JOURNAL OF
EMERGING INVESTIGATORS

Article

The impact of attending a more selective college on future income

Edward Ho[1], Jingying Ye[1]
[1] Newark Academy, Livingston, New Jersey

SUMMARY

There have been many debates regarding legacy preferences, recruited athletes, and affirmative action in the U.S. college admission system. These issues hinge upon the premise that being accepted into a more selective institution would result in higher future earnings for graduates. We hypothesized that graduating from a more selective college could increase a student's future earnings by providing students with more learning resources, connections, and opportunities. If this is the case, then we expected there to be a significant positive correlation between college selectivity and future earnings even after accounting for selection bias, as more selective schools tend to admit students who already demonstrate characteristics correlated with future success. We used weighted linear regressions to determine if college selectivity is a significant indicator of future income based on institutional level data from students entering college in 2001. Before controlling for measurable student characteristics before enrollment, attending a more selective college appeared to have a significant positive effect on future income. However, when we adjusted for measurable student characteristics such as SAT scores, race, gender, major choice, and parental income, the selectivity of the college attended became statistically insignificant. Thus, attending a selective school may not have a significant effect on the median predicted income of students, meaning that programs influencing college admissions, like affirmative action, may not have significant effects of future incomes.

INTRODUCTION

Recently, there has been great controversy over the college admissions process regarding practices including legacy preferences, recruited athletes, and affirmative action. Legacy preferences and recruited athletes have usually served to benefit the already privileged (1, 2). Meanwhile, affirmative action in college admissions was designed to help reduce the racial wealth gap and benefit the least privileged (3). The impact of these issues depends upon the benefits of going to a more selective school.

The often-cited benefits of going to more selective school include smaller class sizes, state-of-the-art facilities, and more research and networking opportunities (4). With these benefits, students attending more selective schools would be expected to have higher salaries after graduation. However, some have argued that graduates of more selective schools have higher salaries because of selection bias, claiming that students accepted into selective schools possess qualities that would lead to future success regardless of the school attended (5).

There exists extensive literature on the effects of college selectivity on future earnings. Recently, Chetty et al. studied these effects by examining the future earnings and successes of students who were waitlisted at Ivy League institutions and accepted compared to those who were waitlisted and then rejected. They claimed that being accepted off a waitlist was based upon random factors, noting that colleges often used waitlists to balance their classes. The students they favor may not exhibit characteristics not predictive of future income, but instead have a specific characteristic, like being able to play the flute or bassoon, that the college needs. The paper found that going to an Ivy League school did not greatly increase the average income of a student but did increase the earnings of the top decile (6).

Dale and Kruger used regression models to examine these effects on the 1976 and 1989 cohorts of college graduates, using survey data from individuals. The authors used many control variables, including race, parental income, and Scholastic Aptitude Test (SAT) scores and notably accounted for unmeasurable student "motivation" using the number of college applications submitted by a student. They found that going to a more selective school increased earnings by a modest but still statistically significant amount before accounting for student "motivation." After accounting for "motivation," that correlation became insignificant (5).

Here in this study we attempted to determine whether going to a more selective college still resulted in higher future earnings after accounting for selection bias. Like Dale and Kruger, we used weighted linear regressions in this study. However, we used institution-level data from college graduates as opposed to the survey data from individuals used in Chetty et al. and Dale and Kruger (5, 6). This allowed us to control for some covariates that were not always available on an individual basis, like major choice. However, we omitted variables that were not directly measurable without surveys, like motivation.

The use of institutional-level data meant that we could only study the median characteristics of each college, which restricted our conclusions to only the median student. However, this process effectively expanded our sample size, as we could then study the median characteristics of students in 607 American colleges where this data was available. While Chetty et al. only compared the effect of going to an Ivy League college to other fairly selective colleges, this paper attempted to measure the effects of selectivity across a broader range

https://doi.org/10.59720/23-317

of colleges (6). The colleges were categorized by selectivity using Barron's selectivity index, which ranked schools into tiers using a variety of factors, including admission rates and perceived prestige (7).

Here we used graduate income data from the College Scorecard, parental income data from Opportunity Insights, and all other student characteristic data from the Integrated Postsecondary Education Data System (IPEDS) (7). All the data were from years that corresponded with the 2001 cohort of college entrants Using this regression model, we found that the impact of college selectivity on the future income of students dropped in degree and significance as we added covariates and eventually grew insignificant.

RESULTS

To study whether attending a selective college increased a student's future earnings, we used linear regressions to analyze school level data on the characteristics and future earnings of graduates from the College Scorecard, Opportunity Insights, and the IPEDS.

We controlled for the earning potential of a student before enrolling to account for selection bias in selective schools. The difficulty with this process was that not all characteristics of a high school student that impact future success can be easily measured. Examples of these characteristics include determination and creativity, which are revealed to admissions officers through extracurricular activities, essays, and interviews but were unknown to us. We used Equation 1 to relate future income to student characteristics,

$$Y_i = \beta_0 + \beta_1 Q_i + \beta_2 X_{1i} + \beta_3 X_{2i} + \varepsilon_i \quad (1)$$

where Q_i is a measure of the selectivity of the college attended, X1s are the measured characteristics that affect earnings, and X2s are the unmeasured characteristics that affect earnings. β1 then represents the estimated effect of attending a more selective college isolated from the selection bias of those colleges. However, within our regression, unobserved characteristics cannot be included, so the equation used is

$$Y_i = \beta_0 + \beta_1 Q_i + \beta_2 X_{1i} + \varepsilon_i \quad (2)$$

which is the same as Equation 1, except β3X2s is now part of εi, the error. A more selective college would be more inclined to admit a student with higher β3 in Equation 1 assuming comparable β2 values, so β3 is assumed to be positively correlated with β0, meaning that removing β3X2i would be expected to bias the equation by increasing the future earnings of students attending more selective colleges. Thus, the β1 value derived in (2) was assumed to be higher than β1 in equation 1, allowing the β1 in our regression equation to serve as an upper bound for the "true" impact of β1 if unmeasured characteristics are included. However, a lower bound of β1 in (1) cannot be established, meaning that this study could not prove that β1 is significant in (1) but could prove that β1 is insignificant in (1) if β1 is insignificant in our regression equation.

A weighted linear regression was used as opposed to an ordinary linear regression because the variances were unequal. The weight of each variable corresponded to the variance of the variable via the formula:

$$\text{Weight} = 1/\text{Standard Deviation}^2 \quad (3)$$

There was no need to transform the data in any way, as though the data were not always linearly distributed, the residuals were normally distributed.

We performed one regression with various covariates (**Table 1**), while the other without. The results of the regressions included predicted relationships between all covariates and the median incomes of students after graduation. These relationships were described in the results by the coefficient of the variable, which represents the variables' impact, as well as the statistical significance of the relationship. Only the relationship between the college selectivity and future median income was meaningful to this study (**Figure 1**).

Before adding any covariates, the importance of going to a selective university was clearly apparent. The Barron's coefficient was $4821, meaning that the regression predicted that for every tier increase in Barron's selectivity index, median yearly earnings increased by more than $4821 (**Table 2**). This would mean that attending a "most selective" college instead of an "unselective" college, which represents a five-tier difference, would increase the median expected earnings of a student by more than $24,000 every year. This increase appeared to be statistically significant, with a p-value of < 2e-16 (**Table 2**). However, after accounting for selection bias by adding various controls, such as parental income, SAT scores, major, and demographics, the Barron's coefficient decreased significantly to $151 (**Table 3**). This slight relationship was found to be statistically insignificant, with a p-value of 0.650 (**Table 3**). The standard deviation error bars overlapped with the x-axis, which suggested that

Covariate	Variance Inflation Factor (VIF)
Barron's Selectivity Score	3.91
Average SAT Score	5.45
Median Household Income	3.61
Percent STEM Major	1.53
Percent Business Major	1.48
Percent Arts/Humanities Major	1.99
Percent Health Major	1.35
Percent Social Science Major	1.33
Percent Public Social Major	1.25
Percent Asian/Pacific Islander	2.26
Percent African American	9.29
Percent Non-White Hispanic	2.63
Percent White	1.38
Percent Female	10.64

Table 1: The multicollinearity of each regression covariate. Table showing the Variance Inflation Factor (VIF) of each covariate used in the regression. The VIF measures the multicollinearity between regression variables. A VIF of above 5 is associated with significant multicollinearity, which can lower the significance of the regression coefficient and p-values associated with the covariate.

https://doi.org/10.59720/23-317

college selectivity likely had no statistically significant effect on the median yearly incomes of graduates (**Figure 1**). Thus, the selectivity of college attended was not a useful predictor for median future income in our model.

The coefficients of variables other than selectivity were not considered here, as they were not the main focus of our investigation and were solely used as controls. Additionally, since some covariates had a high Variance Inflation Factor (VIF), their predictive power could not be determined (**Table 1**) (9). The variable of interest, Barron's Selectivity Index, was not significantly correlated with the controls (VIF < 5), which allowed the p-value to be used in this study (**Table 1**). This allowed us to claim that the covariates as a whole were better predictors of future earnings than Barron's Selectivity Index. Further research would be required to determine which of these covariates were reliable predictors of future income. As some covariates exhibited signs (VIF > 5) of multicollinearity, a different methodology may be required to accomplish this.

DISCUSSION

We examined the 2001 cohort of college entrants from 607 U.S. colleges in this study. Contrary to our hypothesis, our study found that the positive and significant effect of going to a more selective university on median income after graduation became insignificant after we accounted for measurable student characteristics like SAT scores and demographic factors. Therefore, evidence against our hypothesis that graduating from a more selective college would increase the future earnings of students was found. This suggested that a student's characteristics, including their demographic variables and SAT scores, before attending college were better predictors of future income. Thus, programs that impact college admissions, like legacy preferences, recruited athlete programs, and affirmative action in college admissions, may have little impact on the future earnings of students.

Dale and Kruger came across a similar conclusion after using weighted linear regressions to study the 1976 and 1989 cohorts of college graduates. Unlike our study, they used individual-level data from only 27 different schools and slightly different covariates, including omitting the major chosen by a student and using the number of applications made by a high school student as a measure of student motivation (5). Chetty et al. similarly concluded that going to an Ivy League school had little effect on the future average earnings of a student (6). However, they only compared Ivy and Non-Ivy graduates, as opposed to colleges across all tiers, and used individual-level instead of college-level data. Chetty et al. were additionally able to conclude that the most successful Ivy League graduates were more successful than the most successful non-Ivy-League graduates (6). As only the median future income of students was available on a school-by-school basis, our study was unable to affirm or dispute that claim.

Our study suggested that for most students, their personal characteristics and circumstances mattered much more for their future income than which college they attended. We thus attributed the apparent difference in success between graduates of highly selective and unselective schools to selective schools accepting students who were already projected to be successful after graduation.

Based on the results of this study, affirmative action in college admissions might only change the college selectivity variable of a student, which is insignificant to the predicted income of that student. Removing legacy preferences and recruiting athlete programs would introduce more socioeconomic diversity onto the campus, but this study seemingly suggested that they might not significantly impact the future outcomes of students.

Instead, programs to address social inequality before the college application process may be more beneficial. The covariates as a whole were better predictors of future income and contain variables including parental income, race, and SAT score. Programs that promote equitable primary and secondary education and those that directly reduce the racial wage gap by increasing incomes of underprivileged groups, like affirmative action in the workplace (10), may have a more substantial impact in addressing social inequality.

However, certain factors limited our ability to conclusively make this inference. Notably, our study only examined the median incomes of graduates, so there remained the potential that college selectivity may matter more for the top-performing students or the bottom-performing students in the class (6). Programs like affirmative action may also impact students in other ways, as they aim to introduce more perspectives by way of diversity onto the campus. In addition, by studying the effects of college selectivity on students as a whole, it is

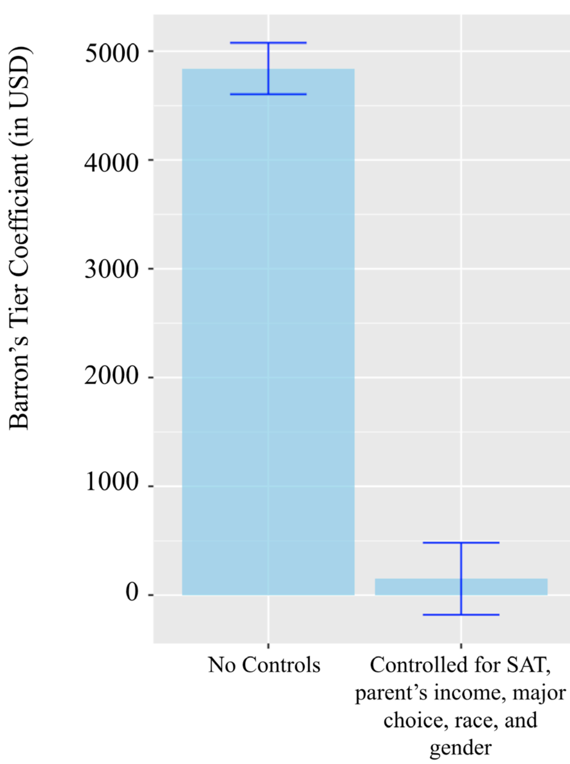

Figure 1: The apparent correlation between college selectivity and future income with and without controls. Bar graph showing Barron's Selectivity Score coefficient with and without controls (SAT scores, parental income, major choice, and demographic characteristics). The Barron's Tier Coefficient on the y-axis represents the increase in median yearly income after graduation per each increase in degree of selectivity. Weighted linear regressions were performed comparing future income with college selectivity with data from 607 American colleges. Error bars present Standard Deviation.

	Estimated variable coefficient	Std. Error	p-value
Intercept	27729.97	727.49	<2e-16 ***
Barron's Selectivity Score	4841.28	236.12	<2e-16 ***

Table 2: The effect of attending a selective college on future income without controls. Table with the intercept and variable coefficients in U.S. dollars, standard errors, and p-values of the weighted linear regression. Median Household income is in U.S. dollars per year. Barron's Selectivity Score is in defined tiers from 1 to 6. The variable coefficient represents the expected change in future expected income when that variable is increased by one unit. The intercept represents the projected income when all variables were zero. The coefficient of the intercept is not relevant to the study. In this weighted linear regression, the superscript *** indicates statistical significance at the 1% level.

possible that trends specific to certain groups were missed. For example, Ge et. al. found that college selectivity only had a causal impact on future earnings for women (11). Future research using data specific to different populations would be useful to determine the effect of college selectiveness on these groups.

Additionally, choosing weighted linear regression as our analytical approach introduces limitations on our results. The regression may be affected by omitted variable bias, which occurs when a covariate correlated with the independent variable is not included and can potentially skew our results (12). Introducing these omitted variables might have impacted the regression, leading to different, potentially more accurate, coefficients. The degree of this bias cannot be measured using statistical tests. Despite including a wide variety of covariates to reduce this bias, a few potentially significant covariates, including the average parental education and ethnic composition of enrolled students, were not included. The parental income and race covariates added do not completely encompass these variables, meaning that omitted variable bias may have affected our results. However, adding more covariates like these would increase the multicollinearity between these covariates and the independent variable, which can decrease the reliability of regression results (9). A future experiment using weighted linear regressions should thus aim to include as many uncorrelated covariates as possible to negate biases induced by both multicollinearity and omitted variables.

The nature of our data may also have limited the relevance of our results, as our study used data from students entering college in 2001. Older data was used because the median incomes of students from various colleges ten years after enrollment is not available for recent classes. As college degrees have become more prevalent, employers might be more inclined to use the selectivity of the institution attended to differentiate between college graduates (13). Thus, there is the potential that college selectivity has become more of a causal predictor of median income in recent years, which cannot be accounted for in our study. A future experiment repeating this study should attempt to use more recent data if it becomes available.

	Estimated variable coefficient	Std. Error	p-value
Intercept	13900.00	4045.50	6.32e-04***
Barron's Selectivity Score	150.67	331.53	0.650
Average SAT Score (points)	8.49	3.56	0.0175**
Median Household Income (USD/year)	0.18	0.01	< 2e-16***
Percent STEM Major	239.01	21.27	< 2e-16***
Percent Business Major	154.51	15.51	< 2e-16***
Percent Arts/Humanities Major	-134.89	28.77	3.41e-06***
Percent Health Major	177.28	24.05	5.67e-13***
Percent Social Science Major	62.61	15.26	4.68e-05***
Percent Public Social Major	49.14	23.06	0.0335**
Percent Asian/Pacific Islander	276.12	64.18	1.98e-05***
Percent African American	-53.35	27.17	0.0501**
Percent Non-White Hispanic	44.83	37.05	0.227
Percent White	-74.88	26.28	0.00453***
Percent Female	-20.48	17.16	0.233

Table 3: The effect of attending a selective college on future income with all controls. The variable coefficient represents the expected change in future expected income when that variable is increased by one unit. The intercept represents the projected income when all variables were zero. The Barron's Selectivity Score is the only independent variable of relevance to this study. The coefficients of the intercept and other covariates are not relevant to the study. Weighted linear regression, superscripts **, and *** indicate statistical significance at the 5% and 1% levels, respectively.

MATERIALS AND METHODS

The study was based on 2011 College Scorecard Income data, which recorded the median income of students from each college ten years after enrollment. Individuals who were unemployed or still in school were excluded from the dataset. These earnings were derived from the National

https://doi.org/10.59720/23-317

Barron's Selectivity Index	Number of Colleges present in Regression	Description	Example Schools
6	67	Most Selective: only admit a very small percentage of their applicants and usually require students to be in the top 10% of their high school class.	Yale University, Amherst University, University of Chicago
5	66	Highly Selective: may admit fewer than one quarter of their applicants and accept most of their students from the top 20% of their class.	Boston University, Grinnell College, University of Florida
4	134	Very Selective: may admit fewer than a third of their applicants, prefer students who rank in the top 35% of their class.	Michigan State University, Loyola Marymount University, New Jersey Institute of Technology
3	270	Selective: may admit fewer than half of their applicants, prefer students to be in the top 50% of their high school class.	CUNY Brooklyn College, Texas Tech University, Hastings College
2	56	Less Selective: generally admit around 85% of applicants who normally rank in the top two thirds of their high school class.	Salem State University, West Liberty University, Northwood University
1	14	Unselective: generally only require evidence of graduation from high school.	Gordon State College, Crown College, Spartanburg Methodist College

Table 4: Colleges in each Barron's Selectivity Tier used in the regression. Table depicting the number of colleges used in the regression in each selectivity tier, with representative example colleges in that tier as well as a description of the colleges within the tier.

Student Loan Data System (NSLDS) records of wages from all W-2 tax forms for each individual, plus self-employment income reported to the Internal Revenue Service (14). The NSLDS only tracked students receiving federal financial aid in the form of Title IV grants and loans. While only 72.7% of students received federal financial aid in 2001, research has shown that the College Scorecard earnings are representative of actual outcomes when compared with survey data from Post-Secondary Employment Outcomes (PSEO) below, which includes all students, regardless of aid status (15). The College Scorecard Income data could thus be used as a proxy for the true earnings of a student graduating from each college.

As the earnings data is for students in the class of 2001, the median SAT scores, racial demographics, and gender demographics were all taken from the 2001 Integrated Postsecondary Education Data System (IPEDS) dataset. As only the 25th and 75th percentiles of SAT scores were released, the median SAT scores were defined as the average of the 25th and 75th percentiles. The choice of major from the students entering college in 2000 was used because the IPEDS did not have data from 2001, as it only recorded this data every other year (8). Student major choice largely remained stable in the early 2000's (16). The percentage of students choosing a "multidisciplinary" major was not included in the regression, as this category was too broad to be useful as a covariate. The parental income data from Opportunity Insights was derived from the tax records of students in the 1983 birth cohort and was defined as the median sum of incomes from both parents (17). As the majority of students enter college when they are 18 years old, the data from the 1983 birth cohort corresponded to the cohort of students entering college in the fall of 2001.

Barron's Selectivity Tiers, which categorized individual schools in terms of selectivity using factors including testing scores and admission rates, were our chosen measure of selectivity (16). For the regression, numbers from 1-6 corresponding to the appropriate Barron's Selectivity tier (**Table 4**) were assigned to each college. Any colleges classified as "special" (usually highly specialized schools, like The Juilliard School), were excluded. Only colleges with data available for every variable used in our regression were included in the dataset. All regressions thus included 607 colleges. Most tiers had large sample sizes (N>50), but only 14 "unselective" schools were in the dataset due to most "unselective" schools lacking available IPEDS data (Table 4). The Barron's coefficient was not greatly impacted after "unselective" schools were removed, so "unselective" schools were included in our regressions to ensure all types of schools were represented. The 607 colleges studied were all based in the U.S. and included 463 private schools, 144 public schools, and all 8 Ivy-League universities.

The data were then merged into a data table within R, at which point any college without complete information for all characteristics was removed from the dataset. A simple linear regression was then performed, with income after graduation as the dependent variable and Barron's selectivity index as the independent variable. The standard deviation of that regression was used to calculate the weight necessary to perform a weighted linear regression. A weighted linear regression was then performed using the weight obtained. The process was repeated similarly but with every characteristic acting as an independent variable as opposed to Barron's selectivity alone. As this variation had multiple covariates, the

https://doi.org/10.59720/23-317

VIF test was performed to test for multicollinearity.

The linear regression was weighted because the variances of the observation errors were not homogenous. This violated the homoscedastic requirement and could therefore skew the results of an unweighted, simple linear regression alone, which necessitated the use of a weighted linear regression (18).

Received: November 25, 2023
Accepted: April 16, 2024
Published: October 16, 2024

REFERENCES

1. Judge-Stasiak, Angela. "Special admission: how college sports recruitment favors white, suburban athletes." *International Journal for Educational Integrity*, vol. 18, no. 1, Jan. 2022, https://doi.org/10.1007/s40979-021-00094-6. Accessed 10 Mar. 2024.
2. Kim, Ashley. "Legacy Admissions: An Insidious Form of Racial Discrimination." *Columbia Undergraduate Law Review*, Oct. 2022, culawreview.org/journal/legacy-admissions-an-insidious-form-of-racial-discrimination, Accessed 10 Mar. 2024.
3. Rios, Edwin. "What was affirmative action designed to do – and what has it achieved?" *The Guardian*. theguardian.com/law/2023/jun/22/what-is-affirmative-action-supreme-court-explainer. Accessed 10 Mar. 2024.
4. Rim, Christopher. "Is An Ivy League Degree Worth The Cost?" *Forbes*. forbes.com/sites/christopherrim/2023/09/19/is-an-ivy-league-degree-worth-the-cost. Accessed 10 Mar. 2024
5. Dale, Stacey and Krueger Alan. "Estimating the Effects of College Characteristics over the Career Using Administrative Earnings Data." *Journal of Human Resources*, vol. 49, no. 2m 2014, pp. 323–358. JSTOR, https://doi.org/10.3368/jhr.49.2.323. Accessed 26 Aug. 2023.
6. Chetty, Raj, et. al. "Diversifying Society's Leaders? The Determinants and Consequences of Admission to Highly Selective Colleges." *National Bureau of Economic Research,* nber.org/papers/w31492. Accessed 22 Aug. 2023. Working Paper.
7. *2009 Barron's Profiles of American Colleges 28th Edition*. Barron's Educational Series, 2008. Accessed 11 Mar. 2024.
8. *Integrated Postsecondary Education Data System*. U.S Department of Education, National Center for Education Statistics. nces.ed.gov/ipeds. Accessed 22 Aug. 2023.
9. Gareth, James, et. al. "Collinearity". *An Introduction to Statistical Learning: with Applications in R*. Springer Science & Business Media, 2013, pp. 106-109. Accessed 16 Mar. 2024.
10. Holzer, Harry, and David Neumark. "Assesing Affirmative Action." *Journal of Economic Literature*, vol. 38, no. 2, 2000, JSTOR, jstor.org/stable/2565419. Accessed 6 Jun. 2024.
11. Ge, Suqin et. Al. "Elite schools and opting in: effects of college selectivity on career and family outcomes." *National Bureau of Economic Research*. doi.org/10.3386/w25315. Accessed 16 Mar. 2023. Working Paper.
12. Wooldridge, Jeffrey. "Omitted Variable Bias: The Simple Case". *Introductory Econometrics: A Modern Approach*. Cengage Learning, 2009 pp. 89–93. Accessed 16 Mar. 2024.
13. Neitzel, Michael. "Percentage Of Adults With College Degrees Edges Higher, Finds New Lumina Report." *Forbes*. forbes.com/sites/michaeltnietzel/2024/02/01/percentage-of-us-adults-with-college-degrees-edges-higher-finds-lumina-report. Accessed 10 Mar. 2024.
14. *College Scorecard*. U.S. Department of Education. collegescorecard.ed.gov/data. Accessed 22 Aug. 2023.
15. "Technical Documentation: College Scorecard Data by Field of Study." *U.S. Department of Education.* collegescorecard.ed.gov/assets/FieldOfStudyDataDocumentation.pdf. Accessed 26 Aug. 2023.
16. "Bachelor's degrees conferred by postsecondary institutions, by field of study: Selected years, 1970-71 through 2019-20. *National Center for Education Statistics*. nces.ed.gov/programs/digest/d21/tables/dt21_322.10.asp. Accessed 16 Mar. 2024.
17. *Data Library*. Opportunity Insights. opportunityinsights.org/data. Accessed 22 Aug. 2023.
18. "Weighted Least Squares." *Penn State Eberly College of Science*. online.stat.psu.edu/stat501/lesson/13/13.1. Accessed 11 Mar. 2024.