

# ACP

Irma Martínez

2022-04-04

## Analisis de Componentes principales

### Introducción

Analisis de Componentes principales (*ACP*) de la base de datos AEROPUERTOS

### Matriz de trabajo

Analisis de base de datos Aeropuertos, extraida del paquete **datos**

#### PASO 1: INSTALACIÓN DEL PAQUETE DATOS

```
install.packages("datos")
```

#### PASO 2: ABRIR LIBRERIA

```
library(datos)
```

#### PASO 3: ELIGE UNA MATRIZ

```
aeropuertos1<-datos::aeropuertos
```

#### PASO 4: EXPLORACIÓN DE LA MATRIZ

```
str(aeropuertos1)
```

```
## tibble [1,458 x 8] (S3: tbl_df/tbl/data.frame)
## $ codigo_aeropuerto: chr [1:1458] "04G" "06A" "06C" "06N" ...
## $ nombre           : chr [1:1458] "Lansdowne Airport" "Moton Field Municipal Airport" "Schaumburg R
## $ latitud           : num [1:1458] 41.1 32.5 42 41.4 31.1 ...
## $ longitud          : num [1:1458] -80.6 -85.7 -88.1 -74.4 -81.4 ...
## $ altura            : num [1:1458] 1044 264 801 523 11 ...
## $ zona_horaria      : num [1:1458] -5 -6 -6 -5 -5 -5 -5 -5 -8 ...
## $ horario_verano    : chr [1:1458] "A" "A" "A" "A" ...
## $ zona_horaria_iana: chr [1:1458] "America/New_York" "America/Chicago" "America/Chicago" "America/N
```

#### PASO 5: CONFIGURACIÓN Y/O FILTRADO DE VARIABLES (ME QUEDO CON LAS CUANTITATIVAS)

```
colnames(aeropuertos1)
```

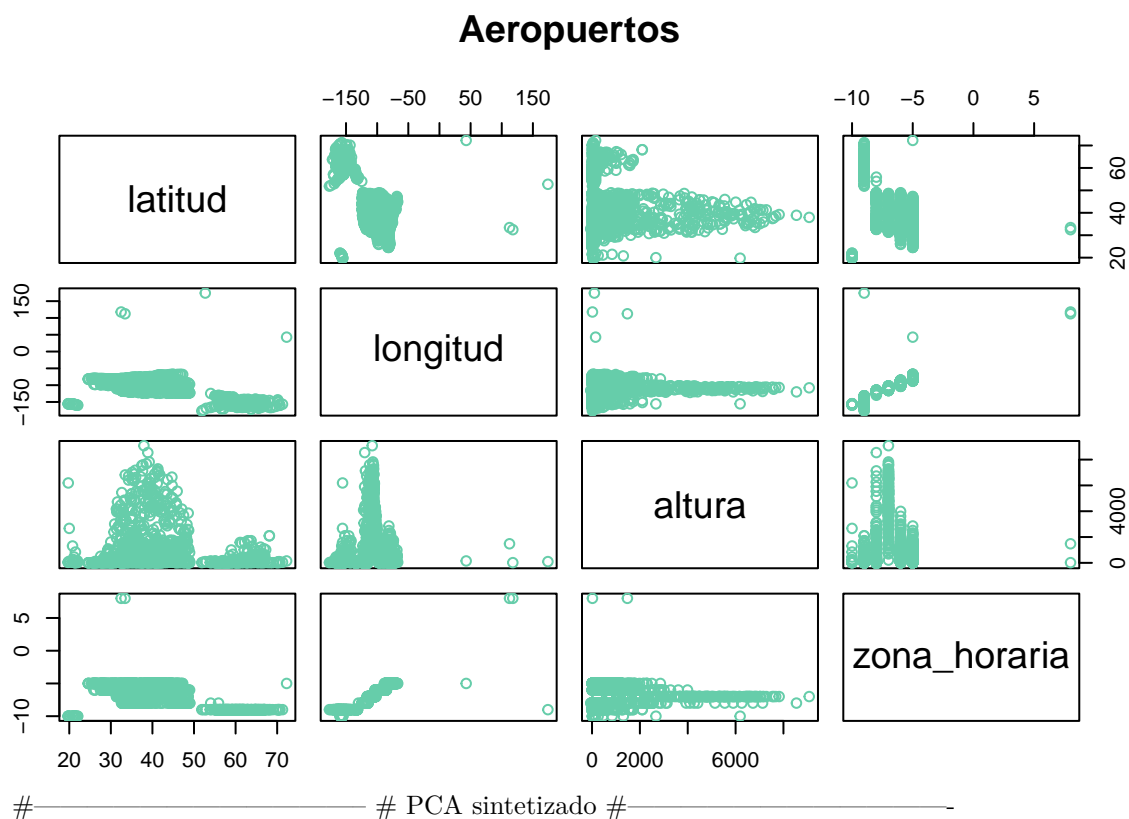
```
## [1] "codigo_aeropuerto" "nombre"          "latitud"
## [4] "longitud"          "altura"         "zona_horaria"
## [7] "horario_verano"    "zona_horaria_iana"
```

```
aeropuertos2<-aeropuertos1[,3:6]
colnames(aeropuertos2)
```

```
## [1] "latitud"          "longitud"         "altura"           "zona_horaria"
```

## PASO 6: DESARROLLAR EL PCA PASO A PASO

```
plot(aeropuertos2,col="aquamarine3",main="Aeropuertos")
```



```
x<-as.data.frame(aeropuertos2)
```

## VISUALIZACIÓN DE LOS DATOS

```
head(x)
```

```
##   latitud longitud altura zona_horaria
## 1 41.13047 -80.61958  1044          -5
## 2 32.46057 -85.68003   264          -6
## 3 41.98934 -88.10124   801          -6
## 4 41.43191 -74.39156   523          -5
## 5 31.07447 -81.42778    11          -5
## 6 36.37122 -82.17342  1593          -5
```

## Aplicar el cálculo de la varianza a las columnas

1=filas, 2=columnas

```
apply(x, 2, var)
```

```
##      latitud      longitud      altura zona_horaria  
## 1.091922e+02 8.906117e+02 2.321437e+06 2.634180e+00
```

## Centrado por la media y escalada por

la desviacion standar (dividir entre sd).

```
acp<-prcomp(x, center=TRUE, scale=TRUE)  
acp
```

```
## Standard deviations (1, .., p=4):  
## [1] 1.5503883 1.0146787 0.6987463 0.2801370  
##  
## Rotation (n x k) = (4 x 4):  
##           PC1      PC2      PC3      PC4  
## latitud    0.50693661 0.19190799 -0.8397303 0.03224415  
## longitud   -0.61286128 0.06759722 -0.3270306 0.71615829  
## altura     -0.01486704 -0.97218246 -0.2321894 -0.02698790  
## zona_horaria -0.60596642 0.11603108 -0.3660495 -0.69667014
```

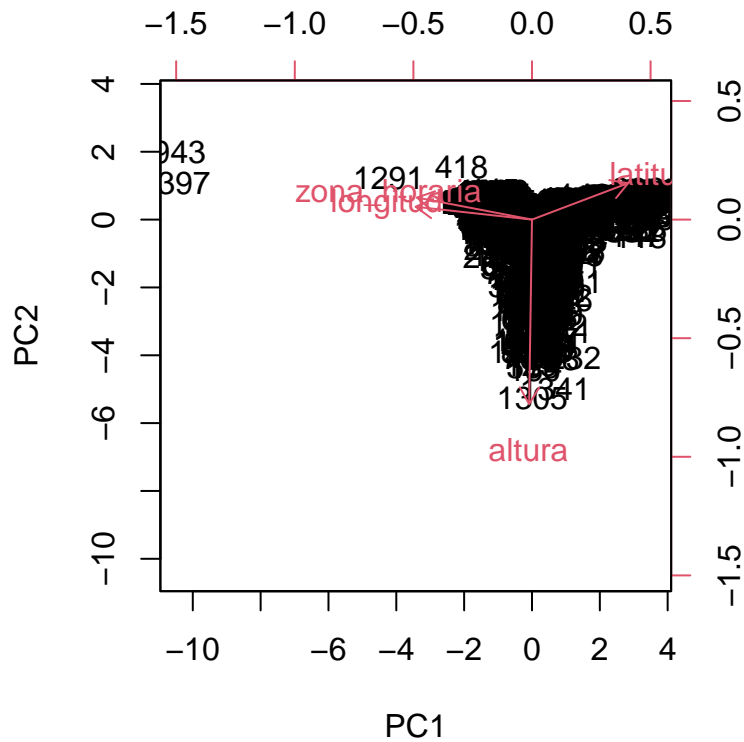
## Visualizar el resumen

```
summary(acp)
```

```
## Importance of components:  
##           PC1      PC2      PC3      PC4  
## Standard deviation    1.5504 1.0147 0.6987 0.28014  
## Proportion of Variance 0.6009 0.2574 0.1221 0.01962  
## Cumulative Proportion 0.6009 0.8583 0.9804 1.00000
```

## Construcción del Biplot

```
biplot(acp, scale=0)
```



Componente principal calculada

Suma del producto de la matriz acp de cada uno

de los componentes por el dato de la matriz original

por filas

filas = 1

columnas = 2

```
pc1<-apply(acp$rotation[,1]*x, 1, sum)
pc2<-apply(acp$rotation[,2]*x, 1, sum)
pc3<-apply(acp$rotation[,3]*x, 1, sum)
```

```
x$pc1<-pc1
x$pc2<-pc2
x$pc3<-pc3
```