-------------------------   --------------------

-------------------------   ----------------

FLIP ROBO

NAME OF THE PROJECT

CAUSE OF DEATH
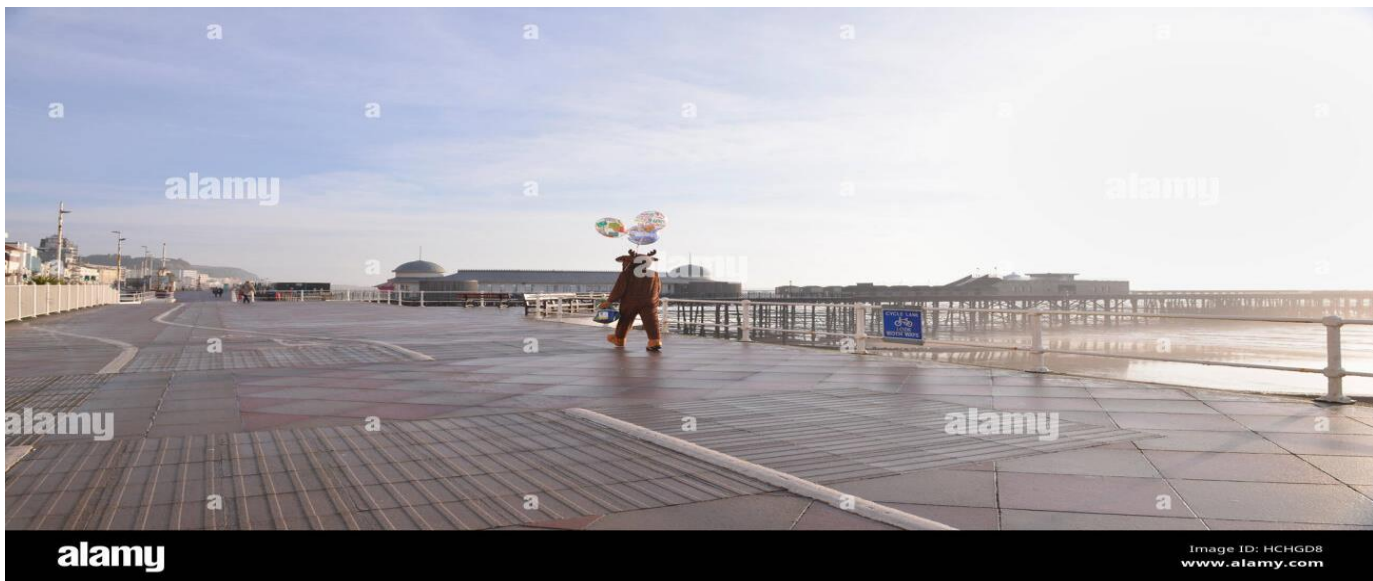
Submitted To: Submitted By:

Ms. Khushboo Garg   **MONIKA LATHAR**

SME   **Intern-34**

ACKNOWLEDGMENT

I have taken efforts in this project. However, it would not have been possible without the kind support and help of many individuals and organizations. I would like to extend my sincere thanks to all of them.

I am highly indebted to **Ms KHUSBOO GARG** for their guidance and constant supervision as well as for providing necessary information regarding the project & also for their support in completing the project.

I would like to express my gratitude towards my parents & member of **Flip Robo Technology** for their kind co-operation and encouragement which help me in completion of this project.

I would like to express my special gratitude and thanks to industry persons for giving me such attention and time.

My thanks and appreciations also go to my colleague in developing the project and people who have willingly helped me out with their abilities.

I have studied lot of research papers and references for completing this project of **Cause Of Death.**

List of all the research papers as given below:

1) https://www.kaggle.com/code/spscientist/a-simple-tutorial-on-exploratory-data-analysis

2) https://en.wikipedia.org/wiki/Exploratory_data_analysis#:~:text=In%20statistics%2C%20exploratory%20data%20analysis,and%20other%20data%20visualization%20methods.

3) Alai, Daniel H., Séverine, Arnold, Madhavi, Bajekal, and Andrés M., Villegas. 2018. Mind the Gap: A Study of CauseSpecific Mortality by Socioeconomic Circumstances. North

4) American Actuarial Journal, 22, no. 2:161-181. American Cancer Society. 2018. Cancer Facts & Figures2018. https://www.cancer.org/research/cancer-factsstatistics/all-cancer-facts-figures/cancer-facts-figures-2018.html (accessed on October 17, 2019).

5) Arnold, Séverine, Alexandre, Boumezoued, Héloïse Labit, Hardy, Nicole, El Karoui. (2018). Cause-of-Death Mortality: What Can Be Learned from Population Dynamics?. Insurance: Mathematics and Economics, 78:301-315.

6) Bergeron-Boucher, Marie-Pier, Vladimir, Canudas-Romo, Jim, Oeppen, and James W., Vaupel. 2017. Coherent Forecasts of Mortality with Compositional Data Analysis. Demographic Research, 37:527-566.

ABSTRACT

The demographic and epidemiological transitions of the past 200 years are well documented at an aggregate level. Understanding differences in individual and group risks for mortality during these transitions requires linkage between demographic data and detailed individual cause of death information. This paper describes the digitization of almost 185,000 causes of death for Ohio to supplement demographic information in the Longitudinal, Intergenerational Family Electronic Micro-database (LIFE-M). To extract causes of death, our methodology combines handwriting recognition, extensive data cleaning algorithms, and the semi-automated classification of causes of death into International Classification of Diseases (ICD) codes. Our procedures are adaptable to other collections of handwritten data, which require both handwriting recognition and semi-automated coding of the information extracted.

INTRODUCTION

WHAT IS A CAUSE OF DEATH?

The International Classification of Diseases (ICD) assigns all human deaths to a cause. The ICD is very granular and contains different levels of classification. The different providers of data on death use these ICD classifications to assign a cause to each. Databases on causes of death are available at granular levels, sometimes with hundreds of causes. The instability of historical data can be a problem at this level of granularity. When forecasting death rates with stochastic mortality, the number of causes of death forecasted must be limited in order to guarantee minimal exposure and therefore, a statistically reliable estimated death rate. The benefits of using cause-specific death rates are that they provide a good split of aggregate death rates and all of the causes sum up to the aggregate.

How does early-life nutrition affect the development of cancer, adult hypertension, and cardiovascular disease? How does childhood exposure to toxins such as lead affect suicide rates and the development of Alzheimer's disease? How did the early 20th century's public health campaigns to clean water, improve sanitation, and deliver vaccines affect later-life health and longevity? Demographic and epidemiological transitions have brought substantial improvements in human welfare over the past 200 years; however, the role of individual genetics, circumstances, local policies, and environmental conditions in shaping health is less well understood. The lack of historical individual-level data on cause of death has been a key limitation.

Our methods are adaptable to other historical sources containing semi-structured handwriting. Our contributions include methods for locating a specific piece of information in structured documents (that change format over time), standardizing terms using Google's spelling correction tools, and mapping causes of death into a classification scheme conformable with historical language and modern death classification. Similar processes could be used for other structured sources with features that need to be classified into finite lists. Because our goal was to take production to scale, we invested in developing algorithms to parse words and phrases into causes of death and a concise and informative classification system. This paper provides a roadmap of the process including lessons learned. At key points, we highlight forks in the road and alternative routes, where other researchers could make different choices.

It is important to understand what is meant by the *cause* of death and the risk factor associated with a premature death:

In the epidemiological framework of the Global Burden of Disease study each death has *one* specific cause. In their own words: 'each death is attributed to a single underlying cause — the cause that initiated the series of events leading to death'.

The Global Burden of Disease is a major global study on the causes of death and disease published in the medical journal *The Lancet*.[4] These estimates of the annual number of deaths by cause are shown here.

This is shown for deaths worldwide. But you can explore data on the annual number of deaths by cause for any country or region using the "change country" toggle.

Non-communicable diseases (NCDs) not only dominate mortality figures at a global level, but also account for the majority of deaths in high-income countries.

Deaths from causes such as infectious disease, malnutrition, nutritional deficiencies, neonatal and maternal deaths are common – and in some cases dominant – across low- and middle-income nations. In Kenya, for example, the leading cause of death remains diarrheal diseases. In South Africa and Botswana, the leading cause of death is HIV/AIDS. In high-income countries however the share of deaths caused by these is very low.

Using the timeline on the chart you can also explore how deaths by cause have changed over time.

Death rates related to disease, illness and other health factors tend to change relatively slowly over time. Whilst death rates may fall or decline from year-to-year as part of a general trend, dramatic changes in such deaths are typically rare. Natural disaster and terrorism-related deaths are an important exception to this rule, as they can vary significantly between countries. This can make the annual comparison of deaths and death rates between health-related factors and volatile events more challenging. Understanding the relative risk of these events can require a longer-term overview of high and low-mortality years. We cover discussion and analysis on this topic in a post here.

CAUSE OF DEATH CLASSIFICATION FOR THIS PROJECT

Here I used the data furnished by the Flip Robo Technology in CSV File Cause-of-Death Database (COD) and In this Dataset, we have Historical Data of different cause of deaths for all ages around the World. The key features of this Dataset are as shown in below :

Meningitis, Alzheimer's Disease and Other Dementias, Parkinson's Disease, Nutritional Deficiencies, Malaria, Drowning, Interpersonal Violence, Maternal Disorders, HIV/AIDS, Drug Use Disorders, Tuberculosis, Cardiovascular Diseases, Lower Respiratory Infections, Neonatal Disorders, Alcohol Use Disorders, Self-harm, Exposure to Forces of Nature, Diarrheal Diseases, Environmental Heat and Cold Exposure, Neoplasms, Conflict and Terrorism, Diabetes Mellitus, Chronic Kidney Disease, Poisonings, Protein-Energy Malnutrition, Road Injuries, Chronic Respiratory Diseases, Cirrhosis and Other Chronic Liver Diseases, Digestive Diseases, Fire, Heat, and Hot Substances, Acute Hepatitis.
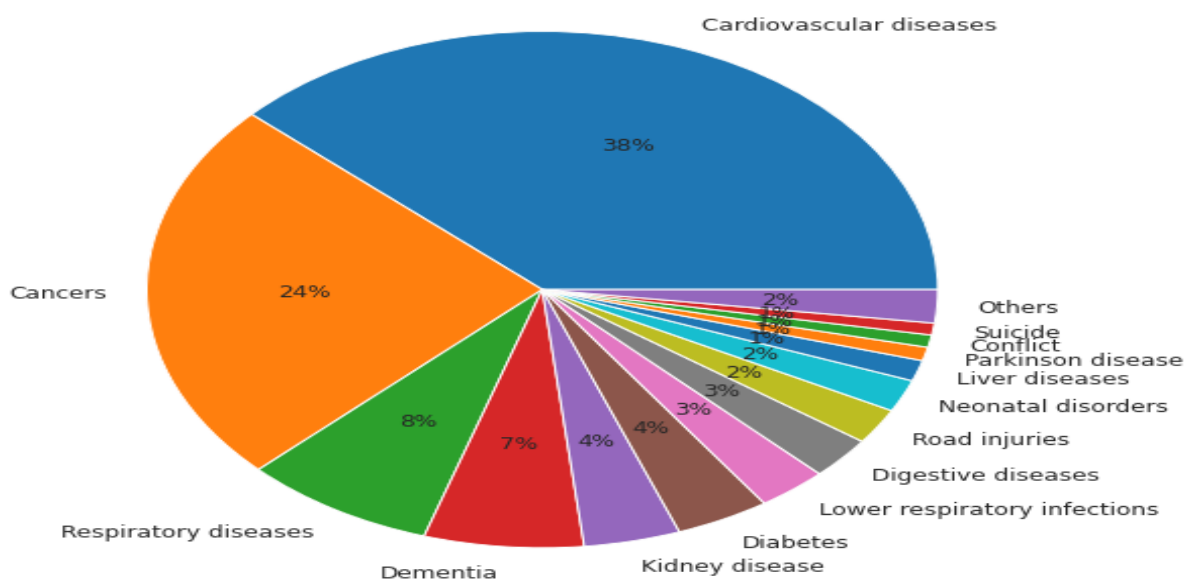
The dataset consists of 34 different cause of death counts by country per year. It covers the years from 1990 to 2019.

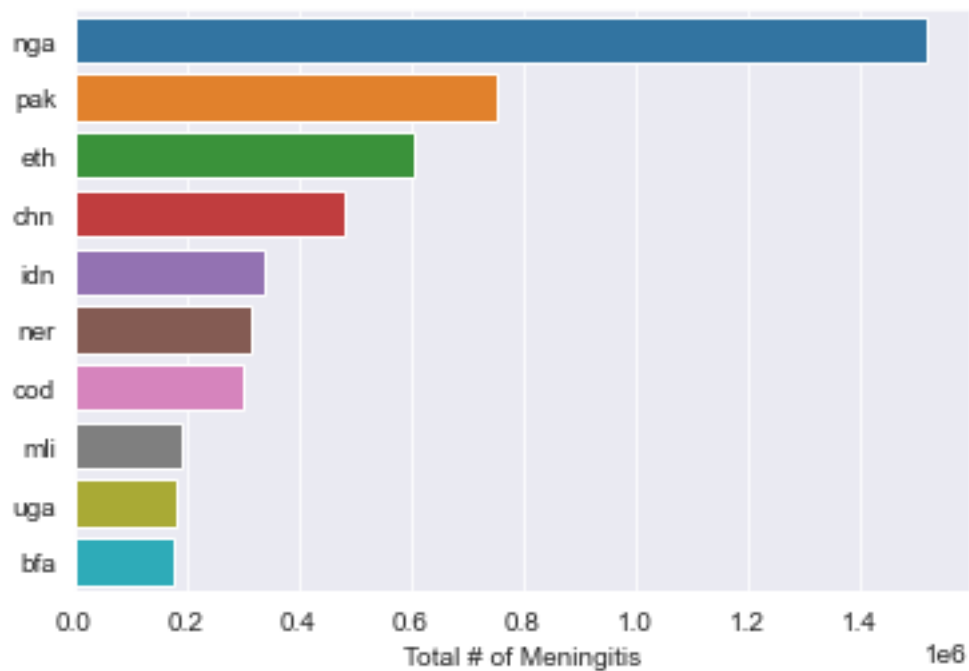Top 10 causes of death for Turkey in the latest available year, 2019:
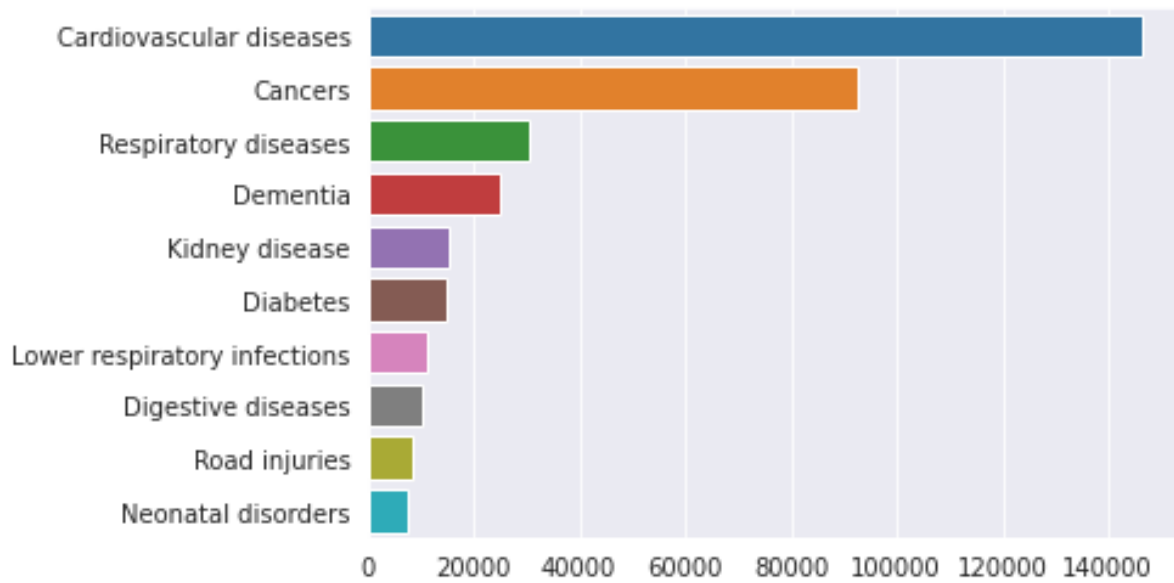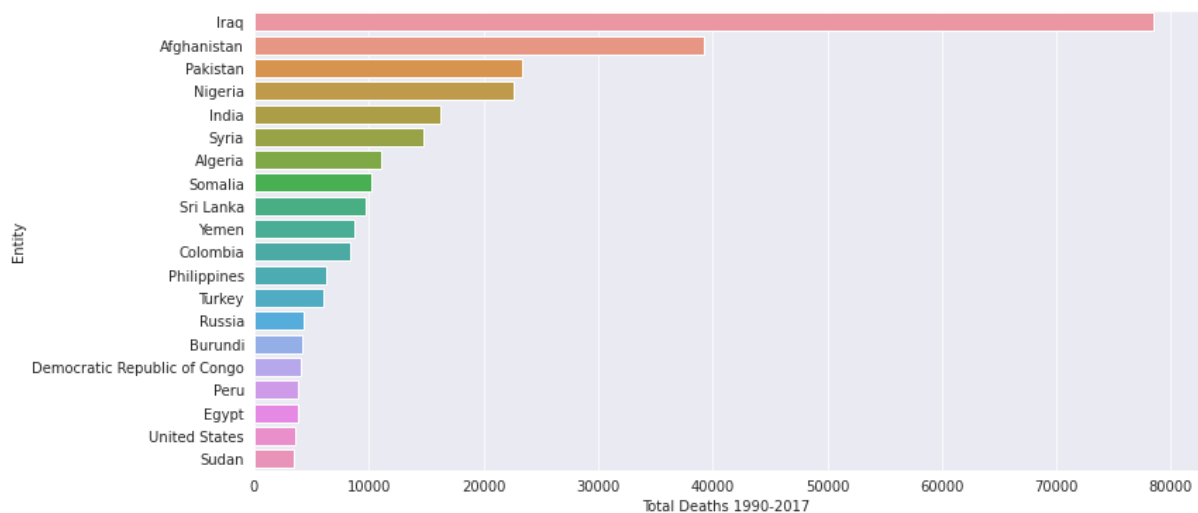
Percentages in pie graph:

- Almost 4 out 10 deaths in Turkey are caused by Cardiovascular diseases in 2019

Top 10 countries with the highest **execution** numbers

Top 20 countries which suffered from **terrorism** the most. Our dataset Entity column contains a mix of country, continent, region, territory information too such as Sub-Saharan Africa, South America etc.

Deadliest year appears to be **2009**, however this is due to the increase of the world population each year. We need to check death rate per year.
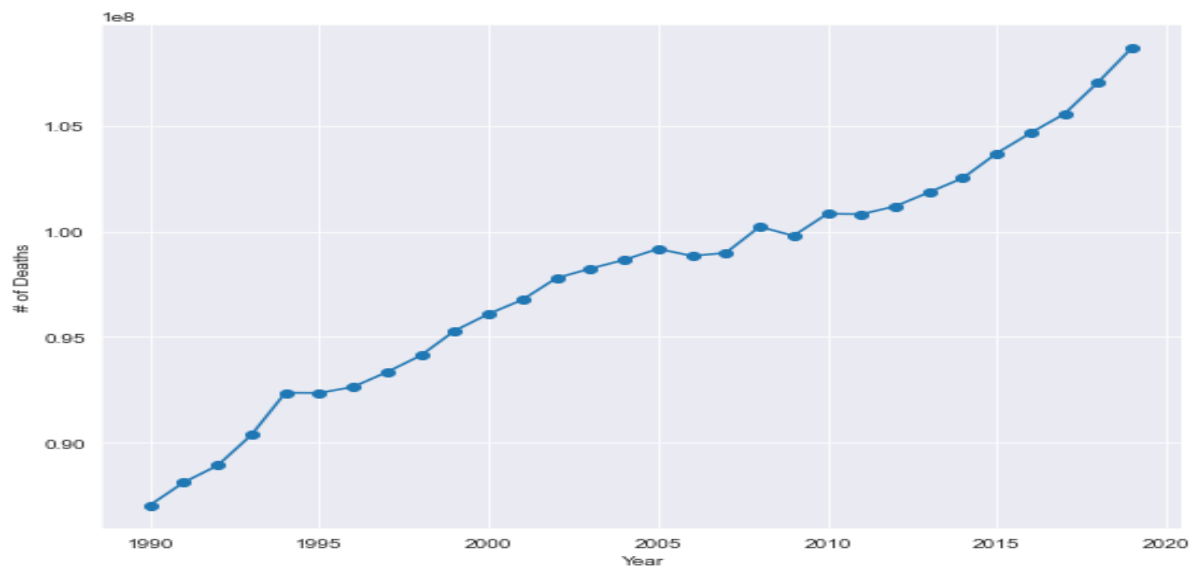

- 2009 was the deadliest year with almost 4% of the population was died.
- Lets examine what was the cause for this high death rate in 2009. Conflict line of the heatmap shows an interesting light colour on year 2009 (the lighter the colour the higher the number of deaths)
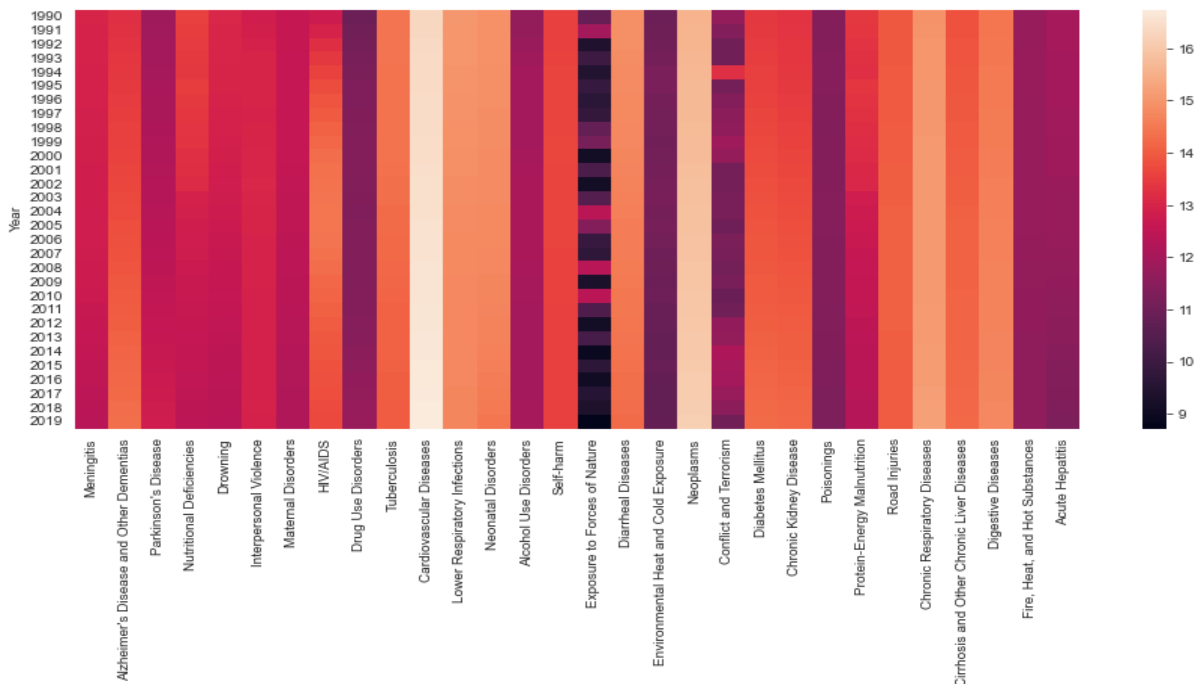
Conflict graph shows that **1994** was the year where the deadliest conflicts took place in the world. Bosnian War in Europe may have contributed to it significantly.
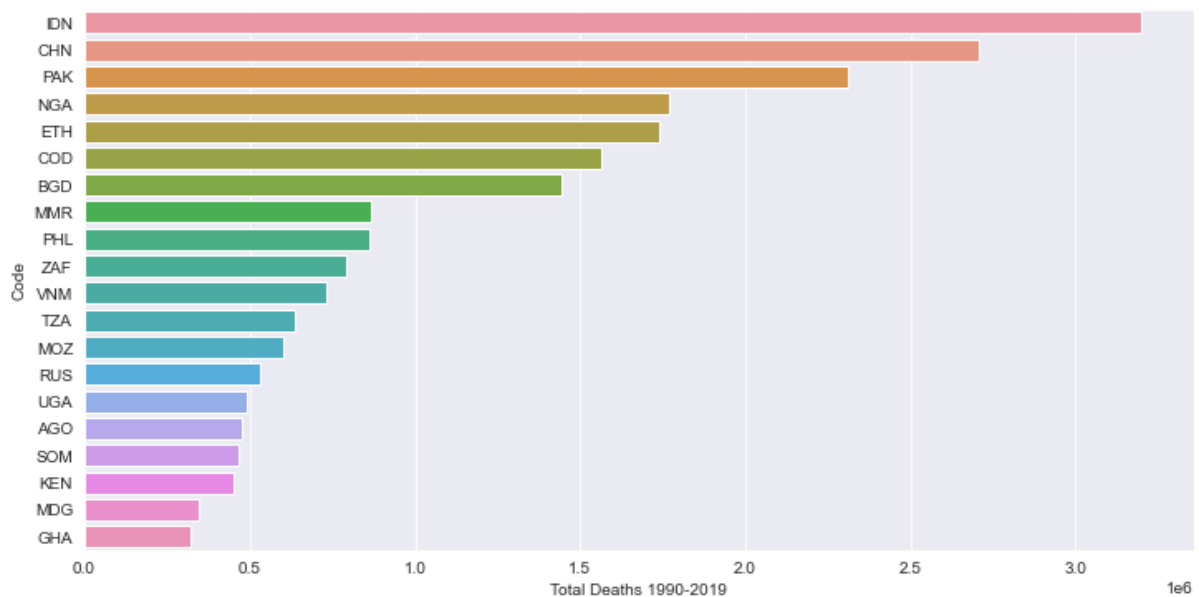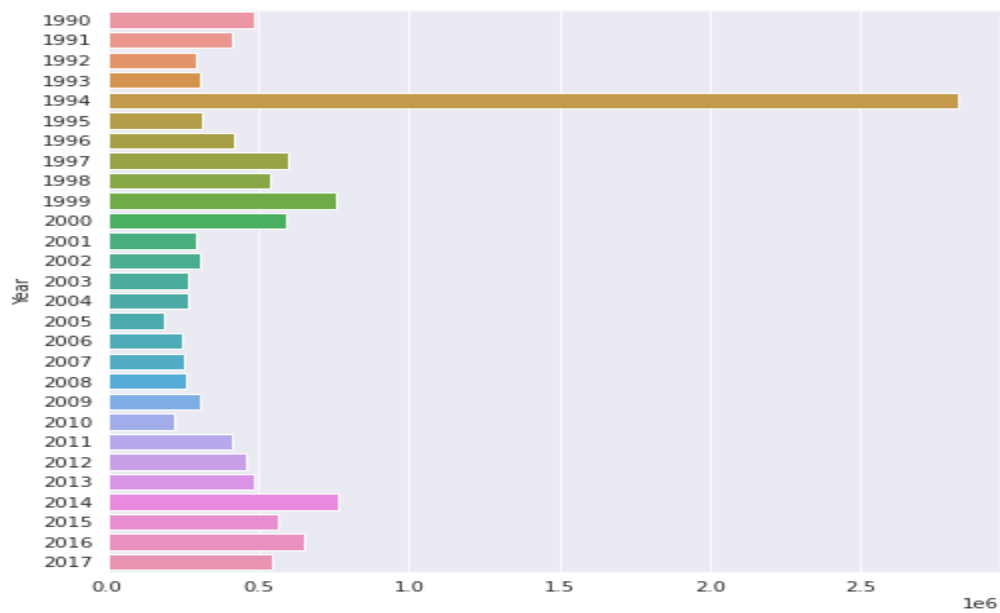
Death by Terrorism and Tuberculosis peaked in 2014, and has been declining since then, however no effect was observed on 1994 death toll
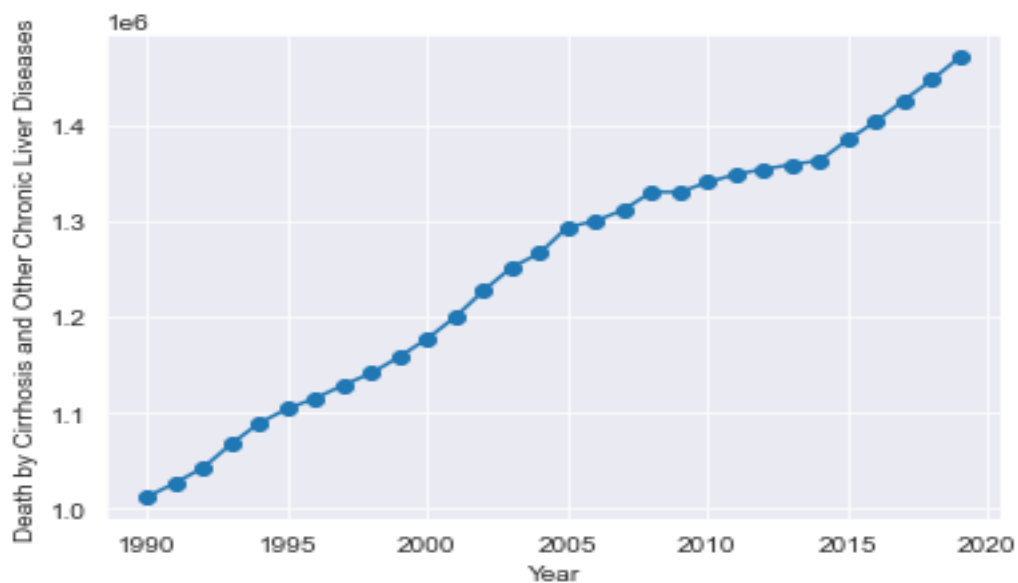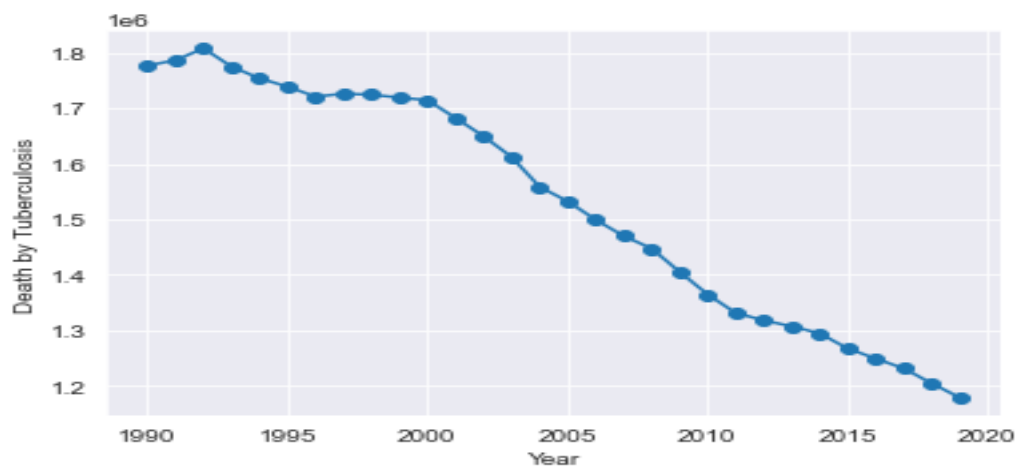
Causes of death by category

The share of deaths from infectious diseases are declining; a larger share is dying from NCDs

 In the visualization we see the distribution of global deaths broken down by three broad categories:

- 1 – in yellow: Injuries caused by road accidents, homicides, conflict deaths, drowning, fire-related accidents, natural disasters and suicides.
- 2 – in blue: Non-communicable diseases. These are often chronic, long-term illnesses and include cardiovascular diseases (including stroke), cancers, diabetes and chronic respiratory diseases (such as chronic pulmonary disease and asthma, but excluding infectious respiratory diseases such as tuberculosis and influenza).
- 3 – In red: Communicable diseases (i.e. infectious diseases) such as HIV/AIDS, malaria, and tuberculosis together with maternal deaths, neonatal deaths and deaths from malnutrition.

This is shown for global deaths as the default, but can be viewed for any country or region using the "change country" toggle on the interactive chart.

At a global level we see that the majority of deaths are caused by non-communicable diseases (NCDs). Collectively NCDs account for more than 73% of global deaths. As the world is making progress in the fight against many infectious diseases, and as populations age, we expect that NCDs will become increasingly dominant as the cause of death.
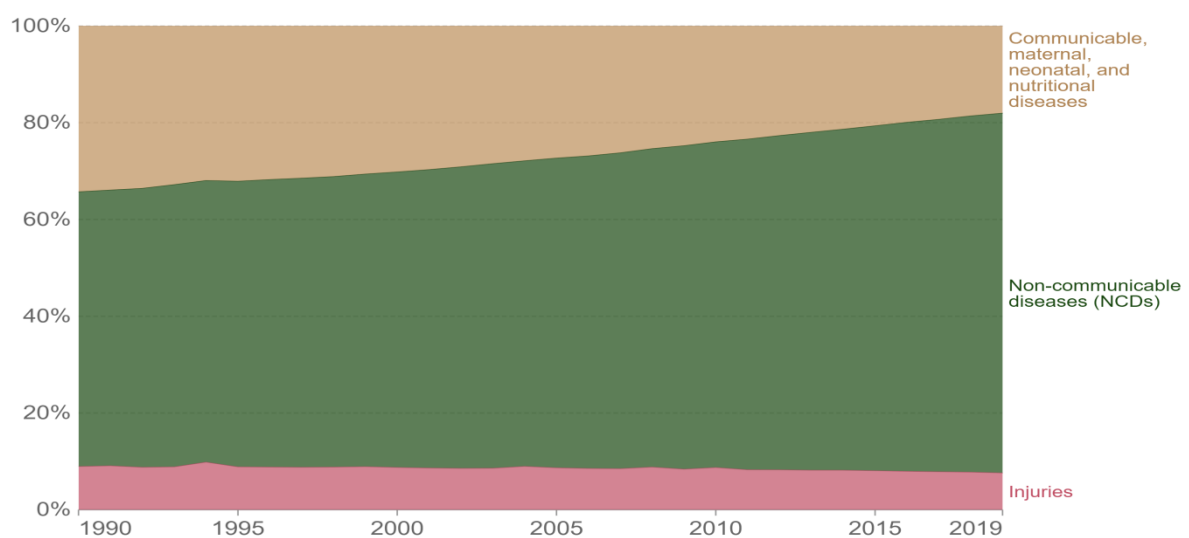
Noncommunicable diseases (NCDs) include cardiovascular disease, cancers, diabetes and res

piratory disease.Injuries include road accidents, homicides, conflict deaths, drowning, fire-related accidents, natural disasters andsuicides.

Number of deaths by cause, World, 2019

Cause	Deaths
Cardiovascular diseases	18.56 million
Cancers	10.08 million
Respiratory diseases	3.97 million
Digestive diseases	2.56 million
Lower respiratory infections	2.49 million
Neonatal disorders	1.88 million
Dementia	1.62 million
Diabetes	1.55 million
Diarrheal diseases	1.53 million
Liver diseases	1.47 million
Kidney diseases	1.43 million
Road injuries	1.2 million
Tuberculosis	1.18 million
HIV/AIDS	863,837
Suicide	759,028
Malaria	643,381
Homicide	415,180
Parkinson's disease	362,907
Nutritional deficiencies	251,577
Drowning	237,242
Meningitis	236,222
Protein-energy malnutrition	212,242
Maternal disorders	196,471
Alcohol use disorders	168,015
Drug use disorders	128,083
Fire	111,292
Hepatitis	79,176
Poisonings	77,162
Conflict and terrorism	62,985
Heat (hot and cold exposure)	47,461
Natural disasters	6,076

Source: IHME, Global Burden of Disease (2019)   OurWorldInData.org/causes-of-death • CC BY

An infection with HIV (human immunodeficiency virus) can lead to AIDS (acquired immunodeficiency syndrome). AIDS results in a gradual and persistent decline and failure of the immune system, resulting in heightened risk of life-threatening infection and cancers.

In the majority of cases, HIV is a sexually-transmitted infection. However, HIV can also be transmitted from a mother to her child, during pregnancy or childbirth, or through breastfeeding. Non-sexual transmission can also occur through the sharing of injection equipment such as needles.

In the map we see death rates from HIV/AIDS across the world.
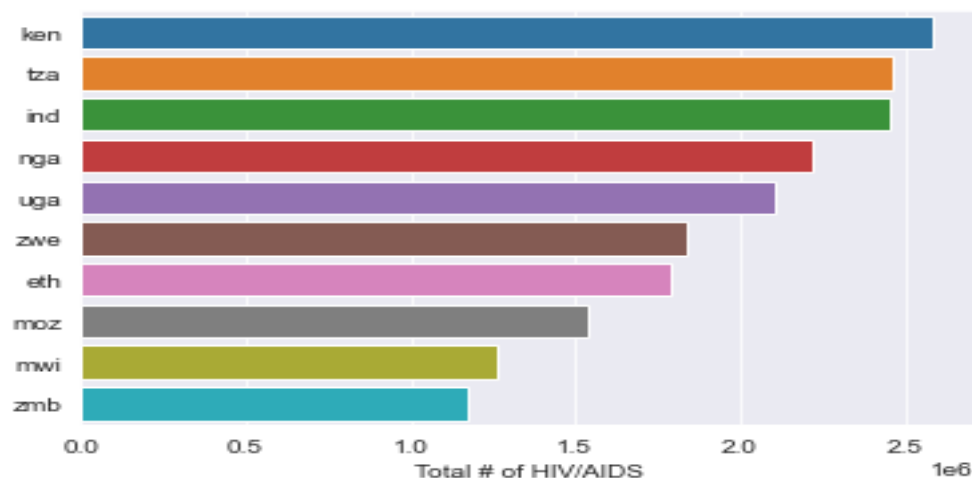
Most countries have a rate of less than 10 deaths per 100,000 – often much lower, below 5 per 100,000. Across Europe the death rate is less than one per 100,000.

Across Sub-Saharan Africa the rates are much higher. Most countries in the South of the region had rates greater than 100 per 100,000. In South Africa and Mozambique, it was over 200 per 100,000.

DROWNING

The World Health Organization (WHO) emphasises that drowning is one of the most overlooked, preventable causes of death across the world.[8] For every country in the world, drowning is among the top 10 killers for children. In some countries, such as Bangladesh, it is the top mortality cause for children under 15 years old.

In the map we see death rates from drowning across the world.

In 2016, death rates were highest in Papua New Guinea and Seychelles, between 10 to 16 deaths per 100,000. Rates were also high in countries such as Bangladesh, Central African Republic, Vietnam, and Haiti.

If we look at death rates we see a significant decline since 1990 — especially in low to middle-income countries. In Bangladesh and China, for example, rates have fallen by more than two-thirds over this period.
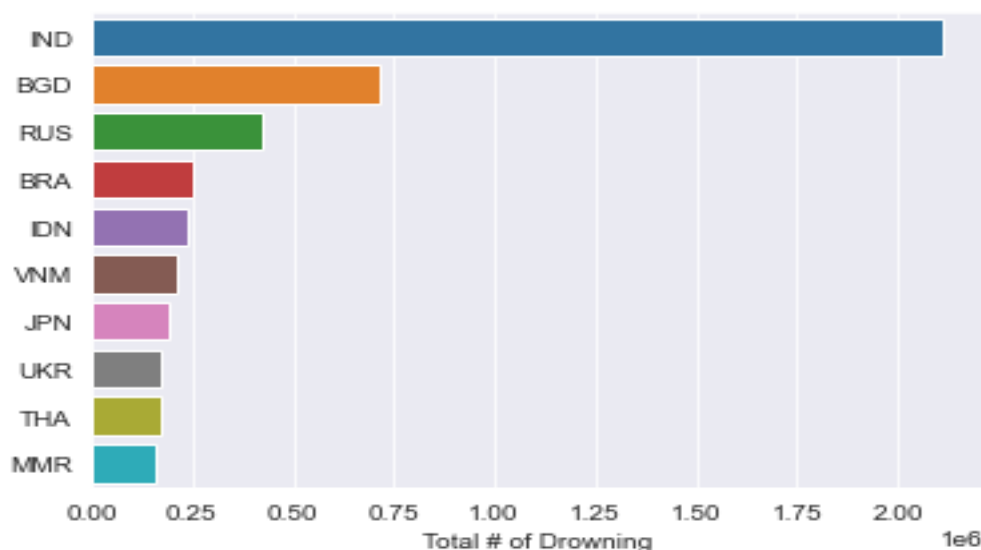
Let's find out which country suffers from what disease the most
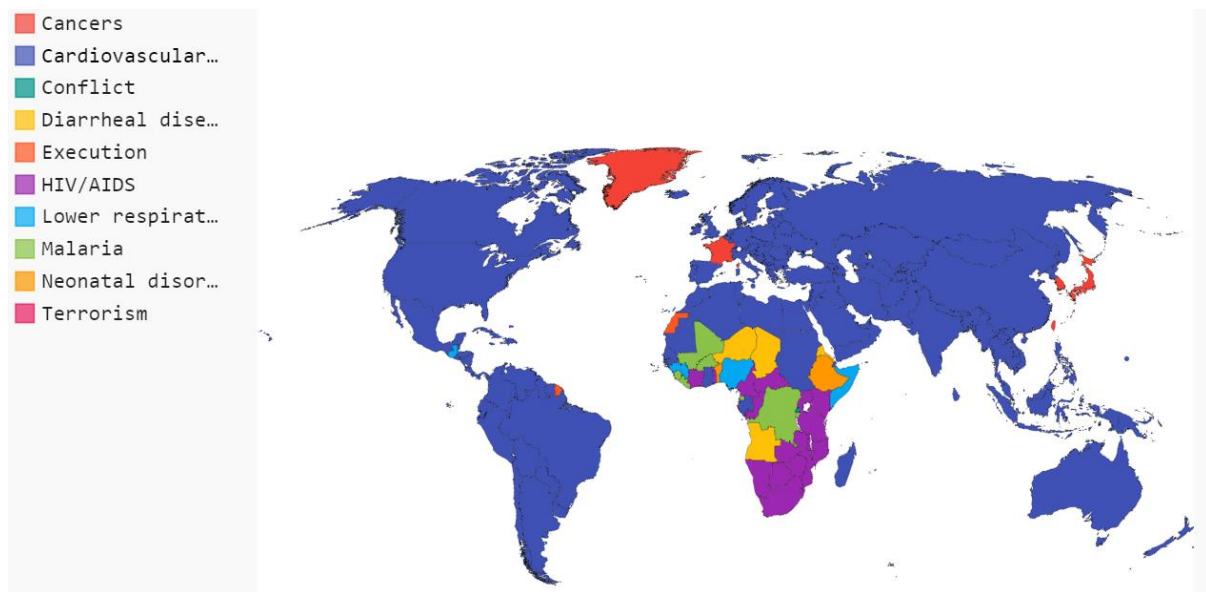
It is obvious that Cardiovascular Disease dominate death toll all over the world, how would the World map look if we took them out?

- It appears that the second most leading cause of death in rich countries like USA, Canada, Europe, Japan and Australia is Dementia (mainly caused by Alzheimer's disease). This could be due to the high frequency of elder people in their population.

- **Conflict** is leading cause for war territories like Syria and Palestine

- For Russia and their neighbours like old Soviet Union countries, Eastern European countries leading cause is **a Digestive disease** like Ulcer, Cirrhosis, and Hepatitis. It could be related to excessive consumption of alcohol.

- **Diarrheal** diseases causes deaths mostly in the mid African region

- **HIV/AIDS** deaths are the most frequent in the South African region

- Some of the South American countries have **Homicide** as the leading cause for death

- **Lower respiratory infections** like Tuberculosis, Pneumonia is the leading cause for mainly South American countries like Brazil and Argentina.

- China, India and surrounding countries suffers from **Respiratory diseases** like Asthma and Lung Cancer etc.

- The countries where **road accidents** are one of the leading cause of death are in the Gulf region like Iran, Saudi Arabia, UAE. It could be related to their habit of car stunt driving.

- Greenland is the only country where **suicide** is the leading cause of death.
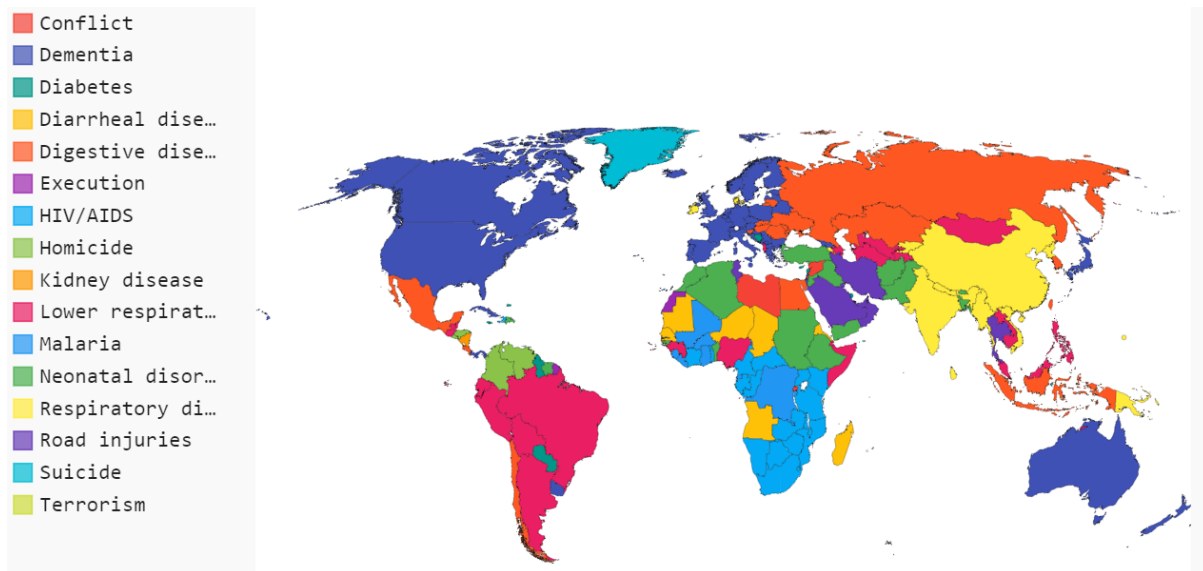
Let's now group the diseases by their historical increasing or decreasing trends. Calculations will be based on deaths per 100000 for particular disease.

Following diseases show monotonic decrease trend over years:

- Even though Digestive and Respiratory diseases were dropping constantly for period of time, they started to enter an increasing trend since 2013. Still way better as compared to 90s

- Suicide rates are dropping since mid-90s

- Peak in the Heat graph could be a mistake in the dataset or it is a very distinguishing event happened in 2010 which caused this sudden jump
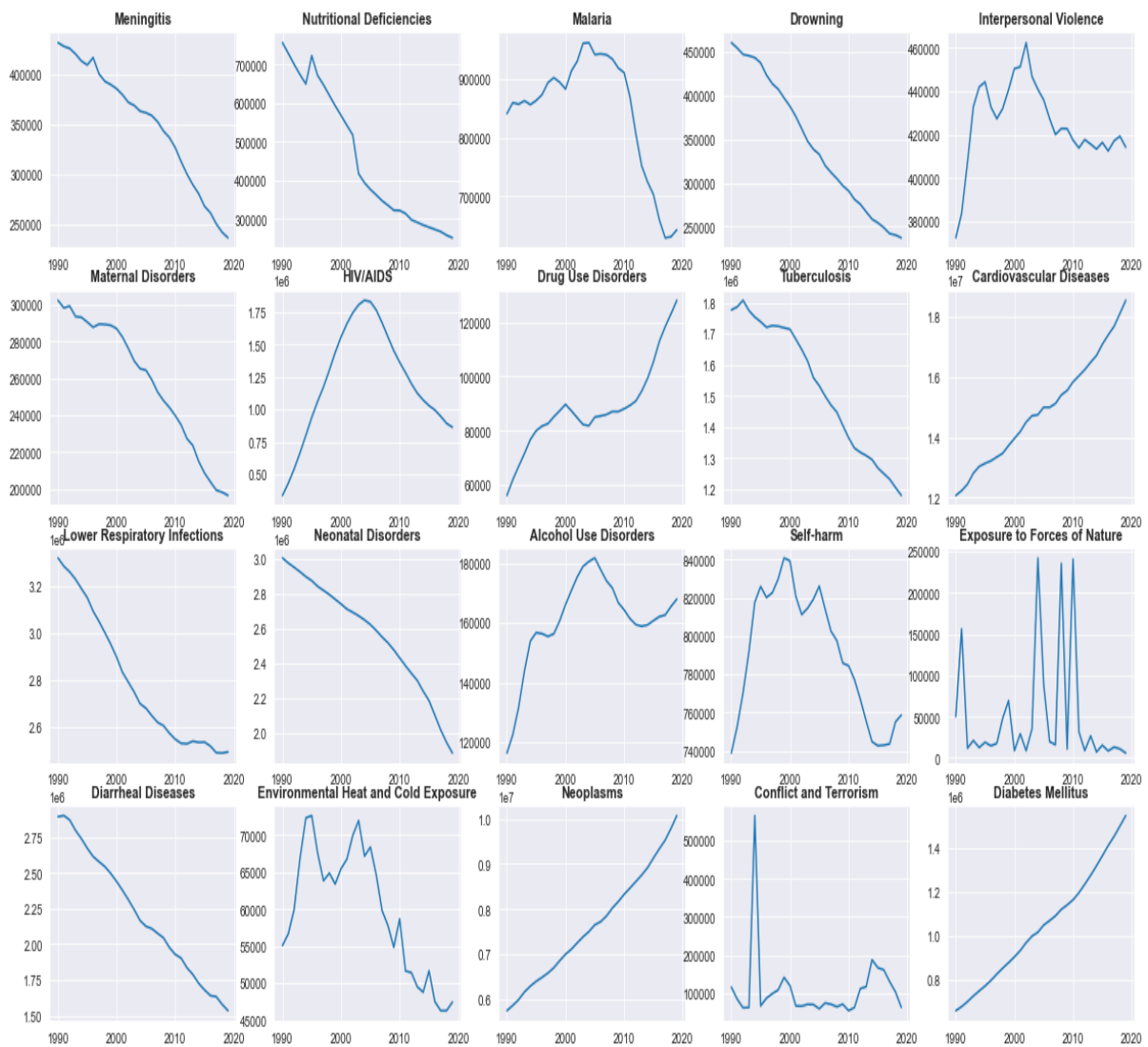
Following are the death causes exhibiting an increasing or no trend over time:

- Dementia, Kidney, Cancer, Parkinson, and Diabetes diseases have been increasing since the beginning of our data span (1990)

- Drug and alcohol use disorders are correlated with Liver disease trend

- HIV/AIDS related deaths rate started to decline since 2005

- Natural disasters show no trend as expected
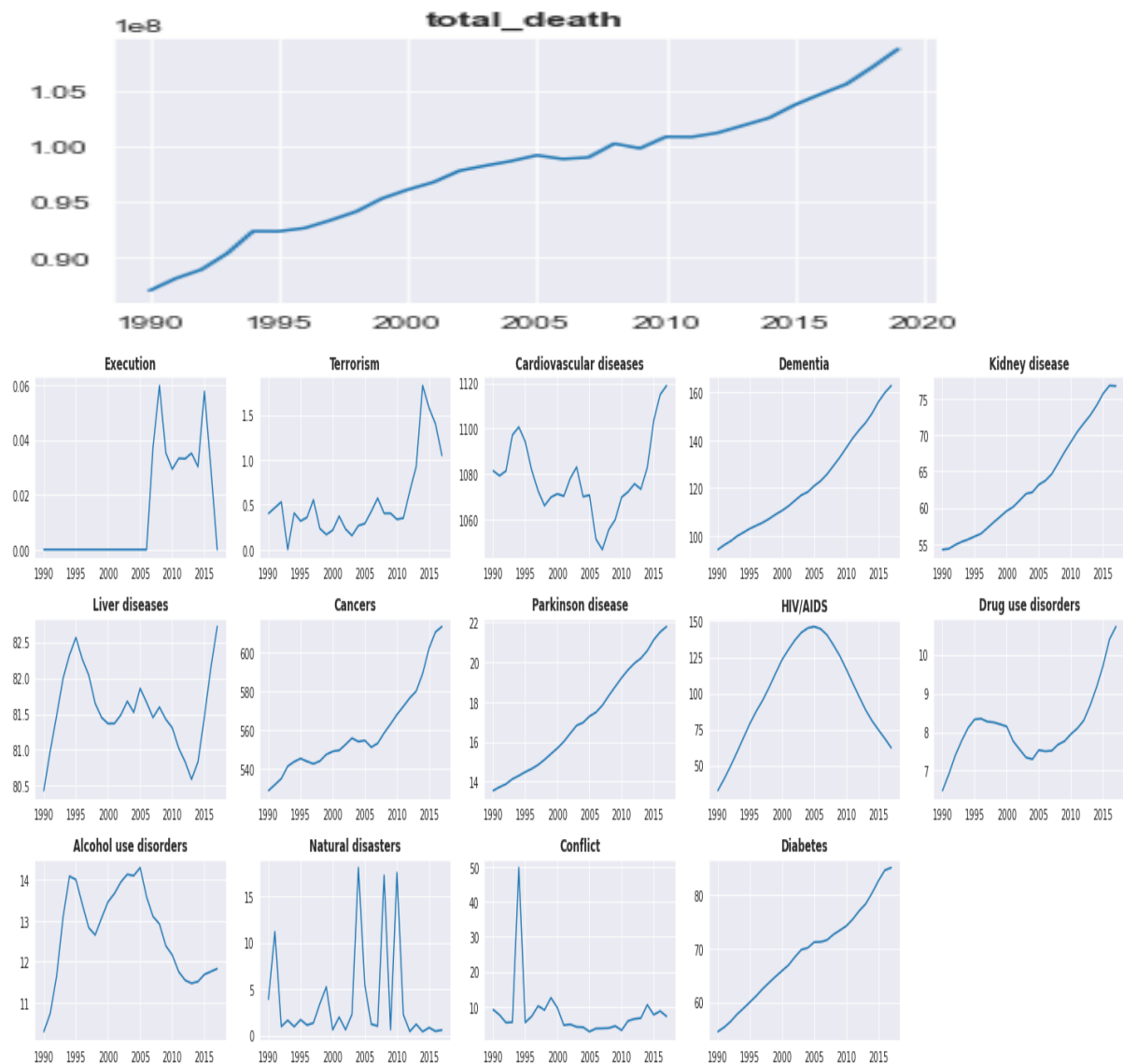
Conclusions

Death rates at ages 80 and over have been calculated for the post-1950 period for 28 low-mortality countries that process sufficiently reliable data on deaths by single years of age to allow the use of the extinct cohorts method. The material available covers most of the countries representative of the low-mortality countries of the world.

The study shows that mortality in old age has undergone in the developed countries during the post-war period a deep-going and fairly general transformation reaching much lower levels than have ever been recorded before. Unprecedented in known demographic history, this decline has made in the prevailing trend a break which had not been foreseen in population projections. These new factors which are distinctively period rather than cohort factors may include further advances in medical practice such as more generalized control of blood pressure but there is on the other hand solid evidence from many countries to the effect that an important and probably the main factor has been a change for a healthier lifestyle - a change which has been shown to have almost immediate effects on the probability of survival. Such behavioural changes prevent or delay the onset of morbid processes which would in time lead to serious health impairment or death. Maybe the aging process itself is thereby delayed.

We do not, as yet, know the ultimate extent of the new mortality transition. But considering that the new lifestyle has not yet been anything like universally adopted, a vast potential exists for further improvement in health. And since this social process seems to be going on and as there is at the present time a strong downward momentum in death rates, the new transition may, barring unforeseen events, still continue for an extended period.