Researching the characteristics of student debt holders in the United States

Alex Wycoff, Data Science Intern

August 2023

Abstract

The goal of this work is to investigate the characteristics of student debt holders in the United States. Because the majority of undergraduates are or were in a bachelor's degree program, we choose to focus on the debt of bachelor's degree recipients to maintain a reasonable scope while staying relevant to the majority of undergraduate students. Before our analysis, we overlay numerous factors related to demographics, education, institutional characteristics, and socioeconomic status with average student debt and develop functions that can clean and plot data using measures of central tendency. We then describe trends in the data and conclude as to their significance. Finally, we use said trends to answer the questions: how can prospective students aiming for a 4-year degree reduce their student debt after graduation, and what characteristics lead to higher average student debt among bachelor's degree holders?

1. Introduction

In total, Americans owe over $1.7 trillion on student loans, an amount that has only risen over the last decade. Many approaches to this subject focus on numbers like this, or categorize student debt as a type of student aid, comparing it to grants and merit aid. Our goal is to instead compare student debt to a broader array of variables. In this study, we compare student debt to both modifiable and unmodifiable factors to develop a clearer understanding of both what patterns exist within the student loan system among varying factors and what modifiable factors students can leverage to reduce their possible debt. Since student debt can have impacts long after graduation, prospective college students stand to benefit from knowing more about these factors that have the potential to reduce their student debt.

2. Data

Data for 2014 to 2021 comes from the Survey of Income and Program Participation (SIPP), which interviews survey members monthly and provides data on their current amounts of educational debt, income, race, gender, and age. Data retaining to students' geographical region, college entrance test scores, parents' educational attainment, institution characteristics, socioeconomic status percentile, and parents' income comes from the National Postsecondary Student Aid Study: 2020 Undergraduates (NPSAS: UG 2020), which also reports current amounts of educational debt. Due to some factors not

meeting reporting standards in the NPSAS: UG 2020, we have substituted the National Postsecondary Student Aid Study: 2016 Undergraduates (NPSAS: UG 2016) for the NPSAS: UG 2020 in some analyses, which are labeled as from 2016. For both sources, we use data only from those who have completed their bachelor's degrees to maintain consistency between the sources and allow for comparison.

3. Data Visualization and Analysis

In this section, we use Python 3.11 and Jupyter Notebook to visualize and generate graphs using age ranges, racial identification, and gender. We then analyze these graphs to determine whether each factor or group of factors affects student debt among bachelor's degree holders.

3.1 Demographics

First, we analyze trends in student debt among differing demographics.

3.1.1 Age

Figure 1 shows us that, by age, the majority of debt among bachelor's degree holders is held by younger people, specifically those 25-34 years old. However, older debt holders tend to have more debt, as shown in Figure 13 in Appendix A. This effect likely stems from those with unrepaid loans accumulating interest in their years out of college. Figure 14 in Appendix A shows that a smaller percentage of 4-year degree holders have student debt in older age groups, while over 30% of 4-year degree holders aged 34 or younger have student debt.

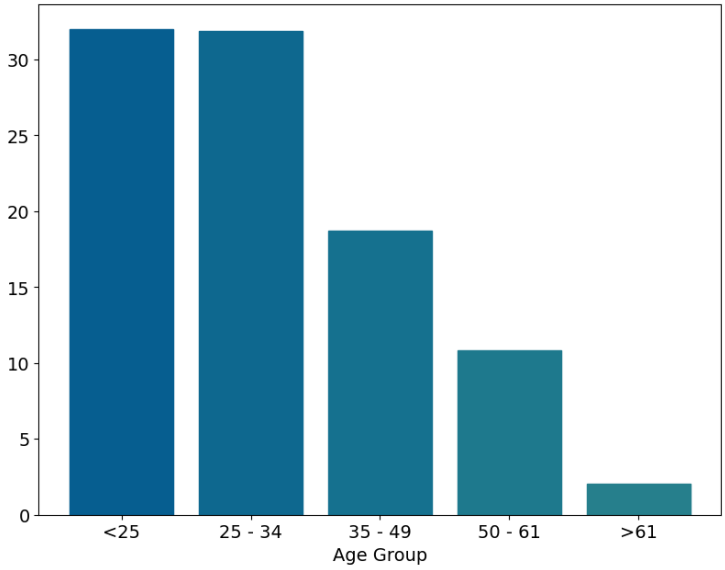

Figure 1: Percent of Age Group with student debt (SIPP 2021)

3.1.2 Race

On average, bachelor's degree holders who identify as black alone who owe student debt owe 15% more on average than the average student who identifies as white alone. Figure 2 shows that those who identify as black alone are 70% more likely to have student debt than those who identify as white alone and nearly 200% more likely than those who identify as Asian alone. Additionally, Figure 15 in Appendix A demonstrates how people who identify as white alone on average owe less in student debt than those who identify as black alone or Asian alone.

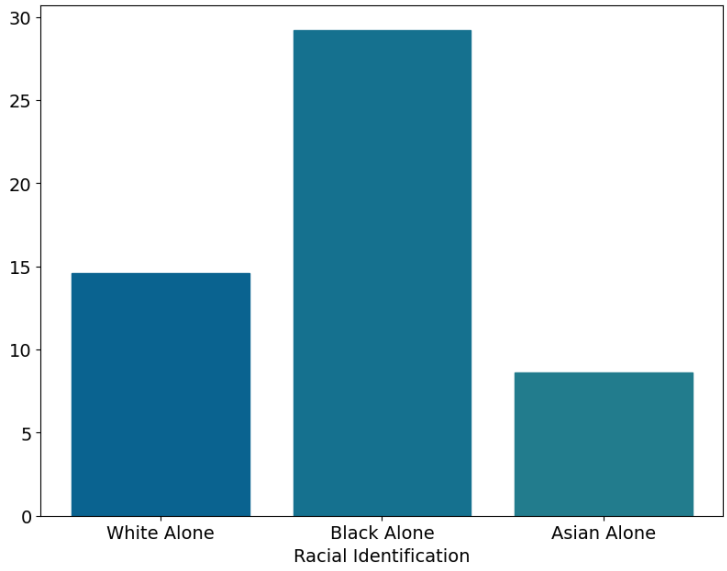

Figure 2: Percent of People with Student Debt by Racial Identification (SIPP 2021)

3.1.3 Gender

Gender is unlikely to correlate to average student debt. Figure 3 shows a slight bias toward males in terms of the percentage of people with student debt. Similarly, Figure 16 in Appendix A shows that data from the NPSAS: UG 2020 study has a small bias toward males, while data from the SIPP 2020 has a small bias toward females. Such insignificant and negligible changes make a connection between gender and average student debt statistically improbable.

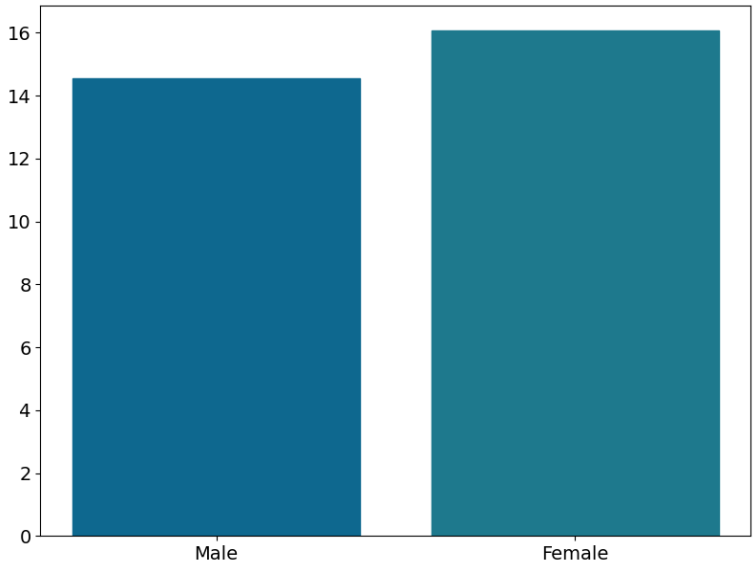

Figure 3: Percent of People with student debt By Gender (SIPP 2021)

3.2 Education

Second, we analyze the impact of a student's educational statistics on his or her student debt.

3.2.1 College Entrance Test Scores

To start, we examine the effect of two standardized college preparedness tests, the Scholastic Aptitude Test (SAT) and the American College Test (ACT), on average student debt. As can be seen in Figure 17 in Appendix A, taking one test, the other, or both have little correlation to average student debt. However, students who take neither test often incur less debt, likely due to their decreased ability to apply to colleges that require standardized test scores and thus increased chance to attend lower-tuition schools. In Figure 4, the length of intervals between scores is based on the standard deviation of scores for each test. While scorers from 12 to 31 and 610 to 1400 for the ACT and SAT, respectively, owe similar amounts of student debt to each other, scorers in the range of 1410 to 1600 for the SAT and 32 to 36 for the ACT owe significantly less than any other group of scorers. Because scores in the ACT and SAT follow Normal-like distributions, scorers on the high end of both tests' score ranges are exceedingly rare depending on how close their score is to perfect. This creates a situation where a few outliers with slightly higher - yet many times rarer - scores deflate the average student debt owed by others in their scoring category, so it is a mistake to assume that any score is a cut-off at which points

students incur less debt. Instead, this demonstrates the role standardized tests play in college admissions and merit selection.

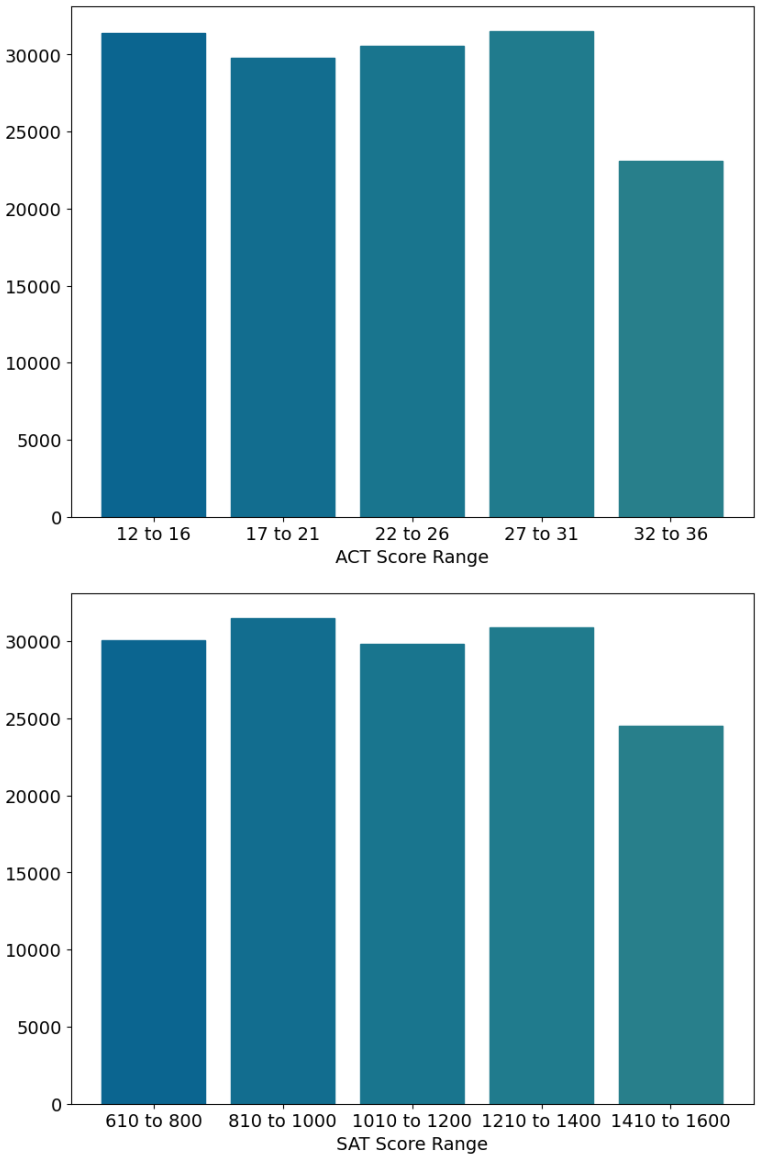

Figure 4: Average student debt by SAT (Top) and ACT (Bottom) Score Range (NPSAS: UG 2020)

3.2.2 High School and College Grade Point Average

When it comes to a student's grade point average, there are similar trends in both college and high school. Figure 5 shows that students who graduated from high school with a grade point average of 2.5-2.9 on average owed about 15% more in student debt than students who graduated with a 3.5-4.0 grade point average. Similarly, students who had an estimated 2.50-2.99 grade point average in college had about 13% more in student debt than students whose estimated grade point average in college was 3.5 or greater.

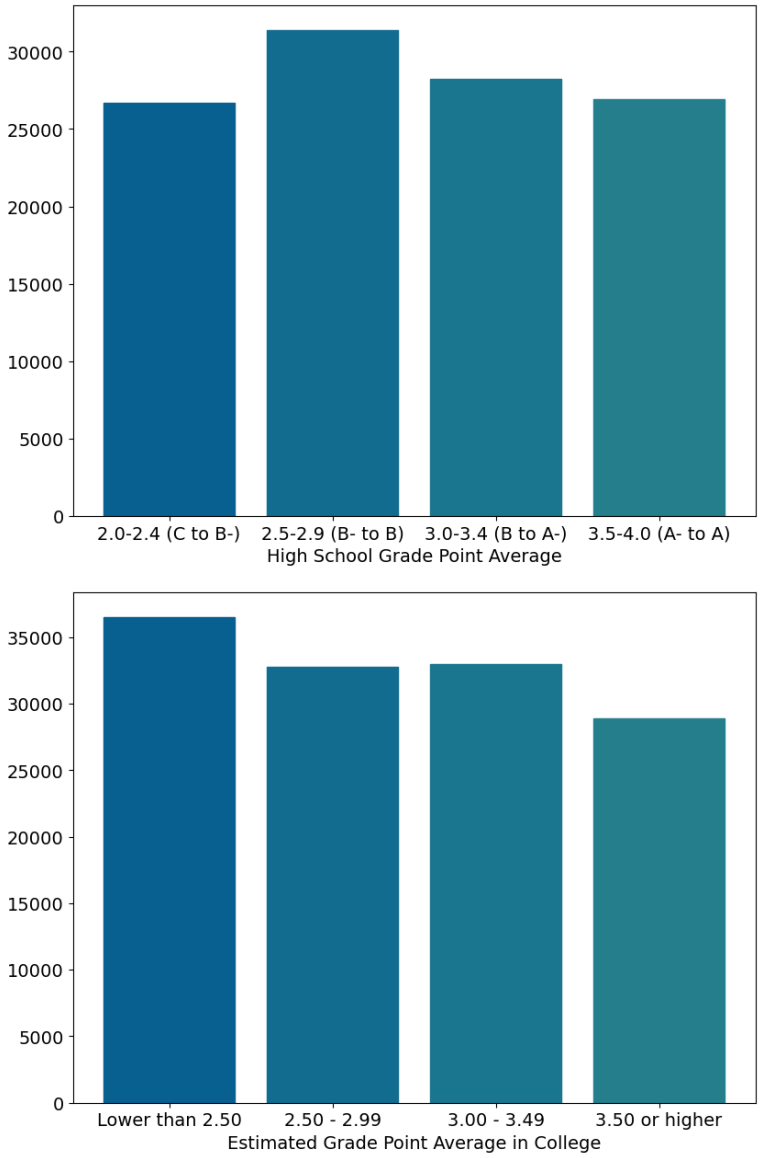

Figure 5: Average student debt by High School Grade Point Average (Top) and Estimated Grade Point Average in College (Bottom) (NPSAS: UG 2020)

3.3 Institution Characteristics

Due to the high degree of control, students have in choosing an institution to attend, prospective students need to understand the characteristics of their institution of choice. Here, we analyze the correlation between these characteristics and the student debt of bachelor's degree holders from such an institution in an attempt to determine the ideal institution characteristics to minimize student debt.

3.3.1 Institution Type and Selectivity

When comparing the student debt of students attending different types of institutions, we can see from Figure 6 that students who attend public schools have less student debt on average than those attending private non-profit or private for-profit schools. This is likely due to the higher tuition of private schools.

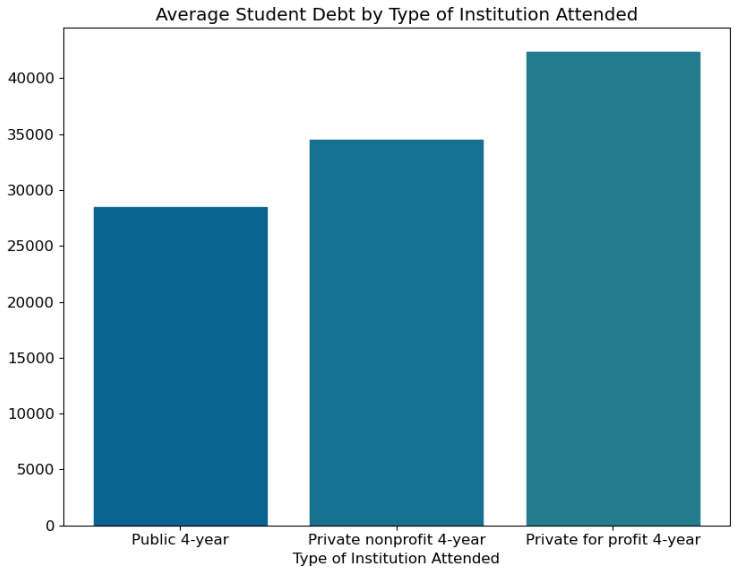

Figure 6: Average Student Debt by Type of Institution Attended (NPSAS: UG 2020)

The selectivity of a student's attended school, however, has a surprising trend. The selectivity of an institution is calculated by using the 25th and 75th percentiles of ACT and/or SAT scores and the admission rate of the institution. Of students who attended four-year institutions, 17.8% of students attended very selective schools, 51.6% of students attended moderately selective institutions, 8.8% of students attended minimally selective institutions, and 21.8% attended open admissions institutions. Generally, students attending more selective schools tend to have less student debt. This is likely due to more selective schools having policies like a 'no-loan' policy, which guarantees a college will match the entirety of a student's demonstrated need. However, 4-year open-admission schools graduates have debt skewed towards the higher end of the spectrum due to private for-profit institutions being the most common type of open-admission schools. This is demonstrated in Figure 8

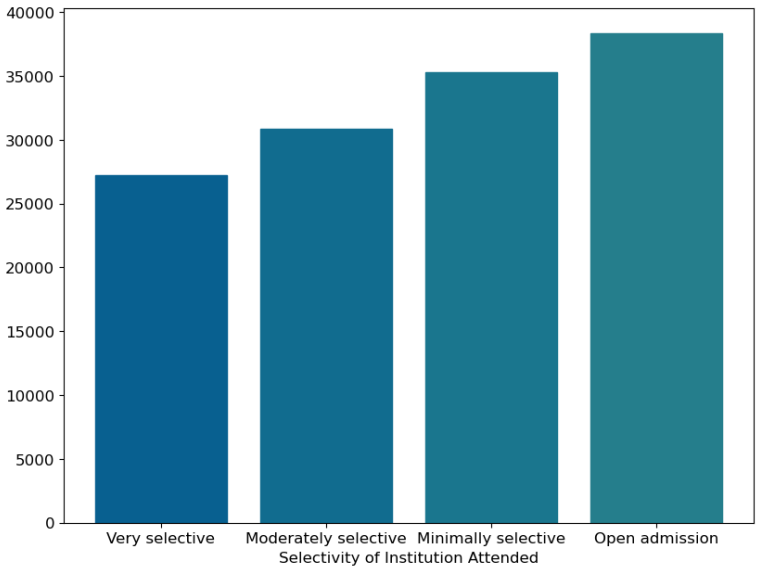

Figure 7: Average Student Debt by Selectivity of Institution Attended (NPSAS: UG 2020)

These data culminate into the idea that attending a very selective public school would result in a student obtaining the least student debt on average. To test this theory, we graph average student debt in terms of the type and selectivity of a student's attended institution in Figure 8.

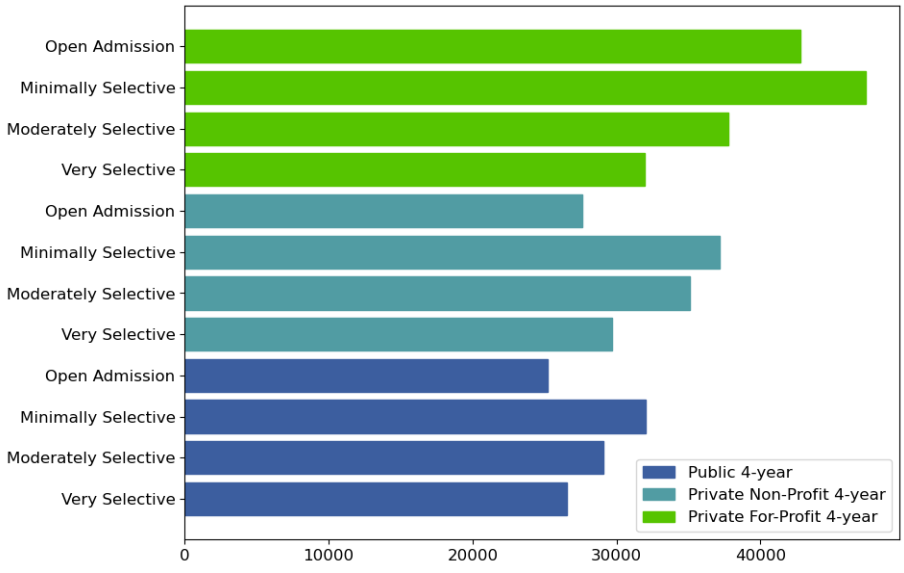

Figure 8: Average Student Debt by Type and Selectivity of Institution Attended (NPSAS: UG 2016)

Within Figure 8, we identify three main institution types as those that graduate 4-year students with the least amounts of average student debt. These are very selective public 4-year schools, open admissions public 4-year schools, and open admissions private non-profit 4-year schools. While both trends from the previous 2 graphs continue to apply, both public and private non-profit open admissions colleges produce significantly less indebted 4-year graduates. Prospective 4-year degree students should be aware of this large difference in indebtedness between students who attend minimally selective and public or

non-profit open admissions institutions. Those who are likely to attend a less selective institution may benefit from applying to public or non-profit open admissions colleges. According to the NCES, 14% of all private non-profit 4-year institutions are open admissions and 28% of all public 4-year institutions are open admissions. For more information and examples of open admissions 4-year public and private non-profit institutions, see Appendix B.

3.3.2 Geographical Region

Figure 9 displays the average student debt of students attending institutions within each BEA (Bureau of Economic Analysis) region; see Figure 18 in Appendix A for a map of the United States by BEA region. We can see that institutions within the Plains, Great Lakes, and New England areas tend to produce graduates with the highest average student debt while those in the Far West and Rocky Mountains produce students with the lowest average student debt. Figure 10 also shows the average student debt by a student's state of residence. We notice that the high average student debt in the Plains region is caused by the above-average amount in Minnesota, North Dakota, and South Dakota. In the Far West region, Nevada is an outlier with $10,000 more in average student debt per student than the regional average.

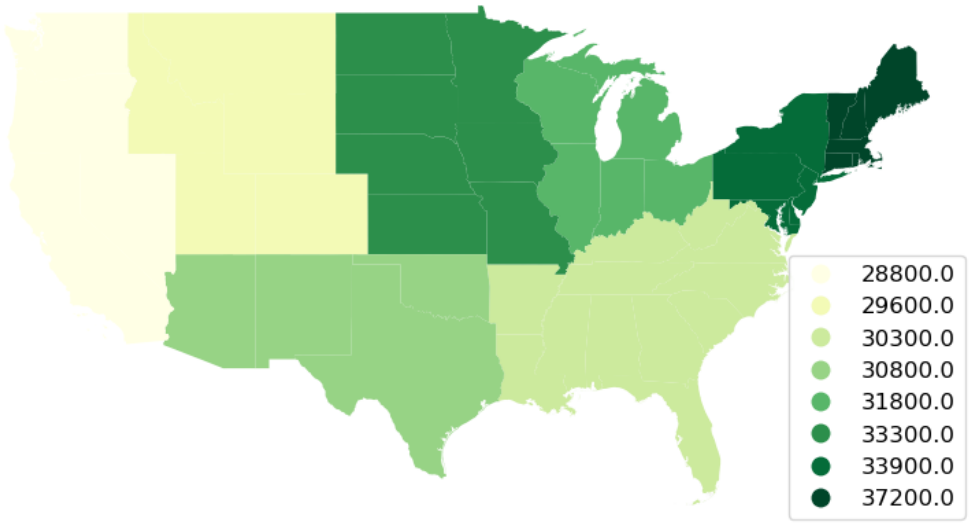

Figure 9: Average Student Debt by Region of Institution Attended (NPSAS: UG 2016)

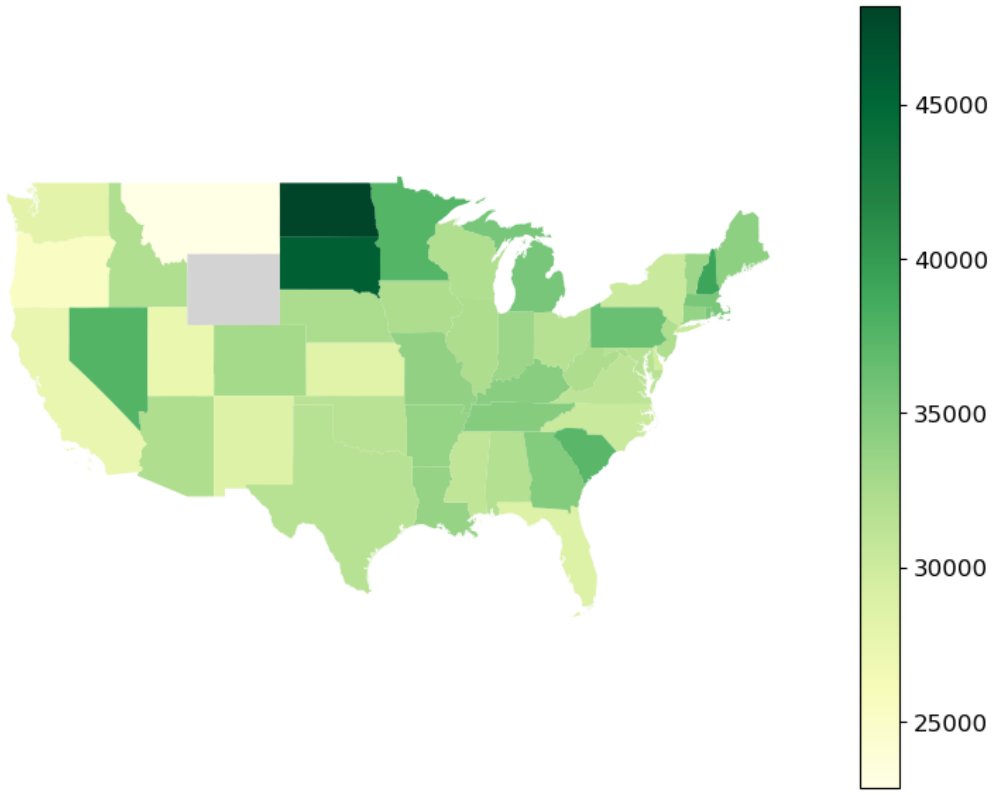

Figure 10: Average Student Debt by Student's State of Residence (NPSAS: UG 2016)

3.4 Income & Socioeconomic Status

3.4.1 Current Income

In Figure 11, the income shown for dependent students is the income of the student's parents while the income shown for independent students is income earned by the student. Figure 11 illustrates how, on average, dependent students owe more in student debt than independent students. Among dependent students, students whose parents have higher incomes owe more student debt on average. Among independent students, it appears that the average amount of student debt owed falls with greater levels of income, but

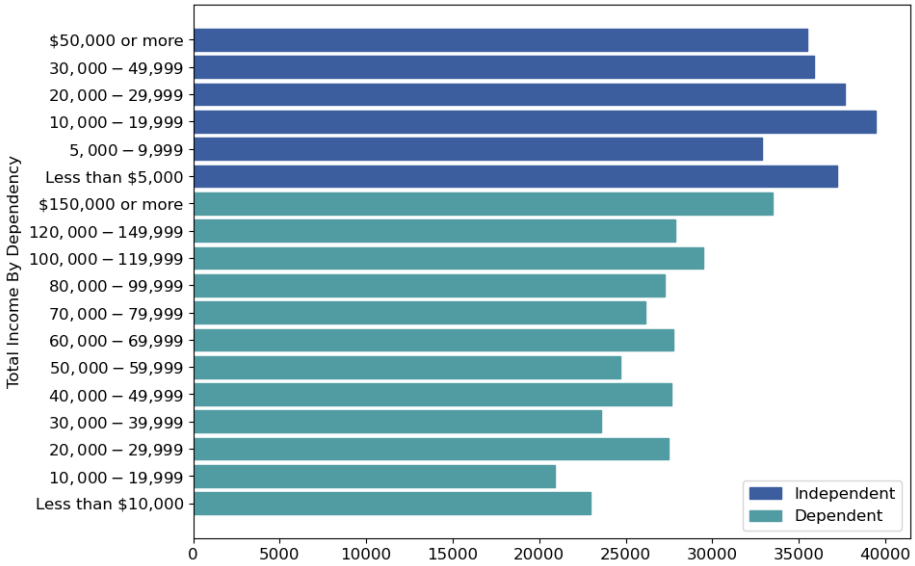

Figure 11: Average Student Debt by Total Income and Dependency (NPSAS: UG 2020)

3.4.2 Socioeconomic Status

A student's composite socioeconomic status score is calculated using both of his or her parents' levels of educational attainment, socioeconomic index based on their occupations, income, and the student's own income. This composite score is then used to calculate their socioeconomic status percentile. Figure 12 displays the average amount of student loan debt held by students in every tenth socioeconomic status percentile. We can see that debt is significantly lower among students on the lower end of socioeconomic status percentiles. Average student loan debt is then at its highest at the 41st to 50th percentile range. It then decreases and plateaus from the 61st to the 100th percentile at around $29,000. These data suggest that the top 40% of students, by socioeconomic status, are less likely to differ in what they owe in student debt while the bottom 30% of students owe the least student debt on average out of any percentile range.

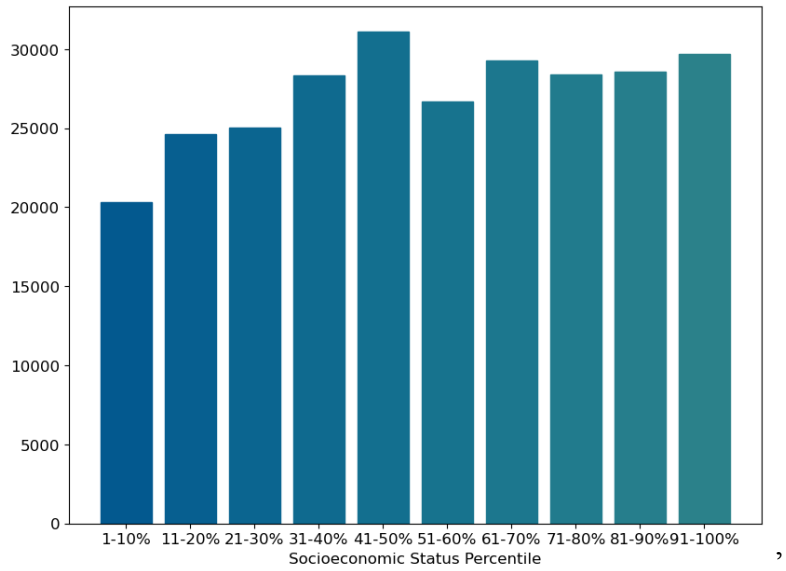

Figure 12: Average Student Loan Debt by Socioeconomic Status Percentile (NPSAS: UG 2020)

4. Findings

In conclusion of our findings, we first note that average student debt varies significantly across different demographics. Those who identify as black are much more likely than the other identifications considered to owe student debt, and those who owe student debt also owe more on average. We also note that while younger people owe more of the total student debt, people who are older and still owe student debt tend to owe more. However, gender has little - if any - impact on average student debt.

Secondly, we note that education both before and during college can impact average student debt. Both a student's high school and college grade point averages have similar effects on average student debt, implying that a student's characteristics in high school are likely to affect their college experience. Additionally, we find that a student's ACT or SAT scores have a significant ability to reduce average

student debt for students that score in the upper percentiles of either test, while whether a student takes one or the other has little to no effect.

Thirdly, we find that the characteristics of a student's attended institution can have surprisingly significant effects on average student debt. We learn that more selective institutions graduate students with lower amounts of student debt on average, while open-admissions institutions often graduate students with some of the most. Furthermore, public schools tend to graduate students with less student debt on average than private non-profit schools, and a significant amount less than private for-profit schools. Despite private for-profit open admissions schools having graduates with some of the highest average student debt, public and private non-profit open admissions schools have some of the lowest average student debt among graduates. We also see that the Plains, Great Lakes, and New England areas of the US tend to have residents with the most student debt on average.

Finally, we observe the impact of socioeconomic status on average student debt, which shows us that students who have lower incomes or have a lower socioeconomic status, especially one in the bottom 40%, have less student debt on average in comparison to students who have higher incomes or socioeconomic statuses.

5. Limitations

In this study, we intentionally limit our scope to bachelor's degree holders who owe student debt in order to produce information relevant to the majority of prospective undergraduate students. However, considering other degrees and students without debt could reveal additional patterns, which could potentially be useful to undergraduate students who intend to complete a 2-year degree or go beyond a 4-year degree. The timeframe of our research also serves as a limitation, as we use data from 2020 and 2021 at the latest, and some missing data from the NPSAS: UG 2020 is replaced with data from the NPSAS: UG 2016.

6. Future Work

In the future, we can potentially use data spanning through 2023 to determine the effects of the student debt payment pause and the COVID-19 pandemic. In another respect, we can attempt to focus on the practices of these private for-profit institutions to determine why graduates of them generally have significantly more debt than those of other types of institutions. On top of that, we can leverage college information databases in order to determine what universities are exemplars of institutions that produce especially indebted students or ones with exceptionally little debt.

7. References

"IPEDS Analytics: Delta Cost Project Database." *National Center for Education Statistics*, Institute of Education Sciences, https://nces.ed.gov/ipeds/use-the-data/delta-cost-project-finance-data. Accessed 10 August 2023.

National Center for Education Statistics. "Characteristics of Degree-Granting Postsecondary Institutions." *U.S. Department of Education*, Institute of Education Sciences, May 2022, https://nces.ed.gov/programs/coe/indicator/csa. Accessed 8 August 2023.

"National Postsecondary Student Aid Study: 2016 Undergraduates (NPSAS:UG)." *National Center For Education Statistics*, Institute for Education Sciences, https://nces.ed.gov/datalab/powerstats/121-national-postsecondary-student-aid-study-2016-undergraduates/averages-medians-percents. Accessed 21 July 2023.

"National Postsecondary Student Aid Study: 2020 Undergraduates (NPSAS:UG)." *National Center for Education Statistics*, Institute of Education Sciences, https://nces.ed.gov/datalab/powerstats/157-national-postsecondary-student-aid-study-2020-undergraduate-students. Accessed 3 August 2023.

"Survey of Income and Program Participation Datasets." *Census Bureau*, https://www.census.gov/programs-surveys/sipp/data/datasets.html. Accessed 10 July 2023.

"U.S. Bureau of Economic Analysis Regions | Iowa Community Indicators Program." *Iowa Community Indicators Program*, Iowa State University, https://www.icip.iastate.edu/maps/refmaps/bea. Accessed 2 August 2023.

Appendix A

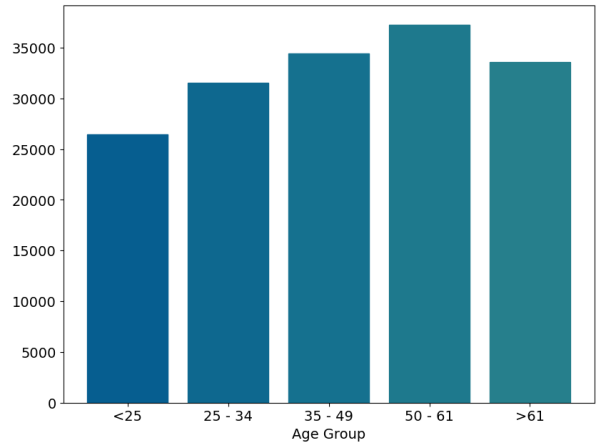

Figure 13: Average student debt by Age Group (SIPP 2021)

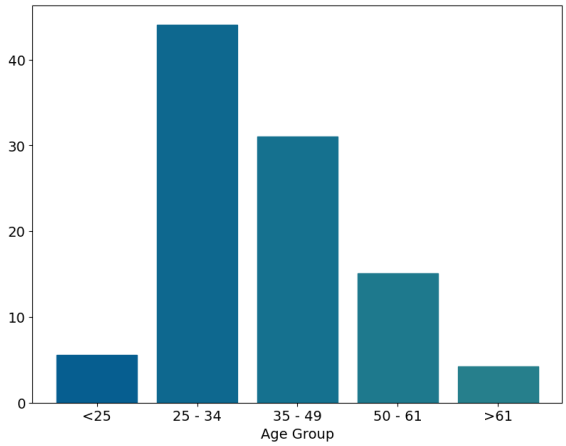

Figure 14: Percent of Total student debt by Age Group (SIPP 2021)

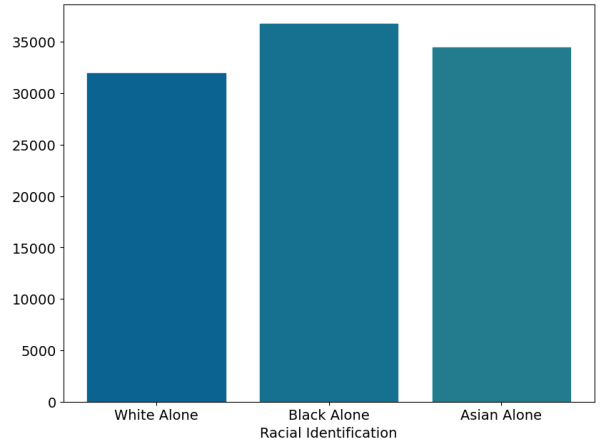

Figure 15: Average Student debt by Racial Identification (SIPP 2021)

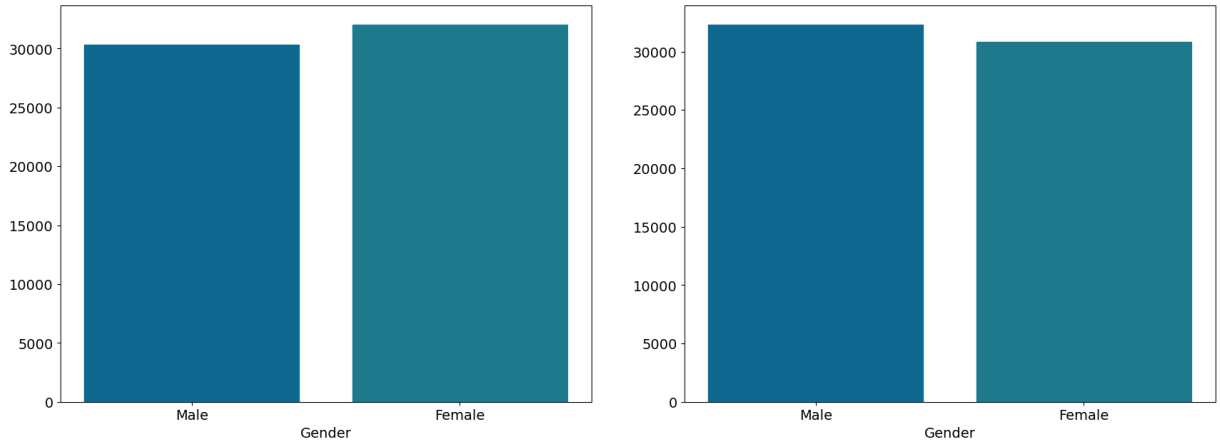

Figure 16: Average Student Debt by Gender (Top: NPSAS: UG 2020, Bottom: SIPP 2020)

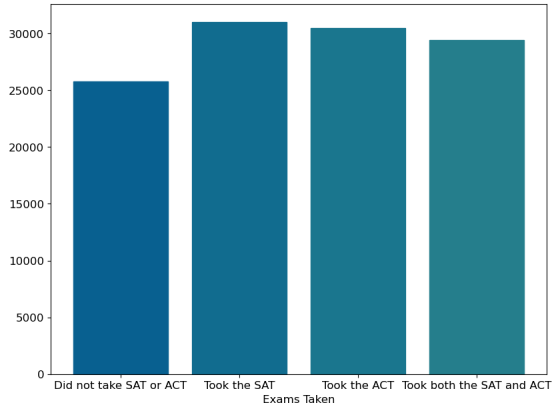

Figure 17: Average Student Debt by Exams Taken (NPSAS: UG 2020)

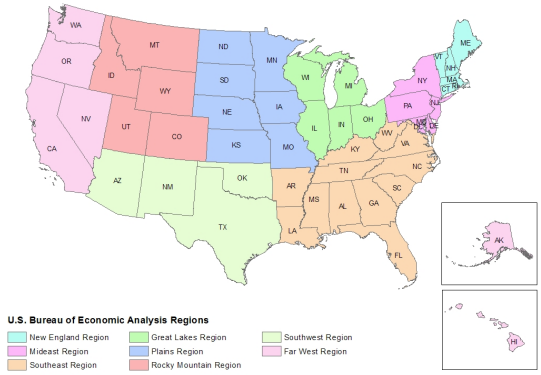

Figure 18: BEA economic regions (Iowa State University)

Appendix B

A full list of institutions and their respective values for the variables mentioned below can be found at github.com/AlexWycoff/studentdebt. All data shown is from the 2000-2012 Delta Cost Project, so some of the institutions listed may no longer be continuing in education. All institutions had a 100% acceptance rate at the time of the survey (given in the variable 'academicyear') and were either public 4-year institutions or private non-profit 4-year institutions. The following is a list of the accompanying variables:

- academicyear: academic year at the time of the survey.
- instname: name of the institution.
- city: city of the institution.
- state: state of the institution.
- zip: zip code of the institution.
- sector_revised: sector of the institution; 1 corresponds to public 4-year schools while 2 corresponds to private non-profit 4-year schools.
- loan_num: number of undergraduate students receiving student loans.
- loan_avg_amount: the average amount of student loans received by undergraduates.
- applicantcount: the number of applicants in the given year.
- admitcount: the number of admittances in the given year.
- enrollftcount: the number of first-time, full-time, degree/certification-seeking students enrolling full-time in the given year.
- actcm25: the institution's ACT Composite 25th percentile score.
- actcm75: the institution's ACT Composite 75th percentile score.
- satmt25: the institution's SAT Math 25th percentile score.
- satmt75:the institution's SAT Math 75th percentile score.
- satvr25: the institution's SAT Critical Reading 25th percentile score.
- satvr75: the institution's SAT Critical Reading 75th percentile score.