

# Project Edu+?

## Project Proposal document

HCDE 511 - Information Visualization; Winter 2020

- **Student names:**

- Anqi Cao: [caoanqi@uw.edu](mailto:caoanqi@uw.edu)
- Amitabh Nag: [amnag@uw.edu](mailto:amnag@uw.edu)
- Susan Zheng: [susanz96@uw.edu](mailto:susanz96@uw.edu)
- Shiyao Chen: [shiyaoc@uw.edu](mailto: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 attainment and enrolment. There are other factors that may relate to education attainment and enrolment. These are:

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

We want to test out our hypotheses:

- Over the years, education attainment has had an upward curve globally, regionally and nationally.
- With the increase in education attainment, suicide rate increases for developed countries and decreases for developing countries, life expectancy increases for all countries.

- At a global level, the average life expectancy has increased. For countries with higher rates of change in education attainment over a certain time period, the increase in the average life expectancies of these countries is larger during the same time period
- Asia-Pacific (excluding Australia and New Zealand), Africa and South America have higher rates of changes in education attainment compared to North America and Europe

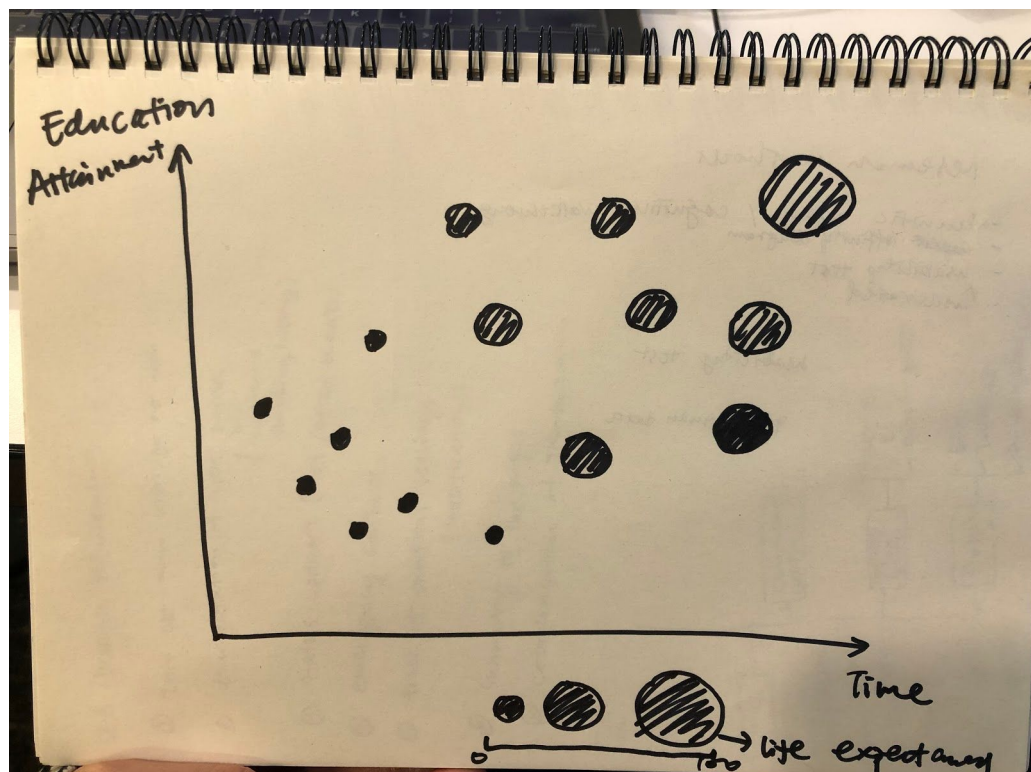
- **Users:**

- The Analyzer - people who need to read and use data visualizations to explore details and identify patterns, and decide what stories they can tell to inform their projects
  - Ziqi: PhD student in Economics; worked on data visualization projects for his internship at a hedge fund; consumes data visualizations on a regular basis.
  - Holly: Researcher at a policy think tank; reads news articles with data visualizations on a regular basis for both personal interest and work.
- The Curious Reader - General Students
  - College students: students who want to learn about if spending all that time and money on education is worth it for their lives.

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



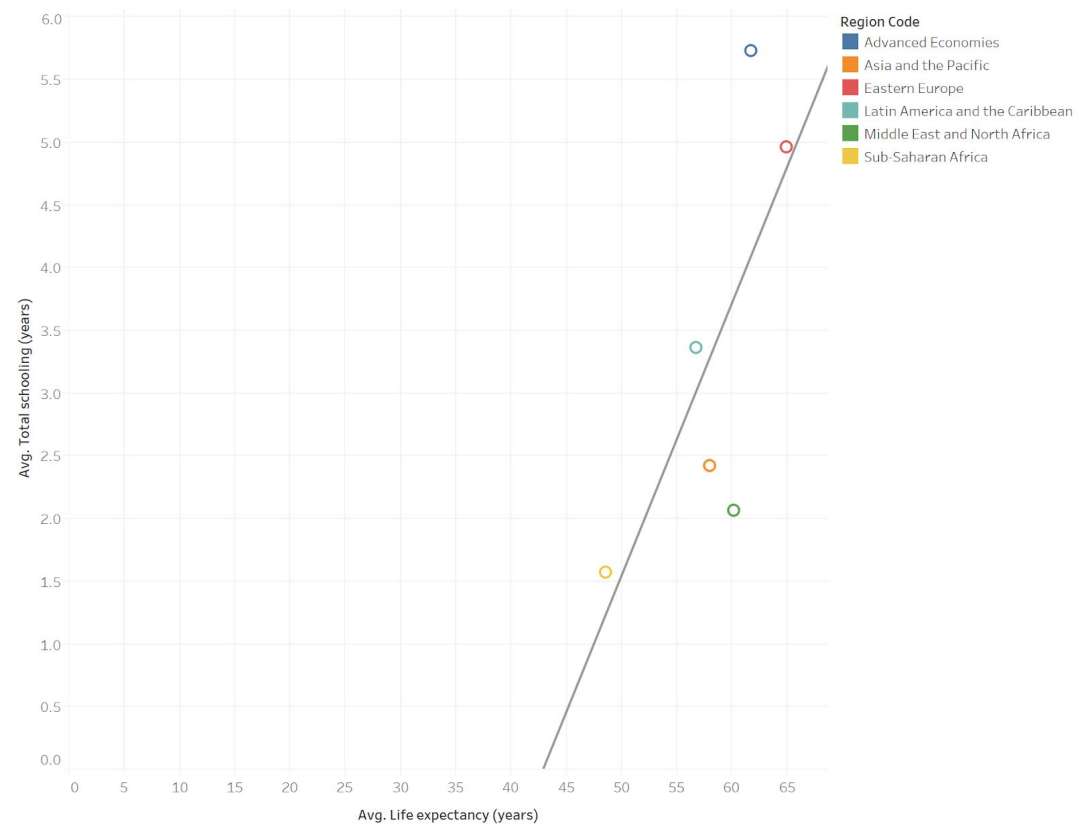
First sketch: visually represent data using a world map, and users can see more detailed visualizations when they hover over a specific country.



Second sketch: possible visualization of relationships among education attainment, time, and life expectancy.

Interaction between years of schooling, life expectancy by region:

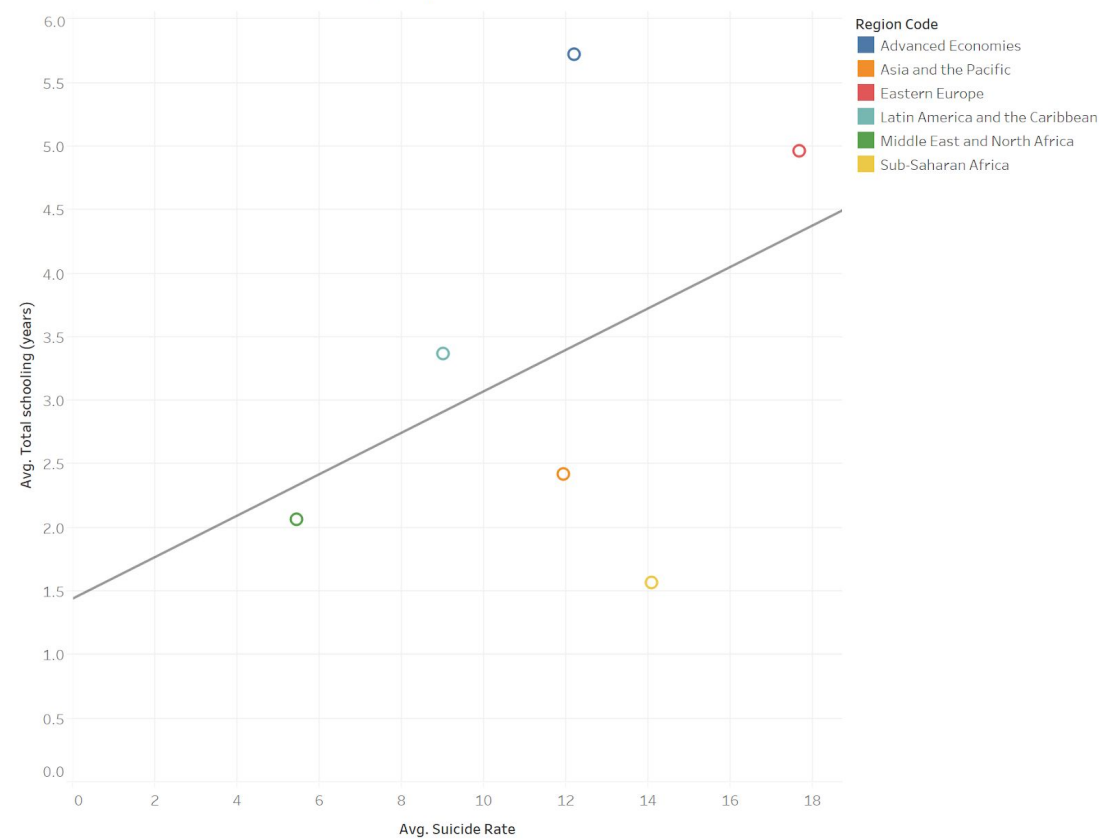
Life expectancy vs Total schooling by Region



Average of Life expectancy (years) vs. average of Total schooling (years). Color shows details about Region Code.

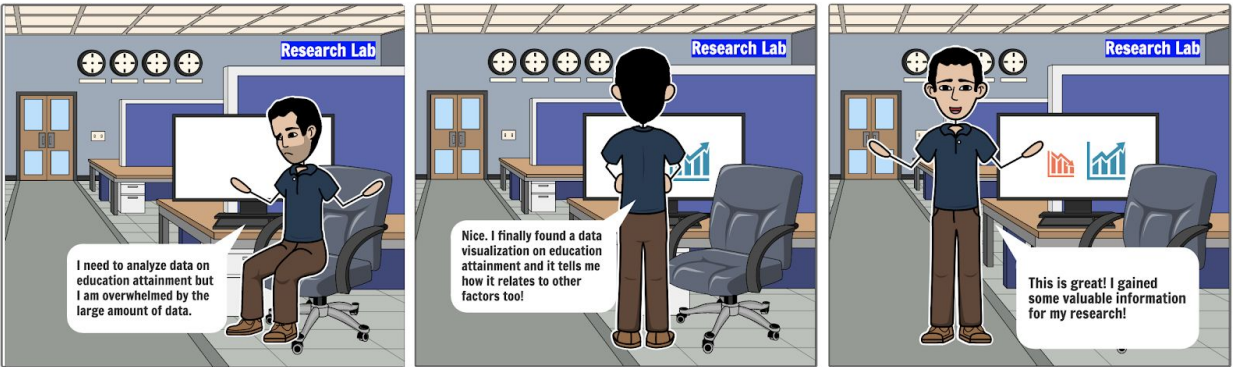
Interaction between years of schooling, suicide rate by region:

Suicide rate vs Total schooling by Region

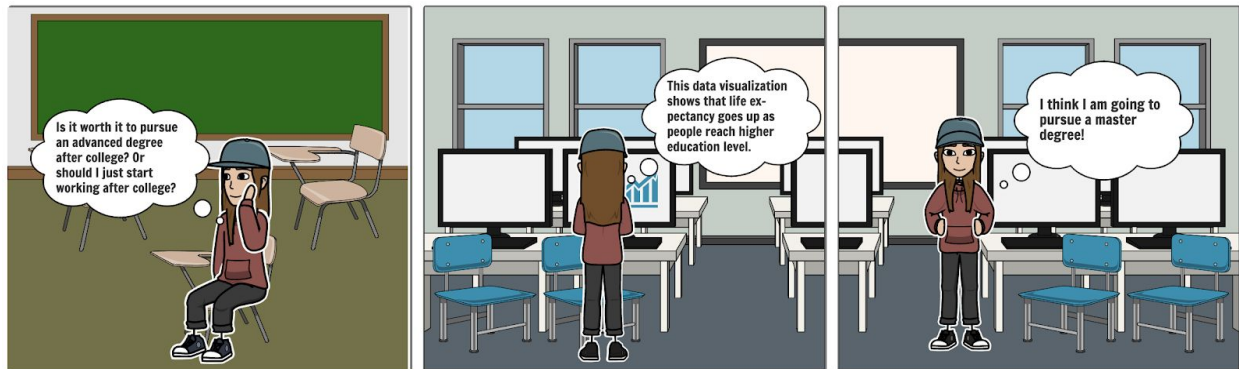


Average of Suicide Rate vs. average of Total schooling (years). Color shows details about Region Code.

Storyboard for the analyzer persona:

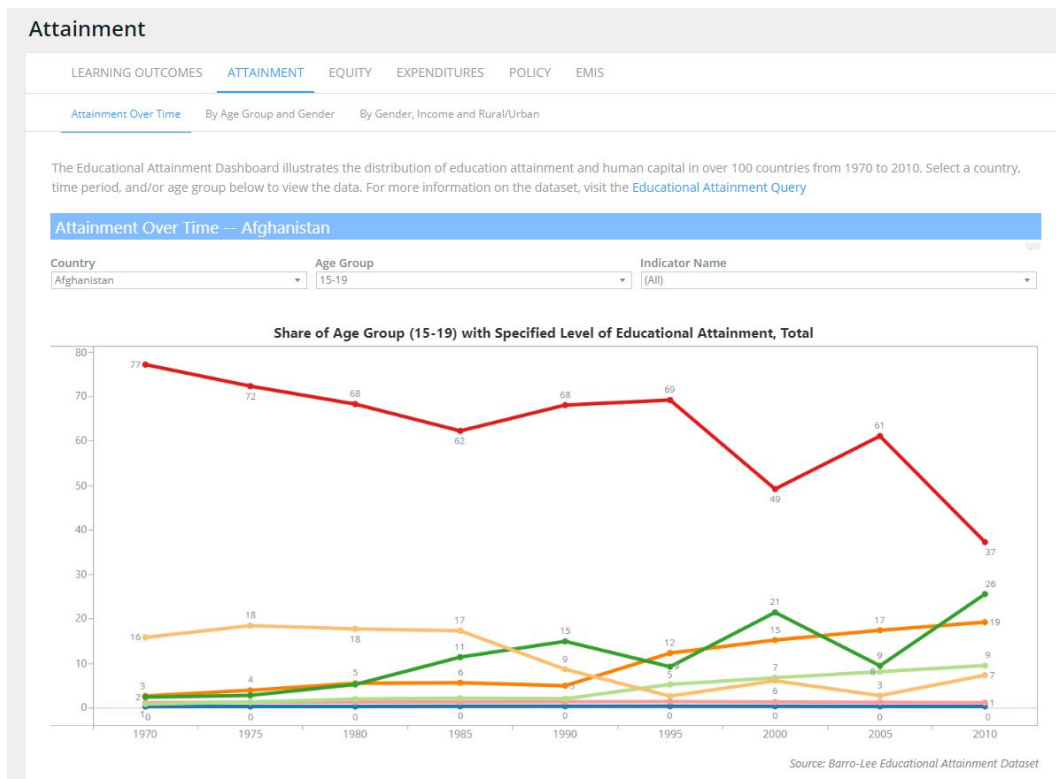


Storyboard for the student persona:



- **Discussion of related work:**

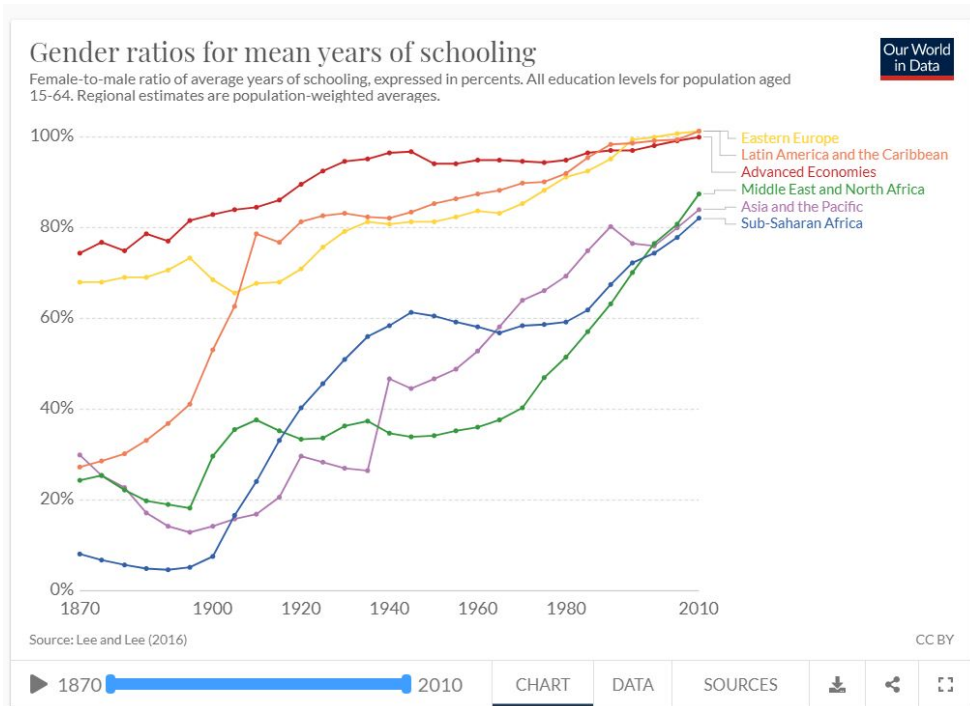
There are many visualizations of global educational attainment using the Barro-Lee's data set. The World Bank has an educational attainment dashboard illustrating the distribution of educational attainment and human capital in over 100 countries from 1970 to 2010. Viewers can select a country, time period,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



<http://datatopics.worldbank.org/education/wDashboard/dqattainment>

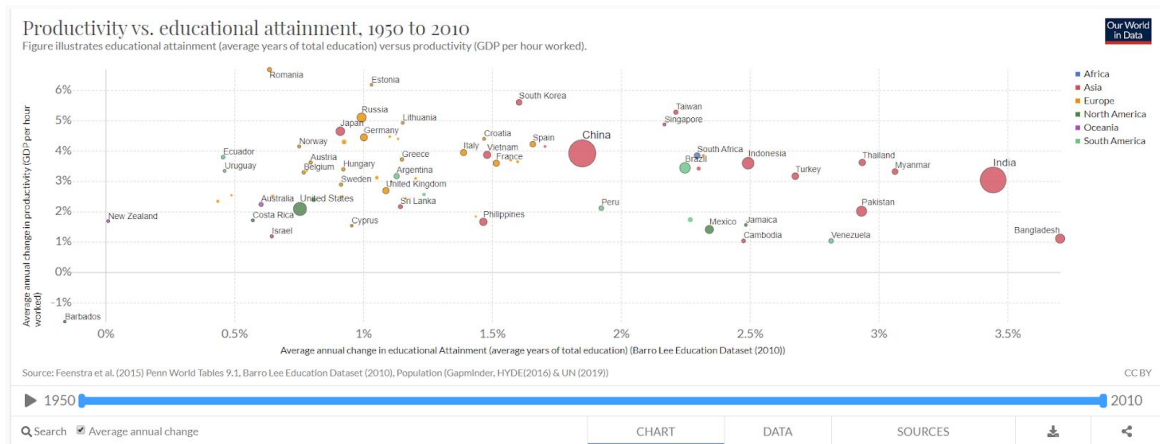


The organization, Our World in Data use it from a more specific variable, like primary school enrollment, mean years of schooling. Map was used to show the distribution and line charts to show the change. They also did some work presenting demographic change in educated people. Other education topics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



<https://ourworldindata.org/global-education#school-enrollment-and-attendance>

This figure illustrates educational attainment (average years of total education) versus productivity (GDP per hour worked). It categorized all the countries by regions, including Africa, Asia, Europe, North America, Oceania and South America. It showed the correlation of these two attributes in line chart and the annual changes in scatter plot.



<https://ourworldindata.org/grapher/productivity-vs-educational-attainment?time=1950..2010>

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

We will follow a modified agile/scrum methodology and we will not have fixed roles. We will sync twice a week in person/Skype and daily on Slack. We will use Github as a platform for data storage, task assignment and documentations. We will assign tasks to each member via Github based on their interests, skillsets and earliest availability. Here is the link to our github repo where we will track tasks, code, Tableau files and documentation:

<https://github.com/edu-infoviz/edu/>

- **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 and data types:**

- Country name (nominal)
- Year (interval)
- Age (ratio)
- Sex (nominal)
- Primary adjusted enrollment ratio (%) (ratio)
- Secondary adjusted enrollment ratio (%) (ratio)
- Tertiary adjusted enrollment ratio (%) (ratio)
- Percentage of no schooling (ratio)
- Percentage of primary (ratio)

- Percentage of primary complete (ratio)
- Percentage of secondary (ratio)
- Percentage of secondary complete (ratio)
- Percentage of tertiary (ratio)
- Percentage of tertiary complete (ratio)
- Years of schooling (ratio)
- Years of primary schooling (ratio)
- Years of secondary schooling (ratio)
- Years of Tertiary Schooling; (ratio)
- Human capital (ratio)
- population aged 15-64 years (ratio)
- Alternative human capital (ratio)
- 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 and data types:**

- Country (nominal)
- country code (nominal)
- Year (interval)
- life expectancy (ratio)

**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 and data types:**

- Country (nominal)
- country code (nominal)
- Year (interval)
- % of deaths from suicide (ratio)

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

We have merged all three datasets by joining them on country and year as keys. This process required country name cleanup and was done in python. Merged data is available here:

[https://github.com/edu-infoviz/edu/blob/master/data/edu\\_merged\\_data.csv](https://github.com/edu-infoviz/edu/blob/master/data/edu_merged_data.csv)

Cleanup and merge script is available here:

<https://github.com/edu-infoviz/edu/blob/master/edu/EduProject.ipynb>

- **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
- 
- **Results we would like to achieve but do not have the time or the tools for?**
    - 1) We wanted to build a web based visualization 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