

Group 5

Grace Austen, gausten@g.clemson.edu, C11076684

Elliot Phillips, cep6@g.clemson.edu, C19199755

Derek Walker, djwalke@g.clemson.edu, C15141932

GitHub Repository: <https://github.com/Grace-Austen/CPSC-4030-Project>

GitHub Page: <https://grace-austen.github.io/CPSC-4030-Project/>

Presentation Video: <https://vimeo.com/779617495>

School Rules, Actually

Overview and Motivation

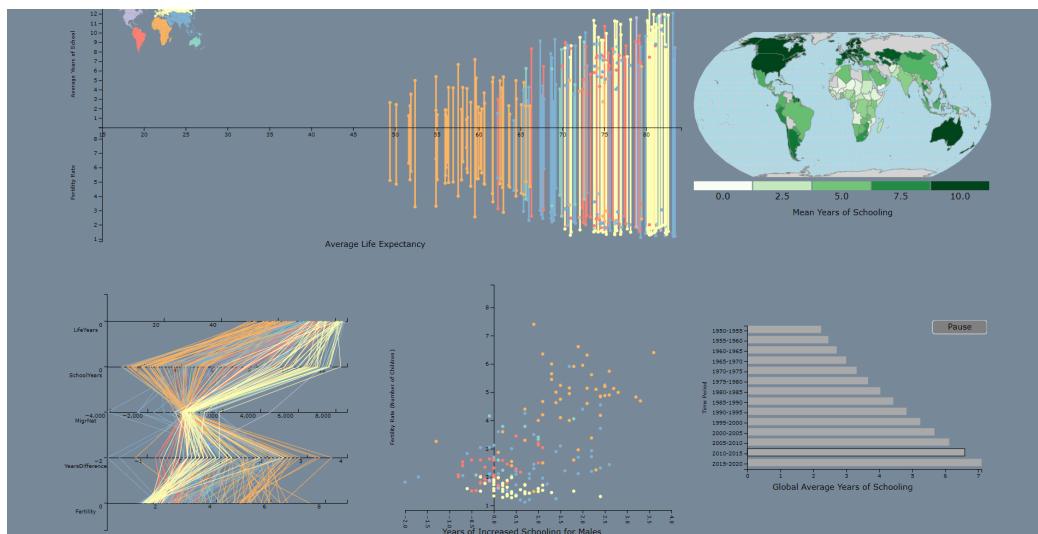
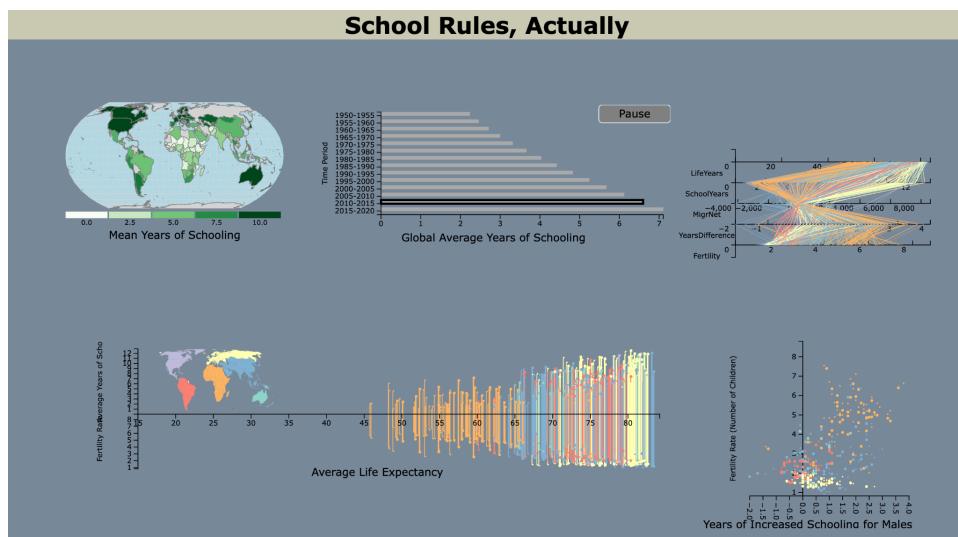
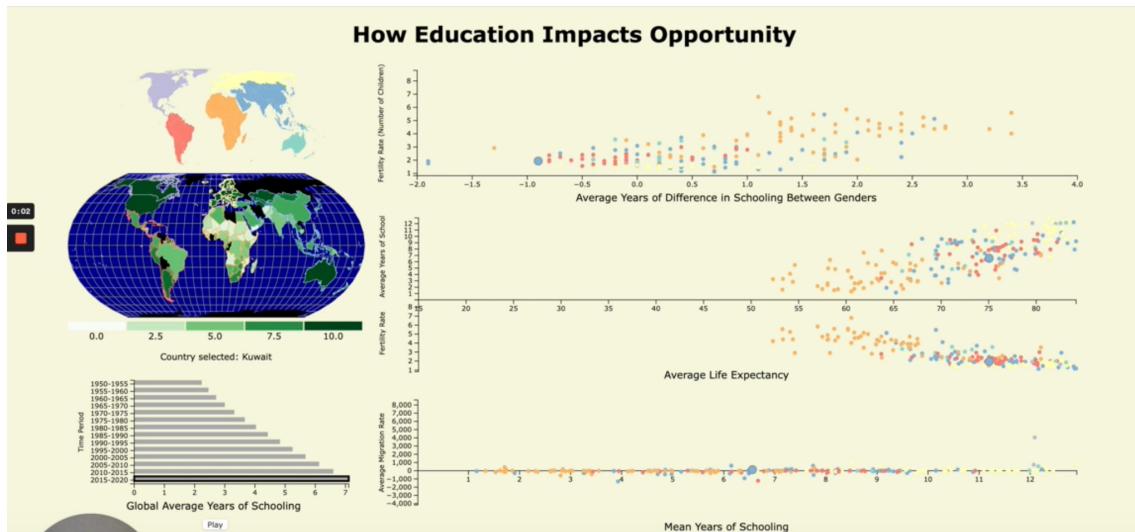
This project serves to provide insight into the importance and impacts of education on a global scale. Achievement in education provides an interesting look at the opportunities of individuals throughout the world and provides a tangible measurement of some attributes of inequality that individuals in certain geopolitical regions struggle with.

Our motivation for this project is that, as U.S. students in higher education, it is often easy to lose sight of the opportunities granted by education as well as just being an individual in the United States of America. Increasing our scope to a global scale provides an opportunity to reflect on the advantages provided to us and note areas that are still struggling or have higher degrees of opportunity and equality.

Related Work

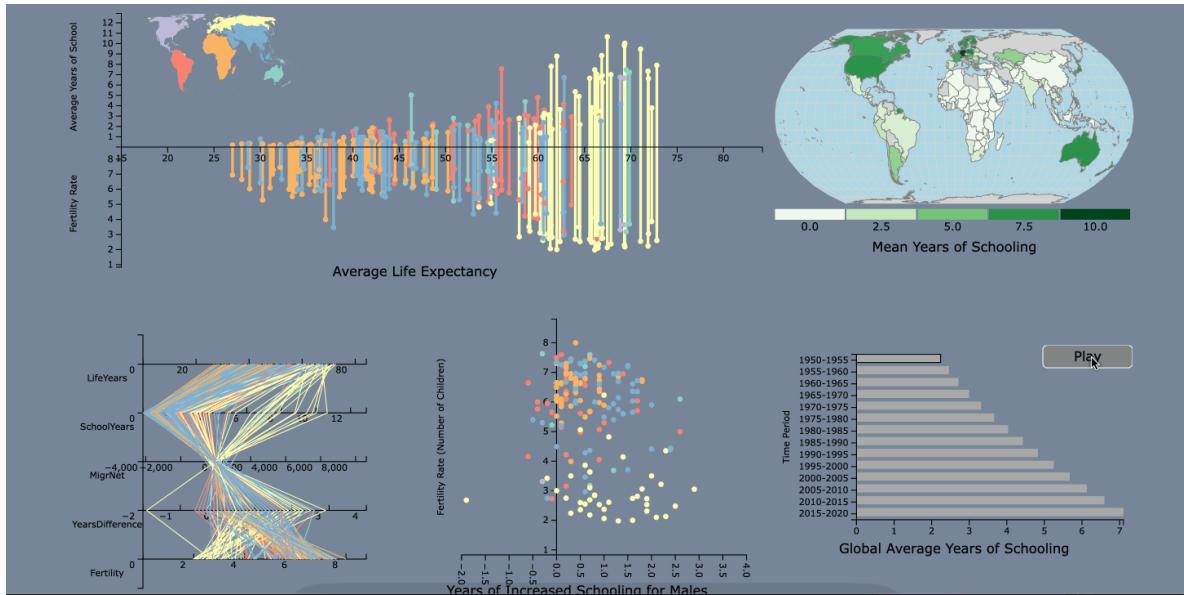
Upon receiving feedback from our prototype of this project, we realized our website layout was not the best for effectively conveying the information we were trying to communicate. We found this article

<https://cxl.com/blog/10-useful-findings-about-how-people-view-websites/#h-1-the-top-left-corner-gets-the-attention-first> from CXL, which is a marketing education platform that made us rethink the placement of our visualizations.



By considering the areas of priority for viewing, we have increased user attention to the colors we chose to represent the countries, highlighted the visualizations we feel most effectively communicate our discoveries, and provided a natural flow through the website.

The TED Talk delivered by Hans Rosling (<https://www.youtube.com/watch?v=hVimVzgtD6w>) that we watched in class was also a significant motivator in some of our choices for interaction and provided motivation that our project was on the right track during the development phase, despite watching this video well into the work for this project. Notably, it showed the power of animation moving through the years.

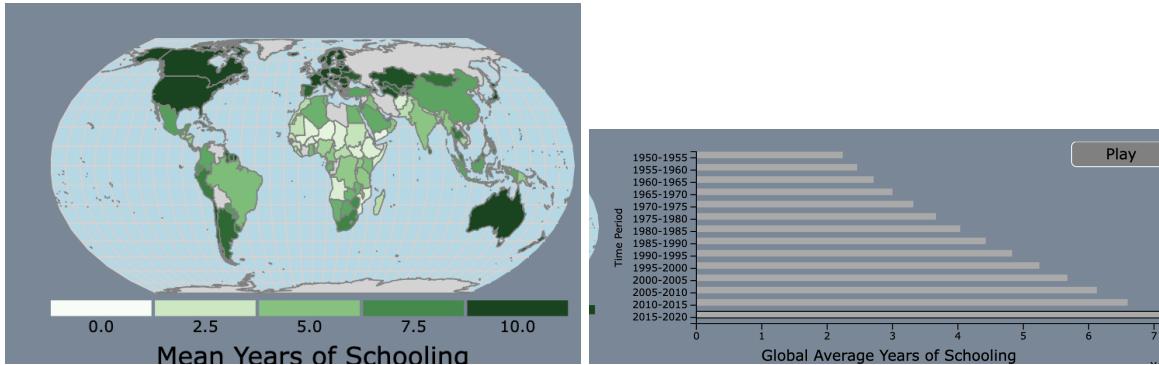


Questions

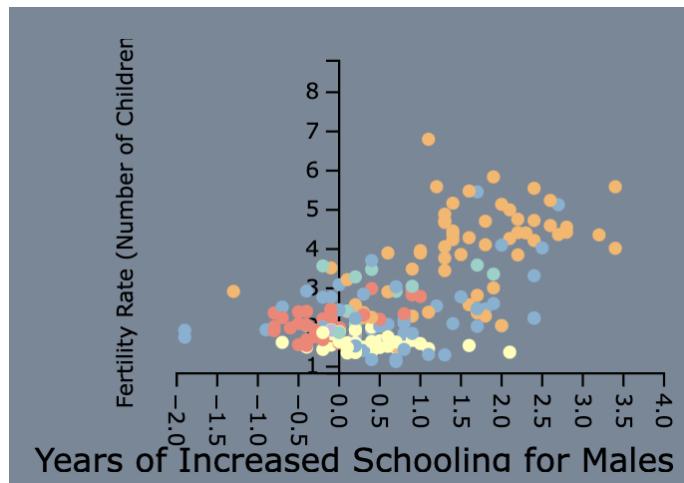
The initial questions we proposed were:

1. What is the average amount of schooling globally, and how has it changed over time?
2. Is fertility rate correlated with the gender gap in education?
3. How does life expectancy impact fertility rate and mean years of schooling?
4. How does education impact migration rate?

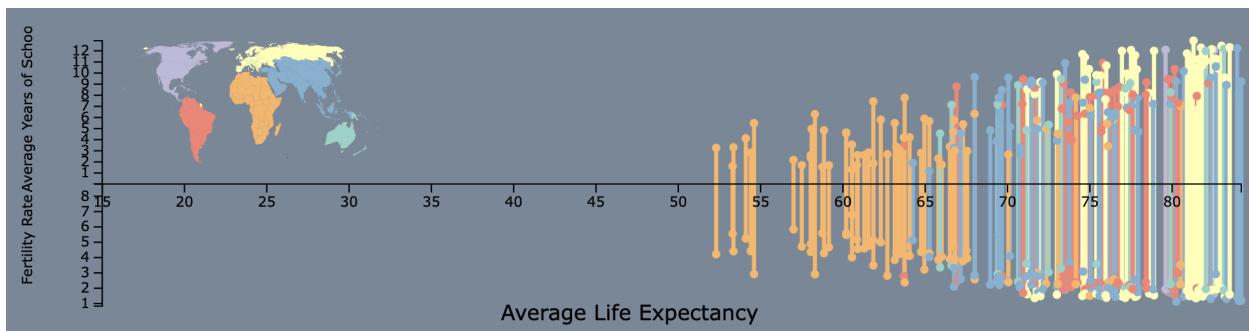
Question 1 ended up splitting into a way to encode the interface for the user interaction in an interesting way, as we have a visualization that serves as a country selector as well as a visualization for the average years of schooling for that country and the global average turned into a bar chart that also serves as a time period selector.



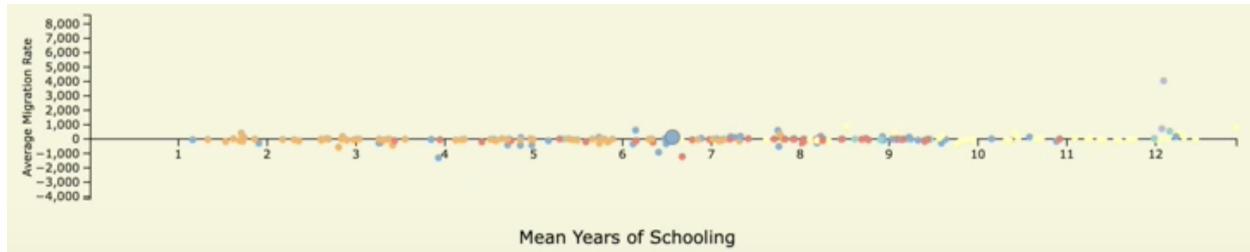
Question 2 remained the same throughout the development process for the project.



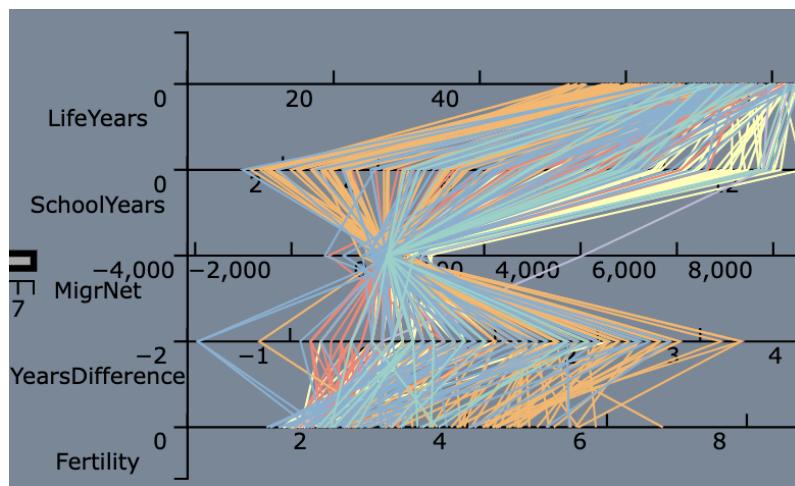
Question 3 remained the same throughout the development process, and the difference between fertility rate and years of schooling ended up being an incredibly interesting analysis within the overall question.



Question 4, unfortunately, did not result in as interesting of a visualization as we had hoped. In its original state, it was very difficult to see any trend.



We still felt that migration was an interesting measure of how individuals responded to the opportunity or lack of opportunity presented to them by their country, so we pivoted to see how all of the measures we looked at interacted with migration. This design more effectively communicates the clustering and highlights the outliers further.



Data

Our data source was the Wittgenstein Centre Human Capital Data Explorer (<http://dataexplorer.wittgensteincentre.org/wcde-v2/>), which is a collection of historical data and future projections about a multitude of attributes for humans by country from 1950 projected to 2100. For our project, we chose not to include any projections and just focus on historical data from 1950-2020. In addition to education, the main driver of our project, we also thought fertility and difference in the education attained between genders would be interesting to investigate. We received feedback that we should look at more attributes to give a more in-depth analysis and chose life expectancy and net migration.

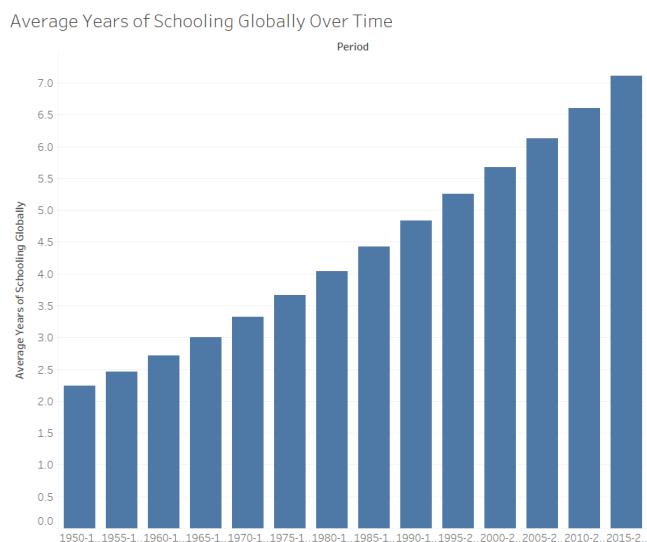
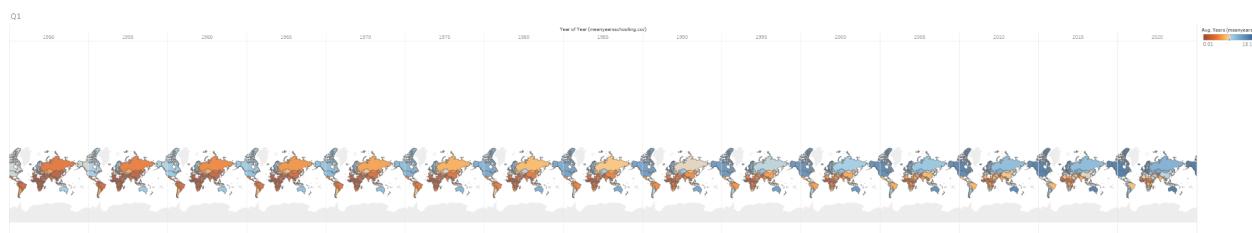
The CSVs provided by the Wittgenstein Centre Human Capital Data Explorer were not set up for us to work with, so we used python to alter them to better fit our needs. First, they came with header information such as source and what data was contained in the CSV that needed to be removed for us to work with it. In addition, there were many attributes, such as the age group of

the humans in the country and gender, which might be useful for future visualizations, but for our purposes, we ended up aggregating. The CSVs also had some inconsistencies with using year or range of years for each data point. CSVs using year were converted to use a range of years going from year to year+5 to make everything consistent. Finally, we needed to add the continents the countries belonged to. At the same time, we added the region as we were unsure if we also wanted to filter by region.

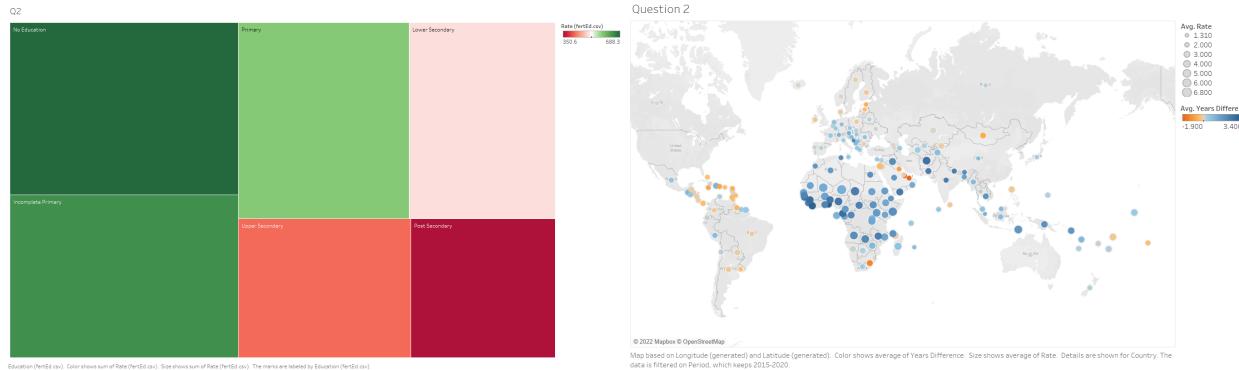
After altering individual CSV files, we had to decide how to deal with visualizations pulling information from multiple files. Initially, we read each CSV we were combining through the javascript files, but this was too slow and took up to a minute to render in all the visualizations. Afterward, we created dedicated CSVs with all the attributes we needed for the visualizations. We considered using a “master CSV” containing all the features for a given country and period of time; however, each visualization would then need to aggregate certain features out, and our run times suffered, so we decided to only use dedicated CSVs for each visualization.

Exploratory Data Analysis

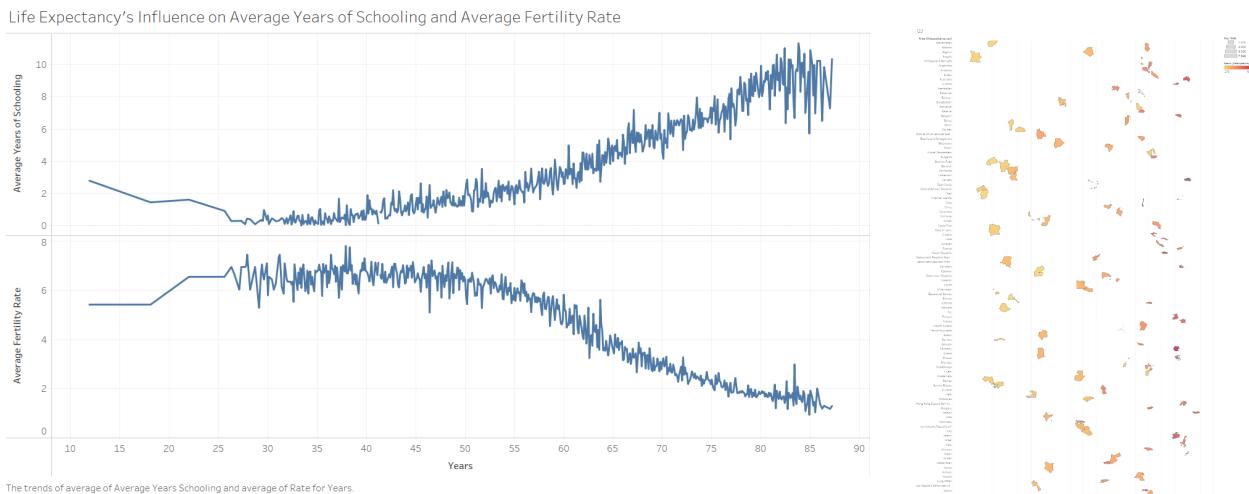
We utilized Tableau for our exploratory data analysis and found that education, time, fertility rate, migration rate, and the interactions between them were the most interesting elements of our dataset. Below are screenshots of our exploratory visualizations.



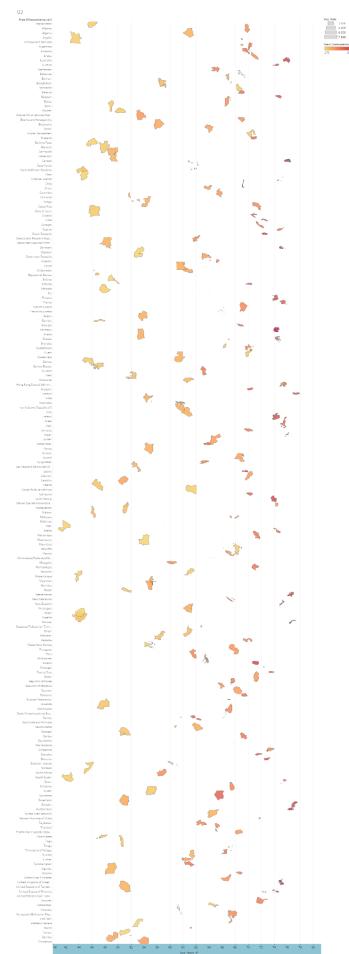
The two visualizations, above and left, answered the question of global education over time, but the first provided much more nuance between countries. We had concerns with the first one due to the amount of space that would be necessary, but we felt that a colormap representation of education and using a geometry would be incredibly effective. Interestingly, the second visualization inspired our time period selector.

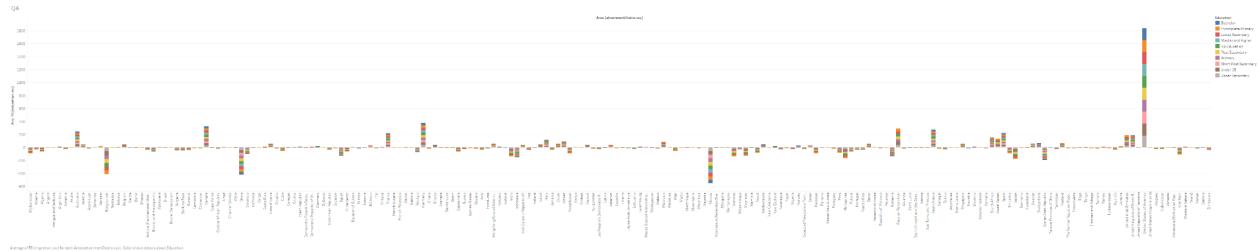


The two above visualizations looked at the relationship between education and fertility rate. While the categorical attribute of education level in terms of no education, primary, secondary, etc., were initially interesting, we found that the gap in education between males and females was far more effective at showing the relationship between education and fertility rate and also provided an opportunity to explore inequality in our visualizations.

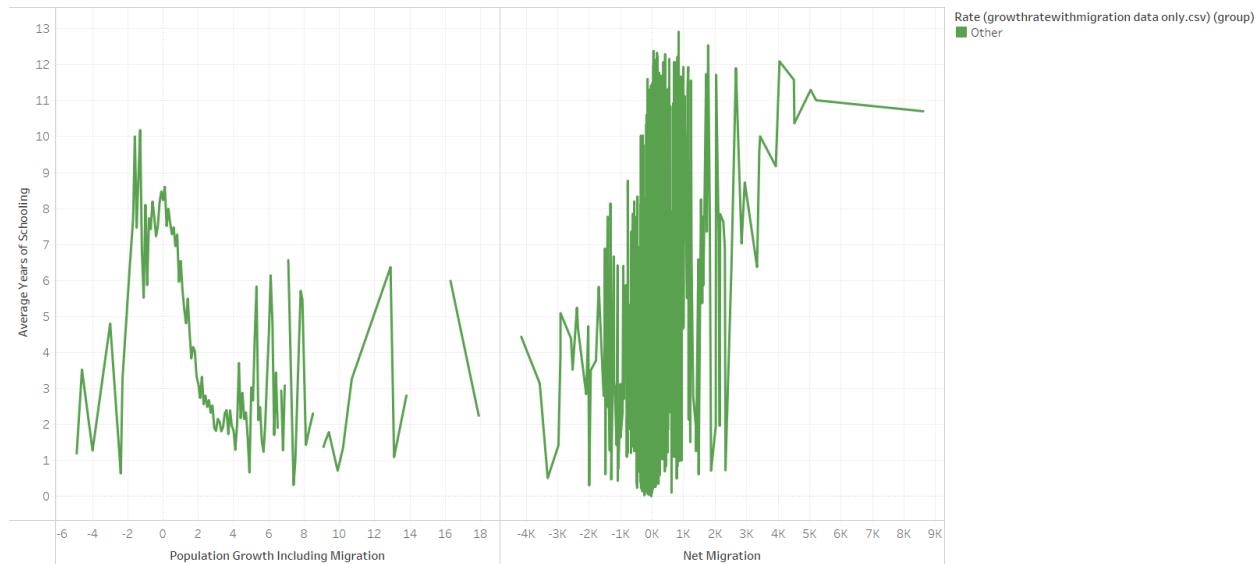


The two visualizations, above and right, started seeing how three attributes, specifically life expectancy, fertility rate, and education, interacted with one another. While the second option was incredibly interesting to us due to our familiarity with the dataset at this point, it was far too complicated to decode as a user and recreate. However, the first ended up being our strongest visualization from this exploratory stage to the final stage of the project and convinced us to move forward with this dataset.





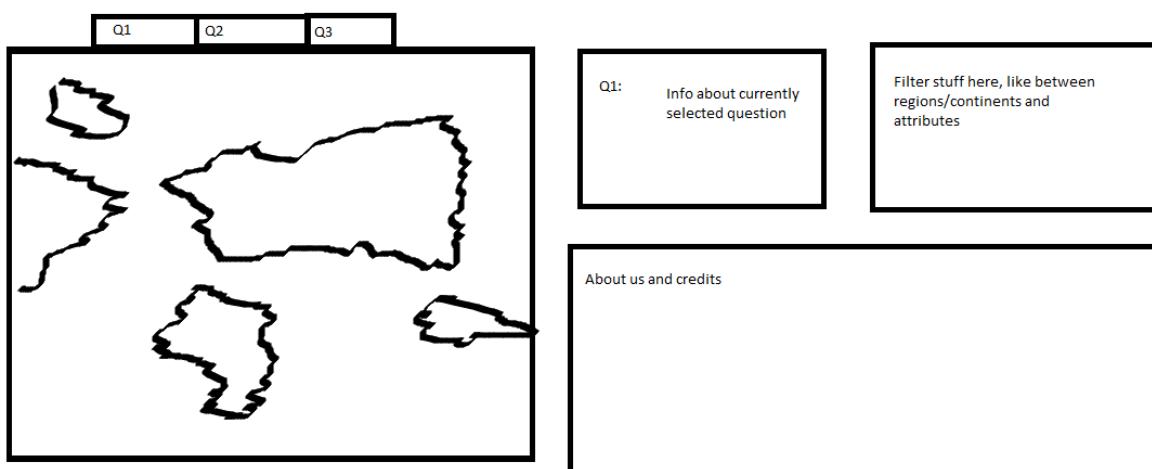
Population Growth and Net Migration's Influence on Average Years of Schooling



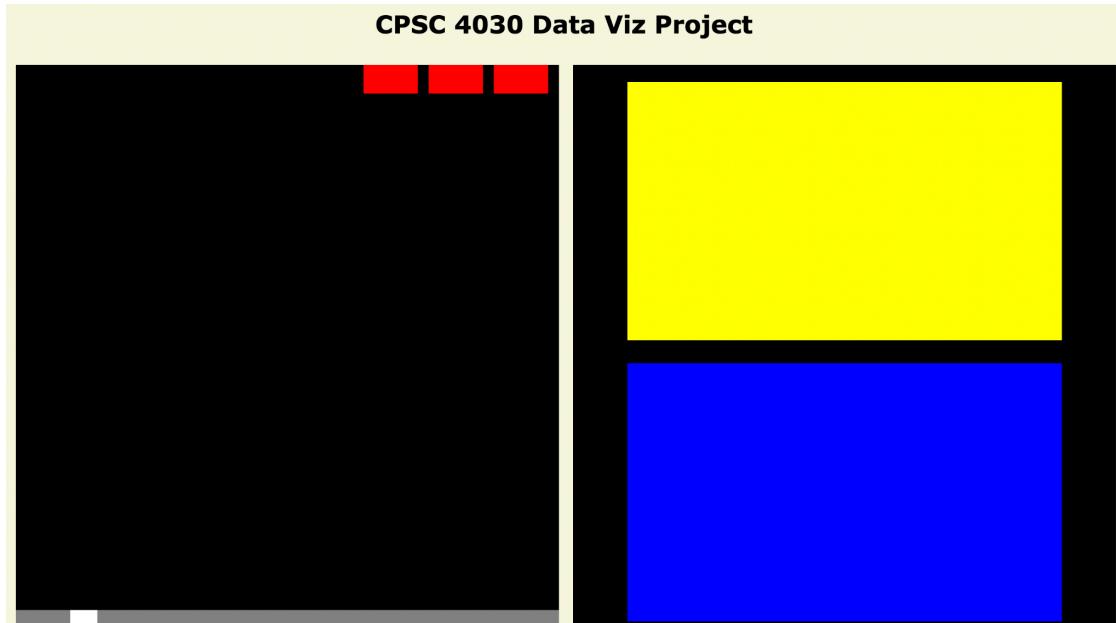
The trends of average of Average Years Schooling for Rate (growthratewithmigration data only.csv) and Net. Color shows details about Rate (growthratewithmigration data only.csv) (group).

The above two visualizations looked at education and migration rate. The first utilized the categorical education levels mentioned earlier, and the second visualization used net migration for each country. At this stage, we thought it would be possible to combine the categorical education level with the succinctness of the second visualization.

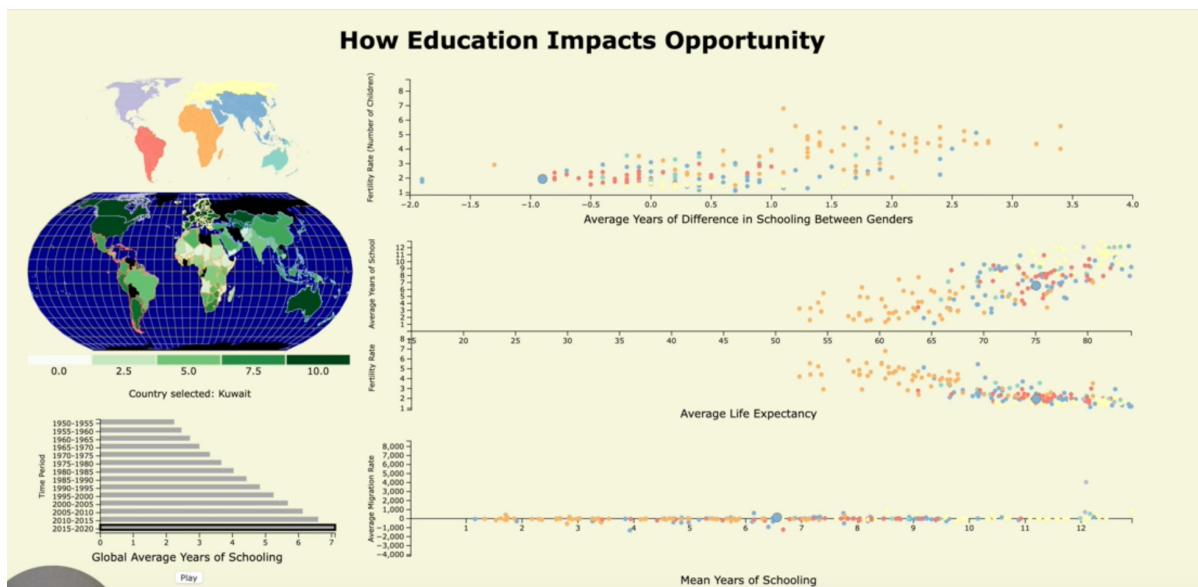
Design Evolution



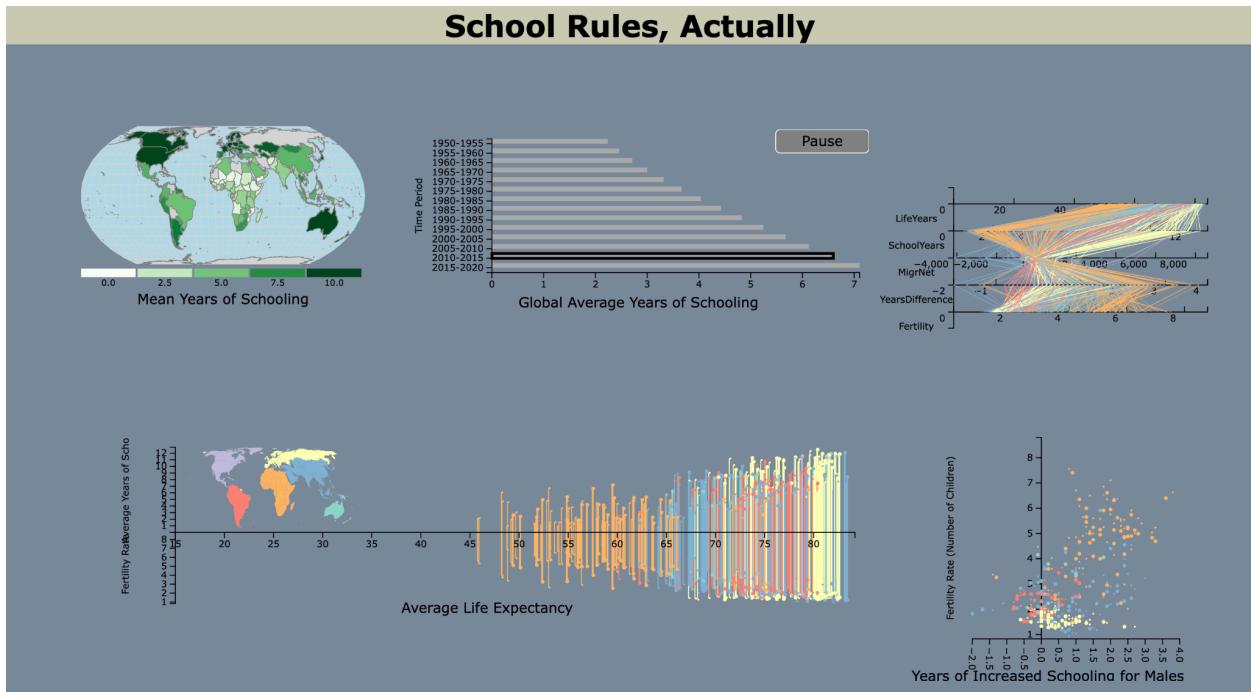
Our initial design had a user-created visualization with filtering and selection to allow for exploration by the users. Our created visualizations were present, but they would have to be selected. This design did not allow for any comparison between the visualizations and relied heavily on the user knowing what they wanted to look for rather than telling a story and allowing for further exploration. Additionally, it relied heavily on text to communicate the information.



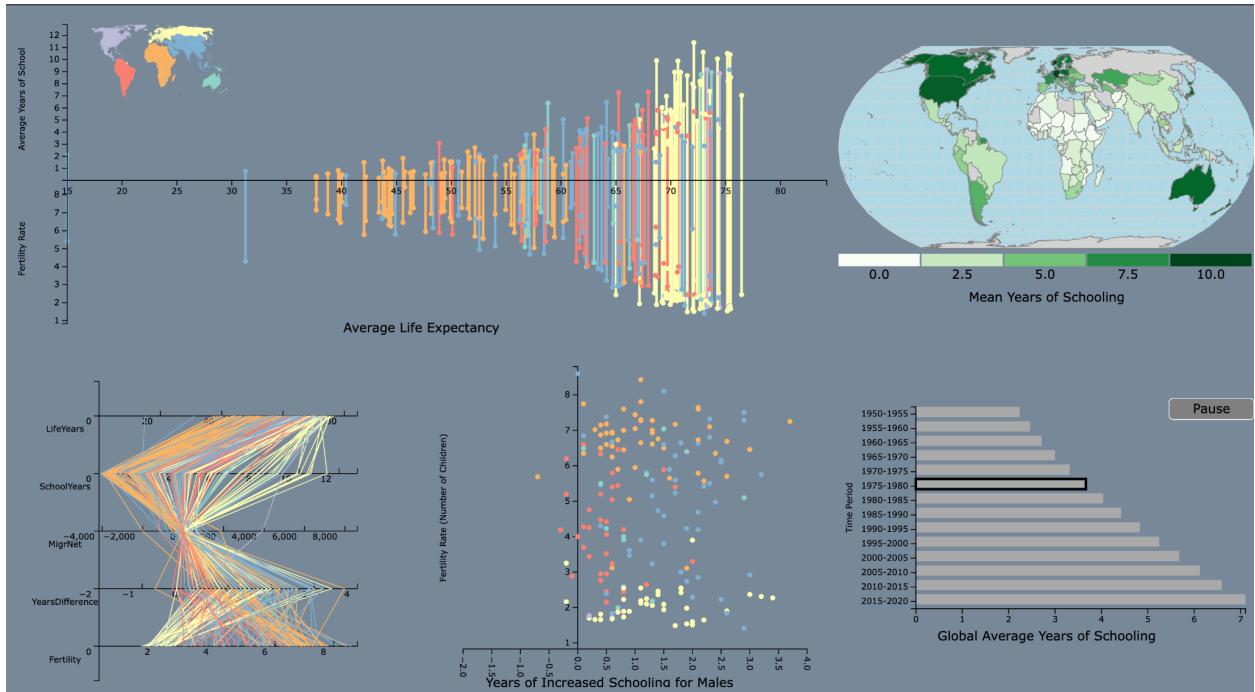
This design had a similar pitfall to the initial proposal, where all of the visualizations were paginated, but attempted to incorporate some of the filtering in the visualization itself. However, the blue rectangle was going to still rely on text to communicate information about what the user was looking at, and the yellow rectangle was going to hold simple filtering options also in the form of text.



This design started to approach the dashboard effect of having everything visible at once, but the scaling and idiom choices were not communicating information effectively. There was too much to decode in the map visualization, which limited the impact of the color map and took away attention from the more powerful visualizations to the right. Additionally, the visualizations that were more for the user interface and as a color legend were more prominently placed than the visualizations themselves. Finally, the choice to use all scatter plot idioms with no distinction between scaling or marks made it hard to see the differences between the visualizations and the message of each individual visualization, and the story as a whole was lost at this stage.



This stage addressed the concerns over using the same idioms for every visualization to make each visualization more visually unique and decrease the conflicts between them. The colors also contrast significantly less to where no individual visualization dominates the overall visualization. However, the layout at this stage had the user noticing the interface visualizations first rather than the visualizations that conveyed information.

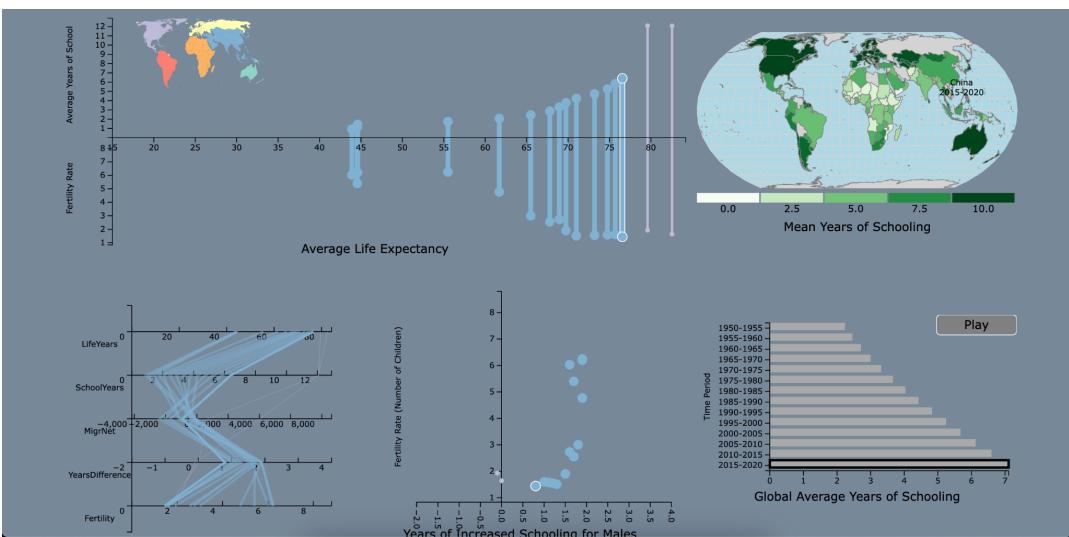
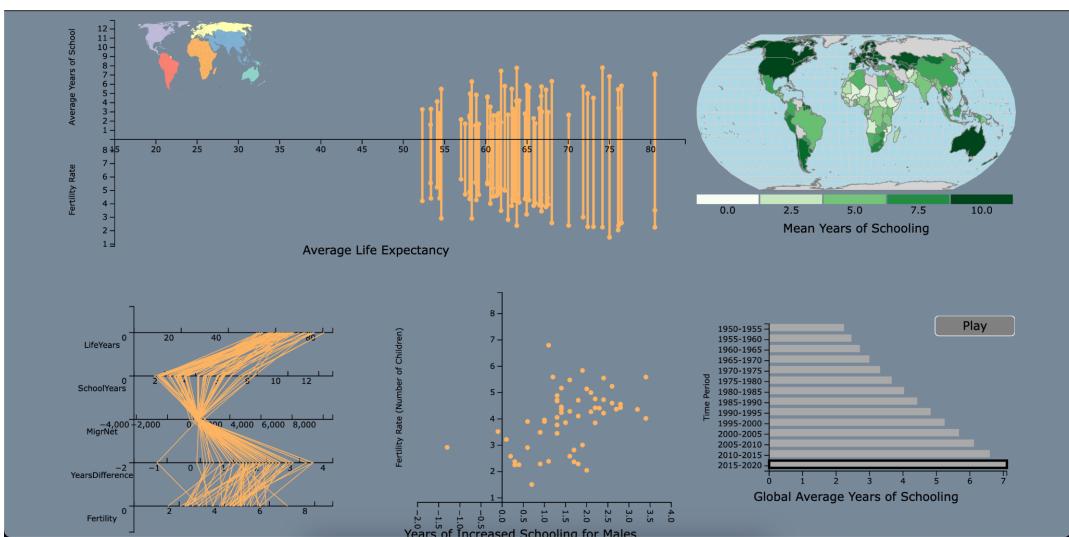
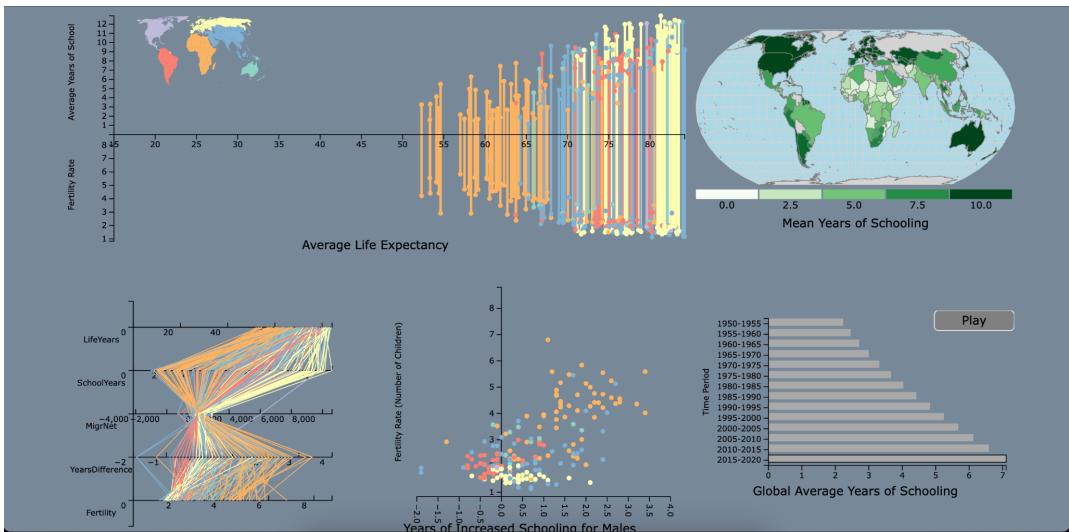


The above design is what we ended up landing on as our final design for the overall visualization. This design follows the principles from the CXL article by highlighting our strongest visualization and immediately introducing the user to the continent color map. It also provides a clear grouping of our interactive visualizations and the user interface/filtering visualizations.

Implementation

Filtering

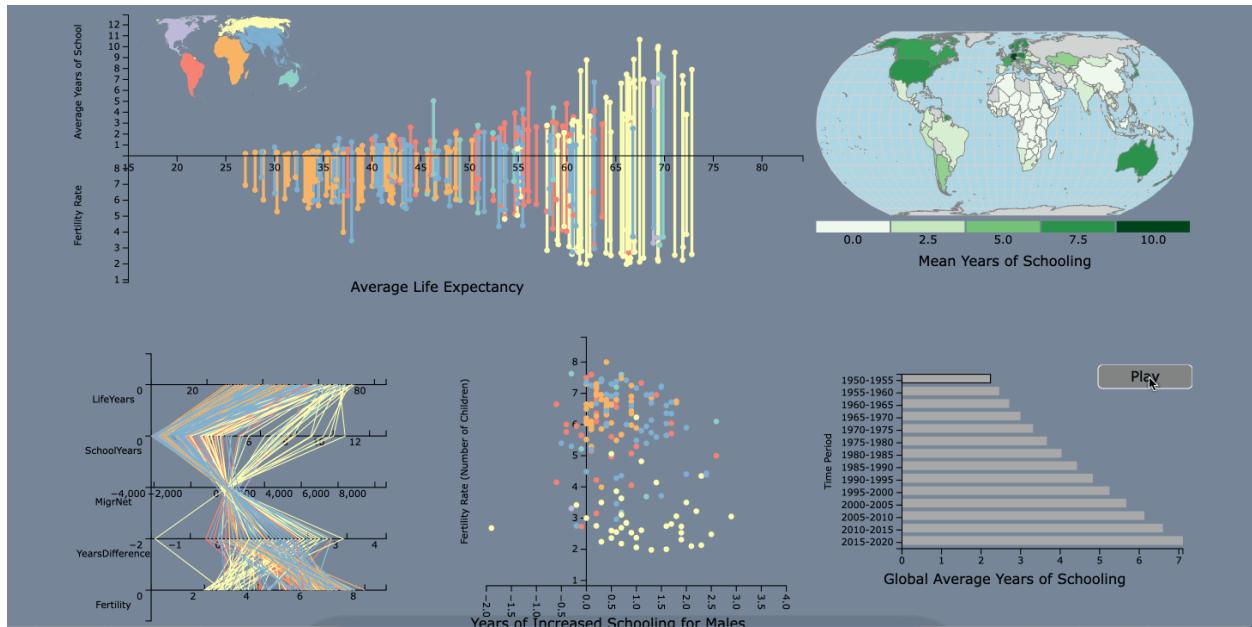
The initial interaction we set out to have was to allow the user to filter every visualization between seeing all of the world's countries, an individual continent, or just a single country. This was expanded to provide the ability to either hover over a particular continent or country and have that selection be popped out in all of the visualizations or click a specific continent and only visualize that continent.



This allows the user to explore the visualization system at various levels of abstraction, which enables the user to create their own paths through the story and make their own discoveries rather than what we, as the creators of the visualization, found interesting.

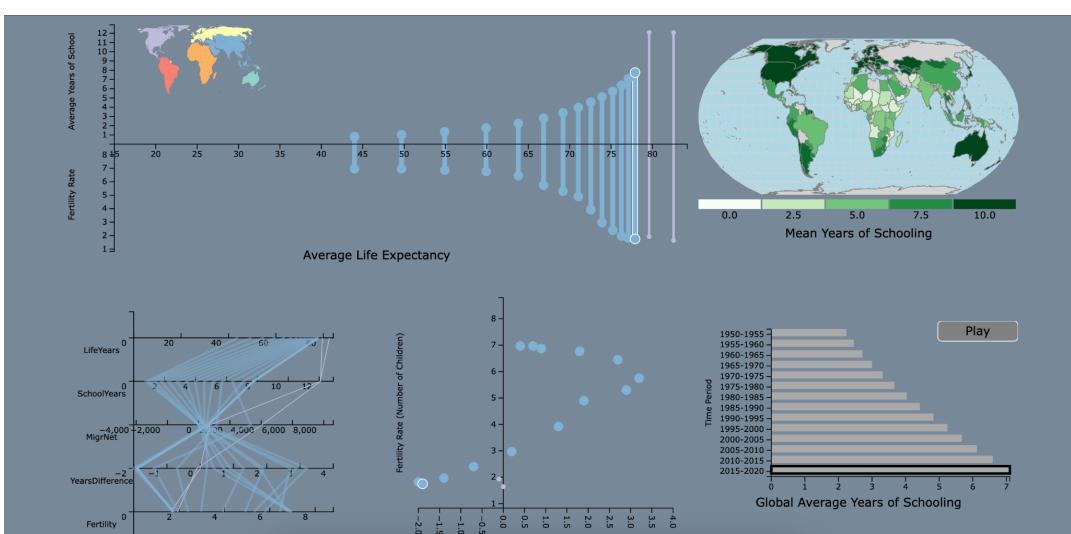
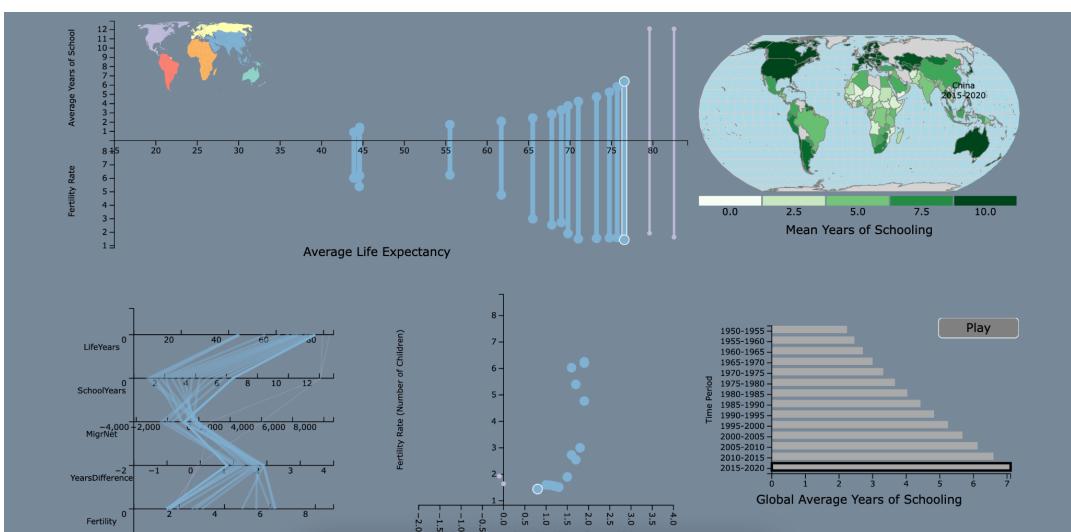
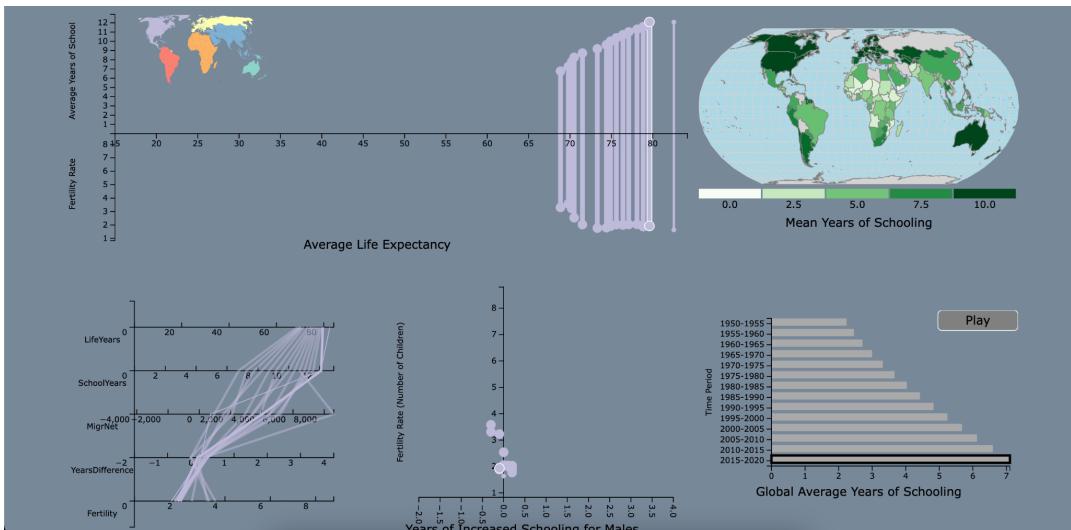
Animation

Our website begins as an animation for the entire page rather than as a static visualization at a time period that we selected. This animation communicates the change over time element of our dataset that we found incredibly interesting and allows the user to enjoy the dataset as a whole initially and then decide if there is a particular time period they would like to explore further.



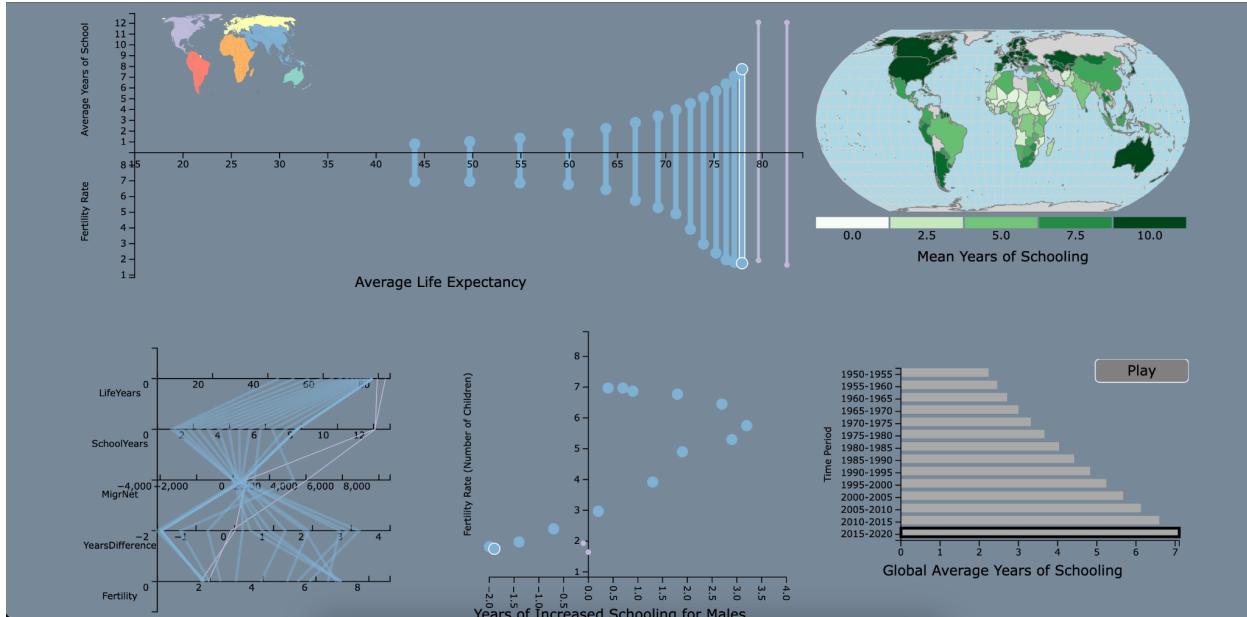
Comparison

The overall visualization was designed to visualize one time period at a time. However, that design limited the user's ability to compare anything past general trends for the world, and even that was limited by the amount of change happening on the website. This led us to incorporate the ability for the user to click on a country from any of the visualizations and visualize all of the entries for that country from every time period and allow for direct comparison of that country over time.

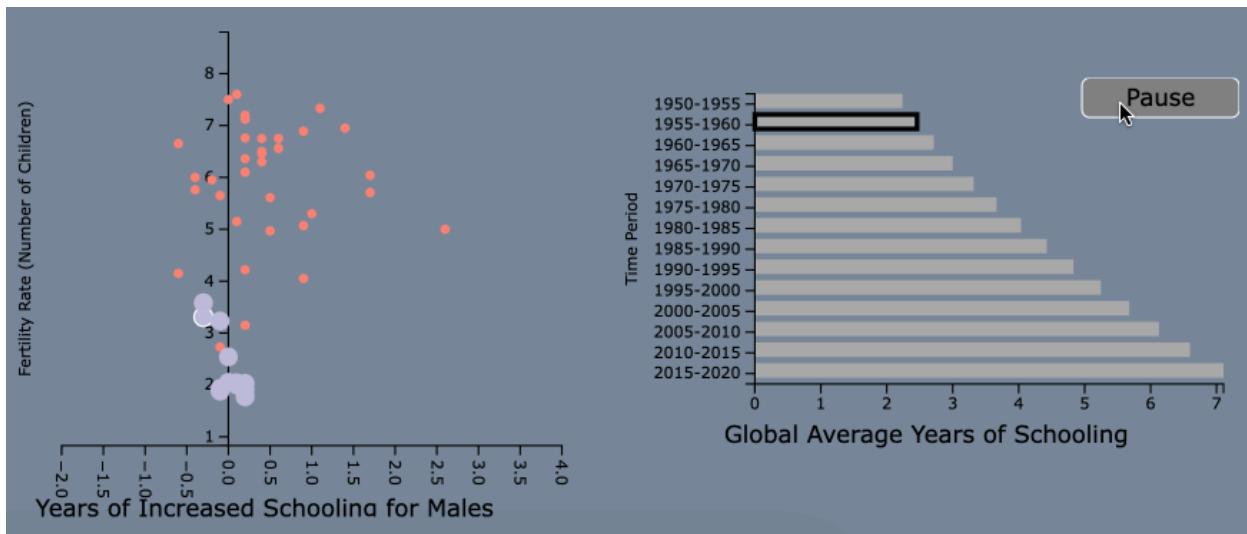


Evaluation

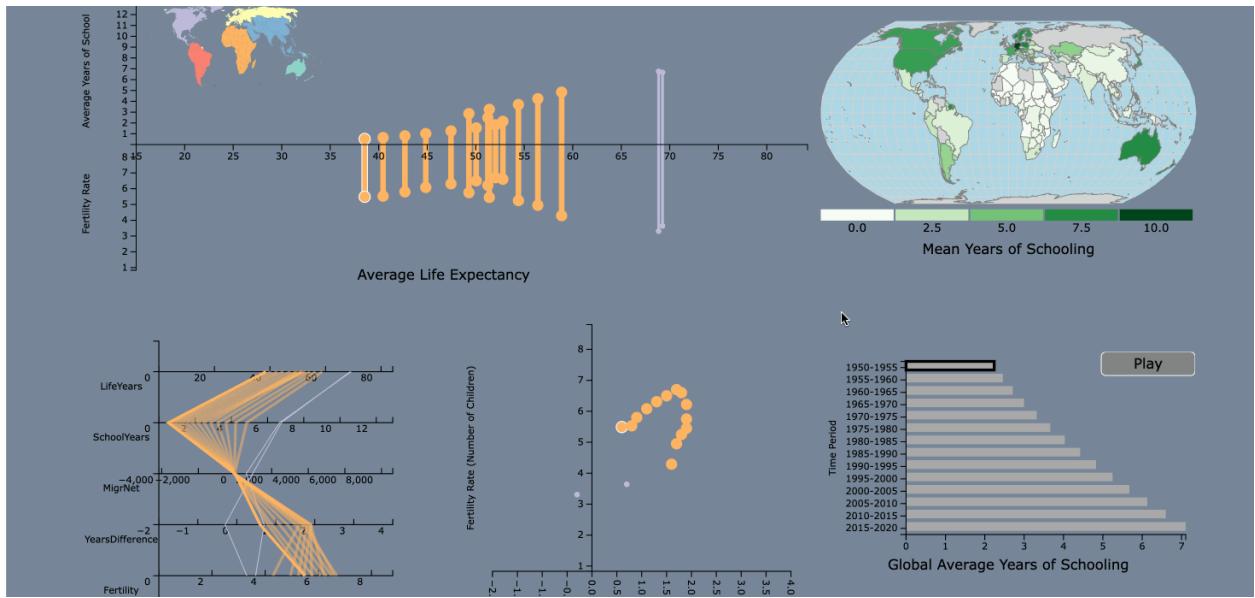
Many preconceived notions from an American-centric view of the world were proven false from our time spent with this visualization system. For example, the United Arab Emirates tends to have more educated females than males,



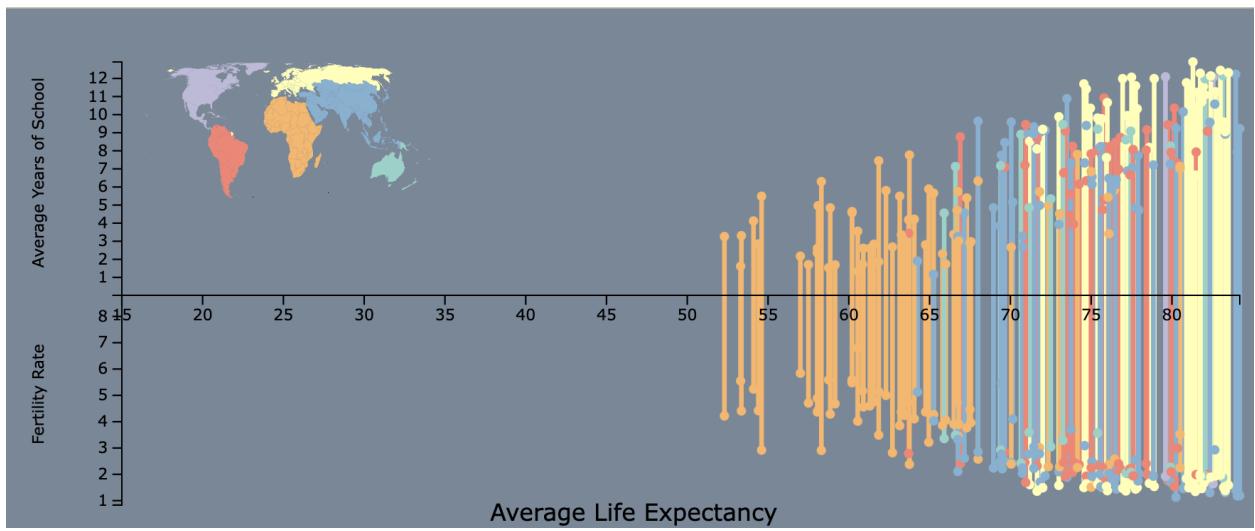
and the U.S. has moved between having more educated males and females throughout history.



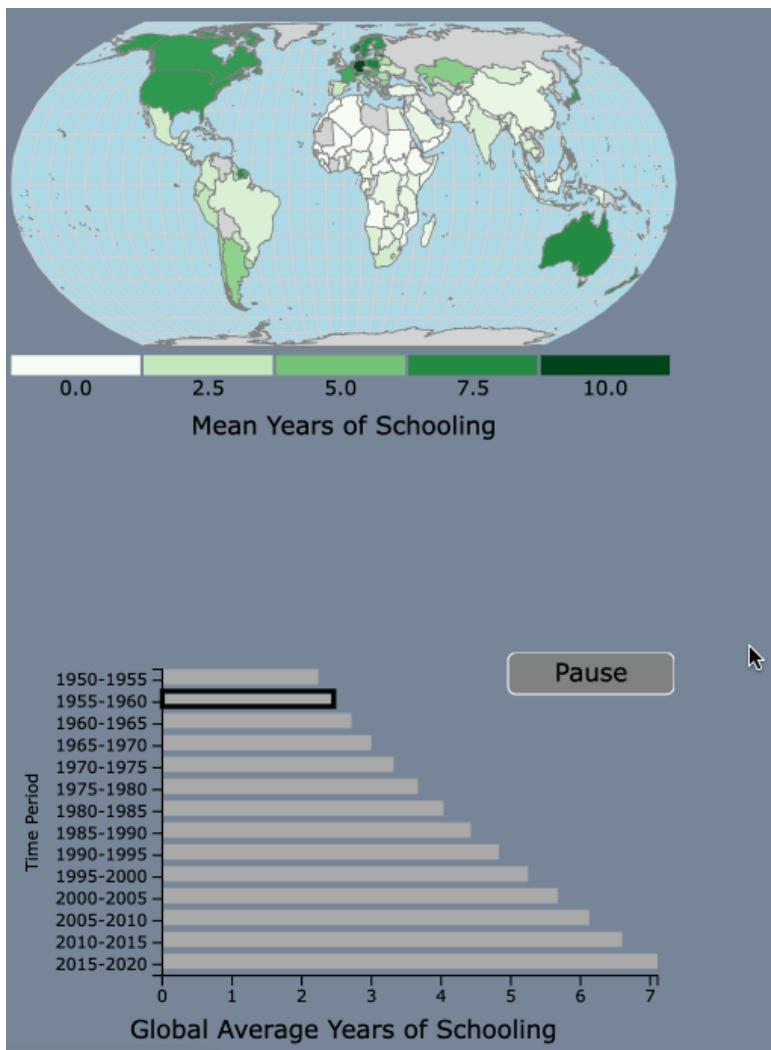
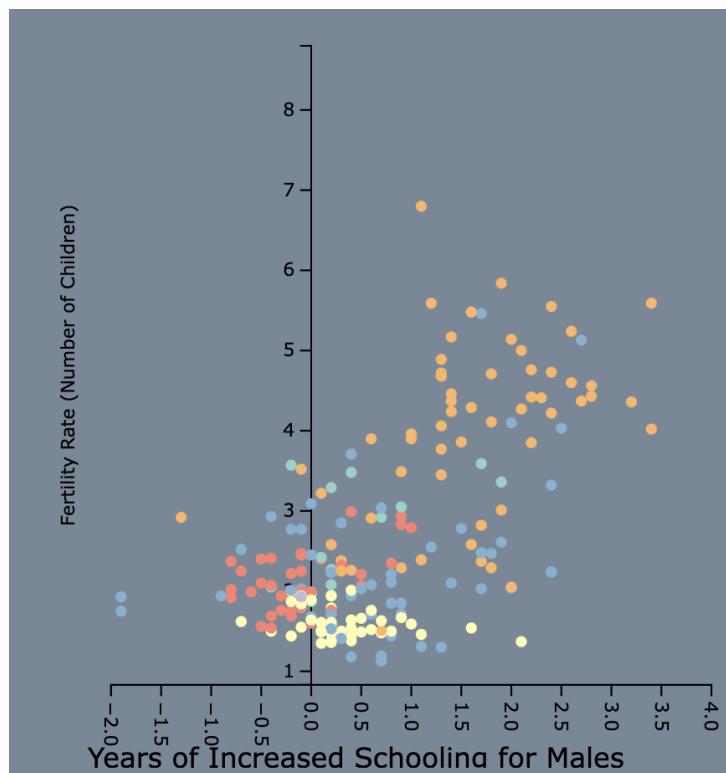
Additionally, not all countries have a linear progression where there is less of a gender gap in education, higher education overall, longer life expectancy, and lower fertility rate as we approach the present.



Longer life expectancy does tend to correlate with an increase of education and decrease fertility rate, communicating that as overall conditions improve and allow for longer lives, more of a priority is placed on education.

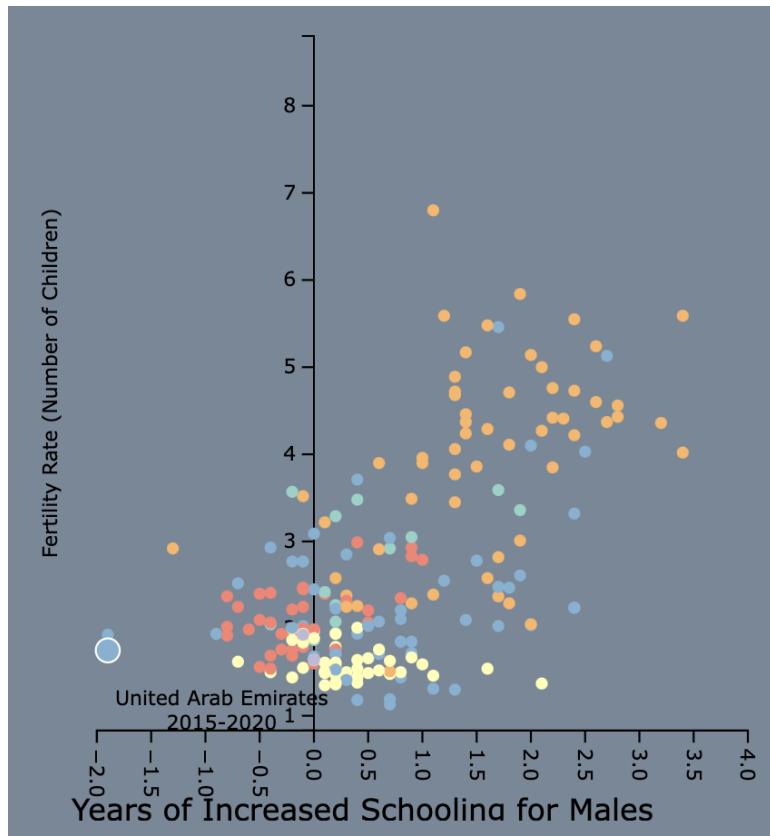


Countries with males that are slightly more educated than females do not generally have a strong relationship between fertility rate and the gender gap in education. However, the extremes of those cases tend to see high fertility rates for areas where females are less educated, and areas with high rates of education for females have lower fertility rates.

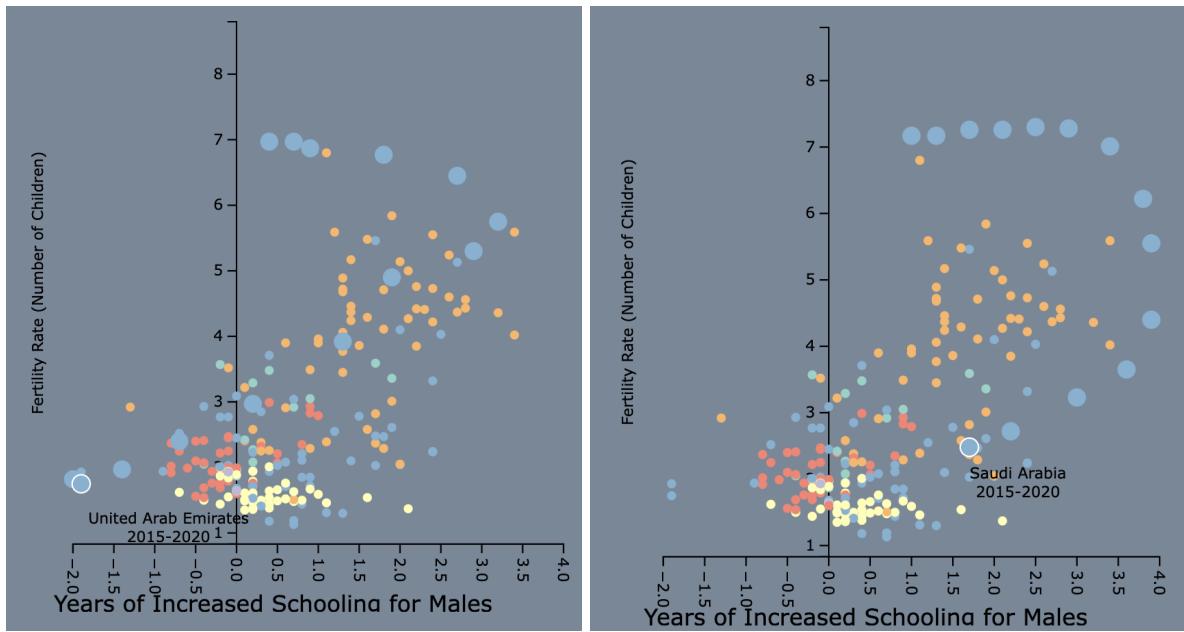


When looking at the change in education over time (left), the geopolitical map with the color scale effectively communicates this when viewing it with the animation running.

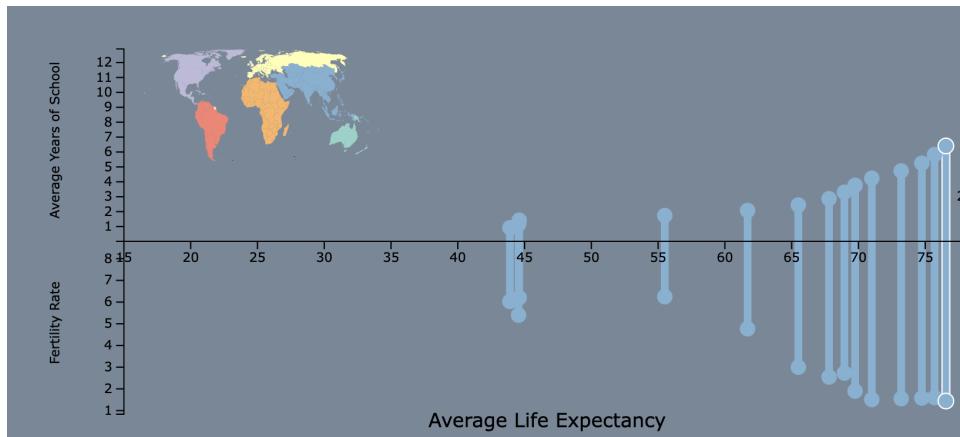
The scatter plot looks at the relationship between fertility rate and the gender gap in education by using horizontal and vertical positioning.



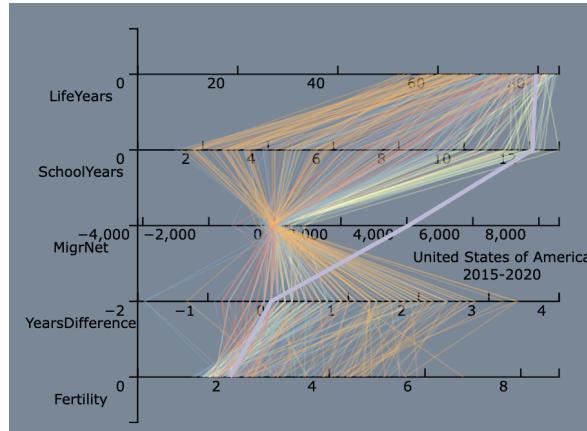
It is made stronger through filtering to look at a particular country over time.



The relationship between life expectancy, fertility rate, and education uses distance as a powerful channel to communicate that as life expectancy increases, the distance between fertility rate and education increases. This growth shows much higher education and much lower education. Countries with lower life expectancies tend to have much less of a distance between these attributes, leading to the conclusion that longer lives lead to a higher priority on education and a lower emphasis on reproduction.



The inclusion of migration on the parallel axis chart communicates that while most countries tend to have very low migration in or out of that country, there are notable exceptions.



Our visualization effectively analyzes the relationship between education and other attributes of life, such as fertility rate and life expectancy, and it even begins to look at measures of inequality through the lens of the gender gap in education and the limitations of a shorter life. We achieved our must-have features of comparison over time, filtering between country and continent and having everything visualized at once while still having a digestible visualization system. We were able to reach our stretch goals of having a geopolitical map, comparing a country over time on the same visualization, and being able to highlight a country in one visualization and have it pop out in all of the other visualizations. Further improvements to this visualization would be incorporating more attributes that could serve as measures of inequality or opportunity and allowing the user to select which attributes they would like to see a direct comparison of would be powerful for exploration.