

A4. Componentes Principales

Oscar Gutierrez

2024-10-08

Parte 1

Análisis descriptivo

```
X = read.csv('corporal.csv')
X = X[, -which(names(X) == 'sexo')]

n=5 #número de variables

d=matrix(NA,ncol=7,nrow=n)
for(i in 1:n){
  d[i,]<-c(as.numeric(summary(X[,i])),sd(X[,i]))
}
m=as.data.frame(d)

row.names(m)=c('edad', 'altura', 'peso', 'muneca', 'biceps')
names(m)=c("Minimo", "Q1", "Mediana", "Media", "Q3", "Máximo", "Desv Est")
m
```

##	Minimo	Q1	Mediana	Media	Q3	Máximo	Desv Est
## edad	19.0	24.750	28.00	31.44444	37.00	65.0	10.554469
## altura	42.0	54.950	71.50	68.95278	82.40	98.2	14.868999
## peso	147.2	164.800	172.70	171.55556	179.40	190.5	10.520170
## muneca	8.3	9.475	10.65	10.46667	11.50	12.4	1.175463
## biceps	23.5	25.975	32.15	31.16667	35.05	40.4	5.234392

En el dataframe se pueden observar valores estadísticos importantes que brindan información general sobre las variables que se estarán manejando.

Varianza y covarianza

```
S <- cov(X) # Matriz de varianza-covarianza
R <- cor(X) # Matriz de correlaciones

print(S)
```

```
##          edad      peso      altura      muñeca      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
## muñeca    7.698095  14.84467  8.156476  1.381714  5.400571
## biceps   26.720952  70.73838  39.021048  5.400571 27.398857
```

```
print(R)
```

```
##          edad      peso      altura      muñeca      biceps
## edad    1.00000000 0.5153847 0.3302211 0.6204942 0.4836702
## peso     0.5153847 1.00000000 0.7973737 0.8493361 0.9088813
## altura   0.3302211 0.7973737 1.00000000 0.6595849 0.7086144
## muñeca   0.6204942 0.8493361 0.6595849 1.00000000 0.8777369
## biceps   0.4836702 0.9088813 0.7086144 0.8777369 1.00000000
```

La matriz de varianza-covarianza mide la relación entre las variaciones de dos o más variables. Cada elemento en la matriz representa la covarianza entre un par de variables, que indica cómo varían conjuntamente. Los elementos diagonales son las varianzas de cada variable individual.

La matriz de correlación es similar, pero los elementos están estandarizados, de modo que los valores de correlación están entre -1 y 1. La correlación indica la fuerza y dirección de la relación lineal entre variables, eliminando la influencia de las escalas de las variables.

Vectores y valores propios

```
# Valores y vectores propios de la matriz de varianza-covarianza S
eigen_S <- eigen(S)

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

print(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
```

```
print(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
```

Estos son los eigenvalores y eigenvectores propios obtenidos. Los eigenvalores representan la variación en ese eje mientras que los vectores propios representan los coeficientes de la combinación lineal.

```
# Calcular la varianza total
varianza_total_S <- sum(diag(S))
varianza_total_S_eigen <- sum(eigen_S$values)

cat('Varianza total calculada a partir de la suma de la diagonal:',varianza_total_S, '\n
Varianza total calculada a partir de los valores propios: ', varianza_total_S_eigen)
```

```
## Varianza total calculada a partir de la suma de la diagonal: 471.9385
## Varianza total calculada a partir de los valores propios: 471.9385
```

```
# Proporción de varianza explicada por cada componente
proporcion_varianza_S <- eigen_S$values / varianza_total_S
cat('\n\nProporción de varianza S', proporcion_varianza_S)
```

```
##
##
## Proporción de varianza S 0.7615357 0.1703099 0.05853072 0.009127104 0.0004965839
```

```
# Acumulación de la varianza explicada
varianza_acumulada_S <- cumsum(proporcion_varianza_S)
cat('\n\nProporción de varianza acumulada S', varianza_acumulada_S)
```

```
##
## Proporción de varianza acumulada S 0.7615357 0.9318456 0.9903763 0.9995034 1
```

```
# Calcular la varianza total de R
varianza_total_R <- sum(eigen_R$values)

# Proporción de varianza explicada
proporcion_varianza_R <- eigen_R$values / varianza_total_R
cat('\n\nProporción de varianza R', proporcion_varianza_R)
```

```
##
##
## Proporción de varianza R 0.7514995 0.1451713 0.06406596 0.02492375 0.0143395
```

```
# Acumulación de la varianza explicada
varianza_acumulada_R <- cumsum(proporcion_varianza_R)
cat('\nProporción de varianza acumulada R', varianza_acumulada_R)
```

```
##
## Proporción de varianza acumulada R 0.7514995 0.8966708 0.9607368 0.9856605 1
```

De acuerdo con estos resultados, los componentes más importantes son los primeros dos, ya que con ellos se logra explicar aproximadamente un 90% de la variación, por lo que si se seleccionan estos componentes, podemos evitar utilizar 5 dimensiones y solo estaríamos trabajando con 2, lo cual facilita la interpretación y reduce el costo computacional.

Combinaciones lineales

```
etiquetas <- colnames(X)
# Combinación lineal de CP1
CP1_coeficientes_S <- eigen_S$vector[,1]
CP1_combinacion_S <- data.frame(Variable = etiquetas, Coeficiente_CP1_S = CP1_coeficientes_S)

# Combinación lineal de CP2
CP2_coeficientes_S <- eigen_S$vector[,2]
CP2_combinacion_S <- data.frame(Variable = etiquetas, Coeficiente_CP2_S = CP2_coeficientes_S)

# Imprimir los resultados
print(CP1_combinacion_S)
```

```
##   Variable Coeficiente_CP1_S
## 1    edad      -0.34871002
## 2    peso      -0.76617586
## 3   altura      -0.47632405
## 4   muneca      -0.05386189
## 5   biceps      -0.24817367
```

```
print(CP2_combinacion_S)
```

```
##   Variable Coeficiente_CP2_S
## 1    edad       0.9075501
## 2    peso      -0.1616581
## 3   altura      -0.3851755
## 4   muneca       0.0155423
## 5   biceps      -0.0402221
```

```
# Combinación lineal de CP1
CP1_coeficientes_R <- eigen_R$vectors[,1]
CP1_combinacion_R <- data.frame(Variable = etiquetas, Coeficiente_CP1_R = CP1_coeficientes_R)

# Combinación lineal de CP2
CP2_coeficientes_R <- eigen_R$vectors[,2]
CP2_combinacion_R <- data.frame(Variable = etiquetas, Coeficiente_CP2_R = CP2_coeficientes_R)

# Imprimir los resultados
print(CP1_combinacion_R)
```

```
##   Variable Coeficiente_CP1_R
## 1   edad      -0.3359310
## 2   peso      -0.4927066
## 3  altura      -0.4222426
## 4  muneca      -0.4821923
## 5  biceps      -0.4833139
```

```
print(CP2_combinacion_R)
```

```
##   Variable Coeficiente_CP2_R
## 1   edad      0.8575601
## 2   peso      -0.1647821
## 3  altura      -0.4542223
## 4  muneca      0.1082775
## 5  biceps      -0.1392684
```

Las variables más relevantes para cada uno de los componentes principales son las siguientes:

Para la matriz de varianza-covarianza: CP1: peso y altura CP2: edad y altura

Para la matriz de correlación CP1: peso y biceps CP2: edad y altura

Parte 2

Gráfica de las dos primeras componentes

Para la matriz de covarianza

```
library(stats)
```

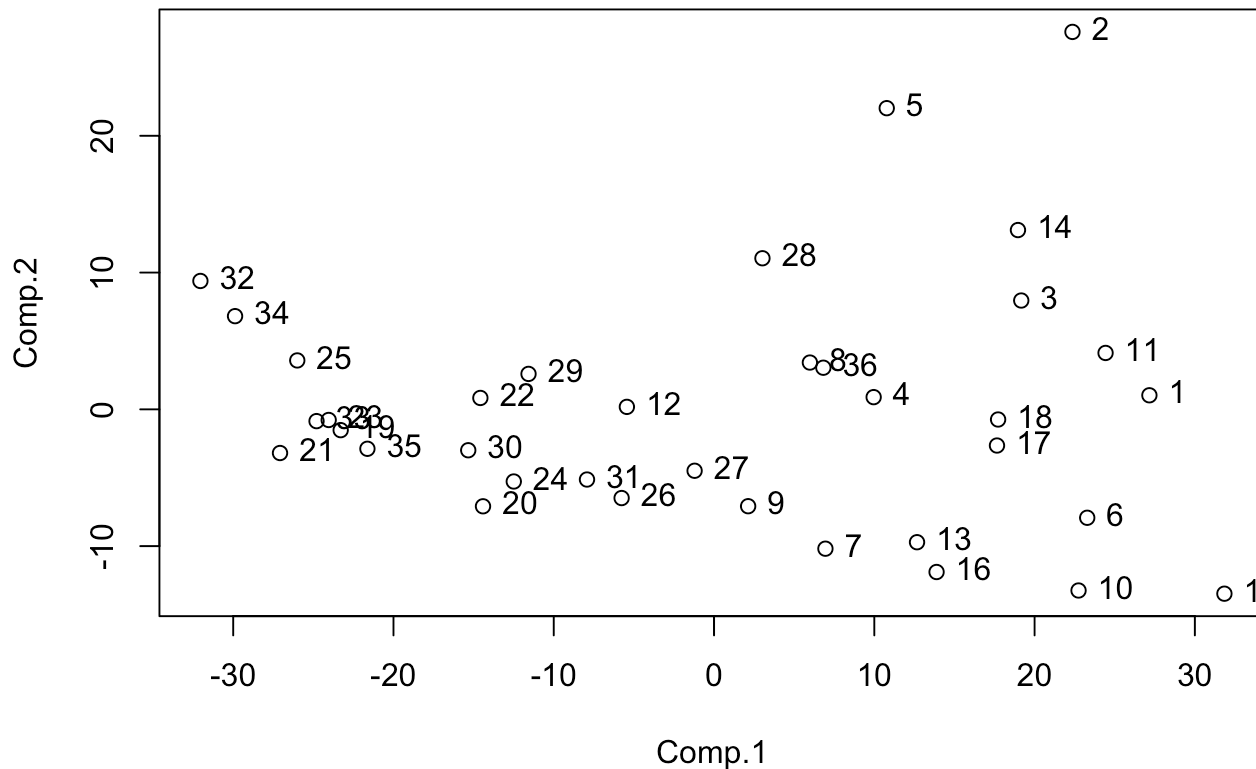
```
# Calcular el análisis de componentes principales para la matriz S
cpS <- princomp(X, cor = FALSE)
```

```
# Graficar las dos primeras componentes principales
```

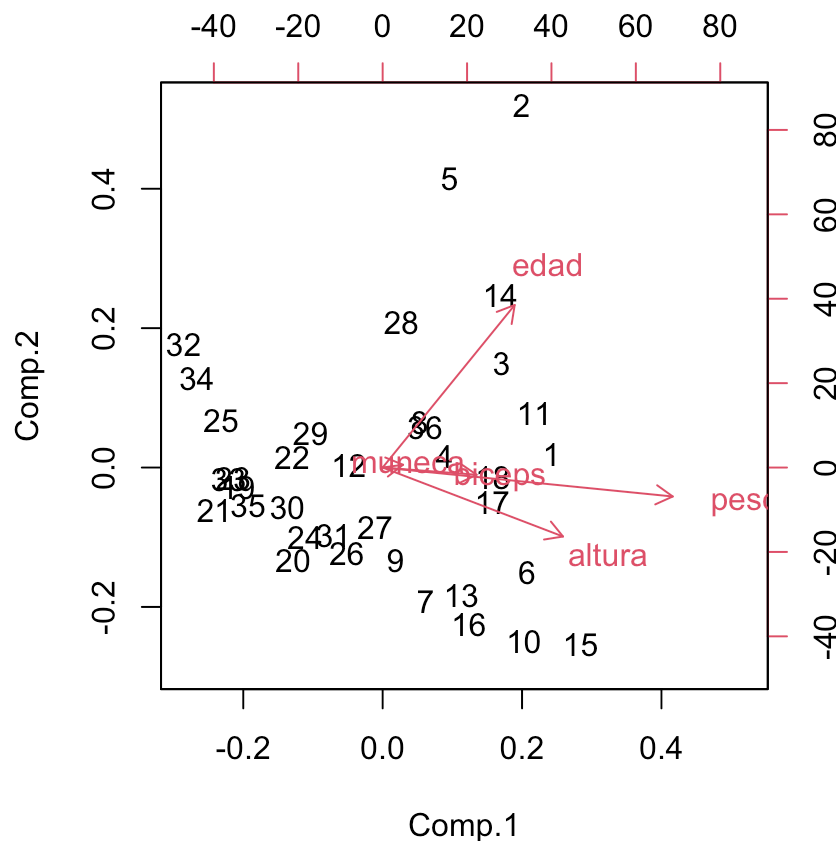
```
plot(cpS$scores[, 1:2], type = "p", main = "Componentes principales usando matriz de varianzas-covarianzas (S)")
```

```
text(cpS$scores[,1], cpS$scores[,2], labels = 1:nrow(X), pos = 4) # Añadir etiquetas
```

Componentes principales usando matriz de varianzas-covarianzas (S)



```
biplot(cpS)
```



Primer gráfico: Este gráfico muestra las proyecciones de las observaciones en el espacio definido por las dos primeras componentes principales. Observaciones como 2, 5, y 14 se encuentran alejadas del centro, lo que sugiere que podrían ser outliers o casos extremos que contribuyen significativamente a la variabilidad capturada por las primeras componentes.

Biplot:

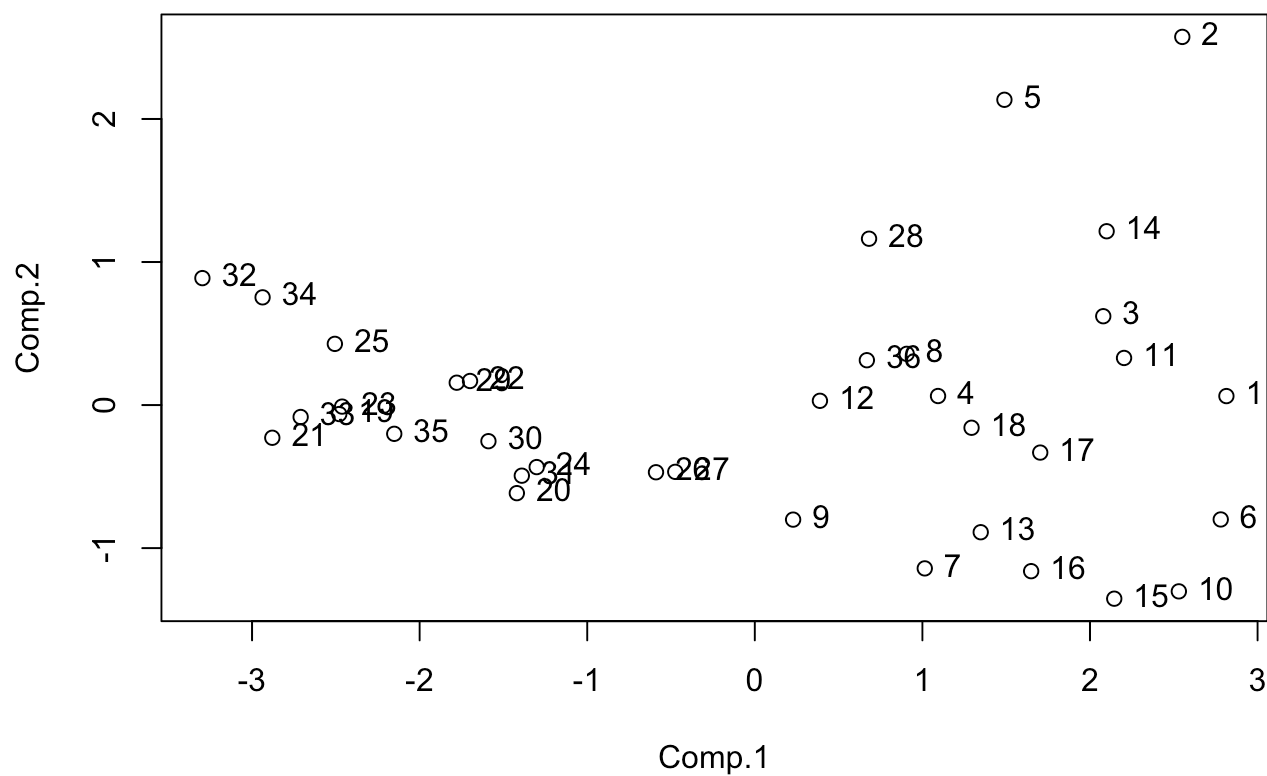
El biplot combina las proyecciones de las observaciones con las proyecciones de las variables originales. Peso y altura son las variables que más contribuyen a la primera componente principal (Comp.1), mientras que edad contribuye mayormente a la segunda componente principal (Comp.2). Las variables peso y altura están fuertemente correlacionadas, como lo indican sus vectores alineados en la misma dirección. Las observaciones 5 y 2 muestran valores altos en edad, mientras que las observaciones cercanas al vector de peso y altura (como 6, 1, y 17) tienen valores altos en estas variables.

Para la matriz de correlación

```
# Calcular el análisis de componentes principales para la matriz R (correlaciones), las
variables se estandarizan automáticamente
cpR <- princomp(X, cor = TRUE)

# Graficar las dos primeras componentes principales
plot(cpR$scores[, 1:2], type = "p", main = "Componentes principales usando matriz de cor
relaciones (R)")
text(cpR$scores[,1], cpR$scores[,2], labels = 1:nrow(X), pos = 4) # Añadir etiquetas
```

Componentes principales usando matriz de correlaciones (R)



```
biplot(cpR)
```

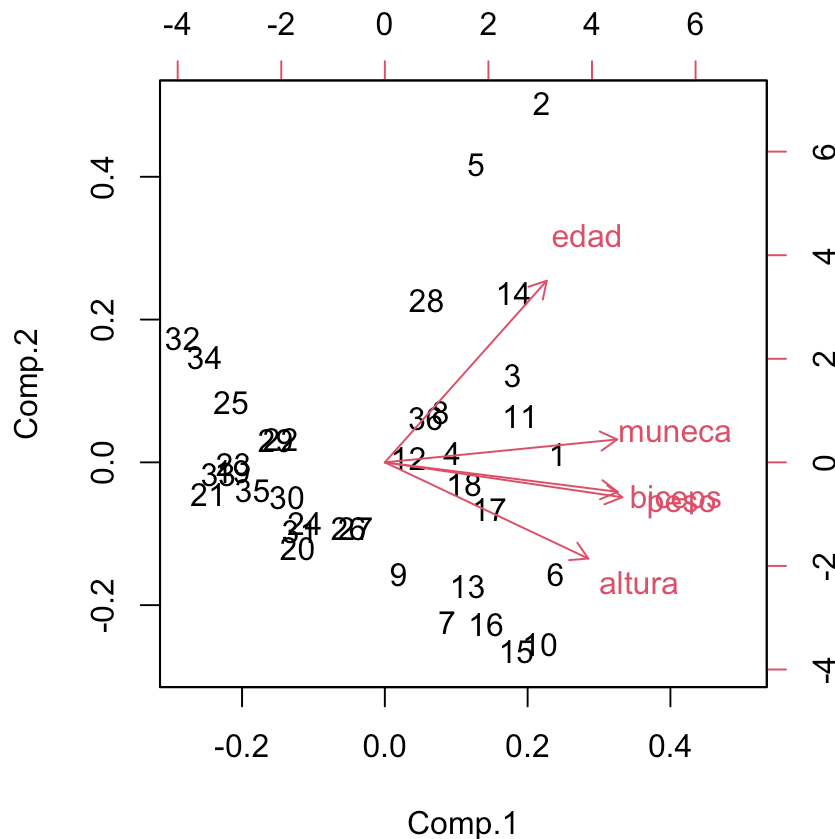



Gráfico de las Dos Primeras Componentes Principales:

Este gráfico muestra las proyecciones de las observaciones en el espacio definido por las dos primeras componentes principales. Las observaciones como 2 y 5 se encuentran alejadas del resto, lo que sugiere que pueden ser outliers o valores atípicos en el conjunto de datos, destacándose en la dirección de edad. Las observaciones que están más agrupadas, como las numeradas 21, 19, 32, 35, sugieren que estas observaciones comparten características comunes que son capturadas por las dos primeras componentes principales.

Biplot:

En el biplot, las flechas representan las variables originales (edad, altura, peso, muñeca, bíceps), y la magnitud y dirección de las flechas nos indican su influencia sobre los componentes principales. Peso, altura y bíceps tienen una correlación fuerte con la primera componente principal (Comp.1), lo que significa que estas variables contribuyen significativamente a la variabilidad explicada por Comp.1. Edad influye principalmente en la segunda componente principal (Comp.2), lo que sugiere que capta información diferente a las otras variables. Las variables peso y altura están altamente correlacionadas, como se observa por la cercanía y alineación de sus vectores en el biplot. Edad no está fuertemente correlacionada con las demás variables, como lo muestra su vector orientado en una dirección diferente, indicando que explica una parte de la variabilidad no capturada por las otras variables.

Explorar funciones de princomp

```
# Resumen del análisis de componentes principales
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
```

```
# Cargar las combinaciones lineales (loadings) de cada componente
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
```

```
# Puntuaciones (scores) de cada observación
cpS$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

```
# Resumen para la matriz de correlaciones
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

```
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
```

cpR\$scores

##		Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
##	[1,]	2.8139915	0.06282760	0.51434516	-0.37618363	-0.161649397
##	[2,]	2.5508161	2.57369731	0.42896223	0.01252075	0.083602262
##	[3,]	2.0792069	0.62112516	-0.12602006	0.51138786	0.430775853
##	[4,]	1.0933160	0.06328171	0.46145821	-0.35236278	-0.008424496
##	[5,]	1.4893629	2.13420572	-0.08620983	-0.19530483	-0.097669770
##	[6,]	2.7801900	-0.79964368	-0.11180511	-0.52796031	0.113681564
##	[7,]	1.0141243	-1.14171806	-0.27787746	0.22743193	0.800375496
##	[8,]	0.9063369	0.35803327	-0.79126430	0.07179533	-0.031461084
##	[9,]	0.2285350	-0.80075813	0.71215644	-0.15394896	0.481123407
##	[10,]	2.5302453	-1.30235901	-0.76205083	0.03215070	0.050616130
##	[11,]	2.2033222	0.32934887	0.10037610	0.49363388	-0.135246631
##	[12,]	0.3885728	0.02978904	-0.70291329	-0.72426251	0.460456523
##	[13,]	1.3480354	-0.88888844	-0.48237353	-0.13878866	-0.248233214
##	[14,]	2.0994018	1.21514134	0.47434543	-0.23319402	-0.019726560
##	[15,]	2.1447355	-1.35354752	0.76511713	0.71259130	-0.587575667
##	[16,]	1.6489148	-1.16117562	-0.85070099	0.08586963	0.111234627
##	[17,]	1.7030809	-0.33209829	0.01673614	0.27827557	0.099895723
##	[18,]	1.2932746	-0.15858301	0.48173868	0.55369253	-0.076249945
##	[19,]	-2.4795617	-0.06280633	0.02839564	-0.11803106	-0.136704692
##	[20,]	-1.4200084	-0.61570309	-0.15277478	-0.25447677	-0.063137788
##	[21,]	-2.8791600	-0.22853227	-0.06023367	-0.03148088	-0.068564803
##	[22,]	-1.6992789	0.16837324	-0.63755548	0.43611800	-0.277172176
##	[23,]	-2.4625686	-0.01072936	-0.59031600	0.26691381	0.024784946
##	[24,]	-1.3015384	-0.43354360	0.20575074	-0.40705451	-0.177314913
##	[25,]	-2.5058729	0.42780280	-0.01308499	-0.30917018	-0.015086855
##	[26,]	-0.5896282	-0.46963951	-0.61738513	-0.25029697	-0.536163469
##	[27,]	-0.4747287	-0.46682854	0.62201914	0.09167385	-0.007586913
##	[28,]	0.6816507	1.16291258	-1.08391248	0.03253793	-0.282947483
##	[29,]	-1.7786024	0.15640801	1.29302710	0.33642964	0.183446578
##	[30,]	-1.5894735	-0.25254138	0.54948615	-0.44020946	-0.006577363
##	[31,]	-1.3903223	-0.49360911	0.76675148	0.17233872	-0.188151664
##	[32,]	-3.2962547	0.88748511	0.06759476	0.35410490	0.371715392
##	[33,]	-2.7100620	-0.08340844	0.02833828	0.31628667	0.201732879
##	[34,]	-2.9371073	0.75312128	-0.93702305	0.36683866	-0.011037680
##	[35,]	-2.1514986	-0.20099407	0.51126095	-0.63846467	-0.074866432
##	[36,]	0.6685529	0.31355440	0.25564126	-0.20140147	-0.201892385

El comando `summary()` brinda un resumen de la varianza explicada por cada componente. El resultado permite ver cómo se distribuye la varianza total entre los diferentes componentes, ayudando a decidir cuántos componentes son importantes.

El objeto `cpS$loadings` contiene las combinaciones lineales de las variables para cada componente. Estos coeficientes te indican qué variables contribuyen más a cada componente principal.

Por otro lado, el comando `cpS$scores` brinda las proyecciones de los puntos en cada una de las componentes.

Parte 3

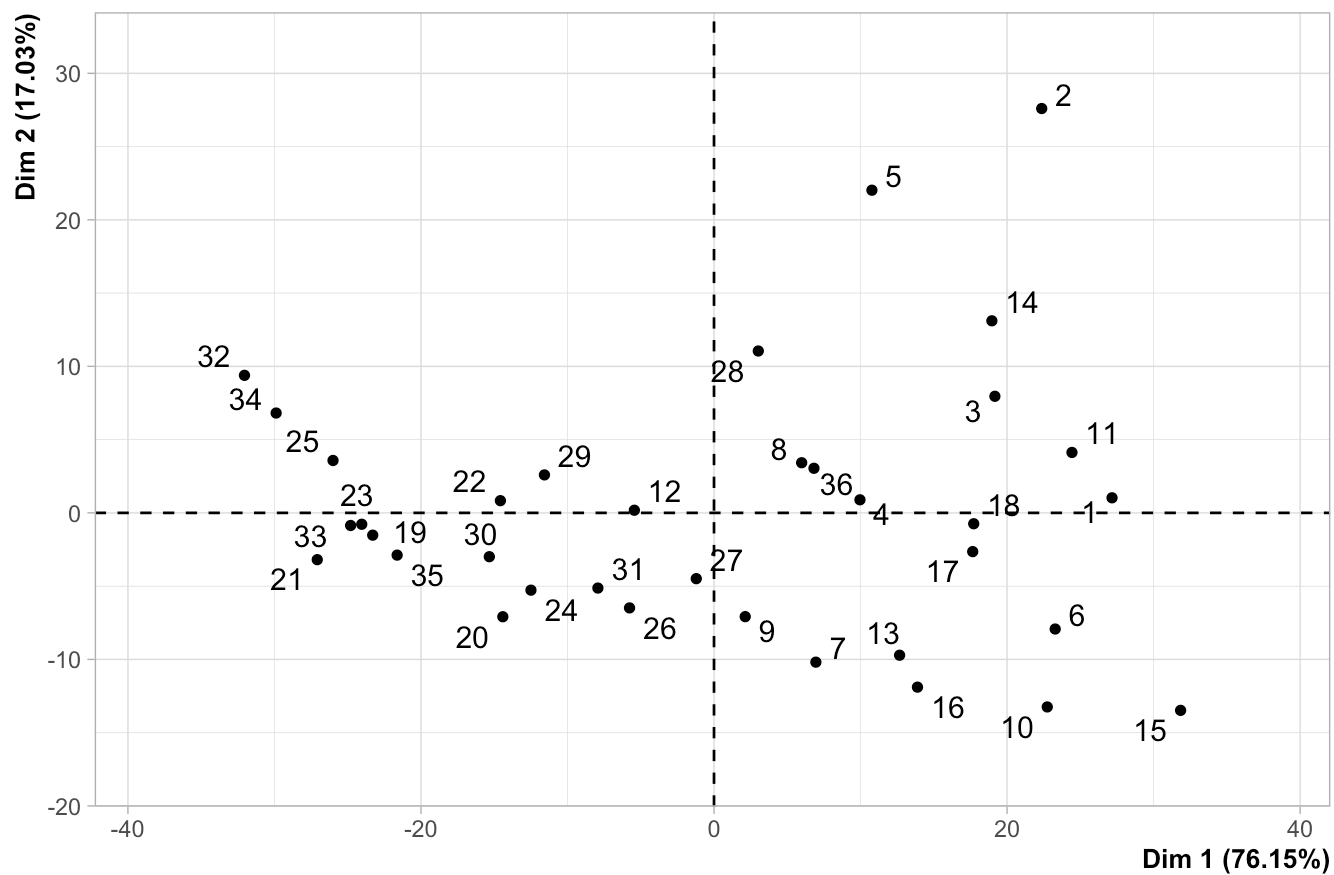
Matriz de covarianza

```
# Cargar las librerías necesarias
library(FactoMineR)
library(ggplot2)
library(factoextra)
```

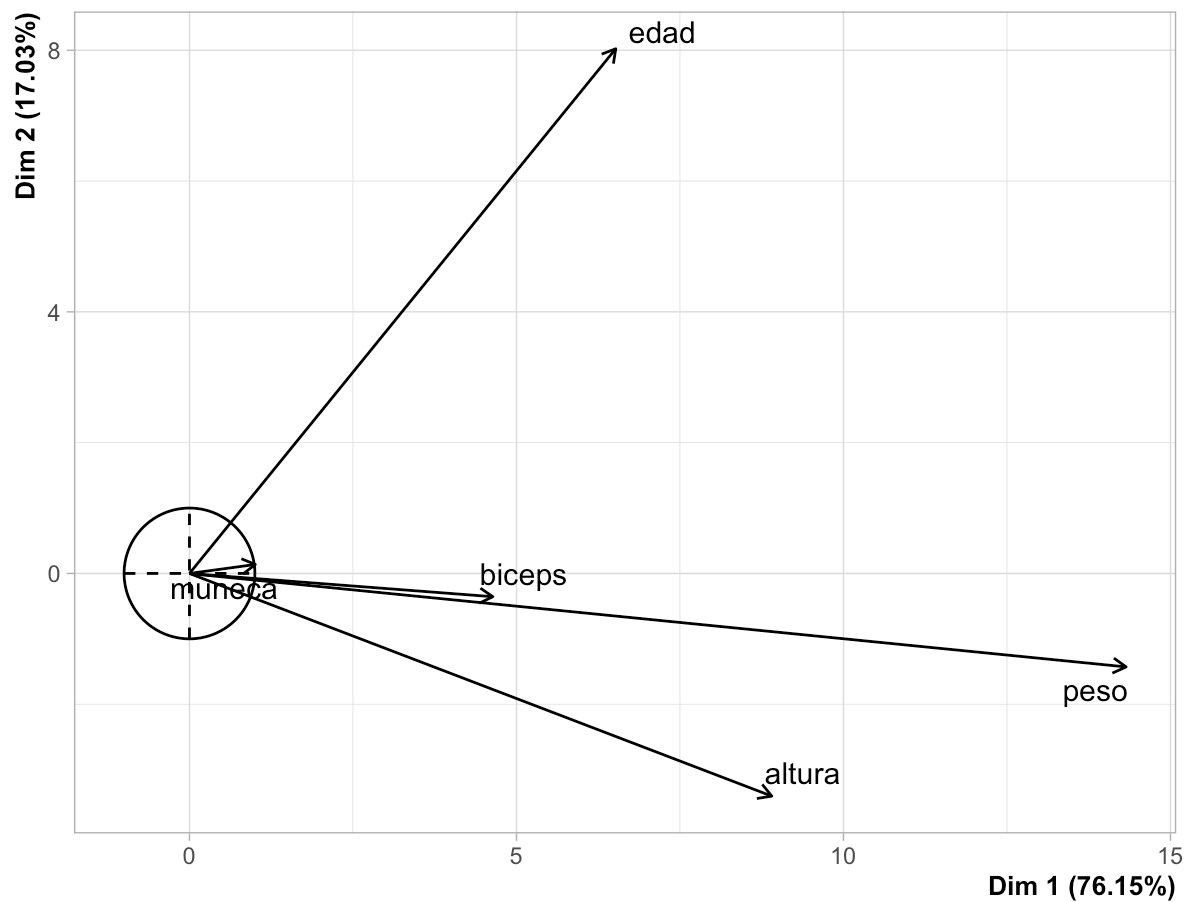
```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```
# Análisis de Componentes Principales con la matriz de varianzas-covarianzas (scale.unit = FALSE)
cpS <- PCA(X, scale.unit = FALSE)
```

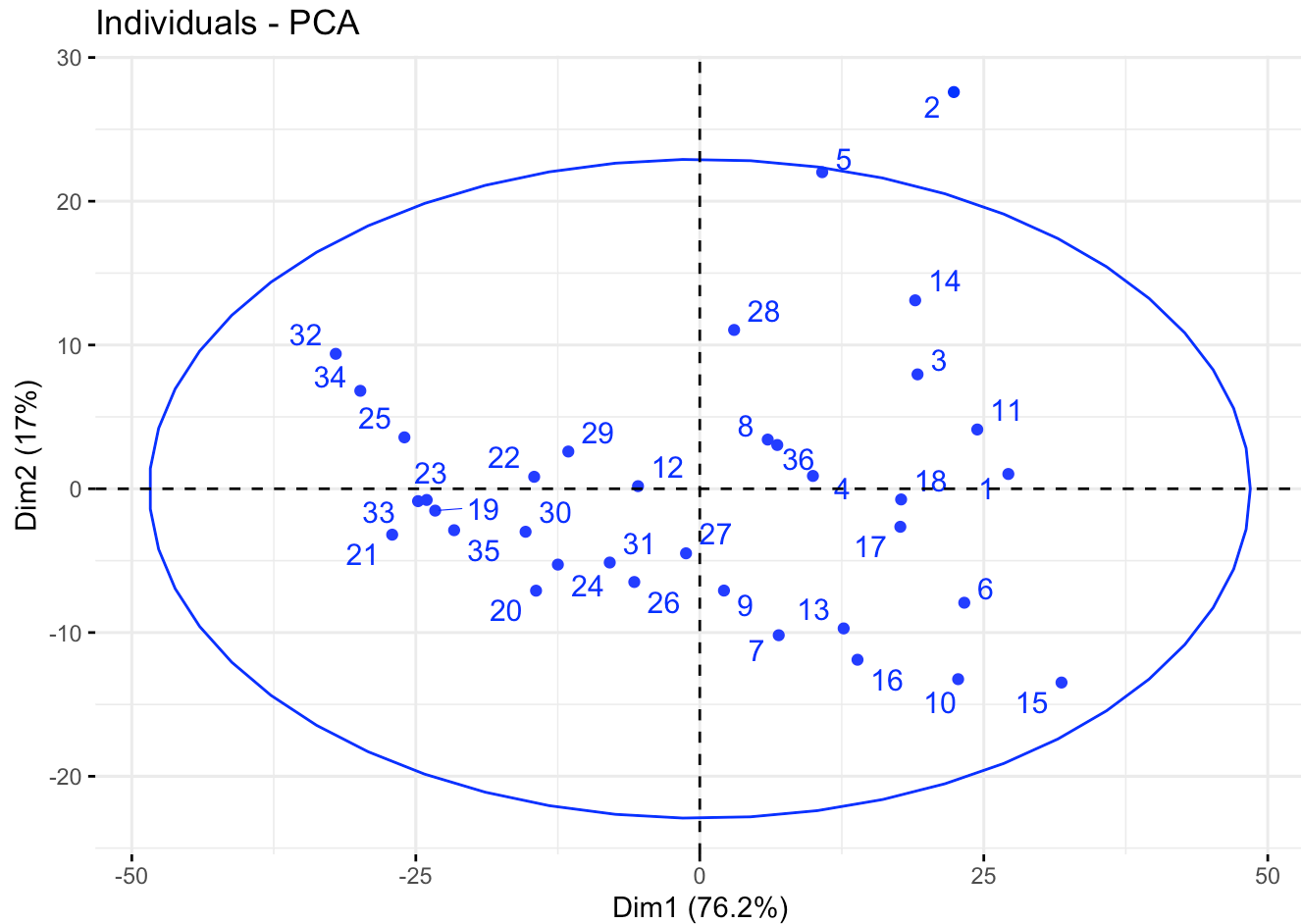
PCA graph of individuals



PCA graph of variables

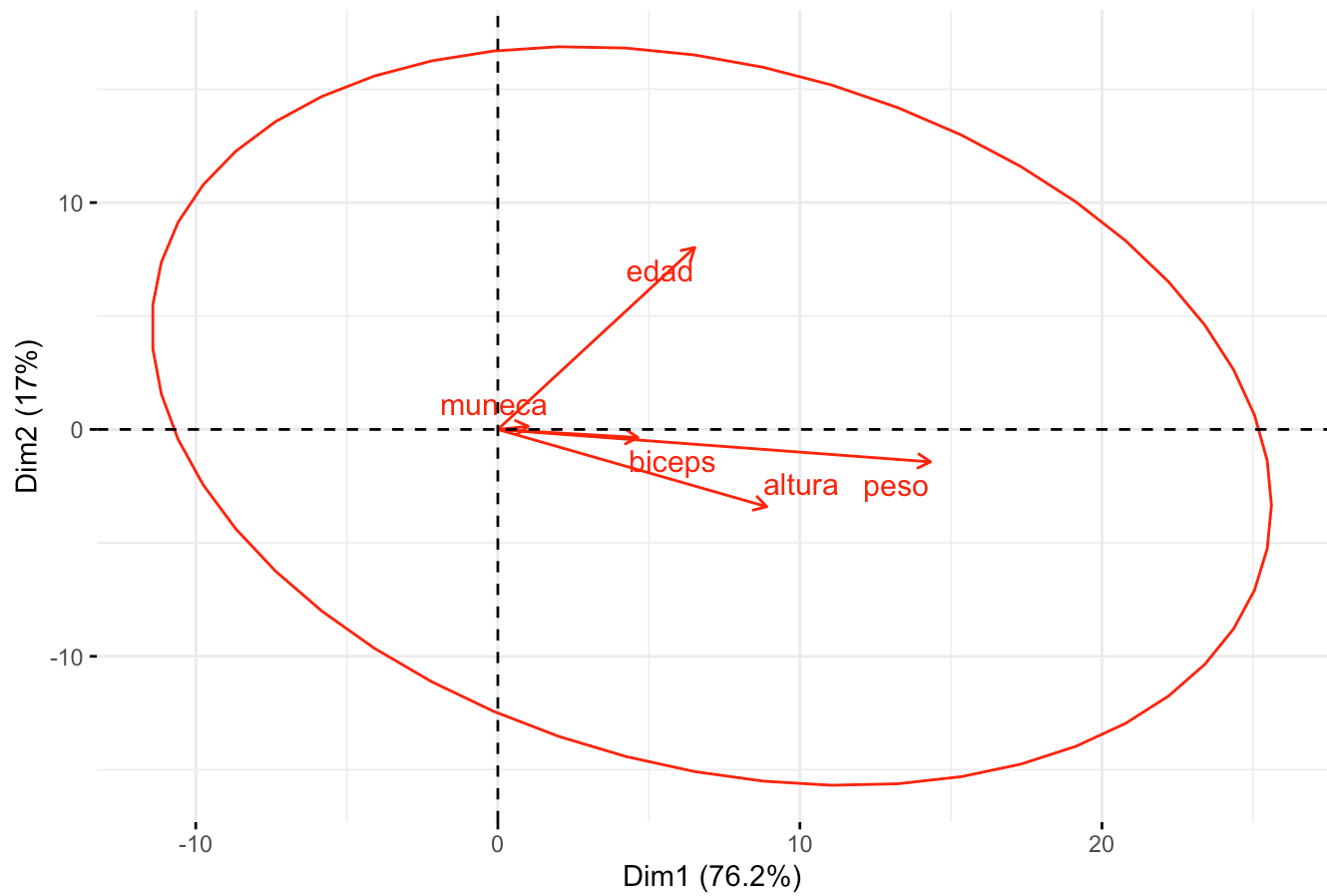


```
# Graficar las observaciones (individuos)
fviz_pca_ind(cpS, col.ind = "blue", addEllipses = TRUE, repel = TRUE)
```

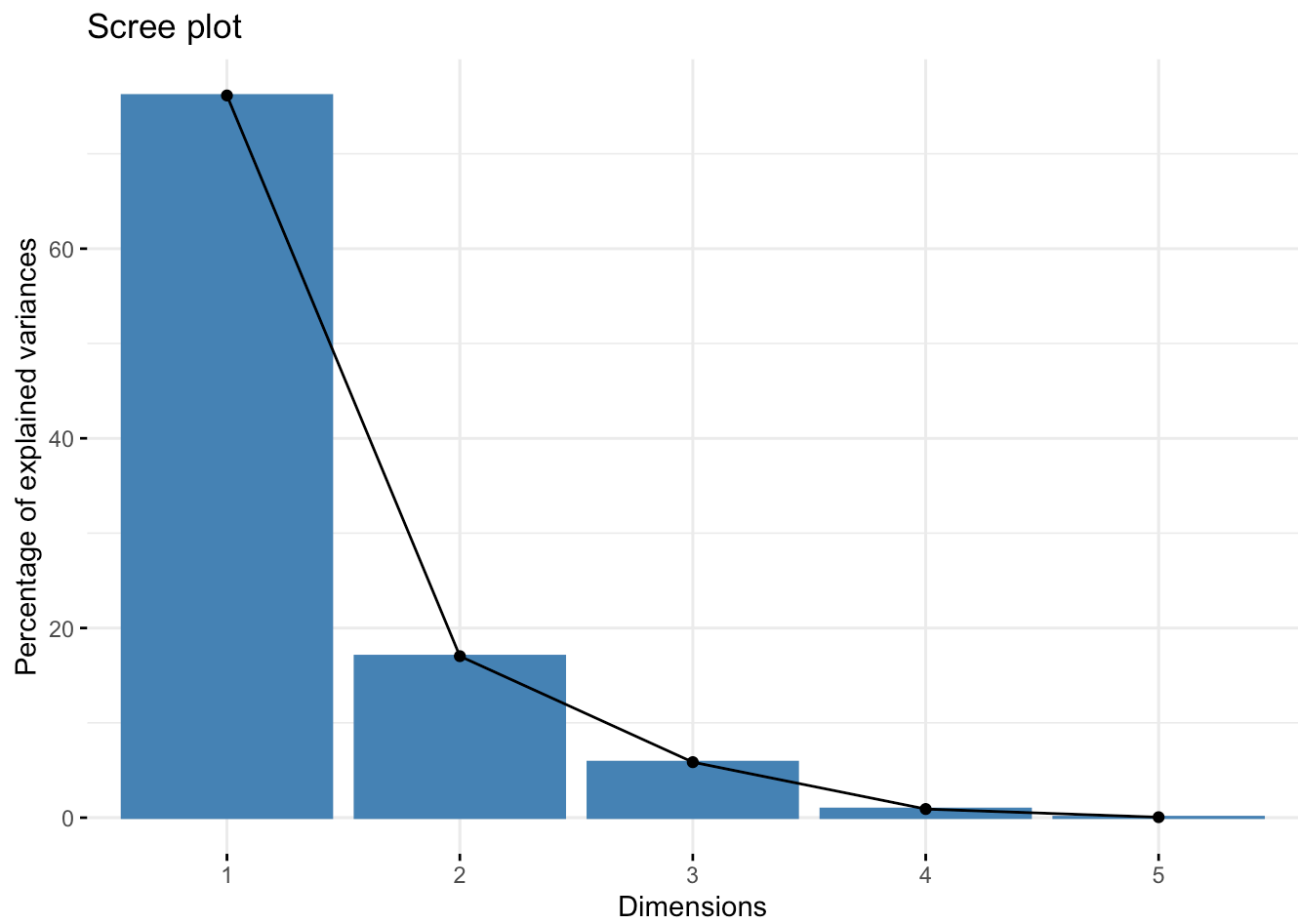


```
# Graficar las variables
fviz_pca_var(cpS, col.var = "red", addEllipses = TRUE, repel = TRUE)
```

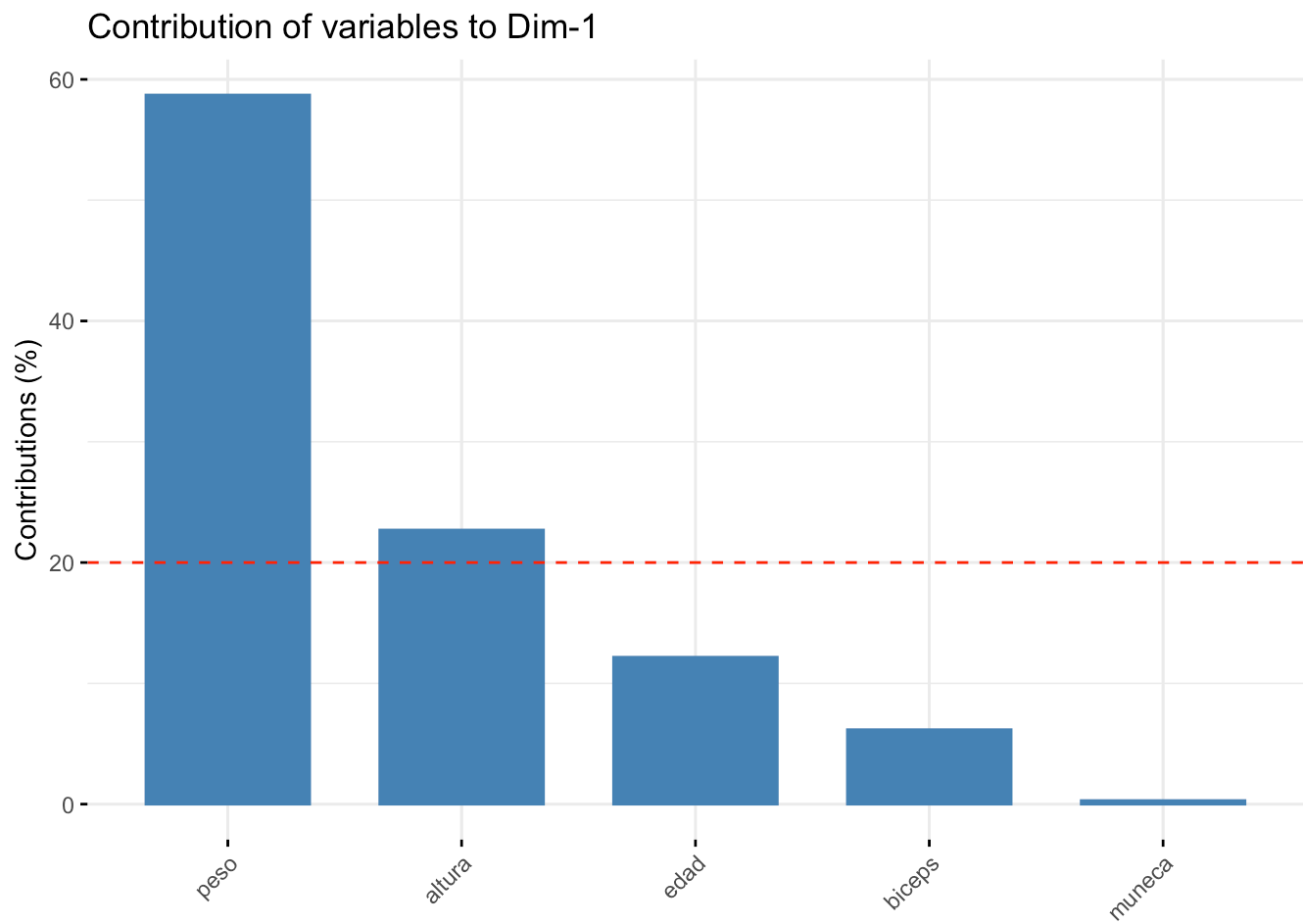

Variables - PCA



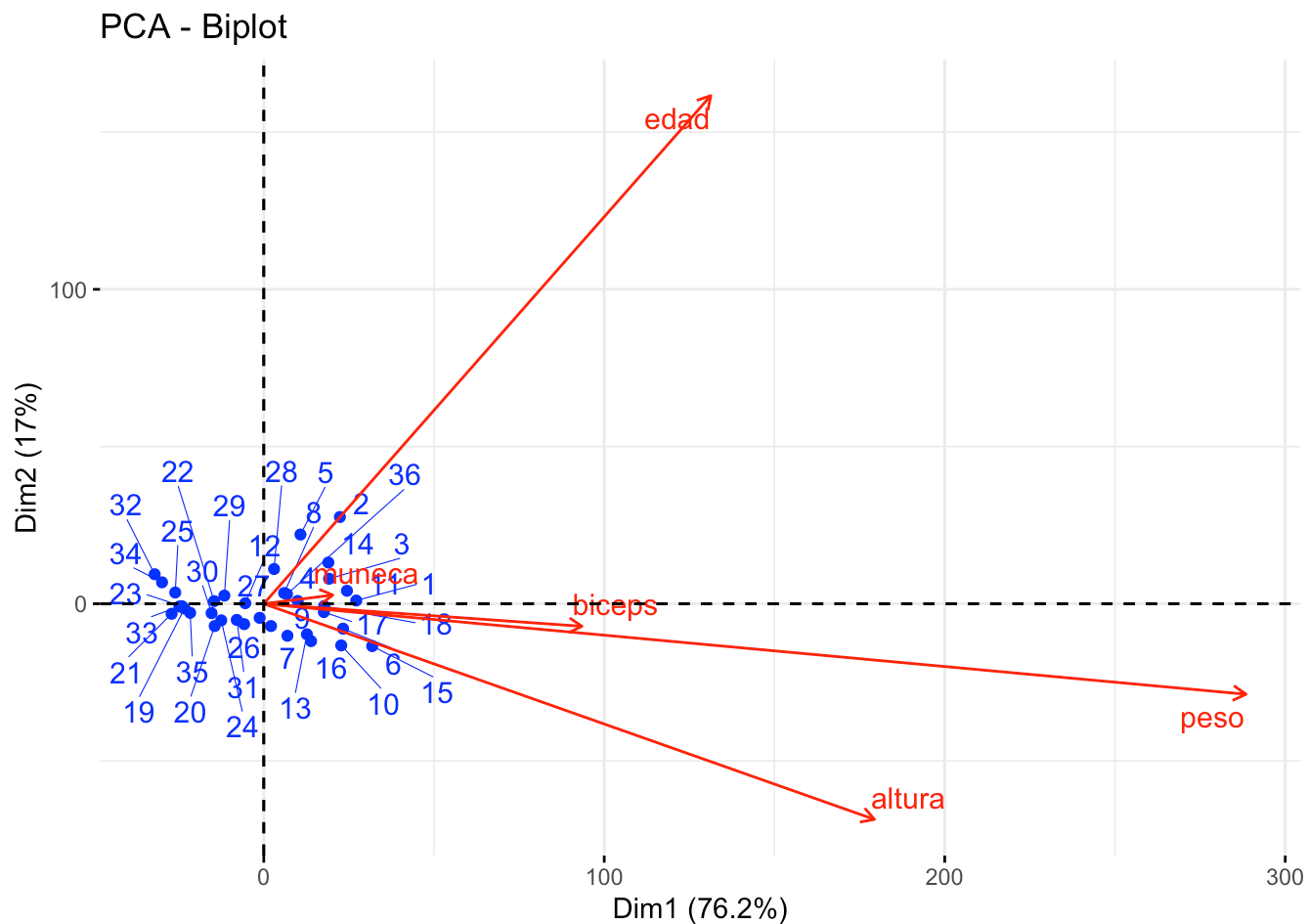
```
# Scree plot para visualizar la proporción de varianza explicada por cada componente  
fviz_screepplot(cpS)
```



```
# Contribución de las variables  
fviz_contrib(cpS, choice = "var")
```



```
# Biplot combinando individuos y variables  
fviz_pca_biplot(cpS, repel = TRUE, col.var = "red", col.ind = "blue")
```



El gráfico de individuos permite observar cada uno de los datos del dataset, lo que permite visualizar si hay agrupaciones (como en el caso de los puntos 33, 23, 19 y 35). Estos grupos representan datos que tienen valores similares en las variables que están presentes en este plano como edad, peso y altura. La gráfica de individuos que contiene la elipse permite identificar outliers más fácilmente, en este caso la observación 2 podría considerarse outlier.

En el gráfico de variables se puede observar que las variables que más contribuyen al CP1 son peso, altura y bíceps, mientras que edad tiene la mayor contribución al CP2. Cuando las variables que tienen un ángulo similar (como peso y altura) significa que tienen correlación entre sí.

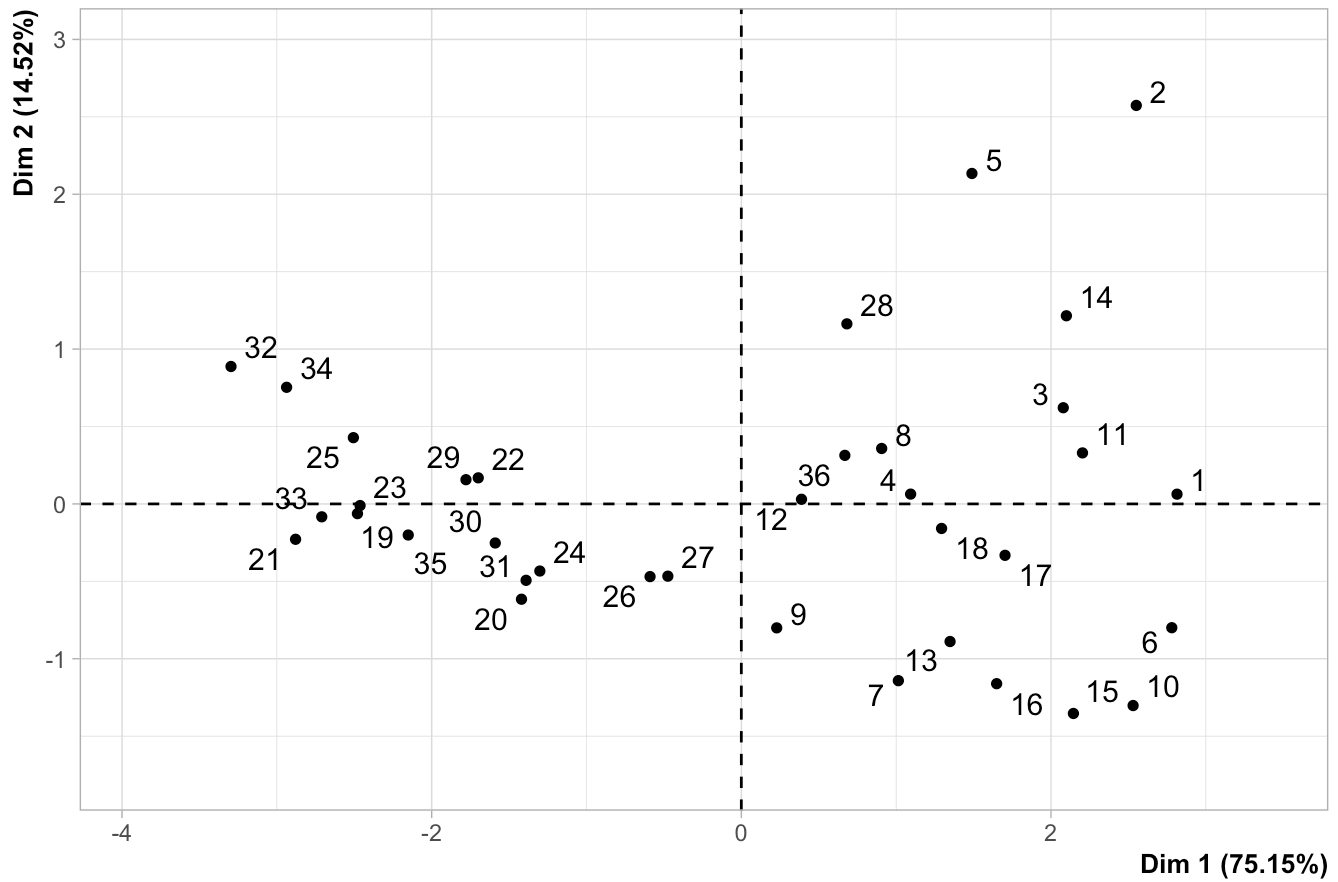
El Scree plot permite observar la variación explicada por cada componente principal en porcentaje, puede ser útil para determinar cuántos componentes principales se desean utilizar, una regla de dedo es basarse en el 'codo' de la gráfica, en este caso éste codo se encuentra en el 2do componente.

En la gráfica de la contribución de variables se puede observar el porcentaje de contribución de cada variable para el primer componente principal, en este caso el que mayor contribución tiene es el peso, con una contribución de casi el 60%

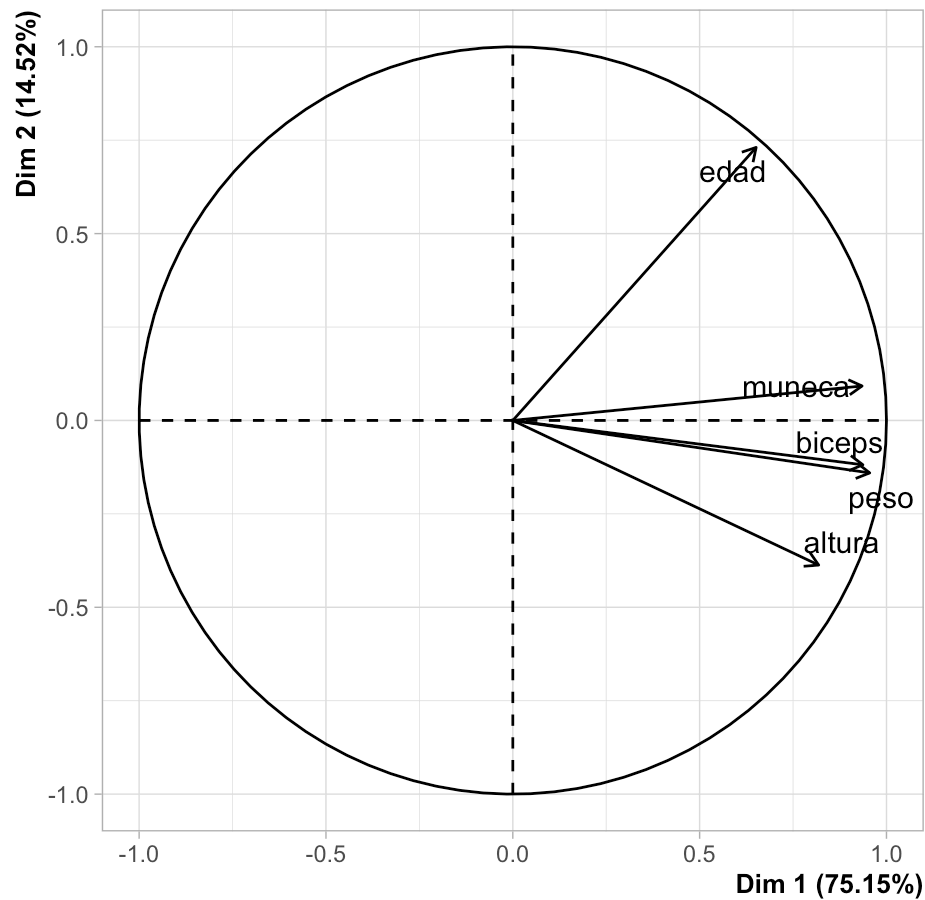
Por último se tiene el biplot, que permite visualizar la gráfica de individuos y de variables al mismo tiempo, esto es útil para ver los valores atípicos y en que dirección lo son, por ejemplo, el valor 2 es atípico en la dirección de edad. ### Matriz de correlaciones

```
# Análisis de Componentes Principales con la matriz de correlaciones (scale.unit = TRUE)
cpR <- PCA(X, scale.unit = TRUE)
```

PCA graph of individuals

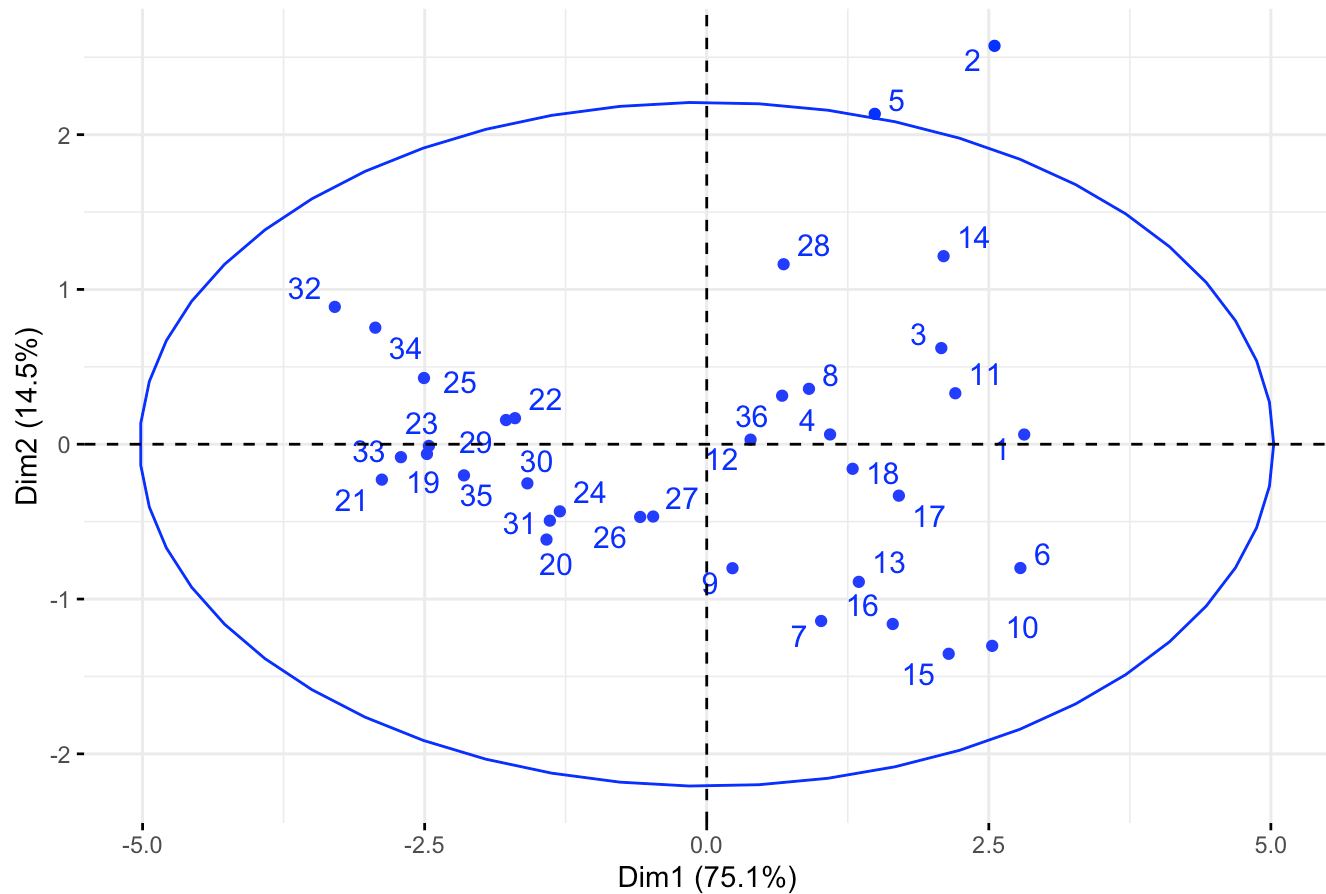


PCA graph of variables



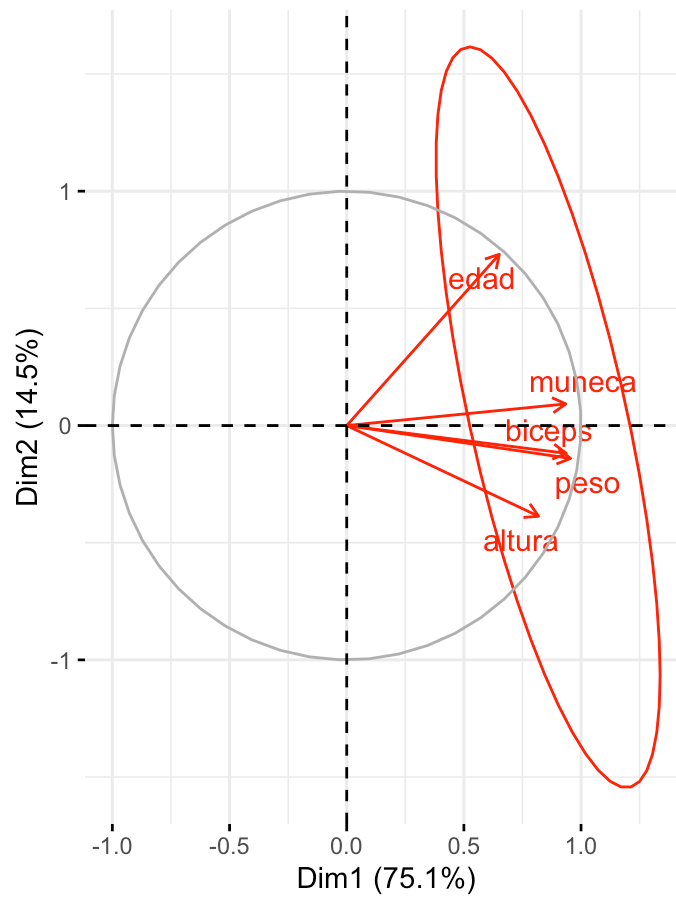
```
# Graficar las observaciones (individuos)
fviz_pca_ind(cpR, col.ind = "blue", addEllipses = TRUE, repel = TRUE)
```

Individuals - PCA

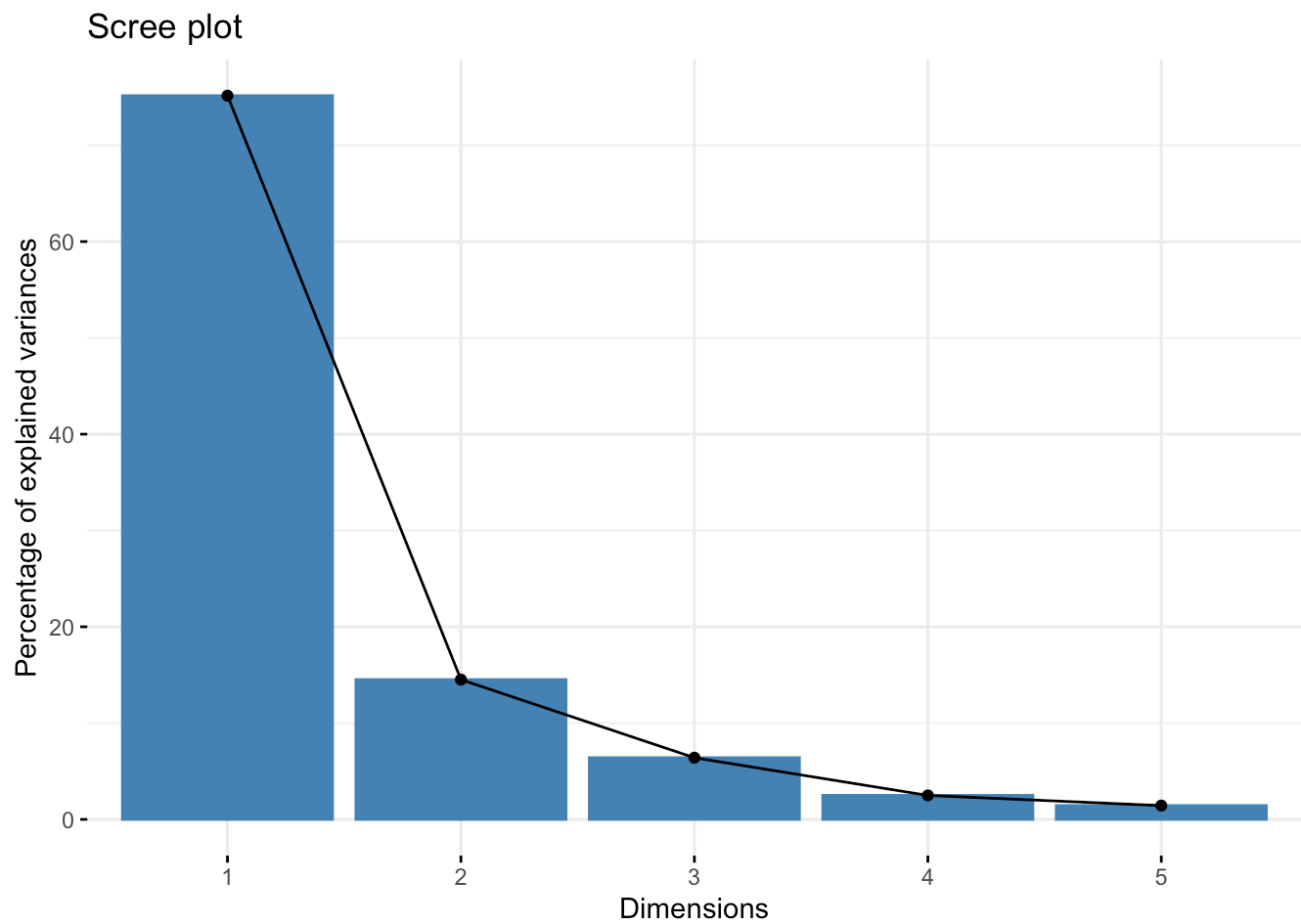


```
# Graficar las variables
fviz_pca_var(cpR, col.var = "red", addEllipses = TRUE, repel = TRUE)
```

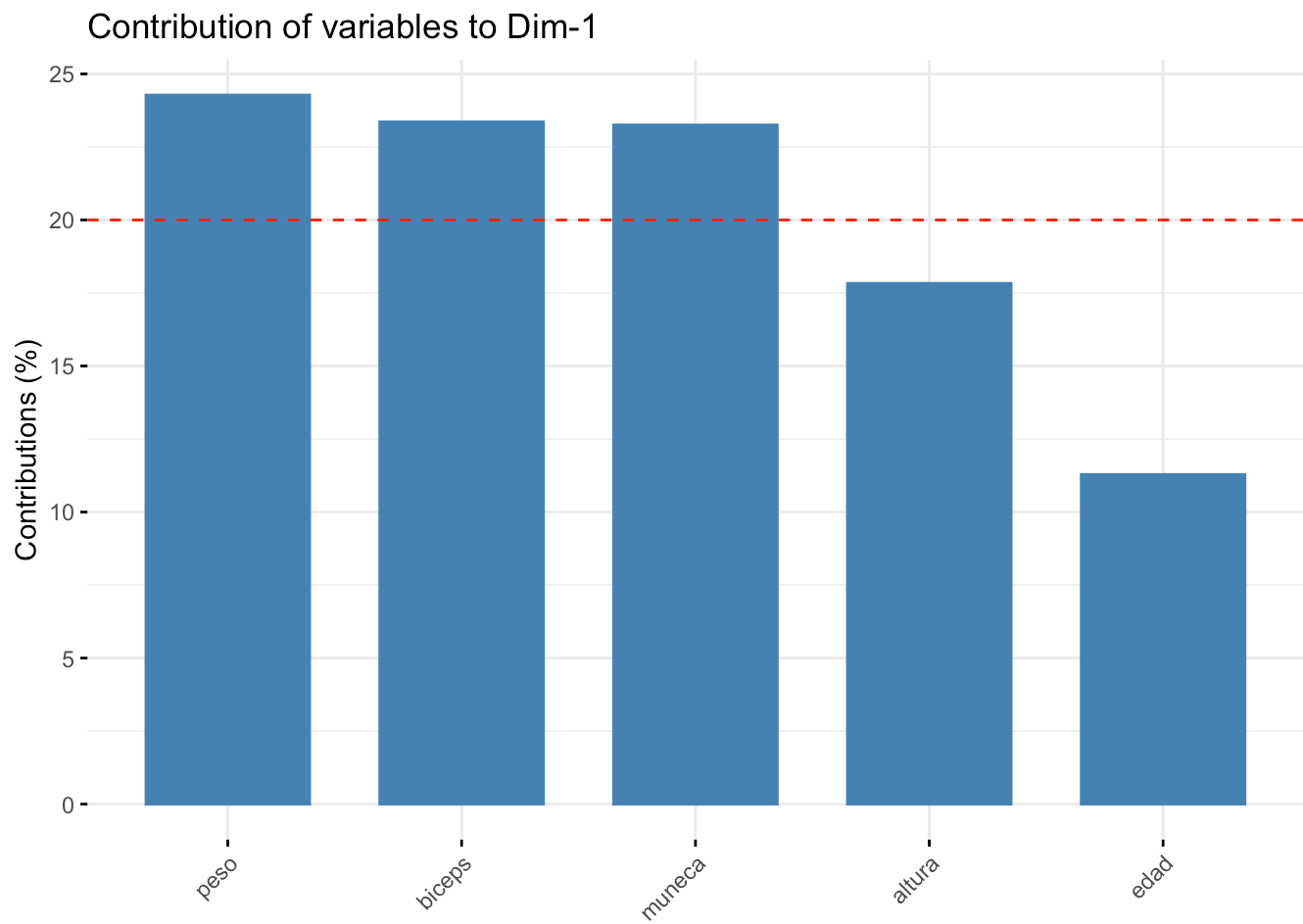
Variables - PCA



```
# Scree plot para visualizar la proporción de varianza explicada por cada componente  
fviz_screepplot(cpR)
```

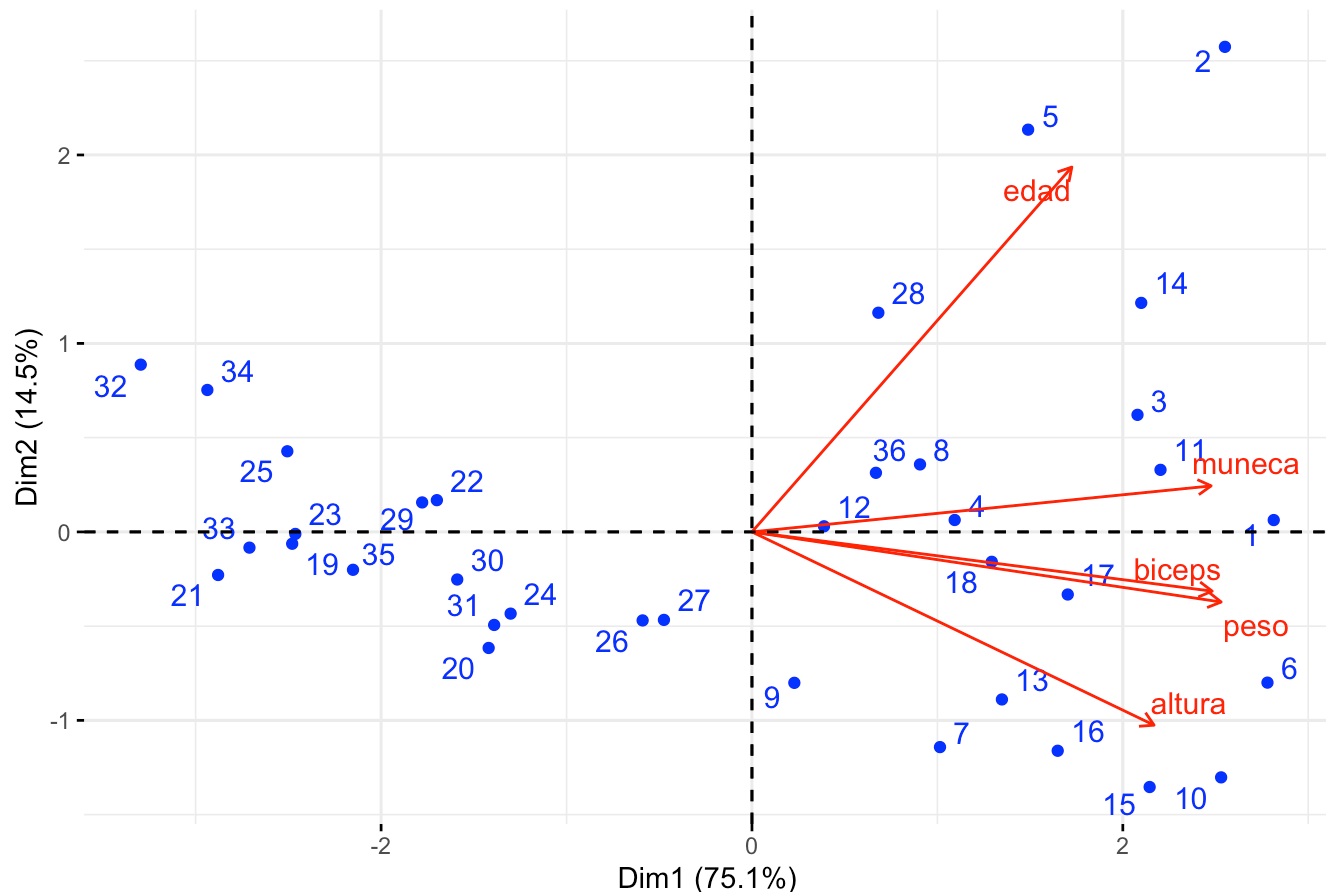


```
# Contribución de las variables  
fviz_contrib(cpR, choice = "var")
```

```
# Biplot combinando individuos y variables  
fviz_pca_biplot(cpR, repel = TRUE, col.var = "red", col.ind = "blue")
```

PCA - Biplot



El valor de individuos permite ver los datos atípicos, los datos 2 y 5 son atípicos en este dataset, además de permitir ver agrupaciones de datos.

En este caso, el gráfico de variables muestra que los vectores tienen magnitudes similares, esto debido a que en este procedimiento se estandarizan los datos. Las variables que más contribuyen al PC1 son muneca, biceps, peso y altura, mientras que para el PC2 edad tiene la mayor contribución con diferencia, esto significa que logra capturar varianza diferente a las demás variables.

El scree plot muestra algo similar al anterior, donde el codo se encuentra en la segunda dimensión, haciendo esta selección la ideal para facilitar el análisis.

De acuerdo con la gráfica de barras de la contribución a la primera dimension principa., las variables muestran una contribución similar, a excepción de edad, que tiene un 12% de contribución aproximadamente.

Por último, el biplot muestra la gráfica de variables y de individuos en uno solo, donde se confirma que el punto 2 y 5 son outliers en dirección de la edad, también hay algunos puntos outliers en otras direcciones como el 1, 6, 10 y 15.

Parte 4

Matriz de Varianza-Covarianza (S):

Este análisis se realiza sin estandarizar las variables, lo que significa que las variables con mayores varianzas (en este caso, variables como peso y altura) tienen una influencia mayor en los componentes principales. En el gráfico de variables y en el biplot, observamos que variables como peso y altura contribuyen significativamente a

la primera componente principal (Comp.1), lo que indica que estas variables explican gran parte de la varianza en los datos. Matriz de Correlación (R):

En este análisis, las variables son estandarizadas, por lo que todas tienen igual peso inicial, independientemente de su escala original. Este enfoque es útil cuando las variables tienen escalas muy diferentes, lo que nos permite hacer comparaciones equitativas.

El análisis con la matriz de correlaciones resulta más adecuado, ya que las variables tienen escalas diferentes y al estandarizarlas podemos hacer una comparación más justa entre ellas.

Las variables que mas contribuyen a los primeros dos componentes de la matriz de correlación son:

CP1: peso y biceps CP2: edad y altura

Las combinaciones lineales de los primeros dos componentes principales son las siguientes:

$$PC1 = -0.336 * edad - 0.493 * peso - 0.422 * altura - 0.482 * muñeca - 0.483 * biceps$$

$$PC2 = 0.857 * edad - 0.165 * peso - 0.454 * altura - 0.108 * muñeca - 0.139 * biceps$$