TIH AUTUMN INTERNSHIP 2025 PROJECT REPORT

VISUALIZING TIME SERIES DATA OF COVID-19 PANDEMIC AND EBOLA OUTBREAK

MAHIGUHAPPRIYAPRAKASH

RESEARCH TRAINEE (OBSERVATIONS)
INDIAN INSTITUTE OF ASTROPHYSICS, BENGALURU

PERIOD OF INTERNSHIP: 25TH AUGUST 2025 - 19TH SEPTEMBER 2025

REPORT SUBMITTED TO:

IDEAS –INSTITUTE OF DATA ENGINEERING, ANALYTICS AND SCIENCE FOUNDATION, ISI KOLKATA

DATE OF SUBMISSION
19 SEPTEMBER 2025

Abstract

This project attempts to explore the use of data visualization techniques offered by python to analyze and interpret epidemic time series data, with a focus on COVID-19 pandemic and Ebola outbreak (2014-2016). The primary objective is to understand trends, patterns, and regional variations in disease spread through effective graphical representations. Using the COVID-19 dataset provided, interactive visualizations were developed to illustrate global and regional case counts, deaths, and temporal dynamics. These visualizations highlight key phases of the pandemic such as major waves driven by emerging variants and the impact of interventions like vaccination. The methods applied to the COVID-19 dataset were then extended to the Ebola outbreak data, enabling comparative insights into the temporal and geographic progression of two distinct infectious diseases. The project demonstrates how visualization not only enhances comprehension of complex data but also supports meaningful communication of public health information.

Contents

List of Figures

1 INTRODUCTION

1.0.1 *What is a Time Series Dataset?*

A *time series* is an ordered collection of measurements taken at regular intervals [IBM]. It can be used to predict an outcome of an event based on previous trend of the data. Studying a time series data can help us determine the patterns and isolate out the sudden pulses to further study the cause of their occurrences. A time series data has the following characteristics:

- **Trends:** Long-term increase or decrease
- **Seasonality:** Repeating patterns
- **Noise:** Random fluctuations in data
- **Autocorrelation:** Correlation with past values
- **Non-stationary:** Changing statistical measures (mean/variance) over time
- **Outliers:** Sudden large deviations
- **Structural Breaks:** Abrupt changes due to interventions
- **Lag Effects:** Delayed impact of events
- **Cyclic Behavior:** Longer irregular cycles.

Figure 1: Medical professionals rejoicing after the pandemic was brought to an end. Source: CNN

In Chapter 2, I will elaborate the tools and methods used for the time series analysis of COVID-19 and Ebola outbreak data. Following that, the whole process of analysis and the results drawn from them are described in detail in Chapter 3. In Chapter 4, I have summarized my work. The Python notebooks, datasets and other supporting documents can be found in my GitHub Repository.

1.1 PROJECT OBJECTIVE

The objective of this project is to apply data visualization techniques to epidemic time series data in order to identify trends, patterns, and regional variations in disease spread. Ultimately, the project seeks to demonstrate how visualization can enhance the understanding of epidemiological datasets and support effective communication of public health insights.

2

METHODOLOGY

The project employed a structured approach to explore epidemic time series datasets for COVID-19 (2020–2023) and the Ebola outbreak in West Africa (2014–2016). Both datasets were provided by the institute in tabular format, containing country-level daily or weekly records of confirmed cases, deaths, and cumulative totals. The analysis was conducted using Python and Jupyter Notebook as the primary environment.

Data Preprocessing

The datasets were ingested and processed using the `pandas` library. Preprocessing steps included:

- Conversion of date fields into standardized `datetime` objects to facilitate time-series operations.
- Handling of missing values through either removal or forward-filling, depending on the context.
- For Ebola, where reporting was sparser and at times weekly, interpolation was avoided to preserve the integrity of limited but high-impact events.

Visualization Tools

The project used:

- **Matplotlib** and **Seaborn** for line charts, bar plots, and stacked visualizations.
- **Plotly** for interactive graphs with hover and zoom functionalities, allowing dynamic exploration of peaks and troughs.
- **Plotly.express** for choropleth maps to display regional variation in case and death counts.

Workflow

The methodology followed a layered structure:

1. Exploratory analysis of distributions and raw counts.
2. Construction of time-series plots to identify temporal dynamics.
3. Comparative visualization across countries and WHO regions.
4. Application of the same visualization pipeline to the Ebola dataset, enabling cross-disease comparison.

3 DATA ANALYSIS AND RESULTS

3.1 DATA ANALYSIS

COVID-19 Pandemic

The COVID-19 dataset revealed a multi-wave progression of the pandemic:

- The first global wave emerged in early 2020, with Europe and North America as the epicenters (Figure 4) (6).

- Subsequent peaks corresponded to the emergence of variants: Alpha (late 2020), Delta (mid-2021), and Omicron (late 2021), each producing distinct spikes in global cases (Figure 2) (8; 4).

- As from the Figure 2 it is evident that there is a sharp spike in the new cases in January 2023. The possible reason might be that the Zero COVID policy was withdrawn by China in late December 2022. From the country-wise analysis we could infer that the major contributor to the spike is China.(9)

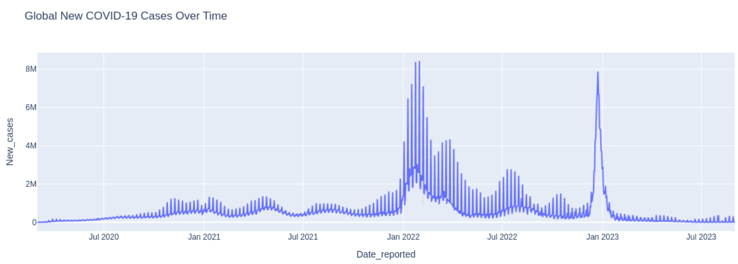

Figure 2: Global daily new COVID-19 cases, 2020–2023. The Omicron wave shows the highest recorded peak.

- Deaths (Figure 3) tracked closely with cases during early waves but decoupled by 2022, reflecting widespread vaccination, prior immunity, and lower intrinsic severity of Omicron (3).

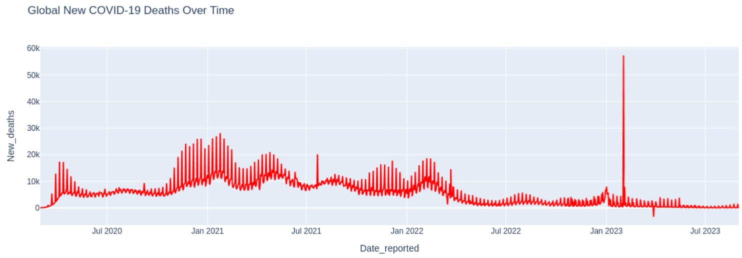

Figure 3: Global daily new COVID-19 death, 2020–2023.

- Geographically, the United States, India, Brazil, and later China dominated global totals, while Africa reported fewer cases, largely due to under-reporting and testing limitations (1).

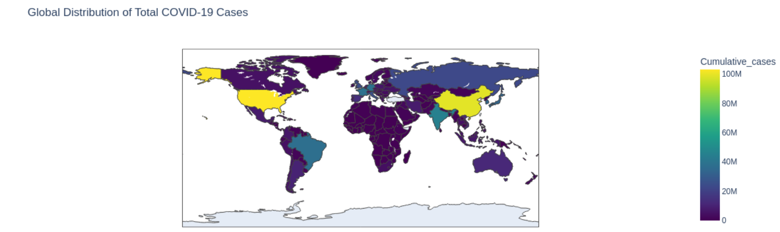

Figure 4: A map showing the global cumulative COVID-19 cases.

Ebola Outbreak (2014–2016)

The Ebola dataset highlighted the different dynamics of a localized epidemic:

- The outbreak was concentrated in Guinea, Liberia, and Sierra Leone, which together accounted for the vast majority of the 28,600 reported cases and 11,325 deaths (5).

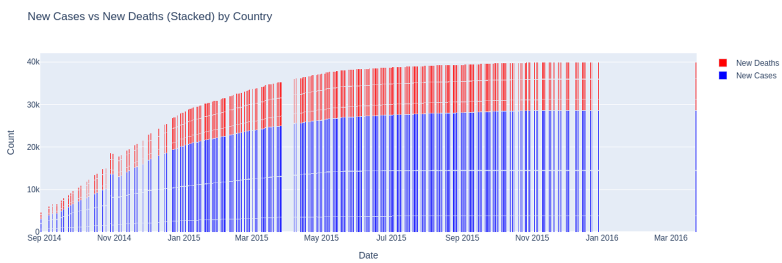

Figure 5: Global cumulative Ebola cases and deaths, 2014–2016.

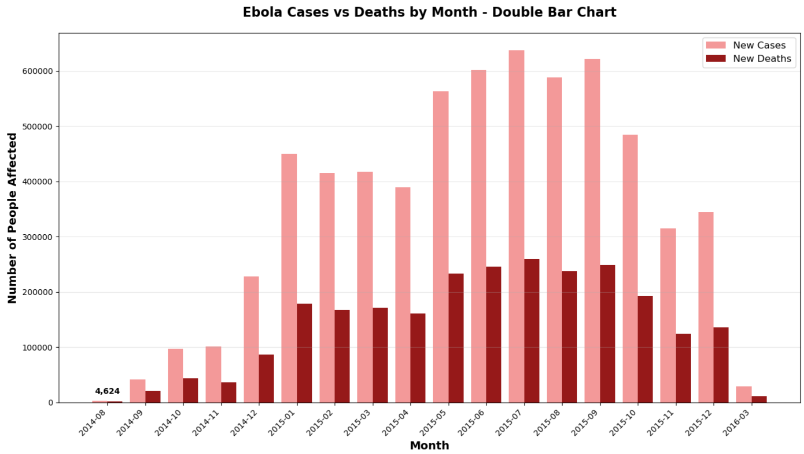

Figure 6: Double bar plot of the global cumulative Ebola cases and deaths.

- Unlike COVID-19, the Ebola epidemic showed fewer but sharper peaks, reflecting localized surges and the effect of containment interventions such as quarantines and community engagement (2).
- Case fatality rates were far higher (averaging around 40–50%) compared to COVID-19's global fatality rate of less than 2% (7).
- Limited geographic spread was due to Ebola's transmission mode (direct contact with bodily fluids), which contrasts with the airborne transmission of SARS-CoV-2(Figure 7).

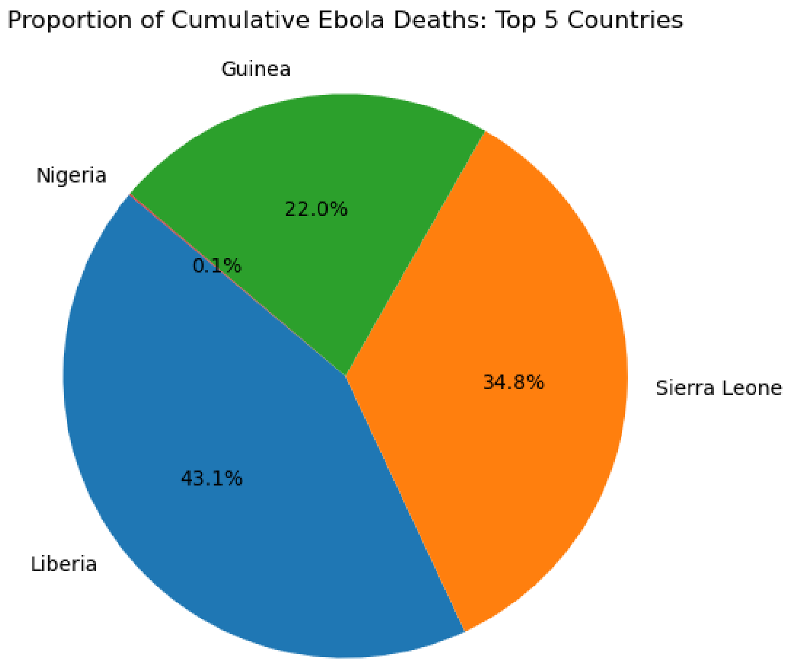

Figure 7: Pie chart distribution of the top countries that were affected by the Ebola outbreak.

Comparative Insights

Comparing the two outbreaks underscores key epidemiological differences:

- COVID-19 spread globally within months, while Ebola remained regionally confined despite its severity.
- Visualization of COVID-19 data shows recurring global waves, whereas Ebola data reflect localized outbreaks with sharp mortality.
- Intervention effects are visible in both datasets: lock downs and vaccination reduced COVID-19 deaths, while community surveillance and international aid curbed Ebola transmission.

3.2 RESULTS

The visualization-driven analysis produced several insights:

1. COVID-19 exhibited recurring global waves linked to viral evolution and human interventions.
2. Mortality burden was highest during the Delta wave, while Omicron produced unprecedented case numbers but proportionally fewer deaths.
3. Ebola, although geographically limited, had a much higher case fatality rate and posed severe challenges to fragile health systems.
4. Visualization proved effective for identifying temporal and geographic patterns, making complex epidemiological data more interpretable.

4 CONCLUSION

This project set out to explore the potential of data visualization in understanding epidemic time series data, with a focus on the COVID-19 pandemic and the 2014–2016 Ebola outbreak. By applying interactive and static visualization methods using Python libraries such as Plotly and Matplotlib, complex epidemiological data were translated into accessible and interpretable graphical representations.

For the COVID-19 dataset, global and regional analyses revealed the multi-wave nature of the pandemic, with distinct surges corresponding to the emergence of major variants such as Alpha, Delta, and Omicron. Interactive line charts and choropleth maps highlighted not only the scale of transmission worldwide but also the uneven regional burden, underscoring disparities in health systems, testing capacity, and policy interventions. Specific patterns, such as the January 2023 surge in China following the end of the zero-COVID policy, demonstrated the importance of connecting data trends with real-world policy and epidemiological developments.

The Ebola outbreak analysis provided a contrasting case study, marked by localized but severe impacts in West Africa. Visualization of the Ebola data highlighted the concentrated nature of cases and deaths, the temporal dynamics of outbreak peaks, and the role of international intervention in controlling spread. Comparing Ebola and COVID-19 emphasized both the diversity of epidemic trajectories and the universal importance of timely, reliable data.

Overall, this project demonstrates that visualization is not merely a tool for presentation but a method of inquiry in its own right. Well-designed visualizations reveal patterns that might remain hidden in raw numbers and allow researchers, policymakers, and the public to grasp the dynamics of health crises more effectively. By applying the same methodological framework to two different epidemics, the project reinforces the generalization of visualization as a strategy for epidemic data analysis.

BIBLIOGRAPHY

World Health Organization Regional Office for Africa. Covid-19 in africa: Data challenges and lessons, 2021. URL: https://www.afro.who.int/news/covid-19-africa-data-challenges-and-lessons. (Cited on page 10.)

Centers for Disease Control and Prevention. 2014â2016 ebola outbreak in west africa, 2019. URL: https://www.cdc.gov/vhf/ebola/history/2014-2016-outbreak/index.html. (Cited on page 11.)

DW News. Who says covid-19 is no longer a global health emergency, 2023. URL: https://www.dw.com/en/who-says-covid-19-is-no-longer-a-global-health-emergency/a-65530462. (Cited on page 9.)

UN News. Who chief declares end to covid-19 as a global health emergency, 2023. URL: https://news.un.org/en/story/2023/05/1136367. (Cited on page 9.)

World Health Organization. Ebola situation report â march 2016 (final), 2016. URL: https://apps.who.int/ebola/current-situation/ebola-situation-report-30-march-2016. (Cited on page 10.)

World Health Organization. Coronavirus disease 2019 (covid-19) situation report â 51, 2020. URL: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200311-sitrep-51-covid-19.pdf. (Cited on page 9.)

World Health Organization. Ebola virus disease: Key facts, 2021. URL: https://www.who.int/news-room/fact-sheets/detail/ebola-virus-disease. (Cited on page 11.)

World Health Organization. Tracking sars-cov-2 variants, 2022. URL: https://www.who.int/activities/tracking-SARS-CoV-2-variants. (Cited on page 9.)

India Today. Unreliable data, new variants? here's why covid outbreak in china is causing concern, 2023. URL: https://www.indiatoday.in/world/story/why-covid-outbreak-in-china-is-causing-concern-2316699-2023-01-03. (Cited on page 9.)