

An Analysis of National Student Loan Data

Surya Menon, Madelyn
Rubenstein, Camellia
Debnath, and Nanditha
Sundararajan



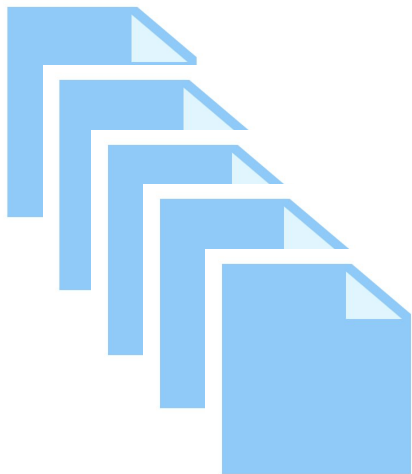
College Scorecard Dataset

- Generated by the US Department of Education to increase transparency on how well different schools are serving their students
- Provides a vast amount of information on various school metrics
- Captures school data from 1996-2017
- Original dataset contains 2,146 variables (!)



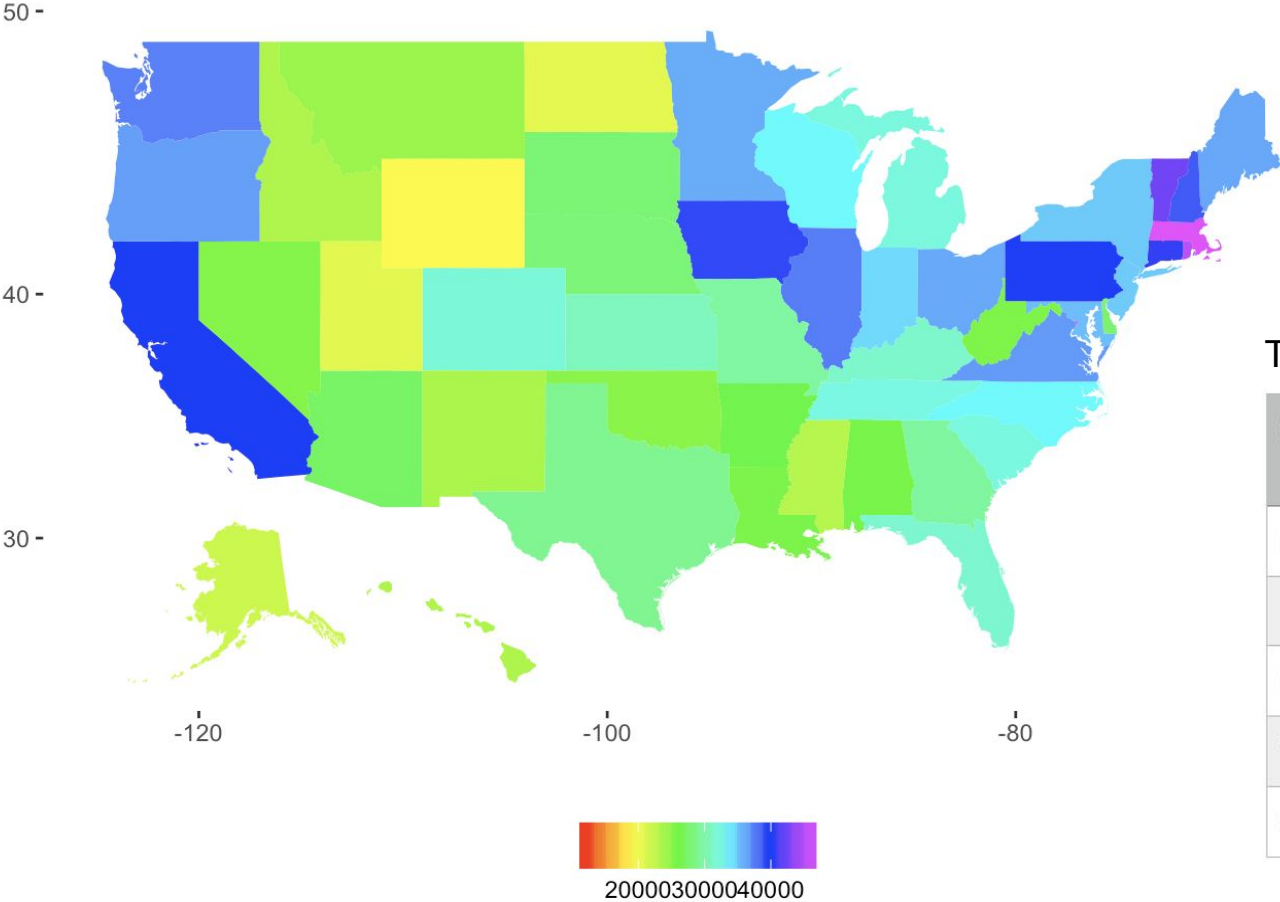
Research Goal: Can we predict the percentage of students at a university that will repay their loans within 3 years?

Exploratory Data Analysis



- Joined data sets for 5 years
 - Filtered on public/private and 4-year institutions
 - Subsetted down to 81 variables
 - Looked at local (top 5 public and private MA universities, New England region) and national data
-

Average cost of attendance by state, 2016-2017



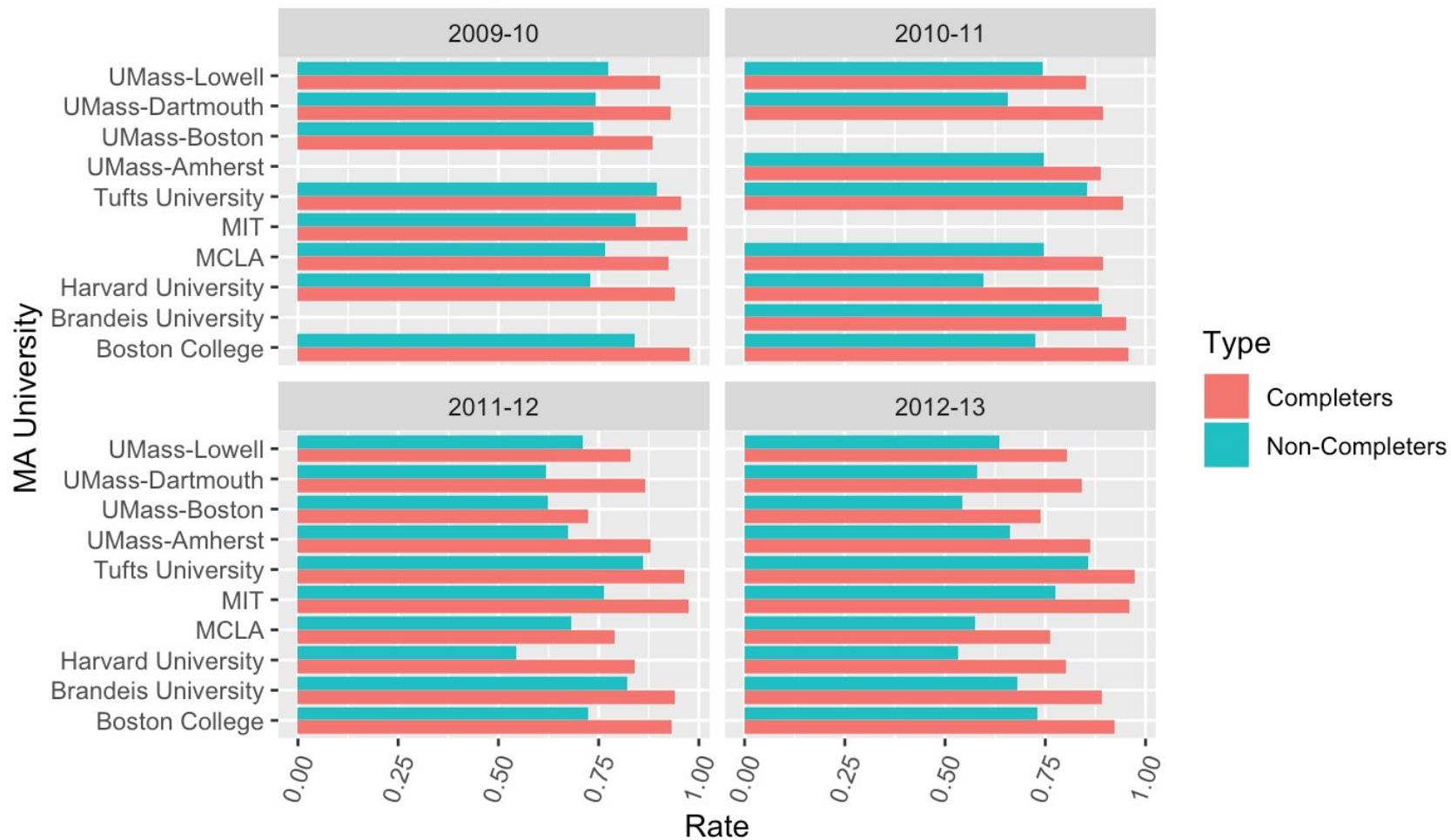
Top 5 States:

State	Average Cost
DC	47597.38
Massachusetts	46977.03
Rhode Island	45472.00
Vermont	42408.13
Connecticut	40956.41

Average % of Pell Grants for New England States



3 Year Repayment Completers vs. Non-Completers



Modeling

Response variable: 3-year repayment rate for completers

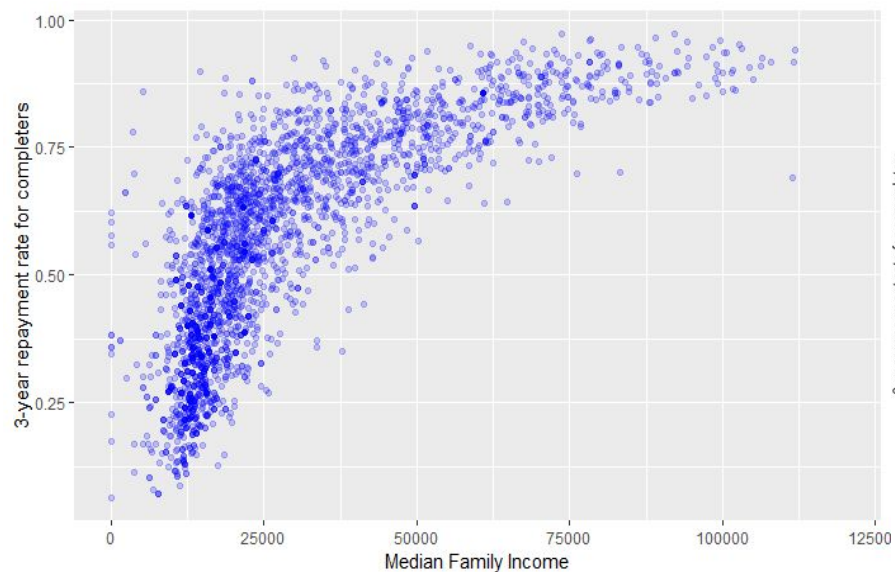
We performed EDA for some potential predictor variables and chose the following:

1. Debt to earnings ratio
2. **Percentage of undergraduate students who received Pell Grants**
3. Median earnings of federally aided students after graduation
4. **Cost of attending the institute**
5. Median family income

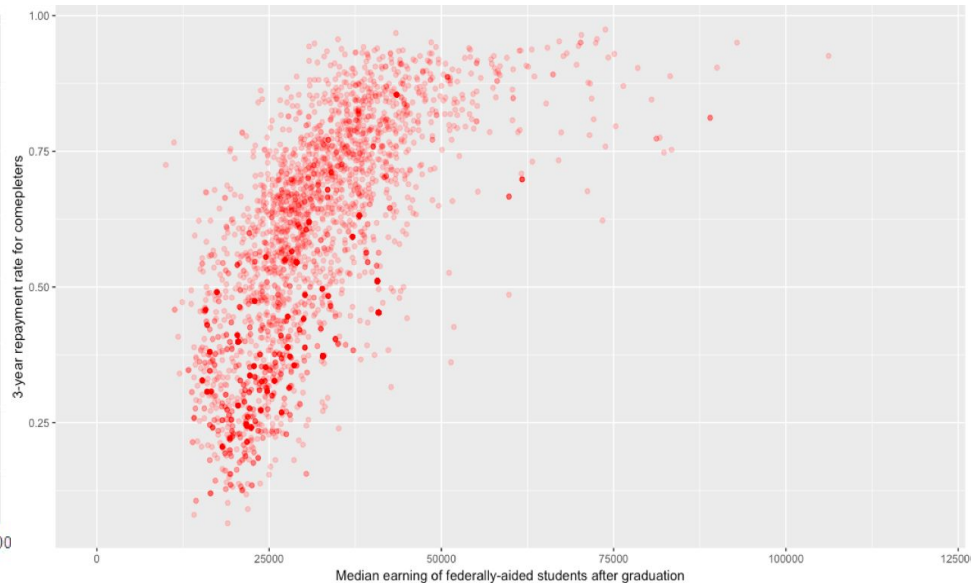
```
MODEL <- lm(COMPL_RPY_3YR_RT ~ DEBT_TO_EARN, + MD_EARN_WNE_P8  
+ PCTPELL + log2(COSTT4_A) + log_MD_FAMINC, data)
```

Variable Selection

Reference year: 2013-14

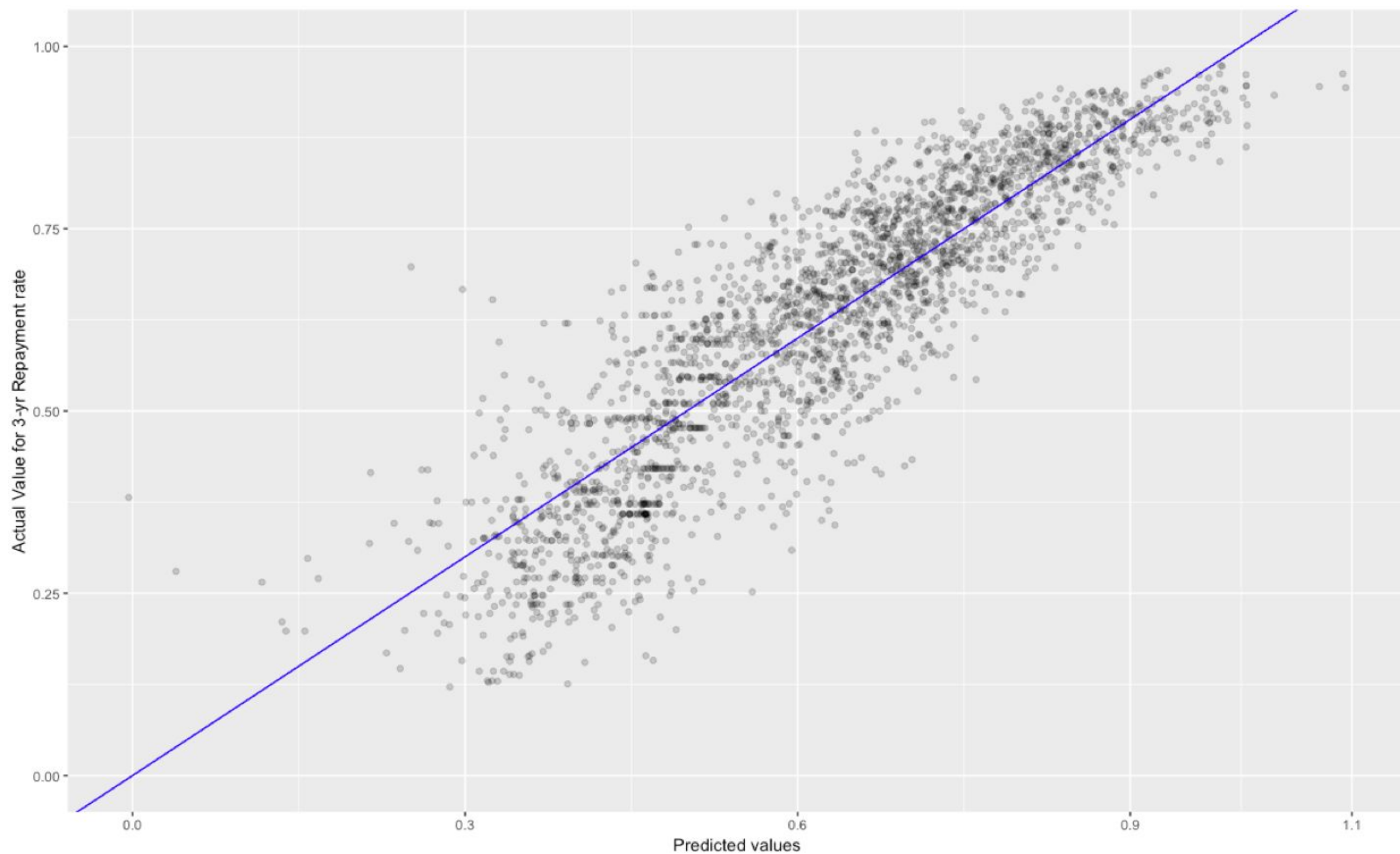


Correlation Coefficient: **0.7642012**



Correlation Coefficient: **0.6575527**

Results



Test RMSE= 0.2108707

Conclusions

Percentage of undergraduate students who received Pell Grants

Students with grants aren't burdened with as much student debt as those who have taken out federal loans

Median earnings of federally aided students after graduation

Higher salary = better ability to pay off loan faster

Cost of attending the institute

Higher cost = more difficult/takes longer to pay off loans

Median family income

Determines how much federal loan money a student will receive and dictate how long it could take them to pay it off

Discussion

- Aimed to choose intuitive variables that are important for predictions
 - Weaknesses of the data set
 - Lack of consistency throughout years
 - Null data
 - Financial information only for those receiving loans
 - Future models: Using different repayment rates in modeling (1, 3, 5, or 7)
 - Obama vs. Trump Administration data capture
 - Plans for the Trump administration to capture more detailed and complete financial data
-

Questions?

According to my
calculations,
I will have my student
loans paid off 5
years after I die.



somee cards
user card

References

College Scorecard:

<https://collegescorecard.ed.gov/data/documentation/>

Getting Repayment Rates Right:

<https://www.americanprogress.org/issues/education-postsecondary/reports/2018/07/10/453199/getting-repayment-rates-right/>
