Data Analysis

HEART ATTACK

María Nieto

Index

Introduction

Heart attacks represent a significant threat to global cardiovascular health, being one of the leading causes of morbidity and mortality worldwide. Data analysis related to the prediction and study of heart attacks has become an essential tool for understanding risk factors, identifying underlying patterns, and developing effective preventive strategies. In this context, the present work focuses on an exhaustive data analysis that covers clinical, lifestyle variables and relevant biomarkers, with the aim of discerning patterns and relationships that may contribute to the prediction and understanding of adverse cardiac events.

The relevance of this analysis lies in the need to advance the predictive capacity of existing models, as well as in the identification of new key variables that can improve the precision in the evaluation of cardiovascular risk. By addressing this data set, we seek to not only predict future events, but also deepen our understanding of the complex interactions between various factors that contribute to the development of heart attacks.

Through advanced statistical analysis and predictive modeling techniques, we aim to identify risk patterns, evaluate the relevance of specific variables, and provide a solid foundation for personalized intervention strategies. The application of advanced analytical approaches not only allows the identification of linear relationships between variables, but also the discovery of non-linear patterns and complex interconnections that may be crucial for an accurate assessment of cardiovascular risk.

In summary, this work aims to contribute to the advancement of research in the prediction and analysis of heart attacks, using a comprehensive approach that leverages advanced data analysis capabilities to improve understanding of risk factors and, ultimately, promote cardiovascular health and disease prevention.

About the Data

These are the categories that are in the data set:

- Age : Age of the patient

- Sex : Sex of the patient
- cp : Chest Pain type:
  - Value 0: typical angina
  - Value 1: atypical angina
  - Value 2: non-anginal pain
  - Value 3: asymptomatic
- trtbps : resting blood pressure (in mm Hg)
- chol: cholesterol in mg/dl fetched via BMI sensor
- fbs: fasting blood sugar > 120 mg/dl:
  - 1 = true
  - 0 = false
- rest_ecg: resting electrocardiographic results:
  - Value 0: normal
  - Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)
  - Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria
- thalach: maximum heart rate achieved
- exang: exercise induced angina:
  - 1 = yes
  - 0 = no
- old peak: ST depression induced by exercise relative to rest
- slp: the slope of the peak exercise ST segment:
  - 0 = unsloping
  - 1 = flat
  - 2 = downsloping
- caa: number of major vessels (0-3)

- over: thalassemia:
  - 0 = null
  - 1 = fixed defect
  - 2 = normal
  - 3 = reversable defect
- output: diagnosis of heart disease (angiographic disease status):
  - 0: < 50% diameter narrowing. less chance of heart disease
  - 1: > 50% diameter narrowing. more chance of heart disease

Descriptive Data Analysis

The database is the following, and with the nrow function, it tells us the number of data it contains.

```
heart_heart_csv=read.csv("heart - heart.csv.csv",head=T,sep=",")
head(heart_heart_csv)
```

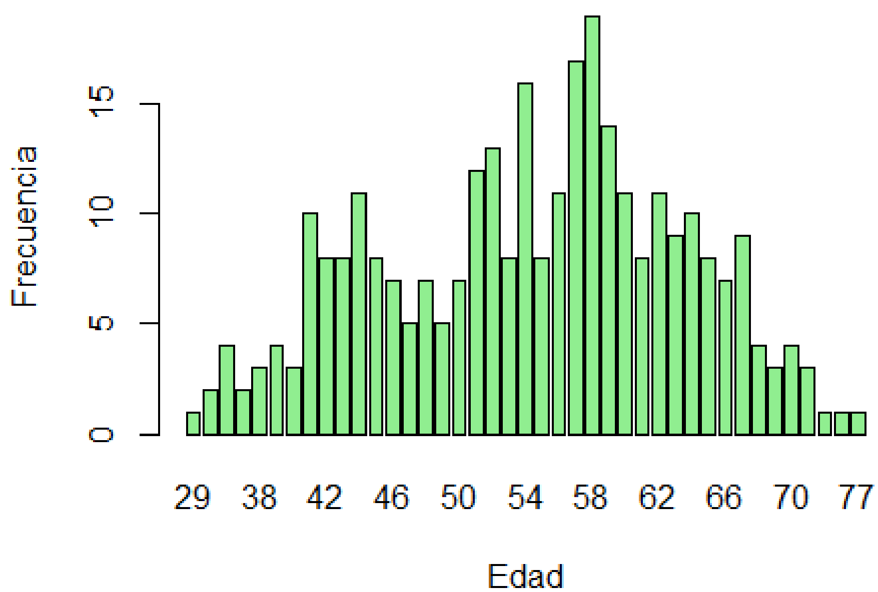
```
##    age     sex  cp  trtbps  chol  fbs  restecg  thalachh  exng  oldpeak  slp  caa
thall
## 1   63    Male   3     145   233    1        0       150     0      2.3    0    0
1
## 2   37    Male   2     130   250    0        1       187     0      3.5    0    0
2
## 3   41  Female   1     130   204    0        0       172     0      1.4    2    0
2
## 4   56    Male   1     120   236    0        1       178     0      0.8    2    0
2
## 5   57  Female   0     120   354    0        1       163     1      0.6    2    0
2
## 6   57    Male   0     140   192    0        1       148     0      0.4    1    0
1
##   output
## 1      1
## 2      1
## 3      1
## 4      1
## 5      1
## 6      1
```

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

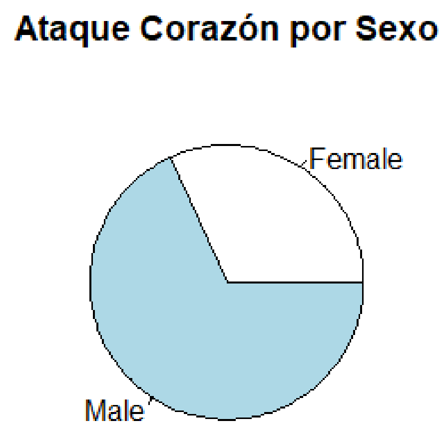
nrow indicates the number of data contained in the heart attack database.

```
##
## 29 34 35 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58
```

```
59
##   1   2   4   2   3   4   3  10   8   8  11   8   7   5   7   5   7  12  13   8  16   8  11  17  19
14
## 60 61 62 63 64 65 66 67 68 69 70 71 74 76 77
## 11  8 11  9 10  8  7  9  4  3  4  3  1  1  1
```




```
##
## Female   Male
##     96    207
```



In view of all these data and graphs related to the qualitative variables of the database, we can say that the highest number of cases of heart attacks occur between 56 and 60 years of age. If we break it down by gender It is found that men are more likely to suffer these acute myocardial infarctions. It's not entirely clear why men face a higher risk of developing heart disease, but on average, a woman's cardiovascular risk is equivalent to that of a man who is 20 years older. One of the crucial factors in this disparity is attributed to hormones. Estrogens produced by the ovaries are presumed to have a protective effect. In fact, it is after menopause that the risk of cardiovascular diseases in women increases significantly.

Numerical Data Summary

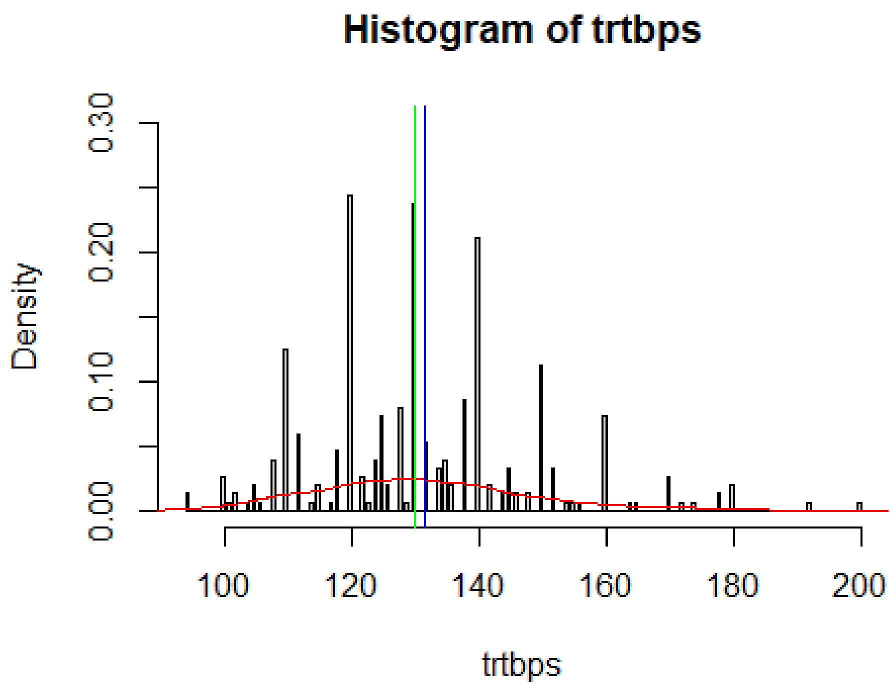
In this section, you can see the relative information of each of the variables that make up the database.

```
##       age             sex                    cp             trtbps
##   Min.   :29.00   Length:303         Min.   :0.000   Min.   : 94.0
## 1st Qu.:47.50 Class:character 1st Qu.:0.000 1st Qu.:120.0
##   Median :55.00   Mode  :character   Median :1.000   Median :130.0
##   Mean   :54.37                      Mean   :0.967   Mean   :131.6
```

```
## 3rd Qu.:61.00 3rd Qu.:2.000 3rd Qu.:140.0
##  Max.   :77.00                      Max.   :3.000   Max.   :200.0
## chol fbs restecg thalachh
## Min. :126.0 Min. :0.0000 Min. :0.0000 Min. : 71.0
## 1st Qu.:211.0 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:133.5
##  Median :240.0   Median :0.0000   Median :1.0000   Median :153.0
##  Mean   :246.3   Mean   :0.1485   Mean   :0.5281   Mean   :149.6
## 3rd Qu.:274.5 3rd Qu.:0.0000 3rd Qu.:1.0000 3rd Qu.:166.0
##  Max.   :564.0   Max.   :1.0000   Max.   :2.0000   Max.   :202.0
##       exng           oldpeak          slp             caa
## Min. :0.0000 Min. :0.00 Min. :0.000 Min. :0.0000
## 1st Qu.:0.0000 1st Qu.:0.00 1st Qu.:1.000 1st Qu.:0.0000
##  Median :0.0000   Median :0.80   Median :1.000   Median :0.0000
##  Mean   :0.3267   Mean   :1.04   Mean   :1.399   Mean   :0.7294
## 3rd Qu.:1.0000 3rd Qu.:1.60 3rd Qu.:2.000 3rd Qu.:1.0000
##  Max.   :1.0000   Max.   :6.20   Max.   :2.000   Max.   :4.0000
##      thall           output
## Min. :0.000 Min. :0.0000
## 1st Qu.:2.000 1st Qu.:0.0000
##  Median :2.000   Median :1.0000
##  Mean   :2.314   Mean   :0.5446
## 3rd Qu.:3.000 3rd Qu.:1.0000
##  Max.   :3.000   Max.   :1.0000
```

In view of the results, it can be seen that the average age is 55 years. Furthermore, the type of chest pain suffered by patients on average is type 1, angina typical. However, two other types exist. Type 2, which is atypical angina; and, type 3, non-anginal pain. On the other hand, the mean arterial pressure at rest (measured in mmHg upon admission to the hospital) is 130.

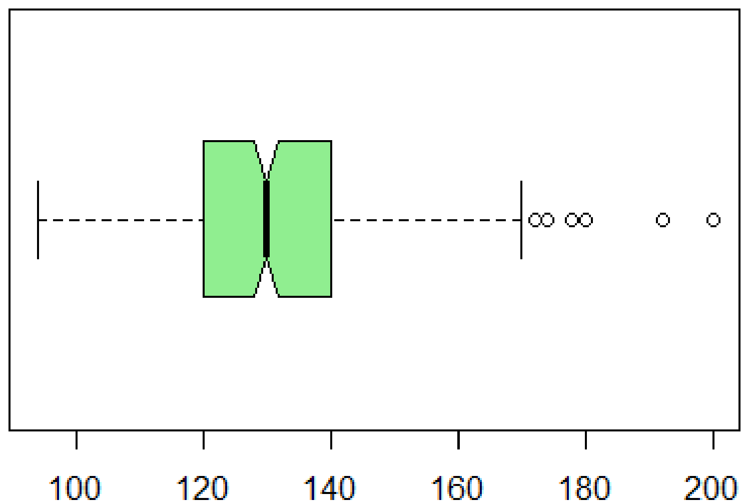
Data Distribution Analysis “trtbps”




```
## [1] 0.7102301
```

At this point, we analyze the quantitative variable of blood pressure at rest and, of which we can say, that we are dealing with a multimodal density function because several local peaks are observed, being between 120 and 140, which indicate where the blood pressure is. the center of distribution. When calculating the skewness coefficient, we obtain a result of 0.7102301, which indicates that the distribution has a right or positive skewness, since this coefficient is positive or greater than 0. Also, it can be seen by comparing the mean and the median , since if the mean takes a value much higher than the median, we are faced with a case of positive asymmetry. In this case, there is not a big difference since the mean value is 131.6, and the median value is 130.

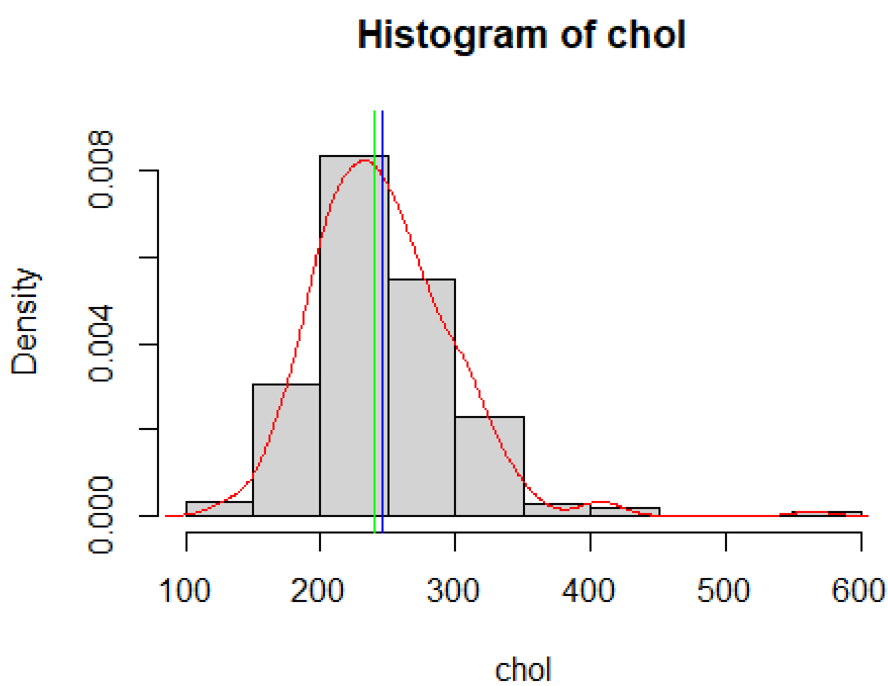
```
## The following objects are masked from heart_heart_csv (pos = 4):
##
##     age, caa, chol, cp, exng, fbs, oldpeak, output, restecg, sex, slp,
## home, over, trtbps
```



On the other hand, the boxplot is similar to the histogram but not identical. While the histogram with its density is good for visualizing the center, spread, tails, and shape of the distribution, it is not useful when comparing distributions. A graph that allows you to observe the above and also compare distributions in the same figure is the boxplot.

When analyzing the boxplot we can observe a large concentration of values around the range between 120 and 140. This confirms, as mentioned above, that in the sample there is a significant proportion of patients with blood pressure in that range. Thanks to the arrangement of this data, we can identify the limits of the box, determined by the interquartile range. Likewise, several outliers are observed, those that are outside said box.

Analysis of the Distribution of “chol” Data



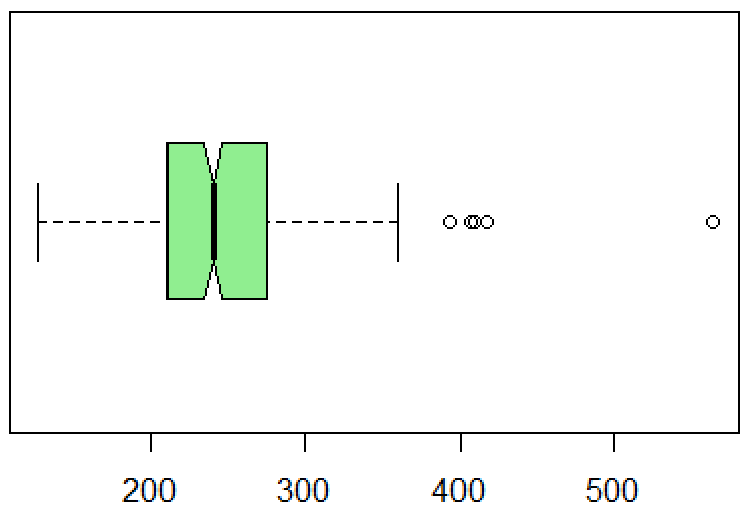
The condition of multimodal density is attributed to the present distribution due to the presence of multiple local points in its representation, indicating the existence of various significant concentrations.

```
## [1] 1.137733
```

When calculating the skewness coefficient, we obtain a result of 1.137733, which indicates that the distribution exhibits a positive skewness or bias to the right, given that said coefficient is positive or greater than 0. This conclusion is also corroborated by comparing the mean and the median; When the mean significantly exceeds the median, a positive skew is suggested. In this particular case, although the mean is 246.3 and the median is 240, the disparity between the two is not considerable.

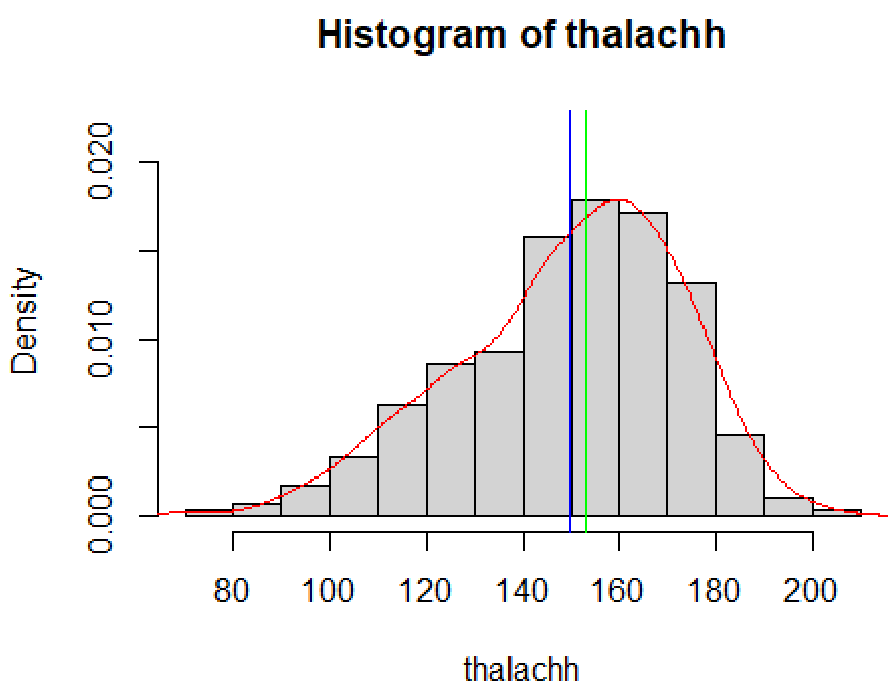
```
## The following objects are masked from heart_heart_csv (pos = 3):
##
```

```
##     age, caa, chol, cp, exng, fbs, oldpeak, output, restecg, sex, slp,
## home, over, trtbps

## The following objects are masked from heart_heart_csv (pos = 5):
##
##     age, caa, chol, cp, exng, fbs, oldpeak, output, restecg, sex, slp,
## home, over, trtbps
```



When examining the box plot, a notable concentration of values is observed in the interval between 200 and 300. In addition, the presence of extreme values can be confirmed, since these are outside the interquartile range.

Data Distribution Analysis “thalachh”



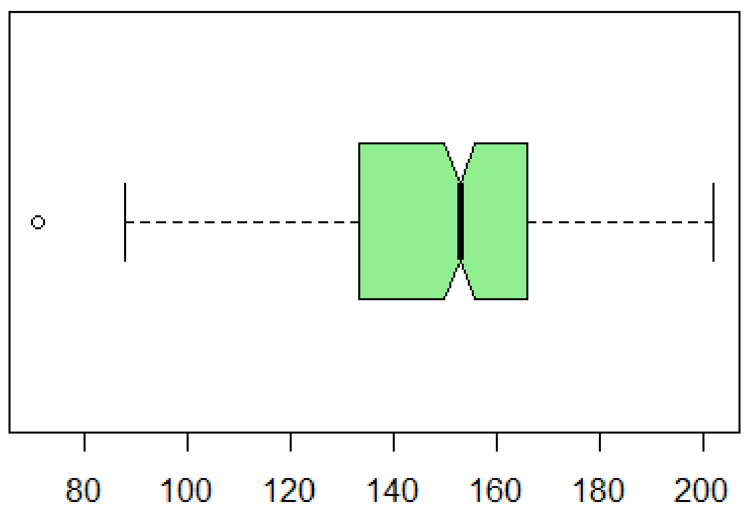
The characterization of this distribution as multimodal is based on the presence of multiple local points in its representation, which indicates the existence of various significant concentrations.

```
## [1] -0.5347455
```

We are facing a negative asymmetry, with a value of -0.5347, this indicates a unilateral distribution that extends towards more values.negatives.

```
## The following objects are masked from heart_heart_csv (pos = 3):
##
##     age, caa, chol, cp, exng, fbs, oldpeak, output, restecg, sex, slp,
## home, over, trtbps

## The following objects are masked from heart_heart_csv (pos = 4):
##
```

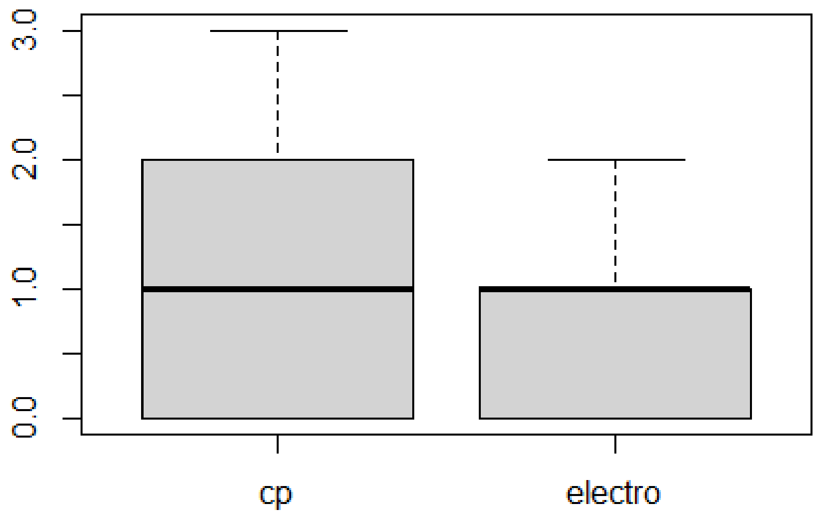
```
##     age, caa, chol, cp, exng, fbs, oldpeak, output, restecg, sex, slp,
## home, over, trtbps

## The following objects are masked from heart_heart_csv (pos = 6):
##
##     age, caa, chol, cp, exng, fbs, oldpeak, output, restecg, sex, slp,
## home, over, trtbps
```



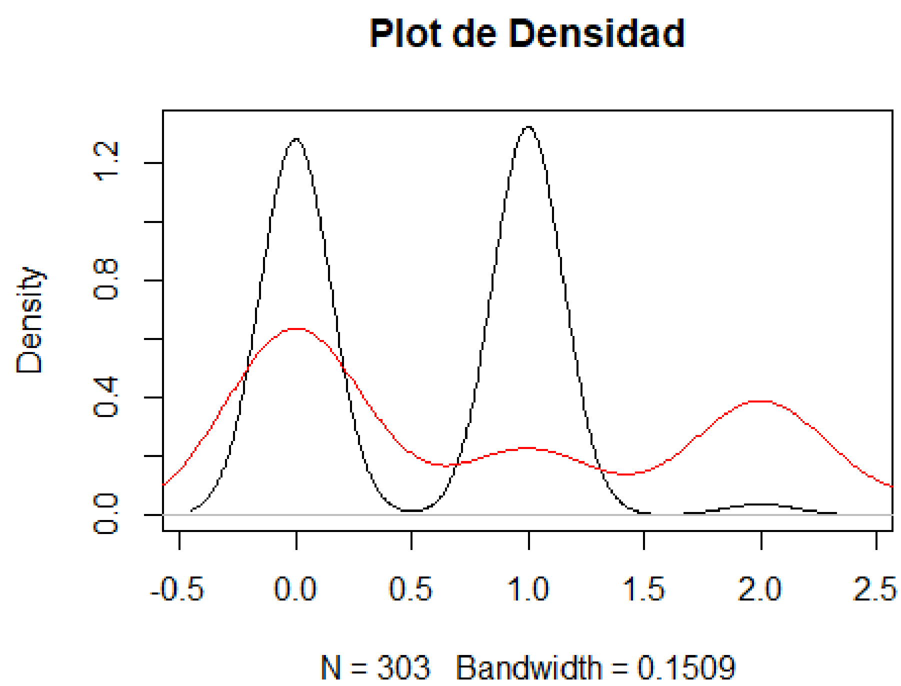
When analyzing the box plot, a notable concentration of values is noted in the interval from 130 to 170. Likewise, the existence of extreme values is confirmed, evidenced by their location outside the interquartile range.

Comparison of Numerical Variables

In this case, we are going to compare the variables “number of shares” cp, which measures chest pain, and “number of sads” restecg, which are the electrographic results at rest, with the aim of being able to relate both magnitudes.



In the box plot provided, the distributions of the variables “cp” (type of chest pain) and “restecg” (resting electrocardiographic results) are compared. Both boxplots present similar interquartile ranges, since the boxes range from the first quartile to the third quartile. It is interesting to note that both boxplots extend to their maximum values.



The density graph shows the distribution of data over a continuous time interval, this graph is an alteration of a Histogram. The peaks on this density graph help show where the values that are concentrated in the interval. Areas of higher density on the graph may indicate regions where the combination of certain levels of chest pain and resting electrocardiographic results is more common. The shape and direction of the joint density curves can reveal patterns in the association between both variables. Low-density areas may point to less common combinations of values, and the presence of outliers could stand out in low-density regions.

Multivariate Data

In this section, we will focus on the analysis and description of a multivariate data set. For the most part, the variables will be numerical or quantitative in nature, so our interest will focus on understanding the relationships between them. In order to quantify the joint variation of two variables, we will use covariance and calculate the matrix of variances and covariances between said variables. This approach will allow us to gain a deeper understanding of how different numerical variables interact with each other, providing a fundamental tool to explore the complexity and interrelationships within the data set.

```
##                  cp      trtbps         chol           fbs      restecg
## cp        1.06513234   0.8617140   -4.1137740  0.034719035  0.024107709
## trtbps    0.86171399 307.5864533  111.9672153  1.109042030 -1.052324438
## chol     -4.11377396 111.9672153 2686.4267480  0.245426529 -4.116702730
## fbs       0.03471903   1.1090420    0.2454265  0.126876926 -0.015769458
## restecg   0.02410771  -1.0523244   -4.1167027 -0.015769458  0.276528315
##thalachh 6.99161804 -18.7591305 -11.8004940 -0.069897056 0.531462418
## exng     -0.19116779   0.5571110    1.6319913  0.004294800 -0.017474264
## oldpeak -0.17882106 3.9344863 3.2467937 0.002376893 -0.035882893
## slp       0.07613708  -1.3128319   -0.1289642 -0.013146679  0.030151028
## caa      -0.19108037   1.8183726    3.7372522  0.050258999 -0.038740629
## over -0.10220095 0.6680218 3.1354884 -0.006983149 -0.003857671
## output    0.22332962  -1.2679496   -2.2038555 -0.004983280  0.035997639
##               thalachh        exng      oldpeak         slp         caa
## cp         6.99161804 -0.19116779 -0.178821061  0.07613708 -0.19108037
## trtbps   -18.75913055  0.55711101  3.934486263 -1.31283195  1.81837257
## chol     -11.80049396  1.63199134  3.246793653 -0.12896422  3.73725220
## fbs       -0.06989706  0.00429480  0.002376893 -0.01314668  0.05025900
## restecg    0.53146242 -0.01747426 -0.035882893  0.03015103 -0.03874063
## thalachh 524.64640570 -4.07629008 -9.153517802 5.45936878 -4.99323542
## exng      -4.07629008  0.22070684  0.157215920 -0.07461806  0.05560291
## oldpeak -9.15351780 0.15721592 1.348095207 -0.41321881 0.26439578
## slp        5.45936878 -0.07461806 -0.413218805  0.37973466 -0.05051035
## caa       -4.99323542  0.05560291  0.264395777 -0.05051035  1.04572378
## over -1.35249055 0.05947151 0.149462330 -0.03952746 0.09506480
## output     4.81876598 -0.10235394 -0.249452495  0.10632090 -0.19982296
##                  thall      output
## cp       -0.102200949  0.22332962
## trtbps    0.668021769 -1.26794964
## chol      3.135488383 -2.20385548
## fbs      -0.006983149 -0.00498328
```

```
## restecg  -0.003857671  0.03599764
##thalachh -1.352490547 4.81876598
## exng      0.059471510 -0.10235394
## oldpeak 0.149462330 -0.24945249
## slp      -0.039527463  0.10632090
## caa       0.095064804 -0.19982296
## over 0.374882521 -0.10507508
## output   -0.105075077  0.24883614
```

In order to evaluate the interdependence between variables, it is studied by calculating the correlation matrix.

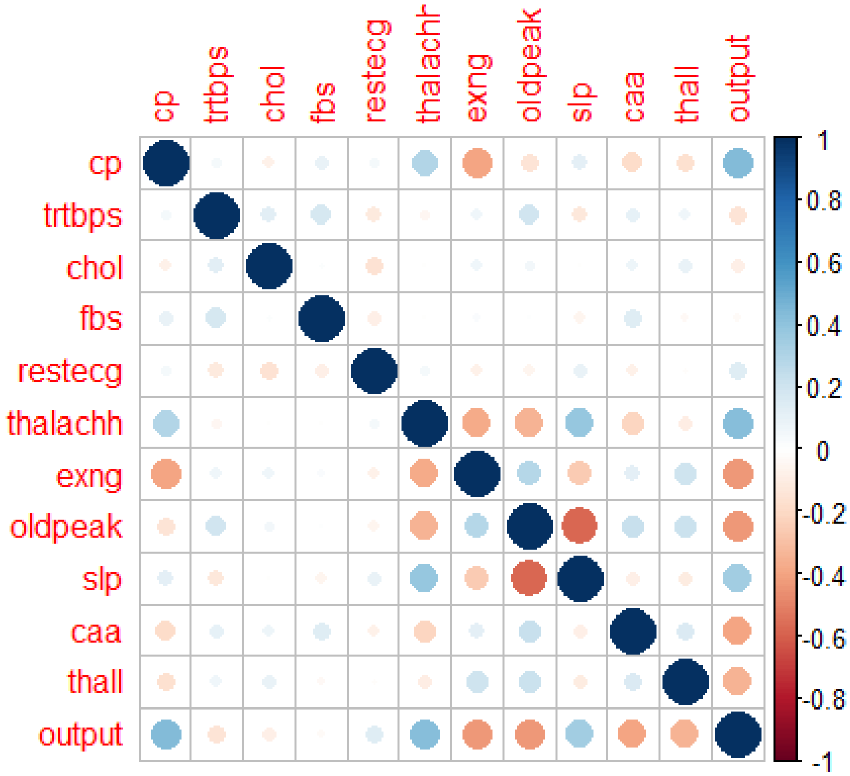
```
##                   cp      trtbps         chol          fbs     restecg
## cp        1.00000000  0.04760776 -0.076904391  0.094444035  0.04442059
## trtbps    0.04760776  1.00000000  0.123174207  0.177530542 -0.11410279
## chol     -0.07690439  0.12317421  1.000000000  0.013293602 -0.15104008
## fbs       0.09444403  0.17753054  0.013293602  1.000000000 -0.08418905
## restecg   0.04442059 -0.11410279 -0.151040078 -0.084189054  1.00000000
## thalachh 0.29576212 -0.04669773 -0.009939839 -0.008567107 0.04412344
## exng     -0.39428027  0.06761612  0.067022783  0.025665147 -0.07073286
## oldpeak -0.14923016 0.19321647 0.053951920 0.005747223 -0.05877023
## slp       0.11971659 -0.12147458 -0.004037770 -0.059894178  0.09304482
## caa      -0.18105303  0.10138899  0.070510925  0.137979327 -0.07204243
## over -0.16173557 0.06220989 0.098802993 -0.032019339 -0.01198140
## output    0.43379826 -0.14493113 -0.085239105 -0.028045760  0.13722950
##              thalachh        exng      oldpeak          slp         caa
## cp        0.295762125 -0.39428027 -0.149230158  0.11971659 -0.18105303
## trtbps   -0.046697728  0.06761612  0.193216472 -0.12147458  0.10138899
## chol     -0.009939839  0.06702278  0.053951920 -0.00403777  0.07051093
## fbs      -0.008567107  0.02566515  0.005747223 -0.05989418  0.13797933
## restecg   0.044123444 -0.07073286 -0.058770226  0.09304482 -0.07204243
## thalachh 1.000000000 -0.37881209 -0.344186948 0.38678441 -0.21317693
## exng     -0.378812094  1.00000000  0.288222808 -0.25774837  0.11573938
## oldpeak -0.344186948 0.28822281 1.000000000 -0.57753682 0.22268232
```

```
## slp       0.386784410 -0.25774837 -0.577536817  1.00000000 -0.08015521
## caa      -0.213176928  0.11573938  0.222682322 -0.08015521  1.00000000
## over -0.096439132 0.20675379 0.210244126 -0.10476379 0.15183213
## output    0.421740934 -0.43675708 -0.430696002  0.34587708 -0.39172399
##                thall      output
## cp       -0.16173557  0.43379826
## trtbps    0.06220989 -0.14493113
## chol      0.09880299 -0.08523911
## fbs      -0.03201934 -0.02804576
## restecg  -0.01198140  0.13722950
##thalachh -0.09643913 0.42174093
## exng      0.20675379 -0.43675708
## oldpeak 0.21024413 -0.43069600
## slp      -0.10476379  0.34587708
## caa       0.15183213 -0.39172399
## over 1.00000000 -0.34402927
## output   -0.34402927  1.00000000

## [1] 0.1287076
```

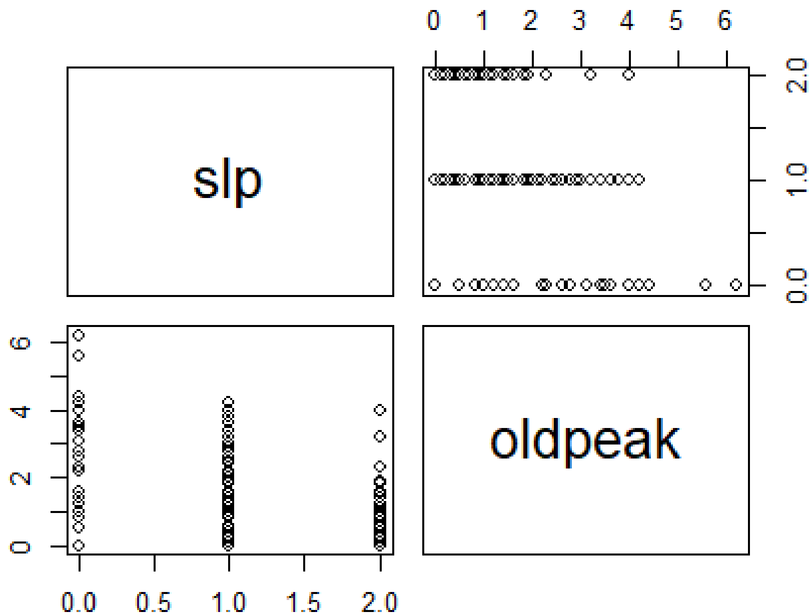
Since the determinant of the correlation matrix is close to 0, this suggests a high degree of dependence between the variables.

Linear Dependence between Variables

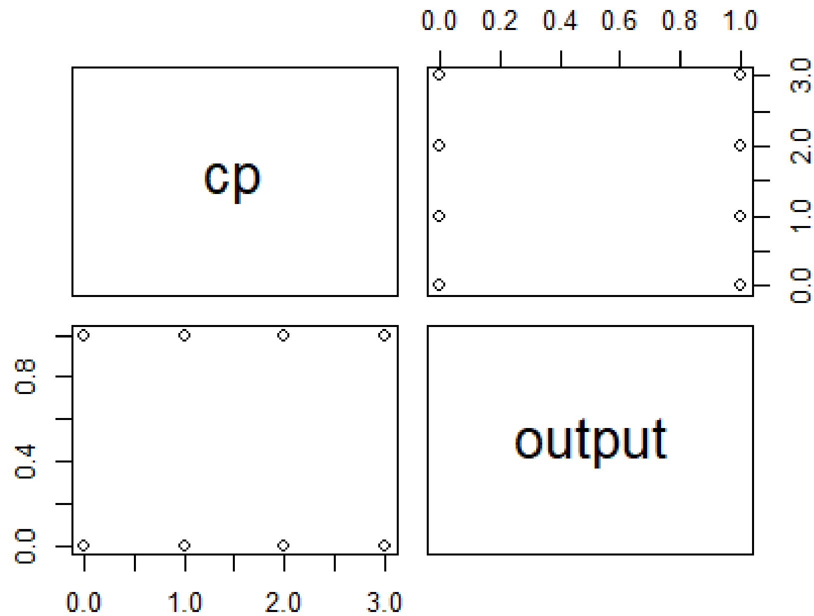
Furthermore, at this point, it is relevant to create scatter plots to visualize the relationships between the variables in pairs. These diagrams are essential to identify possible non-linear relationships, since in such cases the variance-covariance matrix may not be sufficient to summarize the dependence between the variables.



As evident in the scatter diagram, strong linear relationships between the variables are rare. Therefore, below, we will focus on analyzing the strongest relationship identified and the one that presents a minimal connection.



We can conclude that the relationship between both variables is practically null. This is because the “slp” variable, which measures the slope of the ST segment during maximal exercise, and the “oldpeak” variable, which assesses ST segment depression induced by exercise compared to rest, are linked to physical activity and changes in the electrocardiogram, but they measure different concepts.



If “output” assigns probabilities to the chance of having a heart attack, it is reasonable to assume that there could be a connection between “cp” and predictions of heart attack risk. In this sense, it would be plausible to anticipate that higher levels of chest pain would correlate with a greater probability that the models predict a higher risk of heart attack. Therefore, it is feasible that a linear relationship exists between both variables.

Selection of Main Components

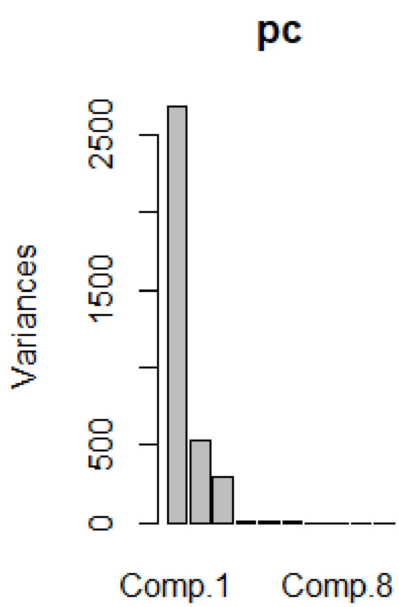
Next, we will proceed to calculate the main components. To carry out this calculation, it is necessary to select the appropriate matrix for the determination of the eigenvalues and eigenvectors.

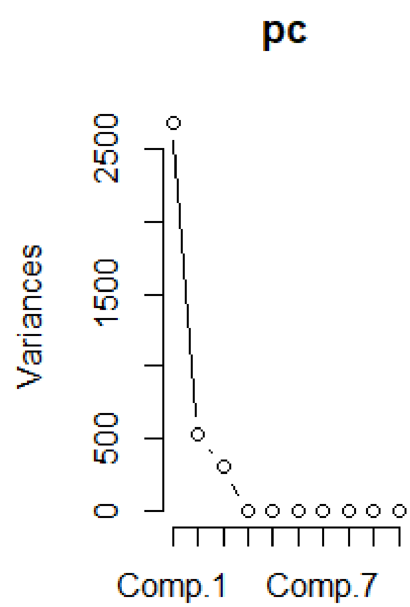
```
##
## Loadings:
##             Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9
Comp.10
##  cp                                            0.353   0.663   0.605   0.165
0.173
## trtbps                 0.996
## chol     -0.999
## fbs                                                           0.104  0.112
0.141
## restecg                                                      -0.917  0.365
0.106
## thalachh 0.996
## exng                            -0.112        -0.130  0.112  0.105  0.105
0.866
## oldpeak -0.737 0.558 -0.113 -0.116 -0.146 -0.306
## slp                              0.217 -0.226                -0.295 -0.858
0.211
## caa                           -0.445 -0.428  0.767
## over -0.165 0.945 -0.219
## output                              0.214                    -0.176 -0.154
-0.311
##          Comp.11 Comp.12
## cp
## trtbps
## chol
## fbs      -0.832   0.509
## restecg
```

```
## thalachh
## exng             -0.416
## old peak
## slp               0.126
## caa              -0.127
## over -0.136 -0.101
## output   -0.516  -0.719
##
##                    Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8
Comp.9
## SS loadings      1.000   1.000   1.000   1.000   1.000   1.000   1.000   1.000
1.000
## Proportion Var 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083 0.083
## Cumulative Var   0.083   0.167   0.250   0.333   0.417   0.500   0.583   0.667
0.750
##                Comp.10 Comp.11 Comp.12
## SS loadings      1.000   1.000   1.000
## Proportion Was 0.083 0.083 0.083
## Cumulative Var   0.833   0.917   1.000
```

To make the decision on which components to retain, we have three study approaches to guide our choice.

Sedimentation Charts





In the left graph, it can be seen that the first three components explain most of the variability in the data. In the right graph, it is observed that a straight line can be approximated from the second component. This visual approach suggests choosing the first two or three components as the most relevant candidates to retain.

Variance Proportion

```
## Importance of components:
##                            Comp.1     Comp.2      Comp.3       Comp.4
## Standard deviation     51.7967661 22.9076588 17.31797723 1.1991291037
## Proportion of Variance  0.7638997  0.1494144  0.08539362 0.0004094143
## Cumulative Proportion   0.7638997  0.9133142  0.99870777 0.9991171859
##                              Comp.5       Comp.6       Comp.7       Comp.8
## Standard deviation     1.0104314122 0.9253485447 5.918983e-01 5.255744e-01
## Proportion of Variance 0.0002906998 0.0002438045 9.975285e-05 7.865013e-05
```

```
## Cumulative Proportion  0.9994078858 0.9996516903 9.997514e-01 9.998301e-01
##                              Comp.9      Comp.10      Comp.11      Comp.12
## Standard deviation     4.582512e-01 4.015061e-01 3.412882e-01 0.3302322094
## Proportion of Variance 5.979134e-05 4.590028e-05 3.316451e-05 0.0000310506
## Cumulative Proportion  9.998899e-01 9.999358e-01 9.999689e-01 1.0000000000
```

From the previous operation, the first row provides us with the standard deviations associated with each component. Furthermore, the eigenvalues reflect the variance related to each component; Therefore, the square of these values (the variance) is equivalent to the eigenvalues. The second row indicates the proportion of variance explained by each component, while the third row displays the proportion of variance accumulated by adding the variance explained by the current component to the previous ones. When focusing on the last row, and as evidenced in the scree plot, it is observed that the first component alone covers 66.3% of the variability. To reach a level of 90%, it would be necessary to retain the first two main components. However, if the goal is to reach 99%, it would be necessary to conserve the first three components.

Calculation of Eigenvalues

```
##          Comp.1          Comp.2          Comp.3          Comp.4          Comp.5
Comp.6
##  2682.9049737    524.7608331    299.9123354       1.4379106       1.0209716
0.8562699
##          Comp.7          Comp.8          Comp.9         Comp.10         Comp.11
Comp.12
##       0.3503437       0.2762284       0.2099941       0.1612071       0.1164776
0.1090533

## [1] 292.6764
```

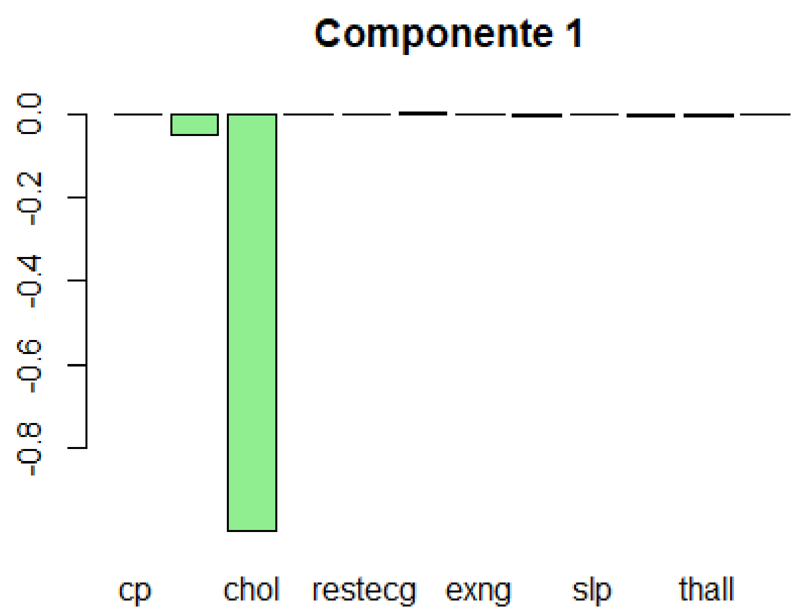
The reference value in this situation is 292.6764. When observing carefully, we notice that the first three components have an associated eigenvalue greater than said value. Therefore, this third method also supports the decision to retain the first three components.

After performing these calculations using the three methods, we can affirm that we chose to retain the first three principal components.

```
##                 Comp.1        Comp.2        Comp.3
## cp        1.527711e-03  0.0130660971  0.0053855273
## trtbps -4.695977e-02 -0.0808990493 0.9955018118
## chol -9.988741e-01 0.0097270963 -0.0463424267
## fbs      -1.105958e-04 -0.0002987658  0.0036155697
## restecg   1.547379e-03  0.0010959148 -0.0027081078
## thalachh 5.863633e-03 0.9962467354 0.0815078358
## exng     -6.245369e-04 -0.0077866839  0.0004926436
## oldpeak  -1.294510e-03 -0.0179360342  0.0100901359
## slp       8.301833e-05  0.0105573834 -0.0028630768
## caa      -1.430361e-03 -0.0096975728  0.0041110936
## over -1.178518e-03 -0.0026186910 0.0013683051
## output    8.509534e-04  0.0092985993 -0.0025613835
```

The first component can be interpreted as follows:

R1 = 1.527711e-03X1 + (-4.695977e-02X2) + (-9.988741e-01X3) + (-1.105958e-04X4) + 1.547379e-03X5 + 5.863633e-03X6 + (-6.245369e-04X7) + (-1.294510e-03X8) + 8.301833e-05X9 + (-1.430361e-03X10) + (-1.178518e-03X11) + 8.509534e-04X12



As evident, the coefficients have both positive and negative values. Consequently, we will proceed to classify the variables into two groups according to the sign of the coefficients that accompany them:

```
R1= (1.527711e-03*given_num[,1] + 1.547379e-03*given_num[,5] +
5.863633e-03*given_num[,6] + 8.301833e-05*given_num[,9] +
8.509534e-04*given_num[,12]) - (4.695977e-02*given_num[,2] +
9.988741e-01*given_num[,3] + 1.105958e-04*given_num[,4] +
6.245369e-04*given_num[,7] + 1.294510e-03*given_num[,8] +
1.430361e-03*given_num[,10] + 1.178518e-03*given_num[,11])
R1

## [1] -238.6661 -254.7282 -208.8682 -240.3250 -358.2820 -197.4896
-299.3478
## [8] -267.3241 -205.9018 -173.8334 -244.3684 -279.9771 -270.8016
-215.0834
## [15] -288.7736 -223.4610 -344.2406 -232.1198 -252.7649 -244.4203
-239.1338
```

```
##     [22]  -237.7907  -231.2760  -248.9659  -204.3054  -308.2253  -217.8836
-179.2448
##     [29]  -422.1851  -201.9918  -201.7225  -181.6165  -223.7556  -277.6719
-217.8978
##     [36]  -182.5317  -308.9973  -237.8168  -275.1059  -366.2222  -313.3966
-249.7736
##     [43]  -211.7877  -268.9710  -326.1441  -329.2592  -240.2526  -262.2746
-221.0893
##     [50]  -239.2801  -260.9418  -306.4117  -235.9932  -144.8844  -257.0456
-206.1406
##     [57]  -226.3898  -264.0242  -186.3128  -307.7402  -269.1048  -312.8085
-190.2134
##     [64]  -208.3342  -216.3677  -188.2088  -225.6065  -238.8158  -224.3889
-213.6317
##     [71]  -262.4828  -230.2556  -208.6904  -266.1918  -217.5189  -255.1158
-249.6223
##     [78]  -226.3628  -209.6995  -243.7585  -253.9252  -312.6671  -321.4918
-303.7568
##     [85]  -268.7784  -567.8288  -281.3446  -200.6060  -217.9971  -251.7041
-259.5095
##     [92]  -211.9821  -228.2409  -292.9441  -164.2685  -231.7661  -399.2131
-236.9529
##     [99]  -319.8008  -250.8161  -249.6323  -277.2082  -200.3064  -244.2266
-200.8781
##  [106]  -215.7236  -240.4805  -241.3258  -248.4102  -257.9486  -332.1843
-131.8874
##  [113]  -318.4403  -214.9849  -266.8992  -219.3930  -218.8797  -197.4684
-207.6908
##  [120]  -248.3176  -308.0536  -276.1095  -271.9481  -270.7901  -202.1373
-214.1782
##  [127]  -208.1912  -282.8156  -201.1734  -273.6249  -207.3297  -276.0359
-299.3514
##  [134]  -239.0021  -310.6149  -273.8459  -182.8687  -212.9557  -205.2017
```

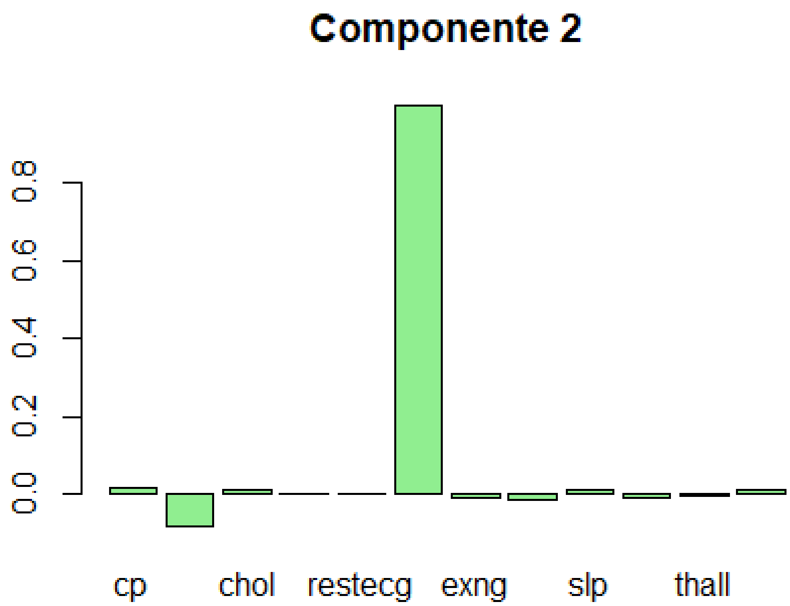
```
-268.1024
## [141] -299.3815 -306.9993 -213.3823 -226.8971 -202.6691 -251.2112
-246.3937
## [148] -245.7674 -230.3865 -185.0193 -234.4508 -153.3607 -233.8175
-283.6517
## [155] -225.3383 -202.1155 -257.7671 -196.4909 -224.7837 -225.9012
-244.3724
## [162] -346.8401 -161.3895 -180.2715 -180.2715 -292.5674 -233.6312
-273.3443
## [169] -258.9635 -208.4452 -260.9849 -232.9244 -288.3803 -228.9394
-211.1089
## [176] -171.3158 -234.3044 -340.2683 -181.7396 -282.0804 -358.0328
-231.1312
## [183] -334.7445 -234.0381 -249.0266 -294.0395 -257.9812 -270.8927
-238.3573
## [190] -176.0488 -309.9328 -221.0107 -192.7661 -287.6696 -190.4603
-332.8038
## [197] -236.9243 -258.6334 -271.7618 -251.9645 -200.9095 -262.7618
-276.0941
## [204] -281.2679 -170.4945 -259.7836 -243.0709 -263.8426 -192.6142
-182.4291
## [211] -233.8758 -264.5302 -223.4772 -312.6128 -253.7513 -346.0254
-268.2417
## [218] -334.9699 -259.3174 -260.9440 -412.6956 -222.6890 -287.1437
-296.3001
## [225] -243.1653 -179.8863 -285.7200 -202.6546 -294.7257 -313.7557
-246.9044
## [232] -295.7050 -295.3445 -250.8026 -327.1128 -304.2278 -304.5358
-298.2555
## [239] -308.5849 -286.6887 -275.5586 -256.0535 -217.8030 -280.3189
-189.3817
## [246] -278.5453 -413.9620 -252.5359 -290.5575 -259.4396 -303.5357
-252.0921
```

```
## [253] -299.5355 -302.6333 -279.4708 -314.4668 -263.9678 -205.8035
-249.8694
## [260] -235.3108 -235.1414 -234.0644 -286.9095 -272.7844 -210.3047
-216.2512
## [267] -334.4027 -153.6391 -290.7368 -288.1885 -253.5149 -239.1813
-241.9540
## [274] -237.5212 -279.1697 -216.6522 -224.0009 -265.7031 -324.1397
-171.5696
## [281] -320.3014 -208.8668 -222.8866 -228.8273 -212.5400 -316.5283
-209.1130
## [288] -238.0111 -338.9577 -210.0228 -208.7807 -322.1825 -231.8821
-218.0211
## [295] -173.6044 -192.5315 -201.8051 -182.9809 -246.5846 -268.0932
-198.7274
## [302] -136.2885 -240.8210
```

Therefore, it can be interpreted as the sum of two weighted averages. Where a high negative weight is seen in the chol variable, which measures cholesterol in mg/dl obtained through the BMI sensor; This negative value separates those affected by heart attacks into those who have a higher cholesterol level and those who do not.

The second component can be interpreted as follows:

R2 = 0.0130660971 0.0179360342X8) + 0.0105573834X9 + (-0.0096975728X10) + (-0.0026186910 X11) + 0.0092985993X12




```
R2=     (0.0130660971*given_num[,1]     +     0.0097270963*given_num[,3]     +
0.0010959148*given_num[,5]        +        0.9962467354*given_num[,6]        +
0.0105573834*given_num[,9]        +        0.0092985993*given_num[,12])       -
(0.0808990493*given_num[,2]        +        0.0002987658*given_num[,4]        +
0.0077866839*given_num[,7]        +        0.0179360342*given_num[,8]        +
0.0096975728*given_num[,10] + 0.0026186910*given_num[,11])
R2

##      [1]  139.97739  178.18155  162.83502  169.94462  156.13145  137.99741
143.96402
##      [8]  165.23774  149.45355  162.86993  150.40314  130.68519  162.45730
136.62680
##     [15]  152.05602  149.84248  165.00617  103.63332  160.62529  141.44596
151.75462
##     [22]  170.11515  168.23065  126.73045  167.97170  151.39847  146.36143
115.38334
##     [29]  149.16803  142.83724  160.83066  131.50483  178.94599  143.98415
116.51211
```

```
##    [36]  149.63080  161.44972  154.52253  137.55981  141.02812  133.15182
171.21477
##    [43]  141.00728  134.52180  173.16475  164.84361  170.33572  146.80159
106.37089
##    [50]  150.53673  140.45636  143.67040  137.15996  167.00869  162.93560
148.53695
##    [57]  177.61680  177.55648  165.63555  151.01094  123.23217  149.71979
181.60647
##    [64]  122.58920  155.15827  171.93406  136.54533  166.11937  161.83336
154.41597
##    [71]  139.28093  148.05784  192.74753  176.53219  156.62116  151.90870
157.64512
##    [78]  154.23984  174.98741  147.28168  171.75165  162.04105  154.28370
167.96441
##    [85]  115.87194  155.59160  143.60373  149.19635  150.60284  115.86120
166.82482
##    [92]  158.72015  159.38439  150.54642  130.00647  101.30882  148.91044
139.97005
##    [99]  153.94776  164.24859  167.76290  132.64737  168.91903  185.91251
153.91116
##  [106]  106.92675  119.88264  142.56427  154.07711  152.00019  142.03996
161.47764
##  [113]  124.26573  153.57290  146.48920  161.79779  158.93905  153.57853
164.88370
##  [120]  142.63595  113.93298  172.81406  164.94413  160.28455  172.71175
183.80276
##  [127]  135.41140  161.79490  159.30890  113.47173  151.44222  153.21632
154.59292
##  [134]  145.85206  155.21077  154.51420   87.71533  131.18064  118.56698
96.80108
##  [141]  149.61289  173.95877  164.71761  135.06238  106.17821  132.26441
141.27540
##  [148]  160.60717  160.90854  140.72342  126.74252  116.90650  142.91389
```

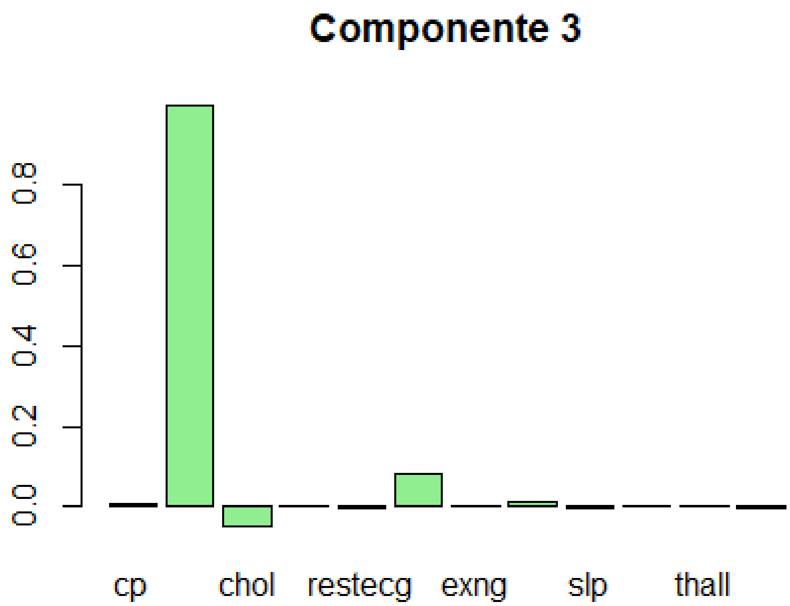
```
142.35343
## [155] 142.44724 121.91264 170.32465 165.38419 135.46731 154.05665
161.01054
## [162] 158.04277 173.17551 162.90247 162.90247  97.37429 120.96433
150.59127
## [169] 138.36997 144.99543 133.44582 160.68643 152.44019 163.81348
122.92395
## [176] 106.25670 152.13328 149.38280 111.51349 102.10915 124.22435
103.58154
## [183] 161.07464 157.53155 117.70448 146.19209 135.37537 101.09258
153.33823
## [190] 150.19441 133.89141 122.18066 104.66568 132.40501 144.86953
128.81611
## [197] 136.52835 154.72704  91.46201 150.90875 169.35919 132.80300
101.06597
## [204] 137.53246 132.95902 152.51793 134.85661 146.72214 130.56407
151.78467
## [211] 141.32140 132.21745 132.04096 136.68486 135.73521 128.06863
88.67599
## [218] 124.14143 118.01543 141.39605 145.14793 101.25331 164.93532
119.01533
## [225] 118.88906 114.43181  95.61980 121.69747 147.49011 123.39147
145.09920
## [232] 112.95276 134.29393  88.26249 101.13932 163.91109 163.15780
160.84790
## [239] 154.21563 147.96972 101.21281 130.80752 121.78902  78.00307
95.66981
## [246] 158.00443 142.51636 108.98385 181.50489 136.56071 113.00650
134.14123
## [253]  97.24080 119.31133 114.29757 137.93210 121.59803 115.80182
143.63407
## [260] 173.82303 152.15684 152.58329  87.37664 162.19214 100.68737
124.50262
```

```
## [267] 105.11854 117.42891 108.39763  94.80467 136.17260 135.87103
63.31943
## [274] 149.60359 111.30307 159.29624  94.87338 133.00012 143.53943
114.90474
## [281] 116.56157 147.02937 125.41000 170.20750 128.13139 111.19320
152.55825
## [288] 153.20194 136.75547 121.10259 150.40192 133.23683 133.81802
139.21698
## [295] 135.33599 133.86696 127.37294  78.07704 113.54914 125.19510
130.62178
## [302] 105.29054 165.13434
```

We can interpret this second principal component as the contrast between two variables, the negative being the pressure arterial resting (trtbps), and the positive one, the maximum heart rate reached (land). This is because, in general, when blood pressure decreases, the body can respond by increasing heart rate to compensate for decreased blood flow and maintain an adequate supply of oxygen to organs and tissues.

The third component can be interpreted as follows:

R3 = 0.0053855273X1 + 0.9955018118X2 + (-0.0463424267X3) + 0.0036155697X4 + (-0.0027081078X5) + 0.0815078358X6 + 0.0004926436X7 + 0. 0100901359X8 + (-0.0028630768X9) + 0.0041110936X10 + 0.0013683051X11 + (-0.0025613835X12)




```
R3=     (0.0053855273*given_num[,1]     +     0.9955018118*given_num[,2]     +
0.0036155697*given_num[,4]              +      0.0815078358*given_num[,6]     +
0.0004926436*given_num[,7]        +         0.0100901359*given_num[,8]        +
0.0041110936*given_num[,10]      +      0.0013683051*given_num[,11]      )    -
(0.0463424267*given_num[,3]        +         0.0027081078*given_num[,5]       +
0.0028630768*given_num[,9]   + 0.0025613835*given_num[,12] )
R3

##       [1]   145.81794   133.11515   133.99469   123.03700   116.33906   142.53294
138.23209
##       [8]   121.37151   175.22098   155.74076   141.33952   128.00519   131.02917
111.49620
##     [15]   149.43895   122.21098   117.72565   148.18363   151.82341   140.63638
136.67240
##     [22]   133.21439   143.39700   149.25021   144.68039   158.49853   152.31968
111.42930
##     [29]   132.86321   132.70164   109.04644   122.66585   134.58955   124.19554
124.78460
```

```
##    [36]  146.22547  134.16711  152.04535  153.91041  154.91799  136.69537
132.73747
##    [43]  105.98420  128.83780  139.33398  118.41742  143.15641  138.18969
126.79015
##    [50]  139.57082  129.70651  117.77384  130.64752  115.25521  136.73902
136.96996
##    [57]  126.31811  117.50708  123.22786  126.34078  107.83342  105.90810
124.34811
##    [64]  135.74289  143.04963  143.73888  100.93576  132.84373  123.11834
127.03418
##    [71]  119.49901   95.61813  136.42580  142.43028  125.03647  135.94672
126.65005
##    [78]  142.49348  132.92222  105.97371  114.50301  127.00693   99.85217
152.04248
##    [85]   99.20776  101.41243  116.95813  104.13356  112.48768   98.00862
125.90313
##    [92]  135.50028  140.83867  131.02742  115.32949  139.93155  133.91749
108.70813
##    [99]  128.04743  132.13702  150.56201  176.56558  144.92873  124.17185
132.62491
##  [106]  119.07858  159.13589  138.83145  121.36516  110.68842  176.67348
157.60061
##  [113]  135.71151  112.84282  129.90436  123.35544  133.21954  123.75441
109.09031
##  [120]  138.50504  125.34044  139.64933  113.10149  108.75780   98.94744
123.39100
##  [127]  113.69072  152.50539  140.08905  116.86301  163.25786  134.04270
118.99060
##  [134]  111.08255  124.53535  130.22664  119.04215  129.19955  110.46922
123.79709
##  [141]  118.59721  115.20058  123.88088  106.76593  139.70947  155.59984
118.41153
##  [148]  152.15791  122.76417  133.30229  159.97859  114.79041  171.37018
```

```
144.86146
## [155] 139.57848 130.96396 132.28302 126.73297 126.00136 132.46054
122.11571
## [162] 129.09666 127.01601 143.38914 143.38914 154.85704 119.39850
140.03902
## [169] 129.64539 142.63595 129.14924 112.60297 119.36364 135.17605
130.66190
## [176] 111.07982 118.87833 136.72885 121.06551 145.67280 125.82164
148.21378
## [183] 127.89407 114.32759 148.52454 110.52872 129.44213 120.02754
141.87772
## [190] 114.41092 126.86604 128.12806 119.97500 142.89152 143.47153
165.57768
## [197] 150.64608 125.96492 115.18148 110.90043 114.80375 124.00808
145.86713
## [204] 178.75094 163.57777 128.72995 110.02143 150.20135 122.11925
144.37219
## [211] 129.05815 118.86175 118.74196 142.03261 124.65579 126.72416
125.15896
## [218] 124.91425 133.00687 129.78846 143.07005 138.41971 138.53093
196.65108
## [225] 108.73074 146.50078 114.85817 120.89561 168.86785 120.82340
108.64725
## [232] 160.99909 157.71988 115.91427 123.41375 139.62703 124.47944
139.66983
## [239] 123.56373 125.07876 155.98622 173.33133 145.30911 145.80639
131.46185
## [246] 124.28099 126.69849 157.67500 193.92334 139.54395 135.55790
131.63808
## [253] 132.42669  95.89994 156.83045 139.03717 126.05779 144.36458
150.58171
## [260] 123.64307 180.10300 113.87711 117.14883 108.84695 108.76598
112.43327
```

```
## [267] 173.60203 120.86237 117.69307 124.71999 119.66453 134.41953
114.27137
## [274] 101.42207 106.39352 128.32043 143.82176 122.84354 133.00844
139.91569
## [281] 130.99331 130.69423 126.28912 155.73050 141.04761 134.76411
137.17441
## [288] 155.92963 105.66951 128.54066 151.04930 110.21409 170.74749
153.73794
## [295] 123.39288 142.48880 125.39549 162.46217 138.22824 108.05663
145.94546
## [302] 132.73302 132.67016
```

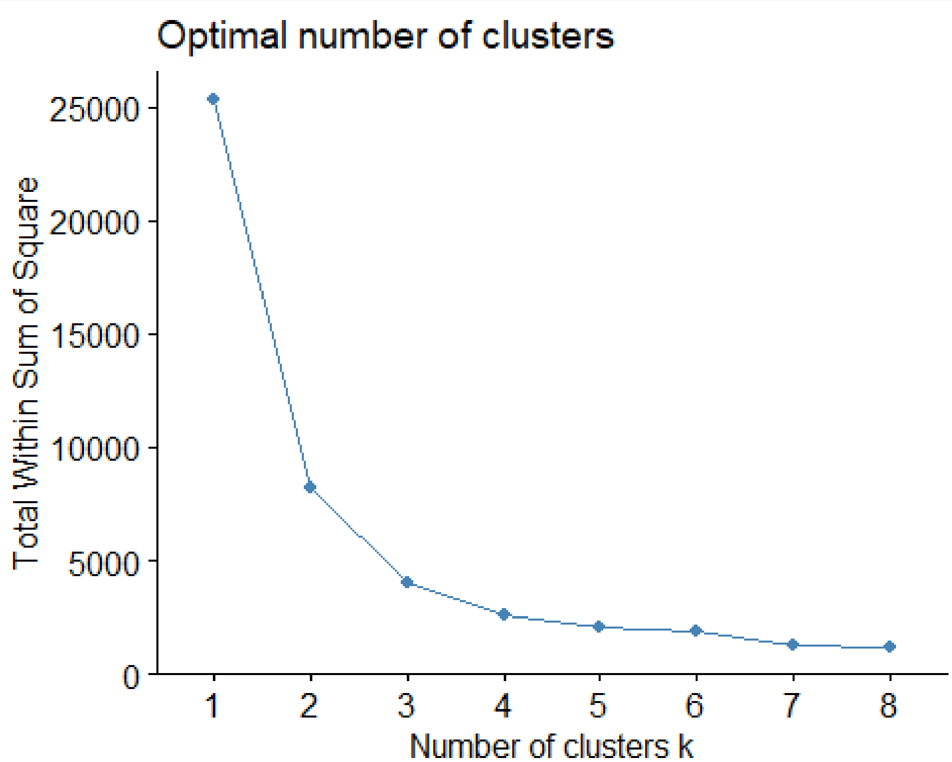
In this last principal component, we observe that the variable “trtbps”, which measures the pressure arterial At rest, it is the one that has the most positive weight. In contrast, the cholesterol is variable heavier on the negatives.

Cluster Analysis

In this section we are going to focus on cluster analysis, whose objective is to group the observations in our database into homogeneous groups that have common characteristics between them.

We are going to begin by grouping quantitative variables that do not measure a pain scale such as the "cp" variable.

```
## Warning in age$oldpeak = subset(heart_heart_csv, select = c("age",
"oldpeak")):
## Performing LHD coercion on a list

## Warning: package 'factoextra' was built under R version 4.1.3

## Loading required package: ggplot2

## Welcome! Want to learn more? See two factoextra-related books at
https://goo.gl/ve3WBa
```



When analyzing the graph generated by this function, it is crucial to identify the point at which an “elbow” is formed, since from that point the reduction in variability decreases. In this scenario, we can conclude that 3 groups could be an appropriate number to represent the underlying structure of this data set. After determining the appropriate number of groups, we implement the algorithm to group the data into each of those groups using R's kmeans() function. This function performs a random assignment of the initial centers, and the number of random assignments is specified by the nstart() parameter of the function

```
## K-means clustering with 3 clusters of sizes 130, 90, 83
##
## Cluster means:
##        age  oldpeak
## 1 54.80000 1.050769
## 2 64.67778 1.392222
## 3 42.50602 0.639759
##
## Clustering vector:
##   1   2   3   4   5   6   7   8   9  10  11  12  13  14  15  16  17  18 
19  20 
##   2   3   3   1   1   1   1   3   1   1   1   3   1   2   1   1   1   2 
3   2 
##  21  22  23  24  25  26  27  28  29  30  31  32  33  34  35  36  37  38 
39  40 
##   1   3   3   2   3   2   1   1   2   1   3   2   3   1   1   3   1   1 
2   2 
##  41  42  43  44  45  46  47  48  49  50  51  52  53  54  55  56  57  58 
59  60 
##   1   3   3   1   3   1   3   3   1   1   1   2   2   3   2   1   3   3 
3   1 
##  61  62  63  64  65  66  67  68  69  70  71  72  73  74  75  76  77  78 
79  80 
##   2   1   1   3   1   3   1   3   3   2   1   1   3   1   3   1   1   1 
1   1
```

```
##  81  82  83  84  85  86  87  88  89  90  91  92  93  94  95  96  97  98 
99 100 
##   3   3   2   1   3   2   2   3   1   1   3   1   1   1   3   1   2   1 
3   1 
## 101 102 103 104 105 106 107 108 109 110 111 112 113 114 115 116 117 118 
119 120 
##   3   1   2   3   1   2   2   3   1   1   2   1   2   3   1   3   3   1 
3   3 
## 121 122 123 124 125 126 127 128 129 130 131 132 133 134 135 136 137 138 
139 140 
##   2   1   3   1   3   3   3   2   1   2   1   1   3   3   3   1   2   2 
1   2 
## 141 142 143 144 145 146 147 148 149 150 151 152 153 154 155 156 157 158 
159 160 
##   1   3   3   2   2   2   3   2   3   3   2   2   2   2   3   1   3   3 
1   1 
## 161 162 163 164 165 166 167 168 169 170 171 172 173 174 175 176 177 178 
179 180 
##   1   1   3   3   3   2   2   2   2   1   1   3   1   1   2   3   2   2 
3   1 
## 181 182 183 184 185 186 187 188 189 190 191 192 193 194 195 196 197 198 
199 200 
##   1   2   2   1   1   3   2   1   1   3   1   1   1   2   2   1   3   2 
2   2 
## 201 202 203 204 205 206 207 208 209 210 211 212 213 214 215 216 217 218 
219 220 
##   3   2   1   2   2   1   1   2   1   1   1   2   3   2   1   3   2   2 
2   3 
## 221 222 223 224 225 226 227 228 229 230 231 232 233 234 235 236 237 238 
239 240 
##   2   1   2   1   1   2   2   3   1   2   3   1   1   2   2   1   1   2 
2   3 
## 241 242 243 244 245 246 247 248 249 250 251 252 253 254 255 256 257 258
```

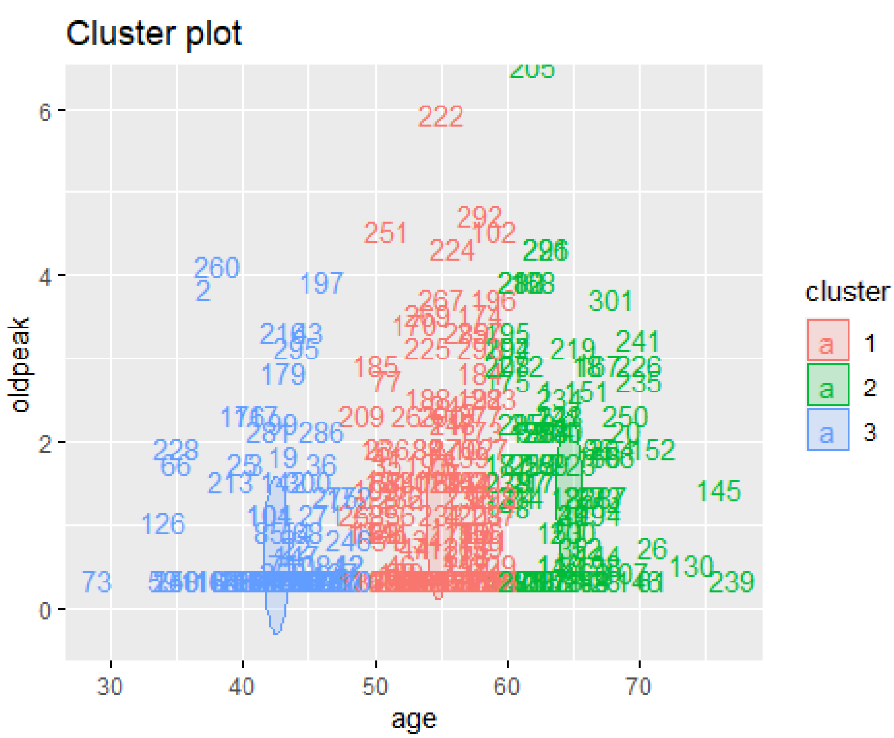
```
259 260
##   2   1   2   1   1   3   1   2   1   2   1   3   2   2   1   3   1   1
2   3
## 261 262 263 264 265 266 267 268 269 270 271 272 273 274 275 276 277 278
279 280
##   2   1   1   2   1   2   1   1   1   1   3   2   2   1   3   1   1   1
1   2
## 281 282 283 284 285 286 287 288 289 290 291 292 293 294 295 296 297 298
299 300
##   3   1   1   3   2   3   1   1   1   1   2   1   1   2   3   2   2   1
1   3
## 301 302 303
##   2   1   1
##
## Within cluster sum of squares by cluster:
## [1] 1353.325 1350.780 1319.746
##  (between_SS / total_SS =  84.1 %)
##
## Available components:
##
##  [1]  "cluster"          "centers"          "totss"              "withinss"
"tot.withinss"
## [6] "betweenss"    "size"         "iter"         "ifault"
```

After data clustering has been performed, it may be relevant, for future research, to add a column in the data set that identifies which cluster each experimental unit has been assigned to.

```
## Warning: package 'dplyr' was built under R version 4.1.3

##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
## filter, layers

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union

##   age oldpeak cluster
## 1  63     2.3       2
## 2  37     3.5       3
## 3  41     1.4       3
## 4  56     0.8       1
## 5  57     0.6       1
## 6  57     0.4       1
```

Next, we will plot the groups on a graph to visualize the assignment of each country to its respective group.



What we've done is grouped people by age and exercise-induced ST depression relative to rest. In the blue zone of the cluster plot are the younger people, in the red zone are the people with an intermediate age, between 50 and 60 years old. Finally, the green area represents the data of the elderly.

Comparison of Means of Two Populations

In this apart, we are going to study the relationship between a quantitative and a qualitative variable in order to analyze the database more deeply.

Are there significant differences between a person's sex and the likelihood of heart disease?

Our variable under study will be the probability of heart disease (output), while sex will be the factor with two levels of study: Female and Male. For this, It is necessary to check that the variable intended as a factor is of class "factor".

```
## [1] "character"
```

When we verify that class is not a factor, we must change it.

```
## [1] "factor"
```

Next, we will analyze the data with the summary command to study the sample size of the variable in each population:

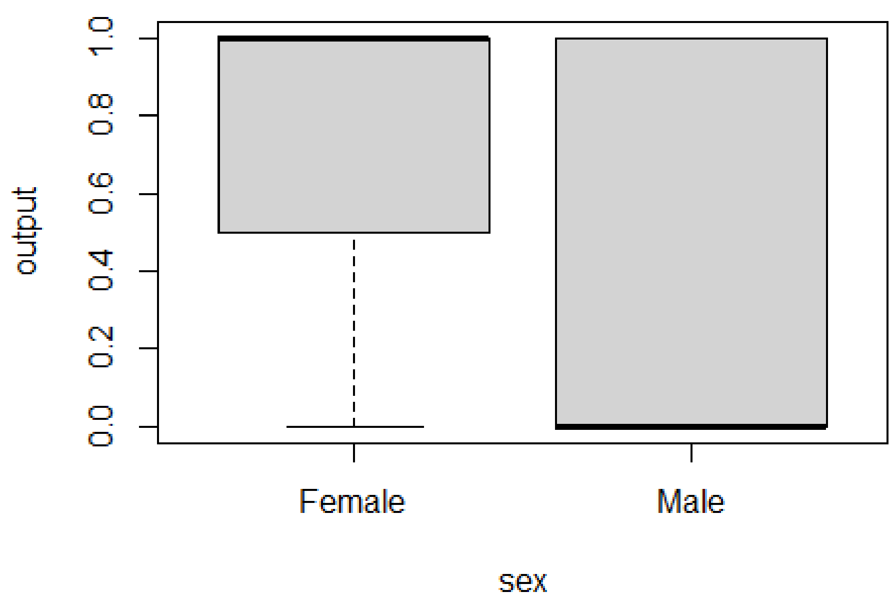
```
##      sex          output
##   Female: 96   Min.   :0.0000
## Male:207 1st Qu.:0.0000
##                Median :1.0000
##                Mean   :0.5446
## 3rd Q.:1.0000
##                Max.   :1.0000
```

As we can see, the sample size for men is 207, while that for women is 96. In both cases the sample size is very large, so it is not necessary to analyze normality.

Before addressing the resolution of the hypothesis tests, we can carry out a descriptive analysis examining the sample means of the variable under study for each level of the factor. Additionally, we can visualize this data using a tiered box plot.

```
## Warning: package 'car' was built under R version 4.1.3
```

```
## Loading required package: carData

## Warning: package 'carData' was built under R version 4.1.3

##
## Attaching package: 'car'

## The following object is masked from 'package:dplyr':
##
##     recode

##    Female      Male
## 0.7500000 0.4492754
```



In view of the box plot and the numerical values of the means, it seems that there is a notable difference in the probability of having heart disease between different sexes. To confirm this we must solve the t-Student test after analysis of equality of variances.

Homoscedasticity or Equality of Variances

To study whether there is equality of variances in both populations, we apply Levene's test with the intention of solving the following test:

$\{H_0\colon \sigma_1 = \sigma_2 \; H_1\colon \sigma_1 \neq \sigma_2$

```
## Levene's Test for Homogeneity of Variance (center = "mean")
##        Df F value   Pr(>F)
## group   1  56.409 6.74e-13 ***
##       301
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Once the Levene test has been carried out and after obtaining a p-value less than 2.2e-16 and less than 0.05, we reject H0 and, therefore, there is no equality of variances in both populations. Now we will perform the Student t test to resolve the equality of means test.

$\{H_0\colon \mu_1 = \mu_2 \; H_1\colon \mu_1 \neq \mu_2$

```
##
##  Welch Two Sample t-test
##
## data:  output by sex
## t = 5.3372, df = 209.95, p-value = 2.44e-07
## alternative hypothesis: true difference in means between group Female and
group Male is not equal to 0
## 95 percent confidence interval:
##  0.1896497 0.4117996
## sample estimates:
## mean in group Female   mean in group Male
##            0.7500000            0.4492754
```

After performing t-test We obtain a p-value less than 2.2e-16, being less than 0.05, so we reject H0 and also the equality of means, stating that there are significant differences between the probability of having heart disease in different sexes.

Conclusion

Once the study of the Heart Attack data is completed, we can say that older people have a greater risk of developing cardiovascular diseases. On the other hand, we have seen that there is a significant difference between men and women when it comes to suffering a heart attack. This may be because women may experience different heart attack symptoms than men, and these symptoms may be less recognized or misattributed to other health conditions. Women may have more subtle or atypical symptoms, such as fatigue, shortness of breath, nausea, or back pain, instead of typical chest pain. These differences in symptoms and risk perception can sometimes lead to underdiagnosis and, consequently, later treatment in women.

Finally, in the main components, especially in the second of them, it is observed that, in general terms, when blood pressure decreases, the body responds by increasing heart rate to counteract the reduction in blood flow.