

Medidas Cuantitativas-Sureste

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("03.csv")
```

```
head(dataset)
```

```
##           X      NO2           O3 PM10      SO2 WSR WDV
## 1 2023-01-01 00:00:00 0.0326 0.003000000 110 0.0035 3.2 257
## 2 2023-01-01 01:00:00 0.0303 0.003000000 116 0.0034 3.3 278
## 3 2023-01-01 02:00:00 0.0288 0.003000000 117 0.0036 3.7 278
## 4 2023-01-01 03:00:00 0.0291 0.003000000 135 0.0038 3.6 197
## 5 2023-01-01 04:00:00 0.0257 0.003333333 132 0.0036 4.9 271
## 6 2023-01-01 05:00:00 0.0231 0.003666667  96 0.0030 6.8 284
```

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          1.379591e-02  0.0114  0.0068
## O3           2.849279e-02  0.0260  0.0050
## PM10         5.244695e+01 45.0000 40.0000
## SO2          3.411492e-03  0.0031  0.0024
## Velocidad Aire 1.013673e+01  9.5000  6.3000
## Dirección Aire 1.383083e+02 109.0000 98.0000
```

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          0.0011  0.0774  0.009441434 8.914067e-05
## O3           0.0020  0.1630  0.018324029 3.357700e-04
## PM10         2.0000 800.0000 37.018415101 1.370363e+03
## SO2          0.0005  0.0392  0.001824003 3.326987e-06
## Velocidad Aire 0.3000 31.2000  5.281862077 2.789807e+01
## Dirección Aire 1.0000 360.0000 84.043251427 7.063268e+03
```

##Resultados de tendencia central y medidas de dispersión

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

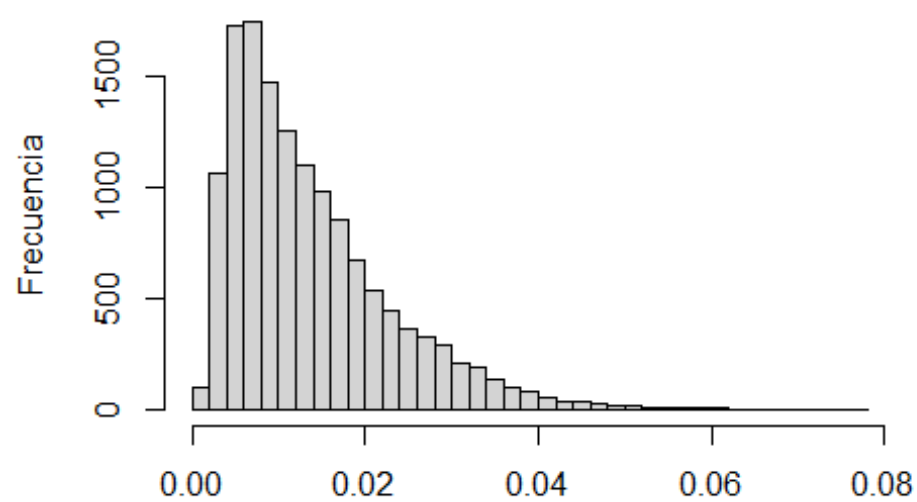
##              Media  Mediana   Moda   Min      Max Desviación
Est.
## NO2          1.379591e-02  0.0114  0.0068 0.0011  0.0774
0.009441434
```

```
## O3                2.849279e-02    0.0260  0.0050 0.0020    0.1630
0.018324029
## PM10             5.244695e+01   45.0000 40.0000 2.0000 800.0000
37.018415101
## SO2              3.411492e-03    0.0031  0.0024 0.0005    0.0392
0.001824003
## Velocidad Aire  1.013673e+01    9.5000  6.3000 0.3000   31.2000
5.281862077
## Dirección Aire  1.383083e+02  109.0000 98.0000 1.0000 360.0000
84.043251427
##                      Varianza
## NO2              8.914067e-05
## O3               3.357700e-04
## PM10             1.370363e+03
## SO2              3.326987e-06
## Velocidad Aire  2.789807e+01
## Dirección Aire  7.063268e+03
```

##Tabla de distribución frecuencia

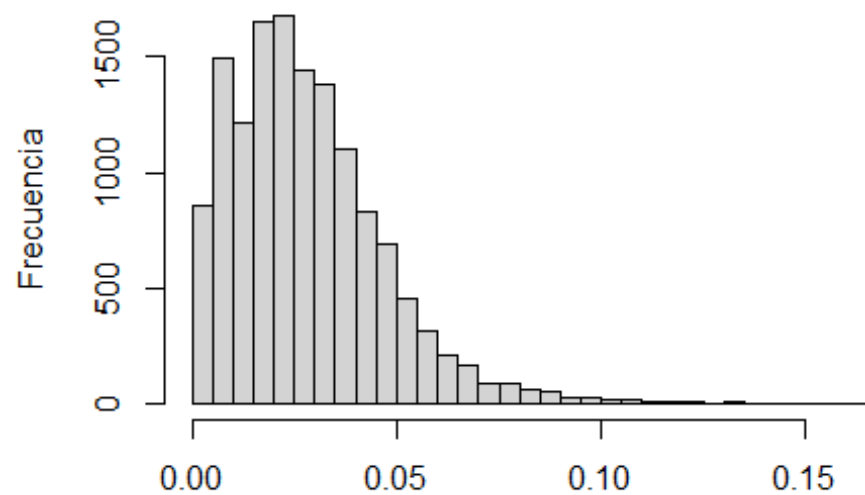
```
titulos <-c("NO2", "O3", "PM10", "SO2", "Velocidad Aire", "Dirección Aire")
for (i in 2:7){
  hist(dataset[,i],breaks = 35, main = titulos[i-1], xlab = titulos[i-1],ylab
= "Frecuencia")
}
```

NO2



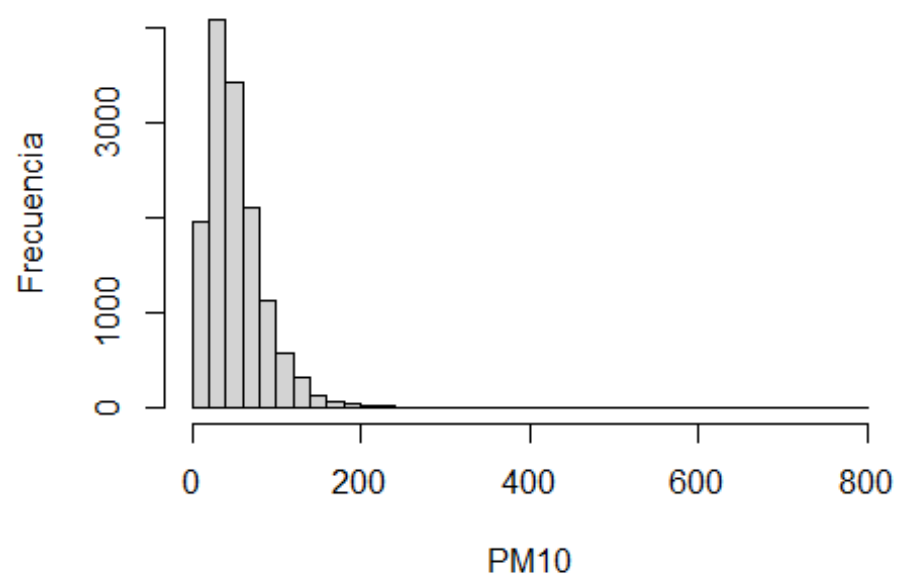
NO2

O3

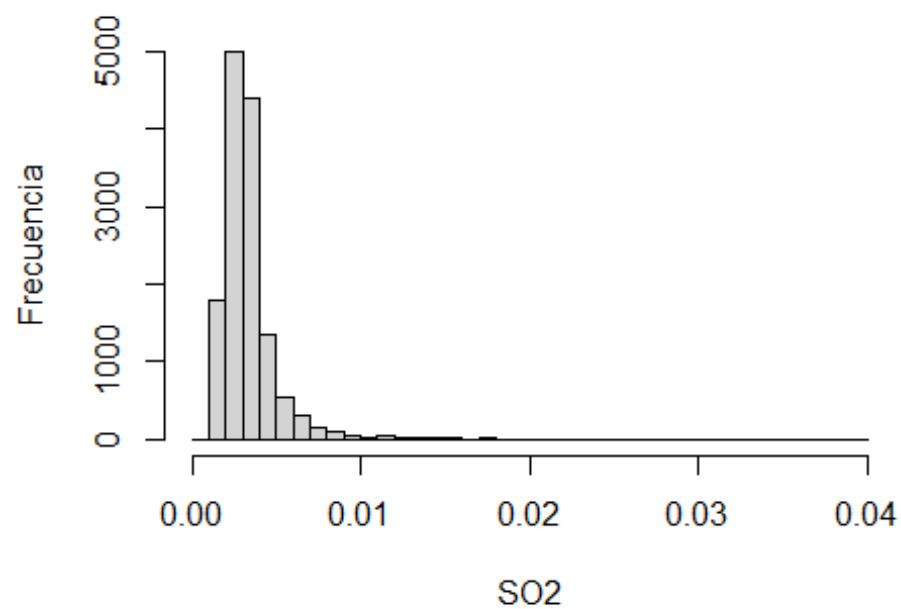


O3

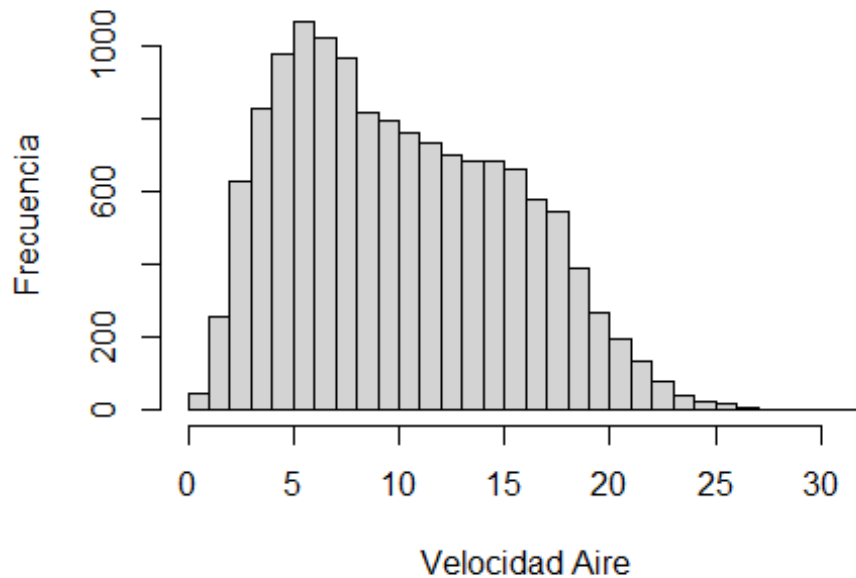
PM10



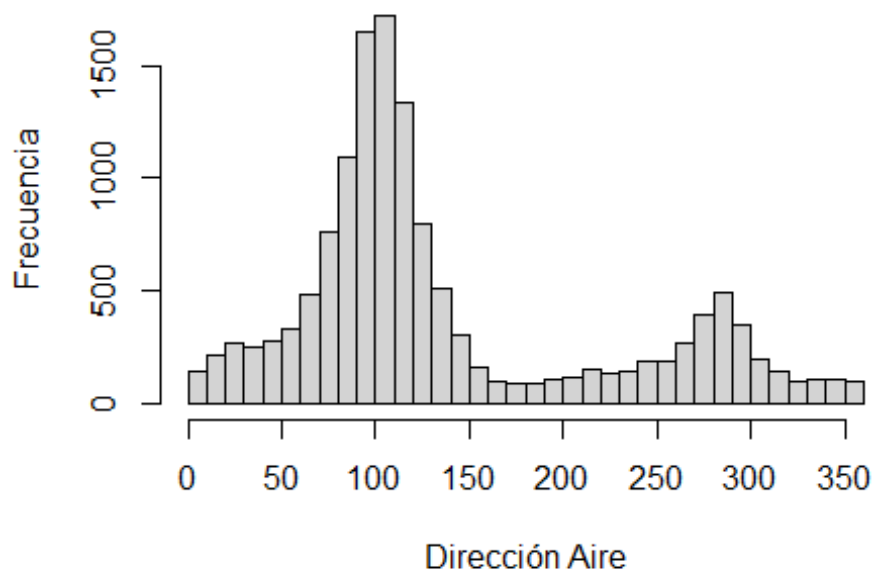
SO2



Velocidad Aire



Dirección Aire

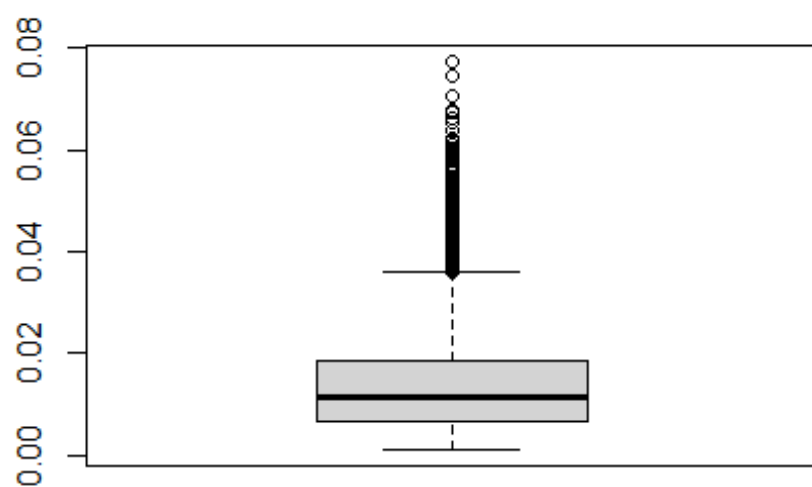


NO₂: Asimétrica
 sesgada a la derecha O₃: Asimétrica sesgada a la derecha PM₁₀: Asimétrica sesgada a la
 derecha SO₂:Asimétrica sesgada a la derecha (Una gran concentración de datos en los
 rangos 0.02 y 0.04) Velocidad aire: Asimétrica sesgada a la derecha(Hay vientos
 mayormente entre 3 y 16 kilómetros por hora) Tiene cierta normalidad. Dirección aire:

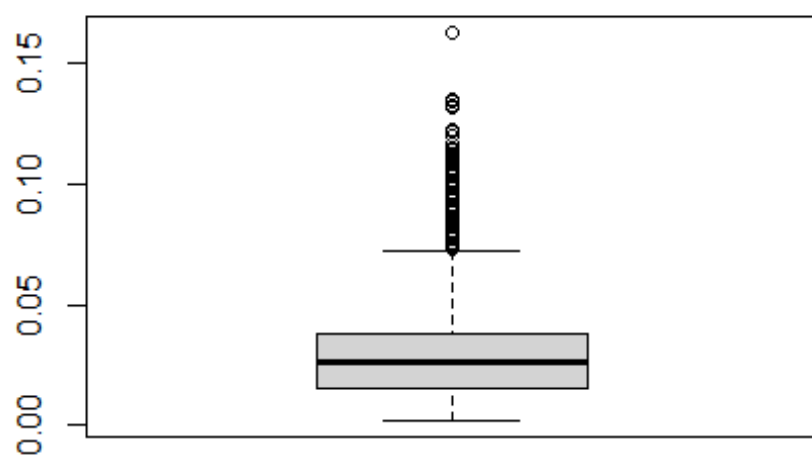
Asimétrica sesgada a la derecha (Se concentra las direcciones en el este) #3.Exploración a través de gráficas

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

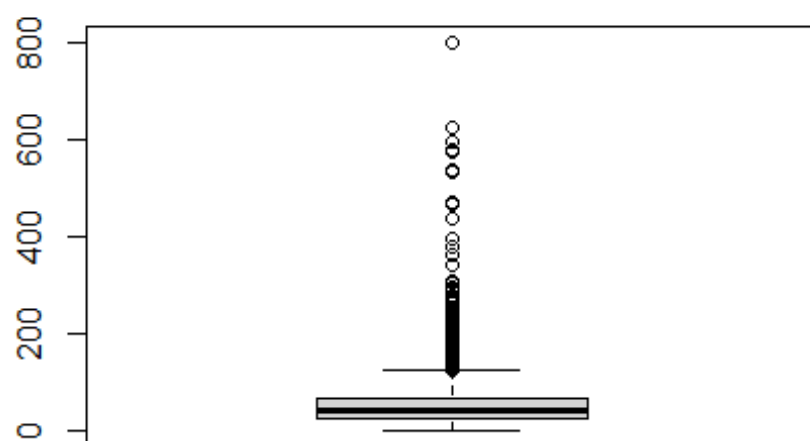
NO2



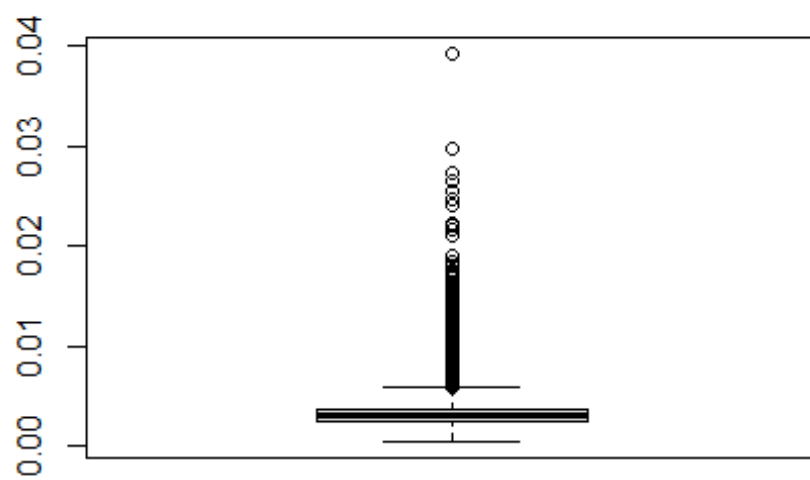
O3



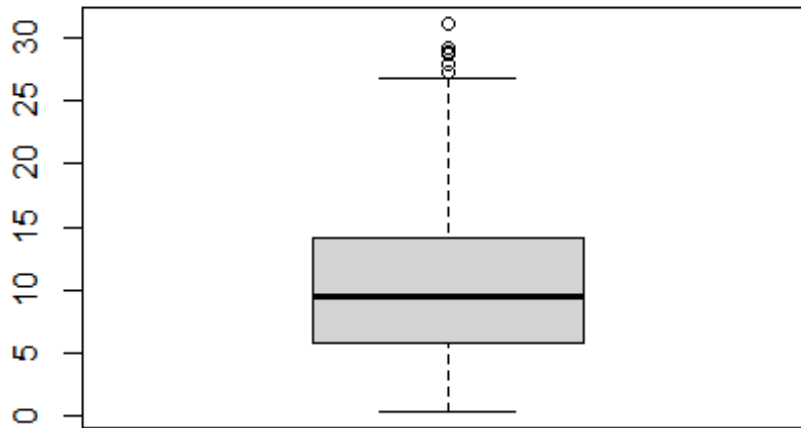
PM10



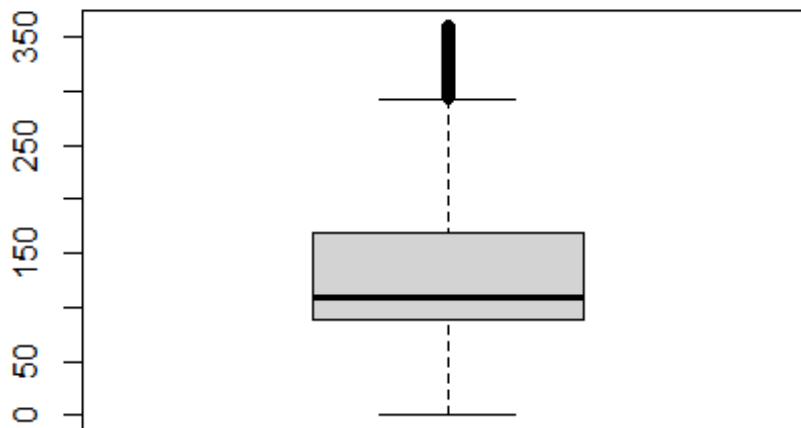
SO2



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
```

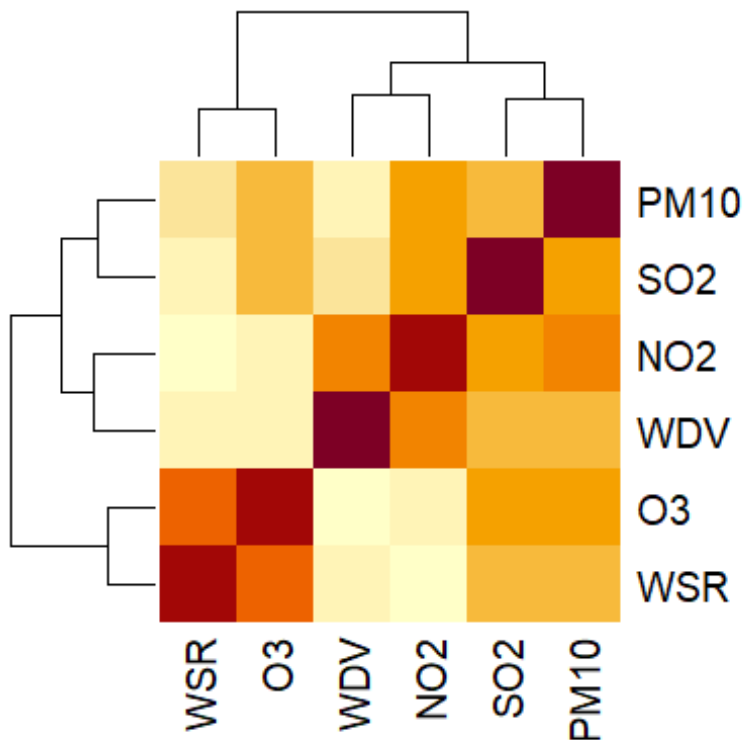
	N02	O3	PM10	SO2	WSR	WDV
N02	1.0000000	-0.2307871	0.36863516	0.338923984	-0.499425638	0.36041783
O3	-0.2307871	1.0000000	0.26454676	0.233549522	0.512077447	-0.32536201
PM10	0.3686352	0.2645468	1.00000000	0.325043331	0.052820489	0.03958864
SO2	0.3389240	0.2335495	0.32504333	1.000000000	0.007139974	0.02398637
WSR	-0.4994256	0.5120774	0.05282049	0.007139974	1.000000000	-0.35937502
WDV	0.3604178	-0.3253620	0.03958864	0.023986371	-0.359375018	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)

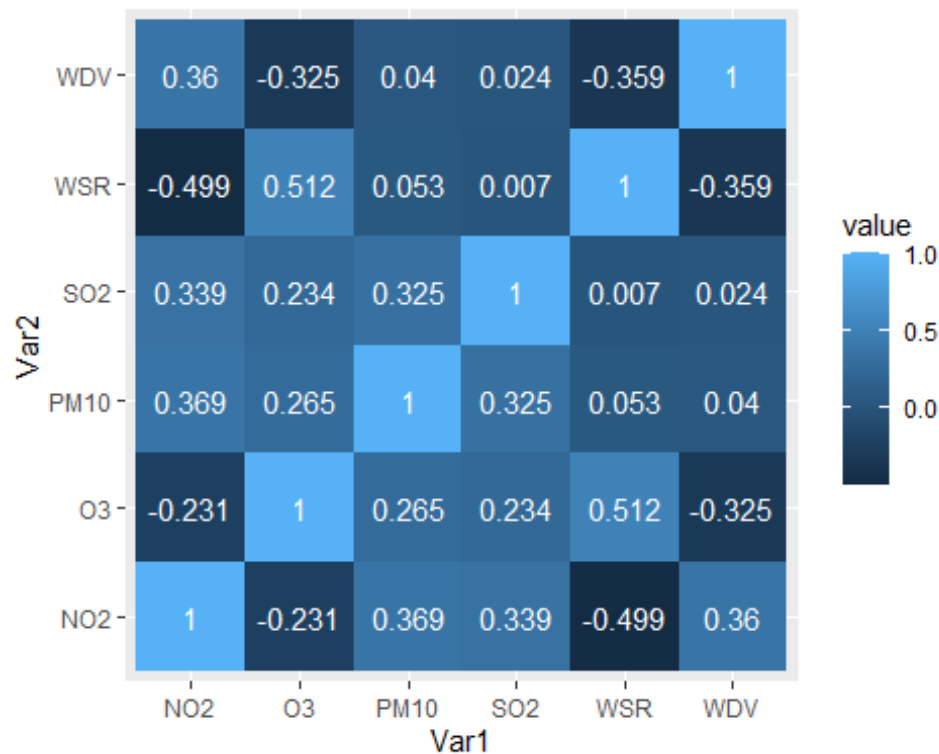
heatmap(cor_matrix)
```



```
titulos <-c("N02", "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                                     # Print
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: 4.389939e-167

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

## Conclusión: Existe una correlación significativa
```