

# U.S. College Data Comparison Tool

CS 171 Project Book  
Jaemin Cheun and Raymond Kim  
TF: Mimi Xu

## Table Of Contents

[Introduction](#)

[Overview and Motivation](#)

[Related Work](#)

[Data](#)

[Tasks](#)

[Users](#)

[Process](#)

[Data Collection](#)

[Map](#)

[Table](#)

[Scatterplot](#)

[Interaction](#)

[Visual Design](#)

[Analysis](#)

[Logistics](#)

[Conclusion](#)

# Introduction

## Overview and Motivation

For our content, we wanted to work with college ranking data because of the great demand for college comparison tools by parents, teachers, guidance counselors, and high school students. An intuitive interface where one could compare college data (e.g., cost, enrollment, location) would be immensely valuable for those who are involved in the college application process.

We began brainstorming the overall structure of our data visualization by looking through some of the previous projects listed under the Hall of Fame tab of [www.cs171.org](http://www.cs171.org). We thought that the data visualizations that incorporated scatterplots and map data were relevant to our topic since common criteria for people searching for colleges include rank, location, cost, and enrollment size. We can aptly visualize these dimensions using a map and a scatterplot to compare the cost and enrollment size for each college. We can then use a scrollable table to visualize rankings.

## Related Work

Among the data visualizations we looked at under the Hall of Fame tab, we thought that the [Streaming the IMDB Top 250 visualization](#) was the most cohesive. We liked how they used a scatterplot with brushing to tie together the table, map, and bar charts in a seamless manner.

## Data

We used data from [Forbes' America's Top Colleges list](#) as our dataset. From this list we extracted the rank, name, cost, and total population.

We then found the coordinates for each school using Google Maps API.

## Tasks

On a high level, we have to create a map, table, and scatterplot that visualizes our college data in a cohesive manner. The user will be able to click on a particular college by clicking on the corresponding circle on the map, row in the table, or dot on the scatterplot. When clicked, it will highlight all three relevant points.

## Users

The target audience of our visualization is students who are beginning their college search process. We want to facilitate their decision process by providing them with useful statistics for each college such as its cost and student population. We hope this visualization can help students to find the right college that fits them best.

## Process

### Data Collection

#### **Ranking, cost, total student population:**

Using javascript, we scraped <http://www.forbes.com/top-colleges/> to retrieve information of top 100 colleges. We have built both a JSON object and a CSV object that contains the information we want to use for our visualization. There were some problems we faced while scraping the data because some of the schools had a comma in their name (e.g. University of California, Berkeley), which made it hard to create the correct CSV object.

#### **Location:**

For the location of each college, we have used Google Maps API and query the name of the college to retrieve the latitude and the longitude. We have included the location data in the JSON object. One problem we have faced was because Sewanee-University of the South contained a dash in its name, Google Maps could not correctly query the location.

### Map

The map visualization represents each college as a circle located at the actual coordinates of the college.

In terms of size of each circle, we made the radius proportionate to the enrollment size of that particular college. That is, the larger the student body at a particular school, the larger the circle will be.

In terms of color, we created a color scale to compare cost across all colleges. That is, our color scale ranges from white to blue, where white is the cheapest schools and solid blue the most expensive schools.

Users can hover over each circle and see a tooltip with further information about the college (e.g., cost, enrollment, name).

### Table

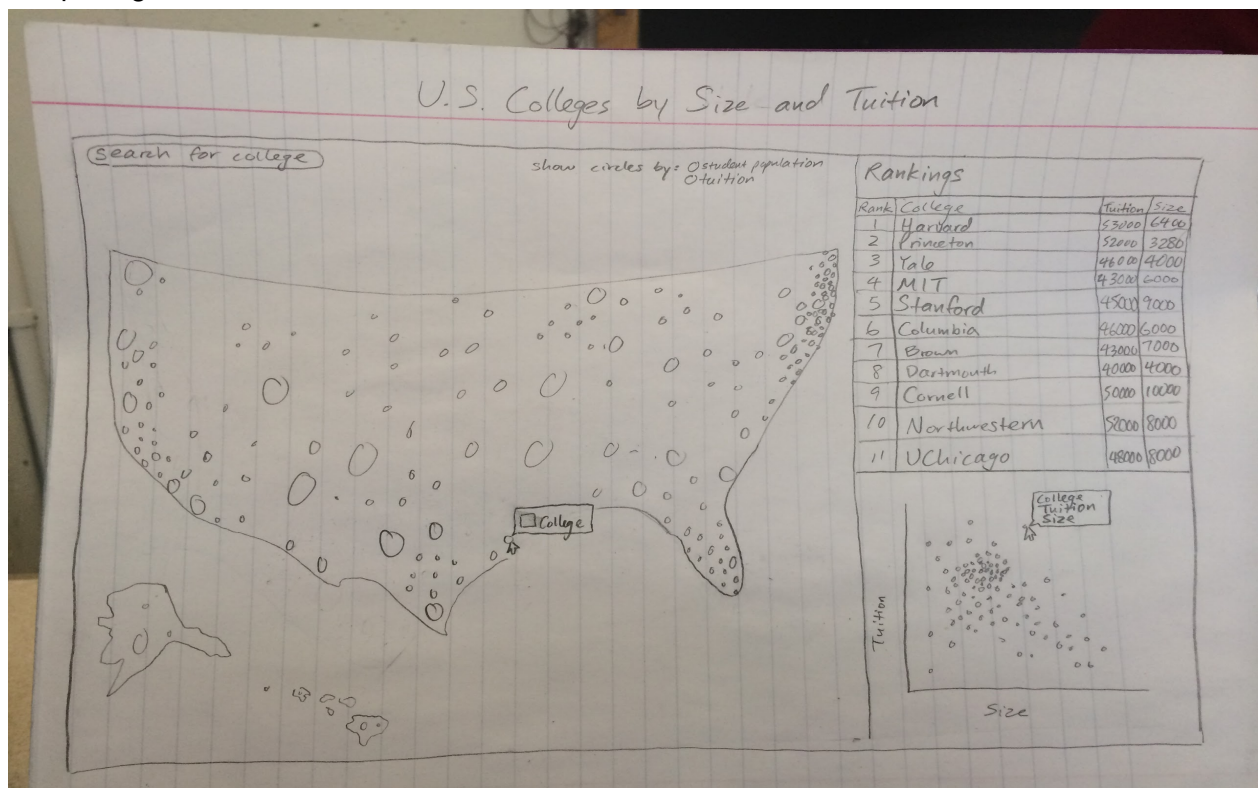
The table initially displays a list of 100 colleges sorted by ranking. It also contains useful statistics about each college such as its cost and student population and allows the user to sort rows by either of them.

## Scatterplot

The scatterplot displays a dot for each college along an x-axis (enrollment) and y-axis (cost). We also shade each dot on the same color scale as the circles on the map to further illustrate the relative cost of each college. Thus the higher up the circle, the more blue it becomes. Again, users can hover over each dot to show a tooltip with further information (e.g., cost, enrollment, name).

## Interaction

We first brainstormed how we can best unify our multiple visualizations into one compelling interface. Below is our sketch of our overall data visualization:



As you can see, a large part of our interaction is based on the user exploring and hovering over relevant points (e.g., circles on the map, rows in the table, dots on the scatterplot). Each of these points can be clicked on for further analysis and insights about the colleges.

Visual Design

Analysis

Logistics

Conclusion