

# Tecnológico de Monterrey

18 de Agosto de 2023

**Aplicación de métodos multivariados en ciencia de datos**

MA2003B.101

## Conocimiento de la naturaleza de contaminantes que influyen en la calidad del aire y sus interrelaciones con el medio

### Profesoras

Monica Guadalupe Elizondo Amaya

Blanca Rosa Ruiz Hernandez

### Autores

#### (Equipo 6)

Rubén Darío Castro Terrazas (A00833945)

Christian Jaffe Alarcón Acosta (A00832881)

José Andrés Orantes Guillén (A01174130)

Juan José Montes Raygoza (A00834630)

Pedro Fernández Merino (A01733006)

### Importar librerías y dependencias

```
##
## Attaching package: 'Hmisc'

## The following objects are masked from 'package:base':
##
##   format.pval, units

## corrplot 0.92 loaded

##
## Attaching package: 'dplyr'

## The following object is masked from 'package:gridExtra':
##
##   combine
```

```
## The following objects are masked from 'package:Hmisc':  
##  
##      src, summarize  
  
## The following objects are masked from 'package:stats':  
##  
##      filter, lag  
  
## The following objects are masked from 'package:base':  
##  
##      intersect, setdiff, setequal, union
```

## Etapa I

La problemática de la contaminación del aire se ha erigido como un tema de suma importancia en la agenda global, debido a sus profundos impactos en la salud humana, el entorno ambiental y el fenómeno del cambio climático. La calidad del aire se refiere a la composición y pureza de los gases que conforman la atmósfera que respiramos. Esta contaminación puede originarse de diversas fuentes, tales como la quema de combustibles fósiles, las emisiones industriales, el tráfico vehicular y las prácticas agrícolas. En el contexto específico de Nuevo León y su zona metropolitana, estas cuatro actividades ejercen un notorio impacto.

A medida que estas fuentes emiten contaminantes al aire, sus efectos pueden acarrear consecuencias graves para la sociedad y el planeta en su conjunto. De acuerdo con la Organización Mundial de la Salud, la contaminación del aire representa uno de los principales riesgos ambientales para la salud, desencadenando una carga significativa de enfermedades como accidentes cerebrovasculares, enfermedades cardíacas, cáncer de pulmón y diversas afecciones pulmonares crónicas y agudas, incluyendo el asma. De hecho, estas enfermedades se asocian con un número anual alarmante de 6.7 millones de muertes prematuras [1].

Por otra parte, estadísticas proporcionadas por la National Weather Service han calculado que los costos económicos relacionados con esta contaminación ascienden a 150 mil millones de dólares anuales, únicamente en los Estados Unidos [2]. De esta manera, resulta claro que por el bienestar de la economía y, lo que es aún más primordial, por la preservación de nuestra salud, se torna imperativo salvaguardar la calidad del aire que nos rodea.

Es fundamental comprender los aspectos cruciales que se consideran al analizar la calidad del aire y familiarizarse con las clasificaciones internacionales de los contaminantes más relevantes, así como identificar los contaminantes más perjudiciales para la salud. Esta información enriquecerá considerablemente el contenido de este informe. Entre los aspectos esenciales para evaluar la calidad del aire se destacan:

**§Partículas en suspensión (PM2.5 y PM10):** Estas partículas diminutas suspendidas en el aire tienen un impacto significativo en la salud. Las PM2.5 son aquellas con un diámetro

igual o menor a 2.5 micrómetros, mientras que las PM10 tienen un diámetro igual o menor a 10 micrómetros [3].

**§ Niveles elevados de ozono (O<sub>3</sub>):** Cuando el ozono, un compuesto que nos protege de los rayos ultravioleta en la estratósfera, desciende a niveles del suelo, provoca contaminación atmosférica perjudicial, conocida como “smog” en inglés. Esto tiene consecuencias negativas para la salud comunitaria [4].

**§ Dióxido de nitrógeno:** Este gas contribuye a la formación del “smog” y está relacionado con problemas de salud como afecciones respiratorias.

**§ Dióxido de azufre:** Principalmente emitido por la quema de combustibles fósiles, este compuesto puede dar lugar a la lluvia ácida y otros impactos ambientales [5].

**§ Plomo:** El plomo representa una preocupación mayor debido a sus graves efectos en la salud y desarrollo humano. Estudios, como los citados en un artículo de The Atlantic, han demostrado una relación directa entre la exposición al plomo y la disminución del coeficiente intelectual, daños auditivos, habilidades cognitivas debilitadas e incluso tasas más altas de criminalidad en diferentes estados [6]. En el siglo XX, el plomo desempeñó un papel determinante en el deterioro del desarrollo infantil en Estados Unidos, lo que resalta su impacto negativo en la sociedad.

Es crucial comprender cómo se mide la calidad del aire, qué variables intervienen en este proceso y cómo se llevan a cabo las mediciones. Para llevar a cabo la medición de la calidad del aire, se emplean sistemas de vigilancia equipados con sensores diseñados específicamente para detectar contaminantes particulares. Estos contaminantes pueden originarse tanto de fuentes naturales, como tormentas de polvo, humo de incendios forestales y erupciones volcánicas, como de fuentes antropogénicas, como la quema de combustibles fósiles [7].

Dentro del conjunto de contaminantes mencionados previamente, existen algunos que tienen impactos significativos en la salud humana y el medio ambiente. Entre estos se encuentran PM2.5, PM10, ozono troposférico, dióxido de carbono, dióxido de nitrógeno y dióxido de azufre [8]. La presencia y concentración de estos contaminantes en el aire se traducen en el Índice de Calidad del Aire (ICA), cuyo rango va desde 0 hasta 500. En este índice, un puntaje de 50 o menos se considera seguro, mientras que valores superiores a 100 indican niveles poco saludables [9].

Sin embargo, la evaluación de la calidad del aire no se limita únicamente a los contaminantes presentes en el aire en un momento determinado. Las estaciones meteorológicas también incorporan variables como la dirección y velocidad del viento, la humedad relativa, la temperatura, la precipitación, la presión atmosférica y la radiación solar en su análisis. Estos factores meteorológicos juegan un papel crucial en la dispersión y concentración de los contaminantes, así como en la formación de patrones de contaminación.

Esta comprensión integral de la medición de la calidad del aire nos brinda una perspectiva más completa sobre cómo se evalúa y monitorea esta cuestión crítica, en concordancia con la información previamente proporcionada.

### § Cuáles son las normas oficiales establecidas de los distintos contaminantes para la protección de la salud.

Las normas oficiales establecidas de los contaminantes para la protección de la salud son las siguientes:

Contaminante	-	Norma	Concentración	Tiempo de exposición(horas)
Monóxido de Carbono	CO	NOM-021-SSAI-2021	26.00 ppm	1
Monóxido de Carbono	CO	NOM-021-SSAI-2021	9.0 ppm	8
Bióxido de Azufre	SO2	NOM-022-SSAI-2019	0.075 ppm	1
Bióxido de Azufre	SO2	NOM-022-SSAI-2019	0.04 ppm	24
Ozono	O3	NOM-020-SSAI-2021	0.090 ppm	1
Ozono	O3	NOM-020-SSAI-2021	0.065 ppm	8
Bióxido de Nitrógeno	NO2	NOM-023-SSAI-2021	0.106 ppm	1
Bióxido de Nitrógeno	NO2	NOM-023-SSAI-2021	0.021 ppm	Promedio Anual
Partículas Menores a 10 micras	PM10	NOM-025-SSAI-2021	70 µg/m3	24
Partículas Menores a 10 micras	PM10	NOM-025-SSAI-2021	36 µg/m3	Promedio Anual
Partículas Menores a 2.5 micras	PM2.5	NOM-025-SSAI-2021	41 µg/m3	24
Partículas Menores a 2.5 micras	PM2.5	NOM-025-SSAI-2021	10 µg/m3	Promedio Anual

**Tabla 1. Normas Oficiales, obtenida de: Aire Nuevo León**

En estas normas se establecen los criterios a tomar en cuenta para evaluar la calidad del aire, con respecto a la concentración de cada contaminante.

La medición de la calidad del aire se encuentra confrontada con desafíos considerables debido a una serie de factores complejos. La falta de estandarización en los métodos de

medición, la variabilidad en distintos lugares y momentos, y la complejidad inherente de los contaminantes y sus impactos en la salud humana se perfilan como obstáculos centrales [10].

La falta de uniformidad en los enfoques de medición dificulta la comparación y evaluación de datos, lo que a su vez complica la toma de decisiones informadas. La variabilidad en la calidad del aire, impulsada por diversas fuentes de contaminación, condiciones climáticas cambiantes y características topográficas, plantea dificultades para obtener mediciones verdaderamente precisas. Además, la amplia gama de contaminantes y sus efectos multifacéticos en la salud requieren la implementación de equipos y metodologías especializadas [10].

Un factor adicional que contribuye a los desafíos es la falta de requisitos legales para la medición de la calidad del aire en un considerable 37% de los países, según datos del Programa de las Naciones Unidas para el Medio Ambiente (PNUMA). Esta carencia limita la disponibilidad de datos confiables y obstaculiza la formulación de políticas efectivas para controlar la contaminación del aire y salvaguardar la salud pública [11]. En resumen, la superación de estas dificultades es de vital importancia para comprender plenamente y mitigar los riesgos asociados con la contaminación del aire.

### **Socio Formador**

En México, se han establecido los programas ProAire para abordar las necesidades de los estados en cuanto a contar con herramientas preventivas y correctivas en relación con la calidad del aire y la protección de la salud [12]. Un ejemplo de ello es el programa SIMA en el Estado de Nuevo León, que representa el Sistema Integral de Monitoreo Ambiental. En este proyecto, SIMA desempeña un papel fundamental como socio formador en el proyecto actual, trabajando conjuntamente para lograr los objetivos que se describen a continuación.

El programa SIMA es una iniciativa respaldada por el Gobierno del Estado de Nuevo León, cuyo propósito central radica en evaluar la calidad del aire mediante la monitorización de las concentraciones atmosféricas a las que está expuesta la población. Específicamente, en situaciones adversas, el programa tiene la función de alertar a la población sobre episodios en los que los niveles de contaminación atmosférica son elevados [13].

Una de las características destacadas de este programa es su capacidad para suministrar información actualizada de manera constante acerca de la calidad del aire en la región donde opera. Su ámbito de actuación principal abarca la zona metropolitana de Monterrey. La entidad responsable de la gestión de este programa es la Secretaría de Medio Ambiente.

### **·Descripción del problema específico (preguntas de investigación)**

El problema es la contaminación del aire producida por diferentes contaminantes en el estado de Nuevo León y diferentes cuestiones climatológicas, por eso se abordarán las siguientes preguntas de investigación para enfocarnos en el problema específico:

- ¿Hay una relación entre los contaminantes que tenga un impacto significativo en la calidad del aire?

- ¿La calidad del aire se ve afectada por una relación entre las cuestiones climatológicas y los contaminantes?
- ¿Qué tanto se ve afectada la calidad del aire con relación a la concentración de cada contaminante?

### · **Objetivos (se definen con base en las preguntas de investigación)**

#### **Pregunta 1:**

1. **Objetivo General:** Identificar si existen relaciones entre los contaminantes atmosféricos que se encuentran en la base de datos y a su vez determinar si estas relaciones tienen un impacto importante en la calidad del aire.
2. **Objetivos Específicos:**
3. Analizar el impacto de cada contaminante en la calidad del aire.
4. Identificar patrones de comportamiento temporales en los contaminantes y su impacto en la calidad del aire.
5. Comparar el impacto de los contaminantes en la calidad del aire.

#### **Pregunta 2:**

1. **Objetivo General:** Observar la relación entre las variables climatológicas y las variables contaminantes para determinar el impacto de esta relación en la calidad del aire.
2. **Objetivos Específicos:**
3. Analizar el impacto de las variables climatológicas en la calidad del aire.
4. Realizar el análisis de la relación entre las variables climatológicas y las contaminantes.
5. Evaluar de qué forma varía el impacto de las variables climatológicas y contaminantes en la calidad del aire, debido a la variación geográfica.
6. Identificar patrones de comportamiento temporales en la relación entre las variables climatológicas y contaminantes.

#### **Pregunta 3:**

1. **Objetivo General:** Determinar el impacto de diferentes niveles de concentración de contaminantes atmosféricos en la calidad del aire, y en la salud humana y ambiental.
2. **Objetivos Específicos:**
3. Indagar sobre el impacto de los contaminantes en la salud humana y ambiental, y definir rangos aceptables para cada contaminante.
4. Comparar el impacto que tiene cada contaminante en la calidad del aire.
5. Verificar si existen variaciones entre los niveles de contaminantes y la calidad del aire dadas por variación geográfica.
6. Elaborar una serie de recomendaciones y medidas de prevención en base a los resultados que se obtengan.

### · **Justificación de los objetivos**

La contaminación es un problema que no se puede dejar de abordar, especialmente cuando cada vez nos acercamos más al año 2030, en el cual deberíamos tener cumplidos varios ODS. En el estado de Nuevo León el principal problema de contaminación, de acuerdo a la ciudadanía, es el aire [14], por lo que resulta de interés hacer un análisis que permita encontrar información que ayude de alguna manera a disminuir este problema. A continuación se presentan las razones para la realización de este trabajo y sus objetivos:

1. **Impacto en la Salud de la Población:** La comprensión de las relaciones entre los contaminantes y la calidad del aire permitirá impactar en gran medida la salud de la población, pues podrá servir para saber hacia dónde dirigir las políticas que buscan la reducción de esta contaminación.
2. **Variabilidad climática:** El análisis de la relación entre las variables climatológicas y contaminantes, permitirá comprender cómo su relación tiene un impacto en la calidad del aire y, a su vez, en la salud humana. La comprensión de esta relación permitirá la toma de decisiones relacionadas con la prevención de problemas de salud de los ciudadanos.
3. **Evaluación e identificación de contaminantes que presentan mayor riesgo para la salud:** El conocimiento de la concentración de los contaminantes en el aire permitirá conocer el daño potencial que presentan para los habitantes del estado. Teniendo en cuenta esto, se podrá realizar la identificación de contaminantes que presenta mayor riesgo, y se podrá buscar una manera en la que se pueda mitigar la emisión de estos contaminantes.
4. **Toma de Decisiones guiada por la información obtenida:** El análisis de esta información permitirá que se puedan tomar decisiones que lleven a la mitigación del problema. Además que con estos datos se podrán planificar nuevas estrategias y políticas, generales y específicas, que permitan la prevención del aumento de la contaminación y un aumento en la afectación de la salud de los ciudadanos y del medio ambiente.

#### · Descripción de las fuentes de información (datos)

Las fuentes de información que han sido utilizadas para la realización de este trabajo hasta el momento son principalmente gubernamentales, puesto que buscamos entregar un trabajo que se sustente de fuentes que a la consideración del equipo se consideren confiables, no olvidando que se quiere primeramente tomar fuentes primarias o secundarias. Algunas de estas fuentes son la Organización Mundial de la Salud, el Servicio Nacional del Clima de Estados Unidos, las Naciones Unidas, y el gobierno de Nuevo León a través del SIMA.

Para realizar el trabajo estadístico que se presentará a continuación se hará uso de una base de datos proporcionada por el Sistema Integral de Monitoreo Ambiental de Nuevo León. Estos datos se encuentran divididos por estaciones de monitoreo, que se pueden encontrar en la página de SIMA en tiempo real. Los datos que se recopilaron son de contaminantes presentes en el aire en ciertas horas y fechas, además que son de igual manera de carácter climatológico.

#### · Impacto social principal

Se espera que el desarrollo del proyecto tenga un impacto positivo para la sociedad a través de la generación de conocimiento e información que permita la generación de planes de mitigación para el problema actual que resulta ser actualmente la contaminación del aire. Estos planes de mitigación se espera que aporten un gran beneficio a la población en general. Además de los planes de mitigación se espera que se puedan generar políticas para problemas específicos que se puedan encontrar a través de la información generada.

## Etapa II: Comprensión y preparación de los datos

Para dirigir la solución al reto, seguirás la metodología CRISP-DM. En esta primera actividad realizarás el segundo paso Comprensión y Preparación de los datos que permitirá añadir el análisis descriptivo e introductorio de las variables.

### 1) Comprensión de los datos del negocio

#### Importar base de datos

```
sheet_names = excel_sheets("datos2022_2023_estaciones.xlsx")

for (i in 1:length(sheet_names)) {
  variable_name = paste("M",sheet_names[i],sep="_")
  value = read_excel("datos2022_2023_estaciones.xlsx",sheet = sheet_names[i])
  assign(variable_name,value) #guardamos cada hoja en una variable
}
```

Lo que se hizo en el código de arriba fue guardar los datos de cada estación en una variable. Por fines prácticos utilizaremos solamente una de ellas para realizar el análisis descriptivo de esta entrega.

#### A Dimensión del dataset

```
cat("Número de observaciones:",dim(M_CENTRO)[1])

## Número de observaciones: 14255

cat("\nNúmero de variables:",dim(M_CENTRO)[2])

##
## Número de variables: 16
```

**Describe claramente cada una de las variables, incluyendo su nombre, descripción, tipo (categórico/Numérico) y valores posibles que puede tomar, valores nulos.**

```
cat("Tipo dato de cada variable\n")

## Tipo dato de cada variable

t(sapply(M_CENTRO,class))

##      date      CO      NO      NO2      NOX      O3
## [1,] character,2 "character" "character" "character" "character"
"character"
```



```

##      PM10      PM2.5      PRS      RAINF      RH      S02
## [1,] "character" "character" "character" "character" "character"
"character"
##      SR      TOUT      WSR      WDR
## [1,] "character" "character" "character" "character"

summary(M_CENTRO)

##      date                                CO      NO
## Min.   :2022-01-01 00:00:00.00   Length:14255   Length:14255
## 1st Qu.:2022-05-29 11:30:00.00   Class :character   Class :character
## Median :2022-10-24 23:00:00.00   Mode  :character   Mode  :character
## Mean   :2022-10-24 23:18:54.42
## 3rd Qu.:2023-03-22 11:30:00.00
## Max.   :2023-08-17 23:00:00.00
##      NO2      NOX      O3      PM10
## Length:14255   Length:14255   Length:14255   Length:14255
## Class :character   Class :character   Class :character   Class :character
## Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##      PM2.5      PRS      RAINF      RH
## Length:14255   Length:14255   Length:14255   Length:14255
## Class :character   Class :character   Class :character   Class :character
## Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##      S02      SR      TOUT      WSR
## Length:14255   Length:14255   Length:14255   Length:14255
## Class :character   Class :character   Class :character   Class :character
## Mode  :character   Mode  :character   Mode  :character   Mode  :character
##
##
##      WDR
## Length:14255
## Class :character
## Mode  :character
##
##
##

```

Si observamos a simple vista la tabla hecha por la función `summary()`, inicialmente vemos que las variables que deberían catalogarse como cuantitativas se identifican como caracteres, lo cual no resulta eficiente para hacer un análisis estadístico y descriptivo. Para solucionar esto, transformaremos los datos de las variables para que sean numéricos.

### Transformación a variables numéricas

```
variables_names = colnames(M_CENTRO)
M_CENTRO[,2:length(variables_names)] =
sapply(M_CENTRO[,2:length(variables_names)],as.numeric)
t(sapply(M_CENTRO,class))

##      date      CO      NO      NO2      NOX      O3      PM10
## [1,] character,2 "numeric" "numeric" "numeric" "numeric" "numeric"
"numeric"
##      PM2.5      PRS      RAINF      RH      SO2      SR      TOUT
## [1,] "numeric" "numeric" "numeric" "numeric" "numeric" "numeric" "numeric"
##      WSR      WDR
## [1,] "numeric" "numeric"

cat("Valores únicos por cada variable\n")

## Valores únicos por cada variable

for (i in 1:length(variables_names)) {
cat("\n-",variables_names[i],":",length(unique(sapply(M_CENTRO
[,i],as.list)))) #Transformamos las columnas a vectores para poder calcular
sus valores únicos
}

##
## - date : 14255
## - CO : 404
## - NO : 655
## - NO2 : 617
## - NOX : 1021
## - O3 : 133
## - PM10 : 1090
## - PM2.5 : 4420
## - PRS : 240
## - RAINF : 2
## - RH : 93
## - SO2 : 226
## - SR : 984
## - TOUT : 3222
## - WSR : 209
## - WDR : 360

cat("\nNúmero de valores nulos (NA) por variable")

##
## Número de valores nulos (NA) por variable

for (i in 1:length(variables_names)){
cat(cat("\n-",
",variables_names[i],":",length(which(is.na(M_CENTRO[,i]),arr.ind=TRUE))))
}
}
```

```
##
## - date : 0
## - CO : 544
## - NO : 1546
## - NO2 : 446
## - NOX : 444
## - O3 : 908
## - PM10 : 586
## - PM2.5 : 2506
## - PRS : 452
## - RAINF : 424
## - RH : 584
## - SO2 : 2258
## - SR : 266
## - TOUT : 424
## - WSR : 426
## - WDR : 458
```

## Exploración de los datos

### 1. Variables cuantitativas

- Medidas de posición no-central: cuartiles, outlier (valores atípicos), boxplots

```
summary(M_CENTRO)
```

```
##          date                      CO                      NO
## Min.      :2022-01-01 00:00:00.00  Min.      :0.070  Min.      : 0.600
## 1st Qu.:2022-05-29 11:30:00.00  1st Qu.:1.150  1st Qu.: 2.500
## Median :2022-10-24 23:00:00.00  Median :1.510  Median : 3.600
## Mean    :2022-10-24 23:18:54.42  Mean    :1.675  Mean     : 8.444
## 3rd Qu.:2023-03-22 11:30:00.00  3rd Qu.:2.030  3rd Qu.: 8.700
## Max.    :2023-08-17 23:00:00.00  Max.     :4.940  Max.    :166.600
##                                     NA's      :272  NA's      :773
##          NO2                      NOX                      O3                      PM10
## Min.      : 0.40  Min.      : 2.70  Min.      : 1.00  Min.      : 3.00
## 1st Qu.: 7.90  1st Qu.: 10.90  1st Qu.: 12.00  1st Qu.: 41.00
## Median :14.20  Median : 18.70  Median : 23.00  Median : 57.00
## Mean    :16.76  Mean     : 25.17  Mean     : 26.65  Mean     : 64.19
## 3rd Qu.:22.60  3rd Qu.: 31.90  3rd Qu.: 37.00  3rd Qu.: 78.00
## Max.    :89.10  Max.    :186.80  Max.    :164.00  Max.    :735.00
## NA's     :223  NA's     :222  NA's     :454  NA's     :293
##          PM2.5          PRS          RAINF          RH
## Min.      : 0.00  Min.      :698.7  Min.      :0  Min.      : 1.00
## 1st Qu.: 14.89  1st Qu.:709.0  1st Qu.:0  1st Qu.:36.00
## Median : 21.68  Median :711.0  Median :0  Median :53.00
## Mean     : 24.86  Mean     :711.4  Mean     :0  Mean     :51.99
## 3rd Qu.: 31.42  3rd Qu.:713.3  3rd Qu.:0  3rd Qu.:69.00
## Max.     :120.10  Max.     :724.9  Max.     :0  Max.     :92.00
## NA's     :1253  NA's     :226  NA's     :212  NA's     :292
##          SO2          SR          TOUT          WSR
## Min.      : 0.500  Min.      :0.0000  Min.      : -1.89  Min.      : 0.700
```

```
## 1st Qu.: 3.200 1st Qu.:0.0010 1st Qu.:18.27 1st Qu.: 5.100
## Median : 4.800 Median :0.0080 Median :23.79 Median : 7.800
## Mean : 5.175 Mean :0.1918 Mean :22.91 Mean : 8.213
## 3rd Qu.: 6.100 3rd Qu.:0.3190 3rd Qu.:28.04 3rd Qu.:10.800
## Max. :73.900 Max. :1.0570 Max. :41.47 Max. :24.800
## NA's :1129 NA's :133 NA's :212 NA's :213
## WDR
## Min. : 1.0
## 1st Qu.: 47.0
## Median : 72.0
## Mean :102.1
## 3rd Qu.:117.0
## Max. :360.0
## NA's :229

#Desviacion estandar por variable
cat("Desviacion estandar por variable")

## Desviacion estandar por variable

for (i in 1:length(variables_names)) {
cat(cat("\n-",variables_names[i],":",sapply(M_CENTRO[,i],var))) #omitimos los
valores faltantes
}

##
## - date : 2.195167e+14
## - CO : NA
## - NO : NA
## - NO2 : NA
## - NOX : NA
## - O3 : NA
## - PM10 : NA
## - PM2.5 : NA
## - PRS : NA
## - RAINF : NA
## - RH : NA
## - SO2 : NA
## - SR : NA
## - TOUT : NA
## - WSR : NA
## - WDR : NA
```

Vemos que obtenemos “NA”s o “valores faltantes” cuando tratamos de obtener la desviación estándar de cada variable-a excepción de la variable de fecha ya que esa es categórica y no aporta valor al análisis. Esto se debe principalmente a que R no acepta valores faltantes al calcular la desviación estandar o la varianza por cada columna. Mucho del análisis requiere que no haya valores faltantes, por lo que tendremos que omitir estos los datos faltantes.

```

#Desviacion estandar por variable
cat("Desviacion estandar por variable")

## Desviacion estandar por variable

for (i in 1:length(variables_names)) {
  cat(cat("\n-",variables_names[i],":",sapply(na.omit(M_CENTRO[,i]),var)))
}

##
## - date : 2.195167e+14
## - CO : 0.5596336
## - NO : 156.9811
## - NO2 : 131.475
## - NOX : 434.7073
## - O3 : 368.5081
## - PM10 : 1192.019
## - PM2.5 : 202.4671
## - PRS : 12.95555
## - RAINF : 0
## - RH : 415.1135
## - SO2 : 8.156039
## - SR : 0.08182824
## - TOUT : 56.2214
## - WSR : 15.14082
## - WDR : 7491.107

```

- Análisis de distribución de los datos (Histogramas). Identificar si tiene forma simétrica o asimétrica

```

#Histograma
list <-lapply(1:ncol(M_CENTRO),
             function(col) ggplot2::qplot(M_CENTRO[[col]],
                                           geom = "histogram",
                                           binwidth = 1))

## Warning: `qplot()` was deprecated in ggplot2 3.4.0.
## This warning is displayed once every 8 hours.
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was
## generated.

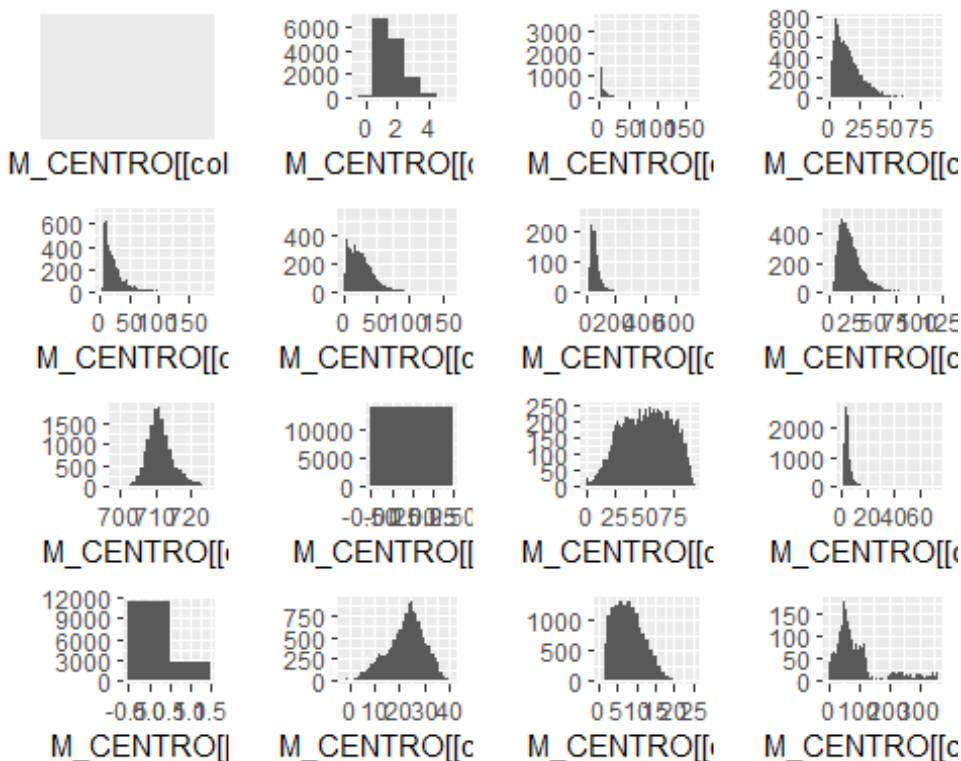
cowplot::plot_grid(plotlist = list)

## Warning: Computation failed in `stat_bin()`
## Caused by error in `bin_breaks_width()`:
## ! The number of histogram bins must be less than 1,000,000.
## i Did you make `binwidth` too small?

## Warning: Removed 272 rows containing non-finite values (`stat_bin()`).
## Warning: Removed 773 rows containing non-finite values (`stat_bin()`).
## Warning: Removed 223 rows containing non-finite values (`stat_bin()`).

```

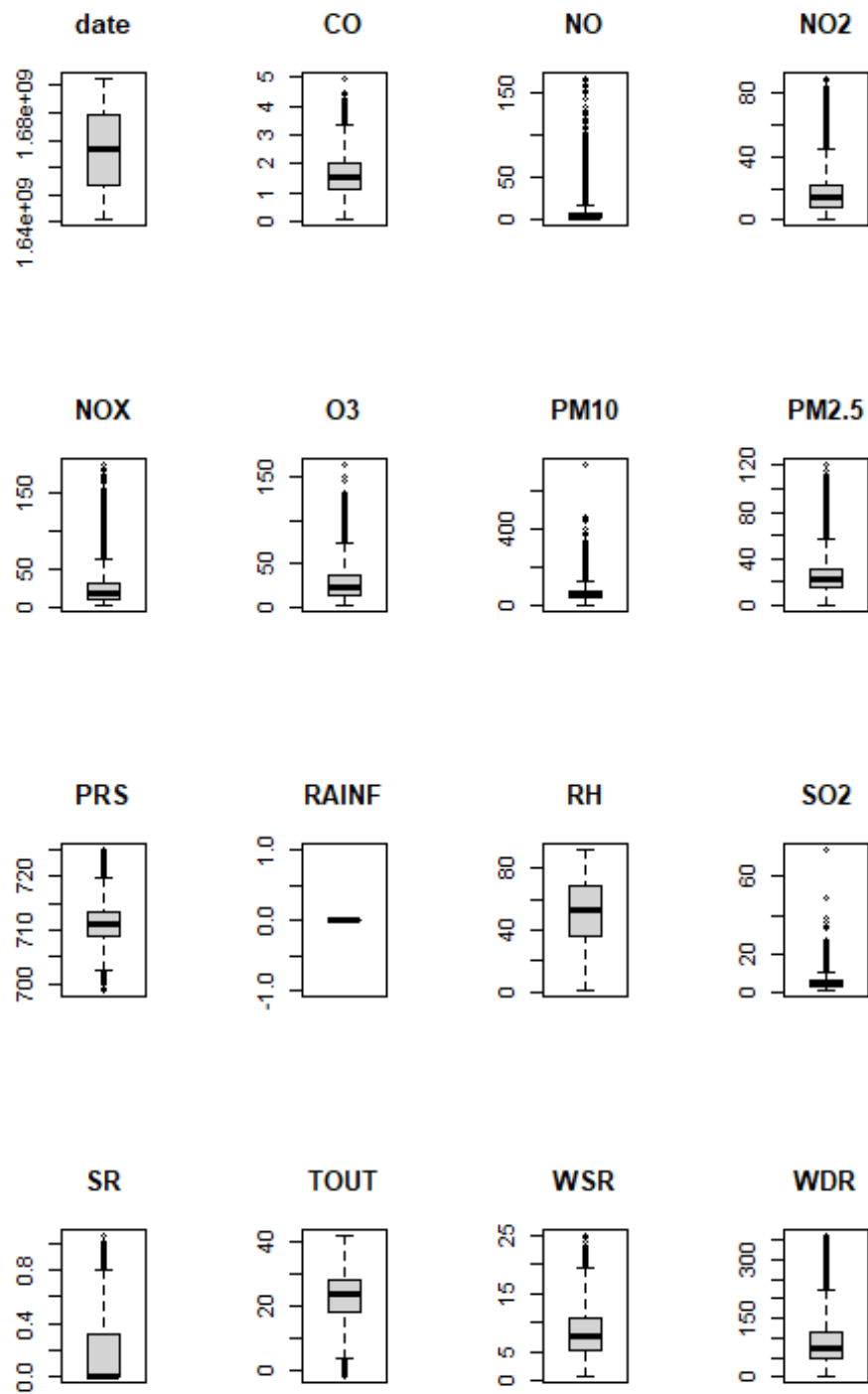
```
## Warning: Removed 222 rows containing non-finite values (`stat_bin()`).
## Warning: Removed 454 rows containing non-finite values (`stat_bin()`).
## Warning: Removed 293 rows containing non-finite values (`stat_bin()`).
## Warning: Removed 1253 rows containing non-finite values (`stat_bin()`).
## Warning: Removed 226 rows containing non-finite values (`stat_bin()`).
## Warning: Removed 212 rows containing non-finite values (`stat_bin()`).
## Warning: Removed 292 rows containing non-finite values (`stat_bin()`).
## Warning: Removed 1129 rows containing non-finite values (`stat_bin()`).
## Warning: Removed 133 rows containing non-finite values (`stat_bin()`).
## Warning: Removed 212 rows containing non-finite values (`stat_bin()`).
## Warning: Removed 213 rows containing non-finite values (`stat_bin()`).
## Warning: Removed 229 rows containing non-finite values (`stat_bin()`).
```



A simple vista, podemos deducir que la mayoría de las variables cauntitativas tienen un sesgo hacia la derecha, a excepción de variables como “PRS”, “RH”, y “TOUT”.

```
par(mfrow = c(2, 4)) #Despliega Los boxplot en una sola hoja
for (col in variables_names) {
```

```
boxplot(na.omit(M_CENTRO[[col]]), main = col)
}
```



- Análisis de correlación de los datos, mapa de calor

```

cor_matrix = cor(na.omit(M_CENTRO[,2:length(variables_names)]))

## Warning in cor(na.omit(M_CENTRO[, 2:length(variables_names)])): the
## standard
## deviation is zero

cor_matrix

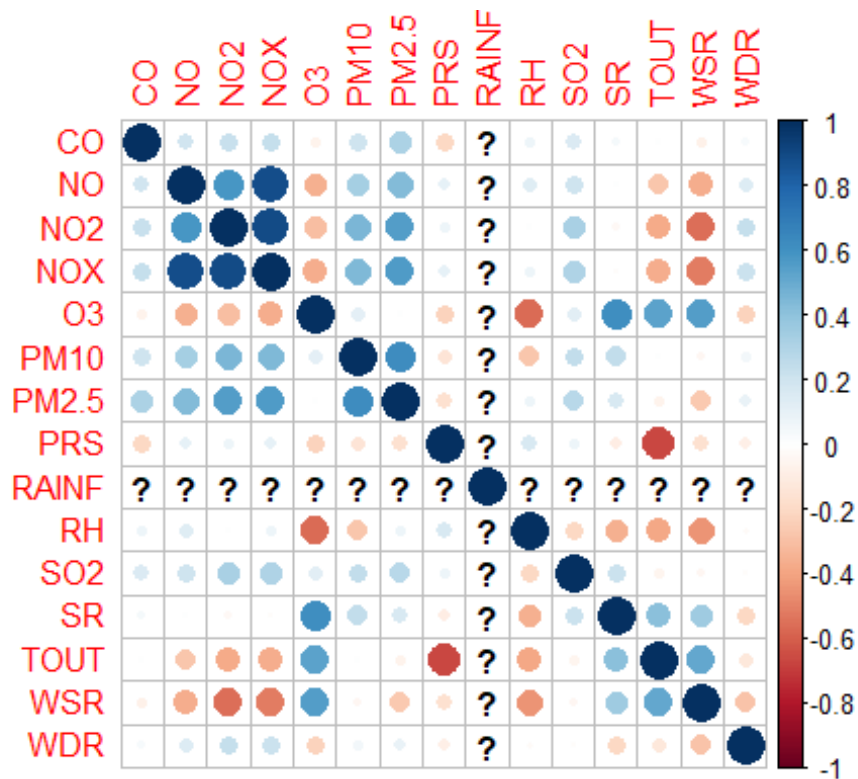
##           CO           NO           NO2           NOX           O3
## CO      1.000000000  0.19353971  0.221381461  0.23363881 -0.060324305
## NO      0.193539705  1.00000000  0.580690148  0.88537926 -0.351398080
## NO2     0.221381461  0.58069015  1.000000000  0.89258016 -0.307119487
## NOX     0.233638815  0.88537926  0.892580156  1.00000000 -0.369923780
## O3      -0.060324305 -0.35139808 -0.307119487 -0.36992378 1.000000000
## PM10    0.207063641  0.33136889  0.456616861  0.44431124 0.116993654
## PM2.5   0.317259624  0.43788977  0.558567133  0.56146053 0.002511103
## PRS     -0.204045029  0.10875113  0.076754243  0.10401368 -0.225746458
## RAINF    NA          NA          NA          NA          NA
## RH       0.070084617  0.13053453  0.006596992  0.07609906 -0.561263150
## SO2      0.157169515  0.20563440  0.327395845  0.30085697 0.120416545
## SR       0.044555682  0.00986736 -0.038940546 -0.01643090 0.613314553
## TOUT     0.004046107 -0.27839827 -0.371887216 -0.36652437 0.530126241
## WSR      -0.074799619 -0.36505226 -0.556834892 -0.51991603 0.557319641
## WDR      0.034489860  0.14405507  0.239185499  0.21629718 -0.228699299
##           PM10        PM2.5        PRS RAINF          RH          SO2
## CO      0.207063641  0.317259624 -0.20404503  NA  0.070084617  0.15716952
## NO      0.331368888  0.437889773  0.10875113  NA  0.130534533  0.20563440
## NO2     0.456616861  0.558567133  0.07675424  NA  0.006596992  0.32739585
## NOX     0.444311239  0.561460534  0.10401368  NA  0.076099058  0.30085697
## O3      0.116993654  0.002511103 -0.22574646  NA -0.561263150  0.12041655
## PM10    1.000000000  0.621287860 -0.14847902  NA -0.275698268  0.24318405
## PM2.5   0.621287860  1.000000000 -0.16403977  NA  0.079819492  0.27154944
## PRS     -0.148479020 -0.164039773  1.00000000  NA  0.162557657  0.07308559
## RAINF    NA          NA          NA      1          NA          NA
## RH      -0.275698268  0.079819492  0.16255766  NA  1.000000000 -0.20790529
## SO2     0.243184050  0.271549440  0.07308559  NA -0.207905288  1.00000000
## SR      0.248180155  0.161793362 -0.09233537  NA -0.359562359  0.21501328
## TOUT    0.003881171 -0.061258671 -0.66167650  NA -0.388931959 -0.05850601
## WSR     -0.046744827 -0.268046001 -0.16906494  NA -0.442403622 -0.04950906
## WDR     0.058628751  0.099633611 -0.08722834  NA -0.022897182 -0.01038542
##           SR          TOUT          WSR          WDR
## CO      0.04455568  0.004046107 -0.07479962  0.03448986
## NO      0.00986736 -0.278398274 -0.36505226  0.14405507
## NO2     -0.03894055 -0.371887216 -0.55683489  0.23918550
## NOX     -0.01643090 -0.366524371 -0.51991603  0.21629718
## O3      0.61331455  0.530126241  0.55731964 -0.22869930
## PM10    0.24818016  0.003881171 -0.04674483  0.05862875
## PM2.5   0.16179336 -0.061258671 -0.26804600  0.09963361
## PRS     -0.09233537 -0.661676498 -0.16906494 -0.08722834
## RAINF    NA          NA          NA          NA

```



```
## RH      -0.35956236 -0.388931959 -0.44240362 -0.02289718
## SO2      0.21501328 -0.058506014 -0.04950906 -0.01038542
## SR       1.00000000  0.412479613  0.35248587 -0.20694581
## TOUT     0.41247961  1.000000000  0.51221300 -0.12470607
## WSR      0.35248587  0.512212995  1.00000000 -0.28785509
## WDR     -0.20694581 -0.124706072 -0.28785509  1.00000000
```

```
# Visualiza la matriz de correlación utilizando corrplot
corrplot(cor_matrix, method = "circle")
```



Visualmente, se

observan los siguientes aspectos:

- Las variables NO y NOX parecen tener una correlación positiva fuerte
- Las variables TOUT y PRS parecen tener una correlación negativa fuerte
- NO y NO2 parecen tener una correlación positiva fuerte o moderada
- La variable O3 parece ser que tiene correlaciones moderadas con NO, NO2, NOX, SR, TOUT, WSR
- La variable WSR tiene correlaciones débiles o moderadas con NO, NO2 Y NOX
- Las variables PM10 Y PM2.5 parecen tener correlaciones moderadas con NO, NO2 y NOX
- La variable RAINF no tiene ninguna relación con ninguna otra variable

Se encuentran varios patrones dentro de los datos, sin embargo, resulta central realizar pruebas de *independencia* y de *colinealidad* entre nuestras posibles variables predictoras para poder reducir la dimensionalidad de nuestra matriz de datos y la complejidad del problema. Del mismo modo, nuestras variables de interés o dependientes tendrán que ser

los contaminantes que describimos en la etapa I-estos incluyen a "O3", "PM10", "PM2.5" o "NO"

## 2. Variables categóricas

- Distribución de los datos (diagramas de barras, diagramas de pastel)

En este caso nadamas tenemos una variable categórica, la cual el la variable "date", por lo que no tiene mucho caso analizar esta misma

## Verificación de datos

### 1. Valores faltantes

Como ya se vió anteriormente, se necesitan eliminar a los valores faltantes y en el proceso de hacerlo, tenemos que asegurarnos de no haber perdido tanta información. Para esto necesitamos ver que porcentaje de los datos totales son "NA".

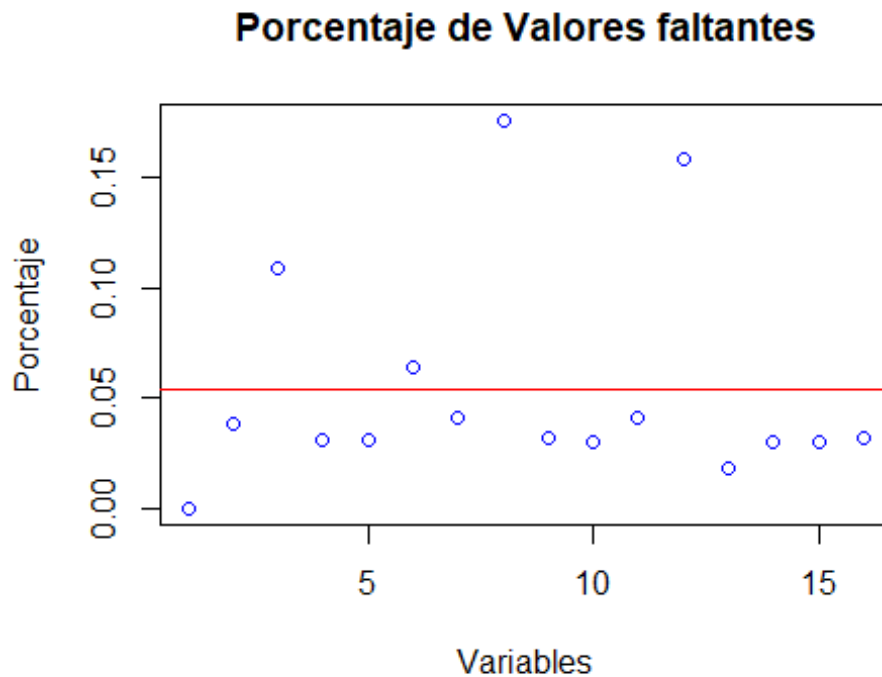
```
cat("\nNúmero de valores nulos (NA) por variable")

##
## Número de valores nulos (NA) por variable

props_NA = c()
for (i in 1:length(variables_names)){
  prop = length(which(is.na(M_CENTRO[,i]),arr.ind=TRUE)) / nrow(M_CENTRO)
  cat("\n-",variables_names[i],":",prop)
  props_NA[length(props_NA) + 1] = prop
}

##
## - date : 0
## - CO : 0.03816205
## - NO : 0.1084532
## - NO2 : 0.03128727
## - NOX : 0.03114697
## - O3 : 0.06369695
## - PM10 : 0.04110838
## - PM2.5 : 0.175798
## - PRS : 0.03170817
## - RAINF : 0.02974395
## - RH : 0.04096808
## - SO2 : 0.1584006
## - SR : 0.01866012
## - TOUT : 0.02974395
## - WSR : 0.02988425
## - WDR : 0.03212908

plot(1:length(variables_names),props_NA,col="blue",xlab="Variables",ylab="Porcentaje",main="Porcentaje de Valores faltantes")
abline(h=mean(props_NA ),col="red") #media de valores faltantes
```



Vemos que el mayor porcentaje que representan los datos faltantes es de 5.3805682%, y el máximo de ellos siendo el de la variable “PM2.5”, “so2” y “NO”,respectivamente. No hay un estandar que diga a que porcentaje es seguro eliminar los valores faltantes sin perder una gran cantidad de información; sin embargo, como en este caso tenemos un gran número de valores válidos con respecto a los inválidos, tomaremos la decisión empírica de eliminarlos

```
#Tratamiento de valores faltantes
M_CENTRO = na.exclude(M_CENTRO)
cat("\nNúmero de valores nulos (NA) por variable")

##
## Número de valores nulos (NA) por variable

for (i in 1:length(variables_names)){
  cat(cat("\n-
",variables_names[i],":",length(which(is.na(M_CENTRO[,i]),arr.ind=TRUE))))
}

##
## - date : 0
## - CO : 0
## - NO : 0
## - NO2 : 0
## - NOX : 0
## - O3 : 0
## - PM10 : 0
## - PM2.5 : 0
```

```
## - PRS : 0
## - RAINF : 0
## - RH : 0
## - SO2 : 0
## - SR : 0
## - TOUT : 0
## - WSR : 0
## - WDR : 0
```

## 2. Valores de los datos

```
summary(M_CENTRO)
```

```
##      date                                CO              NO
## Min.   :2022-01-01 00:00:00.00    Min.   :0.070    Min.   : 1.800
## 1st Qu.:2022-05-21 11:00:00.00    1st Qu.:1.150    1st Qu.: 2.500
## Median :2022-10-30 08:00:00.00    Median :1.520    Median : 3.600
## Mean   :2022-10-26 10:21:54.14    Mean   :1.677    Mean   : 8.034
## 3rd Qu.:2023-03-31 19:00:00.00    3rd Qu.:2.040    3rd Qu.: 8.500
## Max.   :2023-08-17 23:00:00.00    Max.   :4.940    Max.   :126.600
##      NO2              NOX              O3              PM10
## Min.   : 2.00    Min.   : 4.30    Min.   : 1.00    Min.   : 8.94
## 1st Qu.: 8.40    1st Qu.: 11.50    1st Qu.: 11.00    1st Qu.: 43.00
## Median :14.70    Median : 19.40    Median : 23.00    Median : 59.00
## Mean   :17.31    Mean   : 25.55    Mean   : 27.14    Mean   : 66.50
## 3rd Qu.:23.20    3rd Qu.: 32.40    3rd Qu.: 38.00    3rd Qu.: 81.00
## Max.   :89.10    Max.   :173.90    Max.   :164.00    Max.   :735.00
##      PM2.5          PRS          RAINF          RH          SO2
## Min.   : 0.00    Min.   :698.7    Min.   :0    Min.   : 1.00    Min.   :
1.300
## 1st Qu.: 15.00    1st Qu.:709.0    1st Qu.:0    1st Qu.:35.00    1st Qu.:
3.200
## Median : 22.00    Median :711.0    Median :0    Median :51.00    Median :
4.800
## Mean   : 25.13    Mean   :711.3    Mean   :0    Mean   :50.84    Mean   :
5.216
## 3rd Qu.: 31.74    3rd Qu.:713.2    3rd Qu.:0    3rd Qu.:67.00    3rd Qu.:
6.100
## Max.   :120.10    Max.   :724.9    Max.   :0    Max.   :91.00    Max.
:73.900
##      SR          TOUT          WSR          WDR
## Min.   :0.0000    Min.   : -1.89    Min.   : 0.700    Min.   : 1.0
## 1st Qu.:0.0010    1st Qu.:18.51    1st Qu.: 5.200    1st Qu.: 46.0
## Median :0.0270    Median :24.07    Median : 8.000    Median : 71.0
## Mean   :0.2106    Mean   :23.20    Mean   : 8.322    Mean   :101.7
## 3rd Qu.:0.3850    3rd Qu.:28.35    3rd Qu.:11.000    3rd Qu.:116.0
## Max.   :1.0570    Max.   :41.47    Max.   :24.800    Max.   :360.0
```

## Preparación de los datos

### 1) Selecciona el conjunto de datos a utilizar

- a) Decide qué conjunto de datos se utilizará. Explica por qué se incluyeron o excluyeron ciertos datos.

Para este caso en particular utilizaremos nos quedaremos con el siguientes listado de variables predictoras:

- NO
- NO2
- NOX
- RH
- SR
- TOUT
- WSR

Las variables anteriores fueron seleccionadas en base al mapa de calor anterior donde muestra las matriz de correlaciones, y donde se visualiza la magnitud de las correlaciones. Además, se eligieron ese conjunto de contaminantes y variables metereologicas con fines de predecir los niveles de O3, PM2.5 Y PM10. No obstante, quedará pendiente realizar una prueba de independencia entre las variables.

```
selected_names = c("NO", "NO2", "NOX", "RH", "SR", "TOUT", "WSR")
```

```
x_pred = M_CENTRO[,selected_names]
x_pred
```

```
## # A tibble: 11,333 × 7
##       NO    NO2    NOX    RH    SR    TOUT    WSR
##   <dbl> <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
## 1   3.5  44.9  48.5   46 0     21.4   3.4
## 2   2.9  32.9  36     48 0     20.8   3.3
## 3   2.9  27.6  30.7   46 0     20.5   5.9
## 4   2.8  16.4  19.4   37 0     20.0   4.3
## 5   2.9  13.5  16.5   31 0     21.2   5.1
## 6   3     13.1  16.3   26 0.008  22.5   5.9
## 7   6.6  16.5  23.4   27 0.038  23.4   4.2
## 8  17.1  23.5  40.8   23 0.122  27.5   4.5
## 9   5.8  11.7  17.7   18 0.488  30.2   8.5
## 10  5.5  10.8  16.5   17 0.605  31.0  10.7
## # ... with 11,323 more rows
```

- b) Identifica columna objetivo

Para fines prácticos de esta entrega, elegiremos la variable O3 como variable dependiente. Esto se debe a la correlación moderada que tiene tanto con los contaminantes (NO,NOX,NO2) y con las variables metereológicas. Sin embargo, más adelante se planea experimentar con las variables "PM2.5" y "PM10".

```
y = M_CENTRO$O3
```

## 2) Limpieza de datos

a) Elimina duplicados

```
cat("Valores únicos por cada variable\n")

## Valores únicos por cada variable

for (i in 1:length(selected_names)) {
  cat("\n-",selected_names[i],":",length(unique(sapply(x_pred[,i],as.list))))
  #Transformamos las columnas a vectores para poder calcular sus valores únicos
}

##
## - NO : 578
## - NO2 : 596
## - NOX : 953
## - RH : 91
## - SR : 977
## - TOUT : 3076
## - WSR : 202

cat("\n\nNumero de observaciones totales: ",dim(x_pred)[1])

##
##
## Numero de observaciones totales: 11333

x_pred = distinct(x_pred)
cat("\n\nNumero de observaciones totales: ",dim(x_pred)[1])

##
##
## Numero de observaciones totales: 11333
```

Parace ser que no se encontraban observaciones duplicadas

b) Corrige valores erróneos

La corrección de los errores se realizaron al principio de la etapa 2, donde originalmente el tipo de dato de las observaciones de todas las variables eran de carácter.

c) Maneja valores faltantes

En base al analisis que realizamos de los valores faltantes en los puntos anteriores, vimos que no representan un porcentaje significativo en respecto con las observaciones totales, por lo que decidimos eliminarlos. Sin embargo, queda pendiente si hacer otras técnicas de sustitución de NAs mejora el rendiimiento del modelo multivariado-tal y como sustituirlos por la media de cada variable.

d) Maneja datos categóricos: Transforma a datos numéricos si es necesario.

Por el momento, decidimos no incluir la variable de las fechas en nuestro análisis. Todavía no descartamos el potencial que pudiera representar esta variable para realizar un análisis de temporalidad o épocas del año.

- e) Maneja adecuadamente los valores atípicos (outliers) que encuentres en el dataset.

Ahora, trataremos de detectar los outliers o valores atípicos dentro de los datos de la estación central (M\_CENTRO) calculando los rangos intercuartílicos (IQR) de cada variable.

*#Detección de datos atípicos al calcular los límites de los valores atípicos utilizando IQR*

```
outliersTest1 = grubbs.test(x_pred$NO)
outliersTest2 = grubbs.test(x_pred$NO2)
outliersTest3 = grubbs.test(x_pred$NOX)
outliersTest4 = grubbs.test(x_pred$RH)
outliersTest5 = grubbs.test(x_pred$SR)
outliersTest6 = grubbs.test(x_pred$TOUT)
outliersTest7 = grubbs.test(x_pred$WSR)

outliersTest1

##
##  Grubbs test for one outlier
##
## data:  x_pred$NO
## G = 10.63936, U = 0.99001, p-value < 2.2e-16
## alternative hypothesis: highest value 126.6 is an outlier
```

El test de Grubbs resulta útil para detectar valores atípicos univariados, pero una de las desventajas es que nomás detecta uno al correr la función, por lo que se necesitaría hacer un correrse dentro de un ciclo para encontrarlos- si es que hay más de uno.

### Transformación de Datos

- a) Revisa si es necesario discretizar los datos (binning)
- b) Si es necesario escala y normaliza los datos.

```
scaled_X = scale(x_pred)
scaled_X[1:10,]

##           NO           NO2           NOX           RH           SR           TOUT
## [1,] -0.4068103  2.39458670  1.1386319 -0.2399432 -0.7173692 -0.24515713
## [2,] -0.4606503  1.35312842  0.5184848 -0.1408135 -0.7173692 -0.31735136
## [3,] -0.4606503  0.89315102  0.2555424 -0.2399432 -0.7173692 -0.35879620
## [4,] -0.4696236 -0.07887670 -0.3050706 -0.6860269 -0.7173692 -0.42163192
## [5,] -0.4606503 -0.33056245 -0.4489447 -0.9834160 -0.7173692 -0.27456959
## [6,] -0.4516770 -0.36527773 -0.4588670 -1.2312403 -0.6901244 -0.09007321
## [7,] -0.1286370 -0.07019788 -0.1066235 -1.1816754 -0.5879563  0.02757665
## [8,]  0.8135628  0.53731944  0.7566213 -1.3799348 -0.3018856  0.57839192
```

```
## [9,] -0.2004237 -0.48678119 -0.3894106 -1.6277591 0.9445653 0.93401537
## [10,] -0.2273437 -0.56489056 -0.4489447 -1.6773239 1.3430210 1.04096979
##          WSR
## [1,] -1.26160782
## [2,] -1.28724059
## [3,] -0.62078875
## [4,] -1.03091295
## [5,] -0.82585085
## [6,] -0.62078875
## [7,] -1.05654572
## [8,] -0.97964743
## [9,] 0.04566309
## [10,] 0.60958388

cat("\n")
```

c) Construye atributos si es conveniente (atributos derivados).

## Referencias

- [1][[https://www.who.int/es/news-room/fact-sheets/detail/ambient-\(outdoor\)-air-quality-and-health](https://www.who.int/es/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health)]([https://www.who.int/es/news-room/fact-sheets/detail/ambient-\(outdoor\)-air-quality-and-health](https://www.who.int/es/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health))
- [2][<https://www.weather.gov/safety/airquality>](<https://www.weather.gov/safety/airquality>)
- [3][<https://www.ecologistasenaccion.org/17842/que-son-las-pm25-y-como-afectan-a-nuestra-salud/>](<https://www.ecologistasenaccion.org/17842/que-son-las-pm25-y-como-afectan-a-nuestra-salud/>)
- [4][<https://www.epa.gov/ground-level-ozone-pollution/ground-level-ozone-basics>](<https://www.epa.gov/ground-level-ozone-pollution/ground-level-ozone-basics>)
- [5][<https://www.epa.gov/so2-pollution/sulfur-dioxide-basics>](<https://www.epa.gov/so2-pollution/sulfur-dioxide-basics>)
- [6][<https://www.theatlantic.com/health/archive/2016/02/an-american-history-of-lead-poisoning/462576/>](<https://www.theatlantic.com/health/archive/2016/02/an-american-history-of-lead-poisoning/462576/>)
- [7][<https://www.milenio.com/estados/calidad-del-aire-como-se-mide-y-que-es>](<https://www.milenio.com/estados/calidad-del-aire-como-se-mide-y-que-es>)
- [8][<https://www.ecologiaverde.com/que-es-y-como-se-mide-la-calidad-del-aire-2423.html>](<https://www.ecologiaverde.com/que-es-y-como-se-mide-la-calidad-del-aire-2423.html>)
- [9][<https://www.unep.org/es/noticias-y-reportajes/reportajes/como-se-mide-la-calidad-del-aire>](<https://www.unep.org/es/noticias-y-reportajes/reportajes/como-se-mide-la-calidad-del-aire>)



[10][[https://www.epa.gov/sites/default/files/2015-05/documents/aqm\\_200608\\_challenges.pdf](https://www.epa.gov/sites/default/files/2015-05/documents/aqm_200608_challenges.pdf)]([https://www.epa.gov/sites/default/files/2015-05/documents/aqm\\_200608\\_challenges.pdf](https://www.epa.gov/sites/default/files/2015-05/documents/aqm_200608_challenges.pdf))

[11][<https://www.un.org/es/our-work/support-sustainable-development-and-climate-action>](<https://www.un.org/es/our-work/support-sustainable-development-and-climate-action>)

[12][<https://www.gob.mx/semarnat/acciones-y-programas/programas-de-gestion-para-mejorar-la-calidad-del-aire>](<https://www.gob.mx/semarnat/acciones-y-programas/programas-de-gestion-para-mejorar-la-calidad-del-aire>)

[13][<http://aire.nl.gob.mx/>](<http://aire.nl.gob.mx/>)

[14] [https://consejocivico.org.mx/noticias/2023/03/23/aumenta-preocupacion-sobre-mala-calidad-del-aire-ciudadania-respalda-mayores-medidas/#:~:text=La%20contaminaci%C3%B3n%20del%20aire%20es,contaminaci%C3%B3n%20del%20agua%20\(17.6%25\).](https://consejocivico.org.mx/noticias/2023/03/23/aumenta-preocupacion-sobre-mala-calidad-del-aire-ciudadania-respalda-mayores-medidas/#:~:text=La%20contaminaci%C3%B3n%20del%20aire%20es,contaminaci%C3%B3n%20del%20agua%20(17.6%25).)