BỘ GIÁO DỤC VÀ ĐÀO TẠO
TRƯỜNG ĐẠI HỌC TÂN TẠO

DATA ANALYSIS REPORT

CAUSES OF DEATH IN THE UNITED STATES (2009–2019)

Sinh viên thực hiện: Giang Đình Trung
Mã số sinh viên: 2202080
Khoa: Khoa học máy tính
Giảng viên hướng dẫn: TS Nguyễn Xuân Hà

Table Of Contents

I. Introduction to the Topic

1.1 Reason for Choosing the Topic

In the field of public health, understanding the causes of death is a critical foundation for policymakers to develop effective disease prevention and control strategies. In the United States, mortality trends have fluctuated over time, reflecting both the advancements in modern medicine and emerging risks. Recognizing the importance of monitoring these trends, I decided to undertake a project to analyze and visualize the causes of death in the United States from 2009 to 2019, focusing on two main disease groups: infectious diseases and non-communicable diseases.

1.2 Research Objectives

Analysis of Mortality Causes from Infectious and Non-Communicable Diseases in the United States (2009–2019) aims to:

- Identify the leading causes of death over more than a decade.
- Track mortality trends by cause — whether they increase or decrease over time.
- Assess the proportion (percentage) of each cause in the total number of deaths per year.

1.3 Scope and Limitations of the Study

- The data focuses on the number of deaths in the United States from 2009 to 2019.
- Causes are divided into two groups: infectious and non-communicable diseases.
- The report does not delve into social, geographic, or environmental factors, focusing solely on mortality figures and trends over the years.

II. Theoretical Overview

2.1 Concept of Data Visualization

Data visualization is the process of transforming raw data—often numbers or tables—into charts, graphs, or images that are easy to understand. The primary goal is to enable viewers to quickly grasp key information, identify trends, anomalies, or correlations between factors.

2.2 Role in Healthcare Data Analysis

In healthcare, data is often vast and complex. When visualized effectively, dangerous trends become more apparent, aiding healthcare administrators, doctors, and communities in making informed decisions.

2.3 Introduction to the Dataset Used

The dataset used in this report is the CSV file "Cause_of_death_2009_2019.csv," which records the number of deaths by cause in the United States from 2009 to 2019. It includes 16 causes, divided into two main groups for separate analysis.

III. Methodology and Tools

3.1 Data Preprocessing Steps

- Standardize column names: retain the primary cause names.
- Filter data by country and time period.
- Group causes into non-communicable and infectious diseases.

3.2 Tools Used

- Google Colab
- Python (pandas, matplotlib, seaborn, plotly)

3.3 Types of Charts Applied

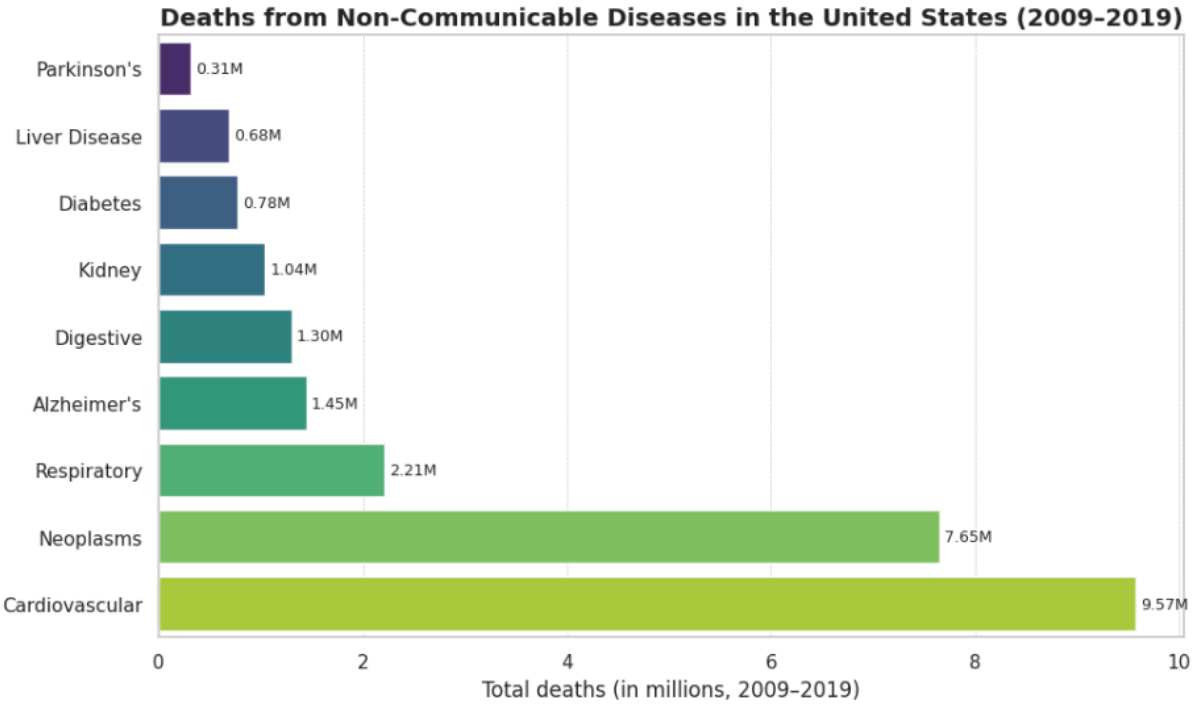
- Horizontal bar chart
- Stacked bar chart
- Heatmap

IV. Data Analysis

4.1. Non-Communicable Diseases

In the non-communicable disease group, I focused on prevalent causes such as cardiovascular diseases, cancer, diabetes, Alzheimer's, Parkinson's, liver diseases, kidney diseases, and chronic respiratory diseases. These are non-contagious diseases, primarily related to lifestyle, age, and genetic factors.

Chart 1: Total Deaths Due to Non-Communicable Diseases in the United States (2009–2019)



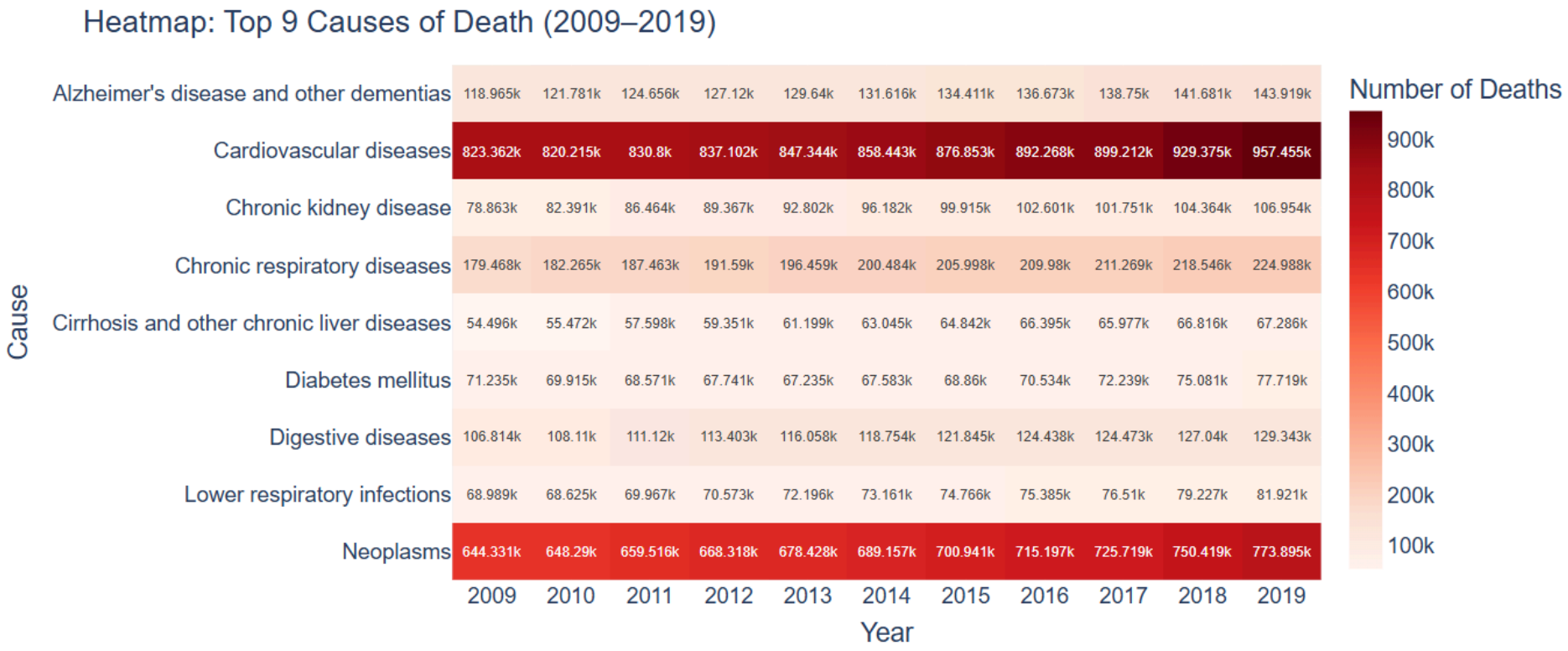
The horizontal bar chart illustrates the total number of deaths caused by non-communicable diseases from 2009 to 2019. Data is converted to millions for clarity, with causes arranged in ascending order to highlight the most dangerous diseases.

Observations from the Chart:

- Cardiovascular diseases lead with over 9 million deaths, confirming their status as the top cause of death in the U.S. during this period.
- Cancer (Neoplasms) follows with approximately 7 million deaths, ranking second.
- Chronic diseases such as respiratory diseases, Alzheimer's, diabetes, liver, kidney, and Parkinson's diseases each caused hundreds of thousands to millions of deaths.
- Parkinson's had the lowest number of deaths in this group, yet still exceeded 0.2 million—a significant figure.

Conclusion:
The chart demonstrates that non-communicable diseases remain the primary cause of death in the United States over the 11-year period. Cardiovascular diseases and cancer dominate, far surpassing other chronic diseases. This highlights ongoing challenges related to non-communicable health issues—such as diet, lifestyle, and long-term healthcare—within the U.S. healthcare system.

Chart 2: Heatmap – 9 Leading Causes of Death in the United States (2009–2019)



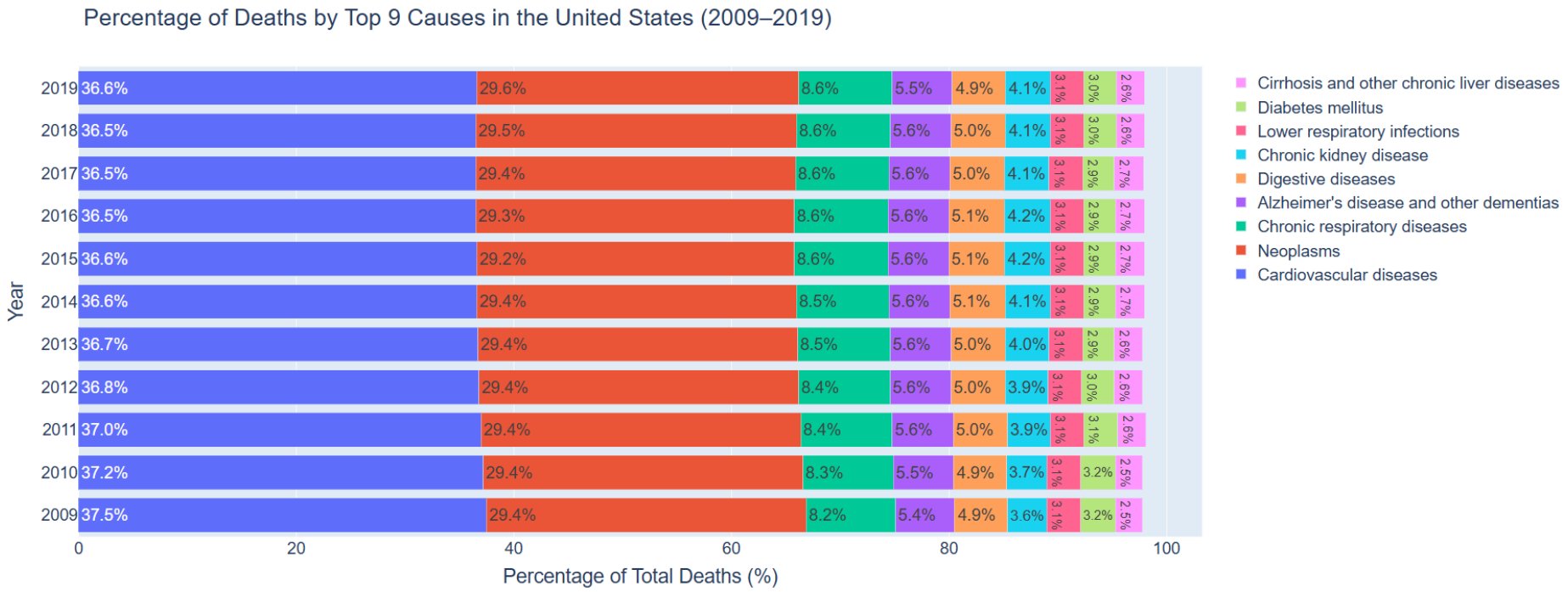
The heatmap displays the number of deaths by cause and year for the 9 leading causes from 2009 to 2019. Darker colors indicate higher death counts, providing a clear way to observe overall trends and changes over time.

Observations from the Chart:

- Cardiovascular diseases and cancer stand out with consistently dark red cells throughout the period, indicating persistently high death counts.
- Chronic respiratory diseases, Alzheimer's, and diabetes maintain significant death tolls, though with slightly lighter colors, indicating a smaller scale compared to the top two causes.
- Diseases like Parkinson's and liver diseases show lighter colors, reflecting lower death counts, but remain consistent each year.
- A slight darkening in Alzheimer's cells in later years suggests a gradual increase over time.

Conclusion:
Non-communicable diseases dominate mortality in the U.S., with cardiovascular diseases and cancer consistently leading. The heatmap also reveals emerging concerns, such as the rising trend in Alzheimer's, which warrants attention in the future.

Chart 3: Proportion of Deaths by 9 Leading Causes in the United States (2009–2019)



The stacked horizontal bar chart shows the percentage contribution of the 9 most common causes to total deaths each year from 2009 to 2019. Unlike previous charts focusing on absolute numbers, this chart highlights the relative contribution of each cause.

Observations from the Chart:

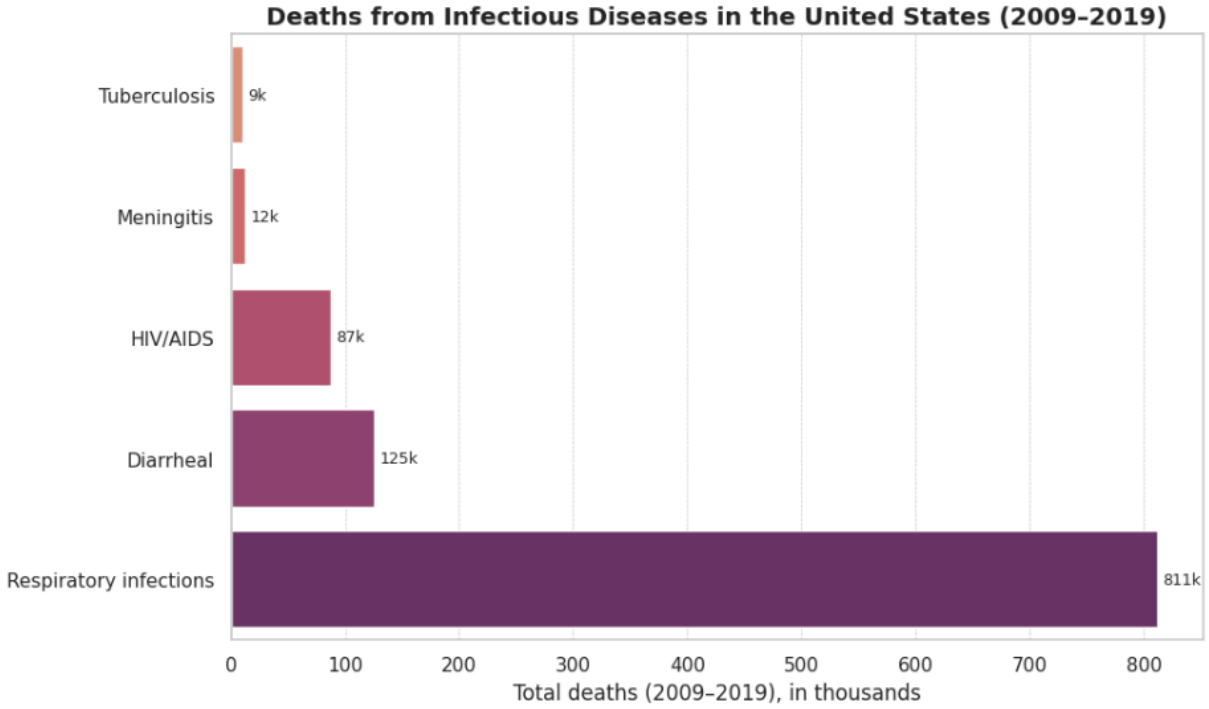
- Cardiovascular diseases consistently account for the highest proportion, hovering around 36%–37% of total deaths.
- Cancer is the second leading cause, contributing nearly 30% and remaining stable.
- Other causes, such as respiratory diseases, diabetes, Alzheimer's, kidney diseases, and Parkinson's, each contribute between 2% and 8%.
- Alzheimer's shows a slight increase in proportion in recent years.

Conclusion:
While the total number of deaths may vary yearly, the proportional structure of causes remains stable. This indicates that non-communicable diseases—particularly cardiovascular diseases and cancer—play a dominant role in the U.S. public health landscape, regardless of annual fluctuations.

4.2. Infectious Diseases

In the infectious disease group, I focused on prevalent causes such as tuberculosis, HIV/AIDS, pneumonia, diarrheal diseases, and meningitis. These are contagious diseases, often linked to hygiene, living conditions, immune status, and access to healthcare.

Chart 4: Total Deaths Due to Infectious Diseases in the United States (2009–2019)



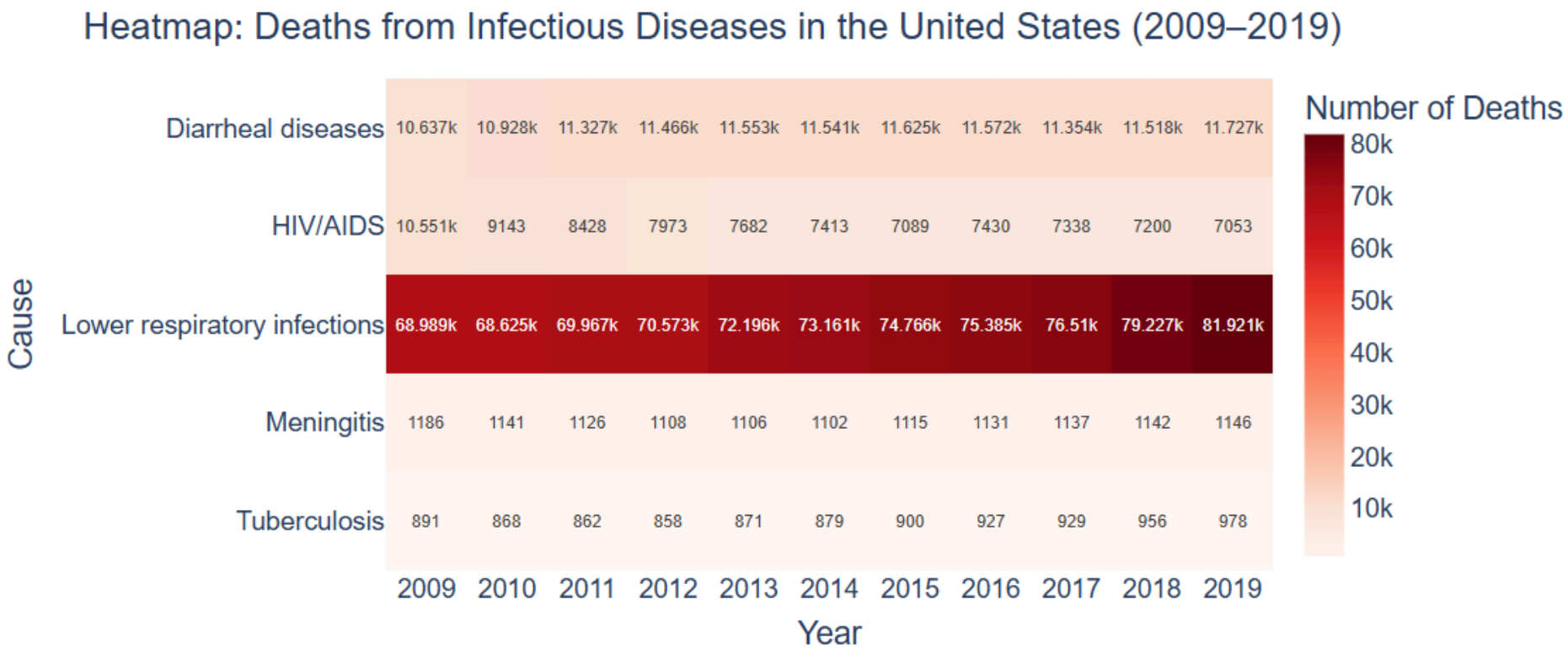
The chart shows the total number of deaths caused by common infectious diseases over 11 years, measured in thousands. It focuses on five main causes: tuberculosis, HIV/AIDS, pneumonia, diarrheal diseases, and meningitis.

Observations from the Chart:

- Pneumonia (respiratory infections) is the leading cause, with over 800,000 deaths from 2009 to 2019.
- Diarrheal diseases and HIV/AIDS follow, with 125,000 and 87,000 deaths, respectively.
- Meningitis and tuberculosis have lower totals, each under 15,000 deaths.

Conclusion:
Compared to non-communicable diseases, infectious diseases cause significantly fewer deaths. However, diseases like pneumonia and HIV/AIDS remain significant within this group and require close monitoring. These contagious diseases can have severe consequences if not controlled through vaccines, treatment, and public education.

Chart 5: Heatmap – Deaths Due to Infectious Diseases in the United States (2009–2019)



The heatmap illustrates the number of deaths each year from five common infectious diseases: tuberculosis, HIV/AIDS, pneumonia, diarrheal diseases, and meningitis.

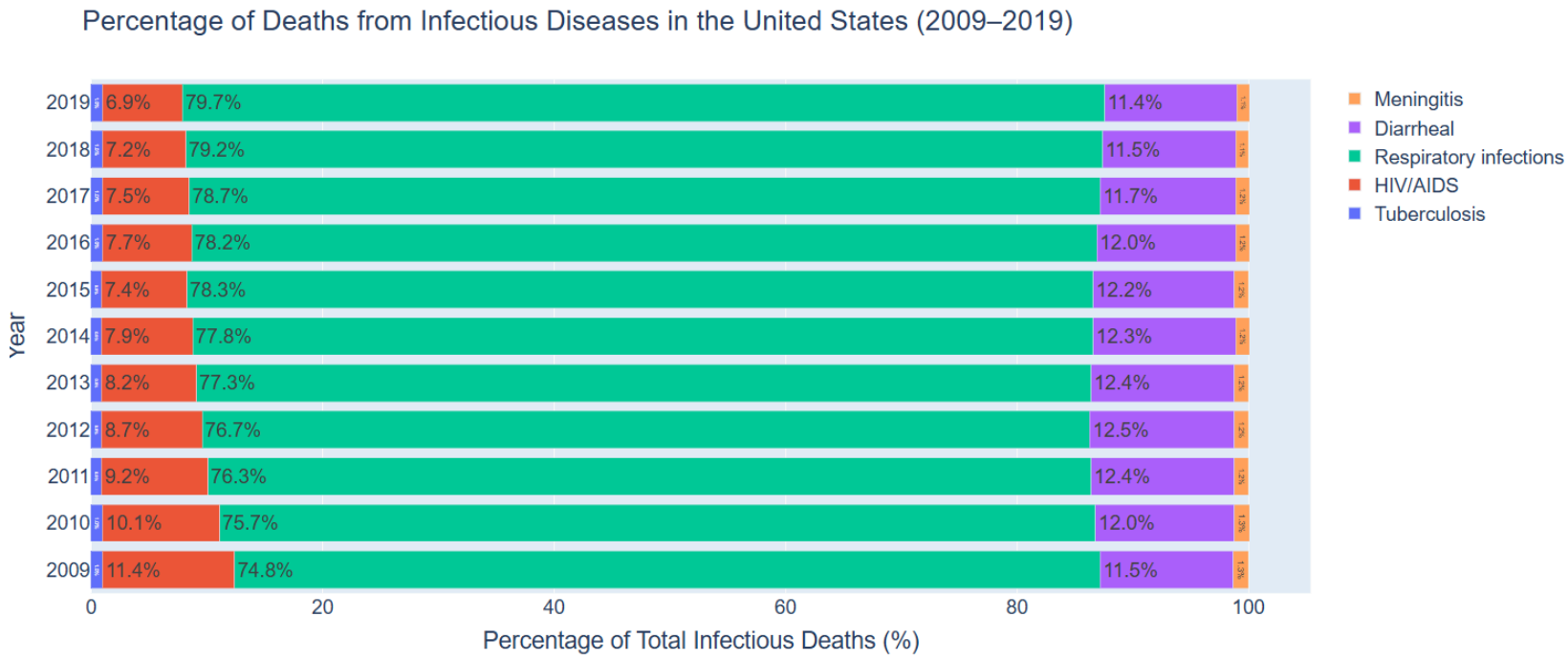
Observations from the Chart:

- Pneumonia consistently shows the darkest cells across all years, indicating a significantly higher death toll compared to other diseases in the group.
- HIV/AIDS cells fade to lighter shades over time, reflecting a positive trend of decreasing deaths.
- Diarrheal diseases, meningitis, and tuberculosis have lighter colors, indicating lower and stable death counts.

- No years show a sudden spike, suggesting stable and improving conditions.

Conclusion:
The heatmap clearly identifies pneumonia as the most concerning infectious disease, while others remain stable or decline. This reflects progress in controlling infectious diseases in the U.S., particularly through preventive healthcare, vaccination, and timely treatment.

Chart 6: Proportion of Deaths by Infectious Diseases in the United States (2009–2019)



The stacked horizontal bar chart shows the percentage of deaths attributed to each infectious disease out of the total infectious disease deaths each year. It includes the five main causes, calculated as percentages rather than absolute numbers.

Observations from the Chart:

- Pneumonia dominates, accounting for nearly 80% of infectious disease deaths each year.
- HIV/AIDS and tuberculosis each contribute around 6%–12%.
- Diarrheal diseases and meningitis have minimal shares, around 1% each year.

Conclusion:
The proportion of infectious disease deaths is heavily skewed toward pneumonia, underscoring the need for prioritized control measures. The declining proportion of HIV/AIDS deaths reflects significant progress in prevention and treatment. Other diseases contribute minimally but require ongoing monitoring to prevent outbreaks.

V. Conclusions and Recommendations

Through the analysis of mortality data in the United States from 2009 to 2019, it is evident that public health is significantly influenced by two major disease groups: non-communicable and infectious diseases. Each group has distinct characteristics, but both include prominent causes that require attention in national healthcare strategies.

5.1 Conclusions

1. **Non-Communicable Diseases:**
    - This group accounts for the vast majority of deaths, dominating both in absolute numbers and proportions each year.
    - Cardiovascular diseases and cancer are the leading causes, each responsible for millions of deaths over 11 years, comprising approximately 50–60% of annual deaths.
    - Other chronic diseases, such as diabetes, Alzheimer's, and chronic respiratory diseases, also contribute significantly, with some showing rising trends.
    - The proportional structure of deaths remains stable, indicating the need for long-term interventions.
2. **Infectious Diseases:**

- Although less deadly than non-communicable diseases, certain infectious diseases, particularly pneumonia, cause significant mortality.
- HIV/AIDS shows a declining trend, reflecting effective prevention and treatment efforts.
- Diseases like diarrheal diseases, meningitis, and tuberculosis maintain stable, low death rates but cannot be overlooked.
- The data suggests good control of infectious diseases, but vigilance is necessary, especially amid climate change, migration, and global pandemics.

5.2 Recommendations

- **Enhance Prevention and Lifestyle Changes:**
  Non-communicable diseases like cardiovascular diseases, cancer, and diabetes are closely tied to diet, physical activity, and lifestyle. Public awareness campaigns should encourage proactive health protection.
- **Invest in Preventive and Long-Term Healthcare:**
  Treating chronic diseases is often prolonged and costly. Healthcare systems should prioritize early detection, regular screenings, and long-term treatment support, especially for the elderly, who are at higher risk.
- **Maintain and Improve Infectious Disease Control:**
  Despite reduced mortality, infectious diseases require sustained efforts. Vaccination programs, clean water access, early warning systems, and rapid response mechanisms must be maintained to address emerging threats.
- **Leverage Data Visualization in Healthcare:**
  Visualization not only helps researchers quickly identify trends but also makes health information more accessible and understandable to the public.