

Visualization Project Proposal

Basic Info:

The project title: Global Health Analysis

Names: Harshitha Manduva, Neha Kherde, Antara Bahursettiwar

e-mail addresses: harshitha.manduva@utah.edu, neha.kherde@utah.edu, antara.bahursettiwar@utah.edu

UIDs: u1080236, u1205893, u1209537

Link to the project repository: <https://github.com/harshi-M/VizAssignment2018>

Background 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 realted health-related factors.

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.

Project Objectives:

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?
2. Compare multiple countries over years for a particular factor
3. What is the cumulative share of a factor in the world across all the countries?
4. What are the different factors and their values for a country and a specified year?

Data Sources:

The data for different factors for countries is gathered from the following 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/>

Data Processing.

Data is taken from multiple sources so we joined the data from different sources according to the features that are required. We required Geo and Topo data that is missing to represent the map, which we plan to collect. Data is sparse for a few factors, so we plan to handle these cases explicitly. Moreover, there is a need to calculate percentages for certain visualizations like Treemap and world map for representing certain features.

Visualization Design:

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.
2. Line Chart:
Represent countries using a line chart where the x-axis has countries and y-axis is the magnitude
3. World Map with varying circle size:
Adding circles on the map with radius scaling to the magnitude of the feature.

Finalized Design:

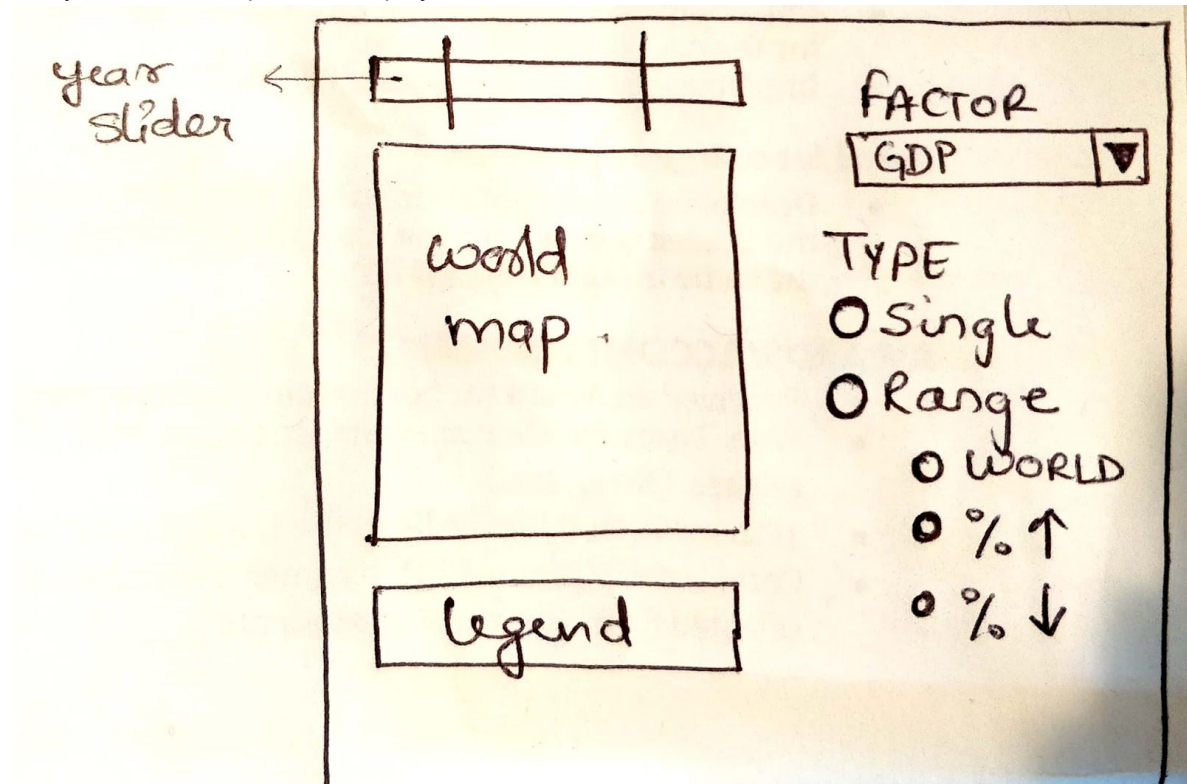
World Map:

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

1. In case of visualizing for a year and a health factor across all the countries, filling the map with respect to the magnitude should be easy to understand as the darker colors would represent huge values and light colors would represent small values

2. Also, if the user wants to view only top 5 countries with highest percentage increase/decrease values for a year and a factor, highlighting those countries with colors would be clear and it also gives the information relative to the size and location of the country.
3. Also for a range of years, it is easy to visualize how the countries are changing in its values on adding a play button which would change the color as the values of the factors change

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 Visualization over the years:

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

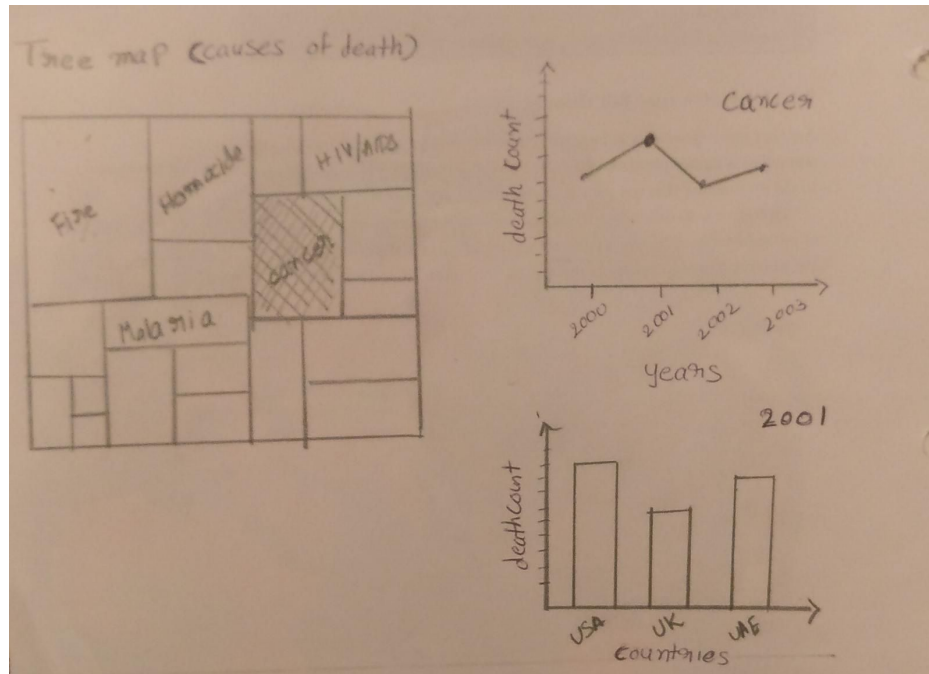
- TreeMap
- SunBurst Chart
- Line graph

Following are the visualizations diagrams and its details and the reason for choosing the treemap in the end.

1. TreeMap

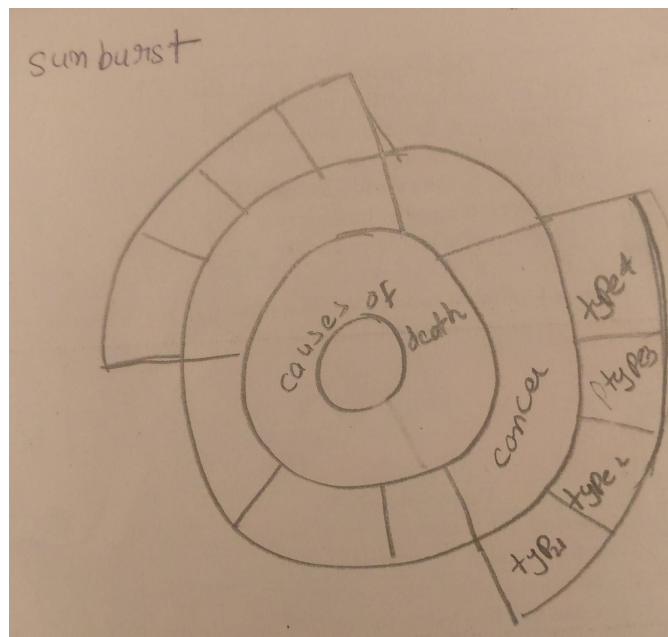
Treemap is one of the most effective visualizations for this data as the whole causes of death can be considered as on whole 100 percent and the all the various causes will account for the count will be represented in the subpart by the percentage of it. The treemap will be representing the causes of 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.



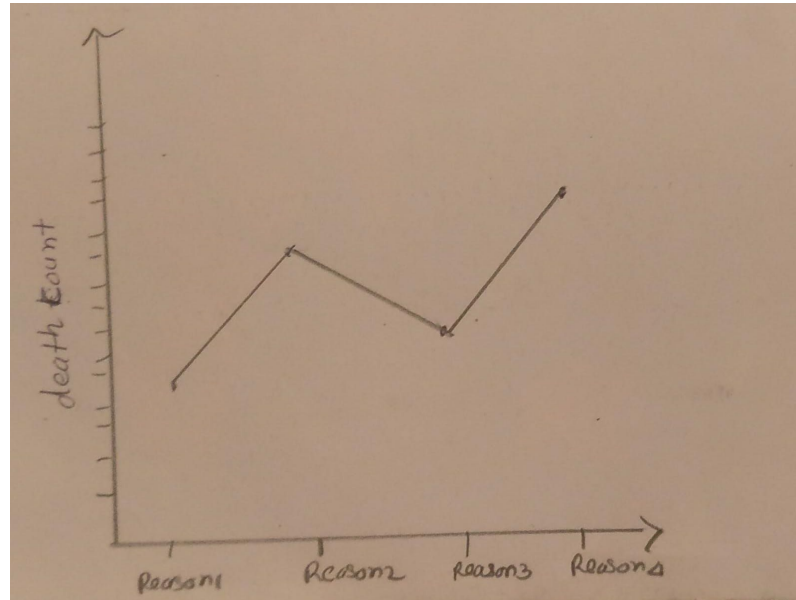
2. Sunburst chart

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.



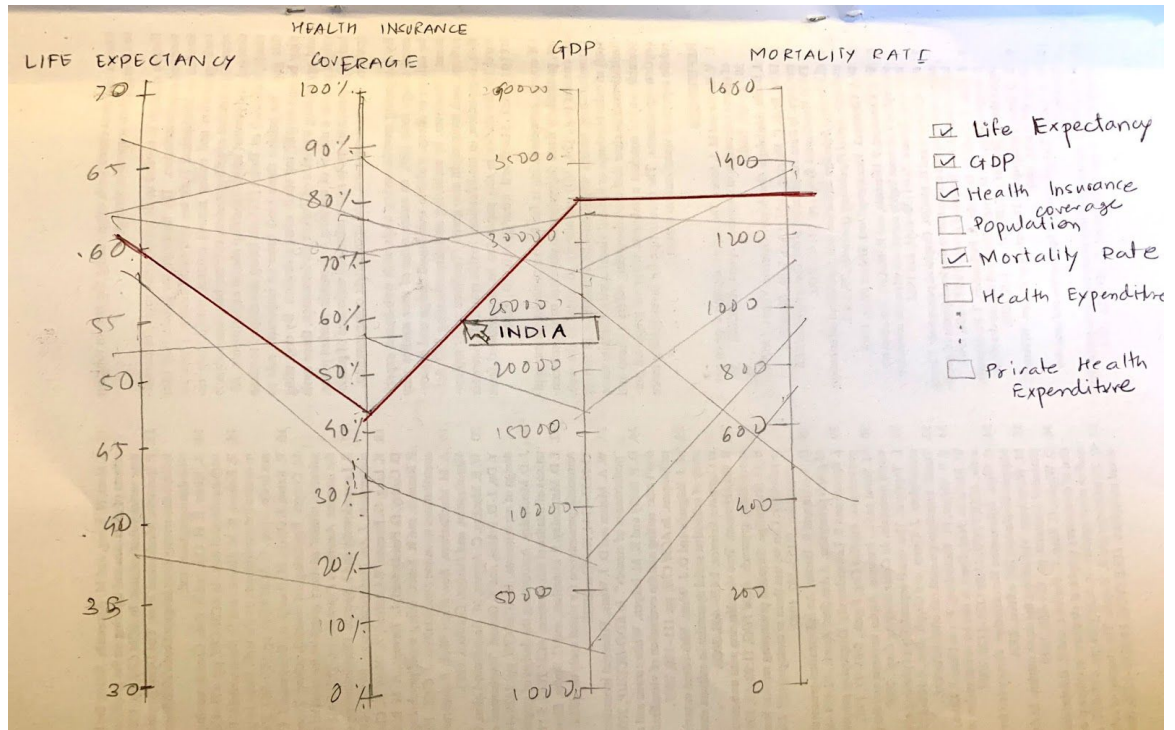
Reason for choosing the tree map:

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.

Multiple features for a country/countries:

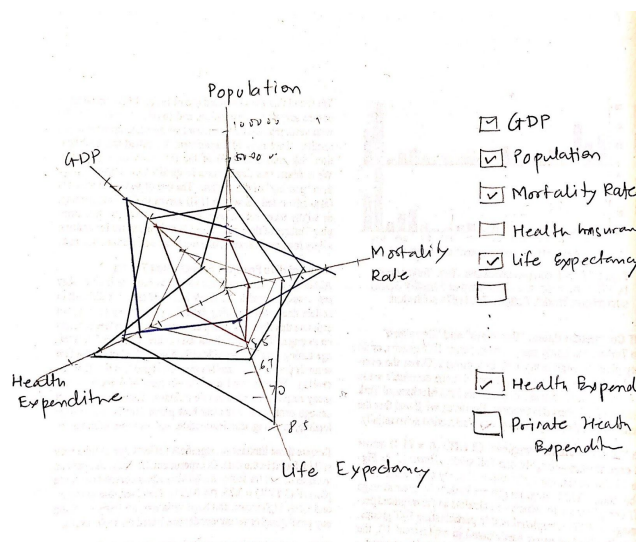
Each country has multiple features like Health Expenditure, Health Insurance Coverage, Life Expectancy, Child Mortality Rate etc. These features have a correlation among them and also are comparable across multiple countries. The features being numbers also provide the option of filters or ranges. The visualizations possible are:

4. Parallel Coordinates



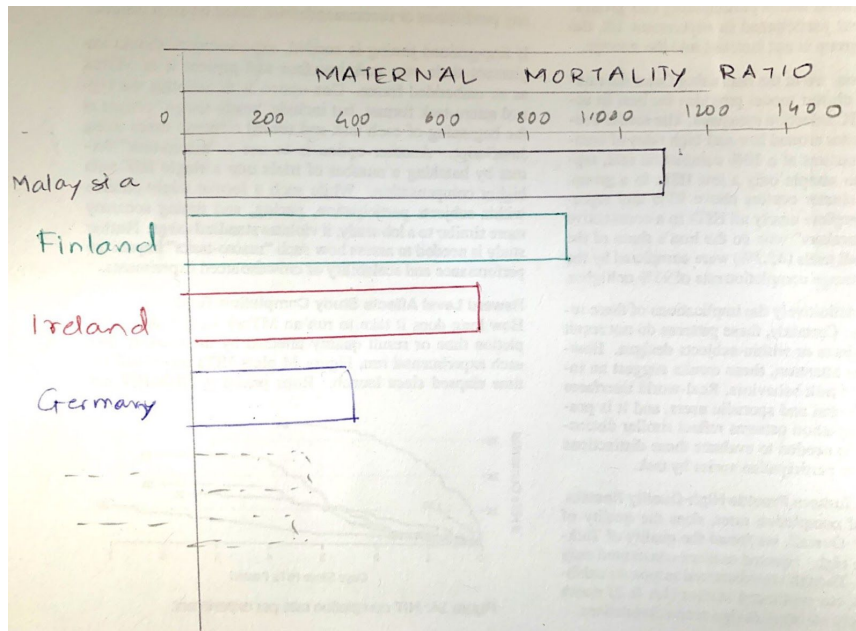
A line would represent an individual country. Each coordinate would be represented as a feature having scale according to the values of the feature. The value on the coordinate would indicate the value of that feature for that country. The features can be changed and accordingly, coordinates get updated.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

5. 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.

6. Horizontal Barchart



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.

Comparison:

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 brushing on axes 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.

Must-Have Features.

World Map:

1. On select of a year and feature, it should show all the countries in the map filled with colors with respect to their values
2. On select of a range of years and a feature, show change in the values across all selected years
3. On select of 2 years, show top 10 countries with max percentage increase or decrease within those two years.

Causes of Death Visualization over the years TreeMap:

1. The data required to display the Treemap must be processed
2. The treeMap being displayed with appropriate data and scales
3. On the selection of a particular year or a set of year and some countries, the treemap must be changed and the appropriate data must be represented
4. The selection of a set of years will be displayed in the slider on brush an that must be implemented.
5. ToolTip to display the data.
6. On click of a particular cause in the treemap displaying a line chart will be drawn with years on the x-axis and the death count on the y-axis for that particular cause.

For the visualization Parallel Coordinates:

1. A selection checkbox on the rightmost side to add or remove features as coordinates.
2. Each coordinate having an exact scale according to the feature domain.
3. Graying of all lines initially and coloring of line on hovering it.
4. A tooltip to be displayed on hover on the line displaying the Country name it indicates.
5. Brushing on each coordinate to filter countries on the values less than the value indicated by brush on a feature.

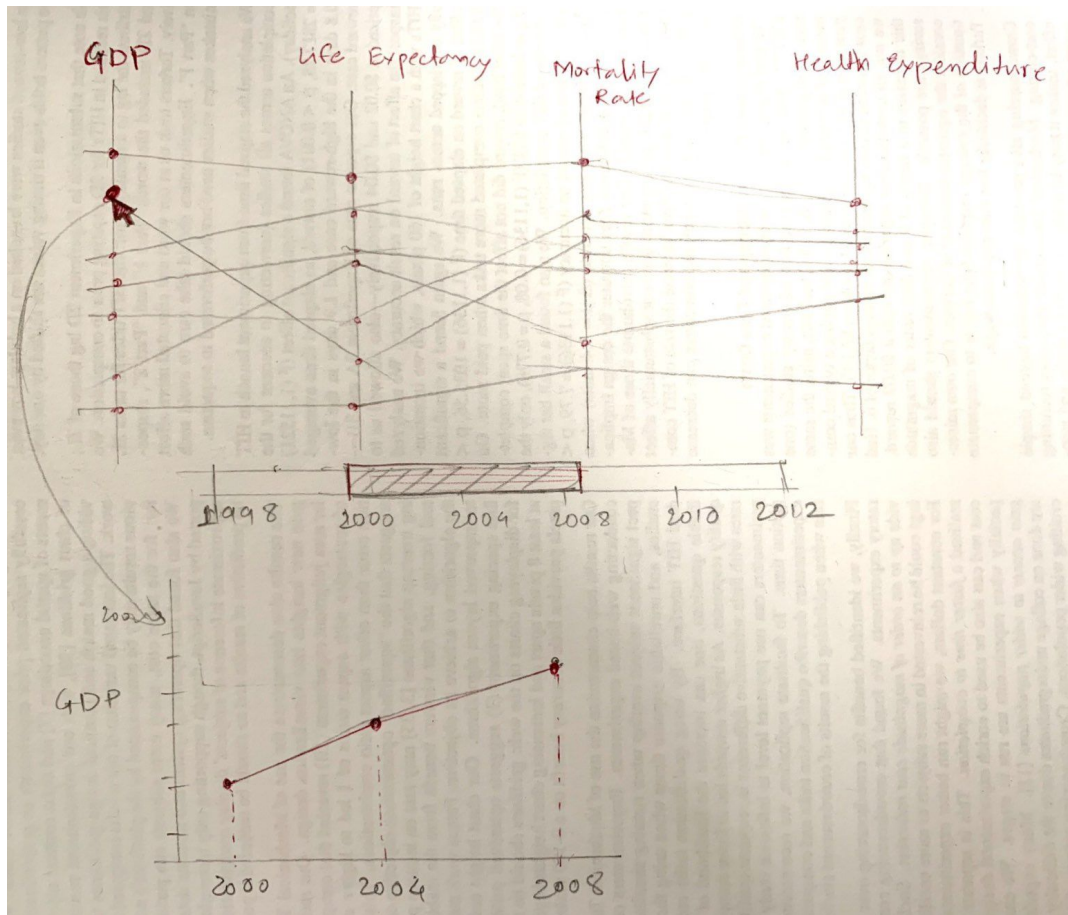
Optional Features.

Causes of Death Visualization over the years TreeMap

On click of a particular cause in the treemap displaying a line chart will be drawn with years on the x-axis and the death count on the y-axis for that particular cause will be implemented mandatory. But on click, if one particular year in the line chart displaying the bar chart with the countries in the x-axis and the death count in the y-axis will be an optional feature.

Parallel Coordinates: Grouping data for a range of years and selective Line charts for details

In this, if a range of years is selected, data will be averaged on the years and this averaged value would be represented on the coordinate. On selection of that value, a new line chart would be shown with years on X-Axis and values on Y-Axis representing the exact change of the values over the selected range of years.



Project Schedule.

10/28/2018: Project Proposal

10/30/2018: Peer review

11/2/2018: Getting data ready, and data selection working. And change in design plan based on peer review.

11/5/2018: On one year displaying the world map, treemap with proper axis

11/7/2018: On one-year display the parallel coordinates. And countries must be selectable.

11/9/2018: All basic views up and running

11/12/2018: Map for years and countries selection.

11/15/2018: Countries election for treemap and parallel coordinates.

11/18/2018: For multiple selections of countries and years, displaying the treemap and parallel coordinates.

11/22/2018: Review

11/25/2018: Adding optional features.

11/28/2018: Review

11/30/2018: Final submission