Data Visualizations of Air Quality and Mortality Indexes from 1990 to 2020

Pol Monné Parera
polmonneparera@lewisu.edu
DATA-53000, Fall 2025
Data Visualization
Lewis University

I. Introduction

Since the begging of the industrial era in the late 18th century and the constant increase in the use of non-renewable energy sources, humanity has been releasing a staggering number of solid particles to the atmosphere which affect air quality. The World Health Organization (WHO) first addressed this issue in 1958 with a publication on air pollution and its effects on health, which was later followed with a vast number of papers and studies investigating this area and advocating for pollution reduction polices to be implemented. The main motivation of the study comes from the fact that the World Health Organization claims that 99% of the global population breaths air that exceeds their guideline limits [1].

This report aims to visualize distribution of air quality worldwide from 1990 to 2020, and if this distribution can be correlated to mortality rates per country. The study also seeks to help determine overall trends in data that could potentially be relevant to take actions in healthcare and environmental areas per country.

II. Data Description

The dataset used for the analysis previously mentioned consisted of a collection of data from a variety of sources, which merged data on both air quality indicators and mortality rates per country on a fixed time range (1990-2020). The data from the different sources has been preprocessed to follow the same timeline and clean undesired outliers or null values.

The first data source consists of PM 2.5 air pollution mean annual exposure per country [2] and can be found on the world bank group website [3]. The second dataset, similarly to the first, can also be found in the world bank group website [3] and consists of data on the mortality rates per country [4]. Finally, to study air quality data on a more specific geospatial dimension (cities) the third dataset consists of information extracted from the Kaggle *Global Air Pollution Dataset* [5]. This last dataset contains quantitative data on different pollution indicators per city as well as categorical data defining the hazard range of air quality per instance. All information was extracted on given dates not specified around November 2022. It is important to note that the Air Quality Index (AQI) studied from this dataset is not the same as the PM 2.5 air pollution mean, as the first one englobes many more pollution indicators such as ozone, CO or NO2.

The data studied consists of around 8000 instances for the bank group datasets, they don't have the exact same number of instances as some smaller island countries had no data collection regarding air quality or mortality rates and where therefore dropped. The Kaggle datasets consists of 23,463 rows with each row representing a different city. The attributes studied are depicted in Table I for the bank group datasets and in Table II for the Kaggle one.

TABLE I. Bank Group Data Attributes

Attribute	Type	Example Value	Description
Year	Categorical	2005	Date (Yearly frequency).
Country Name	Categorical	United States	Name of the country.
Country Code	Categorical	ABW	International code of the country.
Air Quality	Numeric	25.1114	PM 2.5 air pollution mean annual exposure.
Mortality Rate	Numeric	6.6770	Mortality rate (deaths per 1000 habitants).

TABLE II. KAGGLE DATA ATTRIBUTES

Attribute	Type	Example Value	Description
Country	Categorical	Brazil	Name of the country where the city is.
City	Categorical	Presidente Dutra	Name of the city.
AQI Index	Numeric	41	General air quality index.
AQI category	Categorical	Good	Air quality categorization.

III. METHODOLOGY AND RESULTS

To conduct the desired visual exploratory analysis of the dataset all the information was loaded into Tableau and was linked accordingly to be able to work with all the information. The visualizations developed studied the distributions of data with geospatial visualizations of the entire globe.

Fig. 1 shows the average air quality (PM 2.5 air pollution mean) of each country from 1990 to 2020 and allows for pre-attentive detection of outliers and trends. In Fig. 1 we can easily observe how the country with the highest average pollution index is in the upper center of the African continent (to be precise, that country is Niger), it can also be observed how nearby countries share increased air pollution rates. This is a desertic area that might have this higher number of particles in the air in the form of sand. Another cluster of high polluted countries can also be observed in the Asian continent, with India, Pakistan and China among others appearing in a darker orange color. Fig. 1 also allows clear visualization of the countries with the least contaminated air, with Australia, the North American continent and the Northern and Western side of Europe showing a lighter tonality of orange. Finally, Fig. 1 allows the observation of an outlier in the South American territory with Bolivia and Peru showing a much darker shade than their neighboring countries.

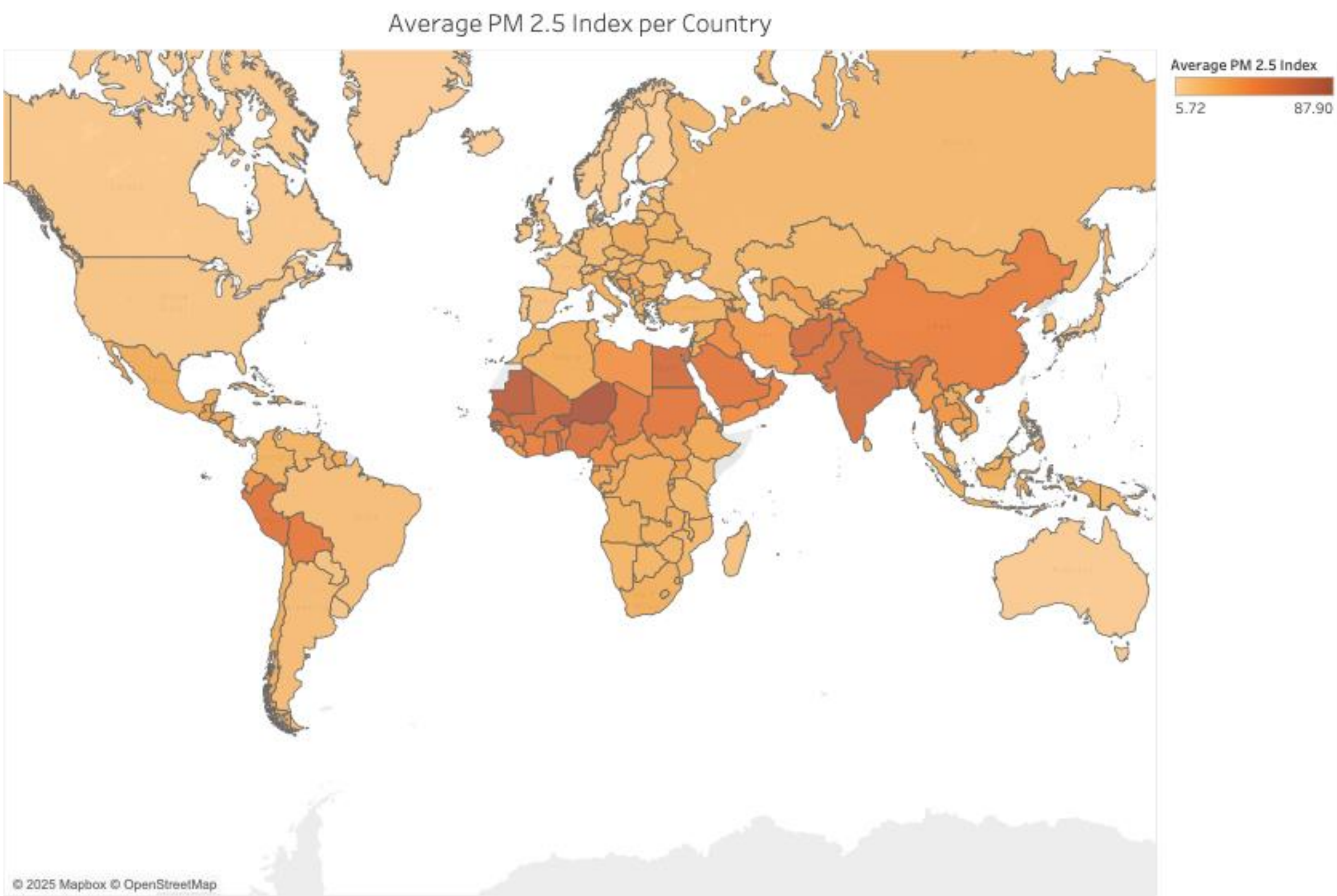

Fig. 1 Average PM 2.5 Index per Country

In Fig. 2 the mortality rate of each country can be studied with another choropleth map, colored with black because of its arbitrary symbology with death. It can be easily observed how there is a cluster of high mortality rate in the middle to lower parts of Africa with South Sudan displaying the highest mortality rate in the world. It is important to note that, even though this area was also studied in the previous paragraph as an area of high air pollution, no causation can be established. Some of these areas of the world have suffered a lot of violent incidents during the past century due to their position of highly valuable resources in their land, which could have been a motive for an increased death rate and maybe even the pollution observed. Fig. 2 also shows Russia and its European neighbors to have an increased mortality rate, which again might be due to past armed conflicts between the countries for the time studied.

The other countries of the world all seem to have a similar value of deaths per 1000 habitants with most of them being at less than 10 deaths. In Fig. 2 it is important to notice that countries in the Arabian Peninsula appear to have the lowest mortality rate in the world, which could be further studied.

Fig. 2 Mortality Rate per Country

Finally, Fig. 3 displays a selected variety of cities and their air pollution statues via a point distribution map, with dot size being determined by the air quality index (AQI) value. The selected cities in Fig. 3 correspond to those classified in to have a more perilous health status related to their AQI index, excluding the cities with a good or moderate classification. The color of the dots in the map is also determined by this classification.

Fig. 3 allows a clear visualization of the highest polluted cities worldwide, with two clear aggregations/clusters, one in India and neighbor countries and the other in China. Nevertheless, the Indian cluster seems to be the one with the worst ratio worldwide as it contains the biggest aggregation of cities with a hazardous classification, while the Chinese one is mainly composed of unhealthy classified cities with only three being classified as hazardous. Fig. 3 also allows a visualization of at least one hazardous classified city in each continent except for Europe and Australia. Nevertheless, it is interesting to notice that only Europe has their highest classification as an unhealthy city with a few of this in a cluster on the French territory.

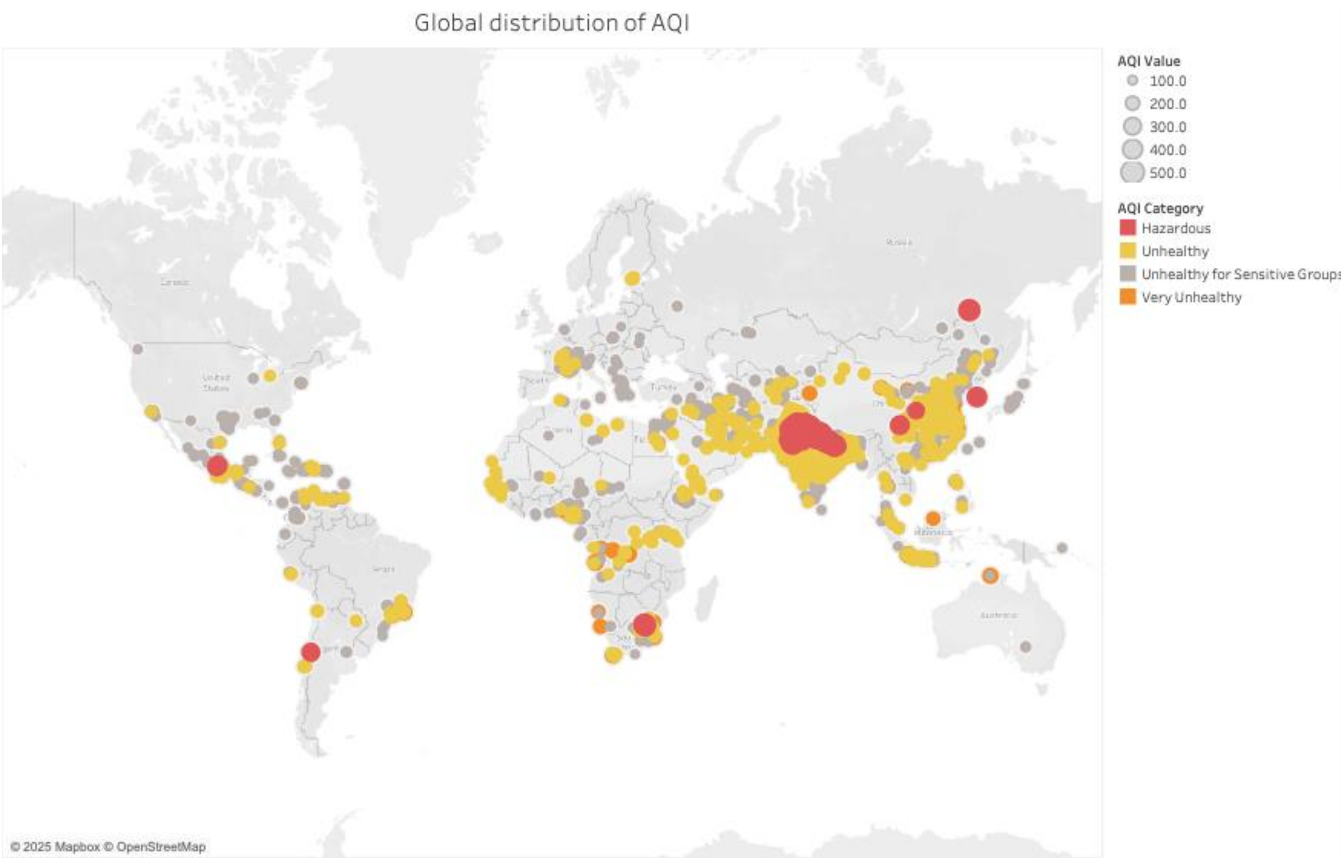

Fig. 3 Global Distribution of AQI

IV. Discussion

In the previous section (Section III), several geospatial visualizations were used to explore the relationships and patterns between air quality indicators and mortality indexes. By looking at the different images clusters and outliers become easily visible to the naked eye, which may uncover hidden trends in the data.

In Fig. 1 it was clear how the mid African and Asian territories had the worse PM 2.5 indices, while also observing a clear outlier among south American countries with a spike in this type of contamination in the countries of Bolivia and Peru. Fig. 1 also showed how North America, Australia and some European Territories were the least affected countries by this type of contamination worldwide.

Fig. 2 allowed for the visualization of two clusters of higher mortality rates, one in Russia and some of its neighboring European territories and the other one in Africa. It is interesting to detect how the African area detected to have a higher PM 2.5 index also seems to have a higher mortality rate.

Finally, Fig. 3 allows a clear visualization of two clusters of highly polluted environments in India and China. With the Indian one having a high density of hazardous classified cities while the Chinese cluster is mostly classified with unhealthy ones. This last figure also displays Europe as the only continent with unhealthy as its worse air quality index category.

V. DISCUSSION OF THE USAGE OF THIS WEEK'S CONTENT

This report uses two of the covered geospatial visualization on this week's material, these being a choropleth and a point distribution map. The first one is great for the display of standardized data like the particles per cubic meter or the mortality rate of each country. With the use of well-thought color encoding the visualizations are great at rapidly showing patterns and outliers in the data. On the other hand, the point distribution map has been used for the AQI values as this kind of visualization works well on both raw and standardized data but, most importantly, can show a graphic distribution of many related phenomena.

VI. CONCLUSIONS

This report displayed some clear clustering of pollution in the global environment, with Africa and Asia being the continents more affected by this kind of contamination. The report also displays some correlation between the air quality of certain African countries and their mortality rate, but this must be studied carefully before drawing any conclusions as lots of factors could be affecting these indices and causation is never implied. It is important to notice also that mortality rates do not only involve deaths derived from air pollution, but traffic, armed conflicts and the economic situation of the country among others are also factors that can have an effect in this indicator. It is also important to consider how air pollution may not influence the overall health of the population today but it may do so in the foreseeable future.

REFERENCES

[1] World Health Organization. (n.d.). Air pollution. Retrieved November 13, 2025, from https://www.who.int/health-topics/air-pollution#tab=tab_1

[2] The World Bank. (n.d.). World Bank Open Data. Retrieved November 13, 2025, from https://data.worldbank.org

[3] The World Bank. (2024, June 6). PM2.5 air pollution, mean annual exposure (micrograms per cubic meter) [Indicator EN.ATM.PM25.MC.M3]. World Bank Open Data. Retrieved November 13, 2025, from https://data.worldbank.org/indicator/EN.ATM.PM25.MC.M3

[4] The World Bank. (n.d.). Crude death rate (per 1,000 people) [Indicator SP.DYN.CDRT.IN]. World Bank Open Data. Retrieved November 13, 2025, from https://data.worldbank.org/indicator/SP.DYN.CDRT.IN

[5] Hasib Al Muzdadid. (2022, November 2). Global Air Pollution Dataset [Data set]. Kaggle. Retrieved November 13, 2025, from https://www.kaggle.com/datasets/hasibalmuzdadid/global-air-pollution-dataset/data