Untitled

2024-10-11

```
library(FactoMineR)
library(ggplot2)
library(factoextra)
```

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

```
data <- read.csv("documents/corporal.csv")

head(data)
```

```
##   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
```

```
numeric_data <- data[, sapply(data, is.numeric)]

S <- cov(numeric_data)

print("Matriz de varianza-covarianza S:")
```

```
## [1] "Matriz de varianza-covarianza S:"
```

```
print(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 <- cor(numeric_data)

print(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
```

```
eigen_S <- eigen(S)

autovectores_S <- eigen_S$vectors
```

```r
print(eigen_S$values)
```

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

```r
print(eigen_S$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
eigen_R <- eigen(R)

print(eigen_R$values)
```

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

```r
print(eigen_R$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
varianza_total_S <- sum(diag(S))

suma_valores_propios_S <- sum(eigen_S$values)

cat("Varianza total (suma de la diagonal de S):", varianza_total_S, "\n")
```

```
## Varianza total (suma de la diagonal de S): 471.9385
```

```r
cat("Suma de los valores propios de S:", suma_valores_propios_S, "\n")
```

```
## Suma de los valores propios de S: 471.9385
```

```r
proporcion_varianza_S <- eigen_S$values / varianza_total_S

cat("Proporción de varianza explicada por cada componente en S:\n")
```

```
## Proporción de varianza explicada por cada componente en S:
```

```r
print(proporcion_varianza_S)
```

```
## [1] 0.7615357176 0.1703098726 0.0585307219 0.0091271040 0.0004965839
```

```r
varianza_total_R <- sum(diag(R))

suma_valores_propios_R <- sum(eigen_R$values)

cat("Varianza total (suma de la diagonal de R):", varianza_total_R, "\n")
```

```
## Varianza total (suma de la diagonal de R): 5
```

```r
cat("Suma de los valores propios de R:", suma_valores_propios_R, "\n")
```

```
## Suma de los valores propios de R: 5
```

```r
proporcion_varianza_R <- eigen_R$values / varianza_total_R

cat("Proporción de varianza explicada por cada componente en R:\n")
```

```
## Proporción de varianza explicada por cada componente en R:
```

```r
print(proporcion_varianza_R)
```

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

```r
varianza_acumulada_S <- cumsum(proporcion_varianza_S)

cat("Varianza acumulada para cada componente en S:\n")
```

```
## Varianza acumulada para cada componente en S:
```

```r
print(varianza_acumulada_S)
```

```
## [1] 0.7615357 0.9318456 0.9903763 0.9995034 1.0000000
```

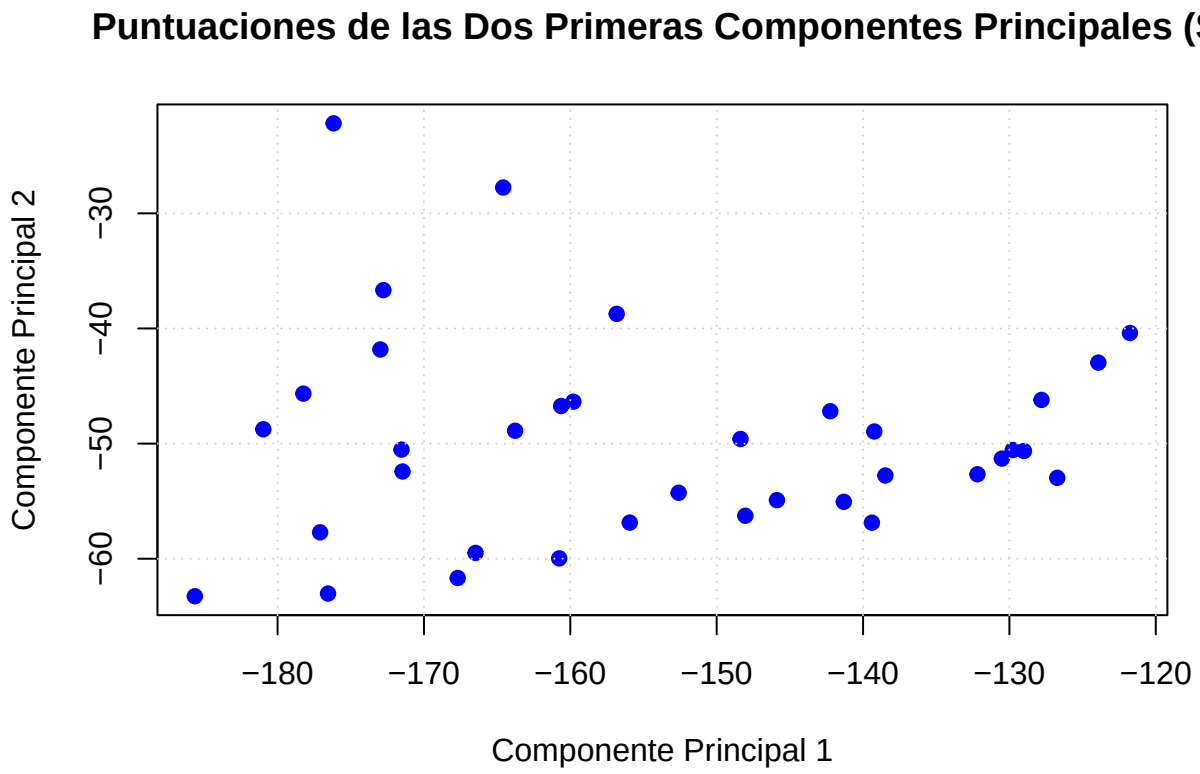
```r
varianza_acumulada_R <- cumsum(proporcion_varianza_R)

cat("Varianza acumulada para cada componente en R:\n")
```

```
## Varianza acumulada para cada componente en R:
```

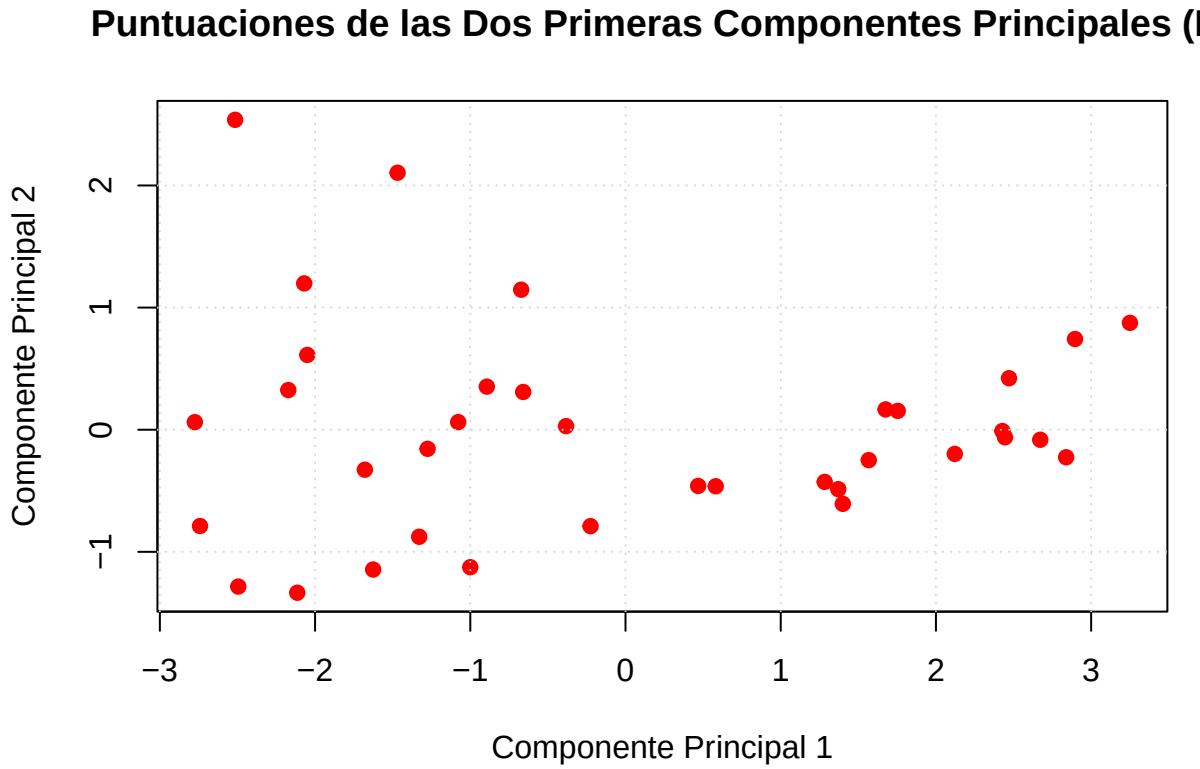
```r
print(varianza_acumulada_R)
```

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

```r
scores_S <- as.matrix(numeric_data) %*% autovectores_S[, 1:2]

numeric_data_standardized <- scale(numeric_data)

autovectores_R <- eigen_R$vectors

scores_R <- as.matrix(numeric_data_standardized) %*% autovectores_R[, 1:2]

plot(scores_S, main="Puntuaciones de las Dos Primeras Componentes Principales (S)",
     xlab="Componente Principal 1", ylab="Componente Principal 2", pch=19, col="blue")
grid()
```




```
plot(scores_R, main="Puntuaciones de las Dos Primeras Componentes Principales (R)",
     xlab="Componente Principal 1", ylab="Componente Principal 2", pch=19, col="red")
grid()
```



```r
cpS <- princomp(numeric_data, cor = FALSE) # cor = FALSE usa la matriz de covarianzas

cpR <- princomp(numeric_data, cor = TRUE)

summary(cpS)
```

```
## 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
```

```r
summary(cpR)
```

```
## 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
```

```r
cpS$loadings
```

```
##
## Loadings:
##        Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
## edad    0.349  0.908  0.232
## peso    0.766 -0.162 -0.522  0.339
## altura  0.476 -0.385  0.789
## muneca                      -0.126 -0.990
## biceps  0.248        -0.225 -0.931  0.138
##
##                Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
## SS loadings       1.0    1.0    1.0    1.0    1.0
## Proportion Var    0.2    0.2    0.2    0.2    0.2
## Cumulative Var    0.2    0.4    0.6    0.8    1.0
```

```r
cpR$loadings
```

```
##
## Loadings:
##        Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
## edad    0.336  0.858  0.349  0.136  0.107
## peso    0.493 -0.165         0.525 -0.671
## altura  0.422 -0.454  0.734 -0.207  0.184
## muneca  0.482  0.108 -0.367 -0.755 -0.226
## biceps  0.483 -0.139 -0.447  0.305  0.674
##
##                Comp.1 Comp.2 Comp.3 Comp.4 Comp.5
## SS loadings       1.0    1.0    1.0    1.0    1.0
## Proportion Var    0.2    0.2    0.2    0.2    0.2
## Cumulative Var    0.2    0.4    0.6    0.8    1.0
```

```r
head(cpS$scores)
```

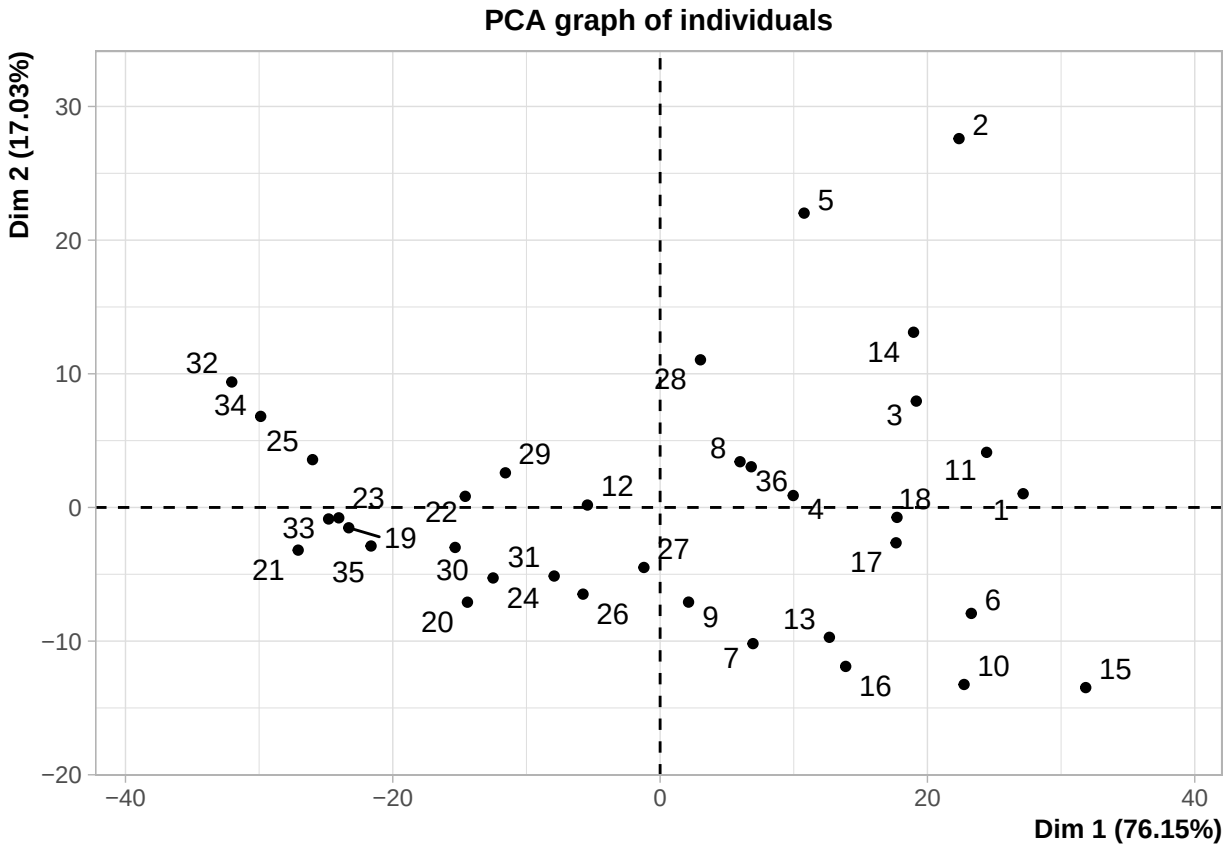
```
##         Comp.1     Comp.2     Comp.3      Comp.4      Comp.5
## [1,] 27.162853  1.0278492  5.0022646  0.93622690 -0.51688356
## [2,] 22.363542 27.5955807  3.0635949 -0.08338126  0.02552809
```

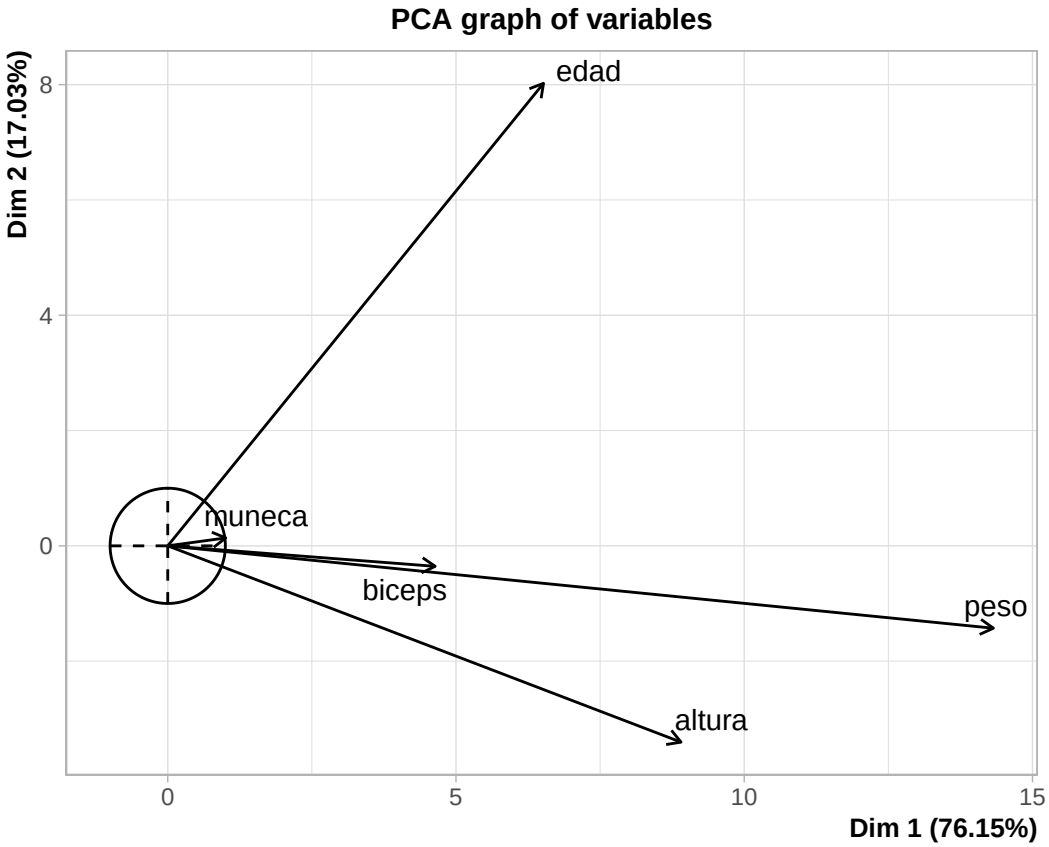
```
## [3,] 19.167874  7.9566157 -1.5770026 -2.61077676  0.80391745
## [4,]  9.959001  0.8923731  5.5146952  0.12345373 -0.35579895
## [5,] 10.775593 22.0203437 -0.7562826  0.17996723 -0.41646606
## [6,] 23.283948 -7.9268214  2.7958617 -2.09339284 -0.62252321
```

```
head(cpR$scores)
```

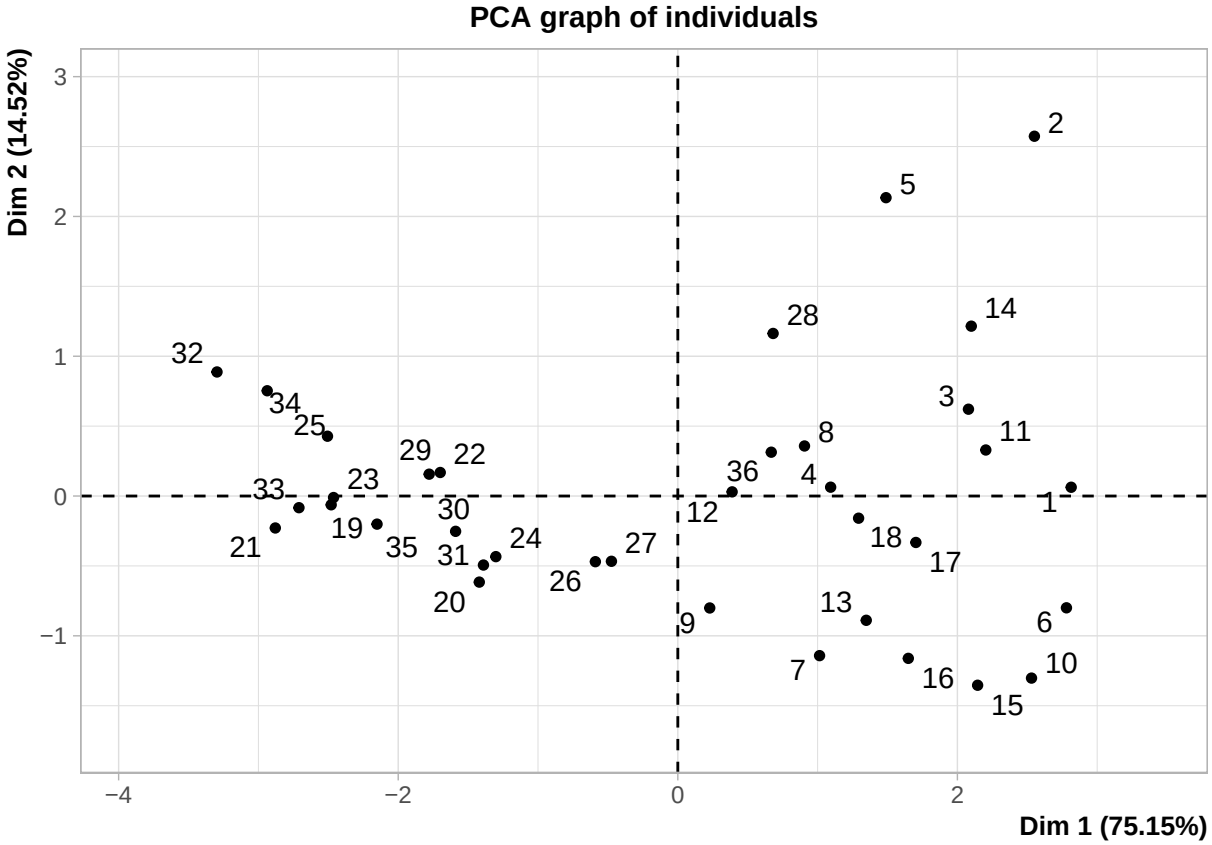
```
##        Comp.1      Comp.2      Comp.3      Comp.4       Comp.5
## [1,] 2.813992  0.06282760  0.51434516 -0.37618363 -0.161649397
## [2,] 2.550816  2.57369731  0.42896223  0.01252075  0.083602262
## [3,] 2.079207  0.62112516 -0.12602006  0.51138786  0.430775853
## [4,] 1.093316  0.06328171  0.46145821 -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
```

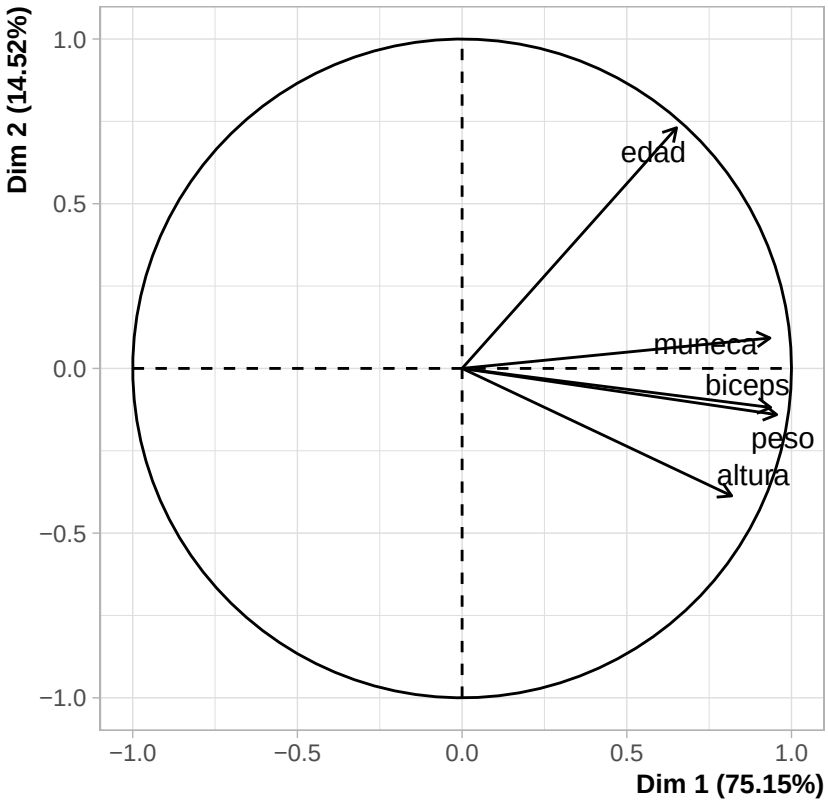
```
cpS <- PCA(numeric_data, scale.unit = FALSE)
```



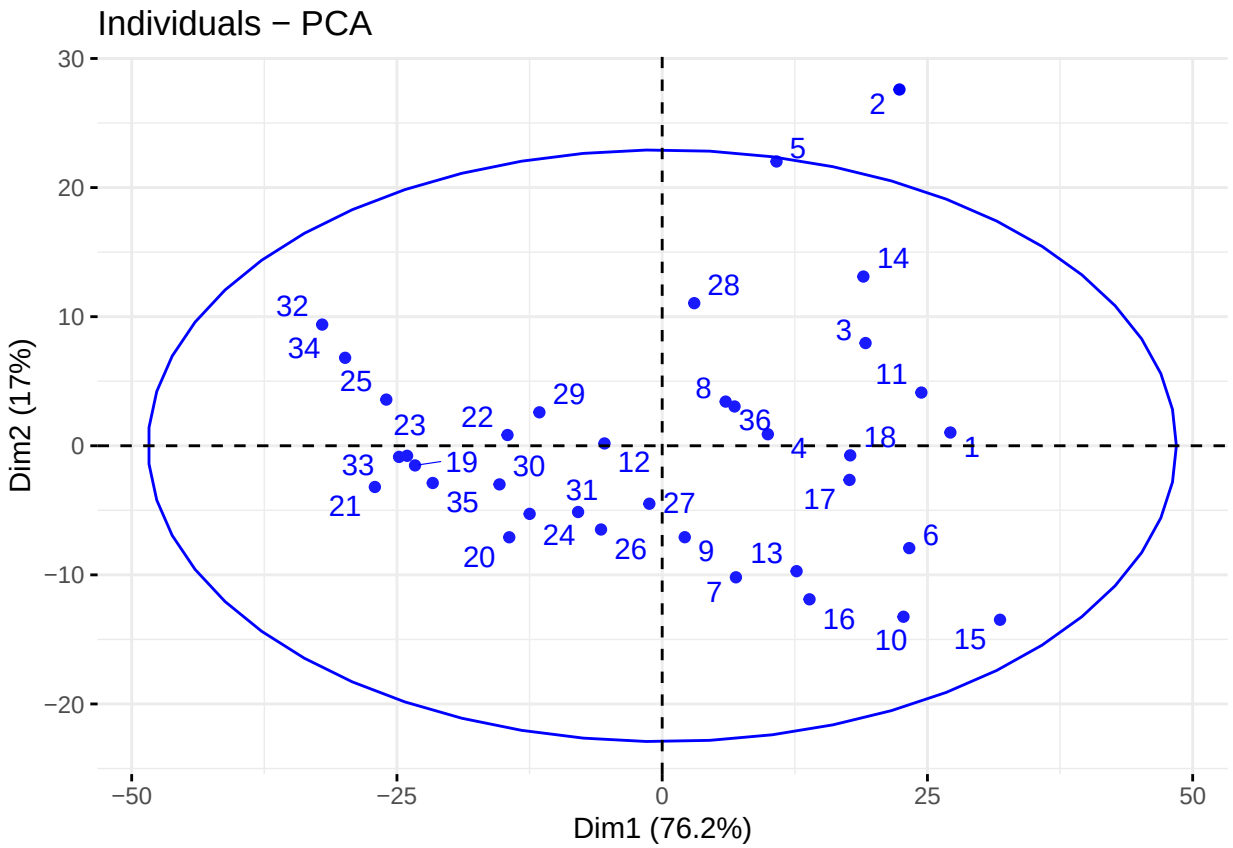



```
cpR <- PCA(numeric_data, scale.unit = TRUE)
```

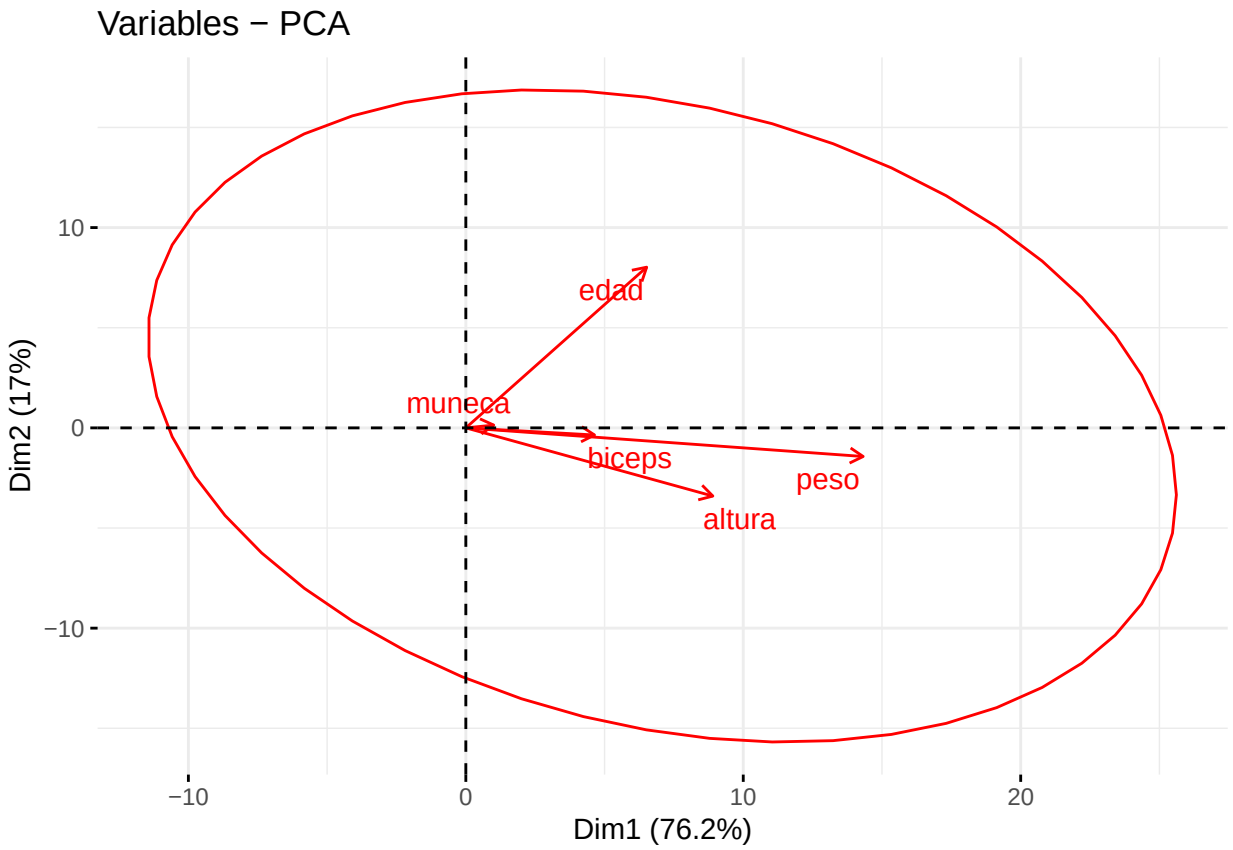

PCA graph of individuals
Dim 2 (14.52%)
Dim 1 (75.15%)
3
2
1
0
−1
−4
−2
0
2
2
5
28
14
32
3
34
25
8
11
29
22
36
4
33
23
12
1
30
18
19
35
24
27
21
31
17
26
20
13
9
6
7
16
10
15


PCA graph of variables
Dim 2 (14.52%)
Dim 1 (75.15%)
1.0
0.5
0.0
−0.5
−1.0
edad
muneca
biceps
peso
altura

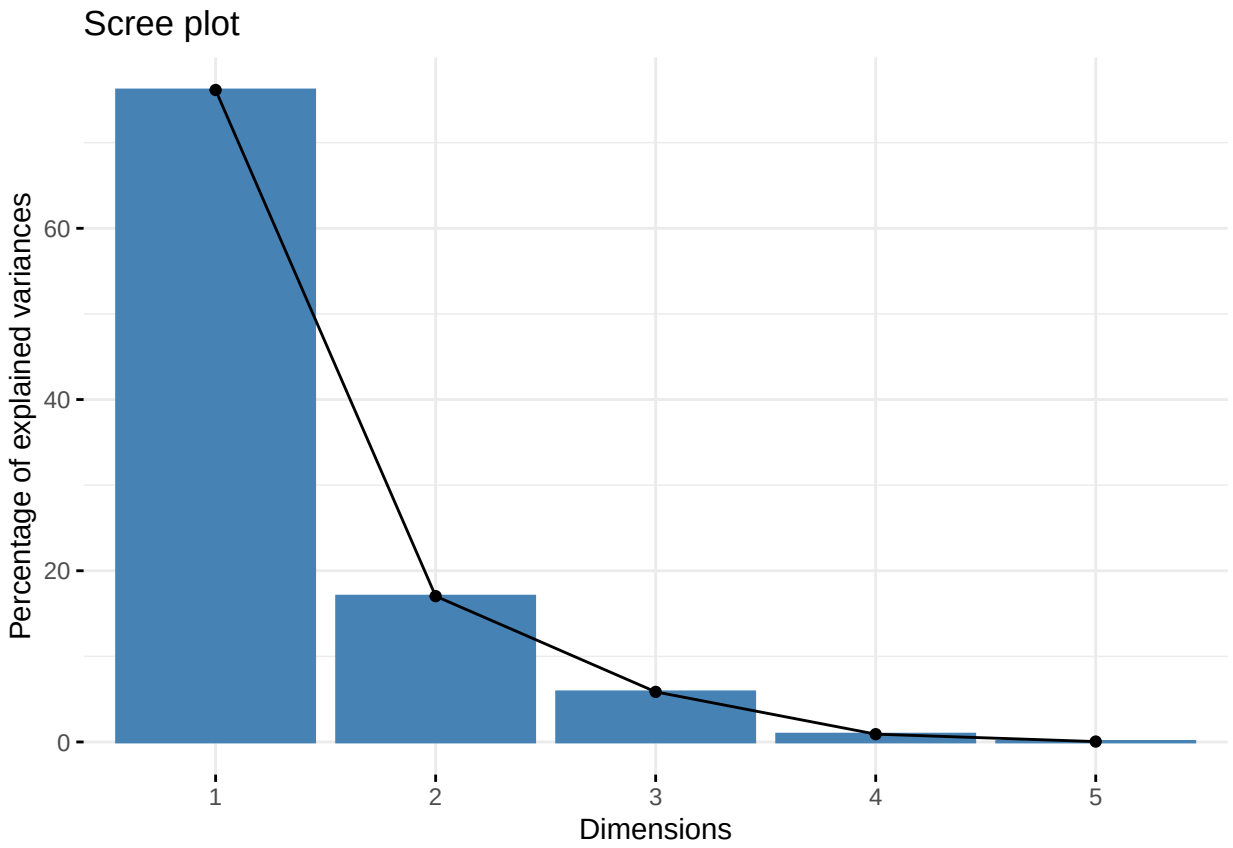
```
fviz_pca_ind(cpS, col.ind = "blue", addEllipses = TRUE, repel = TRUE)
```



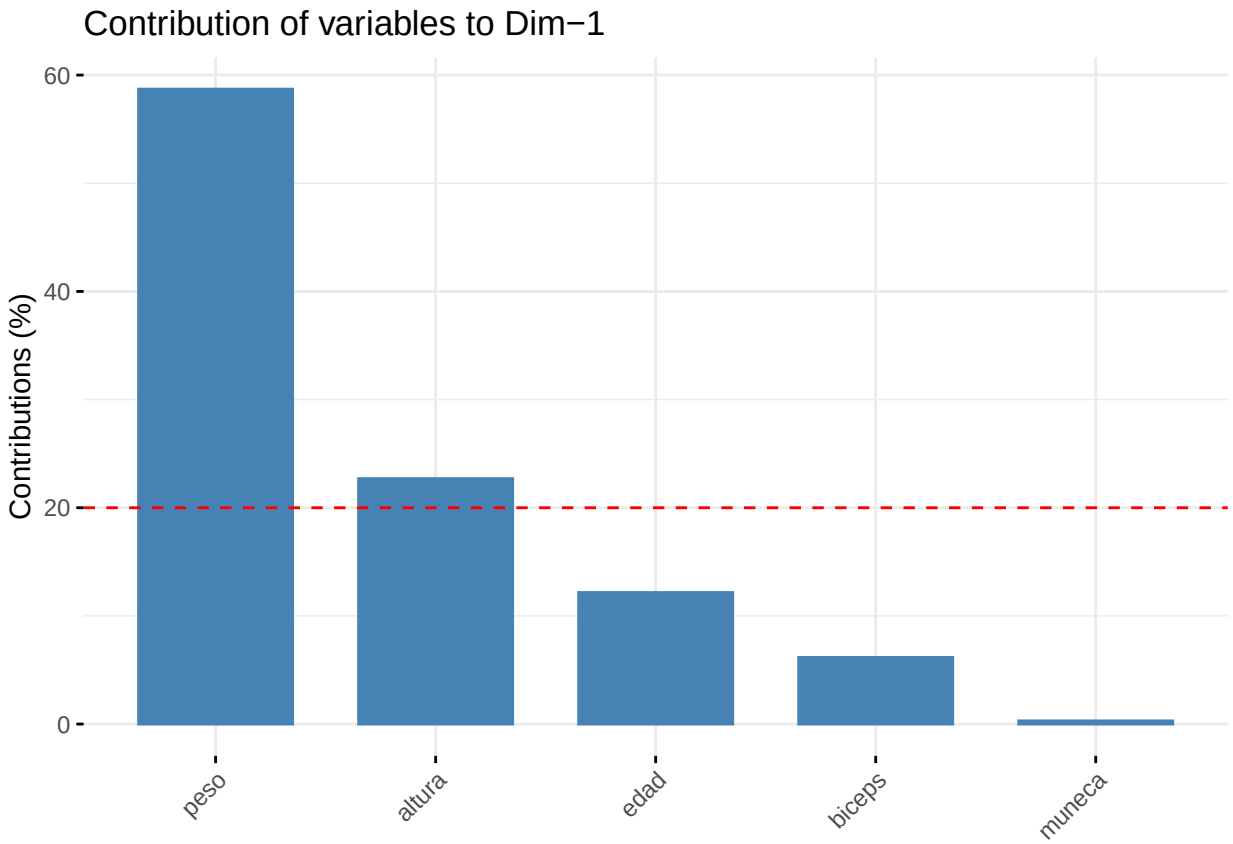

```
fviz_pca_var(cpS, col.var = "red", addEllipses = TRUE, repel = TRUE)
```



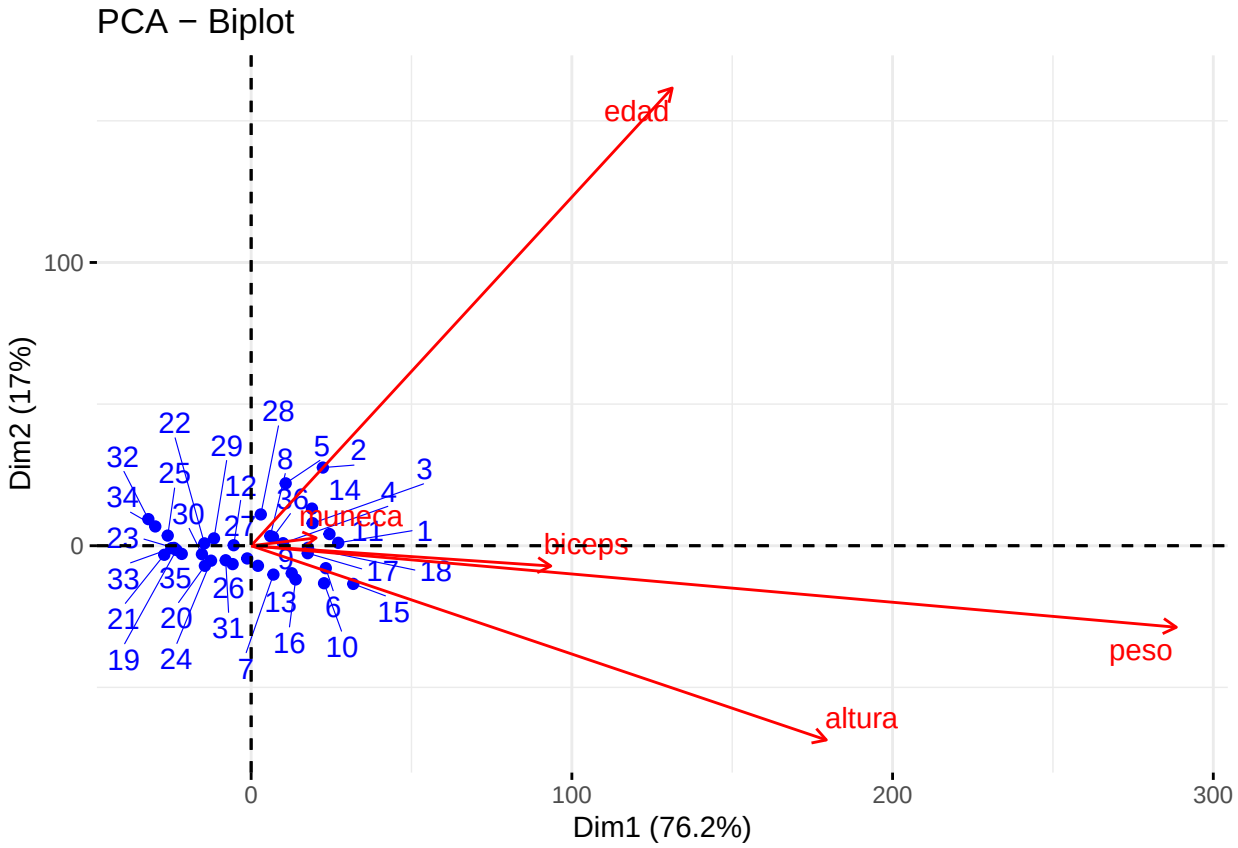

```
fviz_screeplot(cpS)
```




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




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



Las variables con mayor contribucion son peso, altura y edad, esto lo vemos en el componente uno donde peso es la mayor contribucion, mientras tanto en el segundo componente las variables que mas afectan son edad y altura, donde la edad es la variable que mas influye, las otras variables tienen una contribucion relativamente baja a comparacion de las que se escogieron.

Con los datos que se nos dieron la matriz de correlacion es mejor porque las variables tienen diferentes unidades y escalas, esto puede hacer que en la matriz de varianza-covarianza los datos mas grandes dominen.