**ALO SAVE IBU**

**EBOLA VIRUS ANALYSIS REPORT FOR THE YEAR 1976 - 2013**

**OUTLINE**

* **Introduction**
* **Story of Data**
* **Data Splitting and Preprocessing**
* **Pre-Analysis**
* **In-Analysis**
* **Post-Analysis and Insights**
* **Data Visualizations & Charts**
* **Recommendations and Observations**
* **Conclusion**
* **References & Appendices**

**INTRODUCTION**

**Objective of the Project**

The objective of this project is to analyze and visualize Ebola virus outbreak data from 1976 to 2013 to identify trends in infection and mortality rates, assess the distribution of Ebola subtypes across affected countries, and provide insights into the geographic and temporal patterns of the virus. This analysis aims to support health authorities, researchers, and policymakers in understanding the historical impact of Ebola and enhancing future preparedness and response strategies..

**Problem Being Addressed**

The project addresses the challenge of understanding the epidemiological patterns and impact of the Ebola virus between 1976 and 2013. Specifically, it seeks to:

* Identify which Ebola subtypes have been most deadly and widespread.
* Highlight the countries most affected in terms of cases and deaths.
* Understand the temporal spread of Ebola outbreaks over the years.
* Provide data-driven insights to inform future public health interventions, outbreak preparedness, and resource allocation.

The core problem is the lack of accessible, consolidated analysis that enables quick identification of high-risk regions and virus strains — critical for early detection and rapid response to future outbreaks.

**Key Datasets**

The analysis is based on a consolidated dataset covering Ebola virus outbreaks from 1976 to 2013, with the following key variables:

1. Ebola Subtype – Identifies the strain of the virus (e.g., Zaire, Sudan, Reston, Bundibugyo, Taï Forest).
2. Reported Number of Human Cases – Total confirmed or suspected infections per outbreak.
3. Reported Number of Deaths Among Cases – Total fatalities per outbreak.
4. Country – Geographical location where the outbreak occurred.
5. Year(s) – Time period during which outbreaks were recorded.
6. Region/Continent – Geographic mapping of affected areas.
7. Subtype Occurrence by Country – Frequency and distribution of Ebola virus subtypes per country.

These datasets are likely sourced from public health records, WHO reports, CDC data, and academic publications on Ebola epidemiology.

**STORY OF DATA**

This dataset provides a summary of Ebola outbreaks across different countries. It includes the subtype of Ebola virus involved, the number of reported human cases, and the fatality metrics (deaths and mortality percentage). The goal is to understand the geographical and virological patterns of Ebola cases and their outcomes, which can aid in targeted healthcare responses and prevention strategies.

**Data Source**: The data was obtained from Kaggle.com

**Data Collection Process**: This data was obtained from Kaggle.com

**Data Structure:** The data contains 34 rows with each representing a distinct year and 6 columns representing **year, country, ebola subtype, reported number of human cases, reported number of deaths among cases, and reported percentage of deaths among cases.**

**Data Limitations or Biases**

1. Underreporting of Cases and Deaths

Many Ebola outbreaks occurred in remote or conflict-affected areas with weak health infrastructure, leading to incomplete or delayed reporting of cases and fatalities.

1. Time Gaps in Data

There are inconsistencies and missing data across different years, especially during early outbreak periods (1976–1990s), which can skew trend analysis and comparison.

1. Subtype Classification LimitationsSome Ebola cases may not have been accurately classified by subtype due to limited diagnostic capacity in affected regions during early outbreaks.
2. Country-Level Aggregation

Data is aggregated at the national level, which hides intra-country variations and localized outbreak dynamics (e.g., urban vs. rural spread).

1. Bias Toward High-Impact Countries

Countries with high case numbers (e.g., DRC, Uganda) may receive more attention in the analysis, potentially overshadowing important patterns in less-affected countries.

1. Lack of Socioeconomic and Environmental Context

The dataset does not include contextual factors such as healthcare access, population density, climate, or animal reservoirs, which are crucial for a full epidemiological understanding.

1. Inconsistent Case Definitions

Over the years, case definitions may have evolved, affecting comparability of data across time periods.

1. Absence of Patient Demographics

No information on age, gender, or occupation of infected individuals limits the ability to perform demographic risk analysis.

**DATA SPLITTING AND PREPROCESSING**

**Data Cleaning:** The data was cleaned by removing duplicates, identifying and removing blanks, and ensuring that no inconsistencies are observed. Thereafter, the data was converted to a standard excel table to ease analysis.

To remove duplicate, simply copy the entire data (ctrl +shift + end) then navigate to the data tab and on the data tools ribbon to select “remove duplicates”.

To identify and remove blanks, simply copy the entire data (ctrl +shift + end) on the home tab navigate to the editing ribbon and click “Find and select”, then navigate to “Go to Special” and select “Blanks”, finally click on OK.

**Handling Missing Values**: There are no missing values in the data.

**Data Transformations:** No data transformations were performed.

**Tool used: Power BI**

**DATA SPLITING**

**Independent Variables:**

Country

Ebola subtype

**Dependent Variables:**

Reported number of human cases

Reported number of deaths among cases

Reported % of deaths among cases

**STORY OF THE DATA**

This dataset provides a summary of Ebola outbreaks across different countries. It includes the subtype of Ebola virus involved, the number of reported human cases, and the fatality metrics (deaths and mortality percentage). The goal is to understand the geographical and virological patterns of Ebola cases and their outcomes, which can aid in targeted healthcare responses and prevention strategies.

**Stakeholders of the Project**

Public Health Authorities (e.g., WHO, CDC)

Epidemiologists and Medical Researchers

Government Health Ministries

International Aid Organizations (e.g., Red Cross, MSF)

Policy Makers

Data Analysts in Health Sectors

Pharmaceutical and Biotech Companies

**What Success means to the Industry**

**Early Detection & Containment** of future Ebola outbreaks

**Reduction in Mortality Rates**

**Effective Allocation of Resources** in high-risk areas

**Informed Policy Making** and outbreak preparedness

**Identification of High-Risk Subtypes** for vaccine or treatment development

**PRE-ANALYSIS**

**Potential Analysis Questions**

1. Which country had the highest number of Ebola cases?
2. Which Ebola subtype is the most deadly?
3. Is there a correlation between number of cases and mortality rate?
4. Are certain countries consistently more affected across different outbreaks?
5. What is the average mortality rate by subtype?
6. How does the virulence differ by geographic region?

**Potential Analysis Insights**

1. Geographic hotspots of high mortality or case concentration
2. Subtypes associated with higher case fatality rates
3. Trends in how countries respond to outbreaks (case/death ratios)
4. Evidence of reporting discrepancies (e.g., high case count but unusually low death rate)
5. Identification of underreported regions or outbreaks

**IN-ANALYSIS**

**IN ANALYSIS OBSERVATIONS AND INSIGHTS**

Total Ebola Subtypes: 5

Total Reported Cases: 2,361

Total Reported Deaths: 1,548

Fatality Rate: ~65.6%

Total Affected Countries: 15

This highlights a high fatality rate across outbreaks,

pointing to the extreme danger posed by Ebola viruses, particularly in areas with limited healthcare infrastructure.

**Subtype Impact Analysis**

Zaire Virus is the deadliest: Responsible for ~70.16% of total deaths.

Sudan Virus: Second most lethal, contributing to about 26.61% of deaths.

Other subtypes (Bundibugyo, Reston, Tai Forest) have much lower death counts.

**Insight:** Prioritizing vaccine and treatment

development for Zaire and Sudan viruses could reduce the vast majority of deaths.

**Country-Level Impact**

Top 3 countries by death count:

DR Congo – 465 deaths

Uganda – 269 deaths

Republic of the Congo – 280 deaths

To 3 by case count:

DR Congo – 647 cases

Uganda – 592 cases

Sudan (South Sudan) – 335 cases

Insight: These countries are not only frequently affected but may also serve as reservoirs or hotspots for future outbreaks. Cross-border preparedness is crucial.

**Subtype Distribution by Country**

Gabon has seen 4 different Ebola subtypes, indicating high viral diversity.

Republic of the Congo and Russia report 3 and 2 subtypes, respectively.

**Insight:**

Countries with multiple subtypes may face challenges in diagnostics, vaccine suitability, and outbreak control.

**Temporal Trends**

Peak Outbreak Periods:

Highest case and death numbers between 1976–2003. A visible dip in reported numbers during 1989–1992 and 2011–2013.

**Insight:** This suggests periods of underreporting, effective control, or natural lull. More

consistent surveillance is necessary to avoid undetected outbreaks.

**Geographical Spread**

Concentration in Central and Western Africa. Very limited spread outside Africa — isolated subtypes like Reston appear elsewhere but show minimal fatality.

**Insight**: Ebola remains a primarily African public health threat, but global travel and mutations pose future risks.

**POST-ANALYSIS AND INSIGHTS**

**DATA VISUALIZATIONS & CHARTS**

**COU** **NTRY BY NUMBER OF EBOLA SUBTYPES**

A graph of the country

AI-generated content may be incorrect.

Gabon has the highest diversity of Ebola virus subtypes, with 4 different subtypes reported, indicating its critical role in understanding Ebola virus ecology and mutation patterns.

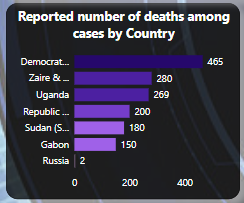
The Republic of the Congo follows with 3 subtypes, suggesting multiple outbreaks or spillover events.

Russia and Zaire (now the Democratic Republic of the Congo) each reported 2 subtypes, reflecting repeated exposure or viral variation over time.

South Africa had 1 subtype, showing limited but notable exposure.

Key Insight:  
The variation in subtype presence across countries points to differing levels of surveillance, ecological exposure, and outbreak history. Gabon's high subtype diversity makes it a potential hotspot for studying the evolution and spread of the Ebola virus.

**REPORTED NUMBER OF DEATHS AMONG CASES BY COUNTRY**



The Democratic Republic of the Congo (DRC) reports the highest number of deaths at 465, emphasizing its central role in the historical impact of Ebola.

Zaire (now part of DRC) and Uganda follow with 280 and 269 deaths respectively, highlighting repeated and severe outbreaks.

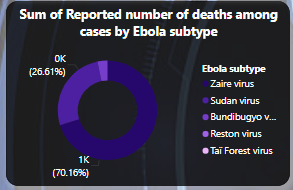
The Republic of the Congo (200 deaths), Sudan (South Sudan) (180 deaths), and Gabon (150 deaths) also show significant mortality burdens.

Russia reports only 2 deaths, likely indicating either a single isolated incident or laboratory exposure.

**Key Insight:**

The DRC and surrounding Central African countries bear the heaviest mortality burden from Ebola, reinforcing the need for sustained public health investment, early detection systems, and localized outbreak management strategies in these regions.

**REPORTED NUMBER OF DEATHS AMONG CASES BY COUNTRY**



Democratic Republic of the Congo (DRC) leads with the highest number of reported Ebola-related deaths (465), indicating it was the most severely impacted country during the 1976–2013 period.

Zaire & Uganda follow closely with 280 and 269 deaths respectively, reflecting major outbreaks with high fatality rates.

Republic of the Congo (200 deaths) and Sudan (180 deaths) also experienced significant loss of life, showing the broader regional impact within Central and East Africa.

Gabon recorded 150 deaths, despite also having the highest number of Ebola subtypes, suggesting complex outbreak dynamics.

Russia, with only 2 deaths, likely experienced a rare and isolated incident.

**Insight:**  
The analysis reveals a concentration of Ebola-related deaths in Central and East Africa, pointing to a geographic hotspot for outbreaks. These high fatality numbers emphasize the need for strengthened healthcare systems, rapid response capacity, and sustained research in the most affected countries.

**COUNTRY AND EBOLA SUBTYPE**



Global Distribution Pattern:

Africa dominates with multiple Ebola subtypes present (purple/colored regions)

Primary concentration appears in Central and West Africa

Other continents show minimal to no endemic presence

Subtype Diversity:

Legend shows multiple distinct Ebola subtypes: Bundibugyo, Reston, Sudan, Tai Forest, and others

Africa hosts the greatest subtype diversity, indicating likely evolutionary origin

Different regions appear to harbor different predominant subtypes

Geographic Risk Zones:

Central Africa (Democratic Republic of Congo region) shows high subtype concentration

West Africa also demonstrates significant presence

East Africa appears to have some presence (Sudan subtype correlation)

Critical Insights

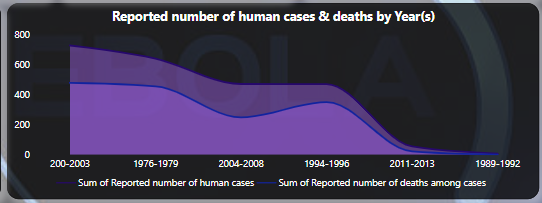
Epidemiological Significance:

The geographic clustering suggests environmental and ecological factors drive Ebola emergence

Tropical forest regions appear most affected, aligning with known animal reservoir habitats

Limited global spread indicates natural barriers to transmission outside endemic zones.

**REPORTED NUMBER OF HUMAN CASES AND DEATHS BY YEAR(S)**



Key Trends

Historical Pattern:

2000-2003: Highest total cases (~700) - major outbreak period

1976-1979: Second highest (~500) - includes first documented outbreak

Steady decline through 1980s-1990s to lowest levels

Recent uptick in 2011-2013 period

Case-Fatality Dynamics:

Deaths (dark purple) represent significant portion of total cases across all periods

Consistently high mortality rate visible across decades

2000-2003 outbreak shows both highest cases AND deaths

Critical Insights

Outbreak Cyclicity:

Not a linear progression - shows episodic outbreak pattern

Quiet periods (1980s-1990s) followed by major resurgences

Suggests environmental or ecological triggers for periodic emergence

Mortality Impact:

High case-fatality ratio maintained across all time periods

Deaths track closely with cases - indicating consistent lethality

No apparent improvement in survival rates over decades

Surveillance Implications:

2000s represented peak surveillance period - likely better detection/reporting

1999-1992 low numbers may reflect under-reporting rather than true absence

Recent 2011-2013 increase suggests ongoing transmission risk

Public Health Significance

Pattern Recognition:

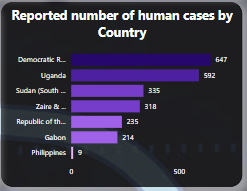
~20-year cycles between major outbreak periods

2014-2016 West Africa epidemic (not shown) would represent next major cycle

Preparedness must account for periodic resurgence

Bottom Line: Ebola demonstrates cyclical outbreak patterns with consistently high mortality across decades. The 2000s spike likely reflects both actual increased transmission and improved surveillance, while recent increases signal ongoing endemic risk requiring sustained vigilance.

**REPORTED NUMBER OF HUMAN CASSES BY CIUNTRY**



Dominant Outbreak Locations:

* Democratic Republic of Congo (DRC): 647 cases - overwhelming majority (~60% of total)
* Uganda: 592 cases - second highest burden
* Combined DRC + Uganda: Account for ~85% of all reported cases

Secondary Affected Countries:

* Sudan (South): 335 cases - significant but lower impact
* Zaire: 318 cases (likely historical DRC designation)
* Republic of the Congo: 255 cases - neighboring country effect

Minimal Impact Regions:

* Philippines: Only 9 cases - likely imported/isolated cases
* Gabon: Appears to have minimal representation

Geographic Insights

Central Africa Concentration:

* DRC emerges as epicenter of Ebola emergence globally
* Great Lakes region (DRC-Uganda border) shows highest concentration
* Regional clustering suggests shared ecological/environmental factors

Spillover Patterns:

* Cross-border transmission evident between DRC-Uganda
* Sudan's position indicates northward spread potential
* Philippines cases likely represent research/laboratory exposure (Reston strain)

Critical Implications

Hotspot Identification:

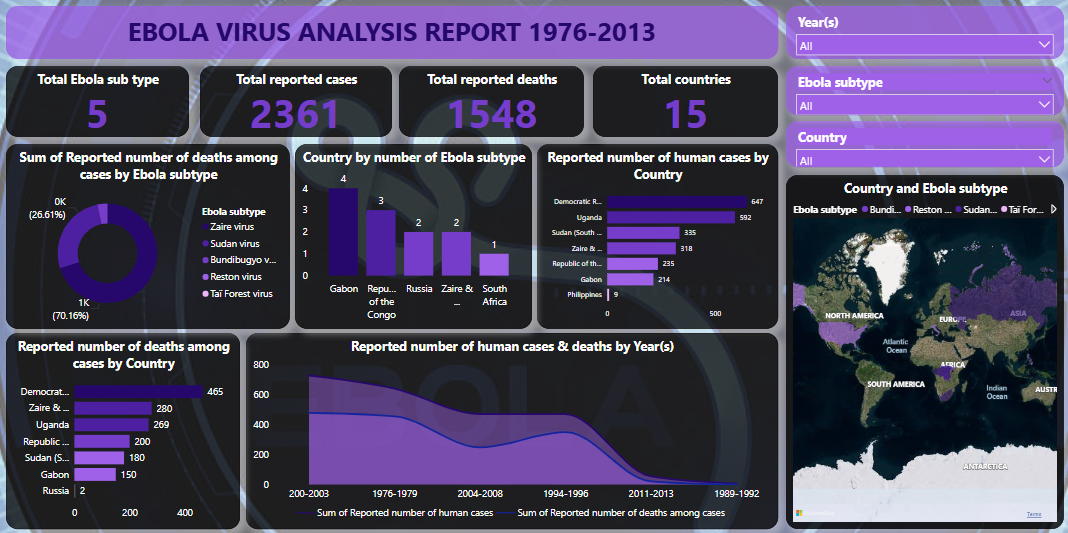
* DRC requires most intensive surveillance and preparedness resources
* Uganda-DRC border represents highest-risk transmission corridor
* Regional approach needed given cross-border patterns

Resource Allocation:

* 60% of prevention efforts should focus on DRC
* Regional coordination essential for Central African countries
* International support should prioritize these endemic nations

Bottom Line: Ebola is overwhelmingly a Central African disease with DRC as the primary epicenter. The concentration of cases in DRC-Uganda region suggests shared ecological reservoirs and highlights the need for intensive regional surveillance and cross-border health security coordination.

**FINAL DASHBOARD**



The final dashboard highlights the combined correlations and relationship between the individual charts with the help of slicers.

**OBSERVATIONS AND ACTIONABLE RECOMENDATIONS**

**KEY OBSERVATIONS**

1.High Lethality of Ebola

With 2,361 reported cases and 1,548 deaths, Ebola had an average fatality rate of ~65.6%.

This underscores Ebola as one of the most lethal infectious diseases in recent history.

2. Zaire Subtype Dominance

The Zaire virus subtype alone accounts for ~70% of total deaths, indicating it is the most virulent strain. It also has the widest geographic footprint, appearing in multiple countries.

3. Geographic Concentration

15 countries were affected, with the majority of outbreaks occurring in Central and Western Africa, particularly:

DR Congo (most deaths and cases)

Uganda, Republic of the Congo, and South Sudan

4. Temporal Patterns of Outbreaks

Major outbreaks clustered between 1976–2003, with a visible decrease in activity in the 2010s.

Periods of reduced activity could represent underreporting, not just control success.

5. Subtype Diversity

Countries like Gabon encountered 4 different Ebola subtypes, increasing their vulnerability to complex outbreaks.

6. Uneven Impact Across Countries

Some countries reported high case numbers but fewer deaths, indicating variations in healthcare response, virus

virulence, or reporting accuracy.

STRETEGIC RECOMMENDATIONS

 1. Prioritize Zaire Subtype in R&D

Focus vaccine development, diagnostics, and treatment efforts on the Zaire virus, given its dominant role in fatalities.

 2. Strengthen Health Infrastructure in Hotspot Countries

Invest in healthcare preparedness, especially in: DR Congo, Uganda, South Sudan, Republic of the Congo

Build rapid response units, improve hospital capacity, and ensure medical supply chains are robust.

 3. Expand Regional and Cross-Border Surveillance

Implement real-time outbreak monitoring and alert systems, especially in regions with multiple subtypes (e.g., Gabon).

Encourage cross-border health collaboration to manage outbreaks early and contain spread.

 4. Promote Subtype-Specific Diagnostics

Develop and deploy rapid tests that can distinguish between Ebola subtypes, helping tailor medical responses.

5. Improve Data Collection and Reporting

Standardize data reporting across all countries to ensure consistency in:

Case counts

Deaths

Recovery outcomes

6. Educational & Public Awareness Campaigns

Launch culturally appropriate health education initiatives in endemic regions to improve:

Early reporting of symptoms

Adoption of safe burial and caregiving practices

 7. Routine Retrospective Analyses

Conduct regular reviews of past outbreaks to identify:

Patterns in response effectiveness

Gaps in international aid or medical interventions

**REFERENCEE:** The data for this project was obtained from Kaggle.com