Project Proposal

CS – 5630 / CS – 6630
University of Utah

Team:
Dylan Zwick
dylanzwick@gmail.com
UID: u0075213

Clark Barnett
twentyonetroy@gmail.com
UID: u1119424

Background and Motivation:

This project grows out of a big data competition in which one member of the team, Dylan Zwick, was a participant. The goal of the competition was to take the College Scorecard Dataset (compiled, published, and made available to the public by the U.S. Department of Education) and use it to construct something useful for high school students as they are trying to decide upon a college.

For the competition, our team used these data to cluster similar schools together, and provide a "distance" between any two schools. For each school this provides an ordered list of most similar schools, and Dylan built a D3 visualization for exploring these similar schools. The idea being that if a student has a school that he or she likes, he or she can discover schools similar to the favorite school about which he or she might not be aware.

This visualization, while a start, can be greatly expanded and improved. For this project we'd like to build upon this visualization to construct a tool that will be useful for anybody, high school student or not, who would like to learn more about colleges in the United States.

Project Objectives:

The visualization should provide a tool for the exploration of colleges in the United States. The primary use case will be for high school students. The tool should allow the student to discover schools that may be a good fit for them, based upon information the student provides about himself or herself and his or her preferences.

In addition, the tool should provide a way of exploring and understanding college data on the state level. So, the tool should provide a way for a user to find out which state has the highest (or lowest) percentage of its population in college, or which school in California has the lowest tuition where the average SAT score is above 1400.

At the conclusion of the project, we'd like to have a tool that is informative, intuitive, and beautiful. We'd like to learn how to use D3 to create this tool.

Data

The primary source of data for this project will be the college scorecard data coming from the U.S. Department of Education:

https://collegescorecard.ed.gov/data/

If we have the bandwidth, we may try to incorporate additional data (like from the US News and World Report college rankings), but all the data we'll need for the visualization as currently planned can be found within this "scorecard".

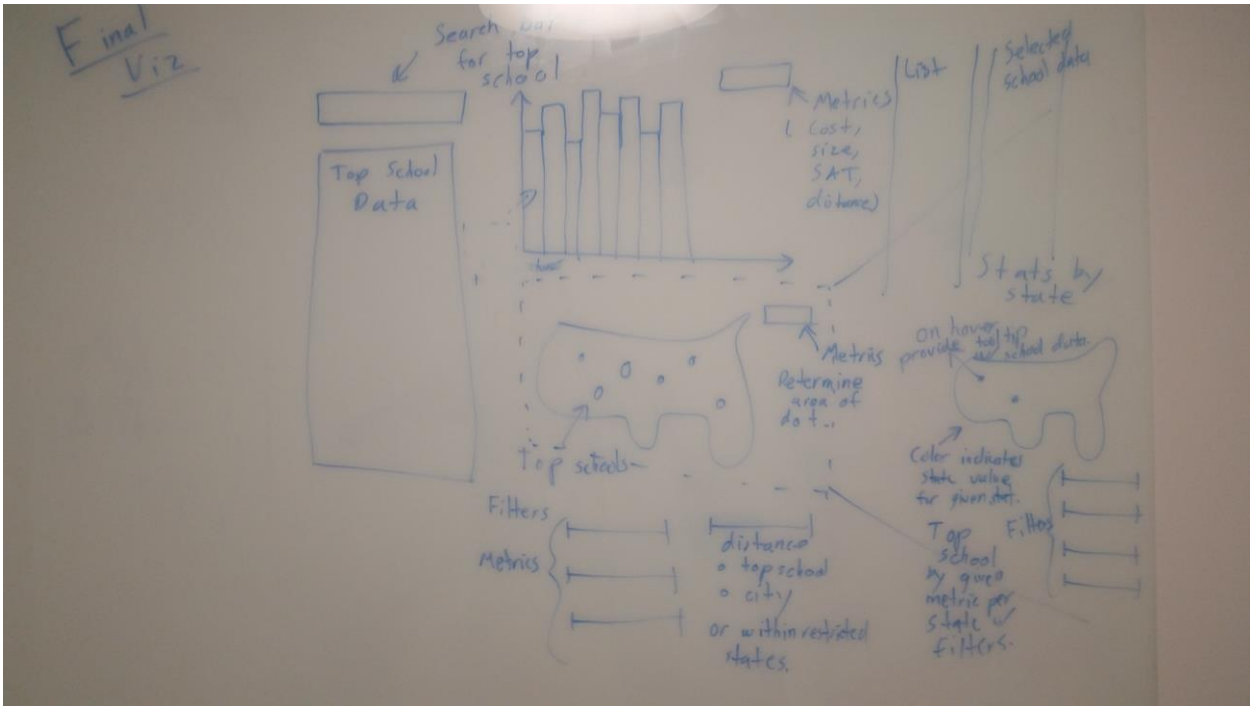
Data Cleanup

There was a substantial amount of data processing and cleanup done on the data for the clustering and ranking algorithm. While we'll be using the output of this processing, none of it was done by us.

The data as it currently exists is not ideal. For example, the cost of the school is not broken out (when appropriate) by in-state vs. out-of-state tuition. Also, there are some problems with missing values. While imputing missing values was necessary for the clustering algorithm, we will not be using any imputed values for the construction of the visualization, as that might provide incorrect information to the end user. For dealing with missing values we will just make sure that our D3 visualization displays something appropriate like "data not available" for these values.

Visualization Design

Sheet 5:



Our plan is to start with a search field that the user can use to input his/her school of interest. Below that will be a module that displays the selected school data as text. These data include: Average School SAT Scores, Tuition Cost, School Size, etc. We think it is important for the user to see these values explicitly displayed.

The user will then be able to view data about schools similar to the selected school of interest. The focus of our design is around a map of the United States and a bar chart. As our data set and visualization objective is highly geographic, it is important that we represent the data with a map so the user knows where the school is, and is able to visualize it geographically in relation to other schools. The map will have each similar school displayed as a circle on the map, with the size of the circle dependent upon a metric of choice. So, the circle will not only represent the location of the school, but also a metric of interest. These metrics can be selected by a drop-down menu. We are double-encoding these circles, using size of the circle and also a color gradient as channels to represent the data. This double-encoding will help the user compare the metrics for each school.

The bar chart in our visualization will also be flexible. Each school metric can once again be selected with a drop-down menu and displayed visually on the chart. Each bar will be a school, and the bars will be ordered according to how similar the school is to the favorite school. Bar charts are always powerful because they use the two most effective channels – position and length – to encode the data.

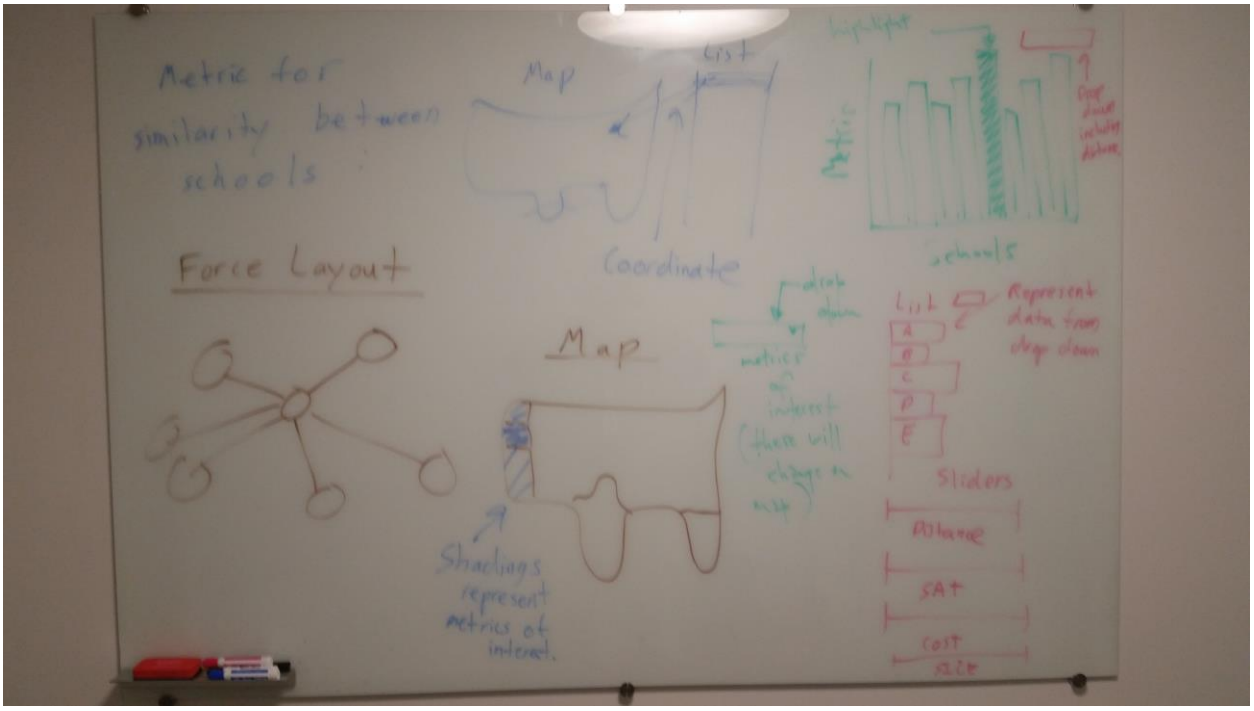
In addition to the bar chart and map visualization there will be list of similar schools, ordered by similarity, on the right of the screen. If the user selects a similar school from the bar chart, the map, or the list, the corresponding element in all three visualizations will highlight, and summary data about the selected similar school will be presented on the far right of the screen.

Our visualization will provide a set of filters that will allow the user to expand or reduce his/her selection by brushing any of the school metric scales. Each metric in our data set will have its own scale that the user can interact with. The initialization of these scales will be based on the similarity score in our data set, and the initial brushed portion will include the range for the similar schools, but the user can then expand or reduce or move that selection and the map and bar chart will adjust accordingly.

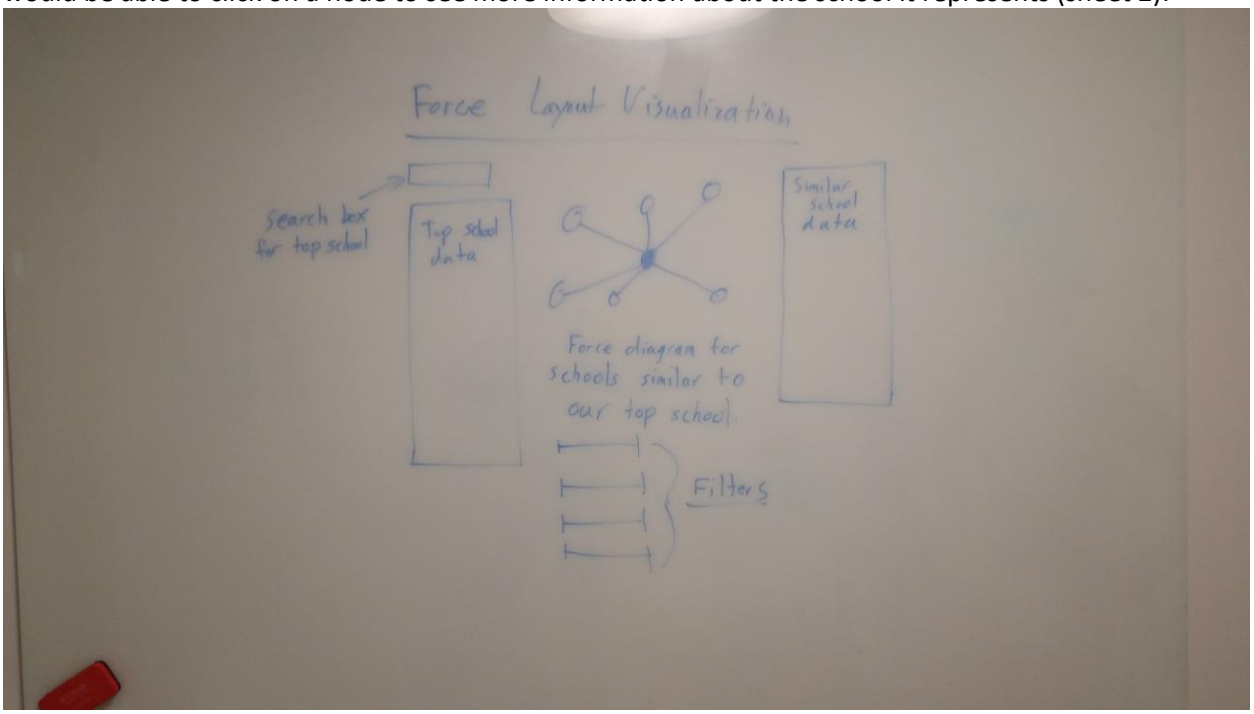
The user will also be able to expand the map visualization, and explore summary data by state. So, for example, the user would be able to explore which state has the highest number of students per capita, or the highest average tuition. This will be available by shading the states, and hovering over a state will introduce a tool-tip with more information.

Additionally, the user will be able to look at top (or bottom) schools by state based upon a given metric. So, the user could look at the school in each state with the highest average SAT, and these schools would all show up on the map visualization (as circles) with the size of the circle representing the average SAT.

In choosing this visualization, we played around with a few ideas. See our brainstorming slide below (sheet1):

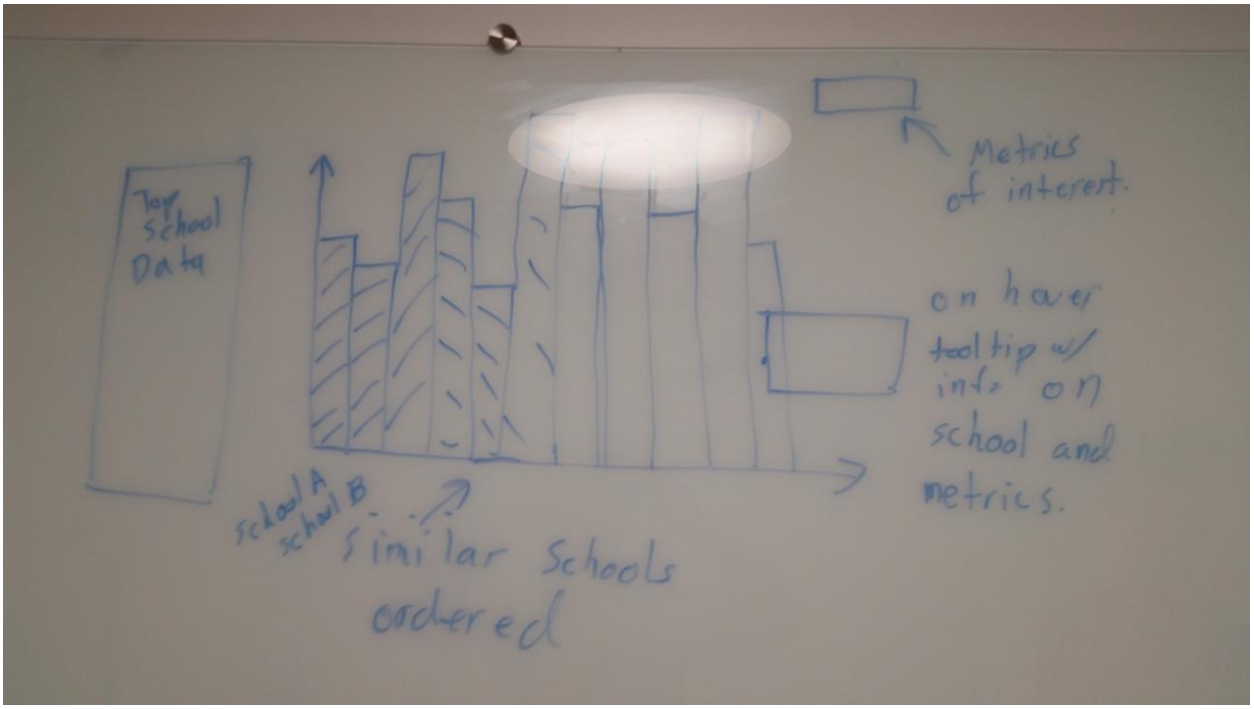


The first visualization we discussed was a force-layout, where the distance between the nodes would represent how similar the school was to the selected school. Filters would also be available, and the user would be able to click on a node to see more information about the school it represents (sheet 2):

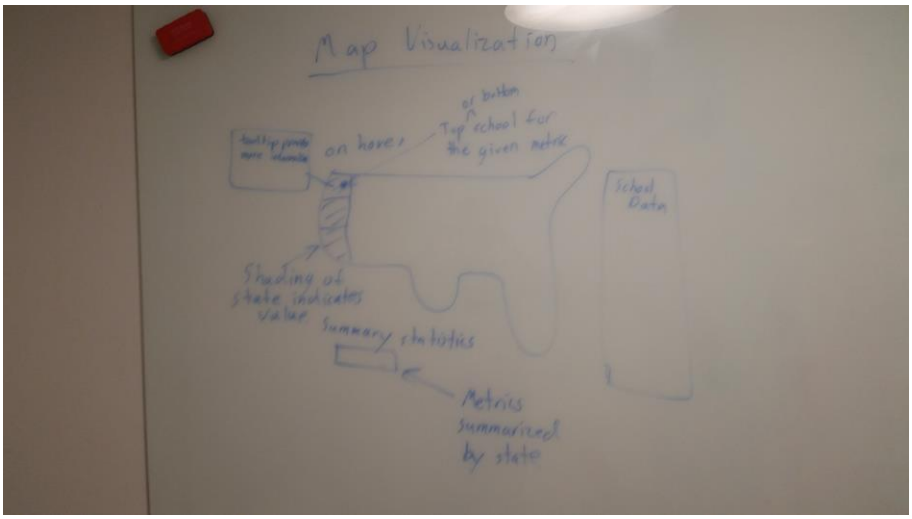


This idea was, essentially, abandoned as we decided (after consultation with the professor) that a force layout visualization didn't do much (the edges would be essentially meaningless), and we figured there would be more effective ways of visualizing an ordered list.

Our second idea was to use a column chart, similar to the one we created in homework 3 (sheet 3 below). We liked this idea more, as it effectively encoded order information by the locations of the bars (which does a better job of specifying order than distance in a force layout), and allowed us the opportunity to display additional data more effectively using the height of the bar (superior to the area of the circle) and the color of the bar.



Finally, as these schools exist at locations within the United States, we figured a map layout would make sense (sheet 4):



This layout would let the user know where the school in question was within the United States (important information for which we don't think there's a better representation), while also providing additional opportunities (through the size and color of the circle indicating the school's location) for displaying metric data about the schools.

As the ordering information is lost by placing the schools on a map, we figured the order of the top similar schools could be provided by an ordered list to the right of the map. Kind of easy, but sometimes the easiest way to do things is the best.

We combined ideas from these three prototypes, particularly the second and third, into our final visualization design, pictured at the start of this section.

Must-Have Features

"List the features without which you would consider your project to be a failure." – That's a somewhat harsh way of putting it!

The project will need to have an interactive visualization with which the student is able to:

- Select a top school
- Filter based upon metrics important to the student
- Visualize similar schools and their corresponding metrics both using a map layout and a chart visualization
- Explore schools that meet filtering criteria, but without reference to a top school.
- Visualize education data by state

Optional Features

Features that it would be nice to have but that aren't critical:

- Filtering based upon state or proximity to a city. So, the ability to filter for only schools within a selected set of states, and the ability to choose those states from a map.
- Filtering visualization. So, when filtering is done using a slider, a chart will pop up that will help the student visualize how many schools are included/excluded by the filter. In the initial design, this number will just be displayed, but there will be no visualized data regarding how much this number changes when the slider is moved.

Project Schedule

October 28, 2016

Dylan

- Skeleton US Map
- Implement search field
- Implement module with selected school metrics

Clark

- Skeleton Bar Chart

November 4, 2016

Dylan

- Implement filters skeleton
- Implement map drop-down menu

Clark

- Implement bar chart drop-down menu
- Create all required CSV files

November 11, 2016 – **PROJECT MILESTONE DUE**

Dylan

- Make filters fully operational

Clark

- Make bar chart fully operational with drop-down menu

November 18, 2016

Dylan

- Implement transitions on map

Clark

- Implement transitions on bar chart

November 25, 2016

Dylan

- Implement zoom-in feature

Clark

- Implement visualization of zoom-in map

December 2, 2016 – **FINAL PROJECT DUE**

Dylan

- Final touch-ups

Clark

- Final touch-ups