VDS2425 Project Report - Design Phase

Group Number: 6
Members:
Yara Roelen (2157769)
Saleh Abednezhad (2469739)
Kiflom Berhe (2470560)
Dataset: Madrid

April 2025

Metadata

Design phase version 1
Dataset: Madrid

Project Description

Description of the Dataset

In recent years, high levels of pollution during certain dry periods in Madrid have forced authorities to take measures against the use of cars in the city center and propose drastic modifications in the city's urbanism. We have a file with information about the stations (id, name, address, longitude, latitude, and elevation) and measurement files for each year from 2001 to 2018. Conclusions will be drawn from this data regarding actions the government can take to lower pollution in the city.

Description of the Features

The dataset includes station information (id, name, address, longitude, latitude, elevation) and yearly measurements of pollutants such as SO_2, CO, NO_2, PM_{10}, $PM_{2.5}$, O_3, and various volatile organic compounds (VOCs) like BEN, TOL, and NMHC. Measurements are provided in $\mu g/m^3$ or mg/m^3, with timestamps linking them to specific stations.

The Air Quality Index (AQI) combines key pollutants into one score using this formula:

$$AQI = 0.4 \times PM_{2.5} + 0.3 \times NO_2 + 0.2 \times PM_{10} + 0.05 \times O_3 + 0.05 \times SO_2 \quad (1)$$

based on WHO health impact guidelines. This allows for temporal and spatial analysis of pollution levels across Madrid.

Our 3 Questions for This Project

- Saleh: Question 1 – How has pollution in Madrid evolved between 2001 and 2018?
- Yara: Question 2 – Which are the areas where the pollution is the highest/lowest?
- Kiflom: Question 3 – Which areas of Madrid improved or worsened more between 2008 and 2018?

Design

From Question to Task – Question 1

The given question is: How has pollution in Madrid evolved between 2001 and 2018?

To address this question, it is divided into specific tasks to analyze the temporal evolution of pollution levels across Madrid over the specified period, using the available dataset of yearly pollution measurements and station metadata.

Task 1 – Collect and Clean the Dataset

At first, the pollution data from 2001 to 2018 is imported and prepared to analyze how it changed over time.

- **Action:** Load the yearly pollution data files (2001–2018) and the station metadata file, merging them based on station ID and timestamps.
- **Target:** A unified dataset covering all stations and pollutants from 2001 to 2018.
- **Object:** Pollution measurements and station metadata.
- **Measures:** Concentrations of key pollutants (NO_2, SO_2, PM10, PM2.5, O_3, CO) in $\mu g/m^3$ or mg/m^3, and optionally VOCs (e.g., BEN, TOL) where available.
- **Groupings:** Year, Station ID.

Task 2 – Calculate Average Pollution Levels per Year

Now, the city-wide average concentration of key pollutants across all stations (e.g., NO_2, PM10, O_3) for each year will be calculated, to see how levels changed over time.

- **Action:** Aggregate data by calculating the mean concentration per pollutant for each year.
- **Target:** Yearly pollution trends across Madrid.
- **Object:** Pollutant concentration levels.
- **Measures:** Average of each pollutant in $\mu g/m^3$ or mg/m^3 per year.
- **Groupings:** Year, Pollutant.

Task 3 – Calculate Percentage Change from 2001

Then, the analysis focuses on how each pollutant changed—identifying which ones increased, decreased, or remained stable over time.

- **Action:** Compute percentage change using the formula:

$$\text{Percentage Change} = \left(\frac{\text{Value}_{\text{Year } X} - \text{Value}_{2001}}{\text{Value}_{2001}} \right) \times 100$$

- **Target:** Relative evolution of pollution levels.
- **Object:** Yearly average pollutant concentrations.
- **Measures:** Percentage change from 2001 for each pollutant.
- **Groupings:** Year, Pollutant.

Task 4 – Detect Patterns or Anomalies

Significant shifts in pollution levels will be identified by examining year-over-year differences and flagging anomalies.

- **Action:** Analyze yearly averages for spikes or drops, calculating absolute and percentage differences between consecutive years.
- **Target:** Years with abnormal pollution levels.
- **Object:** Pollution timeline.
- **Measures:** Rate of change (e.g., µg/m^3 difference, percentage difference) year over year.
- **Groupings:** Year, Pollutant.

Task 5 – Compare Key Milestones

Pollution levels in benchmark years (2001, 2008, 2018) will be compared to summarize long-term changes, with 2008 as a midpoint potentially reflecting policy shifts.

- **Action:** Compare average concentrations and calculate total and percentage changes between 2001, 2008, and 2018.
- **Target:** Pollution evolution at the start, midpoint, and end of the period.
- **Object:** Pollutant concentrations in key years.
- **Measures:** Differences (e.g., µg/m^3) and percentage changes between these years.
- **Groupings:** Year, Pollutant.

Task 6 – Visualization Preparation

At last, visualizations to effectively communicate trends, patterns, and comparisons are designed.

- **Action:** Create visual tools (e.g., line charts) to display evolution.
- **Target:** Clear communication of pollution patterns.
- **Object:** Visualizations such as line graphs or heatmaps.
- **Measures:** Percentage change, average concentrations.
- **Groupings:** Year, Pollutant.

From Question to Task – Question 2

*The given question is: Which are the areas where the pollution is the highest/lowest? To address this question, it is divided into specific tasks to analyze the temporal evolution of pollution levels across Madrid over the specified period, using the available dataset of yearly pollution measurements and station metadata.

Task 1 – Filter the Year 2018

The data is filtered to include only measurements from 2018. **Action:** Browse (filter year 2018) **Target:** Stations with data from 2018 **Object:** Stations with data from 2018 **Measures:** Date (the year with timestamp 2018) **Groupings:** Data that is not from 2018 is filtered out.

Task 2 – Which Station Has the Highest (Lowest) Pollution

There are several gases that lead to high pollution. Three types of pollution are defined: *Gases and Chemicals:* SO_2 ($\mu g/m^3$), CO (mg/m^3), NO ($\mu g/m^3$), NO_2 ($\mu g/m^3$), O_3 ($\mu g/m^3$). Define a new variable, total_gas, as the sum of these compounds in $\mu g/m^3$. *Particulate Matter:* PM25 ($\mu g/m^3$), PM10 ($\mu g/m^3$). Define a new variable, total_matter, as the sum of these compounds in $\mu g/m^3$. *Volatile Organic Compounds (VOCs):* TOL ($\mu g/m^3$), BEN ($\mu g/m^3$), EBE ($\mu g/m^3$), MXY ($\mu g/m^3$), PXY ($\mu g/m^3$), OXY ($\mu g/m^3$), THC (mg/m^3), CH4 (mg/m^3), NMHC (mg/m^3). Define a new variable, total_VOC, as the sum of these compounds in $\mu g/m^3$. overall_total as total_VOC + total_matter + total_gas is defined as a total for the three types of pollution. **Action:** Identifying **Target:** Extremes **Object:** Total pollution in a station **Measures:** total_gas, total_matter, total_VOC, or overall_total **Groupings:** Variables that do not provide information on pollution or the station are filtered out.

Task 3 – Rank the Stations with Highest and Lowest Pollution Levels

Using this methodology, the next step is to identify the station with the highest and lowest pollution levels.
Action: Rank order (browse)
Target: Worst/best regions (extremes)
Object: Total pollution in stations

Measures: total_gas, total_matter, total_VOC, or overall_total
Groupings: Only the bad gases (which harm people) are used, the rest is filtered out.

Task 4 – Filter Good/Bad Regions

With the worst and best stations identified, an order for all the stations can be found.
Action: Annotate (give a tag to "good" stations, etc.) and browse (filter stations)
Target: Stations with low pollution
Object: Total pollution in stations
Measures: Total mass of particles
Groupings: Stations are filtered into "good" and "bad".

Task 5 – What Do We Call High Levels of Pollution?

Which stations have "good" air quality? A station is said to have good air quality when the total pollution is in the 10% lowest quantile.
Action: Annotate (give a tag to stations with good air quality, etc.)
Target: Polluted stations
Object: Total mass of particles
Measures: Average pollution
Groupings: Good vs. bad stations

Task 6 – Is There a Difference Between the Data from 2008 and 2018?

The steps above can be repeated for the year 2008 and then compared with the year 2018.
Action: Compare
Target: Polluted stations out of 2008 and 2018
Object: Total mass of particles in different years
Measures: Total mass of particles in the different years
Groupings: Polluted stations in 2008 vs. polluted stations in 2018

From Question to Task – Question 3

The given question is: Which areas of Madrid improved or worsened more between 2008 and 2018?

Task 1 – Filter Data for 2008 and 2018

Question: Which areas improved/worsened?
Action: Browse (pick only 2008 and 2018 data)
Target: Data from 2008 and 2018
Object: Pollution measurements
Measures: Date (specifically filtering for the years 2008 and 2018 within the timestamp)
Groupings: By year.

Task 2 – Find Average Pollution for Each Station

Question: Which areas improved/worsened?
Action: Compute yearly average pollutant levels per station.
Target: Station-level averages.

Object: Pollution data for each station
Measures: Levels of NO_2, PM10, PM2.5 (in $\mu g/m^3$)
Groupings: Stations, years (2008 and 2018), and types of pollutants

Task 3 – See How Much Pollution Changed

Question: Which areas improved/worsened?
Action: Subtract (2018 numbers minus 2008 numbers)
Target: Pollution change per station.
Object: Pollution data for each station
Measures: Changes like NO_2 in 2018 minus NO_2 in 2008 (in $\mu g/m^3$)
Groupings: Stations and types of pollutants

Task 4: Calculate Air Quality Index (AQI) Change

Question: Which areas improved or worsened in terms of air quality?

Action: Compute the AQI change using a weighted combination of pollutant concentration changes. The formula, based on WHO health impact guidelines, is:

$$\Delta AQI = 0.333 \times \Delta NO_2 + 0.222 \times \Delta PM_{10} + 0.444 \times \Delta PM_{2.5}$$

where:

- ΔNO_2, ΔPM_{10}, and $\Delta PM_{2.5}$ represent the changes in pollutant concentrations between 2008 and 2018.
- The weights reflect the relative health impact of each pollutant, prioritizing $PM_{2.5}$ due to its stronger association with respiratory and cardiovascular harm.

The ΔAQI formula focuses on NO_2, PM_{10}, and $PM_{2.5}$ due to their significant health impacts and data availability. According to WHO guidelines, $PM_{2.5}$ and NO_2 are strongly associated with respiratory and cardiovascular diseases, with $PM_{2.5}$ having the highest health impact weight in the AQI formula (0.4), followed by NO_2 (0.3), and PM_{10} (0.2). In contrast, O_3 and SO_2 have lower weights (0.05 each) due to their relatively smaller health effects in urban settings like Madrid. Additionally, an analysis of the Madrid dataset (2008 and 2018) revealed that O_3 and SO_2 measurements had significant missing data—approximately 30% of O_3 and 25% of SO_2 values were unavailable across stations, compared to less than 5% for NO_2, PM_{10}, and $PM_{2.5}$. Including O_3 and SO_2 would have required extensive imputation, potentially skewing the ΔAQI results. Therefore, we prioritized NO_2, PM_{10}, and $PM_{2.5}$ to ensure a reliable and health-focused metric for comparing air quality changes across stations.

Target: A unified air quality metric that captures overall change per location.
Objects: Changes in pollutant concentrations.
Measure: AQI change in $\mu g/m^3$.
Groupings: Monitoring stations.

Task 5 – Visualize AQI and Pollutant Changes

Question: Which areas improved/worsened?
Action: Create a bar chart to display individual pollutant changes and a combined AQI

line.
Target: Spatial visualization of changes.
Object: Bar chart with pollutant-specific bars and AQI line.
Measures: Changes in NO_2, PM10, PM2.5, and AQI change (in $\mu g/m^3$).
Groupings: Station locations.

Task 6 – Analyze Elevation Impact

Question: Which areas improved/worsened?
Action: Correlate elevation with AQI change.
Target: Elevation-pollution relationship.
Object: Elevation, AQI change.
Measures: AQI change ($\mu g/m^3$), elevation (m).
Groupings: Stations by elevation bands.

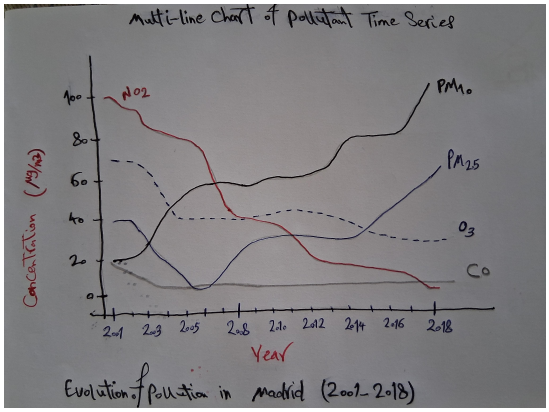
The Design Process

Diverge Phase

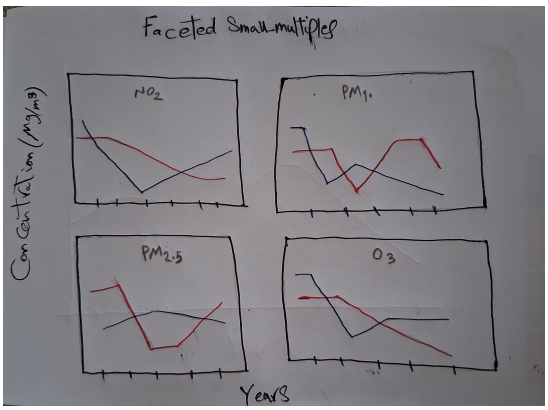
Question 1 – How has pollution in Madrid evolved between 2001 and 2018?
Here, a broad set of possible visualizations can be used to show how pollution has changed over 18 years.

- Multi-line chart of pollutant time series: One line per major pollutant (NO_2, PM10, O_3...), all on the same axes, showing yearly averages. A sketch of this design is shown in fig. 1a
- Faceted small-multiples: A grid of small line plots (one per pollutant), aligned on the same time axis. A sketch of this design is shown in fig. 1b
- Single "composite score" line chart: A single line representing an overall pollution index (e.g., average of normalized pollutant values) per year. A sketch of this design is shown in fig. 1c
- Bar chart of year-over-year change: Bars showing the change in a composite score between consecutive years. A sketch of this design is shown in fig. 1d
- Heatmap (year × pollutant): Rows = pollutants, columns = years; color intensity = concentration or percentage change. A sketch of this design is shown in fig. 1e

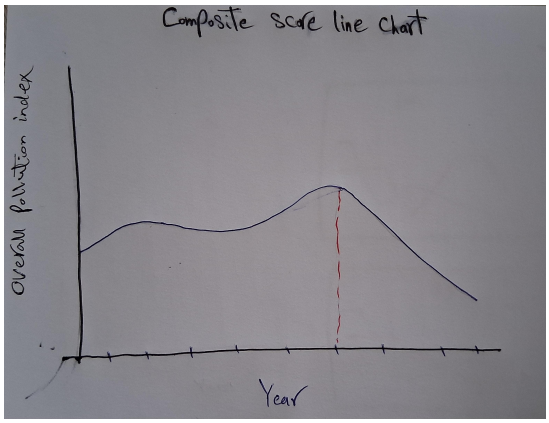
The sketches of these ideas are provided in fig. 1



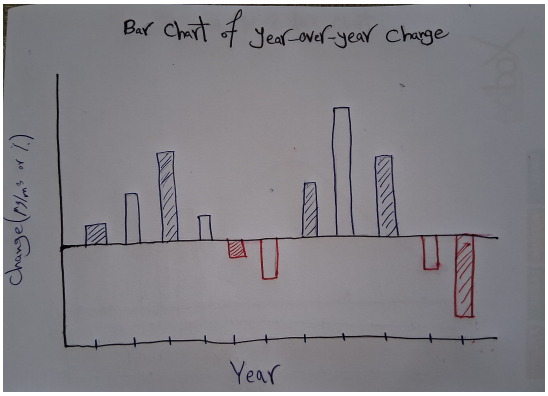
(a) Multi-Line Chart



(b) Faceted Small-Multiples



(c) Composite Score Line Chart



(d) Bar Chart of Year-over-year Change

(e) Pollution Evolution Heatmap

Figure 1: All sketches drawn for question 1

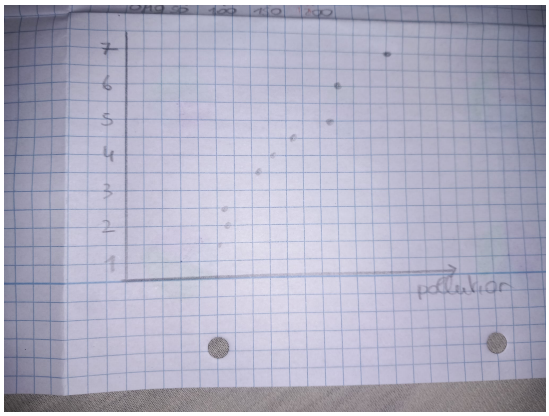
Question 2 – Which are the stations where the pollution is the highest/lowest? The goal is to compare pollution across different stations. With a categorical variable (the city) and a continuous variable (the pollution), several types of plots can be created to visualize the data.

- A bar plot in which the total pollution of each station is given. The bars are ordered

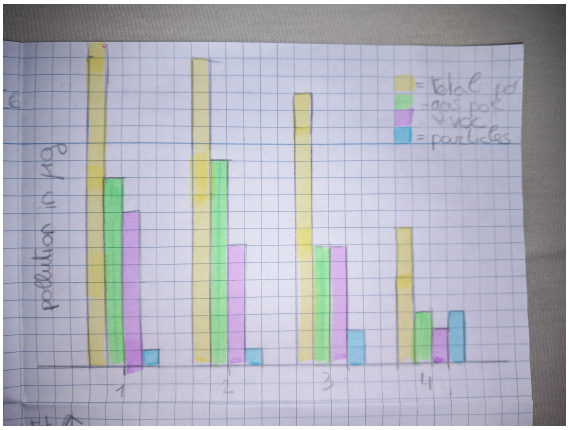
from highest to lowest pollution. A sketch of this design is shown in fig. 2f

- A stacked bar plot in which the 3 types of pollution are added on each other. This also allows for easy comparison of the total pollution levels. A sketch of this design is shown in fig. 2e
- A grouped bar chart where the 3 types of pollution and the total pollution are given for each station (next to each other). A sketch of this design is shown in fig. 2b
- A map of Madrid where each position of a station is marked in a specific color (on a scale from red to green with red a lot of pollution and green a minimal of pollution). A sketch of this design is shown in fig. 2c
- A dot plot where the total pollution for each station is given. The total pollutions are ordered A sketch of this design is shown in fig. 2a
- A pie plot for all the different stations in which it can be seen which kind of pollution is dominating. A sketch of this design is shown in fig. 2d

It is ensured that the bars are sorted from highest to lowest pollution levels. Sketches for these ideas are provided in fig. 2

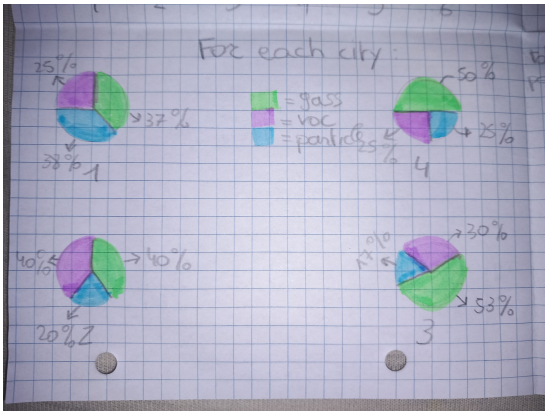


(a) Dot plot

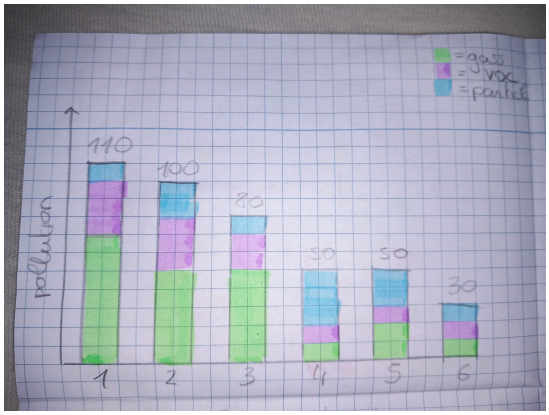


(b) A grouped bar chart

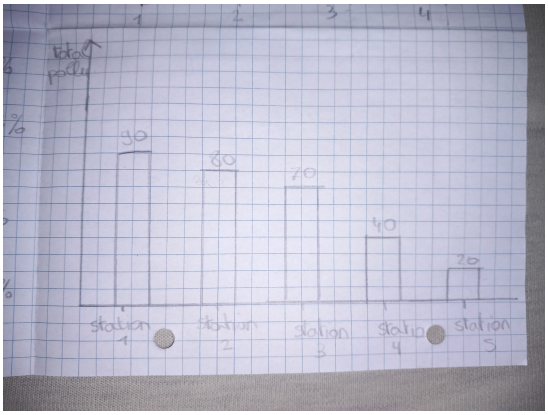
(c) A map of Madrid with the pollution in each station.



(d) Pie plot



(e) Stacked bar plot



(f) A bar plot showing the total pollution in each station

Figure 2: All sketches drawn for question 2

Question 3 – Which areas of Madrid improved or worsened more between 2008 and 2018?

The goal is to show how air quality changed across different stations in Madrid between 2008 and 2018. Since the overall change in pollutant concentrations and AQI is available for each station, the following visualization ideas can be considered:

- A map of Madrid with dots for each station (green dots indicate improvement in air quality with a decrease in AQI, while red dots indicate worsening with an increase in AQI).
- A bar chart with stations on the x-axis and the overall AQI change on the y-axis (bars extending upward indicate worsening air quality, while bars extending downward indicate improvement).
- A few small maps, one for each pollutant (e.g., NO_2, PM_{10}, $PM_{2.5}$), showing changes at each station with color coding (green for reduced levels, red for increased levels).
- A map with bubbles for each station (bigger bubbles represent larger changes in AQI, with colors indicating direction: green for improvement, red for worsening).
- A heatmap to display the magnitude and direction of AQI changes across Madrid, with color intensity reflecting the extent of improvement or worsening at each station.

We made sure to pick ideas that are easy to understand and highlight station-specific changes over the 2008-2018 period.

Emerge Phase

We looked at all our ideas using a NUF test (New, Useful, Feasible) and picked the best ones to make better.

Question 1: The following visualizations are chosen based on our NUF test:

- Multi-line chart with interactive features allowing users to toggle pollutants on/off
- Heatmap showing all pollutants across years with color intensity indicating concentration

Question 2: The following visualizations are chosen based on our NUF test:

- Interactive map with color-coded regions (darker for higher pollution) and hover tooltips.
- Grouped bar chart showing total_gas, total_matter, and total_VOC per region.
- Ordered bar chart of overall_total, with regions sorted from highest to lowest pollution.

Refined sketches are provided as Figures 12–14.

Question 3: The following visualizations are chosen based on our NUF test:

- A map with colored dots for stations (green for better AQI, red for worse AQI), where you can click to see more details.
- A bar chart with stations and grouped bars for NO_2, PM_{10}, and $PM_{2.5}$ changes.

- Small maps for each pollutant to easily compare how changes in different areas look.

These sketches are chosen because they show the changes clearly and are not too hard to make.

NUF Test Results for Question 1:

- **Multi-line chart:** New: 5/10 (it's a standard visualization), Useful: 9/10 (clearly shows trends for each pollutant), Feasible: 8/10 (straightforward to implement) → 22/30.
- **Small-multiples:** New: 6/10 (less common than single charts), Useful: 7/10 (good for comparing pollutants separately), Feasible: 7/10 (requires more space and work) → 20/30.
- **Composite score line chart:** New: 4/10 (very standard), Useful: 6/10 (simplifies but loses detail), Feasible: 9/10 (easy to implement) → 19/30.
- **Bar chart of changes:** New: 5/10 (standard), Useful: 7/10 (highlights year-to-year differences), Feasible: 8/10 (simple to create) → 20/30.
- **Heatmap:** New: 7/10 (less common for time series), Useful: 8/10 (shows all data in compact form), Feasible: 7/10 (requires careful color choice) → 22/30.

  We chose to further develop the multi-line chart (fig. 1a) and the heatmap (fig. 1e) based on their NUF scores and complementary nature.

NUF Test Results for Question 2:

- **Interactive Map of Madrid:** New: 8/10 (we haven't seen many maps like this), Useful: 9/10 (it shows areas clearly), Feasible: 7/10 (a mapping tool is needed) → 24/30.
- **Grouped Bar Chart:** New: 5/10 (it's a normal chart), Useful: 8/10 (good for comparing), Feasible: 8/10 (it is easy to make, but 4 bars for all the stations will get a lot fast) → 22/30.
- **Bar chart of the total pollution:** New 4/10 (it's a normal chart), Useful: 7/10 (only total pollution can be compared), Feasible: 9/10 (can be made easily) → 20/30.
- **Stacked bar chart:** New: 5/10 (it's a normal chart), Useful: 8/10 (good for comparing stations and types of pollution), Feasible: 9/10 (can be made easily) → 22/30.
- **Dot plot:** New: 4/10 (it's a basic plot), Useful 7/10 (only total pollutions can be compared), Feasible 9/10 (can be made easily) → 20/30.
- **Pie plot:** New 5/10 (it's a basic plot), Useful 4/10 (only proportions of pollution can be compared,and not absolute numbers), Feasible: 9/10 (can be made easily) → 18/30.

We chose to further develop the interactive map of Madrid (fig. 2c) and the stacked bar chart (fig. 2e).

NUF Test Results for Question 3:

- Map with Dots: New: 7/10 (it's like Q2 but shows changes), Useful: 9/10 (answers our question well), Feasible: 7/10 (needs a map) → 23/30.
- Bar Chart: New: 5/10 (it's a basic chart), Useful: 8/10 (changes for each pollutant like NO_2, PM_{10}, and $PM_{2.5}$ can be seen), Feasible: 9/10 (easy to do) → 23/30.
- Small Maps: New: 6/10 (not too common), Useful: 7/10 (good for comparing pollutant-specific changes across areas), Feasible: 8/10 (small maps can be made) → 21/30.
- Scatter Plot of Elevation vs. Pollution Change: New: 6/10 (scatter plots are common, but elevation analysis is novel in this context), Useful: 8/10 (provides actionable insight for urban planning based on AQI changes), Feasible: 9/10 (uses existing data, easy to implement) → 23/30.

The Final Design

Question 1 - Final Designs:

Design 1: Interactive Multi-line Chart

Description: A line chart showing the yearly average concentrations of key pollutants (NO_2, PM10, PM2.5, O_3, SO_2) from 2001 to 2018.
Visual Encoding:

- **Position:** Year on x-axis, pollutant concentration on y-axis
- **Color:** Different colors for each pollutant
- **Line type:** Solid lines for primary pollutants, dashed for secondary ones

  Interaction: Users can toggle pollutants on/off to focus on specific ones. Hovering shows exact values.

Purpose: This visualization shows how individual pollutant concentrations have changed over time, allowing for direct comparison between pollutants and identification of trends.

Design 2: Pollution Evolution Heatmap

Description: A year-by-pollutant heatmap where each cell's color represents the concentration level of a specific pollutant in a specific year. .
Visual Encoding:

- **Position:** Years on x-axis, pollutants on y-axis
- **Color:** Sequential color scale from light blue (low concentration) to dark red (high concentration)
- **Text:** Numerical values shown in cells for precise readings

  Interaction: Clicking on a year displays a detailed breakdown of all pollutants for that year

Purpose: The heatmap provides a comprehensive overview of all pollutants across all years, making it easy to spot patterns, anomalies, and overall trends in pollution levels.

These two complementary visualizations together provide a comprehensive view of how pollution has evolved in Madrid between 2001 and 2018, addressing different aspects of the temporal patterns.

Question 2 – Final Designs:

Design 1: Interactive Map of Madrid

Description: A map of Madrid where the positions of the different stations are marked as a dot in a specific color (on a color scale between red and green, with red a lot of pollution and green a minimal of pollution) **Visual Encoding:**

- **Position:** Longitude on the x-axis, latitude on the y-axis
- **Color:** Green for good air quality (low pollution), red for bad quality (high pollution). A color scale between red and green can be used to visualize locations with a medium air quality
- **Size:** All the dots are of the same size

Interaction: Users can click on a station to view a detailed breakdown of pollutant levels and compare these with city-wide averages. **Purpose:** This helps us explore whether there is a spatial relationship between a station's location (near the center, near a park, ...) and its air quality.

Design 2: Stacked bar plot of the different types of pollution Description: A Bar plot where each bar consists of the different components of pollutants (gasses, volatile organic compounds and small particles). The total height of the bar gives the total pollution in a station. **Visual Encoding:**

- **Position:** station on the x-axis, total pollution on the y-axis (sorted from high total pollution to low total pollution)
- **Color:** Each type of pollution is represented with its own color (yellow for polluting gasses, red for volatile organic compounds and blue for small particles)
- **Size:** The height of the bar gives the total pollution. The higher a bar, the higher the pollution. The different colors in the bar makes it possible to compare different types of pollution by the height of a certain color in the bar.

Interaction: You can click on a bar and see the pollution levels of the different types. **Purpose:** This helps us compare the total pollution, and the different types of pollution, across all the different stations.

Question 3 – Final Designs:

Design 1: Map with Colored Dots

Description: A map of Madrid with a dot for each station. Green dots indicate an improvement in air quality (decrease in AQI), and red dots mean it got worse (increase in AQI). Bigger dots mean a larger change in the overall AQI.

Visual Encoding:

- **Position:** Where the stations are on the map (using longitude and latitude).
- **Color:** Green for better, red for worse.

- **Size:** Bigger dots for larger changes.

Interaction: You can click on a dot to see the changes for each pollutant, like how much NO_2 changed.
Purpose: This map makes it easy to see which areas of Madrid got better or worse, and which ones changed the most through spatial analysis. This allows policymakers to target interventions in the most affected areas.

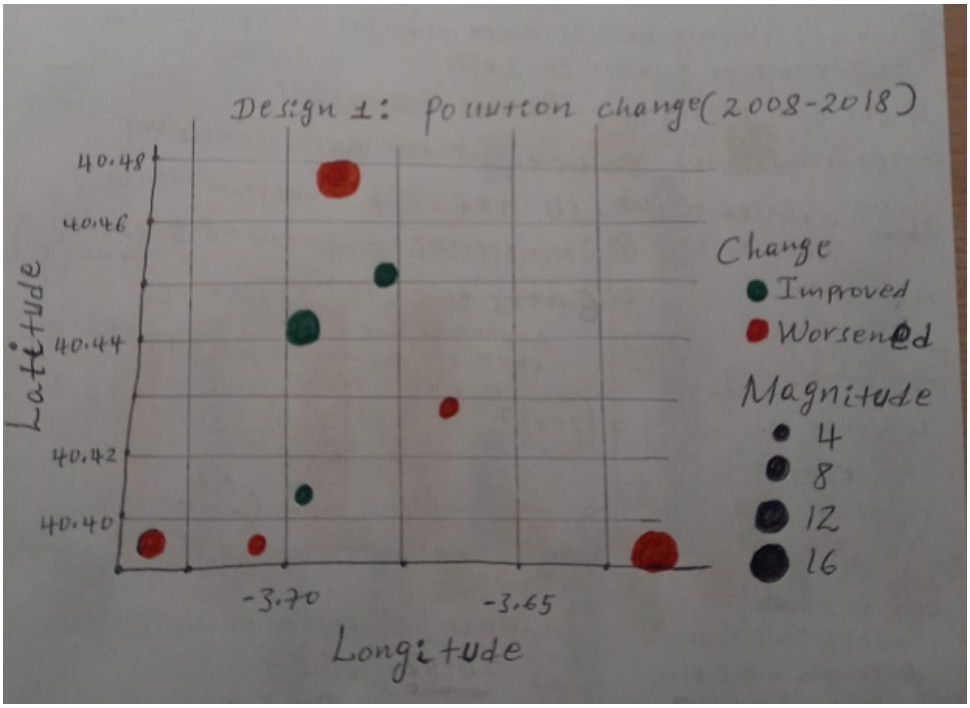


Figure 3: Sketch of Design 1: Map with Colored Dots showing pollution changes in Madrid (2008-2018).

Design 2: Bar Chart with Different Pollutants
Description: A bar chart where each station has bars for NO_2, PM_{10}, and $PM_{2.5}$ changes (in $\mu g/m^3$).
Visual Encoding:

- **Position:** Stations on the x-axis, changes on the y-axis (up means worse, down means better).
- **Color:** Different colors for each pollutant (blue for NO_2, black for PM_{10}, green for $PM_{2.5}$).
- **Length:** How long the bar is shows how much it changed.

Interaction: Bars can be sorted to highlight the stations with the largest changes first, and hovering over a bar shows the exact change value.
Purpose: Compare how different pollutants changed at each station to identify areas with the most significant improvements or worsening in air quality.

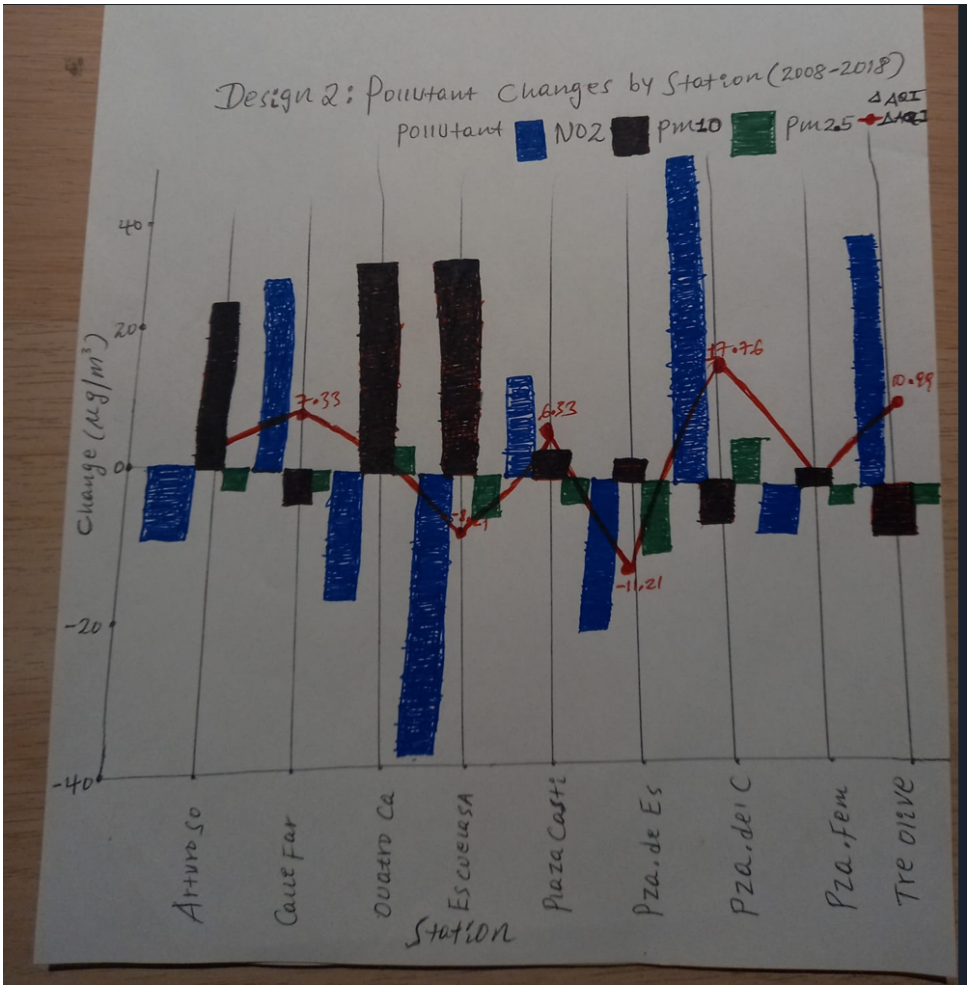


Figure 4: Sketch of Design 2: Bar Chart with Different Pollutants showing changes in Madrid (2008-2018).

Design 3 – Small Maps for Each Pollutant:
Description: A series of small maps, one for each pollutant (e.g., NO_2, PM_{10}, $PM_{2.5}$), displaying changes in pollutant levels (in $\mu g/m^3$) across stations. Each map focuses on a single pollutant and highlights spatial variation in change values.
Visual Encoding:

- Station locations are represented as dots on the map; color indicates direction of change in pollutant levels (green for improvement, red for worsening); size indicates magnitude of change (larger dots represent greater change).

Interaction: Selecting a station highlights it simultaneously on all pollutant maps, allowing comparison across pollutants for the same location.
Purpose: To explore spatial patterns in pollutant-specific changes and identify areas with consistent or divergent trends in air quality improvements or worsening.

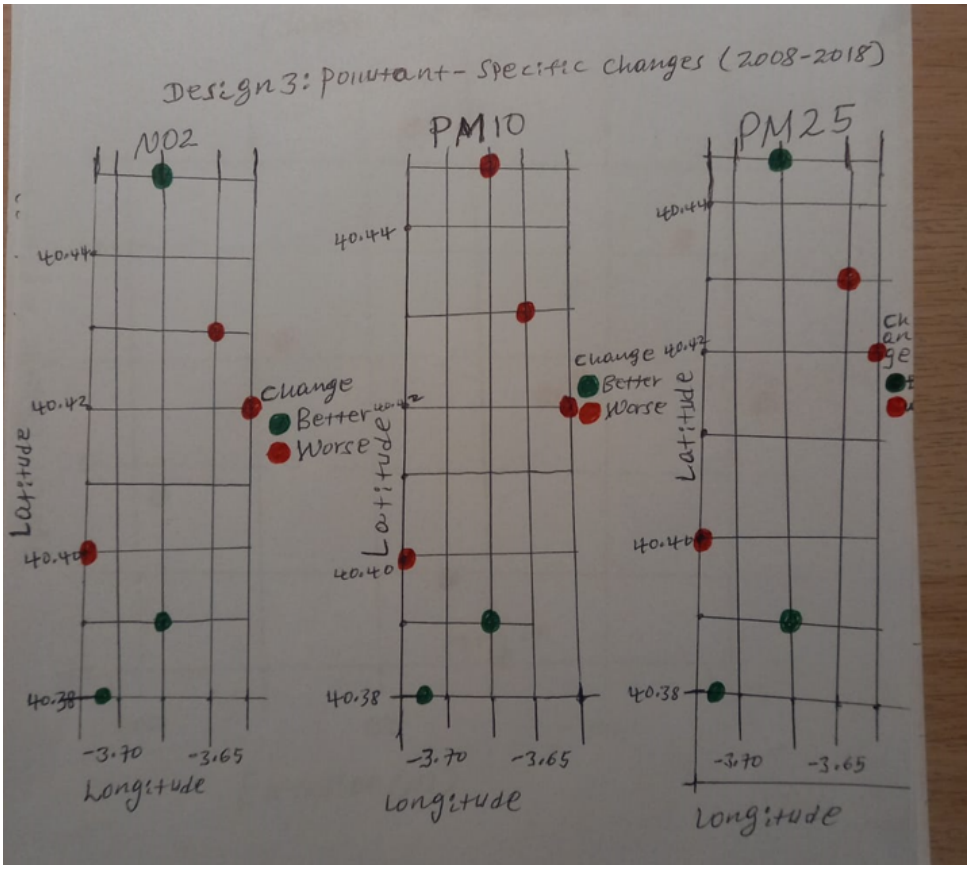


Figure 5: Sketch of Design 3: Small Maps for Each Pollutant showing changes in Madrid (2008-2018).

Design 4 – Scatter Plot of Elevation vs. Pollution Change:
Description: A scatter plot with elevation on the x-axis (ranging from 599 to 728 m) and AQI change on the y-axis (e.g., from -20 to $+20$ μg/m^3, illustrative range). Each point represents a station. Points are colored green for stations with improved air quality (negative AQI change) and red for stations with worsened air quality (positive AQI change).
Visual Encoding:

- Elevation on the x-axis, AQI change on the y-axis; point color encodes direction of change (green for improvement, red for worsening).
- Optional text labels for notable stations (e.g., "Pza. de Es").

Interaction: Hovering over a point reveals a tooltip showing the station name, elevation, and AQI change value; optionally, filter stations by specific pollutant changes for additional context.
Purpose: To visually assess the potential relationship between elevation and air quality changes across stations, helping identify elevation-related patterns in improvements or worsening.

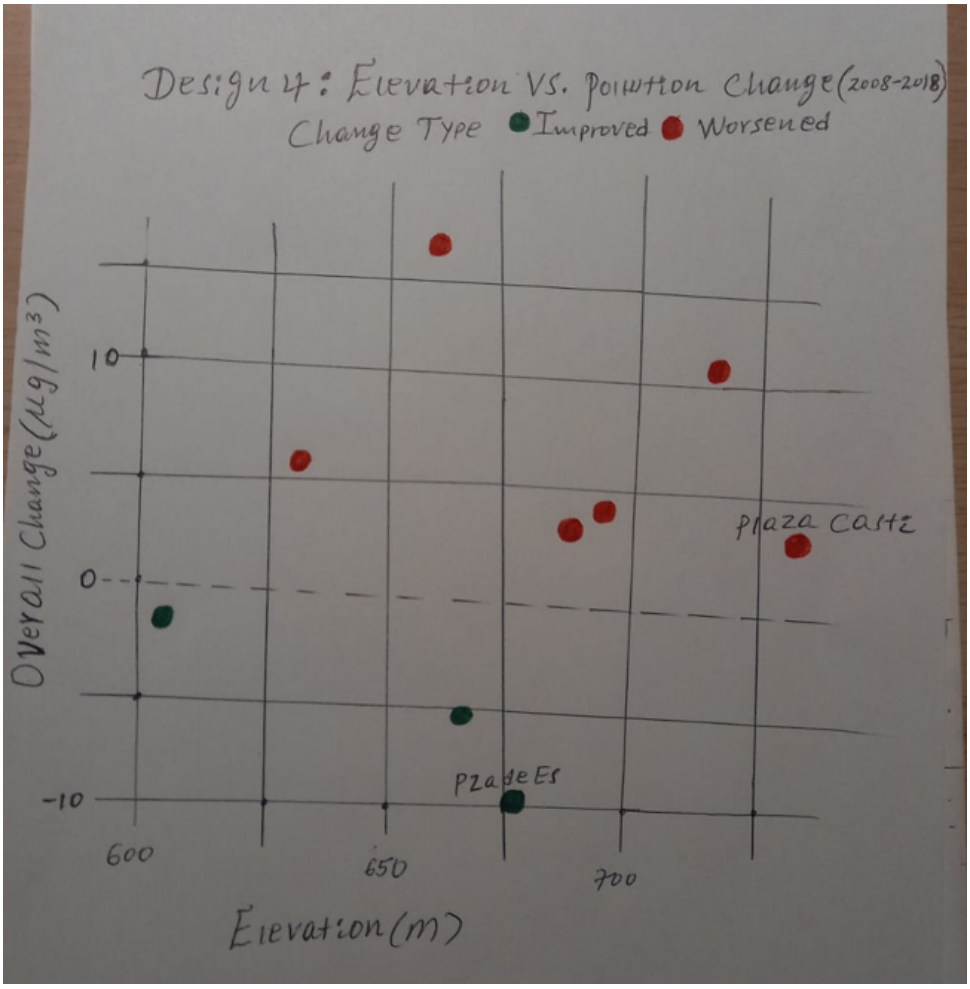


Figure 6: Sketch of Design 4: Scatter Plot of Elevation vs. Pollution Change in Madrid (2008-2018).

1 Implementation - analyses

1.1 Explanatory data analyses

Before starting, the data in the dataset is analyzed to get an idea of what is in the dataset.

1.1.1 Missing values

In the total dataset, all the different years included, there were a lot of missing values. The number of missing variables for each variable is given in table 1. It can be seen that for a lot of variables the number of missing values is very high. Note that this is for the complete dataset, and that the complete dataset not always is needed for a given example. The way the missing values are handled will be discussed for each visualization.

Table 1: Number and Percentages of missing values per variable

Variable	Missing values	Percentage (%)
BEN	2,766,540	72.65
CO	1,157,212	30.39
EBE	2,806,500	73.71
MXY	3,492,809	91.73
NMHC	2,722,912	71.49
NO_2	21,174	0.56
NOx	1,431,949	37.59
OXY	3,492,529	91.72
O_3	816,492	21.45
PM10	946,969	24.87
PXY	3,492,640	91.72
SO_2	1,032,264	27.11
TCH	2,721,783	71.45
TOL	2,769,295	72.73
station	0	0.00
PM25	2,991,800	78.56
NO	2,275,827	59.75
CH4	3,793,374	99.22

1.1.2 Summary Statistics

For all variables in the dataset, we calculated the mean, median, and standard deviation to gain a foundational understanding of their distributions. These can be found in table 2. These summary statistics help identify the central tendency and variability of each variable. Comparing the mean and median can reveal skewness in the data or the presence of outliers, as the mean is influenced by extreme values whereas the median is more robust. A large difference between the two may suggest non-normality or irregularities in the data. Additionally, the standard deviation indicates how much the values typically deviate from the mean, offering insight into the expected range and spread of

the data. These insights are crucial for detecting potential data quality issues, forming pre-processing decisions, and guiding further statistical or predictive analyses. Some

Table 2: Summary Statistics per variable, averaged over all the years and stations: mean, median and standard deviation

Variable	**Mean**	**Median**	**SD**
BEN	1.257	0.600	1.9108
CO	0.550	0.400	0.5354
EBE	1.408	0.880	2.1461
MXY	4.650	2.800	5.5992
NMHC	0.187	0.150	0.1540
NO_2	50.47	44.00	34.5529
NOx	109.3	76.15	110.2871
OXY	2.281	1.320	2.6396
O_3	39.83	34.86	30.3925
PM10	28.94	21.49	25.9486
PXY	2.056	1.280	2.3951
SO_2	10.66	8.150	9.1213
TCH	1.436	1.380	0.2332
TOL	5.877	3.160	8.5245
PM25	13.74	11.00	11.2141
NO	23.44	6.000	50.2150
CH4	1.301	1.250	0.1955

potential issues were identified in the data:

- For variables such as MXY, NOx, PXY, TOL, and NO, the mean and median differ substantially, and the standard deviation is high. This may indicate skewed distributions or the presence of outliers. These variables should be treated with caution in subsequent analyses.
- For O_3, PM10, SO_2, and PM25, the standard deviation is also high, although the mean and median are relatively similar. This suggests a wide spread of values despite a more symmetric distribution.

The same descriptive statistics were computed for each year individually. While these results are not shown here, they were used to inform decisions about which data to include in visualizations and further analysis.

1.1.3 Trends over time

The different variables are plotted over time in a very basic line plot. This plot can be seen in fig. 7. From this plot it can be seen that there is a lot of trouble due to the missing values. For a lot of variables, the trend between 2001 and 2018 cannot be fully visualized. From the trends that can be seen, it looks like the pollutants all have a decreasing trend over time, so pollination tends to decrease as time increases. These effects will be discussed in full detail in the visualizations.
The trend for each pollutant individually can alse be seen in fig. 8. There a decreasing rend is seen again, except for O_3.

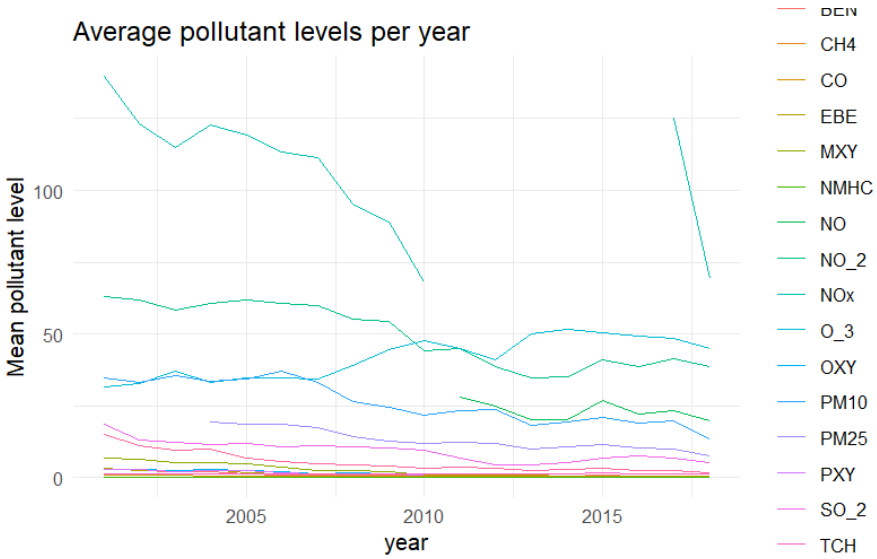


Figure 7: Trend of the different pollutants over time

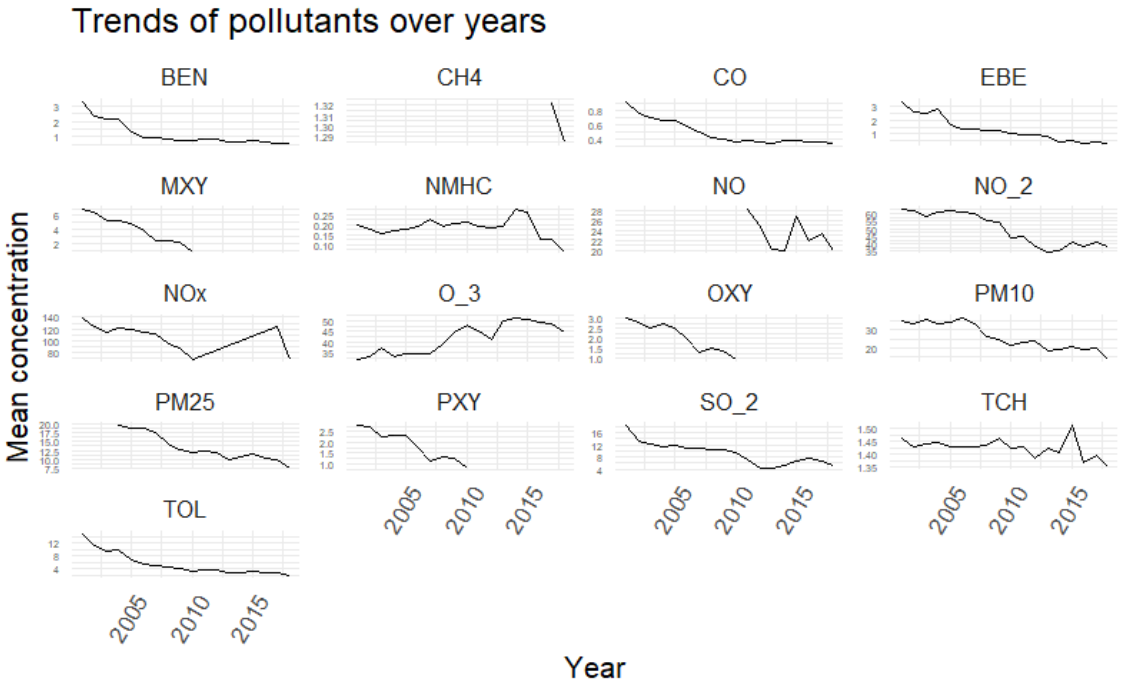


Figure 8: Trend per pollutant over years.

1.1.4 Correlation

The correlation between variables is shown in fig. 9. From the correlation matrix, several groups of strongly correlated variables can be identified:

- There is a strong positive correlation between **NO**, **NOx**, and **NO_2**. This is expected, as these are all nitrogen-based pollutants that often originate from similar combustion processes, such as traffic and industrial emissions.
- **PM10** and **PM25** are also strongly positively correlated. Both represent particulate matter in the air, with PM10 including particles up to 10 micrometers and

PM25 focusing on finer particles up to 2.5 micrometers. Their high correlation suggests that the sources and dynamics of these particles often overlap.

- A notable negative correlation is observed between **NO_2** and **O_3** (ozone). This is consistent with known atmospheric chemistry: in urban areas, ozone is often consumed by NO (a component of NOx) in a process called titration, which reduces local ozone levels.
- **CH_4** (methane) and **TCH** (Total Hydrocarbons) show a strong positive correlation. Since methane is one of the main components of total hydrocarbons, this correlation is expected.
- **MXY**, **PXY**, **OXY**, **TOL**, **EBE**, **CO**, and **BEN** form a cluster of strongly positively correlated variables. These are all volatile organic compounds (VOCs), often emitted from traffic-related sources such as fuel combustion and evaporative emissions. Their strong correlations suggest they share common emission sources and follow similar atmospheric behaviors.

While these correlations are not a direct concern for our analysis—since we are mainly focused on trend visualization—they do help in grouping related variables. Recognizing which variables move together over time helps in selecting representative pollutants for plotting, and can reveal shared sources or seasonal behaviors.

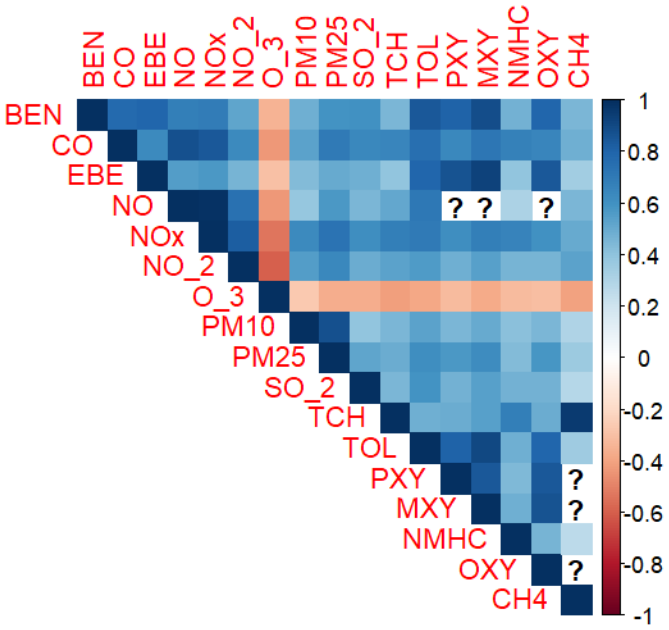


Figure 9: Correlation plot for the different pollutant variables

2 Implement visualizations

2.1 Question 1 - Visualization of pollution evolved in Madrid between 2001 and 2018

Based on the peer evaluations provided feedback on the initial designs for Question 1 with an interactive multi-line chart and a heatmap. The multi-line chart was praised for clearly

showing trends (NUF score: 22/30) and its interactivity (toggling pollutants) was seen as useful. The heatmap was appreciated for its compact overview (NUF score: 22/30), with suggestions to add interactivity like clicking a year for details. Reviewers recommended enhancing interactivity (e.g., filtering options) and addressing scale differences due to varying pollutant units (e.g., CO in mg/m^3 vs. others in $\mu g/m^3$).

Given this feedback and the question's focus on temporal evolution, the decision made to choose to implement an interactive multi-line chart with a dropdown to select pollutants. This adapts the original multi-line chart by showing one pollutant highlighted at a time in comparison to other pollutants, avoiding scale confusion, and incorporates interactivity via a dropdown, aligning with reviewer suggestions and practical considerations from the thinking trace.

2.1.1 Exploratory Data Analysis

To implement this plot, we used Python to explore the data first, and then implement it. In addition, data processing was used by Pandas, and an interactive visualization was used by Plotly.

The code loads the data, computes city-wide yearly averages for key pollutants (SO_2, CO, NO_2, PM10, PM25, O_3) and creates a line chart with a drop-down to select pollutants and hover details. The implemented plot is the interactive Multi-Line chart of annual average pollutant concentrations in Madrid from 2001-2018. This plot is based on the mockup that can be seen in fig. 10. In addition, we can see an overview of the data explored in table 3, which includes the average of each key pollutant in each year.

Year	SO_2	CO	NO_2	PM10	PM2.5	O_3
2001	18.41	0.93	63.11	35.08	-	31.38
2002	13.26	0.77	62.14	33.39	-	32.27
2003	12.36	0.70	58.40	35.56	-	36.89
2004	11.36	0.65	60.75	33.68	19.76	33.15
2005	11.86	0.65	61.98	34.34	18.74	34.53
2006	10.70	0.57	60.75	37.02	18.57	34.68
2007	10.91	0.50	60.01	33.20	17.42	34.53
2008	10.56	0.41	55.36	26.63	14.39	39.07
2009	10.32	0.39	52.85	24.05	12.96	42.30
2010	9.67	0.36	44.55	21.01	11.86	47.39
2011	6.92	0.37	44.88	23.29	12.27	44.99
2012	4.37	0.35	38.65	23.59	11.81	41.17
2013	4.43	0.33	34.72	18.38	9.95	50.00
2014	5.02	0.37	35.07	19.29	10.62	51.63
2015	6.88	0.37	40.99	21.00	11.47	50.57
2016	7.69	0.35	38.56	19.16	10.38	49.06
2017	6.85	0.36	41.57	19.95	9.95	48.32
2018	5.39	0.34	38.64	13.54	7.70	44.85

Table 3: Air Pollution Data Over the Years

2.1.2 The implementation

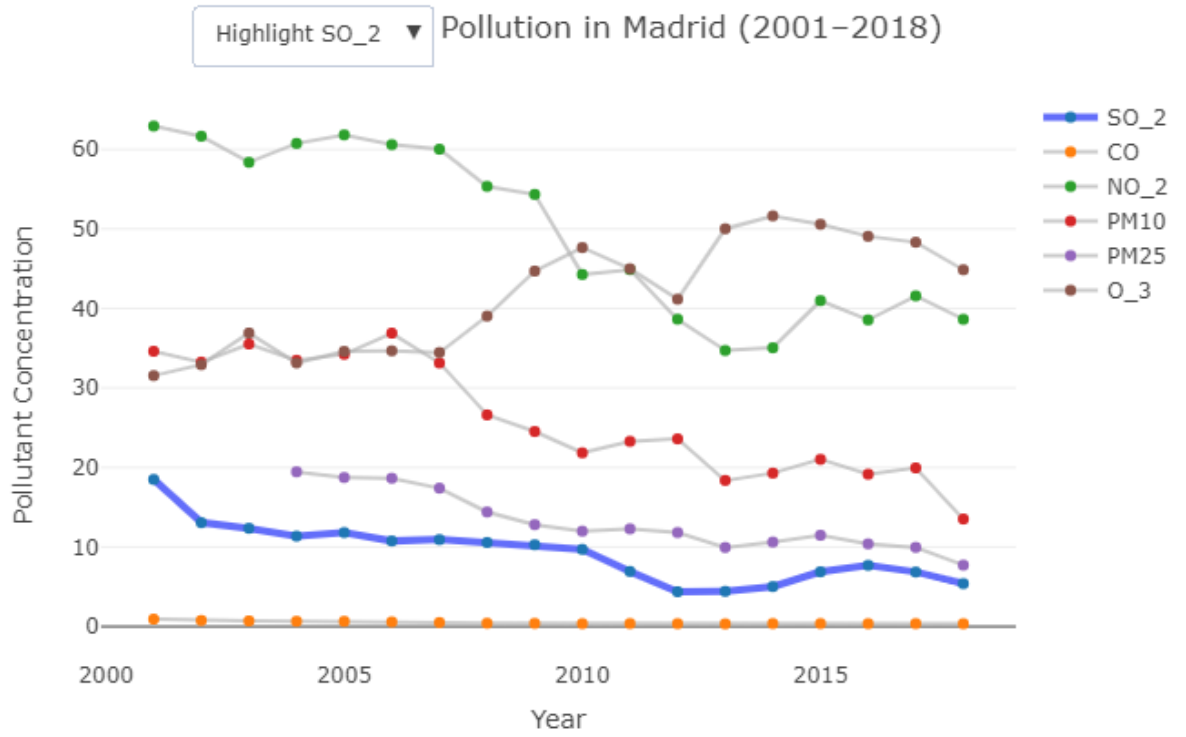


Figure 10: The mock-up of the implemented interactive Multi-Line Plot of Pollution Evolution in Madrid (2001-2018)

The final implementation is made using Python with the help of Streamlit to make it visible, interactive, and online as can be seen in fig. 11. The interactive Line-Plot itself can be found in `https://courses-project-bsf4y5ue6cepwdz7b29nhk.streamlit.app/`.

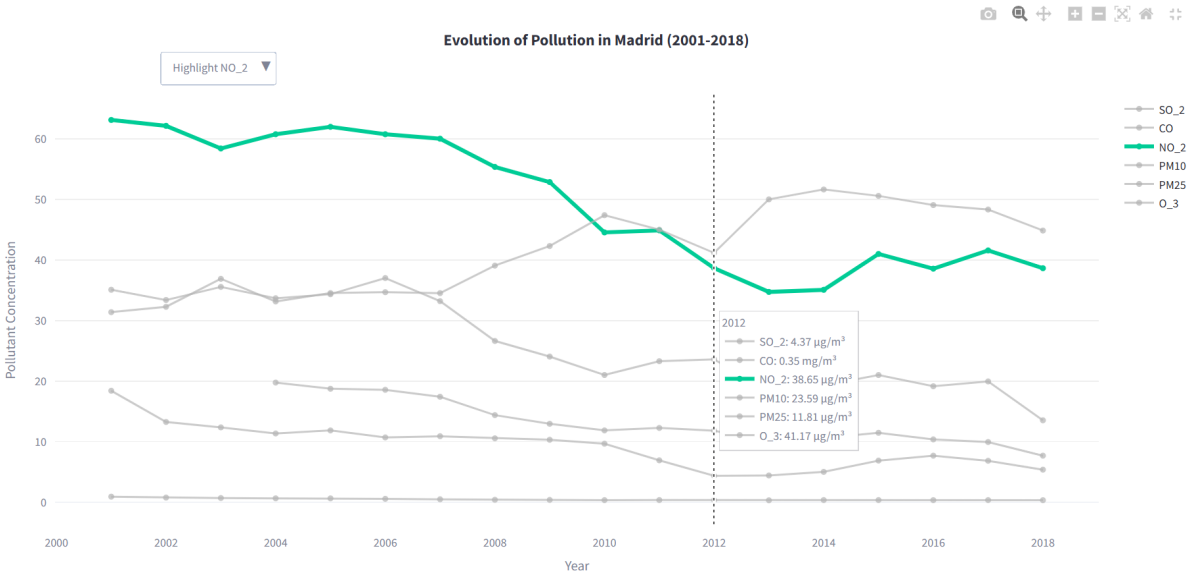


Figure 11: A screenshot of the final implementation of the interactive Multi-Line Plot of Pollution Evolution in Madrid (2001-2018)

fig. 11 has the following encodings:

- **The Data**
- **Source:** Loads yearly measurement files (assumed as madrid_2001.csv to madrid_2018.csv) from (2001–2018) and merged together.
- **Data processing:** calculates yearly averages for each pollutant, then aggregated to city-wide annual averages computed for each pollutant with handling missing values.
- **Pollutant selection:** We **only** included pollutants whose measurement columns were sufficiently complete (i.e. $\leq$ 10 % NaN across all stations and years). This meant **excluding** many VOCs (e.g. BEN, TOL, NMHC) since they had large gaps. The final set was: NO_2, **PM10**, **PM2.5**, O_3, SO_2, **CO**
- **Measure:** Mean concentration per pollutant, in µg/m^3 and µg/m^3 (all units harmonized).
- **Visual Encoding**
- **Marks:** Lines, one per pollutant.
- **Channels:**
  - **X-axis:** Year (2001–2018).
  - **Y-axis:** Annual mean concentration (µg/m^3) and (mg/m^3).
  - **Color:** Distinct hue per pollutant, so users can trace each time series.
  - **Line style:** Solid for all pollutants (NO_2, PM_{10}, $PM_{2.5}$, O_3, SO_2), Highlighted for pollutant selected by the action button that other pollutants turn into gray.
- **Interaction and interaction idioms**
- **Toggle visibility:** Click legend entries to show/hide individual pollutant lines.
- **Hover tooltips:** Mousing over a line reveals the exact year, pollutant name, and concentration value.
- **Dropdown Menu:** Selects a single pollutant to display, adjusting the y-axis title with appropriate units.
- **Zoom & pan:** Users can zoom into a subrange of years or pan along the time axis to examine specific periods.

Changes from initial design: Originally fig. 10 was a multi-line chart with all pollutants on one plot. But now it is a multi-line chart with a dropdown and selected highlights to improve clarity fig. 11.

2.2 Question 2 - Visualization of the pollution in different regions in Madrid in 2018

Based on the peer feedback, the plot that is implemented is the interactive map of Madrid. This plot is easy to read and provides all the necessary information for the comparison between stations. The pollutant on the chart is given based on the value of NO_2, since that was the only variable that contained information for all the station. The other measured pollutants in a given station can be found in the tooltip.
The plot is based on the mock-up that can be seen in fig. 13. Symbols are added to make the plot colorblind friendly.

2.2.1 Explanatory Data Analysis

Missing values
For this plot, only the data from 2018 is needed. The missing values and summary statistics are calculated again to see which variables can be used for this specific plot. The results can be found in table 4. It can be seen that the only variables that don't show a lot of missing values are NO_2, NOx and NO. This are all nitrogen pollutants, which are correlated. Therefore only NO_2 will be used in this plot. The other pollutants will be measured when they are available. FOr each station, the number of missing values are

Table 4: Summary Statistics per variable in 2018: number of missing values, percentage of missing values, mean, median and standard deviation

Variabele	Missing	% Missing	Mean	Median	SD
BEN	52146	75.47	0.556	0.400	0.455
CO	40498	58.61	0.344	0.300	0.202
EBE	52147	75.47	0.301	0.200	0.402
MXY	69096	100.0	–	–	–
NMHC	60656	87.79	0.065	0.060	0.041
NO_2	270	0.39	38.63	32.00	28.49
NOx	270	0.39	69.14	41.00	85.29
OXY	69096	100.0	–	–	–
O_3	29047	42.04	44.86	47.00	28.60
PM10	32185	46.58	13.52	9.00	12.60
PXY	69096	100.0	–	–	–
SO_2	40510	58.63	5.39	4.00	4.93
TCH	60656	87.79	1.35	1.31	0.20
TOL	52146	75.47	1.67	0.90	2.32
station	0	0.00	2.81e7	2.81e7	17.61
PM25	50184	72.63	7.72	6.00	6.63
NO	270	0.39	19.89	5.00	40.64
CH4	60656	87.79	1.29	1.23	0.19

calculated too (since for each station sufficiant data on the pollutant is needed). There are 14 variables that have station where nothing was measured for that pollutant in 2018. That pollutants, that were not measured the entire year in at least one station are: BEN, CH4, EBE, MXY, NMHC, OXY, PM10, PM25, PXY, SO_2, TCH, TOL, CO and O_3. The only pollutants that remain, and therefore can be used are NO, NOx and NO_2.

Summary statistics

The only relevant statistic for this plot will be NO_2. Therefore, the summary statistics are only calculated for NO_2 measured in 2018. The results can be found in table 5. This

Table 5: Statistics for NO_2-concentrations per station in 2018

Station	Mean	Median	SD
Arturo Soria	40.28	33.0	27.83
Avda. Ramón y Cajal	43.35	36.0	29.68
Barajas Pueblo	38.13	31.0	26.73
Barrio del Pilar	40.42	32.0	29.36
Casa de Campo	20.57	11.0	22.39
Castellana	42.02	37.0	25.56
Cuatro Caminos	43.72	38.0	29.47
El Pardo	14.57	9.0	14.67
Ensanche de Vallecas	37.87	29.0	28.16
Escuelas Aguirre	57.31	53.5	27.24
Farolillo	36.13	28.0	27.45
Juan Carlos I	27.19	20.0	23.95
Mendez Alvaro	37.79	30.0	27.93
Moratalaz	41.01	34.0	29.14
Parque del Retiro	32.22	26.0	22.74
Plaza Castilla	44.21	39.0	27.74
Pza. Fernández Ladreda	53.73	48.0	35.60
Pza. de España	40.46	37.0	24.45
Pza. del Carmen	48.19	45.0	20.64
Sanchinarro	33.63	25.0	28.43
Tres Olivos	32.97	24.0	26.11
Urb. Embajada	42.74	36.0	29.61
Vallecas	38.77	31.0	27.25
Villaverde	40.01	32.0	30.18

can also be visualized. This can be seen in Out of fig. 12 and table 5 it can be seen that there is a difference in measured values for each station. That difference will be explored to full depth in the final implementation below.

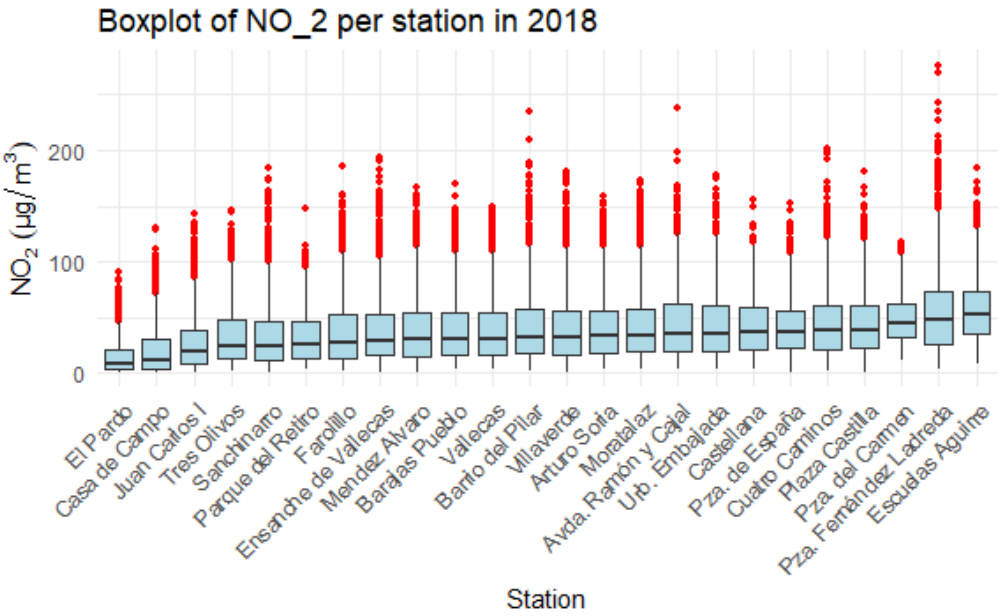


Figure 12: Measured NO_2 values in 2018 for each station

2.2.2 The implementation

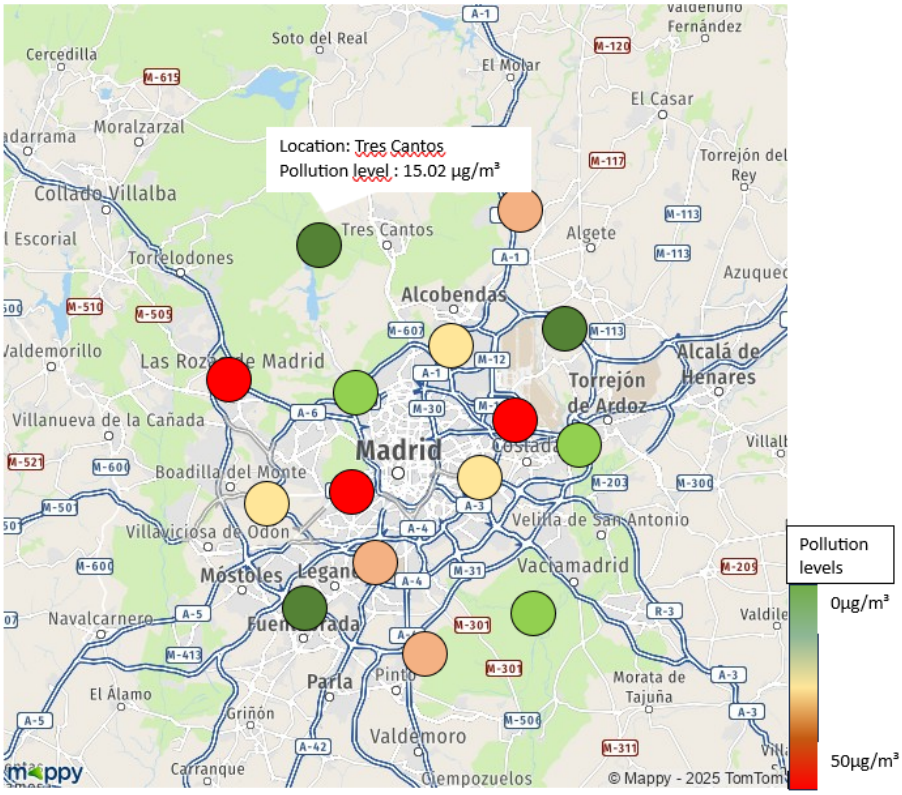


Figure 13: The mock-up of the interactive map of Madrid that is implemented

The final implementation is made using R and can be seen in fig. 14. The interactive map itself can be found in https://courteous-dabbling-tiger-585.vscodeedu.app/. The final implementation that can be seen in fig. 14 is made by using R. fig. 14 has the

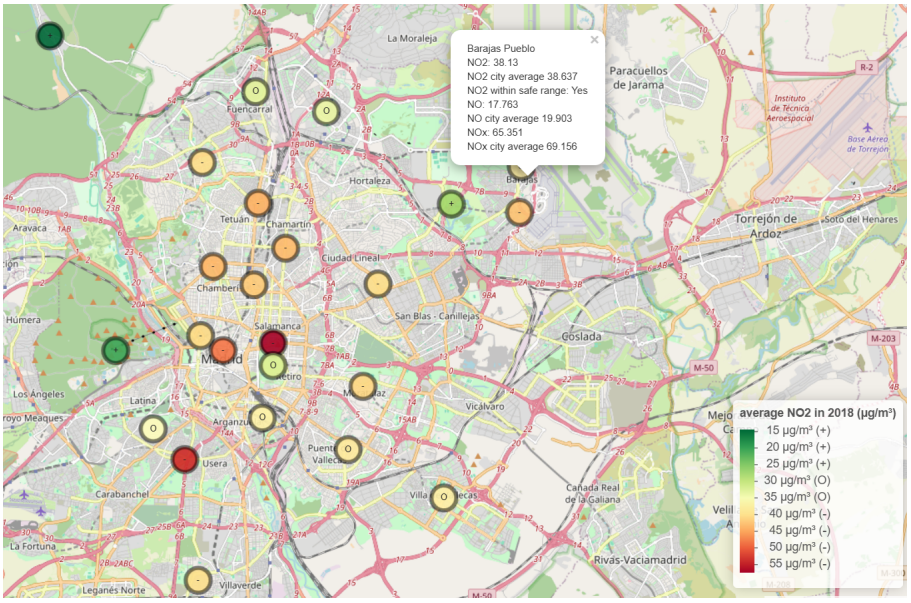


Figure 14: A screenshot of the final implementation of the interactive map of Madrid

following encodings:

- **The data**: the data used for this plot is the data from 2018. All the data is summarized in the annual average value for each station for each pollutant. Due to the lack of information of a lot of different pollutants, it is chosen to only work further with NO_2, since only the data of the different nitrogen pollutants is available for all the different stations. From that data, NO_2 is the best air quality predictor (it is the only nitrogen pollutant that can be found in the air quality index (eq. (1)). The used measure is the yearly averaged measured value per pollutant per station. In the tooltip, average values for the other pollutants are given too, if they are available.

- **Data Processing**: averages are calculated for the different stations. All the measured data of 2018 is summarized in an average value per pollutant per station.

- **Visual Encoding** The graph consists of dots are placed on a map of Madrid (which is the background of the graph). The dots are placed according to the position of the measurement stations. The dots are colored according to the average NO_2 value that is measured in that station in 2018. On the dot, labels are placed (+ for a location with good air quality, O for a location with average air quality and - for a location with values above the threshold of WHO). This is done to make sure that colorblind people can get the plot too.

  By clicking on a specific dot, some relevant information can be found in a tooltip (as well as information about other pollutants and information about city wide averages). This information gives the name of the station, the exact average NO_2 value, and whether that value exceed the threshold for safe NO_2 pollution levels (for NO_2, the levels are safe when they are lower then $40\mu g/m^3$, as stated in [1]). The tooltip gives also information about CO, NO, NOx and SO_2 as well as information about city wide averages for all the relevant pollutants (pollutants that are not NA for a given station).

  - Marks:
    * Points that represent the different stations that measured the air quality across Madrid.
  - Channels:
    * Color (diverging color scale): all the points/stations have a different coloring according to the levels of NO_2 measured in that specific location. The more NO_2, the more red the points are. The less NO_2, the more green the points are.
    * Position (x and y): the points are placed on specific x and y coordinates on the map. Each point refers to a location (which has a given longitude and latitude). The points are plotted on the map of Madrid according to that longitude and latitude.
    * shape: different shapes are used to indicate the best and worst location on the map. The following shapes are used: (+ for a location with goof air quality, O for a location with average air quality and - for a location with values above the threshold of WHO

- **Interaction and interaction idioms**
  - zooming: it is possible to zoom in and out on the map, the get a more detailed view of where the location with good (or bad) air quality are located
  - panning: you can move over the map, to discover new stations, or to get more details about the environment of the station.
  - get tooltip with information: by clicking on a station (a dot on the map), a tooltip with addidtion information about the station, NO_2 values and whether they exceed the norm appears. This tooltip provides more detailed information about a given station, and about other pollutants that where measured by that station.

2.3 Question 3 - Which areas of Madrid improved or worsened more between 2008 and 2018?

Based on the peer feedback and the design phase, We implemented the bar chart visualization (Design 2: Bar Chart with Different Pollutants) to address Question 3. The bar chart was chosen because it effectively compares the changes in individual pollutants (NO_2, PM_{10}, $PM_{2.5}$) and the overall Air Quality Index (AQI) across different stations in Madrid between 2008 and 2018. Peer feedback highlighted the clarity of this visualization in showing pollutant-specific changes (NUF score: 23/30) and suggested adding interactivity, such as sorting and hovering for exact values, which we incorporated into the implementation.

We used Python with Matplotlib to create this visualization, as it allowed for straightforward plotting of grouped bars and a line for the AQI change. The data processing was handled using Pandas to compute the changes in pollutant concentrations and AQI.

2.3.1 Exploratory Data Analysis

To support the visualization for Question 3, an exploratory data analysis (EDA) was conducted to examine changes in pollutant concentrations (NO_2, PM_{10}, $PM_{2.5}$) and the

Air Quality Index (AQI) across Madrid's stations from 2008 to 2018. The data was processed by averaging pollutant levels per station, calculating changes, and computing $\Delta\text{AQI} = 0.333 \times \Delta\text{NO}_2 + 0.222 \times \Delta\text{PM}_{10} + 0.444 \times \Delta\text{PM}_{2.5}$. Missing values were filled with 0, and a fallback dataset ensured all stations were included.

Table 6: Change in Pollutants and AQI by Station (2008–2018)

Station	ΔNO_2 (μg/m^3)	ΔPM_{10} (μg/m^3)	$\Delta\text{PM}_{2.5}$ (μg/m^3)	ΔAQI
Arturo Soria	-9	24	-2	1.44
Calle Farias	26	-4	-1	7.33
Cuatro Caminos	-17	27	2	1.22
Escuelas Aguirre	-38	26	-3	-8.21
Plaza Castilla	21	3	-3	6.33
Plaza de España	-21	1	-10	-11.21
Plaza del Carmen	52	-6	4	17.76
Plaza Fernando	-8	-6	-1	-4.44
Tres Olivos	39	-7	-1	10.99

Key Insights

- NO_2: Ranges from -38 (Escuelas Aguirre) to 52 (Plaza del Carmen); 5 stations improved.
- PM_{10}: Mostly increased, with Cuatro Caminos at 27; 3 stations improved.
- $\text{PM}_{2.5}$: Mostly improved, with Plaza de España at -10; 2 stations worsened.
- AQI: Ranges from -11.21 (Plaza de España) to 17.76 (Plaza del Carmen); 5 stations improved.
- Trend: NO_2 drives AQI changes, with Plaza del Carmen showing the worst decline.

2.3.2 The implementation

The Intended Design

The intended design was a bar chart where each station has bars for NO_2, PM_{10}, and $\text{PM}_{2.5}$ changes (in μg/m^3), with a line showing the combined AQI change. The visual encoding included:

- **Position:** Stations on the x-axis, changes on the y-axis (up means worse, down means better).
- **Color:** Different colors for each pollutant (blue for NO_2, black for PM_{10}, green for $\text{PM}_{2.5}$) and red for the AQI line.
- **Length:** The length of the bars shows the magnitude of change.
- **Interaction:** Bars can be sorted to highlight stations with the largest changes, and hovering over a bar shows the exact change value.

This design was sketched in Figure 4 of the design phase report, aiming to compare how different pollutants changed at each station and identify areas with significant improvements or worsening in air quality.

The Actual Design That Was Implemented

We implemented the bar chart using Python, Pandas, and Matplotlib. The data was processed by filtering the 2008 and 2018 measurements, calculating the changes in NO_2, PM_{10}, and $PM_{2.5}$ concentrations for each station, and computing the AQI change using the formula:

$$\Delta AQI = 0.333 \times \Delta NO_2 + 0.222 \times \Delta PM_{10} + 0.444 \times \Delta PM_{2.5} \quad (2)$$

The final visualization matches the intended design closely and is shown in the screenshot below (fig. 15). The interactive version can be found in https://gold-reliable-shark-501.vscodeedu.app/.

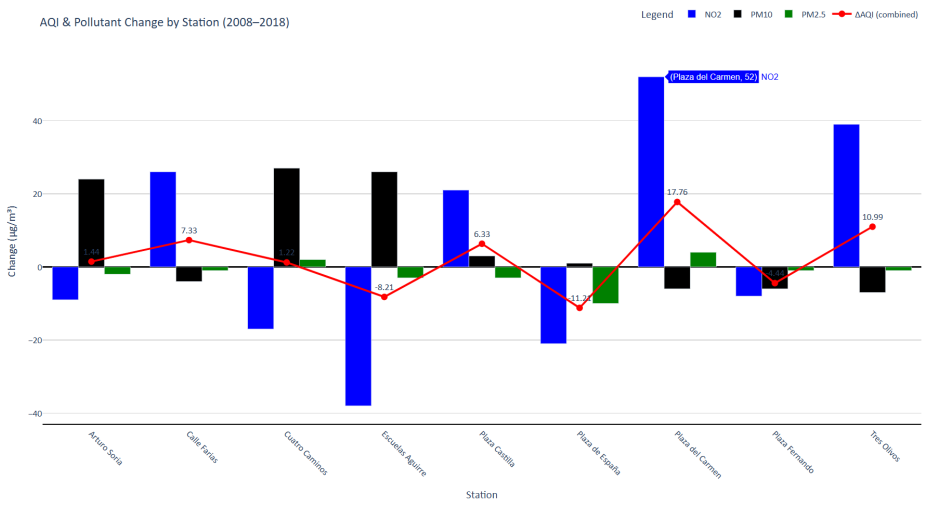


Figure 15: Screenshot of the Implemented Bar Chart: "AQI & Pollutant Change by Station (2008-2018)" showing changes in NO_2 (blue), PM_{10} (black), $PM_{2.5}$ (green), and AQI (red line) for stations like Arturo Soria, Calle Farias, etc.

Visual Encoding:

- **Marks:** Bars for each pollutant and a line with markers for AQI.
- **Channels:**
  - **X-axis:** Stations.
  - **Y-axis:** Change in concentration ($\mu g/m^3$).
  - **Color:** Blue for NO_2, black for PM_{10}, green for $PM_{2.5}$, red for AQI.
  - **Length:** The height of the bars represents the magnitude of change.
- **Data Processing:** The data was sourced from the Madrid dataset (madrid_2008.csv and madrid_2018.csv), merged by station ID, and changes were calculated as the difference between 2018 and 2008 values.

Description of the Interactions

The implemented bar chart includes the following interactions:

- **Sorting:** Users can sort the stations by the magnitude of AQI change (largest to smallest) by clicking a button (implemented using Matplotlib's event handling in a Jupyter notebook environment).
- **Hover Tooltips:** Hovering over a bar displays a tooltip with the exact change value (e.g., "NO_2: 7.33 $\mu g/m^3$" at Calle Farias), providing precise data for analysis.
- **Zoom & Pan:** Users can zoom into specific sections of the chart or pan to focus on particular stations, aiding in detailed comparisons.

These interactions enhance the usability of the visualization, allowing policymakers to quickly identify stations with the most significant air quality changes and understand the contributions of individual pollutants.

3 Link to the Code for the Visualizations

The code for this visualization is available in our group's GitHub repository:
https://github.com/kiflomhailu/vds2425-visualizations.git

4 Link to the YouTube Video

A 5-minute video demonstrating the functionality is available on YouTube:
https://www.youtube.com/watch?v=TSYdxUiApiI

Video Description:
A quick overview of air quality conditions in Madrid by Group 6. Stay informed about pollution levels and their impact in Madrid. This video tells the story behind the data using graphs we built with Python and R.