

# Project Edu Plus

Group #5

Amitabh Nag

Anqi Cao

Shiyao Chen

Susan Zheng

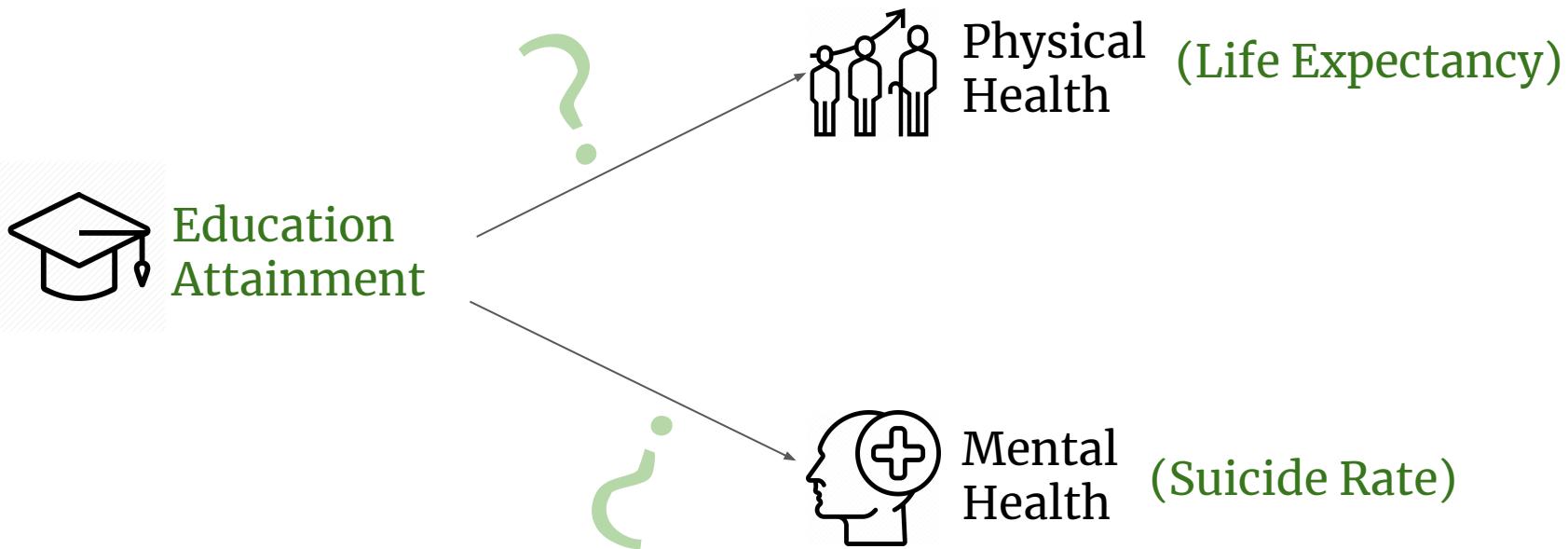
*March 2020*



# Agenda

- Project Motivation
- Project Goals
- Target Users
- Data Overview
- Design Process
- Final System Demonstration
- Final System Evaluation & Future Improvements
- Appendix

# Project Motivation



# Project Goals

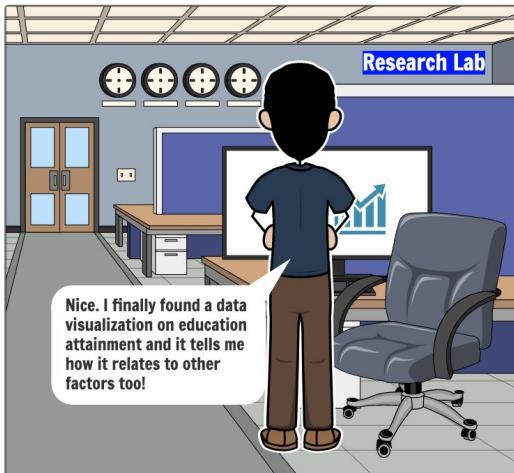
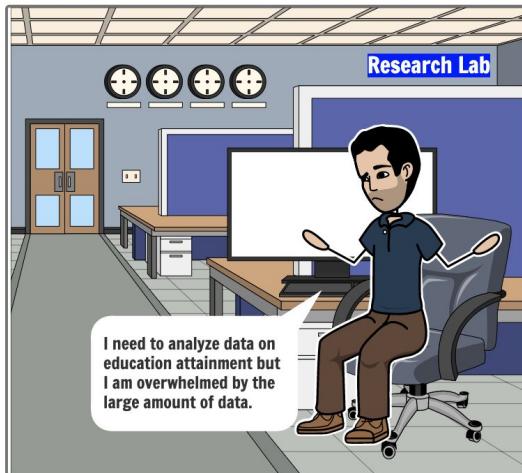
Integrating different visualizations, we aspire to explore, analyze and display answers to the following questions:

- How do regions in the world compare to each other in terms of total years of schooling, primary schooling, secondary schooling and tertiary schooling? What about gender differences?
- Is there a correlation between average life expectancy and education attainment?
- Is there a correlation between suicide rate and education attainment?
- How do we enable exploration at different levels of granularities that are visually aligned with each other?

# Target Users

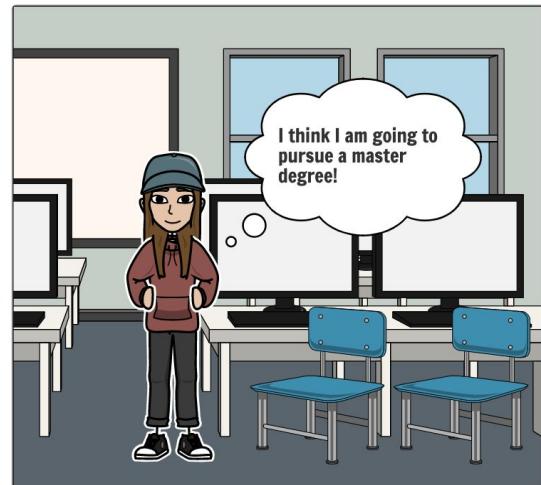
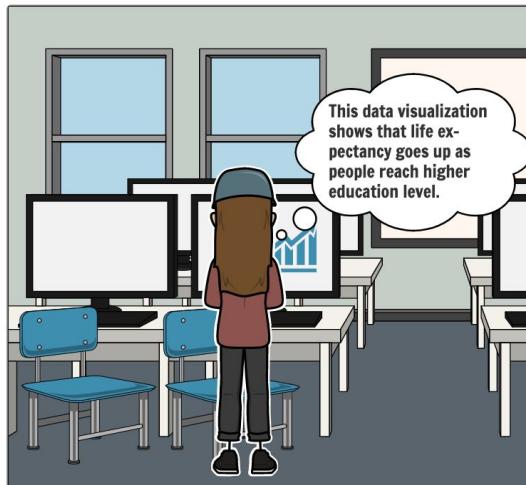
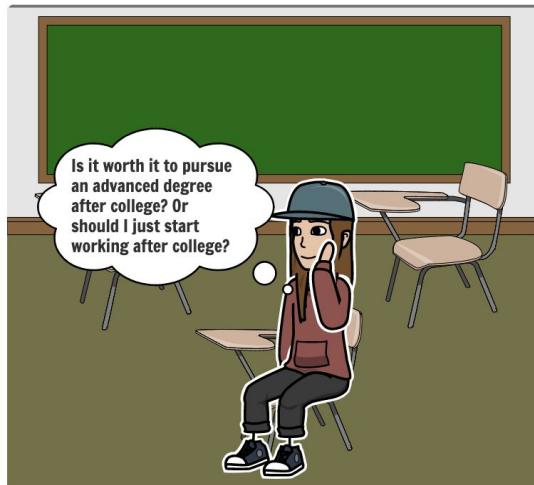
# Target users: The Analyzers

Working professionals and researchers who need to read and use data visualizations to explore details and identify patterns, and decide what stories they can deliver as a part of their work projects



# Target users: The Curious Readers

People whose background, personal interests and concerns draw them to the topic of global education development and its correlation with health. This group of people tend to consist of students, journalists, NGO workers and anyone with a strong curiosity to learn and use the data to inform their decisions/discussions



# Data Overview

# Datasets

## Primary: Long-term education attainment

- Education attainment from 1820 to 2010 by sex, country and region; 5-year intervals
- 12987 rows x 23 columns
- Variable types: 20 ratio; 2 nominal; 1 interval
- <http://www.barrolee.com/>

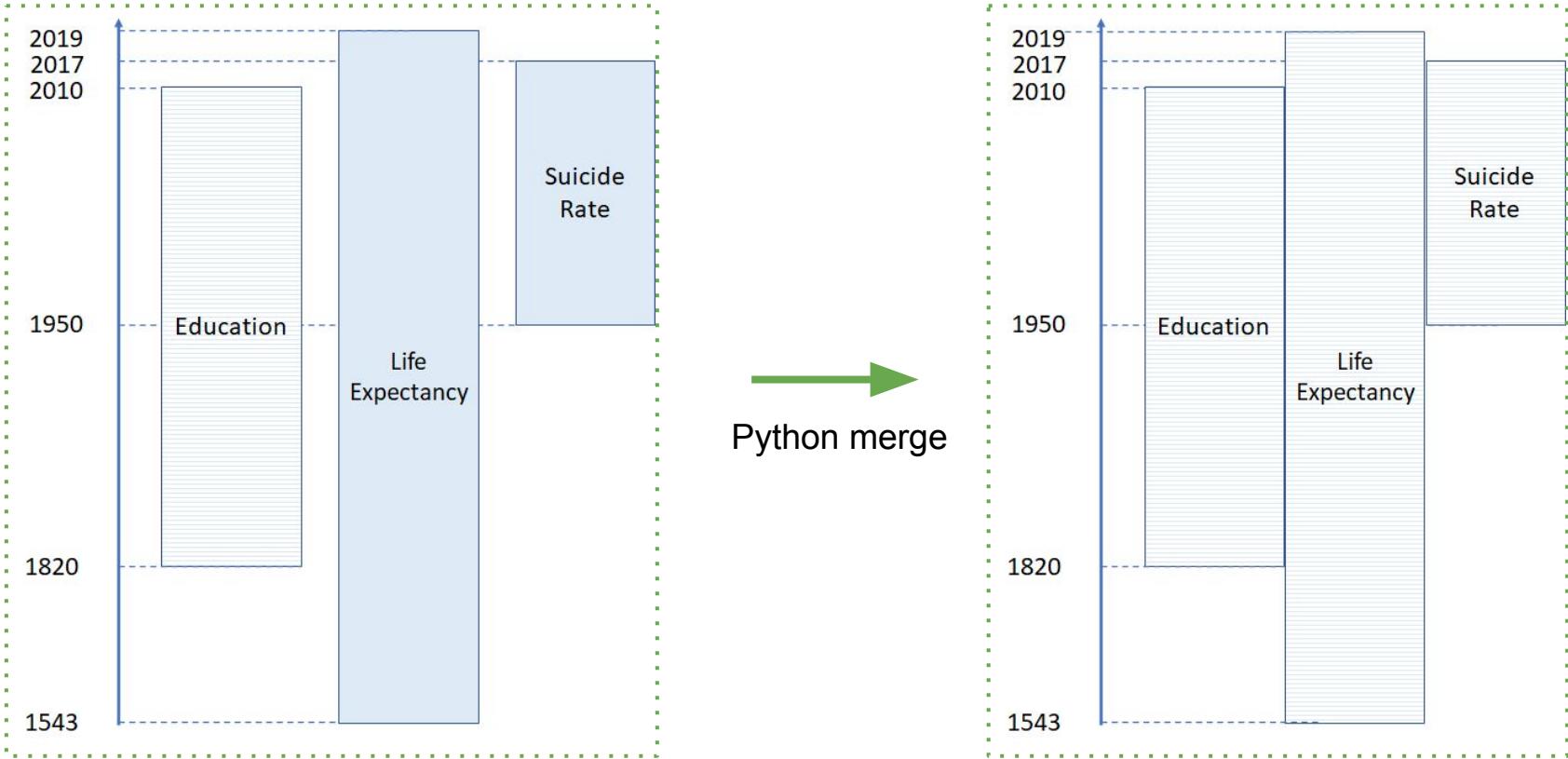
## Secondary: Life expectancy

- Life expectancy from 1543 to 2019 by country
- 19028 rows x 4 columns
- Variable types: 2 nominal; 1 interval; 1 ratio
- <https://ourworldindata.org/life-expectancy>

## Secondary: Suicide rate

- Number of deaths from suicide per 10,000 from 1950 to 2017 by country
- 6469 rows x 4 columns
- Variable types: 2 nominal; 1 interval; 1 ratio
- <https://ourworldindata.org/suicide>

# Data cleanup – Phase 1 (python)



# Data Cleanup – Phase 1 (python)

After some analysis, found that below countries in suicide/life-expectancy are similar, but different than barro lee

'Congo, D.R.', 'Cote d'Ivoire', 'Dominican Rep.', 'Hong Kong, China', 'Republic of Korea', 'Reunion', 'Russian Federation', 'USA'

Renaming these in barrolee data

```
In [68]: #Country cleanup in barrolee
barro_lee_data.loc[(barro_lee_data.country == 'Congo, D.R.'), 'country']= 'Democratic Republic of Congo'
barro_lee_data.loc[(barro_lee_data.country == 'Cote d'Ivoire'), 'country']= 'Cote d'Ivoire'
barro_lee_data.loc[(barro_lee_data.country == 'Dominican Rep.'), 'country']= 'Dominican Republic'
barro_lee_data.loc[(barro_lee_data.country == 'Hong Kong, China'), 'country']= 'Hong Kong'
barro_lee_data.loc[(barro_lee_data.country == 'Republic of Korea'), 'country']= 'South Korea'
#assumption based on wikipedia article: https://en.wikipedia.org/wiki/South_Korea
#"Reunion" and "Hong Kong" is not present at all in suicide_death_rate_countries
barro_lee_data.loc[(barro_lee_data.country == 'Russian Federation'), 'country']= 'Russia'
barro_lee_data.loc[(barro_lee_data.country == 'USA'), 'country']= 'United States'
```

**Renaming a very long column name in suicide data to a simple name**

```
In [69]: suicide_death_rate_data.rename(columns = {'Deaths - Self-harm - Sex: Both - Age: Age-standardized (Rate) (deaths per 100,000 indi
```

**First merge barro lee with life expectancy and then merge this data with suicide rate**

Drop extra column that are generated by the merge like year, country name(Entity) and country code

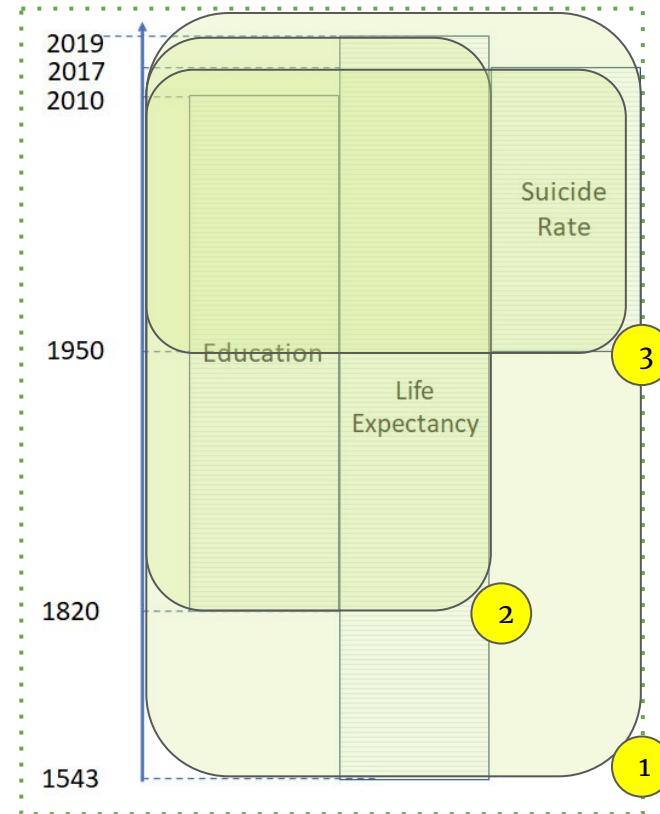
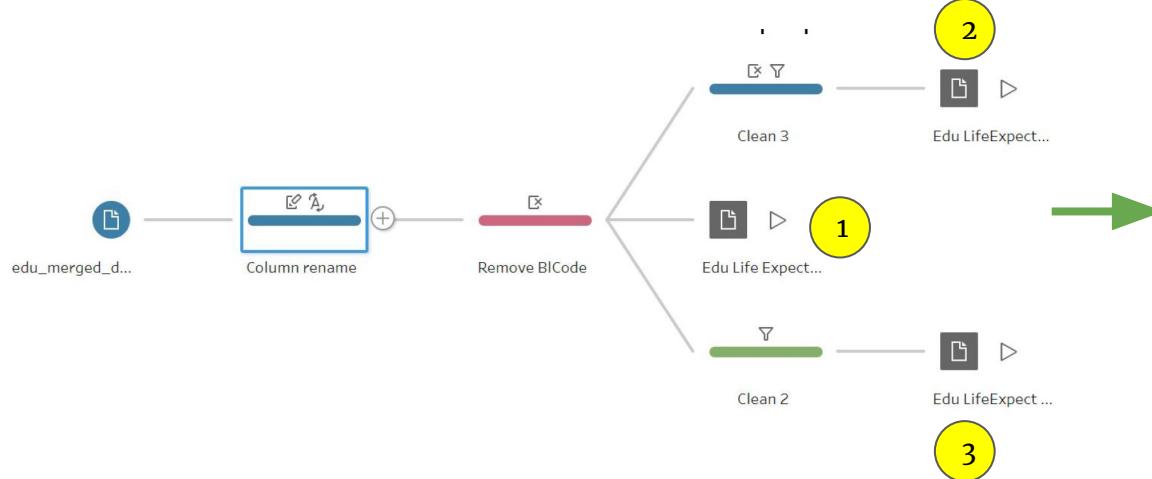
```
In [70]: merged_data = barro_lee_data.merge(life_expectancy_data, left_on=['country', 'year'], right_on=['Entity', 'Year'], how = 'left').reset_index()
merged_data.drop(['Year', 'Entity', 'Code'], axis=1, inplace=True)
merged_data = merged_data.merge(suicide_death_rate_data, left_on=['country', 'year'], right_on=['Entity', 'Year'], how = 'left').reset_index()
merged_data.drop(['Year', 'Entity', 'Code'], axis=1, inplace=True)
```

# Data Cleanup - Phase 2 (Tableau prep)

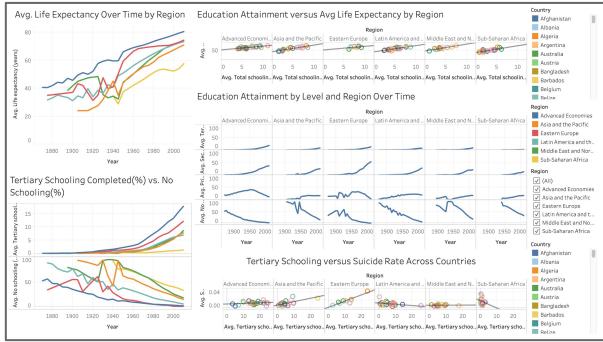
Output from Phase 1 was fed into Tableau prep workflow to generate the three datasets:

- 1) Edu\_suicide\_rate\_life\_expectancy\_nulls (maximum)
- 2) Edu\_life\_expectancy\_no\_null (medium)
- 3) Edu\_suicide\_rate\_life\_expectancy\_no\_null (small)

We used datasets 2) and 3) for our interactive dashboards



# Design Process



## Main Dashboard 1.0

- A collage of individual exploration pieces



First round of usability testing & class feedback & peer review



## Main Dashboard 2.0

- Finalized the topics to cover
- Tried the graphs for each topic
- Made a navigation system for the dashboard



## Dashboard 3.0

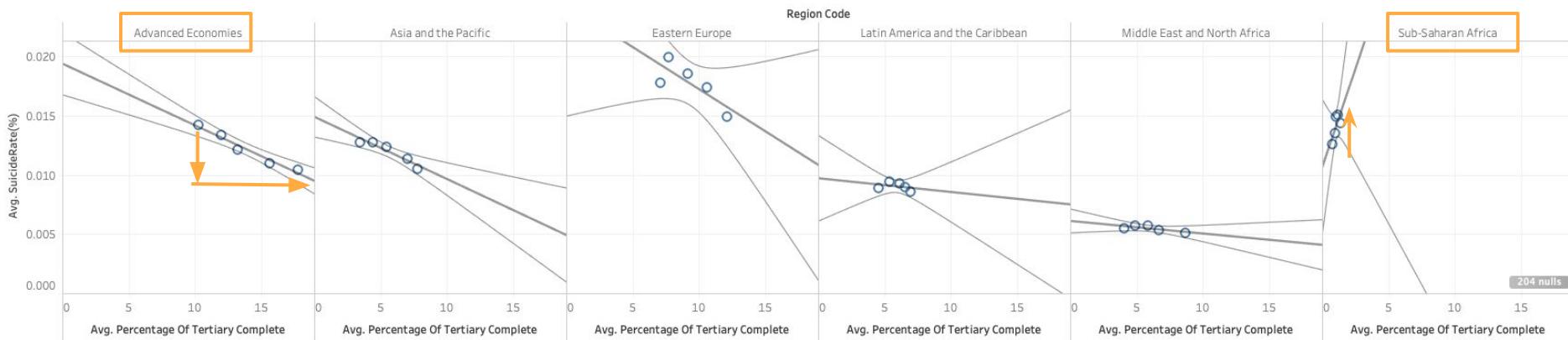
- Refined the navigation
- Refined the graph details



Second round of usability testing

# Stage #1: Synthesis of Insights From Individual Data Exploration

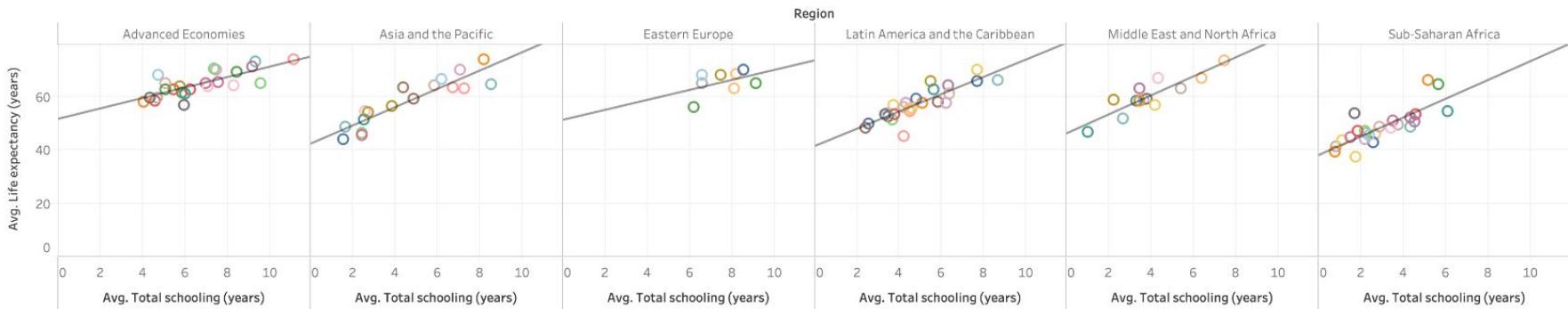
Insight #1: From 1990 to 2010, Advanced Economies experienced the most sharp increase in percentage of tertiary education degrees, whereas Sub-Saharan area had a barely visible amount of growth for the same variable, with suicide rate increased a lot relatively.



# Stage #1: Synthesis of Insights From Individual Data Exploration

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

Education Attainment versus Avg Life Expectancy by Region



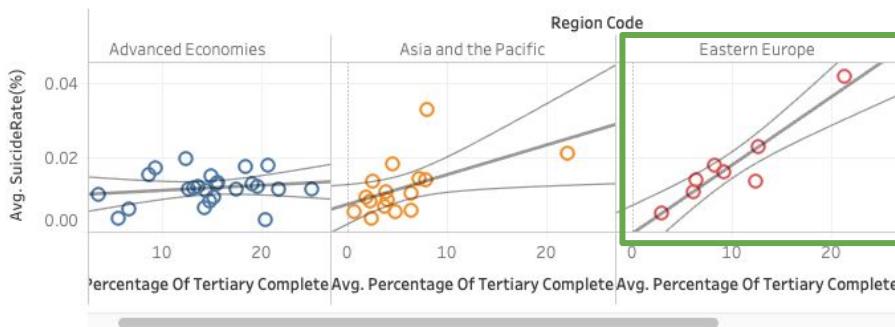
# Stage #1: Synthesis of Insights From Individual Data Exploration

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**

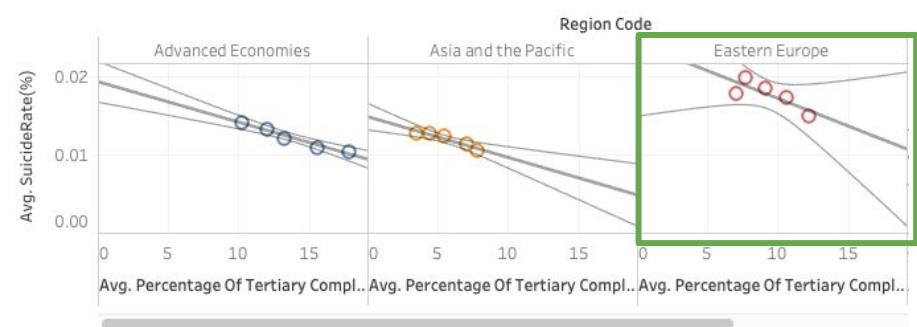


# Stage #1: Synthesis of Insights From Individual Data Exploration

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.



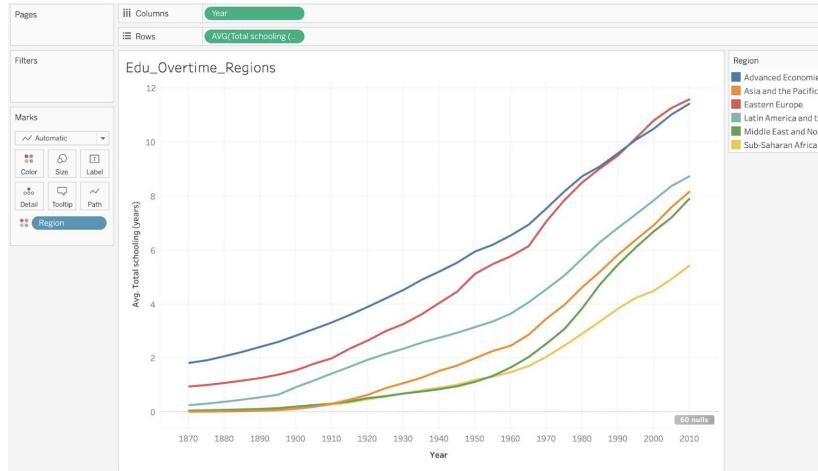
Graph #1: Each dot represents an individual country of the region



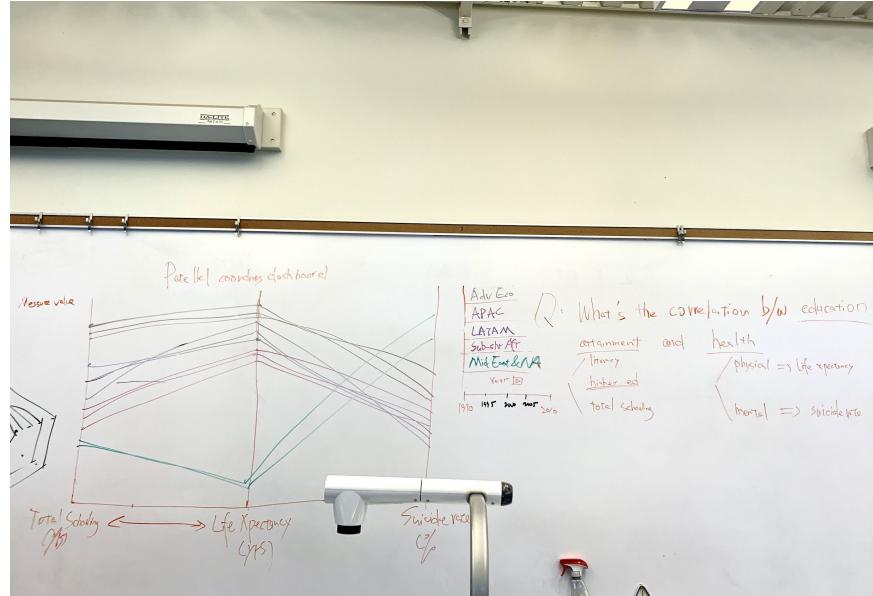
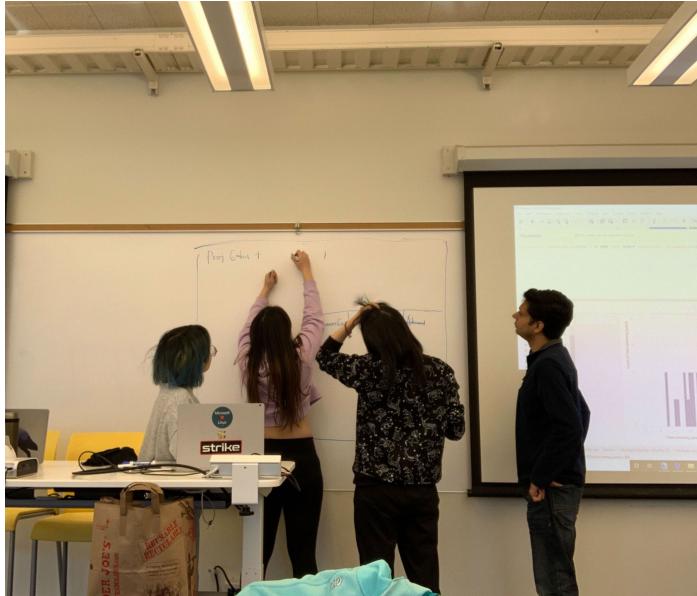
Graph #2: Each dot represents a year (from 1990 to 2010, in 5-year intervals)

# Stage #1: Synthesis of Insights From Individual Data Exploration

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) <Susan>



# Stage #2: Group Brainstorming & Sketching



# Stage #2: Group Brainstorming & Sketching

Challenges	Mitigations
Difficult to create concise charts with multiple education, life expectancy and suicide rate variables	Used parallel coordinates for multivariate analysis
Unable to decide if the dashboard's design be focussed on story-telling or data exploration	Based on user feedback, we designed dashboard with a data exploration focus
Data encoding challenges in encoding nominal variables such as region, sex into the parallel coordinates	We used sex as a filter. In future we will explore using parallel sets for nominal data such as sex, region
We had visualizations about education, suicide rate and life expectancy. Fitting them all on one dashboard was challenging	We liked the idea of “tabs” and created a homepage for the dashboard with links to various tabs for detailed information
Collaborating on a single dashboard is challenging as Tableau does not allow multi-person edits of the dashboard.	Towards the end of the project, we discovered that Tableau supports export and import of sheets. We used that approach to integrate the work

# Usability Testing 1.0

## User #1

**Name:** Jackie

**Gender:** Male

**Age:** 25

**Occupation:** PhD Candidate in Economics



## Experience with Data Visualization:

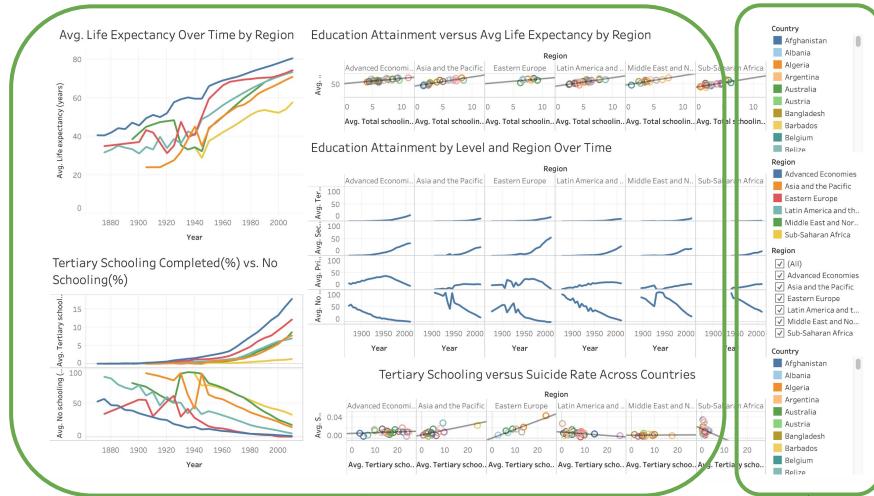
- Intermediate-high level of proficiency in data visualization tools; have worked on data viz projects in the past for his internship

## Knowledge of Education, Physical and Mental Health:

- General interest in learning more about education development
- Armchair psychologist

# Changes Made Based on Feedback

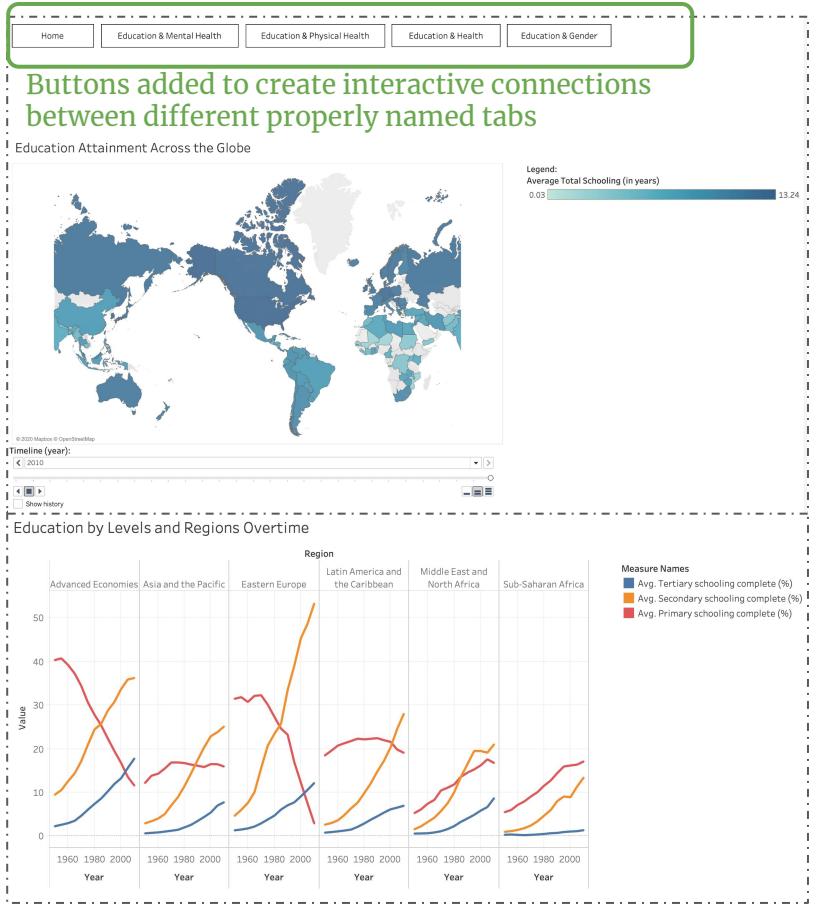
## First Version of Dashboard:



How are the charts connected to each other in a coherent way?  
The small multiples view is overwhelming.

More interactivity is needed -> filter / drop-down menus

## Iterated Version of Dashboard:



A leaner main dashboard

# Usability Testing 2.0

## User #2

Name: Raza  
Gender: Male  
Age: 45  
Occupation: Principal Software Engineer



### Experience with Data Visualization:

- Experience in building dashboards for KPIs
- High level of proficiency in data visualization

## User #3

Name: Daniel  
Gender: Male  
Age: 24  
Occupation: Math Teacher



### Experience with Data Visualization:

- Education background in Business Finance
- Intermediate level of proficiency in data visualization

## User #4

Name: Jason  
Gender: Male  
Age: 22  
Occupation: Undergraduate Student



### Experience with Data Visualization:

- Understands basic data visualization
- Low level of proficiency in data visualization

# Changes Made Based on Usability Test 2.0 Results

Dashboard Navigation: reworded navigation menu

Home

Education & Mental Health

Education & Physical Health

Education & Health

Education & Gender

Overview

Education

Education + Life Expectancy

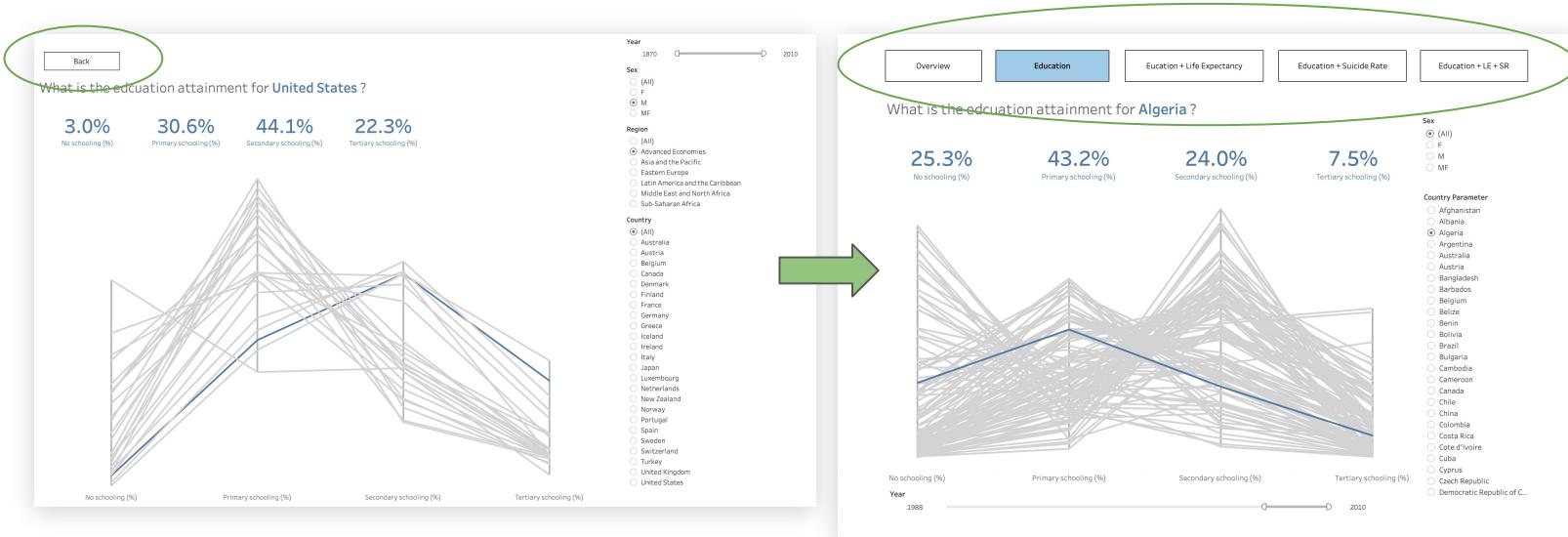
Education + Suicide Rate

Education + LE + SR



# Changes Made Based on Usability Test 2.0 Results

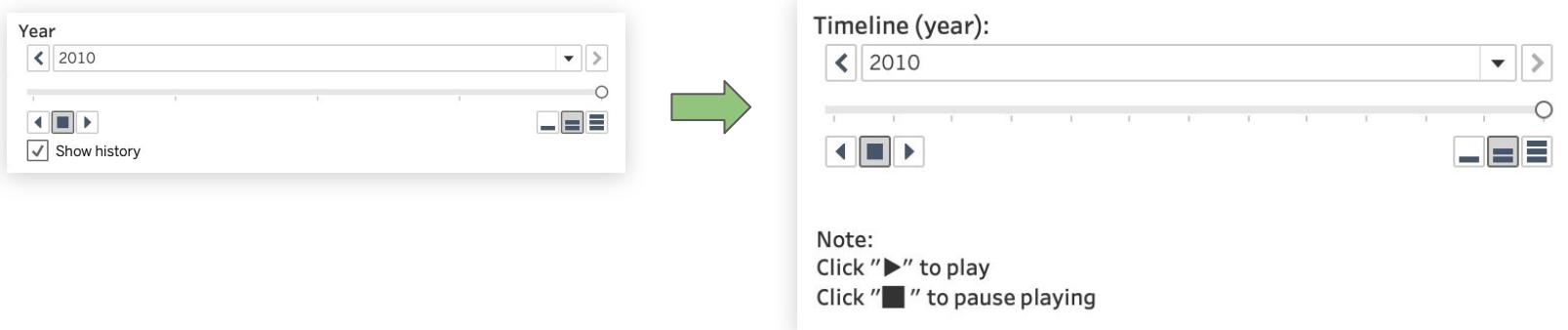
Dashboard Navigation: improved navigation among dashboards



# Changes Made Based on Usability Test 2.0 Results

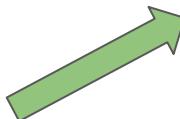
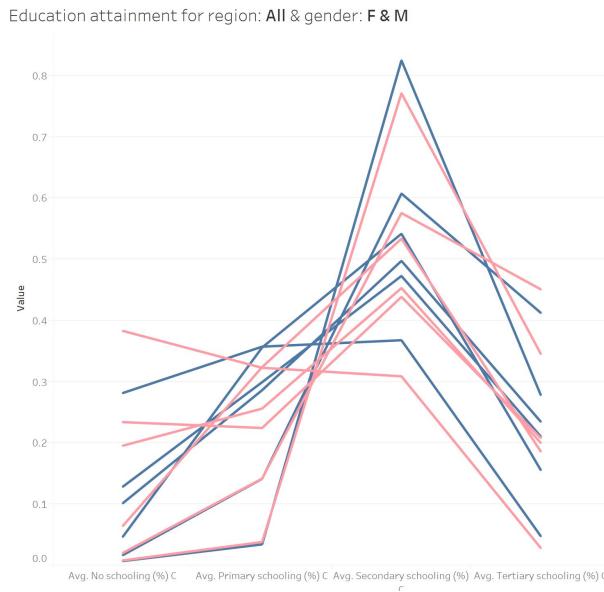
Timeline: users were not sure how to interact with the timeline.

Added hints to users about how to interact with the timeline.



# Changes Made Based on Usability Test 2.0 Results

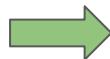
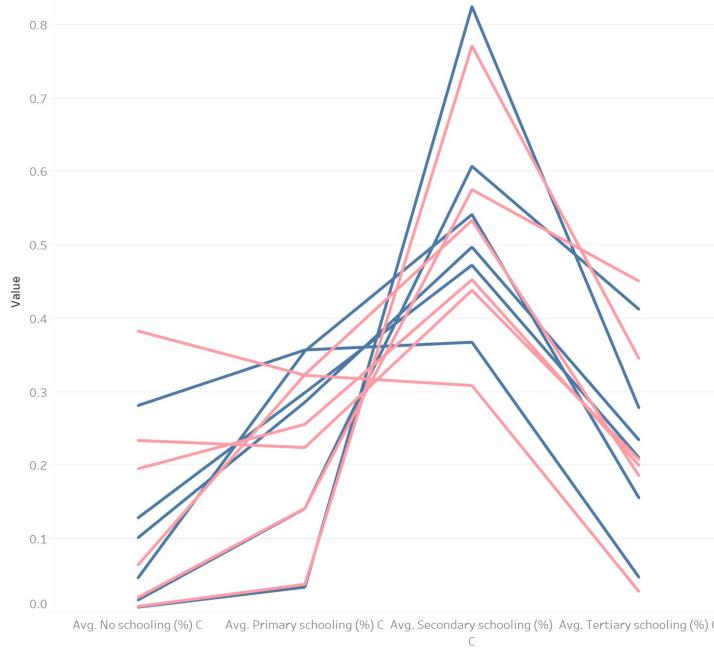
- Keep charts for one variable data exploration - we improvised parallel coordinate for education data exploration and removed sex from the chart instead making it as a filter



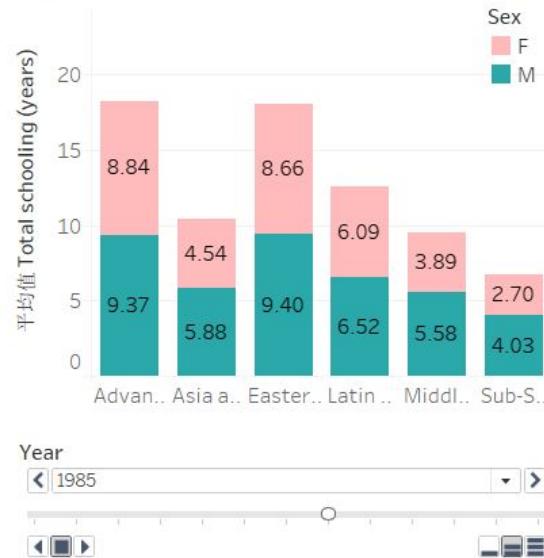
# Changes Made Based on Usability Test 2.0 Results

## Education Attainment- Gender Difference

Education attainment for region: All & gender: F & M

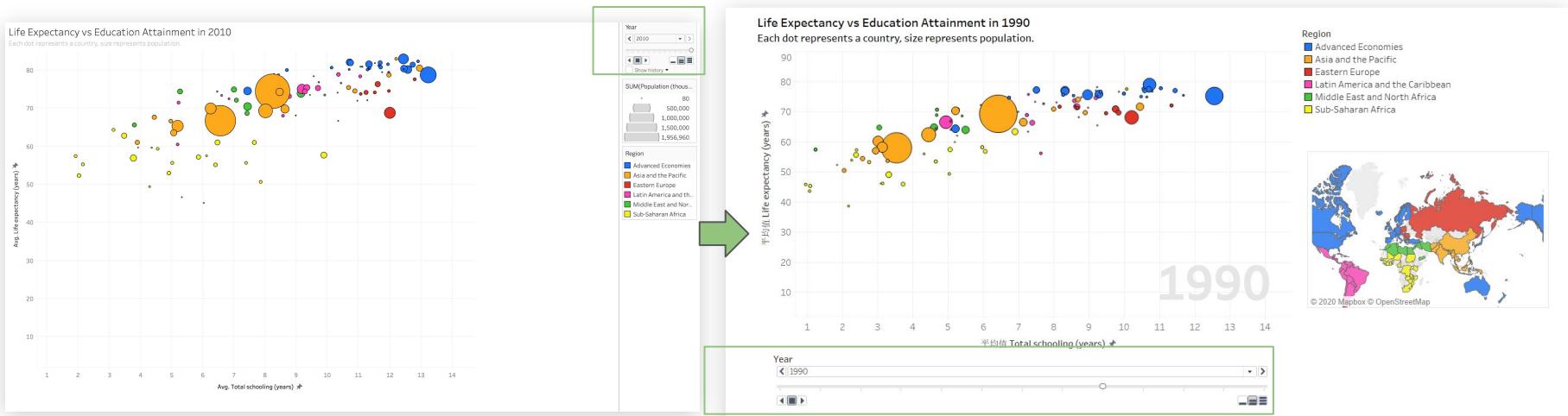


## Education Attainment - Gender Difference - 1985



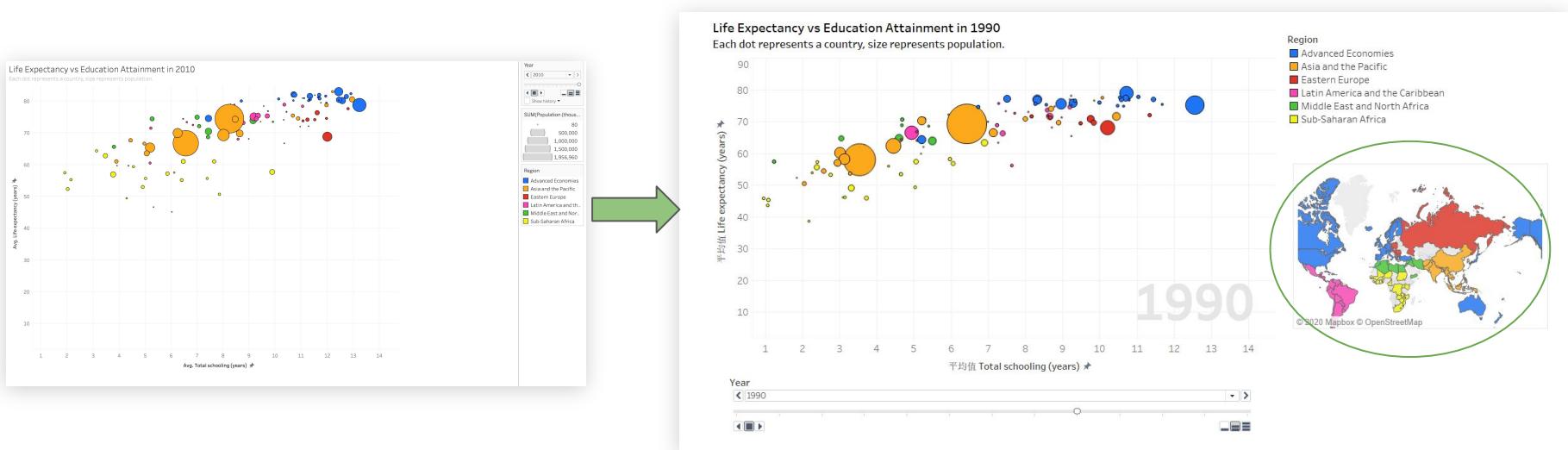
# Changes Made Based on Usability Test 2.0 Results

## Education & Life Expectancy (After)



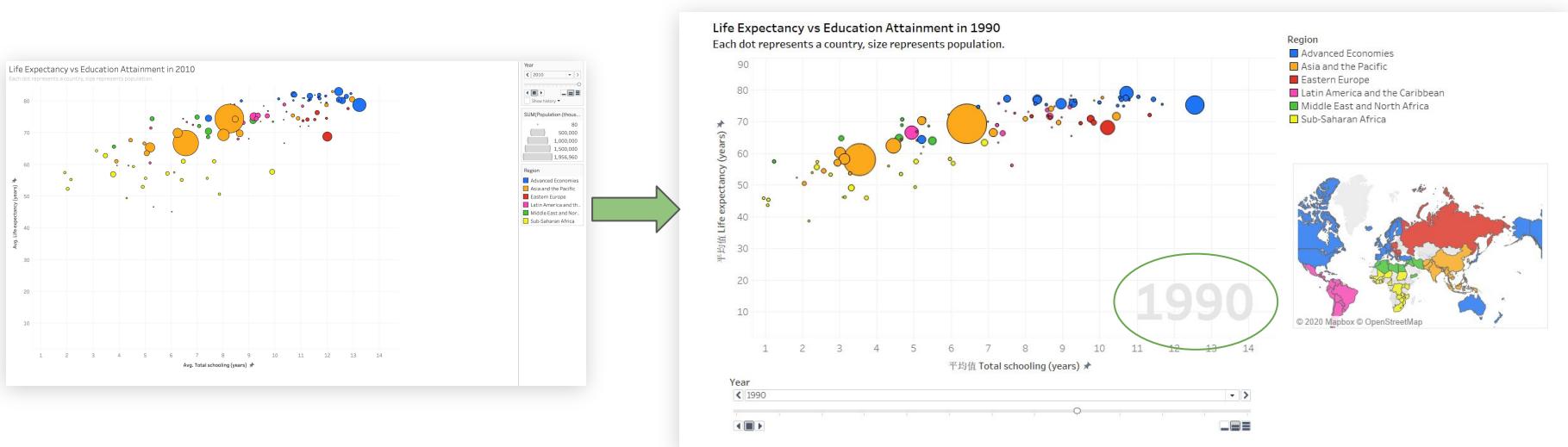
# Changes Made Based on Usability Test 2.0 Results

## Education & Life Expectancy (After)



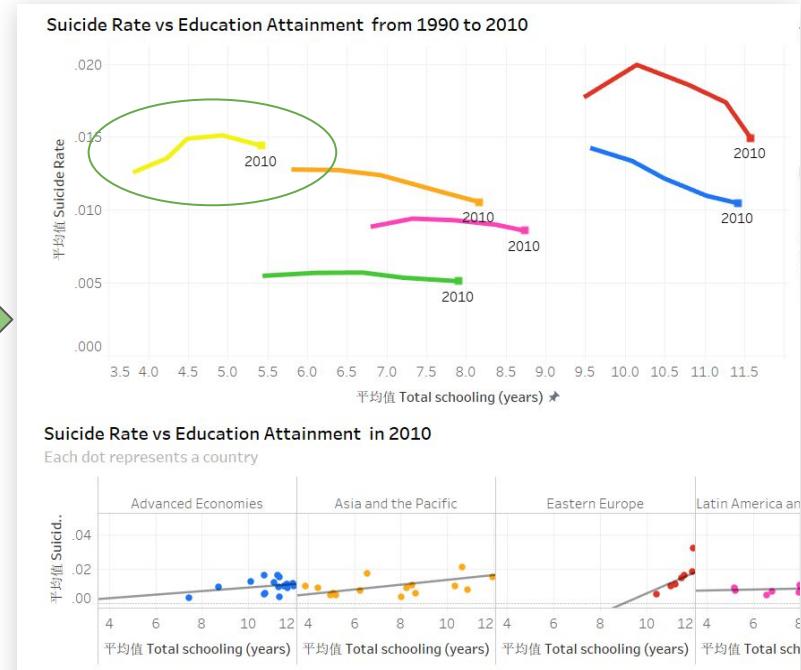
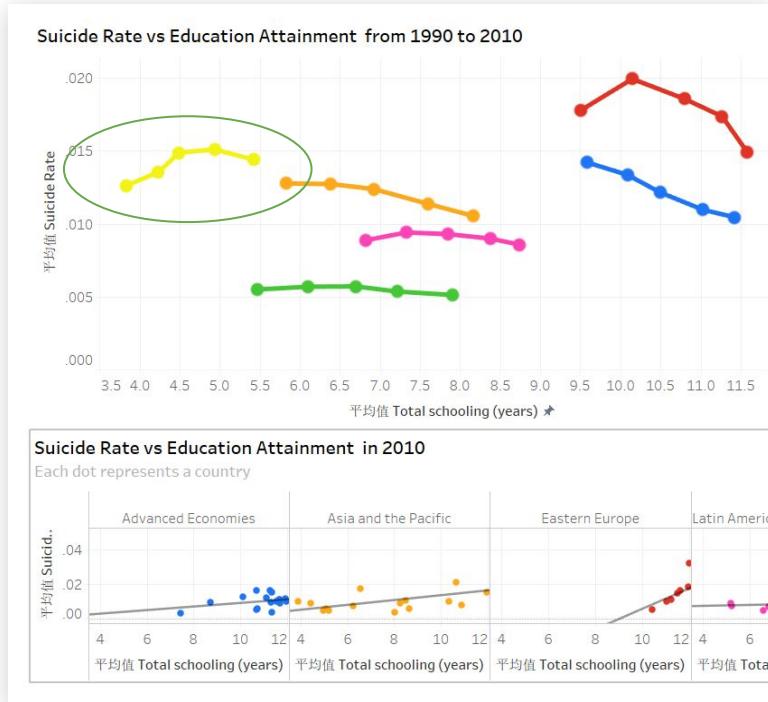
# Changes Made Based on Usability Test 2.0 Results

## Education & Life Expectancy (After)



# Changes Made Based on Usability Test 2.0 Results

## Education & Suicide Rate



# Final System Demonstration

# Final System Evaluation

# Evaluation Rubric

Visualization Principle	Self Evaluation
Interesting data (Complex, multivariate)	
Maximization of Data-Ink Ratio (erase non-data ink & redundant data ink)	
Accurate communication (lie factor)	
Accurate presentation of value relationships (size matches data; avoid area and volume encodings, etc)	
Clear and careful labeling (labels should indicate important events, define units and context, correct for distortion, provide precise values)	
Contextualization of data	
Shneiderman's mantra: Overview first, zoom and filter, then details on demand	

*Future improvements section include areas where we can improve the dashboard*

# Future improvements

- Explore using “parallel sets” to encode nominal education data variables such as sex, region, country
- Dashboard uses a different filter for each chart for filters for same variable (say year). We will like to link filters
- Tableau presented challenges in creating year animation. We wanted a play button that allows year animation, but Tableau did not support this. We will like to use d3 to get more control on visualization animation
- 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.

Thank you!

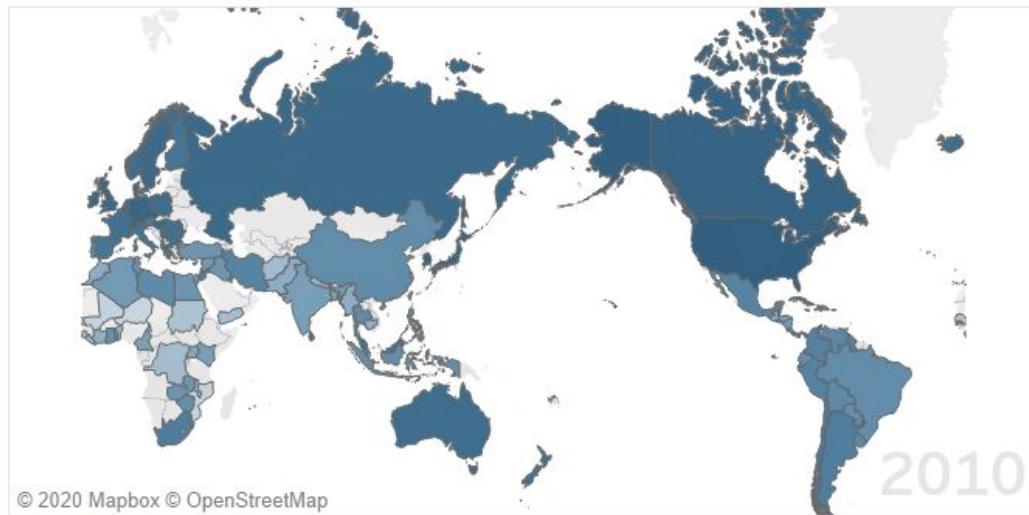
# Appendix

# Finalizing topics to cover

- Education Attainment
  - World Trend
  - Different Education Levels
  - Gender Difference
- Education Attainment vs Life Expectancy
- Education vs Suicide Rate

# Education Attainment - World Trend

Education Attainment Across the Globe From 1950 to 2010



Timeline (year):

◀ 2010 ▶

◀ □ ▶

▢

▢ ▢

Note:

Click "▶" to play

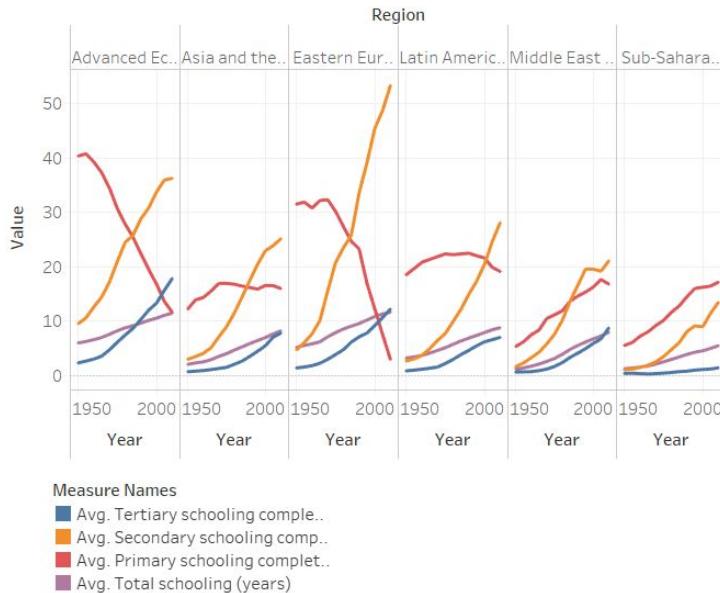
Click "▢" to pause playing

Legend: Average Total Schooling (in years)

0.03  13.24

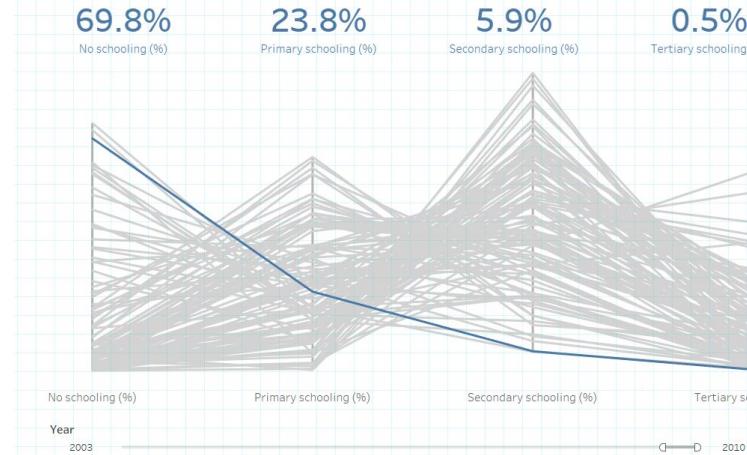
# Education Attainment - Different education levels

## Education by Levels and Regions Overtime



Overview -by region

What is the education attainment for Mozambique?

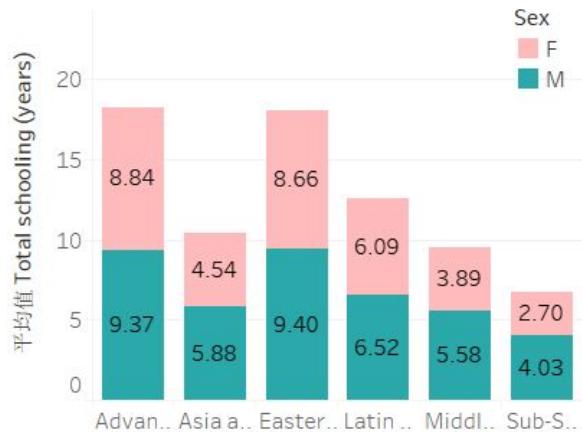


Details - by country



# Education Attainment - Gender Difference

## Education Attainment - Gender Difference - 1985



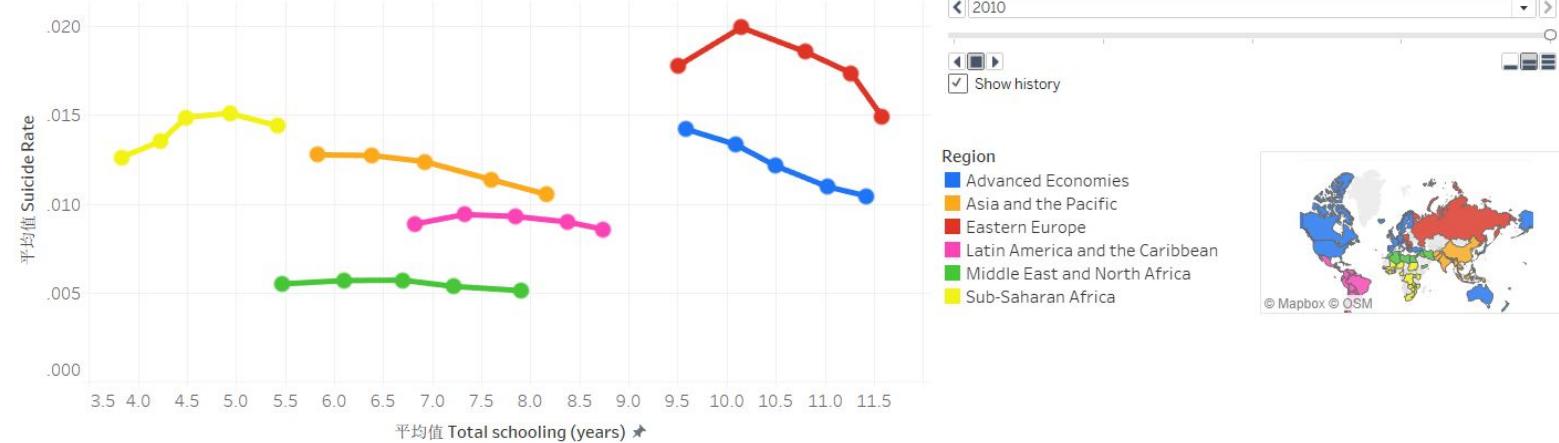
Year

◀ 1985 ▶



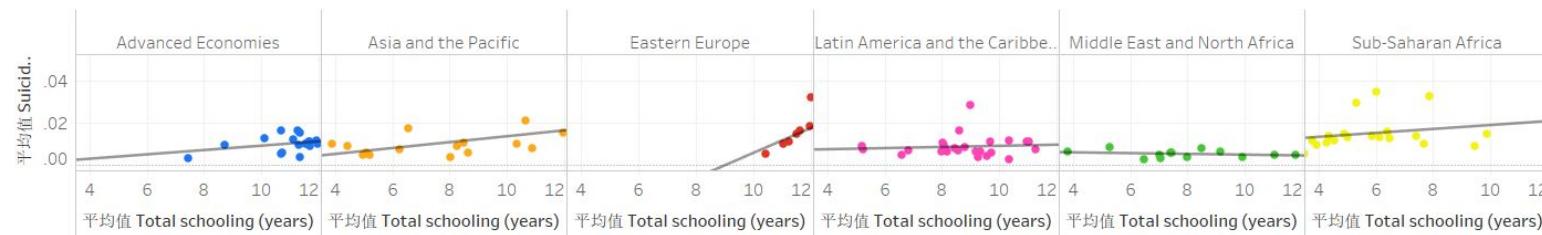
# Education Attainment vs Life Expectancy

Suicide Rate vs Education Attainment from 1990 to 2010



Suicide Rate vs Education Attainment in 2010

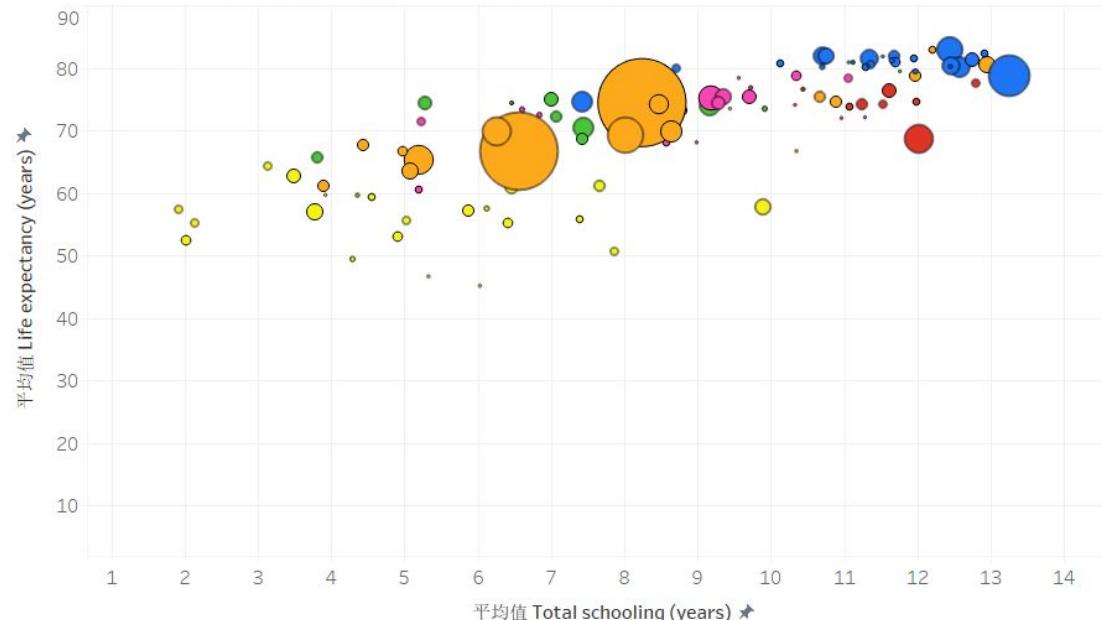
Each dot represents a country



# Education Attainment vs Life Expectancy

## Life Expectancy vs Education Attainment in 2010

Each dot represents a country, size represents population.



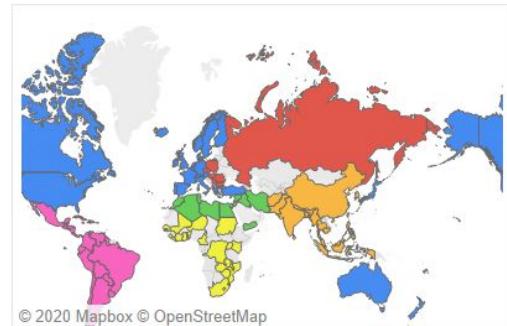
Year

◀ 2010 ▶



## Region

- Advanced Economies
- Asia and the Pacific
- Eastern Europe
- Latin America and the Caribbean
- Middle East and North Africa
- Sub-Saharan Africa



# Education Attainment vs Life Expectancy

How Does **Senegal** Look Like in Terms of Total Years of Schooling, Life Expectancy..

**64.28**

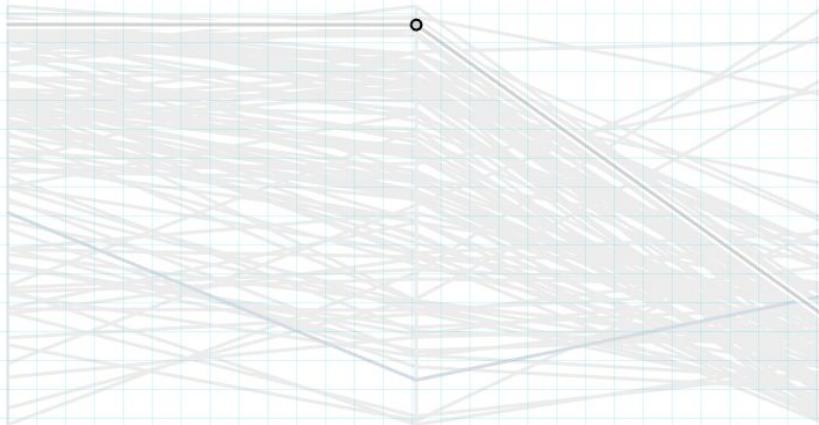
Avg. Life expectancy (years)

**3.13**

Avg. Total schooling (years)

**0.13%**

Avg. Suicide Rate (%adjusted)



Life Expectancy (Years)

Total Schooling (Years)

Suicide Rate(%)

Year

◀ 2010 ▶



Show history

Country Parameter

- Afghanistan
- Albania
- Algeria
- Argentina
- Australia
- Austria
- Bangladesh
- Barbados
- Belgium
- Belize
- Benin
- Bolivia
- Brazil
- Bulgaria
- Cambodia
- Cameroon
- Canada
- Chile
- China
- Colombia
- Costa Rica
- Cote d'Ivoire
- Cuba
- Cyprus



# Usability Testing 2.0

## User #4

**Name:** Raza Naqvi

**Gender:** Male

**Age:** ~45

**Occupation:** Principal Software Engineer, Microsoft



### Experience with Data Visualization:

Raza works at Windows Server division and is responsible for building dashboards for team's key performance indicators. He received tertiary education and is passionate about education space

### Feedback/Questions:

- Is there gender aspect to education attainment and can that be visualized?
- Did you guys had hypothesis or just cobbled the data together?
- Seems like the dashboard is showing charts where hypothesis is proven, and not the ones where hypothesis is rejected. It will be helpful to enable data exploration
- Keep charts for one variable data exploration
- Instead of parallel coordinate for edu, does it make sense to use bar chart?

# Usability Testing 2.0

## User #4

Name: Jason Zhang

Gender: Male

Age: 22

Occupation: Undergraduate Student

### Experience with Data Visualization:

He likes to

### Feedback/Questions:

- At start didn't know it's an animation, because there are two triangle icons, don't know it means play.
- When click the button, expect only change the content below button, not to a complete new screen.
- Hope to see all buttons on all dashboards because going back to home then switching to other screens is troublesome.
- 7. The three-variable parallel coordinates graph is confusing at start, because he thinks a complete table can also show the information. After I suggest him to look at the parallels between education and suicide, he found it's interesting concluded that higher education results in lower suicide rate because of the downward slope of line (actually if looking at parallels their correlation is positive).
- 8. The suicide rate-education line chart is confusing, don't know what it want to convey. The suicide rate-education dot plot is easier to understand and concluded different regions have different correlations.