03_Exploratory_Data_Analysis

06010567

Contents

Firstly, I read the cleaned dataset for further analysis. Then, I use a choropleth map to show an overview of the fertility rate around the world in 2023.

The choropleth map illustrates substantial global variation in fertility rates. Sub-Saharan Africa exhibits the highest fertility levels, often exceeding 5 births per woman, while Europe, East Asia, and North America maintain significantly lower fertility, often below 2. This visualisation emphasises the demographic divide between high- and low-income regions and highlights where population growth remains most rapid.

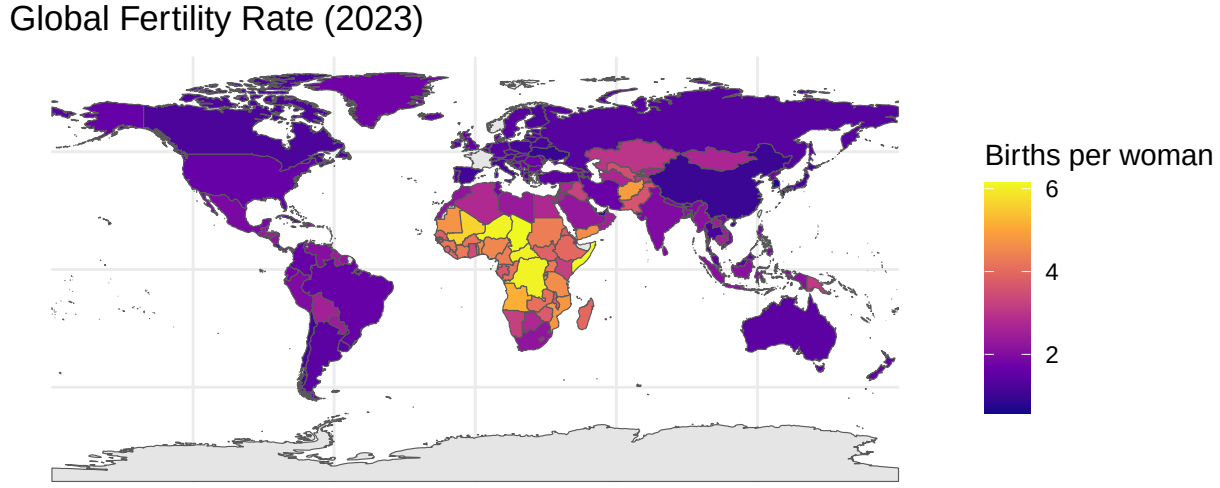

Figure 1: Choropleth map of the fertility rate across the world (2023)

From Table 1, we can see that there are still some missing values even after doing imputation. This happens because that some countries do not have any data of some variables, which means that imputation does not work. Therefore, I will drop these rows containing missing values in some parts of analysis as they account for a small proportion.

Table 1: Summary statistics for year 2023

fertility	edu_upper	gdp	urban	female_labor	contraceptive	gii	mmr	pr_f
Min. :0.586	Min. : 1.841	Min. : 193	Min. : 13.72	Min. : 4.939	Min. : 4.001	Min. :0.0030	Min. : 1.102	Min. : 0.2985
1st Qu.:1.486	1st Qu.:24.103	1st Qu.: 2566	1st Qu.: 43.93	1st Qu.:42.051	1st Qu.:29.877	1st Qu.:0.1630	1st Qu.: 12.148	1st Qu.:16.9014
Median :1.916	Median :49.372	Median : 8143	Median : 64.47	Median :53.205	Median :54.500	Median :0.3515	Median : 59.102	Median :25.0000
Mean :2.376	Mean :49.224	Mean : 21690	Mean : 62.36	Mean :50.785	Mean :49.724	Mean :0.3293	Mean : 143.938	Mean :25.6522
3rd Qu.:3.099	3rd Qu.:74.585	3rd Qu.: 29905	3rd Qu.: 81.82	3rd Qu.:60.126	3rd Qu.:66.700	3rd Qu.:0.4905	3rd Qu.: 194.860	3rd Qu.:33.7838
Max. :6.132	Max. :98.180	Max. :256581	Max. :100.00	Max. :82.561	Max. :86.300	Max. :0.8380	Max. :1222.531	Max. :55.7447
NA	NA's :33	NA's :3	NA's :2	NA's :30	NA's :64	NA's :45	NA's :33	NA's :24

Figure 2 reveals a clear negative relationship between female upper secondary education and fertility rate across countries in 2023. Higher levels of female education are associated with lower fertility rates, especially in high-income countries, where education levels are the highest and fertility the lowest.

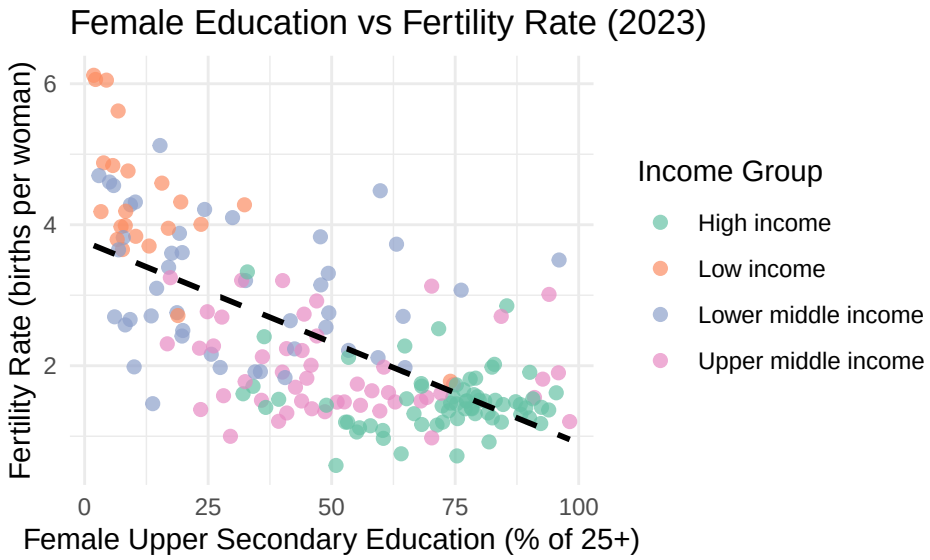

Figure 2: Scatter plot of female education and fertility rate (2023)

Figure 3 demonstrates a strong negative relationship between the expected years of schooling (female) and fertility rate. Countries with longer education tend to have substantially lower fertility.

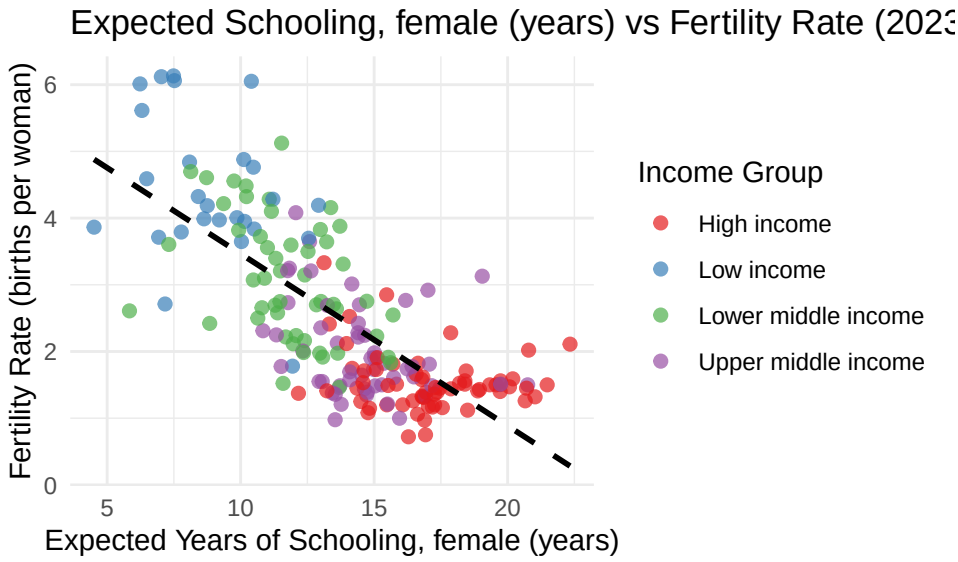

Figure 3: Scatter plot of expected years of schooling (female) and fertility rate (2023)

Figure 4 demonstrates a strong positive relationship between the Gender Inequality Index and fertility rate. Countries with greater gender inequality—reflected by higher GII values—tend to have substantially higher fertility. This relationship is particularly evident among low- and lower-middle-income countries, suggesting that limited access to education, employment, and reproductive autonomy may contribute to elevated fertility levels.

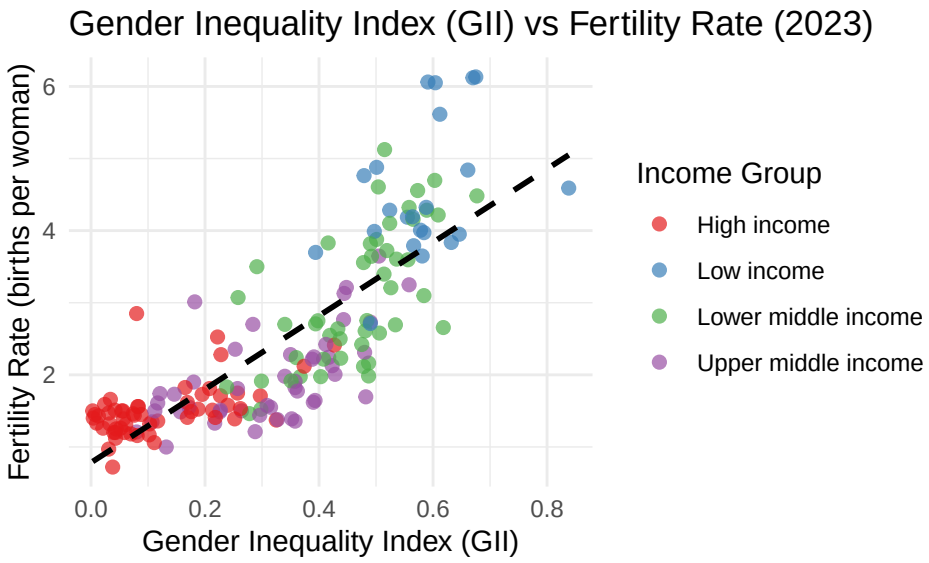

Figure 4: Scatter plot of GII and fertility rate (2023)

Figure 5 shows a similar pattern between MMR and fertility. There is a steep positive gradient: countries with high maternal mortality also tend to have high fertility. The use of a log scale on the x-axis reveals that this relationship is non-linear, with the most dramatic contrasts occurring among countries with MMR above 100 per 100,000 live births. This indicates a shared vulnerability: countries with insufficient maternal health services often also have high population growth pressures.

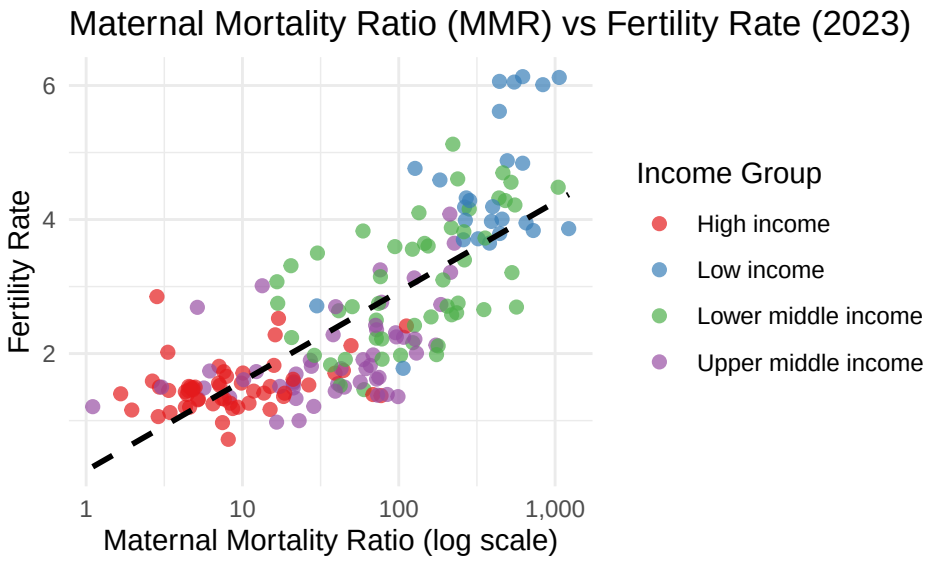

Figure 5: Scatter plot of MMR and fertility rate (2023)

Figure 6 shows a mild negative association between the share of women in national parliaments and fertility rates. While the relationship is weaker than with GII or MMR, it suggests that political representation of women may play a subtle but meaningful role in influencing social policy, gender norms, and eventually reproductive behaviour.

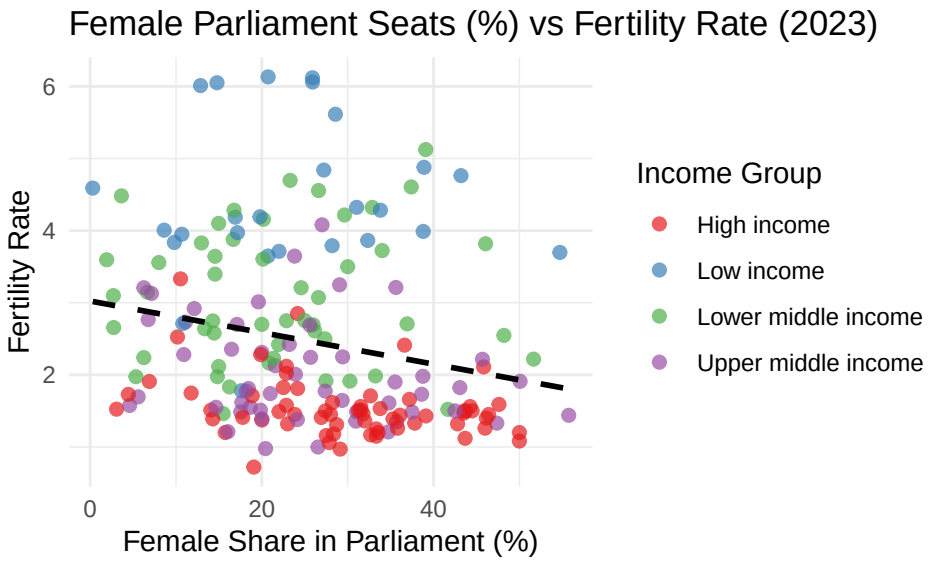

Figure 6: Scatter plot of proportion of female parliament seats and fertility rate (2023)

Figure 7 shows that the negative association between female upper secondary education and fertility is strongest in low-income countries. As the level of education increases, fertility rates drop sharply in these contexts. In contrast, the slope becomes progressively flatter in higher-income groups, indicating a weaker marginal effect of education on fertility. This pattern suggests that while expanding access to secondary education remains a powerful lever for fertility reduction in low-income countries, its effectiveness diminishes in wealthier settings, where fertility may be influenced by a broader set of structural and normative factors. Hence, this relationship requires further analysis in the future modelling step.

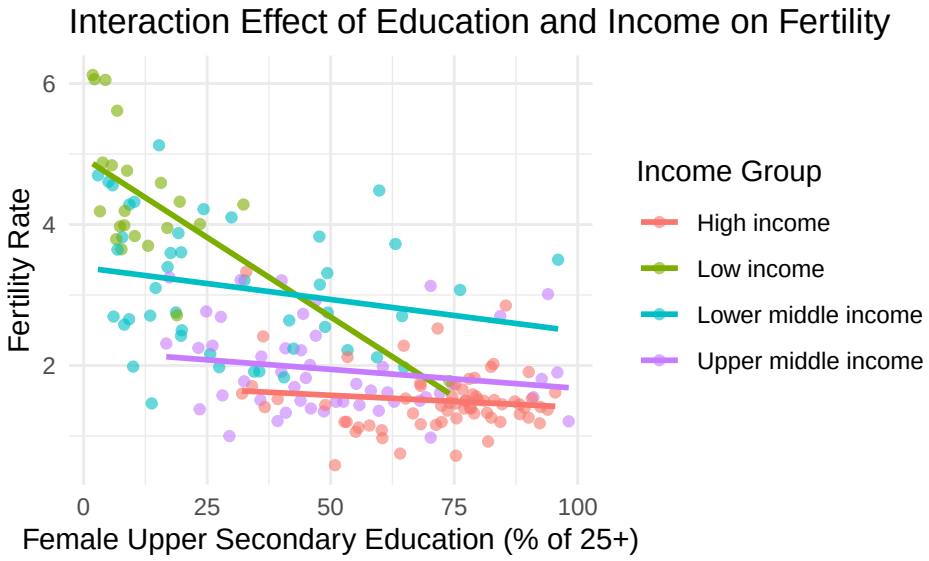

Figure 7: Relationship between education and fertility rate (2023)

From Figure 8, we can see that across all world regions, average female upper secondary education levels have been steadily increasing from 2000 to 2023. North America, Europe and Central Asia lead in educational attainment, while Sub-Saharan Africa and South Asia lag behind but show gradual improvements.

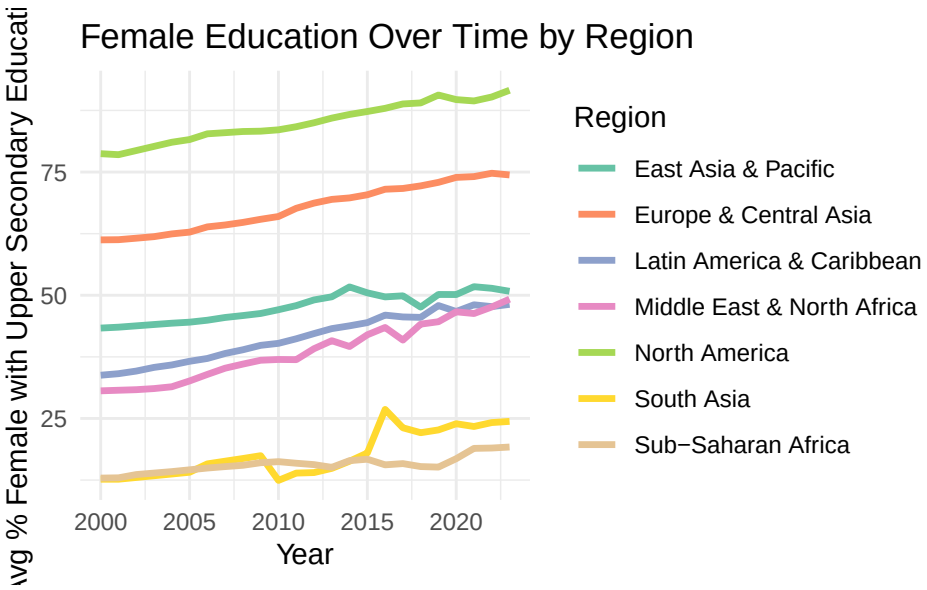

Figure 8: Female education trends plot over regions (2000-2023)

Fertility rates have declined across all regions over the past two decades (Figure 9). Sub-Saharan Africa maintains the highest fertility, though it is decreasing. In contrast, North America, Europe and Central Asia maintain consistently low fertility rates, suggesting regional differences in demographic transition pace.

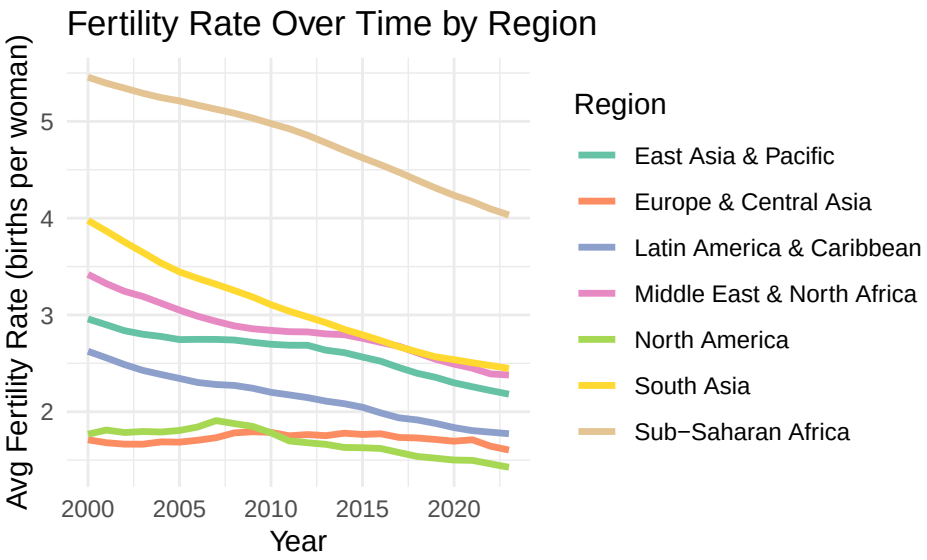

Figure 9: Fertility rate trends plot over regions (2000-2023)

The correlation heatmap (Figure 10) shows strong negative correlations between fertility rate and all levels of female educational attainment. Fertility is also negatively associated with GDP per capita, urbanisation, and contraceptive prevalence. Fertility is positively correlated with both GII and MMR and negatively correlated with PR_F. Additionally, strong multicollinearity is observed between different education variables and GDP/log-GDP, reinforcing the importance of careful variable selection in modelling steps. Together, these

indicators highlight that gender inequality, health system strength, and women's empowerment are jointly linked to fertility variation across countries.

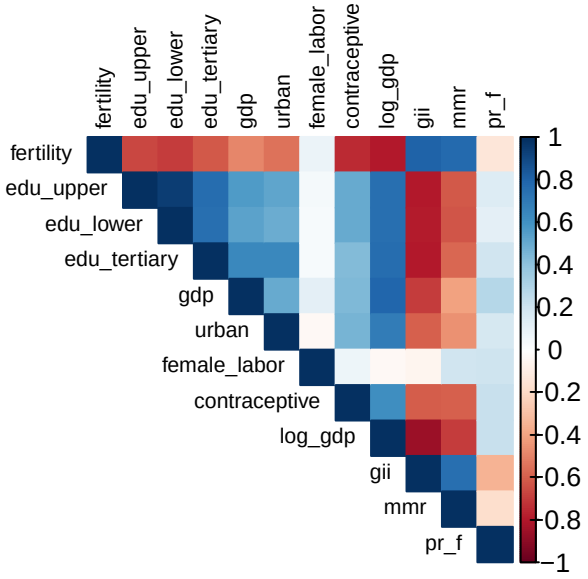

Figure 10: Correlation heatmap between variables

The comparison between fertility rate and GDP (left of Figure 11) versus log-transformed GDP (right of Figure 11) reveals that the raw GDP values are highly skewed, leading to a non-linear and heteroscedastic relationship with fertility. In contrast, the log(GDP) transformation produces a much more linear and evenly spread association, with a clearer negative slope. This suggests that log-transformed GDP is a more appropriate predictor in regression models analysing fertility, as it better captures the diminishing marginal effects of economic development on fertility.

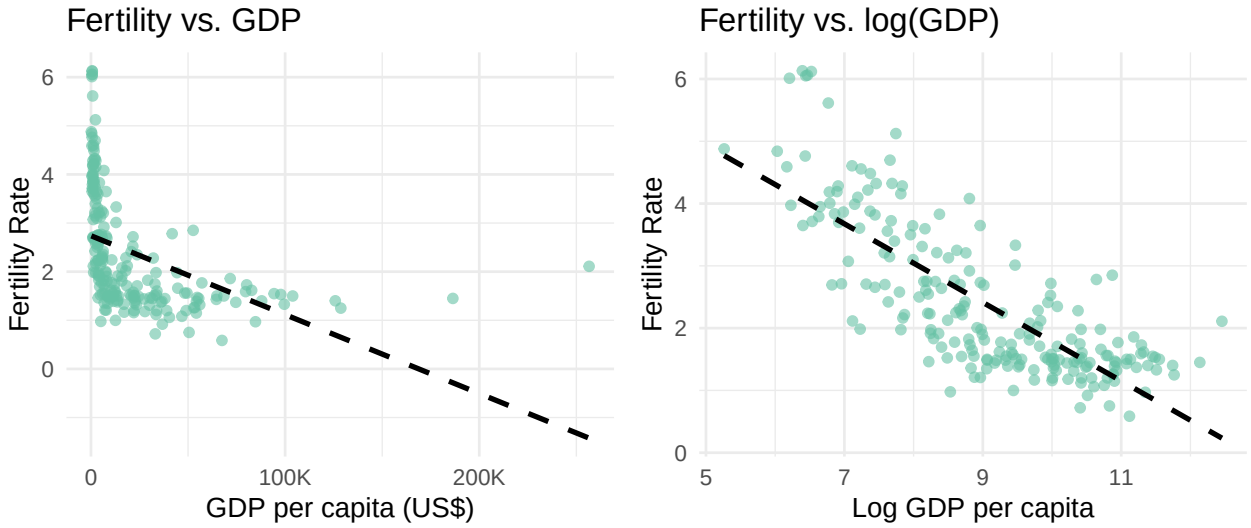

Figure 11: Comparison between fertility rate and GDP versus log-transformed GDP