Cause Of Death Project Report

I N Venkateswara Rao / Flip Robo Techanologies

INTRODUCTION

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. This is the topic of this entry. 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.

ABOUT THE DATASET

The data set has the following Data

- 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

Loading The Major Libraries Required

In [1]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import warnings
warnings.filterwarnings('ignore')
import seaborn as sns
import plotly.express as px
```

Loading the Dataset Given by the Customer:

In [2]:
```
ds = pd.read_csv(r"C:\Users\invra\Downloads\Cause-of-Death\cause_of_deaths dataset.csv")
ds
```

Out[2]:

	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	Ma
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	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
6115	Zimbabwe	ZWE	2015	1439	754	215	3019	2518	770	1302	...	3176	2108	381	
6116	Zimbabwe	ZWE	2016	1457	767	219	3056	2050	801	1342	...	3259	2160	393	
6117	Zimbabwe	ZWE	2017	1460	781	223	2990	2116	818	1363	...	3313	2196	398	
6118	Zimbabwe	ZWE	2018	1450	795	227	2918	2088	825	1396	...	3381	2240	400	
6119	Zimbabwe	ZWE	2019	1450	812	232	2884	2068	827	1434	...	3460	2292	405	

6120 rows × 34 columns

The Data set was loaded to a variable “ds” and it shows that the dataset has 6120 Rows and 34 columns of data

DATA SET INFO

```
In [4]: ds.columns
```

```
Out[4]: Index(['Country/Territory', 'Code', 'Year', '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'],
      dtype='object')
```

There are 34 columns and 6120 rows in this data set

All the 34 column names were displayed with the ds.columns command, as given the earlier dataset explanation there are all the columns mentioned. Ds.info shows the Non null values count and the types variables present in each column.

```
In [5]: ds.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 6120 entries, 0 to 6119
Data columns (total 34 columns):
 #   Column                                        Non-Null Count  Dtype
---  ------                                        --------------  -----
 0   Country/Territory                             6120 non-null   object
 1   Code                                          6120 non-null   object
 2   Year                                          6120 non-null   int64
 3   Meningitis                                    6120 non-null   int64
 4   Alzheimer's Disease and Other Dementias       6120 non-null   int64
 5   Parkinson's Disease                           6120 non-null   int64
 6   Nutritional Deficiencies                      6120 non-null   int64
 7   Malaria                                       6120 non-null   int64
 8   Drowning                                      6120 non-null   int64
 9   Interpersonal Violence                        6120 non-null   int64
 10  Maternal Disorders                            6120 non-null   int64
 11  HIV/AIDS                                      6120 non-null   int64
 12  Drug Use Disorders                            6120 non-null   int64
 13  Tuberculosis                                  6120 non-null   int64
 14  Cardiovascular Diseases                       6120 non-null   int64
 15  Lower Respiratory Infections                  6120 non-null   int64
 16  Neonatal Disorders                            6120 non-null   int64
 17  Alcohol Use Disorders                         6120 non-null   int64
 18  Self-harm                                     6120 non-null   int64
 19  Exposure to Forces of Nature                  6120 non-null   int64
 20  Diarrheal Diseases                            6120 non-null   int64
 21  Environmental Heat and Cold Exposure          6120 non-null   int64
 22  Neoplasms                                     6120 non-null   int64
 23  Conflict and Terrorism                        6120 non-null   int64
 24  Diabetes Mellitus                             6120 non-null   int64
 25  Chronic Kidney Disease                        6120 non-null   int64
 26  Poisonings                                    6120 non-null   int64
 27  Protein-Energy Malnutrition                   6120 non-null   int64
 28  Road Injuries                                 6120 non-null   int64
 29  Chronic Respiratory Diseases                  6120 non-null   int64
 30  Cirrhosis and Other Chronic Liver Diseases    6120 non-null   int64
 31  Digestive Diseases                            6120 non-null   int64
 32  Fire, Heat, and Hot Substances                6120 non-null   int64
 33  Acute Hepatitis                               6120 non-null   int64
dtypes: int64(32), object(2)
memory usage: 1.6+ MB
```

Here all the columns are int64 except the county/ territory and Code.

DATA ANALYSIS

Ds.describe() shows the data set statistical data, so that we can identify if there is any abnormal or huge diffeern ces in the data

In [6]:
```
ds.describe().T.style.bar(subset=["mean"],
                          color='#205ff2').background_gradient(subset=['std'],
                                                               cmap='Reds').background_gradient(subset=['50%'],
                                                                                                cmap='coolwarm')
```

Out[6]:

	count	mean	std	min	25%	50%	75%	max
Year	6120.000000	2004.500000	8.656149	1990.000000	1997.000000	2004.500000	2012.000000	2019.000000
Meningitis	6120.000000	1719.701307	6672.006930	0.000000	15.000000	109.000000	847.250000	98358.000000
Alzheimer's Disease and Other Dementias	6120.000000	4864.189379	18220.659072	0.000000	90.000000	666.500000	2456.250000	320715.000000
Parkinson's Disease	6120.000000	1173.169118	4616.156238	0.000000	27.000000	164.000000	609.250000	76990.000000
Nutritional Deficiencies	6120.000000	2253.600000	10483.633601	0.000000	9.000000	119.000000	1167.250000	268223.000000
Malaria	6120.000000	4140.960131	18427.753137	0.000000	0.000000	0.000000	393.000000	280604.000000
Drowning	6120.000000	1683.333170	8877.018366	0.000000	34.000000	177.000000	698.000000	153773.000000
Interpersonal Violence	6120.000000	2083.797222	6917.006075	0.000000	40.000000	265.000000	877.000000	69640.000000
Maternal Disorders	6120.000000	1262.589216	6057.973183	0.000000	5.000000	54.000000	734.000000	107929.000000
HIV/AIDS	6120.000000	5941.898529	21011.962487	0.000000	11.000000	136.000000	1879.000000	305491.000000
Drug Use Disorders	6120.000000	434.006699	2898.761628	0.000000	3.000000	20.000000	129.000000	65717.000000
Tuberculosis	6120.000000	7491.928595	39549.977578	0.000000	35.000000	417.000000	2924.250000	657515.000000
Cardiovascular Diseases	6120.000000	73160.454575	291577.537794	4.000000	2028.000000	11742.000000	42546.500000	4584273.000000
Lower Respiratory Infections	6120.000000	13687.914706	48031.720009	0.000000	345.000000	2126.500000	10161.250000	690913.000000
Neonatal Disorders	6120.000000	12558.942647	56058.366412	0.000000	131.000000	916.000000	7419.750000	852761.000000
Alcohol Use Disorders	6120.000000	787.421242	3545.823616	0.000000	9.000000	80.000000	316.000000	55200.000000
Self-harm	6120.000000	3874.825327	18425.616418	0.000000	94.000000	533.000000	1882.250000	220357.000000
Exposure to Forces of Nature	6120.000000	243.485621	4717.104377	0.000000	0.000000	0.000000	12.000000	222641.000000
Diarrheal Diseases	6120.000000	10822.795425	65416.174485	0.000000	20.000000	296.500000	3946.750000	1119477.000000
Environmental Heat and Cold Exposure	6120.000000	292.295915	1704.466356	0.000000	2.000000	21.000000	109.000000	29048.000000
Neoplasms	6120.000000	37542.244771	161558.365445	1.000000	809.750000	5629.500000	20147.750000	2716551.000000
Conflict and Terrorism	6120.000000	538.243954	7033.308187	0.000000	0.000000	0.000000	23.000000	503532.000000
Diabetes Mellitus	6120.000000	5138.704575	16773.081040	1.000000	236.000000	1087.000000	2954.000000	273089.000000
Chronic Kidney Disease	6120.000000	4724.132680	16470.429969	0.000000	145.750000	822.000000	2922.500000	222922.000000
Poisonings	6120.000000	425.013399	2022.640521	0.000000	6.000000	52.500000	254.000000	30883.000000
Protein-Energy Malnutrition	6120.000000	1965.994281	8255.999063	0.000000	5.000000	92.000000	1042.500000	202241.000000
Road Injuries	6120.000000	5930.795588	24097.784291	0.000000	174.750000	966.500000	3435.250000	329237.000000
Chronic Respiratory Diseases	6120.000000	17092.374837	105157.179839	1.000000	289.000000	1689.000000	5249.750000	1366039.000000
Cirrhosis and Other Chronic Liver Diseases	6120.000000	6124.072059	20688.118580	0.000000	154.000000	1210.000000	3547.250000	270037.000000
Digestive Diseases	6120.000000	10725.267157	37228.051096	0.000000	284.000000	2185.000000	6080.000000	464914.000000
Fire, Heat, and Hot Substances	6120.000000	588.711438	2128.595120	0.000000	17.000000	126.000000	450.000000	25876.000000
Acute Hepatitis	6120.000000	618.429902	4186.023497	0.000000	2.000000	15.000000	160.000000	64305.000000

Here we found that the Exposure to natural forces is having the least mean and the cardio vascular disease deaths is having the highest and much more than the any other cause. Neoplasms is the second leading cause of death all over the world.

CHECKING FOR NAN VALUES

Checking for NaN values

In [7]:
```
pd.DataFrame(ds.isnull().sum(), columns=["Null values "]).rename_axis("Column Name")
```

Out[7]:

Column Name	Null values
Country/Territory	0
Code	0
Year	0
Meningitis	0
Alzheimer's Disease and Other Dementias	0
Parkinson's Disease	0
Nutritional Deficiencies	0
Malaria	0
Drowning	0
Interpersonal Violence	0
Maternal Disorders	0
HIV/AIDS	0
Drug Use Disorders	0
Tuberculosis	0
Cardiovascular Diseases	0
Lower Respiratory Infections	0
Neonatal Disorders	0
Alcohol Use Disorders	0
Self-harm	0
Exposure to Forces of Nature	0

On checking for the Null and nan values its found that there are no null values found in the data set

In [8]:
```
ds.isin(['NAN','NA','N/A','-',' ','?','--']).sum().any()
```

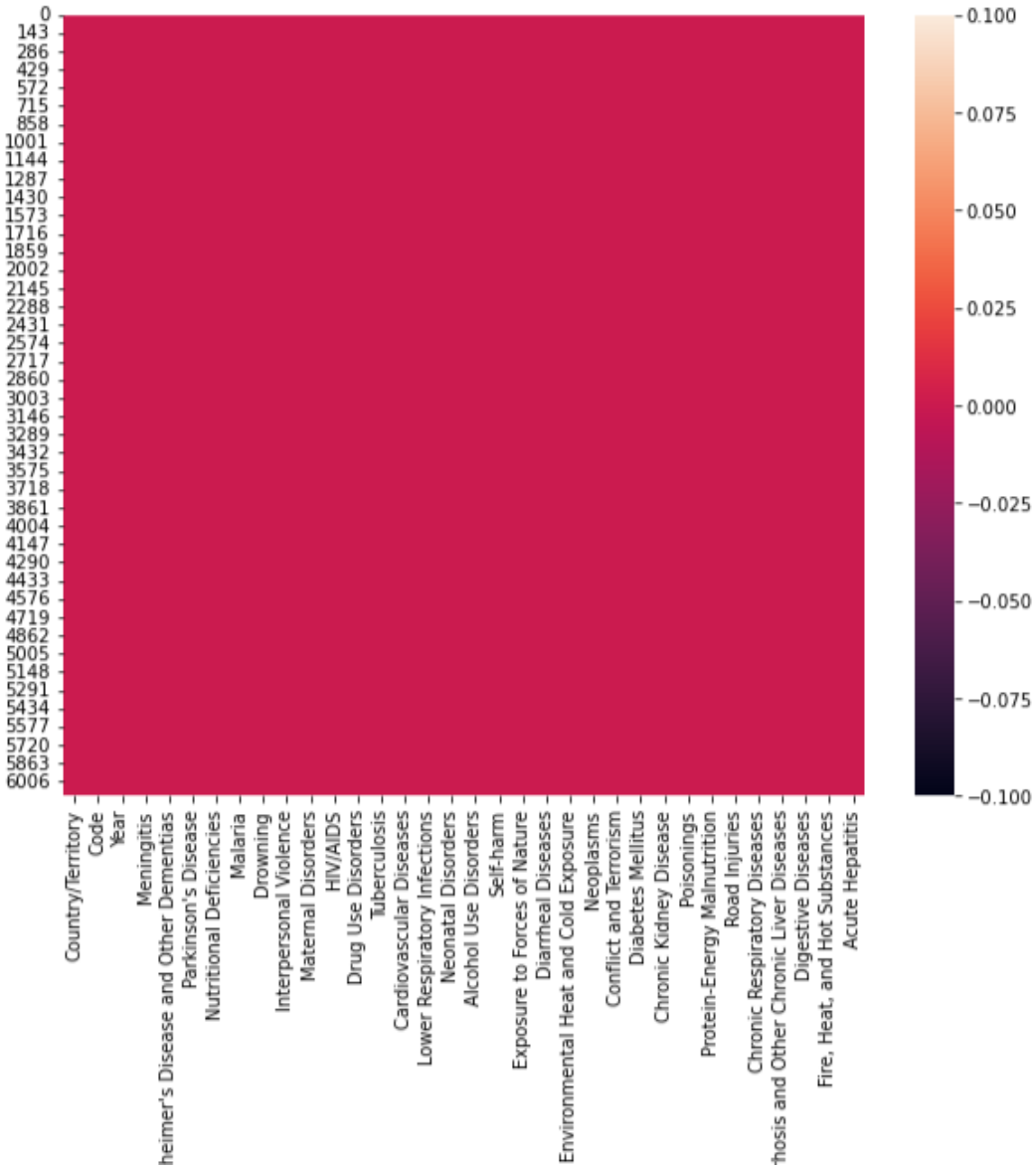
Out[8]:
```
False
```

In [9]:
```
#Checking Null values through heatmap
plt.figure(figsize=(10,8))
sns.heatmap(ds.isnull())
```

NULL VALUES HEAT MAP

```
In [9]: #Checking Null values through heatmap
        plt.figure(figsize=(10,8))
        sns.heatmap(ds.isnull())
```

```
Out[9]: <AxesSubplot:>
```



Even cross checked the same with Heat map, so that we can visually check that there are no null values

DATA ANALYSIS

```
In [10]:
ds['Country/Territory'].unique()
Out[10]:
array(['Afghanistan', 'Albania', 'Algeria', 'American Samoa', 'Andorra', 'Angola',
'Antigua and Barbuda', 'Argentina', 'Armenia', 'Australia', 'Austria', 'Azerbaijan',
'Bahamas', 'Bahrain', 'Bangladesh', 'Barbados', 'Belarus', 'Belgium', 'Belize',
'Benin', 'Bermuda', 'Bhutan', 'Bolivia', 'Bosnia and Herzegovina', 'Botswana',
'Brazil', 'Brunei', 'Bulgaria', 'Burkina Faso', 'Burundi', 'Cambodia', 'Cameroon',
'Canada', 'Cape Verde', 'Central African Republic', 'Chad', 'Chile', 'China',
'Colombia', 'Comoros', 'Congo', 'Cook Islands', 'Costa Rica', "Cote d'Ivoire",
'Croatia', 'Cuba', 'Cyprus', 'Czechia', 'Democratic Republic of Congo', 'Denmark',
'Djibouti', 'Dominica', 'Dominican Republic', 'Ecuador', 'Egypt', 'El Salvador',
'Equatorial Guinea', 'Eritrea', 'Estonia', 'Eswatini', 'Ethiopia', 'Fiji', 'Finland',
'France', 'Gabon', 'Gambia', 'Georgia', 'Germany', 'Ghana', 'Greece', 'Greenland',
'Grenada', 'Guam', 'Guatemala', 'Guinea', 'Guinea-Bissau', 'Guyana', 'Haiti',
'Honduras', 'Hungary', 'Iceland', 'India', 'Indonesia', 'Iran', 'Iraq', 'Ireland',
'Israel', 'Italy', 'Jamaica', 'Japan', 'Jordan', 'Kazakhstan', 'Kenya', 'Kiribati',
'Kuwait', 'Kyrgyzstan', 'Laos', 'Latvia', 'Lebanon', 'Lesotho', 'Liberia', 'Libya',
'Lithuania', 'Luxembourg', 'Madagascar', 'Malawi', 'Malaysia', 'Maldives', 'Mali',
'Malta', 'Marshall Islands', 'Mauritania', 'Mauritius', 'Mexico', 'Micronesia',
'Moldova', 'Monaco', 'Mongolia', 'Montenegro', 'Morocco', 'Mozambique', 'Myanmar',
'Namibia', 'Nauru', 'Nepal', 'Netherlands', 'New Zealand', 'Nicaragua', 'Niger',
'Nigeria', 'Niue', 'North Korea', 'North Macedonia', 'Northern Mariana Islands',
'Norway', 'Oman', 'Pakistan', 'Palau', 'Palestine', 'Panama', 'Papua New Guinea',
'Paraguay', 'Peru', 'Philippines', 'Poland', 'Portugal', 'Puerto Rico', 'Qatar',
'Romania', 'Russia', 'Rwanda', 'Saint Kitts and Nevis', 'Saint Lucia', 'Saint Vincent
and the Grenadines', 'Samoa', 'San Marino', 'Sao Tome and Principe', 'Saudi Arabia',
'Senegal', 'Serbia', 'Seychelles', 'Sierra Leone', 'Singapore', 'Slovakia',
'Slovenia', 'Solomon Islands', 'Somalia', 'South Africa', 'South Korea', 'South
Sudan', 'Spain', 'Sri Lanka', 'Sudan', 'Suriname', 'Sweden', 'Switzerland', 'Syria',
'Taiwan', 'Tajikistan', 'Tanzania', 'Thailand', 'Timor', 'Togo', 'Tokelau', 'Tonga',
'Trinidad and Tobago', 'Tunisia', 'Turkey', 'Turkmenistan', 'Tuvalu', 'Uganda',
'Ukraine', 'United Arab Emirates', 'United Kingdom', 'United States', 'United States
Virgin Islands', 'Uruguay', 'Uzbekistan', 'Vanuatu', 'Venezuela', 'Vietnam', 'Yemen',
'Zambia', 'Zimbabwe'], dtype=object)
```

We checked for total unique country/territories present in the data set and as well as the unique number of years for which the data was given.

```
In [13]:
ds['Year'].unique()
Out[13]:
array([1990, 1991, 1992, 1993, 1994, 1995, 1996, 1997, 1998, 1999, 2000, 2001, 2002,
2003, 2004, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016,
2017, 2018, 2019], dtype=int64)
```

Its found that there are 30 years of data from 1991 to 2019 and data covered for 204 countries

DATA ANALYSIS

```
In [15]:
# Drop Code feature as it same as coutries so will be multicollinarity
ds.drop('Code',axis=1,inplace=True)
In [16]:
ds.head(5)
Out[16]:
```

	Country/Territory	Year	Meningitis	Alzheimer's Disease and Other Dementias	Parkinson's Disease	Nutritional Deficiencies	Malaria	Drowning	Interpersonal Violence	Maternal Disorders	..	Diabetes Mellitus	Chronic Kidney Disease	Poisonings	Protein-Energy Malnutrition
0	Afghanistan	1990	2159	1116	371	2087	93	1370	1538	2655	..	2108	3709	338	
1	Afghanistan	1991	2218	1136	374	2153	189	1391	2001	2885	..	2120	3724	351	
2	Afghanistan	1992	2475	1162	378	2441	239	1514	2299	3315	..	2153	3776	386	
3	Afghanistan	1993	2812	1187	384	2837	108	1687	2589	3671	..	2195	3862	425	
4	Afghanistan	1994	3027	1211	391	3081	211	1809	2849	3863	..	2231	3932	451	

5 rows × 33 columns

Dropped the Code column as its code sign for the Country/Territories column in the dataset. To avoid multi collinearity the column was removed using dop command as shown above.

Adding a new column named TOTAL for adding deaths due to all reasons

```
[17]: ds["Total"] = 0
      ds_cols = ds.columns
      ds_cols=ds_cols.drop(['Year'])
      ds["Total"] = ds[ds_cols].sum(axis=1)

[18]: ds.head(5)
```

t[18]:

	ınal ıies	Malaria	Drowning	Interpersonal Violence	Maternal Disorders	...	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	Total
	087	93	1370	1538	2655	...	3709	338	2054	4154	5945	2673	5005	323	2985	147971
	153	189	1391	2001	2885	...	3724	351	2119	4472	6050	2728	5120	332	3092	156844
	441	239	1514	2299	3315	...	3776	386	2404	5106	6223	2830	5335	360	3325	169156
	837	108	1687	2589	3671	...	3862	425	2797	5681	6445	2943	5568	396	3601	182230
	081	211	1809	2849	3863	...	3932	451	3038	6001	6664	3027	5739	420	3816	194795

Added a new column for total death count summing all the causes in a year at the last column for easy reference.

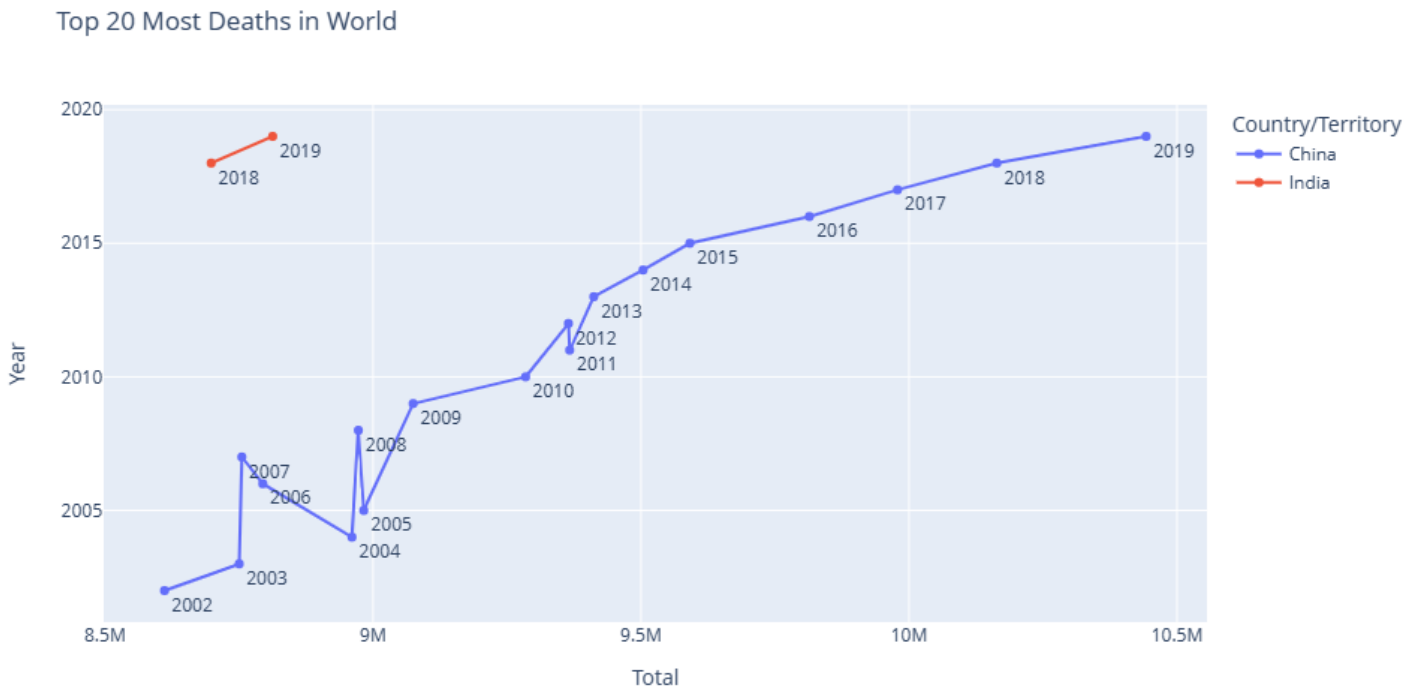
DATA ANALYSIS

We can get the top 20 daeths in the world by following code

Top 20 Most Death counts in World

```
In [19]: ds.sort_values(by=["Total"],inplace=True)
         fig = px.line(ds.tail(20), x="Total", y="Year",color="Country/Territory", text="Year",title="Top 20 Most Deaths in World")
         fig.update_traces(textposition="bottom right")
         fig.show()
```

The graphical representation is as follows:



The Top 20 deaths are from India and China Only. Its mainly due the dense population these countries too.

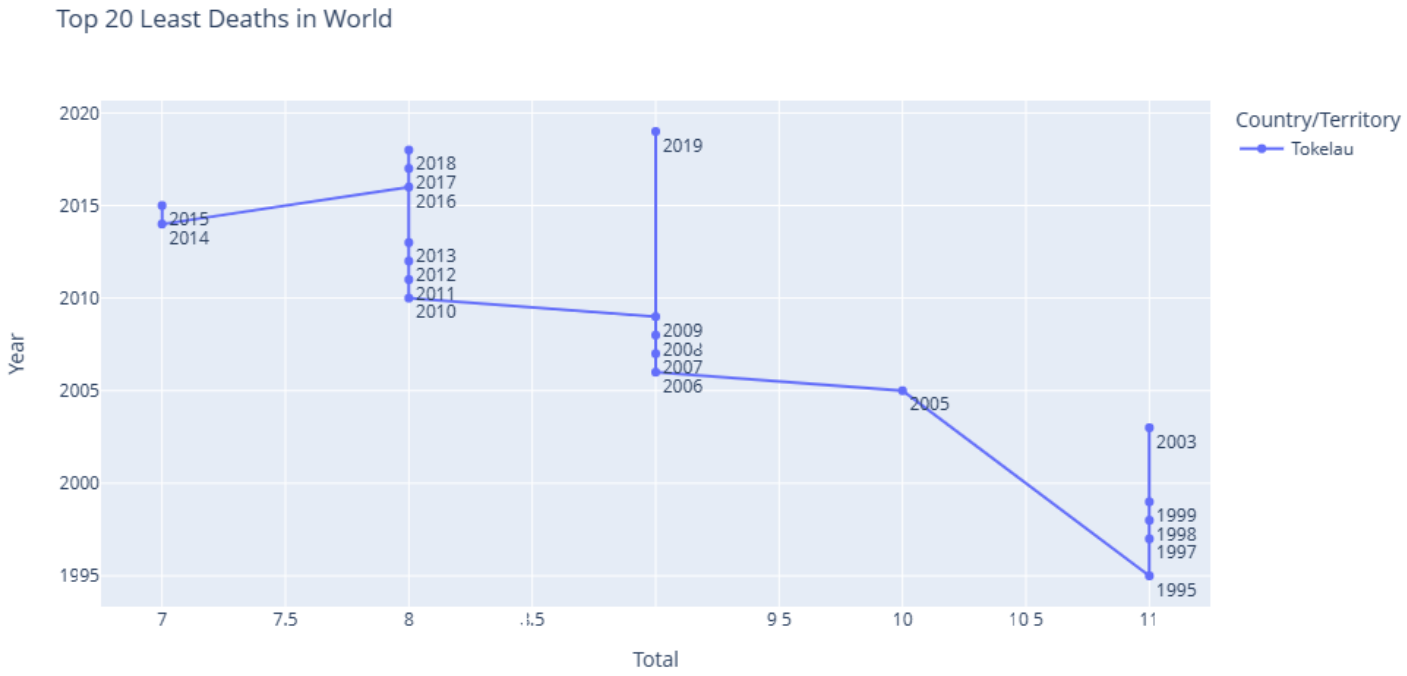
DATA ANALYSIS

Top 20 Least Death counts in World

```
In [20]: fig = px.line(ds.head(20), x="Total", y="Year", color="Country/Territory", text="Year",title="Top 20 Least Deaths in World")
         fig.update_traces(textposition="bottom right")
         fig.show()
```

The lowest death count in the world is present in Tokelau

The Graphicsl representation is given below:

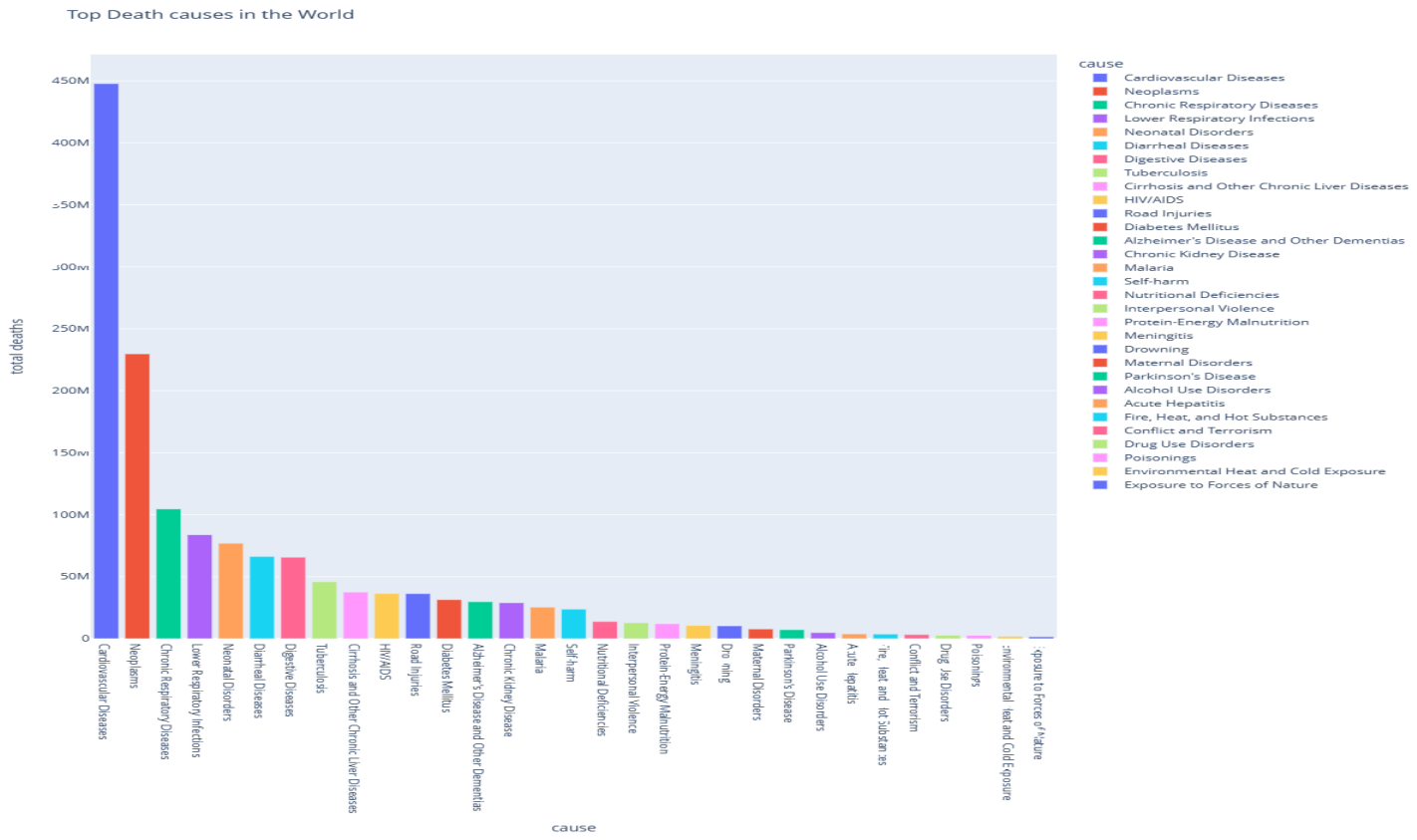


DATA ANALYSIS

The following code was used to get the top 5 causes for the highest death counts

Top 5 Causes of Death¶

```
In [21]:
# which top 5 diseases causes most of deaths ?
ds_new =ds.sum(axis=0).reset_index()
ds_new=ds_new.rename(columns={'index': 'cause', 0:'total deaths'})
ds_new.drop(ds_new.index[[0,1 , -1]],inplace=True)
ds_new.convert_dtypes()
ds_new = ds_new.sort_values(by=["total deaths"],ascending=False)
In [22]:
fig = px.bar(ds_new, x='cause', y='total deaths',
             hover_data=['cause', 'total deaths'], color='cause',
             labels={'deaths':'causes of deaths'},width=1200, height=1200, title="Top
Death causes in the World")
fig.update_traces(textfont_size=12, textangle=0, textposition="outside",
cliponaxis=False)
fig.show()
```



DATA ANALYSIS

As you can see here, the top causes of deaths can be seen in the above graph however the rate of deaths due to cardiovascular diseases are on the top.

Top 5 Causes for High individual Death counts are:

1. Cardiovascular Diseases
2. Neoplasms
3. Chronic Respiratory Diseases
4. Lower Respiratory Infections
5. Neonatal Disorders

Last five Causes for Highest Death Counts Are:

1. Exposure to Forces of nature
2. Environmental Heat and cold exposure
3. Poisonings
4. Drug use disorders
5. Conflict and terrorism

Top 5 Most death causes Impacts which countries in which year¶

In [23]:

```
#which countries suffered from top death causes and in which years
ds2 = ds.iloc[0:0]
for itr in ds_new.head().cause:
    ds2= ds2.append([ds.sort_values(by=[itr] , ascending=False).head(1)])
```

In [24]:

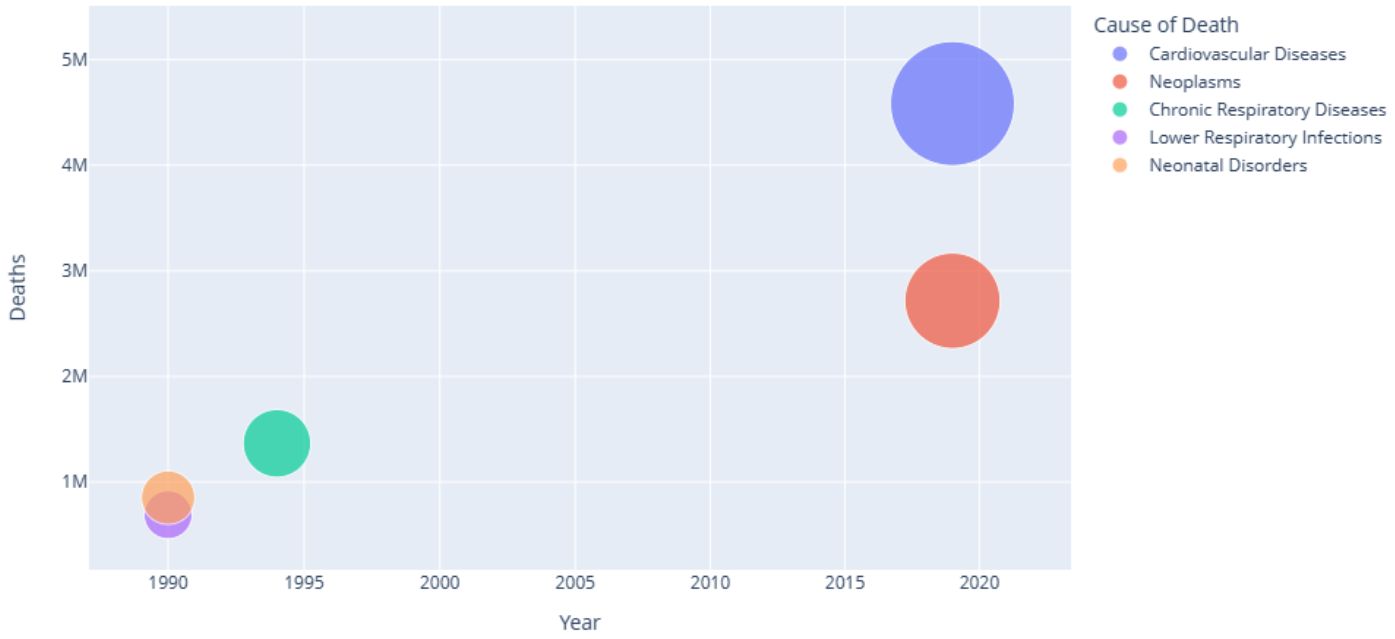
```
ds2
```

Out[24]:

	Country/Territory	Year	Meningitis	Alzheimer's Disease and Other Dementias	Parkinson's Disease	Nutritional Deficiencies	Malaria	Drowning	Interpersonal Violence	Maternal Disorders	...	Chronic Kidney Disease	Poisonings	Protein-Energy Malnutrition
1139	China	2019	6465	320715	76990	16863	0	56524	11970	1537	...	196726	27084	13099
1139	China	2019	6465	320715	76990	16863	0	56524	11970	1537	...	196726	27084	13099
1124	China	1994	31752	103621	34680	32825	1146	138993	38146	14144	...	101945	20528	31078
2436	India	1990	98358	24195	13419	268223	162369	88688	38720	106173	...	92805	7257	202241
2436	India	1990	98358	24195	13419	268223	162369	88688	38720	106173	...	92805	7257	202241

DATA ANALYSIS

In [25]:
```
i=0
data = []
while(i<5):
    ls = [ds2.iloc[i]["Country/Territory"], ds2.iloc[i]["Year"],ds_new.cause.iloc[i],ds2.iloc[i][ds_new.cause.iloc[i]]]
    data.append(ls)
    i+=1
ds_rank = pd.DataFrame(data, columns=['Country', 'Year','Cause of Death', "Deaths"])
ds_rank
fig = px.scatter(ds_rank, x="Year", y="Deaths",
                 size="Deaths", color="Cause of Death",
                     hover_name="Country", log_x=True, size_max=60)
fig.show()
```



As we can see that the Top 5 Causes of death affects which country in which year. The Top cause of death cardiovascular diseases affects China in 2019 with a lot of deaths. Similarly, the Rest of the diseases and their impacts can be seen in the above chart.

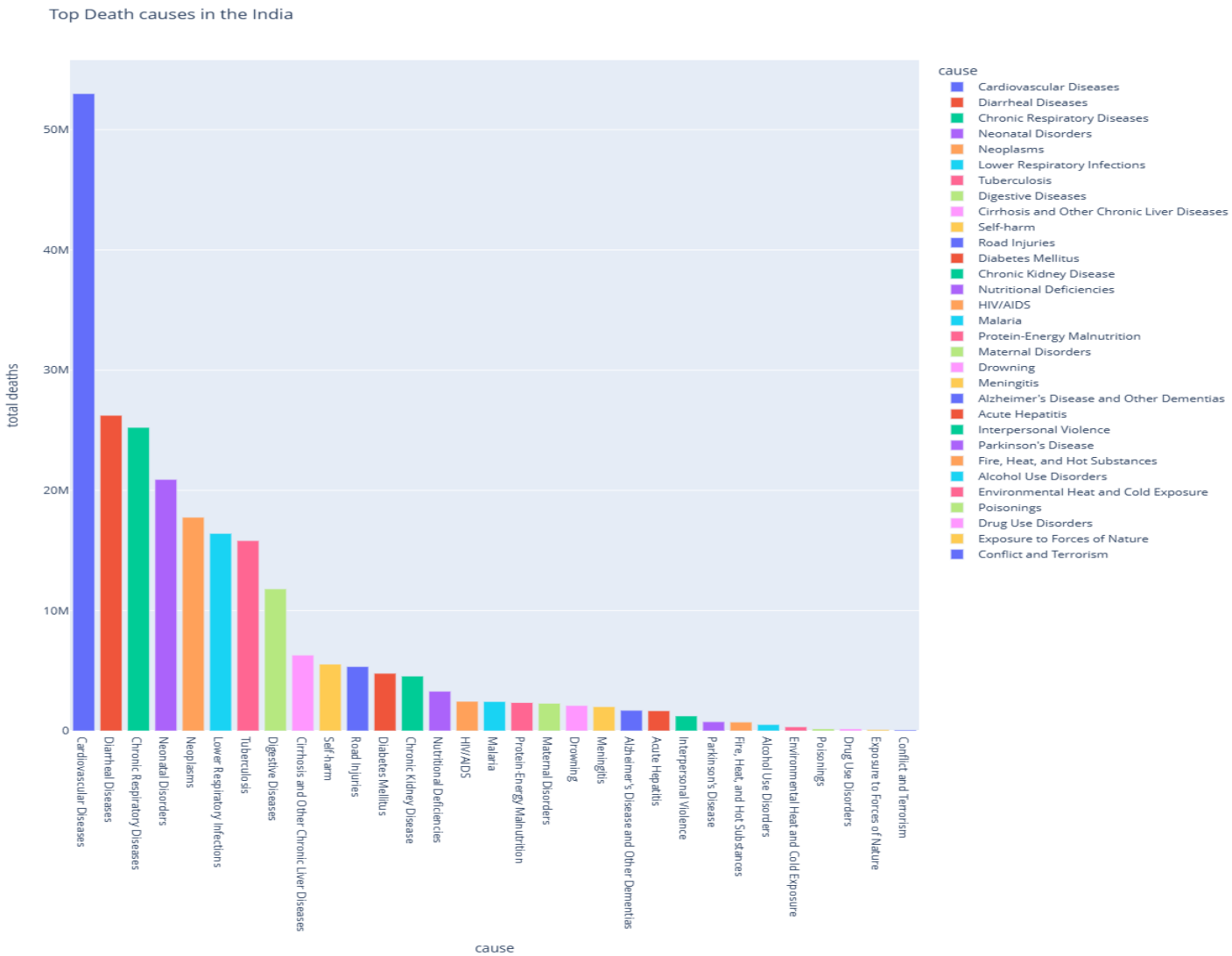
TOP COUNTRIES DISEASES TREND

```
In [38]: India1 =India.sum(axis=0).reset_index()
         India1 =India1.rename(columns={'index': 'cause', 0:'total deaths'})
         India1.drop(India1.index[[0,1 , -1]],inplace=True)
         India1.convert_dtypes()
         India1 = India1.sort_values(by=["total deaths"],ascending=False)



         fig = px.bar(India1, x='cause', y='total deaths',
                      hover_data=['cause', 'total deaths'], color='cause',
                      labels={'deaths':'causes of deaths'},width=1200, height=1200, title="Top Death causes in the India")
         fig.update_traces(textfont_size=12, textangle=0, textposition="outside", cliponaxis=False)
         fig.show()
```

From the above code we got the top diseases causing most deaths in india

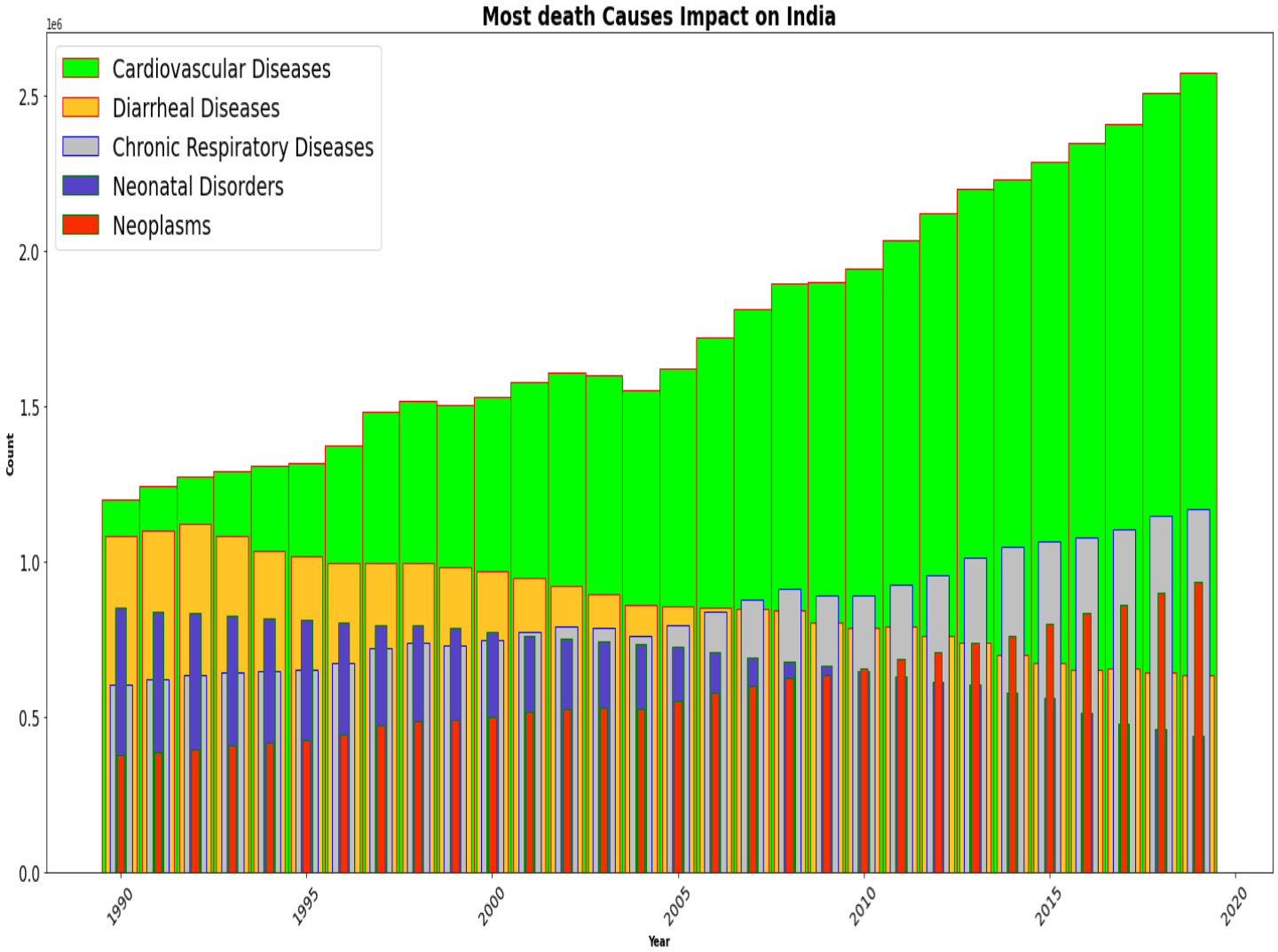


INDIA'S DATA

In India Cardiovascular diseases are the highest cause for deaths and followed by Diarrheal Diseases, Chronic respiratory diseases, Neonatal Disorders and Neoplasms

And the least causes are Terrorism followed by Exposure to forces of nature, drug use disorders, poisonings, Environmental Heat and cold exposure.

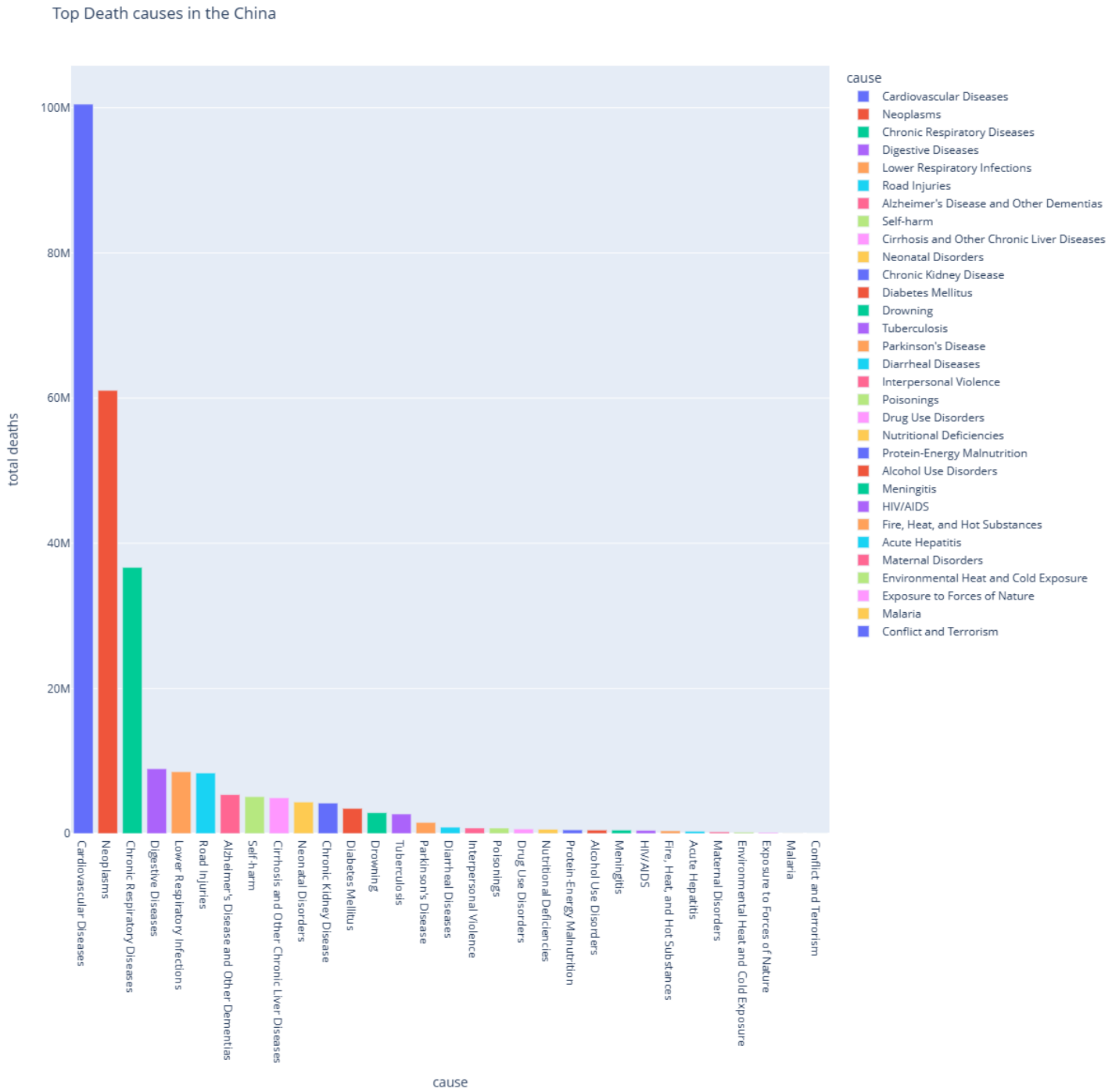
And the trend of the last 30 years for the highest effecting diseases are bas follows



INDIA'S DATA

The Above graph shows how thefive causes of death in india is varying in the past 30 years top that
1. Cardiovascular Diseases are increasing in trend and rapidly increasing
2. Diarrheal Diseases are decreasing in trend
3. Chronic Respiratory diseases are also in increasing trend
4. Neonatal Disorders deaths are decreasing in trend
5. Neoplasms deaths are steadily increasing

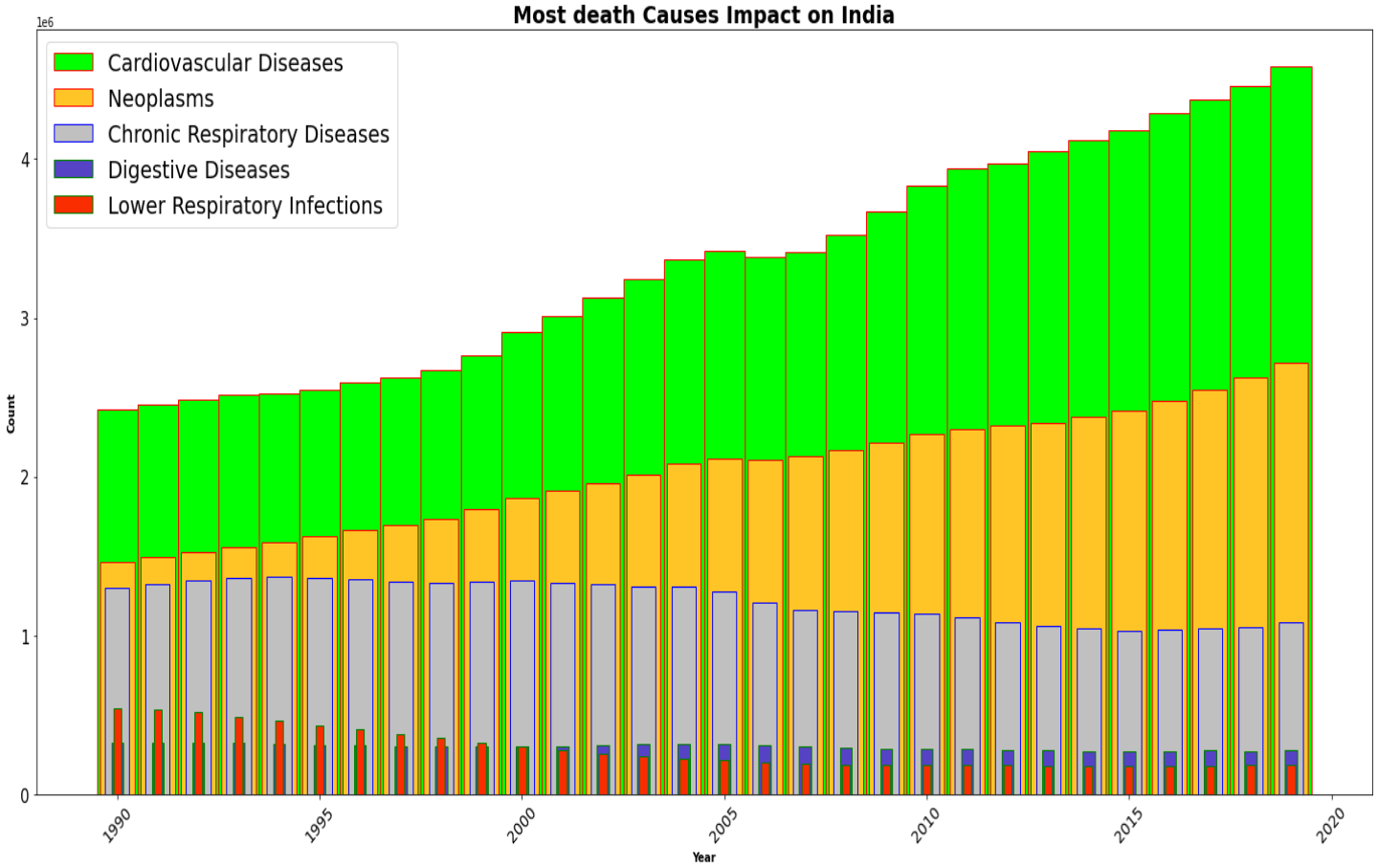
CHINA'S DATA


Top Death causes in the China
total deaths
100M
80M
60M
40M
20M
0
cause
Cardiovascular Diseases
Neoplasms
Chronic Respiratory Diseases
Digestive Diseases
Lower Respiratory Infections
Road Injuries
Alzheimer's Disease and Other Dementias
Self-harm
Cirrhosis and Other Chronic Liver Diseases
Neonatal Disorders
Chronic Kidney Disease
Diabetes Mellitus
Drowning
Tuberculosis
Parkinson's Disease
Diarrheal Diseases
Interpersonal Violence
Poisonings
Drug Use Disorders
Nutritional Deficiencies
Protein-Energy Malnutrition
Alcohol Use Disorders
Meningitis
HIV/AIDS
Fire, Heat, and Hot Substances
Acute Hepatitis
Maternal Disorders
Environmental Heat and Cold Exposure
Exposure to Forces of Nature
Malaria
Conflict and Terrorism
cause

DATA ANALYSIS

From the above graph we can see that the top 5 cause for the highest deaths in china are due to:

	cause	total deaths
13	Cardiovascular Diseases	100505973
21	Neoplasms	61060527
28	Chronic Respiratory Diseases	36676826
30	Digestive Diseases	8924906
14	Lower Respiratory Infections	8525819



DATA ANALYSIS

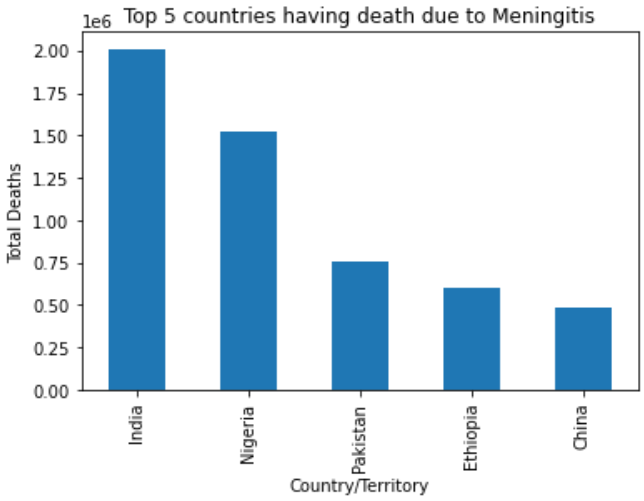
The Above plot shows the trend of top 5 diseases effect on China in the past 30 years
1. cardiovascular diseases are increasing in trend and rapidly increasing
2. Neoplasms deaths are steadily increasing
3. Chronic Respiratory diseases are in decreasing trend
4. Digestive diseases are in a stable trend
5. Lower Respiratory infection deaths are also in decreasing trend

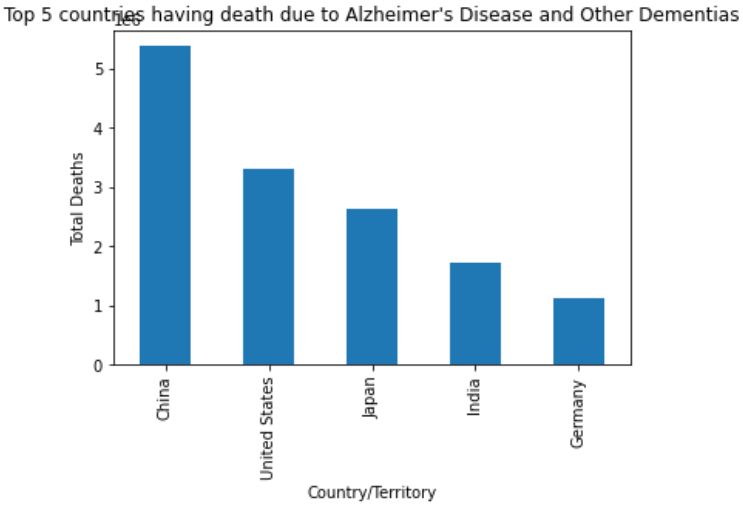
ANALYSIS OF EACH DISEASE EFFECT ON ALL COUNTRIES

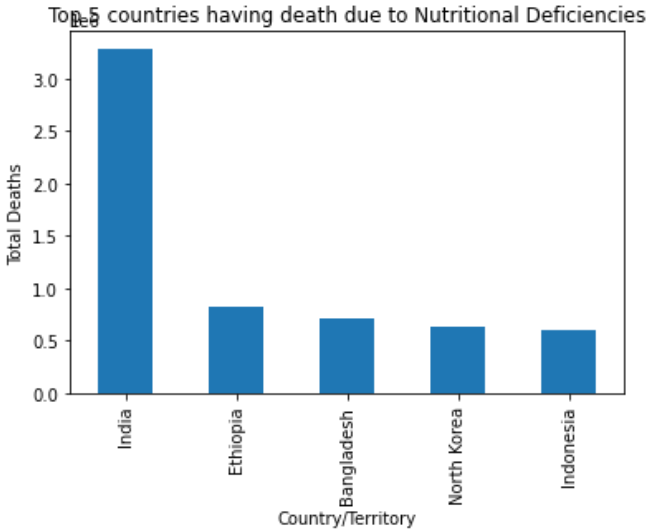
In [52]:

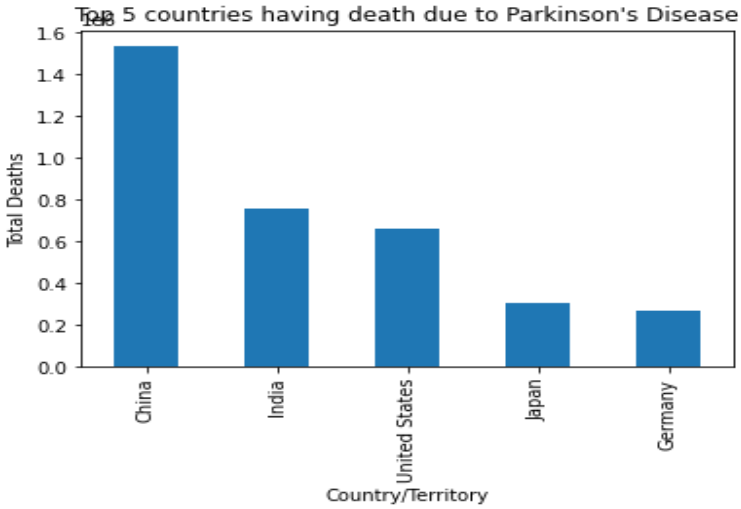
```
ds1=ds.select_dtypes(include=['float64','int64'])
del ds1["Year"]
for i in ds1.columns:

ds.groupby('Country/Territory')[i].sum().sort_values(ascending=False).head().plot(kind='bar')
    plt.title('Top 5 countries having death due to '+i)
    plt.ylabel('Total Deaths')
    plt.show()
```

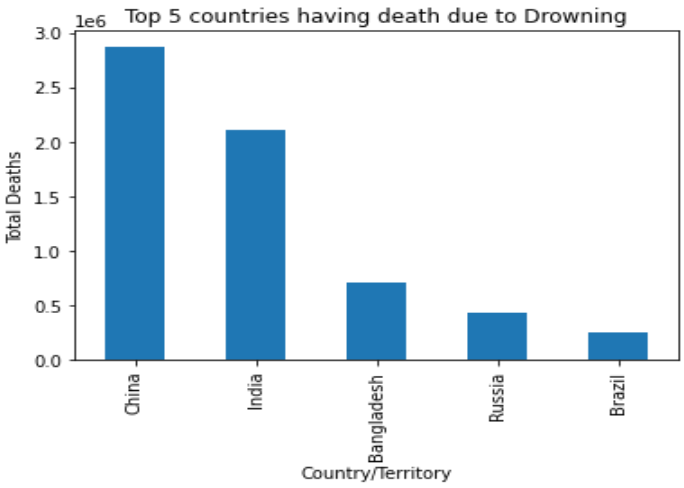


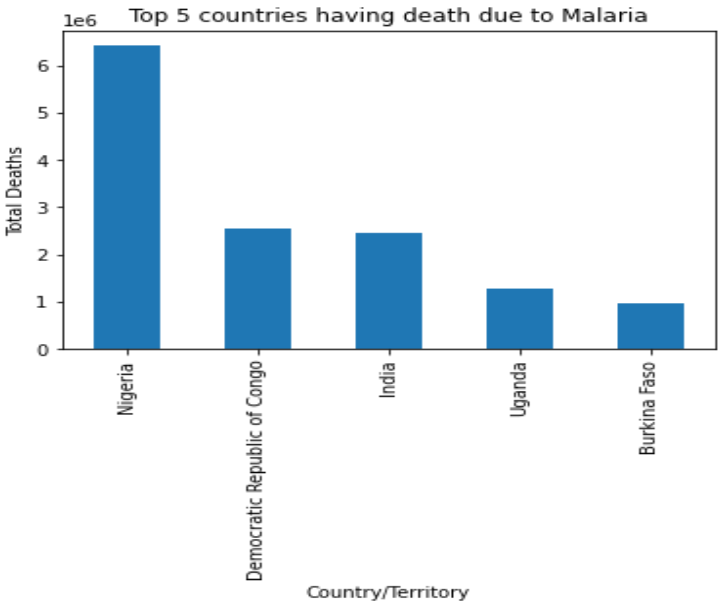


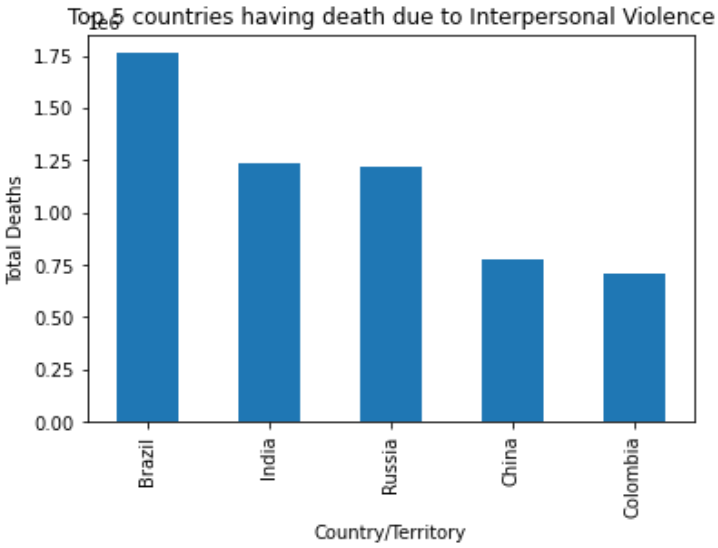


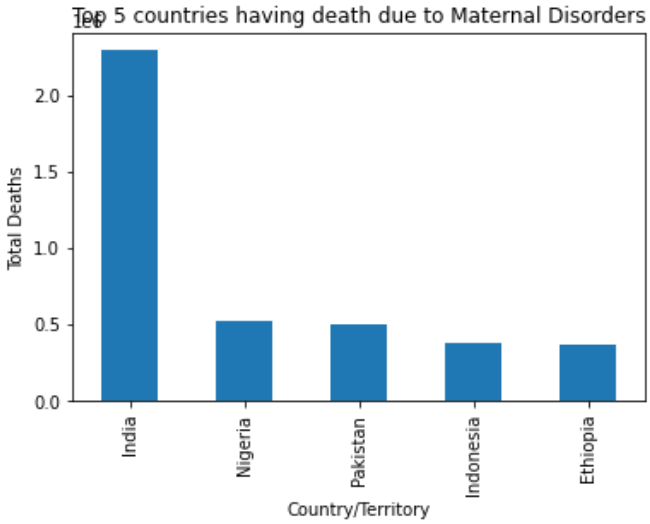


DATA ANALYSIS



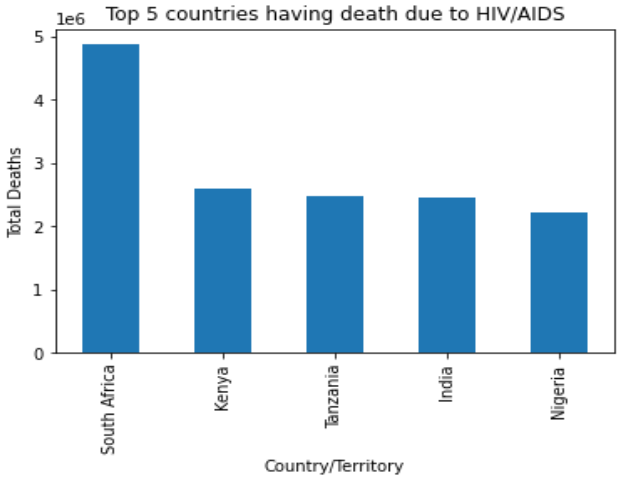


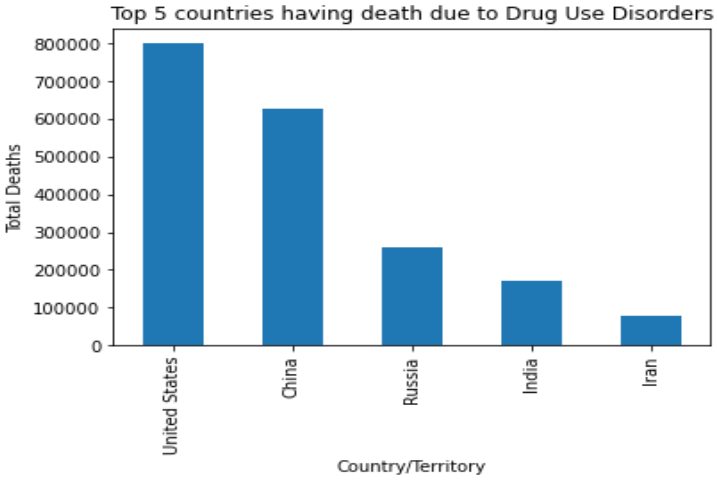


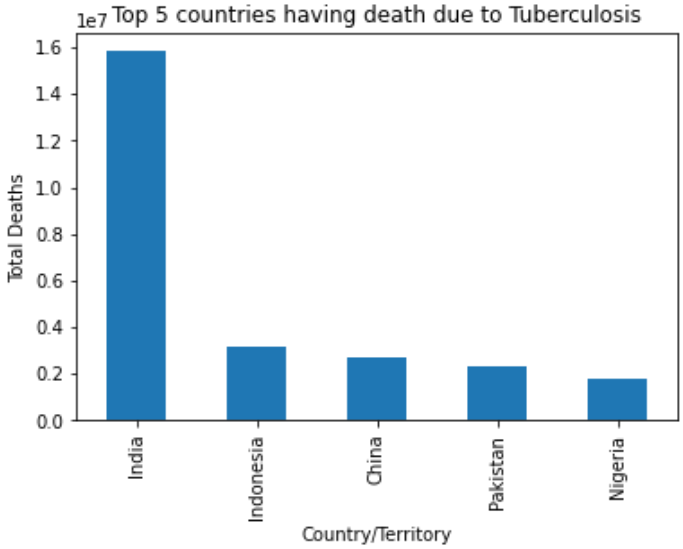


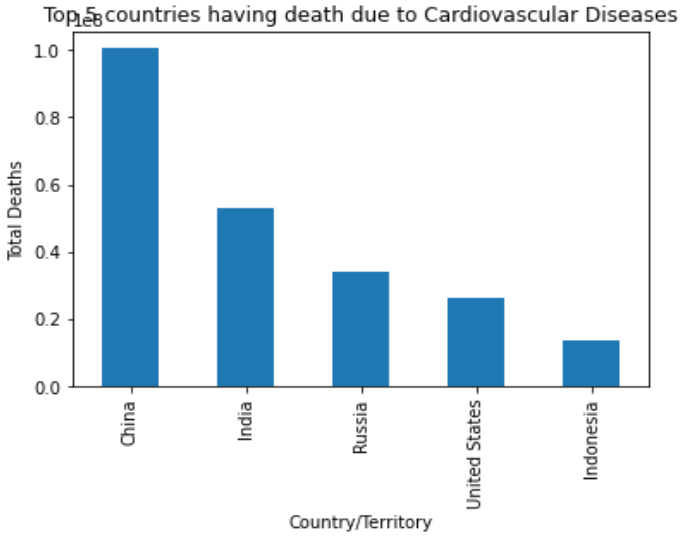
DATA ANALYSIS

From The Above Graphs we can observe that:

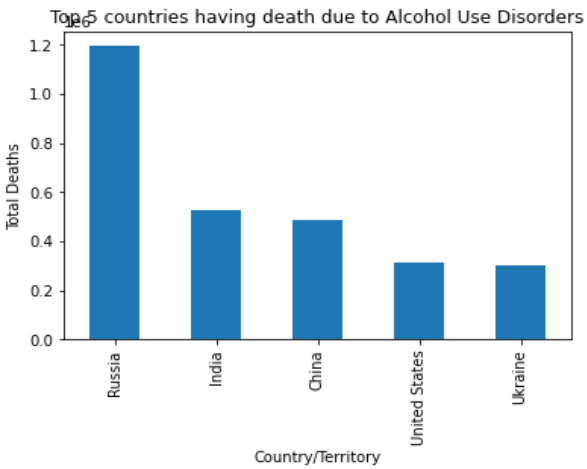


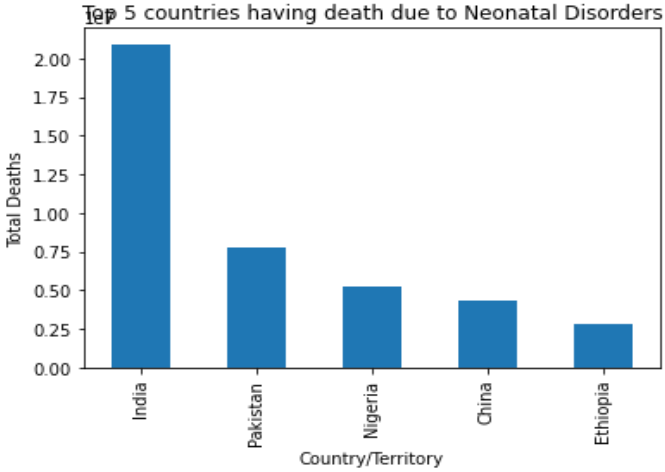


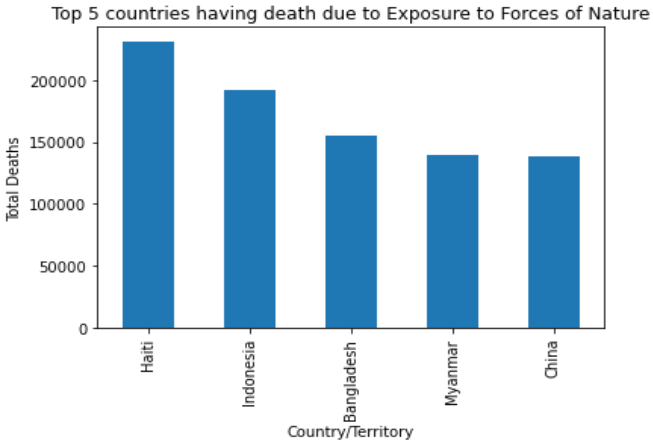


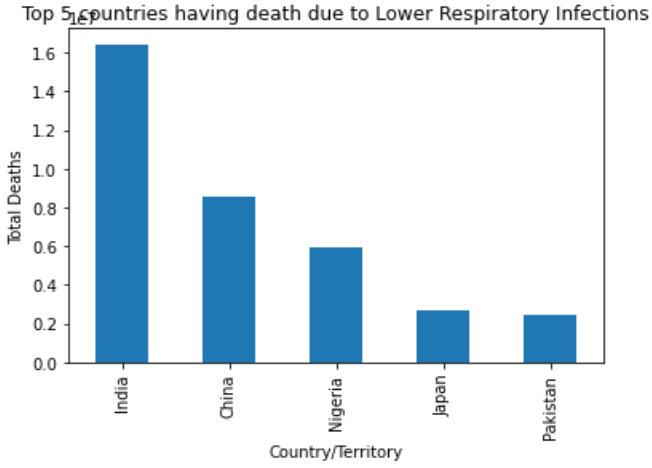


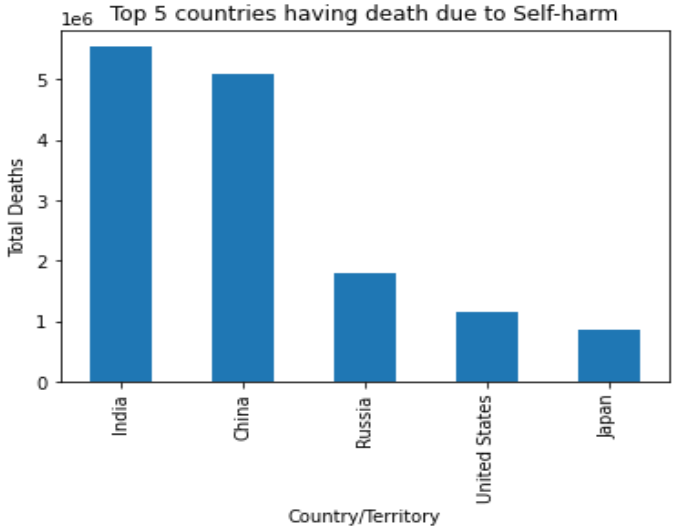
DATA ANALYSIS

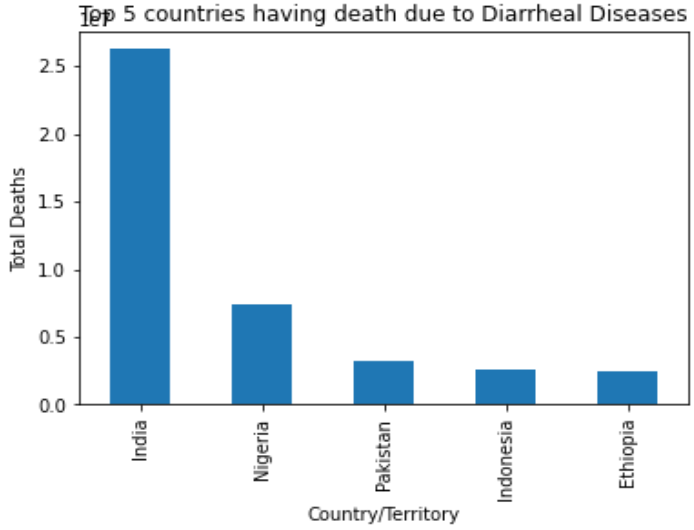




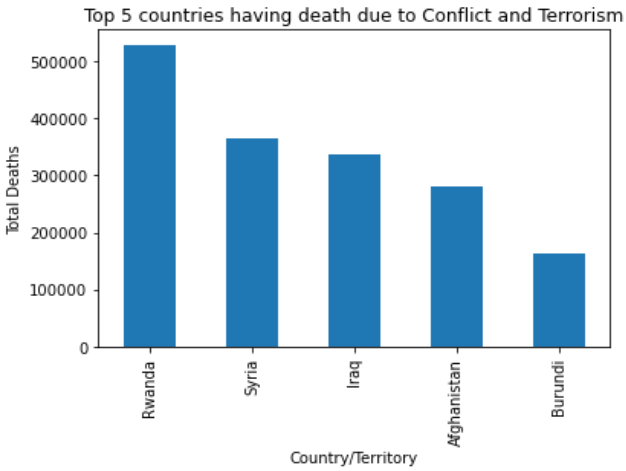


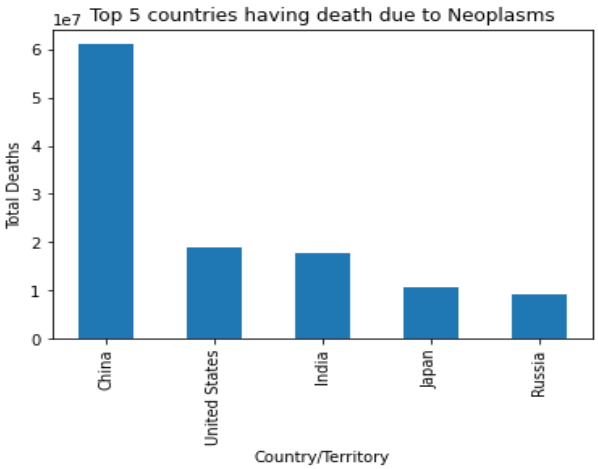


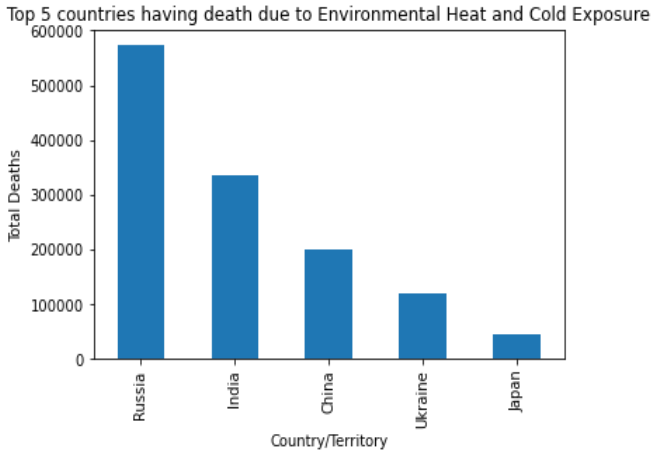


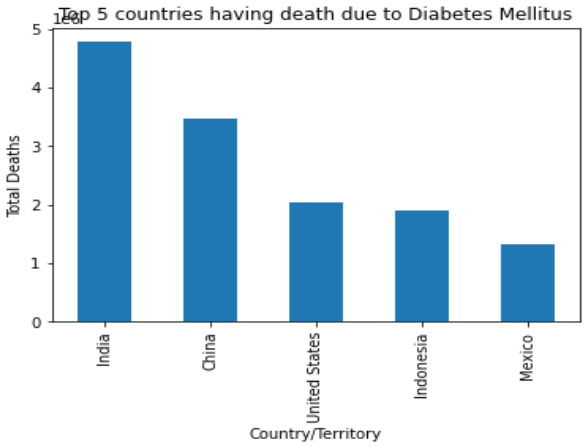


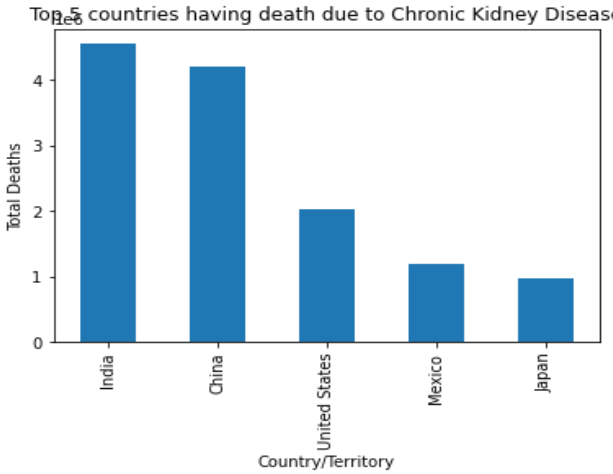
DATA ANALYSIS

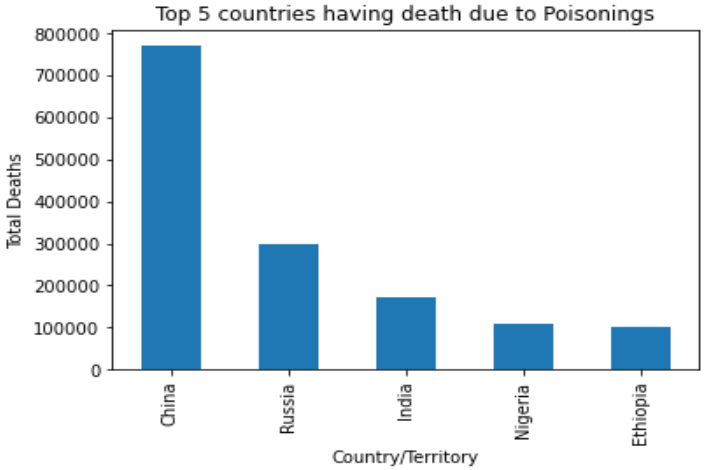




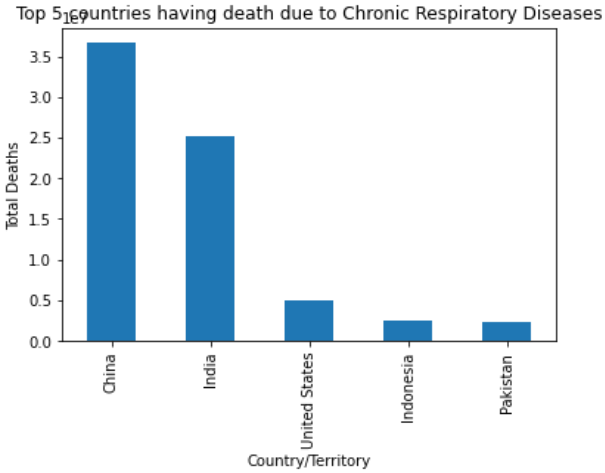


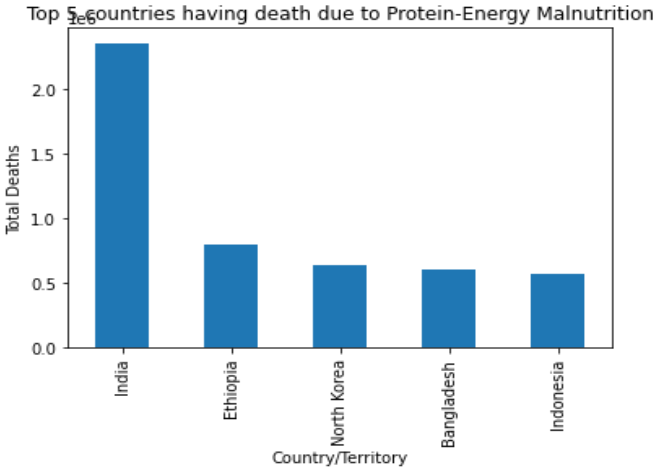


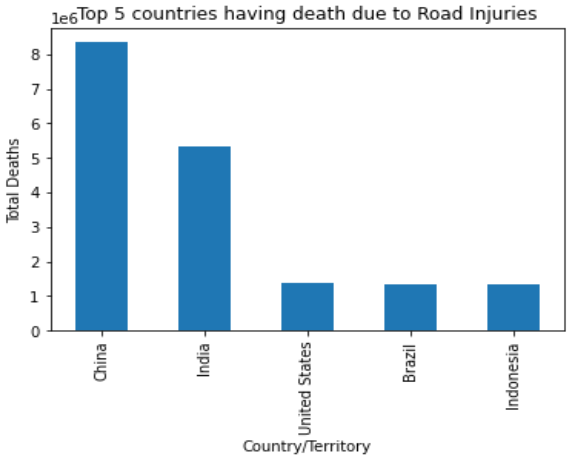


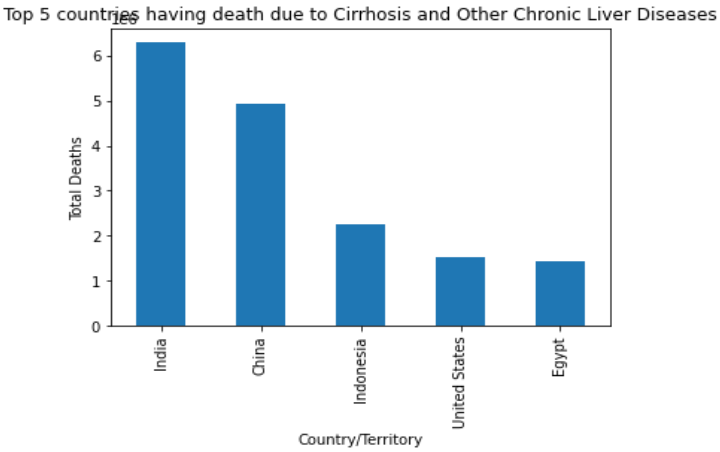


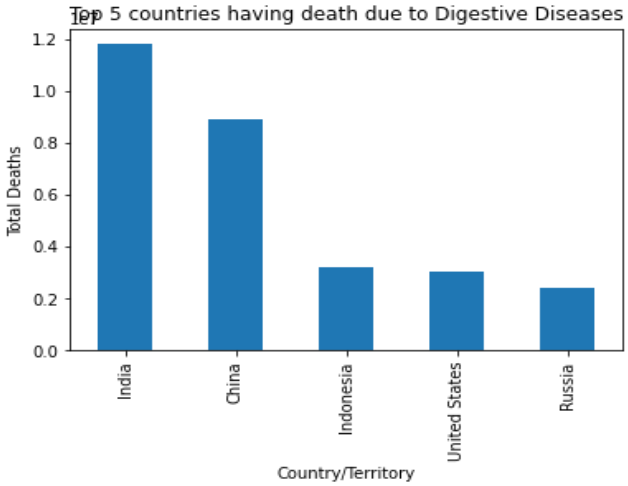
DATA ANALYSIS

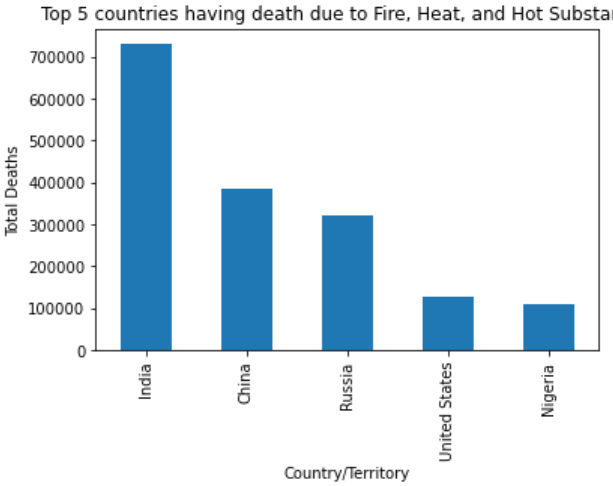




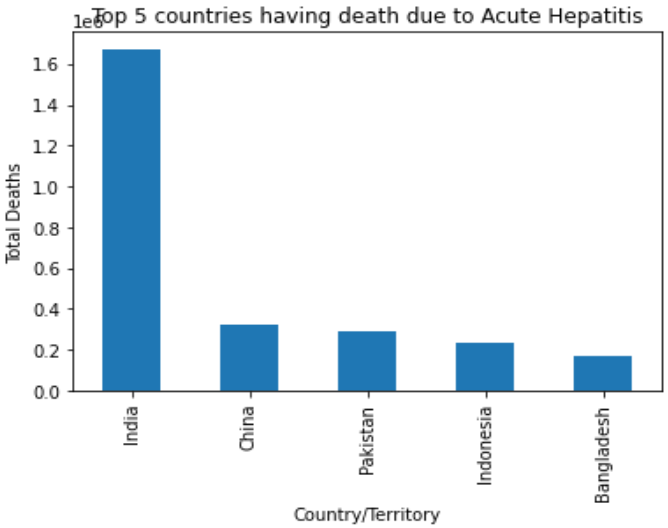


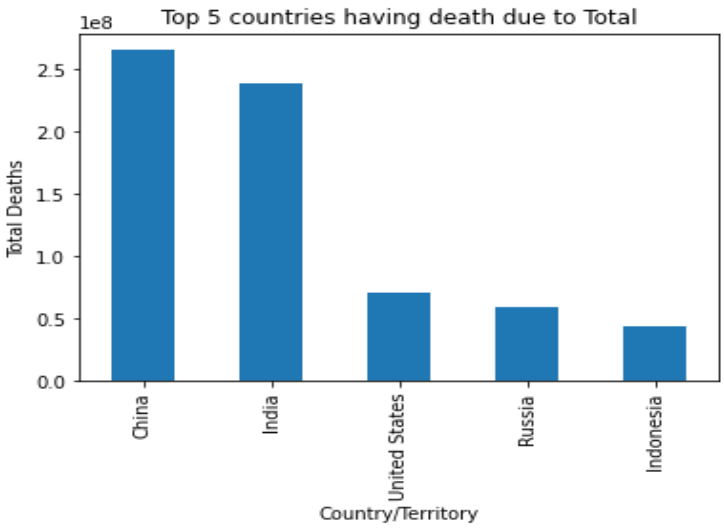






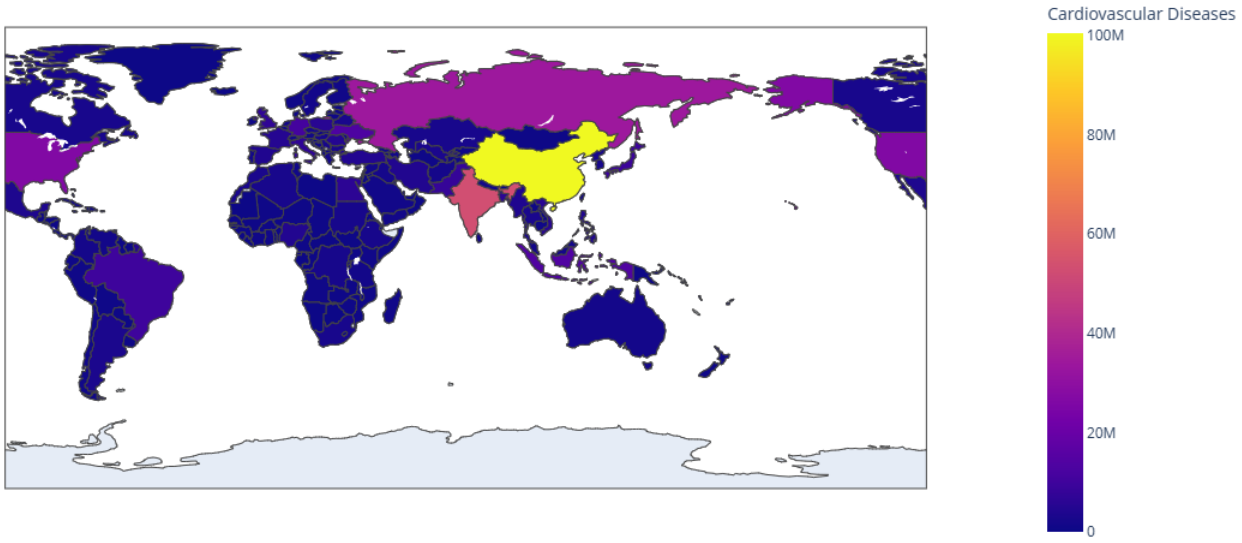
DATA ANALYSIS



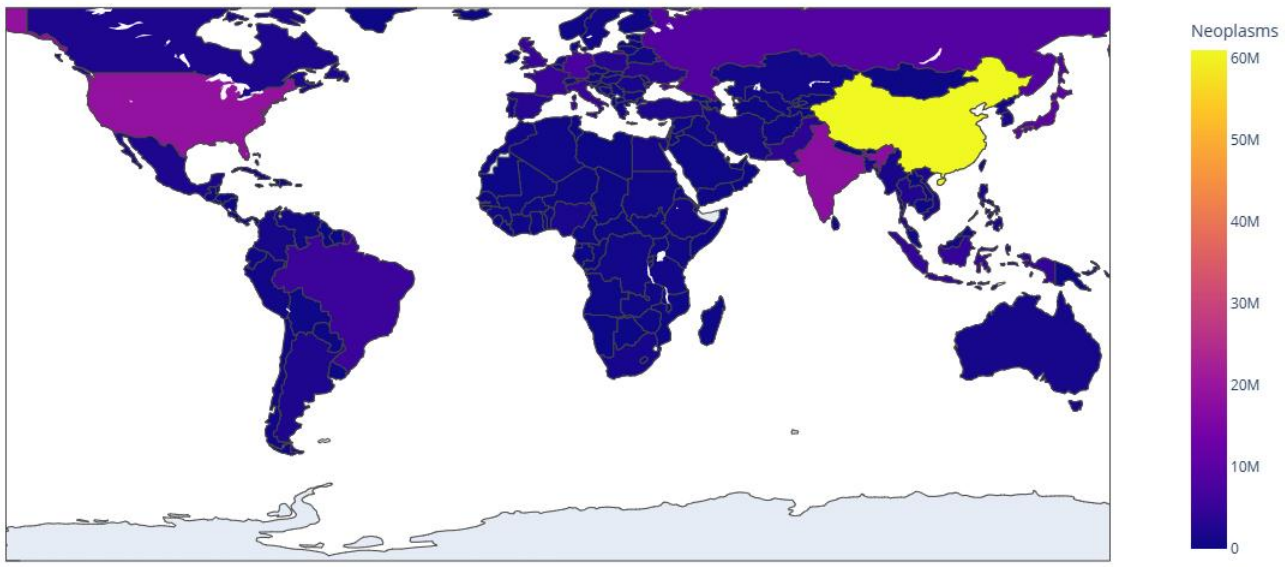


DATA ANALYSIS

Cardiovascular Diseases Deaths

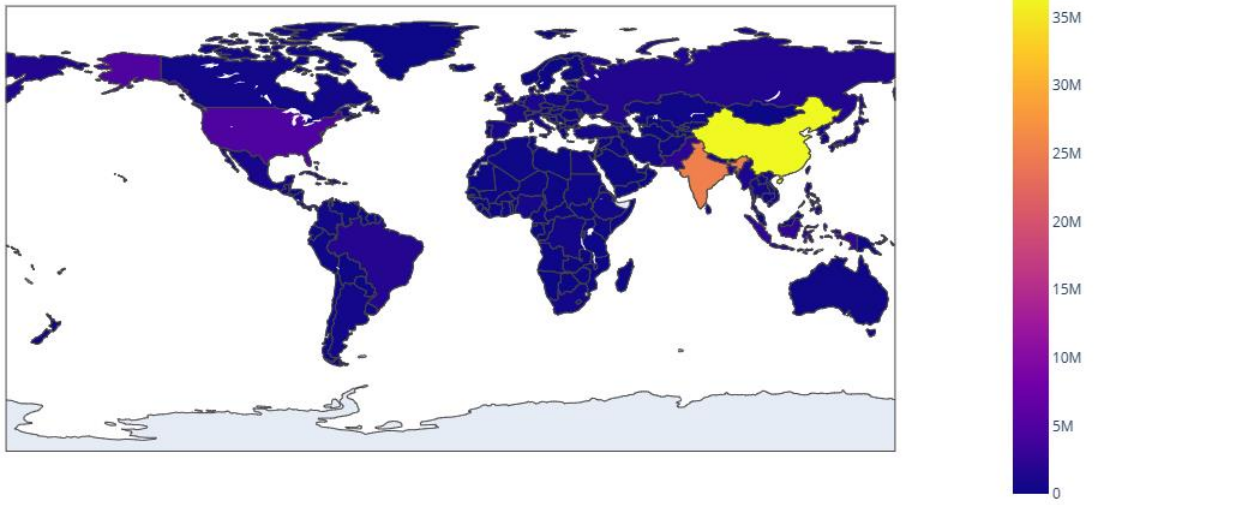


Neoplasms Deaths

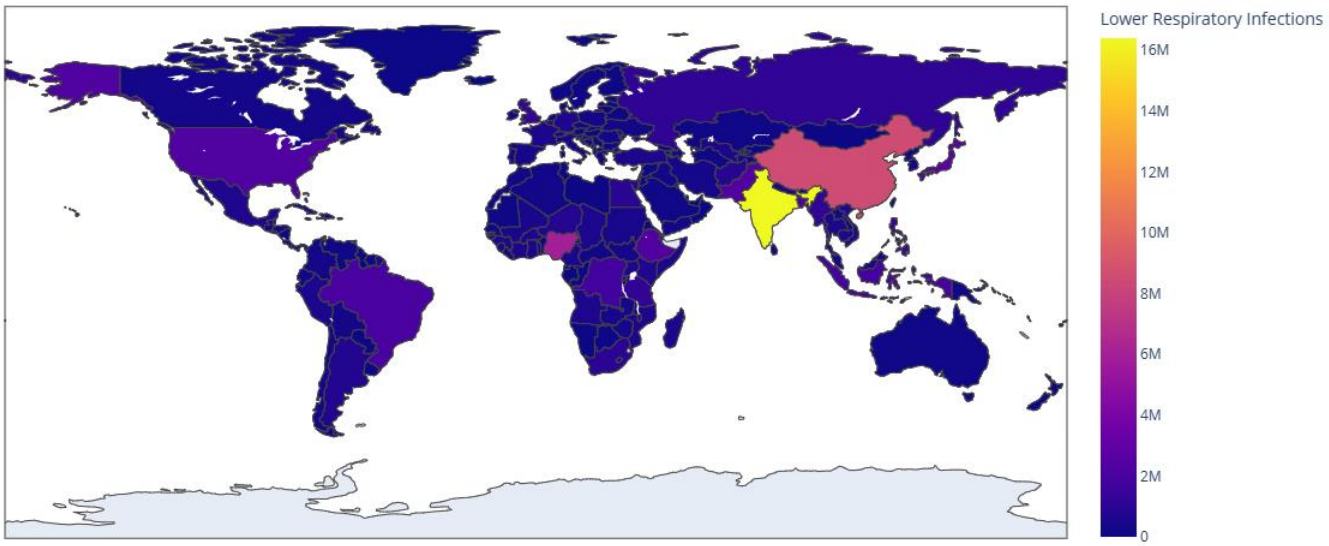


DATA ANALYSIS

Chronic Respiratory Disease Deaths

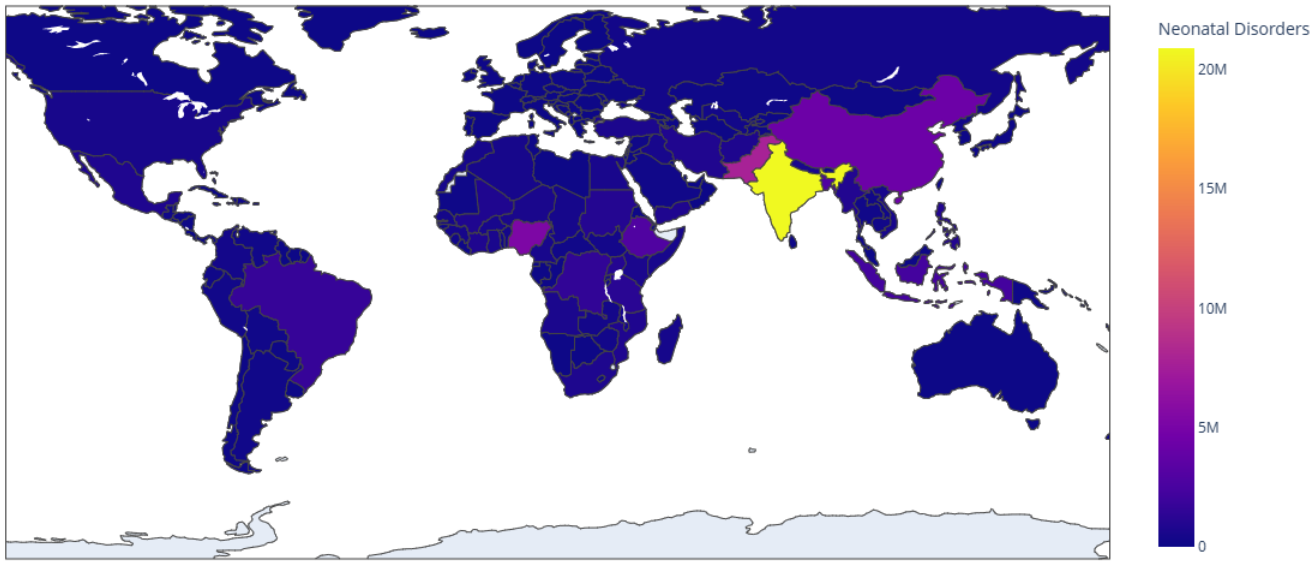


Lower Respiratory Infection Deaths



DATA ANALYSIS

Neonatal Disorder Deaths



DATA ANALYSIS

DATA ANALYSIS

Deaths causes were splitted into 3 different Ctegories like:

1. Death by Diseases
2. Death by natural Causes
3. Death By Human Causes

Deaths By Diseases

In [88]:
```
Diseases_Year=Diseases.groupby(['Year'])[["Meningitis",
                          "Alzheimer's Disease and Other Dementias",
                          "Parkinson's Disease",
                          "Digestive Diseases",
                          "Malaria",
                          "Tuberculosis",
                          "Diabetes Mellitus",
                          "HIV/AIDS",
                          "Acute Hepatitis",
                         "Parkinson's Disease",
                        "Nutritional Deficiencies",
                       "Cardiovascular Diseases",
                       "Neoplasms","Neonatal Disorders","Maternal Disorders",
                       "Diarrheal Diseases","Chronic Kidney Disease",
                     "Chronic Respiratory Diseases",
                     "Cirrhosis and Other Chronic Liver Diseases",
                     "Lower Respiratory Infections"]].sum().reset_index()

Nature_Year=Nature_Cause.groupby(['Year'])[["Environmental Heat and Cold Exposure",
                    "Drowning",
                    "Road Injuries",
                    "Exposure to Forces of Nature",
                    "Protein-Energy Malnutrition"]].sum().reset_index()

Human_Year=Human_Cause.groupby(['Year'])[["Interpersonal Violence",
                                          "Drug Use Disorders",
                                          "Alcohol Use Disorders",
                                          "Self-harm",
                                          "Conflict and Terrorism",
                                          "Poisonings"]].sum().reset_index()
```

In [91]:
```
Nature_Year["Nature_Total"] = 0
Nature_cols = Nature_Year.columns
Nature_cols=Nature_cols.drop(['Year'])
Nature_Year["Nature_Total"] = Nature_Year[Nature_cols].sum(axis=1)

Diseases_Year["Diseases_Total"] = 0
Diseases_cols = Diseases_Year.columns
Diseases_cols=Diseases_cols.drop(['Year'])
Diseases_Year["Diseases_Total"] = Diseases_Year[Diseases_cols].sum(axis=1)

Human_Year["Human_Total"] = 0
Human_cols = Human_Year.columns
Human_cols=Human_cols.drop(['Year'])
Human_Year["Human_Total"] = Human_Year[Human_cols].sum(axis=1)
```

DATA ANALYSIS

In [111]:
```
N1=Nature_Year.drop(["Environmental Heat and Cold Exposure",
                     "Drowning",
                     "Road Injuries",
                     "Exposure to Forces of Nature",
                     "Protein-Energy Malnutrition"],axis=1)
H1=Human_Year.drop(["Interpersonal Violence",
                    "Drug Use Disorders",
                   "Alcohol Use Disorders",
                   "Self-harm",
                   "Conflict and Terrorism",
                   "Poisonings"],axis=1)
D1=Diseases_Year.drop(["Meningitis",
                           "Alzheimer's Disease and Other Dementias",
                           "Parkinson's Disease",
                           "Digestive Diseases",
                           "Malaria",
                           "Tuberculosis",
                           "Diabetes Mellitus",
                           "HIV/AIDS",
                           "Acute Hepatitis",
                          "Parkinson's Disease",
                         "Nutritional Deficiencies",
                        "Cardiovascular Diseases",
                        "Neoplasms","Neonatal Disorders","Maternal Disorders",
                        "Diarrheal Diseases","Chronic Kidney Disease",
                       "Chronic Respiratory Diseases",
                       "Cirrhosis and Other Chronic Liver Diseases",
                       "Lower Respiratory Infections"],axis=1)
```

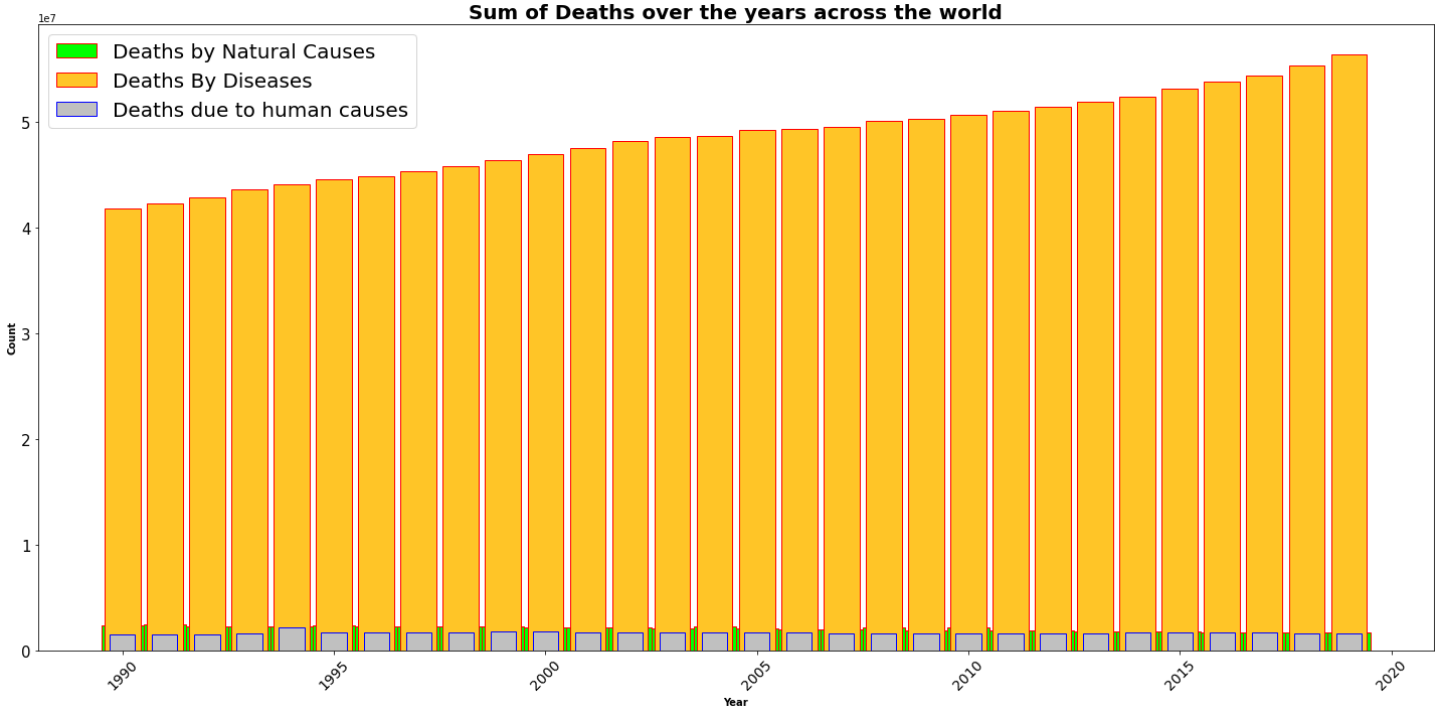
In [116]:
```
Data=N1.merge(D1[["Year","Diseases_Total"]],on="Year")
#df1.merge(df2[['Name', 'Grade', 'Rank']])
Data=Data.merge(H1[["Year","Human_Total"]],on="Year")
Data
```

Here we got the data separated for each year by the three above mentioned death causes, Here are few rows for reference:

Out[116]:

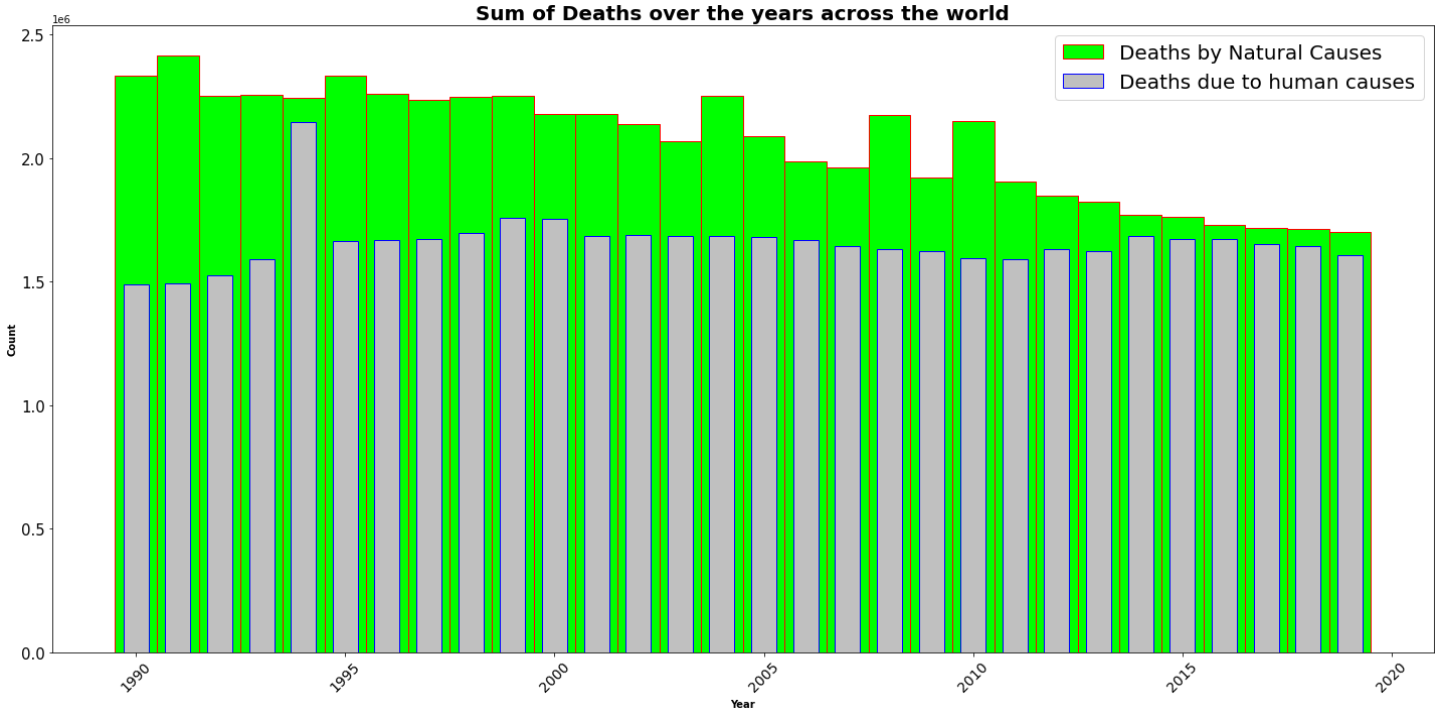
	Year	Nature_Total	Diseases_Total	Human_Total
0	1990	2334493	41780179	1488061
1	1991	2415622	42289829	1493462
2	1992	2250593	42880381	1526451
3	1993	2254978	43613942	1593035
4	1994	2246060	44125449	2147208
5	1995	2332990	44584305	1664430
6	1996	2262399	44871912	1667042
7	1997	2237250	45324451	1674600
8	1998	2247567	45771514	1698716
9	1999	2252869	46387743	1758627
10	2000	2180224	46967373	1754236
11	2001	2179451	47486298	1686426
12	2002	2140006	48165512	1689931
13	2003	2069127	48578884	1684051
14	2004	2252065	48685444	1684436
15	2005	2088897	49245972	1680344
16	2006	1988125	49287291	1669075
17	2007	1963286	49527180	1642275
18	2008	2173407	50097352	1632442
19	2009	1923675	50255006	1621878
20	2010	2148490	50705380	1596396

DATA ANALYSIS

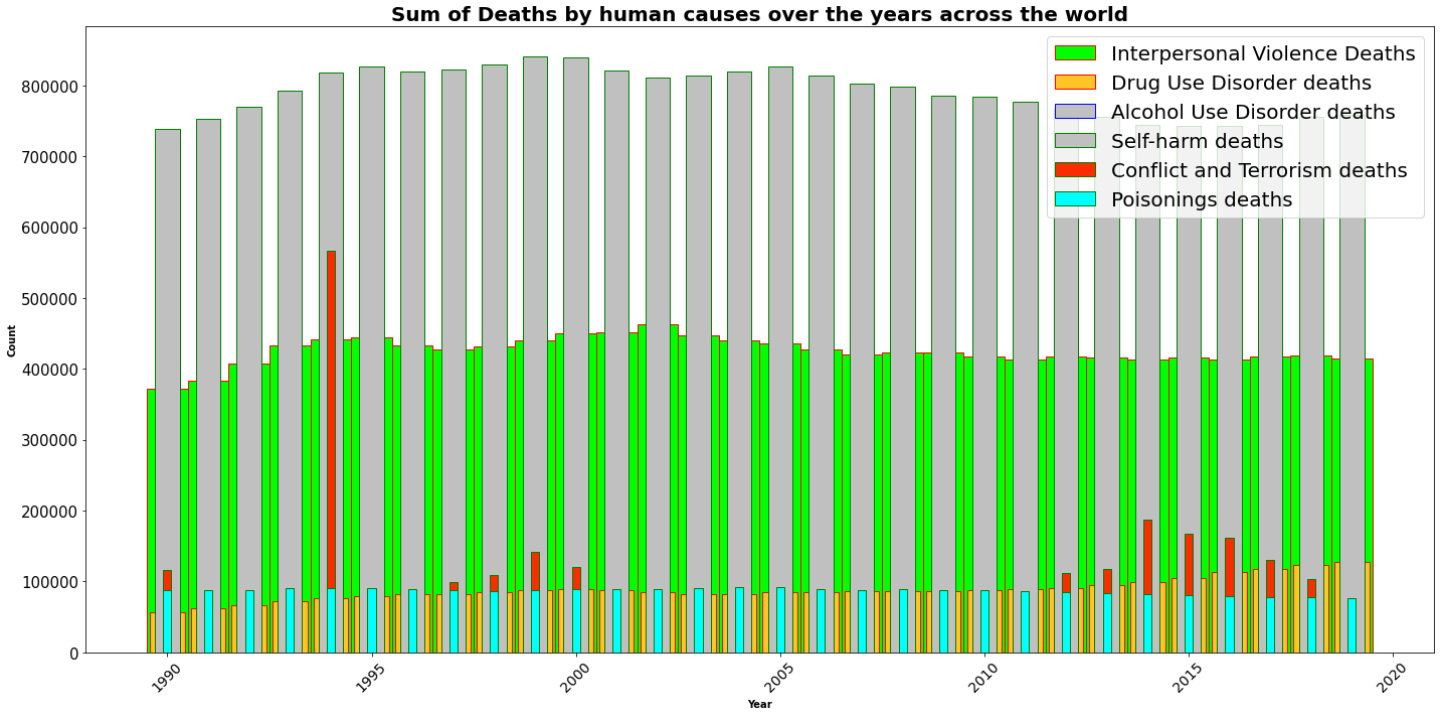


Deaths by diseases are the major for deaths in the world

Lets check the other two causes separately to compare



DATA ANALYSIS



In 1994 the major life loss is due to conflict/terrorism and its reduced ove time and again inceased around 2015.

CONCLUSION

The following Diseases were in continuously decreasing trend since last 30 Years:

1. Meningtis
2. Nutritional Deficiencies
3. Drowining
4. Maternal Disorders
5. Tuberculosis
6. Lower Respiratory infections
7. Neonatal Disorders
8. Diarrheal Disorders
9. Poisonings
10. Protein Energy Malnutrition
11. Acute Hepatitis

The following Diseases were in continuously increasing trend since last 30 Years:

1. Alzheimer's disorders
2. Parkinson's diseases
3. Drug use disorders
4. Cardiovascular Disorders
5. Alcohol use disorders
6. Neoplasms
7. Diabetes Related
8. Chronic kidney Diseases
9. Road Injuries
10. Chronic Respiratory diseases
11. Digestive diseases
12. Cirrhosis and Other Chronic Liver Diseases

HIV/AIDS was increased during 2005 and now its in decreasing trend.
I found out that there are many diseases which continiously increasing such as Neoplasms,Diabetes,Cardiovascular Diseases,Digestive disorder and Alzheimer. I Found out that there are many disease which are continously decreasing too such as Acute Hepatitis,Diarrheal Diseases,Nutritional Diseases and Meningitis. Parkinson Diseases seems to be constants till 1990 to 1993 after that no data is present for the same. We can see that in all the given years i.e 1990 to 2019 ,Road accident have taken Maximum lifes and the least can death can be seen in Exposure to force of Nature. In case of Death by crime ,self-harm and Accident -> Maximum death have been taken place by Conflict and Terroism and the second highest death have been recorded by -Interpersonal Violence.