Ashley He (ach238), Leandro Dorta Duque (ld443)

College Navigator

CS3300 P2 Write-Up

Our Data

We found and selected values to use in our dataset by compiling three Kaggle datasets (https://www.kaggle.com/samsonqian/college-admissions and https://www.kaggle.com/wsj/college-salaries, file names: “college-admissions-complete”, “salaries-by-college-type”, “salaries-by-region”). The first dataset is comprised of detailed data on certain aspects college admission statistics from all colleges across the 48 contiguous United States in 2017. It delineated any particular college’s total enrollment, school type, etc. The second dataset contained salary statistics post-graduation from different colleges based on school type, as classified by the Wall Street Journal. The final dataset contained the region classification for each school, again as classified by the Wall Street Journal. From the extensive list of categories in the original two datasets, we chose to look at seven categories we felt would be most relevant for a user determining which college they should attend, listed below. We compiled the seven categories for each school into the file “College-stats.” We also needed the longitudes and latitudes of each college to help plot points that would represent the schools on the U.S. map visualization, which we found from the National Center for Education Statistics (Under Institutional Characteristics, then Directory Information https://nces.ed.gov/ipeds/datacenter/DataFiles.aspx, file name: “Position Data”). Finally, we randomly selected a subset of 124 colleges and compiled the data for each of the seven categories and the college’s position into one final dataset, swapping out schools as necessary to ensure there was a variety of schools spread across the map. To get the logos for each school to appear in the tooltips, we manually entered an additional column with a link containing the image reference found online.

As the dataset encompasses a large number of variables, we have selectively chosen a subset of variables to take into consideration while creating our visualizations and revealing our “story.”

- ***Total Enrollment:*** For a given school, the total enrollment of undergraduate and graduate students combined
- ***Tuition and Fees***: For a given school, the total tuition and fees for attending the school, excluding cost of living
- ***Starting Median Salary:*** For a given school, the median starting salary after graduation for all majors and degrees
- ***% Admitted (total):*** For a given school, the total percentage of admitted students for a given academic year

- ***Type of School:*** For a given school, the category it falls under, as classified by the Wall Street Journal (Liberal Arts, Party, Engineering, State, or Ivy League)
- ***Geographic Region:*** For a given school, the geographic it falls in, as classified by the Wall Street Journal (California, Midwestern, Northeastern, Southern, or Western)
- ***Private or Public:*** For a given school, whether it is designated as a public or private school

To create the map of the U.S., we used the us.json file available in the INFO 3300 class repository.

Design Rationale

- ***U.S. Map***
  - ***Marks:*** Circles representing colleges on the map
  - ***Channels:***
    - Circle colors
      1. "Blue" = private school
      2. "Yellow" = public school
    - Location of the circles
    - Circle opacity
      1. 0.8 = unfiltered, college (circle) matches given criteria
      2. 0.1 = filtered, college (circle) does not match given criteria

For our visualization of different college statistics, we used a map of the U.S. with circles representing each of the colleges within our dataset. The circles were colored based on whether they were classified as a public or private school, and their positions were determined through scaling their longitude and latitude coordinate values to our map. Having all the points plotted on the same map allows a user to quickly get an idea of what region or state a college might be, and could help in their search for an ideal school. We chose to leave out Hawaii and Alaska, as there was limited or no data in our selected categories for schools in those two states.

- ***Tooltips/Increasing Circle Size***
  - We chose to also add tooltips to our U.S. map visualization to make the map itself a little more interactive. A user can hover their mouse over a college, represented by a circle, and the circle radius would double to more clearly indicate the selected school. The tooltip then pops up to the right of the selected school with more information, such as the school logo, name, type, whether it is public or private, region, admittance, and tuition. For schools that were already filtered out (opacity of 0.1), the tooltips were disabled.
- ***Histogram Sliders***
  - To adjust specifications that involved a range of values (total enrollment, tuition and fees, starting median salary, % admitted), we chose to use a set of histogram sliders. A user can click and drag a range of values and schools with statistics

outside of the selected range would be filtered out. When a user hovers over the selected area, the cursor changes to four arrows pointed in different directions, indicating that the selected area can be moved. The cursor also changes to a double sided arrow on edges of the selected area, indicating that the selected area could be shrunk or expanded. Even so, the sliders weren't the most intuitive to use so we added in a brief description of how to interact with them.

- ***Button Filters***
  - For categories that were not numerical (private/public, type of school, geographic region), we chose to include buttons to help filter those values. Since our dataset did not include multiple categories for each school, e.g. a school is not both a party and state school, we allowed a user to filter schools by only one criteria for each category. We did this by disabling the remaining buttons of a given category if one criteria was already selected from that category. When a button was clicked, we changed its color (for type and region) or its opacity (for private/public) to indicate which of the criteria were currently selected. For the private and public buttons, we colored them to match the circles, acting as a legend for our map. Finally, we had a "Clear All" button in red, to indicate a difference between the rest of the buttons, that cleared all the selected button filters along with the histogram slider selections.

Story Time

The story. What does your visualization tell us? What was surprising about it? What insights do you want to convey to the viewer of your visualization?

Searching for a college that is the "perfect fit" is often a time-consuming and stressful process for juniors and seniors in high school and students searching to transfer to another institution. There are so many different possibilities for schools that the available options can seem overwhelming and one may not even know where or how to begin their college search. In our visualization of college statistics across the United States, the student can begin to narrow down their choices based on a few different criteria that he/she considers are more important. The dataset used in the development of the website includes the most popular criteria used in the process of college searching according to a study conducted by the Wall Street Journal. The student can combine multiple of these different categories to make the search easier until he/she ends up with few options that can check out their websites. Categories included in the website from the datasets utilized are the type of school, the region, % of Admitted students (out of the total number of applicants), tuition and fees, whether the school is public or private, starting median salary, etc. In this way, this student can take in consideration his priorities and find good engineering schools in California or the best parties in the Northeastern region, whichever is his/her interest. This website has been designed to accommodate all types of users that might

find the dataset utilized helpful, not only highschoolers looking for universities but also PhD students looking for options for graduate school. From our visualization and through playing around with some of the filters, we found that schools in the northeast, mostly Ivy League schools, had the highest median starting salaries. Some things we found were surprising were that engineering schools did not always have the highest starting median salary post graduation. This website's intention is to provide the user with a very interactive visualization with the possibility of applying a variety of filters to the used data in order to ease the search for upper-level institutions; it also displays several numerical statistics related to median salaries, tuition and % of admittance that helps the user analyze different trends and also allows him/her to interact with the graphs to again filter the colleges represented by the selected criteria.

Team Contributions

- **Ashley He:** Initial research on college statistics visualization, sketches for the college statistics visualization, finding, filtering, and compiling the four datasets into the final one used in the project (3 hrs), coded the U.S. map, histogram sliders, and button filters, added tooltips and ability to hover over circles (7 hrs), wrote the Our Data, Design Rationale, and Story Time sections of the write-up (1 hr).
- **Leandro Dorta Duque:** Manipulation of the dataset of colleges, representation of the different colleges on the map based on latitude and longitude values, design of tooltip to display the different data for each college, creation of a dataset of school logos to be incorporated in the tooltips, sizing effect of college representation while hovering to denote the selection of the college, and Story Time section of the write-up.

In general, we worked on the development of this project for approximately four days. One of the biggest challenges was the modification that we had to make to the dataset utilized and the different data that we had to add such as the links to the different colleges' logos. In terms of visualization, one of the parts that we spent more time on was the design of the tooltip and the functionality for the filter buttons. Once we had those working properly, we were practically done.