Cause ofdeath

Project title

CAUSES OF DEATH

Submitted by

Mrs.Nupriya saxena

Fliprobo Sme:

Ms. Khushboo Garg

ACKNOWLEDGMENT

I would like to express my special gratitude to “Flip Robo” team, who has given me this opportunity to deal with a beautiful dataset and it has helped me to improve my analyzation skills. And I want to express my huge gratitude to Ms. Khushboo Garg (SME Flip Robo), she is the person who has helped me to get out of all the difficulties I faced while doing the project.

A huge thanks to “Data trained” who are the reason behind my Internship at Fliprobo. Last but not least my parents who have been my backbone in every step of my life.

References use in this project:

1. SCIKIT Learn Library Documentation

2. Blogs from towardsdatascience, Analytics Vidya, Medium

3. Andrew Ng Notes on Machine Learning (GitHub)

4. Data Science Projects with Python Second Edition by Packt

5. Hands on Machine learning with scikit learn

6. Sifei Lu, Zengxiang Li, Zheng Qin, Xulei Yang, Rick Siow Mong Goh, ”A Hybrid Regression Technique for House Prices Prediction”, @2017 IEEE, 2017 IEEE International Conference on Industrial Engineering & Engineering Management , Singapore DOI:10.1109/IEEM.2017.8289904

7. CH.Raga Madhuri, Anuradha G, M.Vani Pujitha “House Price Prediction Using Regression Techniques: A Comparative Study”, IEEE 6th International Conference on smart structures and systems ICSSS 2019

8. Zhongyun, Jiang, Guoxin, Shen, “Prediction of House Price Based on The Back Propagation Neural Network in The Keras Deep Learning Framework”, 2019 6th International Conference on Systems and Informatics (ICSA I2019)

. 9. Zhen Peng, Qiang Huang, Yincheng Han, “Model Research on Forecast of Second-Hand House Price in Chengdu Based on XGboost Algorithm”, 2019 IEEE 11th International Conference on Advanced Infocomm Technology 3 Surprise Housing - Housing Price Predication & Analysis Project

10. J.-G. Liu, X.-L. Zhang, and W.-P. Wu, “Application of fuzzy neural network for real estate prediction,” Advances in Neural Networks - ISNN 2006, vol. 3973, pp. 1187–1191, 2006.

+ 11. H. Kusan, O. Aytekin, and I. Ozdemir, “*e use of fuzzy logic ¨ in predicting house selling price,” Expert Systems with Applications, vol. 37, no. 3, pp. 1808–1813, 2010.

12. Ayush Varma, Abhijit Sarma, Rohini Nair and Sagar Doshi, ”House Price Prediction Using Machine Learning And Neural Networks”, @2018 IEEE, 2018 Second International Conference on Inventive Communication and Computational Technologies(ICICCT), Coimbatore, India, DOI:10.1109/ICICCT.2018.8473231.

13. Arietta, Sean M., et al. "City forensics: Using visual elements to predict non-visual city attributes." IEEE transactions on visualization and computer graphics 20.12 (2014): 2624-2633

Definitions: Cause of death vs risk factors

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'.[2]

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).

What do people die from?

56 million people died in 2017.[3] What did they die from?

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

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

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

Cause ofdeath

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

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

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

Our World in Data's causes of death dataset is examined in this repository. All codes are in the the notebook, please check the .ipynb file in the commits. Link for accessing the data:

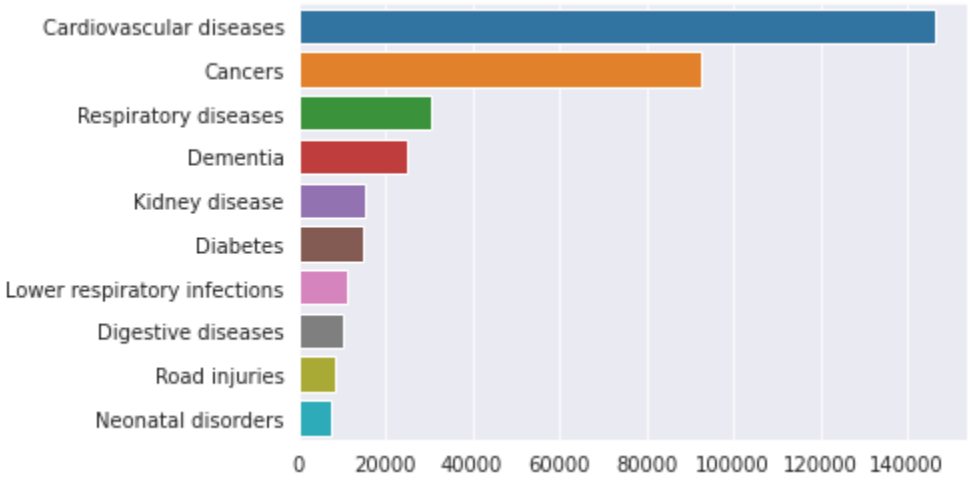
https://ourworldindata.org/causes-of-death

The dataset consists of 34 different cause of death counts by country per year. It covers the years from 1990 to 2017.

Let's start exploring the dataset.

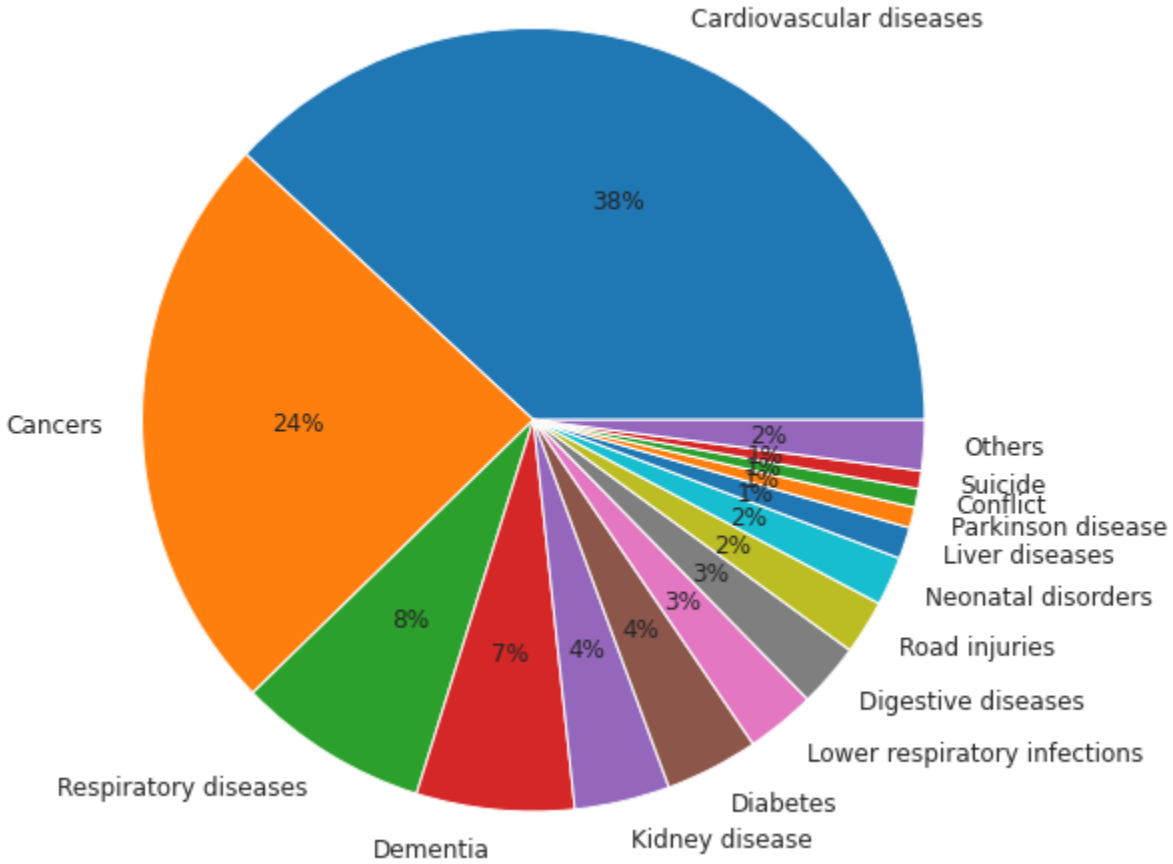
Top 10 causes of death for Turkey in the latest available year, 2017:

Cause ofdeath



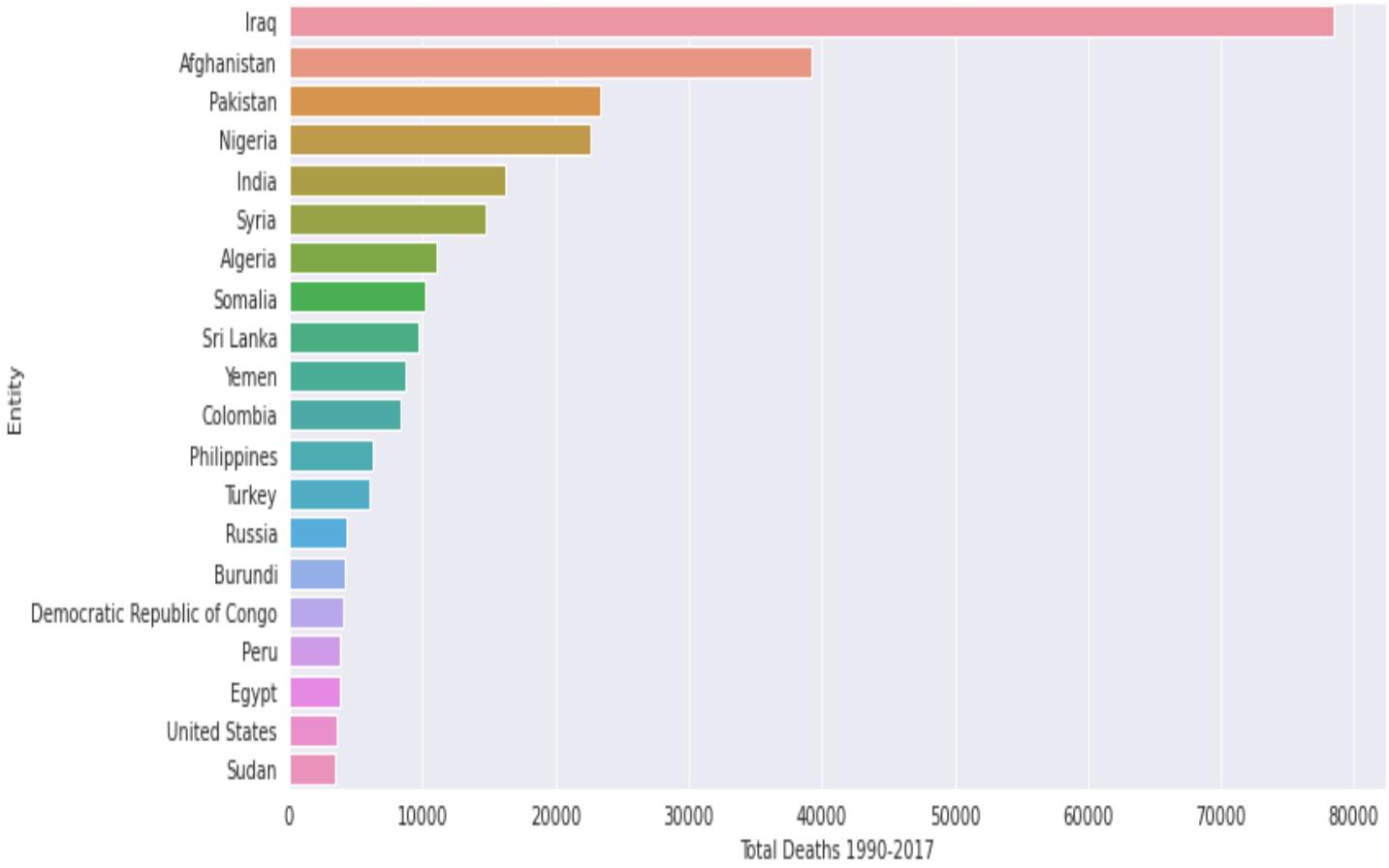
Percentages in pie graph:

- Almost 4 out 10 deaths in Turkey are caused by Cardiovasculare diseases in 2017

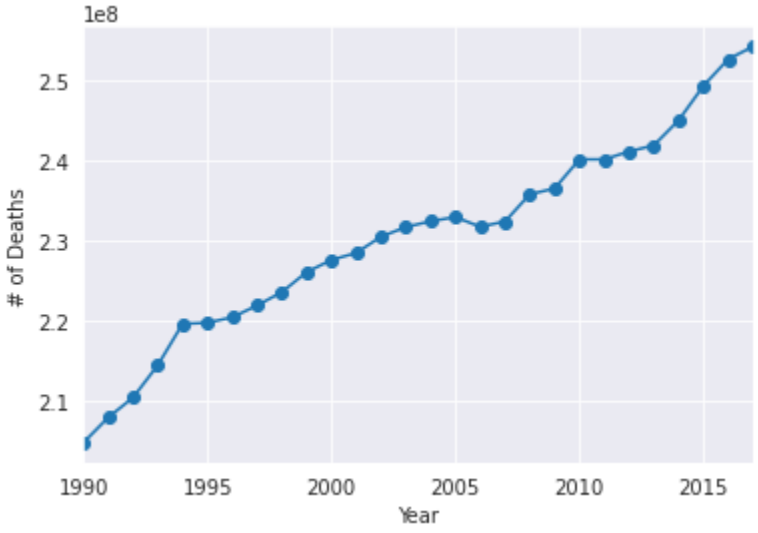


Cause ofdeath

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



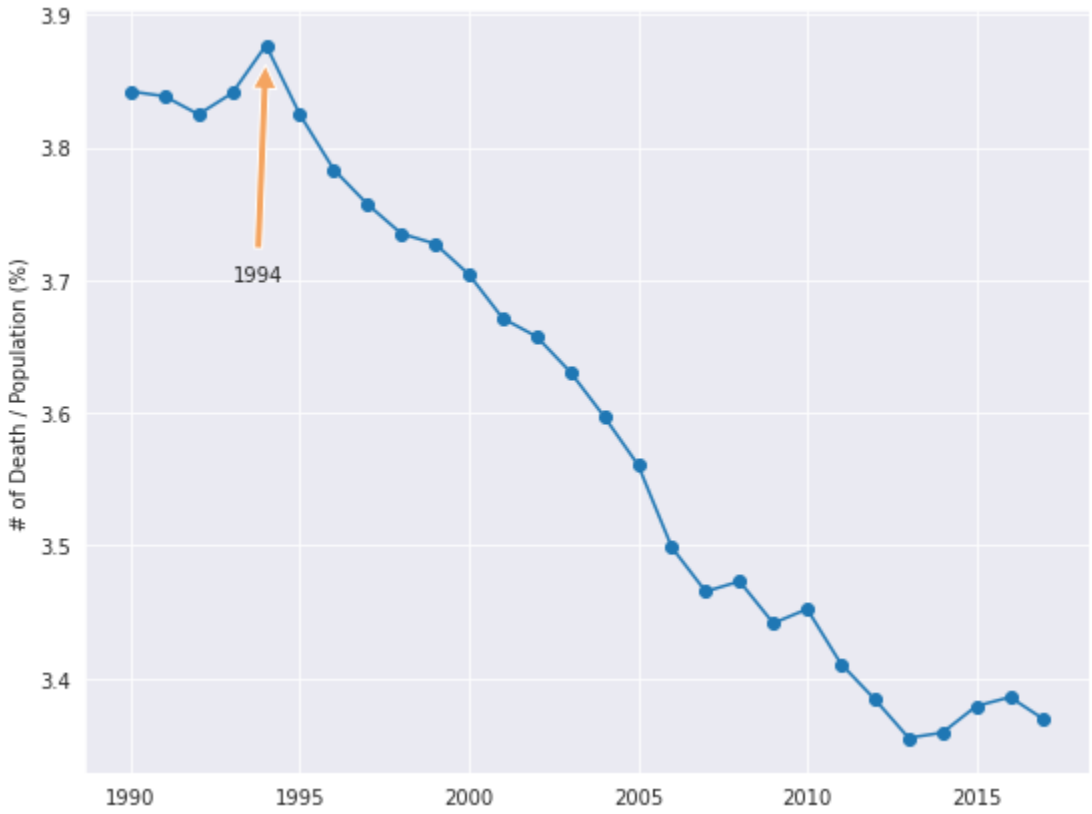
Deadliest year appears to be **2017**, however this is due to the increase of the world population each year. We need to check death rate per year.



Cause ofdeath

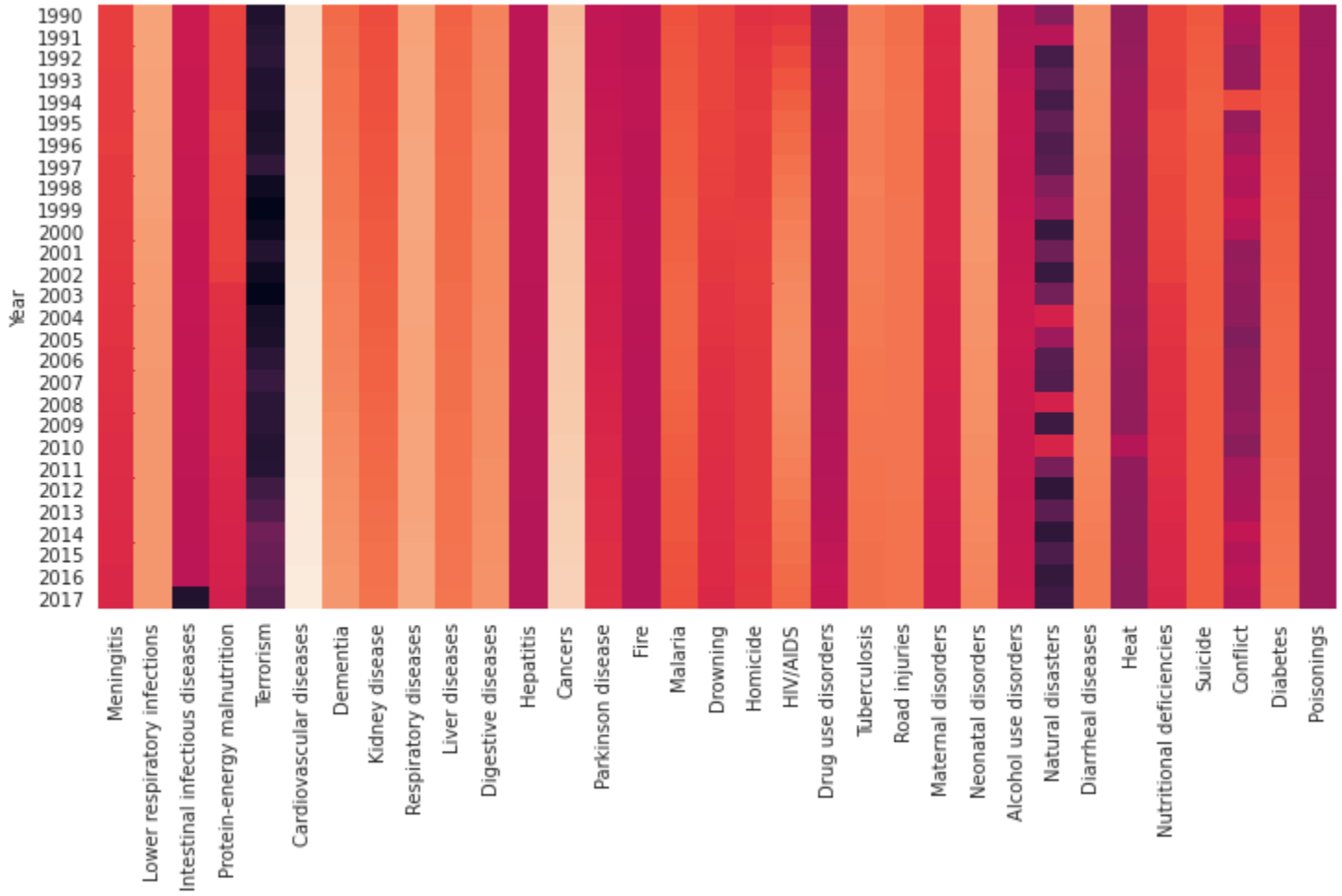
Find world population per year from the internet and plot year vs death rate (%):

- **1994** was the deadliest year with almost 4% of the population was died



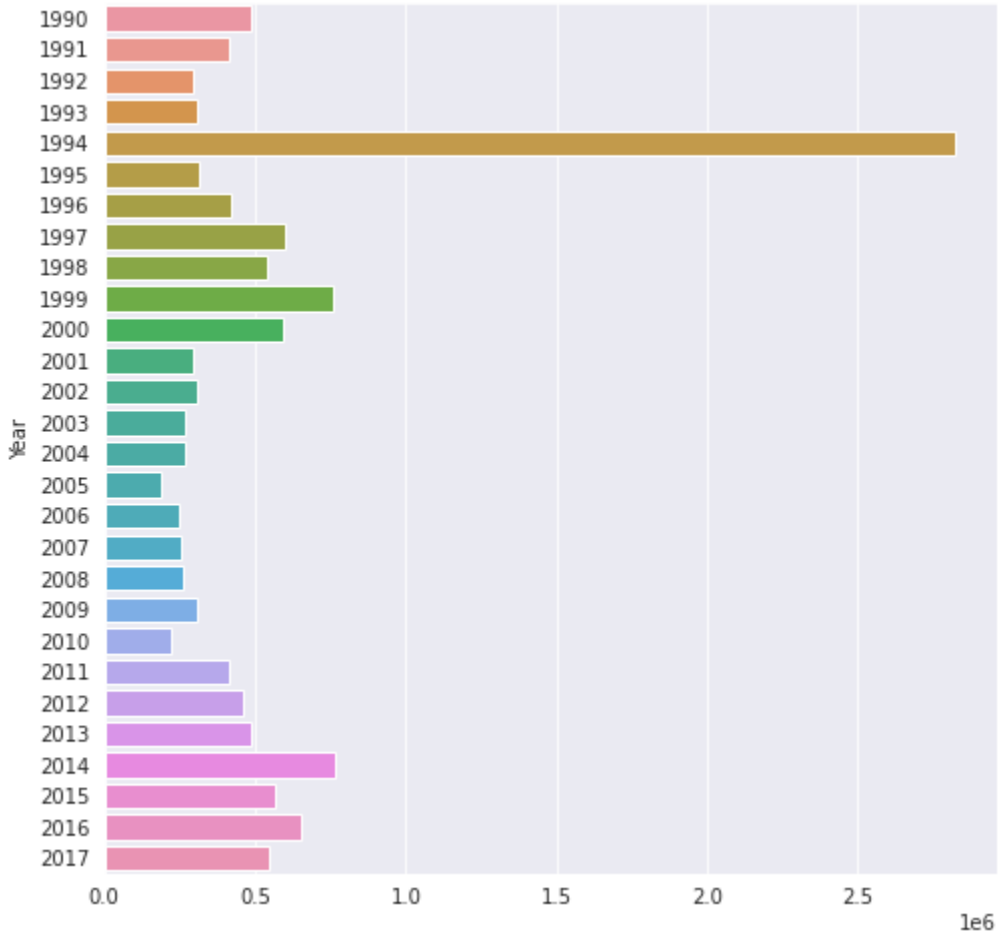
Lets examine what was the cause for this high death rate in 1994. Conflict line of the heatmap shows an interesting light colour on year 1994 (the lighter the colour the higher the number of deaths)

Cause ofdeath

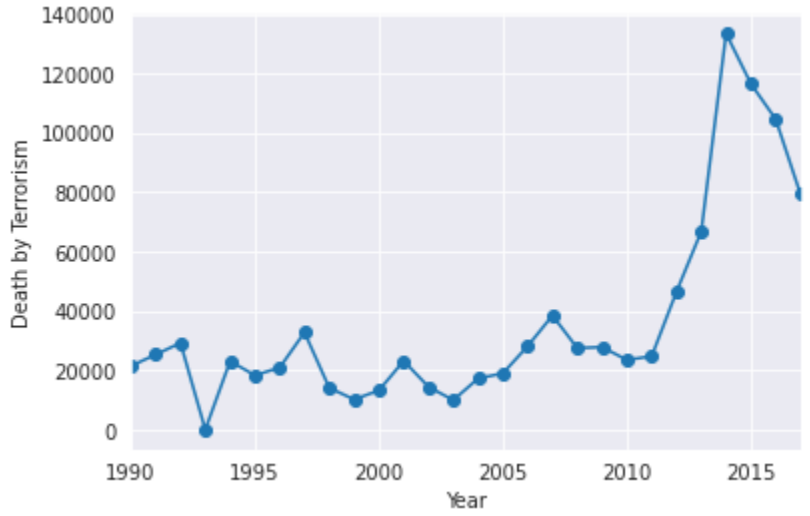


Conflict graph shows that **1994** was the year where the deadliest conflicts took place in the world. Bosnian War in Europe may have contributed to it significantly.

Cause ofdeath

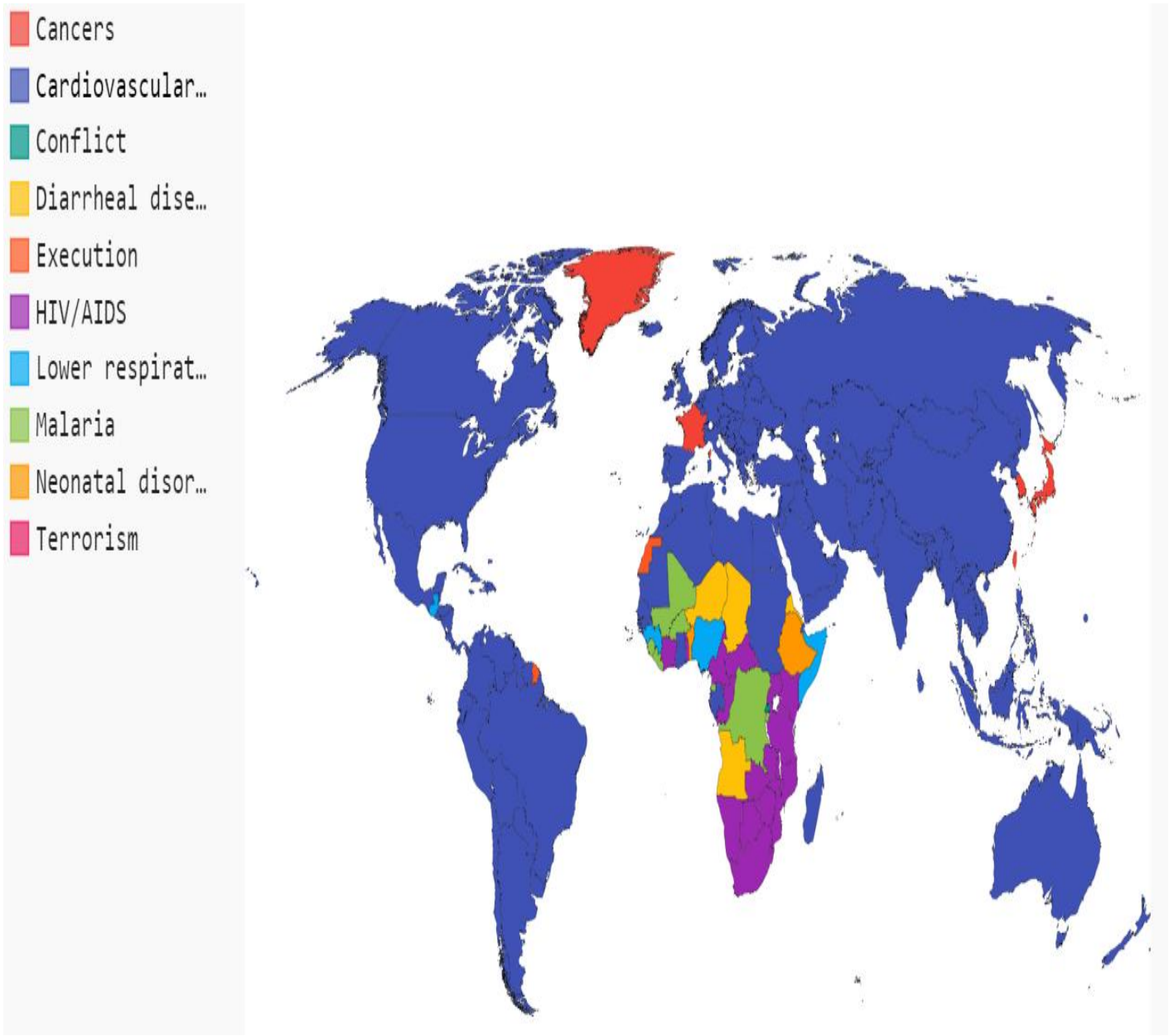


Death by terrorism peaked in 2014, and has been declining since then, however no effect was observed on 1994 death toll



Let's find out which country suffers from what disease the most

Cause ofdeath



It is obvious that Cardiovascular Disease dominate death toll all over the world, how would the World map look if we took them out?

- It appears that the second most leading cause of death in rich countries like USA, Canada, Europe, Japan and Australia is Dementia (mainly caused by Alzheimer's disease). This could be due to the high frequency of elder people in their population.
- **Conflict** is leading cause for war territories like Syria and Palestine
- For Russia and their neighbours like old Soviet Union countries, Eastern European countries leading cause is **Digestive diseases** like Ulcer, Cirrhosis, Hepatitis. It could be related to excessive consumption of alcohol.
- **Diarrheal** diseases causes deaths mostly in the mid African region

Cause ofdeath

- **HIV/AIDS** deaths are the most frequent in the South African region
- Some of the South American countries have **Homicide** as the leading cause for death
- **Lower respiratory infections** like Tuberculosis, Pneumonia is the leading cause for mainly South American countries like Brazil and Argentina.
- China, India and surrounding countries suffers from **Respiratory diseases** like Asthma and Lung Cancer etc.
- The countries where **road accidents** are one of the leading cause of death are in the Gulf region like Iran, Saudi Arabia, UAE. It could be related to their habit of car stunt driving.
- Greenland is the only country where **suicide** is the leading cause of death.

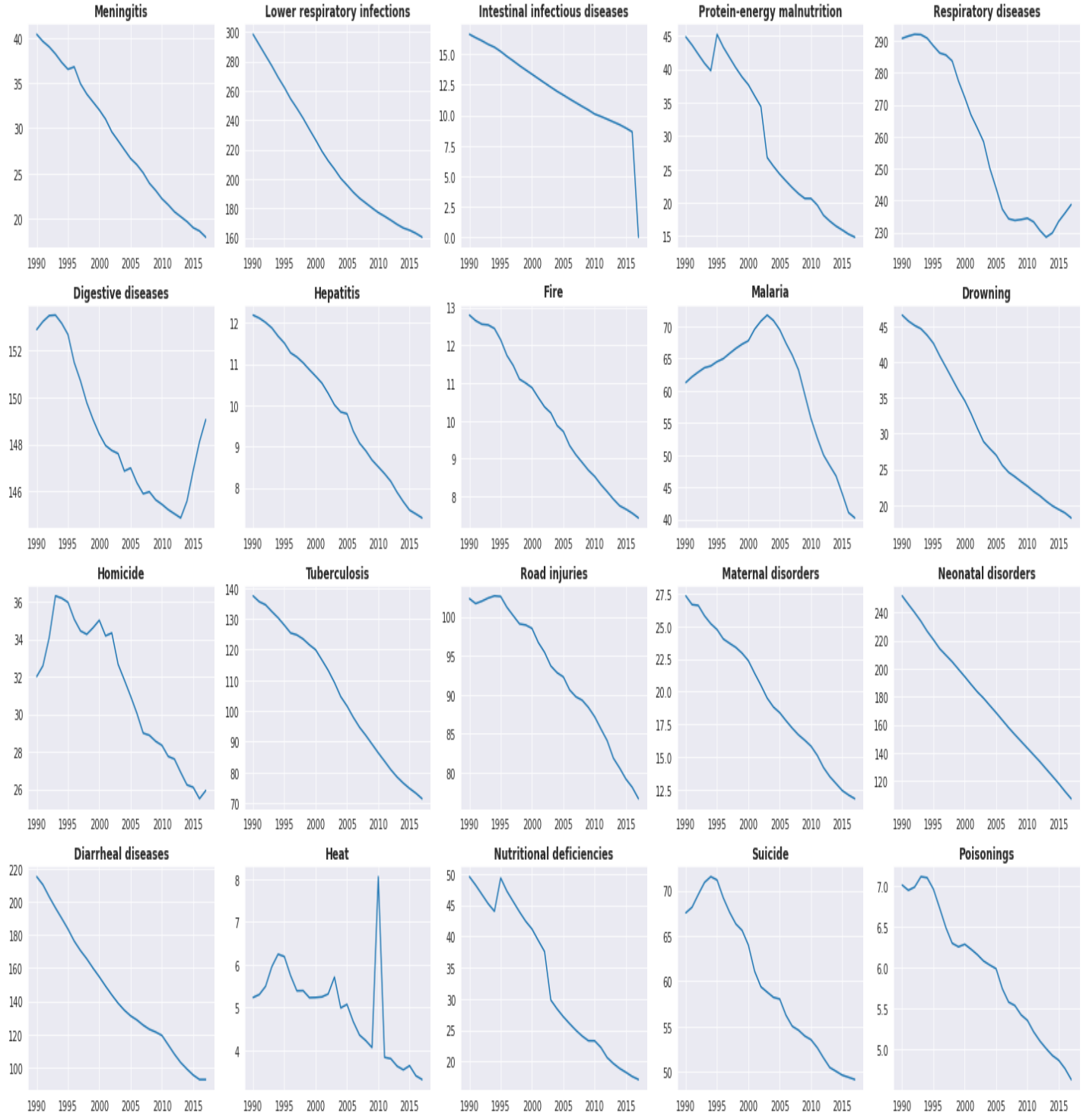
Let's now group the diseases by their historical increasing or decreasing trends. Calculations will be based on deaths per 100000 for particular disease.

Cause ofdeath

Following diseases show monotonic decrease trend over years:

- Even though Digestive and Respiratory diseases were dropping constantly for period of time, they started to enter an increasing trend since 2013. Still way better as compared to 90s
- Suicide rates are dropping since mid-90s
- Peak in the Heat graph could be a mistake in the dataset or it is a very distinguishing event happened in 2010 which caused this sudden jump

Cause ofdeath



Following are the death causes exhibiting an increasing or no trend over time:

- Dementia, Kidney, Cancer, Parkinson, and Diabetes diseases have been increasing since the beginning of our data span (1990)
- Drug and alcohol use disorders are correlated with Liver disease trend
- HIV/AIDS related deaths rate started to decline since 2005

Cause ofdeath

- Natural disasters show no trend as expected