

# A2-Componentes Principales

Diego Alberto Baños Lopez  
A01275100

2023-09-28

## Parte I

A partir de los datos sobre indicadores económicos y sociales de 96 países hacer una análisis de Componentes principales a partir de la matriz de varianzas-covarianzas y otro a partir de la matriz de correlaciones , comparar los resultados y argumentar cuál es mejor según los resultados obtenidos.

1. Cargar los datos y calcular las matrices de varianza-covarianza S con  $\text{cov}(X)$  y la matriz de correlaciones con  $\text{cor}(X)$

```
M <- read.csv("países_mundo.csv")  
# Matriz de varianzas-covarianzas  
cov_matrix <- cov(M)  
# Matriz de correlaciones  
cor_matrix <- cor(M)
```

2. Calcular los valores y vectores propios de cada matriz.

```
vec_cov <- eigen(cov_matrix)  
vec_cor <- eigen(cor_matrix)
```

3. Calcular la proporción de varianza explicada por cada componente.

```
var_cov <- vec_cov$values/sum(diag(cov_matrix))  
var_cor <- vec_cor$values/sum(diag(cor_matrix))
```

4. Acumule los resultados anteriores.

```
cumsum_cov <- cumsum(var_cov)  
cumsum_cor <- cumsum(var_cor)
```

```
## [1] "-----"
```

```
## [1] "Usando matriz de varianza-covarianza:"
```

```
## [1] "-----"
```

```

## [1] "Vectores propios:"

## eigen() decomposition
## $values
## [1] 6.163576e+10 6.581612e+09 4.636256e+06 3.107232e+05 1.216015e+04
## [6] 5.137767e+02 3.627885e+02 4.542082e+01 5.800868e+00 1.438020e+00
## [11] 4.768083e-01
##
## $vectors
##           [,1]           [,2]           [,3]           [,4]           [,5]
## [1,] -1.658168e-06 4.706785e-07 0.0001263736 -1.928408e-05 -0.0055373971
## [2,] -4.048139e-05 -1.774254e-05 0.0082253821 -2.493257e-03 -0.0944030203
## [3,] 5.739096e-06 -1.084543e-05 0.0001318149 5.538307e-03 0.0314036410
## [4,] 8.880376e-01 4.597632e-01 0.0026022071 -3.893588e-04 -0.0003327409
## [5,] 4.597636e-01 -8.880405e-01 0.0005694896 1.096305e-03 0.0002207819
## [6,] 3.504341e-04 4.016179e-04 -0.0619424889 7.641174e-03 0.9921404486
## [7,] 2.625508e-04 -1.122118e-03 -0.0401453227 -9.991411e-01 0.0057795144
## [8,] 4.089564e-06 7.790843e-06 0.0012719918 6.435797e-03 0.0419331615
## [9,] -1.073825e-06 2.350808e-07 0.0001916177 4.043796e-05 -0.0018090751
## [10,] 2.547156e-03 7.126782e-04 -0.9972315499 3.973568e-02 -0.0625729475
## [11,] 4.643724e-06 -1.315731e-06 -0.0020679047 -5.626049e-05 -0.0042367120
##           [,6]           [,7]           [,8]           [,9]           [,10]
## [1,] 1.243456e-02 5.359089e-03 -8.390810e-02 -6.778358e-02 -1.158091e-01
## [2,] 9.917515e-01 2.258020e-02 -7.891128e-02 -1.637836e-02 4.264872e-04
## [3,] 8.552992e-02 -1.136481e-01 9.856498e-01 -1.468464e-02 8.241465e-03
## [4,] -8.621005e-06 -7.566477e-06 1.217248e-05 -3.971469e-07 4.274451e-07
## [5,] 1.955408e-05 1.544658e-05 -2.558998e-05 1.059471e-06 -1.353881e-06
## [6,] 9.109622e-02 4.748682e-02 -3.416812e-02 -5.379549e-03 -3.409423e-03
## [7,] -1.087229e-03 -6.863294e-03 4.698731e-03 7.965261e-05 3.621425e-05
## [8,] 1.721948e-02 -9.920538e-01 -1.169638e-01 1.416566e-03 5.891758e-03
## [9,] 1.758667e-03 -7.455427e-03 1.811443e-02 1.283039e-01 -9.859317e-01
## [10,] 2.639673e-03 -3.764707e-03 1.267052e-03 2.262931e-03 2.672618e-04
## [11,] -1.877994e-02 -1.709137e-03 -5.204823e-03 -9.891529e-01 -1.200519e-01
##           [,11]
## [1,] 9.872887e-01
## [2,] -2.092491e-02
## [3,] 8.344324e-02
## [4,] 2.723996e-07
## [5,] -2.086857e-07
## [6,] 4.944397e-04
## [7,] 4.780416e-04
## [8,] -3.748976e-03
## [9,] -1.052934e-01
## [10,] 5.906241e-05
## [11,] -8.221371e-02

## [1] "-----"

## [1] "Proporción de varianza explicada por cada componente:"

## [1] 9.034543e-01 9.647298e-02 6.795804e-05 4.554567e-06 1.782429e-07
## [6] 7.530917e-09 5.317738e-09 6.657763e-10 8.502887e-11 2.107843e-11
## [11] 6.989035e-12

```

```

## [1] "-----"

## [1] "Varianza acumulada por cada componente:"

## [1] 0.9034543 0.9999273 0.9999953 0.9999998 1.0000000 1.0000000 1.0000000
## [8] 1.0000000 1.0000000 1.0000000 1.0000000

## [1] "-----"

## [1] "Usando matriz de correlaciones:"

## [1] "-----"

## [1] "Vectores propios:"

## eigen() decomposition
## $values
## [1] 4.02987902 1.92999195 1.37041115 0.86451597 0.79414057 0.72919997
## [7] 0.57130511 0.32680096 0.16806846 0.14632819 0.06935866
##
## $vectors
##           [,1]      [,2]      [,3]      [,4]      [,5]      [,6]
## [1,] -0.314119414  0.34835747 -0.07352541 -0.44028717 -0.32972147 -0.18392437
## [2,] -0.392395442 -0.04136238 -0.17759254 -0.13398483  0.08340489 -0.08656390
## [3,]  0.116546319 -0.58283641  0.16686305  0.05865031  0.18654100  0.16835650
## [4,]  0.295393771 -0.17690839 -0.53343025 -0.26248209 -0.14110658  0.04653378
## [5,]  0.258964724 -0.17356372 -0.61438847 -0.17389644 -0.07521971  0.02821905
## [6,]  0.446082934 -0.02719077  0.15177250  0.04959796 -0.05416498  0.02442175
## [7,]  0.092410503  0.32060987 -0.37024258  0.73603097  0.02671021 -0.30940890
## [8,]  0.005692925 -0.45742697  0.16480339  0.04024882 -0.41531702 -0.75356463
## [9,] -0.243652293 -0.15408201 -0.02961449  0.33650345 -0.73261463  0.50894232
## [10,] 0.415029554  0.23286257  0.20608749 -0.06730166 -0.23100421  0.05806466
## [11,] 0.374531032  0.29168698  0.20631751 -0.14843513 -0.24028756 -0.02809233
##           [,7]      [,8]      [,9]      [,10]     [,11]
## [1,]  0.1628974320 -0.09481963  0.52181220  0.34674573 -0.10062784
## [2,]  0.6398040762 -0.32307802 -0.29031618 -0.38959240  0.17487096
## [3,]  0.5310867107  0.05209889  0.23599758  0.42854658 -0.16786800
## [4,] -0.1490207046 -0.44913216 -0.36995675  0.34911534 -0.15247432
## [5,]  0.1082745817  0.50343911  0.30681318 -0.33770404  0.12366382
## [6,] -0.0008501608 -0.56975094  0.44733110 -0.20997673  0.44992596
## [7,]  0.2357666690 -0.05962470  0.08358225  0.20561803 -0.07067780
## [8,] -0.0806036686  0.04275404 -0.07438520 -0.08671232 -0.01493710
## [9,]  0.0112333588 -0.01607505 -0.01868615 -0.03209758  0.07259619
## [10,] 0.2711228006 -0.05023582 -0.04339752 -0.36147417 -0.67912543
## [11,] 0.3352822144  0.30978009 -0.37666244  0.28779437  0.46737561

## [1] "-----"

## [1] "Proporción de varianza explicada por cada componente:"

## [1] 0.366352638 0.175453813 0.124582832 0.078592361 0.072194597 0.066290906
## [7] 0.051936828 0.029709178 0.015278951 0.013302563 0.006305332

```

```
## [1] "-----"

## [1] "Varianza acumulada por cada componente:"

## [1] 0.3663526 0.5418065 0.6663893 0.7449816 0.8171762 0.8834671 0.9354040
## [8] 0.9651132 0.9803921 0.9936947 1.0000000

## [1] "-----"
```

5. Según los resultados anteriores

1. ¿Qué componentes son los más importantes?
  1. Matriz de varianza-covarianza:
    1. Los dos primeros componentes son los más importantes. El primero explica aproximadamente el 90.3% de la varianza, mientras que el segundo explica aproximadamente el 9.6%
  2. Matriz de correlaciones:
    1. Los dos primeros componentes también son los más significativos, pero con proporciones menores. El primero explica el 36.6% de la varianza y el segundo el 17.5%.
2. ¿Qué variables son las que más contribuyen a la primera y segunda componentes principales?
  1. Se identifican observando los valores más altos en magnitud en las columnas [,1] y [,2] de los vectores propios para ambas matrices.
3. ¿Por qué lo dice?
  1. Los vectores propios indican la dirección de máxima varianza. Las variables con coeficientes más altos en estos vectores son las más contribuyentes.
4. ¿Influyen las unidades de las variables?
  1. Matriz de varianza-covarianza:
    1. Sí, las unidades de las variables influyen. Las variables con valores más grandes pueden dominar el análisis.
  2. Matriz de correlaciones:
    1. No, porque al usar la matriz de correlaciones, todas las variables se estandarizan para tener una varianza de 1.
6. Compare los resultados. ¿qué concluye?
  1. Al comparar los resultados, se puede observar que la matriz de varianza-covarianza y la matriz de correlaciones dan diferentes proporciones de varianza explicada para las componentes principales. Esto se debe a que la matriz de correlaciones estandariza las variables, mientras que la matriz de varianza-covarianza no.

## Parte II

- Obtenga las gráficas de respectivas con S (matriz de varianzas-covarianzas) y con R (matriz de correlaciones) de las dos primeras componentes e interprete los resultados en término de agrupación de variables

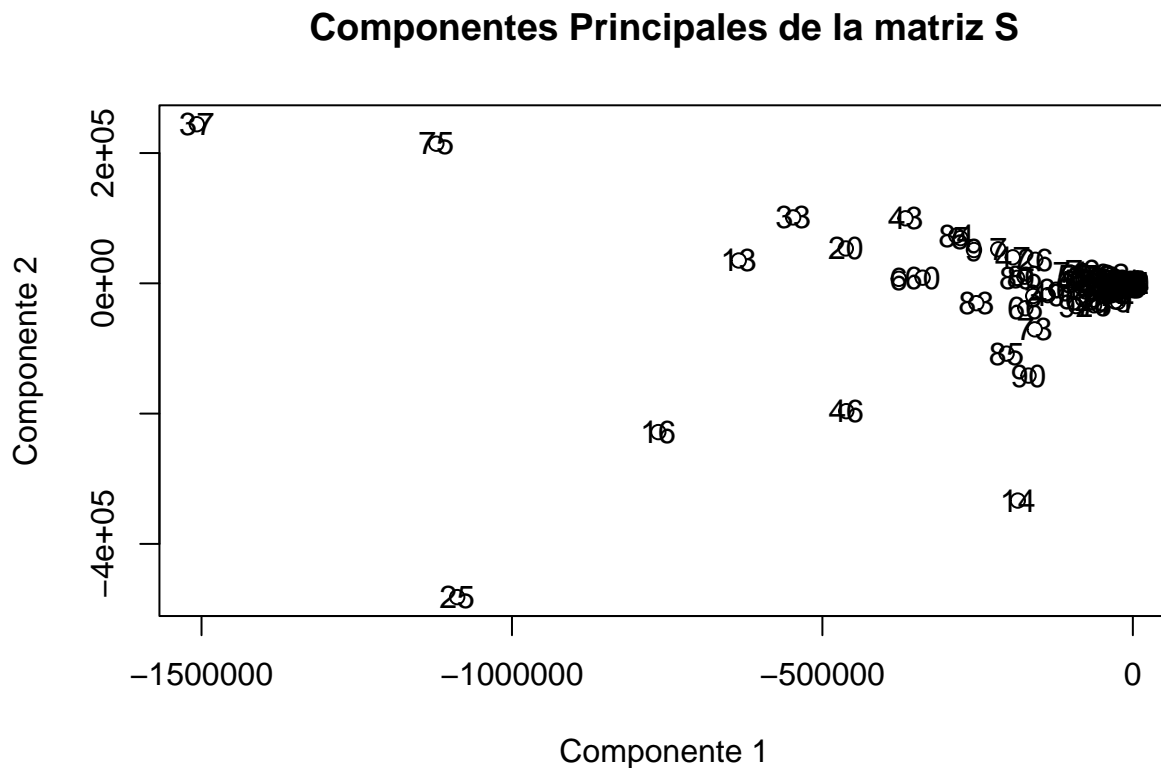
```
library(stats)
library(ggplot2)
library(FactoMineR)
library(factoextra)
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3Wba
```

```

cpS <- princomp(M, cor = FALSE)
cpaS <- as.matrix(M) %*% cpS$loadings
plot(cpaS[, 1], cpaS[, 2], type = "p", main = "Componentes Principales de la matriz S",
     xlab = "Componente 1", ylab = "Componente 2")
text(cpaS[, 1], cpaS[, 2], labels = row.names(M))

```



```
biplot(cpS)
```

```
## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped
```

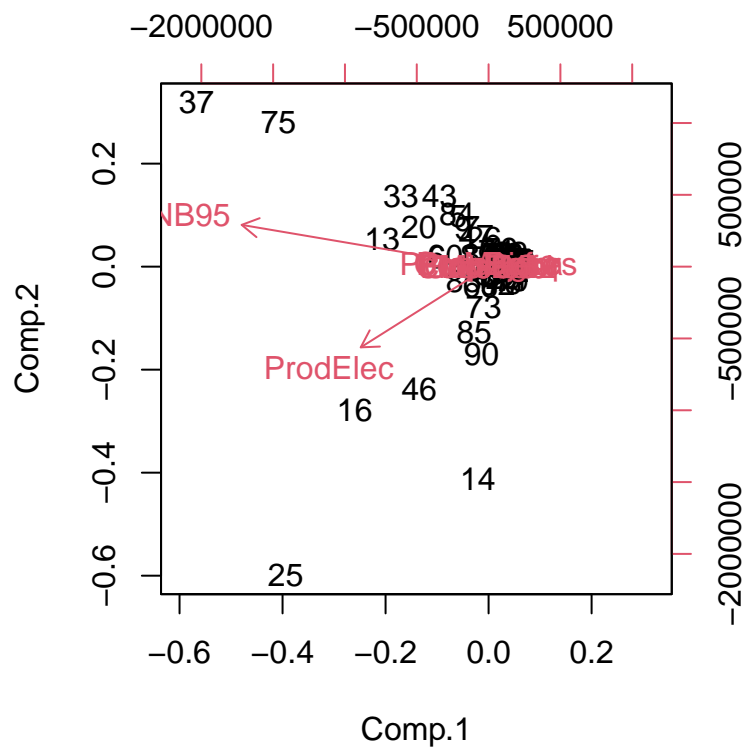
```
## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped
```

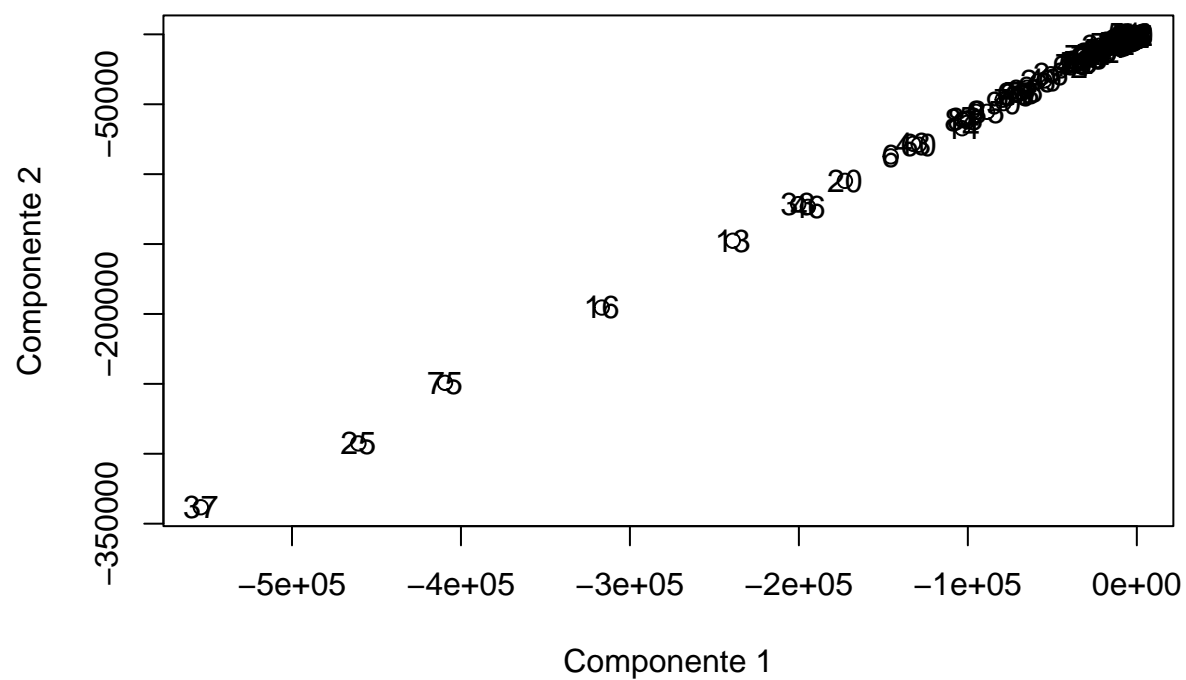
```
## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped
```

```
## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped
```

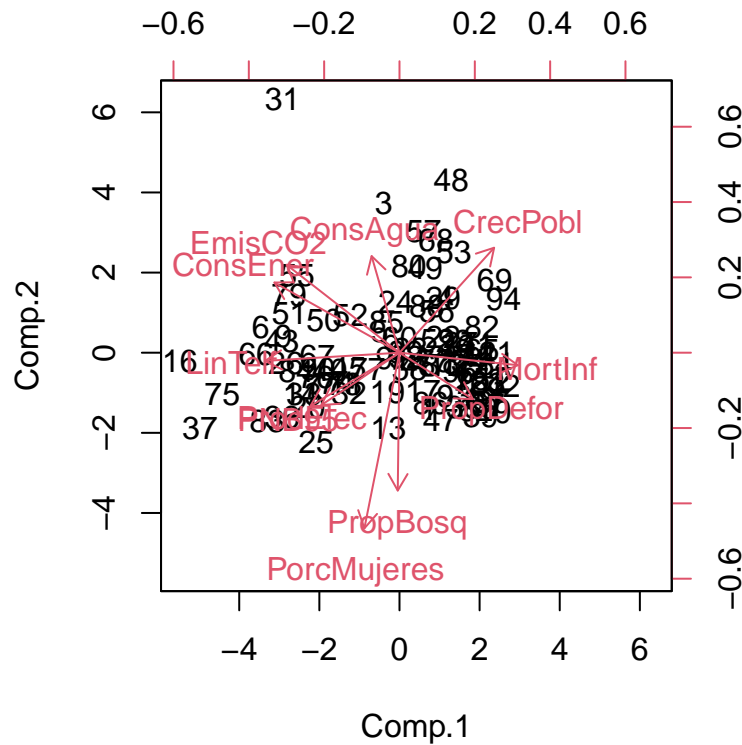


```
cpR <- princomp(M, cor = TRUE)
cpaR <- as.matrix(M) %*% cpR$loadings
plot(cpaR[, 1], cpaR[, 2], type = "p", main = "Componentes Principales de la matriz R",
      xlab = "Componente 1", ylab = "Componente 2")
text(cpaR[, 1], cpaR[, 2], labels = row.names(M))
```

## Componentes Principales de la matriz R



```
biplot(cpR, scale = 0)
```



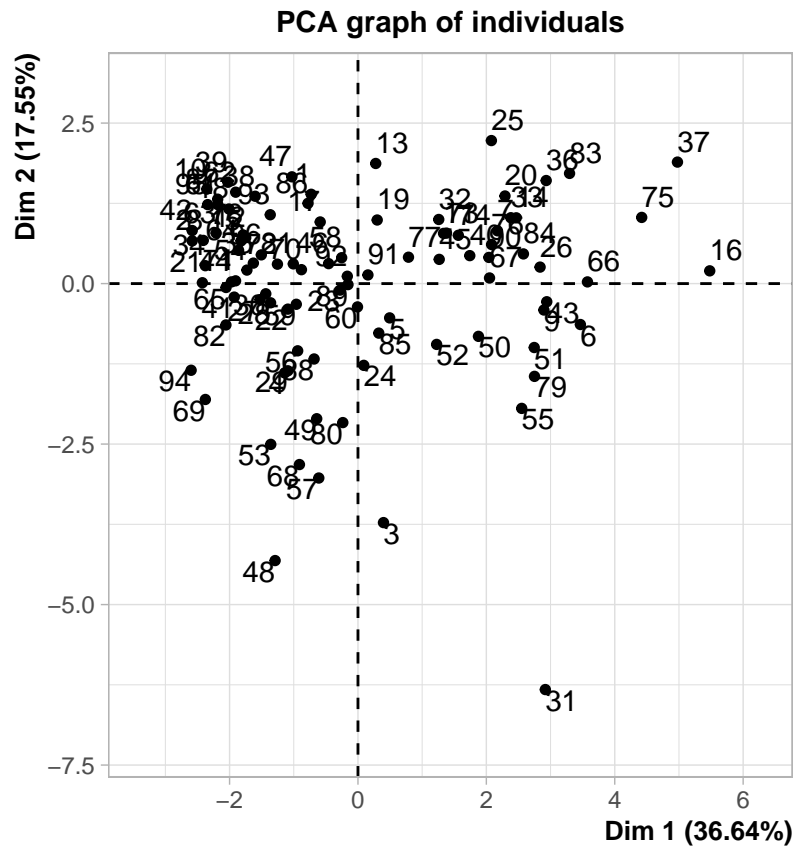
Las gráficas de dispersión de las dos primeras componentes principales, tanto para la Matriz S como para la Matriz R, muestran cómo se agrupan los países según las características económicas y sociales que se analizan. El componente 1 es el principal contribuyente a esta estructura, y el componente 2 la complementa. Un grupo de países en la esquina superior derecha de la gráfica de la Matriz R comparte características comunes, indicando una asociación positiva entre los indicadores representados por ambas componentes. En general, los países con valores altos en una componente tienden a tener valores altos en la otra, revelando correlaciones sistemáticas entre ciertos indicadores en los países analizados.

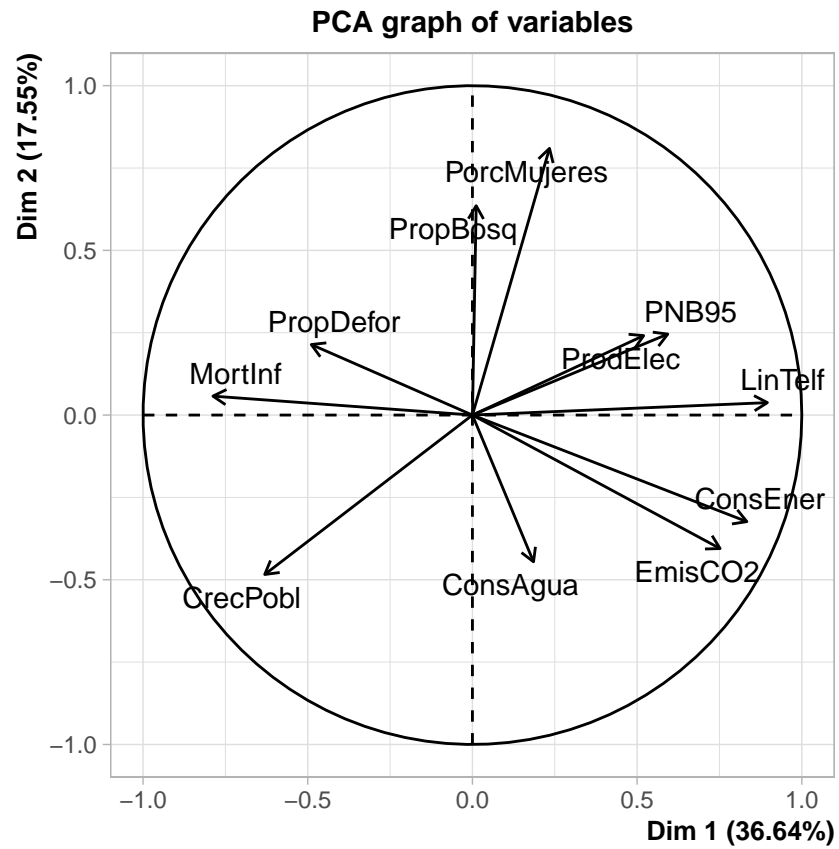
### Parte III

- Explore los siguientes gráficos relativos al problema y Componentes Principales y dé una interpretación de cada gráfico.

```
cp3 <- PCA(M)
```

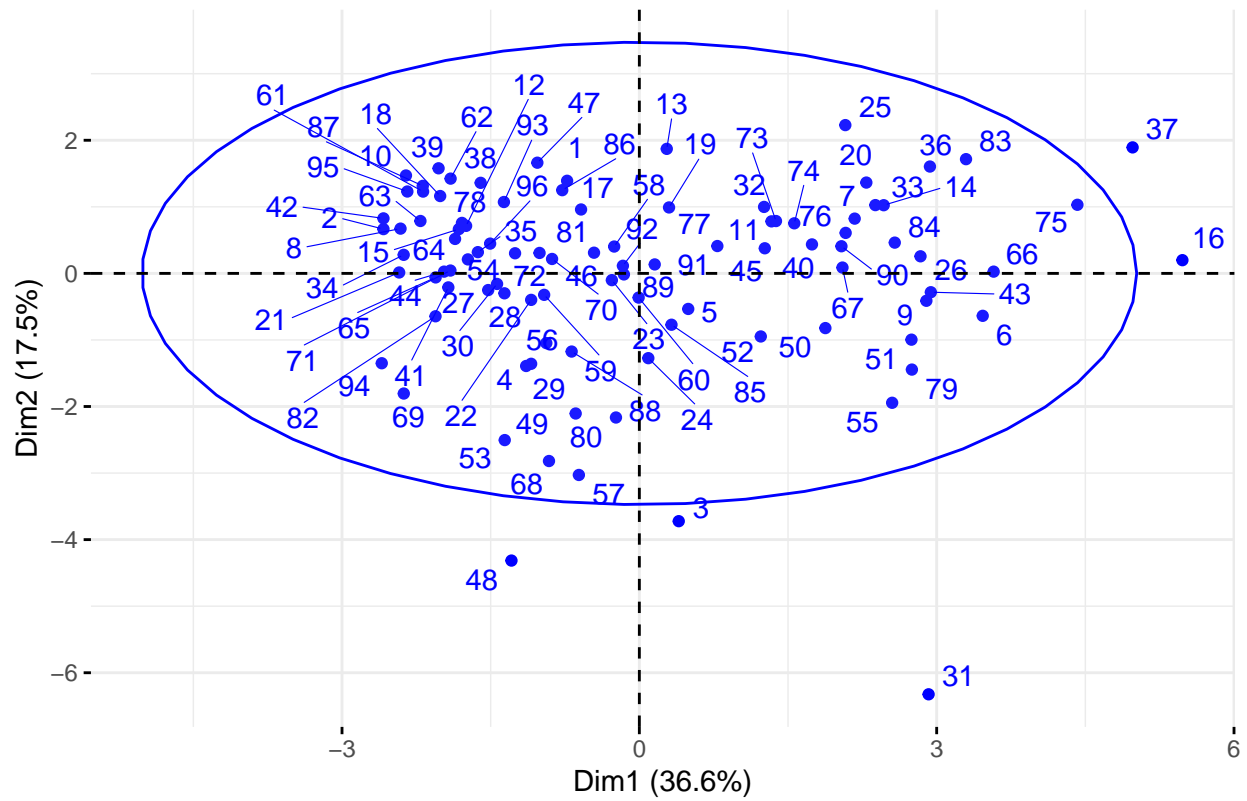






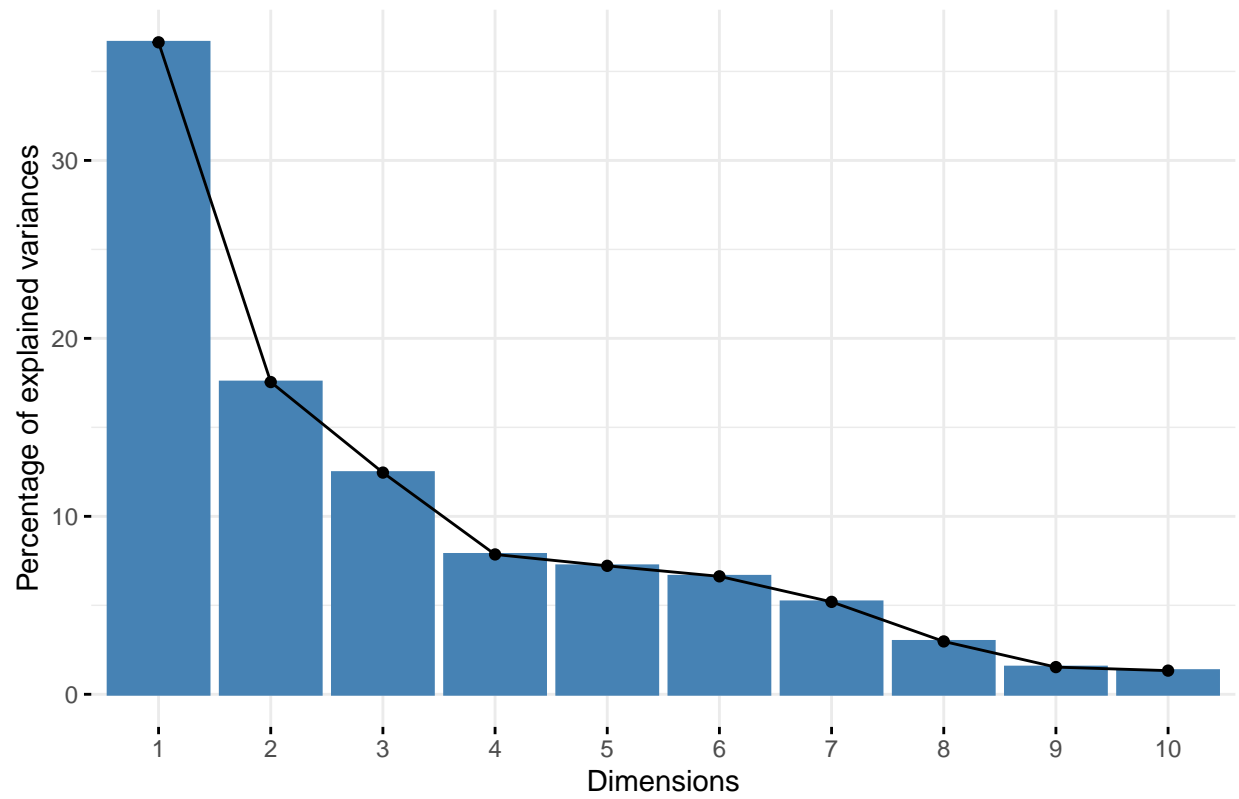
```
fviz_pca_ind(cp3, col.ind = "blue", addEllipses = TRUE, repel = TRUE)
```

## Individuals – PCA

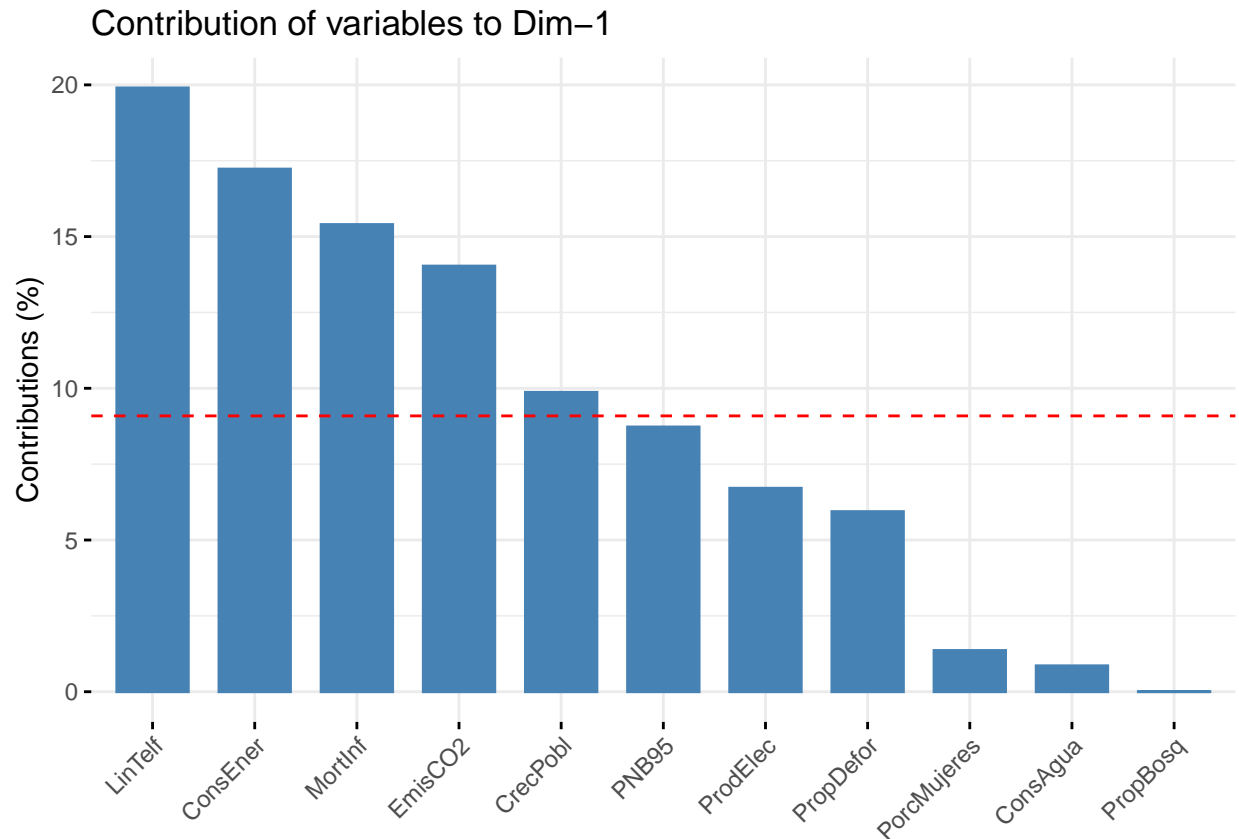


```
fviz_screepplot(cp3)
```

Scree plot



```
fviz_contrib(cp3, choice = c("var"))
```



- Gráfica de Individuos (PCA Individual Plot): Muestra la distribución de países según las dos primeras componentes principales. Los países cercanos entre sí tienen perfiles socioeconómicos similares, mientras que los dispersos presentan mayor variabilidad en sus indicadores.
- Gráfica de Scree Plot (Gráfica de Codo): Representa la varianza explicada por cada componente principal. La primera componente es la más relevante, seguida de la segunda, y su relevancia disminuye con las componentes subsiguientes.
- Gráfica de Contribuciones de Variables: Indica la influencia de cada variable en las dos primeras componentes principales. Las variables cercanas a las componentes tienen mayor contribución, mientras que las alejadas contribuyen menos. Esta gráfica destaca las variables más determinantes en la variabilidad de los datos de los países.