Causes of Death Around the World (1990-2019)

Table of Content:

1. Project Description

Death is an unavoidable phase of everyone's life, understanding and analyzing the causes of death is essential for enhancing public health outcomes. Every year, millions of people die around the world due to various reasons including diseases, accidents, and natural disasters.

1.1. Project Objective

In this project our focus is to analyze and visualize the dataset on causes of deaths around the world from 1990 to 2019. Our aim is to extract some important insights about the leading causes of deaths over the past three decades and identify any significant changes in the mortality patterns.

1.2. Dataset Description

The dataset on "Causes of Deaths around the World" has been taken from the 'Kaggle' website. The dataset has a total of 34 columns and 6121 rows, including the headers which were later converted into a pivot with 8 rows using tableau prep for the convenience of visualization. It contains the data of 204 countries from the year 1990 to the year 2019. It has 31 different causes of deaths including diseases, accidents, and natural disasters.

The dataset used for this project is collected from

https://www.kaggle.com/datasets/iamsouravbanerjee/cause-of-deaths-around-the-world .

1.3. Metadata of dataset

Sr. No.	Field Name	Data type	Description
1	Category	String	Category of the cause of death (non-communicable diseases, communicable diseases or accidents and others)

2	No of death	Number	Number of deaths for the respective cause of death and country
3	Causes of death	String	Causes of death
4	Total death	Number	Total no. of deaths for the country for the given year
5	Decade	String	Includes the decade during which the deaths happened (1990-1999, 2000-2009, 2010-2019)
6	Country/Territory	String	Name of the country
7	Year	Date	The year when the deaths happened
8	Percentage	Number	This is a calculated field that calculates the percentage of the cause of death over the total deaths over all the decades

2. Data Cleaning and Preparation

Causes of Death (1990-2019) is an opensource dataset taken from Kaggle. This had the data about various causes of death across the world for a period of 30 years. The data was downloaded as a csv file. Microsoft excel and Tableau Prep tool have been used to clean our data for analysis. Following are the changes that we have done during data cleaning process:

1. Microsoft Excel was used to remove the column called “Country code” and to filter data only from the European Union and then imported to Tableau Prep as a .csv file.
2. In Tableau Prep, Changed the datatype of “Year” Column from String to Number (Whole).
3. A new calculated field was created by grouping the ‘Year’ column as decades. This calculated field named as ‘Decade’.
4. A new calculated field was created by adding the number of deaths due to all the causes and named as “Total Death”.
5. A pivot table was then created with the cause of death column showing the total deaths, country, year, and decade.

6. A new ‘Category’ calculated field was created in the pivot table. This column has 3 values, ‘Communicable Diseases’, ‘Non-Communicable Diseases’, and ‘Accidents and Others’. This column divides the causes of deaths into the above-mentioned categories.
7. This pivot table was exported as a hyper file, which was then imported to Tableau for further analysis and visualization.

3. Data Visualization Questions

The dataset included the causes of death that occurred during the span of 30 years (1990-2019) in the European Union.

Our objective was to understand the trend of the deaths, the category-wise comparison of the same over the decades, to understand if there were any significant spikes, if any and to identify the countries in EU that contributed to most deaths.

By using Tableau Desktop, we have visualized and analyzed the following questions on the given dataset.

1. What was the trend of deaths from 1990-2019 in the EU?
2. What was the percentage change for the various categories of death over the decades?
3. What were the major causes of death for the highest contributing category?
4. Which countries had the highest number of deaths for the major causes?

4. Design and Visualization

Tableau Desktop provided various charts that could be used to visualize and analyze each of our questions, among them we selected most suitable ones for our objectives.

The Charts that we used for our visualization are the following:

1. Line Graph: This chart is most suitable to answer our first question about the trend of deaths over a period.
2. Grouped bar Graph: It is suitable to find the percentage change in different categories.
3. Bullet Chart: It is a perfect visualization to visualize the major contributor among causes of deaths in a particular category.
4. Geographical Map: This is a perfect fit to find answer to our last question that is, “Which countries had the highest number of deaths for the major causes?”.

In addition, we added a Donut chart to represent the contribution of different categories to the total number of deaths in the EU in the time period of 30 years from 1990 to 2019.

Question 1

1. **What was the trend of deaths from 1990-2019 in the EU?**

A Line graph has been selected for the visualization of this question. Here, horizontal axis represents different years from 1990 to 2019 and the vertical axis represents total number of deaths in EU in the year.

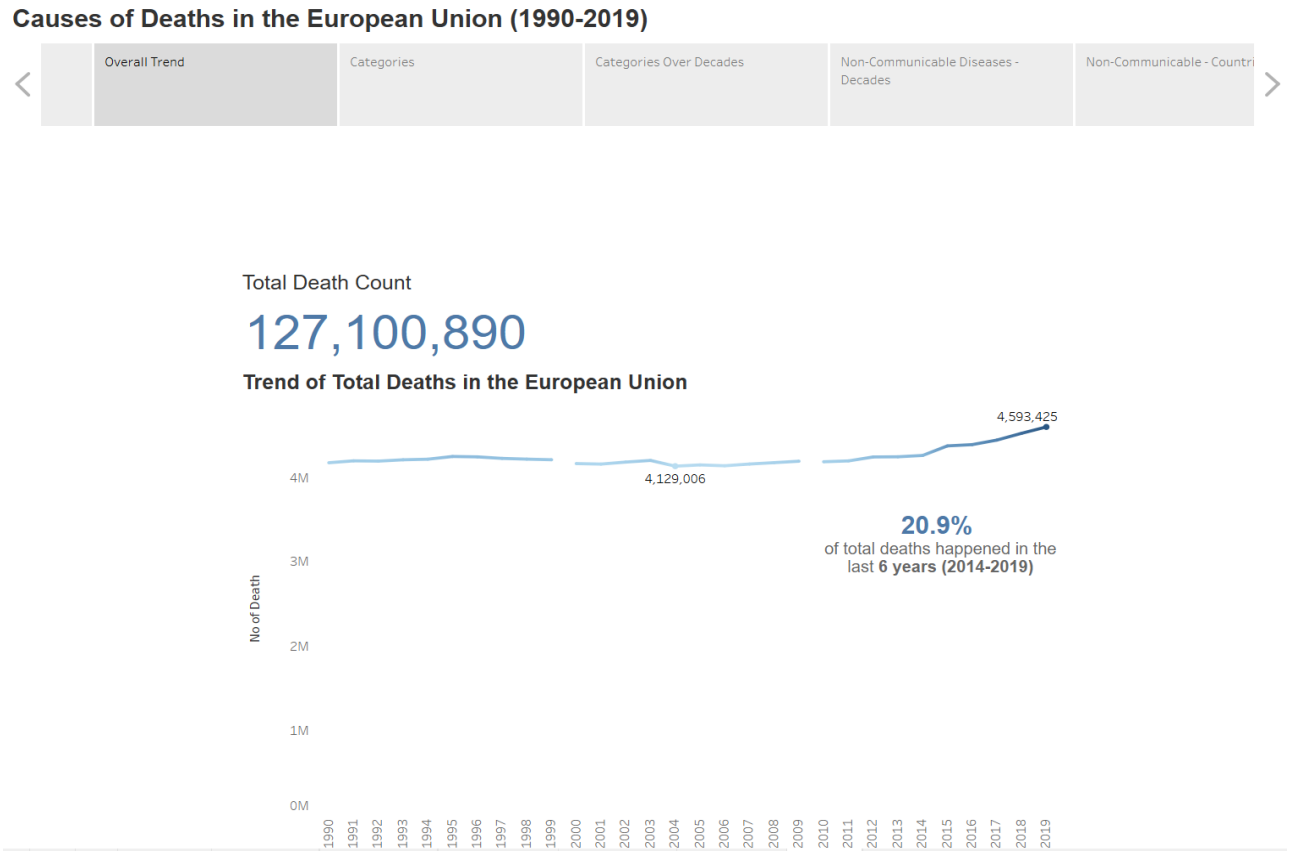

Figure 1

Based on the data, there were 120,100,890 deaths in the 30 years in the EU, and 20.9% of deaths happened in the last decade.

Question 2

2. **What was the percentage change for the various categories of death over the decades?**

A Donut chart has been used to understand the contribution of different categories in the total number of deaths in EU from 1990 to 2019.

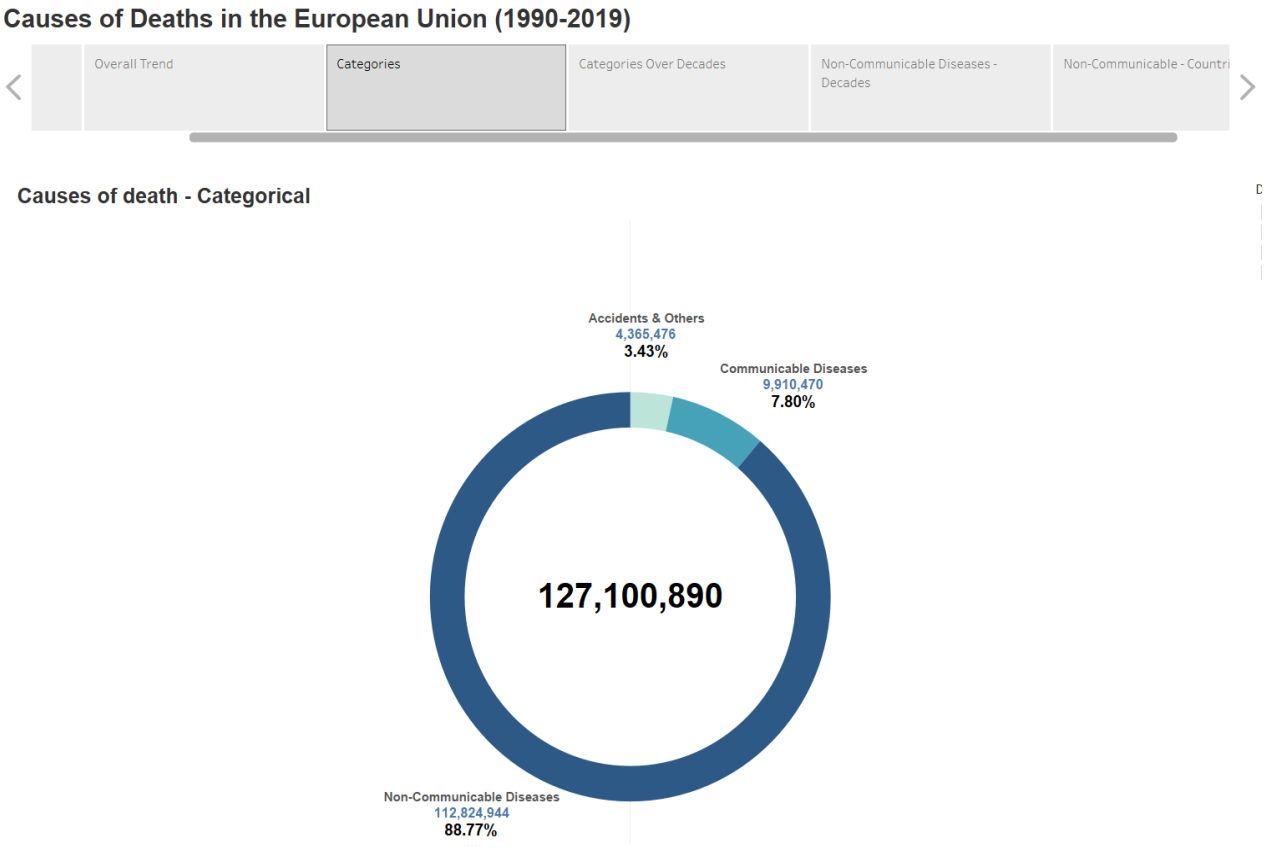

Figure 2

All the causes of death have been categorized into 3: communicable diseases, non-communicable diseases, and accidental deaths. It can be observed that non-communicable diseases contributed to 88.77% of deaths.

Moreover, a Grouped bar chart was selected to find the percentage change for various categories over three decades. The x-axis represents the different decades group in three categories i.e., non-Communicable Diseases, Communicable Diseases, Accidents, and others.

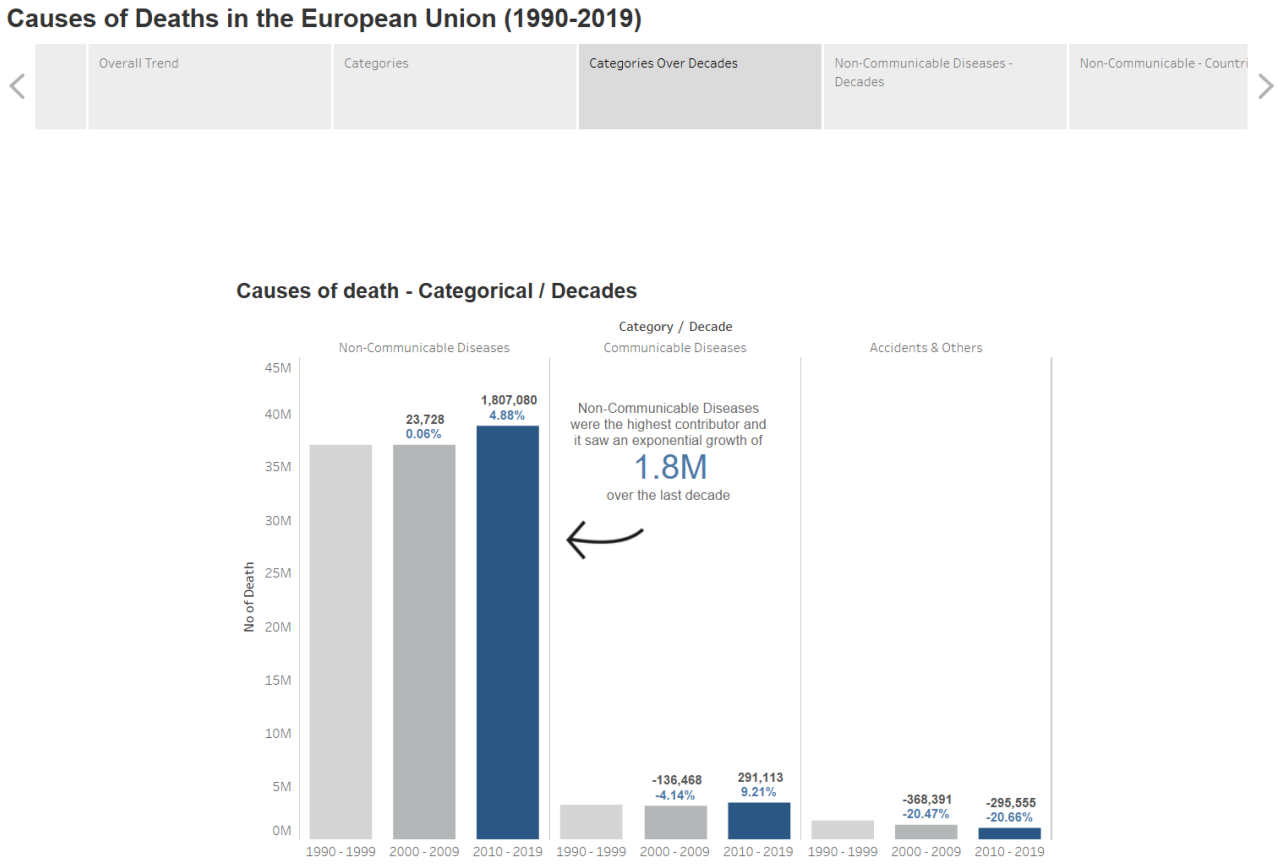

Figure 3

Non-communicable diseases were the highest contributor to the death rate over all 3 decades. The number of deaths due to non-communicable diseases saw an exponential growth of 1.8 million in the last decade. It is followed by deaths caused by communicable diseases and then accidental deaths.

Question 3

3. **What were the major causes of death for the highest contributing category?**

A Stacked horizontal bar chart has been used to visualize our third question. The x-axis represents the number of deaths and y-axis represents top 10 causes of deaths in a particular category. We have also used a dropdown filter for categories to analyze the visualization more effectively. In this chart, the grey horizontal bar represents the decade 2000-2009 and the blue horizontal bar represents the decade 2010-2019.

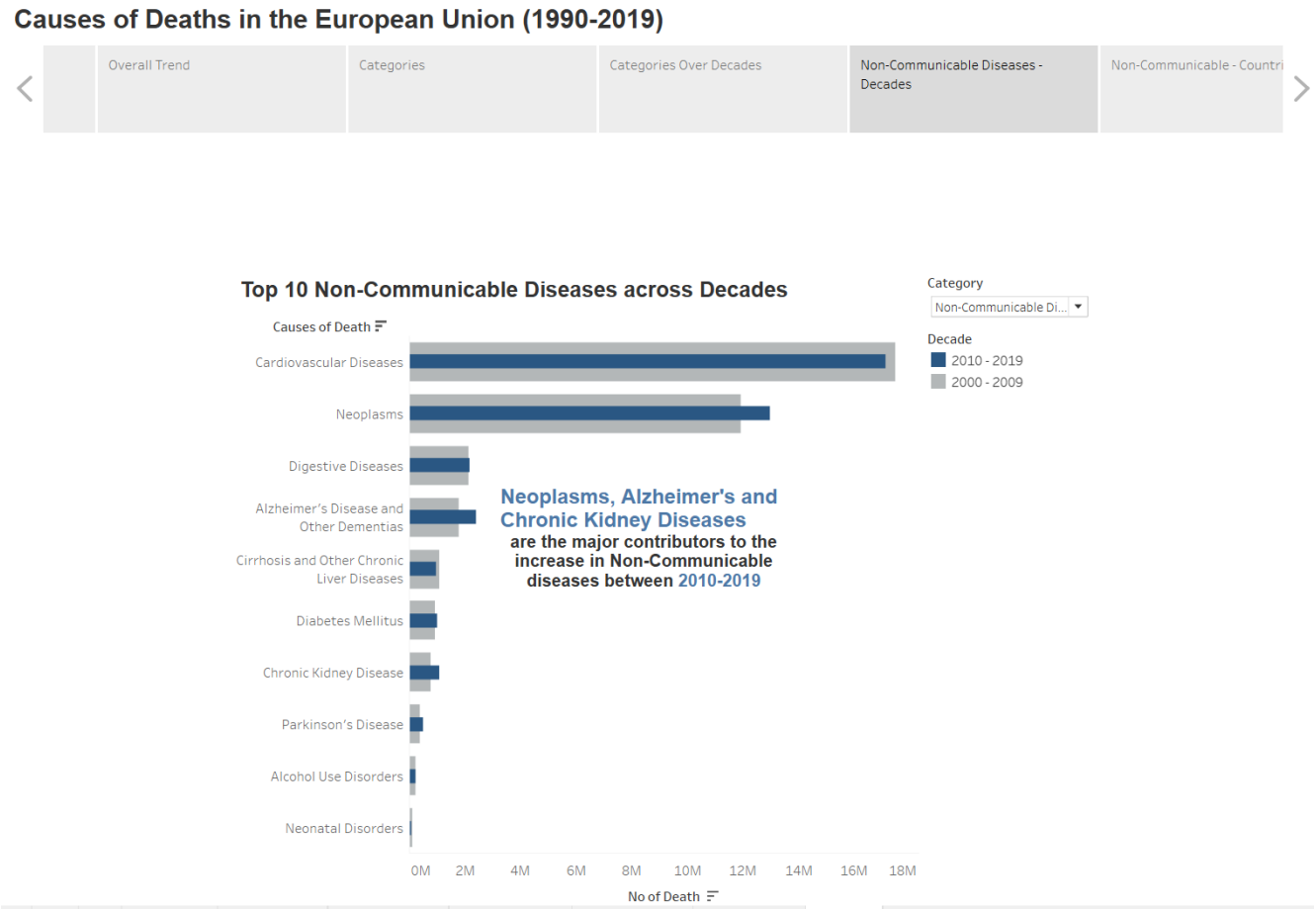

Figure 4

Non-communicable diseases were the highest contributing category, with cardiovascular diseases, neoplasms and digestive diseases being the top 3 contributors in the last 2 decades. Neoplasms, Alzheimer's, and chronic kidney diseases saw the most increase in deaths between 2000-2009 and 2010-2019.

Question 4

4. Which countries had the highest number of deaths for the major causes?

A Geographical map has been used to find the countries in EU that had highest number of deaths due to non-communicable diseases. We have also added a horizontal bar graph in the tooltip to visualize the highest contributing top 7 causes to the total number of deaths in a particular country. Moreover, we have added a dynamic insight about the percentage contributed by the selected country in the total number of deaths in EU during the time period of 30 years.

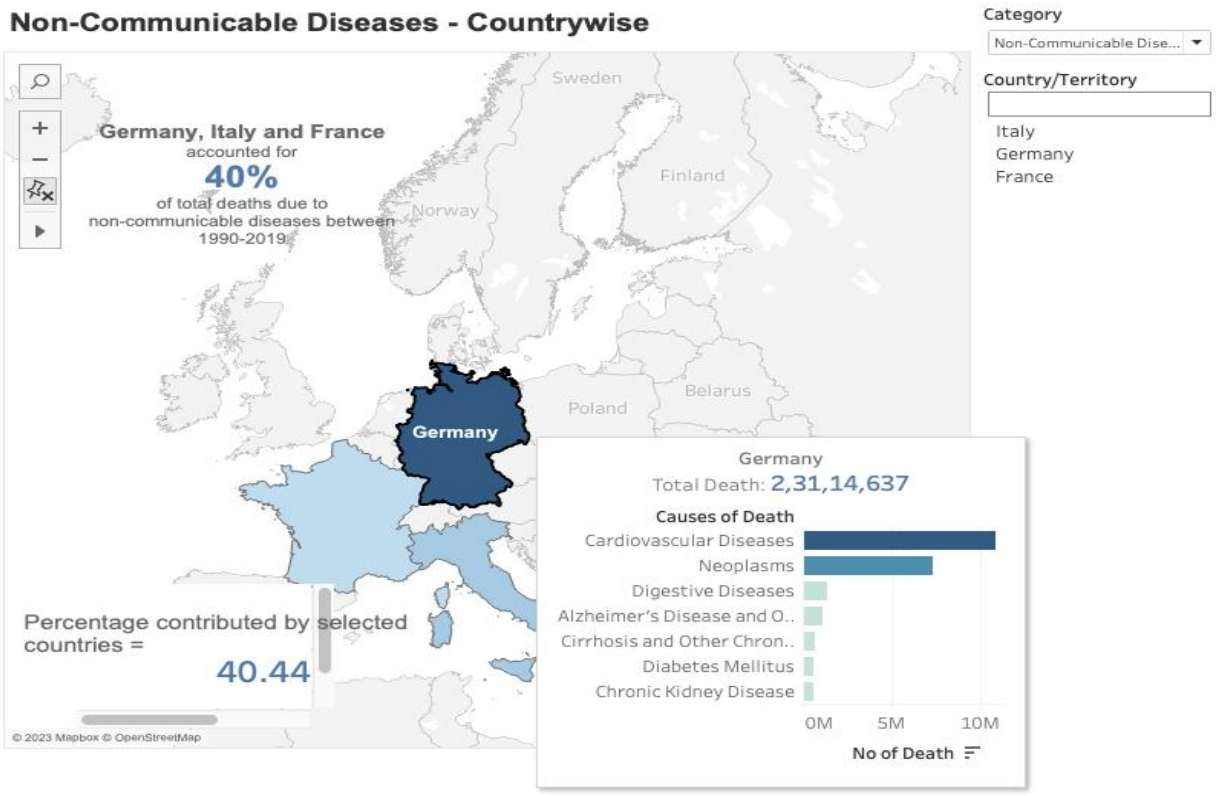

Figure 5

Cardiovascular diseases, neoplasms and Alzheimer's were the major causes of deaths in Germany, Italy, and France. These 3 countries alone accounted for 40% of all deaths in the EU between 1990 and 2019.

5. Conclusion

Visualizing the dataset on causes of death helped us conclude the following –

1. The trend for the number of deaths showed a steady increase. There was a spike observed in the last decade (2010-2019).
2. Further on visualizing the various categories of deaths over the decades, it was observed that the categories of non-communicable diseases and communicable diseases increased in the last decade whereas the opposite was the case for accidents and others.
3. Even though the percentage increase for communicable diseases was higher than that of non-communicable diseases, the number of deaths caused by non-communicable deaths were extremely high (1.8 million).
4. Analyzing the causes of death in the non-communicable category helped us identify that cardiovascular diseases, neoplasms and digestive diseases contributed to the highest number of

deaths whereas neoplasms, Alzheimer's and chronic kidney diseases saw the highest percentage increase in the number of deaths from the previous decade.

5. Germany, Italy, and France saw the largest number of deaths due to cardiovascular diseases, neoplasms and Alzheimer's and totaled to around 40% of all deaths in the EU between 1990 and 2019.