Dimention Reduction. Analysis of Cardio taraining.

Intruducation

In this project we will analys, the information & dataset is a result of the medical examination, based on real patiant information.

Also, in this project we will use: PCA & MCA & CA

Dataset contains the following information:

|  |
| --- |
| Age | Objective Feature | age | int (days) |
| Height | Objective Feature | height | int (cm) | |
| Weight | Objective Feature | weight | float (kg) | |
| Gender | Objective Feature | gender | categorical code | |
| Systolic blood pressure | Examination Feature | ap\_hi | int | |
| Diastolic blood pressure | Examination Feature | ap\_lo | int | |
| Cholesterol | Examination Feature | cholesterol | 1: normal, 2: above normal, 3: well above normal | |
| Glucose | Examination Feature | gluc | 1: normal, 2: above normal, 3: well above normal | |
| Smoking | Subjective Feature | smoke | binary | |
| Alcohol intake | Subjective Feature | alco | binary | |
| Physical activity | Subjective Feature | active | binary | |
| Presence or absence of cardiovascular disease | Target Variable | cardio | binary | |

Data preparation

|  |  |
| --- | --- |
| install.packages("factoextra")  install.packages("gridExtra")  install.packages("tidyverse")  install.packages("ggplot2") | library(factoextra)  library(FactoMineR)  library(ggplot2)  library(dplyr)  library(reshape2)  library(corrplot)  library(gridExtra)  library(grid) |

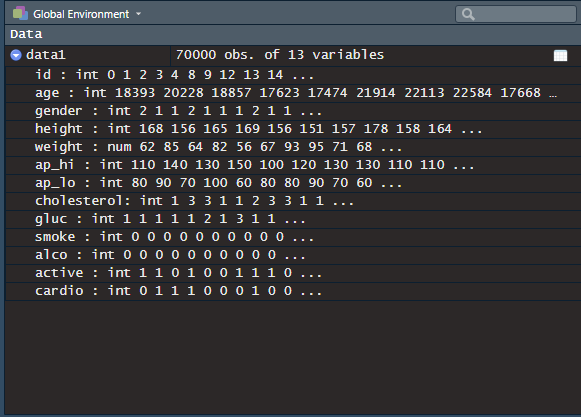
Loading data

setwd("D:\\R and R Studio\\Dimension Reduction\\DimensionReduction")

getwd()

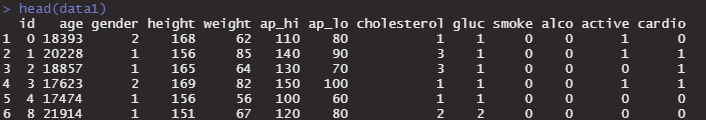
data1 <- read.csv("cardio\_train.csv", sep = ";")

View(data1)

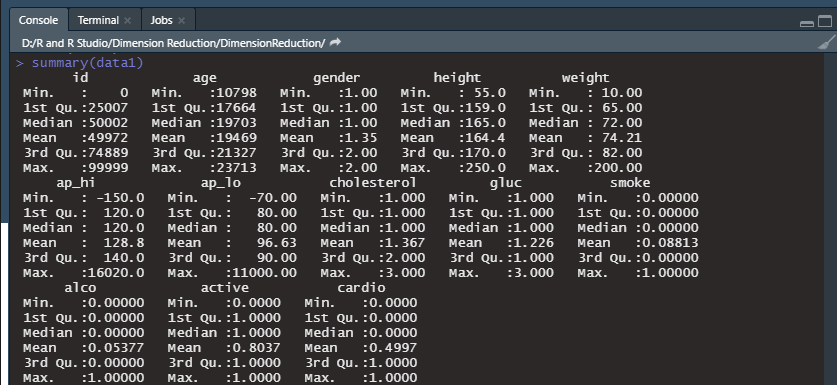


Data Analysis

head(data1)



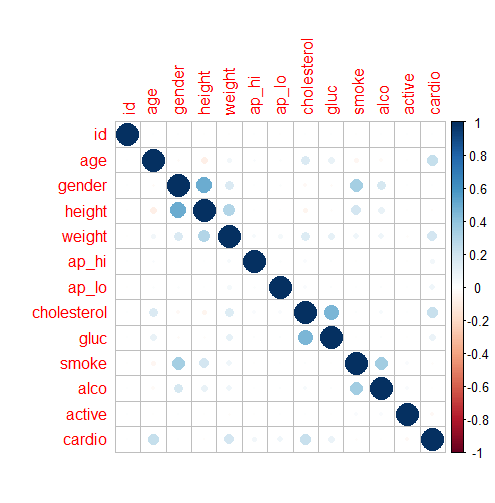
summary(data1)



Undersatnding data viusaly for it we need to show it in correlation.

corr\_data = cor(data1,method='pearson')

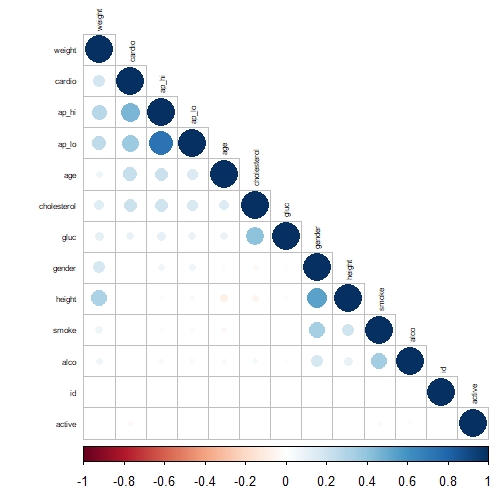
corrplot(corr\_data)



Here, correlation suing matrix Cholesterol.

cor.matrix <- cor(data1, method = c("spearman"))

corrplot(cor.matrix, type = "lower", order = "hclust", tl.col = "black", tl.cex = 0.5)



PCA

Choosing number in Component

> data.pca <- prcomp(data1, center=TRUE, scale=TRUE)

> eigen(cor(data1))$values

[1] 1.9430876 1.7384125 1.1616289 1.0544779 1.0021538 0.9977607 0.9841246 0.9731207 0.8270677 0.7235359

[11] 0.6215278 0.5243303 0.4487716

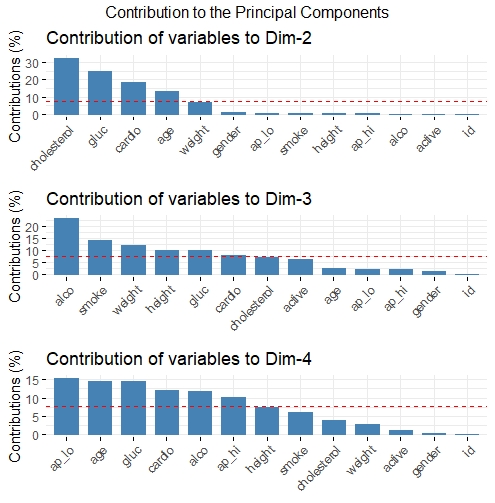
var <- get\_pca\_var(data.pca)

a<-fviz\_contrib(data.pca, "var",axes = 2)

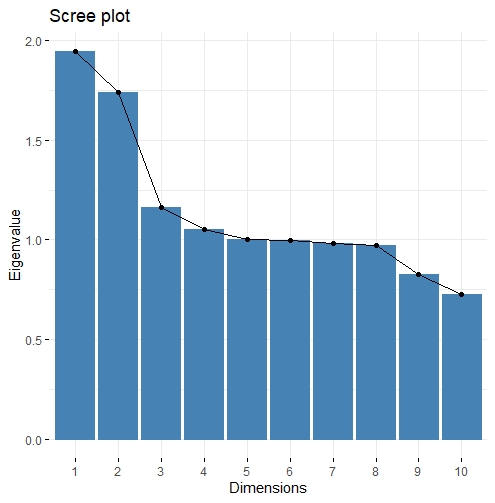
b<-fviz\_contrib(data.pca, "var",axes = 3)

c<-fviz\_contrib(data.pca, "var",axes = 4)

grid.arrange(a,b,c,top='Contribution to the Principal Components')

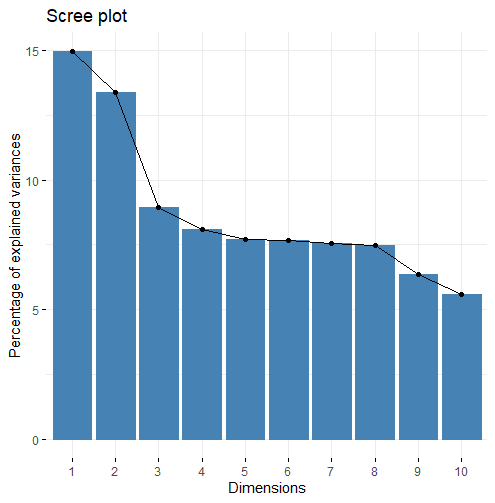


fviz\_eig(data.pca, choice='eigenvalue')

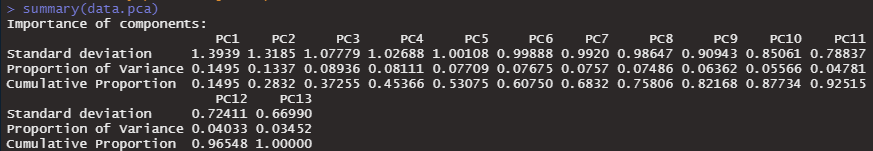


Two diffrent contrubutions to see dimensioanl space result.

fviz\_eig(data.pca)



summary(data.pca)

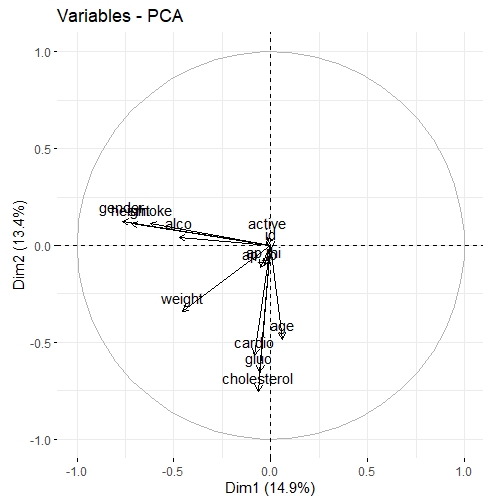


If we look at the plot of components and variance that they explain.In the analysis, the all component will be taken into consideration. (Mainly, 3 component)

Component Analysis

Cleare visoin of issues

fviz\_pca\_var(data.pca, col.ind = "Age")



Conclution

To conculude the project, analysing current data, it gives us more understanding about most case in cardiovascular disease.

References

### Source information; /kaggle/input/cardio\_train.csv

### Movite based on: [RPubs - Dimension Reduction for nominal data and qualitative data](https://rpubs.com/piterii/dimension_reduction)