Subject : CSE 564 - Visualization

Topic : Analysis World Statistics

PROJECT REPORT

By Melita Saldanha, and Neha Shetty

Resources:

- **YouTube video link of Demonstration:** https://youtu.be/A3tSP--bM7Y
- **Link to Individual Dataset Files:** https://drive.google.com/drive/folders/1xjwy416dVxxsNV5zF_SgF8Xh2Iwcpm9G?usp=sharing

Introduction:

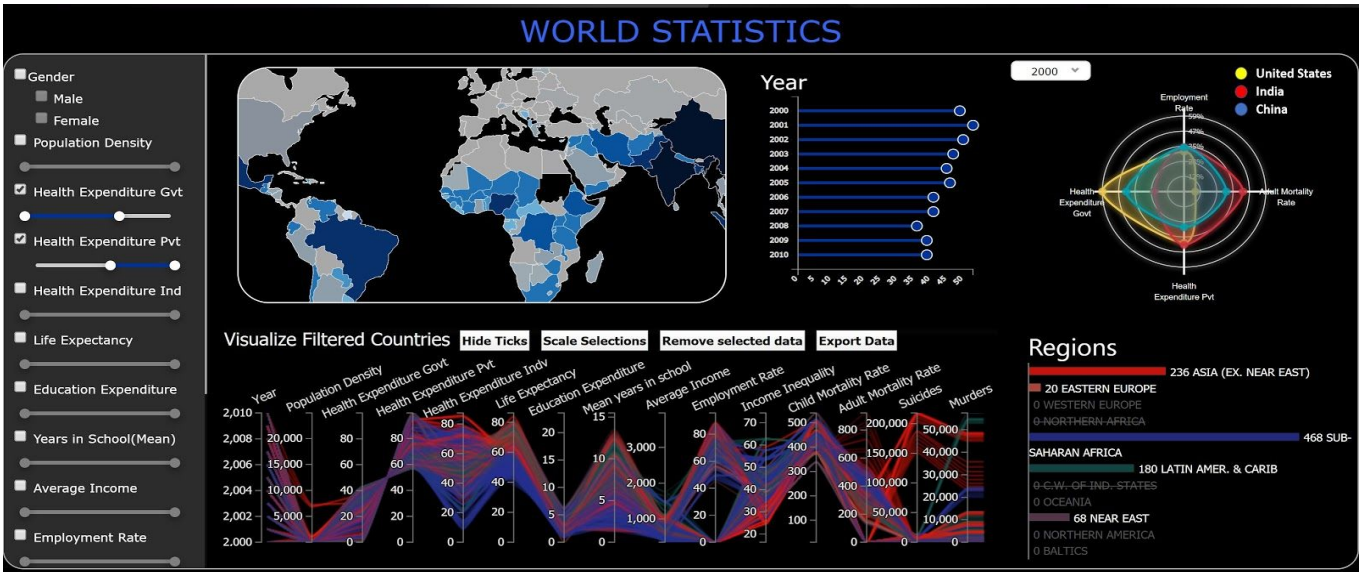
Sometimes we need some hard data to fully understand things; it means a lot more to say "1.3 billion people live in poverty" than to say "a lot." World Statistics is all about the data of each country with respect to various categories like Population density, Health Expenditure, Average Income etc. The goal of this project is to visually analyze the statistics of a country and to perform a comparison between different countries or their regions based on a collection of features.

Dataset:

The dataset for this project is taken from various sources and is a collection of many feature datasets which was merged based on countries in the world. Each feature has an attribute: Country and its value for each year in a certain range. Some of the features includes Health Expenditure by Government, Life Expectancy, Suicide rates and many others.

Dashboard Design:

We created a dashboard in order to perform this analysis and to draw valuable insights from it.



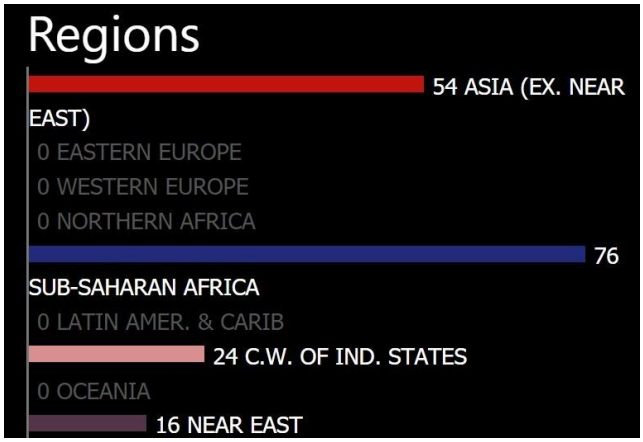
Dashboard to Analyze World Statistics

Elements of the DashBoard:

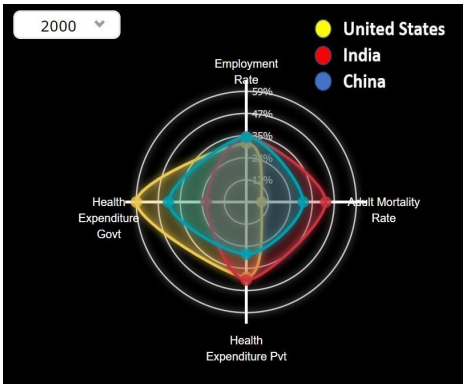
1. **Filters:** This consists of filters that can be applied on the various features of the dataset in order to analyze a specific category of items.
2. **Choropleth Map:** This map shows all the countries in the world. Shades of blue is used as a color scheme to indicate the population of the country. The countries that fall under the filters selected are highlighted in blue and the rest are disabled and shown by using the grey color. Clicking on a country in the map, highlights the data of that country in the parallel coordinate plot.
3. **Lollipop Bar Chart:** This chart shows the number of countries in each year which satisfy the constraints applied by the other filters. In order to filter the data based on the year, we can click the bar corresponding to the year to enable / disable it.
4. **Parallel Coordinates Plot:** This shows the data for each country for the various dimensions simultaneously. Filters applied in the left bar, as well as filtering by the year and region is linked to this plot which shows only filtered data. We can brush a selected portion of a dimension and scale the axis value to show it. Also, you can select points to be excluded from the plot. Other features are rearranging the axis positions, inverting the axis and exporting the filtered data.
5. **Regions Bar Chart:** This chart shows the number of countries in each region and clicking on it filters / removes filtering for that region on toggle. The chart shows the color corresponding to each region which is used in the parallel coordinate plot.
6. **Radar Plot:** This shows the intrinsic dimensions obtained by PCA. We can analyze the values for the United States, India and China for these dimensions by choosing the year.

Observations:

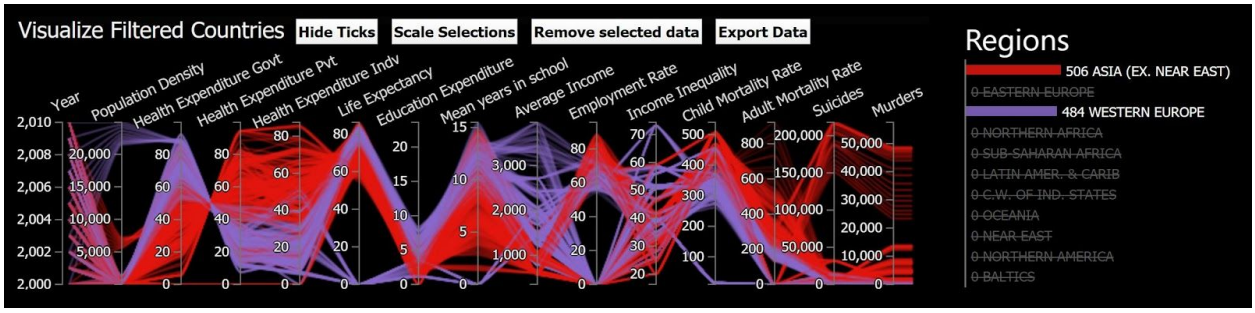
1. **Regions with lowest Health Expenditure by Government**
   We see from the above image that Asia and Sub-Saharan Africa that has the majority of the world's population has the lowest Health Expenditure by the Government.



2. **Comparison of top features obtained by PCA**
   The United States has the highest health expenditure by government while India has the highest Adult Mortality rate. There was an increase in the Adult Mortality Rate in India from the year 2000 to 2010.



3. **Comparison of Asia and Western Europe**
   From the below plot, we can say that Western Europe has high Health Expenditure by the Government with high average Income and Low suicides and murder rates. The mean years in school is also higher whereas when compared to Asia that has comparatively low Health Expenditure by the Government with low average income and higher suicide and murder rates.



4. **Child Mortality Rate V/s Adult Mortality Rate**
   Strangely, we can see that in Asia, the child mortality rate is comparatively higher than the adult mortality rate.

Conclusion:

From the above observations we can come up with some conclusions like regions with low average income have higher Murder and Suicide rates. And regions like Westen Europe that have comparatively higher Education expenditure has higher Employment and Income rate.

References for Datasets:

[1] https://www.gapminder.org/data/
[2] Region: Kaggle - https://www.kaggle.com/fernandol/countries-of-the-world/version/1
[3] Population Density: https://population.un.org/wpp/
[4] Health Expenditure: https://www.who.int/gho/en/
[5] Life Expectancy: https://population.un.org/wpp/
[6] Education Expenditure: UNESCO
[7] Mean years in school: http://ghdx.healthdata.org/record/global-educational-attainment-1970-2015; http://www.healthmetricsandevaluation.org/
[8] Average Income: http://gapm.io/dgdppc
[9] Employment Rate: https://www.ilo.org/ilostat/
[10] Income Inequality: http://gapm.io/ddgini
[11] Child Mortality Rate: https://www.gapminder.org/data/documentation/gd005/

[12] Adult Mortality Rate: (1) United Nations Population Division. World Population Prospects 2017 Revision, (2) University of California, Berkeley and Max Plank Institute for Demographic Research. Human Mortality Database.
[13] Suicides: https://www.healthdata.org/
[14] Murders: https://www.healthdata.org/
[15] Military Expenditure: https://data.worldbank.org/indicator/MS.MIL.XPND.GD.ZS