https://github.com/COGS108/Projects/blob/master/FinalProject_Guidelines.md#Project-Proposal

Research Questions

- Is the country's GDP and GDP per capita correlated with the Global Gender Gap Index? How can we predict the gender pay gap by using this index?
    - Is there a relationship between precipitation and the Global Gender Gap Index in a country from 2009-2019?

Background and Prior Work

A report from the World Economic Forum shows that only "3% of students joining information and communication technology courses across the globe are women." Also, women account for 17.19% of workers in STEM-qualified industries and their salaries are 19% percent less than male workers in Australia. Since we were in the STEM background and we have seen similar news and reports so frequently, we would like to further analyze the factors that contributed to gender gap in the global working place. The Gender Gap Report by World Economic Forum benchmarks the evolution of gender-based gaps among four key dimensions (Economic Participation and Opportunity, Educational Attainment, Health and Survival, and Political Empowerment). The measure of Economic Participation include the labour force participation rates by gender, ratio of estimated female-to-male earned income and wage equality for similar work. Educational Attainment captures the gap between women 's and men 's current access to education through ratios of women to men in primary-, secondary- and tertiary-level education. Health and Survival provides an overview based on sex ratio at birth and the gap between women 's and men 's healthy life expectancy. Political Empowerment measures the gap between men and women at the highest level of political decision-making through the ratio of women to men in minister-level positions and the ratio of women to men in parliamentary positions.

Since Economic participation was usually represented by the country's GDP and GDP per capita, we believe that GDP can be a driving force in the issues related to gender gap and it allowed us to explore the connection between each country's economy and gender equality.

In addition, the gender gap includes the gender wage gap, which is also highly correlated with the country's GDP. In many countries with agriculture, the GDP is always closely related to the annual precipitation. If that country is experiencing either drought or excessive rainfall, the agriculture and the country's GDP will be impacted severely. Besides the economic impact, natural disasters such as flood are also potential threats to most single mothers. Single mothers often rely on government welfare in order to live and raise children. If there is a flood caused by heavy rain that is severe enough to destroy their houses, it will further increase the gender inequality. Thus, we are curious about whether the change in precipitation will impact on gender equality.

Since the World Economic Forum publishes the annual report through its official website, we could download the data from 2009 to 2019 there. We choose to analyze the data in this time frame because we would like to avoid the effect of pandemic as the confounding variables. This decade will allow us to explore the patterns and correlation between the economy and gender gap index so that we can find the factors that influence gender equality.

Reference

Format & Structure:
https://github.com/COGS108/FinalProjects-Wi20/blob/master/FinalProject_group019.ipynb

"Global Gender Gap Report 2021". March 2021. Accessed by Apr 20, 2022. Link

"Explore Australia's data on women in STEM". Australia Government, Department of Industry, Science, Energy and Resources. Link

Johnny Wood, “3 things to know about women in STEM”. Journal of World Economic Forum. Link

Damania, Richard; Desbureaux, Sebastien; Zaveri, Esha. 2019. Does Rainfall Matter for Economic Growth? Evidence from Global Sub-National Data (1990-2014). Policy Research Working Paper;No. 8888. World Bank, Washington, DC. © World Bank. https://openknowledge.worldbank.org/handle/10986/31901 License: CC BY 3.0 IGO. Link

2022, 28 F. (2022, February 28). Explainer: How gender inequality and climate change are

interconnected. UN Women – Headquarters. Retrieved April 21, 2022, from

https://www.unwomen.org/en/news-stories/explainer/2022/02/explainer-how-gender-inequality-a

nd-climate-change-are-interconnected.

Hypothesis

Our hypothesis is there is a correlation between a country's annual precipitation and Global Gender Gap index from 2009 to 2019.

Null Hypothesis: The country’s annual precipitation has no association with the Global Gender Gap index, any association is due to random chance.

Alternative Hypothesis: Countries that have a higher annual precipitation will have a lower Global Gender Gap index in the following year. Countries that have a lower annual precipitation will have a higher Global Gender Gap index in the following year.

The test significance is 5%.

Data

Here, you are to think about and describe the ideal dataset (or datasets) you would need to answer this question:

- What variables would you have?
- How would they be stored?
- How many observations would you have?
- What/who would the observations be? Over what time period? etc.

- etc.

Global Gender Index
https://tcdata360.worldbank.org/indicators/af52ebe9?country=BRA&indicator=27962&viz=line_chart&years=2006,2021

World GDP 2021
https://data.worldbank.org/indicator/NY.GDP.MKTP.CD

Subindex:
https://tcdata360.worldbank.org/indicators/846d20f8?country=BRA&indicator=27960&viz=line_chart&years=2006,2021

Precipitation

Ethics and Privacy

- All the datasets are publicly available online and permission-free.
- Since all the datasets do not contain any identifiable personal information and we would be focused on analyzing worldwide data instead of individuals, there should be no privacy concerns.
- Possible unintended use of the data:
- In the process of analysis, no potential biases
- All the datasets are from authoritative and credible websites
- we scrape the data following the terms and condition
-

Team Expectation

- Everyone is to show up to the weekly meeting on time, unless there is a proper excuse
- Show positive attitude
- Consistent communication with everyone when one faces any issue with tasks they have been given

Project Time Proposal

- 4/20 : Meeting to discuss research questions/expectations/ set up timeline
- 4/27: Discuss data wrangling approach assign tasks and roles
- 5/4: Continue to work and finalize data cleaning

- **5/6 data due**
- 5/11: Discuss data analysis approach and assign tasks and roles
- 5/18: Continue to work and finalize data analysis
- **5/20 EDA due**
- 6/2: Finalize project and complete conclusion
- **6/6 project due**

Project Proposal

The Project Proposal is completed as a group. Your proposal must include the following sections:

NAMES: Be sure to include each member's name

- Names : Replace the lines to list each person's full name. Add lines as needed for your group size, and make sure each name is listed on a separate line.

RESEARCH QUESTION: What is your research question? Include the specific question you're setting out to answer. This question should be specific, answerable with data, and clear. A general question with specific sub questions is permitted. (1-2 sentences)

BACKGROUND & PRIOR WORK: This section will present the background and context of your topic and question in a few paragraphs. Include a general introduction to your topic and then describe what information you currently know about the topic after doing your initial research. Include references to other projects who have asked similar questions or approached similar problems. Explain what others have learned in their projects.

Find some relevant prior work, and reference those sources, summarizing what each did and what they learned. Even if you think you have a totally novel question, find the most similar prior work that you can and discuss how it relates to your project. References can be research publications, but they need not be. Blogs, GitHub repositories, company websites, etc., are all viable references if they are relevant to your project. It must be clear which information comes from which references. (2-3 paragraphs, including at least 2 references)

HYPOTHESIS: What is your main hypothesis/predictions about what the answer to your question is? Briefly explain your thinking. (2-3 sentences)

DATA: Here, you are to think about and describe the ideal dataset (or datasets) you would need to answer this question:

- What variables would you have?
- How would they be stored?
- How many observations would you have?
- What/who would the observations be? Over what time period? etc.
- etc.

Note: For the project proposal, you do NOT have to find the actual dataset(s) needed for your project. For the first checkpoint and onward, you will.

ETHICS & PRIVACY: Acknowledge and address any ethics & privacy related issues of your question(s), proposed dataset(s), and/or analyses. Use the information provided in lecture to guide your group discussion and thinking. If you need further guidance, check out Deon's Ethics Checklist. In particular:

- Are there any biases/privacy/terms of use issues with the data you propsed?
- Are there potential biases in your dataset(s), in terms of who it composes, and how it was collected, that may be problematic in terms of it allowing for equitable analysis? (For example, does your data exclude particular populations, or is it likely to reflect particular human biases in a way that could be a problem?)
- How will you set out to detect these specific biases before, during, and after/when communicating your analysis?
- Are there any other issues related to your topic area, data, and/or analyses that are potentially problematic in terms of data privacy and equitable impact?
- How will you handle issues you identified?

(1-2 paragraphs)

TEAM EXPECTATIONS: Read over the COGS108 Team Policies individually. Then, include your group's expectations of one another for successful completion of your

COGS108 project below. Discuss and agree on what all of your expectations are. Discuss how your team will communicate throughout the quarter and consider how you will communicate respectfully should conflicts arise. By including each member's name above and by adding their name to the submission, you are indicating that you have read the COGS108 Team Policies, accept your team's expectations below, and have every intention to fulfill them. These expectations are for your team's use and benefit — they won't be graded for their details.

PROJECT TIMELINE PROPOSAL: Specify your team's specific project timeline. An example timeline has been provided. Changes the dates, times, names, and details to fit your group's plan.

If you think you will need any special resources or training outside what we have covered in COGS 108 to solve your problem, then your proposal should state these clearly. For example, if you have selected a problem that involves implementing multiple neural networks, please state this so we can make sure you know what you're doing and so we can point you to resources you will need to implement your project. Note that you are not required to use outside methods.

To reemphasize: for the Project Proposal you are not expected to have already done any analyses for the proposed project, but what you submit should be a plan for what you will answer and what data you would ideally have for the project. (Of course, for the final project you will need to actually find the data and do the analyses.)

Project Proposal - Style Guidelines

The proposal should be written clearly and at a level understandable by a typical undergraduate student.

This is a short but detailed proposal meant to give us time to assess and critique your Final Project idea (further described below), in order to give you time to improve upon it throughout the quarter.

You will receive feedback on your project proposal, and you are fully expected to make the changes suggested by the Professor, TAs, IAs, and your classmates on this assignment.

Remember to proofread your Project Proposal. Do not use overly flowery and/or vague language.