

# A2

José Romo - A01197772

2023-11-24

```
# instalacion de libreria
install.packages("stats")

## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)

## Warning: package 'stats' is a base package, and should not be updated

install.packages("factoextra")

## Installing package into '/cloud/lib/x86_64-pc-linux-gnu-library/4.3'
## (as 'lib' is unspecified)

# lectura de archivo
X <- read.csv("países_mundo.csv")
```

## Parte I

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

1. Calcule las matrices de varianza-covarianza  $S$  con `cov(X)` y la matriz de correlaciones con `cor(X)`

```
# Calcular la matriz de varianza-covarianza
matriz_cov <- cov(X)

# Calcular la matriz de correlaciones
matriz_corr <- cor(X)
```

2. Calcule los valores y vectores propios de cada matriz. La función en R es: `eigen()`

```
# Calcular los valores y vectores propios para la matriz de varianza-covarianza
resultados_cov <- eigen(matriz_cov)

# Calcular los valores y vectores propios para la matriz de correlaciones
resultados_corr <- eigen(matriz_corr)
```

3. Calcule la proporción de varianza explicada por cada componente. Se sugiere dividir cada lambda entre la varianza total (las lambdas están en `eigen(S)[1]`). La varianza total es la suma de las varianzas de la diagonal de S. Una forma es `sum(diag(S))`. La varianza total de los componentes es la suma de los valores propios (es decir, la suma de la varianza de cada componente), sin embargo, si sumas la diagonal de S (es decir, la varianza de cada x), te da el mismo valor (¡compruébalo!). Recuerda que las combinaciones lineales buscan reproducir la varianza de X.

```
# Proporción de varianza explicada por cada componente para la matriz de varianza-covarianza
prop_var_cov <- resultados_cov$values / sum(diag(matriz_cov))

# Proporción de varianza explicada por cada componente para la matriz de correlaciones
prop_var_corr <- resultados_corr$values / sum(diag(matriz_corr))
```

4. Acumule los resultados anteriores. (`cumsum()` puede servirle).

```
# Acumular las proporciones de varianza explicada
prop_acum_cov <- cumsum(prop_var_cov)
prop_acum_corr <- cumsum(prop_var_corr)

# Imprimir resultados
print("Proporciones de varianza explicada para la matriz de varianza-covarianza:")

## [1] "Proporciones de varianza explicada para la matriz de varianza-covarianza:"
print(prop_var_cov)

## [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

print("Proporciones de varianza explicada para la matriz de correlaciones:")

## [1] "Proporciones de varianza explicada para la matriz de correlaciones:"
print(prop_var_corr)

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

print("Proporciones acumuladas de varianza explicada para la matriz de varianza-covarianza:")

## [1] "Proporciones acumuladas de varianza explicada para la matriz de varianza-covarianza:"
print(prop_acum_cov)

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

print("Proporciones acumuladas de varianza explicada para la matriz de correlaciones:")

## [1] "Proporciones acumuladas de varianza explicada para la matriz de correlaciones:"
print(prop_acum_corr)

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

En terminos de importancia, la primera componente es significativamente mas importantw que las siguientes. Las variables que mas contribuyen a la primera componente son aquellas que tienen una mayor varianza en la matriz de varianza-covarianza

Las proporciones de varianza explicada son significativamente mas bajas. La interpretacion basada en la matriz de correlaciones esta relacionada con la correlacion entre las variables, no solo con su varianza. Esto puede llevar a una distribucion mas uniforme de la importancia entre las variables

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

Los resultados son los mismos, ya que el analisis de componenetes principales basado en la matriz de correlaciones y en la matriz de varianza-covarianza es equivalente cuando se trata de variables numericas.

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

Los resultados de los incisos 6 y 7 son idenicos, ya que la matriz de correlacion es simplemente una version escalada de la matriz de varianza-covarianza

En resumen, al realizar el analisis de componentes principales, la eleccion entre utilizar entre una matriz u otra depende plenamente de los objetivos especificos del analisis y la interpretacion que se busca. Una es util cuando se quiere explorar las relaciones lineales entre las variables, y la otra para analizar la variabilidad absoluta de las variables/

## Parte 2

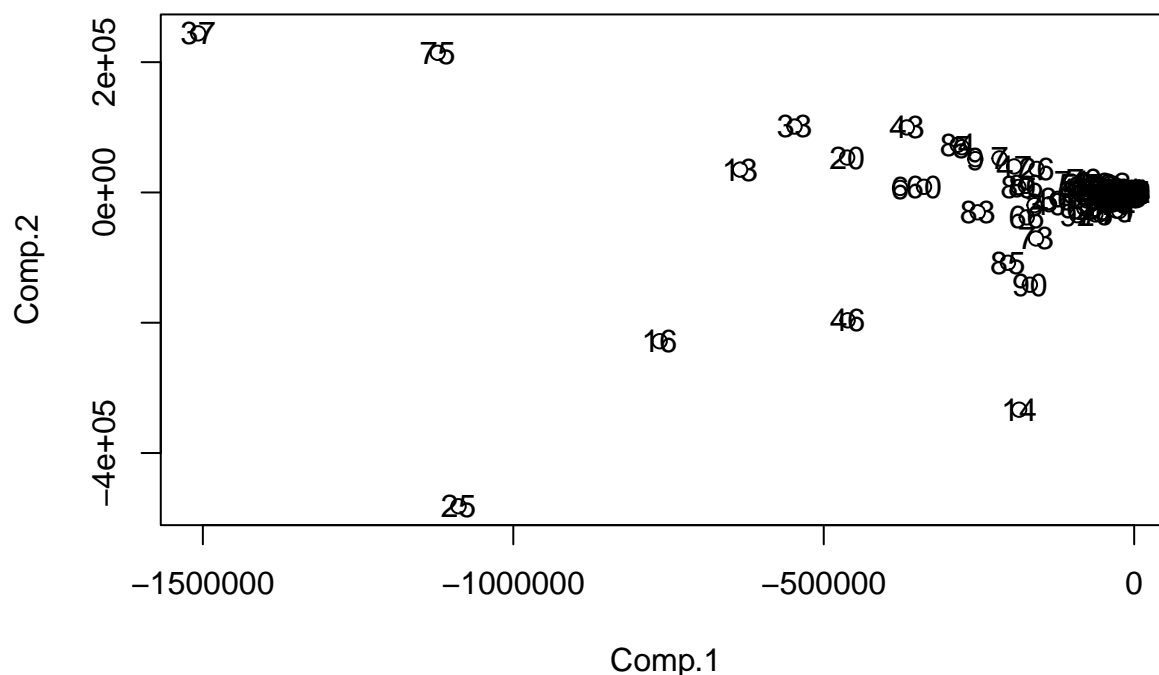
Obtenga las gráficas de respectivas con S (matriz de varianzas-covarianzas) y con R (matriz de correlaciones) de las dos primeras componentes e interprete los resultados en término de agrupación de variables (puede ayudar “índice de riqueza”, “índice de ruralidad”)

```
# Seleccionar solo las columnas numéricas
datos_numericos <- X[, sapply(X, is.numeric)]

# Análisis de Componentes Principales con la matriz de varianzas-covarianzas
cpS <- princomp(datos_numericos, cor = FALSE)
cpaS <- as.matrix(datos_numericos) %*% cpS$loadings

# Gráfica con la matriz de varianzas-covarianzas
plot(cpaS[, 1:2], type = "p", main = "Componentes Principales con S (Varianzas-Covarianzas)")
text(cpaS[, 1], cpaS[, 2], 1:nrow(cpaS))
```

## Componentes Principales con S (Varianzas–Covarianzas)



```
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
```

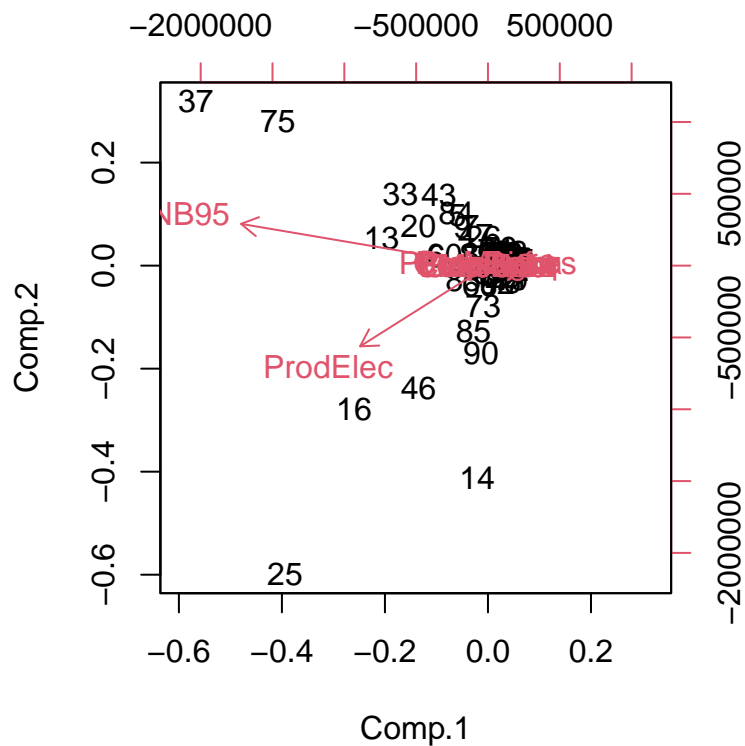
```
## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped
```

```
## 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
```

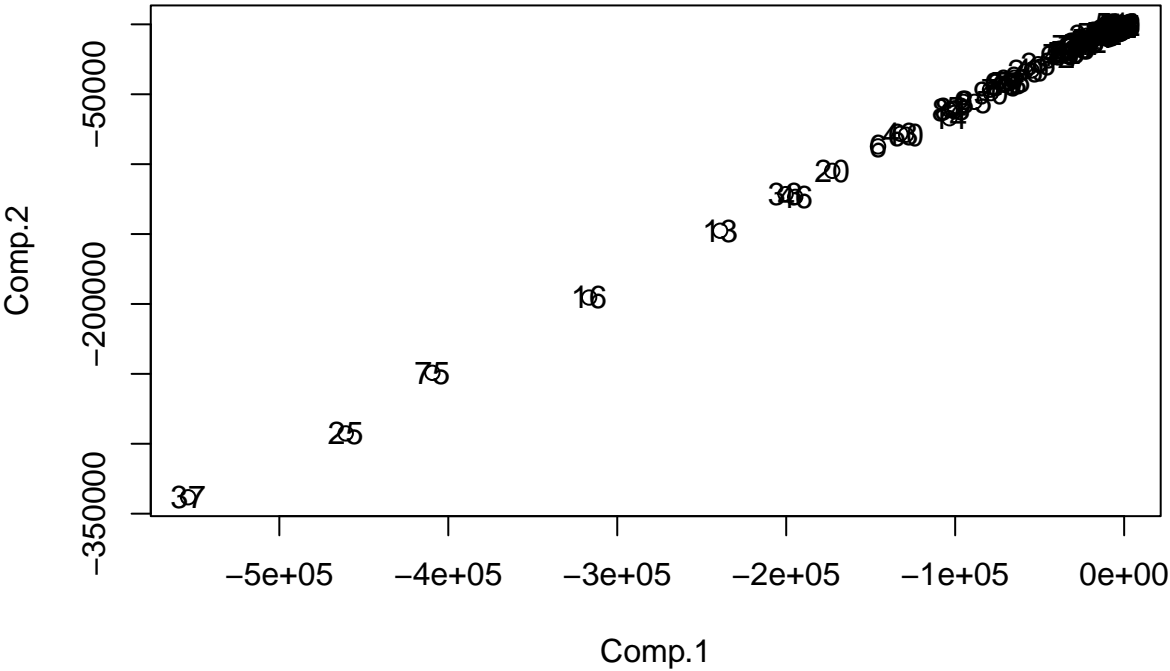
```
## Warning in arrows(0, 0, y[, 1L] * 0.8, y[, 2L] * 0.8, col = col[2L], length =
## arrow.len): zero-length arrow is of indeterminate angle and so skipped
```



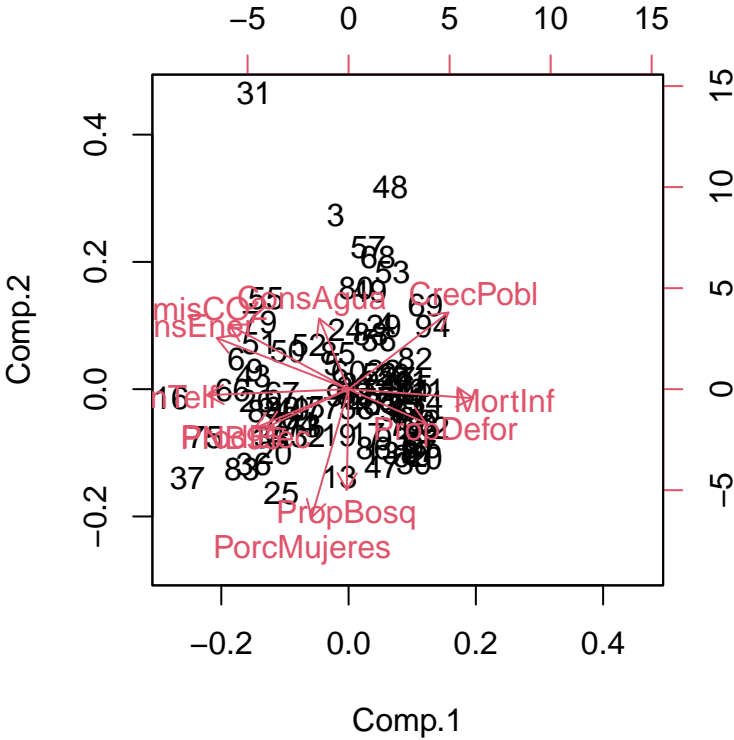
```
# Análisis de Componentes Principales con la matriz de correlaciones
cpR <- princomp(datos_numericos, cor = TRUE)
cpaR <- as.matrix(datos_numericos) %*% cpR$loadings

# Gráfica con la matriz de correlaciones
plot(cpaR[, 1:2], type = "p", main = "Componentes Principales con R (Correlaciones)")
text(cpaR[, 1], cpaR[, 2], 1:nrow(cpaR))
```

Componentes Principales con R (Correlaciones)



```
biplot(cpR)
```



## Parte 3

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

```
library(FactoMineR)
library(factoextra)
```

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

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

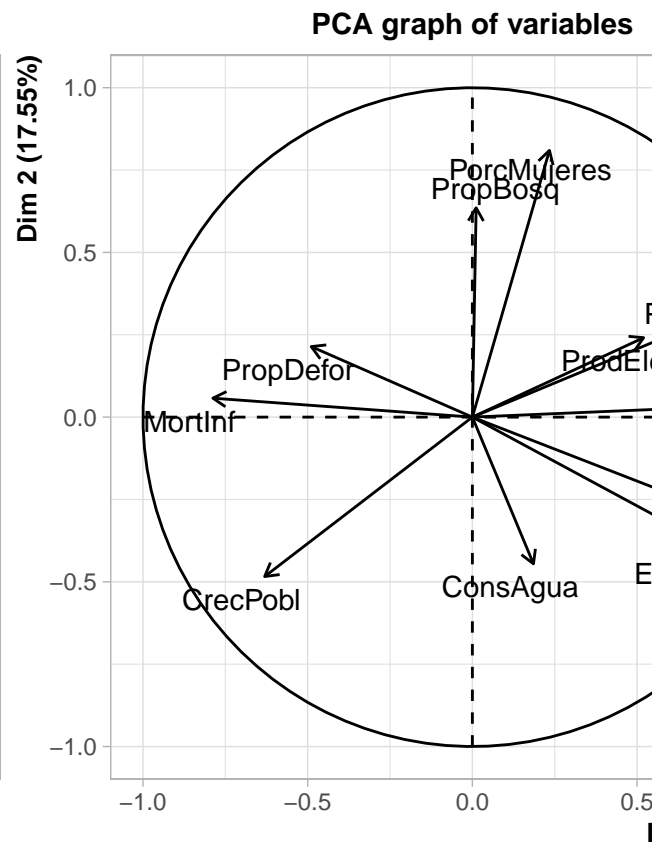
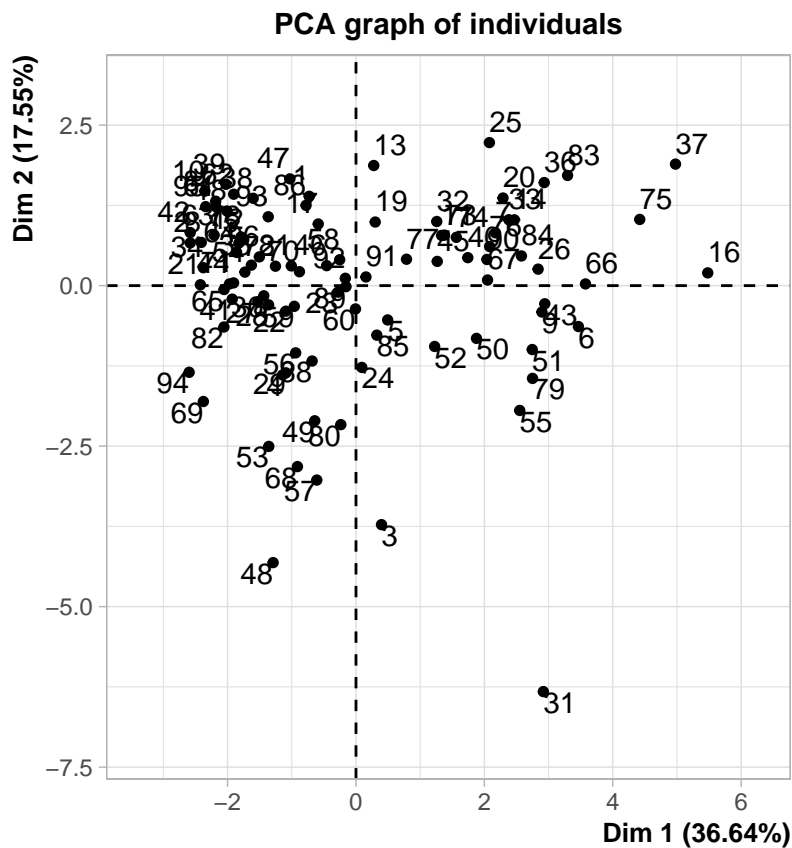
```
library(ggplot2)
```

```
# Seleccionar solo las columnas numéricas
```

```
datos_numericos <- X[, sapply(X, is.numeric)]
```

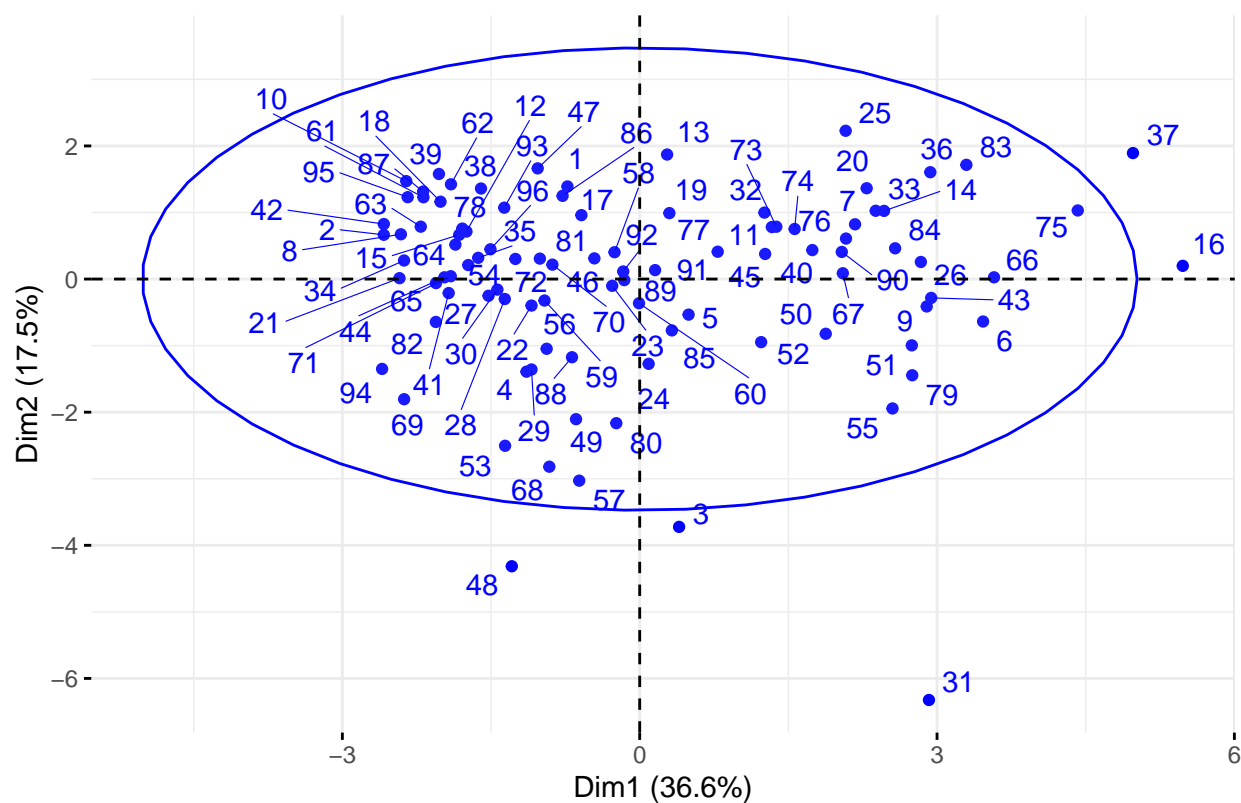
```
# Análisis de Componentes Principales
```

```
cp3 <- PCA(datos_numericos)
```



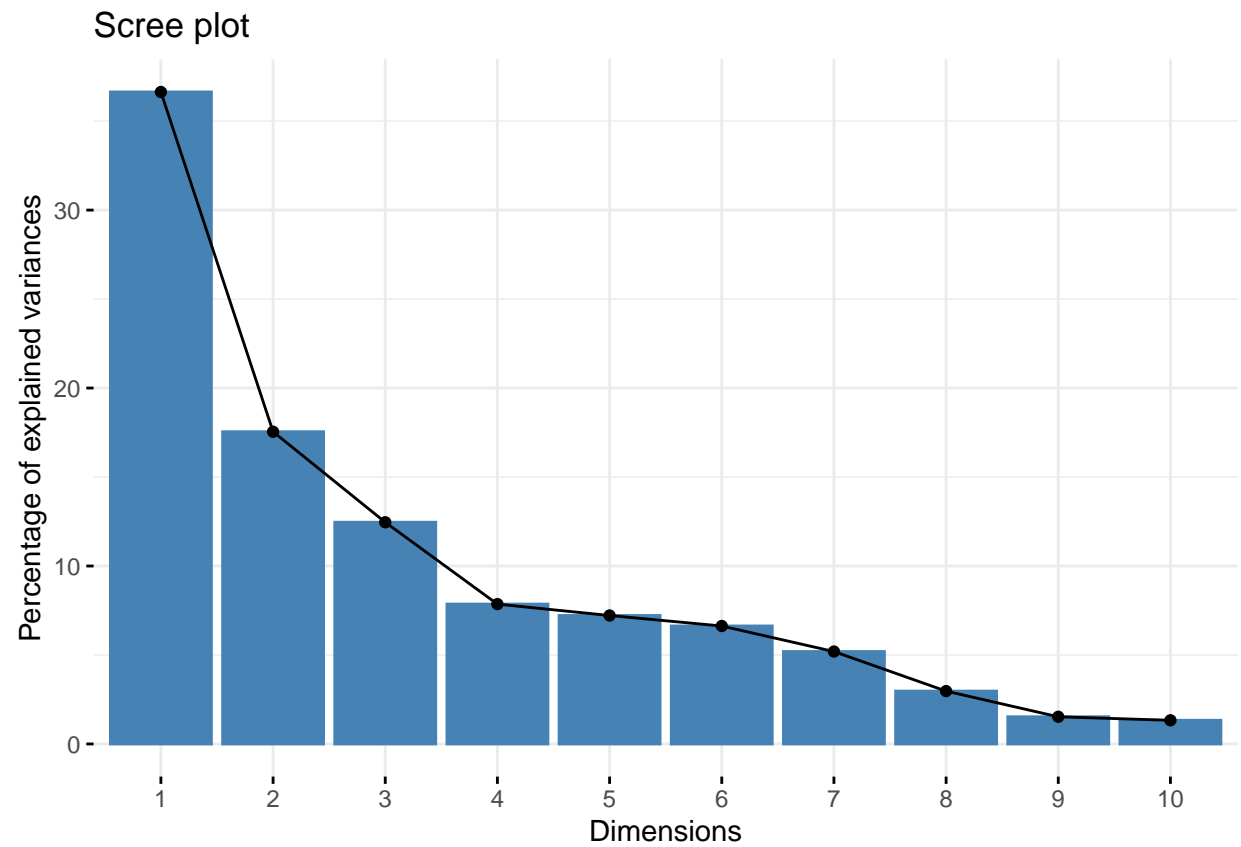
```
# Gráfico de individuos (puntos de observaciones) con elipses de confianza
fviz_pca_ind(cp3, col.ind = "blue", addEllipses = TRUE, repel = TRUE)
```

## Individuals – PCA



# Gráfico del screeplot para visualizar la proporción de varianza explicada por cada componente principal  
 fviz\_screplot(cp3)





```
# Gráfico de contribuciones de las variables a las componentes principales  
fviz_contrib(cp3, choice = c("var"))
```

