

# Actividad 1.6 Componentes principales

Franco Mendoza Muraira A01383399

2023-11-11

## Parte 1

```
# Cargar datos
datos <- read.csv("países_mundo.csv")

# Seleccionar las variables de interés
variables <- datos[, c("CrecPobl", "MortInf", "PorcMujeres", "PNB95", "ProdElec",
                      "LinTelf", "ConsAgua", "PropBosq", "PropDefor", "ConsEner", "EmisCO2")]
head(variables,5)
```

```
##   CrecPobl MortInf PorcMujeres PNB95 ProdElec LinTelf ConsAgua PropBosq
## 1      1.0     30         41   2199    3903     12      94      53
## 2      3.0    124         46  4422     955      6      57     19
## 3      4.3     21         13 133540   91019     96     497      1
## 4      2.5     34         24 44609   19883     42     180      2
## 5      1.3     22         31 278431   65962    160    1043     22
##   PropDefor ConsEner EmisCO2
## 1      0.0      341     1.2
## 2      0.7       89     0.5
## 3      0.0     4566    13.1
## 4      0.8      906     3.0
## 5      0.1     1504     3.5
```

## Matriz de varianzas-covarianzas

```
S <- cov(variables)
S
```

```
##           CrecPobl      MortInf  PorcMujeres      PNB95
## CrecPobl  1.538298e+00  2.195026e+01 -6.078026e+00 -8.933379e+04
## MortInf   2.195026e+01  1.032859e+03 -9.249342e+00 -2.269332e+06
## PorcMujeres -6.078026e+00 -9.249342e+00  7.698322e+01  2.813114e+05
## PNB95      -8.933379e+04 -2.269332e+06  2.813114e+05  4.999786e+10
## ProdElec   -4.973964e+04 -1.043435e+06  2.260248e+05  2.247791e+10
## LinTelf    -1.369079e+02 -4.381366e+03  4.499750e+02  2.039550e+07
## ConsAgua   -4.827092e+01 -1.288211e+03 -1.568313e+03  1.097481e+07
## PropBosq   -3.887018e+00 -1.466316e+01  6.517895e+01  2.474311e+05
```

```
## PropDefor      3.361974e-01  1.276296e+01  2.680592e-01 -5.806203e+04
## ConsEner      -8.384169e+02 -4.442568e+04  2.855207e+02  1.415628e+08
## EmisC02       -1.137877e+00 -9.485500e+01 -2.150132e+00  2.501673e+05
##               ProdElec      LinTelf      ConsAgua      PropBosq
## CrecPobl      -4.973964e+04 -1.369079e+02 -4.827092e+01   -3.887018
## MortInf       -1.043435e+06 -4.381366e+03 -1.288211e+03   -14.663158
## PorcMujeres   2.260248e+05  4.499750e+02 -1.568313e+03    65.178947
## PNB95         2.247791e+10  2.039550e+07  1.097481e+07  247431.122807
## ProdElec      1.821909e+10  7.583050e+06  1.399817e+07  70359.785965
## LinTelf       7.583050e+06  3.841247e+04  1.193110e+04   248.715789
## ConsAgua      1.399817e+07  1.193110e+04  3.301981e+05 -2220.757895
## PropBosq      7.035979e+04  2.487158e+02 -2.220758e+03   401.003509
## PropDefor     -3.180340e+04 -9.940461e+01 -6.743793e+01    2.625263
## ConsEner      6.801296e+07  3.426262e+05  2.092242e+05 -5153.438596
## EmisC02       1.392779e+05  6.385700e+02  4.869328e+02   -12.897193
##               PropDefor      ConsEner      EmisC02
## CrecPobl      3.361974e-01 -8.384169e+02   -1.137877
## MortInf       1.276296e+01 -4.442568e+04   -94.855000
## PorcMujeres   2.680592e-01  2.855207e+02   -2.150132
## PNB95         -5.806203e+04  1.415628e+08  250167.323509
## ProdElec     -3.180340e+04  6.801296e+07  139277.888640
## LinTelf       -9.940461e+01  3.426262e+05   638.570000
## ConsAgua      -6.743793e+01  2.092242e+05   486.932763
## PropBosq      2.625263e+00 -5.153439e+03   -12.897193
## PropDefor     1.817253e+00 -1.051522e+03    -2.632487
## ConsEner     -1.051522e+03  5.014395e+06  10286.159781
## EmisC02      -2.632487e+00  1.028616e+04    27.268614
```

## Análisis de componentes principales para la matriz de varianzas-covarianzas

### Valores y vectores propios

```
eigen_S <- eigen(S)
eigen_S
```

```
## 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
```

### Proporción de varianza explicada por cada componente

```
prop_var_S <- eigen_S$values / sum(diag(S))
prop_var_S
```

```
## [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
```

### Varianza acumulada

```
var_acum_S <- cumsum(prop_var_S)
var_acum_S
```

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

En este caso los componentes mas importantes son los primeros 2, ya que el primero explica el 90.3% de la varianza acumulada, y el segundo explica aproximadamente el 9.6%, en este caso el mas significativo es el primero ya que explica la gran mayoria de la varianza en los datos.

## Coeficientes de las combinaciones lineales de CP1 y CP2 para la matriz de varianzas-covarianzas

```
##
## Coeficientes de las combinaciones lineales de CP1 y CP2 para la matriz de varianzas-covarianzas:

## [1] -1.658168e-06 -4.048139e-05  5.739096e-06  8.880376e-01  4.597636e-01
## [6]  3.504341e-04  2.625508e-04  4.089564e-06 -1.073825e-06  2.547156e-03
## [11]  4.643724e-06

## [1]  4.706785e-07 -1.774254e-05 -1.084543e-05  4.597632e-01 -8.880405e-01
## [6]  4.016179e-04 -1.122118e-03  7.790843e-06  2.350808e-07  7.126782e-04
## [11] -1.315731e-06
```

En este caso, la ecuación de la combinación lineal de los Componentes Principales (CP1 y CP2) para la matriz de varianzas-covarianzas sería:

$$CP1 = -1.6582 \times 10^{-6} X1 - 4.0481 \times 10^{-5} X2 + 5.7391 \times 10^{-6} X3 + 8.8804 X4 + 0.4598 X5 + 0.0004 X6 + 0.0003 X7 + 4.0896 \times 10^{-6} X8$$

$$CP2 = 4.7068 \times 10^{-7} X1 - 1.7743 \times 10^{-5} X2 - 1.0845 \times 10^{-5} X3 + 0.4598 X4 - 0.8880 X5 + 0.0004 X6 - 0.0011 X7 + 7.7908 \times 10^{-6} X8 +$$

Para CP1: - La variable  $X4$  (PNB95) tiene el mayor coeficiente positivo (8.8804), por lo que contribuye significativamente al aumento de CP1. - La variable  $X5$  (ProdElec) también tiene un coeficiente positivo alto (0.4598). - Otras variables como  $X10$  (ConsEner) y  $X6$  (LinTelf) también contribuyen positivamente, pero menos que las otras.

Para CP2: - La variable  $X5$  (ProdElec) tiene el mayor coeficiente negativo (-0.8880), por lo que contribuye significativamente a la disminución de CP2. - La variable  $X4$  (PNB95) también tiene un coeficiente alto (0.4598). - Otras variables como  $X7$  (ConsAgua) y  $X10$  (ConsEner) también contribuyen negativamente, pero menos que las otras.

## Matriz de correlaciones

```
R <- cor(variables)
R
```

	CrecPobl	MortInf	PorcMujeres	PNB95	ProdElec
CrecPobl	1.00000000	0.55067948	-0.55852711	-0.32212154	-0.29711119
MortInf	0.55067948	1.00000000	-0.03280139	-0.31579250	-0.24053689
PorcMujeres	-0.55852711	-0.03280139	1.00000000	0.14338826	0.19085114
PNB95	-0.32212154	-0.31579250	0.14338826	1.00000000	0.74476081
ProdElec	-0.29711119	-0.24053689	0.19085114	0.74476081	1.00000000
LinTelf	-0.56321228	-0.69558922	0.26167018	0.46539599	0.28664508
ConsAgua	-0.06772953	-0.06975563	-0.31106243	0.08541500	0.18047653
PropBosq	-0.15650281	-0.02278415	0.37096694	0.05525919	0.02603078
PropDefor	0.20107881	0.29459348	0.02266339	-0.19262327	-0.17478434
ConsEner	-0.30187731	-0.61731132	0.01453216	0.28272492	0.22501894
EmisCO2	-0.17568860	-0.56520778	-0.04692837	0.21425123	0.19760017

```
##          LinTelf      ConsAgua      PropBosq      PropDefor      ConsEner
## CrecPobl    -0.56321228 -0.06772953 -0.15650281  0.20107881 -0.30187731
## MortInf     -0.69558922 -0.06975563 -0.02278415  0.29459348 -0.61731132
## PorcMujeres 0.26167018 -0.31106243  0.37096694  0.02266339  0.01453216
## PNB95        0.46539599  0.08541500  0.05525919 -0.19262327  0.28272492
## ProdElec     0.28664508  0.18047653  0.02603078 -0.17478434  0.22501894
## LinTelf      1.00000000  0.10593934  0.06337138 -0.37623801  0.78068385
## ConsAgua     0.10593934  1.00000000 -0.19299225 -0.08705811  0.16259804
## PropBosq     0.06337138 -0.19299225  1.00000000  0.09725032 -0.11492480
## PropDefor    -0.37623801 -0.08705811  0.09725032  1.00000000 -0.34833836
## ConsEner     0.78068385  0.16259804 -0.11492480 -0.34833836  1.00000000
## EmisC02      0.62393719  0.16227447 -0.12333592 -0.37396154  0.87965517
##          EmisC02
## CrecPobl    -0.17568860
## MortInf     -0.56520778
## PorcMujeres -0.04692837
## PNB95        0.21425123
## ProdElec     0.19760017
## LinTelf      0.62393719
## ConsAgua     0.16227447
## PropBosq     -0.12333592
## PropDefor    -0.37396154
## ConsEner     0.87965517
## EmisC02      1.00000000
```

## Análisis de componentes principales para la matriz de correlaciones

### Valores y vectores propios

```
eigen_R <- eigen(R)
eigen_R
```

```
## 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
```

### Proporción de varianza explicada por cada componente

```
prop_var_R <- eigen_R$values / sum(diag(R))
prop_var_R
```

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

### Varianza acumulada

```
var_acum_R <- cumsum(prop_var_R)
var_acum_R
```

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

Con estos resultados obtenidos, podemos ver que los primeros 2 componentes son los mas significativos, esto porque explican una proporción considerable de la varianza acumulada, ya que acumulan el 54.18% de la varianza, con porcentajes de 36.64 y 17.75 respectivamente.

### Coefficientes de las combinaciones lineales de CP1 y CP2 para la matriz de correlaciones

```
##
## Coeficientes de las combinaciones lineales de CP1 y CP2 para la matriz de correlaciones:

## [1] -0.314119414 -0.392395442 0.116546319 0.295393771 0.258964724
## [6] 0.446082934 0.092410503 0.005692925 -0.243652293 0.415029554
## [11] 0.374531032

## [1] 0.34835747 -0.04136238 -0.58283641 -0.17690839 -0.17356372 -0.02719077
## [7] 0.32060987 -0.45742697 -0.15408201 0.23286257 0.29168698
```

La ecuación de la combinación lineal de los Componentes Principales (CP1 y CP2) se puede expresar así:

$$CP1 = -0.3141X1 - 0.3924X2 + 0.1165X3 + 0.2954X4 + 0.2590X5 + 0.4461X6 + 0.0924X7 + 0.0057X8 - 0.2437X9 + 0.4150X10 +$$

$$CP2 = 0.3484X1 - 0.0414X2 - 0.5828X3 - 0.1769X4 - 0.1736X5 - 0.0272X6 + 0.3206X7 - 0.4574X8 - 0.1541X9 + 0.2329X10 + 0.0000X11$$

Para CP1: - La variable  $X6$  (LinTelf) tiene el mayor coeficiente positivo (0.4461), por lo que contribuye significativamente al aumento de CP1. - La variable  $X4$  (PNB95) también tiene un coeficiente positivo alto (0.2954). - Otras variables como  $X5$  (ProdElec) y  $X10$  (ConsEner) también contribuyen positivamente, pero menos que las otras.

Para CP2: - La variable  $X3$  (PorcMujeres) tiene el mayor coeficiente negativo (-0.5828), por lo que contribuye significativamente a la disminución de CP2. - La variable  $X8$  (PropBosq) también tiene un alto coeficiente negativo (-0.4574). - Otras variables como  $X1$  (CrecPobl) y  $X7$  (ConsAgua) también contribuyen negativamente, pero menos que las otras.

## Parte 2

**Obtener las dos primeras componentes principales con la matriz de varianzas-covarianzas**

```
cpS <- princomp(variables, cor = FALSE)
cpS
```

```
## Call:
## princomp(x = variables, cor = FALSE)
##
## Standard deviations:
##      Comp.1      Comp.2      Comp.3      Comp.4      Comp.5      Comp.6
## 2.469691e+05 8.070349e+04 2.141953e+03 5.545146e+02 1.096972e+02 2.254828e+01
##      Comp.7      Comp.8      Comp.9      Comp.10      Comp.11
## 1.894754e+01 6.704303e+00 2.395922e+00 1.192913e+00 6.869072e-01
##
## 11 variables and 96 observations.
```

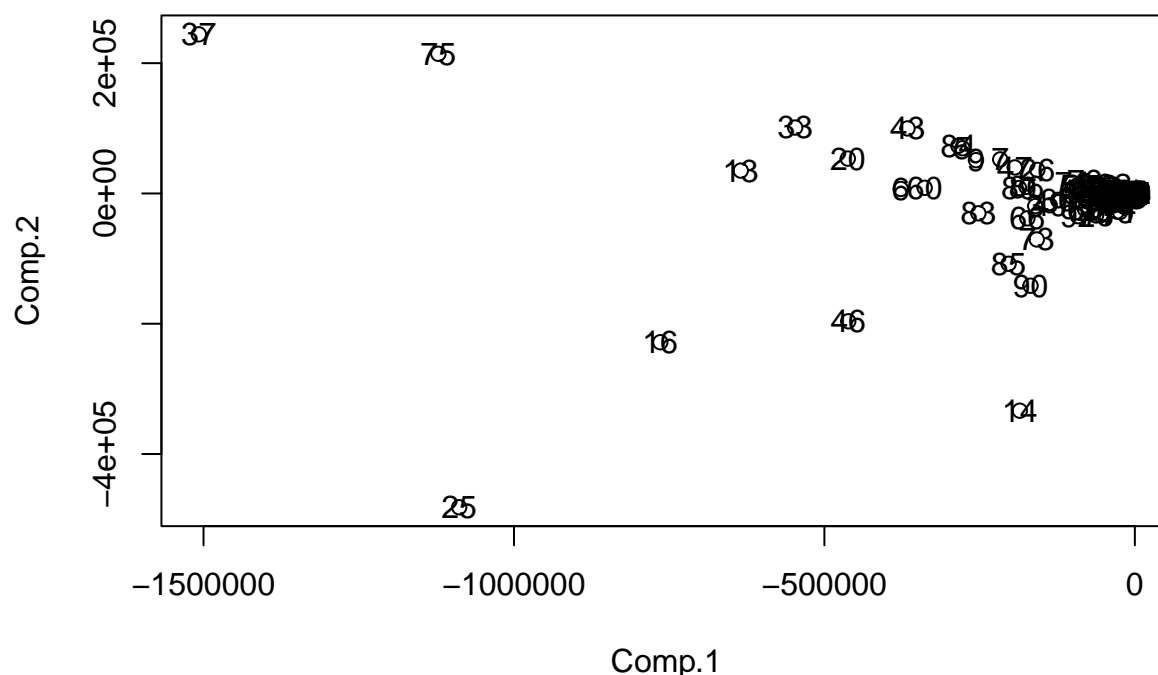
**Calcular las puntuaciones (scores) de las observaciones para los componentes obtenidos con la matriz de varianzas-covarianzas**

```
cpaS <- as.matrix(variables) %*% cpS$loadings
```

**Graficar las dos primeras componentes con la matriz de varianzas-covarianzas**

```
plot(cpaS[, 1:2], type = "p", main = "Gráfica de Componentes Principales (S)")
text(cpaS[, 1], cpaS[, 2], 1:nrow(cpaS))
```

## Gráfica de Componentes Principales (S)



```
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
```

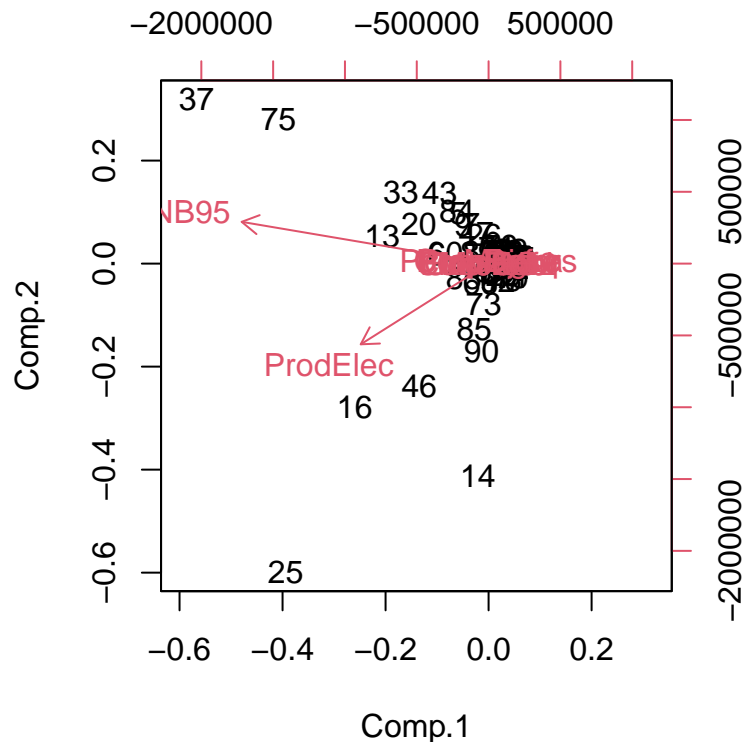
```
## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length =  
## arrow.len): zero-length arrow is of indeterminate angle and so skipped
```

```
## 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
```





En la primera gráfica de componentes principales, se puede observar que los datos en general se encuentran mayormente en un rango de valores cercano a 0 en ambos componentes. En el primer componente, se aprecia que los valores son principalmente negativos y ceros, llegando a valores muy grandes. A diferencia del primer componente, en el segundo se puede ver que también alcanza valores muy grandes, pero en este caso, puede tener valores positivos.

En el segundo gráfico se aprecia que las variables que más afectan a los componentes son ProdElec y PNB95, siendo el primero influyente en ambos de manera similar, mientras que el segundo afecta más al componente 1. También se pueden identificar algunos datos atípicos que se separan del resto, como las observaciones 37, 75 y 25. Además, se puede notar que las observaciones 16 y 14 se distancian significativamente, aunque no tanto como las mencionadas anteriormente.

### Resumen de las componentes principales obtenidas con la matriz de varianzas-covarianzas

```
summary(cpS)
```

```
## Importance of components:
##               Comp.1      Comp.2      Comp.3      Comp.4
## Standard deviation  2.469691e+05 8.070349e+04 2.141953e+03 5.545146e+02
## Proportion of Variance 9.034543e-01 9.647298e-02 6.795804e-05 4.554567e-06
## Cumulative Proportion 9.034543e-01 9.999273e-01 9.999953e-01 9.999998e-01
##               Comp.5      Comp.6      Comp.7      Comp.8
## Standard deviation  1.096972e+02 2.254828e+01 1.894754e+01 6.704303e+00
## Proportion of Variance 1.782429e-07 7.530916e-09 5.317738e-09 6.657763e-10
## Cumulative Proportion 1.000000e+00 1.000000e+00 1.000000e+00 1.000000e+00
```

```
##                               Comp.9      Comp.10      Comp.11
## Standard deviation      2.395922e+00 1.192913e+00 6.869072e-01
## Proportion of Variance 8.502887e-11 2.107843e-11 6.989035e-12
## Cumulative Proportion 1.000000e+00 1.000000e+00 1.000000e+00
```

Calcular las puntuaciones (scores) de las observaciones para los componentes obtenidos con la matriz de correlaciones

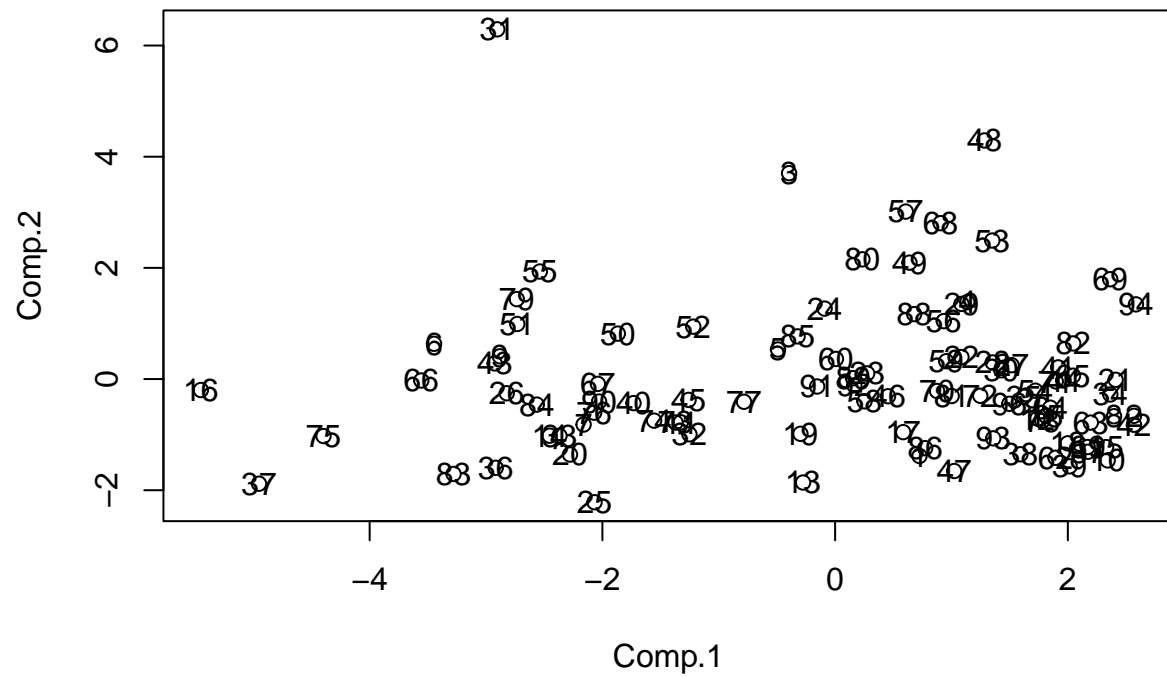
```
cpR <- princomp(variables, cor = TRUE)
cpaR <- as.matrix(scale(variables)) %*% cpR$loadings
cpR
```

```
## Call:
## princomp(x = variables, cor = TRUE)
##
## Standard deviations:
##      Comp.1      Comp.2      Comp.3      Comp.4      Comp.5      Comp.6      Comp.7      Comp.8
## 2.0074559 1.3892415 1.1706456 0.9297935 0.8911456 0.8539321 0.7558473 0.5716651
##      Comp.9      Comp.10      Comp.11
## 0.4099615 0.3825287 0.2633603
##
## 11 variables and 96 observations.
```

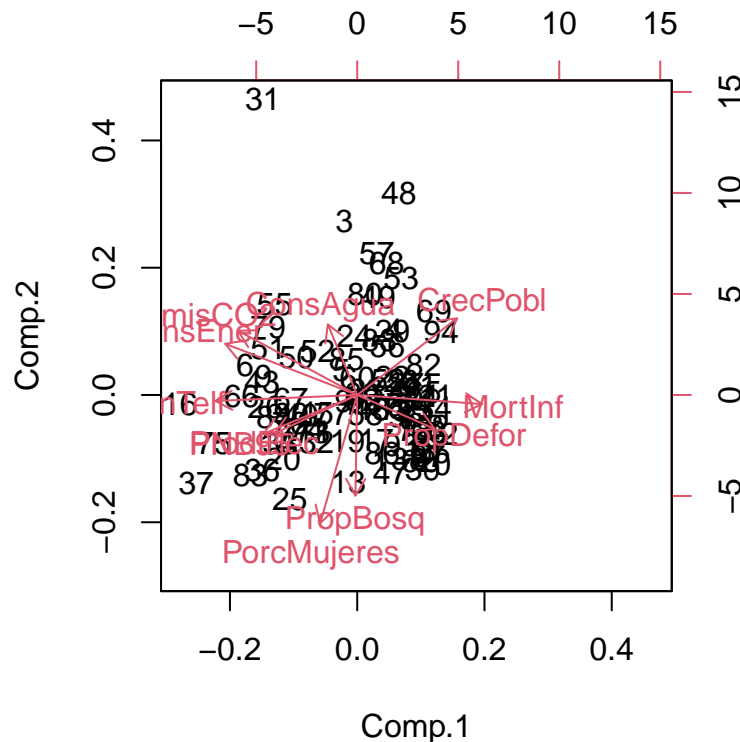
Graficar las dos primeras componentes con la matriz de correlaciones

```
plot(cpaR[, 1:2], type = "p", main = "Gráfica de Componentes Principales (R)")
text(cpaR[, 1], cpaR[, 2], 1:nrow(cpaR))
```

## Gráfica de Componentes Principales (R)



`biplot(cpR)`



En el análisis de la matriz de correlación, podemos ver una gran variabilidad en los datos comparado con la matriz de varianzas, en la primera gráfica de componentes principales, aunque se puede ver una tendencia a los datos en estar positivo y cercano a el valor de 2 en el primer componente, y negativo o cercano a 0 en el segundo componente.

En la segunda gráfica podemos ver que todas las variables tienen una influencia bastante parecida en los 2 componentes, aunque se ve un poco mayor la variable de MortInf al primer componente, y la variable de PorcMujeres en el segundo componente. Los datos atípicos que se pueden ver a simple vista en este gráfico son el 31, 3 y 48 ya que son muy altos en el segundo componente comparado con los otros datos.

## Resumen de las componentes principales obtenidas con la matriz de correlaciones

```
summary(cpr)
```

```
## Importance of components:
##               Comp.1   Comp.2   Comp.3   Comp.4   Comp.5
## Standard deviation  2.0074559 1.3892415 1.1706456 0.92979351 0.8911456
## Proportion of Variance 0.3663526 0.1754538 0.1245828 0.07859236 0.0721946
## Cumulative Proportion 0.3663526 0.5418065 0.6663893 0.74498164 0.8171762
##               Comp.6   Comp.7   Comp.8   Comp.9   Comp.10
## Standard deviation  0.85393206 0.75584728 0.57166507 0.40996153 0.38252868
## Proportion of Variance 0.06629091 0.05193683 0.02970918 0.01527895 0.01330256
## Cumulative Proportion 0.88346715 0.93540398 0.96511315 0.98039210 0.99369467
##               Comp.11
## Standard deviation  0.263360314
```

```
## Proportion of Variance 0.006305332
## Cumulative Proportion 1.000000000
```

## Parte 3

### Matriz de covarianzas

```
library(FactoMineR)
```

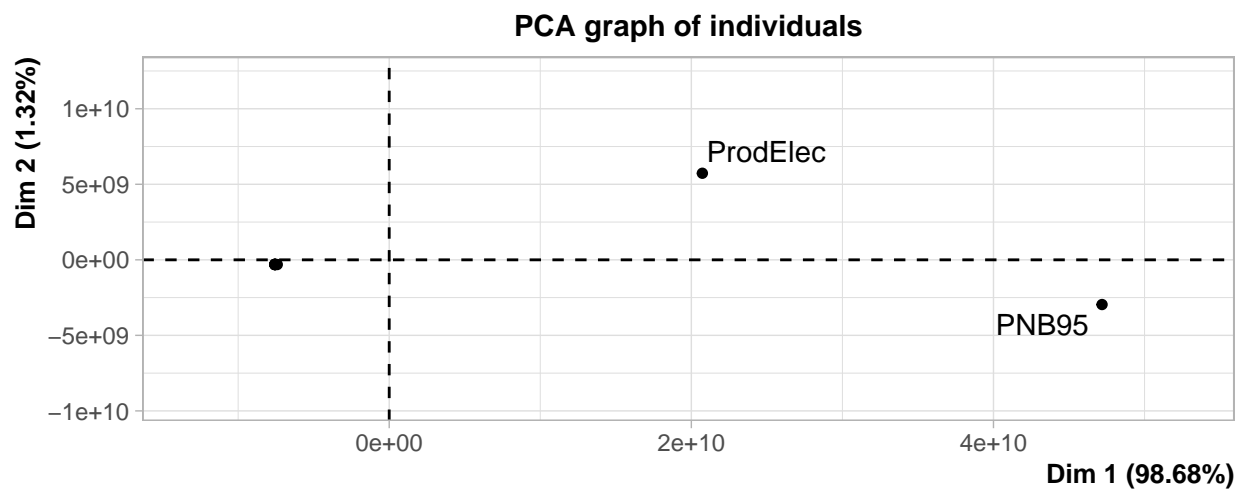
```
## Warning: package 'FactoMineR' was built under R version 4.2.3
```

```
library(ggplot2)
```

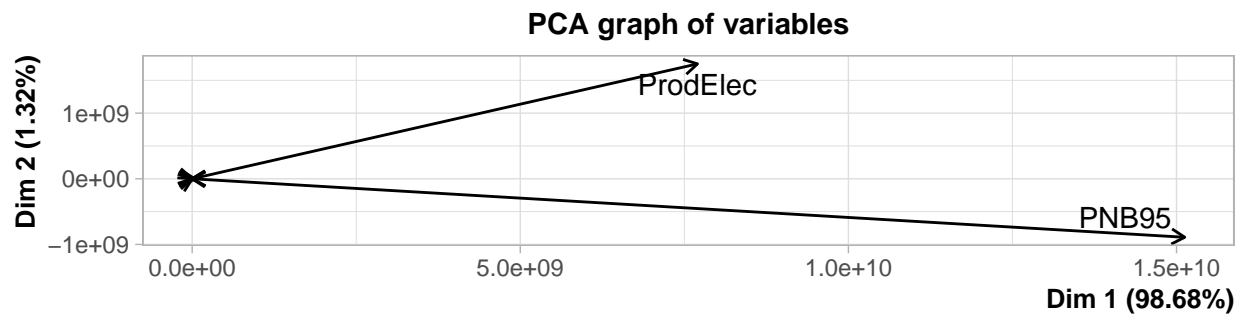
```
## Warning: package 'ggplot2' was built under R version 4.2.3
```

```
cpS = PCA(S,scale.unit=FALSE)
```

```
## Warning: ggrepel: 9 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



```
## Warning: ggrepel: 9 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```

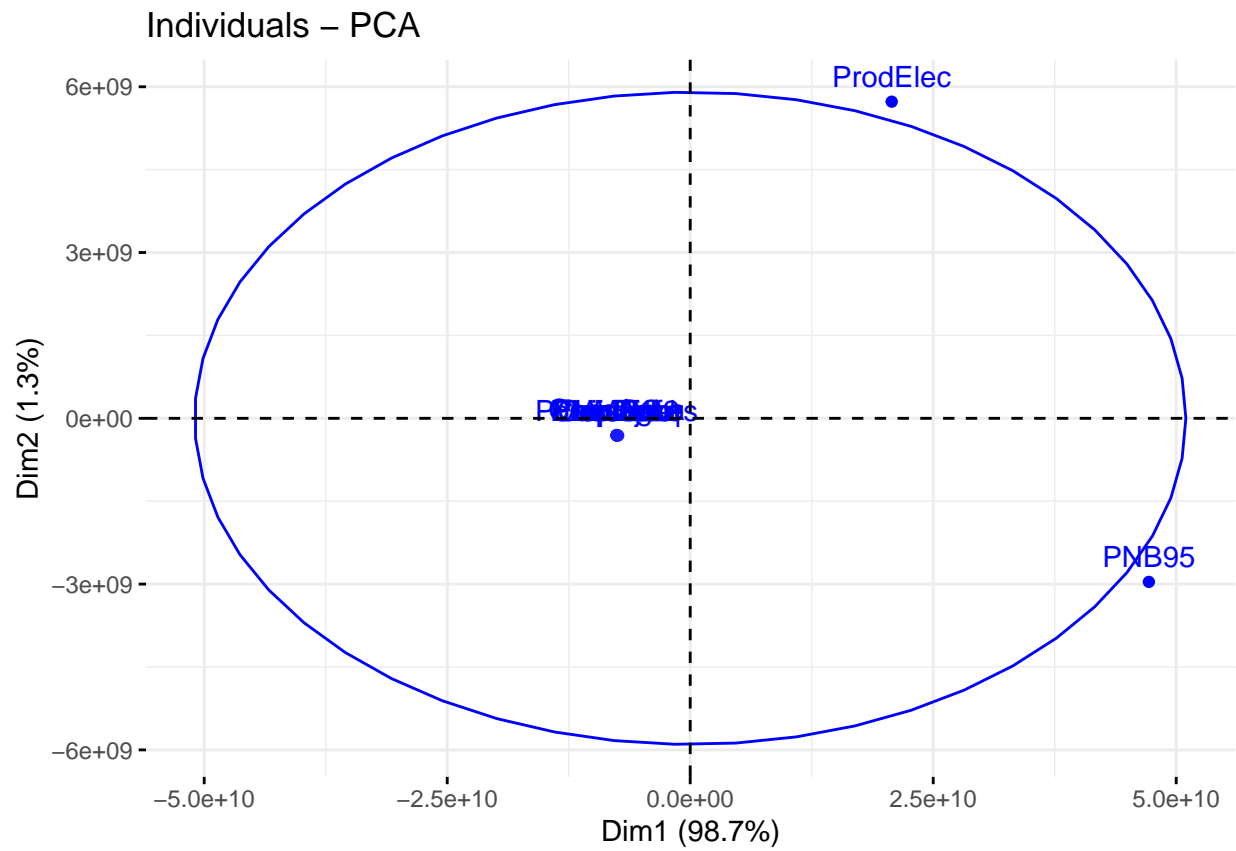


```
library(factoextra)
```

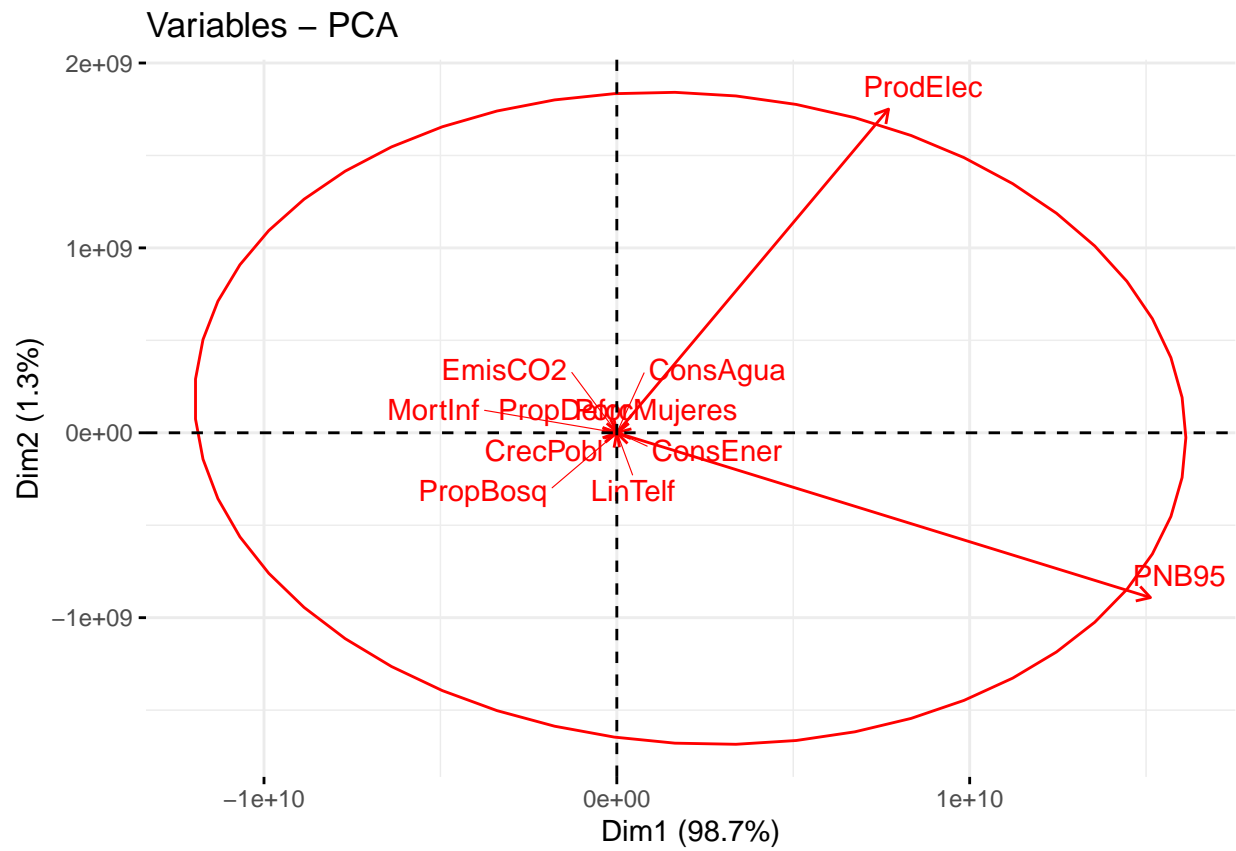
```
## Warning: package 'factoextra' was built under R version 4.2.3
```

```
## 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 = FALSE)
```

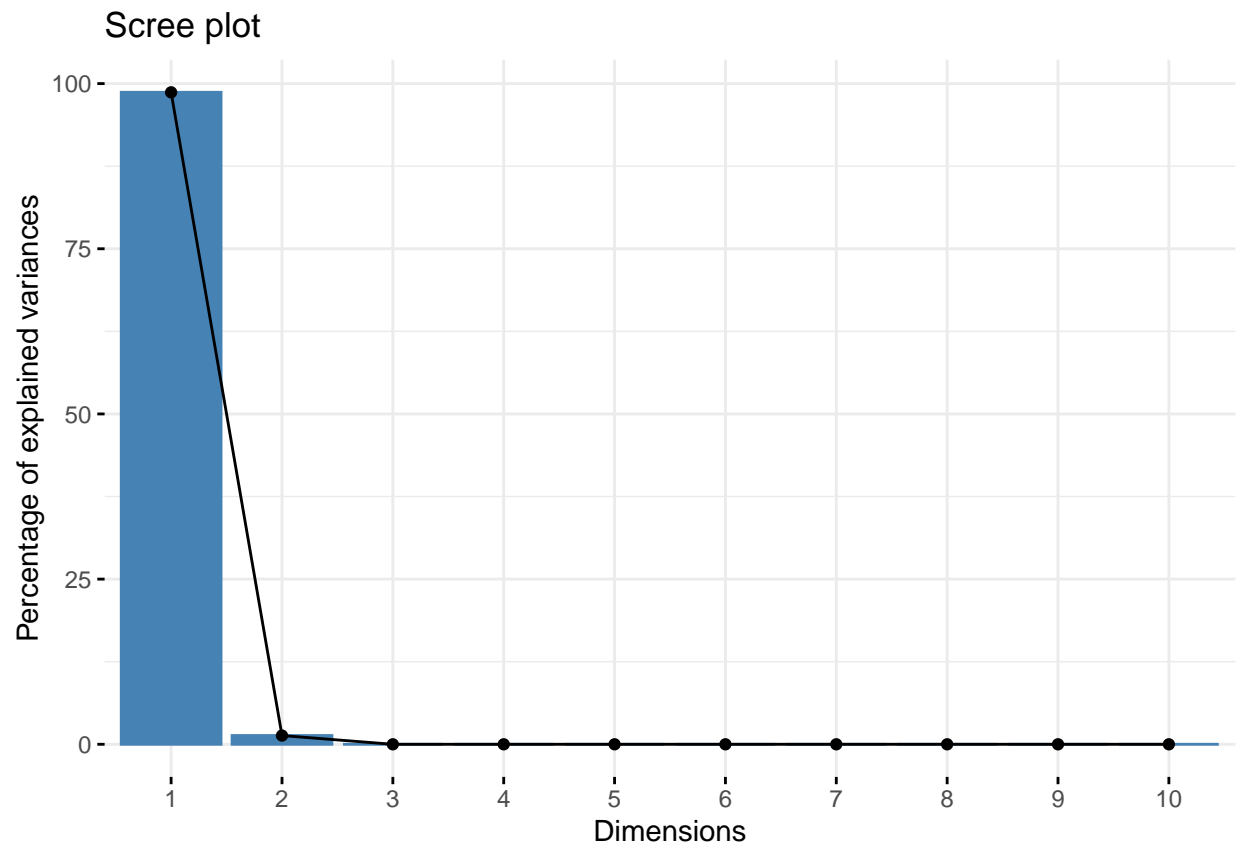


```
fviz_pca_var(cpS, col.var = "red", addEllipses = TRUE, repel = TRUE)
```

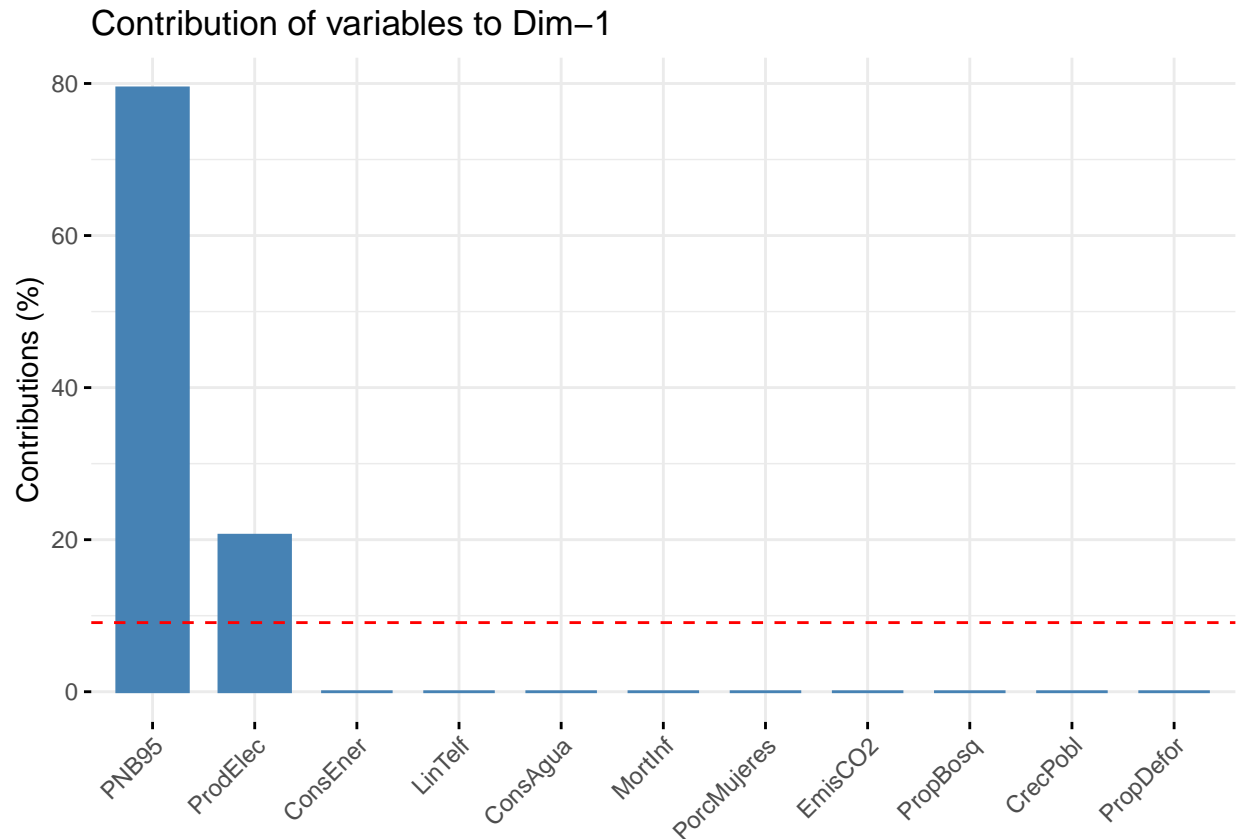


```
fviz_screepplot(cpS)
```





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

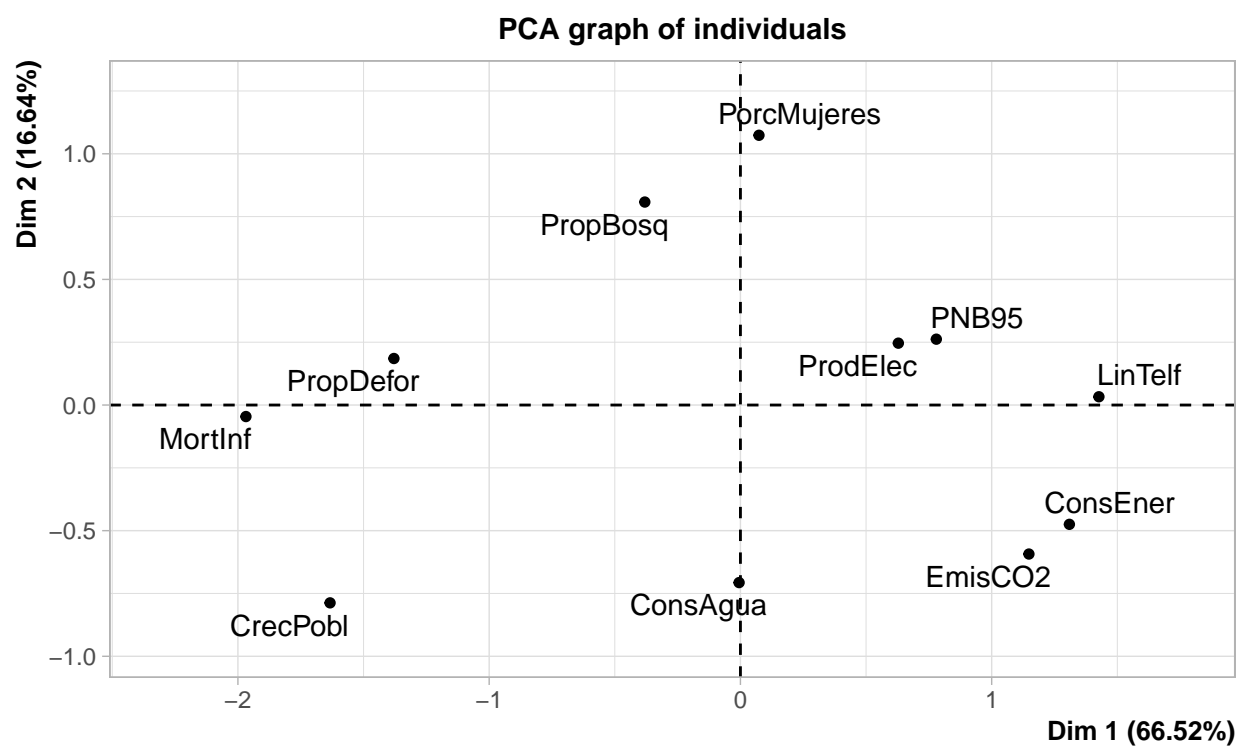


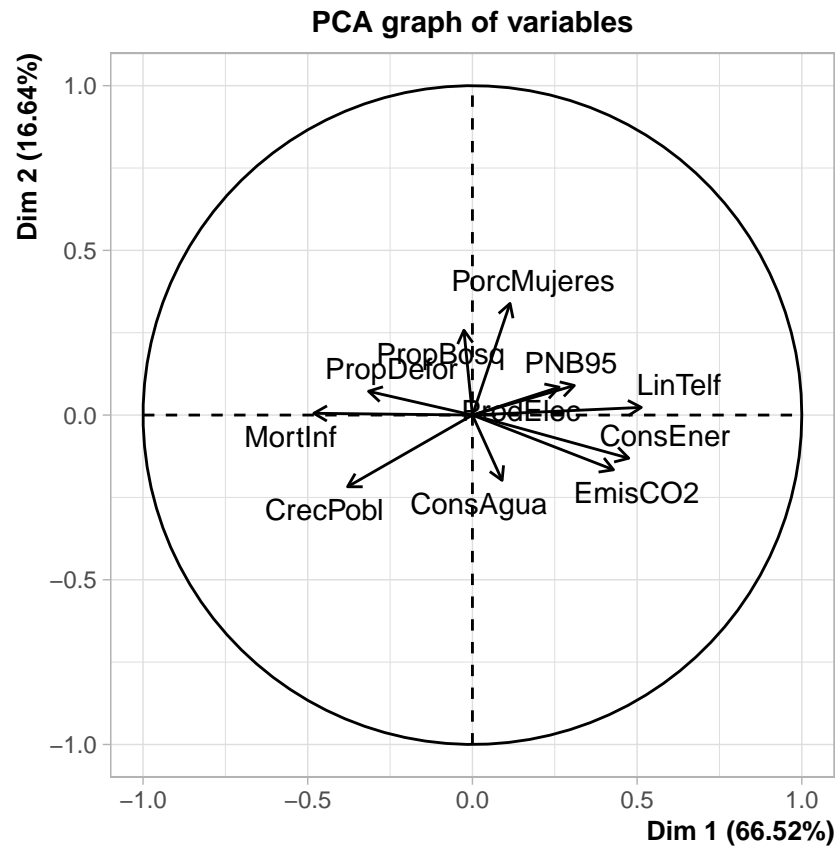
En las gráficas de variables, se puede ver claramente como la variable de PNB95 es la que mas contribuye por mucho porcentaje, con uno de casi 80%, seguido de la de ProdElec, con un poco mas de 20%, las otras variables contribuyendo muy poco a las dimensiones del modelo.

En el scree plot, y en los ejes de las primeras 4 gráficas también se puede ver que la dimensión 1 explica la varianza por un porcentaje muy grande, con un porcentaje de 98.7% comparandolo con el 1.3% de la segunda dimensión, que juntas explican toda la varianza

## Matriz de correlaciones

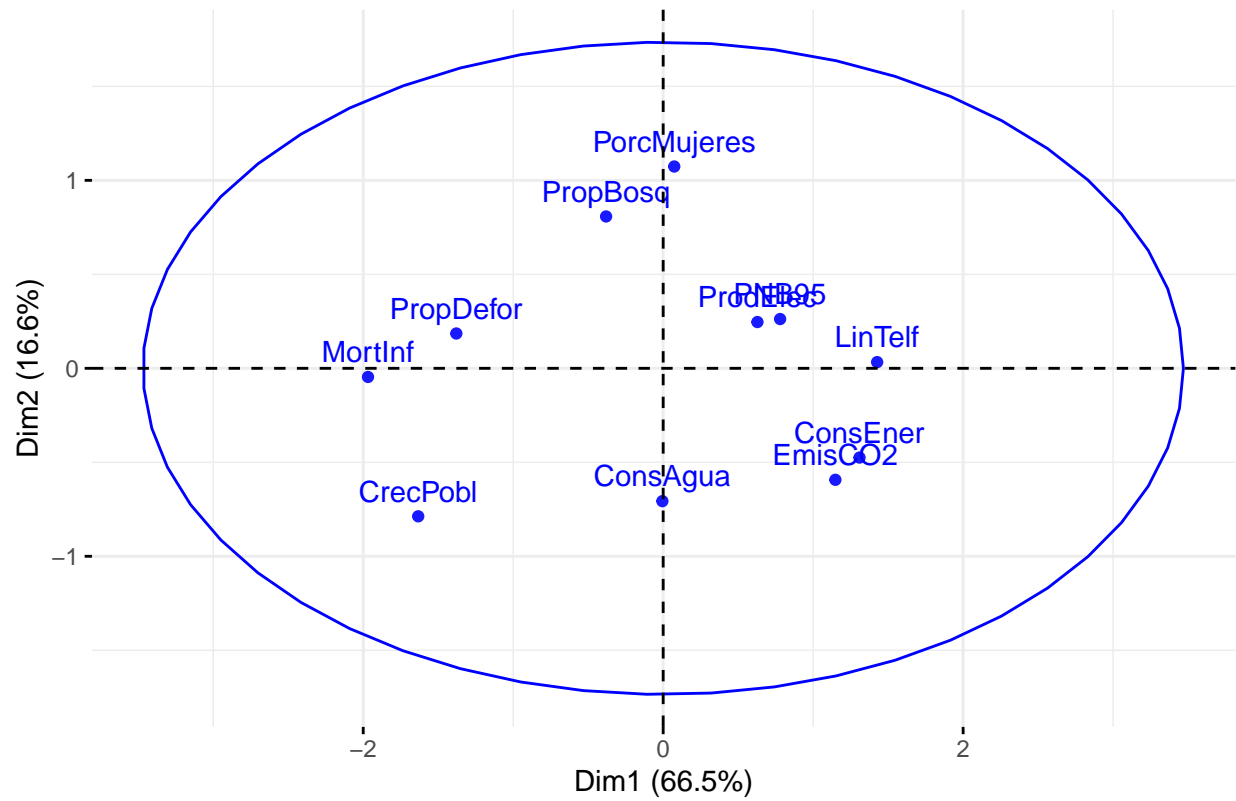
```
cpR = PCA(R,scale.unit=FALSE)
```





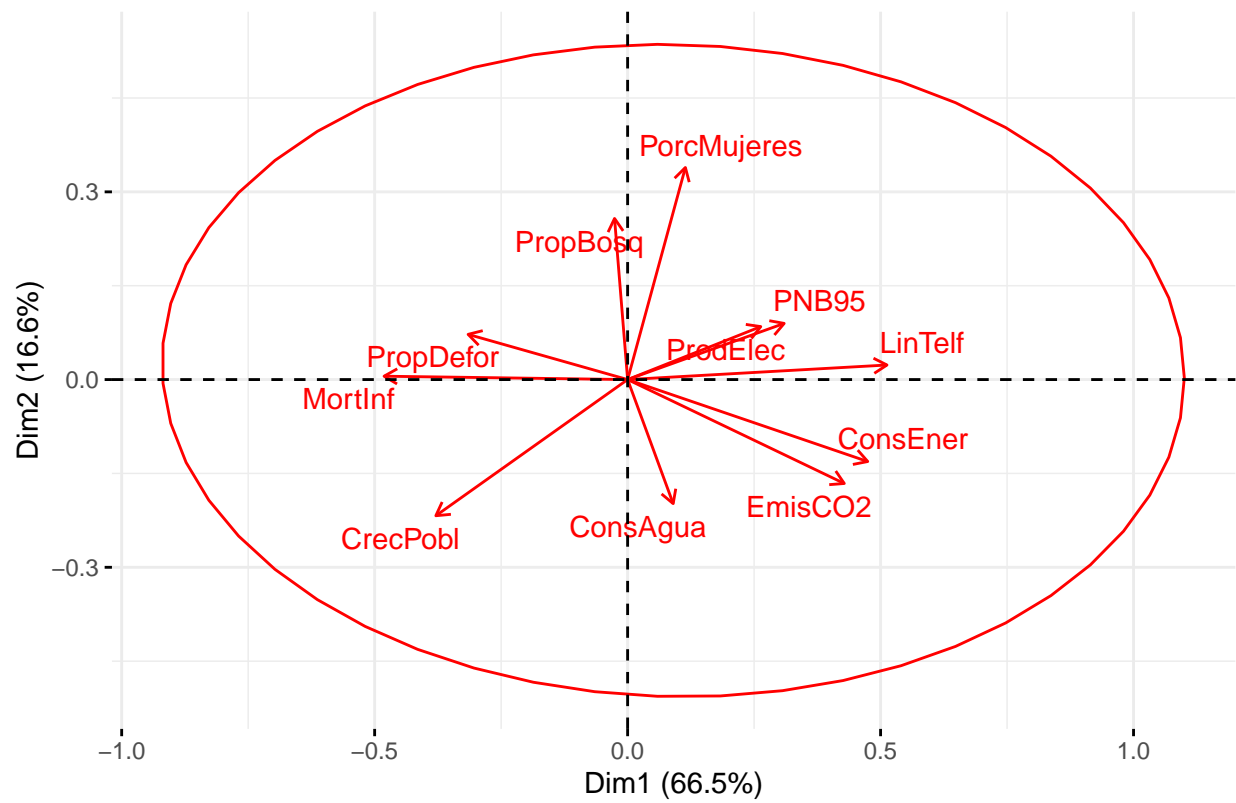
```
library(factoextra)
fviz_pca_ind(cpR, col.ind = "blue", addEllipses = TRUE, repel = FALSE)
```

## Individuals – PCA



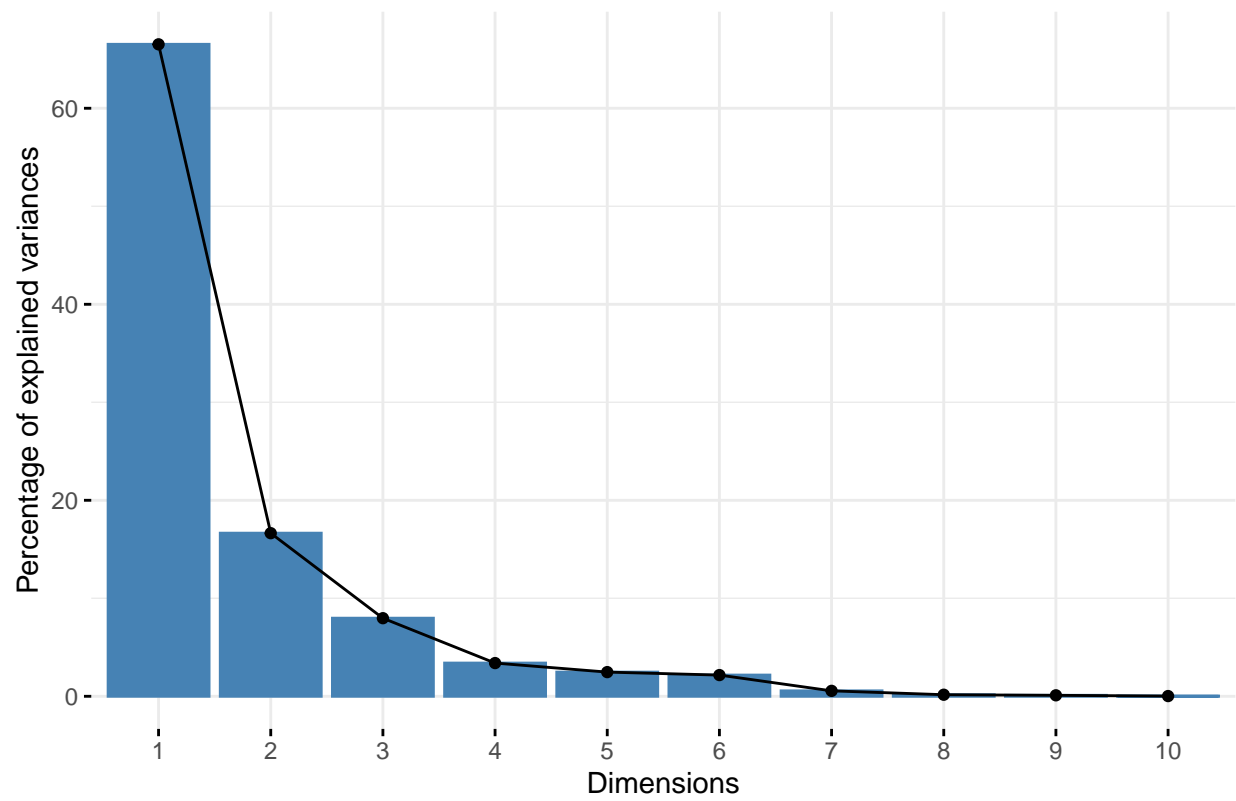
```
fviz_pca_var(cpR, col.var = "red", addEllipses = TRUE, repel = TRUE)
```

Variables – PCA

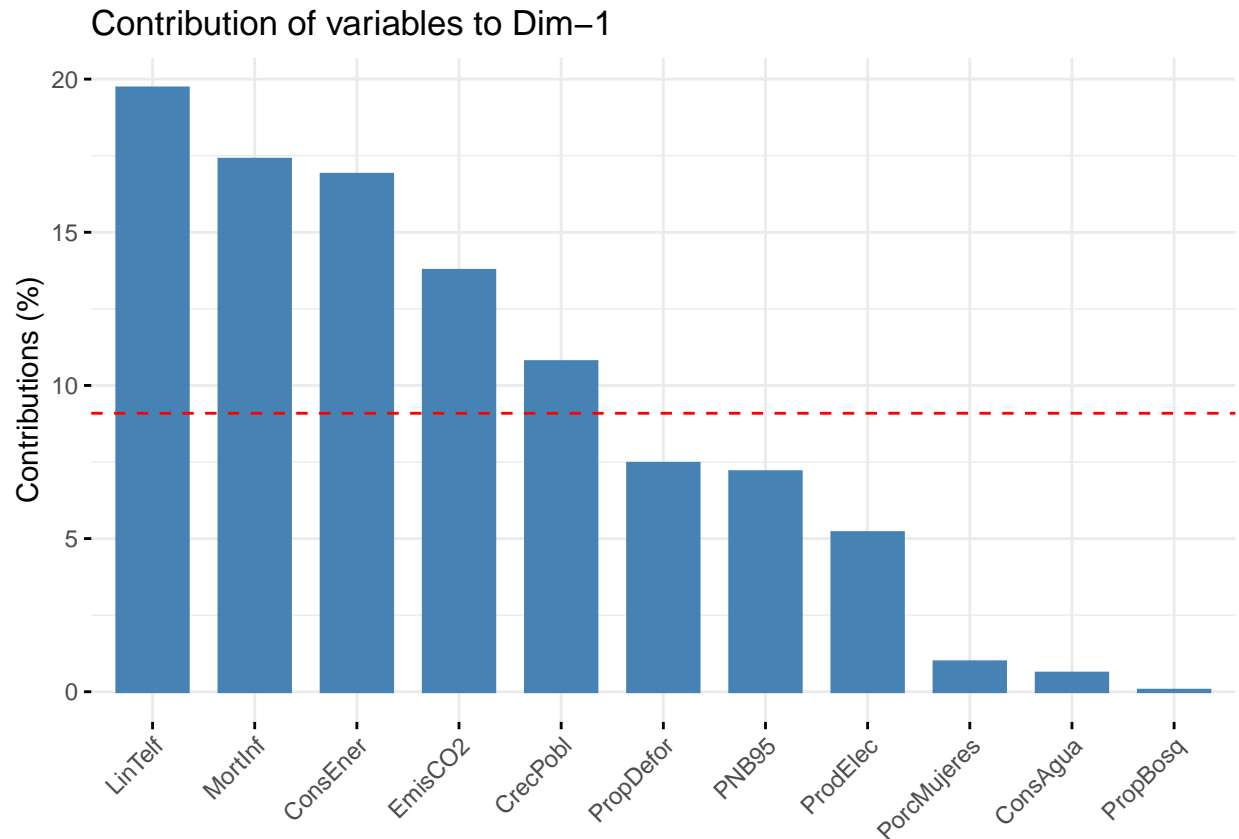


```
fviz_screepplot(cpR)
```

Scree plot



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



En las gráficas de la matriz de correlación se puede ver como las variables en general son mucho más igualadas en cuanto a significancia a las dimensiones del modelo, hay mas variabilidad entre ellas.

También en el scree plot se puede ver como se necesitarían más dimensiones para llegar a explicar un mejor porcentaje de la varianza, ya que la combinación de solo las primeras 2, explica el 82% de la varianza. Por último, en la última gráfica podemos ver las contribuciones de las variables a la dimensión que más explica, y se observa lo que se mencionó previamente, de como contribuyen más variables de manera similar a este modelo.

## Parte 4

Comparando los resultados de las 2 matrices, podemos ver que la matriz de varianza explica mejor la varianza de los datos usando solo 2 componentes, con una explicación de prácticamente el 100%, y se enseña que usa más que nada solo 2 de las variables de los datos, cuando la matriz de correlación ocuparía muchas más dimensiones para explicar lo que explica la de varianza en 2 dimensiones.

En las gráficas se puede ver que las variables de PNB95 y ProdElec fueron las más destacadas al hacer el modelo, y las que en el análisis de varianza servirían más como datos indicadores en el dataset de los 96 países. Las demás variables son mucho más cercanas entre sí, pero esto debido a que son las que menos explican las dimensiones.

En resumen, estas 2 variables de PNB95 y ProdElec se puede concluir que son las que más contribuirían a un estudio como indicadores económicos y sociales en los países que fueron parte de estos datos. En el modelo de componentes principales tuvieron altos coeficientes para influir en él.