Exploring the Relationship between Family Background and Children's Education Level in the United States (2021)*

Ruibo Sun

12 April 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 General Social Survey (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 factors. The findings in this paper suggest that parents' education, family income, and prestige score all affect the children's education level. 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). Ultimately, these may affect the education level of the children.

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

*Code and Data: https://github.com/ruibosun/revised-how-parents-affect-childrens-education 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).

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 my previous paper, there are several findings. The average years of schooling in the US is 14, which is higher than the global average of 8.7 years. A bachelor's degree and a graduate degree are more common than an associate or junior college degree, and the US ranks behind several countries in terms of higher education attainment for 25-34 year-olds. Parental education levels have a positive correlation, and there is a strong link between parents' education levels and their children's academic achievement and societal status. Fathers' education levels appear to have a stronger impact than mothers' on children's academic achievement, but other factors may also play a role. Further research is needed to fully understand the relationship between these variables. Policymakers can use this information to make informed decisions about resource allocation and policy development. Here is a link to view the entire paper. However, the prior study did not appropriately handle the issue of missing values (NAs). Consequently, this new paper aims to conduct a more rigorous analysis by appropriately addressing the NAs. Moreover, it will address the limitations of the previous study to produce more accurate and representative results.

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 `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). The missing values were imputed using `mice` (van Buuren and Groothuis-Oudshoorn 2011). This paper is knitted as a PDF file by the packages of Quarto and `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. 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 limitations will be explained in detail in Section 5.4.

$$\text{Prestige}_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij} \tag{1}$$

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. In this study, the "children" are the respondents and "parent" are the parents of the respondents.

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 100, with higher scores indicating greater prestige. The ideal dataset would have 90,090 ratings from 1,001 raters, but due to various issues, only 82,800 ratings from 946 raters were used. Invalid or inconsistent ratings, such as cases with reversed codes, too few ratings, equal scores for all occupations, or a low standard deviation of ratings, were excluded to ensure the reliability of the occupational scores (Smith, n.d.).

Next, the prestige score is obtained by the Equation 1. Prestige_{ij} is the prestige score for occupational title i and rater j. The α_i represents the occupational differences of interest, while the β_j represents differences among the raters that need to be controlled for in estimating the α_i. The μ is the overall mean prestige score, and ϵ_{ij} represents the random error term. The range of the prestige scores is from 0 to 100, with higher scores indicating a better or more prestigious occupation (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**, **degree**, **born**, **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.
- **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.
- **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.

As mentioned before, this paper is an revised version of the previous paper which has some limitations and weaknesses. In order to improve the representiveness of the data, this paper use the multiple-imputation and dropping methods (Gelman and Hill 2006). There are two types of missing values: numerical and categorical missing values.

For numerical variables (**paeduc**, **maeduc**, **educ**, **mapres10** and **papres10**), multiple imputation techniques are employed to estimate and fill in the missing data points. This method

Table 1: Number of respondents by degree for 2021 survey

Degree	Total	Proportion
graduate	731	19.3%
bachelor's	1010	26.6%
associate/junior college	362	9.5%
high school	1507	39.7%
less than high school	184	4.8%

generates a more accurate dataset without the missing values. Following the imputation, the resulting dataset is transformed into a new one with no NAs for further analysis.

Next, any categorical variables with missing values are addressed, such as **degree**, **born**, **reg16** and **incom16**. In this case, all these variable with NAs are removed from the data, ensuring that the analysis is only performed on complete cases. This helps to minimize any biases that might arise due to incomplete data.

The number of observations before data cleaning is 4032. This number decrease to 3794 because some rows with missing values. are dropped from the data.

In addition, the variable of **degree** is grouped into two levels. For respondents who have obtained at least a degree from an associate/junior college, their levels are denoted as "1", indicating they have completed a post-secondary education. For respondents who have attained only a high school diploma or possess a lower level of education, their levels are designated as "0".

3 Results

This following findings are from the 2021 General Social Survey (GSS) data.

According to the Table 1, the distribution of education levels in terms of degrees shows that the most prevalent category is high school education, with 1507 individuals (39.7%) falling into this group. Next in prevalence is a bachelor's degree, held by 1010 individuals (26.6%). A graduate degree is the third most common category, with 731 individuals (19.3%) having attained this level of education. In contrast, the associate/junior college degree is less frequent, with 362 individuals (9.5%) completing this level of education. The smallest proportion, 4.8%, have less than a high school education.

The second variable utilized to quantify education is relation to the number of years of schooling. Figure 1 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 higher levels of education. In other words, the bulk of the respondents are clustered towards

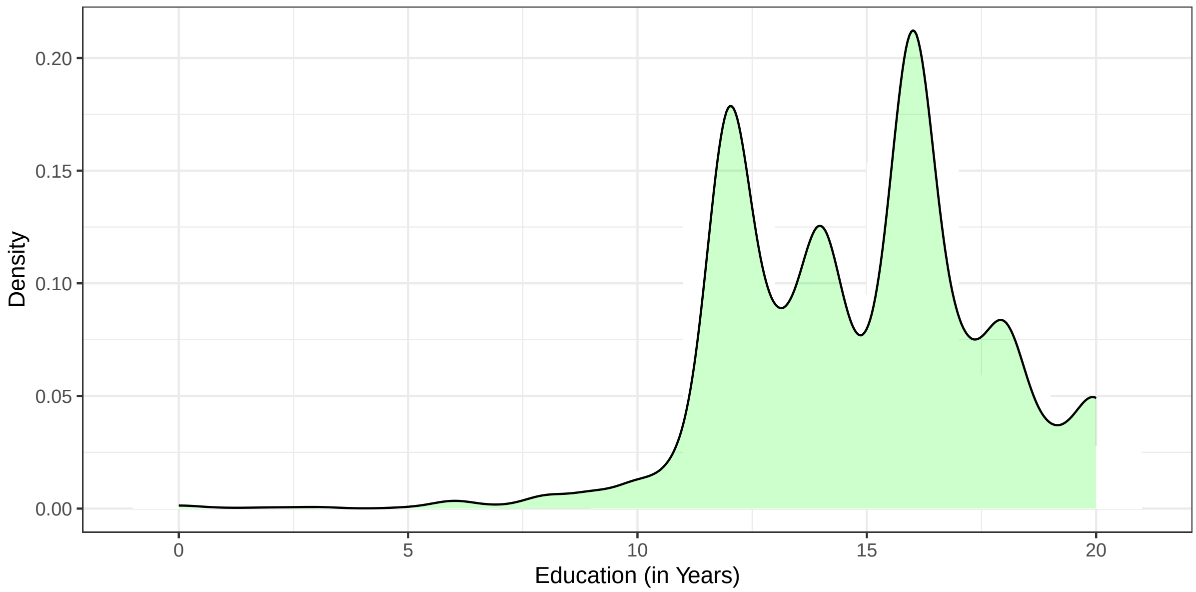

Figure 1: Histogram of Respondents' Education

Table 2: Number of respondents by education for 2021 survey

Min	Median	Max	Mean
0	15	20	14.82

the left side of the histogram (lower levels of education), and there are relatively fewer respondents towards the right side of the histogram (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, 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 14.82 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.

Figure 2 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. Specifically, the distribution of mother's prestige score is bimodal and the distribution of father's prestige score is unimodal.

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

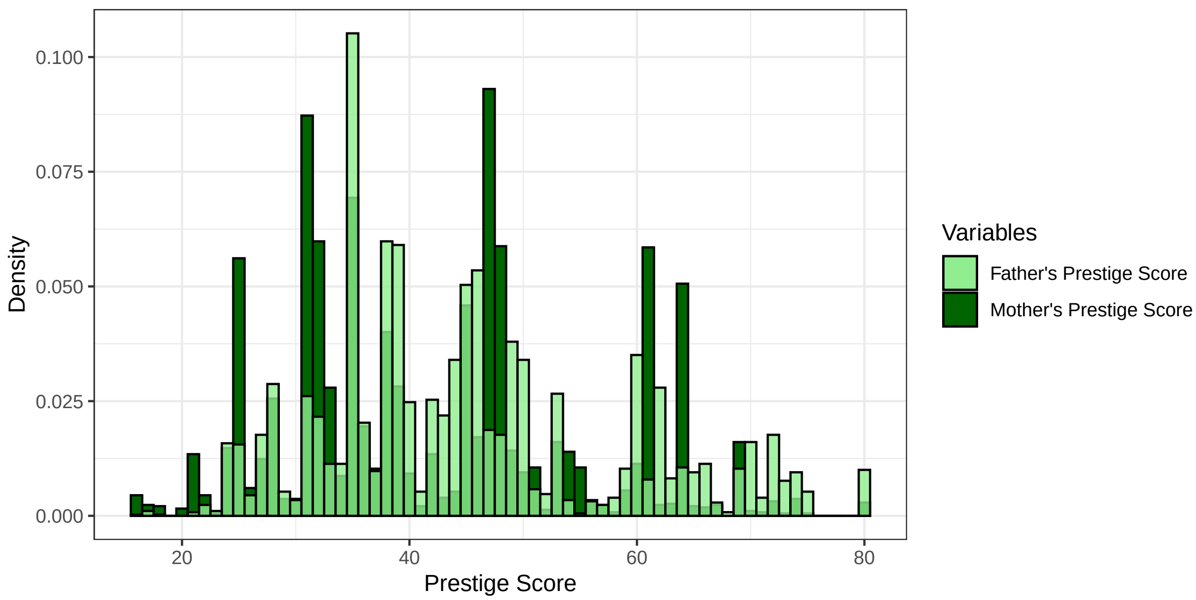

Figure 2: Mother's and Father's Prestige Score

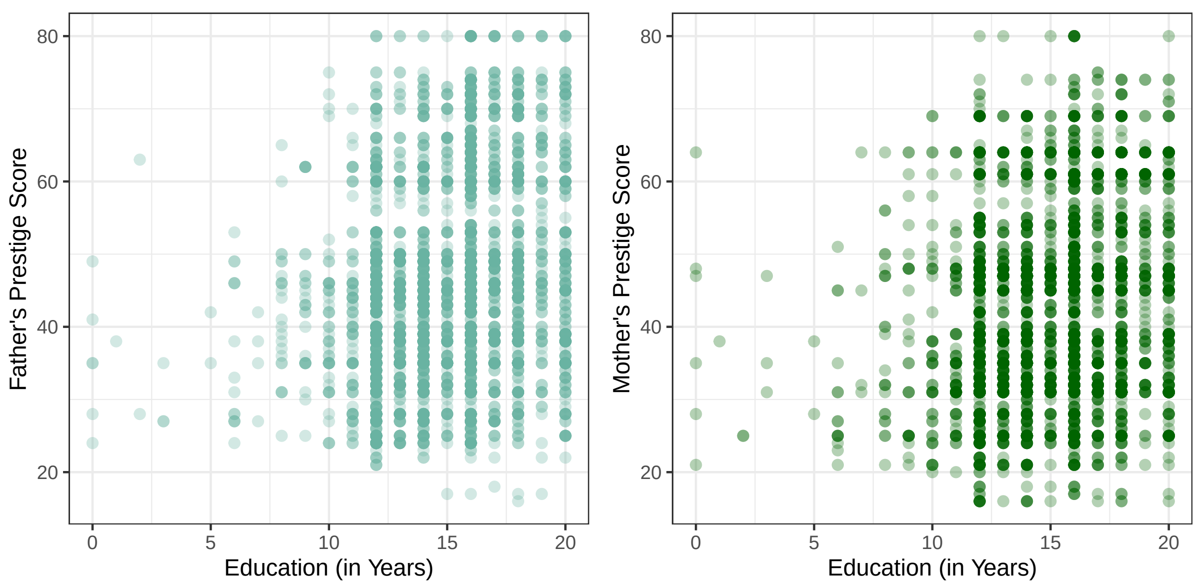

Figure 3: Mother's and Father's Prestige Score with children's education

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.

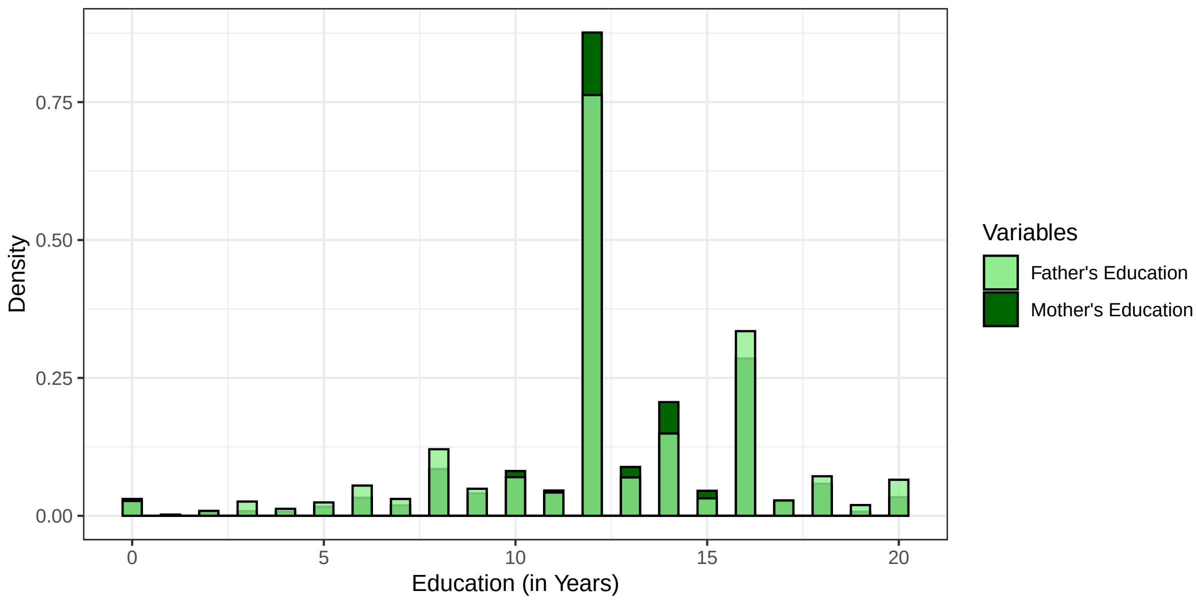

Figure 4: Mother's Education and Father's Education Level

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.

Overman and Bosquet (2016) suggests 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 90% of the respondents were born in the US, and only 10.0% answered "No". This is because respondents who did not answer this question or answered NA are removed during the data cleaning. Given the huge differences in the sample sizes between groups, it is hard to draw a general and representative conclusion of whether birthplace may affect the education level.

(**tb-regandeduc?**) showcases the average number of years of education for individuals in different living regions of the United States when the respondent was 16. The highest average education level is found in the Middle Atlantic region, where the mean is 15.13 years. The East North Central region follows closely, with an average of 14.98 years of education. In

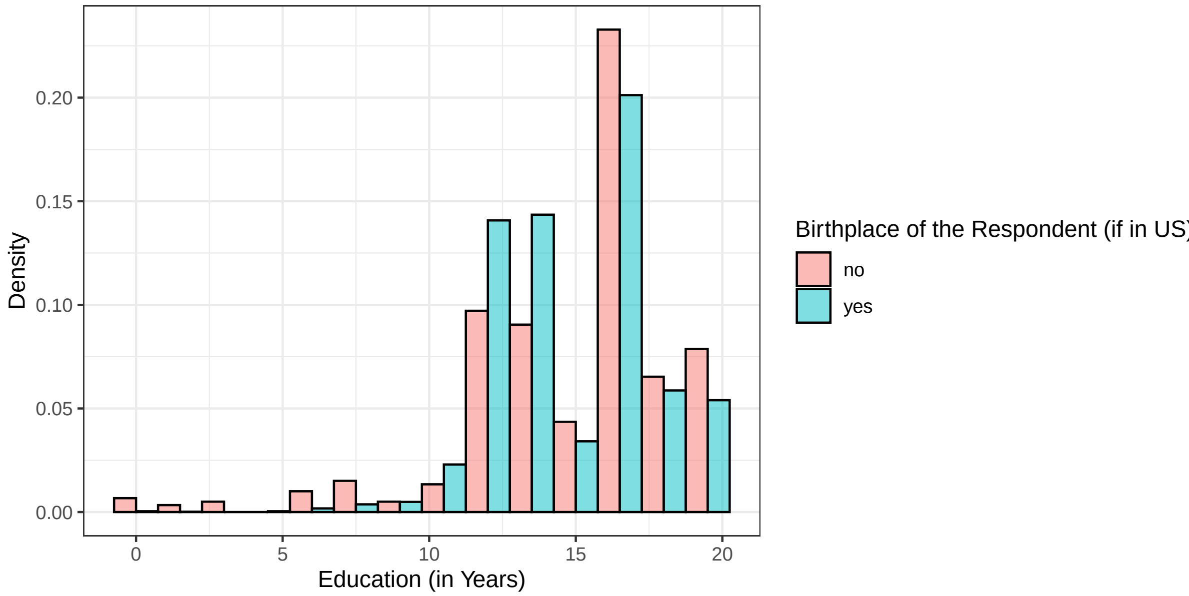

Figure 5: Birthplace of the Respondent (if in US)

Table 3: The proportion and mean of education for different regions

Region	Count	Percent	Mean of Education
east north central	778	20.51%	14.98
east south atlantic	219	5.77%	14.10
middle atlantic	524	13.81%	15.14
mountain	229	6.04%	14.69
new england	165	4.35%	14.85
pacific	715	18.85%	14.94
south atlantic	509	13.42%	14.75
west north central	320	8.43%	14.73
west south central	335	8.83%	14.41

the Pacific and New England regions, the mean years of education are relatively similar, with 14.94 and 14.85 years, respectively. Those regions are above the overall average, which is 14.82 from (**tb_educ?**). On the other hand, the South Atlantic and West North Central regions also display comparable average education levels of 14.75 and 14.73 years, respectively. The Mountain region exhibits a slightly lower mean of 14.69 years. Furthermore, the East South Atlantic and West South Central regions have the lowest average education levels, with 14.10 and 14.39 years, respectively.

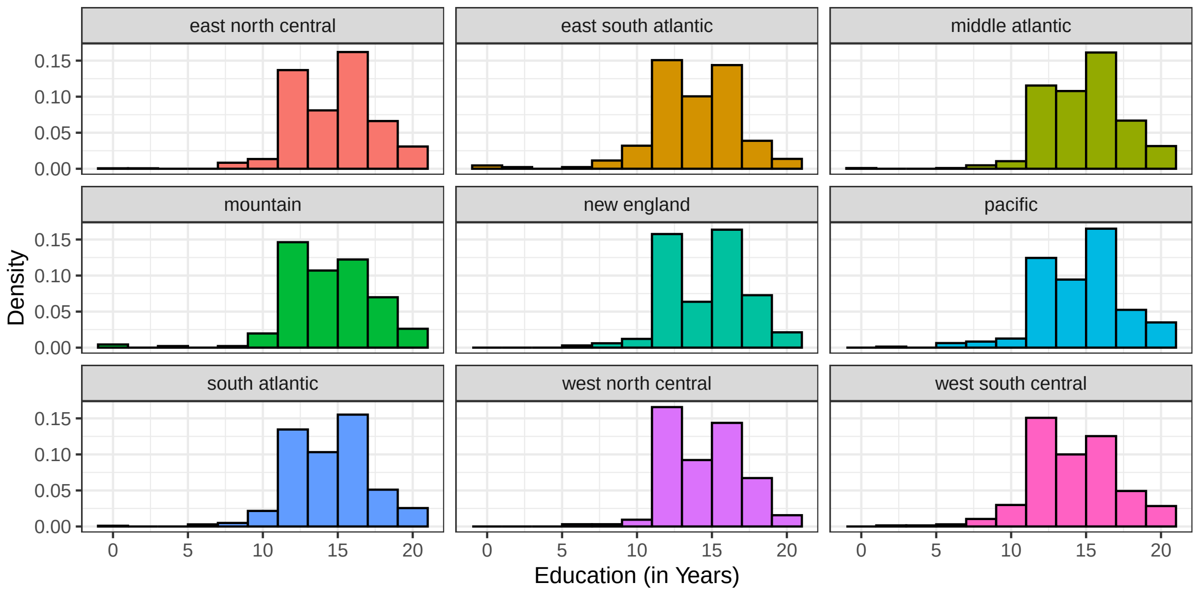

Figure 6: Respondent's Living Region at Age of 16 v.s. Education

Nieuwenhuis and Hooimeijer (2016) finds 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. 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 will be discussed in the Section 5.4.

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**. 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,

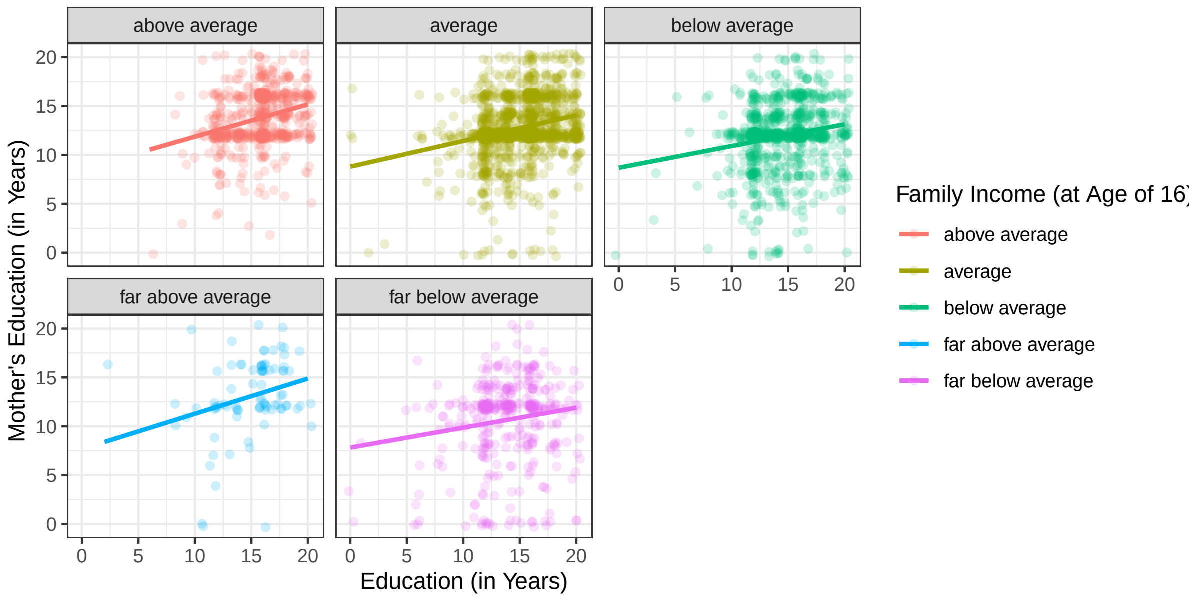

Figure 7: Education vs Mother's Education with respect to Family Income

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.

4 Model

We estimate the following mode:

$$\log\left(\frac{\pi}{1-\pi}\right) = \hat{\beta}_0 + \hat{\beta}_1 X_{\text{mapres10}} + \hat{\beta}_2 X_{\text{papres10}} + \hat{\beta}_3 X_{\text{paeduc}} + \hat{\beta}_4 X_{\text{maeduc}} \tag{2}$$

Table 4: Logistic Model and its Summary Statistics

	Children's Education
(Intercept)	−2.19
	[−2.64, −1.74]
mapres10	0.01
	[0.00, 0.02]
papres10	0.01
	[0.01, 0.02]
paeduc	0.11
	[0.09, 0.14]
maeduc	0.02
	[0.00, 0.05]
incom16average	−0.30
	[−0.51, −0.10]
incom16below average	−0.40
	[−0.62, −0.18]
incom16far above average	−0.03
	[−0.53, 0.48]
incom16far below average	−0.50
	[−0.77, −0.22]
reg16east south atlantic	−0.14
	[−0.45, 0.18]
reg16middle atlantic	0.28
	[0.04, 0.52]
reg16mountain	−0.17
	[−0.48, 0.14]
reg16new england	0.13
	[−0.22, 0.49]
reg16pacific	0.03
	[−0.19, 0.25]
reg16south atlantic	−0.05
	[−0.28, 0.19]
reg16west north central	−0.14
	[−0.42, 0.13]
reg16west south central	−0.28
	[−0.56, −0.01]
Num.Obs.	3794
AIC	4817.8
BIC	4923.9
Log.Lik.	−2391.899
F	22.471
RMSE	0.47

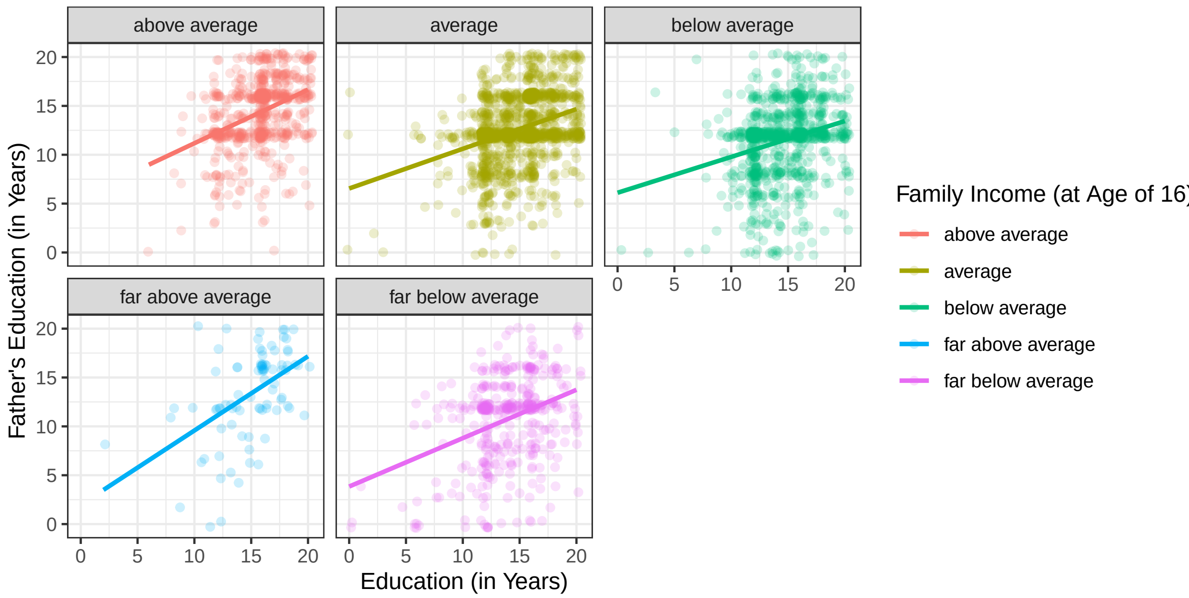

Figure 8: Education vs Father's Education with respect to Family Income

The Table 4 is built based on the Equation 2. In this logistic model, it is used to predict the probability of a child attains a post-secondary degree based on these predictors. *missing the bullet point to explain each exp(coef)*

From Table 4, the positive coefficients for both mother's and father's prestige scores (**mapres10** and **papres10**) suggest a slight positive relationship with children's education levels. Moreover, the positive coefficients for parents' education (**paeduc** and **maeduc**) have a greater positive association with children's education level.

The coefficients for the income level at age of 16 (**incom16**) indicate that higher-income levels are generally associated with lower education levels for children. Particularly, the "far above average" and "far below average" income categories show a significant negative relationship with children's education levels.

Regarding the living regional differences at age of 16 (**reg16**), the results indicate that children from the east south atlantic region may have lower education levels compared to those from other regions. The positive coefficients for middle atlantic and west north central regions suggest a positive association with children's education levels, although the results are not statistically significant.

5 Discussion

5.1 Average years of schooling in the US and Outside of US

Education has been a crucial driver of social and economic progress worldwide, and its expansion over the past two centuries has been a significant phenomenon. The global average year of schooling currently stands at around 8.7 years, according to data from World Economics ("Average Years of Schooling," n.d.).

In this paper, by examining the average year of schooling in the United States, based on 2021 GSS data, it was found that the average year of schooling is 14 in Table 2. This result is consistent with the findings from Our World in Data, which indicate that the average year of schooling in the US is also 14 years (Roser and Ortiz-Ospina 2016).

Furthermore, the data in Table 1 reveals that a bachelor's degree and a graduate degree are the second and third most common educational categories in the US, respectively. This finding suggests that having an associate or junior college degree is less common compared to the bachelor's or graduate degree. Conversely, a smaller proportion of individuals have less than a high school education.

The Organisation for Economic Co-operation and Development (OECD) reports that the US is among the most well-educated countries globally, with 42% of 25-64 year-olds holding tertiary attainment, behind only Canada, Israel, Japan, and the Russian Federation ("Education at a Glance 2012 (Summary in English)" 2012). However, the US ranks 14th out of 37 OECD and G20 countries in the percentage of 25-34 year-olds with higher education, with 42% - above the OECD average but significantly behind Korea, which is the leader with 65%.

This data on education levels is not only important for understanding social and economic trends in the US, but it also has implications for policymaking and decision-making related to education. By accurately understanding the educational distribution of a given population, policymakers can make informed decisions about how to allocate resources and develop policies that address the needs and opportunities of different groups.

5.2 The impact of parental education level

In the Section 3, the Figure 4, Figure 8 and Figure 7, they all show that there is a positive correlation between the education levels of mothers and fathers. If one parent has a high level of education, it is likely that the other parent also has a high level of education.

The correlation between parents' education levels and their children's educational outcomes is well-established in the literature. Studies have consistently shown that parents' education is one of the strongest predictors of their children's academic achievement and educational attainment (2012). Chetty, Friedman, and Rockoff (n.d.) used data from tax records and college enrollment records to analyse the relationship between parents' income, education, and social status, and their children's earnings and educational outcomes. The results showed that children from high-income families were more likely to attend college and earn higher incomes than children from low-income families. In addition, the study found that parental education was a key factor in determining children's educational and economic outcomes. Children with highly educated parents were more likely to attend college and earn higher incomes than children with less educated parents.

Furthermore, the analysis of the statistical model from the Table 4 shows that parental education and prestige scores are positively correlated with their children's academic achievement and societal status. The coefficients for the independent variables father's and mother's education in years indicate that an increase in fathers' education has a greater association with their children's education than that of mothers. This suggests that the father's role in education and investment in their children's future may be particularly important. However, the influence of both parents is still significant. The R-squared value suggests that the independent variables explain only a small portion of the variation in the dependent variable, indicating that other factors beyond parental education and prestige score also play a role. Nonetheless, these findings highlight the importance of parental involvement in their children's education, with a particular emphasis on the role of fathers.

The relationship between parental education and student achievement has been the focus of many studies, with varying results. Some study has found that maternal educational attainment had an impact on student achievement, which contradicts the finding in this paper (Crockett, Eggebeen, and Hawkins 1993). The present study showed that paternal support of education has a significant impact on academic achievement, while maternal education level did not show a significant impact. However, the study of Cabrera et al. (2000) showed that families with active fathers fostered maternal involvement in the household, which resulted in a support system that fosters educational attainment and positive long-term contributions.

Despite these discrepancies, the findings in this paper concur with prior literature that emphasizes the importance of parental education levels and their impact on college student achievement (Arias Ortiz and Dehon 2008). This study focuses on showing that the fathers' education level has a significant relationship with childrens' year of schooling, while no statistically significant relationship was observed with the mothers' education level.

These findings highlight the importance of parental involvement and support in their children's education, with a particular emphasis on the role of fathers. The results suggest that fathers' education levels may have a stronger impact on their children's academic achievement than that of mothers. However, it is important to note that other factors, such as family income and race, may also play a role in student achievement. Therefore, further research is necessary to fully understand the complex relationship between these variables and their impact on academic achievement, which will be carefully discussed in Section 5.4.

5.3 The Effect of Living Region during the Adolescent Period

referring results from the model about how regions affect the education level

5.4 Weaknesses and next steps

5.4.1 Weaknesses

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).

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.

5.4.2 Next Steps

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 the 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.

In the Table 4, a logistic model is employed with the degree as the response variable. However, using a linear regression model with years of education as the response variable may provide a more detailed examination of the factors involved. A linear regression can help analyze specific relationships, such as how years of education change with a five-unit increase in parents' prestige scores. This study opted not to use the linear model due to violations of underlying assumptions. Future research may improve upon this by constructing a more sophisticated linear regression model to elucidate the relationship between these variables in a more precise manner.

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.

Lastly, the study's data was limited to individuals who responded to the 2021 GSS survey in the US, which may not be representative of the entire population. Respondents who chose to participate in the survey may have different characteristics than non-respondents, which could affect the generalizability of the study's findings to the wider population.

6 Appendix

6.1 Link to the Survey

Please view the survey by this link.

6.2 Supplementary Survey

1.What is your gender?

- Male
- Female
- Non-binary
- Woman
- Man
- Prefer not to say

2.What is your race?

- White
- Black or African American
- Hispanic or Latino
- Asian
- Native American or Alaska Native
- Native Hawaiian or other Pacific Islander
- Other

3.Do you consider yourself as an immigrant?

- Yes
- No
- Maybe

4.US citizen

- Yes
- No

5.What is the highest level of education you have achieved?

- Less than high school
- High school diploma or equivalent
- Some college or associate degree
- Bachelor's degree
- Graduate degree or higher

6.What type of school did you attend in your teenage?

- Public school
- Private school
- Homeschool
- Other

7.High school GPA (if applicable)

8.College/University GPA (if applicable)

9.What is the highest level of education achieved by your mother and father?

- Less than high school
- High school diploma or equivalent
- Some college or associate degree
- Bachelor's degree
- Graduate degree or higher

10.What was your parents' occupation in your teenage?

- Professional/managerial
- Technical/sales
- Administrative/clerical
- Blue collar/other manual
- Other

11.What is the highest level of education completed by your mother?

- Less than high school
- High school graduate
- Some college or technical school
- Bachelor's degree
- Graduate degree (Master's or Doctorate)

12.What is the highest level of education completed by your father?

- Less than high school
- High school graduate
- Some college or technical school
- Bachelor's degree
- Graduate degree (Master's or Doctorate)

13.How involved are your parents in your education?

- Not involved
- Somewhat involved
- Moderately involved
- Highly involved

14.What is your family's income level in your teenage?

- Far above average
- Above average

- Average
- Below average
- Far below average

15.Where did you live in your teenage years?

- Rural area
- Suburban area
- Urban area

16.How much pressure do you feel from your parents to excel academically?

- None
- A little
- Moderate
- High
- Very high

References

2012. https://www.annualreviews.org/doi/abs/10.1146/annurev.soc.012809.102503.

Arias Ortiz, E., and C. Dehon. 2008. "What Are the Factors of Success at University? A Case Study in Belgium." *CESifo Economic Studies* 54 (2): 121–48. https://doi.org/10.1093/cesifo/ifn012.

"Average Years of Schooling." n.d. *World Economics*. https://www.worldeconomics.com/Indicator-Data/ESG/Social/Mean-Years-of-Schooling/.

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.

Cabrera, Natasha, Catherine S. Tamis-LeMonda, Robert H. Bradley, Sandra Hofferth, and Michael E. Lamb. 2000. "Fatherhood in the Twenty-First Century." *Child Development* 71 (1): 127–36. https://doi.org/10.1111/1467-8624.00126.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. n.d. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review*. https://www.aeaweb.org/articles?id=10.1257%2Faer.104.9.2633.

Crockett, Lisa J., David J. Eggebeen, and Alan J. Hawkins. 1993. "Father's Presence and Young Children's Behavioral and Cognitive Adjustment." *Journal of Family Issues* 14 (3): 355–77. https://doi.org/10.1177/019251393014003002.

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.

"Education at a Glance 2012 (Summary in English)." 2012. *Education at a Glance 2012*. https://doi.org/10.1787/eag-2012-sum-en.

Gelman, Andrew, and Jennifer Hill. 2006. "Data Analysis Using Regression and Multilevel/Hierarchical Models." https://doi.org/10.1017/cbo9780511790942.

"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.

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/.

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

Roser, Max, and Esteban Ortiz-Ospina. 2016. "Global Education." *Our World in Data*. https://ourworldindata.org/global-education.

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.

van Buuren, Stef, and Karin Groothuis-Oudshoorn. 2011. "mice: Multivariate Imputation by Chained Equations in r." *Journal of Statistical Software* 45 (3): 1–67. https://doi.org/10.18637/jss.v045.i03.

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.