

Information Visualization Final Paper

Kaitlyn Liu and Hanqi Deng

New York University

Overview

Our project's target users are students who are applying for top U.S. colleges with the intention of paying back their expensive tuition by landing a more lucrative job after graduating from college. Data of the median salaries of the college's graduate students are showcased as a sign of the economic benefits of attending that college.

Data

<https://www.kaggle.com/wsj/college-salaries?select=salaries-by-college-type.csv>

We obtained our dataset from Kaggle and the URL link is above. In this Kaggle webpage, there are three CSV files and we picked out the last one for our project after due consideration. The dataset type is a table in a CSV file form. It has 8 attributes and 320 keys. In our data preprocessing procedure, we first removed columns that are of no use to our project and the rows with not applicable data. Then we harnessed Python to conduct our data preprocessing by writing simple lines of file handling code that found the first 30 intersections of top U.S. colleges and the colleges in the CSV file sorted descending. Finally, we also used the Command-F shortcut key to remove measurement units and commas in numerals.

Goals and Tasks

Our intended tasks are to make a geo-map data visualization and a stacked bar chart data visualization which are interactively linked. In the geo-map data visualization, the region column decides the location on the U.S. map, the college name column decides the location on the region, and the hover on each college showcases the payback salaries. In the stacked bar chart data visualization, for each row, the first column becomes the x-axis, and the other columns' values from left to right are stacked up in a bar form. The scenario is that you are a high school senior ready to apply for college. You are elated because you will be the first generation in your family to attend college, but in the meantime, you are also worried about how long it will take you to pay off your student loans. You want to see how the starting salaries of graduates vary among colleges and their corresponding regions. When you open our webpage, you will see a

map of the U.S. classified into 4 regions. Each dot on the map represents a university. One-color hue represents public schools and the other represents private ones. By clicking on the checkbox at the top corner, you can filter out the colleges that you're disinterested in. By clicking on the dots, you may see a tooltip with the name, rank, and median starting salary of the college, and it links to a bar on the second view. You may then refer to the second view to go through details of mid-career salaries of each percentile. You can even sort the stacked bar chart at your preference.

Visualization

Our visualization consists of a geo map and a stacked bar chart.

The geo map and its changed views are demonstrated in Figure 1. The information is mainly encoded by the spatial locality for the location of the universities, color hue for college type, and color luminance for median wages of a specific state. Users can interact with this view using the checkboxes on top of the view to hide the universities of the type they are not interested in. Also, if users hover the circles representing different universities, a tooltip displaying the basic information of the university, including its name, city, and national rank, is popped up.

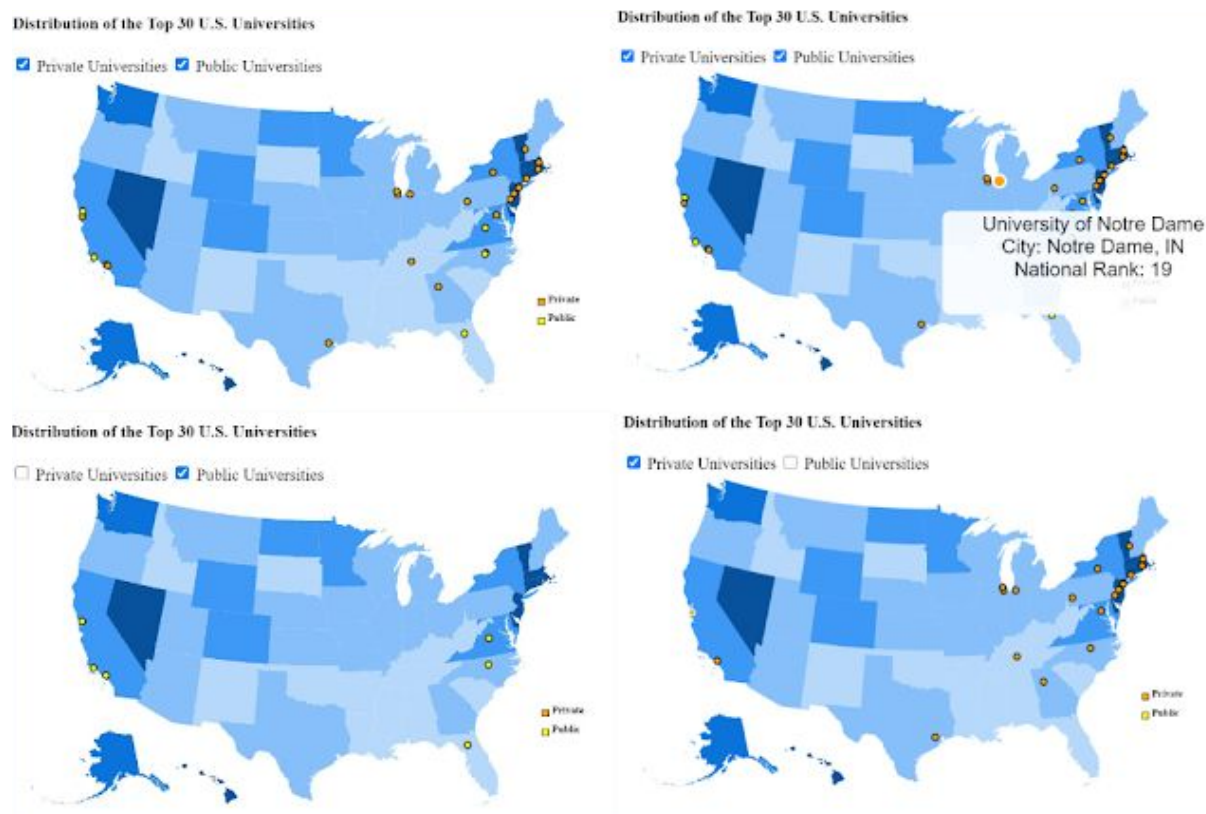


Figure 1: The geo map and its view changes

The stacked bar chart and its change views are shown in Figure 2. The information in this view is mainly encoded by vertical position for the number of graduate salaries and color luminance for the number of salaries within certain percentiles. To interact with this view, users can click on the color legend bar on the top right to hide the percentile bars they are not interested in. By clicking on the bar again, the effect of this previous clicking is canceled out. Besides, users can hover on any bar in the chart to get the exact length of each bar from the popped up tooltip.

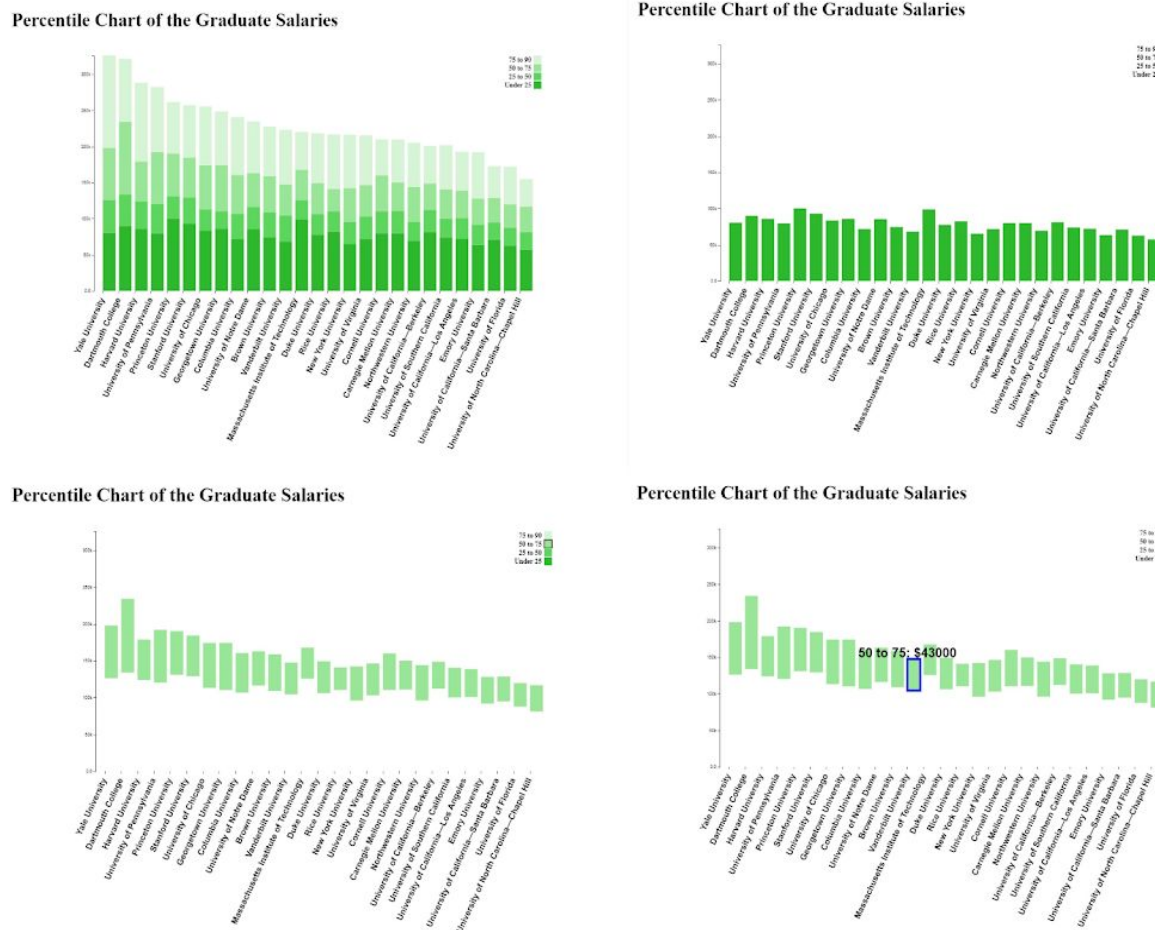


Figure 2: The stacked bar chart and its view changes

The interactions between views are shown in Figure 3. When hovering a circle on the geo map, the stacked bar corresponding to this university is also highlighted in the second view, so that a user can quickly navigate the salary information based on the location of a university. Similarly,

if users hover on any bar in the second view, the circle representing the university that bar belongs to is highlighted, so that a user gets to know the location of a university quickly when looking at the bar chart.

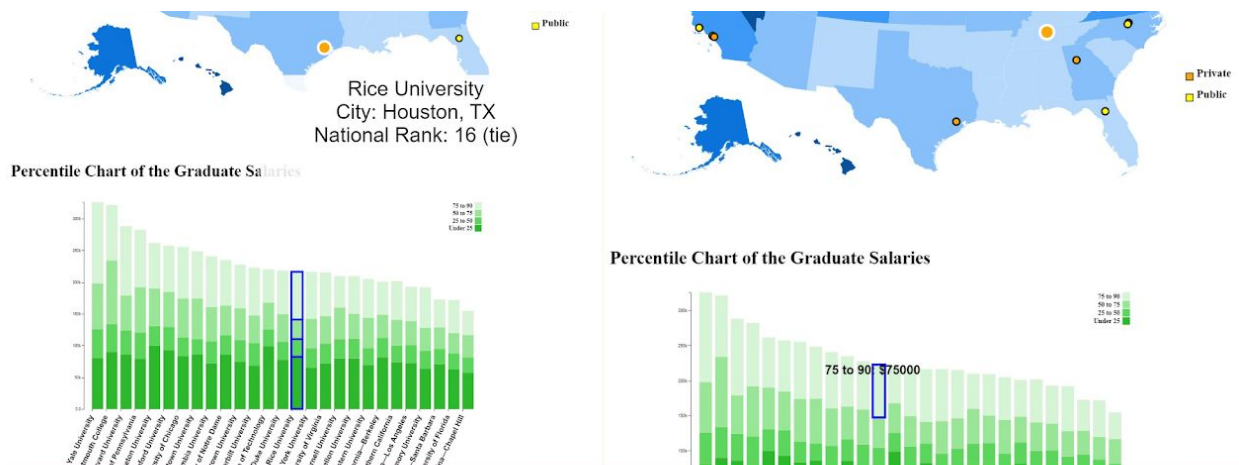


Figure 3: Interaction between views

Reflection

Based on our initial proposal, each of us was responsible for making one view, and we linked the two views together after both of us finished. The “Description of Visualization and Sketch” section of our project proposal made what our project would look like very straightforward to us as well as provided us with a clear, concise thread on how we would implement the visualization. Therefore, we did not spend time figuring out “what to do next” during the project. We did spend a great amount of time debugging JavaScript code, though. The only key in this part was to keep trying and never give up. The process was painful but rewarding.

Our visualization and technical goals mostly remain the same as what we described in our proposal, and we only make some small changes to them. The parts of implementation different from the initial goals are: 1) On the geo map, we use color luminance to encode an additional attribute: median salaries by state, with deeper colors denoting higher salaries; 2) On our bar chart, we use color luminance, instead of color hue described in the initial sketch, to encode the salaries within different percentiles. We add an additional interaction to this view by enabling a selection of the parts of the bars representing a specific percentile so that users can focus only on

the percentile they are interested in. Also, we take away the sorting function we initially presented. Instead, we presort the bars by their *stacked* heights, because after discussing the practicality of sort, we believe that it does not make sense to sort the height between specific percentiles since people using percentile charts care more about the starting and ending points of each percentile.

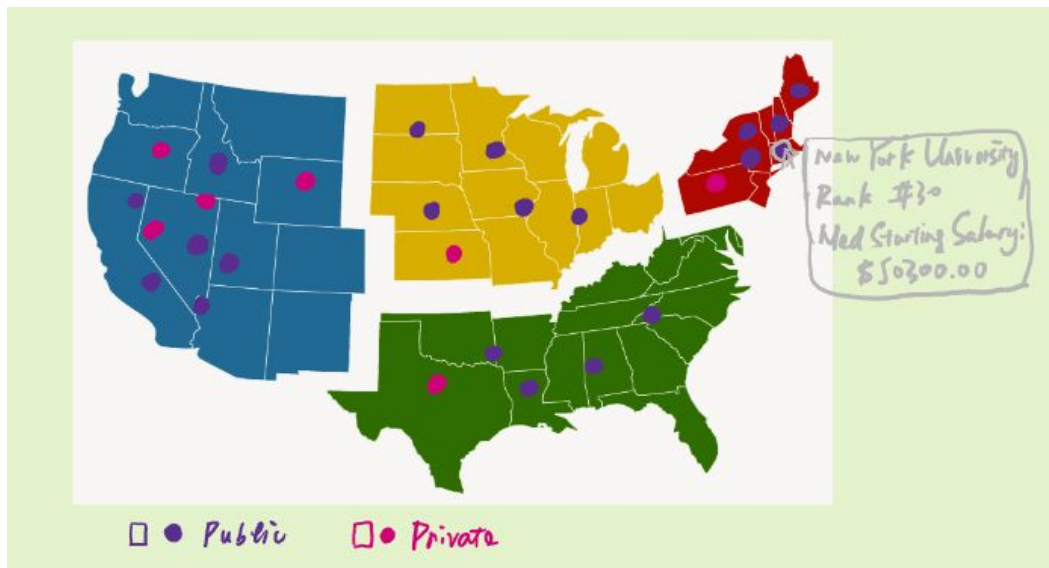


Figure 4: Initial sketch of View 1

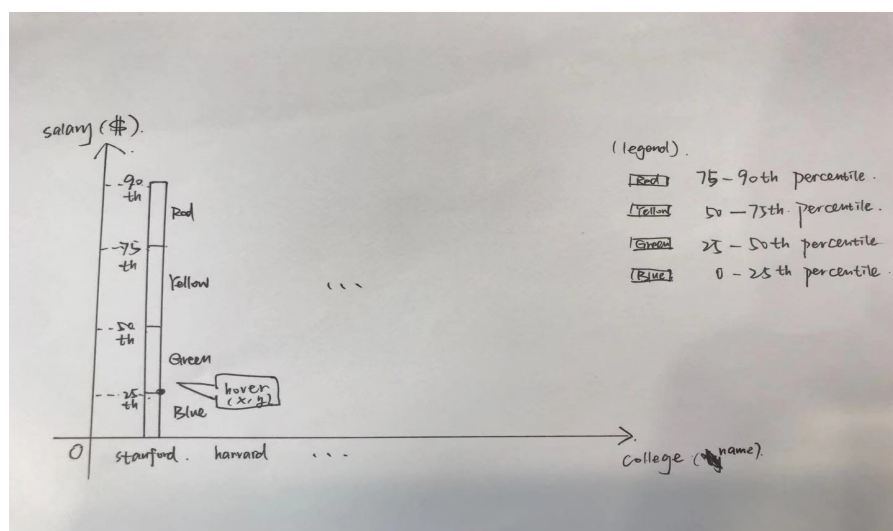


Figure 5: Initial sketch of View 2