

CS 4460 - Project 5

Thinh Nguyen - tnguyen452

Nhat Lan Le Tu - ntu9

College Dataset

Professor Endert

Overview

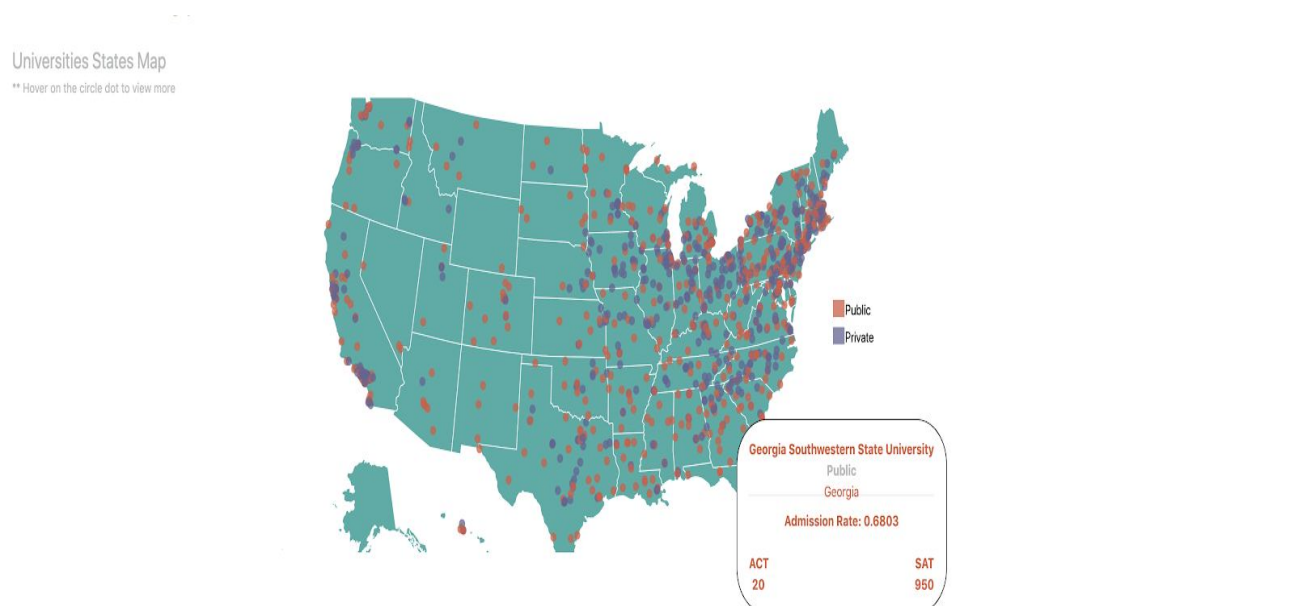
For this assignment, our group decided to use college dataset. The college dataset provides a lot of different attributes such as information of each school in the United States. Our group wanted to see the difference from each school over thousands of other schools in the dataset. The goal from this dataset is to focus on finding out the correlation between the school admission rate including acceptance rate, SAT vs ACT and school type, also the salary of students employed earned after 8 years, and the average tuition of each institute.

Detailed Implementation and analysis

1. Universities state map - School coordinates

The first visualization was implemented using the schools' coordinates in the Map chart. Even though the given raw dataset did not provide any kind of coordinates for each school in the U.S, our group took advantage of Google Map API and integrated it using a helper method written in Python script to implement and collect the coordinates such as longitude and latitude from the dataset. After querying and mapping all the coordinates with the schools'

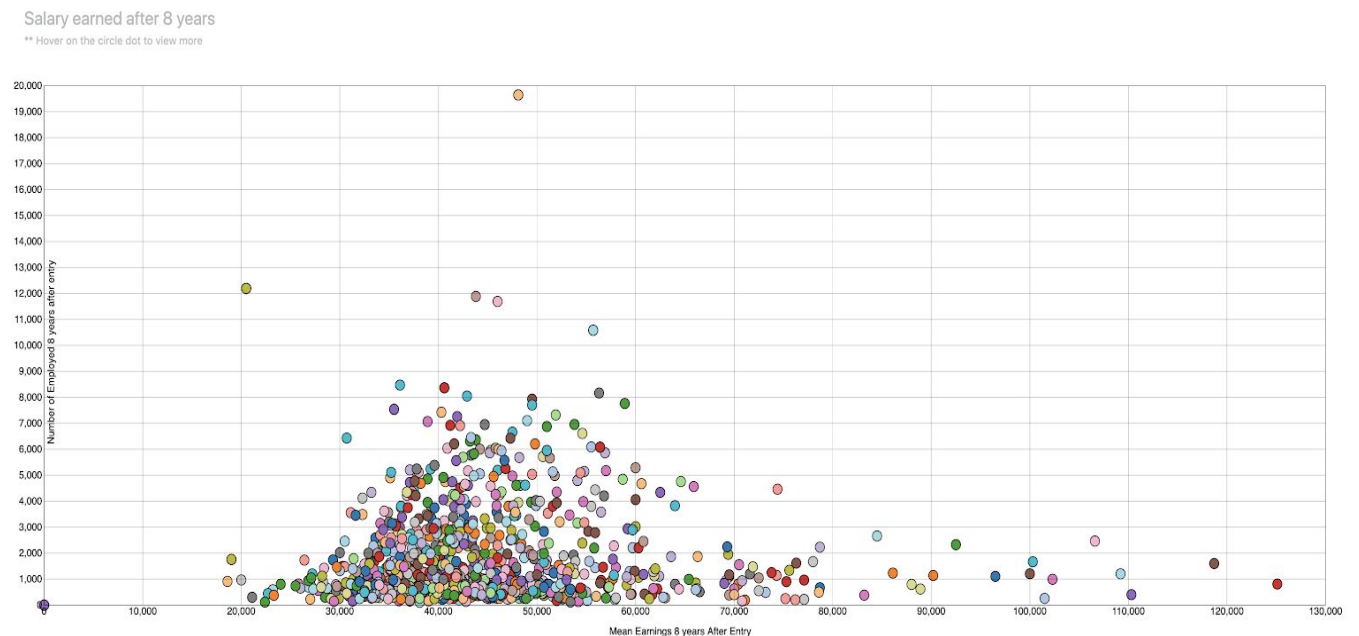
names, our group decided to export data to another CSV file named `collegesLocation.csv` instead of adding to the original dataset. Moreover, I also use `us-states.json` file that consists of all the coordinates range in each state so we can use and map it with the coordinates that we collected in the earlier step. This visualization is composed of two parts. The state map on the left, and the legend information dialog on the right. The legend dialog will just show the type of school in each circle on the map with the color match. The state map will show each component by its own school type all over places in each state. Whenever the users hover the mouse on top of each circle, it will bring up a little dialog with some basic school information and admission rate. Analytical tasks supported by this graph are identified, locate, categorize, distinguish, compare within relations. By using this graph, we can easily show where the schools are located and what the types are.



Some analytical questions could be answered from this visualization:

- What states have the most private or public institution?
- Which region has the most schools?
- For example, Georgia state, which school has the highest admission rate?

2. Employed salaries earned after 8 years - Scatter plot chart



Scatter plot charts are common use to compare the different subject with other attributes. It is highly recommended using when we want to see the correlation of one variable is affected by another variable. Therefore our group decided to use the same idea and migrate our dataset of employed salaries earned after 8 years graduated from school. By plotting those data in x and y coordinates, we can easily visualize which school has the highest salaries earned or which school has the highest number of employed by hovering on each circle dot for few second. In this

visualization, the x-axis represents the mean earning 8 years after entry, on the other hands, the y-axis represents the number employed 8 years after entry. Analytical tasks supported by this graph are locate, compare within relations.

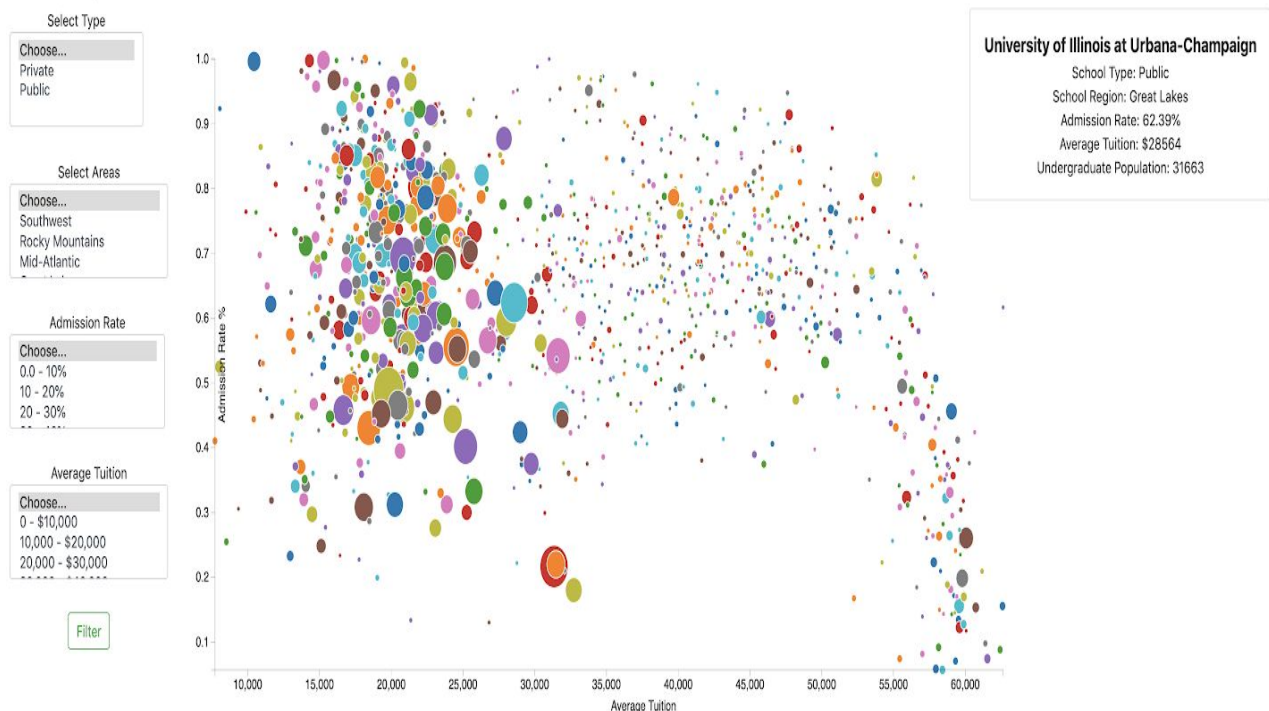
Some analytical questions could be answered from this visualization:

- Which schools have the highest mean earned?
- Which schools have the highest employed number after graduate?
- Which control type has more chance to be employed after graduate?

3. Universities admission vs average tuition fee - Scatter plot + Filter

Universities Admission and Tuition Fee

** To select multiple field hold down Command. To reset select choose... and hit filter



For the last visualization, our group uses the same technique of scatter plot chart to represent our ideas to compare between thousand of school in the dataset. However, in this chart, our group decided to take advantage of bringing up the filter functionality, which allows the users to filter whatever the information they are looking for such as private or public school, region or tuition range. Moreover, each circle that represents a school, and the radius of that circle stand for the population of the school. This visualization is composed of three parts. The first part on the left-hand side, allows the user to add the information they want to filter. The middle part is where the graph represents all school in the U.S in x and y coordinates. The last part on the right side shows the school's detailed information after the user clicks on the circle in the graph. In this chart, the vertical axis represents the admission rate in percentage while the horizontal axis represents the average tuition cost of the school. This graph supports the users to easily determine which school that must fit with their weight and cost. Furthermore, since the graph represented base on the average tuition on the x-axis in the ascending order; therefore, we can tell almost all the schools that lay on the right on the graph are the private institutions. Analytical tasks supported by this graph are locate, categorize, rank, compare within relations. Some analytical questions could be answered from this visualization:

- Which school in the United States has the most expensive tuition?
- Which institute has the largest population?
- What is the average range of tuition and admission rate of all the school in the U.S?