



# Global Health Analysis

- *Look at healthcare over the world, over the years...*

Antara Bahursettiwar - U1209537

Harshitha Manduva - U1080236

Neha Kherde - U1205893

Link: <https://harshi-m.github.io/GlobalHealthAnalysis/>



## Contents:

1. Overview and Motivation
2. Related Work
3. Questions
4. Data
5. Design Evolution
6. Implementation
7. Evaluation
8. Resources

## Overview and Motivation

With the data, we are trying to build a better visualization of world health, how that is changed over the years and to find the co-relations between the various related health-related factors. The aim is to give a global overview and to show changes over the very long run so that we can see where we are coming from and where we are today.

Data has become increasingly valuable in the 21st century and new economies have been shaped by who controls it health data and the healthcare industry are unlikely to be an exception. Health data can be used to benefit individuals, public health, and medical research and development. Even though there are multiple visualizations online for the health data none could depict the data appropriately according to the user requirement. So, we decided to choose this project as can help us to get better knowledge and analysis of the world health and find hidden patterns in it.

## Related Work

1. The data from <https://ourworldindata.org/> (our world in data) inspired us to choose this topic as we thought there is a necessity to represent where we are coming from and where we are today with respect to multiple health factors across the world.
2. We explored a wide range of visualizations from the official website of D3 (<https://d3js.org/>) from which we chose the visualization that best fit our data.

## Questions

The visualizations show fluctuations in the health factors over the last few centuries. Also, find the correlations between relevant patterns and find the story in the overall world health.

Following are the questions we are trying to answer using the visualizations:

1. What is the pattern of the factor on the population over a range of years for a country?

We used a line chart with multiple lines where each line represents the factor of a country. Here a factor is selected from the treemap and a country is selected from the world map. Though only one the year is selected on the rear slider, the bar chart shows the data of 3 of years around the selected year

2. What are the trends in the health factor of a specific country for a year?

On clicking a country on the world map, the corresponding line in the parallel chart gets highlighted which shows all the health factors of a country.

3. What is the difference of the health trends of multiple countries?

User is given a feature where he can compare the factors of multiple countries by selecting a country of interest and hover on other lines to compare. By this, line chart gives a detailed and clear comparison of various factors of the country.

4. What is the effect of political stability of a country on it's life expectancy?

By analysing the change in political stability of the country plotted in a line graph with Years on the Y axis and stability index on X-Axis, against the change in the years of Life Expectancy plotted in bars, we see how Years of Life Expectancy decline or increase in those years and analyse the correlation between the two.

## Data

### I. Data Sources

1. Our World In Data

*Link:* <https://ourworldindata.org/>

2. Global Health Data Exchange.

*Link:* <http://ghdx.healthdata.org>

3. Health Expenditure and Financing Dataset

*Link:* <http://stats.oecd.org>

4. WHO - Global Health Data Expenditure

*Link:* <http://apps.who.int/nha/database>

5. GapMinder Life Expectancy Data

*Link:* <https://www.gapminder.org/data/documentation/gd004/>

6. The Human Mortality Database

*Link:* <https://www.mortality.org/>

## II. Data Preprocessing

We have data related to health factors in the format:

Country Name	Country ID	Year	value
--------------	------------	------	-------

Extracting the data related to a particular country and year is difficult in this case as we have to iterate through the complete file (i.e., iterate over 40,000 records)

Solution:

In order to make the data access process easy, wrote a python script using Pandas to convert the structure of the excel to:

Country Name	Country ID	Year1	Year2	Year3	Year4	Year5	Year6
--------------	------------	-------	-------	-------	-------	-------	-------

Here each cell below the corresponding year has the value for that year.

This way we don't need to iterate through all the all the records. Made this change in order to increase the performance of data extraction.

For all the CSV files thus generated for each factor, different factors had a different range of years of data, so to maintain consistency we extracted only the data for the years 1990-2014 from the given data which was present for all the factors. Also, some factors had data for more countries than others. We filtered data for only those countries which are present in all the CSV's so that visualizations like parallel coordinates don't break.

For the causes of death data, the data was distributed in multiple CSV files so combined all the data into single CSV and the data is processed in the code as well. So that it don't need to make multiple I/O operations while rendering the visualization. Even if any changes are done in the constraints as the selected counties and a selected range of years. All the above decisions are made keeping in mind the volume of the data, i.e, as the data is not huge storing it in the variable is a better option than making multiple I/O operations as it will slow the performance of the visualization.

## Design Evolution

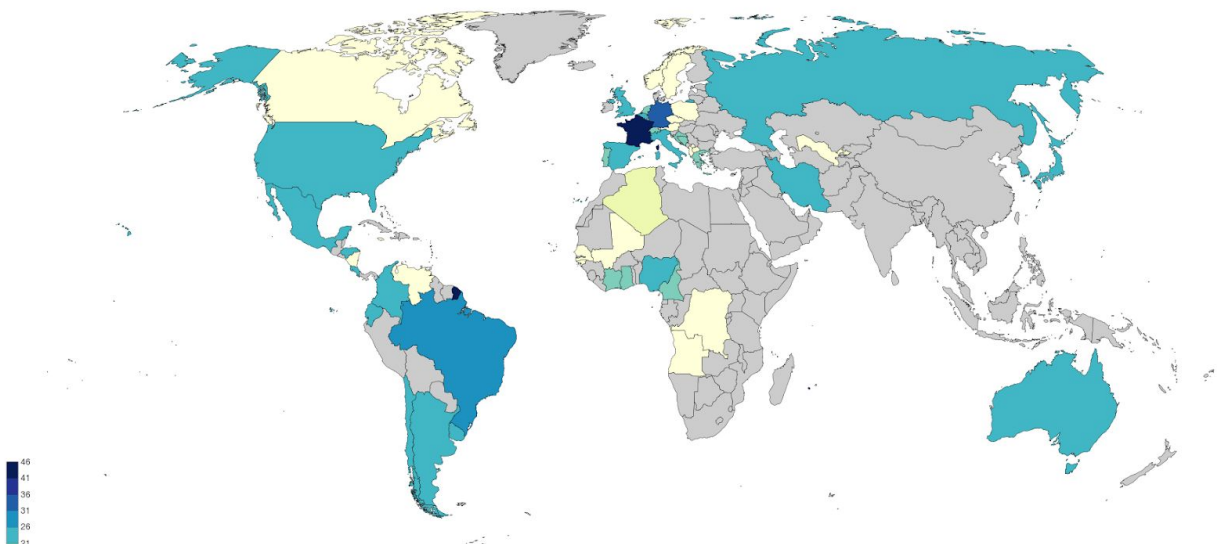
### World Map:

Things to represent are, for a year or range of years, what is the distribution of a health factor across the world and for a given year and a health factor which are the top 10 affected countries.

### Visualization options:

#### 1. World Map with colored countries:

Represent countries using a world map where the countries are filled with colors with respect to its magnitude. Here we draw world map based on the data stored in a JSON-compatible data format and we just need to define the size of the map and the geographic projection to use. By map to select the country, we can also identify the selected country easily by its geographic location.



## 2. World Map with varying circle size:

Using a world map with circles of different radius which represent the magnitude of the factor.



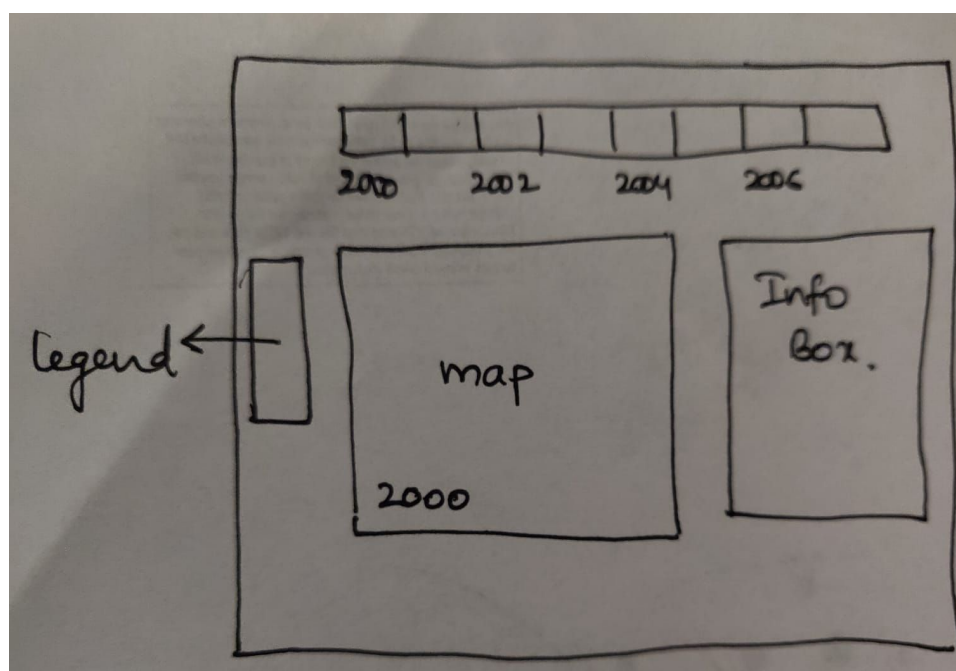
### Why World Map?

Use of world map where the countries are filled with colors with respect to the magnitude of the factor.

Area-value maps allow representing the various segments of the data value through color gradients on the map. As an example in the screenshot above, a darker shade of blue would be representing the higher value segments on the map, whereas a lighter blue would be used to represent the opposite. This visualization definitely helps people understand presentation in a glance! Clicking on the country would give the user to get a more detailed information about that country with respect to the health factors.

This approach is better than using world map with circles and line chart as they may make visualization clutter when many values are required to display.





## Causes of Death Visualisation:

This visualization is for analyzing the different causes of death over the years, in different countries in an effective manner. And in this visualization, there were 3 different alternative visualizations that were considered.

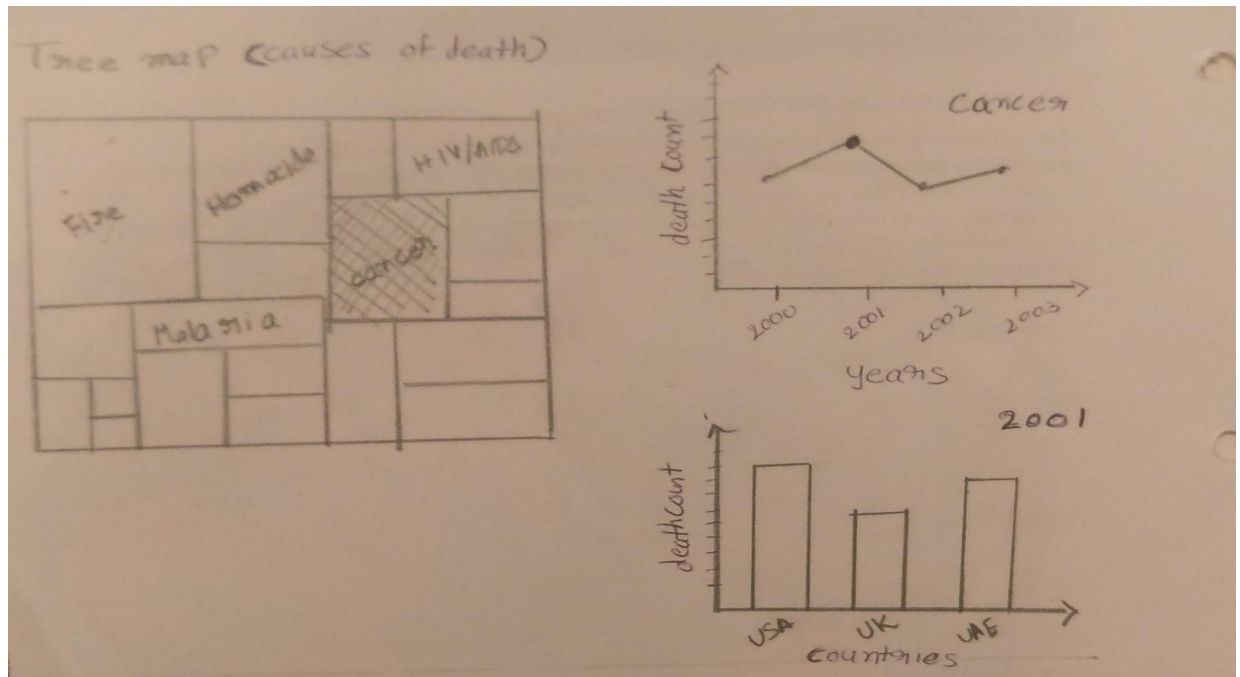
- TreeMap
- SunBurst Chart
- Line graph

## Alternatives:

### 1. TreeMap

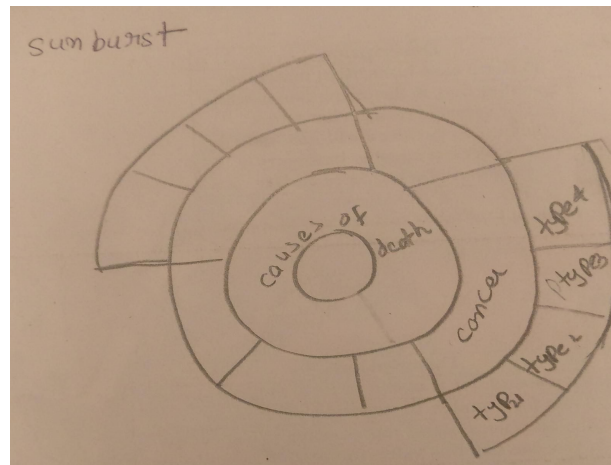
Treemap is one of the most effective visualizations for this data as the whole causes of death can be considered as on whole 100 percent and the all the various causes will account for the count will be represented in the subpart by the percentage of it. The treemap will be representing the causes of death for a single year/ range of years and for the selected countries/for the whole world. And on click of a particular cause, a line chart will be drawn with years on the x-axis and the death count on the Y-axis for that particular

cause. And on click, if one particular year A bar chart will be displayed with the countries in the x-axis and the death count in the y-axis. In this manner, we can analyze all the causes and the cause rate over the years and in different countries.



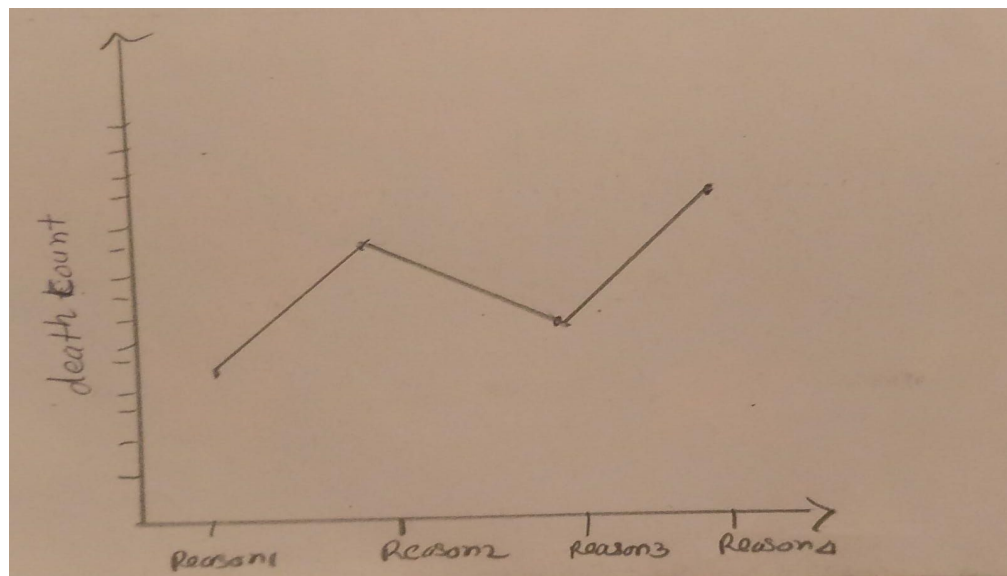
## 2. Sunburst chart

Sunburst chart can be used to represent the different categories of death in a selected country/all over the world and for one particular year or the range of years. We can categorize the causes of death and that can be expanded the causes of death into the subcategories and their percentage of death.



### 3. Line chart

This data can be represented as a simple line chart with all the reasons in the x-axis and the death count for that respective year in the y-axis, and the death count for one year or the range of years.



## Why TreeMap?

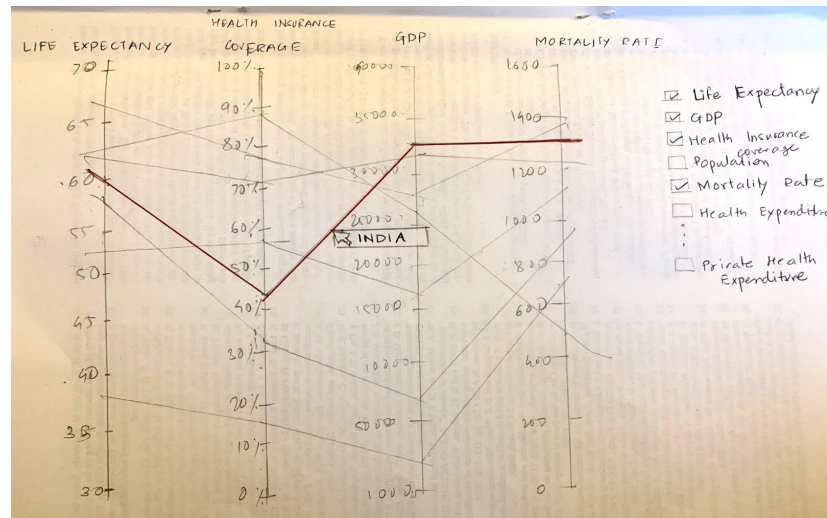
The reason for choosing the treemap, in the end, is that the treemap provides the user with good overall details of the map and various causes of death. And if the details are required the user on click will display him the details in further. The sunburst did seem a good alternative, the data will not be that effective and can be misinterpreted. And the line chart cannot give the detail view of the overall perspective, and the user must see the Y-axis constantly to derive to his results. And each reason is not correlated to other so the line chart will not be effective.

## Health Factors Visualisation:

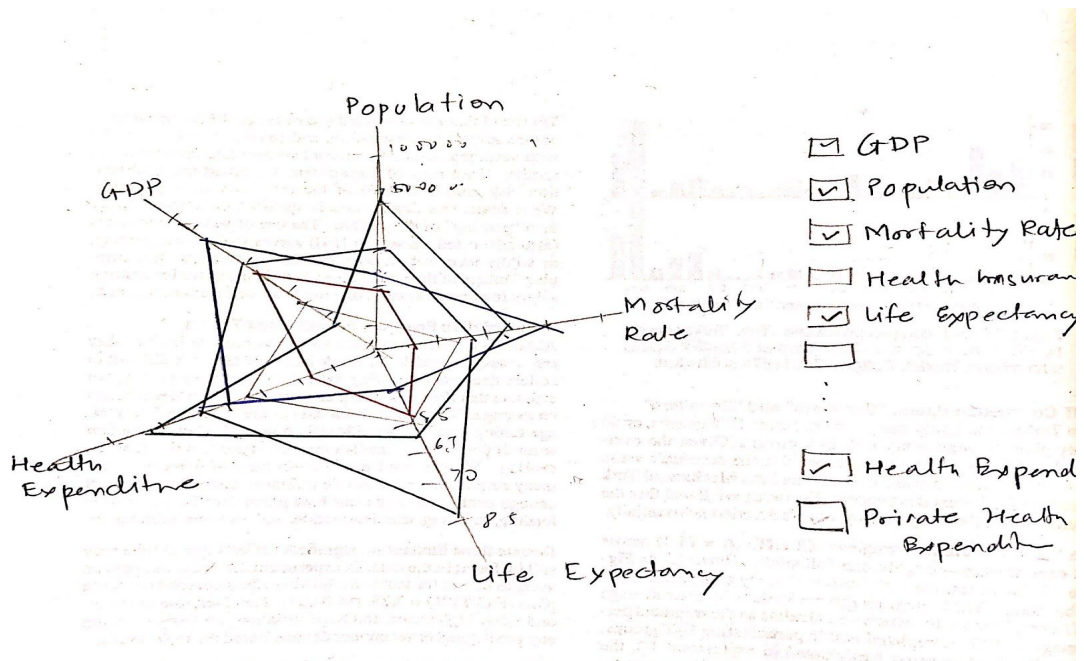
Each country has multiple features like Health Expenditure, Health Insurance Coverage, Life Expectancy, Child Mortality Rate etc. These features have a correlation among them and also are comparable across multiple countries. The visualizations possible are:

### 1. Parallel Coordinates

A line would represent an individual country. Each coordinate would be represented as a feature having scale according to the values of the feature. The value on the coordinate would indicate the value of that feature for that country. The features can be changed and accordingly, coordinates get updated. Filtering can be implemented on brushing on coordinates

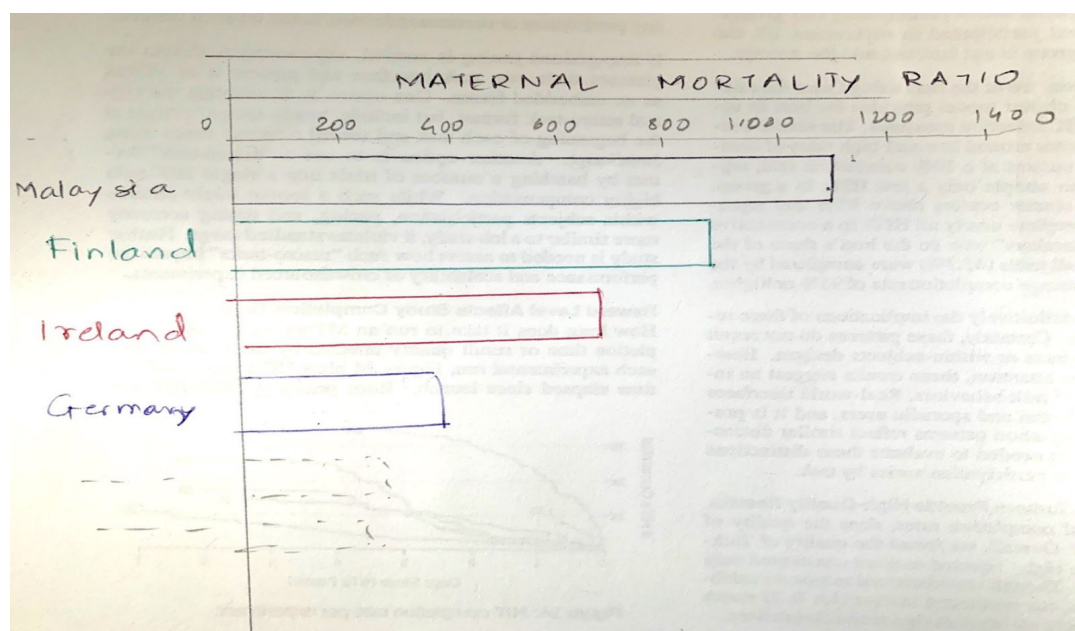


## 2. Radar Chart.



Each axis on the radar chart can be one of the selected features. On load, a default of 5 features would be the spokes. The spokes would be scaled according to the feature for that country. The value on the spoke would indicate the value of the feature for that country. The country would be represented as a line joining values of the country on all spokes.

### 3. Horizontal Bar Chart



For every individual feature a new horizontal having countries on Y-axis and value of the feature on X-axis. Each country represented by a bar of different color.

### Why Parallel Coordinates?

Horizontal Barcharts would make it easier to visualize the ranking of countries for a feature but for a comparative study of a feature among multiple countries, it would be very difficult to analyze. A radar chart would provide a connected line as a country and would allow comparative study well but only to an extent. When many countries would be represented, the chart will look cluttered and difficult to interpret. Parallel coordinates will be the best solution as it clearly represents features, values. Comparison of the value of a feature of multiple countries is clear and map selection will give easy flexibility to filter countries. A tooltip on the line indicating the country and some minor details would make parallel coordinates readable for multiple countries.



## Implementation

### 1. Home Page:

#### Word Cloud:

Implemented a static word cloud to creatively display the health factors like Cancer, different causes of death, Maternal Mortality etc. that are represented as part of multiple visualizations in the project instead of just writing the names sequentially. Moreover, also included few country names in it as small texts so that it represents the representation of factors over multiple countries. This implementation made the homepage look more attractive, informative.



We added the project description following the word cloud on the homepage.

Healthy citizens are the greatest citizens any country can have! - Winston Churchill

#### Health care factors

Health is a fundamental driver of our overall quality of life. We thus focus on factors affecting health, globally and country wise, the expenditure in these directions and trends over these factors in the past few years. How the world population is changing is of great importance for humanity's impact on the Earth's natural environment, but it also gives reasons to hope for a good future. This is because we have a bigger team of better educated people who can contribute to the solutions that improve global well-being.

#### Causes of death

As the population grows, causes of death across the world are changing. The data visualisations attempt to provide a comparison between causes and risk factors of death across countries and age groups. We observe that Non Communicable diseases not only dominate mortality figures at a global level, but also account for the majority of deaths in high-income countries. Deaths from causes such as infectious disease, malnutrition, nutritional deficiencies, neonatal and maternal deaths in high-income countries is typically very low in relative terms.

#### HealthCare Expenditure

Global expenditure on healthcare as a share of world income has been increasing, steadily but slowly over the course of the last couple of decades. In the background, however, there has been substantial cross-country heterogeneity, both in levels and trends. Regionally, high-income countries spend – and have been spending – a much larger share of their income on healthcare than low-income countries (about twice as much).

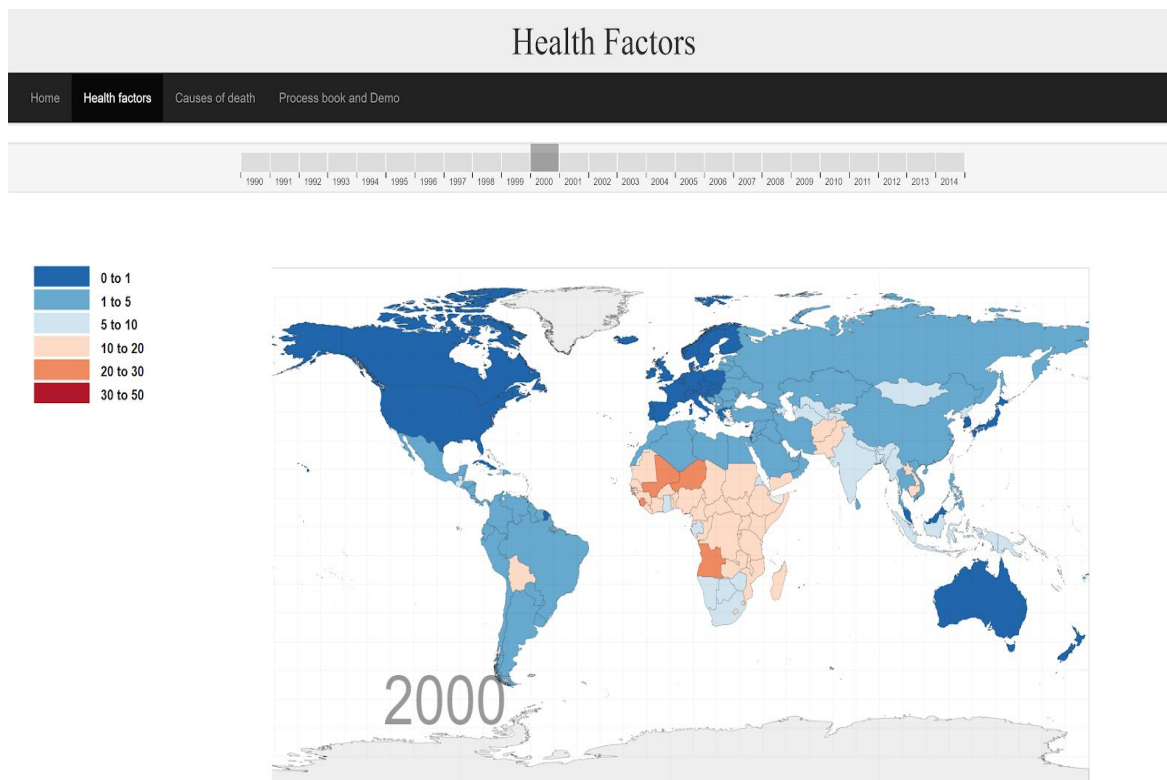
#### About Us

Neha Kherde  
Harshitha Manduva  
Antara Bahursettiwar

## 2. Health Factors Page:

### Year Slider, World Map & Parallel chart

As we wanted to represent multiple factors for a selected year, we implemented a unique style of year slider. To gain a detailed information for a specific year, the user should be clicking on one of the boxes in year slider. On selecting a year from the year slider, parallel chart and world map will get updated with corresponding year's data. By default selected year is 2000



On page load, the world map will represent "Child Mortality" data. The countries are colored with a range of colors depending on its value of child mortality. A color legend is added which represents the color corresponding to the range of values.

And the corresponding info box with details about the factor selected



### Child Mortality

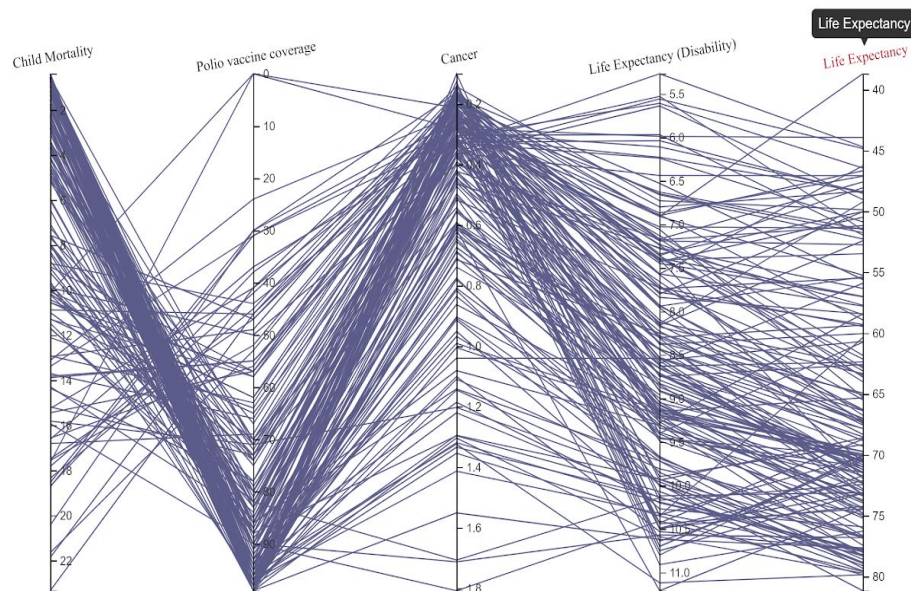
*Shown is the share of children (born alive) who die before they are 5 years old*

The world map shows the estimated level of child mortality for all world regions from 1990 to 2014 in the country borders of today. In all parts of the world child mortality is estimated to be higher than one third.

Similarly, on a load of the page, a parallel chat is displayed. This chart represents a factor as a vertical axis and each line across these axis represents a country. This chart provides a very clear understanding of each country's health factor on hovering.

Reset

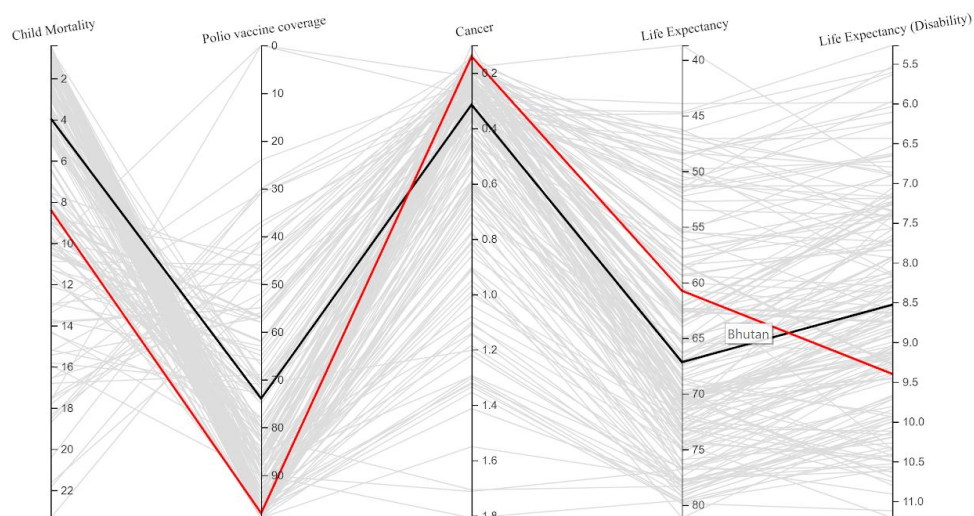
### Study Of Factors For Countries



The parallel chart provides multiple interesting features like, adding of the factors dynamically according to the user's requirement as well as remove the factors from the parallel chart by unchecking the checkbox from "Select health factors" box. By giving this feature, the user is given the flexibility to view only the factors of his interest. The selected factor range varies from 2 to all the factors.

Reset

## Study Of Factors For Philippines



The parallel chart also provides the feature to select a factor and reflect that factor's data on to the world chart. By doing this the user will be able to see the distribution of the factors across the world. Moreover, the selected factor on the parallel chart gets highlighted when clicked so that it is easy to correlate the data that is represented by visualizations.

## Select Health Factors

- ☐ Beer Consumption per person
- ☒ Child Mortality
- ☐ Child mortality by income level of country
- ☒ Expected years of living with disability/disease
- ☒ Life expectancy
- ☐ Maternal Mortality
- ☒ Polio vaccine coverage of one year olds
- ☒ Share of population with cancer
- ☐ Share with alcohol use disorders
- ☐ Share with mental and substance disorders

In order to gain information related to a single country, the user can click on a country in the map or a line in the parallel chart and see the corresponding line highlighted with red color.

One more interesting feature about the parallel chart is that it can be used to compare the data of one selected country with any other country. The user can select the country by one of the ways mentioned previously. This comparison can be done by hovering on the gray lines which also shows a tooltip indicating the country that the line represents.

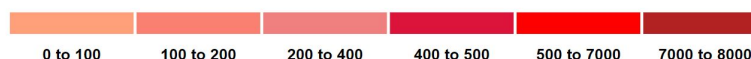
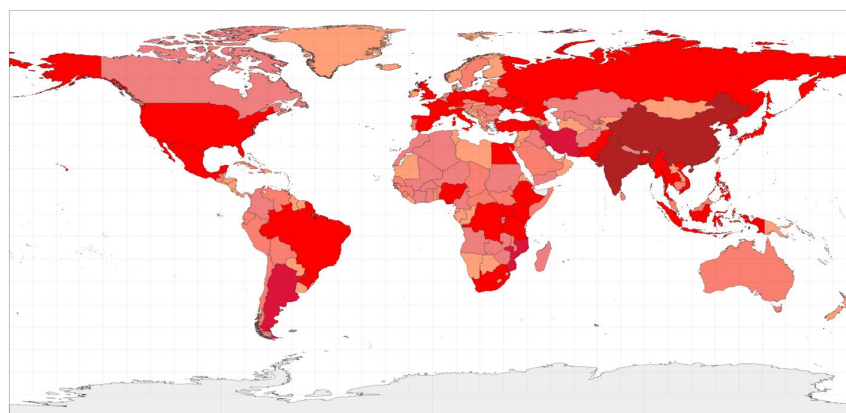
There is a reset button to view all country lines again in the parallel chart. This removes all the highlights of the countries selected for comparison.

From all the above interactions we can say that parallel chart, year slider and world map are tightly interconnected with respect to the interactions.

### 3. Causes of Death Visualization:

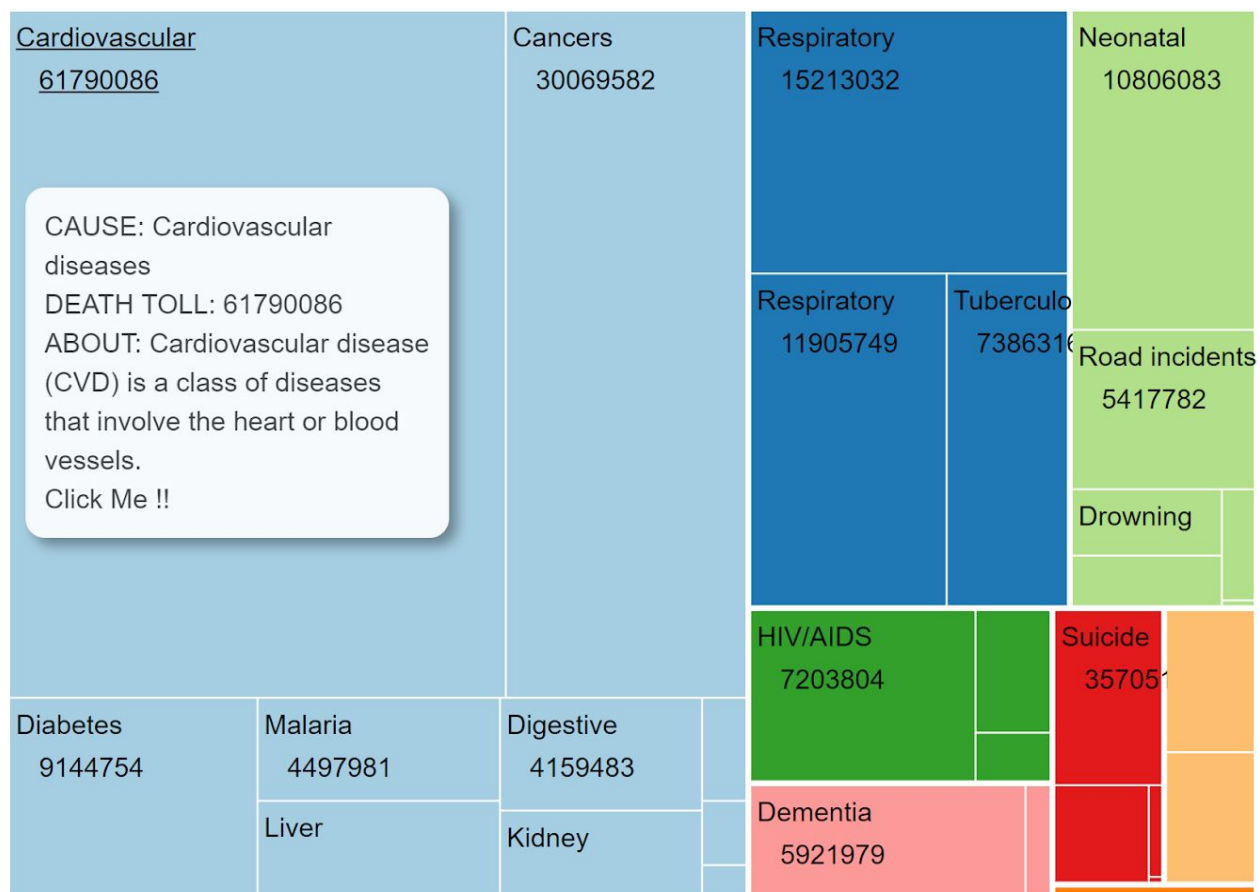
#### **TreeMap, Line Chart, Bar Chart, Year Slider & World Map**

To gain a detailed information for a specific year, the user should be clicking on one of the boxes in year slider. Same as in the earlier page, and on click of a year, the world map changes accordingly displaying the total death toll, due to all the causes of the data in that selected year. In the range of 1000's. On mouse hover on the map, the death toll is displayed in that year.



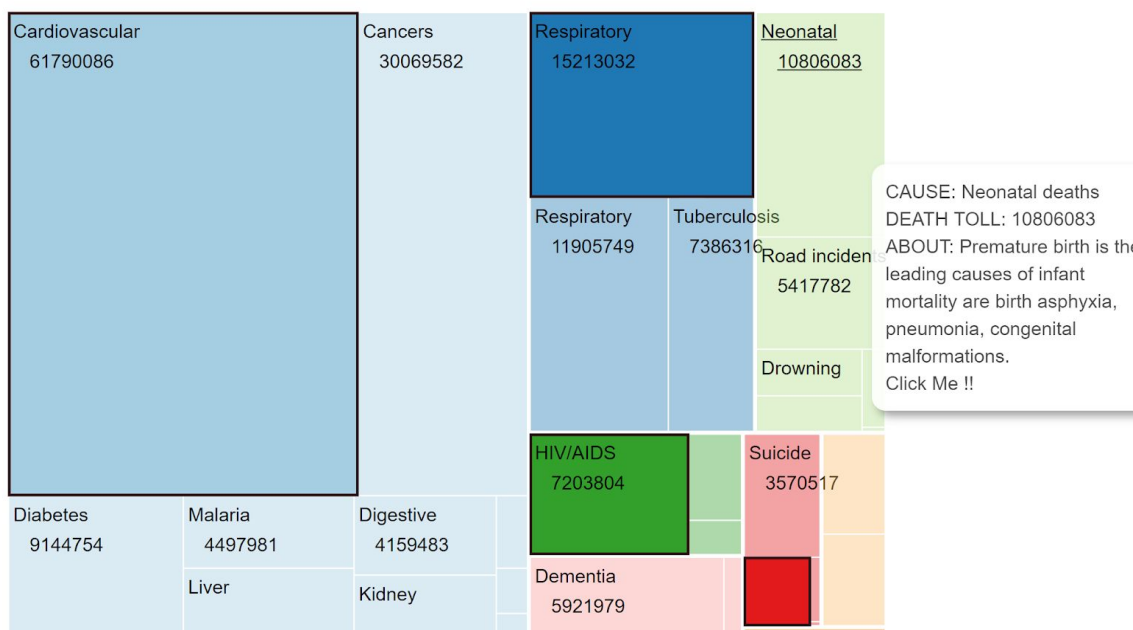
On load on the page, the treemap is displayed with the default year as 2000 and the death toll of all countries. On click of a country or a set of countries, the sum of the death toll of these countries will be displayed in the treemap. On mouse over of each rectangle in the map will be displaying a tooltip with details of the cause and the total no of people affected by the cause and the details of the cause.

### Causes of death distribution

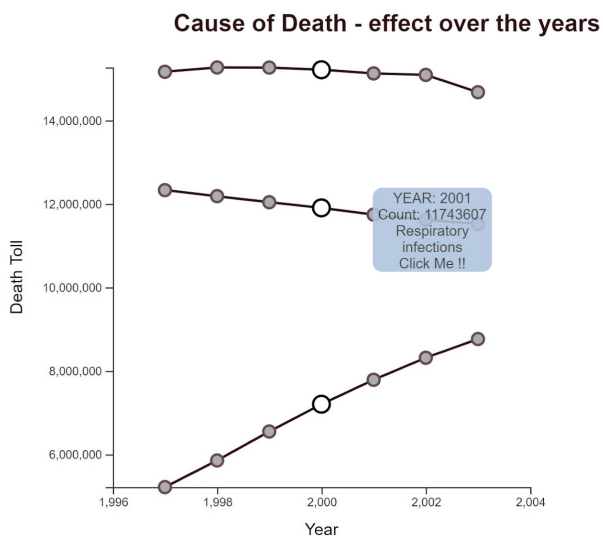


On select of a cause in the treemap, the selected causes will be highlighted in the map and the others will be opaque. On double click, the causes will disappear.

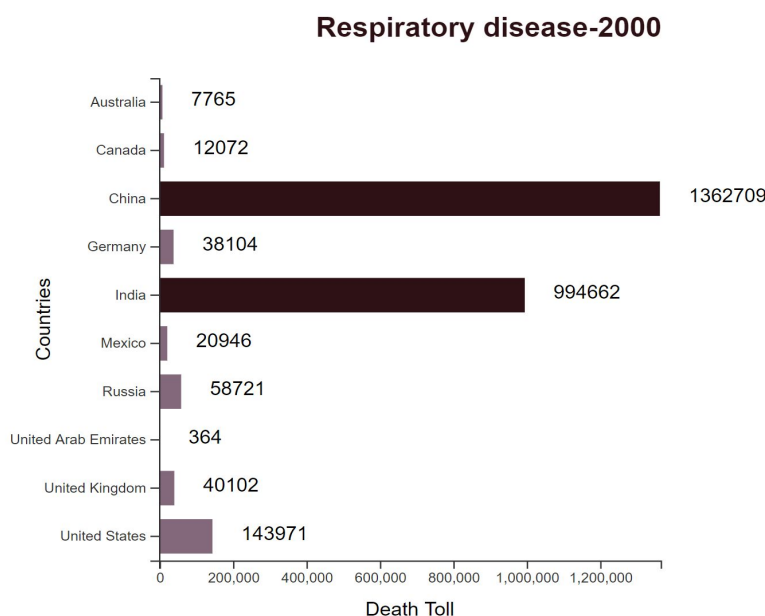
**Causes of death distribution**



On click of a cause in the treemap, the corresponding line chart will be displayed with all the current year and the way the death toll changed over the years. I.e., 3 years before the selected year and the 3 years, after the selected year. On mouse over of any one data point will display the tooltip with the year and the death toll and the cause corresponding to that line.



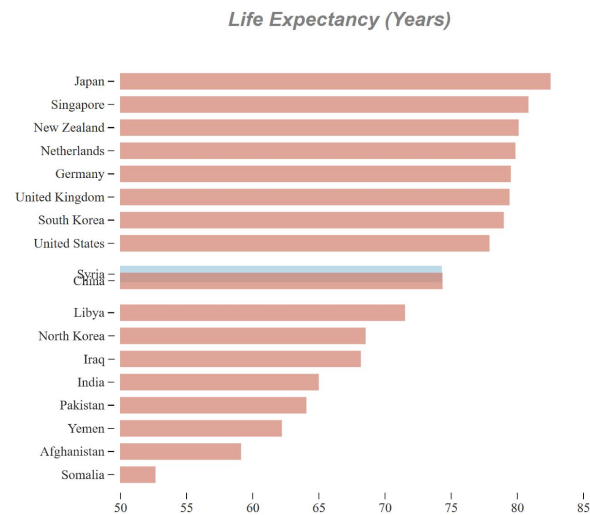
On click of a node in the line map will display a corresponding bar chart. This bar chart will be displaying the death toll of the countries for the selected year and selected cause. If no countries are selected the bar chart will be displayed top 10 famous countries data. But if the countries are selected the bar chart will be displaying the data of the selected countries. If there are more than 10 countries selected the 10 first selected countries data will be displayed in the bar chart.



#### 4. Political Stability Affect on Life Expectancy (2004 - 2014):

##### Bar chart and Line chart

This part of the site will be the story displaying a bar chart will be which will be representing the set of 20 countries and the life expectancy of that country and how it changed over the years. And correspondingly we will be displaying the line chart representing the political stability of a selected country over the years. The visualizations in this page change gradually with respect to the year's change, which is done in a time interval.



**2007**

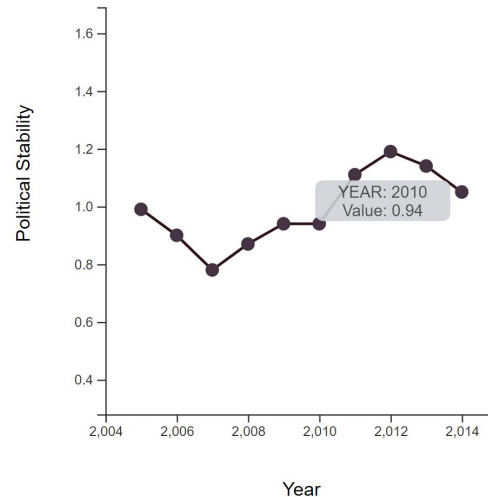
On selecting a particular country the bar chart will be highlighting that country. And the line chart will be representing the data corresponding to that country.

View Political Stability

Syria

The user cannot change the country while the data of some other country will be presented. And the user cannot see the button as well.

After the plotting is the process done the user can hover on the line chart nodes and the year and value will be represented in a tooltip.

*Political Stability*

## Evaluation -

Overall we have 3 tabs -

### - Health factors

This tab represents the health factors of the world which answers questions like what are the different factors of a selected country, comparing multiple factors of different countries. And how is the health factors of all the countries across the world?

Possible Improvements

Things that could have been improved in visualization are, reducing the opacity of all the countries except the one that is selected.


Adding a play button to the world map such that we can see how a factor is changing across the years.

Adding movement in the axis of the parallel chat so that the user can compare 2 factors of interest by placing them adjacently

### - Causes of Death

This visualization helps in understanding the various causes of death all over the the world. And how it changed over the years. This visualization was designed keeping mind the filtering of the data is as simple as possible. I.e, the user can select one year a set of countries and the corresponding data will only be represented. The





user can also select the data that they want to see without any clutter in the visualization.

- **Our Findings** ( Political stability effect on life expectancy (2004 - 2014)

On plotting the change in Life Expectancy along with its Political Stability, we analyzed that a significant change in Political Stability gives a direct impact on the Life Expectancy of a region. For example, the political stability index of Syria declines and its Life Expectancy has a corresponding decrease in years. The visualization works fairly well giving a clear idea about its standing among other countries and the change with its political index over the given range of years.

Possible Improvements:

- Analyzing more correlations and effects of one factor on another. For example, how does the Health Expenditure of a country has an effect on its Mortality Rate, Life Expectancy?