ACP Glaucoma_DB - Tarea Voluntaria 3

Juan Manuel Rodríguez Gómez

2023-11-08

In this work, we are going to apply a dimensionality reduction (PCA) to the dataset Glaucoma_DB.xlsx using the R language. Some changes had to be made to the data in the records so that the analysis could be performed correctly. For example, in a discrete quantitative variable, it appeared in a register "4 scattered", so it has been replaced simply by a 4.

Loading packages and data set

Loading and installing R packages

The following source code module is responsible for loading, if they are already installed, all the packages that will be used in this R session. While an R package can be loaded at any time when it is to be used, it is advisable to optimize its calls with this code chunk at the beginning.

Loading a package into an R session **requires it to be already installed**. If it is not, the first step is to run the sentence:

install.packages("name_of_the_library")

```r
#########################################
# Loading necessary packages and reason #
#########################################

# This is an example of the first installation of a package
# Only runs once if the package is not installed
# Once it is installed this sentence has to be commented (not to run again)
# install.packages("summarytools")

# Package required to call 'freq' and 'descr' functions (descriptive statistics)
library(summarytools)

# Package required to call 'ggplot' function (graphical tools)
library(ggplot2)

# Package required to call 'ggarrange' function (graphical tools)
library(ggpubr)

# Package required to call 'read.spss' function (loading '.spss' data format)
library(haven)

# Package required to call 'read_xlsx' function (loading '.xlsx' data format)
```

```r
library(readxl)

# Package required to call 'cortest.bartlett' function
library(psych)

# Package required to call 'fviz_pca_var, fviz_pca_ind and fviz_pca' functions
library(factoextra)

# Package required to call 'scatterplot3d' function
library(scatterplot3d)
```

Description of the data set *Glaucoma_DB.xlsx*

The following code chunk shows how to load the dataset with Excel format (.xlsx).

Excel format *(xlsx)*

```r
# Loading a .xlsx (excel) file.
# The output of this function is already a data.frame object
# Remember that package 'readxl' is required
data_xlsx<-read_excel("Glaucoma_DB.xlsx", sheet = 2)

# This sentence identifies the type of object that the identifier represents
class(data_xlsx)
```

```
## [1] "tbl_df"     "tbl"        "data.frame"
```

The file *Glaucoma_DB.xlsx* contains, among others, the variables *OJO, TIPO_GLAUCOMA, N_IMPACTOS, CUADRANTES, ENERGIA_IMPACTO, ENERGIA_TOTAL, CIRUJIA_PREVIA, PIO_PRE_SLT, PIO_1_SEMANA, PIO_1_MES, PIO_3_MES, FARMACOS_PRE, FARMACOS_1_MES, FARMACOS_3_MES, DOLOR, SEXO, EDAD, PIO_NORMAL, PIO_NORMAL_CAT*, which are not standardized. Each record in the dataset collects data from a glaucoma operation performed on a patient. Below is a description of each variable in the dataset:

- Intervened eye (OJO)
- Type of glaucoma/disease (TIPO_GLAUCOMA)
- Number of impacts/Number of laser pulses (N_IMPACTOS)
- Number of quadrants made (CUADRANTES)
- Impact laser energy (ENERGIA_IMPACTO)
- Sum of total impact energy (ENERGIA_TOTAL)
- The patient has received any previous surgery (CIRUJIA_PREVIA)
- Intraocular pressure before laser (PIO_PRE_SLT)
- Intraocular pressure after a week of laser (PIO_1_SEMANA)
- Intraocular pressure after one month of laser (PIO_1_MES)
- Intraocular pressure after three months of laser (PIO_3_MES)
- Number of medications before the intervention (FARMACOS_PRE)
- Number of active medications after one month of laser (FARMACOS_1_MES)
- Number of active medications after three months of laser (FARMACOS_3_MES)
- The patient has experienced pain in the intervention (DOLOR)
- Patient sex (SEXO)
- Patient age (EDAD)
- Pressure after intervention below 21 (PIO_NORMAL)
- Positive/negative intervention result (PIO_NORMAL_CAT)

Basic descriptive statistics

In this section, a preliminary exploratory data analysis of the data set is performed. For this purpose, if the variable is **quantitative**, the basic **numerical descriptive statistics** and a representation of its **histogram, density and boxplot** are shown. On the other hand, for the **categorical** variables their **frequency table** and a **sector and bar diagram** are provided.

Exploring the data set

```
# This line loads the variable names from this data.frame
# So that we can access by their name with no refer to the data.frame identifier
attach(data_xlsx)

# Retrieving the name of all variables
colnames(data_xlsx)
```

```
##  [1] "OJO"             "TIPO_GLAUCOMA"   "N_IMPACTOS"      "CUADRANTES"
##  [5] "ENERGIA_IMPACTO" "ENERGIA_TOTAL"   "CIRUJIA_PREVIA"  "PIO_PRE_SLT"
##  [9] "PIO_1_SEMANA"    "PIO_1_MES"       "PIO_3_MES"       "FARMACOS_PRE"
## [13] "FARMACOS_1_MES"  "FARMACOS_3_MES"  "DOLOR"           "SEXO"
## [17] "EDAD"            "PIO_NORMAL"      "PIO_NORMAL_CAT"
```

```
# Displaying a few records
head(data_xlsx, n=10)
```

```
## # A tibble: 10 x 19
##      OJO TIPO_GLAUCOMA N_IMPACTOS CUADRANTES ENERGIA_IMPACTO ENERGIA_TOTAL
##    <dbl>         <dbl>      <dbl>      <dbl>           <dbl>         <dbl>
##  1     0             0        112          4             1.5           174
##  2     1            NA        108          4             1.2           128
##  3     0             1        123          4             1.1           133
##  4     1             2        131          4             1.5           191
##  5     0             2        156          4             1.2           182
##  6     1             1        125          4             1.4           170
##  7     0             1        178          4             1.4           249
##  8     0             3        164          4             1.9           301
##  9     0             1        109          4             1             109
## 10     1             1        116          4             2.2           238
## # i 13 more variables: CIRUJIA_PREVIA <dbl>, PIO_PRE_SLT <dbl>,
## #   PIO_1_SEMANA <dbl>, PIO_1_MES <dbl>, PIO_3_MES <dbl>, FARMACOS_PRE <dbl>,
## #   FARMACOS_1_MES <dbl>, FARMACOS_3_MES <dbl>, DOLOR <dbl>, SEXO <dbl>,
## #   EDAD <dbl>, PIO_NORMAL <dbl>, PIO_NORMAL_CAT <dbl>
```

```
# Displaying basic descriptives of variable 'PIO_3_MES'
summary(PIO_3_MES)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   0.000   0.000   5.347  14.000  43.000
```

```
# Displaying frequency table of variable 'TIPO_GLAUCOMA'
# Absolute
table(TIPO_GLAUCOMA)
```

```
## TIPO_GLAUCOMA
##  0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16
##  1 48  7  1  3  4 30  5  1  1  2  2  2  1  1  3  7
```

```r
# Relative
round(prop.table(table(TIPO_GLAUCOMA)),2)
```

```
## TIPO_GLAUCOMA
##    0    1    2    3    4    5    6    7    8    9   10   11   12   13   14   15
## 0.01 0.40 0.06 0.01 0.03 0.03 0.25 0.04 0.01 0.01 0.02 0.02 0.02 0.01 0.01 0.03
##   16
## 0.06
```

Descriptive analysis (numerical and graphical)

Basic descriptive statistics of quantitative variables

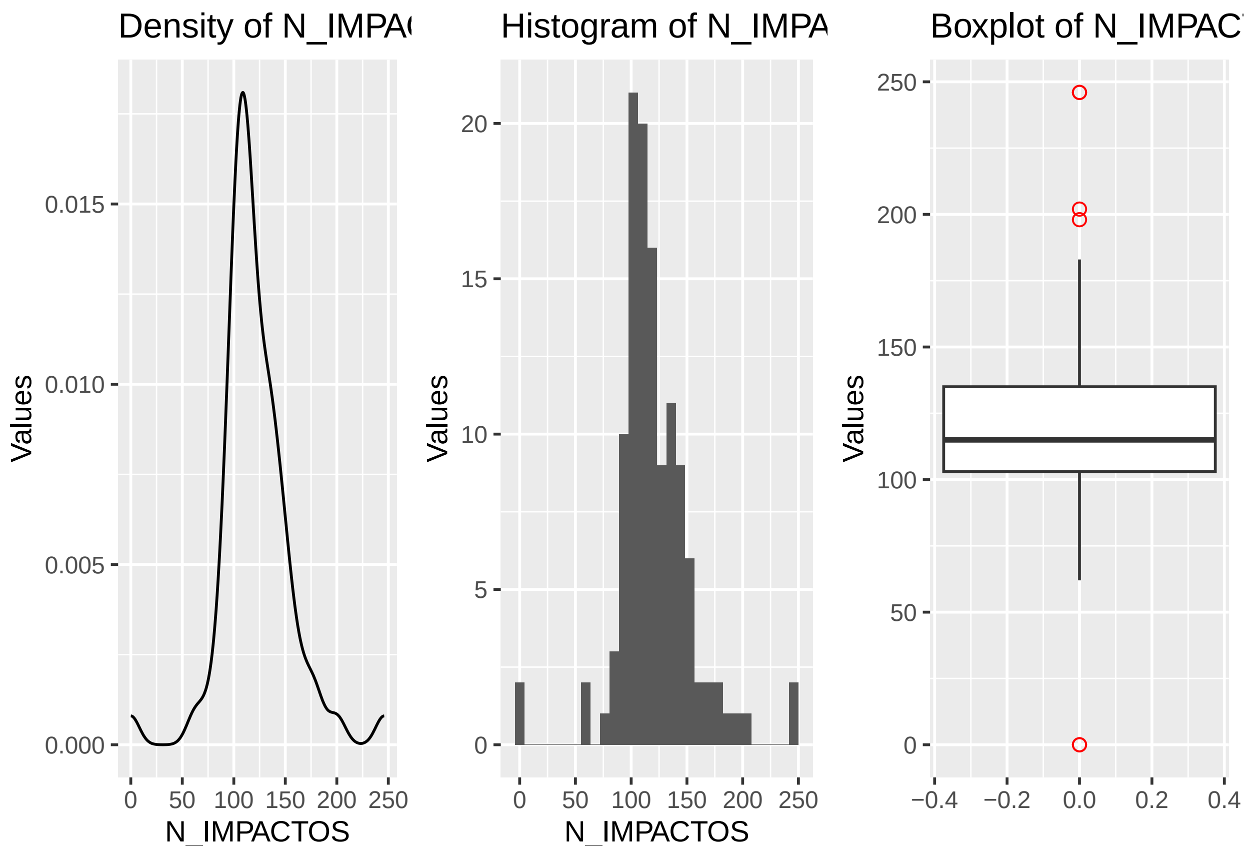
```r
# Basic descriptive statistics
# Remember that package 'summarytools' is required
descr(N_IMPACTOS)
```

N_IMPACTOS

```
## Descriptive Statistics
## N_IMPACTOS
## N: 121
##
##                      N_IMPACTOS
## ----------------- ------------
##              Mean       120.75
##           Std.Dev        33.78
##               Min         0.00
##                Q1       103.00
##            Median       115.00
##                Q3       135.00
##               Max       246.00
##               MAD        22.24
##               IQR        32.00
##                CV         0.28
##          Skewness         0.38
##       SE.Skewness         0.22
##          Kurtosis         4.02
##           N.Valid       121.00
##         Pct.Valid       100.00
```

```r
# Histogram, density and boxplot
# Remember that package 'ggplot2' is required
p1<-ggplot(data_xlsx,aes(x=N_IMPACTOS))+geom_density()+
  labs(title = "Density of N_IMPACTOS",x="N_IMPACTOS",y="Values")

p2<-ggplot(data_xlsx,aes(x=N_IMPACTOS))+geom_histogram()+
  labs(title = "Histogram of N_IMPACTOS",x="N_IMPACTOS",y="Values")

p3<-ggplot(data_xlsx,aes(x=N_IMPACTOS))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of N_IMPACTOS",x="Values",y="")

# This function controls the graphical output device
# Remember that package 'ggpubr' is required
```

```r
ggarrange(p1,p2,p3, nrow=1, common.legend = FALSE)
```



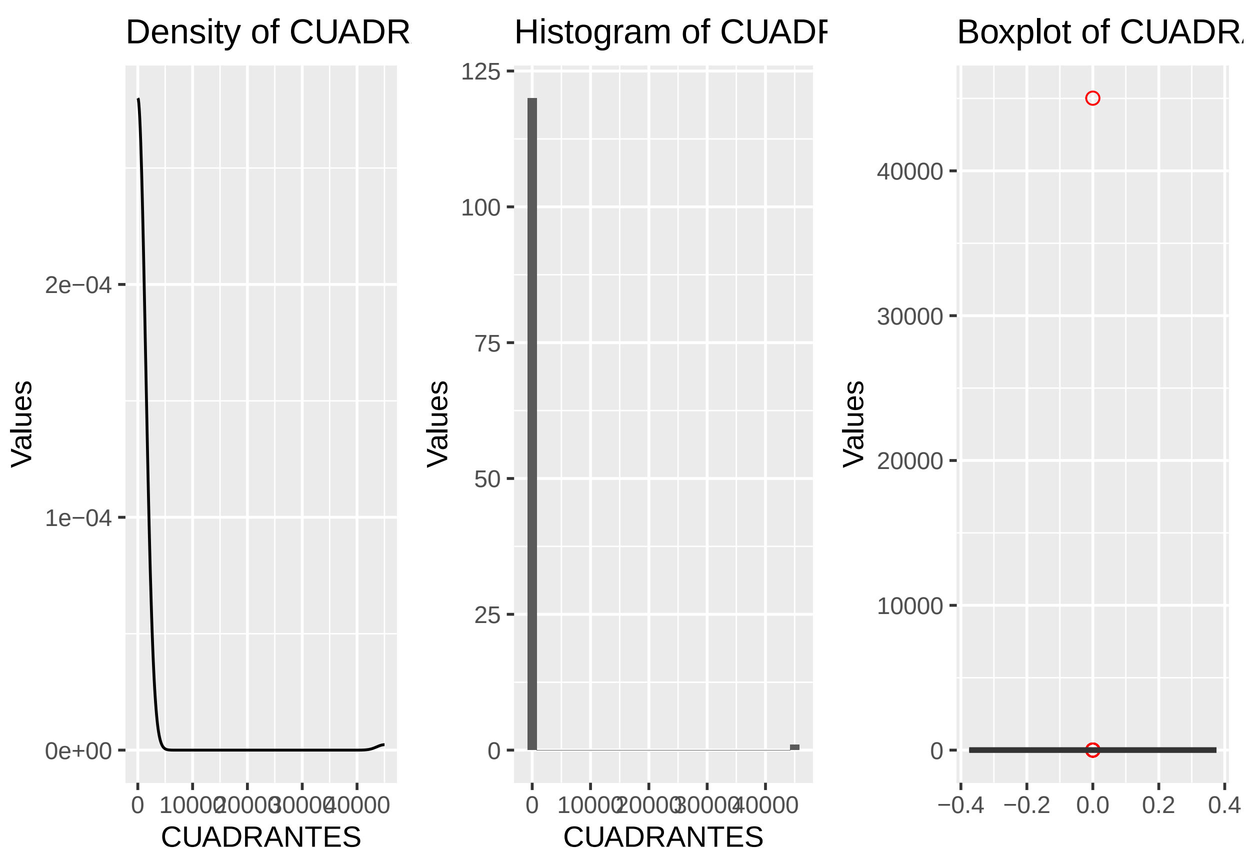

```r
# Basic descriptive statistics
# Remember that package 'summarytools' is required
descr(CUADRANTES)
```

CUADRANTES

```
## Descriptive Statistics
## CUADRANTES
## N: 121
##
##                      CUADRANTES
## ----------------- ------------
##              Mean       375.87
##           Std.Dev      4092.29
##               Min         0.00
##                Q1         4.00
##            Median         4.00
##                Q3         4.00
##               Max     45019.00
##               MAD         0.00
##               IQR         0.00
##                CV        10.89
##          Skewness        10.73
```

```
##      SE.Skewness        0.22
##         Kurtosis      114.05
##          N.Valid      121.00
##        Pct.Valid      100.00
```

```
# Histogram, density and boxplot
# Remember that package 'ggplot2' is required
p1<-ggplot(data_xlsx,aes(x=CUADRANTES))+geom_density()+
  labs(title = "Density of CUADRANTES",x="CUADRANTES",y="Values")

p2<-ggplot(data_xlsx,aes(x=CUADRANTES))+geom_histogram()+
  labs(title = "Histogram of CUADRANTES",x="CUADRANTES",y="Values")

p3<-ggplot(data_xlsx,aes(x=CUADRANTES))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of CUADRANTES",x="Values",y="")

# This function controls the graphical output device
# Remember that package 'ggpubr' is required
ggarrange(p1,p2,p3, nrow=1, common.legend = FALSE)
```

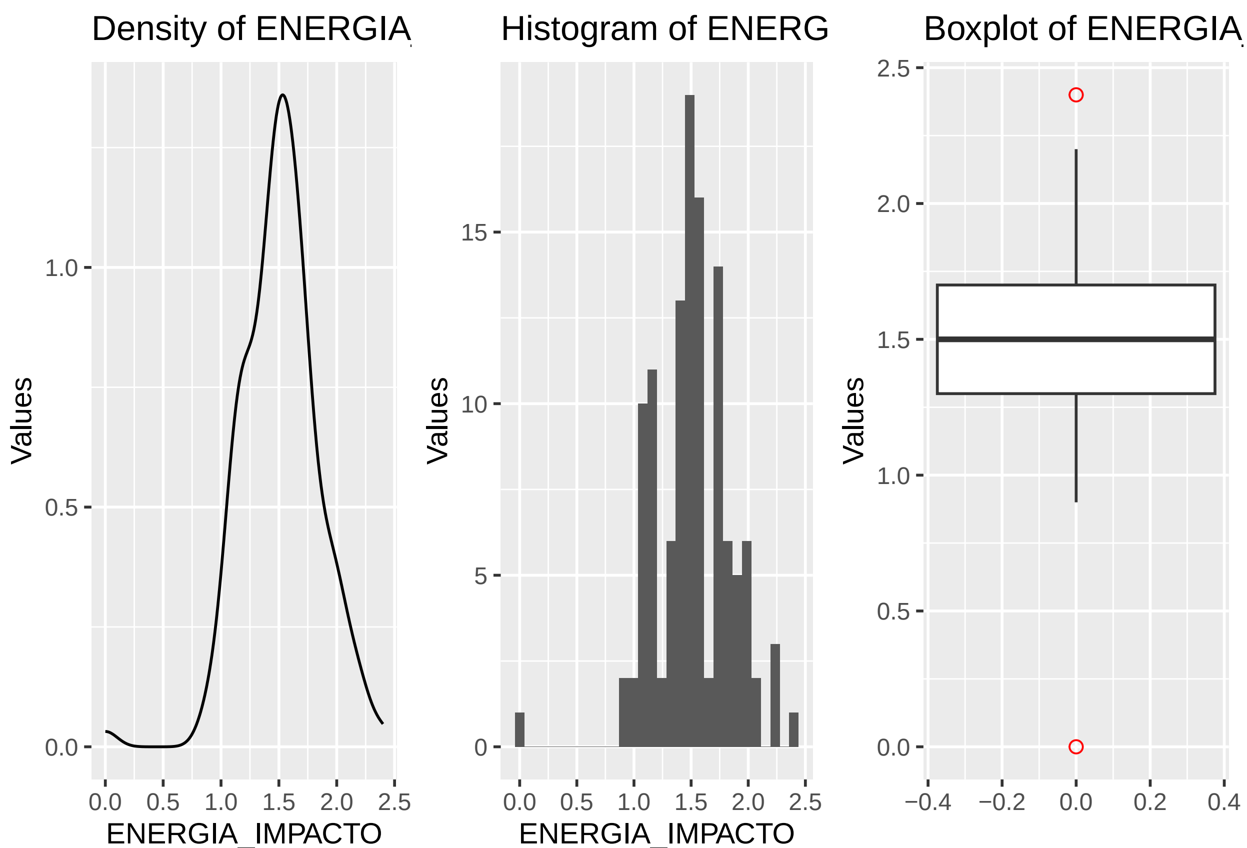



```
# Basic descriptive statistics
# Remember that package 'summarytools' is required
descr(ENERGIA_IMPACTO)
```

ENERGIA_IMPACTO

```
## Descriptive Statistics
## ENERGIA_IMPACTO
## N: 121
##
##                     ENERGIA_IMPACTO
## ----------------- -----------------
##              Mean              1.51
##           Std.Dev              0.33
##               Min              0.00
##                Q1              1.30
##            Median              1.50
##                Q3              1.70
##               Max              2.40
##               MAD              0.30
##               IQR              0.40
##                CV              0.22
##          Skewness             -0.46
##       SE.Skewness              0.22
##          Kurtosis              2.62
##           N.Valid            121.00
##         Pct.Valid            100.00
```

```r
# Histogram, density and boxplot
# Remember that package 'ggplot2' is required
p1<-ggplot(data_xlsx,aes(x=ENERGIA_IMPACTO))+geom_density()+
  labs(title = "Density of ENERGIA_IMPACTO",x="ENERGIA_IMPACTO",y="Values")

p2<-ggplot(data_xlsx,aes(x=ENERGIA_IMPACTO))+geom_histogram()+
  labs(title = "Histogram of ENERGIA_IMPACTO",x="ENERGIA_IMPACTO",y="Values")

p3<-ggplot(data_xlsx,aes(x=ENERGIA_IMPACTO))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of ENERGIA_IMPACTO",x="Values",y="")

# This function controls the graphical output device
# Remember that package 'ggpubr' is required
ggarrange(p1,p2,p3, nrow=1, common.legend = FALSE)
```



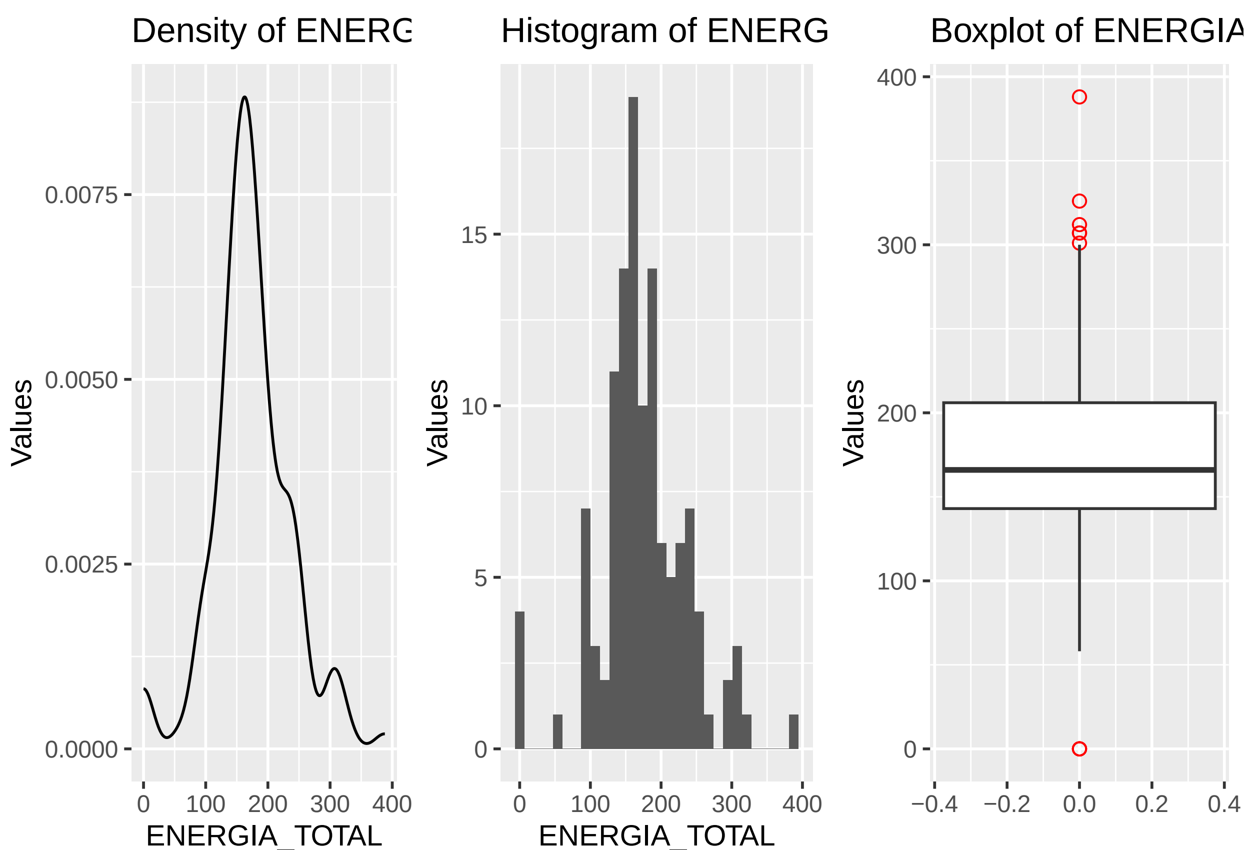

```r
# Basic descriptive statistics
# Remember that package 'summarytools' is required
descr(ENERGIA_TOTAL)
```

ENERGIA_TOTAL

```
## Descriptive Statistics
## ENERGIA_TOTAL
## N: 121
##
##                       ENERGIA_TOTAL
## ----------------- ---------------
##              Mean          173.33
##           Std.Dev           63.44
##               Min            0.00
##                Q1          143.00
##            Median          166.00
##                Q3          206.00
##               Max          388.00
##               MAD           45.96
##               IQR           63.00
##                CV            0.37
##          Skewness            0.10
##       SE.Skewness            0.22
##          Kurtosis            1.51
```

```
##          N.Valid          121.00
##        Pct.Valid          100.00
```

```
# Histogram, density and boxplot
# Remember that package 'ggplot2' is required
p1<-ggplot(data_xlsx,aes(x=ENERGIA_TOTAL))+geom_density()+
  labs(title = "Density of ENERGIA_TOTAL",x="ENERGIA_TOTAL",y="Values")

p2<-ggplot(data_xlsx,aes(x=ENERGIA_TOTAL))+geom_histogram()+
  labs(title = "Histogram of ENERGIA_TOTAL",x="ENERGIA_TOTAL",y="Values")

p3<-ggplot(data_xlsx,aes(x=ENERGIA_TOTAL))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of ENERGIA_TOTAL",x="Values",y="")

# This function controls the graphical output device
# Remember that package 'ggpubr' is required
ggarrange(p1,p2,p3, nrow=1, common.legend = FALSE)
```



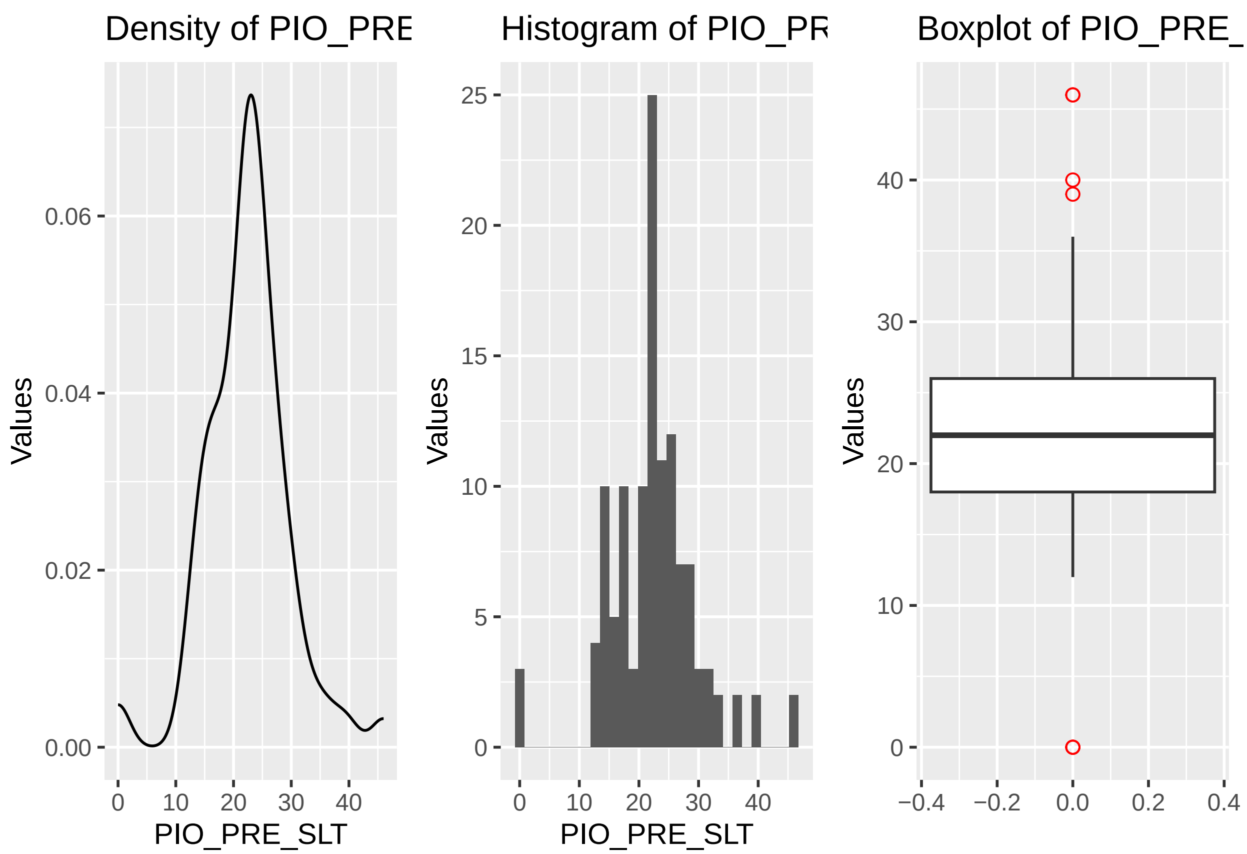

```
# Basic descriptive statistics
# Remember that package 'summarytools' is required
descr(PIO_PRE_SLT)
```

PIO_PRE_SLT

```
## Descriptive Statistics
```

```
## PIO_PRE_SLT
## N: 121
##
##                     PIO_PRE_SLT
## ----------------- -------------
##              Mean         22.45
##           Std.Dev          7.30
##               Min          0.00
##                Q1         18.00
##            Median         22.00
##                Q3         26.00
##               Max         46.00
##               MAD          5.93
##               IQR          8.00
##                CV          0.33
##          Skewness          0.07
##       SE.Skewness          0.22
##          Kurtosis          2.21
##           N.Valid        121.00
##         Pct.Valid        100.00
```

```r
# Histogram, density and boxplot
# Remember that package 'ggplot2' is required
p1<-ggplot(data_xlsx,aes(x=PIO_PRE_SLT))+geom_density()+
  labs(title = "Density of PIO_PRE_SLT",x="PIO_PRE_SLT",y="Values")

p2<-ggplot(data_xlsx,aes(x=PIO_PRE_SLT))+geom_histogram()+
  labs(title = "Histogram of PIO_PRE_SLT",x="PIO_PRE_SLT",y="Values")

p3<-ggplot(data_xlsx,aes(x=PIO_PRE_SLT))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of PIO_PRE_SLT",x="Values",y="")

# This function controls the graphical output device
# Remember that package 'ggpubr' is required
ggarrange(p1,p2,p3, nrow=1, common.legend = FALSE)
```



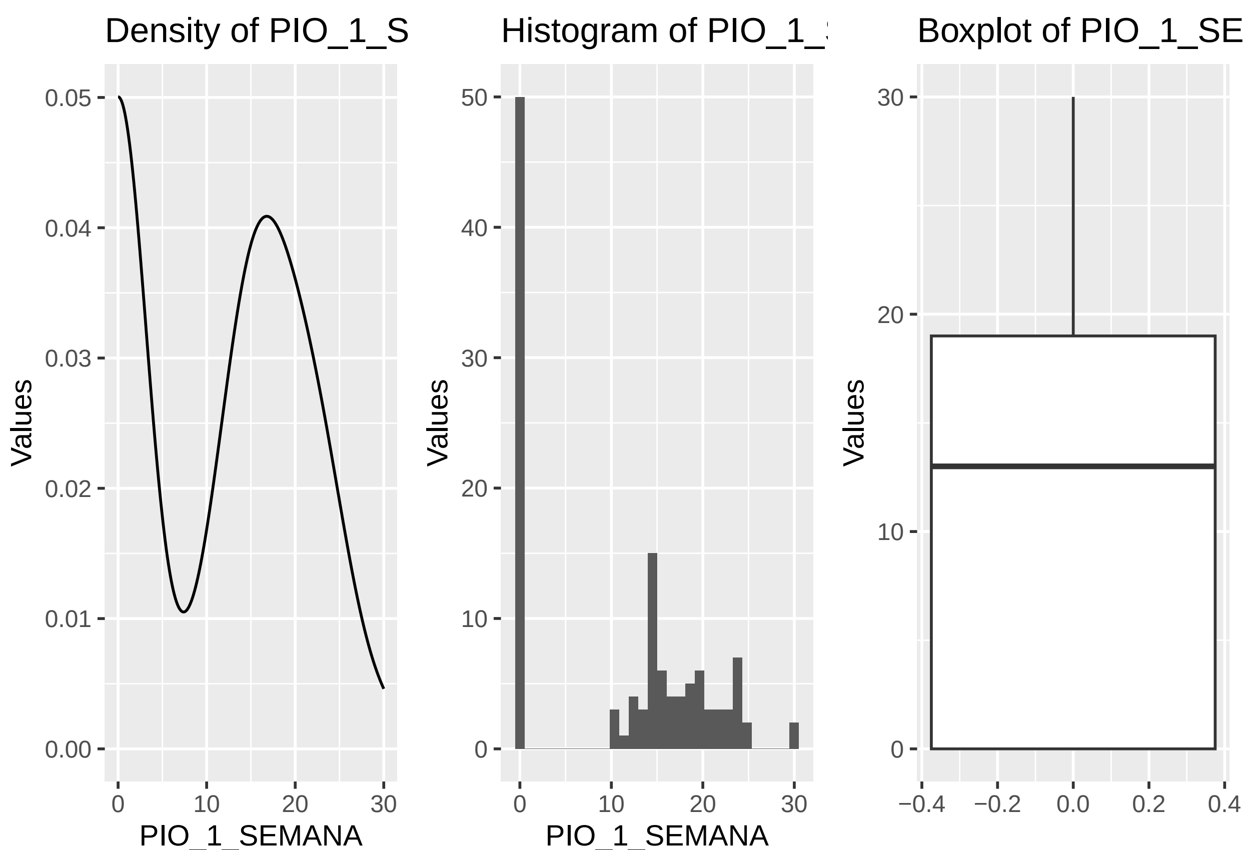

```r
# Basic descriptive statistics
# Remember that package 'summarytools' is required
descr(PIO_1_SEMANA)
```

PIO_1_SEMANA

```
## Descriptive Statistics
## PIO_1_SEMANA
## N: 121
##
##                      PIO_1_SEMANA
## ----------------- --------------
##              Mean          10.56
##           Std.Dev           9.56
##               Min           0.00
##                Q1           0.00
##            Median          13.00
##                Q3          19.00
##               Max          30.00
##               MAD          16.31
##               IQR          19.00
##                CV           0.90
##          Skewness           0.05
##       SE.Skewness           0.22
##          Kurtosis          -1.53
```

```
##           N.Valid         121.00
##         Pct.Valid         100.00
```

```
# Histogram, density and boxplot
# Remember that package 'ggplot2' is required
p1<-ggplot(data_xlsx,aes(x=PIO_1_SEMANA))+geom_density()+
  labs(title = "Density of PIO_1_SEMANA",x="PIO_1_SEMANA",y="Values")

p2<-ggplot(data_xlsx,aes(x=PIO_1_SEMANA))+geom_histogram()+
  labs(title = "Histogram of PIO_1_SEMANA",x="PIO_1_SEMANA",y="Values")

p3<-ggplot(data_xlsx,aes(x=PIO_1_SEMANA))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of PIO_1_SEMANA",x="Values",y="")

# This function controls the graphical output device
# Remember that package 'ggpubr' is required
ggarrange(p1,p2,p3, nrow=1, common.legend = FALSE)
```



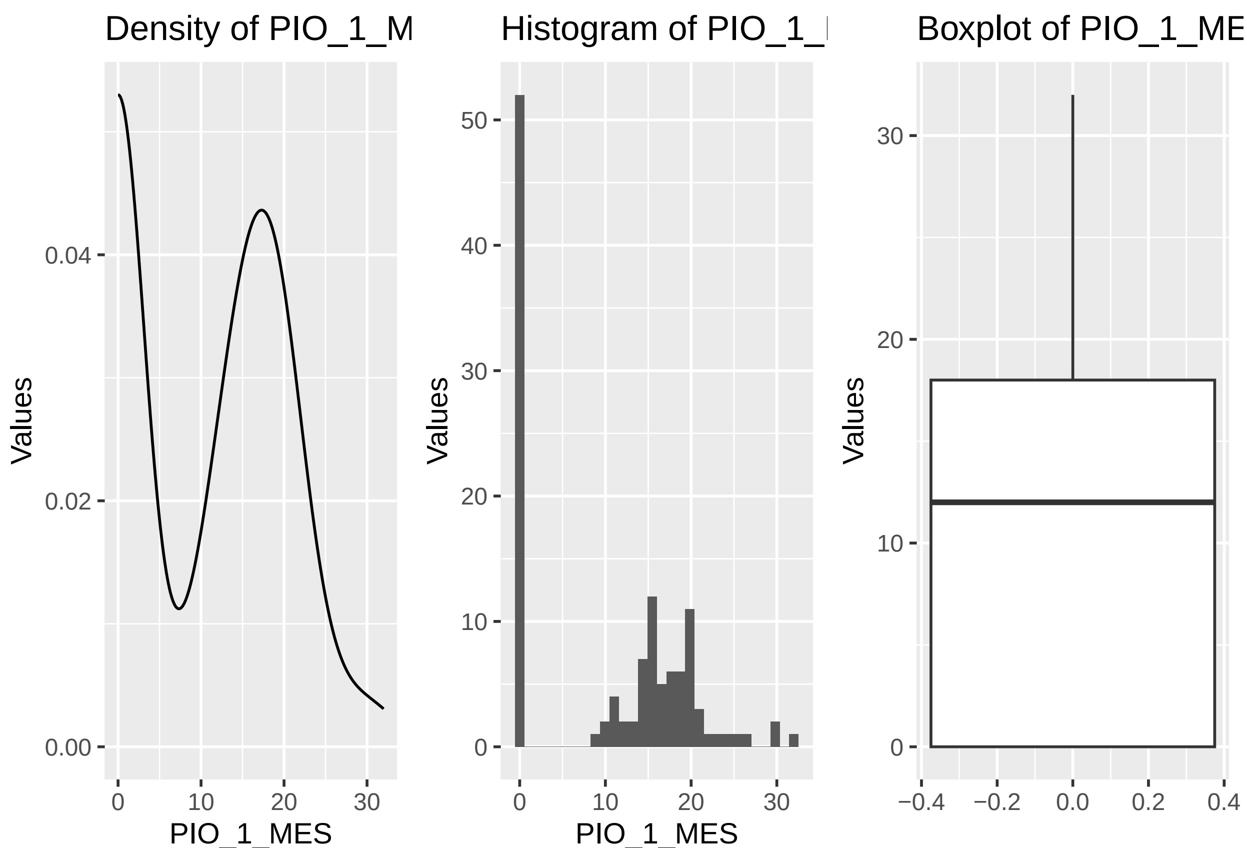

```
# Basic descriptive statistics
# Remember that package 'summarytools' is required
descr(PIO_1_MES)
```

PIO_1_MES

```
## Descriptive Statistics
```

```
## PIO_1_MES
## N: 121
##
##                       PIO_1_MES
## ----------------- -----------
##              Mean       10.00
##           Std.Dev        9.39
##               Min        0.00
##                Q1        0.00
##            Median       12.00
##                Q3       18.00
##               Max       32.00
##               MAD       13.34
##               IQR       18.00
##                CV        0.94
##          Skewness        0.16
##       SE.Skewness        0.22
##          Kurtosis       -1.38
##           N.Valid      121.00
##         Pct.Valid      100.00
```

```r
# Histogram, density and boxplot
# Remember that package 'ggplot2' is required
p1<-ggplot(data_xlsx,aes(x=PIO_1_MES))+geom_density()+
  labs(title = "Density of PIO_1_MES",x="PIO_1_MES",y="Values")

p2<-ggplot(data_xlsx,aes(x=PIO_1_MES))+geom_histogram()+
  labs(title = "Histogram of PIO_1_MES",x="PIO_1_MES",y="Values")

p3<-ggplot(data_xlsx,aes(x=PIO_1_MES))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of PIO_1_MES",x="Values",y="")

# This function controls the graphical output device
# Remember that package 'ggpubr' is required
ggarrange(p1,p2,p3, nrow=1, common.legend = FALSE)
```



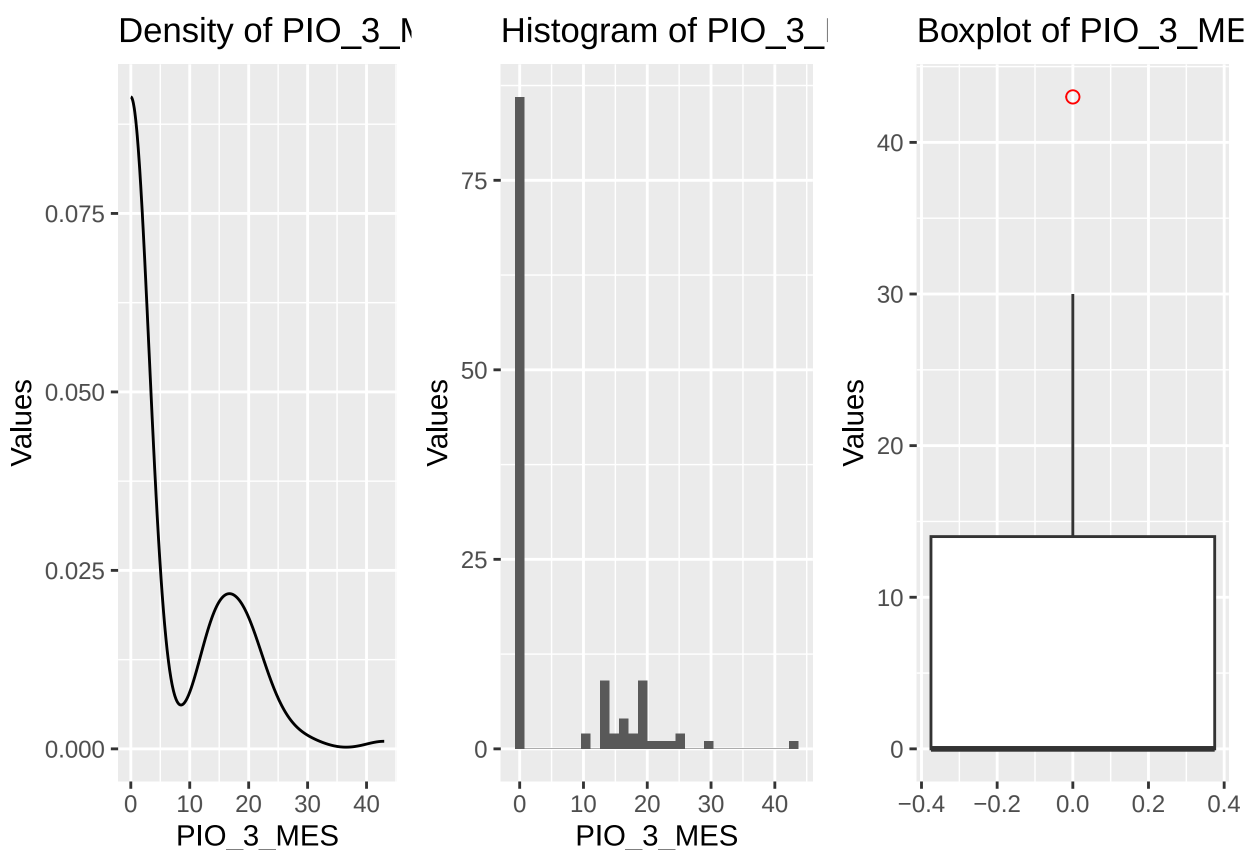

```
# Basic descriptive statistics
# Remember that package 'summarytools' is required
descr(PIO_3_MES)
```

PIO_3_MES

```
## Descriptive Statistics
## PIO_3_MES
## N: 121
##
##                       PIO_3_MES
## ----------------- -----------
##              Mean        5.35
##           Std.Dev        9.00
##               Min        0.00
##                Q1        0.00
##            Median        0.00
##                Q3       14.00
##               Max       43.00
##               MAD        0.00
##               IQR       14.00
##                CV        1.68
##          Skewness        1.47
##       SE.Skewness        0.22
##          Kurtosis        1.50
```

```
##           N.Valid      121.00
##         Pct.Valid      100.00
```

```r
# Histogram, density and boxplot
# Remember that package 'ggplot2' is required
p1<-ggplot(data_xlsx,aes(x=PIO_3_MES))+geom_density()+
  labs(title = "Density of PIO_3_MES",x="PIO_3_MES",y="Values")

p2<-ggplot(data_xlsx,aes(x=PIO_3_MES))+geom_histogram()+
  labs(title = "Histogram of PIO_3_MES",x="PIO_3_MES",y="Values")

p3<-ggplot(data_xlsx,aes(x=PIO_3_MES))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of PIO_3_MES",x="Values",y="")

# This function controls the graphical output device
# Remember that package 'ggpubr' is required
ggarrange(p1,p2,p3, nrow=1, common.legend = FALSE)
```



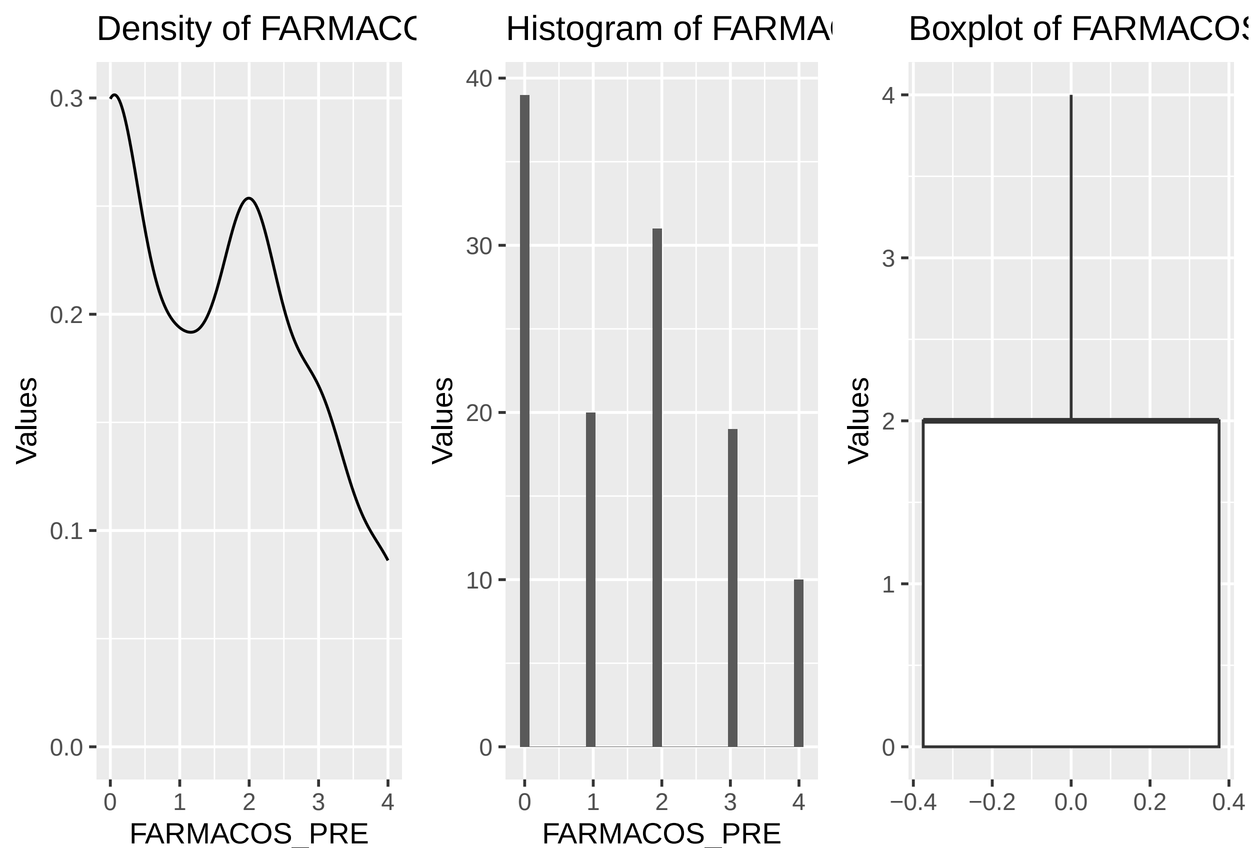

```r
# Basic descriptive statistics
# Remember that package 'summarytools' is required
descr(FARMACOS_PRE)
```

FARMACOS_PRE

```
## Descriptive Statistics
```

```
## FARMACOS_PRE
## N: 121
##
##                      FARMACOS_PRE
## ----------------- --------------
##              Mean           1.50
##           Std.Dev           1.32
##               Min           0.00
##                Q1           0.00
##            Median           2.00
##                Q3           2.00
##               Max           4.00
##               MAD           1.48
##               IQR           2.00
##                CV           0.88
##          Skewness           0.32
##       SE.Skewness           0.22
##          Kurtosis          -1.10
##           N.Valid         119.00
##         Pct.Valid          98.35
```

```r
# Histogram, density and boxplot
# Remember that package 'ggplot2' is required
p1<-ggplot(data_xlsx,aes(x=FARMACOS_PRE))+geom_density()+
  labs(title = "Density of FARMACOS_PRE",x="FARMACOS_PRE",y="Values")

p2<-ggplot(data_xlsx,aes(x=FARMACOS_PRE))+geom_histogram()+
  labs(title = "Histogram of FARMACOS_PRE",x="FARMACOS_PRE",y="Values")

p3<-ggplot(data_xlsx,aes(x=FARMACOS_PRE))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of FARMACOS_PRE",x="Values",y="")

# This function controls the graphical output device
# Remember that package 'ggpubr' is required
ggarrange(p1,p2,p3, nrow=1, common.legend = FALSE)
```



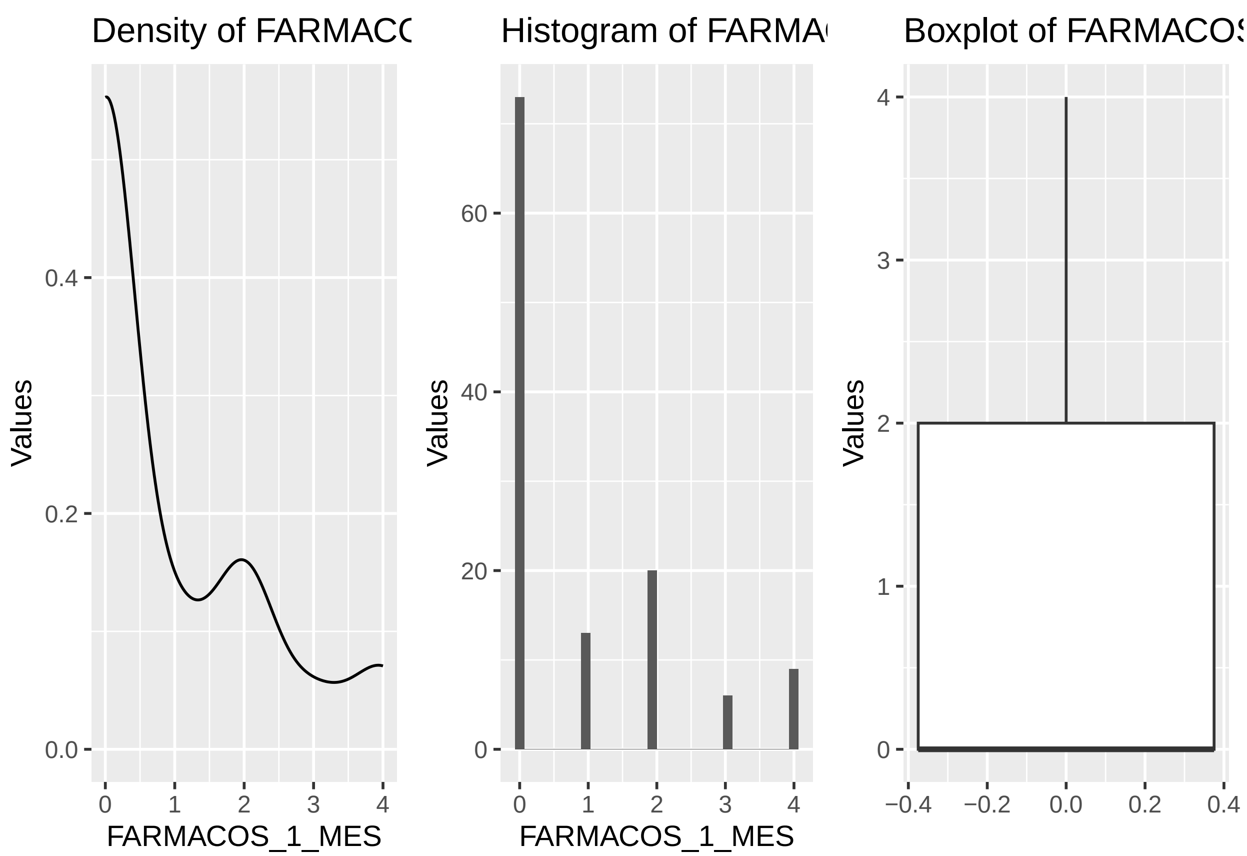

```
# Basic descriptive statistics
# Remember that package 'summarytools' is required
descr(FARMACOS_1_MES)
```

FARMACOS_1_MES

```
## Descriptive Statistics
## FARMACOS_1_MES
## N: 121
##
##                      FARMACOS_1_MES
## ----------------- ----------------
##              Mean             0.88
##           Std.Dev             1.28
##               Min             0.00
##                Q1             0.00
##            Median             0.00
##                Q3             2.00
##               Max             4.00
##               MAD             0.00
##               IQR             2.00
##                CV             1.45
##          Skewness             1.21
##       SE.Skewness             0.22
##          Kurtosis             0.22
```

```
##          N.Valid          121.00
##        Pct.Valid          100.00
```

```
# Histogram, density and boxplot
# Remember that package 'ggplot2' is required
p1<-ggplot(data_xlsx,aes(x=FARMACOS_1_MES))+geom_density()+
  labs(title = "Density of FARMACOS_1_MES",x="FARMACOS_1_MES",y="Values")

p2<-ggplot(data_xlsx,aes(x=FARMACOS_1_MES))+geom_histogram()+
  labs(title = "Histogram of FARMACOS_1_MES",x="FARMACOS_1_MES",y="Values")

p3<-ggplot(data_xlsx,aes(x=FARMACOS_1_MES))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of FARMACOS_1_MES",x="Values",y="")

# This function controls the graphical output device
# Remember that package 'ggpubr' is required
ggarrange(p1,p2,p3, nrow=1, common.legend = FALSE)
```



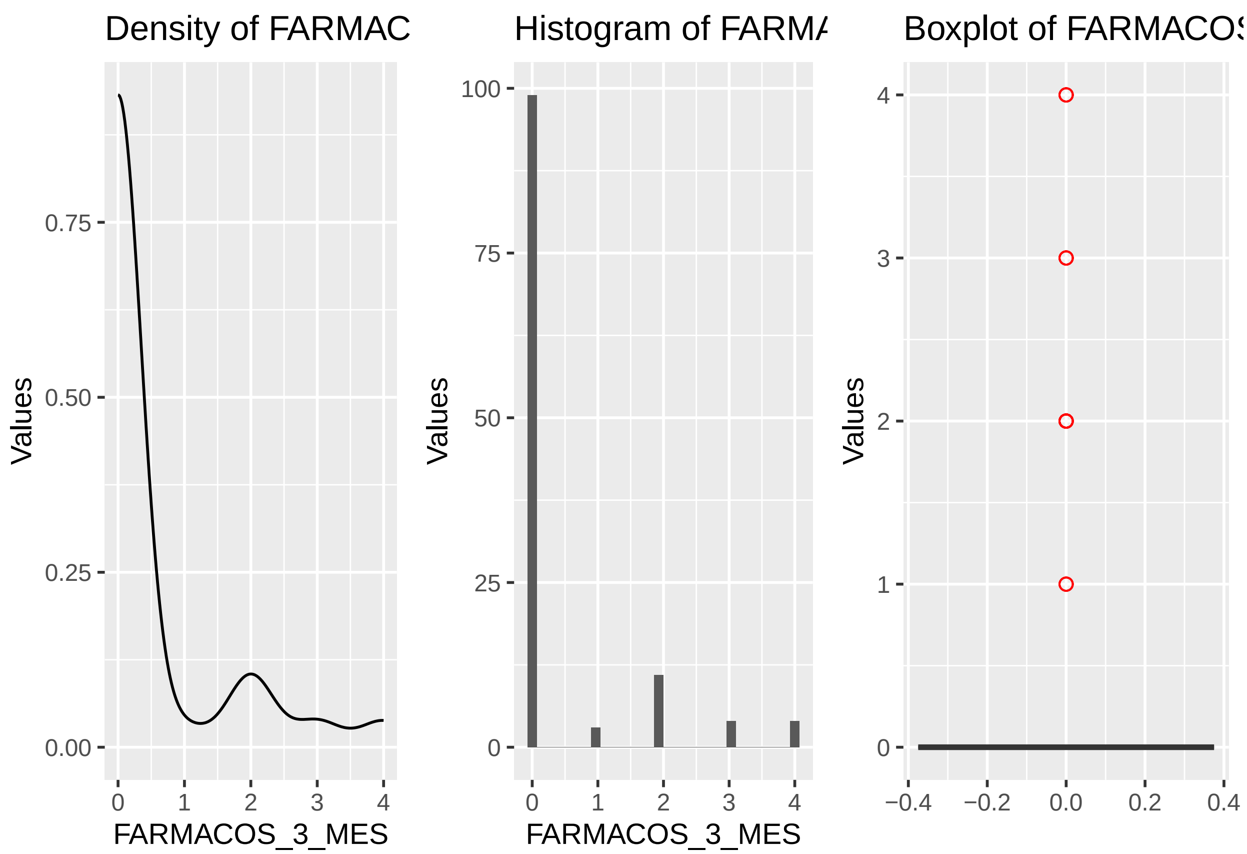

```
# Basic descriptive statistics
# Remember that package 'summarytools' is required
descr(FARMACOS_3_MES)
```

FARMACOS_3_MES

```
## Descriptive Statistics
```

```
## FARMACOS_3_MES
## N: 121
##
##                     FARMACOS_3_MES
## ----------------- ----------------
##              Mean             0.44
##           Std.Dev             1.02
##               Min             0.00
##                Q1             0.00
##            Median             0.00
##                Q3             0.00
##               Max             4.00
##               MAD             0.00
##               IQR             0.00
##                CV             2.32
##          Skewness             2.23
##       SE.Skewness             0.22
##          Kurtosis             3.88
##           N.Valid           121.00
##         Pct.Valid           100.00
```

```r
# Histogram, density and boxplot
# Remember that package 'ggplot2' is required
p1<-ggplot(data_xlsx,aes(x=FARMACOS_3_MES))+geom_density()+
  labs(title = "Density of FARMACOS_3_MES",x="FARMACOS_3_MES",y="Values")

p2<-ggplot(data_xlsx,aes(x=FARMACOS_3_MES))+geom_histogram()+
  labs(title = "Histogram of FARMACOS_3_MES",x="FARMACOS_3_MES",y="Values")

p3<-ggplot(data_xlsx,aes(x=FARMACOS_3_MES))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of FARMACOS_3_MES",x="Values",y="")

# This function controls the graphical output device
# Remember that package 'ggpubr' is required
ggarrange(p1,p2,p3, nrow=1, common.legend = FALSE)
```



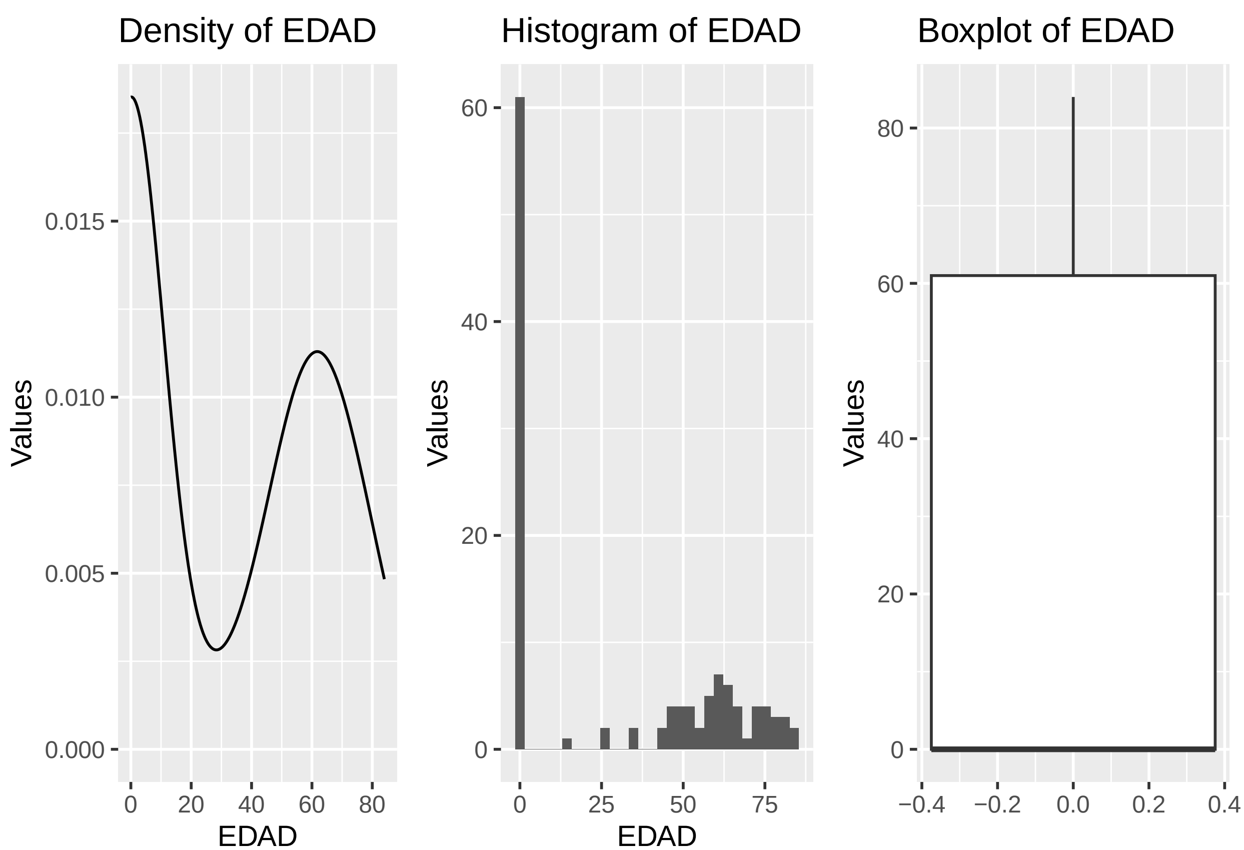

```r
# Basic descriptive statistics
# Remember that package 'summarytools' is required
descr(EDAD)
```

EDAD

```
## Descriptive Statistics
## EDAD
## N: 121
##
##                            EDAD
## ----------------- --------
##              Mean    29.65
##           Std.Dev    31.76
##               Min     0.00
##                Q1     0.00
##            Median     0.00
##                Q3    61.00
##               Max    84.00
##               MAD     0.00
##               IQR    61.00
##                CV     1.07
##          Skewness     0.28
##       SE.Skewness     0.22
##          Kurtosis    -1.69
```

```
##          N.Valid   121.00
##        Pct.Valid   100.00
```

```
# Histogram, density and boxplot
# Remember that package 'ggplot2' is required
p1<-ggplot(data_xlsx,aes(x=EDAD))+geom_density()+
  labs(title = "Density of EDAD",x="EDAD",y="Values")

p2<-ggplot(data_xlsx,aes(x=EDAD))+geom_histogram()+
  labs(title = "Histogram of EDAD",x="EDAD",y="Values")

p3<-ggplot(data_xlsx,aes(x=EDAD))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of EDAD",x="Values",y="")

# This function controls the graphical output device
# Remember that package 'ggpubr' is required
ggarrange(p1,p2,p3, nrow=1, common.legend = FALSE)
```



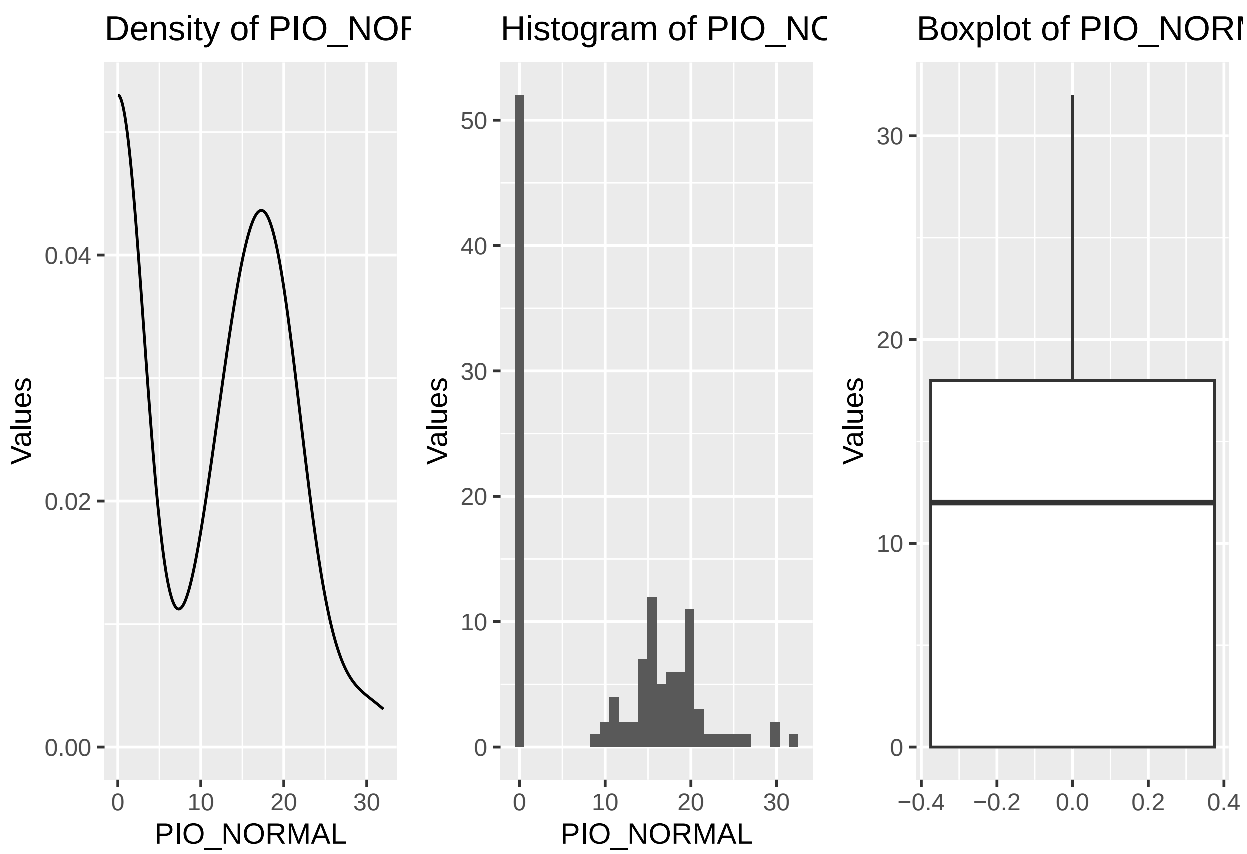

```
# Basic descriptive statistics
# Remember that package 'summarytools' is required
descr(PIO_NORMAL)
```

PIO_NORMAL

```
## Descriptive Statistics
```

```
## PIO_NORMAL
## N: 121
##
##                     PIO_NORMAL
## ----------------- ------------
##              Mean        10.00
##           Std.Dev         9.39
##               Min         0.00
##                Q1         0.00
##            Median        12.00
##                Q3        18.00
##               Max        32.00
##               MAD        13.34
##               IQR        18.00
##                CV         0.94
##          Skewness         0.16
##       SE.Skewness         0.22
##          Kurtosis        -1.38
##           N.Valid       121.00
##         Pct.Valid       100.00
```

```
# Histogram, density and boxplot
# Remember that package 'ggplot2' is required
p1<-ggplot(data_xlsx,aes(x=PIO_NORMAL))+geom_density()+
  labs(title = "Density of PIO_NORMAL",x="PIO_NORMAL",y="Values")

p2<-ggplot(data_xlsx,aes(x=PIO_NORMAL))+geom_histogram()+
  labs(title = "Histogram of PIO_NORMAL",x="PIO_NORMAL",y="Values")

p3<-ggplot(data_xlsx,aes(x=PIO_NORMAL))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of PIO_NORMAL",x="Values",y="")

# This function controls the graphical output device
# Remember that package 'ggpubr' is required
ggarrange(p1,p2,p3, nrow=1, common.legend = FALSE)
```



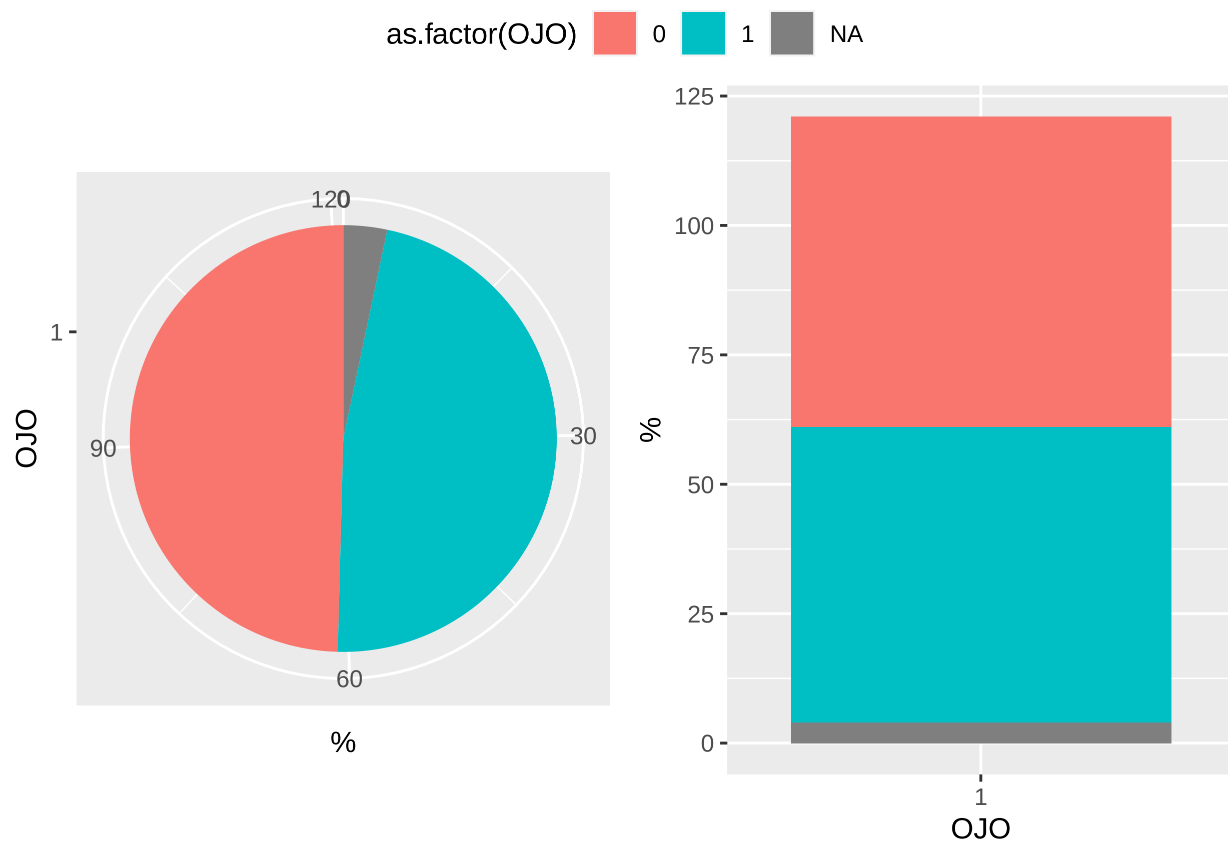
Basic descriptive statistics of categorical variables

```r
# Frequency tables. Descriptive analysis
# Remember that package 'summarytools' is required
freq(as.factor(OJO))
```

OJO

```
## Frequencies
##
##               Freq   % Valid   % Valid Cum.   % Total   % Total Cum.
## ----------- ------ --------- -------------- --------- --------------
##           0     60     51.28          51.28     49.59          49.59
##           1     57     48.72         100.00     47.11          96.69
##        <NA>      4                               3.31         100.00
##       Total    121    100.00         100.00    100.00         100.00
```

```r
# Pie chart and bar graph
p1<-ggplot(data_xlsx,aes(x=factor(1),fill=as.factor(OJO)))+geom_bar()+
  coord_polar("y")+labs(x="OJO",y="%")
p2<-ggplot(data_xlsx,aes(x=factor(1),fill=as.factor(OJO)))+geom_bar()+
  labs(x="OJO",y="%")

# This function controls the graphical output device
# Remember that package 'ggpubr' is required
ggarrange(p1,p2,nrow = 1,ncol=2, common.legend = TRUE)
```



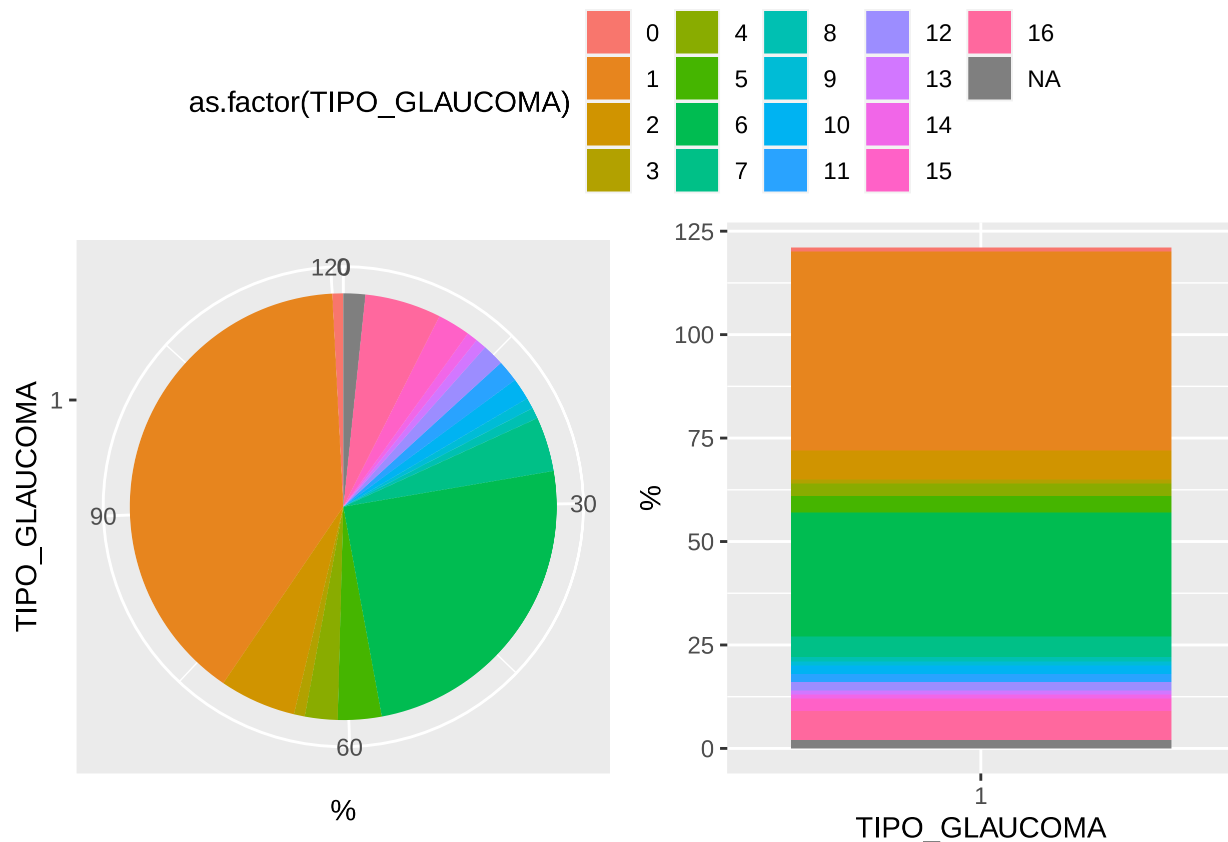

```r
# Frequency tables. Descriptive analysis
# Remember that package 'summarytools' is required
freq(as.factor(TIPO_GLAUCOMA))
```

TIPO_GLAUCOMA

```
## Frequencies
##
##                Freq   % Valid   % Valid Cum.   % Total   % Total Cum.
## ----------- ------ --------- -------------- --------- --------------
##           0      1      0.84           0.84      0.83           0.83
##           1     48     40.34          41.18     39.67          40.50
##           2      7      5.88          47.06      5.79          46.28
##           3      1      0.84          47.90      0.83          47.11
##           4      3      2.52          50.42      2.48          49.59
##           5      4      3.36          53.78      3.31          52.89
##           6     30     25.21          78.99     24.79          77.69
##           7      5      4.20          83.19      4.13          81.82
##           8      1      0.84          84.03      0.83          82.64
##           9      1      0.84          84.87      0.83          83.47
##          10      2      1.68          86.55      1.65          85.12
##          11      2      1.68          88.24      1.65          86.78
##          12      2      1.68          89.92      1.65          88.43
##          13      1      0.84          90.76      0.83          89.26
##          14      1      0.84          91.60      0.83          90.08
```

```
##          15      3      2.52          94.12      2.48           92.56
##          16      7      5.88         100.00      5.79           98.35
##        <NA>      2                              1.65          100.00
##       Total    121    100.00         100.00    100.00          100.00
```

```r
# Pie chart and bar graph
p1<-ggplot(data_xlsx,aes(x=factor(1),fill=as.factor(TIPO_GLAUCOMA)))+geom_bar()+
  coord_polar("y")+labs(x="TIPO_GLAUCOMA",y="%")
p2<-ggplot(data_xlsx,aes(x=factor(1),fill=as.factor(TIPO_GLAUCOMA)))+geom_bar()+
  labs(x="TIPO_GLAUCOMA",y="%")

# This function controls the graphical output device
# Remember that package 'ggpubr' is required
ggarrange(p1,p2,nrow = 1,ncol=2, common.legend = TRUE)
```



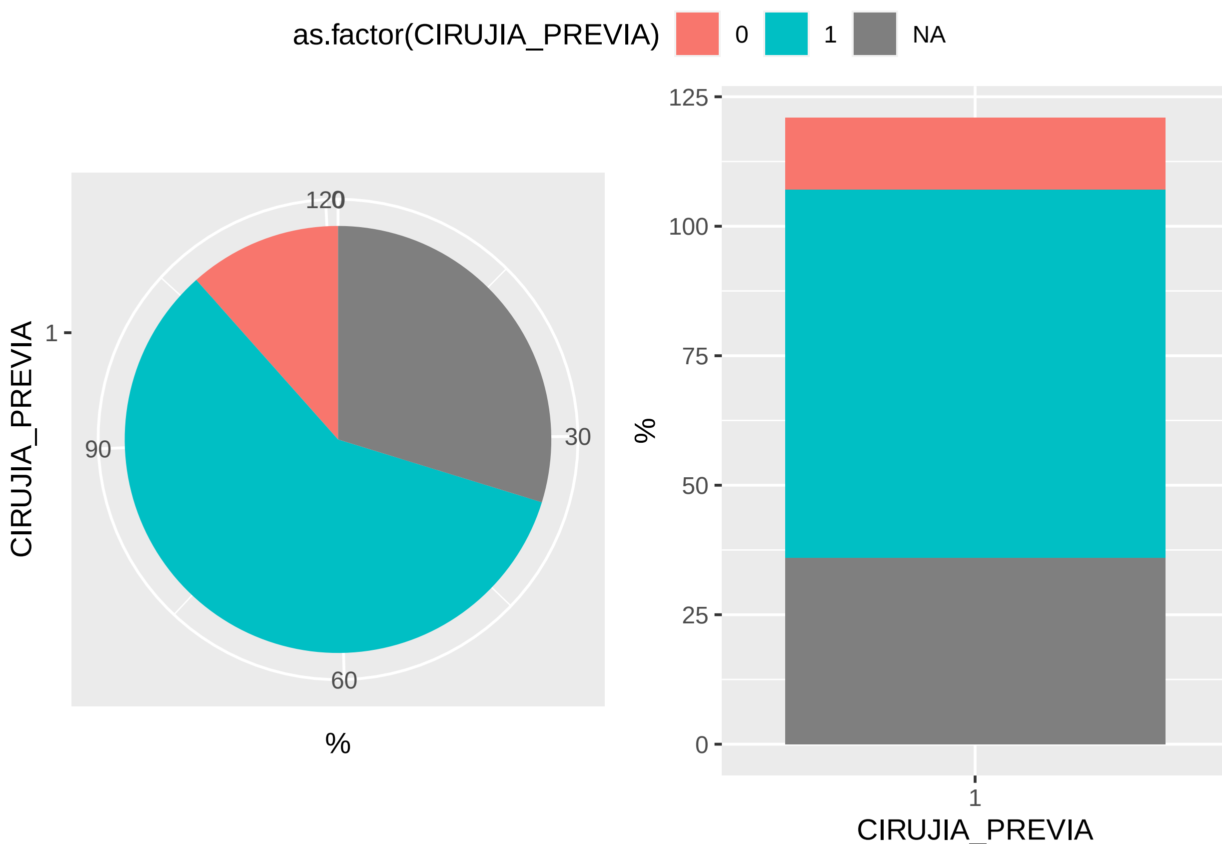

```r
# Frequency tables. Descriptive analysis
# Remember that package 'summarytools' is required
freq(as.factor(CIRUJIA_PREVIA))
```

CIRUJIA_PREVIA

```
## Frequencies
##
##                Freq   % Valid   % Valid Cum.   % Total   % Total Cum.
## ----------- ------ --------- -------------- --------- --------------
##           0     14     16.47          16.47     11.57          11.57
```

```
##           1      71      83.53          100.00      58.68          70.25
##        <NA>      36                                29.75         100.00
##       Total     121     100.00          100.00     100.00         100.00
```

```
# Pie chart and bar graph
p1<-ggplot(data_xlsx,aes(x=factor(1),fill=as.factor(CIRUJIA_PREVIA)))+geom_bar()+
  coord_polar("y")+labs(x="CIRUJIA_PREVIA",y="%")
p2<-ggplot(data_xlsx,aes(x=factor(1),fill=as.factor(CIRUJIA_PREVIA)))+geom_bar()+
  labs(x="CIRUJIA_PREVIA",y="%")

# This function controls the graphical output device
# Remember that package 'ggpubr' is required
ggarrange(p1,p2,nrow = 1,ncol=2, common.legend = TRUE)
```



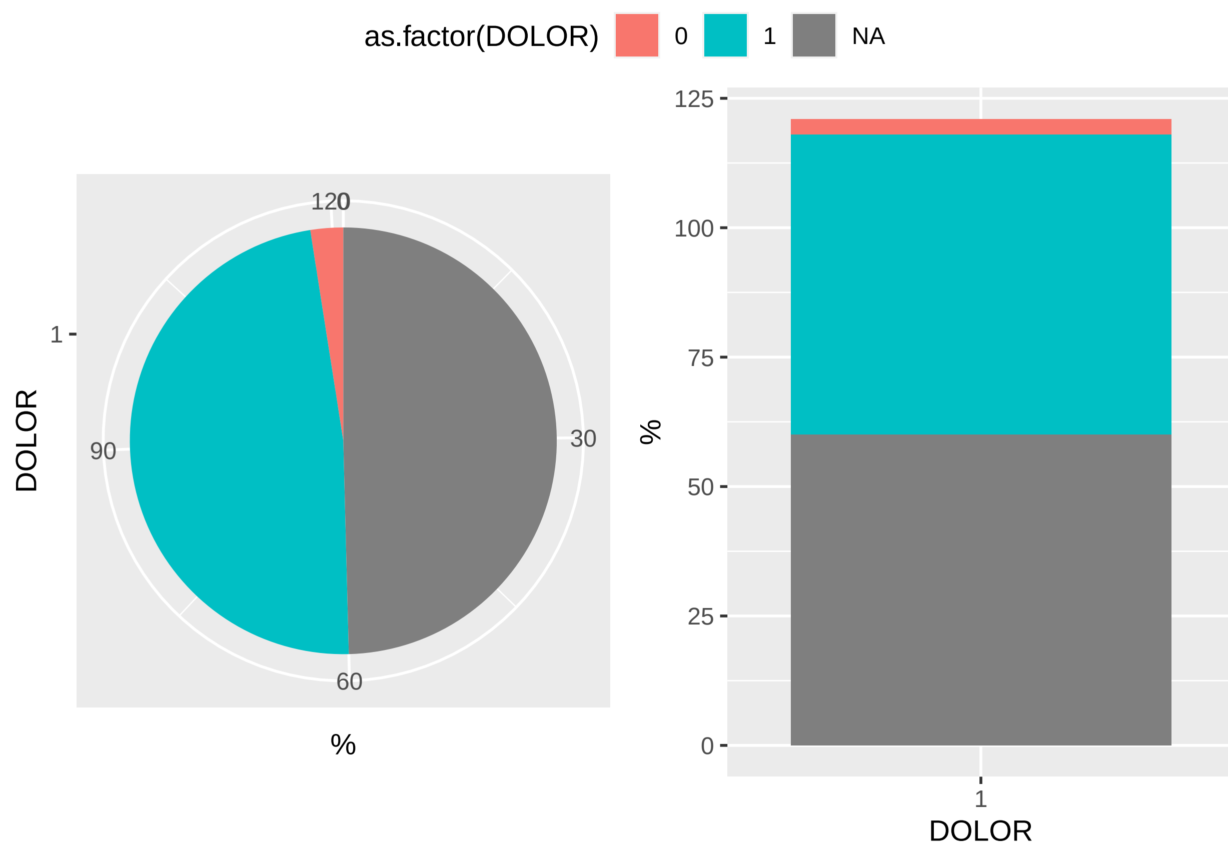

```
# Frequency tables. Descriptive analysis
# Remember that package 'summarytools' is required
freq(as.factor(DOLOR))
```

DOLOR

```
## Frequencies
##
##                Freq   % Valid   % Valid Cum.   % Total   % Total Cum.
## ----------- ------ --------- -------------- --------- --------------
##           0      3      4.92           4.92      2.48           2.48
##           1     58     95.08         100.00     47.93          50.41
```

```
##        <NA>     60                              49.59          100.00
##       Total    121    100.00          100.00   100.00          100.00
```

```
# Pie chart and bar graph
p1<-ggplot(data_xlsx,aes(x=factor(1),fill=as.factor(DOLOR)))+geom_bar()+
  coord_polar("y")+labs(x="DOLOR",y="%")
p2<-ggplot(data_xlsx,aes(x=factor(1),fill=as.factor(DOLOR)))+geom_bar()+
  labs(x="DOLOR",y="%")

# This function controls the graphical output device
# Remember that package 'ggpubr' is required
ggarrange(p1,p2,nrow = 1,ncol=2, common.legend = TRUE)
```



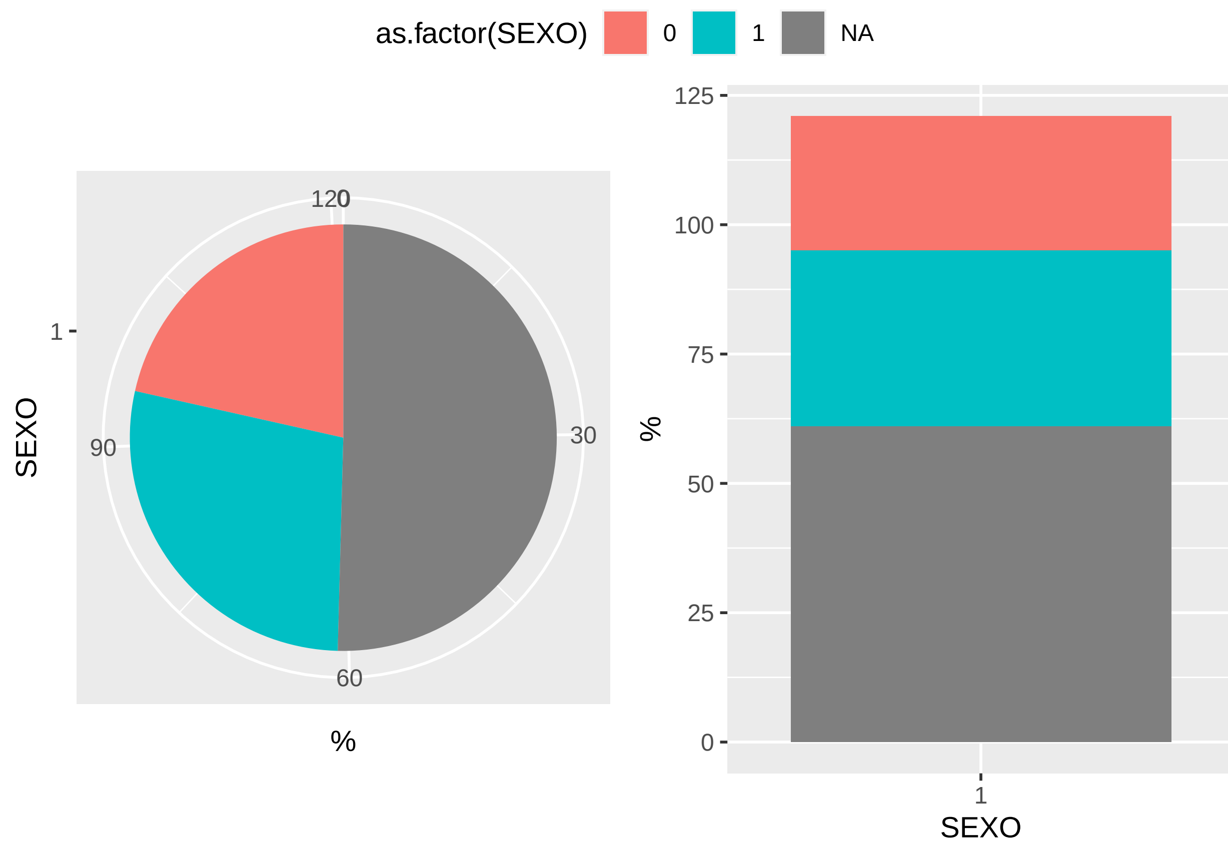

```
# Frequency tables. Descriptive analysis
# Remember that package 'summarytools' is required
freq(as.factor(SEXO))
```

SEXO

```
## Frequencies
##
##                Freq   % Valid   % Valid Cum.   % Total   % Total Cum.
## ----------- ------ --------- -------------- --------- --------------
##           0     26     43.33          43.33     21.49          21.49
##           1     34     56.67         100.00     28.10          49.59
##        <NA>     61                              50.41         100.00
```

```
##       Total    121    100.00          100.00    100.00          100.00
```

```
# Pie chart and bar graph
p1<-ggplot(data_xlsx,aes(x=factor(1),fill=as.factor(SEXO)))+geom_bar()+
  coord_polar("y")+labs(x="SEXO",y="%")
p2<-ggplot(data_xlsx,aes(x=factor(1),fill=as.factor(SEXO)))+geom_bar()+
  labs(x="SEXO",y="%")

# This function controls the graphical output device
# Remember that package 'ggpubr' is required
ggarrange(p1,p2,nrow = 1,ncol=2, common.legend = TRUE)
```

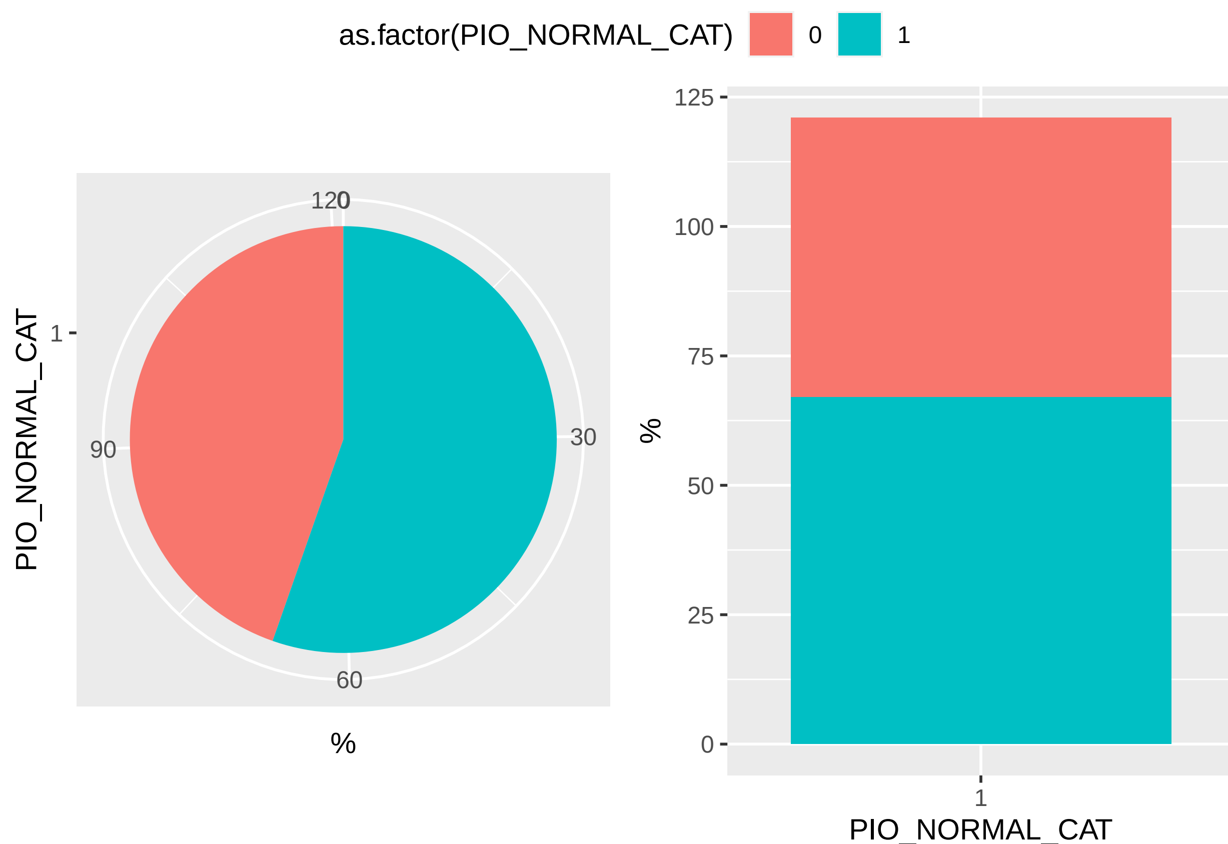



```
# Frequency tables. Descriptive analysis
# Remember that package 'summarytools' is required
freq(as.factor(PIO_NORMAL_CAT))
```

PIO_NORMAL_CAT

```
## Frequencies
##
##                Freq   % Valid   % Valid Cum.   % Total   % Total Cum.
## ----------- ------ --------- -------------- --------- --------------
##           0     54     44.63          44.63     44.63          44.63
##           1     67     55.37         100.00     55.37         100.00
##        <NA>      0                               0.00         100.00
##       Total    121    100.00         100.00    100.00         100.00
```

```
# Pie chart and bar graph
p1<-ggplot(data_xlsx,aes(x=factor(1),fill=as.factor(PIO_NORMAL_CAT)))+geom_bar()+
  coord_polar("y")+labs(x="PIO_NORMAL_CAT",y="%")
p2<-ggplot(data_xlsx,aes(x=factor(1),fill=as.factor(PIO_NORMAL_CAT)))+geom_bar()+
  labs(x="PIO_NORMAL_CAT",y="%")

# This function controls the graphical output device
# Remember that package 'ggpubr' is required
ggarrange(p1,p2,nrow = 1,ncol=2, common.legend = TRUE)
```



Not available data (NA)

Identification and treatment

The decision for not available data is to replace them by the mean of their variable. This decision has been made assuming that the behavior of the *NA* is totally random (this would have to be analyzed in depth to confirm this decision made).

The following source code defines the function *not_available* whose utility is to deal with not available data.

```
# Construction of the function that handles missing values.
not_available<-function(data,na.rm=F){
  data[is.na(data)]<-mean(data,na.rm=T)
  data
}

# We call the not_available function for each variable in the database
```

```r
data_xlsx$OJO<-not_available(data_xlsx$OJO)
data_xlsx$TIPO_GLAUCOMA<-not_available(data_xlsx$TIPO_GLAUCOMA)
data_xlsx$N_IMPACTOS<-not_available(data_xlsx$N_IMPACTOS)
data_xlsx$CUADRANTES<-not_available(data_xlsx$CUADRANTES)
data_xlsx$ENERGIA_IMPACTO<-not_available(data_xlsx$ENERGIA_IMPACTO)
data_xlsx$ENERGIA_TOTAL<-not_available(data_xlsx$ENERGIA_TOTAL)
data_xlsx$CIRUJIA_PREVIA<-not_available(data_xlsx$CIRUJIA_PREVIA)
data_xlsx$PIO_PRE_SLT<-not_available(data_xlsx$PIO_PRE_SLT)
data_xlsx$PIO_1_SEMANA<-not_available(data_xlsx$PIO_1_SEMANA)
data_xlsx$PIO_1_MES<-not_available(data_xlsx$PIO_1_MES)
data_xlsx$PIO_3_MES<-not_available(data_xlsx$PIO_3_MES)
data_xlsx$FARMACOS_PRE<-not_available(data_xlsx$FARMACOS_PRE)
data_xlsx$FARMACOS_1_MES<-not_available(data_xlsx$FARMACOS_1_MES)
data_xlsx$FARMACOS_3_MES<-not_available(data_xlsx$FARMACOS_3_MES)
data_xlsx$DOLOR<-not_available(data_xlsx$DOLOR)
data_xlsx$SEXO<-not_available(data_xlsx$SEXO)
data_xlsx$EDAD<-not_available(data_xlsx$EDAD)
data_xlsx$PIO_NORMAL<-not_available(data_xlsx$PIO_NORMAL)
data_xlsx$PIO_NORMAL_CAT<-not_available(data_xlsx$PIO_NORMAL_CAT)

# We view the data again
head(data_xlsx,n=3)
```

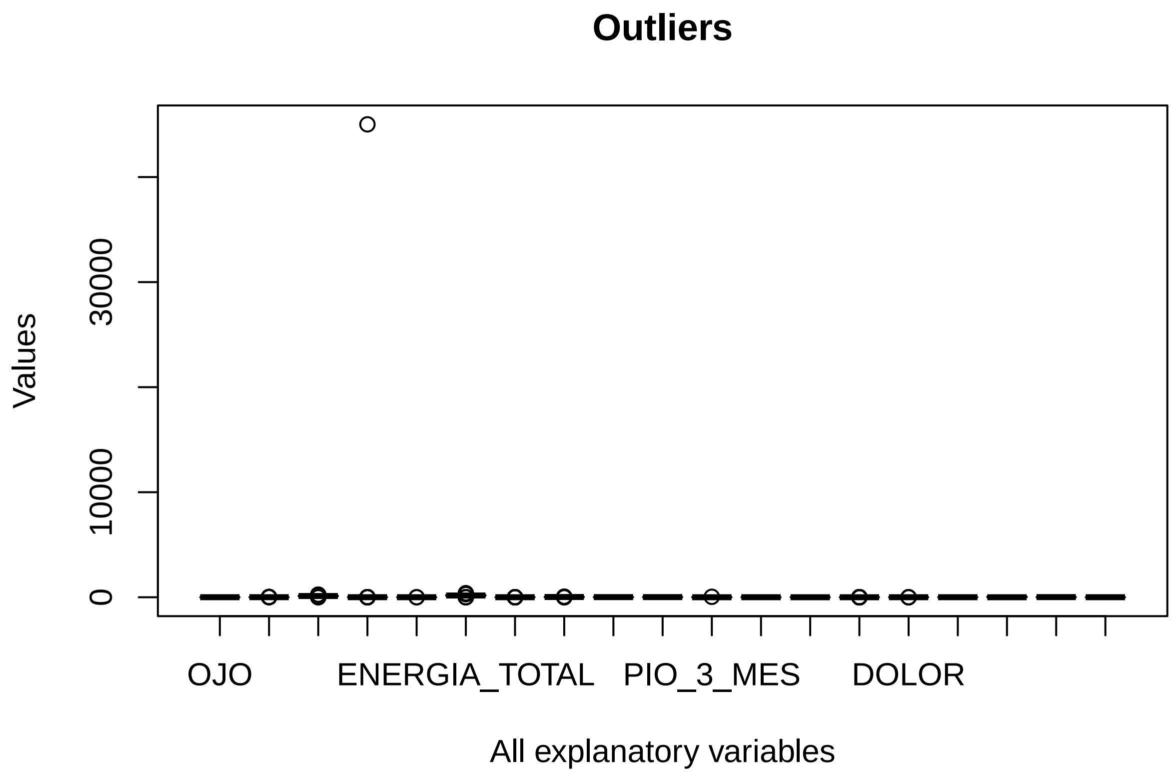
```
## # A tibble: 3 x 19
##     OJO TIPO_GLAUCOMA N_IMPACTOS CUADRANTES ENERGIA_IMPACTO ENERGIA_TOTAL
##   <dbl>         <dbl>      <dbl>      <dbl>           <dbl>         <dbl>
## 1     0          0           112          4             1.5           174
## 2     1          4.87        108          4             1.2           128
## 3     0          1           123          4             1.1           133
## # i 13 more variables: CIRUJIA_PREVIA <dbl>, PIO_PRE_SLT <dbl>,
## #   PIO_1_SEMANA <dbl>, PIO_1_MES <dbl>, PIO_3_MES <dbl>, FARMACOS_PRE <dbl>,
## #   FARMACOS_1_MES <dbl>, FARMACOS_3_MES <dbl>, DOLOR <dbl>, SEXO <dbl>,
## #   EDAD <dbl>, PIO_NORMAL <dbl>, PIO_NORMAL_CAT <dbl>
```

Outliers

Identification

This graphical output shows together the boxplots of all the quantitative variables. Since all the variables are standardized there is no problem with the scales.

```r
# Boxplots of all variables together
# This visualization is not the best due to the difference between the scales
boxplot(data_xlsx,main="Outliers",
        xlab="All explanatory variables",
        ylab="Values",
        col=c(1:15))
```



Making decisions

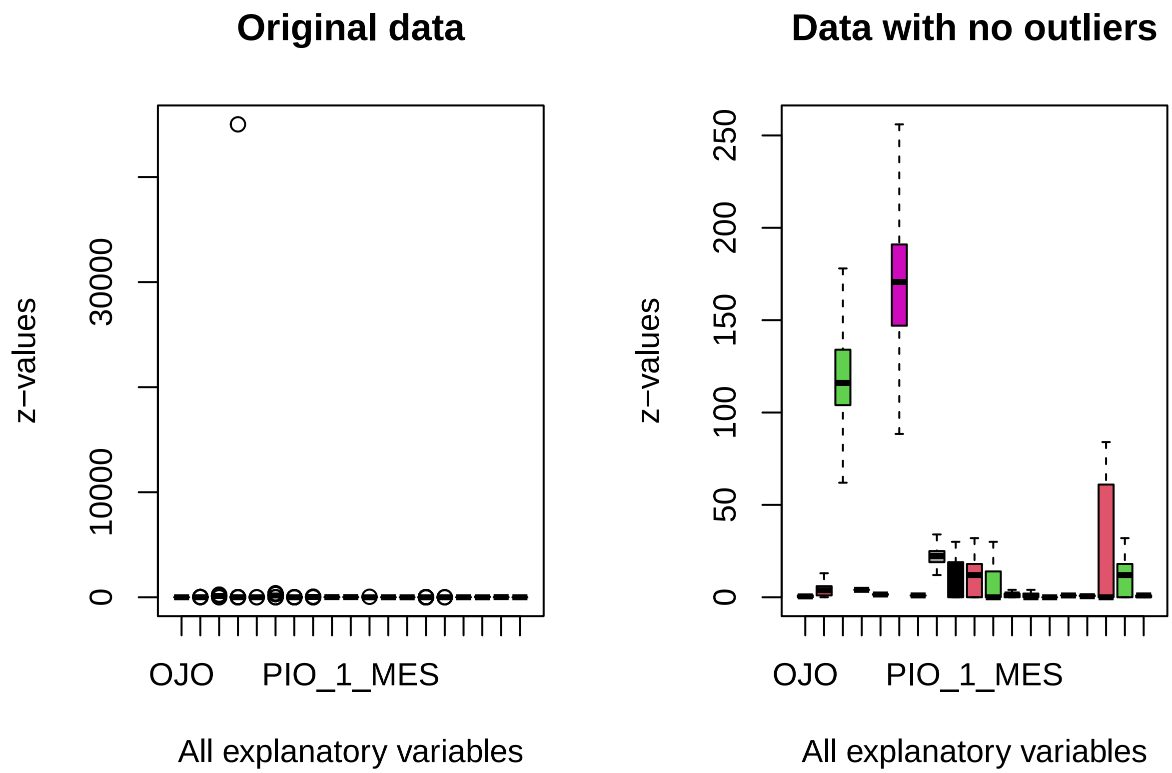
From previous graphical outputs it is noticed the presence of outliers for several variables. It is relevant to take into account these values since multivariate methods, such as principal component analisis (PCA), are sensitive to this fact.

This is not a light topic and it should be analyzed outlier per outlier. However, **the decision for outliers is to replace them by the mean of their variable**.

The following source code defines the function *outlier* whose utility is to deal with the univariate outliers.

```
# Recursive function that modifies outliers by the mean of their variable
outlier<-function(data,na.rm=T){

  H<-1.5*IQR(data)
  data[data<quantile(data,0.25,na.rm = T)-H]<-NA
  data[data>quantile(data,0.75, na.rm = T)+H]<-NA
  data[is.na(data)]<-mean(data, na.rm = T)
  H<-1.5*IQR(data)

  if (TRUE %in% (data<quantile(data,0.25,na.rm = T)-H) |
      TRUE %in% (data>quantile(data,0.75,na.rm = T)+H))
    outlier(data)
  else

    return(data)
```

```
}

# This data.frame is to preserve original data once the outliers are modified
data_xlsx_aux<-data_xlsx

# Called to outlier function for each variable identified with outliers
data_xlsx_aux$OJO<-outlier(data_xlsx_aux$OJO)
data_xlsx_aux$TIPO_GLAUCOMA<-outlier(data_xlsx_aux$TIPO_GLAUCOMA)
data_xlsx_aux$N_IMPACTOS<-outlier(data_xlsx_aux$N_IMPACTOS)
data_xlsx_aux$CUADRANTES<-outlier(data_xlsx_aux$CUADRANTES)
data_xlsx_aux$ENERGIA_IMPACTO<-outlier(data_xlsx_aux$ENERGIA_IMPACTO)
data_xlsx_aux$ENERGIA_TOTAL<-outlier(data_xlsx_aux$ENERGIA_TOTAL)
data_xlsx_aux$CIRUJIA_PREVIA<-outlier(data_xlsx_aux$CIRUJIA_PREVIA)
data_xlsx_aux$PIO_PRE_SLT<-outlier(data_xlsx_aux$PIO_PRE_SLT)
data_xlsx_aux$PIO_1_SEMANA<-outlier(data_xlsx_aux$PIO_1_SEMANA)
data_xlsx_aux$PIO_1_MES<-outlier(data_xlsx_aux$PIO_1_MES)
data_xlsx_aux$PIO_3_MES<-outlier(data_xlsx_aux$PIO_3_MES)
data_xlsx_aux$FARMACOS_PRE<-outlier(data_xlsx_aux$FARMACOS_PRE)
data_xlsx_aux$FARMACOS_1_MES<-outlier(data_xlsx_aux$FARMACOS_1_MES)
data_xlsx_aux$FARMACOS_3_MES<-outlier(data_xlsx_aux$FARMACOS_3_MES)
data_xlsx_aux$DOLOR<-outlier(data_xlsx_aux$DOLOR)
data_xlsx_aux$SEXO<-outlier(data_xlsx_aux$SEXO)
data_xlsx_aux$EDAD<-outlier(data_xlsx_aux$EDAD)
data_xlsx_aux$PIO_NORMAL<-outlier(data_xlsx_aux$PIO_NORMAL)
data_xlsx_aux$PIO_NORMAL_CAT<-outlier(data_xlsx_aux$PIO_NORMAL_CAT)

# We compare the original data and the fixed ones with respective boxplots
par(mfrow=c(1,2))
# Boxplot original data
boxplot(data_xlsx,main="Original data",
        xlab="All explanatory variables",
        ylab="z-values",
        col=c(1:15))
# Boxplot fixed data
boxplot(data_xlsx_aux,main="Data with no outliers",
        xlab="All explanatory variables",
        ylab="z-values",
        col=c(1:15))
```



This is another joint visualization of the boxplots without the effect of the difference in scales.

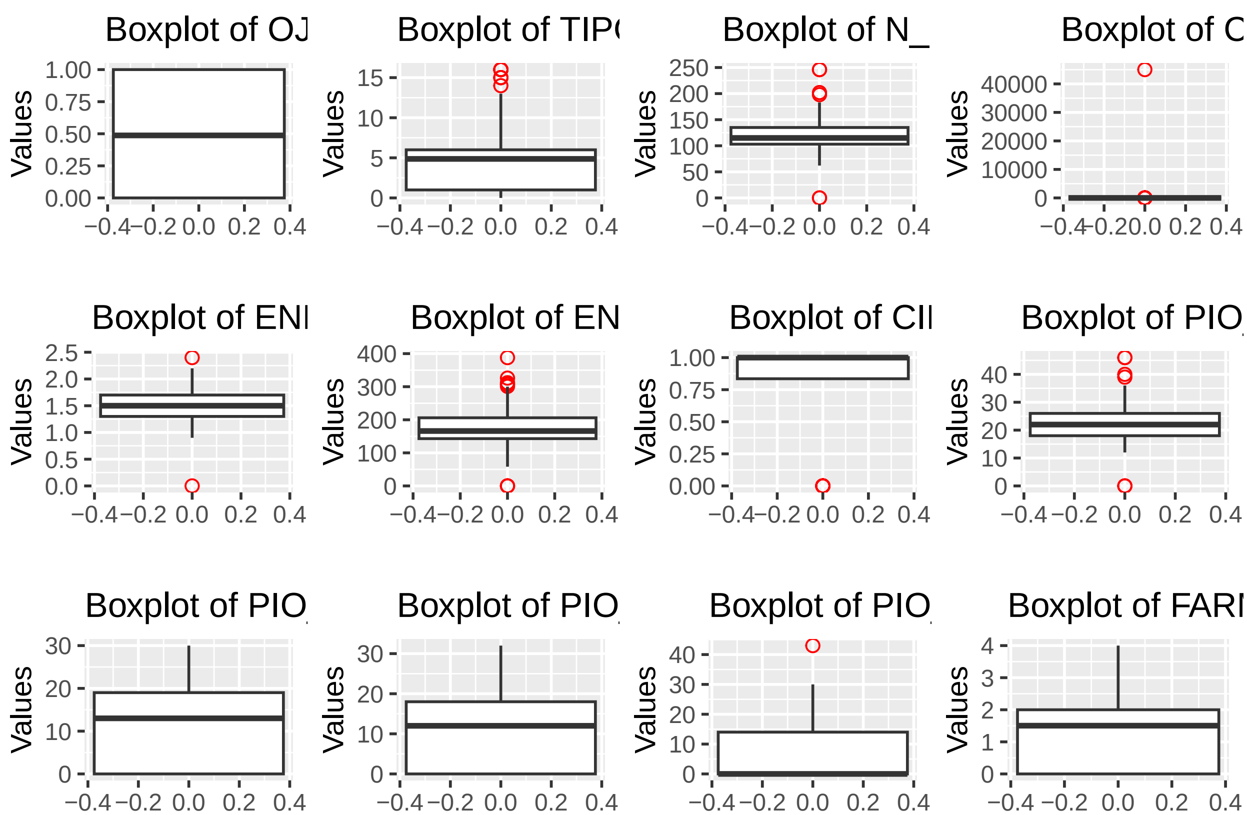
With outliers:

```
# Boxplots of all quantitative variables together
# Remember that package 'ggplot2' is required
p1<-ggplot(data_xlsx,aes(x=OJO))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of OJO",x="Values",y="")

p2<-ggplot(data_xlsx,aes(x=TIPO_GLAUCOMA))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of TIPO_GLAUCOMA",x="Values",y="")

p3<-ggplot(data_xlsx,aes(x=N_IMPACTOS))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of N_IMPACTOS",x="Values",y="")

p4<-ggplot(data_xlsx,aes(x=CUADRANTES))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of CUADRANTES",x="Values",y="")

p5<-ggplot(data_xlsx,aes(x=ENERGIA_IMPACTO))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of ENERGIA_IMPACTO",x="Values",y="")

p6<-ggplot(data_xlsx,aes(x=ENERGIA_TOTAL))+
```

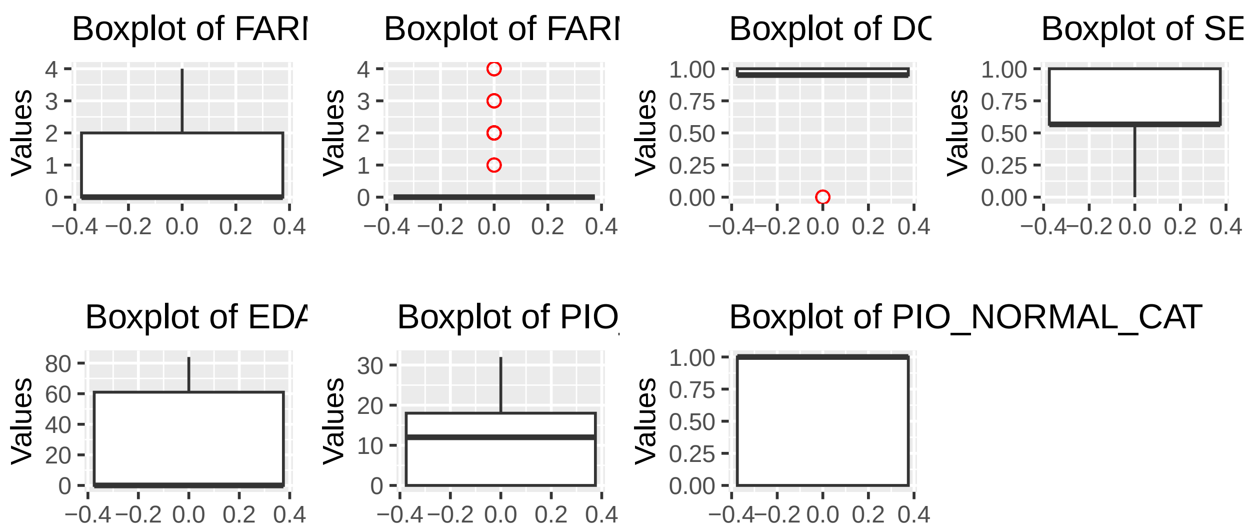
```
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of ENERGIA_TOTAL",x="Values",y="")

p7<-ggplot(data_xlsx,aes(x=CIRUJIA_PREVIA))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of CIRUJIA_PREVIA",x="Values",y="")

p8<-ggplot(data_xlsx,aes(x=PIO_PRE_SLT))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of PIO_PRE_SLT",x="Values",y="")

p9<-ggplot(data_xlsx,aes(x=PIO_1_SEMANA))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of PIO_1_SEMANA",x="Values",y="")

p10<-ggplot(data_xlsx,aes(x=PIO_1_MES))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of PIO_1_MES",x="Values",y="")

p11<-ggplot(data_xlsx,aes(x=PIO_3_MES))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of PIO_3_MES",x="Values",y="")

p12<-ggplot(data_xlsx,aes(x=FARMACOS_PRE))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of FARMACOS_PRE",x="Values",y="")

p13<-ggplot(data_xlsx,aes(x=FARMACOS_1_MES))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of FARMACOS_1_MES",x="Values",y="")

p14<-ggplot(data_xlsx,aes(x=FARMACOS_3_MES))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of FARMACOS_3_MES",x="Values",y="")

p15<-ggplot(data_xlsx,aes(x=DOLOR))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of DOLOR",x="Values",y="")

p16<-ggplot(data_xlsx,aes(x=SEXO))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of SEXO",x="Values",y="")

p17<-ggplot(data_xlsx,aes(x=EDAD))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of EDAD",x="Values",y="")

p18<-ggplot(data_xlsx,aes(x=PIO_NORMAL))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of PIO_NORMAL",x="Values",y="")

p19<-ggplot(data_xlsx,aes(x=PIO_NORMAL_CAT))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
```

```
  coord_flip()+labs(title = "Boxplot of PIO_NORMAL_CAT",x="Values",y="")
# This function controls the graphical output device
# Remember that package 'ggpubr' is required
ggarrange(p1,p2,p3,p4,p5,p6,p7,p8,p9,p10,p11,p12,p13,p14,p15,p16,p17,p18,p19,
          nrow=3, ncol=4,
          commond.legend=FALSE)
```

```
## $`1`
```




```
##
## $`2`
```




```
##
## attr(,"class")
## [1] "list"      "ggarrange"
```

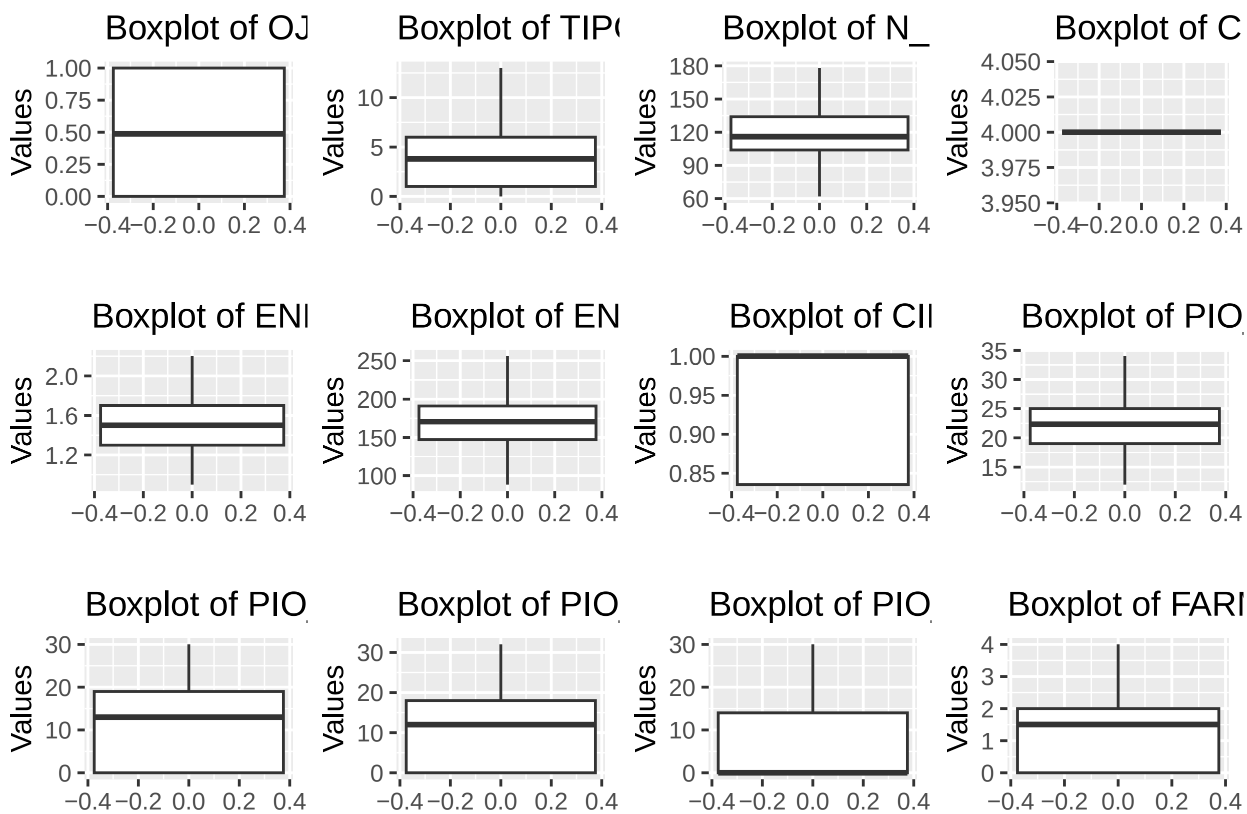
Without outliers:

```
# Boxplots of all quantitative variables together
# Remember that package 'ggplot2' is required
p1<-ggplot(data_xlsx_aux,aes(x=OJO))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of OJO",x="Values",y="")

p2<-ggplot(data_xlsx_aux,aes(x=TIPO_GLAUCOMA))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of TIPO_GLAUCOMA",x="Values",y="")

p3<-ggplot(data_xlsx_aux,aes(x=N_IMPACTOS))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of N_IMPACTOS",x="Values",y="")

p4<-ggplot(data_xlsx_aux,aes(x=CUADRANTES))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of CUADRANTES",x="Values",y="")

p5<-ggplot(data_xlsx_aux,aes(x=ENERGIA_IMPACTO))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of ENERGIA_IMPACTO",x="Values",y="")
```

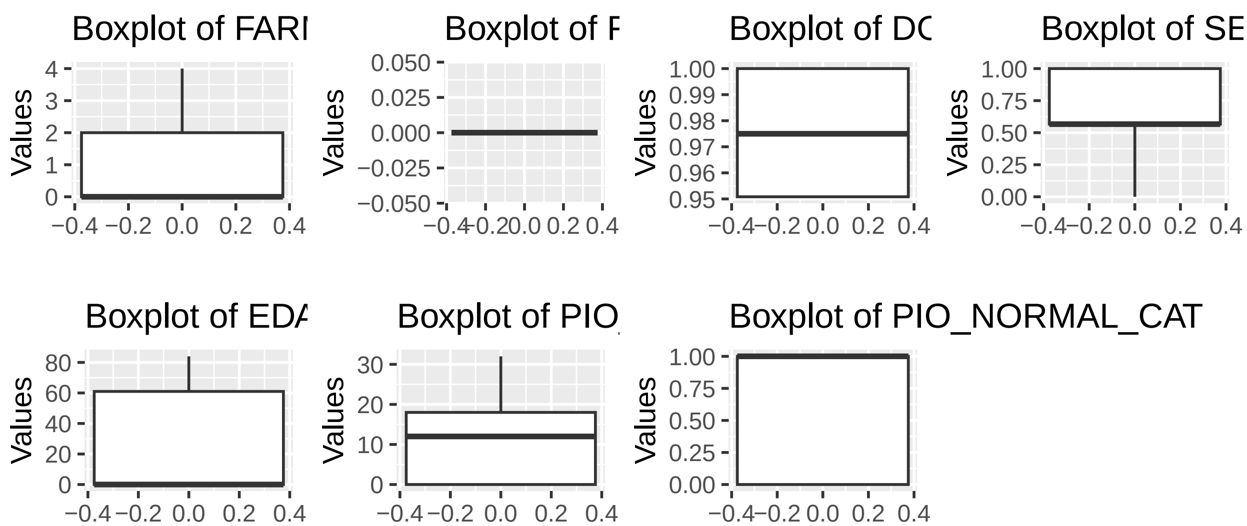
```r
p6<-ggplot(data_xlsx_aux,aes(x=ENERGIA_TOTAL))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of ENERGIA_TOTAL",x="Values",y="")

p7<-ggplot(data_xlsx_aux,aes(x=CIRUJIA_PREVIA))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of CIRUJIA_PREVIA",x="Values",y="")

p8<-ggplot(data_xlsx_aux,aes(x=PIO_PRE_SLT))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of PIO_PRE_SLT",x="Values",y="")

p9<-ggplot(data_xlsx_aux,aes(x=PIO_1_SEMANA))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of PIO_1_SEMANA",x="Values",y="")

p10<-ggplot(data_xlsx_aux,aes(x=PIO_1_MES))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of PIO_1_MES",x="Values",y="")

p11<-ggplot(data_xlsx_aux,aes(x=PIO_3_MES))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of PIO_3_MES",x="Values",y="")

p12<-ggplot(data_xlsx_aux,aes(x=FARMACOS_PRE))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of FARMACOS_PRE",x="Values",y="")

p13<-ggplot(data_xlsx_aux,aes(x=FARMACOS_1_MES))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of FARMACOS_1_MES",x="Values",y="")

p14<-ggplot(data_xlsx_aux,aes(x=FARMACOS_3_MES))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of FARMACOS_3_MES",x="Values",y="")

p15<-ggplot(data_xlsx_aux,aes(x=DOLOR))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of DOLOR",x="Values",y="")

p16<-ggplot(data_xlsx_aux,aes(x=SEXO))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of SEXO",x="Values",y="")

p17<-ggplot(data_xlsx_aux,aes(x=EDAD))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of EDAD",x="Values",y="")

p18<-ggplot(data_xlsx_aux,aes(x=PIO_NORMAL))+
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of PIO_NORMAL",x="Values",y="")

p19<-ggplot(data_xlsx_aux,aes(x=PIO_NORMAL_CAT))+
```

```r
  geom_boxplot(outlier.colour="red", outlier.shape=1,outlier.size=2)+
  coord_flip()+labs(title = "Boxplot of PIO_NORMAL_CAT",x="Values",y="")
# This function controls the graphical output device
# Remember that package 'ggpubr' is required
ggarrange(p1,p2,p3,p4,p5,p6,p7,p8,p9,p10,p11,p12,p13,p14,p15,p16,p17,p18,p19,
          nrow=3, ncol=4,
          commond.legend=FALSE)
```

```
## $`1`
```




```
##
## $`2`
```




```
##
## attr(,"class")
## [1] "list"      "ggarrange"
```

Principal component analysis

Requirements

Correlated variables

If we display the correlation matrix, we will see that there are a couple of rows with the value 'NA':

```
###############################
# Correlation at sample level #
###############################

# Are the variables correlated at sample level?
correlation_matrix<-cor(data_xlsx_aux)
correlation_matrix
```

```
##                            OJO TIPO_GLAUCOMA   N_IMPACTOS CUADRANTES
## OJO              1.0000000000   -0.04047399  0.007229181         NA
## TIPO_GLAUCOMA   -0.0404739875    1.00000000 -0.156088531         NA
## N_IMPACTOS       0.0072291815   -0.15608853  1.000000000         NA
## CUADRANTES                 NA            NA           NA          1
## ENERGIA_IMPACTO  0.0315518927   -0.03332637  0.022476030         NA
## ENERGIA_TOTAL    0.0723942519    0.03670639  0.450430971         NA
```

```
## CIRUJIA_PREVIA   0.0073002686    0.14795950 -0.061184813         NA
## PIO_PRE_SLT      0.0261824909    0.06530760 -0.045190732         NA
## PIO_1_SEMANA     0.0360148309    0.13035711 -0.088090444         NA
## PIO_1_MES       -0.0764291803   -0.03286705  0.121378717         NA
## PIO_3_MES        0.0005921436   -0.16711797  0.174664359         NA
## FARMACOS_PRE    -0.0226470426   -0.23728180  0.101592944         NA
## FARMACOS_1_MES   0.0115065566   -0.25361853  0.139737434         NA
## FARMACOS_3_MES             NA            NA           NA         NA
## DOLOR            0.0008585046   -0.17523363  0.031041769         NA
## SEXO             0.0273059341    0.18732486 -0.111157729         NA
## EDAD            -0.0249843282   -0.24472004  0.104380092         NA
## PIO_NORMAL      -0.0764291803   -0.03286705  0.121378717         NA
## PIO_NORMAL_CAT  -0.0910674076   -0.06503823 -0.036040315         NA
##                 ENERGIA_IMPACTO ENERGIA_TOTAL CIRUJIA_PREVIA  PIO_PRE_SLT
## OJO                 0.031551893   0.072394252    0.007300269  0.026182491
## TIPO_GLAUCOMA      -0.033326365   0.036706392    0.147959496  0.065307602
## N_IMPACTOS          0.022476030   0.450430971   -0.061184813 -0.045190732
## CUADRANTES                   NA            NA             NA           NA
## ENERGIA_IMPACTO     1.000000000   0.469283651    0.163026114  0.104558117
## ENERGIA_TOTAL       0.469283651   1.000000000   -0.011069041 -0.001359308
## CIRUJIA_PREVIA      0.163026114  -0.011069041    1.000000000  0.157872963
## PIO_PRE_SLT         0.104558117  -0.001359308    0.157872963  1.000000000
## PIO_1_SEMANA        0.006786965  -0.106190549    0.358529456 -0.002524094
## PIO_1_MES           0.020785153  -0.038116060    0.386637166  0.157457613
## PIO_3_MES          -0.010003597  -0.059063828    0.327413593  0.127881329
## FARMACOS_PRE        0.098806991   0.052096593   -0.017276419 -0.359018173
## FARMACOS_1_MES     -0.060577672  -0.091832957    0.237016562 -0.081238377
## FARMACOS_3_MES               NA            NA             NA           NA
## DOLOR               0.006916478  -0.095385965    0.501685953  0.086009321
## SEXO               -0.108242035   0.071383759    0.084921303  0.096288973
## EDAD                0.043398254  -0.074554278    0.485099418  0.083414769
## PIO_NORMAL          0.020785153  -0.038116060    0.386637166  0.157457613
## PIO_NORMAL_CAT     -0.047049322  -0.054781161   -0.316821239 -0.135108804
##                 PIO_1_SEMANA   PIO_1_MES     PIO_3_MES FARMACOS_PRE
## OJO              0.036014831 -0.07642918  0.0005921436  -0.02264704
## TIPO_GLAUCOMA    0.130357111 -0.03286705 -0.1671179707  -0.23728180
## N_IMPACTOS      -0.088090444  0.12137872  0.1746643588   0.10159294
## CUADRANTES                NA          NA            NA           NA
## ENERGIA_IMPACTO  0.006786965  0.02078515 -0.0100035968   0.09880699
## ENERGIA_TOTAL   -0.106190549 -0.03811606 -0.0590638279   0.05209659
## CIRUJIA_PREVIA   0.358529456  0.38663717  0.3274135930  -0.01727642
## PIO_PRE_SLT     -0.002524094  0.15745761  0.1278813285  -0.35901817
## PIO_1_SEMANA     1.000000000  0.65662429  0.3569956894   0.24348566
## PIO_1_MES        0.656624289  1.00000000  0.6082639571   0.21962717
## PIO_3_MES        0.356995689  0.60826396  1.0000000000   0.07659788
## FARMACOS_PRE     0.243485665  0.21962717  0.0765978785   1.00000000
## FARMACOS_1_MES   0.459330045  0.53786628  0.4425323708   0.59226275
## FARMACOS_3_MES            NA          NA            NA           NA
## DOLOR            0.529142006  0.79161083  0.6099276923   0.30152167
## SEXO            -0.125506479 -0.10278748 -0.0146488596  -0.16407024
## EDAD             0.455598165  0.75313605  0.5587507386   0.30534137
## PIO_NORMAL       0.656624289  1.00000000  0.6082639571   0.21962717
## PIO_NORMAL_CAT  -0.408158219 -0.68476343 -0.4162236728  -0.13735098
##                 FARMACOS_1_MES FARMACOS_3_MES        DOLOR         SEXO
```

```
## OJO                  0.01150656             NA  0.0008585046  0.027305934
## TIPO_GLAUCOMA       -0.25361853             NA -0.1752336263  0.187324861
## N_IMPACTOS           0.13973743             NA  0.0310417691 -0.111157729
## CUADRANTES                   NA             NA            NA           NA
## ENERGIA_IMPACTO     -0.06057767             NA  0.0069164781 -0.108242035
## ENERGIA_TOTAL       -0.09183296             NA -0.0953859650  0.071383759
## CIRUJIA_PREVIA       0.23701656             NA  0.5016859532  0.084921303
## PIO_PRE_SLT         -0.08123838             NA  0.0860093213  0.096288973
## PIO_1_SEMANA         0.45933005             NA  0.5291420063 -0.125506479
## PIO_1_MES            0.53786628             NA  0.7916108300 -0.102787485
## PIO_3_MES            0.44253237             NA  0.6099276923 -0.014648860
## FARMACOS_PRE         0.59226275             NA  0.3015216686 -0.164070244
## FARMACOS_1_MES       1.00000000             NA  0.6621333589 -0.277624278
## FARMACOS_3_MES               NA              1            NA           NA
## DOLOR                0.66213336             NA  1.0000000000  0.003252437
## SEXO                -0.27762428             NA  0.0032524367  1.000000000
## EDAD                 0.65487613             NA  0.9318865791 -0.057055091
## PIO_NORMAL           0.53786628             NA  0.7916108300 -0.102787485
## PIO_NORMAL_CAT      -0.30338490             NA -0.6381159167 -0.092111666
##                        EDAD  PIO_NORMAL PIO_NORMAL_CAT
## OJO             -0.02498433 -0.07642918    -0.09106741
## TIPO_GLAUCOMA   -0.24472004 -0.03286705    -0.06503823
## N_IMPACTOS       0.10438009  0.12137872    -0.03604032
## CUADRANTES               NA          NA             NA
## ENERGIA_IMPACTO  0.04339825  0.02078515    -0.04704932
## ENERGIA_TOTAL   -0.07455428 -0.03811606    -0.05478116
## CIRUJIA_PREVIA   0.48509942  0.38663717    -0.31682124
## PIO_PRE_SLT      0.08341477  0.15745761    -0.13510880
## PIO_1_SEMANA     0.45559816  0.65662429    -0.40815822
## PIO_1_MES        0.75313605  1.00000000    -0.68476343
## PIO_3_MES        0.55875074  0.60826396    -0.41622367
## FARMACOS_PRE     0.30534137  0.21962717    -0.13735098
## FARMACOS_1_MES   0.65487613  0.53786628    -0.30338490
## FARMACOS_3_MES           NA          NA             NA
## DOLOR            0.93188658  0.79161083    -0.63811592
## SEXO            -0.05705509 -0.10278748    -0.09211167
## EDAD             1.00000000  0.75313605    -0.56855536
## PIO_NORMAL       0.75313605  1.00000000    -0.68476343
## PIO_NORMAL_CAT  -0.56855536 -0.68476343     1.00000000
```

It is observed that for the variables CUADRANTES and FARMACOS_3_MES, 'NA' is obtained. This is because all the records of these variables have the same value (in this case, this is due to the treatment of outliers), making their standard deviation equal to 0 and, therefore, we cannot calculate their correlation with other variables (since we would be dividing by 0, which is not possible). Therefore, we eliminate these two variables.

```r
# The variables 'CUADRANTES' is eliminated.
data_xlsx_aux <- data_xlsx_aux[, -c(4,14)]

# The first three records in the database are displayed
head(data_xlsx_aux,n=3)
```

```
## # A tibble: 3 x 17
##     OJO TIPO_GLAUCOMA N_IMPACTOS ENERGIA_IMPACTO ENERGIA_TOTAL CIRUJIA_PREVIA
##   <dbl>         <dbl>      <dbl>           <dbl>         <dbl>          <dbl>
```

```
## 1     0          0           112             1.5           174           0.835
## 2     1          4.87        108             1.2           128           1
## 3     0          1           123             1.1           133           1
## # i 11 more variables: PIO_PRE_SLT <dbl>, PIO_1_SEMANA <dbl>, PIO_1_MES <dbl>,
## #   PIO_3_MES <dbl>, FARMACOS_PRE <dbl>, FARMACOS_1_MES <dbl>, DOLOR <dbl>,
## #   SEXO <dbl>, EDAD <dbl>, PIO_NORMAL <dbl>, PIO_NORMAL_CAT <dbl>
```

According to the numerical results below, it is observed that the data **are correlated** both **at the sample level** (see correlation matrix) and **at the population level** (Bartlett's sphericity test is significant).

```
################################
# Correlation at sample level #
################################

# Are the variables correlated at sample level?
correlation_matrix<-cor(data_xlsx_aux)
correlation_matrix
```

```
##                          OJO TIPO_GLAUCOMA   N_IMPACTOS ENERGIA_IMPACTO
## OJO             1.0000000000   -0.04047399  0.007229181     0.031551893
## TIPO_GLAUCOMA   -0.0404739875    1.00000000 -0.156088531    -0.033326365
## N_IMPACTOS      0.0072291815   -0.15608853  1.000000000     0.022476030
## ENERGIA_IMPACTO 0.0315518927   -0.03332637  0.022476030     1.000000000
## ENERGIA_TOTAL   0.0723942519    0.03670639  0.450430971     0.469283651
## CIRUJIA_PREVIA  0.0073002686    0.14795950 -0.061184813     0.163026114
## PIO_PRE_SLT     0.0261824909    0.06530760 -0.045190732     0.104558117
## PIO_1_SEMANA    0.0360148309    0.13035711 -0.088090444     0.006786965
## PIO_1_MES      -0.0764291803   -0.03286705  0.121378717     0.020785153
## PIO_3_MES       0.0005921436   -0.16711797  0.174664359    -0.010003597
## FARMACOS_PRE   -0.0226470426   -0.23728180  0.101592944     0.098806991
## FARMACOS_1_MES  0.0115065566   -0.25361853  0.139737434    -0.060577672
## DOLOR           0.0008585046   -0.17523363  0.031041769     0.006916478
## SEXO            0.0273059341    0.18732486 -0.111157729    -0.108242035
## EDAD           -0.0249843282   -0.24472004  0.104380092     0.043398254
## PIO_NORMAL     -0.0764291803   -0.03286705  0.121378717     0.020785153
## PIO_NORMAL_CAT -0.0910674076   -0.06503823 -0.036040315    -0.047049322
##                ENERGIA_TOTAL CIRUJIA_PREVIA  PIO_PRE_SLT PIO_1_SEMANA
## OJO              0.072394252    0.007300269  0.026182491  0.036014831
## TIPO_GLAUCOMA    0.036706392    0.147959496  0.065307602  0.130357111
## N_IMPACTOS       0.450430971   -0.061184813 -0.045190732 -0.088090444
## ENERGIA_IMPACTO  0.469283651    0.163026114  0.104558117  0.006786965
## ENERGIA_TOTAL    1.000000000   -0.011069041 -0.001359308 -0.106190549
## CIRUJIA_PREVIA  -0.011069041    1.000000000  0.157872963  0.358529456
## PIO_PRE_SLT     -0.001359308    0.157872963  1.000000000 -0.002524094
## PIO_1_SEMANA    -0.106190549    0.358529456 -0.002524094  1.000000000
## PIO_1_MES       -0.038116060    0.386637166  0.157457613  0.656624289
## PIO_3_MES       -0.059063828    0.327413593  0.127881329  0.356995689
## FARMACOS_PRE     0.052096593   -0.017276419 -0.359018173  0.243485665
## FARMACOS_1_MES  -0.091832957    0.237016562 -0.081238377  0.459330045
## DOLOR           -0.095385965    0.501685953  0.086009321  0.529142006
## SEXO             0.071383759    0.084921303  0.096288973 -0.125506479
## EDAD            -0.074554278    0.485099418  0.083414769  0.455598165
## PIO_NORMAL      -0.038116060    0.386637166  0.157457613  0.656624289
## PIO_NORMAL_CAT  -0.054781161   -0.316821239 -0.135108804 -0.408158219
##                    PIO_1_MES     PIO_3_MES FARMACOS_PRE FARMACOS_1_MES
```

```
## OJO             -0.07642918  0.0005921436  -0.02264704     0.01150656
## TIPO_GLAUCOMA   -0.03286705 -0.1671179707  -0.23728180    -0.25361853
## N_IMPACTOS       0.12137872  0.1746643588   0.10159294     0.13973743
## ENERGIA_IMPACTO  0.02078515 -0.0100035968   0.09880699    -0.06057767
## ENERGIA_TOTAL   -0.03811606 -0.0590638279   0.05209659    -0.09183296
## CIRUJIA_PREVIA   0.38663717  0.3274135930  -0.01727642     0.23701656
## PIO_PRE_SLT      0.15745761  0.1278813285  -0.35901817    -0.08123838
## PIO_1_SEMANA     0.65662429  0.3569956894   0.24348566     0.45933005
## PIO_1_MES        1.00000000  0.6082639571   0.21962717     0.53786628
## PIO_3_MES        0.60826396  1.0000000000   0.07659788     0.44253237
## FARMACOS_PRE     0.21962717  0.0765978785   1.00000000     0.59226275
## FARMACOS_1_MES   0.53786628  0.4425323708   0.59226275     1.00000000
## DOLOR            0.79161083  0.6099276923   0.30152167     0.66213336
## SEXO            -0.10278748 -0.0146488596  -0.16407024    -0.27762428
## EDAD             0.75313605  0.5587507386   0.30534137     0.65487613
## PIO_NORMAL       1.00000000  0.6082639571   0.21962717     0.53786628
## PIO_NORMAL_CAT  -0.68476343 -0.4162236728  -0.13735098    -0.30338490
##                         DOLOR         SEXO        EDAD  PIO_NORMAL
## OJO              0.0008585046  0.027305934 -0.02498433 -0.07642918
## TIPO_GLAUCOMA   -0.1752336263  0.187324861 -0.24472004 -0.03286705
## N_IMPACTOS       0.0310417691 -0.111157729  0.10438009  0.12137872
## ENERGIA_IMPACTO  0.0069164781 -0.108242035  0.04339825  0.02078515
## ENERGIA_TOTAL   -0.0953859650  0.071383759 -0.07455428 -0.03811606
## CIRUJIA_PREVIA   0.5016859532  0.084921303  0.48509942  0.38663717
## PIO_PRE_SLT      0.0860093213  0.096288973  0.08341477  0.15745761
## PIO_1_SEMANA     0.5291420063 -0.125506479  0.45559816  0.65662429
## PIO_1_MES        0.7916108300 -0.102787485  0.75313605  1.00000000
## PIO_3_MES        0.6099276923 -0.014648860  0.55875074  0.60826396
## FARMACOS_PRE     0.3015216686 -0.164070244  0.30534137  0.21962717
## FARMACOS_1_MES   0.6621333589 -0.277624278  0.65487613  0.53786628
## DOLOR            1.0000000000  0.003252437  0.93188658  0.79161083
## SEXO             0.0032524367  1.000000000 -0.05705509 -0.10278748
## EDAD             0.9318865791 -0.057055091  1.00000000  0.75313605
## PIO_NORMAL       0.7916108300 -0.102787485  0.75313605  1.00000000
## PIO_NORMAL_CAT  -0.6381159167 -0.092111666 -0.56855536 -0.68476343
##                 PIO_NORMAL_CAT
## OJO                -0.09106741
## TIPO_GLAUCOMA      -0.06503823
## N_IMPACTOS         -0.03604032
## ENERGIA_IMPACTO    -0.04704932
## ENERGIA_TOTAL      -0.05478116
## CIRUJIA_PREVIA     -0.31682124
## PIO_PRE_SLT        -0.13510880
## PIO_1_SEMANA       -0.40815822
## PIO_1_MES          -0.68476343
## PIO_3_MES          -0.41622367
## FARMACOS_PRE       -0.13735098
## FARMACOS_1_MES     -0.30338490
## DOLOR              -0.63811592
## SEXO               -0.09211167
## EDAD               -0.56855536
## PIO_NORMAL         -0.68476343
## PIO_NORMAL_CAT      1.00000000
```

```r
det(correlation_matrix)
```

```
## [1] 0
```

```r
#####################################
# Correlation at population level #
#####################################

# Bartlett's sphericity test:
# This test checks whether the correlations are significantly different from 0
# The null hypothesis is H_0; det(R)=1 means the variables are uncorrelated
# R denotes the correlation matrix
# cortest.bartlett function in the package pysch performs this test
# This function works with standardized data.

# Standardization
data_pca_xlsx_scale<-scale(data_xlsx_aux)

# Bartlett's sphericity test
cortest.bartlett(cor(data_pca_xlsx_scale))
```

```
## $chisq
## [1] Inf
##
## $p.value
## [1] 0
##
## $df
## [1] 136
```

Absence of outliers

Done in **Section 2.4.2** in the data.frame *data_xlsx_aux*.

Standardized data

It is not necessary, since the *prcomp* function that obtains the principal components standardizes the data on its own.

Principal components

Obtaining

```r
# The 'prcomp' function in the base R package performs this analysis
# Parameters 'scale' and 'center' are set to TRUE to consider standardized data
PCA<-prcomp(data_xlsx_aux, scale = T, center = T)

# The field 'rotation' of the 'PCA' object is a matrix
# Its columns are the coefficients of the principal components
# Indicates the weight of each variable in the corresponding principal component
PCA$rotation
```

```
##                        PC1          PC2          PC3          PC4
## OJO             0.005832023  0.007306639 -0.109978222  0.007942571
## TIPO_GLAUCOMA   0.056716044  0.408736083 -0.033985311  0.478163290
## N_IMPACTOS     -0.047151620 -0.284640556 -0.410735816 -0.349571462
```

```
## ENERGIA_IMPACTO -0.016651070 -0.055515410 -0.517085076  0.332927439
## ENERGIA_TOTAL    0.025962150 -0.128496082 -0.671079627  0.115953414
## CIRUJIA_PREVIA  -0.214720335  0.266655339 -0.100069817  0.192584421
## PIO_PRE_SLT     -0.043665587  0.389661822 -0.179224155 -0.366215009
## PIO_1_SEMANA    -0.283095888  0.074929577  0.109252946  0.357444285
## PIO_1_MES       -0.384101787  0.081593912 -0.026816852 -0.016604119
## PIO_3_MES       -0.288279630  0.041444966 -0.037108929 -0.309360663
## FARMACOS_PRE    -0.149947474 -0.483257787  0.072329945  0.332814806
## FARMACOS_1_MES  -0.300109657 -0.316120985  0.136167973  0.035608165
## DOLOR           -0.385543210  0.021127892  0.036232373 -0.052824565
## SEXO             0.045518446  0.348352129 -0.071257346 -0.054448993
## EDAD            -0.371415871 -0.036062145  0.006666628 -0.101365183
## PIO_NORMAL      -0.384101787  0.081593912 -0.026816852 -0.016604119
## PIO_NORMAL_CAT   0.293310613 -0.179200098  0.115004475 -0.039025838
##                           PC5         PC6          PC7          PC8
## OJO             -0.4484909710  0.72920108 -0.455941079  0.078093221
## TIPO_GLAUCOMA   -0.1154131030 -0.33761751 -0.284157146  0.162755871
## N_IMPACTOS      -0.1932471356 -0.36237320 -0.244057760  0.306286037
## ENERGIA_IMPACTO  0.3872668814  0.27136940  0.219636592 -0.154569404
## ENERGIA_TOTAL   -0.1271364462 -0.12494971  0.005584447 -0.040981876
## CIRUJIA_PREVIA   0.0925166914  0.14978648  0.263076970  0.688962870
## PIO_PRE_SLT      0.3650848278  0.19101531 -0.085207452 -0.201693099
## PIO_1_SEMANA     0.0351712563 -0.03969276 -0.299966241  0.006409954
## PIO_1_MES        0.0476879248 -0.15541284 -0.146020259 -0.180354327
## PIO_3_MES       -0.0164870727 -0.01672134  0.001124275  0.192704811
## FARMACOS_PRE    -0.1524295860  0.01901875  0.191536071 -0.177151411
## FARMACOS_1_MES   0.0009521849  0.07376171  0.001585313  0.132555410
## DOLOR           -0.0645585441  0.08608671  0.183306846  0.024840896
## SEXO            -0.5969037597 -0.05838134  0.523213838 -0.139988247
## EDAD            -0.0034441944  0.08829121  0.212076103  0.069129533
## PIO_NORMAL       0.0476879248 -0.15541284 -0.146020259 -0.180354327
## PIO_NORMAL_CAT   0.2245864136  0.01451516  0.073642546  0.395603152
##                          PC9         PC10         PC11         PC12        PC13
## OJO             -0.007677210  0.007131246 -0.018888823  0.001300989 -0.13733972
## TIPO_GLAUCOMA    0.232376734 -0.038258177 -0.461770326 -0.151754287 -0.24980049
## N_IMPACTOS       0.135012497 -0.090057665  0.098100509  0.346922070 -0.29553176
## ENERGIA_IMPACTO -0.187821360  0.159377982 -0.140043247  0.144401761 -0.38785703
## ENERGIA_TOTAL    0.042522997  0.036355711  0.129917013 -0.474466313  0.43538661
## CIRUJIA_PREVIA   0.017676269 -0.209997435  0.126699450  0.297077696  0.26707692
## PIO_PRE_SLT      0.646012561  0.065923041 -0.011993019  0.104824548  0.09770397
## PIO_1_SEMANA     0.021563464  0.423946322  0.534367361  0.058619794  0.11643023
## PIO_1_MES       -0.085863404  0.012826394  0.116564431  0.045494526 -0.15096734
## PIO_3_MES       -0.298308206  0.595712404 -0.513766092  0.044140211  0.19760360
## FARMACOS_PRE     0.378748845  0.074517375 -0.179610315  0.407728058  0.08289505
## FARMACOS_1_MES   0.411547953  0.075640837 -0.184544304 -0.312256098  0.08626037
## DOLOR           -0.001454837 -0.150907260 -0.009870329 -0.280996429 -0.16045714
## SEXO             0.151915755  0.307458750  0.162537769  0.056891376 -0.16538486
## EDAD             0.013623848 -0.251214791  0.043684667 -0.289590107 -0.31137555
## PIO_NORMAL      -0.085863404  0.012826394  0.116564431  0.045494526 -0.15096734
## PIO_NORMAL_CAT   0.180824908  0.434266403  0.229523794 -0.268462834 -0.38467712
##                        PC14        PC15         PC16          PC17
## OJO              0.14575909 -0.01544791 -0.008960873  6.839783e-18
## TIPO_GLAUCOMA    0.01361048  0.10384667 -0.016998742 -8.732575e-18
## N_IMPACTOS      -0.23511048  0.01933879  0.067163249  4.388313e-17
```

```
## ENERGIA_IMPACTO -0.21777819 -0.15100768  0.029466709  2.767193e-16
## ENERGIA_TOTAL    0.20572966  0.08888695 -0.024922236 -2.905199e-16
## CIRUJIA_PREVIA   0.16214075 -0.12030187 -0.001095199  5.144961e-17
## PIO_PRE_SLT      0.01188729  0.14058943  0.006722876  7.950915e-17
## PIO_1_SEMANA    -0.38454875  0.22231046 -0.040870374  3.983542e-17
## PIO_1_MES        0.39704109 -0.24678604 -0.002519614  7.071068e-01
## PIO_3_MES        0.01920069  0.16398387 -0.065853413  1.021615e-16
## FARMACOS_PRE     0.29642947  0.30669846 -0.001907977  1.581782e-16
## FARMACOS_1_MES  -0.29209083 -0.60494358 -0.042527015 -5.787184e-17
## DOLOR           -0.08915439  0.29652766  0.759970028  2.692675e-17
## SEXO            -0.06746004 -0.18670452 -0.038192042  5.793660e-17
## EDAD            -0.07031386  0.37326407 -0.634490315 -2.156920e-16
## PIO_NORMAL       0.39704109 -0.24678604 -0.002519614 -7.071068e-01
## PIO_NORMAL_CAT   0.39108642  0.05837616  0.064506298 -4.569891e-17
```

```
# Standard deviations of each principal component
PCA$sdev
```

```
##  [1] 2.400588e+00 1.389594e+00 1.302246e+00 1.089832e+00 1.025068e+00
##  [6] 1.023278e+00 9.706195e-01 8.582354e-01 7.814884e-01 7.222847e-01
## [11] 6.664250e-01 5.707607e-01 5.458230e-01 4.732526e-01 4.105179e-01
## [16] 2.310236e-01 4.788608e-17
```

Each principal component is obtained in a simple way as a linear combination of all the variables with the coefficients indicated by the columns of the rotation matrix.

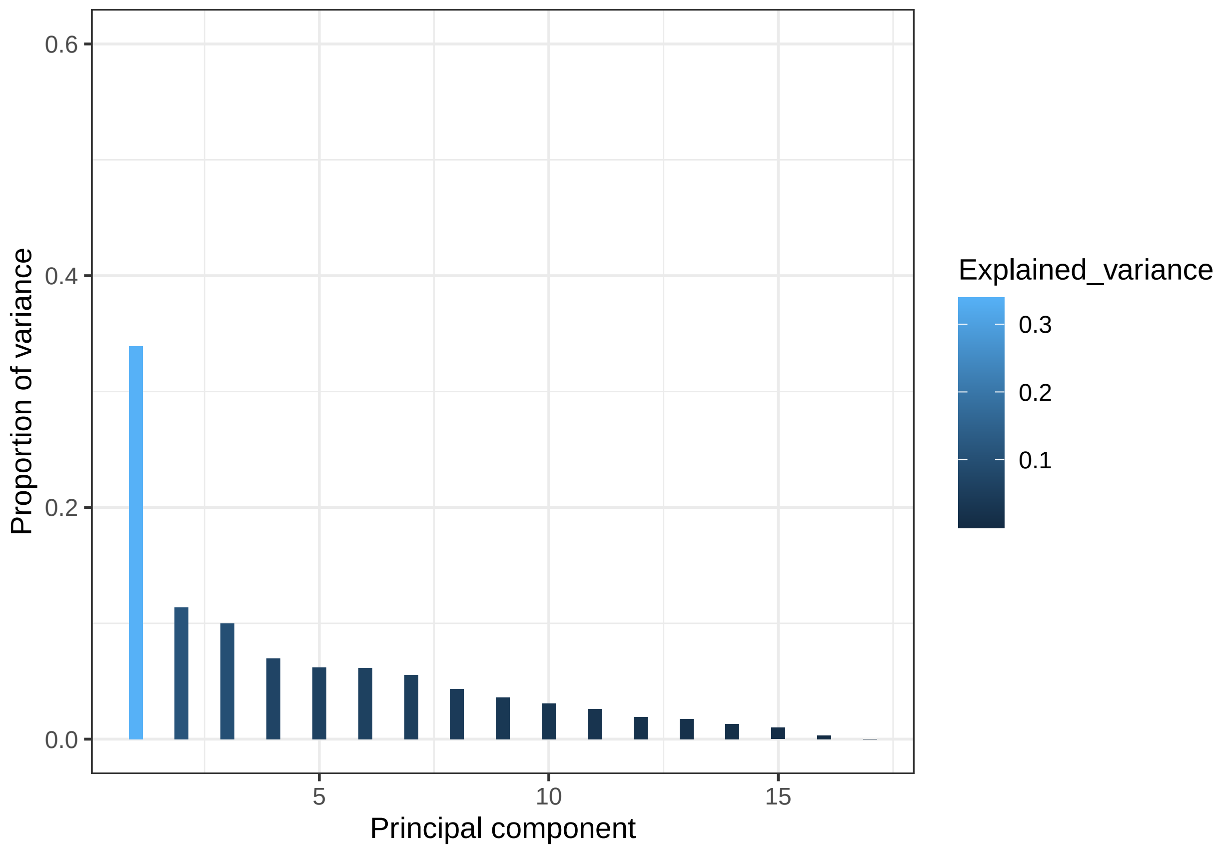
Explained variance rate

```
# The function 'summary' applied to the 'PCA' object provides relevant information
# - Standard deviations of each principal component
# - Proportion of variance explained and cummulative variance
summary(PCA)
```

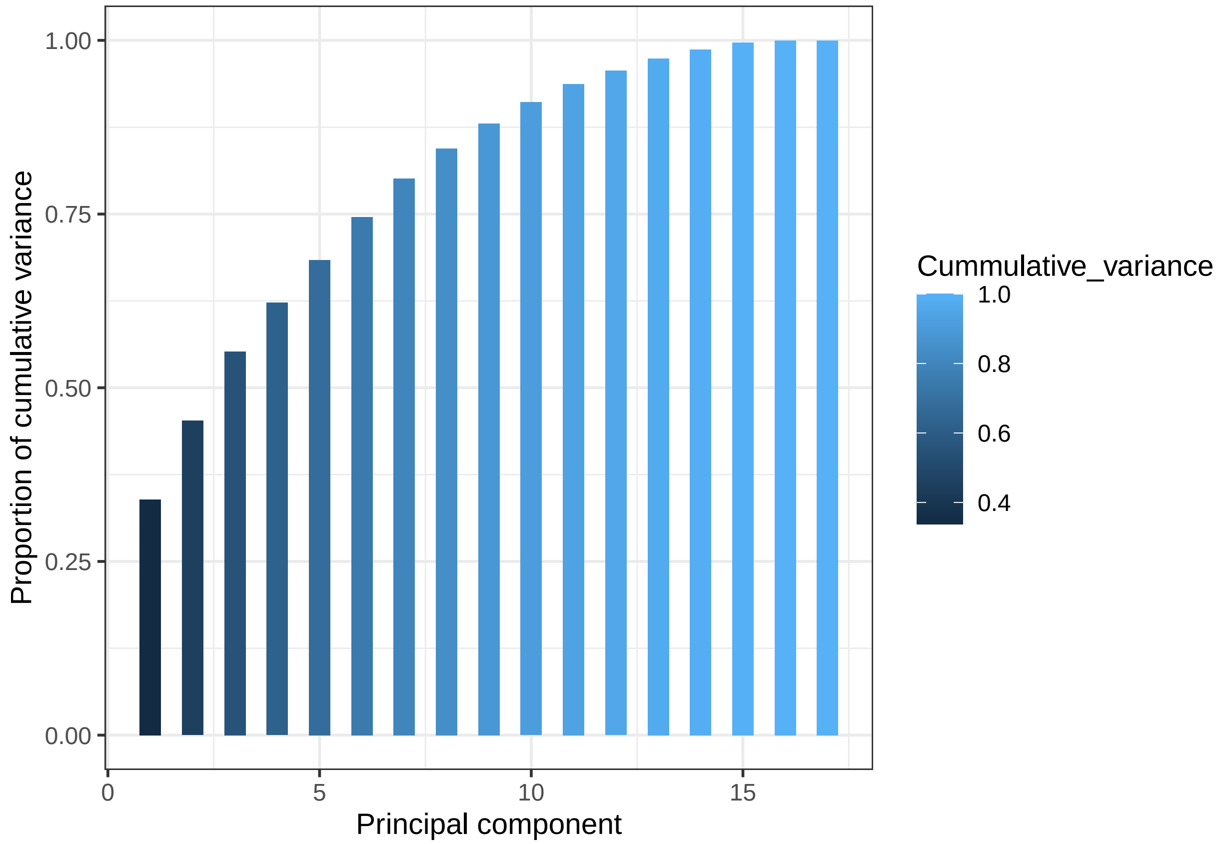
```
## Importance of components:
##                          PC1    PC2     PC3     PC4     PC5     PC6     PC7
## Standard deviation     2.401 1.3896 1.30225 1.08983 1.02507 1.02328 0.97062
## Proportion of Variance 0.339 0.1136 0.09976 0.06987 0.06181 0.06159 0.05542
## Cumulative Proportion  0.339 0.4526 0.55233 0.62220 0.68401 0.74560 0.80102
##                            PC8     PC9    PC10    PC11    PC12    PC13    PC14
## Standard deviation     0.85824 0.78149 0.72228 0.66642 0.57076 0.54582 0.47325
## Proportion of Variance 0.04333 0.03592 0.03069 0.02612 0.01916 0.01752 0.01317
## Cumulative Proportion  0.84435 0.88027 0.91096 0.93708 0.95625 0.97377 0.98695
##                           PC15    PC16      PC17
## Standard deviation     0.41052 0.23102 4.789e-17
## Proportion of Variance 0.00991 0.00314 0.000e+00
## Cumulative Proportion  0.99686 1.00000 1.000e+00
```

```
# The following graph shows the proportion of explained variance
Explained_variance <- PCA$sdev^2 / sum(PCA$sdev^2)

p1<-ggplot(data = data.frame(Explained_variance, pc = 1:17),
  aes(x = pc, y = Explained_variance, fill=Explained_variance )) +
  geom_col(width = 0.3) + scale_y_continuous(limits = c(0,0.6)) + theme_bw() +
  labs(x = "Principal component", y= "Proportion of variance")

# The following graph shows the proportion of cumulative explained variance
```

```r
Cummulative_variance<-cumsum(Explained_variance)

p2<-ggplot( data = data.frame(Cummulative_variance, pc = 1:17),
  aes(x = pc, y = Cummulative_variance ,fill=Cummulative_variance )) +
  geom_col(width = 0.5) +  scale_y_continuous(limits = c(0,1)) + theme_bw() +
  labs(x = "Principal component",
       y = "Proportion of cumulative variance")

p1
```




```r
p2
```



Appropriate number of principal components

There are different methods:

- 1.- **Elbow method** (Cuadras, 2007).
- 2.- **At the discretion of the researcher** who chooses a minimum percentage of variance explained by the principal components (it is not reliable because it can give more than necessary).
- 3.- **Rule of Abdi et al.** (2010). The variances explained by the principal components are averaged and those whose proportion of explained variance exceeds the mean are selected.

For this illustration, applying the rule of Abdi et al., only **six principal components are considered**, as can be deduced from the following code chunk.

```
########################
# Rule of Abdi et al. #
########################

# Variances
PCA$sdev^2
```

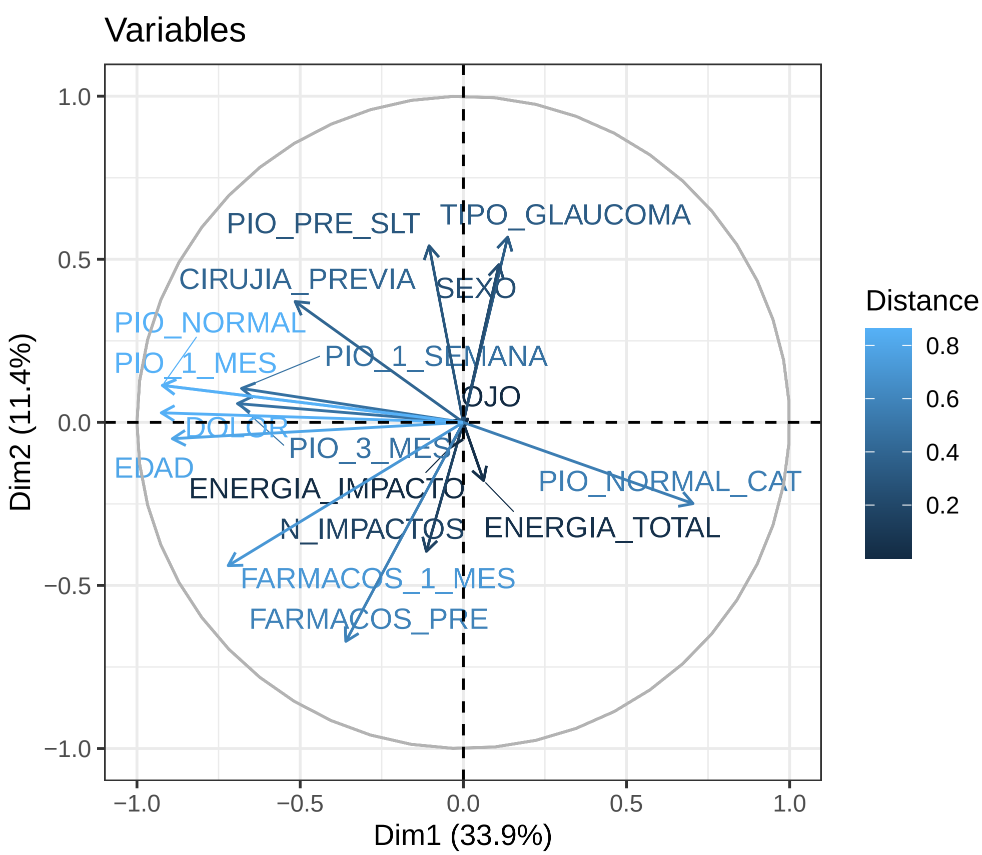
```
##  [1] 5.762821e+00 1.930972e+00 1.695845e+00 1.187734e+00 1.050763e+00
##  [6] 1.047097e+00 9.421021e-01 7.365680e-01 6.107241e-01 5.216951e-01
## [11] 4.441223e-01 3.257678e-01 2.979227e-01 2.239680e-01 1.685249e-01
## [16] 5.337190e-02 2.293077e-33
```

```
# Average of variances
mean(PCA$sdev^2)
```

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

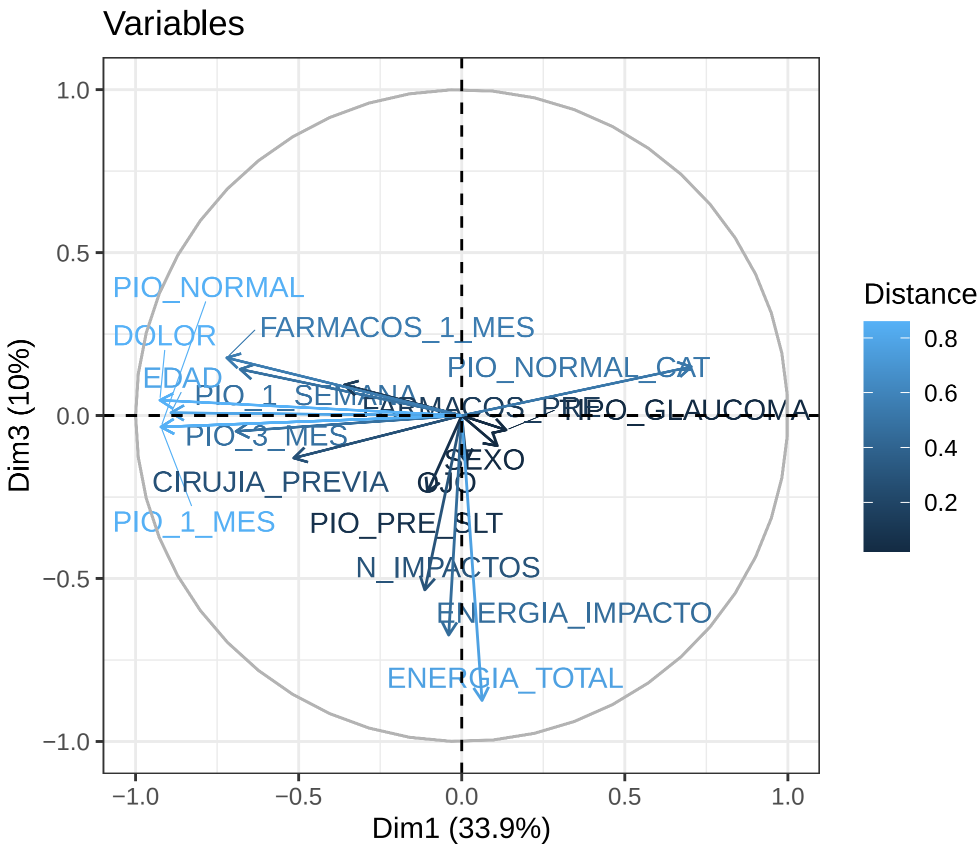
PCA graphical outputs of interest

```
# These graphical outputs show the projection of the variables in two dimensions
# Display the weight of the variable in the direction of the principal component
p1<-fviz_pca_var(PCA,repel=TRUE,col.var="cos2",
                 legend.title="Distance", title="Variables")+theme_bw()

p2<-fviz_pca_var(PCA,axes=c(1,3),repel=TRUE,col.var="cos2",
                 legend.title="Distance", title="Variables")+theme_bw()

p3<-fviz_pca_var(PCA,axes=c(2,3),repel=TRUE,col.var="cos2",
                 legend.title="Distance", title="Variables")+theme_bw()

# Displaying graphics
p1
```

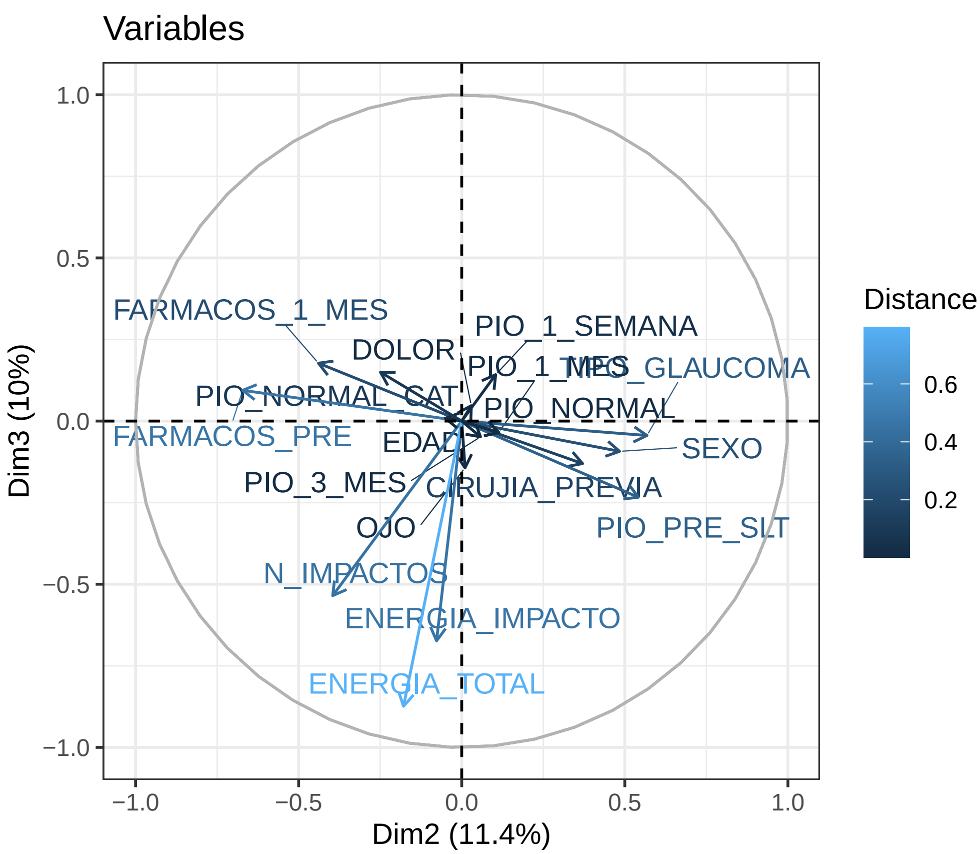


Distances

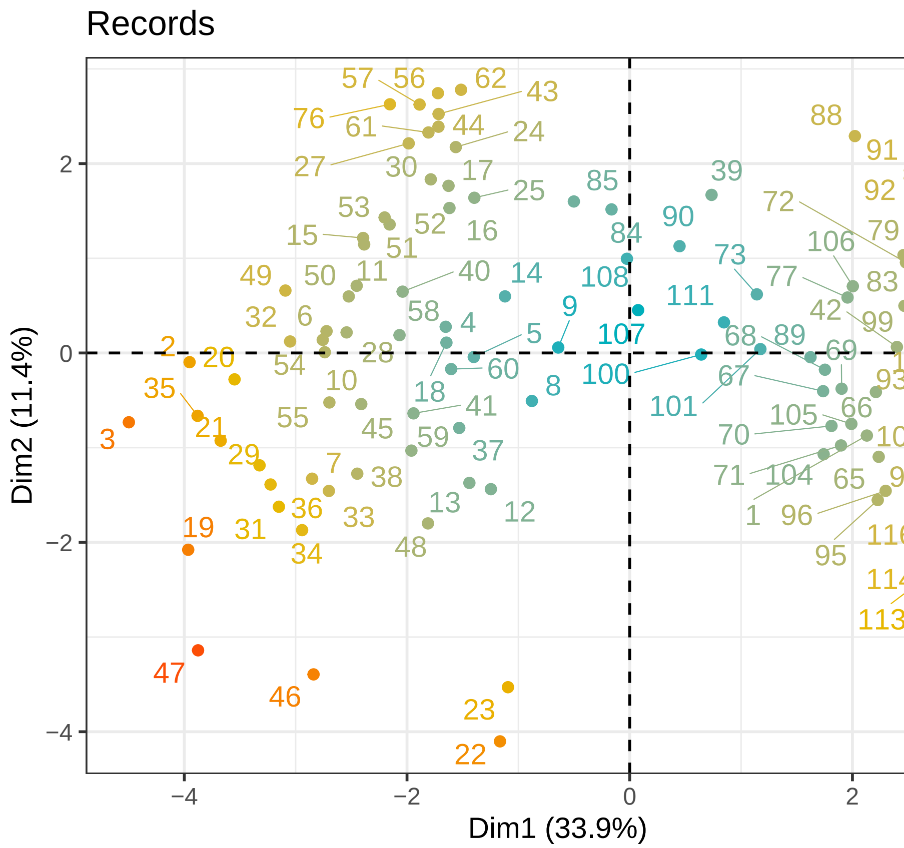
```
p2
```


Variables
Distance
0.8
0.6
0.4
0.2
Dim3 (10%)
Dim1 (33.9%)
PIO_NORMAL
DOLOR
EDAD
FARMACOS_1_MES
PIO_1_SEMANA
FARMACOS_PRE
PIO_NORMAL_CAT
TIPO_GLAUCOMA
PIO_3_MES
SEXO
CIRUJIA_PREVIA
OJO
PIO_1_MES
PIO_PRE_SLT
N_IMPACTOS
ENERGIA_IMPACTO
ENERGIA_TOTAL

```
p3
```

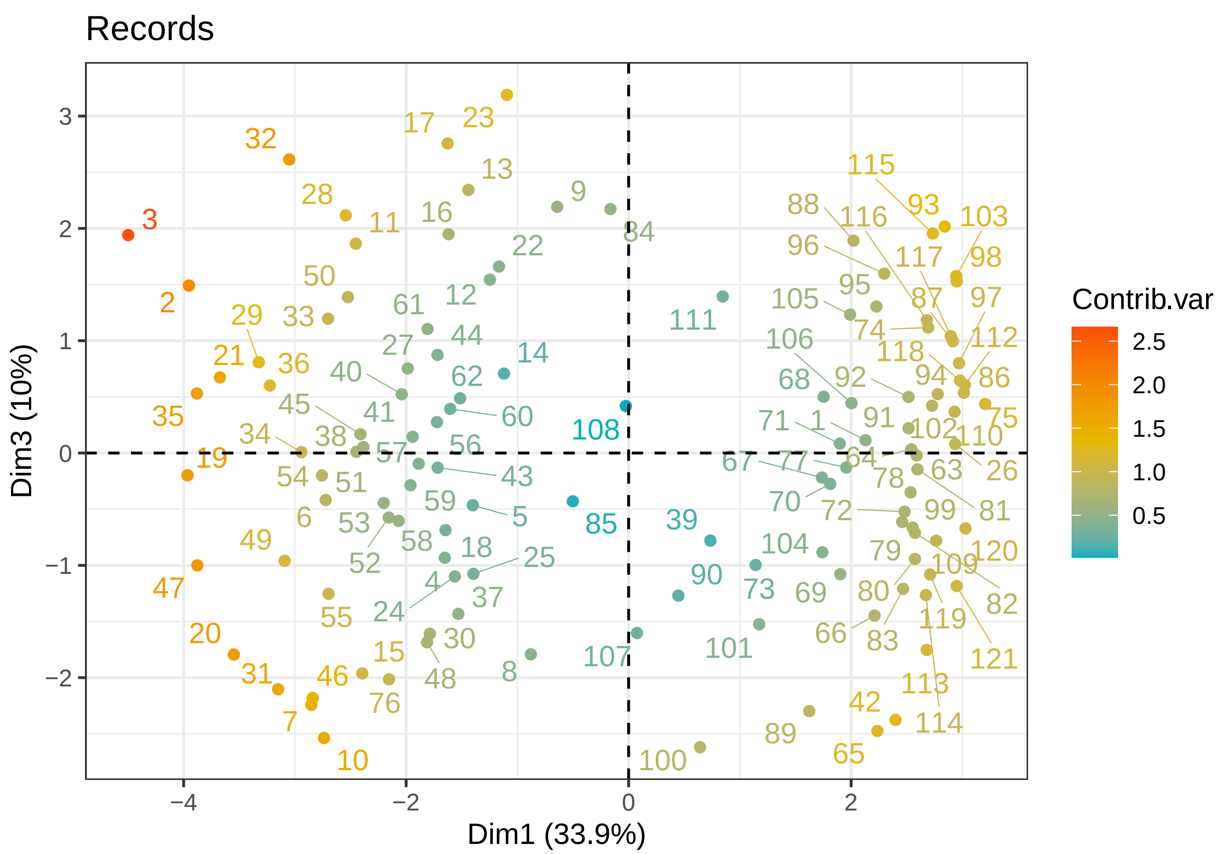



```r
# It is also possible to represent the observations
# As well as identify with colors those observations that explain the greatest
# variance of the principal components
p1<-fviz_pca_ind(PCA,col.ind = "contrib",
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
             repel=TRUE,legend.title="Contrib.var", title="Records")+theme_bw()

p2<-fviz_pca_ind(PCA,axes=c(1,3),col.ind = "contrib",
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
             repel=TRUE,legend.title="Contrib.var", title="Records")+theme_bw()

p3<-fviz_pca_ind(PCA,axes=c(2,3),col.ind = "contrib",
             gradient.cols = c("#00AFBB", "#E7B800", "#FC4E07"),
             repel=TRUE,legend.title="Contrib.var", title="Records")+theme_bw()

# Displaying graphics
p1
```

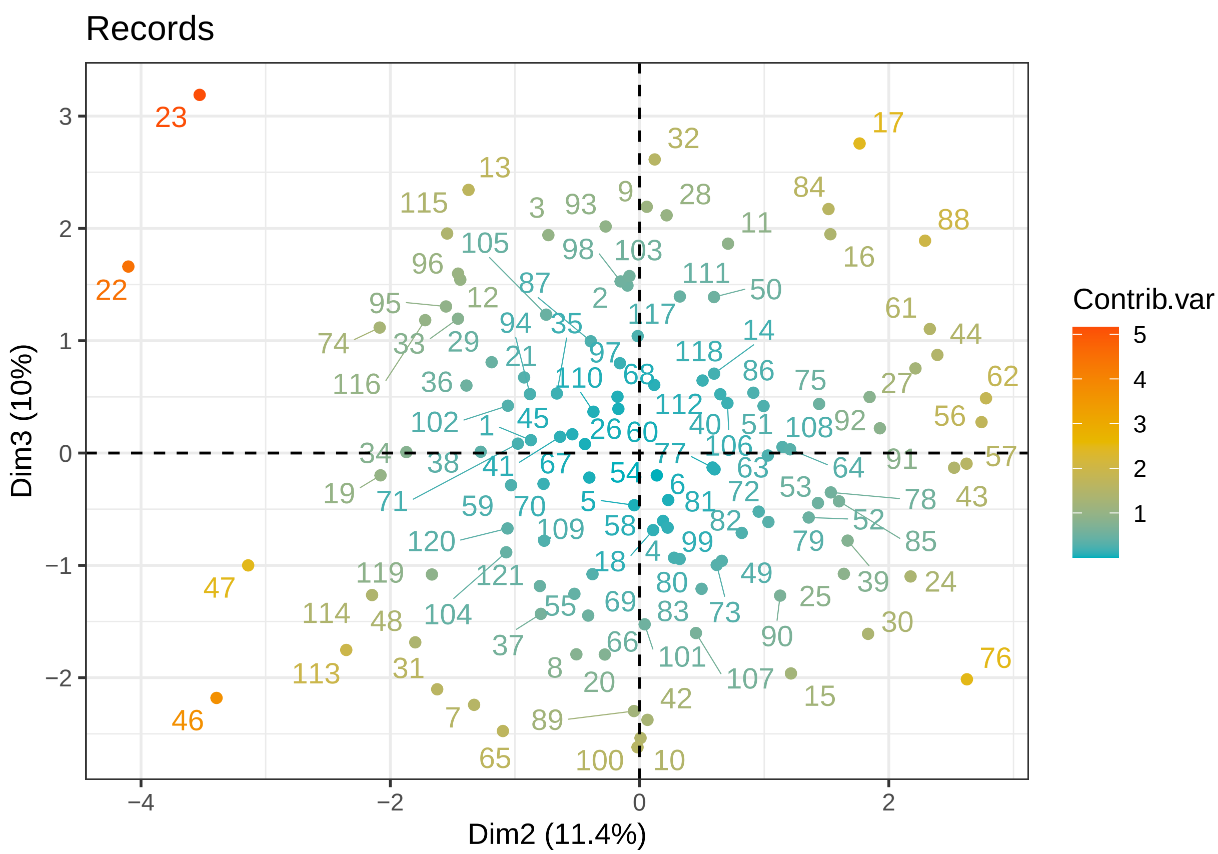


Observations and variance contribution

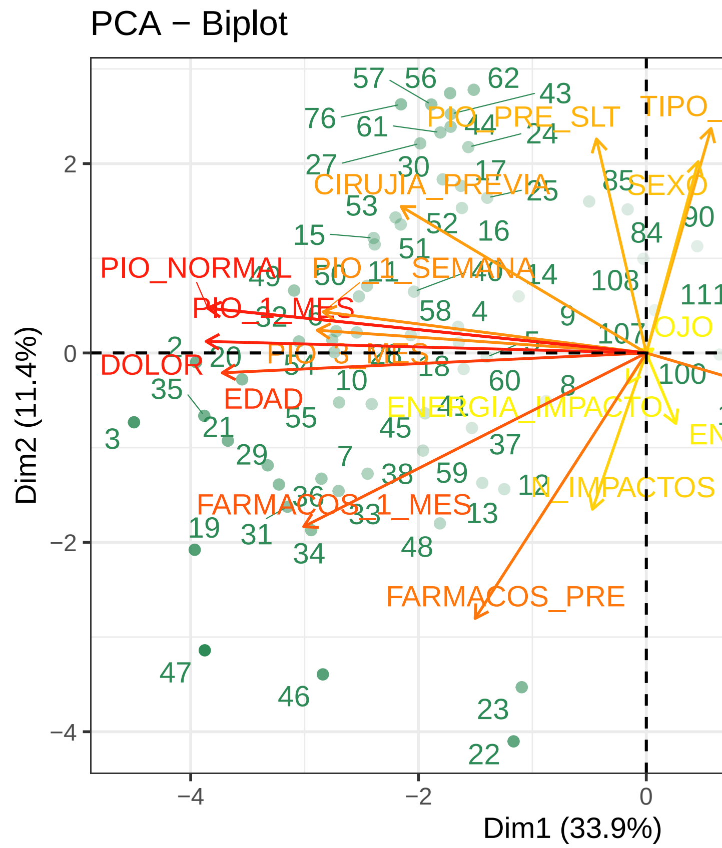
```
p2
```




```
p3
```

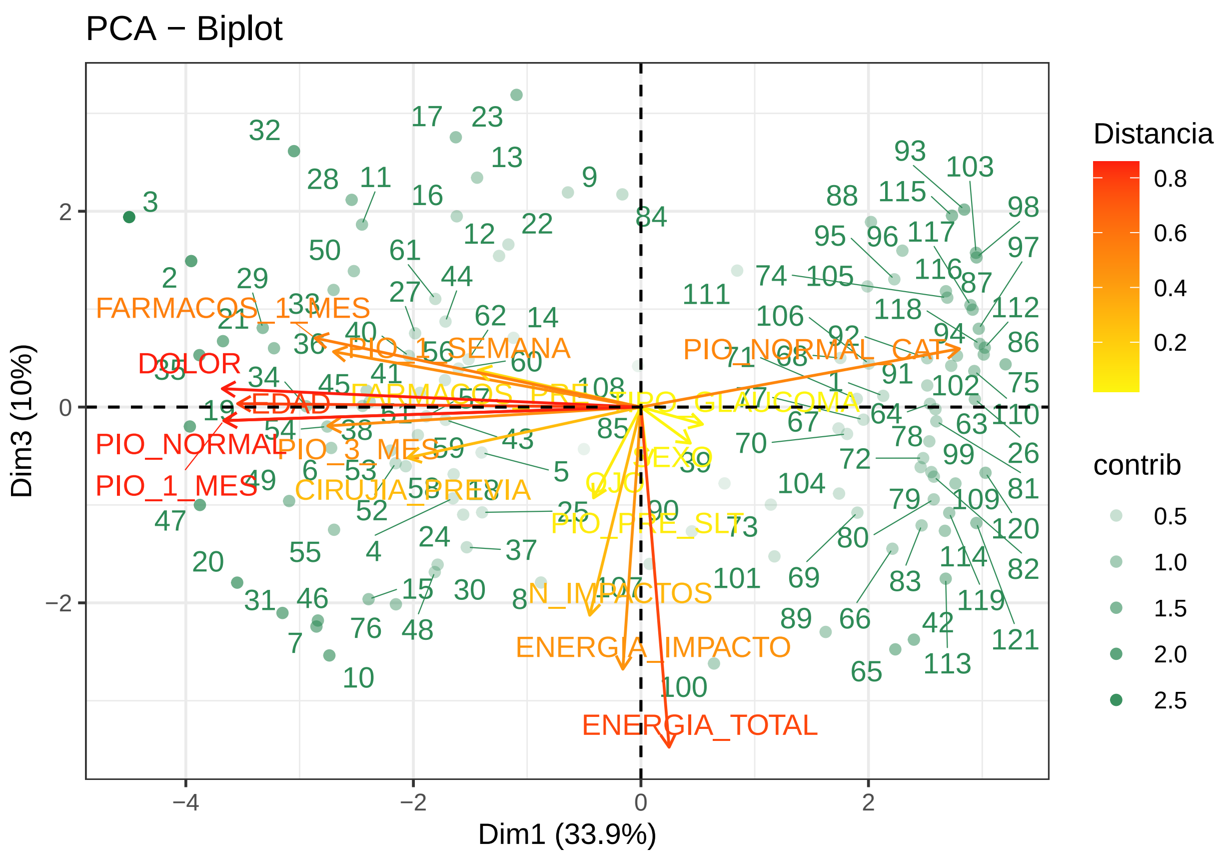



```r
# Joint representation of variables and observations
# Relates the possible relationships between the contributions of the records
# to the variances of the components and the weight of the variables in each
# principal component

p1<-fviz_pca(PCA,alpha.ind ="contrib", col.var = "cos2",
         col.ind="seagreen",
         gradient.cols = c("#FDF50E", "#FD960E", "#FD1E0E"),
         repel=TRUE, legend.title="Distancia")+theme_bw()

p2<-fviz_pca(PCA,axes=c(1,3),alpha.ind ="contrib",
         col.var = "cos2",col.ind="seagreen",
         gradient.cols = c("#FDF50E", "#FD960E", "#FD1E0E"),
         repel=TRUE, legend.title="Distancia")+theme_bw()

p3<-fviz_pca(PCA,axes=c(2,3),alpha.ind ="contrib",
         col.var = "cos2",col.ind="seagreen",
         gradient.cols = c("#FDF50E", "#FD960E", "#FD1E0E"),
         repel=TRUE, legend.title="Distancia")+theme_bw()

# Displaying graphics
p1
```



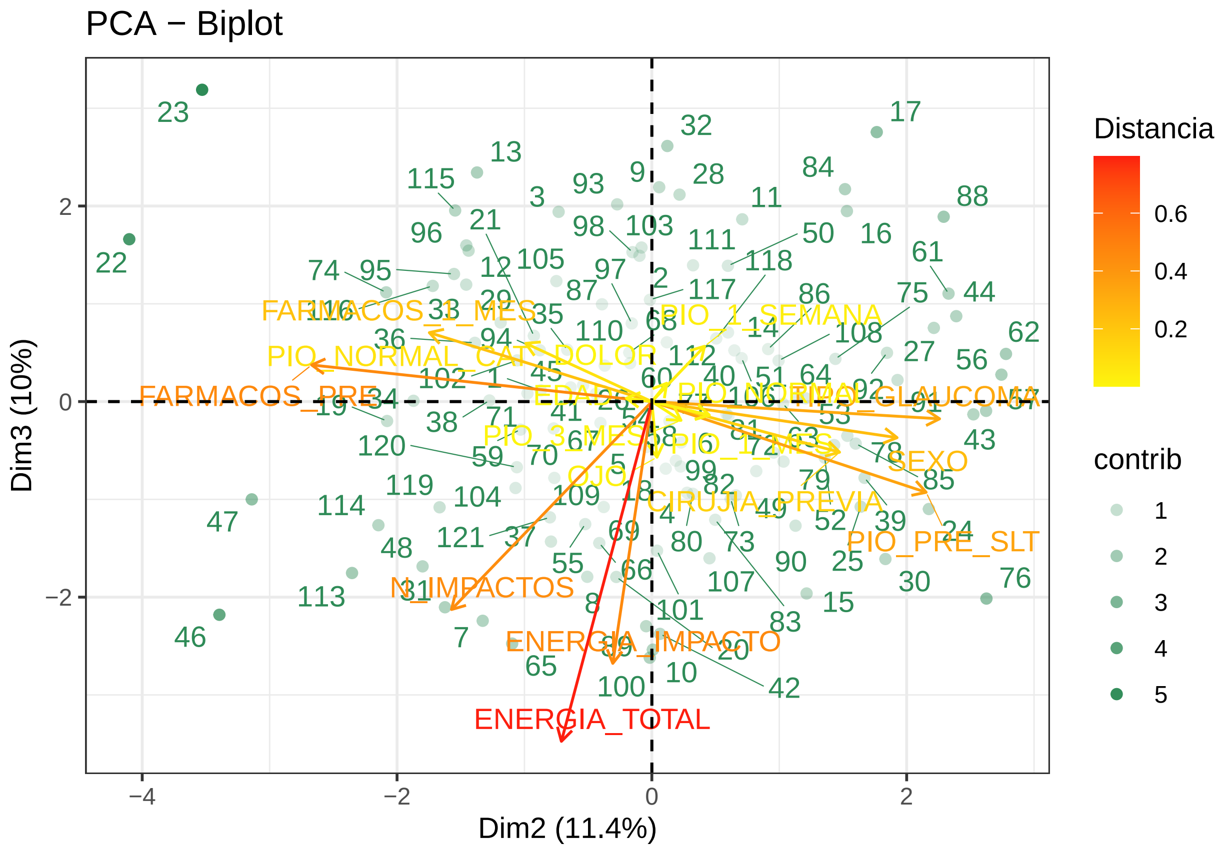
Observations and variables with variance contribution

```
p2
```



PCA − Biplot
Dim1 (33.9%)
Dim3 (10%)
Distancia
0.8
0.6
0.4
0.2
contrib
0.5
1.0
1.5
2.0
2.5
FARMACOS_1_MES
DOLOR
EDAD
PIO_NORMAL
PIO_1_MES
PIO_1_SEMANA
PIO_3_MES
CIRUJIA_PREVIA
PIO_NORMAL_CAT
SEXO
OJO
PIO_PRE_SLT
N_IMPACTOS
ENERGIA_IMPACTO
ENERGIA_TOTAL


```
p3
```



Coordinates in the new reference system Finally, since the object of this study was to reduce the dimension of the data set, it is possible to obtain **the coordinates of the original data in the new reference system**. In fact, they are stored since we used the prcomp function to create the PCA variable.

```
head(PCA$x)
```

```
##            PC1         PC2        PC3        PC4        PC5        PC6
## [1,]  2.129577 -0.87208265  0.1142666 -1.2386028  1.0463051  0.2085975
## [2,] -3.952605 -0.09705269  1.4913943 -0.3324993  0.4638472  1.0881772
## [3,] -4.497718 -0.73171203  1.9404004 -1.3707376  1.5302448 -0.9823441
## [4,] -1.652418  0.27664066 -0.9322884 -0.7348660 -2.0761751  0.1310715
## [5,] -1.400720 -0.04264081 -0.4646457 -1.6096928 -1.8139190 -1.8467109
## [6,] -2.722590  0.23010337 -0.4180876 -2.2892810 -1.0733149  1.1780603
##             PC7        PC8        PC9       PC10        PC11       PC12
## [1,]  1.0535308 -1.5115916  1.4831245  0.2949518 -0.05375524  0.3118375
## [2,] -1.6372462  0.5312332  1.2850596  0.6212151 -1.54149254  0.1576738
## [3,] -1.1688823  1.4877960 -0.3773068  1.9781899  0.27258651 -0.6531470
## [4,]  0.6924509  0.2068482 -2.0919314 -0.5645371 -0.20718058  0.1286191
## [5,]  1.2689228  0.8241372 -1.7558167 -0.6525042 -0.25292302  0.4422838
## [6,]  0.7057618 -0.5567810  1.0624655  0.2326776 -1.21590663 -0.2328782
##            PC13       PC14        PC15        PC16          PC17
## [1,]  0.4425755  0.2449885  0.98409415  0.78783187  2.922102e-16
## [2,]  1.0920922 -0.1181418  0.10878404 -0.11504673  2.727372e-16
## [3,]  0.5230929  0.4886160 -0.76718950 -0.00375731 -2.629013e-16
## [4,] -0.3593564  1.0491916 -0.45008011  0.28360068 -1.713520e-16
## [5,]  0.1216017  0.2524540  0.06133612  0.33124610 -2.823743e-16
```

```
## [6,]  0.0860834  0.2360972 -0.54468570 -0.34908334  5.069257e-17
```