

Medidas Cuantitativas-Norte

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("01.csv")
```

```
head(dataset)
```

```
##           X      NO2      O3 PM10      SO2 WSR  WDV
## 1 2023-01-01 00:00:00 0.0432 0.012  522 0.0042 3.8  18
## 2 2023-01-01 01:00:00 0.0411 0.009  712 0.0040 4.5  15
## 3 2023-01-01 02:00:00 0.0379 0.007  586 0.0038 4.8  28
## 4 2023-01-01 03:00:00 0.0349 0.007  406 0.0037 2.8  26
## 5 2023-01-01 04:00:00 0.0335 0.007  340 0.0038 4.1  28
## 6 2023-01-01 05:00:00 0.0278 0.006  314 0.0037 2.6  27
```

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          0.015054585 0.0110 0.0080
## O3           0.027259697 0.0230 0.0050
## PM10         63.320358688 52.0000 39.0000
## SO2          0.003148688 0.0028 0.0024
## Velocidad Aire 7.589578226 6.9000 4.9000
## Dirección Aire 93.655046864 72.0000 21.0000
```

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         1e-04   0.0873    0.011140131 1.241025e-04
## O3          1e-03   0.1710    0.019264086 3.711050e-04
## PM10        2e+00  712.0000   46.226896971 2.136926e+03
## SO2         5e-04   0.0590    0.001881359 3.539510e-06
## Velocidad Aire 1e-01  27.4000    4.185858651 1.752141e+01
## Dirección Aire 1e+00 360.0000   85.874313473 7.374398e+03
```

##Resultados de tendencia central y medidas de dispersión

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

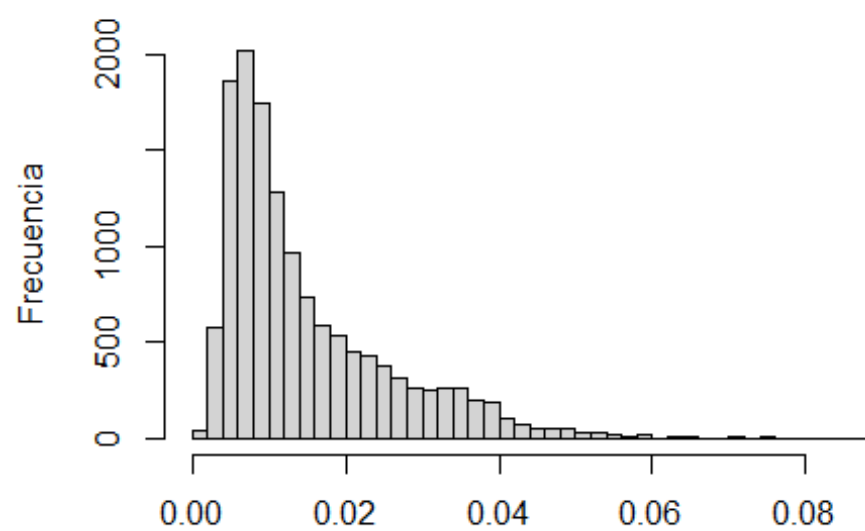
##              Media Mediana   Moda   Min      Max Desviación Est.
## NO2          0.015054585 0.0110 0.0080 1e-04   0.0873    0.011140131
## O3           0.027259697 0.0230 0.0050 1e-03   0.1710    0.019264086
## PM10         63.320358688 52.0000 39.0000 2e+00  712.0000   46.226896971
```

```
## SO2          0.003148688  0.0028  0.0024 5e-04  0.0590  0.001881359
## Velocidad Aire 7.589578226  6.9000  4.9000 1e-01  27.4000  4.185858651
## Dirección Aire 93.655046864 72.0000 21.0000 1e+00 360.0000 85.874313473
##
##          Varianza
## NO2          1.241025e-04
## O3           3.711050e-04
## PM10         2.136926e+03
## SO2          3.539510e-06
## Velocidad Aire 1.752141e+01
## Dirección Aire 7.374398e+03
```

##Tabla de distribución frecuencia

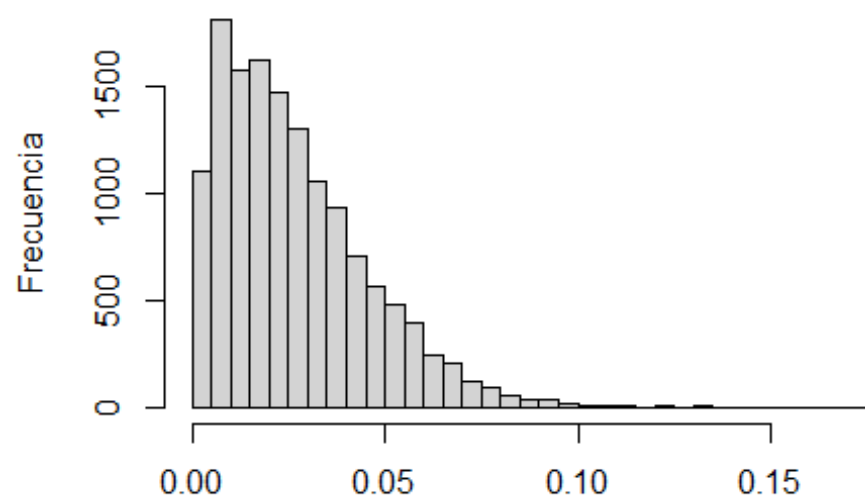
```
titulos <-c("NO2", "O3", "PM10", "SO2", "Velocidad Aire", "Dirección Aire")
for (i in 2:7){
  hist(dataset[,i],breaks = 35, main = titulos[i-1], xlab = titulos[i-1],ylab
= "Frecuencia")
}
```

NO2



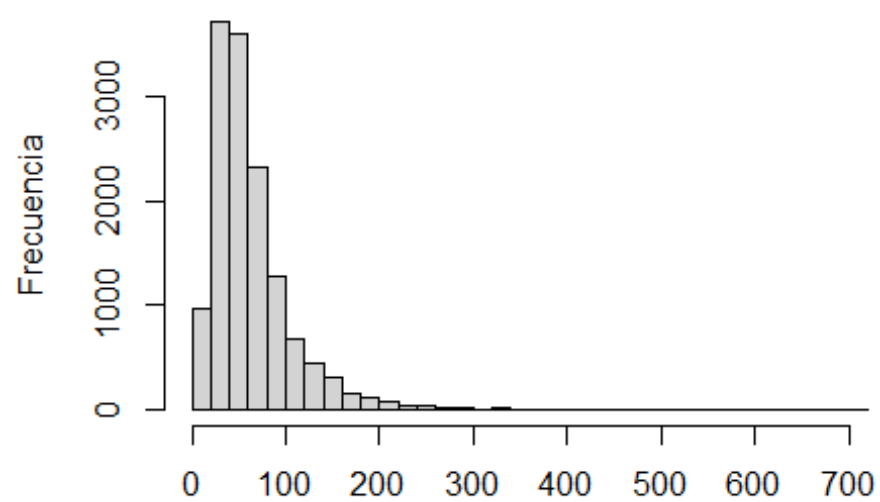
NO2

O3



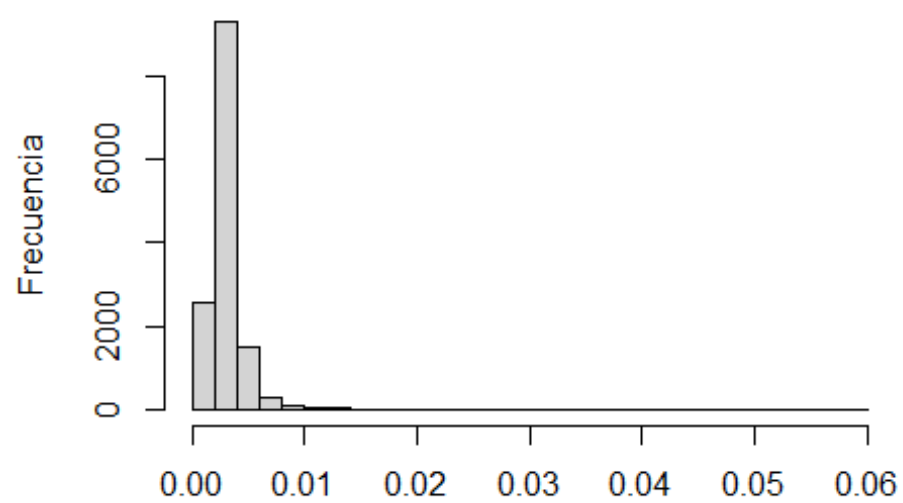
O3

PM10



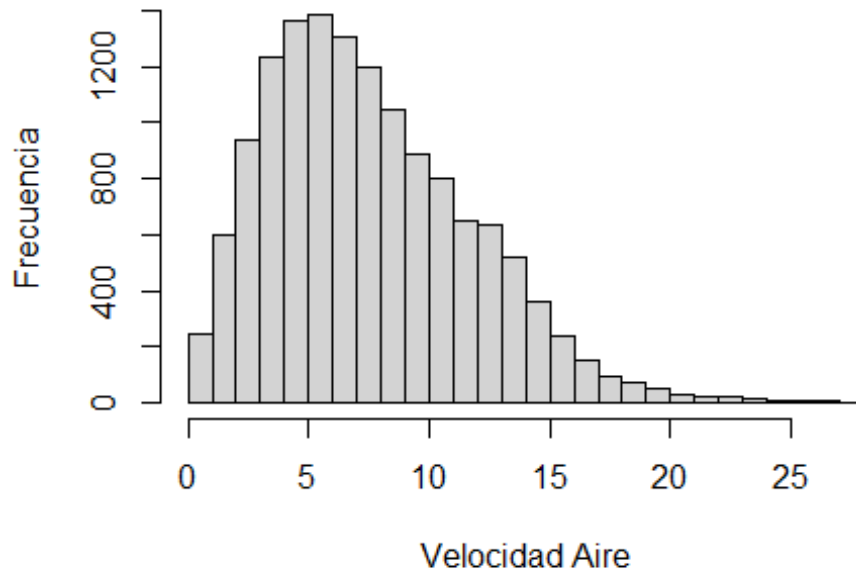
PM10

SO2

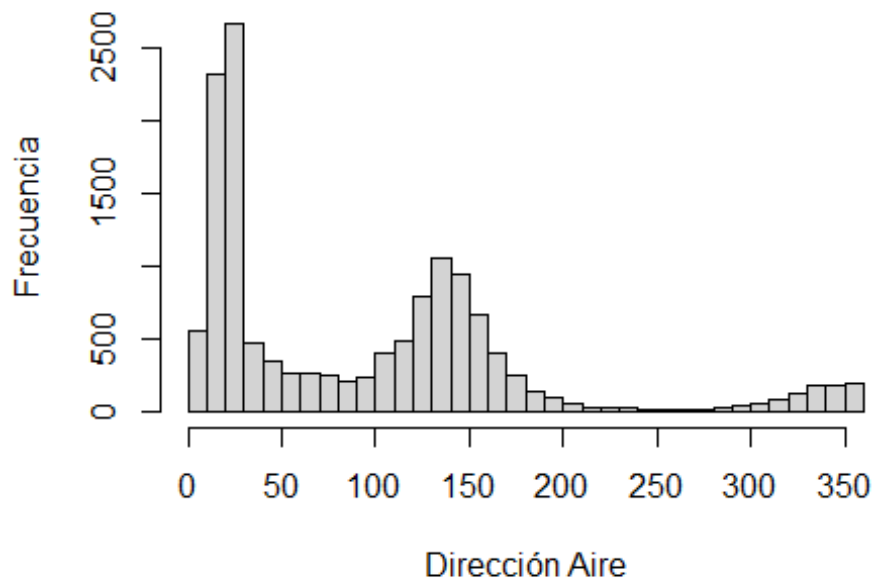


SO2

Velocidad Aire



Dirección Aire

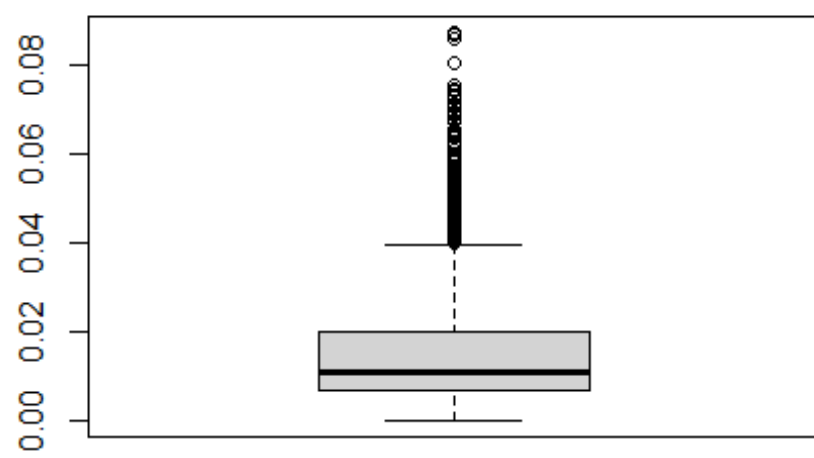


NO₂: Asimétrica
 sesgada a la derecha O₃: Asimétrica sesgada a la derecha PM₁₀: Asimétrica sesgada a la
 derecha SO₂:Asimétrica sesgada a la derecha (Una gran concentración de datos en los
 rangos 0.02 y 0.04) Velocidad aire: Asimétrica sesgada a la derecha(Hay vientos
 mayormente entre 3 y 11 kilómetros por hora) Tiene cierta normalidad. Dirección aire:

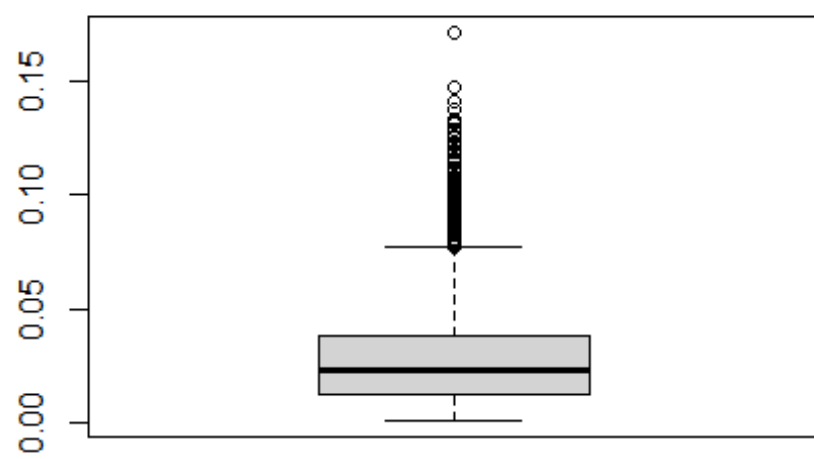
Asimétrica sesgada a la derecha (Se concentra las direcciones entre noreste y sureste)
#3.Exploración a través de gráficas

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

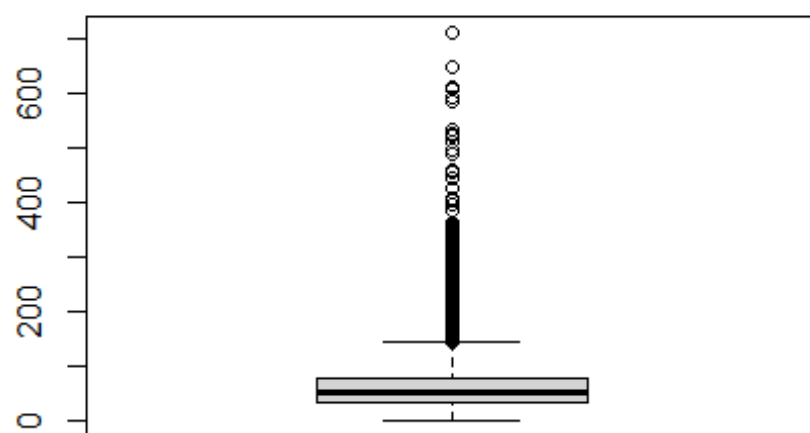
NO2



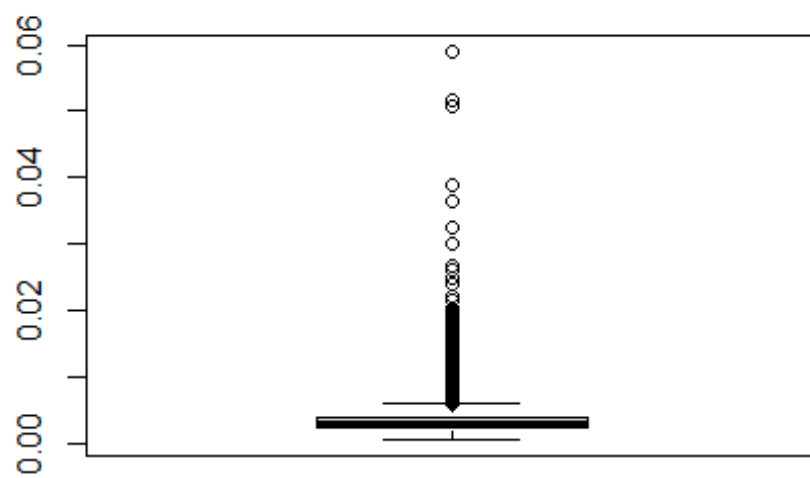
O3



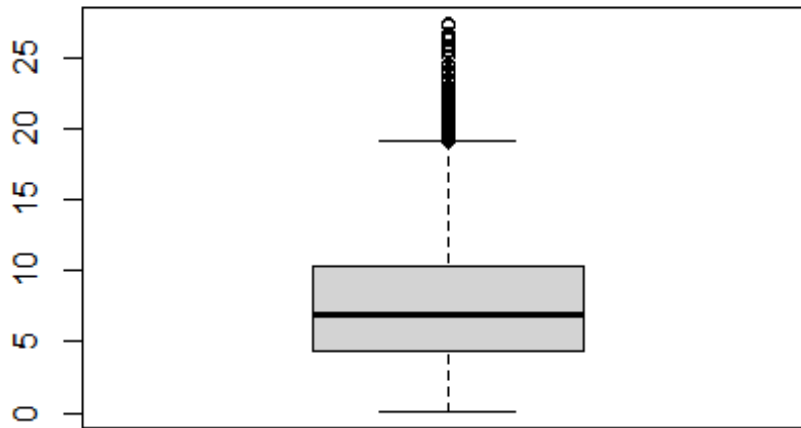
PM10



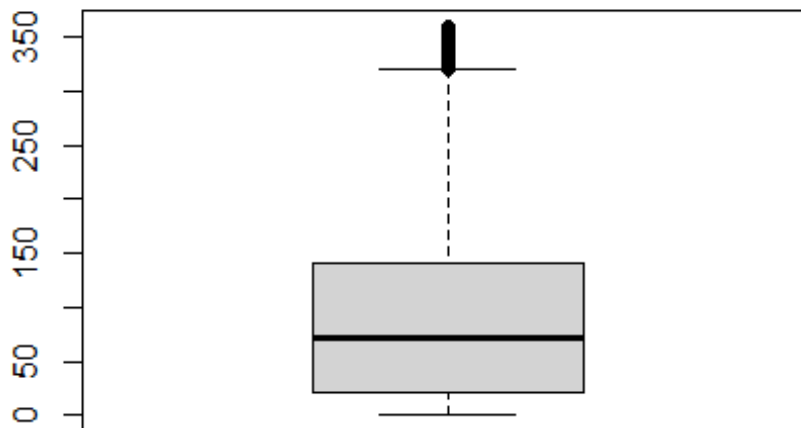
SO2



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
```

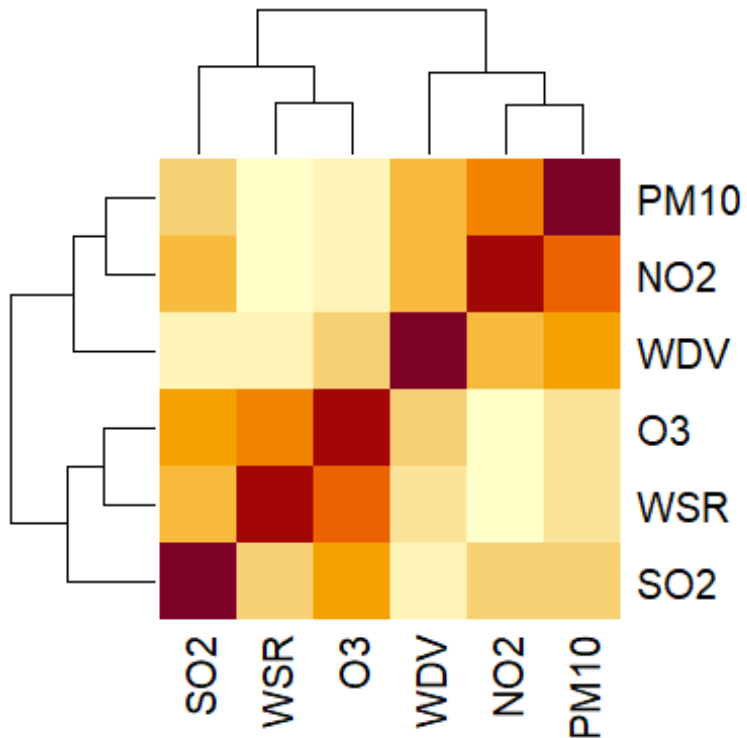
##	N02	O3	PM10	S02	WSR	WDV
## N02	1.00000000	-0.37702117	0.49487058	0.08400291	-0.49060876	0.1389011
## O3	-0.37702117	1.00000000	-0.08032555	0.27678250	0.44155861	0.0266356
## PM10	0.49487058	-0.08032555	1.00000000	0.06038644	-0.22603764	0.2389767
## S02	0.08400291	0.27678250	0.06038644	1.00000000	0.08112461	-0.1486750
## WSR	-0.49060876	0.44155861	-0.22603764	0.08112461	1.00000000	-0.1527164
## WDV	0.13890112	0.02663560	0.23897673	-0.14867499	-0.15271645	1.0000000

```
#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)
```

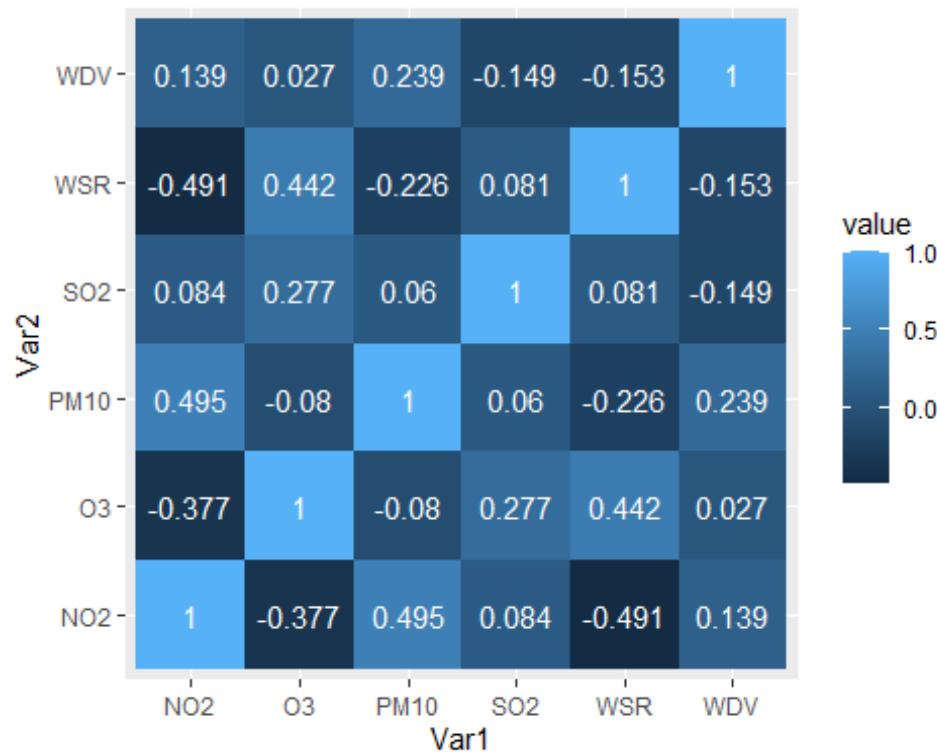


```

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
```