

A4 - Componentes Principales

Héctor Hibrán Tapia Fernández - A01661114

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.

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

```
datos = read.csv("corporal.csv")  
  
summary(datos)
```

```
##      edad      peso      altura      sexo  
## Min.   :19.00  Min.   :42.00  Min.   :147.2  Length:36  
## 1st Qu.:24.75  1st Qu.:54.95  1st Qu.:164.8  Class :character  
## Median :28.00  Median :71.50  Median :172.7  Mode  :character  
## Mean   :31.44  Mean   :68.95  Mean   :171.6  
## 3rd Qu.:37.00  3rd Qu.:82.40  3rd Qu.:179.4  
## Max.   :65.00  Max.   :98.20  Max.   :190.5  
##      muneca      biceps  
## Min.   : 8.300  Min.   :23.50  
## 1st Qu.: 9.475  1st Qu.:25.98  
## Median :10.650  Median :32.15  
## Mean   :10.467  Mean   :31.17  
## 3rd Qu.:11.500  3rd Qu.:35.05  
## Max.   :12.400  Max.   :40.40
```

```
x = datos[, sapply(datos, is.numeric)]  
R = cor(x)
```

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

```
eigen_R = eigen(R)  
lambda_R = eigen_R$values  
vectores_R = eigen_R$vectors
```

- **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.**

```
varianza_total_R <- sum(diag(R))
prop_varianza_R <- lambda_R / varianza_total_R
```

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

```
varianza_acumulada_R = cumsum(prop_varianza_R)
```

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

```
tabla_varianza_R = data.frame(
  Componente = 1:length(lambda_R),
  Lambda = lambda_R,
  PropVarianza = prop_varianza_R,
  VarianzaAcumulada = varianza_acumulada_R
)

print(tabla_varianza_R)
```

##	Componente	Lambda	PropVarianza	VarianzaAcumulada
## 1	1	3.75749733	0.75149947	0.7514995
## 2	2	0.72585665	0.14517133	0.8966708
## 3	3	0.32032981	0.06406596	0.9607368
## 4	4	0.12461873	0.02492375	0.9856605
## 5	5	0.07169749	0.01433950	1.0000000

- **Escriba la ecuación de la combinación lineal de los Componentes principales CP1 y CP2 (e_1X , donde e_1 está en `eigen(S)$vectors[1]`, e_2X para obtener CP2, donde $X = c(X_1, X_2, \dots)$) ¿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.

```
coef_CP1 = vectores_R[, 1]
coef_CP2 = vectores_R[, 2]
variables = colnames(x)
ecuacion_CP1 = paste("CP1 = ", paste(round(coef_CP1, 4), "*", variables, collapse = " +
"))
ecuacion_CP2 = paste("CP2 = ", paste(round(coef_CP2, 4), "*", variables, collapse = " +
"))
cat(ecuacion_CP1, "\n")
```

```
## CP1 = -0.3359 * edad + -0.4927 * peso + -0.4222 * altura + -0.4822 * muneca + -0.483
3 * biceps
```

```
cat(ecuacion_CP2, "\n")
```

```
## CP2 = 0.8576 * edad + -0.1648 * peso + -0.4542 * altura + 0.1083 * muneca + -0.1393
* biceps
```

PARTE II

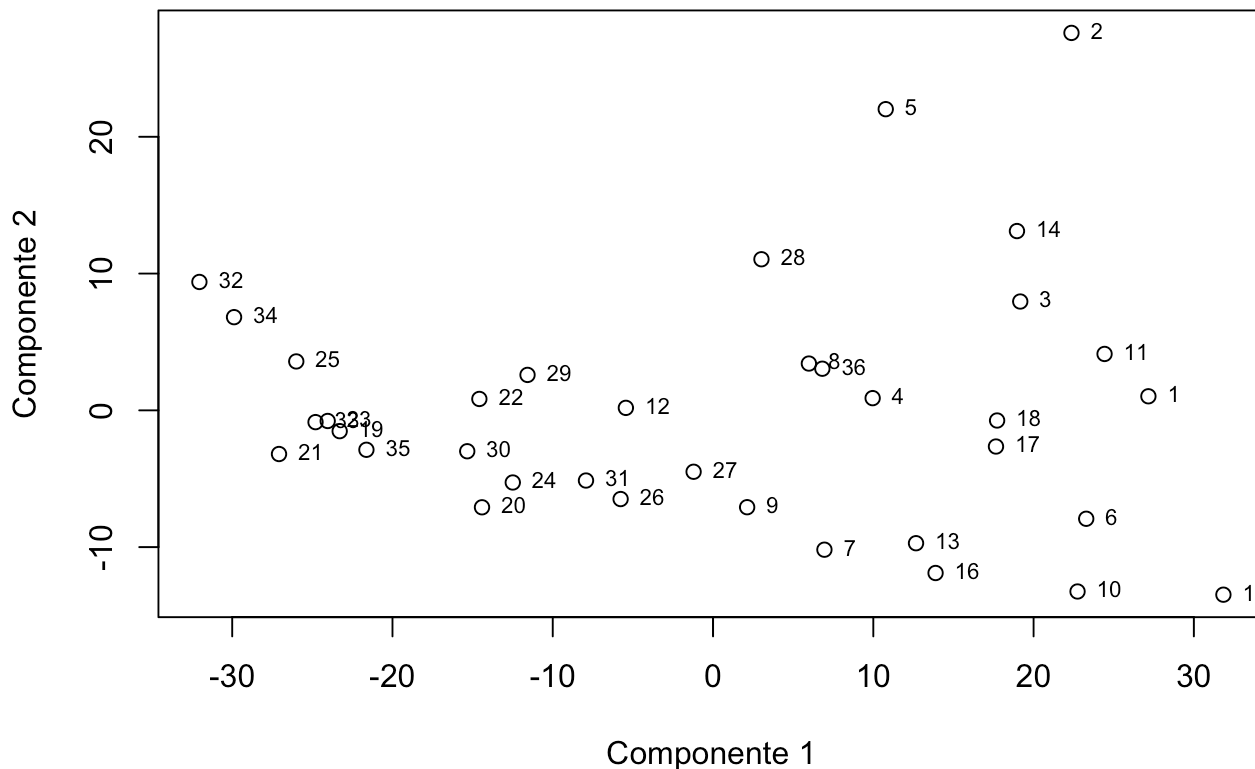
1. Obtenga las gráficas respectivas con S (matriz de varianzas-covarianzas) y con R (matriz de correlaciones) de las dos primeras componentes.

```
x = datos[, sapply(datos, is.numeric)]
S = cov(x)
cp_S = princomp(x, cor = FALSE)
```

- Calcule las puntuaciones (scores) de las observaciones para los componentes obtenidos con la matriz de varianzas-covarianzas

```
scores_S = cp_S$scores
plot(scores_S[,1], scores_S[,2], main = "Componentes Principales (S)", xlab = "Componente 1", ylab = "Componente 2")
text(scores_S[,1], scores_S[,2], labels = rownames(x), cex = 0.7, pos = 4)
```

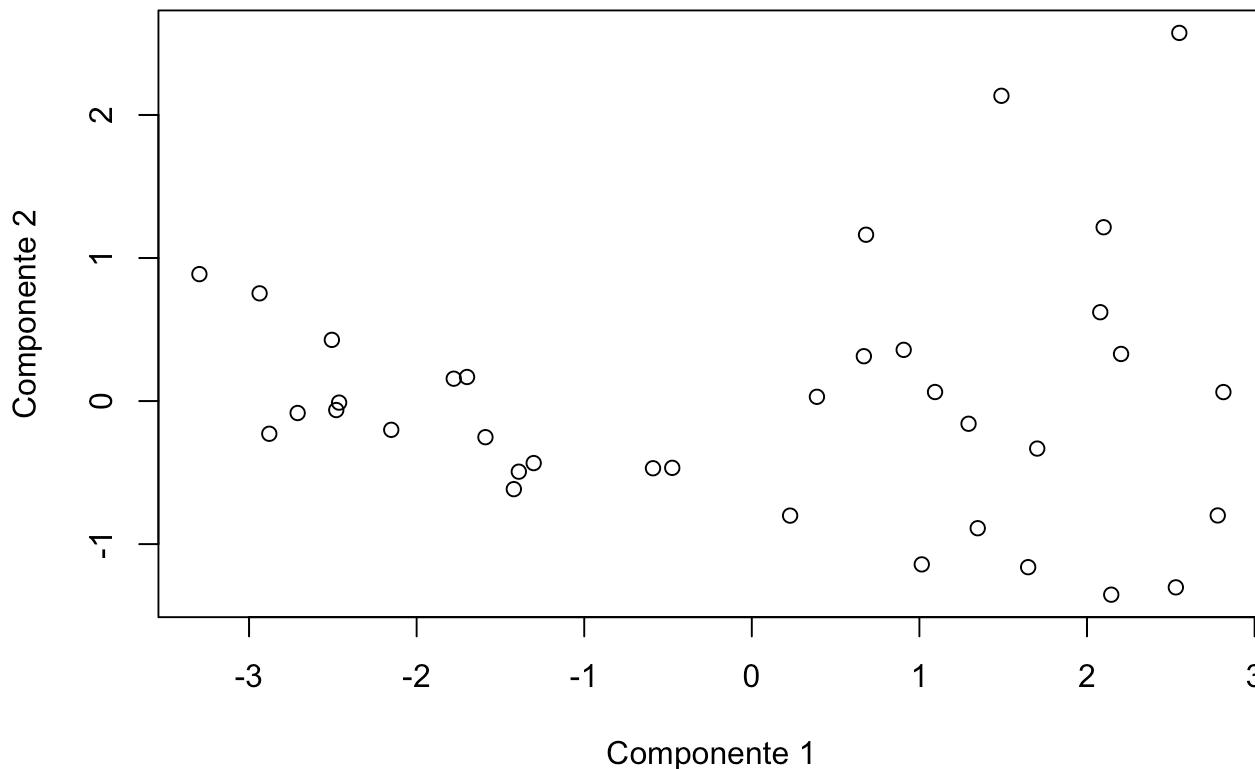
Componentes Principales (S)



- 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.

```
x_scaled = scale(x)
R = cor(x_scaled)
cp_R = princomp(x_scaled, cor = TRUE)
scores_R = cp_R$scores
plot(scores_R[,1], scores_R[,2], main = "Componentes Principales (R)", xlab = "Componente 1", ylab = "Componente 2")
text(scores_R[,1], scores_R[,2], labels = rownames(x_scaled), cex = 0.7, pos = 4)
```

Componentes Principales (R)



2. 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$loadings`, `cpa$Sscores`. ¿Cómo se interpreta el resultado?

```
cp_S = princomp(x, cor = FALSE)
summary(cp_S)
```

```
## 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
```

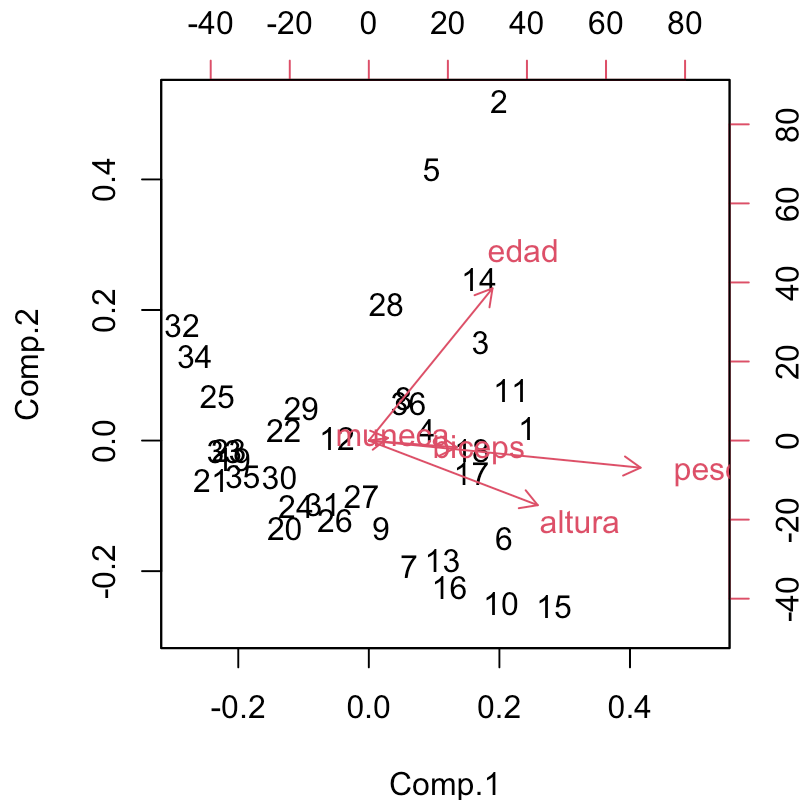
```
cp_S$loadings
```

```
##
## 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
```

cp_S\$scores

##		Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
##	[1,]	27.162853	1.0278492	5.0022646	0.936226898	-0.51688356
##	[2,]	22.363542	27.5955807	3.0635949	-0.083381259	0.02552809
##	[3,]	19.167874	7.9566157	-1.5770026	-2.610776762	0.80391745
##	[4,]	9.959001	0.8923731	5.5146952	0.123453725	-0.35579895
##	[5,]	10.775593	22.0203437	-0.7562826	0.179967226	-0.41646606
##	[6,]	23.283948	-7.9268214	2.7958617	-2.093392841	-0.62252321
##	[7,]	6.949553	-10.1882447	1.5804639	-5.636477243	0.75692216
##	[8,]	5.981213	3.4214568	-7.0113449	-0.999845471	-0.13795746
##	[9,]	2.128453	-7.0823040	9.6199213	-2.402765355	0.30931008
##	[10,]	22.742222	-13.2447241	-5.8006902	-1.900258608	-0.11415400
##	[11,]	24.427931	4.1227827	-3.0914640	1.417935347	0.45836253
##	[12,]	-5.438123	0.1807499	1.3551969	-5.147087631	-0.71928452
##	[13,]	12.665261	-9.7148314	-4.4445147	0.469977365	-0.44199755
##	[14,]	18.962350	13.1080907	4.5325770	0.310839551	-0.27648044
##	[15,]	31.842783	-13.4784052	-1.4672915	5.610391303	0.61177438
##	[16,]	13.884278	-11.8930081	-6.4032979	-2.225813208	-0.01138562
##	[17,]	17.653813	-2.6451319	-0.8986274	-0.529020358	0.37187295
##	[18,]	17.723299	-0.7428241	0.1219847	1.785013852	0.68809035
##	[19,]	-23.293603	-1.5208783	0.2627514	1.143811767	-0.16480880
##	[20,]	-14.414169	-7.0887516	0.1030611	0.006854239	-0.32687435
##	[21,]	-27.078917	-3.1933468	-0.4483831	0.722326288	-0.02028518
##	[22,]	-14.579228	0.8324474	-9.1400445	1.717699742	0.23470254
##	[23,]	-24.042246	-0.7779288	-5.8550300	-0.340341079	0.26832127
##	[24,]	-12.494468	-5.2751971	3.0622990	1.094339917	-0.51675730
##	[25,]	-26.002609	3.5759758	1.6616974	0.054118319	-0.33475598
##	[26,]	-5.766003	-6.4856729	-6.5862305	2.330421808	-0.76268815
##	[27,]	-1.211876	-4.4901315	4.4920764	1.153351801	0.26364518
##	[28,]	3.020501	11.0467489	-10.8052957	0.255974364	-0.43453383
##	[29,]	-11.574038	2.5907341	9.5304169	1.466717121	0.84144772
##	[30,]	-15.335150	-2.9912143	6.9968010	0.493427421	-0.36660212
##	[31,]	-7.926087	-5.1312097	4.1467185	2.808113699	0.29328661
##	[32,]	-32.046176	9.3863372	0.8359798	-1.341797979	0.73976836
##	[33,]	-24.800765	-0.8616289	-0.1246471	-0.477476584	0.58698947
##	[34,]	-29.884003	6.8137270	-9.5237493	-0.372525171	0.27802711
##	[35,]	-21.626441	-2.8831824	7.4391447	0.704477945	-0.64549912
##	[36,]	6.819433	3.0436244	1.8163894	1.375519851	-0.34623005

```
biplot(cp_S, main = " ")
```



PARTE III

1. Explore los siguientes gráficos relativos a Componentes Principales.

2. 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(factoextra)
```

```
## Loading required package: ggplot2
```

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

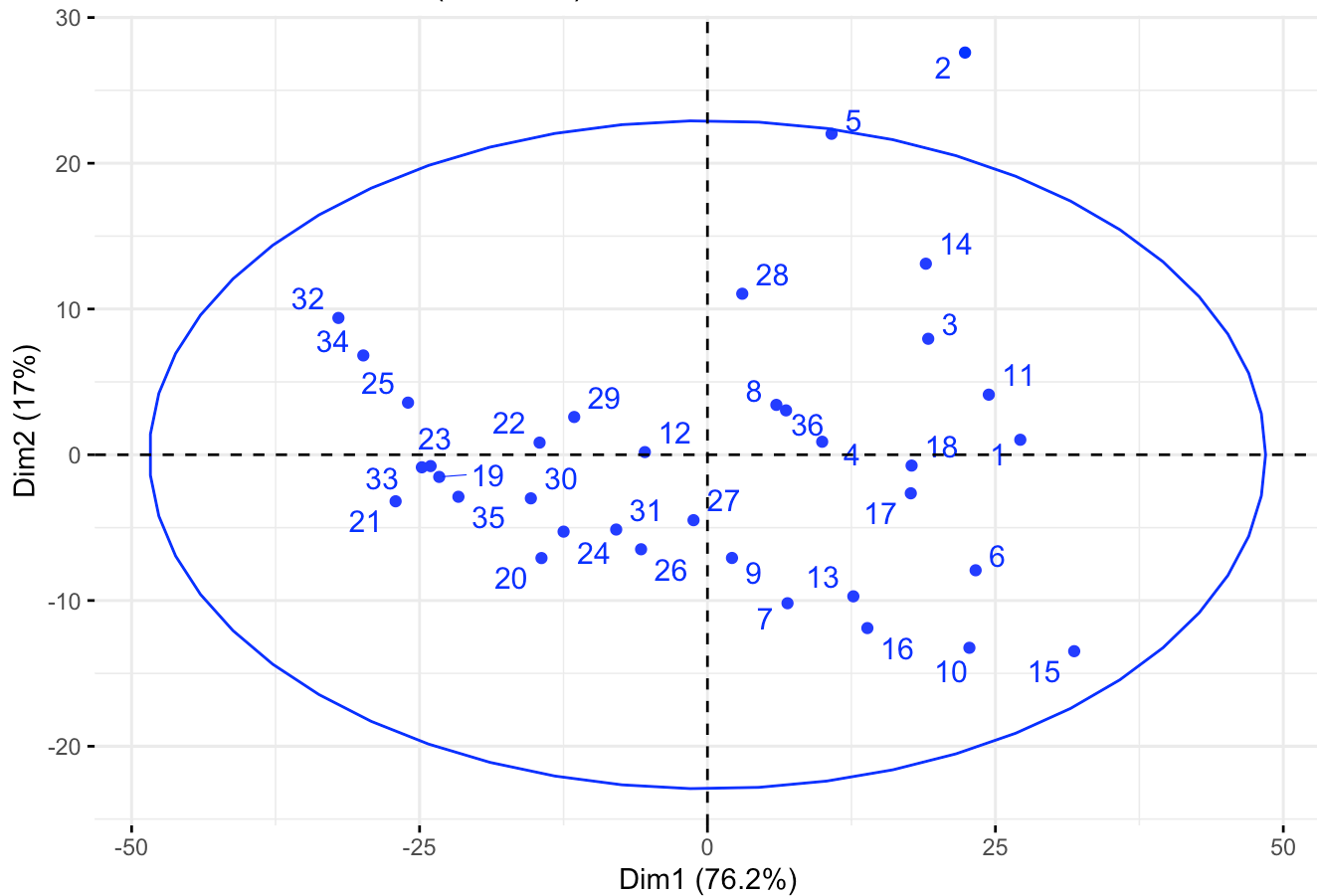


```
library(FactoMineR)
library(ggplot2)

datos = read.csv("corporal.csv")
datos_numericos = datos[, apply(datos, is.numeric)]
```

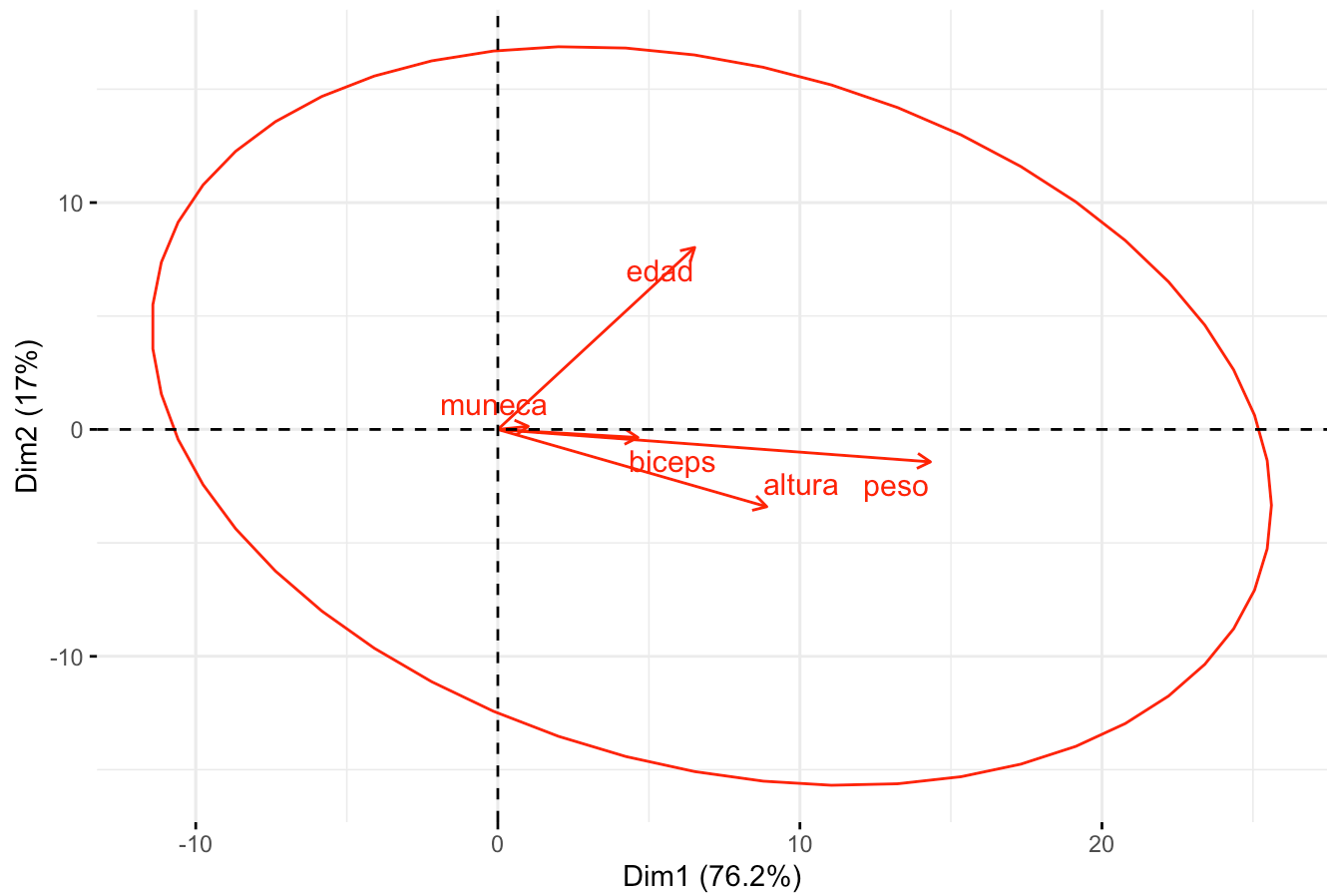
```
pca_S <- PCA(datos_numericos, scale.unit = FALSE, graph = FALSE)
fviz_pca_ind(pca_S, col.ind = "blue", addEllipses = TRUE, repel = TRUE,
             title = "Gráfico de Individuos (Matriz S)")
```

Gráfico de Individuos (Matriz S)



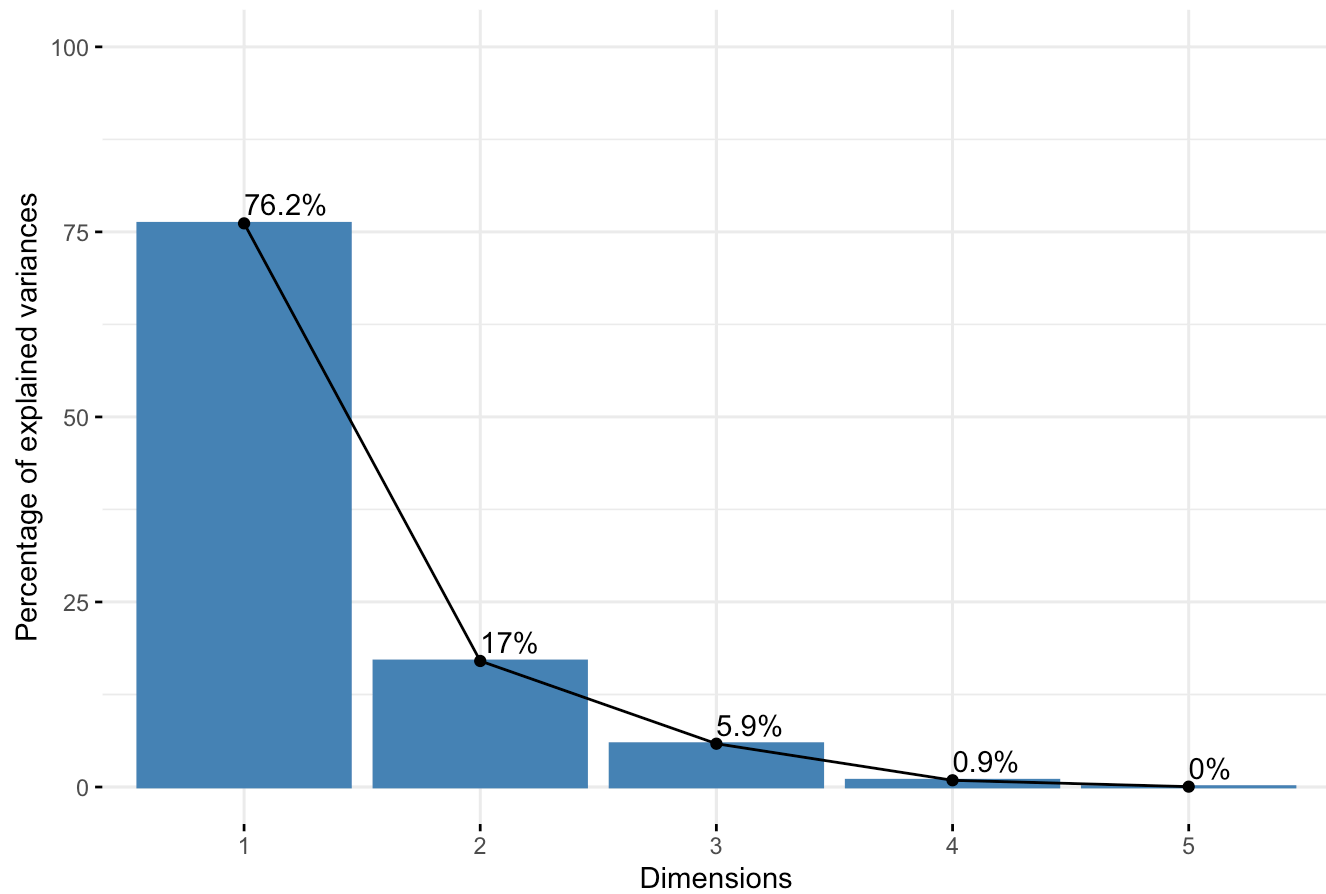
```
fviz_pca_var(pca_S, col.var = "red", addEllipses = TRUE, repel = TRUE,
             title = "Gráfico de Variables (Matriz S)")
```

Gráfico de Variables (Matriz S)



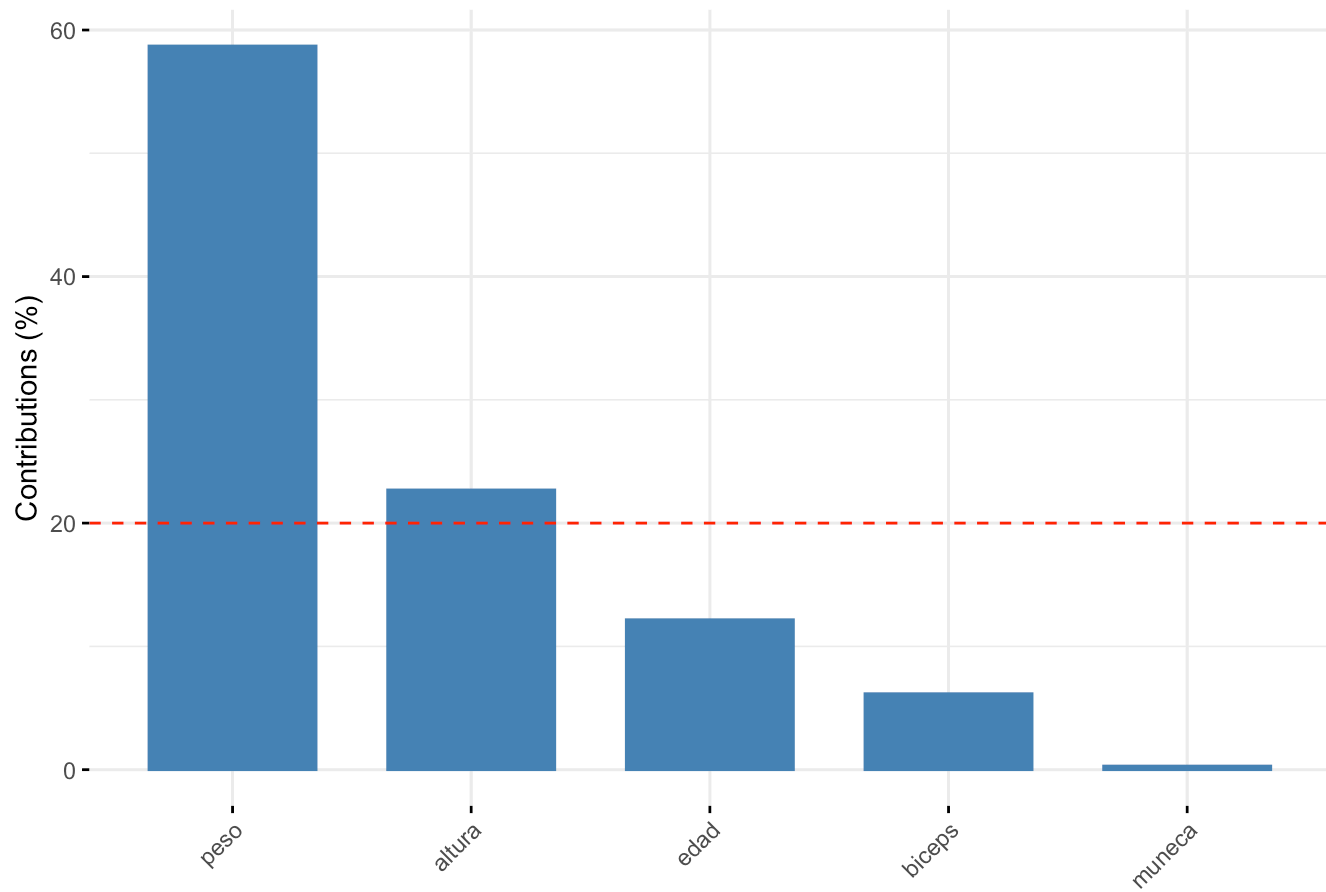
```
fviz_screplot(pca_S, addlabels = TRUE, ylim = c(0, 100),  
              title = "Scree Plot (Matriz S)")
```

Scree Plot (Matriz S)



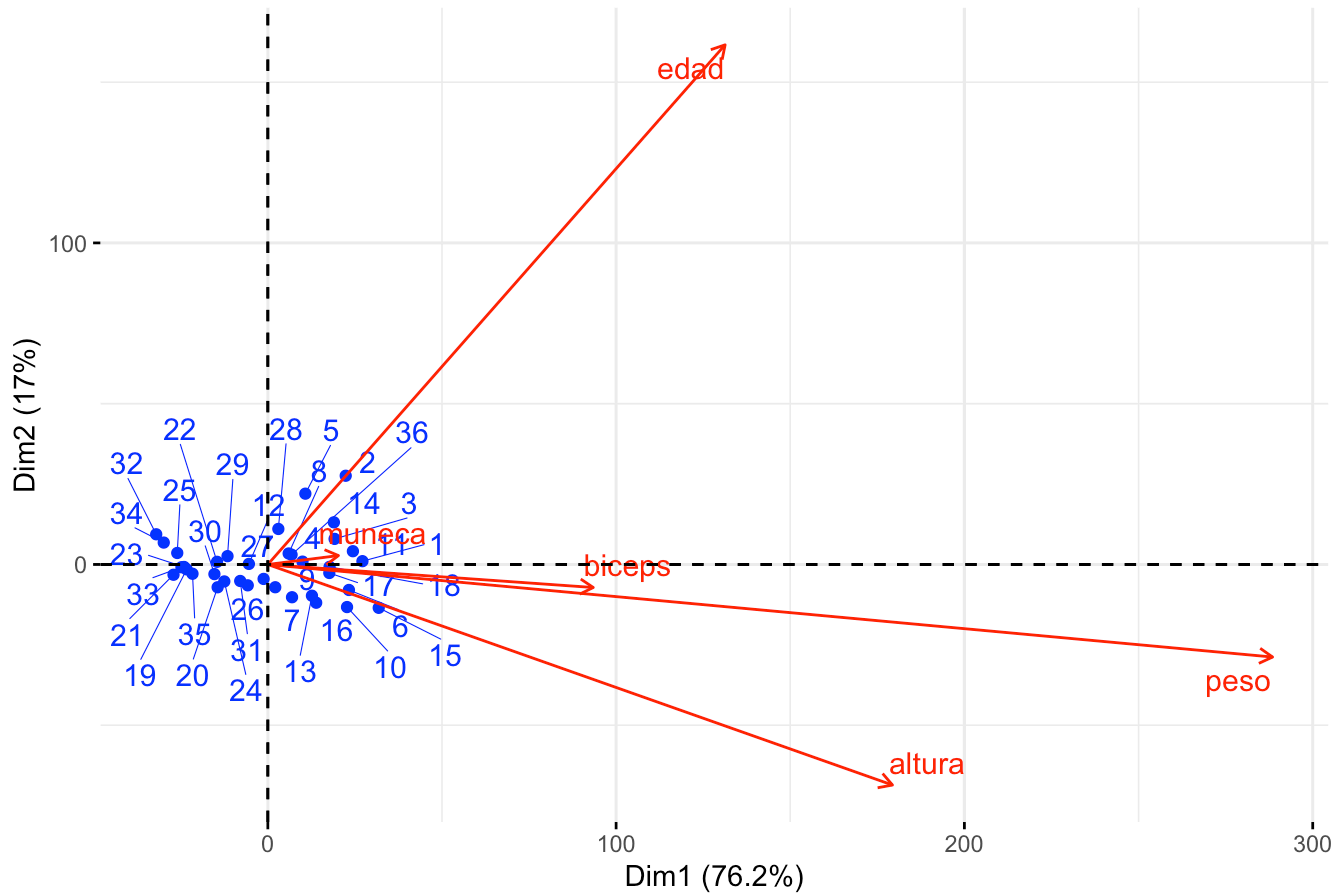
```
fviz_contrib(pca_S, choice = "var", axes = 1, top = 10,  
             title = "Contribución de Variables al PC1 (Matriz S)")
```

Contribución de Variables al PC1 (Matriz S)



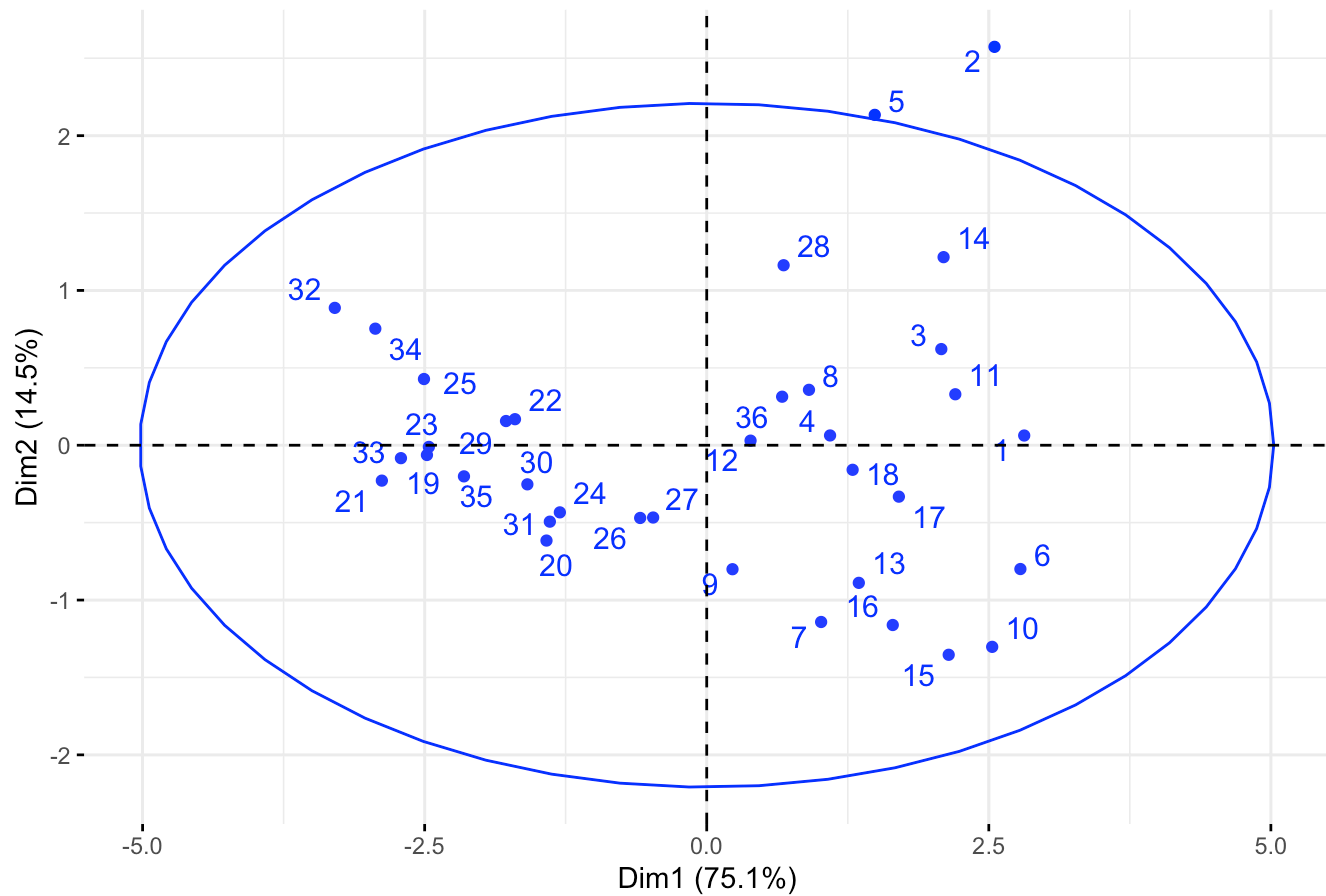
```
fviz_pca_biplot(pca_S, repel = TRUE, col.var = "red", col.ind = "blue",  
  title = "Biplot de PCA (Matriz S)")
```

Biplot de PCA (Matriz S)



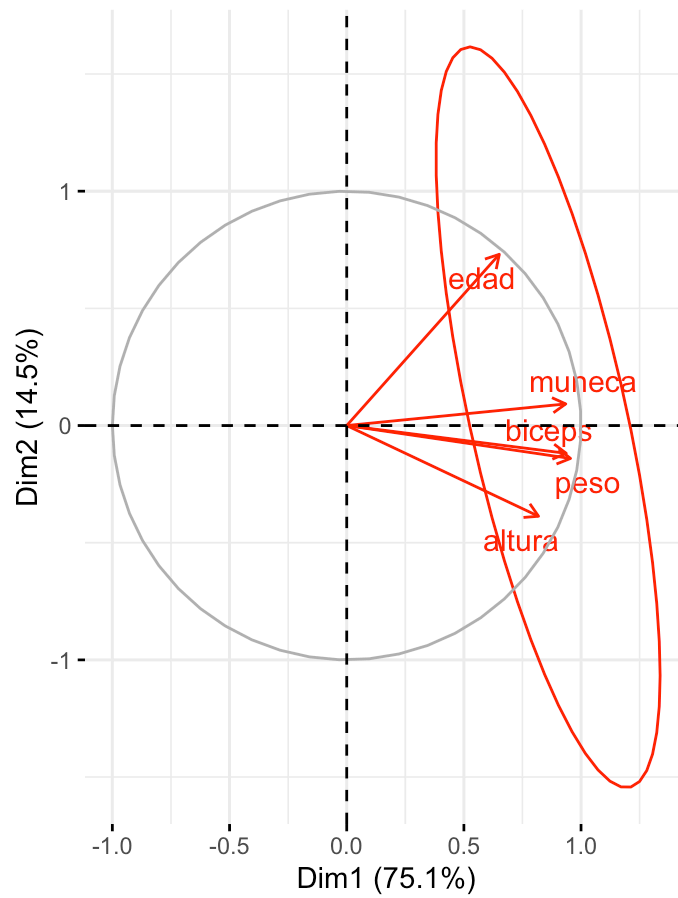
```
pca_R <- PCA(datos_numericos, scale.unit = TRUE, graph = FALSE)
fviz_pca_ind(pca_R, col.ind = "blue", addEllipses = TRUE, repel = TRUE,
             title = "Gráfico de Individuos (Matriz R)")
```

Gráfico de Individuos (Matriz R)



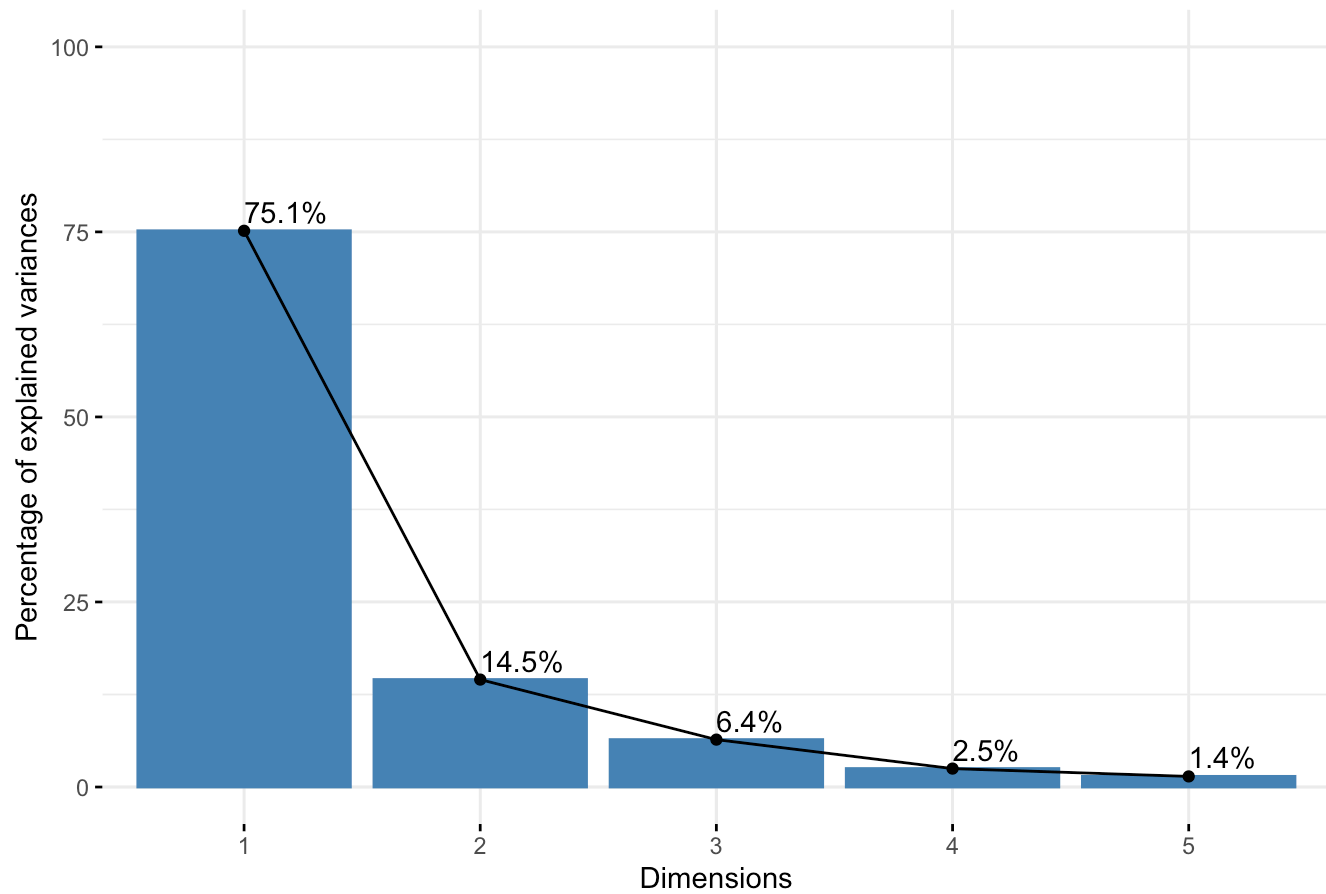
```
fviz_pca_var(pca_R, col.var = "red", addEllipses = TRUE, repel = TRUE,  
             title = "Gráfico de Variables (Matriz R)")
```

Gráfico de Variables (Matriz R)

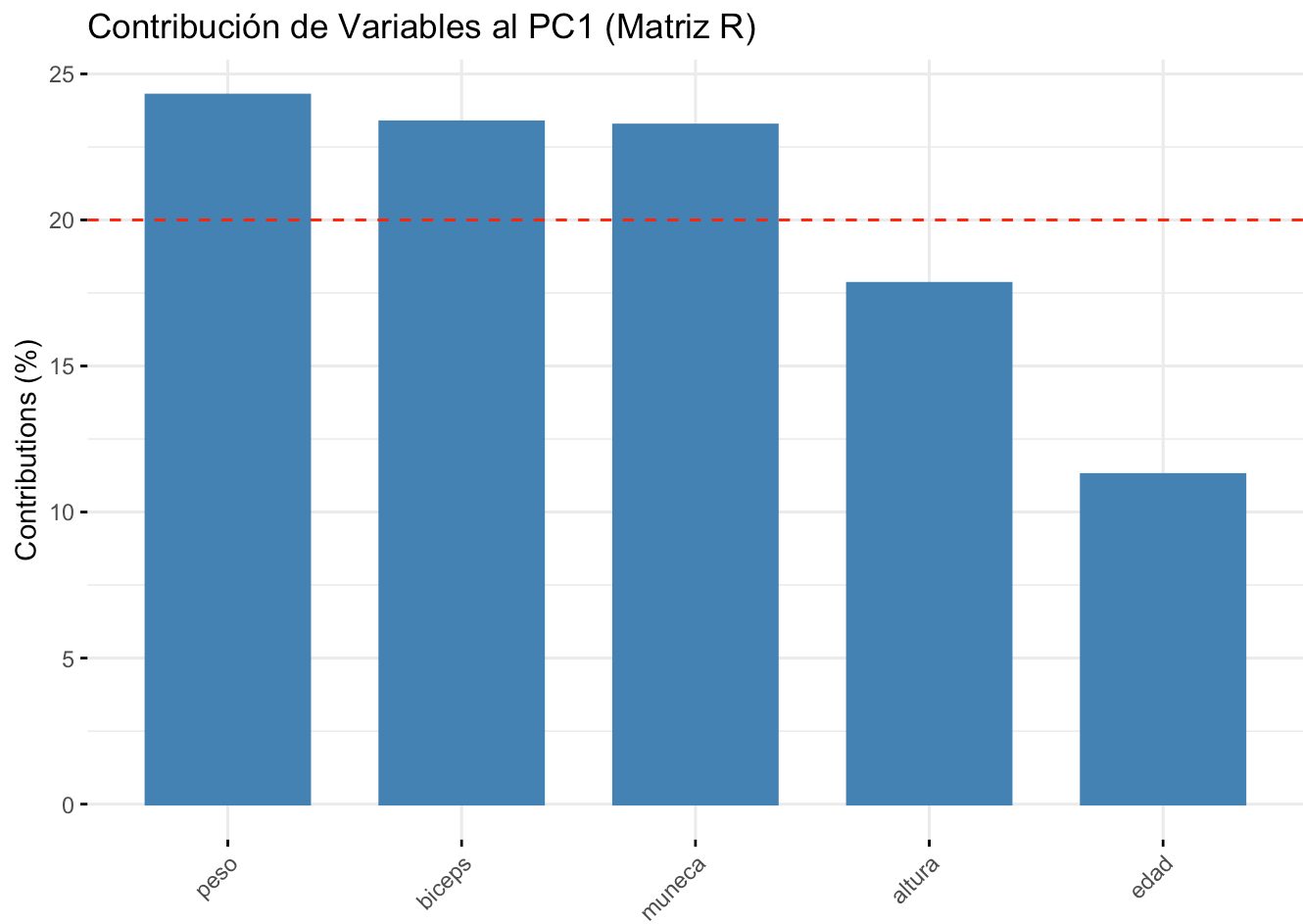


```
fviz_screplot(pca_R, addlabels = TRUE, ylim = c(0, 100),  
              title = "Scree Plot (Matriz R)")
```

Scree Plot (Matriz R)

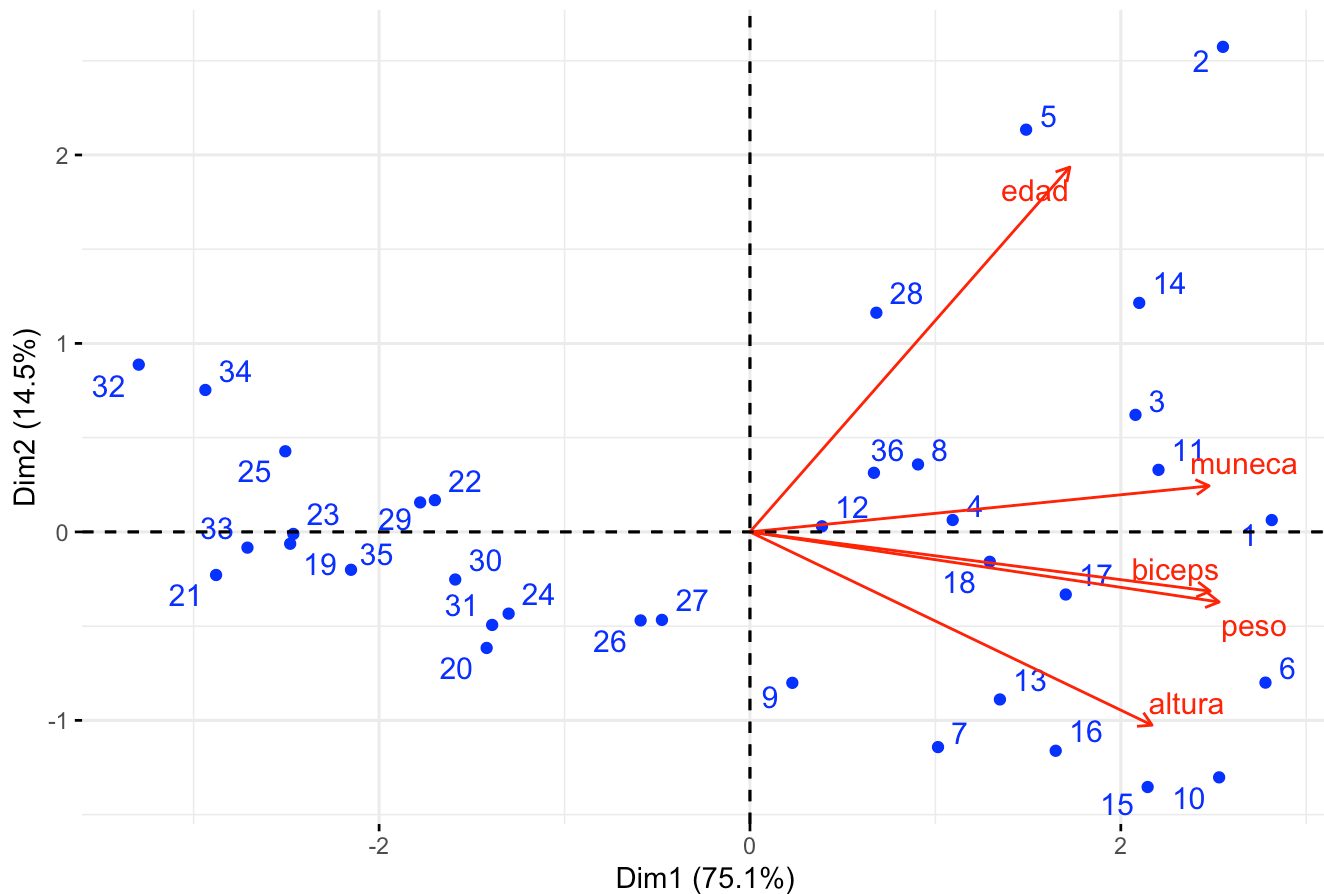


```
fviz_contrib(pca_R, choice = "var", axes = 1, top = 10,  
             title = "Contribución de Variables al PC1 (Matriz R)")
```

```
fviz_pca_biplot(pca_R, repel = TRUE, col.var = "red", col.ind = "blue",  
title = "Biplot de PCA (Matriz R)")
```

Biplot de PCA (Matriz R)



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?

```
datos$sexo <- as.factor(datos$sexo)
nombres_variables <- names(datos)
variables_activas <- which(sapply(datos, is.numeric))
variables_cualitativas <- which(names(datos) == "sexo")
pca_S_sup <- PCA(
  datos,
  scale.unit = FALSE,
  ncp = 5,
  quali.sup = variables_cualitativas,
  quanti.sup = NULL,
  graph = FALSE
)

fviz_pca_ind(
  pca_S_sup,
  col.ind = "cos2",
  palette = "jco",
  addEllipses = TRUE,
  label = "none",
  habillage = "sexo",
  legend.title = "Sexo"
)
```

