

# Project

## Analyzing Data of death causes in the world between 1990- 2017

Ayub Pathan  
Remote Bootcamp, Ironhack, May 2021

### 1. Introduction:

It is important to understand what are the causes behind death and find out relevant reasoning for that.

There are several causes in the world but apart from that certainly there are main dominating factors those are contributing more than 50% of death over the world, such as NCDs, i.e. Non Communicable Diseases.

Apart from infectious and non- infectious diseases, we must consider other aspects such as natural disasters, injuries, forest fires, and so on.

In this project work, I have considered a data comprising over 27 years (from 1990 till 2017).

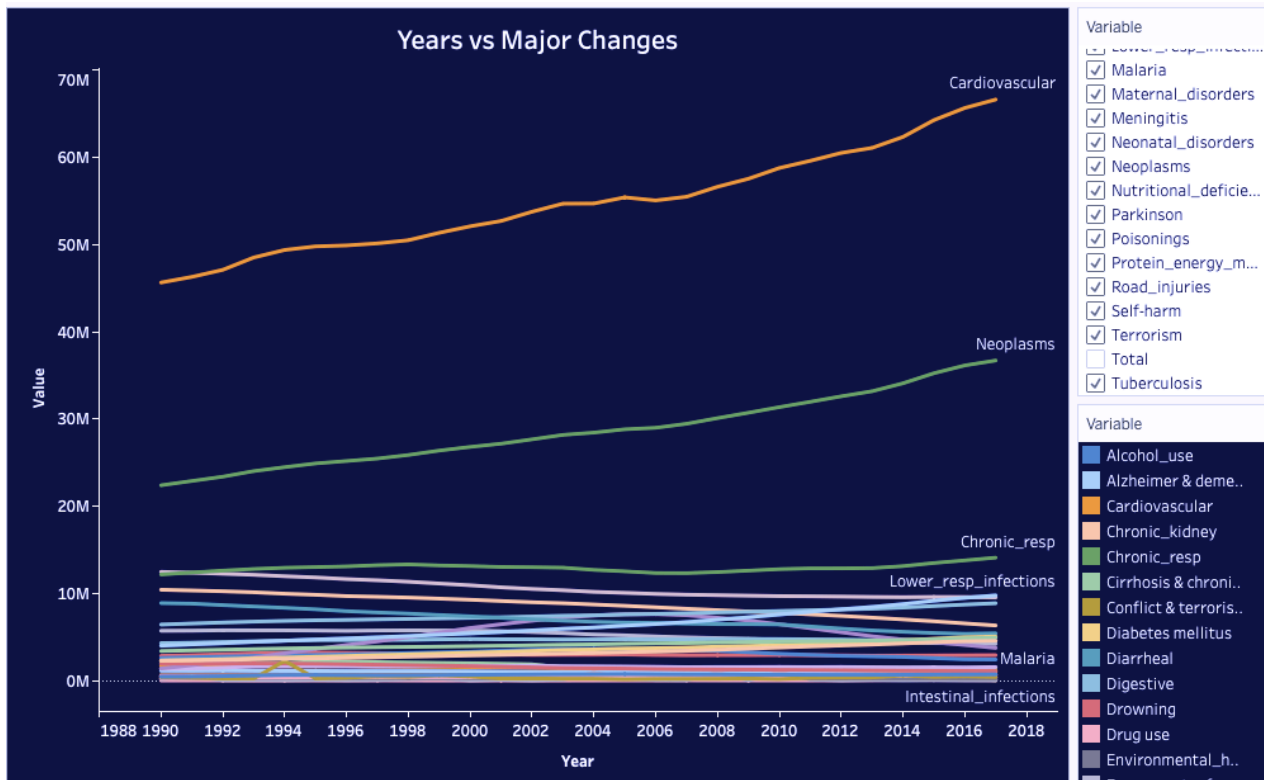
### 2. Data:

To work on relevant subject, I have selected two relevant datasets from open source data webpage\*. Both datasets are of 27 years from 1990 till 2017. Here is an overview of our dataset 1 (after filtering):

	Year	Entity	Meningitis	Lower_resp_infections	Intestinal_infections	Protein_energy_malnutrition	Terrorism	Cardiovascular	Alzheimer & dementias	Chronic_kidney
0	1990	Afghanistan	6470.0	22837.0	295.0	1608.0	12.0	46498.0	1959.0	3155.0
1	1990	Albania	103.0	1960.0	0.0	33.0	0.0	6472.0	512.0	253.0
2	1990	Algeria	731.0	6178.0	254.0	254.0	3.0	42528.0	2302.0	2048.0
3	1990	American Samoa	1.0	12.0	1.0	2.0	0.0	66.0	4.0	9.0
4	1990	Andean Latin America	1526.0	36291.0	12.0	6575.0	0.0	40340.0	5101.0	5281.0

This dataset considers various causes of death such as cardio, neoplasm, HIV, natural disasters, terrorism, and various types of communicable and non-communicable.

Here is a line plot showing all causes over the years:



### 3. Goal:

Goal of this project was to learn at self pace to analyse selected dataset and work out to achieve answers for the raised questions as well as to make them presentable in the form of visualization.

Apart from that, purpose of analyzing these datasets is to understand them thoroughly and do some hypothesis testing as well as to find whether we can achieve answers to our doubts.

### 4. Analysis:

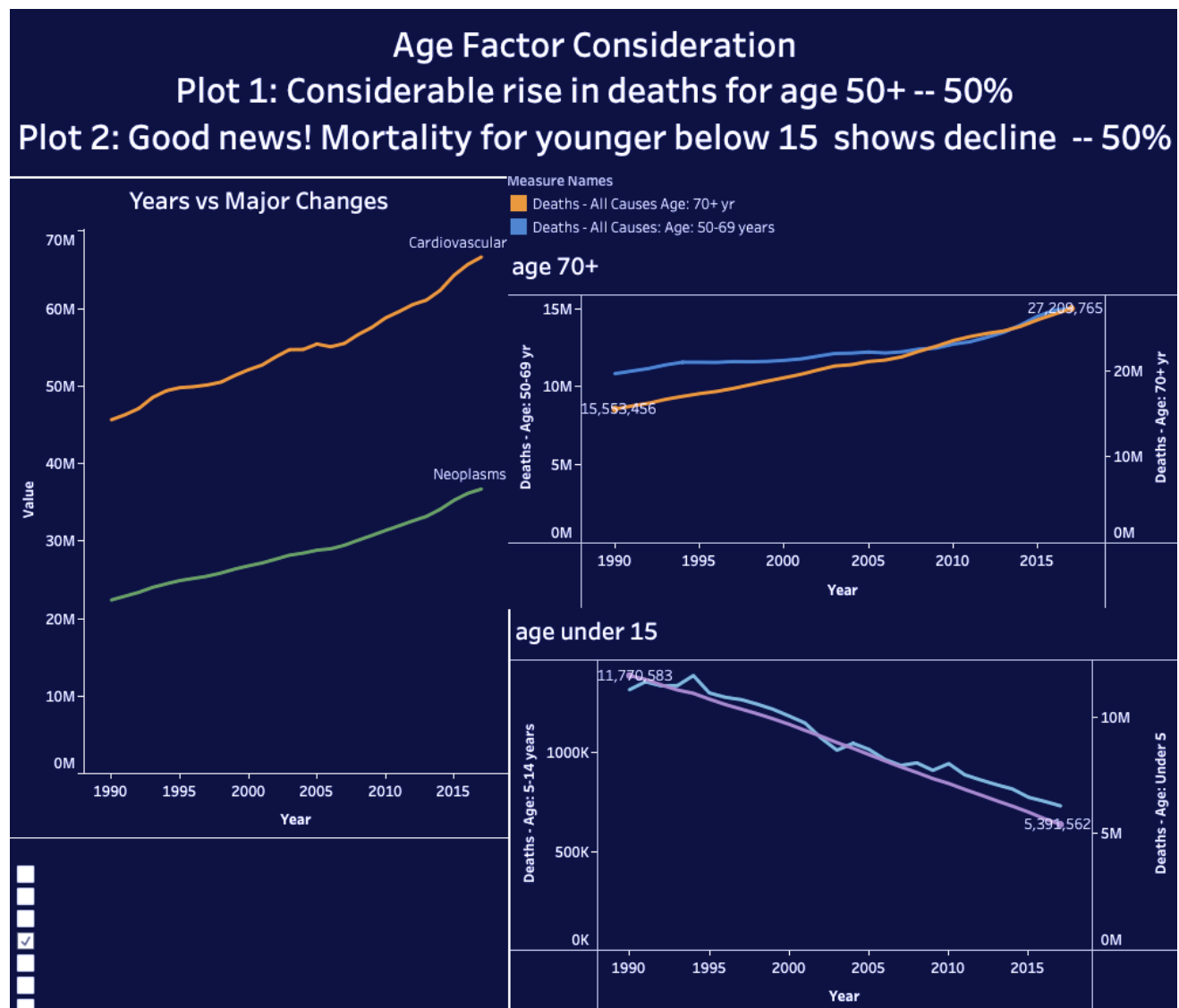
The below procedure was followed in analysis of the data:

- Data mining & understanding
- Data filtering & cleaning
- Finding of nan values if any & their relevancy
- Filtering World data so that not to misinterpret whole dataset with world values
- Melting all filter datasets and saving them separately in the csv files
- Transferring all filtered data to tableau to visualize the results
- Making World Map to figure out special cases of conflicts and natural disasters

### 5. Results:

As this dataset was huge and hence it is not possible to add all the results here which is shown in my presentation. However, I would like to add few of the important findings here:

## Major Contributors and their relevancy with Age:



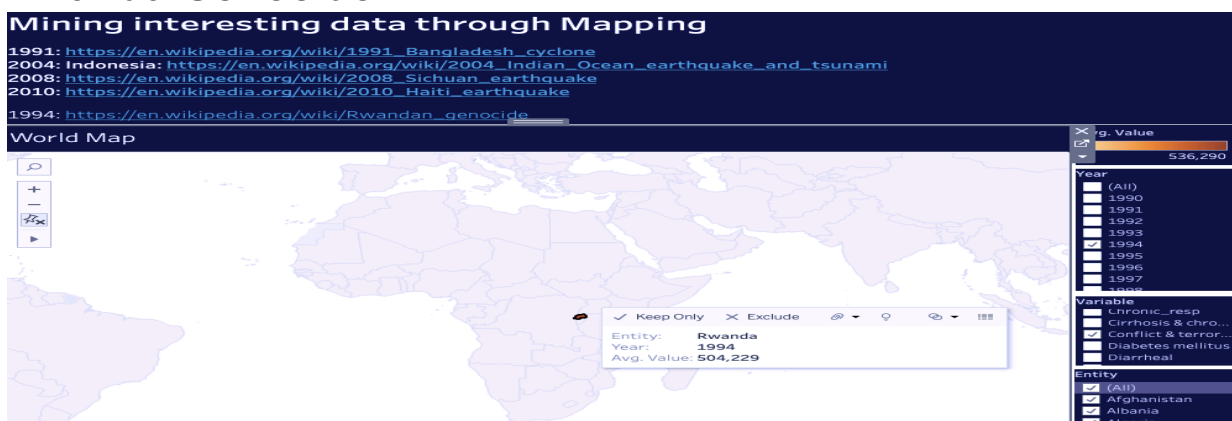
From the major contributors findings, it was clear to see there is relevancy to age and hence I decided to analyse Age dataset\* and found out that 50+ aged people are more prone to die because of cardiovascular as well as Neoplasm. However, in same finding, we can see in plot 2 (Age under 15), there is drastic decline of mortality rate for youngsters less than 15 years. The reason could be said that the improvement in science as well medical researches lead to this change over 27 years.

## Relative change over 27 years for all causes:



From the relative change, we can clearly see that there was a drastic increment in NCD as well as terrorism over these 27 years. This is an alarming fact that all NCD are on rise but terrorism is one of the largest peak in the result. However, good news was that there was 80% decline in natural disasters. Brain related diseases also one of biggest concern to be considered in this modern life structure.

## Rwanda Genocide:



## 6. Conclusion:

From all results, it can be satisfactorily said that all the question were answered through analysing data

- <https://ourworldindata.org/>