Data Analysis of U.S. News 2016 Public Universities Rankings

Xiao Wei, PUBPOL 2018 Winter Computational Thinking for Government Analytics , Final Project

Introduction and the Data

The original dataset is a full version of rankings that include both public and private universities rankings. Since the data of interest is the rankings of U.S. public universities, I extracted those schools who have in-state tuition and limited the rankings to 1st to 200th. As the tuition of public colleges keeps raising in general, the initial motivation of the data choice is to explore the average tution level of public universities and how does it affect their enrollment.

There are eight columns of the data, which are the name of the university, location, rank, description, tuition and fees (out-of-state), in-state tuition, undergraduate enrollment, and the tuition difference between out of and in-state tuition.

Findings:

Data Exploration

count	140.000000
mean	27.404257
std	6.734391
min	11.403000
25%	22.235500
50%	26.462500
75%	31.421000
max	45.066000

The total amount of the universities in the dataset is 140. Among all of them,

the average tuition is $27,404 per year and the maximum tuition is $45,066.

count	133.000000
mean	16.497308
std	5.316482
min	3.231000
25%	13.386000
50%	15.958000
75%	18.930000
max	29.620000

There are 133 schools who have different in-state and out-of-state tuition. Among all of them, the highest difference is $29,620.

To know which school has the highest out-of-tuition, I created a loop to find and show it.

name	University of Michigan--Ann Arbor
location	MI
rank	27
description	The University of Michigan boasts one of the b...
tuition_and_fees	43.476
in_state	13.856
undergrad_enrollment	28312

Here it shows the University of Michigan—Ann Arbor is the one who has the highest out-of-state tuition within the dataset, which is $43,476 per year.

Regression:

As my initial motivation stated before, I am curious that how does the tuition level affect students'choice of universities. Therefore, I need to run a regression model to find out.

First, I need to confirm there is no significant collinearity between two variables, in-state and out-of-state tuition, that I am going to use in the regression. Then I used the Pearson test and the p-value shows there is moderate to high correlation (Pearson value=0.6715) between two variables.

Then I run the regression with undergraduate enrollment as the dependent

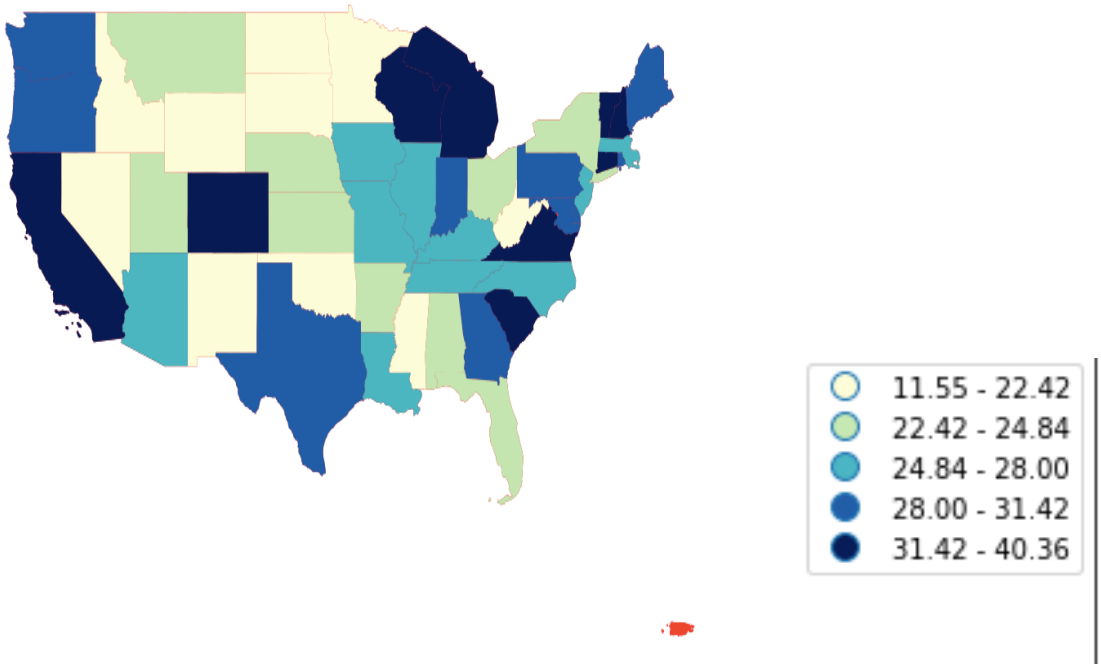
variable, and in-state and out-of-state tuition as two explanatory variables.

	coef	std err	t	P>|t|
Intercept	1.698e+04	3505.918	4.843	0.000
tuition_and_fees	509.8780	160.716	3.173	0.002
in_state	-923.4927	362.680	-2.546	0.012

As the coefficients shown above, the out-of-state tuition has positive correlation with enrollment, and on the contraray, in-state tuition has negative correlation.

Choropleth Map

In the part, I incorprated the geographic information of the dataset with a U.S state map. As the color in the map is darker, the higher the tuition. Therefore, state California, Michigan, Colorado, Virginia and other three states have the highest level of public university tuition in the range of 31.42-40.36 thousand dollars.


11.55 - 22.42
22.42 - 24.84
24.84 - 28.00
28.00 - 31.42
31.42 - 40.36