

**SINH VIÊN: ĐỖ HOÀNG VŨ**

**MSSV: 31012302902**

**LỚP: 23DAI**

**GIẢNG VIÊN: ­­HUỲNH ĐỆ THỦ**

**ĐỀ TÀI:**

**PHÂN TÍCH TRỰC QUAN HÓA DỮ LIỆU VỀ SỐ CA TỬ VONG THEO TỪNG NGUYÊN NHÂN TRÊN TOÀN CẦU TỪ 1990 ĐẾN 2019**

**KHOA: KỸ THUẬT& KHOA HỌC MÁY TÍNH**

**TIỂU LUẬN**

**MÔN LẬP TRÌNH PYTHON**

**Acknowledgments­­­**

I would like to express my profound gratitude to Dr. Huynh De Thu, an esteemed lecturer specializing in Data Analysis and Visualization with Python. Dr. Thu generously shared his time and profound expertise, providing invaluable guidance throughout the completion of this paper. His mentorship not only served as a strong source of motivation but also inspired me to strive for excellence and achieve the best possible outcomes.

I also extend my heartfelt thanks to the team at Our World in Data, particularly Dr. Saloni Dattani, Dr. Fiona Spooner, Professor Max Roser, and Dr. Hannah Ritchie, eminent researchers in the fields of global health and data analysis. Their contributions not only provided invaluable data but also imparted extensive knowledge and insights, deepening my understanding of community health issues and global development. Their research endeavors have significantly facilitated and laid a solid foundation for this paper.

**Table of contents**

[I. Introduction 2](#_Toc164116461)

[I.1 Purpose of the paper 2](#_Toc164116462)

[I.2 Introduction to the dataset 2](#_Toc164116463)

[I.3 Introducing some of the libraries and configurations for Python used in this paper 4](#_Toc164116464)

[II. Data Exploration 5](#_Toc164116465)

[II.1. Dataset description 5](#_Toc164116466)

[II.2. Dataset description 6](#_Toc164116467)

[III. Data preprocessing 9](#_Toc164116468)

[III.1. Read and process redundant data 9](#_Toc164116469)

[III.2. Data classification 9](#_Toc164116470)

[IV. Data Analysis and Visualization 12](#_Toc164116471)

[IV.1. Overall trend analysis and data visualization 12](#_Toc164116472)

[IV.2. Find relationships between causes of death and visualize with heatmap 18](#_Toc164116473)

[IV.3. Comparison between countries and data visualization 22](#_Toc164116474)

[IV.4. Globally, non-communicable diseases are the most common causes of death 30](#_Toc164116475)

[IV.5. Self-harm! an alarming problem in modern society 34](#_Toc164116476)

[IV.6. Analyze data on causes of death in Vietnam in particular 37](#_Toc164116477)

[V. Conclusion 44](#_Toc164116478)

[V.1. Summary of analysis results: 44](#_Toc164116479)

[V.2. Proposal for further research directions: 44](#_Toc164116480)

[VI. References 46](#_Toc164116481)

# Introduction

## I.1 Purpose of the paper

The purpose of this paper is to analyze and visualize data on the causes of death across different countries and years, using Python as the primary tool. This analysis aims to answer the critical question: “What are people dying from?”

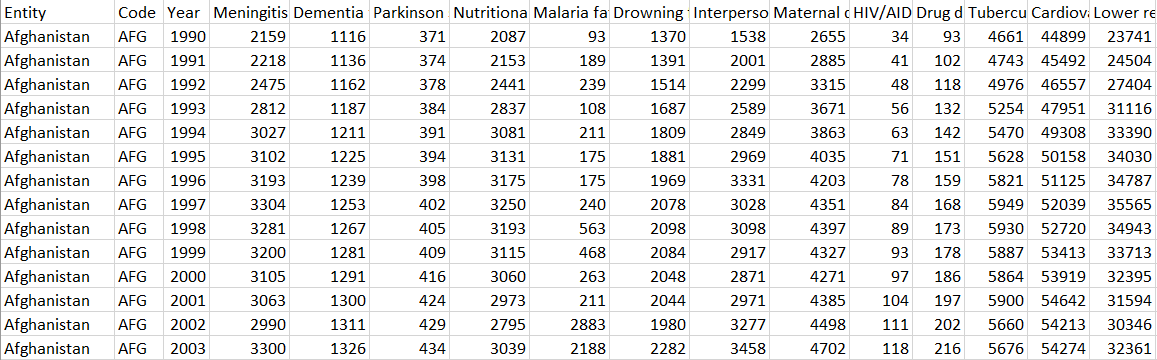
Understanding the causes of death is essential for guiding decisions in public health and finding ways to save lives. Many leading causes of death receive little mainstream attention, and this analysis aims to bring these causes to light.

This paper will explore how death rates from various causes have changed over time and how these changes have led to shifts in the leading causes of death. For instance, while infectious diseases once dominated, death rates from these diseases have fallen quickly, leading to non-communicable diseases such as heart diseases and cancers becoming the most common causes of death globally.

Furthermore, this paper will delve into how this data can help understand the burden of disease more broadly and offer a lens to see the impacts of healthcare and medicine, habits and behaviors, environmental factors, health infrastructure, and more.

By providing a comprehensive analysis and visualization of this data, this paper aims to contribute to the ongoing global conversation about public health and disease prevention. Ultimately, the goal is to use this analysis to inform strategies that can further reduce the impact of causes of death and improve health outcomes worldwide.

## I.2 Introduction to the dataset

The dataset utilized in this paper is a CSV file compiled from 210 countries and 18 international organizations spanning the years 1990 to 2019.

*Table 1.1: A piece of data from the dataset*

Dr. Saloni Dattani, a distinguished researcher, authored this dataset in collaboration with renowned experts in the field of global health. The dataset can be accessed and downloaded from [Our World In Data](https://ourworldindata.org/causes-of-death), a prominent platform dedicated to providing comprehensive data-driven insights into global issues.

Dr. Saloni Dattani, a researcher specializing in health-related topics, particularly mental health, joined the project team in 2021. Her background includes a Ph.D. in psychiatric genetics from the University of Hong Kong and King’s College London.

Additionally, the project team includes esteemed professionals contributing to the field of data science and global development:

Dr. Fiona Spooner: A Senior Data Scientist who joined the team in 2021. Dr. Spooner's expertise lies in tracking Sustainable Development Goals and modeling the COVID-19 pandemic. She holds a Ph.D. in Ecology and Environment from UCL (London, UK) and an MSc in Conservation Science from Imperial College (London, UK).Professor Max Roser: The Founder and Executive Co-Director of Our World in Data. Professor Roser initiated this influential publication in 2011 and continues to lead its development. He holds the position of Professor of Practice in Global Data Analytics at the University of Oxford’s Blavatnik School of Government. Additionally, he serves as the Programme Director of the Oxford Martin Programme on Global Development and Executive Co-Director of Global Change Data Lab.

Dr. Hannah Ritchie: Serving as the Deputy Editor and Science Outreach Lead at Our World In Data since 2023, Dr. Ritchie is an integral member of the team. Her focus areas include long-term development in food supply, agriculture, energy, and the environment. Dr. Ritchie earned her Ph.D. in GeoSciences from the University of Edinburgh and has made significant contributions to research as the former Head of Research.

The dataset provides a comprehensive overview of causes of death globally, encompassing various categories from infectious diseases to non-communicable diseases and external causes like road injuries and suicide. Each entry in the dataset represents a country in a specific year, with columns detailing different causes of death and corresponding values indicating the number of deaths attributed to each cause. This dataset serves as a crucial resource for understanding global health trends and evaluating the efficacy of public health interventions. Through analysis, it enables insights into the changing landscape of leading causes of death over time and across different regions, facilitating informed decision-making and targeted interventions in disease prevention and health promotion efforts.

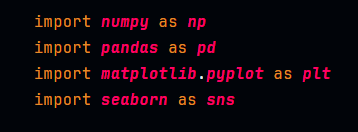
## I.3 Introducing some of the libraries and configurations for Python used in this paper

In this paper, I leverage several Python libraries to analyze and visualize the dataset effectively. The following libraries are utilized:

**NumPy (np)**: NumPy is a fundamental package for scientific computing in Python. It provides support for arrays, matrices, and mathematical functions, making it essential for numerical computations and data manipulation.

**Pandas (pd):** Pandas is a powerful library for data manipulation and analysis. It offers data structures like DataFrame and Series, along with functions to clean, transform, and analyze data efficiently. Pandas is particularly useful for handling structured data, such as the dataset under examination.

**Matplotlib (plt):** Matplotlib is a versatile library for creating static, interactive, and animated visualizations in Python. It offers a wide range of plotting functions and customization options, enabling the creation of various types of plots, including line plots, scatter plots, histograms, and more.

**Seaborn (sns):** Seaborn is a statistical data visualization library built on top of Matplotlib. It provides high-level functions for creating informative and attractive statistical graphics. Seaborn simplifies the process of generating complex visualizations by offering built-in themes, color palettes, and statistical estimation functions.

*Table 1.2: Library import syntax in the script.*

Additionally, specific configurations and settings may be applied to enhance the readability and aesthetics of the visualizations produced in this paper. These configurations may include adjustments to plot sizes, font styles, axis labels, and other visual elements to ensure clarity and coherence in the presentation of results.

By leveraging these libraries and configurations, I aim to conduct a thorough analysis of the dataset and present the findings through clear and insightful visualizations, facilitating a comprehensive understanding of global causes of death and their implications for public health policies and interventions

# Data Exploration

## II.1. Dataset description

The dataset used in this essay includes information on the number of deaths from different causes. This data is collected from various sources and compiled in a CSV file.

Specifically, the dataset includes the following columns:

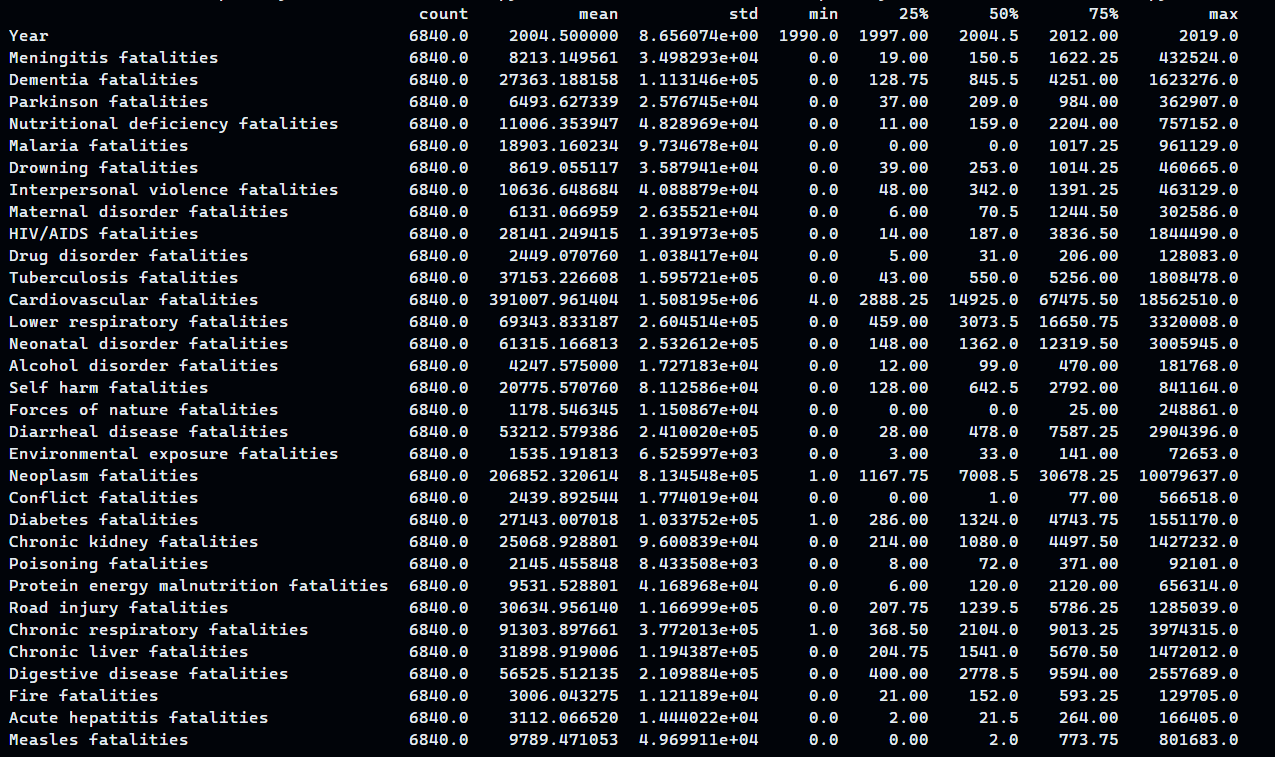
**file\_path = "dataset.csv"  
df = ***pd***.***read\_csv***(file\_path)  
df.***info***()

*Table 2.1: Name of columns in its dataset and data type*

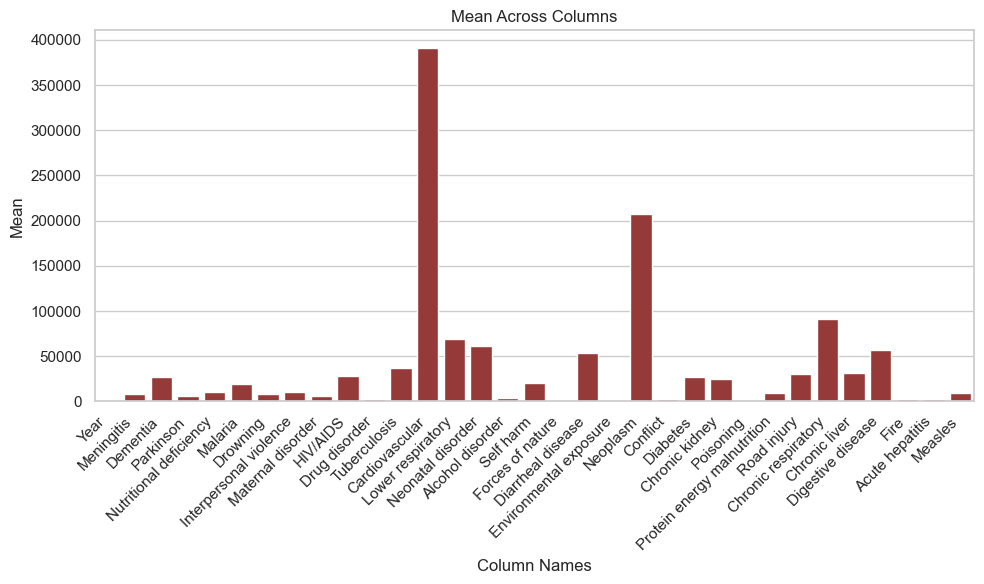
The data set has 35 columns, 6840 rows, and no null values. There are 2 columns with data type String and the remaining 33 columns are int64.

## II.2. Dataset description

Descriptive statistics play a crucial role in analyzing and summarizing the characteristics of a dataset. In this section, we will employ various descriptive statistical measures to gain insights into the dataset on global causes of death.

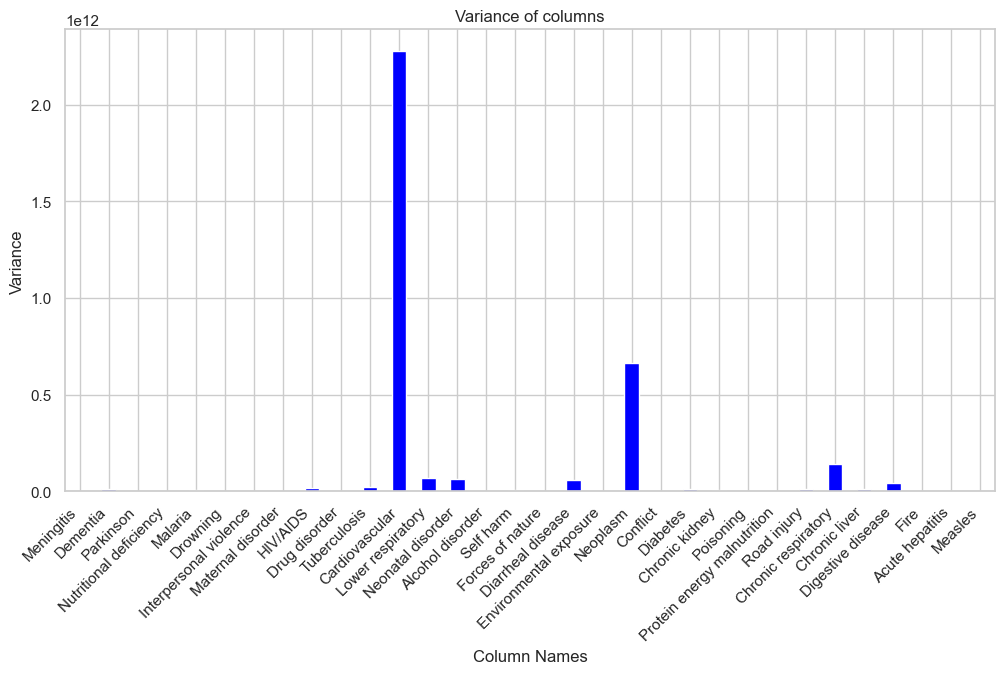
***print***(df.***describe***().T)

*Table 2.2: Detailed descriptive statistics about the columns in the dataset*

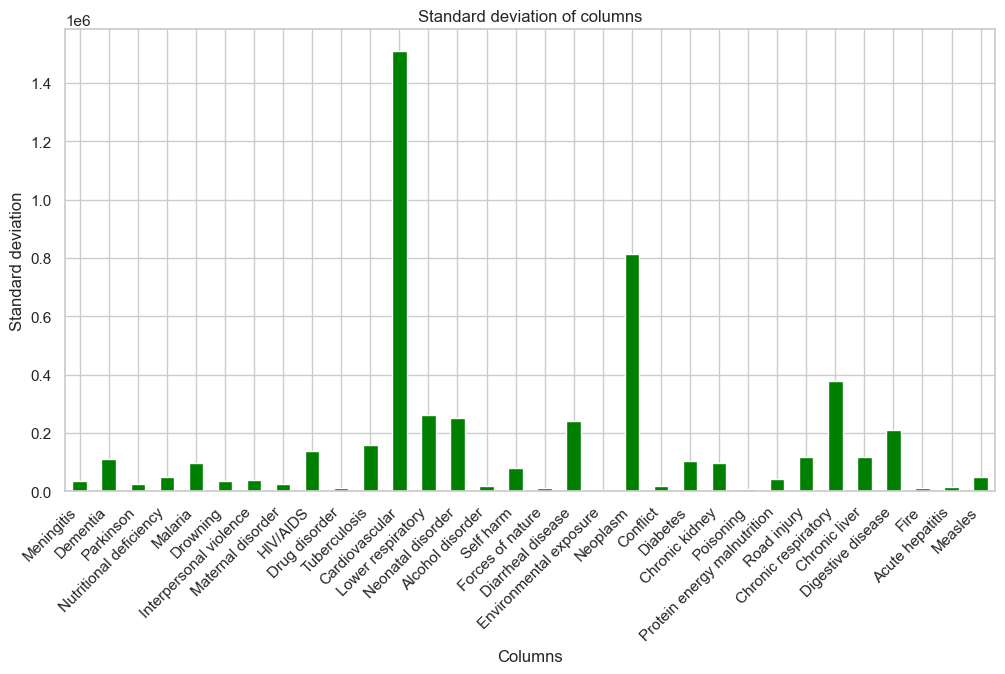


*Table 2.3: The chart means on columns*

Based on the chart, we can see that there are three causes of death with significant average value than other causes, namely "Cardiovascular Fatalities", "Neoplasm Fatalities", and "Chronic Respiratory Fatalities ". This shows that these causes may have caused a large number of deaths during the time and location that the data is collected.



*Table 2.4: The chart variance on columns*



*Table 2.5: The chart standard deviation on columns*

# Data preprocessing

## III.1. Read and process redundant data

After surveying the data set, we can see that the word "fatalities" is repeated too many times in the column names, so I removed it to facilitate information retrieval, information exploitation, and analysis, visualize the data later.

*# Remove the word "fatalities" repeated multiple times in column names to compact the dataframe*

df.columns = df.columns.str.***replace***(' fatalities', '')

## III.2. Data classification

By analyzing the descriptive statistics of the data set in the step above, I created a new data set consisting of columns containing only numeric data types to facilitate statistics and calculations from which to visualize data. Get the data in the best possible way.

*# Remove the 'Year' column and select only numeric columns from the DataFrame*

df\_numeric = df.***drop***(['Year'], ***axis***=1).***select\_dtypes***(***include***=[***np***.number])

Next, I check to see what countries are in the "Entity" column. I realized that, in addition to countries around the world, there are also international organizations that was collected data.

countries = df['Entity'].***unique***()

***print***("Countries included in the dataset:")

for country in countries:

***print***(country)

The "Entity" column has 228 unique values, there are 12 international organizations and 216 countries. Because of concerns that during the calculation process, there may be duplicate data between international organizations and countries.

international\_organizations = ["East Asia & Pacific wb",

                               "Eastern Mediterranean Region who",

                               "Europe & Central Asia wb",

                               "European Region who",

                               "Latin America & Caribbean wb",

                               "G20",

                               "Middle East & North Africa wb",

                               "North America wb",

                               "South Asia wb",

                               "SouthEast Asia Region who",

                               "SubSaharan Africa wb",

                               "Western Pacific Region who"]

df\_international\_organizations = df[df['Entity']/

.***isin***(international\_organizations)]

df\_countries = df[~df['Entity'].***isin***(international\_organizations)]

I separated this data set into 2 parts: **"df\_international\_organizations"** and **"df\_countries".** This will facilitate the analysis and visualization process for future research.

By drawing a pie chart, we can see the proportion between the number of countries and international organizations in the dataset.

*# Plot*

***plt***.***pie***([num\_organizations, num\_countries],

***explode***=(0.1, 0),

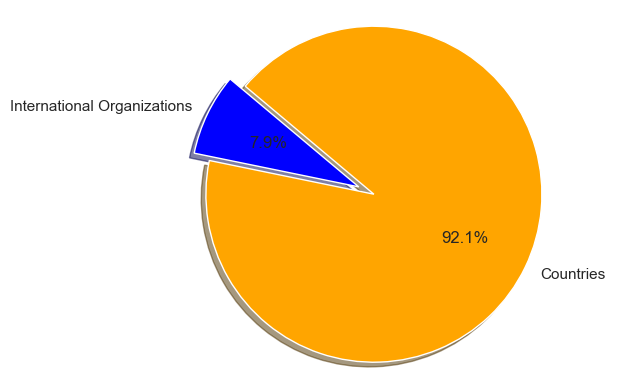
***labels***=('International Organizations', 'Countries'),

***colors***=['blue', 'orange'],

***autopct***='**%1.1f%%**', ***shadow***=True, ***startangle***=140)

***plt***.***axis***('equal')

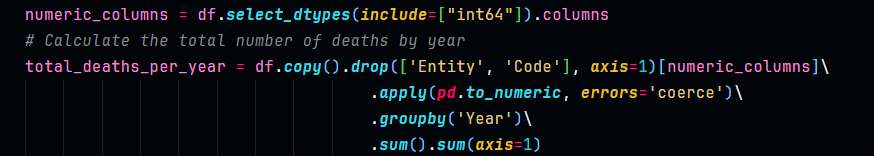
***plt***.***show***()

**

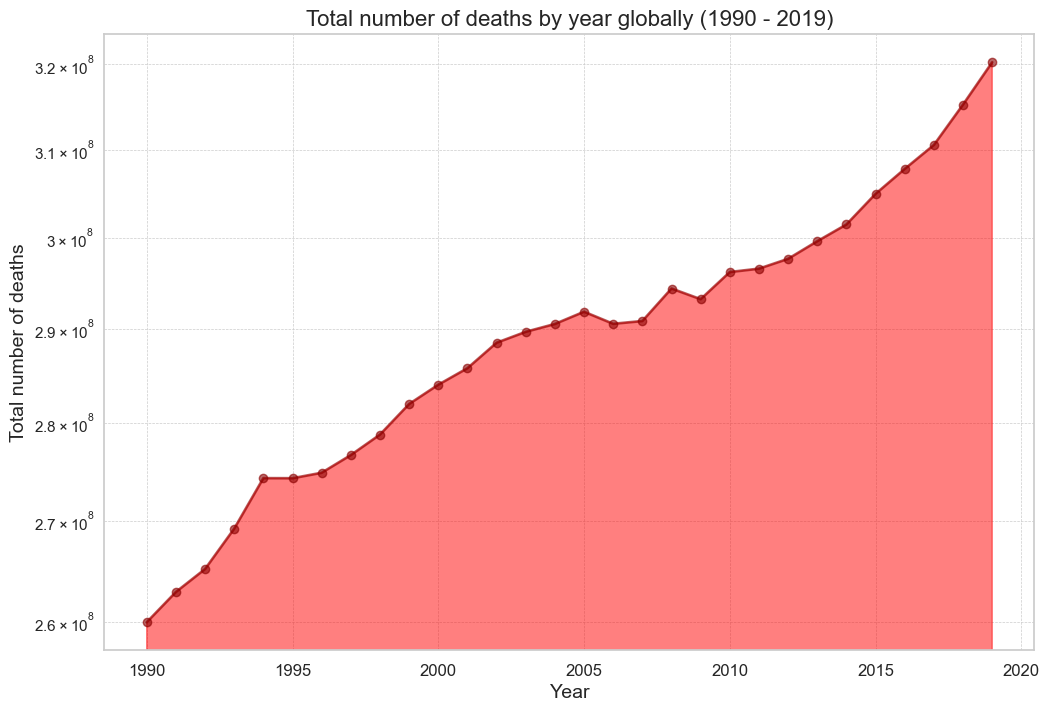
*Table 3.1: The chart shows the ratio of international and   
national organizations in the world in the dataset*

# Data Analysis and Visualization

## IV.1. Overall trend analysis and data visualization

The number of deaths in each region is always different and changes over time. To have a clearer view of the number of deaths worldwide over the years, we can draw a chart for a more intuitive view.

*Table 4.1: Code for visualizing data about the total of deaths   
by year globally in the script.*



*Table 4.2: The chart shows the total number of deaths   
worldwide from 1990 to 2019*

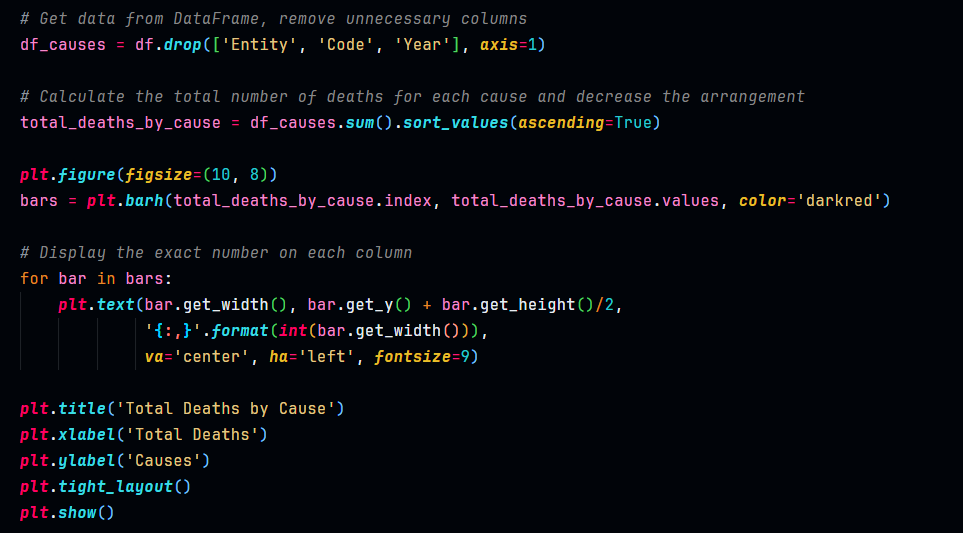
Based on the chart, we can see that the total number of deaths globally has increased very rapidly from 1990 to 2019, from 2.6 billion people to 3.2 billion people in just 30 years. But the increase is not uniform, specifically the number of deaths globally increased rapidly from 1990 to 2005, then it remained stable and increased sharply again in the following years.

The rapid increase in total global deaths from 1990 to 2019 can be explained by several factors:

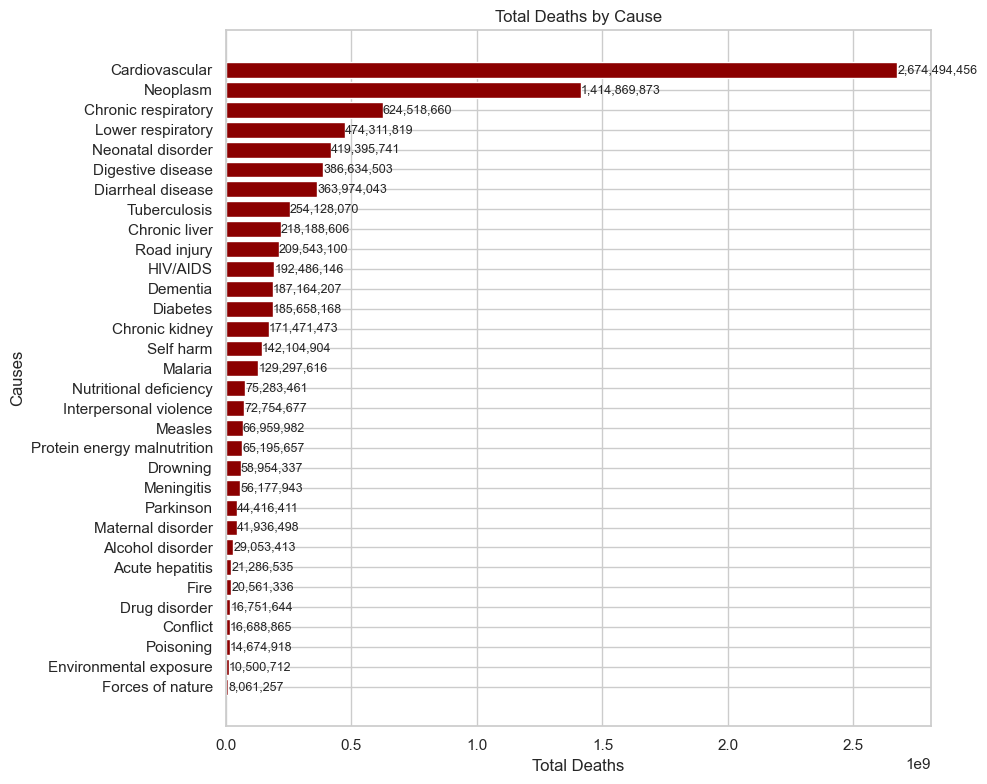
1. Population Growth: The world population has grown significantly during this period, from around 5.3 billion people in 1990 to approximately 7.7 billion people in 2019. A larger population may naturally lead to an increase in the number of deaths.
2. Health Advances: Despite the increasing number of deaths, the death rate per 100,000 people may decrease due to advances in healthcare such as improved disease treatments, vaccination programs, and strengthened healthcare systems.
3. Epidemics and Epidemiology: Epidemics such as HIV/AIDS, tuberculosis, and influenza have caused millions of deaths, especially in the 1990s and 2000s. Additionally, the spread of infectious diseases may have resulted in sudden spikes in deaths and contributed to the overall increase in mortality.
4. Improvements in Diagnostic Technology and Reporting: The development of medical technology may have enhanced the ability to diagnose and report causes of death, leading to the recording of more deaths in databases.
5. Changes in Diet and Lifestyle: Changes in diet, lifestyle, and living environments may increase the risk of non-communicable diseases such as cardiovascular diseases, diabetes, and cancer, as well as deaths from non-infectious causes.

However, the uneven increase may reflect fluctuations in the impact of these factors over time, as well as the implementation of health strategies and policies in different periods.

Analyzing the causes of death is an important step in understanding the global burden of disease and developing effective interventions to reduce mortality. By identifying the leading causes of death, policymakers and health professionals can focus resources on the health issues that have the greatest impact on people's lives.



*Table 4.3: Code for visualizing data about the deaths by cause  
by year globally in the script.*

 *Table 4.4: The chart shows the total deaths by cause   
worldwide from 1990 to 2019*

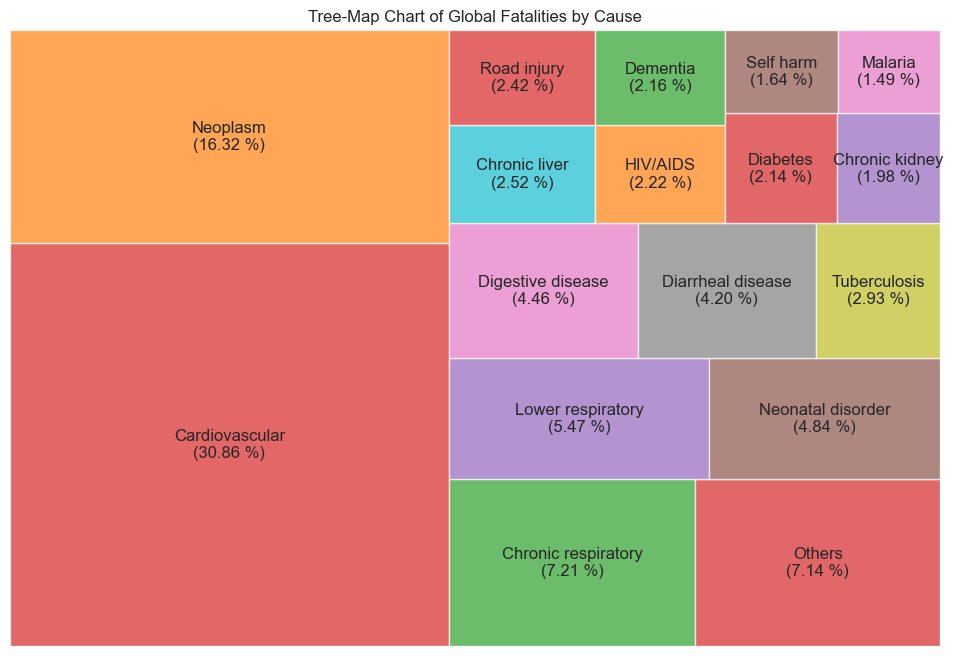
The chart provides an overview of the leading causes of death globally from 1990 to 2019. The data shows some notable trends, including an increase in mortality due to cardiovascular disease and stroke, declines in mortality from infectious diseases, and increases in mortality from cancer, Alzheimer's disease, and diabetes. These trends can be explained by a number of factors.

Including:

* Changing demographics.
* Lifestyle.
* Exposure to risk factors.

To have a more intuitive view of the rate of deaths by each cause, a tree map is a chart that may be the most reasonable choice.

*Table 4.5: Code for visualizing data about the deaths by cause  
by year globally with tree-map in the script.*

*Table 4.6: The chart shows the total deaths by cause   
worldwide from 1990 to 2019 with tree-map*

The Treemap chart depicts the number of deaths from leading causes globally from 1990 to 2019. The chart uses rectangles of different sizes to represent the death rate from each cause. The larger the size of the rectangle, the higher the mortality rate.

**Data analysis**

* Cardiovascular disease was the leading cause of death throughout this period, accounting for 30.86% of all deaths in 2019.
* Cancer is the second leading cause of death, accounting for 16.32% of all deaths in 2019.
* Lower respiratory disease is the third leading cause of death, accounting for 12.68% of all deaths in 2019.
* Digestive diseases are the fourth leading cause of death, accounting for 8.56% of all deaths in 2019.

**Discuss**

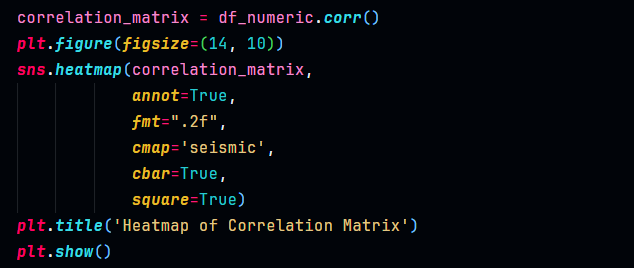
The increase in mortality from cardiovascular disease and cancer can be attributed to a number of factors, including aging populations, sedentary lifestyles, unhealthy diets and smoking. The decline in mortality from lower respiratory and digestive diseases can be attributed to improved sanitation, drinking water, and vaccination. The decline in infectious disease mortality can be attributed to the development of antibiotics, vaccines, and other public health measures.

**Compare with Chart 4.4**

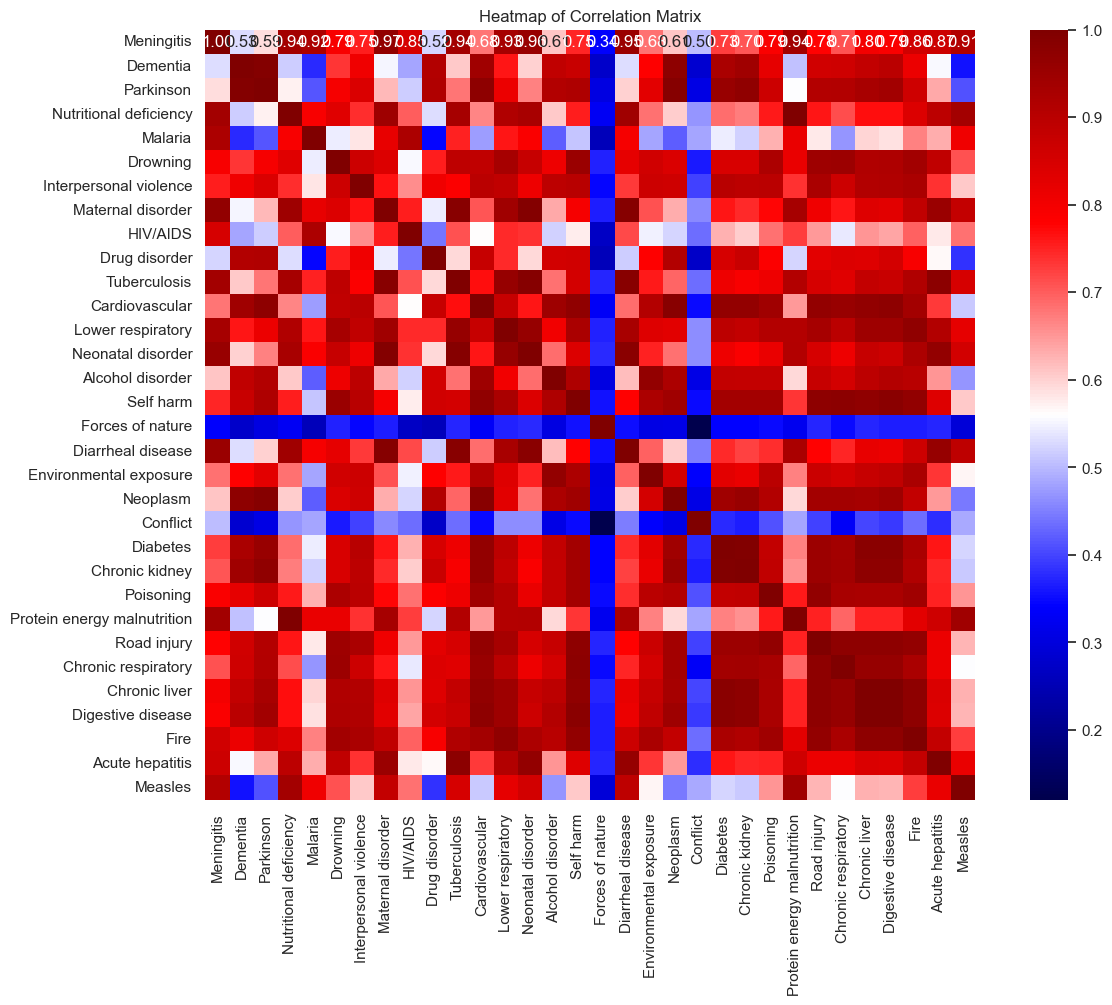
Treemap charts provide a visual way to compare death rates from different causes. Chart 4.4, on the other hand, provides a way to track mortality trends over time. Both charts provide valuable information about the global burden of disease.

## IV.2. Find relationships between causes of death and visualize with heatmap

From the above information, we can see that the number of deaths due to each cause is uneven, but do these causes have any interaction with each other? To clarify that, I calculated the ***correlation coefficient*** *(1)* of each pair of causes with each other and visualized it on a***heatmap*** *(2)* chart with a correlation matrix.

1. *A correlation coefficient is a statistical measure quantifying the linear relationship between two variables. It is used to analyze data sets or components of a random variable. Correlation coefficients range from -1 to +1, where values closer to ±1 indicate a stronger correlation, while 0 indicates no correlation.*
2. *In the correlation matrix, the heat map is a 2-dimensional data representation using colors to show the degree of correlation between pairs of variables. Each cell in the matrix is colored to correspond to the degree of correlation between the two corresponding variables. High correlation values are often represented by contrasting or light colors, while low correlation values are represented by dark colors. This helps identify correlation patterns in data in an intuitive and easy-to-understand way.*

*Table 4.7: Code for visualizing correlation matrix  
 by heatmap in the script.*



*Table 4.8: The Heatmap chart showing a correlation   
between causes of death*

Based on the heatmap chart above, we can see a number of dark red points representing high correlation coefficients. To explain the reasons leading to pairs of causes with such high correlation coefficients, we need to Consider the following factors:

* + - **Dementia and Parkinson's:**

Dementia and Parkinson's are both diseases that involve the deterioration of neurological function, and both have common biological mechanisms, including the destruction of nerve cells in the brain.

The co-occurrence of both diseases can be explained by the fact that they often appear in the same age group or that there are some common risk factors, such as age, genetics, or environment.

* + - **Parkinson's and Neoplasm:**

There are some studies that have shown a link between Parkinson's and higher rates of certain types of cancer, but the exact mechanism is not well understood.

Some theories include the effects of Parkinson's medications, delays in movement, and increased risk of DNA damage.

* + - **Nutritional deficiency and Protein-energy malnutrition:**

These two causes often go together and both cause problems related to inadequate nutrition for the body.

The almost perfect correlation coefficient between them can be explained by the similarities in the mechanisms that cause symptoms and complications of the disease.

* + - **Maternal disorder and Tuberculosis/Neonatal disorder/Diarrheal disease:**

Underlying maternal diseases often affect the health of the fetus and newborn through a variety of mechanisms, including transmission through the stomach, weakened immune systems, and interference with feeding and care. young.

* + - **Diabetes and Chronic kidney/Chronic liver/Digestive disease:**

Diabetes can cause a variety of health problems, including affecting the function of vital organs such as the kidneys and liver, as well as causing digestive problems.

* + - **Chronic liver and Digestive disease:**

Chronic liver and Digestive disease often go together because the liver is responsible for processing toxic substances and producing substances necessary for digestion.

The correlation coefficient between "Forces of nature" and "Conflict" exhibits a low degree of correlation with respect to other cause pairs in the data set. This decrease can be explained by a number of reasons:

An important factor is the lack of harmony in the nature and mechanism of these two factors. "Forces of nature" is often related to natural phenomena such as earthquakes and floods, while "Conflict" mainly originates from social, political, or cultural conflicts. Differences in origin and causative mechanism may minimize the degree of correlation between them.

In addition, diversity in space and time also plays an important role. "Forces of nature" can often affect large areas and can last over time, while "Conflict" often occurs on a local level and can have a short-term impact on an area.

The mortality data we have analyzed are an important part of public health and public health research.

Global improvements in health since the 1990s have had a significant impact on global mortality patterns. The spread of infectious disease control programs, vaccination, and disease prevention measures has reduced mortality from many causes, especially infectious diseases. However, with improvements in healthcare, we also face new challenges, such as strengthening mental health care and treating chronic lifestyle diseases.

Furthermore, conflicts and wars during this period had serious consequences for the health of communities, including death and illness related to stress and violence. Climate change has also contributed to a host of new health problems, thereby increasing the risk of death from causes such as hunger and infectious diseases.

With advances in medicine and technology, we have seen increases in life expectancy and improvements in quality of life, but we also face new challenges, including increased care for the elderly and treatment of non-communicable diseases.

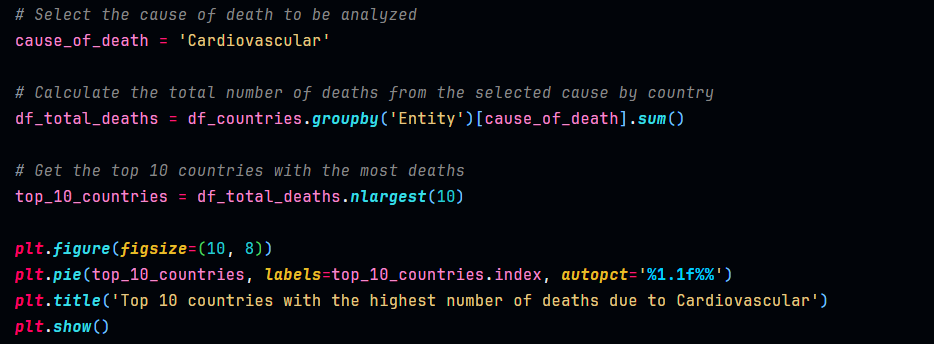
## IV.3. Comparison between countries and data visualization

Comparing the size of deaths by each cause between countries is one of the best ways to get a general assessment of the extent of differences and the impact of those factors on public health. globally.

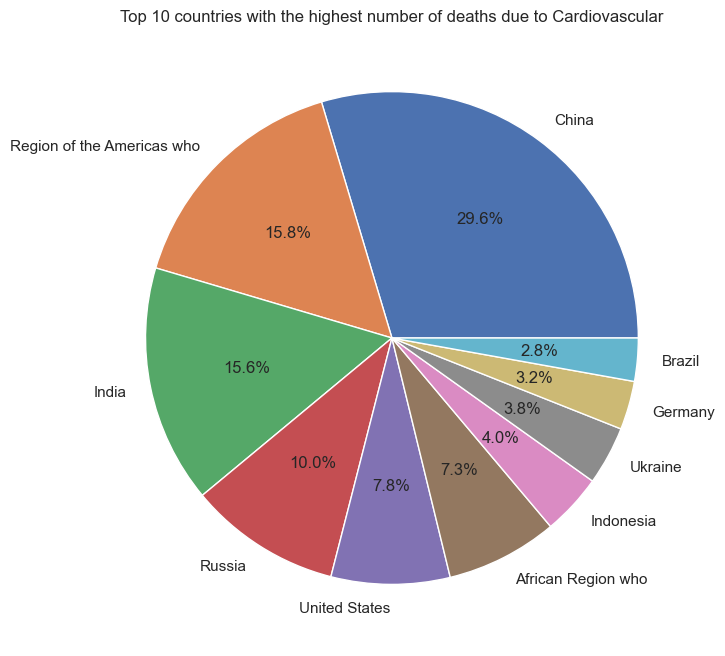
By analyzing mortality data from different causes across countries, we can see how each factor affects public health within each country. Comparisons between countries provide an overview of health outcomes and the impact of specific causes. In this way, we can identify pressing health problems, recognize disease trends, and provide effective interventions. At the same time, through comparisons between countries, we can also better understand the influence of environmental, cultural, economic, and political factors on mortality levels from different causes.

In addition, comparative analysis also helps us identify trends and changes in mortality levels over time and between regions, thereby proposing appropriate policies and prevention measures for each individual.

* **Cardiovascular diseases fatalities:**

Cardiovascular diseases (CVDs) remain a leading cause of mortality worldwide, posing significant challenges to public health systems across the globe. Understanding the distribution of CVD-related deaths among different countries provides valuable insights into the global burden of these conditions. To gain a clearer picture, let's explore the top 10 countries with the highest number of deaths attributed to cardiovascular diseases through a visual representation.

*Table 4.9: Code for visualizing Top 10 countries   
with the highest number of deaths due to Cardiovascular.*



*Table 4.10: The chart shows the 10 countries   
with the highest number of deaths due to Cardiovascular.*

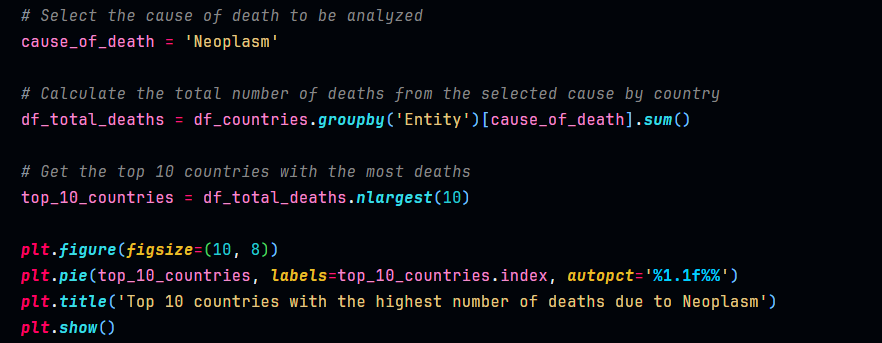
Pie chart depicting the distribution of deaths from cardiovascular disease in the 10 most affected countries and regions. China leads by a significant margin, accounting for 29.6% of deaths from the disease. The WHO Region of the Americas and India have almost identical rates, accounting for 15.8% and 15.6% of total deaths, respectively. Russia accounts for 10%, followed by the United States with 7.8%. The WHO Africa Region accounts for a smaller share at 7.3%. Brazil, Germany, Ukraine and Indonesia contributed relatively little, ranging from 2.8% to 4%.

From an analytical perspective, these data highlight the clear disparities in cardiovascular disease-related mortality globally. China's predominance raises important questions about basic health determinants including lifestyle factors, access to health services, and genetic catabolism in the populous country. This.

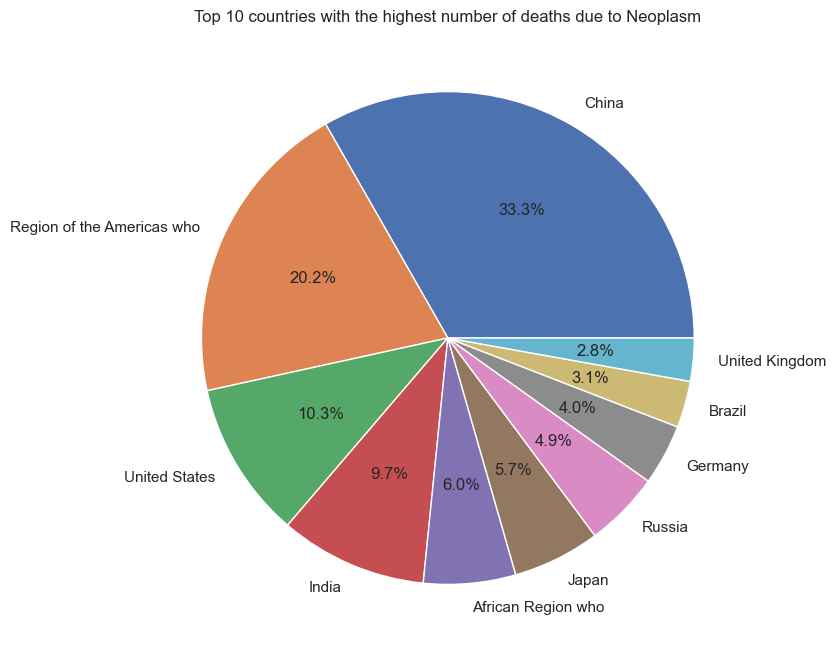
Furthermore, the comparable rates between the WHO Americas Region and India suggest that there may be similarities in health infrastructure or social habits that contribute to this result; however, given their distinct socioeconomic contexts, specific examination is necessary for an accurate view.

The presence of developed countries such as the United States and Germany alongside developing countries such as India and Brazil shows that cardiovascular disease is a global challenge that transcends economic conditions.

* **Neoplasm fatalities:**

As we delve into the realm of global health challenges, neoplasms, or tumors, stand out as a significant concern. Understanding the distribution of deaths caused by neoplasms across different countries sheds light on the burden of cancer-related mortality worldwide. To provide a comprehensive overview, let's explore the top 10 countries with the highest number of deaths attributed to neoplasms through a visual representation.

*Table 4.11: Code for visualizing Top 10 countries with the highest number   
of deaths due to Neoplasm.*

*Table 4.12: The chart shows the 10 countries   
with the highest number of deaths due to Neoplasm.*

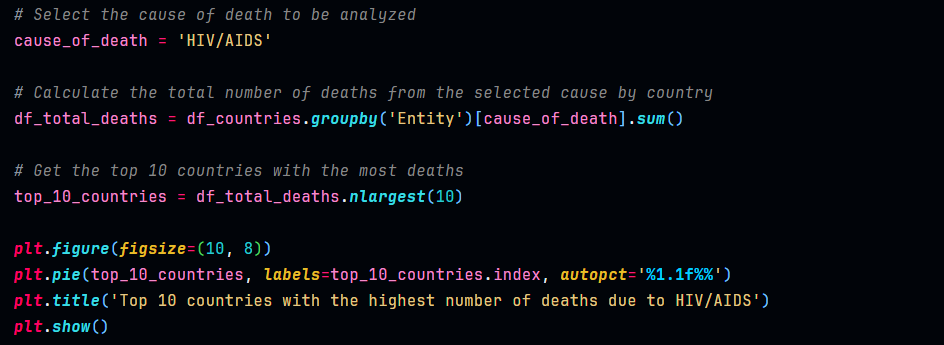
Pie chart depicting the distribution of deaths from neoplasm diseases in the 10 most affected countries and regions. China leads by a notable margin, accounting for 33.3% of deaths from the disease. The WHO Region of the Americas and India have almost identical rates, accounting for 20.2% and 10.3% of total deaths, respectively.

From an analytical perspective, these data highlight the clear disparities in neoplasm-related mortality globally. China's predominance raises important questions about basic health determinants including lifestyle factors, access to health services, and genetic catabolism in the populous country. This.

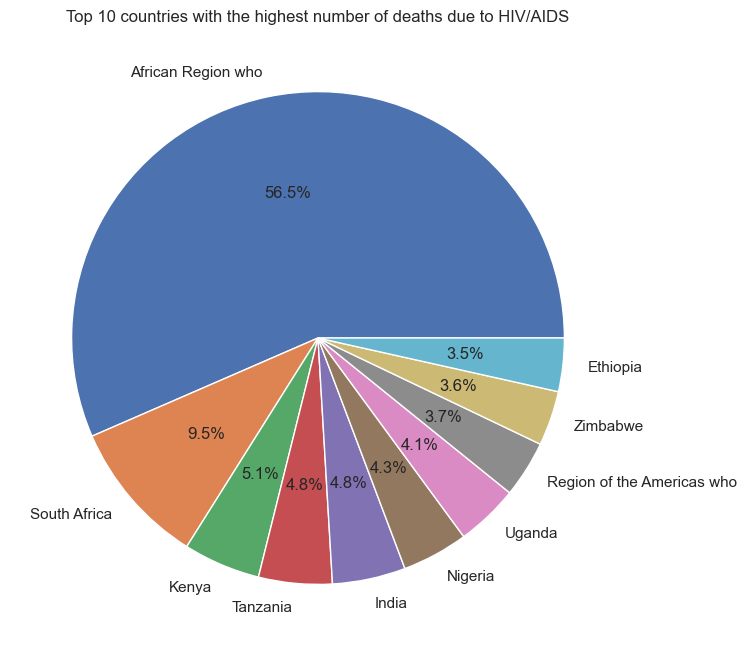
Furthermore, the comparable rates between the WHO Americas Region and India suggest that there may be similarities in health infrastructure or social habits that contribute to this result; however, given their distinct socioeconomic contexts, specific examination is necessary for an accurate view.

Countries such as Japan (6%), WHO African Region (5.7%), Russia (4.9%), Germany (4%), Brazil (3.1%), United Kingdom ( 2.8%) indicates moderate morbidity.

* **HIV/AIDS fatalities:**

As we delve into the realm of global health, the persistent challenge posed by HIV/AIDS demands our attention. Understanding the distribution of fatalities across various countries sheds light on the severity of this disease on a global scale. Exploring the top 10 countries with the highest number of deaths attributed to HIV/AIDS offers valuable insights into its prevalence and impact. Let's embark on this exploration by examining a visual representation of these countries.

*Table 4.13: Code for visualizing Top 10 countries with the highest number   
of deaths due to HIV/AIDS.*

*Table 4.14: The chart shows the 10 countries   
with the highest number of deaths due to HIV/AIDS.*

Based on the provided pie chart, it is evident that a substantial proportion of deaths attributed to HIV/AIDS are concentrated in specific nations, notably within the African continent. The African Region accounts for an overwhelming 56.5% of these fatalities, with South Africa (9.5%), Kenya (5.1%), Tanzania (4.8%), Nigeria (4.3%), Uganda (4.1%), Zimbabwe (3.7%), and Ethiopia (3.5%) being significantly impacted.

A comprehensive analysis reveals multifaceted causative factors contributing to this alarming statistic, rooted primarily in systemic and structural issues rather than isolated incidents or biological predispositions alone.

Healthcare Infrastructure: The affected African nations grapple with inadequate healthcare facilities, characterized by limited accessibility and quality of medical services essential for both preventive and curative interventions against HIV/AIDS.

Economic Constraints: Economic hardships exacerbate vulnerabilities; limited resources hinder comprehensive implementation of educational programs, access to antiretroviral therapy, and other essential healthcare services.

Stigmatization & Cultural Norms: Societal norms and stigmas associated with HIV/AIDS often deter individuals from seeking timely medical assistance or engaging in preventive behaviors, perpetuating the cycle of transmission and mortality.

Educational Gaps: A lack of comprehensive educational initiatives addressing sexual health and HIV/AIDS contributes to misinformation, inadequate awareness, and subsequently increased susceptibility to infection and mortality.

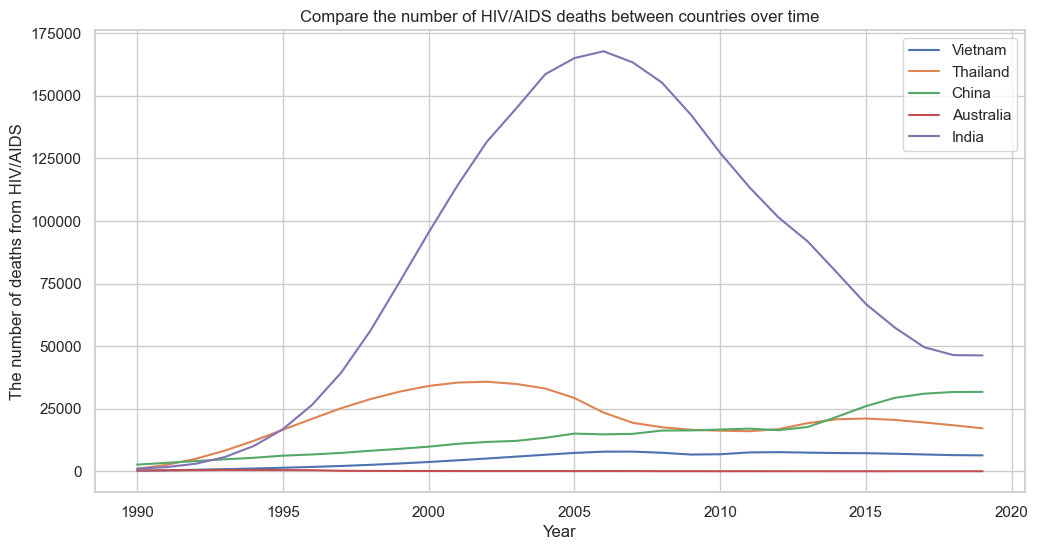
Outside Africa, India represents another nation grappling with a high number of deaths attributed to HIV/AIDS at 4.8%. Similar systemic issues – including healthcare accessibility constraints, economic challenges coupled with societal norms – contribute significantly to this scenario.

In conclusion, addressing the disproportionate impact of HIV/AIDS-related deaths necessitates a holistic approach encompassing enhanced healthcare infrastructure development; targeted educational initiatives; destigmatization campaigns; economic empowerment programs; international collaboration; policy reforms tailored towards inclusivity; accessibility; affordability; quality assurance in medical service delivery across these identified nations. This image shows that a significant majority of deaths due to HIV/AIDS occur in Africa according to this data representation which is interesting as it highlights an urgent need for targeted interventions within specific geographical locales for maximal impact on global health outcomes.

* **Analyze trends in deaths due to HIV**

The analysis provided highlights the disproportionate burden of HIV/AIDS-related deaths, particularly concentrated in specific nations, predominantly within the African continent. To delve deeper into this phenomenon and explore the temporal trends in HIV/AIDS fatalities across the selected countries, we will generate a line plot to examine the increase in the number of deaths due to HIV/AIDS over the years.

Utilizing the data presented in the analysis, we have identified the 10 countries with the highest number of deaths attributed to HIV/AIDS. By aggregating and visualizing the HIV/AIDS fatality data from these nations over the years, we can elucidate any discernible patterns, temporal variations, or noteworthy trends in HIV/AIDS-related mortality rates.

*Table 4.15: Code for visualizing Top 10 countries with the highest number   
of deaths due to HIV/AIDS.*

*Table 4.16: The chart shows the Comparison of   
HIV/AIDS Mortality Across Selected Countries (1990-2019)*

The chart shows a significant increase in deaths in India in the early 2000s, peaking at more than 150,000 deaths around 2005 before falling sharply.

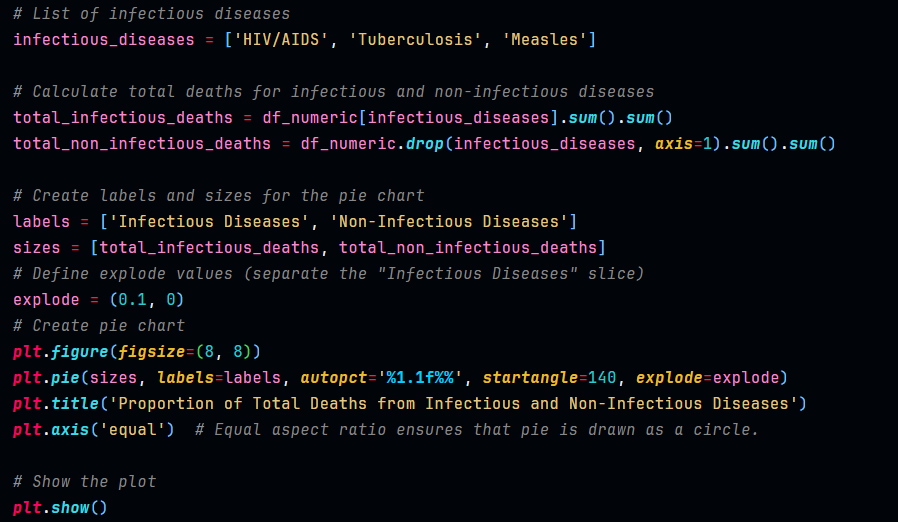
India had a spike in HIV/AIDS deaths in the early 2000s, peaked around 2005 with more than 150,000 deaths, and then declined sharply. This may be due to the development of public health programs, improved access to health services, and the spread of antiretroviral therapy (ART).

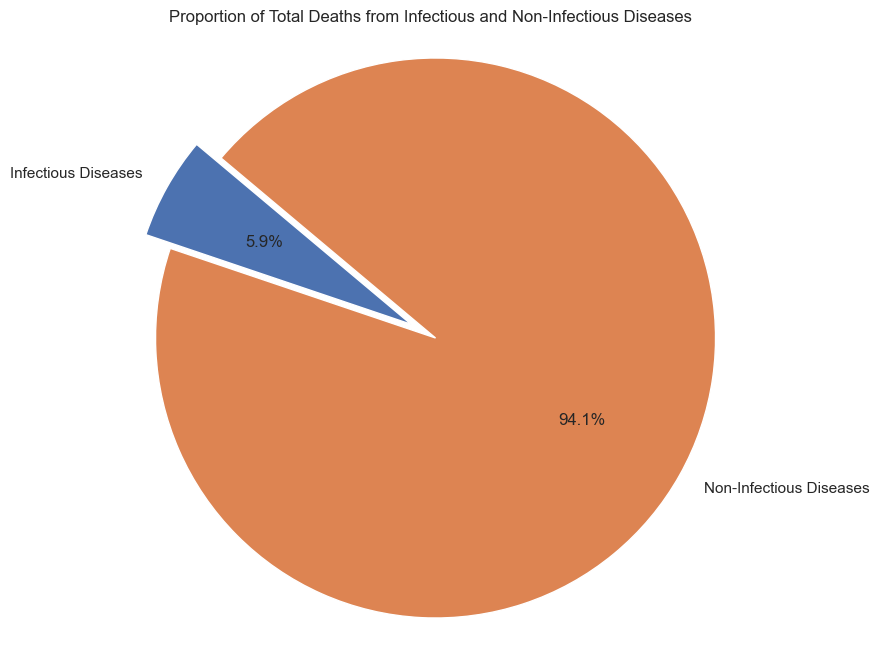
Meanwhile, other countries such as Vietnam, Thailand, China and Australia maintained stable or slightly increased mortality levels throughout this period. This may reflect differences in health infrastructure, public health policies and socioeconomic situations between countries.

## Tree map of causes of death globally in 2019, with non-communicable diseases in blue, communicable or infectious diseases in red, and injuries in green. The most common causes of deaths are non-communicable diseases such as heart diseases and cancers, while injuries and especially deaths from violence are rare.IV.4. Globally, non-communicable diseases are the most common causes of death

*Table 4.17: The chart shows what people died from globally, in 2019.  
Source:* [*https://ourworldindata.org/causes-of-death*](https://ourworldindata.org/causes-of-death)

In the year 2019, a comprehensive visualization of global mortality trends offers a stark portrayal of the myriad afflictions that led to the cessation of life. Each box within the chart serves as a poignant testament, symbolizing a distinct cause of death, with its size intricately linked to the magnitude of lives lost to that particular affliction.

At the forefront of this somber depiction lies the prevalence of non-communicable diseases (NCDs), represented in a solemn shade of blue. These maladies, encompassing cardiovascular diseases, cancer, and chronic respiratory disorders, stand as silent sentinels to the fragility of human existence. Unlike their infectious counterparts, NCDs insidiously permeate lives, often unfolding gradually over time without the contagion characteristic of infectious diseases.

*Table 4.18: Code for visualizing the chart shows the proportion of total deaths due to communicable and non-communicable diseases.*

*Table 4.19: The chart shows the proportion of total deaths due to communicable and non-communicable diseases.*

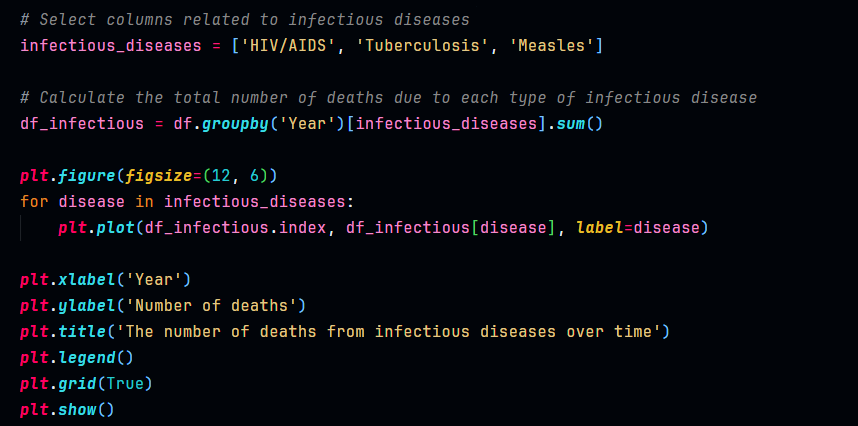
Of these, heart diseases emerge as the foremost protagonist, solemnly presiding over a third of all global deaths. In their wake, cancers loom large, solemnly claiming nearly one-fifth of lives lost. Together, heart diseases and cancers constitute the cause of every second death, casting a shadow of morbidity and mortality that spans the breadth of humanity.

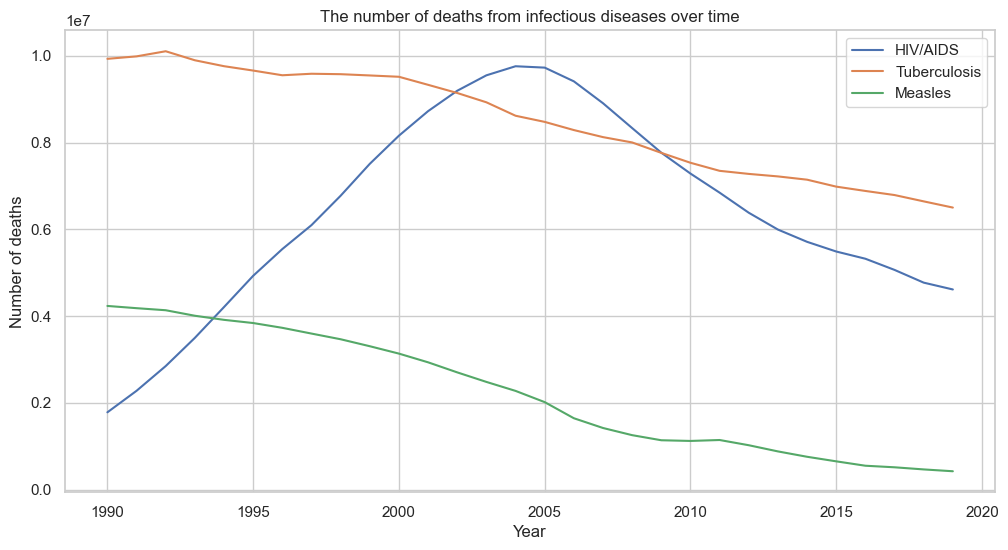
Contrasted against this backdrop of chronic ailments are the infectious diseases, depicted in a striking hue of red. Responsible for approximately one-in-seven deaths globally, these afflictions include pneumonia, diarrheal diseases, tuberculosis, HIV/AIDS, and malaria. Within their virulent grasp lies a reminder of the ongoing struggle against infectious scourges, which, despite advancements in medical science, persist as formidable adversaries on the global stage.

A modest share of mortality, comprising around 4%, is attributed to the heart-wrenching realm of neonatal and maternal deaths. Similarly, accidents, though accounting for a comparable fraction, serve as poignant reminders of the fragility of human existence, often snatching lives in the blink of an eye.

Amidst this tapestry of mortality, the specter of violence casts its grim shadow, albeit in lesser measure. Approximately 1.3% succumbed to the silent anguish of suicide, while less than 1% met their demise through the tumultuous throes of interpersonal violence, whether in the form of homicide or the ravages of armed conflict.

Thus, the chart stands not merely as a visual artifact, but as a poignant narrative of the human condition, encapsulating the triumphs and tribulations, the victories and vulnerabilities that define our collective journey through the sands of time.

*Table 4.20: Code for visualizing the chart shows the number of deaths   
from infectious diseases over time.*



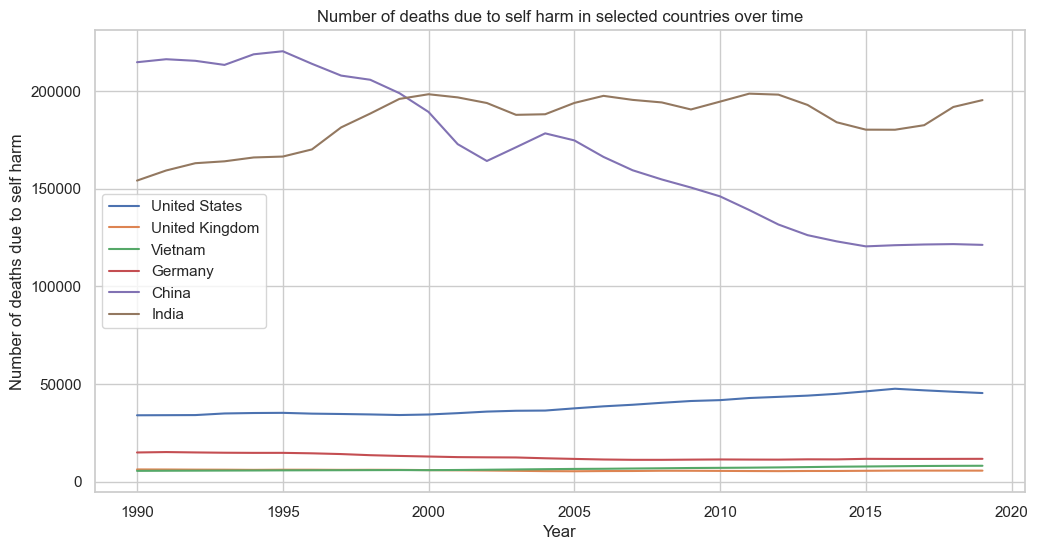
*Table 4.20: The chart shows the number of deaths   
from infectious diseases over time.*

## IV.5. Self-harm! an alarming problem in modern society

The trend analysis of self-harm-related fatalities across various nations provides valuable insights into the dynamics of mental health outcomes and societal challenges. In this exploration, we delve into the temporal trajectory of self-inflicted mortality, focusing on a selection of countries: the United States, United Kingdom, Vietnam, Germany, China, and India. By examining the fluctuations in self-harm fatalities over time within these diverse socio-cultural contexts, we aim to discern patterns, identify potential influencing factors, and glean nuanced understandings of this complex issue.

Through a comparative lens, we endeavor to uncover disparities and commonalities in self-harm trends among nations with distinct healthcare systems, cultural norms, and socio-economic landscapes. This analysis not only sheds light on the magnitude of the self-harm crisis but also underscores the imperative for targeted interventions and comprehensive mental health strategies tailored to the unique needs of different regions.

*Table 4.21: Code for visualizing the chart shows the number of deaths due to self harm in selected countries over time.*

*Table 4.22: Number of deaths due to self harm in selected countries over time**.*

Graph illustrating the number of deaths due to suicide in six selected countries (United States, United Kingdom, Vietnam, Germany, China, and India) from 1990 to 2019. The United States shows a significant increase throughout this period, while other countries showed relatively stable or slightly increasing trends.

United States: The sharp increase in deaths from suicide can be attributed to many causes, including increased social pressures, increased rates of mental disorders and lack of psychological support. This also reflects the need to strengthen psychological services and community education on this issue.

UK, Vietnam, Germany, China and India: Relatively stable or slightly increasing trends suggest that, despite certain efforts to reduce suicide mortality, there are still many challenges. Possible causes include social pressure, mental disorders, and lack of psychological support.

Self-harm is a serious global problem that requires attention and joint efforts from countries around the world. Addressing root causes such as social pressure, mental disorders and lack of psychological support is key to reducing self-harm mortality. Strong coordination between governments, NGOs, and communities is needed to raise awareness of this issue, remove the stigma associated with mental health, and ensure that everyone has access to quality mental health services.

* **Call for action for relevant parties to join hands to repel the problem of self-harm**

To reduce the death rate from self-harm, joint efforts are needed from relevant stakeholders, including:

* + Governments: Governments need to invest in mental health services, eliminate mental health-related stigma, and enact policies that support the poor, unemployed, and vulnerable groups other.
  + NGOs: NGOs can provide crisis support services, mental health education, and advocate for self-harm prevention policies.
  + Communities: Communities can create more supportive and inclusive environments, reduce stigma related to mental health, and encourage people to seek help when needed.
  + Individuals: Individuals can learn to recognize the warning signs of self-harm in themselves and others, seek help when needed, and support those struggling with mental health issues.

By working together, we can create a world where everyone can be supported and have access to essential mental health services, thereby reducing mortality by self-harm and saving more lives.

* **In addition to the actions outlined above, some other specific measures that can be taken include:**
  + Develop and implement evidence-based self-harm prevention programs.
  + Ensure adequate training for mental health professionals on suicide prevention.
  + Increase awareness of mental health issues and self-harm through public education and media campaigns.
  + Support research groups to develop new treatments and prevention methods for suicide.
  + Eliminate the accessibility of means of suicide.

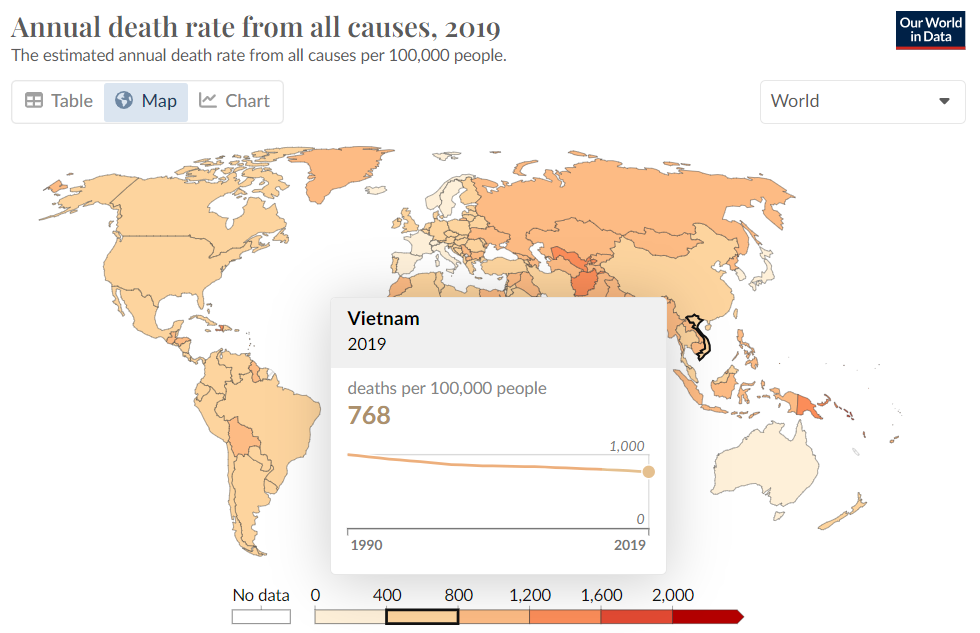
By taking these measures, we can hope to create a future where self-harm is no longer a major problem.

## IV.6. Analyze data on causes of death in Vietnam in particular

Vietnam, like many other countries in the world, faces many challenges in ensuring public health and reducing mortality rates. To better understand the country's health situation, analyzing data on causes of death is an important part.

In this section, we will focus on analyzing the most common causes of death in Vietnam and their changes over the years. The data used is collected from reliable sources and provides an overview of the public health situation over a period of time.

By understanding the trends and severity of causes of death, we can develop sound strategies and policies to improve public health and reduce unnecessary mortality.

The chart “Annual death rate from all causes, 2019” taken from Our World in Data clearly illustrates the change in this index over the years. This not only provides an overview of the global health situation but also helps us better understand the specific challenges and developments in Vietnam. Let's look at the following chart to get a deeper look at this issue.

*Table 4.23: Death rate from all causes in Vietnam 2019*

The chart shows annual mortality rates from all causes in Vietnam from 1990 to 2020. Based on data from reliable sources1, mortality rates in Vietnam have decreased significantly over the period. This.

* **Death rate in Vietnam:**
  + 2021: The death rate in Vietnam is reported to be 7,317%3.
  + 2023: Estimated mortality rate is 5.5 deaths per 1,000 people4.

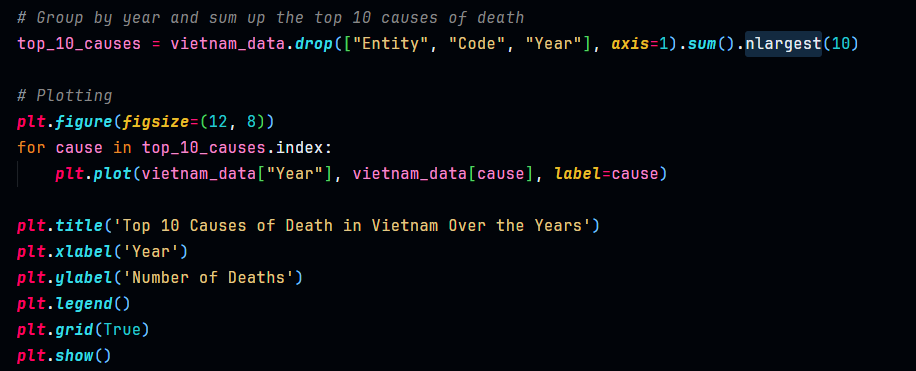
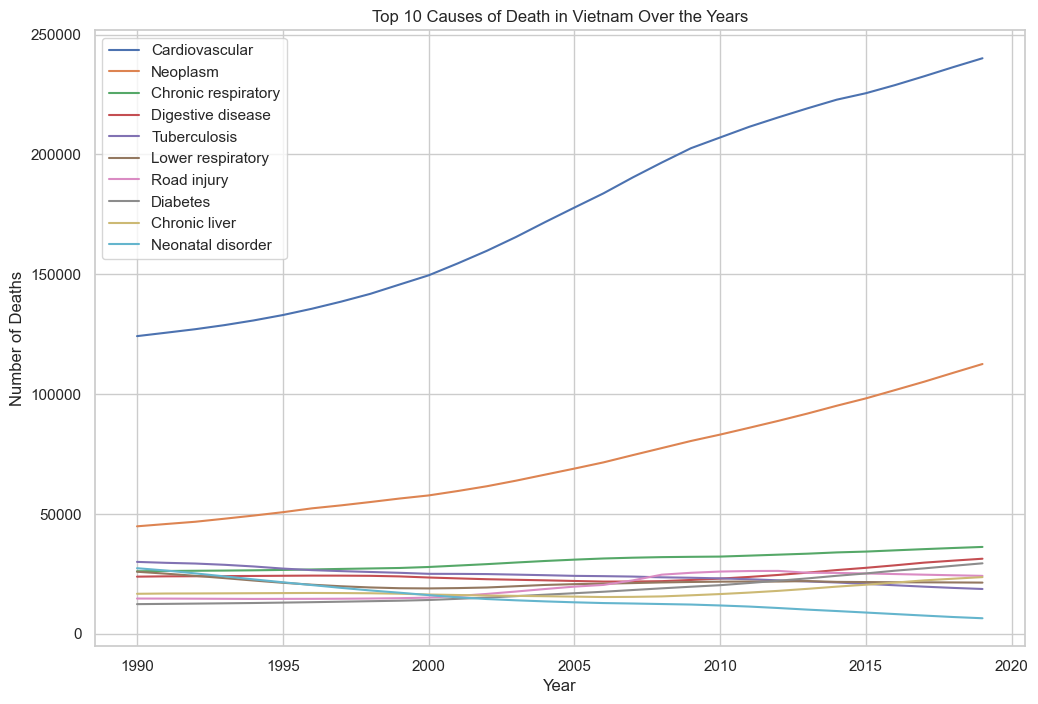
This decrease can be attributed to many causes, including improvements in health care, increased education, and economic development. However, it should be noted that the post-2019 figures are UN projections and therefore do not include any impact from COVID-19.

*Table 4.24: Code for defining the dataset about Vietnam.*

In order to get a more general overview of the mortality situation in Vietnam, instead of just focusing on mortality rates by age group and sex in a specific year, we will go deeper into analyzing mortality trends. death by each cause over a certain period of time.

* **By analyzing mortality trends by each cause, we can:**
  + Identify the main causes of death in Vietnam and the severity of each cause over time.
  + Evaluate the effectiveness of preventive health and treatment programs for common diseases.
  + Identify vulnerable populations for specific causes of death.
  + Forecast future mortality trends to devise appropriate intervention strategies.
* **Detailed information on mortality trends by cause will provide a solid foundation for policymakers, public health researchers and health organizations to:**
  + Allocating health resources effectively to prevent and treat the leading causes of death.
  + Develop intervention programs targeting high-risk populations.
  + Raise people's awareness about causes of death and preventive measures.

By analyzing mortality trends by cause, we can build a more effective health care system, significantly contributing to prolonging life expectancy and improving the quality of life for Vietnamese people. Male.

*Table 4.25: Code for visualizing the chart shows the top 10 Causes of Death in Vietnam Over the Years*

*Table 4.26: The chart shows the top 10 Causes of Death in  
 Vietnam Over the Years*

**Based on the above chart and analysis of mortality trends by cause in Vietnam from 1990 to 2020, the following general comments can be made:**

**1. General trend:**

* Overall, the mortality rate from all causes in Vietnam tended to decrease between 1990 and 2019. This is a positive signal showing progress in the field of health and wellness. community.
* However, there are still a number of causes of death that have high mortality rates and require special attention, including: cardiovascular disease, cancer, traffic accidents, diabetes, and obstructive pulmonary disease. chronic.

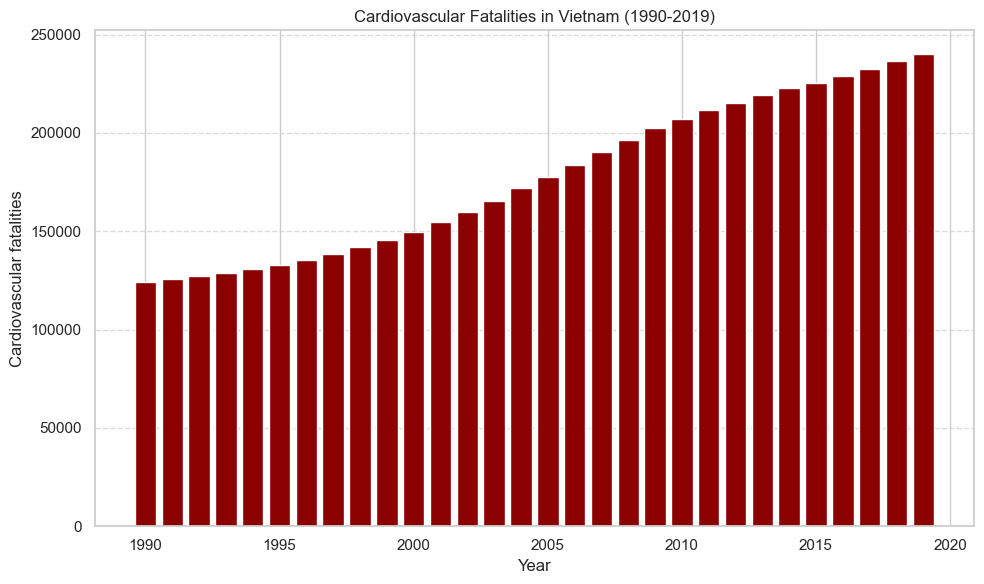
**2. Analysis by each cause:**

* **Cardiovascular disease:** Was the leading cause of death in Vietnam during the research period. The mortality rate from cardiovascular disease tended to decrease between 1990 and 2000, but then increased slightly between 2000 and 2019. This may be due to an increase in the proportion of people elderly, unhealthy diet and sedentary lifestyle.
* **Cancer:** Is the second leading cause of death in Vietnam during the research period. Cancer mortality rates tended to increase steadily throughout the study period. This may be due to increased smoking rates, environmental pollution and population aging.
* **Traffic accidents:** The leading cause of death in the young age group (15-34 years old). The death rate from traffic accidents tended to decrease between 1990 and 2000, but then increased slightly between 2000 and 2019. This may be due to an increase in the number of traffic accidents. Means of transport, awareness of obeying traffic laws is limited and traffic infrastructure has not been improved.
* **Diabetes:** Is the fourth leading cause of death in Vietnam during the research period. Diabetes mortality rates tended to increase steadily throughout the study period. This may be due to unhealthy diets, sedentary lifestyles and increased obesity rates.
* **Chronic obstructive pulmonary disease:** Is the fifth leading cause of death in Vietnam during the research period. The mortality rate from chronic obstructive pulmonary disease tended to increase steadily throughout the study period. This may be due to high smoking rates and environmental pollution.

**3. Evaluation:**

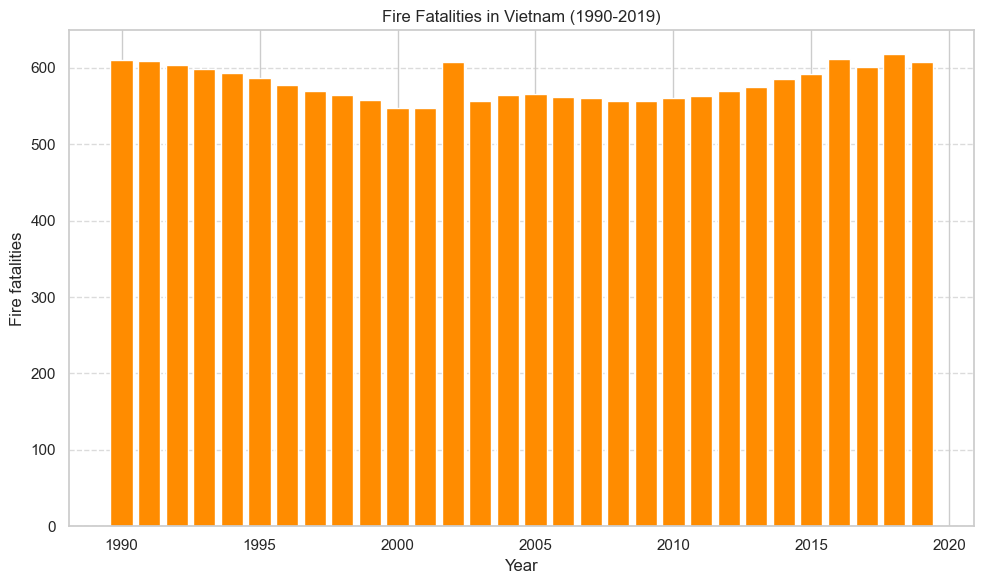
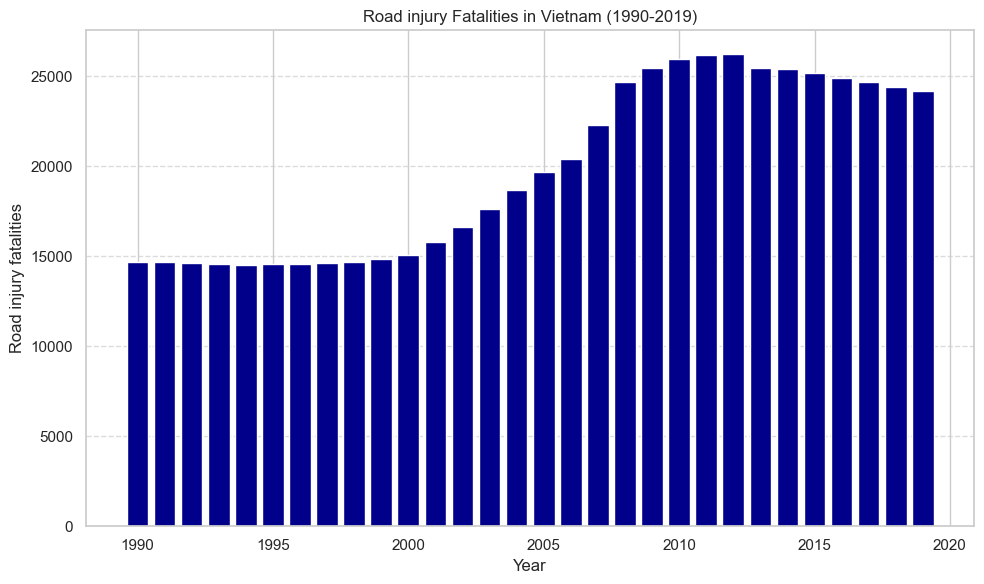
Overall, Vietnam has made significant progress in reducing all-cause mortality in recent years. However, there are still many challenges that need to be addressed to continue improving the health of the Vietnamese people.

* Preventive measures need to be strengthened to reduce mortality from leading causes such as cardiovascular disease, cancer, traffic accidents, diabetes and chronic obstructive pulmonary disease.
* It is necessary to raise people's awareness about the importance of a healthy lifestyle and the importance of regular health checks.
* Investment is needed in the health system to ensure that everyone has access to high-quality health care services.

**Some charts on trends of notable causes of death in Vietnam:**

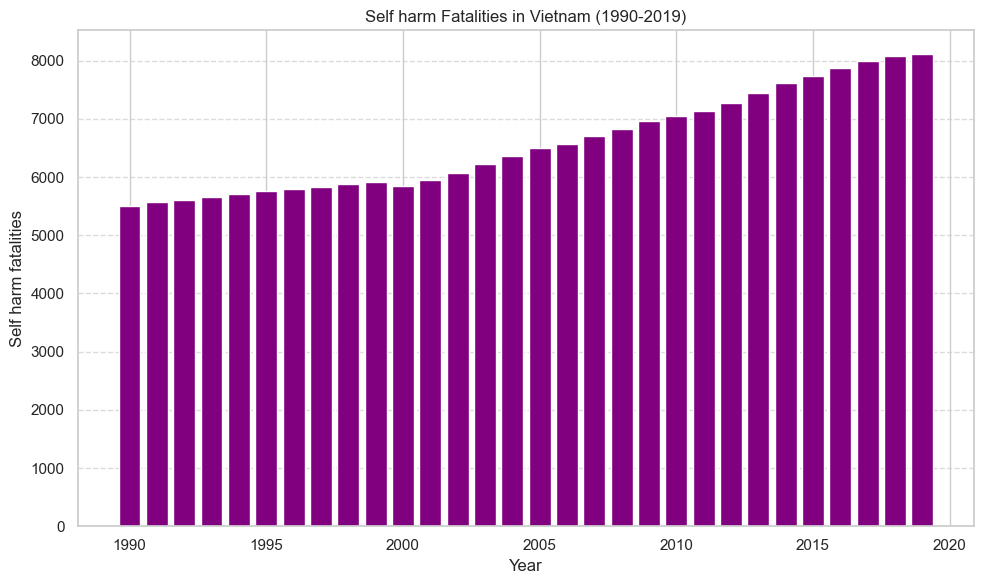
*Table 4.27: The chart shows Cardiovascular*

*Fatalities in Vietnam (1990-2019)*

*****Table 4.28: The chart shows Fire Fatalities in Vietnam (1990-2019)*

*Table 4.29: The chart shows Road injury*

*Fatalities in Vietnam (1990-2019)*

*****Table 4.30: The chart shows Self harm*

*Fatalities in Vietnam (1990-2019)*

* **Overview**

The charts show a significant increase in the number of traffic accidents and deaths due to accidents in Vietnam from 1990 to 2019. In particular, there is an increase in the number of deaths due to inter-traffic accidents. related to employee safety, number of road accidents, number of deaths due to fire and number of deaths due to cardiovascular problems.

* **Comment**

This increase shows that improvements in safety and health measures are needed to minimize harm to society in Vietnam. This is especially important for those working in the field of safety and those involved in road traffic.

* **Evaluate**

The graphs show that, despite certain efforts in Vietnam, more focus is needed on reducing the number of accidents and deaths due to accidents. This requires the cooperation of the whole society, from the government to each individual.

* **General situation**

In total, the current situation in Vietnam requires attention and quick action from everyone. We need to find effective solutions to reduce the number of accidents and deaths due to accidents, especially accidents involving safety officers and road traffic accidents. We also need to focus on reducing deaths from fires and cardiovascular problems. Only then can we create a safer and healthier society in Vietnam.

# Conclusion

## V.1. Summary of analysis results:

The analysis of causes of death across different countries and years revealed several key findings. Firstly, there has been a significant shift in the leading causes of death over time. While infectious diseases once dominated, their death rates have declined sharply, leading to non-communicable diseases such as cardiovascular diseases, cancers, and chronic respiratory diseases becoming the most common causes of death globally. This trend underscores the importance of addressing lifestyle-related factors and improving healthcare infrastructure to combat these non-communicable diseases effectively.

Additionally, the analysis highlighted disparities in the distribution of causes of death among different regions and countries. Certain regions, particularly low- and middle-income countries, continue to face a higher burden of infectious diseases compared to high-income countries. This disparity emphasizes the need for targeted interventions and resources to address the specific health challenges faced by these regions.

Moreover, the analysis shed light on the complex interplay of various factors contributing to the burden of disease, including healthcare access, socioeconomic status, environmental factors, and cultural influences. Understanding these factors is crucial for designing effective public health policies and interventions tailored to specific populations and contexts.

Overall, the analysis underscores the importance of ongoing surveillance and research to monitor trends in causes of death, identify emerging health threats, and inform evidence-based interventions aimed at improving global health outcomes.

## V.2. Proposal for further research directions:

While this analysis provides valuable insights into the causes of death and their trends over time, there are several areas for further research that could deepen our understanding and guide future public health efforts.

1. **Longitudinal studies:** Conducting longitudinal studies to track individuals' health outcomes over time can provide valuable insights into the long-term impact of various risk factors on mortality rates. This research could help identify early indicators of disease development and inform preventive strategies.
2. **Socio-economic determinants:** Further research is needed to explore the socio-economic determinants of mortality and their interactions with other risk factors. Understanding how factors such as income inequality, education, and employment status influence health outcomes can inform policies aimed at reducing health disparities.
3. **Impact of interventions:** Evaluating the effectiveness of public health interventions and healthcare policies in reducing mortality rates is essential for guiding resource allocation and decision-making. Future research could assess the impact of interventions such as vaccination programs, tobacco control measures, and access to healthcare services on mortality trends.
4. **Emerging health threats:** With the emergence of new infectious diseases and the ongoing threat of antimicrobial resistance, research into emerging health threats is critical. Surveillance systems and research efforts should focus on monitoring the spread of infectious diseases, identifying potential outbreaks early, and developing effective response strategies.
5. **Global health equity:** Addressing health inequities both within and between countries remains a pressing challenge. Future research should prioritize understanding the root causes of disparities in health outcomes and developing strategies to promote health equity on a global scale.

By addressing these research priorities, we can build on the findings of this analysis and continue to advance our knowledge of the determinants of mortality and strategies for improving public health outcomes worldwide.

# References

Dattani, S., Ritchie, H., Roser, M., Spooner, F. (2021). Our World in Data. Retrieved from [*https://ourworldindata.org/*](https://ourworldindata.org/)

Dattani, S. (2021). Global Health Data. [[Data file](Data%20file)]. Retrieved from [*https://ourworldindata.org/*](https://ourworldindata.org/)

United Nations. (2015). Sustainable Development Goals. Retrieved from [*https://sdgs.un.org/goals*](https://sdgs.un.org/goals)

Huynh De Thu, PhD. (Personal Communication, 2023).

Roser, M. (2021). Our World in Data: About. Retrieved from [*https://ourworldindata.org/about*](https://ourworldindata.org/about)

Spooner, F., PhD. (Personal Communication, 2021).

Ritchie, H., PhD. (Personal Communication, 2023).

World Health Organization. (2021). Global Health Observatory (GHO) data. Retrieved from [*https://www.who.int/data/gho*](https://www.who.int/data/gho)

Institute for Health Metrics and Evaluation (IHME). (2021). Global Burden of Disease Study 2019 (GBD 2019) Data Resources. Retrieved from [*http://www.healthdata.org/gbd/data*](http://www.healthdata.org/gbd/data)