The Widening Gap: Why Rising Female Secondary Education Fails To Translate into Labor Force Inclusion (1990-2019)

1. Introduction

Exploratory visual analysis helps us discover relationships in large and complex datasets by breaking down the analysis into many steps (known as iterative posing) and answering them individually.

This paper looks at a fundamental relationship in global development: how the female population's access to education relates to their participation in the labor force. This hypothesis is central to global development goals and posits that the higher levels of female secondary school enrollment lead to increased female labor force around the world.

Over the past three decades , global efforts have successfully raised educational access for women. However, educational development has had varying effects on the economy. This goal of this analysis is to move beyond a simple global correlation to understand whether this relationship is context-dependent and how it has evolved.

Research question: How does female education (measured by secondary school enrollment) relate to female labor force participation across countries from 1990 to 2019?

2. Data Profile

The data used in this analysis comes from the World Bank's World Development Indicators (WDI) - a comprehensive dataset containing over 1,000 economic, environmental, and social indicators for more than 200 countries.

The specific extract used for this project included the following indicators:

Indicator	Code	Description
Female labor force participation rate (% of female population ages 15+)	SL.TLF.CACT.FE.ZS	Measures women's participation in the workforce
Female secondary school enrollment (% net)	SE.SEC.NENR.FE	Indicates educational access for women at the secondary level
Population, total	SP.POP.TOTL	Used as a proxy for country size in bubble plots
GDP per capita (current US$)	NY.GDP.PCAP.CD	Used for contextual understanding of income levels

2.1 Data quality and constraints

The initial analysis was meant to explore the timeline from 1990-2020. However, this was not possible on the raw data as valid data was not present for the Female Secondary School Enrollment in 2020. To prevent listwise deletion that would render incorrect visualisations, the decision was made to change the timeline from 1990-2019. Data prep involved standard Python procedures using pandas, including melting year columns into rows, pivoting to align indicators and merging with the World Bank region metadata.

3. Methodology

The analysis followed a three step iterative process. It moves from global assessments to context-specific conclusions. The strategy was meant to focus on regional nuance as opposed to simple global correlations. Python (Pandas and Seaborn) was used for all data wrangling and visualization generation, ensuring reproducibility.

3.1 The Necessity of Regional Grouping

The decision to group analysis by the World Bank Region was strategic.A global average suggests uniform progress, while regional grouping provides the necessary proxy for localized socio-cultural norms, policy environments, and the size of the informal labor sector.

Creating small multiples helps us visually prove that the education-participation link is not a universal law but a context-dependent outcome.

4. Question Exploration

Iteration 1: Cross-Sectional Comparison

The first iteration focused on establishing the relationship between the three key milestones: starting at 1990, a midpoint check-in in 2005 and ending in 2019. This was executed using a small multiples trajectory plot which plots Education Enrollment (x-axis) against Labor Participation (y-axis) and groups the data by region (categorical dimension) .

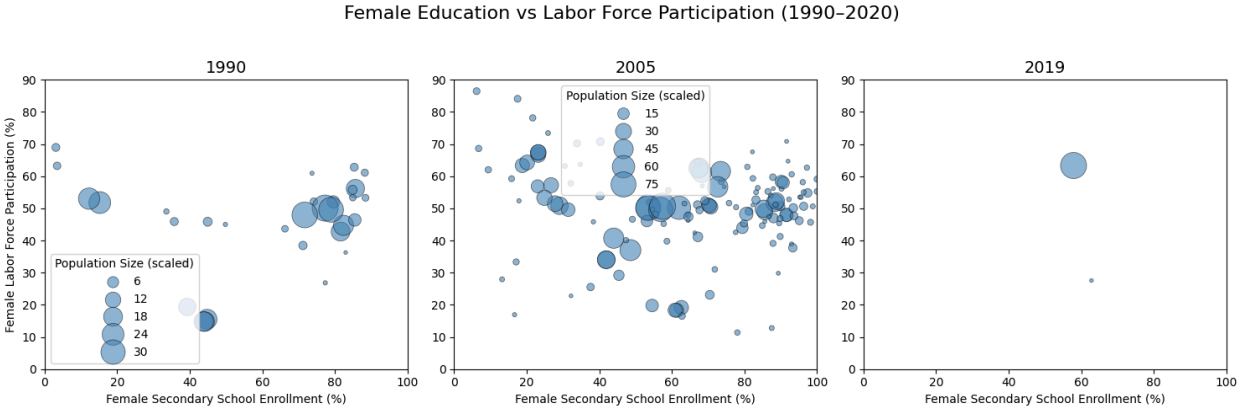

Figure 1: Relationship between female education and labor force participation across countries in 1990, 2005, and 2019.

The core conclusion from this iteration is that the education-participation link is fundamentally context-dependent, negating a single global hypothesis. While the overall data showed a general movement to the upwards right direction (confirming a correlation), the regional trajectories immediately revealed critical differences. Regions like Europe & Central Asia supported the hypothesis with trajectories moving up and to the right from 1990 to 2019 showing simultaneous growth in both indicators (position used as a quantitative channel). In contrast, South Asia provided the most critical contradiction: countries showed a sharp movement to the right (increased education enrollment) but remained largely flat on the y-axis, demonstrating minimum to no labor participation. This visual evidence, combined with the high but also stagnant rates in Sub-Saharan Africa proved the existence of significant structural or cultural barriers that prevent educational gains from translating into economic inclusion. The discovery of this regional plateauing demanded subsequent temporal analysis to quantify this global failure rate. This iteration also made it obvious that a single aggregated scatterplot would suffer from overplotting and insufficient separability of marks, necessitating faceting by region.

Iteration 2: Temporal Trends

Iteration 2 isolated the global trends to define the analytical puzzle: why is the correlation breaking down? A single-axis line chart was used to compare the global average of both indicators over time (using line marks and continuous position channels). The global average for Female Secondary School Enrollment showed strong , near-linear growth, climbing from approximately 50% in 1990 toward 80% in 2019.However, the global average for Female Labor Force Participation remained stagnant, consistently hovering between 50% and 52% over the same three decades. This growing vertical distance between the two lines definitively proved that, at the global level, educational gains are not translating into proportional employment growth, thereby necessitating a shift away from the simple global hypothesis to a detailed regional breakdown. By encoding both indicators in the same coordinate system, the visualization made Gestalt principles like proximity and similarity more salient, helping reveal the widening gap as a pre-attentive pattern.

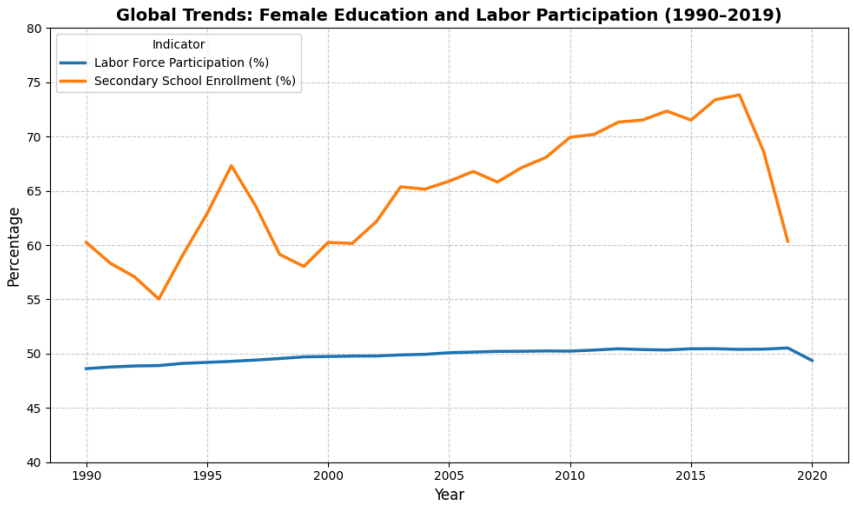

Figure 2: Global trends in female secondary education and labor force participation (1990–2020).

Iteration 3: Regional Participation Trends

To address the participation plateau identified in Iteration 2, we shift focus to contextualizing the Labor Participation rate itself. This iteration utilized a multi-line chart, plotting the time-series trend of the regional average of Female Labor Participation (LFP) (using color as a categorical channel and position as a quantitative channel). The visualization immediately fragmented the global average, showing that Sub-Saharan Africa maintained the highest Female Labor Force Participation (LFP) rates (near 60%), while South Asia and MENA clustered at the lowest end (below 30%). Crucially, only Latin America & Caribbean and Europe & Central Asia demonstrated clear, upward-sloping LFP trends, confirming that the magnitude of participation change is entirely dependent on overcoming region-specific structural barriers. The magnitude and direction of LFP change are entirely dependent on the region. This facet-free multi-line design leveraged the viewer's ability to detect slope differences pre-attentively, making regional divergence perceptually obvious.

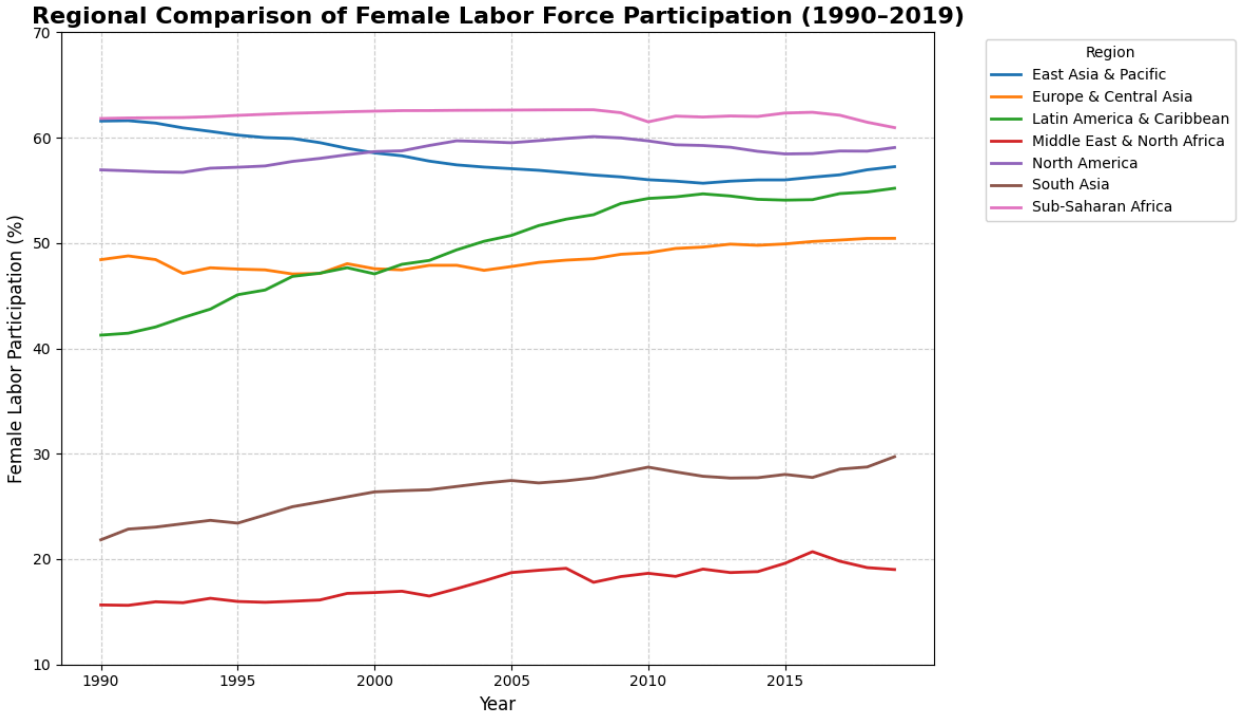

Figure 3. Regional Comparison of Female Labor Participation (1990-2019)

5. Reflection on Iterative process

In iteration 1 , a global scatterplot would have concealed structural issues. Using a Small Multiples Trajectory Plot makes the regional divergence immediately apparent. Similarly, while the initial dual-subplot approach for iteration 2 was statistically sound, the final single axis plot was chosen because it was superior at visually communicating the *magnitude* of the problem—the widening gap—which served as a stronger justification for the regional dive.Ultimately, the process moved from correlation establishment (Iteration 1) to problem definition (Iteration 2) to contextualization (Iteration 3). The final visualization was designed to synthesize these three distinct layers of insight into one conclusive statement.

6. Final Visualisation

The final visualization chosen to conclude the analysis is a Slope Graph . This chart effectively ties together the findings from all three iterations by focusing purely on the net change in the most critical variable (LFP) across the most critical grouping (Region) over the entire time period (1990–2019).

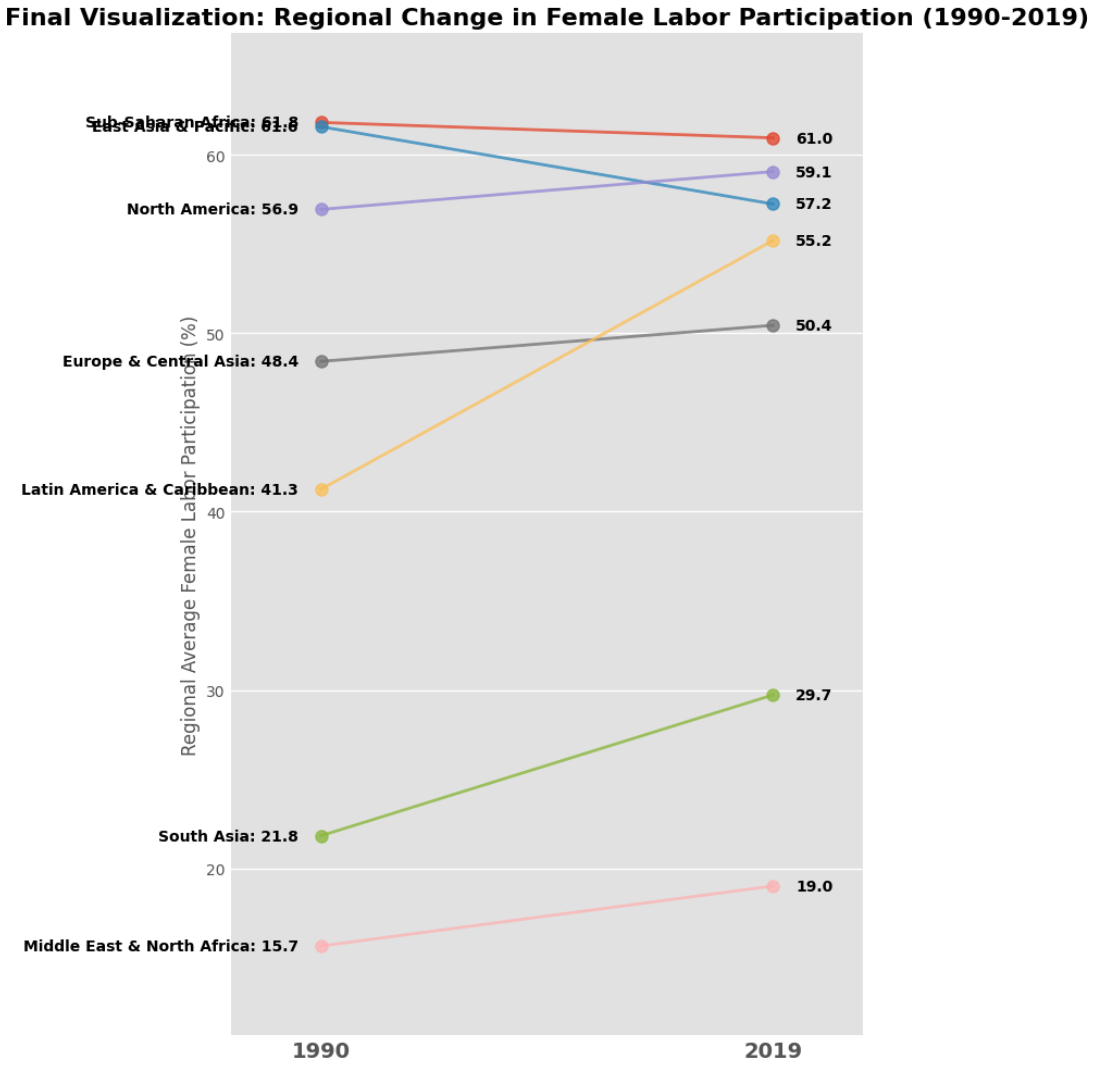

Figure 4: Regional Changes

The slope of each line visually represents the rate of progress in female labor participation for that region. Steep Positive Slopes (like those in Latin America & Caribbean, Europe & Central Asia) show substantial LFP growth, driven by successful economic and policy shifts. Flat or Slightly Negative Slopes (South Asia, Sub-Saharan Africa) demonstrate stagnation or decline, providing the clearest visual evidence that the regional structural barriers observed in Iteration 1 have persisted over three decades.This simple, powerful chart serves as the final, unambiguous proof that a regional policy approach is necessary to unlock the economic potential promised by rising educational attainment

7. Discussion and Conclusion

The exploratory visual analysis successfully demonstrated that the relationship between female education and labor force participation is highly context-dependent and that a simple global average is fundamentally misleading. The analysis supports a nuanced version of the original hypothesis: female education and labor force participation are positively correlated globally, and this relationship has strengthened over the past three decades. However, the data also reveals significant limitations—in regions like South Asia, rising education levels have not yet translated into equivalent labor market inclusion, suggesting persistent and entrenched structural barriers.The stagnation in LFP (Figure 4) in this cluster, despite booming educational enrollment, is the most salient finding of the project. It underscores the fact that education acts as a necessary, but insufficient, condition for economic empowerment. It must be accompanied by supportive policy and economic infrastructure, such as subsidized childcare, flexible work arrangements, and anti-discrimination laws, to effectively bridge the gap.This project affirms the power of iterative visualization to transform a statistical observation (a widening gap in Figure 2) into a policy-relevant conclusion. The clear regional segregation in the final visualizations provides empirical justification for moving toward highly region-specific development strategies to achieve true gender equality in the global workforce.

7. Final Reflection

This project highlighted how essential iterative visualization is to exploratory analysis, especially when dealing with multivariate data where initial encodings often fail to reveal meaningful structure. My early global scatterplots suffered from overplotting and ineffective mark–channel mappings, which pushed me to adopt small multiples and faceted views to reduce cognitive load and reveal regional divergence. Working in Python highlighted the tight link between data wrangling and visualization clarity, as missing or inconsistent data required careful transformation before any meaningful insight could emerge. Overall, the iterative process reinforced that effective exploratory visualization is not about finding the “right chart” immediately, but about refining encodings step by step until the visualization aligns with both the question and human perceptual strengths.