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).

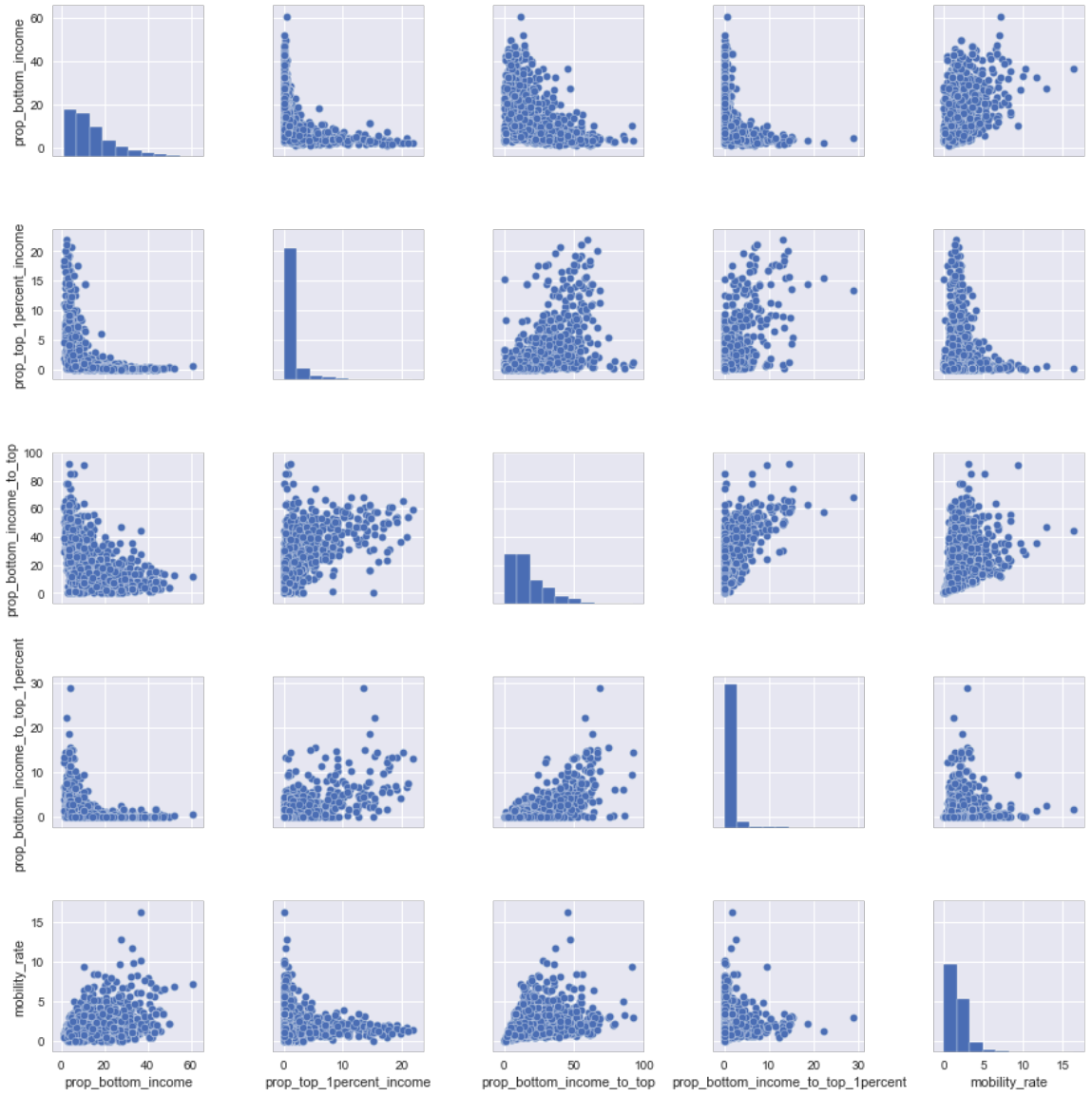
   (b) conditional mobility (% of the low income students who, once admitted reach the top 20% post-graduation)

Data Wrangling

- Imported 120 variables using College Scorecard card's API.
- Plotted missing data on a heat map to look for chunks of missing data
- Focused my analysis on data from 2013, because it is the most complete, recent dataset.
- Neighborhood census data is missing for 2013 and imputed from 2005.
- Replaced integers of categorical variables with meaningful strings from the data dictionary
- Imported the Equal Opportunity Project mobility ranking data from the website. These data are clean so I did not change anything.
- Concatenated the College Scorecard data and the Equal Opportunity Project data using the OPE_ID, which is a unique institution ID in both datasets
- Replaced column names with strings that are more intuitive, shorter and pythonic.
- Used pairplots to visually explore the data and look for any major problems and outliers

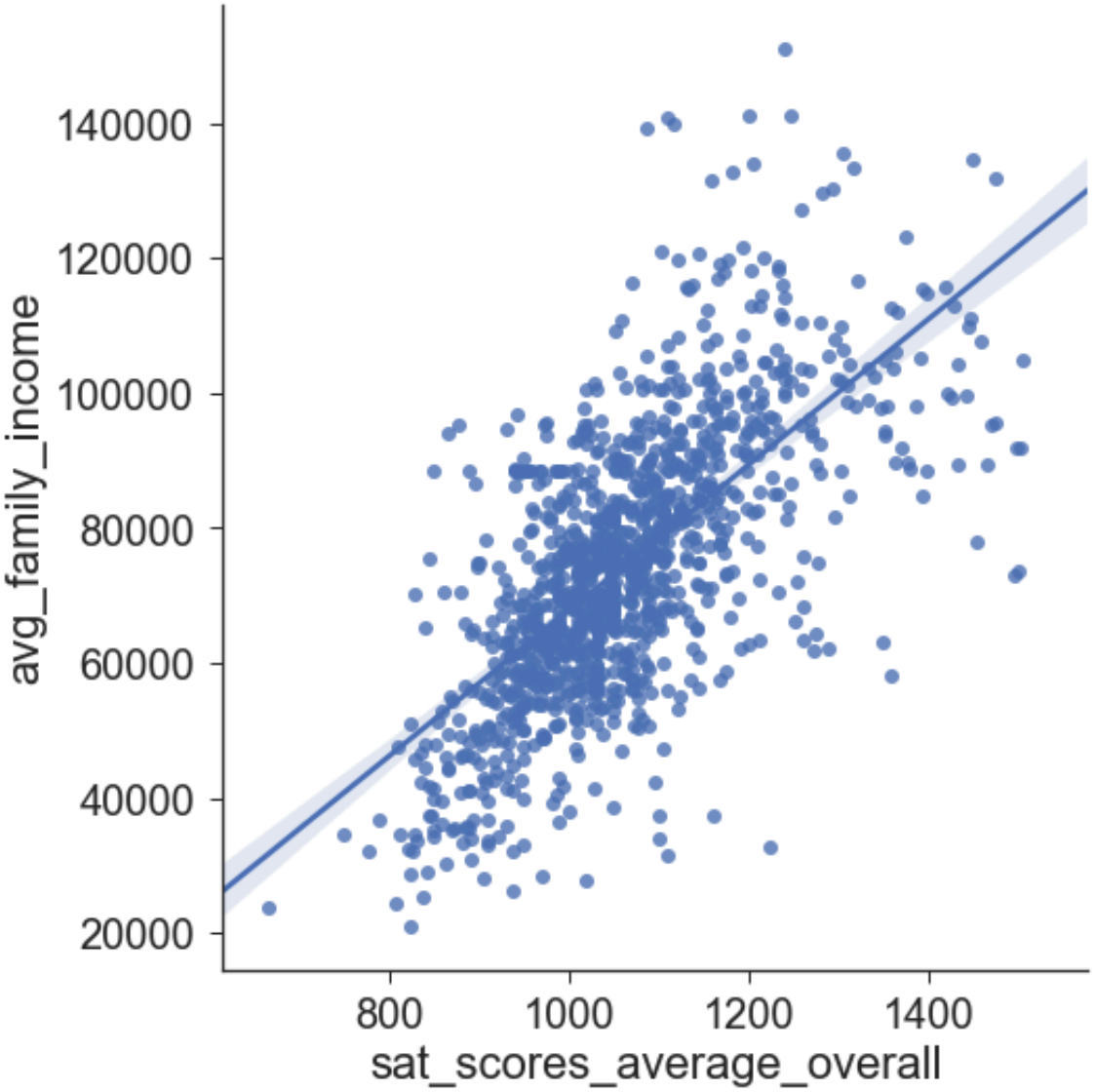
Exploratory Data Analyses

- Many of the metrics are proportions and are non-normal.
- Square-root transformed these for inferential statistics but it is better to fit these with alternative, non-Gaussian distributions



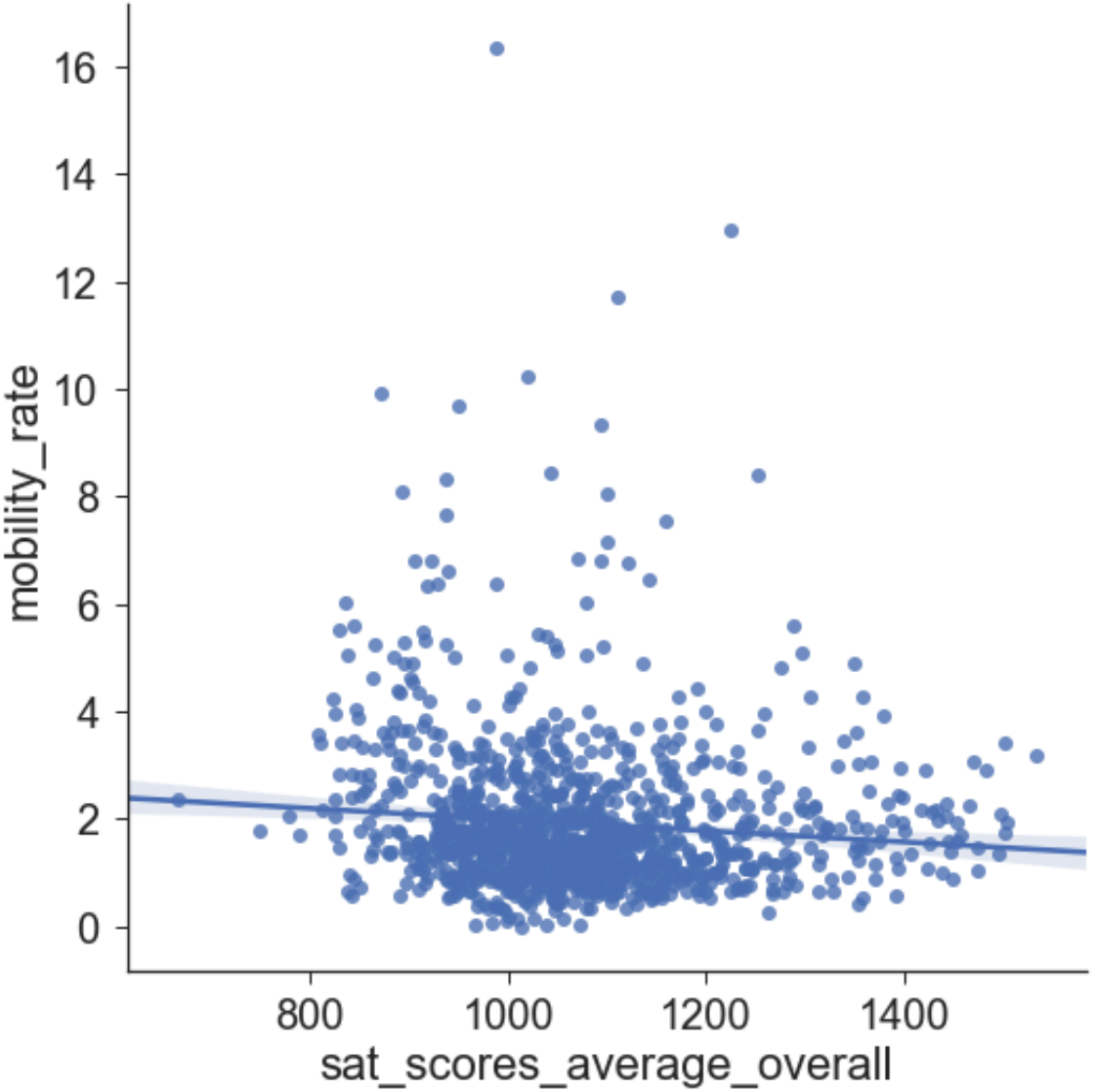
Exploratory Data Analyses

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



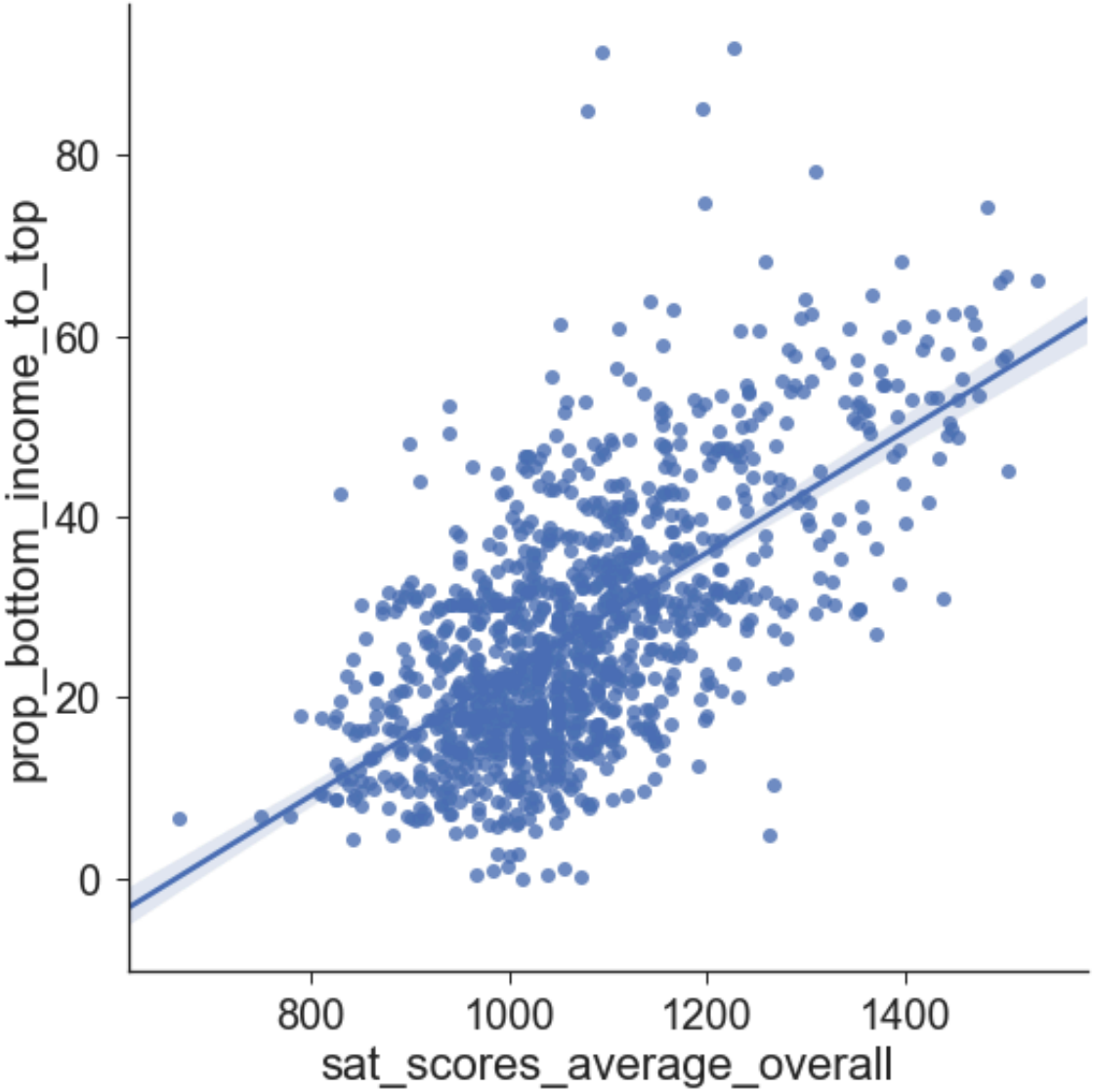
Exploratory Data Analyses

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

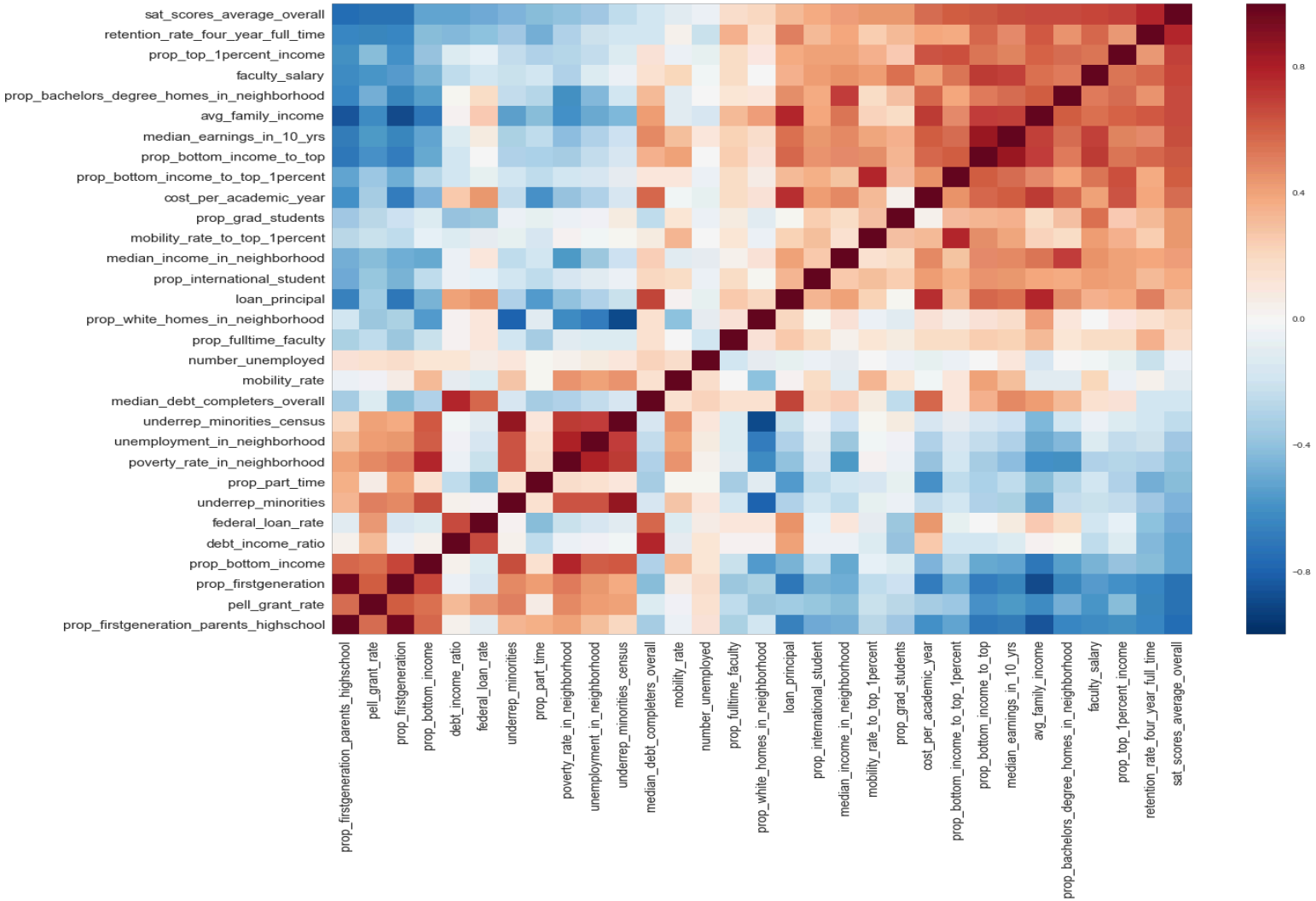


Exploratory Data Analyses

- 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.



Exploratory Data Analyses



Elite universities are doing a poor job of facilitating upward mobility. These universities don't contribute to the American dream.

Exploratory Data Analyses

OLS Regression Results

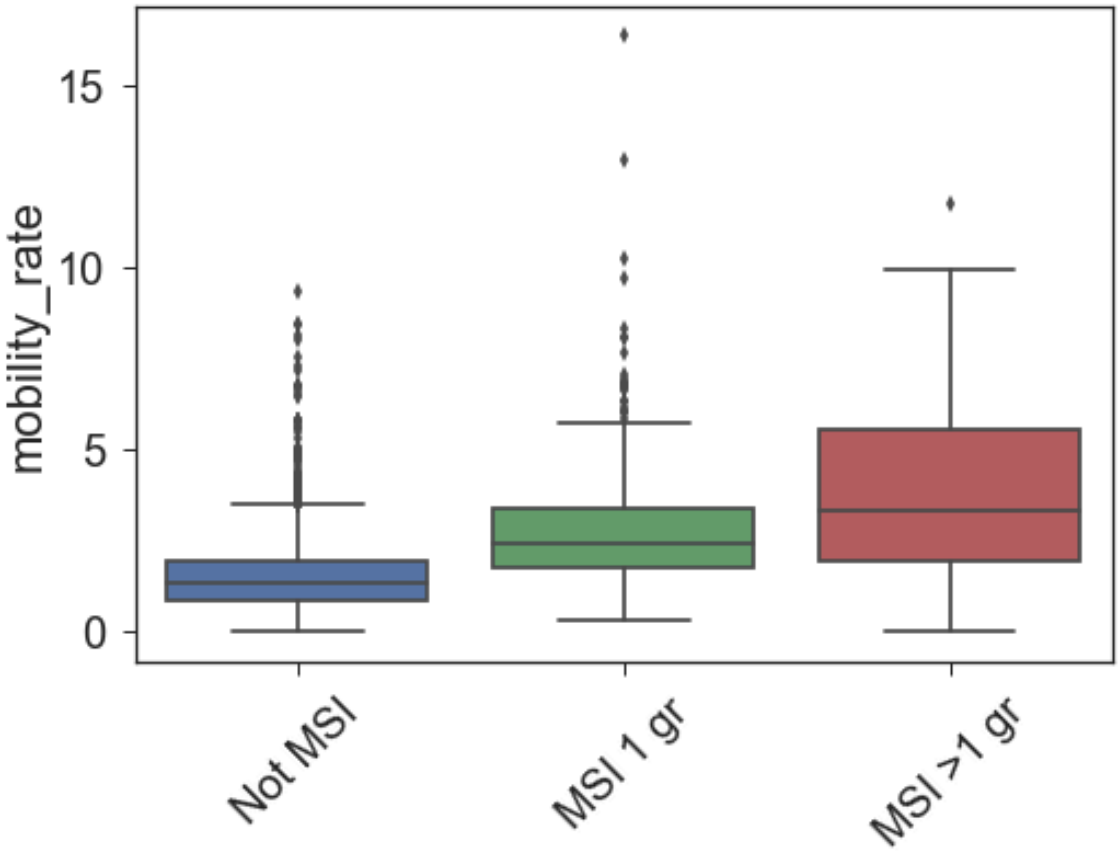
Dep. Variable:	mobility_rate	R-squared:	0.953
Model:	OLS	Adj. R-squared:	0.952
Method:	Least Squares	F-statistic:	2141.
Date:	Tue, 14 Nov 2017	Prob (F-statistic):	0.00
Time:	16:57:52	Log-Likelihood:	-256.37
No. Observations:	1178	AIC:	534.7
Df Residuals:	1167	BIC:	590.5
Df Model:	11		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[95.0% Conf. Int.]	
debt_income_ratio	-0.3049	0.081	-3.782	0.000	-0.463	-0.147
median_earnings_in_10_yrs	8.717e-06	1.4e-06	6.240	0.000	5.98e-06	1.15e-05
loan_principal	2.125e-05	2.91e-06	7.312	0.000	1.55e-05	2.69e-05
prop_firstgeneration	1.0403	0.119	8.778	0.000	0.808	1.273
underrep_minorities	-0.1342	0.102	-1.318	0.188	-0.334	0.066
underrep_minorities_census	0.0054	0.002	3.182	0.002	0.002	0.009
poverty_rate_in_neighborhood	0.0557	0.004	12.899	0.000	0.047	0.064
prop_white_homes_in_neighborhood	-0.0052	0.001	-5.516	0.000	-0.007	-0.003
prop_international_student	0.9024	0.222	4.069	0.000	0.467	1.337
prop_grad_students	-6.007e-06	3.81e-06	-1.576	0.115	-1.35e-05	1.47e-06
faculty_salary	3.458e-05	6.32e-06	5.468	0.000	2.22e-05	4.7e-05

Omnibus:	60.374	Durbin-Watson:	1.814
Prob(Omnibus):	0.000	Jarque-Bera (JB):	211.094
Skew:	0.018	Prob(JB):	1.45e-46
Kurtosis:	5.073	Cond. No.	1.23e+06

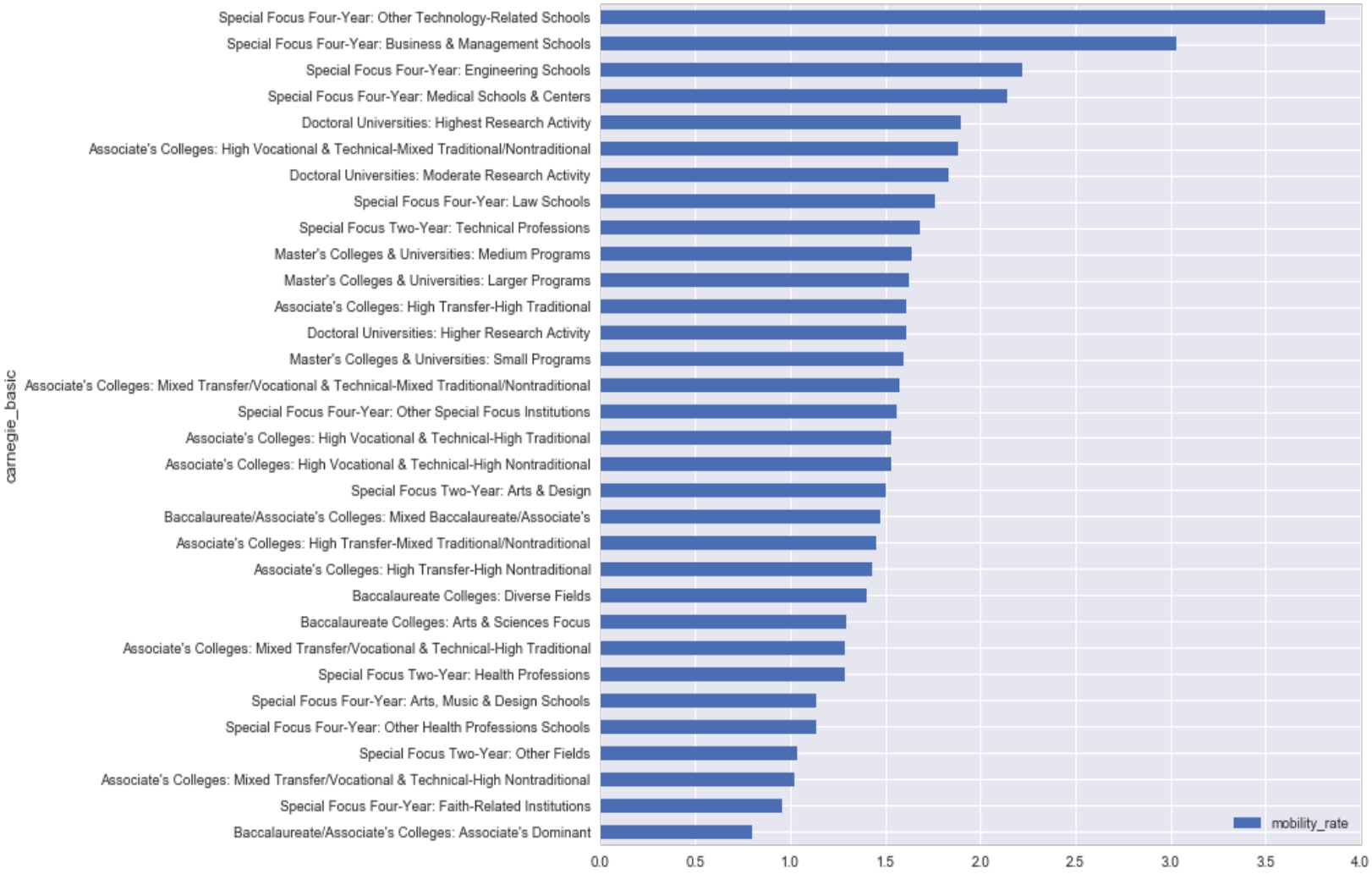
Schools that support a lot of first generation students increase mobility by almost 1%!

Exploratory Data Analyses



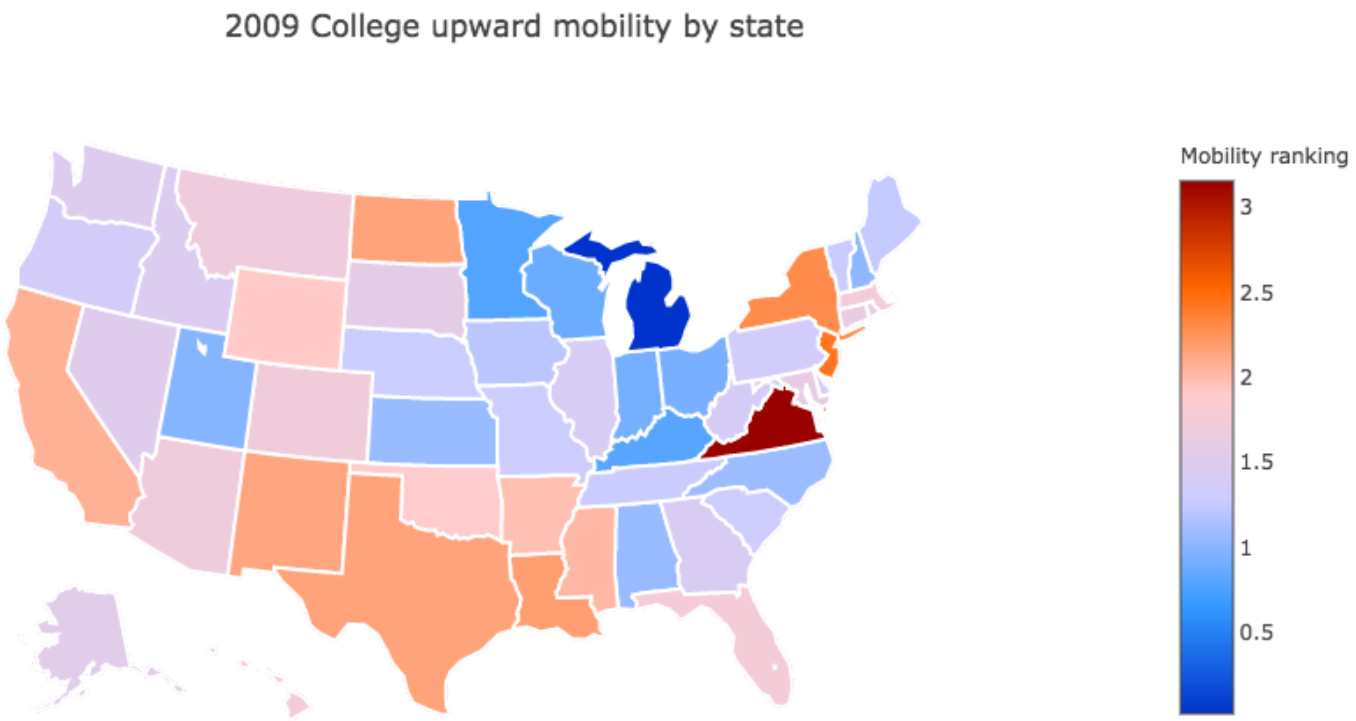
Diversity matters for upward mobility. Minority serving institutions (MSIs) enroll a large proportion of minority students. These institutions make a difference for helping students at lower income levels rise up the economic ladder.

Exploratory Data Analyses



Applied schools (Buisness, Management, Tech related) tend to be better at increasing economic mobility that those focused on the arts and religion

Exploratory Data Analyses



Colleges in some states (CA, TX, NY, ND,VA) have much better mobility ratings

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

Additional exploration

- Data on school academics. The college Scorecard database has information on the academic programs and the proportion of degrees granted in each program by institution. I would like to explore these data as metric for economic mobility

- Plot these data on a map at the county level (still figuring this out – I need to get a dataset that matches zip codes to counties)

Machine Learning methods

- Use ML regressions to identify key predictors of upward mobility

- Develop an algorithm to match students (based on SAT scores, ZIP code, family income, academic interests) to 10 schools based on location, cost and mobility. Not entirely sure how to do this. I've seen this solution that uses Bayes Factors and decision trees. I don't fully understand the code or how I can incorporate mobility as a factor but need to explore this further. https://www.kaggle.com/apollostar/which-college-is-best-for-you-part2/code