

# Medidas Cuantitativas-Noreste

Equipo

2024-11-8

#1. Cargar Dataset Diccionario de Dataset usado (0X.csv): 0: "NORTE 2" 1: "NORTE" 2: "NORESTE" 3: "SURESTE" 4: "CENTRO"

```
#Cargar dataset
dataset = read.csv("02.csv")
head(dataset)

##           X      NO2      O3 PM10      SO2 WSR WDV
## 1 2023-01-01 00:00:00 0.0430 0.007  222 0.0038 4.1 190
## 2 2023-01-01 01:00:00 0.0444 0.008  311 0.0240 1.4 222
## 3 2023-01-01 02:00:00 0.0415 0.007  723 0.0098 2.5   4
## 4 2023-01-01 03:00:00 0.0423 0.006  473 0.0078 1.6 330
## 5 2023-01-01 04:00:00 0.0399 0.005  372 0.0073 4.1 343
## 6 2023-01-01 05:00:00 0.0351 0.006  285 0.0060 5.0 334
```

Se encuentra que hay 7 columnas en el dataset. Para hacer un análisis cuantitativo de las variables numéricas no categóricas se excluye el tiempo como variable de análisis.

#2.Extraer medidas cuantitativas

##Medidas de tendencia central

```
medidas_tendencia_central<-function(column_dataset){
  media = mean(column_dataset)
  mediana = median(column_dataset)

  #Función de moda
  mode <- function(v) {
    uniqv <- unique(v)
    uniqv[which.max(tabulate(match(v, uniqv)))]
  }
  moda = mode(column_dataset)

  return(c(media, mediana, moda))
}

resultados_tendencia_central <- matrix(nrow = 6, ncol = 3)
for(i in 1:6){
  resultados_tendencia_central[i,]<-medidas_tendencia_central(dataset[,i+1])
}
#Definir nombres de filas y columnas
rownames(resultados_tendencia_central) <- c("NO2", "O3", "PM10", "SO2",
"Velocidad Aire", "Dirección Aire")
colnames(resultados_tendencia_central) <- c("Media", "Mediana", "Moda")
```

```
resultados_tendencia_central
```

```
##           Media  Mediana   Moda
## NO2          0.01485351  0.0118  0.0094
## O3           0.02812376  0.0250  0.0070
## PM10         66.61092942 56.0000 50.0000
## SO2          0.00387588  0.0034  0.0028
## Velocidad Aire  8.51191420  8.1000  5.0000
## Dirección Aire 131.49628695 138.0000 144.0000
```

Se puede ver que hay cierta concordancia de media y mediana en los contaminantes NO2,O3,SO2 y en la velocidad del aire. ## Medidas de dispersión

```
medidas_dispersion<-function(column_dataset){
  min = min(column_dataset)
  max = max(column_dataset)
  var = var(column_dataset)
  desvest = sqrt(var)
  return(c(min, max, desvest, var))
}
resultados_medidas_dispersion <- matrix(nrow = 6, ncol = 4)

for(i in 1:6){
  resultados_medidas_dispersion[i, ] <- medidas_dispersion(dataset[, i+1])
}
```

*#Definir nombres de filas y columnas*

```
rownames(resultados_medidas_dispersion) <- c("NO2", "O3", "PM10", "SO2",
"Velocidad Aire", "Dirección Aire")
colnames(resultados_medidas_dispersion) <- c("Min", "Max", "Desviación
Est.", "Varianza")
```

```
resultados_medidas_dispersion
```

```
##           Min      Max Desviación Est.    Varianza
## NO2          5e-04   0.0804    0.009679910 9.370066e-05
## O3           1e-03   0.1550    0.016594674 2.753832e-04
## PM10         2e+00 900.0000    52.401551051 2.745923e+03
## SO2          8e-04   0.1118    0.002518853 6.344619e-06
## Velocidad Aire 2e-01  28.0000    4.228561420 1.788073e+01
## Dirección Aire 1e+00 360.0000    76.256286145 5.815021e+03
```

##Resultados de tendencia central y medidas de dispersión

```
print(cbind(resultados_tendencia_central,resultados_medidas_dispersion))

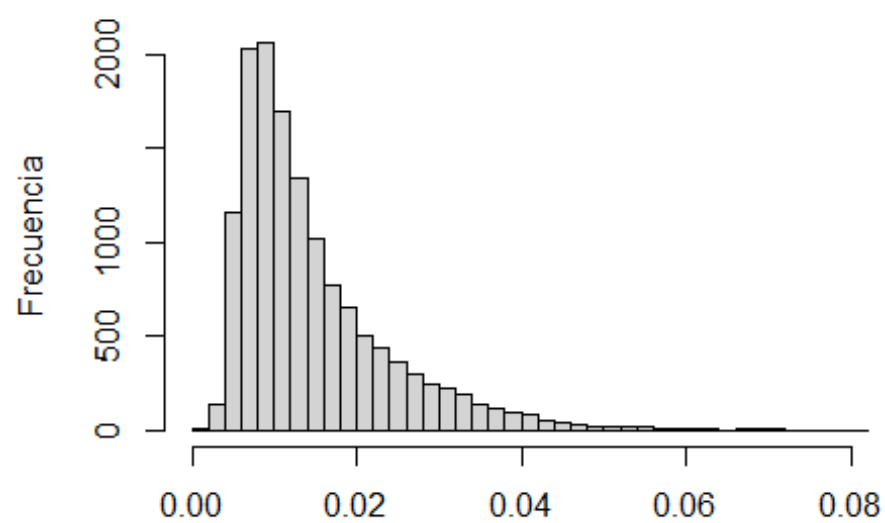
##           Media  Mediana   Moda   Min      Max Desviación
Est.
## NO2          0.01485351  0.0118  0.0094 5e-04   0.0804
0.009679910
```

```
## O3          0.02812376    0.0250    0.0070 1e-03    0.1550
0.016594674
## PM10        66.61092942   56.0000   50.0000 2e+00 900.0000
52.401551051
## SO2         0.00387588    0.0034    0.0028 8e-04    0.1118
0.002518853
## Velocidad Aire  8.51191420    8.1000    5.0000 2e-01   28.0000
4.228561420
## Dirección Aire 131.49628695 138.0000 144.0000 1e+00 360.0000
76.256286145
##              Varianza
## NO2          9.370066e-05
## O3           2.753832e-04
## PM10         2.745923e+03
## SO2          6.344619e-06
## Velocidad Aire 1.788073e+01
## Dirección Aire 5.815021e+03
```

##Tabla de distribución frecuencia

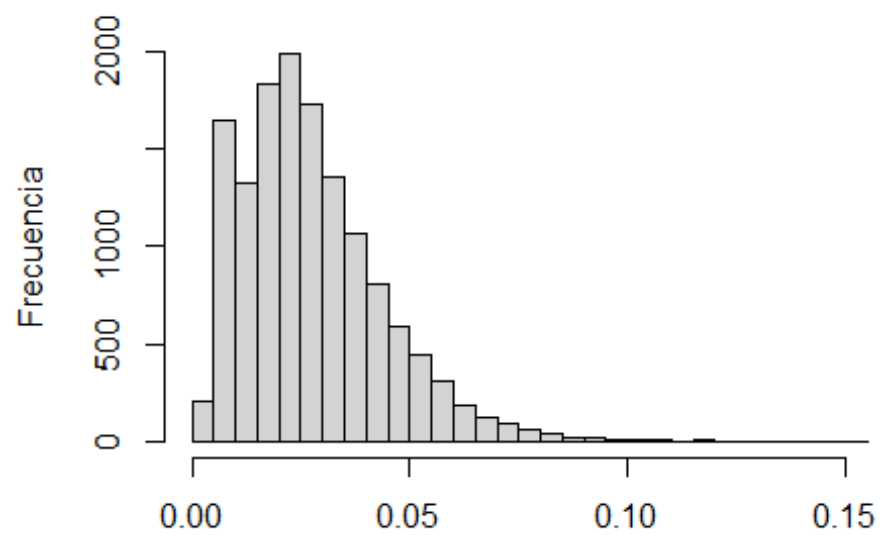
```
titulos <-c("NO2", "O3", "PM10", "SO2", "Velocidad Aire", "Dirección Aire")
for (i in 2:7){
  hist(dataset[,i],breaks = 35, main = titulos[i-1], xlab = titulos[i-1],ylab
= "Frecuencia")
}
```

## NO2



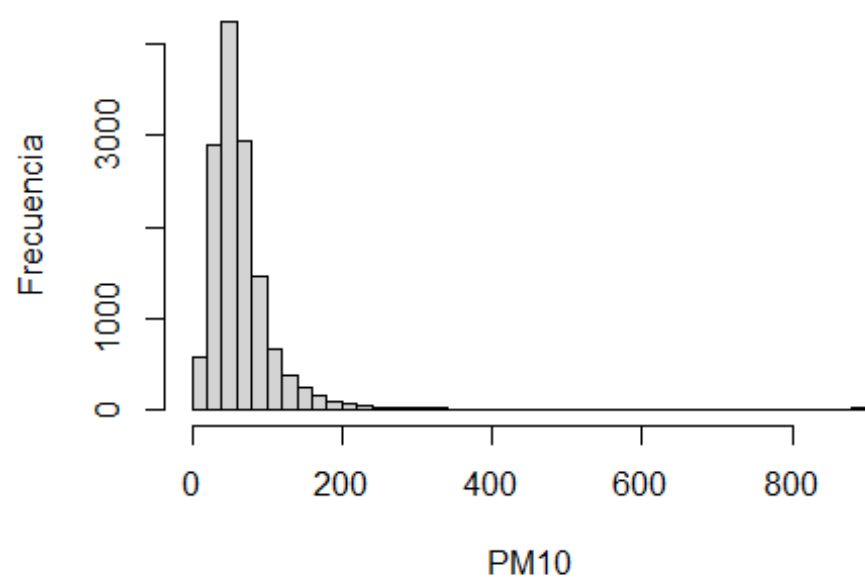
NO2

## O3

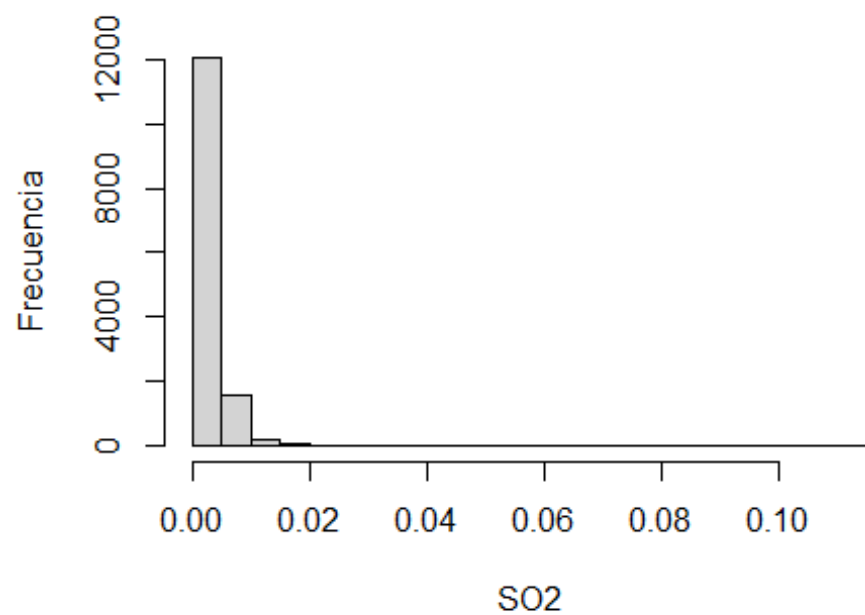


O3

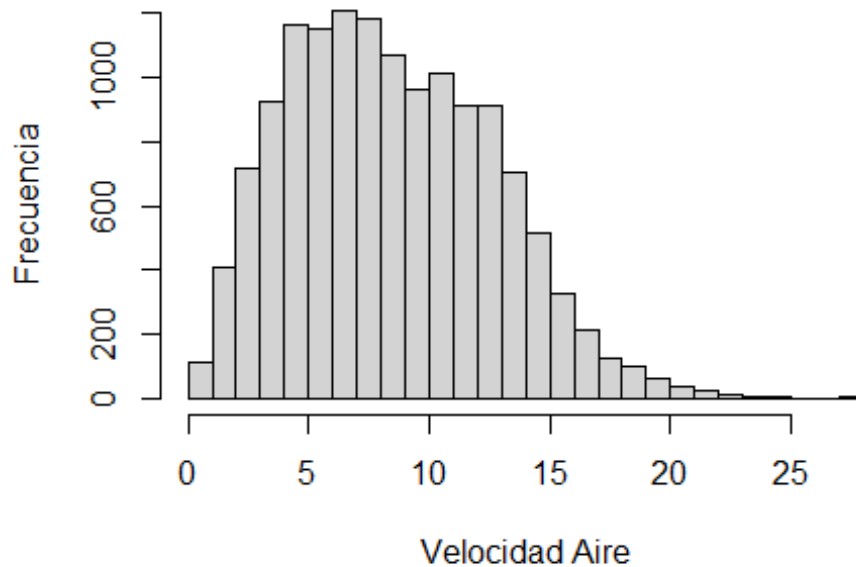
### PM10



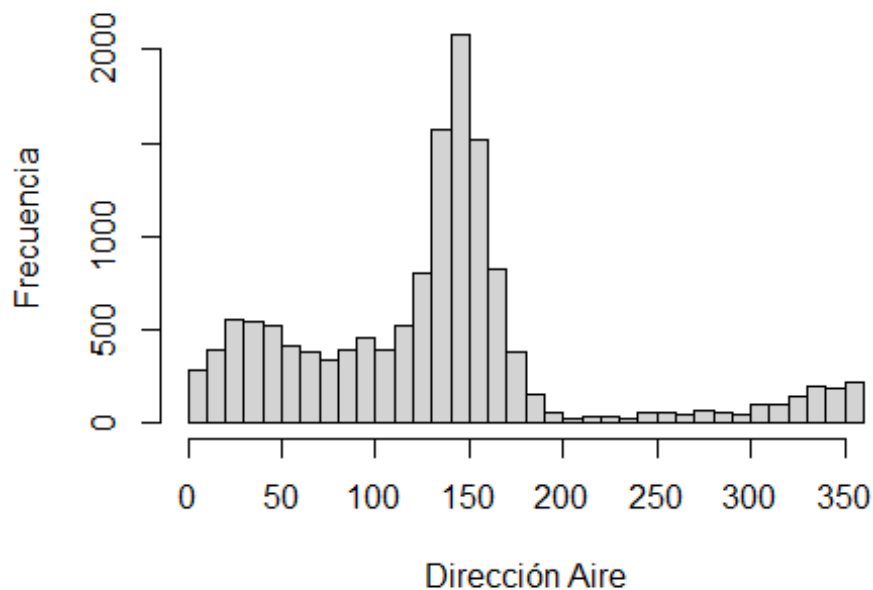
### SO2



### Velocidad Aire



### Dirección Aire



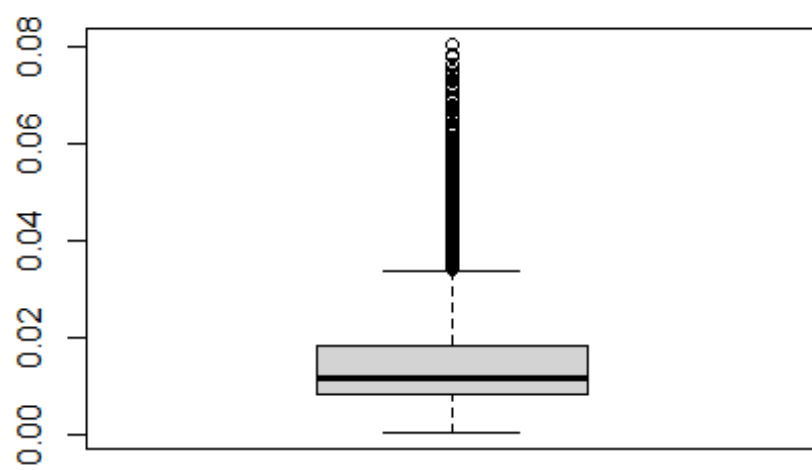
NO2: Asimétrica  
 sesgada a la derecha O3: Asimétrica sesgada a la derecha PM10: Asimétrica sesgada a la  
 derecha SO2:Asimétrica sesgada a la derecha (Una gran concentración de datos en los  
 rangos 0.02 y 0.04) Velocidad aire: Asimétrica sesgada a la derecha(Hay vientos  
 mayormente entre 3 y 14 kilómetros por hora) Tiene cierta normalidad Dirección aire:

Asimétrica sesgada a la derecha (Se concentra las direcciones en el sureste)

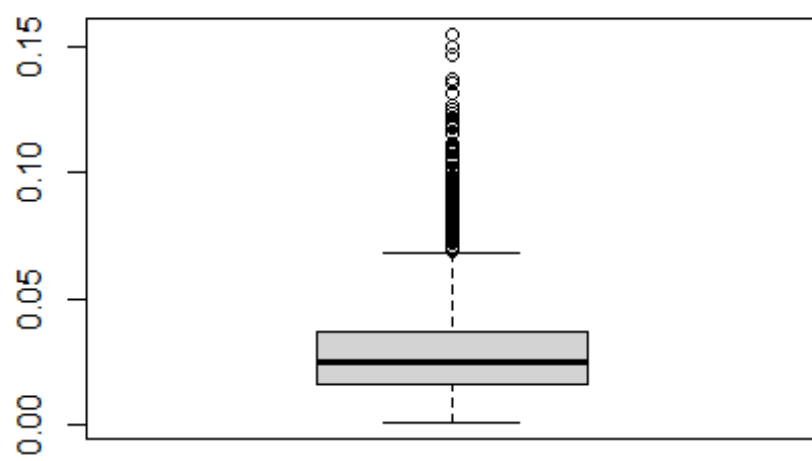
#3.Exploración a través de gráficas

```
titulos <- c("N02", "O3", "PM10", "SO2", "Velocidad Aire", "Dirección Aire")
for (i in 2:7){
  boxplot(dataset[,i], main = titulos[i-1])
}
```

## NO2

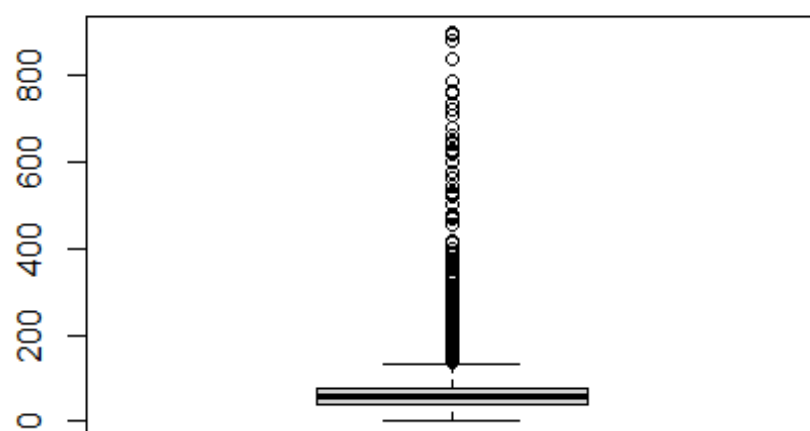


## O3

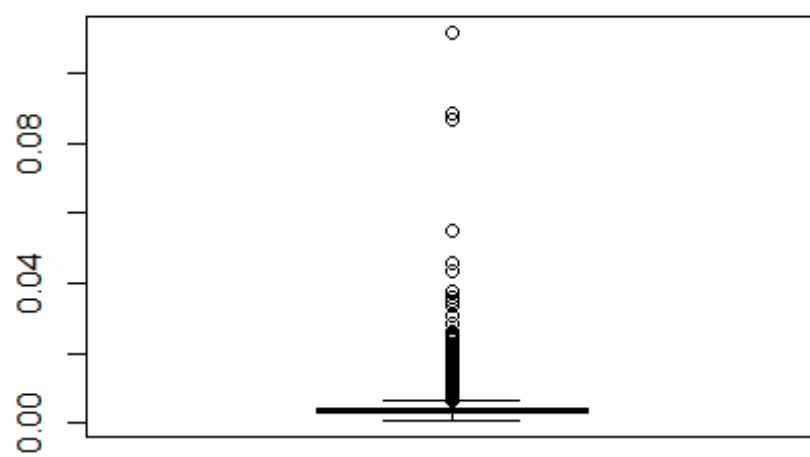




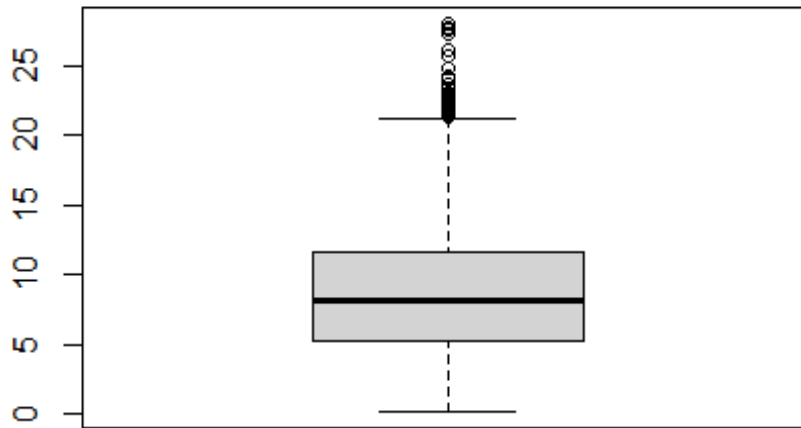
## PM10



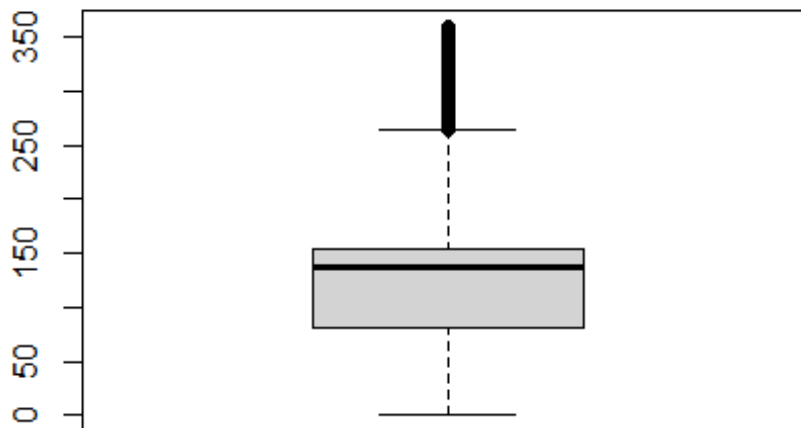
## SO2



### Velocidad Aire



### Dirección Aire



Muchos valores atípicos observados a través de los boxplots, especialmente en la dirección del aire(pero se ven más agrupados). ##Correlación

```

cor_matrix<-cor(data.frame(dataset[,-1]))
cor_matrix

##           NO2           O3           PM10           SO2           WSR
WDV
## NO2    1.0000000 -0.36235094  0.50108300  0.19692360 -0.53392293
0.16953613
## O3     -0.3623509  1.00000000 -0.04374409  0.21698821  0.37989180 -
0.01669908
## PM10   0.5010830 -0.04374409  1.00000000  0.17135237 -0.12238217
0.10904191
## SO2    0.1969236  0.21698821  0.17135237  1.00000000 -0.02976535
0.04721607
## WSR   -0.5339229  0.37989180 -0.12238217 -0.02976535  1.00000000 -
0.01777870
## WDV    0.1695361 -0.01669908  0.10904191  0.04721607 -0.01777870
1.00000000

#chechar corr.test() con librería psych para ver sus p valores.
#new=corr.test(dataset[,-1])
#new$ci

```

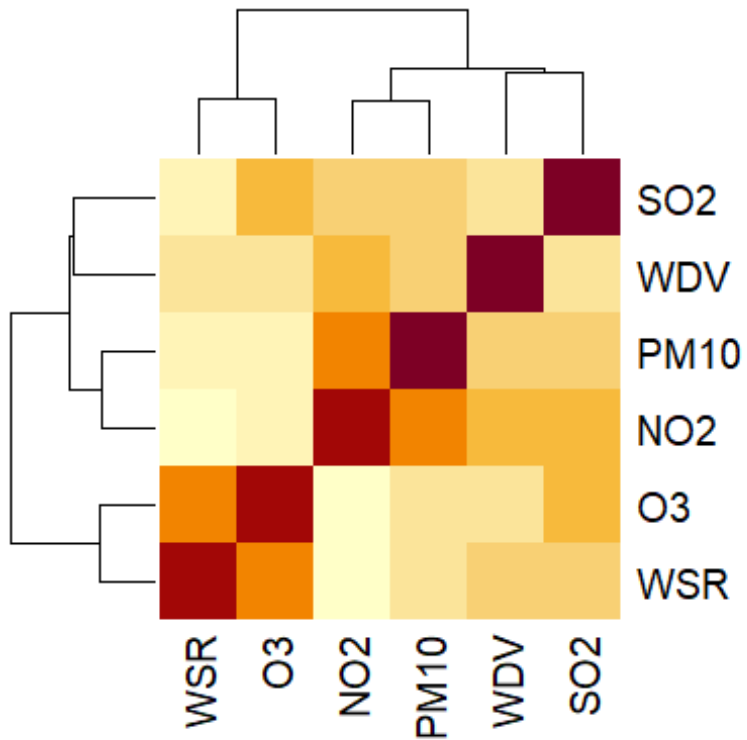
## Mapa de Calor de Correlación

```

#install.packages("corrplot")
library(reshape2)
library(ggplot2)

heatmap(cor_matrix)

```

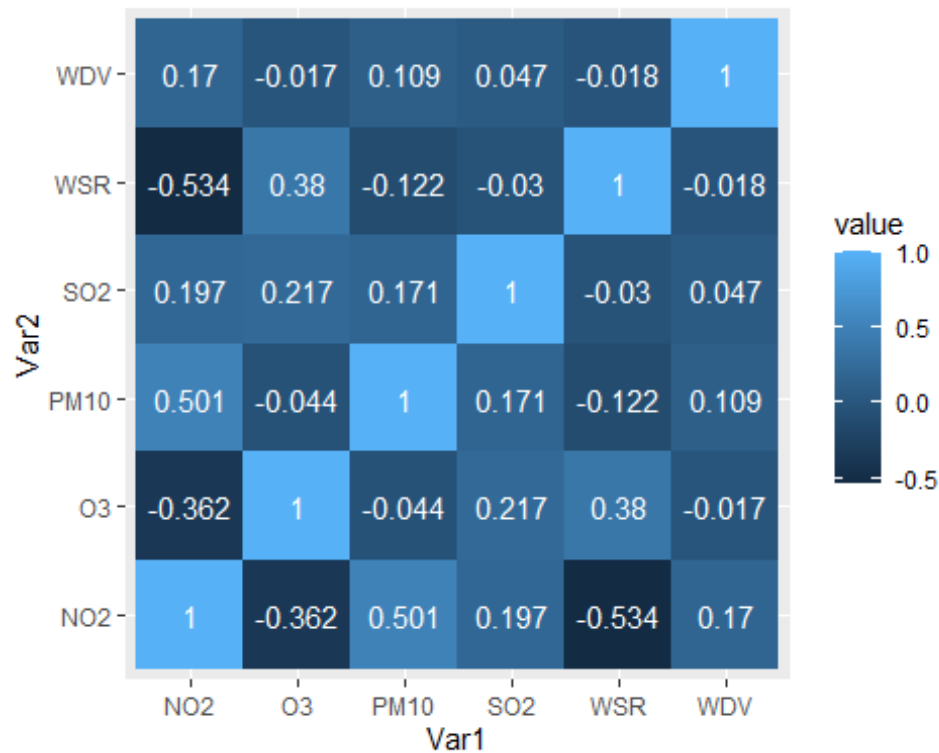


```

titulos <-c("NO2", "O3", "PM10", "SO2", "Velocidad Aire", "Dirección Aire")
data_melt<-melt(round(cor_matrix,3))

ggp <- ggplot(data_melt, aes(Var1,Var2, fill=value))
+geom_tile()+geom_text(aes(Var2,Var1,label = value),color = 'white',size = 4)
ggp
heatmap

```



#Ejemplo

comprobar si hay correlación significativa entre variables

```
alpha = 0.01
corr_dataframe<-data.frame(dataset[-1])
correlation<-cor.test(corr_dataframe[,1], corr_dataframe[,2], method =
"pearson", alternative = "two.sided", conf.level = 1-alpha)
cat("Para las variables", colnames(corr_dataframe)[1:2], "\n")

## Para las variables NO2 O3

cat("Valor p:", correlation$p.value, "\n")

## Valor p: 0

cat("Conclusión:", ifelse(correlation$p.value < alpha, "Existe una
correlación significativa", "No existe una correlación significativa"))

## Conclusión: Existe una correlación significativa
```