Aashka Navale
05-14-2025

FINAL REPORT

Adolescent Fertility & Out of School

DATA 205 | CRN 33334

Introduction

This project analyzes global fertility rate trends using data from the World Bank covering the years 1960 to 2020. The goal is to examine how fertility rates have changed across regions and over time, and to identify differences between geographic areas and how it correlates with adolescents out of school (male and female). The data consists of countries and years from 1960-2020. I created two categorical variables—region and income—that also come from the World Bank. Data from the World Bank is flexible in the sense that you can choose what variables to include in the dataset. My tools for this project include RStudio, Highcharter, and Tableau.

This is a crucial topic considering adolescents across the world are dropping out of school when they have a right and deserve to complete an education. Adolescent fertility rates take part in the decision of dropping out. Spreading awareness is the first step to mitigation so if we are able to address these issues in the countries where it is needed most, we could significantly increase their literacy rates and decrease child fertility rates.

Data Cleaning

Besides the considerable amount of N/A values, my data was already relatively clean. I only had to rename a few variables, remove a row or two, and create my categorical variables. The "Region" variable consists of seven possible values, with every country classified under one of them: East Asia & Pacific, Europe & Central Asia, Latin America & Caribbean, Middle East & North Africa, North America, South Asia, and Sub-Saharan Africa. The "Income" variable consists of four possible values, with every country falling into one of the groups: High income, upper middle income, lower middle income, and low income. I added these variables to all four datasets.

All of my data are time series spanning from 1960-2020. However, all of my adolescent drop-out rate datasets have an extensive amount of N/A values. The data before the year 2000 is sparse and barely workable. I made the decision to omit years 1960 through 1995 to have useful data.

Descriptive Statistics

My datasets consist of 217 observations each. The main fertility rate dataset has 16 variables because the years span from 1960 - 2020. The adolescents out of school datasets have 8 variables instead of 16 because the years span from 2000 - 2020. As stated before, the out of school datasets are heavily composed of N/A values so the years before 2000 were futile. Under each year (which is a variable), is a numerical value which is the rate. Fertility rate is defined as births

per 1000 girls and the out of school datasets are percentages. For example, these are the countries starting with the letter ‘A’ in the fertility rate dataset.

	Country	1960	1965	1970	1975	1980	1985	1990	1995	2000	2005	2010	2015	2020	Region	Income
1	Afghanistan	140.6	140.8	138.5	136.9	131.7	129.2	139.4	158.3	154.3	130.8	105.6	81.0	68.9	South Asia	Low income
2	Albania	62.1	40.2	35.6	30.3	21.4	15.6	15.8	21.9	14.7	18.2	19.7	20.7	14.0	Europe and Central Asia	Upper middle income
3	Algeria	108.7	109.6	108.6	89.5	71.3	42.7	24.4	18.8	8.7	7.9	9.4	10.3	9.9	Middle East and North Africa	Upper middle income
4	American Samoa	52.9	56.8	50.6	40.9	39.3	39.5	52.2	51.5	48.4	38.2	35.8	42.2	34.1	East Asia and Pacific	High income
5	Andorra	9.3	11.4	13.1	19.8	21.5	15.3	12.1	9.2	9.1	9.1	8.0	4.9	3.1	Europe and Central Asia	High income
6	Angola	92.5	102.5	105.8	109.4	112.0	128.1	145.4	151.6	147.7	157.1	160.2	150.2	144.3	Sub-Saharan Africa	Lower middle income
7	Antigua and Barbuda	120.3	119.1	105.4	85.3	86.0	80.8	78.4	66.6	61.9	50.1	42.7	35.9	33.9	Latin America and Caribbean	High income
8	Argentina	57.2	56.2	64.4	73.1	77.9	71.1	70.3	70.3	65.8	63.5	65.9	63.9	31.5	Latin America and Caribbean	Upper middle income
9	Armenia	47.9	40.6	40.1	39.7	50.4	61.7	73.9	70.7	37.3	30.1	31.8	25.1	13.7	Europe and Central Asia	Upper middle income
10	Aruba	44.7	47.4	44.8	43.2	44.5	52.8	59.5	43.2	43.7	40.1	40.1	32.3	20.6	Latin America and Caribbean	High income
11	Australia	43.1	46.9	50.8	40.4	27.9	22.5	22.4	20.5	17.8	16.5	17.1	12.0	7.7	East Asia and Pacific	High income
12	Austria	52.0	54.9	57.7	45.8	34.2	25.2	22.5	17.4	13.9	12.8	10.3	7.8	4.9	Europe and Central Asia	High income
13	Azerbaijan	59.4	37.5	37.7	20.3	20.7	22.7	31.5	38.2	30.1	38.7	48.2	62.7	41.6	Europe and Central Asia	Upper middle income

Region and income were manually created by me sourced from World Bank data. Every country is organized into one of six regions and one of four income groups. All of these edits were made to the rest of the datasets as well.

Final Data Products

1.0 Is there a correlation between adolescent fertility rates and the percentage of female adolescents out of school across different world regions?

For this first main question, I wanted to find a relationship between adolescent fertility rates and the percentage of female adolescents out of school. I also wanted to see if a particular region tends to have a higher/lower fertility rate than others. Using Highcharter, the interactive scatterplot shows the relationship between adolescent fertility rates and the percentage of female adolescents out of school across different countries from 2000 to 2020. Each point stands for a country in a specific year, and the colors and shapes differentiate the years. Overall, there is a clear trend. Countries with higher adolescent fertility rates tend to have more girls out of school. Most data points are clustered in the bottom left of the graph, suggesting that many countries have relatively low fertility and better school attendance. Over time, especially by 2020, more countries appear to have lower fertility rates, showing progress in reproductive health.

1.1. How does this relationship differ between Sub-Saharan Africa and the global average since 2000?

For this sub-question, I wanted to compare Sub-Saharan Africa to the rest of the world. I averaged out the region and global data. I created an interactive line graph using Highcharter. The results are astounding. Sub-Saharan Africa has a significantly larger average fertility and dropout rate than the rest of the world, nearly double. It looks like there's a decline occurring in recent years and based on educated inferences, it will continue to go down as time passes. The highest peak seems to be in 2005. I would have liked to explore data from 1960 as well but it was so sparse that I stated it unworkable.

1.2. Are changes in adolescent fertility associated with increases or decreases in school dropout rates among adolescent girls?

This sub-question aims to explore adolescent fertility with school dropout rates amongst girls. I used ggplot to create this visualization. This is similar to graph 1.0, but the only difference is that it's colored by region instead of year. Now it's much easier to see the spread of the regions. Once again, Sub-Saharan African countries prevail in the most amount of countries with the highest fertility rate and dropout rate. Overall, there is a clear trend, as shown by the regression line. Countries with higher adolescent fertility rates tend to have more girls out of school. Most data points are clustered in the bottom left of the graph, suggesting that many countries have relatively low fertility and better school attendance.

2.0 How does national income level affect adolescents' school attendance rates?

For my second main question, I wanted to incorporate income into fertility rate and adolescents out of school. I chose to facet wrap a box plot to compare the year 2000 vs 2020. This boxplot compares female dropout rates across income groups in the years 2000 and 2020. In both years, dropout rates are highest in low-income countries and decrease as income increases. Over time, there is a noticeable decrease in dropout rates across all income groups, with the most significant drop in low-income and lower-middle-income countries. High-income countries consistently show the lowest dropout rates. The black dots represent the average dropout rate for each group, which also shows improvement from 2000 to 2020.

2.1. Within Europe and Central Asia, is there a consistent trend in school attendance across income groups?

For this sub-question, I wanted to look specifically into Europe and Central Asia since they are the highest income regions who show the lowest fertility rates. This boxplot shows female dropout rates in 2020 across different income levels within the Europe and Central Asia region. The chart includes lower-middle, upper-middle, and high-income countries. If you noticed, low-income countries are not shown because none are a part of this region. Overall, high-income countries had the lowest dropout rates with little variation. Upper-middle-income countries showed a wider range of dropout rates and a slightly higher average compared to lower-middle-income countries. This suggests that while income level generally relates to lower dropout rates, other factors play a part in outcomes.

2.2. How do female dropout rates compare to male dropout rates?

For this sub-question, I wanted to look for male adolescents this time and see how their dropout rates compare with female adolescents. Looking at the visualization, do not immediately trust these results because the data is heavily filled with N/A values. Based

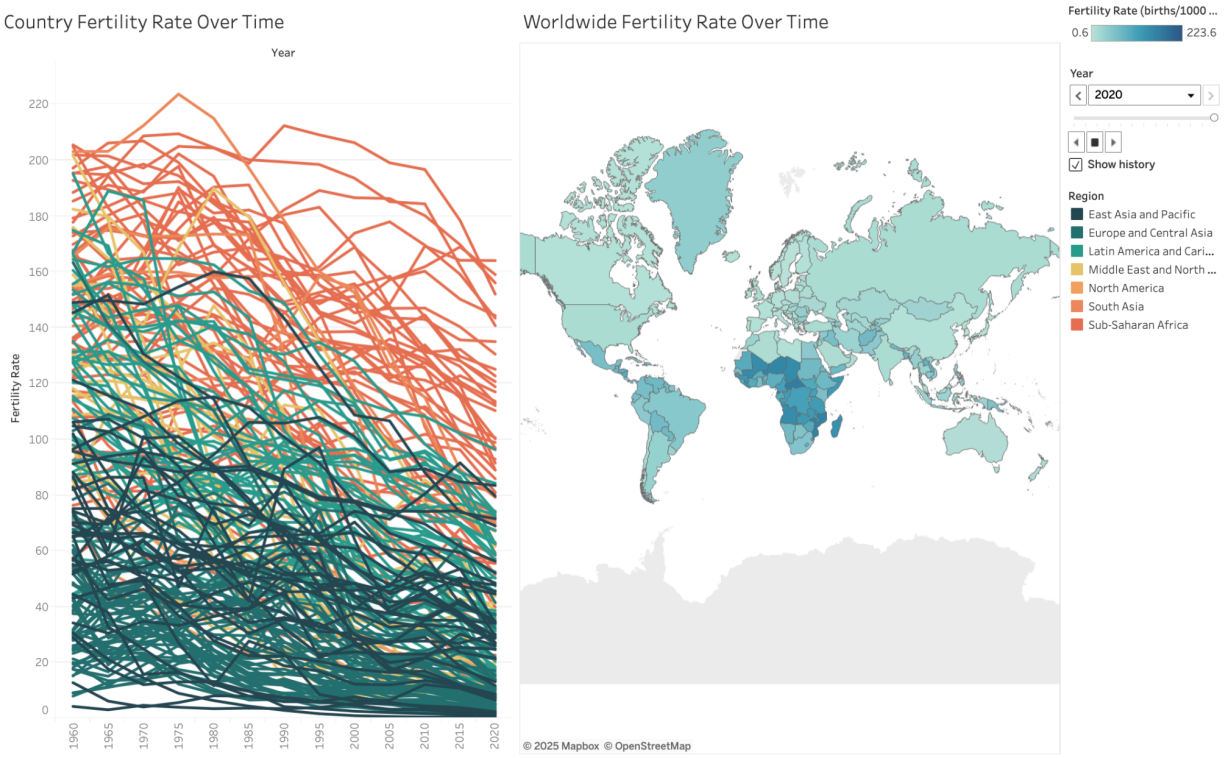
on what we know already, female adolescent dropout rates should be much higher, especially in low income communities, than male dropout rates. In fact, this is the case for most of my visualizations that involve one of the dropout datasets. Since there are so many N/A values, the data was picked at random to whichever country had any data at all. However, a different perspective would tell us that it makes sense that male students would have higher dropout rates for low income groups because families send their sons to work instead of completing an education.

3.0 In 2020, how did the adolescent fertility rate in South Asia compare to East Asia and the Pacific?

I wanted to compare most countries in Asia (since some are a part of the Europe & Central Asia region) by region to see which region has better statistics. This table compares adolescent fertility rates between East Asia and Pacific and South Asia. South Asia has a higher average (mean) fertility rate of 38.5 births per 1,000 girls compared to 29.96 in East Asia and Pacific. The median fertility rate is slightly higher in South Asia (30.05) than in East Asia and Pacific (28.8). While both regions have similar maximum fertility values, 79.2 in South Asia and 83.5 in East Asia and Pacific, the minimum values differ more, with East Asia and Pacific having a very low rate of 0.6, compared to 6.4 in South Asia. This shows that there is a greater variation in fertility rates within the East Asia and Pacific region.

3.1. How do adolescent fertility rates vary across world regions, and how have they changed over time?

This visualization will be made in Tableau as a dashboard so I'm reshaping data to export from RStudio onto my computer. This animated visualization shows countries' fertility rates over time as well as an animated map to show fertility rate over time. The line chart on the left shows fertility trends for individual countries from 1960 to 2020 where there is a consistent overall decline in fertility rates across most regions. Also, Sub-Saharan Africa stands out once again with higher fertility rates compared to other regions, although it is declining like the rest. On the right, the time series map shows regional disparities in fertility rates. Countries in Sub-Saharan Africa still exhibit the highest rates (darker blue), while regions like Europe, East Asia, and North America show much lower rates (lighter shades). If you interact with the animation, you will notice that all the countries' shades get lighter over time, meaning that fertility rate is decreasing. Together, these visuals show the temporal decline and geographic variation in fertility rates.



To conclude, every visualization and every finding in this project leads us to one common point. The fact is that lower income communities struggle with controlling adolescent fertility rates which in turn leads to higher adolescent dropout rates, especially in girls, and especially in Africa. While over time fertility rates and adolescent dropout rates are declining, it is still far too high in certain low-income regions. It is crucial we address this issue to combat girls becoming mothers while in school and in turn having to give up their education.

References & Acknowledgements

Data Sources - World Bank

- https://databank.worldbank.org/source/world-development-indicators
- https://datatopics.worldbank.org/sdgatlas/archive/2017/the-world-by-region.html
- https://datatopics.worldbank.org/world-development-indicators/the-world-by-income-and-region.html

Tools

- RStudio
- Highcharter
- Tableau
    - https://public.tableau.com/app/profile/aashka.navale/viz/FertilityRateOverTime_17472204668760/Dashboard1?publish=yes

First and foremost, I would like to thank Professor Saidi and Professor Alraee for creating such a positive first experience with coding and R. I sincerely enjoyed being their student and loved excelling in their classes.

I would also like to mention my classmate and friend, Charlie Roth, who has helped me numerous times throughout these past two years. Coding is not my strongest suit and every time I was stuck, they were there to help me so I wish to honor that.

Last but not least, I want to acknowledge Professor Perine for her patience and cooperation with me for the time I've been her student. There have been multiple instances where I have not been meeting standards but yet she's shown me great understanding and kindness.

Thank you to everyone else who has helped me throughout my years at Montgomery College.