

College tuition, diversity, and pay

Zahra Khoshmanesh, Vahid Azizi, and Saba Moeinizade

May 07, 2020

Contents

1	Introduction	1
2	Analysis	2
2.1	Diversity dataset	2
2.2	Tuition cost dataset	7
2.3	The historical tuition dataset	17
2.4	Salary potential dataset	20
3	Conclusions	24

1 Introduction

Many people are interested to know about the tuition, costs, diversity and potential salary when searching for college. In this project, we want to analyze tuition costs across different states, and explore diversity in different schools. We have 4 datasets with 23 variables and more than 50000 observations. The variables include different information such as school name, state, type of school, in-state/out-of-state tuition, group/racial/gender category, early/mid-career pay, stem percent and historical tuition information. We are also interested in trends of tuition over time. We will use different summary statistics and visualizations in R to address these problems.

2 Analysis

We start our analysis with exploring the diversity dataset.

2.1 Diversity dataset

In diversity dataset, each row describes a specific university or college in US in terms of enrollment per each diversity group which can be either ethnicity, women, or minority groups.

```
diversity_school %>% glimpse()

## #> #> #> #> #>
```

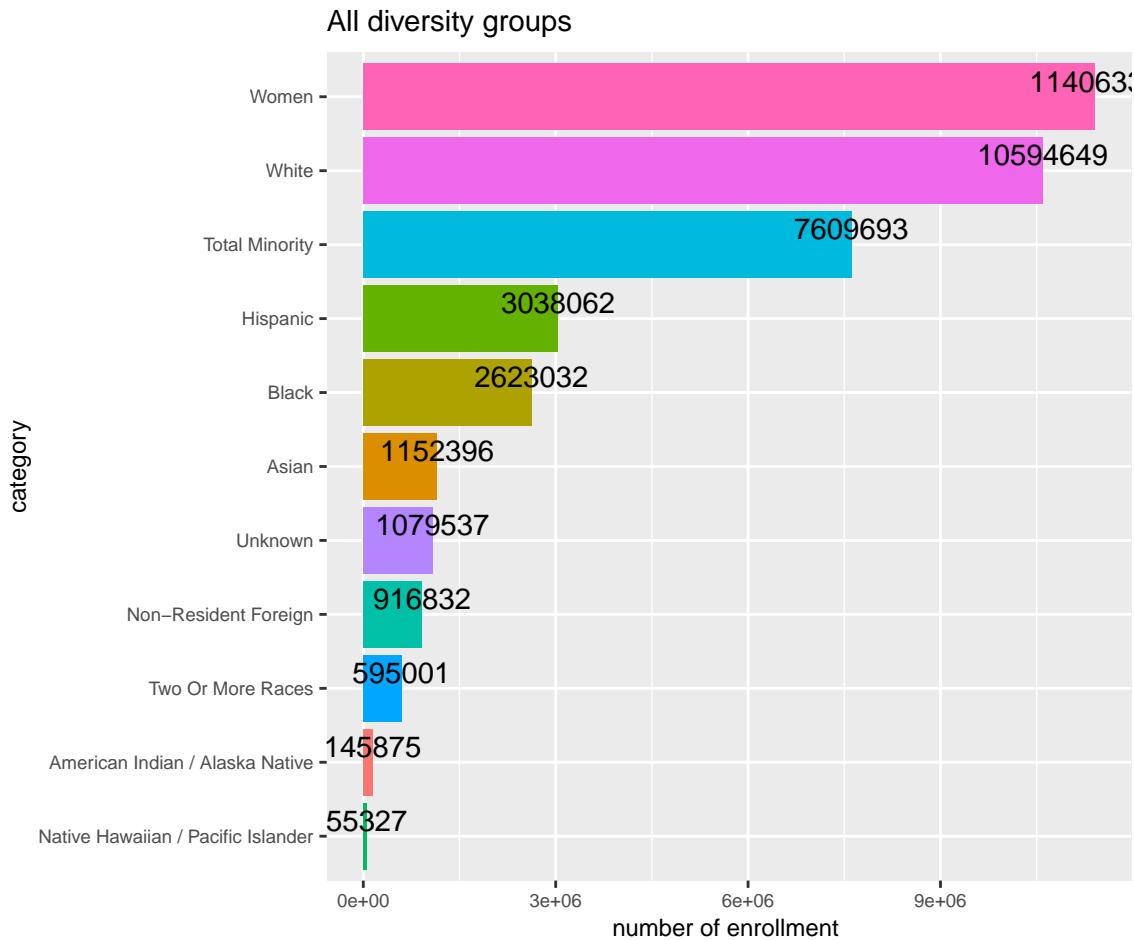
```
## #> Rows: 50,655
## #> Columns: 5
## #>
## #> $ name <chr> "University of Phoenix-Arizona", "University of Ph...
## #> $ total_enrollment <dbl> 195059, 195059, 195059, 195059, 195059, 19...
## #> $ state <chr> "Arizona", "Arizona", "Arizona", "Arizona", "Arizo...
## #> $ category <chr> "Women", "American Indian / Alaska Native", "Asian...
## #> $ enrollment <dbl> 134722, 876, 1959, 31455, 13984, 1019, 58209, 1903...
```

Let's check the different diversity groups provided in this dataset.

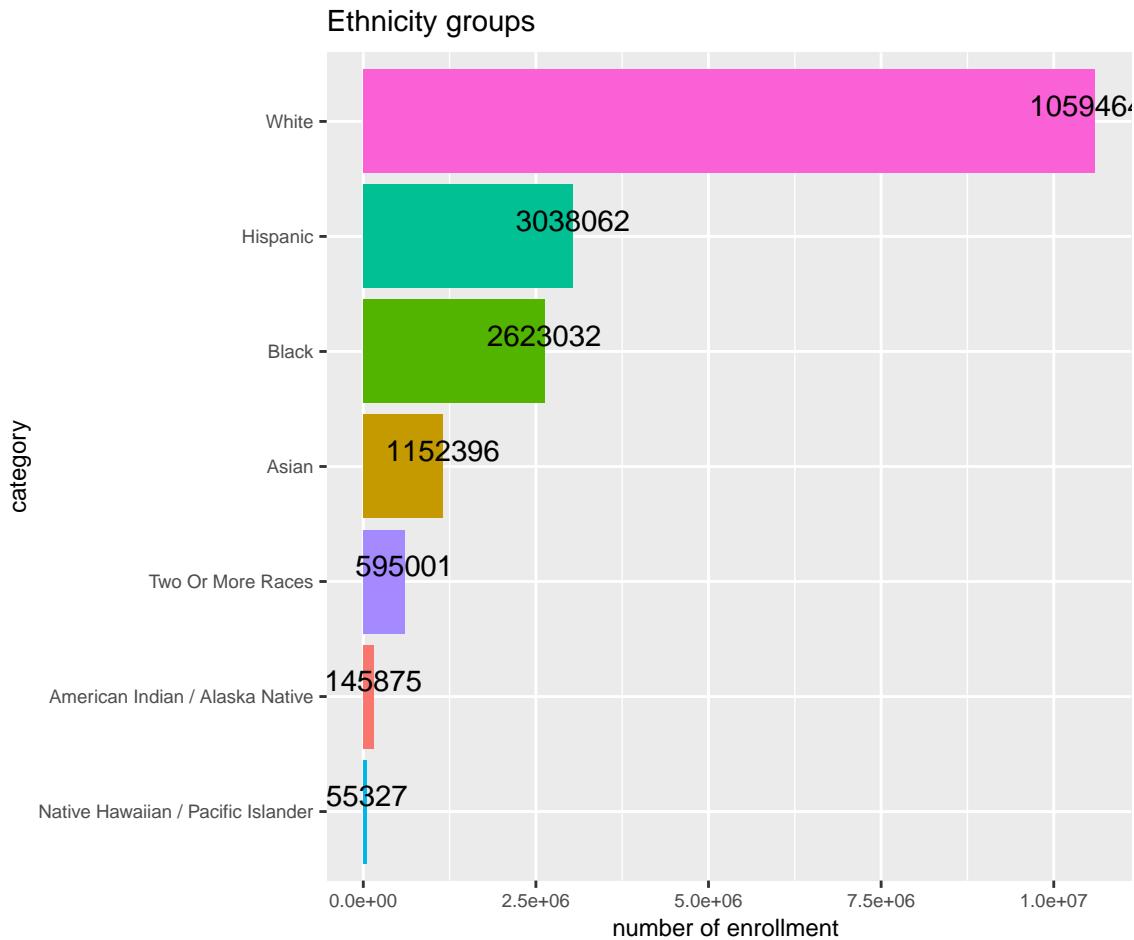
```
## #> [1] "Women" "American Indian / Alaska Native"
## #> [3] "Asian" "Black"
## #> [5] "Hispanic" "Native Hawaiian / Pacific Islander"
## #> [7] "White" "Two Or More Races"
## #> [9] "Unknown" "Non-Resident Foreign"
## #> [11] "Total Minority"
```

2.1.1 Enrollement rate across different diversity groups

We are interested to know the total enrollment across different diversity groups in our dataset. The following figure demonstrates total enrollement across different available diversity groups. Women have the first rank of enrollment among all diversity groups.



The following figure focuses on the ethnicity groups only. As we can see the white ethnicity group has the highest enrollment rate and Native Hawaiian has the lowest rate.



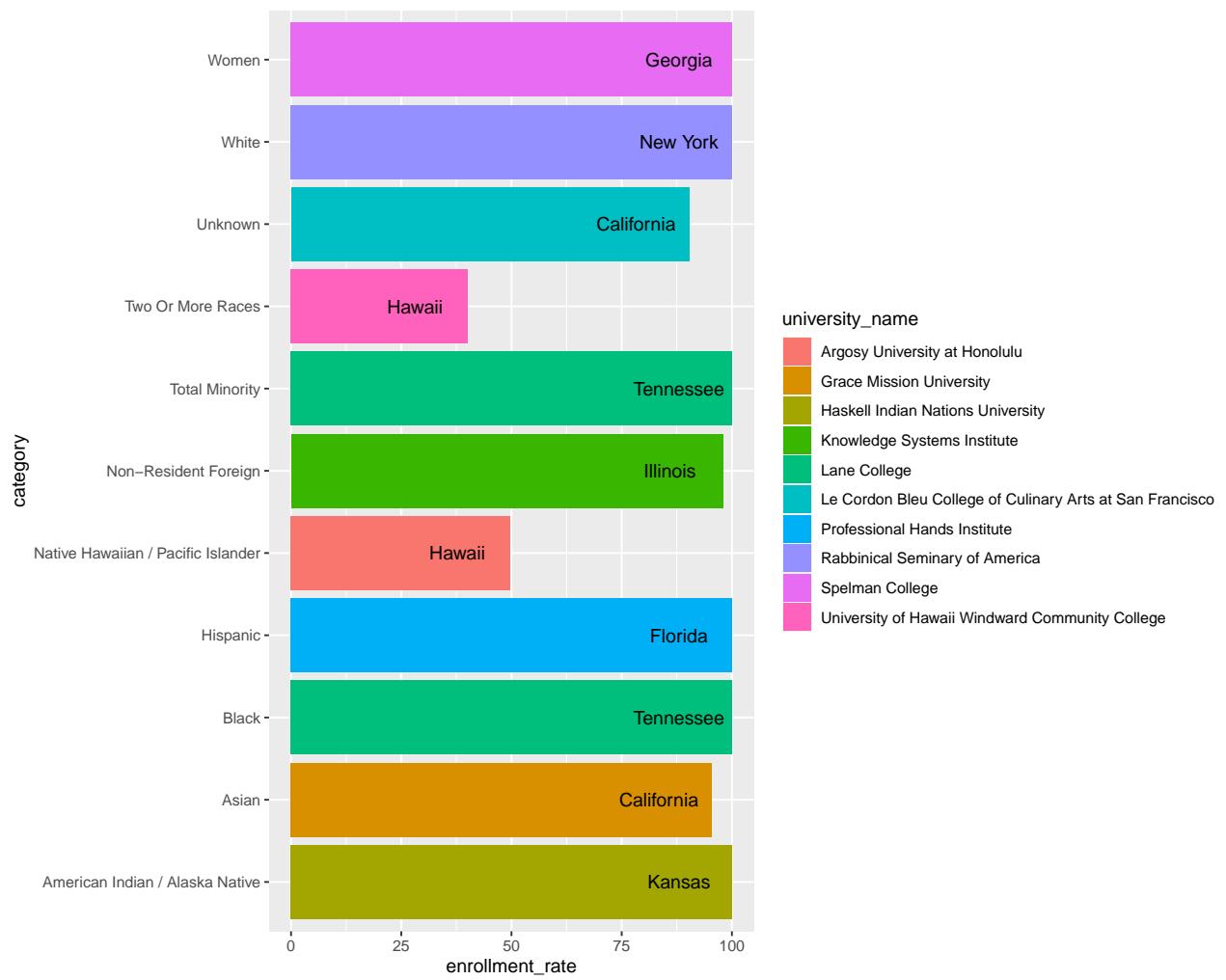
2.1.2 Finding the most diverse campus for each diversity category

In the below dataframe, we extract the name of university that has the highest enrollment per each diversity group. For example, Haskell Indian University in Kansas has 100 enrollment rate all belong to Native American.

```
## # A tibble: 11 x 6
##   category  university_name  state enrollment_rate total_enroll category_enroll
##   <chr>      <chr>        <chr>        <dbl>        <dbl>        <dbl>
## 1 American~ Haskell Indian ~ Kans~        100         808         808
## 2 Asian       Grace Mission U~ Cali~       95.5         88          84
## 3 Black       Lane College    Tenn~        100        1262        1262
## 4 Hispanic    Professional Ha~ Flor~       100         37           37
## 5 Native H~  Argosy Universi~ Hawa~       49.7        933        464
```

## 6 Non-Resident Foreign	Knowledge Systems Institute	Illi-	97.9	193	189
## 7 Total Minority	Lane College	Tenn-	100	1262	1262
## 8 Two Or More Races	University of Hawaii Windward Community College	Hawa-	40.1	2661	1068
## 9 Unknown	Le Cordon Bleu College of Culinary Arts at San Francisco	Cali-	90.3	442	399
## 10 White	Rabbinical Seminary of America	New -	100	500	500
## 11 Women	Spelman College	Geor-	100	2135	2135

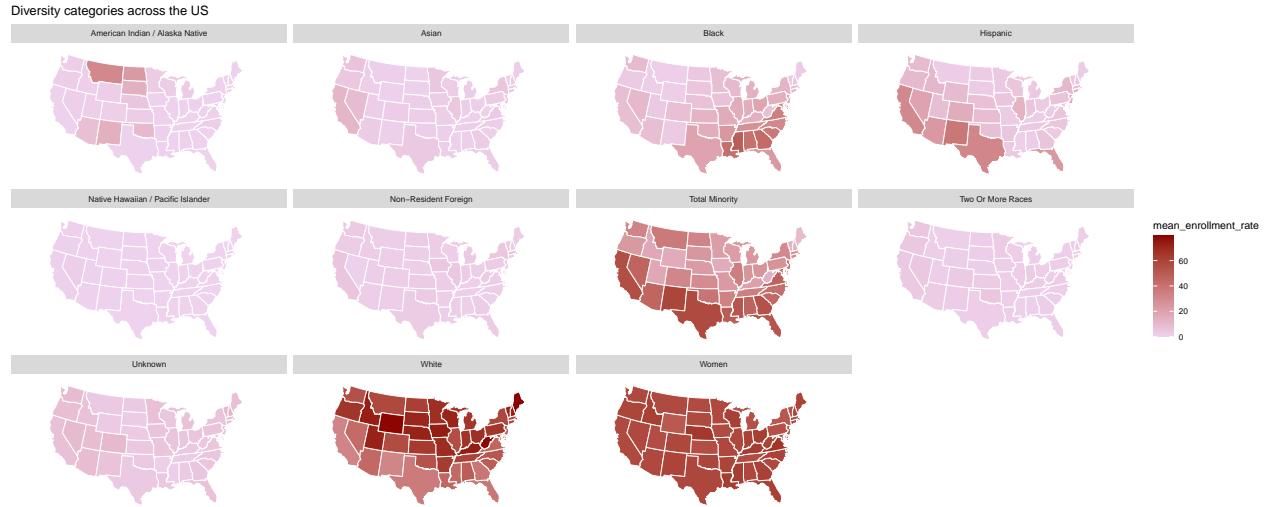
We can also visualize this information using a bar chart as follows.



2.1.3 Diversity map of US for each diversity group

Next, we are interested to know about diversity across different states in US. The following map is provided per each diversity group and the darker colors show places that have greater enrollment

rate belong to the corresponding category. As we can see, the Black diversity group are mostly located in south east of US. In the same logic, midwest shows the higher enrollment of white people and California has the highest enrollment rate of Asian.



2.1.4 Women enrollment rate across US

For women, the enrollment rate of all state seems equally distributed. Virginia (with rate of 65) has the highest enrollment rate and New Jersey has the lowest rank compared to the other states (with rate of 50).



2.2 Tuition cost dataset

Our next dataset is the tuition cost. In this dataset, we have information regarding in-state , and out-of-state tuition for each university.

```
tuition_cost %>% glimpse()

## # Rows: 2,973
## # Columns: 10
## # $ name      <chr> "Aaniiih Nakoda College", "Abilene Christian U...
## # $ state     <chr> "Montana", "Texas", "Georgia", "Minnesota", "C...
## # $ state_code <chr> "MT", "TX", "GA", "MN", "CA", "CO", "NY", "NY"...
## # $ type      <chr> "Public", "Private", "Public", "For Profit", "...
```

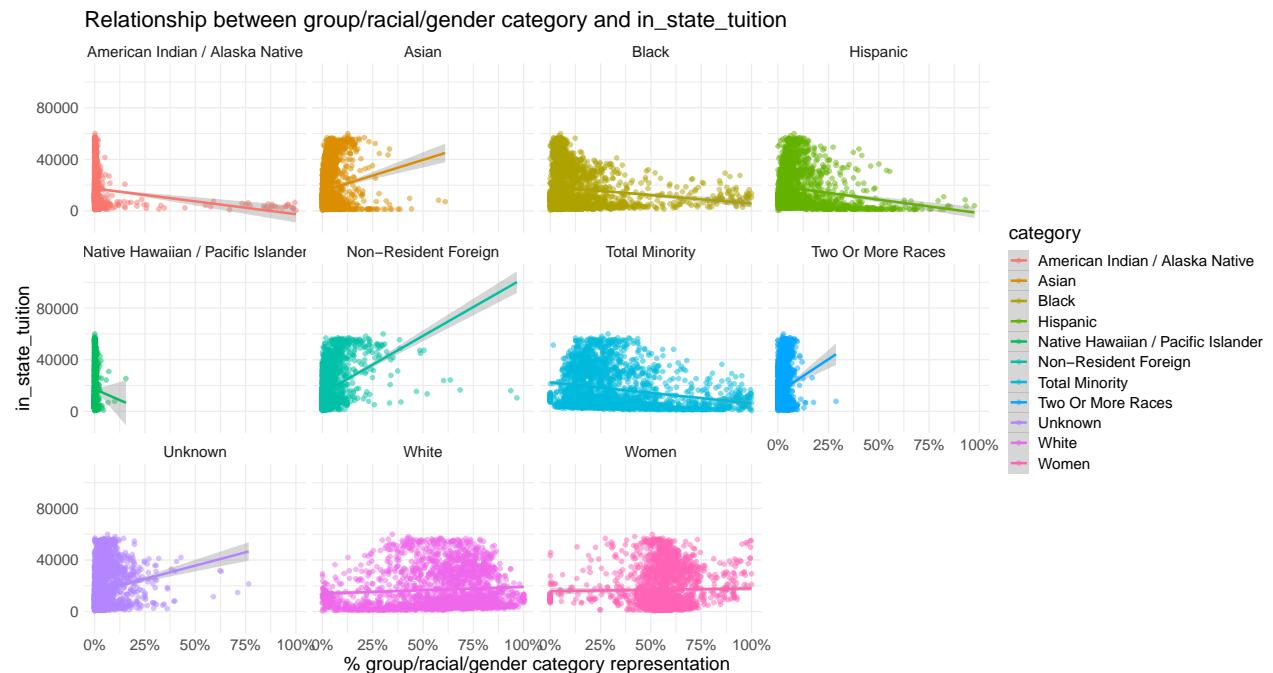
```

## $ degree_length      <chr> "2 Year", "4 Year", "2 Year", "2 Year", "4 Yea...
## $ room_and_board     <dbl> NA, 10350, 8474, NA, 16648, 8782, 16030, 11660...
## $ in_state_tuition   <dbl> 2380, 34850, 4128, 17661, 27810, 9440, 38660, ...
## $ in_state_total      <dbl> 2380, 45200, 12602, 17661, 44458, 18222, 54690...
## $ out_of_state_tuition <dbl> 2380, 34850, 12550, 17661, 27810, 20456, 38660...
## $ out_of_state_total   <dbl> 2380, 45200, 21024, 17661, 44458, 29238, 54690...

```

2.2.1 Relationship between enrollment rate and in/out-of state tuition per each diversity group

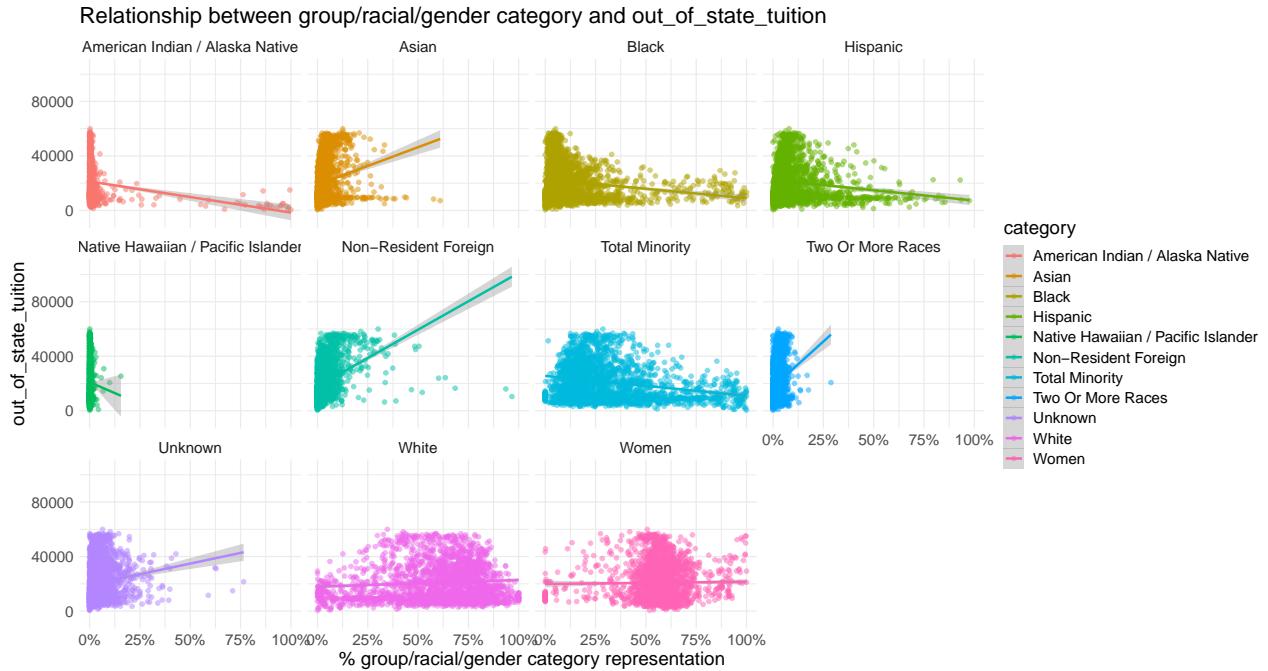
We are interested to know whether there is any correlations between enrollment rate and tuition per each diversity group. So, we merge tuition cost dataset and diversity dataset based on college name.



The above diagram shows the relationship between enrollment rate of each diversity group and in-state tuition. We can see negative correlations in American Indian, Native Hawaiian, Black, Hispanic, and Total Minority groups, which means lower in-state tuition increases the enrollment rate. For white category and women category, there is no strong relationship. In Asian and Non-Resident Foreign categories, we see positive correlations.

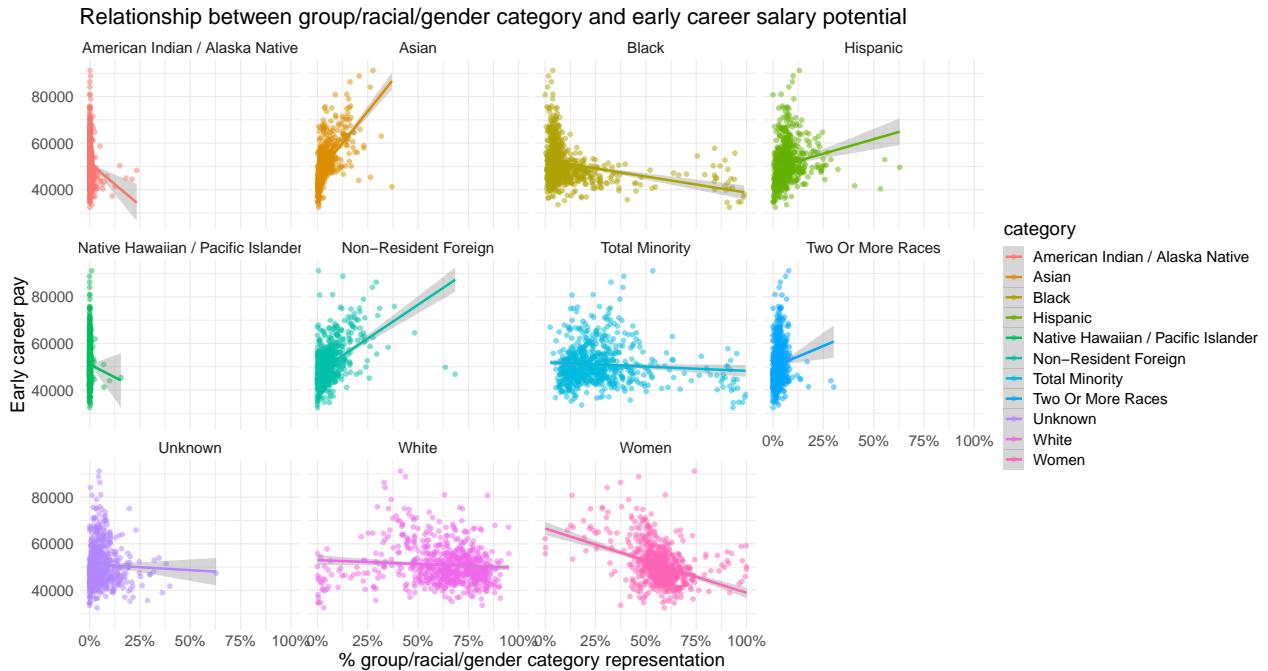
The following figure demonstrates relationship between enrollment rate and out-of-state tuition.

Almost the same trend like in-state tuition exists for out-of state tuition.



2.2.2 Relationship between diversity groups enrollment rate and early career salary

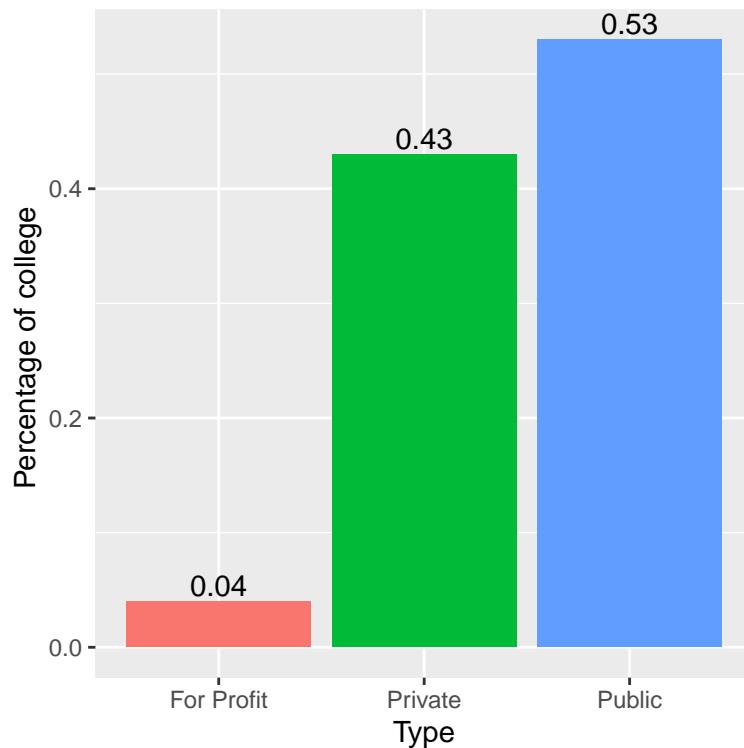
To answer if there is any relationship between salary and diversity, we need to merge diversity and salary datasets.



For Women, American Indian, Black and Native Hawaiian, the enrollment rate has inverse direction with early career salary. For Asian, Hispanic, and Non Resident categories the enrollment rate has direct relationship with early career salary.

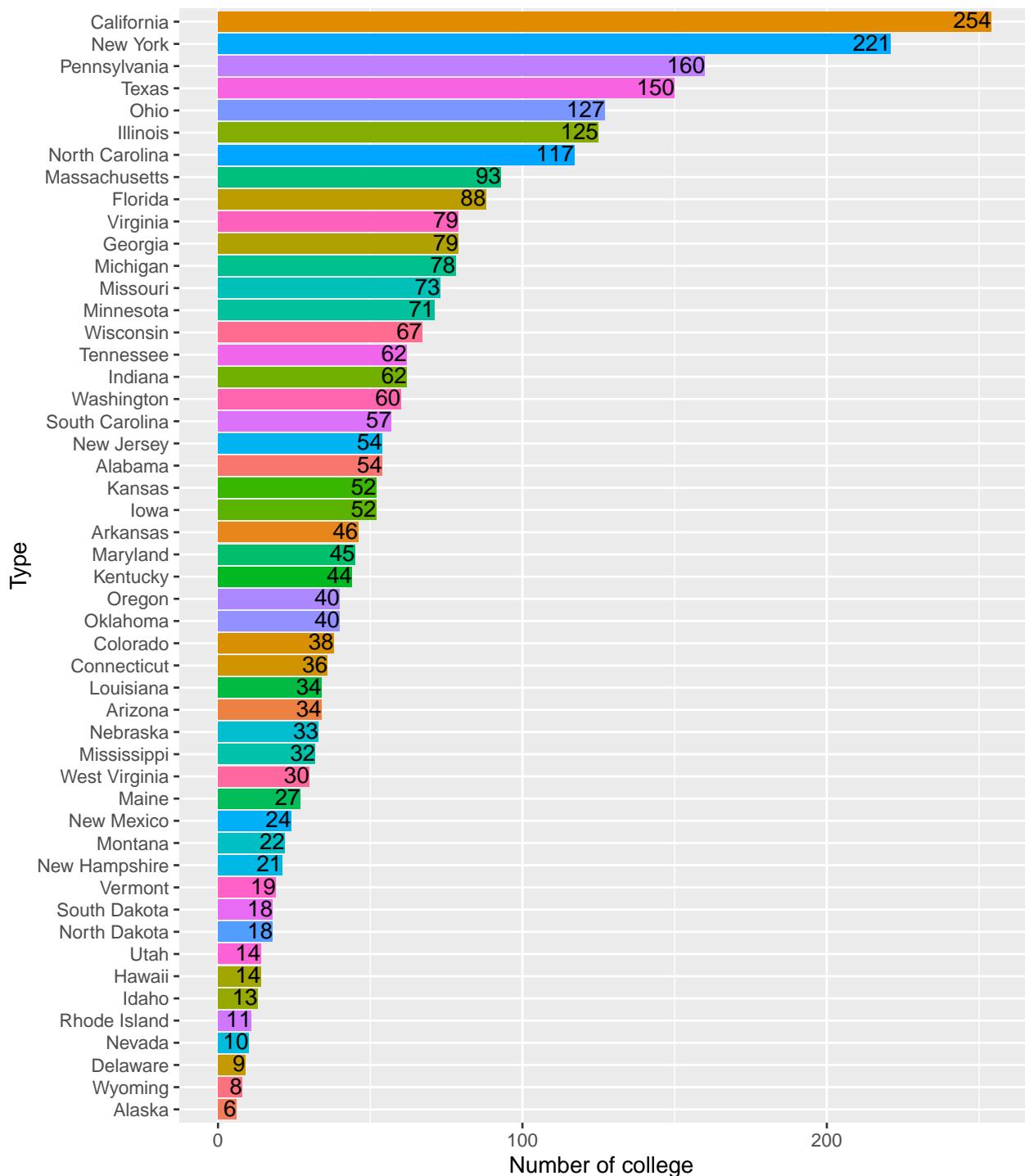
2.2.3 Different types of colleges in US

There are different types of colleges enrolling students across the country. According to this data set, three types of colleges are recognizable: public, private, and for profit. As shown in the following plot, public colleges are the majority category with 53 percent of all colleges and private and for-profit colleges are following public colleges with 43 and 4 percent, respectively.



2.2.4 How many colleges are there in each state?

In this part, we explore the number of colleges in different states. According to the following plot, the top 5 states are California, New York, Pennsylvania, Texas, and Ohio. Rhode Island, Nevada, Delaware, Wyoming, and Alaska are the states with the least number of colleges. Iowa is ranked 19th in the country with 52 colleges.

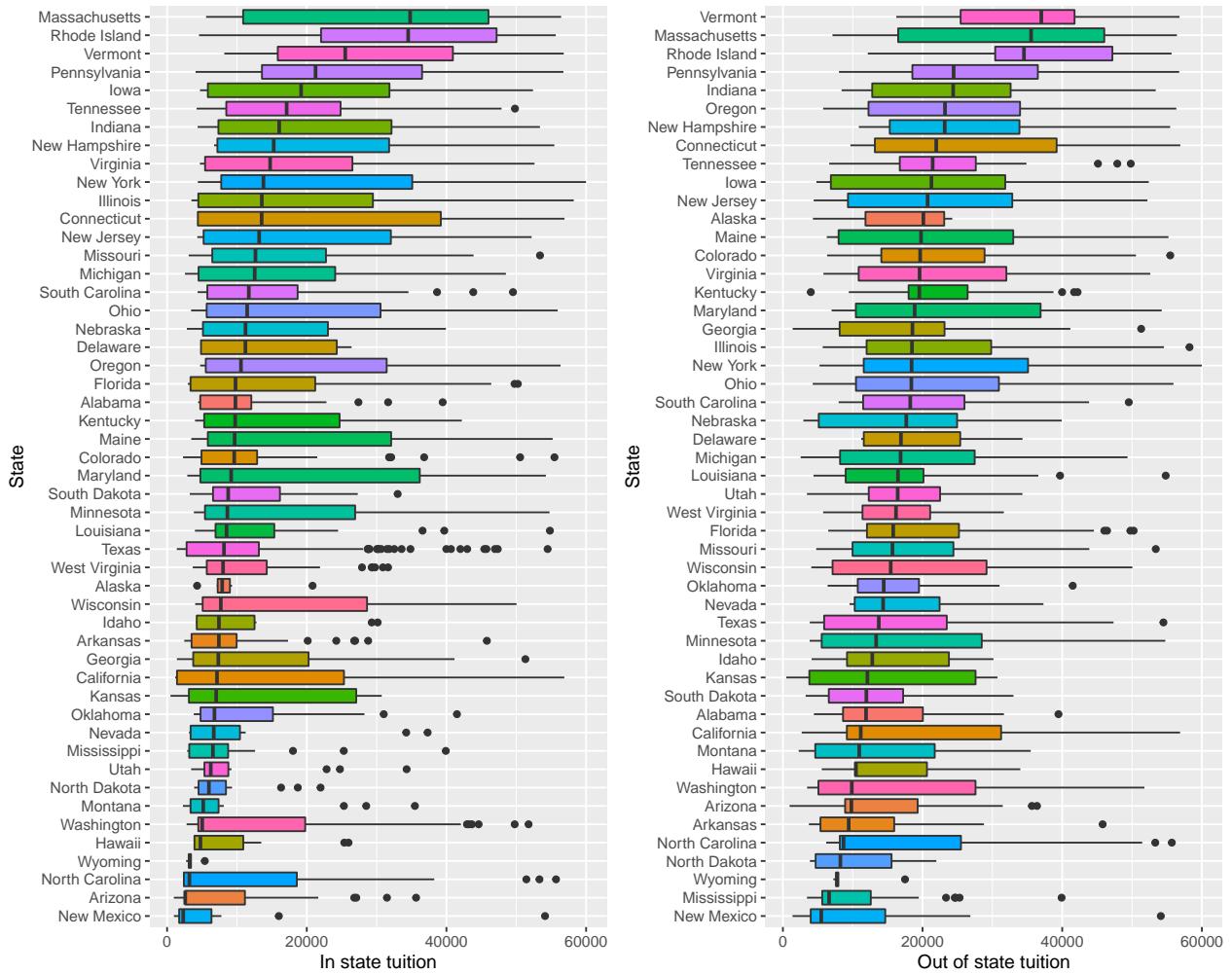


2.2.5 Comparing different states in terms of in-state and out-of-state tuition

Which states have a higher rate of tuition costs? This is a question that most students are interested to find an answer to it when they are going to choose a college to continue their studies. This section provides a comparison for in/out-of-state tuition in different states. As can be seen in the

left boxplot Massachusetts is the most expensive state on average in terms of in-state tuition. On the other hand, North Carolina, Arizona, and New Mexico are the cheapest states on average for those who intend to pay in-state tuition.

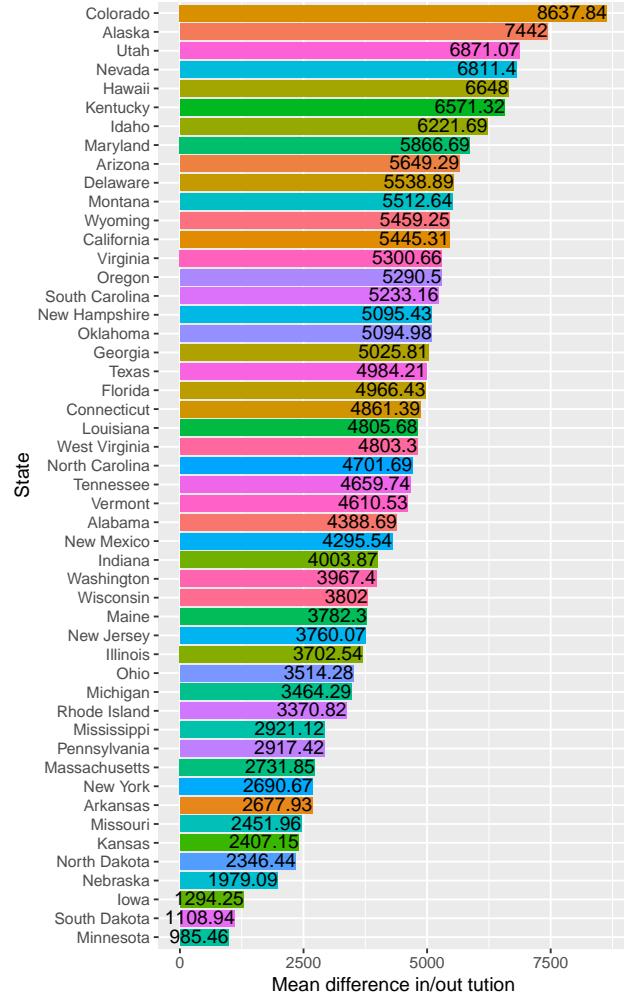
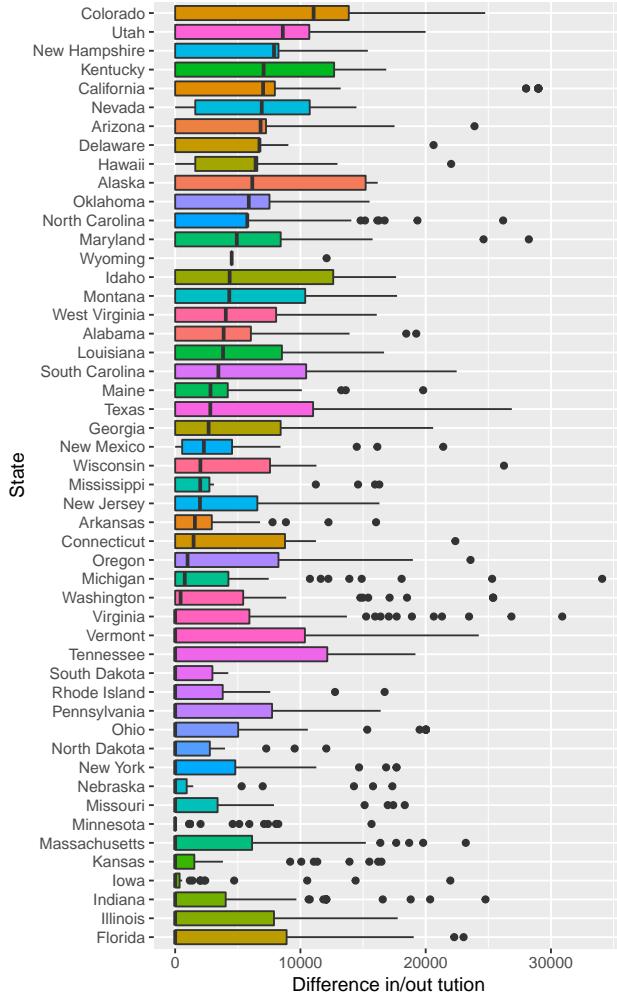
Same comparison is done for out-of-state tuition in right boxplot. This plot shows that Vermont has the largest out-of-state tuition cost on average and New Mexico is the most affordable state.



2.2.6 Comparing in/out-of state tuition

This section compares the difference between in/out-of-state tuition in different states. As can be seen in the following plots, the cost difference of in/out-of-state tuition in Colorado is larger compared to other states. One interesting point in boxplot is the median of difference with zero value for some states. This shows that in these states the average in-state tuition is equal to average

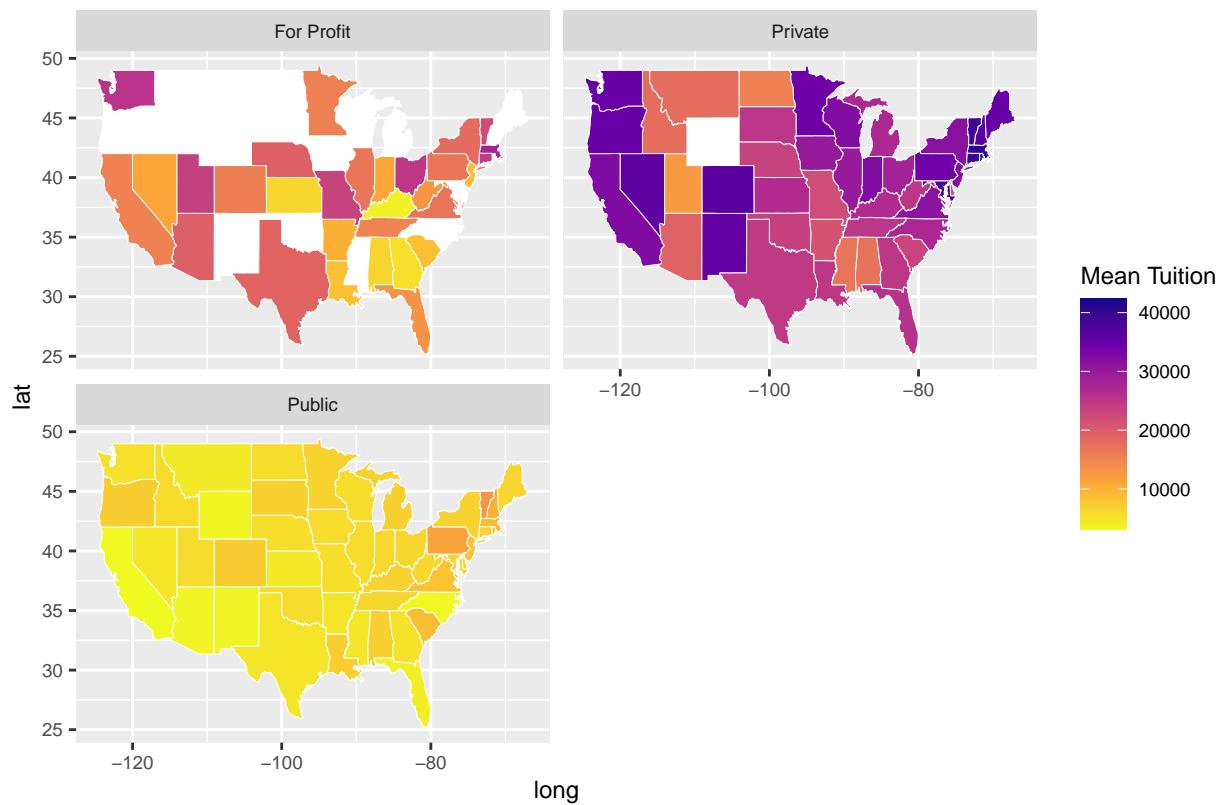
out-of-state tuition.



2.2.7 In-state tuition, different type of colleges; What is the relationship?

Let's take a closer look at the geographical distribution of average in-state tuition costs for different types of colleges in the US. The following plots illustrate the distribution of average tuition costs on the US map. As can be seen, average tuition cost varies in different states for private and for-profit colleges. However, this statistic remains almost the same for public colleges across the country. It should be mentioned that white color in maps means that data is unavailable for the corresponding college type.

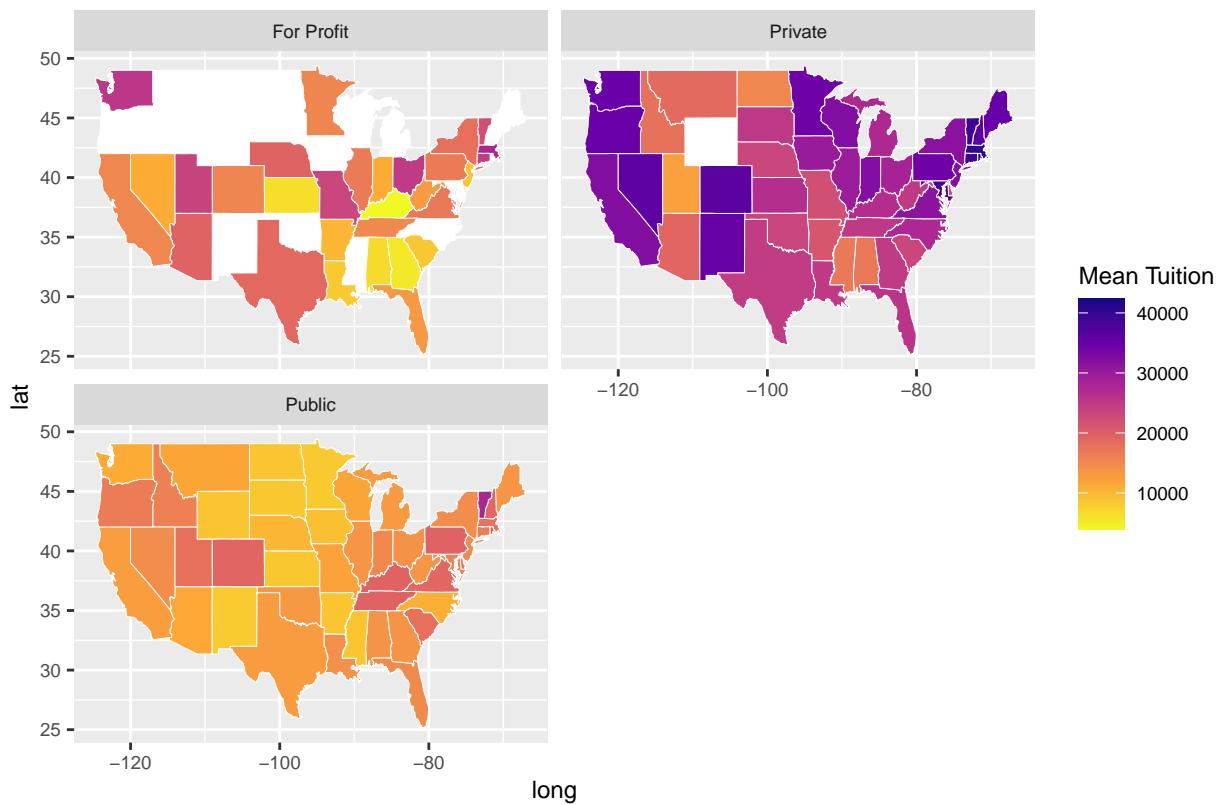
Mean in-state tuition of colleges in U.S. States



2.2.8 Out-of-state tuition, different type of colleges; What is the relationship?

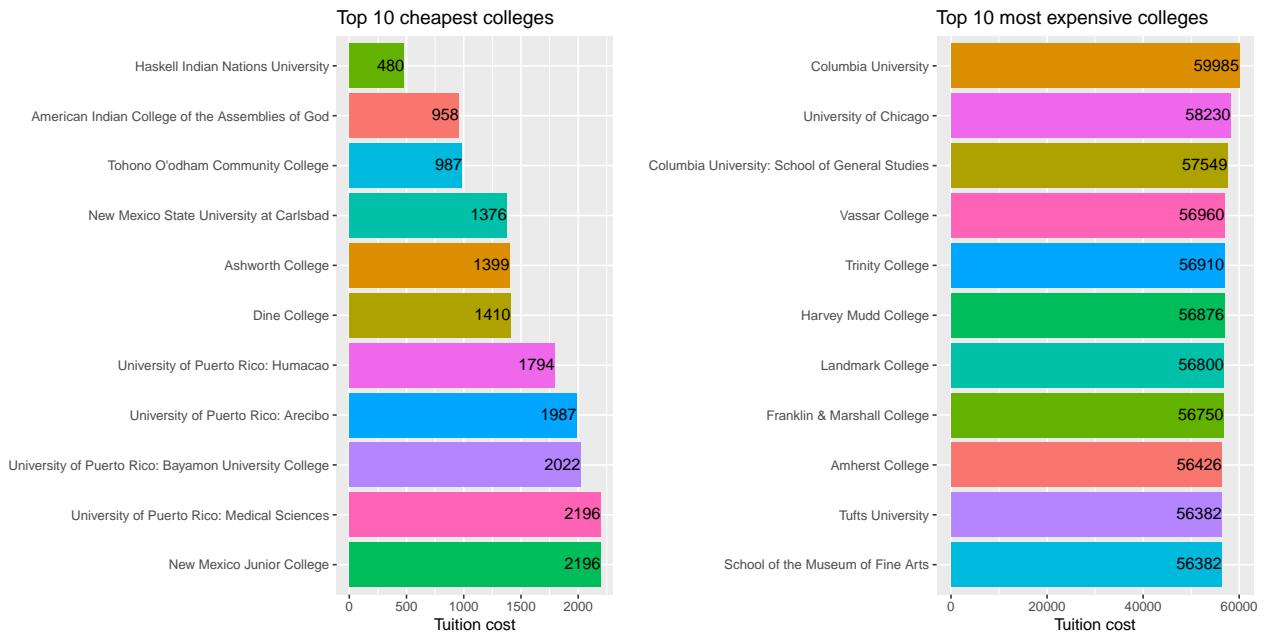
The following plots provide the same analysis for the out-of-state tuition cost. The interesting point is the larger variation of average out-of-state tuition costs in different states.

Mean out-of-state tuition of colleges in U.S. States



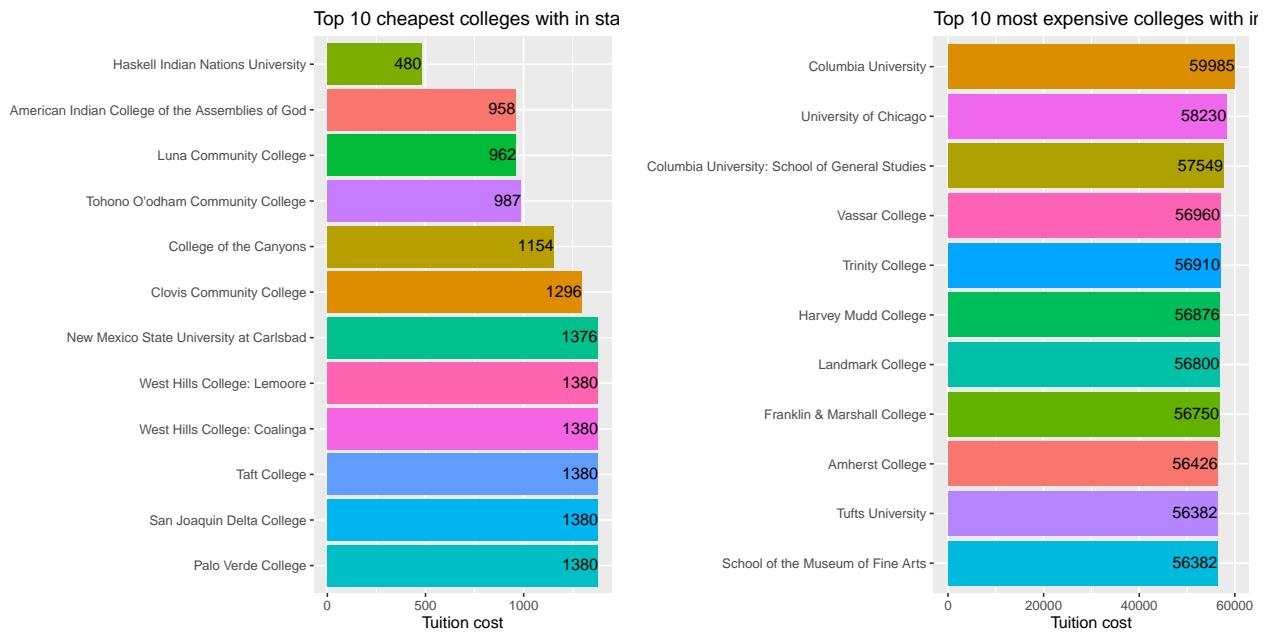
2.2.9 The cheapest/most expensive colleges with out-of-state tuition

“Which one is the cheapest one?” or “which one is the most expensive one?” always are interesting questions. This section provides some insights and helps us to answer these two questions. According to the following plots, Haskell Indian Nations University is the cheapest college in the US, and Columbia University is the most expensive college in terms of out-of-state tuition costs.



2.2.9.1 The cheapest/most expensive colleges with in-state tuition

This section addresses the two questions of the previous section, this time for in-state tuition. According to the following bar charts, the cheapest and the most expensive colleges are the same as out-of-state tuition. The reason is that in-state tuition costs and out-of-state tuition costs are equal in Haskell Indian Nations University and Columbia University.



2.3 The historical tuition dataset

This data set has 270 observations and 4 variables. The following table shows these variables.

```
## Rows: 270
## Columns: 4
## $ type      <chr> "All Institutions", "All Institutions", "All Instituti...
## $ year       <chr> "1985-86", "1985-86", "1985-86", "1985-86", "1985-86",...
## $ tuition_type <chr> "All Constant", "4 Year Constant", "2 Year Constant", ...
## $ tuition_cost <dbl> 10893, 12274, 7508, 4885, 5504, 3367, 13822, 16224, 74...
```

type	year	tuition_type	tuition_cost
All Institutions	1985-86	All Constant	10893
All Institutions	1985-86	4 Year Constant	12274
All Institutions	1985-86	2 Year Constant	7508
All Institutions	1985-86	All Current	4885
All Institutions	1985-86	4 Year Current	5504
All Institutions	1985-86	2 Year Current	3367

The type of college can be either public, private or all institutions and the type of tuition can be either a 2-year, 4-year or all programs. The difference between current and constant is in considering inflation.

2.3.1 Data Manipulation

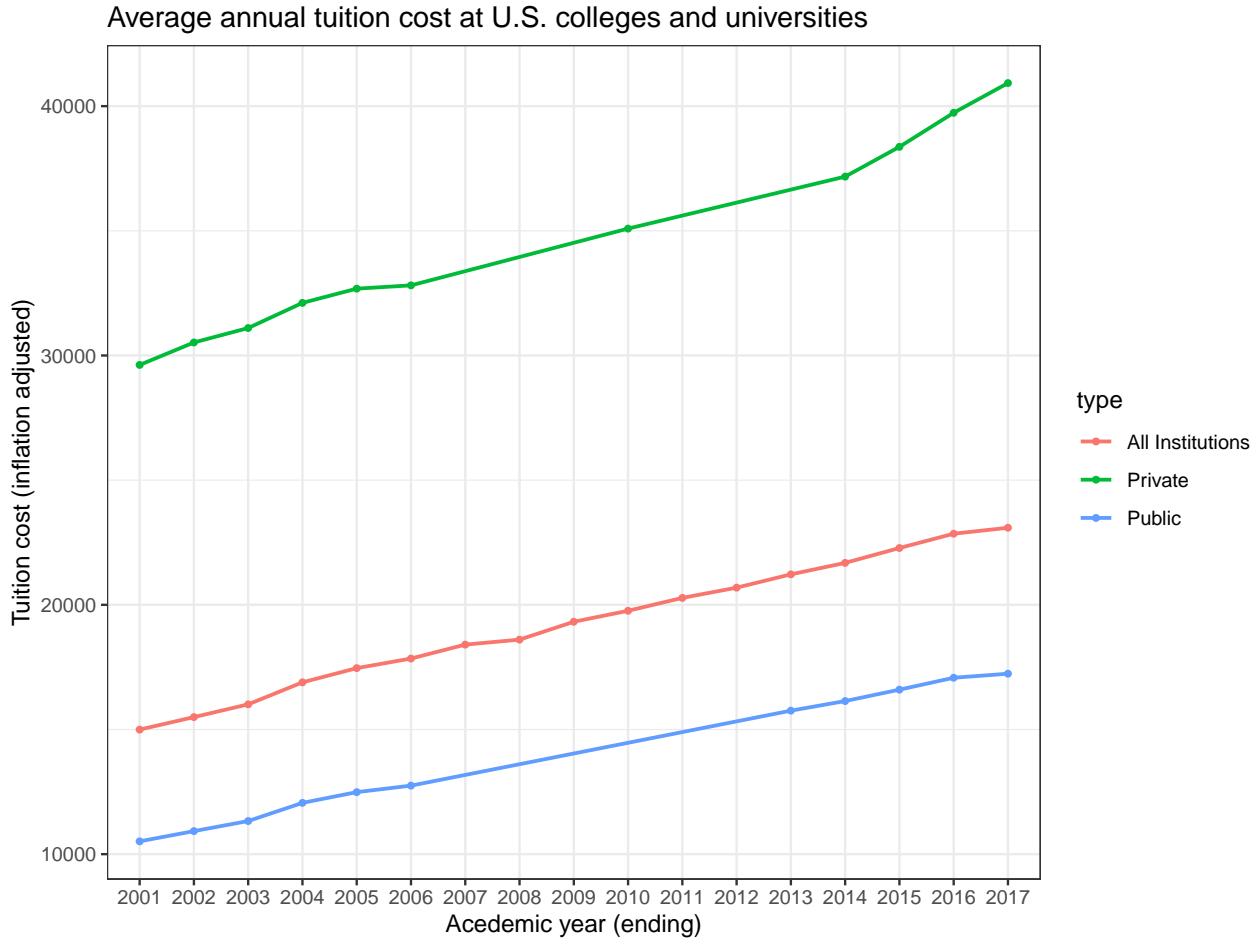
Aftyer investigating the data, we observed that the data is available consistantly from 2000 till 2017. So we decided to just work on data after 2000. For convinience, we will represent each academic year by its end date.

```
## # A tibble: 17 x 1
##   year_end
##       <dbl>
## 1     2001
```

```
## 2    2002
## 3    2003
## 4    2004
## 5    2005
## 6    2006
## 7    2007
## 8    2008
## 9    2009
## 10   2010
## 11   2011
## 12   2012
## 13   2013
## 14   2014
## 15   2015
## 16   2016
## 17   2017
```

2.3.2 Average tuition cost over time

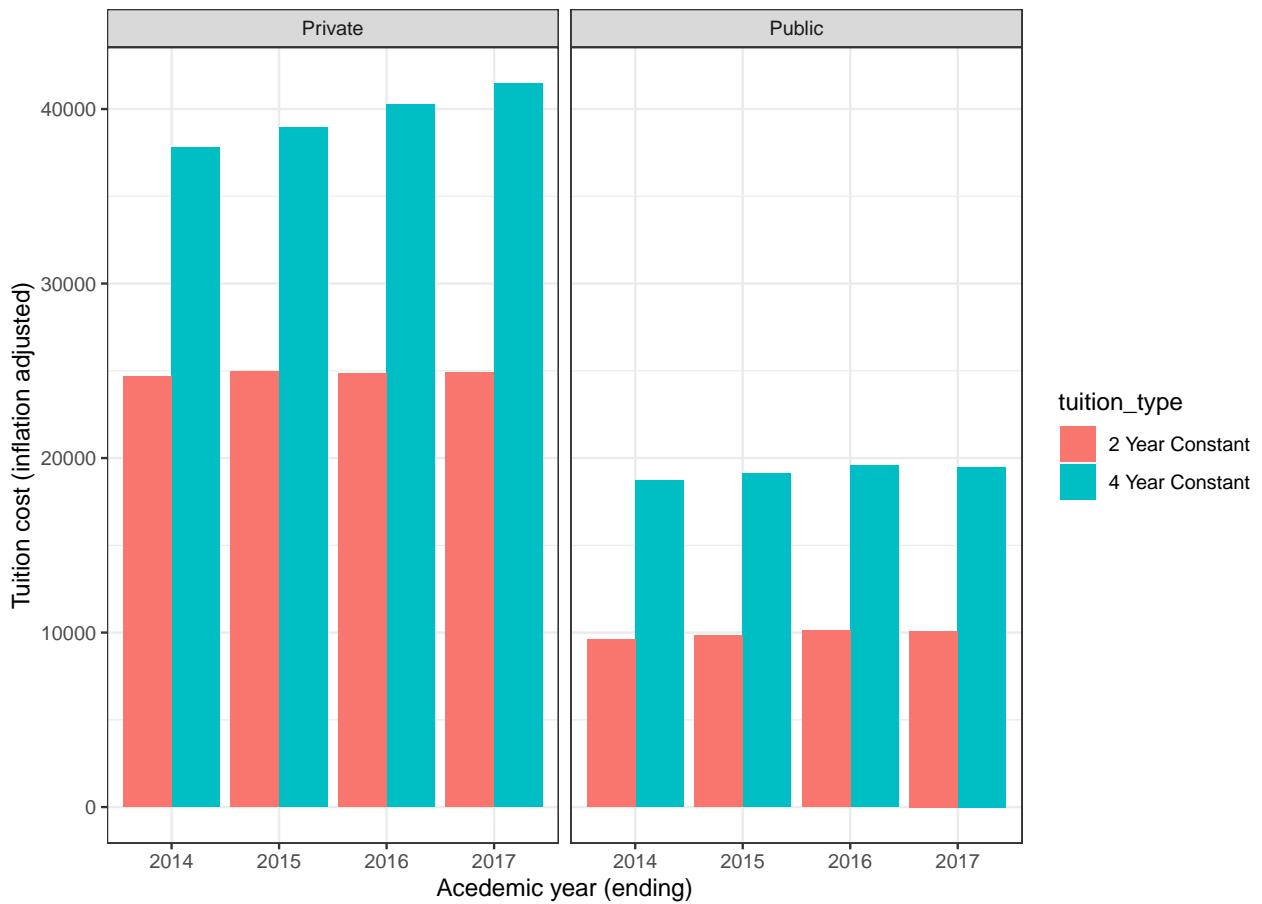
Here, we are looking into the average annual tuition cost at U.S. colleges and universities. We see that as time goes, tuition increases. The private schools have considerably higher average than the public schools.



2.3.3 Tuition comparison between 2 year and 4 year programs

In this plot, we are looking into the difference between 2 and 4 year programs in public and private schools in a time window of 4 years (2014-2017). It should be noted that for tuition cost the inflation has been adjusted. We do not see a big difference among 2 year programs. However, for 4-year programs we see the average tuition has increased over time. This increase is more considerable among private schools.

Public and Private tuitions among 4 year and 2 year progerams



2.4 Salary potential dataset

This data set has 935 observations and 6 variables. For each university in a state, we have the information about average early/mid career pay and also the stem percent.

```
## Rows: 935
## Columns: 7
## $ rank                  <dbl> 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13...
## $ name                  <chr> "Auburn University", "University of Alabama", ...
## $ state_name             <chr> "Alabama", "Alabama", "Alabama", "Alabama", ...
## $ early_career_pay       <dbl> 54400, 57500, 52300, 54500, 48400, 46600, ...
## $ mid_career_pay         <dbl> 104500, 103900, 97400, 93500, 90500, 89100, ...
## $ make_world_better_percent <dbl> 51, 59, 50, 61, 52, 53, 48, 57, 56, 58, 6...
## $ stem_percent           <dbl> 31, 45, 15, 30, 3, 12, 27, 17, 17, 20, 8, ...
```

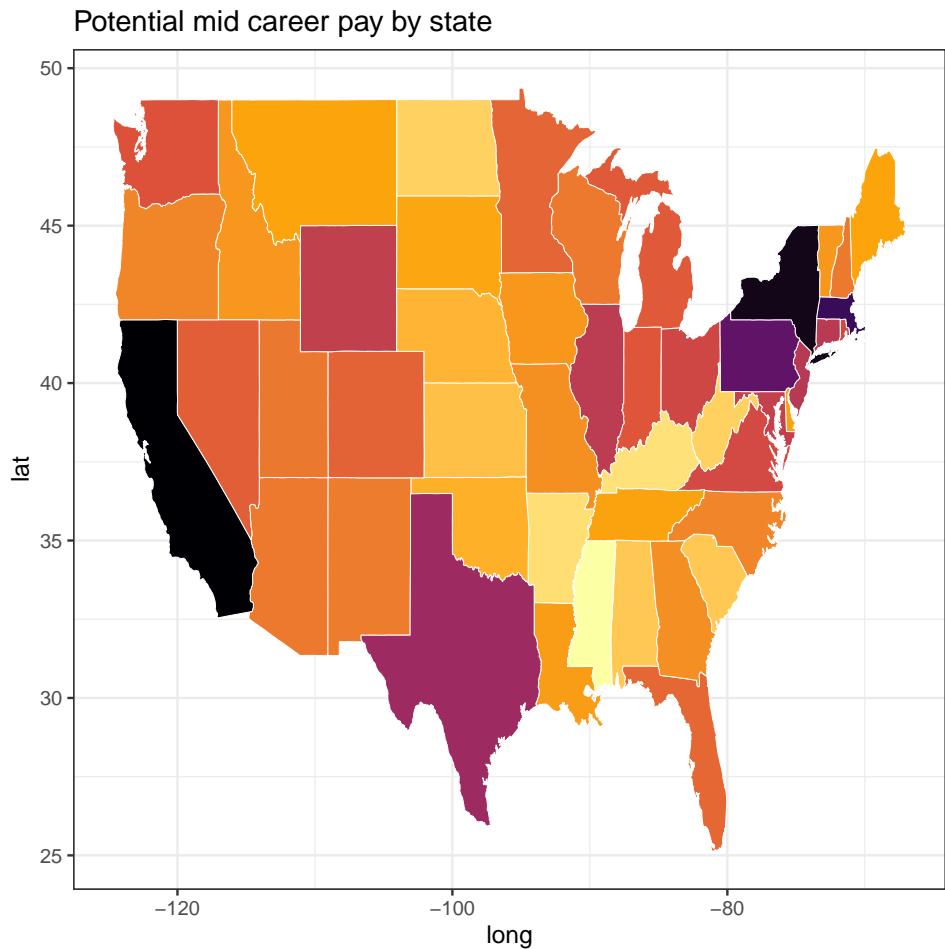
2.4.1 Which states have the highest/least potential salary?

We are interested to know which states are having highest amount of potential salary. Looking at the top 5 states we have California, New-York, Massachusetts, Pennsylvania, and Texas with the highest salary with respect to both mid and early career pay.

state_name	average_mid_career_pay
California	123976
New-York	122328
Massachusetts	115712
Pennsylvania	110884
Texas	103476

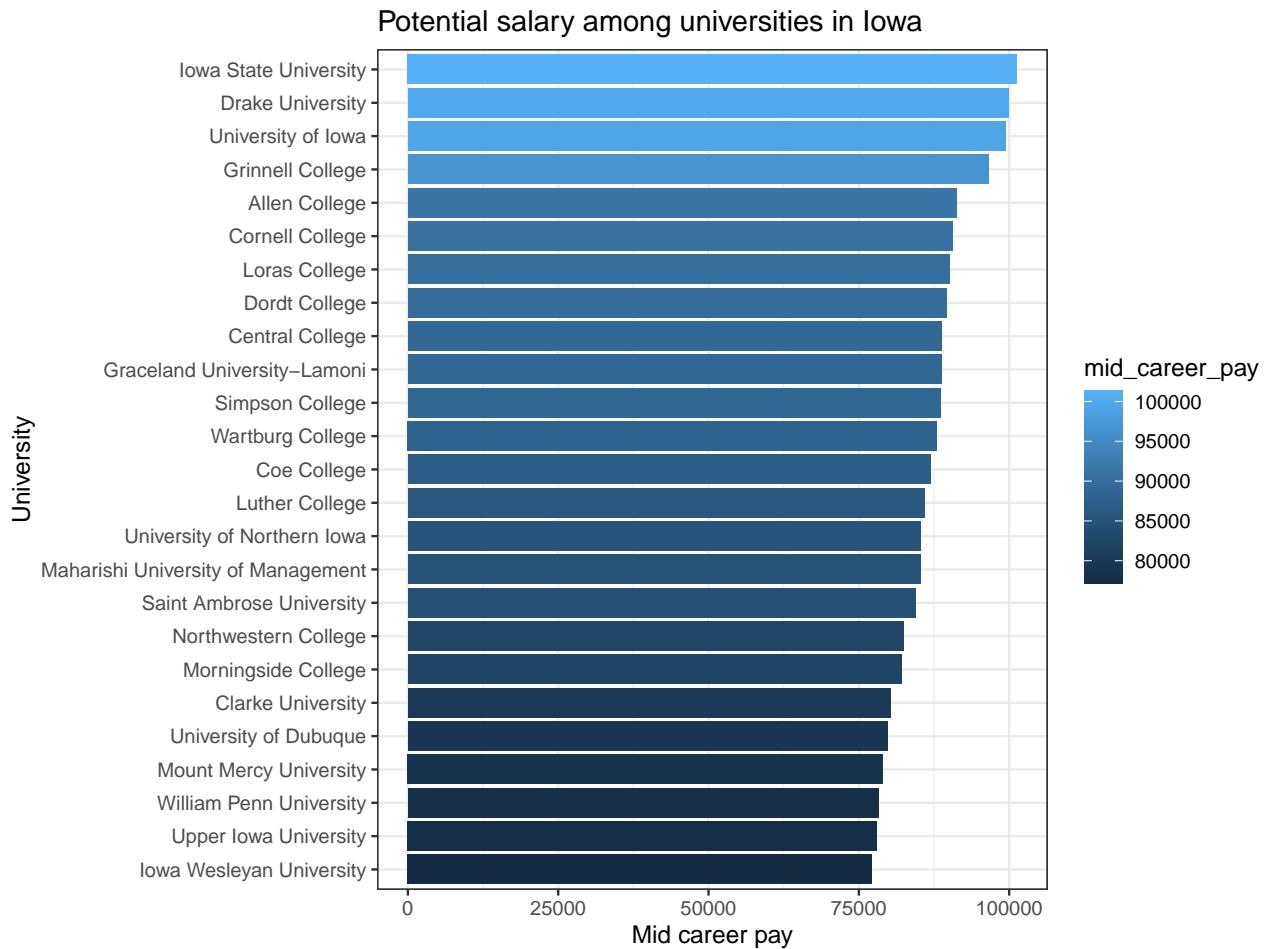
state_name	average_early_career_pay
California	67232
New-York	66688
Massachusetts	63100
Pennsylvania	60644
Texas	57468

The following map demonstrates the potential mid career pay by state which approves our previous findings regarding top 5 states.



2.4.2 Potential salary in Iowa

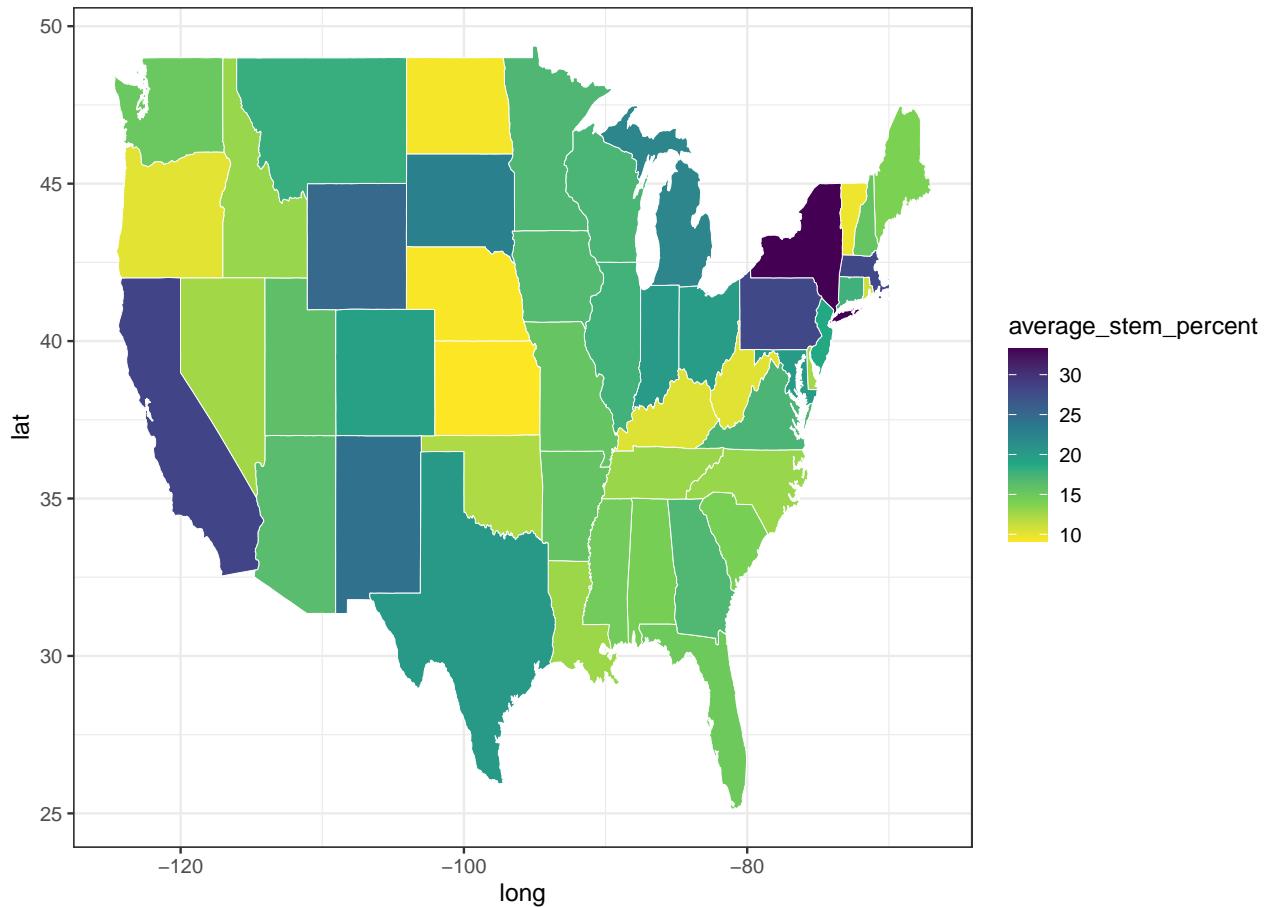
We are interested to know which university has the highest potential salary in Iowa. The following figure demonstrates mid career pay across different universities. We see that ISU has the highest potential salary in Iowa.



2.4.3 What is average stem percent across different states?

Here, we explore the stem percent across different states. It is interesting that the states which had higher amount of potential salary are also the ones with higher stem percent.

Stem percent across different states



3 Conclusions

In this project, we worked with different datasets related to college tuition, diversity and potential salary. We applied different techniques such as mergeing, reshaping, removing missing values and worked with different types of data including categorical, continuous and data time variables. We demonstrated different ideas through visualizations in R. Specifically, we observe that the US colleges are not equally diverse geographically and there is relationship between different diversity groups, tuition and early career salary. We also find out that the average out-of-state tuition differs across the country while average in-state tuition remains almost the same and east coast colleges are more expensive in terms of in-state and out-of-state tuition. Furthermore, we explored the historical tuition and observed that private schools cost more than public schools. The historical trend of tuition showed that tuition is increasing over time. The last data set that we investigated

was the potential salary data which showed universities located in west/east coast have higher stem percent and also higher potential salary.