

Actividad de la clase 3 - A01708119

Erick Alfredo Garcia Huerta - A01708119

2024-05-21

Los datos

```
M <- read.csv("energiafossil_PIB.csv")
length(table(M$entidad))

## [1] 161

names(M)

## [1] "entidad"          "electrd_fosiles" "PIB_per_cap"
"emisiones_CO2"
```

Se observa que hay 161 países en la variable entidad.

DOs filtros M1 la base de datos numéricos y M2 que tiene la base de datos de países

Filtro de 5 países:

```
#[condiciones filas, condiciones columnas]
M2 <- M[M$entidad == 'Mexico' |
        M$entidad == 'Spain' |
        M$entidad == 'Argentina' |
        M$entidad == 'Chile' |
        M$entidad == 'Costa Rica', ]
table(M2$entidad)

##
## Argentina      Chile Costa Rica      Mexico      Spain
##           20           20           20           20           20
```

Ahora haremos un filtro con M2 pero sólo variables numéricas

```
# M1 es de sólo variables numéricas
M1 <- M[-1] # Quita la primera columna, para poner varias columnas, se
hace un vecot de números negativos
head(M1)

## electrd_fosiles PIB_per_cap emisiones_CO2
## 1           0.13      179.4266          1030
## 2           0.31      190.6838          1220
## 3           0.33      211.3821          1030
## 4           0.34      242.0313          1550
## 5           0.20      263.7336          1760
## 6           0.20      359.6932          1770
```

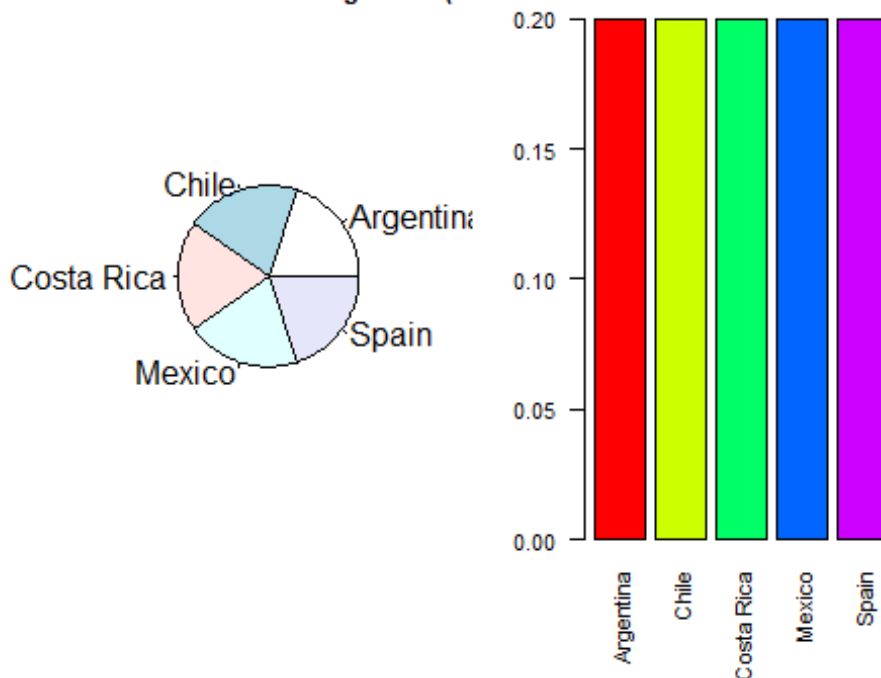
Análisis de las variables categóricas de la base de datos M2 (5 países)

```
tabla <- prop.table(table(M2$entidad))
```

Se observa que los datos se distribuyen uniformemente en los 5 países (20%)

```
par(mfrow = c(1, 2)) #Hace una matriz de gráficos de 1 fila y 2 columnas
pie(tabla)
par(cex = .7)
title("Análisis de las variables categóricas (5 países)")
barplot(tabla, col = rainbow(5), las = 2)
```

Análisis de las variables categóricas (5



Análisis de

las variables numéricas

Tabla de medidas resumen

```
summary(M1)
```

```
## electrd_fosiles    PIB_per_cap    emisiones_CO2
## Min.   : 0.00      Min.   : 111.9    Min.   :   30
## 1st Qu.: 0.30      1st Qu.: 1283.9   1st Qu.:  2210
## Median : 3.18      Median : 4461.9   Median : 11070
## Mean   : 74.45     Mean   : 13012.9   Mean   : 163428
## 3rd Qu.: 27.71     3rd Qu.: 15311.8  3rd Qu.: 60600
## Max.   :5098.22    Max.   :123514.2  Max.   :10707220
```

```
cat("Las desviaciones estándar son: \n")
```

```
## Las desviaciones estándar son:

apply(M1, 2, sd)

## electrd_fosiles      PIB_per_cap      emisiones_CO2
##      358.7617      19375.7730      783605.0117

cv = function(x) {
  sd(x) / mean(x)
}
cat("Los coeficientes de variación son: \n")

## Los coeficientes de variación son:

apply(M1, 2, cv)

## electrd_fosiles      PIB_per_cap      emisiones_CO2
##      4.818712      1.488963      4.794794
```

Matriz de gráficos histogramas

```
par(mfrow = c(1, 3))
hist(M1$electrd_fosiles, main = "Fósiles", col = rainbow(10))
hist(M1$PIB_per_cap, main = "PIB", col = terrain.colors(10))
hist(M1$emisiones_CO2, main = "CO2", col = cm.colors(10))
```

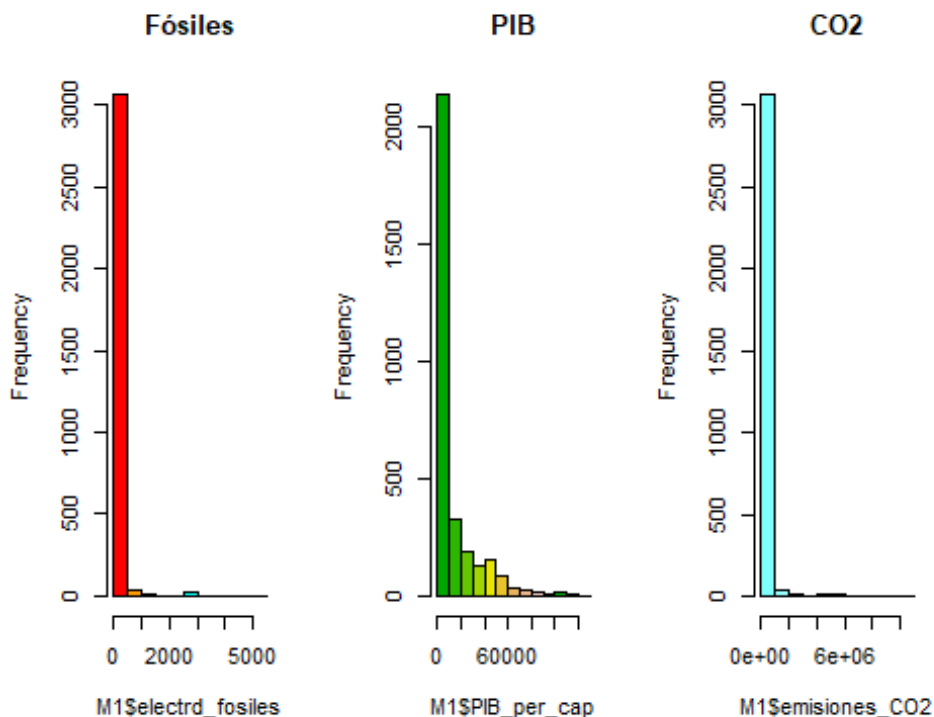
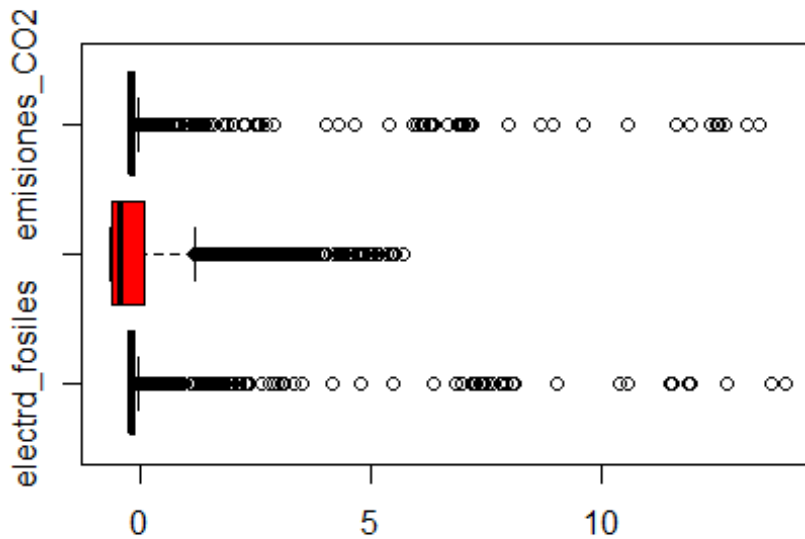


Diagrama de caja y bigote de M1

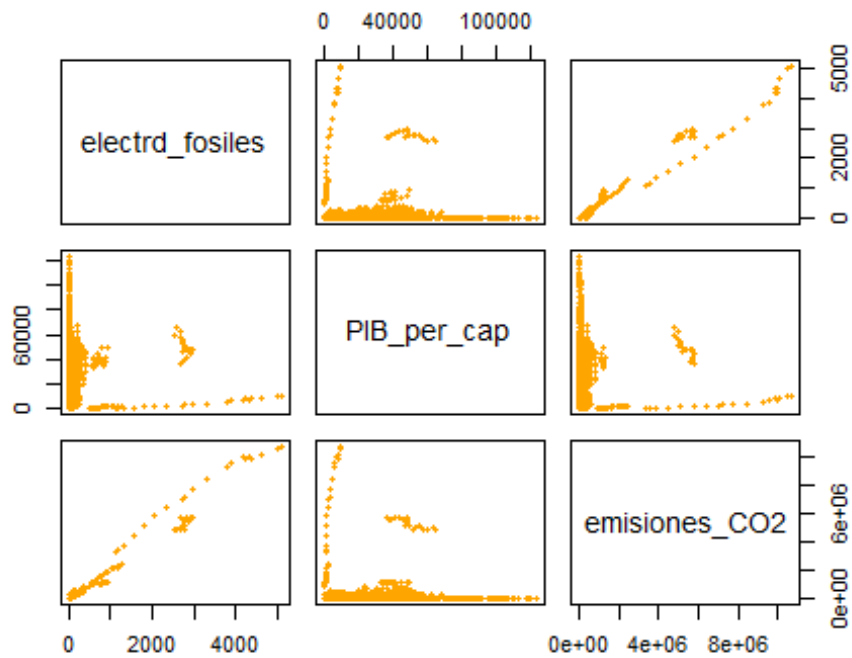
```
boxplot(scale(M1), horizontal = TRUE, col = 'red') # scale escalar los  
datos (estandarizar (x - media / desviación estandar), sólo usar cuando  
las escalas originales son muy diferentes)
```



graficos de dispersión de las variables numéricas

Matriz de

```
plot(M1, pch = 20, col = 'Orange')
```



Matriz de correlación

`cor(M1)`

```
##               electrd_fosiles PIB_per_cap emisiones_CO2
## electrd_fosiles      1.0000000  0.11559112  0.98973515
## PIB_per_cap          0.1155911  1.00000000  0.09761301
## emisiones_CO2        0.9897351  0.09761301  1.00000000
```