Education Statistics Visualization

28th October 2018

BASIC INFO

- Greeshma Mahadeva Prasad, u1141804, greeshma@cs.utah.edu
- Tanvi Gangadhar, u1205740, gangta@cs.utah.edu

Project Repository- https://github.com/magsheer/education-statistics-visualization

BACKGROUND AND MOTIVATION

Initially when we started looking at ideas for this project, we came across many that seemed interesting to us but did not have much success when we tried finding good datasets for them. For example, visualizing black friday shopping trends across various age groups and cities was one of our ideas but we could not find appropriate dataset it.

Later we started looking at datasets itself that seemed interesting for us and we came across one such dataset, education statistics by world bank. This dataset has indicators that describes literacy, enrolment rate, illiterate population among different sexes and age groups across education levels(primary/lower secondary/secondary/tertiary). We decided to go ahead with this dataset as education can be linked with various important parameters such as the country's economic growth, crime rate etc. We also wanted to draw parallels between the economic state of the country (low income vs high income) and education rate to understand the factors influencing education over the years.

PROJECT OBJECTIVES

We have considered a lot of possible questions that we would like to answer as part of our analysis and visualization of data. The primary questions being :

- How the indicators such as literacy rate with respect to males/females, illiterate population etc, has changed in all countries over the last five decades?

- Which are the most literate/illiterate age groups?
- The highest level of education among the countries in particular regions?
- Does the income of the country impact its education rate?

As mentioned earlier, education can be related to various other factors that influence our society. Visualizing such a dataset might help analyze the impact, education has on the growth of the world. In more developing countries, it would be beneficial to see the gender parity in education and whether it is moving towards equality in education among the two sexes.

DATA & DATA PROCESSING

We got out dataset from https://datacatalog.worldbank.org/dataset/education-statistics, The World Bank's Data Catalog.

The dataset includes multiple data subsets and is quite large. It consists of over 4,000 internationally comparable indicators that describe education access, progression, completion, literacy, teachers, population, and expenditures. We do not intend to use all of them for the purpose of this project as this might slow down the visualization. We plan to use indicators that fall under the three essential categories Sex, Age groups and Education level. This will allow us to effectively visualize the Gender Parity Index, highest level of education obtained across regions/countries and drop out rate among different age groups with respect to the economic state of the country. We also intend to show a global trend by comparing country statistics with global data. Since the data we want to use for this visualization is scattered over the data subsets, we need to extract only the data we require depending on the indicators, region and countries. We plan to achieve the data extraction using python or javascript.
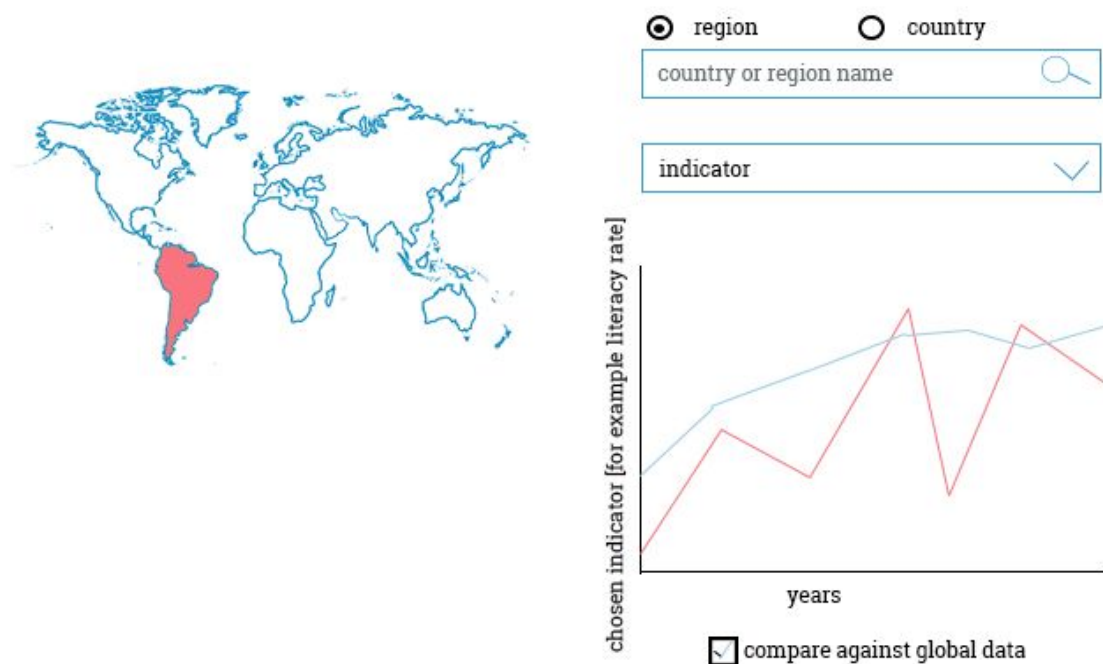
VISUALIZATION DESIGN

Prototype 1

We used some ideas from the assignment 4 for our first prototype. The features include:

Alternative 1:

- Select whether a **region or country** has to be visualized using the radio buttons
- Select the particular country/region using the **map or typing it** in
- Select the **indicator** to observe the trend over years
- Check the **compare** option to compare the global and country/region specific trends.

Alternative 2:

- Select the **country** you want visualize the data for
- Select a single **indicator** from the provided dropdown
- Based on the country and indicator selected, the visualization will show, **the trend** for the particular indicator for last five decades.
- For example, choosing India as the country and sex as the indicator, we show a pie chart with the ratios since there are only two values whereas selecting India as the country and education level as the indicator, will show adjacent bar charts.
- On clicking the bars or sections, details will be shown

Country ⬇ Indicator ⬇

[Ex: India]

Sex
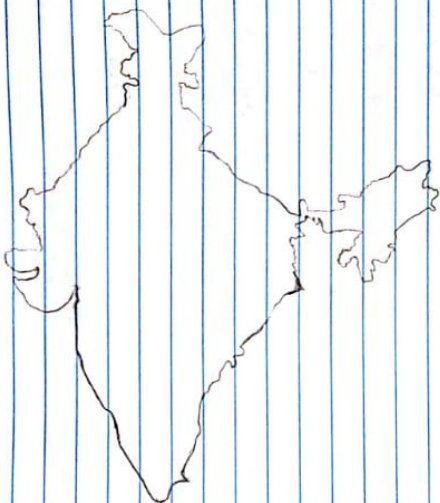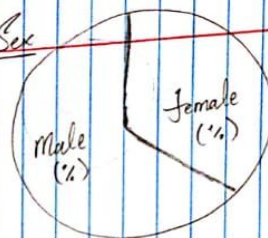
Male (%)   Female (%)

Age Groups

Population

Children   Young Adult   Adult

Age Groups

▨ Male
☐ Female

Level Of Education

▨ Male
☐ Female

Population

Primary   Secondary   College   ...

% Of Collegiate Women

Percent of Women
No. of Women
...

Prototype 2

Prototype 3

Final Design

- World Map that is colored based on the regions such as Sub Saharan Africa, East Asia & Pacific etc
- Sunburst Chart that displays hierarchical data. First level of the hierarchy is represented by regions which drill down to countries, with the World being the innermost circle at the top of the hierarchy
- Bar Chart that shows the trend for the selected indicator over the last 5 decades
- Select the region/country that you want to visualize the trend for
- Select the **indicator** to observe the trend over years
- Check the **compare** option to compare the global and country/region specific trends.

MUST-HAVE FEATURES

- Being able to visualize data region-wise and country-wise for at least 5 indicators
- Being able to compare the trend for a particular country with the world data

OPTIONAL FEATURES

Heat Map that shows :

- How a selected country compares over countries in the same region as the selected country
- How a selected region compares with the other regions in the world

PROJECT SCHEDULE

Week 1 [Oct 29 - Nov 4]

- Do the data processing required for the visualization
- Host the project on github
- Significant work done on getting a working prototype

Week 2 [Nov 5 - Nov 11]

- Have the working prototype ready

Week 3 [Nov 12 - Nov 18]

- Work on incomplete functionalities of the prototype
- Have the basic visualizations working without any glitches

Week 4 [Nov 19 - Nov 26]

- Enhance the visualizations
- Add rest of the planned functionalities
- Fix bugs

Week 5 [Nov 26 - Nov 30]

- Test the visualization
- If there is time, add more extra features and work on making the project better