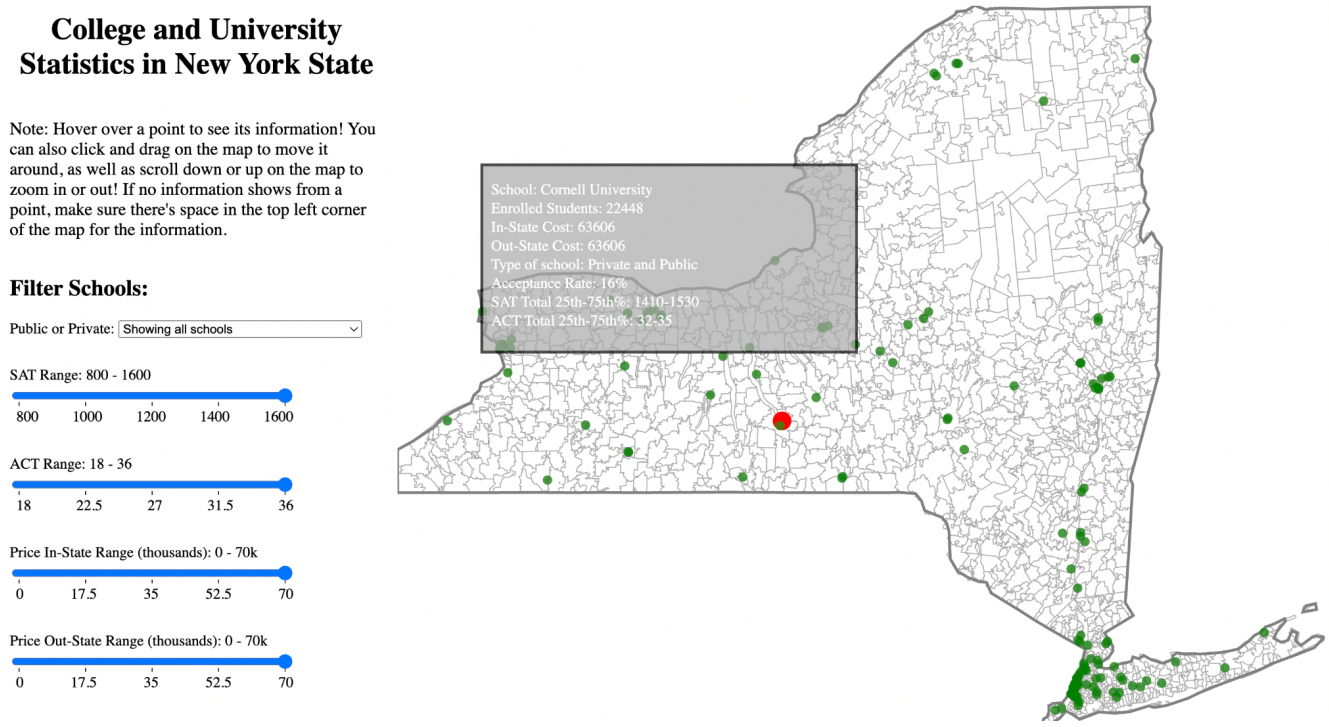
Final Report

I. Finalized Visualization



II. Description of the Data

Data was combined from 3 sources: a list of 238 four-year colleges and universities in NY state (http://www.free-4u.com/Colleges/New-York-Colleges.html), a list of 155 colleges and universities in NY (https://www.universityreview.org/new-york-colleges/), and a list of 170 colleges and universities in the US (https://mytutor.com/blog/admission-statistics-us-colleges/). The first list contained 5 columns, “School”, “City”, “Enroll”, “In-State Cost”, and “Out-State Cost”. The “City” column was dropped. The second list contained 8 columns, but we only kept 4 columns, “Name”, “Address”, “Type”, and “Acceptance Rate”. The third list contained 11 columns, but we only kept 3 columns, “School”, “SAT Total 25th-75th%”, and “ACT Total 25th-75th%”. The datasets were merged on the school name, with a total of 10 columns. Schools without additional statistics were left blank. The final dataset includes 150 schools. A major issue of the merge was that each of these datasets did not always name the school the same way (for example CUNYs and SUNYs) which required lots of manually renaming each school to be the same before the merge.

After merging the three datasets, we decided to manually find the coordinates of each school in New York in order to make it easier for us to place them on the map. To do this, we plugged in the address of each school into a GPS coordinate finder and subsequently added the coordinates to the “Latitude” and “Longitude” fields of the dataset respectively. In addition to finding the GPS coordinates, we also filled in any missing SAT/ACT data that was not present in

the three datasets we used above. Because the dataset for SAT/ACT scores used above was a list for the United States, we found that a lot of the data was missing for schools in the state of New York, with exception to the better known schools (like Cornell and Columbia, for example). Thus, we used online resources, like niche, to fill in the missing data. Although time consuming, we decided this was ultimately crucial to the functionality of our data visualization, as we want to display the data for each school so that prospective college students can have a comprehensive overview. Lastly, we used the ny_income.topo.json file from a homework assignment in order to draw the New York State outline as well as accompanying zip codes.

III. Visual Design Rationale

When deciding how to visualize our data, we knew that we wanted to display the information on a map, since we were focusing on the state of New York specifically, and colleges are each located in an unique location. We had the addresses and thus found the Latitude and Longitude, so it would be easy to visualize the data geographically. Moreover, since we were focusing on one state specifically, we wanted to include a breakdown of the map by zip code as well. This would provide better granularity and more information than a simple outline of the state. In addition, the zip codes inherently outline the finger lakes as well, which is an important visual cue to the user that can help geographically understand where the points are on the map. The map affords the ability to place visual marks for each school, represented by green circles, onto channels of aligned position scales in both the x and y directions. These channels are familiar to the user, since maps are common to us all, and thus the user should have no problem understanding the visual marks and channels in this context.

We also made some important design decisions regarding the colors, size, and opacity of the circles, as well as how these properties would change when the user interacted with them. To begin with, we initiated all circles with a green color, since this is a pretty standard and discernable color for most individuals. We also set the opacity to be 0.7 so that if any two circles overlapped, which was very common for the New York City region, the user would still be able to see dots underneath other dots, rather than circles being cut off. This would allow users to notice that dots were highly concentrated in that region, and use our zoom functionality to further discern the dots from one another. Furthermore, when the user zooms in on the map, the radius of the circles changes accordingly so that the circles aren't excessively large. In addition, the color of the circles changes from green to red, and the radius grows larger, to indicate that you are mousing over the circle. Also, the opacity of the red color is set to 1.0. The red color pops out to the user on the page, and is discernible from the green color used for the other dots. Thus, we ultimately made these design decisions so that it would be easiest for the user to highlight specific schools, and see clearly which one they selected.

Lastly, we also made some design decisions regarding the sliders used to filter the data. We used the visual channel of length and a circle mark at the end of the bar to tell the user where they are on the slider. We also display the range that is currently selected above the bar, so that users are able to see the actual numbers that the sliders are referring to. We used tick marks to

segment the slider into separate sections so that users could easily tell where the slider is in relation to the tick marks. Lastly, we started the SAT and ACT range at the middle values, instead of 0, since all of the schools had a minimum at that value or greater, which allowed for a narrower range of the slider.

IV. Interactive Elements and Their Design Rationale

The interactive elements we have for this project are panning and zooming, sliders to filter data points, a pull-down bar to filter data points, and a mouseover hover on data points to show information about the hovered point. In the visualization, we wanted to be able to show users the information about each school or data point, which led us to the idea of showing the information from hovering over each point. This is the most simple way to interact with the visual that doesn't require much effort from the user and is easily discoverable, but we put it in a note section anyways to make sure it's one hundred percent known. We also included the panning and zooming interactions as part of the note section as well, as these interactions are a lot less obvious to find. When we first created the visualization, we saw how several areas were densely populated with universities and colleges. It's because of this that we found it very difficult to precisely hover over a specific data point to get its data, which is very inconvenient for users. This is why we decided zooming in would be a very helpful tool. Users would be able to zoom in, which will make the points smaller as well, letting the user distinguish between points and actually hover over each point individually. The idea of zooming and panning came hand in hand, as if a user is allowed to zoom, panning would be necessary to go around the map to other places.

The zooming and panning also came in handy for our mouseover hover interaction. When hovering over points, we decided to create a textbox with the information of the point inside of it. In order to make this textbox have space and not cover a point being hovered over, panning and zooming would be necessary to give space for the textbox and data point in order to clearly and concisely convey information. Even with the zooming, panning, and mouseover interactivities, we found that it was still difficult to individually get data points, as near New York City there were still plenty of data points that were either overlapping or too close to each other that it made it hard to hover over them individually. We then decided it would also be a good idea to add different filters to lessen the number of clusters in the visual. These filters would make both hovering easier and would make the visual look cleaner. We wanted to have a bit of diversity in our filters, so we decided to make one a pulldown bar and the others a slider, as it would be less uniform and also make sense for the actual data of each school. The pulldown bar would filter schools by public or private, which makes more sense than making it a slider. The slider filters were for ACT range, SAT range, in-state and out-state price. Since these values deal more with numbers, it made more sense to us to make them sliders as we can create a range of numbers. The visibility for these filters are also very prominent and easy to find, being easy interactivity for users to play around with. Through these interactivities we were able to create a clean visualization that would be easy for users to use and learn from.

V. The Story

Our visualization is a map of New York state with 150 colleges and universities mapped out over it, representing their geographical location within the state. We gave the viewer sliders to use as filters and more information for each college when its dot is hovered over with the cursor. We show information like the size of the student body, the type of school (public or private), how much it costs to go there both in-state and out-of-state, as well as the acceptance rate and test score ranges for the students that go there. We created this to function as a tool that a prospective student or parent could use to learn more about different colleges around the state. This tool can help the student start to explore colleges in New York and ultimately continue research into each institution that piqued their interest.

One thing that we found surprising after playing with our visualization was that when you filter colleges down by the price of tuition, the ratio of in colleges upstate versus in NYC was counter-intuitive. Logically, we expected that as you filter colleges down by lowering the in-state tuition cost on our slider, the ratio of colleges upstate to NYC would increase because the higher cost of living in NYC would be associated with higher tuition costs. However, our intuition was flawed: as we increased the slider for in-state tuition, the number of colleges in NYC seemed to increase much more quickly than upstate. This might be because the significantly higher population density in NYC leads to the need for significantly more schools at all price ranges. Nevertheless, our visualization is ultimately a great tool for both prospective students, and parents of prospective students, to use in order to determine which schools might be a good fit for them in the state of New York specifically.

VI. Outline of Team contributions to the Project

- Isabelle:
    - Manually found coordinates of all addresses of the 150 schools + SAT/ACT scores that were not present in original dataset (4 hrs)
    - Edited previous code (map styling, mouseover, text) and implemented slider filters for the map (5 hrs)
    - Contributed to the final report (2 hr)
- Catherine:
    - Manually cleaned some entries in dataset (2 hrs)
    - Preprocessed and combined datasets (2 hrs)
- Charlie:
    - Helped fix data point hovering interaction (1.5)
    - Implemented and updated pulldown bar functionality to work (2 hrs)
    - Cleaned up and helped organize code (2 hours)
    - Contributed to final report (1.5 hr)
- Christopher:
    - Created structural visualization (background map + plotting of points) (2 hrs)

- Created zoom and pan interaction as well as data point hovering interaction (2.5 hrs)
- Helped with pulldown bar filter (1 hr)
- Contributed to final report (1.5 hr)

Work was broken down as equally as possible, catering to each member's strengths and weaknesses as well as time constraints (if someone isn't available for a certain day they do their part beforehand). We also met once a week, or more if necessary, in order to split the work as well as make sure we were all on the same page. The part of the project that took the most time was manually finding missing data from the datasets we compiled as well as finding coordinates for the address of each school.

Milestone 1

5 Ideas

1. Video games grossing split by platform
    a. Interaction: User chooses which platform the video games will be shown or they can choose which video games are split between platforms. For example, how Minecraft grosses for Xbox compared to PC or how several games like Minecraft and GTA gross for Xbox.
2. Exchange rate of currency versus current events
    a. Interaction: Input a certain number of currency to be exchanged or choose a certain year or time span for the event.
3. Map user selects two countries; Countries currencies are graphed over time in a graph below. (dataset: https://fiscaldata.treasury.gov/datasets/treasury-reporting-rates-exchange/treasury-reporting-rates-of-exchange)
    a. Interaction: Clicking 2 countries will create graph
4. Map with colleges around New York state (possible dataset: http://www.free-4u.com/colleges/New-York-Colleges.html)
    a. Interaction: Clicking on each college will show you stats on the school (avg GPA, avg SAT score, avg % admissions, & tuition cost. Maybe combine with a graph of these stats over time)
    b. Can pan + zoom in on map
5. Map of ski resorts in North America. (www.skiresort.info/ski-resorts/north-america/)
    a. Interaction: click on a specific ski resort to see the resort name, number of lifts, skiing area acreage.

Assigned Tasks

- Isabelle:
    - Brainstorm / sketch visualizations & possible interactions
    - Begin coding the visualization
- Catherine:
    - Find 2+ datasets, clean data and combine
- Charlie:
    - Search for more possible dataset to combine into our current dataset
    - Brainstorm how we intend visualization to look
- Christopher:
    - Sketch different interaction ideas for datasets
    - Brainstorm different visualizations for ideas

Milestone 2

Final Project Idea

A New York State map that shows different colleges around the state (represented by dots on the map). The interaction includes clicking on each college to show different statistics on the school such as avg GPA, avg SAT score, avg % admissions, & tuition cost. We might also maybe combine this visualization with a graph that displays trends over time, or an interface to compare different colleges to one another. The user will also be able to pan and zoom over the map.

Last Week Contributions

- Isabelle:
  - Sketched out visualization and the associated envisioned interactions.
- Catherine:
  - Dataset (with geoJSON data but only for NYC): https://data.cityofnewyork.us/Education/Colleges-and-Universities/4kym-4xw5
  - Statistics (not all NY colleges): https://mytutor.com/blog/admission-statistics-us-colleges/
- Charlie:
  - Found a useful dataset that also has addresses which we could use to place each school on the map: https://www.universityreview.org/new-york-colleges/
  - http://www.free-4u.com/Colleges/New-York-Colleges.html
- Christopher:
  - Brainstormed and sketched new visualizations.
  - Brainstormed different interactions from previous milestones.
  - Dataset (admission rates + ED + regular decision apps) https://www.ivycoach.com/ivy-league-admissions-statistics/

Assigned Tasks

- Isabelle:
  - Finalize sketches of visualization.
  - Come up with a plan for the coding implementation.
- Catherine:
  - Combine datasets and filter columns (we only want geo data, admission stats, and tuition costs)
  - Manually fill in admission data (from school website) if missing
- Charlie:
  - Start exploring different elements of the implementation.
- Christopher:
  - Start coding the visualization and get a basic structure ready
  - Explore how different visualizations would look like in the code.