A4 - Componentes Principales(2)

Russel Rosique

2022-11-14

Primero que nada se lee el archivo de paises mundo.

```
df = read.csv("paises_mundo.csv")
```

A continuación se encuentra la matriz de covarianza y correlación del conjunto de datos, para las cuales se encuentran los eigenvalues.

```
dfcov = cov(df)
dfcor = cor(df)
```

```
v1values = eigen(dfcov)$values
v1vectors = eigen(dfcov)$vectors

v2values = eigen(dfcor)$values
v2vectors = eigen(dfcor)$vectors
```

Para obtener la varianza total de los datos se obtiene la suma de la diagonal de la matriz de covarianza. Para comprobar que este valor es correcto se obtiene también la suma de los eigenvalues de la matriz de covarianza. Dado que el resultado es el mismo, se analiza que es correcto y se puede decir con certeza que esa es la varianza total de los datos.

```
vartotal = sum(diag(dfcov))
vartotal
```

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

```
sum(v1values)
```

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

Para encontrar la proporción de la varianza que le pertenece a cada componente, se divide cada eigenvalue entre la varianza total y se obtienen los siguientes resultados.

```
v1values[1]/vartotal
```

```
## [1] 0.9034543
```

```
v1values[2]/vartotal
```

```
## [1] 0.09647298
```

```
v1values[3]/vartotal
```

```
## [1] 6.795804e-05
```

```
v1values[4]/vartotal
```

```
## [1] 4.554567e-06
```

```
v1values[5]/vartotal
```

```
## [1] 1.782429e-07
```

Posteriormente se siguen los mismos pasos pero ahora utilizando la matriz de correlación. Tras realizar lo mismo hecho anteriormente, se analiza que los resultados finales no son significativos ya que para que este proceso sea útil debe ser hecho con la matriz de covarianza.

```
cortotal = sum(diag(dfcor))
cortotal
```

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

```
v2values[1]/cortotal
```

```
## [1] 0.3663526
```

```
v2values[2]/cortotal
```

```
## [1] 0.1754538
```

```
v2values[3]/cortotal
```

```
## [1] 0.1245828
```

```
v2values[4]/cortotal
```

```
## [1] 0.07859236
```

```
v2values[5]/cortotal
```

```
## [1] 0.0721946
```

```
v2values[6]/cortotal
```

```
## [1] 0.06629091
```

```
v2values[7]/cortotal
```

```
## [1] 0.05193683
```

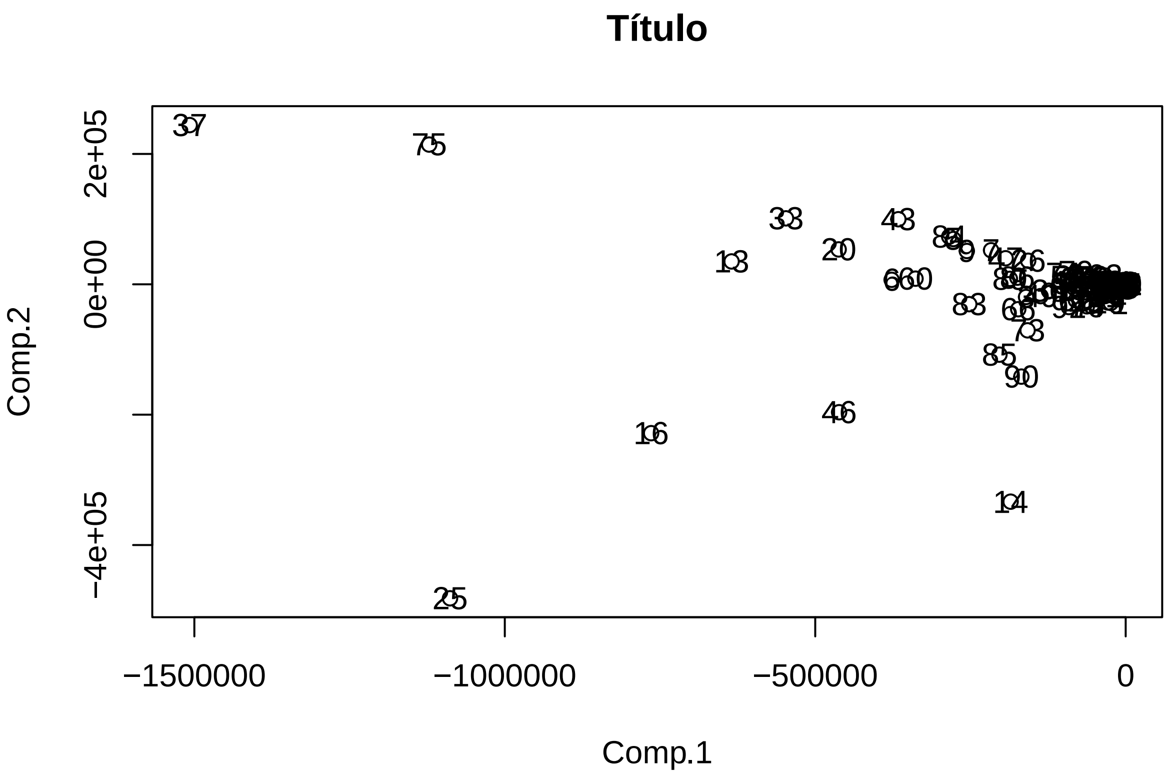
```
library(stats)
library(factoextra)
```

```
## Loading required package: ggplot2
```

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

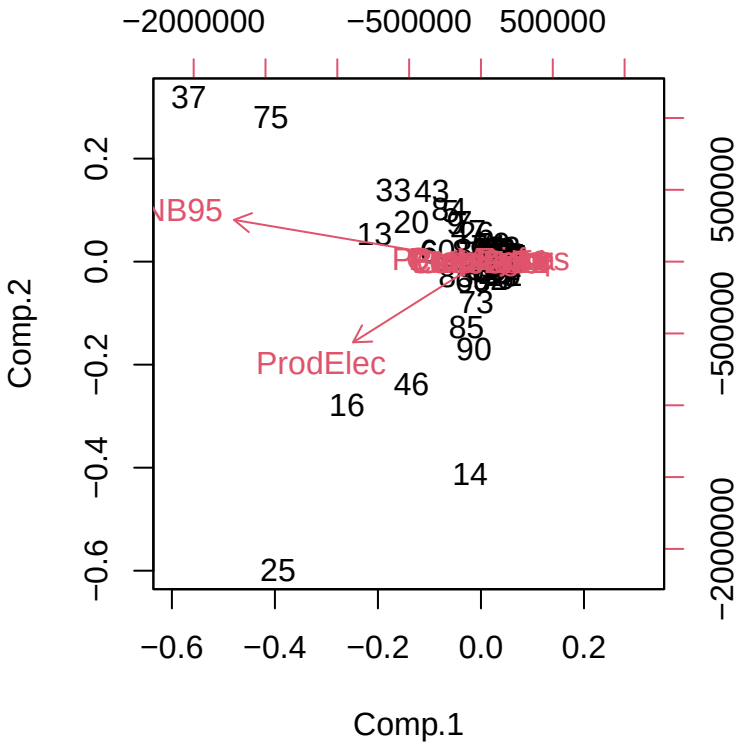
```
library(ggplot2)
```

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



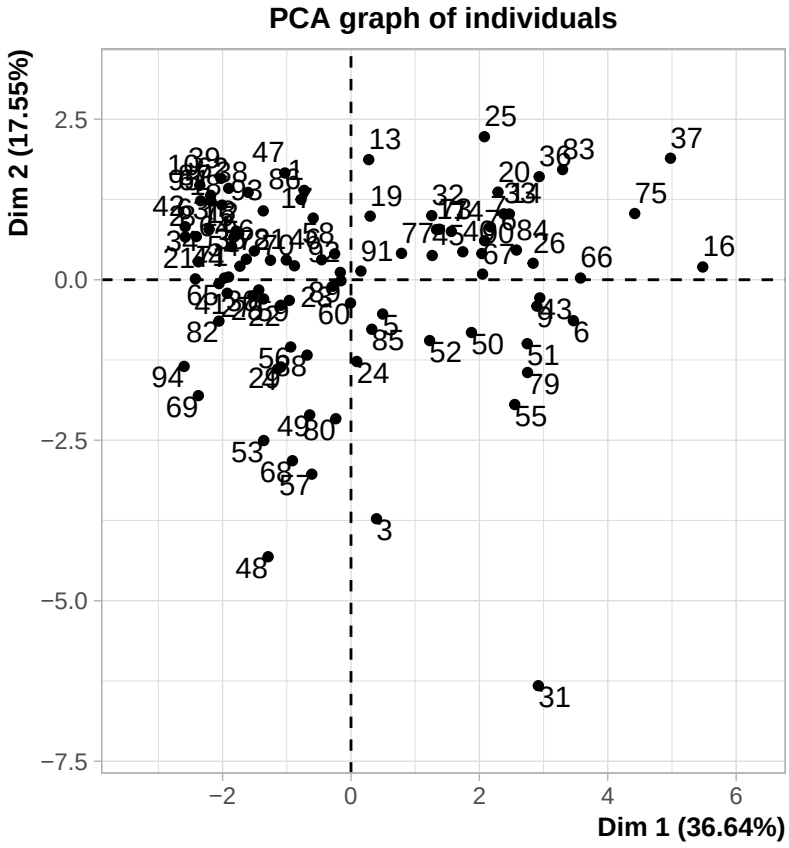

```
biplot(cpS)
```

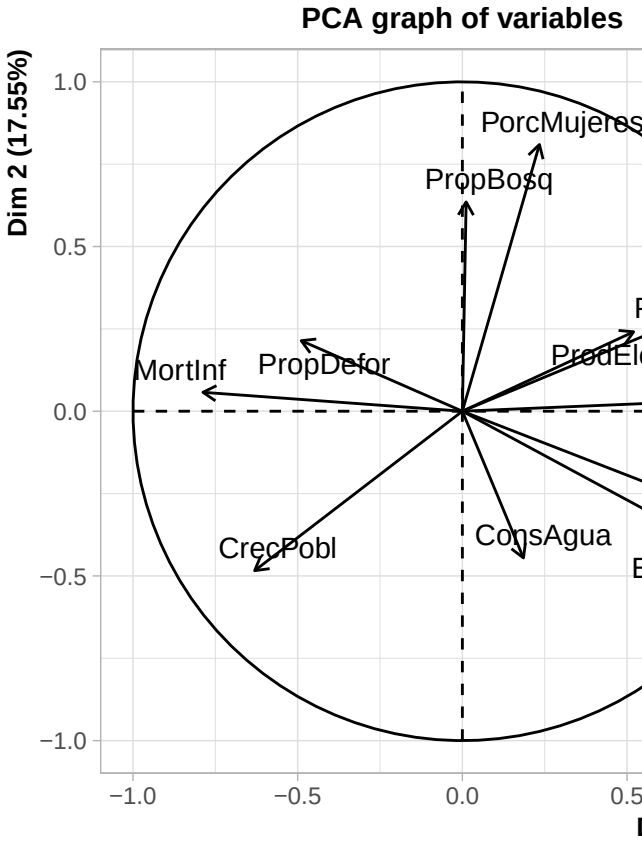
```
## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped

## 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
```




```
library(FactoMineR)
```

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





```r
cpa <- prcomp(df, scale=TRUE)
summary(cpa)
```

```
## Importance of components:
##                           PC1    PC2    PC3     PC4     PC5     PC6     PC7
## Standard deviation     2.0075 1.3892 1.1706 0.92979 0.89115 0.85393 0.75585
## Proportion of Variance 0.3664 0.1754 0.1246 0.07859 0.07219 0.06629 0.05194
## Cumulative Proportion  0.3664 0.5418 0.6664 0.74498 0.81718 0.88347 0.93540
##                           PC8     PC9   PC10    PC11
## Standard deviation     0.57167 0.40996 0.3825 0.26336
## Proportion of Variance 0.02971 0.01528 0.0133 0.00631
## Cumulative Proportion  0.96511 0.98039 0.9937 1.00000
```