

CSC111 Project 2 Written Report: PANDEMIC PATTERNS: NAVIGATING THE SPREAD OF COVID-19

Aryan Aneja
Navnoor Singh Mahal
Vishwesh Manishbhai Patel
Aarav Tandon

April 4, 2024

1 Introduction

For our project, we are going to analyze the effect of the COVID-19 pandemic across geographic regions. The pandemic has significantly impacted most communities worldwide, and through our project, we aim to highlight this spread of the virus across these communities. COVID-19 was a highly contagious respiratory illness that originated because of the novel coronavirus.

This pandemic caused a widespread impact on health and the economy. We have chosen this problem because we believe that there is a need for effective visualization of how viruses spread across communities. By looking at the coronavirus spread globally, we can understand how global economies are affected and gain more insights into the patterns and trends of the pandemic. Moreover, this information will be helpful for health officials, policymakers, and the general public since they would be able to make better decisions regarding disease prevention and measures.

Project Goal: How can we visualize the spread of COVID-19 across geographic regions of different sizes? Our goal is to create an interactive visualization tool that allows users to explore COVID-19 case data with the number of deaths and vaccination records, adjusted for population, at the continent and country levels. By combining population data and COVID-19 case data, we aim to create a detailed and informative visualization that can help users understand the impact of the pandemic and the vaccination drive on different regions.

2 Datasets

We utilized the following dataset for our analysis:

- (Source: <https://github.com/owid/covid-19-data/tree/master/public/data> [3])

We constructed a comprehensive database of COVID-19 statistics, including total cases, deaths, and vaccinations, by country and region by processing the dataset and extracting key values.

3 Computational Overview

In our project, we implemented the tree data structure on our dataset, which helps us in representing the hierarchical structure between the world, its regions, and their respective countries and effectively visualize the transmission of COVID-19.

- **Data Preprocessing:** Reading and wrangling the raw data file and extracting relevant information such as dates, cases, deaths, and vaccinations.
- **Tree Construction:** Building tree structures and implementing all the non-trivial recursive algorithms from scratch to represent the relationship between countries, continents, and regions over the years.
- **Data Analysis:** Computing statistics and trends to visualize them in the form of bar plots and map plots.

- **Visualization:** Developing interactive visualizations using Python libraries such as GeoPandas[1]. Matplotlib [2] to illustrate the spread of COVID-19 and associated factors.

In conclusion, our program utilizes Pandas for data wrangling such as filtering and cleaning, Geopandas for analyzing and visualizing the geospatial data of the tree structure, and Matplotlib to visualize the graph and bar plots and all the other interactive features such as sliders and buttons.

4 Instructions for Running

To run our program, follow these steps:

1. Download all the required files and unzip them to obtain program files and folders
2. Install the required Python libraries listed in the `requirements.txt` file.
3. Execute the `main.py` file.
4. The visualisation screen will pop up and adjust the filters according to your preferences.

Upon running the program, you will be able to see a window pop up showing the interactive visualizations of the spread of COVID-19 statistics for the world. Additionally, there will be a play/pause button, reset button, and two sets of filters that will let you choose the region you want to visualize. Upon choosing one of the regions from the filter, you will also be able to see its respective bar plot showcasing the statistics for the countries present in that region.

5 Changes from Proposal

We made several modifications to our project plan based on feedback and further exploration:

- Expanded the scope of analysis by adding three more filters to include total cases, deaths, and vaccinations, which are adjusted by population based on our algorithm.
- Enhanced visualization techniques by adding a bar plot, which allowed us to effectively see the change in cases for each country of its respective region.
- Switched from implementing a search bar to a radio button for filtering through the regions.

6 Discussion

Our computational exploration yielded valuable insights into the dynamics of COVID-19 spread and its determinants. Key findings include:

- The trend of COVID-19 cases from the epicenter to the rest of the world during the prominent 4 years of the pandemic.
- The important role that vaccines play in mitigating the effect of the pandemic by observing the trends in vaccine-intensive countries.
- The trend of the number of deaths showing compared to the number of cases, indicating the mortality rate in different regions.

Despite these contributions, our analysis has certain limitations, such as we mentioned in the project proposal that we will analyze the COVID data at state level and district level but because of the non-availability of reliable data, we were not able to do so.

7 References

References

- [1] GeoPandas Documentation. Retrieved from <https://geopandas.org/en/stable/>
- [2] Matplotlib Documentation. Retrieved from <https://matplotlib.org/>
- [3] Our World in Data. "COVID-19 Data." GitHub, GitHub, Inc., accessed 2 March 2024 <https://github.com/owid/covid-19-data/tree/master/public/data>