Analysis of Completion Rate and Ownership of U.S. Higher Education Institutions

Group Member:

Kejiang Yao

Xiaozhu Zhao

Yu Qiu

College of Professional Studies, Northeastern University

Professor: Roy Wada

Date: 02/14/2022

Introduction

Target audience:

College admission committee

According to Virginia public universities,

There are negative effects of rising student costs on completion rate in higher education.

According to the U.S. Department of Education,

The completion rate and ethnicity between public colleges and universities is significantly different.

According to the U.S. News & World Report,

Completion rate is one of the key factors to the university's ranking.

Questions to solve:

1. How does students' financial burden influence the completion rate of US higher education institutions?
2. What are the differences between public and private universities regarding ethnicity and completion rate?
3. Does schools' ownership and region have influences on schools' completion rate? (What factors affect completion rate?)
4. Could completion rate and ownership of US universities be well predicted by ethnicity, commitment, degree level and region?

Purpose:

- Solve the questions
- Provide suggestions for U.S. Universities Admission Committee

Dataset Introduction & Variable Selection

Data Set

College Scorecard

from the U.S. Department of education

from 1997 to 2015

- **132,402 observations**
- **66 variables**

Data Cleaning & Preparation

1. **Filter out** variables with missing values over 50%
2. **Impute** numerical variables with column mean or median
3. **Remove** rows with categorical missing values
4. **Create 2 new variables**:
   - Region: transformed from States
   - Ownership: transformed to 2 types
5. **122,443 observations, 19 variables**

2 Outcome Variables

- **Completion Rate**
- **Ownership of Institution: Public or Private**

13 Predictors

- Region: 4 regions
- Median Debt
- Highest degree granted
- Predominant degree
- Average Beginning Age
- % of Female Students
- Undergraduate Enrollment
- % of White Students
- % of Black Students
- % of Hispanic Students
- % of Asian/Pacific Islander
- % of Native American/Alaska Native Students
- % of Part Time Students

Median Debt vs Completion Rate (by year)

Figure 1

Average Median Debt from 2000 to 2015

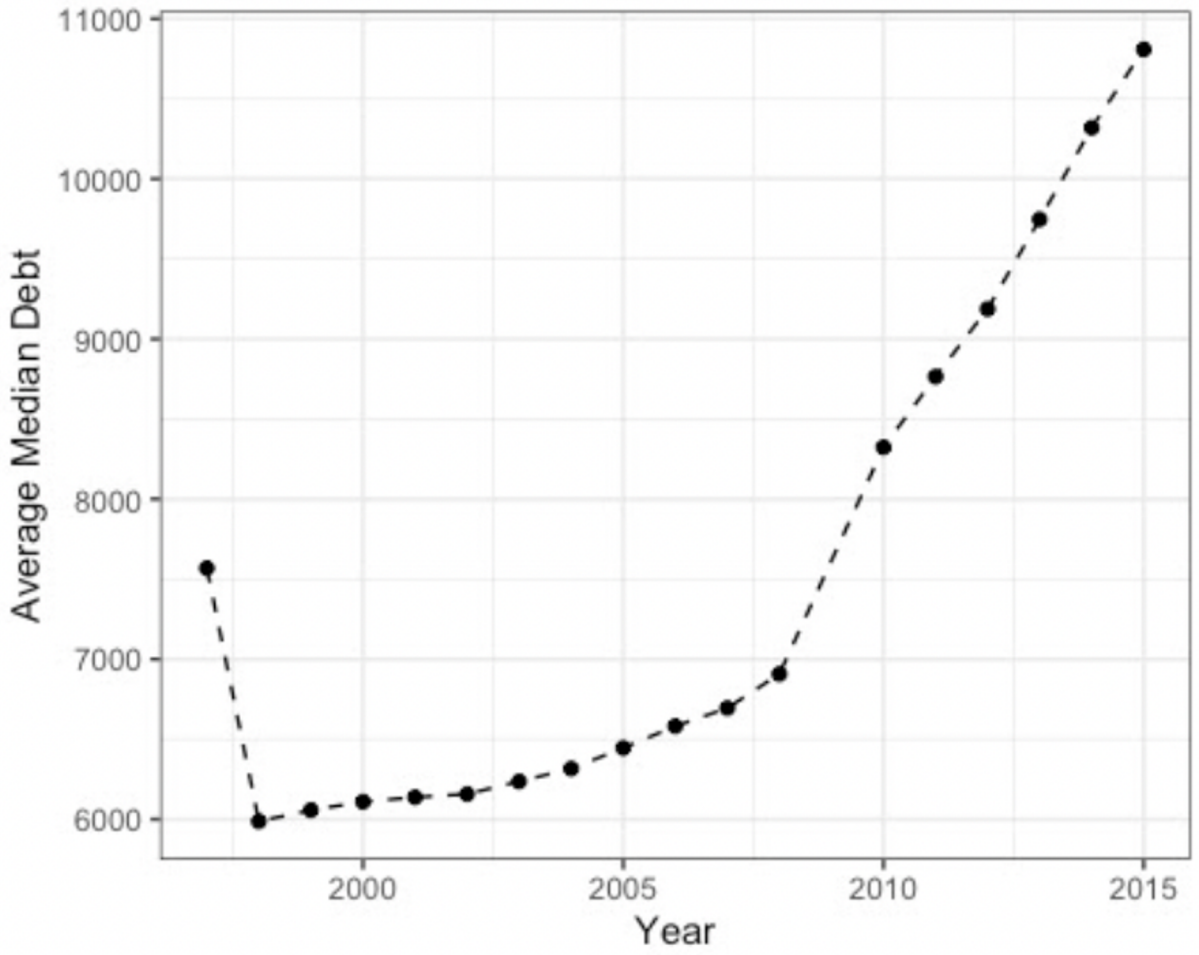
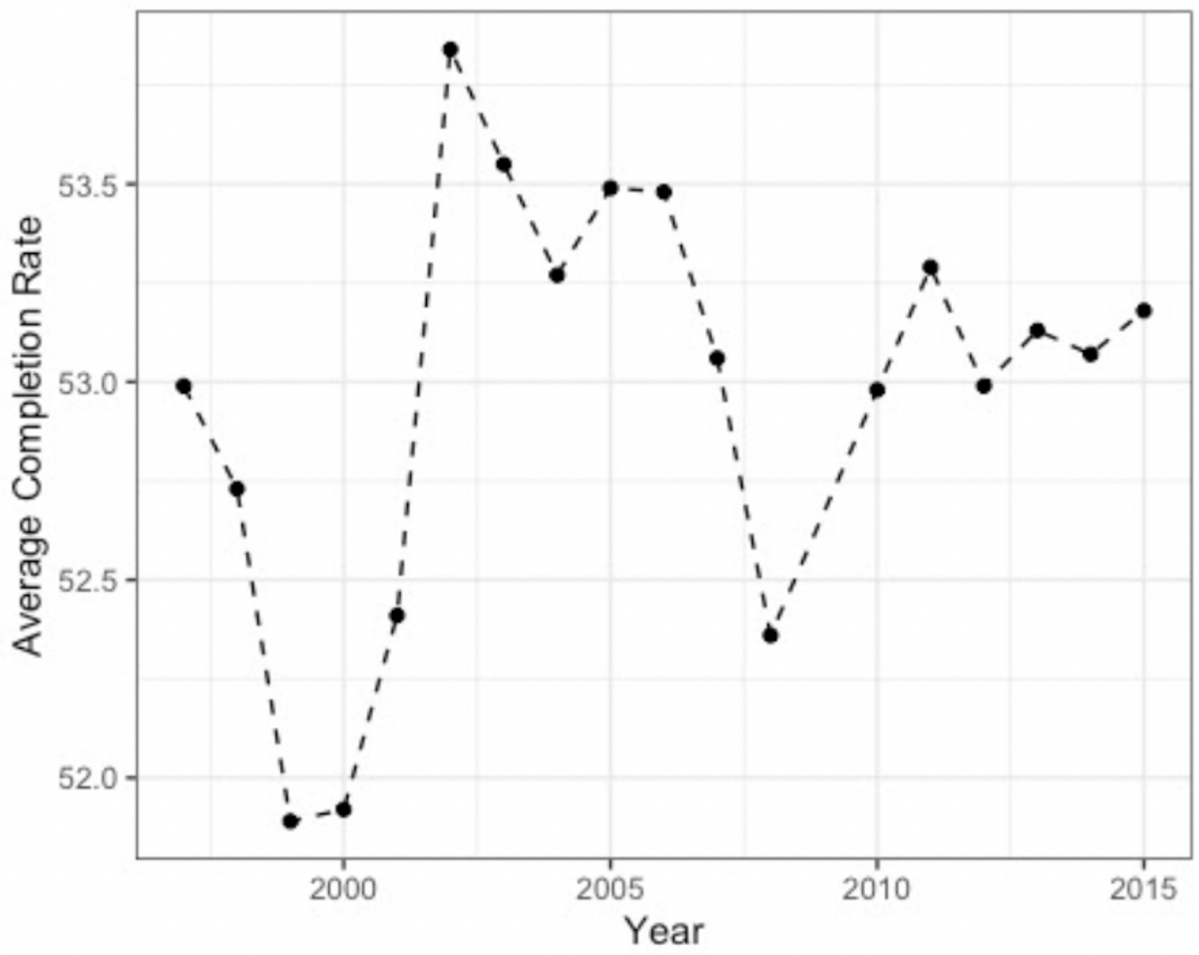


Figure 2

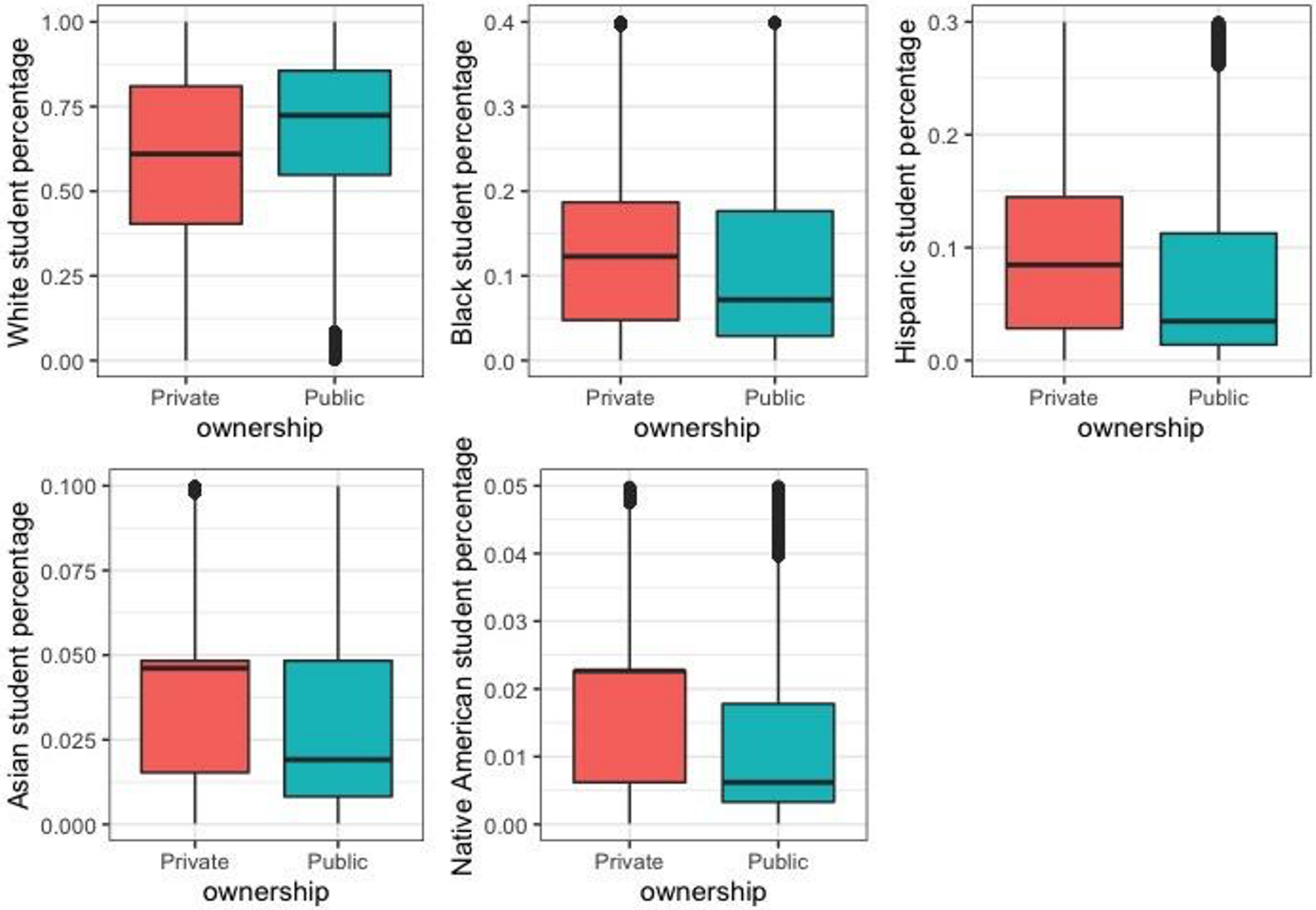
Average Completion Rate from 2000 to 2015



Ethnicity vs Ownership

Figure 3

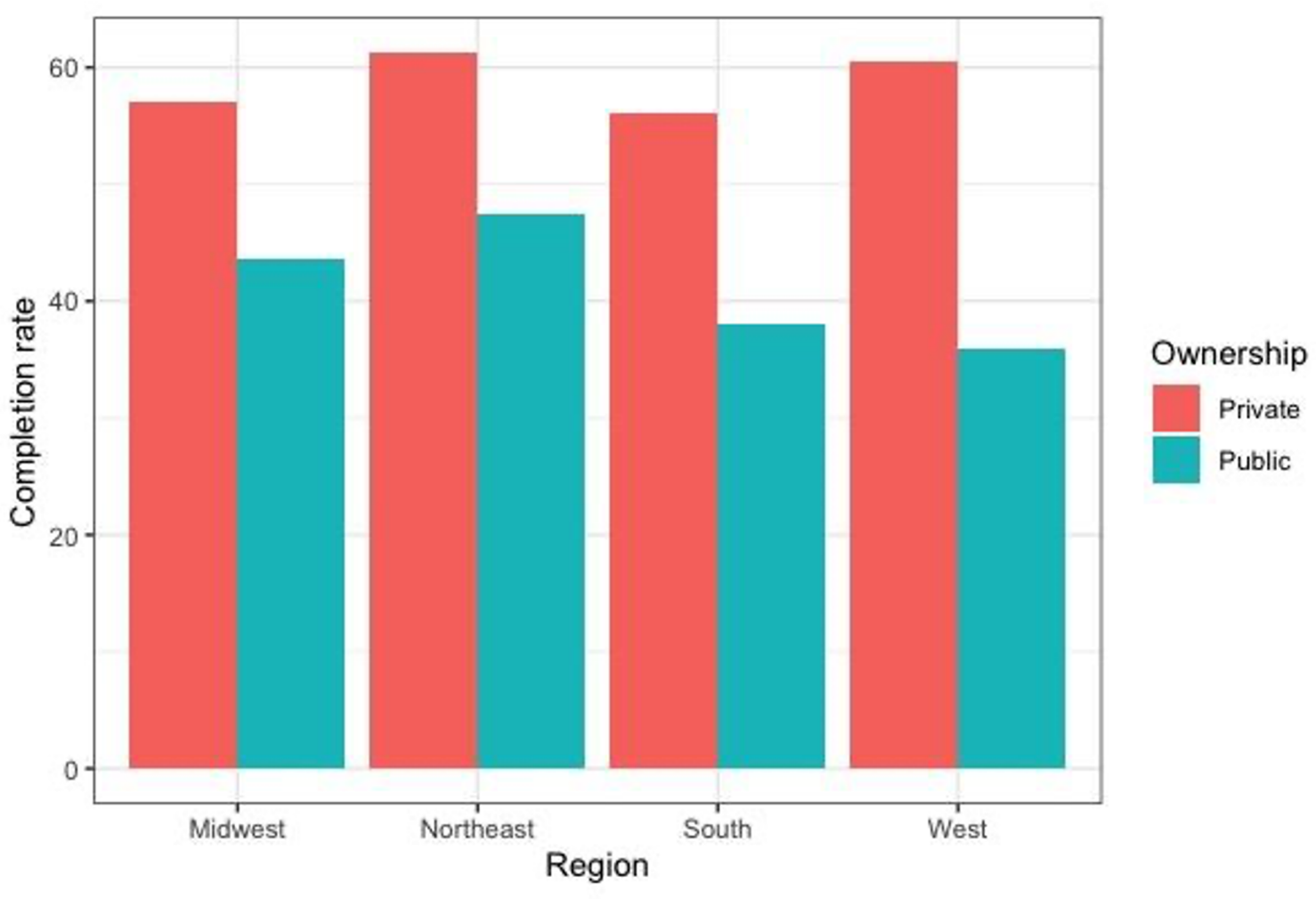
Boxplot of students ethnicity ratios in Private and Public Universities



Ownership vs Completion Rate

Figure 4

Completion rate by region and ownership



ANOVA Test:
How ownership and region influence the completion rate of US universities?

Hypothesis

1. **Regarding the interaction effect, the hypotheses are:**

H0: There is no interaction effect between the 2 factors: ownership and region

H1: There is an interaction effect between the 2 factors: ownership and region

1. **Regarding Ownership, the hypotheses are:**

H0: There is no difference in the average completion rate between private universities and public universities

H1: There is a difference in the average completion rate between private universities and public universities

1. **Regarding Region, the hypotheses are:**

H0: There is no difference in the average completion rate among the 4 regions

H1: There is a difference in the average completion rate among the 4 regions

Table 1

Two-Way ANOVA Test Result

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Ownership	1	817	816.6	20490	<0.01 ***
Region	3	60	20	501.9	<0.01 ***
Ownership:Region	3	43	14.4	361	<0.01 ***
Residuals	122435	4879	0		

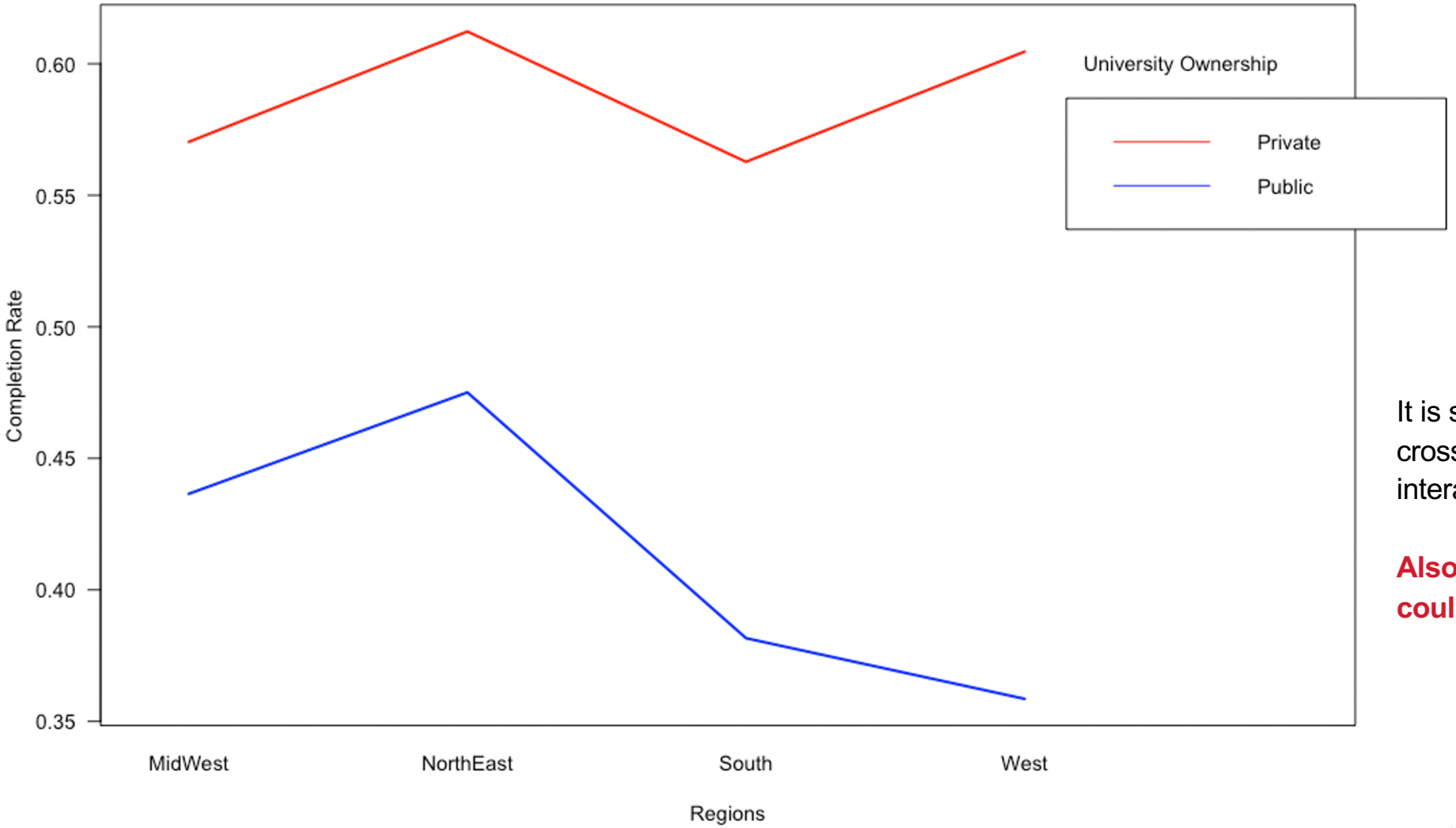
At the significance level of 0.05

- **The interaction effect is significant.**
- **Main effects are significant:**
  - **Ownership:** there is a difference in the average completion rate between private universities and public universities
  - **Regions:** there is a different in the average completion rate among 4 regions

ANOVA Test
Ordinal Interaction: 2 main effects could be explained independently

Figure 5

Graph of Two Variables Indicating an Ordinal Interaction



It is shown that the 2 lines do not cross each other which means the interaction is ordinal interaction.

Also, it means the 2 main effects could be explained independently.

Post-hoc Test: Tukey Test

FInd Groupwise Difference

The confidence interval does not include 0,

which means the pair are significantly different from each other

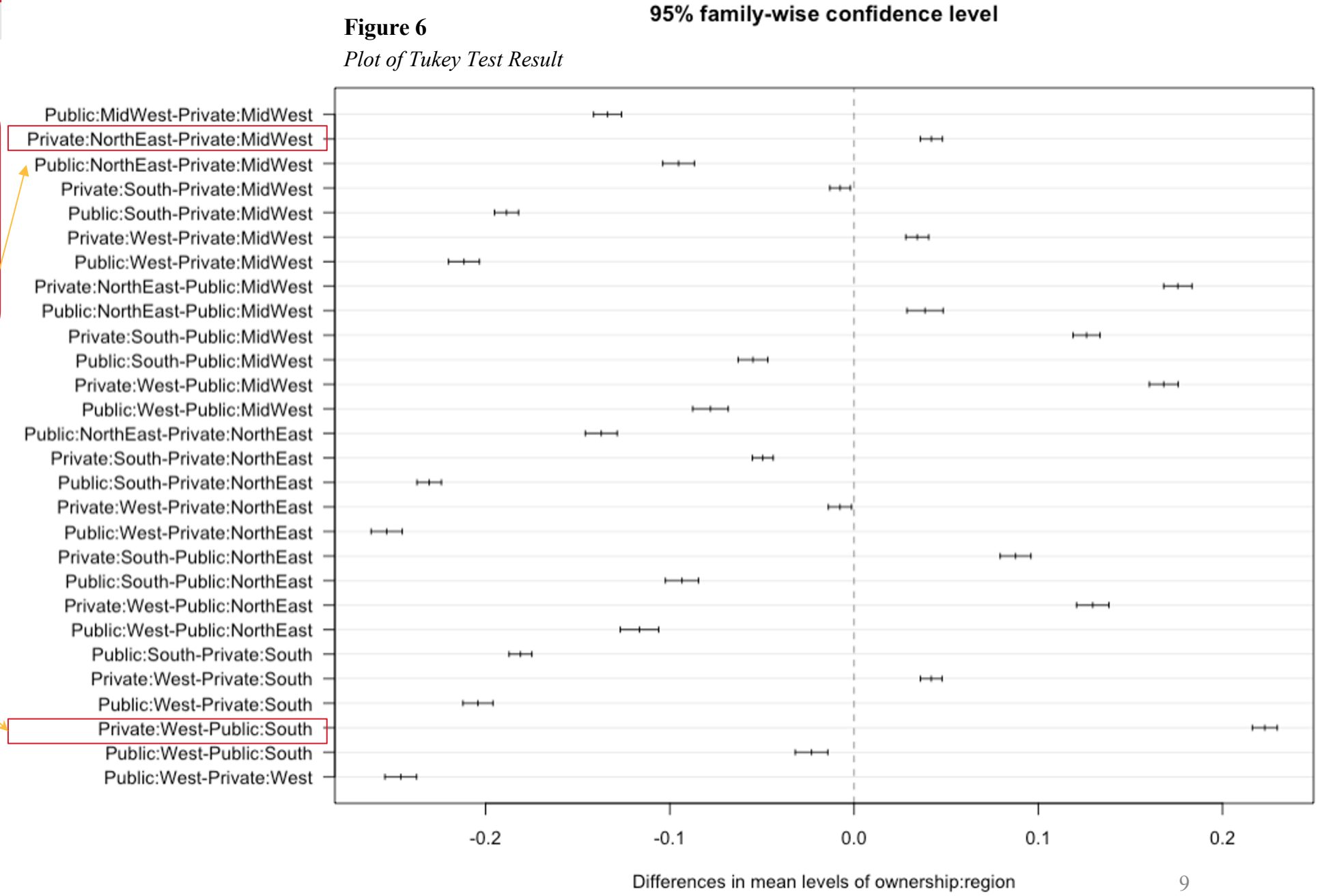
e.g.

There is a significant difference in completion rate between private universities in NorthEast and private universities in MidWest.

There a significant difference in completion rate between private universities in West and public universities in South.

Figure 6

Plot of Tukey Test Result



Linear Regression: Completion Rate Prediction

Dependent Variable:

Completion Rate

Main Predictors:

- Predominant degree
- Highest degree granted
- Ethnicity Ratio
- Commitment
- Ownership
- Region

Approach:

LASSO vs OLS with Stepwise Selection vs K-means + LASSO

Table 2

Coefficients of variables: 23 x 1 sparse Matrix of class "dgCMatrix"

	Test RMSE	Adjusted R-Squared	No. of Predictors	Multicollinearity
LASSO	0.1692	0.39	17	No
OLS + Stepwise	0.1684	0.39	21	Yes
K-means + LASSO	0.1707	0.39	10	No

LASSO model with 17 predictors is our preferred choice

- **Overfitting:** LASSO has relatively low Test RMSE
- **Accuracy:** the 3 models have similar Adjusted R-Squared
- **Simplicity:** LASSO and K-means + LASSO are much simplified than OLS + stepwise
- **Clarity:** LASSO is more interpretable than other 2 models

LASSO: Tuning the model

Figure 7

Plot: Cross Validation : 10-fold CV Mean Squared Error (MSE)

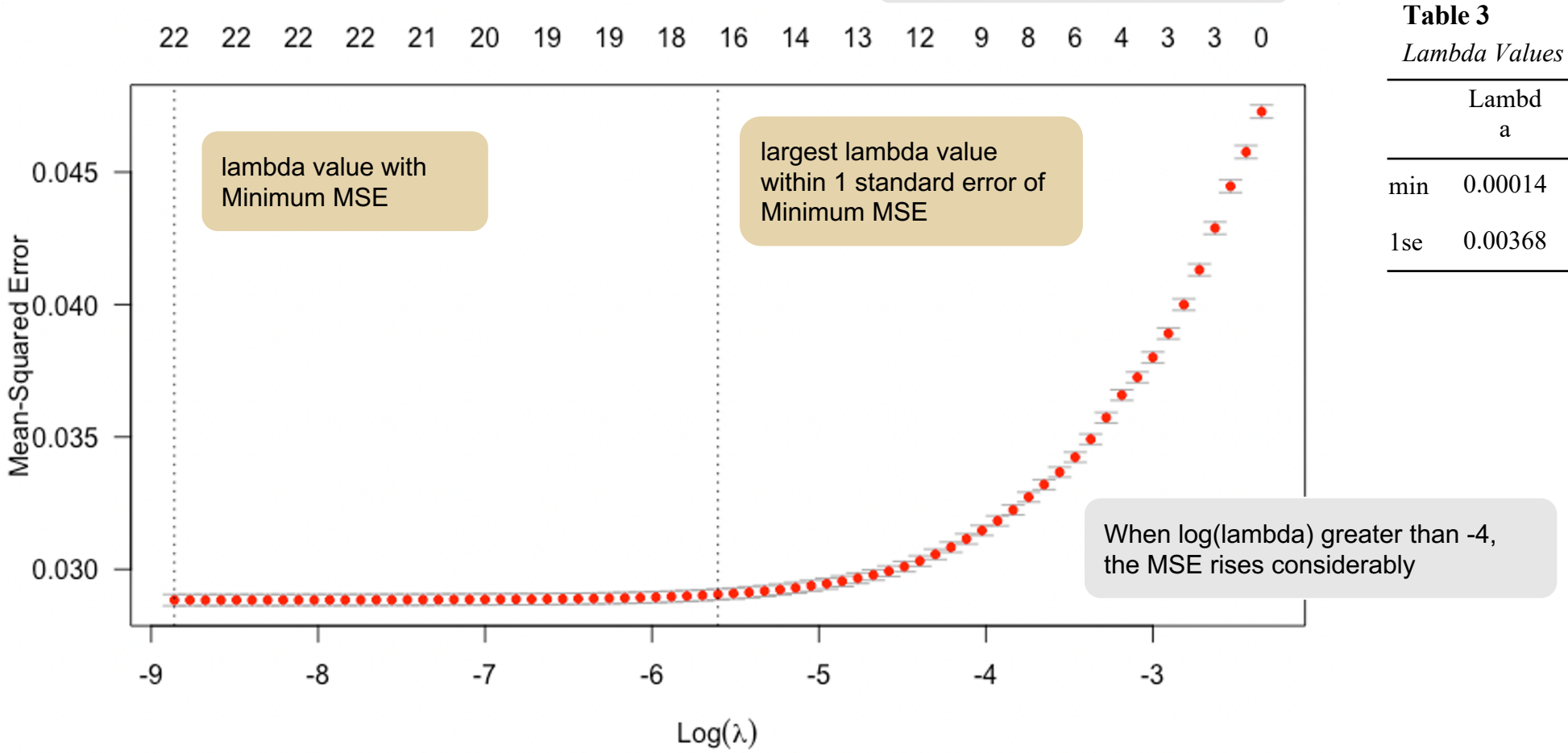


Table 3

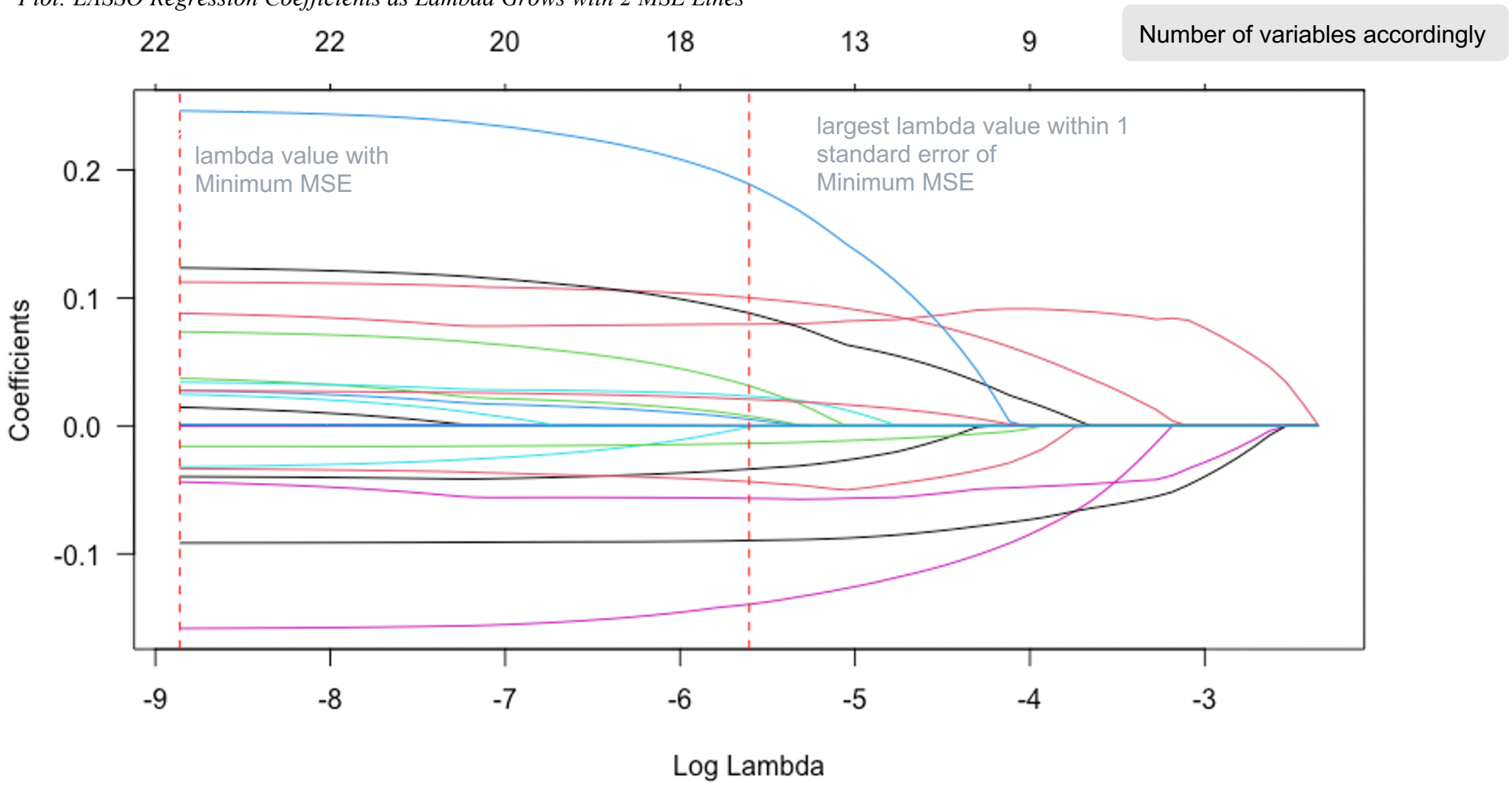
Lambda Values

	Lambda	Log Lambda	SE	Nonzero
min	0.00014	-8.86	0.00021	22
1se	0.00368	-5.6	0.00021	17

LASSO: Tuning the model

Figure 8

Plot: LASSO Regression Coefficients as Lambda Grows with 2 MSE Lines



LASSO Model Interpretation

Table 3

Coefficients of variables: 23 x 1 sparse Matrix of class "dgCMatrix"

	s0
(Intercept)	0.43074
predominant degree Bachelor's degree	.
predominant degree Certificate degree	0.10071
predominant degree Graduate degree	0.00731
predominant degree Non-degree-granting	0.00633
highest degree granted Not classified	0.02193
highest degree granted Predominantly associate's-degree granting	-0.05712
highest degree granted Predominantly bachelor's-degree granting	-0.03400
highest degree granted Predominantly certificate-degree granting	0.07859
median debt	0.00001
age begin	.
female pct	-0.00076
undergraduate enrollment	.
white undergraduate pct	0.08861
black undergraduate pct	-0.04323
hispanic undergraduate pct	0.03172
asian pacific islander	0.18926
native am al undergraduate pct	.
part-time pct	-0.13923
ownership Public	-0.08954
region NorthEast	0.02081
region South	-0.01375
region West	.

More Asian Pacific and white undergraduate students, higher completion rate

- If the % of Asian/Pacific Islander undergraduates students increases by 1, the average completion rate increases by 19%.
- If the % of white undergraduate students increase by 1, the average completion rate increases by 8.9%.

More part-time undergraduate students, lower completion rate

- If the % of part-time undergraduate students increase by 1, the average completion rate decreased by 14%.

Certificate degrees have higher completion rates

- On average, when the most common type of degree granted by the institution is certificate, the completion rate of this institution will be 10% higher than the institutions whose predominant degree are other types.

Private universities have higher completion rates than public universities

- On average, public universities have 8.9% lower completion rate than private universities.

Median debt burden of students following graduation has no influence on completion rate

- The coefficient of median debt in the model is very close to 0.

Logistic Regression

Table 4

Logistic Regression Model

	coefficient estimate	odds ratio	P-value
(Intercept)	-30.27	0.00	<0.001
Bachelor's degree	0.70	2.00	<0.001
Certificate degree	0.23	1.26	<0.001
Graduate degree	0.35	1.42	<0.001
Non-degree-granting	1.94	6.99	<0.001
Not classified	-0.22	0.80	**0.178**
Predominantly associate's-degree granting	2.09	8.12	<0.001
Predominantly bachelor's-degree granting	-0.27	0.77	<0.001
Predominantly certificate-degree granting	1.34	3.83	<0.001
North East	0.23	1.26	<0.001
South	0.64	1.89	<0.001
West	0.12	1.13	<0.001
median debt	0.00	1.00	<0.001
age begin	0.00	1.00	**0.741**
female percentage	-1.28	0.28	<0.001
year	0.01	1.01	<0.001
undergraduate enrollment	0.00	1.00	<0.001
white undergraduate percentage	1.96	7.10	<0.001
black undergraduate percentage	-0.41	0.67	<0.001
Hispanic undergraduate percentage	-1.74	0.18	<0.001
Asian pacific islander percentage	-1.43	0.24	<0.001
native American undergraduate percentage	3.99	53.80	<0.001
part-time percentage	1.62	5.06	<0.001
completion rate	-2.56	0.08	<0.001

There is no big difference between the public and private regarding the students' age to begin the program.

Odds ratio of age begin equals to 1.

More Native American undergraduate students, more probability it is a public university.

The odds ratio of Native American undergrads % is much greater than 1.

Public universities have fewer hispanic undergraduate students.

The odds ratio of Hispanic undergraduate percentage is smaller than 1, which means it has a negative association with the Group 1 (public uni).

If the predominant degree granted by the uni is Associate's degree, more probability it is a public university.

The odds ratio of predominantly associate's – degree granting is much greater than 1, which means it has a positive association with public universities.

Logistic Regression: Model Performance

Figure 9

Confusion Matrix

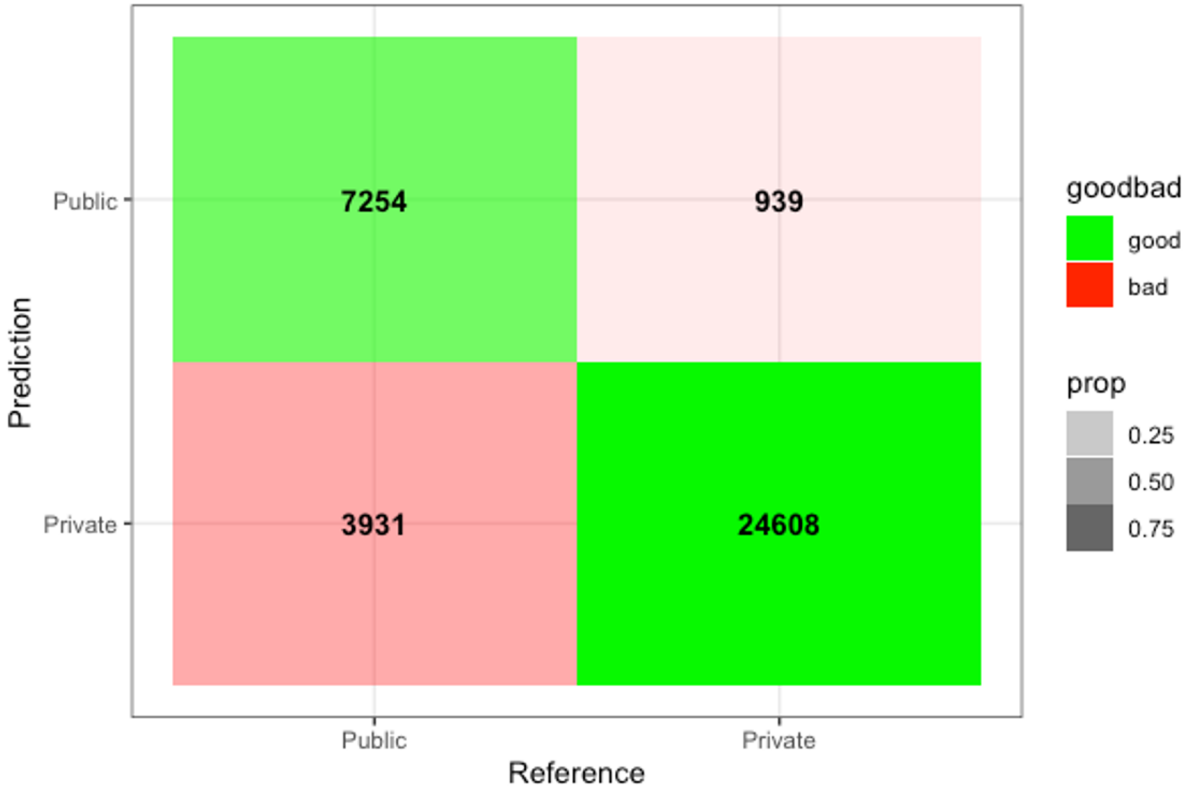
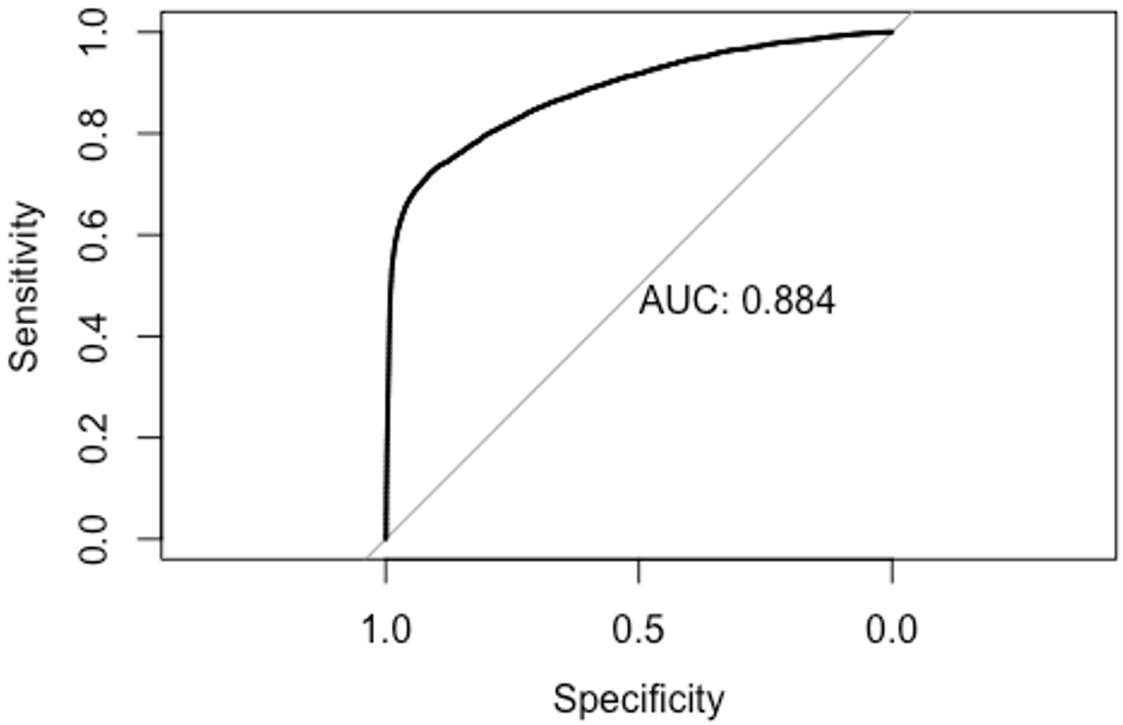


Figure 10

Plot of ROC curve



Accuracy: 86.74%, Sensitivity: 96.32%, Specificity: 64.85%, AUC: 0.884, **THIS MODEL not good at predicting public school**

Conclusion

To conclude the results

- **Student debt burden** has no influence on the **completion rate**;
- The **completion rates** of universities **in different regions** are significant different;
- Public universities have **lower completion rate** than private universities, on average;
- **Institutions' ownership** could be well predicted by students ethnicity structure, degree level, region and commitment (88%);
- Private universities are more **diversified** than public universities.
- **University completion rate** could not be well predicted by using degree level, ethnicity ratio, commitment, ownership and regions;
- By **increasing the diversity** of the students, it would **improve the completion rate.**

Recommendations to University Admission Teams

- Public universities could **recruit more students from Asian Pacific**, which would be beneficial for increasing their completion rate.
- University Admissions **no need to concern about the students' debt** which actually have **no influence on the completion rate**. Therefore, providing more financial supports would help universities to improve their reputation.
- Improve the **commitment of part-time** students would also help universities to increase their **completion rate**.

Thank you!