Team 2
Anjali Sebastian
Ulysses Lin
Zachary Jeffreys
Kruta Prabhu
Date: 6/9/2022
CPSC 5320-01

Final Project Write Up & Reflection

1) A summary of the suggestions from the class from the design feedback session

During the design feedback session, we were still in the planning stage of our project. During this session, we were still deciding between Friends TV series sentiment analysis and dance movement analysis. From the design feedback, we learned that motion graphics data could be interesting. But we felt that we lacked consistent data and advanced processing techniques for capturing the movement details. Following this our team had another brainstorming session and we decided to work on showing the gender gap in the workforce and education system. From the Final_Design_v2.pdf submission, we were told that "You must consider a broader set of existing visualizations that have visualized the gender gap. There are many visualizations and dashboards on this topic, and you should be able to state how your visualization is different from existing ones." This was a great insight because this is a very heavily studied topic, so we needed to ensure ours was novel and different. Another very key suggestion received was that our topic should be more focused, and we incorporated this suggestion by looking at one specific aspect of the data which is the gap in the workforce participation rates.

2) A rigorous rationale for your design decisions including visualization type, size, color, scale, and other visual elements. Document the visual encodings and data transformations you used and why they are appropriate for answering your questions. Also include why you did or did not choose the designs from your design submission for the implementation.

For our design, we wanted to show women's participation in the workforce and the proportion of women with higher education around the world, so some sort of map was our first thought. We wanted more creative freedom than we could achieve with Tableau, so we decided to move towards d3's library. For the map size, we wanted it to be the largest object on the screen, so we wanted most of the screen to be the map, followed by a smaller subsection to show each country's data. Last, we wanted a slider on the bottom where the user can adjust the year. This gives a more interactive and dynamic experience for the audience, who can see changing over the timeframe of three decades for which data was available. We felt that instead of using a choropleth map as seen in many of the related works we felt a Dot Distribution Map would better represent temporal workforce data through the growth of various dots over time. For the color, we decided to overlay four colors of dots and the same colors on the donut charts onto a plain white and grey map. Initially, we used four different hues, but it was difficult to see the distinct colors and the idea that educated women in the workforce are a subset of all the women working in the workforce was not coming through. To show this aspect we used two hues to represent women-men and two value levels to represent education levels. We felt that the dot density style of diagram was visually striking and added the extra bonus of showing how many (in absolute numbers) people of the four groups there were over time. Note that the waffle utilizes a distinctly different color pallet as compared to the map- this would be to distinguish a different dataset, that of employment for women and men by sector. As mentioned after the presentation, we could have made employment by sector something more easily comparable over time instead of needing to hover over a country again after changing the year, but we felt adding say, another set of donuts or dots would take up sparse real estate and that education levels were the priority. We decided to give more real estate space to the linear regression showing the forecast for each country and when a country might reach equity levels in work force participation rates. This was done by minimizing the map and creating the waffle chart as a tool tooltip. Additionally, we decided to not do the forecasting on a waterfall chart as it would be more cluttered, and the waterfall chart does not represent

confidence interval as effectively. The cleaner line chart with simple lines demarcating the past and future with a gradient color(lighter for the future that is unknown) for the line was much more visually appealing.

In terms of data, first, we had to import our data from The World Bank, and we needed to identify the columns/series we are interested in. This was a significant task since The World bank aggregates around 4000 different data series. Even for workforce participation rates there were multiple data series, and we chose the ones that showed participation rate on the population of adults between 15 to 64 years of age. The needed processing since it was in a format that was exceedingly difficult for D3 to directly render. We did multiple transformations (e.g., pivoting, unpivoting) and calculations (e.g., converting percentage to head count, calculating ratio values). There were also various null values for years for certain countries. These values were changed to zero or we took the last year with valid data (see point 4 for more details). To create the trend line, we originally wanted to use a regression line to predict the gender gaps, but we found that Tableau's forecast module was a better option since we are dealing with time series data. The forecast data including 95% confidence interval was downloaded from tableau and then rendered in D3 using a simple line chart.

We decided to choose one of our designs from our v2 submission which was the one inspired by Nadieh Bremer. The minimalistic style of this visualization was perfect for what we wanted to build. As for the second visualization of the dumbbell chart, we decided that this graph would be too busy for the data we wanted to display. If we were to use a dumbbell chart with over 6 countries it would be very crowded and hard to make meaning out of. This graph also introduces bi-variate analysis that would look at what factors affect workforce participation rates and we decided to be more focused on our topic and not include more data on fertility rate, etc. Instead, we decided to create a unique waffle chart to show the different sectors women are working in, which was simple, intuitive, and could be done for a given year. A waterfall (which considers time) per country would require competing for space on the right with the line graph. We also decided to add a donut chart option to be displayed for each country. We made this choice as an alternative to the dot density because it can more clearly distinguish the proportions of the four categories above, though we sacrifice absolute counts of people.

3) images of your design iterations (e.g., sketches, initial design or implementation, final implementation)

These were our hand-drawn visualizations for our first set of ideas. Note that these were on dance and Friends.
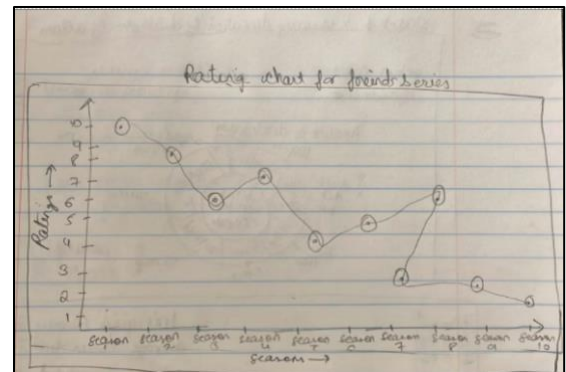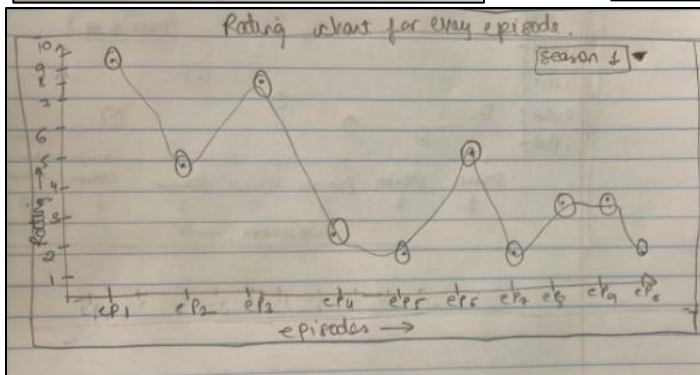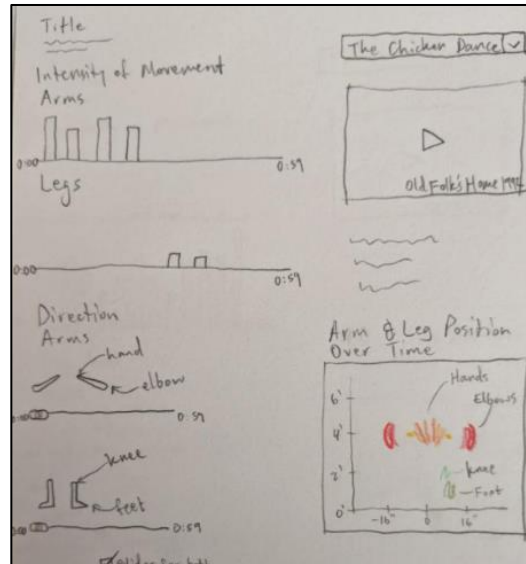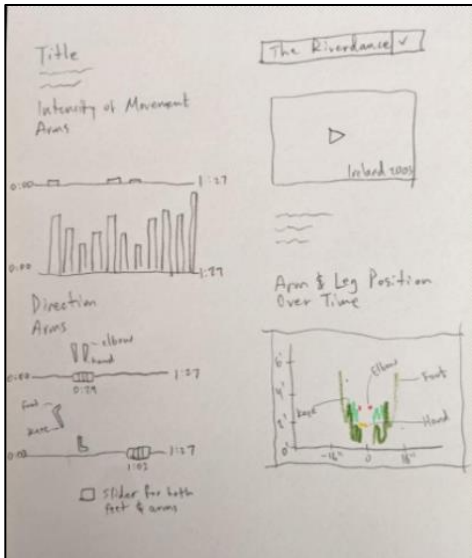
After we decided to change our idea to show a visualization of the gender gap in South and East Asia, this was our original drawing. Like our final visualization, this one shows a map of the region where the user can have mouse interactions with the countries. We also included the waffle chart pop-up and a chart on the right where we wanted to display some sort of regression to try and predict future gender gaps.
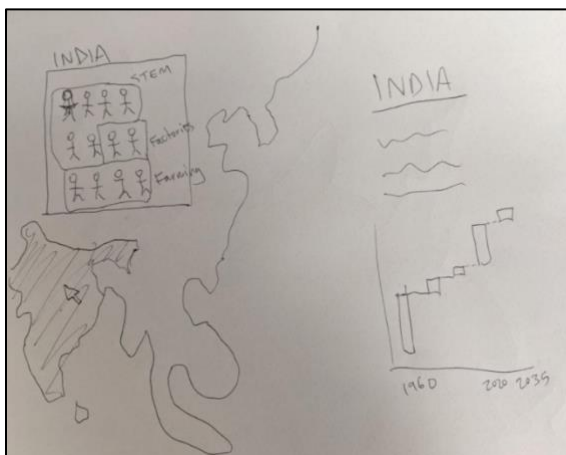
After determining that we wanted to use the d3 library, we had to find the .geojson coordinates for each of the countries that we were interested in visualizing and include smaller countries that were not in the initial
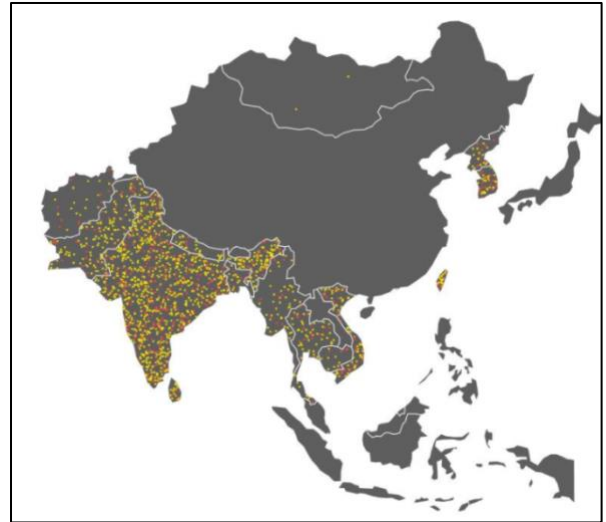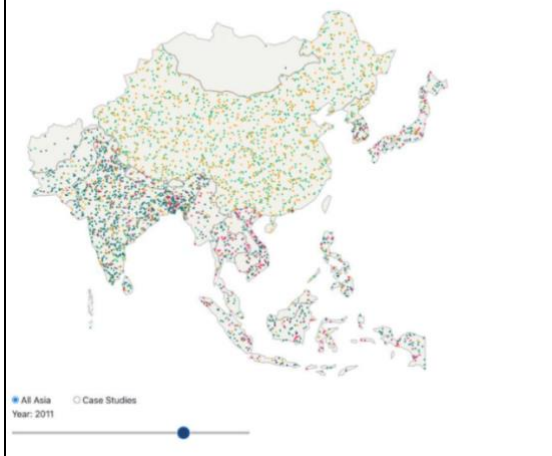
coordinate set. Here is an image of the map before data points were added (choropleth using unrelated population data) and then an image of the map after some of the dots were added. Then the third picture shows our final dot density map before we added the donut charts, and the fourth image is after adding the donut charts, but before creating a more harmonious color scheme. The fifth is the final iteration with better colors (with donuts matching those colors).
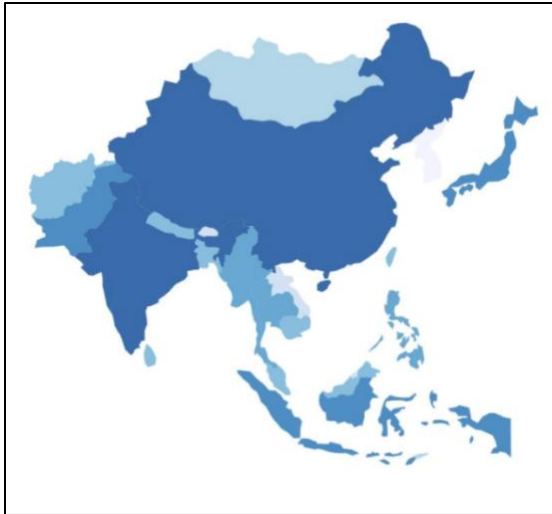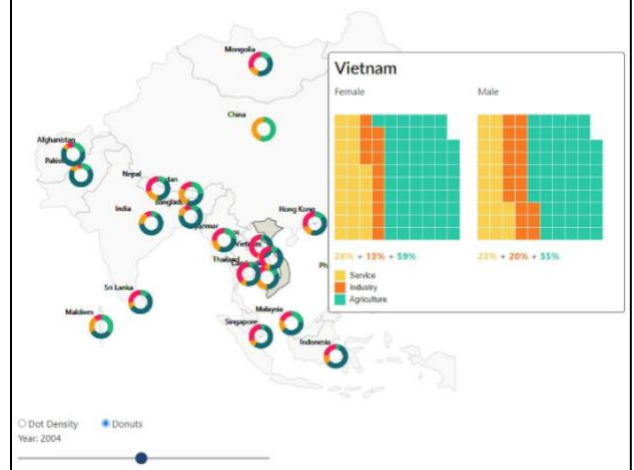
Labor Participation and Education of Women in Asia
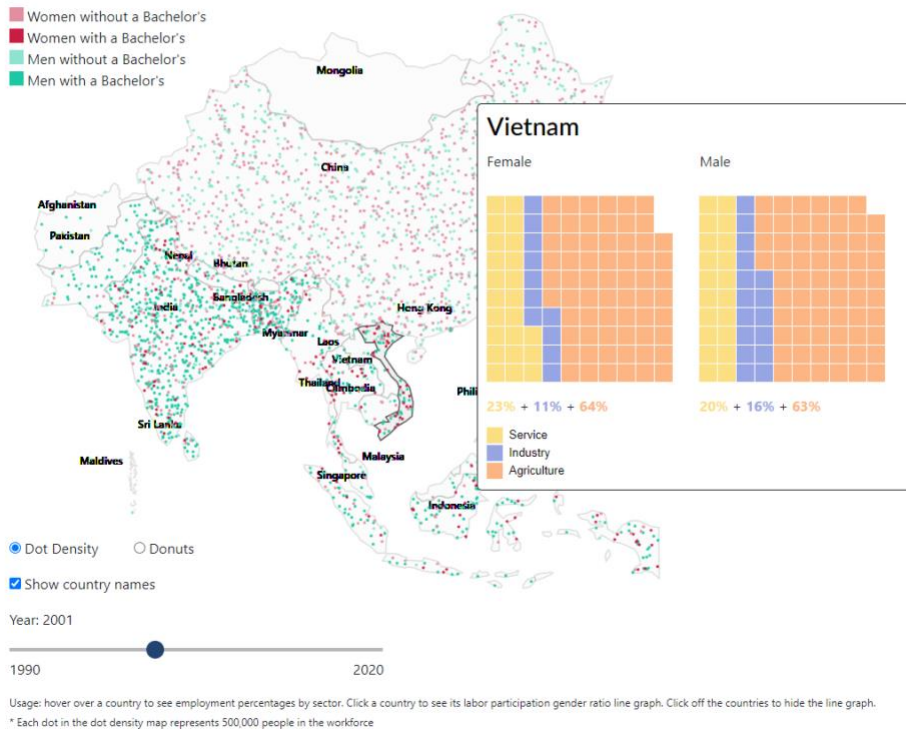Anjali Sebastian, Kruta Arvind Prabhu, Ulysses Lin, Zachary Jeffreys

All Asia

● All Asia   ○ Case Studies
Year: 2011

Donuts

Vietnam
Female   Male

28% + 13% + 59%   23% + 20% + 55%

■ Service
■ Industry
■ Agriculture

○ Dot Density   ● Donuts
Year: 2004

Gender Gap in the Workforce in East and South Asia

Anjali Sebastian, Kruta Arvind Prabhu, Ulysses Lin, Zachary Jeffreys

Dot Density

Here are some pictures of the waffle chart. Our original idea was to have icons representing women in different sectors such as STEM, factories, and farming. Our second iteration was the final waffle chart before the color scheme was updated, and before it was added to the map. The third image shows the waffle chart with service, industry, and agriculture sectors. The final waffle chart visualization with better colors (see above image) also has a percentage to give the audience a clearer understanding of the proportions of the gender gap across sectors.
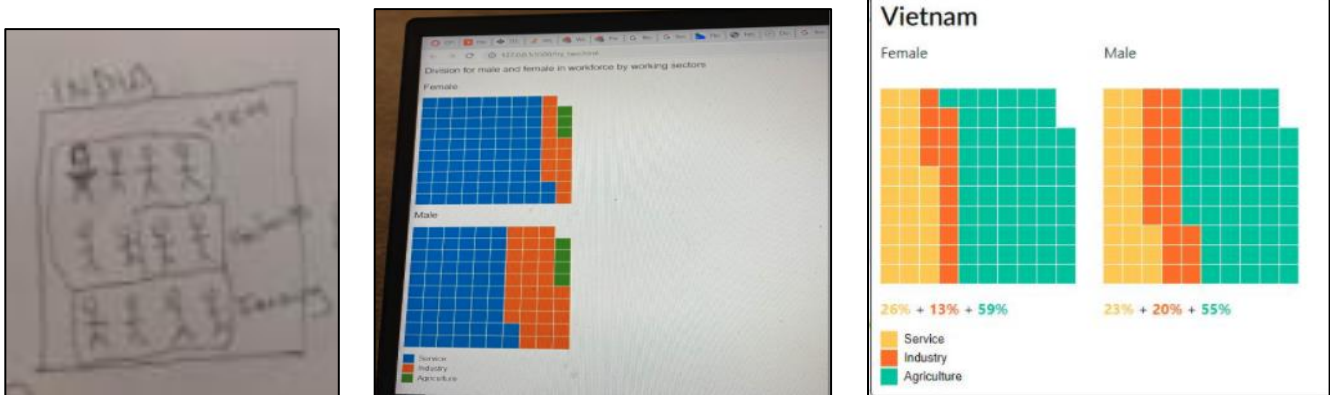
Here are some additional design ideas that we decided not to use. The first one shows a dot density map over the whole world, but we determined that such a "zoomed out" view would hurt visibility of any given country's dots – there would either be massive crowding or we would need less dots, making it harder to compare proportions

accurately, so we decided to focus on a specific region. The second image shows a regression to predict when gender ratios between women and men would be equal, but we decided to use a trend line instead. Last is a dumbbell lollipop chart that we wanted to use but showing more than six countries would have been cluttered.

Storyboard 1: Dot Distribution Map: Showing Literacy and Workforce participation

* We want to show this Dot Distribution Map connected over time
* User click on any country to see specific details for that Country
* On user hover → show tooltip with General Information
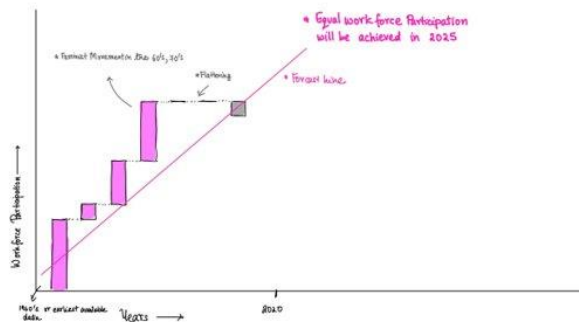

* 1 dot = X number of people
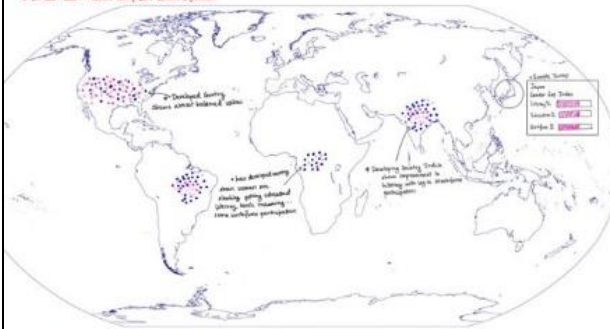
* Women who are literate
* Women who are part of workforce
* Men who are literate
* Men who are part of workforce

Storyboard 2: Changing Women's Involvement in Workforce

* Waterfall chart + Regression line to find when equality is expected (note some countries may have already reached or will work even when achieved)
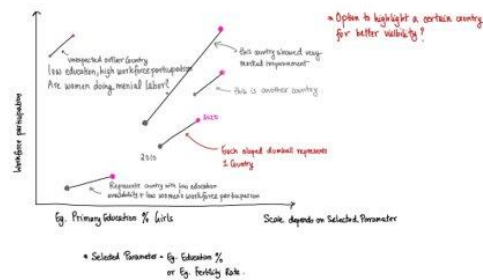
▼ United States   * Country filtered through map selection

* Equal work force Participation will be achieved in 2025

* Feminist Movement in the 60's, 70's
* Flattening
* Forecast line

Workforce Participation

1960's or earliest available data   Years →   2020

Storyboard 3: Factors Affecting women's ability to participate in Workforce

* Select a region/some set of countries
* Select a parameter to see its relation to workforce participation   ▼ Primary Education Girls %

* Option to highlight a certain country for better visibility?

* unexpected outlier Country low education, high workforce participation Are women doing menial labor?
* this country showed very marked improvement
* this is another country
* 2025
* Each sloped dumbbell represents 1 Country
* 2015
* Represents country with low education availability + low women's workforce participation

Eg. Primary Education % Girls   Scale depends on Selected Parameter

* Selected Parameter - Eg. Education % or Eg. Fertility Rate.

* Inspiration for this chart is from Nadieh Bremer's "Visualising Millions of People on the Move"
* However we wanted to show the change in women's workforce participation with respect to their change in circumstance (each of the parameter). Here the change is modelled as movement.

4) the most challenging aspect of the project. e.g., Were there any ethical decisions that you had to make? Any tradeoffs you faced?

The World Bank dataset was incredibly vast in its coverage, even within the realm of gender parity, education, and work; it covered different age ranges, particular trades, and more. As such, we inevitably are sacrificing a plethora of analyses and conclusions in the name of focusing on advanced education, the workforce, and the gender ratio of participation. For example, our tooltip waffle data only shows employment by sector (service, industry, agriculture) as percentages – an unlikely but technically situation is female employment in services has taken a bigger proportion of jobs lately in a country, but the absolute number of women in services for that country has decreased. We are not providing easily accessible data on workforce populations (absolute counts) nor people unemployed; we do provide a dot density diagram and a user could count the dots by color, but this could be difficult and only an estimate and the intent of the dot density diagram was more for comparing proportions and for visual impact/flavor. Our focuses for this project also bring up some biases as well; we are implicitly defining getting advanced education and a service sector job as ideal for women. Indeed, we present education as ordinal data, marking advanced education with stronger color value, and we talk about service employment percentage increases as a sign of progress (as an aside again: we chose not to depict absolute increases in the number of women employed in sectors). We do not talk about increases in industry or agriculture as positively. These are not universally held viewpoints. Also, some of the conclusions our visualizations display may give faulty views. For example, the line graph for Mongolia looks unpromising – equity in labor force participation looks worse in the future. But if you look at the donuts for Mongolia over time, education rates for women (and men) have generally improved. Another design choice we made was if there was missing data for a given year, we would typically repeat the data from the last year we had data on for a country; this is not entirely desirable, but we did this to not leave holes in the visualization. The dot density diagram, by nature of being dependent on workforce population size and geographical size, inherently disadvantages understanding less-populous countries or micronations (ex: Mongolia, Singapore, Hong Kong).

Other challenges we faced were the extensive data-wrangling needed, writing up the case studies and detailed documents for presentation, combining statistics to narrow down the fields to our choices, and the usage of d3.js. Converting a .geojson file of the whole world into just our nations, calculating locations of donuts and labels based on country location, and populating "multi polygons" with dots based on categorical ratios was time-consuming and required understanding complex algorithms.