CAUSE OF DEATH DATASET

PROBLEM STATEMENT:

Around 58 million people die each year worldwide.

What caused their death?

How did the causes of death change over time and differ between different countries and world regions?

These are the big questions we are answering here.

ABOUT THE DATASET:

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'**.

This is different from the deaths that happened due to **risk factors**. These deaths are an **estimation of the reduction of the number of deaths that would be achieved if the risk factors to which a population is exposed would be eliminated** (in the case of tobacco smoking, for example) or reduced to an optimal, healthy level (in the case of body-mass index).

A straightforward way to assess the health status of a population is to focus on mortality – or concepts like child mortality or life expectancy, which are based on mortality estimates. A focus on mortality, however, does not take into account that the **burden of diseases** is not only that they kill people, but that they cause suffering to people who live with them. Assessing **health outcomes by both mortality and morbidity** (the prevalent diseases) **provides a more encompassing view on health outcomes**.

The sum of mortality and morbidity is referred to as the 'burden of disease' and can be measured by a **metric** called **'Disability Adjusted Life Years'(DALYs).** DALYs are **measuring lost health** and are a standardized metric that allow for direct comparisons of disease burdens of different diseases across countries, between different populations, and over time. Conceptually, **one DALY is the equivalent of losing one year in good health** because of either premature death or disease or disability. **One DALY represents one lost year of healthy life.** The first 'Global Burden of Disease' (GBD) was GBD 1990 and the DALY metric was prominently featured in the World Bank's 1993 World Development Report. Today it is published by both the researchers at the Institute of Health Metrics and Evaluation (IHME)

and the ‘Disease Burden Unit’ at the World Health Organization (WHO), which was created in 1998. The IHME continues the work that was started in the early 1990s and publishes the Global Burden of Disease study.

Content

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: 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.

Dataset Glossary (Column-wise)

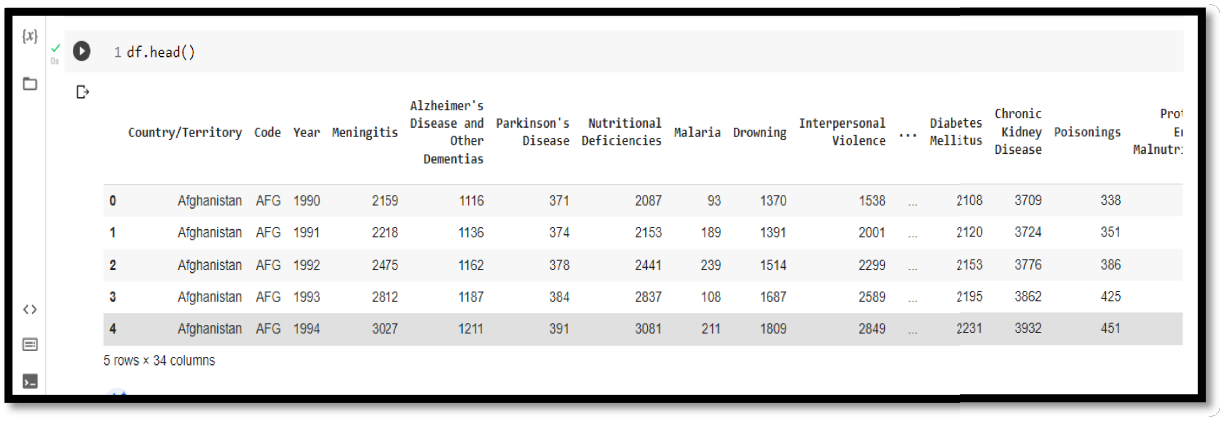
- 01. Country/Territory - Name of the Country/Territory
- 02. Code - Country/Territory Code
- 03. Year - Year of the Incident
- 04. Meningitis - No. of People died from Meningitis
- 05. Alzheimer's disease and Other Dementias - No. of People died from Alzheimer’s disease and Other Dementias
- 06. Parkinson's disease - No. of People died from Parkinson's disease
- 07. Nutritional Deficiencies - No. of People died from Nutritional Deficiencies
- 08. Malaria - No. of People died from Malaria
- 09. Drowning - No. of People died from Drowning
- 10. Interpersonal Violence - No. of People died from Interpersonal Violence
- 11. Maternal Disorders - No. of People died from Maternal Disorders
- 12. Drug Use Disorders - No. of People died from Drug Use Disorders
- 13. Tuberculosis - No. of People died from Tuberculosis
- 14. Cardiovascular Diseases - No. of People died from Cardiovascular Diseases
- 15. Lower Respiratory Infections - No. of People died from Lower Respiratory Infections

- 16. Neonatal Disorders - No. of People died from Neonatal Disorders
- 17. Alcohol Use Disorders - No. of People died from Alcohol Use Disorders
- 18. Self-harm - No. of People died from Self-harm
- 19. Exposure to Forces of Nature - No. of People died from Exposure to Forces of Nature
- 20. Diarrheal Diseases - No. of People died from Diarrheal Diseases
- 21. Environmental Heat and Cold Exposure - No. of People died from Environmental Heat and Cold Exposure
- 22. Neoplasms - No. of People died from Neoplasms
- 23. Conflict and Terrorism - No. of People died from Conflict and Terrorism
- 24. Diabetes Mellitus - No. of People died from Diabetes Mellitus
- 25. Chronic Kidney Disease - No. of People died from Chronic Kidney Disease
- 26. Poisonings - No. of People died from Poisoning
- 27. Protein-Energy Malnutrition - No. of People died from Protein-Energy Malnutrition
- 28. Chronic Respiratory Diseases - No. of People died from Chronic Respiratory Diseases
- 29. Cirrhosis and Other Chronic Liver Diseases - No. of People died from Cirrhosis and Other Chronic Liver Diseases
- 30. Digestive Diseases - No. of People died from Digestive Diseases
- 31. Fire, Heat, and Hot Substances - No. of People died from Fire or Heat or any Hot Substances
- 32. Acute Hepatitis - No. of People died from Acute Hepatitis

---

DATA ANALYSIS:

The dataset looks as below



```
1 df.head()
```

	Country/Territory	Code	Year	Meningitis	Alzheimer's Disease and Other Dementias	Parkinson's Disease	Nutritional Deficiencies	Malaria	Drowning	Interpersonal Violence	...	Diabetes Mellitus	Chronic Kidney Disease	Poisonings	Pro Ei Malnutr
0	Afghanistan	AFG	1990	2159	1116	371	2087	93	1370	1538	...	2108	3709	338	
1	Afghanistan	AFG	1991	2218	1136	374	2153	189	1391	2001	...	2120	3724	351	
2	Afghanistan	AFG	1992	2475	1162	378	2441	239	1514	2299	...	2153	3776	386	
3	Afghanistan	AFG	1993	2812	1187	384	2837	108	1687	2589	...	2195	3862	425	
4	Afghanistan	AFG	1994	3027	1211	391	3081	211	1809	2849	...	2231	3932	451	

5 rows × 34 columns

The above dataset contains **6120 rows and 34 columns**.

The above dataset contains 30 years data of 204 countries.

Adding a new column named **Total_no_of_deaths** which is summation of number of deaths of all the key causes of death given in this dataset. Sorting Top 40 total number of deaths among the observations:

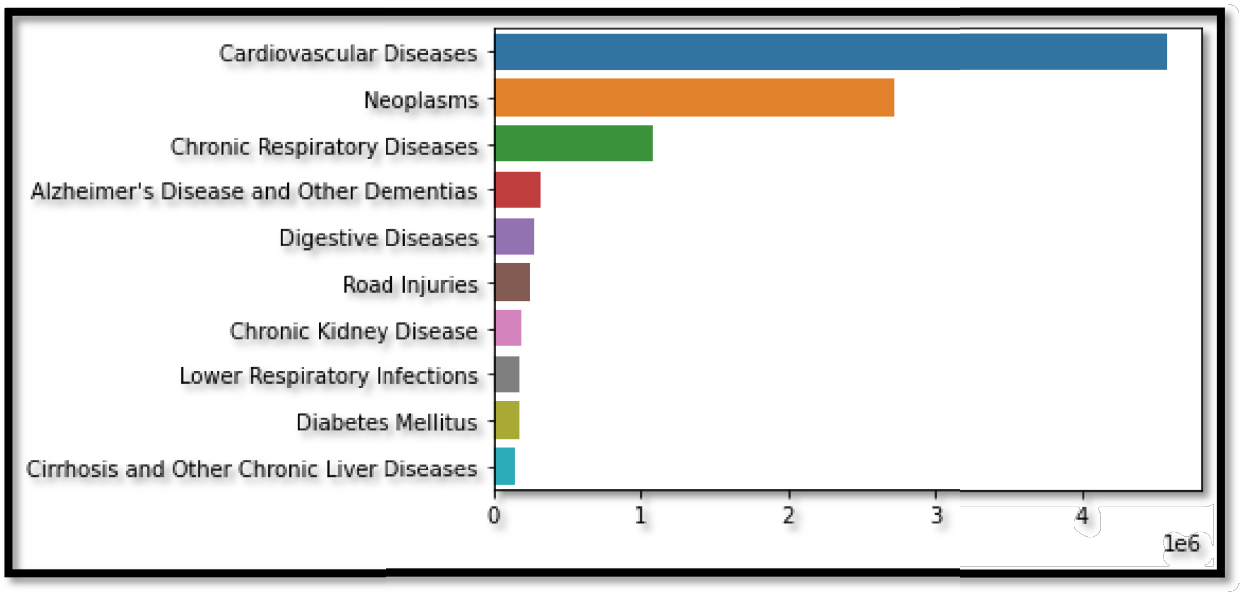
```
# Sorting Top 40 total number of deaths

Top40_Deaths = df.sort_values('Total_no_of_deaths',ascending=False)[:40][['Total_no_of_deaths','Country/Territory','Year']]

Top40_Deaths
```

	Total_no_of_deaths	Country/Territory	Year
1139	10442561	China	2019
1138	10163943	China	2018
1137	9978653	China	2017
1119	9814213	China	2016
1118	9591222	China	2015
1117	9503904	China	2014
1116	9411928	China	2013
1114	9366974	China	2011

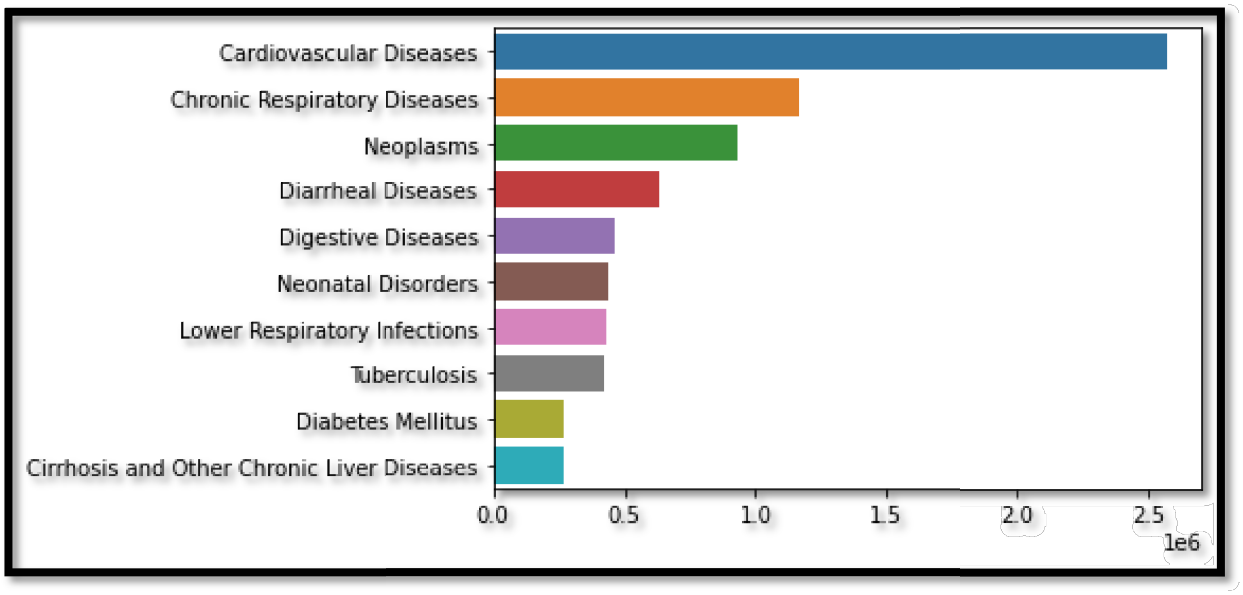
- From the Above table, the "top 40 - Total_no_of_Deaths" belongs to 'China' and 'India' because of the fact **"China" and "India"** are the **most populated countries** of the World

Let's see top 10 causes of death for countries like India and China:

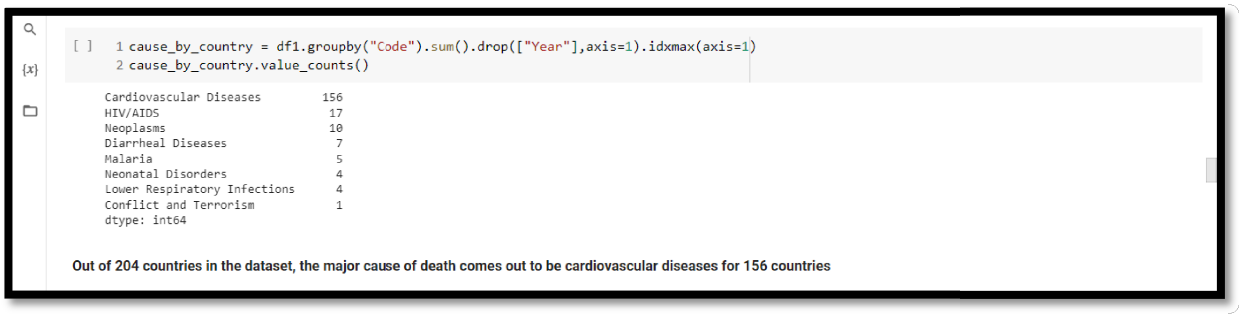
- **Top 10 causes of death in China**



- **Top 10 causes of Death in India**



- Clearly, it can be seen from above two bar plots, cardiovascular diseases are the leading cause of death globally**. Out of 204 countries in the dataset, cardiovascular diseases are a major cause of death for 156 countries**.

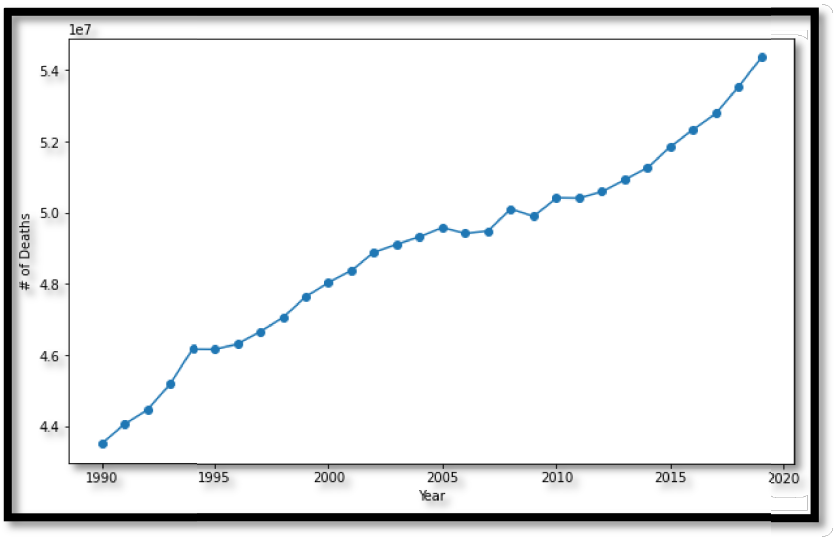


```
[ ]  1 cause_by_country = df1.groupby("Code").sum().drop(["Year"],axis=1).idxmax(axis=1)
     2 cause_by_country.value_counts()
```

```
Cardiovascular Diseases         156
HIV/AIDS                         17
Neoplasms                        10
Diarrheal Diseases                7
Malaria                           5
Neonatal Disorders                4
Lower Respiratory Infections      4
Conflict and Terrorism            1
dtype: int64
```

Out of 204 countries in the dataset, the major cause of death comes out to be cardiovascular diseases for 156 countries

Deadliest Year

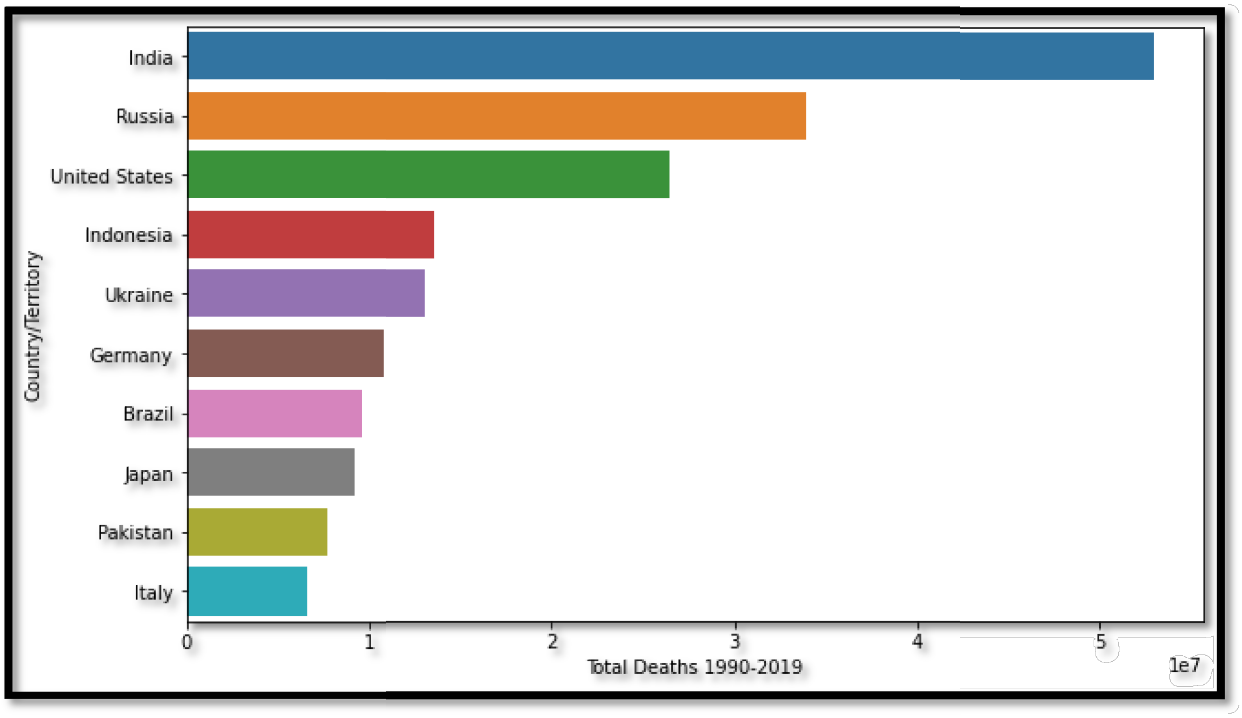


- **Deadliest year appears to be 2019** in terms of death counts; however it was due to the increase of the world population each year.

Cardiovascular diseases

Cardiovascular disease (CVD) is a term used to refer to the range of diseases which affect the heart and blood vessels. These include hypertension (high blood pressure); coronary heart disease (heart attack); cerebrovascular disease (stroke); heart failure; and other heart diseases.

Let's see top 10 countries which suffered from CVD the most:

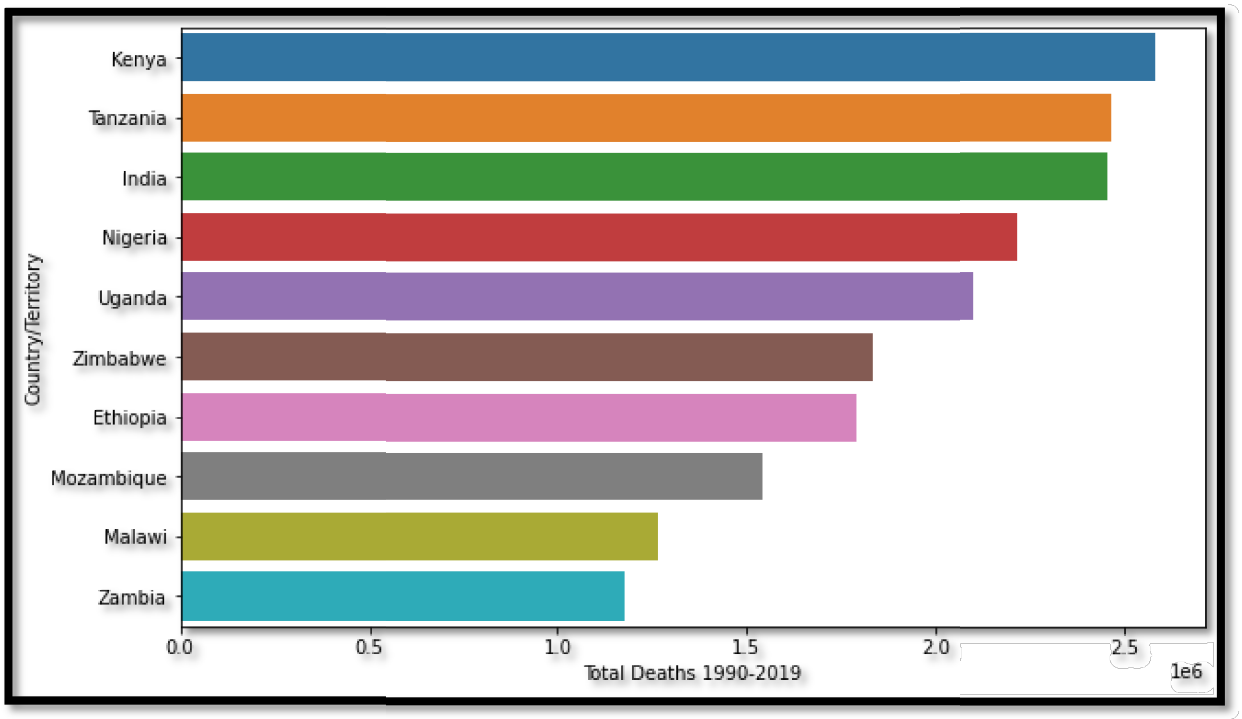


- **India has highest number of deaths due to cardiovascular diseases followed by Russia and US**

HIV/AIDS

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.

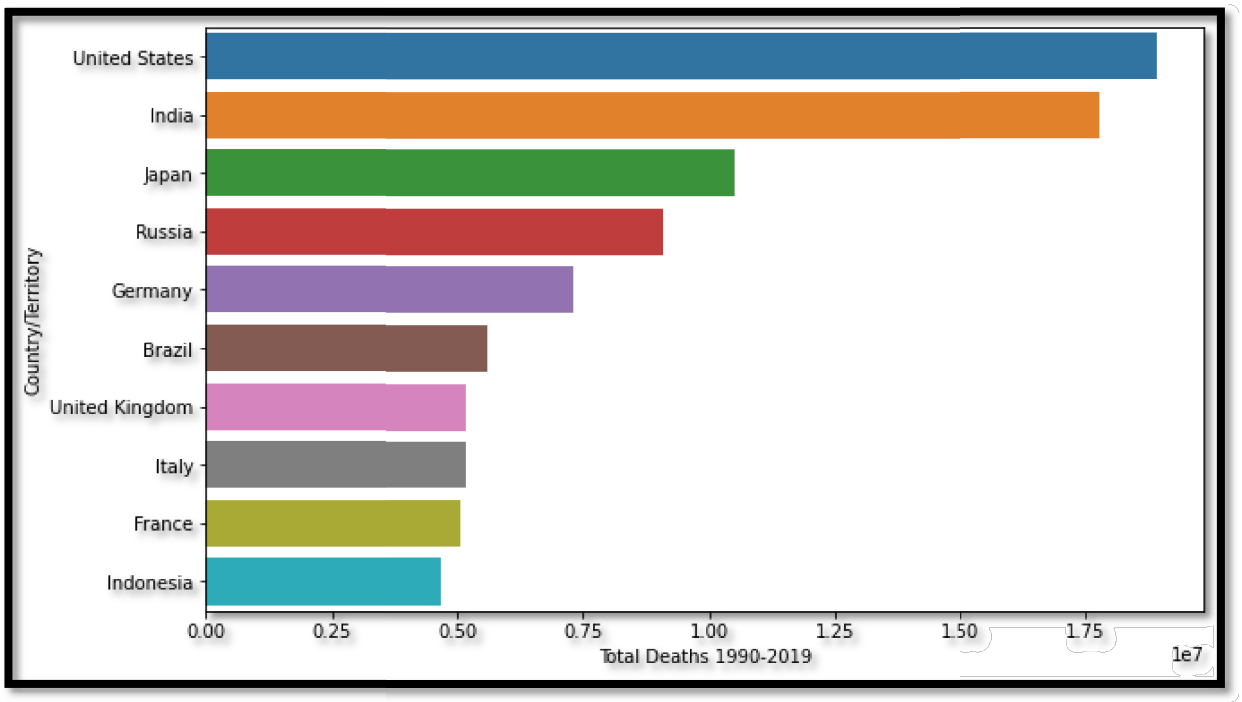


- Across Sub-Saharan Africa the deaths are much higher. Most countries in the South of the region had deaths greater than 1.5 million. In India, the deaths due to HIV/AIDS were nearly 2.5 million.

Neoplasm

An abnormal mass of tissue that forms when cells grow and divide more than they should or do not die when they should. Neoplasm may be benign (not cancer) or malignant (cancer). Cancers can arise in many parts of the body – leading to a range of cancer types – and in some cases spread to other parts of the body through the blood and lymph systems.

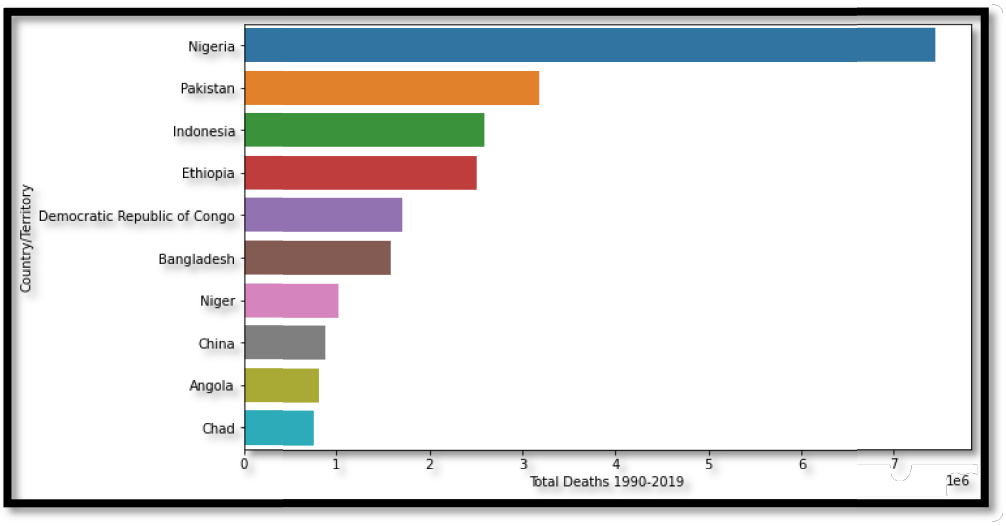
Let's see top 10 countries which suffered from Neoplasm the most:



- Cancer is a particularly common cause of death in richer countries where people are less likely to die of infectious diseases and causes of deaths that lead to very early deaths for people in poverty.

Diarrheal Diseases

Diarrheal diseases are caused primarily by viral and bacterial pathogens. They are particularly dominant at lower incomes where there is poor access to safe sanitation, drinking water and hygiene facilities. Diarrheal diseases are a leading cause of death in children. According to WHO, **each year diarrhoea kills around 525 000 children under five.**

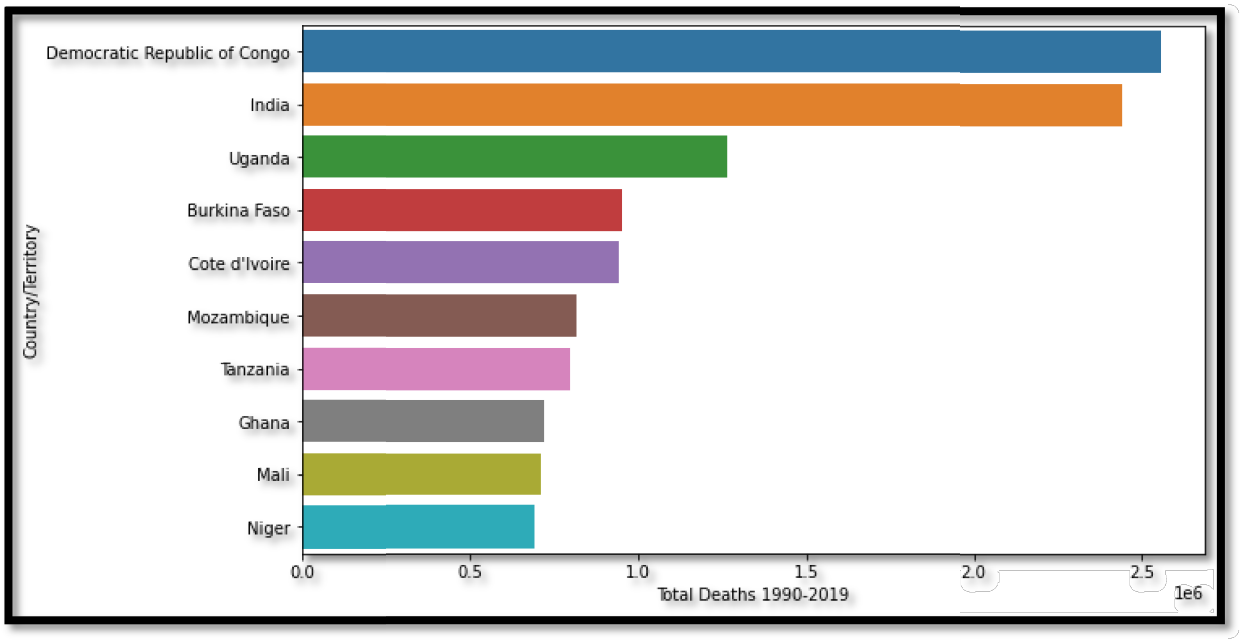


Diarrheal diseases were the fourth leading cause of death in 2019.

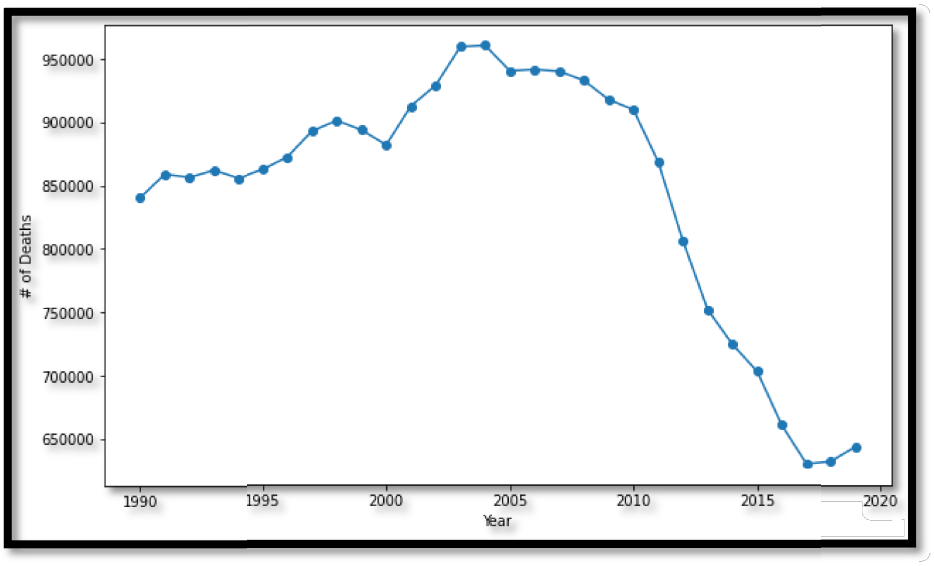
Overall, we see the highest mortality in Sub-Saharan Africa and South Asia.

Malaria

Malaria is a disease that is transmitted from person to person by infected mosquitoes. The bite of an infected Anopheles mosquito transmits a parasite that enters the victim's blood system and travels into the person's liver where the parasite reproduces. The parasite causes a high fever that involves shaking chills and pain. In the worst cases malaria leads to coma and death.



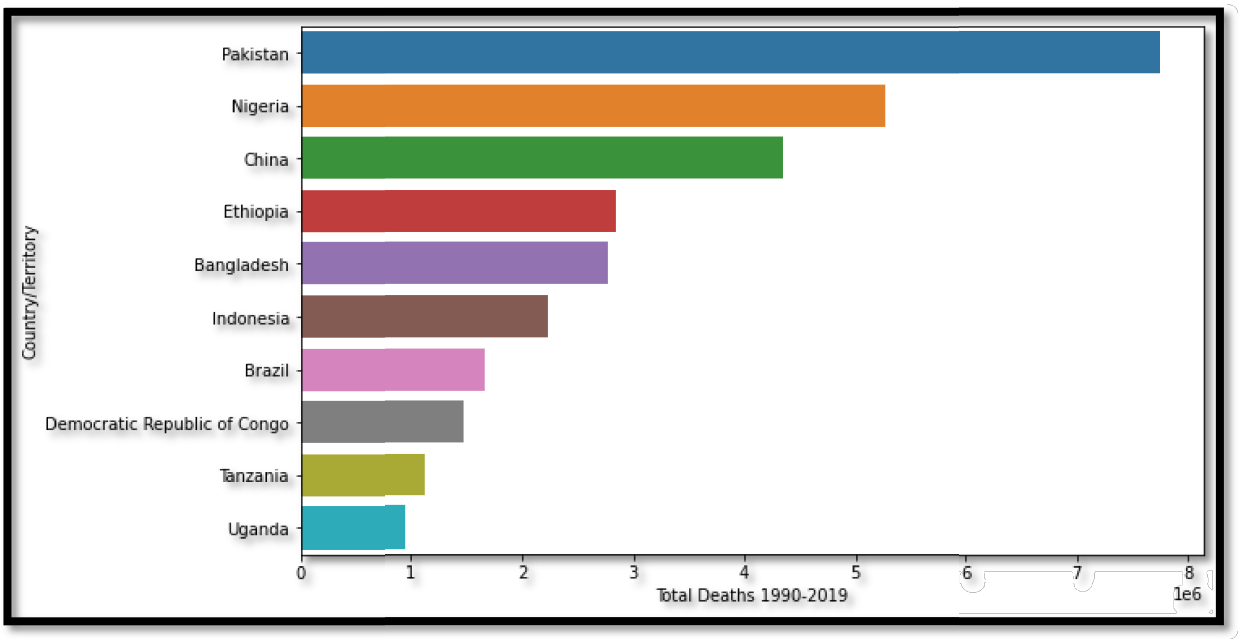
- Africa is the world region that is most affected by malaria in the last 30 years



- Majority of global deaths from malaria occurred on the African continent. But Africa is also the world region that has achieved most progress: since 2000, deaths have fallen significantly

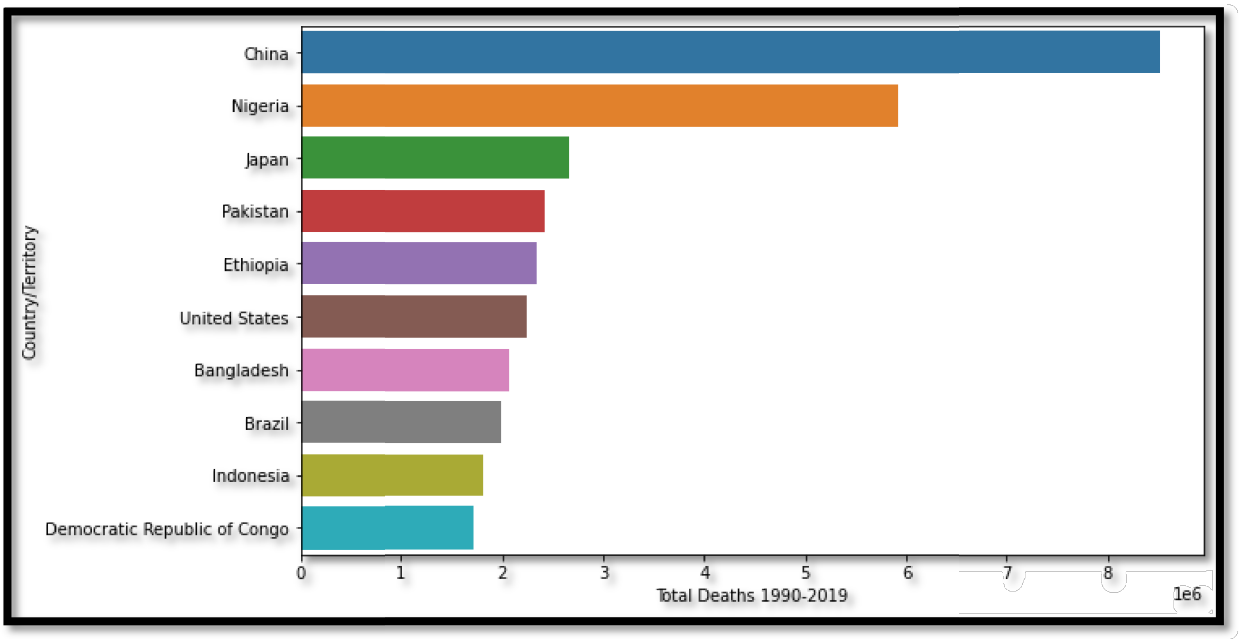
Neonatal Disorders

Its noun form, neonate, is another word for a newborn. Neonatal infections are primarily bacterial in origin, and include pneumonia, sepsis, and meningitis.



- The highest mortality are seen across Sub-Saharan Africa and Asia

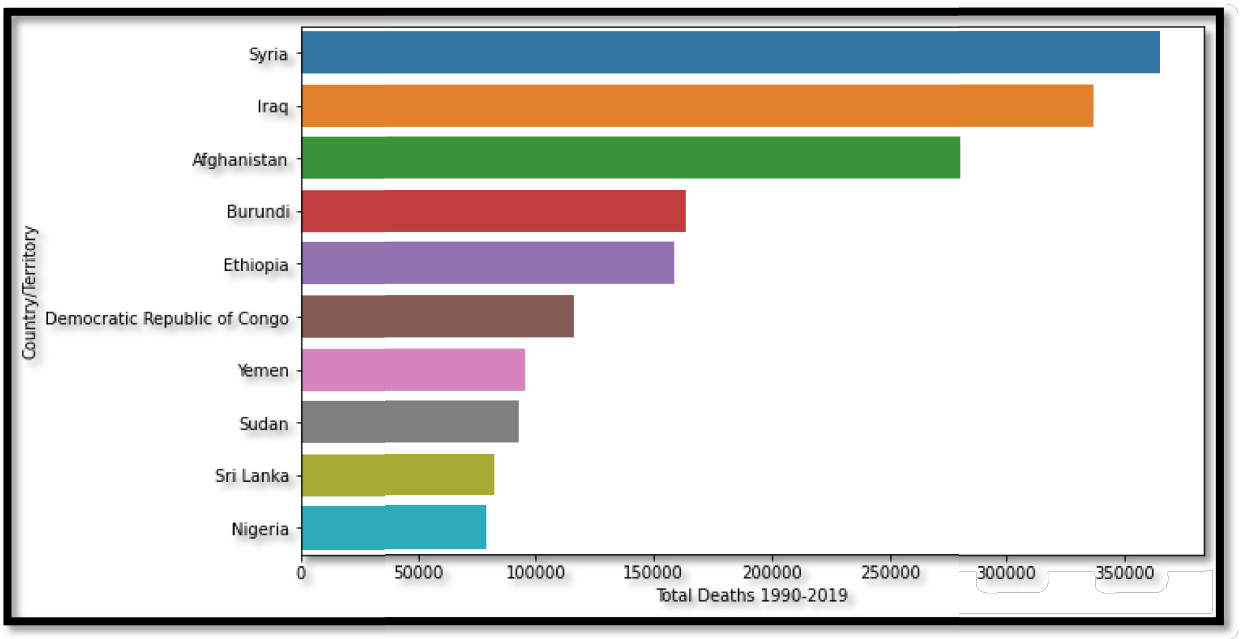
Lower Respiratory Infections



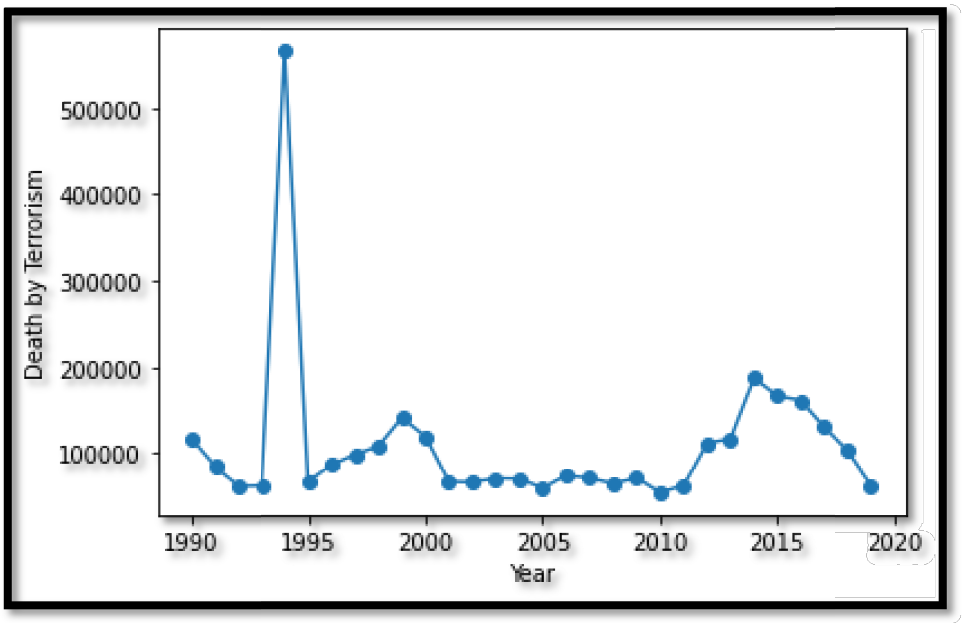
- China has highest number of deaths due to lower respiratory infections like Pneumonia
- USA also in top 6 in number of deaths due to lower respiratory infections

Conflict and Terrorism

Violent, criminal acts committed by individuals and/or groups who are inspired by, or associated with, designated foreign terrorist organizations or nations (state-sponsored).



- Most victims of terrorism die in the Middle East, Africa, and South Asia
- Most are affiliated with well-known terrorist groups, such as Islamic State, Taliban, Boko Haram, and Al-Shabaab



- 1994 was the year where the deadliest conflicts took place in the world. Bosnian War in Europe may have contributed to it significantly
- After 2014, deaths due to conflicts and terrorism has been decreasing which is a good sign for the world

CONCLUSION:

- Causes of death vary significantly between countries: non-communicable diseases dominate in rich countries, whereas infectious diseases remain high at lower incomes.
- 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.