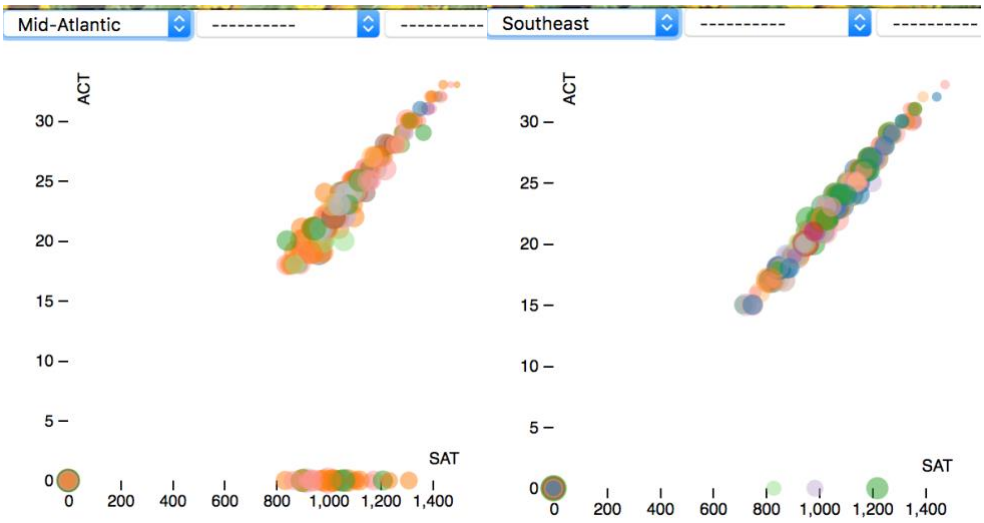
Cayla Vinzons
Yeawon Yoon
CS 4460 P5 Description

Dataset chosen: colleges.csv

Analytic task*:* By selecting a specific school from the 3rd drop down menu, clicking one point in the scatterplot, or clicking one line in the parallel coordinates plot, a user can **retrieve** specific attributes for that case. Using the 1st and 2nd drop down menus, a user can **filter** the data set by locale and/or region. This will filter the data displayed in the scatterplot and parallel coordinates plot. By hovering over the data in the scatterplot and parallel coordinates plot, users can **find extremum** values of ACT, SAT, median family income, average cost, and median earnings. While hovering over clustered and overlapped data points is difficult to execute, it's technically still possible. This hovering interaction can also aid users in **determining range, characterizing distribution, and finding anomalies**. Users can hover to find maximum and minimum values to determine the range and distribution of ACT, SAT, median family income, average cost, and median earnings. Similarly, anomalies can be seen and identified in the scatterplot and parallel coordinates plot. Finally, by design of a scatterplot matrix and parallel coordinates plot, users can **find clusters of data items with similar attributes values** and find **correlations**. Clusters can be seen for the variables represented in the two graphs, and correlations can be seen between ACT and SAT scores, between family earnings and cost, and between cost and earnings after college.

Design overview: Our tool can be used to answer several analytic questions regarding the colleges.csv dataset. Using the first graph, our tool can answer questions regarding the ACT and SAT scores of colleges filtered between locale or region such as "Are there standardized test score differences between the Southeast and Mid-Atlantic?". As shown in figures 1 and 2, Both regions have approximately the same maximum value of test scores, but Southeastern schools have a greater range in scores. The Mid-Atlantic region also contains more schools at the higher end of the ACT/SAT correlation compared to the Southeastern region.



Figures 1, 2. Correlation between SAT and ACT in Mid-Atlantic Colleges and Southeastern Colleges

Using the second graph, our tool can answer questions like “Does the expense of a school make a difference in earnings after college in Mid-Atlantic schools?”. To answer this question, we can generate the parallel coordinates plot for Mid-Atlantic schools. As shown in Figure 3, it appears that the majority of students make between $40-$50k 8 years after graduation, regardless of the amount spent on their education. However, almost all schools with the median earnings above $50k a year had an average cost above $50k. Thus, it appears that in some cases, but not all, a more expensive education was correlated with higher earnings.

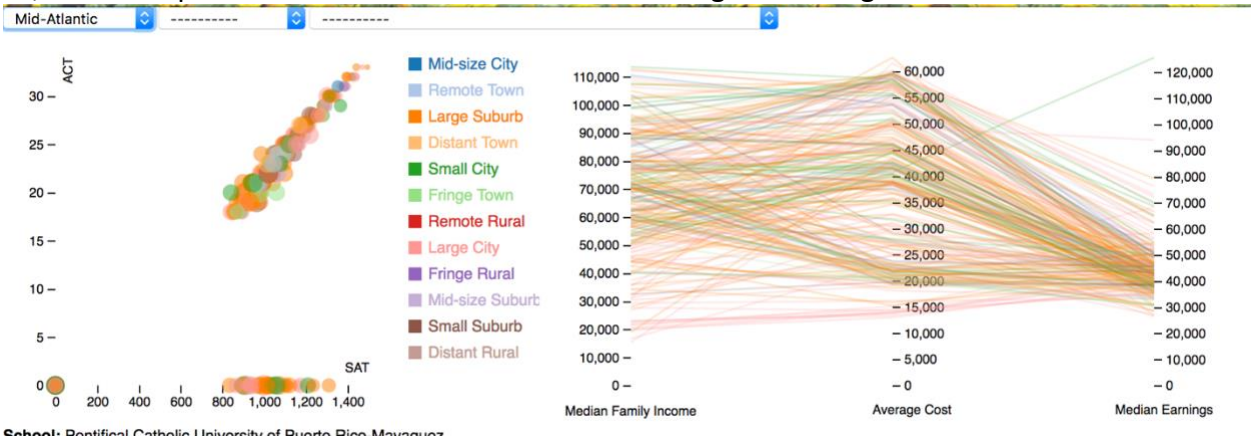
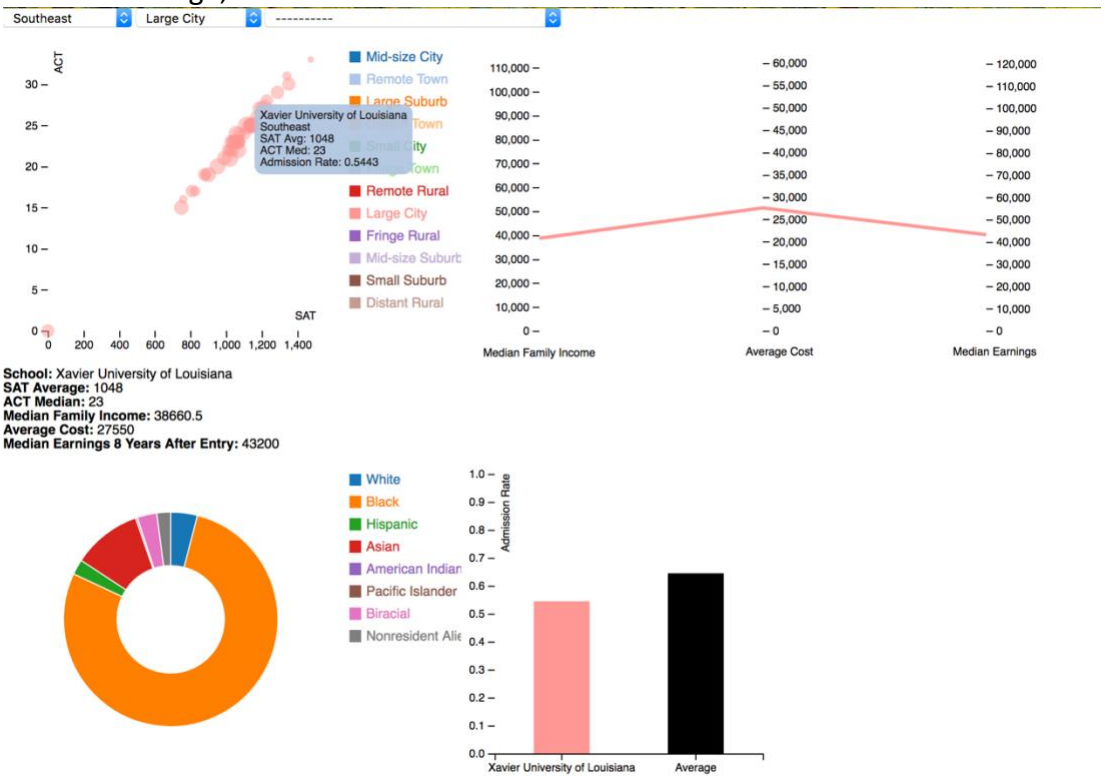


Figure 3. Scatterplot and Parallel coordinates plot for Mid-Atlantic colleges.

For any specific school, our tool can answer the questions such as “What is the race distribution of this school?” or “How does the admission rate of this school compare to the average admission rate?”. Figure 4 answers these question for Xavier University in Louisiana. Xavier University has a mostly Black student body, at over 75% of their student population. The next most popular races are Asian, then White. The admission rate of Xavier is slightly less than the

national average, at about 54%.



Clarifying Details: In general, the first two graphs are meant to serve as an overview while the last two graphs (and the table) are meant to provide details upon clicking on either one point in the scatterplot or one line in the parallel coordinates plot. However, clicking one point on the scatterplot will only show that school's data in the parallel coordinates plot. In this second scenario, the parallel coordinates plot serves as a detail.