

Kevin Tian kt474, Eamon Woods etw33, Matthew Knight msk284  
INFO/CS 3300 Written Description

A. Our main dataset was taken from <https://collegescorecard.ed.gov/data/>, which has been updated as recently as 2018. The subset in our submission is called 'includes\_ugds.csv'. We used the most complete and recent data (see 'Most recent data' link). Since this file was 141 MB, we had to manually reformat it with only the variables we were interested in. After we manually reorganized our data in Excel, our file was only 3.5 MB, which loaded in much faster. The variables/columns in includes\_ugds.csv that we used were:

1. Institution name - INSTNM
2. City - CITY
3. State - STABBR
4. Zip code - ZIP
5. Latitude - LATITUDE
6. Longitude - LONGITUDE
7. School type - CCUGPROF (we used this to detect 4-year, full-time institutions)
8. Admission rate - ADM\_RATE
9. SAT average - SAT\_AVG
10. Undergraduate enrollment (degree seeking) - UGDS
11. Average cost of attendance - COSTT4\_A
12. Public/private - CONTROL
13. Diversity of student body - UGDS\_\* (besides UGDS\_MEN,UGDS\_WOMEN)
14. Gender breakdown of student body - UGDS\_MEN, UGDS\_WOMEN

We chose these variables because we thought they offered a broad but not overwhelming view of each college. We then saved and loaded in our data in a csv format. We used the latitude and longitude columns to project the colleges on a map of the U.S, and we used data from the other columns to display information about each school. For example, we used the information about diversity and gender to create pie charts.

Our complementary data set was stripped manually from <https://www.currentresults.com/Weather/US/weather-averages-index.php> and placed into a csv file. The dataset contains yearly temperature, humidity, and snowfall data for all fifty states as well as the cities in which the data was recorded. This data is used to give some context to the locations of the colleges shown on the map. We decided that this role of weather as context justified a state-level granularity. In other words, since the project is primarily about schools and not weather, we just wanted to give the viewer a general sense of the climate surrounding each school.

We also used the us.json file from the 2/21 lecture in order to create a 2D map of the U.S.

The main data filtering we did (aside from cutting down the 141 MB file) was choosing a subset of the 7,593 schools to display on the map. We used three criteria to select this subset:

1. CCUGPROF value of between 10 and 15, which means that a school is 4-year and full-time.
2. Admittance rate less than 45%. This aligned with our project's mission to display "a look at the best schools in the U.S."
3. Not in Puerto Rico or Guam. These could not fit easily on a U.S. map.

B. In order to map latitude and longitude on a 2d map, we had to use a projection. We used/built on the code from the 2/21 lecture. We used the `geoAlbersUsa` projection with a scale value of 75. We have another variable `pathGenerator` that uses the d3 geo path tool to create a projection. With these two variables, we only needed to map the projection of the longitude as the x value and the latitude as the y value. Each school we mapped with either a white circle if it was a private school and a black circle if it was public. We implemented "details on demand" by having circle-clicks append information to another svg: facts and pie charts about the selected school show up on the side whenever a circle is clicked on. Additionally, hovering over circles shows the name of the school. In terms of the pie charts, we used a d3 scale to show the different percentages of each race at every school. The diversity pie chart had 5 sections for white, black, hispanic, asian, and other. We also added a male/female pie chart with two different colors - blue and pink - to show the percentage of each gender at each school.

We have two svg rectangles disguised as buttons. One shades the 2d map for temperature using a temperature scale to relate the weather data to colors and a set of if statements to find and shade the proper states, and the other shades for snowfall using the same method with a snowfall scale. Accompanying the shading is a legend that shows the values of the colors by adding tick marks to a color gradient bar.

C. Our visualization shows 4 year colleges in the United States with an acceptance rate below 45%. The goal was to give prospective high school students a way to decide which schools would be the best for them. The visualization can be viewed from the perspectives of: location in the U.S., public/private, weather in the area, tuition, admittance rate, average SAT score, number of undergraduates, diversity, and male-female ratio. It was surprising that even when most or all of the schools are mapped, there is still a very high density in the northeast and in California. There do not seem to be many 4 year schools in the mid and northwest. Furthermore, when schools with only low acceptance rates are mapped, there are only schools in the northeast (Ivys) and on the west coast. Lastly, some of the diversity pie charts were interesting to look at. Schools on the west coast generally had a larger percentage of asians while schools on the east coast had more of an even distribution. There were also some regions in the south where there is a mixture of predominantly white and predominantly black schools.

## Citations

We could not have done this project without the code samples and past work of others that we saw. All of these are cited in links at the top of our index.html file as well as below.

Looked at examples from:

Matthew Knight's project 1

<https://bl.ocks.org/mbostock/2206590>,

<http://bl.ocks.org/NPashaP/a74faf20b492ad377312>

<http://bl.ocks.org/michellechandra/0b2ce4923dc9b5809922>

<https://github.com/mimno/info3300-spr2018> (lectures 2/21, 2/16)

<https://bl.ocks.org/mbostock/4090848>

<https://bl.ocks.org/mbostock/3887235>

[https://mimno.github.io/showcase/project2/america\\_air/](https://mimno.github.io/showcase/project2/america_air/)