Analysis Report

Correlation analysis between Female School Education and Female Political Representation in America (Canada, United States, Guatemala, Honduras)

Laura Dauti

Methods of Advanced Data Engineering

Gender inequality in political representation is a significant issue because it stands in the way of creating more inclusive and representative governance systems.

Even though there has been progress in some regions, women are still underrepresented in political positions around the world. This ongoing imbalance makes it important to explore the factors that contribute to or hinder women's participation in politics.

This project focuses on whether there is a connection between female school education and female political representation in the Americas. It looks specifically at four countries: Canada, the United States, Guatemala, and Honduras. By examining these diverse contexts, the project aims to provide a broader perspective on how education might influence women's roles in politics.

To analyze this, the project uses detailed datasets and statistical methods to explore key indicators like expected years of schooling, enrollment rates in tertiary education, and the percentage of women in parliament over several years.

The main question for this project is: *Is there a correlation between a woman's school education and female political representation in the Americas (Canada, United States, Guatemala, and Honduras)?*

Used Data

For this analysis, a dataset was created using an SQLite database containing three tables: *Percentage of Shares of Women in Parliament* [1], *Percentage of Tertiary School Enrollment of Women* [2] and *Expected Years of Schooling for Women* [3]. The dataset was generated through an automated ETL pipeline implemented in Jayvee, ensuring data extraction, transformation and loading processes.

Data Description

The *Shares of Women in Parliament* table provides information on the percentage of seats held by women in parliaments from 2007 to 2022 for the selected countries: Canada, the United States, Guatemala, and Honduras. These countries were chosen to enable a comparative analysis between more developed countries in North America (Canada and the United States) and less developed countries in Central and South America (Guatemala and Honduras).

The *Tertiary School Enrollment* table contains the percentage of female enrollment in higher education institutions from 2008 to 2020, offering a substantial time span for analysis.

Similarly, the *Expected Years of Schooling* table provides data on the average number of years a female student is expected to remain in school. However, this data is only available for four specific years: 2010, 2017, 2018, and 2020, making it less temporally representative compared to the other datasets.

The final dataset output, produced by the pipeline, is largely complete except for minor gaps in data coverage for certain years. All irrelevant features have been removed, and the data has been cleaned to ensure consistency and accuracy for the analysis.

Data Structure

The three tables are standardized in structure, with columns for *Country Name*, *Year*, and *Value*. This consistency enables direct comparisons and consistent interpretation of results.

- **Shares of Women in Parliament:** Captures the proportion of parliamentary seats held by women for over a decade, providing a robust basis for analyzing political representation trends.
- **Tertiary School Enrollment:** Spanning from 2008 to 2020, this table illustrates changes in female participation in higher education across the chosen countries.
- **Expected Years of Schooling:** While limited to four years, this data still offers valuable insights into educational expectations for female students.

A sample from the dataset is provided below in Figure 1.

Country Name	Year	Value
Filtern	2020 ✕	Filtern
Canada	2020	28.994083
Guatemala	2020	19.375
Honduras	2020	21.09375
United States	2020	27.464788

Figure 1: *Sample from Shares of Women in Parliament table*

License

The data set for the *Percentage of Shares of Women in Parliament* table is provided by "World Bank Group" and the data sets for the other two tables are both provided by "World Bank's Gender Data Portal".
All three data sets are completely open access under the Creative Commons BY-4.0 license.

Analysis

To explore how female education relates to political representation in North and Central America, several data analyses, including Exploratory Data Analysis (EDA) and Correlation Analysis, were conducted. The goal was to identify any patterns between women's education levels and their representation in politics. These findings provide useful insights into gender equality in both education and politics.

Exploratory Data Analysis

Exploratory Data Analysis helps understanding the data, spotting patterns, and identifying any outliers. It helps in getting a clearer picture of the data, making it easier to draw conclusions.

- **Methodology:** For this analysis, the *Tertiary School Enrollment* data was compared with the *Shares of Woman in Parliament* data across the four countries with the help of line charts. By plotting one line chart for each country, trends over time were observed to see how education impacts political representation.
  To make the data comparable, the years and values for all countries were aligned.
- **Results:** For all four line charts the years are on the x-axis, and the values are on the y-axis. The plot in Figure 2 provides the analysis for Canada. The same plot charts were done for Guatemala, as shown in Figure 3, for Honduras as in Figure
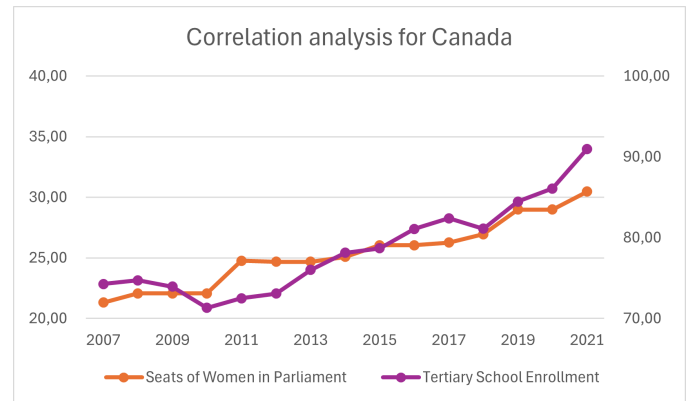

Figure 2: *Line chart for Canada (Tertiary School Enrollment and Shares of Woman in Parliament)*

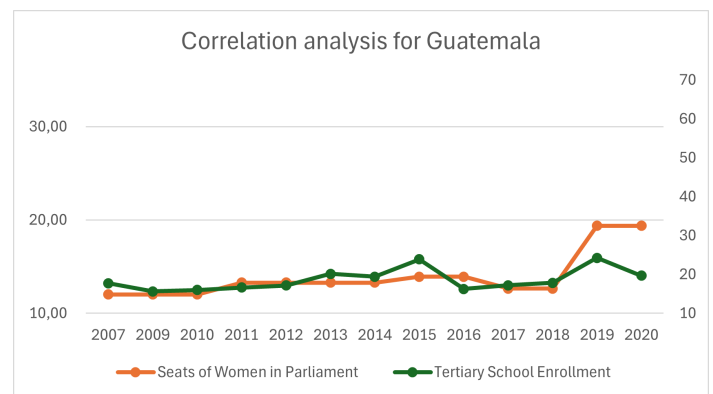4 and lastly for the United States, as provided in Figure 5.

Figure 3: *Line chart for Guatemala (Tertiary School Enrollment and Shares of Woman in Parliament)*
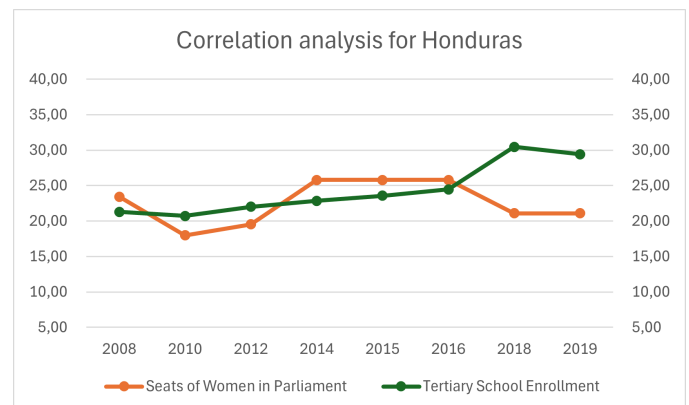

Figure 4: *Line chart for Honduras (Tertiary School Enrollment and Shares of Woman in Parliament)*

- **Interpretation:** The Exploratory Data Analysis reveals interesting insights into the correlation between female representation in parliament and a higher education. For Canada and Guatemala there was a clear positive relationship. As the percentage of female tertiary school enrollment
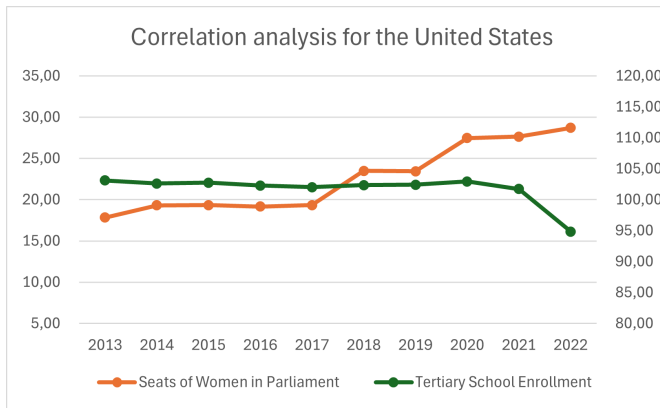
Figure 5: *Line chart for the United States (Tertiary School Enrollment and Shares of Woman in Parliament)*

Figure 6: *Combination chart for 2018 (Expected Years of School and Shares of Woman in Parliament)*

increased, the percentage of seats of women in parliament also increased for both countries. However, Honduras and the United States showed a less clear relationship, with little variations in increasing and decreasing percentage. Both datasets show a difference after the year 2017. While in Honduras the seats of women in parliament after 2017 decreased, the female tertiary school enrollment increased. On the contrary, the seats of women in parliament in the United States after 2017 increased, but the female tertiary school enrollment decreased.

For getting a clearer picture, another Exploratory Data Analysis was done with the *Expected Years of School* dataset. Here, a combination of column charts and line charts was used.

- **Methodology:** For this analysis, the *Expected Years of School* data was compared with the *Shares of Woman in Parliament* data across the four countries. By plotting one chart for 2018 and one for 2020, it is helping to see the correlation between the years, a female student is expected to remain in school and the seats of women in parliament.
- **Results:** The chart for the year 2018 is shown in Figure 6, while Figure 7 provides the chart for 2020.
- **Interpretation:** The charts for this analysis show a similar result as the line charts before. Here, the expected years of school and the seats of women in parliament align for Canada, the United States and more or less for Guatemala for both years. However, as in the line charts before, the expected years of school don't quite align with the seats of women in parliament for Honduras.
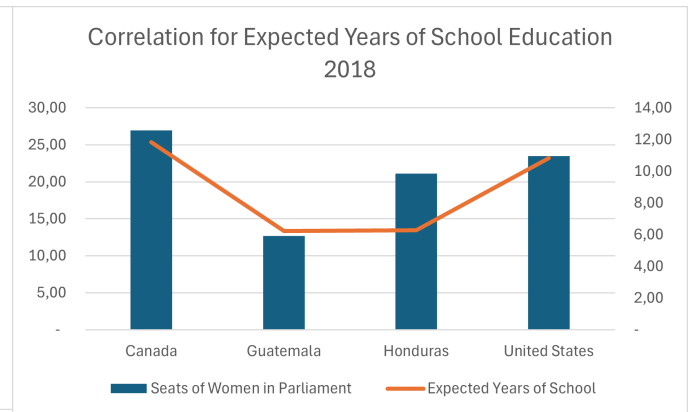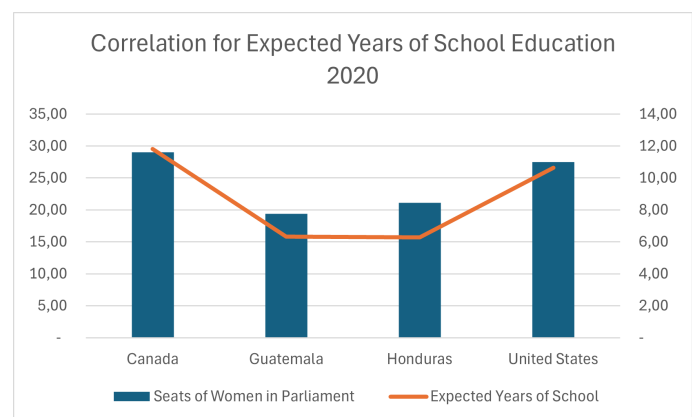

Figure 7: *Combination chart for 2020 (Expected Years of School and Shares of Woman in Parliament)*

Correlation Analysis

After the Exploratory Data Analysis, a Correlation Analysis was calculated. Correlation Analysis helps identifying relationships between numerical features, providing insights into how changes in one feature may relate to changes in another. To achieve this, Correlation Coefficients, the so called Pearson Correlation Coefficients, were calculated to assess the relationship between female education and political representation in the four countries. The result has a value between -1 and 1, where values close to 1 indicate a strong positive correlation, values near -1 represent a strong negative correlation, and values around 0 suggest little to no correlation. The Correlation Coefficients between the values of *Tertiary School Enrollment* and the *Seats of Women in Parliament* were calculated.

- **Canada: 0.91** - A strong positive correlation, indicating that as female education levels increase, political representation also rises significantly.
- **Guatemala: 0.62** - A moderate positive correlation, showing a noticeable, though weaker, relationship between education and political repre-

sentation.

- **Honduras: -0.03** - A very weak negative correlation, indicating that there is almost no relationship between education and political representation.
- **United States: -0.55** - A moderate negative correlation, implying an inverse relationship between education and political representation. In the United States, while education levels may rise, political representation does not always follow.

Conclusions

The analysis addressed the central question: *Is there a correlation between a woman's school education and female political representation in the Americas (Canada, United States, Guatemala, and Honduras)?* The results present a diverse picture that varies considerably across the analyzed countries.

In Canada and Guatemala, the analysis found a positive correlation, suggesting that higher levels of female education are associated with greater political representation.

In contrast, the findings for Honduras and the United States are less straightforward. Honduras displayed no meaningful correlation, while the United States showed a moderate negative correlation, indicating that increases in education levels did not translate into greater political representation. These results highlight that education alone may not be enough to overcome barriers such as political, cultural, or economic factors that hinder women's political participation.

Although the analysis provided valuable insights, some limitations and uncertainties remain to consider. The correlation analysis, though useful, does not imply causation. Factors such as social norms, economic conditions, and institutional frameworks also play crucial roles in shaping political representation. Additionally, the dataset, though comprehensive, had some gaps - especially in the temporal coverage of certain indicators like expected years of schooling. This limitation may have affected the robustness of the conclusions, particularly for Honduras and the United States.

Moreover, the focus on only four countries means that these results cannot be generalized to the entire Americas. Expanding the study to include more nations could offer a broader perspective and more representative trends.

This project shows that education can play a big role in improving gender equality in politics. In countries like Canada and Guatemala, it's a strong factor driving progress. But education alone isn't enough - it must be complemented by other measures, such as cultural and systemic barriers to make politics truly inclusive. For places like Honduras and the United States, it's clear that a more in-depth analysis is needed to unterstand how education, social norms, and political opportunities work together to influence women's participation.

References

[1] Multiple sources compiled by World Bank – processed by Our World in Data. *"Share of women in parliament" [dataset]. Inter-Parliamentary Union (via World Bank), "World Development Indicators" [original data]*. 2022. URL: https : / / ourworldindata . org / grapher / share - of - women - in - parliament - ipu (visited on 11/27/2024).

[2] World Bank Group - Gender Data Portal. *School enrollment, tertiary (% gross)*. 2024. URL: https : //genderdata.worldbank.org/en/indicator/ se-ter-enrr (visited on 11/27/2024).

[3] World Bank Group - Gender Data Portal. *Learning-Adjusted Years of School*. 2020. URL: https : / / genderdata . worldbank . org / en / indicator / hd - hci - lays (visited on 11/27/2024).