



Pandoras Box

Visual solution for Pandemic Control

Project Report

INFO-I590 Data Visualization

under Dr.Sadamori Kojaku

Authors:

Sushmitha Keerthy

Immanuel Savio Donbosco

Sylvia Boddu

UID:2000913990, 2000919040, 2000752346

Fall 2022

Report

Abstract

Cartography, or the study and practice of making maps, has a long history dating back to ancient civilizations. Maps have been used for a variety of purposes, including navigation, military strategy, and resource management. In modern times, cartography has continued to evolve with the use of new technologies. Maps are an important tool for visualizing and understanding spatial relationships and patterns, and they can be used to communicate complex information in an easily understandable format. In the context of pandemics, maps can be used to show the distribution of cases and other relevant data, such as vaccination rates, hospital capacity, and the effectiveness of containment measures. By showing this information in a spatial context, maps can help policymakers and the public understand the impact of the pandemic and make informed decisions about how to respond. When the coronavirus pandemic hit, it affected over 623 million people to date. The World Health Organization has attributed the reason for this spread to the lack of knowledge about the disease itself, mutation, and most importantly the movement of diseases among the public. This project will visualize the history of pandemics and epidemics from around the world using novice analytical tools through spatial maps. This would be a novel approach to learning the pattern of the spread of diseases visually.

Keywords: Visualization, Cartograph, Pandemic, Tableau, Chloropleth, Disease Spread.

Contents

| | | |
|----------|---|-----------|
| 1 | Introduction | 1 |
| 1.1 | Motivation | 1 |
| 1.2 | Existing Visualization | 1 |
| 1.3 | Contribution | 4 |
| 1.4 | Objective | 5 |
| 2 | Data and Methods | 5 |
| 2.1 | Process | 5 |
| 2.2 | Ideas and prototype | 6 |
| 2.3 | Data | 6 |
| 2.4 | Visualization methods selection | 7 |
| 3 | Results and Discussions | 8 |
| 3.1 | Covid 19 - Results | 8 |
| 3.2 | Malaria Results | 11 |
| 3.3 | HIV Results | 14 |
| 4 | Conclusions | 17 |
| 4.1 | Analysis | 17 |
| 4.2 | Future Scope | 18 |
| 5 | References | 19 |

1. Introduction

1.1 Motivation

We all have recently experienced the most impactful pandemic Covid19. This pandemic has changed a lot of people's lives across the globe where a total of 6.5 million were killed. Like covid 19, if we go back in history there have been major pandemics that turned out to be catastrophic. It is hard to contain such outbreaks, but a group of informatics specialists could develop some visualization that could help us derive useful information from the pandemic. We as a team felt the urge to do something about it. Below are some important motivations:

- Understanding the spread and impact of the pandemic: It is often not easy to make such an analysis with just tabled data, having it visualized helps up better in understanding the patterns of origin and the spread patterns.
- Communicating information to the Public: Spreading awareness among the public is the best way of prevention. Visualizations make it easy to communicate than text/tabled data. The severity of the spread is conveyed more easily with help of visualizations.
- Identifying Areas need Help: Visualizations can help policymakers and healthcare professionals identify the most affected areas and know the level of help needed. Some remote places in the world lack proper medical care, such areas will have a tough time dealing with pandemic/disease spread. This motivated us in building a visualization that shows cases of the most impactful pandemics where there is medical care. Once we know about such places there could be NGOs, medical organizations, and WHO to help them.
- Tracking the progress and evaluation intervention: During the pandemic, if we could have some visualization display the severity of the spread, we would know which method of intervention is working the best. For example, when covid 19 hit India, a lockdown was enforced in a few states and the states that did not enforce lockdown had a huge impact, these observations are made with the help of visualization.

1.2 Existing Visualization

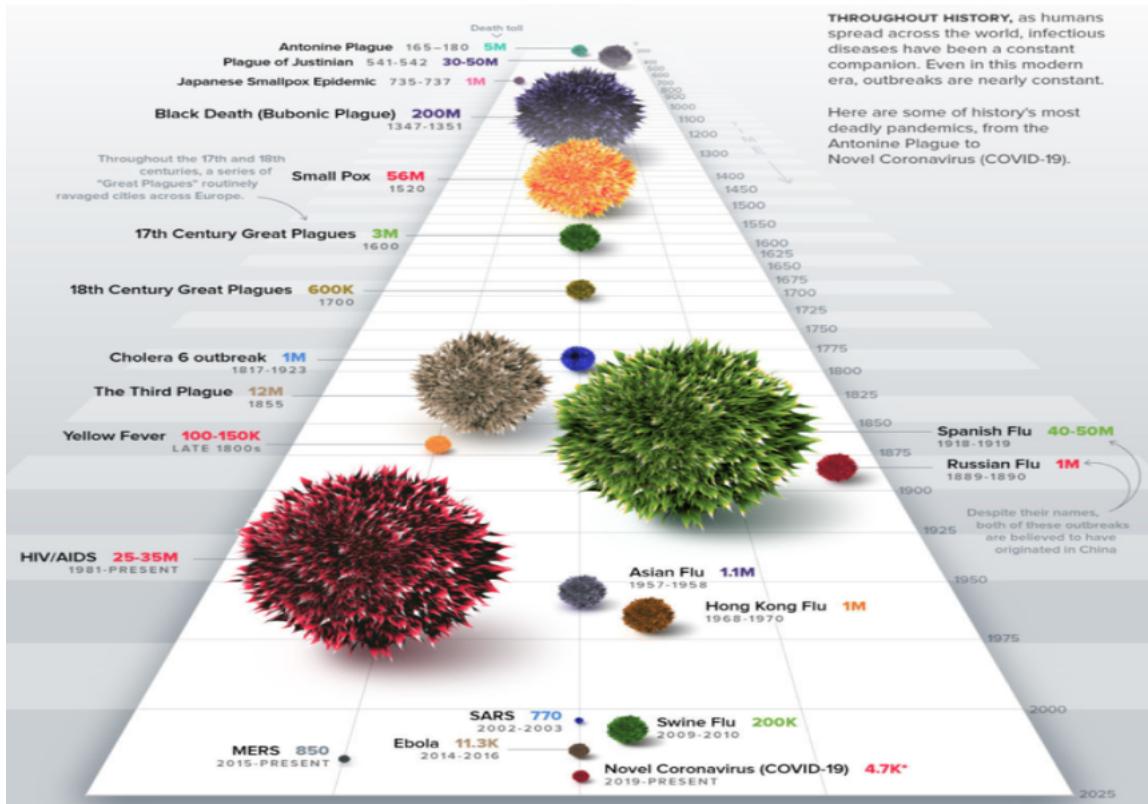
While searching for some of the existing work to structure our project we came across the following, the lapses in these visualizations helped up design wisely and make our project better.

Timeline view of pandemics below gives us a reasonable number of details. The visualization is chronological arrangement of the pandemics that we observed from 1300s till 2003

What is good about this visualization?:

- It covers the information about the pandemics from a really large time frame where the pandemic is indicated as virus globes that are arranged on a timeline ramp.

- The population effected by each pandemic is mentioned explicitly on the left side of the trail and visually by depicting the size of each pandemic.
- The color coding is done well, and it is not creating much confusion in identifying the specific pandemic.
- The time line also shows is that the diseases observed in late 1300s and 1500s and a little ahead of them are not observed at present.



Timeline of Pandemics

What are the missing from this visualization?

There are no details of the location of spread. Which country had the initial breakout? How did it spread across the world? what are the most effected regions. The lack of such information makes it hard nearly impossible to do a proper analysis of the outbreak pattern, effected zones and immune zones.

2. World map canvas view of covid spread: We came across this Jhon Hopkins university dashboard where there are some interesting visualizations of pandemics.

What is good about this visualization?

- There is a world map canvas that makes it very easy to observe the location of spread and outbreak
- The severity of the pandemic is also observed with the size of the circles at the location.
- This is also an interactive visualization where we can zoom into any country and state to observe more details.

- This is a good visualization we consider it to be the initial reference of our project.
- What are the missing from this visualization?
- Diversity, there is only one pandemic analyzed, only covid 19.
 - This visualization does not provide any details of the origin and the way the pandemic has spread etc



John Hopkins Covid Dashboard

3. Heatmap pandemic representation: This is the most relatable existing visualization we came across, most of the issues observed in the previously discussed existing visualizations are not found in this work.

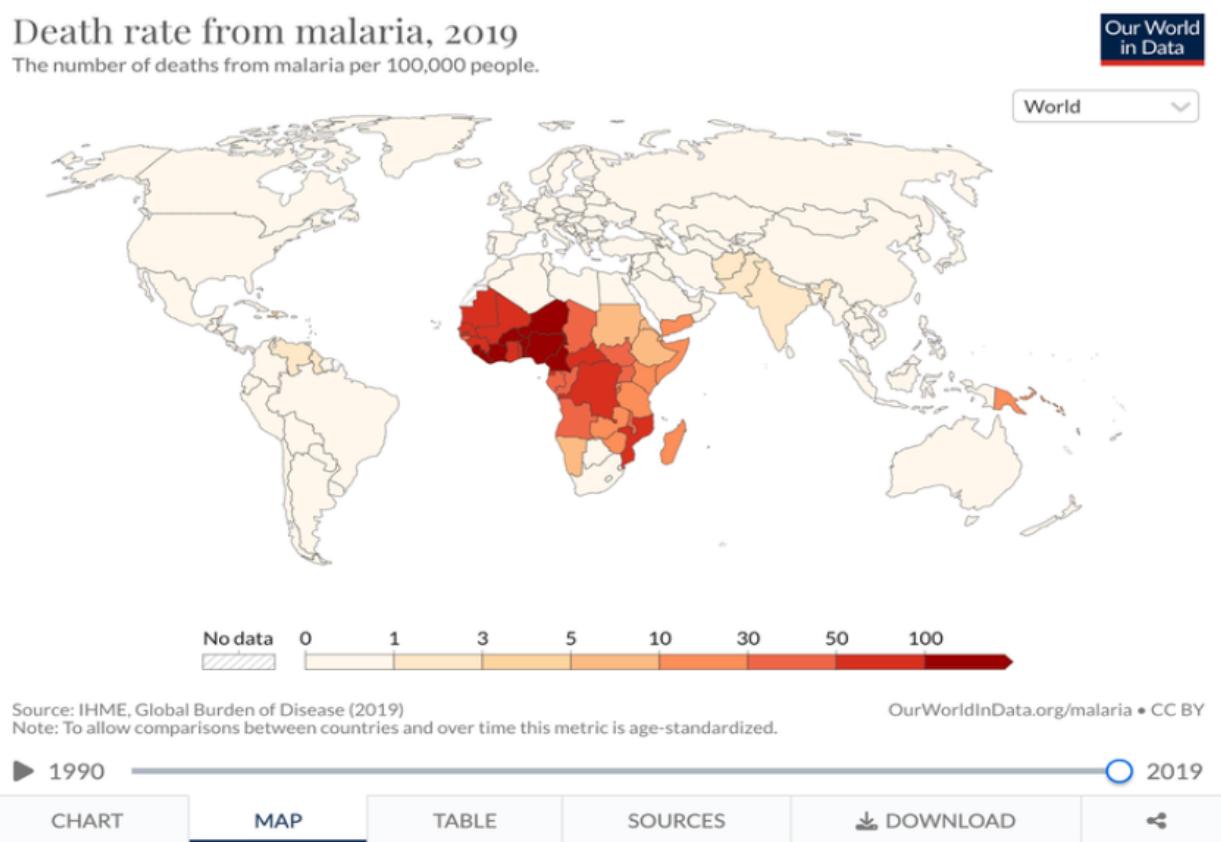
What is good about this visualization?

- It has a world map canvas and that helps us understand the geographical locations of the origin and spread.
- This has a time range bar where the era that we select shows the statistics of the spread in that particular year, rather than showing a cumulative spread.

What are the missing from this visualization?

Diversity, there is only one pandemic analyzed.

Having multiple pandemics visualized on one canvas helps us make constructive analysis of the relation between the pandemics.



Malaria Visualization map

1.3 Contribution

How is our project going to make a difference. We have a clear motivation and objective of what we need to build. We want to build a visualization that would help us analyze various patterns and trends of the various pandemics. From the motivations we conveyed how important the visualization is in analyzing, controlling, and communicating the important statistics of the pandemic spread to the public. There are a few existing visualizations discussed above that have their own limitations. We want our visualization to address all those to the extent possible. Most of the visualizations have only one pandemic analyzed and visualized, this is good enough but it could be better if there are some more pandemics being analyzed and visualized simultaneously. because then, we would have answers to some interesting questions like:

- Are all pandemics getting origin from the same country/continent?
- Which countries take a long time to recover? This helps us know what are the countries that do not have any stable medical support.
- Which countries have the best medical support system?
- What are most impacted countries?
- Is there any relation between all the pandemics?
- Who are targeted audience? This would be a global project, everyone across the globe could benefit from this information.

1.4 Objective

The primary objective of creating a cartograph to visualize pandemic attributes is to provide a clear and concise visual representation of the spread and impact of the pandemic over time. We aim to create a visualization that has the world map as its canvas and the spread of the pandemic is shown with an interactive time scroll bar. This info-graphics is aimed at discovering potential patterns of diseases spread between regions. The visualization represents the spread of pandemics through the timeline. The assumption is that every major pandemic that has occurred in history has a characteristic pattern that can be seen through which it spreads. This visualization will help users look at such complex transmission patterns and help devise new ways to counter them in case we need them in the near future. Thus we aim to produce cartographic visuals that at a glance inform us about the movement of death and spread using colors and shapes as visual encodings, filters such as timelines, and dropdowns to explore more than one pandemic on a single dashboard.

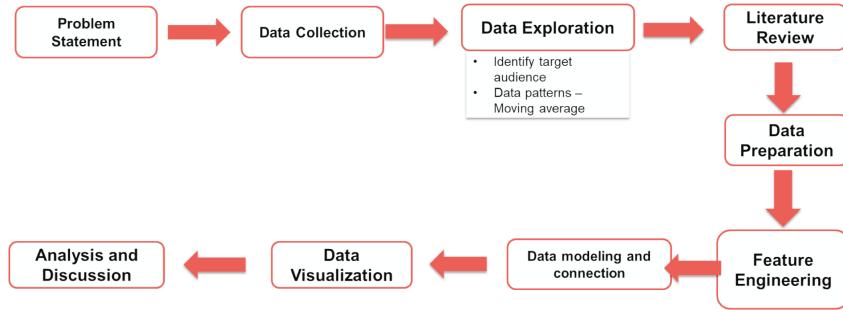
2. Data and Methods

Before we see the methodology, this section would give a brief explanation of some of the tools that we used for this project and what they are and how they work.

2.1 Process

Below is the pipeline diagram of our project, honestly this project feel much more harder than it looks. Major Challenges faced:

- Collecting Data: This was the toughest task, collecting the 3 pandemic data of all the countries in the world.
- Difficulty in accessing data: In some cases, it may be difficult to access the data that is needed for a study. This is due to a variety of factors like, data being proprietary or confidential, or the cost of obtaining the data.
- Data quality: Ensuring the quality of the data is important, as poor-quality data can lead to incorrect conclusions and poor decision-making. Data quality can be affected by a variety of factors, such as errors in data entry or data being collected using flawed methods.
- Required samples: acquiring required samples the 30 years of Pandemic data was tough.
- Learning Tableau was challenging as none of us know Tableau, we learned it from scratch and build all the visualization that were proposed. We also build analytical dashboards for each pandemic.



Overview

2.2 Ideas and prototype

The idea behind this project is to create a visualization using historic data of pandemics as mentioned above. The key idea behind the entire visualization is that every pandemic must follow some kind of spreading pattern. This pattern is what we were mainly looking to identify. Covid was an airborne disease, malaria was a carrier borne and HIV was an STD. We chose 3 different transmission methods to identify the spreading patterns of the 3 different mediums.

This was the hypothesis that we were mainly trying to test. We have maintained the standards structure of the dataframe of all the pandemics that we use for our project.

2.3 Data

To visualise the nature of epidemics throughout history, the timeline will include key information about each outbreak, including the number of people affected, the geographical spread, and the mortality rate and to expand the project we can include, the number of cases and the number of survivors. To acquire the said features we are used multiple methods as data availability in its clean form is sparse. Primarily we collected data using web scraping techniques. We have focused on 3 major pandemics Malaria, HIV, and Covid-19.

We obtained a lot of scattered data, which we later had to clean using python packages and power query. This helped us curate datamodels that cater exactly to our objective.

| Month-Year | location | Death |
|---------------------|----------|--------|
| 2/1/20 Afghanistan | | 0 |
| 3/1/20 Afghanistan | | 21 |
| 4/1/20 Afghanistan | | 812 |
| 5/1/20 Afghanistan | | 4924 |
| 6/1/20 Afghanistan | | 14982 |
| 7/1/20 Afghanistan | | 33439 |
| 8/1/20 Afghanistan | | 42108 |
| 9/1/20 Afghanistan | | 43060 |
| 10/1/20 Afghanistan | | 46274 |
| 11/1/20 Afghanistan | | 48777 |
| 12/1/20 Afghanistan | | 62050 |
| 1/1/21 Afghanistan | | 71891 |
| 2/1/21 Afghanistan | | 67864 |
| 3/1/21 Afghanistan | | 76227 |
| 4/1/21 Afghanistan | | 76300 |
| 5/1/21 Afghanistan | | 85386 |
| 6/1/21 Afghanistan | | 114208 |
| 7/1/21 Afghanistan | | 184478 |
| 8/1/21 Afghanistan | | 216817 |
| 9/1/21 Afghanistan | | 215185 |
| 10/1/21 Afghanistan | | 224427 |
| 11/1/21 Afghanistan | | 218866 |

Data Cleaning

The screenshot shows the Tableau interface with the following details:

- File Structure:** Shows connections to 'hiv_cleaned' (Text file), 'covid_clean' (Text file, currently selected), and 'malaria_clean' (Text file).
- Data Preview:** A preview of the 'covid_clean.csv' file is displayed, showing 5 fields and 7975 rows.
- Row Detail:** A specific row from the preview is expanded to show the mapping between the original field names ('F1', 'Month-Year', 'Location') and the new field names ('covid_clean.csv', 'Month-Year', 'Location').

Data Modelling

2.4 Visualization methods selection

The tidy data is then used to analyse key features which is used to characterize our visualization. The interactive visualization is achieved using the tool *Tableau*. The visualization is the world map indicating areas affected by the disease. The timeline bar is used to change the time period to show the spread of the disease from the origin to the other parts of the map. Filters will be available to add or remove additional disease data. Different colored legends are used to indicate different diseases on the map.

Tableau is a popular data visualization tool that is often used for creating maps and other types of charts and graphics. There are several reasons why Tableau is considered to be a good choice for creating maps:

Ease of use: Tableau is designed to be user-friendly, with an intuitive interface

and a wide range of built-in features that make it easy to create professional-quality maps and other visualizations.

Flexibility: Tableau allows users to customize and style their maps in a variety of ways, including choosing different map projections, customizing the color and style of map elements, and adding layers and data points to the map.

Large dataset support: Tableau can handle large datasets, making it well-suited for creating maps that show trends and patterns across large areas or over long periods of time.

Integration with other tools: Tableau integrates with a variety of other tools and platforms, making it easy to import data from a wide range of sources and to share and publish maps and other visualizations.

Overall, Tableau's combination of ease of use, flexibility, and large dataset support make it a popular choice for creating maps and other types of visualizations.

3. Results and Discussions

After putting in a lot of effort in obtaining the required clean data of the pandemics, it's time for us to visualize them on tableau. Technical Summary:

- We implemented timeline scroll bar using a dynamic timeline function.
- We can press pause and play buttons or drag a scroller to the desired timeline.
- We have chosen colors based on the spread of each pandemic, we tried to maintain the hue and lightness of the color in such a way that there is no overlap of colors.
- The visualizations are arranged in panels where each panel can be switched and visuals of different pandemics can be compared and analyzed

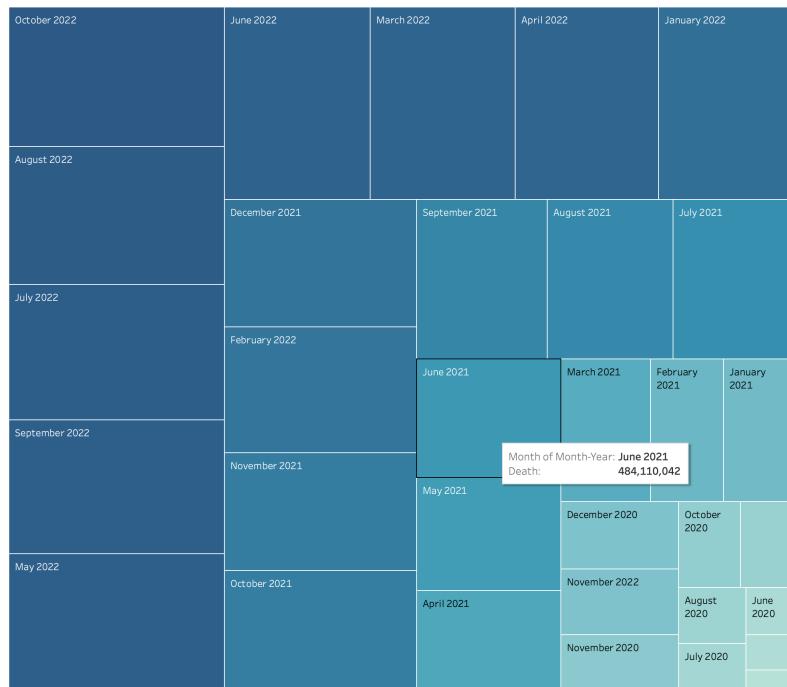
3.1 Covid 19 - Results

This section shows the results and visualizations obtained for the Covid 19 dataset.

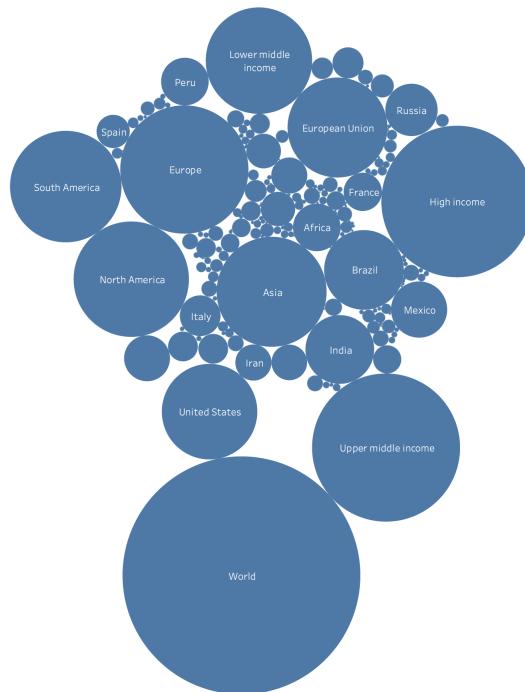
Tree Map: Figure : We have made a tree map of HIV cases observed yearly, this darker color mapping means highest cases and lighter color means few cases. The area also plays a role, the smaller the area means the lesser number of new cases. we could see that the time from 2000-2009 has seen majority of new cases.

Bubble Map: The above Figure : is a visualization to show country wise analysis of HIV impacted regions, we see that some of the African countries have the most largest bubbles compared to other countries.

Histogram: The above Figure : is the visualization of the Death count of the HIV over the years. We can see that the years 2000-2010 are the most impactful years of the pandemic.



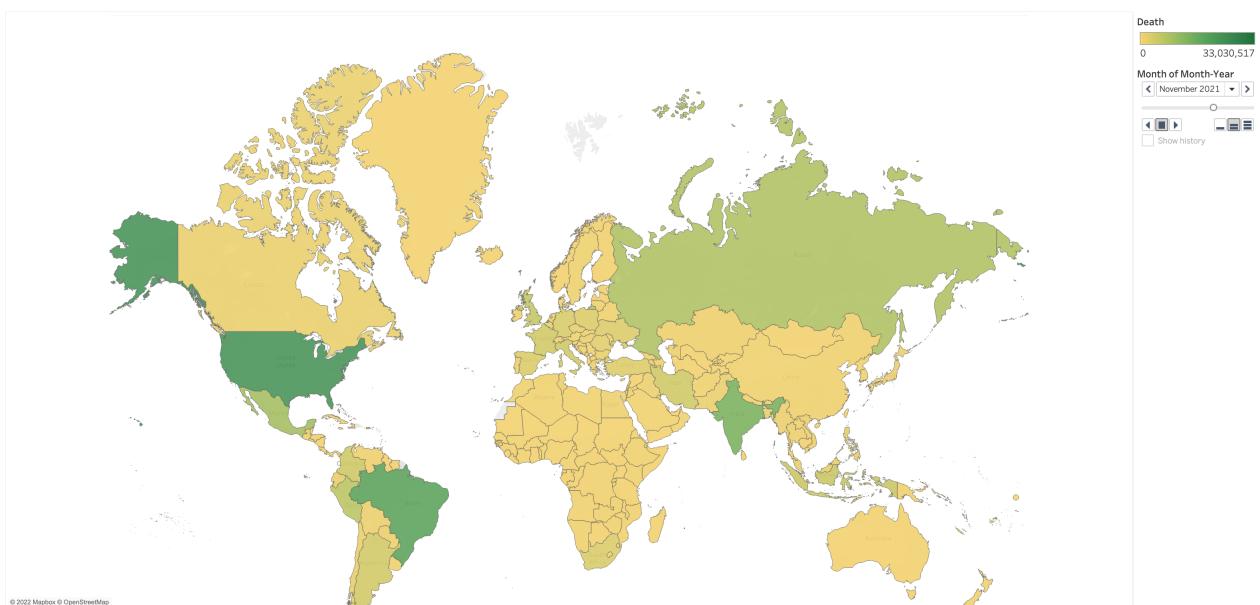
Covid - Tree Map



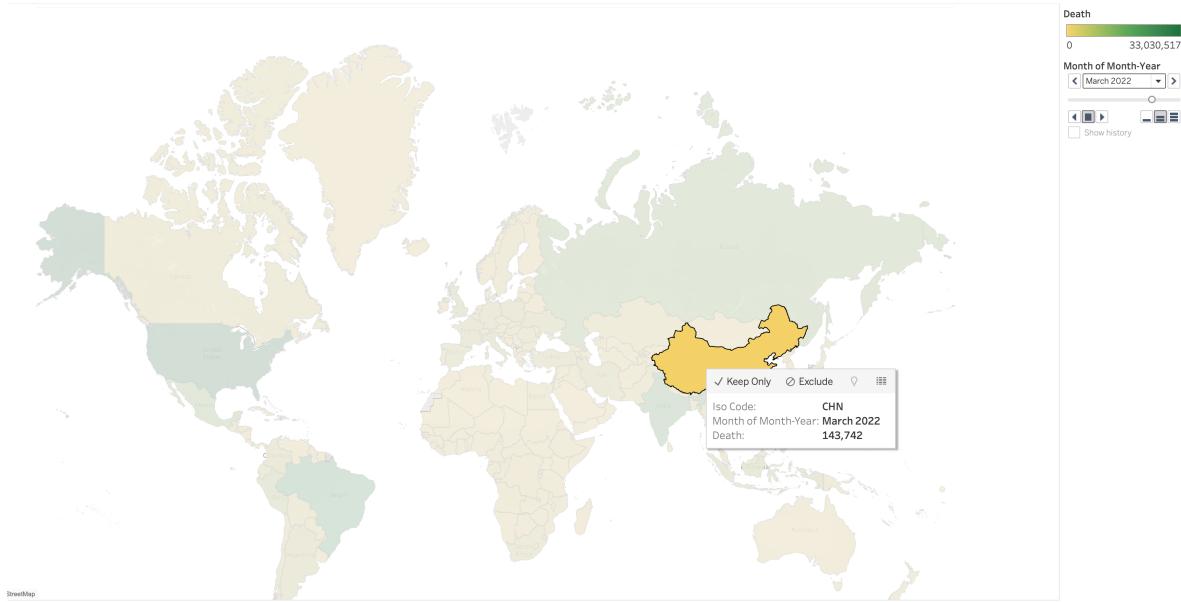
Covid - Bubble Map



Covid - Dot Histogram



Covid Map - Tableau



Covid Map - country specific

3.2 Malaria Results

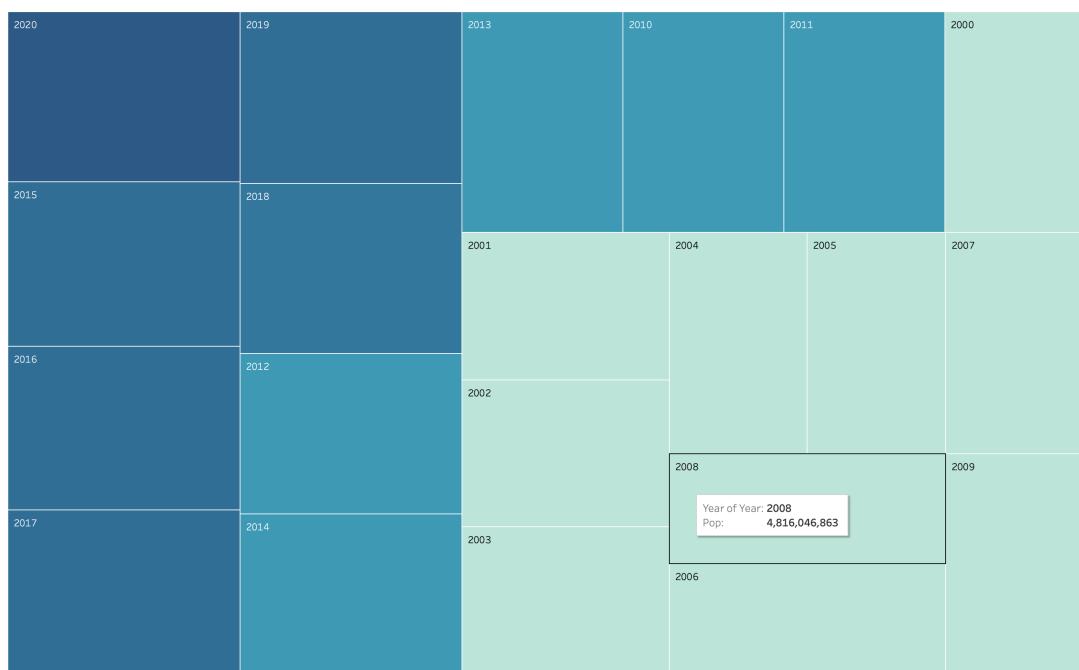
This section shows the results and visualizations obtained for the Malaria dataset.

Tree Map: Figure : We have made a tree map of Malaria cases observed yearly, this darker color mapping means highest cases and lighter color means few cases. The area also plays a role, the smaller the area means the lesser number of new cases. we could see that the time from 2000-2003 and 2010-2012 has seen majority of new cases.

Bubble Map: The above Figure : is a visualization to show country wise analysis of malaria impacted regions, we see that India has the Darkest colored among all the countries

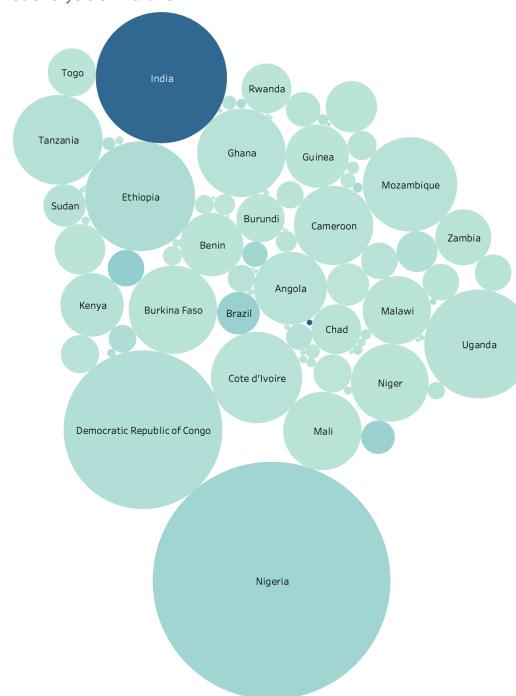
Scatter Plot: Figure : is the scatter plot of the number deaths and the years. Here we can observe the pattern of the increasing cases in 2010 and then the wave of decreases cases till 2020.

Malaria Yearly Heatmap



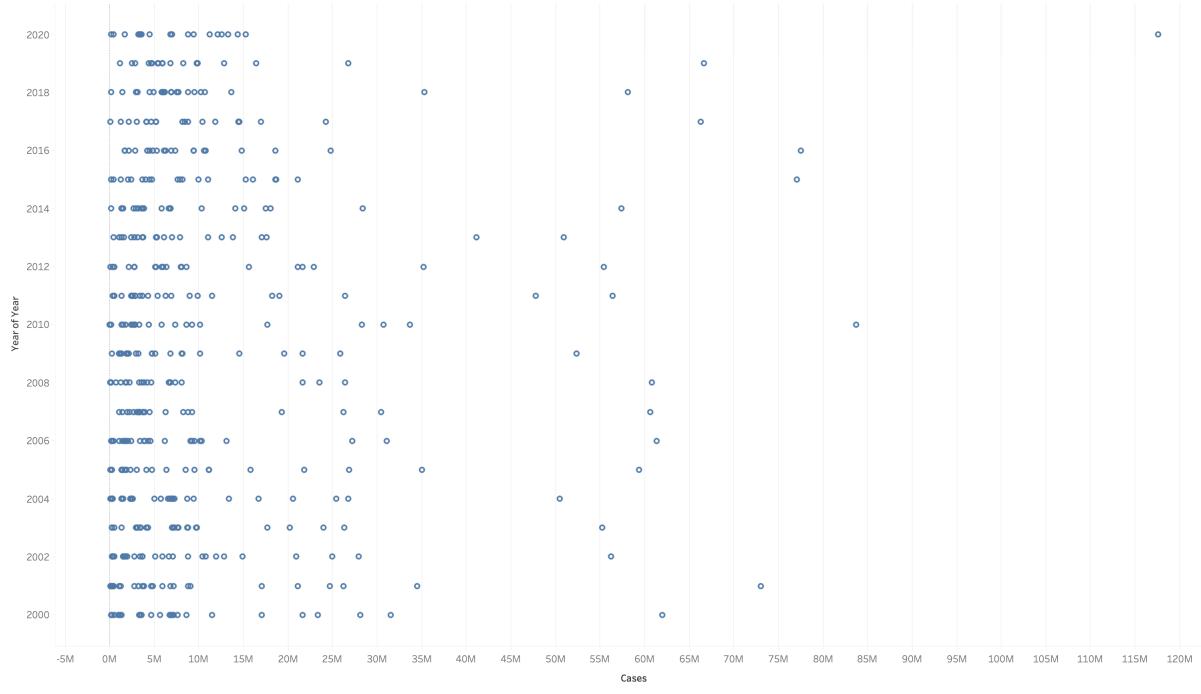
Malaria - Tree Map

Country wise analysis of Malaria

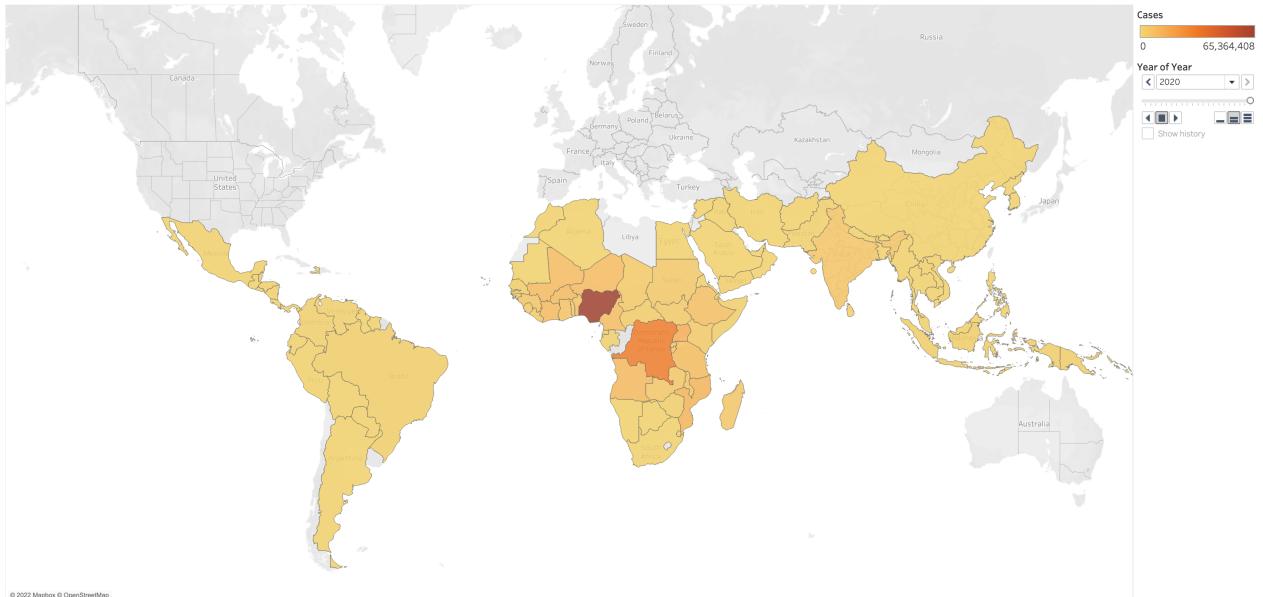


Malaria - Bubble Map

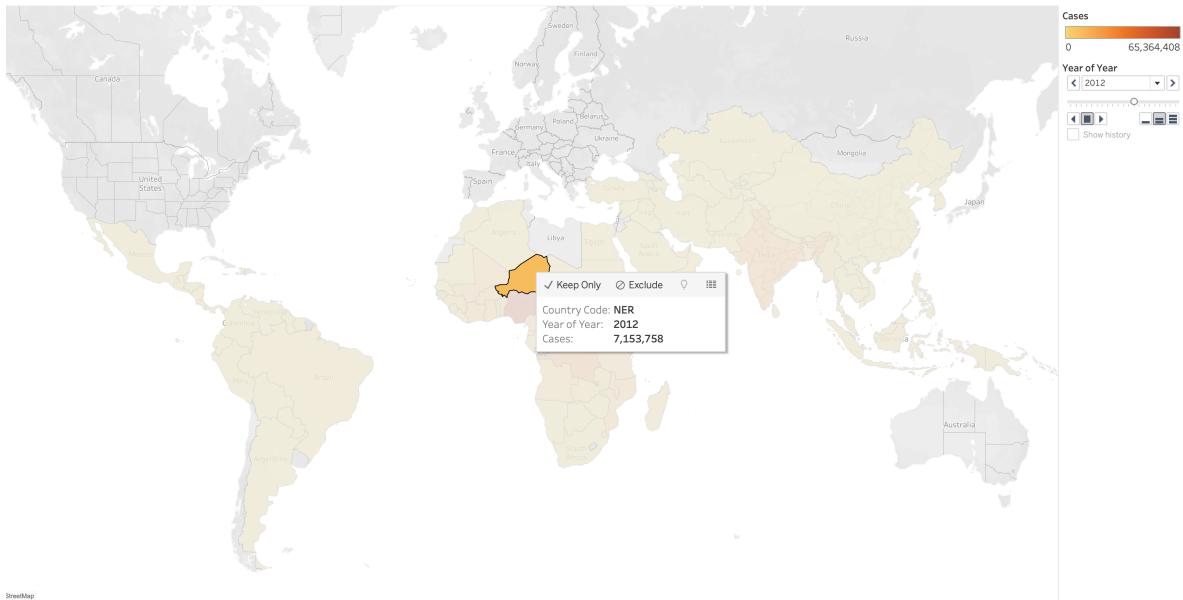
Yearly analysis of Malaria Cases



Malaria - Dot Histogram



Malaria Map - Tableau



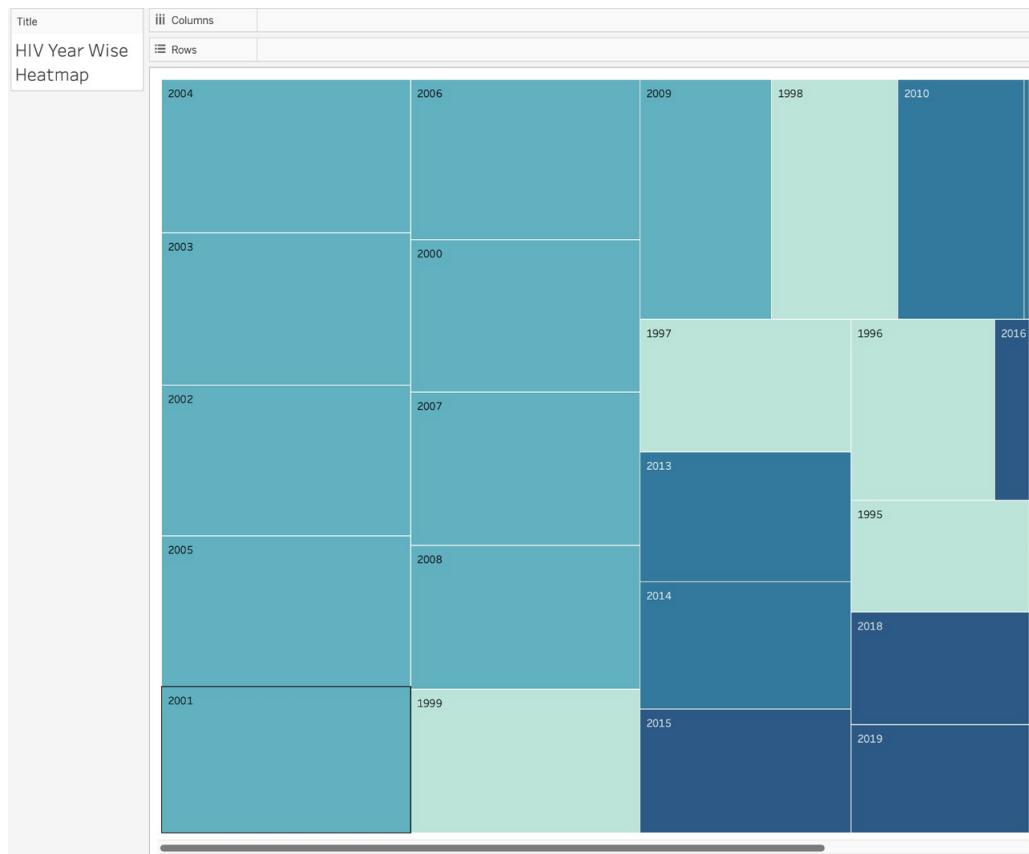
Malaria Map - Country Specific

3.3 HIV Results

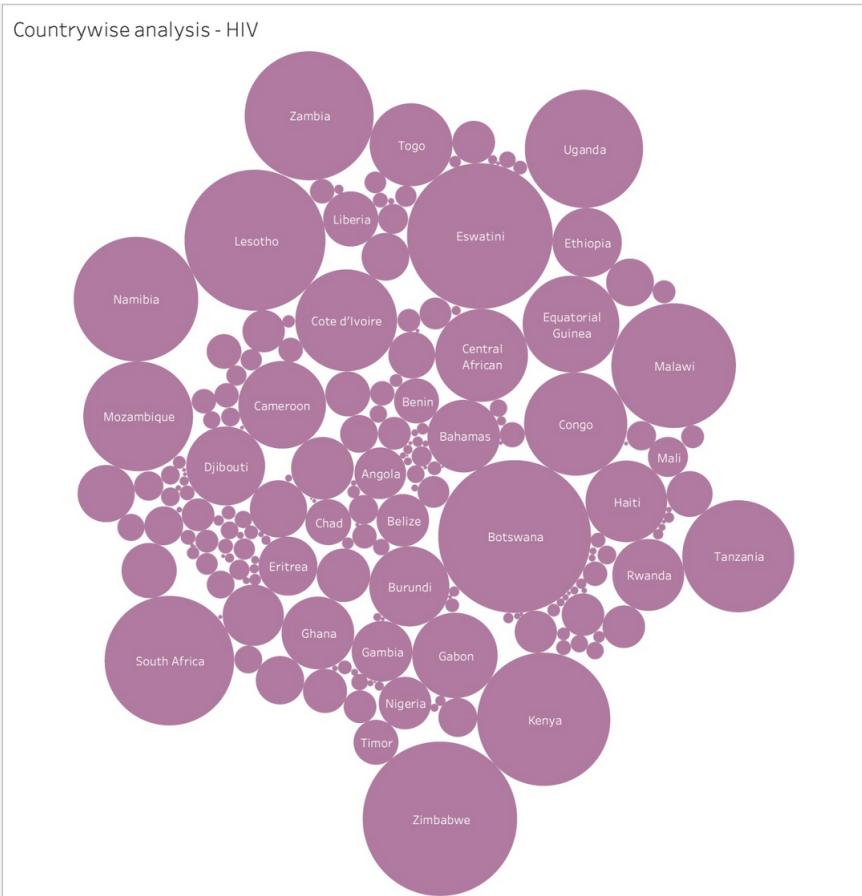
This section shows the results and visualizations obtained for the HIV dataset. Tree Map: Figure : We have made a tree map of covid-19 cases observed monthly, this darker color mapping means highest cases and lighter color means few cases. The area also plays a role, the smaller the area means the lesser number of new cases.

Bubble Map: The above Figure : is a visualization to show country wise analysis of covid 19 impacted regions, we see world where the major part of the world apart from the major countries is represented.

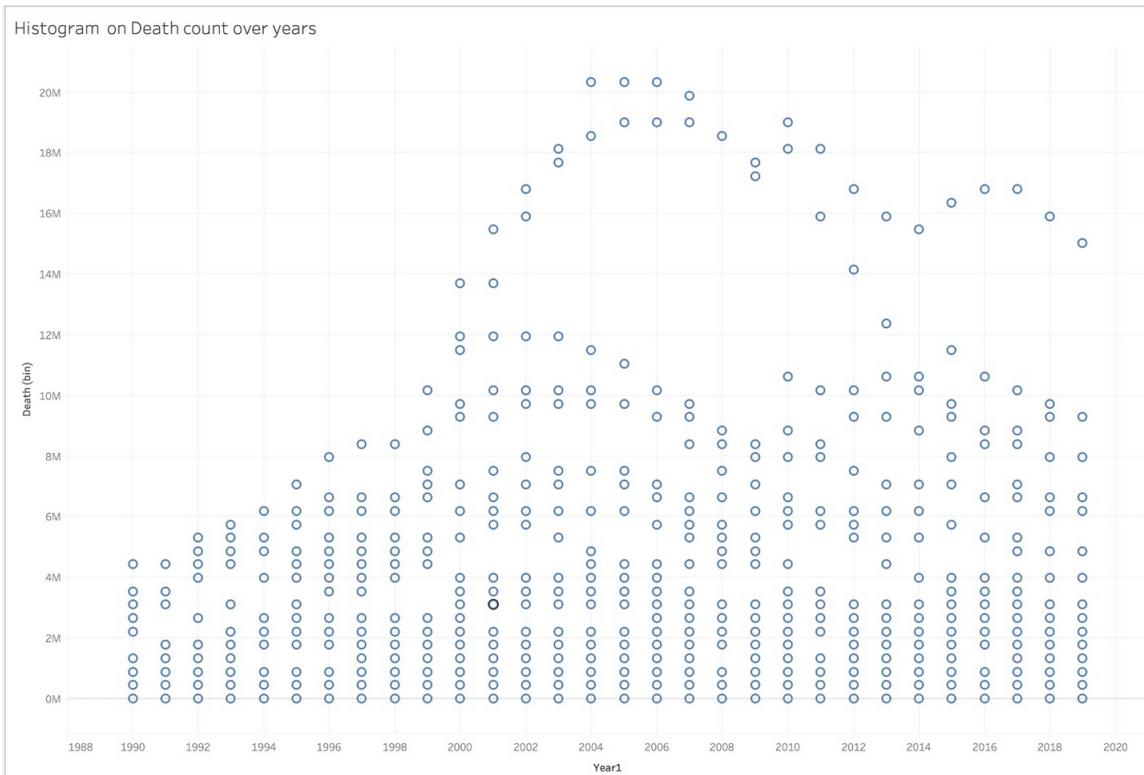
Histogram: Figure : represent the death count of the covid-19 in three years. 2020 -2022. Each circle represents the country, the mortality rate observed for that particular year.



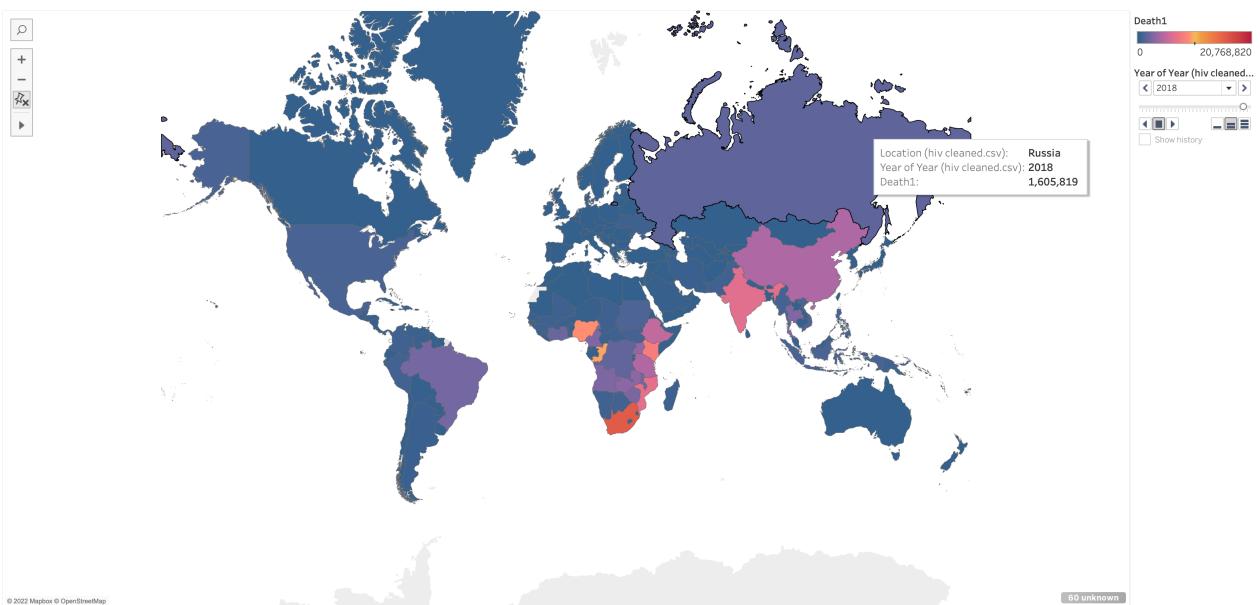
HIV - Tree Map



HIV - Bubble Map



HIV - Dot Histogram



HIV Map - Tableau



HIV Map - Country Specific

4. Conclusions

It is important to effectively visualize pandemic data in order to accurately understand and communicate the spread and impact of the disease. Cartography, or the creation of maps, is a valuable tool for this purpose. By representing data on a map, we can easily see spatial patterns and relationships, such as the distribution of cases within a region or the proximity of cases to each other.

There are many different ways to visualize pandemic data as a cartograph, including choropleth maps, dot density maps, and graduated symbol maps. Each of these techniques has its own strengths and weaknesses, and it is important to choose the appropriate method based on the data and the message that you want to convey.

Effective cartographic visualization of pandemic data can help policymakers, healthcare professionals, and the general public to better understand the situation and make informed decisions. It is important to consider the limitations of the data and the map, and to carefully design and label the map to clearly communicate the information.

Overall, cartographic visualization can be a powerful tool for understanding and communicating pandemic data, and it is an important part of the larger efforts to control and mitigate the spread of the disease.

4.1 Analysis

- Malaria is caused by mosquitoes, and its spread is observed in tropical countries, cold countries are least affected by malaria
- The spread of Covid-19 is coherently observed among the countries bordering the infected countries

- The effect and recovery in the past 30 years among the countries due to HIV are clearly observed proving the measures taken are effective

4.2 Future Scope

- Adding some more Pandemics data to make more analysis, and learn more patterns
- We can host this visualization on a cloud server, and make it available for global public.

5. References

- [1] J. A. Salomon et al., The US COVID-19 Trends and Impact Survey: Continuous real-time measurement of COVID-19 symptoms, risks, protective behaviors, testing, and vaccination. Proc. Natl. Acad. Sci. U.S.A. 118, e2111454118 (2021)
- [2] Kassam, N.A., Kaaya, R.D., Damian, D.J. et al. Ten years of monitoring malaria trend and factors associated with malaria test positivity rates in Lower Moshi. Malar J 20, 193 (2021). <https://doi.org/10.1186/s12936-021-03730-1>
- [3] Riley and Ferguson, 2006, S. Riley, N.M. Ferguson , Smallpox transmission and control: spatial dynamics in Great Britain, Proc. Natl. Acad. Sci. USA (2006), 10.1073/pnas.0510873103
- [4] <https://coronavirus.jhu.edu/map.html>
- [5] <https://press.uchicago.edu/ucp/books/book/chicago/D/bo8490164.html>
- [6] <https://www.visualcapitalist.com/history-of-pandemics-deadliest/>
- [7] <https://ourworldindata.org/malaria>
- [8] <https://ourworldindata.org/hiv-aids>
- [9] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7874133/>
- [10] <https://www.ecohealthalliance.org/2017/10/global-disease-hotspots-2-0>
- [11] <https://www.frontiersin.org/articles/10.3389/fneur.2021.634827/full>