Final Submission

Jaren Ong

2023-11-15

Week 9

What is the topic that you have finalized?

This project aims to assess gender inequality in India by analysing labor and educational statistics over time.

What are the data sources that you have curated so far?

I collected secondary data from the CEIC website, which is a website which compiles data sources from a myriad of reliable data sources such as the World Bank, the International Monetary Fund, data.gov.in and more.

Week 10

What is the question that you are going to answer? (Answer: One sentence that ends with a question mark that could act like the title of your data story),

Is gender equality getting worse in India?

Why is this an important question? (Answer: 3 sentences, each of which has some evidence, e.g., "According to the United Nations. . . " to justify why the question you have chosen is important),

The question of whether gender equality is getting worse in India is significant due to several compelling reasons. According to the United Nations' Human Development Index, India's Gender Inequality Index (GII) ranks it as one of the countries with persistent gender disparities. India's GII, according to the 2020 Human Development Report, stood at 0.501, indicating substantial gender inequality. Furthermore, evidence from the National Family Health Survey-4 (NFHS-4) shows that despite improvements in some areas, India still grapples with gender-based issues like maternal mortality, child marriage, and violence against women, underscoring the need to assess whether the situation is improving or deteriorating. Lastly, tracking changes in gender equality in India is vital as it can inform policy and program development to address disparities and promote a more inclusive and equitable society.

Which rows and columns of the dataset will be used to answer this question? (Answer: Actual names of the variables in the dataset that you plan to use).

I will use "Female.LFPR", "Male.LFPR", "MaleYouthNotInEET", "FemaleYouthNotInEET", "MaleLabourForceWithAdvancedEdu", "MaleLabourForceWithIntermediateEdu", "MaleLabourForceWithBasicEdu", "FemaleLabourForceWithAdvancedEdu", "FemaleLabourForceWithIntermediateEdu", "FemaleLabourForceWithBasicEdu".

Week 11

1. List the visualizations that you are going to use in your project (Answer: What are the variables that you are going to plot? How will it answer your larger question?),

I am going to plot Male LFPR and Female LFPR against Time to show: (1) the stark difference between male and female LFPR and (2) how LFPR for both genders have changed over time in order to answer the question of whether or not gender inequality is worsening in terms of occupational and economic opportunities.

Next, I will plot the difference in educational levels between males and female workers to see how males have unfair access to education when compared to their female counterparts.

2. How do you plan to make it interactive? (Answer: features of ggplot2/shiny/markdown do you plan to use to make the story interactive) ,

I will focus on using ggplot2 to create meaningful visualizations which can walk the user through the issue of gender inequality in India. I have made the decision not to incorporate shiny into my webpage as I feel that the plots generated are sufficient in answering the question of whether or not gender inequality is worsening in India.

3. What concepts incorporated in your project were taught in the course and which ones were self-learnt? (Answer: Create a table with topics in one column and Weeks in the other to indicate which concept taught in which week is being used. Leave the entry of the Week column empty for self-learnt concepts)

Week	Concept Used + How It Was Used
1	How to use Rstudio + Using Rmarkdown
2	Data visualization and use of ggplot2
4	Tidying my data to ensure that it works in ggplot2
7	Creating meaningful visualizations. I mainly used geom_line to show change over time. I used scale_color_manual to adjust color of my geom_lines I used rowMeans function to find mean of row

Include the challenges and errors that you faced and how you overcame them.

I attempted to incorporate shiny into my webpage but was not successful. I managed to run the app in Rstudio but could not get it to work when I published it. After thorough consideration, I decided to move on without the use of shiny as I felt that it did not add significant value to my story.

Week 12

Include the challenges and errors that you faced and how you overcame them.

Over W12, I attempted to plot ggplot2 graphs but struggled to do so for some variables due to the presence of NA values. I was not confident in tidying data in R before plotting it out. I sought help from Prof NV which allowed me to make progress as I was able to plot the graphs I wanted to answer the questions I had in mind.

That being said, given that this is the first time I am creating a website, I remained unsure on how to design it in a way that looks aesthetic. I also struggled to come up with a coherent story which conveyed the main issue that my website is trying to address. I am still in the process of coming up with questions which would allow me to answer the issue of how gender inequality has evolved over time in India.

Final Submission

1. What is the theme of your data story?

The theme of the data story revolves around assessing gender inequality in India by analyzing labor and educational statistics over time. The primary question addressed is whether gender equality is worsening in India, and key variables such as Female LFPR, Male LFPR, and educational levels are utilized to explore this theme. The story aims to shed light on disparities in occupational and economic opportunities, emphasizing the need to understand and address gender-based issues in the country.

2. Why is it important to address this question?

Addressing the question of whether gender equality is worsening in India is crucial for several reasons.

First, global indicators such as the United Nations' Human Development Index and the Gender Inequality Index highlight persistent gender disparities in the country. A high Gender Inequality Index (GII) score, as seen in India's case, signifies substantial gender inequality.

Second, evidence from the National Family Health Survey-4 (NFHS-4) indicates that despite some improvements, India still faces challenges related to maternal mortality, child marriage, and violence against women. These factors influence educational and economic opportunities available to women. Understanding the trajectory of gender inequality is essential to identify areas for improvement and inform policy and program development.

Lastly, tracking changes in gender equality over time provides insights into societal progress and helps create a more inclusive and equitable environment. By addressing this question, the data story aims to contribute to the ongoing dialogue on gender equality and promote initiatives that can lead to positive changes in India's social and economic landscape.

3. Why do you think the data sources that you have curated can help you answer the question?

The curated data sources, obtained from the CEIC website, which compiles information from reliable sources like the World Bank, the International Monetary Fund, and data.gov.in, provide a comprehensive and authoritative foundation for assessing gender inequality in India. Here's why these sources are instrumental in answering the question:

First, the data sources are credible and diverse. The CEIC website aggregates data from globally recognized institutions, ensuring the credibility and reliability of the information. By drawing from diverse sources, the data captures a holistic view of gender-related indicators, enriching the analysis.

Second, a longitudinal perspective is provided. The data sources offer a longitudinal perspective by including historical data. This allows for the examination of trends and changes over time, enabling a nuanced understanding of whether gender equality in India is improving or deteriorating.

Third, a comprehensive range of indicators were included. The dataset includes crucial variables such as Female LFPR, Male LFPR, and educational levels, which are vital metrics for evaluating gender inequality in both economic and educational domains. This diversity of indicators enhances the depth and breadth of the analysis.

4. What are the insights from the data and how are they depicted in plots?

My initial hypothesis going into this data experience was that gender inequality is worsening in India. My impression was formed by media sources which depicted how female labor involvement remained low even as India experienced significant economic growth. Through analysis of my data and visualizations, I drew the following insights:

First, there is a general reduction in economic gap between genders. The data indicates a decreasing gap between Female LFPR (Labor Force Participation Rate) and Male LFPR, hence suggesting that the economic gap between males and females in India has generally reduced over time.

Second, there are increased educational and vocational opportunities for female youths. A larger proportion of female youths in India now have access to education, employment, or training. The gender gap in this regard has decreased from 43% in 1990 to 29% in 2022.

Third, the educational gap between males and females has decreased. The educational gap between males and females has decreased from 56% in 2000 to less than 50% in 2022, this indicates that the average education level of a female worker has increased relative to a male worker.

Considering all of the above points, I would still conclude that there has been minimal improvement in gender inequality in India. While improvements have occurred, the overall progress in reducing gender inequality is considered minimal. The recent Global Gender Gap Index ranking India at 135 out of 146 countries underscores persistent challenges.

5. How did you implement this entire project? Were there any new concepts that you learnt to implement some aspects of it?

First, I utilized the CEIC website to collect secondary data from reputable sources like the World Bank, the International Monetary Fund, and data.gov.in. This process provided me with understanding on how to navigate and extract relevant data from diverse sources. This is a valuable skill that involves assessing data reliability, relevance, and compatibility for analysis.

Second, I formulated the Key Question of whether gender equality is worsening in India. Developing a precise and relevant research question was essential for driving the focus of the project and aligning it with broader social issues.

Third, I had to make decisions on which variables to include and exclude. I ended up choosinng key variables such as Female LFPR, Male LFPR, and educational indicators to answer the central question. The selection of variables required a strategic approach to ensure they effectively address the research question and provide meaningful insights.

Forth, I learnt how to create visualizations using ggplot2 to represent trends in labor force participation and educational disparities. Learning and implementing ggplot2 for data visualization involved understanding its syntax and capabilities, enhancing the storytelling aspect of the data.

Fifth, I learned how to create a website to present the data story, even though I struggled with design aesthetics and narrative coherence. Considerations behind my website design involved considering user experience, aesthetics, and creating a narrative flow that effectively communicates the key insights.

Word Count (for final submission): 964 Words