# Navigating the Crossroads: Understanding Major Choices of First-Generation Students*

Lillian Zhu

Distinguished Majors Program
Department of Economics
University of Virginia
Advisor: Sarah Turner

April 2025

**Abstract**

The choice of college major has profound implications for career opportunities and social mobility, yet first-generation students remain underrepresented in critical fields like humanities and non-healthcare STEM. I estimate logit models to examine how first-generation status relates to major selection, with a focus on the underrepresentation of first-generation students in humanities and non-healthcare STEM fields. I find that first-generation students remain significantly less likely to major in humanities fields, even after controlling for academic and personal factors, while high school GPA is a significant predictor of majoring in a non-healthcare STEM field. This indicates that institutional factors and informational gaps, which are not controlled for, likely influence the decision to major in humanities for first-generation students, whereas academic preparation may be a key factor in accessing STEM fields, regardless of first-generation status. In addition, trends from 2000 to 2024 show that the college-going population has become increasingly diverse, particularly in terms of gender, race, and first-generation status. Together, these results highlight the need for a deeper understanding of the factors and barriers that shape first-generation students' academic paths.

*I would like to thank my advisor, Professor Sarah Turner, for her support and thoughtful comments throughout this process. Thank you to the DMP cohort for the many laughs and words of encouragement. I would also like to thank my closest friends for keeping me sane, and Joon for being there through it all and for the LaTeX help. To my parents and brothers, thank you for your unwavering support. Without you, this would not have been possible.

# Contents

# 1 Introduction

College major choice plays a significant role in future earnings and career paths (Altonji et al., 2016). Yet, disparities persist in the fields that students pursue. First-generation (FG) students — students where neither parent completed a bachelor's degree — are disproportionately underrepresented in fields such as non-healthcare STEM and humanities. These students face unique challenges that shape their academic experience, and later, their career outcomes. These challenges include limited access to advising and academic assistance, networking opportunities, and financial support. Understanding the barriers that impact these major choices is critical to informing higher education policies.

In the United States, about 38% of adults over the age of 25 hold at least a bachelor's degree (U.S. Census Bureau, 2022). Yet, these rates vary across demographic groups: about 53% of these adults are women, whereas about 47% are men. About 42% of White adults and about 61% of Asian adults hold at least a bachelor's degree, compared to about 28% of Black adults and about 21% of Hispanic adults.

Among those who have completed a bachelor's degree, about 70% have parents who have done the same, compared to about 26% of adults whose parents do not hold college degrees (or first-generation students) (Fry, 2021). Underrepresented minority students are more commonly also first-generation students, as shown in Table A1 in the Appendix. In particular, as shown below in Table 1, FG students are underrepresented in non-healthcare STEM and humanities fields. That is, a smaller percentage of FG students major in STEM (besides healthcare) and humanities fields than non-FG, or continuing-generation, students. A more detailed breakdown by field of study can be found in the Appendix (see Table A2).

These patterns, drawn from the National Postsecondary Student Aid Study 2020 (NPSAS:20) using the National Center for Education Statistics (NCES) DataLab, show that even after gaining access to college, FG students may choose different academic pathways than their continuing-generation peers (National Center for Education Statistics, 2020). This motivates the question of why these students choose differently, particularly in fields where they are

Table 1: Percent Distribution of Humanities, STEM (Non-Healthcare), and Other Majors by First-Generation Status.

| **Category** | **Humanities** (%) | **STEM (Non-Healthcare)** (%) | **Other** (%) |
|---|---|---|---|
| Total | 5.6 | 22 | 72.4 |
| First-Gen | 4.3 | 17.8 | 77.9 |
| Non-First-Gen | 7.1 | 27 | 65.9 |

Source: Author's tabulations from NPSAS:20, using the NCES DataLab. Condensed version of Table A2.

underrepresented. Many possible factors can explain these differences, including financial constraints, academic preparation, and type of institution attended.

Financial constraints may influence both whether an individual attends college and if they do, how they navigate college, especially in choosing a major that offers reasonable post-graduate earnings. First-generation students may face several financial constraints, such as lack of parental or other financial support, which may necessitate choosing a major that yields higher post-graduate compensation. Differences in high school academic preparation may also lead to different major choices. For instance, weaker math preparation may deter or prevent students from choosing a STEM major. In addition, liberal arts institutions, which more frequently emphasize humanities, tend to be located in more rural areas, and may be less accessible to FG students who often attend schools closer to home (as shown in Table A3).

This paper explores several possible factors for this gap in major selection. Using data from the National Longitudinal Survey of Youth 1997 (NLSY97), I analyze if the difference in major choice between first-generation and continuing-generation students is driven by financial constraints or academic preparation; specifically, I examine the underrepresentation in humanities and STEM fields (Bureau of Labor Statistics, U.S. Department of Labor, 2024). Then, I use the Current Population Survey (CPS) to document the changing structure of students in college from 2000 to 2024 (Flood et al., 2024). These trends provide important context for understanding the broader shifts of first-generation students over time. I further supplement these analyses with descriptive tables based on data from the NPSAS:20.

The rest of the paper proceeds as follows. Section 2 provides an overview of the existing literature on first-generation students and the factors influencing the academic decisions they make. Section 3 presents the two main data sources used, the NLSY97 and the CPS, as well as the construction of key variables. Section 4 explains the methodology, including the regression models used. Section 5 presents the key results and discussion. Section 6 concludes by outlining the key takeaways from the analysis and providing possible direction for future research.

# 2 Literature Review

A growing body of research has explored the challenges, structures, and outcomes of first-generation students. This research is often centered around the expectations and personal and academic experiences of FG students as compared to non-FG students.

Terenzini et al. (1996) emphasize that not only are first-generation students' educational experiences worse than their continuing-generation peers, but also their cultural and personal experiences. For instance, they tend to work more hours outside of class, are more likely to report experiencing racial or gender discrimination, and have lower degree aspirations. Stephens et al. (2012) propose a cultural mismatch theory that suggests that the independence norms promoted by universities disadvantage FG college students because of gaps in social norms between their mostly working-class backgrounds and the middle-class norms popular in universities.

These broader challenges can have important implications for the college experiences of first-generation students. In the rest of this section, I review the factors that shape the major choices in FG students, specifically focusing on why they are underrepresented in non-healthcare STEM and humanities fields.

First-generation students are more likely to enter college with less academic preparation than their continuing-generation peers. The literature suggests that FG students are less

likely to take AP classes or higher-level math classes in high school (Ives & Castillo-Montoya, 2020). This is also consistent with the findings in Tables A4 & A5, which show that FG students tend to not take any college credits (AP, IB, etc.) in high school and tend to complete a lower level of math.

Academic preparation plays a key role in major choice. Arcidiacono (2004) finds that math preparation (i.e., SAT math scores) is important for both labor market outcomes and major choices. Specifically, those with a lower math score will tend to choose a major that requires less math than the natural sciences, such as business or humanities. However, business is still more lucrative than humanities subjects. Thus, perhaps these relatively underprepared students may choose business over a humanities degree but do not choose natural sciences (or another STEM field) due to a lack of preparation. This is consistent with descriptive patterns from NPSAS:20 (see Table A2), which shows that FG students are more likely to major in business and healthcare fields and less likely to choose humanities or non-healthcare STEM.

Another factor is that FG and non-FG students differ in the types of institutions they attend (See Table A6). Specifically, FG students are more likely to attend two-year institutions and for-profit institutions. In addition, they are less likely to attend private non-profit four-year institutions. The selection of major options differs by type of institution, for instance, a large public university probably offers less humanities options compared to a smaller liberal arts college.

Moreover, over the last few decades, schools in the US have produced increasingly fewer humanities graduates than in other fields. However, liberal arts colleges have continued to produce more humanities graduates than graduates in other fields (Hearn & Belasco, 2015). As shown in Table A6, there is a smaller percentage of FG students at private nonprofit four-year institutions, which tend to be liberal arts colleges, compared to non-first-generation students. Thus, this underrepresentation may be a reason why fewer first-generation students choose a humanities major. In addition, FG students tend to choose schools closer to home,

as shown in Table A3. Since liberal arts institutions tend to be in relatively rural areas, this may also help explain why FG students are underrepresented in humanities fields. Similarly, two-year institutions tend to offer less natural science programs (D'Amico et al., 2019). This means that students in these programs may have less access to opportunities such as research labs and faculty mentorship.

First-generation students may also have different preferences, which may come from varying family circumstances. As shown in Table A7, FG students are more likely to be low-income. Lower income is correlated with college attendance, and Carneiro and Heckman (2002) note that this correlation is likely due to short run credit constraints or long run family effects. Since FG students may face financial constraints, such as coming from a low-income background or needing to pay off any accrued student debt, this may drive them towards a different field. Moreover, FG students may have less access to information about career paths, and this may push them to pursue different majors.

Humanities fields also tend to pay less after college (Arcidiacono, 2004). STEM fields, while they may pay more, can be riskier to pursue if a student lacks the proper academic preparation. So, students may be pushed towards pursuing something "safer", such as business or healthcare. Thus, different financial constraints, coming from both family background and potential future earnings, could influence one's major choices.

Access to information may also differ. Tate et al. (2015) and Toyokawa and Dewald (2020) both find through surveying FG students that FG students tend to lack support and a network that can help with professional development. Without support and guidance from those familiar with career outcomes for certain majors, first-generation students may avoid majors that seem abstract, such as history or philosophy, even if they are interested in those fields.

# 3 Data

Two primary data sources are used in this paper: the National Longitudinal Survey of Youth 1997 (NLSY97) and the Current Population Survey (CPS). The NLSY97 is used to provide an in-depth analysis of how students' academic and personal backgrounds relate to their major choice. The CPS data allows for a broader view of the changing structure and composition of college students over time.

In addition to the NLSY97 and the CPS, I generate tabulations from NPSAS:20 using the NCES PowerStats DataLab Tool for descriptive tables. These provide valuable insights on national-level differences by FG status. In this paper, first-generation status is defined by whether at least one parent completed a bachelor's degree.

## 3.1 NLSY97

The NLYS97 tracks a nationally representative cohort of 8,984 people located in the US in 1997 who were born between 1980 and 1984 (Bureau of Labor Statistics, U.S. Department of Labor, 2024). The participants were all between the ages of 12 and 16 as of December 31, 1996. Extensive information is collected on the participants' educational experiences. The survey tracks the progress of participants over time through their high school, college, and early career experiences.

For this analysis, I use all respondents in the sample who enrolled in college at some point. Thus, these respondents were likely enrolled in college from 1998-2007, if at all. The sample is further restricted to those who declared at least one field of study for any of the institutions recorded.

First-generation status is constructed from parental information. A respondent is coded as FG if both residential parents did not complete a bachelor's degree. The binary variable for humanities major was classified as whether the respondent listed a humanities field as the first field of study for any institution they attended. I defined humanities major based on

the NCES 2010 College Course Map course codes, including subjects such as gender studies, communication, literature, philosophy, arts, and history (Bryan & Simone, 2012). A respondent is defined as a humanities major if their field of study corresponds with a humanities major and if they earned a bachelor's degree at that corresponding institution. STEM majors are similarly defined. Categories included relate to natural sciences, mathematics, and engineering fields. Notably, healthcare related fields, such as nursing, are excluded here.

The survey also contains information on fields of study, which was collected as part of the post-secondary transcript data. So, I use the 2011 sampling weight since this data was collected in 2012-2013, which means it reflects transcript data up to 2011. To best reflect family income around the time of being in college, I use the gross family income in 1999 since those participating in the study would have been around college age. Other variables include binary variables for female, Black, and Hispanic. I also use overall high school GPA, the log of 1999 family income, and a binary variable for whether or not the respondent's ACT score was greater than 24, to measure academic preparation. Because the ACT is scored on a scale of 1-36, and small differences in score may not reflect much of a difference in academic preparation, I create a binary indicator equal to 1 if the student scored higher than a 24.

## 3.2 CPS

Cross-sectional data from the Current Population Survey (CPS) Annual Social and Economic Supplement (ASEC), accessed through IPUMS, is used to examine changes in the composition of students attending college over time (Flood et al., 2024). The CPS is a monthly US household survey conducted by the US Census Bureau to over 65,000 households since the 1940s. The survey gathers information on demographics, education, family structure, and other characteristics for individuals and households. The CPS ASEC contains the same basic monthly demographic information as the main survey, but provides additional data on work experience, income, etc. Even though the CPS has no information on major choice, it still gives useful context for changes in the population of students entering college. For

this paper, the sample was restricted to individuals between ages 18 and 24 and currently enrolled in college either part-time or full-time. I use the ASEC data from 2000 to 2024.

IPUMS allows variables for mother's education level and father's education level to be added when downloading. First-generation status is defined using these. The respondent is flagged as first-generation if neither parent completed a bachelor's degree. However, this is limited to individuals whose parents are still in the household. Thus, those who are no longer in the same household as their parents cannot be represented in the sample.

Additional variables include race and gender. These variables are used to document the changes in demographics over time, both for the overall population of college students and separately by FG status. All estimates use the Annual Social and Economic Supplement household weight to produce representative results.

### 3.3 NPSAS:20

Descriptive tabulations from NPSAS:20 using the NCES PowerStats DataLab tool are also used. Since 1987, NPSAS has collected nationally representative data every 3 to 4 years on undergraduate and graduate students enrolled in post-secondary education. The study focuses on the characteristics of students, with an emphasis on how students and their families pay for their education.

Tables summarizing a variety of educational factors, such as institutional control, test scores, and high school academic preparation, by first-generation status can be found in the Appendix. These are used in mainly Sections 1 and 2 to provide background and motivation for studying the major choices of first-generation students.

## 4 Methodology

The NLSY97 is used to analyze differences between FG and non-FG students' major choices, specifically the gap in choosing a humanities major. The CPS is used for descriptive analyses

to examine how the structure of college students has changed over time.

## 4.1 NLSY97

The main analysis using the NLSY97 is done by constructing two logit regression models, since the dependent variable for both is binary: whether the respondent majored in a humanities field (1) or not (0), and similarly, whether the respondent majored in a non-healthcare STEM field.

The first model estimates the probability that a given respondent majored in a humanities field conditional on first-generation status and a set of control variables. The specification for the first model is as follows:

$$P(\text{Humanities}_i = 1|\text{FGStatus}_i, \mathbf{X}_i) = \frac{1}{1 + exp(-(\beta_0 + \beta_1 \text{FGStatus}_i + \mathbf{X}_i\boldsymbol{\beta} + \epsilon_i))}, \quad (1)$$

where $\text{Humanities}_i$ is a binary variable that indicates whether individual $i$ majored in a humanities field, $\text{FGStatus}_i$ is a binary variable representing the first generation status of individual $i$, and $\mathbf{X}_i$ is a vector of covariates, including gender, race/ethnicity, high school GPA, log of household income, and ACT score. Since the NLSY97 does not contain much on institutional factors, such as variety of majors offered or student support, it is not controlled for in the model.

I also estimate the likelihood that a given respondent majored in a STEM field conditional on FG status and a set of control variables. The model is as follows:

$$P(\text{STEM}_i = 1|\text{FGStatus}_i, \mathbf{X}_i) = \frac{1}{1 + exp(-(\beta_0 + \beta_1 \text{FGStatus}_i + \mathbf{X}_i\boldsymbol{\beta} + \epsilon_i))}, \quad (2)$$

where $\text{STEM}_i$ is a binary variable that indicates whether individual $i$ majored in a humanities field, $\text{FGStatus}_i$ is a binary variable representing if individual $i$ is a first-generation student, and $\mathbf{X}_i$ is a vector of covariates, including gender, race/ethnicity, high school GPA, log of household income, and ACT score.

### 4.2 CPS

While the main analysis relies on the NLSY97, descriptive analysis using the CPS provides context for how the composition of students attending college has changed over time. This analysis focuses on trends from 2000-2024 in order to capture changes in college attendance.

These figures are not used to estimate any causal effects, instead, they give important context on the changing structure of students entering college. Using the 2000 through 2024 CPS ASEC data, I examine how the demographic composition of college students has shifted.

I further replicate these figures conditional on first-generation status. This allows for an exploration of how the characteristics of first-generation college students have changed relative to their continuing-generation peers.

## 5 Results & Discussion

Table 2 below shows the results from the logit regression estimating the likelihood of majoring in a humanities field. The independent variable of interest is first-generation status, and the controls are gender, race/ethnicity, high school GPA, family income logged (i.e., household income while in college), and whether the individual's highest ACT score was greater than 24.

Table 2: Logit Regression: Probability of Majoring in Humanities

| **Variable** | **Coefficient** | **Linearized Std. Error** | **p-value** |
|---|---|---|---|
| First-generation | -1.061 | 0.351 | 0.003 |
| Female | -0.156 | 0.404 | 0.701 |
| Black | -0.279 | 0.484 | 0.566 |
| Hispanic | -0.497 | 0.849 | 0.560 |
| High School GPA | 0.269 | 0.331 | 0.417 |
| Log Family Income | -0.036 | 0.161 | 0.824 |
| ACT $>$ 24 | -0.277 | 0.314 | 0.374 |
| Constant | -1.883 | 1.982 | 0.345 |

*Notes:* Sample size: 795. Other sample restrictions include limiting to BA recipients only and excluding those who did not declare any major. For other variables, if any observations were missing, those were not included as well. Data used is from the NLSY97, as detailed in Section 3.

The results indicate that FG students are significantly less likely to major in a humanities field than non-FG students. Specifically, for a first-generation student, the log-odds of selecting a humanities major is 1.06 lower than for a continuing-generation student. As seen in Table A8, the average marginal effect of FG status is -0.096, which means that FG students are almost 10% less likely than other students to major in a humanities field, on average, holding other variables constant.

None of the other covariates in the model are statistically significant. Surprisingly, the coefficient on high school GPA has a small positive relationship, though this is not statistically significant and has a large standard error. The other covariates all had small, negative, insignificant relationships with choosing a humanities major. While these insignificant relationships could reflect some correlation between these variables and being a humanities major, it is more likely that there is some measurement error or lack of statistical power in these variables. For many observations, ACT score and/or high school GPA were missing, and thus excluded from the regression, which likely contributed to this problem.

These results provide evidence that first-generation college students are less likely to major in a humanities field, even after controlling for factors like academic preparation (high school GPA and ACT score), family income, and demographic characteristics. Since those factors are not significant, this suggests that FG students may be motivated by something else when choosing a major. For instance, the lower salary generally earned by humanities majors may steer them towards something different. Gaps in advising, knowledge about different majors, or professional networks may cause FG students to see humanities majors as a less viable choice.

Although institutional characteristics are not reflected directly in the NLSY97, Table A6 in the Appendix shows that first-generation students are more likely to attend public two year schools and less likely to attend public and private non-profit four-year schools, including liberal arts schools. These two-year institutions often prioritize workplace and career skills over a liberal arts education, and this may offer less options or encouragement

for studying humanities (D'Amico et al., 2019). This suggests that part of the gap in majoring in humanities among FG students may be due to the types of institutions they choose to attend.

Table 3 shows the results from the logit regression estimating the likelihood of majoring in a STEM field. Similar to before, the independent variable of interest is first-generation status, and the controls are gender, race/ethnicity, high school GPA, family income logged (i.e., 1999 household income), and ACT score.

Table 3: Logit Regression: Probability of Majoring in STEM (Non-Healthcare)

| **Variable** | **Coefficient** | **Linearized Std. Error** | **p-value** |
|---|---|---|---|
| First-generation | -0.544 | 0.509 | 0.289 |
| Female | -0.681 | 0.468 | 0.150 |
| Black | 0.443 | 0.696 | 0.526 |
| Hispanic | -0.632 | 0.769 | 0.414 |
| High School GPA | 2.733 | 0.508 | 0.000 |
| Log Family Income | -0.073 | 0.284 | 0.797 |
| ACT $>$ 24 | 0.580 | 0.560 | 0.304 |
| Constant | -10.483 | 3.101 | 0.001 |

*Notes:* Sample size: 757. Other sample restrictions: limited to BA recipients and those who declared a major. For other variables, if any observations were missing, those were not included as well. Data used is from the NLSY97, as detailed in Section 3.

I find that the coefficient for FG status is insignificant, even after controlling for academic and personal background characteristics. Although it is not statistically significant, the negative coefficient (-0.544) suggests a potential underrepresentation of FG students due to FG status. However, the large standard error implies that there is great uncertainty around this estimate. On the other hand, the variable for high school GPA is significant. This suggests that academic preparation may be a more important factor in whether someone chooses a STEM major. The positive coefficient of 2.733 implies that a one unit increase in high school GPA results in a 2.733 increase in the log-odds of majoring in a non-healthcare STEM field. In other words, as shown in Table A9, the average marginal effect of overall high school GPA is 0.2526, which means that a one-point increase in GPA, such as from a 3.0 to a 4.0, results in a 25.26 percentage point increase in the probability of majoring in a non-

healthcare STEM field. The other control variables, race, ethnicity, gender, family income, and ACT score are also insignificant, suggesting that after accounting for high school GPA, these factors do not have an individually significant correlation with the probability of being a STEM major. This highlights the role that academic preparation plays in determining major choices, especially in a STEM field.

While there are valuable insights from these results, there are many limitations to the available dataset. The NLSY97 lacks many institutional characteristics, such as types of support or advising available or used by students. In addition, missingness in the data resulted in many observations being dropped. This likely caused a reduction in statistical power, resulting in less precise estimates.

Although the NLSY97 regressions show evidence of a gap in humanities majors by FG status, and potential underrepresentation for both humanities and STEM majors, this cohort is from a relatively short period of time. To understand how the structure of the population of students entering college has changed over time, I use CPS data from 2000 to 2024, which allows for a broader and more recent view of trends by different demographic groups.

Figure A1 shows the changes in the proportion of first-generation students. The share has mainly decreased over time, but this is likely a reflection of increases in educational attainment among adults in the US. Thus, first-generation students now represent a smaller share of college students in the US.

The proportion of female students has increased over time, as seen in Figure A2. This is consistent with the fact that women have seen growing educational attainment over the last few decades. To further understand whether these trends differ by first-generation status, Figures A3 and A4 present trends for first-generation students and continuing-generation students, respectively. Among first-generation students, the proportion of female students seems to have grown to a bigger share than among non-FG students.

The racial composition of students has also changed. In Figure A5, the share of White students has declined, while the share for the other groups has all increased. Figures A6

and A7 show the trends separately for FG and non-FG students, respectively. These two figures seem largely the same, indicating that trends have been similar among both FG versus non-FG students.

# 6 Conclusion

First-generation students remain less likely to major in humanities fields, even after controlling for academic preparation, test scores, family income, and demographic characteristics. Using data from the NLSY97, I find that this gap for FG students majoring in humanities is statistically significant. This suggests that the gap may be due to other barriers that were not controlled for such as institutional factors and differences in support and advising.

In contrast, for non-healthcare STEM majors, after controlling for academic preparation, FG status was not a significant predictor of the decision to major in a STEM field. This emphasizes the important role academic preparation plays in choosing a STEM major.

Finally, descriptive results from CPS data show that over time, the students in college have become increasingly racially diverse. These shifts suggest that the population of FG students is evolving, and so may the barriers they face.

Future research might expand this work by incorporating institutional characteristics or exploring post-graduation outcomes. Differences in advising, major options, and other institutional factors may play a role in first-generation students' major choices. While this paper highlights a gap in majoring in humanities, it suggests that the gap cannot be fully explained by academic preparation, family income, and demographic differences. On the other hand, I find that majoring in non-healthcare STEM is shaped by academic preparation, in particular, high school GPA. Understanding what determines major selection, both for lower paying fields like humanities and for more quantitative fields such as STEM, is crucial to informing policies that can support more equitable access to different fields of study.

# References


Altonji, J. G., Arcidiacono, P., & Maurel, A. (2016). The analysis of field choice in college and graduate school: Determinants and wage effects. In *Handbook of the economics of education* (pp. 305–396, Vol. 5). Elsevier.

Arcidiacono, P. (2004). Ability sorting and the returns to college major. *Journal of Econometrics*, *121*(1–2), 343–375.

Bryan, M., & Simone, S. (2012). 2010 college course map [National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education. Washington, DC]. http://nces.ed.gov/pubsearch

Bureau of Labor Statistics, U.S. Department of Labor. (2024). National longitudinal survey of youth 1997 cohort, 1997-2021 (rounds 1-20) [Data set. Produced and distributed by the Center for Human Resource Research (CHRR), The Ohio State University.]. https://www.nlsinfo.org

Carneiro, P., & Heckman, J. J. (2002). The evidence on credit constraints in post-secondary schooling. *The Economic Journal*, *112*(482), 705–734.

D'Amico, M. M., Sublett, C. M., & Bartlett II, J. E. (2019). Preparing the workforce in today's community colleges: Issues and implications for higher education leaders [Report].

Flood, S., King, M., Rodgers, R., Ruggles, S., Warren, J. R., Backman, D., Chen, A., Cooper, G., Richards, S., Schouweiler, M., & Westberry, M. (2024). Ipums cps: Version 12.0 [dataset] [https://doi.org/10.18128/D030.V12.0]. https://doi.org/10.18128/D030.V12.0

Fry, R. (2021). First-generation college graduates lag behind their peers on key economic outcomes [Pew Research Center]. Retrieved April 21, 2025, from https://www.pewresearch.org/social-trends/2021/05/18/first-generation-college-graduates-lag-behind-their-peers-on-key-economic-outcomes/

Hearn, J. C., & Belasco, A. S. (2015). Commitment to the core: A longitudinal analysis of humanities degree production in four-year colleges. *The Journal of Higher Education*, *86*(3), 387–416.

Ives, J., & Castillo-Montoya, M. (2020). First-generation college students as academic learners: A systematic review. *Review of Educational Research*, *90*(2), 139–178.

National Center for Education Statistics. (2020). Powerstats [U.S. Department of Education]. https://nces.ed.gov/datalab/

Stephens, N. M., Fryberg, S. A., Markus, H. R., Johnson, C. S., & Covarrubias, R. (2012). Unseen disadvantage: How american universities' focus on independence undermines the academic performance of first-generation college students. *Journal of Personality and Social Psychology*, *102*(6), 1178–1197.

Tate, K. A., Caperton, W., Kaiser, D., Pruitt, N. T., White, H., & Hall, E. (2015). An exploration of first-generation college students' career development beliefs and experiences. *Journal of Career Development*, *42*(4), 294–310.

Terenzini, P. T., Springer, L., Yaeger, P. M., Pascarella, E. T., & Nora, A. (1996). First-generation college students: Characteristics, experiences, and cognitive development. *Research in Higher Education*, *37*(1), 1–22.

Toyokawa, T., & Dewald, C. (2020). Perceived career barriers and career decidedness of first-generation college students. *The Career Development Quarterly*, *68*(4), 332–347. https://doi.org/10.1002/cdq.12240

U.S. Census Bureau. (2022, February 24). U.s. population more educated than a decade ago. Retrieved April 21, 2025, from https://www.census.gov/newsroom/press-releases/2022/educational-attainment.html

# A Appendix

## A.1 Tables

Table A1: Race/Ethnicity of Undergraduate Students by First-Generation Status

| Race/Ethnicity | White | Black | Hispanic/Latino | Asian | Native American | Pacific Islander | More than one |
|---|---|---|---|---|---|---|---|
| Total | 48.1 | 13.0 | 21.1 | 8.0 | 1.0 | 0.7 | 8.0 |
| First-Gen | 40.4 | 15.7 | 28.6 | 6.6 | 1.3 | 0.7 | 6.8 |
| Non-First-Gen | 57.2 | 10.0 | 12.4 | 9.7 | 0.7 | 0.6 | 9.5 |

Source: Author's tabulations from NPSAS:20, using the NCES DataLab

Table A2: Fields of Study of Undergraduate Students by First-Generation Status

| **Major** | **Undecided** (%) | **CIS** (%) | **Eng.** (%) | **Bio & Phys Sci** (%) | **Gen. Studies** (%) | **Soc. Sci.** (%) | **Humanities** (%) | **Health** (%) | **Business** (%) | **Education** (%) | **Other** (%) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Total | 1.0 | 5.8 | 6.4 | 9.8 | 7.6 | 9.0 | 5.6 | 18.6 | 14.6 | 5.2 | 16.5 |
| First-Gen | 0.8 | 5.2 | 4.8 | 7.8 | 8.2 | 8.1 | 4.3 | 22.7 | 14.9 | 5.7 | 17.5 |
| Non-First-Gen | 1.2 | 6.6 | 8.2 | 12.2 | 7.0 | 10.0 | 7.1 | 13.7 | 14.3 | 4.6 | 15.2 |

Source: Author's tabulations from NPSAS:20, using the NCES DataLab

Table A3: Distance from student's home (in miles) to NPSAS institution

| **Distance (in miles)** | **1 ≤ X ≤ 30 (%)** | **31 ≤ X ≤ 100 (%)** | **101 ≤ X ≤ 250 (%)** | **251 ≤ X ≤ 600 (%)** | **601 ≤ X ≤ 3000 (%)** | **3001 ≤ X ≤ 11344 (%)** |
|---|---|---|---|---|---|---|
| Total | 58.7 | 15.3 | 10.3 | 5.9 | 9.0 | 0.8 |
| First-Gen | 64.7 | 14.2 | 7.8 | 4.7 | 8.1 | 0.5 |
| Non-First-Gen | 51.5 | 16.7 | 13.2 | 7.4 | 10.1 | 1.1 |

Source: Author's tabulations from NPSAS:20, using the NCES DataLab

Table A4: Percentage of students who took college credits in high school, by First-Generation Status.

| **Took any college credits in high school** | **No (%)** | **Yes (%)** |
|---|---|---|
| Total | 34.4 | 65.6 |
| FG | 41.1 | 58.9 |
| Non-FG | 27.8 | 72.2 |

Source: Author's tabulations from NPSAS:20, using the NCES DataLab

Table A5: Highest level of math completed or planned by students based on FG Status

| **Level** | **None of these** (%) | **Alg. 1** (%) | **Geometry** (%) | **Alg. 2** (%) | **Trig.** (%) | **Precalculus/Prob. & Stats** (%) | **Calculus/AP Stats** (%) |
|---|---|---|---|---|---|---|---|
| **Total** | 1.2 | 1.3 | 2.4 | 15.7 | 6.6 | 30.6 | 42.3 |
| FG | 1.4 | 1.9 | 3.1 | 20.8 | 8.0 | 30.9 | 34.0 |
| Non-FG | 1.0 | 0.8 | 1.8 | 11.6 | 5.6 | 30.3 | 49.0 |

Source: Author's tabulations from NPSAS:20, using the NCES DataLab

Table A6: Distribution of students by type of institution attended by First-Generation Status.

| **Institution type** | **Public 4-yr** (%) | **Public 2-yr** (%) | **Public < 2-yr** (%) | **Private nonprofit 4-yr** (%) | **Private nonprofit ≤ 2-yr** (%) | **Private for profit 4-yr** (%) | **Private for profit ≤ 2-yr** (%) |
|---|---|---|---|---|---|---|---|
| Total | 41.9 | 32.9 | 0.4 | 16.7 | 0.3 | 4.9 | 2.9 |
| FG | 36.9 | 38.7 | 0.4 | 12.7 | 0.4 | 6.6 | 4.3 |
| Non-FG | 47.8 | 26.0 | 0.3 | 21.4 | 0.2 | 3.0 | 1.3 |

Source: Author's tabulations from NPSAS:20, using the NCES DataLab. This data uses IPEDS to define institution type.

Table A7: Percentage Distribution of Adjusted Gross Income (AGI) by First-Generation Status.

| **AGI ($)** | **$0 \leq X \leq 20000$** (%) | **$20001 \leq X \leq 40000$** (%) | **$40001 \leq X \leq 60000$** (%) | **$60001 \leq X \leq 90000$** (%) | **$90001 \leq X \leq 120000$** (%) | **$120001 \leq X \leq 250000$** (%) | **$250001 \leq X \leq 1000000$** (%) | **$1000001 \leq X \leq 2500000$** (%) |
|---|---|---|---|---|---|---|---|---|
| FG | 27.9 | 29.0 | 15.6 | 13.5 | 7.2 | 6.1 | 0.6 | 0.0 |
| Non-FG | 17.3 | 16.0 | 11.3 | 13.0 | 11.9 | 24.1 | 6.3 | 0.3 |

Source: Author's tabulations from NPSAS:20, using the NCES DataLab

Table A8: Average Marginal Effects on Probability of Majoring in Humanities

| **Variable** | **AME (dy/dx)** | **Std. Err.** | **p-value** |
|---|---|---|---|
| First-generation | -0.0964 | 0.0344 | 0.006 |
| Female | -0.0141 | 0.0370 | 0.704 |
| Black | -0.0251 | 0.0453 | 0.578 |
| Hispanic | -0.0452 | 0.0770 | 0.559 |
| High School GPA | 0.0245 | 0.0305 | 0.426 |
| Log Family Income | -0.0084 | 0.0146 | 0.824 |
| ACT $>$ 24 | -0.0252 | 0.0287 | 0.384 |

Source: Data on college students from the NLSY97. The AMEs represent the average change in the probability of majoring in a humanities field, calculated from the corresponding logit model (Table 2) in the text.

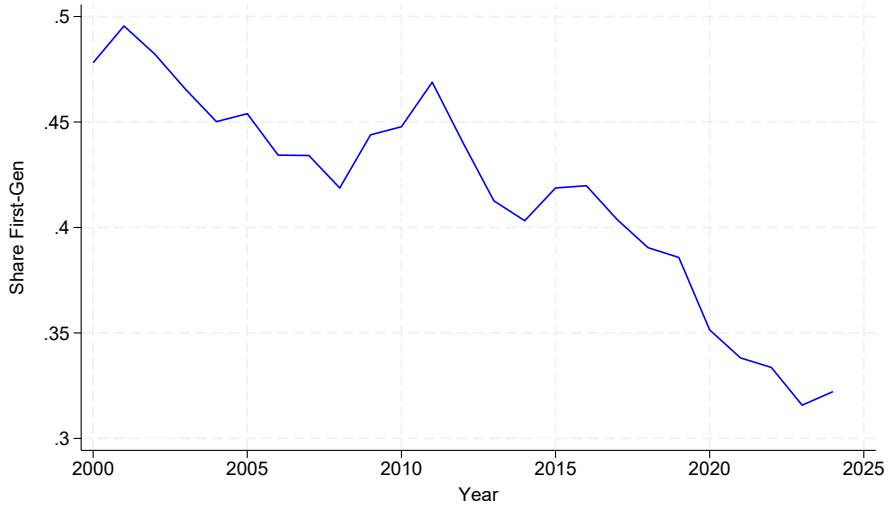
Table A9: Average Marginal Effects on Probability of Majoring in Non-Healthcare STEM

| **Variable** | **AME (dy/dx)** | **Std. Err.** | **p-value** |
|---|---|---|---|
| First-generation | -0.0503 | 0.0468 | 0.286 |
| Female | -0.0630 | 0.0445 | 0.161 |
| Black | 0.0410 | 0.0635 | 0.521 |
| Hispanic | -0.0584 | 0.0719 | 0.420 |
| High School GPA | 0.2526 | 0.0453 | 0.000 |
| Log Family Income | -0.0068 | 0.0260 | 0.795 |
| ACT $>$ 24 | 0.0536 | 0.0486 | 0.274 |

Source: Data on college students from the NLSY97. AMEs represent the average change in the probability of majoring in a non-healthcare STEM field, calculated from the corresponding logit model (Table 3) in the text.

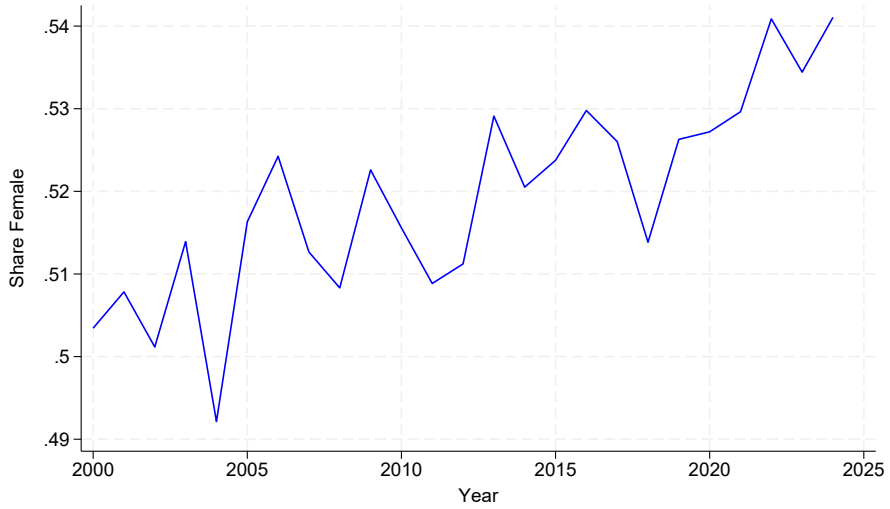
## A.2 Figures

Figure A1: Changes in the Proportion of First-Generation Students Over Time



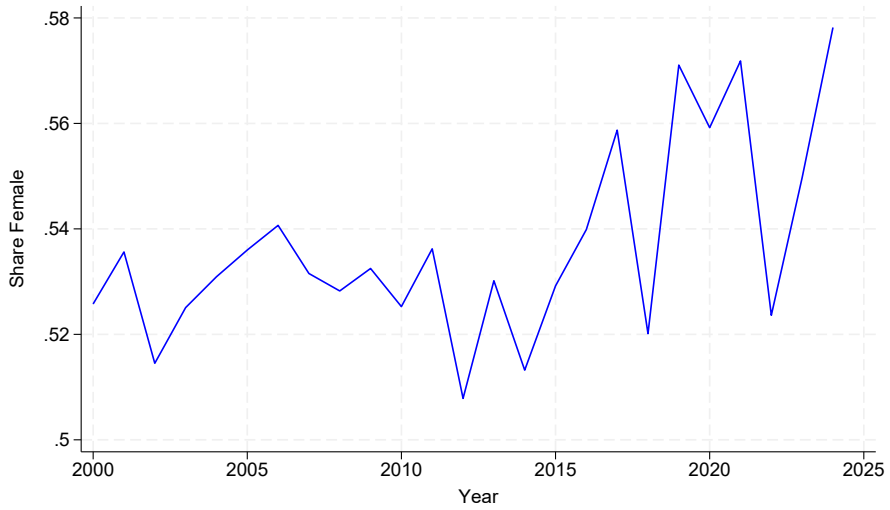

Source: Data on college students ages 18-24 during the period 2000-2024 from the CPS.

Figure A2: Changes in the Proportion of Female Students Over Time



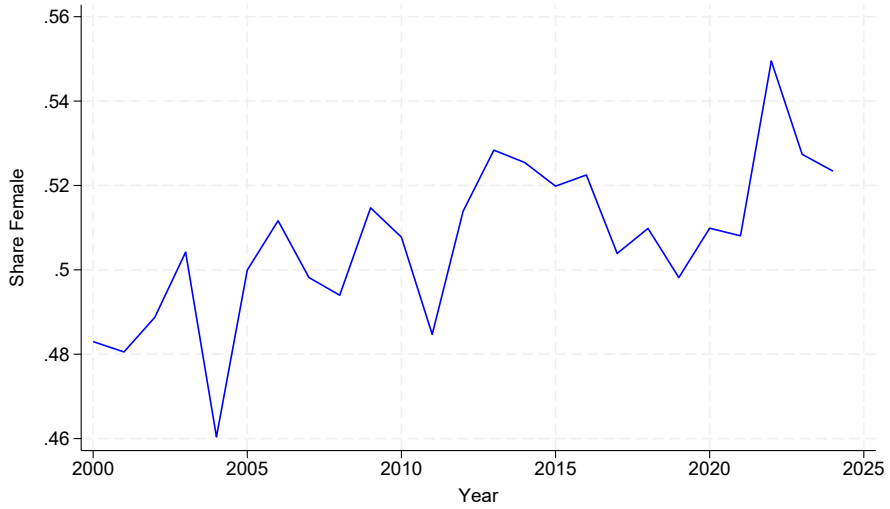

Source: Data on first-generation college students ages 18-24 during the period 2000-2024 from the CPS. First-generation status is calculated as described in the text.

Figure A3: Changes in the Proportion of First-Generation Female Students Over Time



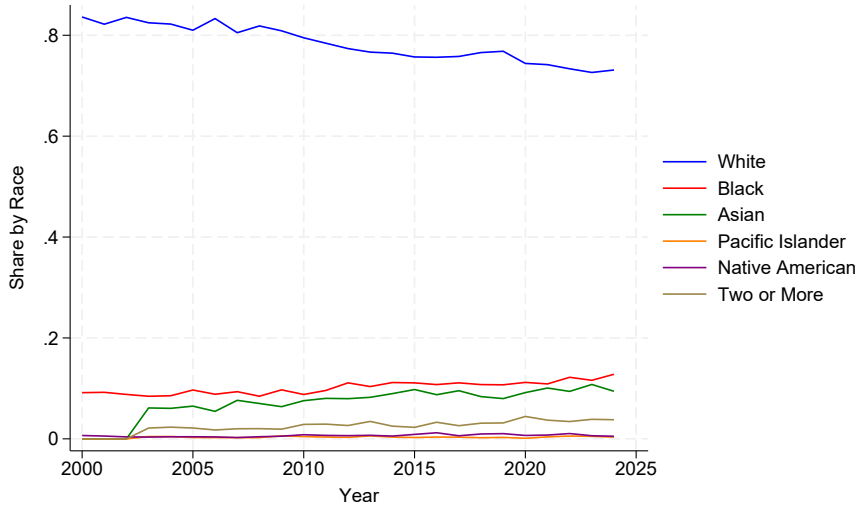

Source: Data on college students ages 18-24 during the period 2000-2024 from the CPS. First-generation status is calculated as described in the text.

Figure A4: Changes in the Proportion of Non-First-Generation Female Students Over Time



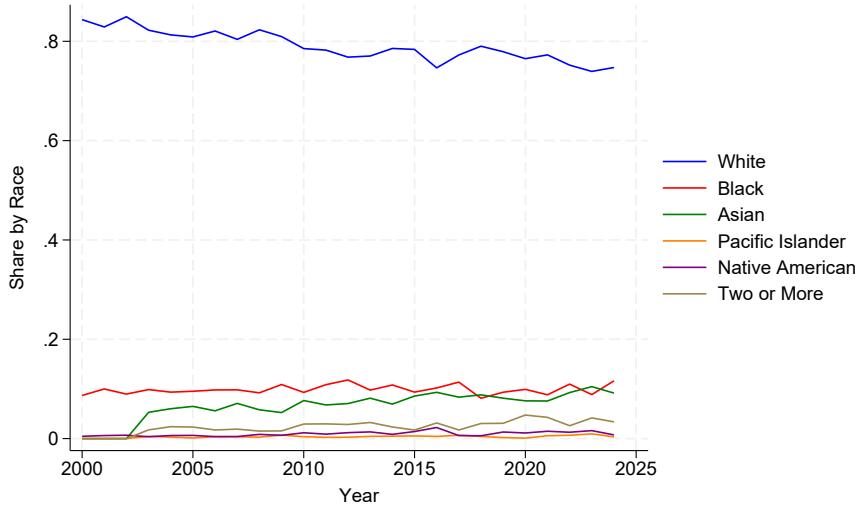

Source: Data on non-first-generation college students ages 18-24 during the period 2000-2024 from the CPS. First-generation status is calculated as described in the text.

Figure A5: Changes in the Proportion of Students by Race Over Time



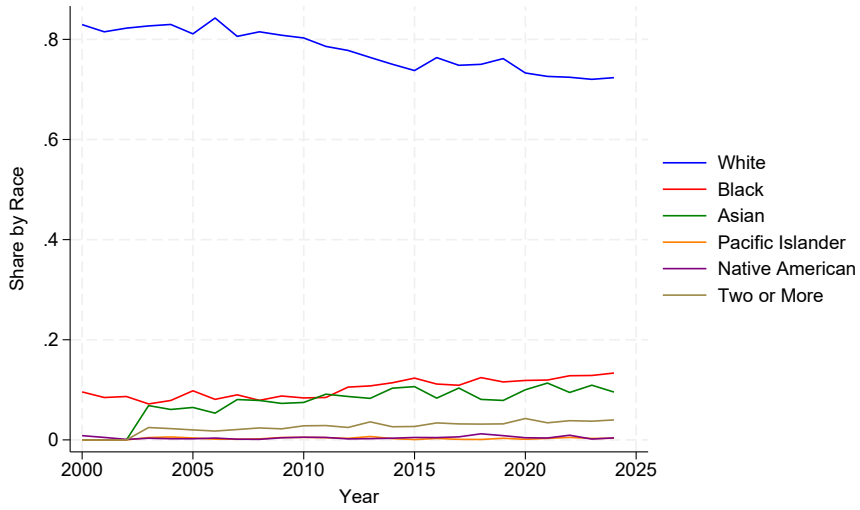

Source: Data on college students ages 18-24 during the period 2000-2024 from the CPS.

Figure A6: Changes in the Proportion of First-Generation Students by Race Over Time




Source: Data on first-generation college students ages 18-24 during the period 2000-2024 from the CPS. First-generation status is calculated as described in the text.

Figure A7: Changes in the Proportion of Non-First-Generation Students by Race Over Time




Source: Data on non-first-generation college students ages 18-24 during the period 2000-2024 from the CPS. First-generation status is calculated as described in the text.