Data Science for Health Systems Final Project Report
Exploratory analysis in R on a World Health Organization Organization dataset
Maternal Mortality

Dimopoulou Sevasti
Undergraduate at EE-CS
University of Perugia
Perugia, Italy
sevasti.dimopoulou@studenti.unipg.it

I. Abstract

This report analyzes maternal mortality data sourced from the World Health Organization (WHO) to understand the distribution and trends of maternal deaths globally. The analysis focuses on examining the distribution of observed maternal deaths, proportional mortality, and environmental factors contributing to maternal mortality. Additionally, temporal trends are assessed to identify improvements and areas needing intervention. The findings highlight significant disparities between countries and show general improvements in maternal health over recent decades.

II. Introduction

Maternal mortality remains a critical public health issue worldwide, reflecting the quality of healthcare systems and socio-economic conditions. Understanding the distribution and trends of maternal mortality can help in identifying regions needing urgent interventions and in formulating policies to improve maternal health outcomes. This report aims to provide a comprehensive analysis of maternal mortality data, focusing on both the distribution of deaths and the temporal trends over the years.

III. Dataset and Features

The dataset used in this analysis, `main_data`, contains information on maternal deaths from various countries over multiple years. It is from the World Health Organization, and it can be accessed at this link:
Key variables include:

- `iso_alpha_3_code`: Country code
- `year_start`: Start year of the data period
- `year_end`: End year of the data period
- `year_mid`: Midpoint year of the data period
- `obs_matdeaths`: Observed maternal deaths
- `final_pm`: Final proportional mortality
- `env_total`: Environmental factors total
- `env_mat`: Environmental factors related to maternal health

The dataset was loaded and explored to understand its structure and key statistics. Relevant columns were selected, and rows with missing values in key columns were filtered out to ensure data integrity.

IV. Exploratory Data Analysis Methodology

The Exploratory Data Analysis (EDA) conducted for this study involved a series of systematic steps aimed at cleaning and visualizing the data to extract meaningful insights. The data cleaning process began with the removal of rows containing missing values to ensure the analysis was based on complete data. Next, appropriate data types were assigned to each column to maintain consistency and accuracy in the analysis. This included converting categorical data to factors and numerical data to integers or numeric types as needed. Duplicate entries were identified and removed to prevent any bias or errors in the analysis.

Handling outliers was a critical step in the data cleaning process. Outliers in key variables, such as `obs_matdeaths` (observed maternal deaths) and `final_pm` (final proportional mortality), were managed by capping them at the 99th percentile. This approach ensured that extreme values did not skew the results while retaining the overall data structure.

For the data visualization, several types of plots were generated to explore the distributions and trends within the dataset. Histograms were used to display the distribution of key variables, providing a clear picture of their spread and central tendency. Line plots were employed to illustrate temporal trends, showing how maternal mortality and related factors have changed over time. Boxplots were created to compare maternal deaths across different countries, highlighting variations and identifying potential outliers.

This comprehensive approach to EDA, combining data cleaning and various visualization techniques, provided a

robust foundation for understanding the patterns and trends in maternal mortality data. Through these methods, key insights into the disparities and improvements in maternal health were uncovered, guiding further analysis and policy recommendations.

V. Detailed Plot Analysis and Interpretations

Distribution of Observed Maternal Deaths

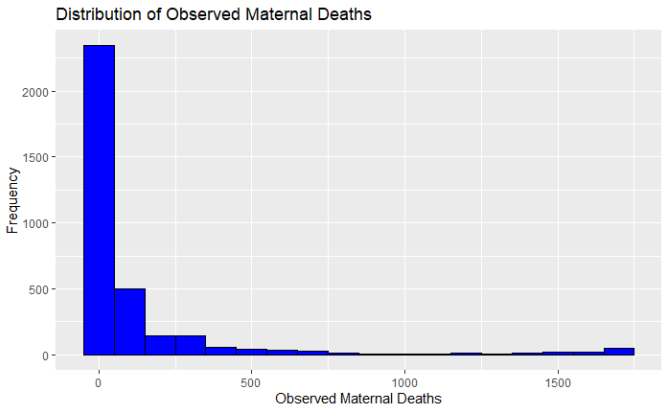

Fig. 1. plot 1

The histogram (plot 1) shows that most countries have relatively low numbers of observed maternal deaths, with a right-skewed distribution indicating a few countries with significantly higher maternal deaths. This highlights the disparity in maternal health outcomes across different regions. The right-skewness suggests that while maternal mortality is generally low, there are outliers with very high mortality rates that need further investigation and targeted interventions.

Distribution of Proportional Mortality

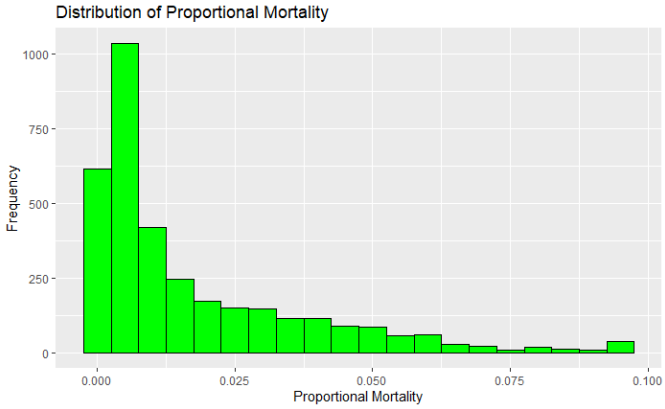

Fig. 2. plot 2

This histogram (plot 2) illustrates that most countries have low proportional mortality rates, with a decreasing frequency as the proportional mortality increases. The right-skewed distribution suggests that higher proportional mortality rates are less common but significant in certain regions. This indicates that in many countries, maternal deaths constitute a small proportion of total deaths, but in a few, they are a substantial percentage, highlighting areas for potential health policy interventions.

Distribution of Environmental Total

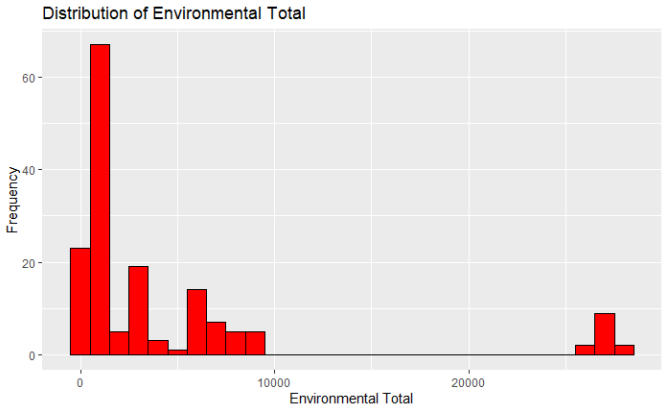

Fig. 3. plot 3

The right-skewed distribution (plot 3) indicates that while most countries have low environmental contributions to maternal mortality, a few countries experience much higher values. This suggests that environmental factors are significant contributors to maternal mortality in certain regions. Addressing these environmental issues could lead to significant improvements in maternal health outcomes in the affected areas.

Distribution of Environmental Maternal

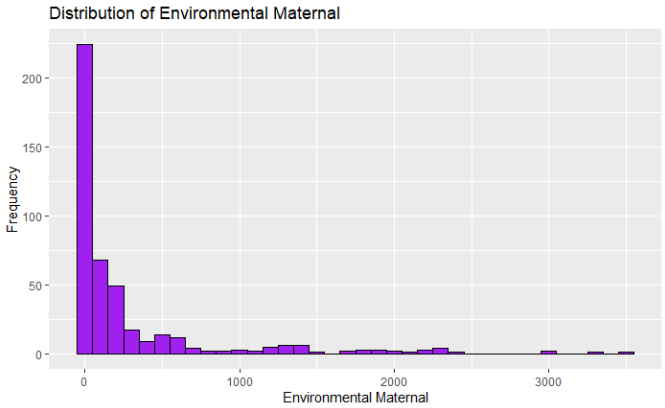

Fig. 4. plot 4

Similar to other distributions, this histogram (plot 4) is right-skewed, with most data points at the lower end. A few countries show extreme values, highlighting significant environmental issues affecting maternal health. The presence of these outliers suggests that specific environmental factors significantly impact maternal mortality in these regions, warranting targeted interventions.

Total Maternal Deaths Over Time

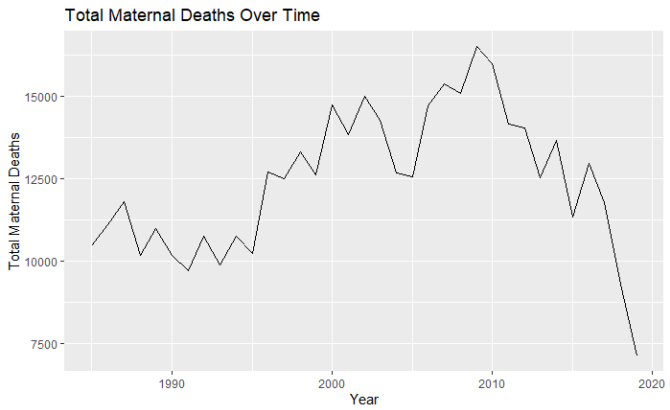

Fig. 5. plot 5

This line plot (5) shows fluctuating trends in total maternal deaths over the years, with a noticeable decline starting around 2010. This suggests possible successful interventions and improvements in maternal healthcare during the later years. The overall decline indicates progress in reducing maternal mortality, but the fluctuations highlight the need for continuous monitoring and intervention.

Average Proportional Mortality Over Time

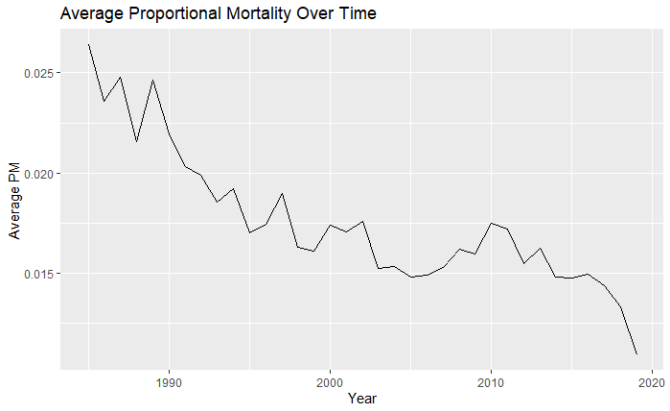

Fig. 6. plot 6

The line plot (6) indicates a steady decline in average proportional mortality from 1990 to 2020. This consistent decline reflects improvements in maternal health and better management of factors contributing to maternal mortality. It suggests that efforts to reduce maternal deaths relative to total deaths have been effective over time.

Boxplot of Maternal Deaths by Top 10 Countries

The boxplot (plot 7) shows significant variation in maternal deaths among the top 10 countries, with Brazil having the highest median and upper range of maternal deaths. The presence of outliers suggests specific years or incidents with unusually high maternal deaths. This variation highlights the need for country-specific strategies to address maternal mortality.

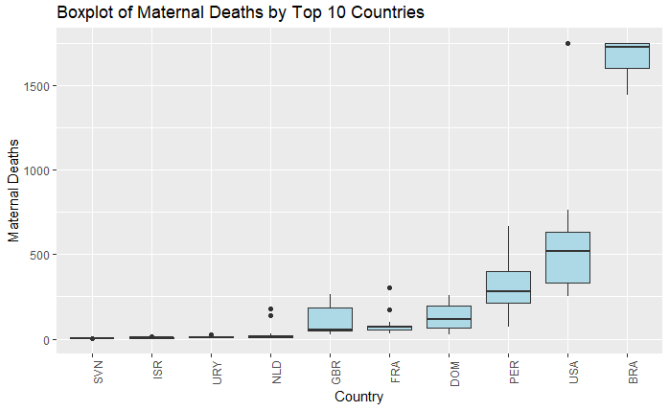

Fig. 7. plot 7

Total Maternal Deaths Over Time (South America)

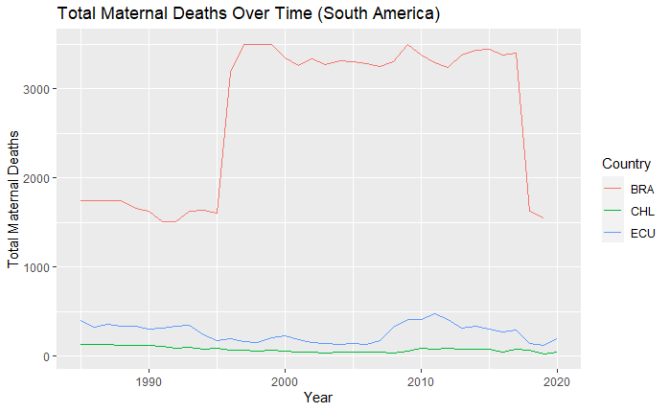

Fig. 8. plot 8

The plot for South America (plot 8) reveals that Brazil experienced a significant increase in maternal deaths starting around 1995, with a subsequent decline around 2015. Chile and Ecuador show relatively stable trends with lower maternal deaths compared to Brazil. This suggests that Brazil's maternal health policies and interventions during these periods had a substantial impact.

Total Maternal Deaths Over Time (Multiple Countries)

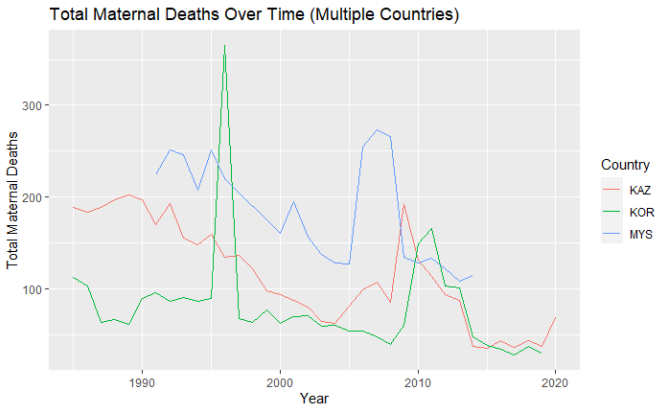

Fig. 9. plot 9

Kazakhstan shows a sharp peak around 2000, while South Korea and Malaysia display declining trends over time. This plot (9) highlights significant regional differences in maternal mortality trends. Understanding the causes of these peaks and declines can help inform targeted interventions.

Total Maternal Deaths Over Time (Multiple Countries)

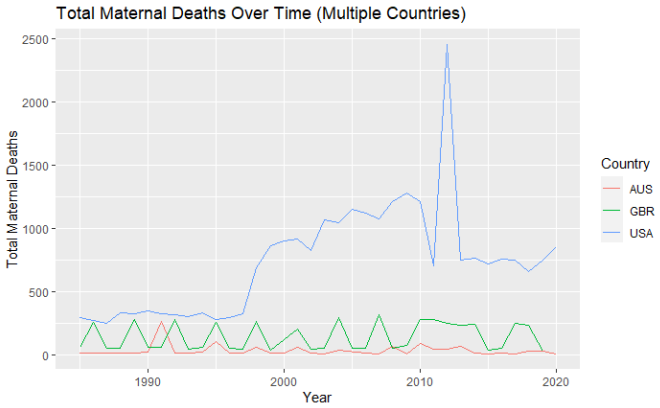

Fig. 10. plot 10

In plot 10 the United States shows a noticeable peak around 2010, with Australia and the UK showing relatively stable and low maternal deaths. This indicates different regional challenges and improvements in maternal health. The peak in the USA around 2010 suggests a period of increased maternal mortality that needs further investigation.

Trend in Proportional Mortality Over Time

The scatter plot (11) with a trend line shows a clear decline in proportional mortality over time, indicating overall improvement in maternal health relative to total mortality rates. Despite the scatter, the downward trend is evident, suggesting sustained efforts to reduce maternal mortality have been effective.

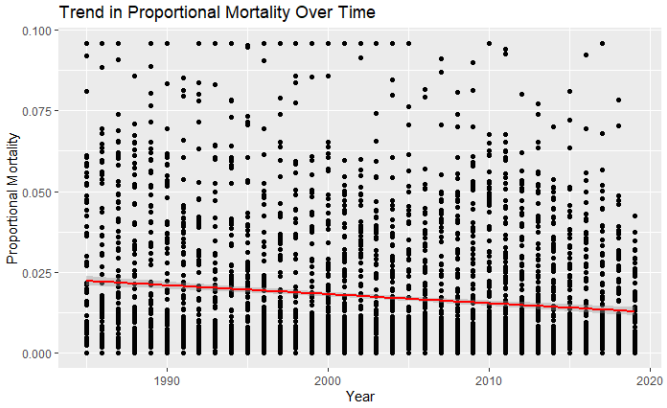

Fig. 11. plot 11

VI. Conclusion

The analysis reveals several key insights into maternal mortality. Significant disparities exist in maternal mortality rates across different countries, with some countries experiencing much higher maternal deaths. There is a general improvement in maternal health over the years, with declining trends in both total maternal deaths and proportional mortality. Certain regions, particularly Brazil and the USA, show periods of high maternal mortality that need targeted interventions. Environmental factors play a significant role in maternal mortality, especially in certain countries, highlighting the need for addressing these issues.

These findings underscore the importance of continued efforts to improve maternal health outcomes, with a focus on regions and factors where maternal mortality remains high.

Bibliography

[1] Fienberg S. E. (2014). *What Is Statistics?*
[2] Wickham H., Çetinkaya-Rundel M., & Grolemund G. (2019). *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data* (2nd ed.).