# The Relationship between Family Status and Academic Achievement in Children*

Ruibo Sun

12-03-2023

This paper explores the relationship between family status variables, such as parents' socio-economic status and educational level, and their impact on children's academic achievement and education level using the data from 2021 GSS. Previous studies have suggested that parents' education can positively influence their children's academic performance due to increased access to resources and greater involvement in their education. However, recent research has demonstrated that this relationship is complex and can be influenced by a variety of psychological and sociological factors. The findings suggest that parents' education, family income, and prestige score all have a positive correlation with children's education. Further research is needed to fully understand the complexities of this relationship and use the results to improve educational policies and programs.

## 1 Introduction

In the past, family status variables, such as the socio-economic status and educational level of parents, were commonly considered as strong predictors of academic achievement in children. More specifically, parents who have lower levels of education, and those with higher levels of education are more inclined to view higher education as desirable and encourage their children to excel academically. They also tend to hold higher expectations for their children's academic performance (Davis-Kean 2014).

One potential explanation for this relationship is that higher levels of education can provide parents with access to resources such as income, time, energy, and community contacts, which can enable greater involvement in a child's education. As a result, the influence of family status variables on children's academic achievement may be best understood as a complex interaction between status and process variables (Khan, Iqbal, and Tasneem 2014).

*https://github.com/ruibosun/how-parents-affect-childrens-education

However, recent research has indicated that the relationship between these factors and academic achievement is not always direct. Instead, socio-economic status and parents' education are part of a larger set of psychological and sociological variables that can impact children's educational outcomes (Khan, Iqbal, and Tasneem 2014).

In short, the findings in this paper demonstrate that parents' education has a positive correlation with children's education. Additionally, the region of living during the teenage years can also impact children's education. Parents' prestige score seems to be another factor in children's education. However, there is some limitation on the definition of prestige score that may influence the overall results. This will be discussed in the Section 5.4.

The data will be presented clearly and succinctly with plots and tables. This following analysis is processed in R (R Core Team 2020) with packages of tidyverse (Wickham et al. 2019), dplyr (Wickham et al. 2022), here (Müller 2020), haven (Wickham, Miller, and Smith 2023) and broom (Robinson, Hayes, and Couch 2022). The tables are constructed via knitr (Xie 2023), scales (Wickham and Seidel 2022) and kableExtra (Zhu 2021). The package inside tidyverse helps to create the plots in ggplot2 (Wickham et al. 2019). This paper is knitted as a PDF file by the packages of R markdown ("R Markdown - Dynamic Documents for r," n.d.) and formatted using patchwork (Pedersen 2022).

# 2 Data

## 2.1 Source

The 2021 General Social Survey (GSS) is a nationally representative survey conducted to collect data on social trends and attitudes among people living in the United States. Data was collected through face-to-face interviews with adult residents, covering a wide range of variables that are of interest to social scientists, policymakers, and the general public. The variables measured in the survey include demographics, employment, education, health, family, and social attitudes ("GSS Data Explorer: NORC at the University of Chicago," n.d.).

The target population is US adults aged 18 and older. The GSS uses a multistage probability sampling approach to recruit its sample from the US Census Bureau's Master Address File. One strength of the GSS is its long history and large sample size, but a potential weakness is a reliance on self-reported data, which can be subject to response bias. The questionnaire includes both closed-ended and open-ended questions, but there may be concerns about biased or leading questions. The GSS employs various methods to adjust for non-response and ensure the representativeness of the sample. One strength of the GSS is its long history and large sample size, but a potential weakness is a reliance on self-reported data, which can be subject to response bias. The questionnaire includes both closed-ended and open-ended questions, but there may be concerns about biased or leading questions. ("General Social Survey (GSS)," n.d.) The GSS employs various methods to adjust for non-response and ensure the representativeness of the sample. The limitation will be explained in detail in the Section 5.4.

This paper will test and explore how parents' social-economics status can affect children's education. The data that will be used in this paper comes from the US General Social Survey from the National Opinion Research Center at the University of Chicago. The factor of socio-economics status is measured using the occupation prestige score. It is a measure used in social science research to assess the level of social status or prestige associated with a particular occupation or profession, which was developed by the National Opinion Research Center at the University of Chicago. As part of the GSS, respondents are asked to rate the prestige or social standing of various occupations on a scale of 1 to 90, with higher scores indicating greater prestige. (Smith, n.d.) Another variable that has been used is the education level, which is a number of variables to indicate the number of years the respondents have spent in school and college. Other variables will be discussed in Section 2.

## 2.2 Data cleaning

For simplicity purposes, only variables that are closely related to the topics are selected for further analysis. These variables are mapres10, papres10, paeduc, maeduc, educ, mawrkslf, pawrkslf, degree, born, sex, reg16, and incom16.

- **mapres10**: A numeric variable indicating the respondent's mother's prestige score.
- **papres10**: A numeric variable indicating the respondent's father's prestige score.
- **paeduc**: A numeric variable indicating the number of years of education completed by the respondent's father.
- **maeduc**: A numeric variable indicating the number of years of education completed by the respondent's mother.
- **educ**: A numeric variable indicating the number of years of education completed by the respondent.
- **mawrkslf**: A categorical variable indicating whether the respondent's mother was employed for wages or salary.
- **pawrkslf**: A categorical variable indicating whether the respondent's father was employed for wages or salary.
- **degree**: A categorical variable indicating the highest degree earned by the respondent.
- **born**: A categorical variable indicating the region of the US where the respondent was born.
- **sex**: A categorical variable indicating the respondent's sex.
- **reg16**: A categorical variable indicating the region of the US where the respondent lived at age of 16.
- **incom16**: A categorical variable indicating the respondent's total household at age of 16.

The level of "no formal schooling" in **maeduc**, **paeduc** and **educ** is modified as the integer zero to make the variables consistent. All the numeric variables are shown as integers, but their class are in terms of character in R. Hence, the variables of **paeduc**, **maeduc**, **educ**, **mapres10** and **papres10** have been converted into a class of numeric for further analysis.

Table 1: Number of respondents by degree for 2021 survey

| Degree | Total | Proportion |
|---|---|---|
| graduate | 760 | 18.8% |
| bachelor's | 1036 | 25.7% |
| associate/junior college | 370 | 9.2% |
| high school | 1597 | 39.6% |
| less than high school | 246 | 6.1% |
| NA | 23 | 0.6% |

None of the missing values (NA) is removed in this paper. The reason is that removing the missing values in every column would result in a huge loss in the dataset. To maintain the representativeness of the data, missing values are automatically filtered by R, rather than removed manually. The possible limitation and effect of the missing values will be discussed in the Section 5.4.

## 3 Results

The respondents in this study are the children and their parents' education has also been collected in the GSS. As for the response variables, children's education level, there are two ways to measure. One is the degree, and the other one is the number of years in school. In Table 1, it shows the distribution of education in terms of degrees. The data reveals that the most common degree category among the group is high school, with 1597 individuals (39.6%) having this level of education. Following this, the next most common category is a bachelor's degree, with 1036 individuals (25.7%) having completed this level of education. The data also indicates that a graduate degree is the third most common category, with 760 individuals (18.8%) having this level of education. Associate/junior college degree is less common than a bachelor's or graduate degree, with 370 individuals (9.2%) having completed this level of education. However, a smaller proportion of individuals (6.1%) have less than a high school education. Additionally, there are also 23 individuals (0.6%) whose educational degree information is not available (NA). Overall, the data provide insights into the educational distribution of the given group and may be useful for drawing conclusions and making informed decisions related to education.

The second variable utilized to quantify education is relation to the number of years of schooling. In Figure 1, this histogram shows an overall distribution of the respondents' education in years. It is clear that the distribution is right-skewed, which indicates that there are more respondents with lower levels of education than with higher levels of education. In other words, the bulk of the respondents are clustered towards the left side of the histogram (lower levels of education), and there are relatively fewer respondents towards the right side of the histogram

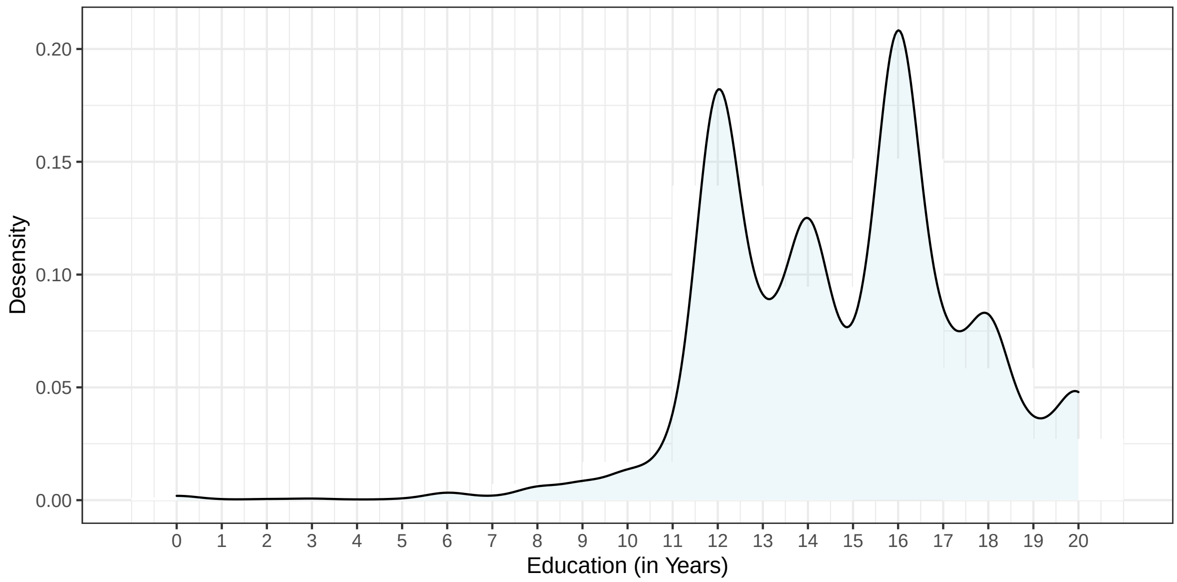


Figure 1: Histogram of Respondents' Education

Table 2: Number of respondents by education for 2021 survey

| Min | Median | Max | Mean |
|---|---|---|---|
| 0 | 15 | 20 | 14.76904 |

(higher levels of education). In short, the majority of respondents in 2021 GSS are highly educated.

To see a more accurate numerical summary, Table 2 is created. In the dataset labelled Section 2, the category denoting "no formal schooling" has been redefined as an integer value of 0. According to the information presented in the Table 2 table, the range of years of schooling reported by respondents spans from 0 to 20 years. The median number of years of schooling reported is 15, and the mean is 14.76904. This mean value roughly corresponds to three years in college. Taken together, these findings suggest that the respondents surveyed in the 2021 GSS have generally completed their high school education and are highly educated.

In Figure 2, it shows the distribution of respondent's father's and mother's prestige scores. The distribution of father's prestige score and the mother's prestige score have quite different distributions.

In order to examine the relationship between parents' prestige scores and children's education level. The next plot, Figure 3, is used to show the relation. These two scatter plots show the father's prestige score and the mother's prestige score with their children's education. It is hard to see a clear relation or pattern based on these two scatter plots. To verify if the prestige score is one of the factors, a linear model will be used in the Section 4.

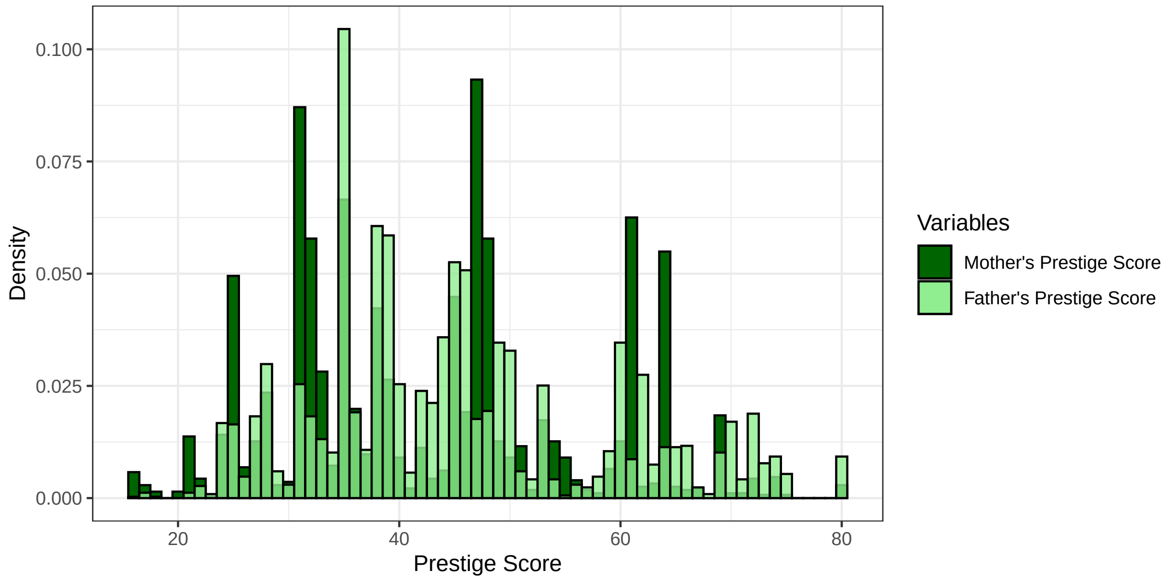


Figure 2: Mother's and Father's Prestige Score

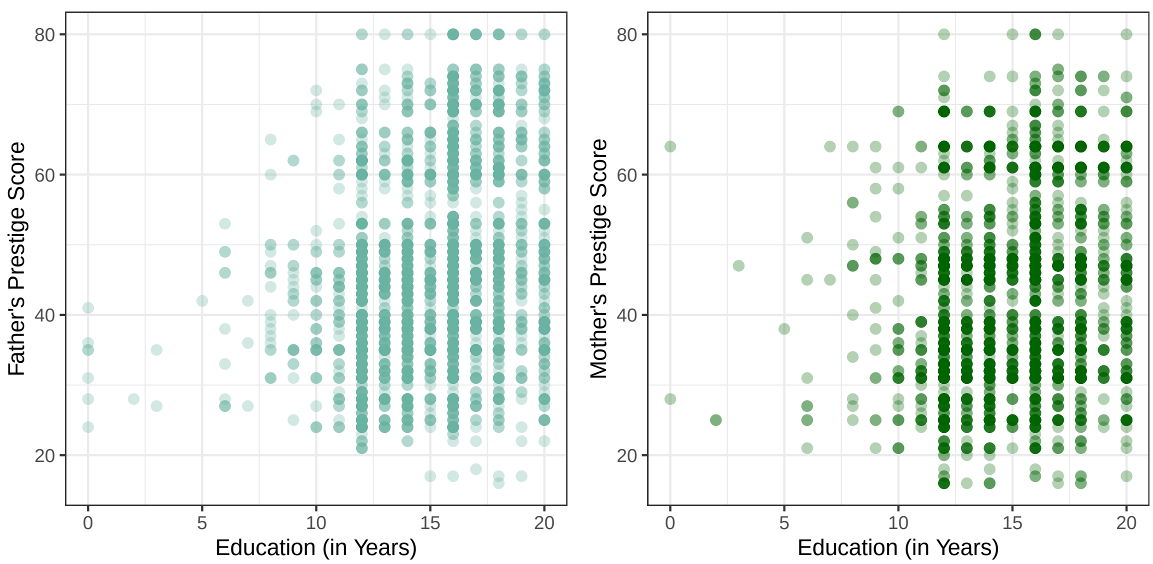


Figure 3: Mother's and Father's Prestige Score with children's education

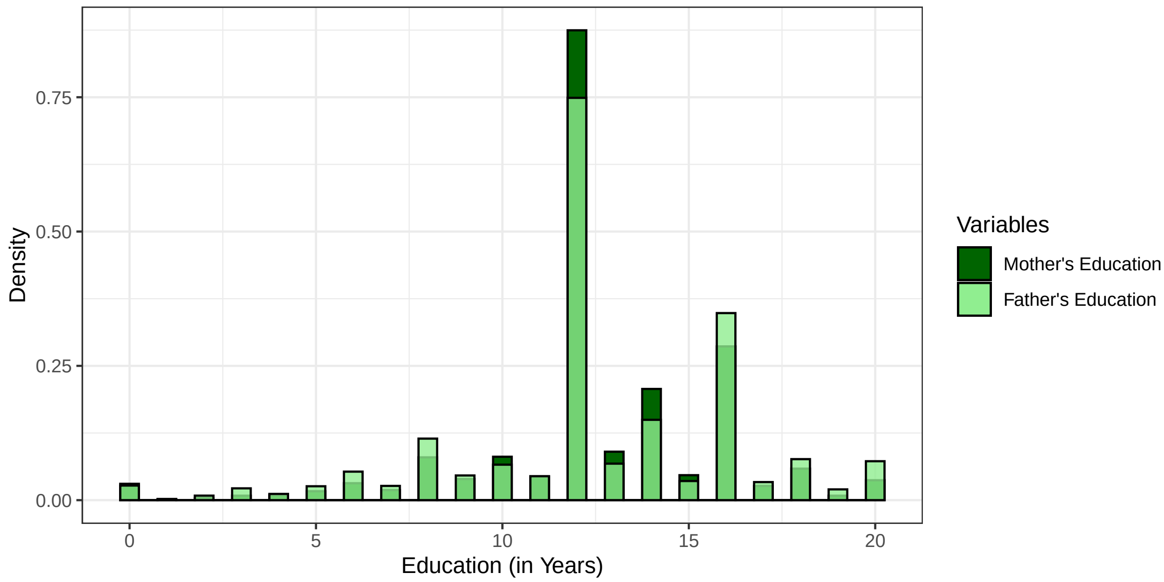


Figure 4: Mother's Education and Father's Education Level

Another key variable is the parent's education level. In Figure 4, it shows the distribution of the mother's education and the father's education. One interesting result is that the distribution of education for mother and father is similar, or even almost identical. This indicates that there is a strong correlation between the education levels of mothers and fathers. In other words, if a mother has a high level of education, it is likely that the father also has a high level of education, and vice versa.

Based on (Overman and Bosquet 2016), the birthplace may have an impact on children's education. In Figure 5, it is used to test the relationship between birthplace and children's education. Note that the question of birthplace only asks if the respondent was born in the US or not. In general, most of the respondents have received at least 10 years of education despite their birthplace. However, the bar for respondents who were born outside of the US is much higher than the other groups. This is because 87.2% of the respondents were born in the US, and only 11.0% answered "no". Additionally, there are 72 NA, which take a proportion of 1.8%. Given the huge differences in the sample sizes between groups, it is hard to draw a conclusion of whether birthplace may affect the education level.

According to the study (Nieuwenhuis and Hooimeijer 2016), there appears to be an association between neighborhood and education. In the 2021 GSS, the neighbourhood variable is being measured by the living region at the age of 16. The distribution for all regions, as shown in Figure 6, follows a similar bimodal shape, with the exception of the NA group. This could indicate that the region of living has a minimal or no impact on children's education levels. However, it is important to note that the measurement of the neighborhood through living regions at age 16 may not fully capture the impact of the neighborhood on education, which

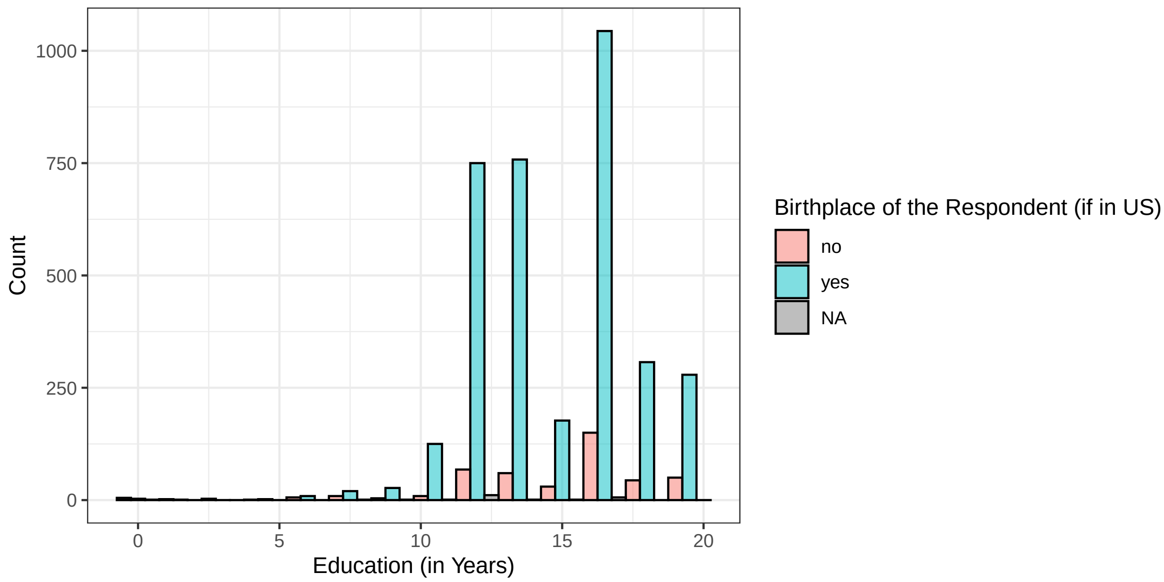


Figure 5: Birthplace of the Respondent (if in US)

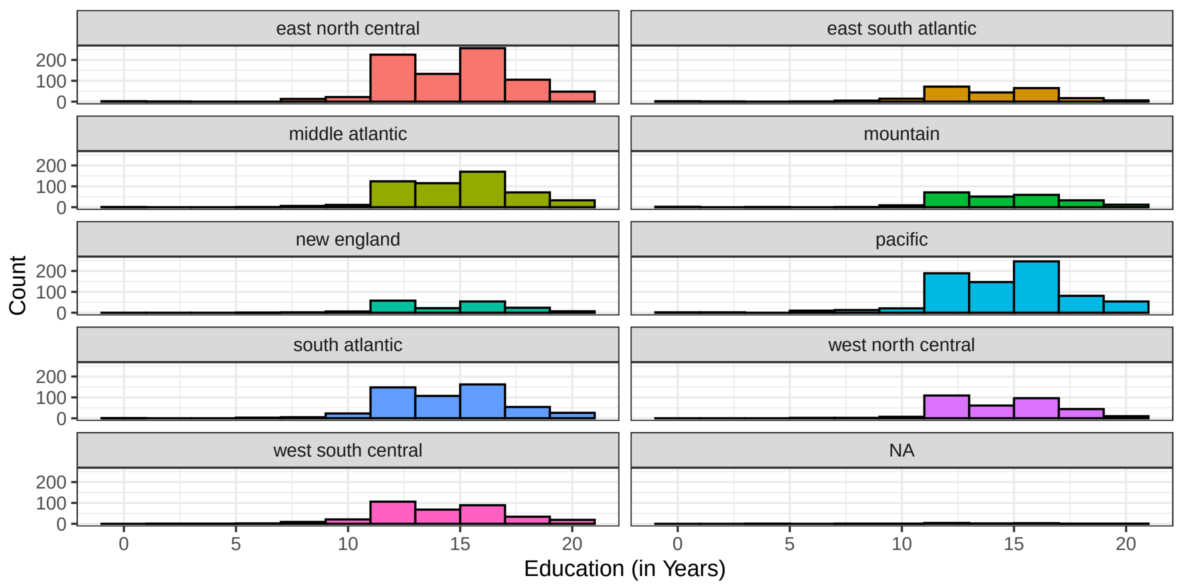


Figure 6: Respondent's Living Region at Age of 16 v.s. Education

will be discussed in the Section 5.4.

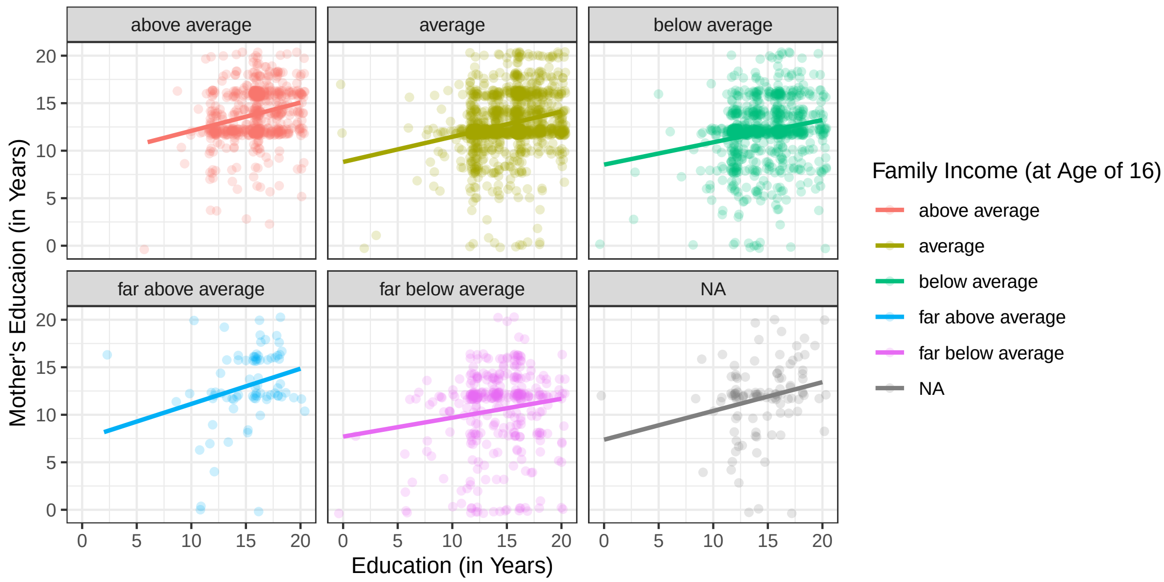


Figure 7: Education vs Mother's Education with respect to Family Income

To examine the relationship between family income and education, the Figure 7 and Figure 8 visualize this relation. In Figure 7, the vertical axis represents the mother's education and the horizontal axis represents the children's education. The respondents were asked to rate their family income at age of 16 in the following category: (1) far above average, (2) above average, (3) average, (4) below average and (5) far below average. Due to missing values, the sixth group is NA, which is coloured grey. Overall, mothers' education and children's education are positively correlated despite the level of income. To be more specific, for respondents whose family income was far above average, the slope seems to be steeper than the other groups, which indicates that for those with higher family income, there may be a stronger relationship between the education of the mother and the education of the children.

In Figure 8, the income groups are the same as before. However, the plot is now examining the correlation between the father's education and the children's education. Once again, there is a positive correlation between the education level of fathers and their children's education, regardless of family income. However, for respondents who reported a "far above average" family income, the relationship between the father's education and children's education appears to be stronger than for other income groups. This suggests that higher family income may be associated with a more pronounced link between paternal and children's education levels. Comparing the two plots, the slopes in Figure 8 for all income groups are much steeper than the slopes in Figure 7. This may suggest that there is a stronger relationship between fathers' education and children's education compared to mothers' education and children's education given the same income level.

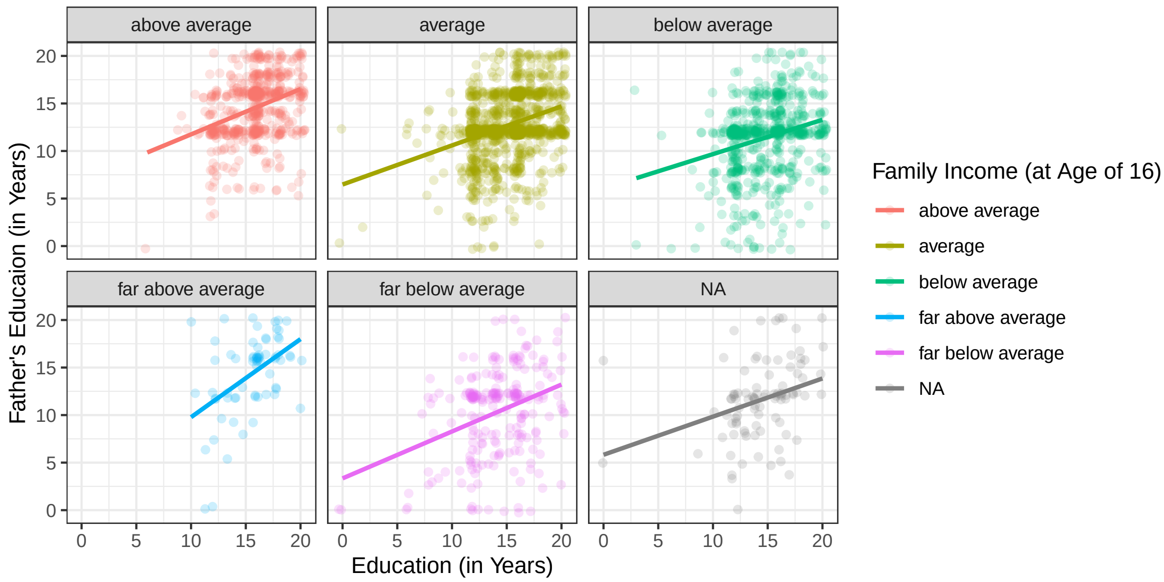


Figure 8: Education vs Father's Education with respect to Family Income

# 4 Model

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_{mapres10} + \hat{\beta}_2 X_{papres10} + \hat{\beta}_3 X_{paeduc} + \hat{\beta}_4 X_{maeduc} \tag{1}$$

The Table 3 is built based on the Equation 1:

- The intercept is 11.32, which means that when all the independent variables are zero, the dependent variable has a value of 11.32.
- The coefficients for the independent variables mapres10 and papres10 are both 0.01, which means that a one-unit increase in either of these variables is associated with an increase of 0.01 in the dependent variable.
- The coefficients for the independent variables paeduc and maeduc are 0.18 and 0.05, respectively. This means that a one-unit increase in paeduc is associated with an increase of 0.18 in the dependent variable, and a one-unit increase in maeduc is associated with an increase of 0.05 in the dependent variable.
- The number of observations used in the model is 1980. The R-squared value is 0.124, which means that the independent variables explain 12.4% of the variation in the dependent variable.
- The adjusted R-squared value is 0.122.
- The AIC and BIC values are measures of the goodness of fit of the model. Lower values indicate a better fit.
- The log-likelihood is a measure of how well the model fits the data.

Table 3: Linear Model and its Summary Statistics

| | Children's Education |
|---|---|
| (Intercept) | 11.32 |
| | [10.79, 11.85] |
| mapres10 | 0.01 |
| | [0.00, 0.02] |
| papres10 | 0.01 |
| | [0.00, 0.02] |
| paeduc | 0.18 |
| | [0.14, 0.22] |
| maeduc | 0.05 |
| | [0.00, 0.09] |
| Num.Obs. | 1980 |
| R2 | 0.124 |
| R2 Adj. | 0.122 |
| AIC | 9156.9 |
| BIC | 9190.5 |
| Log.Lik. | −4572.456 |
| RMSE | 2.44 |

- The RMSE value is 2.44, which is the root mean squared error of the model. It is a measure of how much the predictions of the model deviate from the actual values.

To put it another way, there exists a positive correlation between the educational attainment and societal status of parents with that of their children. However, when examining the specific factors of fathers' education and prestige score, it appears that the correlation with children's education is stronger than that of the mothers' education or prestige score.

This suggests that fathers' educational background may have a particularly strong influence on their children's academic achievement and future opportunities. While the influence of mothers' education and social status is still significant, the impact of fathers' education appears to be even greater in shaping their children's educational outcomes. These findings highlight the importance of parental involvement and investment in their children's education, particularly that of the father's role in this regard.

# 5 Discussion

## 5.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

## 5.2 Second discussion point

## 5.3 Third discussion point

## 5.4 Weaknesses and next steps

As with any research, this study has its limitations and weaknesses. One significant weakness is the use of prestige scores as a variable to measure social status and occupational standing. While attempts were made to capture objective indicators of status, such as income and education, prestige scores are still subject to subjective interpretation. Cultural, social, or personal biases may influence people's perceptions of what constitutes a high-prestige occupation. Furthermore, measures of prestige are not fixed over time and may vary across different societies and historical periods. Heterogeneity within occupations can also lead to different levels of prestige among individuals. Additionally, some measures of prestige may overemphasize certain aspects of an occupation, such as income or education, while overlooking other critical factors like job security or working conditions (Goyder and Frank 2007).

Another weakness of this study is missing data. While the decision to retain all missing values was made intentionally to avoid issues with representativeness, missing data can still affect the statistical power and accuracy of estimates. It can also reduce the representativeness of the sample, potentially limiting the generalizability of the findings. Additionally, missing data can make the analysis of the study more complex, requiring more sophisticated techniques that may increase the risk of errors (Kang 2013).

Moreover, this study did not use hypothesis testing, which could potentially overlook some effects that might be present in the data. Therefore, the generalizability of the findings should be interpreted with caution. Future studies should aim to address these limitations and weaknesses to ensure more accurate and comprehensive research on the topic.

To shed light in future studies, there are some potential improvements. Firstly, it may be beneficial to incorporate additional measurements for social-economic statuses, such as the International Socio-Economic Index of Occupational Status (ISEI). This index is derived from the International Standard Classification of Occupations (ISCO) and comprises comparably coded data on education, occupation, and income from 73,901 full-time employed workers across 16 countries (B. G. Ganzeboom 1 et al. 2004).

Furthermore, new and accurate methodologies could be implemented for data analysis. It is evident that the sample sizes across the groups in Figure 5 are substantially different. Therefore, post-stratification may be a suitable technique to adjust for these discrepancies and verify the results. Moreover, in terms of the statistical model used in the study, a simple linear model was employed with only a few predictors. Future studies could build a more sophisticated model that accounts for potential interaction or quartic terms in the analysis to improve the understanding of the relationship under investigation.

In Figure 6, the results seem to be insignificant. One way to improve is to consider the measurement of the neighbourhood using living regions at age 16 may not provide a comprehensive representation of the impact of the neighbourhood on education. Other factors, such as the quality of schools and access to resources, may also play a significant role in educational outcomes but are not captured by this measure. Therefore, it is possible that the relationship between neighbourhood and education is more complex than what can be inferred from the living region at age 16 alone. While the distribution of education levels among neighbourhoods may appear similar based on Figure 6, it is still possible that there are significant differences in educational outcomes that are not apparent in the bimodal shape of the distribution. Further research using more detailed measures of neighbourhood characteristics and educational outcomes could provide a more nuanced understanding of the relationship between neighbourhood and education.

While this data and study provide valuable insights into children's education and performance, it is important to acknowledge that some key factors have not been included. One such variable is the positive involvement of parents in their children's education. Research has consistently shown that parental involvement can have a significant impact on children's academic success, yet this aspect was not measured in the 2021 GSS (Barger et al. 2019). Therefore, it is essential to include these factors in future studies to fully understand the complex interplay between different variables and their effects on children's education. By including factors like parental involvement, future studies could provide a more comprehensive understanding of the various factors that influence children's education and help to identify effective strategies for improving educational outcomes.

## References


B. G. Ganzeboom 1, Author links open overlay panelHarry, 1, 2, 3, AbstractIn this paper we present an International Socio-Economic Index of occupational status (ISEI), D. J. Treiman, A. Szymanski, et al. 2004. "A Standard International Socio-Economic Index of Occupational Status." *Social Science Research*. Academic Press. https://www.sciencedirect.com/science/article/pii/0049089X9290017B.

Barger, Michael M., Elizabeth Moorman Kim, Nathan R. Kuncel, and Eva M. Pomerantz. 2019. "The Relation Between Parents' Involvement in Children's Schooling and Children's Adjustment: A Meta-Analysis." *Psychological Bulletin* 145 (9): 855–90. https://doi.org/10.1037/bul0000201.

Davis-Kean, Pamela. 2014. "How Does Parents' Education Level Influence Parenting and Children's Achievement." *U. O. M. CAPCA (Center for Analysis of Pathways from Childhood to Adulthood), Editor. In Proc. CDS-II Early Results Workshop*. https://www.academia.edu/2712079/How_does_parents_education_level_influence_parenting_and_childrens_achievement.

"General Social Survey (GSS)." n.d. *NORC at the University of Chicago*. https://www.norc.org/Research/Projects/Pages/general-social-survey.aspx.

Goyder, John, and Kristyn Frank. 2007. "A Scale of Occupational Prestige in Canada, Based on NOC Major Groups." *The Canadian Journal of Sociology* 32 (December): 63–83. https://doi.org/10.1353/cjs.2007.0008.

"GSS Data Explorer: NORC at the University of Chicago." n.d. *GSS Data Explorer | NORC at the University of Chicago*. https://gssdataexplorer.norc.org/gss_data.

Kang, Hyun. 2013. "The Prevention and Handling of the Missing Data." *Korean Journal of Anesthesiology*. U.S. National Library of Medicine. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3668100/.

Khan, Rana Muhammad Asad, Nadeem Iqbal, and Saima Tasneem. 2014. "The Influence of Parents Educational Level on Secondary School Students Academic Achievements in District Rajanpur." *Journal of Education and Practice*. IISTE. No 1 Central, Hong Kong Island, Hong Kong SAR. Tel: +852-39485948; e-mail: JEP@iiste.org; Web site: http://iiste.org/Journals/index.php/JEP. https://eric.ed.gov/?id=EJ1079955.

Müller, Kirill. 2020. *Here: A Simpler Way to Find Your Files*. https://CRAN.R-project.org/package=here.

Nieuwenhuis, Jaap, and Pieter Hooimeijer. 2016. "The Association Between Neighbourhoods and Educational Achievement, a Systematic Review and Meta-Analysis." *Journal of Housing and the Built Environment : HBE*. U.S. National Library of Medicine. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5748572/.

Overman, Henry, and Clément Bosquet. 2016. "Why Birthplace Matters so Much: Sorting, Learning and Geography." *CEPR*. https://cepr.org/voxeu/columns/why-birthplace-matters-so-much-sorting-learning-and-geography.

Pedersen, Thomas Lin. 2022. *Patchwork: The Composer of Plots*. https://CRAN.R-project.org/package=patchwork.

R Core Team. 2020. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

"R Markdown - Dynamic Documents for r." n.d. https://rmarkdown.rstudio.com.

Robinson, David, Alex Hayes, and Simon Couch. 2022. *Broom: Convert Statistical Objects into Tidy Tibbles*. https://CRAN.R-project.org/package=broom.

Smith, Tom William. n.d. "Measuring Occupational Prestige on the 2012 General Social Survey." *Google Books*. NORC at the University of Chicago. https://books.google.com/books/about/Measuring_Occupational_Prestige_on_the_2.html?id=Y2LfsgEACAAJ.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D'Agostino McGowan, Romain François, Garrett Grolemund, et al. 2019. "Welcome to the tidyverse." *Journal of Open Source Software* 4 (43): 1686. https://doi.org/10.21105/joss.01686.

Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2022. *Dplyr: A Grammar of Data Manipulation*. https://CRAN.R-project.org/package=dplyr.

Wickham, Hadley, Evan Miller, and Danny Smith. 2023. *Haven: Import and Export 'SPSS', 'Stata' and 'SAS' Files*. https://CRAN.R-project.org/package=haven.

Wickham, Hadley, and Dana Seidel. 2022. *Scales: Scale Functions for Visualization*. https://CRAN.R-project.org/package=scales.

Xie, Yihui. 2023. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. https://yihui.org/knitr/.

Zhu, Hao. 2021. *kableExtra: Construct Complex Table with 'Kable' and Pipe Syntax*. https://CRAN.R-project.org/package=kableExtra.