

Trabajo realizado por: Marcos Tudela, Hugo Lucas, Joan Olaso y Joan Rodriguez.

Ciclismo y Sostenibilidad Urbana

INFORME

Índice

1. Introducción
2. Metodología
 - 2.1 Análisis exploratorio de los datos
 - 2.2. Preprocesado de Datos
 - 2.3. Preprocesado de la información geográfica
 - 2.4. Elección de gráficas para los distintos tipos de datos
 - 2.5. Diseño del mapa e interactividad
 - 2.6. Diseño del cuadro de mandos
 - 2.7. Implementación
3. Resultados
4. Discusión
5. Conclusiones
6. Referencias

1. Introducción

A día de hoy vivimos en un mundo muy urbanizado, en el que nuestro principal recurso para desplazarnos son los vehículos motorizados. A pesar de ser muy útiles, entrañan diversos peligros, entre estos está la contaminación que producen y el riesgo hacia la vida humana que provocan en forma de accidentes si se utilizan de forma irresponsable. Una forma menos dañina de desplazamiento es ir en bici. A pesar de haber múltiples carriles bicis en Valencia, no es del todo seguro ir por zonas muy transitadas o con mucho tráfico, lo que puede ser un gran inconveniente para su uso.

Debido a esto hemos decidido realizar un estudio en el que se determine qué zonas de Valencia son las más y menos contaminadas, por qué y qué rutas de Valencia son las más seguras y más recomendables para los ciclistas de esta ciudad.

2. Metodología

2.1. Análisis exploratorio de los datos

Para realizar nuestro estudio, hemos utilizado los siguientes conjuntos de datos:

Un dataset con los datos de la calidad aire 2004-2022, y otro conjunto de datos con la localización de las estaciones de medición y los datos recabados por estas. Un conjunto de datos con la información de los distintos distritos de Valencia. Otro con los datos de la intensidad media diaria y mensual de vehículos motorizados. Un conjunto con los puntos de medida bicis espiras electromagnéticas. Y por último otro con la información del itinerario ciclista.

2.2. Preprocesado de Datos

Para poder mostrar la información de los conjuntos de datos de forma correcta en qgis, primero tuvimos que reproyectar cada una de las capas en formato EPSG:25830.

Más tarde cargamos todos los dataframes que vamos a utilizar en R y aplicamos algunas técnicas de limpieza y preparación de datos que incluyen la conversión de tipos de datos, eliminación de columnas innecesarias, ordenación de datos y la imputación de valores faltantes;

- Lectura de datos:

Los datos se leen desde un archivo CSV llamado "rvvcca.csv" utilizando la función `read_delim`. Se especifica que el delimitador es ";" y que la columna "PM1" se debe leer como valores numéricos.

- Conversión de tipos de datos:

Se convierten las columnas relevantes a tipo numérico utilizando `mutate(across(...))`. La columna "Fecha" se convierte a formato de fecha utilizando `as.Date()`.

- Eliminación de columnas innecesarias:

Se eliminan las columnas que no son necesarias para el análisis utilizando `select(-all_of(columnas_eliminar))`, donde `columnas_eliminar` es un vector que contiene los nombres de las columnas a eliminar.

- Ordenación de datos:

Los datos se ordenan por estación y fecha utilizando `arrange(Estacion, Fecha)` para asegurar que los valores imputados provengan de la misma estación en las fechas más cercanas posibles.

- Imputación de valores faltantes:

Se define una función llamada `impute_with_closest_date` que realiza la imputación de valores faltantes utilizando el método LOCF/NOCB (último valor observado hacia adelante/atrás). La función itera sobre las columnas especificadas en `variables_imputar` y, para cada columna, encuentra el valor no faltante más cercano hacia adelante y hacia atrás en el tiempo. Luego, selecciona el valor más cercano para imputar el valor faltante. Se crean columnas auxiliares para calcular la diferencia en días entre las fechas de los valores faltantes y los valores no faltantes más cercanos. Finalmente, se eliminan las columnas auxiliares creadas durante el proceso de imputación.

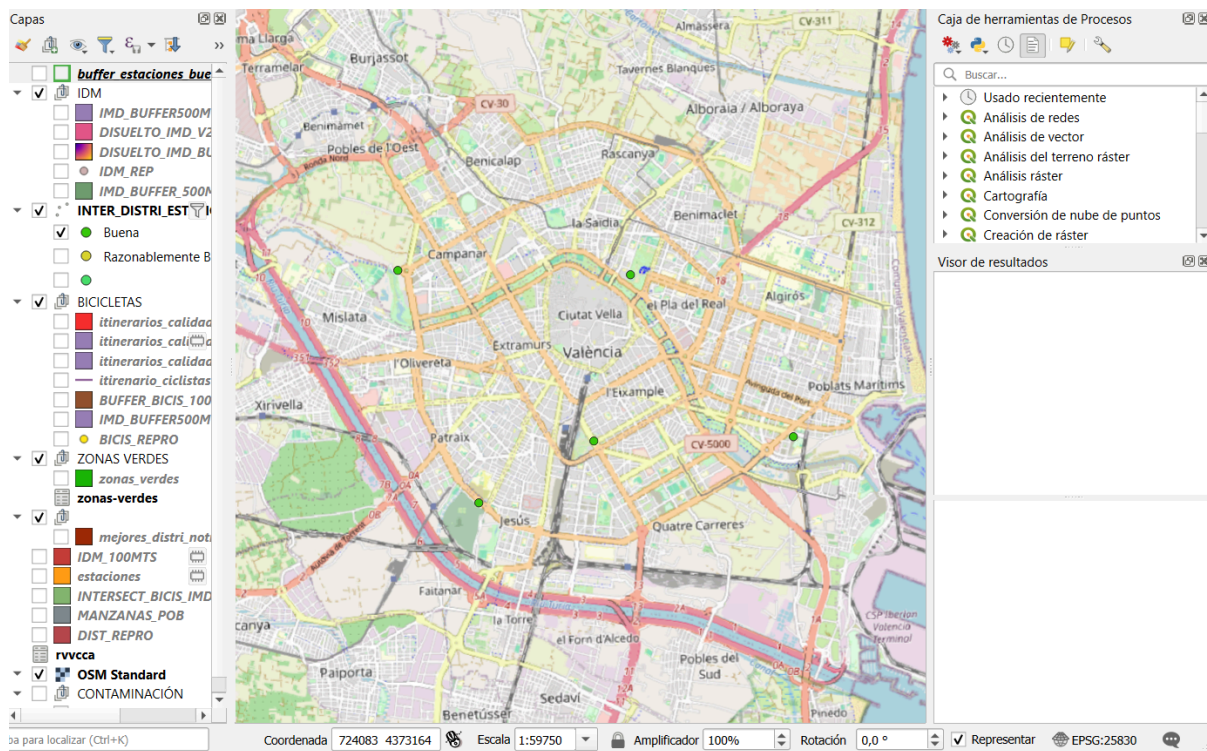
- Aplicación de la función de imputación:

Se aplica la función `impute_with_closest_date` a los datos utilizando `datos`.

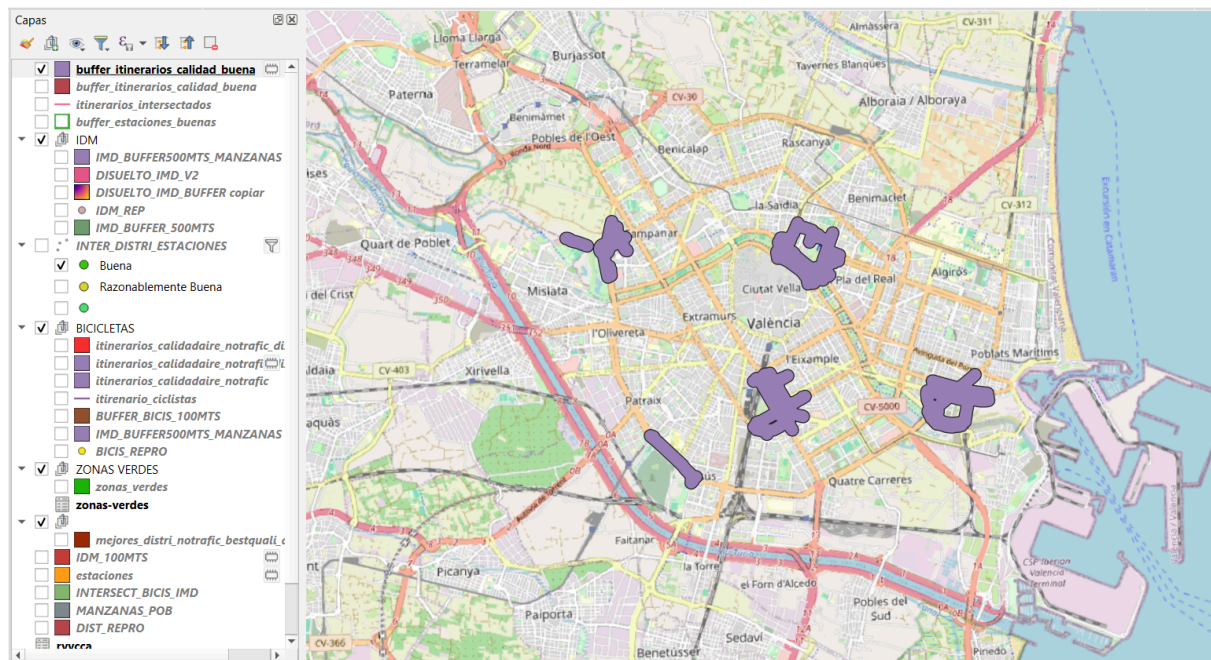
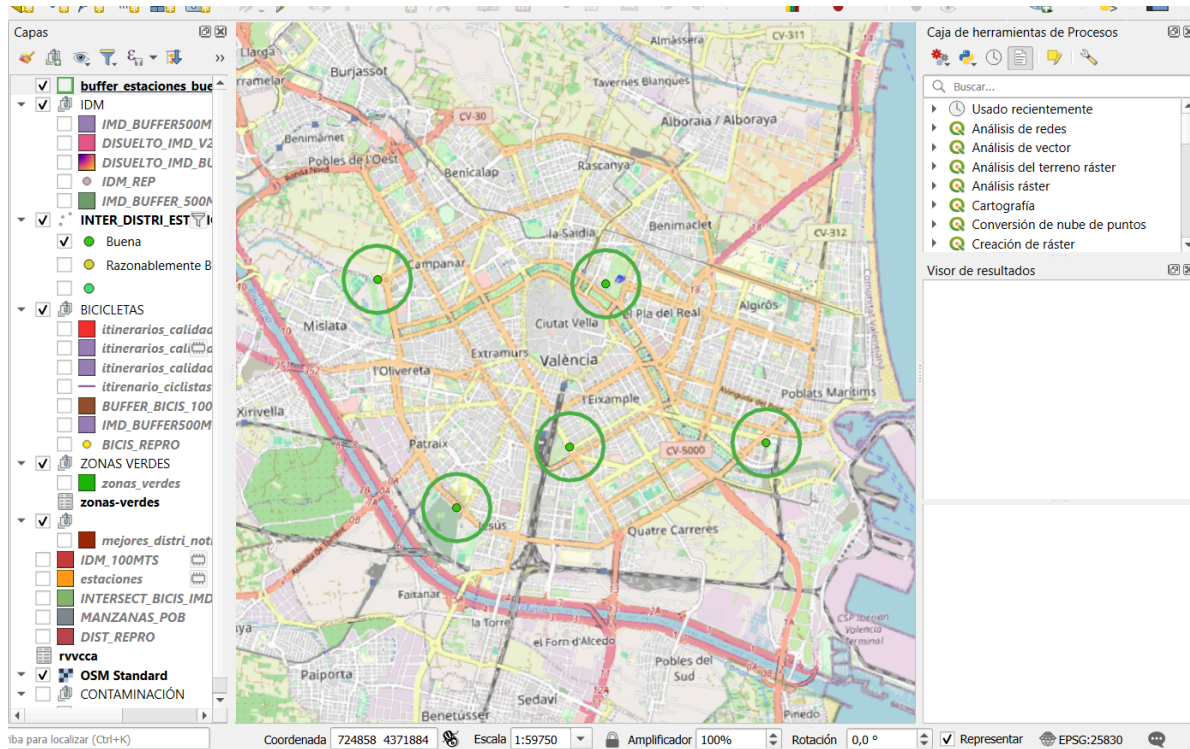
2.3. Preprocesado de la información geográfica

Para la primera vista obtenemos como resultado los mejores distritos de Valencia para ir en bicicleta, según nuestro estudio. Para ello hemos tenido en cuenta diferentes factores y variables como la calidad del aire, medida por las diferentes estaciones de medición de contaminación repartidas por distintos puntos de Valencia. Junto con esto, también hemos tenido en cuenta la intensidad diaria de tráfico motorizado, evitando aquellas zonas y distritos que presentan unos valores altos en este campo. Buscando a su vez también y premiando a la par, la proximidad a las zonas verdes. También como se ha mencionado anteriormente hemos tenido en cuenta el itinerario ciclista de bicicletas en Valencia.

Lo primero que hemos hecho ha sido mostrar la variable categórica *calidad_ambiental* la cual encontramos en los datos recogidos en las estaciones de contaminación meteorológicas, esta variable mostraba únicamente dos tipos de valores diferentes, Buena o Razonablemente Buena. Para buscar lo mejor para los ciclistas decidimos quedarnos únicamente con aquellas estaciones las cuales recogen una calidad de aire Buena y no Razonablemente Buena.

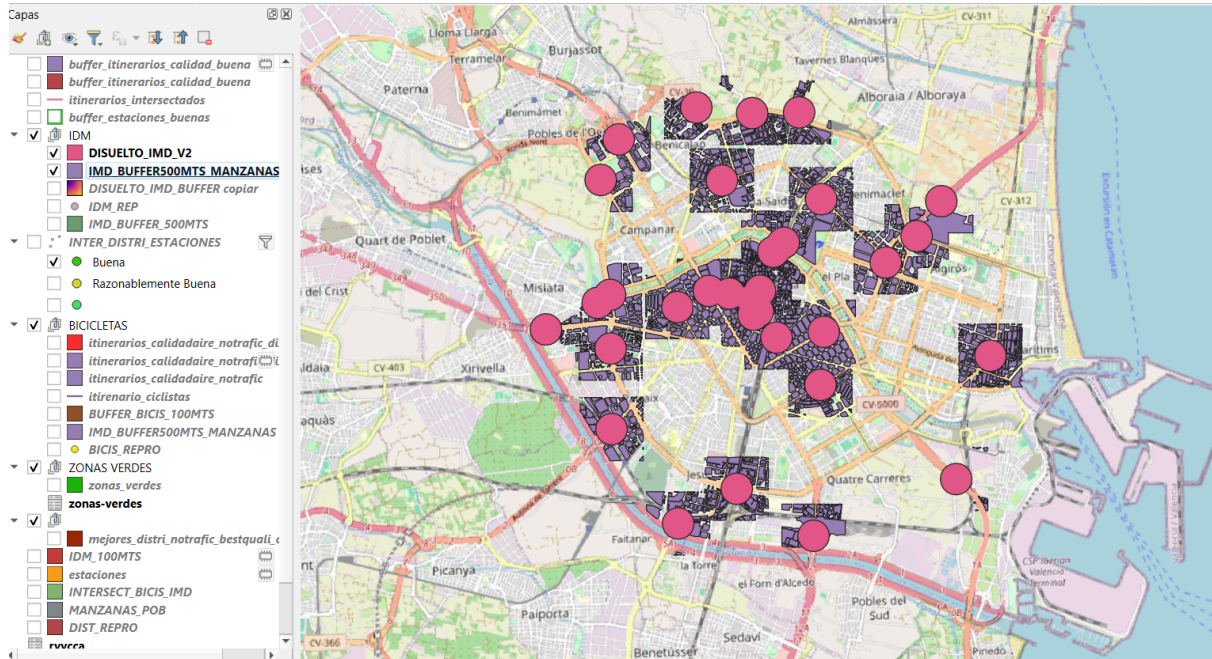


Una vez hecho esto hacemos un buffer para aquellos puntos para los cuales la calidad ambiental sea Buena y lo intersectamos con la capa correspondiente al itinerario que siguen los ciclistas, de esta forma ya tenemos únicamente aquellos itinerarios para los cuales la calidad del aire es Buena.

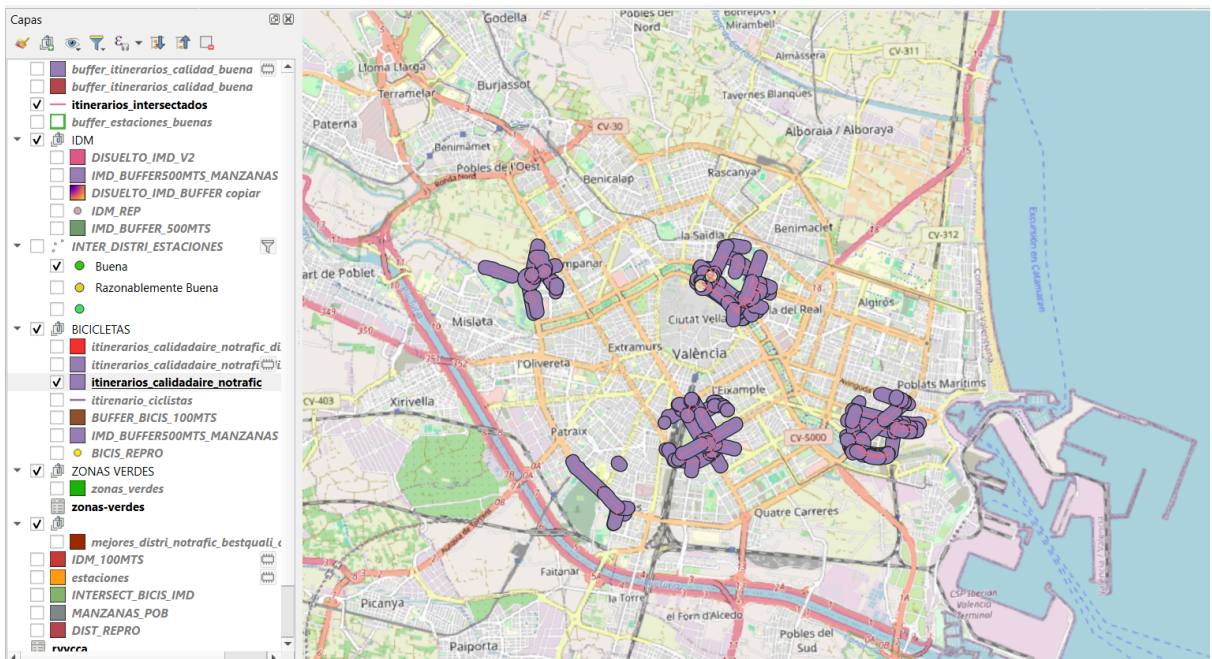


Teniendo en cuenta esta nueva capa mencionada, lo que hacemos a continuación es un nuevo buffer para los datos de intensidad diaria de tráfico en vehículos motorizados. Observamos cómo están repartidos dichos puntos a lo largo de Valencia, nos fijamos que en la zona céntrica de Valencia es donde mayor tráfico hay y donde la calidad del aire baja su

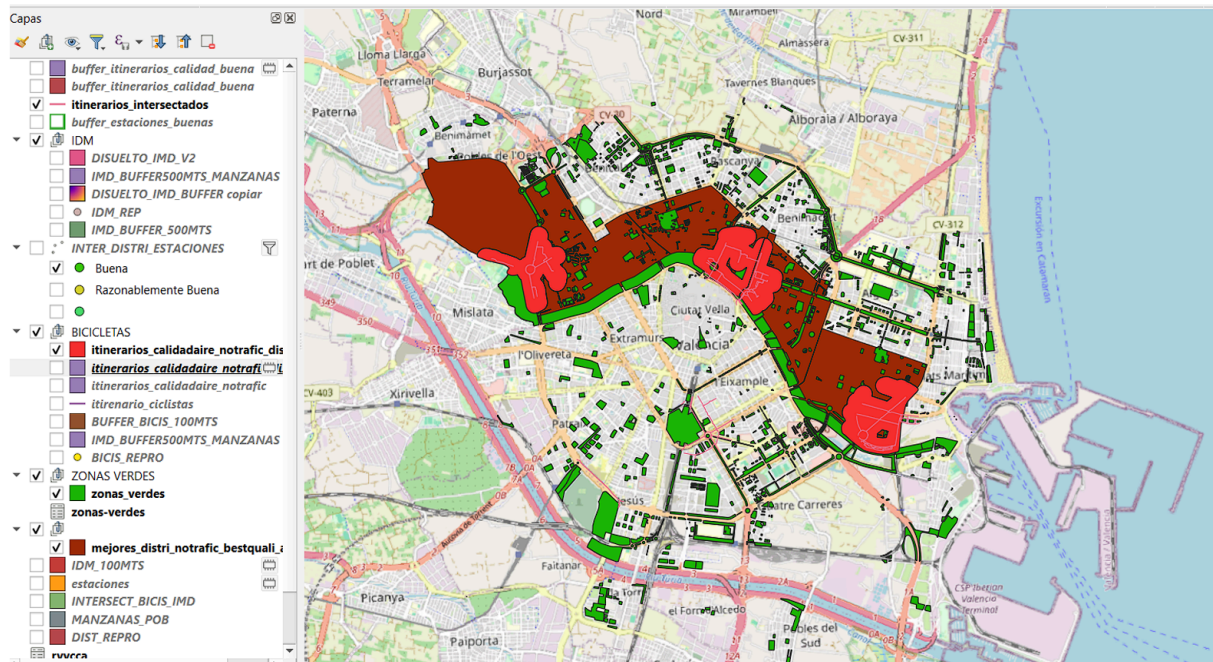
calidad, sirviéndonos esta información para descartar estos distritos céntricos y del sur, como no candidatos para ser los idóneos para ir en bicicleta.



A continuación intersectamos nuestros datos de intensidad diaria de tráfico a nuestra capa la cual contiene los itinerarios de los ciclistas para los cuales la calidad del aire es buena, de esta forma evitamos aquellas zonas con mucho tráfico por los problemas que esto supone para los ciclistas ya que nuestro objetivo es encontrar los mejores distritos de Valencia para ir en bicicleta.



Finalmente de estos distritos resultantes ,los cuales presentan una mejor calidad de aire y un menor tráfico, nos quedamos con aquellos que tengan una mayor densidad de zonas verdes a su alrededor esto lo hacemos mediante una calculadora de campos, dividiendo la superficie total de zonas verdes en cada distrito por la superficie total del distrito.Concluimos que los mejores distritos para ir en bicicleta son : LA SAIDIA, CAMINS AL GRAU, CAMPANAR y finalmente EL PLA DEL REAL.

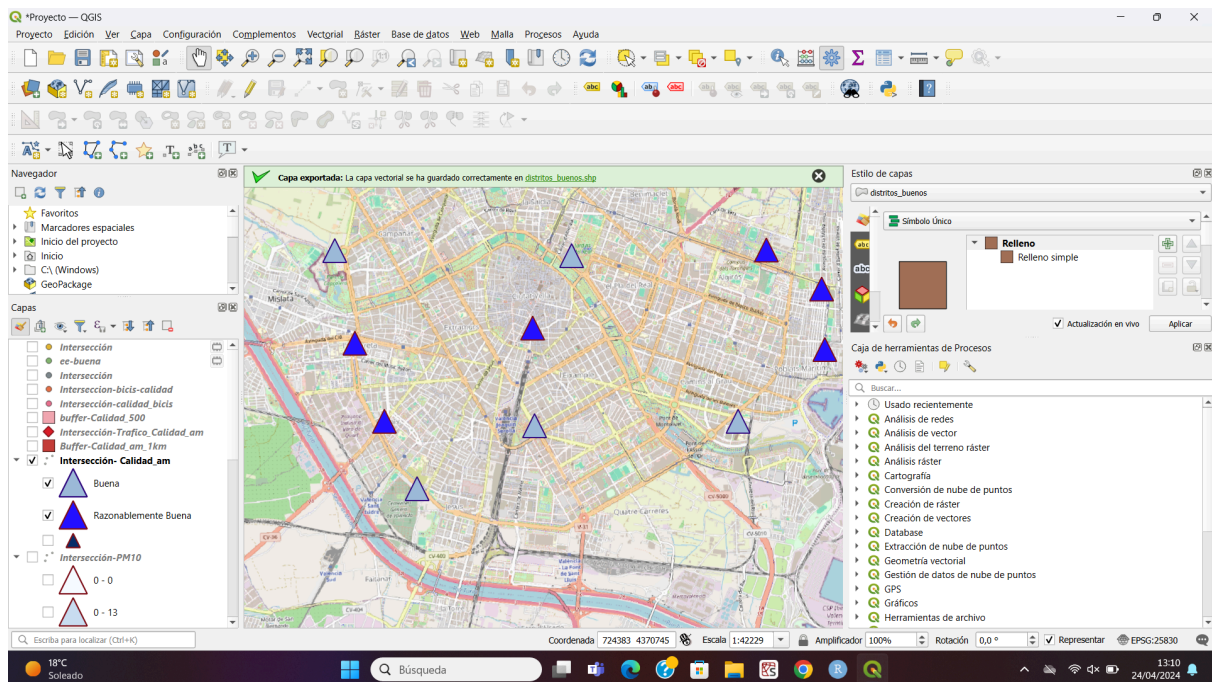


Los distritos mencionados son aquellos para los cuales la calidad del aire , la intensidad del tráfico y la densidad de zonas verdes encuentran un equilibrio favorable , cosa que tiene sentido, además teniendo en cuenta que todos estos distritos están en frontera con el río Turia. que es la zona verde más extensa.

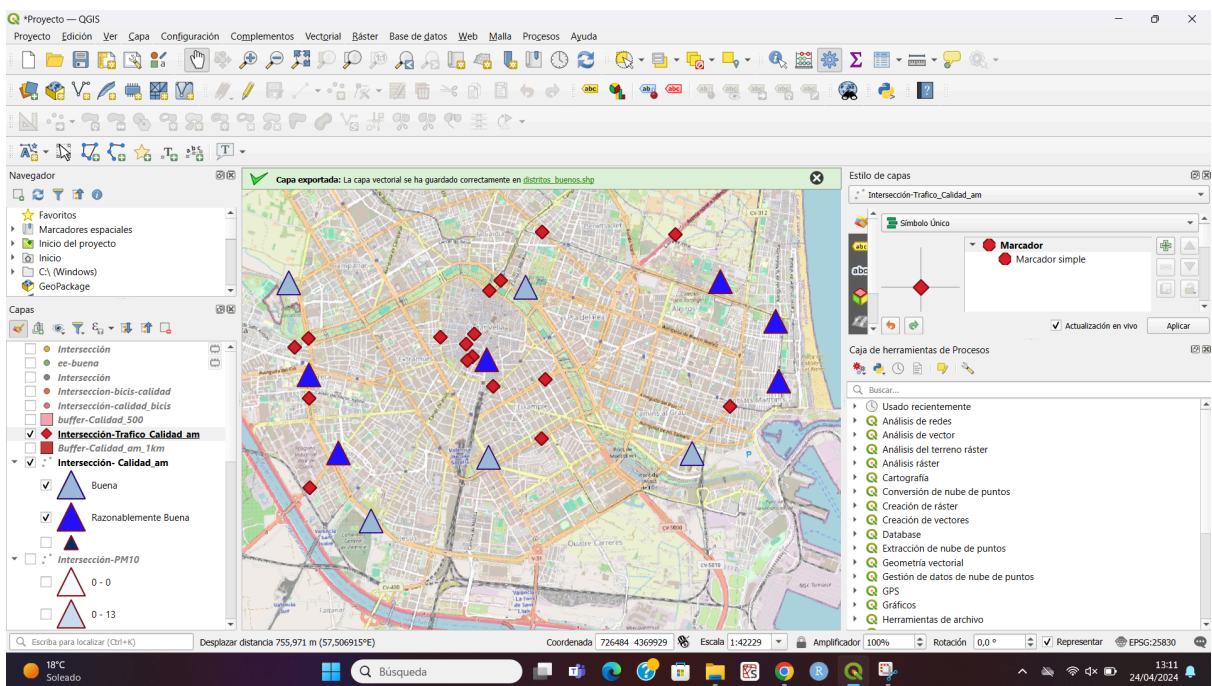
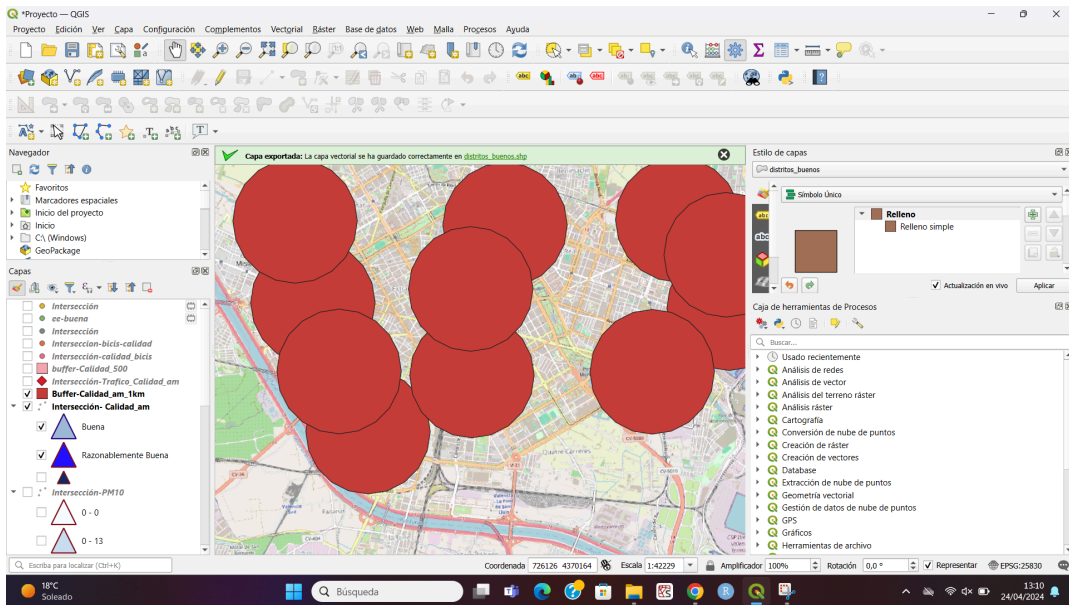
Para la segunda vista visualizamos los distritos con menor contaminación ambiental en Valencia y su relación con el tráfico y el uso de las bicicletas. Para realizar el siguiente estudio nos hemos basado en los datos recogidos por las estaciones de medición de contaminación.

Para realizar este trabajo, hemos realizado diversas operaciones.

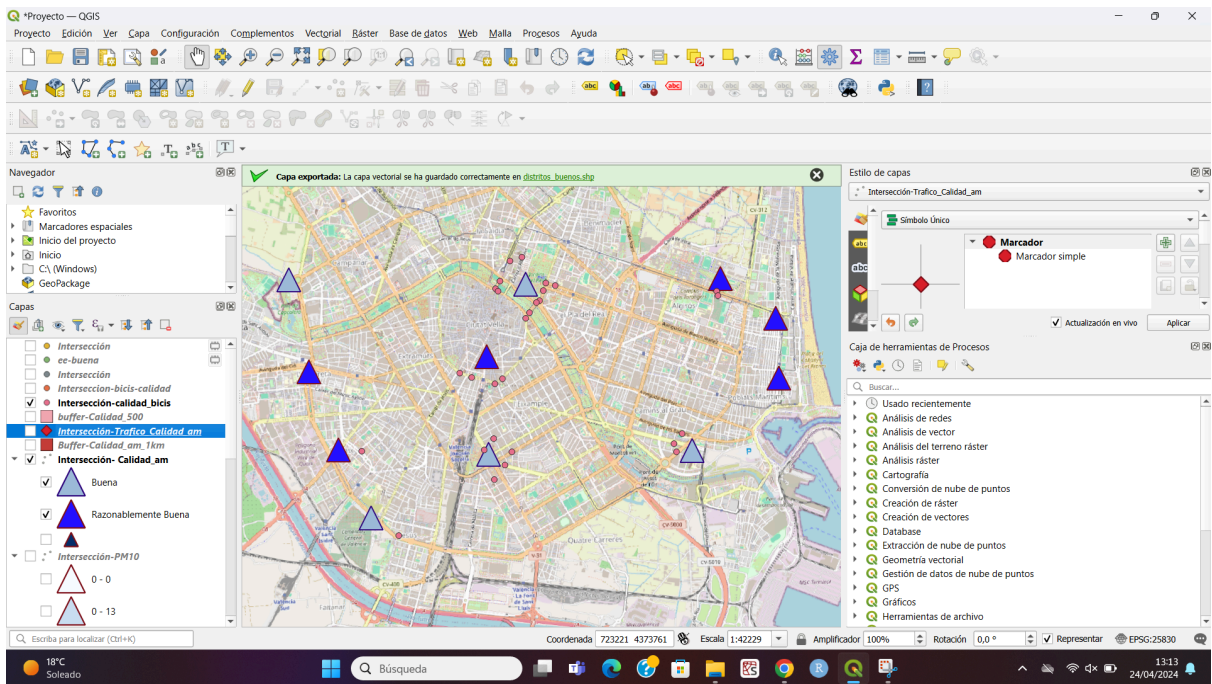
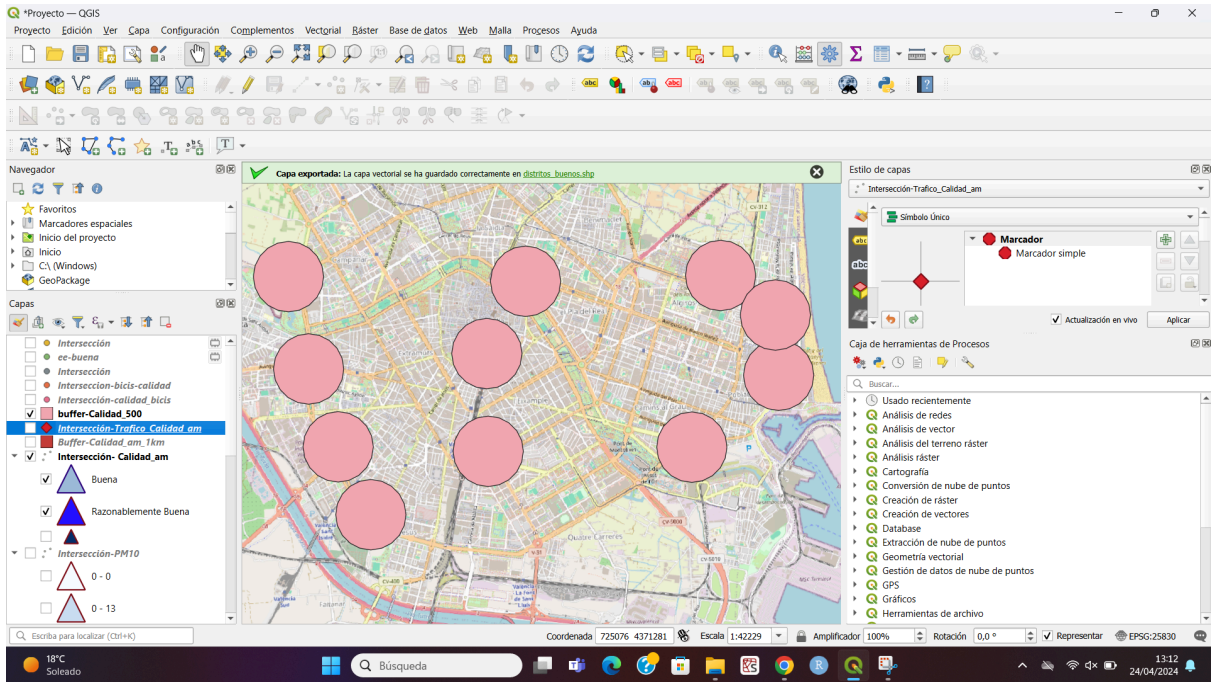
Primero hemos intersectado la capa con la información de las estaciones de medida de contaminación, con la capa de distritos, de esta forma relacionamos cada estación con un distrito. Después entrando en el apartado de simbología, podemos categorizar dichas estaciones, según la variable de la calidad ambiental. De esta forma se ve representado en el mapa cada estación de contaminación con un color distinto dependiendo de su calidad del aire. En nuestro caso cuando descargamos los datos la calidad ambiental sólo era buena o razonablemente buena. Por tanto en nuestro mapa las estaciones estarán representadas por azul clarito las que tienen una calidad buena y en azul oscuro las que tienen una calidad razonablemente buena.



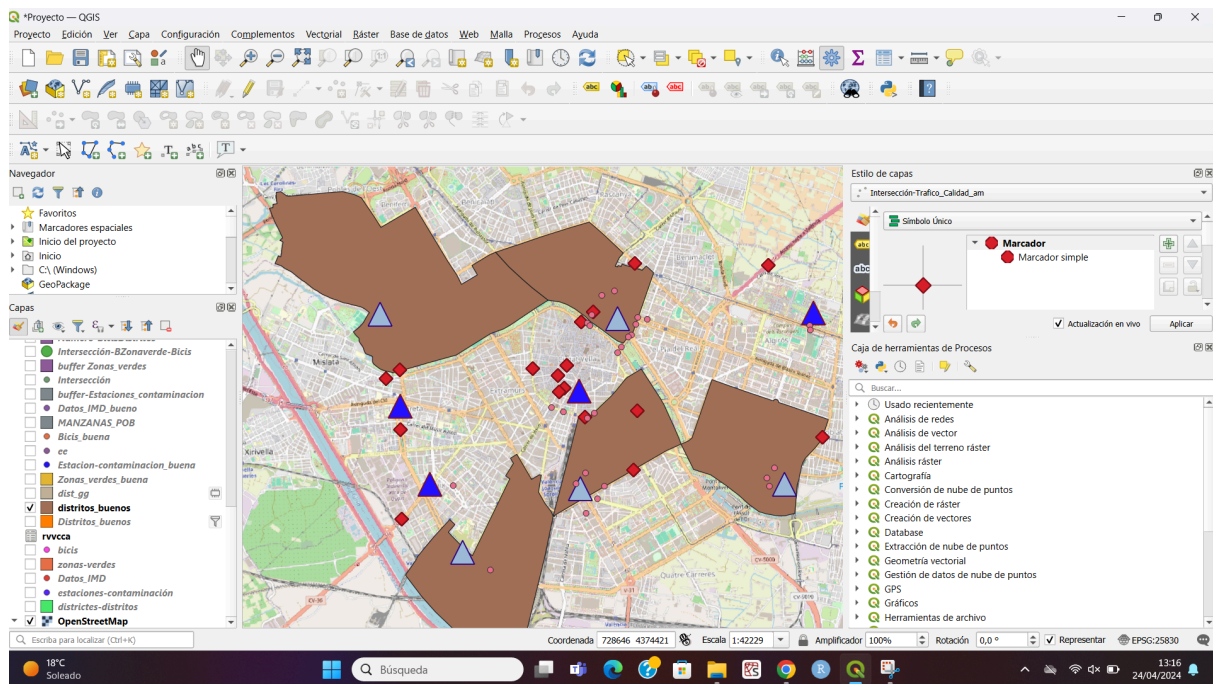
Después hemos creado un buffer en esta intersección con una distancia de 1 kilómetro, para poder determinar la densidad de tráfico alrededor de estas estaciones. Una vez creado el buffer intersectamos la capa de la intensidad diaria de tráfico, teniendo así los puntos con más intensidad de tráfico cercanos a las estaciones.



De nuevo realizamos los mismo pasos que antes pero cambiando la distancia del buffer de 1 kilómetro a 500 metros e intersectando el buffer por la capa de las bicis, consiguiendo los puntos de bicis más cercanos.



Por último seleccionamos los distritos con mejor calidad ambiental:



2.4. Elección de gráficas para los distintos tipos de datos

El primer gráfico que presentamos es una representación de barras que muestra la cantidad de un contaminante presente en los diversos distritos. Los gráficos de barras son herramientas visuales sumamente efectivas para ilustrar datos categóricos y discretos de manera clara y precisa. Optamos por este tipo de representación debido a su capacidad para facilitar la comparación de la cantidad de dicho contaminante entre las diferentes zonas de Valencia

El segundo gráfico que presentamos es una representación de líneas que muestra la evolución de un contaminante a lo largo del tiempo. Optamos por este tipo de gráfico debido a la naturaleza continua de la representación, ya que estamos analizando el factor temporal. Además, este formato nos facilita la identificación de los períodos con mayor concentración de contaminantes, gracias a la intensidad visual de las líneas, lo que permite una interpretación rápida y precisa de los datos.

2.5. Diseño del mapa e interactividad

En el primer mapa se leen 4 capas;

La capa "amb.shp" representa los puntos donde se encuentran las estaciones que recogen datos sobre la contaminación, la capa "coches.shp" representa los puntos donde mayor intensidad media diaria de tráfico hay que están a un km o menos de las estaciones de contaminación, la capa "bicis.shp" representa los puntos donde hay puntos de acceso a bicicletas que están a 500 metros o menos de una estación de contaminación y la capa "distritos_hugo.shp" representa los distritos con mejor calidad ambiental.

Para la cartografía base se usa la cartografía de OpenStreetMap. Se agregan polígonos de los distritos de la zona, que se muestran en color naranja y se proporciona un popup que muestra el nombre del distrito y la media del contaminante seleccionado. Se agregan marcadores para las estaciones de calidad ambiental, puntos IMD y bicicletas, cada uno tiene un popup que muestra el nombre de la estación/punto IMD. También hemos representado los puntos de cada capa con iconos diferentes para que se puedan diferenciar bien. Se establece la vista del mapa en una ubicación específica (longitud -0.375, latitud 39.4667) con un nivel de zoom de 12. Se agrega una leyenda en la esquina inferior derecha que describe los colores y etiquetas utilizados en el mapa.

En el segundo mapa se cargan 4 capas;

La capa "zonas-verdes.shp" representa las zonas verdes presentes en toda Valencia, la capa "mejores_distri_notrafic_bestquali_air.shp" representa los distritos con mejor calidad de aire y de tráfico, además de ser altamente densos en zonas verdes; es decir nuestros distritos ideales para hacer ciclismo en Valencia.

La capa "itinerarios_intersectados.shp" representa los carriles de bici ideales en relación al tráfico y la calidad ambiental de la zona y la capa "INTER_DISTRI_ESTACIONES.shp" representa aquellas estaciones las cuales presentan una calidad ambiental "Buena"..

Para la cartografía base se usa la cartografía de OpenStreetMap. Se agregan polígonos para los itinerarios ideales, distritos ideales y zonas verdes, cada capa de polígono tiene un color diferente (azul para itinerarios, verde para distritos, amarillo para zonas verdes) y un popup que muestra el nombre de la zona y la media del contaminante seleccionado. Se agregan líneas de color rojo que representan itinerarios correctos. Se agregan círculos para representar las estaciones de calidad ambiental, Los círculos son de color púrpura y tienen un radio de 50 metros. Se establece la vista del mapa en una ubicación específica (longitud -0.375, latitud 39.4667) con un nivel de zoom de 12. Se agrega una leyenda en la esquina inferior derecha que describe los colores y etiquetas utilizados en el mapa.

2.6. Diseño del cuadro de mandos

El cuadro de mandos está diseñado por diferentes elementos que permiten a los usuarios interactuar con los datos ambientales seleccionando contaminantes, cargando archivos CSV y visualizando mapas con sus respectivos gráficos relacionados. Estos elementos son;

- Panel de Encabezado (header-panel):

Este panel contiene el logotipo y el título de la aplicación. El logotipo se define como una imagen (`imageOutput("logolmage")`) y el título se define como un encabezado (`tags$h1()`).

- Panel Lateral (sidebarPanel):

Este panel se encuentra a la izquierda de la aplicación y generalmente contiene controles de entrada para que el usuario interactúe con los datos. En este caso, contiene un menú desplegable (`selectInput`) que permite al usuario seleccionar un tipo de contaminante y un botón para cargar un archivo CSV (`fileInput`).

- Panel Principal (mainPanel):

Este panel ocupa el espacio central de la aplicación y se utiliza para mostrar los resultados de las interacciones del usuario. Está dividido en pestañas (`tabsetPanel`), cada una de las cuales contiene un tipo diferente de visualización de datos.

Las pestañas incluyen:

"Mapa 1": Muestra un mapa interactivo (`leafletOutput`) con datos ambientales relacionados con el tráfico de coches y bicicletas.

"Mapa 2": Muestra otro mapa interactivo con información sobre los mejores distritos para andar en bicicleta.

"Gráfico 1": Muestra un gráfico de líneas (`plotOutput`) que representa el nivel de los diferentes contaminantes para todos los distritos.

"Gráfico 2": Muestra un gráfico (`plotOutput`) que representa la evolución en el tiempo de los diferentes contaminantes.

- Pie de Página (footer):

Este panel se encuentra en la parte inferior de la aplicación y generalmente contiene información adicional, como créditos o enlaces.

En este caso, muestra el texto "Alumnos Segundo ciencia de datos".

2.7. Implementación

Detallar la implementación que se ha realizado en R.

El código es una aplicación Shiny que proporciona visualizaciones interactivas de datos de contaminación atmosférica. Aquí dejamos una descripción detallada de la implementación:

- Cargamos librerías:

Cargamos las librerías necesarias para la aplicación, como shiny, leaflet, sf, raster, ggplot2, readr, tidyverse, plotly, y zoo.

- Carga y lectura de datos:

Los datos se leen desde un archivo CSV (rvvcca.csv) y se limpian para su procesamiento posterior. Las columnas relevantes se convierten a tipo numérico y la fecha se convierte a tipo Date.

- Eliminamos columnas innecesarias:

Eliminamos columnas del conjunto de datos que no se necesitan para el análisis posterior.

- Imputación de datos faltantes:

Definimos una función llamada `impute_with_closest_date` que realiza la imputación de datos faltantes utilizando el método LOCF/NOCB (Last Observation Carried Forward/Next Observation Carried Backward).

- Interfaz de usuario:

Se define la interfaz de usuario utilizando el framework Shiny. Esto incluye la disposición de los elementos de la página, como paneles laterales, paneles principales y pestañas, así como la definición de entradas como selectores de contaminantes y carga de archivos CSV.

- Servidor:

Se define la lógica del servidor que procesa las entradas del usuario, carga los datos, realiza cálculos y genera visualizaciones. Esto incluye la lectura de datos, cálculo de medias de contaminantes, definición de iconos para el mapa, carga de datos espaciales, y la generación de mapas y gráficos.

- Mapas:

Se generan dos mapas interactivos utilizando la librería Leaflet. El primer mapa muestra los distritos y ubicaciones de estaciones de calidad ambiental, puntos IMD (Intensidad media diaria) y bicicletas, mientras que el segundo mapa muestra los distritos ideales para ir en bicicleta.

- Gráficos:

Se generan dos gráficos interactivos utilizando ggplot2. El primer gráfico muestra el nivel de contaminante seleccionado en cada estación, y el segundo gráfico muestra la evolución temporal de un contaminante seleccionado.

- Mensaje de carga:

Se muestra un mensaje modal cuando se carga correctamente un archivo CSV.

- Renderizado del logo:

Se muestra un logo en la parte superior izquierda de la aplicación.

3. Resultados

Como ya hemos explicado, los resultados de nuestro trabajo son principalmente unos mapas realizados en QGIS y unos gráficos realizados en R.

En QGIS creamos las visualizaciones que queremos representar sobre un mapa, por ejemplo, resaltar los distritos con mejor calidad ambiental, mostrar la intensidad de tráfico que hay en estos o el itinerarios de bicicletas que hay en valencia, etc.

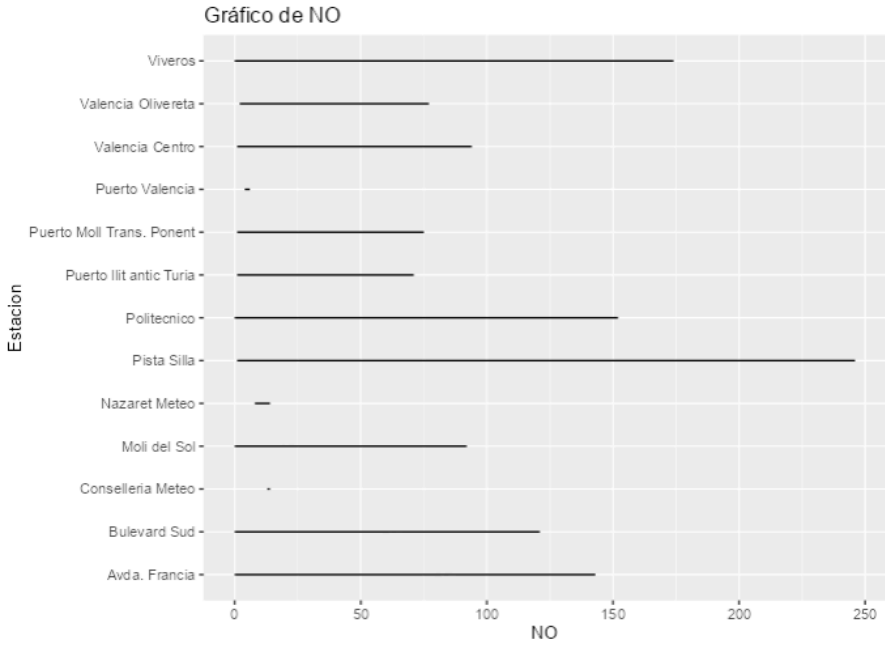
Por otro lado desarrollamos gráficos, para mostrar los estadísticos de los datos que nos encontramos en los distintos dataframes. Estos estadísticos muestran o bien la evolución de un contaminante con el paso de los años o la cantidad que ha habido de este en los diferentes distritos.

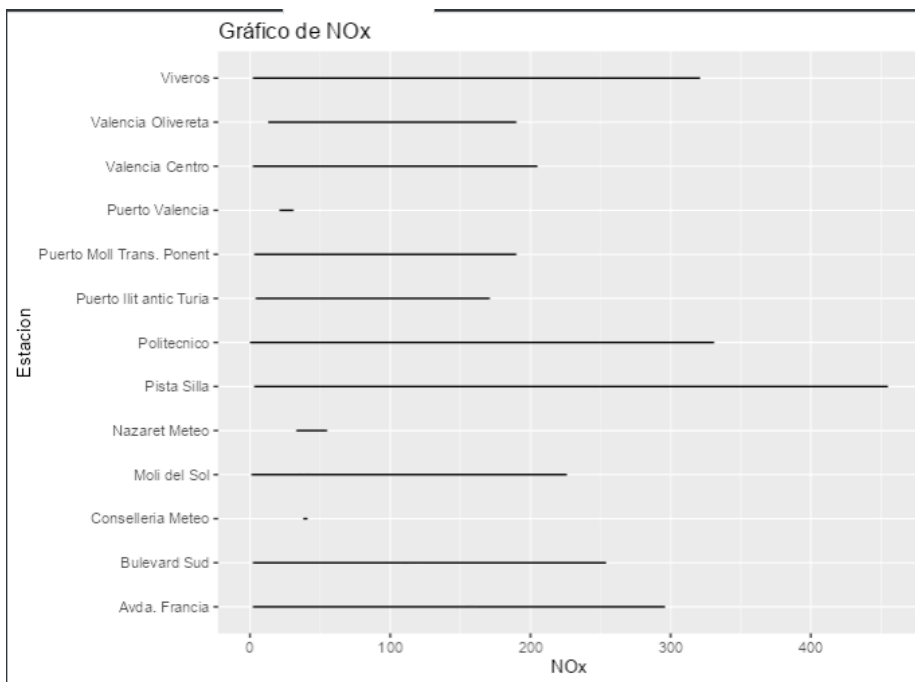
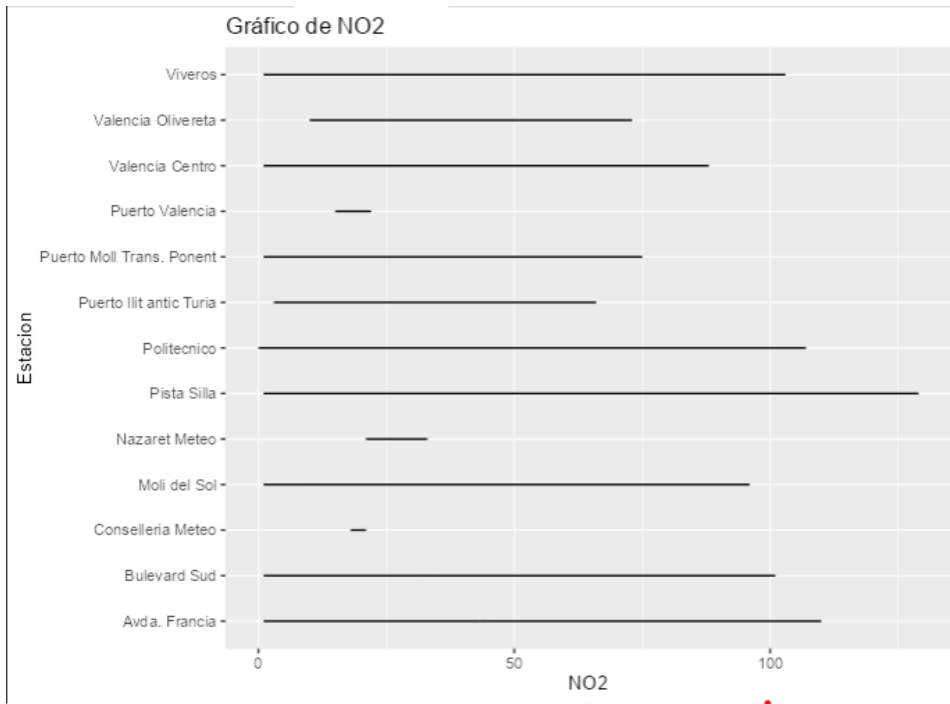
Hay que añadir que los gráficos son interactivos, ya que tanto en el primer gráfico como en el segundo se puede elegir el contaminante que se desea representar y se puede observar información adicional en cualquier punto del gráfico que deseese.

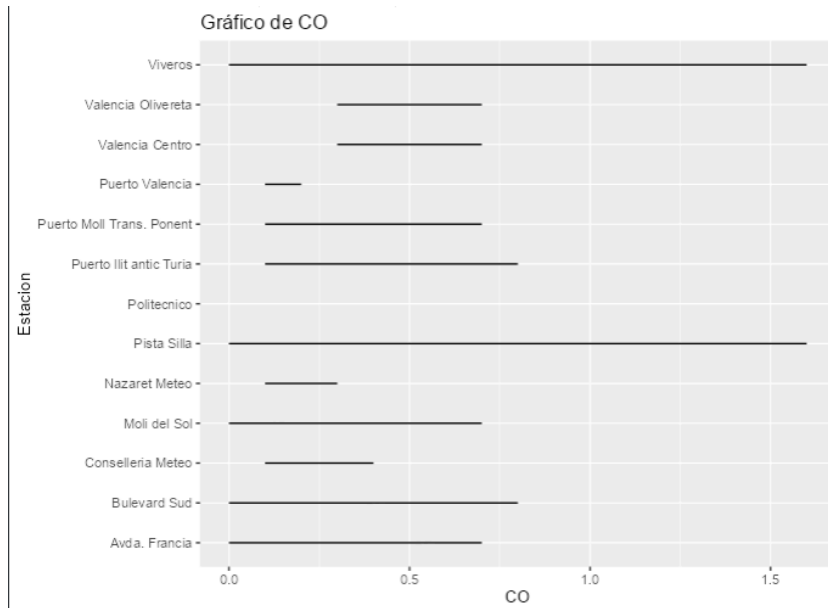
El primero de los mapas hace referencia a los distritos con mejor calidad ambiental y examina la posible relación entre la contaminación ambiental y el tráfico. En dicho mapa podemos observar claramente que varios distritos están seleccionados, estos son La Saida, Camins al Grau, Campanar, L'eixample y Jesús, todos los cuales se destacan por su calidad ambiental favorable. Es visible también que en la mayoría de estos distritos no nos encontramos con una gran densidad de tráfico y los que sí tiene una intensidad mayor, están claramente compensados con varios puntos de bicis. Por otro lado, los distritos que no están destacados, es decir, aquellos que tienen una calidad ambiental más desfavorable, sí que tienden a tener una gran intensidad de tráfico y pocos puntos de recogida de bicis.

El segundo mapa muestra los mejores distritos para ir en bicicleta, junto con la ubicación de los carriles bici disponibles en estos. Podemos observar como los distritos destacados son La Saida, Camins al Grau, Campanar y Pla del Real, todos estos presentan una calidad ambiental buena, por lo que son perfectos para actividades como el uso de bicicletas. También están representadas todas las zonas verdes de estos distritos, estas complementan la experiencia del ciclista, puesto que aportan paz y naturaleza.

En el primer gráfico, podemos observar claramente cuales son los distritos más contaminados.

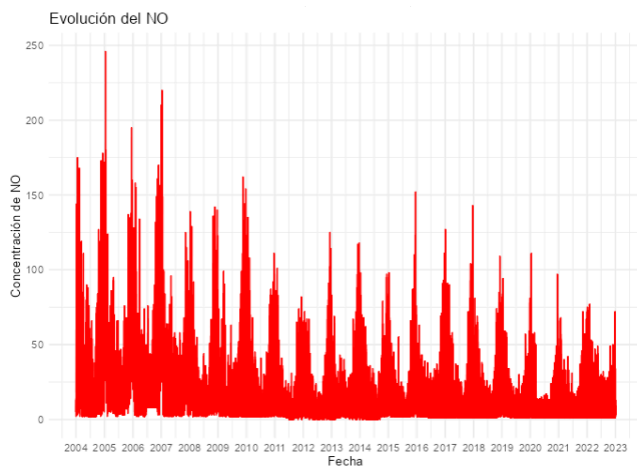


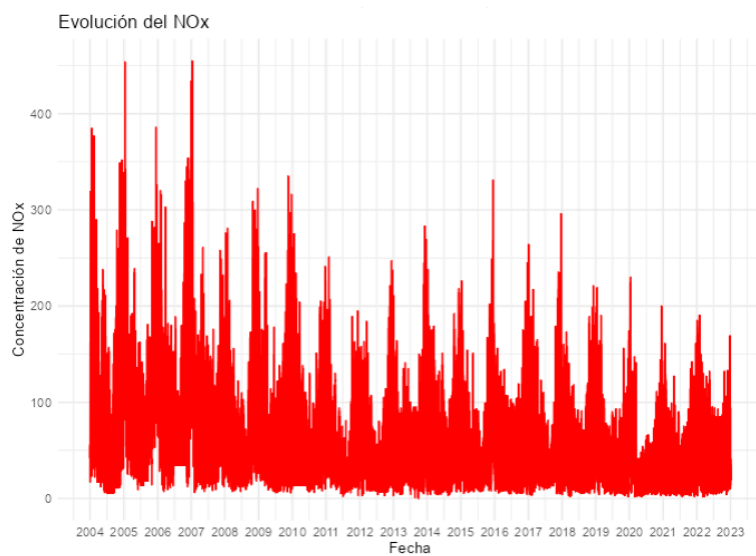
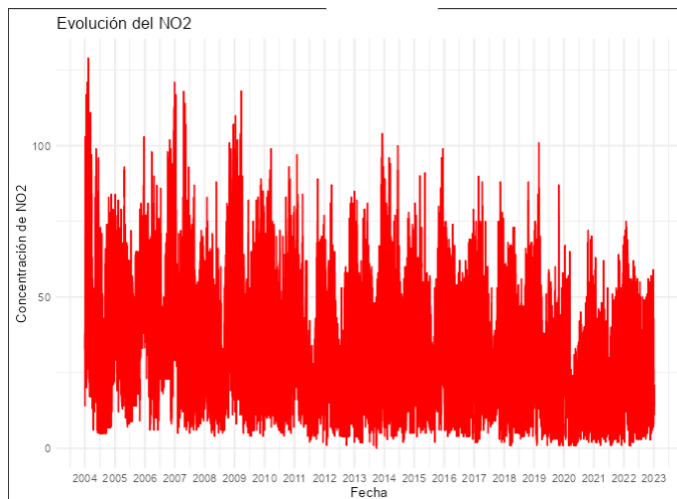




Los distritos más contaminados en todas las gráficas son Viveros y Pista silla.

Y en el segundo gráfico podemos visualizar en qué años ha habido más presencia de gases contaminantes y en qué cantidad.





Gracias a estos gráficos podemos asumir que ha habido una disminución de la contaminación en Valencia. Esto puede ser debido a varios factores.

La pandemia de 2020 pudo haber desempeñado un papel importante al reducir el tráfico y otras fuentes de contaminación, lo que probablemente contribuyó a mejorar la calidad ambiental en la ciudad.

Además, en los últimos años, Valencia ha implementado diversas campañas contra la contaminación, promoviendo el uso del transporte público y aspirando a convertirse en la capital verde de Europa

4. Discusión

Uno de nuestros principales objetivos al realizar este trabajo es averiguar si existe una relación directa entre la intensidad de tráfico que existe en nuestra ciudad y la calidad ambiental. Como hemos podido corroborar en el primer mapa, Los hallazgos de este mapa apoyan la existencia de esta relación, ya que los distritos con una calidad ambiental superior suelen caracterizarse por una menor concentración de tráfico vehicular

Nuestro análisis, también resalta el impacto positivo del uso de la bicicleta en lo referente a la calidad ambiental. Los distritos que destacan como ideales para el ciclismo, como La Saida, Camins al Grau, Campanar y Pla del Real, no solo presentan una calidad ambiental superior, sino que también están diseñados para fomentar el uso de la bicicleta como medio de transporte alternativo. Por tanto se debería considerar más a fondo el fomentar e implementar más puntos de recogida de bicicletas, con tal de disminuir la contaminación atmosférica.

5. Conclusiones

Los resultados de nuestro estudio muestran una clara relación entre la intensidad del tráfico y la calidad ambiental en nuestra ciudad. Mediante la visualización de datos en mapas y gráficos, hemos identificado distritos con una calidad ambiental superior, como La Saida, Camins al Grau, Campanar y Pla del Real, que también se destacan por su infraestructura favorable para el ciclismo y la presencia de zonas verdes.

Nuestros hallazgos respaldan los beneficios del transporte sostenible y la importancia de considerar la calidad ambiental en la planificación urbana. Sin embargo, reconocemos algunas limitaciones en nuestro estudio, como la falta de datos detallados sobre otras fuentes de contaminación y la existencia de muchos datos faltantes dentro de los dataframes.

En conclusión, nuestros resultados subrayan la necesidad de políticas urbanas que promuevan un transporte más sostenible y reduzcan la dependencia del automóvil. Además, resaltan la importancia de integrar consideraciones ambientales en la planificación urbana para crear ciudades más saludables y habitables para todos.

6. Referencias

[Ayuntamiento de Valencia - Portal de Datos Abiertos](#)

<https://programminghistorian.org/es/lecciones/creacion-de-aplicacion-shiny>

[OpenStreetMap](#)

https://diegokoz.github.io/intro_ds/clase_6/06_explicacion.nb.html

[QGIS.org](#)

[Stackoverflow](#)

[Rstudio](#)