

# Medidas Cuantitativas

Diego Gutierrez Vargas

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("00.csv")
head(dataset)
```

```
##              X      NO2      O3 PM10      SO2 WSR WDV
## 1 2023-01-01 00:00:00 0.0498 0.014  228 0.0045 5.5 221
## 2 2023-01-01 01:00:00 0.0473 0.013  478 0.0058 2.2 272
## 3 2023-01-01 02:00:00 0.0431 0.011  566 0.0060 3.7 221
## 4 2023-01-01 03:00:00 0.0421 0.011  498 0.0061 2.5 290
## 5 2023-01-01 04:00:00 0.0404 0.011  483 0.0050 1.9 280
## 6 2023-01-01 05:00:00 0.0373 0.009  452 0.0045 4.4 246
```

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      0.023706510 0.0215 0.0138
## O3       0.026728947 0.0220 0.0110
## PM10     66.317051190 58.0000 68.0000
## SO2      0.004066471 0.0034 0.0030
## Velocidad Aire 7.632725306 7.3000 6.5000
## Dirección Aire 99.421124730 78.0000 92.0000
```

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 A
colnames(resultados_medidas_dispersion) <- c("Min", "Max", "Desviación Est.,"Varianza")

resultados_medidas_dispersion
```

```
##           Min      Max Desviación Est.    Varianza
## NO2      0.0004   0.1260   0.011382177 1.295539e-04
## O3       0.0010   0.1630   0.017363726 3.014990e-04
## PM10     2.0000 705.0000  44.246743875 1.957774e+03
## SO2      0.0012   0.0696   0.002245962 5.044345e-06
## Velocidad Aire 0.1000 27.7000   3.305688702 1.092758e+01
## Dirección Aire 1.0000 360.0000  89.429733460 7.997677e+03
```

##Resultados de tendencia central y medidas de dispersión

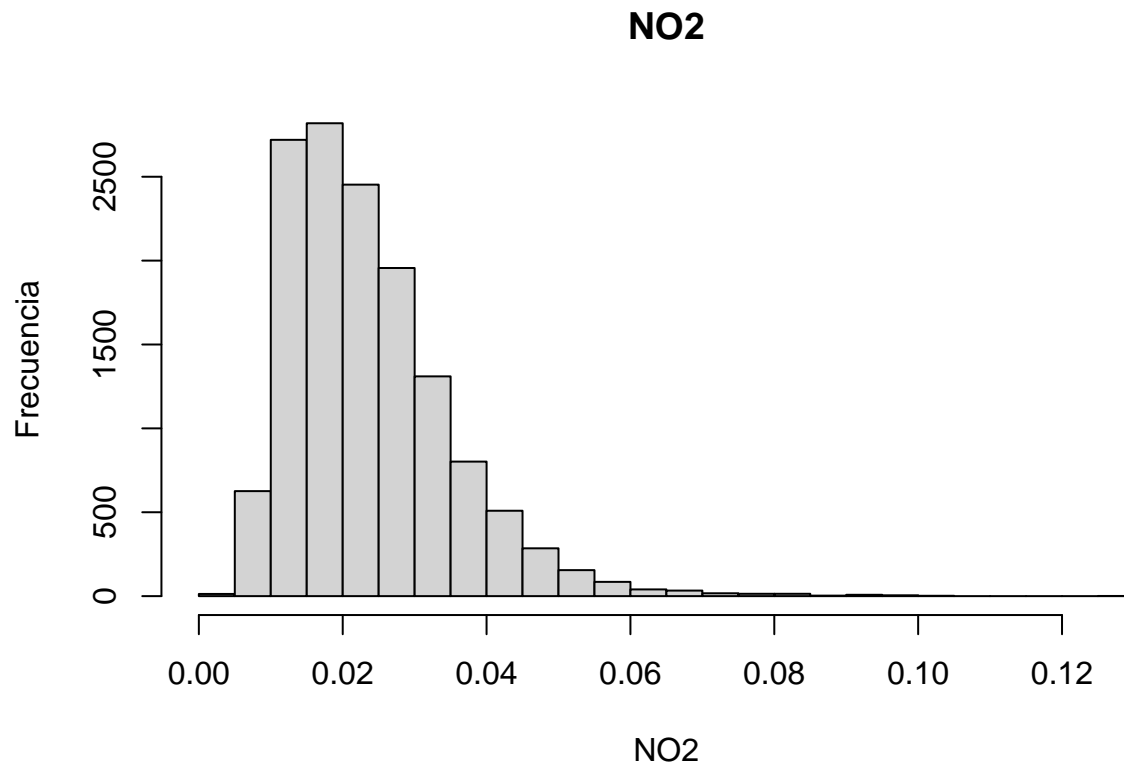
```
print(cbind(resultados_tendencia_central,resultados_medidas_dispersion))
```

```
##           Media Mediana   Moda   Min      Max Desviación Est.
## NO2      0.023706510 0.0215 0.0138 0.0004   0.1260   0.011382177
## O3       0.026728947 0.0220 0.0110 0.0010   0.1630   0.017363726
## PM10     66.317051190 58.0000 68.0000 2.0000 705.0000  44.246743875
## SO2      0.004066471 0.0034 0.0030 0.0012   0.0696   0.002245962
## Velocidad Aire 7.632725306 7.3000 6.5000 0.1000 27.7000   3.305688702
## Dirección Aire 99.421124730 78.0000 92.0000 1.0000 360.0000  89.429733460
##           Varianza
## NO2      1.295539e-04
## O3       3.014990e-04
```

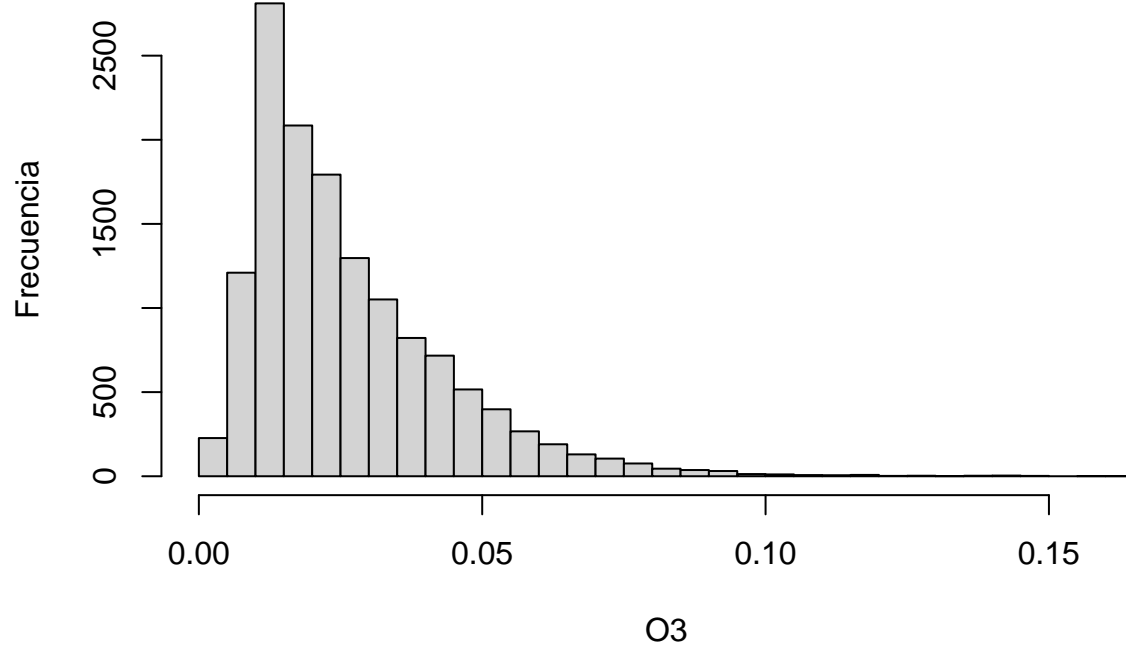
```
## PM10          1.957774e+03
## SO2           5.044345e-06
## Velocidad Aire 1.092758e+01
## Dirección Aire 7.997677e+03
```

```
##Tabla de distribución frecuencia
```

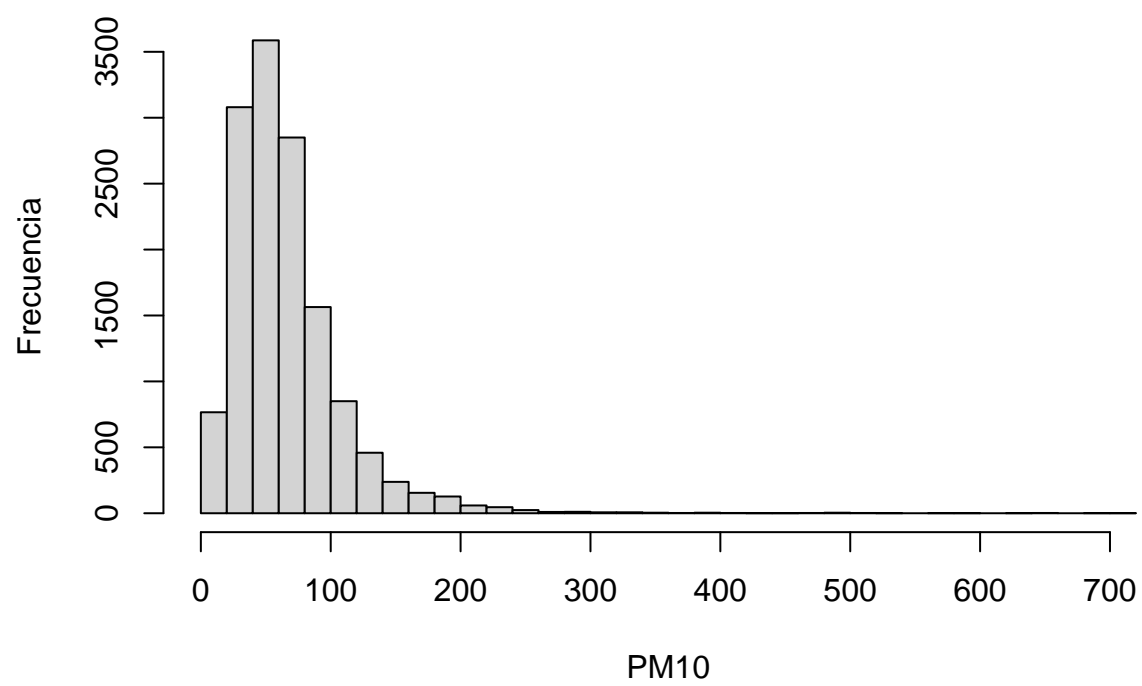
```
titulos <-c("NO2", "O3", "PM10", "SO2", "Velocidad Aire", "Dirección Aire")
for (i in 2:7){
  hist(dataset[,i],breaks = 35, main = titulos[i-1], xlab = titulos[i-1],ylab = "Frecuencia")
}
```



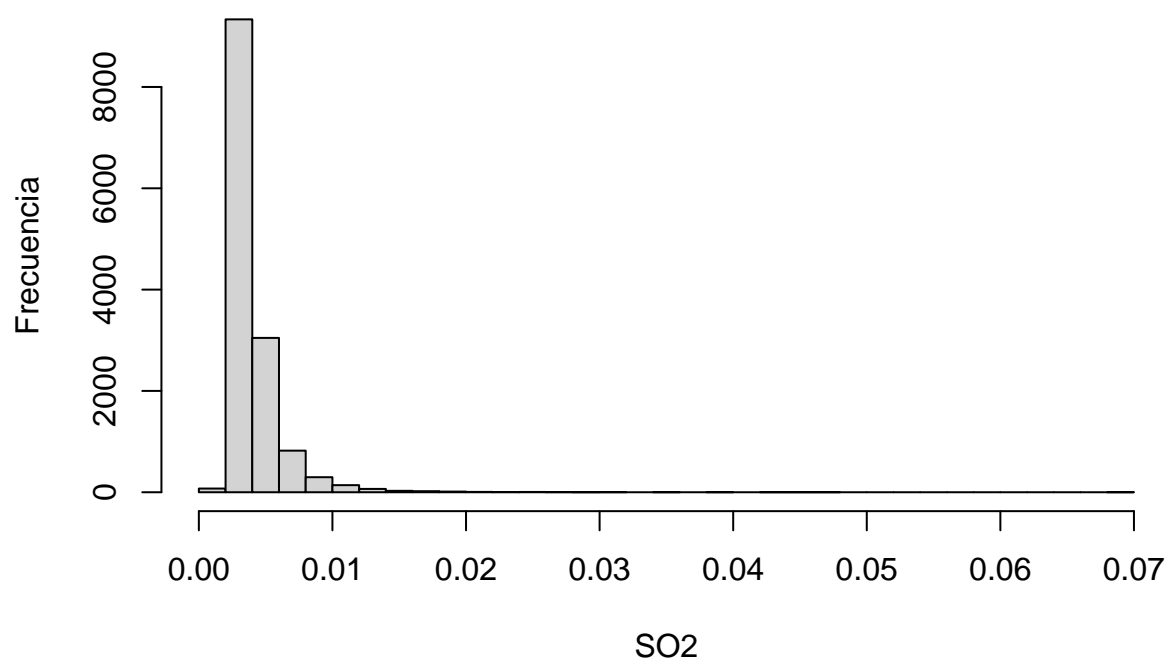
**O3**



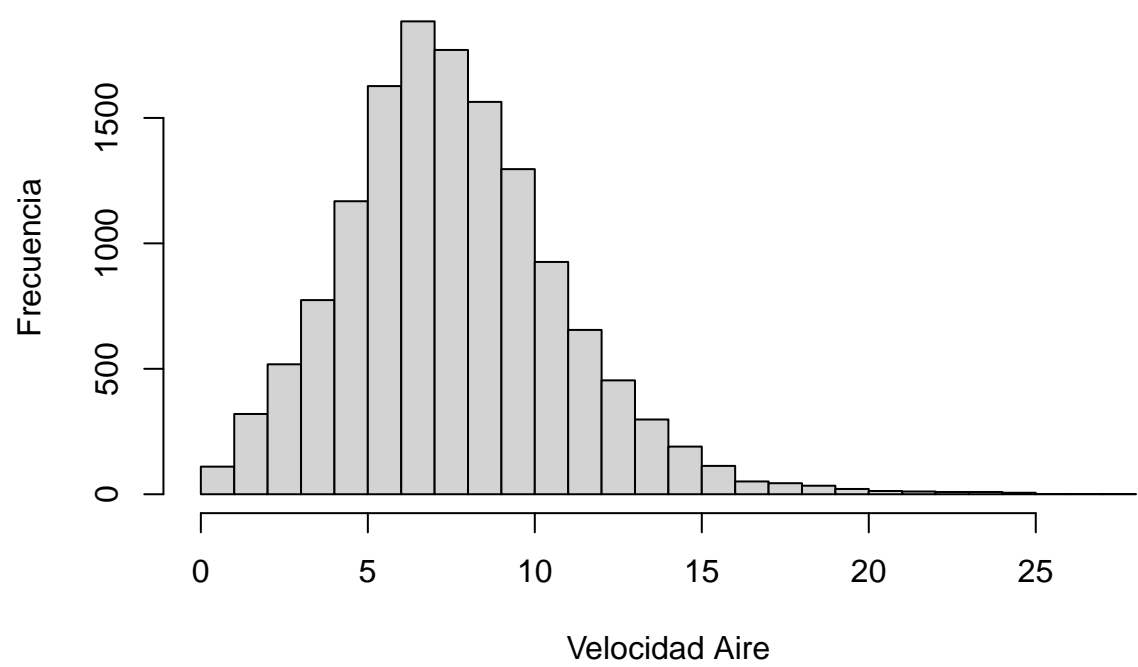
## PM10

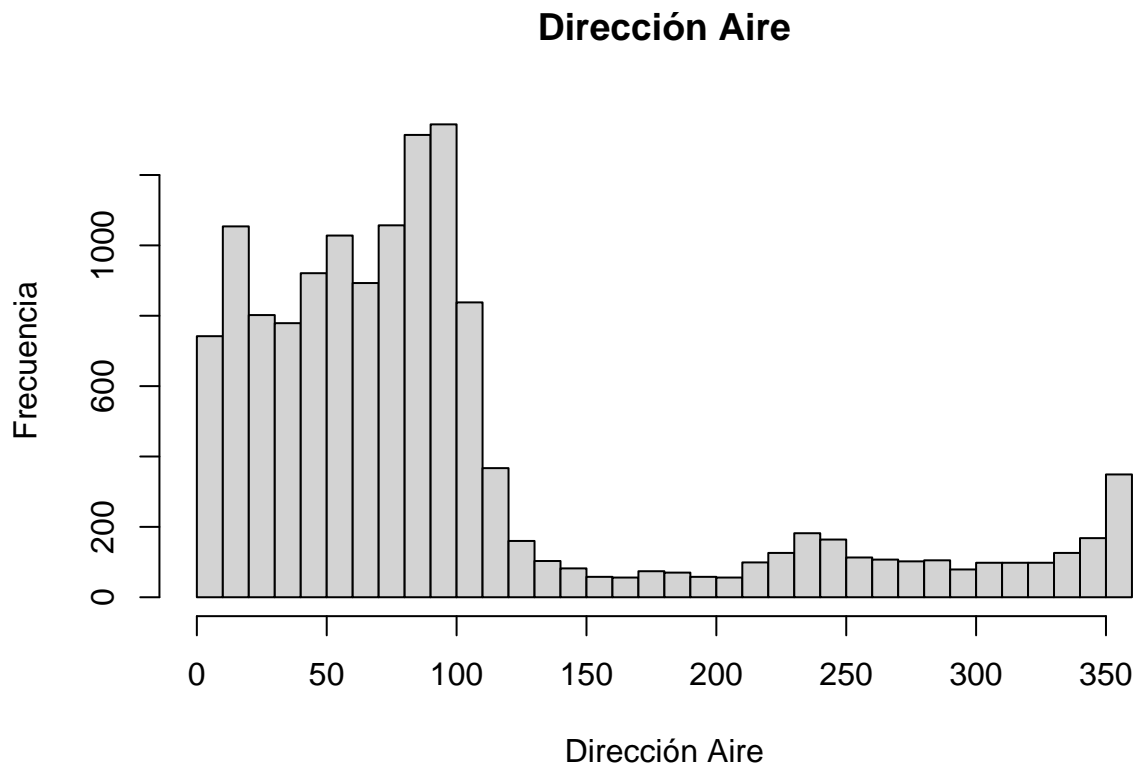


**SO2**



**Velocidad 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 0.2 y 0.4)  
 Velocidad aire: Asimétrica sesgada a la derecha(Hay vientos mayormente entre 4 y 11 kilómetros por hora)  
 Dirección aire: sesgada a la derecha (Se concentra las direcciones entre norte y el este) #3.Exploración a través de gráficas

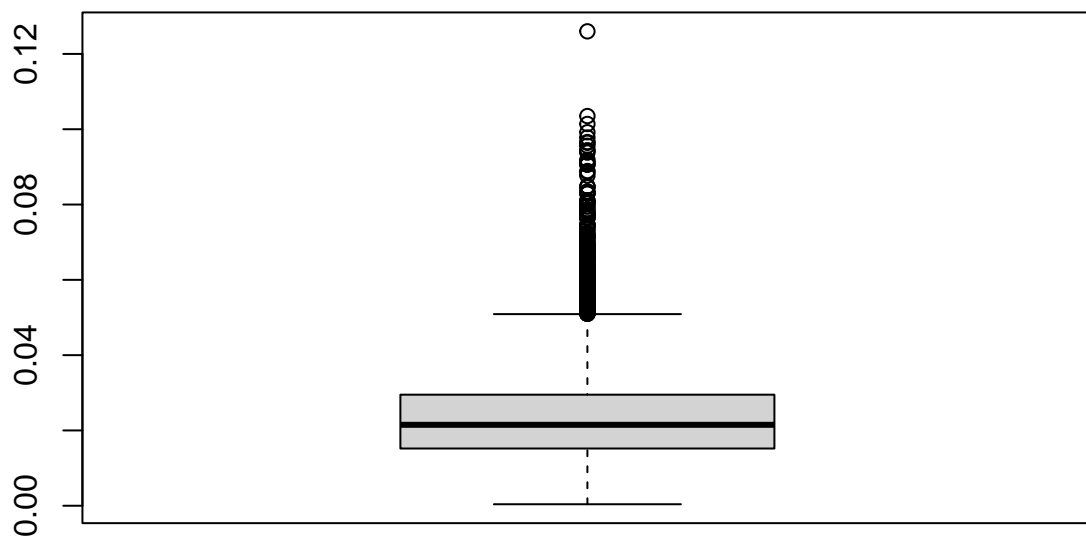
```

titulos <- c("NO2", "O3", "PM10", "SO2", "Velocidad Aire", "Dirección Aire")
for (i in 2:7){
  boxplot(dataset[,i], main = titulos[i-1])
}

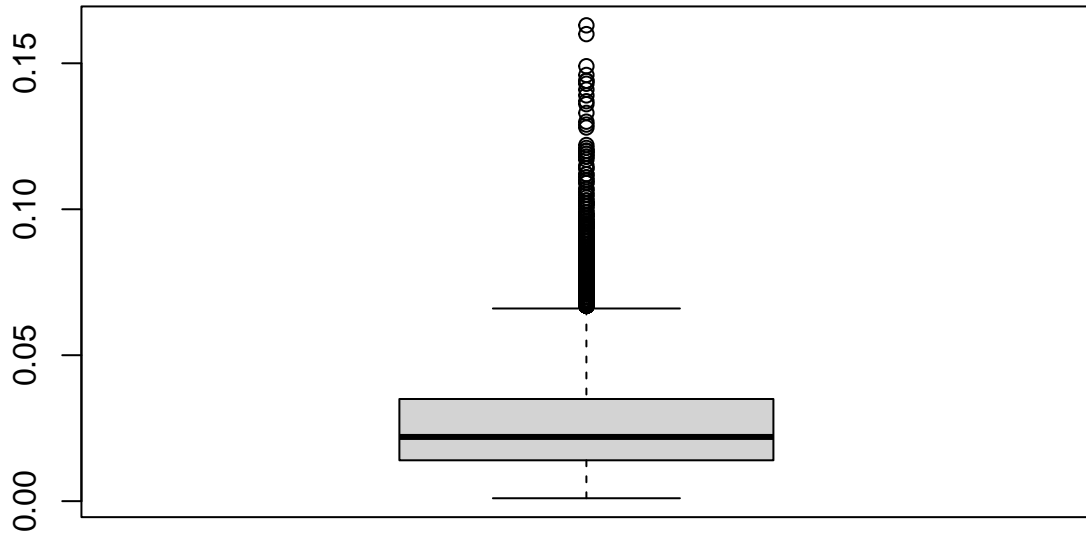
```



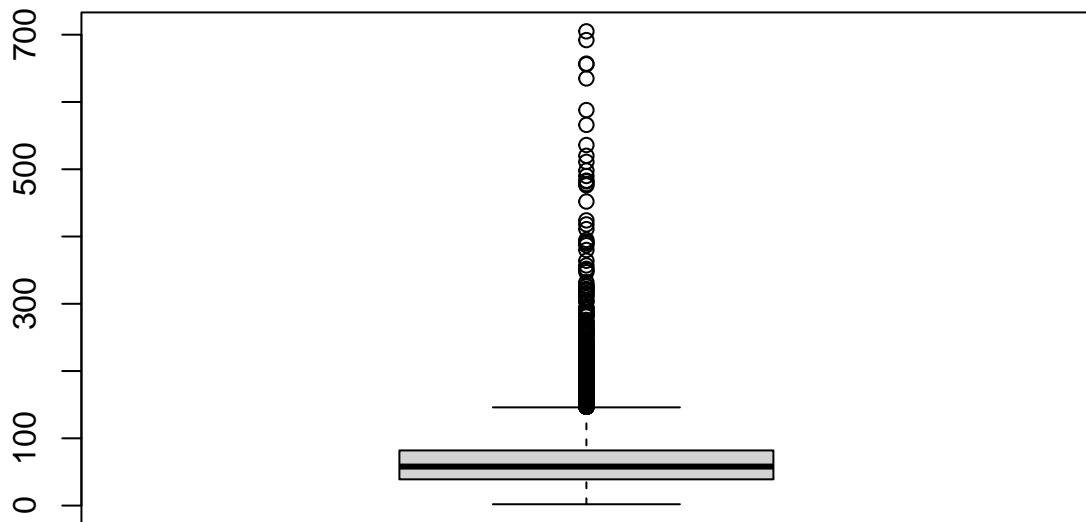
NO2



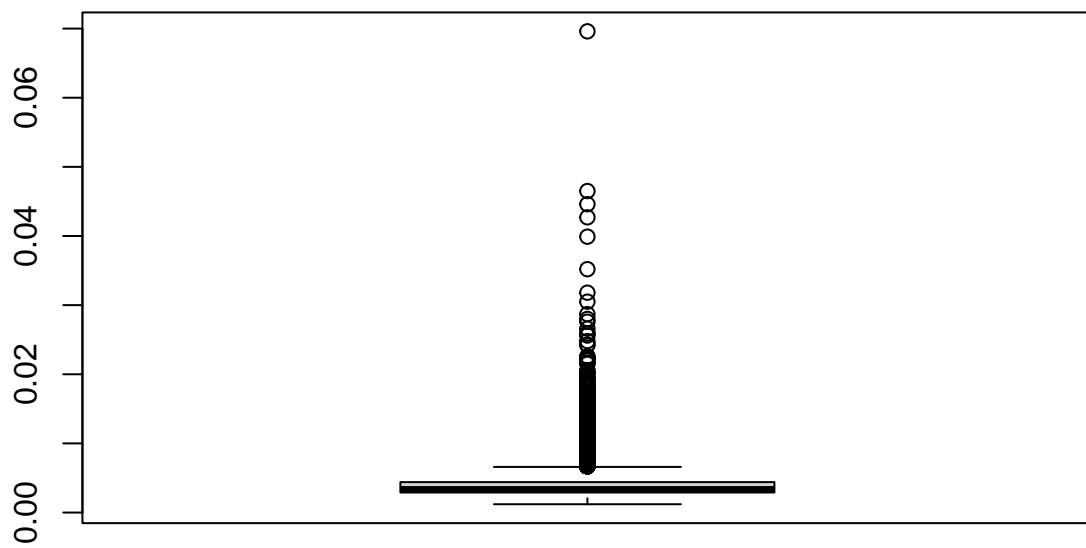
O3



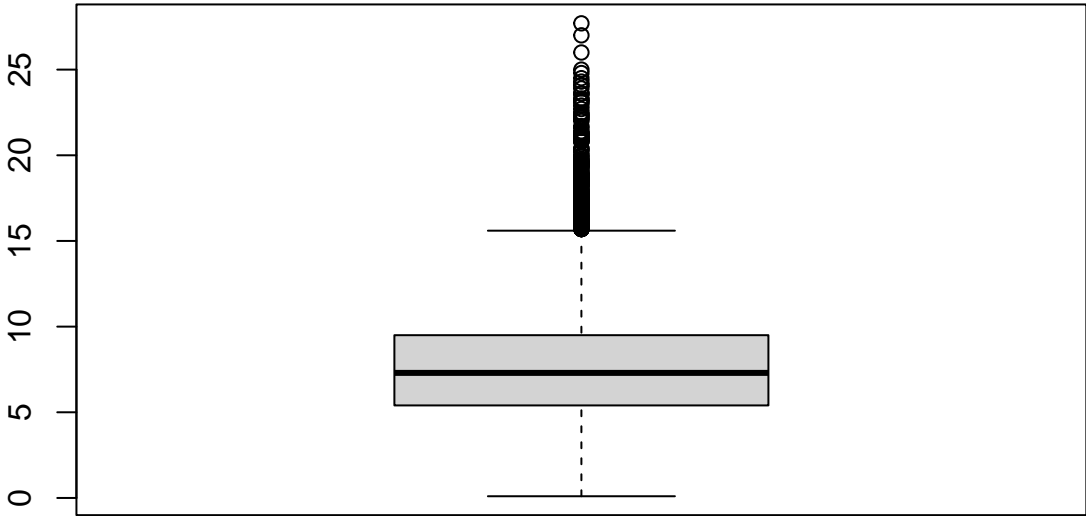
## PM10



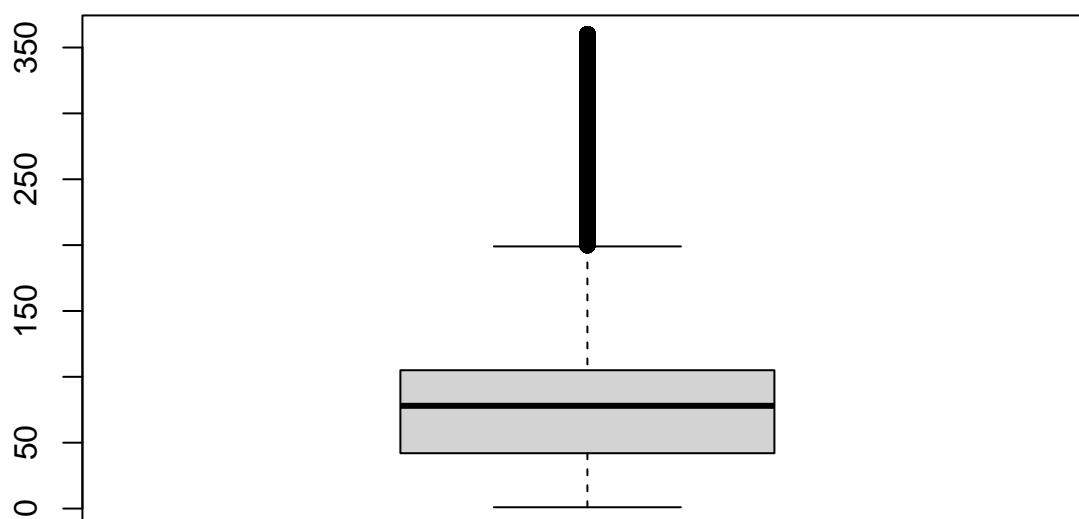
SO2



Velocidad Aire



## Dirección Aire



Muchos valores atípicos observados a través de los boxplots, especialmente en la dirección del aire.  
##Correlación

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

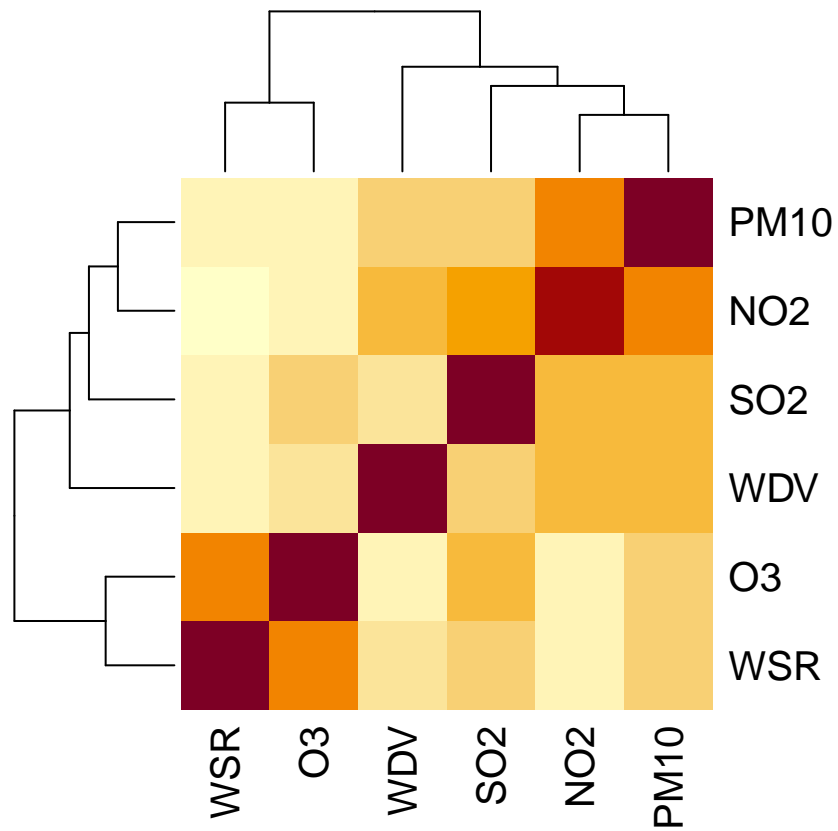
```
##          NO2          O3          PM10          SO2          WSR          WDV
## NO2    1.0000000 -0.1842415  0.53330803  0.29457645 -0.37750623  0.19995157
## O3     -0.1842415  1.0000000  0.01198340  0.18455484  0.38049757 -0.12991144
## PM10   0.5333080  0.0119834  1.00000000  0.23463937 -0.04272143  0.18702230
## SO2    0.2945764  0.1845548  0.23463937  1.00000000 -0.06277424  0.02639781
## WSR   -0.3775062  0.3804976 -0.04272143 -0.06277424  1.00000000 -0.19615804
## WDV    0.1999516 -0.1299114  0.18702230  0.02639781 -0.19615804  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)
```

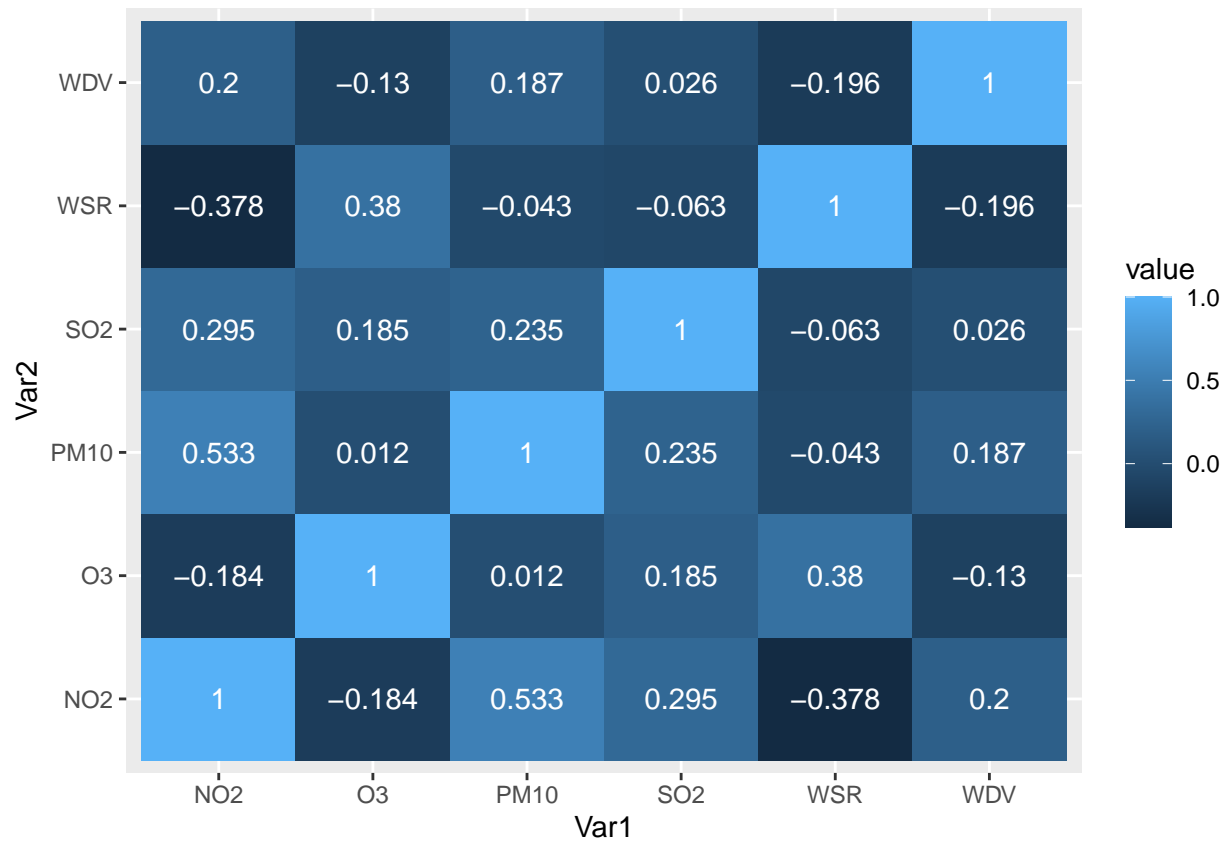


```

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))
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")
cat("Para las variables",colnames(corr_dataframe)[1:2],"\n")
```

## Para las variables NO2 O3

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

## Valor p: 3.696116e-106

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

## Conclusión: Existe una correlación significativa