Project Edu+?

Project Proposal document

HCDE 511 - Information Visualization; Winter 2020

- **Student names:**
  - Anqi Cao: caoanqi@uw.edu
  - Amitabh Nag: amnag@uw.edu
  - Susan Zheng: susanz96@uw.edu
  - Shiyao Chen: shiyaoc@uw.edu

- **Project name**

  Edu+?

- **Project concept and goals. What is the purpose of the visualization?**

  Knowledge is power and education is key for people to improve social mobility. We are very interested in examining the historical data in world education. There are other factors that may relate to education attainment and we listed them below:

  - Life expectancy (https://ourworldindata.org/human-development-index)
  - Suicide (https://ourworldindata.org/suicide)
  - Happiness (TBD)

  We want to show the correlation between education attainment and life expectancy/suicide rate/happiness level. Our hypothesis is that the higher education level leads to higher life expectancy, lower suicide rate, and higher happinese level. By visualizing the data, we want to reassure the purpose of education, and investigate how the world population is learning

from a historical standpoint, and how life expectancy, suicide rate & happiness level change overtime in relation to education

- **Users:**
  - Elementary school teacher: <Susan to find teacher from China>
  - Teacher: Educationist ;Redmond High school teacher: Mr. Garing <Amitabh>
  - Education department at UW: <Shiyao to contact UW Education department for teacher>
  - Author from New York Times: <Anqi>

- **Early sketches and storyboards of initial ideas of visualizations and interactions**

We plan to visually represent data using a world map, and users can see more detailed visualizations when they hover over a specific country.

- **Discussion of related work:**

There are many visualizations of global educational attainment using the Barro-Lee's data set. The World Bank has an educational attainment dashboard illustrating the distribution of educational attainment and human capital in over 100 countries from 1970 to 2010. Viewers can select a country, time period, and/or age group below to view the data.

The organization, Our World in Data from a more specific variable, like primary school enrollment, mean years of schooling. Map was used to show the distribution and line charts to show the change. They also did some work presenting demographic change in educated people. Other education topics are also analysed like determinants and consequences of education.

- **Roles to be performed by team members**

  We will follow agile/scrum methodology and we will not have fixed roles. There will be a backlog and each team member will pull tasks from that backlog

- **Week-by-week schedule**

Steps	Weeks	Tasks	Deliverable
Define	Week 3	Determine the rough topic Gather data sets	Preliminary Project proposal (1/24 5pm, Sat)
	Week 4	Try different data joining Determine the databases Data cleaning	Revised proposal (2/1, Sat)
Explore	Week 5	Data analysis Initial visual visualization Interview intended users	Exploratory Data Analysis (2/5, Wed)
Develop	Week 6	Further data analysis Complete visual mapping	Midterm presentations (2/13 Thurs)
	Week 7	Complete interactive visualization tool prototype	Prototypes
Test & refine	Week 8	Usability testing sessions and quick iterations on the prototype	Usability issues & prototypes
	Week 9	Final work	Final version of interactive visualization tool
Wrap-up	Week 10	Prepare for presentation Paper writing	Final presentation (3/12 Thurs)

			Final Paper (3/16 5pm)

- **Data that will be used in the visualization**

  Dataset #1: Primary dataset

  Name: Barro-Lee Education Attainment Data

  Description: The Barro-Lee Data set provides education attainment data for 146 countries from 1820 to 2010 disaggregated by sex and 5-year intervals. It also provides information about the distribution of educational attainment of the adult population over age 15 and over age 25 by sex at seven levels of schooling— no formal education, incomplete primary, complete primary, lower secondary, upper secondary, incomplete tertiary, and complete tertiary. Average years of schooling at all levels—primary, secondary, and tertiary—are also measured for each country and for regions in the world.

  Source: http://www.barrolee.com/

  Dimensions:

  Country name; Year; Age; Sex;

  Primary adjusted enrollment ratio (%); Secondary adjusted enrollment ratio (%); Tertiary adjusted enrollment ratio (%)

  Percentage of no schooling; Percentage of primary;Percentage of primary complete; Percentage of secondary; Percentage of secondary complete; Percentage of tertiary; Percentage of tertiary complete

  Years of schooling; Years of primary schooling; Years of secondary schooling; Years of Tertiary Schooling;

  Human capital, population aged 15-64 years; Alternative human capital, population aged 15-64 years pop; Population (thousands)

  How we intend to use it: Play around with the dataset to generate a geospatial graph that allows the audience to have an overview of the rate of education attainment growth globally, filter by time period and type of education attainment and zoom in on the details associated with each geographical region.

Size: 12987 rows x 23 columns

Dataset #2: Life Expectancy: Secondary dataset

Name: Time series data of differences in life expectancy across the world

Description: The dataset aggregate the life expectancies of countries from 1543 to 2019 with gaps for certain countries and certain time periods

Source: https://ourworldindata.org/life-expectancy

Size: 19028 rows x 4 columns

Dimensions: Country; country code, year; life expectancy

How we intend to use it: Correlate the time series data of life expectancy by country with the corresponding country's data of education attainment

Dataset #3: Suicide: Secondary dataset

Description: The dataset provides share of suicide deaths from 1950 to 2017 by country

Size: 6469 rows x 4 columns

Dimensions: Country; country code, year; % of deaths from suicide

How we intend to use it: Correlate the time series data of suicide % by country with the corresponding country's data of education attainment

Dataset #4: Happiness: Secondary dataset

This is a stretch goal - Dataset TBD

- **Which tools are you considering using to accomplish the goals (this can change if needed).**
    - Visualization tool: Tableau
    - Ad-hoc data exploration: Excel
    - Data cleanup tool: Tableau prep
    - Data cleanup (programmatic): Python
    - Data/Script/visualization repository: Github and/or shared google drive
    - Visualization hosting: Tableau public

- **What kinds of results we anticipate achieving?**

  There are these aspects of results we anticipate achieving:

  1) Identify the trend of the rate of changes and the type of changes in education attainment globally
  2) Identify the correlation between changes in education attainment over time and suicide rate / terrorism-caused death over time
  3) Identify the correlation between changes in education attainment over time and life expectancy over time
  4) Identify correlation between education attainment of a country and happiness (stretch goal)

- **Results we would like to achieve but do not have the time or the tools for?**
  1) We wanted to build a web based visuzlation using tools such as d3. However due to lack of time, we are opting for Tableau
  2) We would have loved to have a user from the United Nations. However, neither of the team members have contacts and United Nations is located at New York. Instead we are reaching out to teachers from US/China and NY Times author who can evaluate our work
  3) We would have loved to work on homelessness. However, we did not find a dataset that is large enough to meet the project's data size requirements