Report on NYC Air Pollution Visualization

Haotong Wu hw2933

April 2024

1 Introduction

Air pollution in urban areas poses significant risks to public health, environmental stability and overall quality of life. In densely populated, industrialized New York, understanding and monitoring air quality is critical for policymakers, researchers and the public.
In this project, I visualized and analyzed air pollution in New York City. The visual analysis is built on a dataset of reliable air quality measurement bases, including fine particulate matter (PM2.5), nitrogen dioxide (NO2), and ozone (O3). And I deployed an application website that utilizes Dash and Plotly, which provides interactive and intuitive visualizations that reveal spatial and temporal trends in air pollutants.

2 Data Processing

2.1 Dataset Info

The New York City Air Quality dataset is a comprehensive collection of air quality surveillance data, crucial for assessing the environmental health within urban spaces of NYC. The dataset includes measurements of various pollutants, revealing the disparities in pollutant emissions and exposure across different neighborhoods. It offers valuable insights into the prevalence of respiratory and cardiovascular diseases, cancers, and premature deaths linked to air quality, aiding in the characterization of air quality and associated health outcomes over time and across NYC geographies.

2.2 Data Cleaning

The data cleansing phase is essential to improve the usability of the dataset. First, I read the original dataset and found that some of the rows contained missing values. In order to maintain the integrity of the data, we removed rows with such inconsistencies. Redundant columns with no analytical value, especially the 'Message' column, were also discarded.

I normalized the "Start Date" column to a consistent date format to facilitate time analysis. And the complexity in the "Time Period" column was resolved by dividing it into "Season" and "Year" sections, thus simplifying the analysis of longitudinal and seasonal trends. This phase culminated in the merging of the cleaned data into a new file, which provided a solid foundation for the subsequent analysis process.

	Unique ID	Indicator ID	Name	Measure	Measure Info	Geo Type Name	Geo Join ID	Geo Place Name	Time Period	Start_Date	Data Value	Message
0	172653	375	Nitrogen dioxide (NO2)	Mean	ppb	UHF34	203	Bedford Stuyvesant - Crown Heights	Annual Average 2011	12/01/2010	25.30	NaN
1	172585	375	Nitrogen dioxide (NO2)	Mean	ppb	UHF34	203	Bedford Stuyvesant - Crown Heights	Annual Average 2009	12/01/2008	26.93	NaN
2	336637	375	Nitrogen dioxide (NO2)	Mean	ppb	UHF34	204	East New York	Annual Average 2015	01/01/2015	19.09	NaN
3	336622	375	Nitrogen dioxide (NO2)	Mean	ppb	UHF34	103	Fordham - Bronx Pk	Annual Average 2015	01/01/2015	19.76	NaN
4	172582	375	Nitrogen dioxide (NO2)	Mean	ppb	UHF34	104	Pelham - Throgs Neck	Annual Average 2009	12/01/2008	22.83	NaN

Figure 1: original dataset

	Unique ID	Indicator ID	Name	Measure	Measure Info	Geo Type Name	Geo Join ID	Geo Place Name	Start_Date	Data Value	Season	Year
0	172653	375	Nitrogen dioxide (NO2)	Mean	ppb	UHF34	203	Bedford Stuyvesant - Crown Heights	2010-12-01	25.30	Annual	2011
1	172585	375	Nitrogen dioxide (NO2)	Mean	ppb	UHF34	203	Bedford Stuyvesant - Crown Heights	2008-12-01	26.93	Annual	2009
2	336637	375	Nitrogen dioxide (NO2)	Mean	ppb	UHF34	204	East New York	2015-01-01	19.09	Annual	2015
3	336622	375	Nitrogen dioxide (NO2)	Mean	ppb	UHF34	103	Fordham - Bronx Pk	2015-01-01	19.76	Annual	2015
4	172582	375	Nitrogen dioxide (NO2)	Mean	ppb	UHF34	104	Pelham - Throgs Neck	2008-12-01	22.83	Annual	2009

Figure 2: cleaned dataset

3 Code Application

These functions collectively enable a comprehensive analysis of air pollution trends in New York City, providing insights into spatial and temporal variations. By integrating these visualizations into an interactive web application, users can explore and understand the complex patterns of urban air pollution.

3.1 Data Loading Functions

The program starts by loading the necessary data sets using Pandas and GeoPandas. The following data are loaded:

- Cleaned air pollution data from `clean_data.csv`.
- New York City borough boundaries from `nyc_boroughs.geojson`.
- UHF42 health regions from `UHF42.geo.json`.

3.2 Visualization Functions

The core functionality of the application lies in its ability to visualize average pollution levels across different regions of New York City.

3.2.1 plot_average_levels

This function generates a choropleth map showing the average pollution levels for a specified pollutant across the boroughs of New York City.

3.2.2 plot_average_levels_by_uhf42

Similar to `plot_average_levels`, this function visualizes pollution levels by UHF42 regions, providing a more granular view:

3.2.3 plot_average_levels_by_season

This function creates a bar chart comparing average pollution levels across different seasons for each borough.

3.2.4 plot_average_levels_by_year

This function displays a bar chart of the different results across New York City.

3.2.5 plot_average_levels_by_year_borough

This function generates a line chart showing how average pollution levels have changed over time within each borough.

3.3 Interactive Web Application

For the web application, I use Dash, an interactive web application to built, which includes:

- Dropdown menus for selecting the type of pollution data and the geographic area to display.
- A visual representation of the data through choropleth maps, which are updated based on the user's selections.

The application employs callback functions to dynamically update the content and charts based on user interaction, enhancing the user experience and providing insightful data visualizations.

4 Analysis and Example Pages

4.1 Example 1

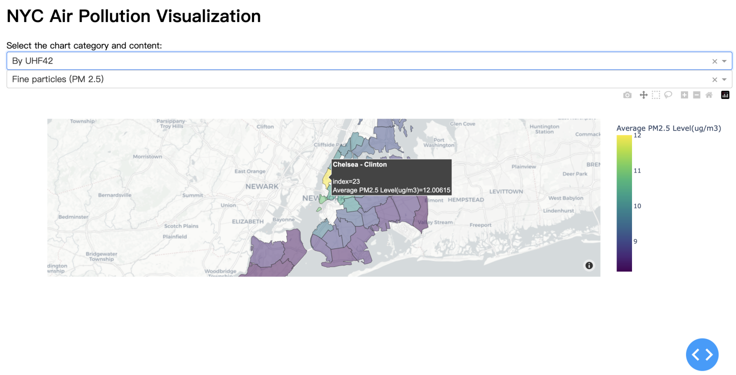

Figure 3: UTF42

The first example shows air pollution levels measured as PM2.5 concentrations in the UHF42 area. The color-coded map provides a direct visualization of areas with higher pollution levels. For example, the Chelsea-Clinton area has higher average PM2.5 levels than surrounding areas, which is clearly colored on the scale. Hovering over areas to view detailed data including PM2.5 averages adds an interactive element that helps users better view specific data.

4.2 Example 2

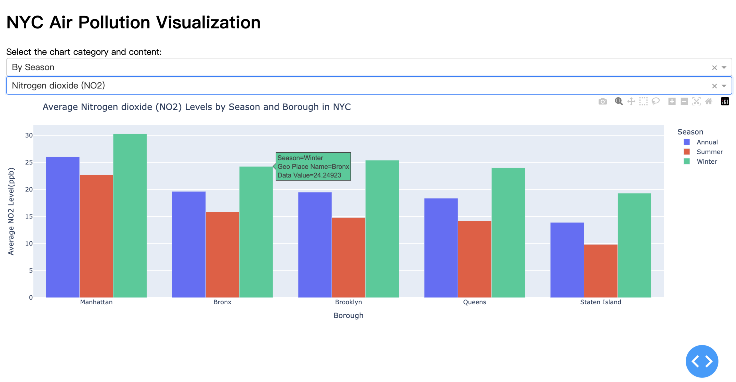

Figure 4: season

The second example converts the map into a bar chart showing nitrogen dioxide (NO2) levels by season for each borough. This visualization helps to analyze seasonal trends and identify the times of the year when air pollution is most severe. The interactive bar charts and seasonally differentiated colors provide a clear visual comparison between boroughs, highlighting which boroughs have higher pollution levels during which season.

4.3 More Examples

These are some other examples that show the different categories of my diagrams.

Figure 5: results

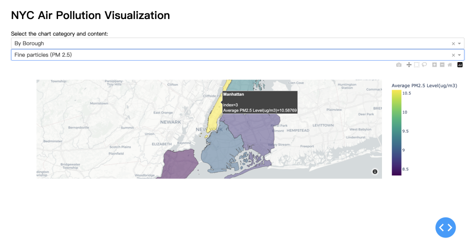

Figure 6: borough

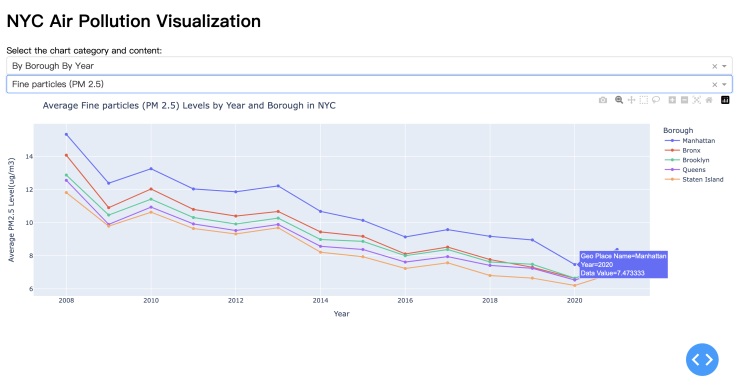

Figure 7: borough by years

5 Conclusion

This report presents a comprehensive picture of air pollution in New York City through a series of interactive visualizations. By utilizing the capabilities of Dash and Plotly, the application serves as a tool for public engagement and detailed analysis, enabling users to explore various dimensions of air quality data. Future work could extend this project by integrating more detailed data, exploring correlations with public health outcomes, and considering the impact of changes over time and related policies. In addition, the framework established

by this project could be applied to other urban areas by simply replacing the dataset to support broader applications for improving urban air quality and public health on a global scale.
Ultimately, this visualization project could provide builders with an easier way to take informed action to build a cleaner, healthier New York City.