

Project 5 Description

Team: Joseph Avocado and Nichole Deeb

Dataset Chosen:

colleges.csv

Supported Analytic Tasks

- Cluster colleges together that have similar values for a specific selected key
 - Rapidly analyze proximity or lack thereof between colleges based on population and acceptance rate
 - Easily change the selected key with a dropdown that instantly begins to regenerate the clusters for the new key
 - View the relationship that acceptance rate and undergraduate population has on the location of colleges in the clusters
- Use brushing to select a rectangular area and receive average information about the cluster
 - Select the entire data set and view average information about the sample as a whole
 - Shows averages of acceptance rate, average cost, ACT and SAT scores, and the selected key if it falls outside of these specifications (i.e. Poverty Rate if and only if it selected as the key)
- Live search filtering for specific colleges to see their location in the clusters (i.e. "Georgia" will highlight all colleges that have "Georgia" in their name)
- Highlight Ivy League colleges and see their locations in the clusters
- View data of a single data point, showing all relevant information as well as the selected key

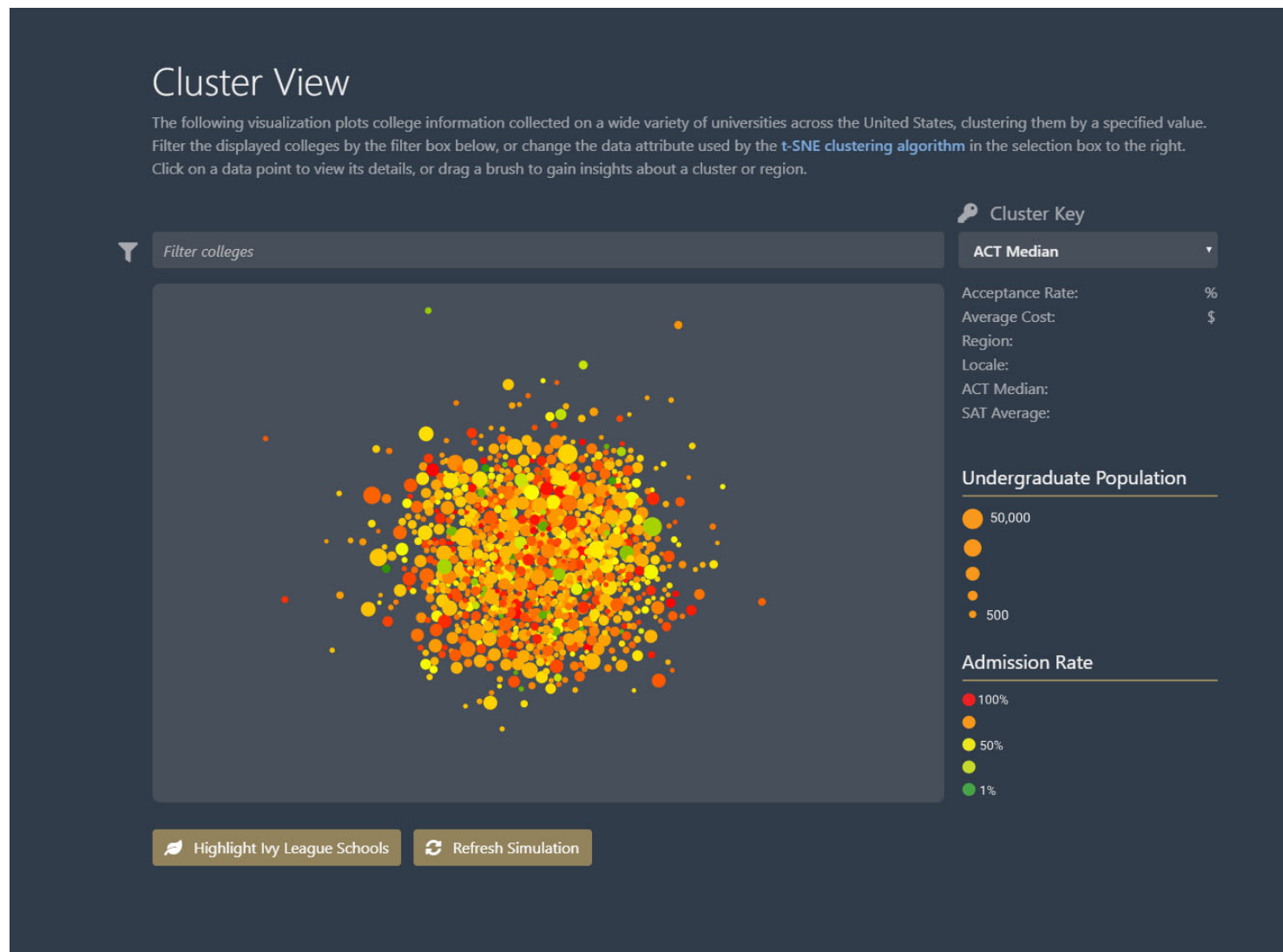
Design Overview

We focused our design mainly around the correlation of colleges' undergraduate population and admission rate (reflecting how "prestigious" the school is) versus a variety of different variables that we allowed the user to select themselves. With this selected variable, we utilized the [t-SNE library](#) to simulate forces and enforce clustering, meaning that the points representing schools with similar values for these variables will tend to be closer together.

Overall, our design is intended to show relationships between colleges on a large-scale graph with over a thousand data points. We allowed individual searches to separate out specific colleges, but our design

was generally more geared to large clusters of colleges. For instance, we wanted to allow the ability to glance at a cluster of ACT scores, and easily see that typically smaller colleges with lower admission rate (more green in color) were in clusters with higher ACT averages, and so forth.

Screenshots of User Interface



Initial view upon starting the cluster

Cluster View

The following visualization plots college information collected on a wide variety of universities across the United States, clustering them by a specified value. Filter the displayed colleges by the filter box below, or change the data attribute used by the **t-SNE clustering algorithm** in the selection box to the right. Click on a data point to view its details, or drag a brush to gain insights about a cluster or region.

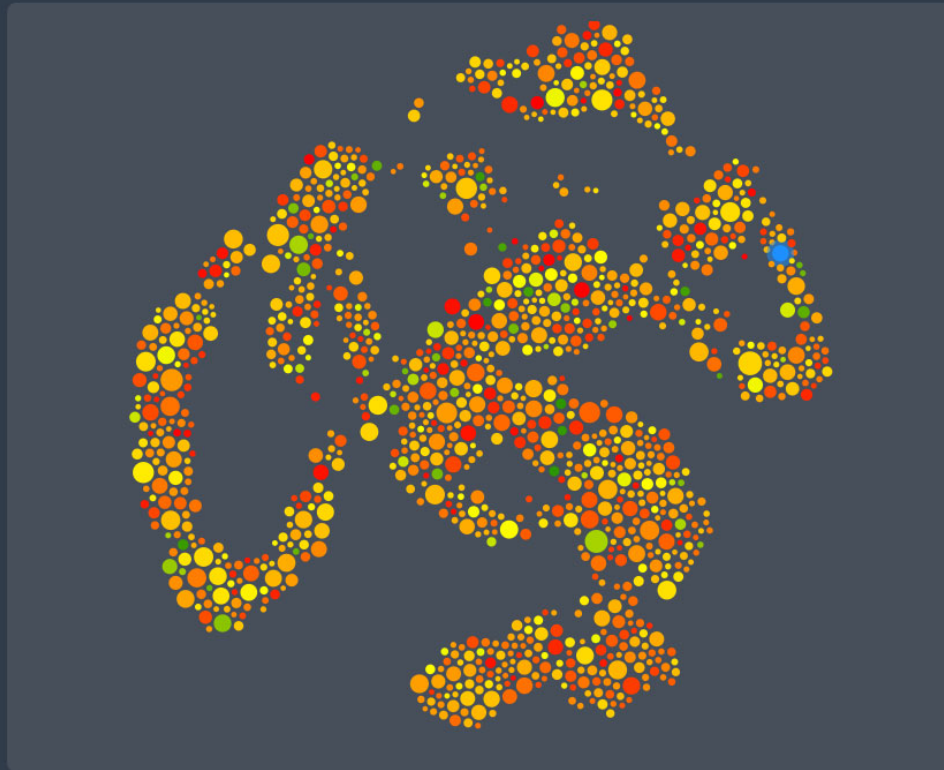


Filter colleges



Cluster Key

Poverty Rate



Kent State University at Kent

Public

Acceptance Rate: 83.02 %

Average Cost: \$ 23,995

Region: Great Lakes

Locale: Large Suburb

ACT Median: 23

SAT Average: 1048

Poverty Rate: 6.920000076

Undergraduate Population

50,000

10,000

5,000

500

Admission Rate

100%

50%

1%



Highlight Ivy League Schools



Refresh Simulation

Selection interface, where a data item turns blue and its attributes are displayed to the right

Cluster View

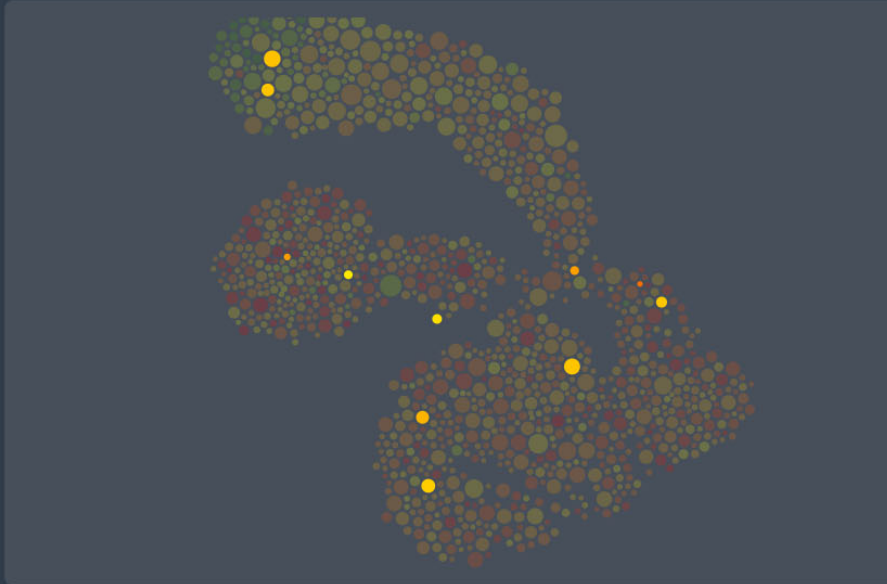
The following visualization plots college information collected on a wide variety of universities across the United States, clustering them by a specified value. Filter the displayed colleges by the filter box below, or change the data attribute used by the **t-SNE clustering algorithm** in the selection box to the right. Click on a data point to view its details, or drag a brush to gain insights about a cluster or region.



Brushing interface shown when the mouse is clicked and dragged in the data view area

Cluster View

The following visualization plots college information collected on a wide variety of universities across the United States, clustering them by a specified value. Filter the displayed colleges by the filter box below, or change the data attribute used by the **t-SNE clustering algorithm** in the selection box to the right. Click on a data point to view its details, or drag a brush to gain insights about a cluster or region.



Highlight Ivy League Schools



Refresh Simulation



Cluster Key

Retention Rate (First Time Stud ▾)

Acceptance Rate: **44.45 %**

Average Cost: **\$ 46,074**

Region:

Locale:

ACT Median: **28**

SAT Average: **1274**

Retention Rate (First Time Students):

Undergraduate Population

50,000

500

Admission Rate

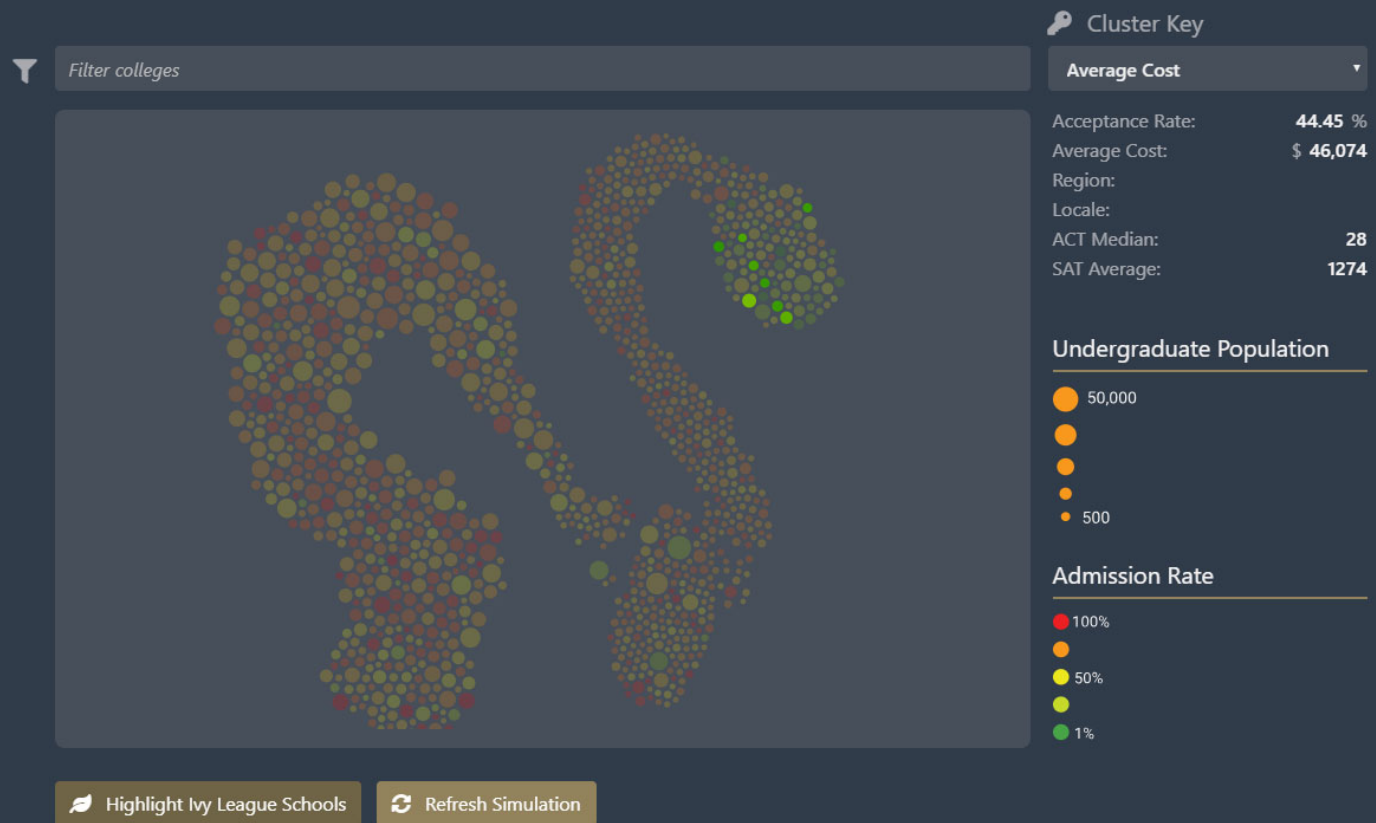
100%

50%

1%

Cluster View

The following visualization plots college information collected on a wide variety of universities across the United States, clustering them by a specified value. Filter the displayed colleges by the filter box below, or change the data attribute used by the **t-SNE clustering algorithm** in the selection box to the right. Click on a data point to view its details, or drag a brush to gain insights about a cluster or region.



Ivy league highlighting interface that shows when the **Highlight Ivy League Schools** button is pressed

Additional Notes/UI Explanations

We wanted the first two characteristics, undergraduate population and admission rate, to always be visible in the visualization, as they were our primary motivating factor. Because of this, we encoded their values in the data points themselves, deciding on the following encodings:

- **Undergraduate population** - represented as the size of each data mark (a visual cue to the real-life idea of university "size")
- **Admission rate** - represented as the color of each data mark, with red being the greatest admissions rate (lowest selectivity) and green being the least admissions rate (greatest selectivity). This was designed as a cue to the idea that more selective colleges are usually ranked better, and are therefore better in general than less selective colleges (utilizing the green/red good/bad visual association)

Inclusion of Ivy League Highlighting

We wanted to use a baseline to compare other colleges against, so we used very notorious and prestigious colleges as a baseline: the Ivy League. The Ivy League has 8 member colleges, all of which are listed below, and are world renown for their academics.

- Brown University
- University of Pennsylvania
- Cornell University
- Harvard University
- Yale University
- Princeton University
- Columbia University
- Dartmouth College