NAME: Dipesh Ramesh Limaje

INTERNSHIP BATCH : 33

SME : Mr. Shwetank Mishra

Report On Cause Of Death

OBJECTIVES

1) To apply analytical skills and to give findings and conclusions in detail about data analysis.

2) **Causes of Death** is an official determination of conditions resulting in a human's death that resulted in or contributed to death, and the circumstances of the accident or Disease.

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

EDA STEPS

1) Check data shape(Num. of Rows, Num. of Columns)

(6120,34)
So there are 6120 Rows and 34 Columns.

2) Check each data type of columns and missing Values

So from data.info, we get all the information related to data types, and after observing all data types they are all ok.

And after observing all values from columns there are no missing values.

3) Change the data types

After all observations, the data looks good we don't need to change any of the data types.

4) Summary Statistics

```
In [6]: df.describe()
```

Out[6]:

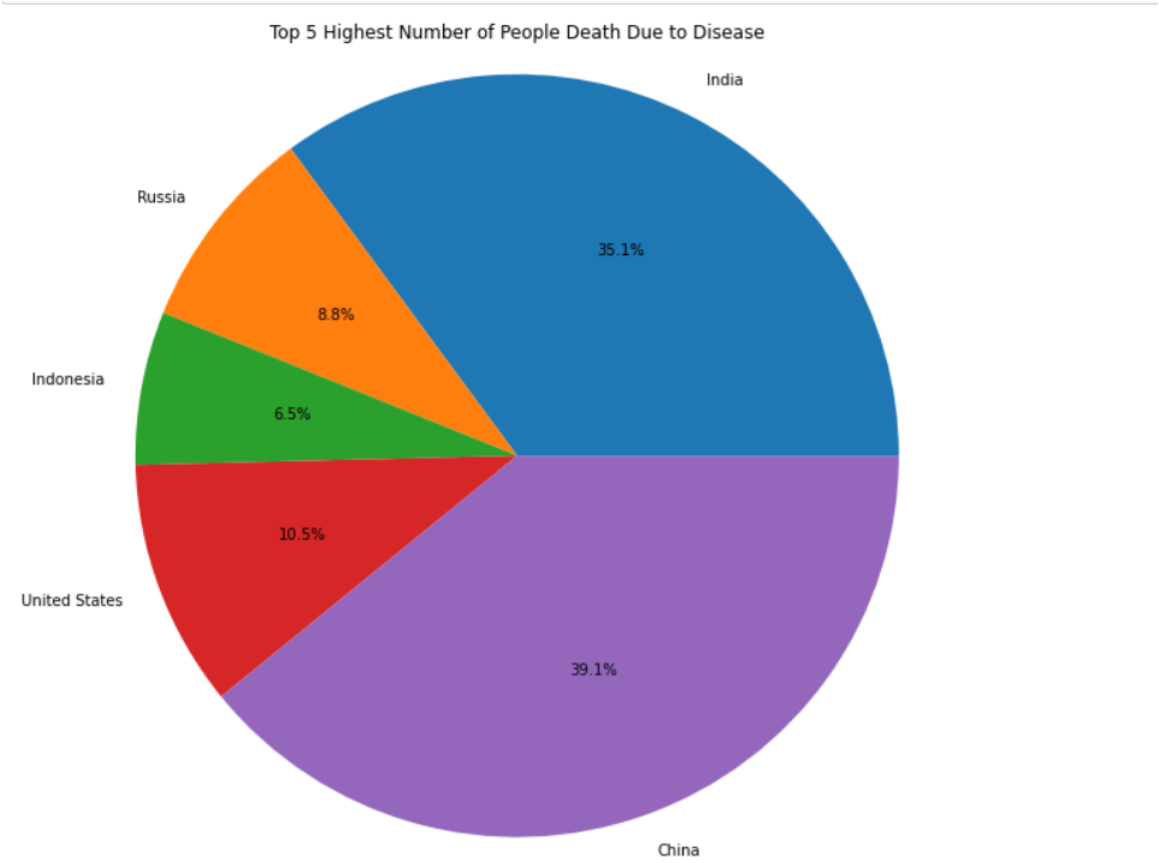
	Year	Meningitis	Alzheimer's Disease and Other Dementias	Parkinson's Disease	Nutritional Deficiencies	Malaria	Drowning	Interpersonal Violence	Maternal Disorders	HIV/AIDS	...
count	6120.000000	6120.000000	6120.000000	6120.000000	6120.000000	6120.000000	6120.000000	6120.000000	6120.000000	6120.000000	...
mean	2004.500000	1719.701307	4864.189379	1173.169118	2253.600000	4140.960131	1683.333170	2083.797222	1262.589216	5941.898529	...
std	8.656149	6672.006930	18220.659072	4616.156238	10483.633601	18427.753137	8877.018366	6917.006075	6057.973183	21011.962487	...
min	1990.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000	...
25%	1997.000000	15.000000	90.000000	27.000000	9.000000	0.000000	34.000000	40.000000	5.000000	11.000000	...
50%	2004.500000	109.000000	666.500000	164.000000	119.000000	0.000000	177.000000	265.000000	54.000000	136.000000	...
75%	2012.000000	847.250000	2456.250000	609.250000	1167.250000	393.000000	698.000000	877.000000	734.000000	1879.000000	...
max	2019.000000	98358.000000	320715.000000	76990.000000	268223.000000	280604.000000	153773.000000	69640.000000	107929.000000	305491.000000	...

From this, we can see the data distribution that you have for each and determine whether there are outliers or not.

5) Check Duplicate values

Here we don't have any specific unique ID so here we can't remove the duplicate, and if try to remove it the data will become useless.

6) See the data distribution and data anomaly



The above pie chart shows the Top 5 Countries in which the Total Number of death is the highest among all countries.

1] China 39.1%

2] India 35.1%

3] United States 10.5%

4] Russia 8.8%

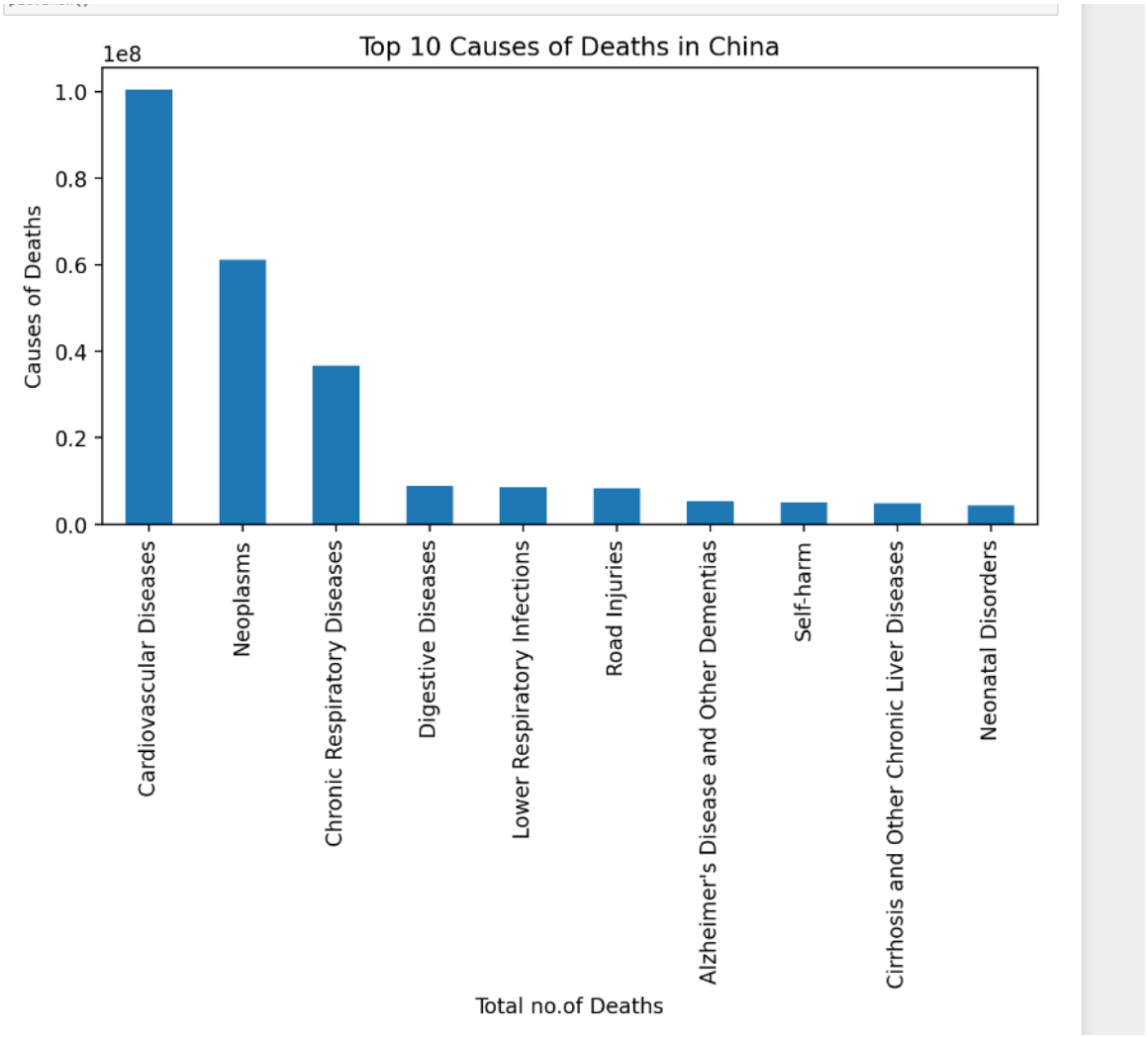
5] Indonesia 6.5%

Now we select these top 5 countries for further analysis

Now we are going to analyze the top 10 reasons why people die so much in the top 5 country

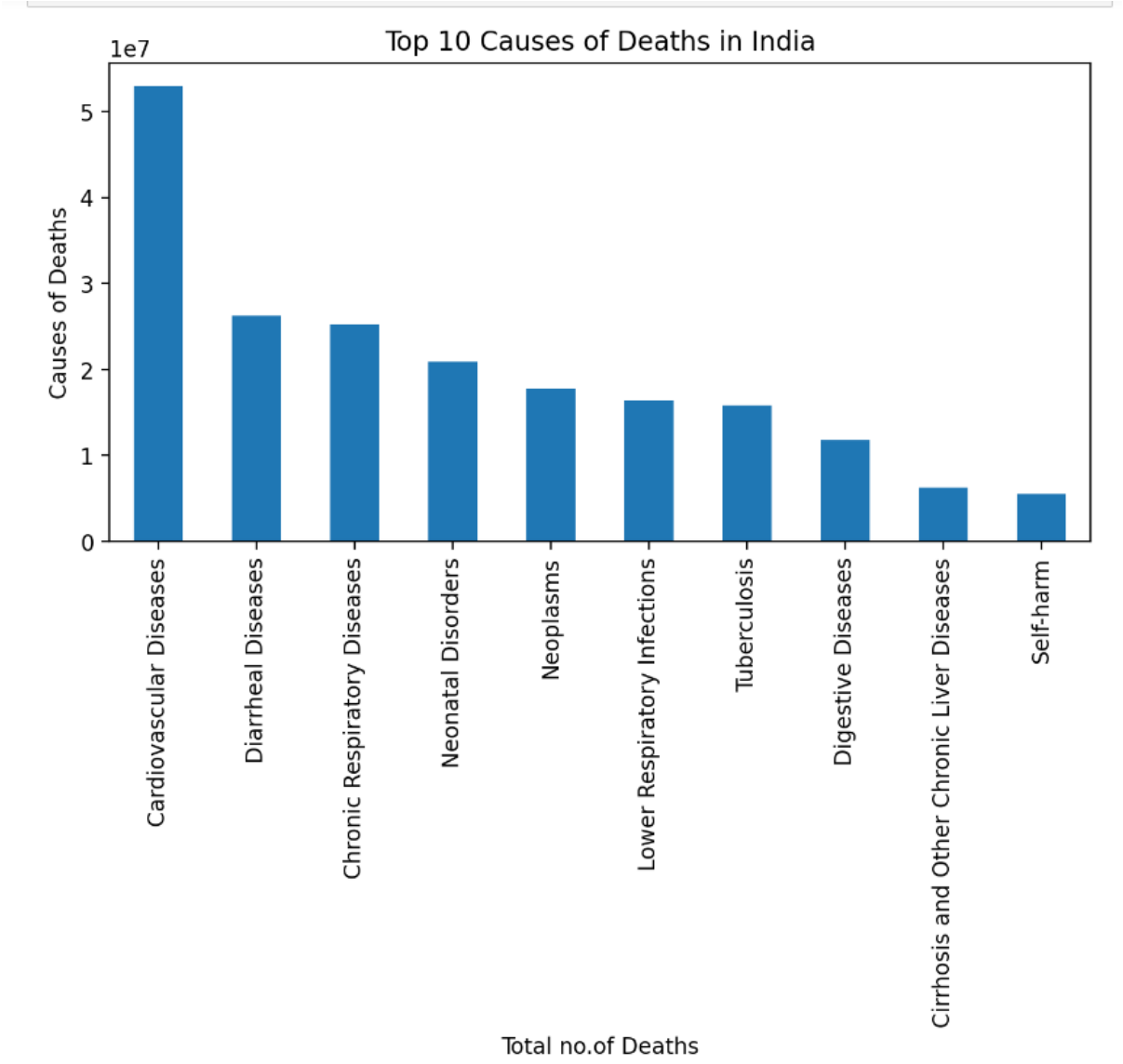
GRAPHICAL VISUALIZATION

Top 10 cause of death in china



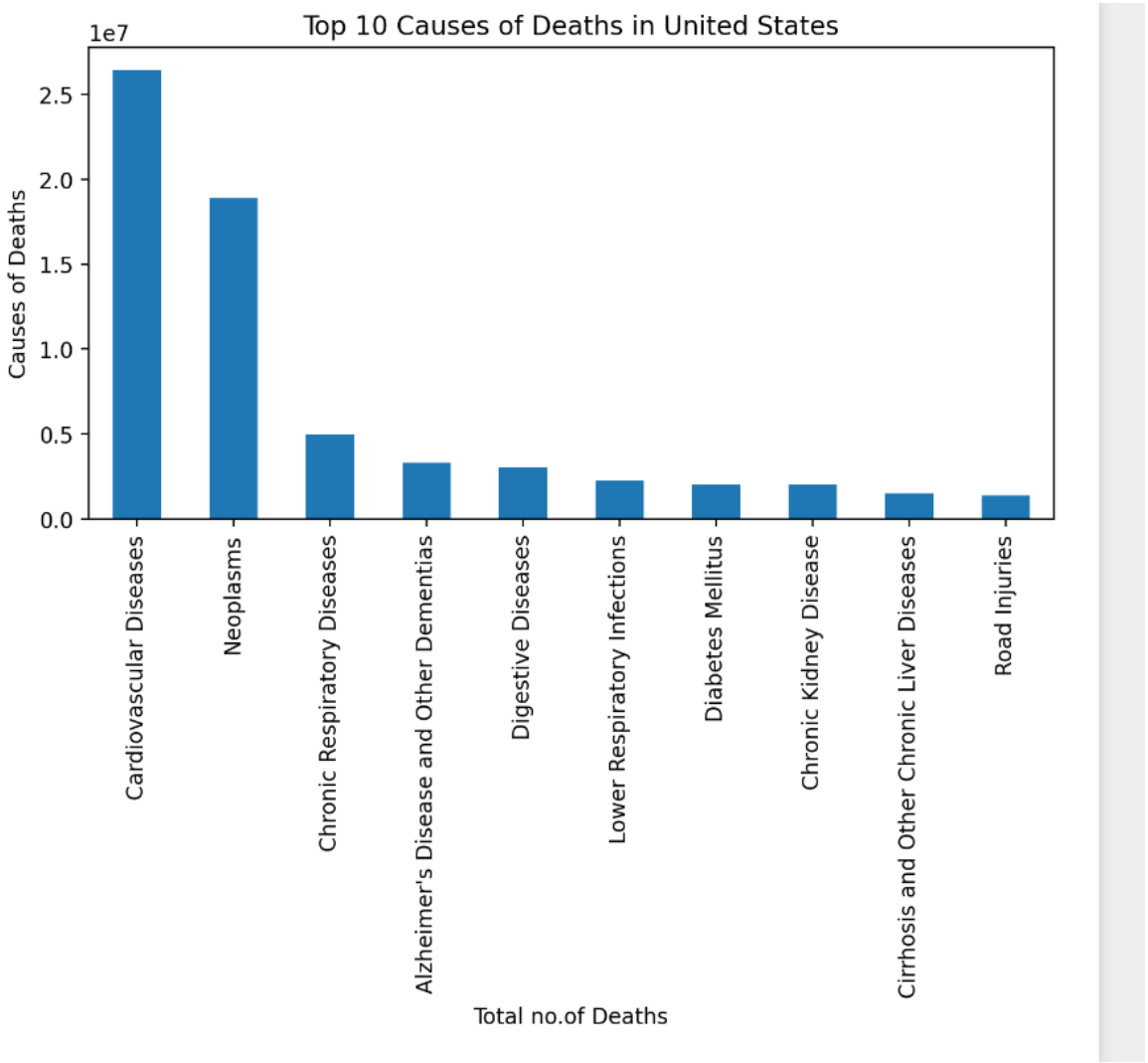
The above bar chart shows the top 10 diseases why people death is more in china, first and the main reason is due to Cardiovascular Disease followed by Neoplasms and Chronic Respiratory Disease.

Top 10 cause of death in India



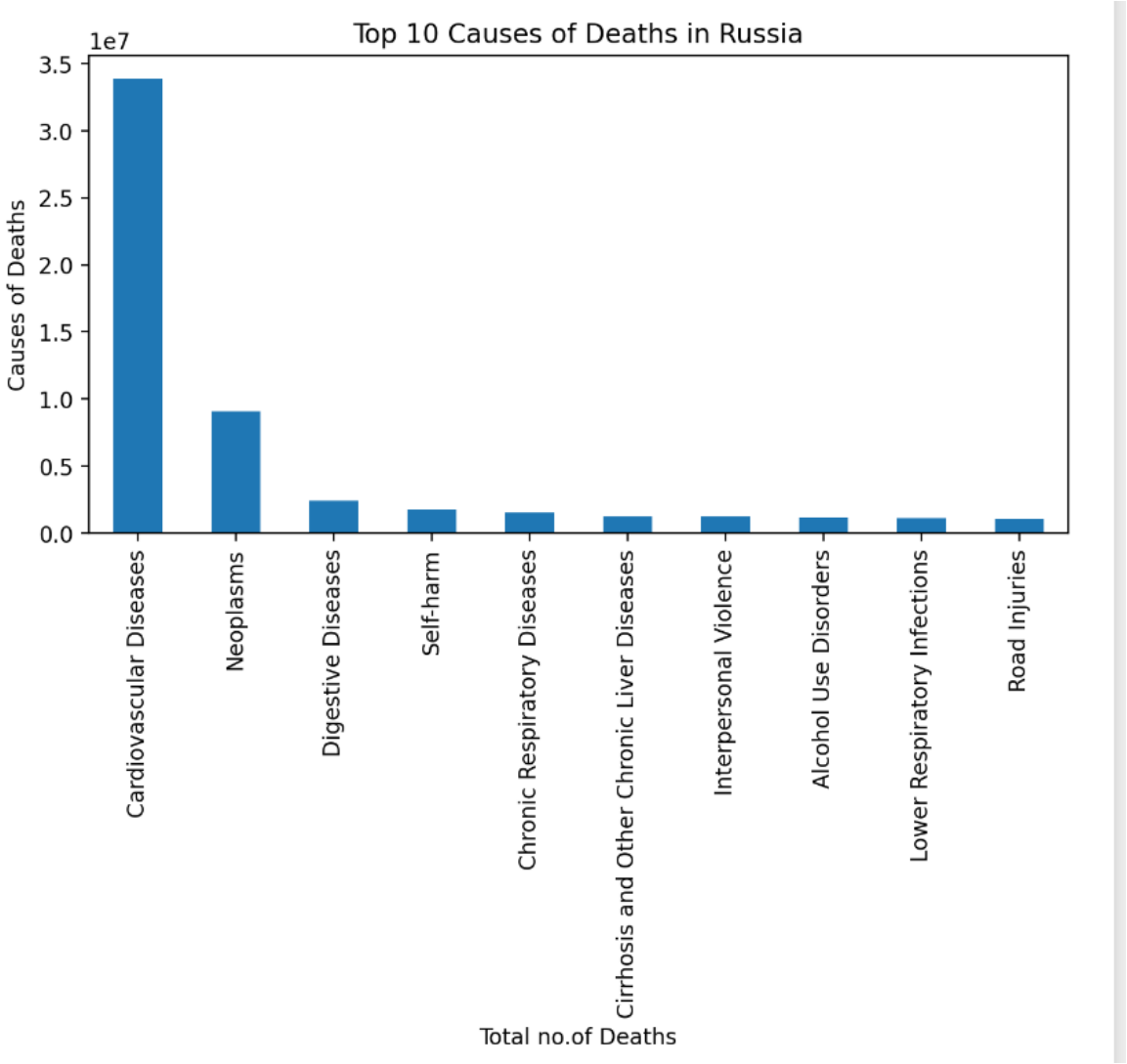
The above bar chart shows the top 10 diseases why people death is more in India, first and the main reason is due to Cardiovascular Disease followed by Diarrheal Disease, Chronic Respiratory Disease, and Neonatal Disorders.

Top 10 cause of death in United States



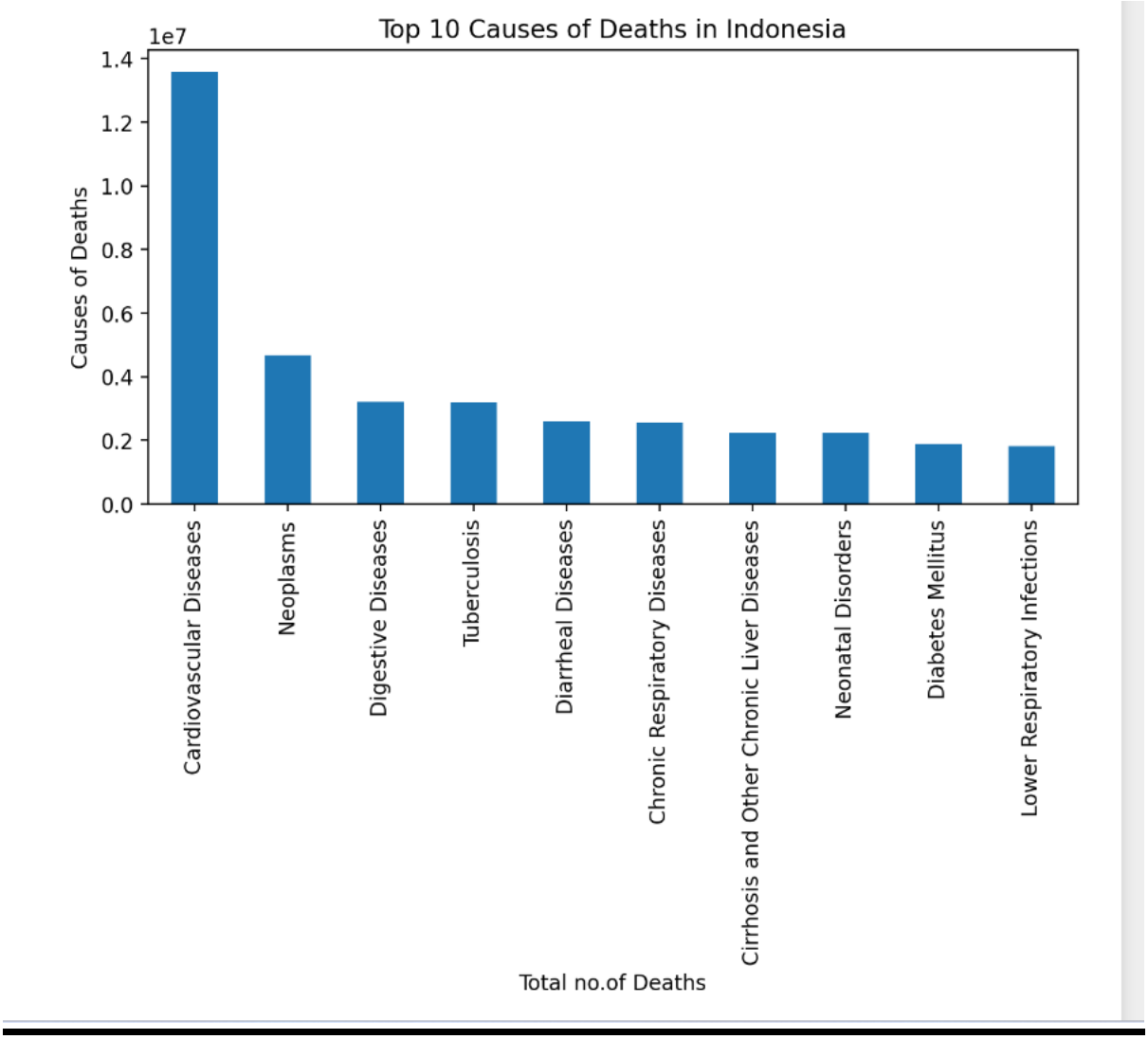
The above bar chart shows the top 10 diseases why people death is more in United States, first and the main reason is due to Cardiovascular Disease followed by Neoplasms.

Top 10 cause of death in Russia



The above bar chart shows the top 10 diseases why people death is more in Russia, first and the main reason is due to Cardiovascular Disease followed by Neoplasms.

Top 10 cause of death in Indonesia

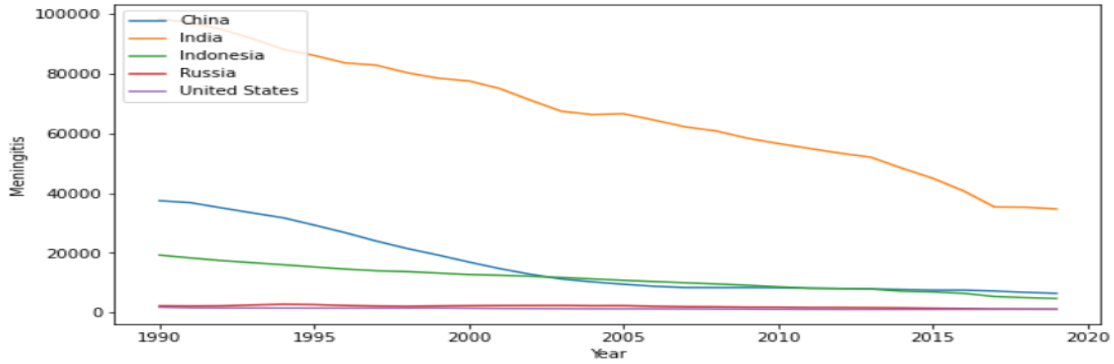


The above bar chart shows the top 10 diseases why people death is more in Indonesia, first and the main reason is due to Cardiovascular Disease followed by Neoplasms and Digestive Diseases.

Now Comparing All Disease With this top 5 country to observe which Country is most affected by which Disease.

Meningitis VS Country

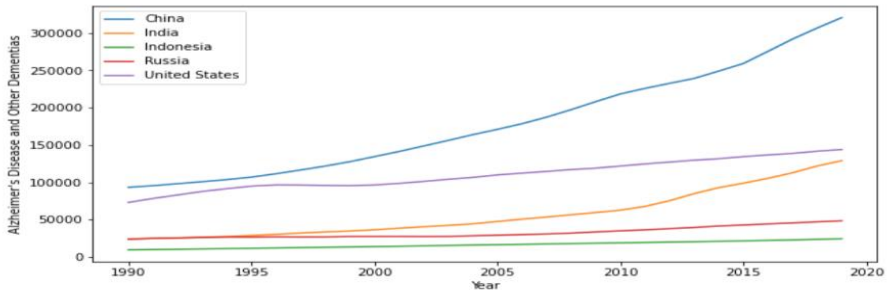
```
plt.figure(figsize=(10,5))
sns.lineplot(x='Year' , y="Meningitis" , data=result ,hue='Country/Territory')
plt.legend( loc ="upper left")
plt.show()
```



From the above Line graph, we observe people dying due to Meningitis was high in India in starting 1990 and now in 2019 it gradually decreases, the same condition in Indonesia and china but we also observe very few people die due to Meningitis in Russia and the United states. and we also observe except for India in 2019 the death ratio of people is more than in the rest of the 4 countries.

Alzheimer's Disease and other Dementias VS Country

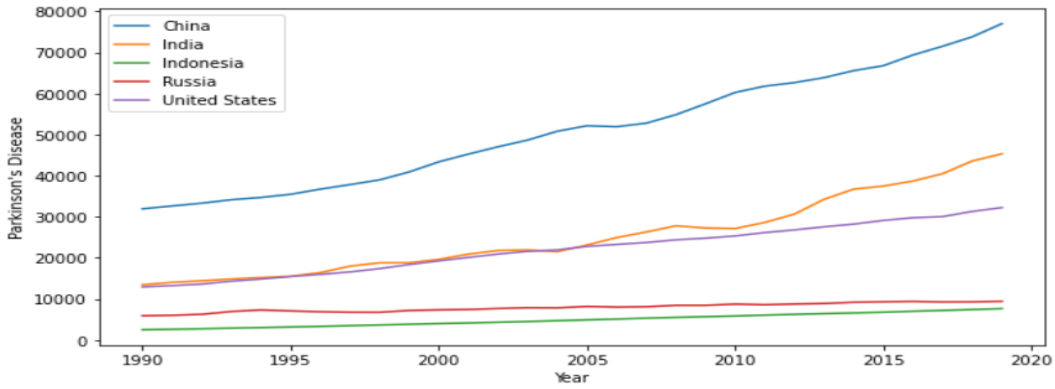
```
plt.figure(figsize=(10,5))
sns.lineplot(x='Year' , y="Alzheimer's Disease and Other Dementias" , data=result ,hue='Country/Territory')
plt.legend( loc ="upper left")
plt.show()
```



From the above Line graph, we observe people dying due to Alzheimer's Disease and Other Dementias were low in starting year (1990) but it is gradually increasing as time passes, From the above graph we also observe china is more affected as compared to the other 4 countries.

Parkinson Disease VS Country

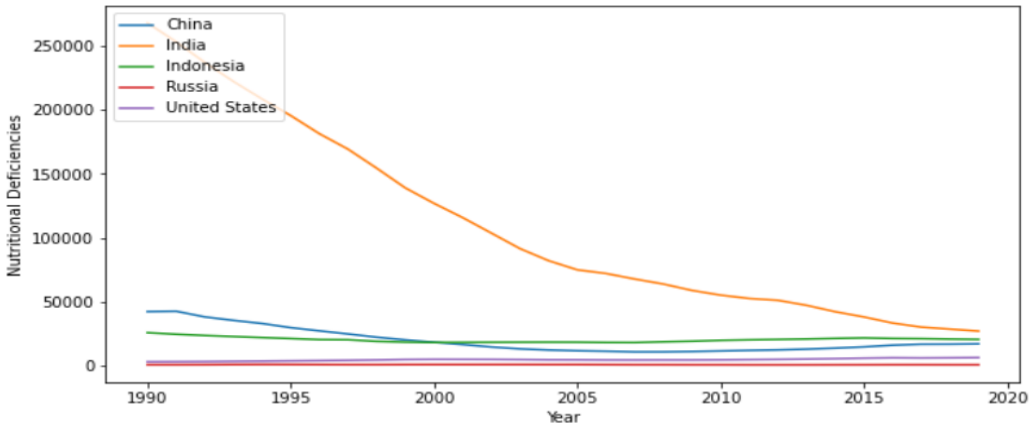
```
plt.figure(figsize=(10,5))
sns.lineplot(x='Year' , y="Parkinson's Disease" , data=result ,hue='Country/Territory')
plt.legend( loc ="upper left")
plt.show()
```



From the above Line graph, we observe people dying due to Parkinson's Disease were low in starting year (1990) but it is gradually increasing as time passes, From the above graph we also observe china is more affected as compared to the other 4 countries.

Nutritional Deficiencies VS Country

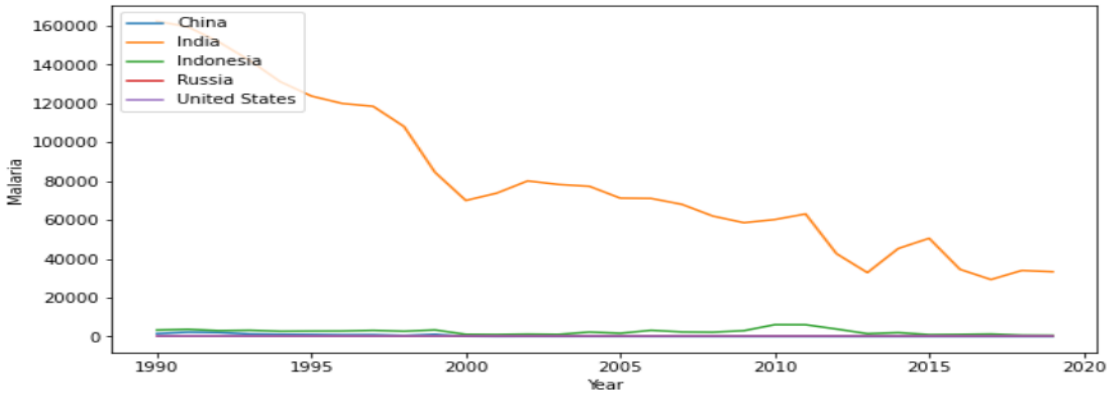
```
plt.figure(figsize=(10,5))
sns.lineplot(x='Year' , y='Nutritional Deficiencies' , data=result ,hue='Country/Territory')
plt.legend( loc ="upper left")
plt.show()
```



From the above Line graph, we observe people dying due to Nutritional Deficiencies was very high in India but we observe there is a sudden decrease from the year 2004 and now people dying due to Nutritional Deficiencies in 2019 are almost equal to all other 4 countries which are less than 50000 people.

Malaria VS Country

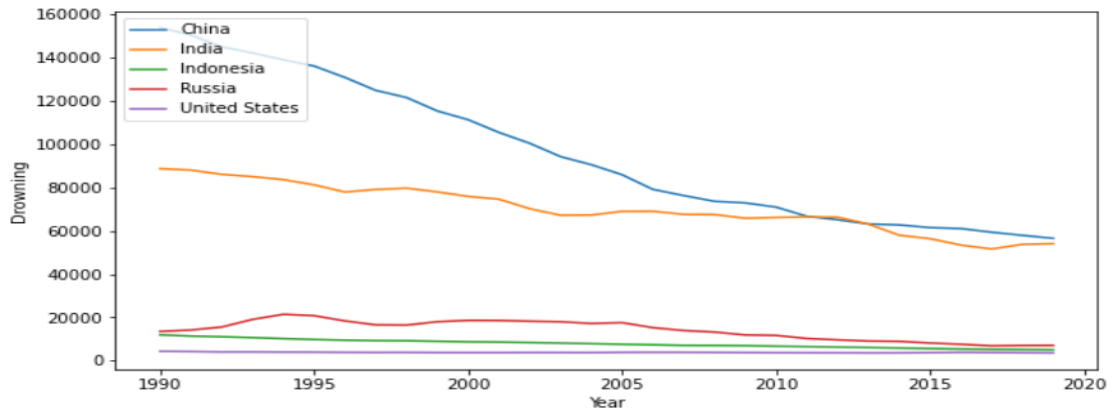
```
plt.figure(figsize=(10,5))
sns.lineplot(x='Year' , y='Malaria' , data=result ,hue='Country/Territory')
plt.legend( loc ="upper left")
plt.show()
```



From the above Line graph, we observe people dying due to Malaria is very high in India there is a gradually decrease from the year 1990 to 2019 but people dying due to Malaria is high as compared to the remaining 4 countries in the year 2019, but we also observe country affected by Malaria is only India.

Drowning VS Country

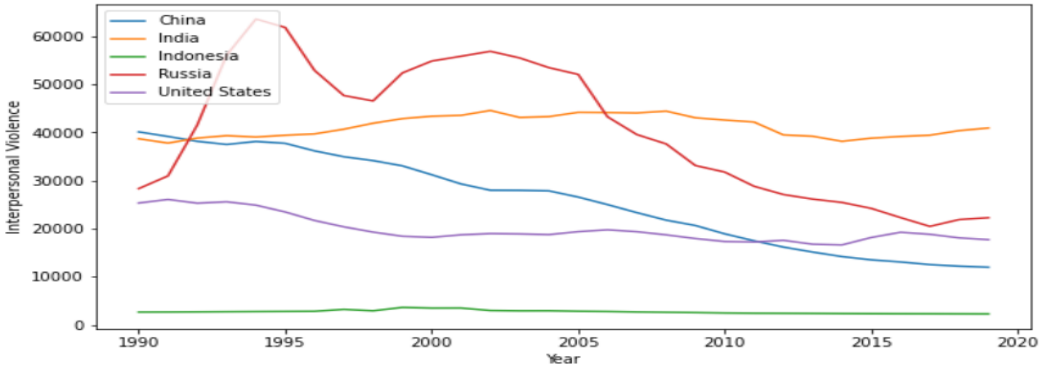
```
plt.figure(figsize=(10,5))
sns.lineplot(x='Year' , y='Drowning' , data=result ,hue='Country/Territory')
plt.legend( loc ="upper left")
plt.show()
```



From the above Line graph, we observe people dying due to Drowning is very high in China and Followed by India but they gradually decrease as time passes but not enough in the remaining countries the number of people dying due to Drowning is less than 20000.

Interpersonal Violence VS Country

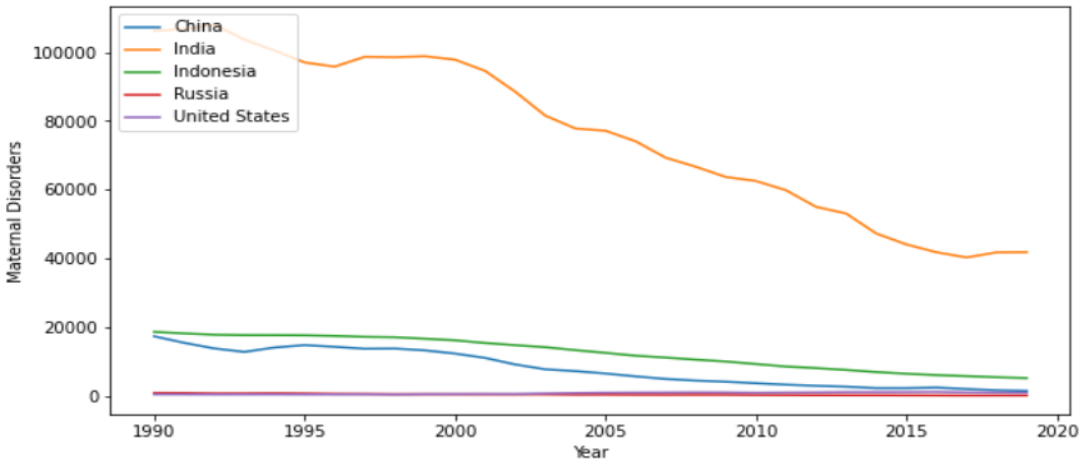
```
plt.figure(figsize=(10,5))
sns.lineplot(x='Year' , y='Interpersonal Violence' , data=result ,hue='Country/Territory')
plt.legend( loc ="upper left")
plt.show()
```



From the above Line graph, we observe people dying due to Interpersonal Violence is high in all Countries except Indonesia, as from the graph we observe the number of people who died due to Interpersonal Violence is Russia till 2005 and then it is gradually decreasing, but for India, it is constant at 40000 and for china and united states it is gradually decreasing from the year 1990 to the year 2019.

Maternal Disorders VS Country

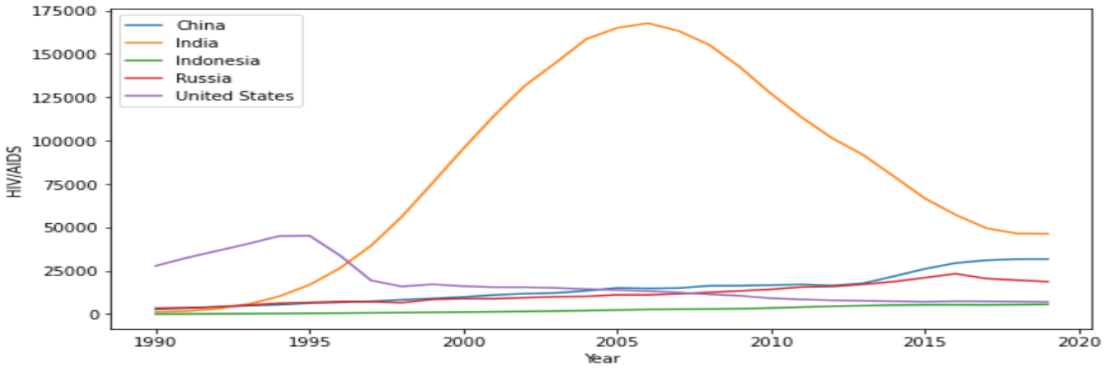
```
plt.figure(figsize=(10,5))
sns.lineplot(x='Year' , y='Maternal Disorders' , data=result ,hue='Country/Territory')
plt.legend( loc ="upper left")
plt.show()
```



From the above Line graph, we observe people dying due to Maternal Disorders are high in India and are gradually decreasing followed by Indonesia and China but as compared to other countries India's death is more in the year 2019.

HIV/AIDS VS Country

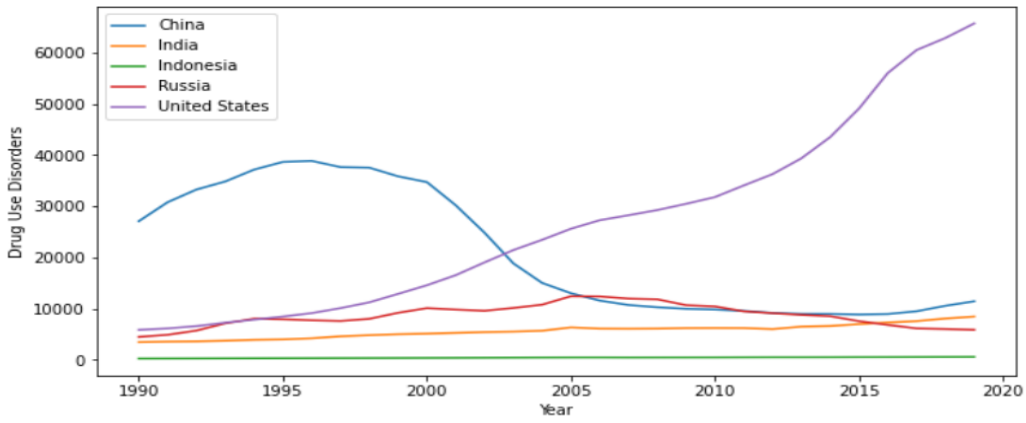
```
plt.figure(figsize=(10,5))
sns.lineplot(x='Year' , y='HIV/AIDS' , data=result ,hue='Country/Territory')
plt.legend( loc ="upper left")
plt.show()
```



From the above Line graph, we observe people dying due to HIV/AIDS are less starting the year 1990 except in the United States but after 1997 the people death is decreasing till 2019, India after 1992 the number of people die due to HIV/AIDS is sudden increase and continues till 2006 and after 2007 the people die due to HIV/AIDS id gradually decrease.

Drugs Use Disorders VS Country

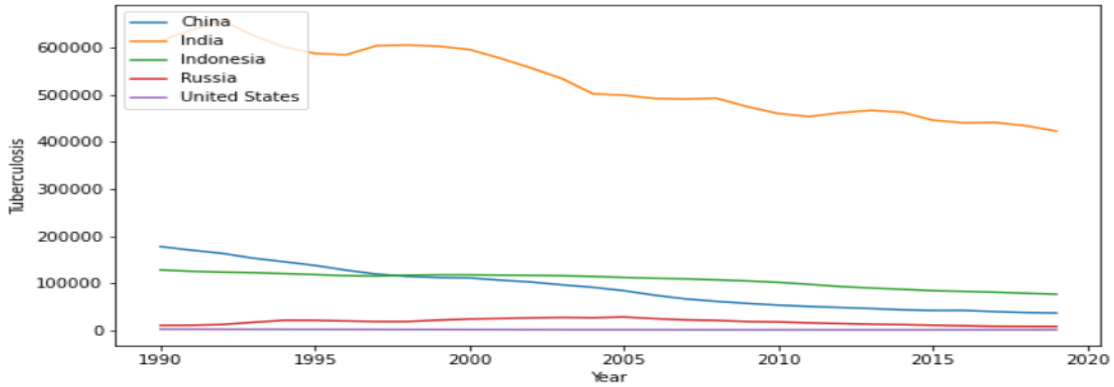
```
plt.figure(figsize=(10,5))
sns.lineplot(x='Year' , y='Drug Use Disorders' , data=result ,hue='Country/Territory')
plt.legend( loc ="upper left")
plt.show()
```



From the above Line graph, we observe people dying due to Drug Use Disorders was high in China but suddenly decrease after the 2000 year but in the united states the death of people due to Drug Use Disorders is increasing since the year 1990 and sudden increase in the year 2005 and continues increasing at last and for other 3 countries it remains constant under 10000.

Tuberculosis VS Country

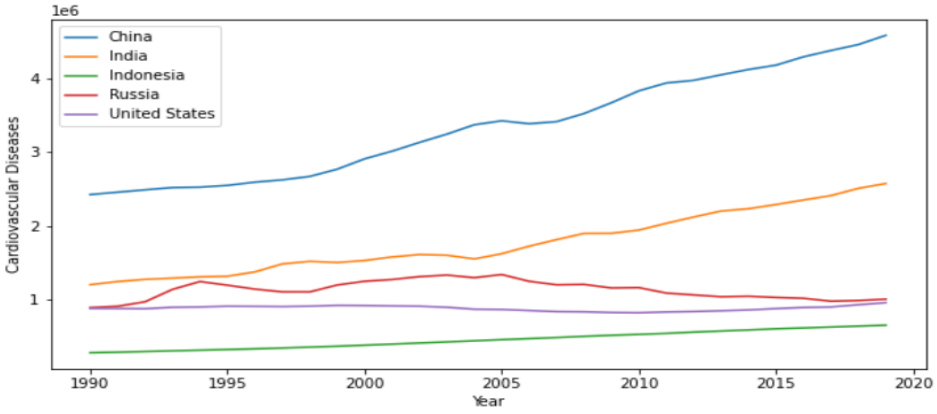
```
plt.figure(figsize=(10,5))
sns.lineplot(x='Year' , y='Tuberculosis' , data=result ,hue='Country/Territory')
plt.legend( loc ="upper left")
plt.show()
```



From the above Line graph, we observe people dying due to Tuberculosis is high in India there is a very slow decrease of death of people due to Tuberculosis in India, and for china and Indonesia, we observe a gradually decrease in the death ratio of people and till the year 2019 the remaining 4 countries except for India the number of death is less than 10000.

Cardiovascular Disease VS Country

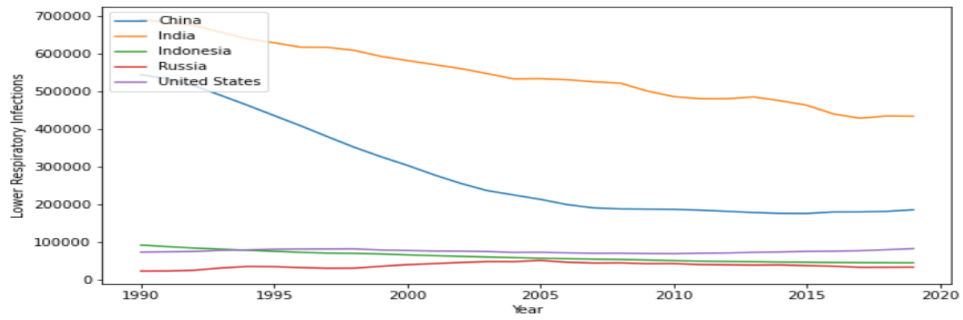
```
plt.figure(figsize=(10,5))
sns.lineplot(x='Year' , y='Cardiovascular Diseases' , data=result ,hue='Country/Territory')
plt.legend( loc ="upper left")
plt.show()
```



From the above Line graph, we observe people dying due to Cardiovascular Diseases are increasing in China from the Year 1990 to the year 2019, Same condition is in India but the number of people dying due to Cardiovascular Diseases is less as compared to china and for other 3 countries the count remains constant.

Lower Respiratory Infections VS Country

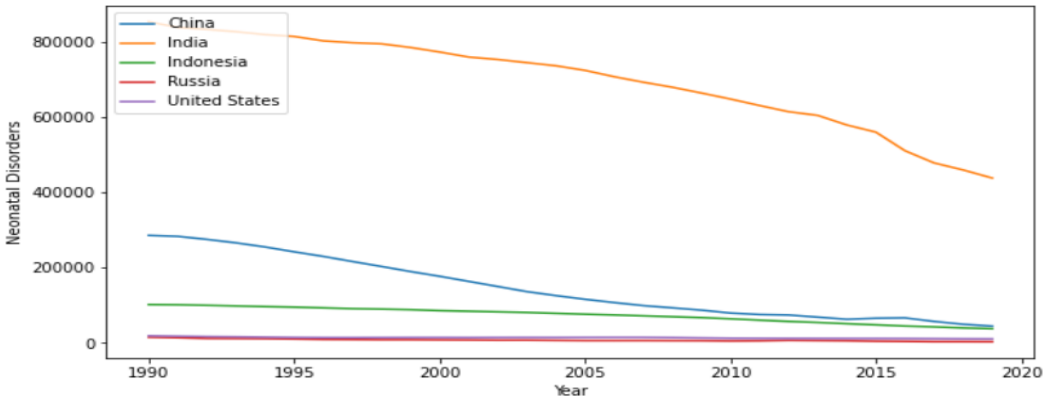
```
plt.figure(figsize=(10,5))
sns.lineplot(x='Year' , y='Lower Respiratory Infections' , data=result ,hue='Country/Territory')
plt.legend( loc ="upper left")
plt.show()
```



From the above Line graph, we observe people dying due to Lower Respiratory Infections was High in India and China in the year 1990 but as time passes it gradually decrease till the year 2019 for the other 3 countries the count remains constant, but for india if we observe the count for the year 2019 the people die due to Lower Respiratory Infections is very much high as compare to other 4 countries.

Neonatal Disease VS Country

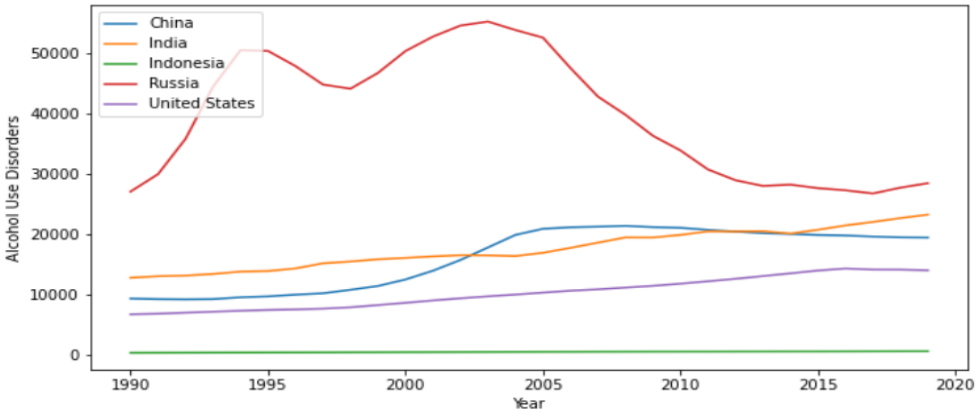
```
plt.figure(figsize=(10,5))
sns.lineplot(x='Year' , y='Neonatal Disorders' , data=result ,hue='Country/Territory')
plt.legend( loc ="upper left")
plt.show()
```



From the above Line graph, we observe people dying due to Neonatal Disorders was High in India and gradually decreasing, the count in china for starting the year 1990 was near about 300000 and the count decreased till the year 2019 which was less than 100000, and for other 3 countries the count remains constant, but for India, if we observe the count for the year 2019 the people die due to Neonatal Disorders is very much high as compare to other 4 countries.

Alcohol Use Disorders VS Country

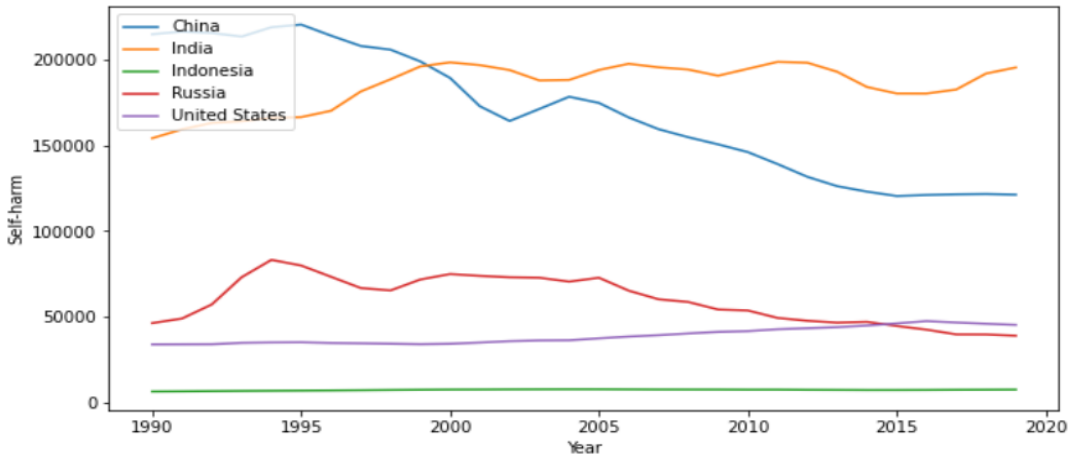
```
plt.figure(figsize=(10,5))
sns.lineplot(x='Year' , y='Alcohol Use Disorders' , data=result ,hue='Country/Territory')
plt.legend( loc ="upper left")
plt.show()
```



From the above Line graph, we observe people dying due to Alcohol Use Disorders were High in Russia, it remains COnstant but for the other 3 countries, it gradually increased, and for Indonesia it remains constant very less people die due to Alcohol use disorders.

Self-harm VS Country

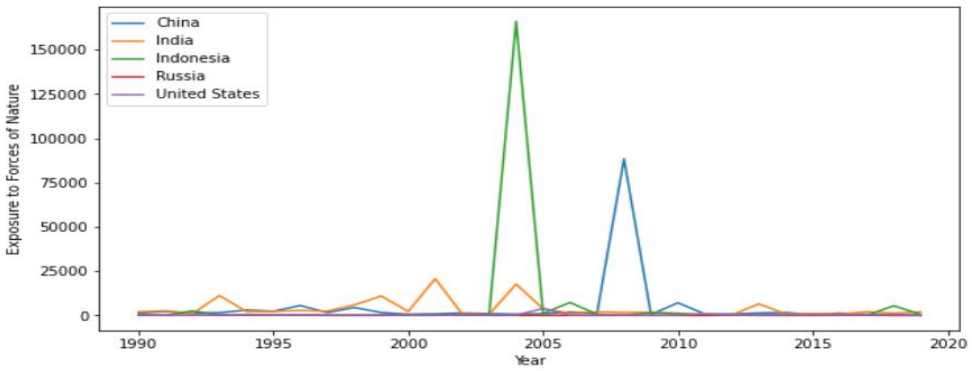
```
plt.figure(figsize=(10,5))
sns.lineplot(x='Year' , y='Self-harm' , data=result ,hue='Country/Territory')
plt.legend( loc ="upper left")
plt.show()
```



From the above Line graph, we observe people dying due to Self-harm is high in China and India for Russia it was increasing but later it decrease for the united states and Indonesia it remains constant.

Exposure to Forces of Nature VS Country

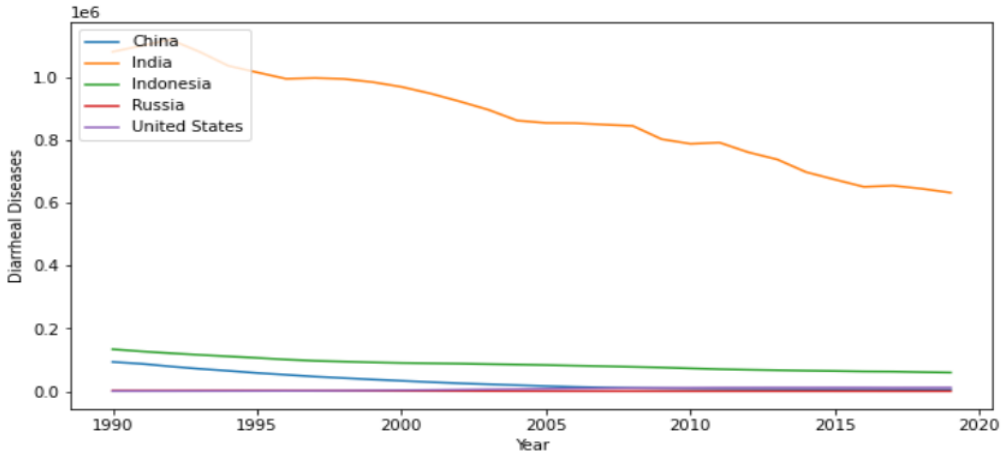
```
plt.figure(figsize=(10,5))
sns.lineplot(x='Year' , y='Exposure to Forces of Nature' , data=result ,hue='Country/Territory')
plt.legend( loc ="upper left")
plt.show()
```



From the above Line graph, we observe people dying due to Exposure to Forces of Nature is high in China and Indonesia for china in 2007 the people die most due to Exposure to Forces of Nature and for Indonesia in 2004 the people die most Exposure to Forces of Nature and for other 3 countries, it remains constant.

Diarrheal Disease VS Country

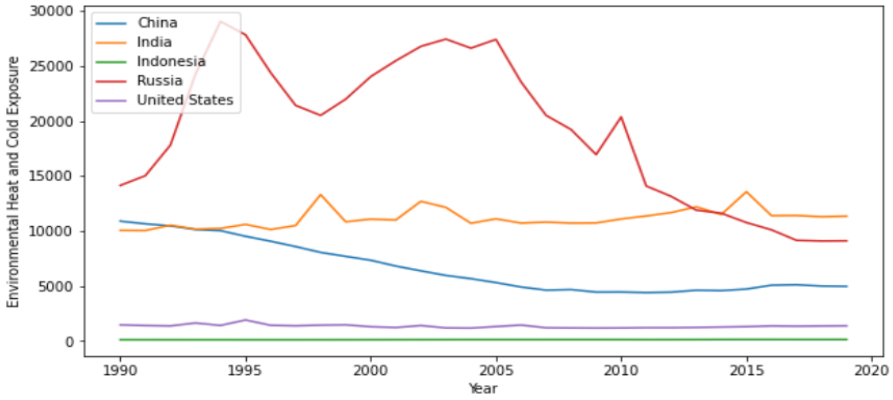
```
plt.figure(figsize=(10,5))
sns.lineplot(x='Year' , y='Diarrheal Diseases' , data=result ,hue='Country/Territory')
plt.legend( loc ="upper left")
plt.show()
```



From the above Line graph, we observe people dying due to Diarrheal Diseases was High in India and gradually decreasing, and for the other 4 countries the count remains constant, but for India, if we observe the count for the year 2019 the people die due to Neonatal Disorders is very much high as compare to other 4 countries.

Environmental Heat and cold Exposure VS Country

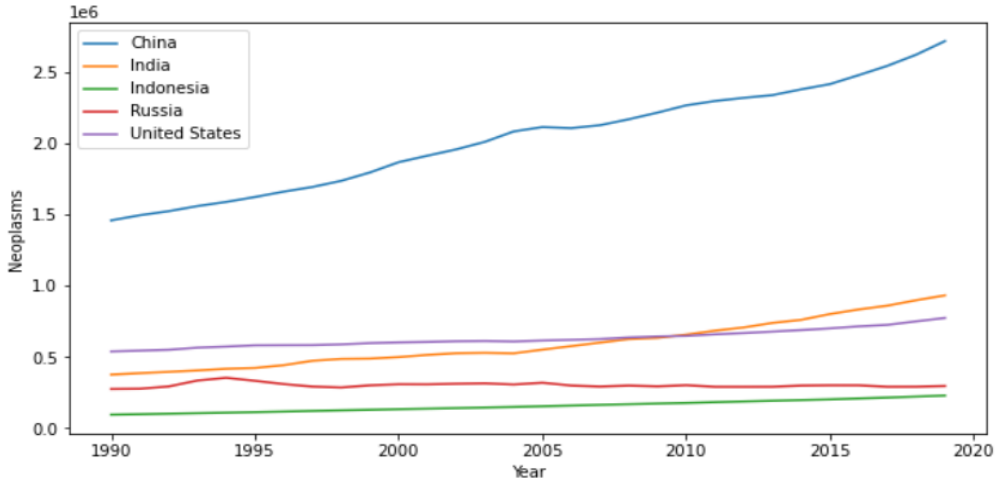
```
plt.figure(figsize=(10,5))
sns.lineplot(x='Year' , y='Environmental Heat and Cold Exposure' , data=result ,hue='Country/Territory')
plt.legend( loc ="upper left")
plt.show()
```



From the above Line graph, we observe people dying due to Environmental Heat and Cold Exposure was High in Russia and followed by India for china it is gradually decreasing, and for the remaining 2 countries it remains constant.

Neoplasms VS Country

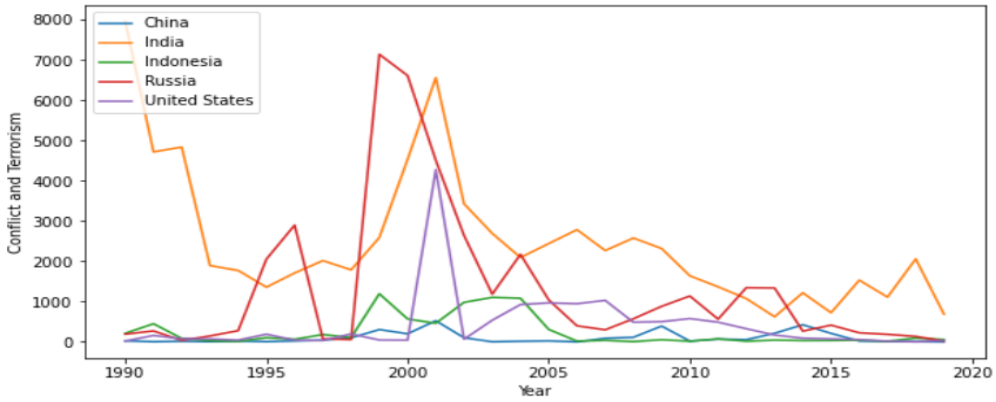
```
plt.figure(figsize=(10,5))
sns.lineplot(x='Year' , y='Neoplasms' , data=result ,hue='Country/Territory')
plt.legend( loc ="upper left")
plt.show()
```



From the above Line graph, we observe people dying due to Neoplasms are increasing in China from the Year 1990 to the year 2019, and for the remaining 4 countries it remains constant till the year 2010, and very little increase in death of people after the year 2010.

Conflict and Terrorism VS Country

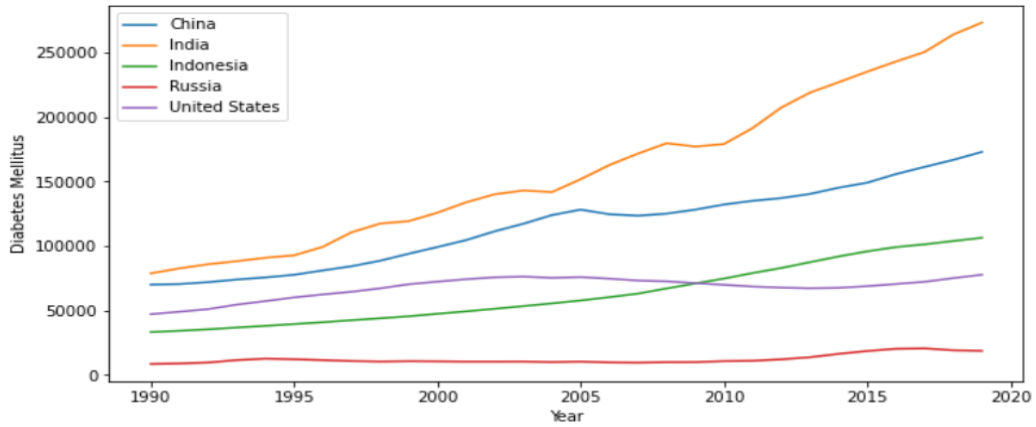
```
plt.figure(figsize=(10,5))
sns.lineplot(x='Year' , y='Conflict and Terrorism' , data=result ,hue='Country/Territory')
plt.legend( loc ="upper left")
plt.show()
```



From the above Line graph, we observe people dying due to Conflict and Terrorism is high in India and Russia for the united states the number of people dying due to Conflict and Terrorism was high in 2001 and after that, it decreased same for Indonesia and China.

Diabetes Mellitus VS Country

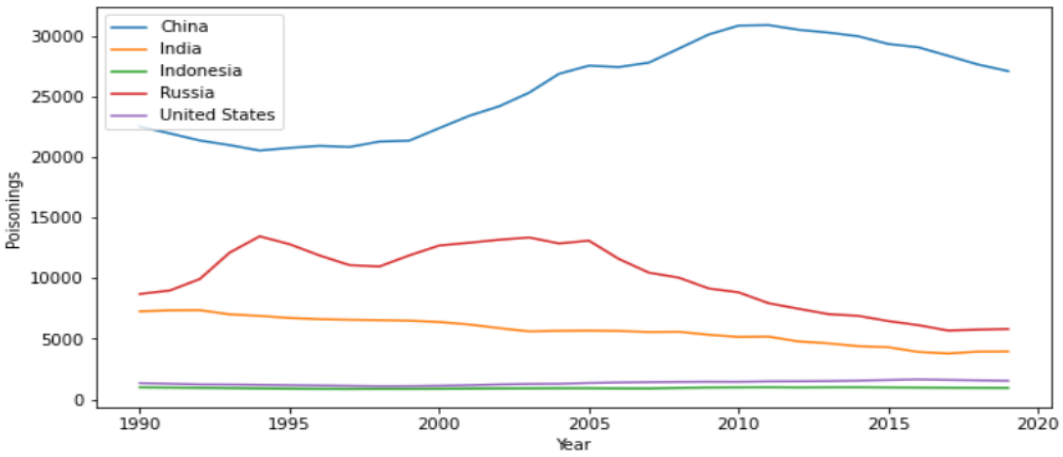
```
plt.figure(figsize=(10,5))
sns.lineplot(x='Year' , y='Diabetes Mellitus' , data=result ,hue='Country/Territory')
plt.legend( loc ="upper left")
plt.show()
```



From the above Line graph, we observe people dying due to Diabetes Mellitus was less for every country starting the year 1990 but then it increase every year till 2019 except for Russia, The most people dying due to Diabetes Mellitus was high India followed by china followed by Indonesia in the year 2019.

Poisonings VS Country

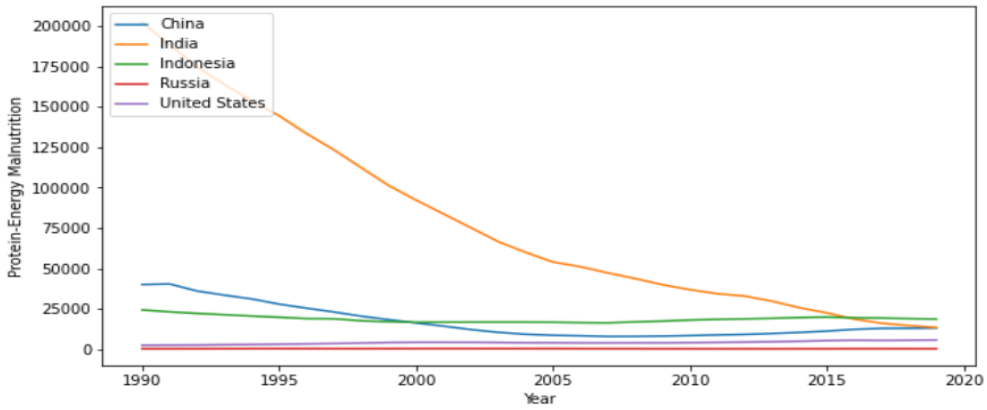
```
plt.figure(figsize=(10,5))
sns.lineplot(x='Year' , y='Poisonings' , data=result ,hue='Country/Territory')
plt.legend( loc ="upper left")
plt.show()
```



From the above Line graph, we observe people dying due to Poisoning was high in China and for Russia, it was high till the year 2005 after that it gradually decrease same with India for Indonesia and the united states it is very low and remains constant for all 30 years.

Protein-Energy Malnutrition VS Country

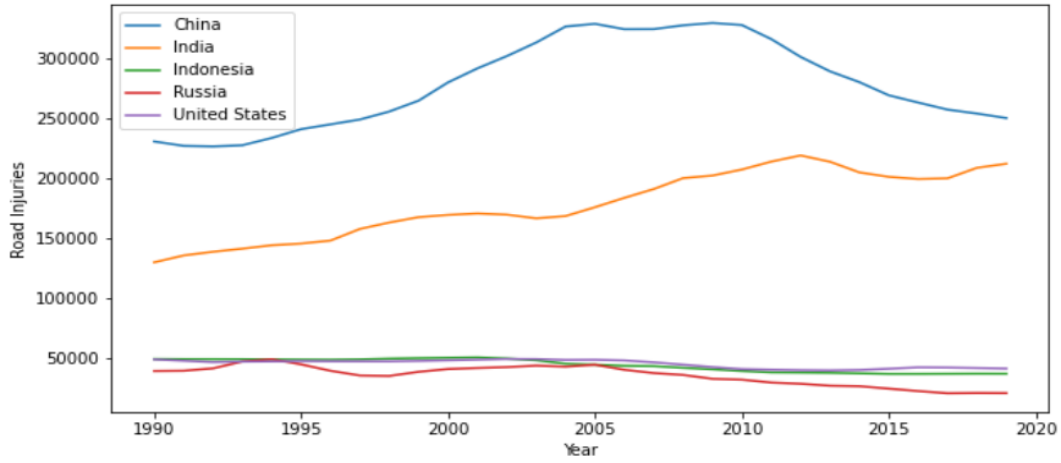
```
plt.figure(figsize=(10,5))
sns.lineplot(x='Year' , y='Protein-Energy Malnutrition' , data=result ,hue='Country/Territory')
plt.legend( loc ="upper left")
plt.show()
```



From the above Line graph, we observe people dying due to Protein-Energy Malnutrition was high in India but it rapid decrease and for all countries, the number of people dying due to Protein-Energy Malnutrition was less than 25000 in the year 2019.

Road Injuries VS Country

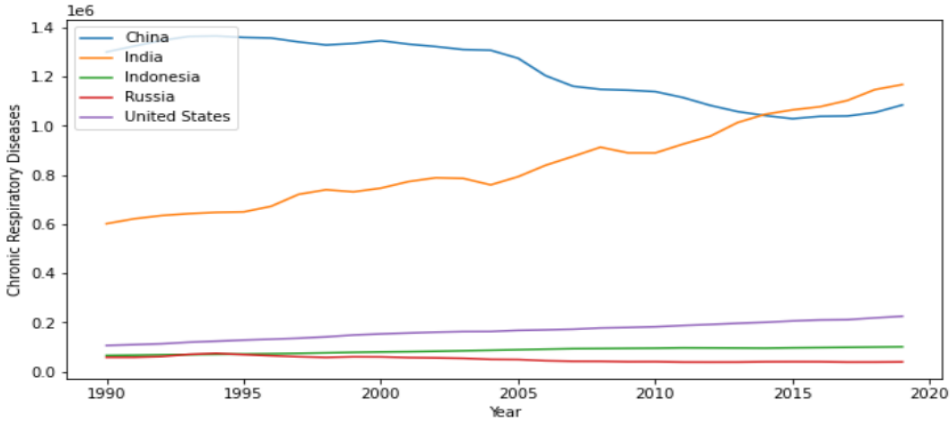
```
plt.figure(figsize=(10,5))
sns.lineplot(x='Year' , y='Road Injuries' , data=result ,hue='Country/Territory')
plt.legend( loc ="upper left")
plt.show()
```



From the above Line graph, we observe people dying due to Road Injuries were high in India and China and it gradually increased for the other 3 countries the number of people dying from Road Injuries was less and remains constant.

Chronic Respiratory Disease VS Country

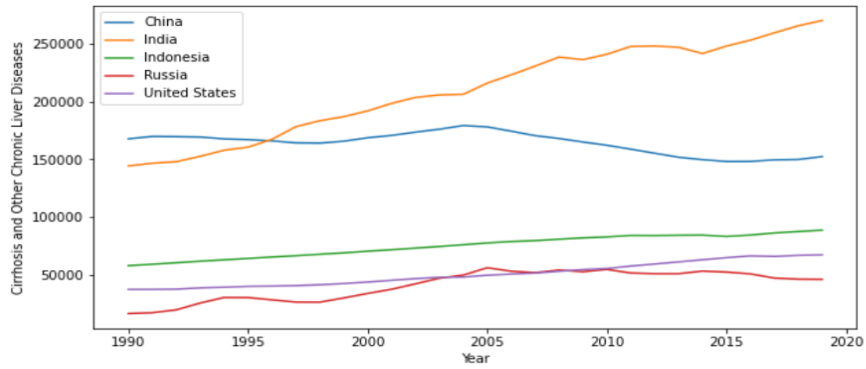
```
plt.figure(figsize=(10,5))
sns.lineplot(x='Year' , y='Chronic Respiratory Diseases' , data=result ,hue='Country/Territory')
plt.legend( loc ="upper left")
plt.show()
```



From the above Line graph, we observe people dying due to Chronic Respiratory Diseases were high in India and China, As for India it is gradually increasing and for china, after the year 2005 it is gradually decreasing and for the other 3 countries, it is increasing very slowly.

Cirrhosis And Other Chronic Liver Disease VS Country

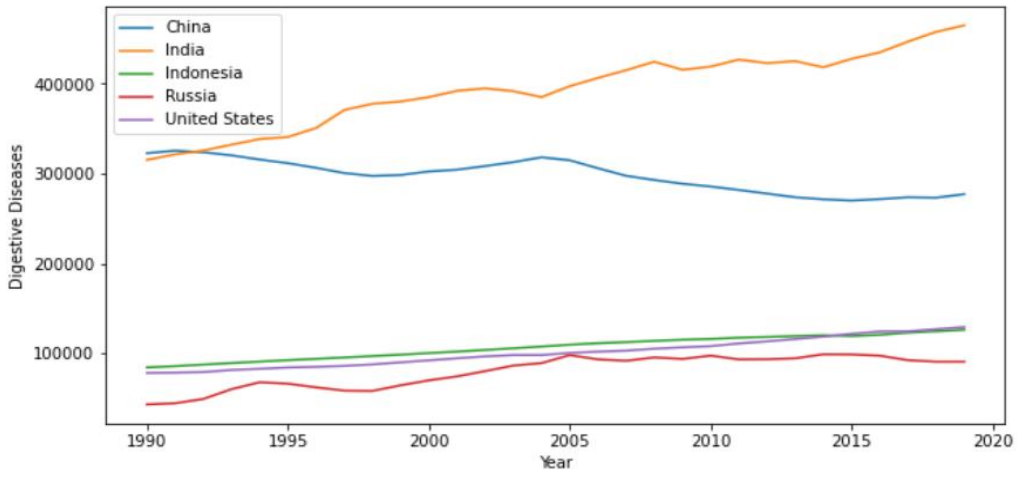
```
plt.figure(figsize=(10,5))
sns.lineplot(x='Year' , y='Cirrhosis and Other Chronic Liver Diseases' , data=result ,hue='Country/Territory')
plt.legend( loc ="upper left")
plt.show()
```



From the above Line graph, we observe people dying due to Cirrhosis and Other Chronic Liver Diseases were high in India and China, As for India it is gradually increasing and for china, after the year 2005 it is gradually decreasing and for the other 3 countries, it is increasing very slowly.

Digestive Disease VS Country

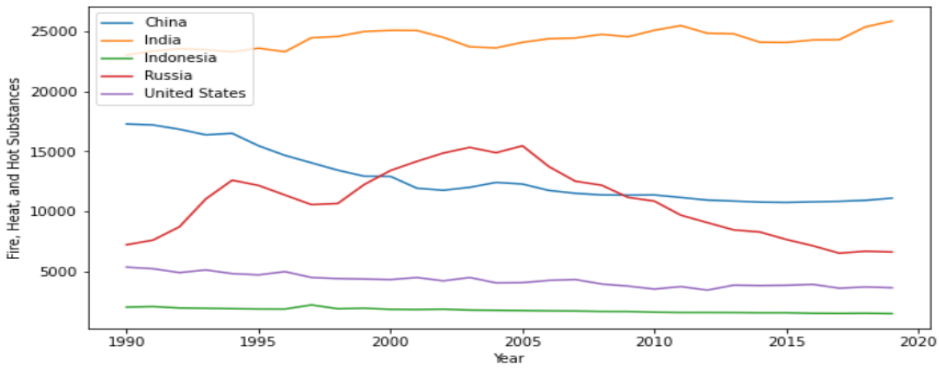
```
plt.figure(figsize=(10,5))
sns.lineplot(x='Year' , y='Digestive Diseases' , data=result ,hue='Country/Territory')
plt.legend( loc ="upper left")
plt.show()
```



From the above Line graph, we observe people dying due toDigestive Diseases were high in India and China, As for India it is gradually increasing and for china, after the year 1995 it is gradually decreasing and for the other 3 countries, it is increasing very slowly.

Fire , Heat and Hot Substances VS Country

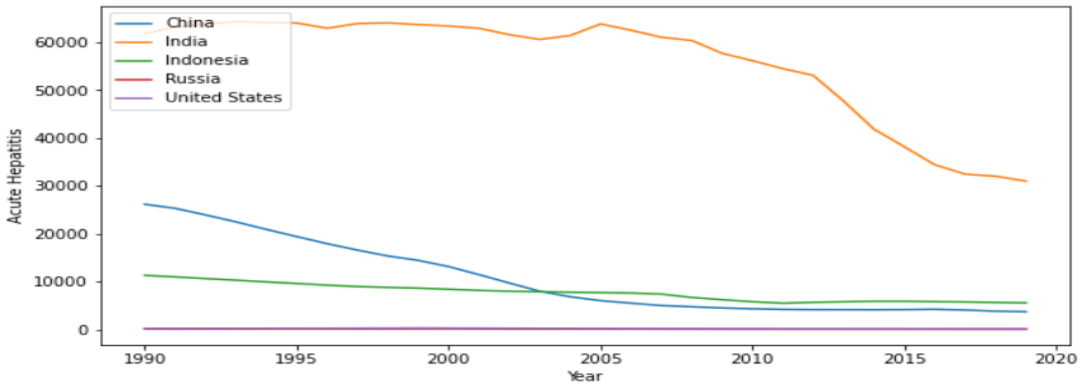
```
plt.figure(figsize=(10,5))
sns.lineplot(x='Year' , y='Fire, Heat, and Hot Substances' , data=result ,hue='Country/Territory')
plt.legend( loc ="upper left")
plt.show()
```



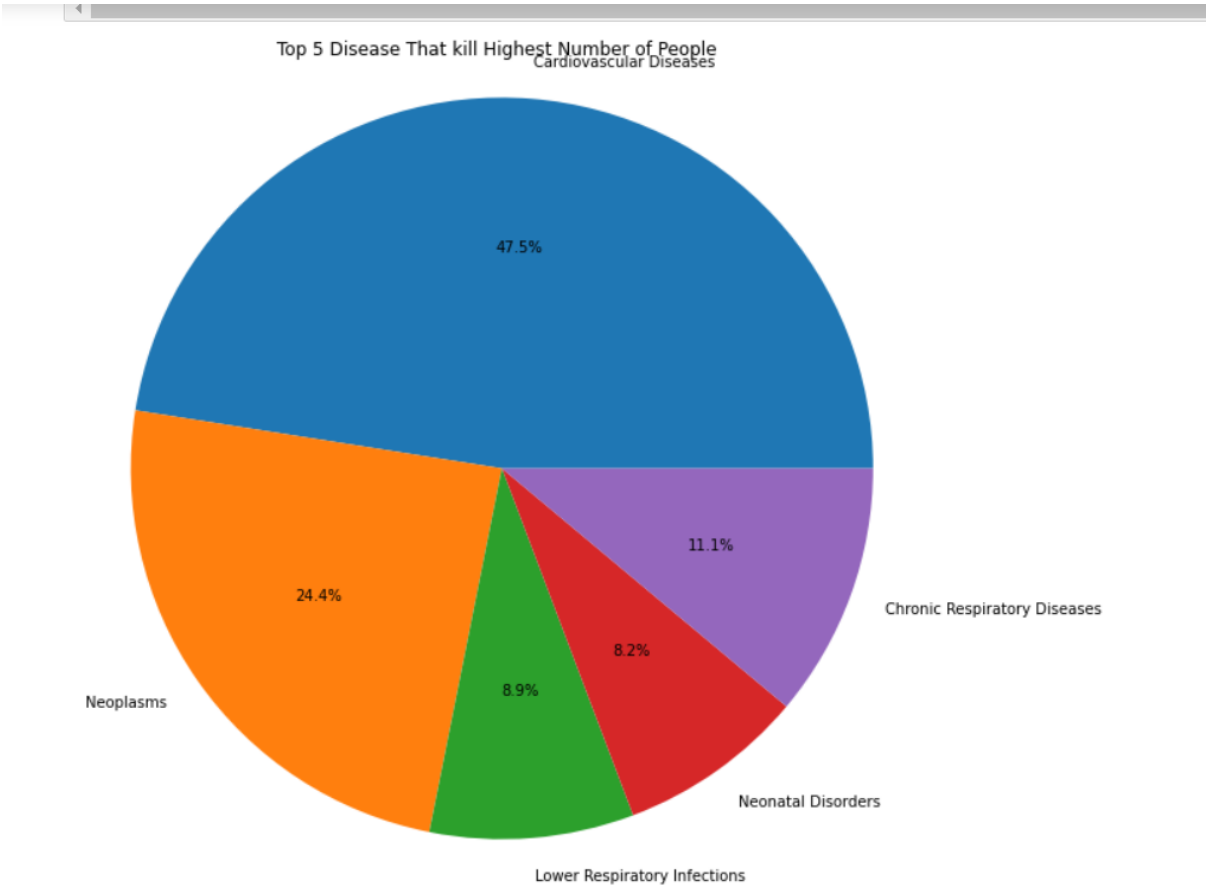
From the above Line graph, we observe people dying due to Fire, Heat, and Hot Substances were high in India and China, As for India it is gradually increasing and for china, after the year 1995 it is gradually decreasing and for Russia, it was increasing in the year 1997 but after the year 2005 it tends to decrease rapidly and for the other 2 countries, it is decreasing slowly.

Acute hepatitis VS Country

```
plt.figure(figsize=(10,5))
sns.lineplot(x='Year' , y='Acute Hepatitis' , data=result ,hue='Country/Territory')
plt.legend( loc ="upper left")
plt.show()
```



From the above Line graph, we observe people dying due to Acute Hepatitis was High in India and gradually decreasing, the count in china for starting the year 1990 was near about 25000 and the count decreased till the year 2019 which was less than 100000, and for other 3 countries the count remains constant, but for India, if we observe the count for the year 2019 the people die due to Acute Hepatitis is very much high as compare to other 4 countries.



The above Pie chart shows the Top 5 Diseases that kill a high number of people

1] Cardiovascular Disease 47.5%

2] Neoplasms 24.4%

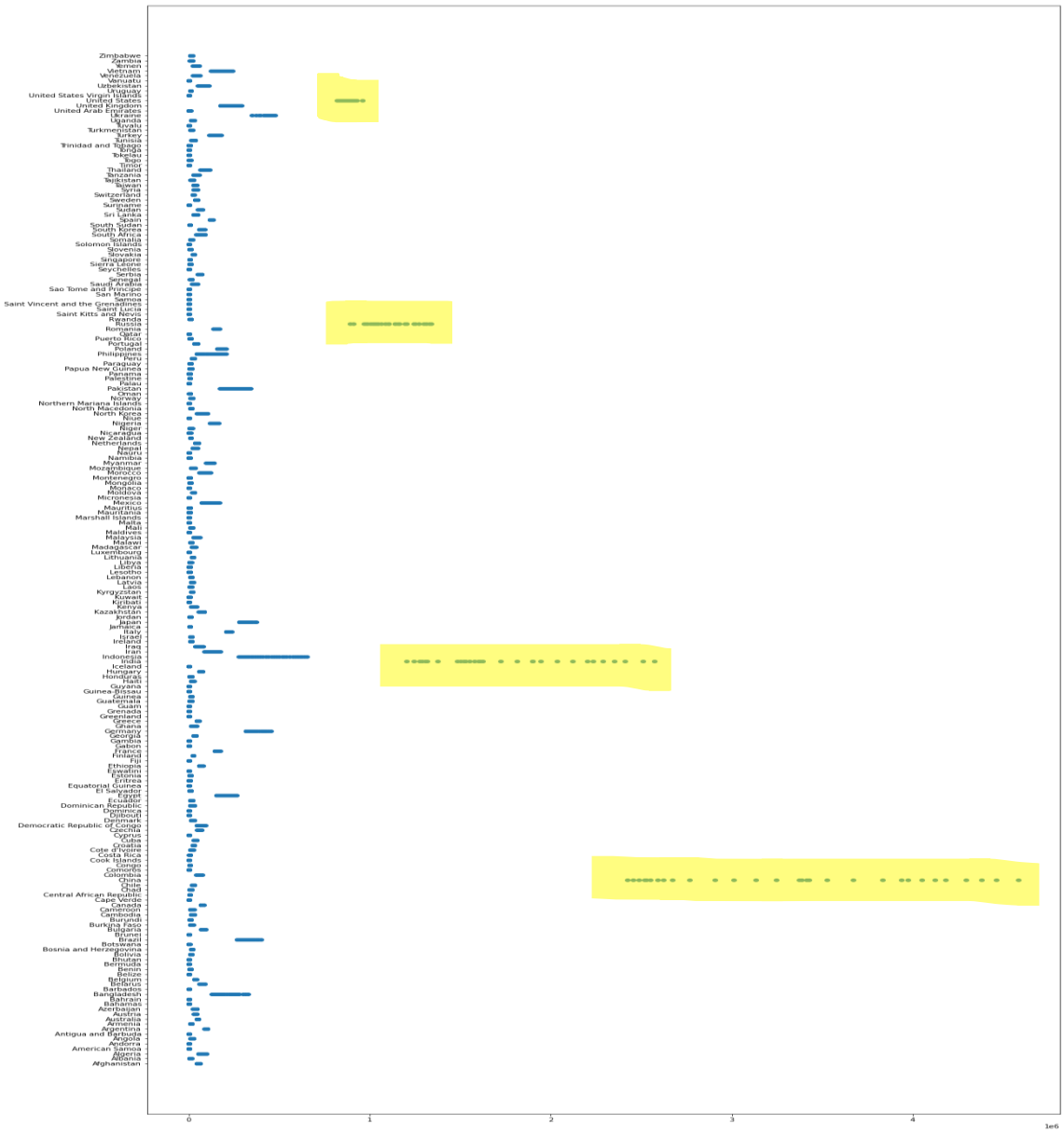
3] Chronic Respiratory Diseases 11.1%

4] Lower Respiratory Infections 8.9%

5] Neonatal Disorders 8.2%

Now we select these top 5 countries for further analysis in which Country it is most affected.

Country/Territory VS Cardiovascular Diseases

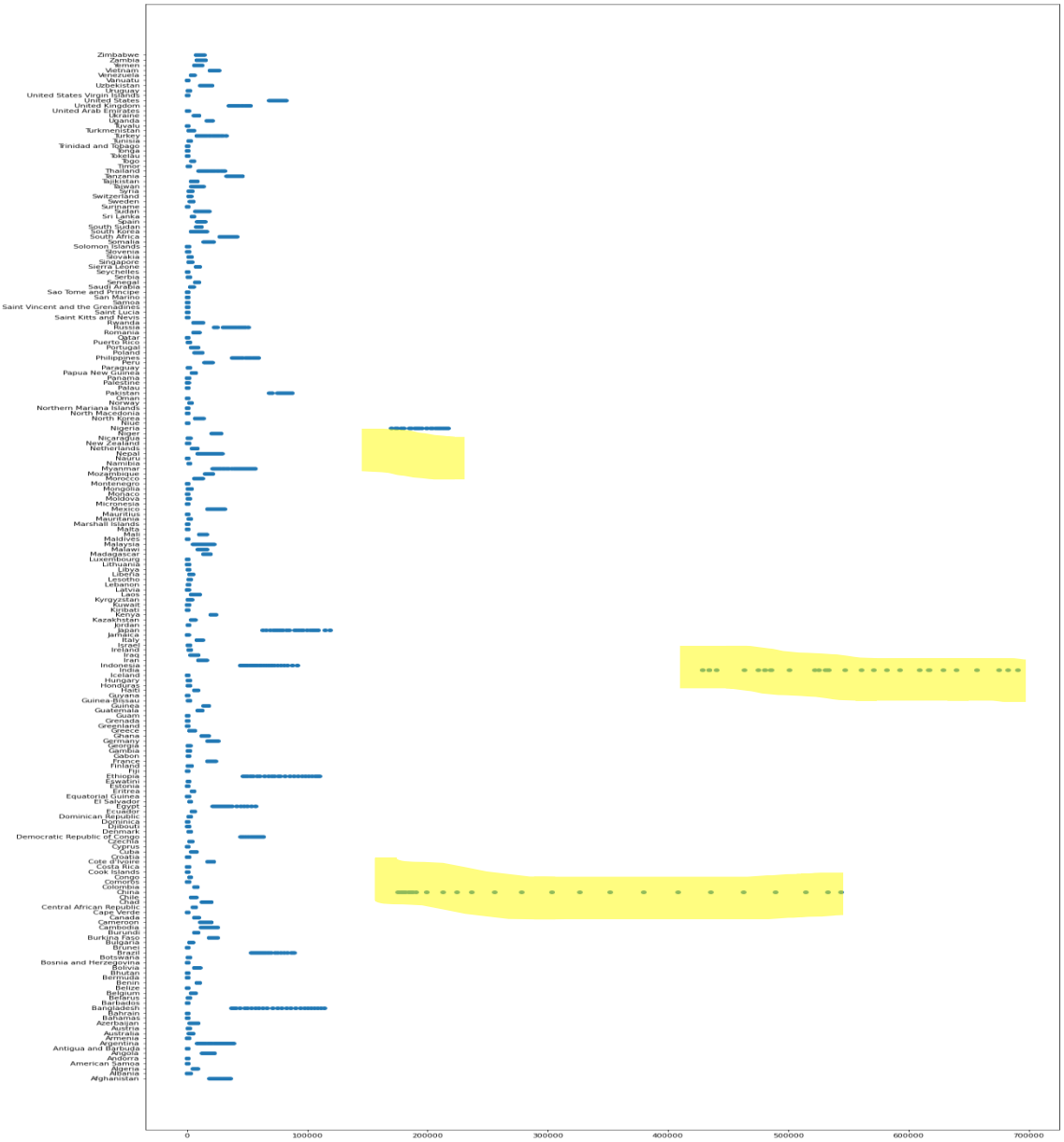


From the above Scatter plot (Country/Territory VS Cardiovascular Diseases) we observe the most number of people who died due to Cardiovascular Diseases in the countries like United States , Russia, India , China.

Country/Territory VS Chronic Respiratory Disease

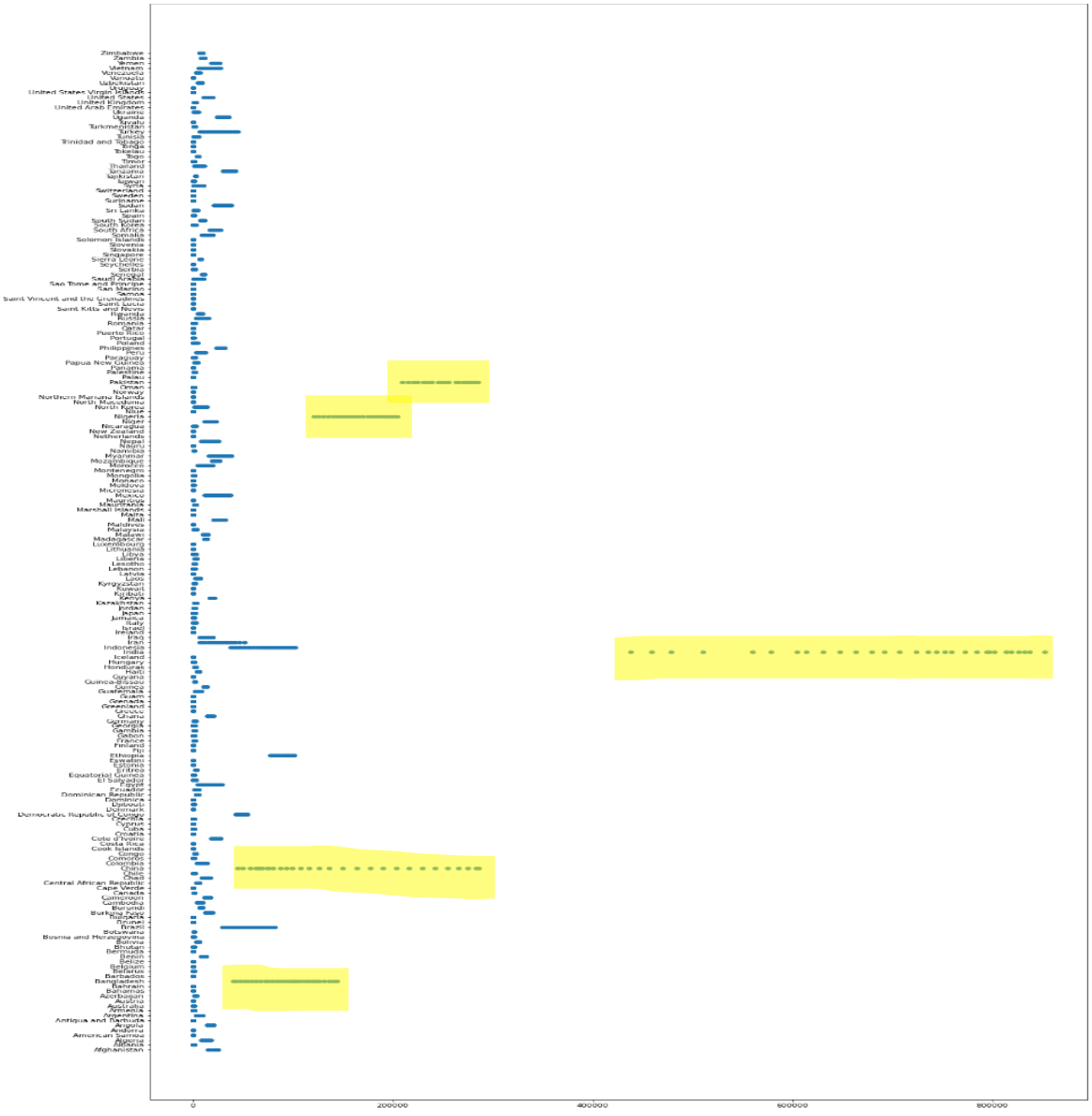
From the above Scatter plot (Country/Territory VS Chronic Respiratory Disease) we observe the most number of people who died due to Chronic Kidney Disease in the countries like United States, Pakistan, Mexico, Indonesia, India China, Brazil, and Japan.

Country/Territory VS Lower Respiratory Infections



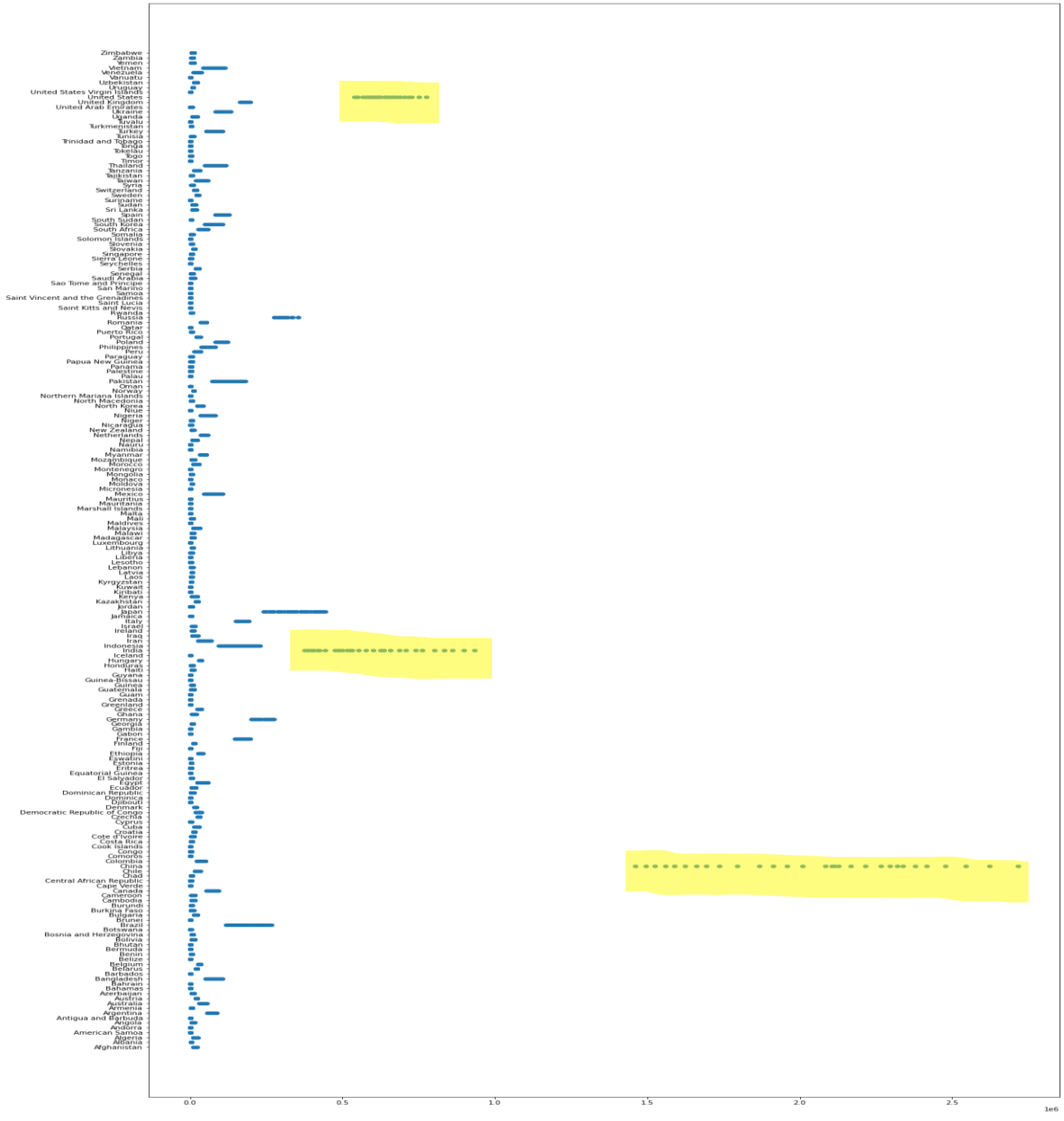
From the above Scatter plot (Country/Territory VS Lower Respiratory Infections) we observe the most number of people who died due to Lower Respiratory Infections in the countries like Nigeria, India , China.

Country/Territory VS Neonatal Disorders



From the above Scatter plot (Country/Territory VS Neonatal Disorders) we observe the most number of people who died due to Neonatal Disorders in the countries like Nigeria, Pakistan, India , China , Bangladesh.

Country/Territory VS Neoplasms



From the above Scatter plot (Country/Territory VS Neoplasms) we observe the most number of people who died due to Neoplasms in the countries like United States ,India , China.

Conclusion

1] From further analysis, we observe the top 5 country that has high mortality and morbidity that is China, India, Russia, Unites States, and Indonesia.

2] From further analysis, we observe the top 5 diseases that kill most of people in the country are Cardiovascular Diseases, Neoplasms, Chronic Respiratory Diseases, Lower Respiratory Infections, Neonatal Disorders, etc.

3] We also further analyzed the top 5 countries why the death of people is most from which disease each country.