

## Diversity in the NESCAC Colleges, 2005-2015

NOTE: "unknown" ethnic background could also represent people who identified as multiracial (e.g. in Amherst before 2010).

### //About the Project:

This project was created by Jamie Hand ('16.5) and Emily Sarich ('16) for CS465, Information Visualization.

To deploy, make sure the file 'college\_data.csv' is in the same directory as the main visualization file, 'diversity\_vis.html'. The main visualization file can be opened locally or deployed on a server. It is also available through its github page at <http://jamiechand.com/infovis-collegediversity/>.

### //Data:

Our data is comprised of the Common Data Sets (CDS) of the NESCAC colleges. All colleges are required to fill out the CDS every year to give statistics on admissions and the student body. This data was originally in the form of PDFs published each year by each college, so we converted the sections that we wanted into a .csv file that we could manipulate and visualize with D3. Some colleges only had certain years available online, so we reached out to those colleges and requested the data. We were somewhat successful, but couldn't get all of the data from all of the colleges, so there are some gaps in the data.

### //Audience:

Our visualization is targeted towards a high school student who is investigating the diversity of the student body at the colleges they are applying to. It is designed to show the ethnic breakdown of each NESCAC college so a student can compare the diversity of each school, and show the ethnic breakdown of a college over time so a student can see how diversity has changed over time.

We focused on a student who is looking to find an ethnically diverse student body, so we chose to exclude the students who identify as 'white'; otherwise, unfortunately, the other categories would be significantly smaller and harder to analyze.

### //Description of the Vis:

Our first and main visualization is a stacked bar chart that shows all colleges for a specific year. It shows each ethnic category and the total number of students that make up the incoming first-year class for that year. Each bar can be brought down to the axis so that a student can get a better comparison of how one college compares to the other. We also added a tooltip that shows the actual number of students in each ethnic category as well as the percentage of the first-year class that

the category makes up.

In this visualization, a user can use a drop-down menu to select the year they wish to visualize. We initially considered several options for this selection (check boxes, radio buttons, a text box, etc.) but decided that a drop-down was best for this selection feature because it was the least cluttered and made it clear that a user could only select one year at a time.

Our second visualization is a different type of visualization that chooses to focus on a specific school rather than a specific year. The idea behind this is that a user could look at our first visualization, see something they want to investigate in a specific college, and click on it to show that college's ethnic breakdown over the years to see if it has changed over time.

For example:

A student looking at the data from Trinity College notices that it has the highest number of students who identify as Hispanic compared to other colleges. They are curious as to whether or not Trinity always has a high number of Hispanic students, so they look at Trinity's ethnic breakdown over the years and see that the number of Hispanic students has been steadily increasing since 2005.

//Evaluation

One of our biggest difficulties in this process was the collection of the data. Collecting data from pdfs is, not surprisingly, a pain. We were initially going to include many more colleges (other small liberal arts colleges similar to Middlebury), but we found that the process of collecting data was very time-consuming and not worth the amount of time necessary to gather data from a larger number of schools. Our code is scalable and could handle a larger data set, but because of our limited time frame we chose to reduce the list down to only NESCAC schools.

Another slightly frustrating part of the data was that in 2010, the ethnic breakdown of the CDS changes slightly to be more inclusive. This changes some of the categories, and it was difficult to figure out what to do when the data suddenly changed formats.

One of our biggest successes was the ability to draw our second visualization based on data in our main visualization. Adding the action listener to the tick marks on the x-axis of our main vis makes it possible for a user to visualize a college without having to use a less intuitive drop-down or button selection. It keeps the user in the visualization and keeps the interface cleaner.

If we could change anything about the project, we would have obviously included more data. If we had more time, we would have also added some more detail to our second visualization. The main goal of this visualization is to show general trends and not specific instances, so we thought that this was an acceptable amount of detail.