Intermediate Project Report Date: 11/09/2018

Global Health Analysis

1. Data Processing:

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

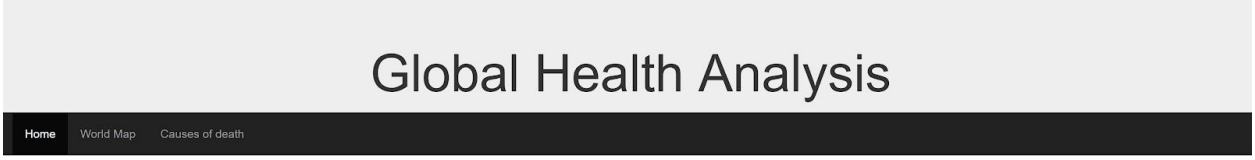
Here each cell below the corresponding year has the value for that year.
This way we don't need to iterate through all the all the records. Made this change in order to increase the performance of data extraction.

For all the CSV's 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's so combined all the data into single CSV and the data is processed in the code as well. So that it dont 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.

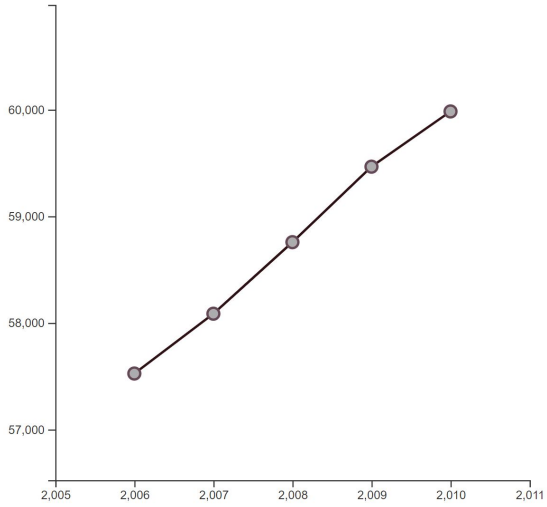
2. Implementation so far:

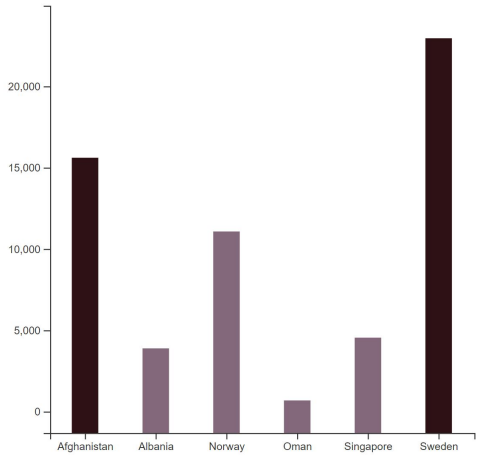
Started off with designing the basic user interface with 3 tabs.



a. **Home Page:**
This tab would have an introduction section, a button to navigate to project video, etc.,

b. Causes of Death:
Treemap is being displayed for a set of counties and the tooltip on mouse over it. On click of any of the disease the corresponding line chart is being displayed ie., the death toll due to that disease for the selected range of years (which is hardcoded as of now but will integrate the year slider in next phase) And on click of one year on the line chart the corresponding bar chart is being displayed. The bar chart will represent the death toll to that disease in that year for the selected countries (which is hardcoded as of now). Below are the screenshots of the treemap, line chart and bar chart. Even though the selected counties and years are hard-coded changing it and making it dynamic will not be requiring major code changes.



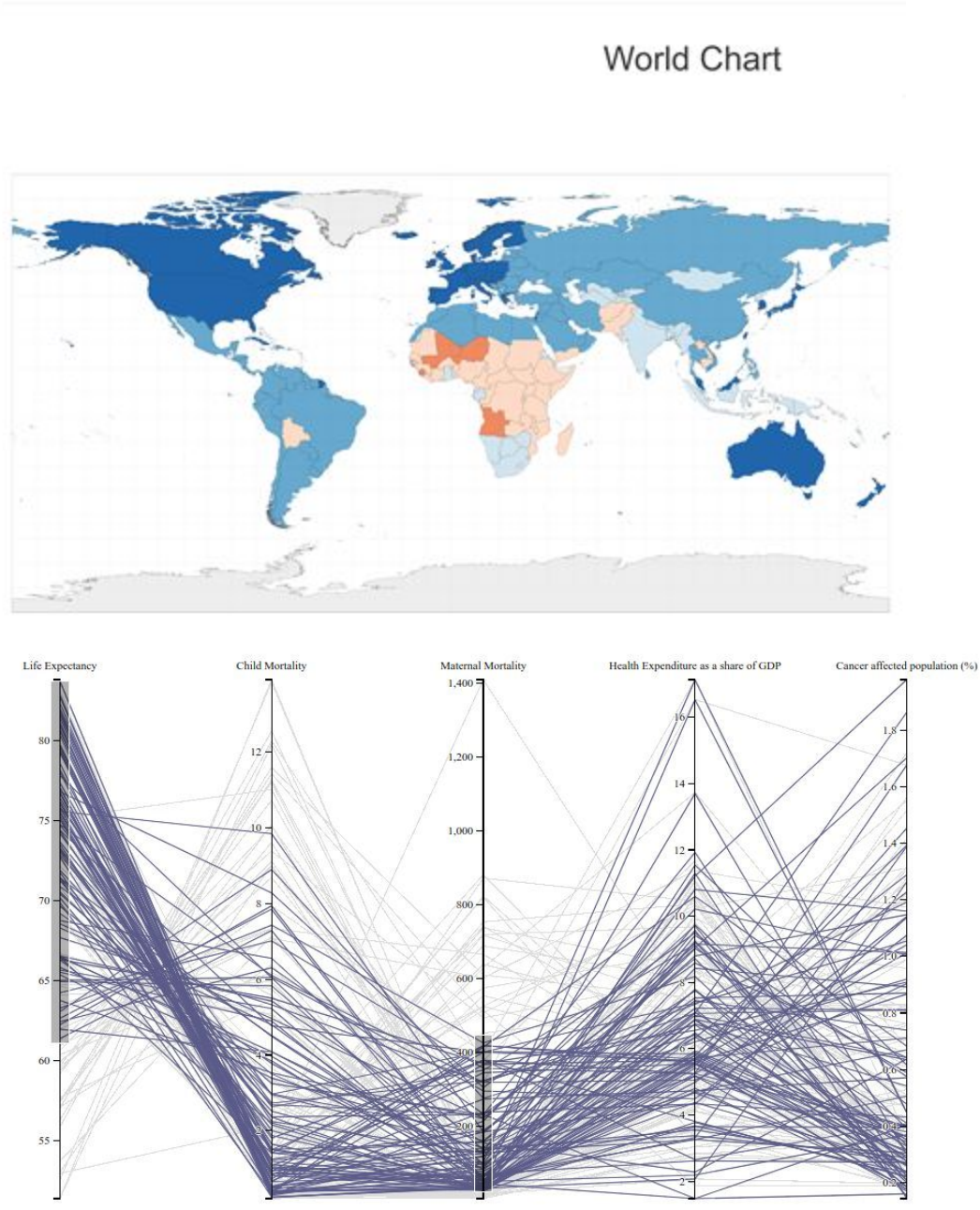


c. Other health factors:

As part of this tab, Designed a year slider bar by using d3 ticks for the markings. Here each rectangle represents a year. Clicking on a rectangle leads to creating of a brush on that rectangle. And also allows brushing across the years to selecting a range of years. Added a feature such that the brush is moveable and resizable.

Implemented the world map using a geo-rectangular map such that it can fetch the data related to one factor and a year. Also, filled with colors depending on the value of the factor.

A basic implementation of parallel coordinates using 5 factors as coordinates and representing each country as a line. Implemented brushing to filter over the range on the coordinate. The data now represented is for a default year, which we plan to change according to the year selected in the year slider.



3. Project Schedule:

Following are the tasks to be completed by 19th Nov:

In Other Health Factors Tab:

1. Add interaction between the year slider and map such that on selecting a year corresponding map should be displayed
2. Add drop down box for selecting a health factor

3. Add color legend indicator below the world map
4. Add on hover tooltip on the world map
5. Add a play button to give a video effect on the map when a range of years are selected.
6. Coordinate parallel coordinates visualization with year selected.
7. Coordinate parallel coordinates visualization highlighting the country selected in Map.
8. Change coordinates in parallel coordinates visualization according to the factors selected by the user.

In Causes of death Tab:

1. Year slider to select the countries must be added.
2. The map must be introduced to select the countries.
3. Add appropriate color scale to the treemap.
4. Make the small CSS changes in the rest of the page.