Education in Another's Shoes
CS 171 Process Book

Faran Sikandar, Gaew Lertsuridej, Lydia Kim

Project proposal

What would my life look like if I had been born in a different country? What more or less opportunities would I have in my life? Our project aims to help users think about these questions by offering perspectives on what life in different countries is like. We hope that the audience will be able to compare their personal background to the lives of individuals of the same (or different) gender and income quantiles in different countries. Our plan is to provide a condensed yet informative view of life in a country for an individual of a specific gender and economic background. By looking at the life trajectory as a whole as well as in distinct phases, we hope to answer questions such as:

- How long am I expected to live?
- How many siblings would I have?
- What kind of home would I live in and what kind of possessions would I have?
- What are my chances of surviving past 5 years old?
- When will I enter school and how many years of education will I obtain?
- When can I expect to get married and/or have children?
- When will I start my first job and how much would I make?

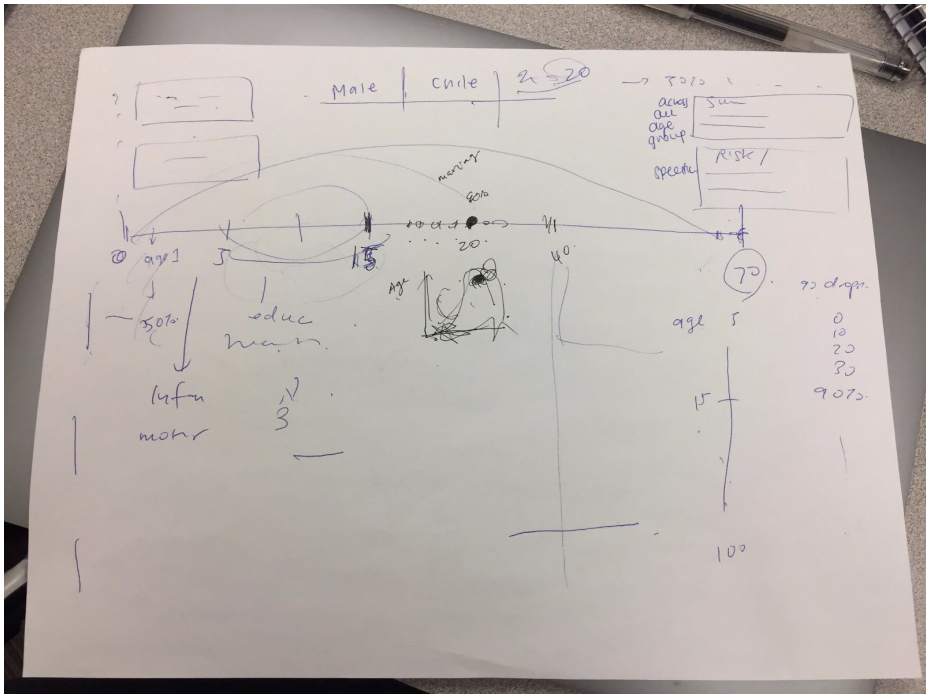
We plan to use the Demographic Health Surveys (DHS) to obtain nationally representative data on health outcomes, family composition, household assets, marriage, employment, wealth, etc. The DHS are mostly conducted in developing countries, so in order to ensure a wider range of perspectives, we plan to draw from census data available on websites such as IPUMS to include more developed countries.

Current TO-DO list:

- Look at DHS to determine which countries and indicators are available
- Look into other potential data sources for developed countries

Designs

Initial sketch



Ideas

Potential indicators

- Age-specific
    - Infant mortality
    - Maternal mortality
    - Education - age of entry, attainment, drop out
    - Likelihood of stunting
    - Pregnancy, marriage
- General
    - Household assets - type of home, ownership of assets, electricity, water source, etc.
    - Household composition - siblings, size
    - Literacy
    - Meals

Update: Project Detailed Proposal

CS171 Project plan

Gaew Lertsuridej, Faran Sikandar & Lydia Kim

November 6, 2017

1 Goals & Tasks

1.1 Goals

The main objective of our project is to provide the audience with perspectives on what life is like in countries around the world. We expect that the majority of our audience will be US-based or from another developed country, so we hope to give them the opportunity to explore what life might have been like if they had been born or grown up in a different country. Specifically, we hope that the audience will be able to compare their personal background to the lives of individuals of the same (or different) gender and wealth categories in different countries.

Our plan is to provide a condensed yet informative view of life in a country for an individual of a specific gender and economic background. As there are various facets of life that we could potentially explore in our visualization, we plan to organize our visualization by breaking the human life trajectory into 4 broad age groups and look at topics particularly relevant to that age range. By doing this, we hope to make put the audience "in the shoes" of another individual and make the visualization more relatable than just a presentation of a number of facts. The 4 life phases and their corresponding indicators are:

- **Ages 0-4**: Household composition (parents, relatives, siblings), household size, infant mortality, stunting, maternal mortality, parent literacy, HH assets, make of house, availability of electricity, water, etc.
- **Ages 5-19**: education (age of first enrollment, years of education, age of dropout/graduation) & potentially age when leaves education to enter the labor force or work at home
- **Ages 20-45**: type of job, occupation industry, season/migrant worker, marriage, pregnancy, birth etc.
- **Ages 45+**: health (likelihood of contracting X disease), life expectancy

At the beginning of the visualization, the user would be able to choose a specific country, wealth category, and gender. Based on these filters, the data displayed in each of the following frames would be updated. In addition, throughout the entire visualization, the user would be able to access the filters in case they want to make quick comparisons. A timeline would also be displayed on the side to guide the user through the various phases of life.

1.2 Tasks

Our immediate tasks include:

- Assessing which time frame, countries, and indicators we would like to include in our visualization
- Cleaning the data to extract relevant variables in usable format using Stata

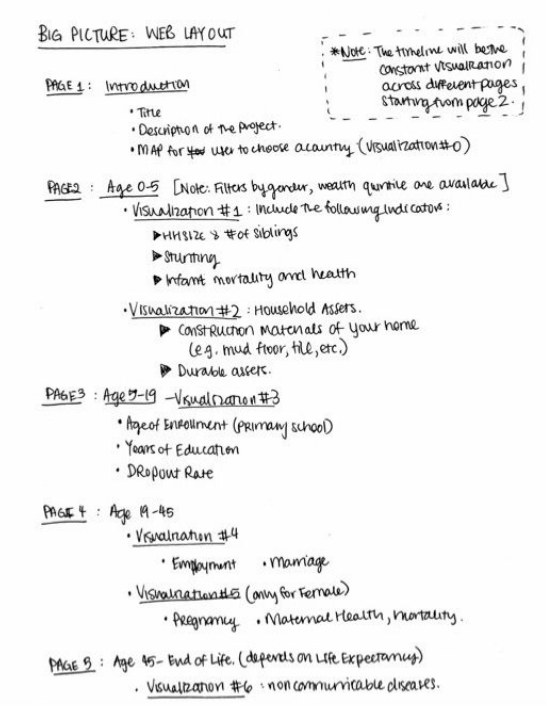
- Meet with Fritz to discuss how to better encode some of our key indicators
- Somewhat finalize our visualization sketches and interactive storyboard
- Implement!

2 Description of data

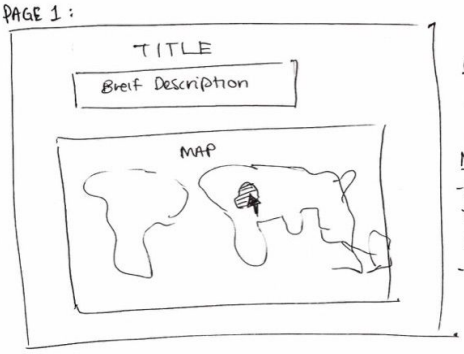
We plan to use the Demographic Health Surveys (DHS) to obtain nationally representative data on health outcomes, family composition, household assets, marriage, employment, wealth, etc. The DHS are mostly conducted in developing countries, so in order to ensure a wider range of perspectives, we plan to draw from census data available on websites such as IPUMS to include more developed countries.

The data will require a bit of cleaning, as there are many different recodes with different respondents. Fortunately, the variable names for the same indicators are more or less the same across DHSs of different countries of the same survey number (i.e., surveys that happened in approximately the same time frame). We have some of the data in Stata .dta format and plan to carry out the initial variable extraction and cleaning in Stata.

3 Visualization sketches



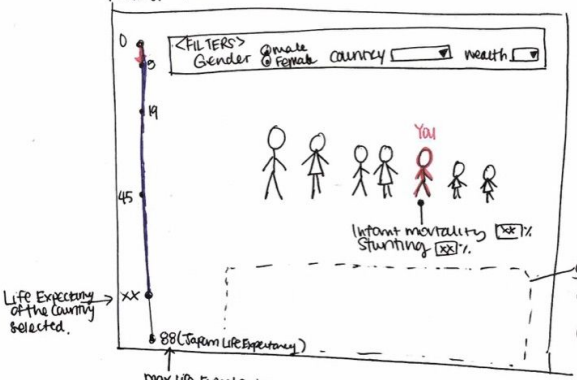
PAGE 1:



Option:
Select the country
→ either zoom or disappear (toggle out)
→ Then page 2

Note:
- Grey out the countries that we don't have data for
- Hover over and country would change colors.

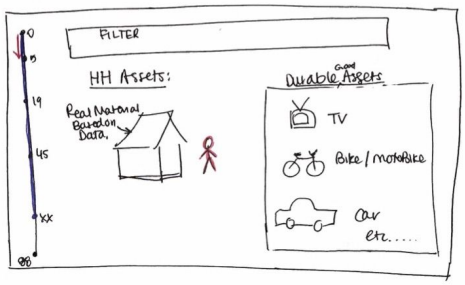
PAGE 2: Visualization #1



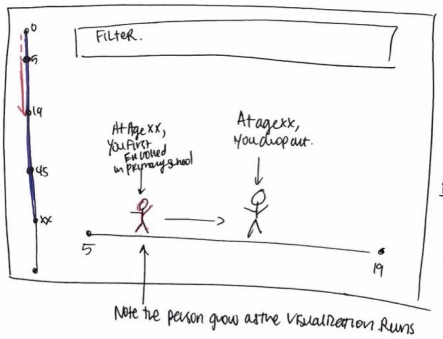
- Image of people = household size + # of siblings.

Optional [TBD]
add a comparison visualization. (using contrast to narrow the attention to the main vis.)
e.g. # of HH members in US, # of siblings in US, etc.

Page 2: Visualization #2 (Age 0-5)



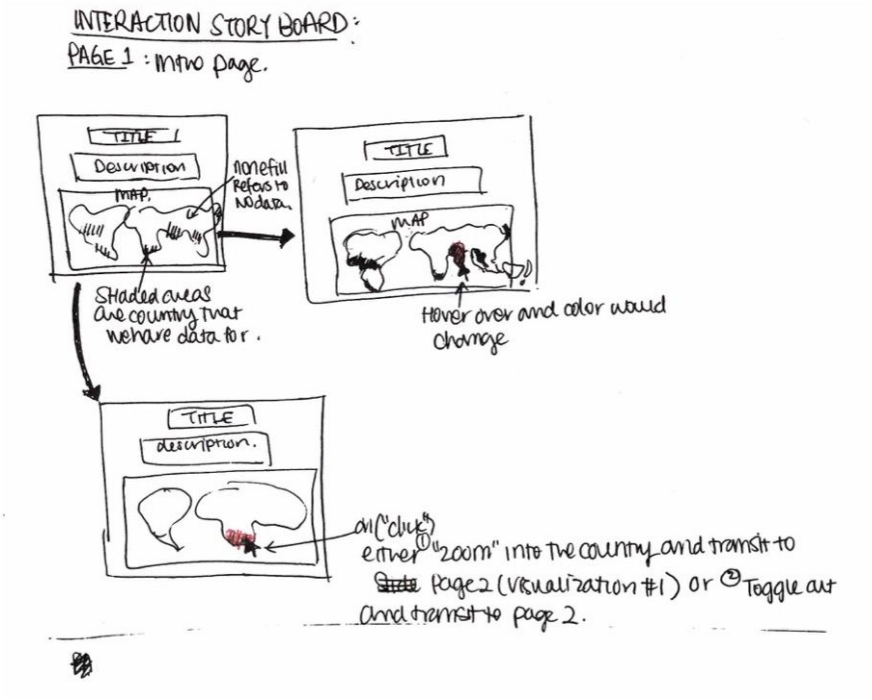
PAGE 3: Visualization #3 (Age 5-19)

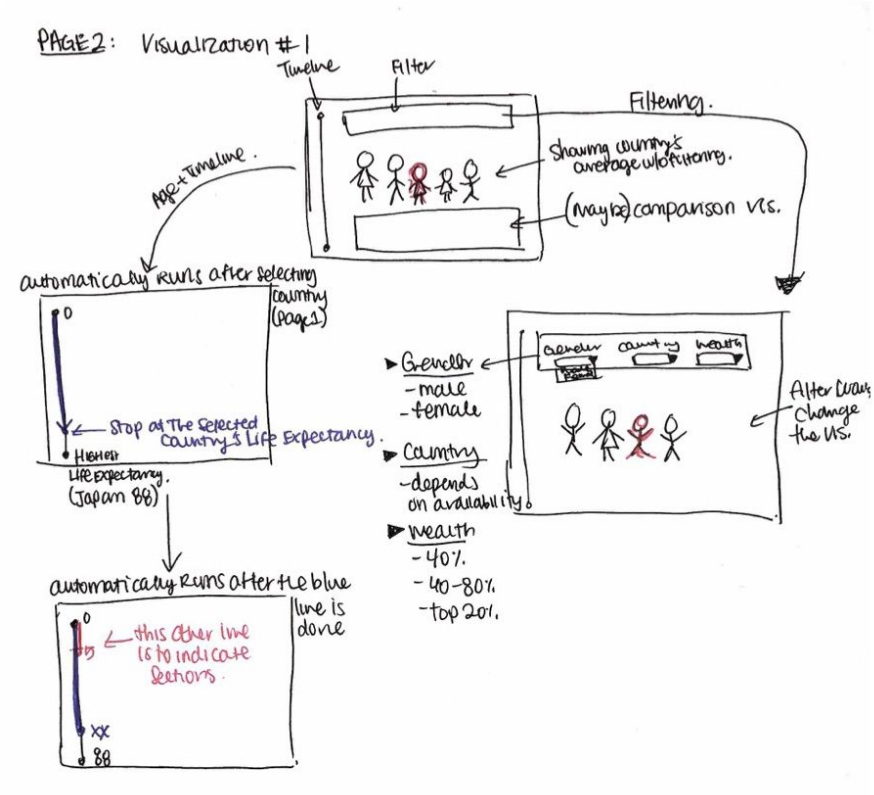


PAGE 4 & 5
↳ TBD

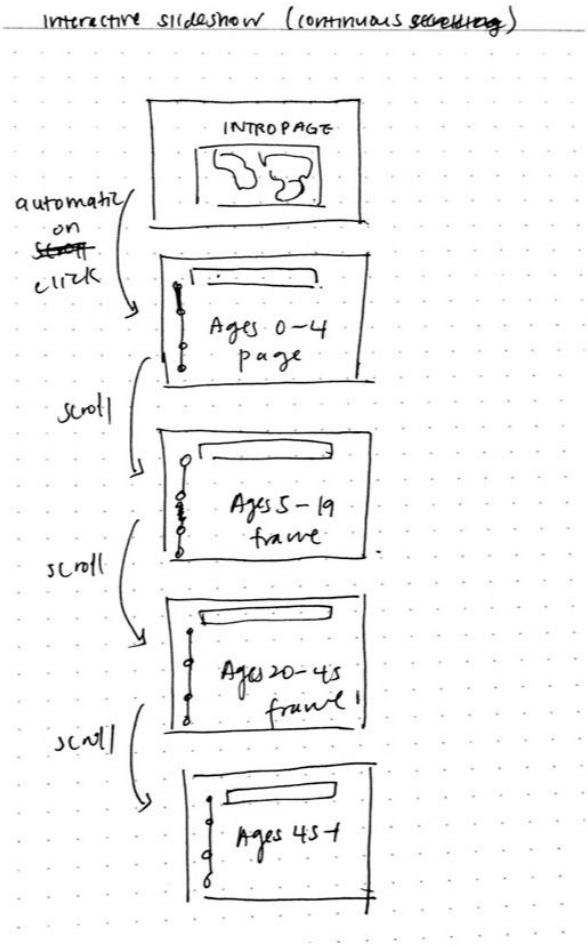
Note the person grow as the visualization runs

4 Interaction storyboard





5 Webpage layout



6 Project timeline

- **Nov 8**: Set up repository, finish choosing indicators/countries & data cleaning
- **Nov 9 or 10**: Meet with Fritz to go over project proposal & finalize visualization sketches
- **Nov 10-13**: Work on prototype v.1 (introduction and visualization 1)
- **Nov 13**: Prototype v.1 due
- **Nov 17**: Finish introduction, age group 0-4
- **Nov 24**: Finish age group 5-19, age group 20-45
- **Nov 27**: Prototype v.2 due
- **Dec 1**: Finish age group 45+ and test visualization on first-timers through thinking aloud studies

7 Feature list

Must-have features

- Filters that allow user to select gender, country, wealth category
- Timeline that provides broad overview of the visualization and is present in all frames to connect one to another and guide the user through the visualization

Good-to-have features

- Visualization/statistics in the bottom/side of each frame that allows comparison of selected country with the US/another country
- A map in the beginning that allows the user to see what countries in which regions are available to select
- Annotations and text to guide the reader

Optional features

-

8 Description of team roles

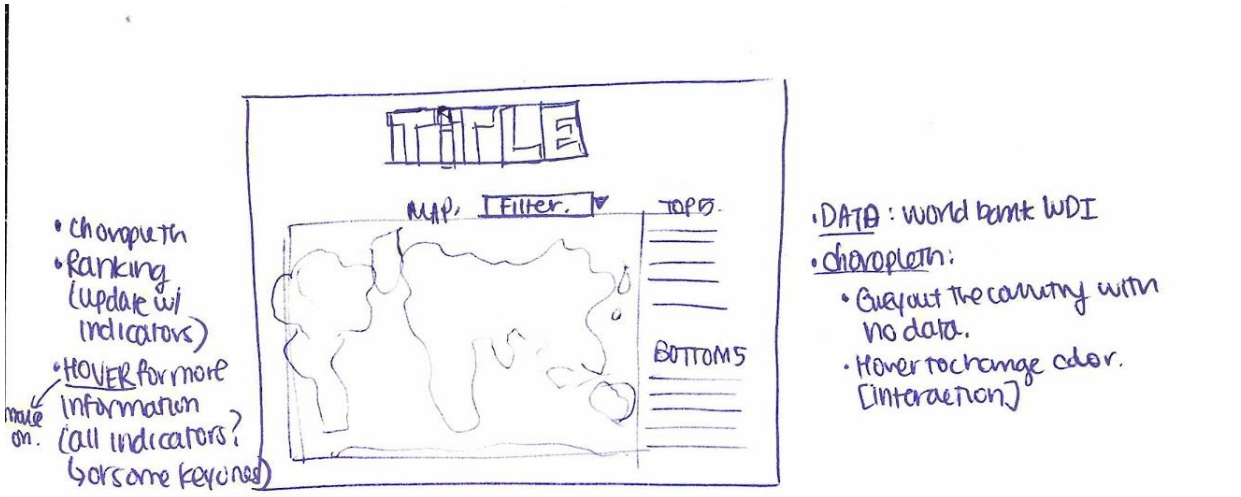
We plan to work collectively at every stage, but the point person for each task will be as follows:

- Faran – Target, making sure our visualizations are true to their original objectives
- Lydia – Data wrangling, updating process book
- Gaew – Evaluate, making sure team is on schedule
- Design/implementation – All together

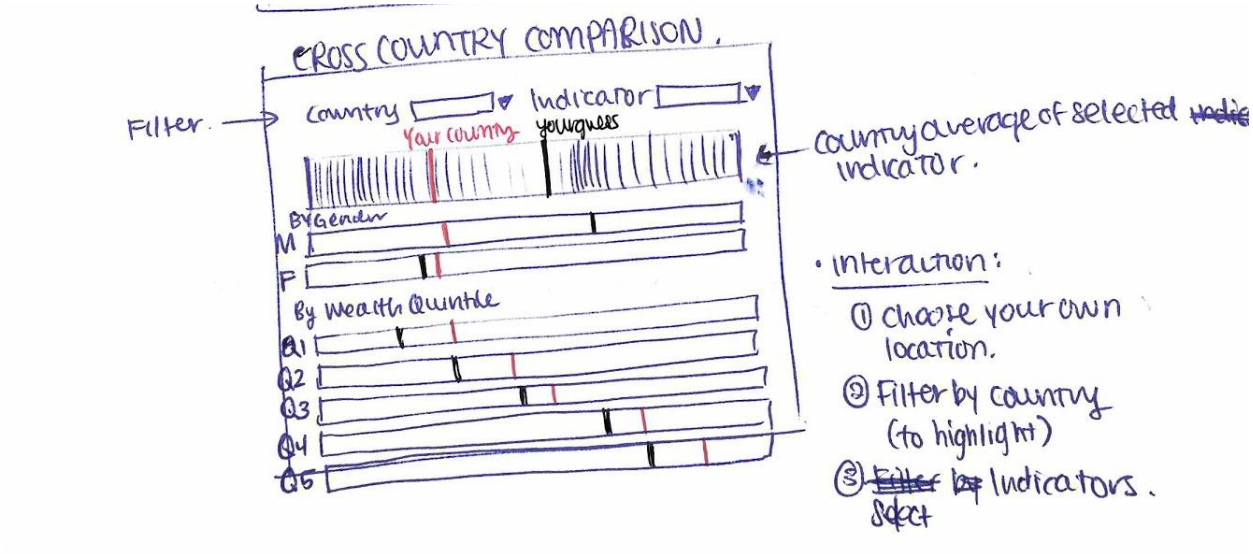
Update: 11/10/2017 After a meeting with Fritz

- No clear objective and storyline based on the previous plan, and therefore the team decided to change the visualization to the following:

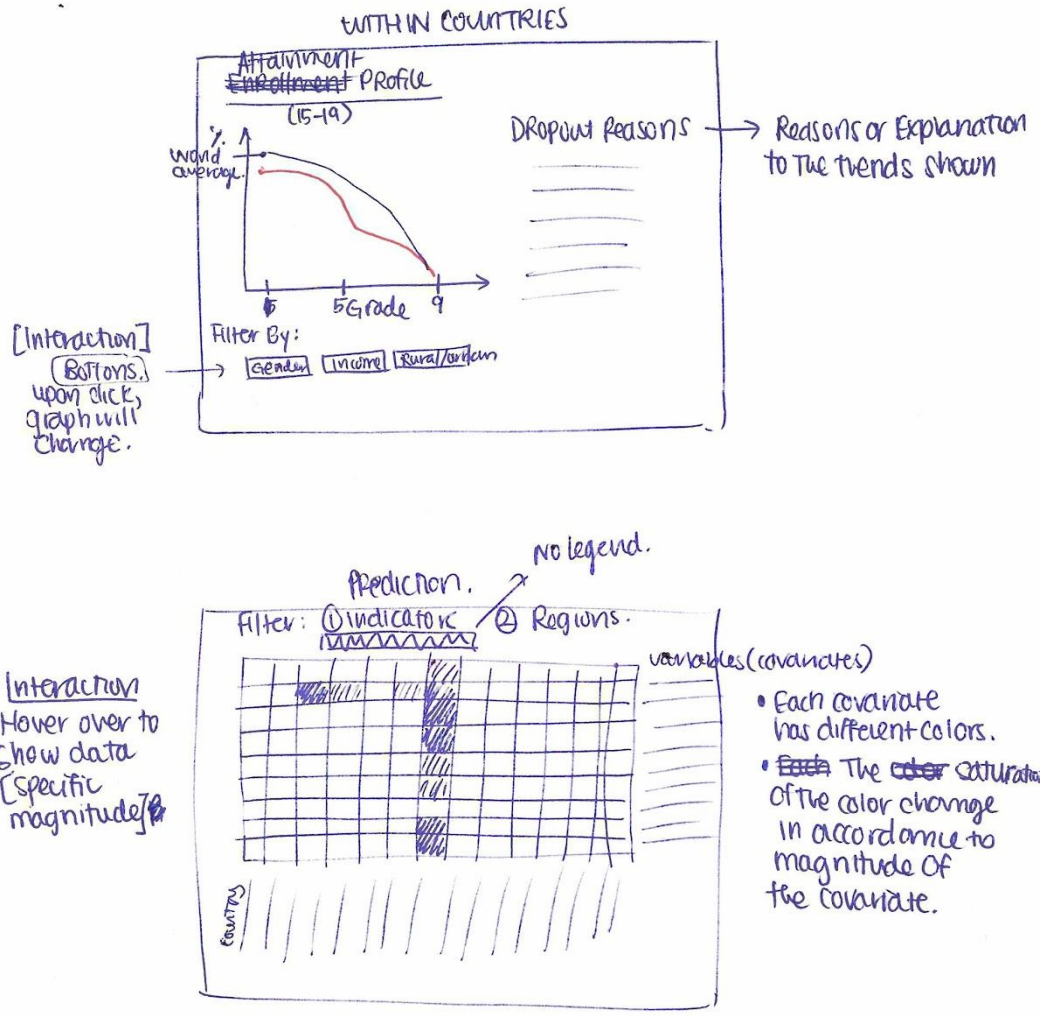
Visulization 1:



Visulization 2:



Visualization 3 and 4:



- Will also be using WDI dataset (world bank)

- Instead of focusing on different variable for different individual, we will only look into details for education sector.
- We will first look at the overall cross-country comparison.
  → Then dig deeper into distribution & how that distribution looks differently for (i) male and female and (ii) across different income group
  → Then we will look at education attainment profile of different countries
  → Then we will end with prediction of labor outcome (e.g. likelihood of having waged jobs) or tertiary education outcome (e.g. likelihood of obtaining an undergraduate degree)

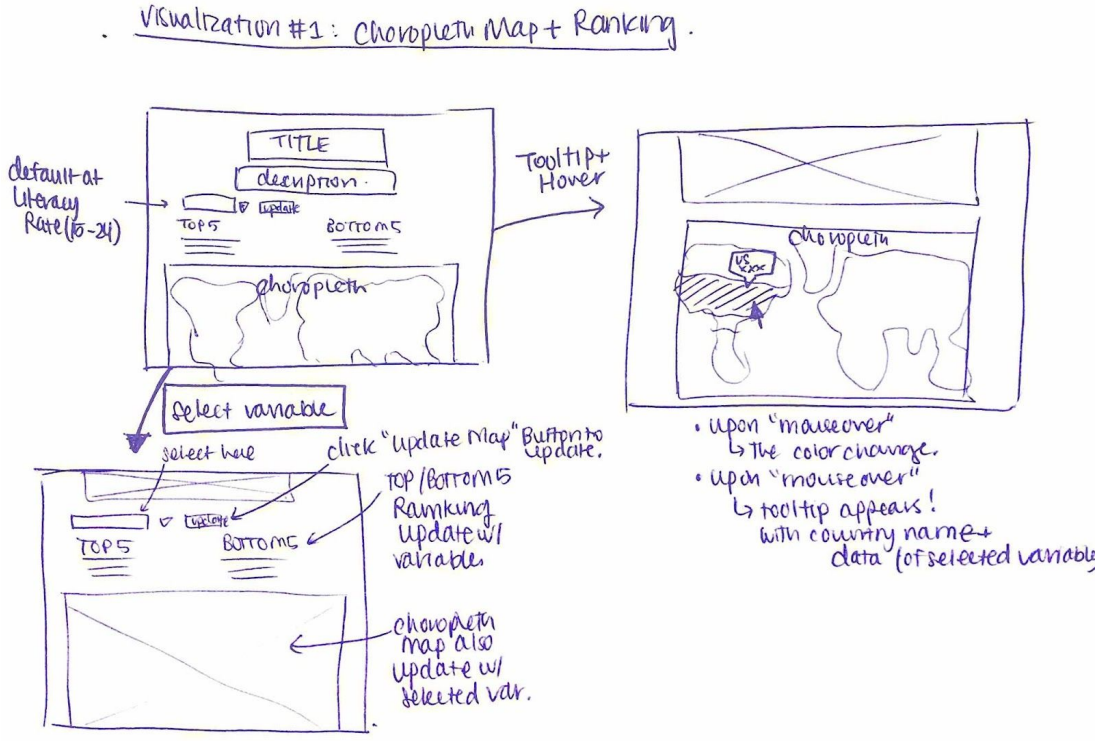
For the prototype v1:

- Gaew: Selection of variables for Choropleth + ranking (first page) & Implement the visualization
- Faran: Overall website layout
- Lydia: Cleaning the data (WDI, DHS)

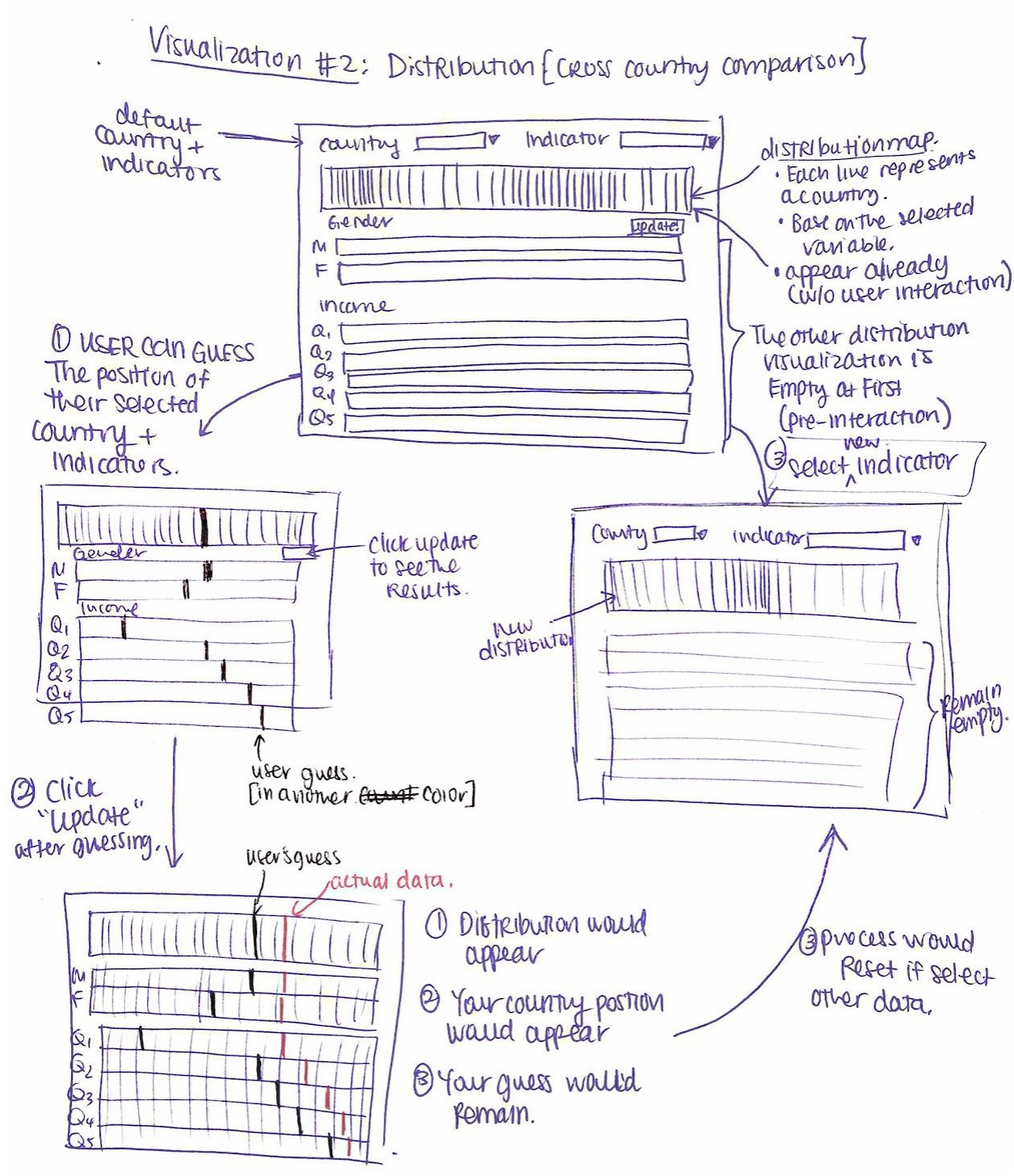
For prototype v1 submission:

1. Implementation of visualization no. 1 (see testvis.html)
2. Cleaned Data: mapeducation.csv, worldmap.json, DHS_data.csv
3. Draft for the other visualization (see figures under updated from 11/8/2017) and new storyboard (see below)
4. Website layout (see index.html)

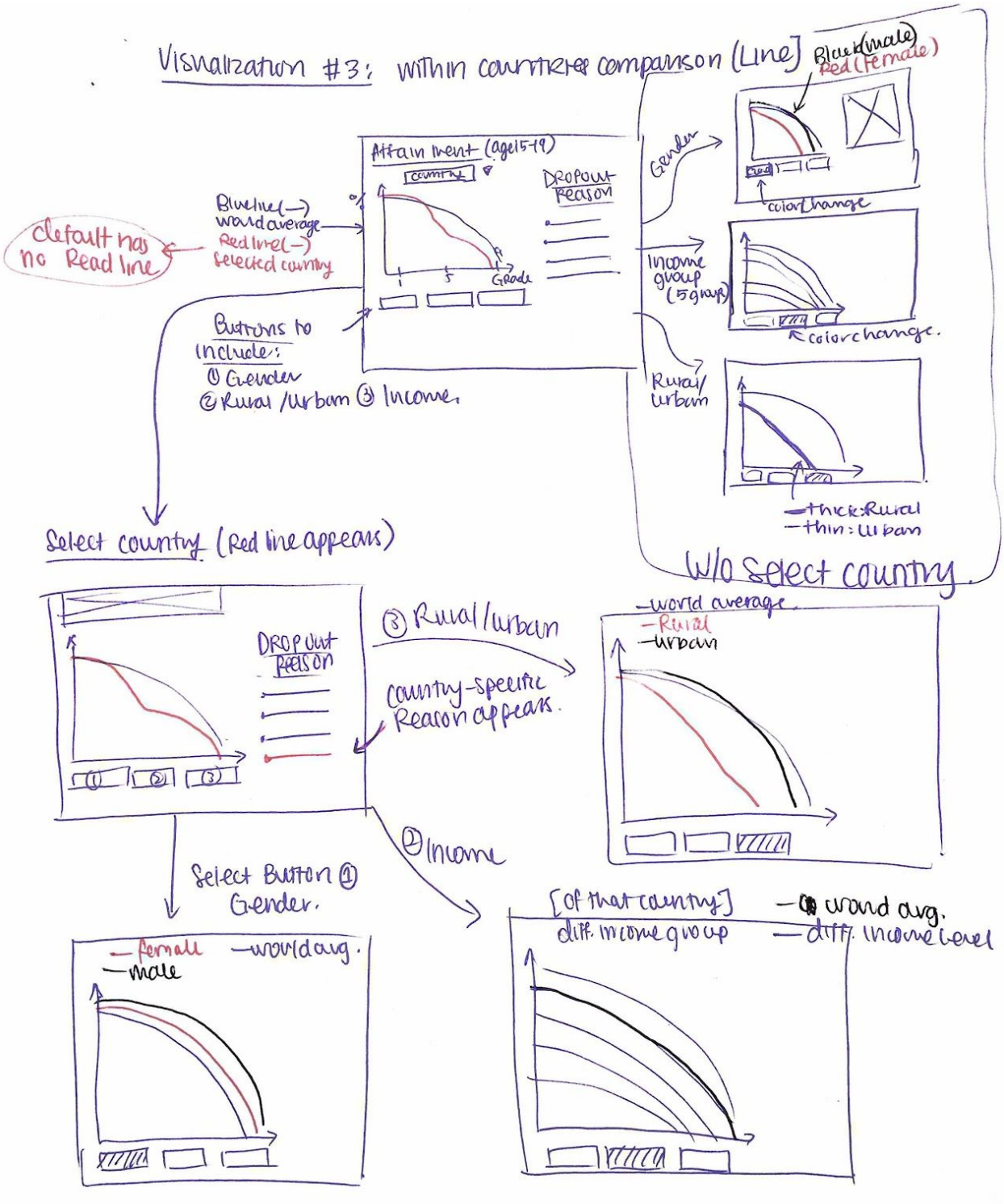
Visualization 1:



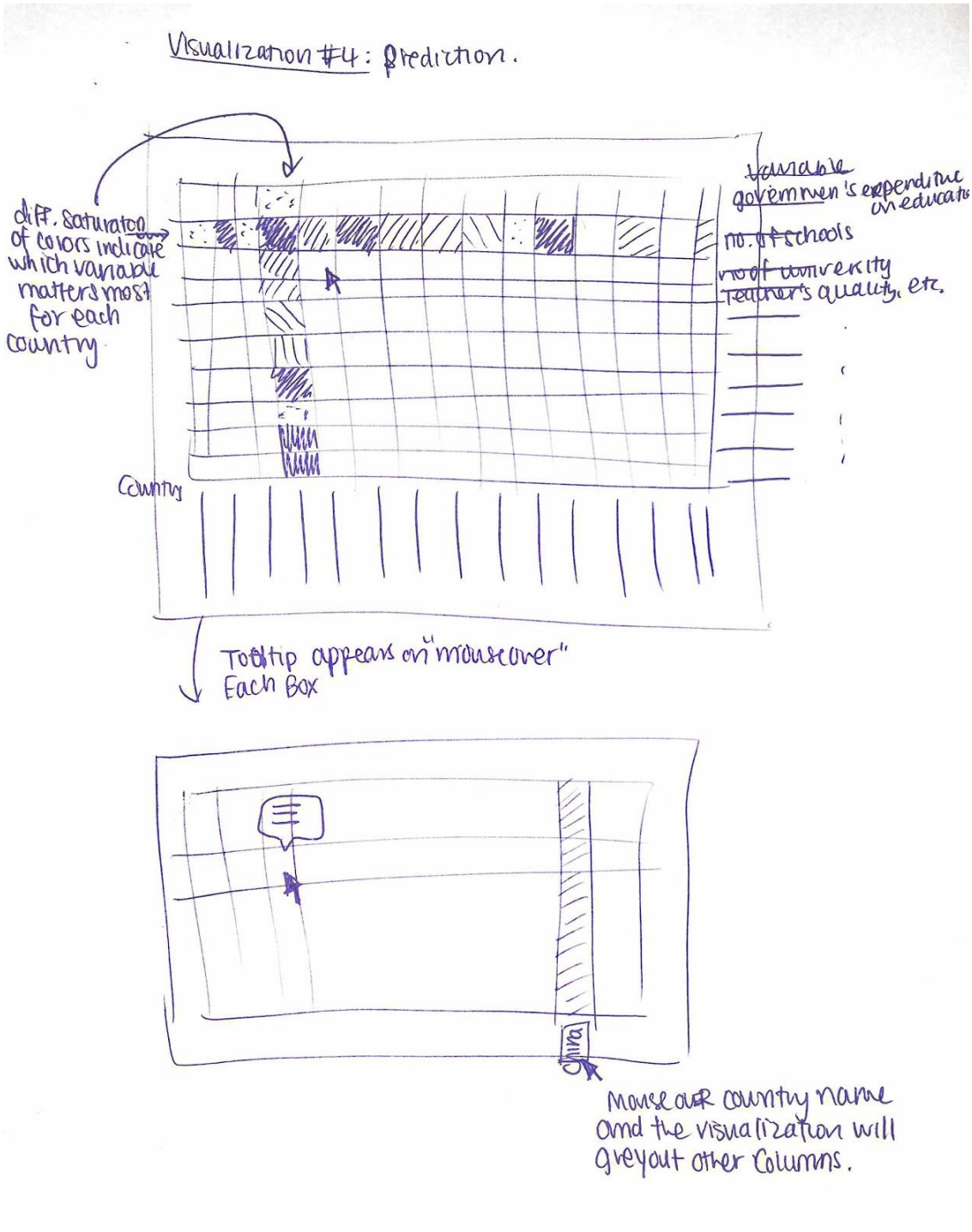
Visualization 2:



Visualization 3:



Visualization 4:



11/16/2017 Meeting: Updates and Discussion

Visualization #3:

- DHS (perfect and all the data needed)
    - 51 countries, mostly in Africa and developing countries
- IPUMS
    - Mostly developed countries (European countries + US, etc.)
    - But variation in variables
    - 32 countries
- Is this case, shall we generalize a bit more for the attainment profile visualization (visualization #3)? Or still use both (and change the visualization)?
    - Conclusion: For visualization #3, we are going to impute data for EU countries and visualization #2, we are going keep it at a higher level (completion of primary school, completion of secondary school, etc. and not year by year).
    - For Vis #3, since the attainment profile of the developed countries are similar across different countries and the ones that are more interesting is developing countries. Then we are gonna have one representative line (as comparison).

Visualization #2:

- We also have Education data from Ed School at Harvard (Barro-Lee)
    - 150 countries
    - Shall we also include the data in Visualization #2?
        i. Yes, we are going to use this data instead for vis #2
- Changes to visualization #2:
    - Keep Male/Female breakdown
    - Instead of the income quintile, we are doing age
    - Only focusing on 1 variable (year of education)
    - Add the story line:
        i. User guess first on the blank distribution visualization (can guess for overall, male and female)
            - "Countries started with very different level of education in 1950. Can you guess where you country was in 1950?"
            - "How about 2010?"
        ii. Then:
            - Bring in different region one by one
            - Animate over time
            - Highlight the actual data for the selected country
        iii. User can guess again for male and female
        iv. Then:
            - Drop all data points on the distribution at the same time (don't do by region)
            - Play change over time for one at the time upon clicking "Play"
            - Show where the country actually is at 2010 and 1950

Delegating the work towards Prototype v2:

Work to do overall:

- Layout
- Visualization #2
    - Gitter plot
    - Populate one region at the time
    - Guessing (interaction part)
    - Animation over time
- Visualization #3
- Visualization #4

To-Do:

Lydia:

- Finalize the analysis (PCA for generating income quintile) for other visualization
- Push the submitted prototype v1
- Run regressions
- Visualization #4
- Layout for website

Faran:

- Visualization #2
    - Skeleton by Sunday (Jitter plot)
    - Animation by Friday after Thanksgiving

Gaew:

- Finalize Visualization #1 by Monday
    - Including frame + box/table to the ranking
    - Grey out data that is not available
    - Keep the variable list for now
    - Think about the story and the objective of this study and add some title and text.
- Visualization #3 by Friday after Thanksgiving
- Redraw the storyboard for Visualization #2

11/28/2017: Prototype v2 Presentation

Feedback from Class and studio

Choropleth:

- Legend on top instead // after adjusting the layout, it looks ok on the side.
- Zoom (maybe)
- Ranking more attractive
- Guide user to use the tooltips
- Ranking on the side rather than on top
- Adjust the layout to fit within the full page.
- Transition for the choropleth maps // broke the tooltip and chose tooltip > transition

Distribution Vis:

- Still line rather circle.
- Animations
- Add option on top

Attainment Vis:

- Adjust the layout a bit to make it work within the full page.
- Remove the legend entirely
- Input for country selection + autofill
- Reduce the size of data (other ages, there's no need to include them if the data is not being used).

Prediction Vis:

- Explanation of the colors and the numbers
- Change the colors?
- Sort based coefficient?

Overall:

- Add a storyline
- More images and icons
- Explain data (some data is not available)
- Explain vis that only use data from developing countries, etc.

Specific to-do for distribution visualization

Distribution Vis

- You Draw it (user engagement) (Gaew and Lydia to explore)
- Animation/ visualization (Faran)
- Select the country (Gaew, dropdown menu)
- Animation: playing over time
- Animation: pulling in data
- Male/female animation

In class feedback

CS 171 Project Presentations

(Give the completed form to the team you gave feedback on. They will have to scan it in and attach it to their final submission.)

Your Names: Tansaya, [illegible], Jamin

Your E-mail: tansaya_leenaratskul@hks19.harvard.edu

Name of group you evaluated: Benefactors
Education for Central Planning

What is good about the group's visualization?

- ~~[illegible]~~ a lot of data, good for reference
- colors are great

What could be improved?

- small text
- have legend up front
- sort by predictor
- more explanations for data interpretation

Is the message clear? What is the message?

CS 171 Project Presentations

(Give the completed form to the team you gave feedback on. They will have to scan it in and attach it to their final submission.)

Your Names: Michelle Ng
Celina Qi

Your E-mail:

michellng@ college ; celinaqi@ college

Name of group you evaluated:

Education in Another's Shoes

What is good about the group's visualization?

- so clean! we love your color palette
- the factors square vis is so cool and encourages exploration
- we like how you can click on different categories for the line chart

What could be improved?

- maybe more transition text to guide the story?
- a bit of explanation about regressions (you said this yourself so you already know !!)

Is the message clear? What is the message?

yes! huge educational disparities worldwide, due to a number of factors

12/4/2017 Meeting with Fritz

- Vis#1: choropleth
    - Fixed data (if data is missing, just grey them out)
    - Consider changing hover methods (opacity can be confusing depending on the vis)
- Vis#2: Distribution
    - Focus on finishing this one
        - Since it's currently confusing, consider
        - Doing one colors first (to see the overall distribution)
        - Be able to separate different regions
        - Consider doing stacked
        - Add Tooltips
    - Low priority: user interface
- Vis#3: Attainment Profile
    - Add developed country data
    - Add framing text to ask the user to compare between developed and developing countries
- Vis#4: Matrix
    - Point out interesting pattern
    - Add sorting
- Add conclusion slide

12/10/2017 Call/Meeting to discuss final touch-ups and issues

Gaew:

General

- Page 1: “Have you ever wondered what your educational outcomes would be if you were born in another country?”
- Add a page for conclusion (slider wise)
- Process book

Choropleth:

- Ranking - text too big
- Title to shift into left // Note that the map looks cut off currently..

Distribution

- Take a look at Distribution vis. / button / dropdown menu

Attainment:

- Tooltip remove information
- Remove 0 from x axis.
- Change size or color mouseon

Lydia:

- Fix the scrolls of the slides
- Stacked scatter plot
- Look at questions to frame both vis 3 and vis 4
- Look at data for attainment + buttons
- Add instruction
- Tooltip on distribution?

Faran:

- Tooltip “mouseout”
- Cambodia, PNG in Europe? Data
- Axis (year labels)
- Photos
- Conclusion

Other updates:

Distribution: Drop user interface

Final:

Link: https://fsikandar.github.io/CS171_FinalProject_Education/