

Determinants of income after college graduation



Alice Chang, Ben Thomas, Olek Wojcik



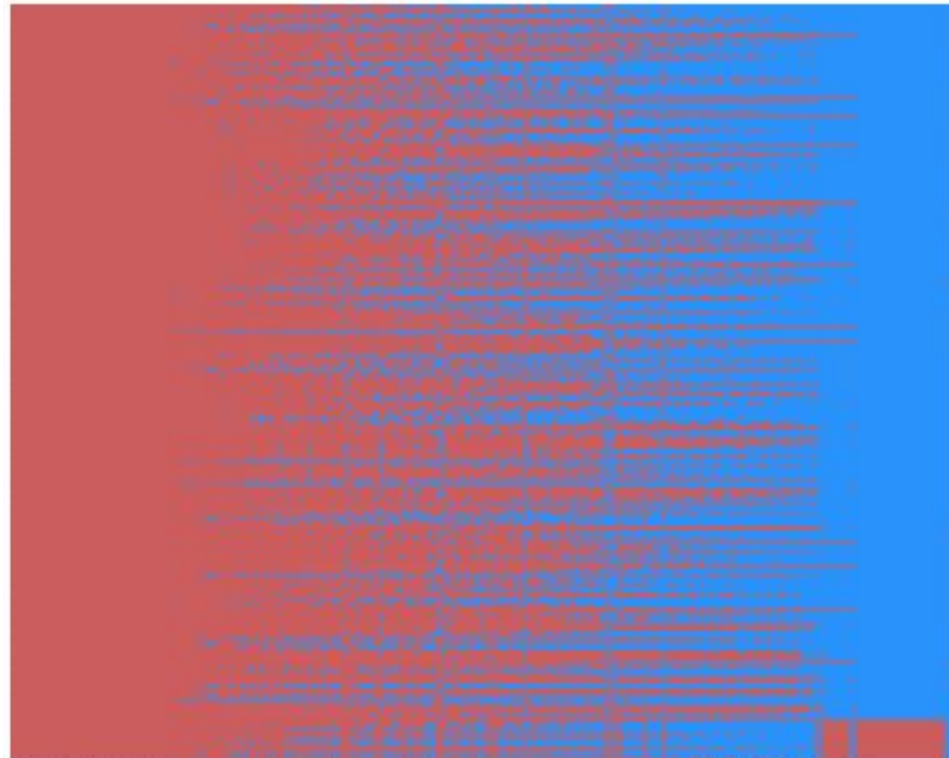
The question

What factors are important for determining the income of graduates from a given college in the years following their graduation?

The data

- College Scorecard (US Department of Education)
 - 1996-1997 through 2016-2017
 - Released in 2013
 - Reports earnings 6-10 years post-graduation
- 2014-15 dataset
 - 1,977 variables
 - 7,703 observations (individual colleges)
- Required some heavy-duty cleaning
- Several limitations

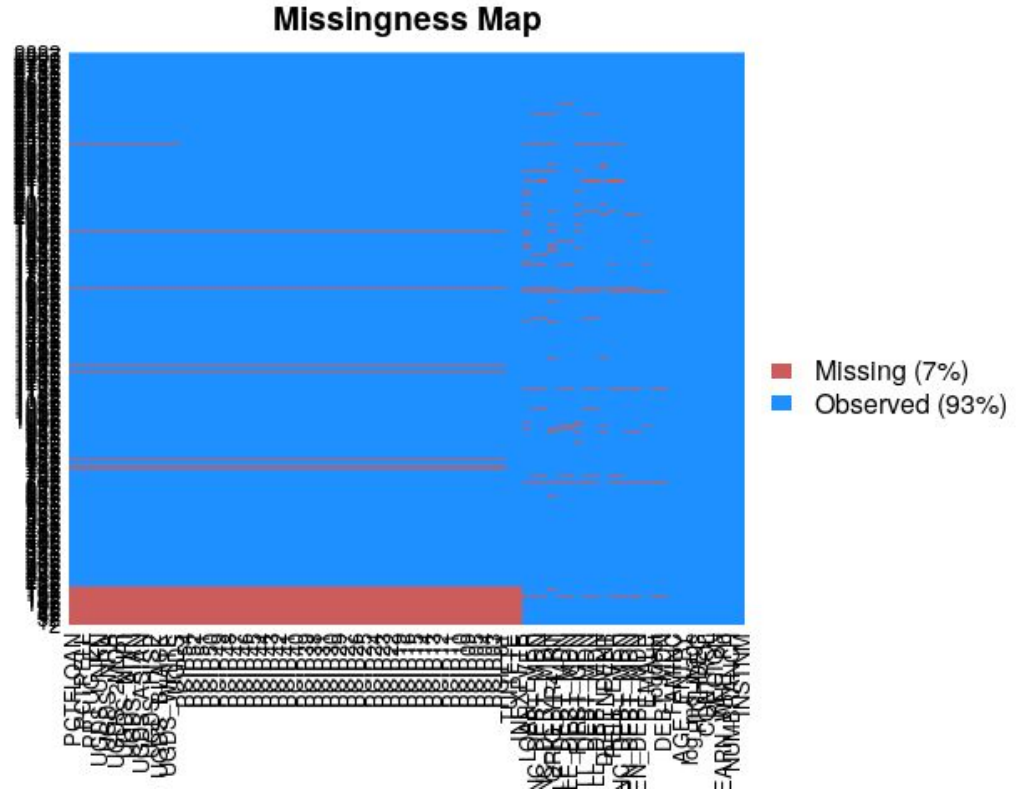
Missingness map



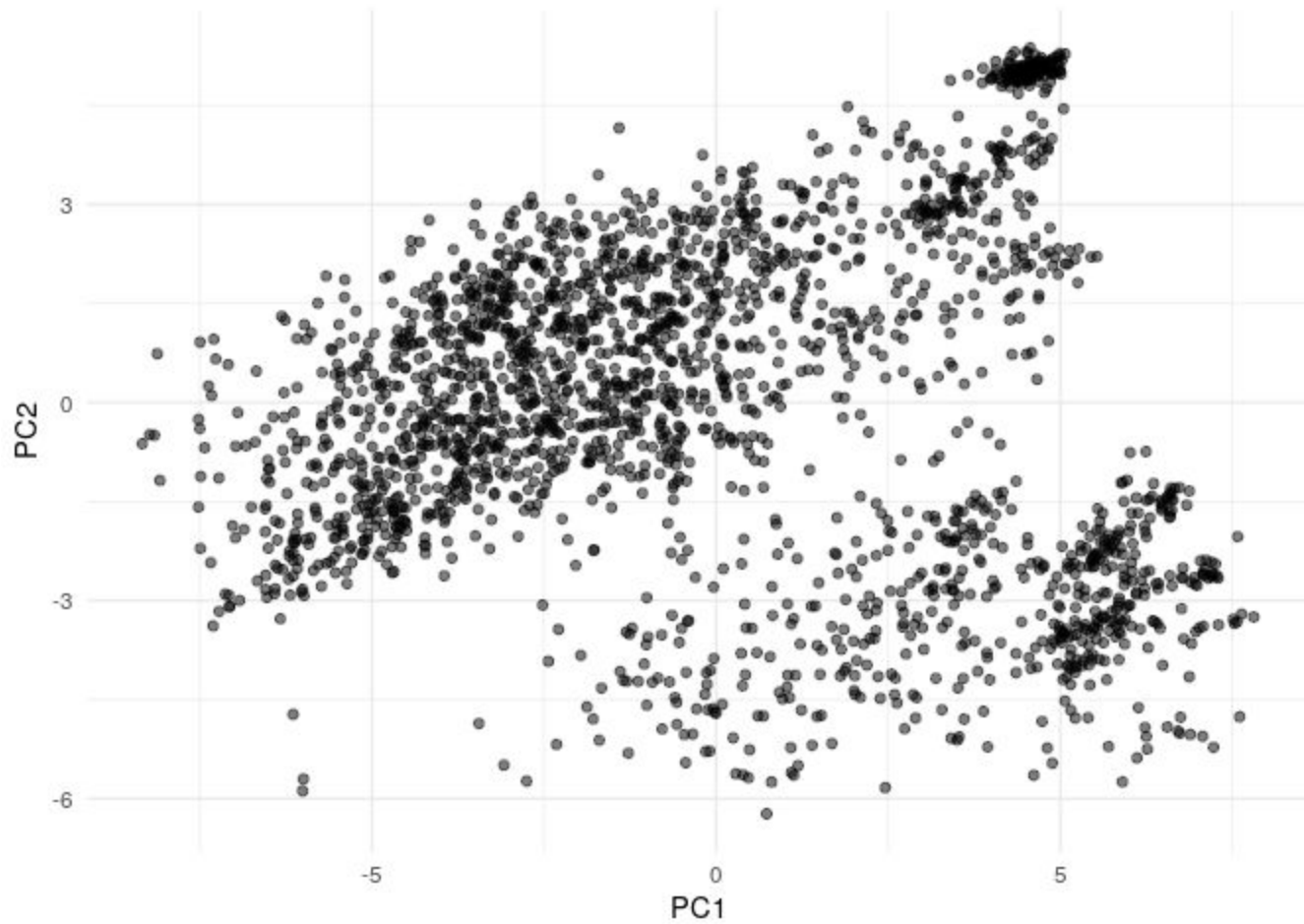
■ Missing (61%)
■ Observed (39%)

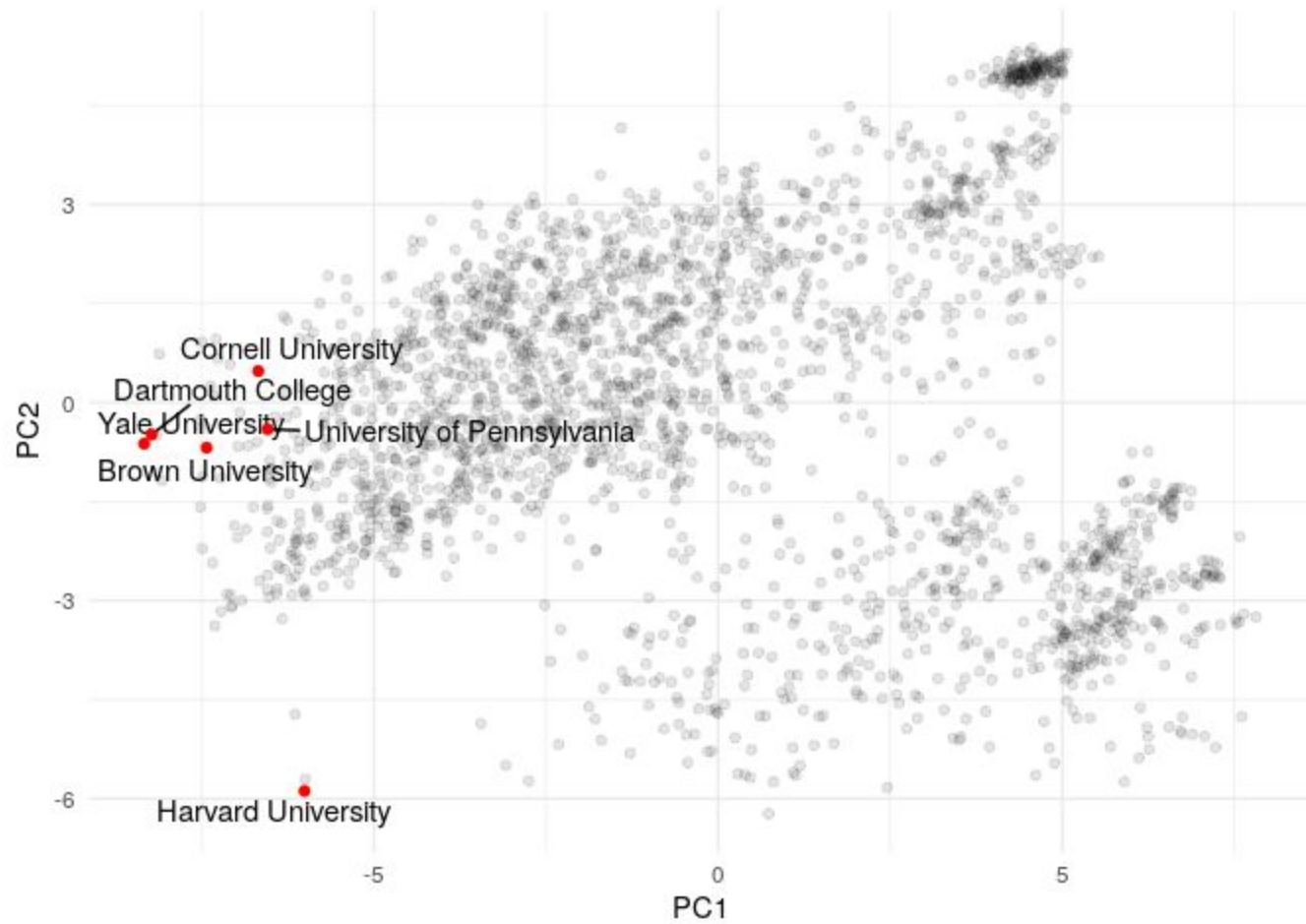
Choosing Variables

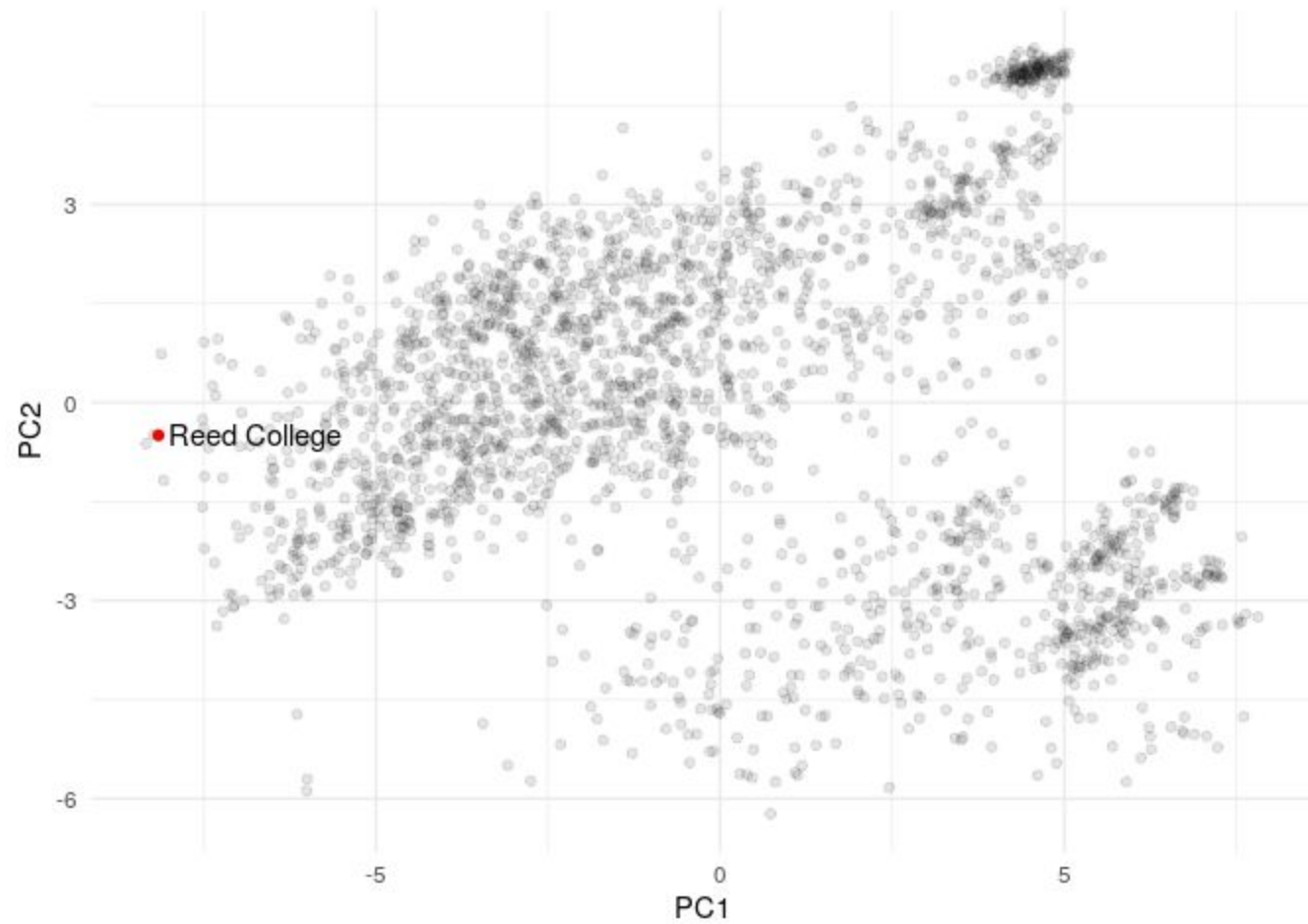
- We went through the data and chose relevant variables with a decent amount of data in order to have working models
- Our final dataset had a much better ratio of missingness



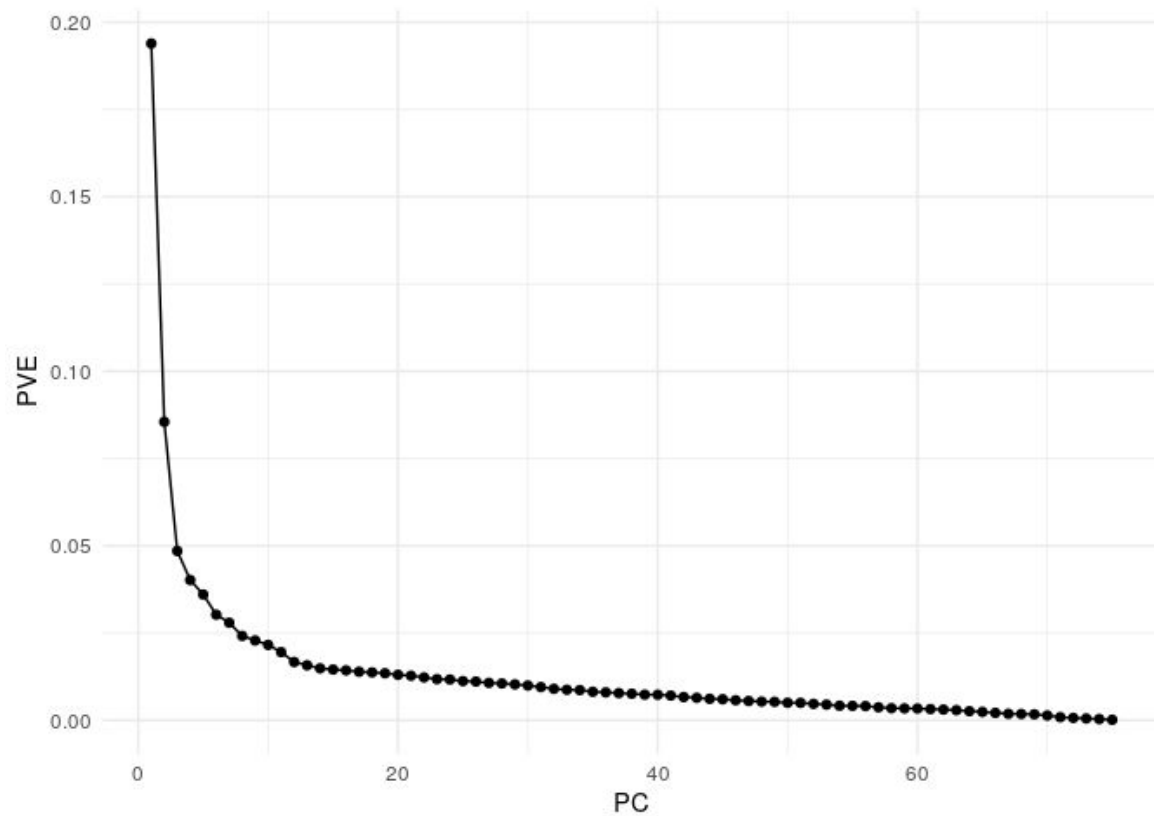
PCA







Scree Plot



The Two First PCs

PC1

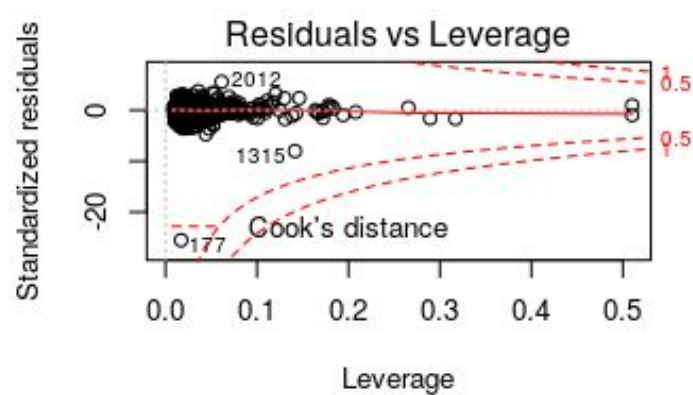
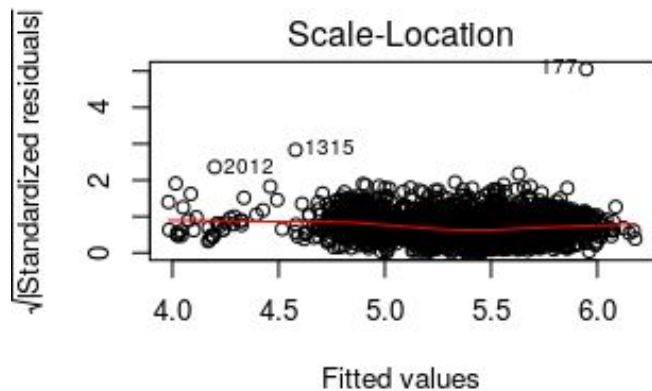
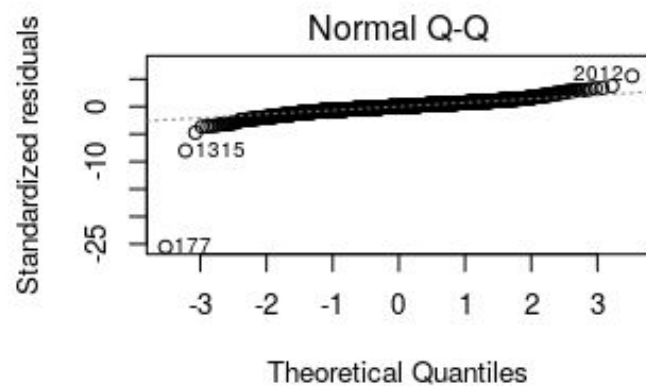
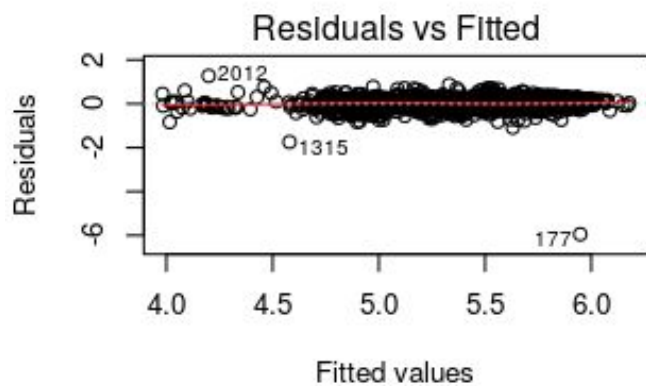
1. FIRST_GEN
2. AGE_ENTRY
3. PELL_EVER
4. PCTPELL
5. WDRAW_ORIG_YR4_RT
6. LO_INC_DEBT_MDN

PC2

1. PCIP15
2. NUMBRANCH
3. PCIP11
4. PCTFLOAN
5. PCTPELL
6. PCIP10

Linear Model Most Significant Variables

- Percentage of degrees awarded in Natural Resources And Conservation.
- +Percentage of degrees awarded in Computer And Information Sciences And Support Services.
- Percentage of degrees awarded in Education.
- +Percentage of degrees awarded in Engineering.
- +Percentage of degrees awarded in Engineering Technologies And Engineering-Related Fields.
- Percentage of degrees awarded in Theology And Religious Vocations.
- Percentage of degrees awarded in Visual And Performing Arts.
- Share of white students
- +Age of entry
- Region 9, Outlying Areas (AS, FM, GU, MH, MP, PR, PW, VI)
- +share of part-time students
- share of students with a federal loan

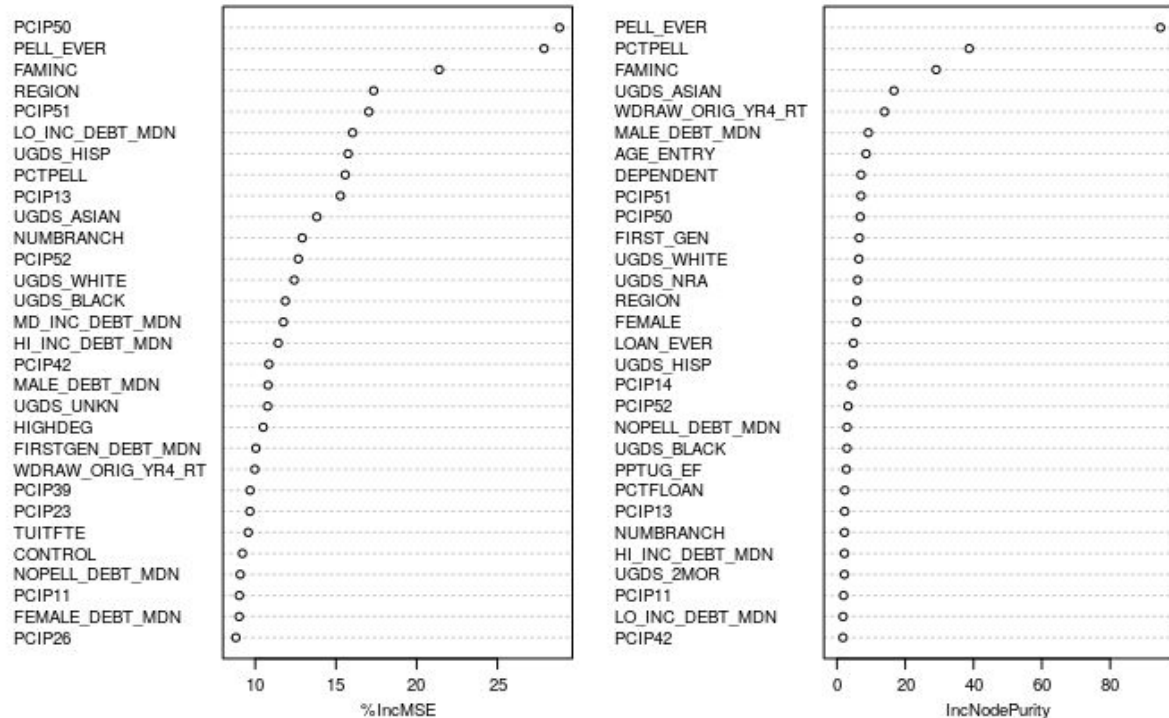


Lasso Model

- 12 share of degree variables
- 3 racial demographic variables (Asian+, biracial-, unknown+)
- Share of female students
- Pell grant/debt variables
- Age
- Region dummy variables (Southeast-, Southwest+, Outlying Areas-)
- Graduate school dummy variables

Random Forest

rf.college





Conclusions

Further Study

-
- Other years
 - Other datasets
 - Other countries
 - In-state comparisons
 - Micro individual student analysis
 - Other kinds of educational institutions
-