

Education
Calibration
May 4, 2016

CS171 PROCESS BOOK

*by Richa Chaturvedi,
Erik Godard, and
Waverley He*

Education Calibration is a website designed to make the daunting process of picking a college more manageable. We do so by introducing three visualizations that help select, compare, and learn more about different schools.

Past entries

Entry 1

Submitted Project Proposal

Every year, millions of students across the country decide where to apply to college. While there is data about each school readily available, there is no one place where all the factors are listed, measured, and compared together. As college students, we want to make this information more accessible and understandable so that other students have the tools to pick the right schools for themselves.

Education Calibration is a place where all of this information comes together. Using a data set entitled “College Scorecard” from data.gov (<http://catalog.data.gov/dataset/college-scorecard>), we are hoping to create interactive visualizations that make the uncertain and daunting task of choosing a college a bit more clear. For those who already have a prospective school in mind, they will be able to see how their college fits into the larger dataset. Others will be able to decide on potential colleges by filtering the data to adjust to their own needs.

Some of the data we will consider are admission rate, six year graduation rate, SAT/ACT scores, cost of the school/average cost of attendance, race of the student body, number of undergraduates, percentage of Pell students, mean and median earnings, and default rates. These different factors will allow users to see how their priorities are reflected in the characteristics of various schools.

Team Roles

We divided up roles within our team. We decided that Richa would be the team CEO, who is responsible for submitting deadlines, scheduling meetings, and ensuring the overall functioning of the team. We decided that Erik would be the team CTO, who is in charge of the technical aspect of the project. We also decided that Waverly would be the team CDO, who is in charge of the design and

overall appearance of our visualization and submitted materials. Clearly, we will all be performing a variety of roles throughout the course of the project, but this list clearly spells out who is primarily responsible for each area.

We also made several technical decisions at this meeting. We decided that we would use git for source control. Erik has used git extensively, while Waverly and Richa have minimal experience with it. We will continue to check this area to make sure that nobody has any trouble with it. Erik also created an empty Github repository to store the project. We also discussed potentially purchasing a custom domain name on which to host the visualization project in order to make it appear more polished and professional. Finally, we decided on methods of communication and formulated a tentative weekly meeting plan. We established a Facebook group message as well as an iMessage group message in which we will communicate regularly. We also decided to have a 24 hour response policy for all incoming communications. We also decided that we would have a check-in meeting on Google Hangouts once each weekend, with potential group work times to also be added later.

Data Description

The dataset we will use is the college scorecard, obtainable from <http://catalog.data.gov/dataset/college-scorecard>. It is a comprehensive dataset of metrics about US colleges over the last 15 years. There is a comprehensive summary of the dataset in the PDF included with the download of the dataset. However, after inspecting the data, there are several categories that we are particularly interested in, which are summarized below.

- Name - Clearly an important data point, this is how we will refer to the school when presenting it to the user.
- Public/Private Nonprofit/ Private For-Profit - This determines the type of the school, from the given options.
- Admission rate - According to the data description, this is “number of admitted undergraduates divided by the number of undergraduates who applied.”
- SAT/ACT scores - Average scores on standardized tests for students enrolled at the school. Also includes 25th and 75th percentiles.
- Average Cost of Attendance - This is the average cost to attend the school, without factoring in financial aid or scholarships.
- Average Net Prices - This is the average cost to attend the school, after aid and scholarships have been factored in.
- Race of the student body
- Number of Undergraduates - The number of currently enrolled degree seeking undergraduates.
- Share of First Generation Students - Percentage of currently enrolled students who are the first in their family to attend college.
- Percentage of Pell students - The percentage of students receive Pell grants, which are federal grants for those whose families make below a certain amount.
- Cumulative Median Debt - The median debt of those who are in debt to the school
- Mean and Median earnings - Mean and median amount earned by those who were once enrolled but are no longer enrolled in the school.

- Undergraduate student body by race - Racial composition of the current student body.
- Threshold earnings - Percentage of students earning over \$25k.
- Default rates - Percentage of those who owe money to the school who are currently in default.
- 150 Percent Completion Rate - The percentage of student who graduate within 150% of the expected time. For example, for a four year program, this is the number of people who graduate in six years.
- 200 Percent Completion Rate - Similar to above, except for 200% of the time.

Goals and Tasks

Our main goal at Education Calibration is to make our website accessible and helpful to different types of users. The website caters to two different populations: prospective students looking to learn more about different colleges and people with interests in the college system as a whole. We will create different visualizations and features to serve these two groups of people.

For students deciding where to apply to college, there is a wealth of information available about each school but a lack of centralization of that information. Education Calibration hopes to make the college search easier by bringing all of the data to one place and allowing users to search for schools based on what is important to them. Using data from data.gov, Education Calibration will show users schools that reflect their interests. This will be made possible by a comprehensive filtering system that allows users to input their optimal class size, tuition, and type of education. From this, we hope that students will begin to look more closely at schools that are tailored to their interests and needs that they may not have known about before. If a particular school interests them even more, students can click on it to learn more. This is the core of Education Calibration - completing anything else will be a bonus.

The other demographic that we anticipate will use Education Calibration is people who are interested in learning more about the American college system. For these users, we hope to create an interface that allows them to design their own scatterplots to find trends in the data. For example, if a particular user is interested in learning how tuition is linked to the four-year retention rate of a college, they can select these two variables and plot them against one another. This allows for an interactive way for users to examine what interests them more closely.

Education Calibration is a website made by students, so we know how daunting choosing where to apply, or the college system as a whole, can be. By creating this tool, we hope to make a seemingly overwhelming process more manageable, but will need to accomplish some tasks in order to do so.

To implement this visualization project, there are many technical tasks that we will need to accomplish. Below is an overview of what we believe we will need to do.

- **Determine hosting for website** - Up until this point, all of the projects that we have done have been run on our local machine. However, for a project of this scale, we will want it to be accessible to a large number of users. To accomplish this, we will need to put our visualization on a public

website and potentially acquire a domain name to make the location more memorable.

- **Process data** - Currently, the data is located across several files, separated by year. We will need to determine exactly what data we will need for our visualizations and process it appropriately to make it usable for our program. This may involve making choices including what years we wish to feature as well as merging the data into one file for easier analysis. We may also want to remove schools with large amounts of missing and incomplete data. This may allow us to focus on more reputable schools as well as decreases the computational load on our servers.
- **Implement visualizations** - Clearly, this is a critical component of our project. We will need to leverage the data we have obtained and our knowledge of D3 and effective visualization design to design meaningful visualizations for our data. The specifics of our visualizations are covered elsewhere and subject to change.
- **Create additional website content**
 - Website text - We will need text explaining the context of each visualization as well as aims of the project as a whole. While we already have some text about broad goals of the project, we will need to consolidate and refine our current material in order to clarify and strengthen our message.
 - Website visual content - This is a broad category, comprising images, layouts, and colors schemes. Essentially, we will need to make the website hosting the visualization intuitive and aesthetically pleasing to use. We will need to make decisions about potential website templates to use as well color schemes and the flow of visualizations and information. We also may want to consider separating different parts of the website to make it easier to use for people who use it for different purposes.

Testing our project - After designing a preliminary version of our project, we will want to watch real people interact with it and observe their thoughts. We would like to know if they find the visualization easy to use and informative, as well as any suggestions and feedback that they. This will be very helpful to our project. Since we will eventually become well-versed with the data, it will be critical that we obtain the perspective of someone who is not as familiar with it.

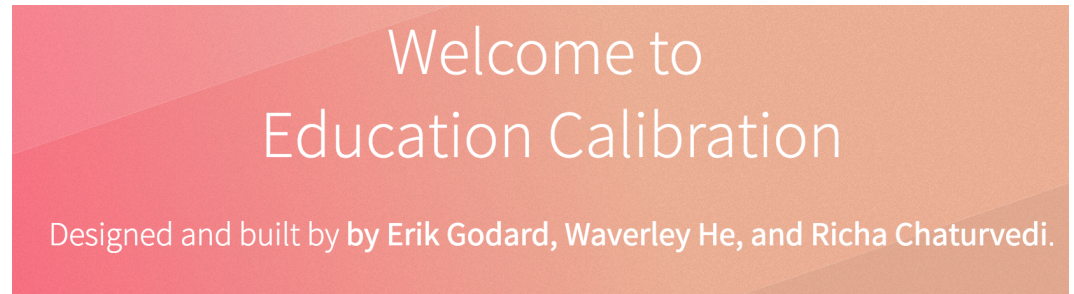
Entry 2

Before implementing our visualizations, we returned to the expert evaluations and peer feedback to think critically about the practicality and usefulness of each visualization. We decided to implement the College Choropleth/College Mapper visualization, the Make Your Own Scatter Plot visualization, and the Segmented Bar Chart visualization. Our current implementations are functional, but can still be improved upon to allow greater user convenience and flexibility.

We decided to combine the choropleth and college mapper because it felt redundant to have them both when instead they could dynamically play off of one another. This is the implementation that will be presented in the final iteration.

Overview and Motivation

In the past couple of years, data.gov compiled a comprehensive dataset about all of the colleges in America. The sheer amount of data is overwhelming. As college students, we understand that the college selection process is overwhelming enough. That is why we created Education Calibration, a place to find and select a college through the use of 3 interactive visualizations.



***There is no place
where all this data has
come together... until
now.***

The College Mapper, School Visualization, and Make Scatterplot pages present the data in a clear, interesting, and engaging way.

Related Work

One of the things that shocked us was the lack of resources on this very topic. That is what inspired us to create Education Calibration. We investigated the visualization “New York City Schools” from the Fame page of the CS171 website, but were actually very confused about how to use it and found it overwhelming. We set out to create an alternative way of comparing schools.

In class, we were inspired by the FIFA World Cup lab to graph different trends. This eventually ended up becoming the Make Scatterplot page.

Questions

How do I find a college that fits my criteria?

We answered this by implementing the College Mapper, which transitions from a choropleth to a node map as different filters are applied. From this map, you can click on a node to learn more about a school.

How do I learn more about a college?

This is answered on the Visualize School page, where data is presented both in tables and in pie charts that show the breakdown of the student body.

How do I compare two colleges?

This is answered on the Visualize School page as well. When users enter a different school, another column of the table appears describing that school. The pie charts become segmented bar charts so that users can easily compare both schools.

How do I learn more about college trends across the board?

We created the Make Scatterplot page so that users could map different variables against one another and color the dots based off of categorical variables.

Data and Exploratory Data Analysis

We found the data entitled “College Scorecard” on data.gov. No cleanup was required but, due to the massive amount of data included, we decided to focus on certain categories of data over others. Initially, we did no basic exploratory visualizations to learn more about the data. Instead, we

learned more and more about our data through different iterations of and commentary on our visualizations.



Design Evolution

The biggest learning process for us over the course of this process was the evolution of the designs of the various visualizations, particularly that of the visualize schools page. Originally, the choropleth map and the node map were two different visualizations. We decided to innovatively combine them when we realized that it would be interesting to see the colors change as different filters were applied. We also made this decision after the design critique in studio. The group we worked with pointed out that a node map could be too crowded based off of the sheer number of schools in the US. From this, we decided that it would be more efficient to combine the two and only show nodes if under 100 schools matched the filters.

Our Make Scatterplot stayed fairly consistent throughout the process. We made the decision to add the option of choosing a categorical variable to change the colors of the dots after realizing that some information, like public v. private schools, is incredibly important and therefore cannot be left out.

The visualization that experienced the greatest amount of change is the School Visualization. Originally, it was a “Create your own Chord Diagram” page. The feedback we received in the studio helped us realize that this was a very confusing way to communicate the data about the student body breakdown. We then decided on a “Make your own Segmented Barchart” page because length is the most easily understood metric. Later, we decided against this because it was more intuitive to have pie charts there instead. Finally, we settled on a combination of the two. If only a single school has been selected, the information is shown in a pie chart. If two are chosen, it is shown in a side by side segmented barchart. This helps students learn about individual schools and compare them in the most optimized way.

Implementation

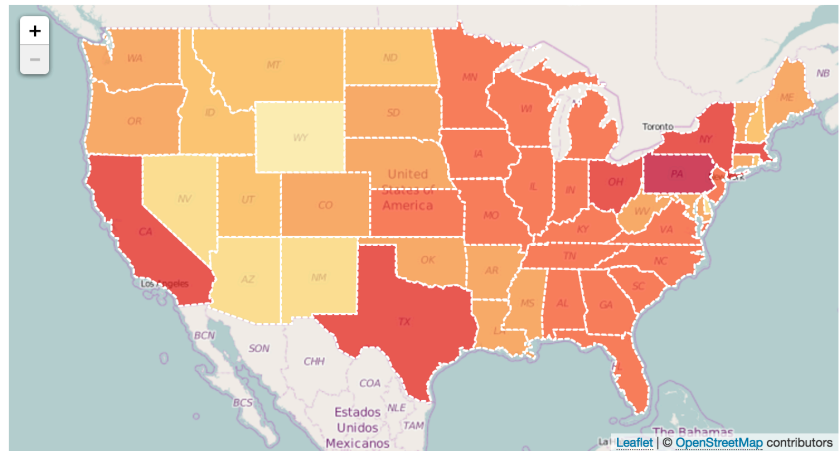
Home page

This page gives very basic information about our website. Clicking “Get Started” takes you to the College Mapper page.

College Mapper

Here, you can see a choropleth that shows the distribution of schools in America. As you slide different filters, the colors will change and eventually nodes will show up. Clicking on a node will show you the school’s name, and clicking on that will take you to the school visualization page.

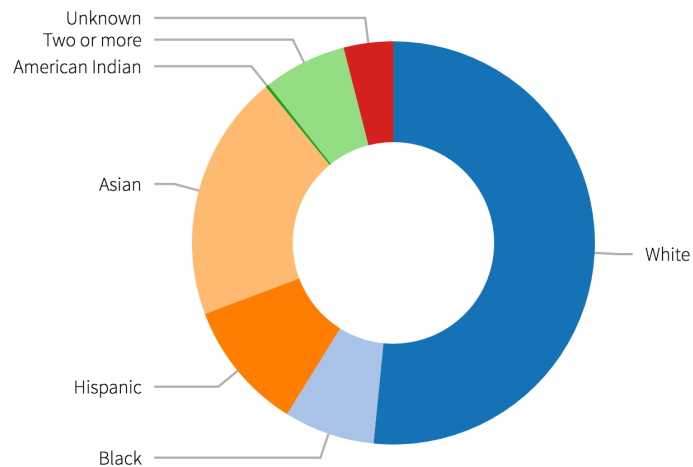
The colors in the map below are shown in proportion of the number of schools fitting your criteria in each state. Once you have filtered out enough results, individual schools will appear.



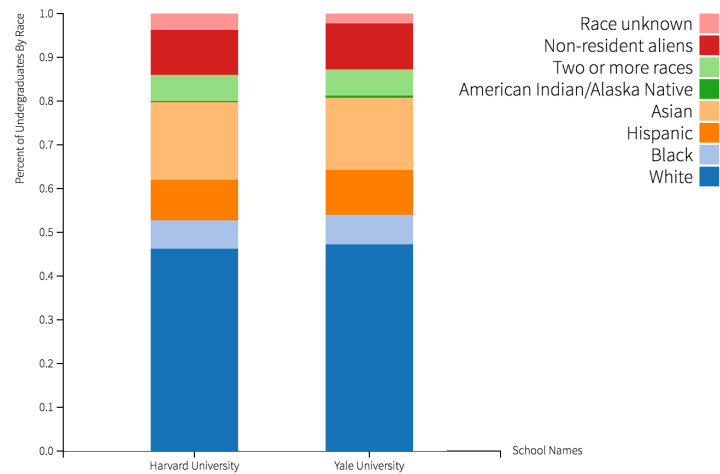
School Visualization

The school visualization page presents data in a table about a certain school and pie charts about the student body breakdown. Adding a school in “Compare to” will add another column in the table about that school and transform the pie chart to segmented bar charts that will help you compare the schools. There is a suggested autofill feature in the “Compare to” box for the ease of the users. Clicking “Make Scatterplot” will take you to that page and compare all of the schools as opposed to just two.

Student Body by Race

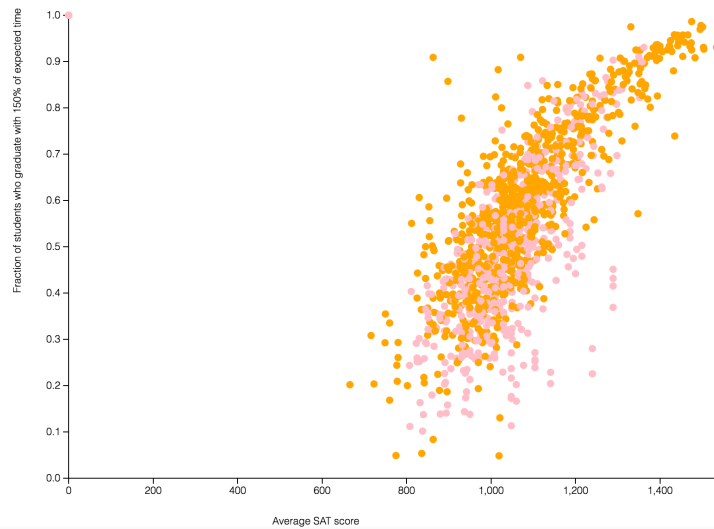


Student Body by Race



Make Scatterplot

Here, you can pick two variables and plot them against each other. You can also color the dots based off of different categorical variables.



Evaluation

We at Education Calibration are proud of the product that we have created. We believe that we have presented the data in a clear and understandable way with each visualization offered on the website. By working with this data, we learned about different trends across colleges – an example of this is the interesting relationship between the Average SAT score and the Percent of Students that Graduate within 150% of the expected time that can be seen on the Make Scatterplot page. Another thing that struck us as interesting was the sheer number of colleges in America. No where else do prospective students have this vast number of options, and that is why this tool needed to be creative. While our visualization works well now, there is potential for greater functionality in the future. We could allow for filtering on the scatterplot page, which would make it less crowded but show less about nation wide trends. We could allow students to compare 3 or 4 schools side by side as well. All of these suggestions would be improvements on an already well-functioning and useful website.

