An analysis of student diversity and in-state tuition factors in United States post-secondary institutions

Jasmine Kwok, Trung Bui

June 12th, 2020

1 Abstract

In this paper, we examined different factors that contributes to the cost of in-state tuition across the United States post-secondary institutions and finding the representation of African-American students in those institutions with comparison to other ethnic groups. Our data provides us insight on the population of students and their respected ethnic group across different universities. Additionally, the data reveals the major expenses for attending colleges in and out of states and the world ranking for the examined institutions. Through using a combination of principle and exploratory data analysis, we came up with several findings that confirmed our initial belief. We found that type of universities and their room and board cost played a major role in determining the in-state tuition cost. We also discovered that the enrollment rate for Caucasian students are generally higher than African-American students. Consequently, we concluded that institution types and room and board cost played the most important role in determine the rate of in-state tuition as well as African-American students will generally have the lowest enrollment rate across different institutions

2 Introduction

The primary goal of this project is to explore and analyze factors that determines the cost of in-state tuition. In doing so, this project hopes to generate a model that can predict the in-state tuition cost for a four year degree institution. Additionally, we wanted to understand the composition of African-American students in different institutions in contrast with other major ethnic groups such as Asian, Caucasian and Hispanic.

The decision on whether or not to attend college has a lasting impact on an individual's life. Therefore, it is crucial to understand the key factors corresponding to the cost of tuition. This project is relevant for high school students, college students, high school counselors, and parents as they all play a role in either advising or decision making. attending college has become a crucial step for an individual's education and preparation for work in the future. In the United States, individuals of any age can attend college for bachelors, master's degree or more. Thus, this project may be applicable to any young adults and students in or out of the United States about the cost of attendance in colleges in the United States in 2018. Although this information may be inapplicable to the current tuition cost, it still provides a general idea of the key school expenses which may help with budgeting and pre-planning for financial aid.

Furthermore, implicit bias in education has been an ongoing debate for many decades as African-American students were often seen as the minorities in post-secondary institutions. For the majority of colleges, Caucasian has been seen as the most represented group in the student body with other ethnic groups trailing behind, followed by African-American student body which usually has the least amount of people. As a result, we wanted to explore and confirm how different ethnic group are represented in universities by analyzing the different type of race as a percentage of the total students residing at the institution. Thus, our analysis seeks to determine if there is an equal representation of African-American students in colleges as other ethnicity.

A study conducted by Allen and Wolniak (8) examines the implications of tuition increase at public 4 year institutions and racial compositions. Multivariate analyses were used to estimate the relationship between changes in tuition cost and racial diversity across 2 year and 4 year public colleges(Allen D. and Wolniak G., 18). The findings suggest that tuition increases at the both 2 year and 4 year year institutions are negatively associated with racial diversity of enrolled students. This helps further develop our question of interest to look at the relationship between in-state tuition and racial diversity in more depth. Based on this study, we would expect a negative relationship between in-state tuition and racial diversity since as colleges increase their tuition fees, there is a negative association with racial diversity of enrolled students.

Moreover, according to Kelchen R., Goldrick-Rab S., and Hosch B., Kelchen, R. (9), the living expenses including room, board, and other expenses account for more than half of the total cost of attending college (Kelchen et al. 947). Given this information, we would expect the ratio of room and board to be around half or more of either in-state or out-of state tuition cost. Their study examined the institutional variation and consistency in living cost allowances in comparison to the region estimated living cost from a report in 2013 (Kelchen et al. 947). The results suggest that larger cities had greater allowances in comparison to smaller cities and suburbs (Kelchen et al. 960). In context of our study, we would expect that universities

located in more urbanized states would require a higher cost of living, room and board, as well as tuition cost with the given information.

The relevant background needed from the readers to better comprehend the project would be someone who is interested in attending post-secondary institution or individuals that went to college. Our dataset is straightforward because it provides all the features and clearly articulates what each of them represents. Thus, for people with no prior college background could understand what each feature is explaining.

Our primary dataset describes the major possible university-related fees that students have to pay during a regular academic year at different institutions throughout the United States. Furthermore it categorized different institutions by their respective states while stating the type of degree offered and whether the institution is private, public or for profit. Additionally, our second dataset tells us the world ranking of each institutions while our third dataset describes the total amounts of student enrolled in a specified university and categorized them by their ethnic groups. These datasets gives us a variety of features such as the total enrollment, ranking and major expenses to examine and determine how in-state tuition cost is constructed. Furthermore, the third dataset gives us an insight as to how diverse a student group is for an university. This consequently allows us to see the representation of African-American students across different institutions.

3 Questions of Interest

1. What are some factors that contribute to the cost of in state college tuition in the United States?
   (a) What is a possible multi linear regression model that can predict in-state college tuition, and how accurate is it?
2. What is the relationship between the general composition of the student body, more specifically, African-Americans with other races and in-state college tuition?

4 Data Background

The tuition cost dataset, which is our primary dataset was obtained from Kaggle (1) and was collected by the College Board for its 2018 Annual Survey (2). The numerical figures collected in the dataset typically represent charges to full-time undergraduates for a nine-month academic year with a total of either 30 semester hours or 45 quarter hours. The inspiration behind this dataset is to help researchers explore the cost of college tuition in different geographical areas and by degree type to determine if there is a casual relationship between these attributes toward the college cost.

All three datasets are under MIT License, which states that the dataset is available for commercial use, modification, distribution, and private use (1). Both diversity and tuition cost dataset are cleaned and downloaded by Thomas Mock and placed in the TidyTuesday repository on March 10th, 2020. The student diversity dataset is obtained from the website The Chronicle of Higher Education and it includes race, ethnicity and gender of students at 4605 colleges in fall of 2014. A flaw in this dataset is that one person could only be counted in a single category. Hence, it does not capture the complexity of international students who could be from any race.

The datasets used in this study are relevant to our questions of interest as we would need information on tuition cost, cost of room and board, university ranking, and diversity in colleges. Precision of the data set which is measured by the closeness of our data points to the true value of cost on each college website is relevant to our questions of interest because it directly affects the accuracy of our analysis. In our primary dataset (tuition cost), random sampling was done to compare the values given on the institution website to the actual listed value on the dataset. As a result, we have confirmed that the existing values in the tuition cost dataset are reported precisely. There is no distortion caused by measurement in the tuition cost dataset as we believed that it is due to the tuition cost being publicly available and thus, it is challenging to modify the data and no incentive to falsify it. Since the data was collected using a survey method by the College Board (2), it is not costly since the main platform of measurement only involves the institution themselves answering survey questions.

Furthermore, in our secondary dataset, we would expect low distortion in figures for the diversity data set since the figures are collected from the Education Department's Integrated Post-secondary Education Data

System (4) and that the total number of enrollments are already on the institution website and thus, there is no incentive to falsify the data. In terms of precision, we have verified that the dataset is precise enough to be used for our analysis through random sampling and testing the number of enrollment. Additionally, the diversity dataset was collected as a mandatory survey from institutions that participate "in any federal student financial aid programs"(5). Thus by collecting data as surveys as the main measurement method, this is not likely to be costly.

Moreover, for our third dataset about world college rankings, we expected a low distortion in figures as the measurements were conducted by examining statistical figures that have been already accomplished (7). The methodology used was clearly stated out (6) where different factors such as the quality of education, alumni employment, quality of faculties and research performance are considered when determining ranking of universities. While there is no way to precisely measure the accuracy of ranking, Center for World University Ranking (CWUR) is known for its credibility as it is one of the largest global college ranking institutions and has experience in conducting rigorous academic ranking research for eight years. Hence, we would expect an unbiased and high level of precision from this dataset. Since the form of measurement includes active research on various areas of the examined institution while maintaining accurate data scraping and analysis, we expect this form of measurement to be costly.

Only universities that are associated with the College Board or took the College Board 2018 are represented in the data. This means that colleges that are not accredited by a national association recognized by the U.S Department of Education or do not offer associate and Bachelor's degrees are not represented in the data set since these are the conditions to be on College Board Search. Also, colleges that are outside of the United States are not represented in this data. From counting the colleges by state, we are able to determine that our dataset may over-represent colleges in California (254), New York (221), and Texas (150) and underrepresented colleges located in Alaska (6), Wyoming (8), and Delaware (9). There is also a possibility that these states consist of larger or smaller numbers of college institutions in comparison to other states in the United States.

One of the concerns that we have about analyzing the dataset would be a potential for students to shift their applications or transfer from their original university which may be more costly to a university who offers the same degree that is more affordable. This may bring harm to the participating universities that have a higher tuition cost as losing current and potential students leads to a loss in millions of dollars per year which may be detrimental to the finances of the college. Thus, this is a potential ethical issue that may arise through the analysis. To the viewer who is looking to enroll in a university, they might defer from attending the university in a certain state just simply by looking at the average tuition or room and board of the state. The information analyzed from this dataset is using data from the survey two years ago (2018) which is not up to the current date. This may be harmful to students and families as the information may be misleading for those who may base off their college and financial plans on the analysis. The total in-state and out-of-state cost here does not take into consideration the grants or other forms of student financial aid given to attending students. As a result, these amounts have a potential to lower the actual cost of attending an institution. Thus, for institutions that are represented in the dataset, it does not reflect the students' actual aid or grants given and thus, might portray cost of attendance and boarding in universities more expensive than it actually is.

The tuition cost data is composed of a total of 2973 rows and 10 column variables. There are 5 numerical variables which are cost of room and board, in state tuition, out of state tuition, in state total, and out of state total and 5 categorical variables: name, state, state code, type, and degree length. There are 50 unique variables for state and state code which correspond to the state names and codes in the United States. The total number of rows, 2973, which equates to the number of individual colleges in the United States included in this dataset. There are two different categories for degree length which is 4 year or 2 year degree and three categories for type of college which is public, private, and for-profit. Our dataset also consists of 1146 null values. One strange value in the dataset is the minimum cost of room and board which is 30 USD for Lassen Community College in California. We preprocessed the "Other" in their type of college column by using external resources to classify it into either a public, private or for-profit college. To further extend our analysis, we included two sets of supplementary data which are the world university rankings data set and student diversity data set, both of which are from the Kaggle website.

5 Methods of Analysis

In the beginning of this project, exploratory analysis was conducted using scatter-plot matrix (Figure 11) to identify interesting relationships between different numerical variables. There were no obvious non-linear relationships displayed by the scatter plots so no transformation was required for any existing variables. We also conducted principal component analysis (Figure 10), a dimensionality-reduction method on our main data set to understand variation explained by each principal component and find interesting patterns of how variables are correlated between institution in each state.

To answer our first question of interest, a combination of histogram and scatter plots were used to further explore interesting findings found from the scatter plot matrix and confirm whether the relationship holds. Depending on the type of variables, continuous, numerical, or categorical, various plots were used to better visualize and better understand our data. Based on the findings, a multi-linear regression model is constructed with the features as predictors for in-state college tuition. A mean-squared error function was utilized to calculate the accuracy of our regression model. In order to answer the second question of interest, multiple scatter plot diagrams were constructed to reflect the differences between the general composition of student body and in-state college tuition.

6 Results

6.1 Exploratory Analysis

There is a positive linear relationship between room and board with in state tuition and out of state tuition. This suggest that the cost of living at universities with higher tuition cost is more expensive. There is a negative relationship between room and board and ranking as well as out of state total and ranking. This indicates that the lower the world ranking (larger number) of an institution, the cheaper the room and board costs and out of state total tuition costs. Some unique plots include in state tuition and ranking whether there are two distinguishable parallel decreasing relationships between in state total and ranking. Another interesting finding is two non parallel positive linear relationship which are present between in state tuition and out of state tuition. One positive linear relationship forms an x=y line which suggest that the price of in state tuition is equivalent to out of state tuition. The other positive linear relationship has a steeper slope which indicates that the price of out of state tuition increase by a greater amount in comparison to an increase in in-state tuition.

Furthermore, based on our principle components analysis plot (Figure 10), we found that the first principal component captures the relationship of room and board cost across different states. States such as South Dakota and Missouri, which are on the far right, have their average college cost for room and board to be $9061 and $8734 respectively. However, states on the far left like Tennessee and Maryland, the average room and board cost is $16234 and $15836 respectively. Additionally, states located around the origin such as Oklahoma and Delaware have their average cost for room and board to be $11116 and $12862 respectively. Therefore, from the figures in our data and our interpretation, we concluded that the first principle component distinguish different states by comparing their average room and board cost‘

Moreover, the second principle component seems to be differentiating states by their total enrollment where states such as Oklahoma is to the far left because its average college enrollment is very minimal, 4682 students. On the other hand, Arizona has the largest average total enrollment of 46278 students, hence making it located on the very far right. Furthermore, for states located around the origin like New York, their average total enrollment is between the Oklahoma and Arizona total enrollment range which is 20,073 students on average. Thus, based on our data, we interpreted that the second principle component is differentiating different states by their varying number of enrollment.

6.2 Multiple linear regression

From the scatter plot matrix, five predictor variables: total enrollment, ranking, room and board cost, type of institution, and degree length offered by institution were chosen to be explored further. Figure 8 shows a negative correlation between in-state tuition and ranking. This illustrates that as school ranking increases (lower value), the cost of in-state tuition decreases. A regression line was created to confirm the

inverse relationship between the variables. From the graph, we discovered that the two distinct decreasing trend were due to the type of institution as private institutions have a higher average in-state tuition cost than public institutions. The apparent negative linear relationship suggest that school ranking is a possible predictor to determine the cost of in-state tuition.

The graph of in-state tuition and total enrollment (Figure 9) depicts no obvious patterns between the variables. Hence, there is no correlation between total enrollment and in-state tuition cost. The points are mostly cluttered towards lower total enrollment, mostly below 20,000 students, regardless of in-state tuition cost. This indicates that there is no linear relationship between in-state tuition and total enrollment.

Figure 6 depicts a significant difference between the in-state tuition cost and the length of the degree offered at different universities. The graph suggests that 2 year degree programs are more affordable than 4 year programs in which most of the 2 year degrees only cost around $\$3,000$ to $\$6,000$. 4 year degree programs, however, are more spread out with most in-state tuition cost greater than $\$16,000$. This suggest that the type of degree length is an important factor in determining the cost of in-state tuition.

A histogram was generated to determine the relationship between the type of institution, private or public, and in-state tuition cost (Figure 7). The distribution for public institution is right skewed while private institutions follow a normal distribution. This illustrates that the majority of public institutions have a low cost of tuition price around the lower end of the range $\$1000$ to $\$21,000$ with most public institutions priced about $\$4,000$ to $\$6,000$. The price range for private institutions are more spread out with a range from around $\$1,000$ to $\$59,000$. Most private institutions, however, have an in-state tuition cost priced around $\$30,000$ - $\$32,000$. This indicates that on average private institutions will most likely cost 6 times more than a normal public institution, and more importantly, the type of institution does contribute to determine the in-state tuition cost.

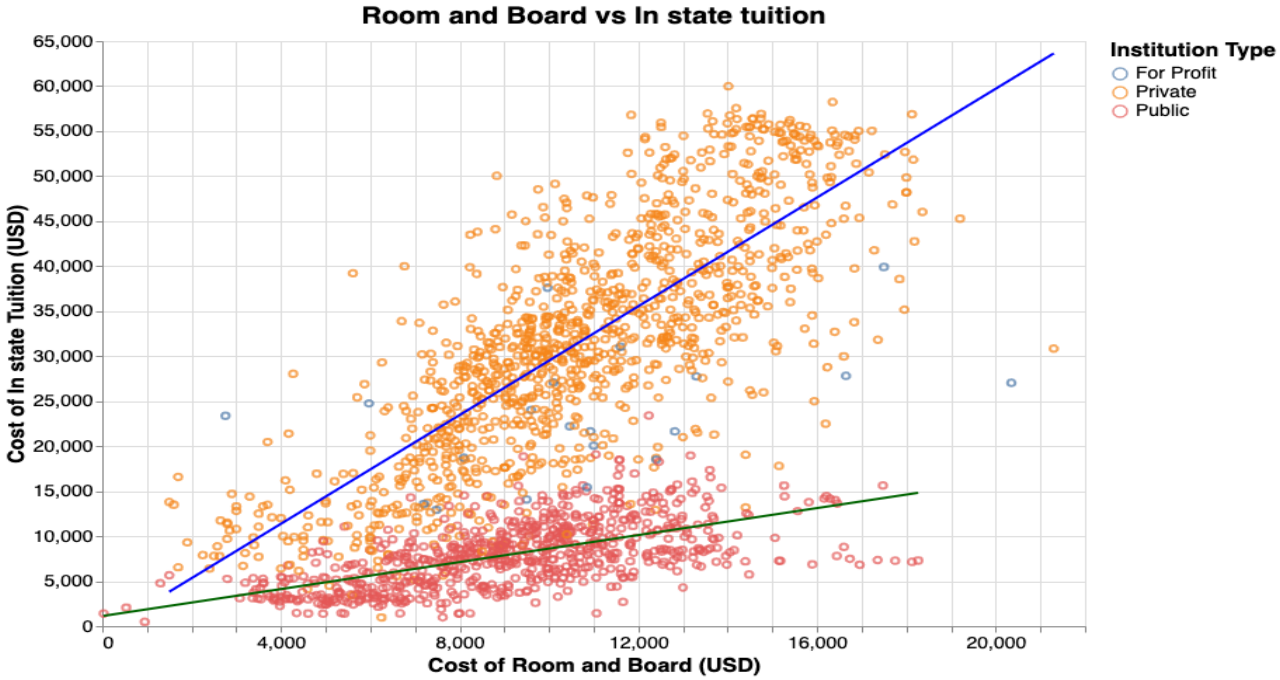

Figure 1: Private institution sees the sharpest rate of increase between cost of room and board to in-state tuition

The scatter plot diagram between in-state tuition and room and board cost (Figure 1) depicts the relationship between the cost of living at the school and its tuition cost. Figure 1 reflects two distinct positive correlation between our two variables. Thus, this indicates that a higher cost of room and board , the higher the cost of in-state tuition. The diagram above color-coded the points belonging to distinct institution types: for profit, private, and public, with the colors blue, orange, and red. Due to the lack of data on for profit institutions, there is no clear distinguishable pattern between the two variables. However,

the Altair fitted linear regression lines reflect a steeper slope for private institutions in comparison to public institutions. This suggest that private institutions are generally more expensive and have a much higher increase in cost of in-state tuition for every unit increase in cost of room and board.

Through utilizing the scatter plots and histogram, we are able to determine that four out of the five features have a relationship with in-state tuition. With that, a sci-kit learn linear regression model was constructed to predict mean in-state tuition cost for institutions. The correlation coefficient for type of institution, degree length, room and board, ranking, and total enrollment is 35087.47, 1.12, 1.01, -5.48, and 0.04. The mean difference of in-state tuition for private institution is $35087.47 more than public institution when all other predictors held constant. The in-state tuition cost for 4 year degree length costs $1.12 more on average than a 2 year degree length when all other predictors held constant. For one unit increase in room and board cost, the cost of in state tuition increases by $1.01 with all other predictors held constant. For one unit increase in university world ranking, which means the university is of a lower rank, the cost of in state tuition decreases by $5.48 with all other predictors held constant. The correlation coefficient for total enrollment is 0.04 which is close to 0. This confirms our previous interpretation that there is no linear relationship between total enrollment and in-state tuition cost. The mean in-state tuition for public institutions when the value of all predictors are 0 is $3115.22. Using the regression model, the mean in-state tuition cost for all institutions in the United States using mean values for continuous columns and median values for categorical columns is $13631.67. The mean squared error loss calculated for our model is 10165433.9 which is significantly large. This indicates that our regression model is inaccurate and our predictors are does not predict and explain the variability of in-state tuition well.

6.3 Diversity and In-state tuition cost

To understand the composition of African-American students and in-state college tuition, a new data frame which combines diversity data and in-state tuition was created. A new column which is the percentage of total enrollment was made by performing calculations on existing columns to take into account the differences among enrollment numbers in different schools. This is more accurate in comparison to sheer count of total enrollment because larger institutions may have more enrollment of a particular student race than smaller institutions but that may not represent a high composition due to the large amounts of total enrollment. Hence, percentages of total enrollments are more statistically significant than the enrollment counts.

A scatter plot between average in-state tuition cost vs African-American students enrollment percentage with different types of institution was created to portray the composition of African-American student body and in-state tuition. The graph reflects a negative relationship between in-state tuition cost and African-American student enrollment percentage. This means that the higher the cost of in-state tuition, the lower the percentage of African-American students in the institution. There is a clear difference in the variability of in-state tuition in which the private institution is more expensive and has a wider tuition range of 5,000 USD to 65,000 USD in comparison to public institutions which only ranges between 1,000 USD to around 15,000USD. In general, institutions with tuition costs less than 30,000 USD have a significant range of African-American student enrollment percentage from 0% to 100%. Private institutions with in-state tuition cost greater than 45,000 USD have a much lower African-American student enrollment percentage ranging from 0% to 15%. (Figure 5)

By plotting other student races such as Asian, Hispanic, and Caucasian, we are able to compare their respective student enrollment percentage at institutions with different in state tuition costs to African-American student composition. An interesting finding is the percentage of enrollment between Asian and African American students. In general, African American students have a higher enrollment percentage at institutions with in-state tuition from 1,000 USD to around 40,000 USD (Figure 3) . For colleges with higher in-state tuition costs, however, there are more institutions with a higher percentage of Asian students enrollment, some with almost double or more in comparison to African American students. The range of Asian student enrollment in more expensive institutions is around 2% to 35%.

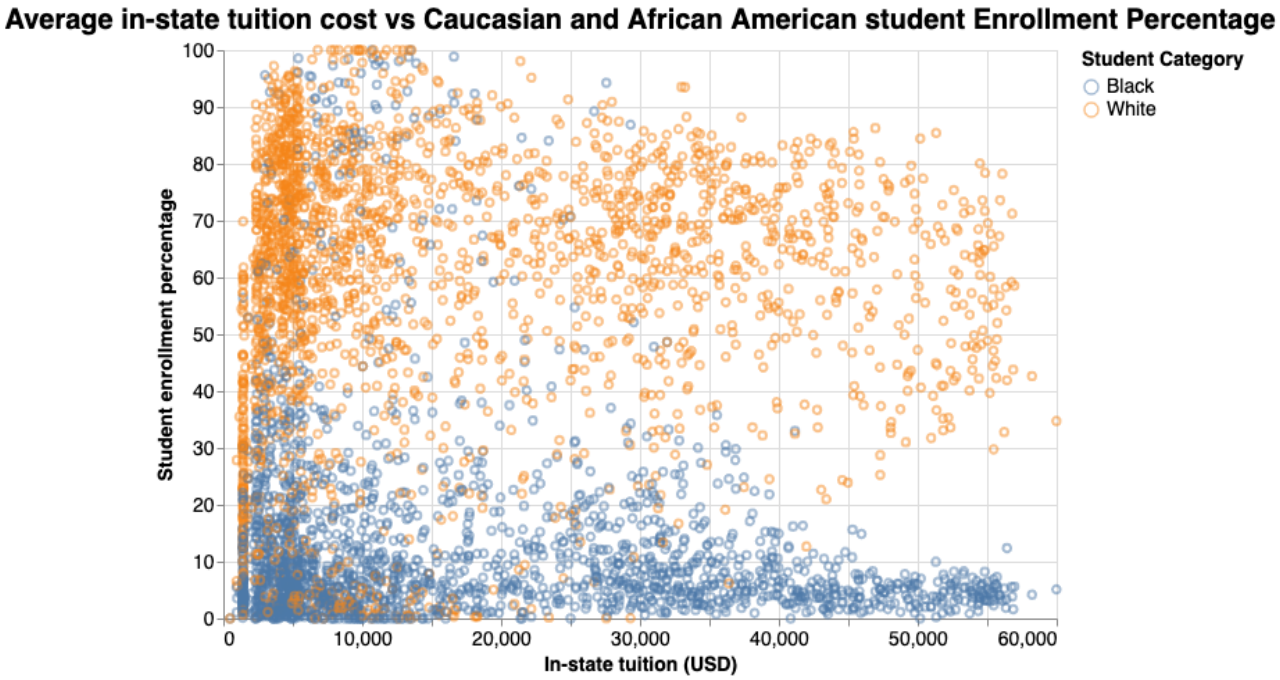

Figure 2: The enrollment percentage for African-American are very low compared to Caucasian students.

Another interesting finding is the percentage of enrollment between Caucasian and African American students. Regardless of in-state tuition amount, the average percentage of Caucasian student enrollment is much higher than the average percentage of African American students (Figure 2). The student enrollment percentage for Caucasian students is heavily concentrated around 40% to 90% while the student enrollment for African American students is concentrated around 0% to 10%. We can infer that there is a much lower number of African American students who attend college in general compared to Caucasian students. Furthermore, regardless of the cost of in-state tuition fee, we would expect the average institution to have a higher number of Caucasian students (30% or more).

In contrast to the comparisons with other races, the scatter plot for Hispanic students is very similar to African American students (Figure 4) and there is a large amount of overlap between the points. There is the same negative decreasing trend between Hispanic student enrollment percentage and in-state tuition. We can conclude that the higher the price of in-state tuition, the lower the percentage of enrollment for both Hispanic and African American students. Overall, the Hispanic student enrollment percentage is concentrated in the range of 0 to 10% which is similar to African American students enrollment percentage. A minor difference present in the graph is for higher in-state tuition institutions, there is an evident higher amount of Hispanic student enrollment by 2% to 5% in comparison to African American students enrollment percentage. Another notable difference is the greater variability and a higher average of African American student enrollment percentage for institutions with in-state tuition ranging around 10,000 USD to 30,000 USD. This suggests that for more costly institutions, we would expect a higher number of Hispanic students compared to African American students while the reverse is expected for middle priced institutions which is around 10,000 USD to 30,000 USD.

7 Discussion

Throughout our project, we encountered several challenges. The process of combining different datasets together were very challenging because there are different institutions that are in one dataset but not in the other. As a result, we have empty values inside our combined dataset, which makes it inefficient when analyzing our question of interest. Empty values are removed during the cleaning process which results in a significant loss of data. A lesson that we learned from this is to be selective of which dataset to combine. Instead of adding world rankings to our original dataset, a better option would be to add United States

universities ranking which better matches our tuition cost dataset. One problem in visualization is the issue of having too many data points. Since Altair can only plot 5000 data points at a time, we could not include every racial group in a single graph. This makes visualizing the racial group in comparison with the other 3 racial groups much more challenging and inefficient since we have to plot the comparison one by one.

From our analysis, we found some possible factors that could influence the cost of in-state tuition, we found that four out of the five analyzed features have a strong correlation with tuition in state. These features are the length of degree offered, type of institution, room and board cost and ranking. This results reflects our understanding of colleges, since private institutions are not federally funded, their in-state tuition would cost more than public institutions, since they are funded by the states and taxpayers. Furthermore, the negative correlation between ranking and in-state tuition tells us that competitive schools would generally be more expensive because of the higher quality of education, world-renowned faculties and the opportunity to network with other talented individuals. There are generally more opportunities to be successful at a more competitive institution rather than an average one and thus, this is reflected in higher tuition rates. Moreover, room and board cost is a good indicator of how expensive the cost of living in the institution area will be and consequently influenced the cost for tuition. Urbanized areas such as New York has a higher cost of living, which corresponds to higher in-state tuition rates compared to suburban states such as Florida or Texas where it is generally cheaper to live. Moreover, the rate of tuition between a 4 year and 2 year degree differs drastically, where a 2 year degree will most likely be cheaper. Since 2 year degrees are offered at community colleges where the quality of education is less advanced compared to the 4 year degree curriculum offered at colleges and universities. As a result, the cost of tuition for an associate degree will generally be cheaper than a bachelor degree. Although we have achieved our goal in finding the possible factors to predict in-state tuition, the linear regression model has a large mean squared error which illustrates its inaccuracy to explain the variability of our response. Thus, an extension to this study would be to improve our model by including more predictors such as the institution diversity, the quality of faculties, university geographical area and the median income of the staff employed there. This would allow us to build a better model with lower mean squared error corresponding to a more accurate regression model. Consequently, these addition features will allow more correlation to be discovered and give us a better understanding of the possible factors that affect in-state tuition.

Through the use of visualizations, we are able to understand the relationship between the general composition of African American student body with other races and in-state college tuition. There is an inverse relationship between African American student enrollment and in-state college tuition which suggests that colleges with high tuition rates would have a smaller composition of African American student bodies. We found that Hispanic student enrollment percentage is very similar to African Americans. In comparison to African American student enrollment percentage, Asians have a lower enrollment percentage for colleges with lower in-state tuition cost and a higher enrollment percentage for colleges with higher tuition cost. There is also a distinct difference in enrollment percentage for African American students and Caucasian students. Regardless of the in-state tuition costs, Caucasian students, in general, have at least 30% or higher enrollment percentage compared to African American students. A possible explanation for the low enrollment percentage for African American students may be the enrollment quota that is placed on this racial group being much lower than other racial groups in more expensive institutions reflecting systemic racism in the institutions. An explanation for other racial groups with higher percentages of enrollment for institutions with higher in-state tuition cost compared to African American students may be due to affordability. There may be more African American families who are living in poverty or with lower incomes compared to other racial groups, hence, are unable to afford to send their children to more expensive institutions. Furthermore, institution with higher in-state tuition may have a smaller number of African American student applications and thus, there will be a smaller application pool leading to a lower chance of applications who meet the admission criteria. We are able to conclude that universities with higher in-state tuition tend to accept a larger number of Caucasian students and a much smaller number of Asian, Hispanic, and African American students. A possible extension would be to collect another diversity data from the same universities in 2020 and compare if there are improvements and changes in the enrollment percentages since our current data is collected in 2014. In doing so, it is possible to determine whether human rights movements in recent years such as the Black Lives Matter movement have an impact on the embedded inequality within the educational system.

A Appendix

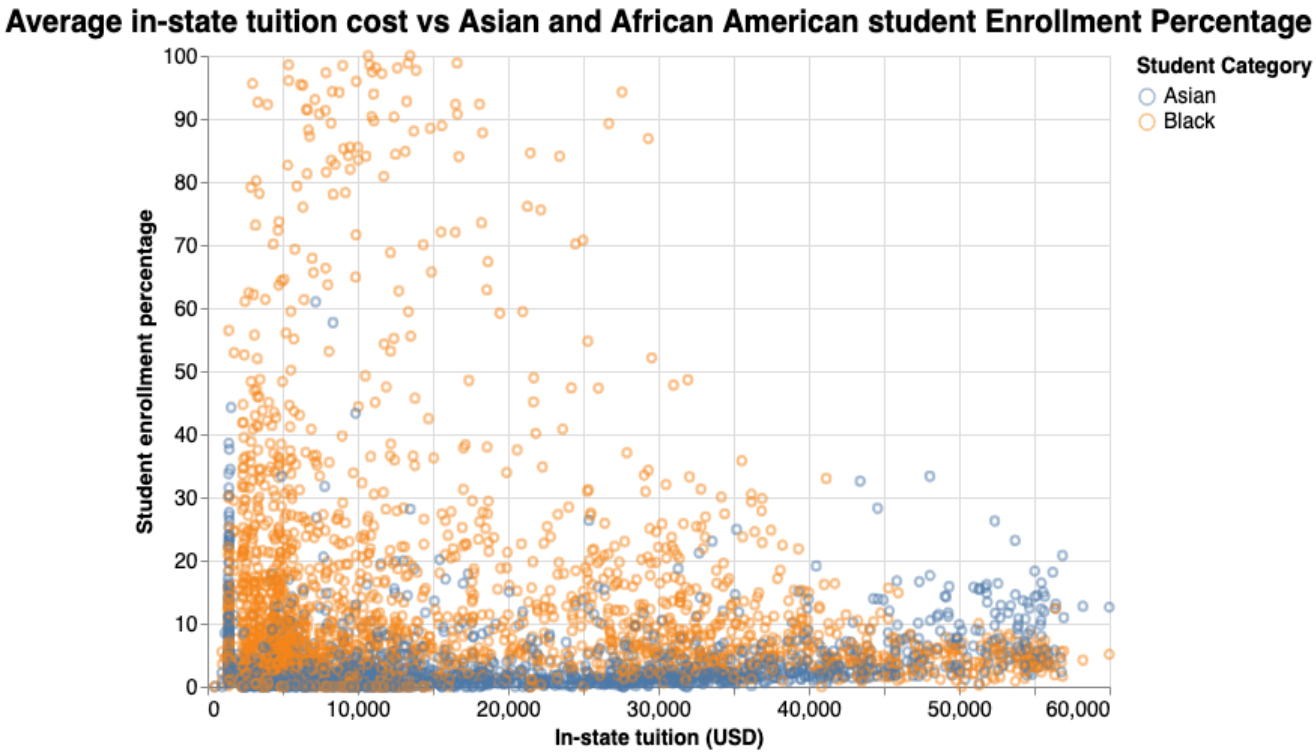

Figure 3: Median Asian and African-American population on campus are 1.77% and 7.16% respectively

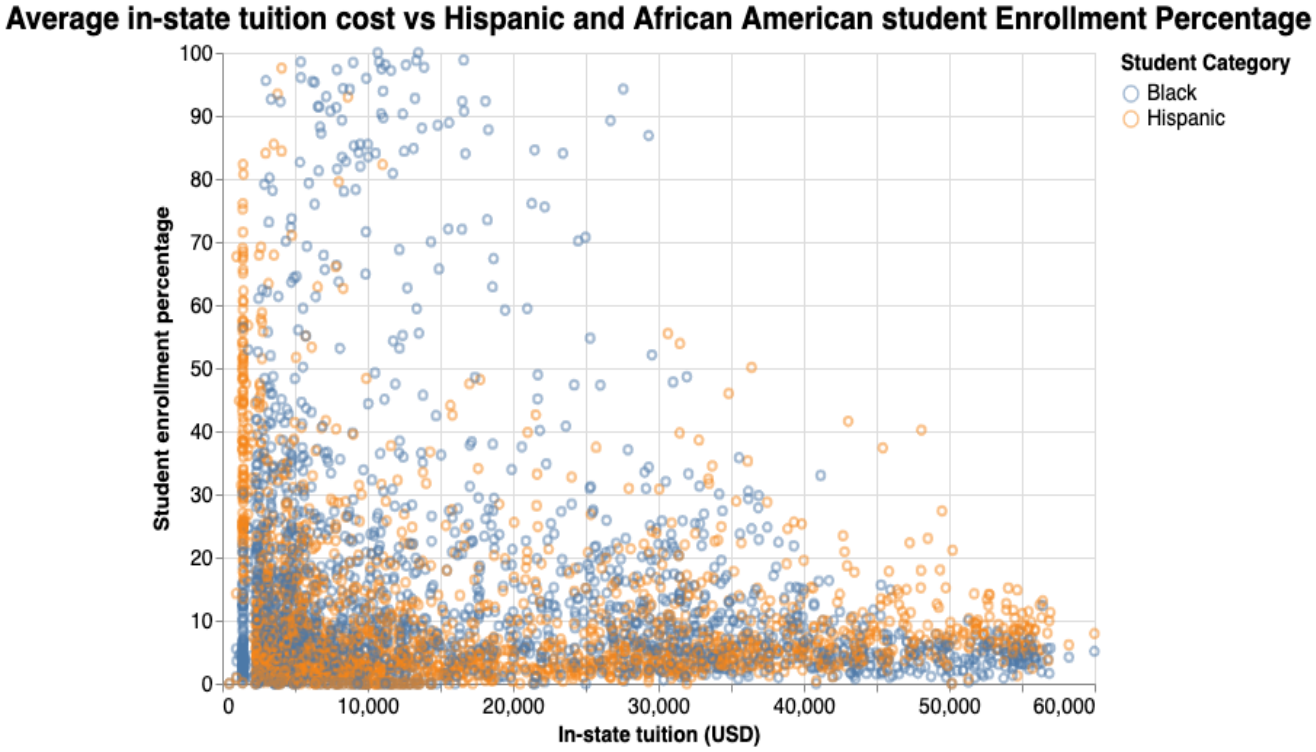

Figure 4: African-American and Hispanic have similar population distribution on college campus.

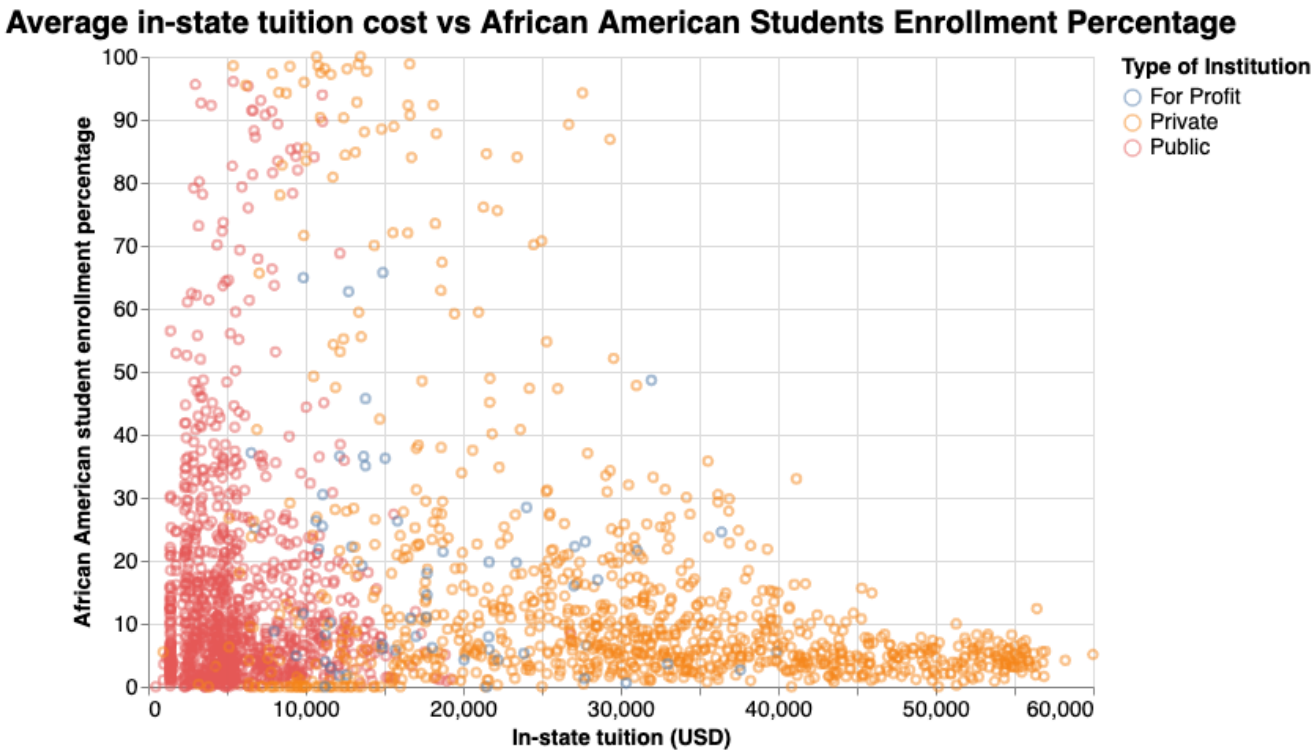

Figure 5: The majority of African-American chooses low-cost tuition, public school

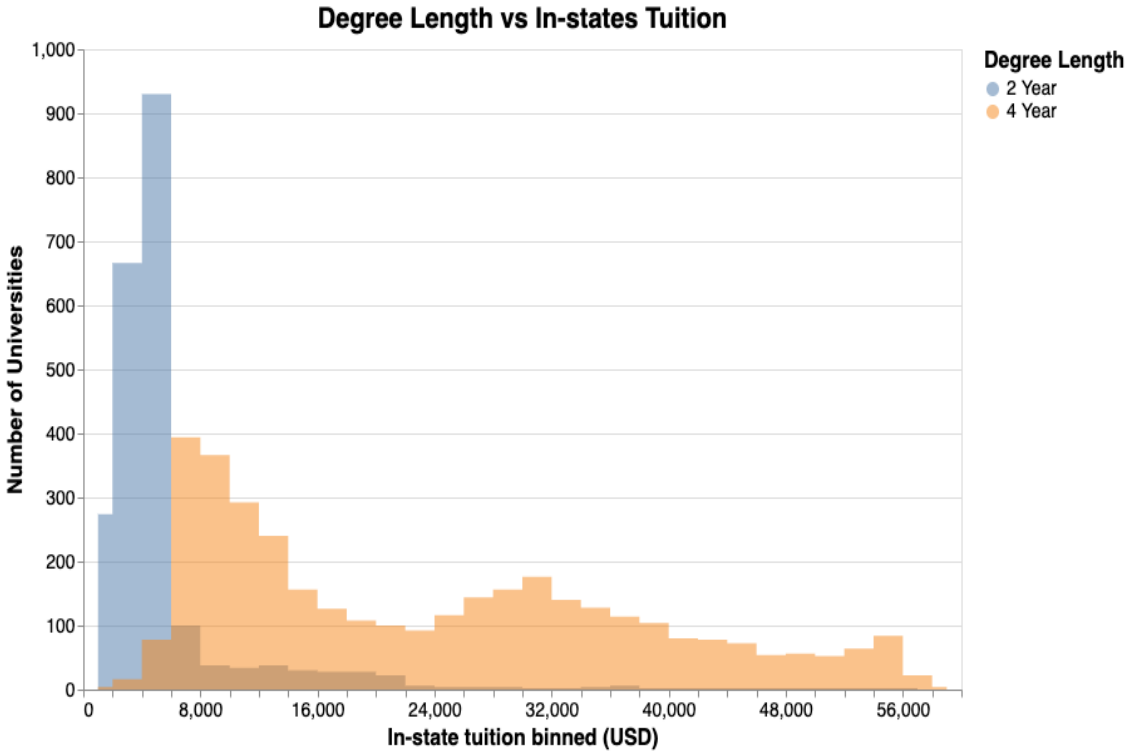

Figure 6: 2 year degree are most likely to cheaper compared to 4 year degree.

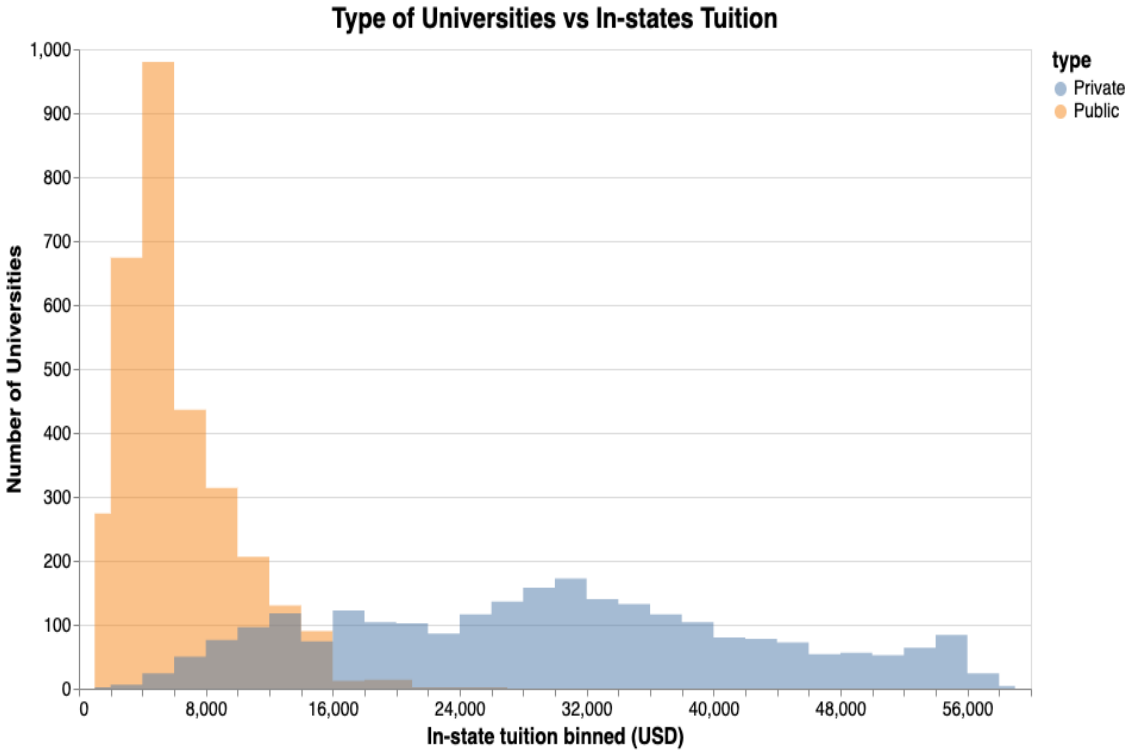

Figure 7: Public universities will tend to be cheaper compared to private universities

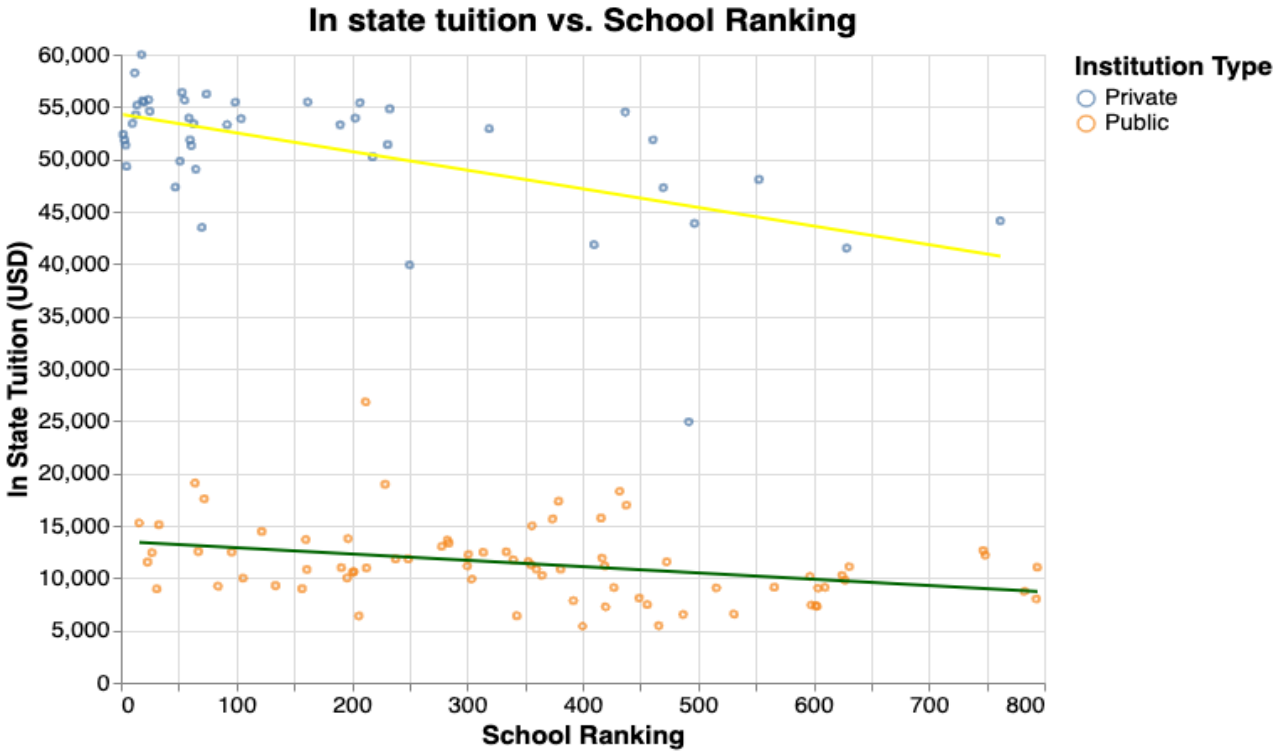

Figure 8: There is a sharper rate of decrease between school ranking and in-state tuition for private institution compared to public institution

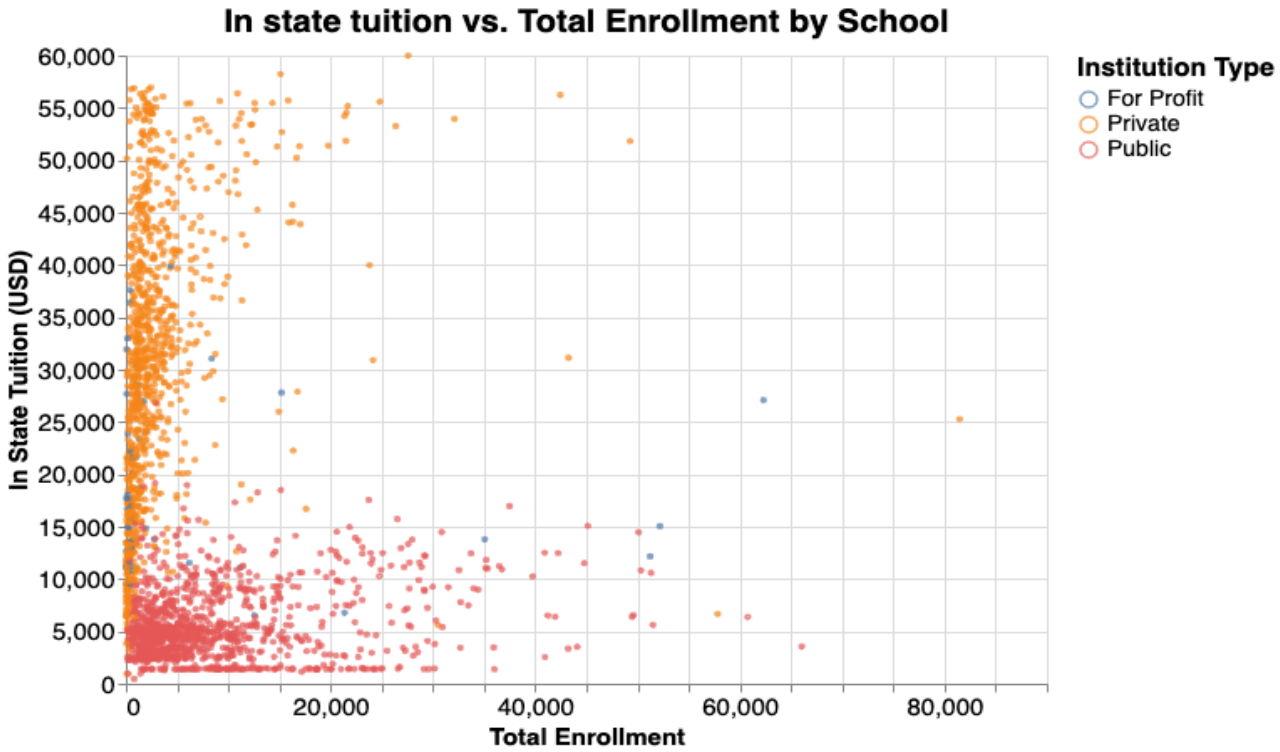

Figure 9: Little to none correlation between total enrollment and in-state tuition depicted by the vertical line around y axis.

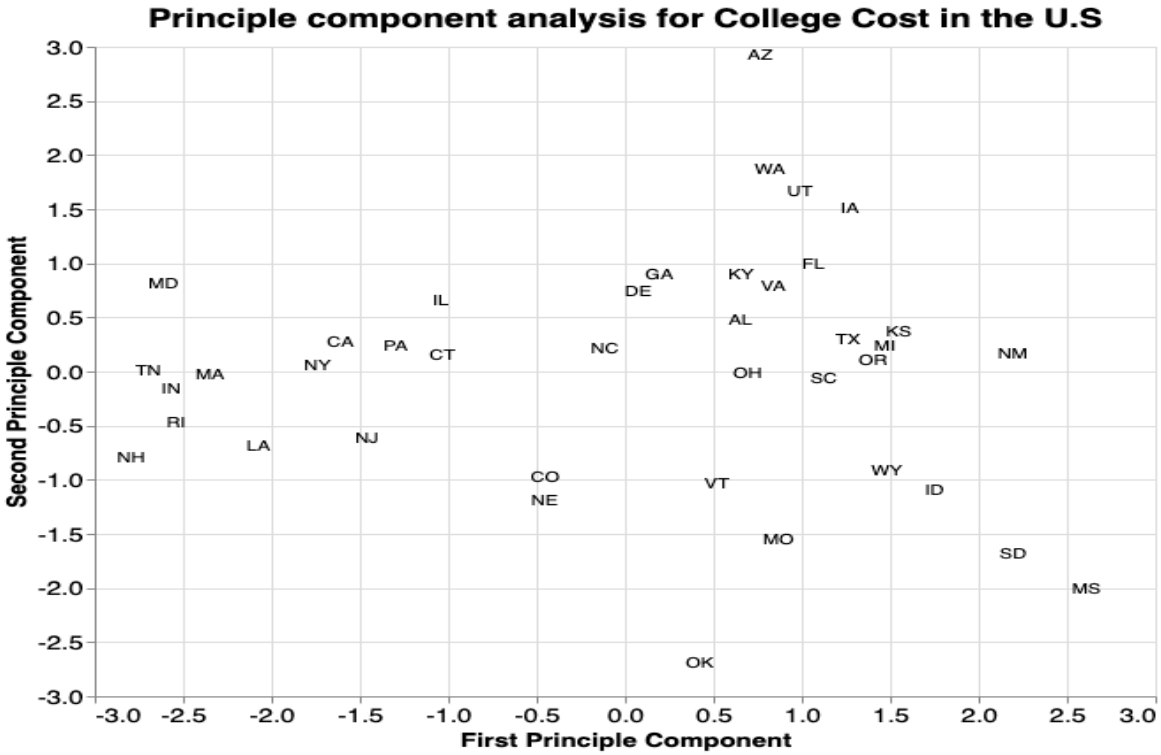

Figure 10: 1st principle component around the origin shows states with average room and board of $11,000

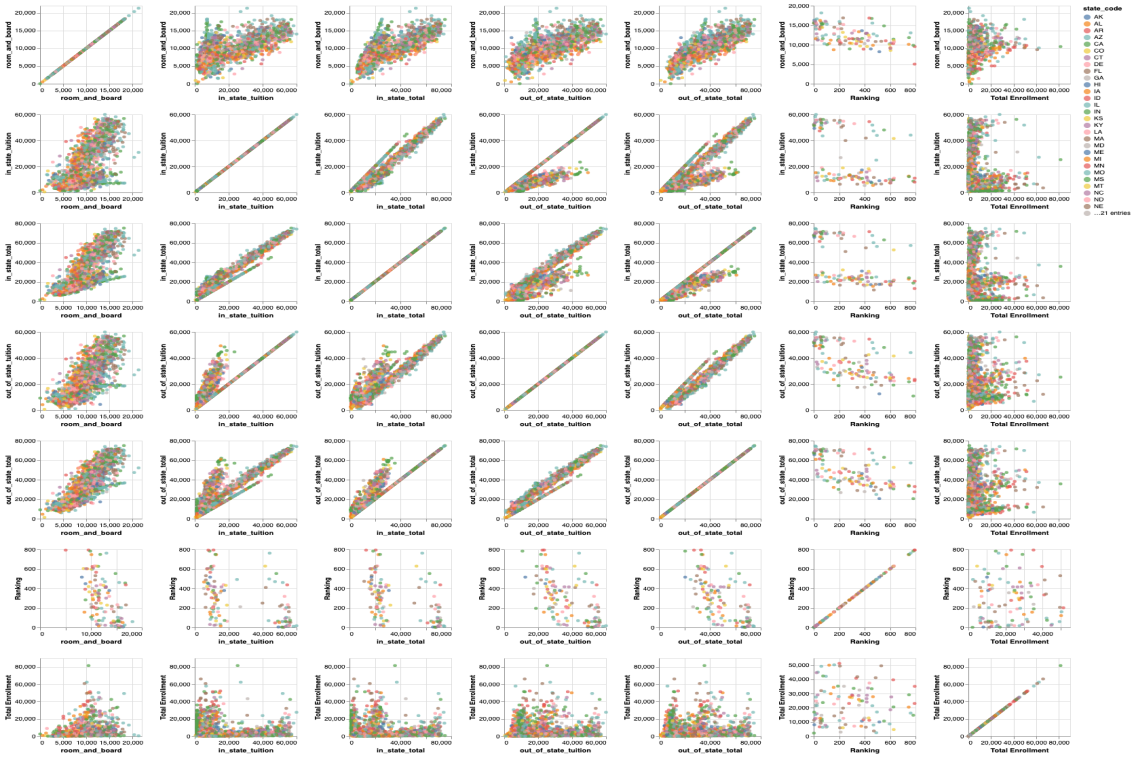

Figure 11: Diagonal plots are linear as the same variable are plot against each other

B Citations

1. Tuition Cost Dataset - https://www.kaggle.com/jessemostipak/college-tuition-diversity-and-pay
2. Tuition Cost context - https://www.chronicle.com/interactives/tuition-and-fees
3. Diversity School Dataset - https://www.kaggle.com/jessemostipak/college-tuition-diversity-and-pay
4. Diversity School Dataset Context - https://www.chronicle.com/interactives/student-diversity-2016
5. Diversity School Dataset Context 2 - https://nces.ed.gov/ipeds/about-ipeds
6. World University Ranking Dataset - https://www.kaggle.com/mylesoneill/world-university-rankings
7. World University Ranking Methodology - https://cwur.org/methodology/world-university-rankings.php
8. Allen, D., and Wolniak, G. (2019). Exploring the Effects of Tuition Increases on Racial/Ethnic Diversity at Public Colleges and Universities. Research in Higher Education, 60(1), 18-43. https://link.springer.com/article/10.1007%2Fs11162-018-9502-6
9. Kelchen, R., Goldrick-Rab, S., & Hosch, B. (2017). The Costs of College Attendance: Examining Variation and Consistency in Institutional Living Cost Allowances. The Journal of Higher Education, 88(6), 947-971 http://web.b.ebscohost.com.proxy.library.ucsb.edu:2048/ehost/pdfviewer/pdfviewer?vid=1&sid=cc41a448-6d85-49bd-9558-69c0427b2b2e%40pdc-v-sessmgr05