

Medidas Cuantitativas-Centro

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("04.csv")
head(dataset)

##           X      NO2      O3 PM10      SO2 WSR WDV
## 1 2023-01-01 00:00:00 0.0303 0.009  118 0.0033 4.9 236
## 2 2023-01-01 01:00:00 0.0289 0.009   97 0.0034 3.4 336
## 3 2023-01-01 02:00:00 0.0232 0.011  103 0.0034 3.4   3
## 4 2023-01-01 03:00:00 0.0209 0.011   83 0.0032 2.6 315
## 5 2023-01-01 04:00:00 0.0207 0.009   65 0.0030 3.9 270
## 6 2023-01-01 05:00:00 0.0306 0.004   67 0.0031 3.4 242
```

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          1.673600e-02  0.0143  0.0060
## O3           3.043861e-02  0.0260  0.0100
## PM10         6.443397e+01 57.0000 46.0000
## SO2          3.754427e-03  0.0033  0.0023
## Velocidad Aire 7.713825e+00  7.4000  6.8000
## Dirección Aire 1.054995e+02 72.0000 52.0000
```

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          4e-04   0.0922    0.010790959 1.164448e-04
## O3           1e-03   0.1640    0.020396784 4.160288e-04
## PM10         4e+00 791.0000   37.144550875 1.379718e+03
## SO2          5e-04   0.0254    0.001990614 3.962545e-06
## Velocidad Aire 1e-01  24.8000    3.872819203 1.499873e+01
## Dirección Aire 1e+00 360.0000   89.438938902 7.999324e+03
```

##Resultados de tendencia central y medidas de dispersión

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

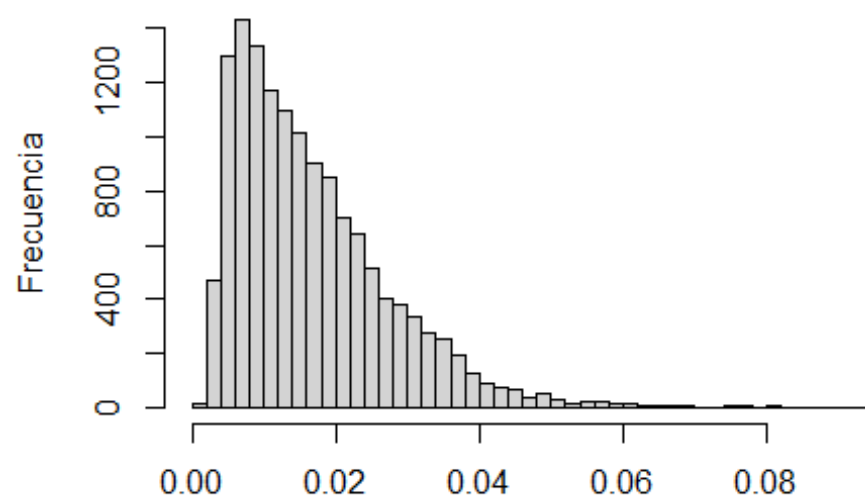
##              Media Mediana   Moda   Min      Max Desviación Est.
## NO2          1.673600e-02  0.0143  0.0060 4e-04   0.0922    0.010790959
## O3           3.043861e-02  0.0260  0.0100 1e-03   0.1640    0.020396784
## PM10         6.443397e+01 57.0000 46.0000 4e+00 791.0000   37.144550875
```

```
## SO2          3.754427e-03  0.0033  0.0023 5e-04  0.0254  0.001990614
## Velocidad Aire 7.713825e+00  7.4000  6.8000 1e-01  24.8000  3.872819203
## Dirección Aire 1.054995e+02 72.0000 52.0000 1e+00 360.0000 89.438938902
##
##          Varianza
## NO2          1.164448e-04
## O3           4.160288e-04
## PM10         1.379718e+03
## SO2          3.962545e-06
## Velocidad Aire 1.499873e+01
## Dirección Aire 7.999324e+03
```

##Tabla de distribución frecuencia

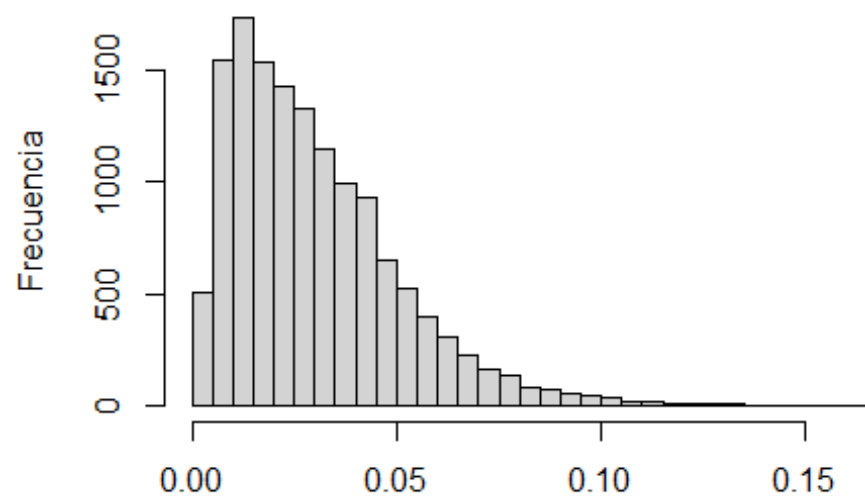
```
titulos <-c("NO2", "O3", "PM10", "SO2", "Velocidad Aire", "Dirección Aire")
for (i in 2:7){
  hist(dataset[,i],breaks = 35, main = titulos[i-1], xlab = titulos[i-1],ylab
= "Frecuencia")
}
```

NO2



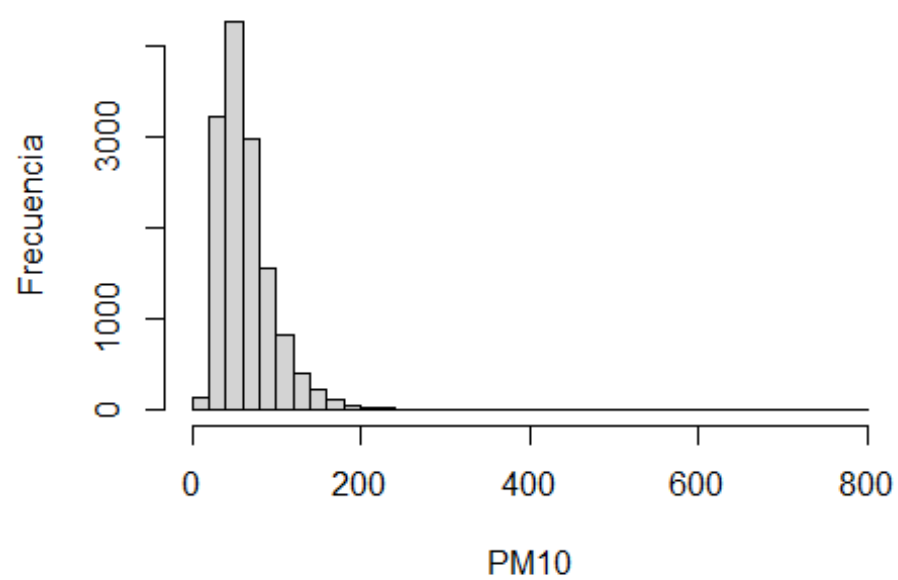
NO2

O3

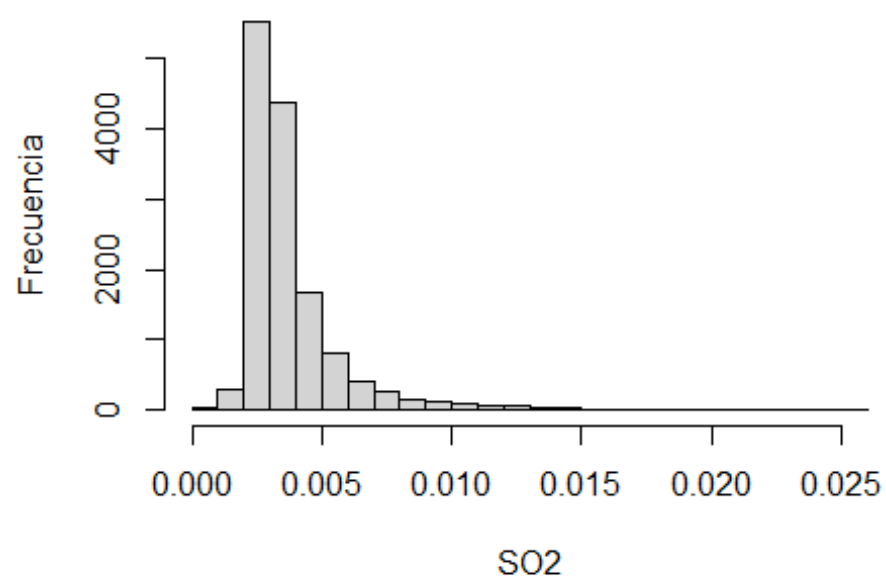


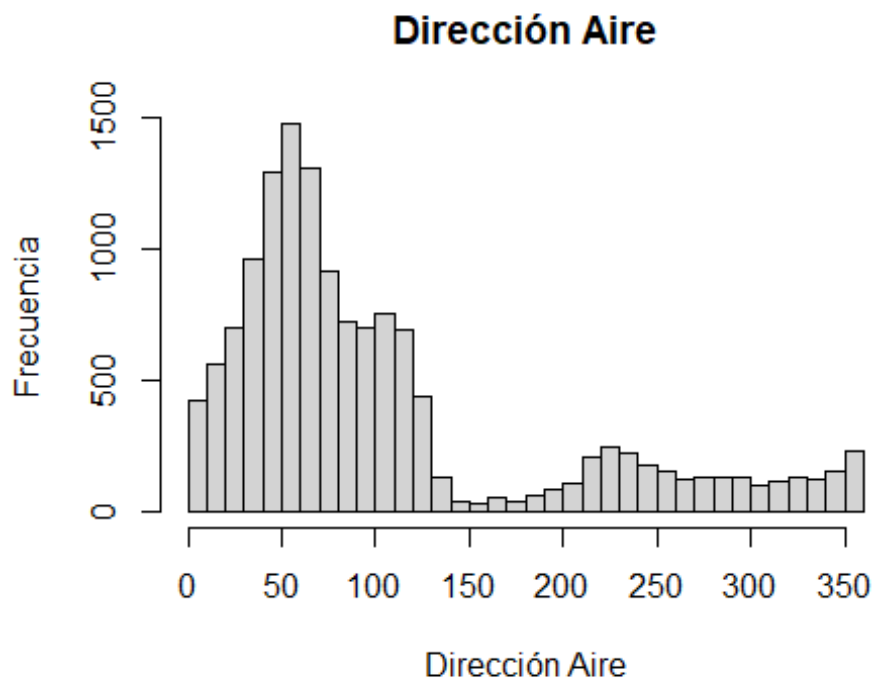
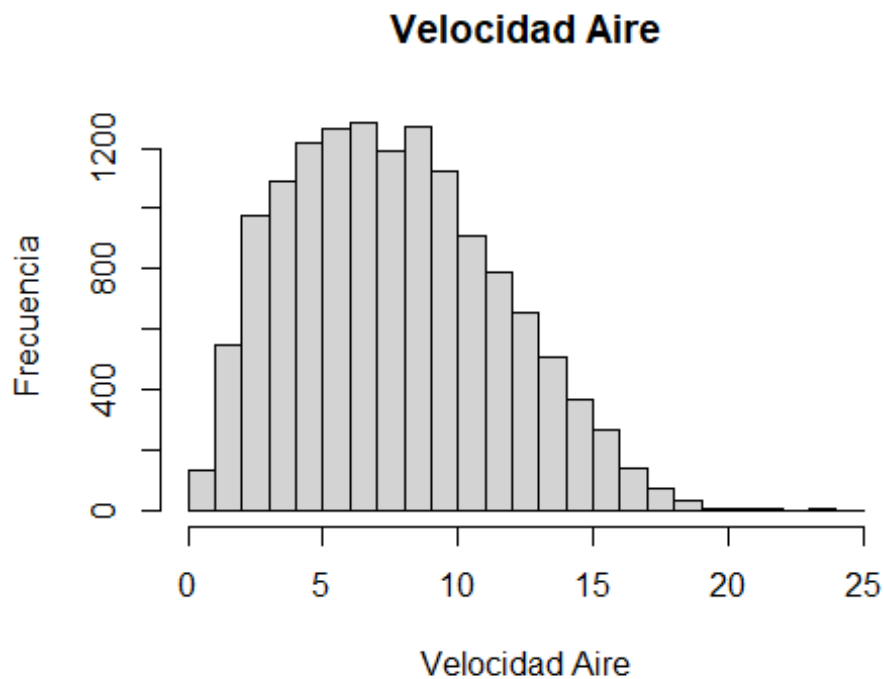
O3

PM10



SO2





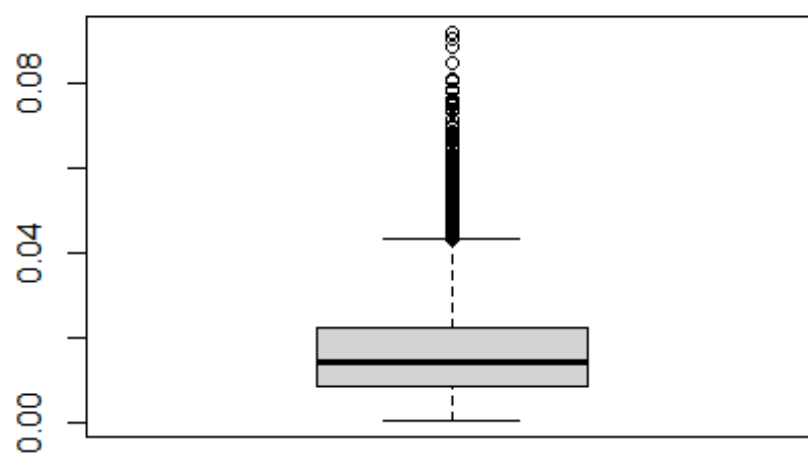
NO₂: Asimétrica
 sesgada a la derecha O₃: Asimétrica sesgada a la derecha PM₁₀: Asimétrica sesgada a la
 derecha SO₂:Asimétrica sesgada a la derecha (Una gran concentración de datos en los
 rangos 0.02 y 0.04) Velocidad aire: Asimétrica sesgada a la derecha(Hay vientos
 mayormente entre 3 y 12 kilómetros por hora) Tiene cierta normalidad Dirección aire:

Asimétrica sesgada a la derecha (Se concentra las direcciones en el noreste y este)

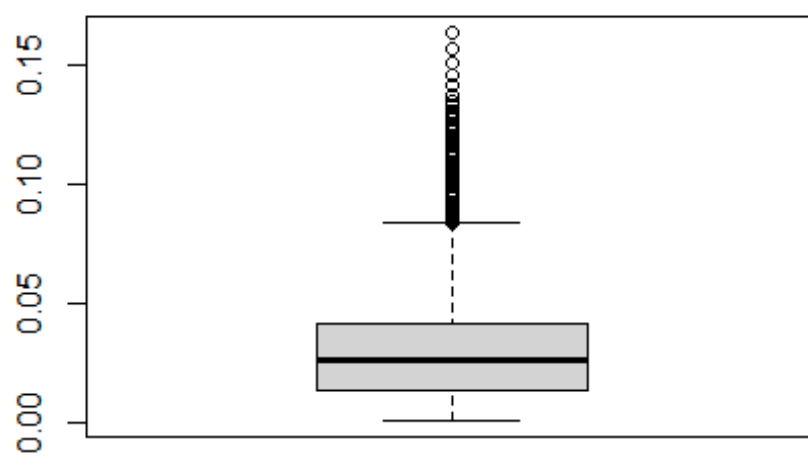
#3.Exploración a través de gráficas

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

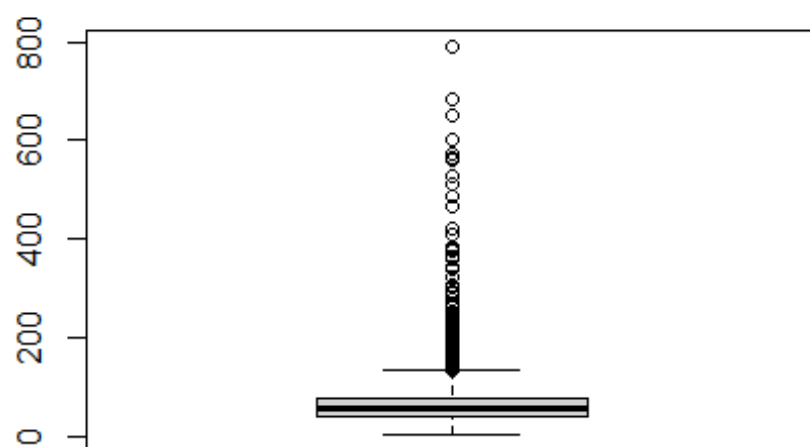
NO2



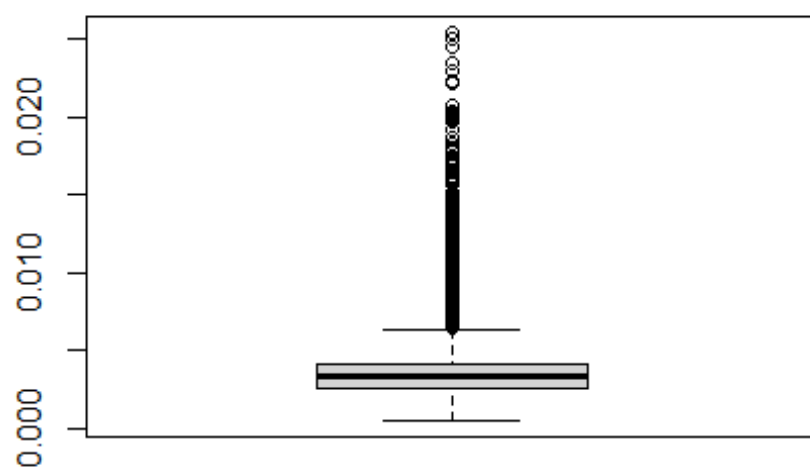
O3



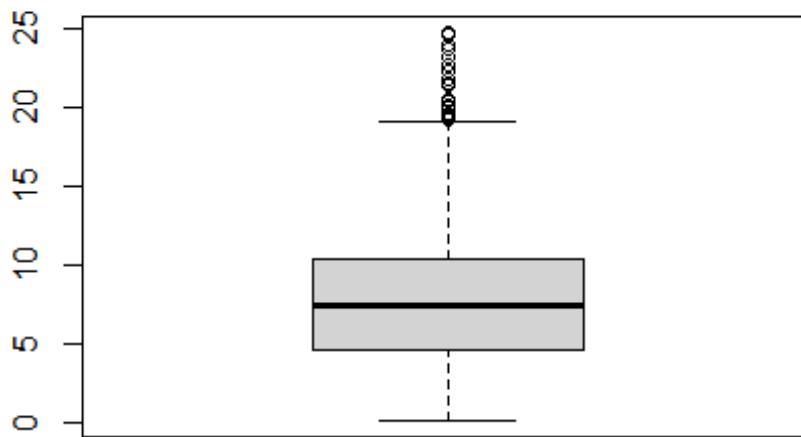
PM10



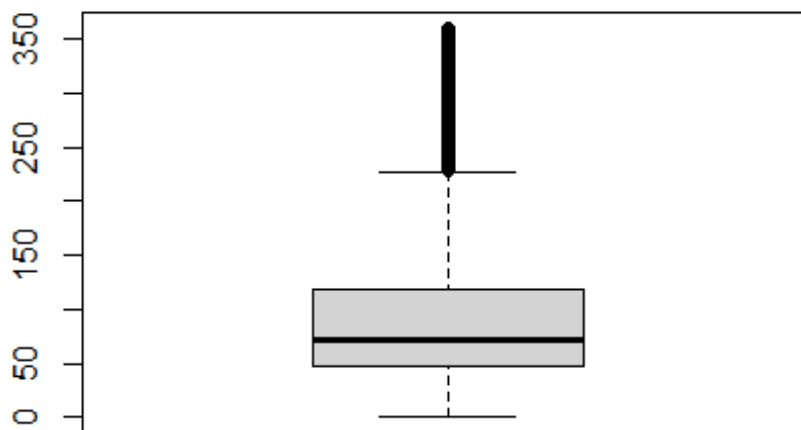
SO2



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
```

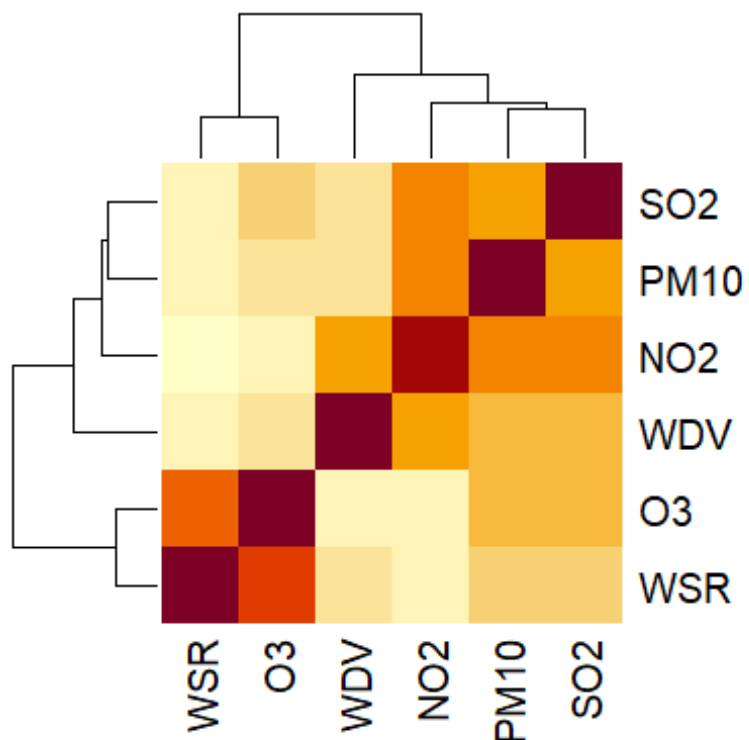
##		N02	O3	PM10	S02	WSR	
WDV							
## N02	1.0000000	-0.28594036	0.47025350	0.48894297	-0.52441292		
	0.23817978						
## O3	-0.2859404	1.00000000	0.07750848	0.10709085	0.55254500	-	
	0.22472648						
## PM10	0.4702535	0.07750848	1.00000000	0.38237290	-0.07155057		
	0.06230386						
## S02	0.4889430	0.10709085	0.38237290	1.00000000	-0.07625926		
	0.04247430						
## WSR	-0.5244129	0.55254500	-0.07155057	-0.07625926	1.00000000	-	
	0.31006567						
## WDV	0.2381798	-0.22472648	0.06230386	0.04247430	-0.31006567		
	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)
```

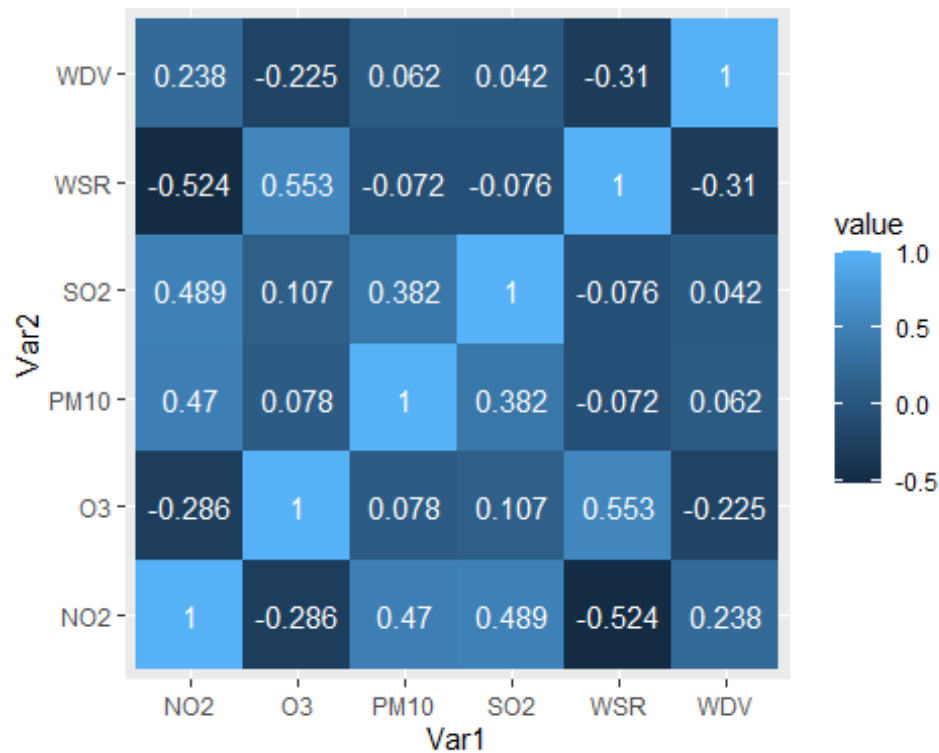


```

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: 3.209548e-259

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

## Conclusión: Existe una correlación significativa
```