

Intuitive Comparison of University Statistics

Samuel Leventhal and Unnikrishnan Rajagopalan

1 Index

1. Initial Proposal
2. Peer Feedback
3. Data Process
4. Peer Feedback
5. Data Processing
6. Meeting with the TA
7. Obstacles, Improvements, and Design Changes
8. To Do

2 Initial Proposal

2.1 Basic Info

Project Title: An Intuitive Look on College Comparisons

Authors: Samuel Leventhal & Unnikrishnan Rajagopalan

e-mail Addresses: Samuel (samlev@cs.utah.edu) Unnikrishnan (unniar@cs.utah.edu)

UIDs: Samuel (u0491567) Unnikrishnan (u1010114)

Repository: <https://github.com/sam-lev/2016-dataviscourse-homework-Sam-Leventhal/tree/master/project>

2.2 Background and Motivation

. Discuss your motivations and reasons for choosing this project, especially any background or research interests that may have influenced your decision.

The project proposed results from (a) an interest in designing an approachable and informative means for scouring potential colleges when applying for undergraduate or graduate schools and (b) to develop a measure and means of comparison between schools and departments based on “academic progress” and other measures of achievements such as success in sports etc.

2.3 Project Objectives

Provide the primary questions you are trying to answer with your visualization. What would you like to learn and accomplish? List the benefits. The backbone of the project is a visualization which allows the user to easily understand and compare numerous college statistics such as tuition cost, ranking, average salary after graduation and so on. Among others a specific college attribute of interest would be for example the number of publications produced for each school department within a year. Using rate of publications as a measure will allow the user to compare schools based on academic progress as well as see the change in academic progress between departments within a school over time. Secondary to this if the data is available we hope to incorporate other important school attributes like funding which would be essential factor for choosing a school.

In terms of utility the visualization will allow for an intuitive alternative to the what is now a tedious and difficult process of determining which schools are a best fit for a student moving into higher education. Much of the web crawling now required during the school selection process will be well contained in an approachable and information rich visualization in that users are able to display similar schools based on their criteria as well as compare specific schools by selection or school attributes of interest.

A second result of interest will be the ability to gauge the academic progress, tuition fees, acceptance rate, etc... of a school or schools over time. In this the development and future direction of school departments will be better estimated by the user.

2.4 Data

From where and how are you collecting your data? If appropriate, provide a link to your data sources.

- College Statistics (Database: <https://nces.ed.gov/ipeds/datacenter/>)
 - College names
 - College ranking
 - Cost of tuition
 - Number of Scholarships awarded with respect to number of students
 - Acceptance rate
 - Average GPA, SAT, GRE, ACT, MCAT, ect for accepted students
- Funding Attributes (Database: As above)
 - Admissions during Fall, Spring, Summer
 - Funding per department.
 - Source of funding.
- Quality Of Life Statistics (Database:<https://www.oecdregionalwellbeing.org>)
 - Quality of life for city
 - Cost of living
 - average salary

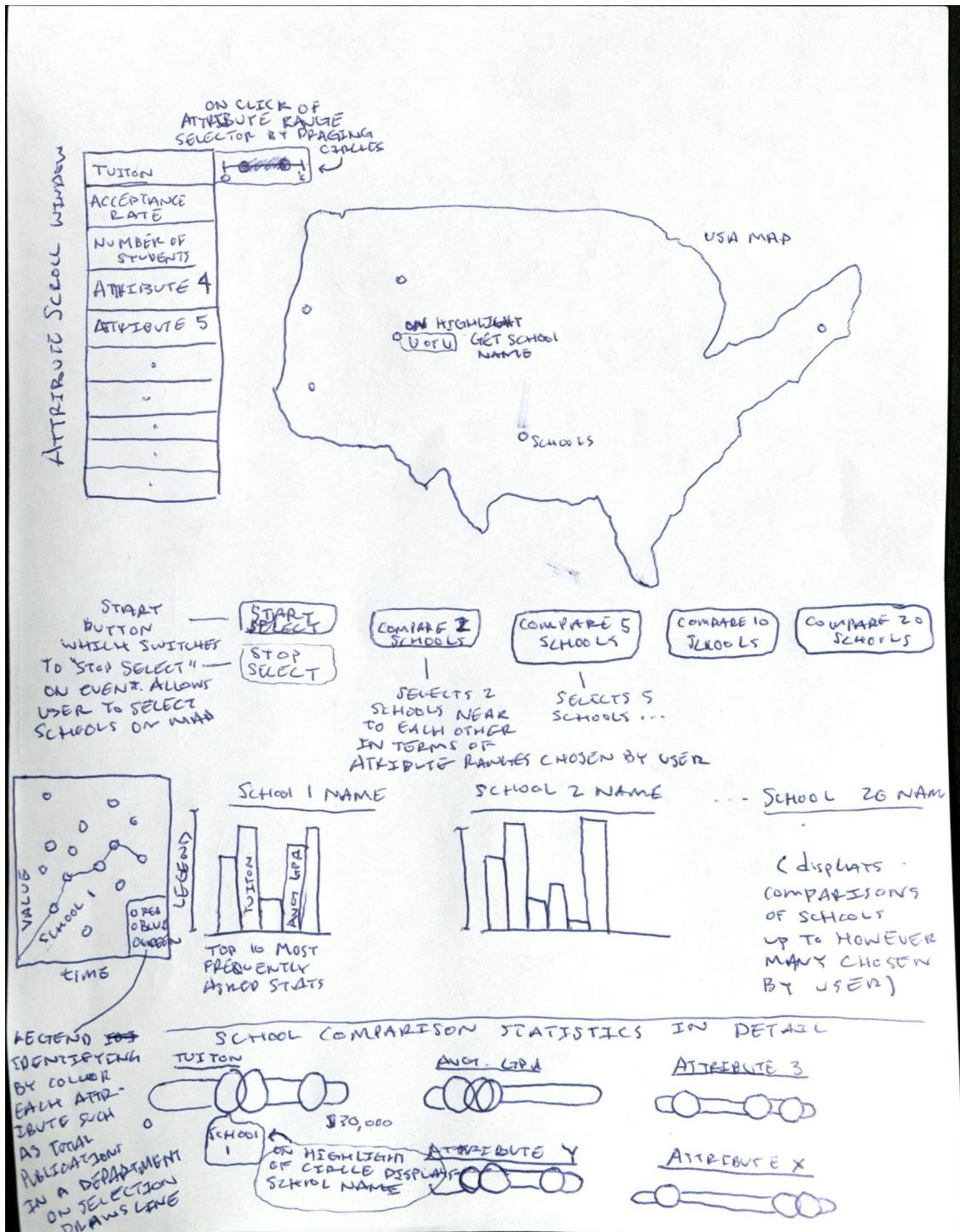
- Academics Progress Statistics (Database: <https://www.nsf.gov/statistics/data-tools.cfm>)
 - Total publications for entire school in one year
 - Total publications for each department in one year

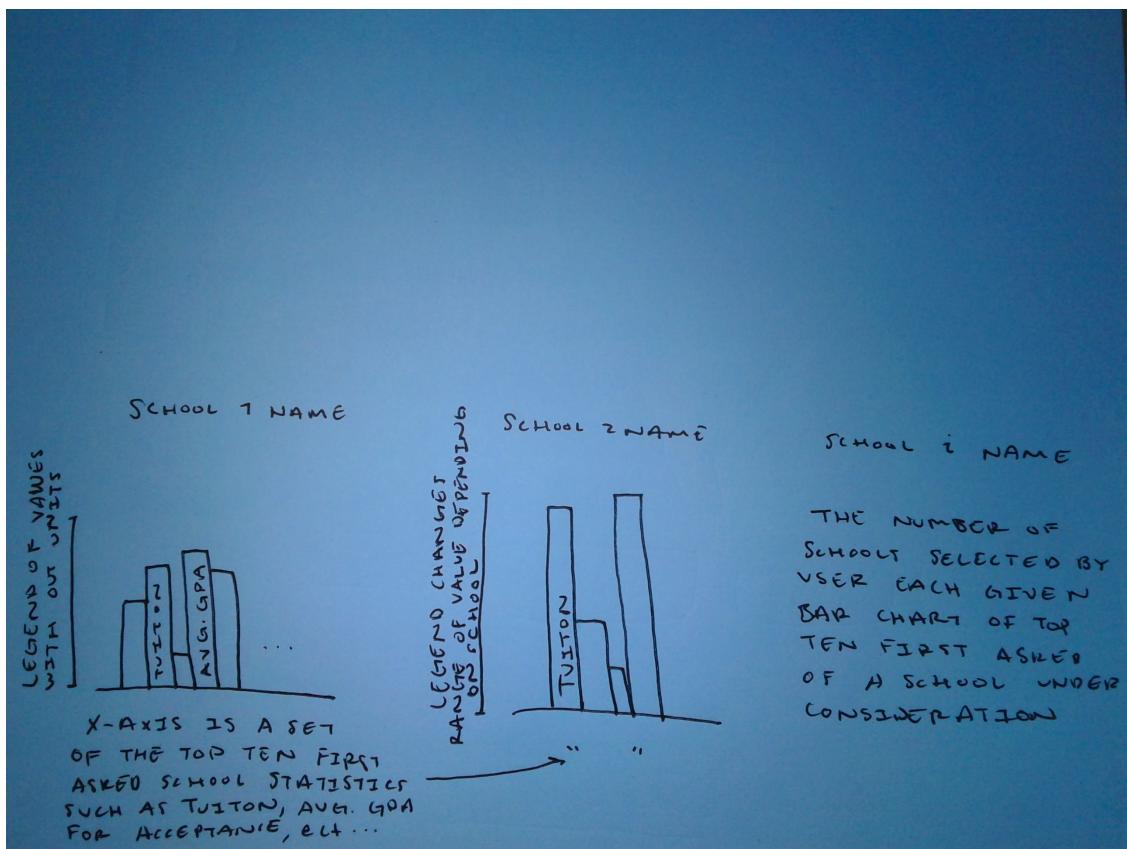
2.5 Data Processing

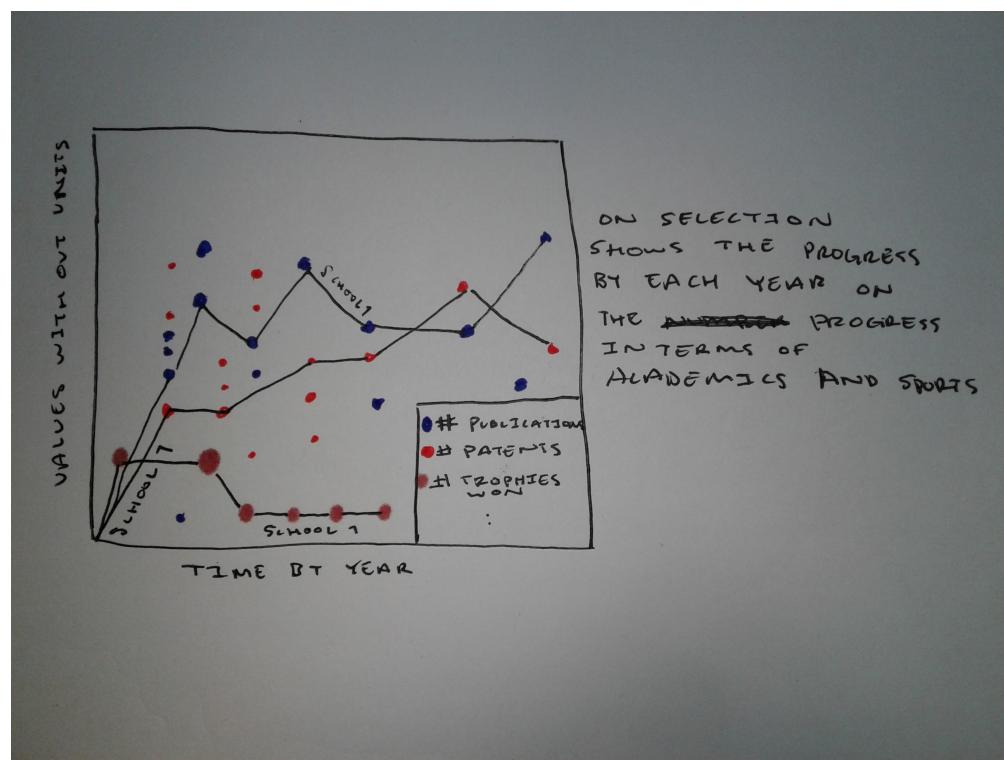
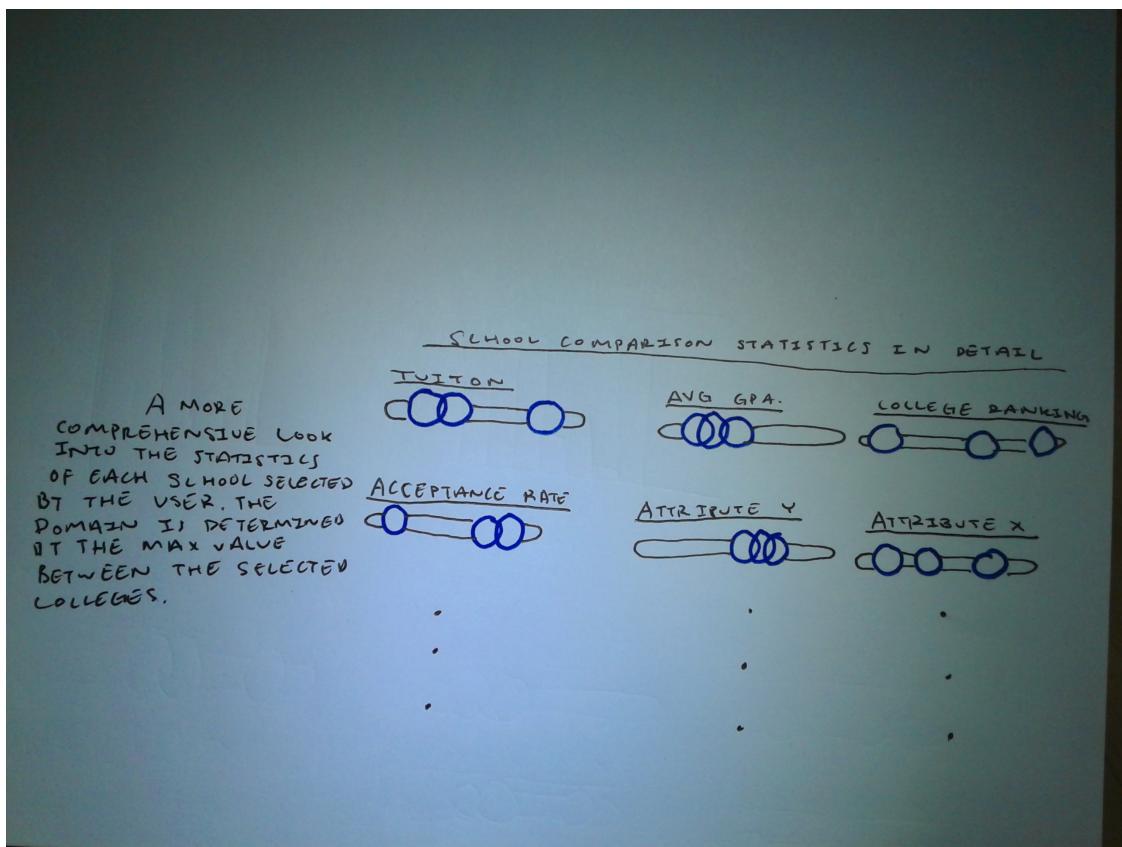
Do you expect to do substantial data cleanup? What quantities do you plan to derive from your data? How will data processing be implemented? There are near 1016 universities in the United States. In order to visualize these universities as circles on a map we may need to create a criteria on which schools will be included in the visualization. We may therefore need to do a basic parsing of the data to only incorporate schools satisfying these attributes. Second to this we may also need to data mine sites such as arxiv.org to organize the data by school, school department, and publication year. For funding we will also need to cluster the data in this way however also include source of funding and amount.

2.6 Visualization Design

How will you display your data? Provide some general ideas that you have for the visualization design. Develop three alternative prototype designs for your visualization. Create one final design that incorporates the best of your three designs. Describe your designs and justify your choices of visual encodings. We recommend you use the Five Design Sheet Methodology. The project will consist of (a) a selection bar on the left hand side where the user is able to determine which features and for what range values they would like to consider, (b) a map to it's right, (c) circled points on the map where each school is located, (d) five buttons beneath the map one of which allows the user to select specific schools to compare and the rest of which compare preset number of schools, i.e. if ten schools are selected then the ten nearest schools matching the user set criteria will be compared. Beneath will be (e) histograms demonstrating the top 10 attributes most often asked about schools such as tuition, general well being, acceptance rate, and so on. The x-axis of the histograms will be ambiguous due to the attributes not being correlated however each rectangle will be labeled as the attribute it represents. The scales of each rectangle will then be set with linear scaling whose domain is determined by the maximum values of the schools selected. And lastly, (f) beneath the histograms will be oval ranged gauges with circles representing where the selected schools are placed accordingly for all school statistics. A similar visualization can be seen on www.oecdregionalwellbeing.org. For a sketch of our intended visualization see below.







2.7 Must-Have Features

List the features without which you would consider your project to be a failure. Necessary visualizations and features include a map projection, hoverable and clickable points on the map for schools, legend with school statistics, click feature for each school statistic in the legend which opens a range selector in which one drags to points along a legend to represent what range to be considered, the option to select specific schools to compare, the option to compare a preset number of similar schools which fall into the user defined features of interest and their ranges, histograms for each selected school displaying it's specific attributes, a more detailed set of all features for each school consisting of oval bars representing an attribute's range and circles for the value of each selected school.

2.8 Optional Features

List the features which you consider to be nice to have, but not critical. The feature which we feel nice to have would be to provide a visualization similar to GapMinder by Hans Rosling which was discussed in the class. Using this type of visualization we would be able to show the change in the school's performance over time. The attributes we would consider are academic progress in terms of patents, publications, sports, endowment etc.

2.9 Project Schedule

. Make sure that you plan your work so that you can avoid a big rush right before the final project deadline, and delegate different modules and responsibilities among your team members. Write this in terms of weekly deadlines.

- Week 1 after proposal - Mine data from the sources
- Week 2 - Map projections with Universities plotted
- Week 3 - Attribute selection Legend and selection comparison
- Week 4 - Gauge visualizations for selected universities
- Week 5 - Work on GapMinder type visualization

3 Peer Feedback

Feedback Group:

- Shane Brown (u0852900)
- Sigmund Chow (u0597938)
- My Huynh (u0729654)

We found the feedback given very helpful and informative. Discussing our project illuminated aspects that were convoluted, allowing for new approaches to be either recommended by our reviewers or derived by us. Below are the suggestions or resolved confusions which we believe will provide for a more informative and intuitive visualization:

1. **Issue:** The bar charts for each school selected by the user add an extra component of confusion by grouping all top ten attributes for each school since attributes do not share the same units so height comparison is meaningless.

Solution: We have decided to instead of creating a bar chart for each school we can better preserve the main goal of school comparison by making ten bar charts for each of the “top ten first asked school attribute” within which is a bar for each of the schools selected by the user. By doing so the user will be able to see how well any of the selected schools performs with respect to the other schools.

2. **Issue:** There is a level of redundancy by adding a “Start” and “Stop” button for when a user decides to select a school on the map.

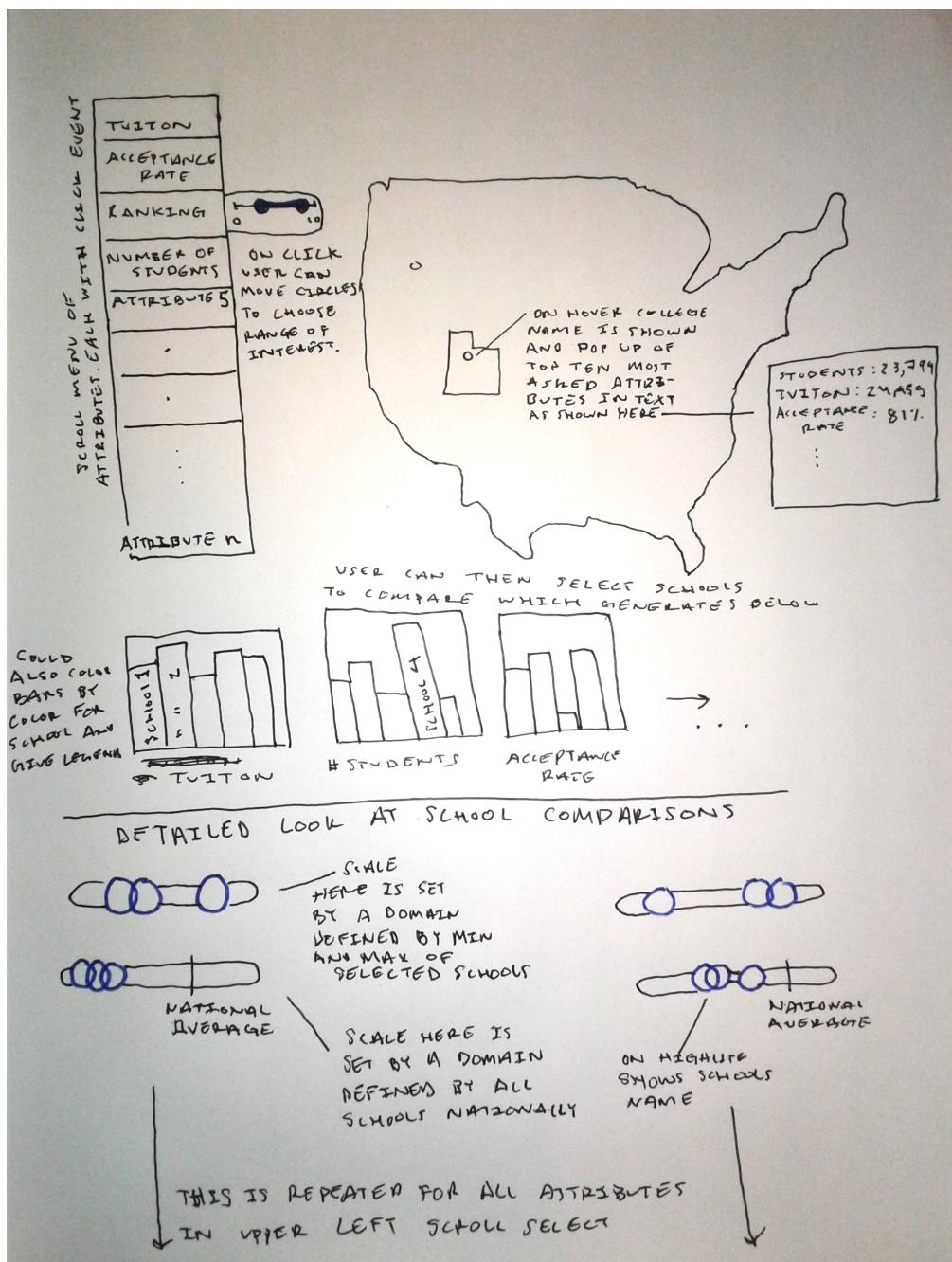
Solution: It was suggested instead the user may select a school on the map which triggers the event of having that schools histogram as well as more detailed attributes to be displayed. After selecting a number of schools the comparisons between schools will all be displayed lower. The User can then remove a school by again selecting the school on the map or selecting a small “x” icon by the schools name lower down.

3. **Issue:** After explaining the attribute comparison in which a range, defined by the max and min value of the selected schools attributes, and each of the selected schools displayed within that range as a circle it was pointed out to us that as more schools are selected it will be difficult to maintain an understanding of how a school falls into a national perspective, i.e. if the range changes for the selected schools picking top 5 worst schools might look similar to the top 5 best schools.

Solution: We can decide that not only would it be more informative but clearer to add to each attribute comparison in the “detailed” section two ranges: the first defined by the domain of the schools selected and the second defined by all schools. In this way as a user selects schools they will be able to see how a school from their selection performs with respect to the other selected schools as well as how the school compares nationally. To better illuminate this on the national scale we will also put a marker along the range indicating national average.

3.1 Other changes decided on

We no longer will include the option to compare a preset number of schools. Our incentive for this is due to the ambiguity on which schools to select from the total range of each and between all attributes. We are also considering including a protip functionality on which where rather than only displaying the schools name when mousing over its representative circle on the map it also displays in plain text the exact values of the top ten most asked questions about a school such as ranking, tuition, average acceptance rate, and so on. A sketch is given below on the direction we intend to take this project.



4 Data Processing

5 Meetings With the TA

6 Obstacles, Improvements, Added Features, and Design Changes

6.1 Added Features

While implementing our design we realized certain functionality was needed to make our visualization more dynamic and accessible. One such feature was the ability to deselect an attribute after having been selected by the brush selection. Originally a user was able to select multiple attributes in order to narrow down colleges of interest which fell into the value ranges desired. However if the user wished to no longer narrow the colleges displayed on the map for some attribute they would have to manually brush select the attributes entire range. Because of this we added the ability for the user to select again an attribute, after which that attributes range was rest to incorporate all values.

6.2 Design Changes

While developing our visualization certain design changes were made to avoid a clutter as well as allow a more immediately understandable user interface. Specifically, our original design had intended to include all data from the data collected. Including all attributes however would not be practical due to the amount of attributes each college has. Instead we therefore sided with choosing key features applicants search when selecting schools. As a result of having less attributes we therefore no longer needed a scroll feature but instead could stack the attributes below the map and allow each attribute it's own brush selection bar to determine value range.

7 To Do

- Legend indicating school and attribute chosen.
- Selected schools displayed in attribute range scale
- Aesthetic touches
- Comment Code
- Zoom functionality to map with counties