Analysis of The Leading Causes of Death in the United States over the Years (2000-2019)

Razan Hussien
University of Tennessee at Chattanooga

Chattanooga, Tennessee, United States

December 11, 2023

***Abstract*— This research comprehensively analyzes mortality trends in the United States from 2000 to 2019, integrating extensive data to uncover patterns in death causes across gender and age groups. The study reveals an upward trend in mortality, notably spiking in 2017. Gender-based analysis indicates comparable mortality rates for males and females, yet 2016 marked an unusual increase in male deaths. The elderly, particularly those aged 70-89, exhibited the highest mortality rates, while the under-5 age group also showed significant vulnerability. The top causes of death in 2019 were consistent across genders, with non-communicable diseases dominating, including cardiovascular diseases, various cancers, and chronic respiratory conditions. The analysis extends to mortality distribution across different demographics, highlighting the disproportionate impact of specific diseases on certain age groups. For example, the elderly were predominantly affected by chronic diseases, while young adults faced more deaths from injuries and non-communicable diseases. This granular view of mortality, segmented by age and gender, underscores the need for tailored healthcare strategies. The study's findings are instrumental in shaping public health policies and interventions, providing a nuanced understanding of the evolving health challenges in the American population. It emphasizes the critical role of demographic-focused approaches in healthcare planning and resource allocation, offering vital insights for future health initiatives and research.**

Keywords—United States, Mortality Trends, Public Health, Demographic Analysis, Non-Communicable Diseases, Gender, and Age Disparities.

I. INTRODUCTION

Research in public health and disease studied often focuses on understanding how and why death rates vary. This includes looking at the main reasons people die and how these reasons change across different ages and groups of people. Knowing this information is very important for creating successful public health plans and actions.

The 2021 report by the National Academies of Sciences, Engineering, and Medicine highlights a critical public health issue: the rising death rates among young and middle-aged adults in the United States. This trend, observed since 2010, is primarily driven by increases in deaths from drug overdoses, alcohol-related issues, suicides, and cardio-metabolic diseases such as diabetes and hypertension-related heart diseases. This alarming rise in mortality rates, particularly among working-age adults, has profound implications for families, the workforce, and the economy. The mentioned report highlights that this issue is not isolated to specific demographic groups; it affects young adults, and many racial and ethnic groups, and spans both rural and urban areas[1].

A key aspect of the report is its focus on socioeconomic differences in mortality rates. It notes that death rates are higher and increasing more rapidly among working-age adults with lower education and income levels. Additionally, the report also points out a worrying change: the gap in death rates between Black and white Americans, which was getting smaller, is now increasing again. This suggests ongoing problems with deep-seated racism and differences in income and social status[1].

Recent research on death trends shows a big shift towards studying different age groups more closely. People now widely recognize that age is a key factor in how likely someone is to die. This means it's crucial to tailor healthcare solutions for different ages. The 2023 Morbidity and Mortality Weekly Report takes a detailed look at U.S. death data for 2022. It shows that the overall death rate, adjusted for age differences, went down from the previous year. The main causes of death were heart disease, cancer, accidents, and COVID-19. The report points out that death rates are higher in men, older people, and Black individuals, stressing the need for public health actions that really address these groups' needs. This research is pivotal for understanding and addressing U.S. mortality trends[2].

The 2017 National Vital Statistics Report looked at how long people live and how this is changing. It focused on how things like age, the cause of death, and people's backgrounds affect life expectancy. The report found that overall, people were living slightly shorter lives due to increases in deaths from accidents, suicide, diabetes, Alzheimer's, and the flu. However, fewer deaths from cancer and heart disease helped balance this out a bit. These changes in how long people live were different for men and women, and also varied among white, Black, and Hispanic groups. This report helps us understand more about why and how life expectancy is changing, which is really important for current research[3].

This research offers an in-depth analysis of mortality in the United States from 2000 to 2019. It thoroughly examines not only the overall death rates but also analyzes the leading causes of death, delivering insights into the differences and consistencies over the years. The research includes a gender-focused analysis, comparing mortality causes between males and females, and it delves into age-specific mortality rates, offering detailed insights into the predominant causes of death across different life stages, from early childhood to the elderly.

This study undertakes an exploratory analysis to investigate key questions regarding mortality trends, specifically examining the impact of gender. It examines how mortality rates differ between genders over time and across different life stages. The aim is to identify patterns within the mortality data, providing a detailed understanding of these trends. Such insights are crucial for informing public health strategies and interventions, contributing to a more comprehensive view of mortality dynamics within the population.

The motivation for this study lies in its potential to uncover critical insights into demographic-specific health risks and trends. By analyzing mortality across different genders and ages and investigating the leading causes of death over an extensive period, the study seeks to contribute to a deeper, more nuanced understanding of health vulnerabilities within the population. These insights are vital for shaping healthcare priorities, addressing specific health concerns effectively, and ultimately enhancing community health outcomes.

This paper is divided into sections, section I of the study provides an introduction, emphasizing the importance of understanding mortality trends in the U.S., including the primary causes of death and their variation between demographics. Section II explores the methodology used detailing the data sources, analytical techniques, and visualization tools used. Section III engages in an in-depth discussion of the study’s findings, examining trends in mortality by gender and age. The concluding Section IV summarizes the study's findings and suggests avenues for future research, offering a comprehensive perspective on mortality trends and their broader implications. Section V presents all the coding scripts used for plotting the graphs.

II. METHODOLOGY

This study utilizes two main datasets [4], [5]. The first dataset was used to explore the primary causes of death in the United States over two decades, from 2000 to 2019[4]. This dataset is the cornerstone of the research, providing detailed insights throughout the project. The second dataset, focusing on the population across various age groups, has been sourced from the Kaiser Family Foundation (KFF)[5].

The ‘Death_Causes’ dataset for this study provides an extensive overview of mortality in the United States from 2000 to 2019[4]. It includes detailed categorizations of death causes, segmented by diverse age groups and gender, using columns such as 'location_id' and 'location_name' for geographical data, 'sex_id' and 'sex_name' for gender, and 'age_id' and 'age_name' for age groups. The dataset, sourced from the Global Burden of Disease, comprises additional columns like 'measure_id' and 'measure_name' for mortality measures, 'cause_id' and 'cause_name' for specific causes of death, and 'val', 'upper', and 'lower' for numerical data regarding death counts, offering a comprehensive framework for analyzing death patterns and trends.

The dataset for this study was imported into SQL Server using SQL Server Management Studio (SSMS)(Appendix A). A dedicated database was created to ensure structured data management. The SSMS Import and Export Wizard facilitated the migration of the dataset into the database. Initial SQL queries were used to understand the dataset's structure and the range of values in each column. Data type adjustments were made to certain columns for optimal data alignment. Additionally, the cleaning phase involved modifying 'Val', 'lower', and 'upper' columns to accurately represent null values.

The data transition from SSMS to Python was managed using SQL Alchemy and Pandas libraries (Appendix B). SQL Alchemy facilitated the connection to the SQL database, allowing for efficient data extraction. The retrieved data was then imported into Python as a Pandas Data Frame, leveraging Pandas' powerful data manipulation and visualization capabilities. This combination of SQL Alchemy for database interaction and Pandas for data handling and visualization enabled a seamless workflow, enhancing the analysis and visual representation of the study's findings.

The initial analysis for Fig.1 shows the overall trend of total annual deaths from 2000 to 2019 was charted. The data, grouped by year, showed total deaths annually without differentiating causes. A line graph, created using Python's matplotlib library, depicted these mortality trends. The y-axis was scaled to display death counts in millions for clearer visualization, while the x-axis represented the years, ensuring a chronological display of mortality over the two decades. This approach offered a comprehensive view of the overall mortality trends, foundational for further detailed analysis.

The analysis of death rates for the top 15 causes over time shown on Fig.2 was conducted using a rate-based approach. Focusing on 'Rate' as the metric, this method allowed for a standardized comparison, accounting for variations in population size and demographics. This standardization is crucial to accurately reflect the risk of death within a population, facilitating fair comparisons across different groups and time periods. After identifying the top causes through aggregation, the data was pivot-transformed for clarity. The resulting graph, with rates per 100,000 on the y-axis, provided a detailed and comprehensive visualization of trends in major causes of death, highlighting changes in public health risks over the years.

The analysis of the total deaths by gender illustrated in Fig. 3, examines gender-specific mortality differences. The dataset was segmented by gender using the pandas library, aggregating total death counts. A bar chart, developed with matplotlib, visually contrasts the mortality figures for females and males, utilizing distinct color schemes for each gender. This approach, coupled with bar labels denoting millions, offers a professional and clear depiction of gender disparities in mortality, enhancing the understanding of gender-related mortality trends.

Expanding on the gender-focused mortality overview in Fig.3, Fig. 4, and 5 further detail the distribution of deaths by specific causes, separately for males and females. These figures provide a nuanced look at how various causes of death differently impact each gender, offering a clearer understanding of gender-specific mortality trends.

The study further delves into the year-over-year mortality rates between genders as seen in Fig.6, utilizing a bar graph. This visualization highlights annual mortality distributions for males and females from 2000 to 2019. By grouping the data by year and gender and calculating total deaths, the Python-generated plot features two distinct, color-coded lines for each gender. This setup allows for an immediate and clear comparison of gender-specific mortality trends across the two-decade span.

In the analysis of the death rates by age group in Fig.7, original age groups were consolidated into broader categories to simplify the data and enhance interpretability. This was done using a mapping dictionary, where specific age ranges were grouped into four larger categories: 'Children', 'Young', 'Middle Age', and 'Elderly'. The grouping was as follows:

1. Children: This category includes several younger age groups. It includes infants and toddlers under 5 years old ('<5 years'), as well as older children aged 5-9 years and 10-14 years.

2. Young: This group includes teenagers aged 15-19 years and extends to include young adults in their early to late 20s, covering the age ranges of 20-24 years and 25-29 years, and going up to those in their early 30s (30-34 years).

3. Middle Age: This category represents the midlife period. It begins with individuals in their late 30s (35-39 years) and includes those in their early to late 40s, covering the 40-44 year and 45-49 years age groups.

4. Elderly: The final category includes the older age groups. It includes adults aged 50-69 years and extends to those aged 70-89 years, representing the senior and older elderly population.

This reclassification is applied using the map() function. The dataset is filtered for the 'Rate' metric, and death rates are aggregated by these new age groups and years using groupby() and sum(). For visualization, the data is pivoted using pivot(), setting years as the index and age groups as columns. The line chart is created with plt.figure() and plt. plot(), each line representing a different age group for easy comparison. The chart includes a title, axis labels, and a legend for clarity and accessibility.

The only age group that was taken out is the 80+ age group as some of it is already included in the 70-89 age group. The grouping of the different age groups simplifies the data into more understandable groups, enhancing the analysis of mortality trends across different life stages.

Moving forward for an age-focused analysis, Fig.8 includes a bar graph to investigate mortality across different age groups. This graph, created using Python's pandas and matplotlib libraries, categorizes death counts by age, shown in descending order to highlight the groups with the highest mortality. The graph effectively conveys this information, with total deaths in millions on the y-axis and distinct age groups on the x-axis, providing a clear visual representation of age-specific mortality trends.

Transitioning to the second dataset [5] for further analysis, this dataset stands out for its accuracy and comprehensive health and demographic data. It includes annual population figures by age group from 2008 to 2019, consolidated into one file for analysis. The population information is organized into several age brackets: from newborns to 18 years, young adults from 19 to 25 years, adults from 26 to 34 years, the middle-aged group from 35 to 54 years, older adults from 55 to 64 years, and seniors aged 65 and above.

To analyze the data, the study uses Python's Pandas library for managing the datasets and Matplotlib for creating visual representations. Initially, the population data is imported into a data frame, a fundamental step in preparing the data for further analysis. The process then involves the calculation of average population size for each age group by removing the 'Year' column, focusing solely on population figures. This process reveals the typical population distribution across the age groups for the period 2008-2019.

The data is visualized using a pie chart, referred to in Fig.9 in the study, which provides an integrated view of the average population distribution by age. This pie chart is designed to be circular, making the data easy to understand and interpret.

Ensuring the data's accuracy and representativeness is vital to the study's validity. The statistical approach is primarily based on calculating average values, simplifying the complex data into actionable information.

In addition to the population analysis, Table-I incorporates a combined analysis of the Death Causes and Population datasets. The study harmonizes the age groups from both datasets using a mapping strategy, age groups from the Death Causes dataset were mapped into the population dataset as follows:

- The age ranges of '5 years and under,' '5-9 years,' '10-14 years,' and '15-19 years' from the Death Causes dataset are grouped together as '0-18' in the Population dataset.
- The age range '20-24 years' from the Death Causes dataset is grouped as '19-25' in the Population dataset.
- The age ranges '25-29 years' and '30-34 years' from the Death Causes dataset are grouped together as '26-34' in the Population dataset.
- The age ranges '35-39 years,' '40-44 years,' and '45-49 years' from the Death Causes dataset are grouped together as '35-54' in the Population dataset.

- The age range '50-69 years' from the Death Causes dataset falls within the '55-64' category in the Population dataset.

People aged '70 years and over' are categorized as '65+After aligning the age groups, the study focuses on the years 2008-2019 to match the time frame of the population data. The data is then restructured using the Melt function in Pandas, facilitating a direct comparison between the population and death figures for each age group.

Finally, in Fig.10, 11, 12, and 13 the study examines the top 15 causes of death for each age group. The death rate is calculated per 100,000 people to ensure a uniform standard for mortality rates across different population sizes. This metric is essential for assessing the frequency of various causes of death in different demographic groups. By concentrating on the most significant causes of death, the research aims to inform public health initiatives and guide the efficient allocation of health resources.

III. Results

A. Analyzing the Total Number of Deaths Throughout the Period 2000-2019

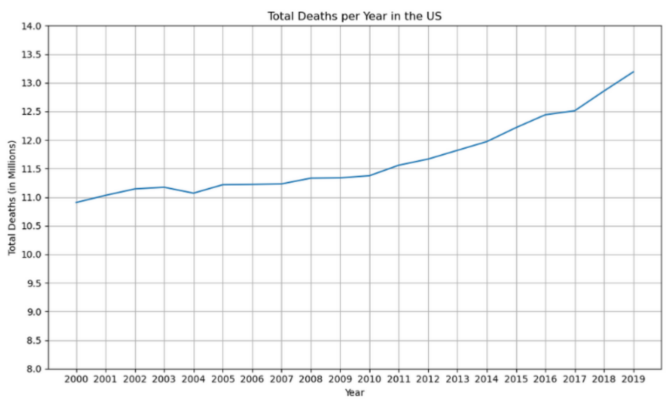


Figure 1 Total Number of Deaths 2000-2019

The above graph offers a comprehensive view of the total deaths in the United States spanning from the year 2000 to 2019. Represented on the y-axis in millions, the data paints a picture of an overarching upward trend in mortality rates. Starting in 2000, the deaths were approximately 10.9 million, and by the end of 2019, this number had risen to about 13.2 million, marking an increase of around 2.3 million deaths in two decades. While the overall trajectory showcases an increase, the journey is punctuated with minor fluctuations. For example, 2004 experienced a slight decrease in deaths, with the number dropping to 11.1 million compared to the previous year. However, this decrease was temporary, with the numbers reemerging in the years that followed. Particularly noticeable are the sharper increments between 2014 to 2016 and again from 2017 to 2018. The graph culminates in 2019, showcasing the highest recorded mortality with an estimated 13.2 million deaths.

The data indicates a notable increase in mortality beginning in 2017, a surge largely attributed to the severe flu season of 2017-2018 in the U.S., primarily driven by the H3N2 virus strain. The CDC's reported estimates highlight the intensity of this season, with significant case numbers and deaths. The crisis was exacerbated by factors such as medical supply shortages due to Hurricane María. The situation's severity is further highlighted by the potential underreporting of certain demographics, implying that the actual impact may surpass reported statistics[6].

This visual summary underscores the importance of understanding the various factors contributing to this rise, whether they are demographic shifts, healthcare challenges, or other societal and environmental influences.

B. Analyzing the Top 15 Causes of Deaths Throughout the Period 2000-2019

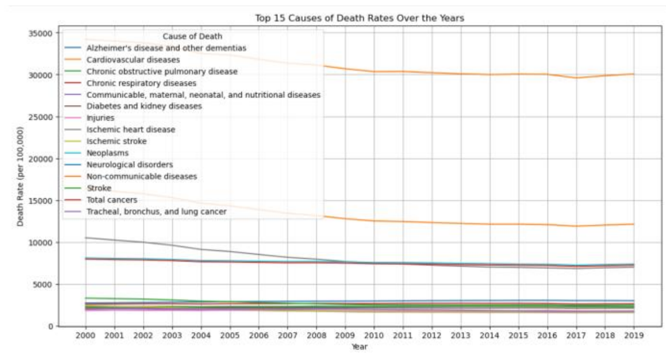


Figure 2 Top 15 Causes of Death Rates over the Years

The analysis then shifts focus to the top 15 causes of death, revealing a significant decline in cardiovascular disease rates. Similarly, notable decreases in non-communicable diseases, ischemic heart disease, and stroke mortality rates reflect enhanced treatment options and increased health awareness. Advances in cancer detection and treatment, alongside targeted public health campaigns, have also contributed to reduced cancer mortality rates. These trends indicate a shift in the public health landscape. While traditional diseases are being managed more effectively, there's an increasing number of neurological and chronic illnesses. This calls for a strategic shift in healthcare priorities to address these emerging health challenges.

High blood pressure is a major cause of death, significantly elevating the risk of heart disease and stroke. It affects nearly half of U.S. adults, leading to arterial damage and increased plaque buildup, which are primary factors in many cardiac and cerebral vascular incidents. This condition is often exacerbated by the American diet, where excessive sodium intake is common. Diabetes, another prevalent condition, not only doubles the risk of heart disease and stroke but also often coexists with high blood pressure, compounding the risks. These factors together underline the critical importance of blood pressure and diabetes management, along with dietary modifications, in reducing mortality from these leading causes[7].

The study 'Changes in Mortality in Top 10 Causes of Death from 2011 to 2018' provides pivotal insights into the shifting dynamics of public health over this period. A key finding is the substantial decrease in cancer mortality rates, a trend primarily driven by significant reductions in deaths from lung, colorectal, breast, and prostate cancers. However, the rate of decline varied among these cancers, with some showing slowed reductions [8].

In contrast to the progress in combating cancer, the study highlights a concerning trend in heart disease mortality. This trend is particularly pronounced in the older population, underscoring the influence of an aging demographic on health patterns[8].

Additionally, the study points out notable increases in mortality due to Alzheimer's disease and accidental deaths, including suicides. These rising figures bring to light emerging public health challenges that require attention [8].

C. Analyzing The Causes of Death by Gender (Female/Male)

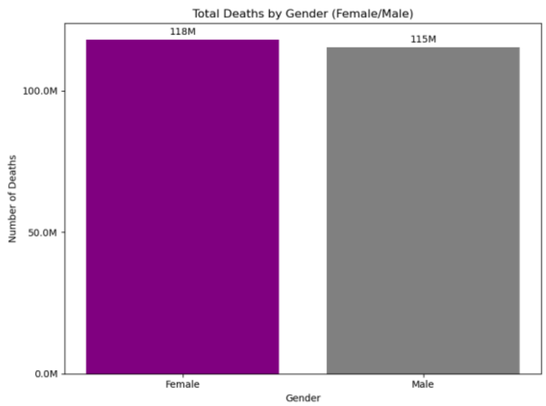


Figure 3. Total Deaths by Gender (Female/Male)

The bar graph illustrates the total mortality in the U.S. by gender, with the counts for females and males being nearly equal. Specifically, it shows females with around 118 million and males with about 114 million deaths, indicating similar overall mortality rates across genders. This suggests that gender-specific factors might not significantly differentiate mortality rates. Nonetheless, a deeper analysis including annual trends, causes, and age-specific data would be required for a more granular understanding of gender-related mortality patterns.

D. Analysis of The Total Distribution of Causes of Death for Males & Females

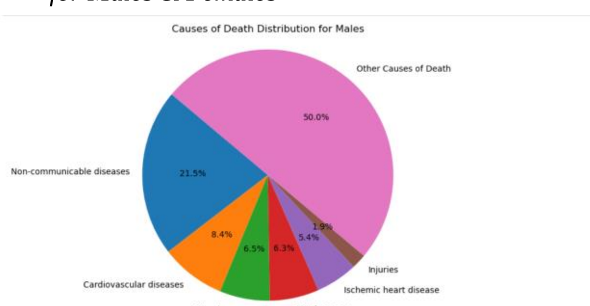


Figure 4. Causes of Death Distribution (Male)

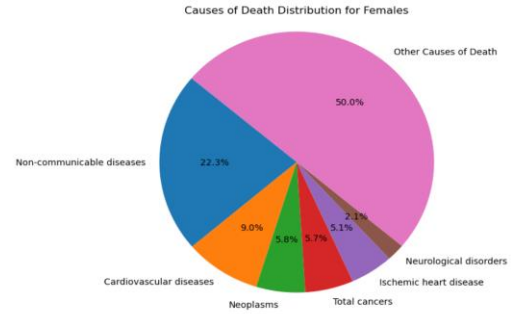


Figure 5. Causes of Death Distribution (Female)

These two charts represent the Distribution of the different causes of death between Males and females, Figure 3 provides an overview of overall mortality rates in males and females, without specifying death causes. In contrast, Figures 4 and 5 enhance this analysis by detailing how different causes of death are distributed between genders, offering a deeper insight into gender-specific mortality patterns.

In both charts, the same six causes dominate, but their proportions vary slightly between males and females. Non-communicable diseases are the most significant cause for both sexes, but they account for a marginally higher percentage of deaths in females (22.3%) compared to males (21%). This suggests that non-communicable diseases might have a slightly greater impact on female mortality.

Cardiovascular diseases, the second most common cause, also show a higher frequency in females (9%) than in males (8.4%). Neoplasms and total cancers, on the other hand, are more prevalent causes of death in males, with neoplasms accounting for 6.5% of male deaths versus 5.8% in females, and total cancers responsible for 6.3% of male deaths compared to 5.7% in females. This indicates a higher burden of these conditions in the male population.

Ischemic heart disease presents a relatively balanced impact, contributing to 5.1% of deaths in females and 5.4% in males, suggesting a similar level of risk for both genders. Lastly, neurological disorders, though less common overall, affect a slightly higher proportion of females (2.1%) than males (1.9%).

Overall, these charts not only underscore the commonality of certain health challenges across genders but also highlight the importance of considering gender differences in health risks and outcomes. The subtle variations in the percentages indicate that while males and females share many of the same health risks, the degree to which these risks manifest can vary, underscoring the need for gender-sensitive approaches in health care and disease prevention.

E. Analyzing The Causes of Death over the Years 2000-2019 for Males and Females

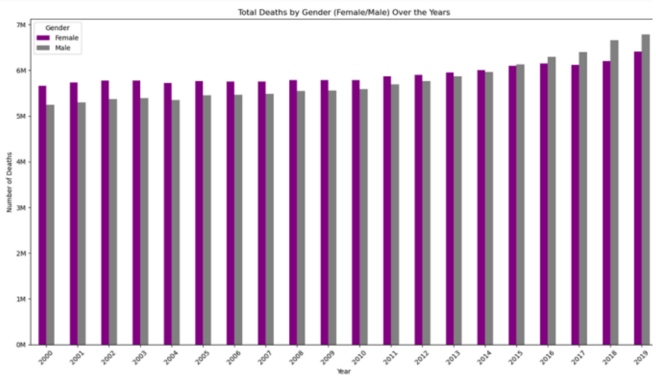


Figure 6. Total Deaths by Gender (Female/Male) over the Years

The above graph presents annual mortality rates for males and females, illustrating historical improvements in both genders' mortality, with a more pronounced decrease for women, leading to a widening life expectancy gap. The graph shows a higher male mortality rate for most years, but a significant abnormality occurred around 2016 when male deaths briefly exceeded female deaths. This deviation where women typically enjoy longer life expectancy, stimulates further investigation into the underlying causes. The chart suggests evolving factors may now be impacting gender-specific mortality rates.

in 1920, women's life expectancy surpassed that of men by almost two years. This gap widened significantly by the 1970s, with women living over seven years longer than men. However, by the turn of the millennium, in 2000, this advantage had slightly diminished to six years, signifying a life expectancy of 74 years for men and 80 years for women [9].

A key factor in the past improvement of women's survival rates was a big drop in deaths from childbirth. Better prenatal and obstetric care greatly reduced these deaths. Nowadays, women generally have lower death rates at all ages. In contrast, men are three times more likely to die from injuries, whether accidental, self-inflicted, or due to violence. Over the last fifty years, progress in reducing these types of deaths hasn't kept pace with others. Also, men tend to be less likely than women to seek medical help and follow medical advice, which affects these trends [9].

F. Death Rates by Age Group Over the Years 2000-2019 (per 100,000)

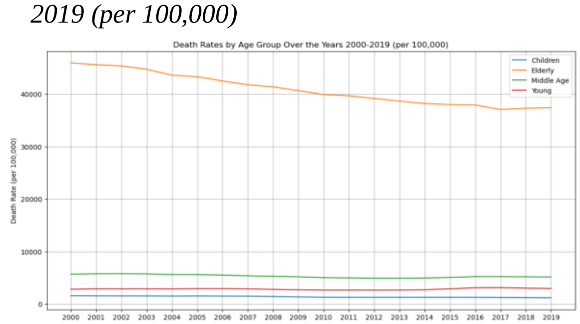


Figure 7. Death Rates by Age Group Over the Years 2000-2019(per 100,000)

The line chart in Fig 7 depicting death rates by age group from 2000 to 2019 reveals several key insights into mortality trends across different stages of life. By categorizing age groups as 'Children' (under 15 years), 'Young' (15-34 years), 'Middle Age' (35-49 years), and 'Elderly' (50-89 years), the chart offers a clear comparison of mortality impacts at various life stages.

Observing the chart, we can see how death rates have evolved over two decades. Each line, representing a specific age group, allows us to track the changes in death rates over time. For instance, lines with upward trends indicate an increase in death rates within those age groups, while downward trends suggest a decrease.

Such a comparative analysis is crucial for understanding public health dynamics. It highlights the age groups most affected by mortality over time and can signal shifts in health risks or the effectiveness of health interventions. This information is invaluable for public health officials and policymakers, aiding in the development of targeted health strategies and resource allocation to address the most pressing health challenges of different age groups.

G. Analyzing the Total Deaths per age group

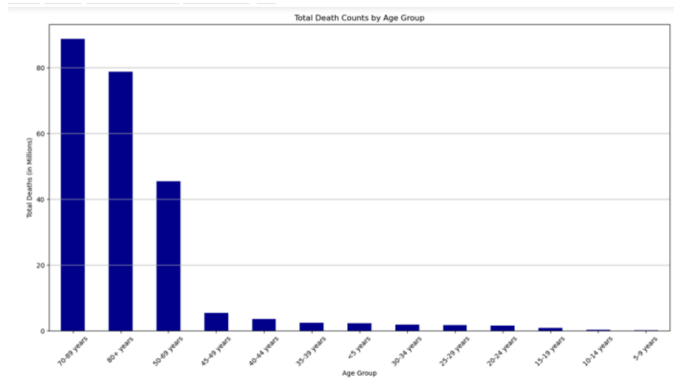


Figure 8. Total Deaths Counts by Age Group

The bar chart provides a clear representation of mortality rates across different age groups. Most prominently, the elderly age brackets of "70-89 years" and "80+ years" record the highest death counts, suggesting that these segments of the population are naturally more susceptible to health challenges leading to mortality. Following them, the "50-69 years" category also witnesses a considerable number of deaths, although not as high as the elderly groups.

Interestingly, the data reveals a spike in the mortality rate for the "<5 years" group, a trend not observed in subsequent younger age categories. However, it's comforting to note that the death counts for this group are still much lower than for older individuals. The age segments "5-9 years" and "10-14 years" experience the lowest mortality, highlighting these years as some of the healthiest phases of life.

The Stanford University study on life expectancy in the United States offers important insights into death patterns. Life expectancy, a crucial measure in health, is calculated by looking at death rates across different ages and genders. According to the study, people born in the U.S. in 2015 are expected to live about 79 years on average. A significant observation from the study is that over 20% of this group may not live to be 70. This matches what we see in the graph, where the number of deaths is particularly high among those aged 70-89 [10].This observation emphasizes the importance of health strategies and interventions tailored to specific age groups, especially those at higher risk.

The report from the National Academies reveals a concerning trend: since 2010, adults in the U.S. between the ages of 25 and 64 have experienced increasing mortality rates. This uptick in deaths, particularly among working-age individuals, is attributed to a range of causes, including drug overdoses, heart disease related to hypertension, and other health conditions that have been worsening since the 1990s. It also underscores the need for comprehensive action to address the multifaceted drivers of increased mortality among working-age adults in the U.S.[1].

H. Analyzing the Distribution of the Population Between the Different Age Groups

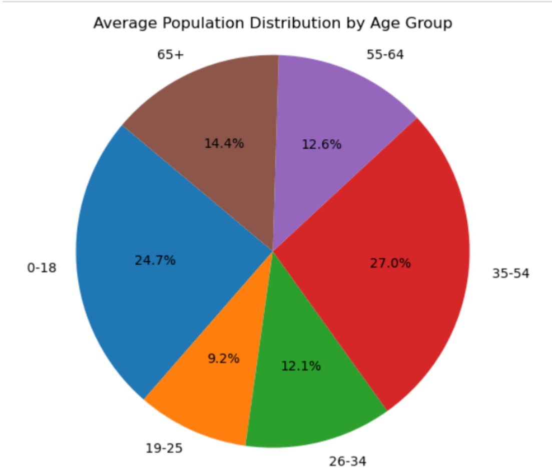


Figure 9. Average Population Distribution by Age Group

The pie chart in this analysis showcases the average population distribution across various age groups, highlighting significant demographic variances. The largest shares are held by the 0-18 and 35-54 age groups, indicating their dominant role in the demographic composition. Meanwhile, the 19-25, 55-64, and 65+ age groups, constituting 9%, 12.6%, and 14.4% of the population respectively, represent smaller yet vital segments. This detailed breakdown provides a nuanced understanding of the demographic landscape, reflecting the diversity and varying contributions of each age group to the overall population.

This visual representation gives us a clear look at current population trends. For instance, if younger age groups take up bigger parts of the chart, it might mean the average age of the population is getting younger. This suggests that the population is not only increasing but also becoming younger. On the other hand, if older age groups make up the bigger sections, it could indicate that the population is aging. This is key to understanding how society is changing.

Moreover, the methodology employed in the pie chart, namely the aggregation of average values over several years, offers a broad-scale view of the population distribution. This approach is instrumental in capturing overall trends, providing a streamlined narrative of demographic evolution. However, it's crucial to recognize that such an averaging technique might overlook the minor details, yet significant, year-to-year variations in the demographic structure. These shifts, although not immediately apparent in the chart, are pivotal in comprehending the dynamic and ever-evolving nature of population trends.

The research by Preston and Vierboom provides a detailed look at the age distribution in the United States between 2013 and 2018, revealing key insights that align with the findings from this research's analysis. Over these years, despite numerous demographic shifts—such as births, deaths, and migrations—the overall pattern of age distribution remained remarkably consistent. However, this pattern did not remain static; it demonstrated a discernible upward shift along the age axis by five years. This shift reflects the natural aging of the population over time.

Intriguingly, age groups that saw peak population numbers in 2013, specifically those aged 20–24 and 50–54, experienced a notable progression to higher age brackets by 2018. As a result, these peak populations were observed in the age groups of 25–29 and 55–59, respectively, by 2018. Similarly, there was a shift in the population trough: the age group 40–44, which represented a population low point in 2013, had moved to the 45–49 age bracket by 2018[11].

I. Age-Group Specific Mortality Rates and Their Relationship to Population Distribution (2008-2019)

TABLE I. RELATIONSHIP BETWEEN THE AVERAGE POPULATION OF AN AGE GROUP AND ITS CORRESPONDING AVERAGE DEATHS

	Age Group	**Average Deaths**	**Average Population**	**Mortality Rate (%)**
	Age Range	***Number of Deaths***	***Population Size***	***Percentage of Population by Deaths***
Demographic Analysis of Population and Mortality (2008-2019)	0-18	105	76,278,775	0.000138
	19-25	156	28,266,033	0.000552
	26-34	190	37,361,317	0.000509
	35-54	359	83,404,658	0.000430
	55-64	5,042	38,863,742	0.012974
	65+	9,345	44,347,417	0.021072

This analysis maps and aggregates death data across different age groups, comparing these figures with their respective average population sizes from 2008 to 2019, revealing several critical insights about mortality rates and their relation to population distribution.

Firstly, the analysis shows that the youngest age group, '0-18', experiences the lowest mortality rate, averaging 105 deaths. This figure represents a minuscule percentage (0.000138%) of the total population in this age group, indicating a significantly high survival rate during the early stages of life. As age increases, there is a noticeable uptick in mortality rates. The '19-25' and '26-34' age groups have higher average deaths, 156 and 190 respectively, yet these numbers still constitute a small fraction of their total populations. This gradual increase in mortality with age is a common demographic trend.

However, the data shows a substantial jump in mortality rates among older age groups. For instance, the '55-64' age group records an average of 5,042 deaths, which is about 0.012973% of its population. This increase becomes even more pronounced in the '65+' age group, with an average of 9,345 deaths, accounting for 0.021072% of its population. These findings align with the known trend of demographic aging, where older age groups are more susceptible to higher mortality rates, primarily due to age-related health issues.

The importance of this analysis lies in its ability to provide a clear and detailed picture of how mortality rates vary significantly across different age groups. It highlights the increasing risk of death with advancing age, information that is crucial for healthcare planning, resource allocation, and policymaking. This data is particularly vital for designing age-specific health interventions and support systems. Moreover, understanding these mortality trends is key in anticipating future healthcare needs and preparing for the challenges of an aging population.

J. Top 15 Causes of Mortality in Early Childhood (Age Group: <5 years, 5-9 years, 10-14 years)

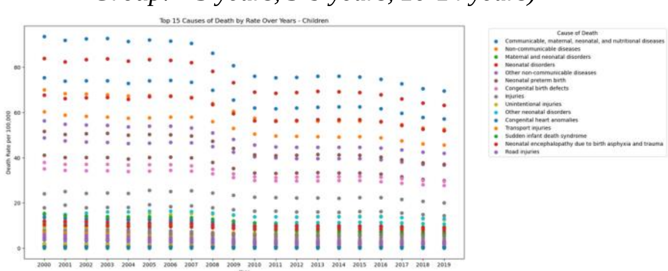


Figure 10. Top 15 Causes of Death -Children

In analyzing the top causes of death among children, as revealed by this scatter plot, several critical health issues emerge, painting a complex picture of the challenges faced in pediatric healthcare.

The category of "Communicable, Maternal, Neonatal, and Nutritional Diseases" stands out prominently. This broad grouping underscores the interconnected nature of health outcomes in children, where communicable diseases, often aggravated by nutritional deficiencies, pose significant risks. Moreover, the health of the mother during pregnancy is intrinsically linked to neonatal health, highlighting the importance of comprehensive maternal and child healthcare services. Following closely is the prevalence of "Non-Communicable Diseases" in young children. The presence of these diseases, which range from hereditary conditions to chronic illnesses that manifest early in life, emphasizes the need for early and effective pediatric care. This includes not only treating the diseases but also managing the long-term care requirements of children living with chronic conditions.

Another critical concern is the high incidence of "Maternal and Neonatal Disorders." Health complications during pregnancy, childbirth, and immediately after birth have a profound impact on both maternal and infant mortality and morbidity rates. This finding underlines the necessity of quality prenatal and postnatal care, ensuring that both mothers and newborns receive the support and medical attention they need during these crucial stages.

K. Top 15 Causes of Mortality in Young People (Age Group: 15-19 years, 20-24 years, 25-29 years, 30-34 years)

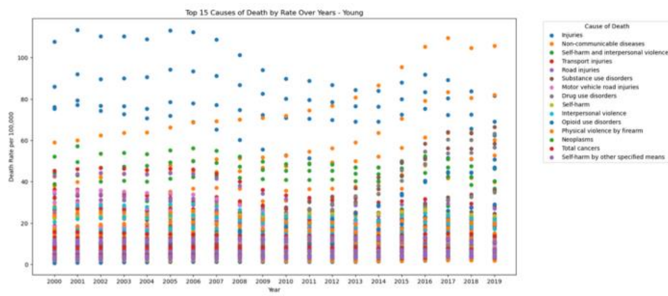


Figure 11. Top 15 Causes of Death -Young

In the Young age category, the top causes of death as revealed by the analysis highlight a distinct shift from primarily health-related issues to a combination of injuries, lifestyle-related disorders, and mental health concerns. Injuries emerge as a leading cause of death, indicating a significant risk factor in this age group. The prevalence of injuries, often resulting from accidents or risky behaviors, underscores the vulnerability of young adults to external, preventable factors. This finding calls for increased awareness and preventive measures, including safety education and the promotion of safer environments.

Non-communicable diseases continue to be a significant concern, reflecting a transition from infectious to chronic conditions as individuals age. This shift necessitates a focus on early detection, lifestyle modifications, and managing long-term health conditions, which can profoundly impact the quality of life. Self-harm and Interpersonal Violence are particularly concerning, highlighting the importance of mental health and social well-being. These causes point to underlying issues such as depression, anxiety, and the impact of social and familial environments. Addressing mental health, providing support systems, and creating awareness about mental well-being becomes crucial in this age group.

Transport and Road Injuries, including Motor Vehicle Road Injuries, represent a substantial risk, often linked to lifestyle and behavioral factors. These findings emphasize the need for road safety measures, responsible driving education, and stricter enforcement of traffic laws to prevent such incidents. The mention of substance use disorders, with a specific focus on

Drug Use and Opioid Use Disorders, sheds light on the growing challenge of substance abuse among young adults. This issue calls for comprehensive strategies encompassing education, rehabilitation, and support services, along with policies aimed at controlling the availability and use of harmful substances.

Physical Violence by Firearm and other forms of violence reflect societal and environmental factors influencing youth mortality. These issues require a multi-faceted approach involving law enforcement, community engagement, and conflict resolution education. Moreover, the presence of Neoplasms (cancers) in the list, while lower, indicates that health-related concerns like cancer are also relevant in this demographic.

The trends identified by the Population Reference Bureau (PRB), note that young adults are particularly vulnerable to injuries and violence, often tied to risky behaviors. These are largely preventable causes that account for a significant portion of deaths in this age group. PRB also emphasizes the importance of public health measures in reducing injury mortality and the need for broad-based responses involving many aspects of society to prevent deaths from violence, suicide, and substance abuse. This analysis illustrates the complex interplay of physical health, mental well-being, and social factors in shaping the health outcomes of young adults[12].

L. Top 15 Causes of Mortality in Middle-Aged (Age Group: 35-39 years, 40-44 years, 45-49 years)

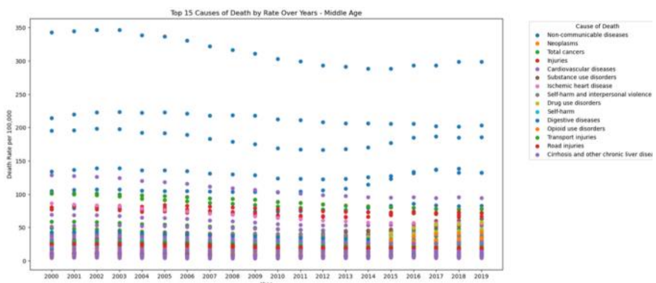


Figure 12. Top 15 Causes of Death -Middle Age

The analysis of the middle age group's leading causes of death presents several key findings. Most notably, cardiovascular diseases, including ischemic heart disease, emerge as top causes, underscoring the need for public health initiatives focusing on heart health and lifestyle changes. Cancers, grouped under 'Neoplasms' and 'Total cancers', are also prevalent, pointing to the impact of environmental, lifestyle, and genetic factors, and highlighting the need for diverse cancer prevention and treatment strategies.

Substance use disorders, including drug and opioid use disorders, are significant concerns, reflecting growing issues related to substance abuse. These disorders suggest the necessity for educational, rehabilitation, and policy interventions. The presence of various injury-related causes, such as 'Self-harm and interpersonal violence', 'Transport injuries', and 'Road injuries', highlights the importance of safety measures and mental health support to reduce preventable deaths.

Additionally, digestive diseases and chronic liver conditions like cirrhosis indicate the role of diet, alcohol consumption, and viral infections in health, calling for efforts to promote healthy eating, moderation in alcohol use, and vaccination programs. This analysis overall highlights the dominance of non-communicable diseases in the middle-aged population, with a significant focus on cardiovascular diseases and cancers, alongside the impact of substance abuse and injuries. This information is crucial for guiding public health policies and healthcare interventions aimed at reducing mortality and improving health outcomes in this demographic.

According to a report from the National Academies of Sciences, Engineering, and Medicine, there has been an increase in death rates among young and middle-aged adults (25-64 years old) in the U.S. since 2010. This rise in mortality rates is particularly noticeable among working-age Americans, whose risk of dying from certain conditions, such as drug overdoses or hypertensive heart disease, has been increasing since the 1990s[1].

The report highlights that this public health crisis has significant implications for families, employers, and the U.S. economy. The rising death rate is primarily due to drug overdoses, alcohol, suicides, and cardiometabolic conditions, which include diabetes and heart diseases caused by high blood pressure and other conditions. It also points out that the increase in premature death is widespread, affecting young adults, all racial and ethnic groups, and both rural and metropolitan areas. Death rates are higher and are increasing faster among working-age adults with less education and income.[1]

M. Top 15 Causes of Mortality in Elderly (Age Group: 50-69 years, 40-44 years, 70-89 years)

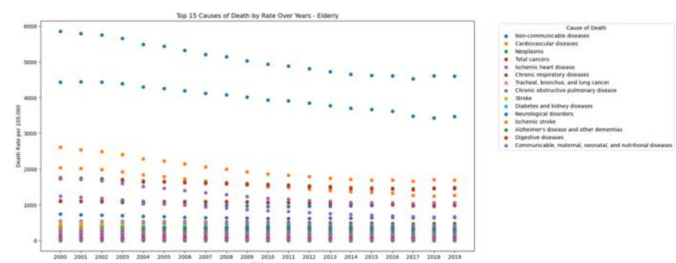


Figure 13. Top 15 Causes of Death -Elderly

In the elderly group, the analysis reveals a dominance of non-communicable diseases, with a notable and positive trend of decreasing death rates over the years. This decline, particularly in cardiovascular diseases, cancers, and neurological disorders, is a promising sign of advancements in healthcare and preventive measures.

Cardiovascular diseases, including ischemic heart disease, show a significant reduction in mortality. This improvement could be due to better risk factor management, advancements in medical treatments, and heightened awareness of heart health. The decline in deaths from cancers, encompassing 'Neoplasms', 'Total cancers', and specific types like tracheal, bronchus, and lung cancer, suggests progress in early detection and treatment, as well as possible lifestyle changes. Chronic respiratory diseases, including chronic obstructive pulmonary disease, also indicate a downward trend. In the realm of neurological disorders, the reduction in deaths from stroke, Alzheimer's disease, and other dementias is a positive development, reflecting advancements in stroke care and better management of neurological conditions.

The analysis also shows a decrease in deaths due to diabetes and kidney diseases, pointing to improved management and treatment of these conditions. Similarly, a decline in mortality from digestive diseases could be linked to healthier dietary habits and better healthcare access. However, the presence of communicable, maternal, neonatal, and nutritional diseases in the elderly underscores the ongoing need for effective prevention and treatment strategies for infectious diseases, especially considering the increased vulnerability of the elderly due to age-related changes in their immune systems.

Based on the information provided by Key Stone Health [13] it's evident that elderly individuals face a unique set of health challenges. Common conditions like arthritis, cancer, chronic kidney disease, COPD, dementia (including Alzheimer's and Parkinson's), diabetes, osteoporosis, and stroke is prevalent in this age group. Your analysis showing a decrease in death rates for major diseases like cardiovascular diseases, cancers, and neurological disorders aligns well with these insights. It reflects the advancements in geriatric medicine and healthcare, emphasizing the importance of continuous care and lifestyle management to mitigate these conditions' impact on the elderly population. This trend towards better management and treatment of age-related diseases is a positive development in geriatric healthcare[13].

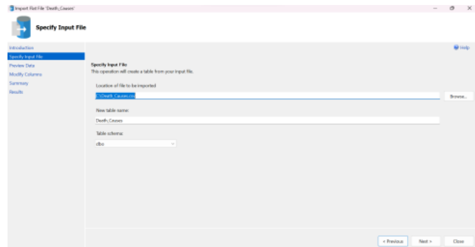
IV. DISCUSSION

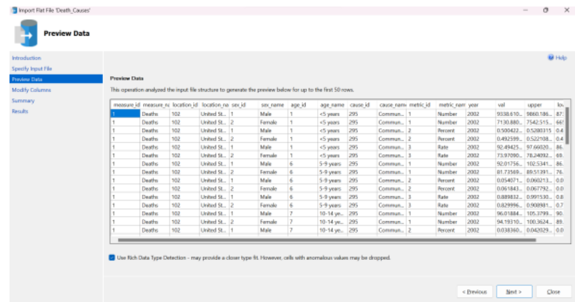
This study's comprehensive analysis of U.S. mortality trends from 2000 to 2019 uncovers critical insights. It highlights an overall rise in death rates, with pronounced increases in 2017 and gender-specific mortality shifts in 2016. The elderly, particularly those 70-89 years old, experienced the highest mortality rates, while the under-5 age group also showed notable vulnerability. A key finding is the dominance of non-communicable diseases as primary death causes in 2019, underscoring the need for public health strategies tailored to different demographics. These insights are vital for future healthcare planning and policy-making.

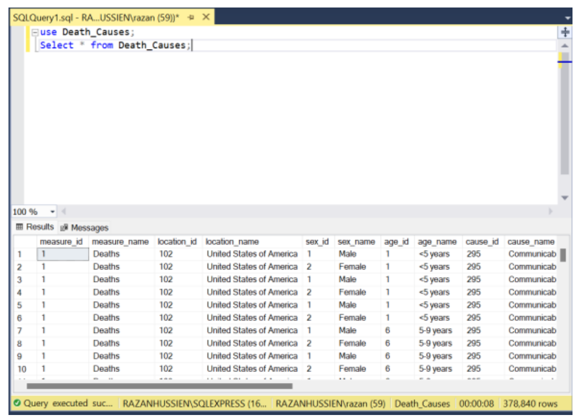
This research effectively identifies key mortality trends across gender and age groups in the U.S. However, it leaves room for further exploration into how socioeconomic status, race, and regional factors impact these trends. Future studies could benefit from a more detailed examination of these elements, potentially uncovering deeper insights into health disparities. Investigating the role of policy changes, healthcare access, and socio-economic shifts could also provide a more nuanced understanding of mortality patterns and guide targeted public health interventions.

APPENDIX

Appendix A: Demonstration of Loading Data Set into SSMS

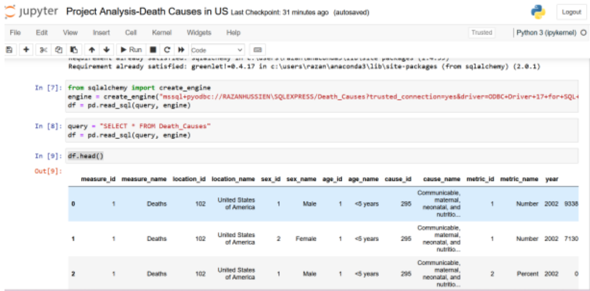






Appendix B: Demonstration of Loading Data Set into Jupyter Notebook

The study's transition from SQL to Python was streamlined using Jupyter Notebook and the SQL alchemy library. This approach enabled efficient data retrieval from SQL into Python, combining SQL's querying strength with Python's analytical versatility for effective data analysis.



References

[1] “Death Rates Rising Among Middle-Aged and Younger Americans - Report Recommends Urgent National Response | National Academies.” Accessed: Nov. 21, 2023. [Online]. Available: https://www.nationalacademies.org/news/2021/03/death-rates-rising-among-middle-aged-and-younger-americans-report-recommends-urgent-national-response

[2] F. B. Ahmad, J. A. Cisewski, J. Xu, and R. N. Anderson, “Provisional Mortality Data — United States, 2022,” *Morb. Mortal. Wkly. Rep.*, vol. 72, no. 18, pp. 488–492, May 2023, doi: 10.15585/mmwr.mm7218a3.

[3] “National Vital Statistics Reports Volume 68, Number 9 June 24, 2019 Deaths: Final Data for 2017”.

[4] “GBD Results,” Institute for Health Metrics and Evaluation. Accessed: Dec. 10, 2023. [Online]. Available: https://vizhub.healthdata.org/gbd-results

[5] “Population Distribution by Age | KFF.” Accessed: Nov. 21, 2023. [Online]. Available: https://www.kff.org/other/state-indicator/distribution-by-age/?dataView=1&currentTimeframe=2&selectedRows=%7B%22wrapups%22:%7B%22united-states%22:%7B%7D%7D%7D&sortModel=%7B%22colId%22:%22Location%22,%22sort%22:%22asc%22%7D

[6] “2017–2018 United States flu season,” *Wikipedia*. Oct. 19, 2023. Accessed: Dec. 09, 2023. [Online]. Available: https://en.wikipedia.org/w/index.php?title=2017%E2%80%932018_United_States_flu_season&oldid=1180926253

[7] “Heart Disease and Stroke | CDC.” Accessed: Dec. 09, 2023. [Online]. Available: https://www.cdc.gov/chronicdisease/resources/publications/factsheets/heart-disease-stroke.htm

[8] J. S. Rana, S. S. Khan, D. M. Lloyd-Jones, and S. Sidney, “Changes in Mortality in Top 10 Causes of Death from 2011 to 2018,” *J. Gen. Intern. Med.*, vol. 36, no. 8, pp. 2517–2518, Aug. 2021, doi: 10.1007/s11606-020-06070-z.

[9] “The Gender Gap in U.S. Mortality,” PRB. Accessed: Dec. 09, 2023. [Online]. Available: https://www.prb.org/resources/the-gender-gap-in-u-s-mortality/

[10] “Life Expectancy and Inequality in Life Expectancy in the United States | Stanford Institute for Economic Policy Research (SIEPR).” Accessed: Dec. 09, 2023. [Online]. Available: https://siepr.stanford.edu/publications/policy-brief/life-expectancy-and-inequality-life-expectancy-united-states

[11] S. H. Preston and Y. C. Vierboom, “The Changing Age Distribution of the United States,” *Popul. Dev. Rev.*, vol. 47, no. 2, pp. 527–539, 2021, doi: 10.1111/padr.12386.

[12] “Young U.S. Adults Vulnerable to Injuries and Violence,” PRB. Accessed: Nov. 21, 2023. [Online]. Available: https://www.prb.org/resources/young-u-s-adults-vulnerable-to-injuries-and-violence/

[13] “Geriatric Diseases: Age-Related Medical Conditions & Illnesses,” Keystone Health. Accessed: Nov. 21, 2023. [Online]. Available: https://keystone.health/geriatric-diseases

V. Code

This section includes the codes used for the analysis of different factors throughout the project; starting from a demonstration of importing the data up to graphs analysis (SQL & Python) up to the last graph plotted.

--Code used to load the data into Jupyter Notebook:

```
!pip install pyodbc
import pyodbc
conn = pyodbc.connect('Driver={ODBC Driver 17 for SQL Server};'

'Server=RAZANHUSSIEN\SQLEXPRESS;'

'Database=Death_Causes;'

'Trusted_Connection=yes;')
import pandas as pd

query = "SELECT * FROM Death_Causes"
df = pd.read_sql(query, conn)
!pip install sqlalchemy
from sqlalchemy import create_engine
engine = create_engine("mssql+pyodbc://RAZANHUSSIEN\SQLEXPRESS/Death_Causes?trusted_connection=yes&driver=ODBC+Driver+17+for+SQL+Server")
df = pd.read_sql(query, engine)
query = "SELECT * FROM Death_Causes"
df = pd.read_sql(query, engine)
```

```
df.head()
```

--Code used to plot Figure 1:

```
#Total deaths per year

plt.figure(figsize=(10, 6))
(yearly_deaths / 1_000_000).plot()
plt.title('Total Deaths per Year in the
US')
plt.xlabel('Year')
plt.ylabel('Total Deaths (in Millions)')
plt.yticks(np.arange(8,
max(yearly_deaths/1_000_000) + 1, 0.5))
--This line creates a series of y-axis
tick marks starting at 8 million deaths,
ending just above the maximum number of
deaths in the data set (converted to
millions), with each tick 0.5 million
deaths apart.
plt.xticks(yearly_deaths.index.astype(int)
)  # This line to make sure the years are
displayed as whole numbers
plt.grid(True)
plt.tight_layout()
plt.show()
```

--Code used to plot Figure 2:

```
#Overall analysis of the top 15 Causes of
Death
import matplotlib.pyplot as plt

df_rate = df[df['metric_name'] == 'Rate']

grouped_data=
df_rate.groupby(['cause_name',
'year'])['val'].sum().reset_index()

# Finding the top 5 causes of death by
summing across all years
top_causes = grouped_data.
groupby('cause_name')['val'].sum().nlarges
t(15).index

# Filter the grouped data for only the top
causes
top_causes_data =
grouped_data[grouped_data['cause_name'].is
in(top_causes)]

# Pivoting the data to get years as the
index and causes as columns
pivot_top_causes_data =
top_causes_data.pivot(index='year',
columns='cause_name', values='val')

plt.figure(figsize=(14,7))
for cause in
pivot_top_causes_data.columns:
plt.plot(pivot_top_causes_data.index.astyp
e(str), pivot_top_causes_data[cause],
label=cause)

plt.title('Top 15 Causes of Death Rates
Over the Years')
plt.xlabel('Year')
plt.ylabel('Death Rate (per 100,000)')
plt.legend(title='Cause of Death',
loc='upper left')
plt.grid(True)
plt.show()
```

--Code used to plot Figure 3:

```
import matplotlib.pyplot as plt
import numpy as np
deaths_by_gender =
df.groupby('sex_name')['val'].sum()
#Formatting Labels
formatted_labels =
[f"{value/1_000_000:.0f}M" for value in
deaths_by_gender.values]
fig, ax = plt.subplots(figsize=(8,6))
bars = ax.bar(deaths_by_gender.index,
deaths_by_gender.values, color=colors)
ax.bar_label(bars,
labels=formatted_labels,
label_type='edge', padding=3)

plt.title('Total Deaths by Gender
(Female/Male)')
plt.ylabel('Number of Deaths')
plt.xlabel('Gender')
plt.xticks(rotation=0)
plt.yticks(np.arange(0, max_val + 1,
50_000_000), [format_millions(x) for x in
np.arange(0, max_val + 1, 50_000_000)])
plt.tight_layout()
plt.show()
```

--Code used to plot Figure 4:

For Males:

```
import matplotlib.pyplot as plt


# Filter the data for males
male_data = df[df['sex_name'] == 'Male']
```

```
# Group by cause of death and sum across all years
grouped_male_data = male_data.groupby('cause_name')['val'].sum().reset_index()

# Get the top 5 causes of death
top_causes = grouped_male_data.nlargest(6, 'val')

# Sum the values of all other causes
other_causes_sum = grouped_male_data[grouped_male_data['cause_name'].isin(top_causes['cause_name'])]['val'].sum()

# Create a new row for 'Other Causes of Death'
other_causes_row = pd.DataFrame({
'cause_name': ['Other Causes of Death'],
'val': [other_causes_sum]
})

# Combine the top causes with the 'Other Causes' row
pie_data = pd.concat([top_causes, other_causes_row], ignore_index=True)

# Plot the pie chart
plt.figure(figsize=(10, 6))
plt.pie(pie_data['val'], labels=pie_data['cause_name'], autopct='%1.1f%%', startangle=140)
plt.title('Causes of Death Distribution for Males')
plt.axis('equal')  # to ensure that pie is drawn as a circle.
plt.show()
```

For Females:

```
#Females
import matplotlib.pyplot as plt


# Filter the data for females
female_data = df[df['sex_name'] == 'Female']


# Group by cause of death and sum across all years
grouped_female_data = female_data.groupby('cause_name')['val'].sum().reset_index()

# Get the top 5 causes of death
top_causes = grouped_female_data.nlargest(6, 'val')

# Sum the values of all other causes
other_causes_sum = grouped_female_data[grouped_female_data['cause_name'].isin(top_causes['cause_name'])]['val'].sum()

# Create a new row for 'Other Causes of Death'
other_causes_row = pd.DataFrame({
'cause_name': ['Other Causes of Death'],
'val': [other_causes_sum]
})

# Combine the top causes with the 'Other Causes' row
pie_data = pd.concat([top_causes, other_causes_row], ignore_index=True)

# Plot the pie chart
plt.figure(figsize=(10, 6))
plt.pie(pie_data['val'], labels=pie_data['cause_name'], autopct='%1.1f%%', startangle=140)
plt.title('Causes of Death Distribution for Females')
plt.axis('equal')  # Equal aspect ratio ensures that pie is drawn as a circle.
plt.show()
------------------------------------------
import matplotlib.pyplot as plt
import numpy as np
grouped_data =df.groupby(['year', 'sex_name'])['val'].sum().unstack()
colors = ['purple', 'grey']
fig, ax = plt.subplots(figsize=(14,8))
grouped_data.plot(kind='bar', ax=ax, color=colors)
#Formatting y-Label
yticks_values = np.arange(0, 7_000_001, 1_000_000)
yticks_labels = [f"{x//1_000_000}M" for x in yticks_values]

plt.title('Total Deaths by Gender (Female/Male) Over the Years')
plt.ylabel('Number of Deaths')
plt.xlabel('Year')
plt.xticks(rotation=45)
plt.yticks(yticks_values, yticks_labels)
plt.legend(title='Gender')
plt.tight_layout()
plt.show()
------------------------------------------
```

--Code used to plot Figure 6:

#Total Deaths of Females & males per Year

```
import matplotlib.pyplot as plt
import numpy as np
grouped_data =df.groupby(['year',
'sex_name'])['val'].sum().unstack()
colors = ['purple', 'grey']
fig, ax = plt.subplots(figsize=(14,8))
grouped_data.plot(kind='bar', ax=ax, color=colors)
#Formatting y-Label
yticks_values = np.arange(0, 7_000_001, 1_000_000)
yticks_labels = [f"{x//1_000_000}M" for x in yticks_values]

plt.title('Total Deaths by Gender (Female/Male) Over the
Years')
plt.ylabel('Number of Deaths')
plt.xlabel('Year')
plt.xticks(rotation=45)
plt.yticks(yticks_values, yticks_labels)
plt.legend(title='Gender')
plt.tight_layout()
plt.show()
```

--Code used to plot Figure 7:

```
#Regrouping age groups and analyzing over
the years
#Defining the revised age-groups:
import pandas as pd
import matplotlib.pyplot as plt

age_group_mapping = {
'<5 years': 'Children',
'5-9 years': 'Children',
'10-14 years':'Children',
'15-19 years': 'Young',
'20-24 years': 'Young',
'25-29 years': 'Young',
'30-34 years': 'Young',
'35-39 years': 'Middle Age',
'40-44 years': 'Middle Age',
'45-49 years': 'Middle Age',
'50-69 years': 'Elderly',
'70-89 years': 'Elderly'
}

# Filter the data for the "Rate" metric
df_rate = df[df['metric_name'] == 'Rate']

# Apply the age group mapping
df_rate['age_group'] =
df_rate['age_name'].map(age_group_mapping)

# Group by age_group and year
grouped_data_rate =
df_rate.groupby(['age_group',
'year'])['val'].sum().reset_index()

# Pivot the data for plotting
pivot_data_rate =
grouped_data_rate.pivot(index='year',
columns='age_group', values='val')

# Plotting
plt.figure(figsize=(14,7))
for column in pivot_data_rate.columns:
plt.plot(pivot_data_rate.index.astype(str)
, pivot_data_rate[column], label=column)

plt.title('Death Rates by Age Group Over
the Years 2000-2019 (per 100,000)')
plt.xlabel('Year')
plt.ylabel('Death Rate (per 100,000)')
plt.legend()
plt.grid(True)
plt.show()
```

--Code used to plot Figure 8:

```
#Total Deaths per Age Group
import matplotlib.pyplot as plt
import numpy as np
death_age_group=df.groupby('age_name')['va
l'].sum()
death_age_group=death_age_group.sort_value
s(ascending=False)


# Formatting the graph
plt.figure(figsize=(14, 8))

(death_age_group / 1e6).plot(kind='bar',
color='darkblue')

plt.title("Total Death Counts by Age
Group")
plt.xlabel("Age Group")
plt.ylabel("Total Deaths (in Millions)")
plt.xticks(rotation=45)
plt.grid(axis='y')
plt.tight_layout()

plt.show()
```

--Code used to plot Figure 9:

```
population_df = pd.read_csv("C:\different
ages population.csv")
population_data_types =
population_df.dtypes
population_data_types

# removing leading and trailing spaces
from column names
```

```
population_df.columns =
population_df.columns.str.strip()
print(population_df.columns)

# This is the code used to draw the pie
chart that shows the average distribution
of population
import pandas as pd
import matplotlib.pyplot as plt

# Load the CSV file
file_path = "C:\different ages
population.csv"
population_data = pd.read_csv(file_path)

# Calculating the mean population for each
age group
average_population =
population_data.drop('Year',
axis=1).mean()

# Creating a pie chart
plt.figure(figsize=(10, 6))
plt.pie(average_population,
labels=average_population.index,
autopct='%1.1f%%', startangle=140)
plt.title('Average Population Distribution
by Age Group')
plt.axis('equal')  # Equal aspect ratio
ensures that pie is drawn as a circle.
plt.show()
-------------------------------------------
```

--Code used to plot Table I:

```
#  mapping from original age groups(Death
data) to broader categories (Population
data)--df is for Death Causes file
age_group_mapping = {
'<5 years': '0-18',
'5-9 years': '0-18',
'10-14 years': '0-18',
'15-19 years': '0-18',
'20-24 years': '19-25',
'25-29 years': '26-34',
'30-34 years': '26-34',
'35-39 years': '35-54',
'40-44 years': '35-54',
'45-49 years': '35-54',
'50-69 years': '55-64',
'70-89 years': '65+',
'80+ years': '65+'
}
# Ensuring that I'm only using the years
from 2008 which is the start year in the
population data
#Adding .copy() to work with a copy of the
data and not a view to avoid potential
issues with inadvertent changes to the
original dataframe.
# Filtering out rows with missing 'val'
values to ensure entegrity of data
representation.

filtered_death_data =
filtered_death_data.dropna(subset=['val'])

filtered_death_data =
df[(df['metric_name'] == 'Number') &
(df['year'] >= 2008) & (df['year'] <=
2019)].copy()

filtered_death_data['new_age_group'] =
filtered_death_data['age_name'].map(age_gr
oup_mapping)

# Aggregating the death data by new age
group

avg_death_data =
filtered_death_data.groupby('new_age_group
')['val'].mean().reset_index()

avg_death_data.columns =
['age_group','avg_deaths']

# Using Melt function on the population
data so as to make it similar to the death
causes data

columns_to_convert = ['0-18', '19-25',
'26-34', '35-54', '55-64', '65+']
population_melted =
population_df.melt(id_vars=['Year'],
value_vars=columns_to_convert,
var_name='age_group',
value_name='population')

# Aggregating population data by age group
avg_population_data =
population_melted.groupby('age_group')['po
pulation'].mean().reset_index()

# Merge the average death and population
data
avg_data_merged = pd.merge(avg_death_data,
avg_population_data, on='age_group')

# Calculating the percentage of population
by death

avg_data_merged['percentage_population_by_
death'] = (avg_data_merged['avg_deaths'] /
avg_data_merged['population']) *100
```

```
# For formatting the numbers for better
readability
avg_data_merged['population'] =
avg_data_merged['population'].apply(lambda
x: "{:,.0f}".format(x))
avg_data_merged['avg_deaths'] =
avg_data_merged['avg_deaths'].apply(lambda
x: "{:,.0f}".format(x))
avg_data_merged['percentage_population_by_
death'] =
avg_data_merged['percentage_population_by_
death'].apply(lambda x:
"{:.6f}%".format(x))
print(avg_data_merged)
```

--Code used to plot Figure 10, 11, 12 and 13:

```
#yearly death rates (15 top causes) for
each grouped age-
group(elderly,young,middle age and
children)
import pandas as pd
import matplotlib.pyplot as plt

# age group mapping
age_group_mapping = {
    '<5 years': 'Children',
    '5-9 years': 'Children',
    '10-14 years':'Children',
    '15-19 years': 'Young',
    '20-24 years': 'Young',
    '25-29 years': 'Young',
    '30-34 years': 'Young',
    '35-39 years': 'Middle Age',
    '40-44 years': 'Middle Age',
    '45-49 years': 'Middle Age',
    '50-69 years': 'Elderly',
    '70-89 years': 'Elderly'
}
# Applying the mapping to the df
df['age_group'] =
df['age_name'].map(age_group_mapping)

# Filter the dataframe for the 'Rate'
metric
rates_data = df[df['metric_name'] ==
'Rate']



# Looping through each age group and
creating a scatter plot
for age_group in
rates_data['age_group'].dropna().unique():
# Exclude NaN values in 'age_group'
because i have values that are not
included in the mapping like 80+
    # Filter data for the specific age
group
    age_group_rates_data =
rates_data[rates_data['age_group'] ==
age_group]

    # Determine the top 15 causes for this
age group
    top_causes =
age_group_rates_data.groupby('cause_name')
['val'].sum().nlargest(15).index

    # Filter for only the top causes
    top_causes_data =
age_group_rates_data[age_group_rates_data[
'cause_name'].isin(top_causes)]
    top_causes_data =
top_causes_data.sort_values(by='year')

    # Generate a scatter plot for each
cause
    plt.figure(figsize=(15, 8))
    for cause in top_causes:
        cause_data =
top_causes_data[top_causes_data['cause_nam
e'] == cause]

plt.scatter(cause_data['year'].astype(str)
, cause_data['val'], label=cause)

    # Add title and labels
    plt.title(f'Top 15 Causes of Death by
Rate Over Years - {age_group}')
    plt.xlabel('Year')
    plt.ylabel('Death Rate per 100,000')

    # Add legend
    plt.legend(title='Cause of Death',
bbox_to_anchor=(1.05, 1), loc='upper
left')

    # Show the plot
    plt.show()
```