DATA 611
PREDICTIVE ANALYTICS

PROJECT 1
REPORT

GLOBAL MORTALITY ANALYSIS

GROUP – 2

BY
SHAHZAIB RAHAT
JUAN BYJU
HARLEEN KAUR

INTRODUCTION

Understanding the leading causes of death worldwide provides critical insights into public health priorities, resource allocation, and intervention strategies. Our project examines mortality trends from 2015 to 2019, focusing on key causes such as cardiovascular diseases, nutritional deficiencies, substance abuse, and conflict-related deaths. These causes were chosen for their global relevance and significant impact on various populations.

The purpose of our visualizations is to identify patterns, correlations, and trends in mortality data, enabling a deeper understanding of health challenges and disparities across regions and countries. For instance, exploring the link between alcohol use disorders and self-harm highlights pressing mental health concerns, while analyzing deaths from malnutrition emphasizes the urgent need for addressing food insecurity.

The intended audience for these analyses includes policymakers, healthcare professionals, researchers, and organizations working toward global health improvement. By providing clear, data-driven insights, our visualizations aim to inform strategies for reducing preventable deaths and promoting global health equity.

DATA

The data for this project was sourced from the Kaggle dataset titled *"Cause of Deaths around the World (Historical Data)"* (2024, February 12), available at: https://www.kaggle.com/datasets/iamsouravbanerjee/cause-of-deaths-around-the-world,

This dataset contains historical mortality data from 1990 to 2019, detailing the number of deaths caused by various diseases and external factors across different countries.

During data preparation, we encountered some challenges, notably countries reporting a value of 0 for deaths due to certain diseases like malaria and exposure to forces of nature. Since these entries likely indicated missing or inaccurate data, we excluded them from our study to ensure the results were meaningful.

Once the data was filtered and cleaned, it was structured to provide a clear breakdown of the number of deaths across various health categories and years. The analysis focused on examining trends in **mortality rates** and **disease burden** across different countries during the selected time. We filtered the dataset to include only data from 2015 to 2019 for this selection.

Limitations of the Dataset

- The dataset has limitations, including potential underreporting and inconsistent data collection practices in certain regions. For instance, diseases like malaria, malnutrition, and self-harm are reported as having zero cases in several countries, which may reflect gaps in data rather than their actual absence. Such discrepancies can impact the accuracy and completeness of the dataset.

- Countries with the most missing data include Tokelau, Niue, San Marino, Cook Islands, and Andorra, often for diseases like malaria, Alzheimer's, tuberculosis, and malnutrition, reflecting gaps in data collection for smaller nations.

- **Zero death entries** were excluded from the study, which might have impacted the analysis for certain regions with lower mortality rates from specific causes

RESULTS

Graph A:

From the given dataset, we first analyzed the leading causes of death globally. Cardiovascular diseases and neoplasms are the top two causes of death. The numbers for these two causes are also growing faster every year compared to the other three causes ('Chronic Respiratory Diseases,' 'Lower Respiratory Infections,' and 'Digestive Diseases') in our dataset.

Graph B:

India and China have a very high number of deaths due to self-harm, based on data from 2015 to 2019. India stands at almost 0.9 million deaths, and China at 0.6 million. The other countries in the top five list are the USA, Russia, and Japan, with 231 thousand, 205 thousand, and 126 thousand deaths, respectively, during the same period.

Graph C:

When we compared the number of deaths due to self-harm with the country's population to analyze how many people died per hundred thousand, Russia and Japan led the list. The other top five countries were the United States, India, and China, as shown in the horizontal bar graph.

Graph D:

It was discovered during the analysis that the number of deaths due to self-harm has a strong correlation with the number of deaths from alcohol use. A correlation coefficient of 0.81 was calculated between these two variables. While we are not claiming any causal relationship, the correlation graph clearly suggests that a higher number of deaths due to drinking is associated with a higher number of deaths due to self-harm.

Graph E:

Nutritional deficiencies are another significant leading cause of death. The overall numbers suggest that India and Indonesia are the top two countries with the highest number of deaths from 2015 to 2019 due to nutritional deficiencies and malnutrition.

Graph F:

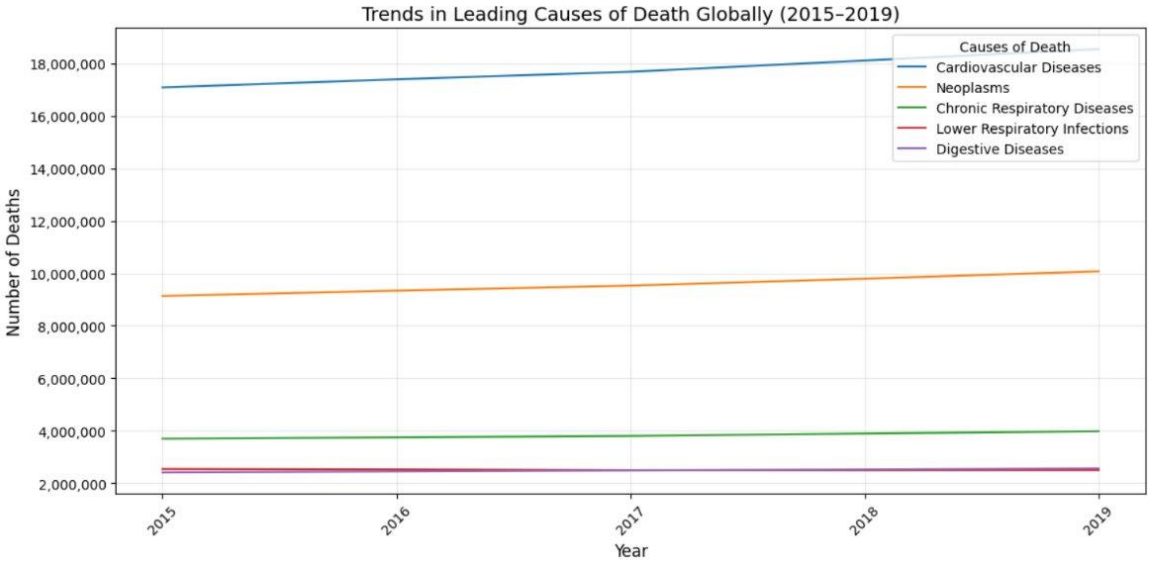
If we consider the number of deaths per one hundred thousand people due to nutritional deficiencies and malnutrition in each country, Mali is staggeringly high, with more than 750 deaths per one hundred thousand people. The remaining countries in the top five list are Indonesia, Pakistan, India, and China, with deaths lower than 100 per hundred thousand.

Graph I:

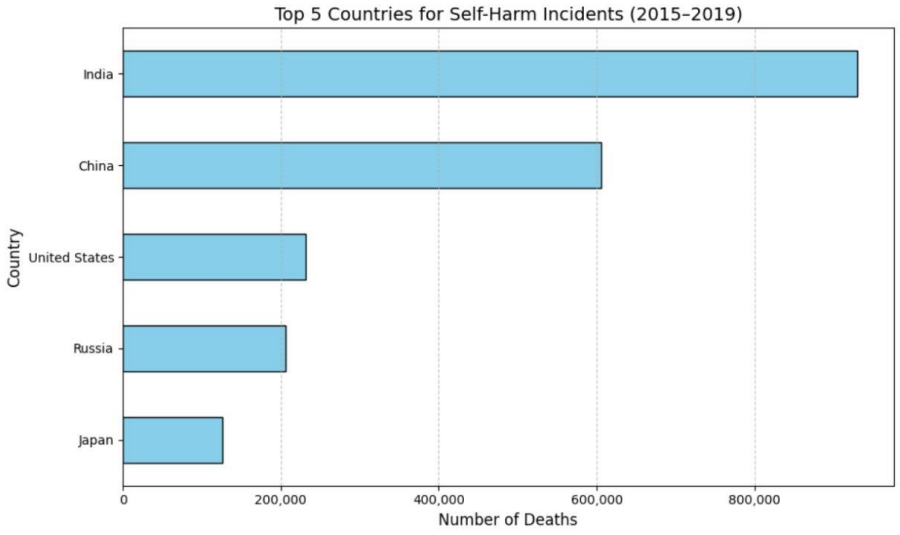
The heat map indicates that Syria witnessed the highest number of deaths in 2015 due to conflict and terrorism, far exceeding any other country. These numbers for Syria declined significantly, from 61 thousand in 2015 to just 6.9 thousand in 2019. The other top five countries that faced the highest number of deaths due to conflict and terrorism are Iraq, Afghanistan, Nigeria, and Yemen. For Iraq and Nigeria, the total number of deaths each year declined from 2015 to 2019.

APPENDIX

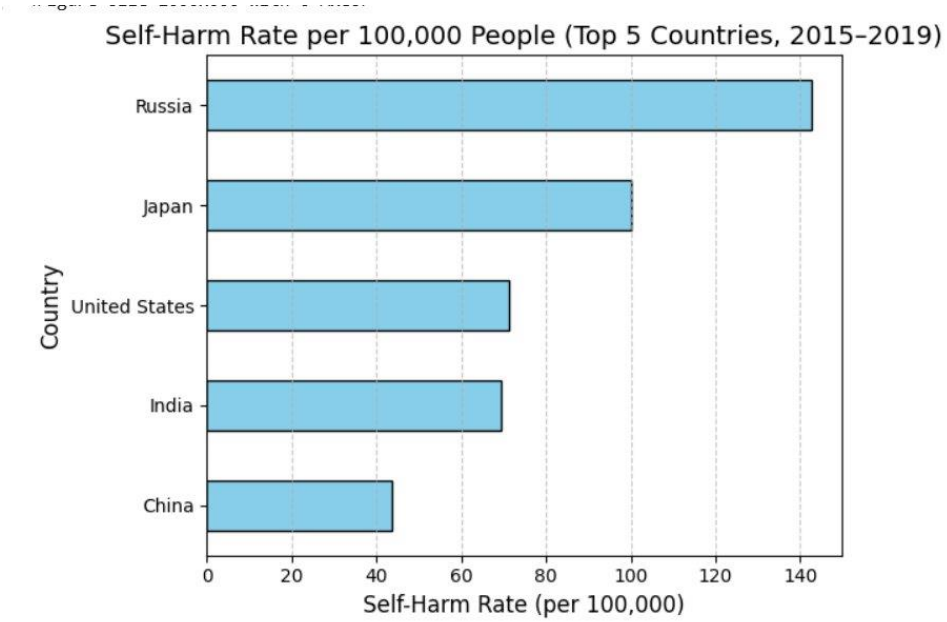
A) Primary Cause of Death Worldwide



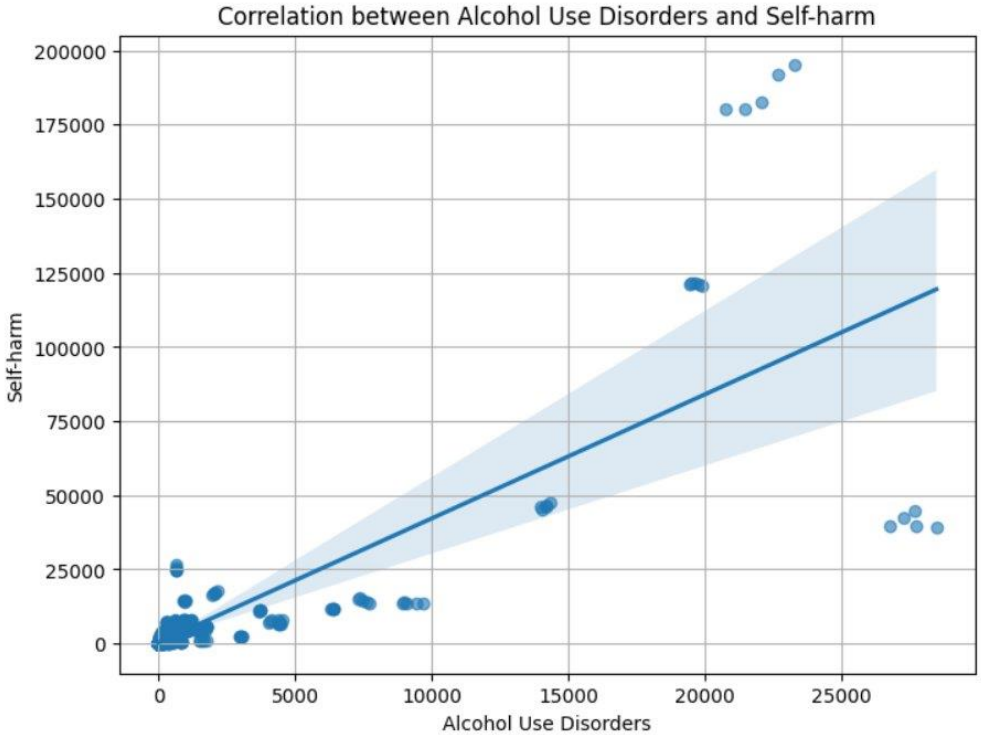
B) Deaths Related to Self-Harm



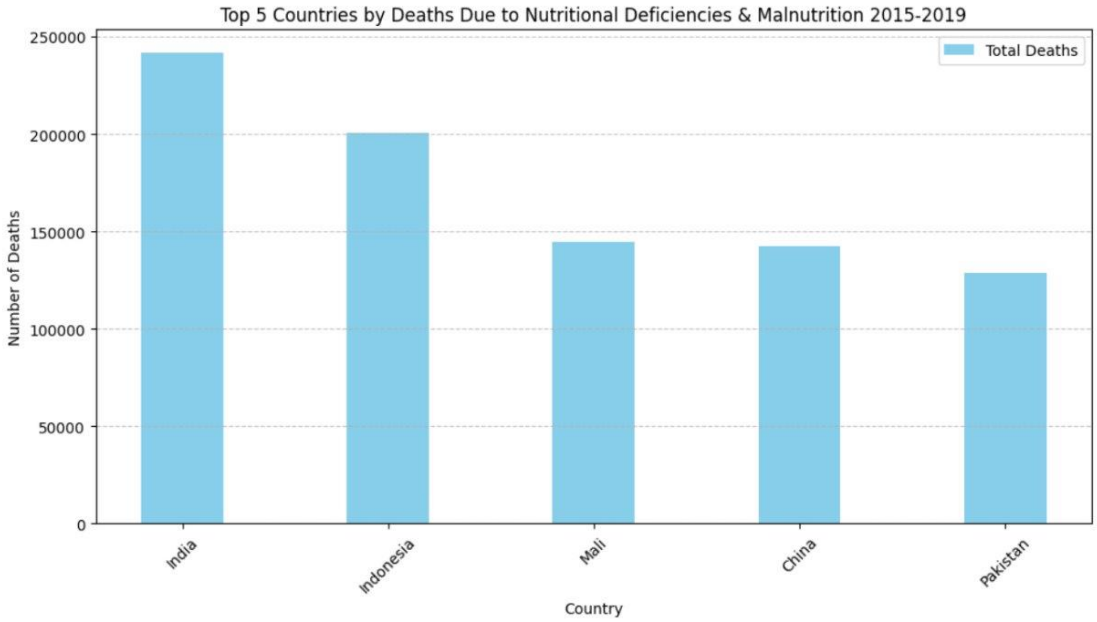
C) Deaths Related to Self-Harm per 100k People



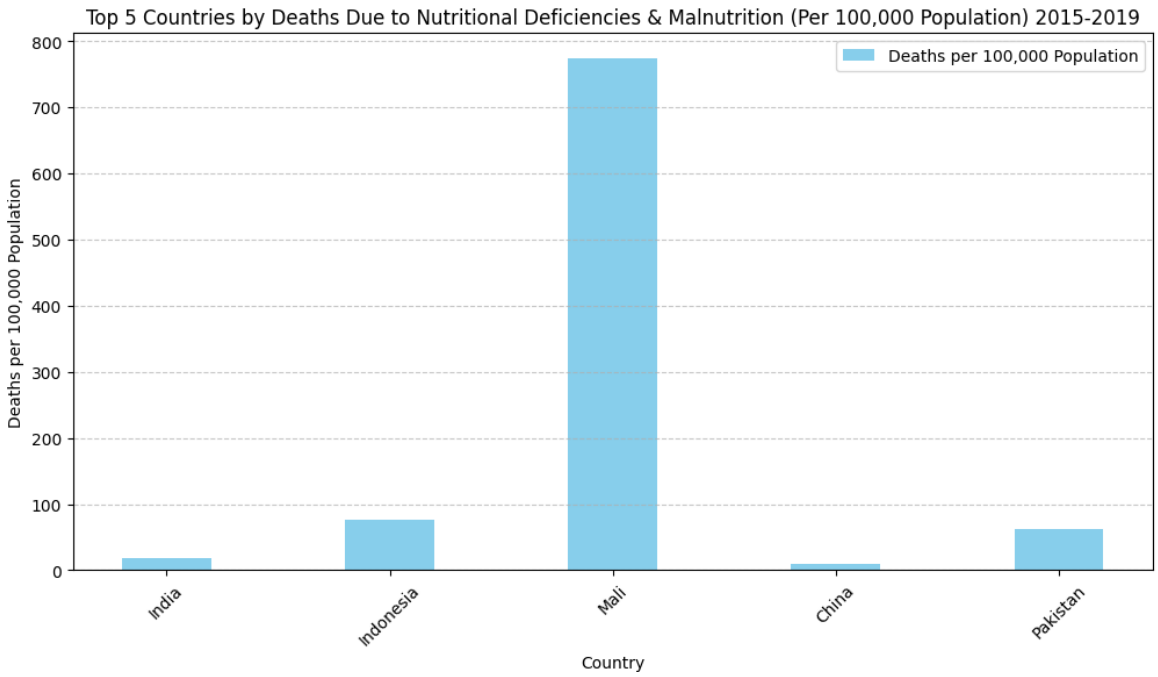
D) Correlation Between Alcohol Use Disorders and Self-harm



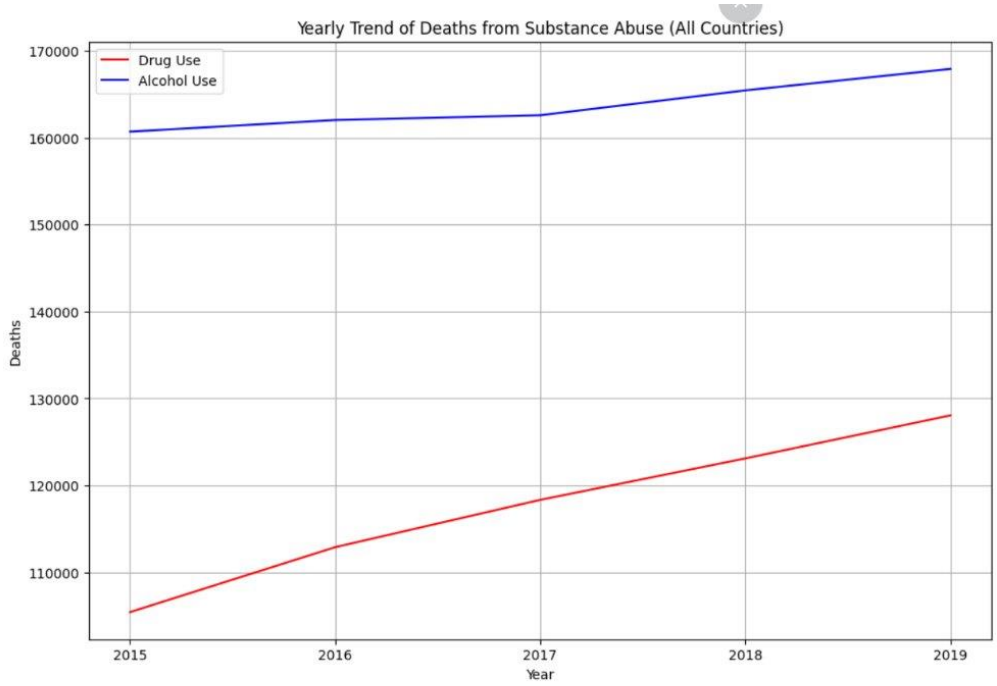
E) Deaths Due to Nutritional Deficiencies & Malnutrition



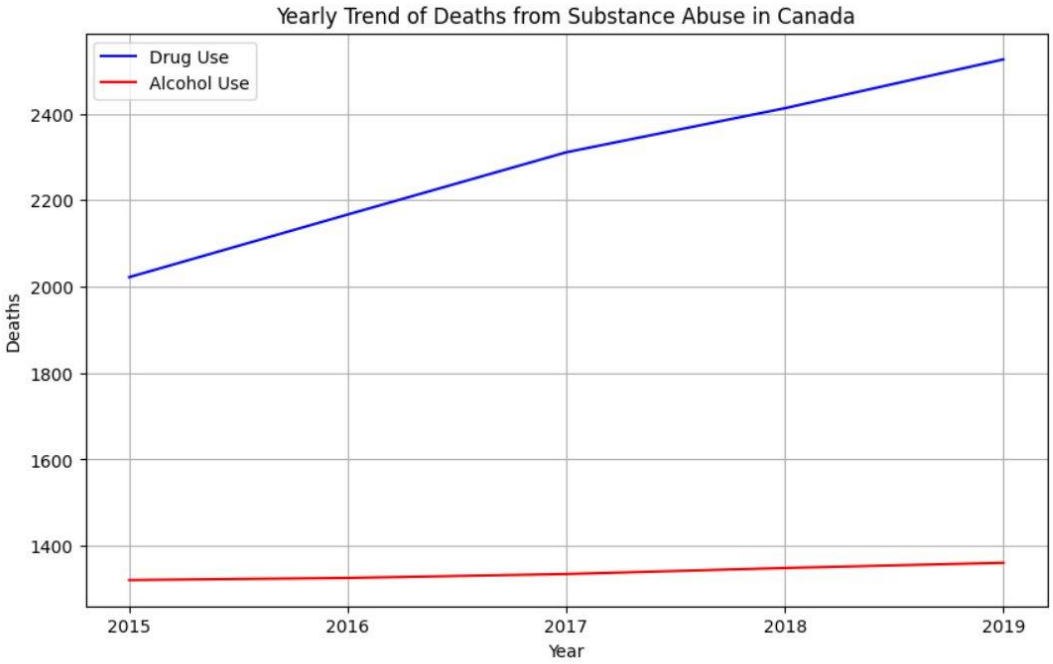
F) Deaths Due to Nutritional Deficiencies & Malnutrition per 100k People



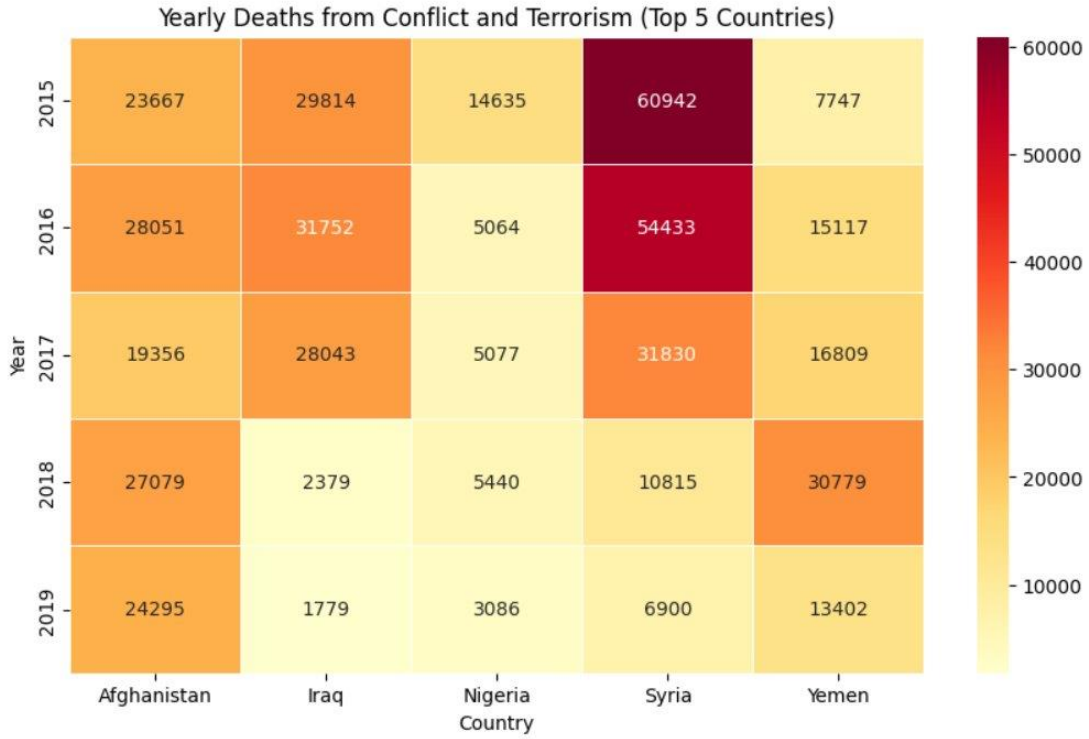
G) Yearly Trend of Deaths from Substance Abuse



H) Yearly Trend of Deaths from Substance Abuse in Canada



I) Yearly Deaths from Conflict and Terrorism (Top 5 Countries)



REFERENCES

1. Banerjee, S. (2024, February 12). Cause of deaths around the world (historical data). Kaggle. https://www.kaggle.com/datasets/iamsouravbanerjee/cause-of-deaths-around-the-world

2. World Bank. (n.d.). World Bank Open Data. Retrieved January 24, 2025, from https://data.worldbank.org