

Samuel Shapiro & Nick Patel
CS 4460
P5
4/21/19

For this project, we chose to create a visualization using the colleges dataset. This data set includes nominal fields such as a college's name and region, as well as quantitative fields such as a college's average cost, average family income of students, and median ACT score of students. With this dataset, we decided to create two visualizations that would allow users to explore relations between variables and discover correlations through their own exploratory analysis.

The first of our visualizations is the focus of the user interface and features a scatterplot with an alterable x-axis variable. The y-axis of this scatterplot shows a college's average cost, while many of the other quantitative variables are options for the user to apply to the x-axis, such as average family income (shown by default), mean earnings 8 years after entry, and median debt after graduation. When the user chooses a new x-axis from the dropdown menu below the x-axis and presses the button to apply changes, the x-axis will rescale and all of the points on the scatterplot will animate and transition to their new appropriate spot. If a data point does not have a proper value for that variable, the point will shrink to a size of zero and animate into the y-axis. When the user hovers over a point on the scatterplot, the border of the point will darken, and the point will fill with a light blue color. When the user clicks on a point, the border will darken, the point will fill with a light green color, and statistics of the selected college will appear to the right of the scatterplot.

The second visualization uses a linking feature which appears only when the user selects a school. The chosen visual is a pie chart showing the breakdown of the selected school's diversity. The chart considers all 8 diversity factors as well as a derived variable called "Other", which is a representation of the percentage needed for the total breakdown of each school to add up to 100% (many schools did not have diversity statistics adding up to 100%). This was done to ensure a well-balanced and fully complete pie chart for each data case. Each slice is represented by a different color and is labeled on either the left or right side of the chart. There is a sorting feature in place to ensure each variable will always be represented by its corresponding color, and when a new point is clicked, the slices will animate, and the corresponding labels will move to represent the diversity of the newly selected school.

When creating a data visualization, it is always important to keep in mind what the objective of the visual is and the message it is trying to get across. The first step in approaching this problem was to answer the question, “What information do we care about?” Our approach to this was to first implement a scatter plot so that the user has a view of all data points. The y-axis for the scatter plot is always a representation of the average cost of attending a school. When comparing schools, average cost can be a major and sometimes deciding factor of whether or not the student wants to attend that school, so we decided to keep it as a constant part of the visualization. The next objective was to organize the data across an x-axis. To do this and allow the user some freedom, we chose to implement a dropdown box with 8 key factors that can be used to create entirely new scatterplot visuals.

Median Debt on Graduation

ACT Median

SAT Average

Admission Rate

Average Faculty Salary

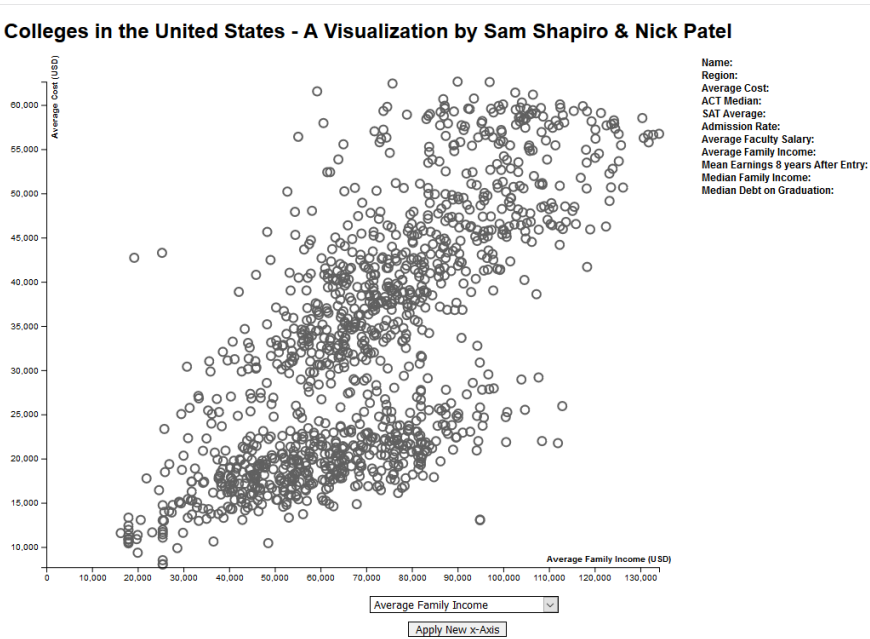
Average Family Income

Mean Earnings 8 years After Entry

Median Family Income

Median Debt on Graduation

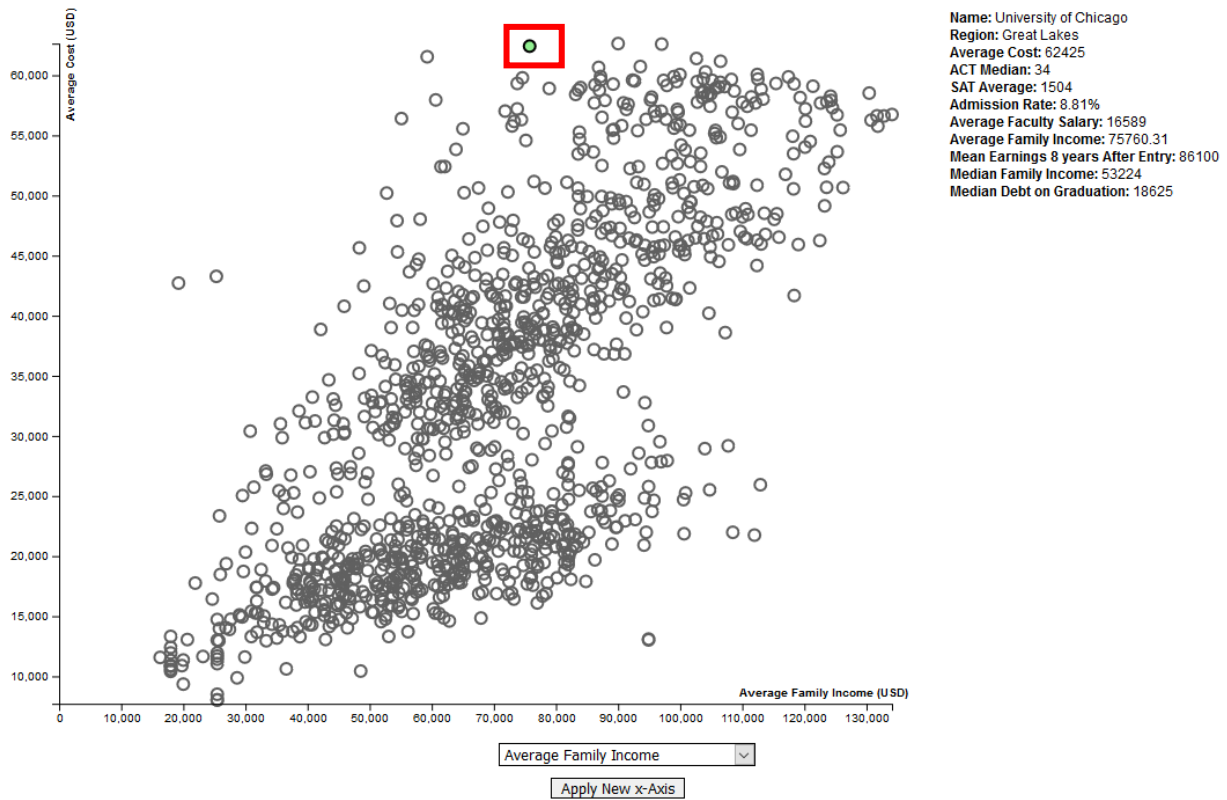
Dropdown box showing the 8 different ways to sort the data across the x-axis.



Default scatterplot without any changes.

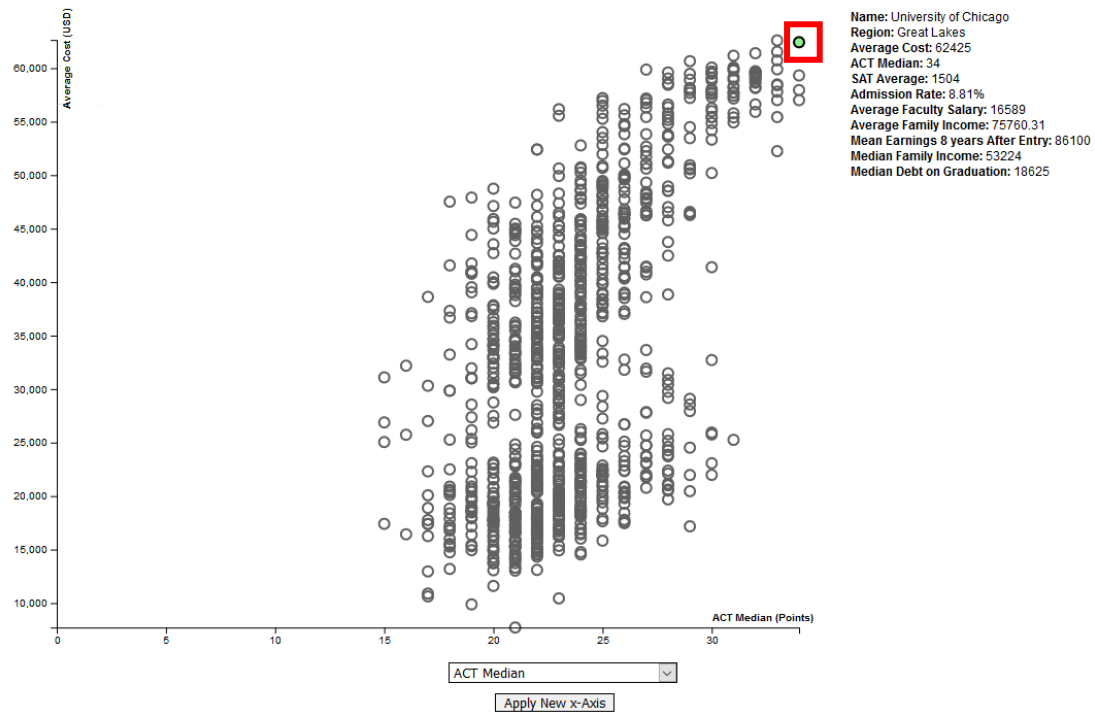
This dropdown was created to allow a dynamic and interactive visual experience which, through exploratory analysis, allows the user to see how colleges relate to each other and follow multiple types of trends.

Colleges in the United States - A Visualization by Sam Shapiro & Nick Patel



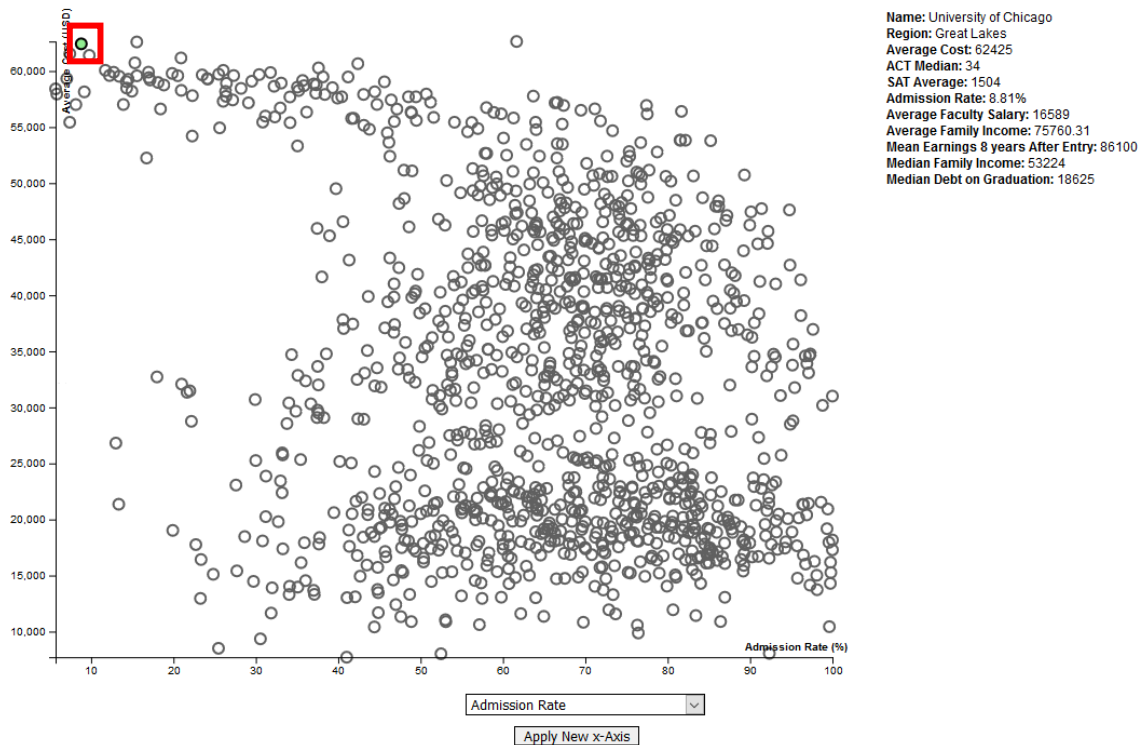
Default scatter plot with selected school node highlighted and showing that schools' information.

Colleges in the United States - A Visualization by Sam Shapiro & Nick Patel



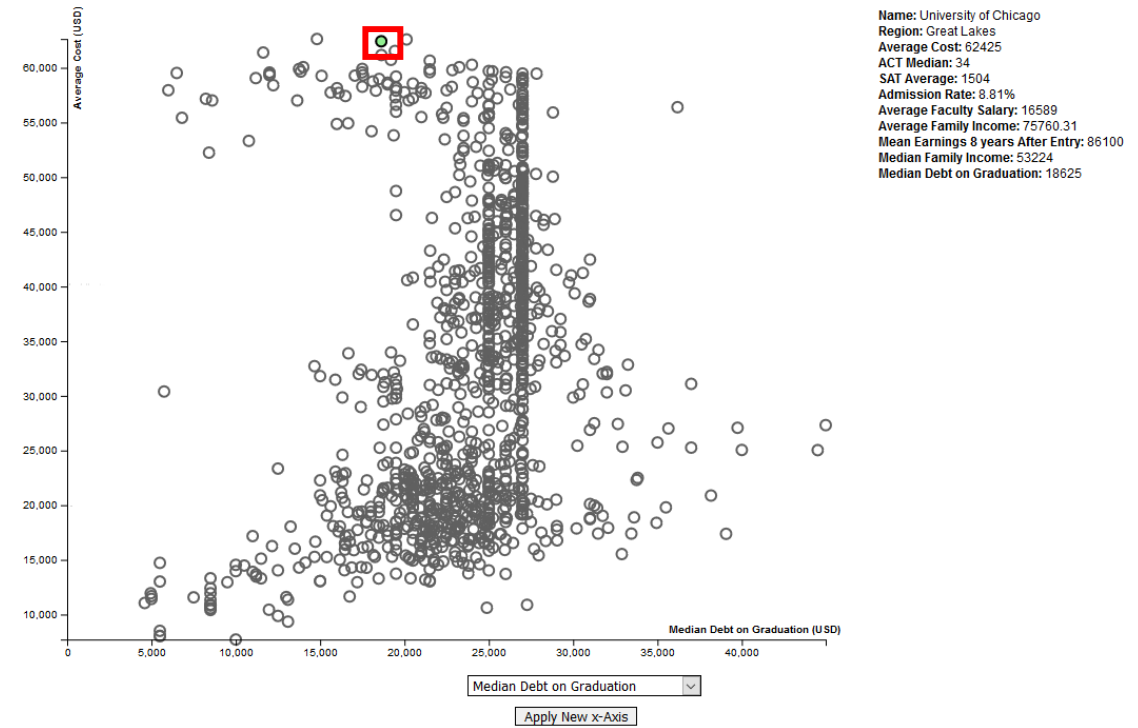
Scatter plot sorted by Median ACT scores with the selected school highlighted.

Colleges in the United States - A Visualization by Sam Shapiro & Nick Patel



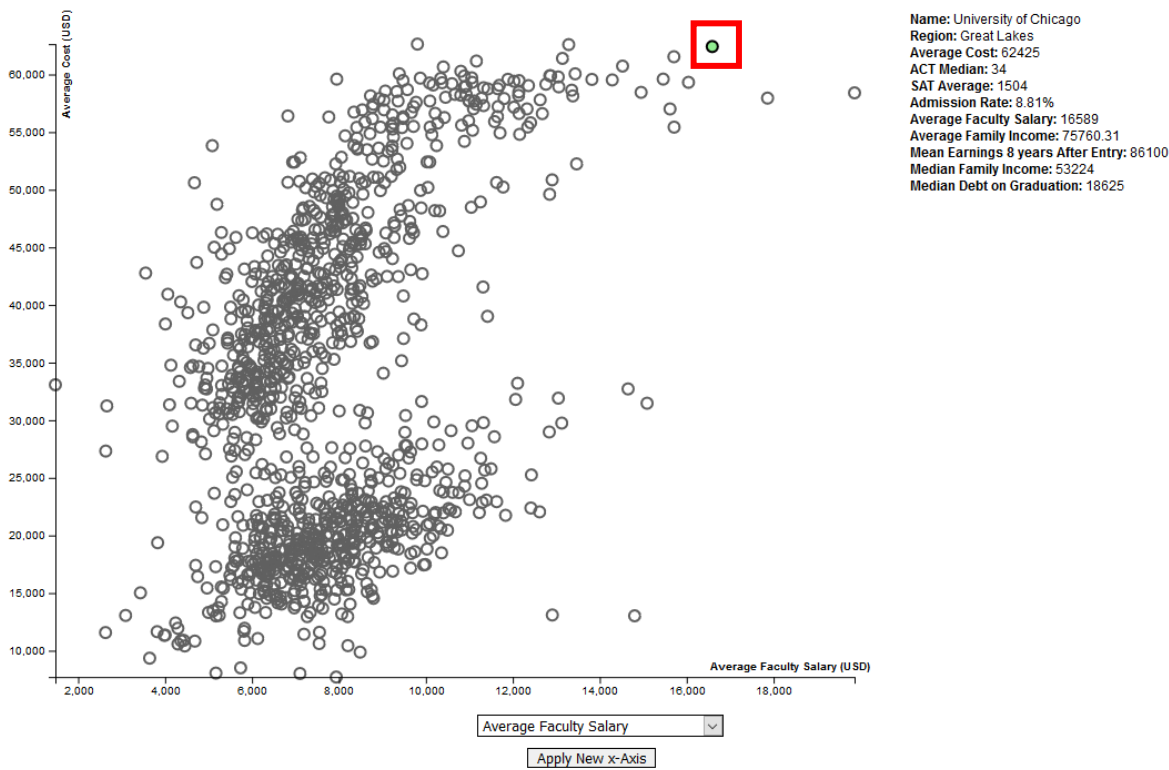
Scatter plot sorted by Admission Rate with the selected school highlighted.

Colleges in the United States - A Visualization by Sam Shapiro & Nick Patel



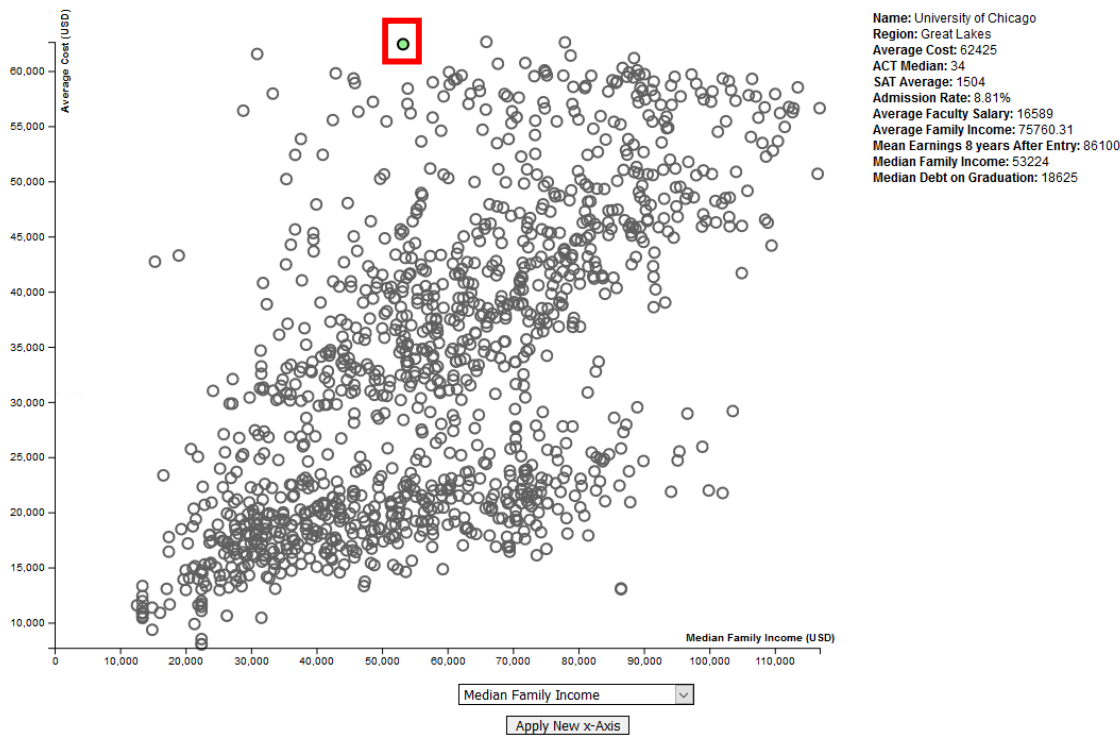
Scatter plot sorted by Median Debt on Graduation with selected school highlighted.

Colleges in the United States - A Visualization by Sam Shapiro & Nick Patel



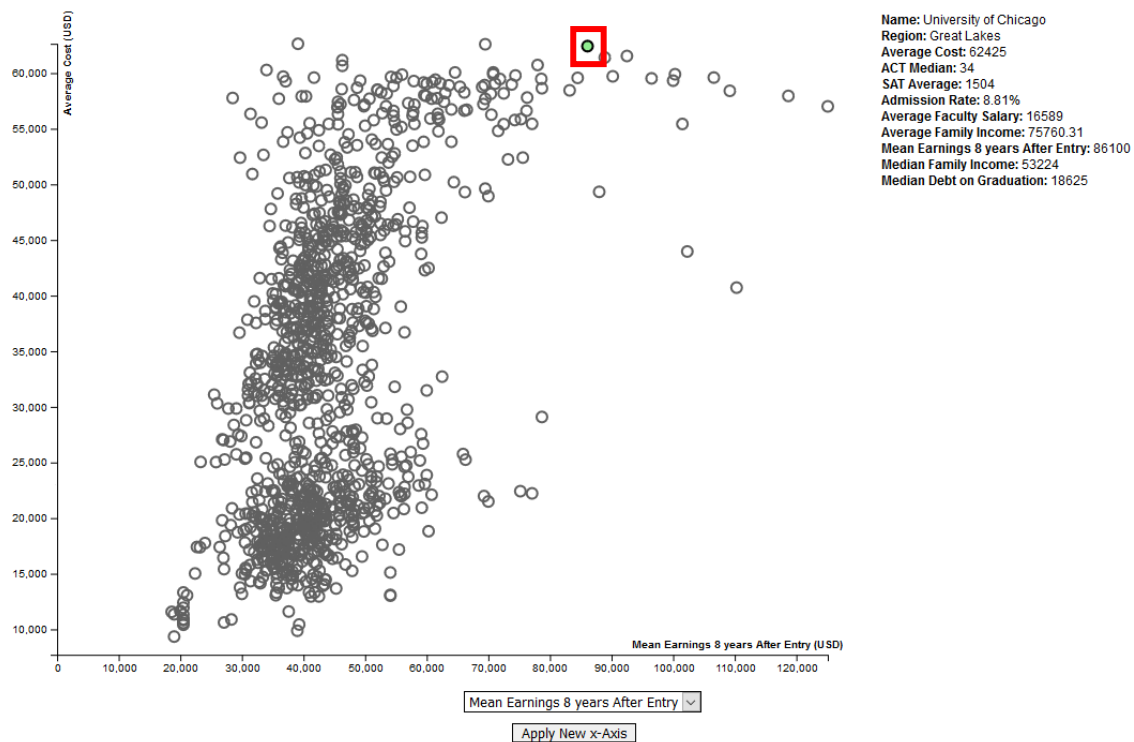
Scatter plot sorted by Average Faculty Salary with the selected school highlighted.

Colleges in the United States - A Visualization by Sam Shapiro & Nick Patel



Scatter plot sorted by Median Family Income with the selected school highlighted.

Colleges in the United States - A Visualization by Sam Shapiro & Nick Patel



Scatter plot sorted by Mean Earnings 8 years After Entry with the selected school highlighted.

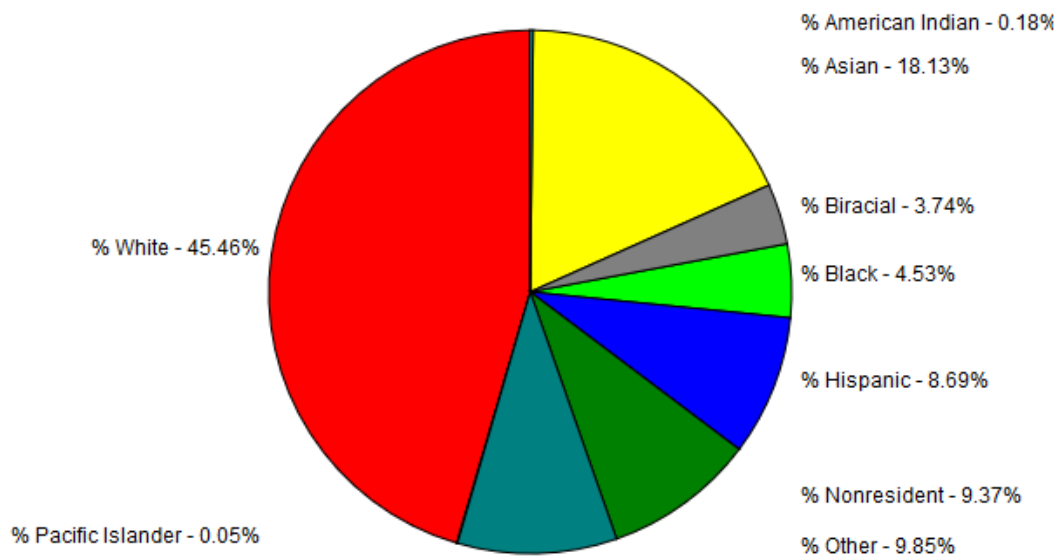
Colleges in the United States - A Visualization by Sam Shapiro & Nick Patel



Scatter plot sorted by SAT Average with the selected school highlighted.

Using the provided images, it is easy to follow the selected school and track the position as well as visually see how that school compares to every other school under different criteria.

After the implementation of the scatterplot, we wanted to explore a single data point more in depth. When a user selects a school by clicking its data point on the scatterplot, a second SVG is generated and appears on the screen as a pie chart showing the selected school's diversity breakdown. Adding this feature allows any user to have greater interactivity with and insight into each node. As users interact with the data, not only will they be able to see how diverse a single school is, but they will be able to draw new conclusions about how diversity factors into the correlations visible on the scatterplots.



Pie chart of the demographics for the selected school, in this case the selected school is the University of Chicago.