Examining female employment as an indicator for promoting economic growth in the United States

Introduction

Increased women's participation in employment is often considered a driver of economic growth, thereby increasing a nation's Gross Domestic Product. In fact, a 2016 McKinsey Global Institute claimed that attaining gender parity in the labor force could add as much as $4.3 trillion to annual GDP in the United States by 2025 (Ellingrud 2016). The report goes on to claim that "About 40 percent of the potential can come from higher female labor-force participation [and} about 30 percent can come from narrowing the gap between men and women who work part time and full time…" (Ellingrud 2016). These claims do not differ greatly from those that are being a made on a global scale. For example, UN Women - an organization which exclusively focuses on gender equality and women's empowerment - asserts on its website the following: "When more women work, economies grow. An increase in female labor force participation—or a reduction in the gap between women's and men's labor force participation—results in faster economic growth (UN Women website)." Building on this body of work, for this project I aim to examine the relationship between women's labor force participation rate and economic growth using existing data sources within the context of the United States.

Client Description

The primary client for this capstone project is Data2X, a collaborative technical and advocacy platform dedicated to improving the quality, availability and use of gender data in order to make a practical difference in the lives of girls and women worldwide. Hosted by the UN Foundation, Data2X works with UN agencies, governments, civil society, academics, and the private sector to close gender data gaps, promote expanded and unbiased gender data collection, and use gender data to improve policies, strategies, and decision-making in support of gender equality. This project is specifically designed to respond to the Data 2x's recent Big Data for Gender Challenge, which aims "to catalyze innovative solutions to fill global gender data gaps on the well-being of women and girls." For this data challenge, Data2X is requesting applicants to use conventional and digital data sources to conduct gender analysis on a specific research question.

Problem Statement

"Data can be a powerful tool in the hands of women's advocates. Without it, the ability to influence policy, track progress and demand accountability has been hampered."

-Ruth Levine & Mayra Bavinic, authors of "Closing the Gender Data Gap"

While the United Nations (UN) has renewed its commitment to gender equality and the empowerment of women and girls in its new Sustainable Development Goals (SDGs) for 2030, there are still significant gender data gaps that make it difficult to monitor progress toward these goals. Although investing in innovations to fill these data gaps is paramount, it is also critical to assess how countries can leverage existing data sources and indicators to launch baseline data collection for SDGs.

In an April 2016 report, Data2X identified sixteen, ready to measure gender indicators to track progress towards eight SDGs (see table I-A) that mapped to five domains: health, education, economic opportunities, political participation, and human security. According to the report, all of these indicators have internationally agreed definitions[1]; are produced through available data collection instruments; already have comparative wide coverage; do not have built in gender biases; and are informed by recommendation from various UN agencies.

Table I-A : Eight UN Sustainable Development Goals relevant to proposed Ready to Measure gender Indicators

1. End poverty in all its forms everywhere.
2. End hunger, achieve food security and improved nutrition, and promote sustainable agriculture.
3. Ensure healthy lives and promote well-being for all at all ages.
4. Ensure inclusive and equitable quality education, and promote lifelong learning opportunities for all.
5. Achieve gender equality and empower all women and girls.
6. Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all.

[1] However, it is unclear what is the internationally agreed definition of "employed." See table I - D for the U.S. definition of employed.

7. Reduce inequality within and among countries.
8. Promote peaceful and inclusive societies for sustainable development, provide access to justice for all and build effective, accountable and inclusive institutions at all levels.

Capstone Project Summary

Within the economic opportunities domain, Data2x identified one of the Ready to Measure gender indicators as the following: females employed as a ratio of the working-age female population (15 to 59), and female employment ratio to male employment ratio. This identified indicator maps to the SDG noted in the table below.

Table I - B: Ready to Measure Indicator and relevant Sustainable Development Goal

Sustainable Development Goal	**Ready to Measure Indicator**
Promote sustained, inclusive and sustainable economic growth, full and productive employment and decent work for all.	Females employed as a ratio of the working-age female population (15 to 59), and female to male ratio (i.e. female employment ratio to male employment ratio).

For this capstone project, I will examine data available for above described indicator and compare it to the Gross Domestic Product (GDP) over a 20-year period (1996 - 2016) within the context of the United States. In my analysis, I plan to identify any interesting patterns and/or correlations in the data to determine if and to what extent the indicator is able to accurately measure economic growth. Moreover, I plan to outline the advantages and limitations of using this indicator for developing baseline data for the relevant SDG. This analysis will inform my recommendations for improving existing conventional data sources.

Data set

For this capstone project, I obtained data from the U.S. Bureau of Labor Statistics and the U.S. Department of Labor Women's Bureau on female employment as well as the female to male ratio. In addition, I will use the GDP data from the Bureau of Economic Analysis. Moreover, I obtained labor force participation rates from the Department of Labor's Women's Bureau website. All three data sources are publicly available, and their links are provided in the table below:

Table I-C: Relevant data sets and descriptions

Data Source	Description & Limitations
Bureau of Labor Statistics(BLS); Current Population Survey	One of the key limitations of using this data set is that because U.S. counts employed persons (see glossary below for more details) who are the ages of 16 years or older. For this reason, it was near possible to obtain data over the time frame for which I was seeking. The data obtained from this source is quite messy/untidy, and will require a lot manipulation to clean data. **Relevant tables:** Employment status of the civilian noninstitutional population by age, sex, and race (only for year) Employment status of the civilian noninstitutional population by sex, 1940s to date https://www.bls.gov/cps/tables.html
Bureau of Economic Analysis(BEA)/ Statistica	Statista has nicely packaged the data file that is derived from BEA, so that I can easily convert the Excel file into a CSV file. This file does not require much cleaning/wrangling for my analysis. When trying to access the data directly on BEA, I found it quite difficult to determine to extract the data I need for my analysis. https://www.statista.com/statistics/188105/annual-gdp-of-the-united-states-since-1990/
Department of Labor(DOL)/Women's Bureau	Though the data visualized in the tables below are extracted from the BLS Current Population Survey, the tables listed below serve as a good references for which direction I should take in my preliminary exploration.

	Relevant tables: Women in the Labor Force Labor Force Participation Rates Earnings and Earnings Ratios https://www.dol.gov/wb/stats/facts_over_time.htm#earn

Methodology & Approach

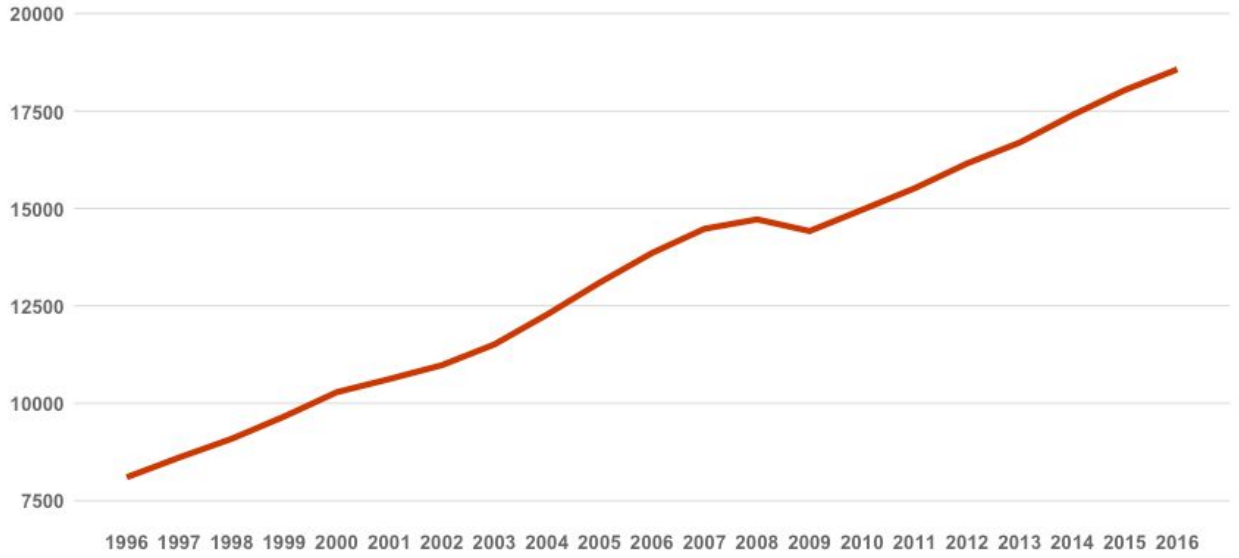
My approach involved downloading the the datasets as CSV files for a 20-year period of data for employment and GDP. I also extracted data on labor force participation rates from the Department of Labor Women Bureau's website. In some instances in which all the data necessary for analysis could not be extracted from one data source, I extracted data from more than one data source to create a complete data set. Next, I wrangled and cleaned the data for analysis. Subsequently, I analyzed and visualized the data to determine if there were any interesting patterns in the data. Specifically, I created the following plots when analyzing the data:

Table I-D: Plots for data analysis

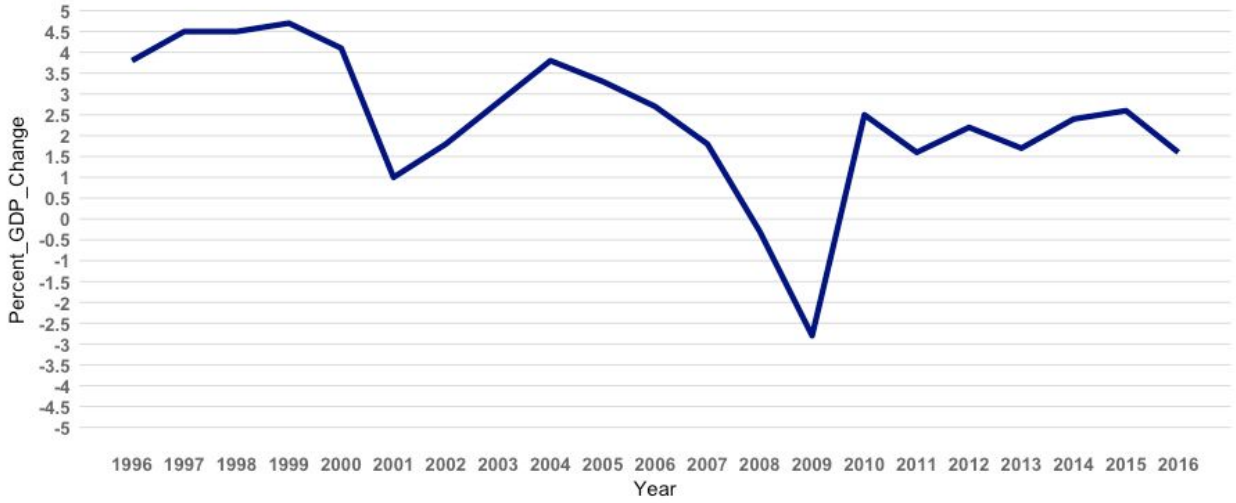
Plot	**Description**
Plot 1: U.S. GDP between 1996 and 2016	I plotted the U.S. Gross Domestic Product (GDP) on y-axis against a 20-year period (i.e. 1996 -2016) on the x-axis.
Plot 2: U.S. GDP Growth rate between 1996 and 2016	I plotted the U.S. GDP annual growth rate on the y-axis against a 20-year period (i.e. 1996 - 2016) on the x-axis.
Plot 3: Female and Male Employment Rates	I plotted both the male and female employment rates on the y-axis against a 20-year period (i.e. 1996 - 2016) on the x-axis.
Plot 4: Female to Male Employment Ratio	I plotted the female to male ratio (see "Mathematical Model" section for further

	explanation) on the y - axis against a 20-year period (i.e. 1996 - 2016) on the x-axis.
Plot: Linear regression model of GDP Growth Rate vs. Female to Male Employment Ratio	I plotted the female to male ratio (see "Mathematical Model" section for further explanation) on the y-axis against GDP Growth Rate on the x-axis.

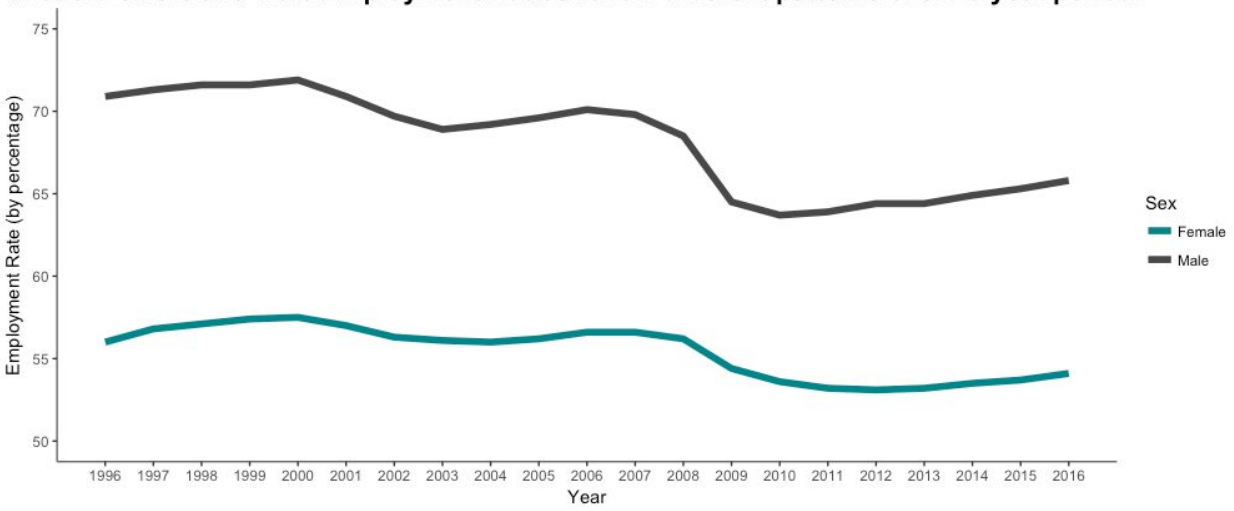
Plot 1: US GDP has steadily increased from 1996 to 2016.



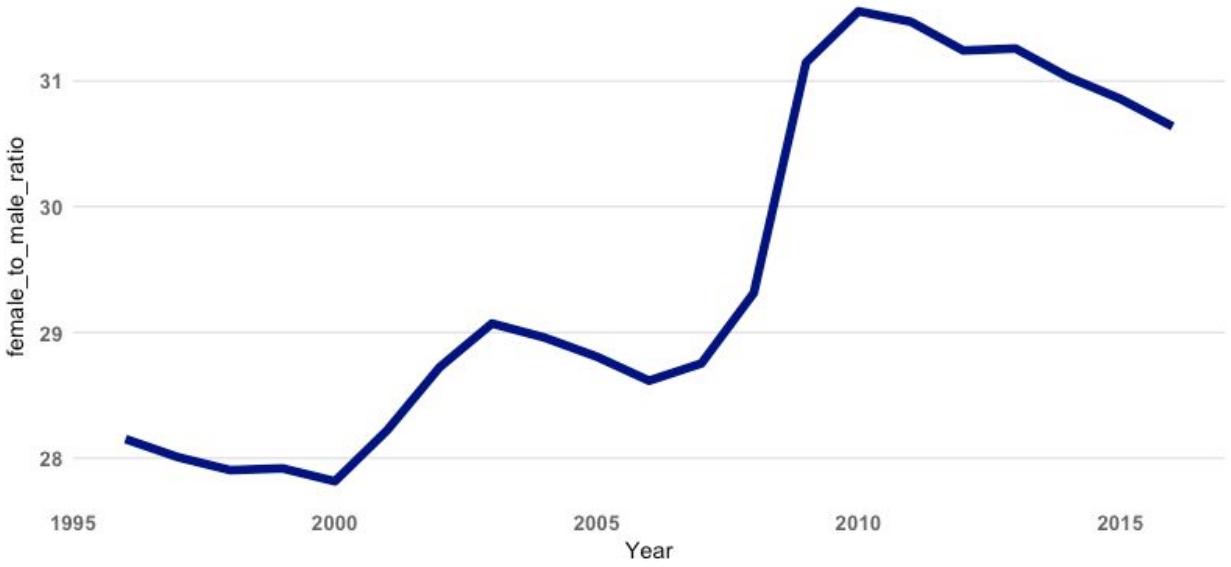
Plot 2: US GDP growth rate plummets during the Great Recession from 2007 to 2009.



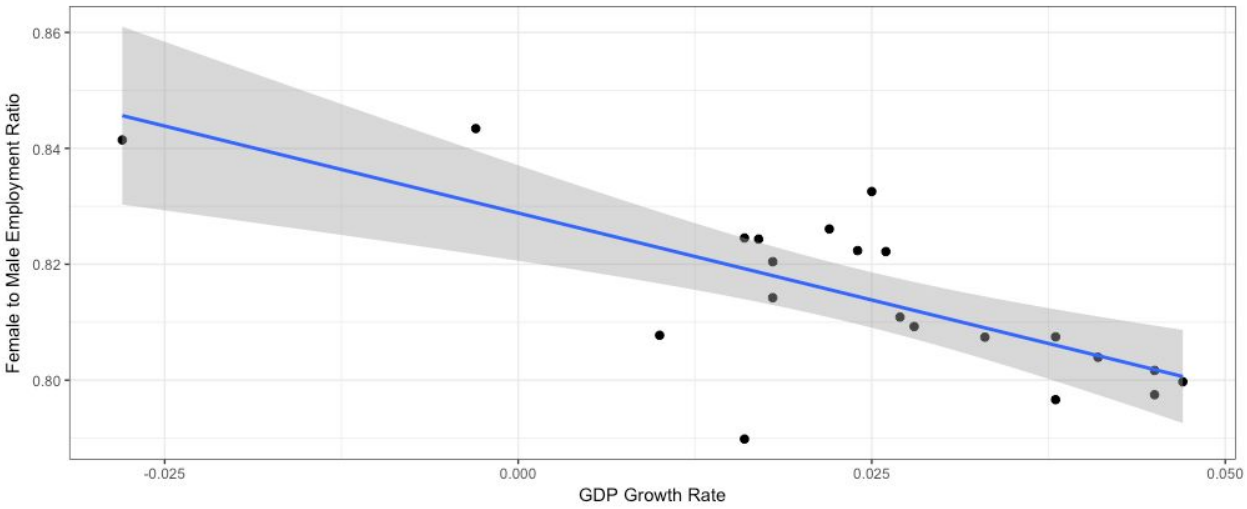
Plot 3: Female and Male Employment Ratios follow different patterns over 20-year period.



Plot 4: US Female to Male Employment Ratio peaks during recession in 2009.



Plot 5: There is a strong negative correlation between GDP Growth Rate and Female to Male Employment Ratio.



I also conducted linear regression analyses to determine if there were any correlations between GDP and women's employment rates as well as labor force participation rates. I primarily used R studio for my analysis.

Table I-E: Description of Linear Regression Analyses

Linear Regression Model	**Description**
Linear Regression 1: Female to Male Employment Ratio and GDP Growth Rate	I conducted to a linear regression analysis with GDP growth rate being the independent variable and female to male employment ratio being the dependent variable.
Linear Regression 2: Female to Male Labor Force Participation Rate and GDP Growth Rate	I conducted a linear regression analysis with GDP growth rate being the independent variable and female to male labor force participation rate being the dependent variable.

To serve as the basis for my analysis, it is important to translate the Ready to Measure Indicator (see Table I-B) into a mathematical model derived from data collected from the Current Population Survey,

U.S. Bureau of Labor Statistics. The mathematical model is outlined below as well as the Ready to Measure Indicators and their relevant BLS indicators.

Mathematical model:

1. Female Employed Persons / Female Civilian non-institutional population;
2. (Female Employed Persons/Female Civilian non-institutional population) / (Male Employed Persons/Male Civilian non-institutional population)

Table I-F: Ready to Measure Indicators vs. BLS Indicators

Ready to Measure Indicator	**BLS Indicator**	**Differences**
Employed	**Employed persons (Current Population Survey)** Persons 16 years and over in the civilian noninstitutional population who, during the reference week, (a) did any work at all (at least 1 hour) as paid employees; worked in their own business, profession, or on their own farm, or worked 15 hours or more as unpaid workers in an enterprise operated by a member of the family; and (b) all those who were not working but who had jobs or businesses from which they were temporarily absent because of vacation, illness, bad weather, childcare problems, maternity or paternity leave, labor-management dispute, job training, or other family or personal reasons, whether or	It is important to highlight the discrepancies between Ready to Measure and the relevant BLS indicator. Specifically, BLS has a rather nuanced definition of how it defines “employed”, where the UN and other international standards for how employed is defined seems less clear.

	not they were paid for the time off or were seeking other jobs. Each employed person is counted only once, even if he or she holds more than one job. Excluded are persons whose only activity consisted of work around their own house (painting, repairing, or own home housework) or volunteer work for religious, charitable, and other organizations. ** Employed included both agriculture and non-agriculture workers.	
The working-age population (15 to 59 years)	**Civilian noninstitutional population (Current Population Survey)** Included are persons 16 years of age and older residing in the 50 states and the District of Columbia who do not live in institutions (for example, correctional facilities, long-term care hospitals, and nursing homes) and who are not on active duty in the Armed Forces.	There are differences in what is considered working age population. In the context of the U.S, individuals ages 16 years or older are considered . Although it is possible to extract the appropriate ages specified by Ready to Measure, this can only be done for years 2016 and 2015, thus making it difficult to chart trends over time.

Analysis & Results

My initial findings from the data analyses revealed the following:

- **U.S. GDP has been steadily increasing between 1996 and 2016.** For my preliminary exploration, I charted the U.S. GDP from 1996 and 2016 (Plot 1). This GDP plot shows that the U.S. GDP has been steadily increasing since 1996 with a slight dip between 2008 and 2009; this trend correlates with the Great Recession—which officially lasted from December 2007 to June 2009. During the twenty-year period shown in the graph, the GDP reaches its peak in 2016, indicating that the United States has experienced continuous economic growth for the past 20 years.
- **U.S. GDP growth rate sharply declined between 2008 and 2009.** Moreover, it was the only occurrence within the 20-year period that the GDP had a negative growth rate, which correlates with the Great Recession. In this vein, this plot (Plot 2) provides a clearer picture of when Great Recession occurred in comparison to the U.S. GDP plot.
- **Extracting labor force participation data from the U.S. Bureau of Labor Statistics database proved to be quite challenging.** To build on my analysis of women's employment rates, I analyzed labor force participation rates, especially since However, my attempts at obtaining the data from BLS to conduct this analysis proved to be quite challenging. As a result, I downloaded data on labor force participation rates from the Department of Labor Women's Bureau in pdf format and manually entered the data into a csv file for my analysis.
- **Extracting BLS data to match the UN's definition of "working-age population" proved to be quite challenging.** As noted above in Table 1-F, the working age population as definition in the *Ready to Measure* report, are individuals who are between the ages of 15 and 59 years old. In contrast, the BLS considers individuals who are 16 years or older. As a result, it was quite challenging to extract the data to match Ready to Measure's working age population. In fact, the only years that this data could be extracted for age subgroups were the years of 2015 and 2016, thereby making it difficult to track changes over time. Moreover, even with the ability to extract age groups, the manipulated data set still excluded individuals who are 15 years old since data was not collected on these individuals.
- **The difference between measuring employment rate for populations that are 16 years or older versus 16 to 59 years old is quite significant.** When analyzing the disaggregated data for 2016, there were notable differences between analyzing 16 years or older versus 16 to 59 years old. As shown in the table below, the differences between the employment rates between the two age groups ranged from 16 percent to over 20 percent. These differences strongly suggest that if available, the data for the UN-defined working-age population within the U.S. would yield different analyses and findings. Moreover, it suggests it would be highly difficult to obtain baseline data across various countries given the challenges I encountered in obtaining such data for the United States. This is underscored by the fact that United States has stronger infrastructure in place to collect and analyze population and labor force data in comparison to other developing countries.

Sex	Age	Employment Rate	Difference[2]
All	16 years and over	59.7	+ 18.9 %
All	16 to 59 years	71.0	
Male	16 years and over	65.8	+ 16.1 %
Male	16 to 59 years	76.4	
Female	16 years and over	54.1	+ 20.7 %
Female	16 to 59 years	65.3	

- **Female employment ratio does not match the trend of U.S. GDP between 1996 and 2016.** Interestingly, the female employment ratio plot and does not follow the same pattern as the GDP plot. During the 20-year period shown in the graph, female employment ratio peaks in 2000. The female employment ratio has its sharpest decline (2.2 percent) between 2008 and 2009 and reaches its lowest point in 2012. While the female employment ratio has been increasing slowly between 2012 and 2016, it has been doing so at a slower rate in comparison to the male employment rate during this time frame. Moreover, the female employment rate has yet to reach pre-recession levels.
- **Male employment ratio and female employment ratio diverge between the years of 2010 and 2012.** Like the female employment ratio, the male employment ratio peaks in 2000. Its sharpest decline is between the years of 2007 and 2008. What is most interesting when comparing the female to male employment ratio, is between the years of 2010 and 2012 during which the female employment ratio decreases and male employment ratio increases, thereby widening the gap between female and male employment ratios. This divergence suggests that males have recovered from the economic recession at a faster rate than females. It is also noteworthy that the male employment ratio lowest point (63.7 percent) is still higher than the female employment ratio highest point (57.5) during the 20-year time period.
- **The U.S. female to male employment ratio peaks during the Great Recession in 2009.** To help better understand the relationship between female and male employment ratios during the 20-year time period, I created a plot of female employment ratio to male employment ratio. As this plot indicates, gender parity where the female employment ratio equals the male employment ratio has not been reached during this 20-year time

[2] This was calculated by taking the difference between employment rate for 16 years and over from the employment rate for 16 to 59 years and then dividing this difference by the employment rate for 16 years and older.

period. Moreover, it shows that the female to male employment reaches its peak in 2009 during the Great Recession.

- **The correlation between the GDP growth rate and female to male employment ratio was negative and significant.** A linear regression analysis of GDP growth rate and female to male employment ratio yielded a correlation coefficient of -.72, indicating a strong negative correlation.
- **The correlation between the GDP growth rate and female to male labor force participation ratio.** A linear regression analysis of the GDP growth rate and the female to male employment labor force participation ratio yielded a correlation coefficient of -.58, also indicating a strong negative correlation.

Discussion and Conclusion

Overall, the analyses of U.S. labor force and employment data presented in this paper do not align with existing literature explaining on relationship gender parity in employment and economic growth. In fact, in many ways this paper shows the *opposite*: gender parity seems more attainable during periods economic decline and vice versa. Given this rather disturbing and perplexing finding, it is paramount to discuss these findings more in-depth and provide recommendations further analysis and future research.

- **Challenges with obtaining data in United States indicate potentially significant difficulties with obtaining baseline data across multiple countries.** The challenges noted earlier with obtaining both employment and labor force data from U.S. databases revealed it was near impossible to obtain the data necessary to conduct an analysis of the data according to definitions outlined in the *Ready to Measure* report. Indeed, these differences in how employed individuals and labor force are defined and counted allude to potentially significant challenges with data gaps and inconsistencies when extracting data across multiple countries. Most importantly, this challenge highlights that the indicator examined in the project may not be "ready to measure," thus making it difficult to obtain robust baseline data.

- **The Ready to Measure indicator may not be well suited for measuring economic growth.** The strong negative correlation between the GDP growth rate and the female to male employment ratio suggest that it may not be a reliable indicator for economic growth. In this vein, the UN may need to consider alternative measures for examining the relationship between gender equality in employment and economic growth.

These recommendations are in many ways informed by the challenges and limitations of conducting this research project.

- **Establish a universal definition of employed persons to ensure more consistent data collection and analyses across countries.** Because this paper revealed how discrepancies in how employed persons and the labor force is defined and counted can significantly impact findings, it is paramount to establish a universal definition of these terms to ensure consistencies in data. Moreover, it is critical to ascertain that such definitions can be consistently applied across multiple country data sets.
- **Extend the time period beyond 20 years to help ensure more robust analyses.** Given that only a 20-year period was examined in this paper, it is recommended to analyze a data set that covers a longer time period. Having a longer time frame will allow for a more robust analysis because it will be inclusive of more economic cycles. When conducting cross-country analyses it is critical to check that it is feasible to collect data for the selected time frame from all the countries to be included in the analyses.
- **Account for other contributing factors of economic growth when examining the relationship between female employment and economic growth.** It is important to note that the other significant contributing factors that may need to be accounted for when analyzing the relationship between female employment participation rates and economic growth. For example, in a 2009 *Feminist Economics* article, the authors found that investments, population growth, and growth in the working-age population were identified as factors that impacted economic growth. This suggests that perhaps a more sophisticated, multiple-regression analyses are needed so these variables can be accounted for and considered in future research.
- **Conduct more sub-group analyses to better understand the relationship between female employment and economic growth.** While this paper did not conduct subgroup analyses to understand how race, age, and education level may impact the relationship between female employment and economic growth, it is recommended that such analyses be conducted in future research. Such analyses for example would a more robust, intersectional analysis of how race and gender interact when analyzing employment rates and economic growth. Indeed, the inclusion of sub-group analyses add another element of complexity that will be difficult to conduct for cross-country analyses. Nonetheless, such subgroup analyses may yield interesting findings that may otherwise be overlooked when examining patterns within a particular context.

While the data and analysis presented in this paper may not be sufficient to completely dismantle previous claims that gender equality in employment leads to economic growth, the findings indicate that further research is needed to empirically prove such claims. As a feminist who leverages research and data to catalyze social change and who firmly believes I hope you will join me in my journey to find answers to the perplexing questions raised in this paper.

Appendix A: Glossary

Below is a glossary terms that will be helpful to reference for my analysis. Most notable among these is the definition of “employed persons”, which does not exactly match the UN’s definition of “working-age population.”

Civilian noninstitutional population (Current Population Survey)
Included are persons 16 years of age and older residing in the 50 states and the District of Columbia who do not live in institutions (for example, correctional facilities, long-term care hospitals, and nursing homes) and who are not on active duty in the Armed Forces**.**

Labor force (Current Population Survey)
The labor force includes all persons classified as employed or unemployed in accordance with the definitions contained in this glossary.

Labor force participation rate (Current Population Survey)
The labor force as a percent of the civilian noninstitutional population.

Employed persons (Current Population Survey)
Persons 16 years and over in the civilian noninstitutional population who, during the reference week, (a) did any work at all (at least 1 hour) as paid employees; worked in their own business, profession, or on their own farm, or worked 15 hours or more as unpaid workers in an enterprise operated by a member of the family; and (b) all those who were not working but who had jobs or businesses from which they were temporarily absent because of vacation, illness, bad weather, childcare problems, maternity or paternity leave, labor-management dispute, job training, or other family or personal reasons, whether or not they were paid for the time off or were seeking other jobs. Each employed person is counted only once, even if he or she holds more than one job. Excluded are persons whose only activity consisted of work around their own house (painting, repairing, or own home housework) or volunteer work for religious, charitable, and other organizations.

Unemployed persons (Current Population Survey)
Persons aged 16 years and older who had no employment during the reference week, were available for work, except for temporary illness, and had made specific efforts to find employment sometime during the 4-week period ending with the reference week. Persons who were waiting to be recalled to a job from which they had been laid off need not have been looking for work to be classified as unemployed.

Appendix B: Code for Plots and Linear Regression Models

Load R Packages

```
#Load appropriate libraries
library(gridExtra)
library(ggplot2)
library(dplyr)
library(grid)
library(readr)
library(tidyverse)
library(tidyr)
```

Plot 1: Annual U.S. Gross Domestic Product

```
#Plot 1: Annual GDP
#Load csv file
annual_us_gdp <- read_csv("~/Desktop/Foundations of Data Science/Capstone
Project/annual-us-gdp-1990-2016.csv")
#View data file to check
View(annual_us_gdp)
#Convert to dataframe
plot_gdp <- as.data.frame(annual_us_gdp)
#Drop rows of irrelevant years
plot_gdp <- dplyr::filter(plot_gdp, Year > 1995)
#View data file to check
View(plot_gdp)
#Assign color to line
line_color <- "orangered3"
#Plot data
plot <- ggplot() +
  geom_line(data = plot_gdp, aes(x = Year, y = GDP),
            color = line_color, size = 3/2)
plot
#Touch up scale and theme options
plot1 <- plot +
  scale_x_continuous(limits = range(plot_gdp$Year),
                     breaks = 1996:2016,
                     labels = 1996:2016) +
  scale_y_continuous(limits = c(7500, 20000),
                     labels = seq.int(from = 7500, to = 20000, by = 2500),
                     breaks = seq.int(from = 7500, to = 20000, by = 2500)) +
  theme_minimal() +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major.x = element_blank(),
```

```
        panel.grid.major.y = element_line(color = "gray85", size = 1/3),
        axis.text = element_text(color= "gray40", face = "bold", size = 10),
        plot.margin = unit(c(4/3, 0.5, 1, 0.5), units = "lines"))
plot1

#Add Graph Title
grob_plot <- ggplot_gtable(ggplot_build(plot1))
grob_plot$layout$clip[grob_plot$layout$name == "panel"] <- "off"
grid.arrange(grob_plot)
g_title <- textGrob(
  label = "Plot 1: US GDP has steadily increased from 1996 to 2016.",
  x = unit(0.5, "lines"),
  y = unit(0, "lines"),
  hjust = 0, vjust = 0,
  gp = gpar(
    fontsize = 16,
    fontface = "bold"
  ))
gg <- arrangeGrob(grob_plot, top = g_title)
grid.arrange(gg)
```

Plot 2: GDP Growth Rate

```
#Plot 2: GDP growth rate
#Load csv file
gdp_growth <- read_csv("~/Desktop/Foundations of Data Science/Capstone
Project/GDP_GrowthRate.csv")
#View data file to check
View(gdp_growth)
#Convert to data frame
plot_gdp_growth <- as.data.frame(gdp_growth)
str(gdp_growth)
#Assign line color
line_color <- "navy"
#Plot data
plot2 <- ggplot() +
  geom_line(data = plot_gdp_growth, aes(x = Year, y = Percent_GDP_Change),
            color = line_color, size = 3/2)
plot2

#Touch up scale and theme options
plot3 <- plot2 +
  scale_x_continuous(limits = range(plot_gdp_growth$Year),
                     breaks = 1996:2016,
                     labels = 1996:2016) +
  scale_y_continuous(limits = c(-5, 5),
                     labels = seq.int(from = -5, to = 5, by = .5),
                     breaks = seq.int(from = -5, to = 5, by = .5)) +
  theme_minimal() +
```

```
  theme(panel.grid.minor = element_blank(),
        panel.grid.major.x = element_blank(),
        panel.grid.major.y = element_line(color = "gray85", size = 1/3),
        axis.text = element_text(color= "gray40", face = "bold", size = 10),
        plot.margin = unit(c(4/3, 0.5, 1, 0.5), units = "lines"))
plot3

#Add Graph Title
grob_plot1 <- ggplot_gtable(ggplot_build(plot3))
grob_plot1$layout$clip[grob_plot1$layout$name == "panel"] <- "off"
grid.arrange(grob_plot1)
g_title1 <- textGrob(
  label = "Plot 2: US GDP growth rate plummets during the Great Recession from 2007 to 2009.",
  x = unit(0.5, "lines"),
  y = unit(0, "lines"),
  hjust = 0, vjust = 0,
  gp = gpar(
    fontsize = 16,
    fontface = "bold"
  ))
gg1 <- arrangeGrob(grob_plot1, top = g_title1)
grid.arrange(gg1)
)
```

Plot 3: Female and Male Employment Ratios

```
#Plot 3: Female and male employment ratios
#Read csv file downloaded from BLS website
BLS_data <- read_csv("~/Desktop/Foundations of Data Science/Capstone Project/BLS
data_1976_2016.csv")
#View file to check data
View(BLS_data)
#Convert to data frame
BLS_data <- as.data.frame(BLS_data)
#Dropping rowing with missing values and assign to new file name
BLS_data1 <- drop_na(BLS_data)
#View data file to check changes
View(BLS_data1)
# Renaming columns in data file
BLS_data2 <- dplyr::rename(BLS_data1, Civilian_Population = X2, Civilian_LaborForce_Total =
X3, CLF_PercentofPop = X4, CLF_TotalEmployed = X5, CLF_Employed_PercentofPop = X6,
CLF_Employed_Agriculture = X7, CLF_Employed_NonAgriculture = X8, CLF_TotalUnemployed = X9,
CLF_Unemployed_PercentofCLF = X10, NotInLaborForce = X11)
# Checking column names
head(BLS_data2)
View(BLS_data2)
# Renaming first column in data file
colnames(BLS_data2)[1] <- 'Year'
View(BLS_data2)
```

```
#Convert columns to numeric
BLS_data2[] <- lapply(BLS_data2, function(x) as.numeric(as.character(x)))
#Create new column for Sex; assign Male and Female to appropriate values
BLS_data2$Sex <- c(replicate(41,"Male"),replicate(41,"Female"))
View(BLS_data2)
str(BLS_data2)
#Drop rows so that only years remaining are 1996 - 2016
BLS_data2 [1:20, ]
BLS_data3 <- BLS_data2 [-(1:20), ]
View(BLS_data3)
BLS_data3 [22:41, ]
BLS_data4 <- BLS_data3 [-(22:41), ]
View(BLS_data4)
#Plot data
plot_employment <- ggplot(data = BLS_data4, aes(x= Year, y = CLF_Employed_PercentofPop, group
= Sex, color = Sex)) +
  geom_line(size = 2)
plot_employment
#Create a new data frame that includes all 2016 observations
label_data <- subset(BLS_data4, Year == 2016)
#Display the data set
label_data
#Create a column named "palette" for each color.
label_data$palette <- c("turquoise4", "gray29")
#Create another new column named "labels" for each label.
label_data$labels <- c("Female Employment Ratio", "Male Employment Ratio")
#View the updated label set.
label_data
#Plot female and male employment ratios
plot_employment1 <- ggplot(data = BLS_data4, aes(x = Year, y = CLF_Employed_PercentofPop,
group = Sex, color = Sex)) +
  geom_line(size = 2) +
  annotate("text", x = label_data$Year, y = label_data$CLF_Employed_PercentofPop, label =
label_data$labels, color = label_data$palette, hjust = -0.5) +
  scale_color_manual(values = label_data$palette) +
  theme_bw() +
  labs(x = "Year", y = "Employment Rate (by percentage)") +
  theme(panel.border = element_blank(),
    panel.grid = element_blank(),
    axis.line = element_line(color = "gray40", size = 0.5)) +
  scale_x_continuous(limits = range(BLS_data4$Year),
                     breaks = 1996:2016,
                     labels = 1996:2016) +
  scale_y_continuous(limits=c(50,75), breaks=seq(50,75,5))
plot_employment1
#Add Graph Title
grob_plot1 <- ggplot_gtable(ggplot_build(plot_employment1))
g_title1 <- textGrob(
  label = "Plot 3: Female and Male Employment Ratios follow different patterns over 20-year
period.",
```

```
  x = unit(0.5, "lines"),
  y = unit(0, "lines"),
  hjust = 0, vjust = 0,
  gp = gpar(
    fontsize = 16,
    fontface = "bold"
  ))
gg1 <- arrangeGrob(grob_plot1, top = g_title1)
grid.arrange(gg1)
```

Plot 4: Female to Male Employment Ratio and GDP

```
#Plot 4: Female to male employment ratio

#Create subset data for female employment ratio
female_employment_ratio <- BLS_data4[c('Year','CLF_Employed_PercentofPop','Sex')]
View(female_employment_ratio)
female_employment_ratio <- female_employment_ratio[-c(1:21), ]
View(female_employment_ratio)
female_employment_ratio <- female_employment_ratio[-c(1,3)]
View(female_employment_ratio)
colnames(female_employment_ratio)[1] <- "Female_Employment_Ratio"
View(female_employment_ratio)
#Create subset data for male employment ratio
male_employment_ratio <- BLS_data4[c('Year','CLF_Employed_PercentofPop','Sex')]
View(male_employment_ratio)
male_employment_ratio <- male_employment_ratio[-c(22:42), ]
View(male_employment_ratio)
male_employment_ratio <- male_employment_ratio[-c(1,3)]
View(male_employment_ratio)
colnames(male_employment_ratio)[1] <- "Male_Employment_Ratio"
View(male_employment_ratio)
#Dropped year and sex columns to isolate male employment ratio
BLS_data4$female_employment_ratio <- female_employment_ratio$Female_Employment_Ratio
View(BLS_data4)
BLS_data4$male_employment_ratio <- male_employment_ratio$Male_Employment_Ratio
View(BLS_data4)
str(BLS_data4)
#Convert values from character to numeric
BLS_data5 <- within(BLS_data4, {
  male_employment_ratio <- as.numeric(as.character(male_employment_ratio))
  female_employment_ratio <- as.numeric(as.character(female_employment_ratio))
})
View(BLS_data5)
#Create new column for female ratio/male ratio
BLS_data6 <- mutate(BLS_data5, female_to_male_ratio =
female_employment_ratio/male_employment_ratio)
View(BLS_data6)
```

```
#Drop sex column
BLS_data7 <-BLS_data6[-c(12)]
View(BLS_data7)
#Drop rows to remove reductant values and years
BLS_data8 <- BLS_data7[-c(22:42), ]
View(BLS_data8)
#Assign color to line
line_color <- "navy"
#Plot data
BLS_data9 <- transform(BLS_data8, Year=as.numeric(Year))
plot_employment_ratio <- ggplot(data = BLS_data9, aes(x= Year, y = female_to_male_ratio, group
= 1)) +
  geom_line(size = 2, color = line_color)
plot_employment_ratio
#Touch up plot and scales
plot_employment_ratio1 <- plot_employment_ratio +
  theme_minimal() +
  theme(panel.grid.minor = element_blank(),
        panel.grid.major.x = element_blank(),
        panel.grid.major.y = element_line(color = "gray85", size = 1/3),
        axis.text = element_text(color= "gray40", face = "bold", size = 10),
        plot.margin = unit(c(4/3, 0.5, 1, 0.5), units = "lines"))
plot_employment_ratio1
#Add Graph Title
grob_plot <- ggplot_gtable(ggplot_build(plot_employment_ratio1))
grob_plot$layout$clip[grob_plot$layout$name == "panel"] <- "off"
grid.arrange(grob_plot)
g_title <- textGrob(
  label = "Plot 4: US Female to Male Employment Ratio peaks during recession in 2009.",
  x = unit(0.5, "lines"),
  y = unit(0, "lines"),
  hjust = 0, vjust = 0,
  gp = gpar(
    fontsize = 16,
    fontface = "bold"
  ))
gg <- arrangeGrob(grob_plot, top = g_title)
grid.arrange(gg)
```

Linear Regression Model for Female to Male Employment Ratio and GDP

```
#Calculate linear regression of female to male employment ratio and GDP growth rate

#Load csv file
gdp_growth_employment_ratio <- read_csv("~/Desktop/Foundations of Data Science/Capstone
Project/GDP_GrowthRate and Female Ratio.csv")
#View file to check data
```

```
View(gdp_growth_employment_ratio)
#Delete columns
gdp_growth_employment_ratio<- gdp_growth_employment_ratio[,-5]
#View file to check
View(gdp_growth_employment_ratio)
#linear regression of gdp growth and female to male employment ratio
gdp.modl <- lm(Female_to_Male_Ratio ~ Converted_GDPPercent, data =
gdp_growth_employment_ratio)
summary(gdp.modl)
gdp.modl$residuals
SSE = sum(gdp.modl$residuals^2)
SSE
cor(gdp_growth_employment_ratio$Female_to_Male_Ratio,gdp_growth_employment_ratio$Converted_GDP
Percent)
```

Plot 5: Linear Regression Model for Female to Male Employment Ratio and GDP

```
#Plot 5: Linear Regression Model for GDP Growth Rate and Female to Male Employment Ratio

# Save the graph as an object so that it can be added to or modified as needed.
plot <- ggplot(data = gdp_growth_employment_ratio, aes(x = Converted_GDPPercent, y =
Female_to_Male_Ratio)) +
  geom_point(size = 2, shape = 19) +
  theme_bw() +
  stat_smooth(method='lm') +
  labs(x = "GDP Growth Rate", y = "Female to Male Employment Ratio") +
  ggtitle("There is a strong negative correlation between GDP Growth Rate and Female to Male
Employment Ratio.")
plot
```

Linear Regression Model for Female to Male Labor Force Participation Rate and Gross Domestic Product

```
#Calculate linear regression of female to male labor force participation rate and GDP growth
rate
#Load csv file
gdp_growth_lfpr_ratio <- read_csv("~/Desktop/Foundations of Data Science/Capstone
Project/LFPR_Female_Ratio.csv")
#View file to check data
View(gdp_growth_lfpr_ratio)
#linear regression of GDP growth and female to male employment ratio
gdp.modl1 <- lm(Female_to_Male_LFPR ~ Converted_GDPPercent, data = gdp_growth_lfpr_ratio)
summary(gdp.modl1)
gdp.modl$residuals
```

```
SSE = sum(gdp.modl1$residuals^2)
SSE
cor(gdp_growth_lfpr_ratio$Female_to_Male_LFPR,gdp_growth_lfpr_ratio$Converted_GDPPercent)
```

Appendix C: References

Appel, D. et. al, (2017). Data2x and Open Data Watch. Retrieved from: http://data2x.org/wp-content/uploads/2017/07/R2M-Report-Phase-2.pdf

Buvinic, M., et. al, (2016). *Ready to Measure Progress on Gender Equality.* Data2x and Open Data Watch. Retrieved from: http://data2x.org/wp-content/uploads/2014/08/Ready_to_Measure.pdf

Buvinic, M. et. al, (2014). *Mapping Gender Data Gaps.* Data2x. Retrieved from: http://data2x.org/wp-content/uploads/2014/11/Data2X_MappingGenderDataGaps_FullReport.pd

Buvinic, M. & Levine, R. (2016). *Closing the Gender Data Gap.* Significance. Retrived from: http://data2x.org/wp-content/uploads/2017/03/Closing-the-Gender-Data-Gap-Mayra-Buvinic-and-Ruth-Levine.pdf

Ellingrud, K., et. al, (2016). *The Power of Parity: Advancing Women's Equality in the United States.* McKinsey Global Institute. Retrieved from: http://www.mckinsey.com/global-themes/employment-and-growth/the-power-of-parity-advancing-womens-equality-in-the-united-states

Klasen, S. and F., Lammanna. (2009). *Feminist Economics* 15(3), 91–132. Retrieved from: http://chicagopolicyreview.org/wp-content/uploads/2014/09/Gender-and-Economic-Growth.pdf

Organization for Economic Cooperation and Development (OECD). (2012). Gender Equality in Education, Employment and Entrepreneurship: Final Report to the MCM 2012. Retrieved from: http://www.oecd.org/employment/50423364.pdf

UN Women (2015). *Facts and Figures on Economic Empowerment.* Retrieved from: http://www.unwomen.org/en/what-we-do/economic-empowerment/facts-and-figures