Visualization of Ebola Data

**Miss. Riddhi Palorkar,**

**B.Sc -2026, Hislop College, Nagpur.**

Period of Internship: 25th August 2025 - 19th September 2025

**Report submitted to: IDEAS – Institute of Data**

**Engineering, Analytics and Science Foundation, ISI, Kolkata.**

**1.Abstract**

The Western African Ebola virus epidemic (2013–2016) was the largest recorded outbreak of Ebola virus disease (EVD), beginning in Guinea and rapidly spreading to Liberia and Sierra Leone. The epidemic caused severe health, social, and economic disruption across the region. By its end, over 28,000 suspected, probable, and confirmed cases were reported, including more than 11,000 deaths, with an overall case fatality rate of about 40%.

This project analyses the Ebola dataset to explore the progression and impact of the epidemic. The dataset includes variables such as reporting date, country, suspected cases, and deaths. Using this data, we examine temporal trends, compare country-level impacts, and calculate case fatality rates.

The goal is to provide a data-driven understanding of outbreak dynamics and highlight how statistical analysis and visualization can support public health decision-making.

**2.Introduction**

The purpose of this research is to analyse the 2013–2016 Western African Ebola epidemic using reported data in order to understand how the outbreak progressed over time, how it affected different countries, and how severe it was (through measures like case fatality rates). By applying statistical analysis and visualization, the study aims to provide data-driven insights that can support public health decision-making, improve outbreak preparedness, and highlight lessons for managing future epidemics.

We received training on the following topics during the 2 week training session:-

Python & Programming Basics

* Variables, Lists, Loops – Think of variables as containers for information, lists as collections of items, and loops as a way to repeat tasks without writing the same code again.
* Data structures – Tools like lists, sets, and dictionaries help organize data so it’s easier to store, search, and use.
* Classes & OOP – Object-Oriented Programming lets us model real-world things in code, making programs more reusable and easier to manage.
* NumPy & Pandas – NumPy speeds up math with arrays, while Pandas makes handling tables of data simple and powerful.

Machine Learning Essentials

* Regression Lab – Learn how to predict numbers, like house prices or temperatures, using simple models.
* Classification Lab – Explore how to sort things into groups, like whether an email is spam or not.
* LLM Fundamentals – Get to know Large Language Models (like ChatGPT) and how they understand and generate human-like text.
* Communication Skills – Practice sharing ideas and results clearly so both technical and non-technical people can understand them.

**3.Project Objective**

The main objectives of this project are:

* **To analyze the progression of the Ebola outbreak (2013–2016)** using time-series data on suspected cases and deaths.
* **To compare the impact across countries** and identify the most severely affected regions.
* **To examine temporal patterns** (monthly and quarterly trends) to highlight peak phases of transmission and fatalities.
* **To visualize the geographical spread** of Ebola using heatmaps and interactive dashboards for better interpretability.
* **To assess outbreak severity through case fatality rates (CFRs)** and provide data-driven insights into the epidemic’s public health impact.

**4. Methodology**

The Ebola dataset was analyzed using Python with libraries such as **pandas**, **numpy**, **matplotlib**, **seaborn**, and **plotly**. A structured approach was followed, with each step explained below:

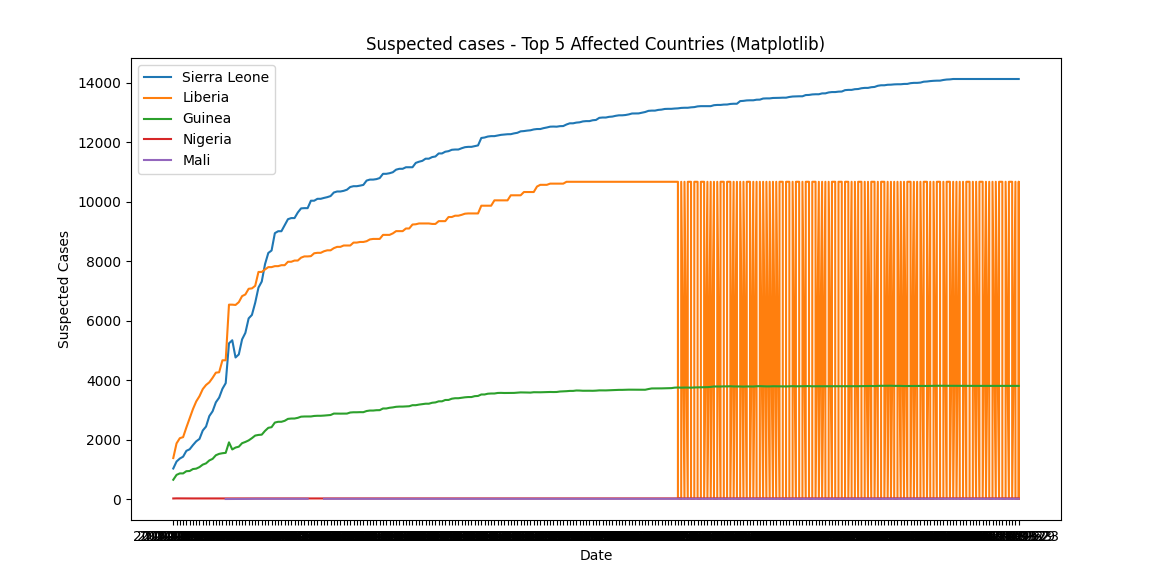
1. **Data Import and Preparation**
   * The dataset was loaded into a Pandas DataFrame using read\_csv().
   * To make analysis easier, long column names were renamed to shorter and more descriptive labels such as Sus\_cases (for suspected cases) and Sus\_death (for suspected deaths).
   * The Date column was converted into a proper datetime format so that time-based operations like grouping by month, quarter, or year could be performed accurately.
2. **Data Cleaning**
   * The dataset was inspected for missing or inconsistent values.
   * A copy of the DataFrame was created for grouped and transformed analyses, ensuring the raw dataset remained unchanged.
   * Time-based features such as Month-Year and Quarter were derived from the Date column, enabling more granular trend analysis over different time periods.
3. **Exploratory Analysis**
   * Initial descriptive analysis was carried out to identify the **top five most affected countries** based on the maximum cumulative suspected cases reported.
   * For these countries, line plots were generated using **Matplotlib** to visualize the trajectory of suspected cases over time, providing an early view of how the outbreak spread differently across regions.
4. **Quarterly Aggregation**
   * To capture epidemic dynamics over larger time frames, the dataset was aggregated by Quarter.
   * The total suspected cases and deaths were computed for each quarter.
   * A stacked bar chart was then created to compare the quarterly counts of cases and deaths, with totals annotated on top of each bar. This highlighted both the scale and severity of the outbreak across different phases.
5. **Heatmap Visualization**
   * Monthly aggregation was performed by grouping data by Country and Month-Year.
   * A pivot table was created with countries as rows and months as columns, with values representing total cases.
   * Using **Seaborn**, a heatmap was plotted to visually capture the geographic and temporal spread of Ebola cases. The color intensity represented outbreak severity, making it easy to identify peak months for different countries.
6. **Interactive Dashboards**
   * To enhance interpretability, interactive visualizations were created using **Plotly**.
   * A line chart was plotted to show **global suspected cases over time**, providing a cumulative picture of outbreak growth.
   * A second line chart was plotted for **global deaths**, using a red color scale for emphasis.
   * A stacked bar chart was built with cases and deaths plotted together by country and date, allowing for comparison of regional impacts interactively. Users could hover over data points to see exact counts.
7. **Insight Extraction**
   * From the aggregated datasets, the **quarter with the highest number of suspected cases** was identified.
   * Key insights such as temporal peaks, country-level differences, and the case-to-death ratio were extracted.
   * These findings provided evidence of the outbreak’s timeline, the most severely impacted countries, and the scale of loss during the epidemic.

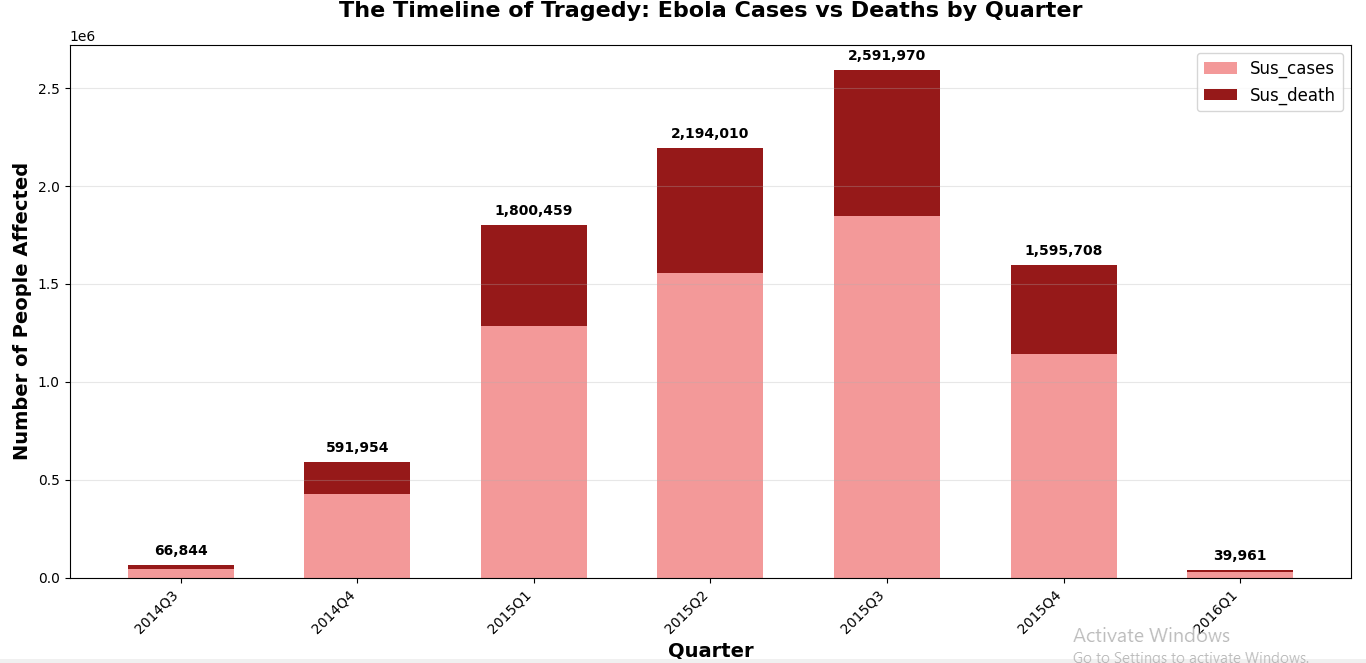
The link to Github:-

https://github.com/riddhipalorkar/Visualizing-\_Time\_Series\_Data.git

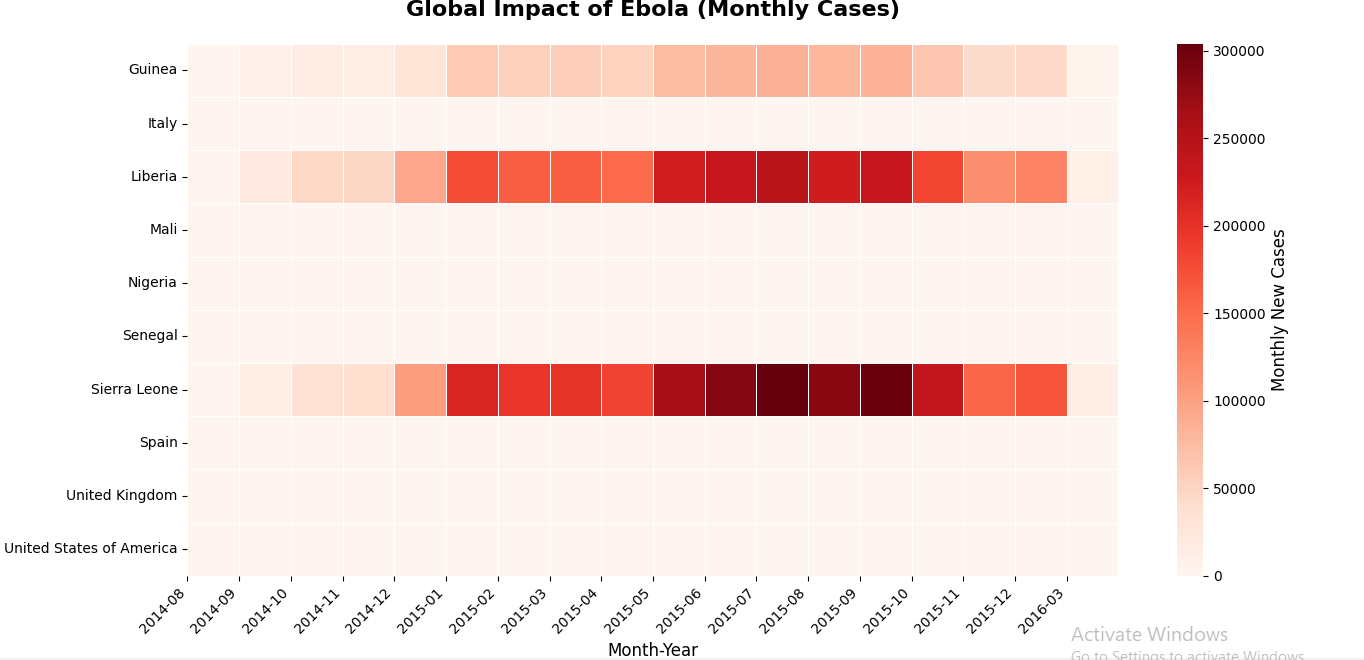
5. Data Analysis and Results

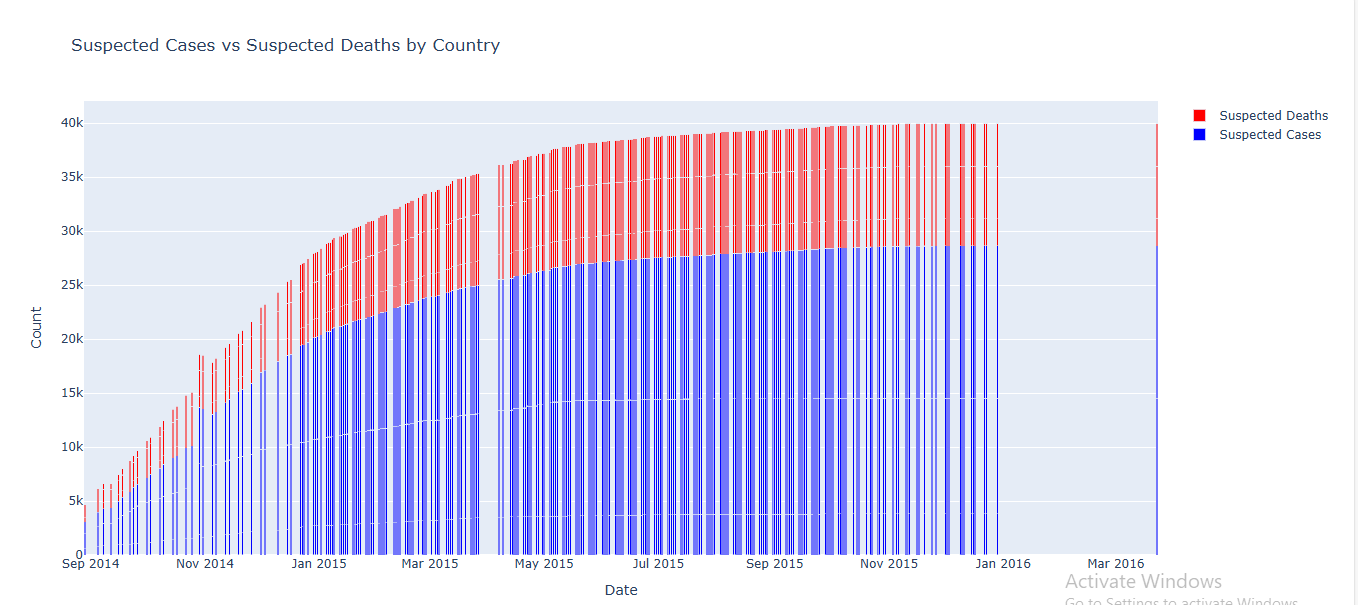
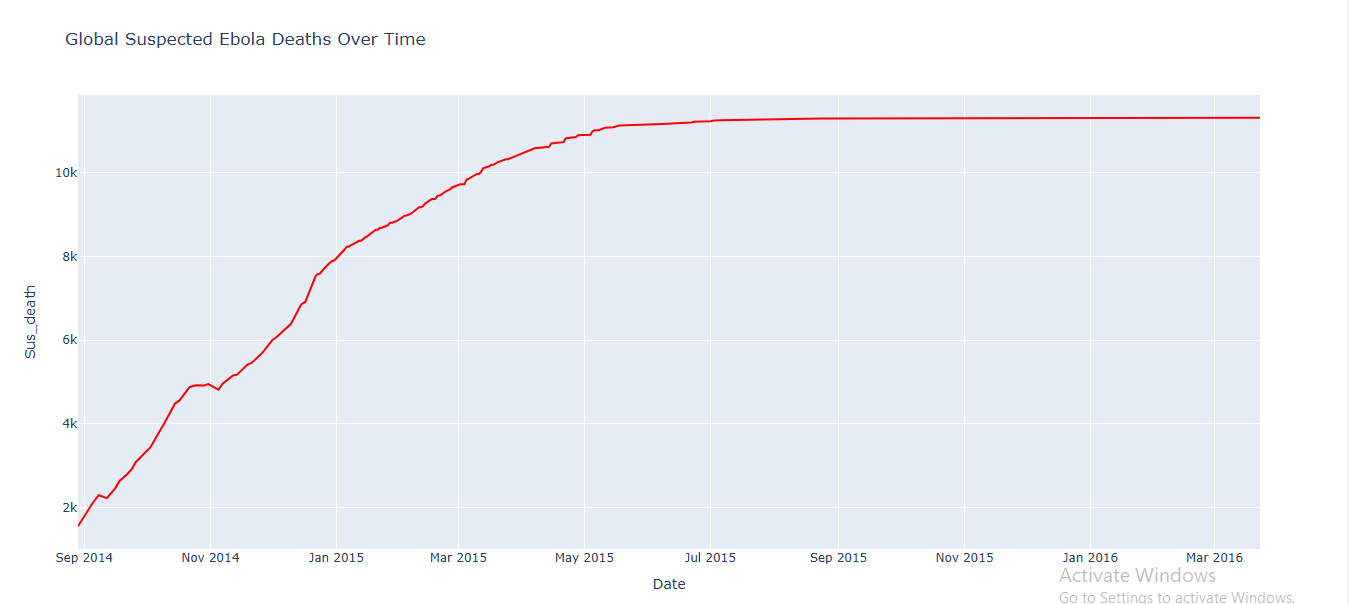
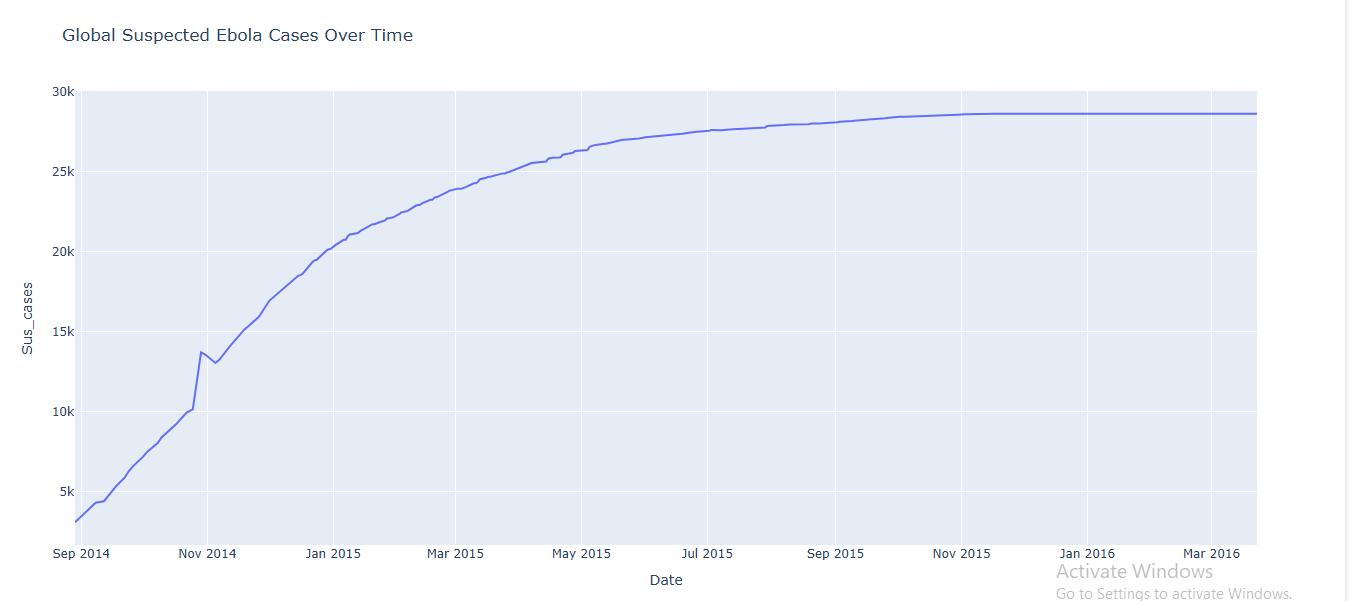
1. Line Plot

2.Line Plot



1. HeatMap



4.Graphs Using Plotly

6. Conclusion

This project on the Ebola outbreak (2013–2016) helps paint a clearer picture of how the epidemic unfolded and the toll it left behind.

1. The outbreak was **concentrated in just a few countries**, with **Sierra Leone, Liberia, and Guinea** bearing the heaviest burden. Together, they accounted for the majority of cases, while countries like Nigeria and Mali faced smaller outbreaks.
2. The **worst phase of the epidemic came in late 2014 (Q4)**, when suspected cases peaked sharply before the situation slowly came under control.
3. Ebola proved to be **extremely deadly**, with nearly **4 in every 10 patients dying**. This high case fatality rate exposed the fragility of healthcare systems in the affected regions.
4. The **spread was uneven across geography and time**. Heatmaps showed that while Sierra Leone and Liberia faced months of sustained transmission, others saw brief but intense spikes.
5. Finally, interactive visualizations revealed how **quickly the epidemic escalated** and how closely deaths tracked cases, reminding us of the devastating speed and severity of the disease.

7.APPENDICES

Dataset sourced from https://www.kaggle.com/datasets/imdevskp/ebola-outbreak-20142016-complete-dataset