Jennifer Stark

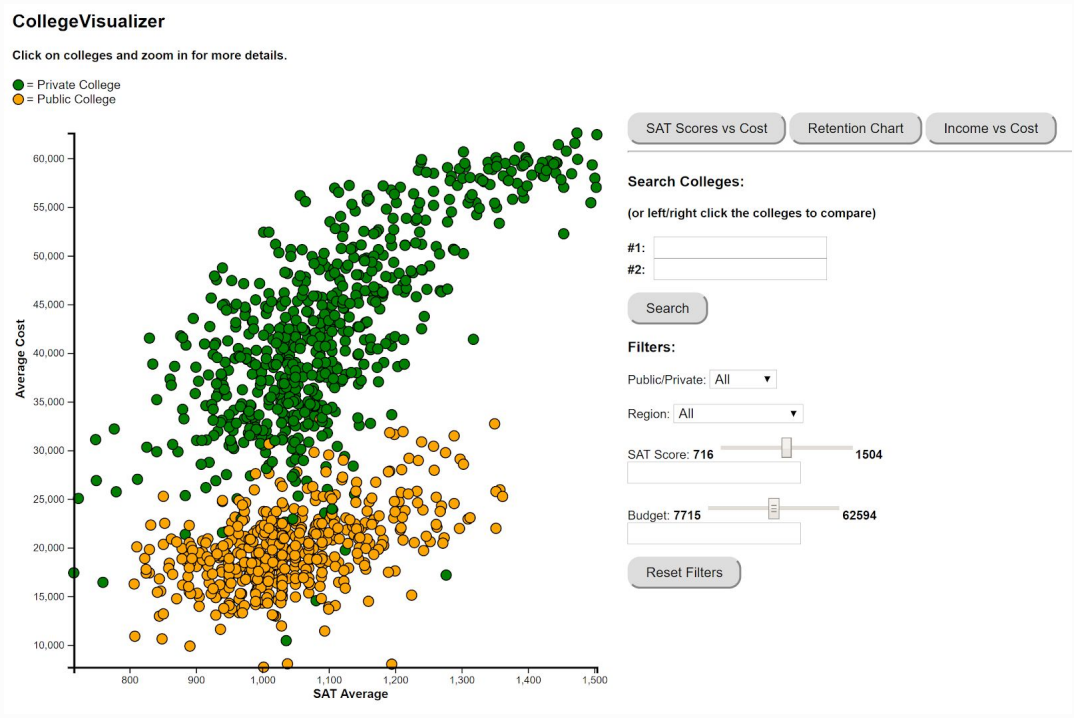
Dong Son Trinh

P5 Report

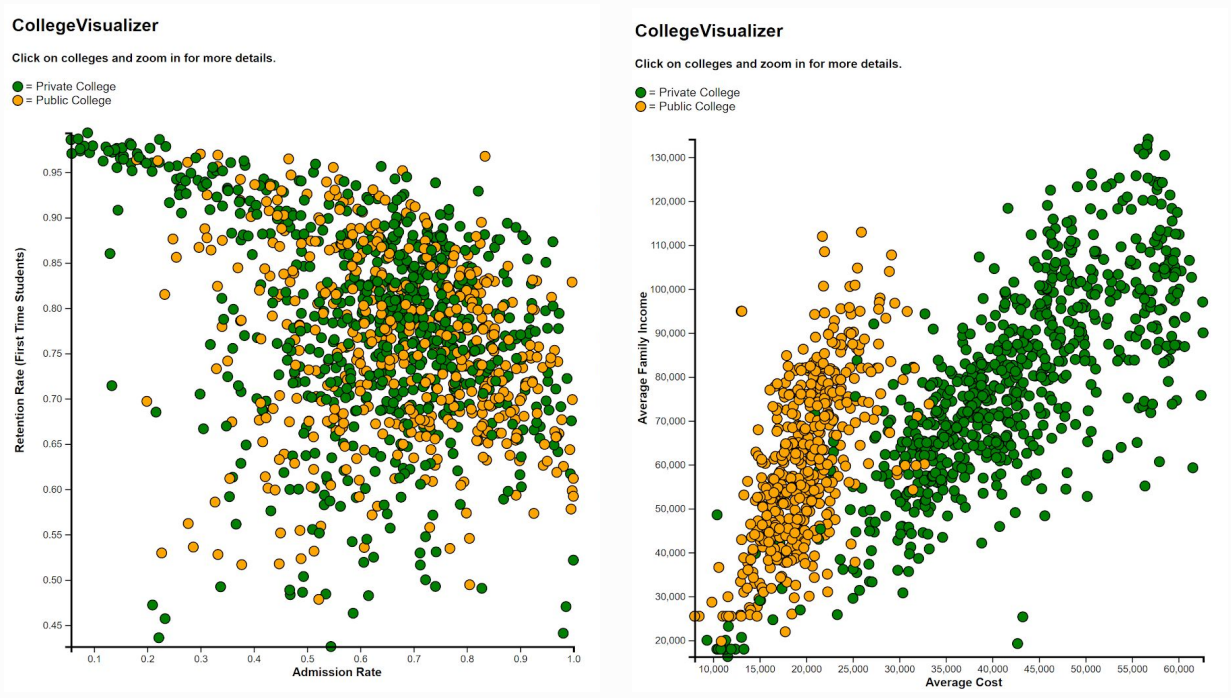
Design

For this assignment, we decided to use the 'colleges' dataset. Our target user is someone searching for colleges to apply to. Using our experience in searching for colleges we came up with some analytical questions such as, 'Which colleges could I attend based on my SAT scores and budget?', 'What colleges are the most expensive/least expensive?', 'Which college has the highest/lowest admission rate?', 'Is there any correlation between retention rate and admission rate?', 'Using my family's income, which colleges are realistically within our price range?', and 'How do these two colleges compare to each other?'.

We were able to narrow down the attributes we wished to focus on in our design because of these questions and the fact that our application decisions were affected the most by a subset of attributes. These attributes were cost, SAT scores, college retention rate, admission rate, whether the college was private/public, and where the college was located. This subset of attributes were the most important to include in our visualizations and we created several different scatterplots to plot these key attribute values. Each visualization embodies a different important aspect of the college search. Users want to see where they can get in with their scores and how much it might cost (plot 1), how easy it will be to apply to/stay in these colleges (plot 2), and if they have no idea how much they should be spending then plot 3 comes into play. We felt that scatterplots were especially useful here so that users could quickly see all the data cases (no aggregation) and really search for individual colleges easily. All plots are shown below.

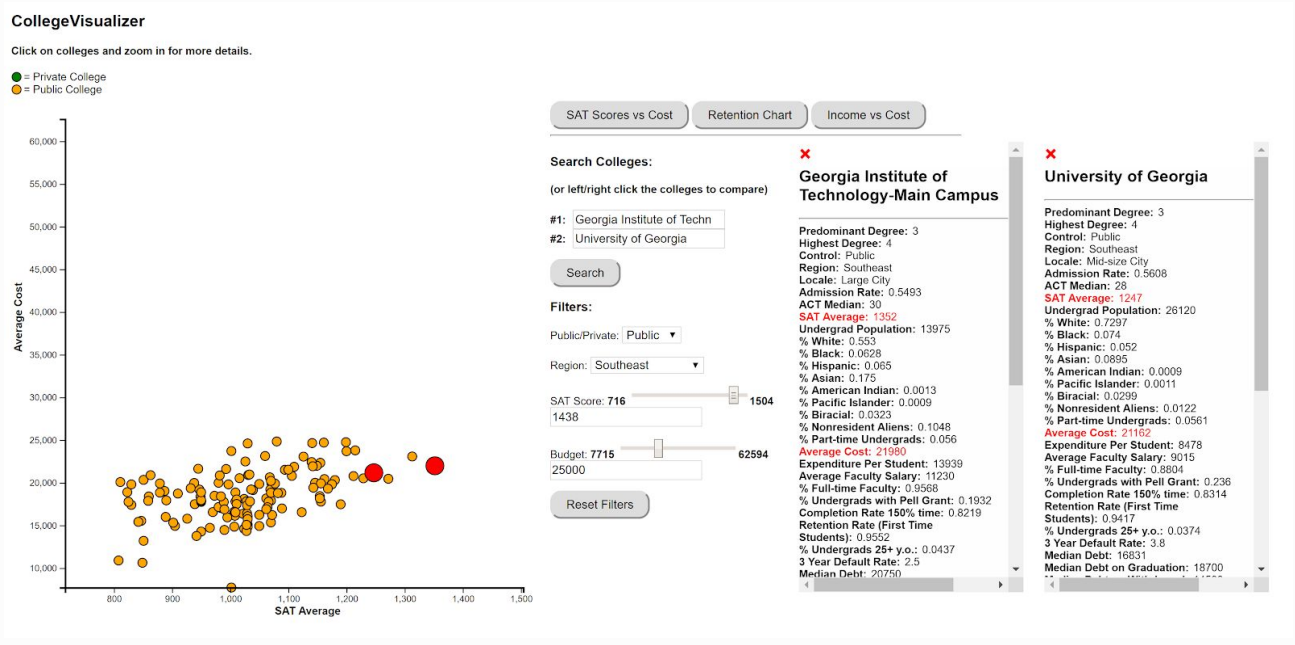


Main page shown with controls on the side with filtering and college search. Other plots can be viewed by clicking on the buttons at the top of the control panel.



Cropped view of other plots available to view.

In addition, we realized that cost, SAT scores, college location, and college type (private/public) were prominent attributes to filter on because here the user can have needs or restrictions that limit where they can consider to apply to before even seeing other data about a college. Lastly, users might want to explore specific colleges so in our design we included search functionality to get details for one or two colleges at a time while highlighting them in the plot.



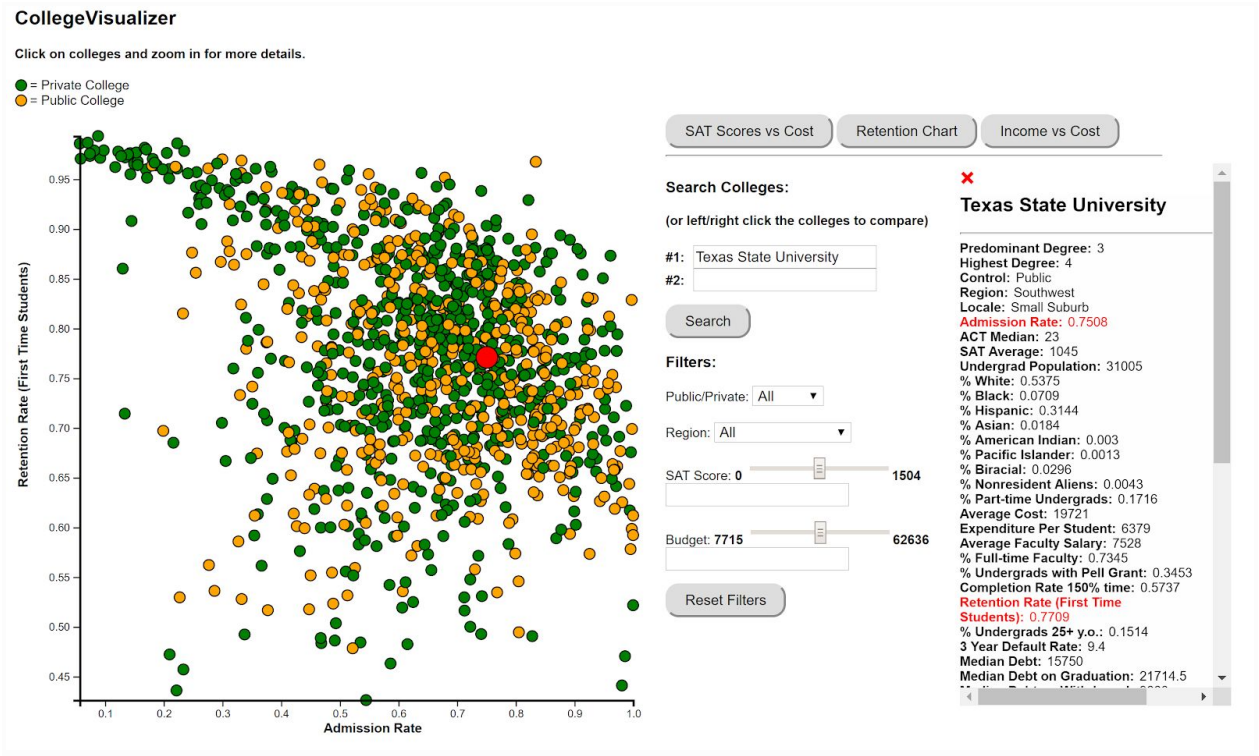
Search functionality and filtering functionality shown. Compare to first graphic.

Analytic Tasks Supported

With our design, users can perform analytic tasks such as retrieve value, filter, find extremum, determine range, characterize distribution, find outliers, cluster, and find correlations between certain attributes. For retrieving attribute values of specific cases, users can click on data points for more details or use the search functionality to find a specific data case. To filter, the user has many options including filtering by public/private schools, by region, by SAT

scores, and by the average cost. For finding extreme attribute values, a user can use the scatterplots to quickly see which data point produces the highest/lowest value for attributes displayed on the chart. However, extreme values cannot be found for every possible attribute because the scatterplots do not include all attributes. Again, users can use the scatterplots to determine the range of attribute values but only for those attributes plotted in the charts. The distribution of certain attributes can clearly be seen in the scatterplots as there is no aggregation of the data cases being done. Outliers along the trend lines in the scatterplots can also be quickly seen by the user. Data cases with similar attribute values will be plotted next to each other and zooming in can help a user find data item clusters. Correlations between attributes (namely the two attributes used in one scatterplot) are shown by simply plotting the data against these two attributes. The scatterplots allow us to support a decent amount of analytic tasks but the attributes a user can analyze in each task has been limited.

In addition, our visualization follows the mantra, ‘Overview first, zoom and filter, then details on demand’. Users are given an overview that shows where colleges fit on a relevant scatterplot and at any time they can zoom and pan on the visualization (shown below) or change options to filter out colleges. If a user wants more details they can hover over points to display a tooltip with the college name and they can see all the details of a specific college by simply clicking on the data point or using the search functionality next to the visualization.



Above: User searches for a college and the college is highlighted.
Below: User zooms and pans on visualization after searching for college.