A4-Componentes Principales

Eliezer Cavazos

2024-10-08

```
oCorporal =
read.csv("C:\\Users\\eliez\\OneDrive\\Desktop\\Clases\\corporal.csv") #leer
la base de datos

oCorporal = oCorporal[, -4]
oCorporal
```

```
##    edad peso altura muneca biceps
## 1    43 87.3  188.0   12.2   35.8
## 2    65 80.0  174.0   12.0   35.0
## 3    45 82.3  176.5   11.2   38.5
## 4    37 73.6  180.3   11.2   32.2
## 5    55 74.1  167.6   11.8   32.9
## 6    33 85.9  188.0   12.4   38.5
## 7    25 73.2  180.3   10.6   38.3
## 8    35 76.3  167.6   11.3   35.0
## 9    28 65.9  183.0   10.2   32.1
## 10   26 90.9  183.0   12.0   40.4
## 11   43 89.1  179.1   11.3   36.5
## 12   30 62.3  170.2   11.5   34.2
## 13   26 82.7  177.8   11.5   35.2
## 14   51 79.1  179.1   11.8   34.0
## 15   30 98.2  190.5   10.7   34.8
## 16   24 84.1  177.8   11.5   38.6
## 17   35 83.2  180.3   11.1   36.4
## 18   37 83.2  180.3   10.5   34.0
## 19   22 51.6  161.2    9.2   24.3
## 20   20 59.0  167.5    9.9   27.8
## 21   19 49.2  159.5    8.9   24.0
## 22   25 63.0  157.0    9.5   28.0
## 23   21 53.6  155.8    9.1   26.9
## 24   23 59.0  170.0   10.0   26.5
## 25   26 47.6  159.1    9.4   24.1
## 26   22 69.8  166.0   10.7   29.2
## 27   28 66.8  176.2    9.8   29.0
## 28   40 75.2  160.2   11.5   33.6
## 29   32 55.2  172.5    8.6   24.8
## 30   25 54.2  170.9    9.7   25.4
## 31   25 62.5  172.9    9.2   25.9
## 32   29 42.0  153.4    8.3   24.0
## 33   22 50.0  160.0    8.6   25.6
## 34   25 49.8  147.2    9.0   26.0
```

```
## 35   23 49.2  168.2    9.6   23.5
## 36   37 73.2  175.0   11.0   31.0
```

Parte 1

1. Calcule las matrices de varianza-covarianza S con cov(X) y la matriz de correlaciones R con cor(X) y realice los siguientes pasos con cada una:

```
cov(oCorporal)

##              edad      peso     altura    muneca    biceps
## edad   111.396825  80.88159  36.666032  7.698095 26.720952
## peso    80.881587 221.08713 124.728698 14.844667 70.738381
## altura  36.666032 124.72870 110.673968  8.156476 39.021048
## muneca   7.698095  14.84467   8.156476  1.381714  5.400571
## biceps  26.720952  70.73838  39.021048  5.400571 27.398857

cor(oCorporal)

##             edad      peso    altura    muneca    biceps
## edad   1.0000000 0.5153847 0.3302211 0.6204942 0.4836702
## peso   0.5153847 1.0000000 0.7973737 0.8493361 0.9088813
## altura 0.3302211 0.7973737 1.0000000 0.6595849 0.7086144
## muneca 0.6204942 0.8493361 0.6595849 1.0000000 0.8777369
## biceps 0.4836702 0.9088813 0.7086144 0.8777369 1.0000000
```

1.1 Calcule los valores y vectores propios de cada matriz.La función en R es: eigen().

Covarianza:

```
oCovEigen = eigen(cov(oCorporal))

oCovEigenValues = oCovEigen$values
oCovEigenVectors = oCovEigen$vectors
```

Correlacion:

```
oCorEigen = eigen(cor(oCorporal))

oCorEigenValues = oCorEigen$values
oCorEigenVectors = oCorEigen$vectors
```

1.2 Calcule la proporción de varianza explicada por cada componente en ambas matrices. Se sugiere dividir cada lambda entre la varianza total (las lambdas están en eigen(S)$values). La varianza total es la suma de las varianzas de la diagonal de S. Una forma es sum(diag(S)). La varianza total de los componentes es la suma de los valores propios (es decir, la suma de la varianza de cada componente), sin embargo, si sumas la diagonal de S (es decir, la varianza de cada x), te da el mismo valor (¡compruébalo!). Recuerda que las combinaciones lineales buscan reproducir la varianza de X.

```
iCovVariablesTotal = sum(diag(cov(oCorporal)))

oCovPropVariablesExp = oCovEigenValues / iCovVariablesTotal
oCovPropVariablesExp

## [1] 0.7615357176 0.1703098726 0.0585307219 0.0091271040 0.0004965839

iCorVariablesTotal = sum(diag(cor(oCorporal)))

oCorPropVariablesExp = oCorEigenValues / iCorVariablesTotal
oCorPropVariablesExp

## [1] 0.75149947 0.14517133 0.06406596 0.02492375 0.01433950
```

1.3 Acumule los resultados anteriores (cumsum() puede servirle) para obtener la varianza acumulada en cada componente.

```
oCovVariablesAcumulada = cumsum(oCovPropVariablesExp)
oCovVariablesAcumulada

## [1] 0.7615357 0.9318456 0.9903763 0.9995034 1.0000000

oCorVariablesAcumulada = cumsum(oCorPropVariablesExp)
oCorVariablesAcumulada

## [1] 0.7514995 0.8966708 0.9607368 0.9856605 1.0000000
```

1.4 Según los resultados anteriores, ¿qué componentes son los más importantes?

1.5 Escriba la ecuación de la combinación lineal de los Componentes principales CP1 y CP2 (eiX, donde ei está en eigen(S)$vectors[1], e2X para obtener CP2, donde X = c(X1, X2, …)) ¿qué variables son las que más contribuyen a la primera y segunda componentes principales? (observe los coeficientes en valor absoluto de las combinaciones lineales). Justifique su respuesta.

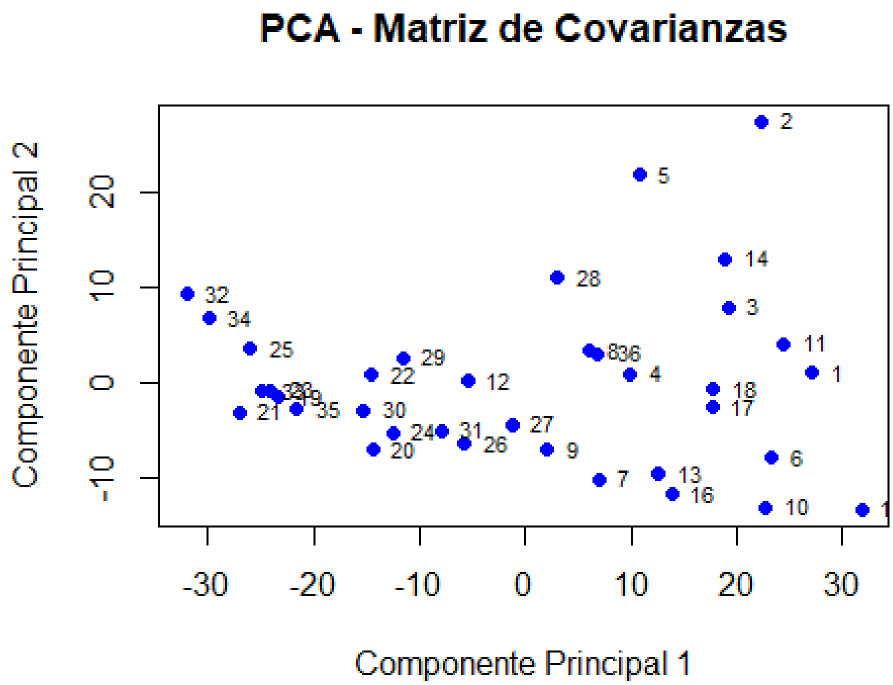
```
oCovCP1 = oCovEigenVectors[,1]
oCorCP1 = oCorEigenVectors[,1]
oCovCP1
```

```
## [1] -0.34871002 -0.76617586 -0.47632405 -0.05386189 -0.24817367

oCorCP1

## [1] -0.3359310 -0.4927066 -0.4222426 -0.4821923 -0.4833139

oCovCP2 = oCovEigenVectors[,2]
oCorCP2 = oCorEigenVectors[,2]
oCovCP2

## [1]  0.9075501 -0.1616581 -0.3851755  0.0155423 -0.0402221

oCorCP2

## [1]  0.8575601 -0.1647821 -0.4542223  0.1082775 -0.1392684
```

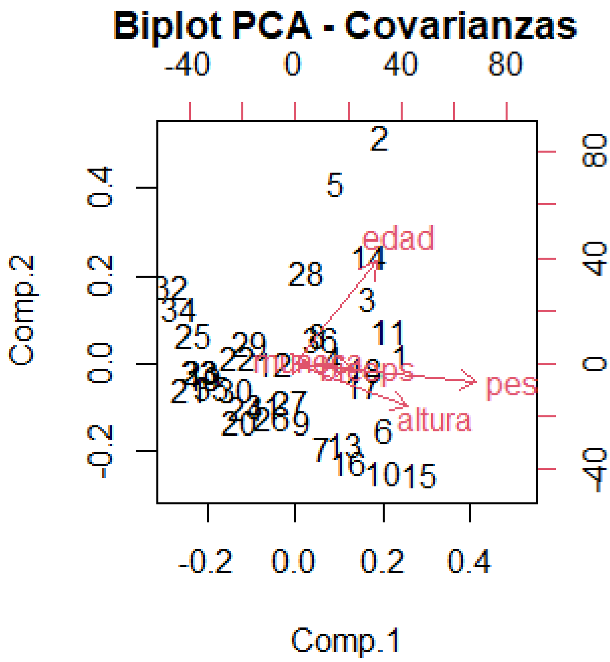
Parte 2

```
# Cargar la librería necesaria
library(stats)

oCPSCov <- princomp(oCorporal, cor = FALSE)
oCPSCor <- princomp(oCorporal, cor = TRUE)

summary(oCPSCov)

## Importance of components:
##                            Comp.1    Comp.2     Comp.3      Comp.4
Comp.5
## Standard deviation     18.6926388 8.8398600 5.18223874 2.046406827
0.4773333561
## Proportion of Variance  0.7615357 0.1703099 0.05853072 0.009127104
0.0004965839
## Cumulative Proportion   0.7615357 0.9318456 0.99037631 0.999503416
1.0000000000

summary(oCPSCor)

## Importance of components:
##                           Comp.1    Comp.2     Comp.3     Comp.4    Comp.5
## Standard deviation     1.9384265 0.8519722 0.56597686 0.35301378 0.2677639
## Proportion of Variance 0.7514995 0.1451713 0.06406596 0.02492375 0.0143395
## Cumulative Proportion  0.7514995 0.8966708 0.96073676 0.98566050 1.0000000

plot(oCPSCov$scores[,1], oCPSCov$scores[,2],
 xlab = "Componente Principal 1",
 ylab = "Componente Principal 2",
 main = "PCA - Matriz de Covarianzas",
 pch = 19, col = "blue")
```

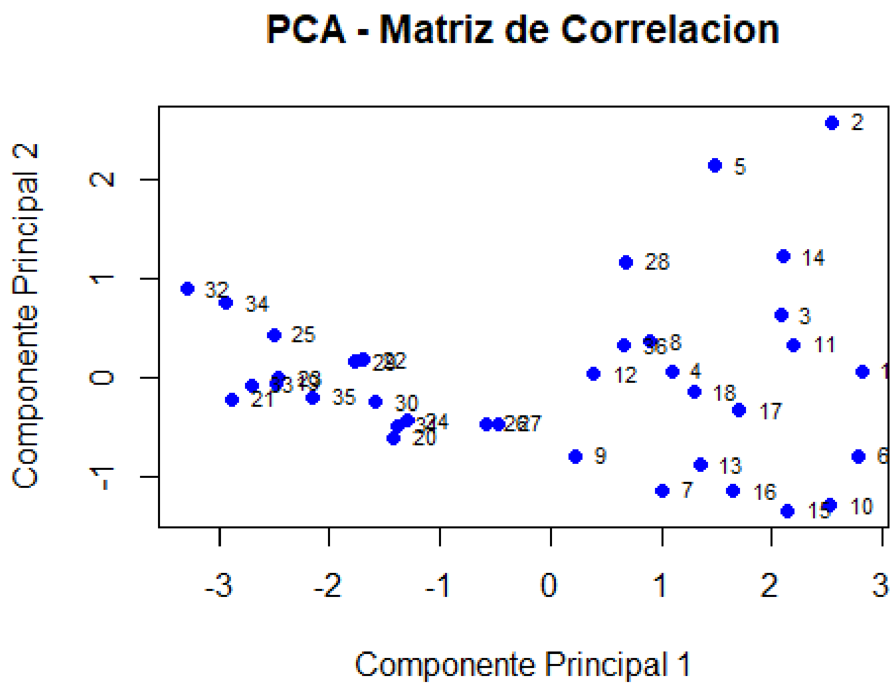
```r
text(oCPSCov$scores[,1], oCPSCov$scores[,2],
 labels = 1:nrow(oCorporal), pos = 4, cex = 0.7)
```



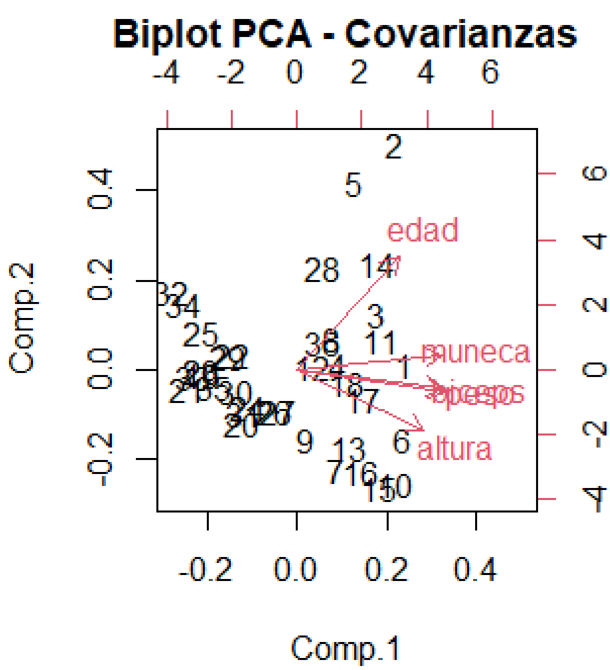

```r
# Biplot para visualizar variables y observaciones
biplot(oCPSCov, main = "Biplot PCA - Covarianzas")
```




```r
plot(oCPSCor$scores[,1], oCPSCor$scores[,2],
 xlab = "Componente Principal 1",
 ylab = "Componente Principal 2",
 main = "PCA - Matriz de Correlacion",
 pch = 19, col = "blue")

text(oCPSCor$scores[,1], oCPSCor$scores[,2],
 labels = 1:nrow(oCorporal), pos = 4, cex = 0.7)
```




```r
# Biplot para visualizar variables y observaciones
biplot(oCPSCor, main = "Biplot PCA - Covarianzas")
```



```r
# Puntuaciones (scores) para la matriz de covarianzas
oScoreS <- oCPSCov$scores
# Ver las primeras puntuaciones
head(oScoreS)
```

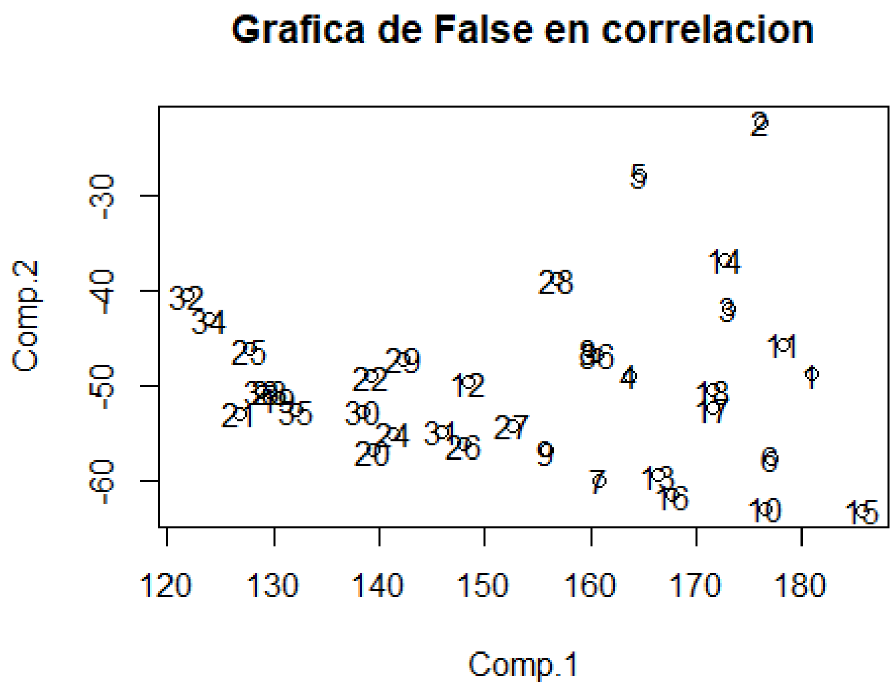
```
##         Comp.1     Comp.2     Comp.3      Comp.4      Comp.5
## [1,] 27.162853  1.0278492  5.0022646  0.93622690 -0.51688356
## [2,] 22.363542 27.5955807  3.0635949 -0.08338126  0.02552809
## [3,] 19.167874  7.9566157 -1.5770026 -2.61077676  0.80391745
## [4,]  9.959001  0.8923731  5.5146952  0.12345373 -0.35579895
## [5,] 10.775593 22.0203437 -0.7562826  0.17996723 -0.41646606
## [6,] 23.283948 -7.9268214  2.7958617 -2.09339284 -0.62252321
```

```r
# Puntuaciones (scores) para la matriz de correlaciones
oScoreR <- oCPSCor$scores
# Ver las primeras puntuaciones
head(oScoreR)
```

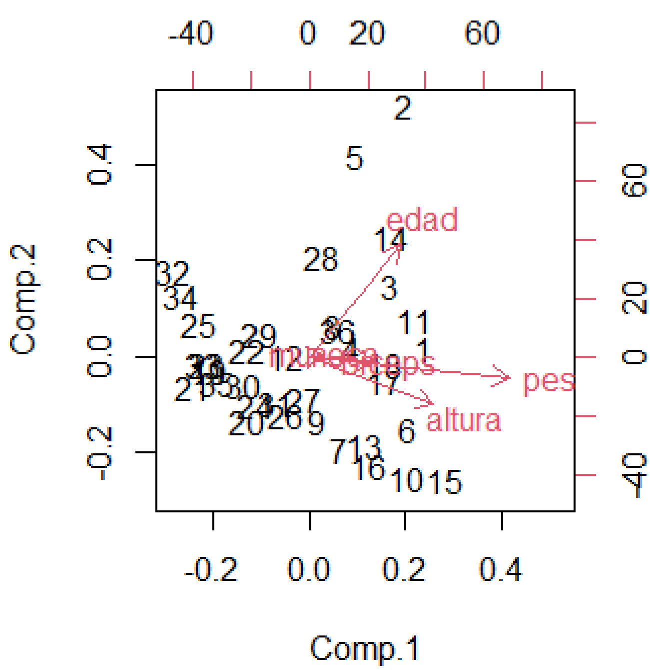
```
##        Comp.1      Comp.2      Comp.3      Comp.4       Comp.5
## [1,] 2.813992  0.06282760  0.51434516 -0.37618363 -0.161649397
## [2,] 2.550816  2.57369731  0.42896223  0.01252075  0.083602262
## [3,] 2.079207  0.62112516 -0.12602006  0.51138786  0.430775853
## [4,] 1.093316  0.06328171  0.46145821 -0.35236278 -0.008424496
## [5,] 1.489363  2.13420572 -0.08620983 -0.19530483 -0.097669770
## [6,] 2.780190 -0.79964368 -0.11180511 -0.52796031  0.113681564
```

2.

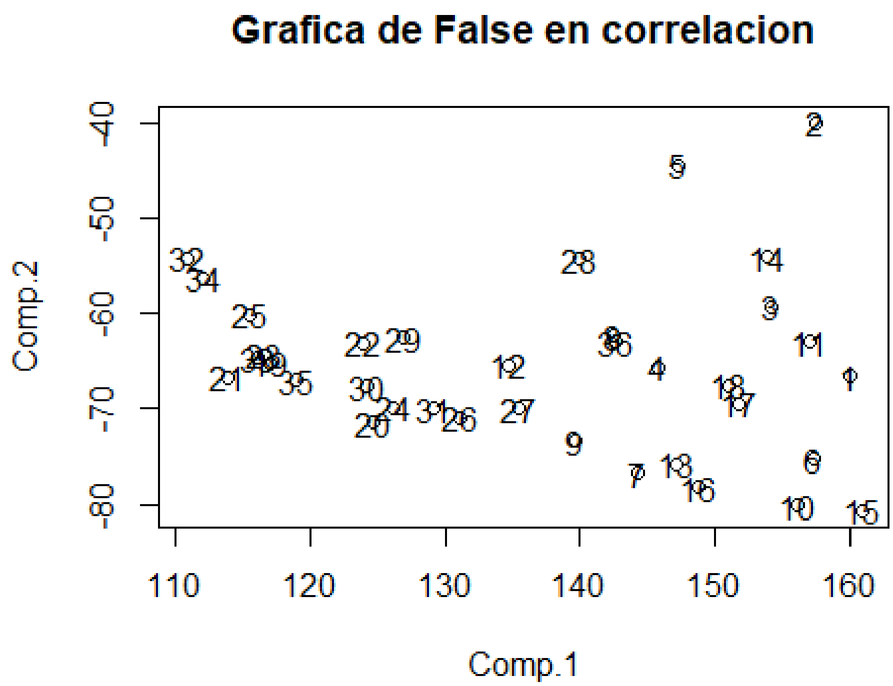
```r
library(stats)

oCPS=princomp(oCorporal,cor=FALSE) #Para la matriz de correlación usa
cor=TRUE
oCPAS=as.matrix(oCorporal)%*%oCPS$loadings #Calcula las puntuaciones
plot(oCPAS[,1:2],type="p", main = "Grafica de False en correlacion")
text(oCPAS[,1],oCPAS[,2],1:nrow(oCPAS))
```



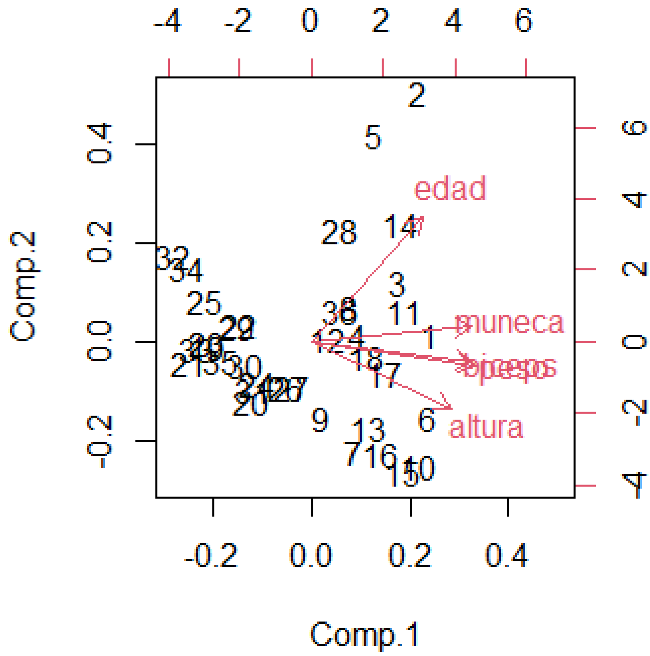

```r
biplot(oCPS)
```



```r
library(stats)

oCPS=princomp(oCorporal,cor=TRUE) #Para la matriz de correlación usa cor=TRUE
oCPAS=as.matrix(oCorporal)%*%oCPS$loadings #Calcula las puntuaciones
plot(oCPAS[,1:2],type="p", main = "Grafica de False en correlacion")
text(oCPAS[,1],oCPAS[,2],1:nrow(oCPAS))
```

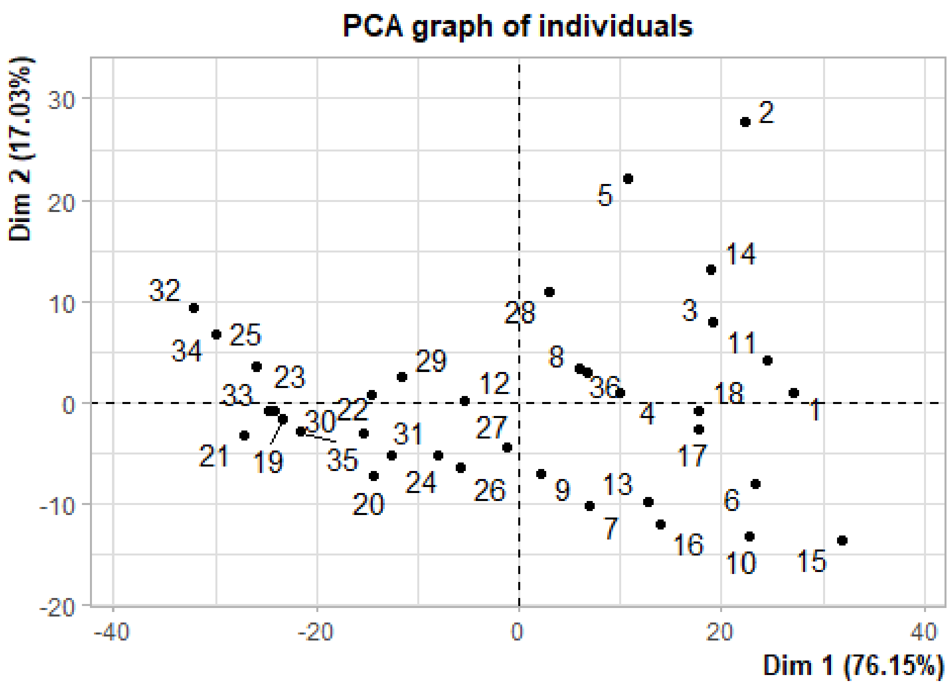


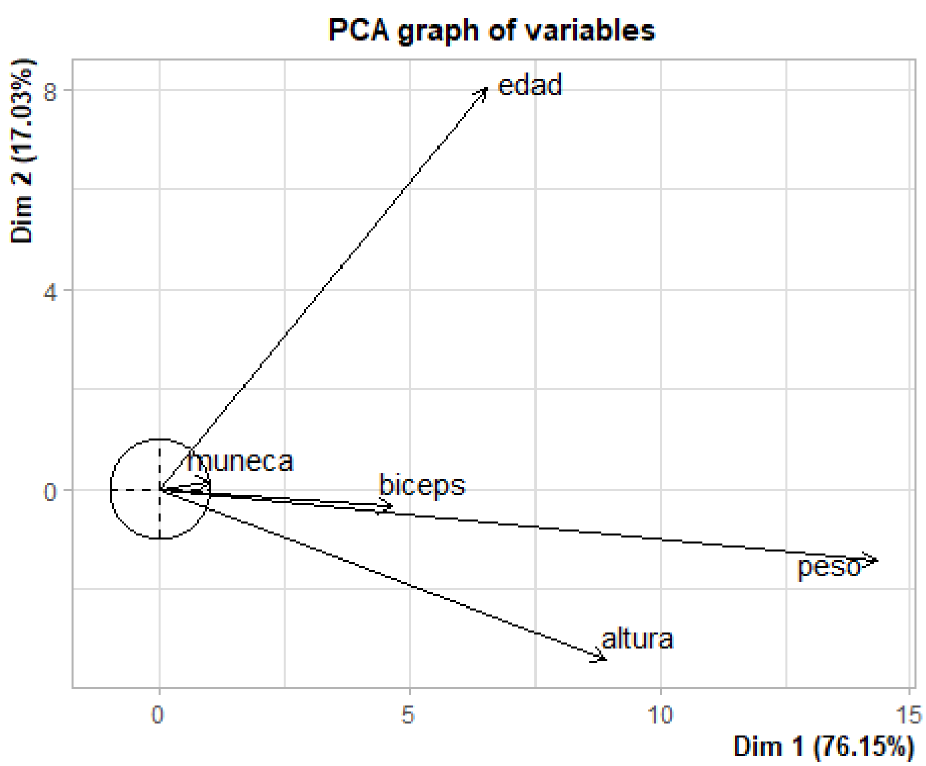

```r
biplot(oCPS)
```



Parte 3

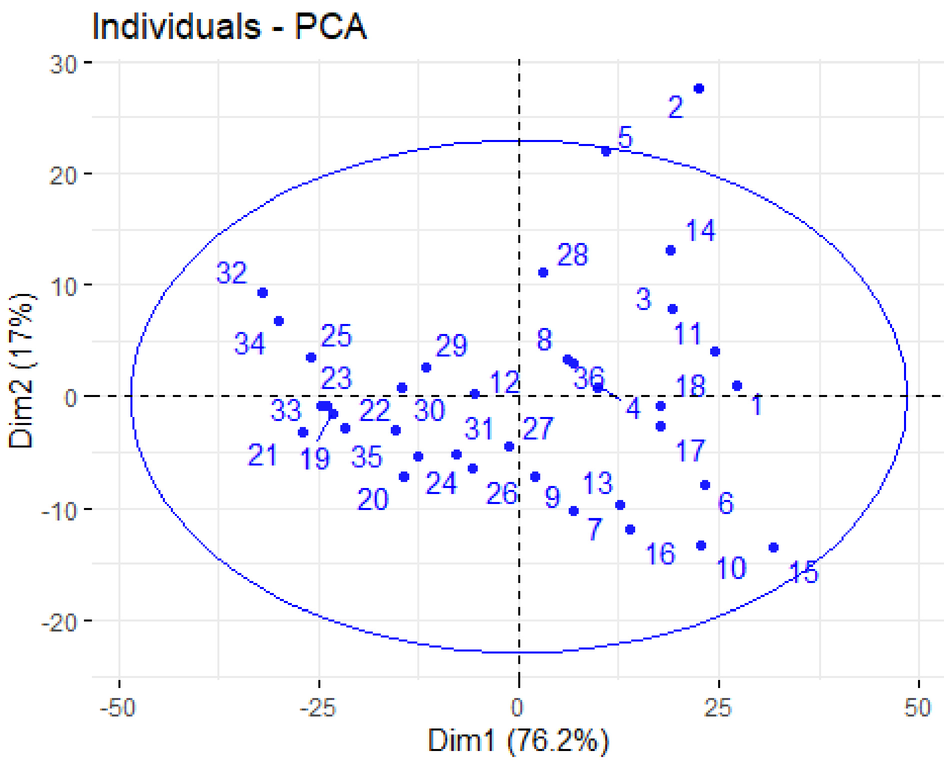
```
library(FactoMineR)
library(ggplot2)

oCPS = PCA(oCorporal,scale.unit=FALSE)
```



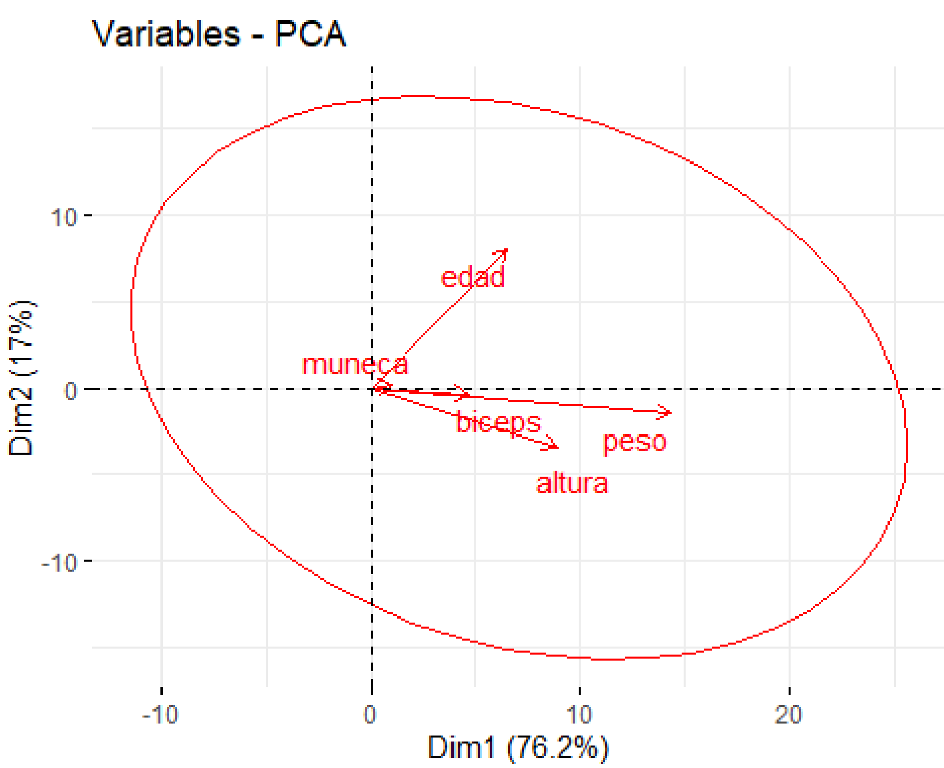



```r
library(factoextra)

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

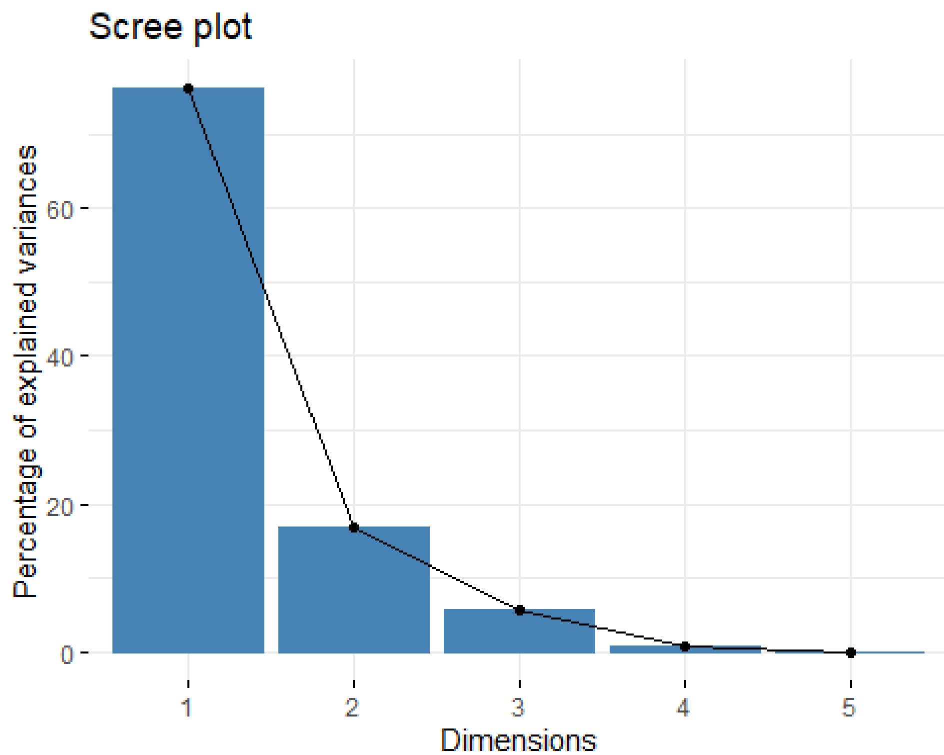
```
fviz_pca_ind(oCPS, col.ind = "blue", addEllipses = TRUE, repel = TRUE)
```



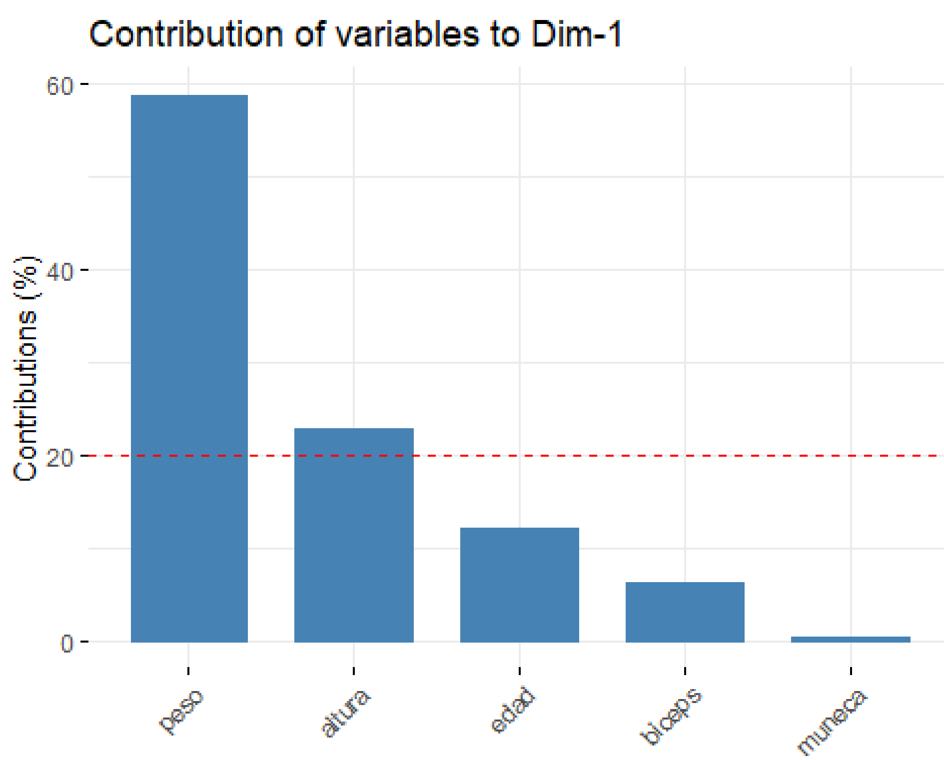

```
fviz_pca_var(oCPS, col.var = "red", addEllipses = TRUE, repel = TRUE)
```



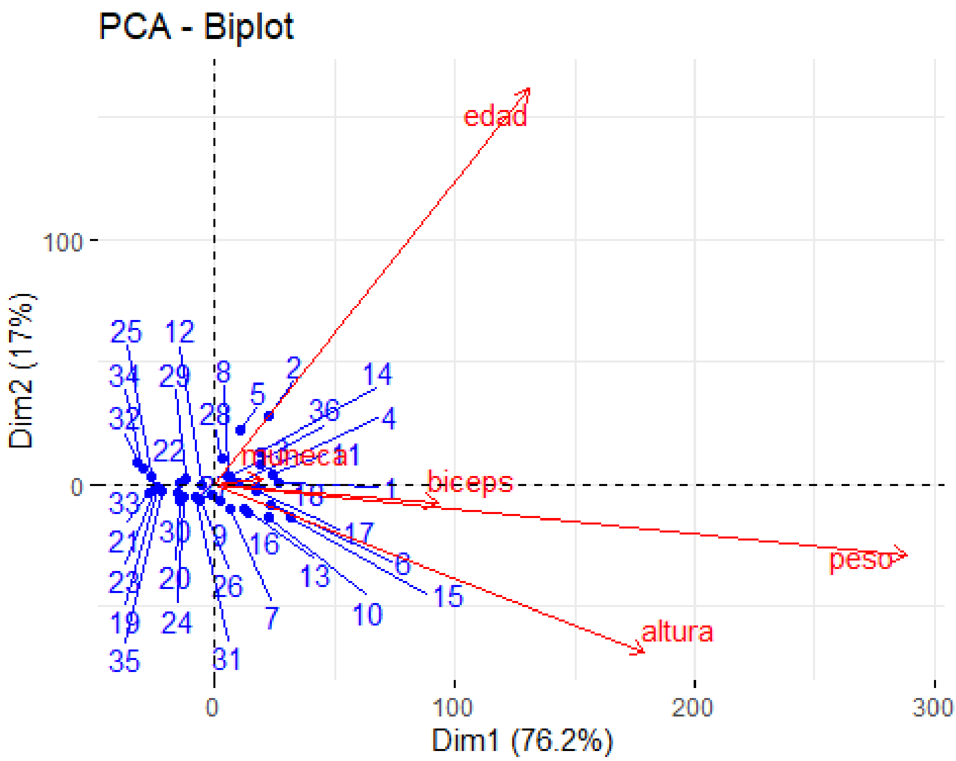
```r
fviz_screeplot(oCPS)
```



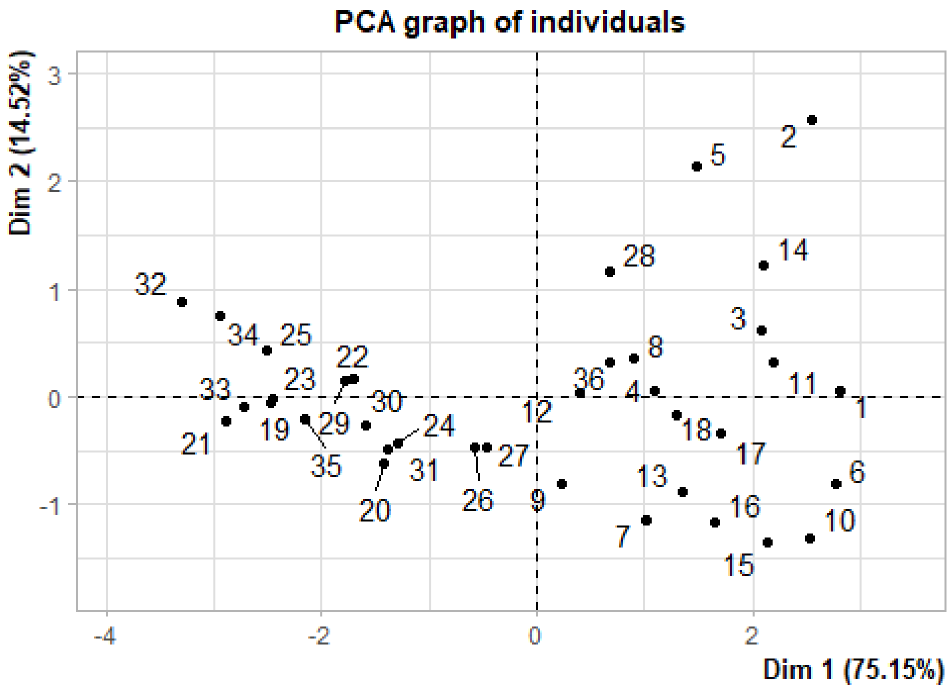

```r
fviz_contrib(oCPS, choice = "var")
```

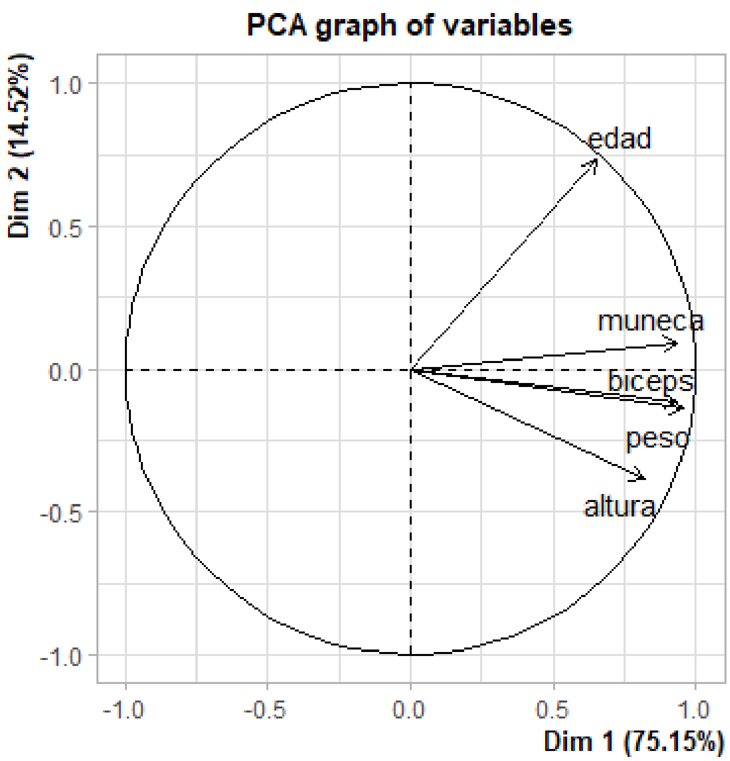


```r
fviz_pca_biplot(oCPS, repel=TRUE, col.var="red", col.ind="blue")
```

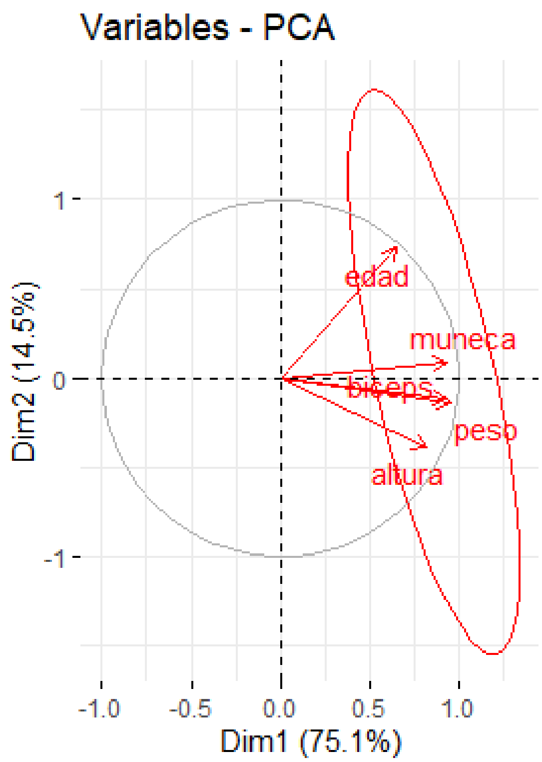



```r
oCPR = PCA(oCorporal,scale.unit=TRUE)
```

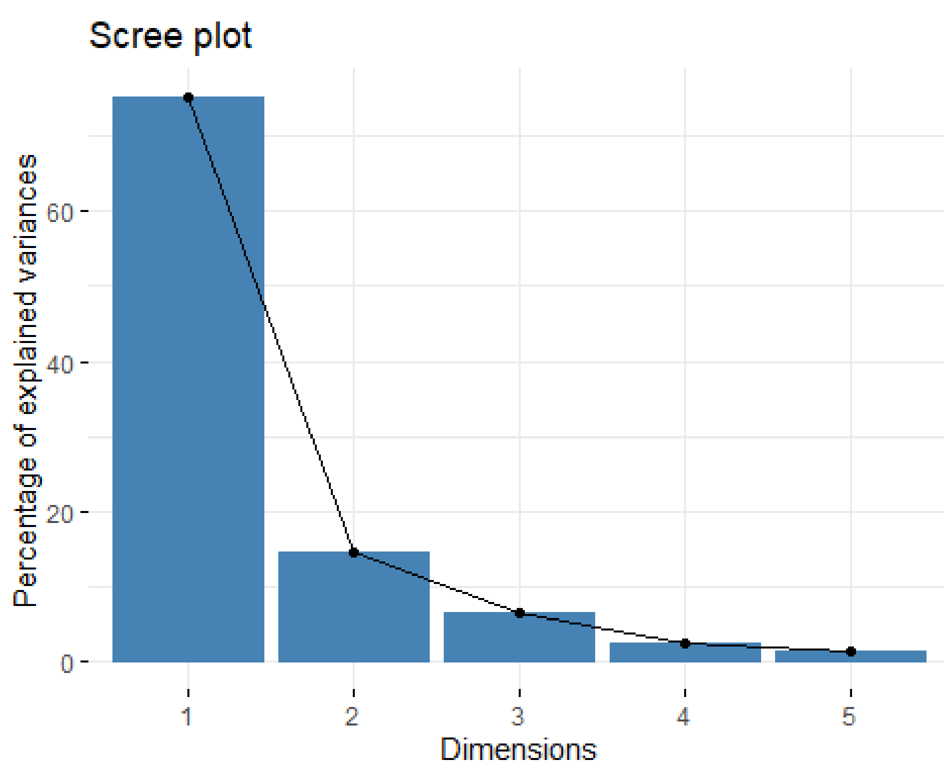




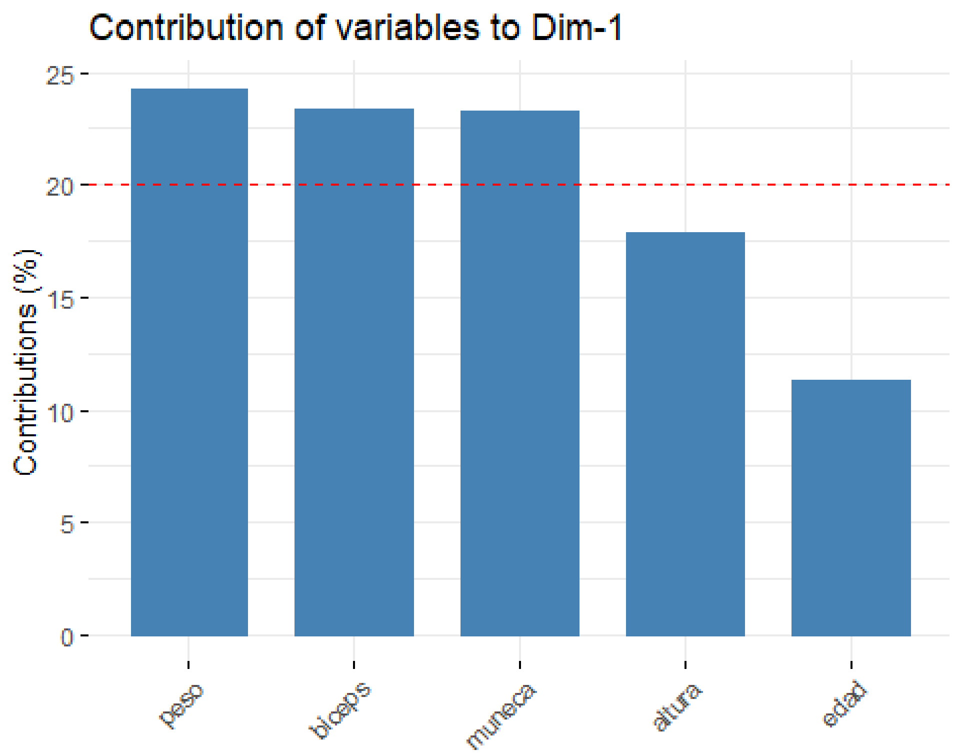

```
fviz_pca_var(oCPR, col.var = "red", addEllipses = TRUE, repel = TRUE)
```



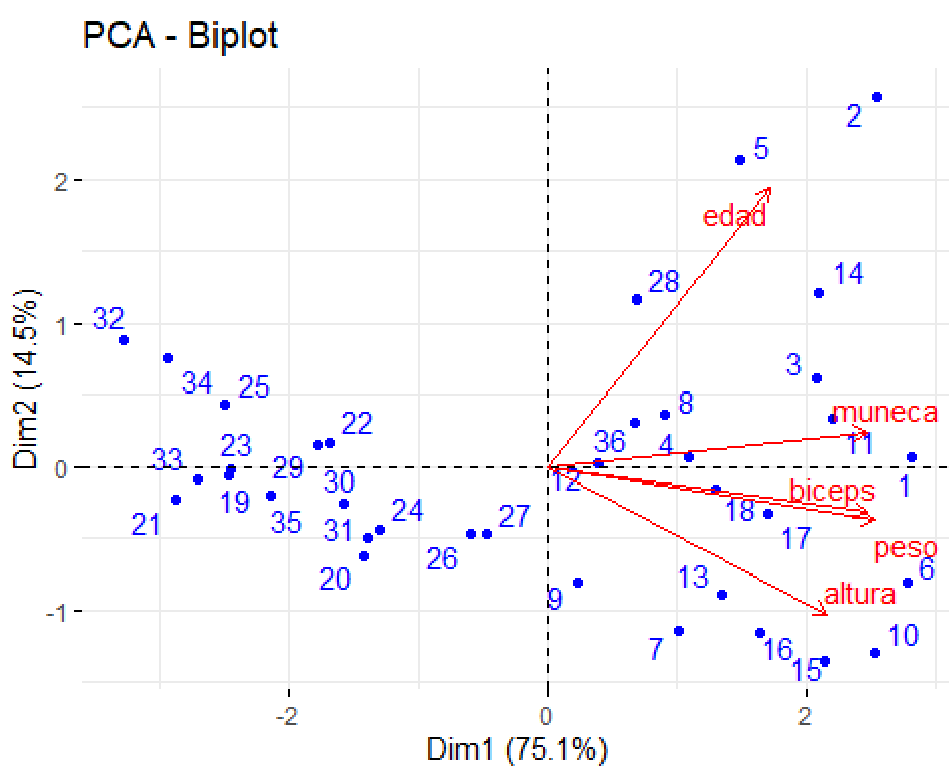

```r
fviz_screeplot(oCPR)
```




```r
fviz_contrib(oCPR, choice = "var")
```




```
fviz_pca_biplot(oCPR, repel=TRUE, col.var="red", col.ind="blue")
```



Parte 4

Compare los resultados obtenidos con la matriz de varianza-covarianza y con la correlación . ¿Qué concluye? ¿Cuál de los dos procedimientos aporta componentes con de mayor interés?

Viendo los resultados puedo concluir que la matriz de correlacion es el mejor, ya que los datos ya no tienen tanta diferencia de tamaño como la varianza-covarianza

Indique cuál de los dos análisis (a partir de la matriz de varianza y covarianza o de correlación) resulta mejor para los datos indicadores económicos y sociales del 96 países en el mundo. Comparar los resultados y argumentar cuál es mejor según los resultados obtenidos.

La Matriz de correlacion es el mejor analisis para indicadores economicos y sociales, ya que con la matriz de correlacion se pueden escalar los valores ya que los valores economicos y sociales normalmente trabajarian en diferentes escalas.

¿Qué variables son las que más contribuyen a la primera y segunda componentes principales del método seleccionado? (observa los coeficientes en valor absoluto de las combinaciones lineales, auxíliate también de los gráficos)

Las variables que más contribuyen para el primer componente pueden ser el PIB, inversion en infraestructura.

Y para el segundo componente estarian la esperanza de vida y algunos indicadores sociales.

Escriba las combinaciones finales que se recomiendan para hacer el análisis de componentes principales.

Para el primer componente un indice de riqueza y desarrollo economico y para el segundo componente los indices sociales e indicadores de bienestar social.

Interpreta los resultados en término de agrupación de variables (puede ayudar “índice de riqueza”, “índice de ruralidad”, etc)

Para agrupar los paises para una segmentacion se puede realizar por inversion en infraestructura, desarrollo economico, esperanza de vida y bienestar social