

# 36-315 Final Project

Nanditha Annapureddy, Josiah Smith, Karishma Kulshrestha

2024-04-30

## Contents

How does notoriety impact school status and applications? . . . . .	2
What impact does selectivity have on the application process? . . . . .	3
How does a university being public or private impact its application population? . . . . .	6

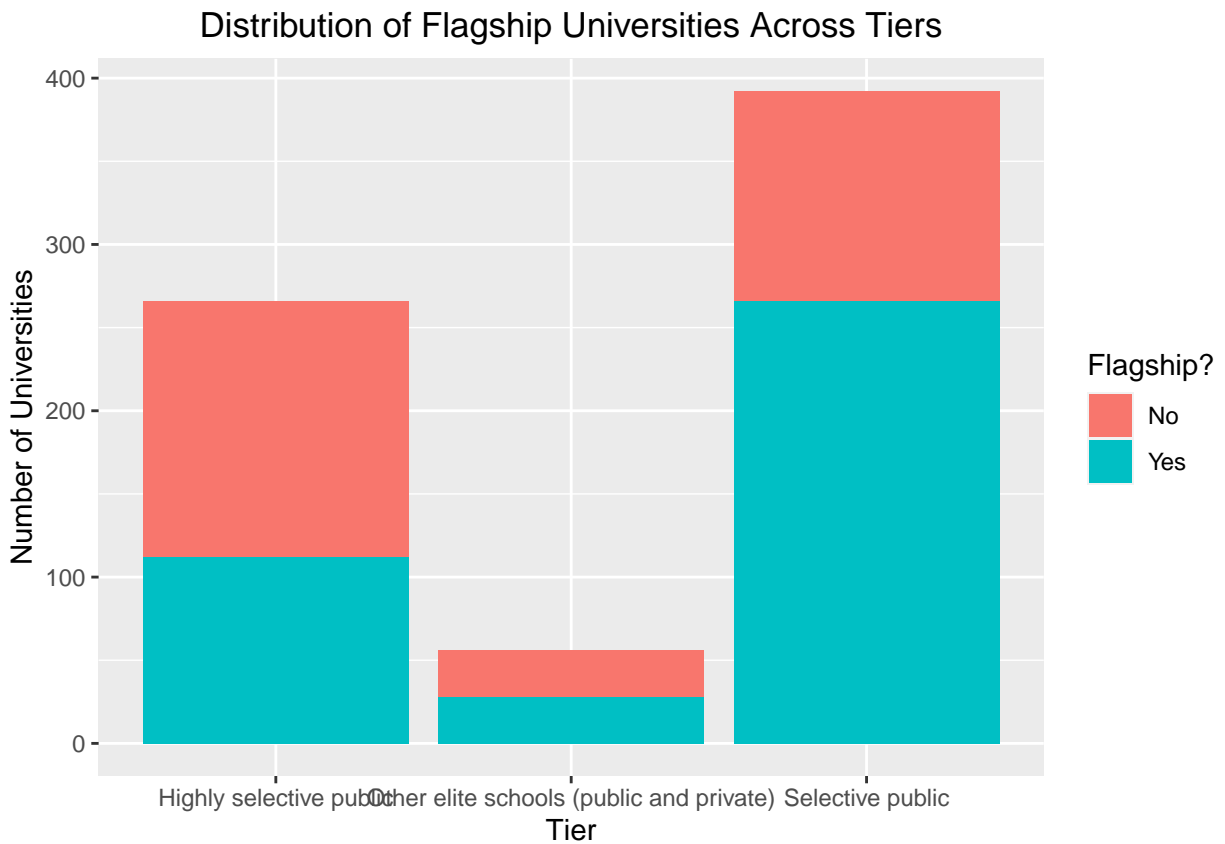
**Dataset** This dataset was uploaded to the CMU S&DS Data Repository on November 8th, 2023 by Jessica Zhiyu Guo. This data was collected to answer the question “How do admissions practices vary by institution, and are wealthy students overrepresented?” (Guo, 2023). This data explores educational equality by looking at 13 different income bins, 39 selective universities, and 2.4 million students entering the classes from 2010-2015. The variables in this dataset include “super\_opeid,” “name,” “par\_income\_bin,” “par\_income\_lab,” “rel\_apply,” “attend,” “rel\_attend,” “rel\_att\_cond\_app,” “rel\_apply\_sat,” “attend\_sat,” “rel\_att\_cond\_app\_sat,” “rel\_apply\_instate,” “rel\_apply\_oostate,” “tier,” “tier\_name,” “flagship,” “public,” and “test\_band\_tier.” The variable “super\_opeid” contains the institution ID, which comes from the Office of Postsecondary Education Identifiers, each with a 6 digit code, and “name” contains the name of the university. The variable “par\_income\_bin” discloses the parental household income group and “par\_income\_lab” shows the label, which are both based on percentile in the income distribution of all the households with children born during the same year. The variable “attend” illustrates the fraction of students attending the institution within all students in that income bin, and “rel\_attend” contains the value of “attend” divided by the overall mean attendance rate across all parental income bins for this school. The variable “rel\_apply” dictates the test-score-reweighted relative application rate. Further, for “rel\_att\_cond\_app” there is the ratio of “rel\_attend” to “rel\_apply.” Another variable for application rates is “rel\_apply\_sat” which shows the relative application rate for the specific test score band contingent on the schools tier. The variable “rel\_att\_cond\_app\_sat” has the relative attendance rate, relying on the application, for the specific test score band based on the tier of the institution. “Rel\_apply\_instate” provides the test-score-reweighted relative application rate for in-state students, where “rel\_apply\_oostate” has the same for out-of-state students. The variable “tier” demonstrate the category of this institution, where “tier\_name” dictates the name of the tier, where the options are “Ivy Plus,” other elite schools (public and private), “highly selective public/private,” and “selective public/private” (where public/private represents two categories). Additionally, the “public” variable acts as an indicator for public universities, where the “flagship” variable specifically identifies public flagship universities. Finally, the variable “test\_band\_tier” serves as the group of the institution for the test-score band statistics.

**Research Questions** The research questions that this report intends to answer are: 1. How does notoriety impact school status and applications? 2. What is the impact of selectivity on the application process? 3. How does a university being public or private impact its application population? Within the realm of notoriety, there is the analysis of the “flagship” variable within the public universities, looking specifically at tier and application rates. Further, in terms of selectivity there is the analysis of “tier” and application rates and for the analysis of public versus private universities the analysis is centered around the “public” variable and looks into income brackets, application rates, and relative acceptance rates.

## How does notoriety impact school status and applications?

For our first research question we want to analyze a school's brand or notoriety in relation to how it affects school status and applications, In our first graph we will analyze first how a flagship university is defined amongst different tiers of universities.

A flagship university can be defined as the most well known university in a given state (Fernandez, 2023).



In this graph, it is shown that across the highly selective and other elite schools, there is a fairly even split between flagship and non-flagship universities, whereas in the selective public tier it is shown to be composed mostly of flagship universities. This allows us to see the difference that is made by the prominence of these public institutions in terms of their tier, that is to say that there is a difference made because the proportions within the different groups do not all appear equal.

Within this data, we chose to make a separate dataset for only public universities when analyzing the flagship variable to maximize the validity of our analysis.

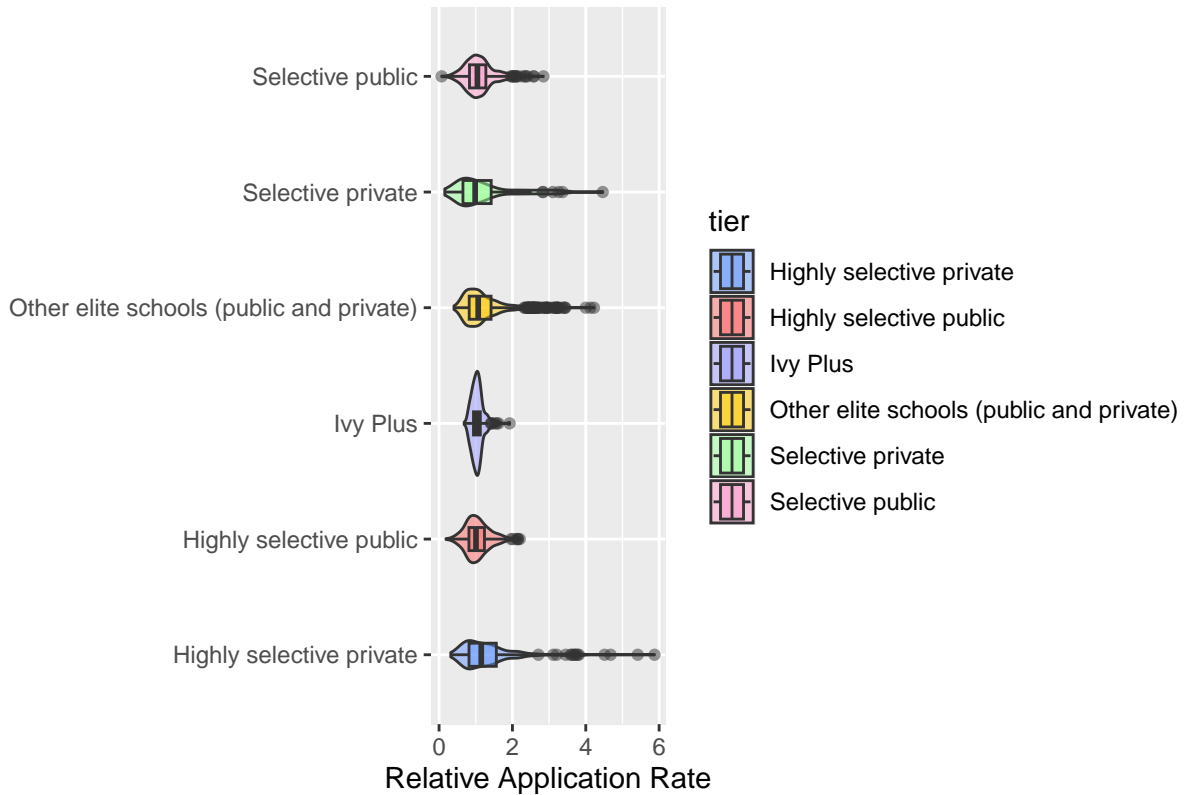
```
##  
## Welch Two Sample t-test  
##  
## data: rel_apply by flagship  
## t = -2.8229, df = 710.84, p-value = 0.004892  
## alternative hypothesis: true difference in means between group 0 and group 1 is not equal to 0  
## 95 percent confidence interval:  
## -0.13298866 -0.02388465  
## sample estimates:  
## mean in group 0 mean in group 1  
## 1.038135 1.116572
```

In this analysis, it is shown that the relative application rates across the flagship and non-flagship public

universities are not equal. Further, it is shown that the mean relative application rate for flagship universities is higher than that of non-flagship universities. This shows that the awareness of the university does impact the application rate, because the p-value within the t-test is .004892 which is less than 0.05, meaning that there is a less than 5% chance of this happening due to random chance if the two means were truly equal.

Lastly in our final graph for this particular research question we want to look at the relative application rate based on the tier of the university. This can help us gather information about whether a school's name or reputation impacts students decisions to apply.

### Relative Application Rate Based on Tier

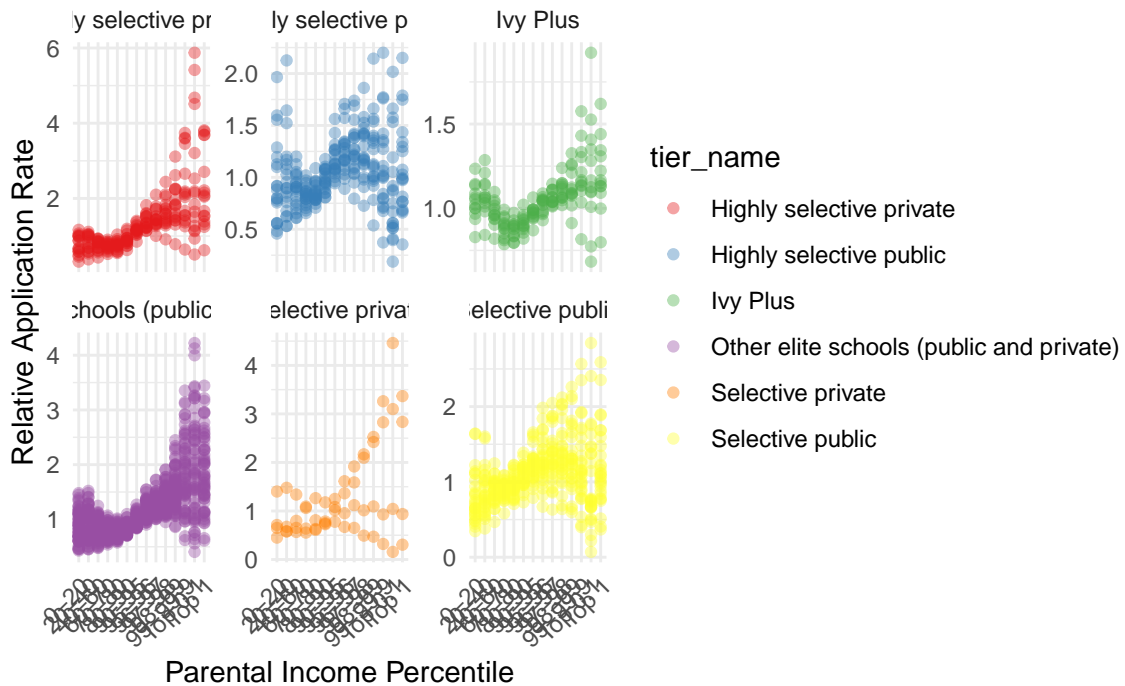


In this chart it is shown that the distributions of relative application rate based on tier have overlapping inter-quartile ranges, but the spreads of the data do not appear to be the same, however since the IQRs overlap, there is not a way to say definitively that these two variables have an interaction.

### What impact does selectivity have on the application process?

In this research question we want to analyze the numbers of selectivity and how it affects application decisions. We first look at how parental income numbers affect application rates for each tier of selectivity.

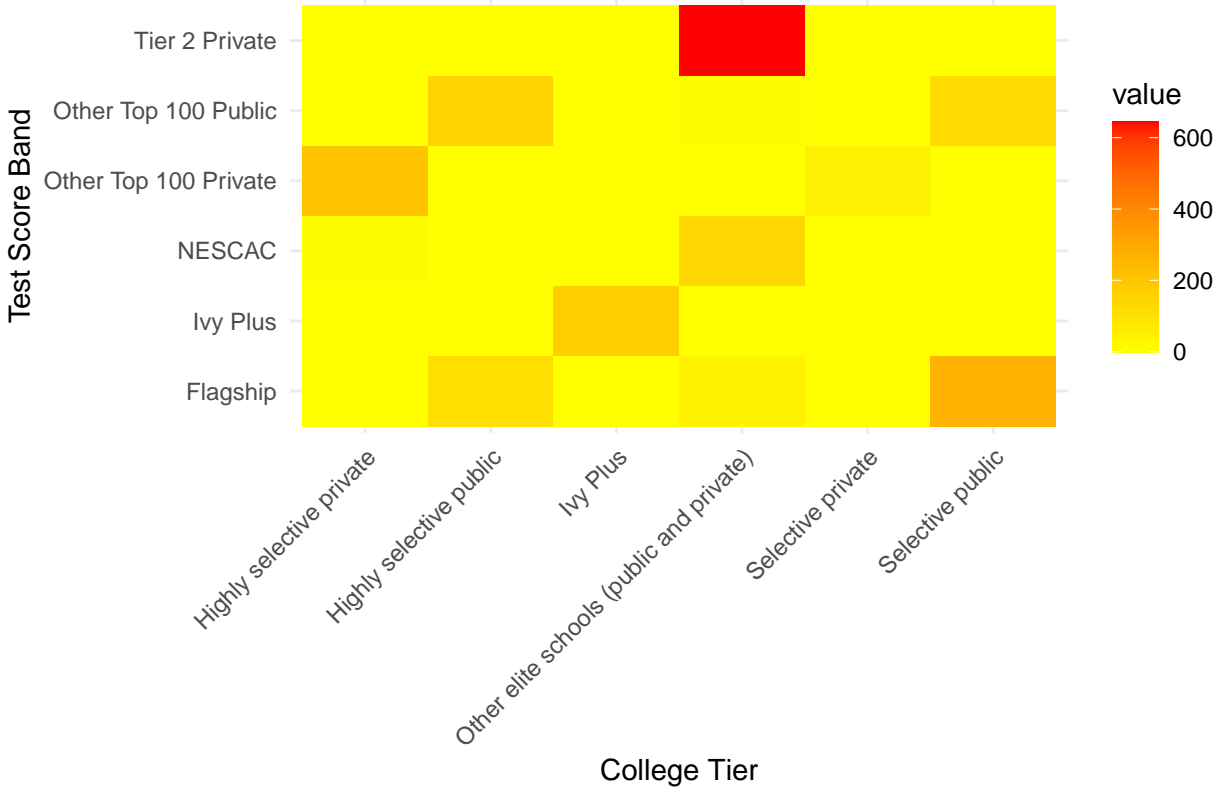
## Relative Application Rate vs. Parental Income Percentile



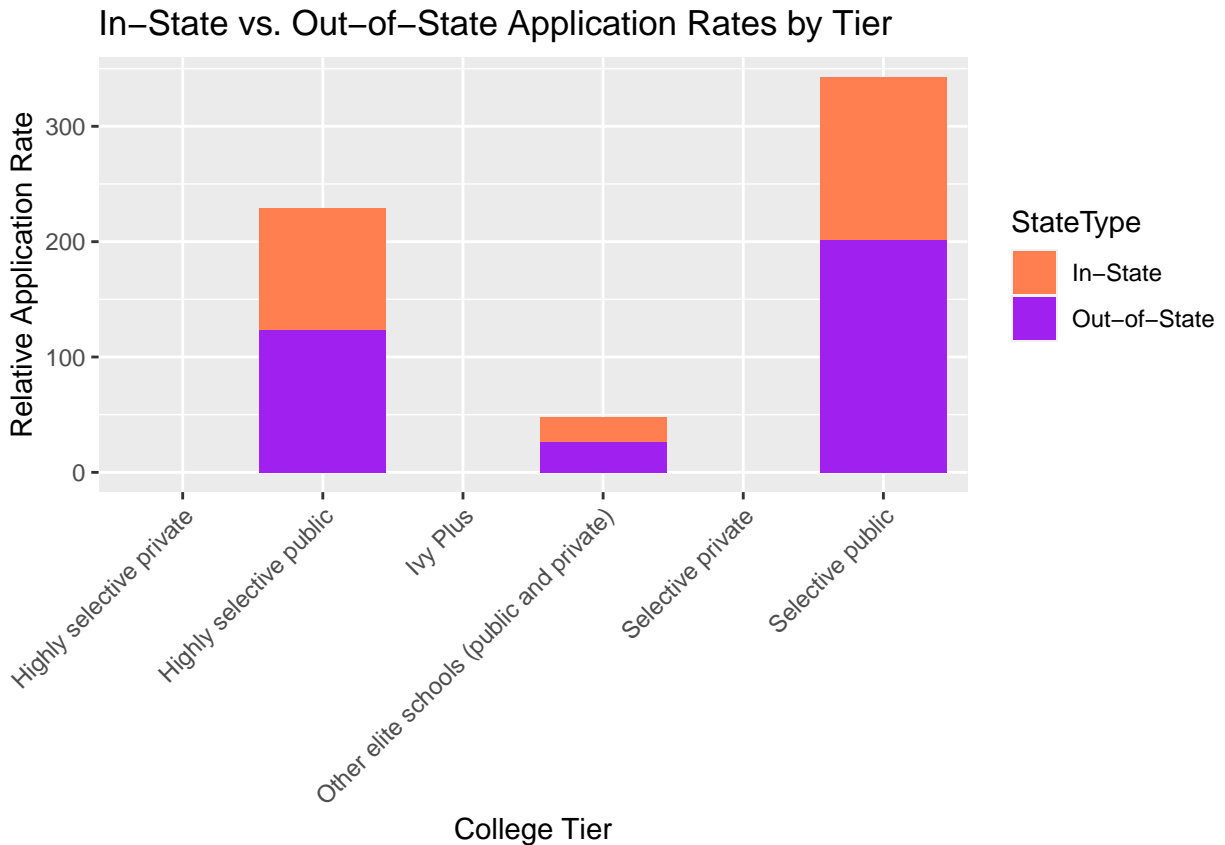
The scatter plot provided below shows the relationship of parental income percentile against the relative application rate to different tiers of colleges. For most college tiers, there appears to be a trend where the relative application rate increases with the parental income percentile. The pattern is most visible in the Ivy Plus category, where the data points rise more sharply at higher income levels. This implies that students who come from high-income families are the ones who apply the most to prestigious colleges. This implies that the economic background is a significant factor in the pursuit of prestigious higher education. When compared, the highly selective private colleges, along with their public counterparts, showed a dense cluster of their points at high echelons of parental income, along with a relatively higher application rate. This reflects perceptions that private institutions are often considered to be more prestigious and are therefore a preference for wealthy households. However, the graph also demonstrates variability within each tier, as evidenced by the spread of data points across the income percentiles. This broad dispersion of family income suggests that, while income is a factor, it is not the only determining factor for the rate of application.

In our second graph for this research question we will be looking at a heatmap of application rates by test score bands and college tier. This will help to answer questions about how test scores factor into a student's decision to apply based on each tier.

Heatmap of Application Rates by Test Score Band and College



In the bar chart below we will be looking at the relative application rates for in-state and out of state.

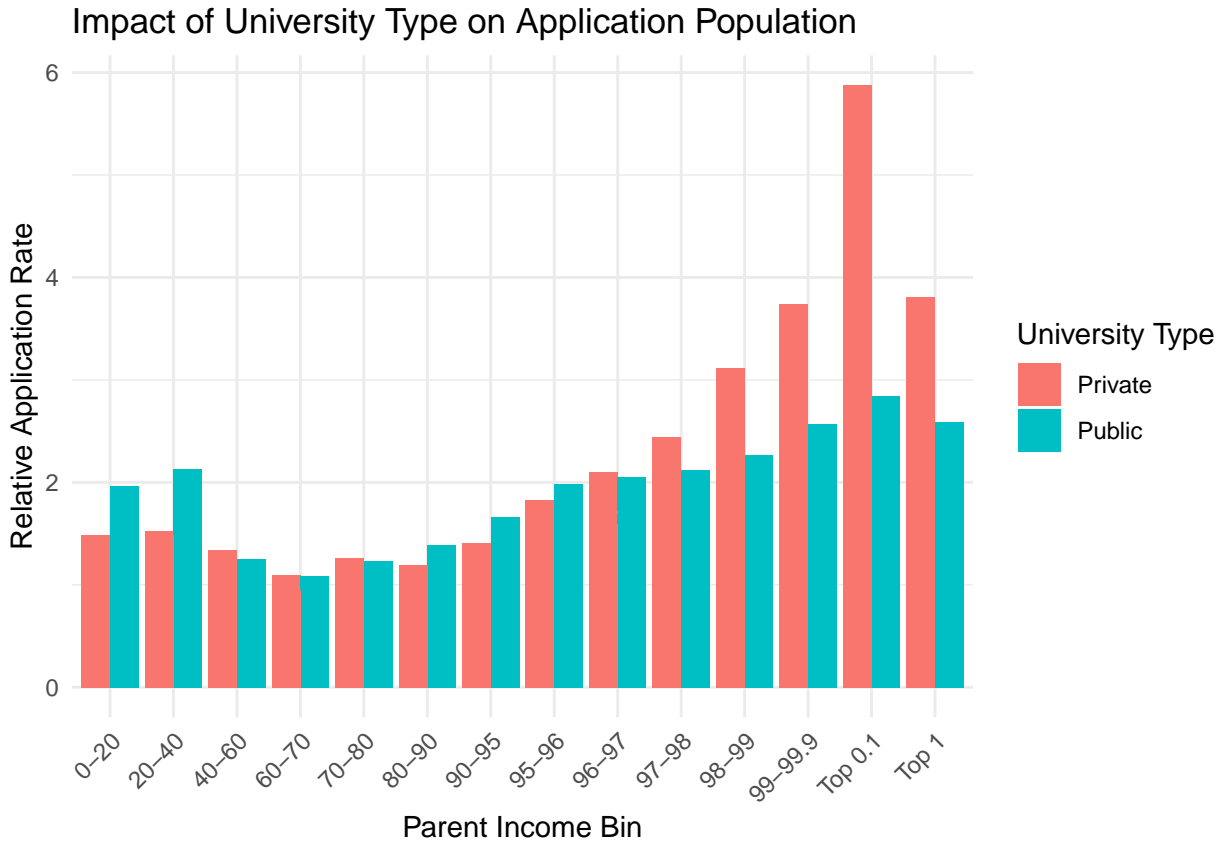


The stacked bar chart in question serves as a visual tool to explore the patterns of in-state versus out-of-state application rates to colleges of varying levels of prestige. A key finding is the close balance between in-state and out-of-state applications, with just a slightly higher rate of out-of-state applicants. This suggests that these colleges are nearly as attractive to non-local students as they are to local ones. The margin gets much slimmer for the highly selective public colleges, which show a comparable split between in-state and out-of-state applicants but where out-of-state applications are somewhat more common. Notable amongst these are highly selective private colleges, Ivy plus, and selective private colleges for which data is missing. There is no data for this category, which does not allow us to get a complete idea of the application pattern in this group. The visualization also highlights that the selective public colleges experience the highest volume of relative application rates, indicating their broad appeal. Overall, the chart suggests that the allure of college prestige extends beyond state lines, attracting applicants nationwide, though the extent of this appeal varies by the prestige and public or private status of the institution.

### How does a university being public or private impact its application population?

In our last research question we want to determine how a university's public or private status impacts its application population. This means that we should look at the `rel_apply`, `public`, and other associated variables.

In our first graph below we look at how relative application rates for public and private schools differs across the different parental income bins.



We notice that for the 0-40th percentile bins public universities have a far higher relative application rate. After that in bins from 40th to 96th percentile bins, no clear pattern is present for how private and public application rates are treated differently. Finally in bins from 96th to Top 1 percentile. We see that private universities overwhelmingly have higher relative application rates than public university. Some reasons could be theorized as to why we see more of a clear pattern towards the two ends of the chart. In state-public universities tend to offer a discounted tuition as well as potentially more financial aid, so it would follow that students whose parents fall into the 0 to 40th percentile would be likely to apply to these schools. On the other end of the spectrum students whose parents are in the 96th percentile and above likely do not have to worry much about potential tuition payments which eliminates a potential barrier entry present for other students. There might also be some cases of legacy admissions at the higher levels of parental income, however more data is needed to determine conclusively.

### Conclusion

Conclusion Our analysis of college application trends provides substantive insights into the roles of institutional type, prestige, geographical origin of applicants, and parental income in shaping application behaviors. Our findings are as follows: 1 Impact of Flagship Status: We demonstrated that flagship universities attract higher application rates compared to their non-flagship counterparts 2 Role of University Tier: The data indicates a strong correlation between the tier of a university and its application rates, which highlights the fact that prestigious tiers such as Ivy Plus are predominantly chosen by applicants from higher parental income brackets. 3 Public vs. Private Universities: Our analysis reveals distinct patterns where public universities are favored by lower-income applicants, likely due to financial considerations, whereas private universities appeal more to higher-income applicants. This pattern supports the hypothesis that economic factors significantly influence college choice, especially in the context of public versus private schooling. 4 Geographical Influences on Applications: The comparative analysis of in-state versus out-of-state application rates highlighted that prestigious institutions maintain a broad appeal, and that they attract a significant proportion of out-of-state applicants.

**Discussion Questions** In future research, experimenters could look into some questions that were not able

to be answered by this dataset, such as: what is the impact of legacy status on application and admission rates? Similarly, it would provide a meaningful comparison between public and private universities if there were values for “rel\_attend\_instate” as well as for out-of-state private institutions to see if there was a significance in terms of a difference between public and private with regard to in/out of state applications. Further, it would be compelling to have more data that expanded beyond elite universities to provide a baseline for comparison.