunk$set(echo = TRUE)
# path of the dataset
setwd("/home/chino/Documentos/17_materias_IS/1_mineria_de_datos/9_semana_miniproyecto2/")

# read the dataset
pacientes_heart_failure <- read.csv("heart_failure_clinical_records_dataset.csv", stringsAsFactors = FA
```

Basic summary statics

- It shows the first 10 records of the dataset.

```
head(pacientes_heart_failure, 10)
```

```
##    age anaemia creatinine_phosphokinase diabetes ejection_fraction
## 1   75       0                     582        0                20
## 2   55       0                    7861        0                38
## 3   65       0                     146        0                20
## 4   50       1                     111        0                20
## 5   65       1                     160        1                20
## 6   90       1                      47        0                40
## 7   75       1                     246        0                15
## 8   60       1                     315        1                60
## 9   65       0                     157        0                65
## 10  80       1                     123        0                35
##    high_blood_pressure platelets serum_creatinine serum_sodium sex smoking time
## 1                    1    265000              1.9          130   1       0    4
## 2                    0    263358              1.1          136   1       0    6
## 3                    0    162000              1.3          129   1       1    7
## 4                    0    210000              1.9          137   1       0    7
## 5                    0    327000              2.7          116   0       0    8
## 6                    1    204000              2.1          132   1       1    8
## 7                    0    127000              1.2          137   1       0   10
## 8                    0    454000              1.1          131   1       1   10
## 9                    0    263358              1.5          138   0       0   10
## 10                   1    388000              9.4          133   1       1   10
##    DEATH_EVENT
## 1            1
## 2            1
## 3            1
## 4            1
## 5            1
## 6            1
## 7            1
## 8            1
## 9            1
## 10           1
```

- It shows the structure of the data and/or the data types of the attributes.

```
str(pacientes_heart_failure)
```

```
## 'data.frame':    299 obs. of  13 variables:
##  $ age                     : num  75 55 65 50 65 90 75 60 65 80 ...
##  $ anaemia                 : int  0 0 0 1 1 1 1 1 0 1 ...
##  $ creatinine_phosphokinase: int  582 7861 146 111 160 47 246 315 157 123 ...
##  $ diabetes                : int  0 0 0 0 1 0 0 1 0 0 ...
##  $ ejection_fraction       : int  20 38 20 20 20 40 15 60 65 35 ...
##  $ high_blood_pressure     : int  1 0 0 0 0 1 0 0 0 1 ...
##  $ platelets               : num  265000 263358 162000 210000 327000 ...
##  $ serum_creatinine        : num  1.9 1.1 1.3 1.9 2.7 2.1 1.2 1.1 1.5 9.4 ...
##  $ serum_sodium            : int  130 136 129 137 116 132 137 131 138 133 ...
##  $ sex                     : int  1 1 1 1 0 1 1 1 0 1 ...
##  $ smoking                 : int  0 0 1 0 0 1 0 1 0 1 ...
##  $ time                    : int  4 6 7 7 8 8 10 10 10 10 ...
##  $ DEATH_EVENT             : int  1 1 1 1 1 1 1 1 1 1 ...
```

- summary with basic statistical measures.

```
summary(pacientes_heart_failure)
```

```
##       age           anaemia       creatinine_phosphokinase    diabetes
##  Min.   :40.00   Min.   :0.0000   Min.   :  23.0           Min.   :0.0000
##  1st Qu.:51.00   1st Qu.:0.0000   1st Qu.: 116.5           1st Qu.:0.0000
##  Median :60.00   Median :0.0000   Median : 250.0           Median :0.0000
##  Mean   :60.83   Mean   :0.4314   Mean   : 581.8           Mean   :0.4181
##  3rd Qu.:70.00   3rd Qu.:1.0000   3rd Qu.: 582.0           3rd Qu.:1.0000
##  Max.   :95.00   Max.   :1.0000   Max.   :7861.0           Max.   :1.0000
##  ejection_fraction high_blood_pressure   platelets      serum_creatinine
##  Min.   :14.00     Min.   :0.0000      Min.   : 25100   Min.   :0.500
##  1st Qu.:30.00     1st Qu.:0.0000      1st Qu.:212500   1st Qu.:0.900
##  Median :38.00     Median :0.0000      Median :262000   Median :1.100
##  Mean   :38.08     Mean   :0.3512      Mean   :263358   Mean   :1.394
##  3rd Qu.:45.00     3rd Qu.:1.0000      3rd Qu.:303500   3rd Qu.:1.400
##  Max.   :80.00     Max.   :1.0000      Max.   :850000   Max.   :9.400
##   serum_sodium        sex            smoking            time
##  Min.   :113.0   Min.   :0.0000   Min.   :0.0000   Min.   :  4.0
##  1st Qu.:134.0   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.: 73.0
##  Median :137.0   Median :1.0000   Median :0.0000   Median :115.0
##  Mean   :136.6   Mean   :0.6488   Mean   :0.3211   Mean   :130.3
##  3rd Qu.:140.0   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:203.0
##  Max.   :148.0   Max.   :1.0000   Max.   :1.0000   Max.   :285.0
##   DEATH_EVENT
##  Min.   :0.0000
##  1st Qu.:0.0000
##  Median :0.0000
##  Mean   :0.3211
##  3rd Qu.:1.0000
##  Max.   :1.0000
```

Describe the distribution of the data.

Exploring the Variables

```
table(pacientes_heart_failure$anaemia)
```

Anemia patients

```
##
##   0   1
## 170 129
```

```
anaemia_table <- table(pacientes_heart_failure$anaemia)
anaemia_pct <- prop.table(anaemia_table) * 100
round(anaemia_pct, digits = 1)
```

Percentage of patients with anemia

```
##
##    0    1
## 56.9 43.1
```

```
table(pacientes_heart_failure$diabetes)
```

Diabetes patients

```
##
##   0   1
## 174 125
```

```
diabetes_table <- table(pacientes_heart_failure$diabetes)
diabetes_pct <- prop.table(diabetes_table) * 100
round(diabetes_pct, digits = 1)
```

Percentage of patients with diabetes

```
##
##    0    1
## 58.2 41.8
```

```r
table(pacientes_heart_failure$high_blood_pressure)
```

Patients with high blood pressure

```
##
##   0   1
## 194 105
```

```r
hbp_table <- table(pacientes_heart_failure$high_blood_pressure)
hbp_pct <- prop.table(hbp_table) * 100
round(hbp_pct, digits = 1)
```

Percentage of patients with high blood pressure

```
##
##    0    1
## 64.9 35.1
```

```r
table(pacientes_heart_failure$sex)
```

Sex

```
##
##   0   1
## 105 194
```

```r
sex_table <- table(pacientes_heart_failure$sex)
sex_pct <- prop.table(sex_table) * 100
round(sex_pct, digits = 1)
```

Percentage of sex

```
##
##    0    1
## 35.1 64.9
```

```r
table(pacientes_heart_failure$smoking)
```

Smoking patients

```
##
##   0   1
## 203  96
```

```r
smoking_table <- table(pacientes_heart_failure$smoking)
smoking_pct <- prop.table(smoking_table) * 100
round(smoking_pct, digits = 1)
```

Percentage of smoking patients

```
##
##    0    1
## 67.9 32.1
```

```r
table(pacientes_heart_failure$DEATH_EVENT)
```

Patient survival during the follow-up period

```
##
##   0   1
## 203  96
```

```r
death_event_table <- table(pacientes_heart_failure$DEATH_EVENT)
death_event_pct <- prop.table(death_event_table) * 100
round(death_event_pct, digits = 1)
```

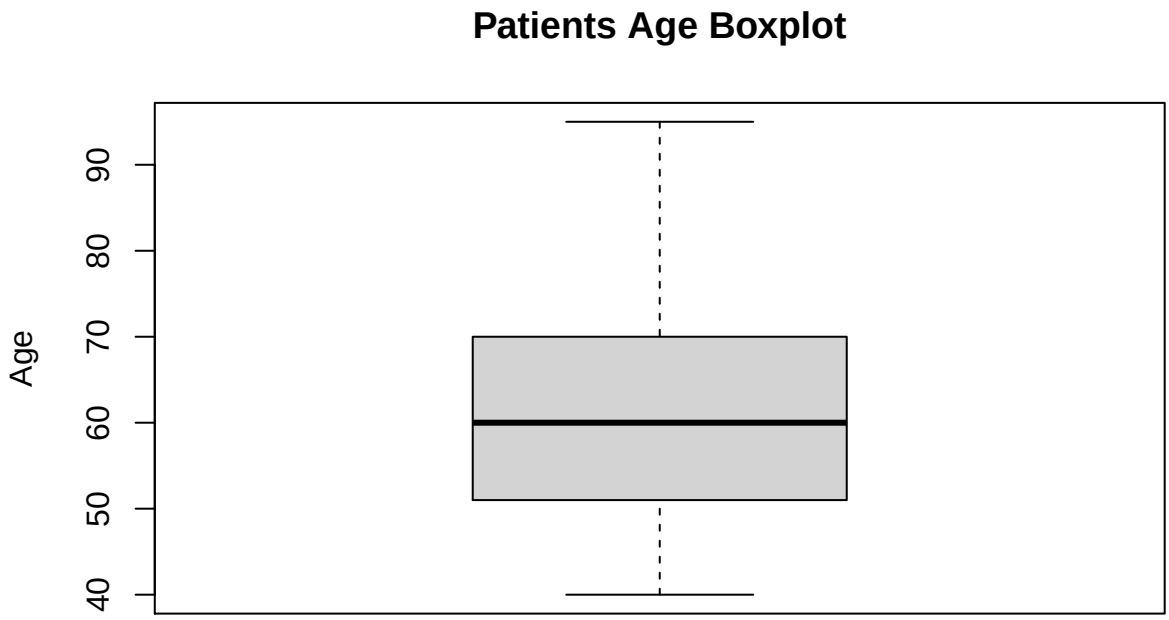
Percentage of patient survival during the follow-up period

```
##
##    0    1
## 67.9 32.1
```

Boxplots - Interpretation

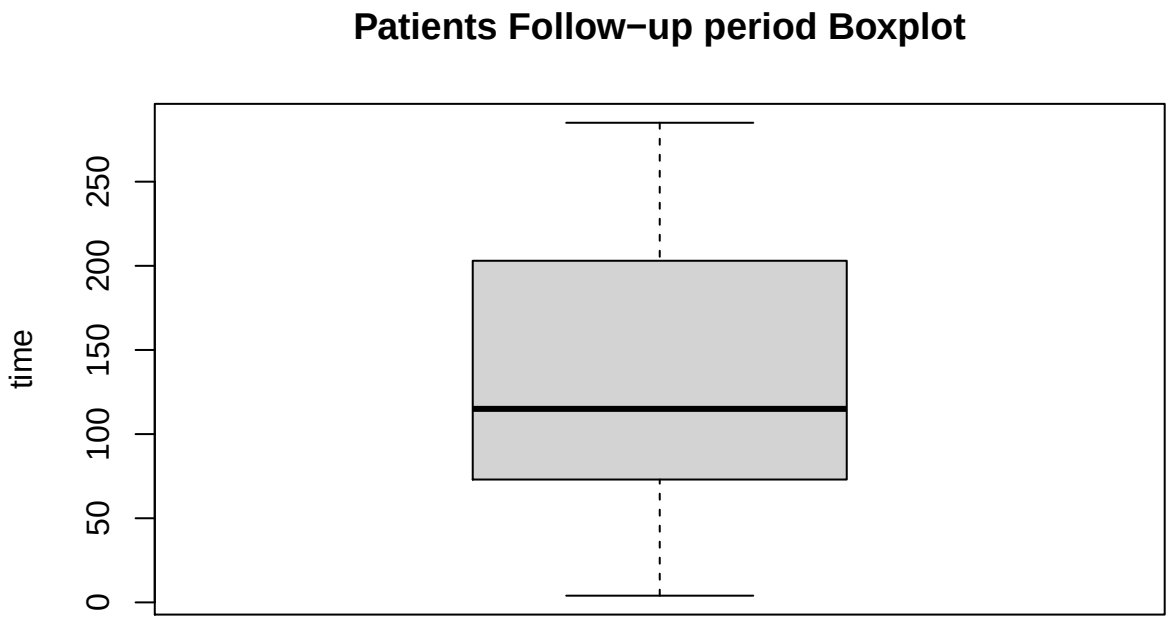
We can see that the average and median age is 60 years and that there are no outliers.

```
boxplot(pacientes_heart_failure$age, main = "Patients Age Boxplot", ylab ="Age")
```



We can see that the period is averaged over 130 days and that there are no outliers.

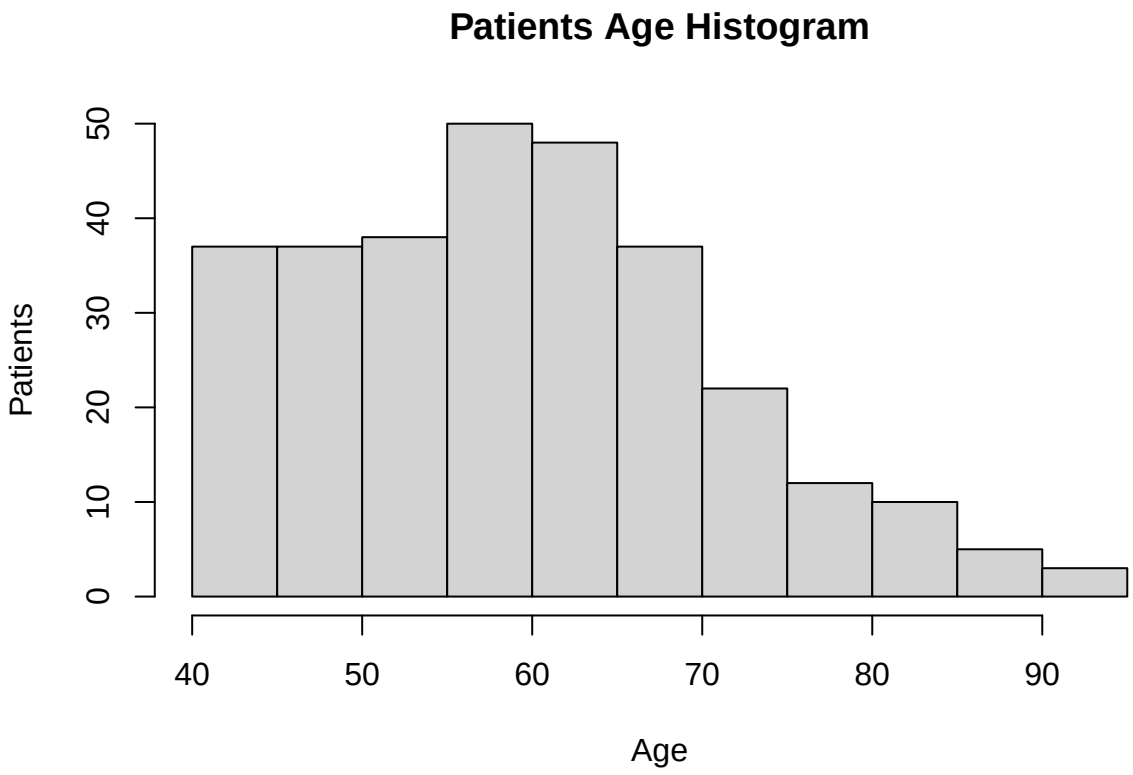
```
boxplot(pacientes_heart_failure$time, main = "Patients Follow-up period Boxplot", ylab ="time")
```



Histograms-Interpretation

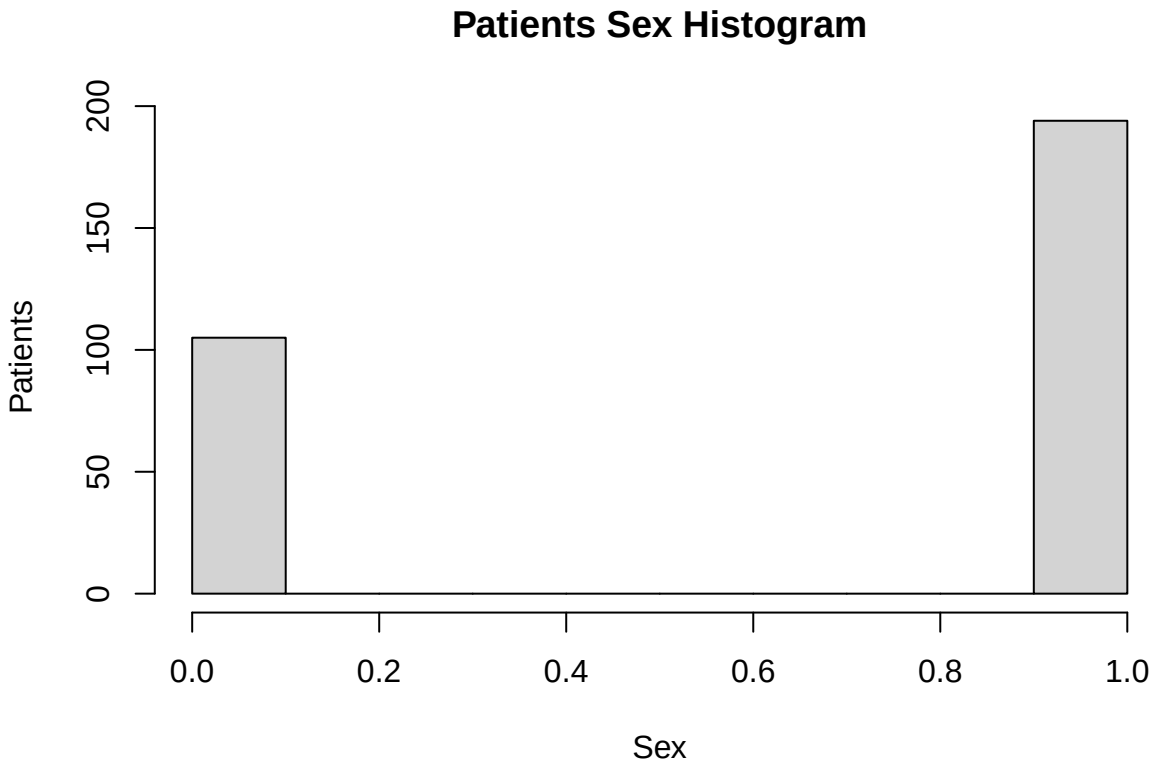
As seen in the graph, the majority of the patients are approximately 60 years old. And it is a **non-symmetric distribution** since it is skewed to the right, because the mean age is greater than the median.

```
hist(pacientes_heart_failure$age, main = "Patients Age Histogram", xlab = "Age", ylab = "Patients")
```



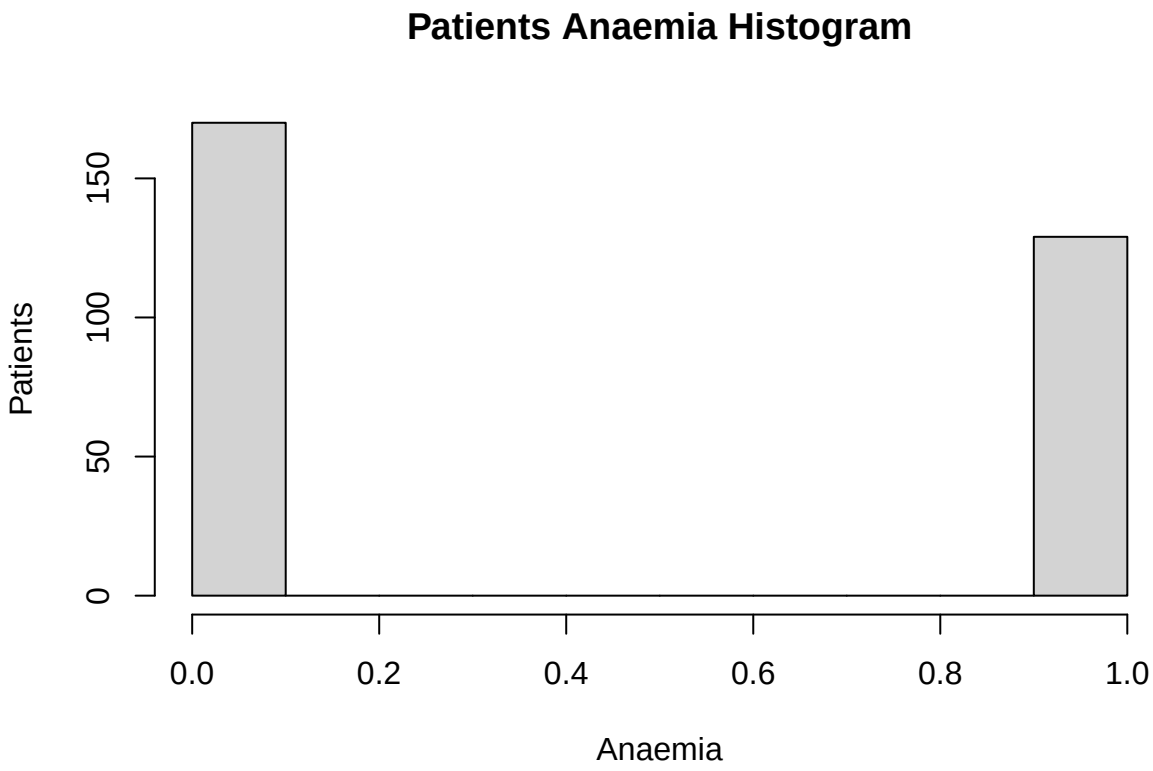
As we can see in the graph, most of the patients are men.

```
hist(pacientes_heart_failure$sex, main = "Patients Sex Histogram", xlab = "Sex", ylab = "Patients")
```



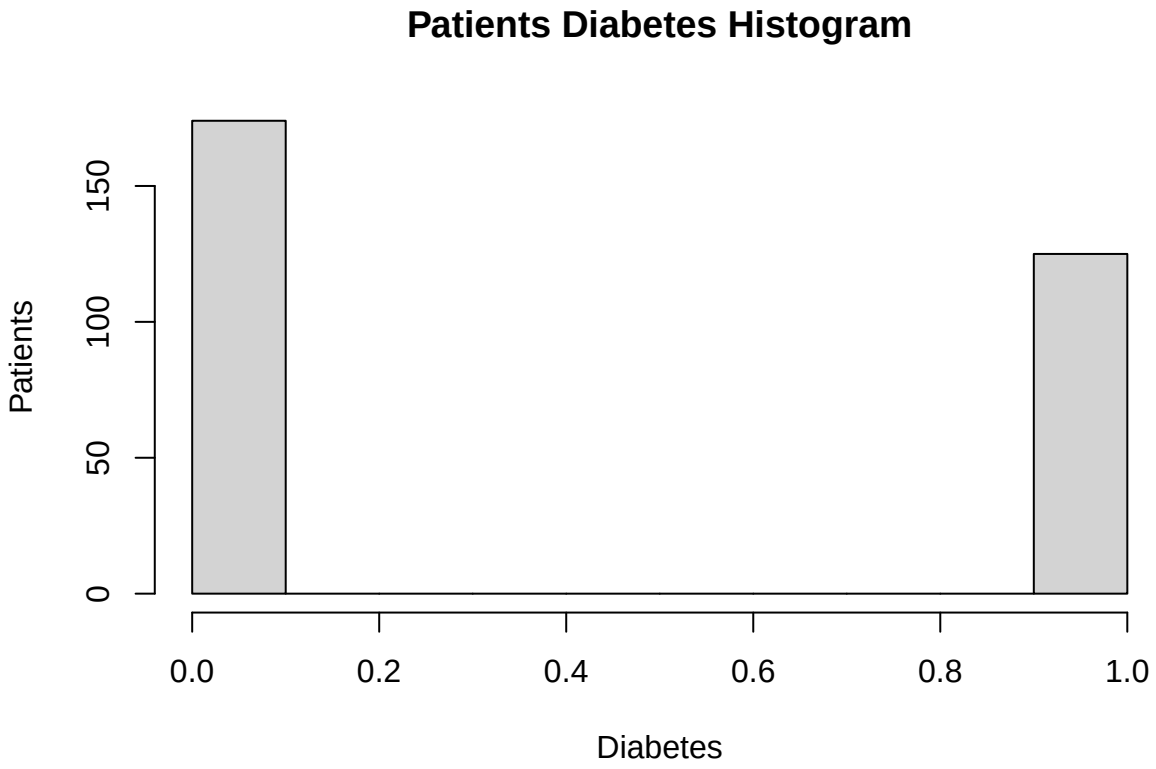
As we can see in the graph, most of the patients do not have anemia.

```
hist(pacientes_heart_failure$anaemia, main = "Patients Anaemia Histogram", xlab = "Anaemia", ylab = "Pa
```



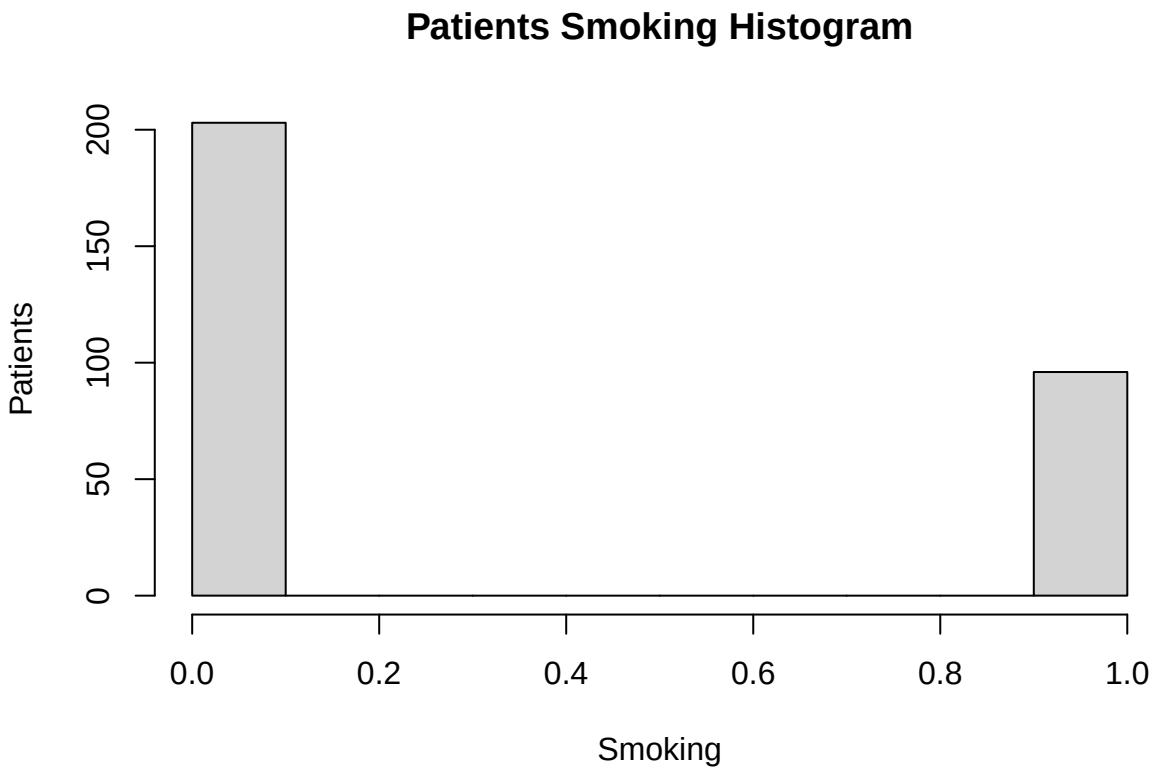
As we can see in the graph, most of the patients do not have diabetes.

```
hist(pacientes_heart_failure$diabetes, main = "Patients Diabetes Histogram", xlab = "Diabetes", ylab =
```



As we can see in the graph, most of the patients do not smoke.

```
hist(pacientes_heart_failure$smoking, main = "Patients Smoking Histogram", xlab = "Smoking", ylab = "Pa
```



- Applying knn to predict survival of patients with heart failure

$$dist(p,q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \ldots + (p_n - q_n)^2}$$

$$dist(tomato, greenbeen) = \sqrt{(6-3)^2 + (4-7)^2}$$

```
### necessary library
# install.packages("class")
# install.packages("gmodels")
# library(class)
# library(gmodels)
```

step 1.- Loading dataset

```
# path of the dataset
setwd("/home/chino/Documentos/17_materias_IS/1_mineria_de_datos/9_semana_miniproyecto2/")

# read the dataset
pacientes_heart_failure <- read.csv("heart_failure_clinical_records_dataset.csv", stringsAsFactors = FA
```

step 2.- Check the structure the dataset

```
str(pacientes_heart_failure)
```

```
## 'data.frame':    299 obs. of  13 variables:
##  $ age                     : num  75 55 65 50 65 90 75 60 65 80 ...
##  $ anaemia                 : int  0 0 0 1 1 1 1 1 0 1 ...
##  $ creatinine_phosphokinase: int  582 7861 146 111 160 47 246 315 157 123 ...
##  $ diabetes                : int  0 0 0 0 1 0 0 1 0 0 ...
##  $ ejection_fraction       : int  20 38 20 20 20 40 15 60 65 35 ...
##  $ high_blood_pressure     : int  1 0 0 0 0 1 0 0 0 1 ...
##  $ platelets               : num  265000 263358 162000 210000 327000 ...
##  $ serum_creatinine        : num  1.9 1.1 1.3 1.9 2.7 2.1 1.2 1.1 1.5 9.4 ...
##  $ serum_sodium            : int  130 136 129 137 116 132 137 131 138 133 ...
##  $ sex                     : int  1 1 1 1 0 1 1 1 0 1 ...
##  $ smoking                 : int  0 0 1 0 0 1 0 1 0 1 ...
##  $ time                    : int  4 6 7 7 8 8 10 10 10 10 ...
##  $ DEATH_EVENT             : int  1 1 1 1 1 1 1 1 1 1 ...
```

PREPROCESING

- We exclude the id

```
# not applicable because the dataset has no id
# pacientes_heart_failure <- pacientes_heart_failure[-1]
```

- Variable that we are going to predict » DEATH_EVENT

```
table(pacientes_heart_failure$DEATH_EVENT)
```

```
##
##   0   1
## 203  96
```

- Transform to factor the categorical variable that we are going to predict » DEATH_EVENT

```
pacientes_heart_failure$DEATH_EVENT <- factor(pacientes_heart_failure$DEATH_EVENT,
                                              levels = c("0", "1"),
                                              labels = c("No", "Si"))
```

- Table of proportions

```
round(prop.table(table(pacientes_heart_failure$DEATH_EVENT)) * 100, digits = 1)
```

```
##
##   No   Si
## 67.9 32.1
```

- Summary of the main variables

```
summary(pacientes_heart_failure[c("age", "time", "serum_creatinine", "serum_sodium", "anaemia", "diabet
```

```
##       age             time       serum_creatinine  serum_sodium
##  Min.   :40.00   Min.   :  4.0   Min.   :0.500    Min.   :113.0
##  1st Qu.:51.00   1st Qu.: 73.0   1st Qu.:0.900    1st Qu.:134.0
##  Median :60.00   Median :115.0   Median :1.100    Median :137.0
##  Mean   :60.83   Mean   :130.3   Mean   :1.394    Mean   :136.6
##  3rd Qu.:70.00   3rd Qu.:203.0   3rd Qu.:1.400    3rd Qu.:140.0
##  Max.   :95.00   Max.   :285.0   Max.   :9.400    Max.   :148.0
##     anaemia          diabetes      high_blood_pressure    smoking
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000      Min.   :0.0000
##  1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.0000      1st Qu.:0.0000
##  Median :0.0000   Median :0.0000   Median :0.0000      Median :0.0000
##  Mean   :0.4314   Mean   :0.4181   Mean   :0.3512      Mean   :0.3211
##  3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:1.0000      3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000      Max.   :1.0000
```

- Normalizing Min-Max numeric data

Min-max normalización:

- atrae los atributos al mismo rango
- deja los datos en escala de 0 a 1

fórmula:

$$X_{new} = \frac{X - min(X)}{max(X) - min(X)}$$

```
normalize <- function(x){
  return ((x - min(x)) / (max(x) - min(x)))
}
```

- We apply the normalize function to all the columns of the dataset

```
pacientes_heart_failure_n <- as.data.frame(lapply(pacientes_heart_failure[1:12], normalize))
summary(pacientes_heart_failure_n)
```

```
##       age            anaemia       creatinine_phosphokinase    diabetes
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.00000          Min.   :0.0000
##  1st Qu.:0.2000   1st Qu.:0.0000   1st Qu.:0.01193          1st Qu.:0.0000
##  Median :0.3636   Median :0.0000   Median :0.02896          Median :0.0000
##  Mean   :0.3788   Mean   :0.4314   Mean   :0.07130          Mean   :0.4181
##  3rd Qu.:0.5455   3rd Qu.:1.0000   3rd Qu.:0.07132          3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.00000          Max.   :1.0000
##  ejection_fraction high_blood_pressure   platelets      serum_creatinine
##  Min.   :0.0000    Min.   :0.0000      Min.   :0.0000   Min.   :0.00000
##  1st Qu.:0.2424    1st Qu.:0.0000      1st Qu.:0.2272   1st Qu.:0.04494
##  Median :0.3636    Median :0.0000      Median :0.2872   Median :0.06742
##  Mean   :0.3649    Mean   :0.3512      Mean   :0.2888   Mean   :0.10044
##  3rd Qu.:0.4697    3rd Qu.:1.0000      3rd Qu.:0.3375   3rd Qu.:0.10112
##  Max.   :1.0000    Max.   :1.0000      Max.   :1.0000   Max.   :1.00000
##   serum_sodium         sex            smoking            time
##  Min.   :0.0000   Min.   :0.0000   Min.   :0.0000   Min.   :0.0000
##  1st Qu.:0.6000   1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.:0.2456
##  Median :0.6857   Median :1.0000   Median :0.0000   Median :0.3950
##  Mean   :0.6750   Mean   :0.6488   Mean   :0.3211   Mean   :0.4493
##  3rd Qu.:0.7714   3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:0.7082
##  Max.   :1.0000   Max.   :1.0000   Max.   :1.0000   Max.   :1.0000
```

PROCESSING AND RESULTS

- We divide the normalized dataset into training data and test data

70% for training

30% for test

```
pacientes_heart_failure_train <- pacientes_heart_failure_n[ 1: 209, ]
pacientes_heart_failure_test <- pacientes_heart_failure_n[ 210: 299, ]
```

- We observe the dimensions of the dataset

```
dim(pacientes_heart_failure_train)
```

```
## [1] 209  12
```

```
dim(pacientes_heart_failure_test)
```

```
## [1] 90 12
```

- We extract the labels or the variable we are trying to predict (Dependent Variable)

```
# we extract the column DEATH_EVENT
pacientes_heart_failure_train_labels <- pacientes_heart_failure[ 1:209, 13]
pacientes_heart_failure_test_labels <- pacientes_heart_failure[ 210:299, 13]
```

step 3.- Training a model on data

```
library(class)
# test prediction
# Nota: pasar solo las etiquetas de los datos de entrenamiento
pacientes_hf_test_pred <- knn(train = pacientes_heart_failure_train,
                              test = pacientes_heart_failure_test,
                              cl = pacientes_heart_failure_train_labels,
                              k = 9)
```

step 4.- Evaluating model performance

- We compare reality (Test labels) against predictions

```
library(gmodels)
 CrossTable(x = pacientes_heart_failure_test_labels,
            y = pacientes_hf_test_pred,
            prop.chisq = FALSE)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  90
##
##
##                                    | pacientes_hf_test_pred
## pacientes_heart_failure_test_labels |        No |        Si | Row Total |
## -----------------------------------|-----------|-----------|-----------|
##                                 No |        74 |         9 |        83 |
##                                    |     0.892 |     0.108 |     0.922 |
##                                    |     0.914 |     1.000 |           |
##                                    |     0.822 |     0.100 |           |
## -----------------------------------|-----------|-----------|-----------|
##                                 Si |         7 |         0 |         7 |
##                                    |     1.000 |     0.000 |     0.078 |
##                                    |     0.086 |     0.000 |           |
##                                    |     0.078 |     0.000 |           |
## -----------------------------------|-----------|-----------|-----------|
##                       Column Total |        81 |         9 |        90 |
##                                    |     0.900 |     0.100 |           |
## -----------------------------------|-----------|-----------|-----------|
##
##
```

```r
table(pacientes_heart_failure_test_labels, pacientes_hf_test_pred)
```

```
##                                  pacientes_hf_test_pred
## pacientes_heart_failure_test_labels No Si
##                                 No 74  9
##                                 Si  7  0
```

step 5.- Improving model performance

- Using Transforming-z core normalization

z-core normalización:

fórmula:

$$X_{new} = \frac{X - \mu}{\sigma} = \frac{X - Mean(X)}{StdDev(X)}$$

```r
# [-13] omite la columna DEATH_EVENT en la normalización
pacientes_hf_z <- as.data.frame(scale(pacientes_heart_failure[-13]))
summary(pacientes_hf_z) # normaliza en desviaciones standard
```

```
##       age              anaemia        creatinine_phosphokinase
##  Min.   :-1.75151   Min.   :-0.8696   Min.   :-0.575952
##  1st Qu.:-0.82674   1st Qu.:-0.8696   1st Qu.:-0.479589
##  Median :-0.07011   Median :-0.8696   Median :-0.342001
##  Mean   : 0.00000   Mean   : 0.0000   Mean   : 0.000000
##  3rd Qu.: 0.77060   3rd Qu.: 1.1460   3rd Qu.: 0.000165
##  Max.   : 2.87235   Max.   : 1.1460   Max.   : 7.502063
##     diabetes       ejection_fraction   high_blood_pressure   platelets
##  Min.   :-0.8462   Min.   :-2.034976   Min.   :-0.7345     Min.   :-2.43607
##  1st Qu.:-0.8462   1st Qu.:-0.683035   1st Qu.:-0.7345     1st Qu.:-0.52000
##  Median :-0.8462   Median :-0.007065   Median :-0.7345     Median :-0.01388
##  Mean   : 0.0000   Mean   : 0.000000   Mean   : 0.0000     Mean   : 0.00000
##  3rd Qu.: 1.1779   3rd Qu.: 0.584409   3rd Qu.: 1.3570     3rd Qu.: 0.41043
##  Max.   : 1.1779   Max.   : 3.541779   Max.   : 1.3570     Max.   : 5.99812
##  serum_creatinine     serum_sodium           sex              smoking
##  Min.   :-0.864061   Min.   :-5.35423   Min.   :-1.3570   Min.   :-0.6865
##  1st Qu.:-0.477404   1st Qu.:-0.59500   1st Qu.:-1.3570   1st Qu.:-0.6865
##  Median :-0.284076   Median : 0.08489   Median : 0.7345   Median :-0.6865
##  Mean   : 0.000000   Mean   : 0.00000   Mean   : 0.0000   Mean   : 0.0000
##  3rd Qu.: 0.005916   3rd Qu.: 0.76478   3rd Qu.: 0.7345   3rd Qu.: 1.4517
##  Max.   : 7.739045   Max.   : 2.57782   Max.   : 0.7345   Max.   : 1.4517
##       time
##  Min.   :-1.6268
##  1st Qu.:-0.7378
##  Median :-0.1966
##  Mean   : 0.0000
##  3rd Qu.: 0.9372
##  Max.   : 1.9937
```

```r
pacientes_hf_train <- pacientes_hf_z[ 1: 209, ]
pacientes_hf_test <- pacientes_hf_z[ 210: 299, ]
pacientes_hf_train_labels <- pacientes_heart_failure[ 1: 209, 13]
pacientes_hf_test_labels <- pacientes_heart_failure[ 210: 299, 13]
pacientesHF_test_pred <- knn(train = pacientes_hf_train,
                             test = pacientes_hf_test,
                             cl = pacientes_hf_train_labels,
                             k = 9)
CrossTable(x = pacientes_hf_test_labels, y = pacientesHF_test_pred, prop.chisq = FALSE)
```

```
##
##
##    Cell Contents
## |-------------------------|
## |                       N |
## |           N / Row Total |
## |           N / Col Total |
## |         N / Table Total |
## |-------------------------|
##
##
## Total Observations in Table:  90
##
##
##                          | pacientesHF_test_pred
## pacientes_hf_test_labels |        No |        Si | Row Total |
## -------------------------|-----------|-----------|-----------|
##                       No |        78 |         5 |        83 |
##                          |     0.940 |     0.060 |     0.922 |
##                          |     0.951 |     0.625 |           |
##                          |     0.867 |     0.056 |           |
## -------------------------|-----------|-----------|-----------|
##                       Si |         4 |         3 |         7 |
##                          |     0.571 |     0.429 |     0.078 |
##                          |     0.049 |     0.375 |           |
##                          |     0.044 |     0.033 |           |
## -------------------------|-----------|-----------|-----------|
##             Column Total |        82 |         8 |        90 |
##                          |     0.911 |     0.089 |           |
## -------------------------|-----------|-----------|-----------|
##
##
```