College Ranking Predictions

Report

Introduction

Research Background

Every year, millions of students apply to colleges across the United States, and many of them use college rankings lists from sources such as *US News and World Report*, Forbes.com, and Niche.com to to help them decide where to apply and where to go. In recent years, these lists have been heavily criticized for focusing on "exclusivity and resources, rather than accessibility and economic mobility"(1).The system can also be easily manipulated by a university if that university prioritize certain metrics to raise their rankings, as seen in Northeastern University meteoric rise from #163 to #49 on the US News and World Report list in only 17 years. The president of Northeastern University even explicitly stated that it was a top priority of the university to raise its ranking (2).

(1) https://thehill.com/changing-america/enrichment/education/3641004-the-scandal-facing-college-ranking-lists-explained/

(2) https://www.bostonmagazine.com/news/2014/08/26/how-northeastern-gamed-the-college-rankings/

These lists are important because students applying to college trust these rankings and weigh them into their college decisions. Due to the large impact that a college has on a student's life, it is important to know where these rankings come from and what they actually measure. In this project, we will explore how influential different metrics are in determining a college's ranking.

In general, we want to examine how different variables affect a school's ranking on the **Niche College Ranking List** and determine which are most important to a high ranking. We plan to look at variables that are typically thought to influence school rank such as average SAT score and admission rate, but we also want to look at variables that aren't typically thought of such as geographic region or endowment size.

For the sake of clarity, when we say a "low rank," we are referring to schools with a lower numerical rank, such as #1 and #2. When we say a "high rank," we a referring to schools with a high numerical rank, such as #497 and #498.

Research Question and Hypothesis

Question: Which characteristics of a university are most associated with rankings on the Niche College Ranking list? Of these characteristics, what is the relationship between high and low rank?

Hypothesis: We hypothesize that SAT/ACT scores, acceptance rate, and family income will have the strongest association with rank because since Niche's audience is in large part students applying to college, we believe that they prioritize variables important in the college admissions process. Of these variables, we predict that SAT/ACT score will have strong negative relationship, acceptance rate will have a strong positive relationship, and family income will have a strong negative relationship with rank.

In our project, we will join two data sets: The Niche College Rankings list and the US Department of Education College Scorecard

Data

Data Set #1: Niche

- The first data set comes from Niche's "2023 Best Colleges in America" list

  (https://www.niche.com/colleges/search/best-colleges/)
- Niche aggregates data from a variety of sources, including the US Department of Education and reviews from students and alumni, to build their list of college rankings. The rankings list is updated monthly to reflect new data that Niche receives. However, Niche only receives data from the US Department of Education on an annual basis. The Niche data was scraped by Maia on October 17-19 2022.
- There are 500 observations, representing the top 500 schools in the United States. Each observation has two variables: `college` and `rank`.

Data Set #2: US Department of Education

- The second data set comes from the US Department of Education's College Scorecard, which is an exhaustive summary of characteristics and statistics for all colleges and universities in the United States.

  (https://collegescorecard.ed.gov/data/)

- The College Scorecard is updated by the Education Department as it collects new data. Data used in the scorecard comes comes from data reported by the institutions, data on federal financial aid, data from taxes, and data from other federal agencies.
- There were 2,989 variables in the original data set, many of which we don't need to answer our question, and since this data set was too large to load into RStudio, we used Excel to narrow this to 31 variables. We chose any variables that we thought would have an impact on college rank, and excluded redundant ones like specific breaksdowns of test scores. There are 6681 observations in the data set, representing all of the colleges and universities in the United States.

Data summary

Here is a summary of the variables we will be using in our analysis.

Variable Name	Categorical (C) Quantitative (Q)	Description	Levels of Cat. Variable
college	C	Institution name	
rank	Q	Niche rank	
REGION	C		New England, Mid East, Great Lakes, Plains, Southeast, Southwest, Rocky Mountains, Far West, Outlying Areas
ACCREDAGENCY	C	Accreditor for Institution	
CONTROL	C		Public, Private nonprofit, private for-profit
CCBASIC	C	Carnegie Classification --basic	
ADM_RATE	Q	Admission rate	
UGDS	Q	Enrollment of undergraduate certificate/degree-seeking students	
UGDS_WHITE	Q	Total share of enrollment of undergraduate degree-seeking students who are white	
UGDS_BLACK	Q	Total share of enrollment of undergraduate degree-seeking students who are black	
UGDS_HISP	Q	Total share of enrollment of undergraduate degree-seeking students who are Hispanic	

Variable Name	Categorical (C) Quantitative (Q)	Description	Levels of Cat. Variable
UGDS_ASIAN	Q	Total share of enrollment of undergraduate degree-seeking students who are Asian	
UGDS_AIAN	Q	Total share of enrollment of undergraduate degree-seeking students who are American Indian/Alaska Native	
UGDS_NHPI	Q	Total share of enrollment of undergraduate degree-seeking students who are Native Hawaiian/Pacific Islander	
UGDS_2MOR	Q	Total share of enrollment of undergraduate degree-seeking students who are two or more races	
UGDS_NRA	Q	Total share of enrollment of undergraduate degree-seeking students who are non-resident aliens	
UGDS_UNKN	Q	Total share of enrollment of undergraduate degree-seeking students whose race is unknown	
NPT4_PUB	Q	Average net price for Title IV institutions (public institutions)	
NPT4_PRIV	Q	Average net price for Title IV institutions (private for-profit and nonprofit institutions)	
COSTT4_A	Q	Average cost of attendance (academic year institutions)	
COSTT4_P	Q	Average cost of attendance (program-year institutions)	
AVGFACSAL	Q	Average faculty salary	
PCTPELL	Q	Percentage of undergraduates who receive a Pell Grant	
C150_4	Q	Completion rate for first-time, full-time students at four-year institutions (150% of expected time to completion)	
AGE_ENTRY	Q	Average age of entry	
FEMALE	Q	Share of female students	

Variable Name	Categorical (C) Quantitative (Q)	Description	Levels of Cat. Variable
MARRIED	Q	Share of married students	
FIRST_GEN	Q	Share of first-generation students	
FAMINC	Q	Average family income	
MD_FAMINC	Q	Median family income	
ENDOWBEGIN	Q	Value of school's endowment at the beginning of the fiscal year	
SAT_AVG	Q	Average SAT equivalent score of students admitted	
ACTCMMID	Q	Midpoint of the ACT cumulative score	

Methodology

Data Preparation

1. To get the data, we scraped from Niche.com and downloaded data from the US Department of Education, and imported. The steps were done in an R script titled `niche-scrape.R`

2. Some of the college names were slightly different between datasets, so we had to individually change names in the Department of Education dataset to match those in the Niche one before joining the two. The Observations for `University of South Florida - Sarasota-Manatee` and `University of South Florida - St. Petersburg` have been dropped from the `colleges` data set due to their non-existence in the `us_dep_of_ed` data set. They were not present in the data set because these two universities were combined with the main University of South Florida campus. This is why the `colleges` data set only has 498 observations.

3. We joined the US Department of Education dataset to the Niche dataset by the college names. The first 10 rows are displayed below.

college	rank	[illegible]	[illegible]	[illegible]	[illegible]
Massachusetts Institute of Technology	1	[illegible]	New England Commission on Higher Education	-71.09323	[illegible]
Stanford University	2	[illegible]	Western Association of Schools and Colleges Senior Colleges and University Commission	-122.16736	[illegible]

college	rank	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Harvard University	3	[illegible] Cambridge MA	New England Commission on Higher Education	42.37447 -71.11831	[illegible] 5 2	0.05 [illegible] 33 35 35 36 31 35 8	10 34 36 33 9 15 [illegible] NA 0.3 [illegible]
Yale University	4	[illegible] New Haven CT	New England Commission on Higher Education	41.311 -72.92669	[illegible] 5 2	0.06 [illegible] 33 35 34 36 31 35 NA NA 34 35 33 NA	15 [illegible] NA 0.3 [illegible]
Princeton University	5	[illegible] Princeton NJ [illegible] 0070	[illegible] States Commission on Higher Education	[illegible] -74.65936	[illegible] 5 2	0.05 [illegible] 32 35 34 36 31 35 8	10 34 35 33 9 15 [illegible] NA 0.3 [illegible]

4. Then, we selected the variables that we thought could have an impact on college ranking as a starting point for our analysis. These are listed in our data summary.

5. Some of the categorical variables in the US Department of Education dataset (REGION, CONTROL, and CCBASIC) used numbers to represent the different levels, so we looked at

the data dictionary and replaced each number with the words that it represents.

6. All of the numerical variables are on different scales—for example, SAT scores can range from 400 to 1600, while admittance rate can only range from 0 to 1—and so we standardized them for easier analysis and comparison. We used the `scale()` functions (found at https://www.statology.org/standardize-data-in-r/) to make the mean value of each numeric variable 0, and the standard deviation 1. The first 10 rows of our standardized dataset are printed below.

college	rank	REGION	CONTROL	CCBASIC	ACCREDAGENCY	[illegible]
Massachusetts Institute of Technology	1	New England	Private, nonprofit	Universities: Very High Research Activity	New England Commission on Higher Education	[illegible]

college	rank	REGION	CONTROL	CCBASIC	ACCREDAGENCY	[illegible]
Stanford University	2	Far West	Private Non-profit	Doctoral Universities: Very High Research Activity	Western Association of Schools and Colleges Senior Colleges and University Commission	[illegible]
Harvard University	3	New England	Private Non-profit	Doctoral Universities: Very High Research Activity	New England Commission on Higher Education	[illegible]

college	rank	REGION	CONTROL	CCBASIC	ACCREDAGENCY	[illegible]
Yale University	4	New England	Private, Non-profit	Doctoral Universities: Very High Research Activity	New England Commission on Higher Education	[illegible]
Princeton University	5	Mid East	Private, Non-profit	Doctoral Universities: Very High Research Activity	Middle States Commission on Higher Education	[illegible]

Preliminary Exploration and Visualization

Means of Different Variables by Rank Group

Interval	Mean Admission Rate	Mean SAT Average	Mean ACT Median	Mean % White Students	Mean % Asian Students	Mean Cost of Attendance
1 to 100	0.2767340	1422.703	32.05495	0.4865770	0.1500930	60691.49
101 to 200	0.6181010	1266.932	27.45455	0.6271786	0.0698429	46371.74
201 to 300	0.6818970	1215.864	26.02469	0.6142470	0.0729400	43882.71
301 to 400	0.7157737	1160.091	24.14474	0.5629051	0.0548111	40799.30

Interval	Mean Admission Rate	Mean SAT Average	Mean ACT Median	Mean % White Students	Mean % Asian Students	Mean Cost of Attendance
401 to 498	0.7430135	1145.662	23.86076	0.6184102	0.0436867	36583.50

After grouping the schools into five groups of 100 by their rank, we see that different metrics vary considerably across the group. As the rank level gets higher, the mean admission rate increases, the mean SAT Average decreases, the mean ACT Median decreases, and the cost of attendance decreases. As far as demographic statistics, 1-100 ranked schools have considerably fewer White students and considerably more Asian students than schools ranked above 100.

Means of Rank by Categorical Variables

Below, we group the schools by the different categorical variables in our analysis and then take the mean rank for each of those groups.

Control	Mean Rank
Private, Non-profit	236.568
Public	269.670
Pivate, For-profit	360.750

We observe that private non-profit colleges have a higher mean rank than public colleges or private for-profit colleges.

Region	Mean Rank
NA	81.5000
New England	162.9070
Far West	222.6102
Mid East	229.7416
Great Lakes	260.5362
Southeast	273.7627
Rocky Mountains	273.8000
Southwest	285.5122
Plains	293.2632

As far as region, schools from New England have the highest mean rank, while schools from the Plains have the lowest mean rank.

Carnegie Classification (Basic)	Mean Rank
Special Focus Four-Year: Business & Management Schools	55.0000
Doctoral Universities: Very High Research Activity	123.8629
Special Focus Four-Year: Engineering Schools	148.5000
Baccalaureate Colleges: Arts & Sciences Focus	199.4045
Special Focus Four-Year: Faith-Related Institutions	213.0000
Baccalaureate/Associate's Colleges: Mixed Baccalaureate/Associate's	248.0000
Special Focus Four-Year: Other Health Professions Schools	268.0000
Doctoral Universities: High Research Activity	279.5663
Master's Colleges & Universities: Small Programs	307.0000
Baccalaureate Colleges: Diverse Fields	313.3500
Special Focus Four-Year: Arts, Music & Design Schools	329.5000
Doctoral/Professional Universities	331.5962
Special Focus Four-Year: Other Special Focus Institutions	351.0000
Master's Colleges & Universities: Larger Programs	363.2727
Master's Colleges & Universities: Medium Programs	365.2069
Special Focus Four-Year: Other Technology-Related Schools	385.0000
NA	430.6667
Tribal Colleges	453.0000

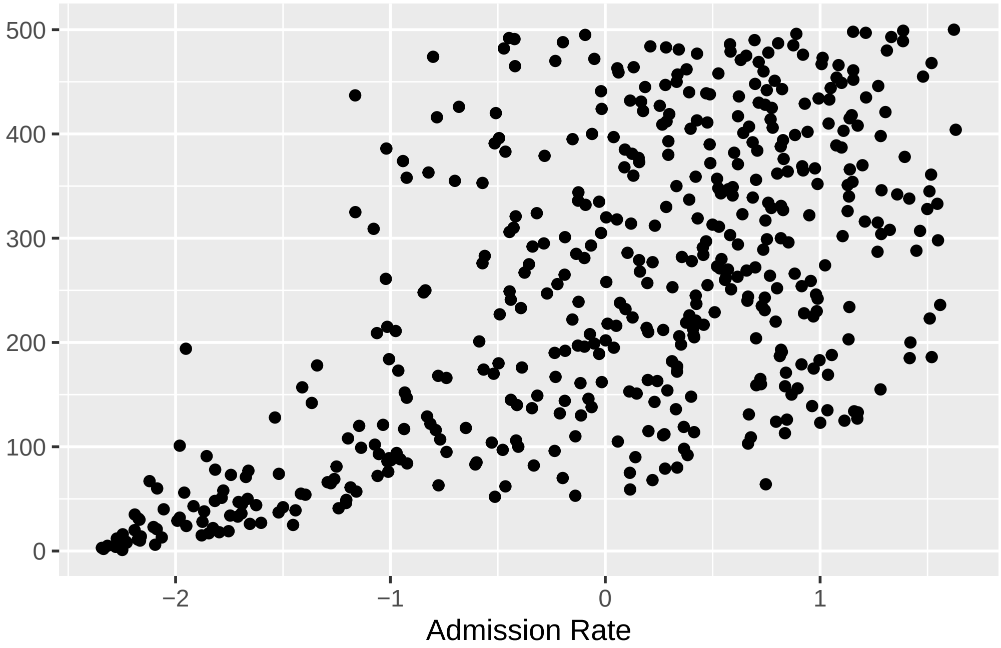
Looking at the Carnegie Classification, Special Focus Four-Year: Business & Management Schools have the highest mean rank, followed by Doctoral Universities: Very High Research Activity. Tribal Colleges have the lowest mean rank.

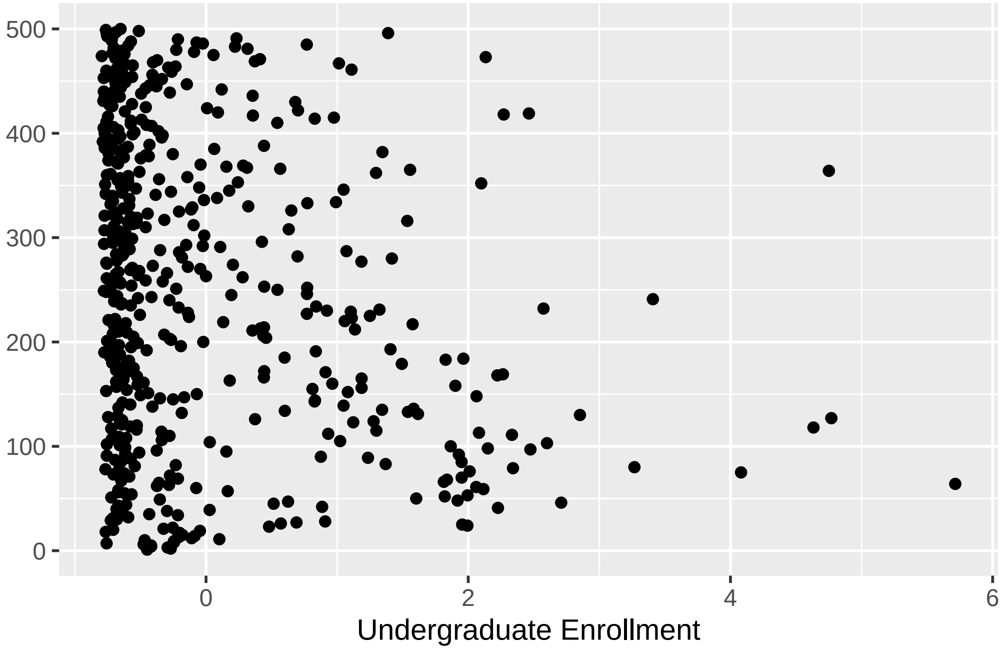
State	Mean Rank
ME	107.0000
MA	123.5238
DE	143.0000
CT	166.4286
VT	167.5000
CA	172.3333
DC	174.2000
GA	191.6667
OK	196.0000
WY	200.0000

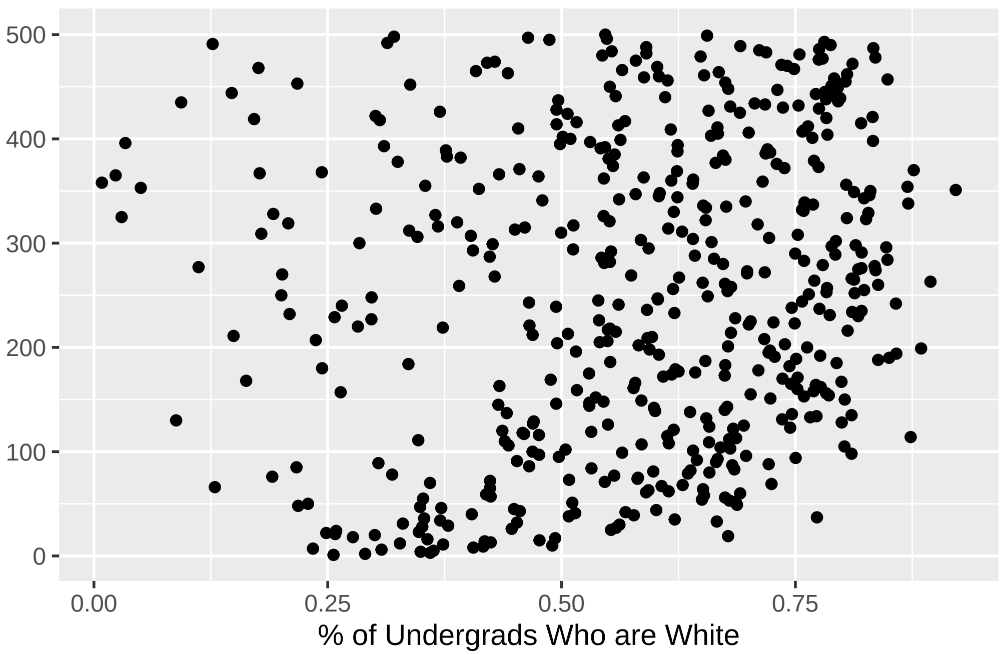
When looking at states, Maine has the highest mean rank, followed by Massachusetts.

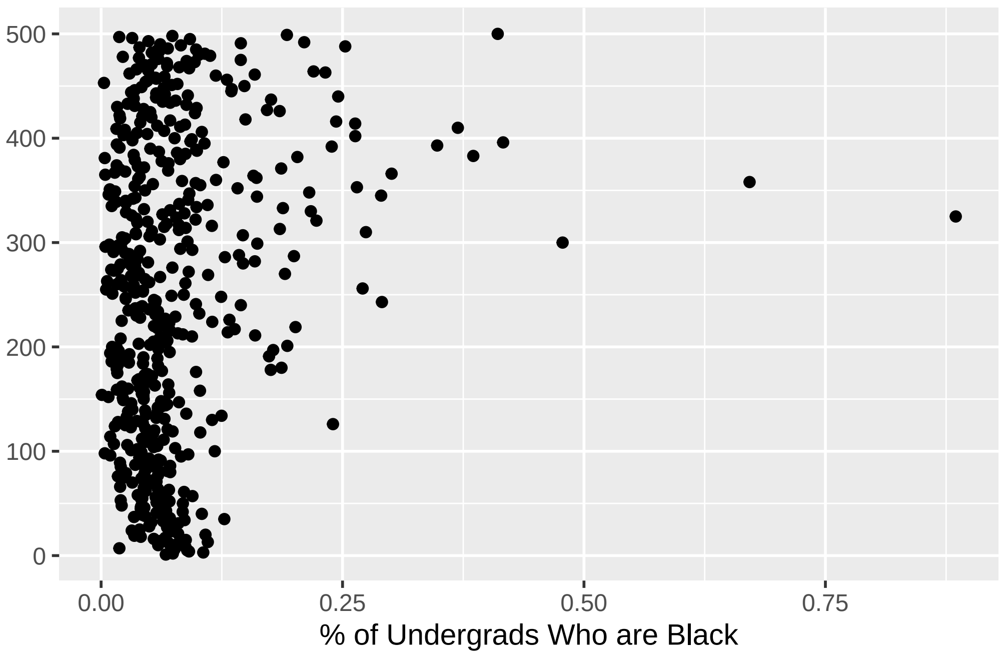
Numerical Variables vs. Rank

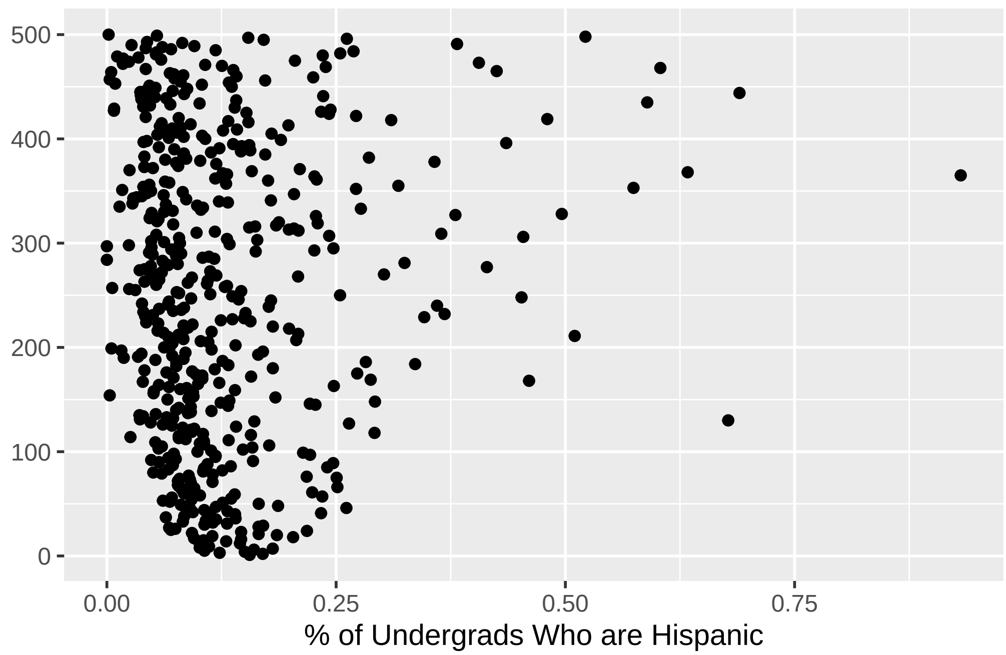
We have plotted rank versus several of the numerical variables. This gives us an idea of how different metrics affect college rank.

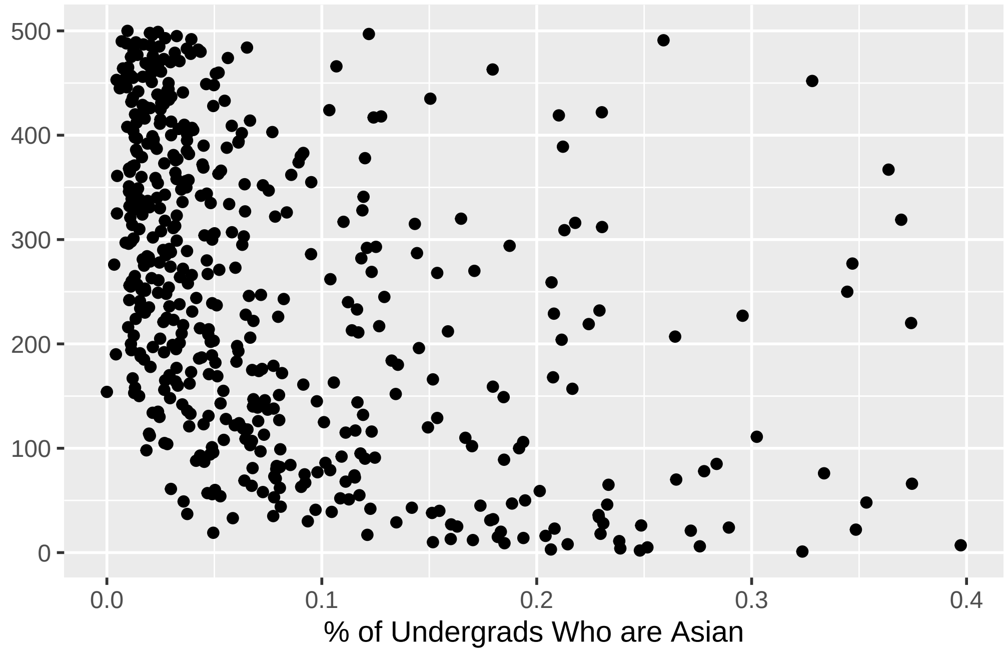


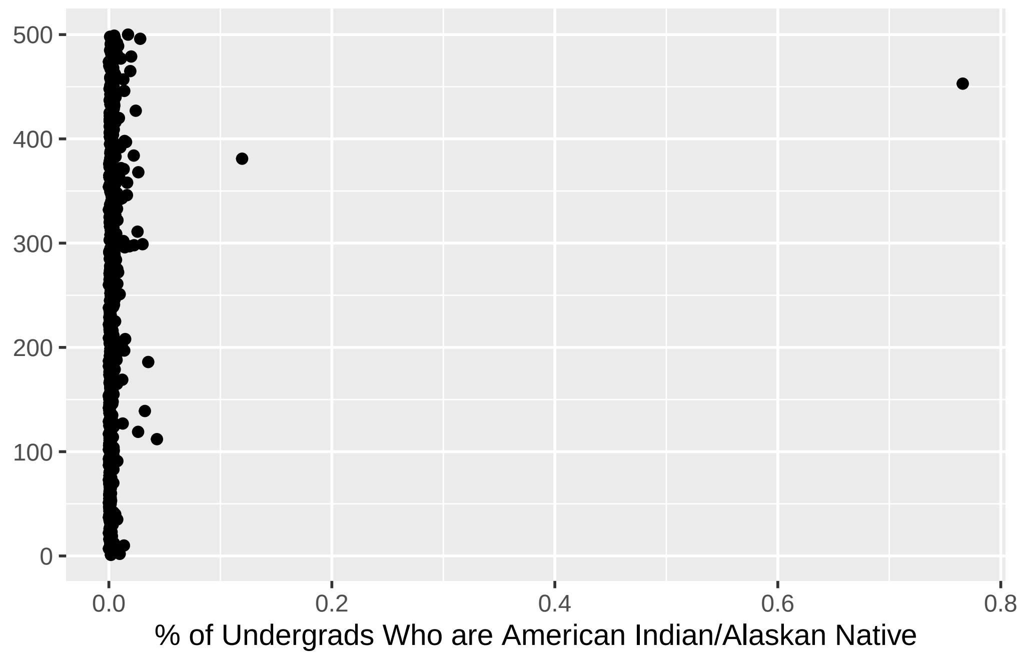


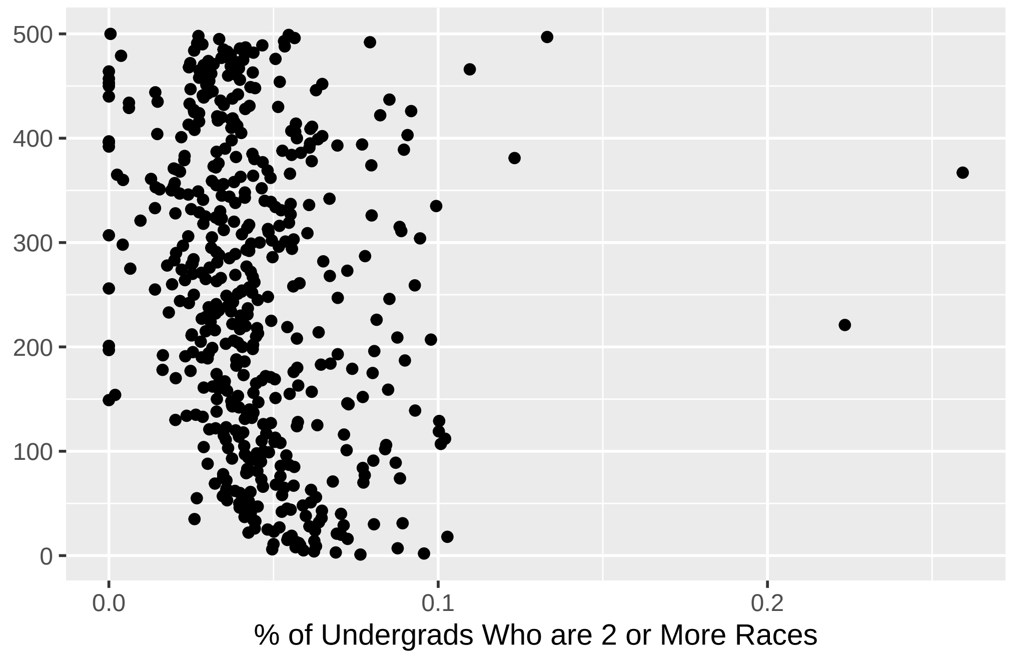

500
400
300
200
100
0
0.00
0.25
0.50
0.75
% of Undergrads Who are White

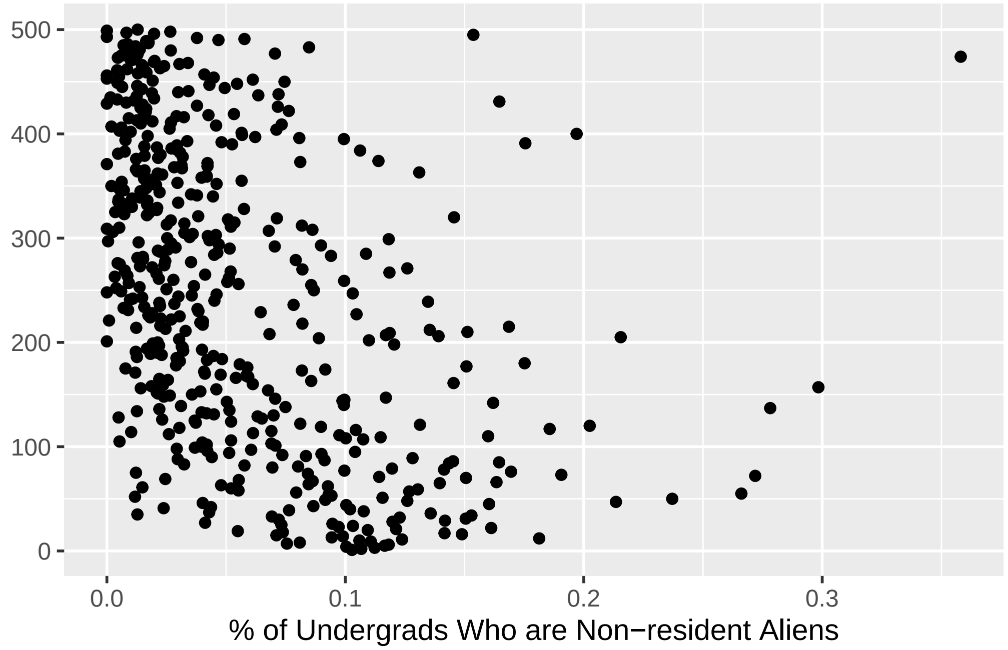

500
400
300
200
100
0
0.00
0.25
0.50
0.75
% of Undergrads Who are Black

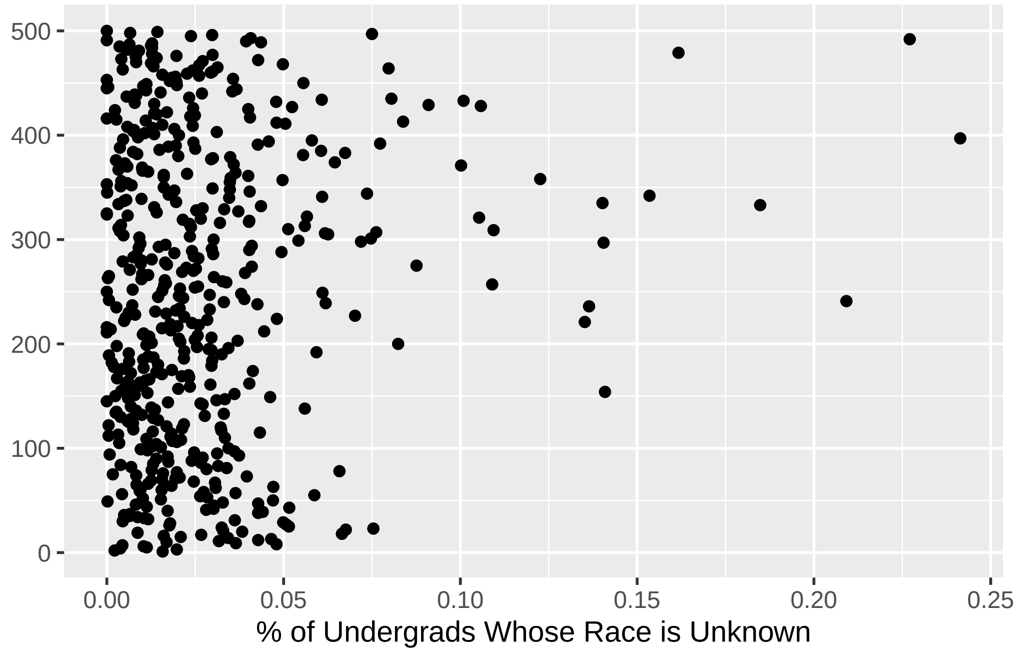

500
400
300
200
100
0
0.00
0.25
0.50
0.75
% of Undergrads Who are Hispanic

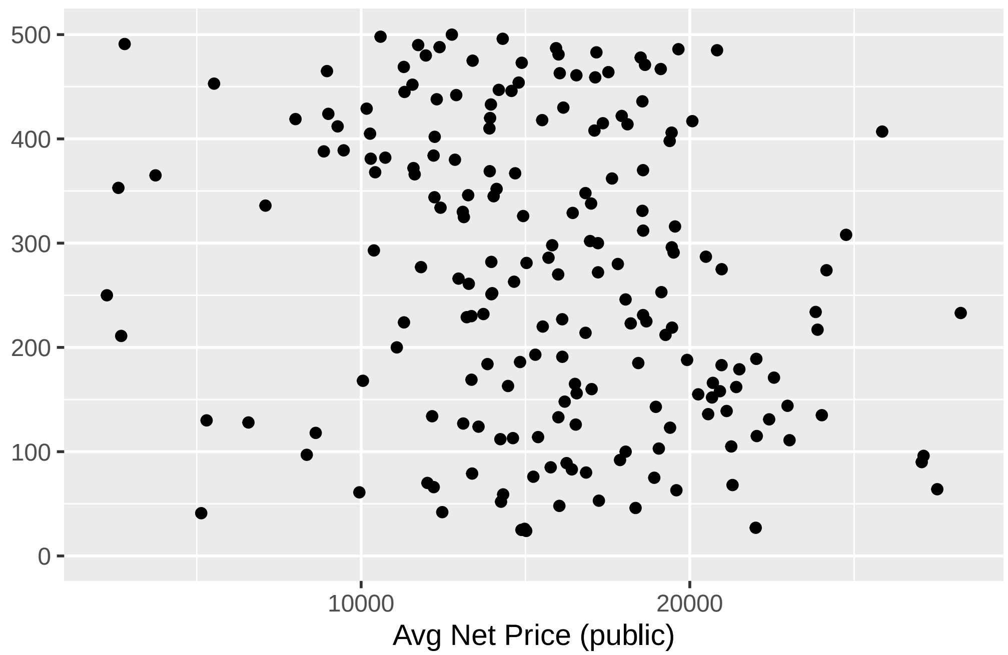

500
400
300
200
100
0
0.0
0.1
0.2
0.3
0.4
% of Undergrads Who are Asian

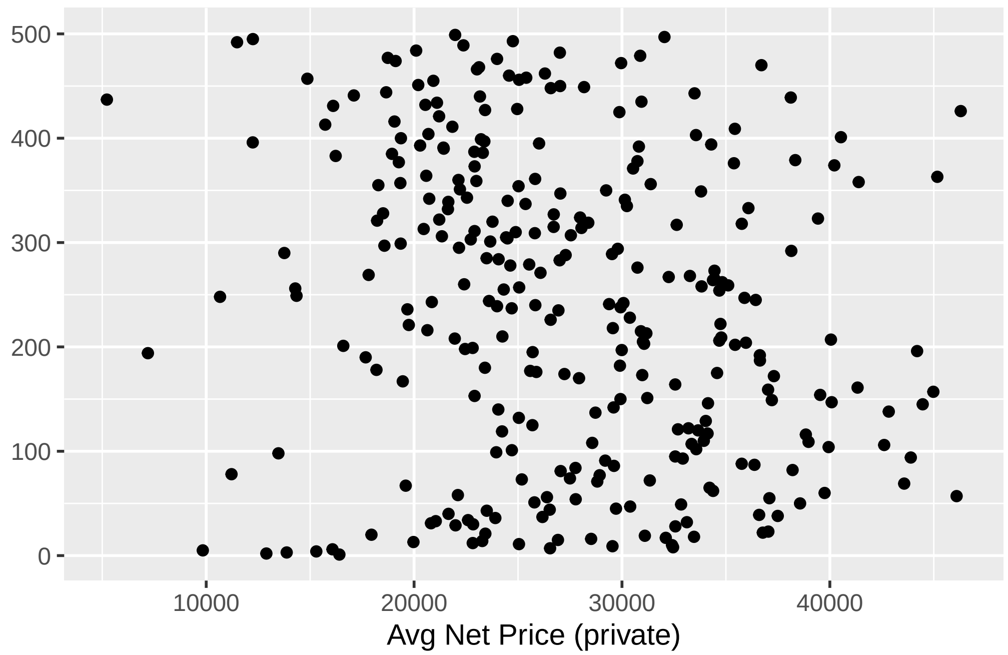

500
400
300
200
100
0
0.0
0.2
0.4
0.6
0.8
% of Undergrads Who are American Indian/Alaskan Native

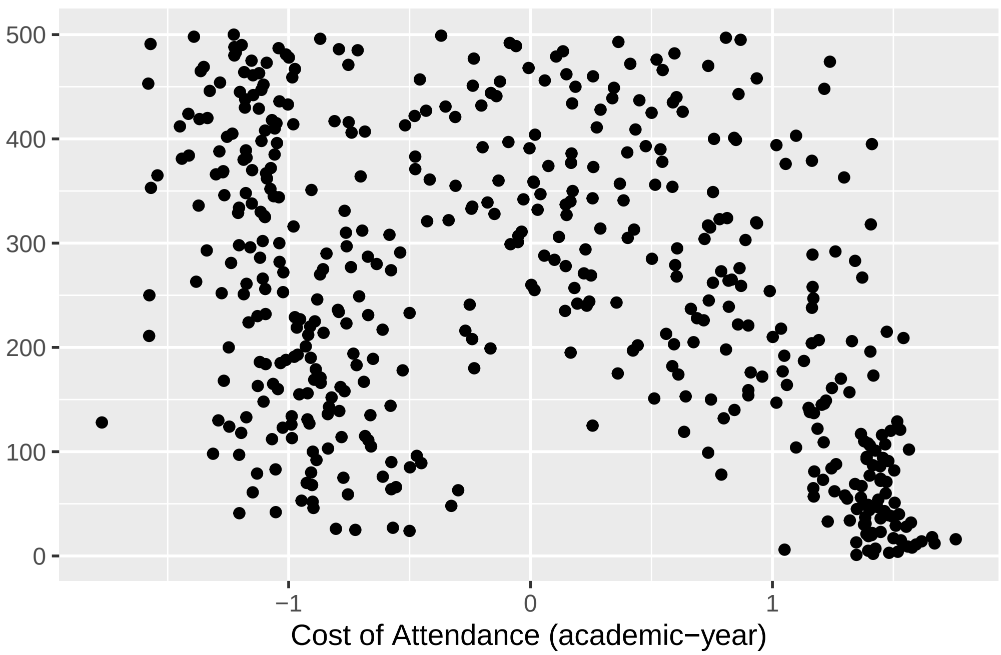

500
400
300
200
100
0
0.0
0.1
0.2
% of Undergrads Who are 2 or More Races

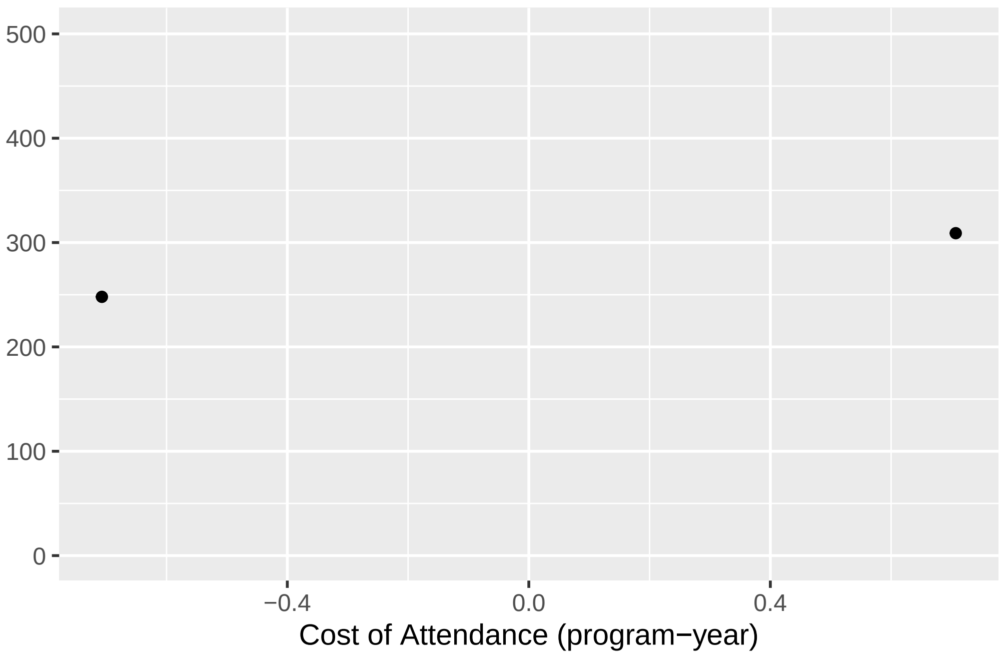

500
400
300
200
100
0
0.0
0.1
0.2
0.3
% of Undergrads Who are Non−resident Aliens

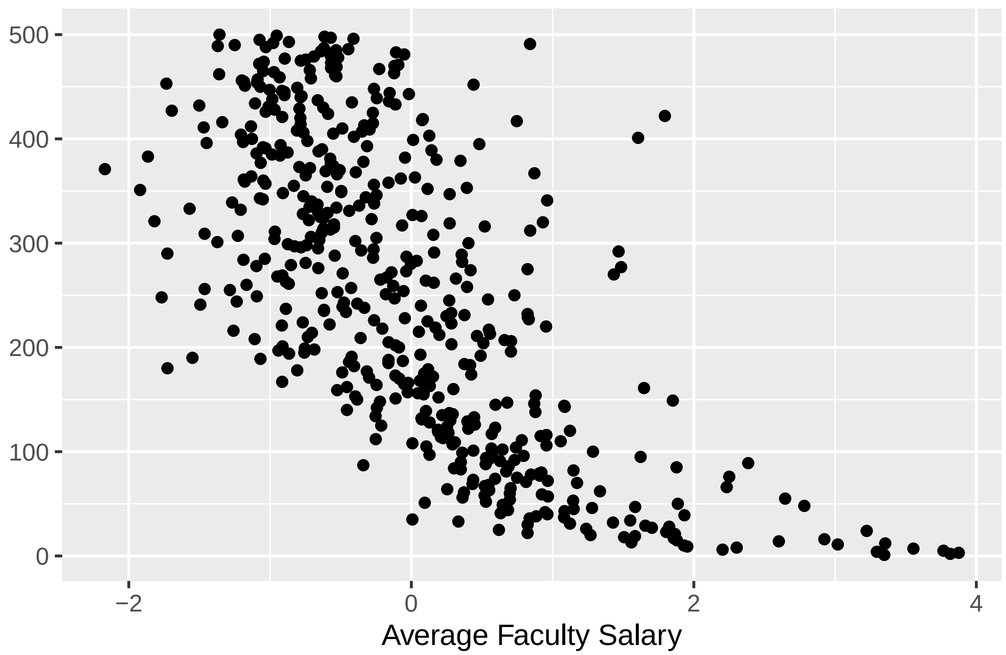

500
400
300
200
100
0
0.00
0.05
0.10
0.15
0.20
0.25
% of Undergrads Whose Race is Unknown

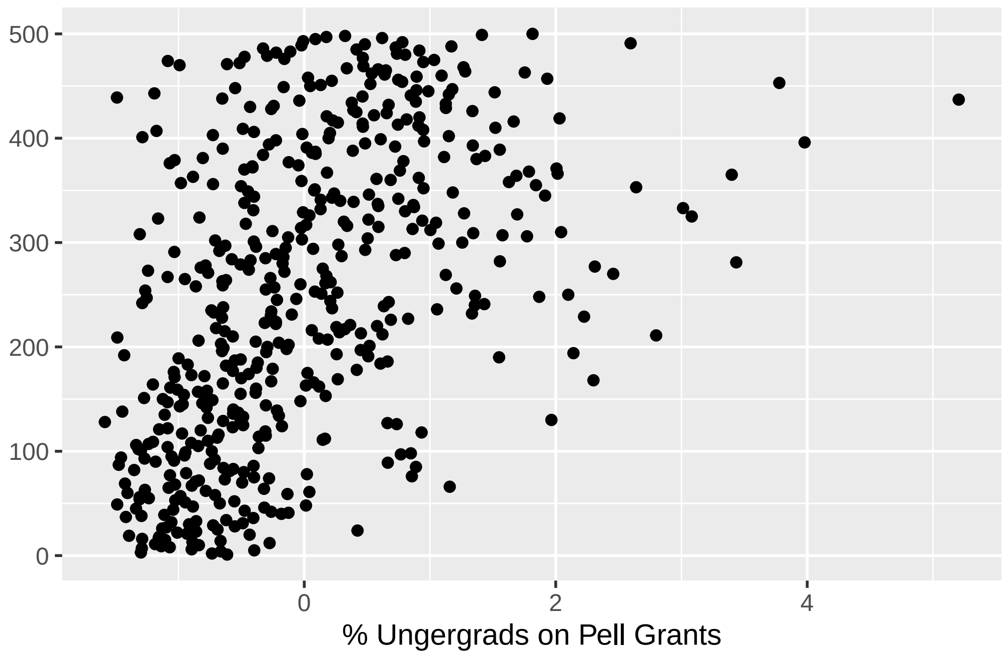

500
400
300
200
100
0
10000
20000
Avg Net Price (public)

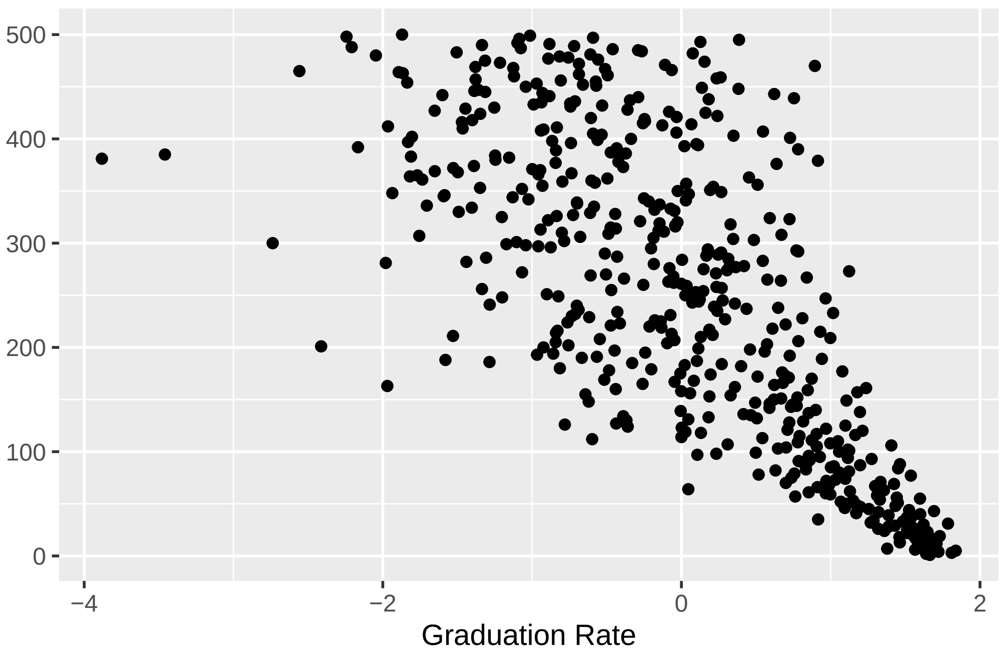

500
400
300
200
100
0
10000
20000
30000
40000
Avg Net Price (private)

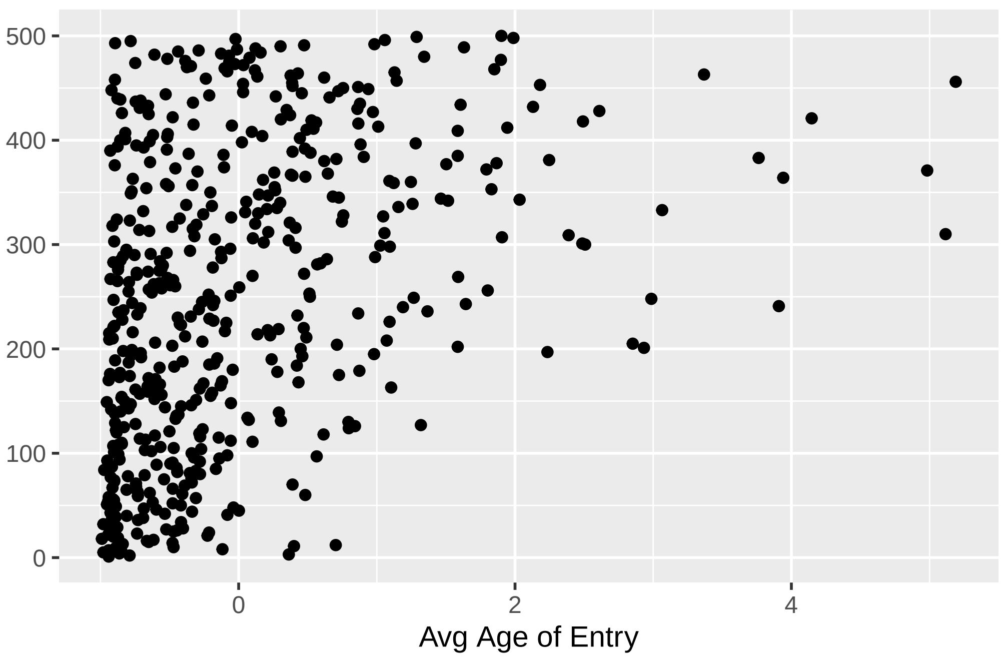

500
400
300
200
100
0
−1
0
1
Cost of Attendance (academic−year)

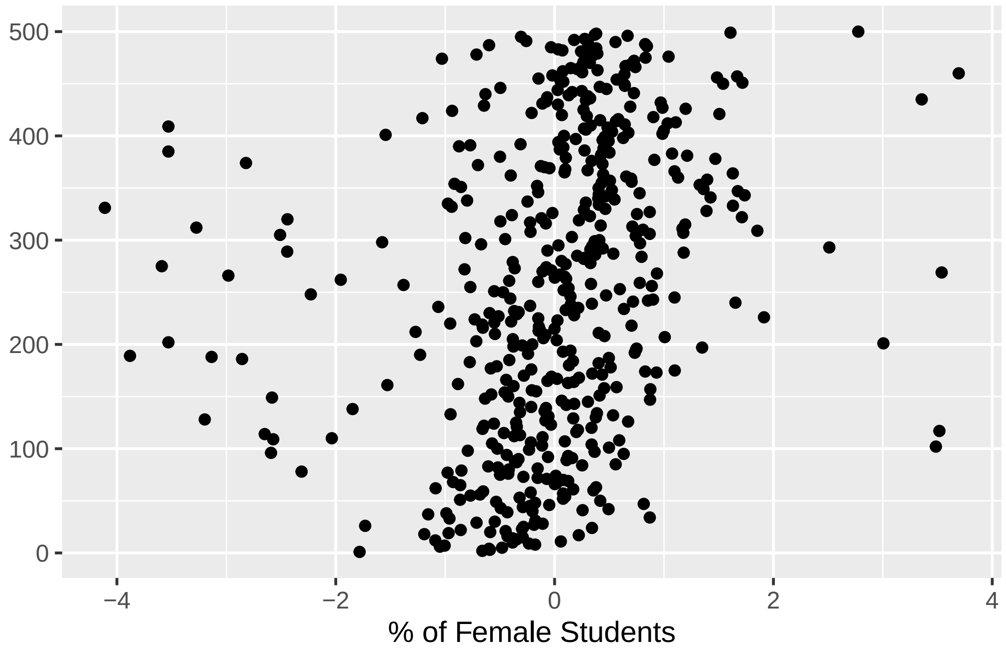

500
400
300
200
100
0
−0.4
0.0
0.4
Cost of Attendance (program−year)

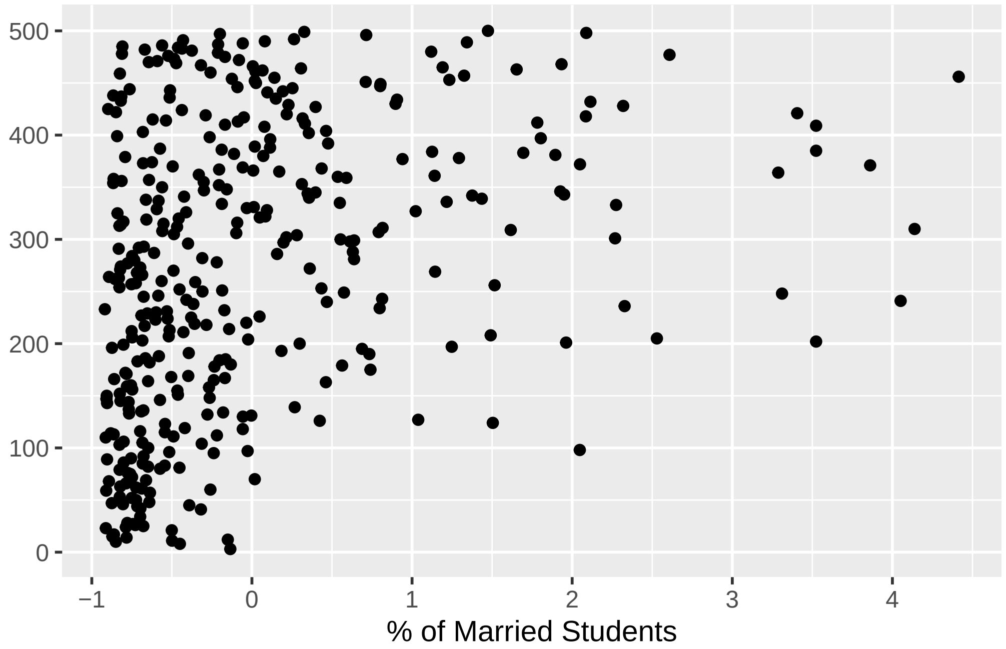

500
400
300
200
100
0
−2
0
2
4
Average Faculty Salary

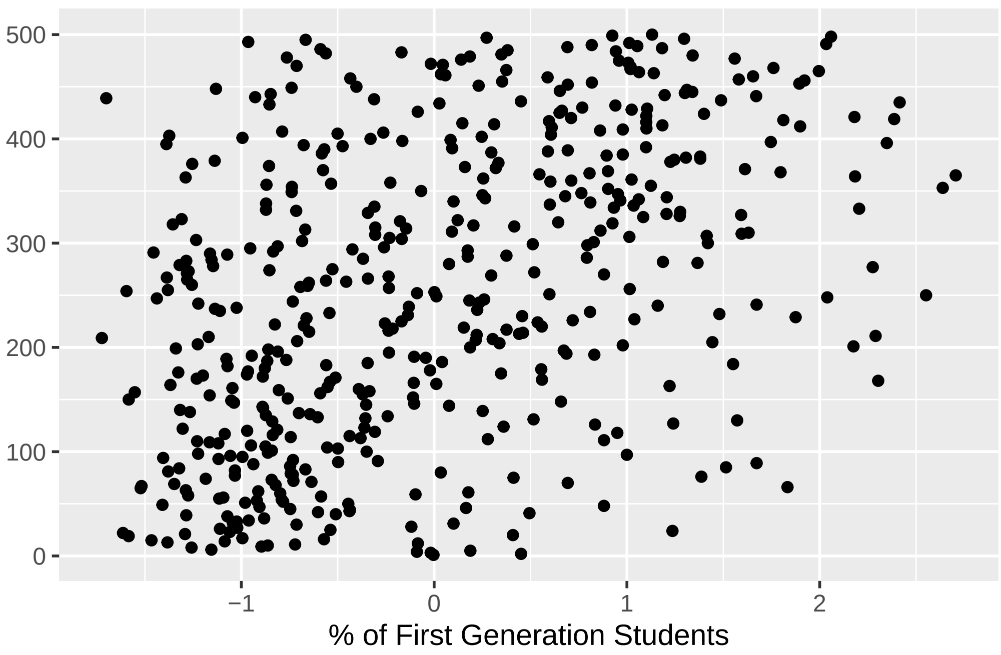

500
400
300
200
100
0
0
2
4
% Ungergrads on Pell Grants

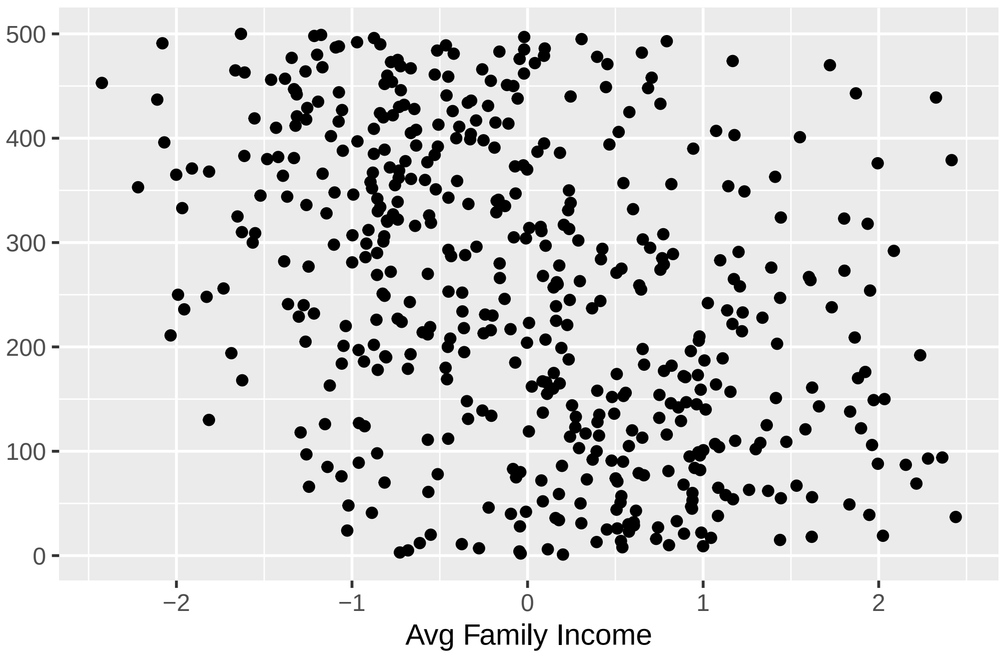

500
400
300
200
100
0
−4
−2
0
2
Graduation Rate

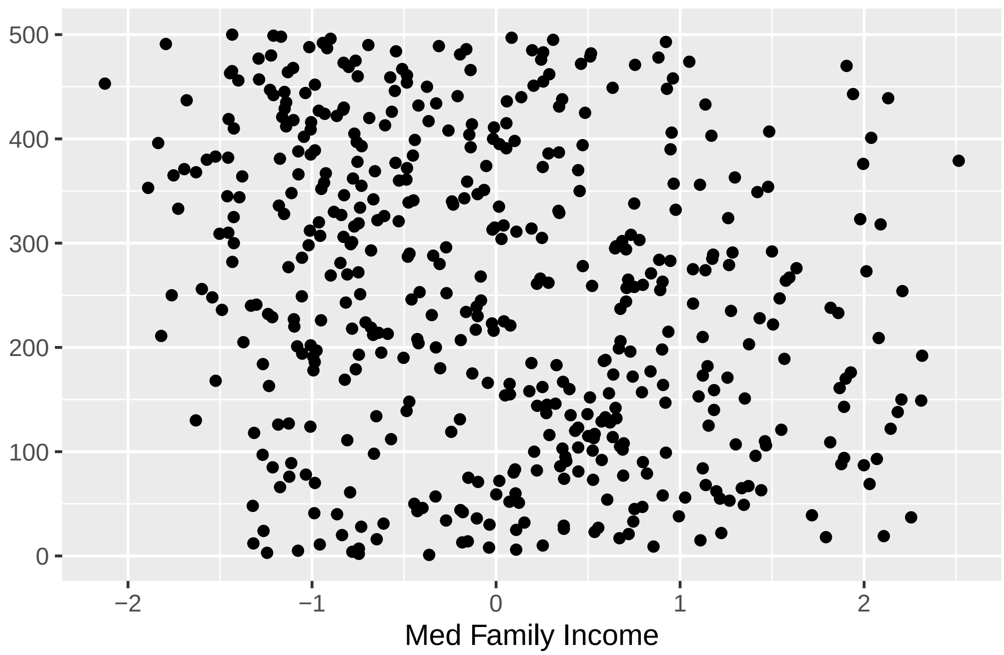

500
400
300
200
100
0
0
2
4
Avg Age of Entry

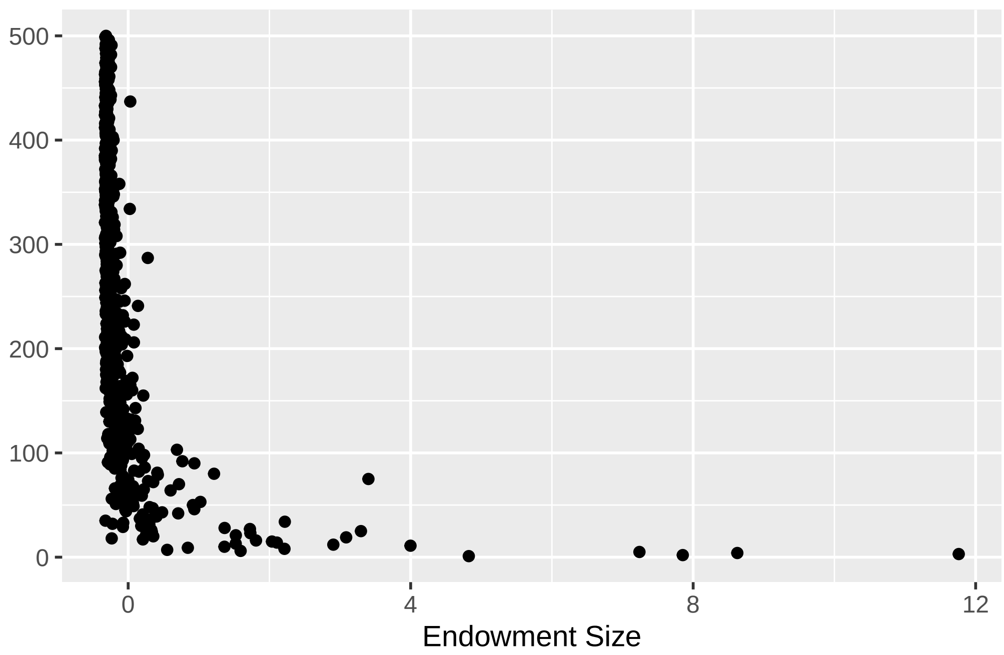

500
400
300
200
100
0
−4
−2
0
2
4
% of Female Students

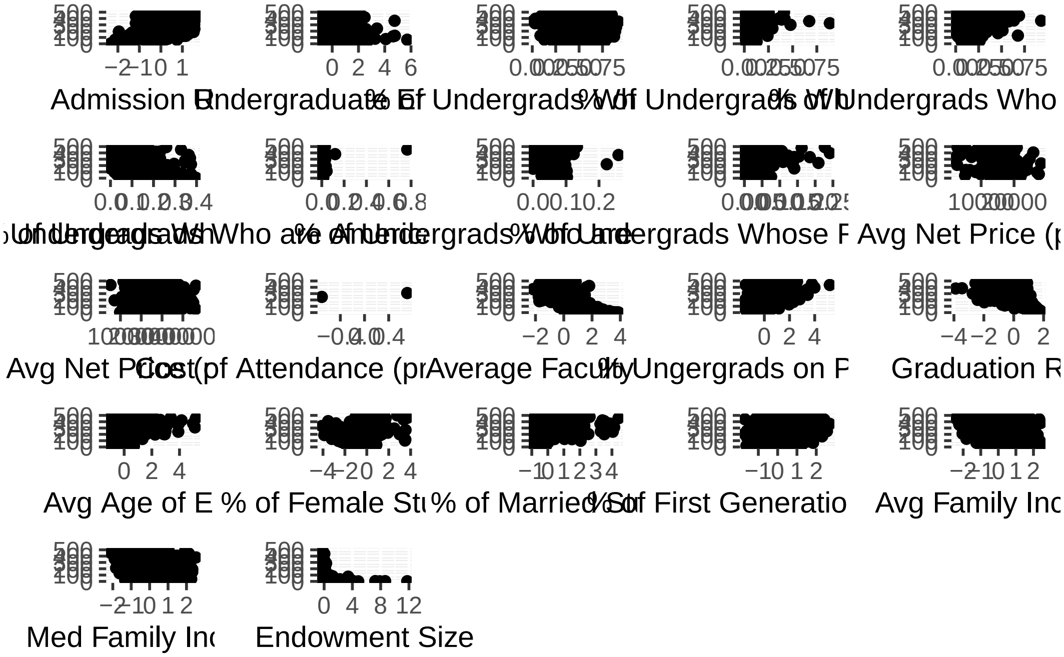

500
400
300
200
100
0
−1
0
1
2
3
4
% of Married Students


500
400
300
200
100
0
−1
0
1
2
% of First Generation Students


500
400
300
200
100
0
−2
−1
0
1
2
Avg Family Income

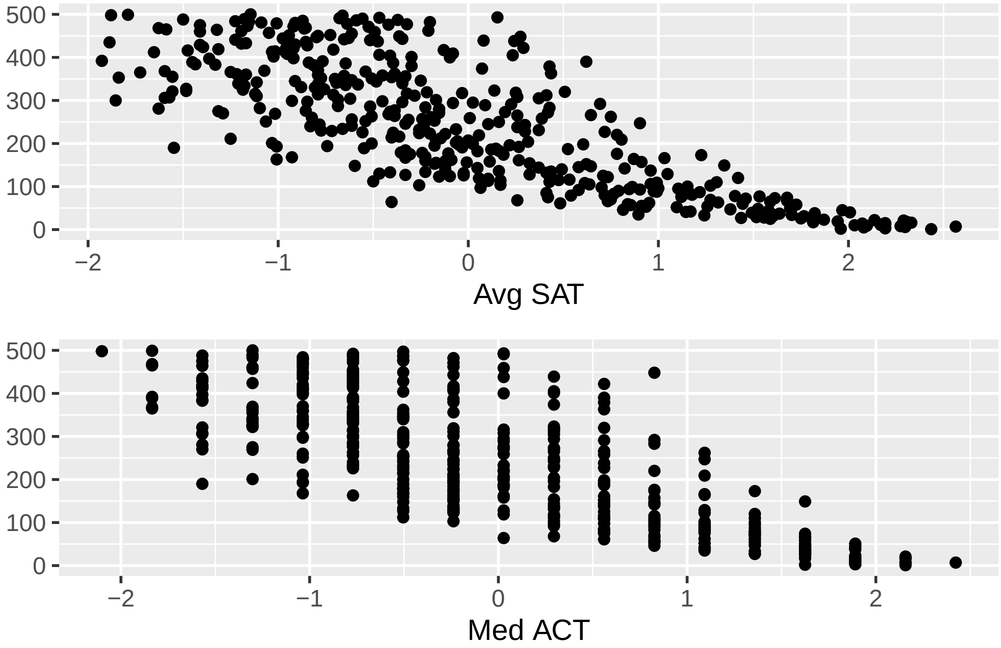

500
400
300
200
100
0
−2
−1
0
1
2
Med Family Income


500
400
300
200
100
0
0
4
8
12
Endowment Size



From these graphs, there appears to be a very strong positive association between admission rate/ % of undergrads on Pell grants and college rank. There appears to be somewhat of a positive association between % of undergrads who are white/% of first generation students and college ranks. This means that as these metrics decrease, the rank of a school is expected to decrease

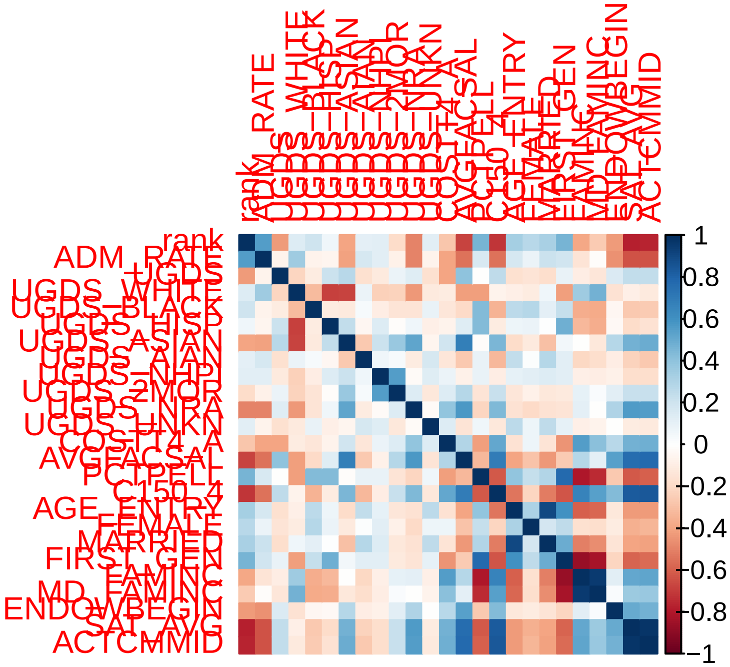
There appears to be a strong negative association between % of undergrads who are Asian/% of undergrads who are non-resident aliens/average faculty salary/graduation rate and college rank. There appears to be somewhat of a negative relationship between average net price for private universities/cost of attendance/average family income/median family income and college rank. This means that as these metrics increase, the rank of a school is expected to increase.



For all type of standardized test scores, there appear to be a strong negative relationship between score and rank, indicating that as the test scores increases, the rank of a school is expected to to increase.

Check Correlation Coefficients

We checked the correlation coefficients between the variables so we don't use variables that are too similar in our model. For the ones that have a absolute value of r greater than 0.8, we picked only one to put in the model. We removed the columns that has more than 100 NA and used the rest for calculating correlations.



According to the correlation matrix, we need to remove variable pairs that have a high correlation (the absolute value > 0.8). In order to do so, we filter the following correlation table to only display entries with value greater or equal to 0.8.

Variable Pairs with r > 0.8	Correlation Coefficients
C150_4, SAT_AVG	0.8389
C150_4, ACTCMMID	0.8494
AGE_ENTRY, MARRIED	0.9059
FAMINC, MD_FAMINC	0.9538
SAT_AVG, ACTCMMID	0.9756

According to the filtered table, these variable pairs are (MD_FAMINC, FAMINC), (C150_4, SAT_AVG), (C150_4, ACTCMMID) and (ACTCMMID, SAT_AVG). Therefore, we will drop the variables C150_4, MD_FAMINC, ACTCMMID and preserve SAT_AVG and FAMINC to represent all other variables.

Build model

Because we want to predict college rank, a number from various variables, we decided that a linear regression model would be the best statistical method to answer our question. We began our modeling by removing the variables with a correlation coefficient above 0.8.

Selected Variables
college
rank
REGION
CONTROL
CCBASIC
ACCREDAGENCY
ADM_RATE
UGDS
UGDS_WHITE
UGDS_BLACK
UGDS_HISP
UGDS_ASIAN
UGDS_AIAN
UGDS_NHPI
UGDS_2MOR
UGDS_NRA
UGDS_UNKN
COSTT4_A
AVGFACSAL
PCTPELL
AGE_ENTRY
FAMINC
ENDOWBEGIN
SAT_AVG
FEMALE
MARRIED
FIRST_GEN

Then we need to split dataset into two parts: a training and a test set. We will use 80% of the college data for training our model and the other 20% for testing our model.

Results

Iteration #1

We begin with creating the first model based on the numerical variables that we chose for the data and created an additive model that considered how the variables influenced the rank of the college.

term	estimate	std.error	statistic	p.value
(Intercept)	246.2027925	4.703871	52.3404663	0.0000000
ADM_RATE	35.6018194	7.629838	4.6661307	0.0000046
UGDS	-41.1973274	6.928463	-5.9460994	0.0000000
UGDS_WHITE	10732.7749048	10599.772816	1.0125476	0.3120710
UGDS_BLACK	4533.2629889	4474.426742	1.0131494	0.3117840
UGDS_HISP	6313.6053442	6232.516748	1.0130106	0.3118502
UGDS_ASIAN	4541.8380285	4477.565288	1.0143544	0.3112098
UGDS_AIAN	1926.4404448	1973.880100	0.9759663	0.3298471
UGDS_NHPI	163.9398183	167.303350	0.9798956	0.3279067
UGDS_2MOR	1407.8496454	1388.644877	1.0138299	0.3114597
UGDS_NRA	2973.0798562	2948.226876	1.0084298	0.3140397
UGDS_UNKN	1735.7967205	1705.950405	1.0174954	0.3097164
COSTT4_A	0.8428333	10.525477	0.0800755	0.9362292
AVGFACSAL	-19.8005704	8.998855	-2.2003434	0.0285247
PCTPELL	2.8162537	10.541378	0.2671618	0.7895234
AGE_ENTRY	-12.0862558	7.334480	-1.6478682	0.1003994
FAMINC	-22.9537106	11.800114	-1.9452109	0.0526593
ENDOWBEGIN	1.3023057	6.035923	0.2157592	0.8293183
SAT_AVG	-67.4876184	12.876049	-5.2413296	0.0000003

At a first glance, variables with negative slopes mean that they have a negative correlation to ranking. For example, the higher the `SAT_AVG`, the lower the college is ranked on a numerical scale. So the college is technically ranked higher.

Variables UGDS_NRA, UGDS_UNKN, COSTT4_A, PCTPELL, AGE_ENTRY, UGDS_WHITE, UGDS_BLACK, UGDS_HISP, UGDS_ASIAN, UGDS_AIAN, UGDS_NHPI, UGDS_2MOR and ENDOWBEGIN are have a relative high p-value around or above 0.3, which is much greater than 0.05. Therefore, we will drop these variables for the next model iteration.

Then, we can evaluate our first primitive model by R-squared:

adj.r.squared
0.7199576

As we can see the model's adjusted R-squared is around 72.0%.

Iteration #2

First, we will avoid using the variables with high p-values as mentioned above to train a new model.

term	estimate	std.error	statistic	p.value
(Intercept)	249.37044	4.506181	55.339648	0.0000000
ADM_RATE	27.21518	6.659181	4.086866	0.0000551
UGDS	-33.32915	4.895297	-6.808403	0.0000000
AVGFACSAL	-10.80730	7.134635	-1.514766	0.1308007
SAT_AVG	-86.86569	7.796687	-11.141359	0.0000000

Then we can evaluation our second model by Rsqure:

```
glance(fit2) |>
  select(adj.r.squared) |> knitr::kable()
```

Since the R-squared value is 0.703 which is even smaller than 0.7200 of our first model, we need to implement forward selection by adding variables back to our model.

Iteration #3

From our first and second iteration, AGE_ENTRY, among the variables removed, has the smallest p-value of 0.1. Therefore we are going to add AGE_ENTRY back to our model for the iteration.

term	estimate	std.error	statistic	p.value
(Intercept)	249.7168722	4.523580	55.2033783	0.0000000
ADM_RATE	26.2874687	6.797196	3.8673990	0.0001330
UGDS	-33.2547774	4.923043	-6.7549237	0.0000000
AVGFACSAL	-11.4385472	7.165074	-1.5964311	0.1113674
SAT_AVG	-86.6323357	8.515896	-10.1730147	0.0000000
AGE_ENTRY	0.4599436	5.802186	0.0792707	0.9368662

adj.r.squared
0.701619

The model accuracy is still not as good as the first model which means we need further forward selection by adding more variables back.

Iteration #4

From our first and second iteration, UGDS_UNKN, among the variables removed, has the smallest p-value of 0.310. Therefore we are going to add UGDS_UNKN back to our model for the iteration.

term	estimate	std.error	statistic	p.value
(Intercept)	249.7153091	4.521131	55.232924	0.0000000
ADM_RATE	27.2202655	6.840754	3.979132	0.0000854
UGDS	-32.7123939	4.942456	-6.618652	0.0000000
AVGFACSAL	-11.1617082	7.165156	-1.557776	0.1202656
SAT_AVG	-86.3294688	8.515275	-10.138190	0.0000000
AGE_ENTRY	-0.8911943	5.914406	-0.150682	0.8803207
UGDS_UNKN	5.4065047	4.651444	1.162328	0.2459603

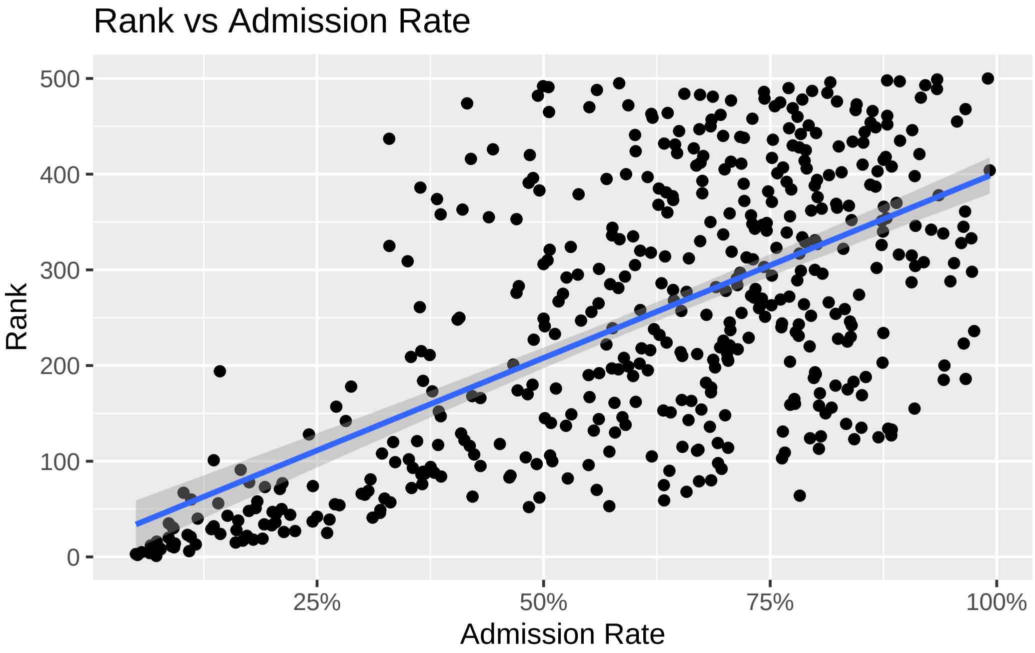
adj.r.squared
0.7019419

While this model proved better than the previous one, it is still not as good as our initial model.

Model results

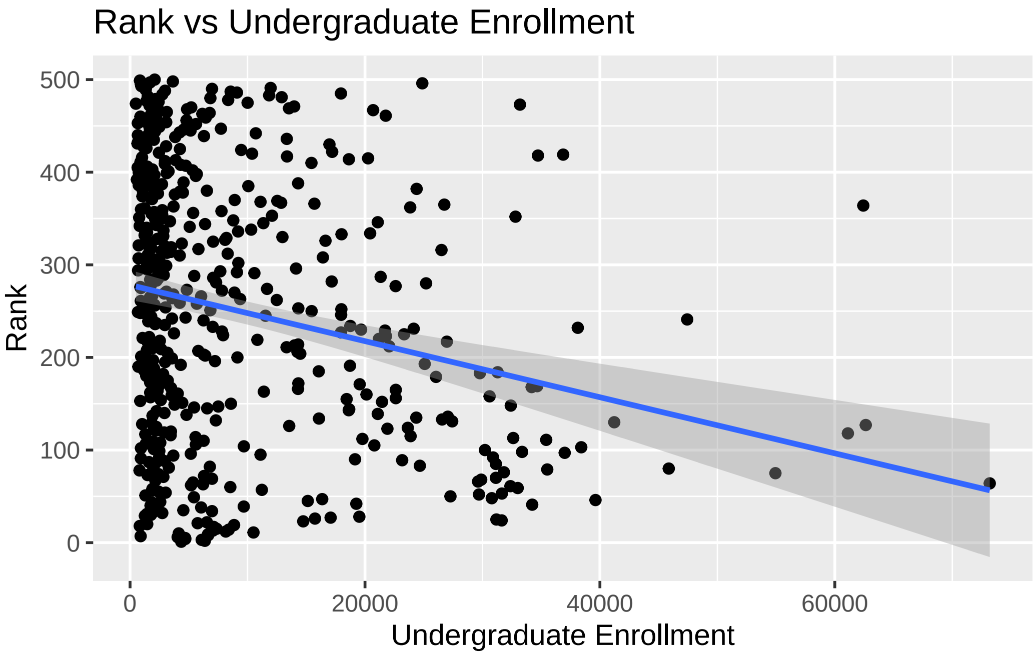
Because our initial model had the highest r-squared value of 0.7354, we can consider that to be our best model for predicting college ranking. However, in an effort to determine which variables have the most significant impact on college ranking, we can look at those with the lowest p-values: `ADM_RATE`, `UGDS`, `AVGFACSAL`, and `SAT_AVG`. Then, we can further single out the most significant variable with R-squared values.

R-Squared Values



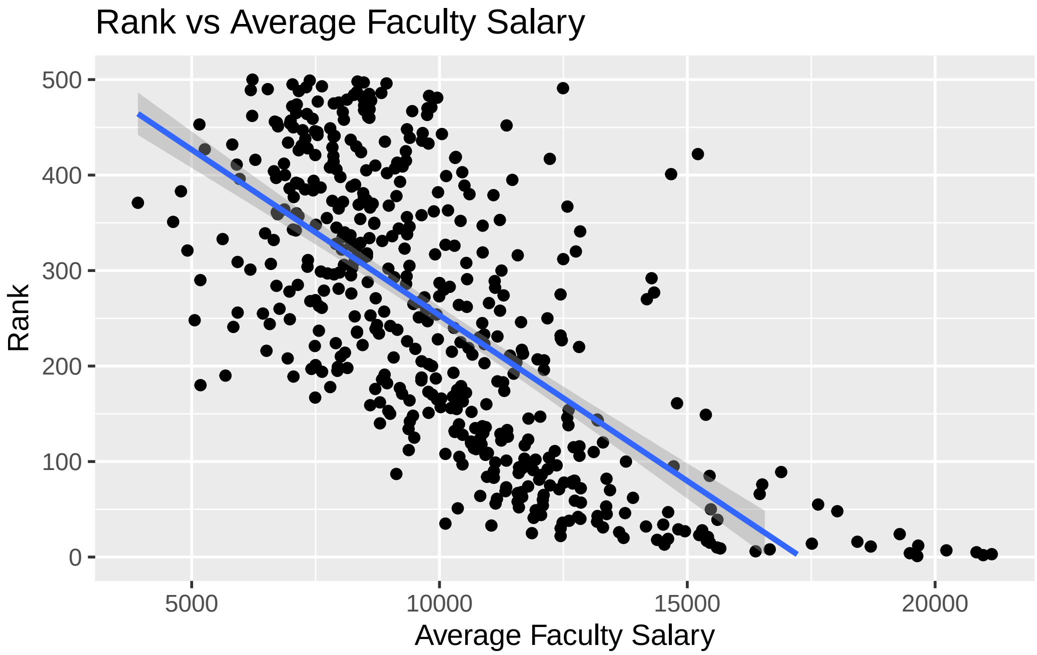
r.squared
0.4153932

The relationship between admission rate and college ranking gives us an R-squared value of 0.415 which is not that good and shows that admission rate might not be as significant as we originally thought.



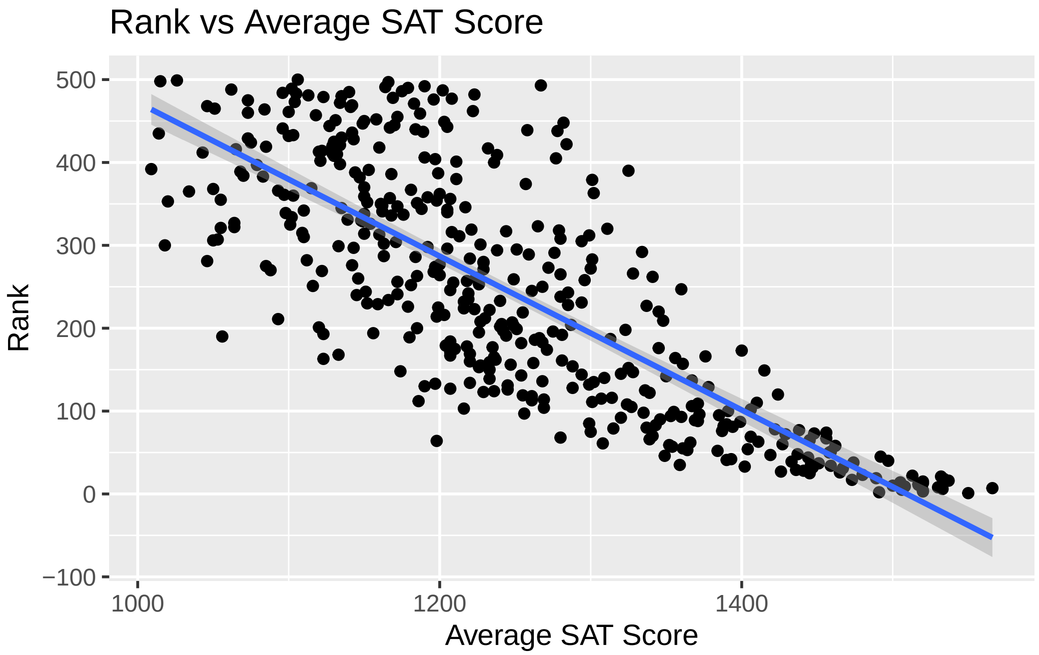
r.squared
0.0479294

As evident from the graph and the given R-squared value of 0.048, undergraduate enrollment has almost no impact on a college's ranking.



r.squared
0.4768954

While still relatively low, the R-squared value of 0.477 for the relationship between average faculty salary and ranking is higher than that of admission rate vs rank. This signifies that average faculty salary is a slightly more significant variable in predicting college ranking than admission rate.



Based on the plot above, a college's average SAT score has an almost semi-linear impact on it's ranking. Typically, as average SAT score increases, a college's ranking improves. Furthermore, we can see how well correlated these two variables are by creating a linear model.

r.squared
0.6410464

While an R-squared value of 0.641 is not the best, it is still the highest out of these variables which means that we can conclude that average SAT score has the greatest impact on a college ranking.

Final Conclusion

Works Cited