COLLEGES THAT MAKE THE AMERICAN DREAM A REALITY

The data

1) The College Scorecard data, made freely available by U.S. Department of Education

   (a) the characteristics of the institutions

   (b) academic offerings

   (c) the median and mean values of student SAT scores, family income, earnings
   (d) characteristics of the neighborhood where the institution is located from the census data

2) The Equal Opportunity Project

   (a) mobility ratings (% of students from the bottom 20% who reach the top 20% post-graduation).

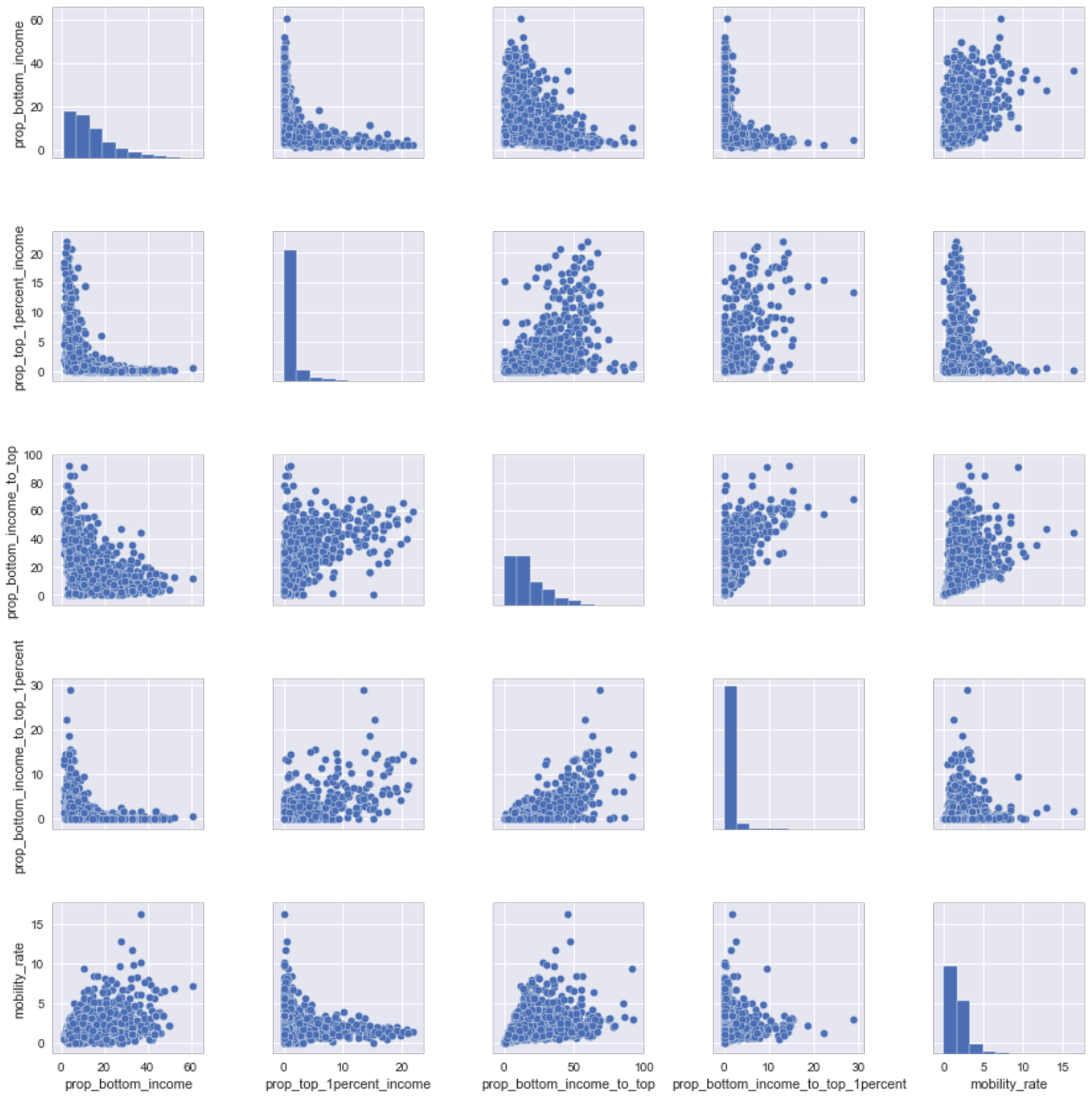
   (b) conditional mobility (% of the low income students who, once admitted reach the top 20% post-graduation)

Data Wrangling

- Imported 120 variables using College Scorecard card's API.
- Imported the Equal Opportunity Project mobility ranking data from the website.
- Concatenated the College Scorecard data and the Equal Opportunity Project data using the OPE_ID, which is a unique institution ID in both datasets
- Data from 2013 is the most complete, recent dataset.
- Imputed missing census data from 2005.
- Pair plots to visually explore the data and look for problems and outliers

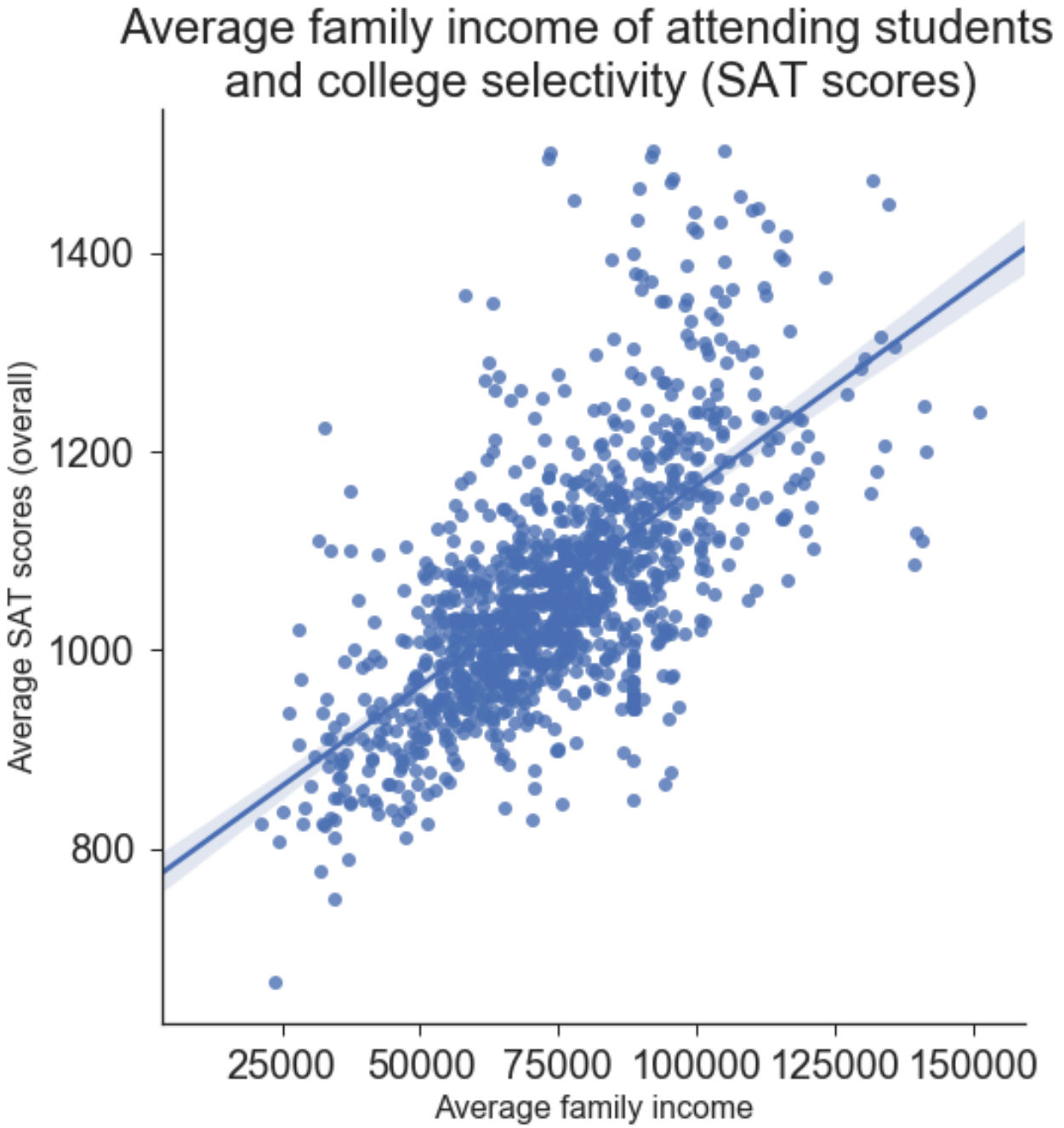
Data issues

- Many of the metrics are proportions and are non-normal. These were square-root transformed for inferential statistics but treated as is for ML methods

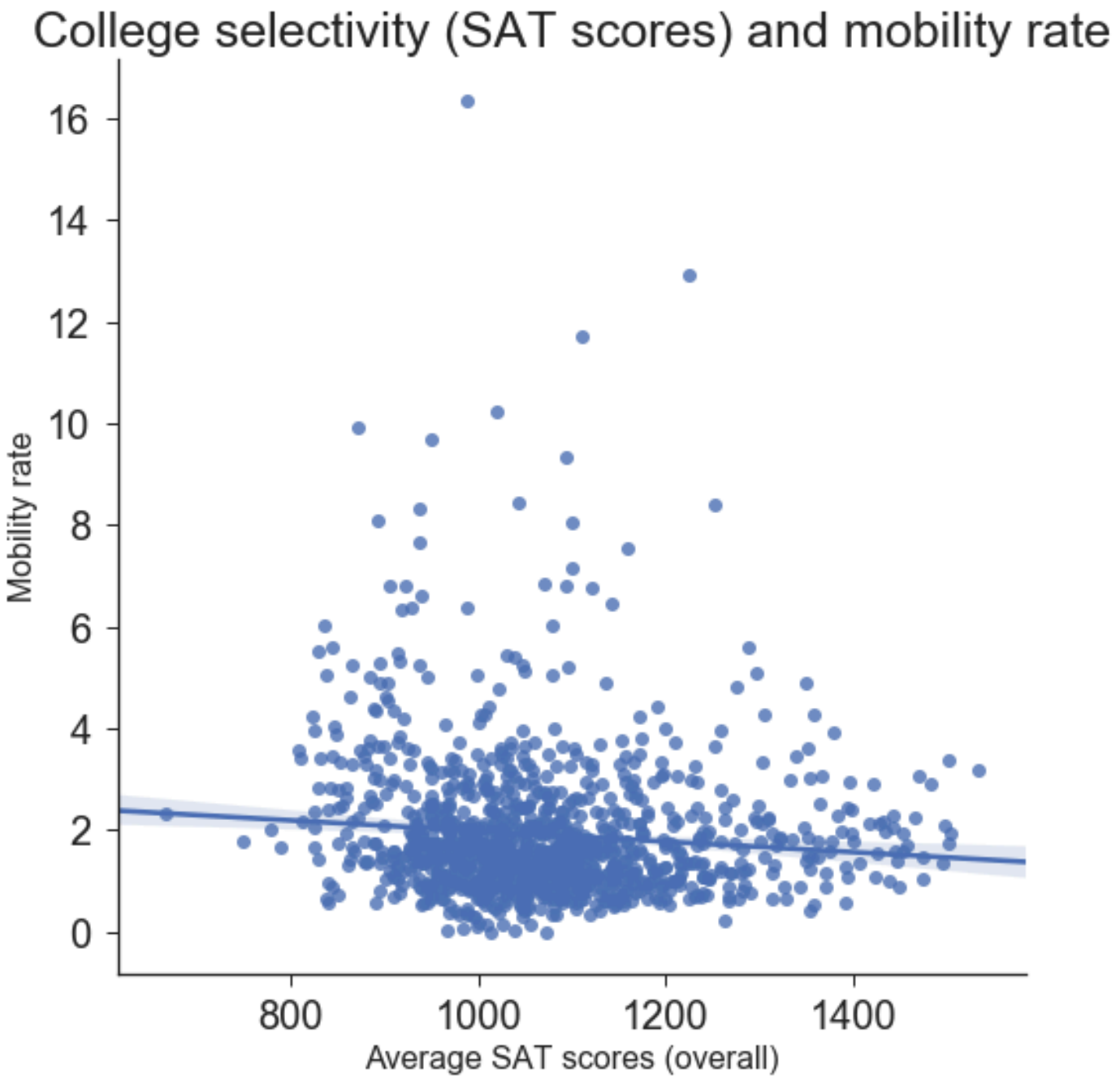


Rich students go to more selective schools.

A 10,000 dollar increase in family income increases SAT scores by on average 13.6 points and this relationship is statistically significant

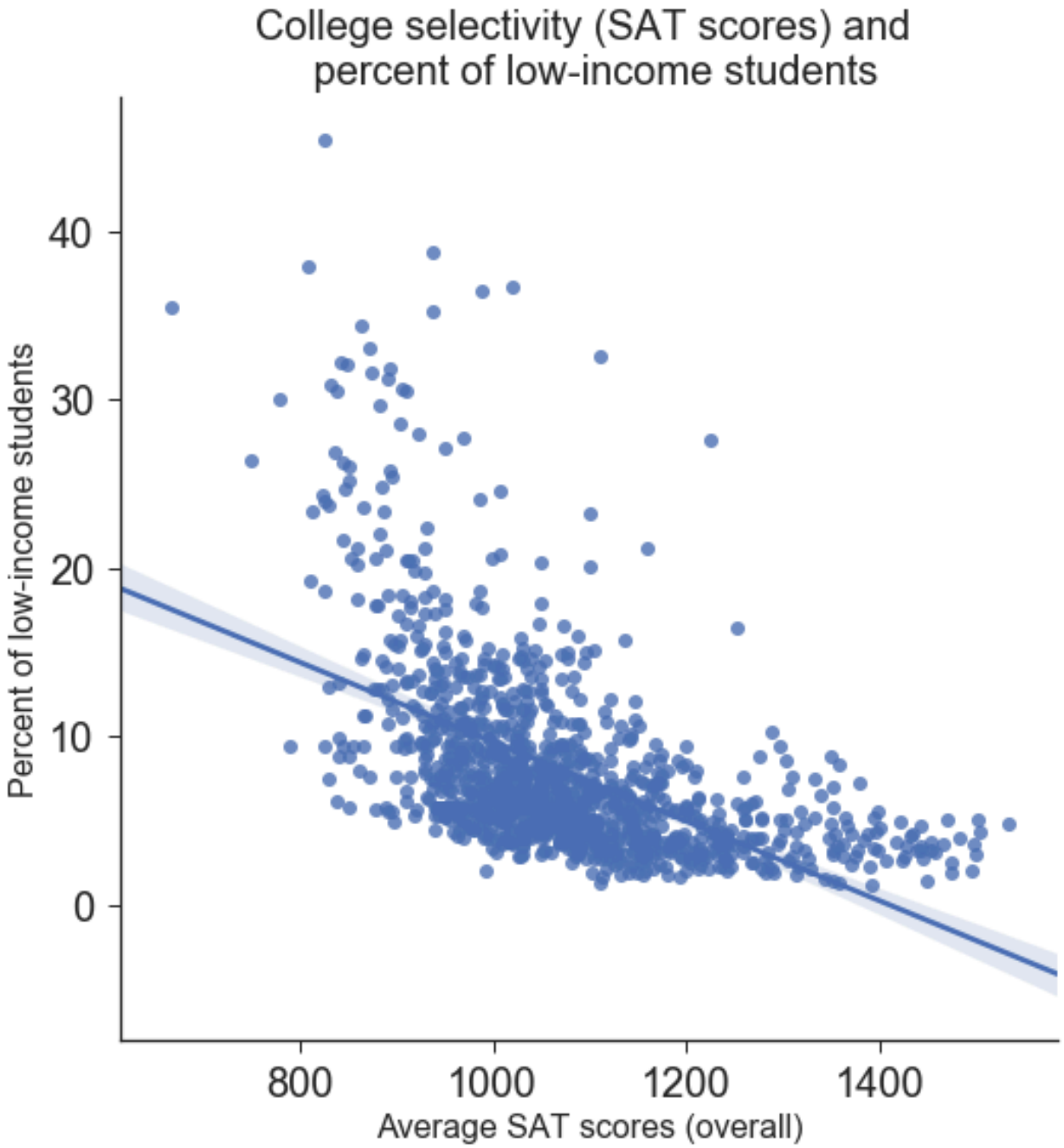


College selectivity (SAT scores) has a small effect on mob rate



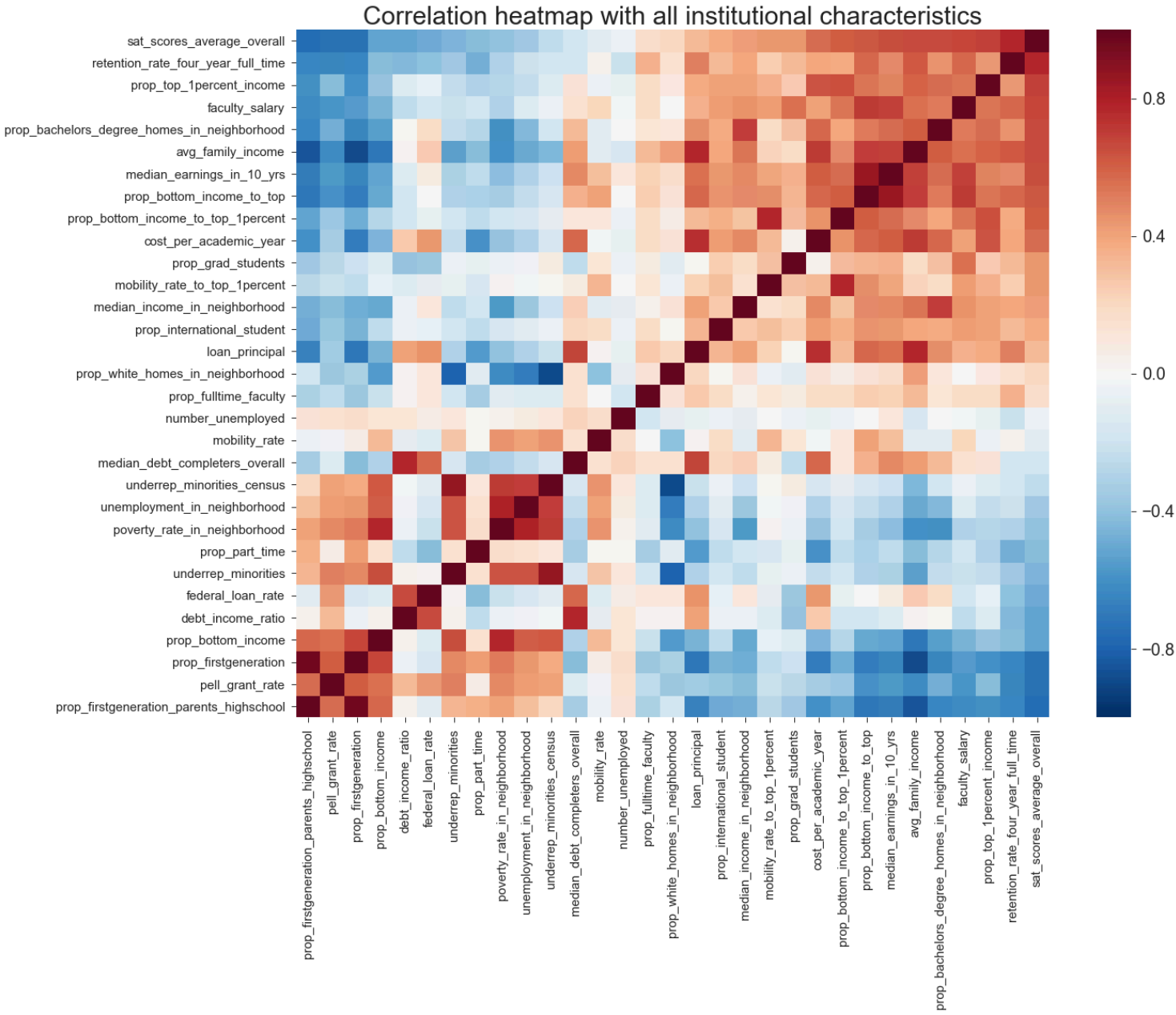
College selectivity (SAT scores) has large positive effect on conditional mobility rate

- Low-income students who can get into more selective colleges have a much higher probability of upward mobility.



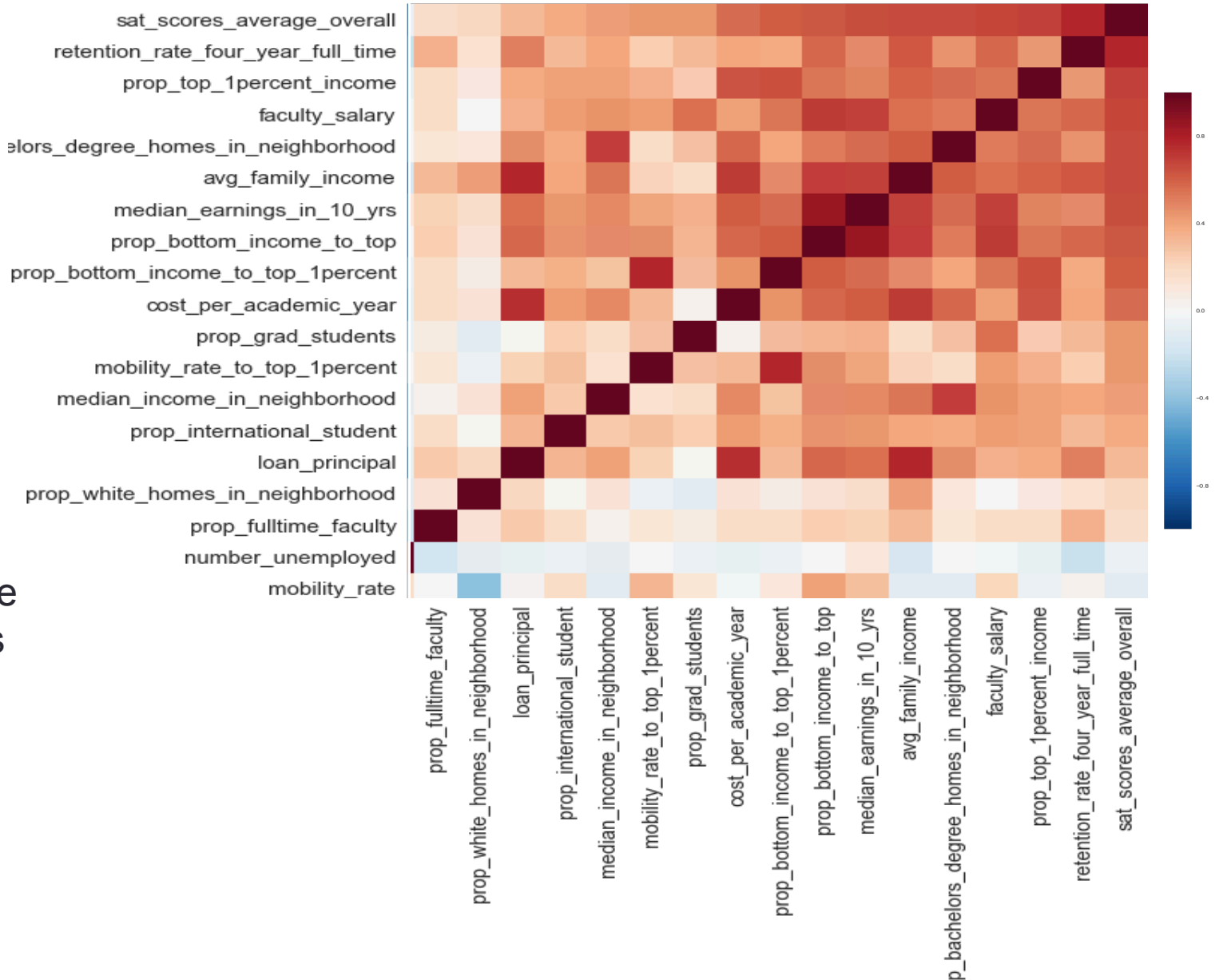
Two categories of college

A correlation heat map between all institutional characteristics depicts two data quadrants representing essentially two types of educational institutions.



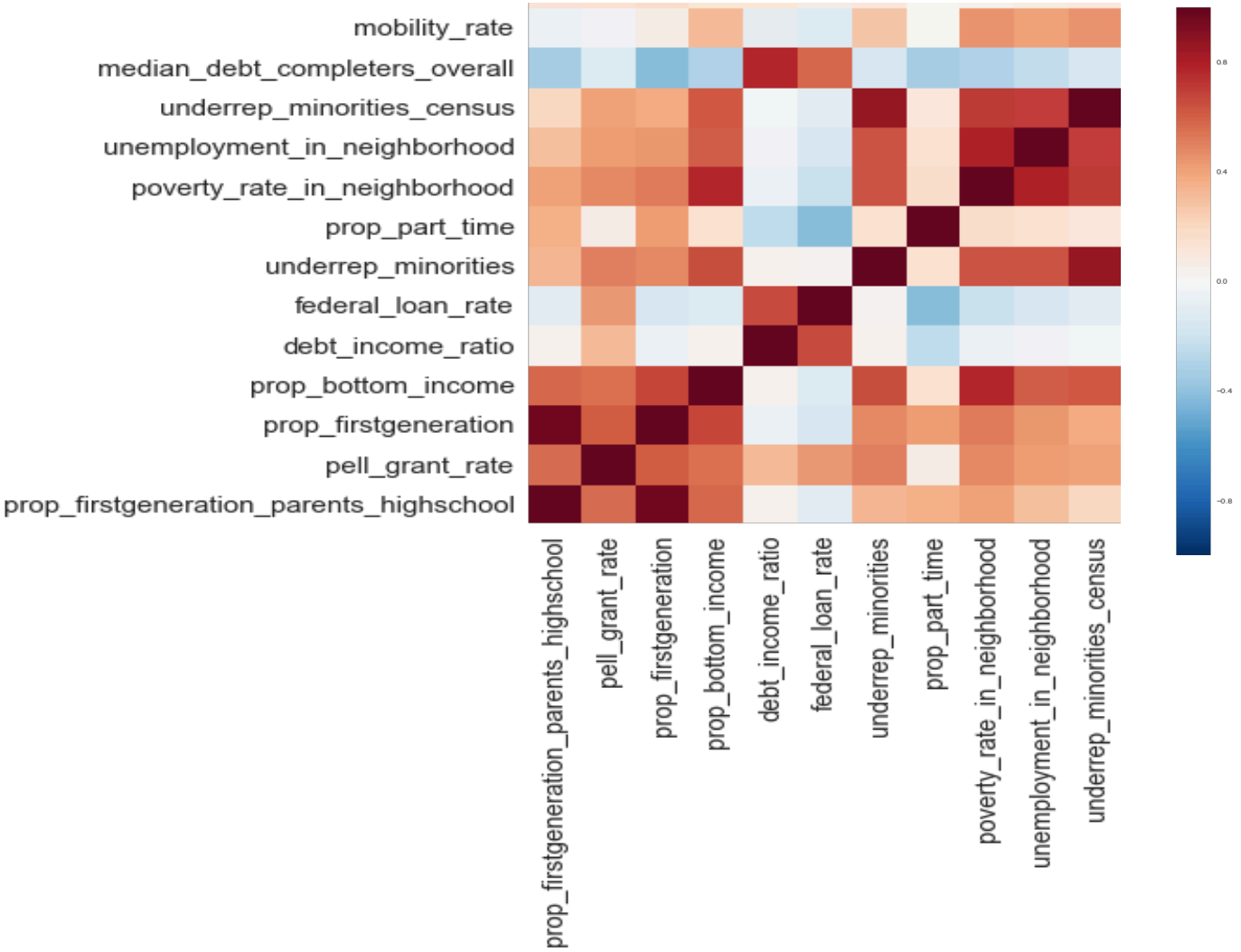
Elite colleges

Elite selective universities: those that are selective (high average SAT scores) which have students from top income brackets, high tuition and are in affluent white neighborhoods with low diversity.

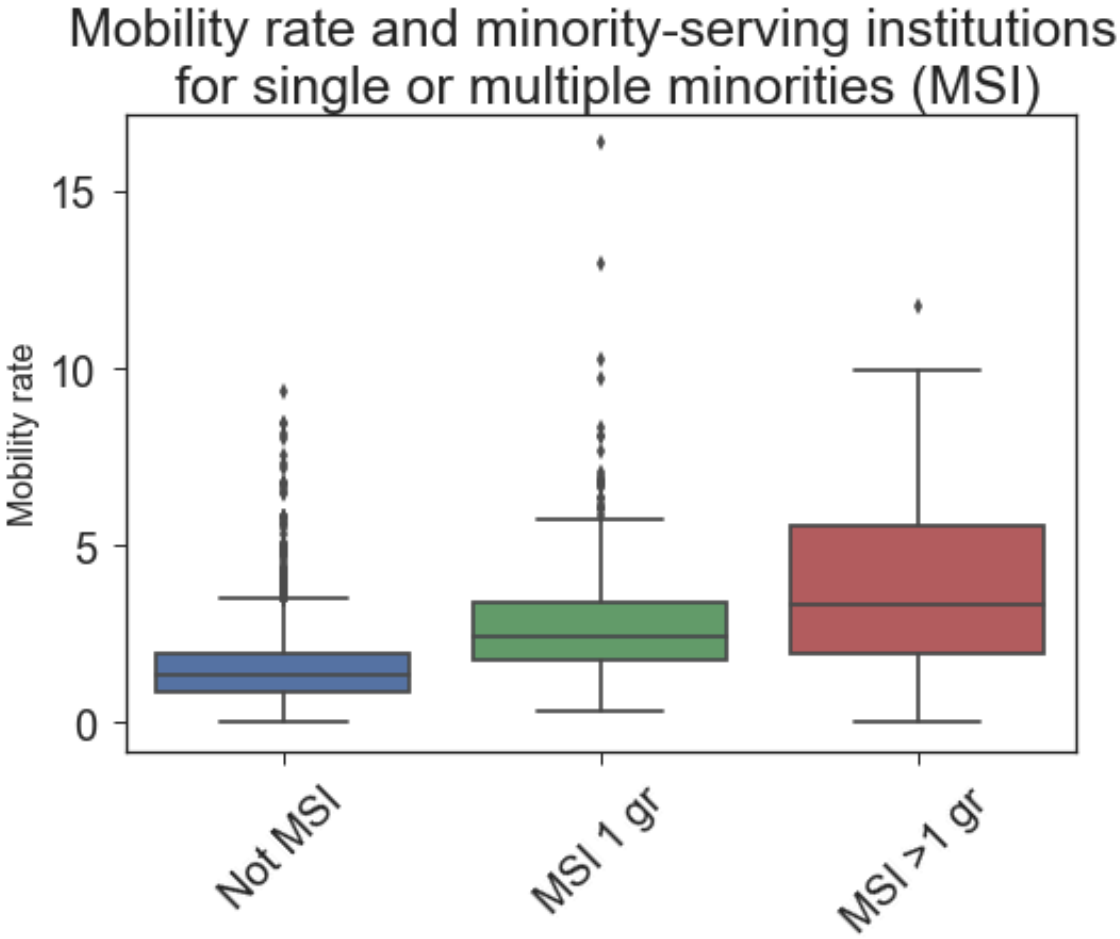


High-access colleges

High-access institutions: The colleges with more minorities, more first generation students, lower family incomes and in lower income neighborhoods.

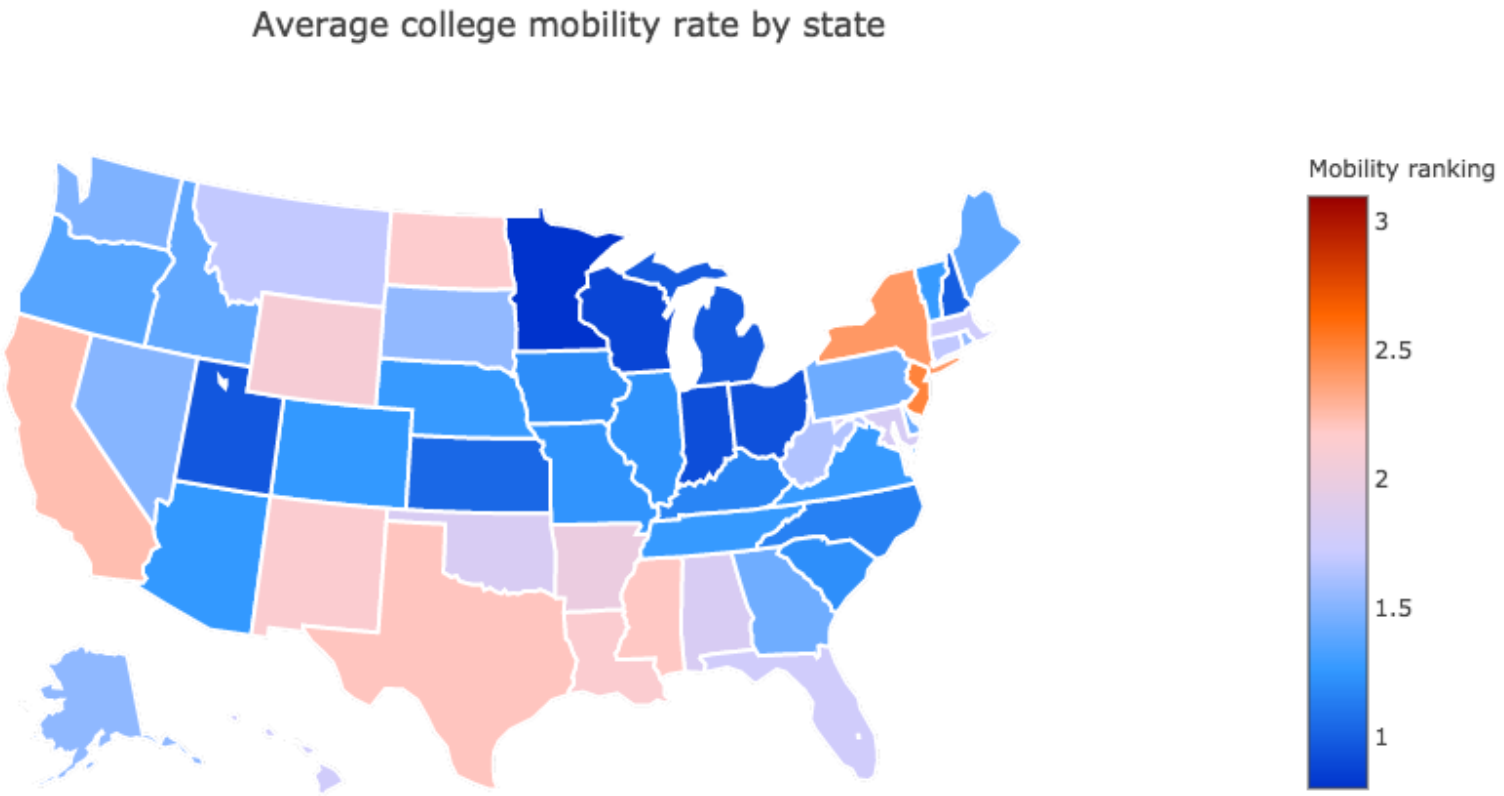


Diversity matters for mobility



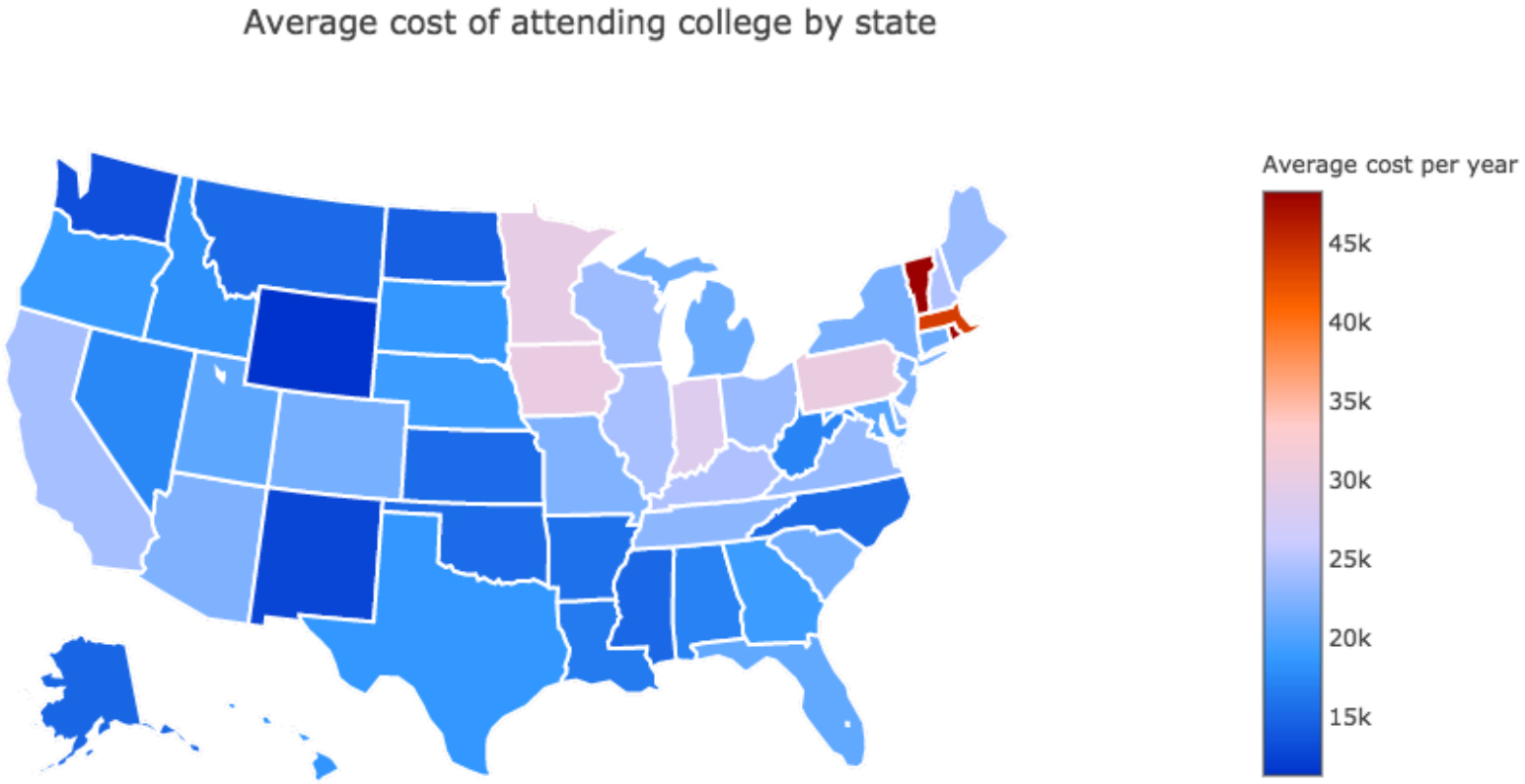
Minority serving institutions (MSIs) enroll a large proportion of minority students. These institutions make a difference for helping students at lower income levels rise up the economic ladder.

Geographic patterns in mobility rate



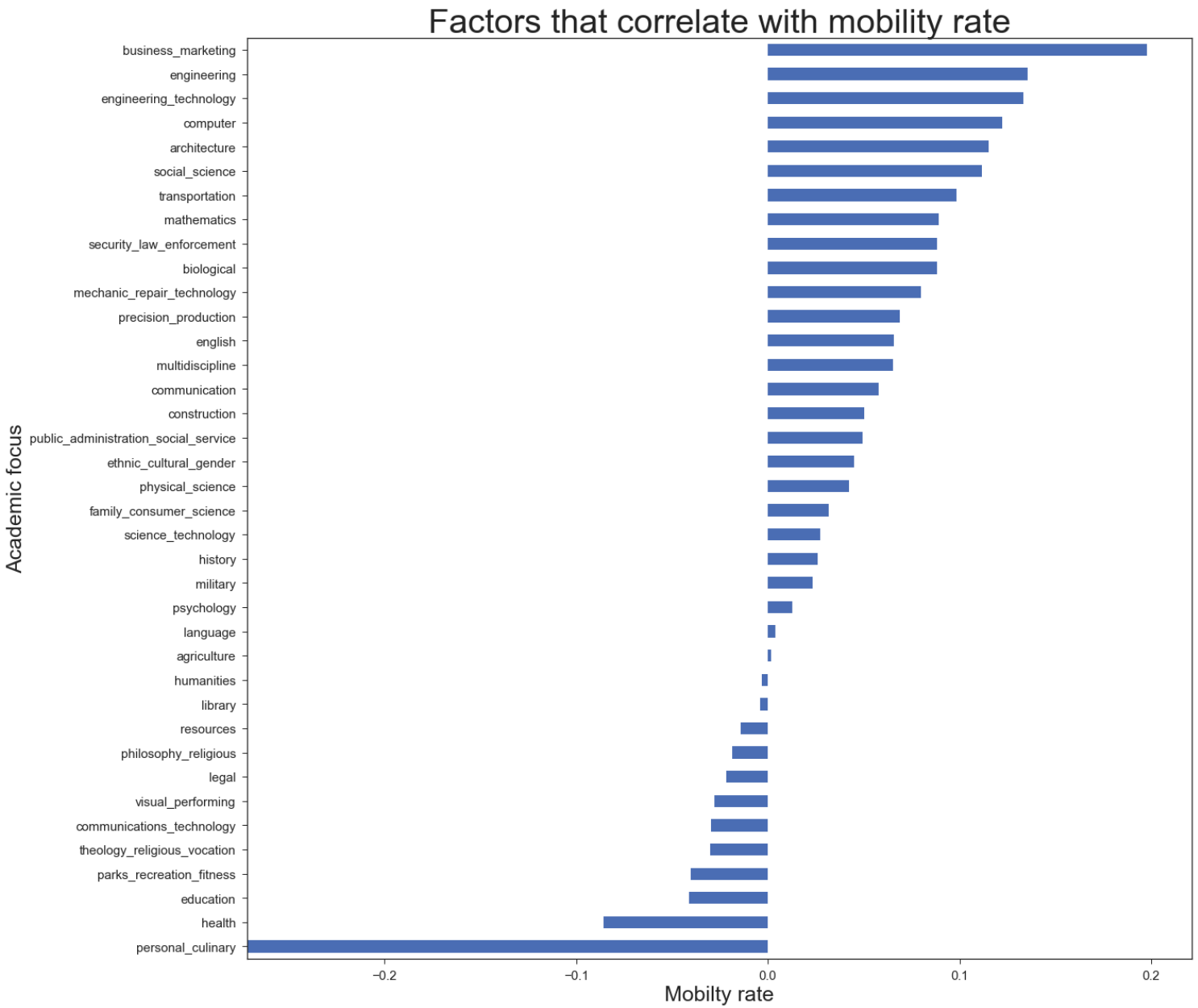
Colleges in mid-western state have surprisingly low economic mobility whereas those with high immigrant populations (NY, TX, CA) have higher mobility on average

Geographic patterns in college costs



Interestingly, colleges in mid-western states are more expensive on average! This might be why mobility rates are lower for institutions in these states

Colleges with an academic focus on business and science are better at promoting upward mobility



Top Colleges for Economic Mobility

Vaughn specializes in engineering and is right next to La Guardia in NY. Rated as the most racially and economically diverse college in America from the *US News Report Ranking*

	name	mobility_rate
	Vaughn College Of Aeronautics And Technology	16.357975
	CUNY Bernard M. Baruch College	12.938586
	City College Of New York - CUNY	11.723747
	CUNY Lehman College	10.235138
	California State University, Los Angeles	9.918455
	CUNY John Jay College Of Criminal Justice	9.691438
	MCPHS University	9.343507
	Pace University	8.432647
	State University Of New York At Stony Brook	8.412747
	New York City College Of Technology Of The Cit...	8.334076

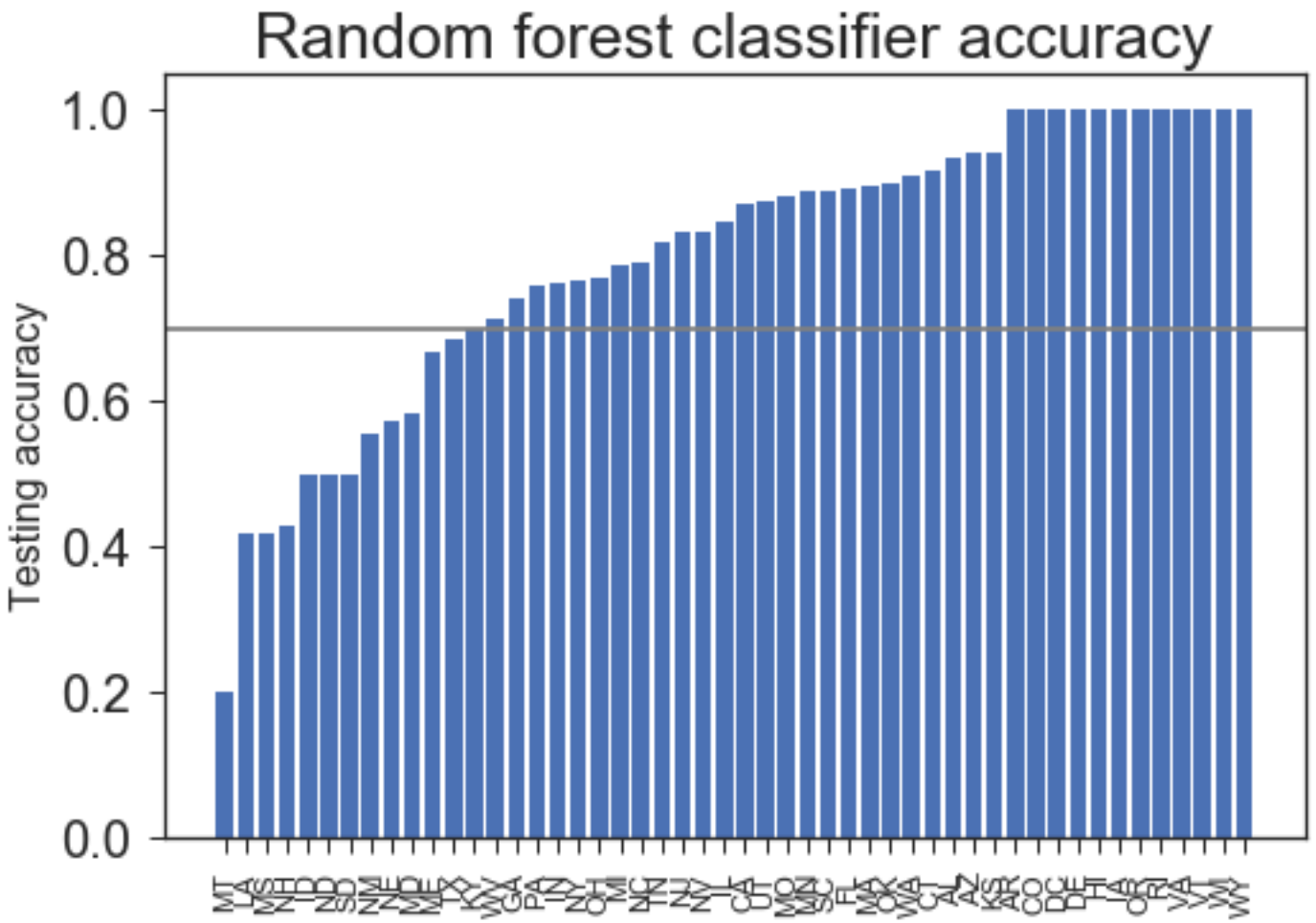
College recommender

Pipeline:

- K means clustering to produce even clusters of similar colleges on:
  - SAT scores
  - Cost of attending
  - Programs offered
- Run model for each state, splitting colleges into approximately 2-20 clusters per state
- Train a Random Forest classifier on these clusters to match student data to colleges
- Display the options with the mobility opportunities that the college provides.

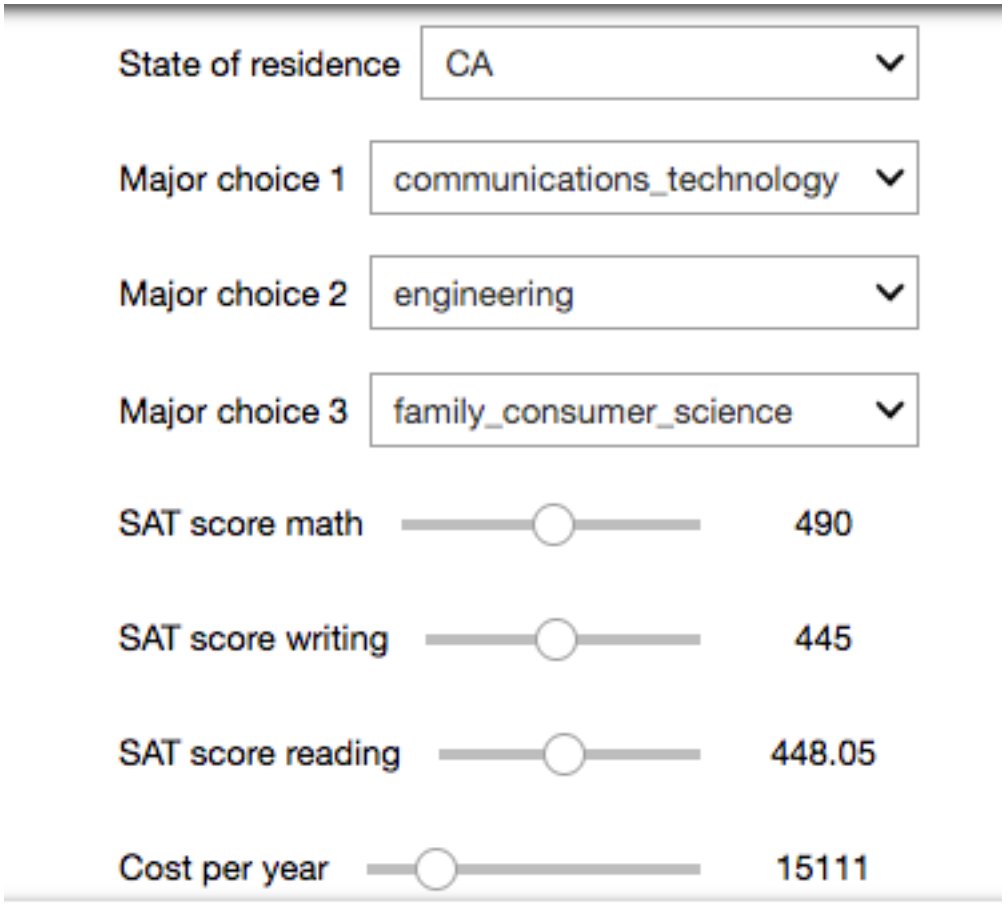
Random forest classifier

Clustering and classification work really well (70-100% accuracy) for most states (high accuracy) but poorly for a few



College recommender app

Interactive app (ipywidgets) for the college recommeder



College recommender output

Interactive plot_ly graph displaying output colleges sized by mobility (name on hover