

Explore the Relationship Between Global Female Education Levels and Fertility Rates

06010567

Compiled: May 11, 2025

Github Repo: [REPLACE-WITH-LINK-TO-YOUR-TAGGED-RELEASE](#)

1 Project Description (approx. 250 words)

This project investigates the relationship between female education and fertility rates globally from 2000 to 2023. Motivated by demographic and development policy implications, the project leverages data from the World Bank (WDI) and UNDP Human Development Report (HDR), incorporating indicators such as fertility, female education at multiple levels, GDP, urbanisation, labour participation, contraceptive prevalence, and gender equality metrics.

The data pipeline includes robust wrangling, interpolation-based imputation, and feature engineering (e.g., log-transformed GDP). Exploratory analysis reveals strong negative associations between education and fertility, with notable variation across income groups and continents.

Modeling is conducted via multivariate linear models and panel regressions using ‘plm’ in R. Interaction terms (e.g., income group \times education) highlight heterogeneous effects. Model predictions are visualised to illustrate policy-relevant thresholds and scenarios.

A Shiny app allows users to explore education–fertility trends interactively by country and year. The app includes scatter plots with optional linear trend lines, world maps, radar profiles of gender-related indicators, and a simple simulation-based forecasting feature for 2024–2026 based on user-defined inputs.

The project is fully reproducible with a ‘make.R’ script, and documented through modular scripts, structured folders, and an annotated ‘README.md’. Overall, the product combines analytical depth, usability, and reproducibility to support insight generation for global education and fertility dynamics.

2 Assessment Criteria

Technical Competence: Proficiency in data collection, processing, analysis, and coding.

- Integrated and cleaned panel data using ‘WDI’, ‘readxl’, and ‘zoo::na.approx’, generating new variables (e.g., log-GDP).
- Built linear and panel regression models with interactions to account for heterogeneous effects.

User Interface: Design, functionality, and usability of the final data product.

- Designed a Shiny app with tabs for trend analysis, mapping, radar profile, and statistical summary.
- Allowed users to simulate forecasts for 2024–2026 based on editable socio-demographic parameters.

Analysis and Interpretation: Depth of analysis, appropriate use of statistical methods, and meaningful interpretation.

- Identified strong inverse associations between education and fertility, with variation across income groups.
- Used modeling outputs and ‘ggeffects’ to interpret nonlinear and interactive effects in accessible form.

Presentation and Communication: Clarity, organisation and effectiveness of written and visual communication.

- Organised all deliverables into a clear file structure, rendered with ‘bookdown’, and summarised via Markdown and LaTeX.
- Produced clear plots with appropriate color schemes, labels, and interactive enhancements (e.g., Plotly).

Reproducibility and Documentation: Clarity and completeness of documentation for product use and reproducibility.

- Complete pipeline automation via ‘make.R’; modular and readable R scripts.
- Clear ‘README.md’ with setup instructions, dependency list, and visual project tree.

Project Management: Considered and effective use of project management and version control systems.

- Managed progress in stages (data, EDA, modelling, dashboard), with each part committed to GitHub.
- Final product released as tagged GitHub version and zipped project archive.

3 Project Reflection

Reflect on the experience of creating your data product. In 6 bullet points and at most 1 page total, summarise the following.

- *3 things you have learned as part of this process,*
- *2 aspects of the project that you found challenging or would approach differently with hindsight,*
- *1 aspect of the project that you would like to learn more about in the future.*

Learnings:

- I deepened my understanding of how to build and validate panel regression models using ‘plm’.
- I learned how to design and deploy an interactive Shiny dashboard that integrates modeling and forecasting.
- I improved my ability to clean and impute real-world multi-source data with non-trivial structure.

Challenges:

- Balancing generalisability and overfitting in panel models with interactions was difficult.
- Designing intuitive but flexible UI components in Shiny required several iterations and testing.

Further Development:

- I would like to explore dynamic panel models and more advanced forecasting techniques for longitudinal data.