ComponentesPrincipales

Cleber Perez

2024-10-08

Leer archivo

```
C = read.csv("corporal.csv")
C = subset(C, select = -sexo)
summary(C)
```

```
##       edad            peso           altura          muneca      
##  Min.   :19.00   Min.   :42.00   Min.   :147.2   Min.   : 8.300  
##  1st Qu.:24.75   1st Qu.:54.95   1st Qu.:164.8   1st Qu.: 9.475  
##  Median :28.00   Median :71.50   Median :172.7   Median :10.650  
##  Mean   :31.44   Mean   :68.95   Mean   :171.6   Mean   :10.467  
##  3rd Qu.:37.00   3rd Qu.:82.40   3rd Qu.:179.4   3rd Qu.:11.500  
##  Max.   :65.00   Max.   :98.20   Max.   :190.5   Max.   :12.400  
##      biceps     
##  Min.   :23.50  
##  1st Qu.:25.98  
##  Median :32.15  
##  Mean   :31.17  
##  3rd Qu.:35.05  
##  Max.   :40.40
```

```
head(C)
```

```
##   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
```

Parte 1

matrices de varianza-covarianza S y la matriz de correlaciones R

```
S = cov(C)
S
```

```
##              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
```

```r
R = cor(C)
R
```

```
##             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
```

calcula los valores y vectores propios

```r
# Para la matriz de varianza-covarianza
eigen_S = eigen(S)
eigen_S
```

```
## eigen() decomposition
## $values
## [1] 359.3980243  80.3757858  27.6229011   4.3074318   0.2343571
##
## $vectors
##             [,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
valores_propios_S = eigen_S$values
valores_propios_S
```

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

```r
vectores_propios_S = eigen_S$vectors
vectores_propios_S
```

```
##             [,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
# Para la matriz de correlaciones
eigen_R = eigen(R)
eigen_R
```

```
## eigen() decomposition
## $values
## [1] 3.75749733 0.72585665 0.32032981 0.12461873 0.07169749
##
## $vectors
```

```
##             [,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
```

```r
valores_propios_R = eigen_R$values
valores_propios_R
```

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

```r
vectores_propios_R = eigen_R$vectors
vectores_propios_R
```

```
##             [,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
```

Calcular la proporcion de varianza explicada

```r
varianza_total_S = sum(diag(S))
proporcion_varianza_S = valores_propios_S / varianza_total_S
varianza_acumulada_S = cumsum(proporcion_varianza_S)

varianza_total_R = sum(diag(R))
proporcion_varianza_R = valores_propios_R / varianza_total_R
varianza_acumulada_R = cumsum(proporcion_varianza_R)

data.frame(
  Componentes = 1:length(valores_propios_S),
  Proporcion_Varianza = proporcion_varianza_S,
  Varianza_Acumulada = varianza_acumulada_S
)
```

```
##   Componentes Proporcion_Varianza Varianza_Acumulada
## 1           1        0.7615357176          0.7615357
## 2           2        0.1703098726          0.9318456
## 3           3        0.0585307219          0.9903763
## 4           4        0.0091271040          0.9995034
## 5           5        0.0004965839          1.0000000
```

```r
data.frame(
  Componentes = 1:length(valores_propios_R),
  Proporcion_Varianza = proporcion_varianza_R,
  Varianza_Acumulada = varianza_acumulada_R
)
```

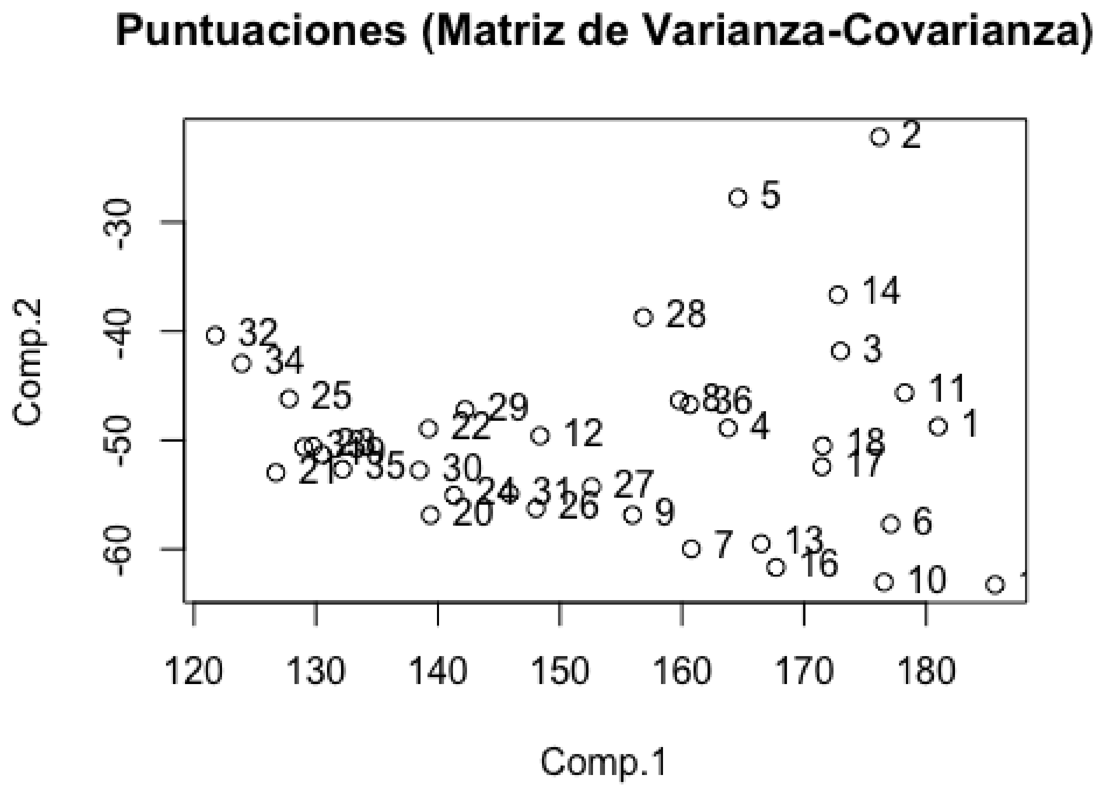
```
##   Componentes Proporcion_Varianza Varianza_Acumulada
## 1           1          0.75149947          0.7514995
## 2           2          0.14517133          0.8966708
## 3           3          0.06406596          0.9607368
## 4           4          0.02492375          0.9856605
## 5           5          0.01433950          1.0000000
```

Componentes mas importantes

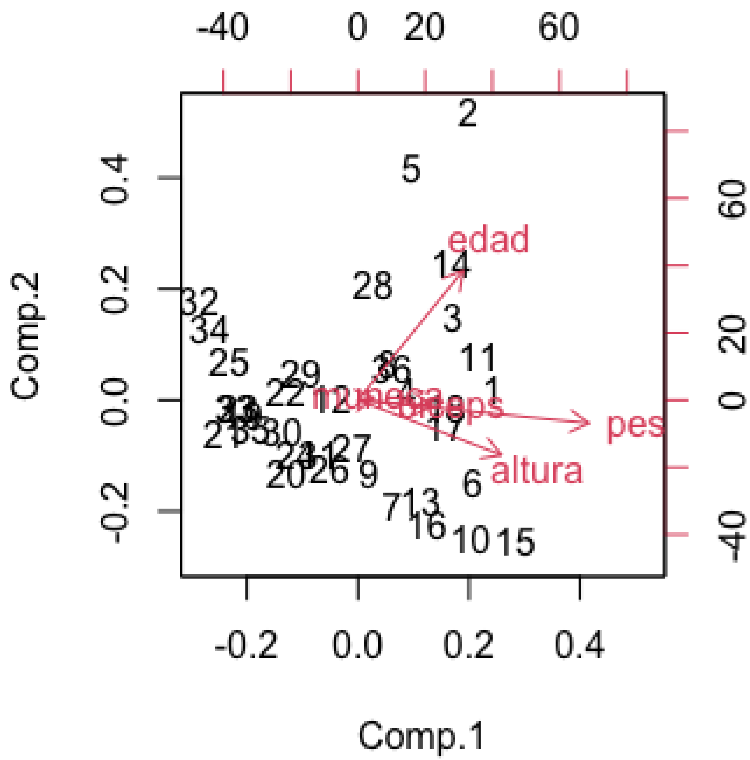
```r
coeficientes_CP1 <- eigen_S$vectors[, 1]
coeficientes_CP2 <- eigen_S$vectors[, 2]

contribucion_CP1 <- abs(coeficientes_CP1)
contribucion_CP2 <- abs(coeficientes_CP2)

variables_CP1 <- names(C)[order(contribucion_CP1, decreasing = TRUE)]
variables_CP2 <- names(C)[order(contribucion_CP2, decreasing = TRUE)]

top_variables_CP1 <- head(variables_CP1, n = 3)  # Las 3 variables que más
contribuyen a CP1
top_variables_CP1

## [1] "peso"   "altura" "edad"

top_variables_CP2 <- head(variables_CP2, n = 3)  # Las 3 variables que más
contribuyen a CP2
top_variables_CP2

## [1] "edad"   "altura" "peso"
```

Grafica de Varianza-Covarianza

```r
datos = C

cpS = princomp(datos, cor = FALSE)
cpaS = as.matrix(datos) %*% cpS$loadings

plot(cpaS[, 1:2], type = "p", main = "Puntuaciones (Matriz de Varianza-
Covarianza)")
text(cpaS[, 1], cpaS[, 2], 1:nrow(cpaS), pos = 4)
```

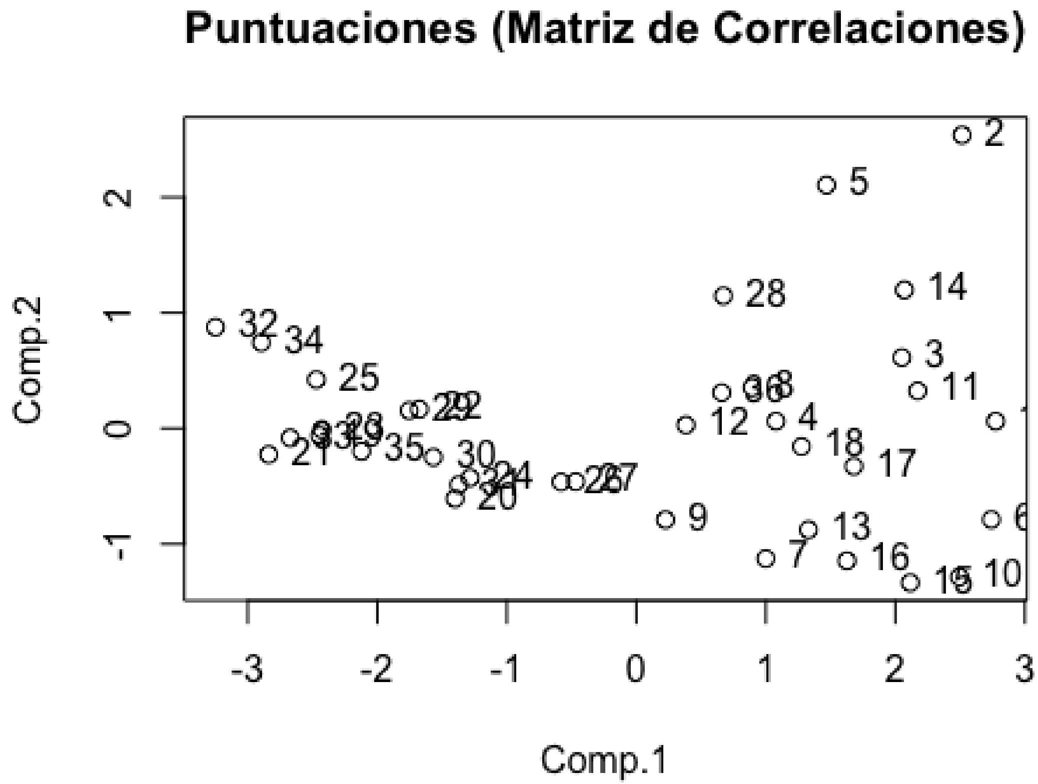

Puntuaciones (Matriz de Varianza-Covarianza)
Comp.2
-30
-40
-50
-60
120
130
140
150
160
170
180
Comp.1
2
5
14
28
3
32
34
25
11
29
8
36
4
1
22
12
18
17
21
35
30
27
24
31
26
9
20
6
7
13
16
10

```r
biplot(cpS)
```

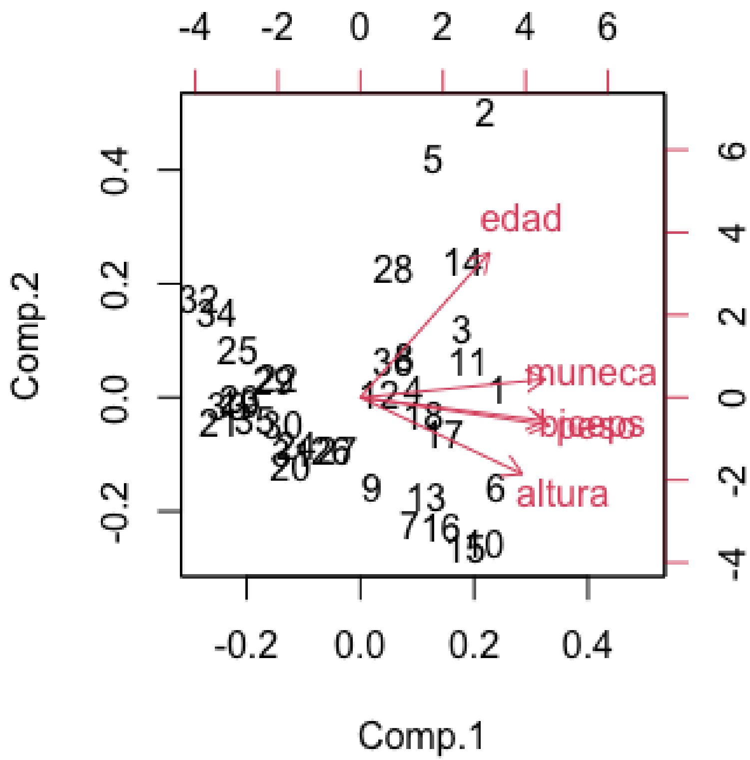


Grafica Matriz de Correlaciones

```
datos_estandarizado = scale(datos)

cpR = princomp(datos_estandarizado, cor = TRUE)
cpaR = as.matrix(datos_estandarizado) %*% cpR$loadings

plot(cpaR[, 1:2], type = "p", main = "Puntuaciones (Matriz de
Correlaciones)")
text(cpaR[, 1], cpaR[, 2], 1:nrow(cpaR), pos = 4)
```

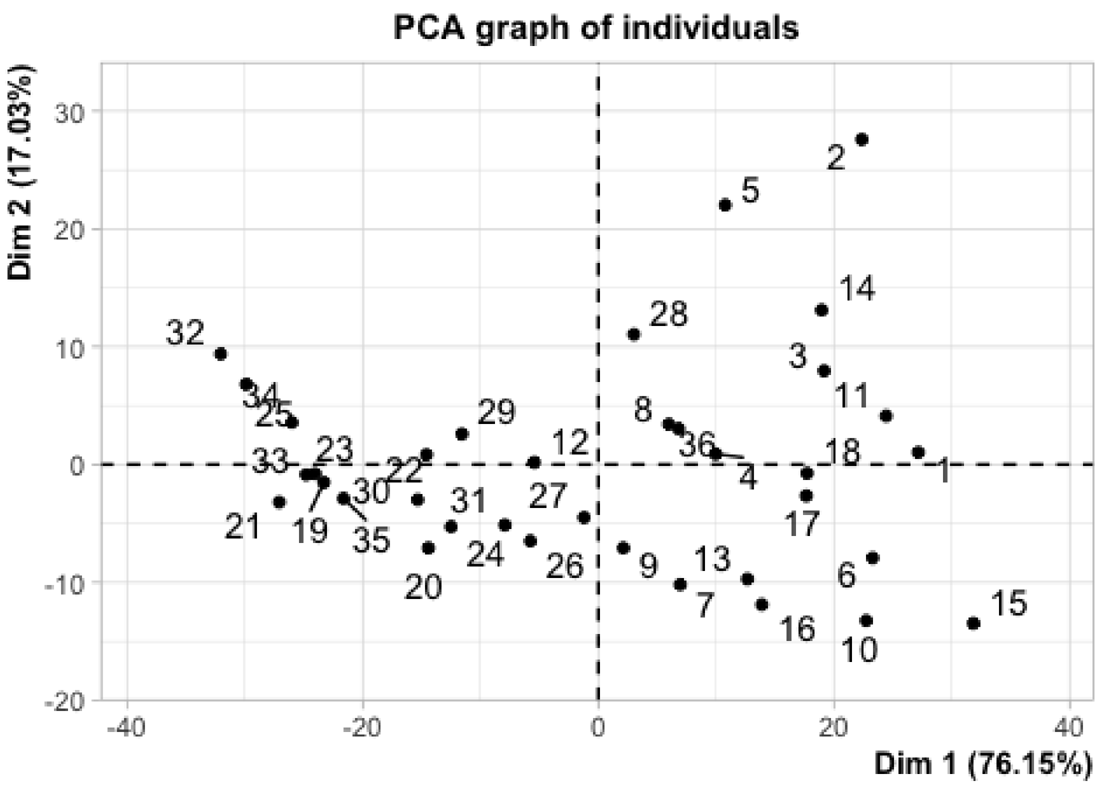


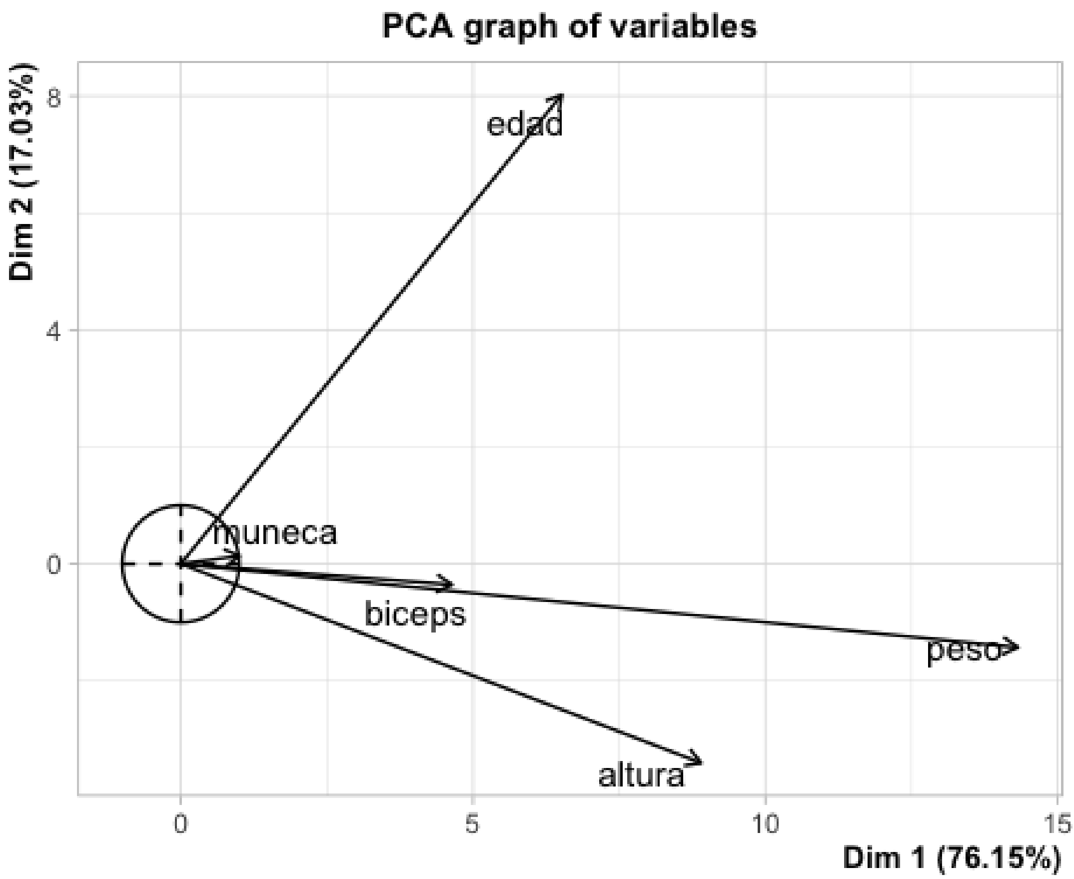

```
biplot(cpR)
```



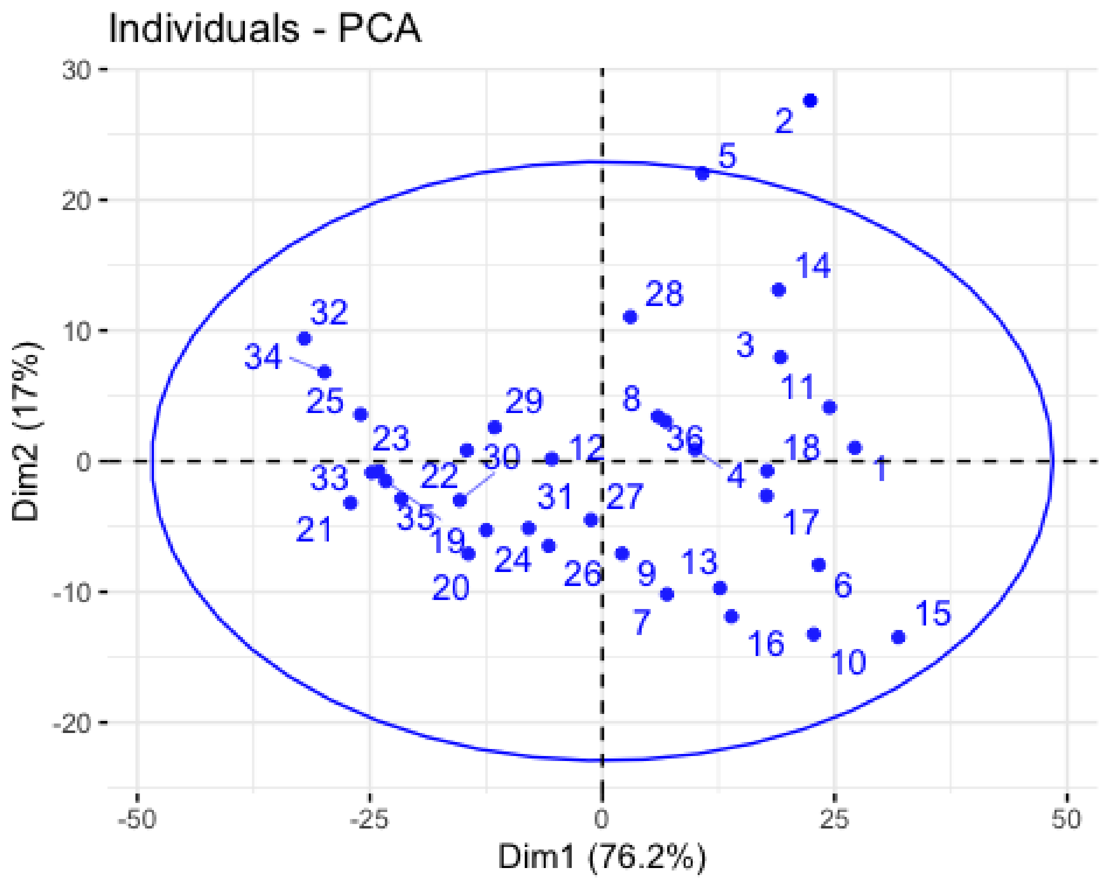
Parte 3

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

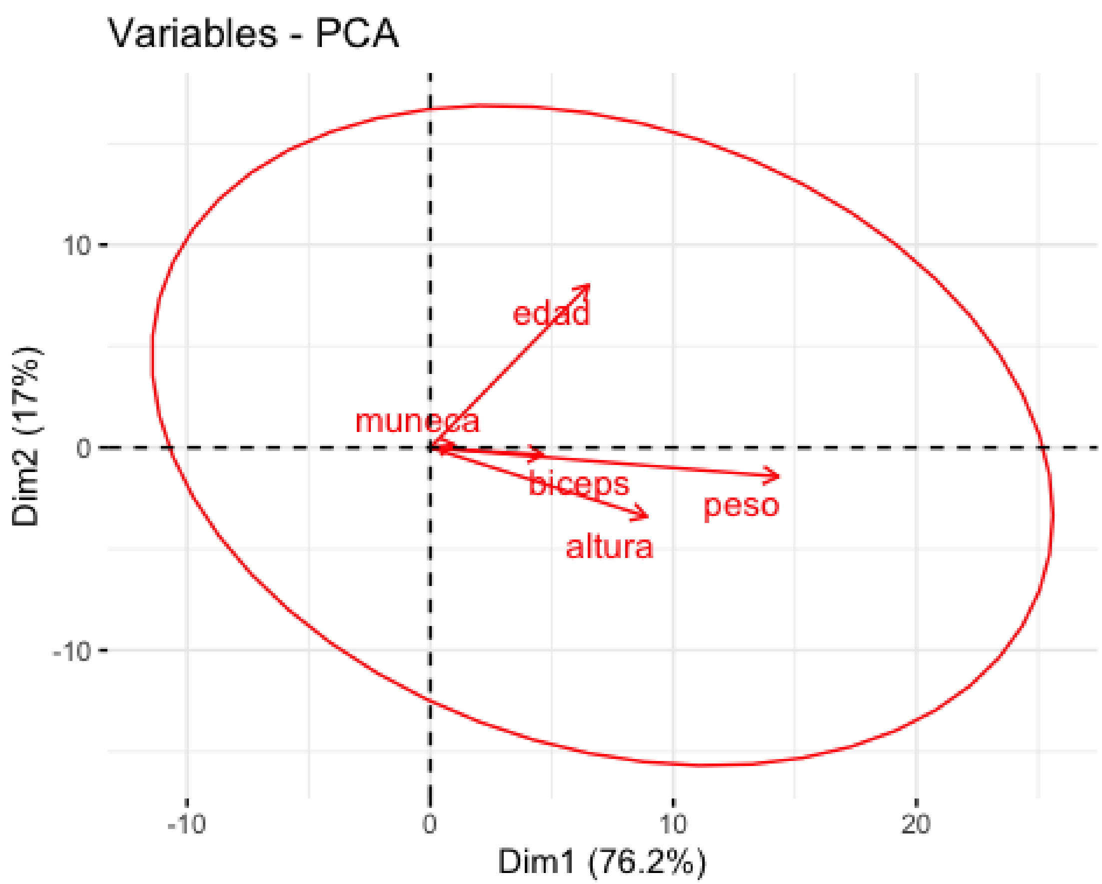

PCA graph of individuals
Dim 1 (76.15%)
Dim 2 (17.03%)
-40
-20
0
20
40
30
20
10
0
-10
-20
32
34
25
33
23
22
21
19
35
30
29
12
31
27
20
24
26
9
13
7
16
10
6
15
17
18
1
4
36
8
11
3
14
28
5
2



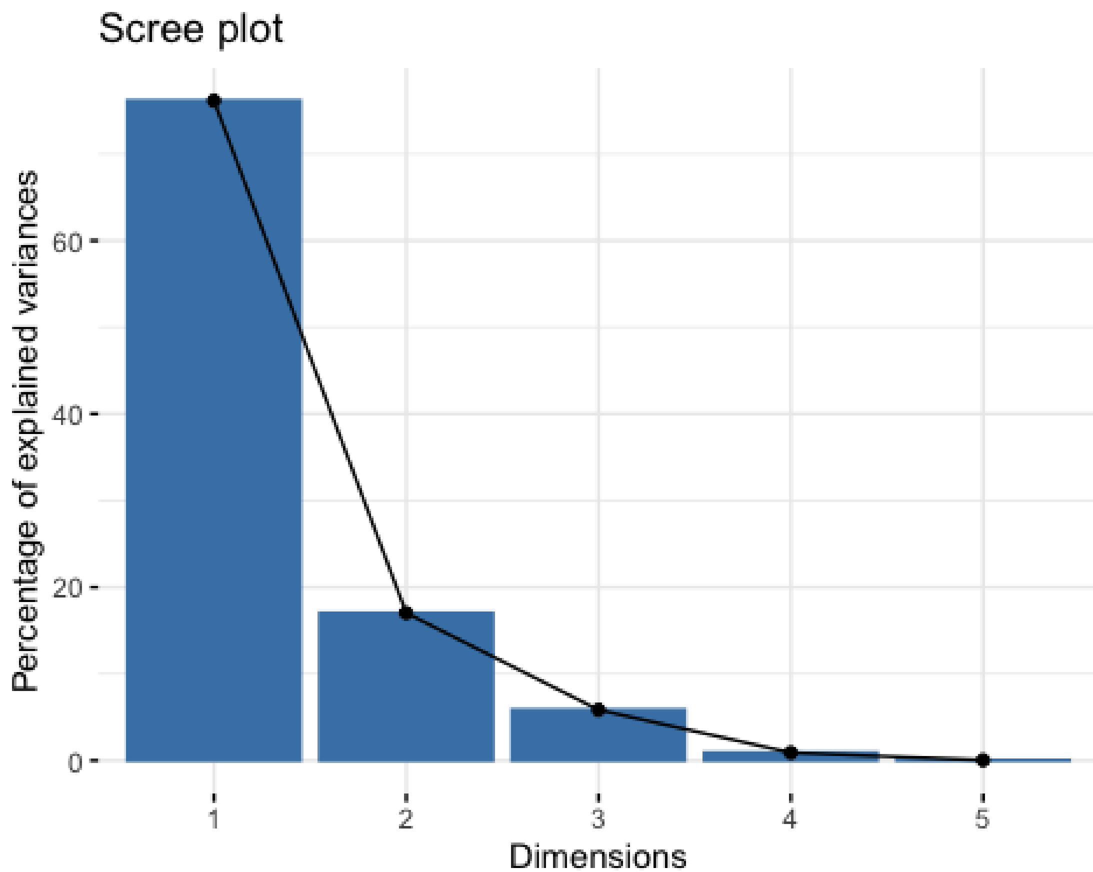

```r
library(factoextra)

## 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)
```



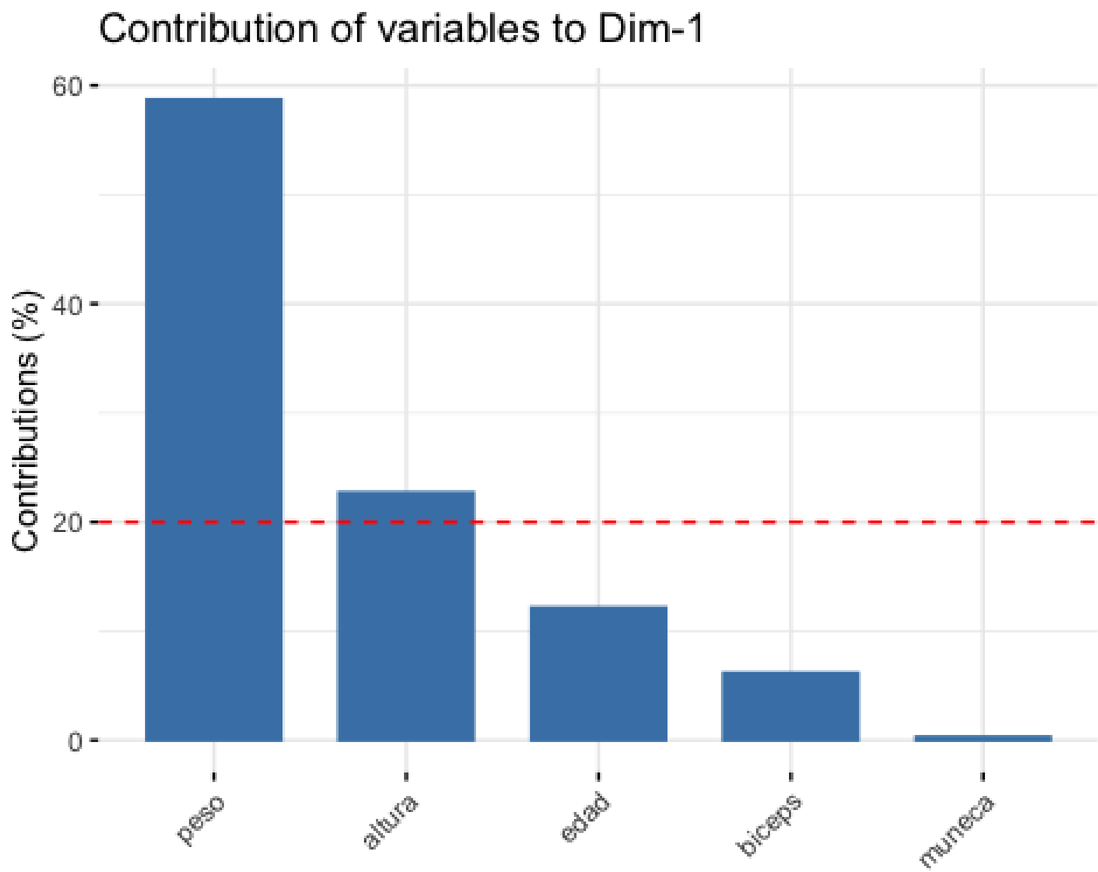

```
fviz_pca_var(cpS, col.var = "red", addEllipses = TRUE, repel = TRUE)
```



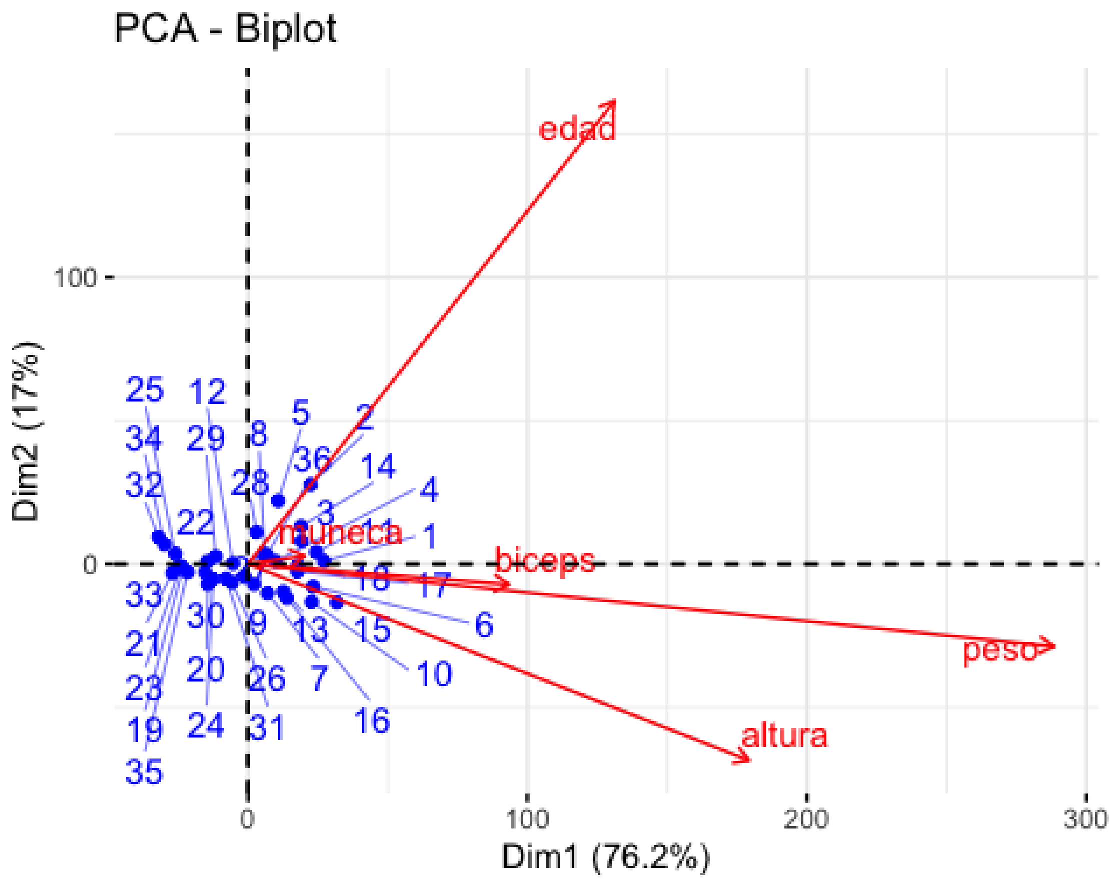

```
fviz_screeplot(cpS)
```



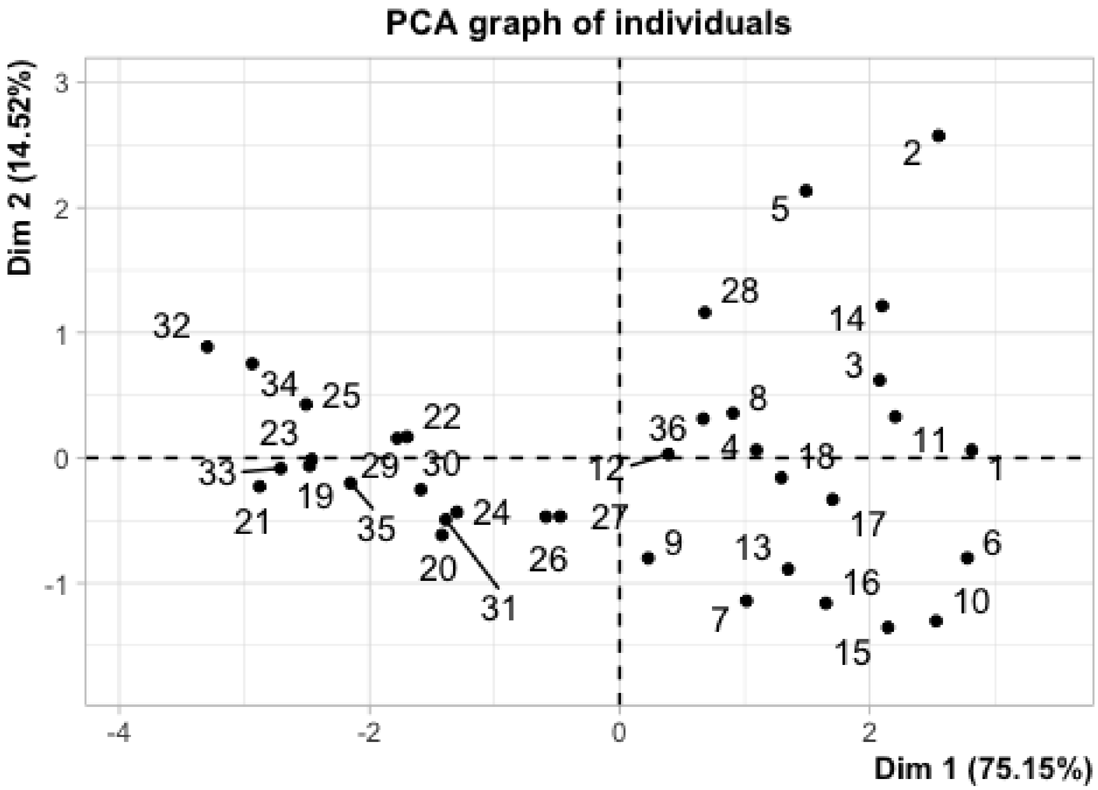

```
fviz_contrib(cpS, choice = c("var"))
```

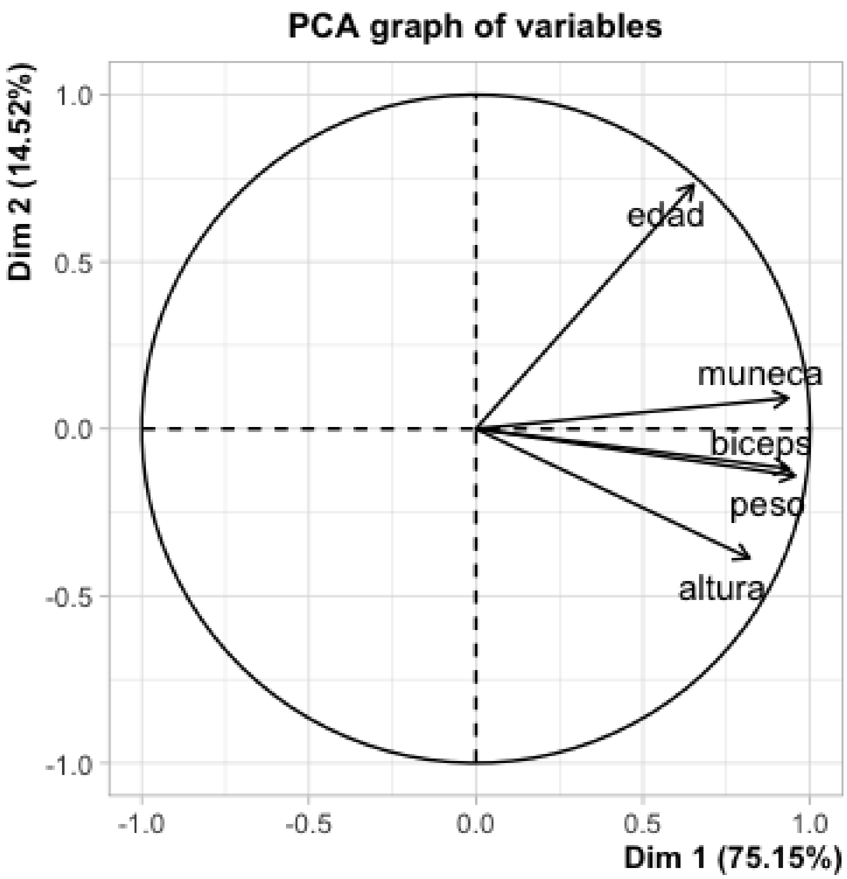



```r
fviz_pca_biplot(cpS, repel=TRUE, col.var="red", col.ind="blue")
```

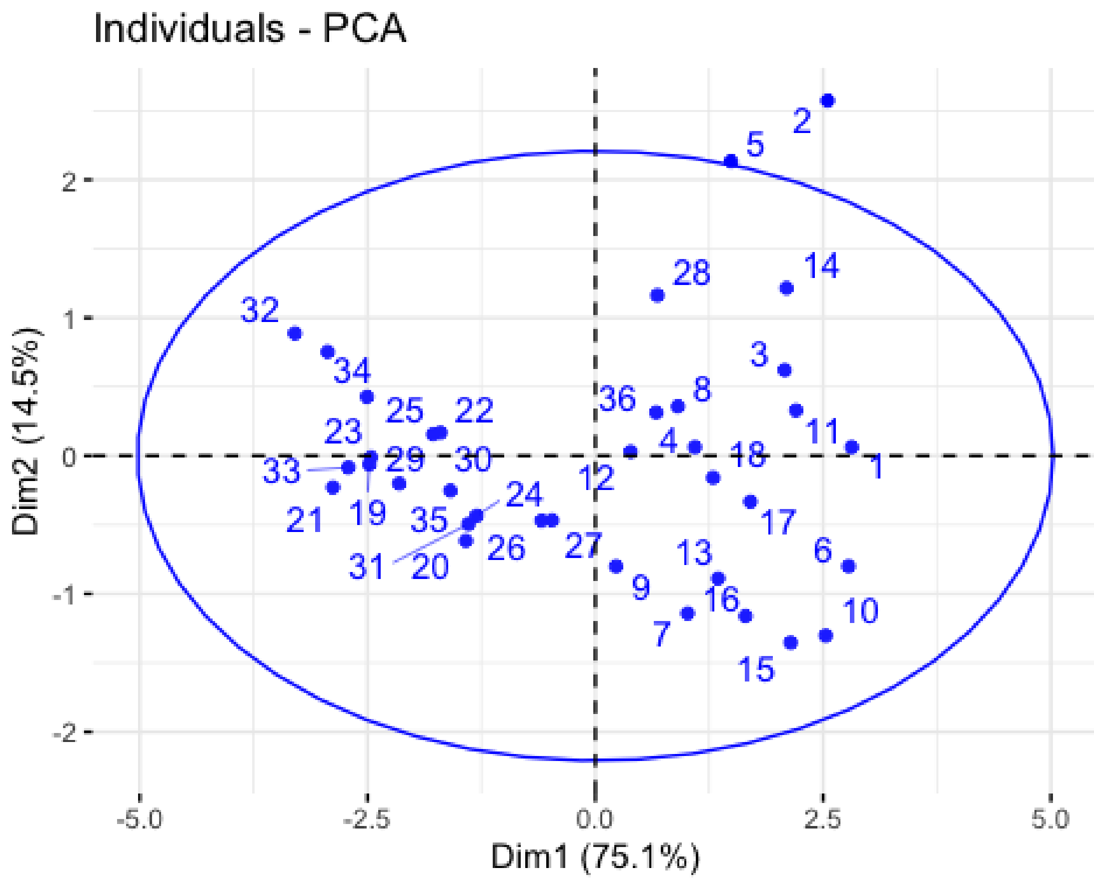



```r
library(FactoMineR)
library(ggplot2)
cpR = PCA(datos,scale.unit=TRUE) #Para matriz de correlaciones usa
scale.unit=TRUE
```

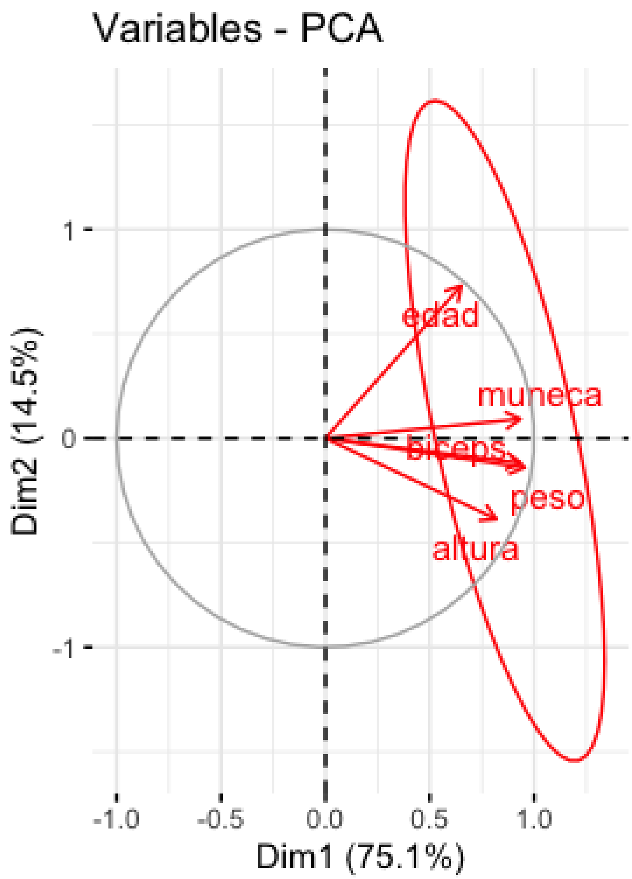

PCA graph of individuals
Dim 2 (14.52%)
Dim 1 (75.15%)
3
2
1
0
-1
-4
-2
0
2
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36



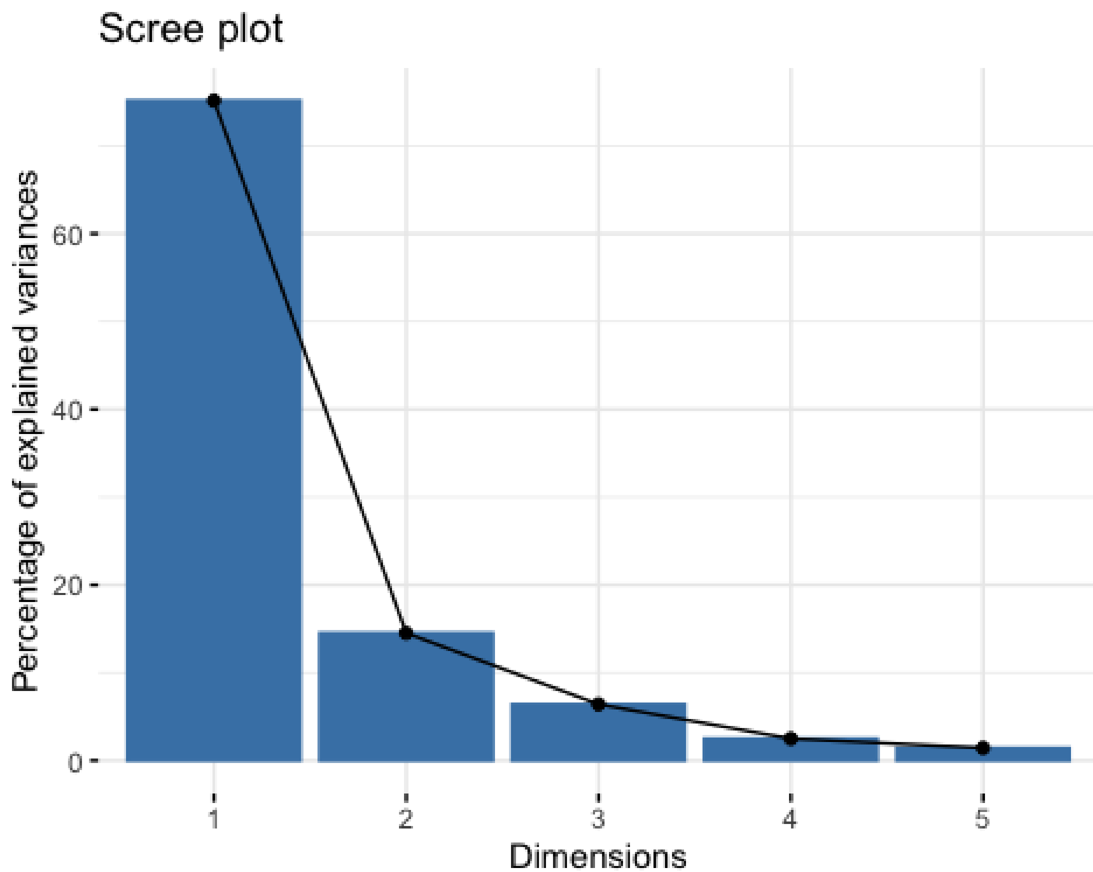

```
library(factoextra)
fviz_pca_ind(cpR, col.ind = "blue", addEllipses = TRUE, repel = TRUE)
```



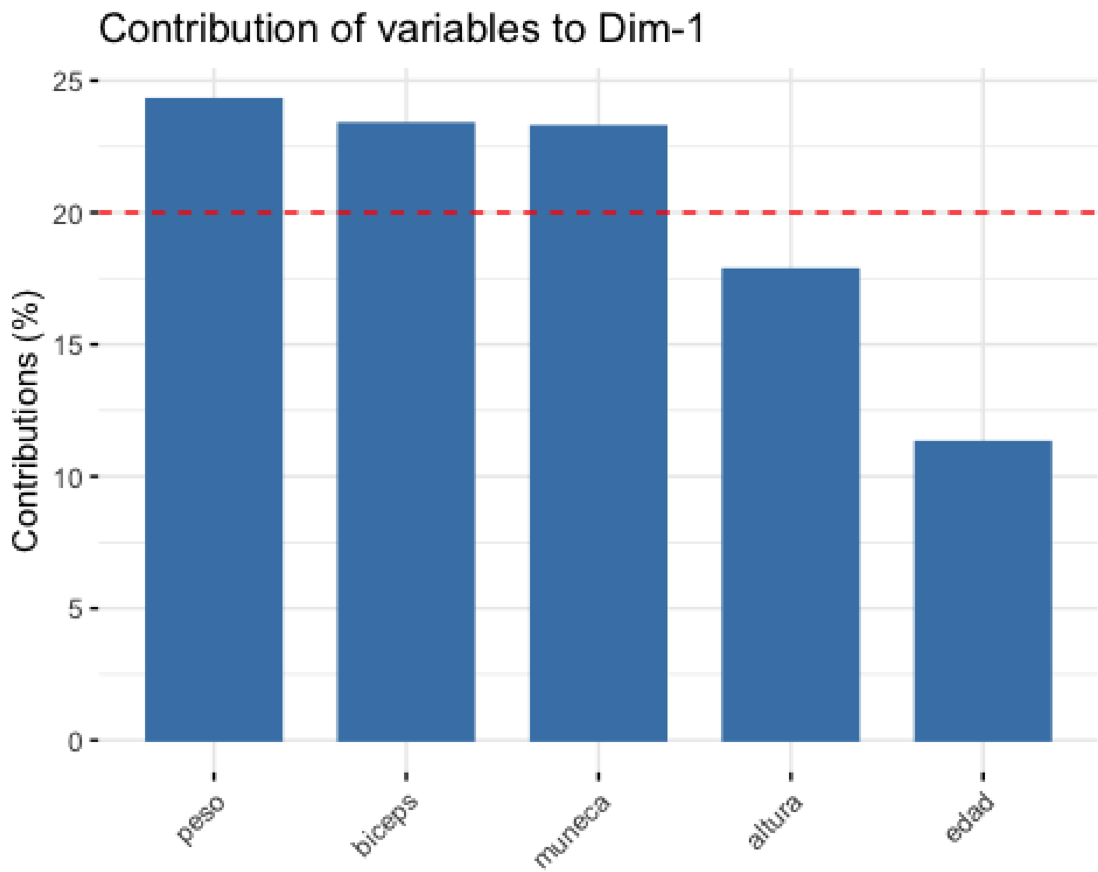

```r
fviz_pca_var(cpR, col.var = "red", addEllipses = TRUE, repel = TRUE)
```



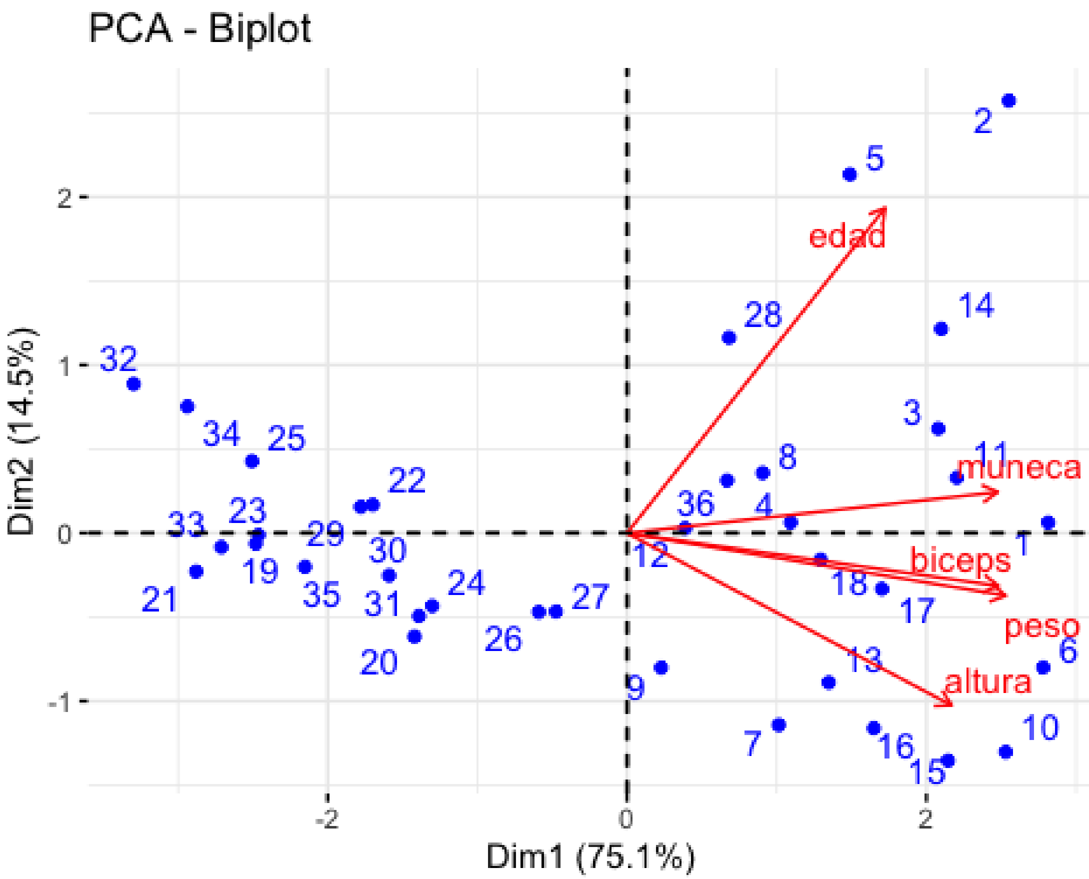

```
fviz_screeplot(cpR)
```




```
fviz_contrib(cpR, choice = c("var"))
```




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



Parte 4

Como podemos observar en ambas matrizes la mayoria de la varianza se explica dentro de los dos primeros componentes, un 76 y 17% para la matriz de varianza-covarianza y 75 y 14% en la matriz de correlacion, pero en ambas la mayor explicacion se encuentra dentro de la matriz de varianza-covarianza.

La mas eficaz para este problema es la Matriz de varianza-covarancia.

Para el primer componente principal, las variables que mas contribuyen son: peso, altura y edad. Para el segundo componente principal, las variables que mas contribuyen son: edad, altura y peso. El hecho que estas variables se repitan en ambos casos indica su gran correlacion y son sumamente importantes para entender la variabilidad.

Para datos de indicadores economicos y sociales de los paises, el analisis basado en la matriz de correlacion es generalmente el mas adecuado, debido a la diversidad de escalas y unidades en las variables. Esto permite que todas las variables tengan un peso equitativo en la identificación de las componentes principales.