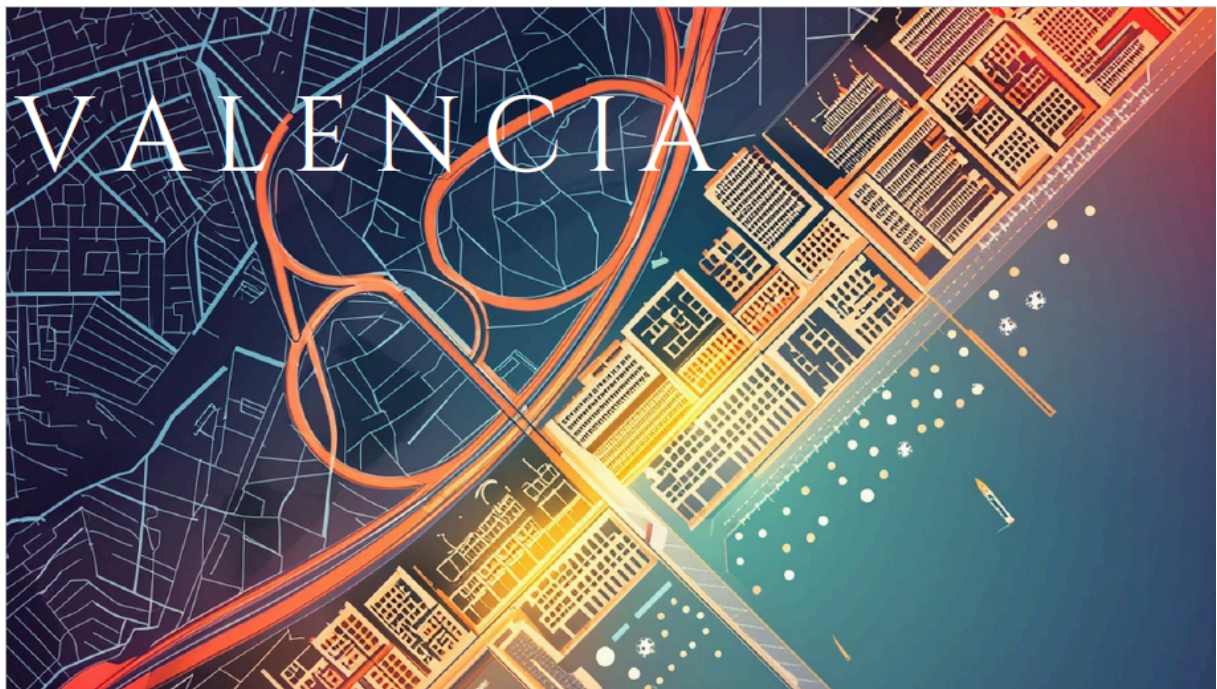


# **INFORME TÉCNICO**

## **PORTADA**

**Título**

**Proyecto Visualización.**



### **Autores.**

Roger Pastor Juan,

Claudia Min García,

David Camarena Sánchez,

Gonzalo Sánchez Muñoz,

Sebastián Gutierrez Ferreira

**Asignatura y curso: Segundo Ciencia de datos.**

# INFORME

## Índice.

- **Introducción:**
  - Presentación del trabajo.
  - Objetivos.
  
- **1. Metodología:**
  - 1.1. Análisis exploratorio de los datos
  - 1.2. Preprocesado de Datos.
  - 1.3. Preparación de Datos.
  - 1.4. Elección de gráficas para los distintos tipos de datos
  - 1.5. Diseño del mapa
  - 1.6. Diseño del cuadro de mandos e Implementación.
    - Implementación de los gráficos interactivos
    - Implementación de los mapas interactivos
  
- **2. Resultados:**
  - Muestra de mapas y resultados
  
- **3. Discusión**
  - Interpretación de los resultados y análisis de los hallazgos

## **INTRODUCCIÓN.**

Nuestro objetivo es a través de los datos abiertos proporcionado por el ayuntamiento de valencia:

- Contaminación:  
<https://valencia.opendatasoft.com/explore/dataset/rvvcca/information/>  
<https://valencia.opendatasoft.com/explore/dataset/estacions-contaminacio-atmosferiques-estaciones-contaminacion-atmosfericas/table/>
- Distritos:  
<https://valencia.opendatasoft.com/explore/dataset/districtes-distritos/informatio/n/?location=10,39.42291,-0.35395&basemap=e4bf90>

El siguiente informe técnico detalla el trabajo realizado en el marco del proyecto de Visualización de Datos, centrado en el análisis y representación gráfica de la contaminación atmosférica en la ciudad de Valencia.

El objetivo principal de este proyecto ha sido explorar, procesar y visualizar los datos relacionados con la contaminación atmosférica, utilizando herramientas como RStudio, QGIS y Shiny. A través de diversas técnicas de análisis exploratorio, preprocesamiento de datos y creación de gráficos interactivos, hemos buscado entender y analizar mejor la distribución y evolución de los contaminantes en distintas áreas de la ciudad.

Este informe ofrece una visión detallada de la metodología empleada, desde la importación y limpieza de datos hasta la implementación de gráficos interactivos en Shiny. Se presentan los resultados obtenidos, incluyendo muestras de los mapas generados y los hallazgos más relevantes. Se interpretan los resultados, destacando las tendencias observadas y las implicaciones para la calidad del aire en Valencia.

A lo largo del informe, se destacan las decisiones tomadas durante el proceso de análisis y visualización de datos, así como las limitaciones encontradas y las posibles áreas de mejora. En última instancia, este proyecto no solo proporciona una representación visual de la contaminación atmosférica en Valencia, sino que también sirve como una plataforma para futuras investigaciones y acciones en el campo de la ciencia de datos y la calidad ambiental.

## 1. METODOLOGÍA

### o 1.1. Análisis exploratorio de los datos

Queremos saber qué datos tenemos y poder trabajar más fácilmente con ellos. para ello miramos la estructura interna de los datos con funciones como son summary y str.

```
## {r}
#Analizamos

# Mostrar la estructura de un objeto
print(str(rvvcca))

# Mostrar un resumen de un objeto
print(summary(rvvcca))

...

spec_TDI_ [45,388 x 34] (>): spec_TDI_at/TDI_at/TDI/data.TFrame)
 $ Id                : num [1:43388] 1 3 5 7 8 12 15 16 19 27 ...
 $ Fecha            : Date[1:43388], format: "2004-01-01" "2004-01-02" "2004-01-03" "2004-01-04" ...
 $ Día de la semana : chr [1:43388] "Jueves" "Viernes" "Sabado" "Domingo" ...
 $ Día del mes      : num [1:43388] 1 2 3 4 4 6 8 8 10 14 ...
 $ Estacion         : chr [1:43388] "Pista Silla" "Pista Silla" "Pista Silla" "Pista Silla" ...
 $ PM1              : num [1:43388] NA NA NA NA NA NA NA NA NA NA ...
 $ PM2.5           : num [1:43388] NA NA NA NA NA NA NA NA NA NA ...
 $ PM10            : num [1:43388] NA NA NA NA 27.8 28.3 NA 57.4 NA NA ...
 $ NO               : num [1:43388] 14 13 60 143 59 37 69 43 58 29 ...
 $ NO2             : num [1:43388] 33 38 70 99 51 45 76 50 52 38 ...
 $ NOx             : num [1:43388] 55 57 162 319 142 103 182 115 141 83 ...
 $ O3              : num [1:43388] 33 53 35 5 9 9 15 15 15 22 ...
 $ SO2             : num [1:43388] 3 3 3 4 6 5 3 8 3 3 ...
 $ CO              : num [1:43388] 0.3 0.3 0.6 1.2 0.9 0.6 0.7 0.5 0.5 0.4 ...
 $ NH3             : num [1:43388] NA NA NA NA NA NA NA NA NA NA ...
 $ C7H8           : num [1:43388] NA NA NA NA 2.4 1.3 NA 2.9 NA NA ...
```

## ○ 1.2.Preprocesado de Datos.

Observamos que nuestro fichero de campos tiene muchos datos faltantes y datos sobrantes también hay columnas que carecen de importancia. Según encontramos el problema lo vamos tratando.

Inicialmente el mayor problema viene dado por el formato de los datos, para ello lo primero que hacemos es transformarlos o cambiarles el nombre para que sean más cómodos de manejar.

El preprocesado de datos realizado tiene como objetivo limpiar y preparar los datos para su posterior análisis y visualización. A continuación listamos los pasos seguidos para este punto:

- **Importación de datos:** Se importan los datos de varios archivos CSV y archivos shapefile (.shp) utilizando las funciones `read_delim()` y `st_read()`. Estos archivos contienen información sobre estaciones de monitoreo de contaminación atmosférica, puntos de medición de bicicletas, puntos de medición de tráfico, movilidad urbana, zonas verdes, entre otros.
- **Análisis exploratorio inicial de los datos:** Se realiza un análisis exploratorio inicial para comprender la estructura y los resúmenes de los datos importados. Esto incluye el uso de funciones como `str()` y `summary()` para examinar la estructura y resumir los datos.
- **Limpieza de datos:** Se lleva a cabo una serie de acciones para limpiar los datos, incluyendo:
  - **Transformación de variables:** Se transforman algunas variables de carácter a factor para facilitar su manipulación.
  - **Renombrado de variables:** Se cambian los nombres de las variables para mayor claridad y consistencia en el análisis.

```

{r}
#observamos que es de tipo character
str(rvvcca$Estacion)

#transformamos a factor para facilitarnos el trabajo.

rvvcca <- rvvcca %>%
mutate("Dia de la semana" = factor(rvvcca$`Dia de la semana`)) %>%
mutate(Estacion = factor(rvvcca$Estacion)) %>%
mutate("Dia del mes" = factor(rvvcca$`Dia del mes`))
...

**LIMPIEZA DE DATOS DE FICHERO CON LOS REGISTROS DE CONTAMINACIÓN**

Vamos a cambiar el nombre de las variables para mayor manejabilidad

{r}
#Filtrar datos
rvvcca <- rvvcca %>% rename(Dia_semana = `Dia de la semana`)%>%
  rename(Dia_mes = `Dia del mes`) %>%
  rename(PM2_5 = `PM2.5`) %>%
  rename(Velocidad_viento = `Velocidad del viento`) %>%
  rename(Direccion_viento = `Direccion del viento`) %>%
  rename(Humedad_relativa = `Humidad relativa`) %>%
  rename(Radiacion_solar = `Radiacion solar`) %>%
  rename(Velocidad_max_viento = `Velocidad maxima del viento`)%>%
  rename(As = `As (ng/m³)`) %>%
  rename(Ni = `Ni (ng/m³)`) %>%
  rename(Cd = `Cd (ng/m³)`) %>%
  rename(Pb = `Pb (ng/m³)`) %>%
  rename(Bap = `B(a)p (ng/m³)`) %>%
  rename(Fecha_creacion = `Fecha creacion`) %>%
  rename(Fecha_baja = `Fecha baja`)

```

- **Filtrado de datos:** Se filtran los datos para trabajar únicamente con las estaciones de monitoreo de contaminación atmosférica que coinciden con las estaciones disponibles en un archivo shapefile, lo que implica seleccionar solo las estaciones relevantes para el análisis.

```

{r}
#sacamos un vector para poder estudiar las estaciones coincidentes
unique(rvvcca$Estacion)

#filtramos solos las estaciones coincidentes con el ggis

rvvcca_sr <- rvvcca %>%
  filter(Estacion %in% c("Viveros", "Politecnico", "Avda. Francia", "Valencia Centro", "Bulevard Sud",
    "Pista Silla", "Moli del Sol", "Valencia Olivereta"))

#Comprobamos que se han eliminado correctamente
unique(rvvcca_sr$Estacion)

#Cambiamos los nombres para que sean comunes a las estaciones del otro fichero
rvvcca_sr <- rvvcca_sr %>% mutate(Estacion = recode(Estacion,
  "Avda. Francia" = "Francia",
  "Bulevard Sud" = "Boulevard Sur",
  "Moli del Sol" = "Moli del Sol",
  "Pista Silla" = "Pista de Silla",
  "Politecnico" = "Universidad Politécnica",
  "Puerto llit antic Turia" = "Puerto llit antic Turia(s)",
  "Puerto Moll Trans. Ponent" = "Puerto Moll Trans. Ponent(s)",
  "Valencia Centro" = "Centro",
  "Valencia Olivereta" = "Olivereta",
  "Viveros" = "Viveros"))

unique(rvvcca_sr$Estacion)

#También queremos eliminar columnas con 95% de nas

# Contar los NA en todas las columnas del dataframe df
num_na_por_columna <- colSums(is.na(rvvcca_sr))
num_na_por_columna

```

- **Eliminación de columnas con alto porcentaje de valores faltantes (NA):** Se eliminan las columnas que tienen más del 85% de valores faltantes para simplificar el análisis.

```

Limpieza de NAs
{r}
#También queremos eliminar columnas con 95% de nas

# Contar los NA en todas las columnas del dataframe df
num_na_por_columna <- colSums(is.na(rvcca_sr))
num_na_por_columna

length(rvcca_sr$NH3)
#36236

# Calcular el porcentaje de NA por columna
porcentaje_na <- colSums(is.na(rvcca_sr)) / nrow(rvcca_sr) * 100

# Identificar las columnas que tienen un 85% o más de NA
columnas_a_eliminar <- names(porcentaje_na[porcentaje_na >= 85])

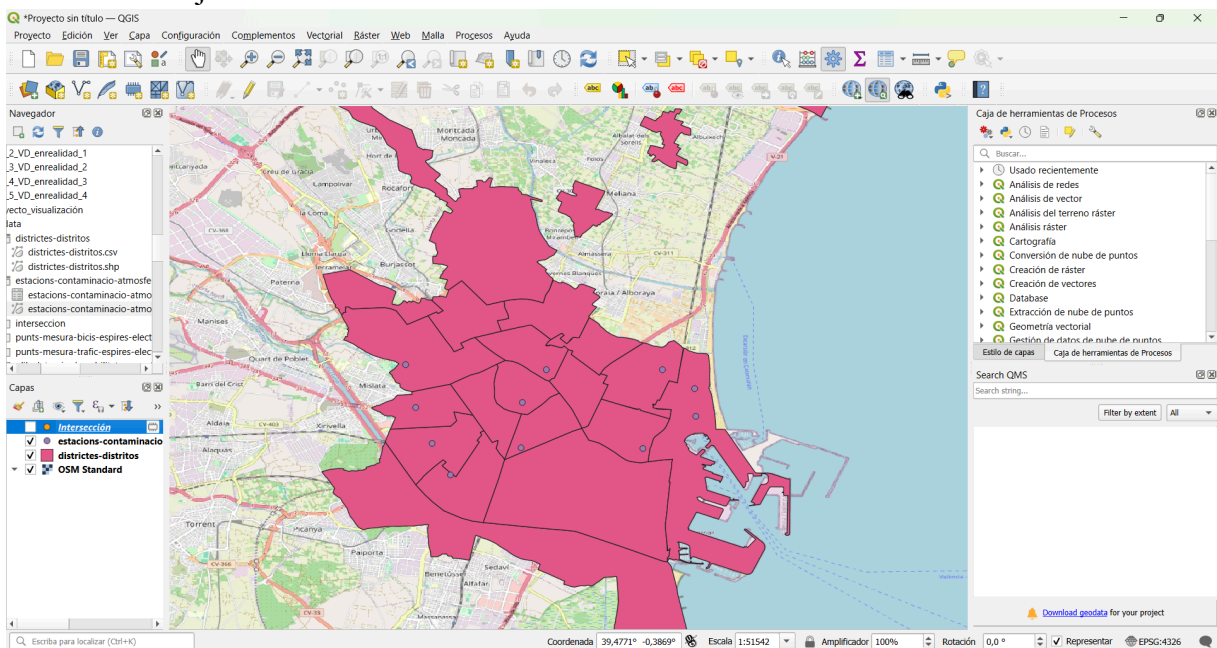
# Eliminar las columnas identificadas
rvcca_sr <- rvcca_sr[, !names(rvcca_sr) %in% columnas_a_eliminar]

```

- **Creación de un nuevo conjunto de datos conjunto:** Se crea un nuevo conjunto de datos combinando la información de contaminación atmosférica con la información de los distritos geográficos, lo que facilitará la visualización en un mapa interactivo.

Queremos comparar nuestros valores del data frame con los registros para cada estación y conseguir mapear con ellos, para ello lo que haremos es enlazar la información que tenemos de los registros en el csv por distritos ya que tenemos información de aquellos donde se registran contaminación.

Para ello en primera instancia hacemos una intersección desde qgis de nuestros campos distritos en conjunto con las estaciones.



Al hacer la intersección esta es la tabla de datos que nos deja, nos deja las estaciones y su distrito.

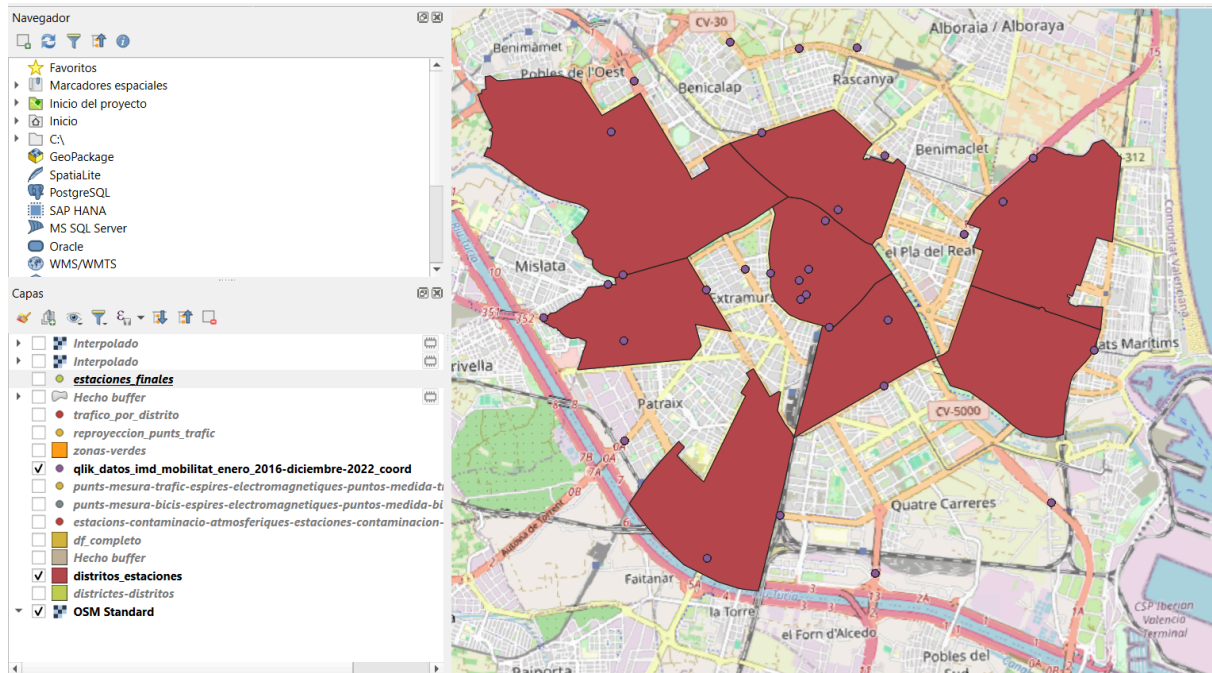
objectid	nombre	direccion	tipozona	tipoemision	parametros	mediciones	so2	no2	o3	co	pm10	pm25	fecha_carga	calidad_amb
24	Boulevard Sur	Avda. Tres Cruc...	Urbana	Tráfico	Dióxido de azuf...	<a href="http://mapas.va...">http://mapas.va...</a>	2,0000000000	4,0000000000	91,0000000000	NULL	NULL	NULL	2024-04-20 20...	Razonablement... 14
430	Dr. Lluch	Calle Dr. Lluch, ...	Urbana	Tráfico	Óxidos de nitró...	NULL	NULL	52,0000000000	NULL	NULL	27,0000000000	10,0000000000	2024-04-20 20...	Razonablement... 18
431	Olivereta	Av. del Cid amb...	Urbana	Tráfico	Óxidos de nitró...	NULL	NULL	15,0000000000	NULL	NULL	23,0000000000	10,0000000000	2024-04-20 20...	Razonablement... 15
432	Patraix	Archiducque Car...	Urbana	Tráfico	Óxidos de nitró...	NULL	NULL	8,0000000000	NULL	NULL	22,0000000000	10,0000000000	2024-04-20 20...	Razonablement... 23
26	Pista de Silla	C/ Filipinas, s/n	Urbana	Tráfico	Dióxido de azuf...	<a href="http://mapas.va...">http://mapas.va...</a>	4,0000000000	12,0000000000	77,0000000000	0,7	8,0000000000	2,0000000000	2024-04-20 20...	Razonablement... 12
25	Molí del Sol	Avda. Pio Baroj...	Suburbana	Tráfico	Dióxido de azuf...	<a href="http://mapas.va...">http://mapas.va...</a>	2,0000000000	15,0000000000	85,0000000000	0,3	6,0000000000	4,0000000000	2024-04-20 20...	Razonablement... 10
28	Viveros	Jardines de Viv...	Urbana	Fondo	Dióxido de azuf...	<a href="http://mapas.va...">http://mapas.va...</a>	1,0000000000	13,0000000000	91,0000000000	NULL	NULL	NULL	2024-04-20 20...	Razonablement... 6
23	Francia	Avda. de Franci...	Urbana	Tráfico	Dióxido de azuf...	<a href="http://mapas.va...">http://mapas.va...</a>	3,0000000000	5,0000000000	77,0000000000	0,1	9,0000000000	4,0000000000	2024-04-20 20...	Razonablement... 13
27	Universidad Pol...	Campus de la U...	Suburbana	Fondo	Dióxido de azuf...	<a href="http://mapas.va...">http://mapas.va...</a>	2,0000000000	8,0000000000	86,0000000000	NULL	19,0000000000	11,0000000000	2024-04-20 20...	Razonablement... 71
22	Centro	Plaza Ayuntami...	Urbana	Tráfico	Óxidos de nitró...	<a href="https://mapas.v...">https://mapas.v...</a>	NULL	16,0000000000	NULL	NULL	19,0000000000	8,0000000000	2024-04-20 20...	Buena 9
829	Cabanyal	Carrer del Prog...	Urbana	Fondo	NULL	NULL	NULL	24,0000000000	NULL	NULL	NULL	NULL	2024-04-20 20...	Razonablement... 18

o3	co	pm10	pm25	fecha_carga	calidad_amb	objectid_2	nombre_2	coddistr	gis_gis_dis	latitud	longitud
91,0000000000	NULL	NULL	NULL	2024-04-20 20...	Razonablement... 14	JESUS	9	2984760,00000...	39.4482014775...	-0.3917921797...	
NULL	NULL	27,0000000000	10,0000000000	2024-04-20 20...	Razonablement... 183	POBLATS MARI...	11	NULL	39.4519525304...	-0.3263356826...	
NULL	NULL	23,0000000000	10,0000000000	2024-04-20 20...	Razonablement... 151	L'OLIVERETA	7	NULL	39.4688834120...	-0.4033652971...	
NULL	NULL	22,0000000000	10,0000000000	2024-04-20 20...	Razonablement... 23	PATRAIX	8	NULL	39.4583977479...	-0.4004125691...	
77,0000000000	0,7	8,0000000000	2,0000000000	2024-04-20 20...	Razonablement... 12	L'EIXAMPLE	2	1733140,50000...	39.4641145233...	-0.3704292703...	
85,0000000000	0,3	6,0000000000	4,0000000000	2024-04-20 20...	Razonablement... 103	CAMPANAR	4	NULL	39.4854436586...	-0.4059322619...	
91,0000000000	NULL	NULL	NULL	2024-04-20 20...	Razonablement... 6	LA SAIDIA	5	1943931,50000...	39.4850946796...	-0.3750723425...	
77,0000000000	0,1	9,0000000000	4,0000000000	2024-04-20 20...	Razonablement... 13	CAMINS AL GR...	12	2367473,25000...	39.4627947452...	-0.3478024542...	
86,0000000000	NULL	19,0000000000	11,0000000000	2024-04-20 20...	Razonablement... 71	ALGIROS	13	NULL	39.4776270293...	-0.3427393297...	
NULL	NULL	19,0000000000	8,0000000000	2024-04-20 20...	Buena 9	CIUTAT VELLA	1	1689851,62500...	39.4744154278...	-0.3767596965...	
NULL	NULL	NULL	NULL	2024-04-20 20...	Razonablement... 183	POBLATS MARI...	11	NULL	39.4519525304...	-0.3263356826...	

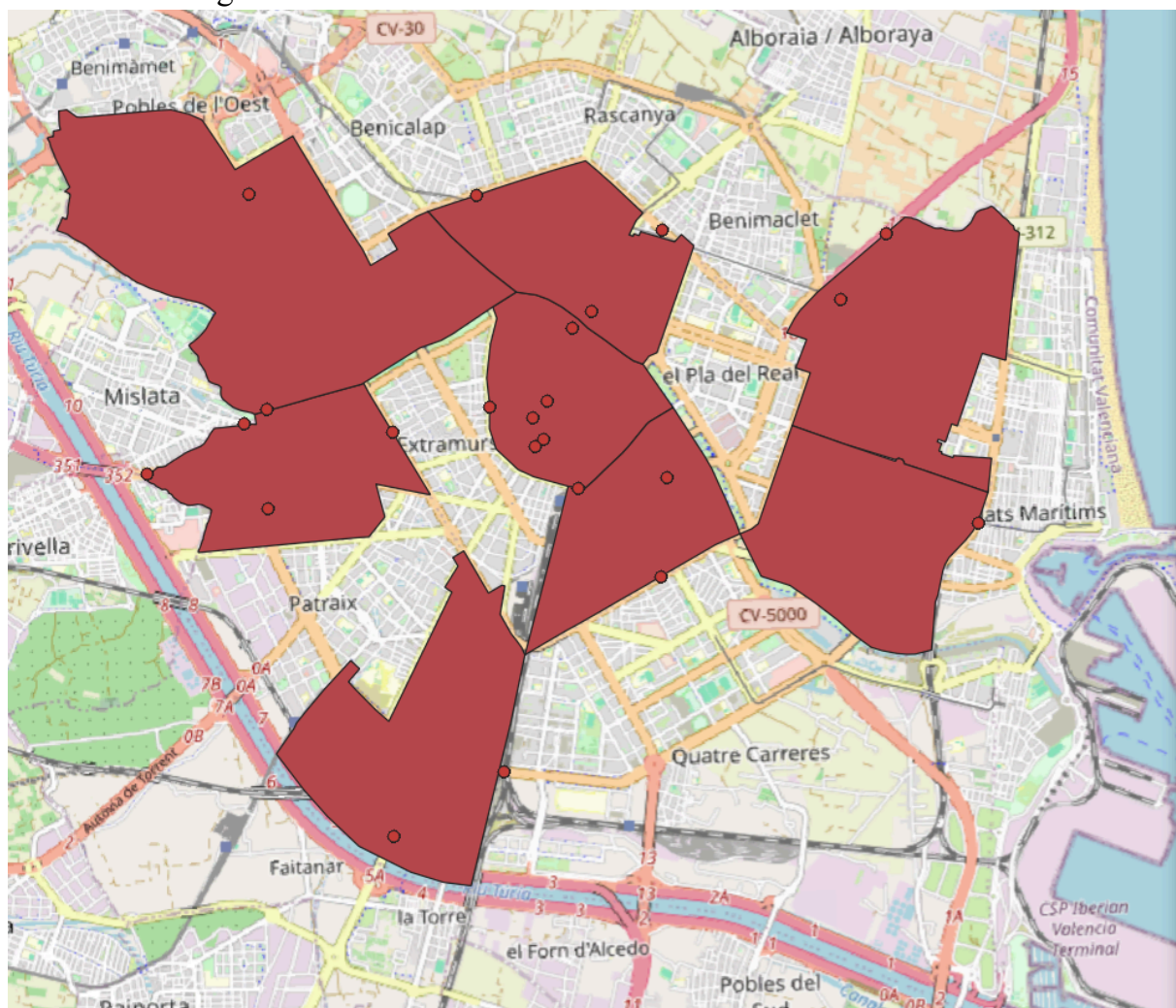
Esta nueva capa la utilizaremos para crear un nuevo campo en nuestro dataframe para poder así representar más fácilmente.

Por otra parte vamos a tratar el shp que mide la contaminación del tráfico, (qlik\_datos\_imd\_mobilitat\_enero\_2016-diciembre-2022\_coord.shp), haciendo primero una intersección para quedarnos los datos que estén dentro de los distritos que tenemos datos.





obteniendo el siguiente resultado.



De esta forma tenemos los puntos de medida de tráfico por distrito.

Vamos a calcular el IMD medio de cada distrito, para ello vamos a exportar la capa para trabajar con ella en R, en donde creamos un código para cada calcular la media de cada distrito y posteriormente añadirlo a los nuevos datos al data frame en un nuevo campo.

```
``{r}
trafico_distrito <- st_read(dsn="data/trafico_por_distrito/trafico_por_distrito.shp")
#Lo hacemos sobre el campo IMD (intensidad media diaria)
#vamos a estimar la media de coches por distrito

#tratamos los datos porque tiene un formato conflictivo al pasarlo a numerico
# ej: 23 455 para el numero 23455
#eliminamos los espacios en blanco entre los numeros
trafico_distrito$imd <- gsub(" ","", trafico_distrito$imd)

#convertimos a numerico
trafico_distrito$imd <- as.numeric(as.character(trafico_distrito$imd))

#creamos la media por distrito
media<- trafico_distrito %>%
  group_by(nombre) %>%
  summarize(media_imd_por_distrito = mean(imd))

media <- st_drop_geometry(media)
#media

#asignamos la media a cada distrito
trafico_distrito <- merge(trafico_distrito, media, by ="nombre")
``
```

- **Tratamiento de valores faltantes:** Se realiza un tratamiento de los valores faltantes, sustituyéndolos por la media de cada variable por distrito para preservar la integridad de los datos.
- **Visualización de datos:** Se generan diferentes gráficos y mapas para explorar y visualizar los datos, utilizando herramientas como ggplot2 para gráficos y leaflet para mapas interactivos. Se crean gráficos de líneas, gráficos de caja (boxplots) y mapas de calor para visualizar la evolución temporal y la distribución espacial de la contaminación atmosférica.

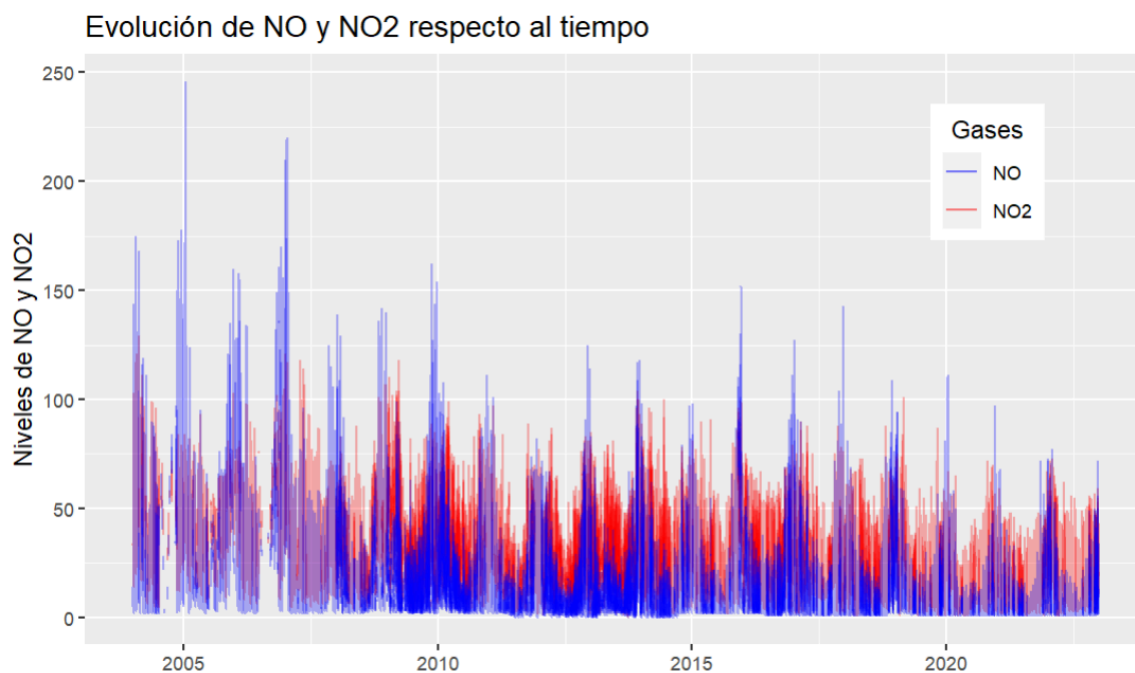
#### ○ 1.4. Elección de gráficas para los distintos tipos de datos

El diseño del cuadro de mandos está enfocado en la visualización de datos geoespaciales y series temporales relacionadas con la contaminación y el tráfico en Valencia. Se utilizan diversas librerías de R, como ggplot2, leaflet, plotly y DT, para la manipulación, visualización y presentación de datos.

El proceso comienza importando datos de múltiples fuentes, incluidos archivos CSV y archivos shapefile (shp). Luego se realiza un análisis exploratorio de los datos para comprender su estructura y contenido. Se llevan a cabo tareas de limpieza de datos, como la eliminación de valores nulos y la normalización de nombres de variables.

Una parte significativa del proyecto implica la creación de visualizaciones interactivas. Se generan mapas interactivos utilizando la librería leaflet, lo que permite explorar la distribución de la contaminación por distrito. También se crean gráficos interactivos utilizando plotly, lo que permite una visualización dinámica de la evolución temporal de la contaminación y el tráfico.

Además de las visualizaciones interactivas, se incluyen gráficos estáticos y tablas para proporcionar una representación más completa de los datos. Se implementan múltiples pestañas en la aplicación Shiny para organizar y presentar diferentes tipos de visualizaciones y análisis de datos.



En este gráfico de líneas generado en RStudio utilizando la librería ggplot2, estamos comparando la evolución de dos gases contaminantes, el NO (óxido de nitrógeno) y el NO2 (dióxido de nitrógeno), a lo largo del intervalo de tiempo entre el año 2000 y el 2024.

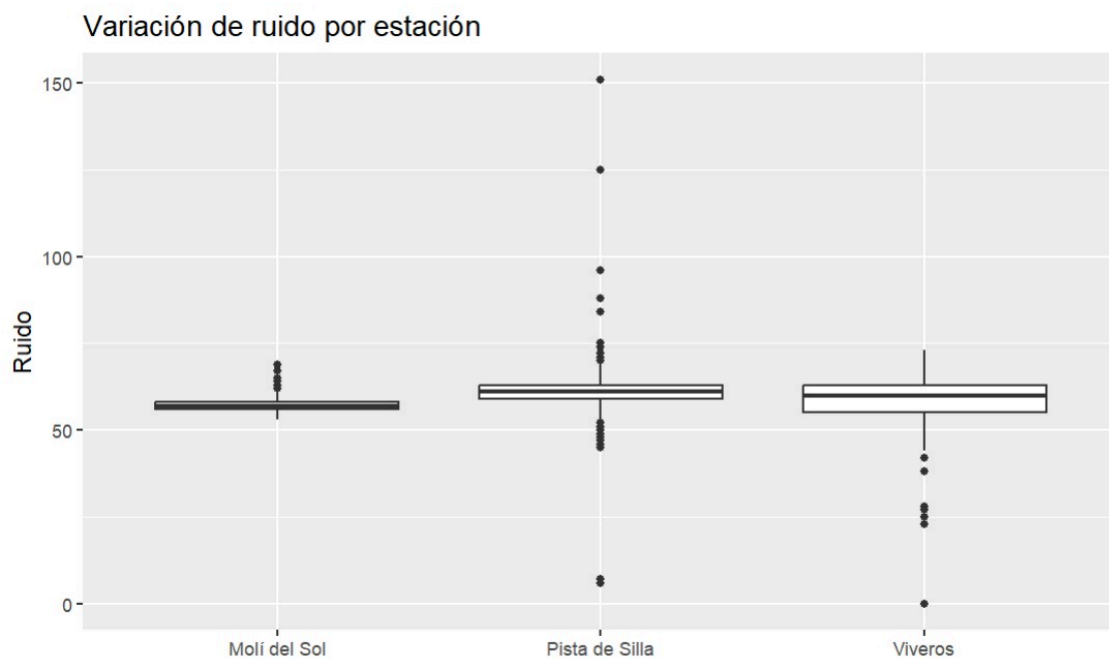
Ambos gases se representan como líneas en el mismo gráfico. Para distinguir entre ellas, hemos asignado colores diferentes: el NO está representado en azul, mientras que el NO2 está representado en rojo.

La variable en el eje X representa el tiempo, con valores desde el año 2000 hasta el 2024. Esto nos permite observar la evolución de los niveles de contaminación a lo largo de los años. Cabe recalcar que en los datos utilizados hay separación por días. Sin embargo, hemos decidido que en el gráfico se muestre solamente por años con el fin de no saturar las etiquetas del eje.

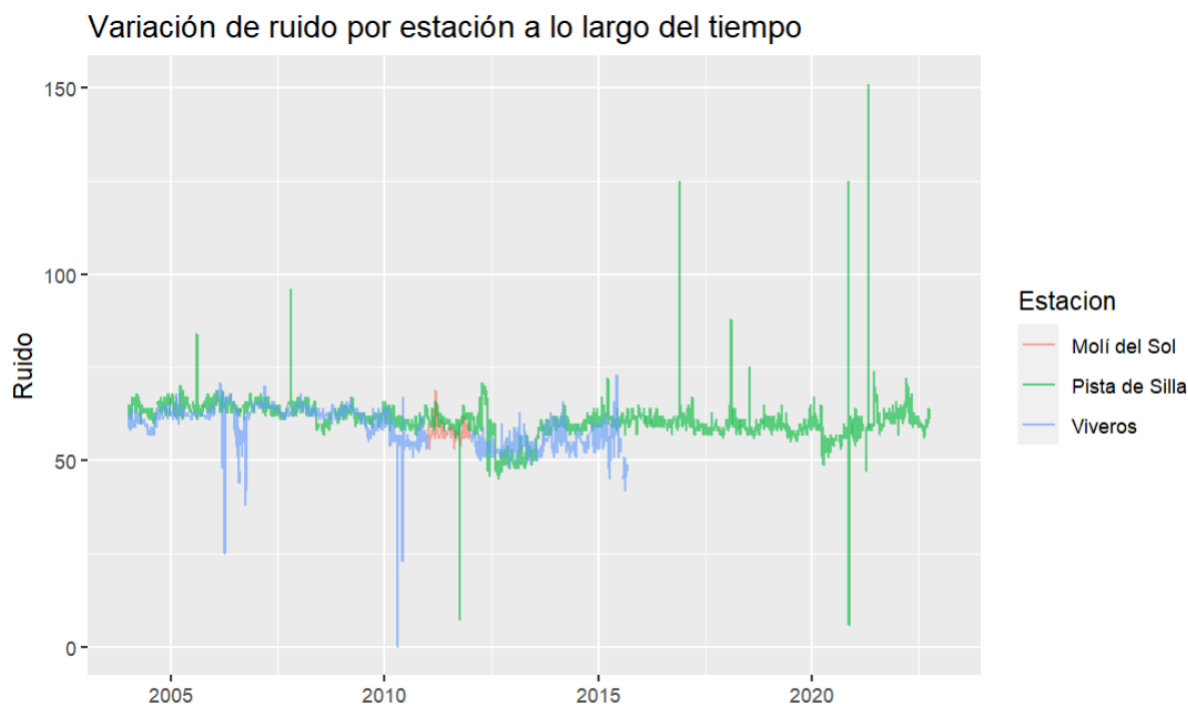
Por otro lado, en el eje Y, tenemos la concentración de cada gas contaminante. No obstante, desconocemos qué tipo de medidas se han utilizado ya que en el data frame empleado no se menciona.

La forma de las líneas nos muestra la tendencia general de la concentración de cada gas a lo largo del tiempo. Si las líneas son ascendentes, indica un aumento en la concentración del gas a medida que avanza el tiempo, lo que podría indicar un empeoramiento de la calidad del aire. Por otro lado, si las líneas son descendentes, sugiere una disminución en la concentración del gas, lo que sería una señal positiva en términos de calidad del aire. Los resultados para este caso han resultado en un descenso notable de ambos gases, lo que indica una mejoría en la calidad del aire para este caso, aunque para llegar a esta conclusión tendríamos que analizar otros parámetros.

Además, podemos observar cualquier patrón estacional o tendencias a largo plazo en la concentración de estos gases. Por ejemplo, si hay picos regulares en ciertos momentos del año, podría indicar factores estacionales, como el aumento de la actividad industrial o el tráfico durante ciertas estaciones.



En este boxplot, cada estación se representa de forma independiente en el gráfico. Cabe recalcar que hemos realizado un filtro solamente de estas tres estaciones debido a que para el resto no había valores, por lo que no aportan información. La altura del boxplot muestra la dispersión de los niveles de ruido registrados en cada estación, donde la línea en el medio representa la mediana. Se comparan entre sí para evaluar si hay diferencias significativas en los niveles de ruido entre las tres estaciones, siendo el caso de Pista de Silla el más relevante, ya que hay muchos valores atípicos y siendo también el más ruidoso de las tres estaciones.



Este lineplot nos permite visualizar cómo varían los niveles de ruido en cada estación a lo largo del tiempo y también comparar la evolución entre las diferentes estaciones. Podemos identificar tendencias generales en los niveles de ruido, como aumentos o disminuciones consistentes, así como patrones estacionales o picos anómalos que podrían requerir una mayor investigación. Con este gráfico y el anterior podemos observar que la estación con mayor información es la de Pista de Silla, además de ser la más ruidosa y con mayor cantidad de valores atípicos. También podemos observar que entre los años 2010 y 2015 hubo un pequeño descenso de los niveles de ruido. Sin embargo, parece que en los últimos años está habiendo un incremento del mismo.

### ○ 1.5. Diseño del mapa.

Tomando de referencia que nos centramos en los niveles de distintos contaminantes por lo que son los distritos hemos decidido implementar mapas que reflejen este hecho.

Para nosotros los más representativos o los que más reflejaban estos problemas que queríamos mostrar eran sin duda:

- un mapa de calor que pudiera variar con el tiempo para diferente tipo de contaminantes y que pudiera mostrar claramente la concentración de dicha sustancia contaminante.
- En segundo lugar uno más intuitivo y general esta vez centrándonos en los distritos mostrar la presencia de cada contaminante en unas fechas concretas.

Para ello hemos utilizado nuestro de banco de datos creado a través de la superposición de capas espaciales que hemos explicado en puntos anteriores .

### ○ 1.6. Diseño del cuadro de mandos e Implementación

- Implementación de los gráficos interactivos

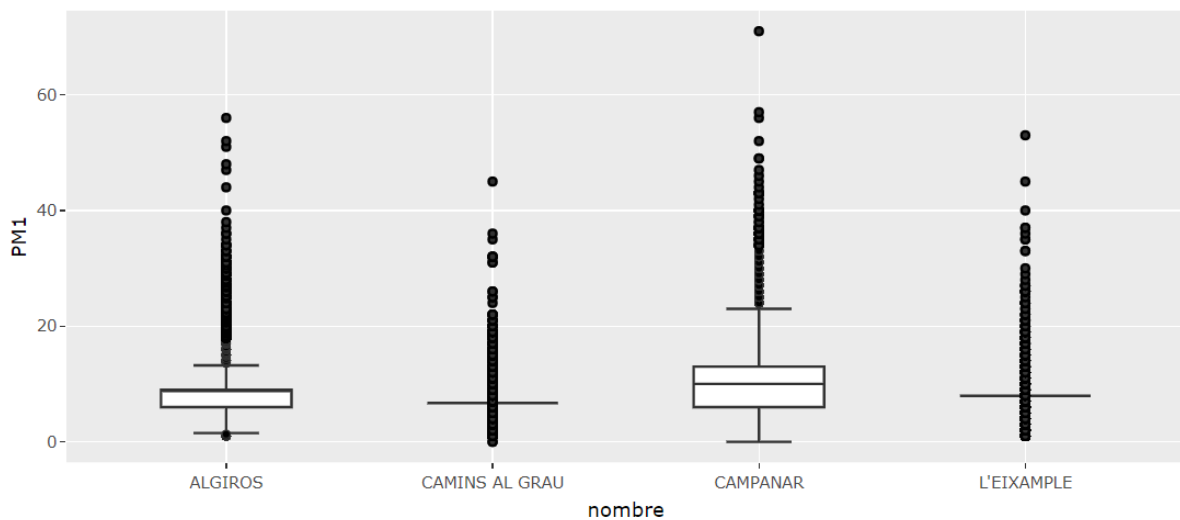
Los gráficos presentados anteriormente han sido introducidos en Shiny en una pestaña de “gráficos predeterminados” dentro de la pestaña de gráficos. Nuestro objetivo a continuación ha sido poder realizar gráficos “automáticos”, en los que el usuario puede elegir qué variables quiere observar. Hemos hecho un generador de gráficos de línea y dos generadores de tipo boxplot, uno a nivel temporal y otro en base a las estaciones, pero a nivel individual. En la segunda pestaña de boxplots también está la opción de realizar una observación en base a las estaciones, pero esta vez mostrando al mismo tiempo para poder sacar conclusiones a partir del mismo contaminante. Para este caso en específico hemos conseguido que de forma automática se visualizan solamente aquellas estaciones que contengan datos, ya que según el contaminante hay estaciones que no tienen datos, y por tanto no son importantes para los gráficos. A continuación un par de ejemplos:

Selecciona el eje x

nombre

Selecciona el eje y

PM1

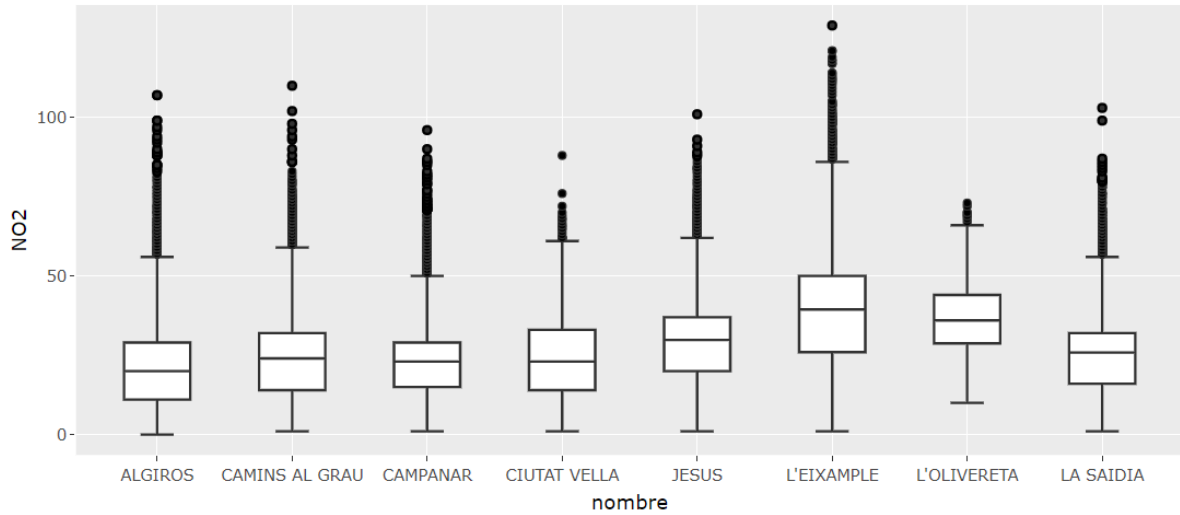


Selecciona el eje x

nombre

Selecciona el eje y

NO2



Como podemos observar según el contaminante elegido solamente se muestran las estaciones que contengan datos, filtrando así aquellos que son innecesarios.

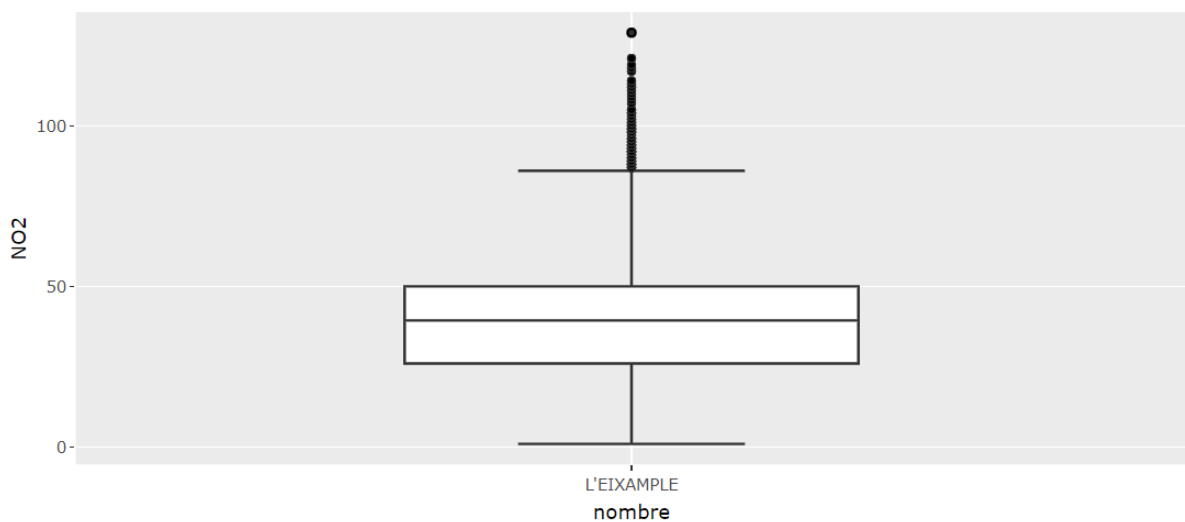
Para el boxplot individual nos centramos más en una sola variable a analizar:

Selecciona estación

L'EIXAMPLE

Selecciona contaminante

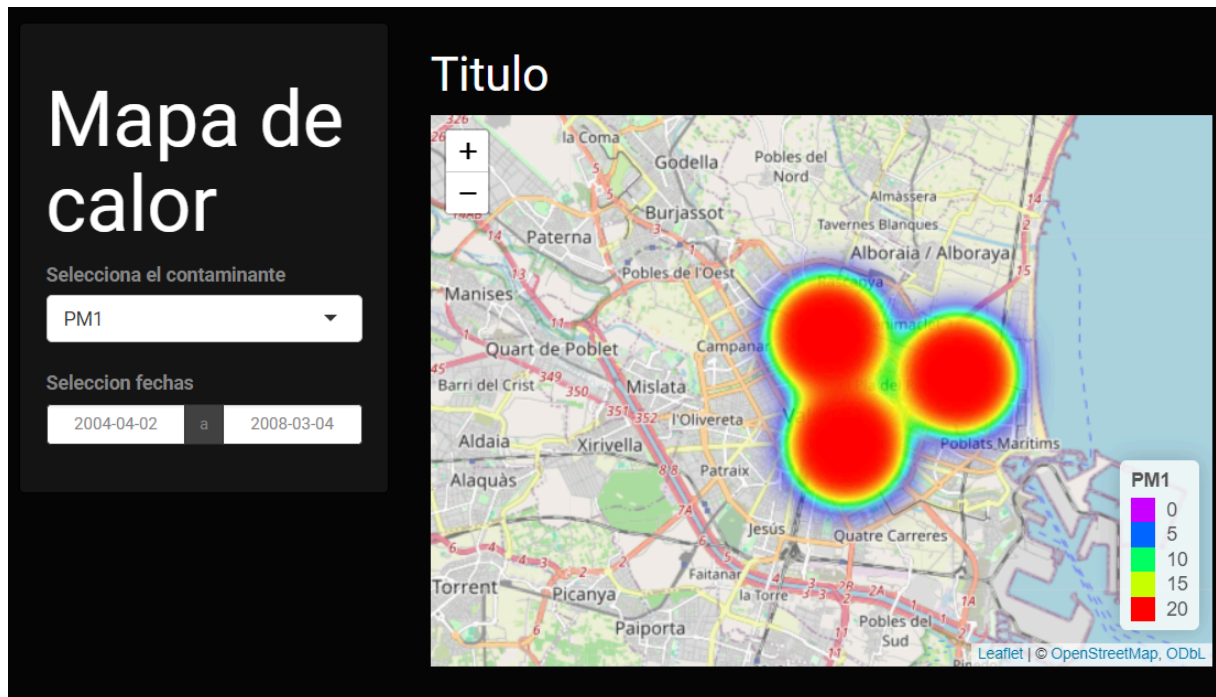
NO2



■ Implementación de los mapas interactivos

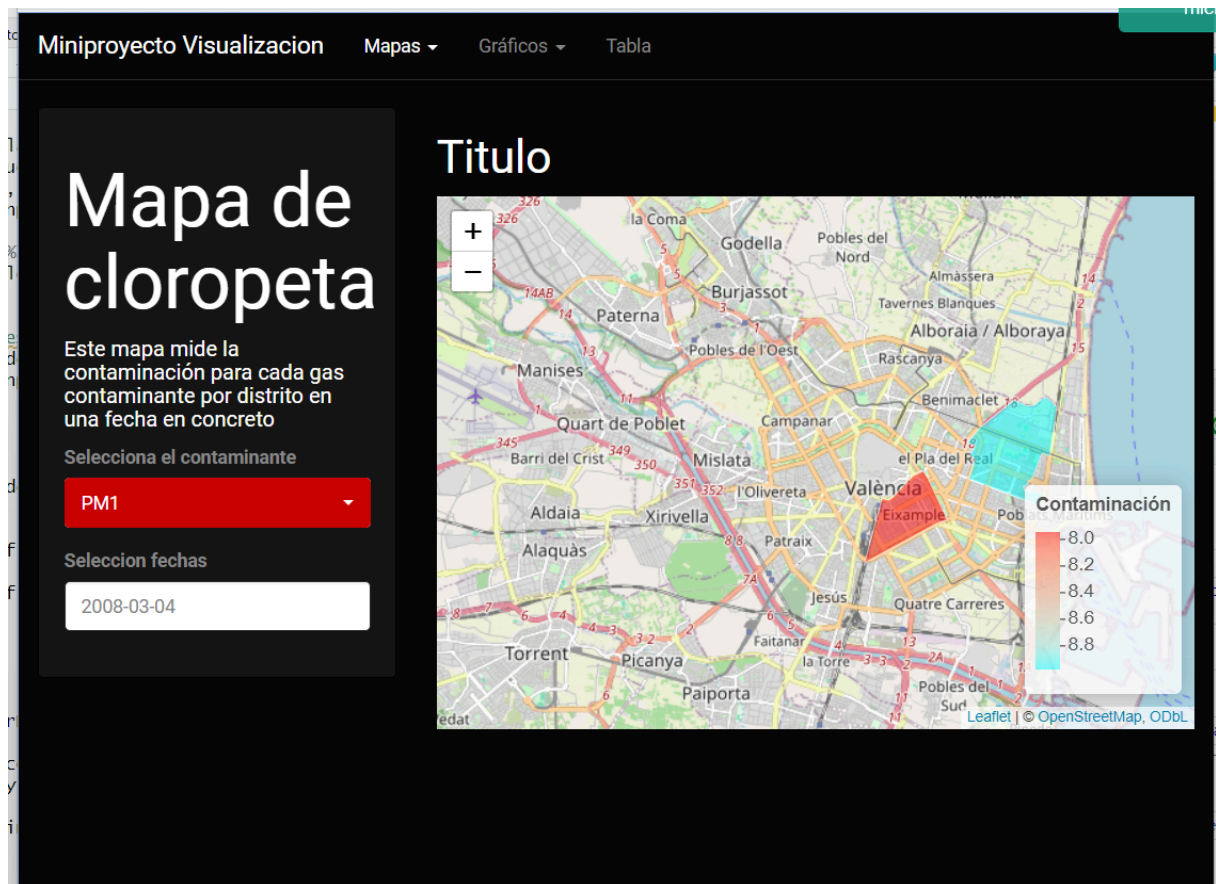
Para que sean interactivos hemos dado opcionalidad al cliente para elegir el elemento contaminante en el que esté interesado en un rango de fechas disponibles (hemos especificado que sea desde la menor fecha a la mayor fecha que tenemos.) para mayor facilidad en shiny también hemos creado algunas funciones para que el código quede más ordenado.

Dentro del código también hemos creado nuevos dataframes que nos ayuden a filtrar la información.



- **2. Resultados:**
  - Muestra de mapas y resultados





Gráficas resultantes implementadas en shiny:

### ● **3. Discusión**

- Interpretación de los resultados y análisis de los hallazgos.

Podemos llegar a afirmar que nuestro banco de datos contiene datos de sumo interés con respecto a la contaminación atmosférica de Valencia, esto nos ha proporcionado una base sólida para el análisis y para toma de decisiones que se ha ido explicado paso a paso a la largo.

Podemos afirmar que analizando todos estos datos, hemos observado que a lo largo de los años se han registrado diferentes niveles de contaminantes, a veces incluso para algunos años no existen registros por lo tanto no aparecen en ninguno de nuestros mapas ni gráficos.

Centrándonos solo en los distritos los más antiguos presentan obviamente más datos, por lo tanto en las representaciones tienen valores más altos de niveles de contaminación, por lo normal.

En cuanto a las tendencias temporales que puede ser uno de los puntos también en los que más hemos basado nuestro trabajo observamos una variabilidad significativa en los niveles de contaminantes, aunque realmente no sabemos la tendencia de estas de manera exhaustiva esto

daría paso a estudios más profundos para poder mejorar la calidad o prever posibles daños futuros.

En general, nuestro proyecto ofrece una visión detallada y completa de la situación de la contaminación atmosférica en Valencia de manera generalizada en distritos dónde sin duda hay más estudios precisos, proporcionando información útil para la gestión ambiental y la toma de decisiones en la ciudad.

Enlace a página web: <https://dacasan5.shinyapps.io/proyectovd/>

Enlace video:  ProyectoVD.mp4