College Data:

Investigating Student Debt Among First-Generation College Students

Introduction

Higher education plays a pivotal role in shaping the futures of individuals, offering the promise of upward mobility and career opportunities. However, for many students, particularly those who are the first in their families to attend college, the pursuit of higher education comes with a unique set of challenges, such as the burden of student debt. The financial decisions made during college can have lasting consequences, impacting financial well-being well into adulthood.

Understanding the dynamics of student debt among first-generation students is crucial for policymakers, educators, and institutions of higher learning. It provides valuable insights into the challenges faced by this demographic group and serves as a foundation for initiatives aimed at forming more equitable opportunities for educational achievement.

This study investigates the financial landscape of first-generation college students, utilizing quantitative analysis to explore patterns and connections within various percentiles of this student demographic. Its primary objective is to unveil disparities and trends in student debt across distinct segments while identifying the statistical factors contributing to these variations. By doing so, this analysis offers valuable insights into the intricate relationship between first-generation students and student debt, enabling the formation of data-driven conclusions.

Hypothesis

Significant disparities exist in student debt levels among first-generation college students, and these disparities are influenced by a complex interplay of factors that can be identified and quantified through statistical analysis.

Key questions addressed in this report include:

- **Distribution of Colleges and Debt Accumulation:** How is student debt distributed among different quarterly percentiles of first-generation students, and how does it vary across colleges?
- **Variation in Debt Levels:** How does the variation in student debt levels differ among different categories of first-generation student percentages?
- **Correlation Analysis:** Is there a correlation between the percentage of first-generation students and student debt? How does student debt correlate with factors of median income, net price of attendance, and enrollment size?
- **Locale and Debt:** Does the locale of a college impact the mean debt for first-generation students?
- **State-Level Analysis:** How does the relationship between the percentage of first-generation students and average debt differ across states?

Methodology

The methodology for the analysis involves R code to generate various graphs and conduct descriptive statistical analyses on a dataset consisting of 2012 colleges. The dataset included the following variables: FirstGen, Debt, MedIncome, NetIncome, Enrollment, Locale, and State. Below is a brief overview of the methodology steps:

- **Data Collection**: Data extracted from the accurate Lock5 sourced Excel file, which includes information concerning 2012 colleges, forms the basis for the report's analysis. It is important to note that some colleges had incomplete data.
- **Data Cleaning**: Prior to analysis, the dataset underwent a data cleaning process involving handling missing values appropriately and addressing outliers that could skew the results.
- **Descriptive Graphs & Statistical Methods**: The methods encompassed data tables, box and whisker plots, bar graphs, scatter plots, central line of tendency, correlation coefficients, and calculations of mean, median, and range. Various graphs and visualizations were created to represent the findings, aiding in interpreting the results and making them more accessible.

Analysis

What is the number of colleges and average debt accumulation amounts within each quarterly percentile of first-generation students?

This dataset offers meaningful insights into college distribution and average debt accumulation across quarterly percentiles of first-generation students. Figure 1 presents the number of colleges within each percentile category (0-25%, 25-50%, 50-75%, and 75-100%) and their corresponding average debt levels. Notably, there is an overall upward trend in average debt as we progress from lower to higher first-

generation student percentiles. The 75-100% percentile category stands out with the highest average debt and the highest number of colleges featuring a substantial proportion of first-generation students. In contrast, the 0-25% percentile exhibits the lowest average debt and fewer colleges. This suggests that, on average, colleges with a higher percentage of first-generation students tend to have higher average debt accumulation among their students.

FirstGenPercentile	Count	AvgDebt
0-25%	448	1384
25-50%	446	2408
50-75%	450	1797
75-100%	668	3622

Figure 1

How does the variation in student debt levels differ among different categories of percentages of first-generation students?

The graph presented in Figure 2 offers a visual exploration of the relationship between different categories of first-generation student levels and student debt. In both the 8.82% to 27.9% and 27.9% to 47% first-generation student percentage ranges, notable outliers are present, with whiskers extending above, indicating a skew toward higher values. In the 47.6% to 66% range, a single outlier stands out above the upper whisker, signifying an exceptional data point. Additionally, the box size suggests a wider data distribution, and the whisker extends above the box, highlighting a concentration of data points with higher values. Conversely, in the 66% to 85.2% first-generation student percentage range, there is no visible box, which indicates an absence of first-generation student data within this category. In the "NA" category, outliers are apparent, implying a lack of first-generation student data but the presence of student debt information. Clearly, there are distinct patterns of student debt variation across various first-generation student percentage categories. Some ranges exhibit skewed distributions, while others reveal empty categories, highlighting potential data limitations.

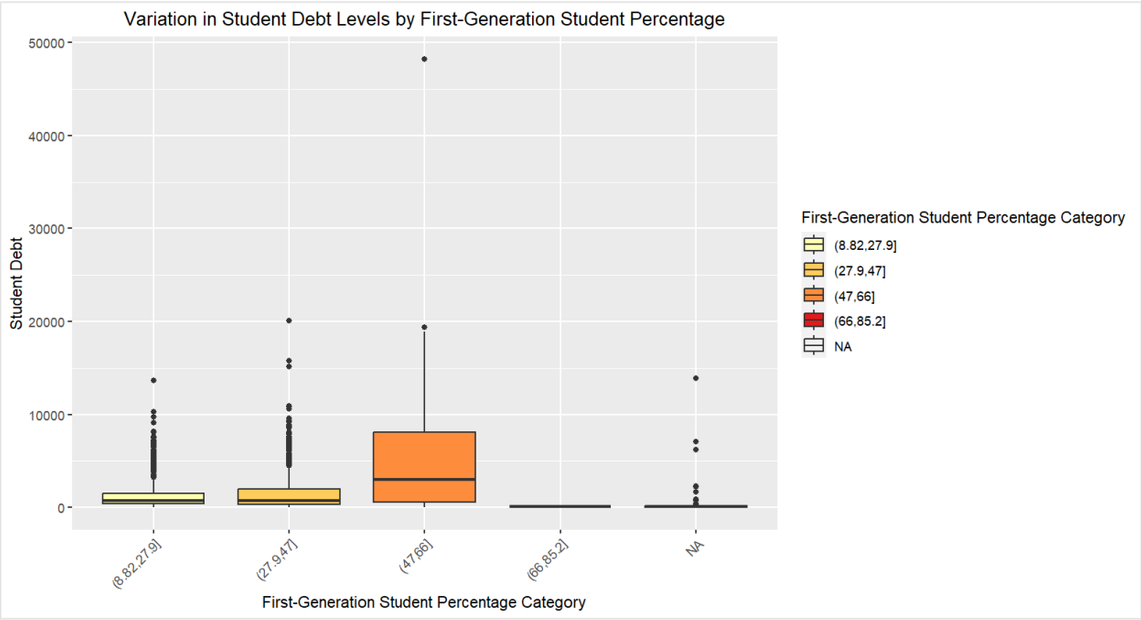
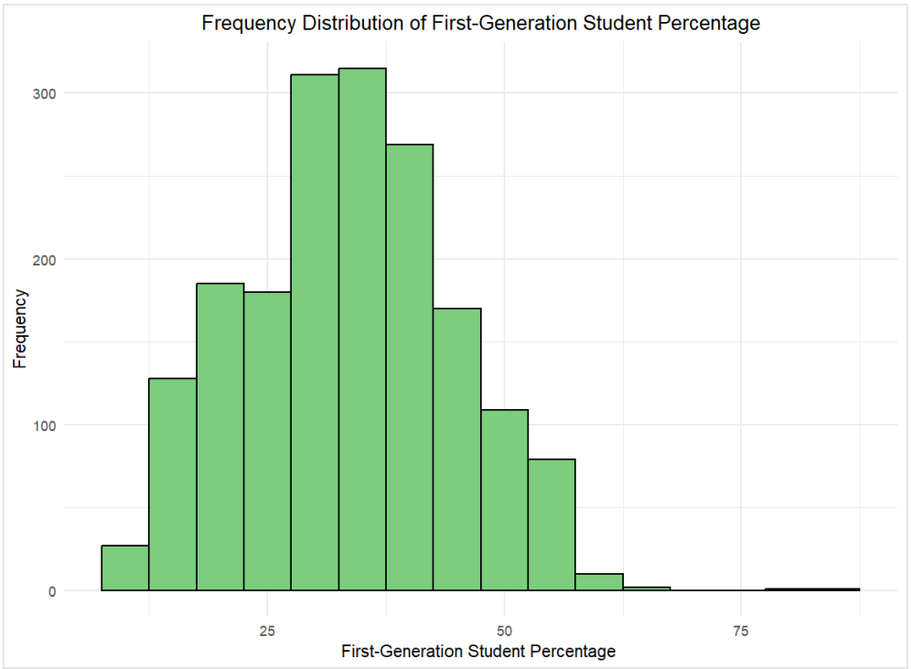


Figure 2

What is the frequency distribution of the percentages of first-generation students, and what are the median and mean percentage values?

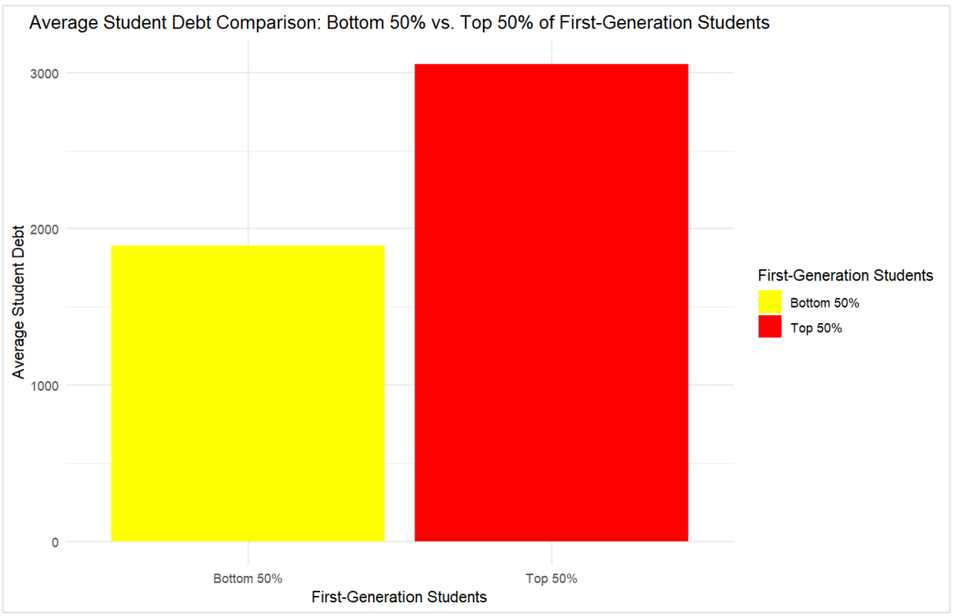
Numerical and visual representations, as demonstrated in Figure 3, effectively illustrate the distribution of first-generation student percentages. This distribution exhibits a nearly symmetrical, normal pattern similar to a bell curve, with colleges predominantly congregating around the middle of the distribution. The median, at 33.8%, signifies that approximately half of the colleges have percentages below this value, while the mean, slightly lower at 33.6%, serves as the average percentage across all colleges. Notably, there may be a few outliers with percentages exceeding 75%. Overall, this combination of visual symmetry, median, and mean provides a broad understanding of the distribution of first-generation student percentages among colleges, highlighting both the typical clustering and the presence of a few higher outliers in this dataset.



Median: 33.8
Mean 33.6
Figure 3

Based on the above question, distributing the top and bottom percent of first-generation students upon the median value of 33.8, how does the average student debt among the bottom and top 50% of first-generation students compare?

Analyzing student debt distribution among the top and bottom 50% of first-generation students using a median value of 33.8% reveals a significant gap. From Figure 4: On average, the bottom 50% carry a debt of $1,893.069, while the top 50% have a much higher average debt of $3,054.965. This results in a difference, or range, of approximately $1,161.896 in average student debt, with the top 50% having the higher burden. Translating to a substantial percentage difference of about 61.42%, it highlights the financial disparities between these groups based on their position relative to the median value.

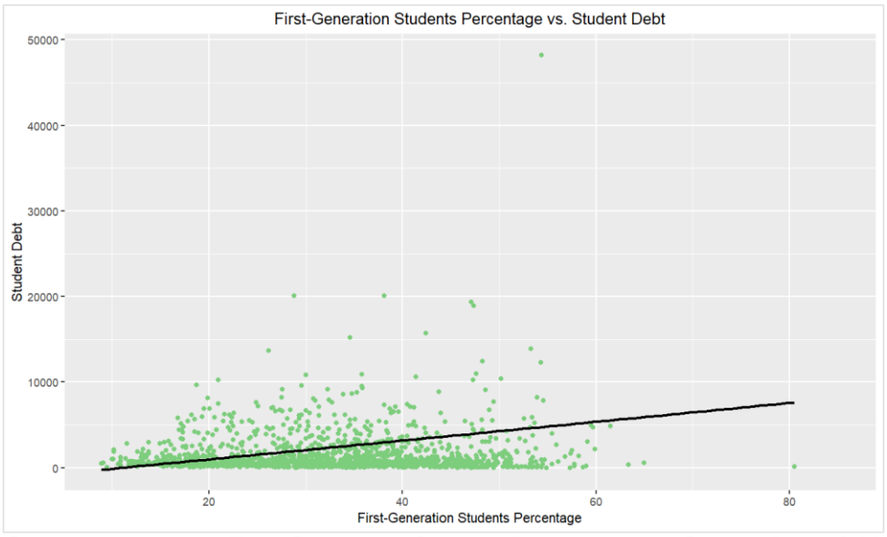


Average Student Debt for Bottom 50%: 1893.069
Average Student Debt for Top 50%: 3054.965
Figure 4

Is there a positive or negative correlation between the percentage of first-generation students and the debt accumulated?

Figure 5 explores the correlation between the percentage of first-generation students and their average debt accumulation. The calculated correlation coefficient of 0.22 indicates a positive but relatively weak linear relationship between these variables. In practical terms, as the percentage of first-generation students increases, there is a slight tendency for average debt to increase, although the connection is not particularly strong. While there is a positive correlation between the percentage of first-generation

students and average debt, other factors also play a role in influencing debt levels, and the relationship is not evident.

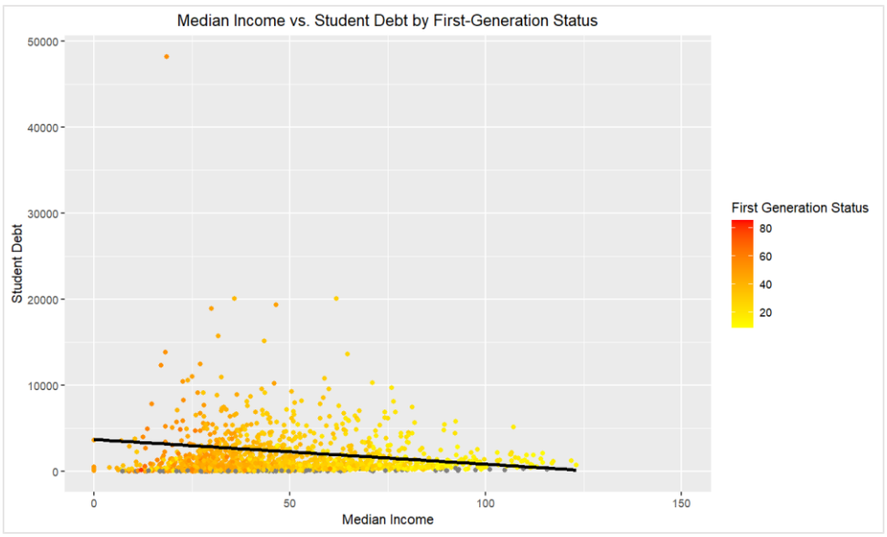


Correlation between First-Generation Students Percentage and Student Debt: 0.22

Figure 5

Amongst median income and student debt, is there a connection involving the percentages of first-generation college students?

The scatterplot provides a visual representation of the relationship between median income, student debt, and the percentage of first-generation students at colleges. Figure 6 reveals a negative correlation of -0.121 between median income and student debt, indicating a weak inverse relationship. As median income increases, there is a tendency for student debt levels to decrease. This aligns with the expected trend, implying that colleges in areas with higher median incomes tend to have students with lower levels of debt. Additionally, the color coordination of points based on the percentage of first-generation students, transitioning from red (higher percentages) to yellow (lower percentages), reinforces the negative correlation. This demonstrates that colleges with higher percentages of first-generation students, often located in areas with lower median incomes, face greater financial challenges and typically carry higher student debt burdens. Conversely, areas with higher median incomes, hosting fewer first-generation students, tend to exhibit lower student debt levels, highlighting the financial disparities faced by different student populations.



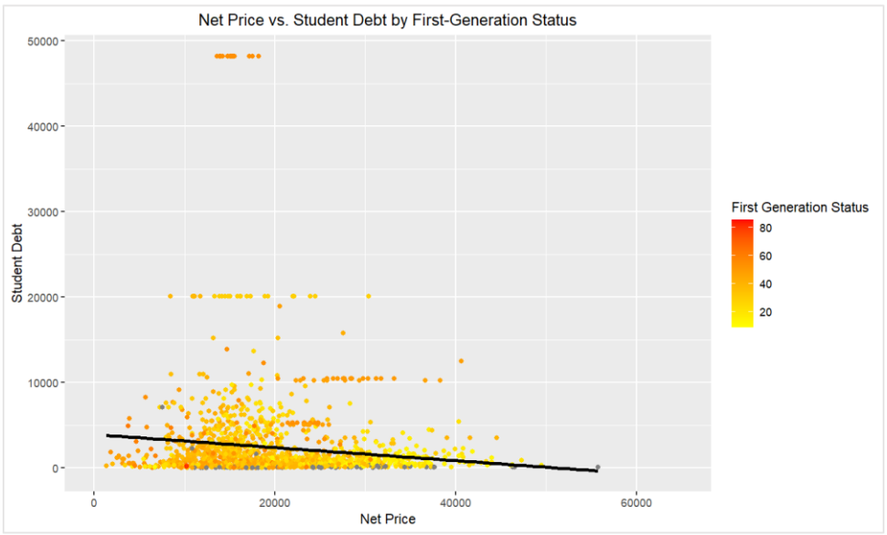
Correlation between Median Income and Student Debt for First-Generation Students: -0.121

Figure 6

Amongst the net price of attendance and student debt, is there a connection involving the percentages of first-generation college students?

Figure 7's scatterplot visually depicts the intricacy among net price, student debt, and the percentage of first-generation students at colleges. The negative correlation, denoted by a coefficient of -0.109 between net price and student debt, signals a somewhat feeble and adverse connection between these two factors. As net price increases, student debt generally decreases, though this association lacks strength. The incorporation of color-coding, representing the percentage of first-generation students, introduces an intriguing layer to the analysis. Moving from left to right on the graph, there appears to be a somewhat organized yet simultaneously sporadic transition from red to yellow. This weak negative correlation

implies that, while a tendency exists for institutions with higher net prices to exhibit lower student debt, the relationship is not particularly strong.

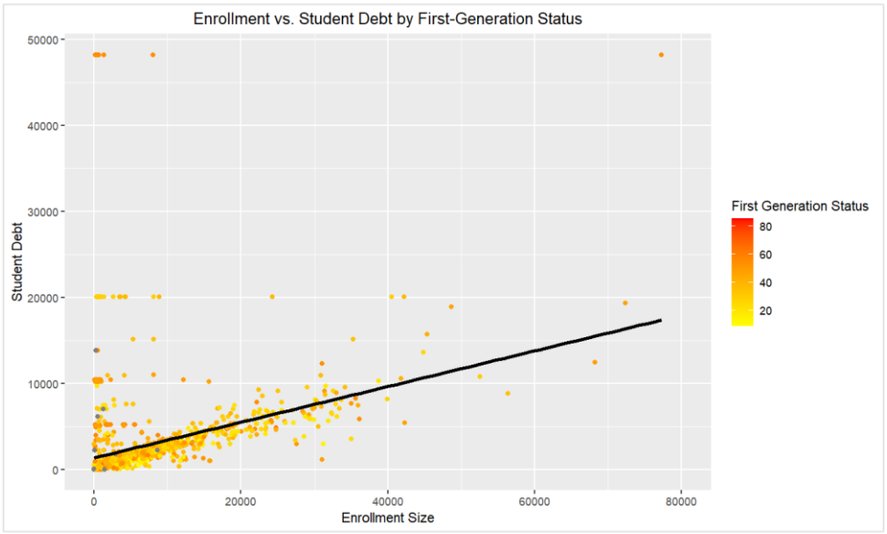


Correlation between Net Price and Student Debt for First-Generation Students: -0.109

Figure 7

Amongst the enrollment size and student debt, is there a connection involving the percentages of first-generation college students?

Scatterplot visualization of Figure 8 shows a positive correlation of 0.297 between enrollment size and student debt, indicating a slightly weak yet positive relationship. However, the color-coding of first-generation student percentages doesn't reveal a clear pattern, suggesting a limited association with enrollment size or student debt. The positive correlation implies larger institutions tend to have higher student debt, but it's not very strong and varies considerably. With insights into enrollment size, student debt, and first-generation students' percentages, there are cautions against inferring causation, especially amongst first-generation student relations. Further research and advanced modeling are required for definitive conclusions on influencing factors.



Correlation between Enrollment and Student Debt for First-Generation Students: 0.297

Figure 8

What is the mean debt for colleges located in different locales for each of the top and bottom 50% of first-generation students?

Examining numerical data across different locales offers valuable insights into mean debt levels for both low and high percentages of first-generation students. Figure 9 presents two sets of locales with similar mean debt patterns and two sets with contrasting trends. Specifically, when comparing City and Town locales, we observe remarkably close mean debt figures, with City range differing by only $3 and Town by $121. Conversely, the mean debt patterns for Rural and Suburb locales exhibit significant variations. In Rural areas, the mean debt for the bottom 50% is over triple that of the top 50%, while in Suburb locales, the mean debt for the bottom 50% is more than 1.5 times higher than that of the top 50%. The data highlights diverse mean debt levels across locales, with City and Town areas showing similarity, while Rural and Suburban locales display significant differences, implying potential disparities in educational access and financial burdens.

Locale	Mean Debt Bottom 50%	Mean Debt Top 50%
City	2237	2234
Rural	2144	693
Suburb	3022	1817
Town	1195	1074

Figure 9

How does the relationship between the percent of first-generation students and average debt differ across states?

Figure 10.0 reveals varying tendencies across states. In a few state graphs, the dataset appears insufficient to draw meaningful inferences due to its limited size, potentially obscuring any detectable trends or patterns. Considering outliers, approximately half of the state graphs depict scatter dots forming a rectangular pattern from left to right, indicating a lack of significant rise or fall and suggesting a relatively stable or linear relationship between the analyzed variables. Nearly one-quarter of the state graphs can be observed to obtain a subtle rise, indicating a mild positive correlation between the variables or a more pronounced ascent, signifying a stronger positive correlation.

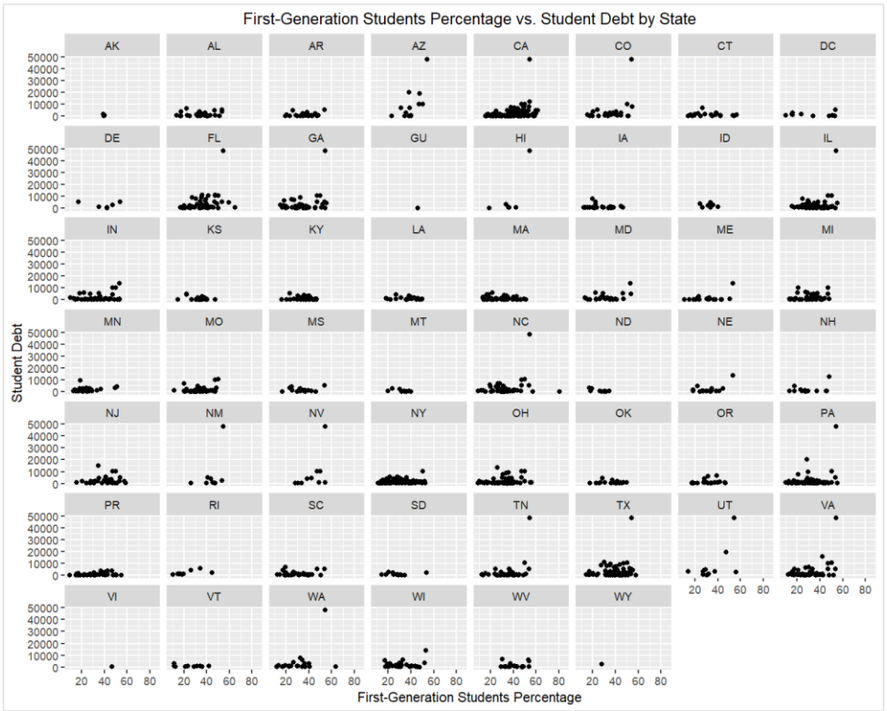


Figure 10.0

Numerical data offers concrete insights into various state-level metrics. Figures 10.1-10.4 provide a snapshot of key statistics, highlighting both the highest and lowest average debt figures and the states with the highest and lowest percentages of first-generation students. The dataset reveals a wide range of average debt amounts, spanning from a minimum of $141 in Guam (GU) to a maximum of $11,855 in Arizona (AZ). Similarly, mean percentages of first-generation students exhibit variability, with Rhode Island (RI) at the lowest end with 23.7% and the Virgin Islands (VI) at the highest with 46.6%. Notably, Nevada (NV) holds a unique position, ranking third with highest mean debt at $7,329 and highest mean percentage of first-generation students at 43.4%, showcasing an interesting overlap in these categories.

Top 3 States with:

Highest Mean Debt

State	Mean Debt
AZ	11855
UT	8249
NV	7329

Figure 10.1

Lowest Mean Debt

State	Mean Debt
GU	141
VI	267
VT	599

Figure 10.2

Top 3 States with:

Highest Mean % First-Gen. Students

State	Mean %
VI	46.6
NM	43.9

Lowest Mean % First-Gen. Students

State	Mean %
RI	23.7
MN	23.8

NV	43.4

Figure 10.3

ND	26.2

Figure 10.4

Results & Conclusion

In this comprehensive analysis, a hypothesis was formulated: Significant disparities exist in student debt levels among first-generation college students, and these disparities are influenced by a complex interplay of factors that can be identified and quantified through statistical analysis.

Key findings include:

- **Debt Variation Across Percentiles:** Average debt showed an upward trend across percentiles of first-generation students, with the 75-100% range standing out for having both the highest debt levels and the largest number of colleges.
- **Significant Debt Concentration:** Notably, the 47-66% category of first-generation college students carried the highest student debt amounts and was the most substantial in terms of size compared to other categories.
- **Symmetrical Distribution:** The distribution of first-generation student percentages displayed a symmetrical, bell-curve shape, featuring a median of 33.8% and a mean of 33.6%.
- **Wide Gap in Debt:** The top 50% of first-generation students bore an average debt that was 61% higher than that of the bottom 50%, indicating a significant disparity.
- **Positive but Weak Correlation:** A positive, weak correlation (0.22) was observed between the percentage of first-generation students and their debt levels.
- **Income Influence:** There was a negative correlation (-0.121) between median income and student debt, signifying that as median income increased, student debt tended to decrease.
- **Net Price Impact:** Similarly, as net price increased, student debt decreased, as indicated by a negative correlation (-0.109).
- **Enrollment and Debt:** A slightly weak positive correlation (0.297) emerged between enrollment size and student debt.
- **Locale Disparities:** Notable disparities in mean debt levels were identified among different locales, with City and Town areas displaying similarities, while Rural and Suburban locales exhibited substantial differences.
- **State-Level Insights:** At the state level, wide ranges in average debt and first-generation student percentages were evident. Notably, Nevada ranked in the top three for having the highest mean debt and the highest mean percentage of first-generation students among states.

In summary, these findings emphasize the intricate nature of student debt among first-generation college students, providing substantial support for the initial hypothesis. They illuminate the substantial disparities in debt levels within this demographic. Overall, these findings accentuate the crucial need to address financial inequities among first-generation students to ensure equal access to educational opportunities.

References

This report utilized the accurate *CollegeScores4yr* dataset sourced from Lock5stat.com, harnessed combination and cross-reference of ChatGPT and Claude AI to develop and shape report questions, context, and supporting R script, and applied STAT 353 PowerPoint notes from Chapter 6, Descriptive Statistics, all of which ensure report reliability.