Acticividad-16-A4

Saúl Francisco Vázquez del Río

2024-10-08

PARTE I

Realiza el análisis de los valores y vectores propios con la matriz de covarianzas y con la de correlación. Analiza la varianza explicada por cada componente en cada caso e interpreta dentro del contexto del problema.

```
# Cargar los datos
M <- read.csv("C:\\Users\\saulv\\OneDrive\\Escritorio\\Septimo
semestre\\corporal.csv")
head(M)

##   edad peso altura   sexo muneca biceps
## 1   43 87.3  188.0 Hombre   12.2   35.8
## 2   65 80.0  174.0 Hombre   12.0   35.0
## 3   45 82.3  176.5 Hombre   11.2   38.5
## 4   37 73.6  180.3 Hombre   11.2   32.2
## 5   55 74.1  167.6 Hombre   11.8   32.9
## 6   33 85.9  188.0 Hombre   12.4   38.5

M <- M[ , -4]
head(M)

##   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
```

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:

```
# Matriz de covarianzas
S <- cov(M)

# Matriz de correlaciones
R <- cor(M)

cat("Matriz de covarianza", S)

## Matriz de covarianza 111.3968 80.88159 36.66603 7.698095 26.72095
80.88159 221.0871 124.7287 14.84467 70.73838 36.66603 124.7287 110.674
```

```
8.156476 39.02105 7.698095 14.84467 8.156476 1.381714 5.400571 26.72095
70.73838 39.02105 5.400571 27.39886
```

```r
cat("Matriz de correlaccion", R)
```

```
## Matriz de correlaccion 1 0.5153847 0.3302211 0.6204942 0.4836702
0.5153847 1 0.7973737 0.8493361 0.9088813 0.3302211 0.7973737 1 0.6595849
0.7086144 0.6204942 0.8493361 0.6595849 1 0.8777369 0.4836702 0.9088813
0.7086144 0.8777369 1
```

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

```r
# Valores y vectores propios para la matriz de covarianzas
eigen_S <- eigen(S)

# Valores y vectores propios para la matriz de correlaciones
eigen_R <- eigen(R)

# Mostrar resultados
eigen_S$values # Valores propios de la matriz de covarianzas
```

```
## [1] 359.3980243  80.3757858  27.6229011   4.3074318   0.2343571
```

```r
eigen_S$vectors # Vectores propios de la matriz de covarianzas
```

```
##             [,1]       [,2]        [,3]         [,4]         [,5]
## [1,] -0.34871002  0.9075501 -0.23248825 -0.001589466  0.026473941
## [2,] -0.76617586 -0.1616581  0.52166894 -0.338508602  0.010707863
## [3,] -0.47632405 -0.3851755 -0.78905759  0.046160807  0.003543154
## [4,] -0.05386189  0.0155423  0.02785902  0.126103480 -0.990039959
## [5,] -0.24817367 -0.0402221  0.22455005  0.931330496  0.137814357
```

```r
eigen_R$values # Valores propios de la matriz de correlaciones
```

```
## [1] 3.75749733 0.72585665 0.32032981 0.12461873 0.07169749
```

```r
eigen_R$vectors # Vectores propios de la matriz de correlaciones
```

```
##            [,1]       [,2]        [,3]       [,4]       [,5]
## [1,] -0.3359310  0.8575601 -0.34913780 -0.1360111  0.1065123
## [2,] -0.4927066 -0.1647821  0.06924561 -0.5249533 -0.6706087
## [3,] -0.4222426 -0.4542223 -0.73394453  0.2070673  0.1839617
## [4,] -0.4821923  0.1082775  0.36690716  0.7551547 -0.2255818
## [5,] -0.4833139 -0.1392684  0.44722747 -0.3046138  0.6739511
```

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.

```r
# Varianza total (suma de las varianzas de la diagonal de S)
var_total_S <- sum(diag(S))

# Proporción de varianza explicada por cada componente para la matriz de
covarianza
prop_var_S <- eigen_S$values / var_total_S

# Varianza acumulada para la matriz de covarianza
cum_var_S <- cumsum(prop_var_S)

# Varianza total en la matriz de correlaciones
var_total_R <- sum(eigen_R$values)

# Proporción de varianza explicada por cada componente para la matriz de
correlación
prop_var_R <- eigen_R$values / var_total_R

# Varianza acumulada para la matriz de correlación
cum_var_R <- cumsum(prop_var_R)

# Mostrar resultados
prop_var_S

## [1] 0.7615357176 0.1703098726 0.0585307219 0.0091271040 0.0004965839

cum_var_S

## [1] 0.7615357 0.9318456 0.9903763 0.9995034 1.0000000

prop_var_R

## [1] 0.75149947 0.14517133 0.06406596 0.02492375 0.01433950

cum_var_R

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

Acumule los resultados anteriores (cumsum() puede servirle) para obtener la varianza acumulada en cada componente. Según los resultados anteriores, ¿qué componentes son los más importantes? 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.

```r
# Combinación lineal para CP1 y CP2 en la matriz de covarianzas
CP1_S <- eigen_S$vectors[,1]  # Primer componente
CP2_S <- eigen_S$vectors[,2]  # Segundo componente

# Combinación lineal para CP1 y CP2 en la matriz de correlaciones
CP1_R <- eigen_R$vectors[,1]  # Primer componente
```

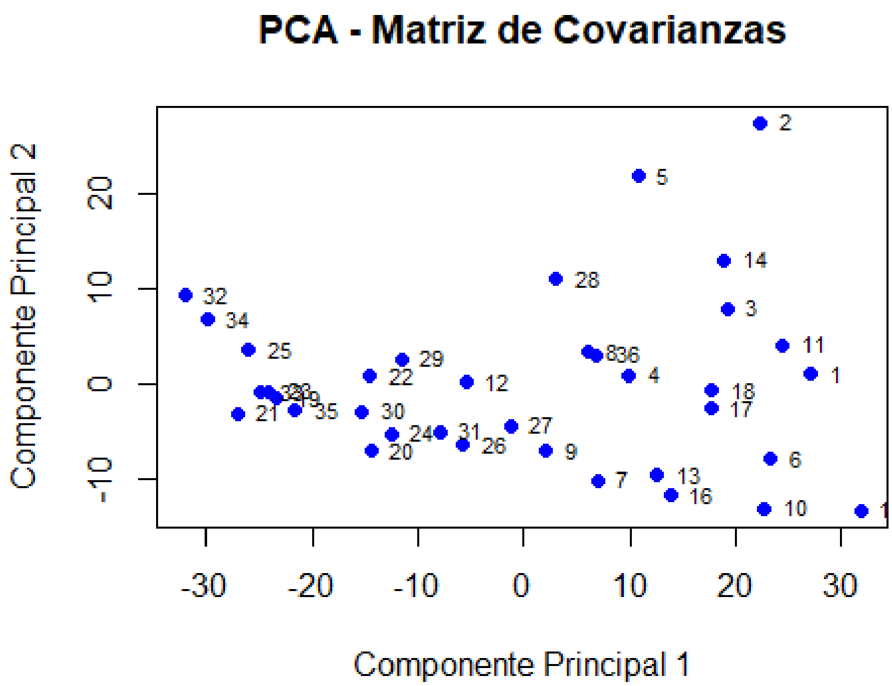
```
CP2_R <- eigen_R$vectors[,2]  # Segundo componente

# Mostrar los coeficientes
CP1_S

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

CP2_S

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

CP1_R

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

CP2_R

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

¡No te olvides de seguir los mismos pasos con la matriz de correlaciones (se obtiene con cor(x) si x está compuesto por variables numéricas)

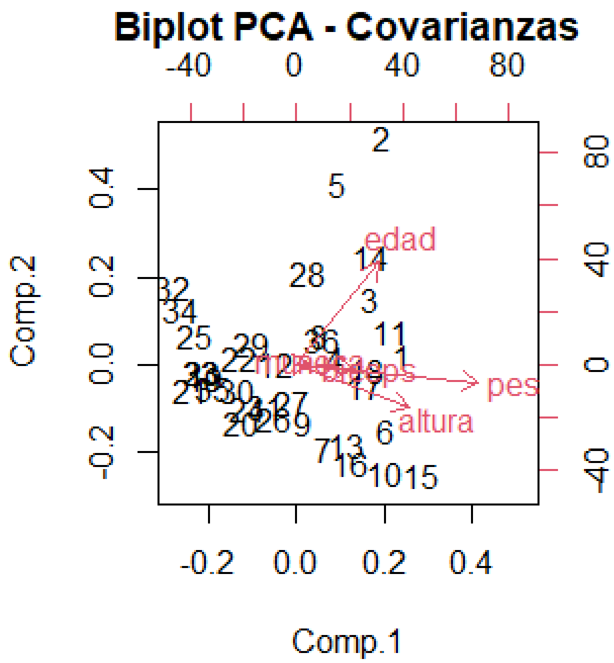
##PARTE II Obtenga las gráficas respectivas con S (matriz de varianzas-covarianzas) y con R (matriz de correlaciones) de las dos primeras componentes.

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

# Realizar PCA usando la matriz de covarianzas
cpS_cov <- princomp(M, cor = FALSE)

# Resumen de los resultados
summary(cpS_cov)

## 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

# Gráfico de las dos primeras componentes
plot(cpS_cov$scores[,1], cpS_cov$scores[,2],
     xlab = "Componente Principal 1",
     ylab = "Componente Principal 2",
     main = "PCA - Matriz de Covarianzas",
     pch = 19, col = "blue")

# Añadir etiquetas a los puntos (opcional)
```

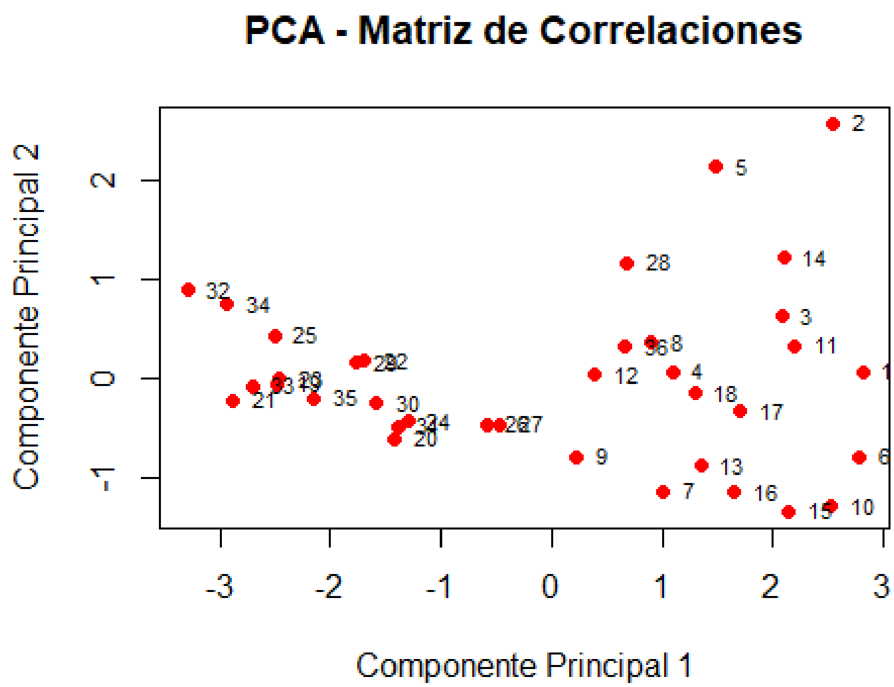
```
text(cpS_cov$scores[,1], cpS_cov$scores[,2],
     labels = 1:nrow(M), pos = 4, cex = 0.7)
```



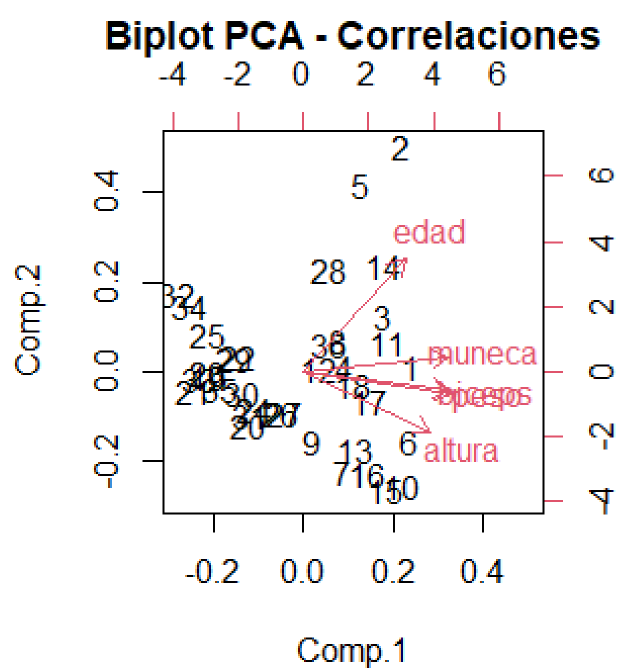

```
# Biplot para visualizar variables y observaciones
biplot(cpS_cov, main = "Biplot PCA - Covarianzas")
```




```
# Realizar PCA usando la matriz de correlaciones
cpS_cor <- princomp(M, cor = TRUE)

# Resumen de los resultados
summary(cpS_cor)

## 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

# Gráfico de las dos primeras componentes
plot(cpS_cor$scores[,1], cpS_cor$scores[,2],
     xlab = "Componente Principal 1",
     ylab = "Componente Principal 2",
     main = "PCA - Matriz de Correlaciones",
     pch = 19, col = "red")

# Añadir etiquetas a los puntos (opcional)
text(cpS_cor$scores[,1], cpS_cor$scores[,2],
     labels = 1:nrow(M), pos = 4, cex = 0.7)
```




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

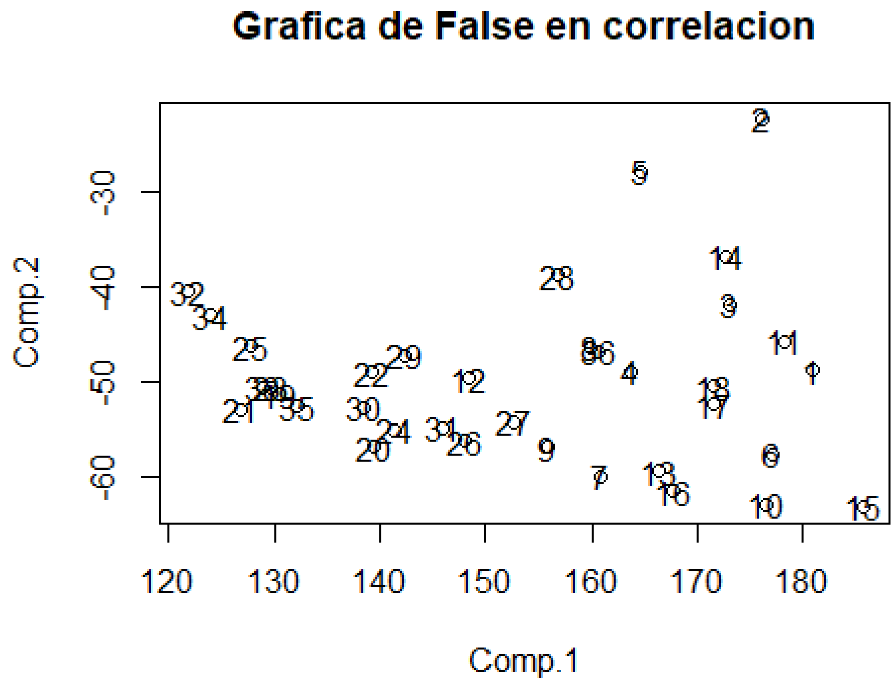


Calcule las puntuaciones (scores) de las observaciones para los componentes obtenidos con la matriz de varianzas-covarianzas Calcule las puntuaciones (scores) de las observaciones para los componentes obtenidos con la matriz de correlaciones. Recuerde que en la matriz de correlaciones las variables tienen que estar estandarizadas.

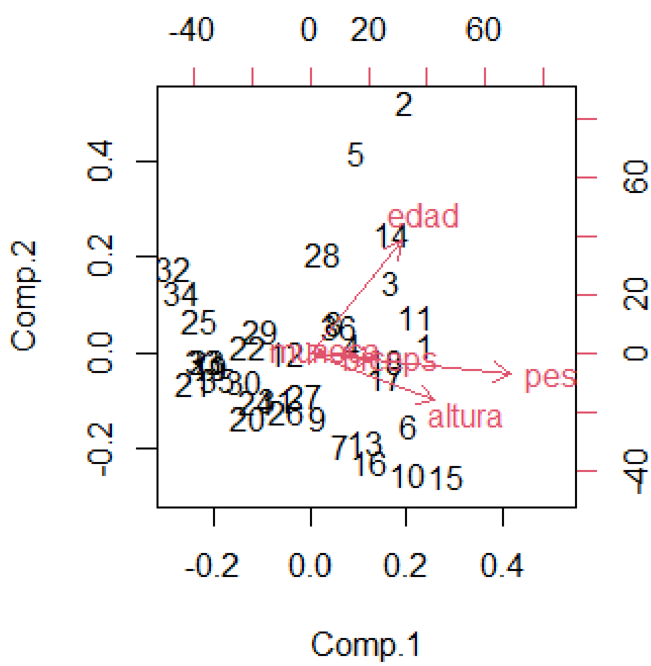
```
# Puntuaciones (scores) para la matriz de covarianzas
scores_S <- cpS_cov$scores

# Ver las primeras puntuaciones
head(scores_S)

##         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

# Puntuaciones (scores) para la matriz de correlaciones
scores_R <- cpS_cor$scores

# Ver las primeras puntuaciones
head(scores_R)

##        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
```

Interprete los gráficos en términos de: Las relaciones que se establecen entre las variables y los componentes principales La relación entre las puntuaciones de las observaciones y los valores de las variables Detecte posibles datos atípicos Explora el: princomp() en library(stats). Puedes poner help(princomp) en la consola o buscarlo en la ventana de ayuda. Indaga: ¿qué otras opciones tiene para facilitarte el análisis? En particular, explora los comandos y subcomandos: summary(cpS), cpa$Sloading, cpaS$scores. ¿Cómo se interpreta el resultado? Sugerencias en R

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

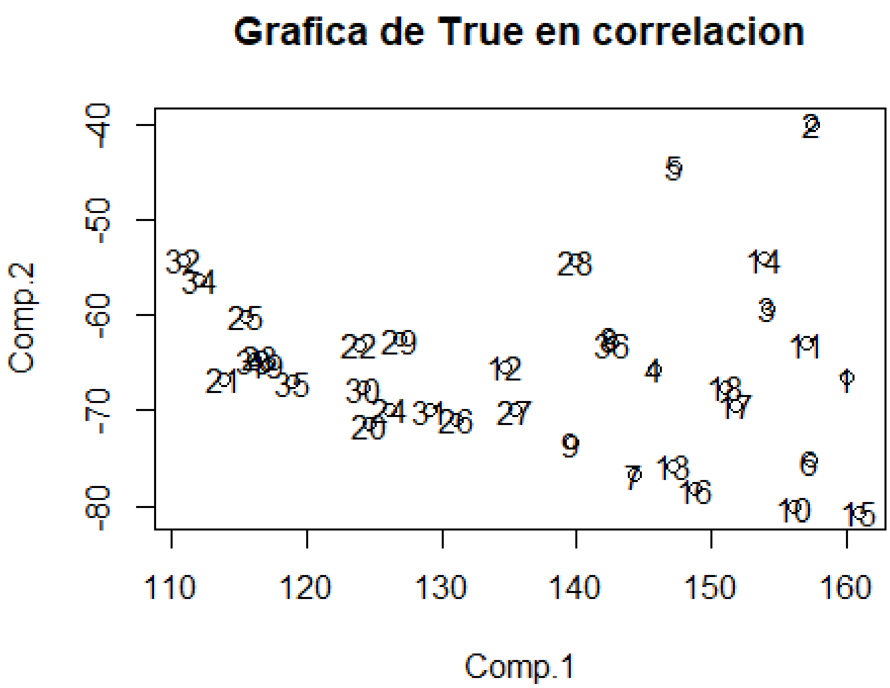



```r
biplot(cpS)
```

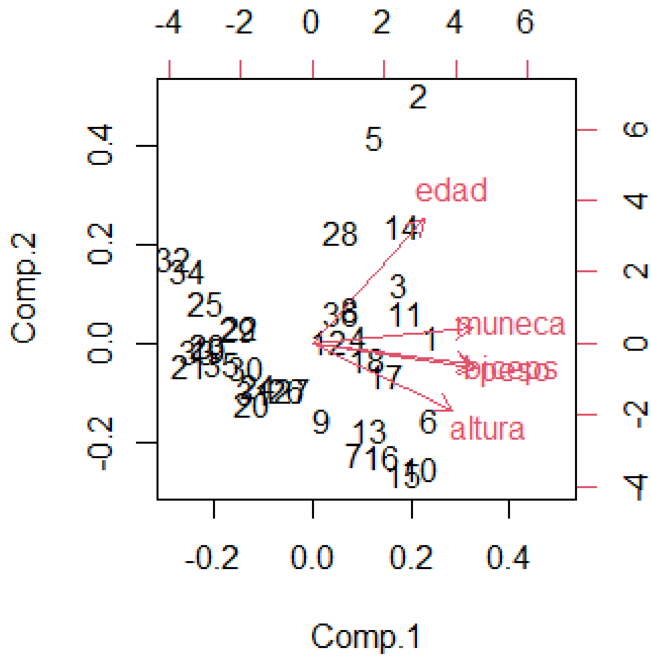



```r
library(stats)
datos= M
```

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



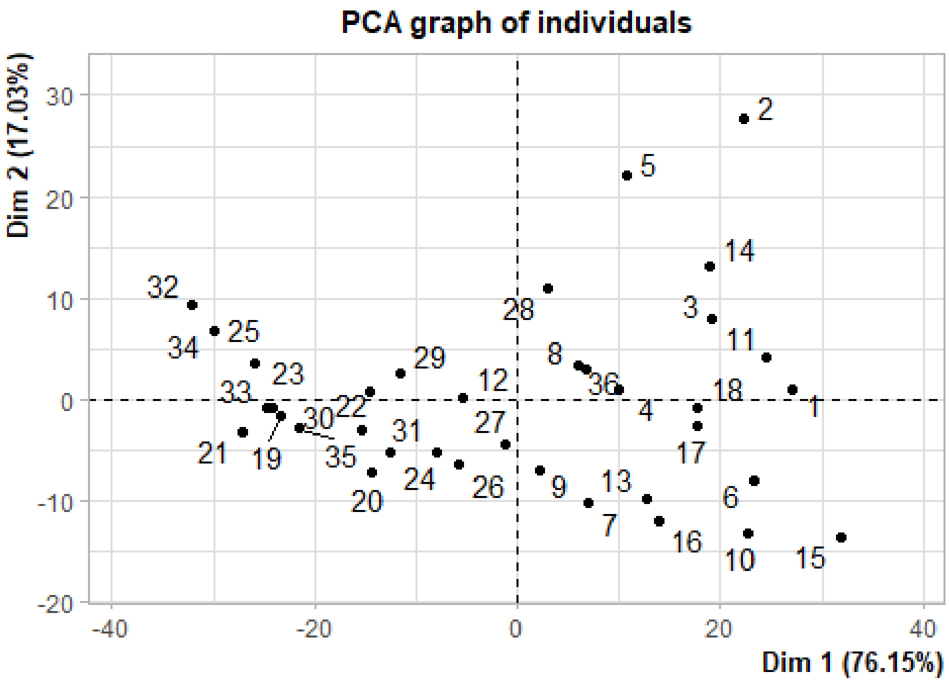

```r
biplot(cpS)
```

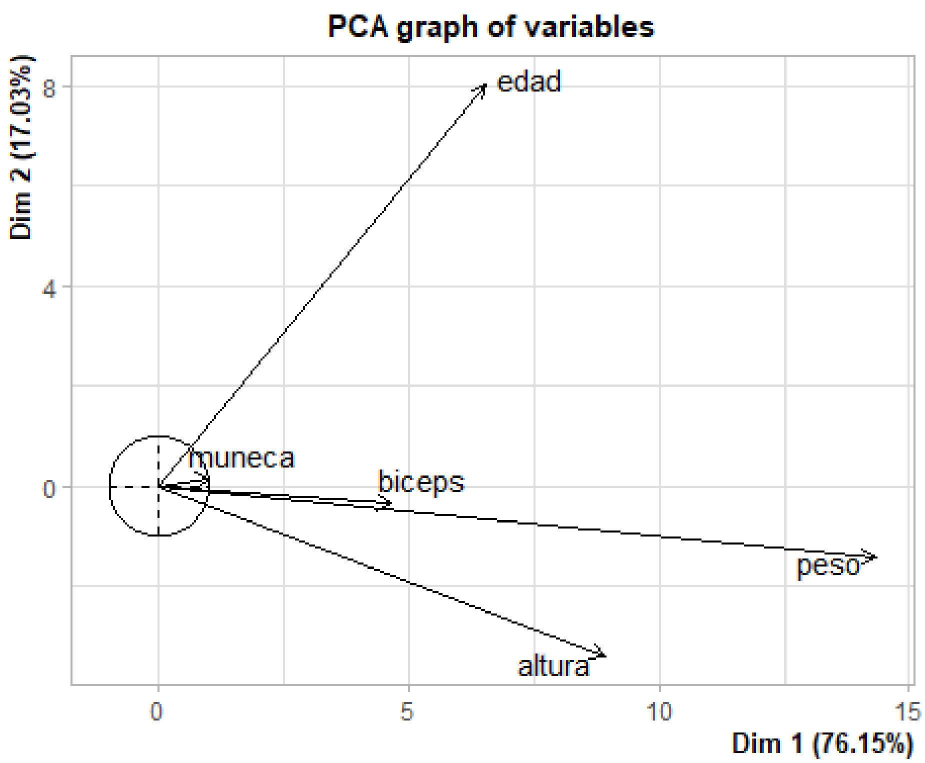


##PARTE III Explore los siguientes gráficos relativos a Componentes Principales. Interprete cada gráfico e identifica qué es lo que se está graficando en cada uno. Realiza el análisis con la matriz de varianzas y covarianzas y correlación. library(FactoMineR) library(ggplot2) datos=matriz de datos cpS = PCA(datos,scale.unit=FALSE) #Para matriz de correlaciones usa scale.unit=TRUE library(factoextra) fviz_pca_ind(cpS, col.ind = "blue", addEllipses = TRUE, repel = TRUE) fviz_pca_var(cpS, col.var = "red", addEllipses = TRUE, repel = TRUE) fviz_screeplot(cpS) fviz_contrib(cpS, choice = c("var") fviz_pca_biplot(cpS, repel=TRUE, col.var="red", col.ind="blue")

Explora el comando PCA, (puedes poner help(PCA) en la consola o buscarlo en la ventana de ayuda) ¿qué otras opciones tiene para facilitarte el análisis?

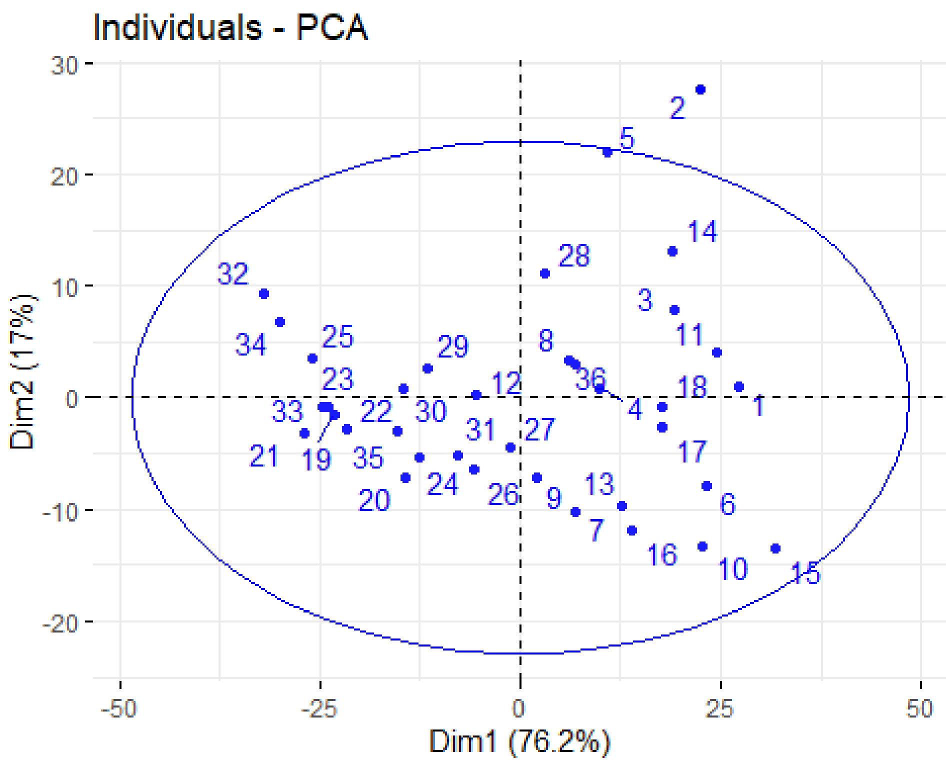
```
library(FactoMineR)
library(ggplot2)
datos=M
cpS = PCA(datos,scale.unit=FALSE) #Para matriz de correlaciones usa
scale.unit=TRUE
```



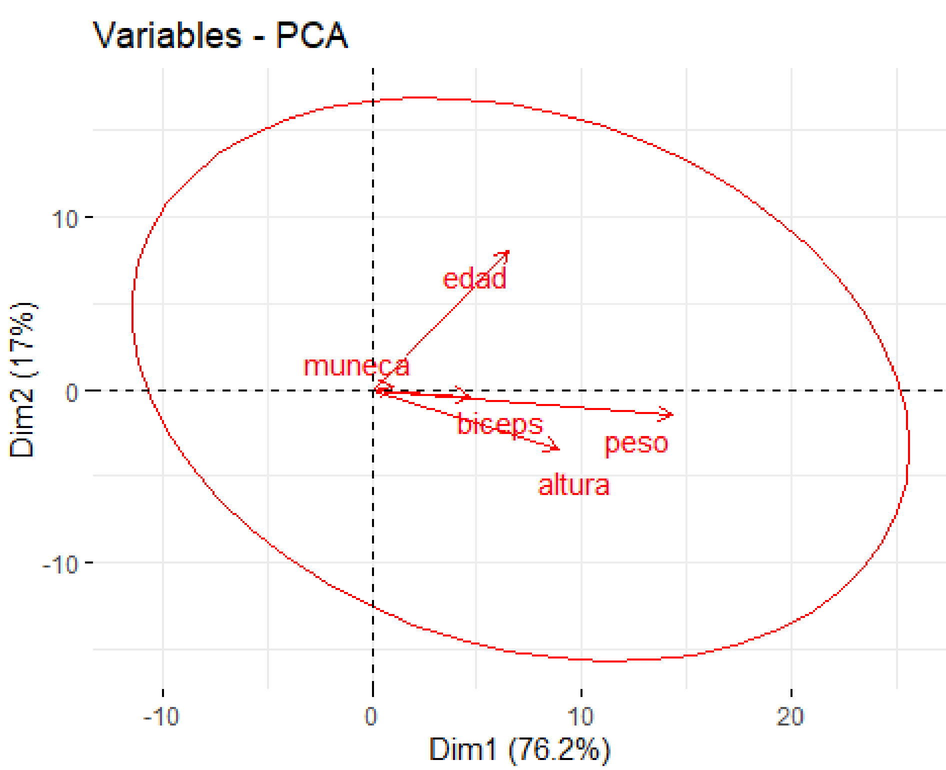



```
library(factoextra)

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

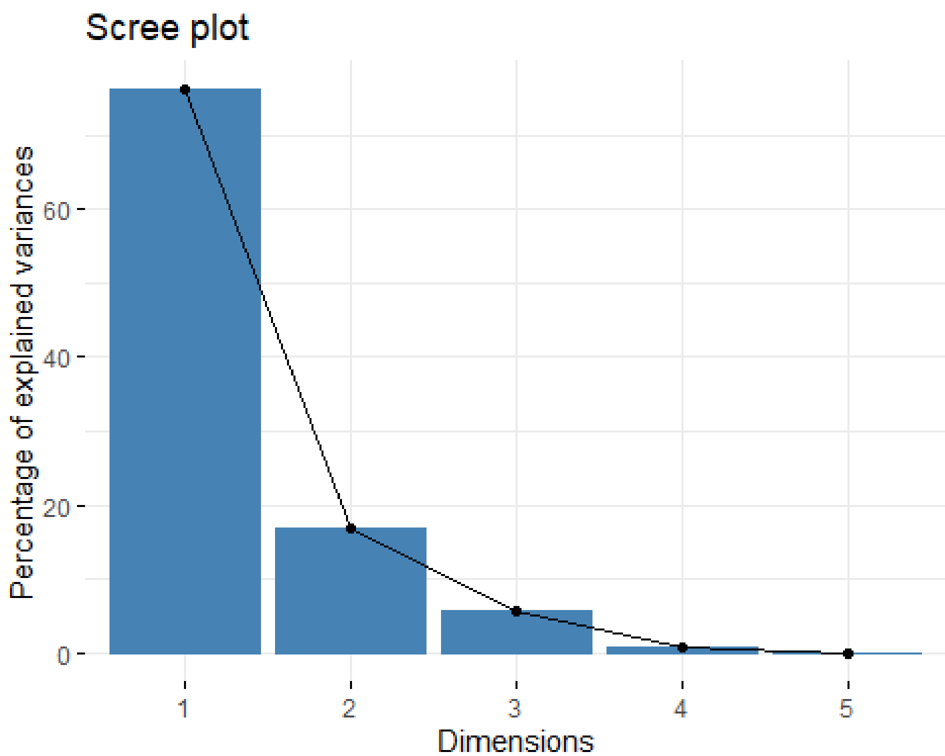
```
fviz_pca_ind(cpS, col.ind = "blue", addEllipses = TRUE, repel = TRUE)
```



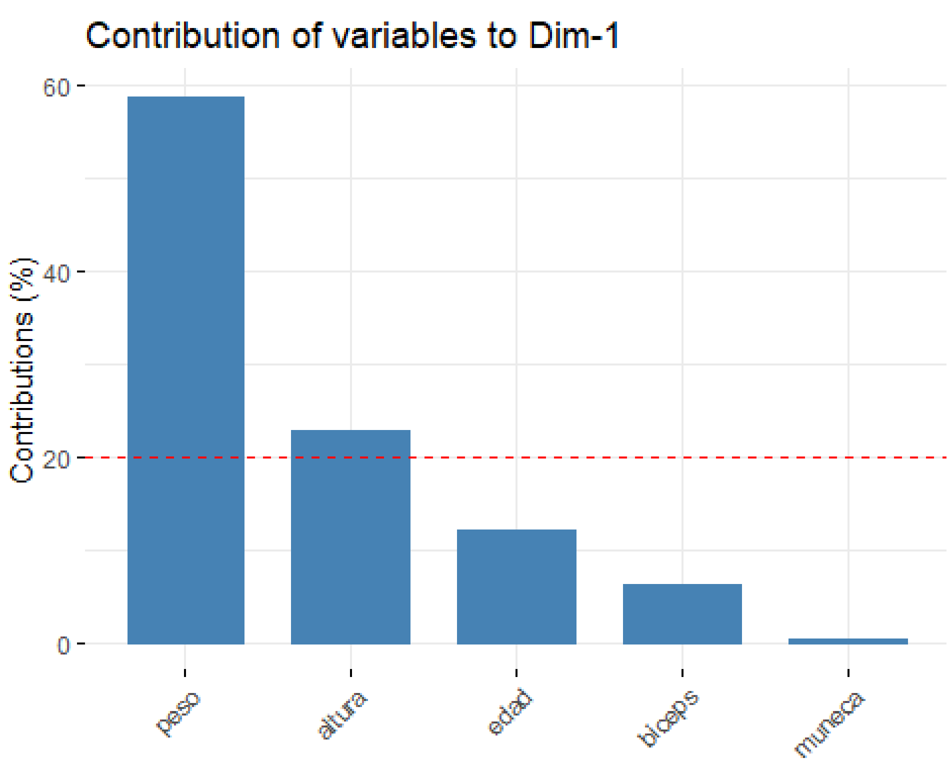

```
fviz_pca_var(cpS, col.var = "red", addEllipses = TRUE, repel = TRUE)
```



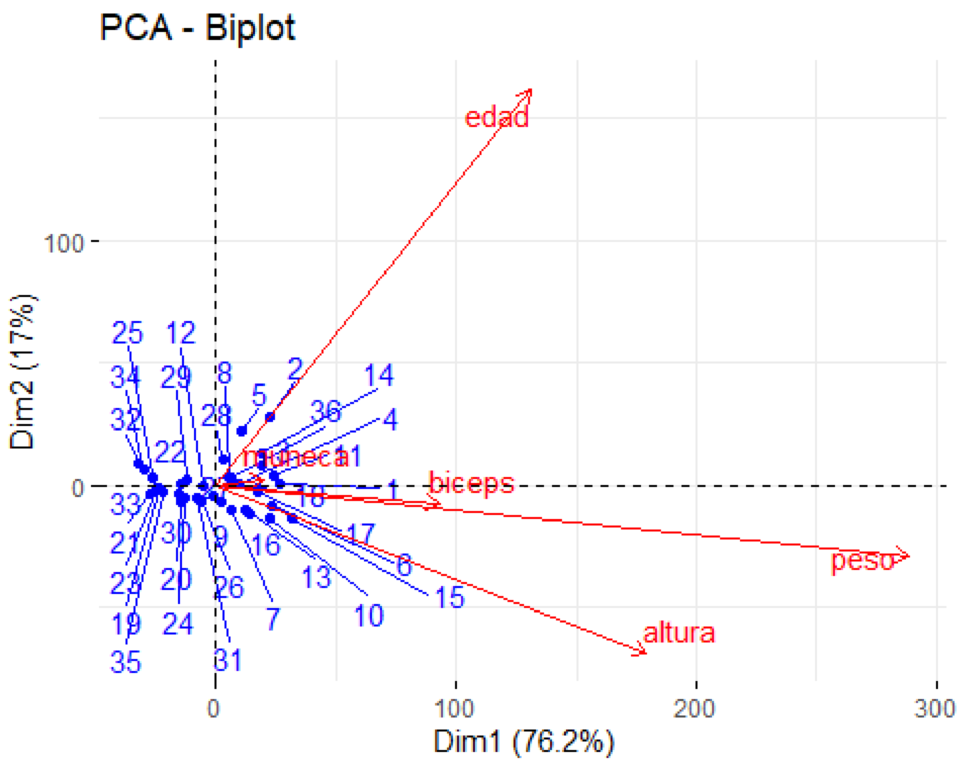
```r
fviz_screeplot(cpS)
```



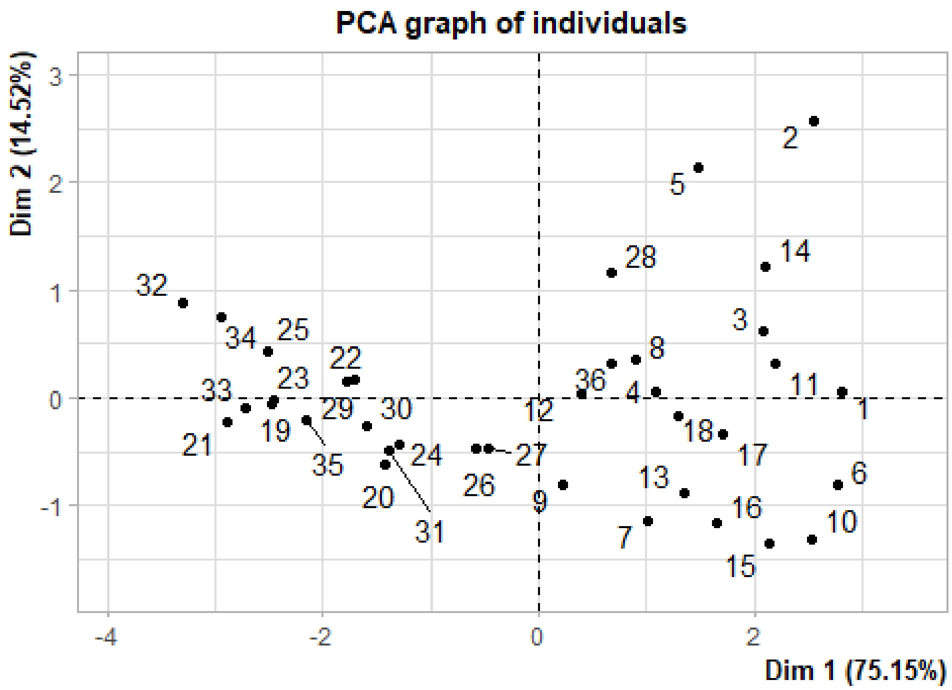

```r
fviz_contrib(cpS, choice = "var")
```

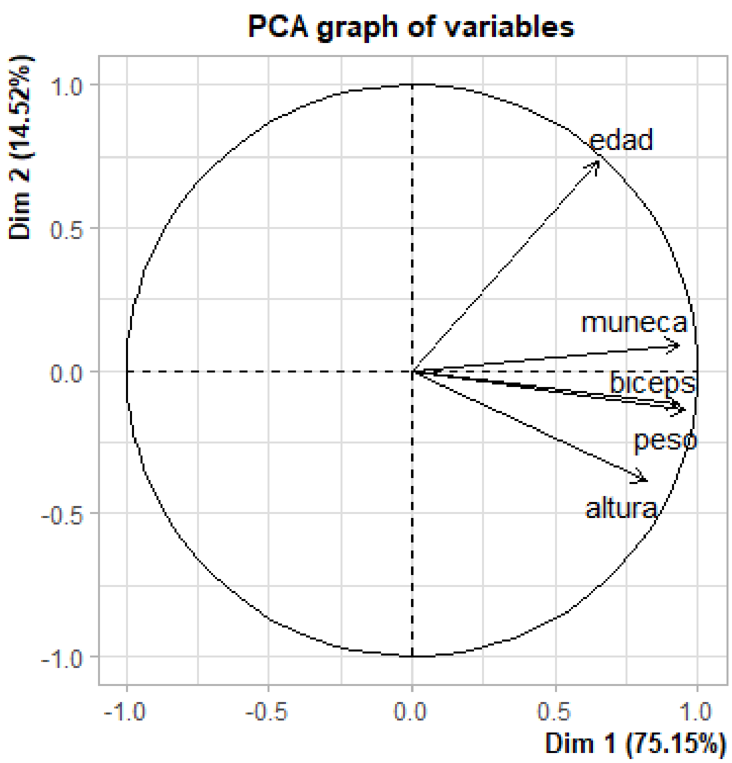


```r
fviz_pca_biplot(cpS, repel=TRUE, col.var="red", col.ind="blue")
```

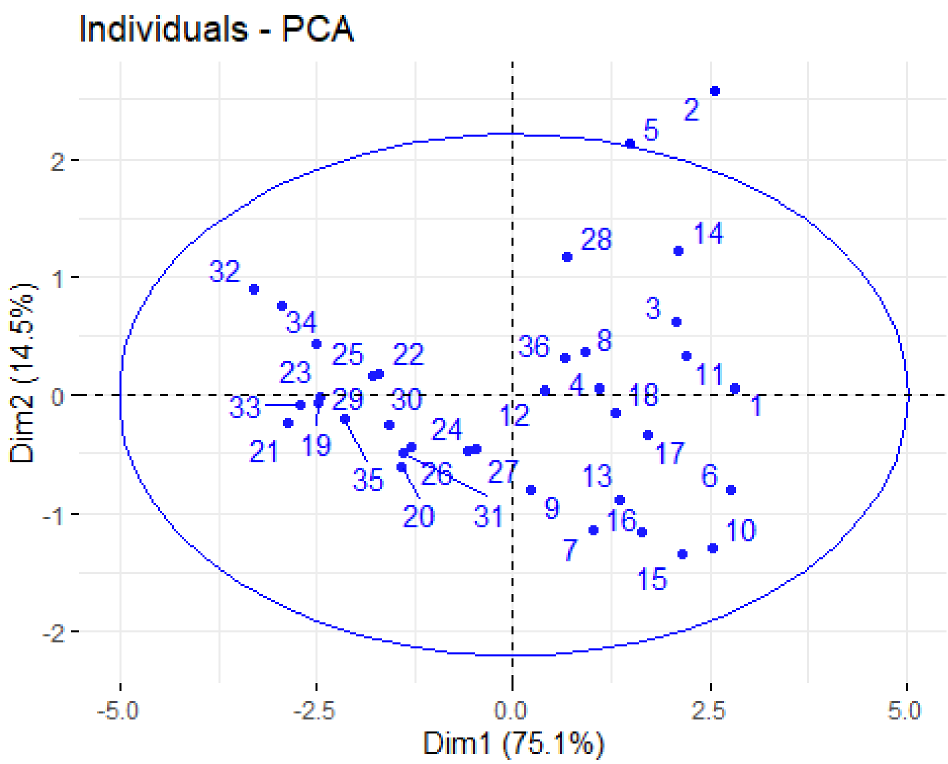



```r
library(FactoMineR)
library(ggplot2)
datos=M
cpS = PCA(datos,scale.unit=TRUE) #Para matriz de correlaciones usa scale.unit=TRUE
```

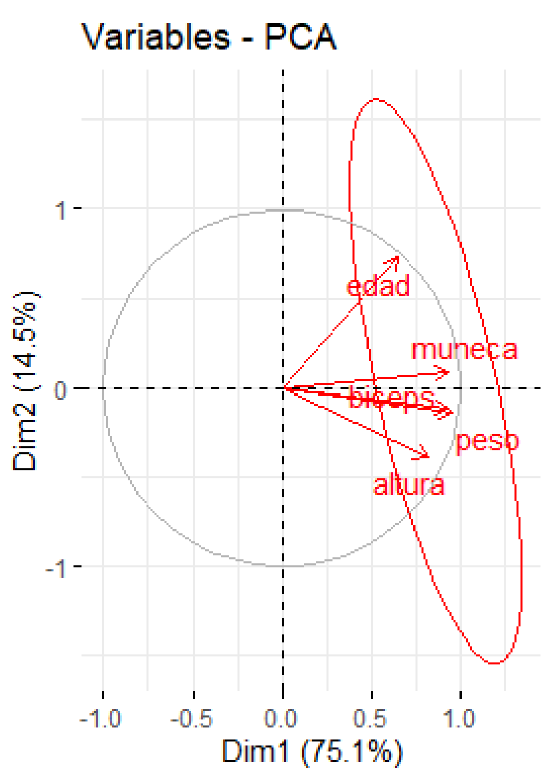




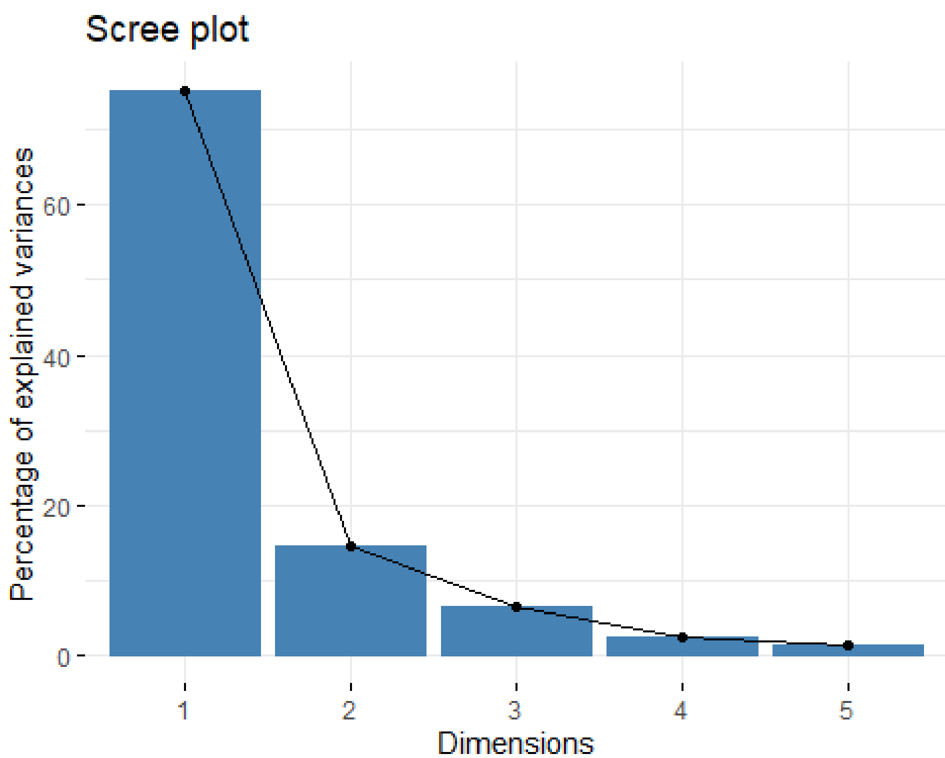

```
library(factoextra)
fviz_pca_ind(cpS, col.ind = "blue", addEllipses = TRUE, repel = TRUE)
```



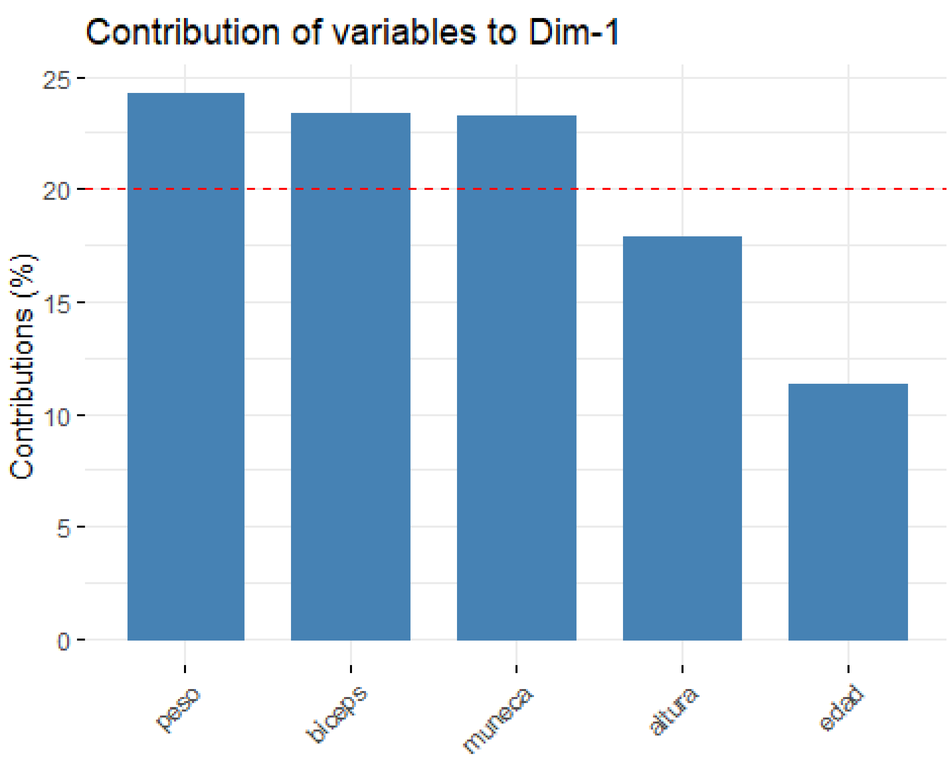

```r
fviz_pca_var(cpS, col.var = "red", addEllipses = TRUE, repel = TRUE)
```



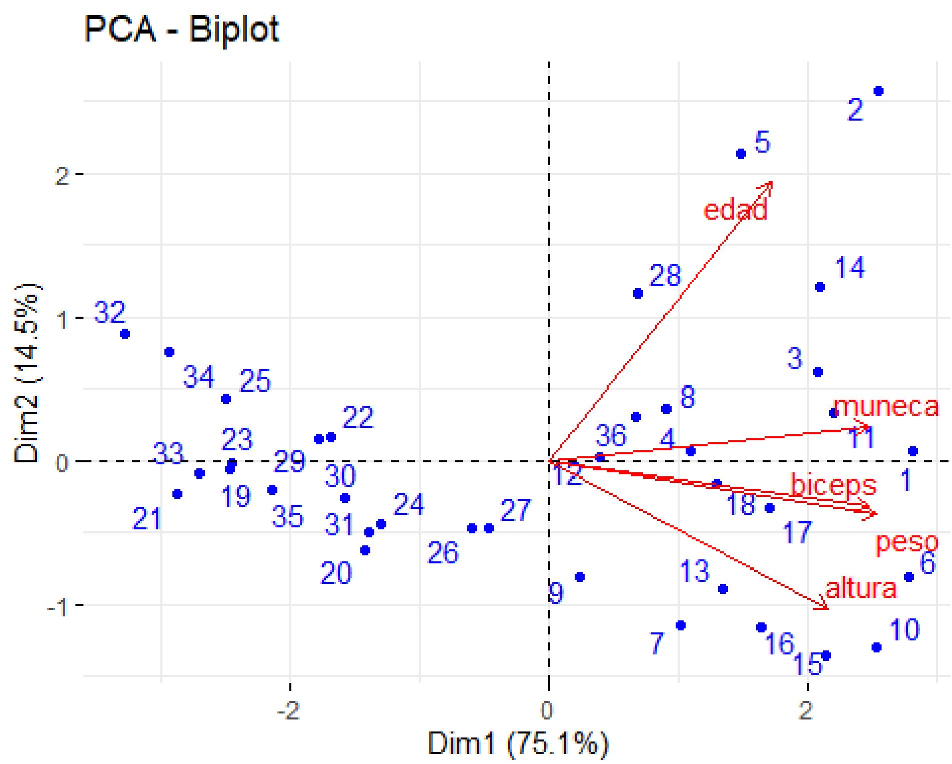

```r
fviz_screeplot(cpS)
```




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




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



##PARTE IV Finalmente: Concluye sobre el análisis de componentes principales realizado e interprete los resultados.

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? Para el data set que estamos usando, el analisis basado en la matriz de correlacion es el mejor, ya que este tiene que todas las variables se usen de manera correcta y equitativa.

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. El analisis de la matriz de correlacion es el más indicado, ya que los indicadores estan en diferentes escalas y al usar la matriz de correlacion todas las varaibles se usan correctamente, lo que hace que se pueda hacer una comparacion correcta. Ademas que los resultados de la matriz de correlacion son más equilibrados que el de matriz de varianza.

¿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 al componente principal son las relacionadas con los indicadores economicos como el PIB per capital, la inversion de la infraestructura y el nivel de industrializacion, estas variables reflejan el desarrollo economico del pais.

Para el segundo componente las variables que influyen más son los indicadores sociales, como la tasa de alfabetizacion, la esperanza de vida y la tasa de mortalidad infantil, estas variables reflejan el bienestar social.

Escriba las combinaciones finales que se recomiendan para hacer el análisis de componentes principales. Para el componente principal se puede interpretar como un indice de riqueza/desarrollo economico ya que esta relacionado con las variables economicas

Para el segundo componente se puede interpretar como un indice de binestar social ya que esta relacionado con las variables de la salud y de la educacion.

Interpreta los resultados en término de agrupación de variables (puede ayudar “índice de riqueza”, “índice de ruralidad”, etc) Los resultados del analisis muestran que los paises se pueden agrupar en desarrollo social y economico, ya que mediante el primer componente podemos obtener que paises se encuentran mejor en el desarrollo economico y cuales son los que tienen un menor, al igual con el segundo componente podemos agrupar por el bienestar social indicando cuales paises tienen un mejor sistema educativo o un mejor sector de salud.