Basic wrangling tasks and questions

Ques 1.1: How many GADM2 regions are present in India?
Ans: Number of GADM2 regions in India: 684

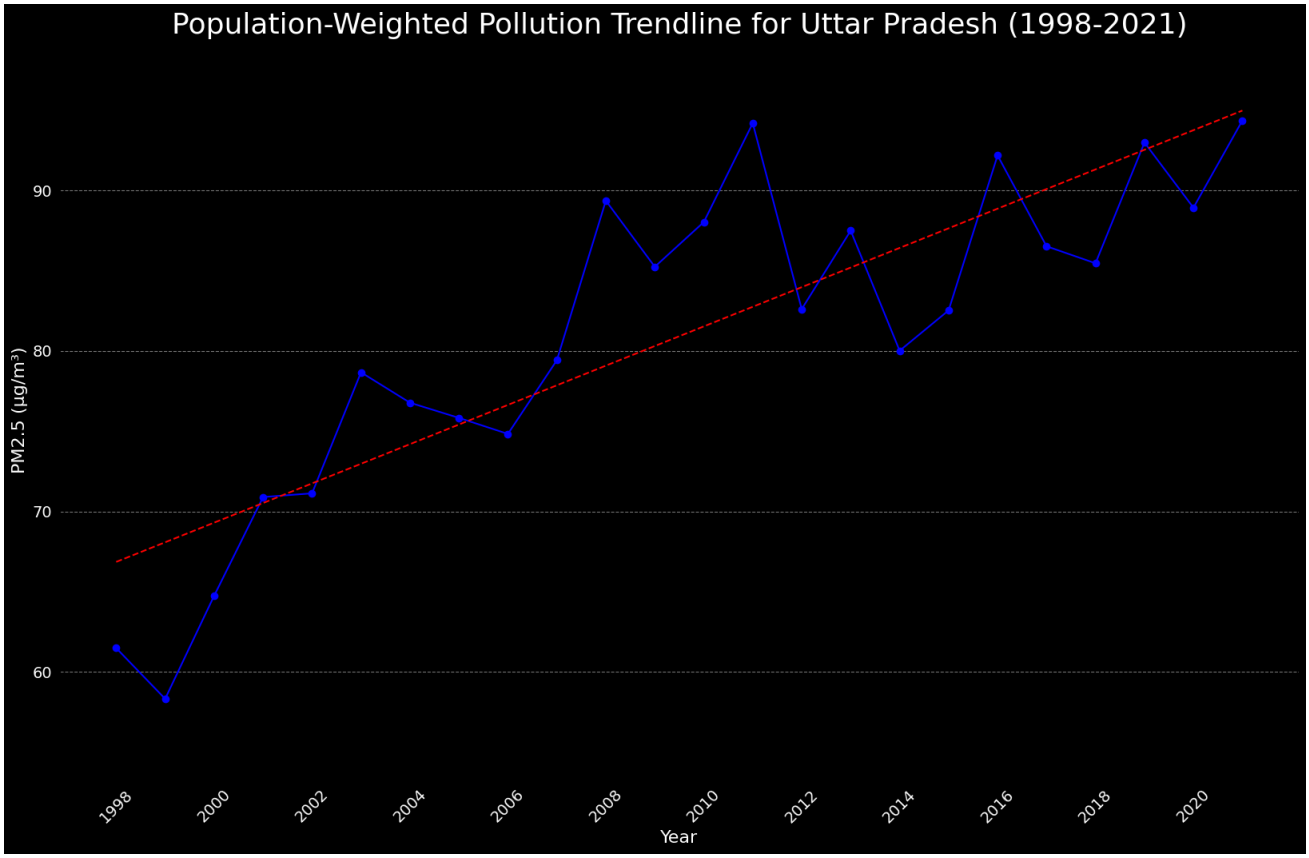
Ques 1.2: Calculate the population-weighted pollution average of all years at the country (GADM0) level. Save the country-level file as a CSV, and What are the 10 most polluted countries in 2021?
Ans: These countries are arranged in descending order.

1. Bangladesh
2. India
3. Nepal
4. Pakistan
5. Mongolia
6. Myanmar
7. Democratic Republic of the Congo
8. Republic of the Congo
9. Rwanda
10. Burundi

Ques 1.3: What was the most polluted GADM2 region in the world in 1998, 2005, and 2021?

1. The most polluted region in 1998:
   Country: India
   Region: Unnao
   PM2.5 Level: 78.55
2. The most polluted region in 2005:
   Country: India
   Region: NCT of Delhi
   PM2.5 Level: 98.75
3. The most polluted region in 2021:
   Country: India
   Region: NCT of Delhi
   PM2.5 Level: 126.51

Ques 1.4: Plot a population-weighted pollution average trendline plot for Uttar Pradesh from 1998 to 2021. Save this plot as a high-quality PNG file.
Ans:

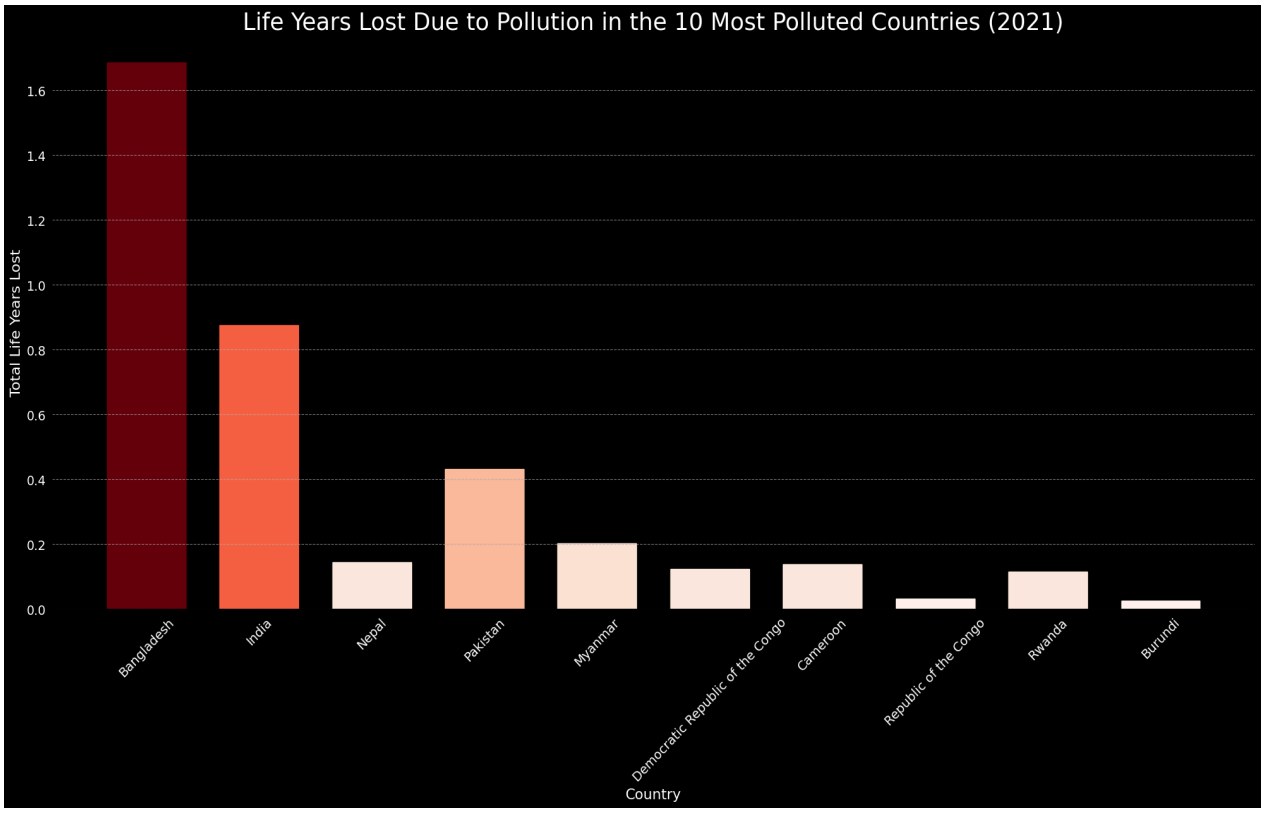


What can be improved:

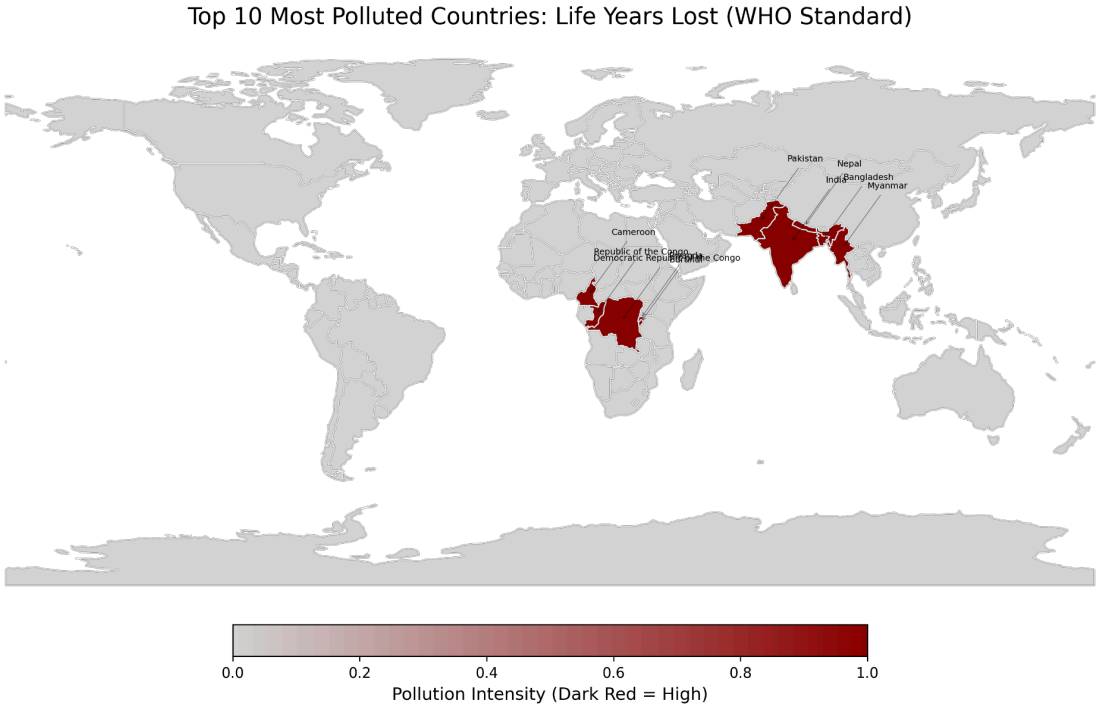
1. I can use an interactive cursor in the current Matplotlib visualization for real-time analysis and understanding
2. Some additional libraries I can use, such as Plotly or Bokeh, for zooming, hovering, and filtering data dynamically.

Ques 2.1: Plot a bar graph for the life years lost relative to the WHO guideline in the 10 most polluted countries in the world and plot them on a global country-level map. For the map, the 10 most polluted country boundaries should be filled in with "dark red," and the rest of the map should be grayed out. Save both the bar graph and the map as high-quality PNG files.

Ans: Bar Graph :


Life Years Lost Due to Pollution in the 10 Most Polluted Countries (2021)
Total Life Years Lost
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1.6
Bangladesh
India
Nepal
Pakistan
Myanmar
Democratic Republic of the Congo
Cameroon
Republic of the Congo
Rwanda
Burundi
Country

World Map :


Top 10 Most Polluted Countries: Life Years Lost (WHO Standard)
Pakistan
Nepal
India
Bangladesh
Myanmar
Cameroon
Republic of the Congo
Democratic Republic of the Congo
Burundi
0.0
0.2
0.4
0.6
0.8
1.0
Pollution Intensity (Dark Red = High)

Some modifications can be made to improve this :

1. Since a single color is specified, introducing a color bar may enhance clarity and interpretation.
2. Label key geographic areas to improve readability and contextual understanding.
3. Use arrows to label the most polluted areas precisely to ensure clear visualization.
4. I can enable region/country selection so users can explore data relevant to their interests.
5. I can use Dash/Streamlit to create interactive dashboards with sliders to see changes over time.

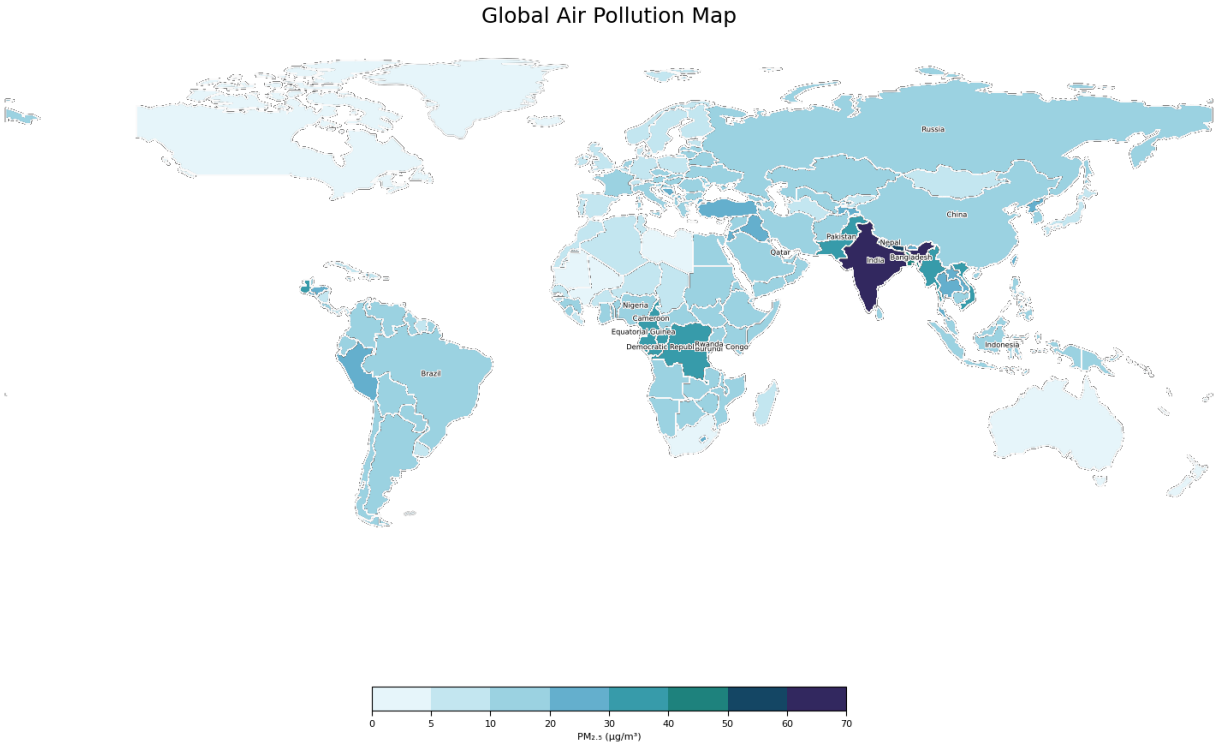
Ques 2.2: Create a potential gain in life expectancy (relative to the WHO guideline) map of Eastern vs. Western Europe at GADM level 2 and save it as a high-quality PDF.

- **The plot should be on the AQLI "Potential gain in life expectancy" color scale. Visit AQLI website Index page > See legend for "Potential gain in life expectancy" and infer "exact" colors from that.**
- **Define Eastern and Western Europe using a credible source and explicitly mention the reference.**
- **Incorporate annotations, labels, or text boxes to enhance clarity and explain key aspects of the visualization.**
- **Ensure the visualization effectively communicates the intended insights while maintaining accuracy and readability.**

Ans:

1. I first selected the continent column and filtered the data for the European region to plot the graph. However, I couldn't derive a clear conclusion from the visualization.
2. To enhance the analysis, I considered using a shapefile to map the regions accurately.
3. I referred to the UN website to categorize Europe into East and West regions, defining them as separate lists for better geographical representation.
4. For the final plot, I will merge the GADM Level 2 shapefile with AQLI life expectancy gain data.
5. In the end, I will plot the Eastern Europe vs. Western Europe on the same map.

Ques 2.3 Look at the AQLI website > switch to the Air Pollution tab > plot a static version of the global pollution map you see there, in those "exact" same colors. Export it as a high-quality (320 dpi) SVG file.



What can we improve?

1. To highlight regions with greater and lesser impact on pollution in your map
2. Improve label readability by adjusting font size and contrast for country names.
3. Add interactive tooltips to show country names and exact PM2.5 values on hover.
4. Use dynamic filtering to highlight specific continents or pollution levels.

4. Verbal reasoning and writing

Please read the following excerpt from the AQLI Annual Update 2024 carefully and summarize your key takeaways in three clearly written bullet points. Each bullet should not exceed 30 words.

1. Most of the highly polluted countries have poor air quality, leading to a reduction of 2.7 years in life expectancy compared to cleaner regions. Additionally, many countries fail to meet air quality standards.
2. To effectively control pollution, countries must implement and enforce specific guidelines and policies related to weather, geography, and industries. Without strong regulations, pollution levels will remain high.
3. If polluted regions met their air quality standards, the world could gain 3 billion life-years, leading to significant health benefits. Achieving this goal requires global efforts, including adherence to the Paris Agreement 2030 in most countries.