

Actividad 1.6 Componentes principales

Raúl Correa Ocañas

2024-08-18

```
# install.packages("FactoMineR")
```

IMPORTACION DE DATOS

```
data_path = "../data/paises_mundo.csv"  
df = read.csv(data_path)  
scaled_df = scale(df)
```

PARTE I

1. Análisis de Matriz de Covarianza

```
cov_df = cov(df)  
eigen_cov_df = eigen(cov_df)
```

```
cov_df
```

##	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
## EmisCO2	-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
## EmisCO2	1.392779e+05	6.385700e+02	4.869328e+02	-12.897193
##	PropDefor	ConsEner	EmisCO2	
## 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	
## EmisCO2	-2.632487e+00	1.028616e+04	27.268614	

```
eigen_cov_df$values / sum(eigen_cov_df$values)
```

```
## [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
```

```
sum(eigen_cov_df$values)
```

```
## [1] 68222335253
```

```
sum(diag(cov_df))
```

```
## [1] 68222335253
```

```
cumsum(eigen_cov_df$values / sum(eigen_cov_df$values))
```

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

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

Los primeros dos explican el 99.99% de la variación de los datos, por lo que se puede argumentar que con estos dos se puede reducir la dimensionalidad de los datos y perder únicamente menos de un 0.01% de la variación.

Escriba la ecuación de la combinación lineal de los Componentes principales CP1 y CP2. ¿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.

```
eigen_cov_df$vectors[,c(1,2)]
```

```
##           [,1]           [,2]
## [1,] -1.658168e-06  4.706785e-07
## [2,] -4.048139e-05 -1.774254e-05
## [3,]  5.739096e-06 -1.084543e-05
## [4,]  8.880376e-01  4.597632e-01
## [5,]  4.597636e-01 -8.880405e-01
## [6,]  3.504341e-04  4.016179e-04
## [7,]  2.625508e-04 -1.122118e-03
## [8,]  4.089564e-06  7.790843e-06
## [9,] -1.073825e-06  2.350808e-07
## [10,] 2.547156e-03  7.126782e-04
## [11,] 4.643724e-06 -1.315731e-06
```

$PC_1 = V_1^T X$ y $PC_2 = V_2^T X$ donde X es el vector de variables y V_n es el n -ésimo vector propio.

Para el primer componente principal, las variables 4, 5, y 10 fueron las que más contribuyeron a la variación a lo largo de este eje. Por otro lado, el segundo componente principal tiene como variables con mayor contribución la 4, 5, y 7. Se puede argumentar que estas son las variables con mayor contribución ya que sus magnitudes son las mayores a comparación al resto. Las variables 4 y 5 tienen un orden de 10^{-1} , y las variables 10 y 7 tienen un orden de 10^{-3} . El resto pueden llegar a ser hasta ordenes de exponente -7, las cuales sabemos que no contribuyen de la misma forma al resultado como los previamente explorados.

Con esto podemos inferir que las variables de PNB95 (Var. 4), ProdElec (Var. 5), ConsEner (Var. 10) y ConsAgua (Var. 7) son aquellas que mejor explican la variación de los datos.

2. Análisis de Matriz de Correlación

```
cor_df = cor(scaled_df)
eigen_cor_df = eigen(cor_df)
```

```
cor_df
```

```
##          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
## PropDeform   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      PropDeform      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
## PropDeform -0.37623801 -0.08705811  0.09725032  1.00000000 -0.34833836
## ConsEner    0.78068385  0.16259804 -0.11492480 -0.34833836  1.00000000
## EmisCO2     0.62393719  0.16227447 -0.12333592 -0.37396154  0.87965517
##          EmisCO2
## CrecPobl    -0.17568860
## MortInf     -0.56520778
## PorcMujeres -0.04692837
## PNB95       0.21425123
## ProdElec    0.19760017
## LinTelf     0.62393719
## ConsAgua    0.16227447
## PropBosq    -0.12333592
## PropDeform -0.37396154
## ConsEner    0.87965517
## EmisCO2     1.00000000
```

```
eigen_cor_df$values / sum(eigen_cor_df$values)
```

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

```
sum(eigen_cor_df$values)
```

```
## [1] 11
```

```
sum(diag(cor_df))
```

```
## [1] 11
```

```
cumsum(eigen_cor_df$values / sum(eigen_cor_df$values))
```

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

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

En esta ocasión, los primeros 8 componentes explican el 96.5% de la variación de los datos. Esta vez ha sido más complicado reducir la dimensionalidad de los datos basandose en la correlación de estos mismos.

Escriba la ecuación de la combinación lineal de los Componentes principales CP1 y CP2. ¿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.

```
eigen_cor_df$vectors[,c(1,2)]
```

```
##           [,1]      [,2]  
## [1,] -0.314119414  0.34835747  
## [2,] -0.392395442 -0.04136238  
## [3,]  0.116546319 -0.58283641  
## [4,]  0.295393771 -0.17690839  
## [5,]  0.258964724 -0.17356372  
## [6,]  0.446082934 -0.02719077  
## [7,]  0.092410503  0.32060987  
## [8,]  0.005692925 -0.45742697  
## [9,] -0.243652293 -0.15408201  
## [10,] 0.415029554  0.23286257  
## [11,] 0.374531032  0.29168698
```

$PC_1 = V_1^T X$ y $PC_2 = V_2^T X$ donde X es el vector de variables y V_n es el n -ésimo vector propio.

Para el primer componente principal, las variables 6, 2, y 10 fueron las que más contribuyeron a la variación a lo largo de este eje. Por otro lado, el segundo componente principal tiene como variables con mayor contribución la 3, 1, y 8. A diferencia de los resultados obtenidos al hacer el análisis sobre la matriz de covarianzas, cada coeficiente en sus respectivas combinaciones lineales están más distribuidos sobre todas las variables. Es decir, en la matriz de covarianzas dos variables podrían representar el 99.99% de los datos, mientras en el análisis de correlación es necesario utilizar 8 para 96%.

Con esto podemos inferir que las variables asociadas en el primer componente principal están más correlacionadas entre si, unicamente las variables 8 (PropBosq), 2 (MortInf) y 6 (LinTelf) son las que menos influyen en los componentes principales de la matriz de correlación.

3. Conclusiones

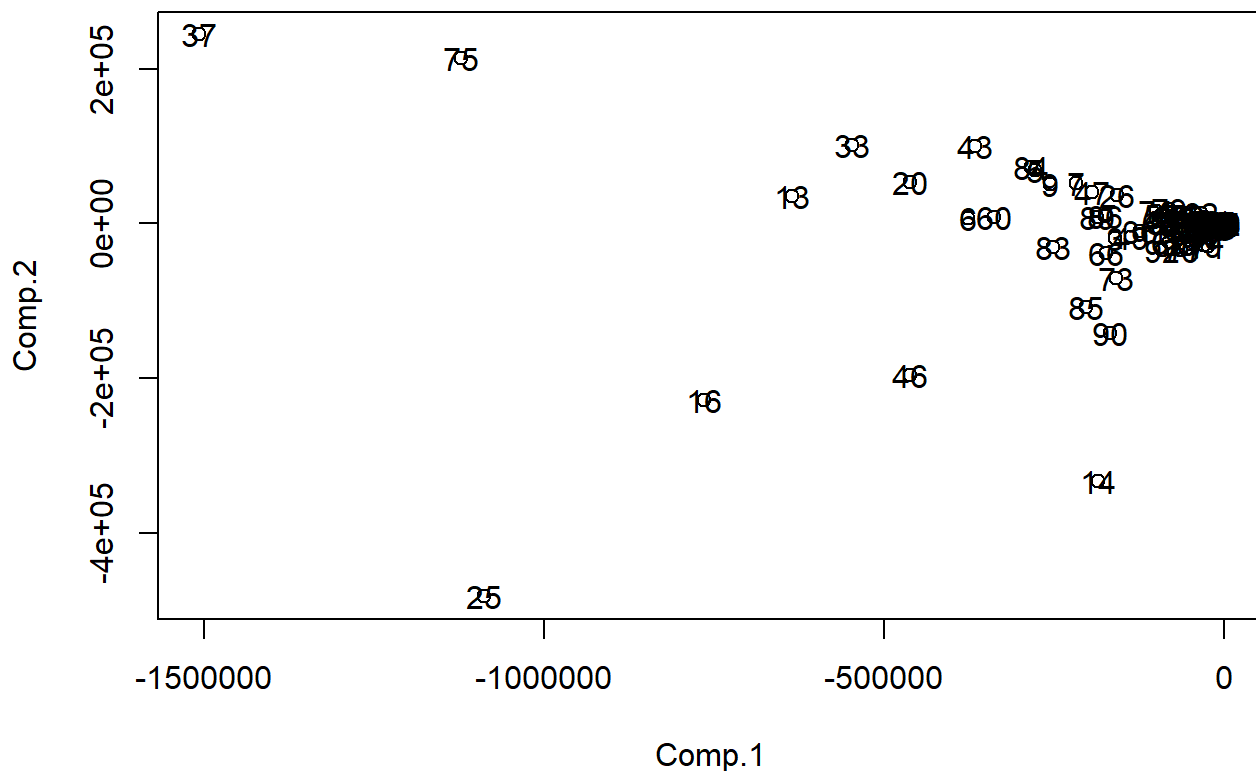
En términos de reducción de dimensionalidad y simplificación de los datos, el análisis de la matriz de covarianza es mejor, ya que permite capturar casi toda la variación de los datos con solo dos componentes principales. Sin embargo, este análisis se hizo sin estandarizar las variables, por lo que existe un sesgo por parte de la magnitud de las covarianzas si el orden de unas variables son considerablemente más grandes que el resto. Por otro lado, un análisis de vectores propios para la matriz de correlación permite un mejor análisis de la relación entre los datos, ya que su análisis no se sesga con la magnitud de la varianza.

PARTE II

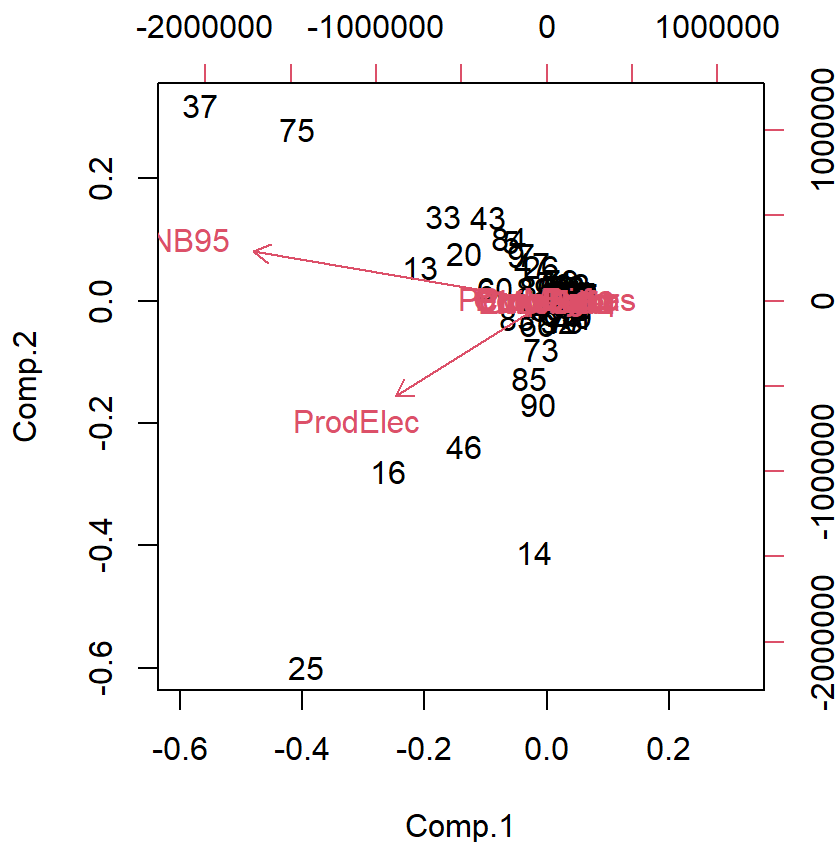
```
cpS=princomp(df,cor=FALSE)
cpaS=as.matrix(df)%%cpS$loadings

plot(cpaS[,1:2],type="p", main = "PCA en S")
text(cpaS[,1],cpaS[,2],1:nrow(cpaS))
```

PCA en S



```
biplot(cpS)
```

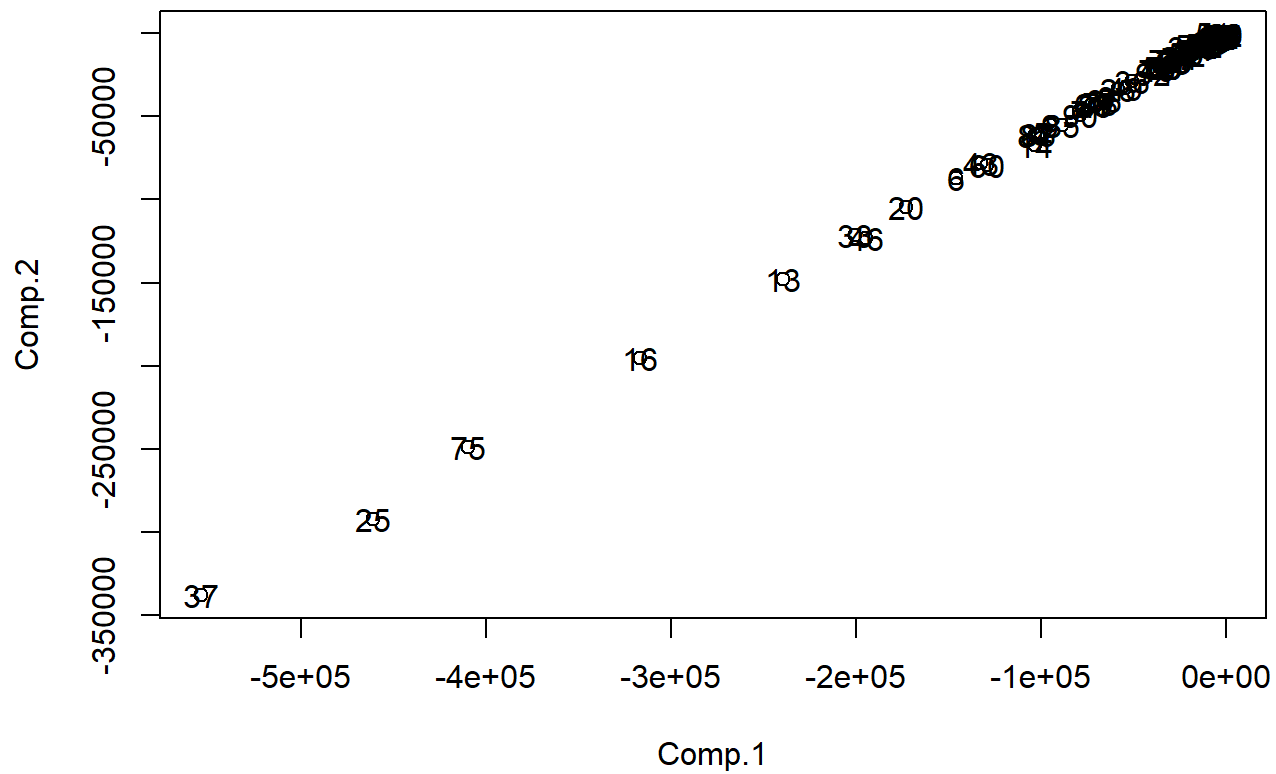
[illegible]

En estos gráficos podemos observar como las variables PNB95 y ProdElec tienen un gran peso sobre los dos componentes principales graficados. Datos que están numericamente más cercanos unos a otros tienen puntuaciones similares, y por lo tanto se observa que los datos atípicos afectan con fuerza la variación de los datos. Esto implica que es probable que existan datos atípicos en las dos variables previamente mencionadas que afecten el análisis.

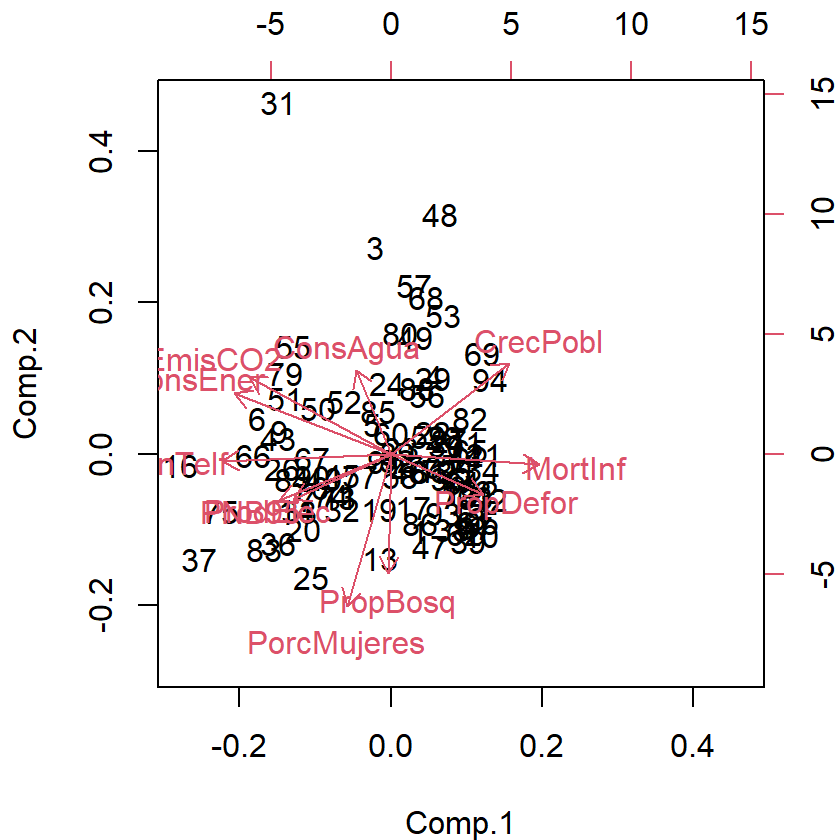
```
cpR=princomp(df,cor=TRUE)
cpaR=as.matrix(df)%*%cpR$loadings
```

```
plot(cpaR[,1:2],type="p", main = "PCA en R")
text(cpaR[,1],cpaR[,2],1:nrow(cpaR))
```

PCA en R



```
biplot(cpR)
```

En este caso, se observa una relación lineal entre el componente 1 y componente 2. Hay una gran concentración de valores tendiendo hacia la esquina superior derecha, contando con algunos datos atípicos tendiendo hacia la esquina inferior izquierda. Nuevamente, el valor de las variables afecta significativamente según su valor en scoring tanto para el componente 1 y componente 2. También se observa que ahora todas las variables contribuyen aproximadamente lo mismo para la explicación de la variación de los datos.

```
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
```

```
cpR$loadings
```

```

##
## Loadings:
##      Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9
## CrecPobl    0.314  0.348      0.440  0.330  0.184  0.163      0.522
## MortInf     0.392      0.178  0.134      -0.187 -0.168  0.531      0.236
## PorcMujeres -0.117 -0.583 -0.167      -0.187 -0.168  0.531      0.236
## PNB95       -0.295 -0.177  0.533  0.262  0.141      -0.149  0.449 -0.370
## ProdElec    -0.259 -0.174  0.614  0.174      0.108 -0.503  0.307
## LinTelf     -0.446      -0.152      0.570  0.447
## ConsAgua      0.321  0.370 -0.736      0.309  0.236
## PropBosq     -0.457 -0.165      0.415  0.754
## PropDefor    0.244 -0.154      -0.337  0.733 -0.509
## ConsEner     -0.415  0.233 -0.206      0.231      0.271
## EmisCO2     -0.375  0.292 -0.206  0.148  0.240      0.335 -0.310 -0.377
##      Comp.10 Comp.11
## CrecPobl    0.347  0.101
## MortInf     -0.390 -0.175
## PorcMujeres 0.429  0.168
## PNB95       0.349  0.152
## ProdElec    -0.338 -0.124
## LinTelf     -0.210 -0.450
## ConsAgua     0.206
## PropBosq
## PropDefor
## ConsEner     -0.361  0.679
## EmisCO2     0.288 -0.467
##
##      Comp.1 Comp.2 Comp.3 Comp.4 Comp.5 Comp.6 Comp.7 Comp.8 Comp.9
## SS loadings  1.000  1.000  1.000  1.000  1.000  1.000  1.000  1.000  1.000
## Proportion Var 0.091  0.091  0.091  0.091  0.091  0.091  0.091  0.091  0.091
## Cumulative Var 0.091  0.182  0.273  0.364  0.455  0.545  0.636  0.727  0.818
##      Comp.10 Comp.11
## SS loadings  1.000  1.000
## Proportion Var 0.091  0.091
## Cumulative Var 0.909  1.000

```

cpR\$scores

##		Comp.1	Comp.2	Comp.3	Comp.4	Comp.5
##	[1,]	0.72890936	-1.39126215	-0.8481719765	-0.073591469	-0.50000388
##	[2,]	2.58182998	-0.66529257	0.0179932618	0.913863832	-0.70274437
##	[3,]	-0.39758336	3.72302269	0.3096903005	1.554743749	1.04487413
##	[4,]	1.14209467	1.39125042	0.1114407253	0.515864365	-0.19046119
##	[5,]	-0.49342134	0.53434151	0.8220319829	-0.647715270	-0.34315035
##	[6,]	-3.46470568	0.63740991	-0.0293781576	-0.021619259	0.49219502
##	[7,]	-2.17279383	-0.82275000	-0.9534192864	-0.012317943	-0.02374627
##	[8,]	2.40954488	-0.67271390	0.1893189948	-0.609690624	0.81447947
##	[9,]	-2.89614082	0.41145447	-0.4269278400	-0.659365647	-0.27183378
##	[10,]	2.35476695	-1.47108866	-0.3902829509	0.598405346	0.17348791
##	[11,]	-1.33332637	-0.78169368	-1.3300508041	-0.203195064	-0.78143504
##	[12,]	1.74798207	-0.71437745	-0.3245715218	0.150290637	0.31714208
##	[13,]	-0.27673336	-1.87094672	1.8524186372	0.882179892	0.91987182
##	[14,]	-2.46559681	-1.02444368	1.0362013577	-1.901655475	-0.86715910
##	[15,]	1.82092277	-0.66531702	-0.4037001540	0.645618315	0.07706329
##	[16,]	-5.47945955	-0.19850119	2.1647894791	0.092647569	1.02359328
##	[17,]	0.58800627	-0.96003954	-0.4452885409	0.080025112	0.31416462
##	[18,]	2.00956372	-1.16203039	-0.5415577544	0.955974054	0.09303764
##	[19,]	-0.29960643	-0.99049133	-1.0045506677	-0.280330574	0.62872500
##	[20,]	-2.29026695	-1.36456367	0.2998275524	-0.079379433	0.84785417
##	[21,]	2.42342143	-0.01335946	-0.0174695036	0.874888055	0.26315536
##	[22,]	1.09159218	0.39795846	-0.0439162727	-1.144318845	1.33542588
##	[23,]	0.27721583	0.10152968	-0.1633837359	-1.284972099	-0.53139906
##	[24,]	-0.09057074	1.27372875	0.6121938800	-1.614481500	-0.86558253
##	[25,]	-2.07827375	-2.22775388	5.6186567173	1.564866235	0.09492346
##	[26,]	-2.83676974	-0.25680289	-1.3306480217	-0.141143402	-0.53189153
##	[27,]	1.52494532	0.25061327	0.0009542767	-0.722723381	0.96154821
##	[28,]	1.36116760	0.29929119	-0.0423868022	-0.508911262	0.98024651
##	[29,]	1.09171875	1.35835050	0.8553152377	-0.337523648	-1.03137054
##	[30,]	1.43567654	0.15862883	-0.0973768627	-0.409179682	0.11254747
##	[31,]	-2.91856734	6.32387288	-1.4410303861	2.099274610	2.95453565
##	[32,]	-1.25841968	-0.99859923	-1.3689008333	-0.511086368	-0.41255092
##	[33,]	-2.38136140	-1.02584549	0.8617364555	-0.374852903	0.33526958
##	[34,]	2.37757571	-0.27883039	0.0898678751	0.864186236	-1.04035821
##	[35,]	1.62924651	-0.31905705	0.3218882832	-1.008235534	1.21487735
##	[36,]	-2.93201062	-1.60509857	-1.8221637073	-0.498125335	0.89912187
##	[37,]	-4.97695562	-1.89271250	4.1962700899	1.413711348	0.65488368
##	[38,]	1.60108643	-1.36044031	-0.8476057434	0.817930511	0.72998871
##	[39,]	2.02573098	-1.57874581	-0.5077464810	0.474183687	0.17674226
##	[40,]	-1.74189010	-0.43529312	-0.9986529400	-0.597507389	0.06932396
##	[41,]	1.92730323	0.21051414	-0.1359020129	0.270911450	0.95125102
##	[42,]	2.58183172	-0.82644680	-0.0485795308	-0.661749495	1.23444969
##	[43,]	-2.94112439	0.28283423	-0.2406149332	0.065528355	-0.37454058
##	[44,]	1.97035275	-0.02637320	-0.1370511806	-0.034020125	1.13573594
##	[45,]	-1.26584011	-0.37726758	-0.7179472075	-0.936617036	-1.37681681
##	[46,]	0.45737247	-0.31082621	2.7668873314	0.580986554	-0.22076791
##	[47,]	1.02905780	-1.66278015	-0.0241003703	0.257818163	0.54211183
##	[48,]	1.29041169	4.31461747	3.4113319177	-4.541053875	-0.70875187
##	[49,]	0.64233821	2.10518139	1.1998337755	-0.332462604	-0.14125232
##	[50,]	-1.87699034	0.82187675	-1.0794918131	0.061270693	-1.37447486
##	[51,]	-2.74565229	0.99587622	-1.5953715464	-0.502187456	-0.37860128

##	[52,]	-1.22534779	0.94751833	-0.7366871197	0.517680285	-0.37650037
##	[53,]	1.35824642	2.50454513	0.0487011933	1.703897528	-0.54173093
##	[54,]	1.72981419	-0.21037981	-0.1696853204	0.620766731	-0.97516034
##	[55,]	-2.55079364	1.94419512	-1.2474348173	-0.441138048	-0.23570010
##	[56,]	0.93875305	1.04736721	-0.1643462450	0.259761534	-0.32986368
##	[57,]	0.61006775	3.02862639	0.3821171418	0.858292932	-0.78790783
##	[58,]	0.25381239	-0.40313078	-0.3128631161	-0.772144430	1.54179628
##	[59,]	0.96038716	0.32233749	0.0024228115	0.447983712	-1.61424357
##	[60,]	0.00707317	0.36342432	1.3096416444	-0.274294666	0.63131714
##	[61,]	2.18204075	-1.22904854	-0.1634955036	0.381150450	-0.94683317
##	[62,]	1.90443422	-1.42355950	-0.2800808748	0.104741384	-0.00753101
##	[63,]	2.20882137	-0.78724553	-0.1395782436	0.393222741	-0.11263353
##	[64,]	1.85935981	-0.51767977	-0.2873304104	-0.070459571	1.05847860
##	[65,]	2.05518046	0.06056508	0.0681132364	0.846945427	-0.39300739
##	[66,]	-3.57493689	-0.02632257	-1.2799437562	0.178545515	-0.50690362
##	[67,]	-2.04952706	-0.08749635	-1.2232495387	-0.496158533	-0.16896364
##	[68,]	0.91162733	2.81893150	0.0038602925	0.938768137	0.71849709
##	[69,]	2.37683247	1.80522003	2.0764427054	-2.187118641	0.98740177
##	[70,]	0.88052404	-0.21674355	-0.2543159839	-1.037091983	0.76930260
##	[71,]	1.90627528	-0.04277925	-0.0350846412	0.006192558	1.26838851
##	[72,]	1.25408854	-0.30225699	-0.1582001867	0.168806565	0.25818413
##	[73,]	-1.37863309	-0.78615264	-0.4642643142	-0.062492415	-0.56369689
##	[74,]	-1.56342191	-0.75210208	-0.7152880209	-1.036375898	-0.98312502
##	[75,]	-4.41996326	-1.03079641	2.3962157660	1.711696753	-0.41107455
##	[76,]	-2.08169541	-0.60827843	-1.5080138636	-0.240859352	-0.20913315
##	[77,]	-0.78715421	-0.41090107	-0.1940655951	-1.531446326	-0.98352127
##	[78,]	1.78980987	-0.75828097	-0.3315174897	0.379615060	-0.10100451
##	[79,]	-2.75004895	1.44528470	-1.9388585015	0.351721612	1.89959676
##	[80,]	0.23588358	2.16520539	0.0065318694	1.621403830	-2.81583098
##	[81,]	1.00712247	-0.30703667	-0.1961378200	-0.830208326	-0.11174702
##	[82,]	2.05664438	0.64428712	0.4730228419	-0.294878896	-0.25800943
##	[83,]	-3.29612254	-1.71608400	-1.2139057475	-0.149068245	0.81578722
##	[84,]	-2.57717085	-0.46226995	-0.7611608503	0.366500658	-0.31731947
##	[85,]	-0.32437979	0.77155643	0.6514731417	1.065007975	-0.90312023
##	[86,]	0.77791978	-1.25024262	0.3172685877	-1.158234561	0.99686114
##	[87,]	2.18495598	-1.31478099	-0.3978946507	0.599470131	0.01455104
##	[88,]	0.68304922	1.17435335	-0.0013716187	0.664575735	-1.99923201
##	[89,]	0.15569371	0.01587790	0.4388495055	0.084294479	-0.43670486
##	[90,]	-2.03915283	-0.40767668	-0.1487843291	-0.403476043	-0.85085412
##	[91,]	-0.15522297	-0.13457766	-0.5667862325	-0.204660451	-1.79985294
##	[92,]	0.16590692	-0.11407437	-0.3876432529	0.103835008	1.19048308
##	[93,]	1.36639109	-1.07234718	-0.2707095806	-0.390566877	-0.18970797
##	[94,]	2.59902204	1.35007750	0.5179565112	1.204499498	-0.55737622
##	[95,]	2.34078019	-1.23017338	-0.1460885190	0.403483869	-0.03543444
##	[96,]	1.50544837	-0.44756342	-0.3782396979	0.512657083	-0.35065555
##		Comp.6	Comp.7	Comp.8	Comp.9	Comp.10
##	[1,]	0.80784207	-0.72615475	-0.553295622	-0.372120272	0.009735613
##	[2,]	-0.30411695	1.77179319	0.632057863	-0.101403269	-0.190772201
##	[3,]	-0.03554062	-0.53348677	-0.500188250	-0.170540320	-0.086366759
##	[4,]	-0.83711734	-1.06464694	-0.149317913	-0.148553797	-0.216248463
##	[5,]	0.21957753	-0.78309862	0.261968758	-0.357345257	0.231231778
##	[6,]	-0.28913091	0.87484985	0.176436007	0.156187633	0.512396414

```
## [7,] 0.56506892 -0.50997667 0.483911034 0.038236929 -0.122478244
## [8,] -2.19919325 0.61274661 0.230942132 -0.205580834 -0.078777185
## [9,] -0.22433419 0.19352855 0.452655837 -0.222311239 -0.064624562
## [10,] 0.31829564 1.16678742 0.293326845 0.135894167 0.110885930
## [11,] -0.09206031 0.32948245 -0.772366915 -0.369509066 0.537101899
## [12,] 0.47890587 0.05666572 0.098709487 -0.197800926 -0.284492348
## [13,] 1.20582370 -0.83810828 0.049421534 -0.693620942 -0.160337074
## [14,] 0.34966758 0.75339256 -1.393658316 0.736756389 -0.635168391
## [15,] 0.62156093 -0.21788549 -0.057754709 0.176749246 0.089073246
## [16,] 1.36076167 1.53024840 -0.482034850 0.861240622 -0.428367427
## [17,] 0.69560009 -0.84230922 -0.272244658 0.029935726 0.003763258
## [18,] 1.32169639 0.84656886 0.162832294 -0.052238265 -0.050058802
## [19,] 1.97424726 0.46395446 -1.121917163 -0.480987787 0.921048365
## [20,] 1.28855332 -0.52225306 0.457696290 -0.101452235 0.021737578
## [21,] 0.35991162 0.21819852 0.339568958 0.001836263 -0.323058649
## [22,] -0.54164051 -1.10644700 0.076339374 0.606464915 0.191532884
## [23,] -0.57783431 -0.70051174 -0.611412626 -0.126162315 0.458044276
## [24,] 0.29700519 -0.60727267 -0.071589728 0.295096981 0.562822006
## [25,] -1.14874670 0.31385725 -2.340665068 0.860771424 -0.642782584
## [26,] -1.10866992 0.17523743 0.679895328 0.283948640 -0.025845950
## [27,] -0.88504115 -0.88148812 -0.002275161 -0.081181943 -0.243308727
## [28,] 0.49685891 -1.03070745 -0.085380563 -0.147354950 -0.304067738
## [29,] -0.27828837 -0.14711698 0.071582115 -0.012261324 -0.133160657
## [30,] -1.50062573 -0.68874101 -0.078477089 0.004773382 -0.042451809
## [31,] 0.31389423 1.79846338 -0.865024767 -0.639726390 0.561743846
## [32,] -0.03288863 0.12323827 -0.540373423 -0.218696491 0.146894425
## [33,] 0.79732387 -0.98634219 0.606983781 -0.561898666 0.014905177
## [34,] -0.37846276 1.15849517 0.529363585 -0.249955324 -0.374707065
## [35,] -0.91022629 -0.29279310 -0.074681781 0.115000153 0.341407723
## [36,] 1.39939265 0.28442074 0.349855733 0.337594246 -0.593377278
## [37,] -0.64136716 -0.55453119 1.795851675 -0.614817452 0.818750131
## [38,] 1.63932685 1.11677938 -0.046782922 -0.390122681 0.044297877
## [39,] 0.05079416 0.94287493 0.040988951 0.423111744 0.526939604
## [40,] 0.67237074 -0.74318963 0.395816596 0.142007257 -0.385775889
## [41,] 0.22586185 -1.08096047 -0.016160723 -0.008762247 -0.336690506
## [42,] -2.91415661 0.47380860 0.159012570 -0.096848585 -0.047617977
## [43,] -0.80532332 -0.12974034 0.890730263 -0.049672941 0.078815154
## [44,] 0.17631320 -0.73914063 -0.019616357 0.163592334 -0.103628873
## [45,] -0.38718992 -0.18241372 -0.487852458 -0.422922696 0.116933173
## [46,] -0.23680895 0.03500398 -0.682681683 -0.006010450 -0.780217265
## [47,] 0.78467638 -0.38845677 -0.062061798 -0.508888888 -0.004746820
## [48,] 2.22699289 1.95876247 0.984881920 -0.259220916 0.165133205
## [49,] 0.53919612 -0.16872044 0.088792408 0.219651120 0.128893381
## [50,] -0.57416857 -0.77499407 -0.016289614 -0.380886269 -0.432427873
## [51,] -1.21595396 0.66390581 0.771686993 0.947266954 -0.578229091
## [52,] -0.57261068 0.01299615 0.384302087 1.053599868 0.652511517
## [53,] 0.18513575 -0.99753299 0.059679942 0.853591766 0.253900056
## [54,] -1.09037365 0.50123670 0.019646725 0.553770703 0.647792224
## [55,] -1.15691254 -0.02713003 -0.398285811 -1.227516460 -1.406710901
## [56,] -0.59424470 -0.86479549 -0.103806095 0.007616588 -0.114610323
## [57,] 0.58481604 0.25226911 0.023365749 -0.200465341 -0.092577022
## [58,] 0.60157558 -0.53038684 -0.155341728 0.475284945 0.430865365
```

```

## [59,] 0.59227384 -0.22056478 0.025257483 -0.097405988 -0.023370987
## [60,] -0.06580998 -0.44367397 -0.030575514 -0.193804769 0.139657342
## [61,] -0.47668276 1.47270350 0.420816724 -0.440723504 -0.326570016
## [62,] 0.19781823 0.48265459 0.169621903 -0.436540290 -0.252571563
## [63,] 0.26481583 0.61508736 0.287303057 -0.207637436 -0.285912411
## [64,] 0.57087654 -0.33746779 -0.068467673 0.325569748 0.191512055
## [65,] -0.33716018 0.30145553 0.207780549 -0.051007291 -0.157266305
## [66,] 0.30238469 0.68622182 -0.010544273 0.181442707 -0.053944439
## [67,] -0.13264117 0.17967343 0.355506820 0.604054699 -0.058624024
## [68,] 0.72517160 -1.24460717 -0.128334788 0.512953778 0.011816254
## [69,] -0.66449161 0.82560208 0.609413693 0.034499074 -0.123063484
## [70,] 0.25987394 -0.77897360 -0.111189619 0.176691075 0.126236503
## [71,] -0.53903427 -0.91881973 -0.179169095 0.116397211 -0.325253419
## [72,] 1.18936670 -0.91844565 -0.021325863 -0.328096383 -0.412367556
## [73,] -0.28507300 0.15495329 -1.009649014 -0.379210587 0.255501560
## [74,] 0.27459788 -0.50080155 0.146623315 0.013970576 0.067617690
## [75,] -0.94852774 -0.47367599 1.225320719 -0.851968468 0.761795007
## [76,] -0.21723604 0.42968620 -0.934246610 -0.692207611 0.203152337
## [77,] 0.11554583 0.17875686 -0.554745236 -0.394814669 0.247468320
## [78,] 0.40901207 0.23421172 0.005183614 0.230688589 0.267011734
## [79,] -1.89495693 0.84653919 0.037694161 -0.111205241 -0.220356117
## [80,] 1.36120504 -0.80352618 -0.169255190 0.340197654 0.225836392
## [81,] -0.36376265 -1.14055365 -0.456193339 -0.091461818 0.202785333
## [82,] -0.08320825 -0.15506670 0.289465711 -0.398964245 -0.535973454
## [83,] 0.97640870 0.14153963 0.757251710 0.875169218 -0.737133301
## [84,] -0.07248439 -0.58018484 1.106807049 0.447026886 -0.262964869
## [85,] -0.36287581 0.59243429 -0.631900107 0.028823059 -0.034823375
## [86,] -1.40268519 0.12218895 -0.189582873 -0.092020410 0.501596029
## [87,] 0.05858001 1.01606846 0.169646893 0.349575394 0.382113102
## [88,] 0.21742413 -0.72010273 -0.065452089 0.019806473 -0.022309377
## [89,] 0.26616269 -0.19446115 0.422260464 0.099927140 -0.125123839
## [90,] -0.58800302 0.85591391 -1.405122234 -0.338474864 0.301556812
## [91,] -0.84346603 -0.77258207 -0.050819744 0.125788252 0.065285820
## [92,] 0.75061244 -0.50724353 -0.491486703 0.014333639 -0.056686757
## [93,] -0.55871643 0.30210302 -0.243226616 0.439609888 0.795801515
## [94,] 0.13482549 0.59717736 0.630402314 0.256238635 -0.249099990
## [95,] 0.43667147 1.31783394 0.414310430 -0.202844408 -0.278568141
## [96,] -0.29283061 0.38971270 -0.136145073 0.312480830 0.499767972
##      Comp.11
## [1,] 0.223373687
## [2,] -0.095230192
## [3,] 0.047421821
## [4,] -0.118617537
## [5,] 0.126234519
## [6,] -0.309741838
## [7,] -0.216184668
## [8,] -0.144290019
## [9,] 0.097350314
## [10,] 0.101538828
## [11,] -0.129132307
## [12,] -0.040376789
## [13,] 0.187387018

```

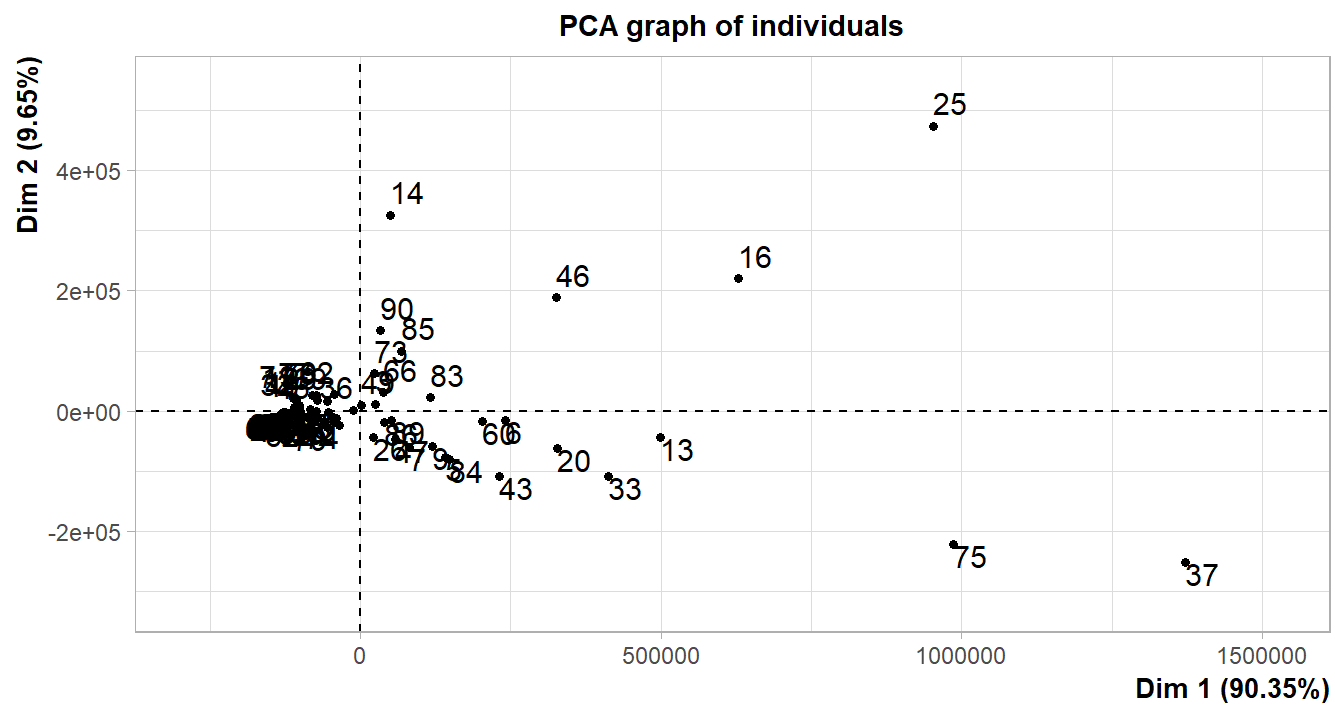
[14,] -0.430681164
[15,] 0.173612748
[16,] 0.352964790
[17,] 0.043746053
[18,] 0.061800852
[19,] -0.287843723
[20,] -0.086046570
[21,] -0.106944188
[22,] -0.150026253
[23,] 0.288646156
[24,] 0.182799247
[25,] -0.112940501
[26,] -0.662042329
[27,] -0.270493114
[28,] -0.129959219
[29,] -0.025978573
[30,] -0.055360944
[31,] -0.235951010
[32,] 0.318973262
[33,] -0.130906320
[34,] -0.116594079
[35,] 0.106155534
[36,] 0.373402377
[37,] 0.394228914
[38,] -0.240042226
[39,] 0.286173751
[40,] -0.785934356
[41,] -0.100733902
[42,] -0.196118242
[43,] -0.088994774
[44,] -0.035427491
[45,] 0.159466951
[46,] -0.218021553
[47,] 0.189598395
[48,] 0.018387493
[49,] 0.074359834
[50,] -0.440199915
[51,] 1.019077493
[52,] -0.318280561
[53,] 0.080310512
[54,] 0.280253716
[55,] 1.107551623
[56,] -0.095833840
[57,] -0.055189007
[58,] 0.182109146
[59,] 0.090315789
[60,] 0.131322998
[61,] -0.071580946
[62,] 0.015076764
[63,] -0.040136572
[64,] 0.157383476
[65,] -0.064619019

```
## [66,] -0.384335022
## [67,] -0.008832707
## [68,]  0.217073172
## [69,] -0.208152702
## [70,] -0.021041507
## [71,] -0.122934331
## [72,] -0.036255778
## [73,] -0.054397723
## [74,] -0.246188168
## [75,] -0.040832956
## [76,] -0.152317046
## [77,]  0.093037662
## [78,]  0.186185058
## [79,]  0.057716847
## [80,]  0.225695416
## [81,]  0.179895648
## [82,] -0.163105349
## [83,]  0.134734491
## [84,] -0.370649029
## [85,] -0.108197371
## [86,]  0.165000832
## [87,]  0.210201888
## [88,]  0.096317929
## [89,] -0.221207154
## [90,] -0.095062679
## [91,] -0.115379366
## [92,]  0.100593586
## [93,]  0.366045293
## [94,] -0.126562475
## [95,] -0.027926947
## [96,]  0.206310165
```

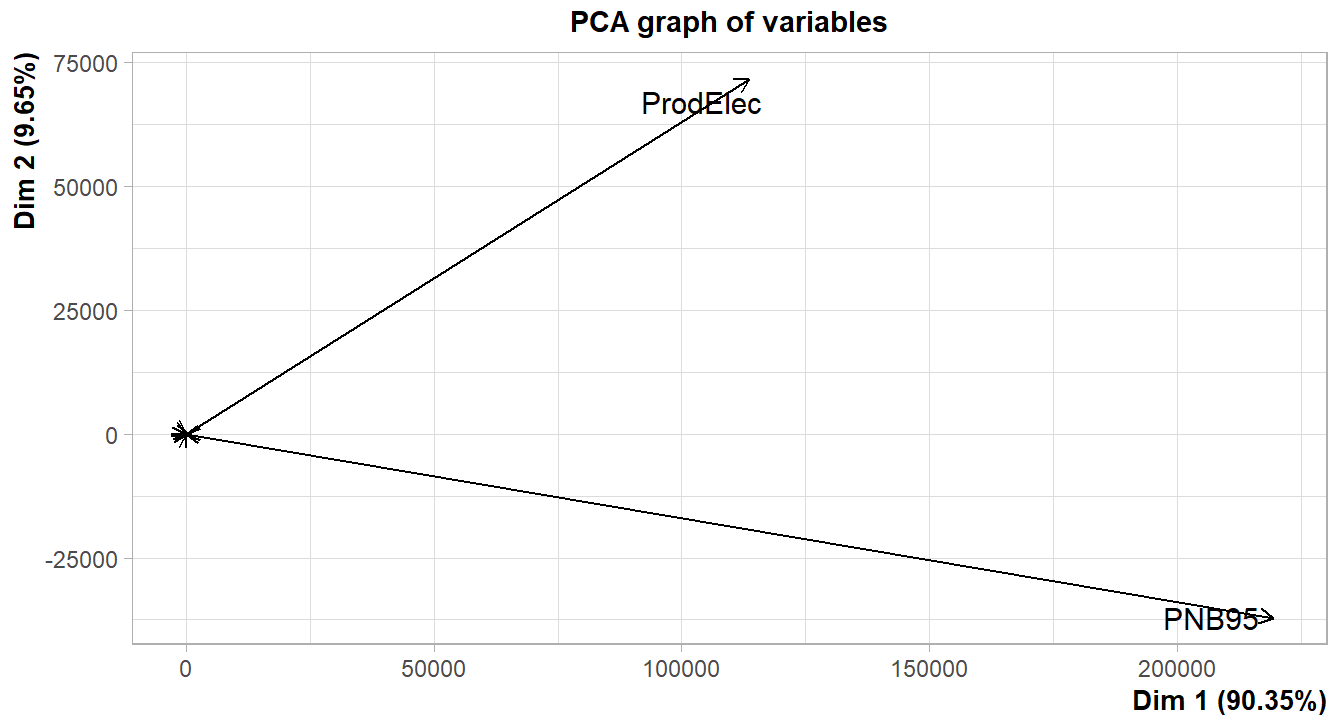
Usando el comando `summary` en el análisis de componentes principales sobre la matriz de correlaciones, se obtiene de manera automatizada los hallazgos de la parte 1. La matriz de loadings nos permite encontrar que variables y coeficientes son los utilizados para llegar a cada combinación lineal correspondiente a cada componente principal. La tabla de scoring permite encontrar cuanto se evalúa cada componente para las variables de cada registro.

PARTE III

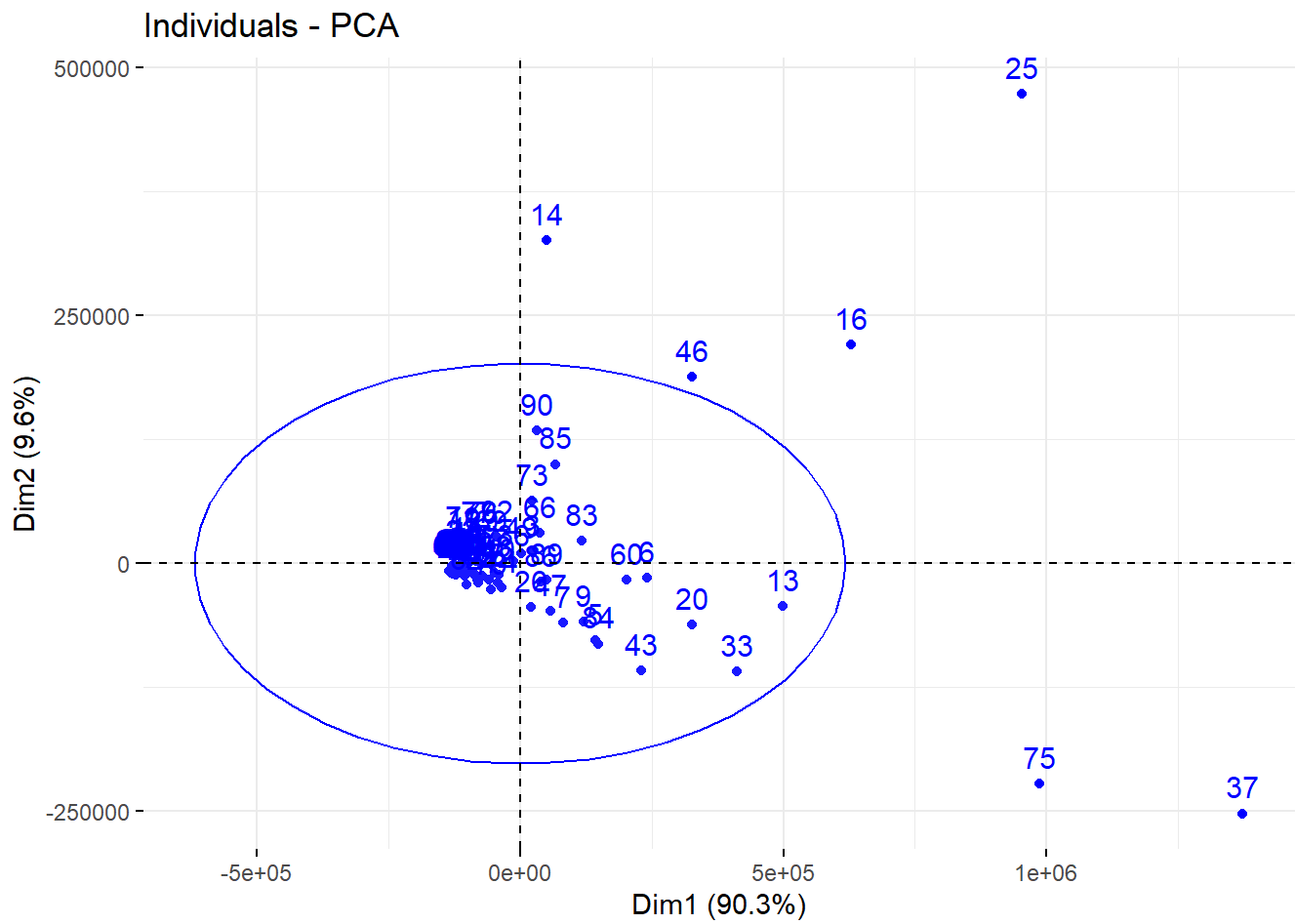
```
cpS = PCA(df, scale.unit=FALSE)
```

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

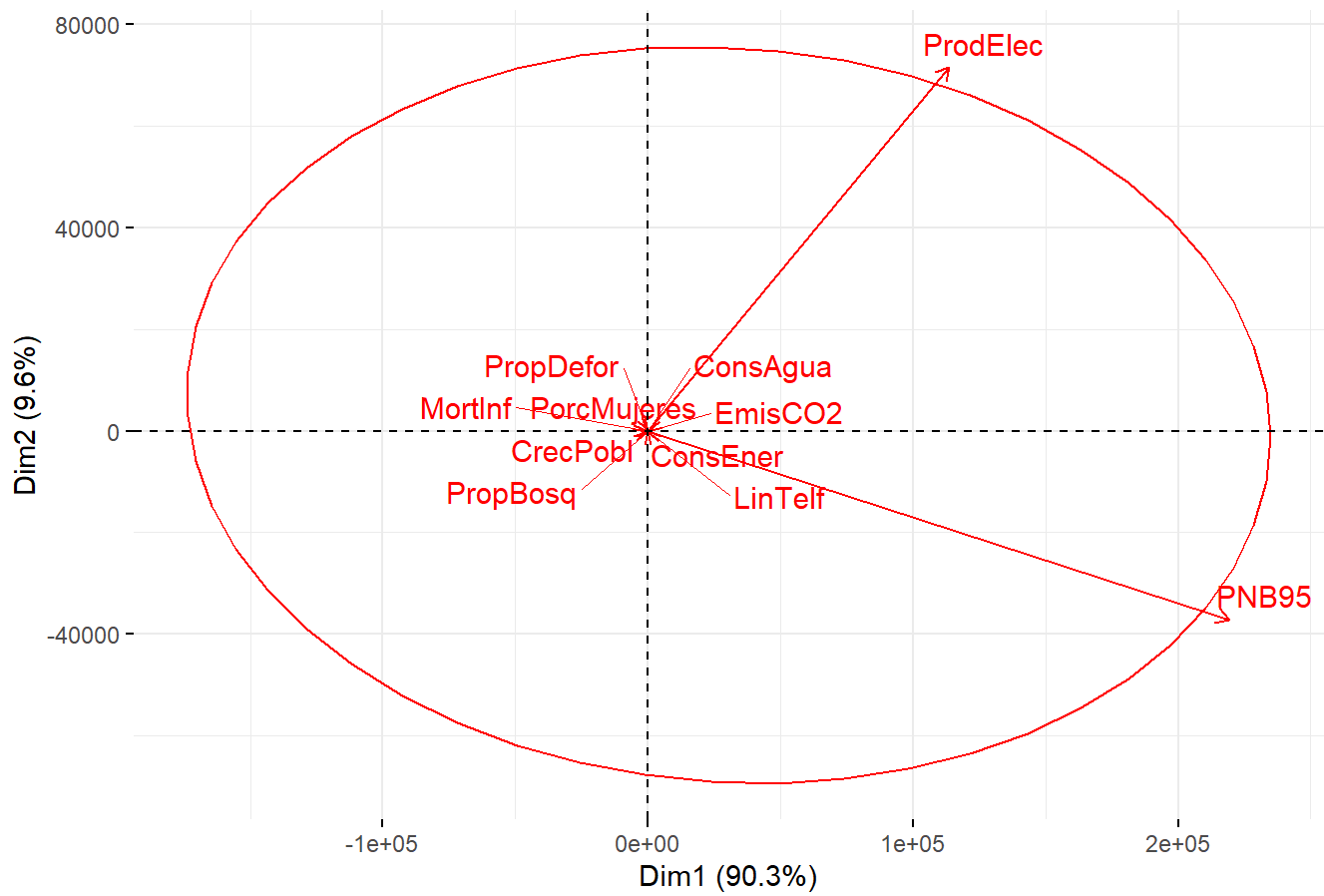


```
fviz_pca_ind(cpS, col.ind = "blue", addEllipses = TRUE, repel = FALSE)
```



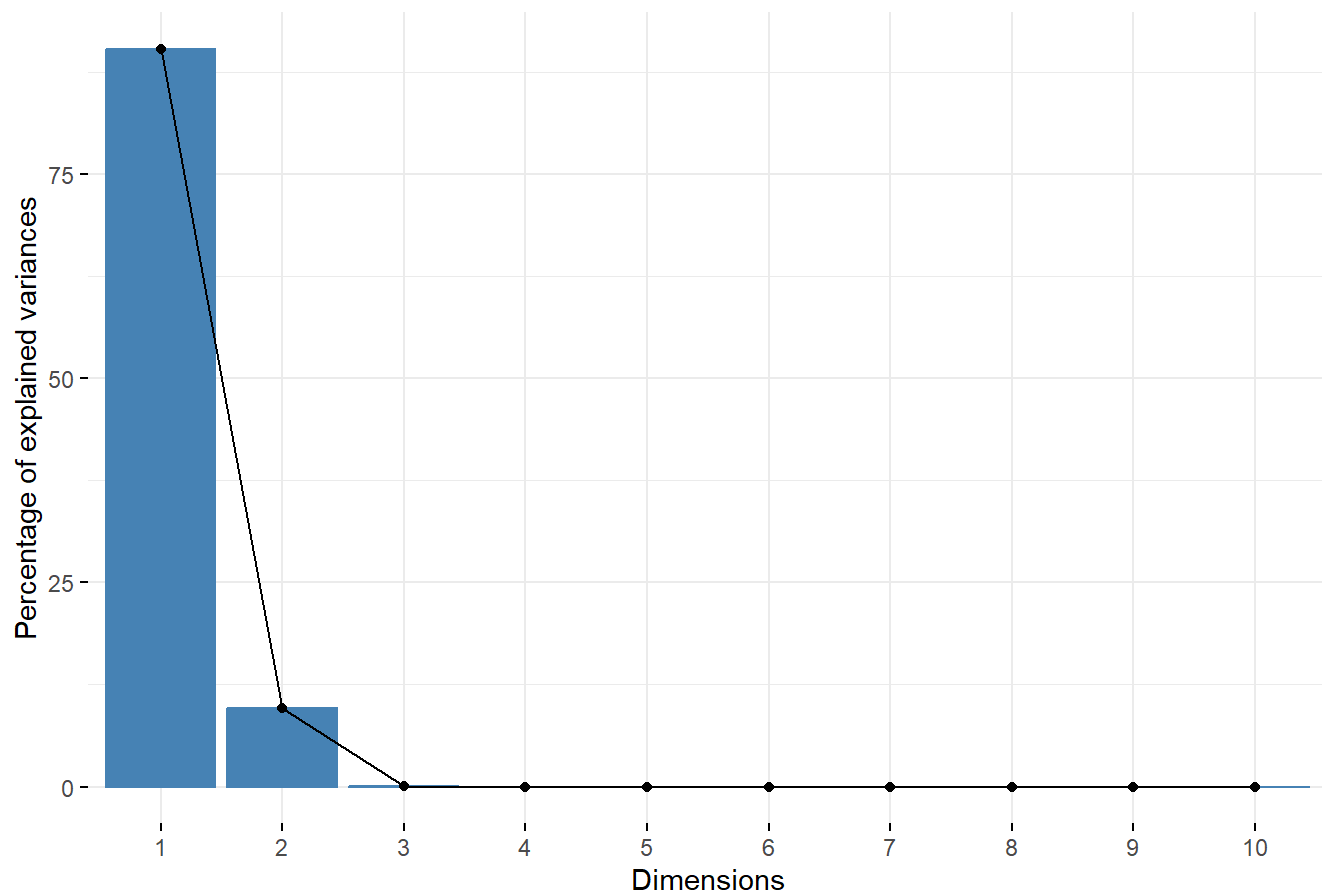
```
fviz_pca_var(cpS, col.var = "red", addEllipses = TRUE, repel = TRUE)
```

Variables - PCA

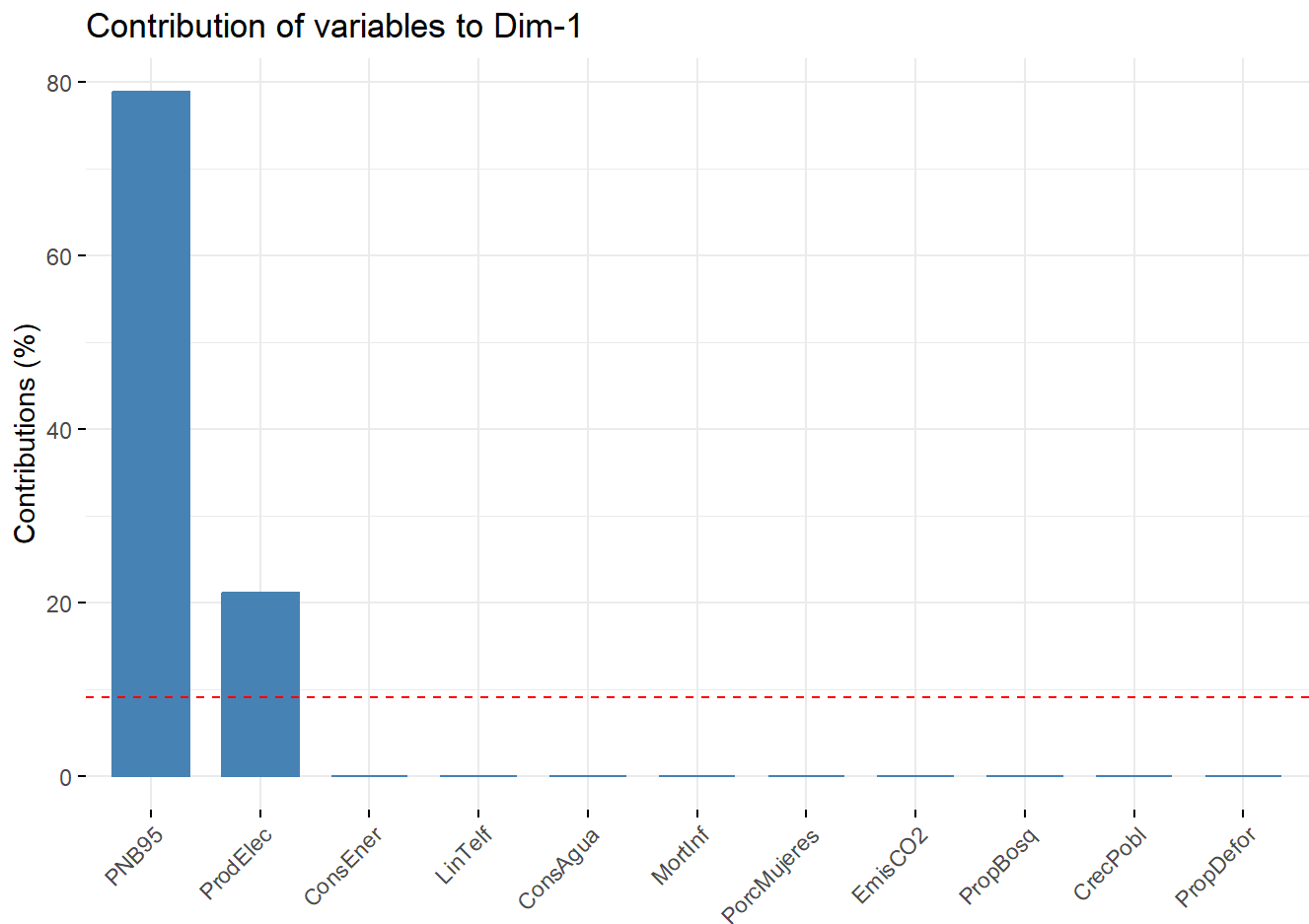


```
fviz_screplot(cpS)
```

Scree plot



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



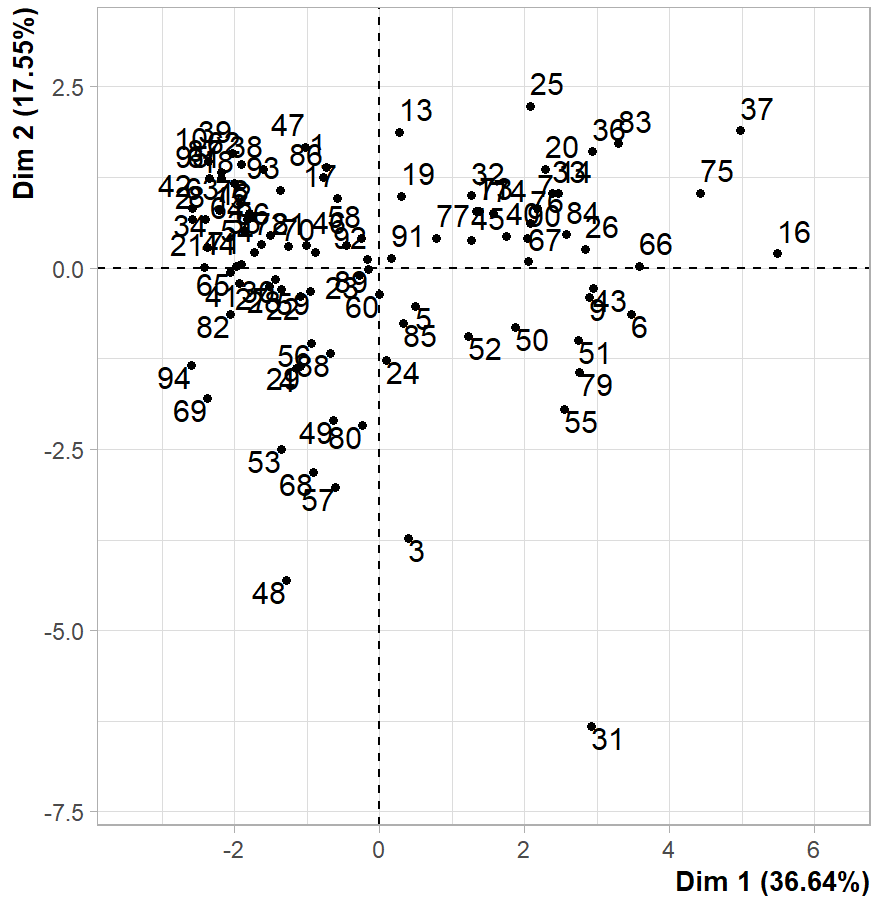
El primer gráfico muestra las proyecciones de los datos sobre los ejes principales 1 y 2 sobre la matriz de covarianzas. Se puede observar que la gran mayoría de los datos están en los cuadrantes 2 y 3, y se tienen algunos datos más esparsos hacia la derecha. Observando la magnitud del eje de la dimensión 1, hace sentido que represente el 90% de la variabilidad. El segundo gráfico afirma que hay dos variables que están sesgando la variabilidad de los datos, probablemente por la magnitud de estos mismos. El resto parecen ser de magnitudes similares y en direcciones relativamente distintas.

El tercer gráfico es el mismo que el primero con una elipse que contendría el 95% de los datos en el caso de que se asemeje a una distribución normal. Lo mismo ocurre para la gráfica cuatro, donde la elipse contiene el 95% de las variables modeladas.

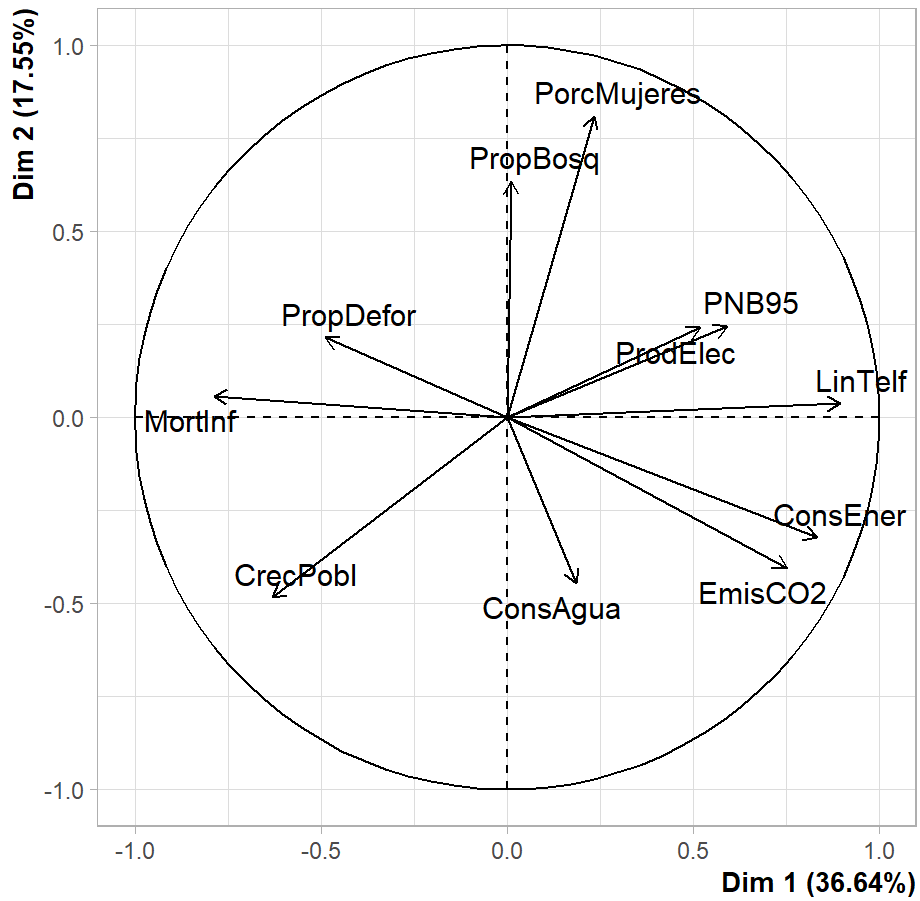
Las gráficas cinco y seis muestran la contribución de la explicación de la variabilidad por componente principal y la contribución por componente por cada variable. Se observa nuevamente que los primeros dos componentes aportan la mayor contribución, mientras que las primeras dos variables son las que más contribuyen al componente 1.

```
cpR = PCA(df, scale.unit=TRUE)
```

PCA graph of individuals

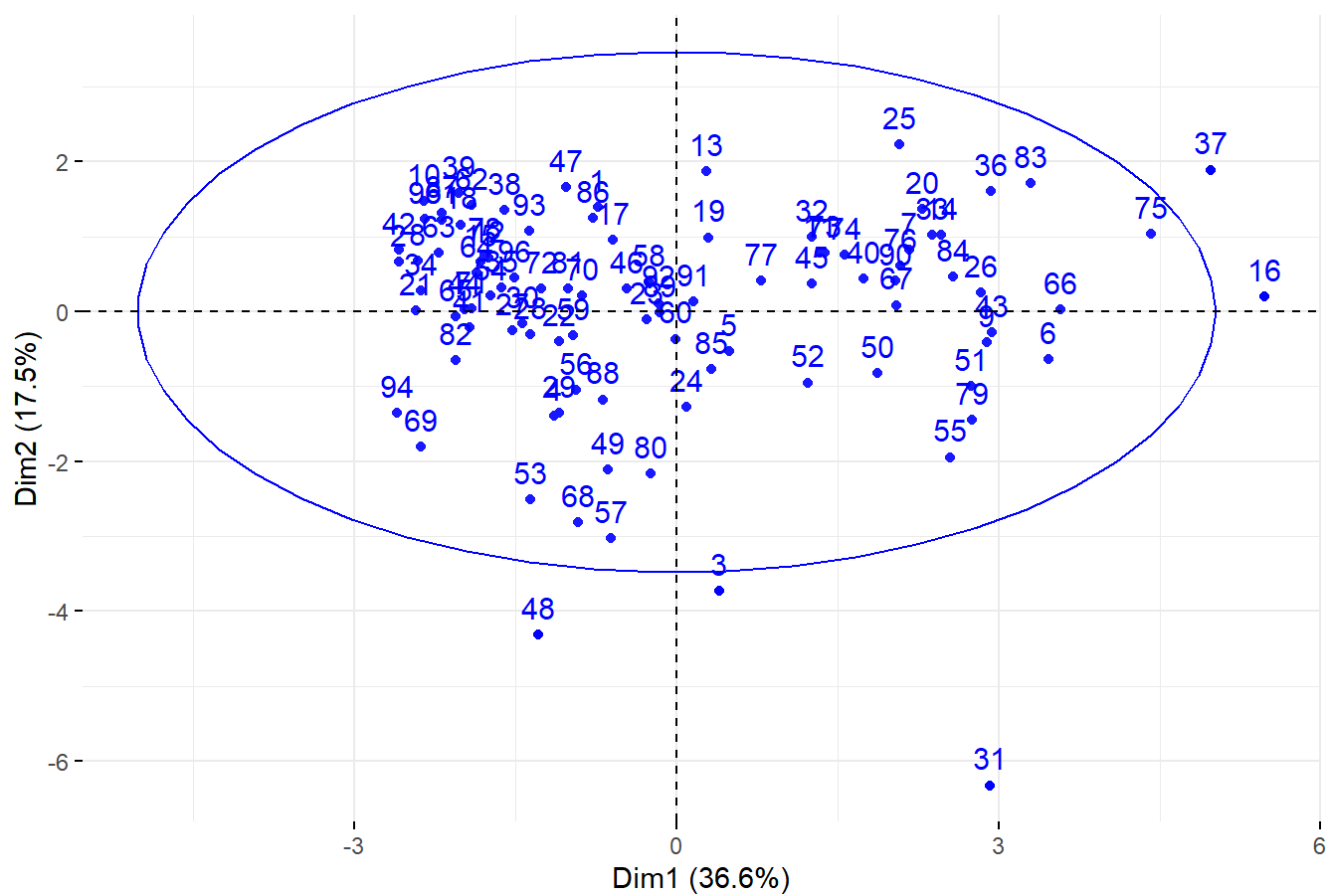


PCA graph of variables



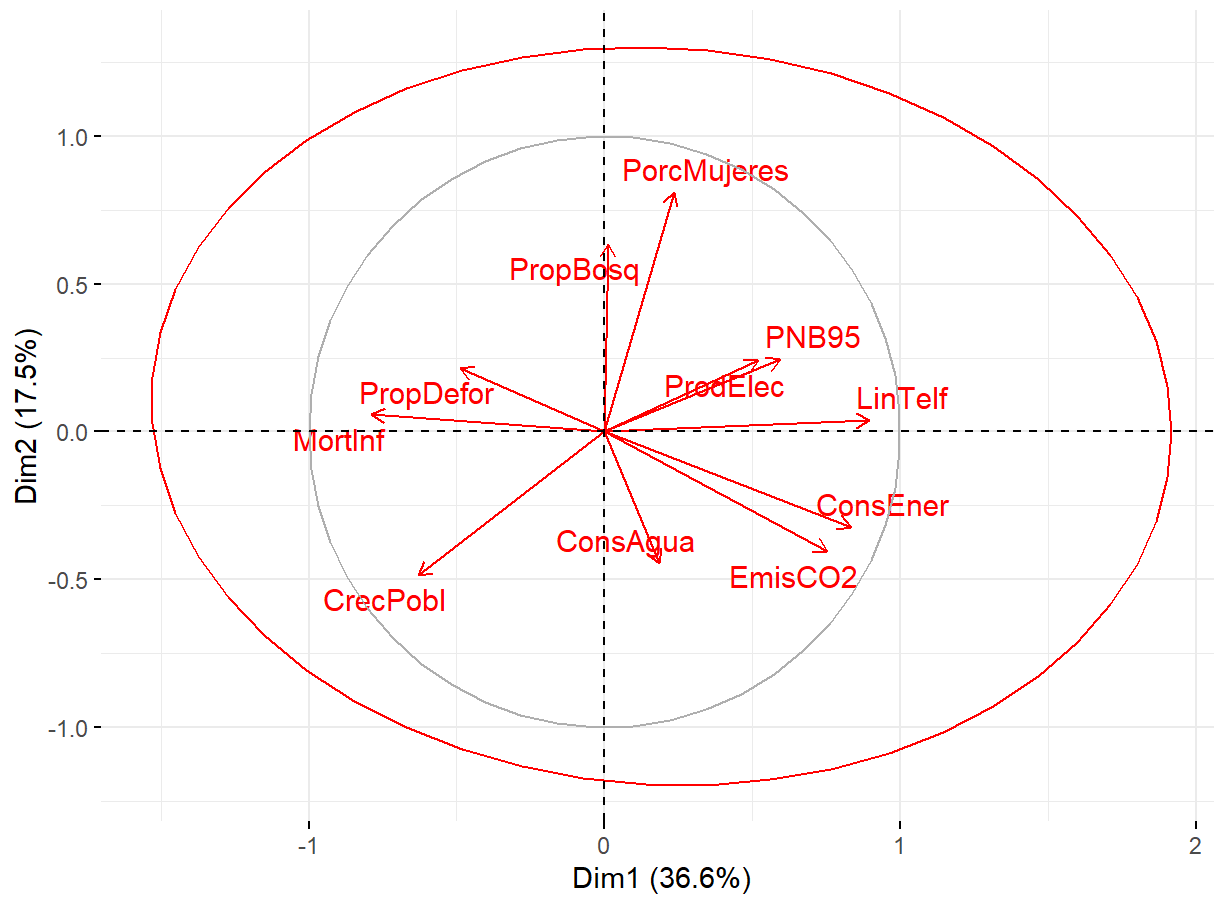
```
fviz_pca_ind(cpR, col.ind = "blue", addEllipses = TRUE, repel = FALSE)
```

Individuals - PCA



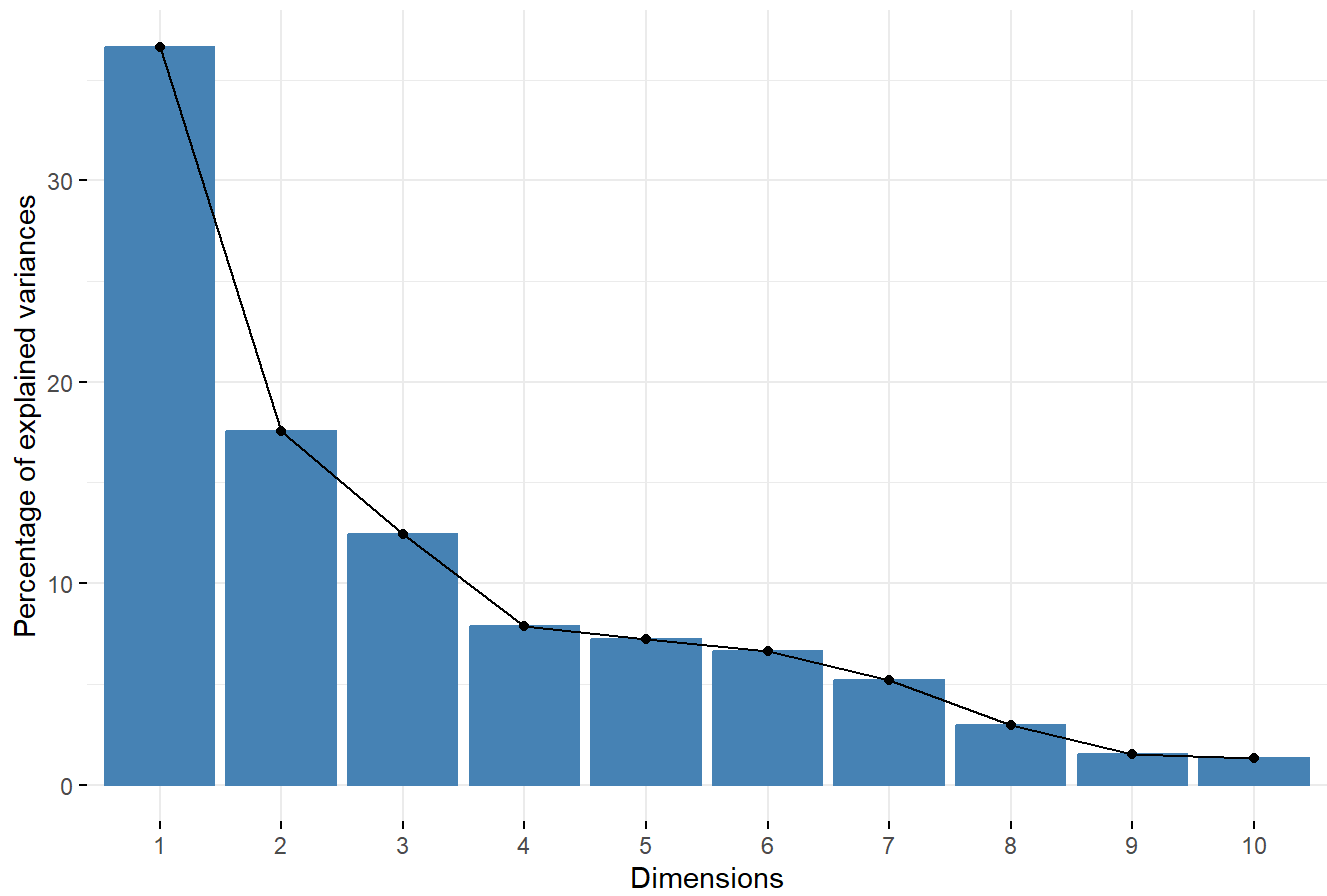
```
fviz_pca_var(cpR, col.var = "red", addEllipses = TRUE, repel = TRUE)
```


Variables - PCA



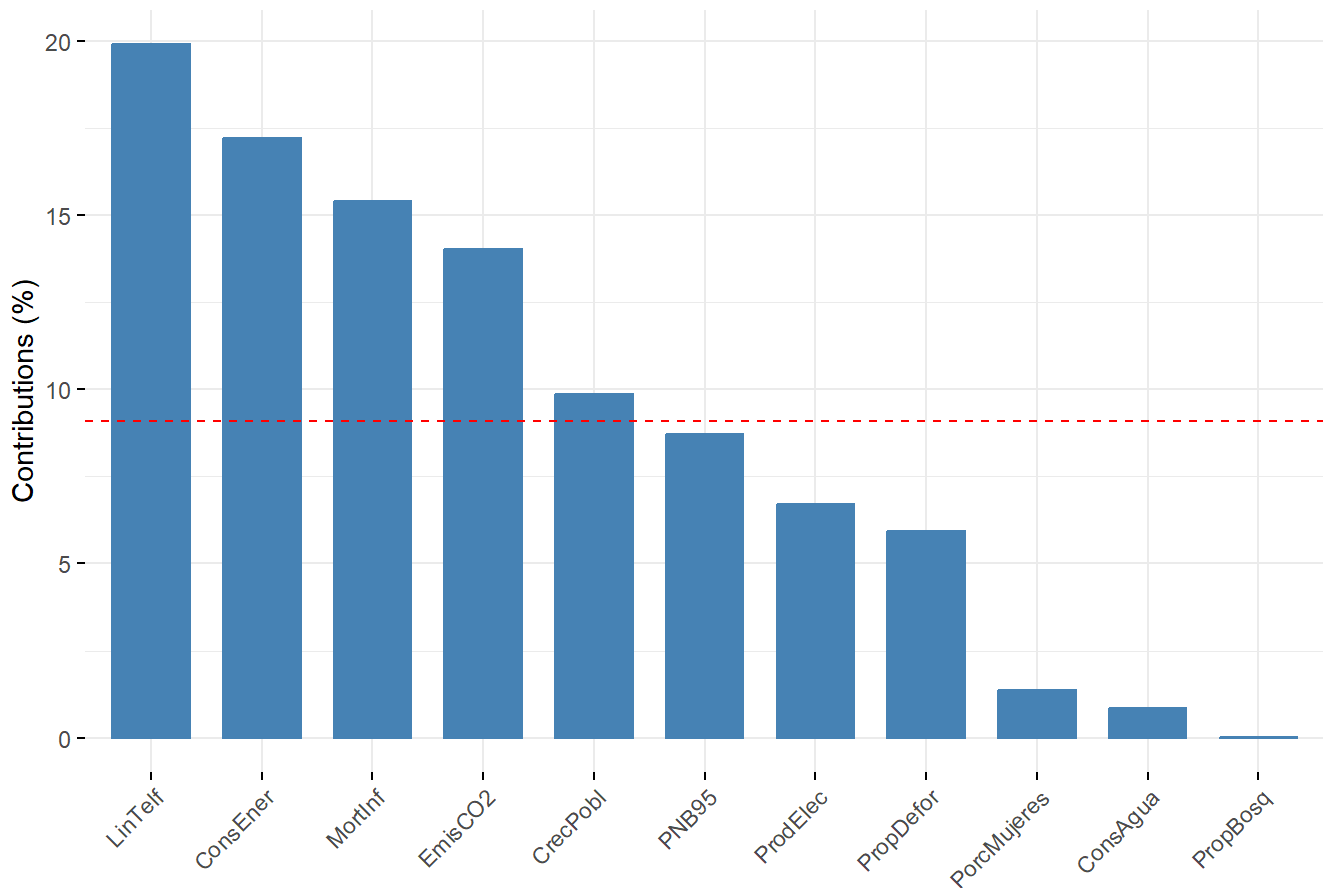
```
fviz_screplot(cpR)
```

Scree plot



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

Contribution of variables to Dim-1



El primer gráfico muestra las proyecciones de los datos sobre los ejes principales 1 y 2 sobre la matriz de correlación. A comparación al gráfico hecho con la matriz de covarianzas, los datos están mejor distribuidos sobre los cuadrantes. Se puede seguir observando casos muy puntuales de datos atípicos, pero su efecto es menor sobre la forma en la que se resume el comportamiento general de los datos

El tercer gráfico es el mismo que el primero con una elipse que contendría el 95% de los datos en el caso de que se asemeje a una distribución normal. Lo mismo ocurre para la gráfica cuatro, donde la elipse contiene el 95% de las variables modeladas.

Las gráficas cinco y seis muestran la contribución de la explicación de la variabilidad por componente principal y la contribución por componente por cada variable. En este caso las variables están mejor distribuidas, siendo necesario más variables y componentes para explicar la variación y aportación para el componente 1.

PARTE IV

Compare los resultados obtenidos con la matriz de varianza-covarianza y con la correlación . ¿Qué concluye? ¿Cuál de los dos procedimientos aporta componentes con de mayor interés?

El método de PCA sobre la matriz de covarianza captura casi toda la variabilidad de los datos con solo dos componentes principales. Sin embargo, este análisis puede estar sesgado si las variables tienen diferentes escalas, ya que la covarianza se ve afectada por la magnitud de las variables. Cuando se trabaja con la matriz de correlación, se pueden tener resultados más adecuados ya que la correlación ya estandariza las variables y permite una comparación de una misma escala. No disminuye la dimensionalidad de la misma forma, pero permite un análisis más concreto de las variables.

Indique cuál de los dos análisis (a partir de la matriz de varianza y covarianza o de correlación) resulta mejor para los datos indicadores económicos y sociales del 96 países en el mundo. Comparar los resultados y argumentar cuál es mejor según los resultados obtenidos.

En este caso, la matriz de correlación es más apropiada para analizar los datos de indicadores económicos y sociales, ya que asegura que las variables están en una escala comparable.

¿Qué variables son las que más contribuyen a la primera y segunda componentes principales del método seleccionado? (observa los coeficientes en valor absoluto de las combinaciones lineales, auxíliate también de los gráficos)

PC1) Línea telefónica, Consumo de Energía per Cápita, Mortalidad Infantil, Emisiones de CO2, Crecimiento de Poblacion.

PC2) Porcentaje de Mujeres, Proporción de Bosques, Crecimiento de Población, Emisiones de CO2, Consumo de Energía.

Escriba las combinaciones finales que se recomiendan para hacer el análisis de componentes principales.

```
cpR=princomp(df,cor=TRUE)
cpR$loadings[,c(1,2)]
```

##	Comp.1	Comp.2
## CrecPobl	0.314119414	0.34835747
## MortInf	0.392395442	-0.04136238
## PorcMujeres	-0.116546319	-0.58283641
## PNB95	-0.295393771	-0.17690839
## ProdElec	-0.258964724	-0.17356372
## LinTelf	-0.446082934	-0.02719077
## ConsAgua	-0.092410503	0.32060987
## PropBosq	-0.005692925	-0.45742697
## PropDeFor	0.243652293	-0.15408201
## ConsEner	-0.415029554	0.23286257
## EmisCO2	-0.374531032	0.29168698

Interpreta los resultados en término de agrupación de variables (puede ayudar “índice de riqueza”, “índice de ruralidad”, etc)

El primer componente podría estar resumiendo desarrollo económico, ya que variables como el consumo de energía per cápita, el número de líneas telefónicas, y las emisiones de CO2 son indicativos de desarrollo industrial y económico. La mortalidad infantil también refleja el nivel de desarrollo, ya que una mayor mortalidad infantil suele asociarse con un menor desarrollo.

El segundo componente podría estar relacionado con la sostenibilidad ambiental. La proporción de bosques y el crecimiento de la población indican factores ambientales y demográficos. Este componente parece capturar la relación entre desarrollo y sostenibilidad, donde un alto porcentaje de bosques y un menor consumo de energía podrían indicar países más rurales o con políticas más sostenibles.