



Final Report: Project Edu +
HCDE 511 Information Visualization
Winter 2020, Group #5

Prepared by Amitabh Nag, Anqi Cao, Shiyao Chen, and Susan Zheng

Keywords:
Data visualization, education attainment, suicide rate, and life expectancy.

Table of Content

1. [Abstract](#)
2. [Introduction](#)
3. [Previous work](#)
4. [Data quality](#)
5. [Target user personas](#)
6. [Overview of Design Milestones](#)
7. [Detailed Design Process](#)
8. [Final system evaluation](#)
9. [Future work](#)
10. [References](#)

Abstract

Project Edu Plus is a course project for HCDE 511 Information Visualization. As a team of four, we used Tableau to create an interactive visualization system of longitudinal education data at regional and national levels and two secondary datasets on dimensions of mental and physical health for correlational study. There are two purposes of this project: 1) To allow users to explore and analyze patterns of education variables and their relationships with health variables. 2) To craft stories for interpretations by curating, visualizing and presenting time-series, multivariate data. Following the User-Centered Design process, we started by generating target user personas via secondary research and iterated on our prototypes based on the feedback we gleaned from two rounds of usability testing sessions with target users interspersed throughout the design process. The final visualization system "[Edu + ?](#)" can be viewed on Tableau Public.

Introduction

Education, as a fundamental resource for both individual and societal development, has been a field studied and researched by a plethora of scholars and professionals. The world has witnessed a remarkable expansion of education over the past two hundred years, with the growth in total years of schooling, highest degree of education completed and reduction in illiteracy rate. Our team was interested in segmenting education data by different dimensions such as the ones listed above and exploring how discrepancies have played out across regions, countries and gender from a historical perspective. With the rich educational dataset provided by [Barro-Lee](#), we can ask and answer interesting questions including but not limited to: How has tertiary education attainment changed historically, especially during WWII, compared to primary and secondary education attainments? What's the gender disparity in total years of schooling for Advanced Economies (as defined by the International Monetary Fund) compared to Sub-Saharan Africa?

After settling on the primary dataset, we started to discuss potential secondary datasets. As one of the Sustainable Development Goals outlined by the United Nations in striving for a more sustainable future, Education is also closely interlinked with the other sixteen goals such as Economic Growth, Gender Equality, Good Health and Well-Being, etc. This interconnectedness has naturally led us to wonder to what extent the Education data at our disposal shows correlation with the other facets of human development across the globe. In choosing the secondary datasets to focus on, we became interested in the interaction between Education and Health/Well-Being. Based on the availability of datasets and research, we decided to unpack Health/Well-Being by operationalizing it as two important indicators of public health: Life Expectancy and Suicide Rate.

Existing research literature has already established a strong positive correlation between educational investments and life expectancy, but the intensity of interaction between the two can vary. For instance, the degree of association between education attainment and health is supposedly mediated by income and moderated by demographic background such as ethnicity

and gender. Past research documented how Black people experienced diminishing returns of education attainment on subjective health compared to White people. Inspired by such existing findings, we were eager to enable exploration / storytelling about the variation in the interaction between education attainment and life expectancy across time, regions and countries.

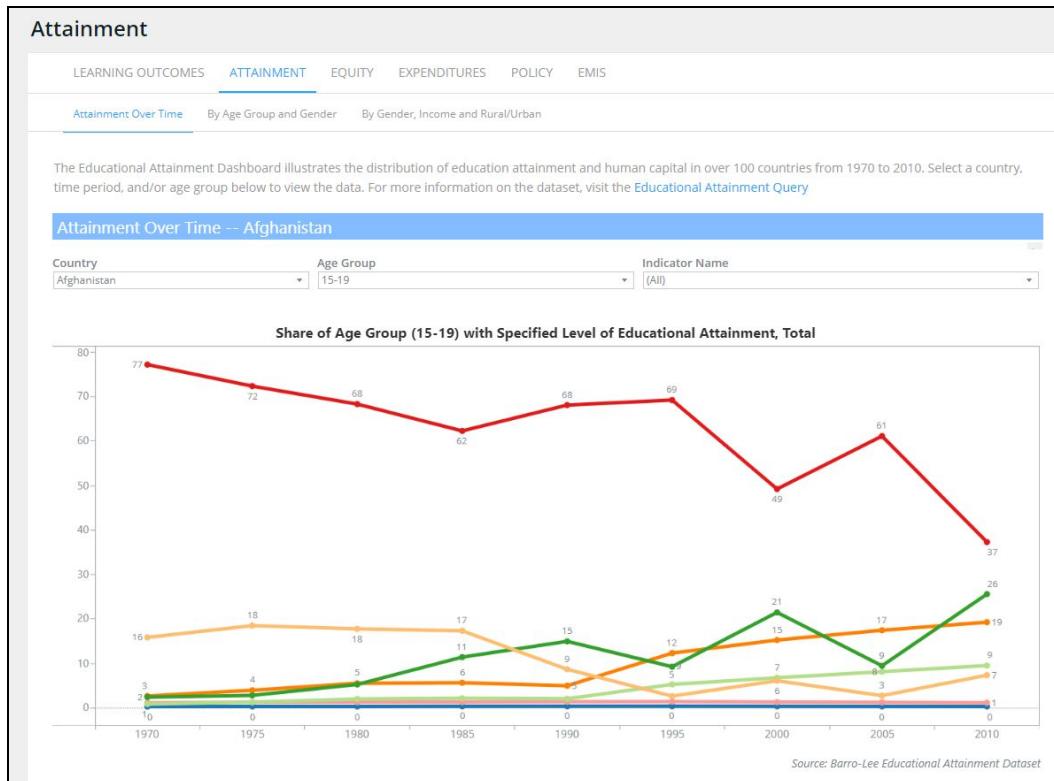
In the meantime, the correlation between education attainment and suicide rate has been less studied compared to the two topics above, despite the fact that suicide has been a leading cause of mortality worldwide. Past research suggested that risk factors of suicide are embedded in the socioeconomic, cultural and familial contexts within which the individual lived prior to committing suicide. It is reasonable to assume in our project that income and psychological knowledge, which have been proven to be consequences of education attainment, correlate with suicide rate. Here we eliminated these mediating factors to reduce the multitude of variables involved in the visualization system and directly correlated suicide rate and education. In doing so, we could spot answers to questions such as: What are the countries on the higher end of education attainment with high suicide rates? What about those that are the opposite - low education attainment with low suicide rates?

We were, at the same time, aware of the fact that the global trends in suicide have been under-documented and our database, therefore, may be far less exhaustive and representative than we hoped for a rigorous visualization system to display. In addition, suicide as a complex social phenomenon warrants explanatory factors more immediate to its operating context than education. However, we were hoping that the dimension of mental health brought by suicide data, when put in dialogue with physical health (life expectancy) as both are correlated with education, could add to the diversity and interactivity of the whole project.

Previous Work

Global educational attainment dashboard | The World Bank

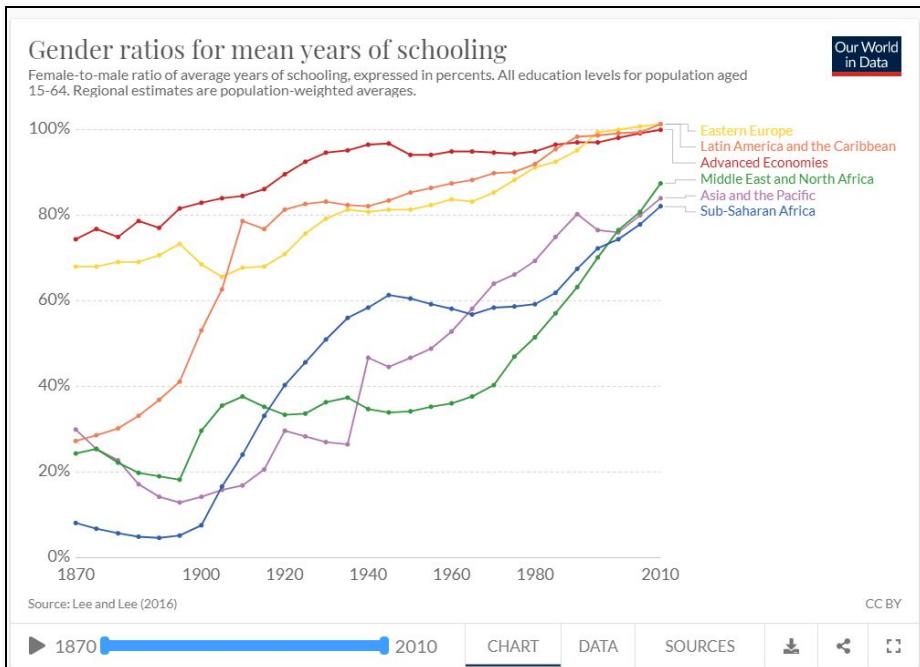
There are many visualizations of global educational attainment using the Barro-Lee's data set. The World Bank has an educational attainment dashboard illustrating the distribution of educational attainment and human capital in over 100 countries from 1970 to 2010. Viewers can select a country, time period,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



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

Gender ratio in educational attainment | Our World in Data

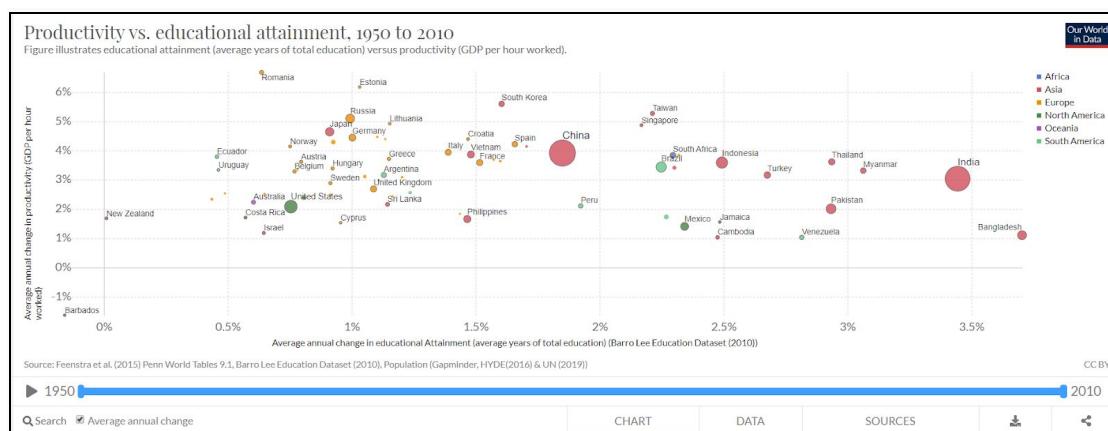
It shows the evolution of female-to-male ratios of educational attainment (mean years of schooling) across different world regions. There has generally been a strong upward trend in the gender ratios across all world regions, which indicates that the inequality between men and women in access to education has been declining.



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

Annual change in productivity vs educational attainment | Our World in Data

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 the line chart and the annual changes in scatter plot.



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

Data Quality

We used three datasets for this project. All three datasets were available as csv. We downloaded the csv and processed them using python and Tableau prep. Final output of Tableau prep was three csv. We used two of the three in our project. Here are the three datasets we used in the project:

1. Primary: Long-term education attainment
 - Education attainment from 1820 to 2010 by sex, country (N=146) and region (N=6); 5-year intervals
 - 12987 rows x 23 columns
 - Variable types: 20 ratio; 2 nominal; 1 interval
 - <http://www.barrolee.com/>
2. Secondary: Life expectancy
 - Life expectancy from 1543 to 2019 by country (N=236)
 - 19028 rows x 4 columns
 - Variable types: 2 nominal; 1 interval; 1 ratio
 - <https://ourworldindata.org/life-expectancy>
3. Secondary: Suicide rate
 - Number of deaths from suicide per 10,000 from 1950 to 2017 by country (N=200)
 - 6469 rows x 4 columns
 - Variable types: 2 nominal; 1 interval; 1 ratio
 - <https://ourworldindata.org/suicide>

Original datasets are available on our github repo here:

<https://github.com/edu-infoviz/edu/tree/master/data/original>. Processed datasets are available here: <https://github.com/edu-infoviz/edu/tree/master/data/processed>

Although the data sets are very rich, they posed following quality challenges:

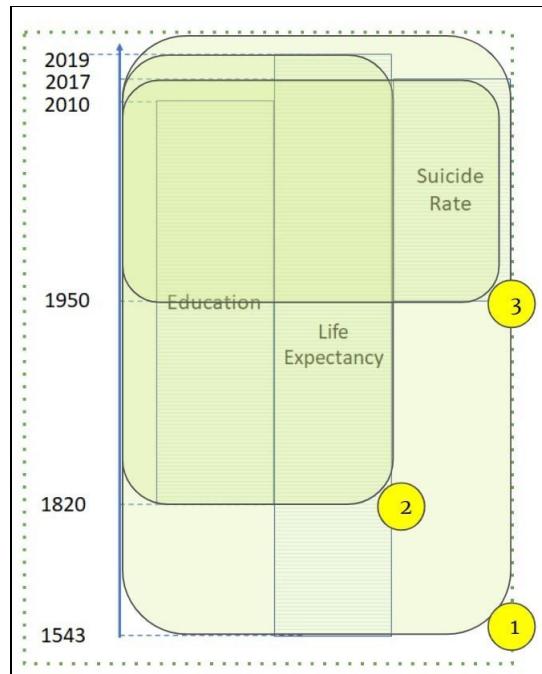
- Seven country names are found to be mismatched between the datasets. For example, the education dataset has “Congo, D.R.” whereas Life Expectancy and Suicide Rate has “Democratic Republic of Congo”
- Years in the three datasets were not aligned. For example, education data is from 1820 to 2010, in five-year intervals. Life expectancy dataset ranges from 1543 to 2019 and Suicide rate data ranges from 1950 to 2017
- Education data had a partly informative data dictionary. www.barrolee.com/ did provide a data dictionary, however the data dictionary only provided an expansion of acronyms in the data variables. For example, “lu” maps to “Percentage of no schooling”. However, the data dictionary or the website did not provide relationship between the data attributes. For example, one would expect following to add up to 100%:
 - Percentage of no schooling
 - Percentage of primary
 - Percentage of secondary

- Percentage of tertiary

However, this was not true for all rows of the dataset. Numbers were close to 100%, but not exactly 100%.

We solved the country name challenge by data cleanup using python. We then cleaned data further in Tableau prep to align data on year and country. We created three datasets from Tableau prep:

- 1) Edu_suicide_rate_life_expectancy_nulls (large)
- 2) Edu_life_expectancy_no_null (medium)
- 3) Edu_suicide_rate_life_expectancy_no_null (small)



Conceptual representation of three datasets created from Tableau prep. 1,2,3 refer to the csvs mentioned above

We used datasets 2) and 3) for our interactive dashboard. Cleanup python and Tableau prep files are available on our repo here: <https://github.com/edu-infoviz/edu/>

Target User Personas

We conducted a research on the potential users of data visualization projects and identified 4 main types of users: the Doers, the Decision Makers, the Analyzers and the Curious Readers. Considering that ours is a class project that is not situated in a specific work/business setting to inform immediate actions and decisions aligned with any top-level strategies, we decided to target the Analyzers and Curious Readers among the four categories.

The Analyzers are people who care deeply about why's and what if's. Their primary need to satisfy is **exploration**. They love data visualization tools that allow them to explore the details by adjusting the metrics and dimensions and discover patterns. It's often their responsibility to figure out what narratives can be built from the datasets being studied.

The Curious Readers are people whose personal interests and concerns draw them to the topic of global education development and its correlation with health. Their primary need is to learn about the **stories** that the data show them intuitively. This group of people tend to consist of students, journalists, NGO workers and anyone with a strong curiosity to learn, make unexpected connections of storylines and expand their knowledge base. Their engagement with the data visualization is not necessarily tied to any problem-solving scenario.

With these two target user personas in mind, we recruited 4 real users for our usability testing sessions throughout the design process. Here's a summary of the profiles of the real users as they are mapped to the personas. In the following sections, the testing users will be referred to by their respective ID number (P1, P2, P3, P4).

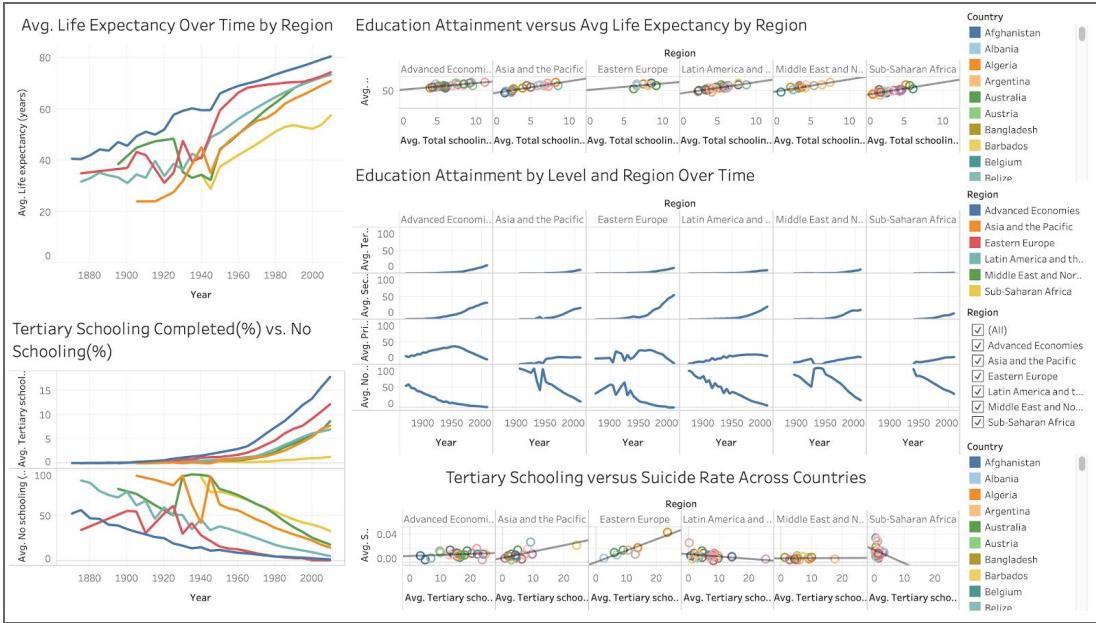
	The Analyzer		The Curious Reader	
ID	P1	P2	P3	P4
				
Occupation	PhD Candidate	School Teacher	Principal Software Engineer	Undergraduate Student

Experience with Data Viz	Education & Work experience in Development Economics; Intermediate-high level of proficiency in data viz tools	Education background in Business Finance; Intermediate level of proficiency in data visualization	Experience in building dashboards for KPIs; High level of proficiency in data visualization	Understands basic data visualization; Low level of proficiency in data visualization
Purpose of Use	Exploration		Story-telling	
Testing Round	#1	#2	#2	#2

Overview of Design Milestones

Home Dashboard 1.0

The first version of our dashboard was a collage of individual exploration pieces on one canvas. The two dot plots show the correlations of education attainment and life expectancy, suicide rate by region. The two line charts on the left show the trends of life expectancy and education attainment by region. The 4x6 line chart compares different levels of education by region.



Home Dashboard 2.0

This is our initial interactive dashboard, with a navigation system and a more clear information architecture compared to Dashboard 1.0. Our focus for this stage is to allow exploration as well as perceivable stories. For exploration, we extended each topic from overview and details. Users can explore each topic from a global trend, comparison between regions, diving into a specific country on a time series. For a more comprehensible storytelling, different topics were discussed on separate dashboards with a clearly-defined navigation system made by several buttons. Each button directed the user to a separate dashboard.

Data Visualization Project: Education + ?

Context: HCD 511 Class Project
Team: Anqi Cao, Shiyao Chen, Amitabh Nag, Susan Zheng

Introduction:
Knowledge is power and education is the key for people to improve social mobility. We want to present the historical data in world education attainment and enrollment across the world. During our project, we also explored the correlation between education attainment and health (physical health and mental health). The data representing education attainment is total schooling in years, the one representing physical health is life expectancy, and the one representing mental health is suicide rate.

[Home](#) [Education & Mental Health](#) [Education & Physical Health](#) [Education & Health](#) [Education & Gender](#)

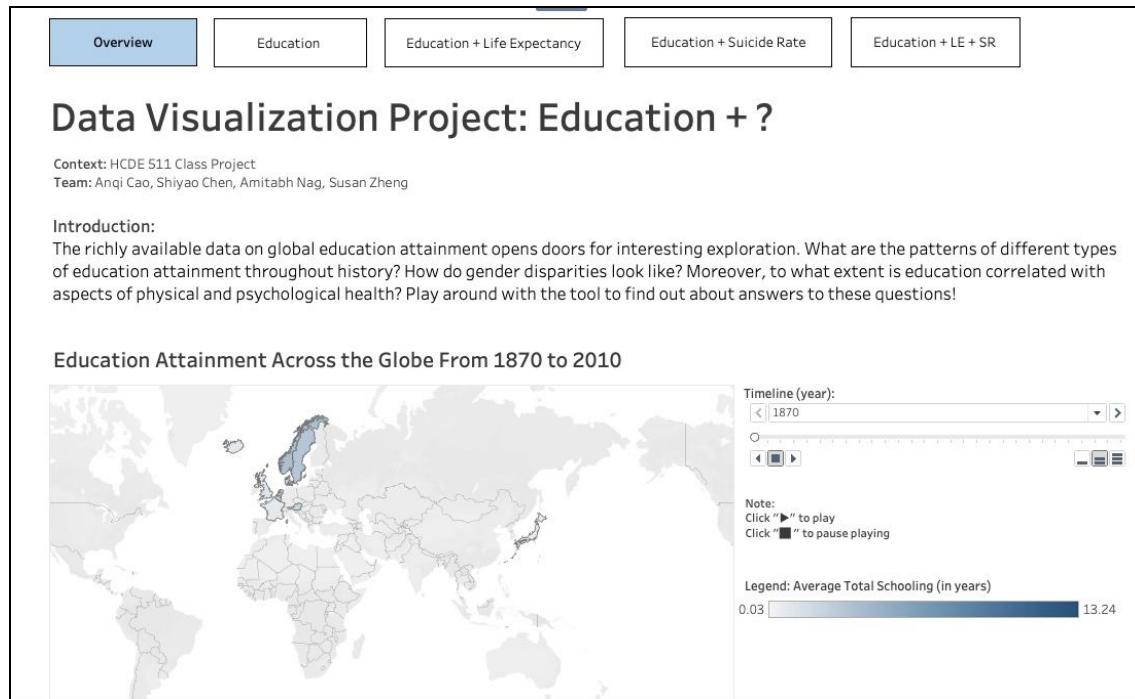
Education Attainment Across the Globe

Legend:
Average Total Schooling (in years)

0.03	13.24
------	-------

Home Dashboard 3.0

This was the final dashboard with polished navigation, visual layout to make it easier for users to interact and grasp the key information. Before users always need to go back to the home screen to switch between different sub-dashboards, the final version keeps the navigation bar on the top, so all the dashboard titles are visible on all screens.



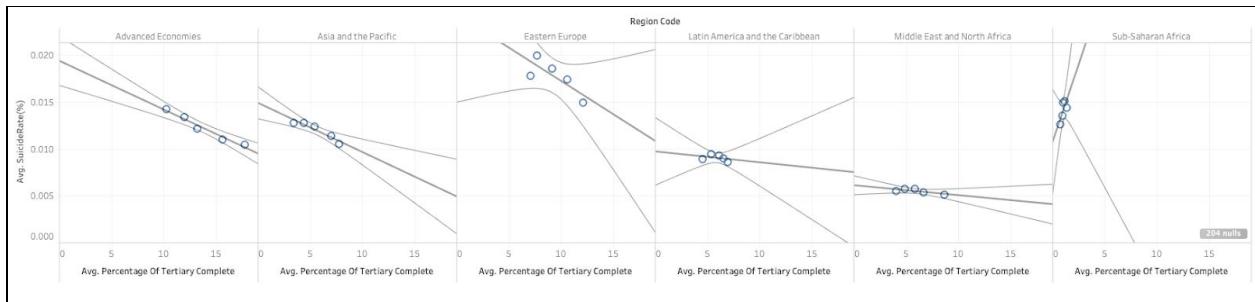
Detailed Design Process

Prototyping Stage 1.0

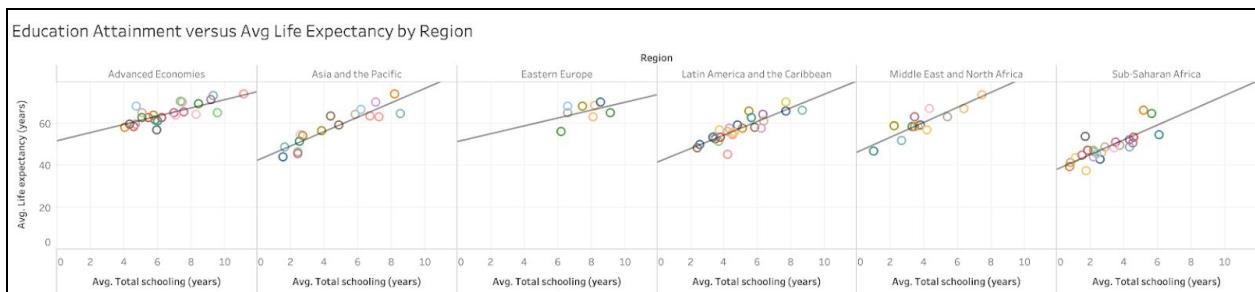
Synthesis of Individual Exploration & Findings

Our first step in the design process was to go over the artifacts we made individually/ in pairs for the data visualization exercise using our data. We decided to put together interesting insights and their corresponding charts first. Here's a list of some examples of insights we uncovered:

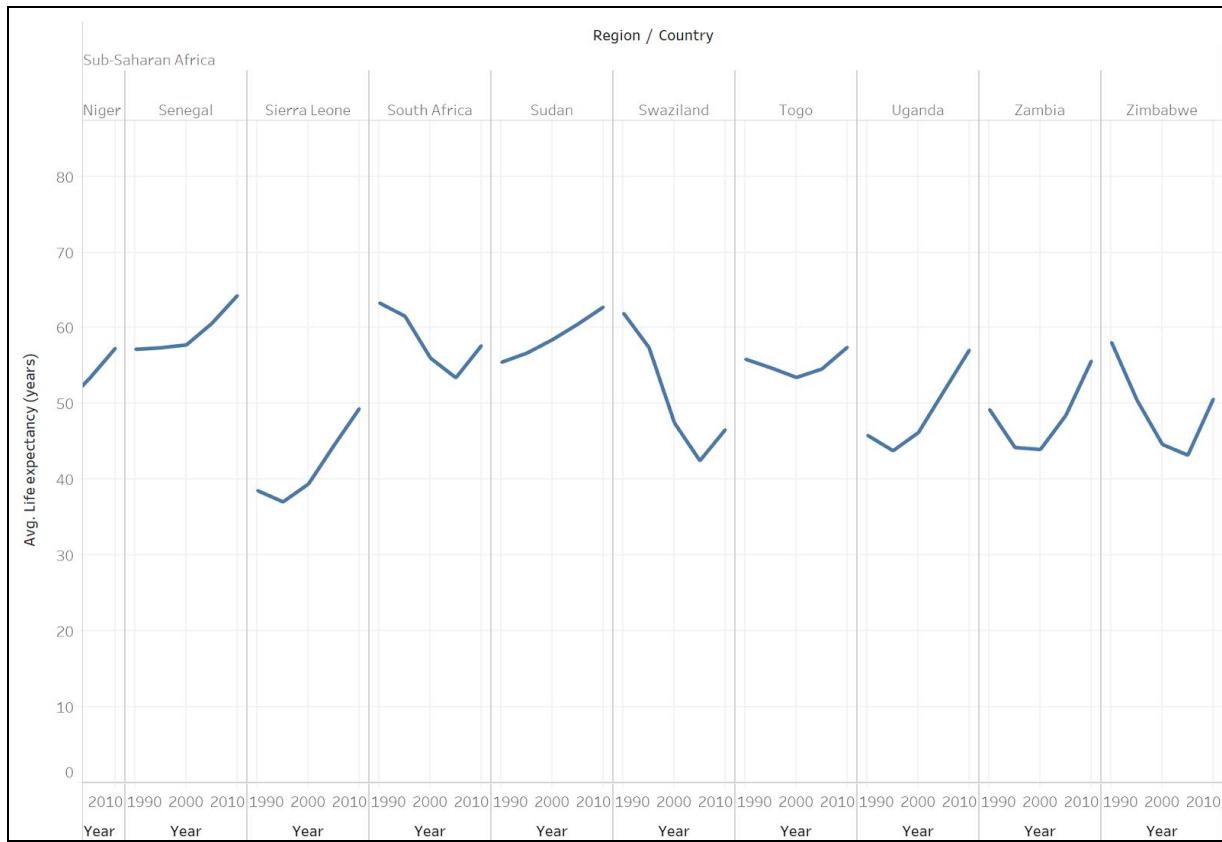
Insight #1: From 1990 to 2010, Advanced Economies experienced the most sharp increase in percentage of tertiary education degrees, whereas Sub-Saharan area had a barely visible amount of growth for the same variable, with suicide rate increased a lot relatively.



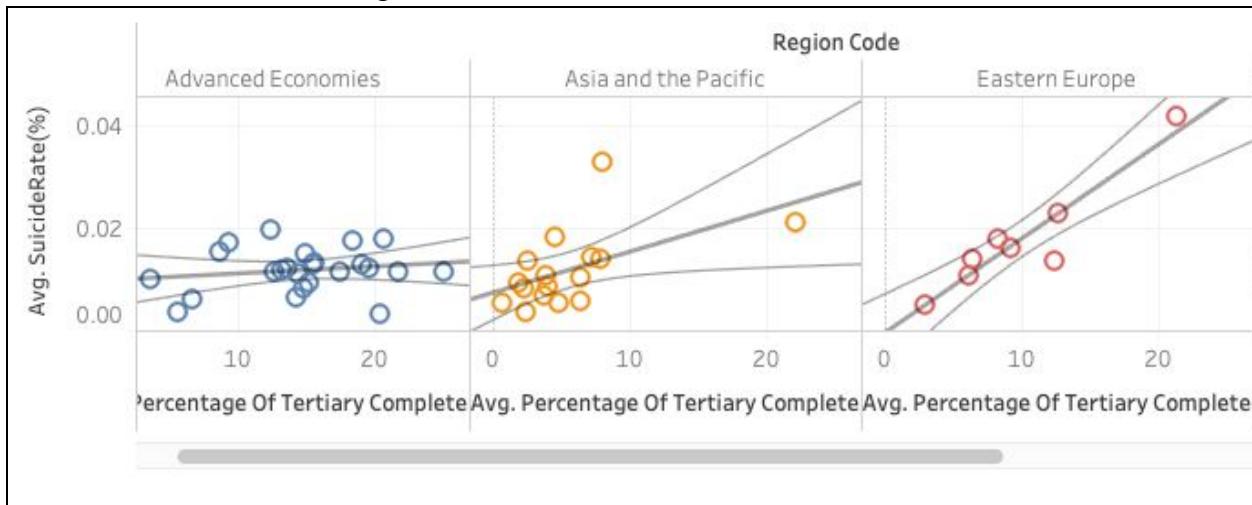
Insight #2: Correlation between education attainment (avg total years of schooling) and avg life expectancy has been positive for all regions throughout the years.



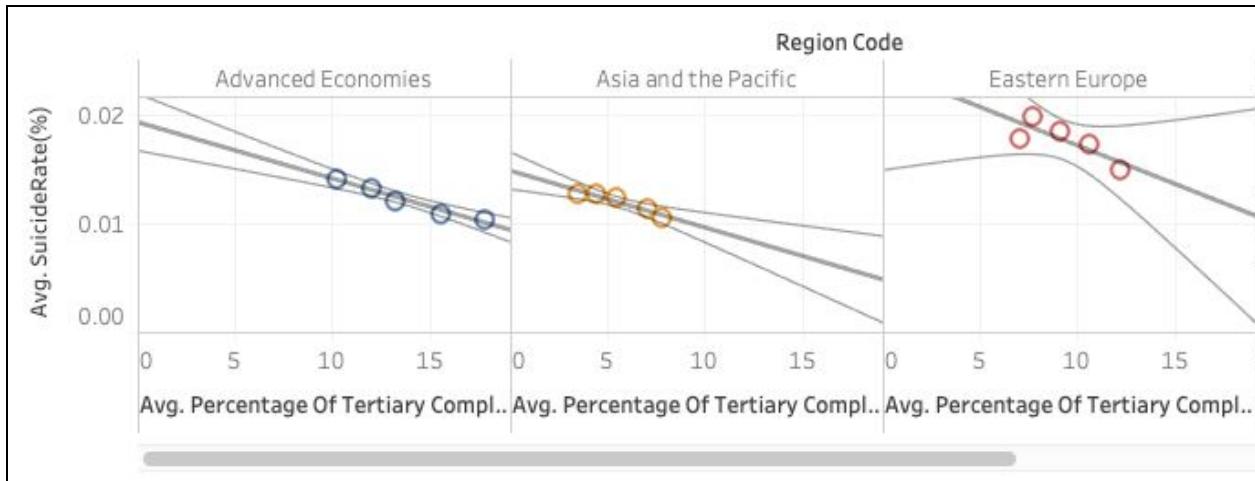
Insight #3: Life expectancy has a hockey stick pattern for mostly sub-saharan african countries, where it decreased in 1990s (due to the HIV epidemic) and picked up in 2000s.



Insight #4: Correlation between Tertiary Education Attainment and Suicide Rate was positive for Eastern Europe when individual country was used as the categorical variable, and negative when year was used as the categorical variable, indicating opposite patterns when the same data is looked at from different angles.

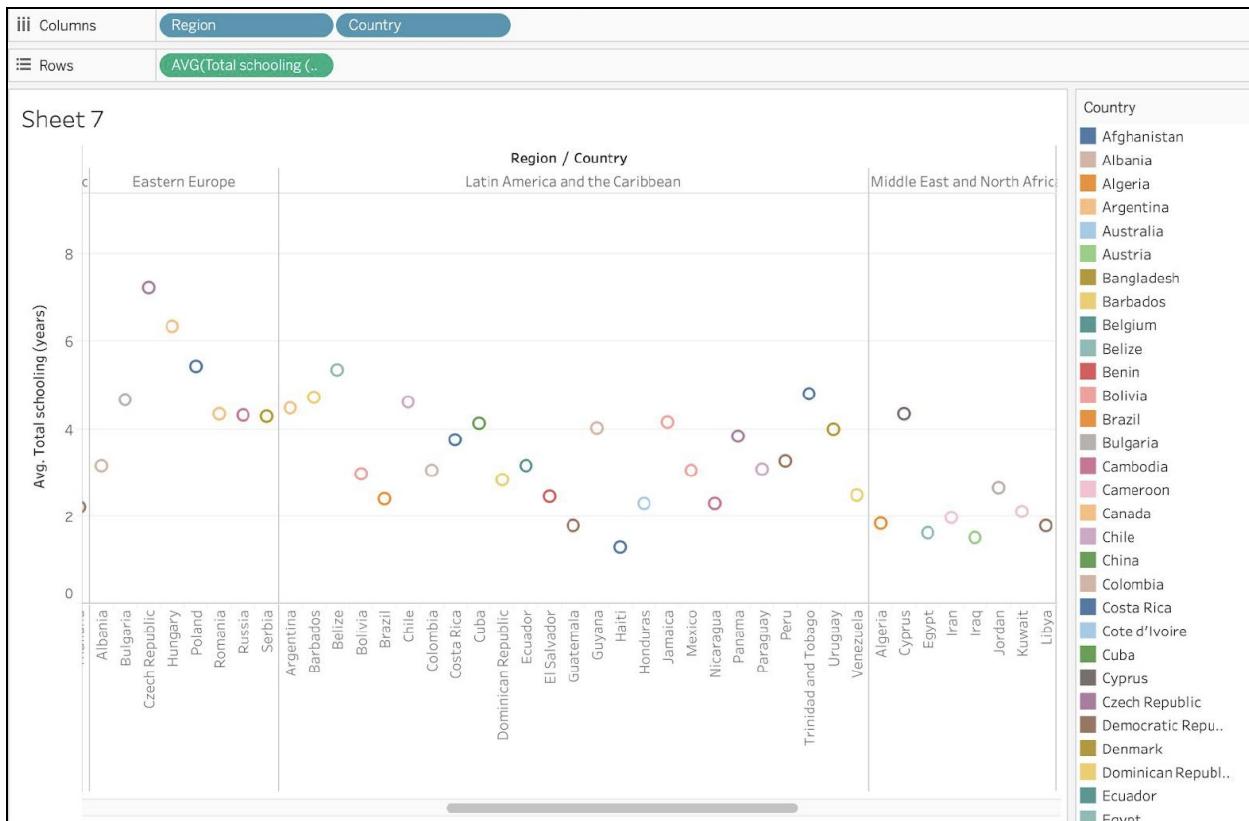


Graph #1: Each dot represents an individual country of the region



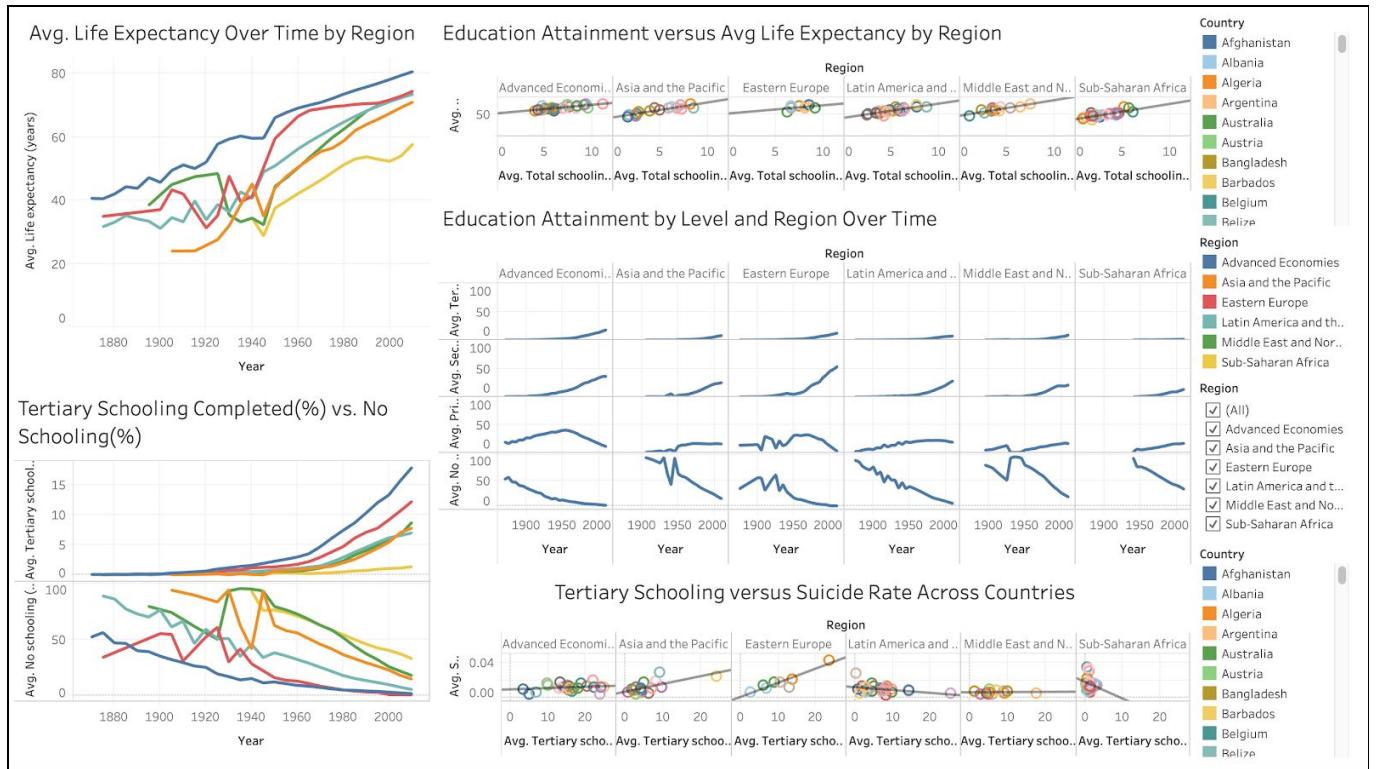
Graph #2: Each dot represents a year (from 1990 to 2010, in 5-year intervals)

Insight #5: Trends were also sometimes driven by outliers (e.g. Czech Republic's change in avg total years of schooling driving that of the whole eastern europe region)



First Version of the Main Dashboard

We decided to take a stab at creating a main dashboard, which turned out to be essentially a collage of different charts of sub-variables randomly selected from all three datasets. It combined all topics together on the same page, but was still a rough patchwork in need of further deliberation and refinement. Here's a screenshot of the draft dashboard 1.0:



Usability Testing on Dashboard 1.0

We conducted a usability testing session with P1 on our draft dashboard for feedback. Based on P1's responses, we generated a list of feedback and suggestions:

Issue with Dashboard Structure:

- P1 thought that usually when you put things together you want to compare them against each other, but currently there exists no link between these graphs.

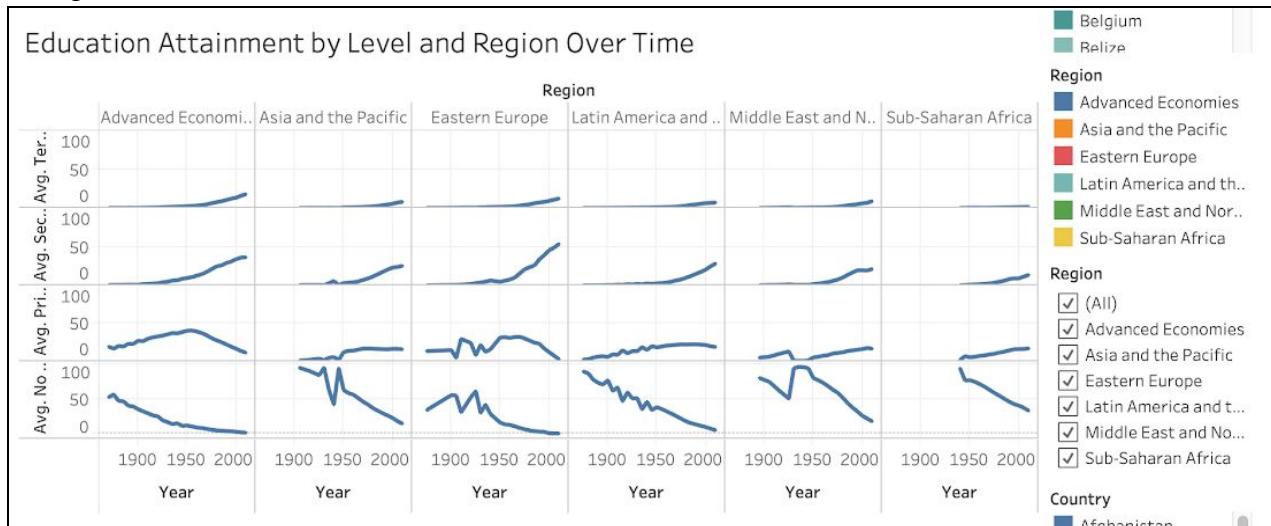
Example: See the chart above.

Recommendation: He proposed that maybe it's possible to show one graph only, and allow people to customize x and y variables (drop-down menus) depending on what they want to view.

Issues with Education Attainment by Level and Region Chart

- P1 was interested in the topic but not sure what you can learn from it, because the chart is disorienting when too many small sections are jumbled together.
- If P1 were to do it he would show who finished at least primary, secondary, tertiary schooling instead of at most, because it sounds more intuitive to him that way.

Example:



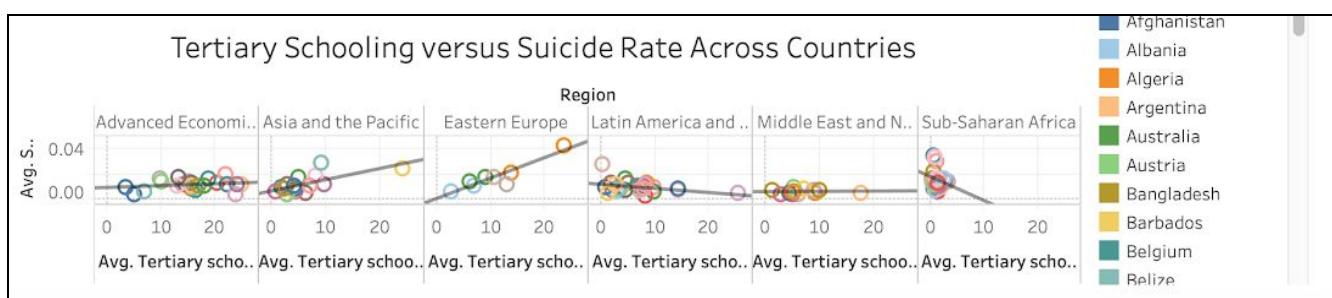
Recommendation:

- Present it interactively, allow users to select for themselves what they want to see, compare different curves within one region, or aggregate within each row, with drop-down menus of different types of data.

Issue with Tertiary Education Attainment vs Suicide Rate Chart:

- P1 was not sure about the take-aways from this chart: What kind of story can you tell?

Example:



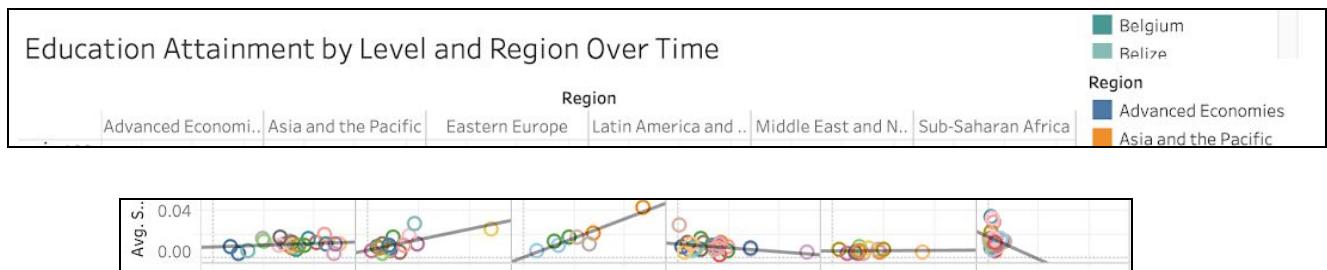
Recommendation:

- P1 proposed two potential narratives that can be made by displaying education and suicide rate data in two different ways: 1) When country is used as the categorical variable: for the countries where people are more educated, there might have been some intrinsic cultural and other factors that have pressured them to get more education and more mentally distressed to the point that they ended their own lives. Hence, they have been more inclined to commit suicide across time. 2) time as the categorical variable: when time goes forward, people in each region in general are getting more educated, and less likely to kill themselves, regardless of their country.

Issues with UI Design:

- The title is left aligned and doesn't look good
- Hollow circles in economics/social sciences are often used in such a way that their size is manipulated to denote the third dimension (mostly population or sample size)

Examples:



Recommendation:

- The title would look better if it's center aligned as opposed to left aligned
- If you are just denoting countries, you don't need to use circles, maybe use solid dots.

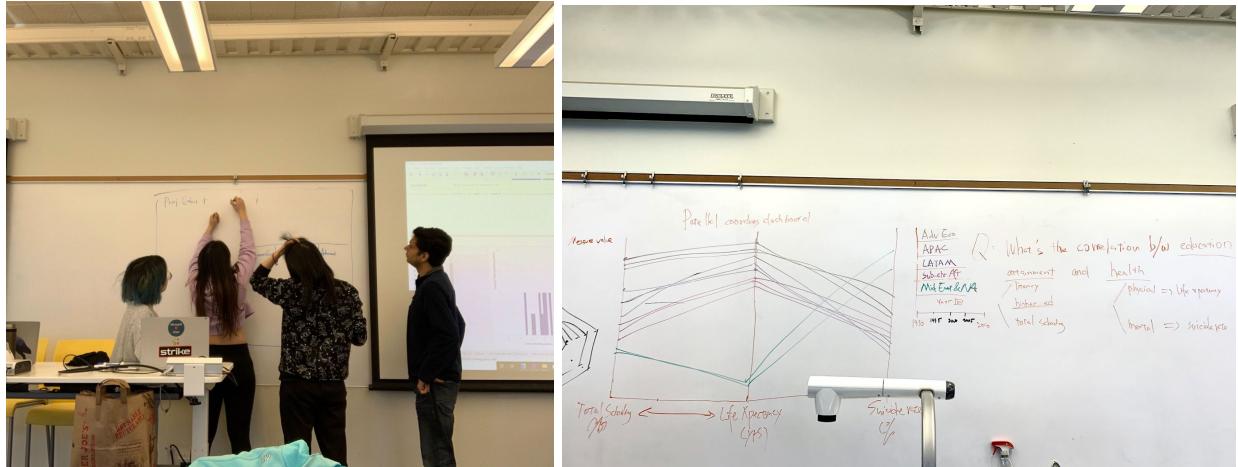
Prototyping Stage 2.0

Delineating changes to be made based on the first round of usability testing & class feedback

Based on the feedback gained from classmates, teacher and P1, we decided to incorporate 3 types of changes into our new prototype: 1) We added interactivity to the dashboards we designed by creating filters and animation buttons. In this way, users have control and freedom over which data variables they want to explore. 2) We also improved the interconnectedness among the dashboards by creating a narrative weaved into the navigation menu attached to the top of each dashboard. 3) We adopted a minimalistic approach to our Home Dashboard, eliminating all the other charts to leave only a data map and an integrated line chart for an overview of time-series education data. The users could then decide on their own which part(s) of the visualization they would love to dig deeper into.

Brainstorming & Sketching

We brainstormed which graphs can achieve the purposes mentioned above via group brainstorming sessions during which we sketched out different charts and interface layout on the whiteboard.



Selected photos of group work

Eventually we arrived at the following charts.

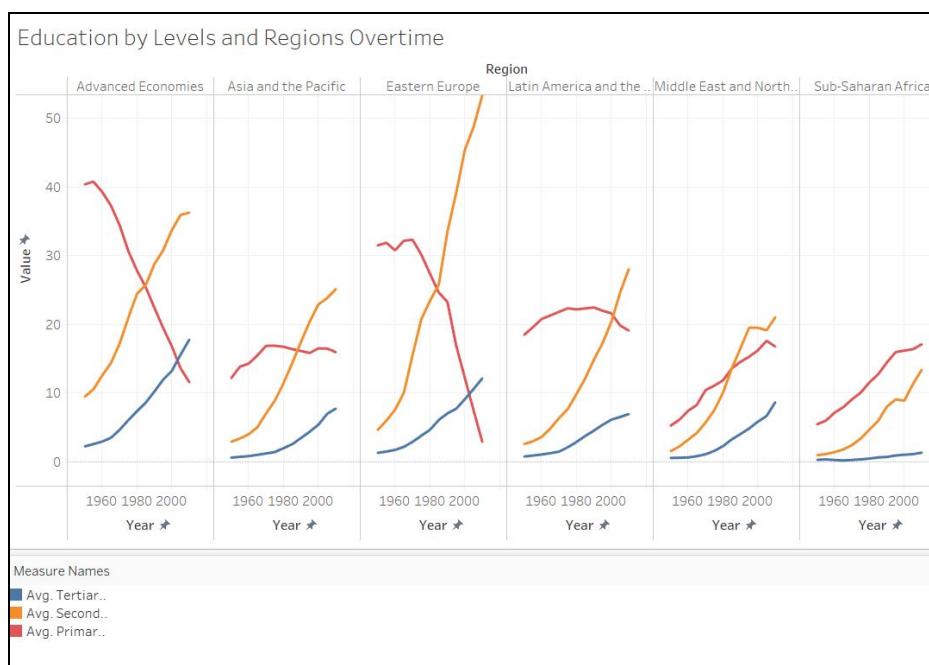
Education Data Map - Overview of Global Trend

Map was used here as maps are a good choice to visualize the distribution and change of a single variable across the world. It clearly shows which parts of the world have higher education attainment and which are less educated. In the animated state, viewers can easily get education attainment growing from the color getting darker with time.

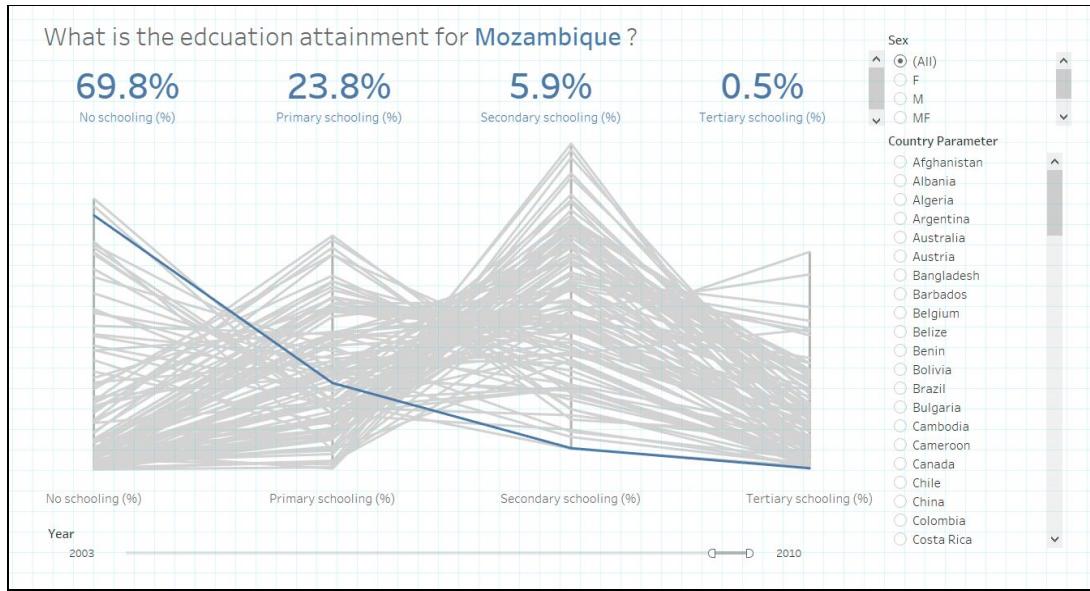


Education - Comparison Between Different Education Levels

There are several variables to represent different education levels, primary, secondary and tertiary. This side-by-side line chart was used to allow users to compare in regions. In dashboard 1.0, This line chart consist 4*6 units as each education variable of each region occupies one unit. Users found it's hard to compare so we made the three variable lines of each region in one graph, so it's easier to compare and get facts like all regions experience a relatively high increase in primary education, but slow in tertiary education, especially sub-saharan african areas.

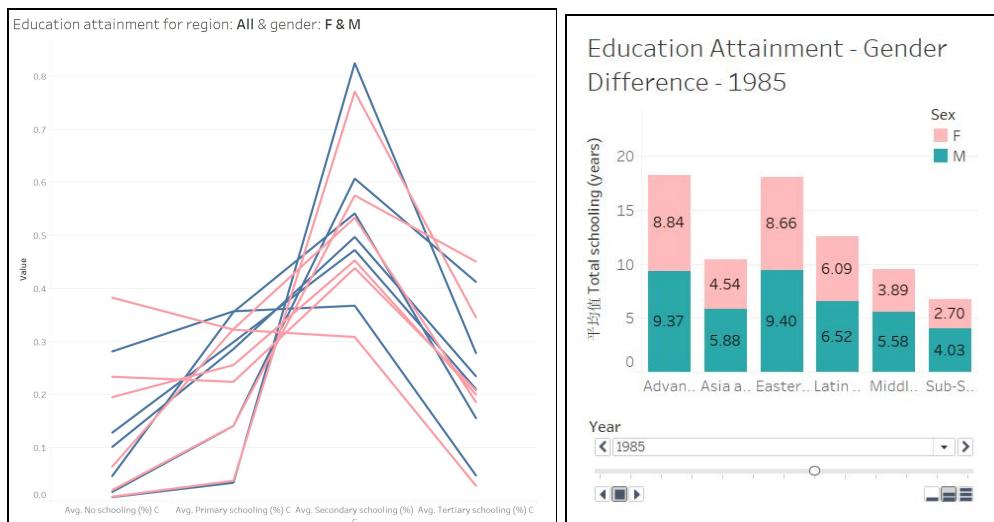


While the line chart was used to give overview, this parallel coordinates plot was more for exploration. Users can select one specific country to see the development of different levels of education and the corresponding numbers are shown above the graph.



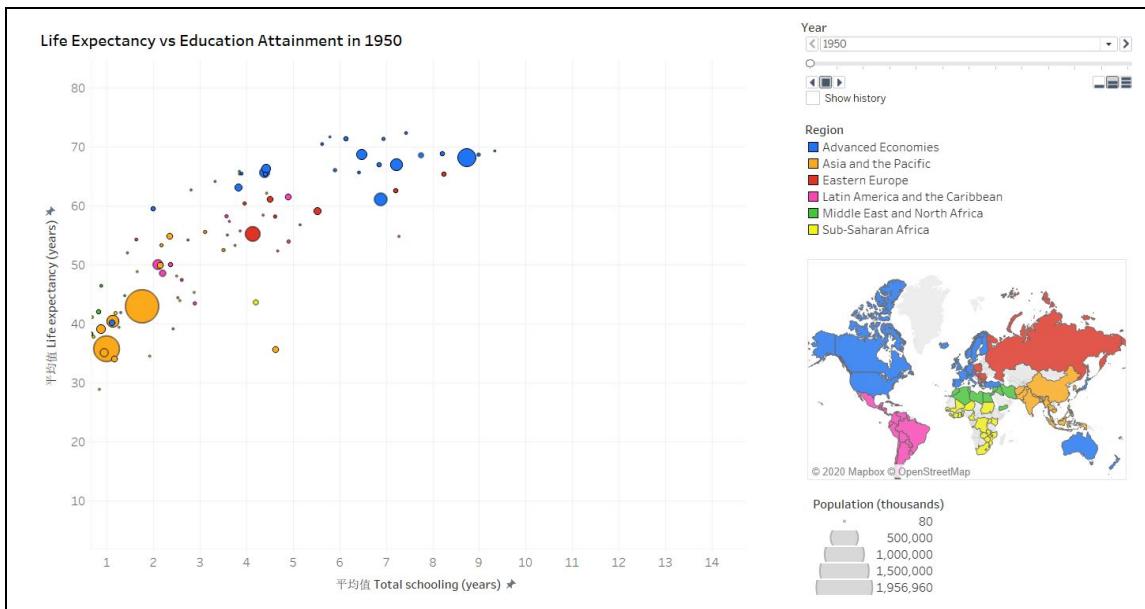
Education - Gender Difference

We first showed gender differences in education attainment in the parallel coordinates plot used to compare different education levels. Then we also made a separate graph for this topic using a stacked bar chart. In the animation, it can be seen for each region, the gender difference is decreasing. We would like to know which graph is more intuitive for them in the next round of usability testing.



Education Attainment vs Life Expectancy

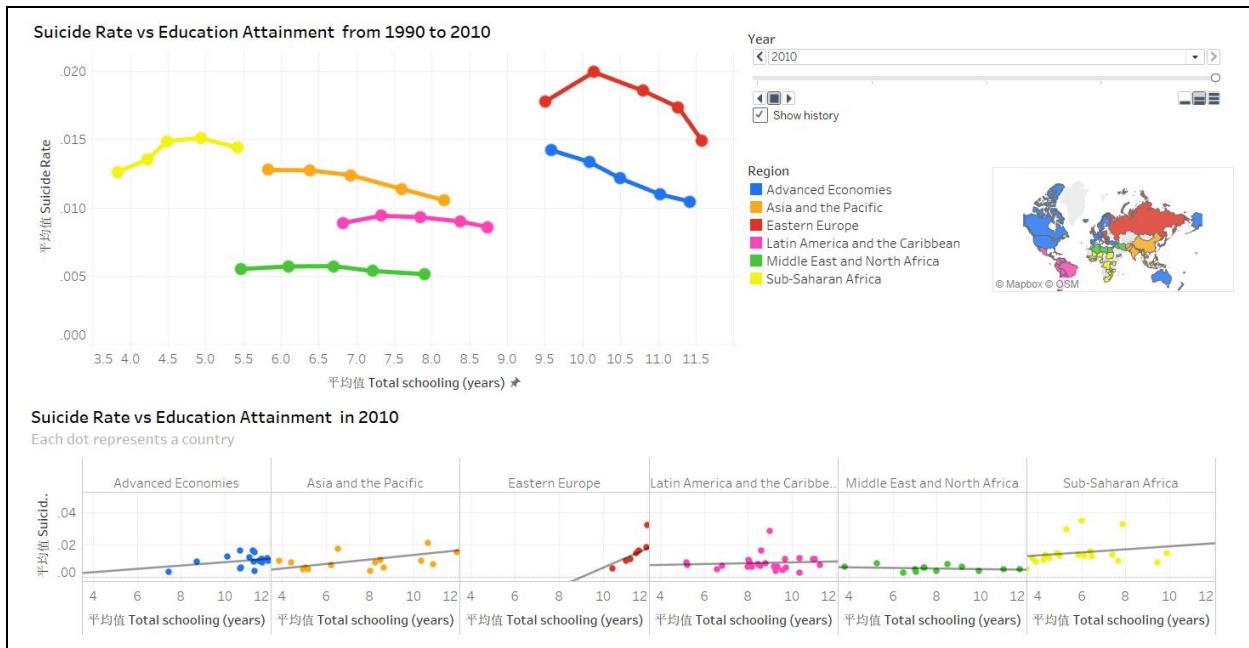
In this graph, two key variables, total schooling years and life expectancy were encoded in two axes of the plot (position), population in size of the bubble (like all bubble/dot charts do), and different regions in color. The change from a static dot plot to an animated bubble graph shows how these two variables change in the scopes of world, regions and countries. It allows users to see the positive correlation from two perspectives. On a time series, it can be seen life expectancy increases with total schooling as all the dots move up and right in the animation. If comparing among countries, it can be seen the countries with higher total schooling also have higher life expectancy. It also allows users to dig into more detailed information in regions and specific countries.



Education Attainment vs Suicide Rate

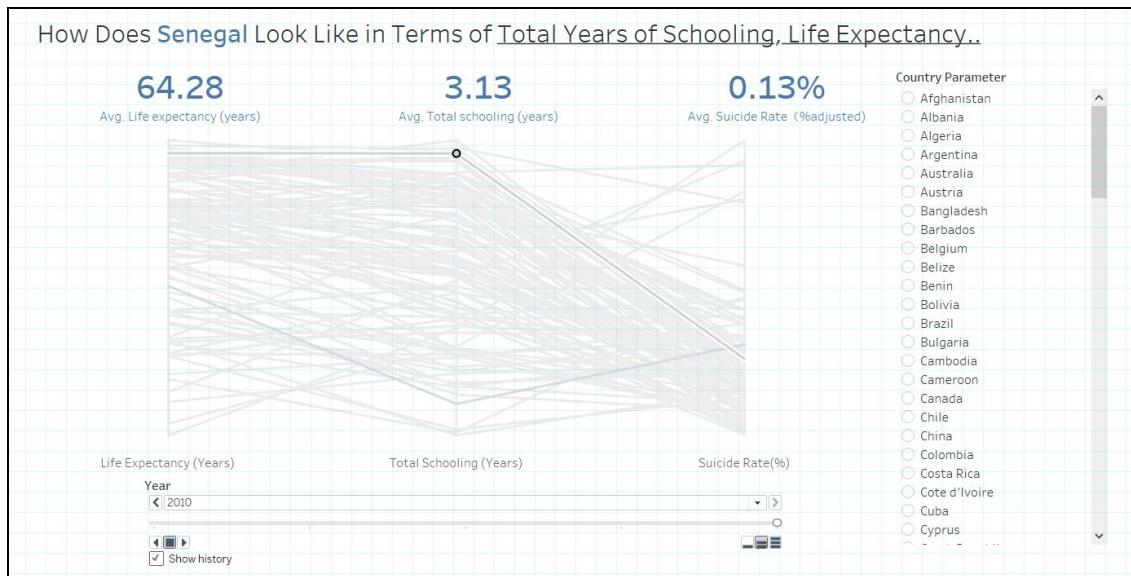
At first we tried the bubble chart and found it's hard to see any stories from it because dots just distributed and moved randomly. So we took the side by side dot chart in the dashboard 1.0 back. The side by side plot chart can show different correlations of these two variables in different regions but a plot chart on a global scale just tells nothing. Then we tried to show the changes of the theses two variables in different regions using trails. On this dashboard, the first trail graph indicates the suicide rate decreases with total schooling except sub-saharan africa. The second graph, however, shows that in most regions, countries with higher eudctaion attainment have higher suicide rate, especially Eastern Europe, the positive correlation is very

obvious.



Education Attainment vs Life Expectancy & Suicide Rate

We still want to have a graph to show the three factors on one graph, after discussing the correlations of education with other two factors separately. Then we chose parallel coordinates plots for this multivariable exploration. Education (Total Schooling) was put in the middle so it's easy to see how it correlates with other two factors put on two sides.



Navigation System

We have 5 dashboards and each one discusses one topic. On the Home page, we have the overview of global education and all the buttons which take you to other dashboards. Once users are on other dashboards, they need to go back to home screen than switch to other dashboards.

Data Visualization Project: Education + ?

Context: HCDE 511 Class Project
Team: Anqi Cao, Shiyao Chen, Amitabh Nag, Susan Zheng

Introduction:
Knowledge is power and education is the key for people to improve social mobility. We want to present the historical data in world education attainment and enrollment across the world. During our project, we also explored the correlation between education attainment and health (physical health and mental health). The data representing education attainment is total schooling in years, the one representing physical health is life expectancy, and the one representing mental health is suicide rate.

[Home](#) [Education & Mental Health](#) [Education & Physical Health](#) [Education & Health](#) [Education & Gender](#)

Education Attainment Across the Globe

Legend:
Average Total Schooling (in years)

Back

Life Expectancy vs Education Attainment in 1960
Each dot represents a country, size represents population.

Year: 1960

Region: Advanced Economies

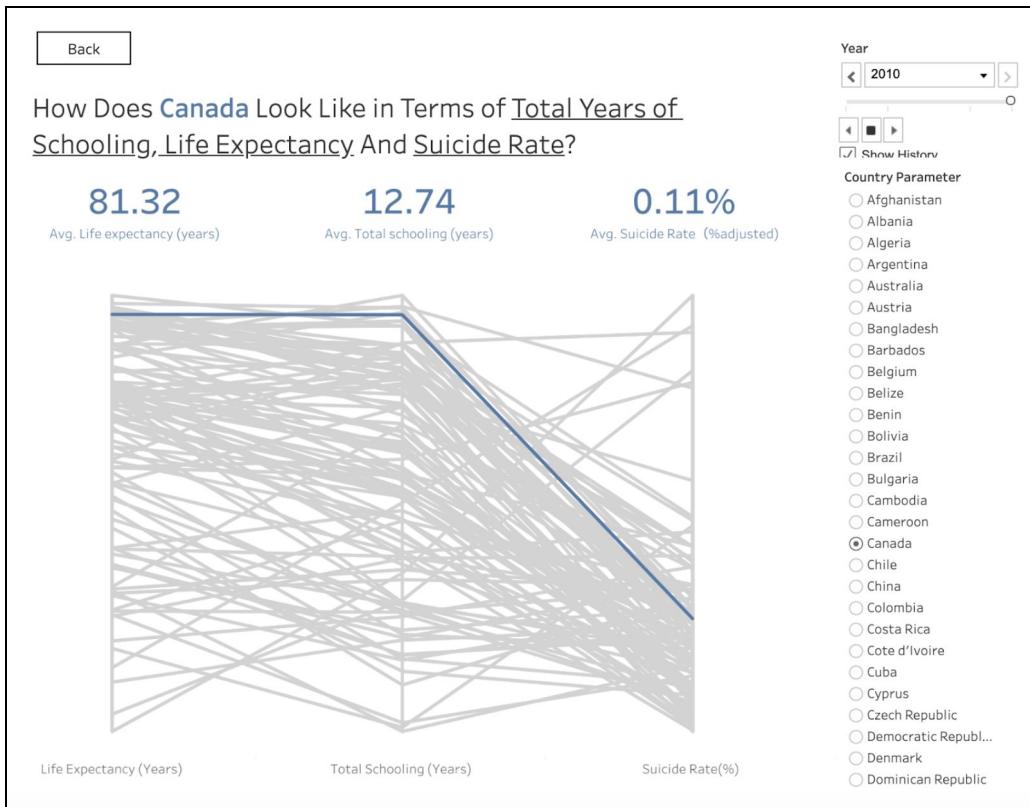
Usability Testing on Dashboard 2.0

The problem with our first prototype (dashboard 1.0) is that it lacked interactivity and story-telling. For our dashboard 2.0, we added interactivity to all the visualizations we created by incorporating timeline. The purpose of our visualization is to tell a compelling story and gives users the option to explore the relationships of education between other variables like life expectancy and suicide rate. We conducted the second round of usability tests on our new Dashboard 2.0 and other tabs. Below are the main findings, which guided our design in the next stage.

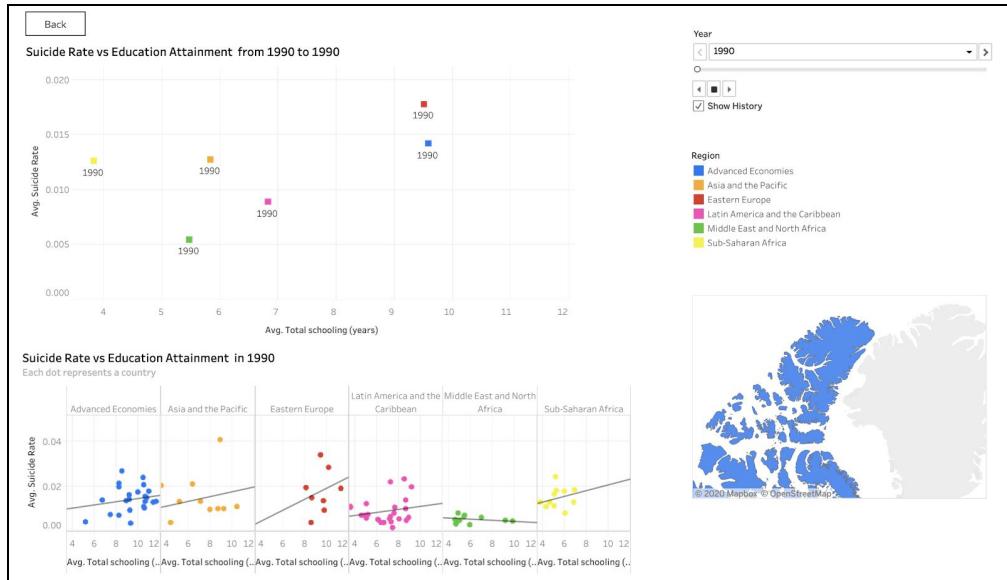
Issues with the timeline control:

- Users did not know how to use the animation feature on the timeline – the “play” button is confusing to users;
- The size of some timelines is so small compared to the corresponding visualizations that users did not even interact with the timeline.
- The size of the timeline is not consistent for all the visualizations.

Examples:



The timeline control is placed too far off to be noticed



Inconsistency of the size of timeline

Recommendation:

- Make the timeline consistent for all the visualizations in terms of size and position in the interface.
- Add annotation to inform users how to use the animation feature on the timeline.

Issues with the navigation menu:

- Naming of the navigation menu is confusing to users – users didn't understand the difference between terms “physical health” and “health”; users also expressed that physical health is a large topic and using the data of life expectancy alone couldn't tell a holistic story.
- Users expressed the concern that it was inconvenient for them to go back and forth on clicking the navigation to see the other visualizations. (Users had to click the “back” button to go back to the main page in order to navigate to another visualization.)
- The position of the navigation caused confusion to users – they expected a change in the content below the navigation, not a redirection to a new page.

Examples:

Data Visualization Project: Education + ?

Context: HCDE 511 Class Project

Team: Anqi Cao, Shiyao Chen, Amitabh Nag, Susan Zheng

Introduction:

Knowledge is power and education is the key for people to improve social mobility. We want to present the historical data in world education attainment and enrollment across the world. During our project, we also explored the correlation between education attainment and health (physical health and mental health). The data representing education attainment is total schooling in years, the one representing physical health is life expectancy, and the one representing mental health is suicide rate.



Education Attainment Across the Globe



Navigation in the main page



Back button on the top left corner of visualizations

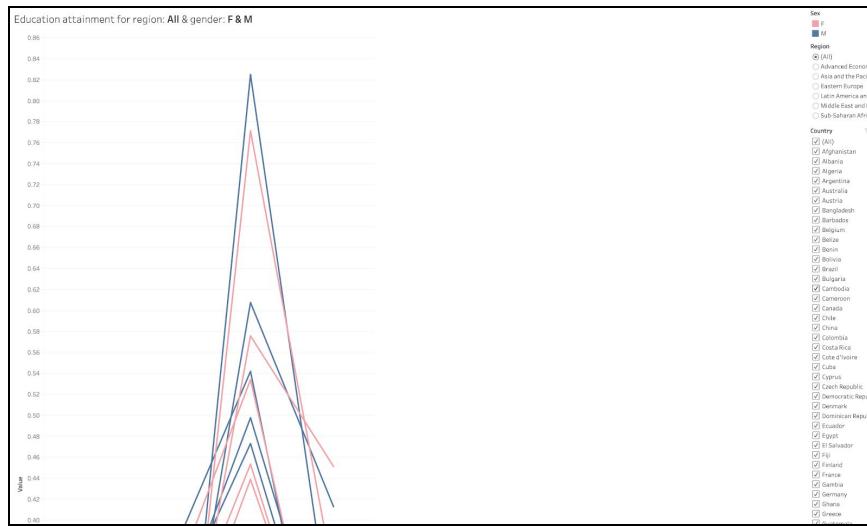
Recommendations:

- Rename the navigation and give user the context regarding the visualizations they are going to see before clicking a button
- Move the navigation to the top and make it consistent across pages

Issues with understanding the Parallel Coordinates Chart:

- The visualization is out of proportion. Users couldn't see the whole picture at once.
- Users were not sure about what the difference in color represents.

Example:



Screenshot of the parallel coordinates chart

Recommendation:

- Give users more contextual information regarding the visualization
- Fit the entire visualization into one screen dimension

Prototyping Stage 3 (Final System Demonstration)

Live dashboard is available here: <http://bit.ly/hcde511eduplus>. Please view the dashboard in presentation mode for best experience.

Based on the findings and recommendations from prototyping stage 3, we made final changes to our design and below are the screenshots of the final design. Below are the five screenshots corresponding to the five buttons on our final dashboard:

tableau public

GALLERY AUTHORS BLOG RESOURCES ACTIVITY SIGN UP SIGN IN

UWEduPlusTeam - Profile

Overview Education Education + Life Expectancy Education + Suicide Rate Education + LE + SR

Data Visualization Project: Education + ?

Context: HCDE 511 Class Project
Team: Anqi Cao, Shiyao Chen, Amitabh Nag, Susan Zheng

Introduction:
The richly available data on global education attainment opens doors for interesting exploration. What are the patterns of different types of education attainment throughout history? How do gender disparities look like? Moreover, to what extent is education correlated with aspects of physical and psychological health? Play around with the tool to find out about answers to these questions!

Education Attainment Across the Globe From 1870 to 2010

Timeline (year): 2010

Note: Click ▶ to play
Click ■ to pause playing

Legend: Average Total Schooling (in years)
0.03 13.24

© 2020 Mapbox © OpenStreetMap

Overview of Education Attainment By Type and Region

Region

Value

Measure Names

- Avg. Tertiary schooling complete (%)
- Avg. Secondary schooling complete (%)
- Avg. Primary schooling complete (%)
- Avg. Total schooling (years)

Year

Overview of Gender Differences in Total Years of Schooling

Avg. Total schooling (years)

Sex

- F
- M

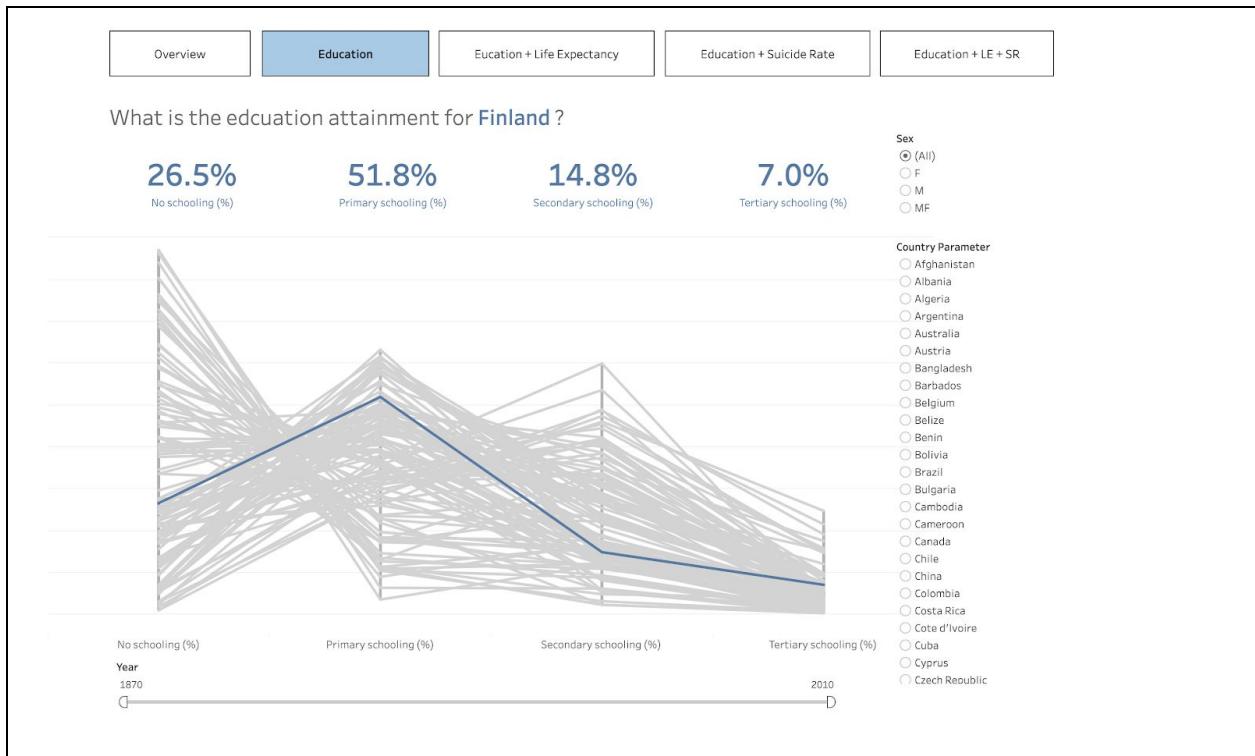
Region	Sex	Avg. Total schooling (years)
Advanced Economic	F	11.45
Advanced Economic	M	11.40
Asia and the Pacific	F	7.76
Asia and the Pacific	M	8.56
Eastern Europe	F	11.61
Eastern Europe	M	11.56
Latin America and the Caribbean	F	8.78
Latin America and the Caribbean	M	8.69
Middle East and North Africa	F	7.61
Middle East and North Africa	M	8.22
Sub-Saharan Africa	F	4.97
Sub-Saharan Africa	M	5.89

Year

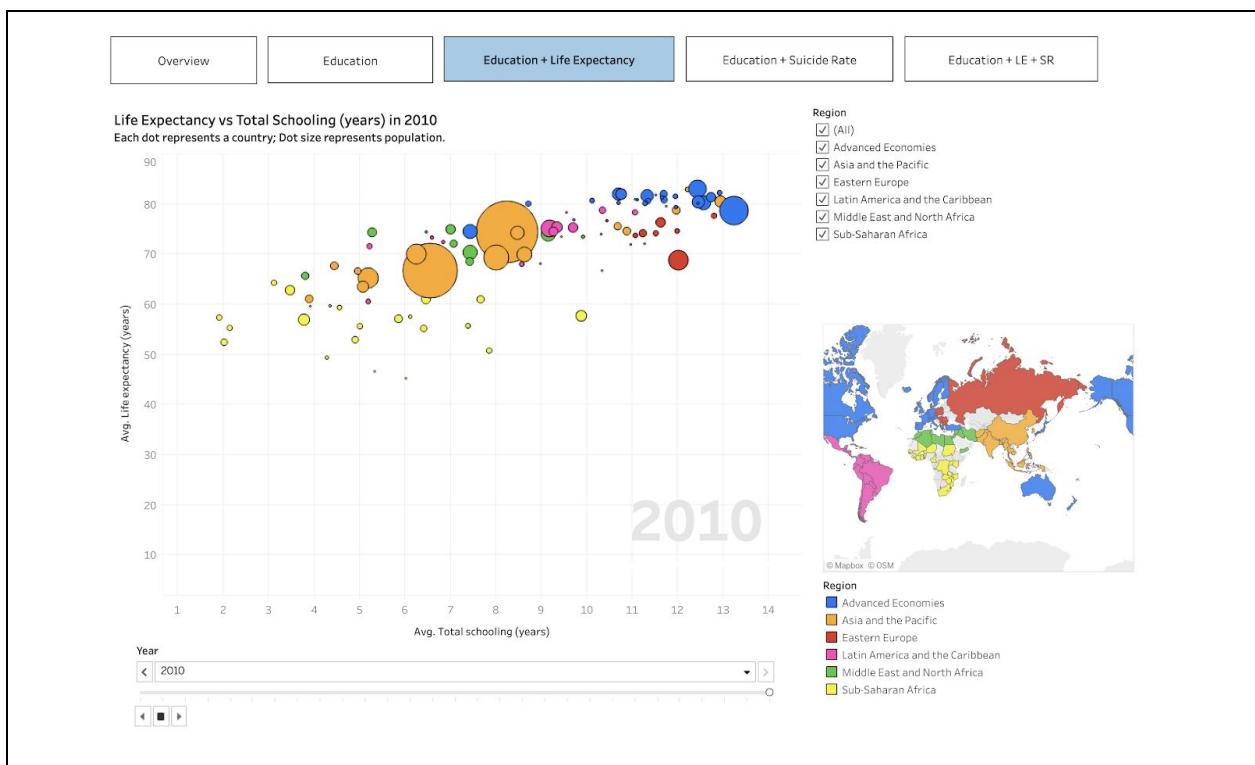
ProjectEduPlus

95 views | UWEduPlusTeam

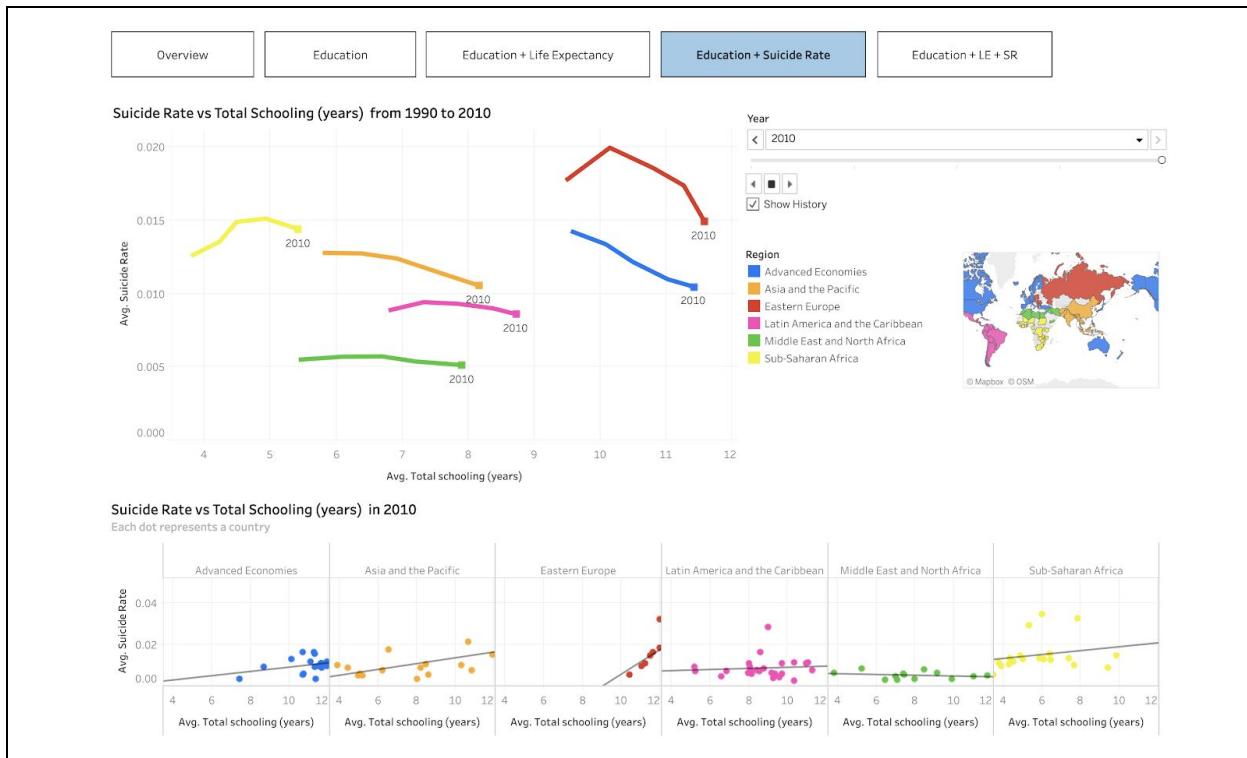
Home page



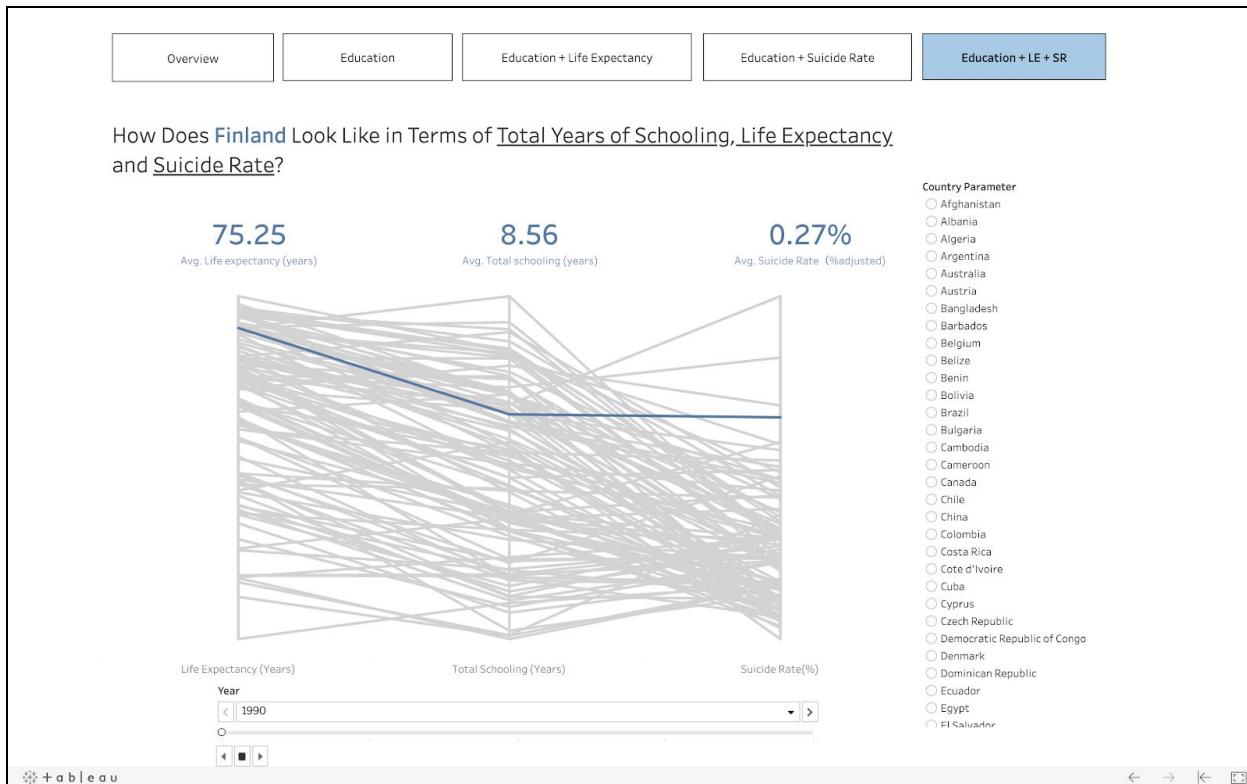
Education attainment by country



Correlation between education attainment and life expectancy



Correlation between education attainment and suicide rate



Correlation between education attainment, life expectancy and suicide rate

Final System Evaluation

We assigned our dashboard star rating out of five based on our perception of how the dashboard follows Tufte's and Shneiderman's visualization principles. Here is the overview of our self-evaluation:

#	Visualization Principle	Self-Evaluation
1	Interesting data	★★★★★
2	Maximization of Data-Ink Ratio	★★★★
3	Accurate communication	★★★★★
4	Accurate presentation of value relationships	★★★★★
5	Clear and careful labeling	★★★★★
6	Contextualization of data	★★★★
7	Shneiderman's principle - overview first; details later	★★★★★

Details of Self-Evaluation:

1. Interesting data (complex, multi-variate): ★★★★★

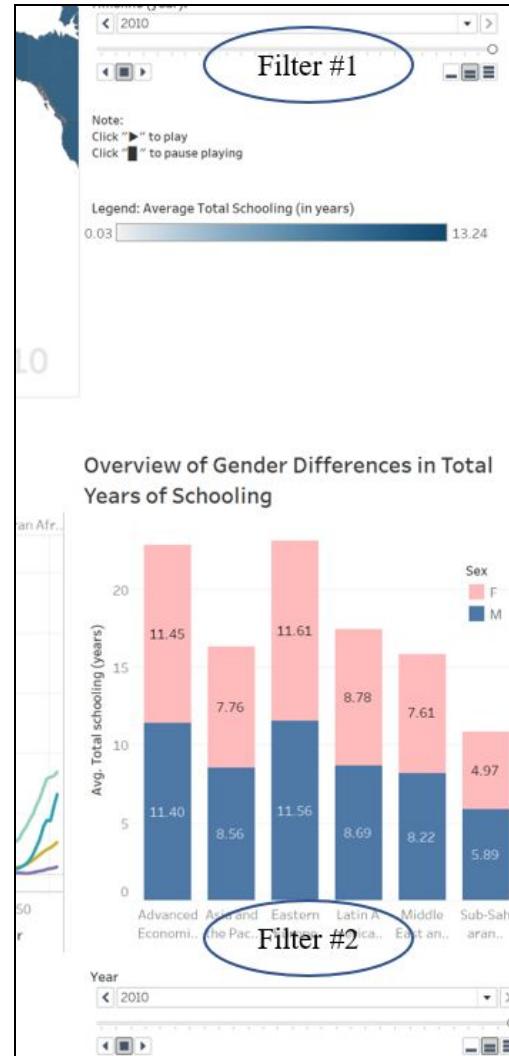
Our dataset is very rich multi-dimensional and has a combination of nominal and quantitative variables. Our dataset was created by merging these three individual datasets:

- Long term education attainment from 1820 to 2010 by sex, country and region; 5-year intervals. Variable types: 20 ratio; 2 nominal; 1 interval
- Life expectancy from 1543 to 2019 by country. Variable types: 2 nominal; 1 interval; 1 ratio
- Number of deaths from suicide per 10,000 from 1950 to 2017 by country. Variable types: 2 nominal; 1 interval; 1 ratio

2. Maximum data-ink ratio: ★★★

We have these redundancies on the dashboard that can be reduced to improve data-ink ratio:

- The overview page has two filters for year, one for each chart. In future, we would like to link them for improved data-ink ratio



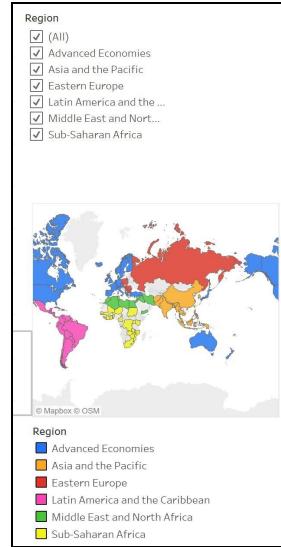
Dashboard overview tab with two year filters

- The “Education” tab on the dashboard has Sex filter with these options: All, M, F and MF. The MF option is redundant and can be removed to just keep All, M, F

A dropdown menu titled "Sex" with four options: "(All)" (selected), "F", "M", and "MF".

Dashboard Education tab with sex filter

- The “Education+Life expectancy” and “Education+Suicide rate” tabs have “Region” repeatedly as a list, map and legend. We could keep only one of these three as a multi-purpose legend and filter



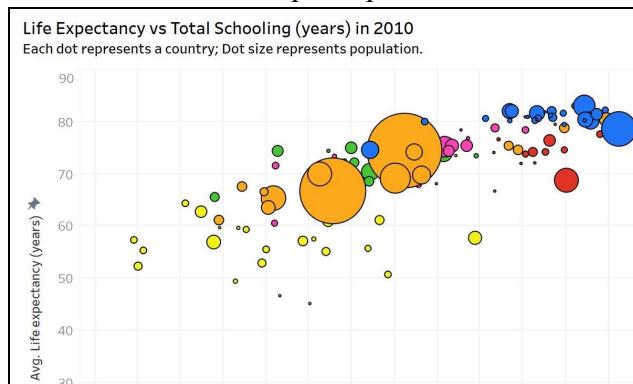
Multiple region filter/legend

3. Accurate communication (Lie factor):

We believe our final dashboard has a lie factor of ~1 as we did not distort sizes of any data. We used Tableau as our visualization tool which decided the visualization sizes based on the data values.

4. Accurate presentation of value relationships (size matches data; avoid area and volume encodings, etc):

In the “Education + Life” expectancy tab, we encoded the country's population as the area of the bubble. We found that it is common practice to encode population as area, hence went ahead with this encoding. We would like to explore other encoding options that are aligned with Tufte’s visualization principle.



Population encoded as bubbles

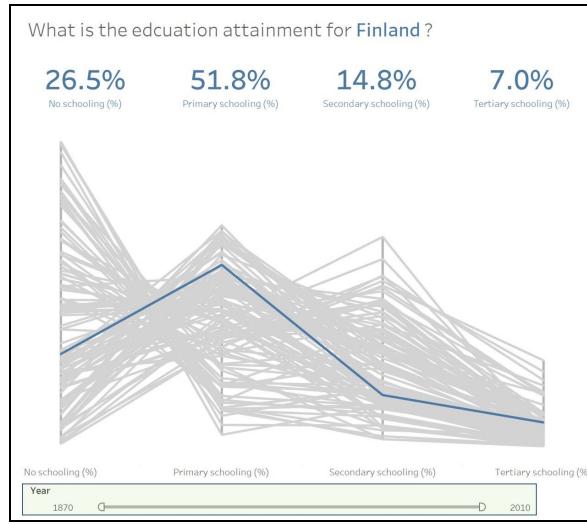
5. Clear and careful labeling:

All the charts on the dashboard provide labeling for the variable encoding used.

6. Contextualization of data:

Our dataset is a time series. However, on the visuals we have aggregated the data and provided a filter for the year. For example, parallel coordinates on “Education” and

“Education + LE + SR ” tabs filter data for a year or a year range. Due to this filtering, the data context in the filtered-out years is missing.



Year as a filter

7. Shneiderman's principle: ★★★★★

Final dashboard has an overview tab, and there are multiple tabs with details. Two tabs with “Education” and “Education + LR + SE ” allow zooming-in on a country. Specific values for a country are displayed as a tooltip and also on a top banner. All detailed tabs provide filters such as year, sex, country and region. Tooltips on visuals provide details on demand as per Sheiderman’s principle.

Future Work

We intend to improve the EduPlus dashboard by making these improvements:

- We used parallel coordinates for exploration of education, life expectancy and suicide rate data. Parallel coordinates pose a challenge in that we are able to encode only one nominal variable: country. However, our dataset has other nominal data such as sex and region that we were unable to encode. As a workaround, we added a filter on the dashboard for sex. In future, we want to explore “parallel sets” to encode nominal data variables such as sex, region, country
- Following changes to improve data-ink ratio:
 - Overview tab of the dashboard uses two filters for a year, one filter per chart. We would like to keep one filter to improve data-ink ratio
 - In Education tab, keep only M, F and MF as sex options
 - In “Education+Life expectancy” and “Education+Suicide rate” tabs, make region as a list, map and legend
- In the “Education+Life expectancy” tab we are using area to encode population. We would like to explore alternate encoding options that are aligned with Tufte’s visualization principle

- For the “overview of gender differences in schooling” chart, it’s better to convert all numbers to percentages on the same scale for a more effective display of gender disparities
- Tableau presented challenges in making the year animation button visually forward. The current play control looks disconnected from the chart and makes it difficult for the readers to notice or understand it. We would like to use d3 to get more control on the visualization animation

References

- Lee, Jong-Wha. “Barro-Lee.Com.” Www.Barrolee.Com, www.barrolee.com/. Accessed 15 Mar. 2020.
- Lee, Lindsay, et al. “Suicide.” Our World in Data, 2015, ourworldindata.org/suicide.
- Roser, Max. “Life Expectancy.” Our World in Data, 2013, ourworldindata.org/life-expectancy.
- Assari, Shervin. “Blacks’ Diminished Return of Education Attainment on Subjective Health; Mediating Effect of Income.” *Brain Sciences*, vol. 8, no. 9, 1 Sept. 2018, p. 176, www.mdpi.com/2076-3425/8/9/176, 10.3390/brainsci8090176. Accessed 17 Mar. 2020.
- Bulled, Nicola L., and Richard Sosis. “Examining the Relationship between Life Expectancy, Reproduction, and Educational Attainment.” *Human Nature*, vol. 21, no. 3, Oct. 2010, pp. 269–289, 10.1007/s12110-010-9092-2. Accessed 10 Jan. 2020.
- dpicampaigns. “About the Sustainable Development Goals - United Nations Sustainable Development.” United Nations Sustainable Development, 2018, www.un.org/sustainabledevelopment/sustainable-development-goals.
- Gastineau, Dan. “The Five Types of People Who Use Visualization.” Medium, 6 Sept. 2019, medium.com/nightingale/the-five-types-of-people-who-use-visualization-733d3998f6ef. Accessed 17 Mar. 2020.
- Kriebel, Andy. “How to Create a Parallel Coordinates Plot.” YouTube, 9 July 2019, www.youtube.com/watch?time_continue=527&v=ZjVK9Dzzk2E . Accessed 15 Mar. 2020.
- Vilorio, Dennis. “Education Matters : Career Outlook: U.S. Bureau of Labor Statistics.” Bls.Gov, 29 Mar. 2016, www.bls.gov/careeroutlook/2016/data-on-display/education-matters.htm.