Edward Chiao (echiao6) / Eugene Ahn (eahn33)
CS 4460 - Alex Endert
Project 5: Putting It All Together

Project Description: A Closer Look at Ethnicities in Colleges

Visualization Design Overview

The main focus of our visualization is the two compared pie charts. We decided to focus on the racial proportions of the students that are at two different colleges. It is to our understanding that each college will differ somewhat in their student demographic, so being able to see the exact percentages reveals differences between the campuses. In implementing the colleges.csv file that we were provided, we implemented a drop-down selection menu for the user to go through. We anticipated that it is a lot of information so parse through, so the user is able to jump to a specific college simply by just clicking on the menu and typing in the words—the menu should automatically jump to the most alphabetically proximal college names. To select a college just click on the name from the menu, and the pie chart should change its appearance to match the new college's demographic. Moving one's mouse over the pie chart highlights the section and shows the percentage of that race at that college. We also added in a new data attribute, "Other", that accounts for the remaining percentage, so users can also see that in the pie chart. Along with the high versatility of the drop-down menu also comes a factor of exploration: the user may not know what college to look at, but has full freedom to explore the other colleges in the menu.

As the ethnic representations are the most noticeable part of the visualization, we also made sure to include other pieces of information about the selected colleges. This way the user is not fed just that singular attribute regarding the college, but has the ability to read through other attributes/statistics for those colleges. We thought that it would be best to not limit the knowledge that is presented, but to give flexibility in what the user can continue to look at. At the same time we did choose to exclude many of the attributes listed in the CSV file because presenting all attributes is the other extreme in which there may be too much data shown, which would also clutter the visualization. Presenting the information side by side on two sides is useful because users are able to just look across the screens and analyze differences.

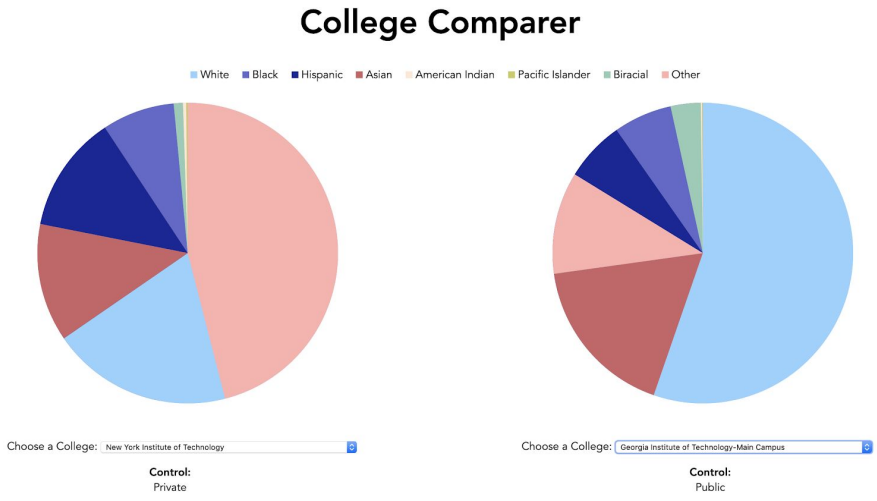
In looking at potential tasks a user may have in navigating this visualization we can definitely prioritize comparing and contrasting. We believe that on their own, colleges and the different statistics/numerical values they have can be quite meaningless if not compared to other colleges. This way users have an idea of the stats in relation to each other, and this opens the door for many other questions to ask. We definitely highlighted ethnicity, but one goal we

wanted for users would be to have them formulate questions of their own. Some examples could be: “Is there a trend in ethnicity among colleges in a particular region? Why?”, “Do public schools differ from private schools in terms of ethnicity or other characteristics?” “What could be the explanation for why ethnicity X is overrepresented or underrepresented in college demographics? Does this representation aris only in relation to other races?”

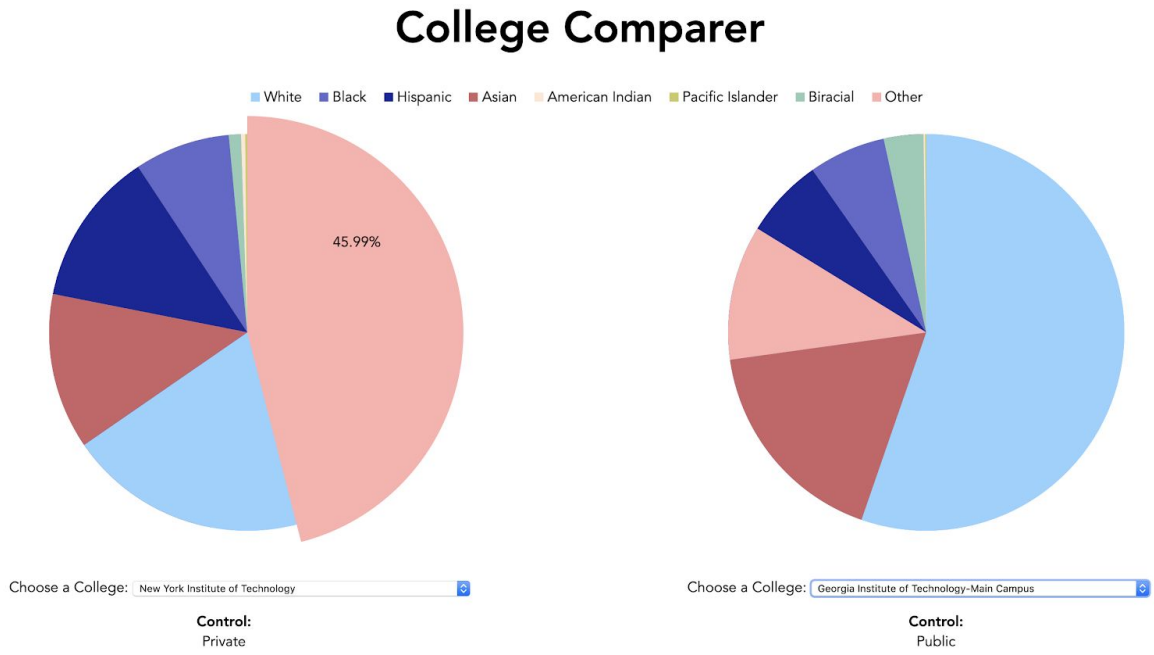
There are many other analytic tasks that are handled by this vis. Retrieving value is done simply by selecting colleges (albeit not all attributes are given). Because of the easily accessible comparisons that could take place, users can also characterize the distribution (with enough comparison), and can find correlations between data with similar attributes. Having a disproportionate amount of one race highlights anomalies in the data (given we are seeking colleges with more “balanced” ethnic demographics. These, of course, are much lower level analytic tasks, but as stated before, users have full freedom to deviate to higher analytic tasks (such as the questions listed above).

We discussed in lecture the mantra of “overview first, zoom and filter, details on demand” The visualization that we created represents this. We have an overview of all the colleges to choose from, as well as a general pie chart representing the proportions of different races. There is a general overview of the names of all the colleges the user can select from. The step of zooming and filtering comes when the user selects a specific college from the menu. The filtering process is just by name, not by other attributes, but zooming would be focusing in on that singular college. Details on demand comes in with the chart display of the college information. We’ve explained the reasoning for why we chose not to show all the details, but the visualization is very straightforward in that selecting the college gives a clear representation of the data we want users to see.

<u>Visualization Characteristics</u>



Above is a screenshot of the visualization (some of the bottom is cut off). This visualization is good in that there isn't an overwhelming amount of information presented at once, and the main focus of the visualization (the pie charts) is clearly visible to the viewers. The options to change the college shown is easily accessible, and the pie charts themselves are easily interpretable based off the key. Hovering over the pie chart sections will reveal more information like the specific percentages, so this way if the user wants to know more about the specifics, they will take the time to hover their mouse over it. You can see this in the image below.



In all, we made sure to reveal enough information in our visualization "at a glance" but not too much to where the user has too much information to process. We know that for all people, they can only process some amount of data at a time, so we designed our visualization to accomodate for a good amount of data reveal at a time. Our end result is a simple visualization that allows for deep comparison.