Stacee Freeman
Scott Higgins

Data Set: Colleges

Project Overview:
The college dataset had a large quantity of data that focused on over 1200 colleges and detailed almost 40 variables for each college. In order to start forming our visualization, we decided that we would focus the colleges within the United States. We wrote a script to search for and scrape the latitude and longitude for each university/institution. This allowed us to set up our central map view. Afterwards, we selected a primary focus of the visualization: providing details to the user and a way to filter and explore/search through the college dataset in a meaningful and coordinated way. While there are several views that make up our overall view, they work together to provide an overview of the dataset first, then allow the user to filter based on some primary attributes, and finally provide a user the way to get more details on demand via hove and click interactions (following the information visualization set up by David Schneiderman). We will first discuss the analytical tasks that our visualization supports, and then we will use screenshots to highlight and discuss the salient features.

Analytical Tasks:
- Identify Physical Distribution of Colleges
    - The geo-spatial layout of colleges can be an important factor for many students or teachers researching a college. Identifying whether a college is isolated, close to a large number of colleges, or near/far from their ideal location can be a very important task. The map component of our visualization makes this task simple.
- Find the relationships between standardized test scores and admission rates
    - Via the parallel coordinates graph at the bottom of our visualization, the user can easily see this relationship. The inverse relationship between admission ratings and ACT scores demonstrates this principle by crossing lines. Obviously, really high scores tend to correlate with schools that have lower admission rates. This brushing along the axes of the parallel coordinates graph allows a user to select their scores on the ACT/SAT (or just any scores they are interested in) and see the admission rates for schools that fall into that category.
- Identify racial diversity within a school
    - Selecting a school from the map or the list of school names on the right of our visualization brings up a detail view (on the left-hand side of the visualization) that shows a bar graph distribution of the racial composition of a school.
- Identify key financial statistics for a school
    - By the same process as the task above, users can easily view the most pertinent financial statistics related to a school, including: cost of attendance, mean earnings after school, average instructor salary, etc.
- What region does a school belong to

- ○ If the map doesn't make it clear enough, there is a consistent color scheme throughout all of the views of the visualization that categorizes each of the schools to a region of the United States (e.g. Southeast, Far West, New England, etc.).
- View how common statistics relate to college size
  - ○ It was interesting to see that most schools with high standardized test scores also had smaller student body sizes. That was just one of the many patterns we identified based on the parallel coordinates chart and it's unique description of many connected statistics.
- Selecting a college
  - ○ At a high level, this visualization provides users with the ability to make a decision on a college based on all of the information contained within the visualization.

Design Overview:
Our information visualization centers around the interactions of 4 distinct views to create the overall visualization. Each view is connected and linked to another, so that it is a unified system as opposed to a piece-by-piece conglomeration of views. Upon loading the visualization, users will see a map of the United States with all of the colleges plotted in their geographical region as dots (colored by the region the school is in). In the upper right-hand corner there is a scroll box that contains a list of the names of all of the schools that are visible on the map. The bottom half of the map contains a parallel coordinates view that plots each school as a line, with the path being determined by the school's admission rate, median ACT score, mean SAT score, and undergraduate population. Finally, in the upper left-hand corner is a school detail view, which starts out blank but is filled with multiple views (controlled from a drop-box picker beneath it) as soon as a user clicks on a dot on the map or a name in the list.

The intention is for a user to start out with a view of all of the colleges and then begin to filter and decrease the size of the group based on the four main attributes below. By clicking and dragging along the axes, a user can create a filter brush that will update the schools shown in the map and names list as the brush is manipulated. The user can create a brush for each of the 4 axes independently and the visualization will respond accordingly (up to 4 active brushes, but only 1 per axis). After the user has filtered out the total number of schools, they can proceed in a few ways. They can choose to click on the name of a school in the list, which will populate the detail box in the top-left with a graph on diversity. They can then choose to change the diversity graph to a financial overview if they would like. Instead of clicking on a name, the user can hover over any dot in the map at any time and get a tooltip table that displays main stats and school name. Clicking on a school in the map has the same effect as clicking on a name in the list. With this flow of actions, users can filter the large group of schools down to a more specific group and learn about each school within that customized group. The intention would be that a user can quickly group schools based on similar characteristics, region, name, or however they would like. After filtering, the user can then learn a lot of the details about the school through a few interactions, and then return back to overall visualization to adjust the filters to show more schools or refine the criteria and focus on an even smaller group. Transitions and visual effects have been included to make the appearance/disappearance of map dots smoother, and the d3.queue() functionality is used to make sure all datasets (college and geographic) load together instead of separately.
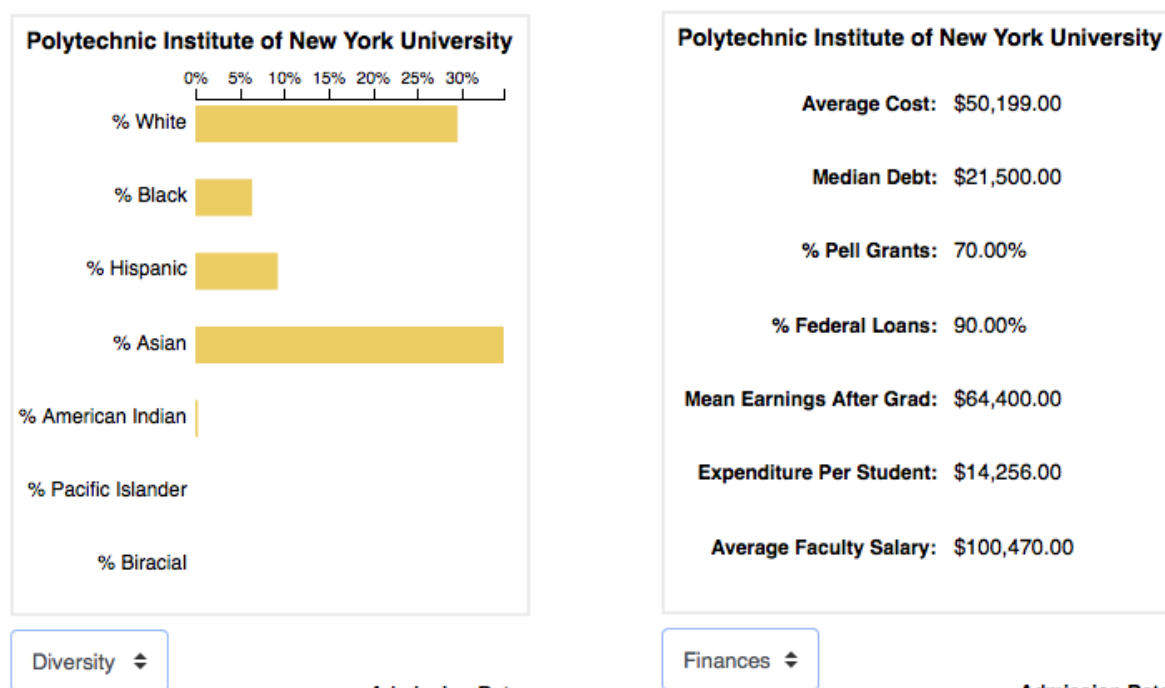
Screenshots:
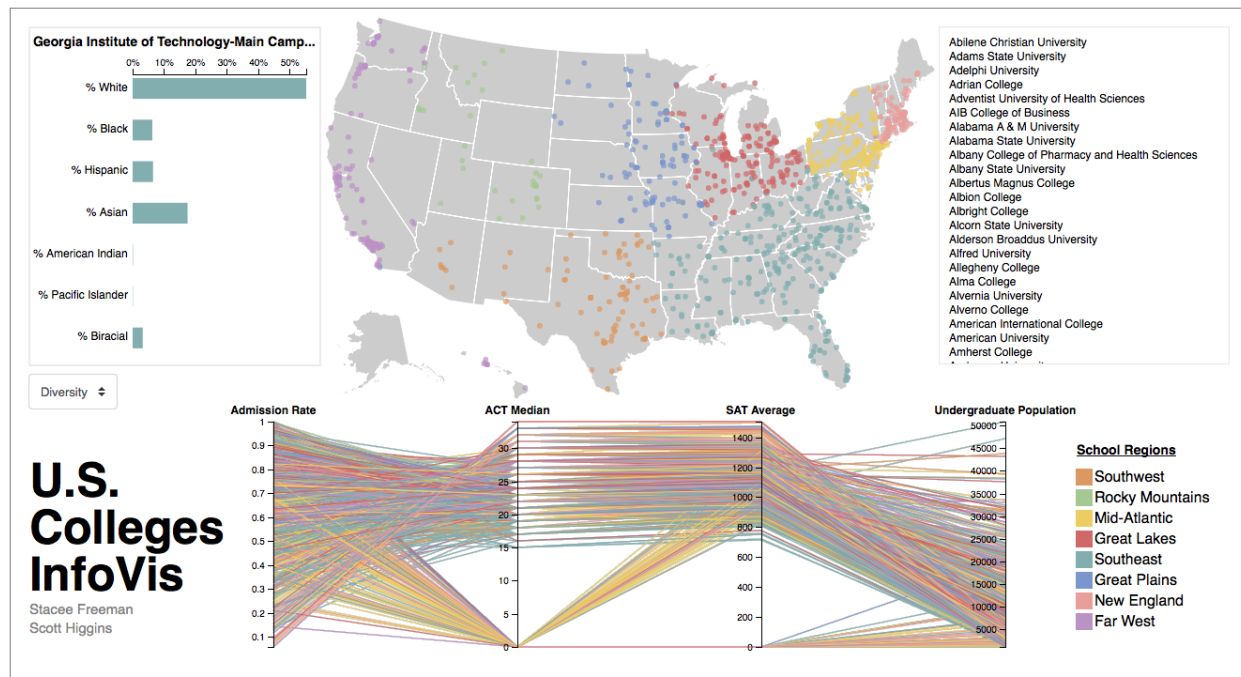
The screenshot above shows the initial visualization displaying data for all of the colleges from the dataset (located within the United States, as that was our focus).
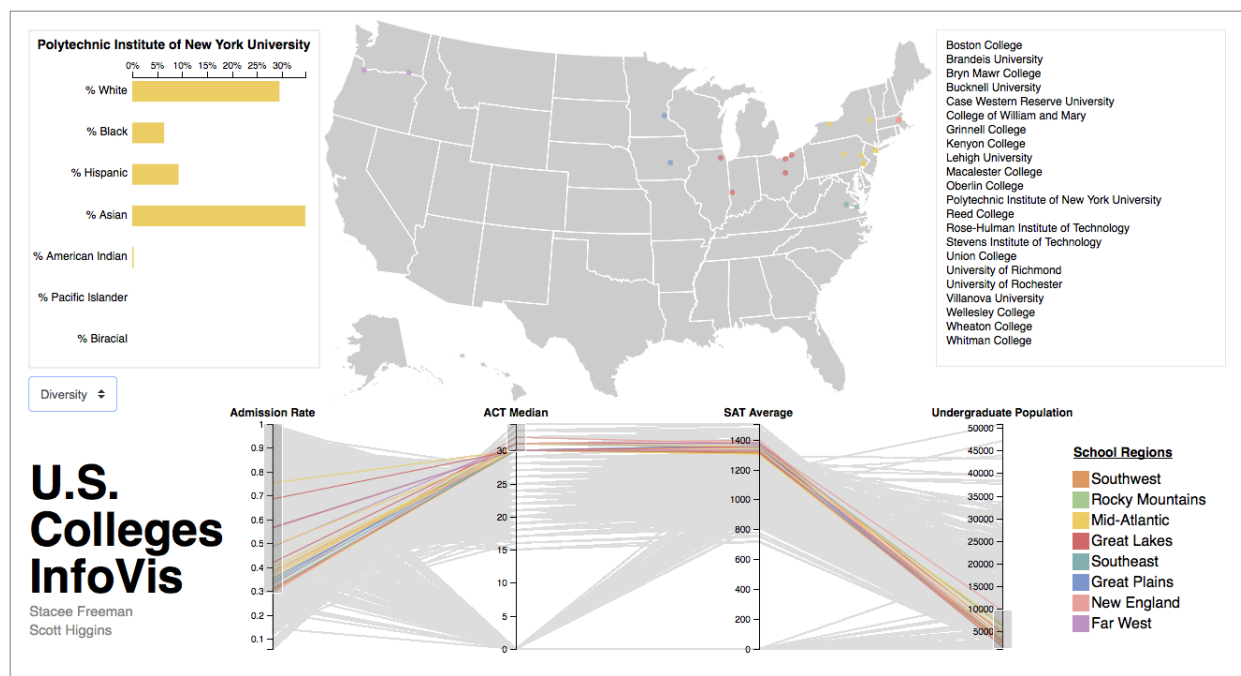

The screenshot above demonstrates how a user could utilize the system to research a specific type of college. The user used the parallel coordinates chart to filter down the colleges by those with a median ACT score higher than a 30, but a smaller student population (< 10,000) and a reasonable admission rate

(30-100%). We see that the list of names has adjusted accordingly, as have the dots displayed on the map. The user clicked on the Polytechnic Institute of New York University in the names list and they were presented with a detail view of the racial composition of the school.

The screenshots above show the 2 potential detail views that can be seen in in the upper left-hand box, racial diversity and financial information.
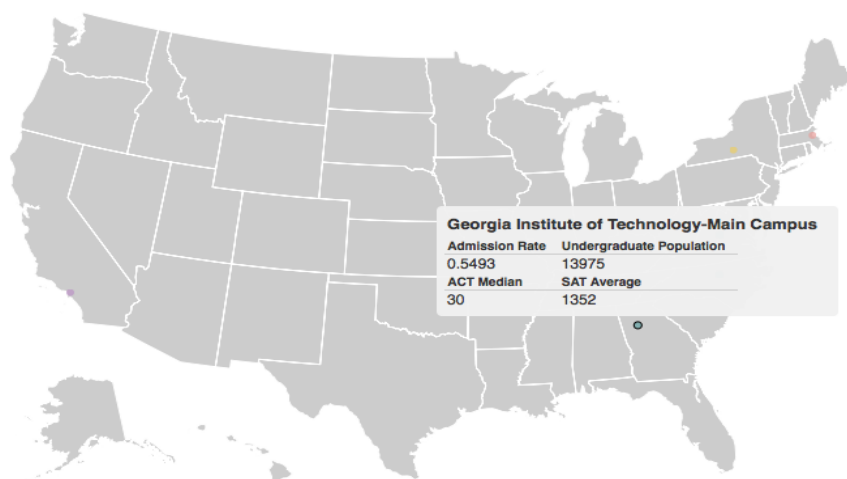
While the parallel coordinates view displays a good overview, the users can get the details of all 4 major attributes by hovering over the school in the map (you can't see it in the screenshot but the mouse is hovering over Georgia Tech on the map, trust us!).

The screenshot above demonstrates the effects of hovering over a school in the names list. The school name is bolded, the school dot becomes larger and is circled in black on the map, and the school line in the parallel-coordinate view becomes thicker and more bright in color, allowing a user to easily learn about the school from all of the views at once.