Global Total Fertility Rates versus GDP per Capita in 2020

Ibari J. Nwosu

August 24, 2025

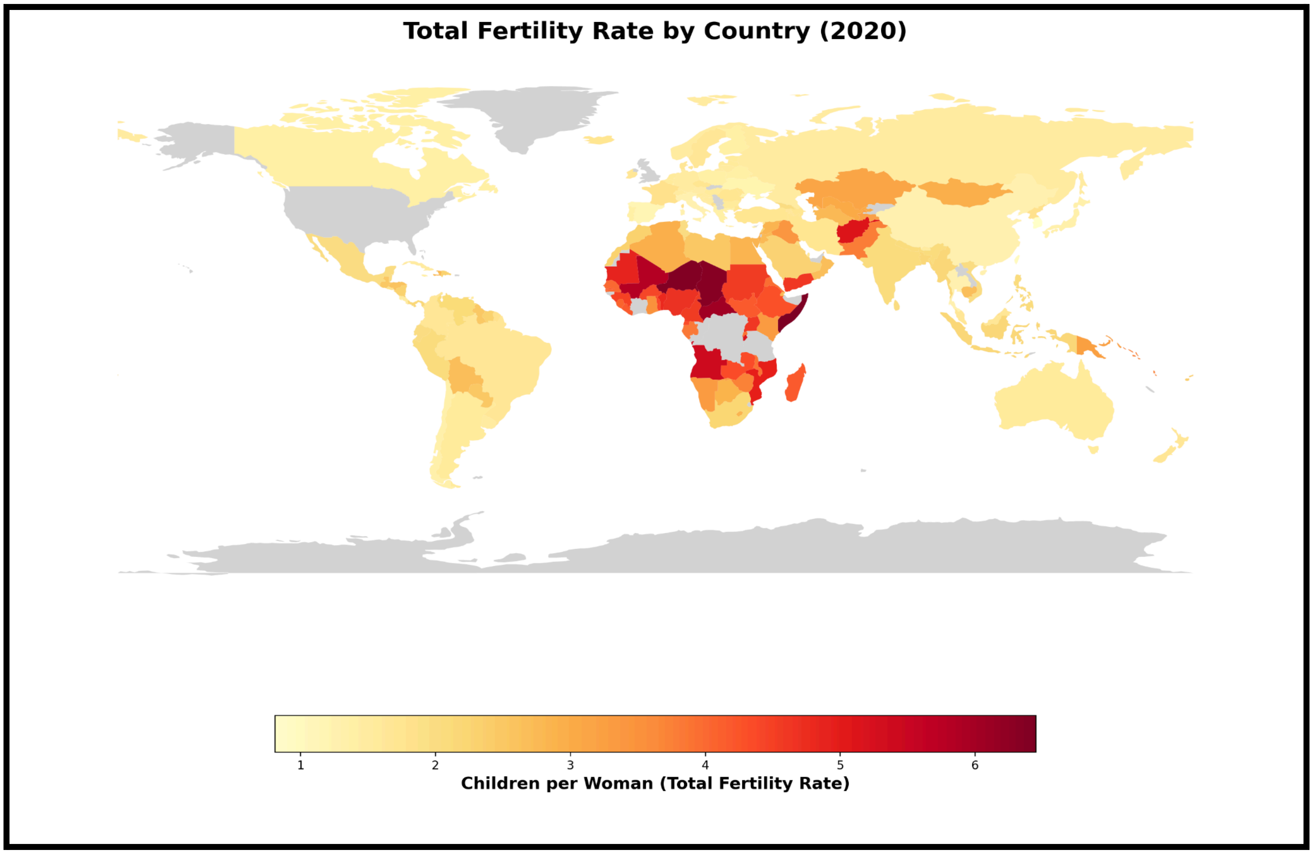
Introduction: This report analysed the relationship between GDP per Capita and Total Fertility Rates (# children per woman) globally in 2020, using data downloaded from the Gapminder website.

Methods: The analysis was carried out using Python in Julius, working with the Claude IV Sonnet LLM, and using the prompts provided in the course materials.

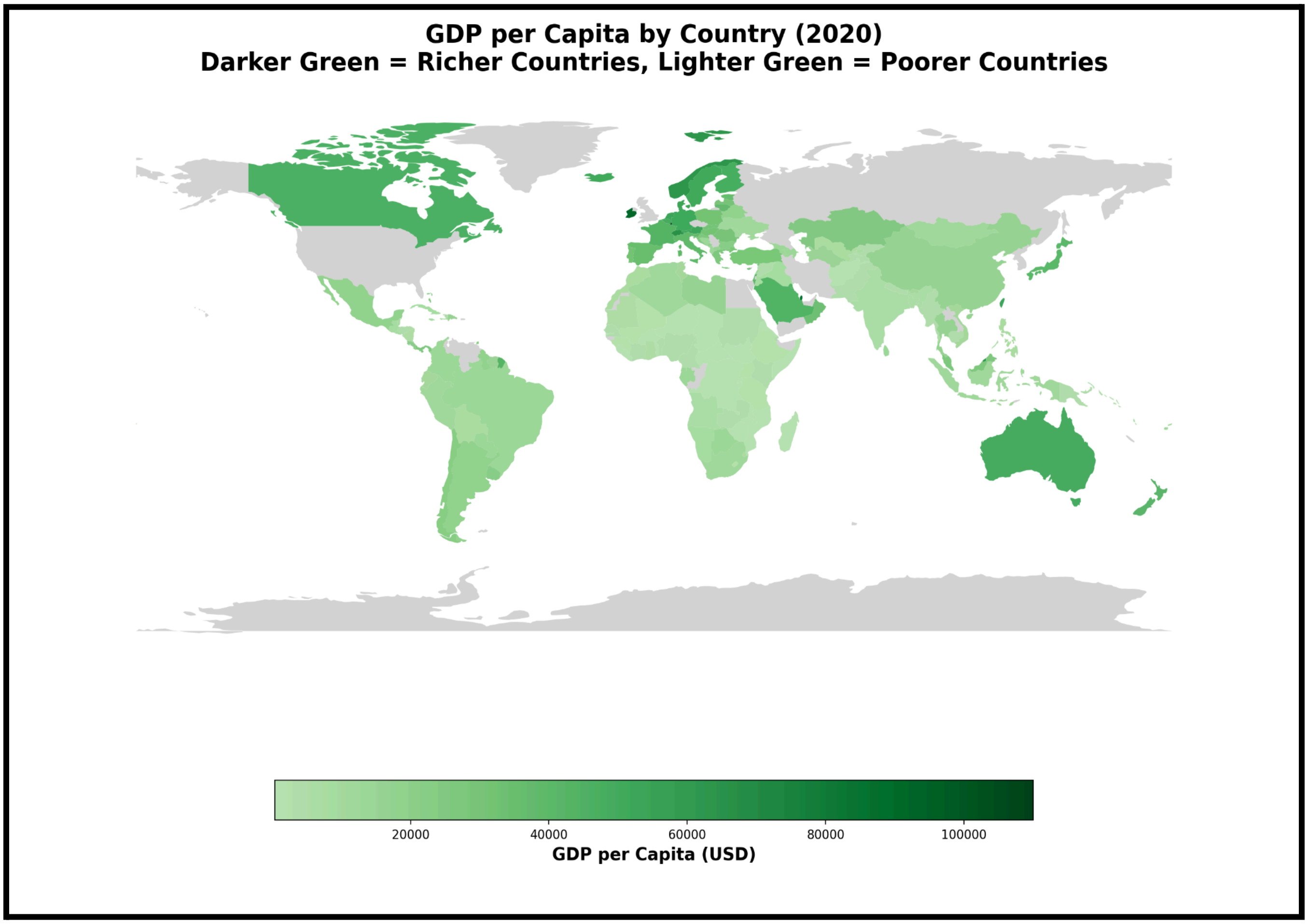
General:
The fertility dataset contained 197 countries and the GDP dataset contained data for 195 countries, spanning from 1800 to 2100. For 2020, the fertility data was complete with no missing values, while 150 countries had GDP data.

Analysis:

a. **Indicator 1: Total Fertility Rates (2020):** Higher fertility rates were generally observed in Sub-Saharan Africa and some parts of Asia, with lower rates in Europe and the Americas.

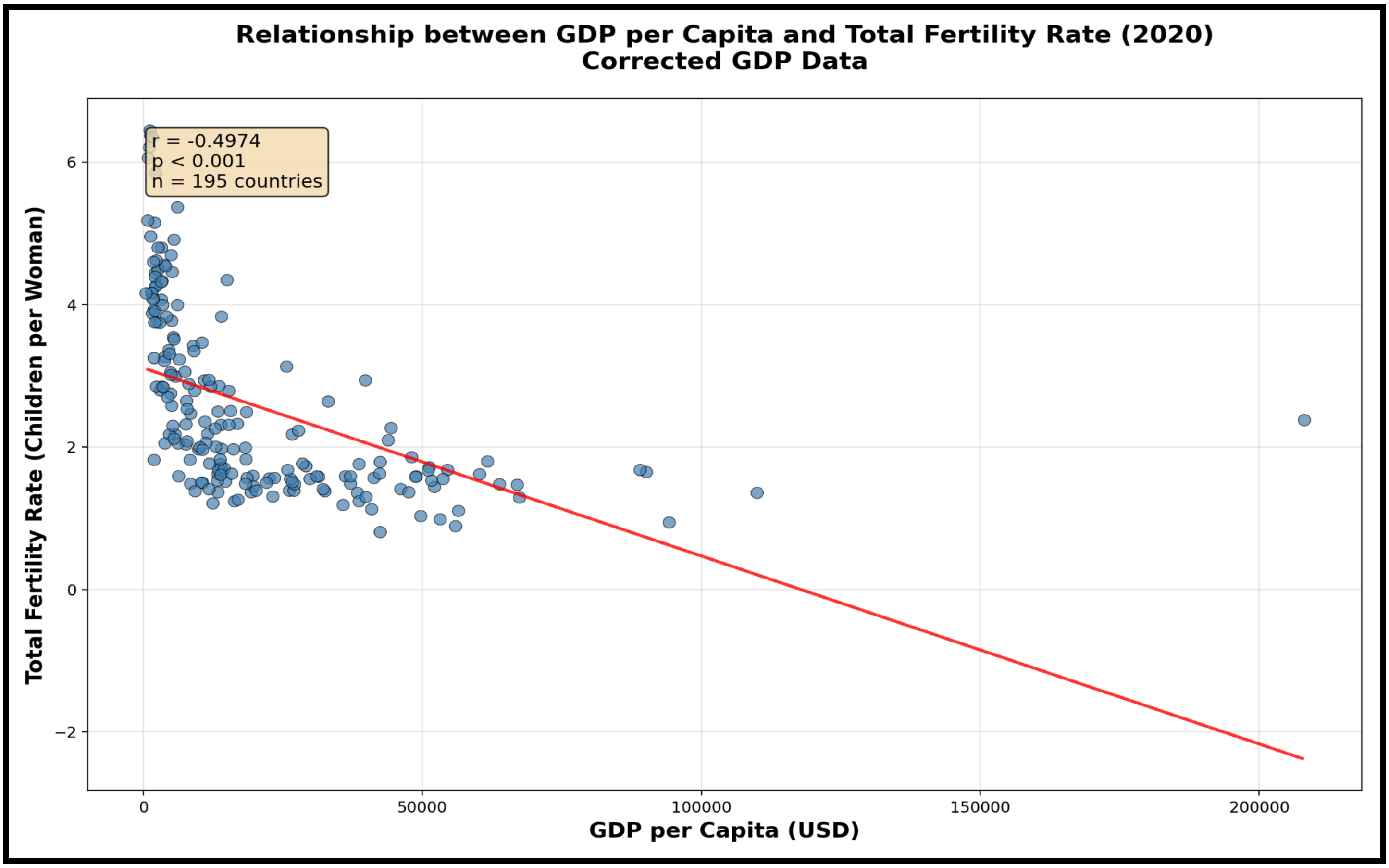

Total Fertility Rate by Country (2020)
1
2
3
4
5
6
Children per Woman (Total Fertility Rate)

b. **Indicator 2**: **GDP per capita (2020):** The global wealth distribution showed the wealthiest nations concentrated in North America, Europe, and parts of Asia and Oceania. Lower income countries predominated in Africa, Asia and to a lesser degree South America.


GDP per Capita by Country (2020)
Darker Green = Richer Countries, Lighter Green = Poorer Countries
20000
40000
60000
80000
100000
GDP per Capita (USD)

c. **Scatterplot showing the relationship between GDP per Capita and Total Fertility Rate (2020) (diagram overleaf)** .

- Number of countries analyzed: 195
- Pearson correlation coefficient (r): -0.4974
- R-squared (explained variance): 0.2474
- P-value: 1.39e-13
- Statistical significance: highly statistically significant ($p < 0.001$)

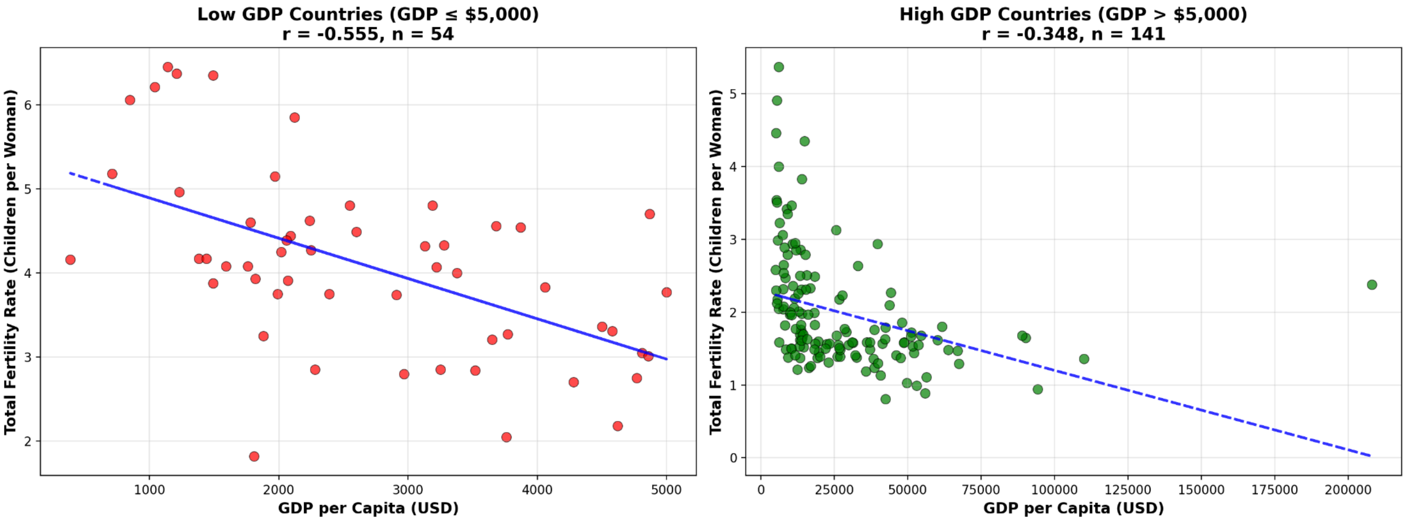


Julius provided the following narrative explanation:

"*There is a moderate negative correlation between GDP per capital and Total Fertility Rate (r ~ 0.5). This correlation is statistically significant (p < 0.001). About 25% of the variation in fertility rates can be explained by GDP per capita alone (R² = 0.247), which is substantial for a single indicator.*"

The correlation appeared to be skewed to the left, suggesting that the observed association is stronger at the lower end of the GDP axis. I asked Julius to explore this further by carrying out some stratified analyses.

d. Correlation between GDP per Capita and Total Fertility Rate, comparing countries with a GDP per capita of $5000 or less, and countries with a GDP per capita of over $5000 (diagram overleaf).



The stratification revealed differing strengths of correlation for the two groups:

- Low GDP countries: r = -0.555 (moderate-to-strong negative correlation)
- High GDP countries: r = -0.348 (weak-to-moderate negative correlation)

Explained Variance:

- Low GDP countries: 30.8% of fertility variation explained by GDP
- High GDP countries: Only 12.1% of fertility variation explained by GDP

When prompted, Julius provided the following interpretation of the relationship shown in the comparative analysis.

“The Economic Development Threshold Effect: This analysis reveals that the GDP-fertility relationship operates differently across income levels:

1. *In poorer countries (GDP ≤ $5,000): Economic improvements may have a substantial impact on fertility decisions. Even small increases in GDP per capita are associated with meaningful decreases in fertility rates.*
2. *In richer countries (GDP > $5,000): The relationship becomes weaker, suggesting that once basic economic needs are met, other factors (cultural values, education, urbanization, social policies) may become more important drivers of fertility decisions than pure economic wealth.*

The Demographic Transition Sweet Spot: The stronger correlation in low-GDP countries suggests this is where the most dramatic phase of the demographic transition occurs - the initial shift from high fertility to lower fertility as societies move from subsistence to more developed economies.

Policy Implications: This finding suggests that economic development policies may have the greatest impact on fertility rates in the world's poorest countries, while in wealthier nations, other social and cultural interventions might be more relevant for influencing demographic outcomes.”

My Reflection:

a. **What was easy?**

- Getting set up on Julius, downloading the datasets and prompting the model were all super simple.
- It was relatively easy to identify when the output was not quite right. I am not sure if this is because of my prior exposure to working with data.
- I found the little ‘code explanation’ at the bottom of each snippet of code in Julius super helpful and very intuitive as a person that does not know any Python..

b. **What was difficult?**

- I ran into a challenge with the formatting of original GDP data. Values above $10,000 were formatted in the dataset with a 'k' representing thousands (e.g., "13.3k", "49.7k", "67.4k"). Julius did not recognise this at first, so it read these values as strings and excluded them. I noticed that the GDP map had a lot of missing values, but was not able to figure out why, and needed to call for help from the course instructor.
- In the ‘old way’ of data analysis, I would have started off inspecting the data using some simple tables, and would easily have detected this challenge before I started my analysis. With these new tools it can be easy to skip those basics and not notice when some variables are missing.

c. **What was surprising?**

- Once we identified where the error mentioned above came from, I was surprised how easy it was to get Julius to fix this. I expected to have to write/edit some code, but a simple conversational prompt was enough to get the model to identify the values containing 'k’, define k = 1000', and multiply the values by the constant, leaving the other numbers unchanged.
- I found the narrative explanations Julius provided surprisingly insightful. They were a bit presumptive considering that this is just a correlation analysis, but they are useful as a starting point for examining what the data might be saying.
- One thing Julius did not do was adjust the correlation part of the task after we identified the ‘k’ omissions. I had to give it another prompt to

use the updated GDP per capita dataset for the correlation exercise. I was slightly surprised by this, as I expected it to at least ask me if it should use the modified dataset for the task.

- It is quite astonishing how powerful these models are. About 20 years ago, when I studied for an MSc in Epidemiology, I spent 3 months learning data analysis with Stata, and another 3 months cleaning and analysing a DHS dataset and writing up my findings as part of my dissertation. Today, the cleaning and analysis part of the same task would take me a couple of weeks at most using these tools. This opens up huge possibilities to use data better for learning, decision making and policy formulation.

d. **What did I learn?** It is essential to have some level of knowledge of statistics and data analysis to be able to use these models, in order to:
    - give them the right prompts
    - identify when there might be errors in the output
    - interpret what the output is telling us

    This means that data scientists are still very much needed; they just need to learn to work differently, making use of the new tools available to them.