# Gender Inequality and Female Literacy

Casey Lei, Chiraag Balu, Chris Zhan, Heather Yuan, Kaitlyn Phan

# Team Introductions

**Chiraag Balu**
Sophomore
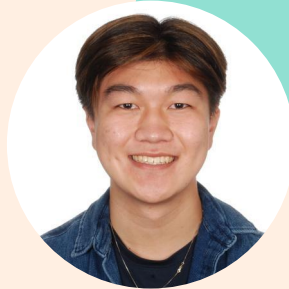Economics

**Kaitlyn Phan**
Sophomore
Cognitive Science

**Heather Yuan**
Sophomore
Media Studies

**Chris Zhan**
Sophomore
Molecular Cell Biology

**Casey Lei**
Sophomore
Theater

# Agenda

**01** Problem

**02** Data Visualization

**03** Analysis & Insights

**04** Next Steps

# 01

# Problem

# Problem Statement

*"Women account for two thirds of the 750 million adults without basic literacy skills." - UNESCO[1]*

Under the theme of **Analyzing + Accelerating Global Development with Data Science**, we focused on accelerating global development in **Gender Equality** through the outcome of **female literacy rates**.

Using data science, we answered the following question: **What factors should we prioritize in improving Gender Equality outcomes?**

# Why Literacy Rate?

**Definition:** Percentage of adult females (15 or older) who can read and write.

The UN defines Female Literacy as a key outcome for the **Sustainable Development Goal (SDG) of Gender Equality**

This statistic provides a **tangible outcome** to **accelerate for global development** in gender equality.
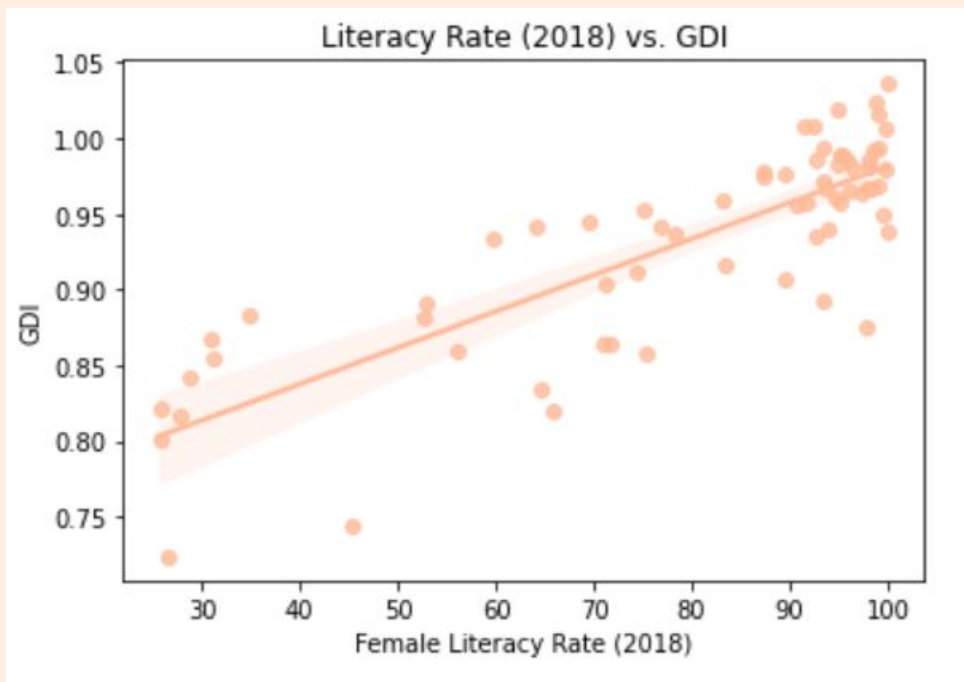
**Source:** World Bank

# Correlating Literacy Rates and GDI

This section **confirms** that female literacy rate is an outcome of gender equality by **determining its correlation** with the UN metric of Gender Development Index using **linear regression**.

# Female Literacy vs Gender Development Index (GDI)



Literacy Rate (2018) vs. GDI

**Definition:** GDI is the ratio of female to male HDI values, which acts as a metric for gender equality.

**Source:**
Human Development Report, 2020

**Correlation Coefficient**: 0.82

# Methodology

1. Establish that female literacy rate is an **outcome** of gender equality by determining its correlation with the UN **metric** of Gender Development Index using linear regression
2. Use linear regressions to find **correlations** between female literacy rates and indicators defined by the United Nations' Sustainable Development Goals for Gender Equality[1]
3. Use **multiple linear regression** to calculate the weight factor of each indicator in correlation with female literacy rates
4. Identify the indicators that are most heavily weighted with female literacy rates to pinpoint which ones should be **prioritized** for global development

1. UN Gender Equality SDG

# UN Gender Equality SDG Indicators Considered

## Unemployment
Total unemployment rate (female to male ratio)
SDG 5.1 - Discrimination

## Violence
Violence from partner (% of women ages 15+)
SDG 5.2 - Gender Violence

## Child marriage
Women married by 18 (% of married women ages 20–24)
SDG 5.3 - Harmful Practices

## Maternity Leave
Mandatory paid maternity leave (days)
SDG 5.4 - Value Unpaid Care

## Leadership
Share of seats in parliament (% held by women)
SDG 5.5 - Women Leaders

## Sexual Health
Contraceptive prevalence (% of women ages 15–49)
SDG 5.6 - Reproductive Rights

## Money
Women with bank accounts (% of women ages 15+)
SDG 5.a - Financial Independence

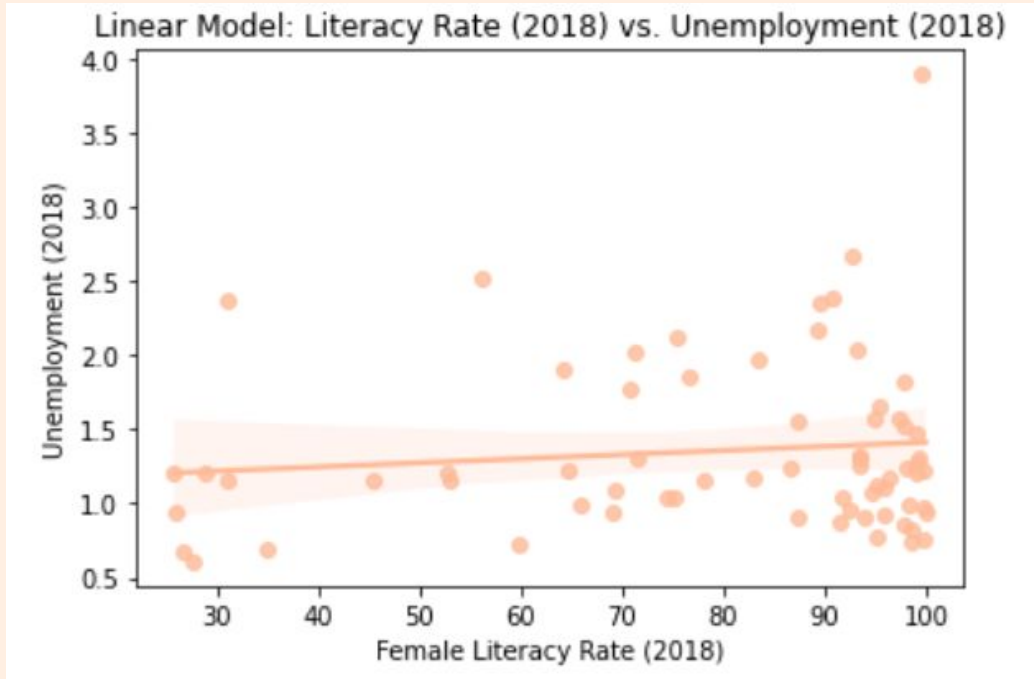# Data Visualization

**02**

# Correlating Literacy Rates and Indicators

This section contains **analyses** for each of the
**datasets** associated with the **UN SDG Indicators**
to see how they are **individually correlated** with
our outcome variable of **female literacy rates**
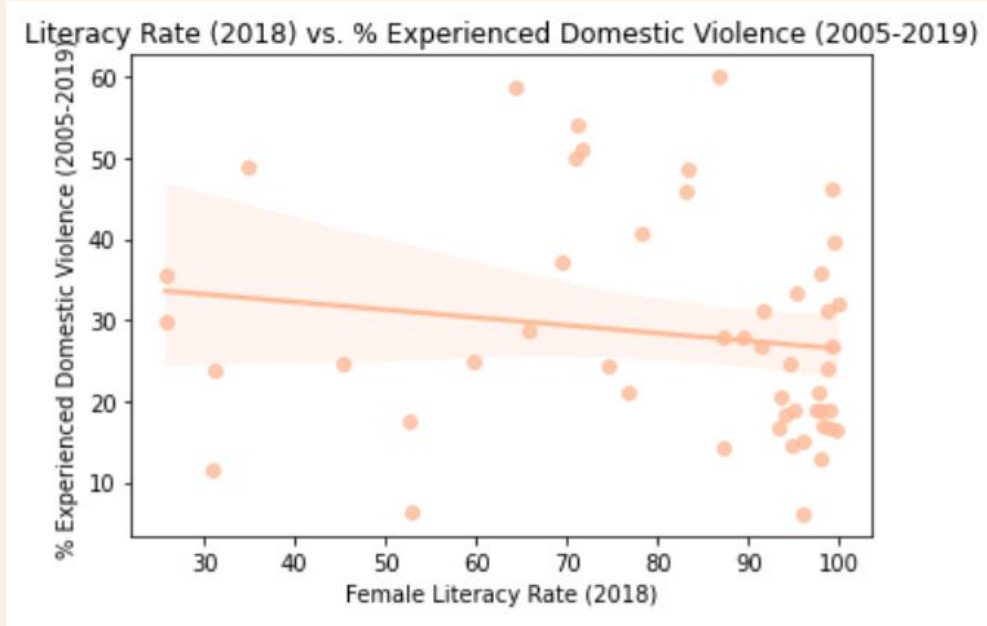
# Unemployment vs Female Literacy



Linear Model: Literacy Rate (2018) vs. Unemployment (2018)

**Definition**: Ratio of unemployed women (15 or older) to ratio of unemployed men.

**Source**: ILO (International Labour Organization)

**Correlation Coefficient**: 0.11

# Violence vs Female Literacy



Literacy Rate (2018) vs. % Experienced Domestic Violence (2005-2019)

**Definition:** Percentage of the female population ages 15 and older that has ever experienced physical and/or sexual violence from an intimate partner.

**Source:** UN Women (United Nations Entity for Gender Equality and the Empowerment of Women), 2019

**Correlation Coefficient:** -0.16

# Child Marriage vs Female Literacy



Literacy Rate (2018) vs. % Women Married by 18 (2005-2019)

**Definition:** Percentage of women ages 20-24 who were first married or in union before age 18.

**Source:** United Nations Statistics Division (2020)

**Correlation Coefficient:** -0.71

# Maternity Leave vs Female Literacy



Literacy Rate (2018) vs. Maternity Leave (Total Days) (2019)
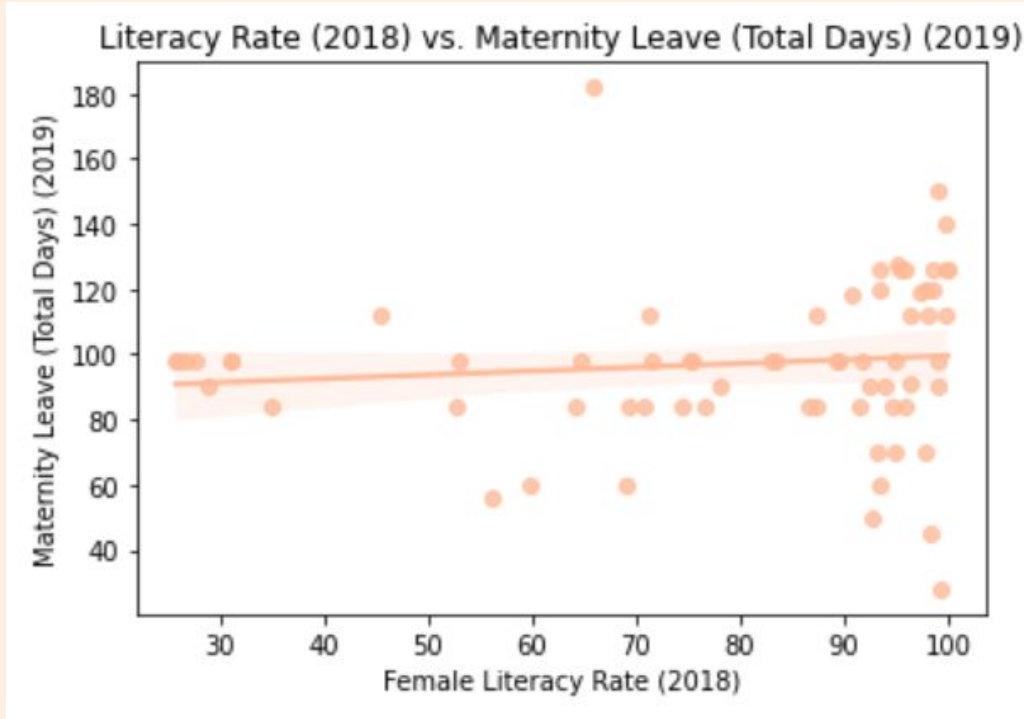
**Definition:** The mandatory minimum number of calendar days that legally must be paid by the government, the employer or both. It refers to leave related to the birth of a child that is available only to the mother; it does not cover parental leave that is available to both parents.

**Source:** World Bank (2020) Gender Statistics database.

**Correlation Coefficient:** 0.11

# Leadership vs Female Literacy



Literacy Rate (2018) vs. % of Parliament Seats held by women (2018)
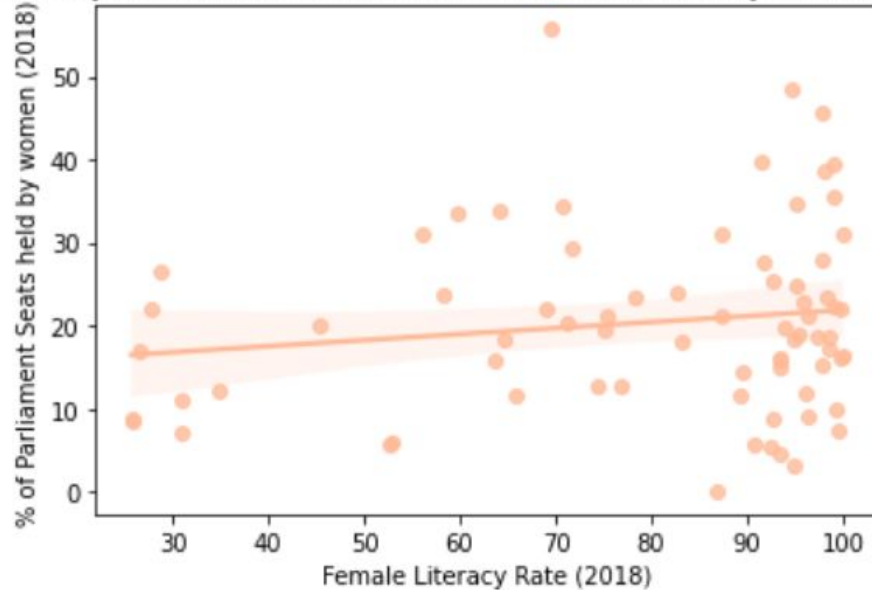
**Definition:** Proportion of seats held by women in the national parliament expressed as a percentage of total seats. For countries with a bicameral legislative system, the share of seats is calculated based on both houses.

**Source:** IPU (2020)

**Correlation Coefficient:** 0.15

# Reproductive Health vs Female Literacy



Literacy Rate (2018) vs. # deaths per 100,000 births (2017)
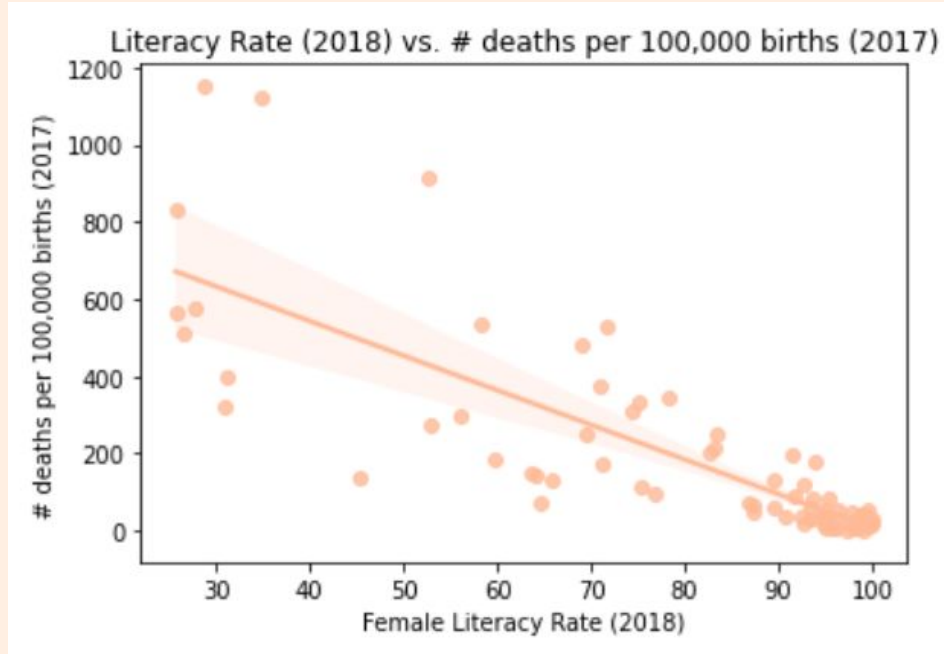
**Definition:** Number of deaths due to pregnancy-related causes per 100,000 live births.

**Source:** WHO, UNICEF, UNFPA, World Bank Group and United Nations Population Division (2019)

**Correlation Coefficient:** -0.80

# Sexual Health vs Female Literacy


Literacy Rate (2018) vs. % using contraceptives (2009-2019)
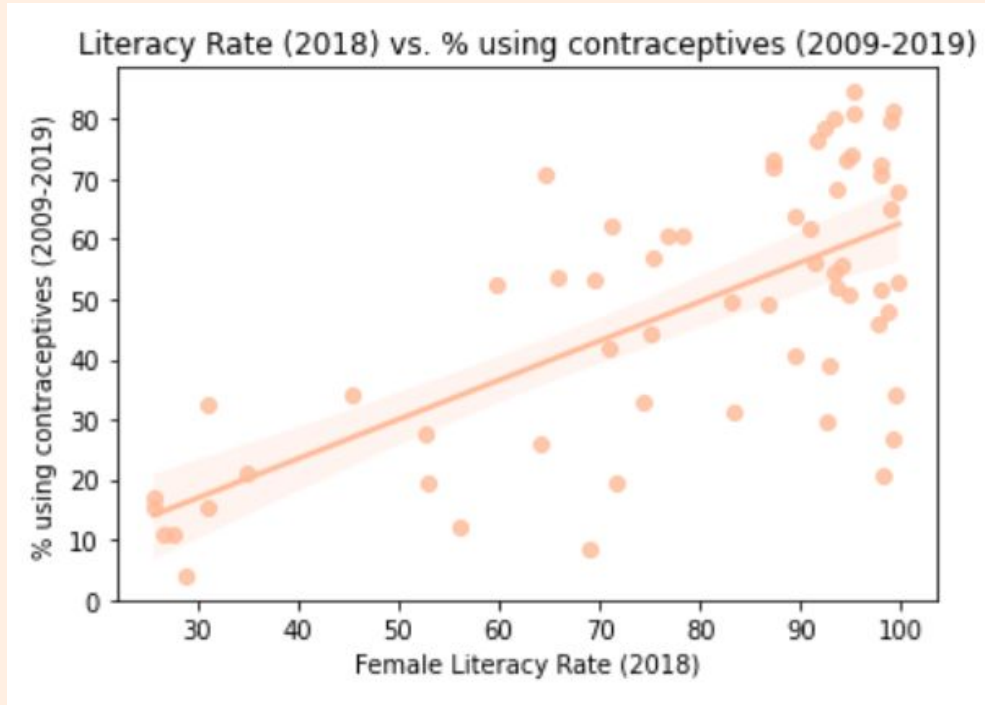
**Definition:** Percentage of married or in-union women of reproductive age (15–49 years) currently using any contraceptive method.

**Source:** UNDESA (2020). World Contraceptive Use 2020.

**Correlation Coefficient:** 0.68

# Financial Independence vs Female Literacy



Literacy Rate (2018) vs. % of women with bank accounts (2017)
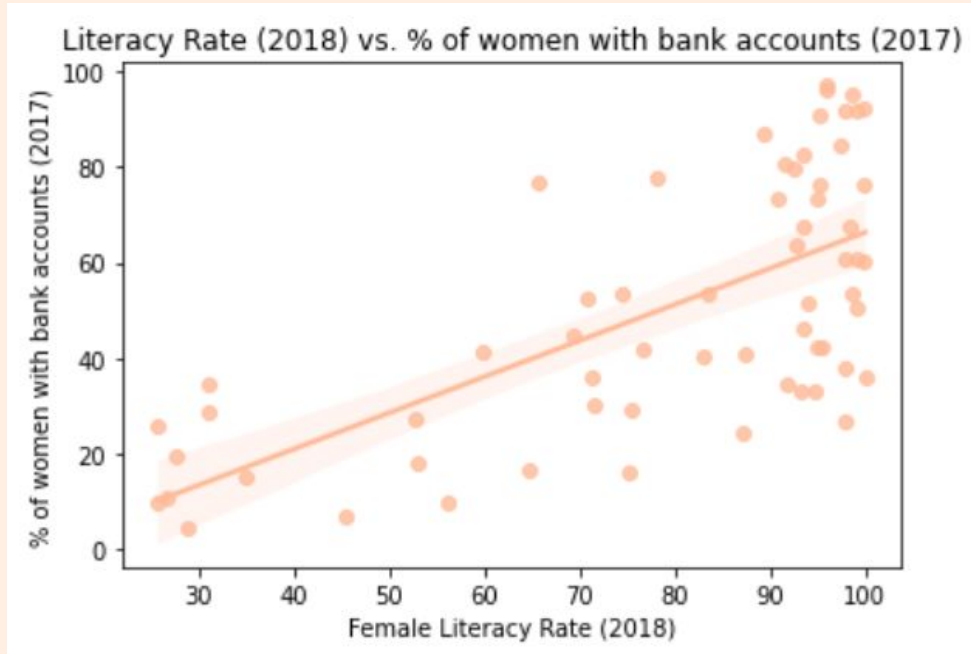
**Definition:** Percentage of women ages 15 and older who report having an account alone or jointly with someone else at a bank or other type of financial institution or who report personally using a mobile money service in the past 12 months.

**Source:** World Bank (2020b). Gender Statistics database. Washington, DC.

**Correlation Coefficient** = 0.6704994940060236

# Summary
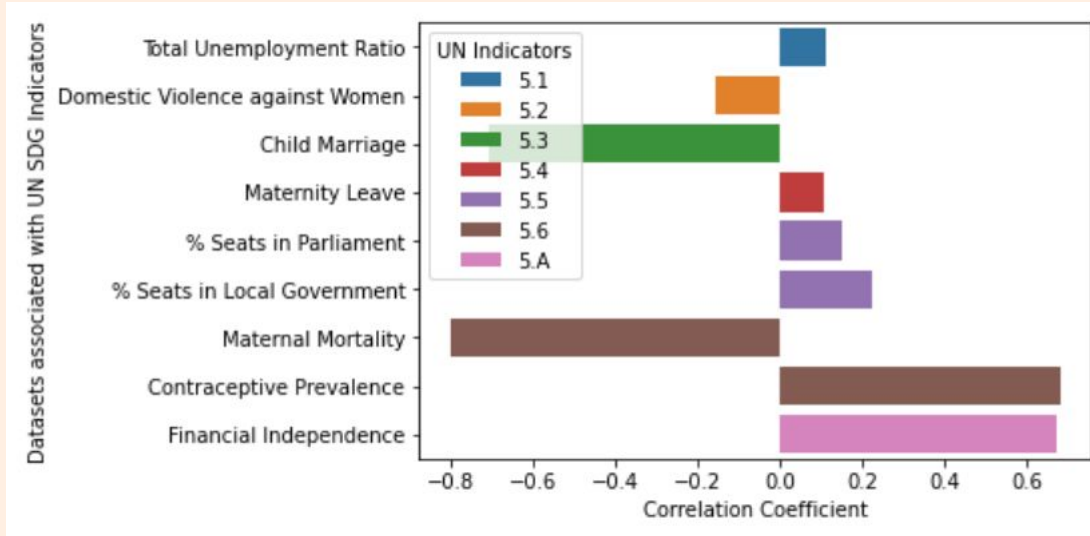


SDG Indicator 5.6 (Reproductive Health)
- ⊖ **Maternal Mortality** showed a strong negative correlation
- ⊕ **Contraceptive Prevalence** showed a strong positive correlation.

SDG Indicator 5.3 (Eliminate Harmful Practices)
- ⊖ **Child Marriage** showed a strong negative correlation

# 03 Analysis & Insights

# What is Multiple Linear Regression?

Multiple Linear Regression is a Machine Learning Model that allows us to take **many independent variables** ('features') and analyze their correlation to a **single dependent variable** ('target').

**MLR equation : $y = b_1x_1+b_2x_2+...+b_nx_n+b_0+e$**

$b_i$ are the **correlation weights** for each of the independent variables $x_i$ : if $b_1$ is twice the value of $b_2$, then an increase of 5 units in $b_1$ has the same effect on y as an increase of 10 units in $b_2$

# Correlation Weight of Each Indicator
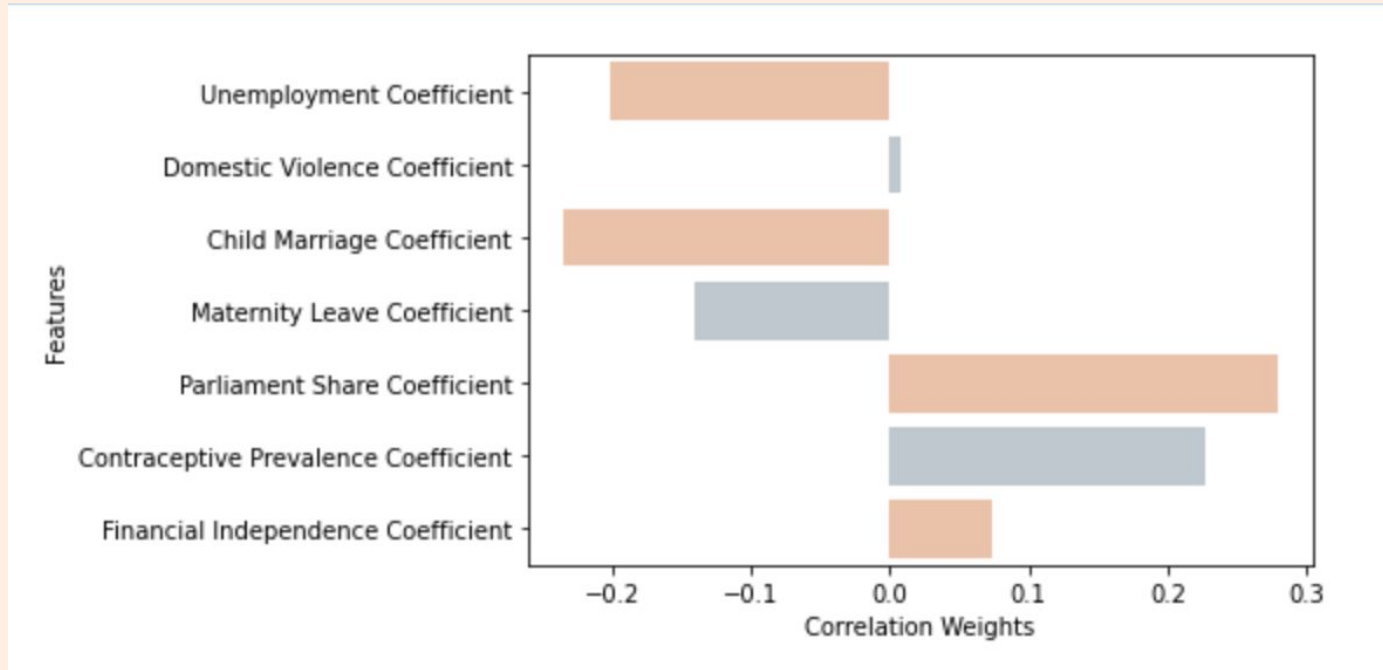


*all data in Standard Units

# Correlation Weight of Each Indicator
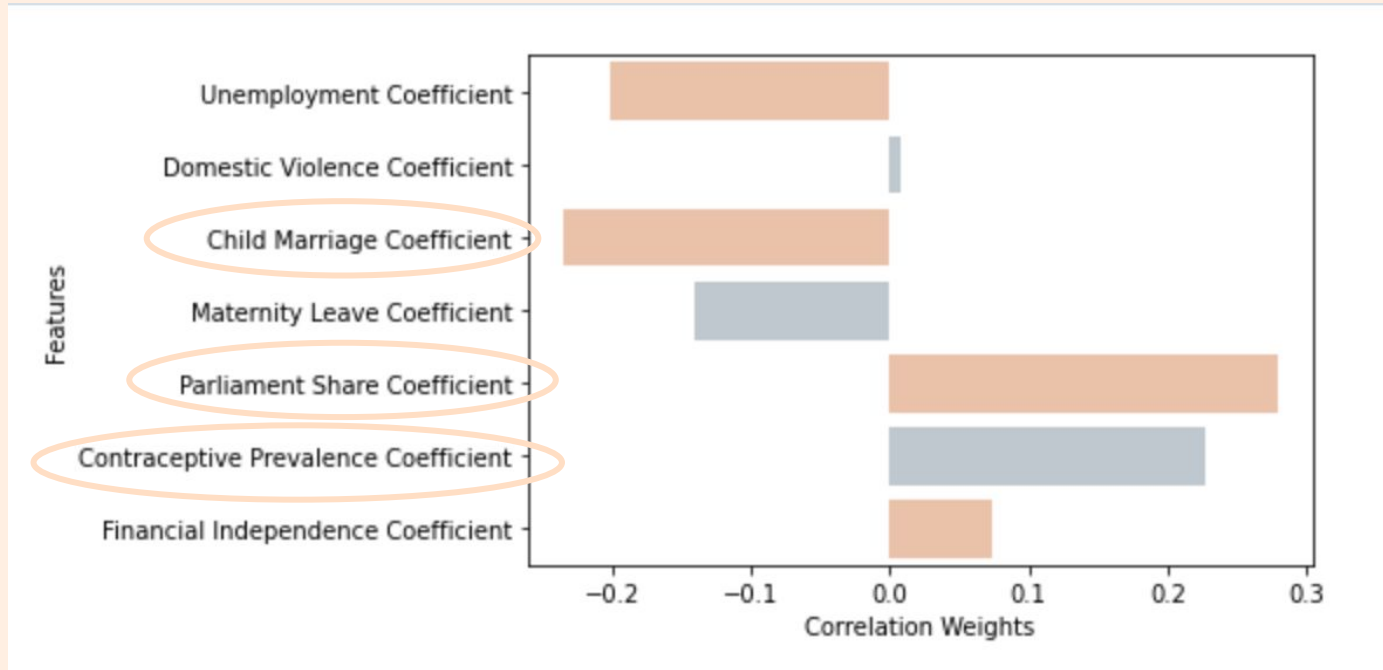


*all data in Standard Units

# Analysis & Insights

According to the correlation weights used by the best-fit multiple linear regression, we see that **child marriage**, **parliament share**, and **contraceptive prevalence** most affect female literacy rates globally.

# Challenges We Faced

### Narrowing Project Scope

Started with large theme of global development, specified to gender based on our skill sets

### Finding Data

Searched for trustworthy sources with relatively recent and comprehensive data

### Data Processing

Needed to clean missing and problematic values from data to prepare for our analysis

### Unintuitive Data

Due to limitations with our data and analytical techniques, some of our results are unintuitive

# Limitations

## Data Limitations

- We used the most recent data for each indicator, but not all were from the same year
- When merging data, some countries didn't have data in common, so those rows were eliminated in some data comparisons

## MLR Limitations

- Standardizing values helps compare across independent variables
- Differing distributions of these variables nonetheless impacted our final results

# Limitations

Visualization of standardized distributions of indicator data used in multiple linear regression
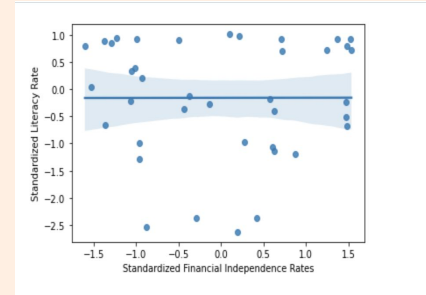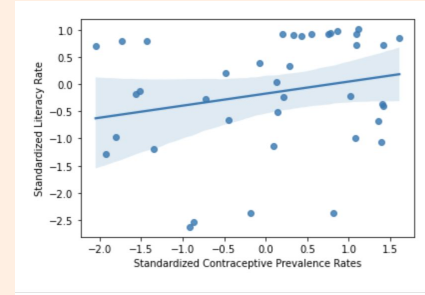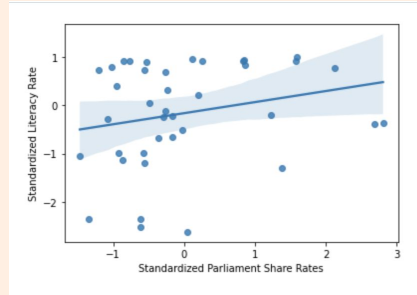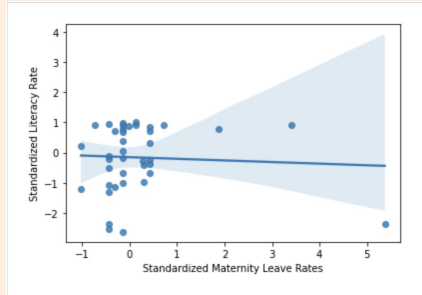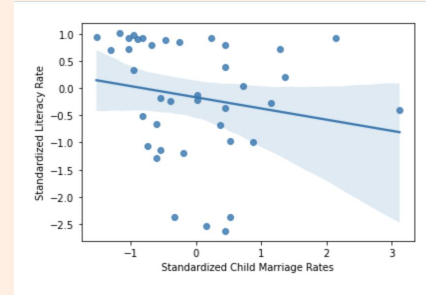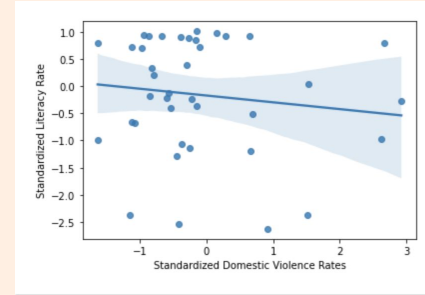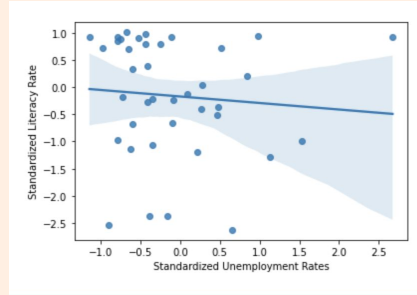
# Next Steps

04

# Next Steps

We recommend developing countries to prioritize UN Gender Equality Indicators 5.3, 5.4, 5.5: **reducing harmful practices,** ensuring **participation in leadership,** and universal access to **reproductive rights/health**, in order to accelerate global gender equality outcomes like female illiteracy.

# Thanks!