College Campus Security

Elisabeth Dewailly, Dennis Sun, Ariel Camperi, Myles Novick
{edewailly, dennissun, acamperi, mylesnovick}@college.harvard.edu

Final Project Process Book cs171 Spring 2014

Table of Contents:

Crime, tuition, financial aid, race, college money (use finances total assets for all types of colleges), crime by county (get it)
Large map of crimes per state; average of crimes per capita for universities scaled to color
Append each school at its longitude and latitude as a circle.
Zoom to states, give crime by county option
Toggle schools based on school type, give all demographic/tuition options for detail view

Overview and Motivation

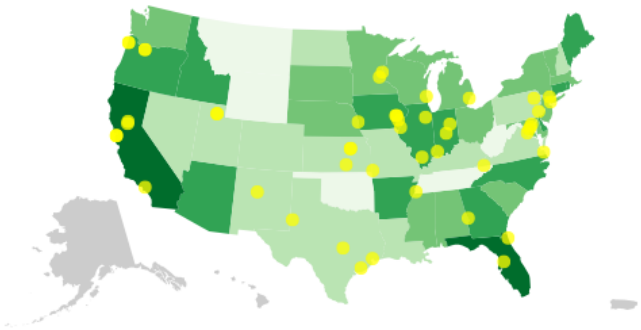
Provide an overview of the project goals and the motivation for it. Consider that this will be read by people who did not see your project proposal.

When searching for a project idea, we knew we wanted to make something functional and pretty, that was actually useful (i.e. not just showing a list of data points as a bar chart). Thus we looked for datasets that were big and had lots of different categories and variables, so that we could combine these into meaningful visualizations, which showed \new" information that couldn't easily be seen by just reading through the data in table format. As we live on a college campus, and periodically receive those emails from HUPD about crimes in and around the campus, we thought it would be interesting to look at campus security data for college campuses in general. This led us to the government data sources suggested by the course staff, where we found a treasure trove of data we can process and from this make a beautiful visualization.
Upon getting some feedback, we decided to change the scope of our visualization from a simple visualization of a map leading into multiple visualizations of crime data with various subcategories into a map which is based on displaying information about crime (again, with multiple subcategories) per capita on a state level for the entire population and average for universities. In addition to crime, we decided to add tuition, financial aid, race demographics, and endowment as a detail visualization for

each state overall and for each university. We wanted a user to be able to see the relevant crime statistics for these universities while still being able to look at other relevant information about the university which might interest a potential student/concerned parent.

Related Work

Anything that inspired you, such as a paper, a web site, visualizations we discussed in class, etc.



Our inspiration for the appearance and choropleth techniques for the map was first the above figure 12-6 from Interactive Data Visualization for the Web by Scott Murray. The dots for cities could be applied in the same way to mark universities, and we could have detail visualizations with crime and other university data on both a state level and university level.
Our inspiration for the zooming techniques we used to be able to focus on each university came from this website: http://techslides.com/demos/d3/worldmap-template.html. Our inspiration for creating a nice little "detail visualization" which appears beneath or to the side of the graph upon clicking on a state or dot came from Homework 3 for cs171. We were planning on having some sort of detail visualization for the prior to Homework 3, but we were not yet quite sure how to present it.

Questions

What questions are you trying to answer? How did these questions evolve over the course of the project? What new questions did you consider in the course of your analysis?

We were initially only interested in knowing the crime statistics for each university in the U.S. So, our questions were limited to: What are the relative amounts of crime at universities for each state? What are the relative amounts of crime at each university? How do individual universities and state averages for universities compare with respect to specific crime subtypes?

We then expanded our scope to finding out about crime for the entire population and finding some other demographic and financial information about the schools. So, we are now also asking: What are the relative amounts of crime for each state? What is university average of the overall demographic and

financial data for each state? What is the demographic and financial data for each university? Do universities with the least crime tend to have certain demographic or financial characteristics?

Data

Source, scraping method, cleanup, etc

All college specific data except crime was gathered from The National Center for Education Statistics (this includes tuition, financial aid, demographics, and endowment assets as well as several other fields). We gathered college locations (longitude and latitude info was found using the Google Maps API) from The Campus Safety and Security Analysis Tool (http://ope.ed.gov/security/Index.aspx). We also gathered college specific crime statistics from The Campus Safety and Security Analysis Tool. We gathered statewide crime statistics from the Uniform Crime Reporting Statistics site (http://ucrdatatool.gov/Search/Crime/Crime.cfm), this information is provided by the U.S. Department of Justice and the Federal Bureau of Investigation. We used this information to generate an object whose properties are unique identifiers for various colleges. Each of those colleges' ID, once indexed into, yields objects containing object data relevant to several things. First, crime stats (an object containing fields such as total, types, locations of crimes, etc.), then various things such as location of school, name, endowment info, then an object containing info on demographics, then an object containing info on financial aid, and lastly an object containing info on tuition. All this was compiled into a large nested JSON object. A secondary super-JSON object exists containing state aggregate data. This data is the average endowment assets on a state-by-state basis, the average crime rates broken up categorically using the schools within the state, the overall average demographics of schools within the state, and the average financial aid and tuition of schools within the state. Refer to data_schema.txt (within data folder) for more detailed information on the architecture of these JSON objects.

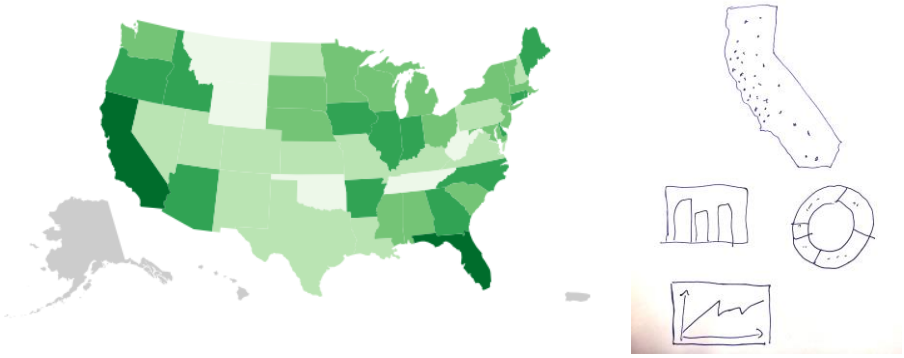
Exploratory Data Analysis

What visualizations did you use to initially look at your data? What insights did you gain? How did these insights inform your design?

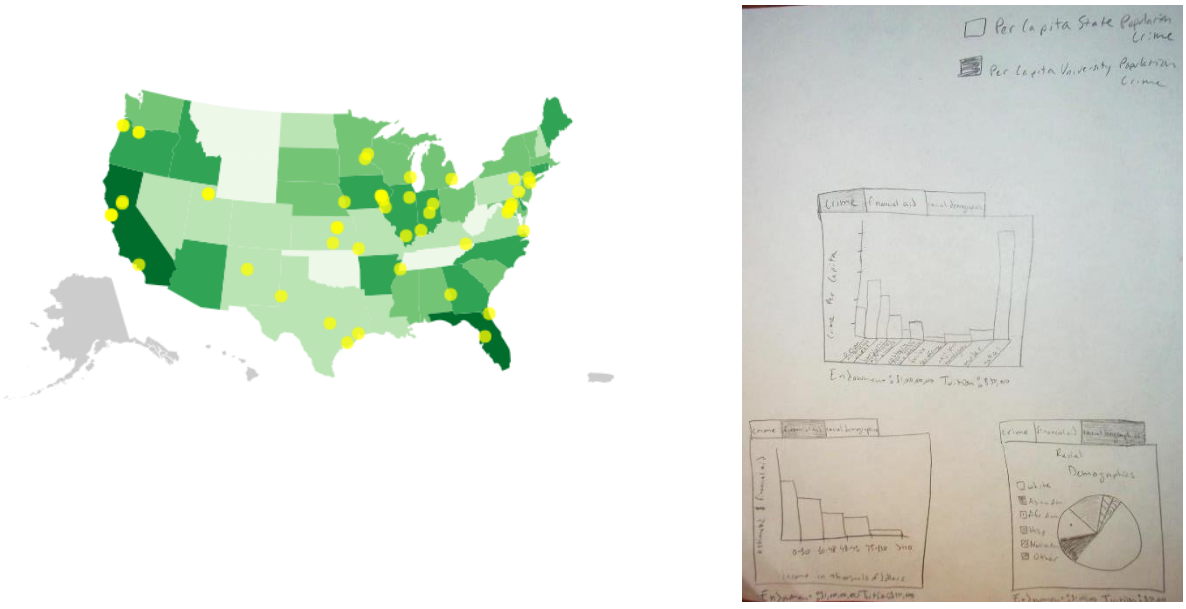
We initially looked at all of our crime data for the population and universities on a state level using a choropleth US map to see if there was some variance between states, and we saw that there was. This led us to believe that there can be significant differences in crime rates between many universities across the United States, but potentially not between universities in the same state, so we decided not to color the dots marking each university on a scale dependent on total crime for that university.

Design Evolution

What are the different visualizations you considered? Justify the design decisions you made using the perceptual and design principles you learned in the course.

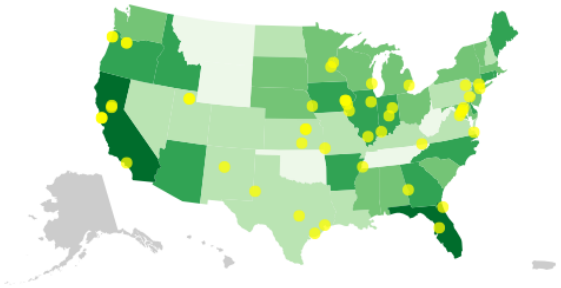


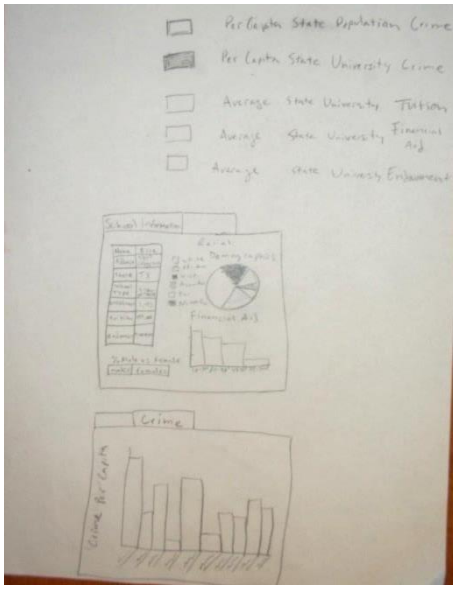
We initially had the above visualization in mind, with with dots representing the different college campuses on the state. Each dot would have varying color based off of total crime for that university. Additionally, under the state map will be several bar charts or other graphics representing safety statistics for the college campuses in that state. Selecting a specifc college campus would filter the data for these graphs to only show data for the selected campus.



We scrapped the idea of varying the color of the dots, and instead decided to just have dots of a consistent color which only appear upon state zoom. The varying color would look too messy, as would the huge amount of dots that we would see on the map when viewing the entire US. Also, since the states are all variations of the same color, this visualization would pose no problem for people who are colorblind. Our detail visualization became more defined to represent crime data as a bar graph, and we added a financial aid bar graph and a racial demographic pie chart, with endowment and tuition below

the detail visualization. We had both overall population and university average as options for the variation in color of the states (choropleth).



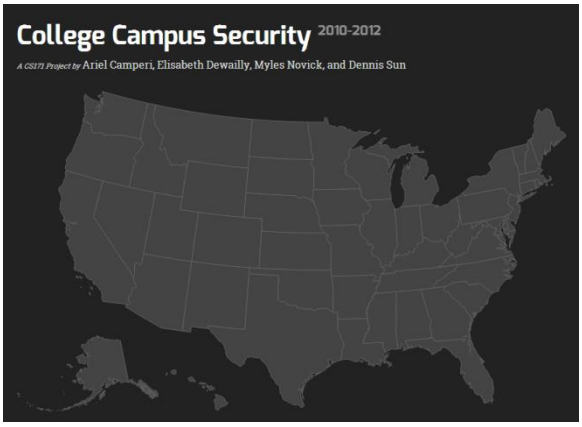


We then decided to put financial aid and racial demographics in the same tab, while adding general information about the school on the left such as name, address, state, tuition, endowment, total enrollment, and school type. This general information tab would be the first, and the second (last) tab would contain the crime data for that university. This would aggregate all the generally significant university data and allow the crime data to be the only other main tab at a larger size, emphasizing the importance of crime data in our visualization. We also want to use the same detail visualization for general information and crime at the state level. We also decided to add tuition, financial aid, and endowment as options for the color scale of the states. The option of varying the color scale would allow for easy comparison between states for most of our relevant dimensions that complement crime.

Implementation

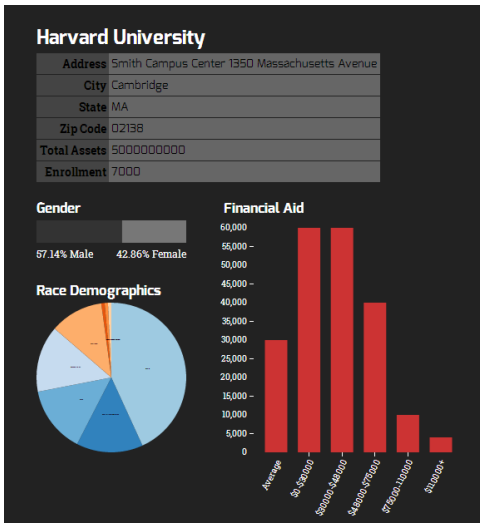
Describe the intent and functionality of the interactive visualizations you implemented. Provide clear and well-referenced images showing the key design and interaction elements.

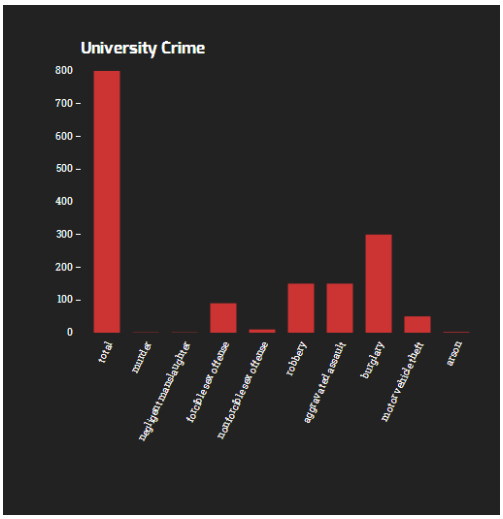
1. We created a zoomable, draggable map of the US, where the zooming is possible beyond the state level, and you can drag the map to move your perspective.



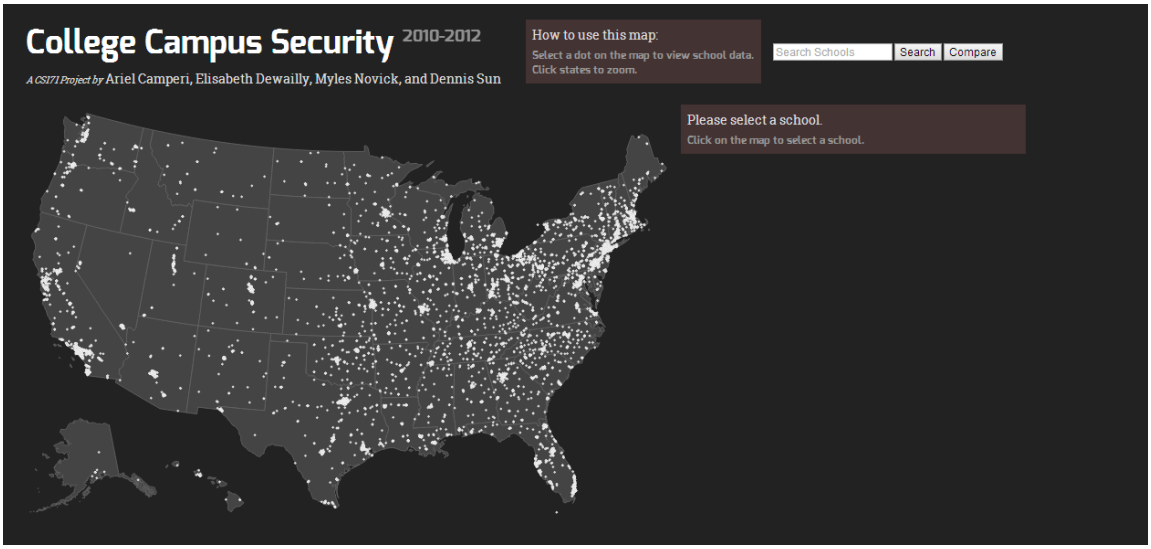
2. We created a map of the US which contains dots which appear upon state zoom to represent each university in the US

3. Upon clicking each dot, a detail visualization appears showing some demographic and general financial information about the school (financial aid, racial demographics, endowment, enrollment, address, etc.), with a toggle/tab to also show the crime data for each university in the categories of total crime, murder, negligent manslaughter, forcible sex offense, non forcible sex offense, robbery, aggravated assault, burglary, motor vehicle theft, and arson





4. We finished collecting and placing dots for all of the data for each university in the US, and we also added some storytelling elements, such as instructions on how to use the map

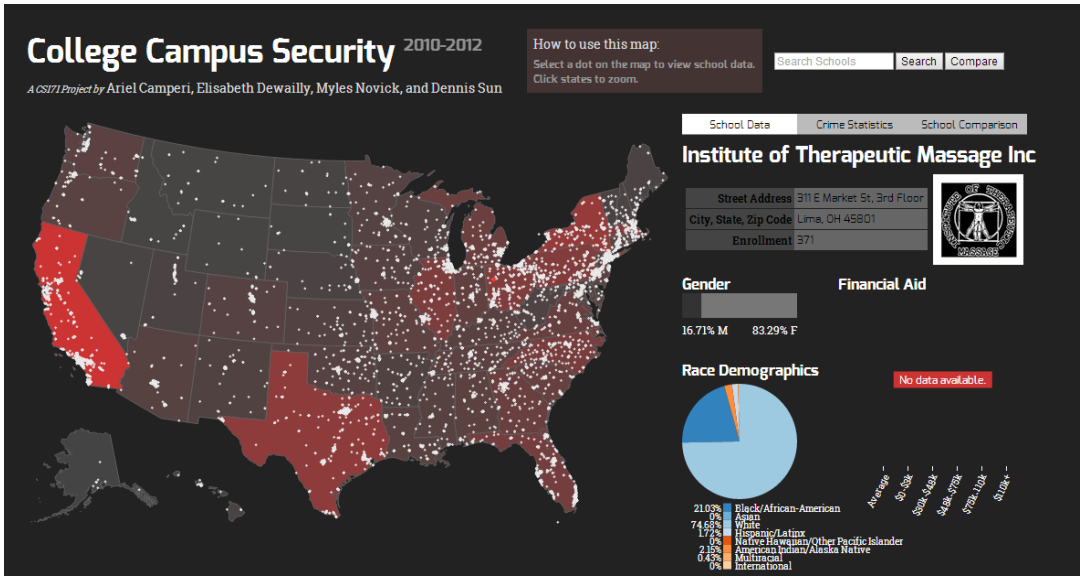


5. We implemented a tabular view of the data so that there are only two tabs (one for school data and the other for crime statistics) on the side of the US map to minimize scrolling and emphasize the importance of crime statistics

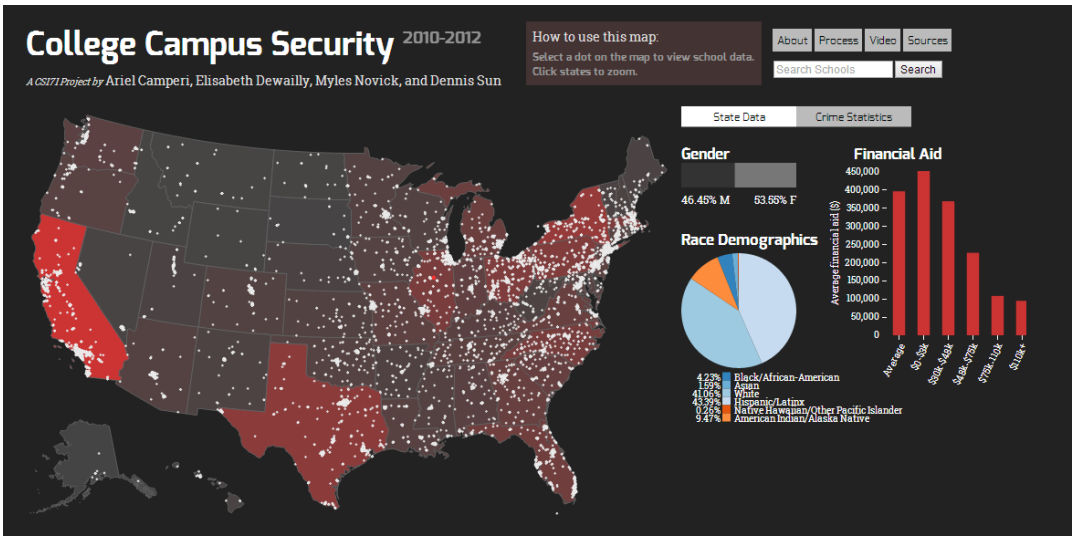
6. We implemented tooltips upon scrolling over each university and a “no data available” message when a certain university does not have crime data or financial aid data available

7. We created a map of the US in which the state varies in color based off of the average of the total crime for the universities within that state OR (toggle between both options) varies in color based off of the per capita crime rate for the entire state population.

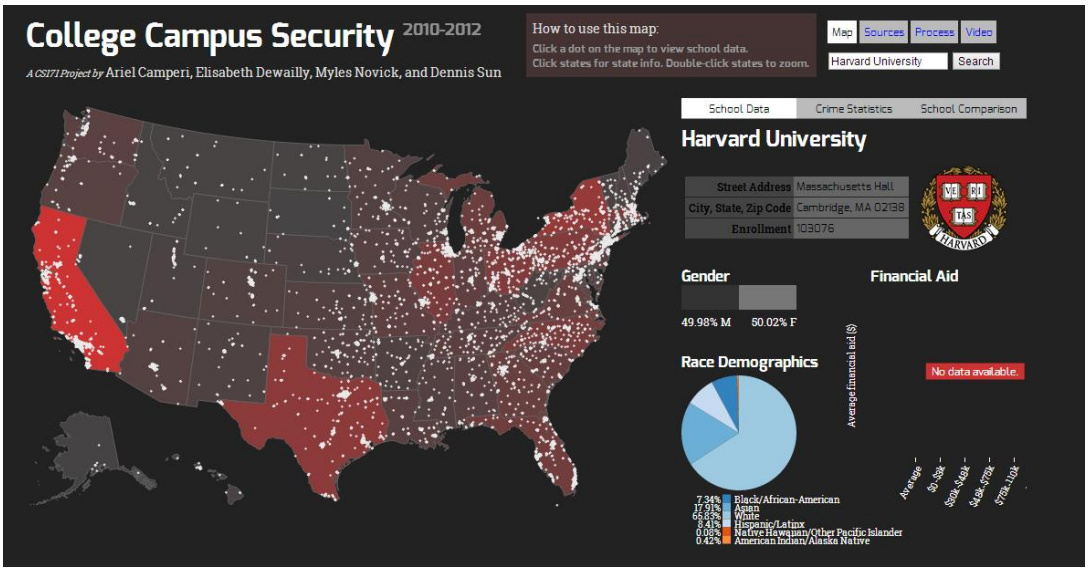
8. We also added images corresponding to the logo for each university and improved upon our pie chart detail visualization. The logos have a link to the source of the image.



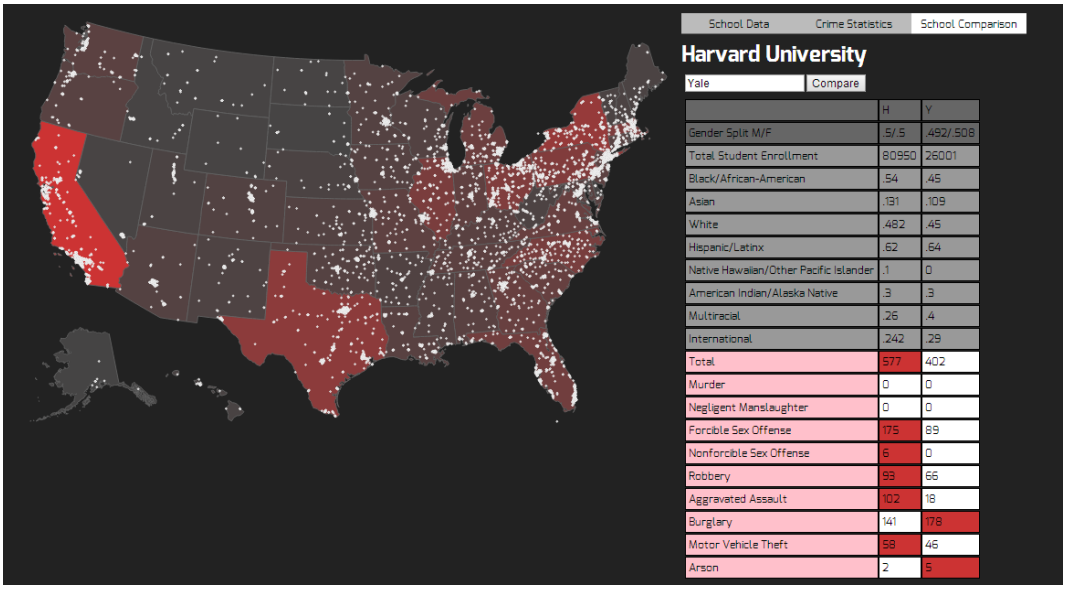
9. We then added a detail visualization for each state showing the overall state data and crime statistics



10. Next, we implemented the search functionality so that we could search for a given school based off entered text, such as “Harvard University”



11. When you click on the School Comparison tab, you can enter another school, such as Yale, and compare gender, enrollment, demographics, and crime data between both universities. For the crime data, dark red highlights the greater of the two compared values.



	H	Y
Gender Split M/F	.5/.5	.492/.508
Total Student Enrollment	80950	26001
Black/African-American	.54	.45
Asian	.131	.109
White	.482	.45
Hispanic/Latinx	.62	.64
Native Hawaiian/Other Pacific Islander	.1	0
American Indian/Alaska Native	.3	.3
Multiracial	.26	.4
International	.242	.29
Total	577	402
Murder	0	0
Negligent Manslaughter	0	0
Forcible Sex Offense	175	89
Nonforcible Sex Offense	6	0
Robbery	93	66
Aggravated Assault	102	18
Burglary	141	178
Motor Vehicle Theft	58	46
Arson	2	5

Evaluation

What did you learn about the data by using your visualizations? How did you answer your questions? How well does your visualization work, and how could you further improve it?

We learned that universities with the least crime tend to be in more rural areas, but they also tend to be more prestigious universities. With this response and visualization, we were able to answer the following questions: “What is the demographic and financial data for each university? Do universities with the least crime tend to have certain demographic or financial characteristics?”. We also learned that Harvard has more campus crime than Yale by directly comparing them using our search and compare functionality. Upon plotting all of our dots for each university on the map, we were able to see that the majority of universities tended to cluster around the more populous areas in the U.S., which gave us an indication of the completeness of our data set.

We were also able to answer the question of what are the relative amounts of crime at universities for each state. With the choropleth functionality we noticed the high incidence of crime in California. Looking at Stanford, we saw that the majority of crime came from burglaries. We posited that these burglaries were largely bicycle theft, so we went on to look at other large universities in which many of the students own bicycles for comparison. Texas & M in College Station showed that the majority of thefts there came from burglary as well, so we came to the conclusion that our hypothesis about crime at universities was fairly likely.

To improve our visualization, we could collect and represent crime data corresponding to the general population of each county. We could also create a detail visualization for the entire country.