



GEO HEALTH VIS REPORT

(Geospatial based Health Data Visualization)

ITCS 5121 – INFORMATION VISUALIZATION

Professor - Dr. Kalpathi Subramanian

GROUP – 10

TEAM MEMBERS:

- | | | |
|------------------------------|---|------------|
| 1. Pavan Gopi Praneeth Gidda | - | 801328702 |
| 2. Rakesh Dulam | - | 801328344 |
| 3. Ramya Reddy | - | 801337485. |

CONTENTS

| | |
|-------------------------|----|
| Abstract | 4 |
| 1. Introduction | 5 |
| 1.1. Problem Statement | 5 |
| 1.2. Motivation | 5 |
| 1.3. Challenges | 6 |
| 2. Datasets | 7 |
| 3. Visualization Design | 15 |
| 4. Implementation | 17 |
| 5. Result Analysis | 23 |
| 6. Limitations | 35 |
| 7. GitHub Link | 36 |
| 8. Future Scope | 37 |
| 9. Conclusion | 38 |
| 10. References | 39 |

ABSTRACT

The lack of accessible and easy-to-understand healthcare data visualization tools impedes the effective communication and comprehension of global health statistics, hindering the development of evidence-based public health policies. Our team aims to tackle this problem by creating Geo Health Vis, a geospatial-based healthcare data visualization tool that provides a clear and concise representation of various medical diseases and their prevalence across the world. Geo Health Vis also allows users to compare healthcare statistics between different locations using scatterplots, histograms, bar plots, and other visualization techniques. This tool will be valuable for public health professionals, researchers, and policymakers who need to understand global health trends to make informed decisions about resource allocation and policy development.

1. INTRODUCTION

1.1. Problem Statement:

- The lack of accessible and easy-to-understand healthcare data visualization tools impedes the effective communication and comprehension of global health statistics, hindering the development of evidence-based public health policies.

This is a significant problem because it prevents public health professionals, researchers, and policymakers from making informed decisions about resource allocation and policy development. For example, if a public health professional does not have access to easy-to-understand data about the prevalence of a particular disease in their community, they may not be able to identify the need for a new public health intervention. Similarly, a researcher who does not have access to easy-to-understand data about the risk factors for a particular disease may not be able to develop a new treatment or prevention strategy. Finally, a policymaker who does not have access to easy-to-understand data about the cost-effectiveness of different public health interventions may not be able to make the best decisions about how to allocate resources.

- Geo Health Vis is a geospatial-based healthcare data visualization tool that aims to address this problem.

Geo Health Vis provides a clear and concise representation of various medical diseases and their prevalence across the world. It also allows users to compare healthcare statistics between different locations using scatterplots, scatterplot matrices, histograms, line plots, and other visualization techniques. This tool will be valuable for public health professionals, researchers, and policymakers who need to understand global health trends to make informed decisions about resource allocation and policy development.

1.2. Motivation:

The global health landscape is complex and ever-changing. To make informed decisions about resource allocation and policy development, public health professionals, researchers, and policymakers need to be able to understand global health trends. However, the lack of accessible and easy-to-understand healthcare data visualization tools makes it difficult to communicate and comprehend these trends.

Geo Health Vis is a geospatial-based healthcare data visualization tool that aims to address this problem. Geo Health Vis provides a clear and concise representation of various medical diseases and their prevalence across the world. It also allows users to compare healthcare statistics between different locations using scatterplots, scatterplot matrices, histograms, line plots, and other visualization techniques.

1.3. Challenges:

There are several challenges that need to be addressed in order to create a successful healthcare data visualization tool. These challenges include:

- The sheer volume and complexity of healthcare data.
- The need to make healthcare data understandable to a wide range of stakeholders, including public health professionals, researchers, and policymakers.
- The need to develop visualization techniques that are both accurate and engaging.
- The need to make healthcare data visualization tools accessible and affordable.

Despite these challenges, there is a great need for healthcare data visualization tools. By providing a clear and concise way to visualize healthcare data, Geo Health Vis can help to improve the understanding of global health trends and lead to better decision-making.

Here are some additional challenges that Geo Health Vis will need to address:

- **Data quality:** The data used to create Geo Health Vis must be of high quality to ensure the accuracy of the visualizations. This means that the data must be collected and processed carefully, and it must be free of errors and biases.
- **Data accessibility:** Geo Health Vis must be easy to use and accessible to a wide range of stakeholders. This means that the tool must be user-friendly and that it must be available in a variety of languages.
- **Data privacy:** Geo Health Vis must protect the privacy of the individuals whose data is used to create the visualizations. This means that the data must be anonymized or aggregated in a way that does not allow individuals to be identified.

Geo Health Vis is a complex project with several challenges. However, the team is confident that they can overcome these challenges and create a successful tool that will improve the understanding of global health trends.

2. DATASETS

Covid-19 Dataset:

A COVID-19 dataset is a collection of data that is used to study the COVID-19 pandemic. This data can include information about cases, deaths, hospitalizations, and other factors related to the virus. Datasets can be used to track the spread of the virus, identify trends, and develop strategies to prevent and control the pandemic.

There are many different types of COVID-19 datasets available. Some datasets are collected by governments, while others are collected by private companies or organizations. Some datasets are publicly available, while others are only available to researchers or other authorized users.

The COVID-19 pandemic is a global health crisis, and datasets are an important tool for understanding and responding to the pandemic. Datasets can be used to track the spread of the virus, identify trends, and develop strategies to prevent and control the pandemic.

Some of the benefits of using COVID-19 datasets include:

- **Tracking the spread of the virus:** Datasets can be used to track the number of cases, deaths, and hospitalizations in different countries and regions. This information can be used to identify trends and hotspots, and to develop strategies to prevent the spread of the virus.
- **Identifying trends:** Datasets can be used to identify trends in the spread of the virus. For example, datasets can be used to identify the types of people who are most likely to be infected, the symptoms that are most common, and the factors that are associated with the severity of the disease.
- **Developing strategies to prevent and control the pandemic:** Datasets can be used to develop strategies to prevent and control the pandemic. For example, datasets can be used to identify which interventions are most effective, and to target these interventions to the populations that are most at risk.

| Country | Other name | ISO 3166-1 | Population | Continent | Total Case | Total Death | Total Case: Total Death | Death percentage |
|------------------------|------------------------|------------|------------|---------------|------------|-------------|-------------------------|------------------|
| Afghanistan | Afghanistan | AFG | 40462186 | Asia | 177827 | 7671 | 4395 | 190 4.313743 |
| Albania | Albania | ALB | 2872296 | Europe | 273870 | 3492 | 95349 | 1216 1.275058 |
| Algeria | Algeria | DZA | 45236699 | Africa | 265691 | 6874 | 5873 | 152 2.587216 |
| Andorra | Andorra | AND | 77481 | Europe | 40024 | 153 | 516565 | 1975 0.382271 |
| Angola | Angola | AGO | 34654212 | Africa | 99194 | 1900 | 2862 | 55 1.915438 |
| Anguilla | Anguilla | AIA | 15237 | Latin Amer | 2700 | 9 | 177200 | 591 0.333333 |
| Antigua and Barbuda | Antigua and Barbuda | ATG | 99348 | Latin Amer | 7493 | 135 | 75422 | 1359 1.801682 |
| Argentina | Argentina | ARG | 45921761 | Latin Amer | 9041124 | 128065 | 196881 | 2789 1.416472 |
| Armenia | Armenia | ARM | 2972939 | Asia | 422574 | 8617 | 142140 | 2898 2.039169 |
| Aruba | Aruba | ABW | 107560 | Latin Amer | 34051 | 212 | 316577 | 1971 0.622596 |
| Australia | Australia | AUS | 26017767 | Oceania | 4680816 | 6384 | 179908 | 245 0.136386 |
| Austria | Austria | AUT | 9096360 | Europe | 3887355 | 15985 | 427353 | 1757 0.411205 |
| Azerbaijan | Azerbaijan | AZE | 10299156 | Asia | 792061 | 9697 | 76905 | 942 1.224274 |
| Bahamas | Bahamas | BHM | 399822 | Latin Amer | 33295 | 788 | 83275 | 1971 2.366722 |
| Bahrain | Bahrain | BHR | 1804995 | Asia | 556241 | 1471 | 308168 | 815 0.264454 |
| Bangladesh | Bangladesh | BGD | 1.68E+08 | Asia | 1951770 | 29122 | 11648 | 174 1.492082 |
| Barbados | Barbados | BRB | 287991 | Latin Amer | 59938 | 375 | 208125 | 1302 0.625647 |
| Belarus | Belarus | BLR | 9443882 | Europe | 965322 | 6844 | 102217 | 725 0.708986 |
| Belgium | Belgium | BEL | 11677924 | Europe | 3851048 | 30826 | 329772 | 2640 0.800457 |
| Belize | Belize | BLZ | 410260 | Latin Amer | 57289 | 656 | 139641 | 1599 1.145071 |
| Benin | Benin | BEN | 12678949 | Africa | 26952 | 163 | 2126 | 13 0.604779 |
| Bermuda | Bermuda | BMU | 61875 | Northern Amer | 12564 | 128 | 203055 | 2069 1.018784 |
| Bhutan | Bhutan | BTN | 786480 | Asia | 31437 | 12 | 39972 | 15 0.038172 |
| Bolivia | Bolivia (Pl) | BOL | 11951714 | Latin Amer | 902448 | 21896 | 75508 | 1832 2.426289 |
| Bosnia and Herzegovina | Bosnia and Herzegovina | BIH | 3245097 | Europe | 375693 | 15719 | 115773 | 4844 1.484001 |

Figure 2.1: Covid-19 Dataset

Cancer Dataset:

This dataset provides information on the age-standardized prevalence of various types of cancer in both sexes across different countries and years, from 1990 to an unknown endpoint. The dataset includes data on 22 types of cancer, including liver cancer, kidney cancer, larynx cancer, breast cancer, thyroid cancer, bladder cancer, uterine cancer, ovarian cancer, stomach cancer, prostate cancer, cervical cancer, testicular cancer, pancreatic cancer, esophageal cancer, nasopharynx cancer, colon and rectum cancer, non-melanoma skin cancer, lip and oral cavity cancer, brain and nervous system cancer, tracheal, bronchus, and lung cancer, and gallbladder and biliary tract cancer. The dataset also includes information on neoplasms in general.

The dataset includes data on 195 countries, represented by their three-letter country codes, and covers a span of several decades. The prevalence rates are reported in percentages, age-standardized to allow for comparisons across different age distributions. The data can be used to identify trends in cancer prevalence over time and across different regions, to inform public health policies and interventions, and to guide research efforts in cancer prevention and treatment.

One potential limitation of the dataset is that it only reports prevalence rates, rather than incidence rates or mortality rates. Prevalence rates reflect the proportion of individuals in a population who have been diagnosed with a particular type of cancer at a given point in time, and may be influenced by factors such as screening practices and survival rates. Incidence rates, on the other hand, reflect the number of new cases of cancer diagnosed in a population over a specified time period, and can provide a better measure of the true burden of the disease. Mortality rates reflect the number of deaths due to cancer in a population over a specified time period and provide another important measure of the impact of cancer on a population.

Overall, this cancer dataset provides a valuable resource for researchers, policymakers, and public health practitioners working to address the global burden of cancer. While there are some limitations to the dataset, such as the lack of incidence and mortality data, the information provided can still be used to identify patterns and trends in cancer prevalence over time and across different populations, and to guide efforts to prevent and treat this devastating disease.

AutoSave

share-of-population-with-cancer-types

Read

Search

Raketh Dulam

FileHomeInsertPage LayoutFormulasDataReviewViewAutomateHelp

CutCopyFormat Painter

Clipboard

Calibri11

B I U

Font

Wrap TextMerge & Center

Alignment

General

Number

Conditional FormattingFormat as TableCell Styles

Styles

InsertDeleteFormat

Cells

AutoSumFillClear

Editing

Sort & FilterFind & SelectAnalyze Data

Analysis

CommentsShare

A1

fx

Entity

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W |
|----|-----------------|------|------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|------------|
| 1 | Entity | Code | Year | Prevalence | Prevalence | Prevalence | Prevalence | Prevalence | Prevalence | Prevalence | Prevalence | Prevalence | Prevalence | Prevalence | Prevalence | Prevalence | Prevalence | Prevalence | Prevalence | Prevalence | Prevalence | Prevalence | Prevalence |
| 2 | Afghanistan AFG | 1990 | 0.01 | 0 | 0.01 | 0.07 | 0.01 | 0.01 | 0 | 0 | 0.04 | 0.02 | 0.02 | 0 | 0 | 0.01 | 0 | 0.02 | 0 | 0 | 0.016165 | 0.01 | |
| 3 | Afghanistan AFG | 1991 | 0.01 | 0 | 0.01 | 0.07 | 0.01 | 0.01 | 0 | 0 | 0.04 | 0.02 | 0.02 | 0 | 0 | 0.01 | 0 | 0.02 | 0 | 0 | 0.015804 | 0.01 | |
| 4 | Afghanistan AFG | 1992 | 0.01 | 0 | 0.01 | 0.06 | 0.01 | 0.01 | 0 | 0 | 0.04 | 0.02 | 0.02 | 0 | 0 | 0.01 | 0 | 0.02 | 0 | 0 | 0.015175 | 0.01 | |
| 5 | Afghanistan AFG | 1993 | 0.01 | 0 | 0.01 | 0.06 | 0.01 | 0.01 | 0 | 0 | 0.04 | 0.02 | 0.02 | 0 | 0 | 0.01 | 0 | 0.02 | 0 | 0 | 0.014904 | 0.01 | |
| 6 | Afghanistan AFG | 1994 | 0.01 | 0 | 0.01 | 0.06 | 0.01 | 0.01 | 0 | 0 | 0.04 | 0.02 | 0.02 | 0 | 0 | 0.01 | 0 | 0.02 | 0 | 0 | 0.015134 | 0.01 | |
| 7 | Afghanistan AFG | 1995 | 0.01 | 0 | 0.01 | 0.06 | 0.01 | 0.01 | 0 | 0 | 0.04 | 0.02 | 0.02 | 0 | 0 | 0.01 | 0 | 0.02 | 0 | 0 | 0.015665 | 0.01 | |
| 8 | Afghanistan AFG | 1996 | 0.01 | 0 | 0.01 | 0.06 | 0.01 | 0.01 | 0 | 0 | 0.04 | 0.02 | 0.02 | 0 | 0 | 0.01 | 0 | 0.02 | 0 | 0 | 0.015546 | 0.01 | |
| 9 | Afghanistan AFG | 1997 | 0.01 | 0 | 0.01 | 0.06 | 0.01 | 0.01 | 0 | 0 | 0.04 | 0.02 | 0.02 | 0 | 0 | 0.01 | 0 | 0.02 | 0 | 0 | 0.015562 | 0.01 | |
| 10 | Afghanistan AFG | 1998 | 0.01 | 0 | 0.01 | 0.06 | 0.01 | 0.01 | 0 | 0 | 0.04 | 0.02 | 0.02 | 0 | 0 | 0.01 | 0 | 0.02 | 0 | 0 | 0.015518 | 0.01 | |
| 11 | Afghanistan AFG | 1999 | 0.01 | 0 | 0.01 | 0.06 | 0.01 | 0.01 | 0 | 0 | 0.04 | 0.02 | 0.02 | 0 | 0 | 0.01 | 0 | 0.02 | 0 | 0 | 0.015693 | 0.01 | |
| 12 | Afghanistan AFG | 2000 | 0.01 | 0 | 0.02 | 0.06 | 0.01 | 0.01 | 0 | 0 | 0.04 | 0.02 | 0.02 | 0 | 0 | 0.01 | 0 | 0.02 | 0 | 0 | 0.015925 | 0.01 | |
| 13 | Afghanistan AFG | 2001 | 0.01 | 0 | 0.02 | 0.06 | 0.01 | 0.01 | 0 | 0 | 0.04 | 0.02 | 0.02 | 0 | 0 | 0.01 | 0 | 0.02 | 0 | 0 | 0.01598 | 0.01 | |
| 14 | Afghanistan AFG | 2002 | 0.01 | 0 | 0.02 | 0.06 | 0.01 | 0.01 | 0 | 0 | 0.04 | 0.02 | 0.02 | 0 | 0 | 0.01 | 0 | 0.02 | 0 | 0 | 0.015821 | 0.01 | |
| 15 | Afghanistan AFG | 2003 | 0.01 | 0 | 0.02 | 0.06 | 0.01 | 0.01 | 0 | 0 | 0.04 | 0.02 | 0.02 | 0 | 0 | 0.01 | 0 | 0.02 | 0 | 0 | 0.016511 | 0.01 | |
| 16 | Afghanistan AFG | 2004 | 0.01 | 0 | 0.02 | 0.07 | 0.01 | 0.01 | 0 | 0 | 0.04 | 0.02 | 0.02 | 0 | 0 | 0.01 | 0 | 0.02 | 0 | 0 | 0.016848 | 0.01 | |
| 17 | Afghanistan AFG | 2005 | 0.01 | 0 | 0.02 | 0.07 | 0.01 | 0.01 | 0 | 0 | 0.04 | 0.02 | 0.02 | 0 | 0 | 0.01 | 0 | 0.02 | 0 | 0 | 0.016575 | 0.01 | |
| 18 | Afghanistan AFG | 2006 | 0.01 | 0 | 0.02 | 0.07 | 0.01 | 0.01 | 0.01 | 0 | 0.04 | 0.02 | 0.02 | 0 | 0 | 0.01 | 0 | 0.02 | 0 | 0 | 0.016311 | 0.01 | |
| 19 | Afghanistan AFG | 2007 | 0.01 | 0 | 0.01 | 0.07 | 0.01 | 0.01 | 0.01 | 0.01 | 0.04 | 0.02 | 0.02 | 0 | 0 | 0.01 | 0 | 0.02 | 0 | 0 | 0.016137 | 0.01 | |
| 20 | Afghanistan AFG | 2008 | 0.01 | 0 | 0.01 | 0.07 | 0.01 | 0.01 | 0.01 | 0.01 | 0.04 | 0.02 | 0.02 | 0 | 0 | 0.01 | 0 | 0.02 | 0 | 0 | 0.015595 | 0.01 | |
| 21 | Afghanistan AFG | 2009 | 0.01 | 0 | 0.01 | 0.07 | 0.01 | 0.01 | 0.01 | 0.01 | 0.04 | 0.02 | 0.02 | 0 | 0 | 0.01 | 0 | 0.02 | 0 | 0 | 0.015135 | 0.01 | |
| 22 | Afghanistan AFG | 2010 | 0.01 | 0 | 0.01 | 0.07 | 0.01 | 0.01 | 0.01 | 0.01 | 0.04 | 0.02 | 0.02 | 0 | 0 | 0.01 | 0 | 0.02 | 0 | 0 | 0.015094 | 0.01 | |
| 23 | Afghanistan AFG | 2011 | 0.01 | 0 | 0.01 | 0.07 | 0.01 | 0.01 | 0.01 | 0.01 | 0.04 | 0.02 | 0.02 | 0 | 0 | 0.01 | 0 | 0.02 | 0 | 0 | 0.014854 | 0.01 | |
| 24 | Afghanistan AFG | 2012 | 0.01 | 0 | 0.01 | 0.07 | 0.01 | 0.01 | 0.01 | 0.01 | 0.04 | 0.02 | 0.02 | 0 | 0 | 0.01 | 0 | 0.02 | 0 | 0 | 0.014594 | 0.01 | |
| 25 | Afghanistan AFG | 2013 | 0.01 | 0 | 0.01 | 0.07 | 0.01 | 0.01 | 0.01 | 0.01 | 0.04 | 0.02 | 0.02 | 0 | 0 | 0.01 | 0 | 0.02 | 0 | 0 | 0.014669 | 0.01 | |
| 26 | Afghanistan AFG | 2014 | 0.01 | 0.01 | 0.01 | 0.07 | 0.01 | 0.01 | 0.01 | 0.01 | 0.04 | 0.02 | 0.02 | 0 | 0 | 0.01 | 0 | 0.02 | 0 | 0 | 0.014551 | 0.01 | |
| 27 | Afghanistan AFG | 2015 | 0.01 | 0.01 | 0.01 | 0.08 | 0.01 | 0.01 | 0.01 | 0.01 | 0.04 | 0.02 | 0.02 | 0 | 0 | 0.01 | 0 | 0.02 | 0 | 0 | 0.014418 | 0.01 | |

<

share-of-population-with-cancer

+

Figure 2.2: Cancer Dataset.

Heart Disease Dataset:

The heart disease dataset provided includes information about the number of deaths caused by cardiovascular diseases (CVD) in various countries and regions around the world between 1990 and 2019. The data are presented as the percentage of total deaths caused by CVD in the population of all ages and both sexes. This dataset provides valuable information for researchers and policymakers to assess the burden of CVD and plan effective interventions to reduce mortality.

The dataset includes information for 195 countries and 21 regions, including Afghanistan and the African Region (WHO) which are listed in the sample. The data for Afghanistan shows a consistent decline in the percentage of deaths caused by CVD from 1990 to 1999, followed by a fluctuating trend with a slight increase in 2017. The data for the African Region (WHO) show a stable percentage of CVD deaths during the early 1990s, followed by a gradual increase until 2017.

The dataset can be used to compare the burden of CVD between different countries and regions. For instance, researchers can use the data to identify countries or regions with a high burden of CVD and prioritize interventions to reduce the mortality. The dataset can also be

used to assess the impact of different interventions and policies on the burden of CVD in different countries and regions.

The dataset can also be used to analyze trends in the burden of CVD over time. Researchers can use the data to identify countries or regions where the burden of CVD is increasing or decreasing and investigate the reasons behind the trends. This information can help policymakers to plan interventions and policies to reduce the burden of CVD.

Overall, the heart disease dataset provides valuable information for researchers, policymakers, and healthcare professionals to understand the burden of CVD in different countries and regions and plan effective interventions to reduce the mortality. However, it is important to note that the data are presented as percentages, and the actual number of deaths caused by CVD may vary between countries and regions due to differences in population size and age structure.

| Entity | Code | Year | Deaths - Cardiovascular diseases - Sex: Both - Age: All Ages (Percent) |
|-----------------|------|------|--|
| Afghanistan AFG | | 1990 | 24.63 |
| Afghanistan AFG | | 1991 | 23.58 |
| Afghanistan AFG | | 1992 | 22.33 |
| Afghanistan AFG | | 1993 | 21.27 |
| Afghanistan AFG | | 1994 | 20.37 |
| Afghanistan AFG | | 1995 | 20.32 |
| Afghanistan AFG | | 1996 | 20.39 |
| Afghanistan AFG | | 1997 | 20.09 |
| Afghanistan AFG | | 1998 | 19.32 |
| Afghanistan AFG | | 1999 | 20.75 |
| Afghanistan AFG | | 2000 | 21.01 |
| Afghanistan AFG | | 2001 | 21.16 |
| Afghanistan AFG | | 2002 | 21.07 |
| Afghanistan AFG | | 2003 | 21.59 |
| Afghanistan AFG | | 2004 | 21.88 |
| Afghanistan AFG | | 2005 | 21.92 |
| Afghanistan AFG | | 2006 | 22.01 |
| Afghanistan AFG | | 2007 | 22.16 |
| Afghanistan AFG | | 2008 | 22.49 |
| Afghanistan AFG | | 2009 | 22.82 |
| Afghanistan AFG | | 2010 | 22.91 |
| Afghanistan AFG | | 2011 | 23.14 |
| Afghanistan AFG | | 2012 | 23.21 |
| Afghanistan AFG | | 2013 | 23.38 |
| Afghanistan AFG | | 2014 | 23.08 |
| Afghanistan AFG | | 2015 | 22.94 |

Figure 2.3: Heart Disease Dataset

Monkey Pox Dataset:

The Monkeypox disease dataset contains information about the reported cases of Monkeypox virus infection in different countries around the world. The dataset contains information about the country, location, date, new cases, new cases smoothed, total cases, new cases per million, total cases per million, and new cases smoothed per million.

The first column, `iso_code`, contains the International Organization for Standardization (ISO) code for the country where the cases were reported. The second column, `location`, contains the name of the country where the cases were reported. The third column, `date`, provides the date when the cases were reported.

The fourth column, `new_cases`, provides the number of new cases reported on the specific date. The fifth column, `new_cases_smoothed`, provides the seven-day rolling average of new

cases, which is useful to identify the trend of the outbreak. The sixth column, `total_cases`, provides the total number of cases reported in the country until the specific date.

The seventh column, `new_cases_per_million`, provides the number of new cases per one million population. This is useful to compare the outbreak intensity across different countries with varying population sizes. The eighth column, `total_cases_per_million`, provides the total number of cases per one million population.

The ninth column, `new_cases_smoothed_per_million`, provides the seven-day rolling average of new cases per one million population. This is useful to compare the trend of the outbreak across different countries with varying population sizes.

Overall, this dataset provides valuable information about the spread of the Monkeypox virus in different countries and regions around the world. The data can be used to identify trends and patterns in the outbreak, which can help public health officials and policymakers to develop effective strategies to control the spread of the virus. The dataset can also be used to compare the intensity and trend of the outbreak across different countries, which can help to identify areas that require immediate attention and intervention.

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W |
|----|----------|------------|-----------|-----------|-------------|-----------------------|-------------------------|------------|--------------|---------------------------------|-----------------------------------|---|---|---|---|---|---|---|---|---|---|---|---|
| | iso_code | location | date | new_cases | total_cases | new_cases_per_million | total_cases_per_million | new_deaths | total_deaths | new_deaths_smoothed_per_million | total_deaths_smoothed_per_million | | | | | | | | | | | | |
| 1 | AND | Andorra | 7/25/2022 | 1 | 1 | 12.653 | 12.653 | | | | | | | | | | | | | | | | |
| 2 | AND | Andorra | 7/26/2022 | 2 | 3 | 25.306 | 37.958 | | | | | | | | | | | | | | | | |
| 3 | AND | Andorra | 7/27/2022 | 0 | 3 | 0 | 37.958 | | | | | | | | | | | | | | | | |
| 4 | AND | Andorra | 7/28/2022 | 0 | 3 | 0 | 37.958 | | | | | | | | | | | | | | | | |
| 5 | AND | Andorra | 7/29/2022 | 0 | 3 | 0 | 37.958 | | | | | | | | | | | | | | | | |
| 6 | AND | Andorra | 7/30/2022 | 0 | 3 | 0 | 37.958 | | | | | | | | | | | | | | | | |
| 7 | AND | Andorra | 7/31/2022 | 0 | 0.43 | 3 | 0 | 37.958 | 5.441 | | | | | | | | | | | | | | |
| 8 | AND | Andorra | 8/1/2022 | 0 | 0.29 | 3 | 0 | 37.958 | 3.669 | | | | | | | | | | | | | | |
| 9 | AND | Andorra | 8/2/2022 | 0 | 0 | 3 | 0 | 37.958 | 0 | | | | | | | | | | | | | | |
| 10 | AND | Andorra | 8/3/2022 | 0 | 0 | 3 | 0 | 37.958 | 0 | | | | | | | | | | | | | | |
| 11 | AND | Andorra | 8/4/2022 | 0 | 0 | 3 | 0 | 37.958 | 0 | | | | | | | | | | | | | | |
| 12 | AND | Andorra | 8/5/2022 | 0 | 0 | 3 | 0 | 37.958 | 0 | | | | | | | | | | | | | | |
| 13 | AND | Andorra | 8/6/2022 | 0 | 0 | 3 | 0 | 37.958 | 0 | | | | | | | | | | | | | | |
| 14 | AND | Andorra | 8/7/2022 | 0 | 0 | 3 | 0 | 37.958 | 0 | | | | | | | | | | | | | | |
| 15 | AND | Andorra | 8/8/2022 | 0 | 0 | 3 | 0 | 37.958 | 0 | | | | | | | | | | | | | | |
| 16 | AND | Andorra | 8/9/2022 | 1 | 0.14 | 4 | 12.653 | 50.611 | 1.771 | | | | | | | | | | | | | | |
| 17 | ARE | United Ara | 5/24/2022 | 1 | 1 | 0.107 | 0.107 | | | | | | | | | | | | | | | | |
| 18 | ARE | United Ara | 5/25/2022 | 0 | 1 | 0 | 0.107 | | | | | | | | | | | | | | | | |
| 19 | ARE | United Ara | 5/26/2022 | 0 | 1 | 0 | 0.107 | | | | | | | | | | | | | | | | |
| 20 | ARE | United Ara | 5/27/2022 | 0 | 1 | 0 | 0.107 | | | | | | | | | | | | | | | | |
| 21 | ARE | United Ara | 5/28/2022 | 0 | 1 | 0 | 0.107 | | | | | | | | | | | | | | | | |
| 22 | ARE | United Ara | 5/29/2022 | 3 | 4 | 0.32 | 0.427 | | | | | | | | | | | | | | | | |
| 23 | ARE | United Ara | 5/30/2022 | 0 | 0.57 | 4 | 0 | 0.427 | 0.061 | | | | | | | | | | | | | | |
| 24 | ARE | United Ara | 5/31/2022 | 0 | 0.43 | 4 | 0 | 0.427 | 0.046 | | | | | | | | | | | | | | |
| 25 | ARE | United Ara | 6/1/2022 | 4 | 1 | 8 | 0.427 | 0.854 | 0.107 | | | | | | | | | | | | | | |
| 26 | ARE | United Ara | 6/2/2022 | 0 | 1 | 8 | 0 | 0.854 | 0.107 | | | | | | | | | | | | | | |
| 27 | ARE | United Ara | 6/3/2022 | 0 | 1 | 8 | 0 | 0.854 | 0.107 | | | | | | | | | | | | | | |

Figure 2.4: Monkey Pox Dataset.

Covid Dataset for India:

The India Covid dataset is a collection of daily reports of new cases and deaths due to Covid-19 in India, starting from January 3rd, 2020. The dataset contains information such as the date, the WHO region, new cases, cumulative cases, new deaths, and cumulative deaths. The dataset is particularly useful for analyzing the progression of the pandemic in India and understanding the impact of various interventions and policies on the spread of the virus.

The dataset starts on January 3rd, 2020, and continues until the present day. This timeline is useful for understanding the timeline of the pandemic in India and how it has progressed over time. The dataset contains information about the WHO region, which is important because it allows for comparisons to be made between different regions of the world.

The new cases and cumulative cases columns provide information about the number of new cases and the total number of cases reported in India. These numbers are essential for tracking the spread of the virus and understanding how it is spreading in India. The new deaths and cumulative deaths columns provide information about the number of new deaths and the total number of deaths reported in India. These numbers are essential for understanding the impact of the virus on the population of India and for tracking changes in mortality rates over time.

The India Covid dataset is an excellent resource for researchers and policymakers who are interested in understanding the impact of the pandemic on India. By analyzing the data, researchers can identify trends and patterns in the spread of the virus and the impact of interventions such as lockdowns and vaccination campaigns. Policymakers can use this information to make informed decisions about how to allocate resources and implement policies that can help control the spread of the virus.

In conclusion, the India Covid dataset is a valuable resource for understanding the impact of the pandemic in India. It provides information about new cases, cumulative cases, new deaths, and cumulative deaths, and covers the timeline of the pandemic from January 3rd, 2020, until the present day. By analyzing this data, researchers and policymakers can gain insights into the spread of the virus and the impact of various interventions on controlling its spread.

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V |
|-----|------------|------|------|-------|----|-----|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 152 | 06-01-2020 | EMRO | 656 | 15836 | 15 | 269 | | | | | | | | | | | | | | | | |
| 153 | 06-02-2020 | EMRO | 742 | 16578 | 5 | 274 | | | | | | | | | | | | | | | | |
| 154 | 06-03-2020 | EMRO | 775 | 17353 | 24 | 298 | | | | | | | | | | | | | | | | |
| 155 | 06-04-2020 | EMRO | 624 | 17977 | 5 | 303 | | | | | | | | | | | | | | | | |
| 156 | 06-05-2020 | EMRO | 1078 | 19055 | 10 | 313 | | | | | | | | | | | | | | | | |
| 157 | 06-06-2020 | EMRO | 582 | 19637 | 18 | 331 | | | | | | | | | | | | | | | | |
| 158 | 06-07-2020 | EMRO | 791 | 20428 | 30 | 361 | | | | | | | | | | | | | | | | |
| 159 | 06-08-2020 | EMRO | 575 | 21003 | 12 | 373 | | | | | | | | | | | | | | | | |
| 160 | 06-09-2020 | EMRO | 305 | 21308 | 7 | 380 | | | | | | | | | | | | | | | | |
| 161 | 06-10-2020 | EMRO | 920 | 22228 | 29 | 409 | | | | | | | | | | | | | | | | |
| 162 | 06-11-2020 | EMRO | 748 | 22976 | 21 | 430 | | | | | | | | | | | | | | | | |
| 163 | 06-12-2020 | EMRO | 656 | 23632 | 20 | 450 | | | | | | | | | | | | | | | | |
| 164 | 6/13/2020 | EMRO | 556 | 24188 | 5 | 455 | | | | | | | | | | | | | | | | |
| 165 | 6/14/2020 | EMRO | 664 | 24852 | 20 | 475 | | | | | | | | | | | | | | | | |
| 166 | 6/15/2020 | EMRO | 761 | 25613 | 5 | 480 | | | | | | | | | | | | | | | | |
| 167 | 6/16/2020 | EMRO | 106 | 25719 | 8 | 488 | | | | | | | | | | | | | | | | |
| 168 | 6/17/2020 | EMRO | 1241 | 26960 | 20 | 508 | | | | | | | | | | | | | | | | |
| 169 | 6/18/2020 | EMRO | 463 | 27423 | 42 | 550 | | | | | | | | | | | | | | | | |
| 170 | 6/19/2020 | EMRO | 541 | 27964 | 2 | 552 | | | | | | | | | | | | | | | | |
| 171 | 6/20/2020 | EMRO | 419 | 28383 | 0 | 552 | | | | | | | | | | | | | | | | |
| 172 | 6/21/2020 | EMRO | 536 | 28919 | 33 | 585 | | | | | | | | | | | | | | | | |
| 173 | 6/22/2020 | EMRO | 310 | 29229 | 17 | 602 | | | | | | | | | | | | | | | | |
| 174 | 6/23/2020 | EMRO | 338 | 29567 | 20 | 622 | | | | | | | | | | | | | | | | |
| 175 | 6/24/2020 | EMRO | 159 | 29726 | 21 | 643 | | | | | | | | | | | | | | | | |
| 176 | 6/25/2020 | EMRO | 535 | 30261 | 36 | 679 | | | | | | | | | | | | | | | | |

Figure 2.5: Covid -19 dataset from India.

Covid-19 Dataset from USA:

The given dataset contains information about the number of COVID-19 cases and deaths globally, starting from January 3, 2020, to May 9, 2023. The dataset has eight columns, namely date, new_cases, new_deaths, total_cases, total_deaths, weekly_cases, weekly_deaths, biweekly_cases, and biweekly_deaths.

The 'date' column contains the date on which the data is recorded. The 'new_cases' column contains the number of new COVID-19 cases reported on a particular day. The 'new_deaths' column contains the number of deaths due to COVID-19 reported on a particular day. The 'total_cases' column contains the total number of COVID-19 cases reported until a particular day. Similarly, the 'total_deaths' column contains the total number of deaths due to COVID-19 reported until a particular day.

The 'weekly_cases' column contains the number of COVID-19 cases reported in the previous week leading up to a particular day. The 'weekly_deaths' column contains the number of deaths due to COVID-19 reported in the previous week leading up to a particular day. The 'biweekly_cases' column contains the number of COVID-19 cases reported in the two weeks leading up to a particular day. Similarly, the 'biweekly_deaths' column contains the number of deaths due to COVID-19 reported in the two weeks leading up to a particular day.

The dataset provides a comprehensive record of COVID-19 cases and deaths globally, which can be used to analyze the spread and impact of the pandemic. The new_cases and new_deaths columns provide daily updates on the pandemic's progression. The total_cases and total_deaths columns provide a cumulative view of the pandemic's impact.

The weekly_cases, weekly_deaths, biweekly_cases, and biweekly_deaths columns provide a view of the pandemic's progression over longer periods. These columns can be used to identify trends in the pandemic's spread and impact. The dataset can be used to analyze the effectiveness of various interventions and measures taken to control the spread of the pandemic.

In conclusion, the given dataset provides a comprehensive record of COVID-19 cases and deaths globally. The dataset can be used to analyze the spread and impact of the pandemic and to identify trends in its progression. The dataset can also be used to analyze the effectiveness of various interventions and measures taken to control the spread of the pandemic.

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V |
|-----|------------|-----|----|-------|-----|------|----|------|-----|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 122 | 05-02-2020 | 298 | 8 | 2469 | 72 | 1139 | 29 | 1561 | 42 | | | | | | | | | | | | | |
| 123 | 05-03-2020 | 0 | 0 | 2469 | 72 | 1006 | 23 | 1536 | 42 | | | | | | | | | | | | | |
| 124 | 05-04-2020 | 0 | 0 | 2469 | 72 | 938 | 22 | 1473 | 39 | | | | | | | | | | | | | |
| 125 | 05-05-2020 | 0 | 0 | 2469 | 72 | 766 | 12 | 1443 | 36 | | | | | | | | | | | | | |
| 126 | 05-06-2020 | 755 | 23 | 3224 | 95 | 1397 | 35 | 2132 | 59 | | | | | | | | | | | | | |
| 127 | 05-07-2020 | 168 | 9 | 3392 | 104 | 1565 | 44 | 2216 | 64 | | | | | | | | | | | | | |
| 128 | 05-08-2020 | 171 | 1 | 3563 | 105 | 1392 | 41 | 2337 | 65 | | | | | | | | | | | | | |
| 129 | 05-09-2020 | 0 | 0 | 3563 | 105 | 1094 | 33 | 2233 | 62 | | | | | | | | | | | | | |
| 130 | 05-10-2020 | 839 | 14 | 4402 | 119 | 1933 | 47 | 2939 | 70 | | | | | | | | | | | | | |
| 131 | 05-11-2020 | 262 | 1 | 4664 | 120 | 2195 | 48 | 3133 | 70 | | | | | | | | | | | | | |
| 132 | 05-12-2020 | 303 | 7 | 4967 | 127 | 2498 | 55 | 3264 | 67 | | | | | | | | | | | | | |
| 133 | 5/13/2020 | 0 | 0 | 4967 | 127 | 1743 | 32 | 3140 | 67 | | | | | | | | | | | | | |
| 134 | 5/14/2020 | 372 | 9 | 5339 | 136 | 1947 | 32 | 3512 | 76 | | | | | | | | | | | | | |
| 135 | 5/15/2020 | 714 | 17 | 6053 | 153 | 2490 | 48 | 3882 | 89 | | | | | | | | | | | | | |
| 136 | 5/16/2020 | 349 | 15 | 6402 | 168 | 2839 | 63 | 3933 | 96 | | | | | | | | | | | | | |
| 137 | 5/17/2020 | 233 | 1 | 6635 | 169 | 2233 | 50 | 4166 | 97 | | | | | | | | | | | | | |
| 138 | 5/18/2020 | 437 | 4 | 7072 | 173 | 2408 | 53 | 4603 | 101 | | | | | | | | | | | | | |
| 139 | 5/19/2020 | 583 | 4 | 7655 | 177 | 2688 | 50 | 5186 | 105 | | | | | | | | | | | | | |
| 140 | 5/20/2020 | 490 | 10 | 8145 | 187 | 3178 | 60 | 4921 | 92 | | | | | | | | | | | | | |
| 141 | 5/21/2020 | 531 | 6 | 8676 | 193 | 3337 | 57 | 5284 | 89 | | | | | | | | | | | | | |
| 142 | 5/22/2020 | 540 | 12 | 9216 | 205 | 3163 | 52 | 5653 | 100 | | | | | | | | | | | | | |
| 143 | 5/23/2020 | 736 | 10 | 9952 | 215 | 3550 | 47 | 6389 | 110 | | | | | | | | | | | | | |
| 144 | 5/24/2020 | 716 | 7 | 10668 | 222 | 4033 | 53 | 6266 | 103 | | | | | | | | | | | | | |
| 145 | 5/25/2020 | 512 | 1 | 11180 | 223 | 4108 | 50 | 6516 | 103 | | | | | | | | | | | | | |
| 146 | 5/26/2020 | 737 | 1 | 11917 | 224 | 4262 | 47 | 6950 | 97 | | | | | | | | | | | | | |
| 147 | 5/27/2020 | 548 | 6 | 12465 | 230 | 4320 | 43 | 7498 | 103 | | | | | | | | | | | | | |
| 148 | 5/28/2020 | 637 | 9 | 13102 | 239 | 4426 | 46 | 7763 | 103 | | | | | | | | | | | | | |

Figure 2.6: Covid -19 Dataset from USA

3.VISUALISATION DESIGN

The web page that provides multiple visualizations related to health. The name of the webpage is “Geo Health Viz”. The visualization designs in the code include a header and footer, a top navigation bar, and a dropdown menu. The webpage has a comparison feature that allows users to compare the health statistics of two countries. The webpage also provides visualizations for population trends and disease trends like Cancer, Heart disease, Monkey Pox, Covid-19.

The webpage has two different sections with visualizations. The first section shows population trends, and the second section shows disease trends. Both sections have similar visualization designs with slight variations in data.

The following are the details of the design:

- Header and Footer: The header and footer are important elements of the webpage design that provide a consistent look and feel throughout the website. In the header section of the Geo Health Viz webpage, the name of the webpage, "Geo Health Viz", is prominently displayed in a clear and easy-to-read font. This immediately communicates to the user what the purpose of the webpage is and what they can expect to find.
-
- The header may also contain additional information such as a logo, a navigation menu, or a search bar. Depending on the webpage, the header may be fixed in position the top of the screen or scroll along with the rest of the webpage.
- The footer section of the webpage typically contains additional information such as copyright notices, disclaimers, and contact information. In the case of Geo Health Viz, the footer provides users with ways to get in touch with the team, such as an email address or a contact form. This helps to establish credibility and trust with the user, as they know they can easily reach out to the team with any questions or concerns.
- Overall, a well-designed header and footer can greatly enhance the user experience by providing consistent branding and easy access to important information.
- Top Navigation Bar: The top navigation bar contains links to different sections of the webpage, including disease trends, and the compare feature. The “Disease” dropdown menu contains links to different diseases patterns and trends visualisation.
- Comparison Feature: The comparison feature has a separate section on the webpage. It includes a form with two text fields for entering the names of the countries to be compared. The form has a “Compare” button to submit the data. Once the data is submitted, the webpage displays a visualization that compares the health statistics of the two countries.
- Population Trends Visualization: The population trends visualization includes a world map with different colours representing the population density of different countries. The map has tooltips that display the name of the country.

- Cancer Disease Trends Visualization: The cancer disease trends visualization includes a world map with different colors representing the cancer death rates of different countries. The map has tooltips that display the name of the country and the cancer death rate.
- Heart Disease Trends Visualization: The heart disease trends visualization include a world map with different colors representing the heart disease death rates of different countries. The map has tooltips that display the name of the country and the heart disease death rate.
- Monkey Pox Disease Trends Visualization: The Monkey Pox disease trends visualization include a world map with different colors representing the Monkey Pox death rates of different countries. The map has tooltips that display the name of the country and the Monkey Pox death rate.
- COVID-19 Disease Trends Visualization: The COVID-19 disease trends visualization include a world map with different colors representing the COVID-19 death rates of different countries. The map has tooltips that display the name of the country and the COVID-19 death rate.
- General: The webpage has a responsive design that adapts to different screen sizes. The visualization is interactive and user-friendly, with clear labels and instructions. The colours used in the visualization have been chosen carefully to ensure that the trends are visualized clearly.

4. IMPLEMENTATION

Overview:

The Geo Health Viz is a program that uses HTML, SVG, D3JS and displays two visualizations: a globe and a map of the world with data on population and diseases. The globe visualization displays a 3D model of the world, colored by country, and shows markers at locations with data. 2D map of the world visualization displays map colored by trends related to population and diseases. The script uses the D3.js library to create the visualizations.

The web page allows users to compare population trends and disease patterns between two countries. To do this, users can select two countries from the drop-down menus and then click on the "Compare" button. This will display a scatter plot that shows the population of the two countries over time. Similarly, selecting two diseases from the drop-down menus and then clicking on the "Compare" button will display a scatter plot that shows the number of cases of the two diseases in the two countries over time.

The Geo Health Viz is a powerful tool for public health researchers and policymakers. It can be used to identify areas with high rates of disease and population growth, track population growth and decline, and compare population and disease trends between different countries. The visualizations are interactive, allowing users to explore data in a more engaging way and gain deeper insights into population and health trends.

3D Globe Visualization:

This implementation is a JavaScript code that creates a population globe visualization using D3.js library. The visualization displays the world map as a 3D globe that rotates. The locations of several countries are highlighted by different colors based on a categorical color scale. The colors are assigned to the countries using their ISO country codes. A legend is also included to provide information about the color scale.

The implementation consists of four functions: drawGlobe, drawGraticule, enableRotation, and drawMarkers. The drawGlobe function uses D3 queue to load two data sets, the world map in JSON format and the country location data in a separate JSON file. It then appends the world map to the SVG element and sets the styles of the map. The drawMarkers function draws markers on the globe to indicate the locations of the countries. The enableRotation function animates the rotation of the globe using the D3 timer. The drawGraticule function adds the latitudinal and longitudinal lines to the globe, allowing the user to see the spatial orientation of the countries.

Overall, this implementation provides an interactive and engaging way to visualize global population and diseases data. It demonstrates the use of D3.js library to create complex and dynamic data visualizations in the browser.

Chloropleth Map Visualization:

This is a JavaScript implementation of a choropleth map visualization of the world³. The code starts by hiding other pages and displaying the home page.

The implementation uses the D3.js library to create an SVG element and projections that will map the world map onto it. The data is loaded from two external sources: a GeoJSON file that provides information about the shape of the countries and a CSV file that contains population data for each country. The data is then used to color-code each country on the map using a color scale ranging from light blue to dark blue based on the population or disease threshold.

The implementation also includes mouseover and mouseleave events that highlight the country being hovered over and displays the name of the country in a tooltip. When a country is clicked, a function called `loadComparePage` is called. Finally, the implementation creates a color scale legend to help the viewer interpret the map.

Color Scale Creation:

We created a color scale legend using D3.js, a popular JavaScript library for data visualization. The legend is created within an SVG element and is positioned at coordinates (1300, 400) using the `transform` attribute.

The legend is composed of a series of groups (`g` elements) created using the `selectAll()` and `data()` methods. For each group, a rectangle element is appended with a position based on the index of the data (`i`) and a fill color determined by the `colorScale` function.

Lastly, a text element is appended to each group with the label for each color in the scale. The `attr()` method is used to set the position and alignment of the text element. The resulting legend will display a series of rectangles with colors representing values within a domain, along with labels for each color. This type of visualization is commonly used to help viewers understand the meaning of data represented by color.

Scatterplot Visualization:

We created a scatter plot using data from a CSV file. It starts by selecting an SVG element to append the plot to. Then it uses the D3.js library to load the data from the CSV file.

The plot is created by defining the scales for the x and y axes using the D3.js `scaleLinear()` function. These scales are then used to set the position of the circles on the plot. The color of the circles is set to red.

Finally, the code creates x and y axes using the D3.js `axisBottom()` and `axisLeft()` functions. These axes are drawn on the plot using the D3.js `select()` and `call()` functions.

Overall, this implementation creates a simple scatter plot using D3.js and demonstrates the use of scales and axes.

Density Plot Visualization:

We created a density plot using D3.js library. It first sets the dimensions and margins of the graph and appends the SVG object to the body of the page. Then, it uses D3's **d3.csv** function to load the data from a CSV file.

The density plot is created using the kernel density estimation algorithm, which is implemented in the **kernelDensityEstimator** function. The x-axis is created using D3's **scaleLinear** function, which scales the data to fit within the given width. The y-axis is created similarly, with the range set to the maximum density value.

After setting up the axes, the density plot is drawn using the **path** element, which is filled with a blue color and given an opacity of 0.8. The **stroke** and **stroke-width** properties define the color and width of the line. Finally, the **curve** property defines the curve type of the line.

Overall, this implementation creates an interactive and informative visualization of the density plot for cardiovascular disease deaths, with a clear indication of the maximum density value.

Histogram:

This implementation is done using D3.js, a JavaScript library for data visualization, to create a histogram plot. The function **histoPlotForMonkeyPox()** sets the dimensions and margins of the graph, appends an SVG element to the HTML document, and defines the scales for the X and Y axes.

The function then uses the **d3.csv()** method to read a CSV file with data about Monkey Pox cases worldwide, and applies the **d3.histogram()** function to this data to generate bins for the histogram. The **d3.axisBottom()** and **d3.axisLeft()** methods are used to draw the X and Y axes, respectively. Finally, the **svg.selectAll()** and **svg.append()** methods are used to create and style the bar rectangles of the histogram.

In summary, this implementation creates a histogram plot using D3.js, based on data about Monkey Pox cases worldwide. It sets the dimensions and margins of the graph, defines the scales for the X and Y axes, generates bins for the histogram, and creates and styles the bar rectangles. The resulting plot has been displayed in an HTML document using an SVG element with the ID **svgAreaForMonkeyPox**.

Bar Plot Visualization:

We implemented a bar plot using D3.js library to visualize cancer cases. The dimensions and margins of the graph are set and an SVG object is appended to the HTML body. The **svg** object is then transformed to fit the margins using a translate transform.

The data is then parsed using **d3.csv** function which reads the CSV file containing cancer cases data. The X-axis is defined using a **scaleBand()** function which divides the domain of the graph (cancer types) into evenly spaced bands. The Y-axis is defined using a **scaleLinear()** function that maps the cancer case values to the height of the SVG object.

The **selectAll("mybar")** function selects all **rect** elements that are not yet created, binds the data to them, and creates a rectangle for each data point. The rectangles are then positioned using the **x** and **y** attributes, and their height and width are defined using the **height** and **width** attributes. Finally, the **fill** attribute sets the color of the bars to blue.

In summary, this implementation reads in data on cancer cases, uses D3.js to create a bar plot with cancer types on the X-axis, cases on the Y-axis, and colored bars representing the number of cases for each cancer type.

Home Page Visualization:

The block of HTML code creates a webpage with multiple sections that can be navigated through using a tabbed interface. The main body of the webpage is enclosed in a div tag with the ID "mainBody". This tag contains multiple nested div tags, each with a unique ID and different display properties.

The first nested div tag with the ID "tooltip" is responsible for displaying a tooltip on the webpage. The tooltip is positioned absolutely and has an initial opacity of 0, meaning it is not visible until activated by user interaction.

The remaining nested div tags are used to create different sections of the webpage, each with a unique ID and display properties. The first section with the ID "homePage" contains two SVG tags for visualizing population trends. The next four sections with IDs "CancerPage", "HeartDiseasePage", "MonkeyPoxPage", and "COVIDPage" are each associated with a particular disease and contain two SVG tags each for visualizing disease trends.

Each section has a title enclosed in an h2 tag and a sub-title enclosed in an h3 tag. The titles and sub-titles are centered on the page with margins set using inline CSS. The SVG tags are also centered on the page with margins set using inline CSS. The SVG tags have unique IDs associated with them, allowing JavaScript code to manipulate them individually.

The display property of the nested div tags associated with the different disease sections is set to "none" by default, meaning they are hidden until activated by user interaction. This allows the user to navigate between sections of the webpage using the tabbed interface without seeing all the content at once.

Overall, the HTML code creates a tabbed webpage with multiple sections for visualizing disease trends, each with a unique ID and display properties. The webpage uses inline CSS to center content and create margins, and unique IDs to allow for manipulation of individual SVG tags using JavaScript.

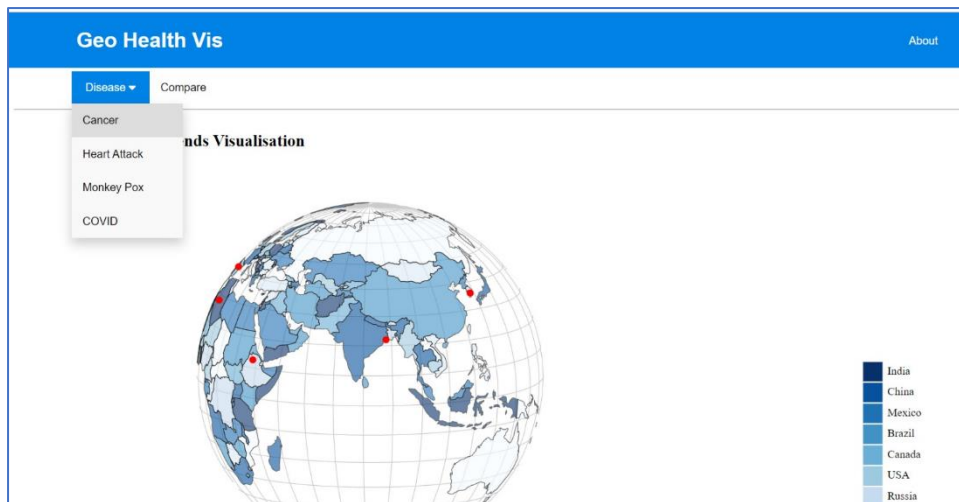


Figure 4.1: Geo Health Vis

Comparison between Countries Visualization:

We implemented a comparison of various disease trends between countries. The page loads with the mainBody element hidden and the compareBody element displayed, along with four different visualizations, each one related to a different disease.

The first function, **barplotForCancer**, creates a bar chart of cancer cases using the d3.js library. The chart is created by parsing the data from a CSV file and plotting the number of cases on the Y axis and the type of cancer on the X axis. The function sets the dimensions of the graph, appends an SVG element to the body of the page and creates the X and Y axes, as well as the bars of the chart. The bars are color-coded in blue.

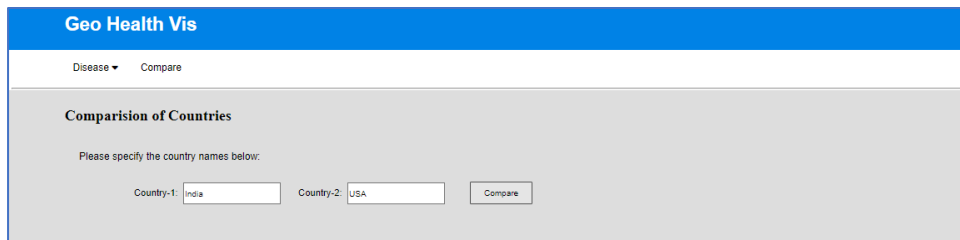
The second function, **histoPlotForMonkeyPox**, creates a histogram of monkey pox cases using the d3.js library. The chart is created by parsing the data from a CSV file and plotting the total number of cases on the Y axis and the number of bins on the X axis. The function sets the dimensions of the graph, appends an SVG element to the body of the page and creates the X and Y axes, as well as the bars of the chart. The bars are color-coded in blue.

The third function, **densityPlotForHeartStroke**, creates a density plot of heart stroke cases using the d3.js library. The chart is created by setting the dimensions of the graph, appending an SVG element to the body of the page and plotting the density curve of the data on the X and Y axes. The function sets the dimensions of the graph, appends an SVG element to the body of the page and creates the X and Y axes, as well as the density curve of the chart. The curve is color-coded in blue.

Finally, the fourth function, **plotForCOVID**, creates a plot of COVID-19 cases using the Plotly.js library. The chart is created by parsing the data from a CSV file and plotting the number of cases over time. The function sets the dimensions of the graph, appends a Plotly.js graph to the body of the page, and creates the X and Y axes. The curve is color-coded in blue.

Overall, this implementation provides a clear comparison of the different geo health visualizations. The user can easily compare the number of cases of different diseases in

different regions and over time, as well as the distribution of cases. This implementation could be useful for researchers, public health officials, and anyone interested in visualizing and comparing health data.



The screenshot shows a web application titled "Geo Health Vis" with a blue header. Below the header is a white bar containing a "Disease" dropdown menu and a "Compare" button. The main content area has a light gray background and is titled "Comparison of Countries". It includes a prompt "Please specify the country names below:" followed by two input fields: "Country-1:" with "India" entered and "Country-2:" with "USA" entered. A "Compare" button is positioned to the right of the second input field.

Figure 4.2: Comparison of Countries.

5. RESULT ANALYSIS

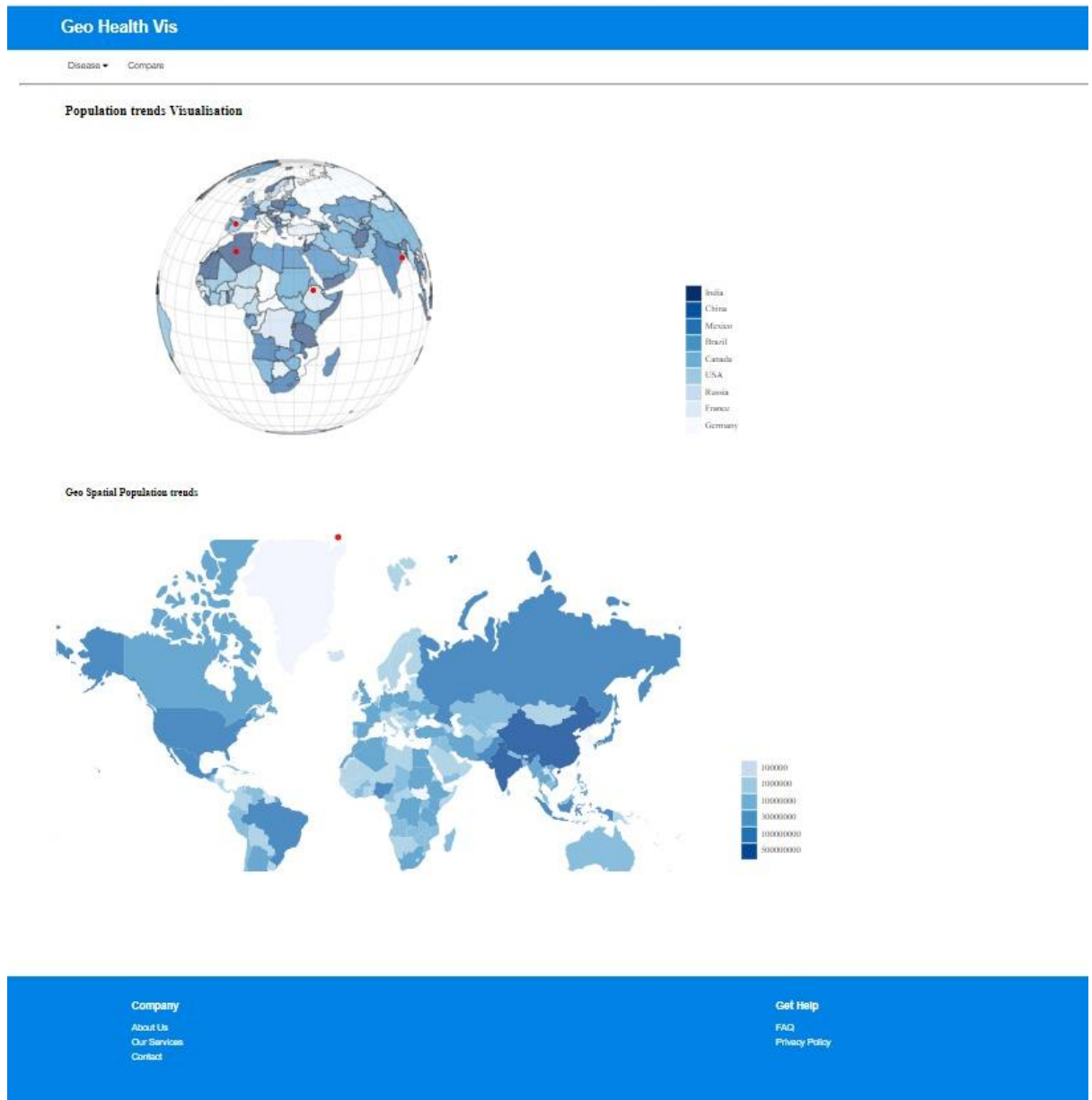


Figure 5.1: Geo Health Vis home page.

The figure 5.1 shows the geo health vis home page which shows the globe visualization of the population trends, if we hover the mouse on the choropleth map it displays the population data. In this home page we can move from one disease to other disease visualisations.

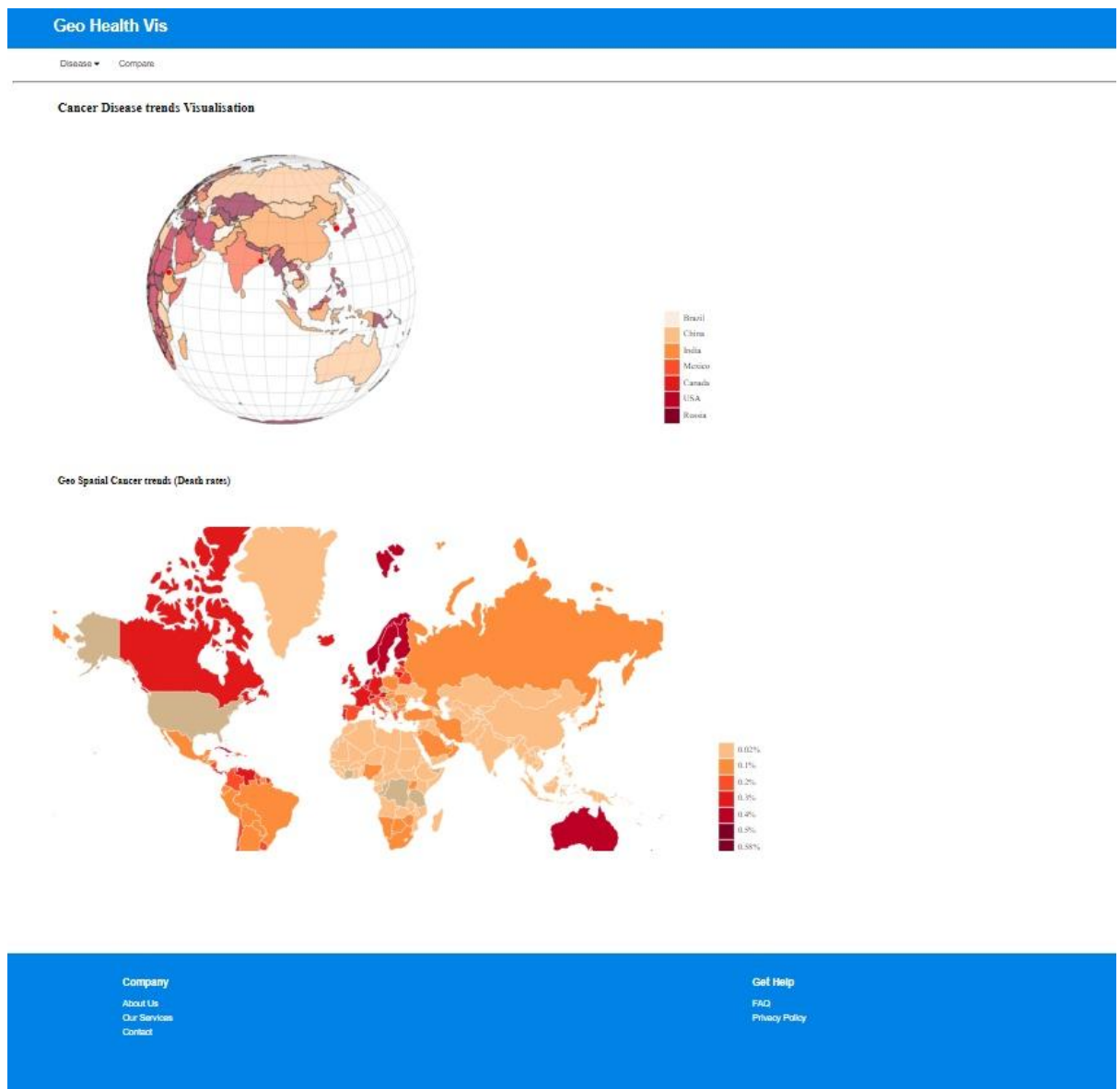


Figure 5.2: Cancer Disease Visualisation.

The figure 5.2 shows the cancer disease visualisation all over the world, when we hover the mouse on map it shows the death rate of the country.

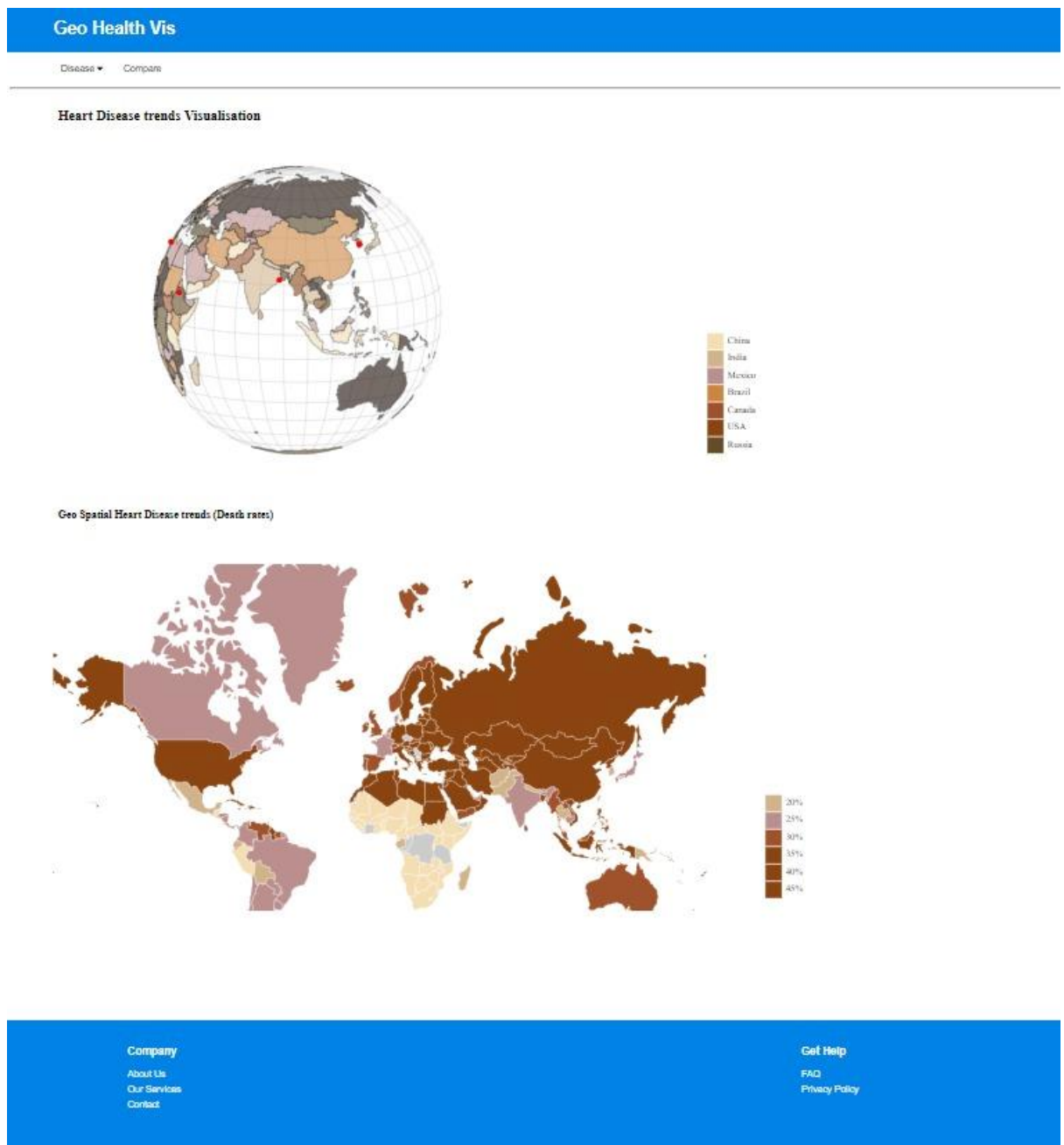


Figure 5.3: Heart Disease Visualisation

This visualisation shows the data which country has highest death rate due to heart disease. When we hover the mouse over the map it displays the percentage of death rate of the specific country.

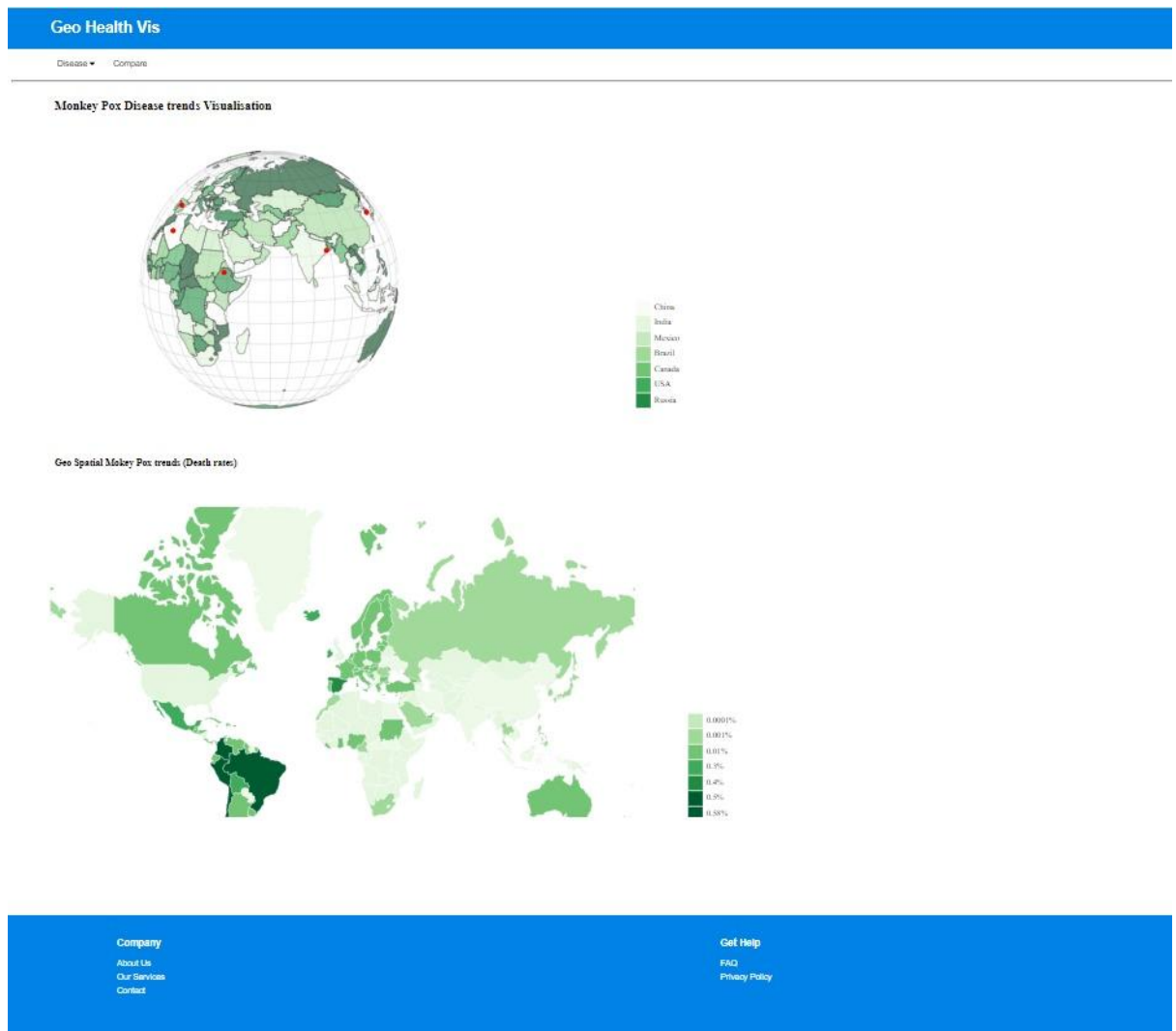


Figure 5.4: This visualisation shows the data which country has highest death rate due to monkey pox. When we hover the mouse over the map it displays the percentage of death rate of the specific country.

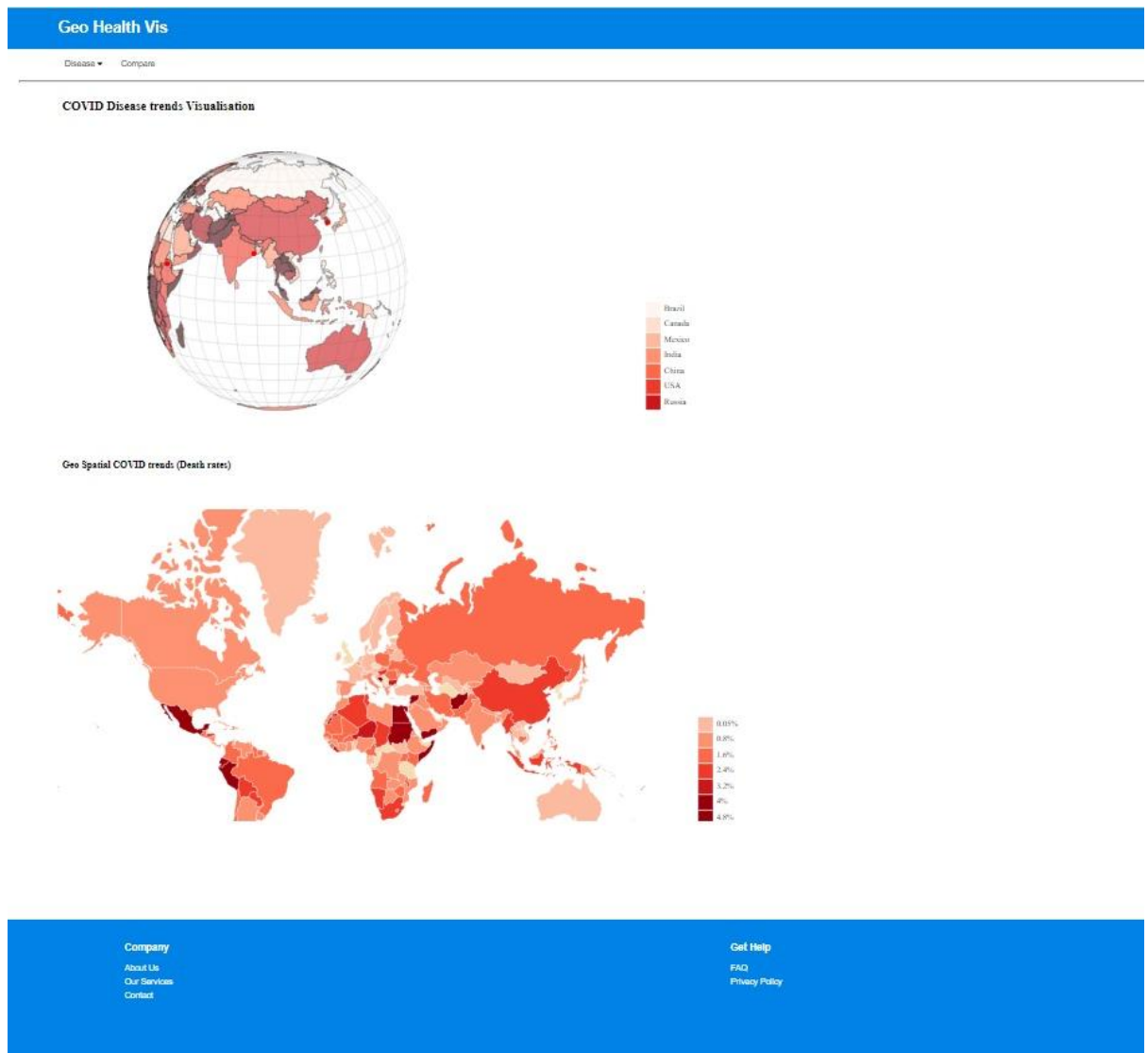
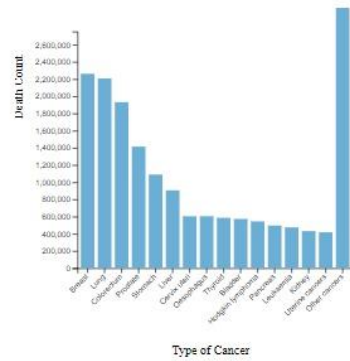


Figure 5.5: This visualisation shows the data which country has highest death rate due to covid-19. When we hover the mouse over the map it displays the percentage of death rate of the specific country.

Statistical Analysis of Health records

Bar Distribution for Cancer Disease



Density Distribution for Heart Stroke Disease

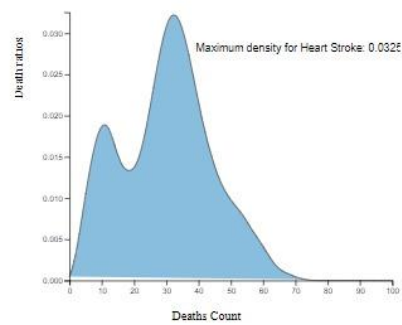


Figure 5.6: Bar distribution and Density distribution.

This figure 5.6 shows the bar distribution and density distribution. These are graphs displayed when we click on the specific country.

Statistical Analysis of Health records

Bar Distribution for Cancer Disease

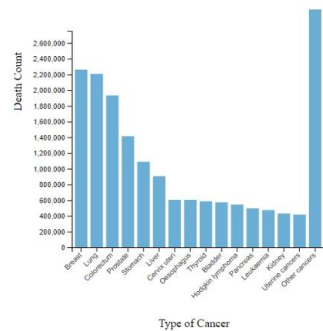


Figure 5.7: Bar distribution

This figure 5.7 shows the bar distribution for the cancer disease it plots death rate on y-axis and type of disease on x-axis.

Density Distribution for Heart Stroke Disease

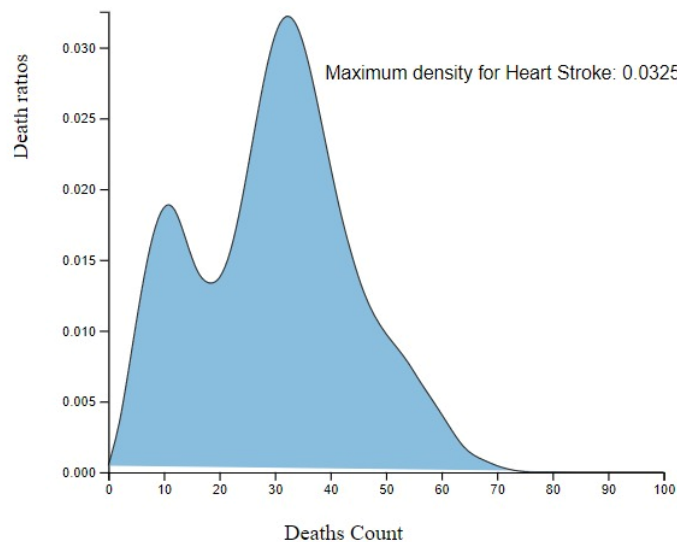


Figure 5.7: Density distribution.

The Density distribution shows the death count on x-axis and death ratio on y-axis. These are graphs are displayed when we click on the specific country. Here the maximum density for heart stroke is 0.0325.

Recording of Monkey Pox cases

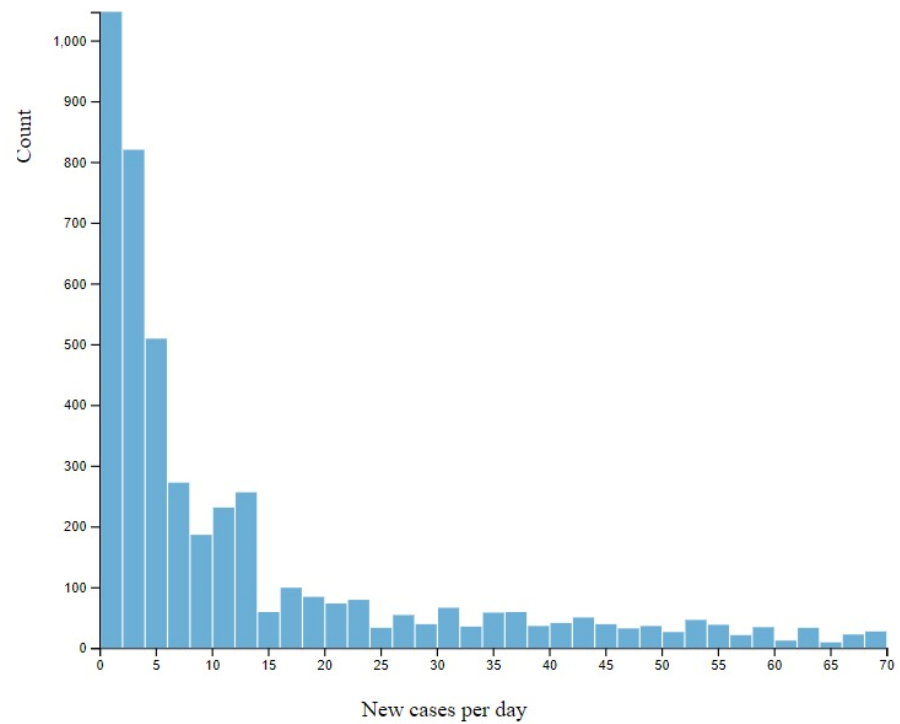


Figure 5.8: Plot for monkey pox cases.

It shows the new cases per day on x-axis and count on the y-axis.

Total cases noted vs Total Deaths for COVID

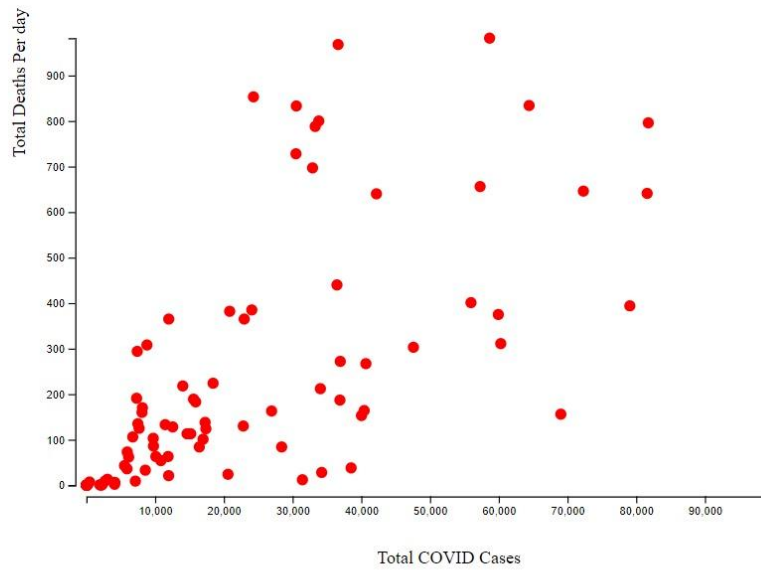


Figure 5.9: Scatterplot which shows the total covid cases on x-axis and total deaths on y-axis.

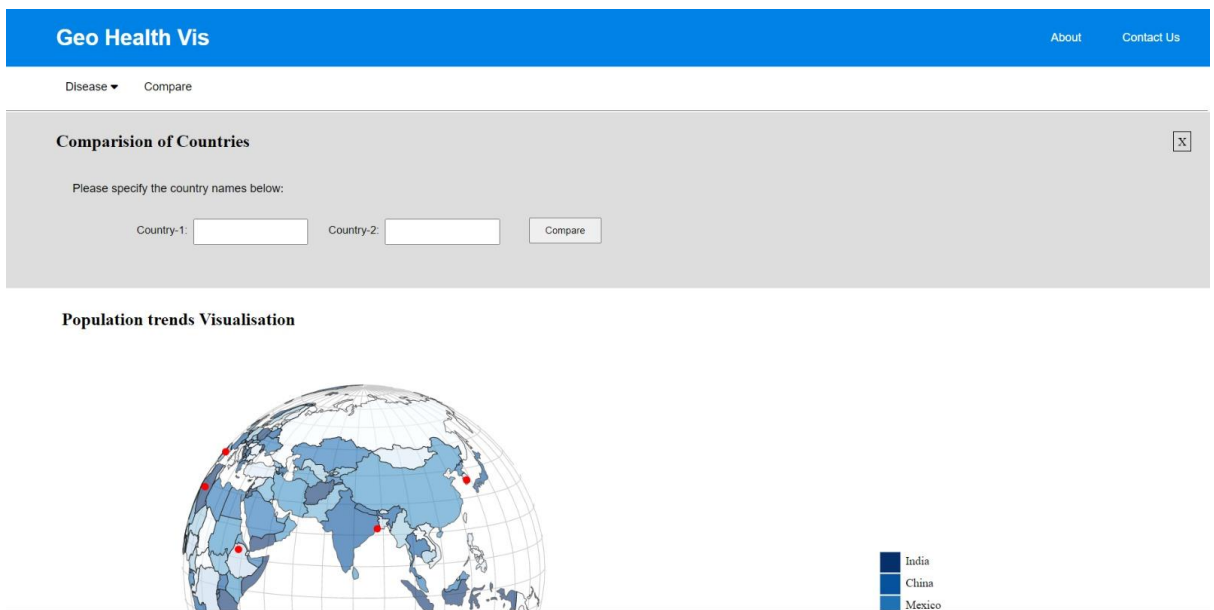


Figure 5.10: Comparison of two countries.

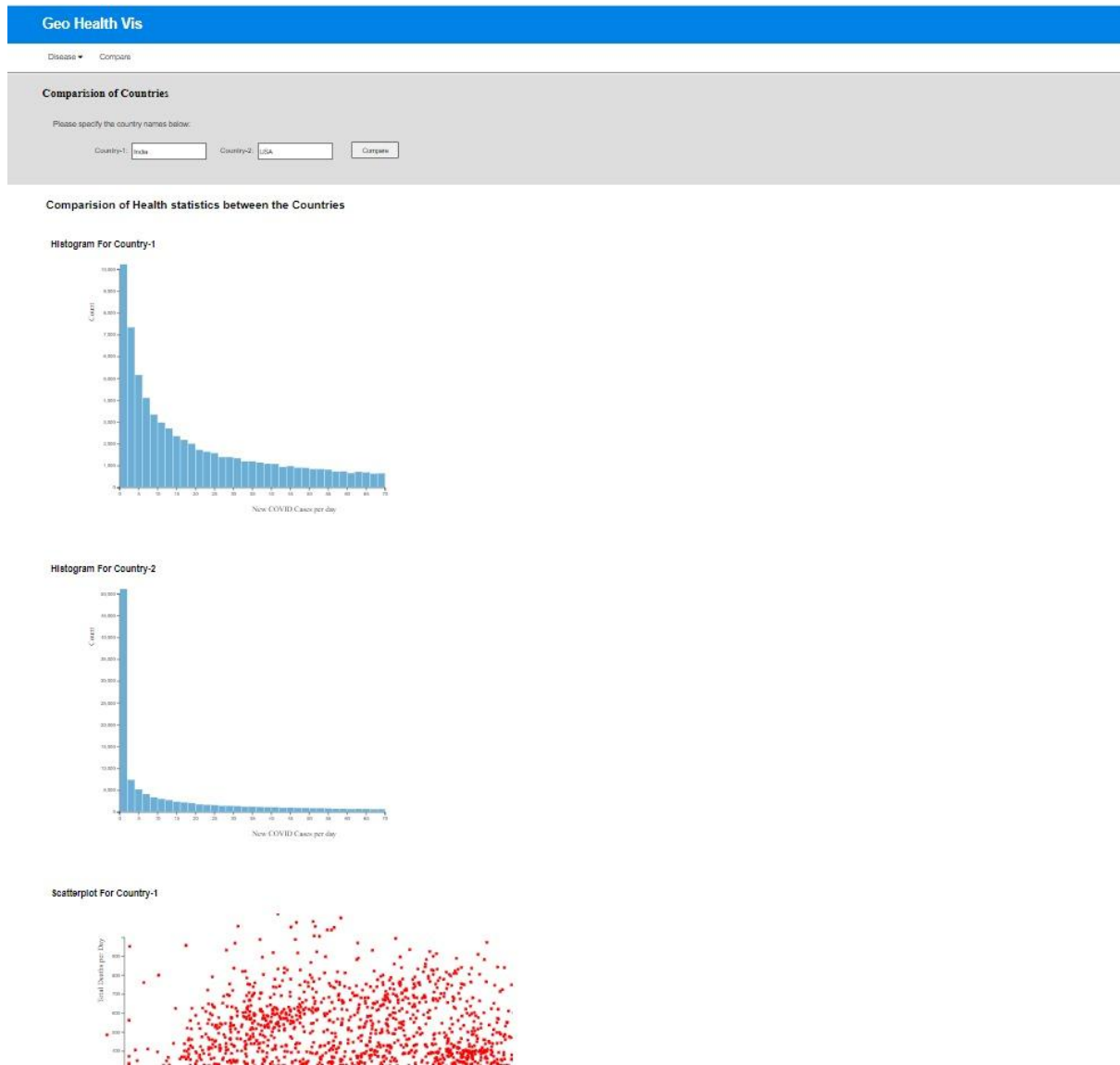
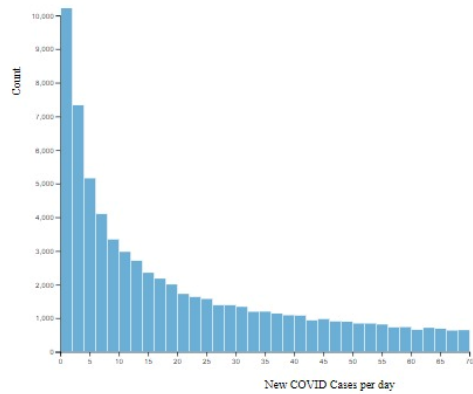


Figure 5.11: Comparison of visualisation between two countries.

When we select two countries and click on compare button it shows the histogram for two countries as well as scatterplots for both the countries.

Comparison of Health statistics between the Countries

Histogram For Country-1



Histogram For Country-2

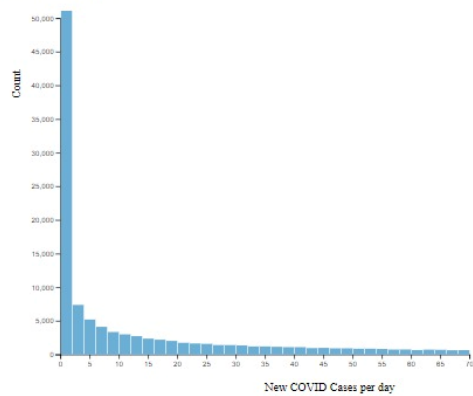
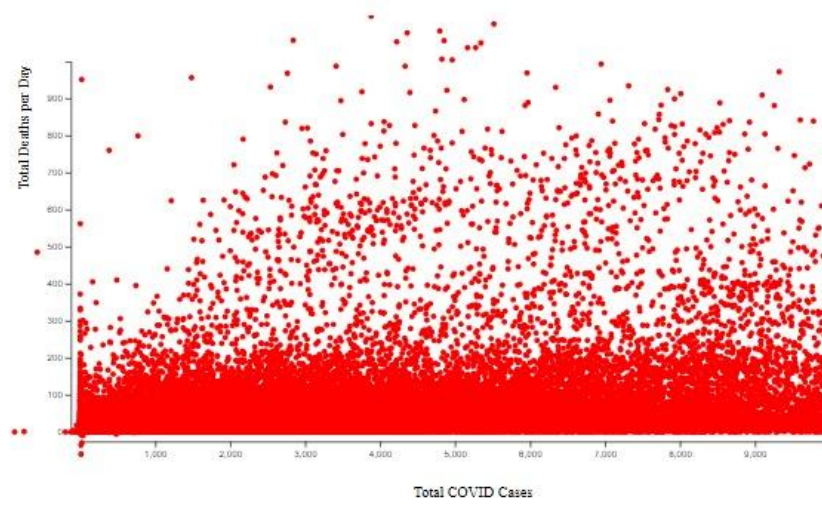


Figure 5.12: Histogram of health statistics comparison between two countries.

The plot shows the new covid cases per day on x-axis and count on the y-axis.

Scatterplot For Country-1



Scatterplot For Country-2

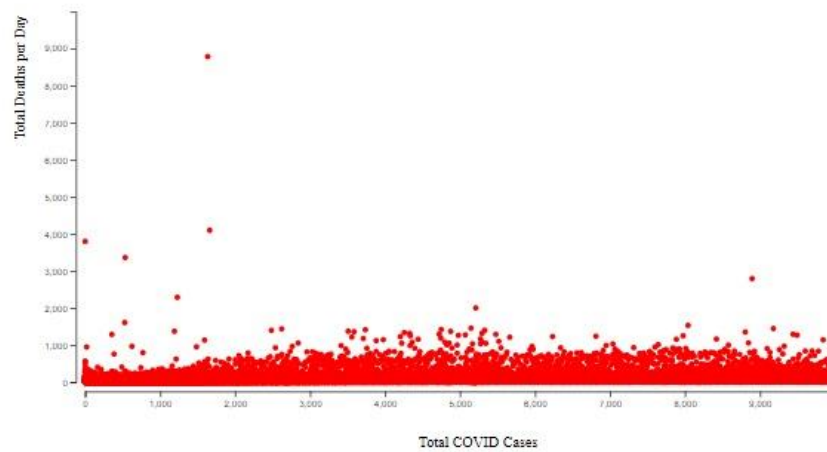


Figure 5.13: Scatterplot comparison between two countries.

6. LIMITATIONS

There may be several limitations to the Geo Health Viz healthcare data visualization tool. Here are some potential limitations to consider:

Data Availability: The tool's usefulness is dependent on the availability and quality of healthcare data that is being visualized. If data is not available for certain regions or countries, or if the data is outdated or inaccurate, the tool may not provide an accurate representation of healthcare statistics.

Platform Compatibility: The tool may not be compatible with all web browsers or mobile devices, which could limit its accessibility to certain users.

User Feedback: The tool may not meet the needs and expectations of all users. User feedback and user testing may be necessary to identify potential issues and areas for improvement.

It is important to consider these limitations when designing and implementing the tool to ensure that it is as effective and useful as possible.

7.GITHUB LINK

https://github.com/PavanGidda99/IV_Group10_FinalProject.git

8. FUTURE SCOPE

- Geo Health Vis will be used by a wide range of stakeholders, including public health professionals, researchers, and policymakers.
- Geo Health Vis will help to improve the understanding of global health trends.
- Geo Health Vis will lead to better decision-making about resource allocation and policy development.
- Geo Health Vis will improve public health outcomes, such as reduced disease prevalence and improved quality of life.

Specifically, Geo Health Vis will be used to:

- Identify and track health trends.
- Monitor the effectiveness of public health interventions.
- Allocate resources more effectively.
- Conduct epidemiological studies.
- Identify risk factors for disease.
- Develop new treatments and prevention strategies.
- Make informed decisions about resource allocation and policy development.

Geo Health Vis has the potential to make a significant contribution to the field of global health. By providing a clear and concise way to visualize healthcare data, Geo Health Vis can help to improve the understanding of global health trends and lead to better decision-making. This could lead to improved public health outcomes, such as reduced disease prevalence and improved quality of life.

9. Conclusion

In conclusion, the lack of accessible and easy-to-understand healthcare data visualisation tools has been a significant impediment to the development of evidence-based public health policies. The creation of Geo Health Vis, a geospatial-based healthcare data visualisation tool, offers a promising solution to this challenge. The tool provides a clear and concise representation of various medical diseases and their prevalence across the world, while also enabling the comparison of healthcare statistics between different locations using different visualization techniques. Geo Health Vis has the potential to enhance the communication and comprehension of global health statistics, thereby supporting evidence-based public health policies. Our team believes that this tool will be a valuable resource for policymakers, public health professionals, and researchers in addressing global health challenges.

What have we learned?

Throughout the project, we learned about the importance of effective communication and comprehension of global health statistics, the value of geospatial-based healthcare data visualization tools in informing public health policies, and the potential limitations that must be considered when designing and implementing such tools. We also learned about the importance of user feedback and user testing to identify areas for improvement and ensure that the tool meets the needs and expectations of its intended audience.

Overall, the Geo Health Viz project provides a valuable contribution to the field of public health by making healthcare data more accessible and easier to understand, and by empowering public health professionals, researchers, and policymakers with the tools they need to make informed decisions about resource allocation and policy development.

10. References

- [1] Ranganathan P, Pramesh CS, Aggarwal R. Common pitfalls in statistical analysis: measures of agreement. *Perspect Clin Res.* 2017;8(4):187-191. doi:10.4103/picr.PICR_127_17
- [2] Grimes DA, Schulz KF. An overview of clinical research: the lay of the land. *Lancet.* 2002;359(9300):57-61. doi:10.1016/S0140-6736(02)07283-5
- [3] Gliklich RE, Dreyer NA. *Registries for Evaluating Patient Outcomes: A User's Guide.* 2nd ed. Rockville (MD): Agency for Healthcare Research and Quality (US); 2010. (Methods Guide for Comparative Effectiveness Reviews). PMID: 21413307
- [4] World Health Organization. Global Health Observatory (GHO) data. <https://www.who.int/gho/en/>.
- [5] Institute for Health Metrics and Evaluation. Global Burden of Disease Study 2019 (GBD 2019) Results. <https://vizhub.healthdata.org/gbd-compare/>.