R Project Summary:
Analyzing Air Quality Index (AQI) and Population Data

Introduction

Given the increasing frequency of air quality alerts during the summer, I became concerned about the global air quality situation. To gain deeper insights into this issue, I delved into AQI data spanning the past year from various regions worldwide. By analyzing this data, I aimed to gain a better understanding of how population and air quality are intertwined and whether specific factors contribute to AQI fluctuations.

In this R project, my primary objective was to investigate the connection between Air Quality Index (AQI) and population data across different countries. I aimed to uncover potential relationships between population size and AQI levels, while also exploring the possible impact of geographical location and weather patterns on AQI variations.

Data Import and Preparation

To begin my analysis, I imported two datasets, namely AQI_Data.csv and World_Pop_2022.csv, into Google Sheets. For the World Population data, I focused solely on the 2022 population, country, and continent information. To streamline the data, I removed unnecessary columns, which I saved for future analysis. On the other hand, I retained all the data in the AQI dataset, as it was vital for my analysis.

The first dataset provided daily AQI values for different countries, while the second dataset contained population data specifically for the year 2022. Once I had made these adjustments, I imported the updated datasets into RStudio.

With the necessary packages, including tidyverse, lubridate, and ggplot2, loaded, I merged the two datasets based on the "Country" column. This merging process enabled me to compare AQI and population data for each country on the same date, setting the foundation for my subsequent analysis.

Data Analysis

To begin the analysis, I checked the structure of the merged dataset and verified that there were no missing values. Next, I identified the most recent date available in the dataset, which happened to be 7/27/2023.

I also used the distinct() function to remove any duplicate dates and keep only the first occurrence of each country on each date. This ensured that the dataset was clean and ready for further analysis.

Findings

I found that there was no significant correlation between population and AQI. Geographic location and weather patterns appeared to be more influential factors in AQI variation.

- **The Country with the Worst AQI Overall: India**

  Based on the available data, India had the highest overall AQI levels.

- **The country with the Best AQI Overall: Guam**

  Guam had the lowest overall AQI levels among all countries in the dataset.

Visualizations

To better understand the AQI trends for specific countries, I created line plots for China, the USA, and Gibraltar. These line plots showed how AQI levels changed over time for each country.

For a broader view of AQI and population data, I used ggplot2 to create a scatter plot comparing the average AQI with the population for each country. However, due to the large number of data points, the scatter plot was not visually appealing and challenging to read.

Interactive Visualization

To address the challenges with the scatter plot, I utilized the Plotly package to create an interactive scatter plot. The interactive plot allowed me to visualize AQI and population data for each country more effectively. Each country was represented as a colored dot, with the dot size corresponding to its population size. Hovering over the dots displayed country-specific information.

Solutions?

Addressing air quality issues on a global scale is undoubtedly a complex and challenging endeavor. The Air Quality Index (AQI) is influenced by a multitude of factors, and achieving significant improvements would necessitate numerous changes and concerted efforts.

However, with the dissemination of more accessible information to the public and a strong focus on education, there is hope for progress. Empowering individuals with knowledge about air quality and its impact on health can lead to increased awareness and more active participation in seeking solutions.

While it is uncertain whether I will witness a profound impact in my lifetime, the collective dedication to tackling this issue gives me reason for optimism. Effectively improving AQI requires a comprehensive approach, involving governmental regulations, technological innovations, heightened public consciousness, and collaborative actions across all sectors of society.

Ultimately, by working together as governments, industries, communities, and individuals, we can make strides toward cleaner air and better public health, making the aspiration of a healthier future a tangible possibility.

Future Exploration

This project provides valuable insights into the relationship between AQI and population data for different countries. However, there are numerous opportunities for further exploration. For a more comprehensive analysis, additional factors such as weather patterns, CO2 levels, and geographical locations could be incorporated. Moreover, creating a map visualization to display the geographical locations of each country and their AQI levels would enhance the understanding of global air quality patterns.

Conclusion

This project offers valuable insights into the intricate relationship between AQI and population data, providing a foundation for future research and exploration of the factors influencing air quality on a global scale. By examining the data, we have begun to unravel the complexities of AQI variations and how they may be linked to population dynamics. This work serves as a stepping stone for continued investigations, opening the door to a deeper understanding of the dynamics at play in the realm of air quality worldwide.