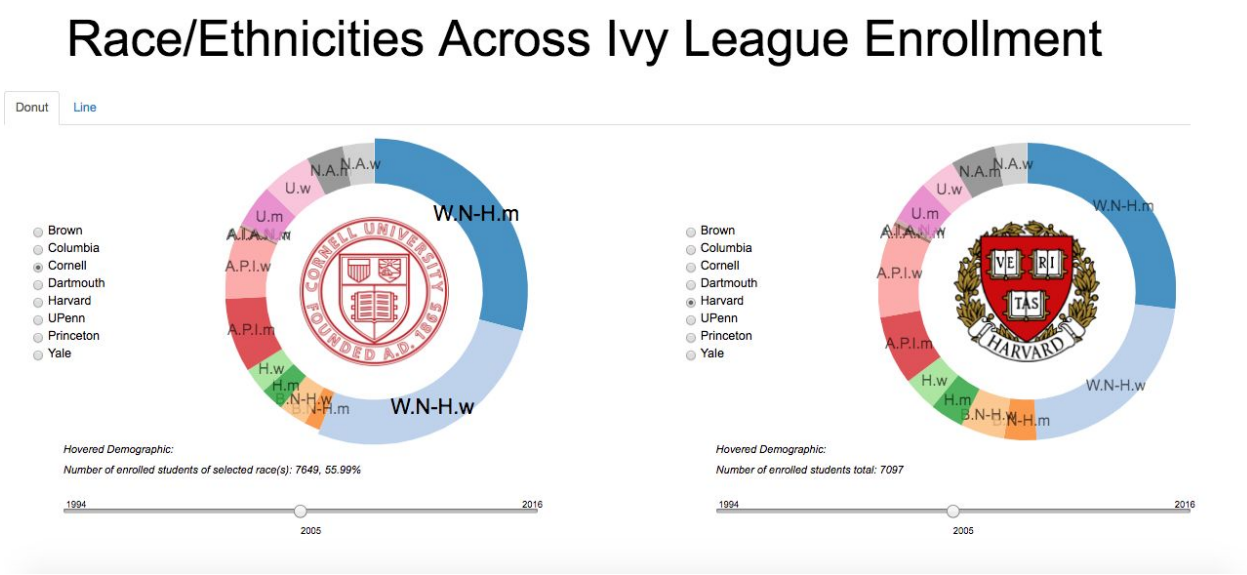
CS 3300: Project Two

William Lin (wrl48) , *Pui Lam Cheng* (pc592) , *Joseph Nechleba* (jdn64)



Description of data

> *A description of the data. Report where you got the data. Describe the variables. If you had to reformat the data or filter it in any way, provide enough details that someone could repeat your results. If you combined multiple datasets, specify how you integrated them. Mention any additional data that you used, such as shape files for maps. Editing is important! You are not required to use every part of the dataset. Selectively choosing a subset can improve usability. Describe any criteria you used for data selection.*

Data retrieval (https://nces.ed.gov/ipeds/datacenter/CDS.aspx)

The data was retrieved from the Institute of Education Sciences (IES), subsidiary National Center for Education Statistics (NCES), within the Integrated Postsecondary Education Data System (IPEDS), using the Compare Institutions tool (link above). The tool allows us to select data by institution and variables of interest. In this case, we chose to focus on data from the Ivy League schools, with school IDs {217156, 190150, 190415, 182670, 166027, 186131, 215062, 130794}. The variables of interest were total full-time undergraduates from Fall Enrollment, divided by race/ethnicity and male/female, for each year.

Data processing

We were able to obtain data for all 8 Ivy League schools with the data of interest from 1994 to 2016. However, the data labels were difficult to work with, more for human readability than data processing. So we split out the year and sex of each row as its own column, and shortened the labels. Additionally, the criteria for reporting, was changed in 2008, and again in 2010. In 2008 and 2009, the data obtained included new, old, and derived values, with some schools not reporting derived or new values. So we updated one set of value, updating it with new values for schools that reported an update, and removed the other two sets.

In 2010, a new category was added, another category was split into two, and the label names updated. The label names were not crucial at this stage, with label processing occurring during mapping from data to visual elements, and so didn't affect the data parsing--however, the change in categories was concerning, as it would affect the color scheme and we worried that it would cause confusion to have categories change color without changing content. To address this, we added placeholder rows in the data where the new data would be, with values of 0. We also sorted the rows for race/ethnicity in each year so that they would be consistent in the visualization.

Data types and additional files

For the donut chart visualization, the primary form of data storage was as a javascript object, with keys corresponding to a year, and values corresponding to an array of objects. Each object represents a race/ethnicity of a gender of a specific year, and contains each Ivy League school's corresponding value. For the line graph visualization, an additional calculated data, diversity index, was stored using a simplified object, with keys corresponding to a school, and values containing an array of diversity indexes, in order of year. An additional file used was a css style sheet in creating the visualization tabs at the top.

Mapping from data to visual elements

> *Description of the mapping from data to visual elements. Describe the scales you used, such as position, color, or shape. Mention any transformations you performed, such as log scales.*

Donut charts

For every donut chart, the number of a certain race/ethnicity and gender that is present in the Fall Enrollment class for a particular school and year is represented by a different arc section. Every ethnicity is scaled down to a portion of a donut chart and transformed around the center of the graphic. The visualizations won't reflect schools with larger overall enrollments than other

Ivies. Every race is assigned a color according to the d3 color scheme “schemeCategory20” which assigns men and women of one race the same color family, varying slightly in shading.

Line graph

For the line graph, the percentage of non-white students enrolled by year is mapped to a d3 line for each Ivy League. We used a linear scale for both the year and the diversity index as the data is fairly evenly distributed, and a simple color list (using official school colors) where there is a 1:1 mapping between the school and the color. The nodes on the lines represent the data points, scaled to the position denoted by the year and diversity index. The tool-tip that shows on hover over a node has the exact value of the percentage of non-white students enrolled.

The story

The story. What does your visualization tell us? What was surprising about it?

How do demographics for fall enrollment compare across the Ivy League in the last twenty years? How have the demographics for any given Ivy League school changed in the same time frame? Does Cornell live up to its recent heralded claims of “most diverse class yet”? These are the questions we answered with our visualizations.

Donut charts

First, the donut charts. In this interactive you can see the breakdown of demographic data for two Ivy Leagues side by side for two respective years of your choosing. In other words you can select Cornell:1994, visualize its demographics on the first donut chart, and then select Princeton:2005, and visualize its demographics on the second chart. This lets you directly compare how two ivy league’s (or even the same ivy league but at different years) races/ethnicities in enrollment were distributed, and whether they had similar trends or marked differences (in terms of percentages). On top of that, you can also click on any section(s) of either donut chart, and for the given school it will tell you how many students of those selected demographics were enrolled at that university during the selected year to further give you insight into the exact numbers behind the visualization.

With the donut charts, we discovered that in fact all Ivies have very similar trends, and follow more or less the same percentages of distribution, with white men and women having by far the largest representation, typically followed by asian men and women, hispanic men and women, and black/african men and women. Also, there is a noticeable drop in white enrollment when comparing any of the two schools from 1994 to 2016.

Line graph

In our second visualization, we added the ability to compare diversity for all the Ivies at once across the entire year span from 1994 to 2016. The line graph displays what we defined as a diversity index, the percentage of non-white students enrolled, over time by year. We thought this would be a good measure of diversity as Ivy League institutions are historically known as being dominated by a white demographic.

For each of the Ivies, you see a respective increase in diversity during the twenty year span; all of the schools bottleneck to 50%-60% in 2011, with some schools even dipping, and then fluctuate in about same range in the years following, with Columbia on top with 63.81% and Dartmouth on the bottom with 50.61% as of 2016. What's also interesting are the differences in diversity increase. For example, Harvard's percentage of non-white students only increased by about five percent, whereas Princeton's nearly increased by twenty-five percent, even though both finish in 2016 with a percentage in the upper fifties. Cornell had a significant increase as well, changing from 38.14% in 1994 to 60.74% in 2016, second only to Columbia University (which has consistently been the most diverse Ivy for the last ten years), meriting its "greater diversity" claims.

Conclusion

In summary, yes, every Ivy League has become more diverse–especially Cornell–but varying in increase between institutions. That said, in terms of individual demographic percentages, each still retains a majority of white men and women.