Justina Chen (jjc382)
Ayah Dweik (asd95)

A.

The data set displayed in our static visualization is the employment statistics of computer science students from Cornell University and University of California, Berkeley. The data set surveyed seniors studying computer science and graduating in 2013, 2014, and 2015. Our data set includes their future employer, the headquarters of their future employer, and the sector they are working in.

UC Berkeley data was found online from the published results of the Career Center's annual Career Destination Survey. Similarly, the Cornell data was published in Post Graduate Reports on the Cornell Engineering career services site. On the Cornell maps, each data point is a Cornell University graduate from the class of 2013, 2014, or 2015. Similarly, the each data point is a University California, Berkeley graduate from the class of 2013, 2014, or 2015 on the Berkeley maps.

Data from both schools included students who chose to pursue further education. However, we chose to exclude these data points, focusing instead on students who chose to go into industry. Furthermore, the reported data included the company that the student was working for and the location they were moving to, but we had to individually tag each data point with the sector of the employer, and the latitude and longitude coordinates of the employer headquarters. Tagging headquarter locations was especially important for the detailed maps of the San Francisco Bay Area and the New York Metro Area. We had three datasets, one for each year, for each school. We concatenated these three datasets into one using Excel. We used a JSON file with shape files of the US states, for the US and San Francisco Bay Area projections. We also used a JSON file of the NYC boroughs for the Manhattan-centered projection.

B.

Using d3 Geo Projections, we mapped the latitude and longitude of employer headquarters to a location on a map projection. The number of students working for one employer was scaled to the length of the bar representing the employer. Additionally, the sector of an employer was mapped to the color of the bar.

We wanted to map the saturation of students in US regions on a choropleth map. The mapping of the percent of students employed in a particular region to the saturation and value of color is only relative. In order to preserve the best visibility of the student employment bars, we chose four arbitrary colors of different value within one hue to represent the four most distinct regions; Berkeley graduates in the Pacific, Cornell graduates in the Pacific and Northeast, and graduates from both schools in all other regions (which were at max at about 5% saturation). These arbitrary differences sufficiently show the relative differences between these four most distinct saturation levels.

C.

Our visualization depicts, on a map, of the United States, where each variable is employed after school. The shading on the map of the United States represents the number of

students who work in that region of the US. Darker shading indicates more students employed in that region. The height of the bar represents the number of people working in that specific office. Our visualization demonstrates that majority of students working in technology end up in the Bay Area and Seattle, while the majority of the students working in financial services work in Manhattan or in the NY Metro Area. It was fairly surprising how few students moved to the Midwest and Southwest. However, after considering the intelligence and strengths in academia, from both universities, it makes sense that students working technology and financial services work in the Bay Area and Manhattan.