



Visual Analytics for Health Data:

Statistical Analysis on factors influencing life expectancy across the globe.

04.17.2020

Sai Pramod Kudapa

112686280

Ghanendra Piplani

113070787

Background

Life expectancy is defined as the average time an organism is supposed to survive, based on various factors such as the year of its birth, current age and health conditions, the environment it lives in among others. Through this project we try to find the factors which affect life expectancy the most, either positively or negatively, using which a country can decide the correct way to spend its resources (both man and money).

Problems

1. What are the predicting factors which affect life expectancy?
2. Should a country having a lower life expectancy value (<65) increase its healthcare expenditure in order to improve its average lifespan?
3. How do Infant and Adult mortality rates affect life expectancy?
4. What is the impact of schooling on the lifespan of humans?
5. Does Life Expectancy have a positive or negative relationship with drinking alcohol?
6. Do densely populated countries tend to have lower life expectancy?
7. What is the impact of Immunization coverage on life Expectancy?

Approach

Data:

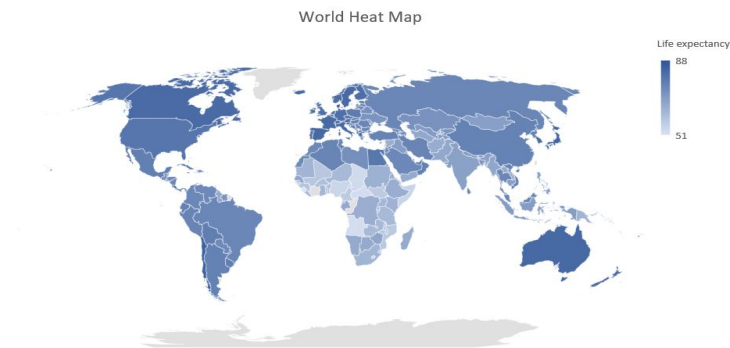
We have acquired following datasets from the internet:

1. [Global Health Observatory \(GHO\)](#) which is a subsidiary of World Health Organization data from kaggle which has 22 columns such as Country, Year, Life expectancy, Adult mortality, immunizations such as for polio, measles, Hepatitis B, GDP, etc.
2. [Current health expenditure \(% of GDP\)](#) data from Worldbank official website which has columns Country and years ranging from 1960 to 2019 which shows the percentage of GDP a country spends on healthcare.
3. [Data](#) for number of alcohol servings per capita per year for 15 years of age or older (for beer, wine, and spirit) across various countries.

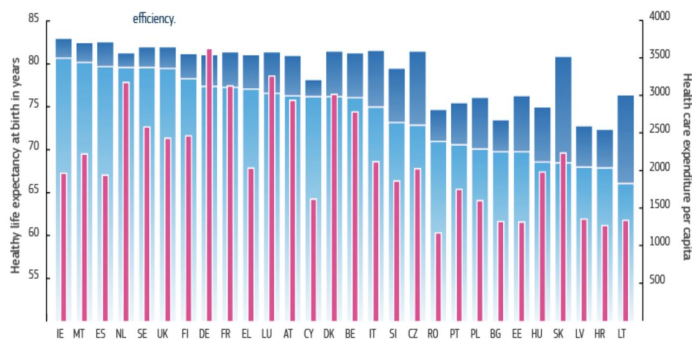
Visualization techniques to be used

We would be creating a single page web application, which would have a dashboard on the landing page. The dashboard would help the user in navigating between different visualizations from the following:

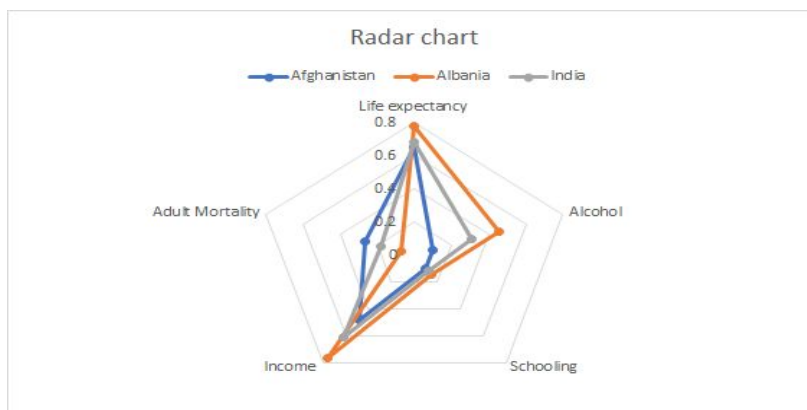
1. **World heat map** consisting of the life expectancy value distribution.



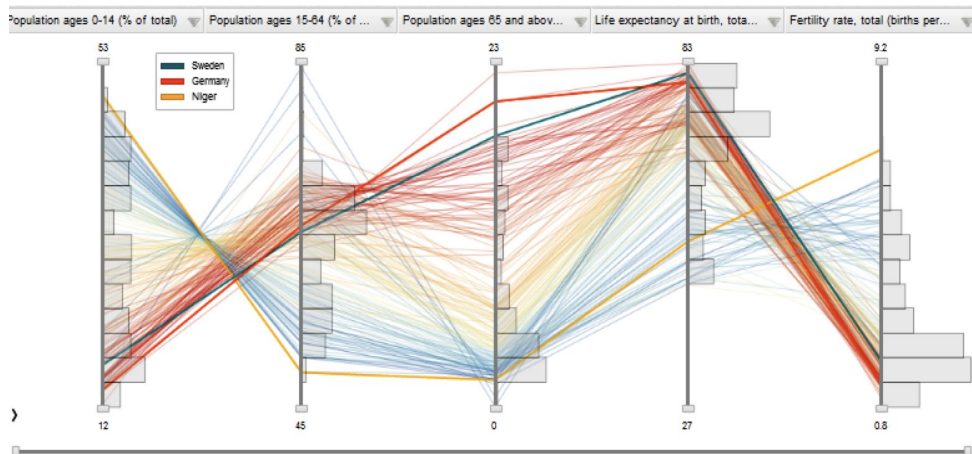
2. **Stacked Bar chart** to visualize life expectancy across countries and years.



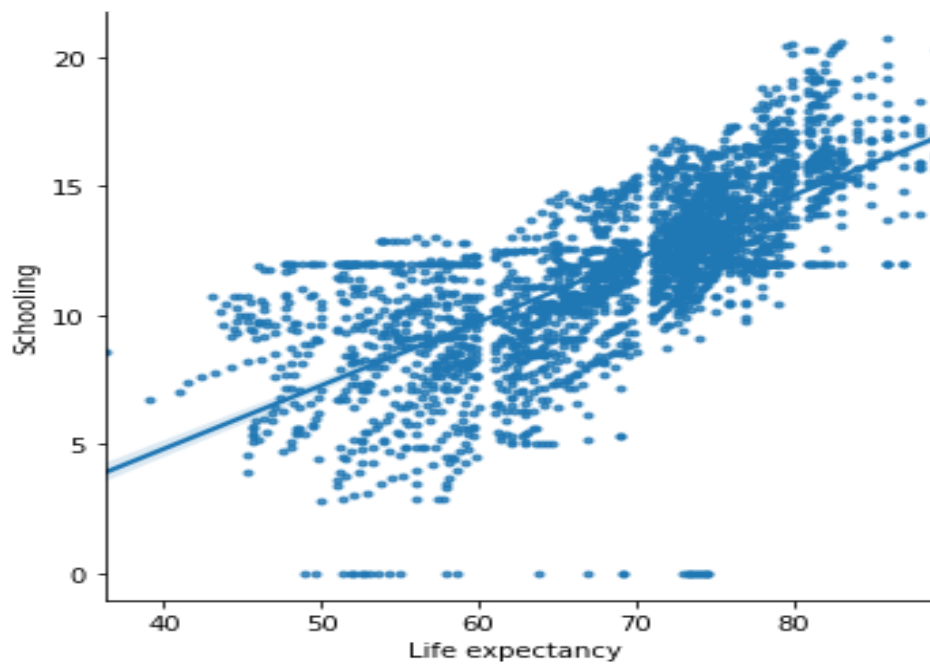
3. **Radar chart** to visualize top 4 parameters (found using PCA) along with life expectancy.



4. **Parallel coordinates plot** to visualize country, life expectancy, age, etc.



5. **Scatter plot** to visualize life expectancy versus various factors.



Tech Stack

To implement the above visualizations we would be using the following technologies:

I. Data cleaning

We would use Pandas and numpy python libraries to clean our data of garbage values.

II. Data analysis

We would use Sklearn python library to implement required statistical tools such as sampling, clustering, dimensionality reduction.

III. Visualization

To achieve the above mentioned visualizations we intend to use the D3 JS library along with Bootstrap, JQuery.

IV. Application Framework

In order to create a back end for processing data and serve the same to a front end in order for D3 to render, we would use Flask web development framework in Python.