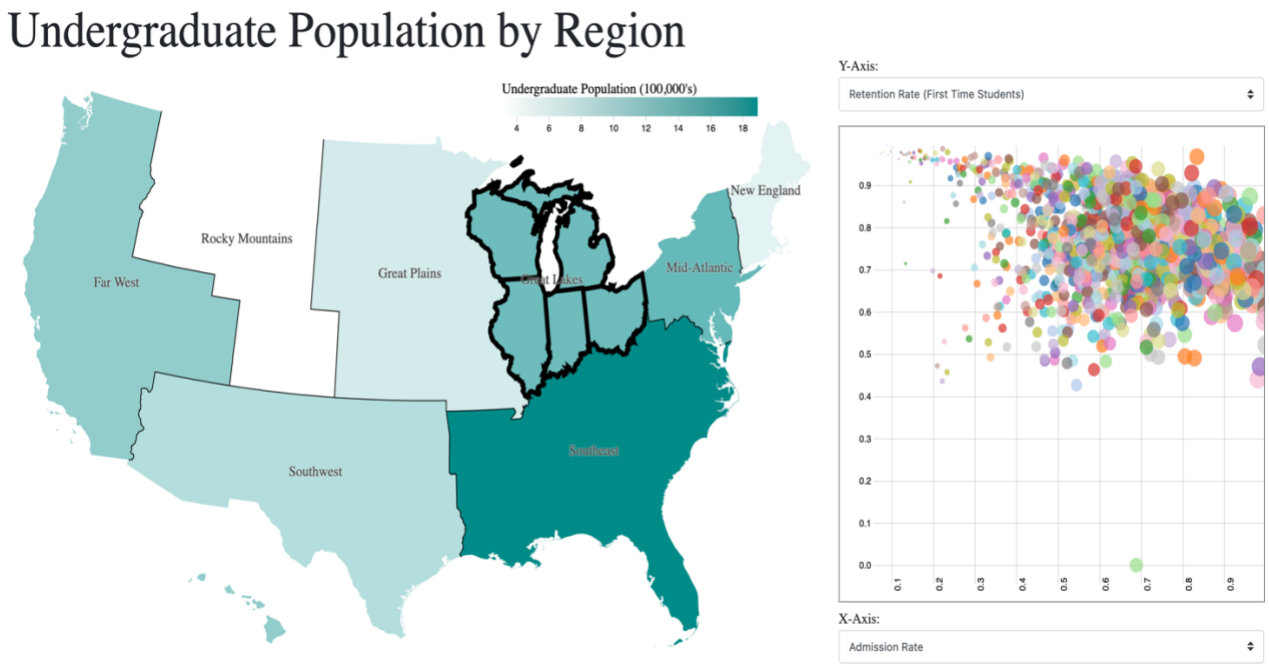
Lab7: Final Project
Name: Tam Nguyen – CS4460 Spring 2021
Dataset chosen: Colleges.csv

Part1: Overall description

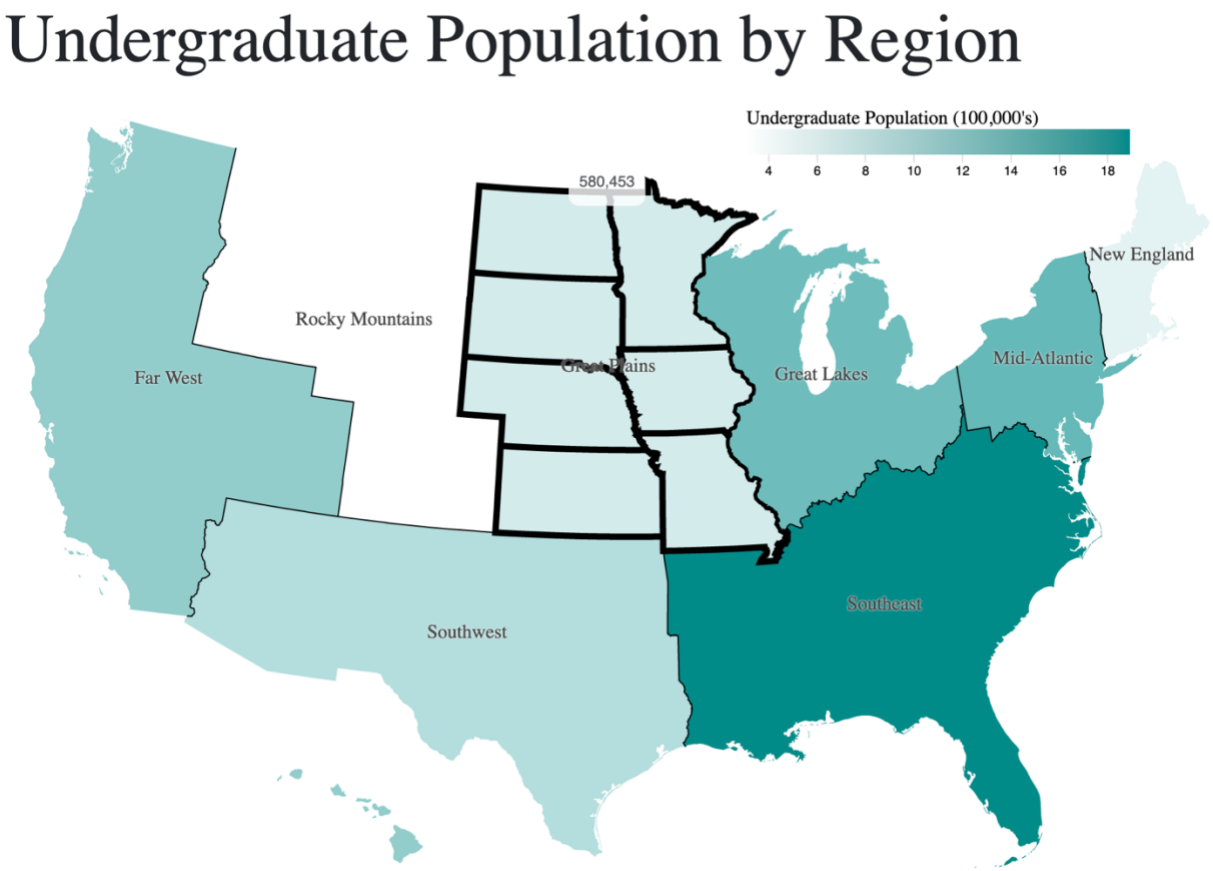
I have decided to use the college dataset to determine how many college students are in the U.S. This dataset comprises a wide variety of university, such as public and private universities. It's come up with the idea to get a full background of universities in the U.S like they have been showed on the scatterplot. Comparing the information of XY axis and visualized in another format such as U.S map is becoming popular. If you're searching for how the number of college students in the U.S is, it's a helpful way to see how university at each region when large volumes of complex statistical data need to be made comprehensible.

List of analytic tasks: Retrieve values, read in the data and present an interactive visualization, integrate data mining and its information, allow users to specify what they are seeking using multiple feature selections and computational techniques, so that the number of attributes used to describe the dataset is less than the original file. For example, we transform data file with lots of attributes to principal components, which is mapped in pair of XY axes in the scatterplot below.



Part2: Design Overview

What is the total number of undergraduate populations in each U.S region?



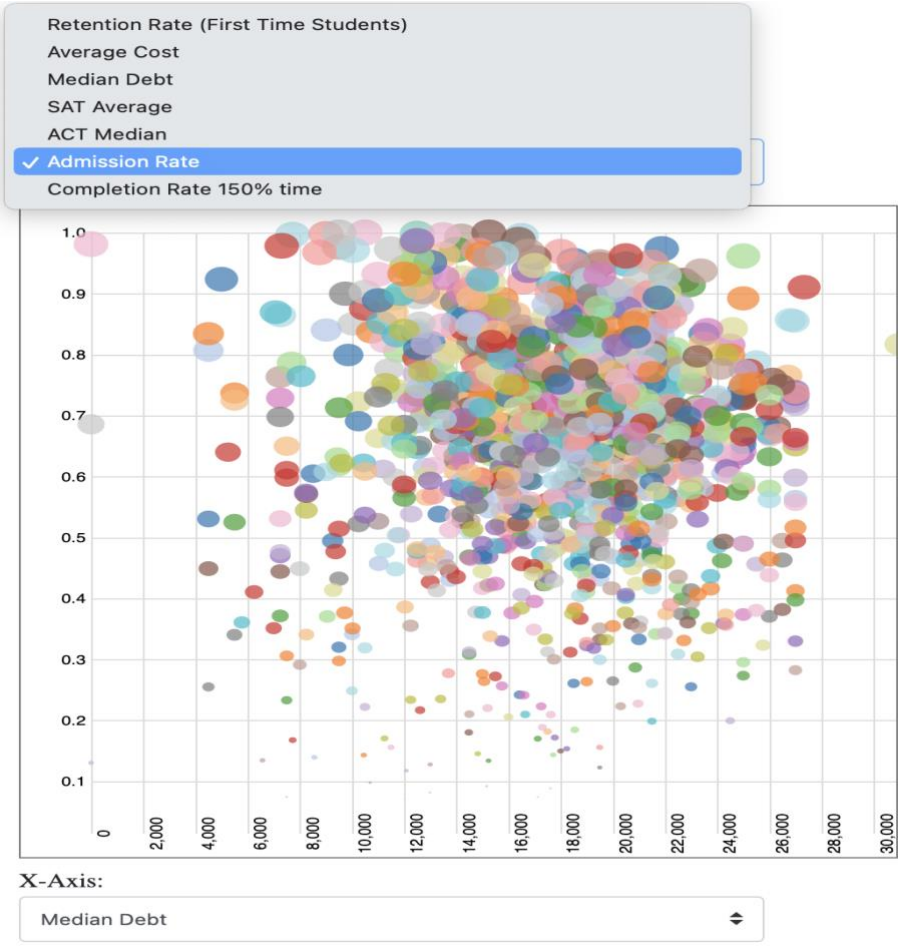
This map depicts the total number of undergraduate college students in the United States. It divides the populations into different regions, which are color-coded, and based on the legend bar chart next to the map, we can check the trend of the populations in the United States. Using this interactive, the colorful map shows details exactly where the most of the United States' undergraduate students comes from and how much population is generated. Each part in the map is bounded by a black border represents an individual region, with the key value demonstrating total number of undergraduate in the United States and the number count is showed when we click on the appropriated region.

Roughly ¾ of undergraduate populations in the United States are proper above 100,000's, with mostly in the East side. When looking at the count figured, we see that about 18,000,000 of

undergraduate population in the United States are of the Southeast region, so this data makes sense when we can say that Southeast has the greatest number of universities in the U.S. In comparison, Far West comes at a low 100,000 counts and Rocky Mountains comes at approximately 5% of the total populations.

Visually, using this colorful map helps to easily navigate and analyze in general user interface more specifically with multiple sub-set of a big dataset. Each visualization has its different strength, and this interactive is built like a summative data object.

How each university in the U.S is compared to each other with different strategies?

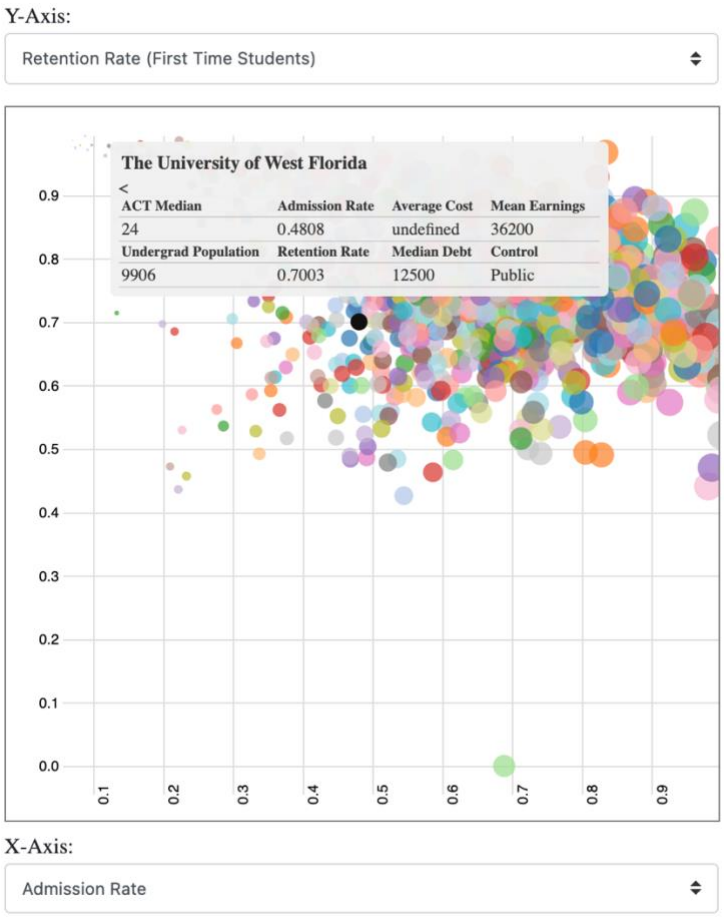


This visualization breaks down the dataset into different categories, so we can compare universities more specially. By picking a specific task with the XY axis selection, it helps the users to easily

get the big picture when the chart is glowing since it allows for sorting the data by the different features.

For example, look at the image above, we can say that the median debt strategy is mostly range from 8,000 to 26,000 versus to the admission rate of the most universities will range from 0.5% to 0.9%.

Otherwise, we can simply pick a random school by the moving the mouse. For example:



And all the information of that appropriated school is displayed on the table with the ACT Median, Admission Rate, etc... It is quite simple to use and has a lot of information for the user. Otherwise, this visualization makes use of color to encode information. Having a different color mapped to each university, and is consistent through the vis.