A4 Componentes Principales

Héctor San Román Caraza

18/10/2022

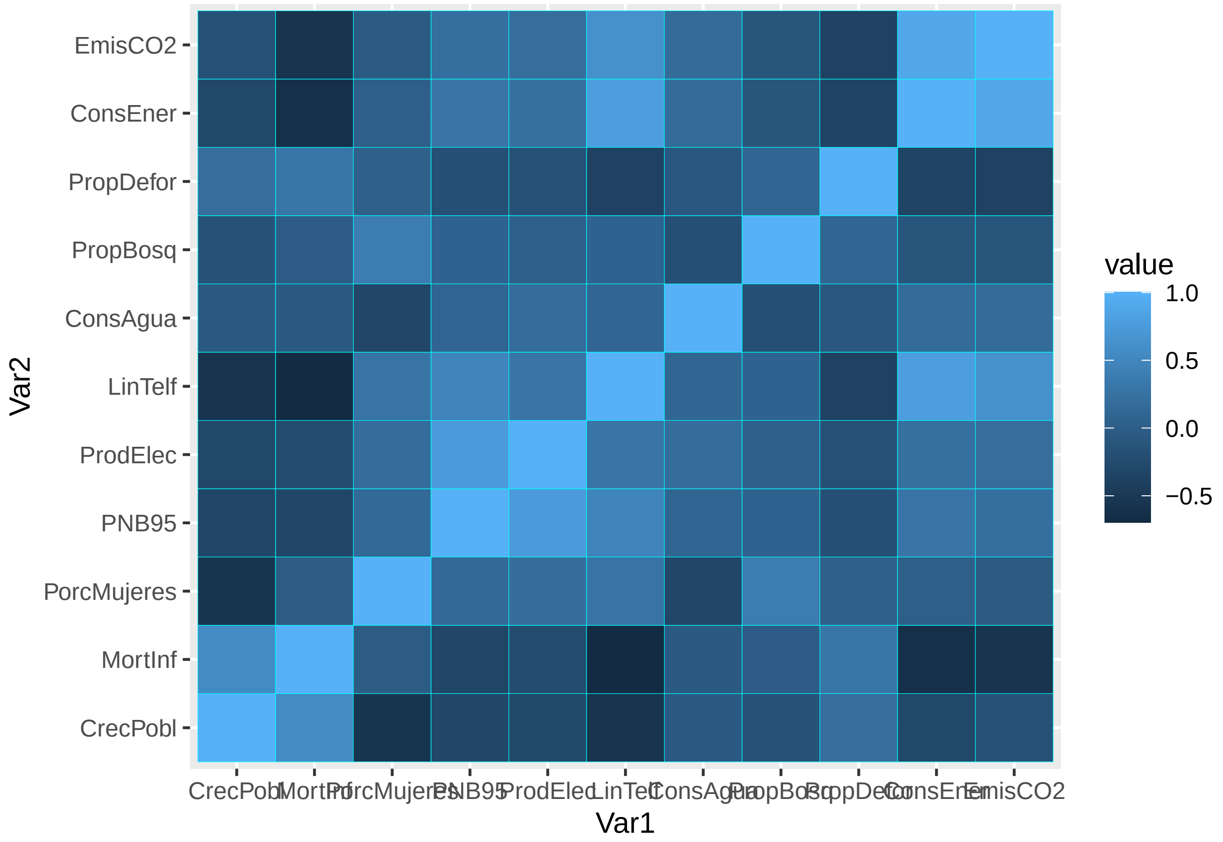
Parte 1

```
setwd("C:/Users/hsrc1/Downloads")
df <- read.csv("paises_mundo.csv")
df
```

```
##    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
## 6       1.4       6          43  337909   167155     510      933       19
## 7       0.6       6          41  216547    53259     465      304       47
## 8       2.0      79          42   28599     9891       2      220        6
## 9       0.3       8          40  250710    72236     457      917       20
## 10      3.0      95          48    2034        6       5       26       45
## 11      0.4      13          49   21356    31397     190      295       31
## 12      2.3      69          37    5905     2824      35      201       45
## 13      1.6      44          35  579787   260682      75      246       66
## 14     -0.6      15          48   11225   381333     335     1544       33
## 15      2.9      56          38    8615     2740       4       38       44
## 16      1.3       6          45  573695   554227     590     1602       49
## 17      1.8      26          37   70263    43354     100      174       52
## 18      3.1      90          43    1784      435       8       20       58
## 19      1.8      26          45   12870    38000      47      687       74
## 20      0.9      10          40  435137   164993     415      632       66
## 21      3.4      86          33    9248     2305       8       66       34
## 22      2.5      13          30    8884     4772     164      780       28
## 23      0.9       9          38    7150    10982      32      870       16
## 24      1.6      12          32   59151    25276     132     1626       12
## 25      1.3      34          45  744890   928083      34      461       13
## 26      0.2       6          46  156027    40097     613      233       12
## 27      2.0      37          29   11390     6182      79      446       22
## 28      2.3      36          26   15997     8256      61      581       43
## 29      2.2      56          29   45507    51947      46      956        0
## 30      1.8      36          34    9057     3211      53      245        6
## 31      5.8      16          13   42806    18870     283      884        0
## 32      0.3      11          48   15848    24740     208      337       38
## 33      0.2       7          36  532347   161654     385      781       51
## 34      2.6     112          41    5722     1293       2       51       13
## 35      2.3      39          37   71865    27062      21      686       26
## 36      0.4       5          48  105174    65546     550      440       77
## 37      0.5       6          44 1451051   476200     558      665       25
## 38      2.9      89          44    3759      933      30       57       71
## 39      3.0      73          51    6719     6115       4       35       42
## 40      0.5       8          36   85885    40623     493      523       47
## 41      2.9      44          26   14255     3161      27      139       39
## 42      2.0      72          43    1777      362       8        7        1
## 43      0.6       6          40  371039    79647     525      518       10
## 44      3.0      45          30    3566     2672      29      294       41
## 45     -0.3      11          44   42129    33486     185      661       18
## 46      1.9      68          32  319660   386500      13      612       17
## 47      1.7      51          40  190105    53414      17       96       60
## 48      2.7     108          18   24600    27060      33     4575        4
## 49      3.2      45          24  113400    79128      79     1362       11
## 50      0.1       6          33   52765    17105     365      233        6
## 51      1.1       4          44    6686     4780     555      636        1
## 52      2.7       8          40   87875    32781     418      408        6
## 53      4.7      31          21    6354     5076      73      173        1
## 54      2.9      58          46    7583     3539       9       87        2
## 55     -0.3      11          28   28941    22798     230      525        0
## 56      2.3      32          28   10673     5184      82      271        8
## 57      3.6      61          21   23400    17800      59      880        0
## 58      2.5      12          37   78321    39093     166      768       54
## 59      2.0      55          35   29545    11100      43      427       20
## 60      2.1      33          31  304596   147926      96      899       25
## 61      1.8     113          48    1353      490       3       55       22
## 62      1.8      83          43   35840     3500       3      101       44
## 63      2.5      91          40    4391      927       4      150       37
## 64      3.1      46          36    1659     1688      23      367       50
## 65      2.9      80          36   28411    15530       4       41       17
## 66      0.5       5          46  136077   113488     556      488       31
## 67      1.0       7          44   51655    35135     479      589       28
## 68      4.5      18          15   10578     6187      77      564       19
## 69      3.0      90          26   59991    58529      16     2053        2
## 70      1.9      23          34    7253     3380     114      754       42
## 71      2.7      41          29    8158    36415      31      109       32
## 72      2.1      47          29   55019    15563      47      300       53
## 73      0.4      14          46  107829   135347     148      321       28
## 74     -0.1       7          43   96829    31380     361      739       34
## 75      0.3       6          43 1094734   325383     502      205       10
## 76      0.0       8          47   39990    58705     236      266       34
## 77      0.0      23          44   33488    55136     131     1134       27
## 78      2.8      62          42    5070     1002      10      202       39
## 79      1.8       4          38   79831    20046     478       84        7
## 80      3.1      32          26   15780    15186      63      435        4
## 81      1.3      16          35   12616     4387      11      503       27
## 82      2.2      77          28    7510     1333       3      633       18
## 83      0.6       4          48  209720   142895     681      341       68
## 84      0.8       6          40  286014    65724     613      173       30
## 85      2.3      50          37  130918   189316      95      359        4
## 86      1.3      35          46  159630    71177      59      602       25
## 87      3.1      82          49    3703     1913       3       40       38
## 88      2.1      39          30   16369     6714      58      381        4
## 89      1.9      48          35  169452    78322     212      585       26
## 90      0.1      15          49   84084   202995     157      673       16
## 91      0.6      18          40   16458     7617     196      241        4
## 92      2.4      23          33   65382    73116     111      382       52
## 93      2.2      41          49   17634    12270      11      414       26
## 94      4.2     100          29    4044     2159      12      335        8
## 95      2.6     109          45    3605     7785       8      186       43
## 96      2.8      55          44    5933     7334      14      136       23
##    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
## 6        0.0     5341    15.3
## 7       -0.4     3301     7.2
## 8        4.1       64     0.2
## 9       -0.3     5120    10.1
## 10       1.3       20     0.1
## 11      -0.4     2392     9.9
## 12       1.2      373     1.0
## 13       0.6      718     1.4
## 14      -0.2     2438     6.4
## 15       0.6      103     0.2
## 16      -1.1     7854    14.4
## 17       0.7      622     1.8
## 18       0.2      331     1.6
## 19       0.0     1129    11.2
## 20       0.1     2982     6.6
## 21       1.0      103     0.5
## 22       3.0      558     1.2
## 23       1.0      923     2.6
## 24      -0.1     1012     2.6
## 25       0.7      664     2.3
## 26       0.0     3977    10.4
## 27       2.9      337     1.4
## 28       1.8      565     1.8
## 29       0.0      600     1.5
## 30       2.3      370     0.7
## 31       0.0    10531    33.9
## 32       0.1     3243     7.0
## 33       0.0     2458     5.7
## 34       0.3       22     0.1
## 35       3.4      316     0.8
## 36       0.0     5997     8.2
## 37      -0.1     4042     6.3
## 38       0.6      652     5.5
## 39       1.4       93     0.2
## 40       0.0     2260     7.2
## 41       1.8      210     0.6
## 42       5.1       29     0.1
## 43      -0.3     4580     9.2
## 44       2.2      204     0.6
## 45      -0.5     2383     5.8
## 46       0.6      248     0.9
## 47       1.1      366     1.0
## 48       0.1     1213     3.4
## 49       0.0     1505     4.0
## 50      -1.2     3137     8.7
## 51       0.0     7932     6.8
## 52      -0.3     2717     8.1
## 53      -1.0     1067     3.0
## 54       0.6      110     0.2
## 55       0.0     8622    11.2
## 56       0.6      964     2.9
## 57      -1.4     2499     8.1
## 58       2.1     1699     3.8
## 59      -1.4      327     1.1
## 60       1.3     1561     3.8
## 61       0.8       40     0.1
## 62       1.3       49     0.1
## 63       1.0       28     0.1
## 64       1.9      300     0.6
## 65       0.7      162     0.9
## 66      -1.4     5318    14.1
## 67       0.0     4245     7.6
## 68       0.0     2392     5.3
## 69       3.5      254     0.6
## 70       1.9      618     1.7
## 71       2.8      299     0.6
## 72       0.4      367     1.0
## 73      -0.1     2401     8.9
## 74      -0.5     1827     4.8
## 75      -1.1     3732     9.8
## 76       0.0     3868    13.1
## 77       0.0     1733     5.4
## 78       0.7       97     0.4
## 79       2.3     8103    17.7
## 80      -4.3      997     3.3
## 81       1.4       97     0.3
## 82       1.1       66     0.1
## 83       0.0     5723     6.6
## 84      -0.6     3629     6.4
## 85      -0.8     2146     7.5
## 86       3.5      769     2.0
## 87       1.2       34     0.1
## 88      -1.9      595     1.6
## 89       0.0      957     2.5
## 90      -0.3     3180    11.7
## 91      -0.6      629     1.6
## 92       1.2     2186     5.7
## 93       1.5      101     0.3
## 94       0.0      206     0.7
## 95       1.1      149     0.3
## 96       0.7      438     1.8
```

Matriz Correlación

```
cormat = round(cor(df),4)
library(reshape2)
melted_cormat <- melt(cormat)

library(ggplot2)
ggplot(data = melted_cormat, aes(x=Var1, y=Var2, fill=value)) +
  geom_tile(color="cyan")
```

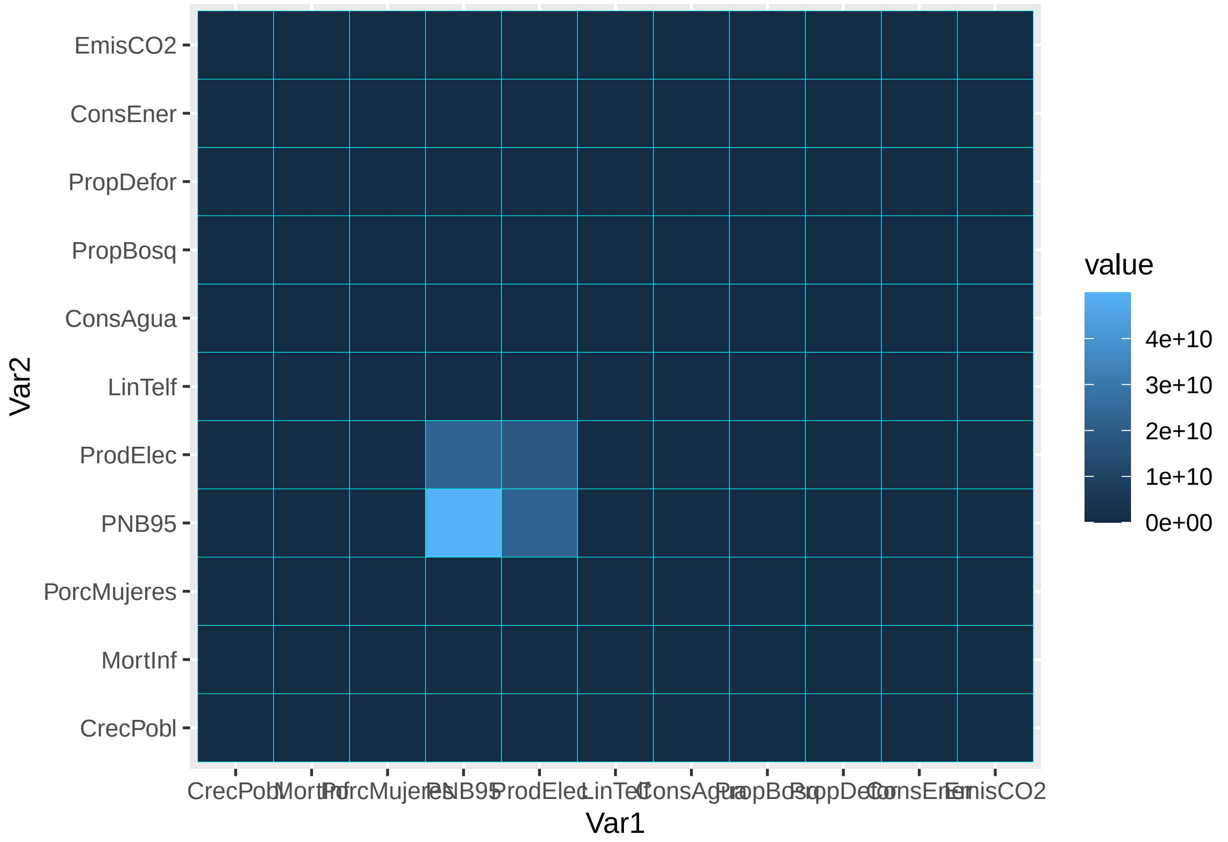


```
round(cor(df),2)
```

```
##             CrecPobl MortInf PorcMujeres PNB95 ProdElec LinTelf ConsAgua
## CrecPobl        1.00    0.55       -0.56 -0.32    -0.30   -0.56    -0.07
## MortInf         0.55    1.00       -0.03 -0.32    -0.24   -0.70    -0.07
## PorcMujeres    -0.56   -0.03        1.00  0.14     0.19    0.26    -0.31
## PNB95          -0.32   -0.32        0.14  1.00     0.74    0.47     0.09
## ProdElec       -0.30   -0.24        0.19  0.74     1.00    0.29     0.18
## LinTelf        -0.56   -0.70        0.26  0.47     0.29    1.00     0.11
## ConsAgua       -0.07   -0.07       -0.31  0.09     0.18    0.11     1.00
## PropBosq       -0.16   -0.02        0.37  0.06     0.03    0.06    -0.19
## PropDefor       0.20    0.29        0.02 -0.19    -0.17   -0.38    -0.09
## ConsEner       -0.30   -0.62        0.01  0.28     0.23    0.78     0.16
## EmisCO2        -0.18   -0.57       -0.05  0.21     0.20    0.62     0.16
##             PropBosq PropDefor ConsEner EmisCO2
## CrecPobl       -0.16      0.20    -0.30   -0.18
## MortInf        -0.02      0.29    -0.62   -0.57
## PorcMujeres      0.37      0.02      0.01    -0.05
## PNB95            0.06     -0.19      0.28     0.21
## ProdElec         0.03     -0.17      0.23     0.20
## LinTelf          0.06     -0.38      0.78     0.62
## ConsAgua        -0.19     -0.09      0.16     0.16
## PropBosq         1.00      0.10     -0.11    -0.12
## PropDefor        0.10      1.00     -0.35    -0.37
## ConsEner        -0.11     -0.35      1.00     0.88
## EmisCO2         -0.12     -0.37      0.88     1.00
```

Matriz covarianzas

```
covmat = round(cov(df),4)
library(reshape2)
melted_covmat <- melt(covmat)

library(ggplot2)
ggplot(data = melted_covmat, aes(x=Var1, y=Var2, fill=value)) +
  geom_tile(color="cyan")
```



```
round(cov(df),2)
```

```
##                 CrecPobl      MortInf PorcMujeres        PNB95       ProdElec
## CrecPobl            1.54        21.95       -6.08 -8.933379e+04 -4.973964e+04
## MortInf            21.95      1032.86       -9.25 -2.269332e+06 -1.043435e+06
## PorcMujeres        -6.08        -9.25       76.98  2.813114e+05  2.260248e+05
## PNB95          -89333.79 -2269332.29   281311.42  4.999786e+10  2.247791e+10
## ProdElec    -49739.64 -1043435.44   226024.81  2.247791e+10  1.821909e+10
## LinTelf       -136.91    -4381.37      449.98  2.039550e+07  7.583050e+06
## ConsAgua       -48.27    -1288.21    -1568.31  1.097481e+07  1.399817e+07
## PropBosq        -3.89      -14.66       65.18  2.474311e+05  7.035979e+04
## PropDefor        0.34       12.76        0.27 -5.806203e+04 -3.180340e+04
## ConsEner      -838.42   -44425.68      285.52  1.415628e+08  6.801296e+07
## EmisCO2         -1.14      -94.86       -2.15  2.501673e+05  1.392779e+05
##                 LinTelf    ConsAgua   PropBosq PropDefor     ConsEner   EmisCO2
## CrecPobl        -136.91      -48.27      -3.89      0.34      -838.42     -1.14
## MortInf        -4381.37    -1288.21     -14.66     12.76    -44425.68    -94.86
## PorcMujeres      449.98    -1568.31      65.18      0.27       285.52     -2.15
## PNB95       20395504.89 10974812.99 247431.12 -58062.03 141562781.13 250167.32
## ProdElec     7583050.03 13998171.50  70359.79 -31803.40  68012960.42 139277.89
## LinTelf        38412.47    11931.10     248.72    -99.40    342626.21    638.57
## ConsAgua       11931.10   330198.07   -2220.76    -67.44    209224.16    486.93
## PropBosq         248.72    -2220.76     401.00      2.63     -5153.44    -12.90
## PropDefor        -99.40      -67.44       2.63      1.82     -1051.52     -2.63
## ConsEner      342626.21   209224.16   -5153.44  -1051.52   5014395.13  10286.16
## EmisCO2          638.57      486.93     -12.90     -2.63     10286.16     27.27
```

Eigen valores y vectores

```
# Covarianza

EVal_Cov <- eigen(covmat)$values
EVec_Cov <- eigen(covmat)$vectors

# Correlación

EVal_Cor <- eigen(cormat)$values
EVec_Cor <- eigen(cormat)$vectors
```

Varianza por cada componente

```
var_t <- sum(diag(covmat))
var_t
```

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

```
listaV <- list()

for (i in 1:length(EVal_Cov)) {
  Variacion <- EVal_Cov[i] / var_t
  cat("Variación total explicada por variable " , i, " : ", Variacion, "\n")
  listaV <- append(listaV, Variacion)
}
```

```
## Variación total explicada por variable  1  :  0.9034543
## Variación total explicada por variable  2  :  0.09647298
## Variación total explicada por variable  3  :  6.795804e-05
## Variación total explicada por variable  4  :  4.554567e-06
## Variación total explicada por variable  5  :  1.782429e-07
## Variación total explicada por variable  6  :  7.530917e-09
## Variación total explicada por variable  7  :  5.317738e-09
## Variación total explicada por variable  8  :  6.65776e-10
## Variación total explicada por variable  9  :  8.50287e-11
## Variación total explicada por variable  10  :  2.107903e-11
## Variación total explicada por variable  11  :  6.989131e-12
```

Acumule los resultados anteriores

```
cumsum(listaV)
```

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

5. Según los resultados anteriores, ¿qué componentes son los más importantes? ¿qué variables son las que más contribuyen a la primera y segunda componentes principales? ¿por qué lo dice? ¿influyen las unidades de las variables?

Con lo que podemos observar de los resultados, la mayor variación de los componentes se puede ver explicada en el primero, con el 90% de la variabilidad. El segundo componente contiene tan solo el 9% de la varianza. Después de eso la varianza ya no está representada en gran proporción. Las unidades de las variables sí influyen dentro de la relación por lo que es importante hacer un reescalamiento de datos para no sesgar el resultado.

6. Hacer los mismos pasos anteriores, pero con la matriz de correlaciones (se obtiene con cor(x) si x está compuesto por variables numéricas)

Varianza por cada componente

```
var_t <- sum(diag(cormat))
var_t
```

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

```
listaV <- list()

for (i in 1:length(EVal_Cor)) {
  Variacion <- EVal_Cor[i] / var_t
  cat("Variación total explicada por variable " , i, " : ", Variacion, "\n")
  listaV <- append(listaV, Variacion)
}
```

```
## Variación total explicada por variable  1  :  0.3663526
## Variación total explicada por variable  2  :  0.1754581
## Variación total explicada por variable  3  :  0.124587
## Variación total explicada por variable  4  :  0.07858922
## Variación total explicada por variable  5  :  0.07219864
## Variación total explicada por variable  6  :  0.06628812
## Variación total explicada por variable  7  :  0.05193783
## Variación total explicada por variable  8  :  0.02970794
## Variación total explicada por variable  9  :  0.01528042
## Variación total explicada por variable  10  :  0.01330355
## Variación total explicada por variable  11  :  0.006296621
```

Acumule los resultados anteriores

```
cumsum(listaV)
```

```
##  [1] 0.3663526 0.5418107 0.6663977 0.7449869 0.8171855 0.8834736 0.9354115
##  [8] 0.9651194 0.9803998 0.9937034 1.0000000
```

7. Compare los resultados de los incisos 6 y 7. ¿qué concluye?

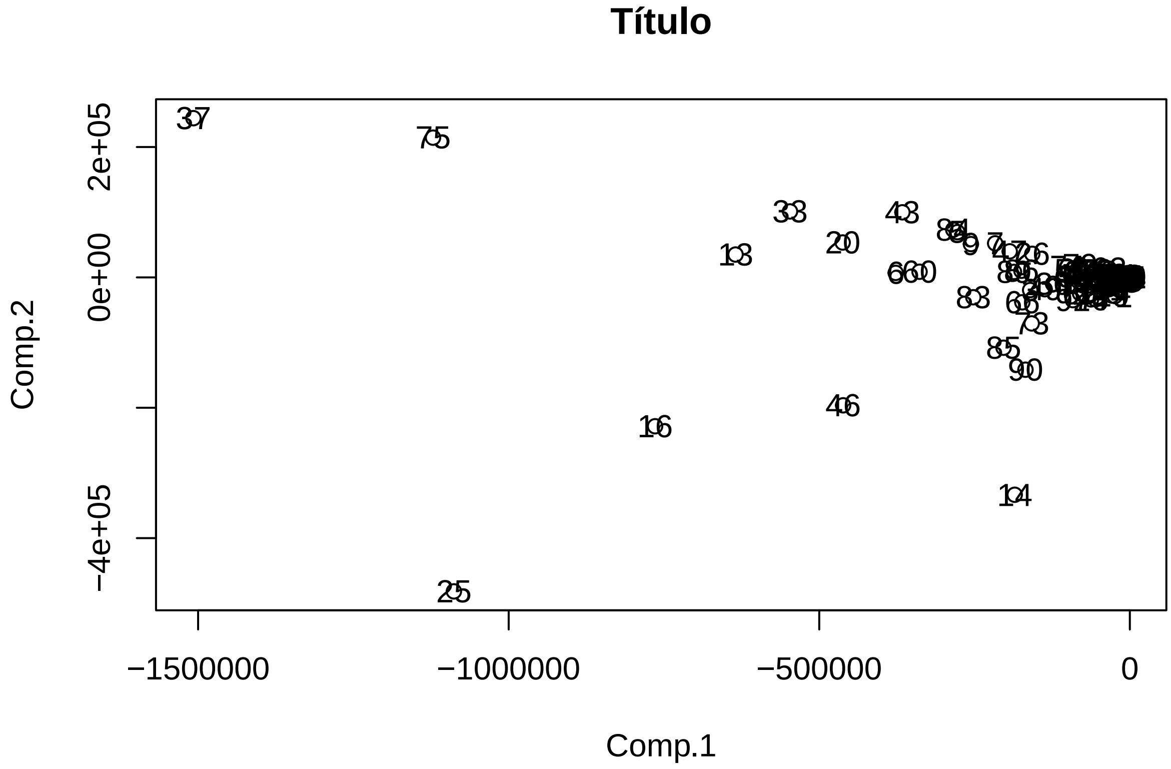
Vemos que la variabilidad de los componentes ya es más dispersa. En el acummulado necesitamos 7 componentes para representar el 90% de la variabilidad. La matriz de correlación se comporta de forma diferente, distribuyendo la variabilidad de los componentes.

Parte II

```
library(stats)
library(factoextra)
```

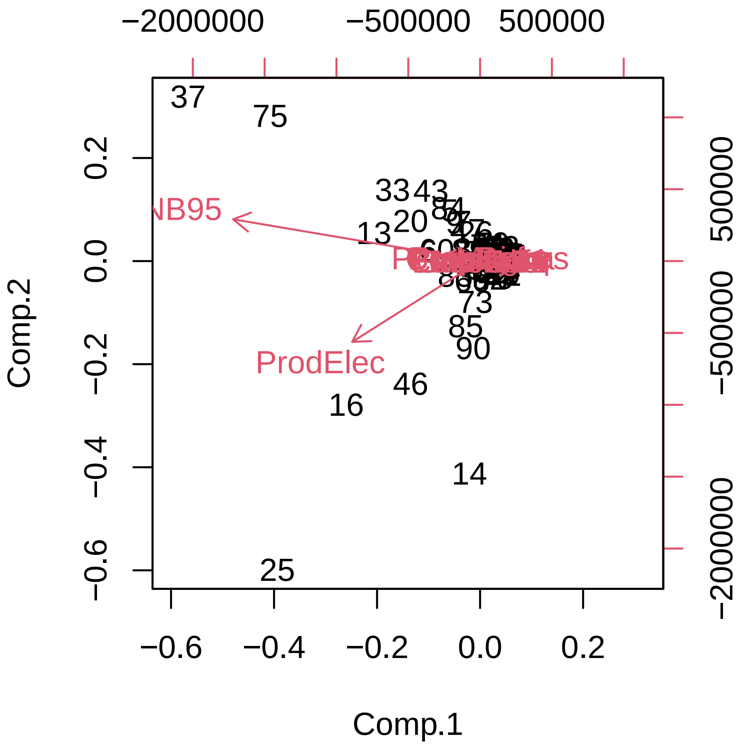
```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```
library(ggplot2)
datos= df
cpS=princomp(datos,cor=FALSE)
cpaS=as.matrix(datos)%*%cpS$loadings
plot(cpaS[,1:2],type="p", main = "Título")
text(cpaS[,1],cpaS[,2],1:nrow(cpaS))
```



```
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
```



Aquí tenemos una representación gráfica de como las variables afectan a lso componentes principales. La mayor parte dde las variables se encuentran en el (0,0) redondeando. Vemos que existen algunas variables que destacan en variabilidad hacia distintos componentes, haciendonos ver cuales se compoortan de forma similar.

Parte III

```
library(FactoMineR)
```

```
## Warning: package 'FactoMineR' was built under R version 4.1.3
```

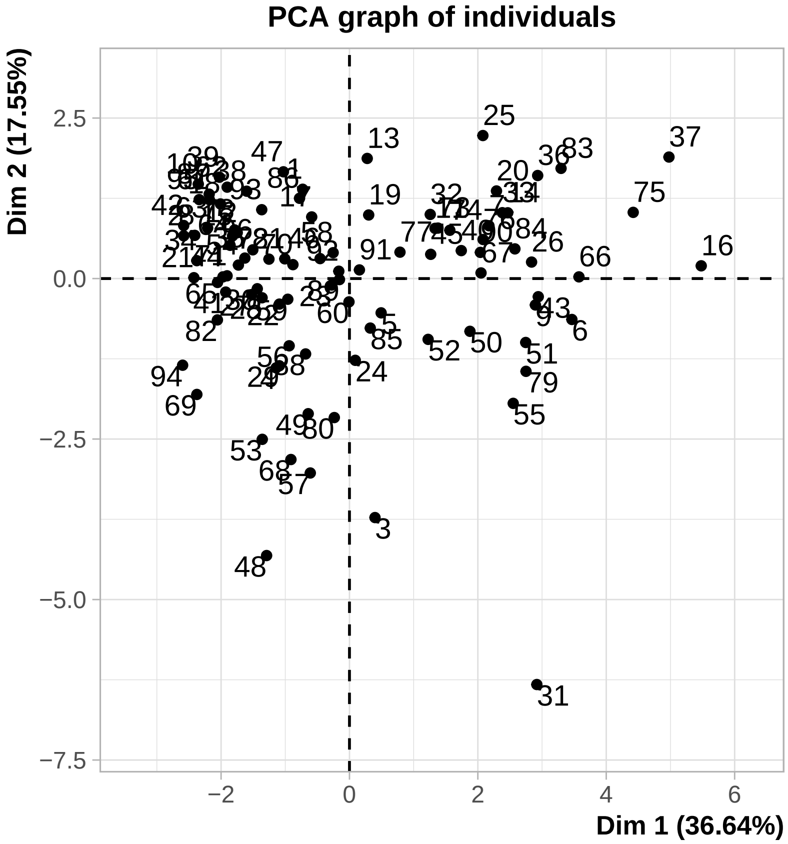
```
library(factoextra)
library(ggplot2)
library(dplyr)
```

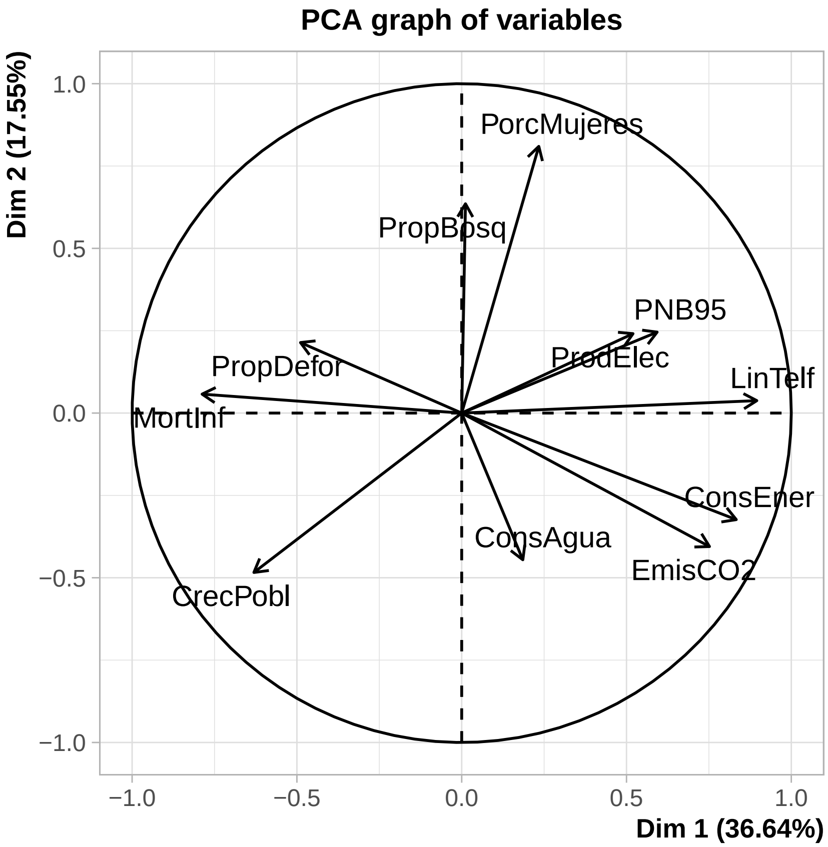
```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

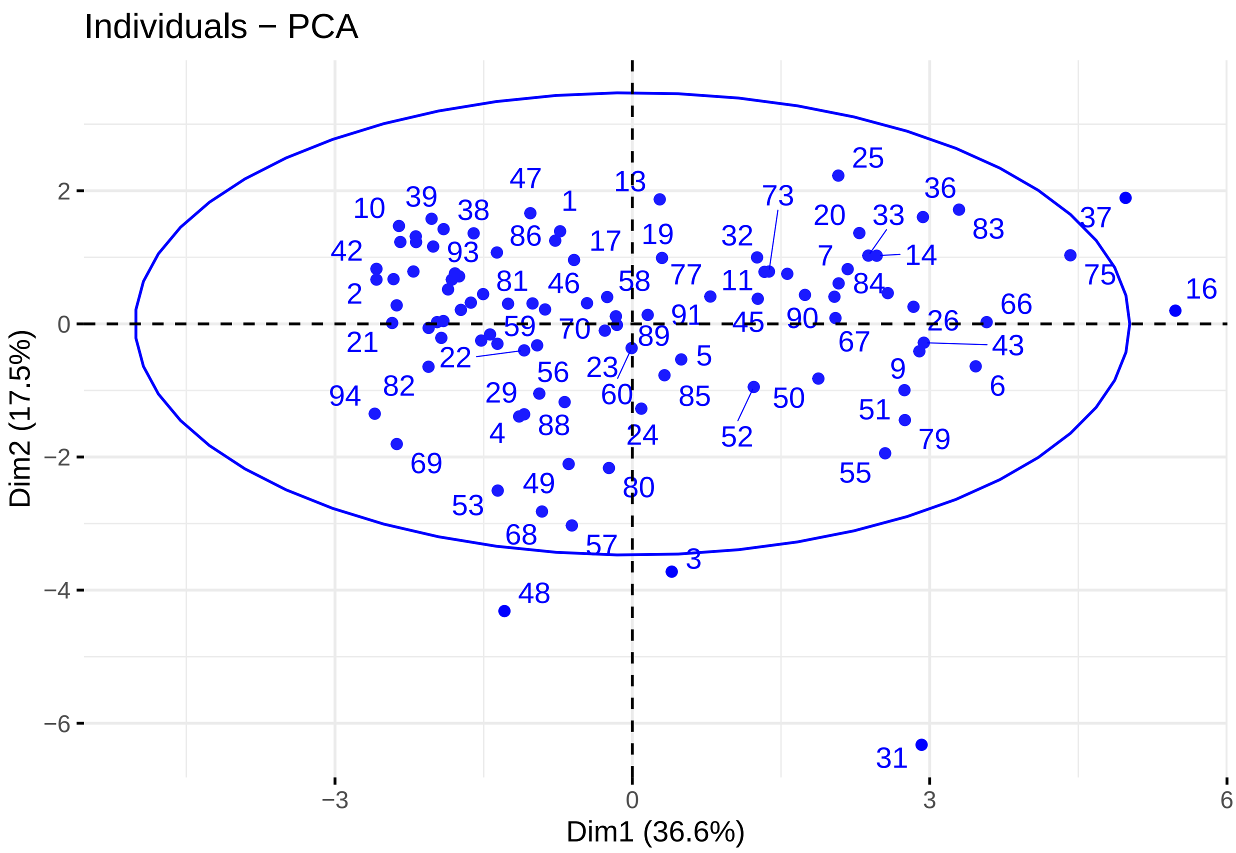
```
datos=df
cp3 = PCA(datos)
```



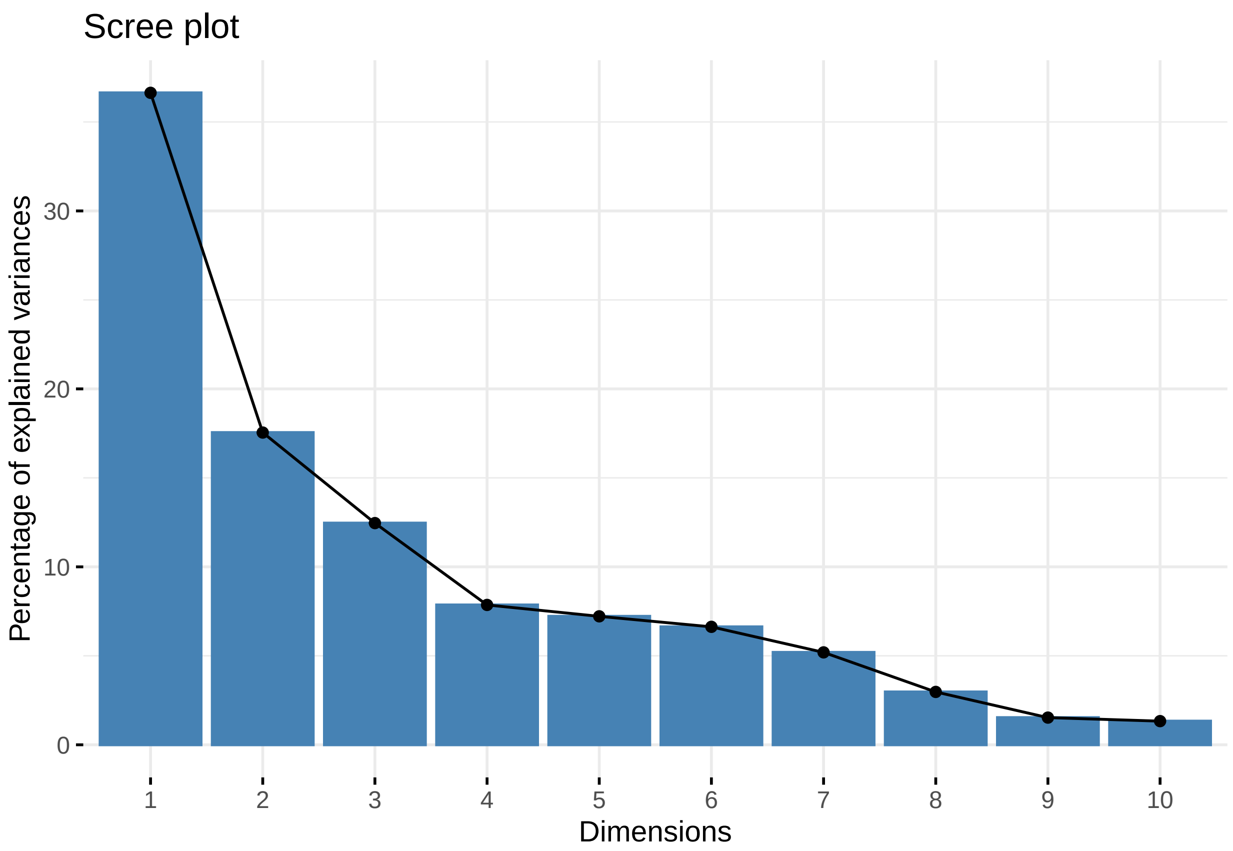


```
fviz_pca_ind(cp3, col.ind = "blue", addEllipses = TRUE, repel = TRUE)
```

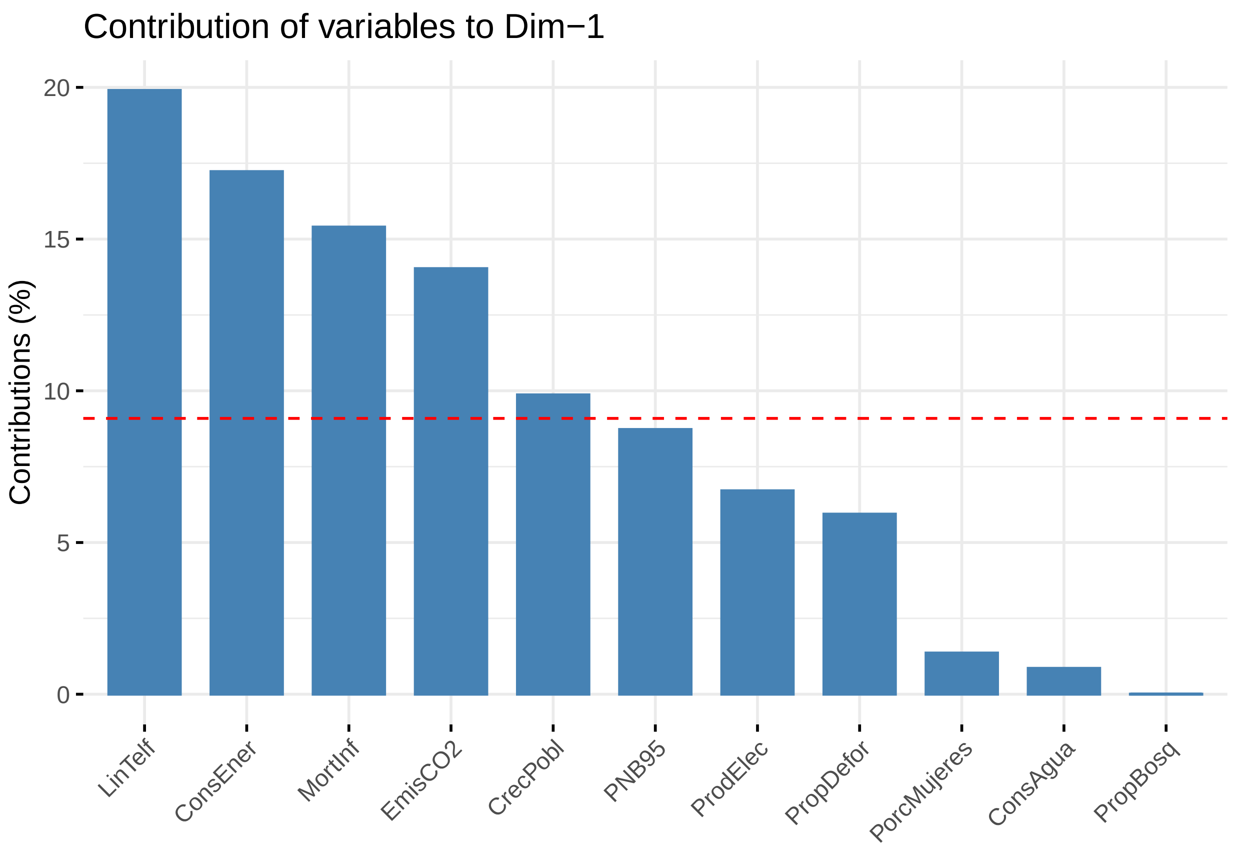
```
## Warning: ggrepel: 27 unlabeled data points (too many overlaps). Consider
## increasing max.overlaps
```



```
fviz_screeplot(cp3)
```



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



En la primera grafica vemos como las variables afectan a los componentes o las dimensiones. Aquí podemos ver como interactuan las variables entre sí. La segunda gráfica es una representación similar solo que es una representación en circulo, normalizada, ayudandonos a interpretar en igualdad las variables. La tercera grafica es una mezcla entre la primera y segunda gráfica, dandonos la representación no normalizada, pero con una ayuda que nos hace el mismo trabajo. Los últimos histogramas nos ayudan a verlo de una forma de barras.