COLLEGE COMPARATOR: COMPARISON BETWEEN UNDERGRADUATE COLLEGES IN THE US

Team:

	Name	WPI Email Id	Github Id
1.	Neha Mahajan	nsmahajan@wpi.edu	nsmahajan
2.	Kirthika Muthukumaran	kmuthukumaran@wpi.edu	kirthim13
3.	Kartik Shetty	kshetty@wpi.edu	arsene2803

Project Repository:

https://github.com/nsmahajan/cs573FinalProject

Overview and Motivation

Before taking admission into a college, students tend to do a lot of research in the internet to get information about universities/colleges. There are different factors that students tend to compare about a college/university such as the number of students accepted a year, tuition fees for an academic year and various degrees offered by a college. Finding such data requires one to look at various websites.

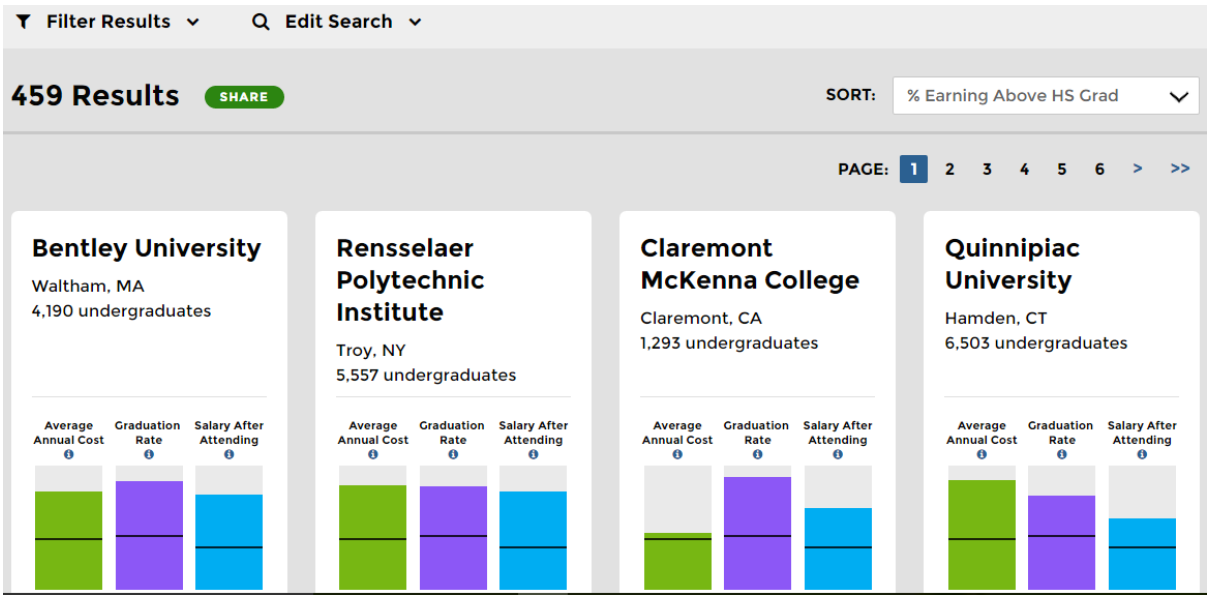
In this project we aim to provide a single destination which lets students to compare different colleges/universities across the US. In specific, this site will allow students to compare in/out state tuition fees, average earnings, degrees offered by a college and percentage of students who have gotten funding. By comparing different degrees offered by a college, one can understand the popularity of a particular degree in a college.

WEEK 1

Related Work

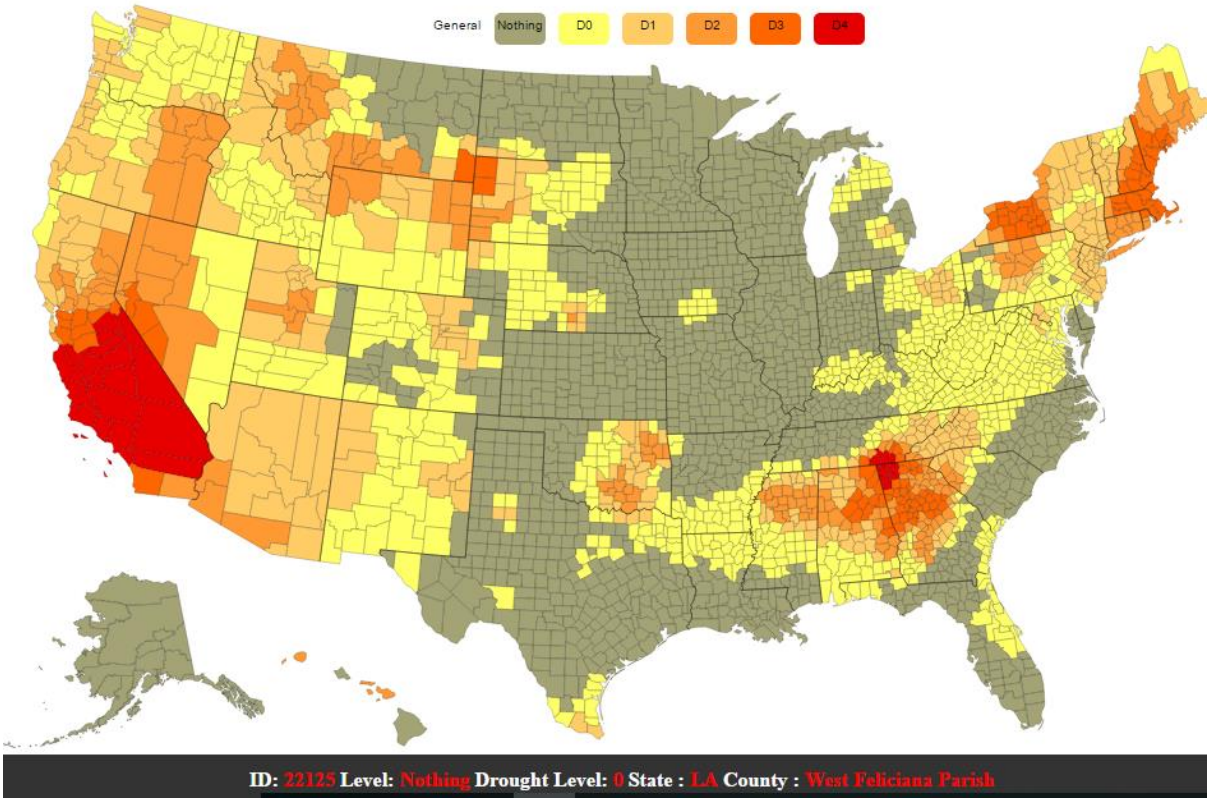
When we were searching for datasets in different resources, we found a dataset related to undergraduate colleges, specifically college scorecard dataset. Few sources that inspired our design are below.

1. https://collegescorecard.ed.gov

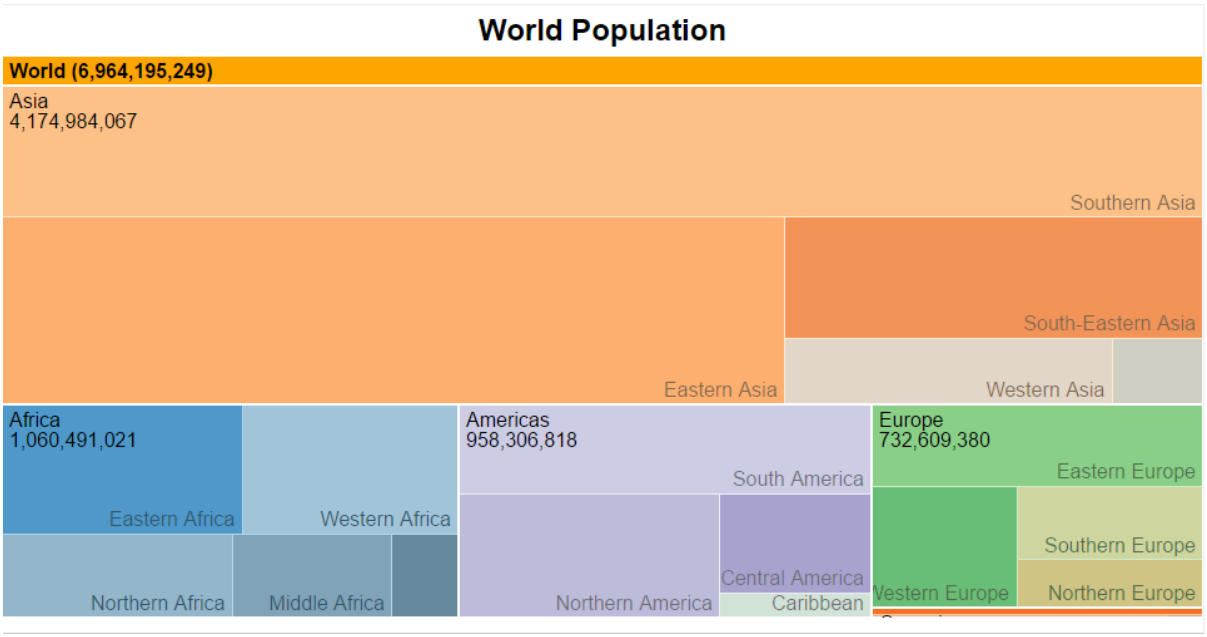


The website on College scorecard inspired us in deciding on the different factors about a college that a student would be interested in comparing.

2. Lab assignment on maps that we did in class inspired us to use choropleth map to understand the distribution of colleges across different states in the US.



3. When we were working on Assignment 4, we came across this zoomable tree map on world population.



We wanted to incorporate this in our design. With zoomable tree map we wanted to understand popular degrees in a particular college/university.

Questions

To decide on what factors about a college are to be compared, we took help from several websites that had a list of factors which influence a student's decision. After a brief research, we decided to answer the following questions about a college.

1. How the undergraduate colleges in the US are distributed across states?
2. What are the public and private (profit/non-profit) undergraduate colleges available in a particular city?
3. What is the annual cost of attendance for a particular college?
4. What degrees are offered in a college?
5. What is the percentage of students getting financial assistance?

Data

Source and Link – US Department of Education (https://collegescorecard.ed.gov/data/)

The data is available in the .csv format. We have data for about 7704 colleges with more than 1800 unique columns. The data is available from the year 1996 to 2015. For our project we are using the year 2014-2015 and to answer the queries we have filtered the required columns using Java and combined the data into a new single master csv file.

Exploratory Data Analysis:

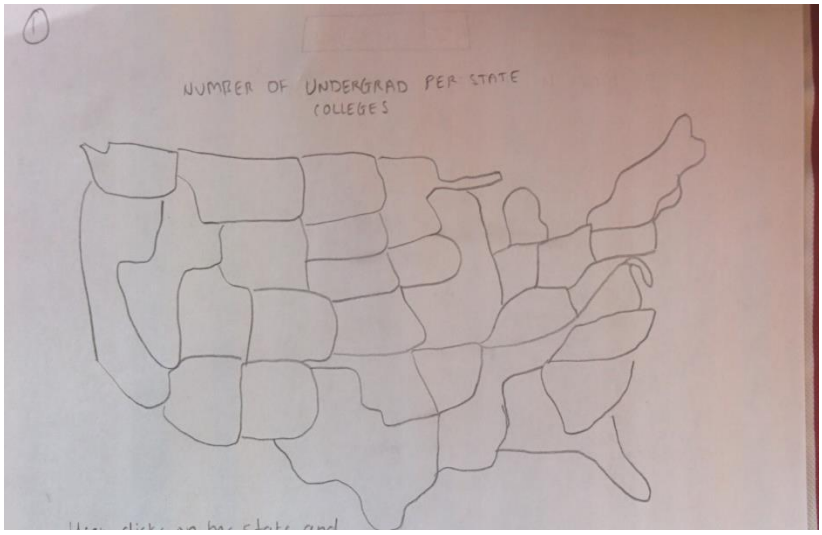
To visualize the data for displaying the in-state and out-state tuition fee the initial design was to use parallel co-ordinates chart and while working on the visualization we found that for public colleges the in-state and out-state tuition fee was same, making the data overlap. So we decided on the multi-series bar chart. We used Excel and d3.js for initial data visualization.

Design Evolution:

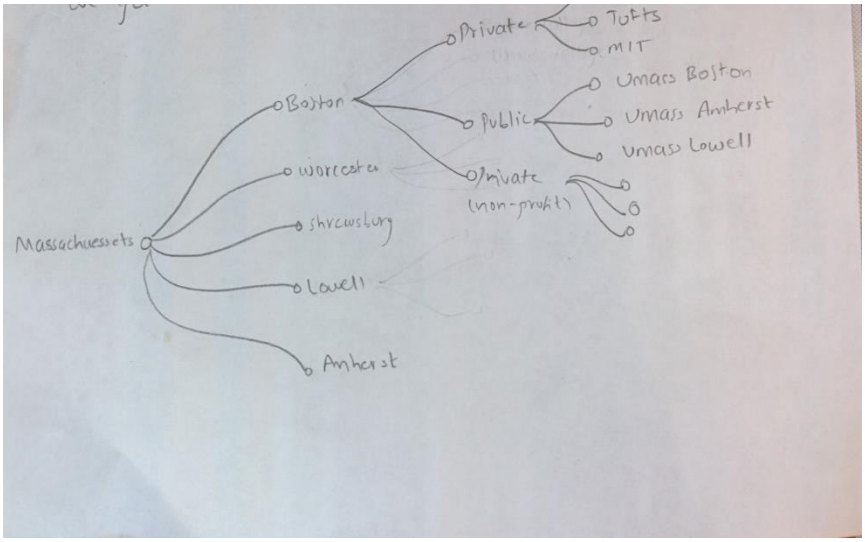
Before starting with the actual coding, we planned layouts using hand sketching to help us understand what visualization we would require and how to connect the data between the layouts.

Idea 1:

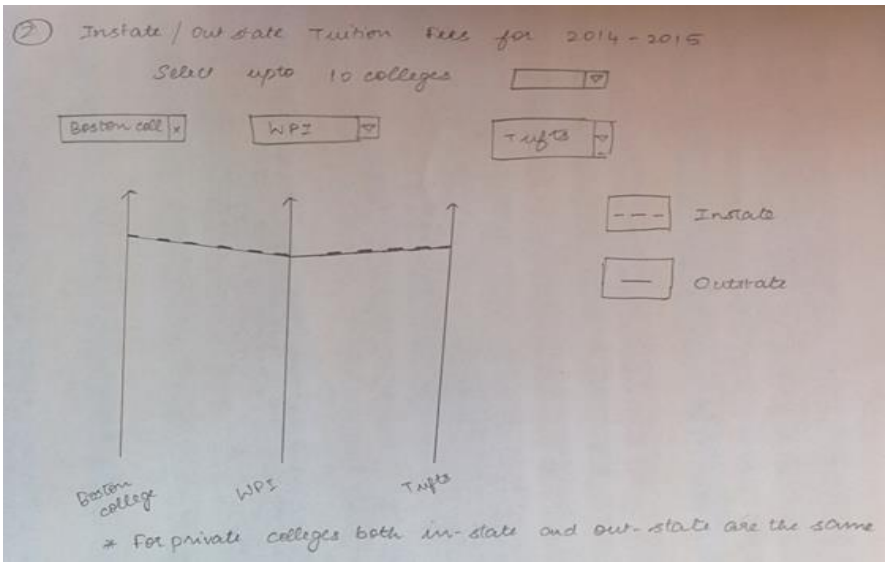
1. **Distribution of undergraduate colleges across the US, state wise**



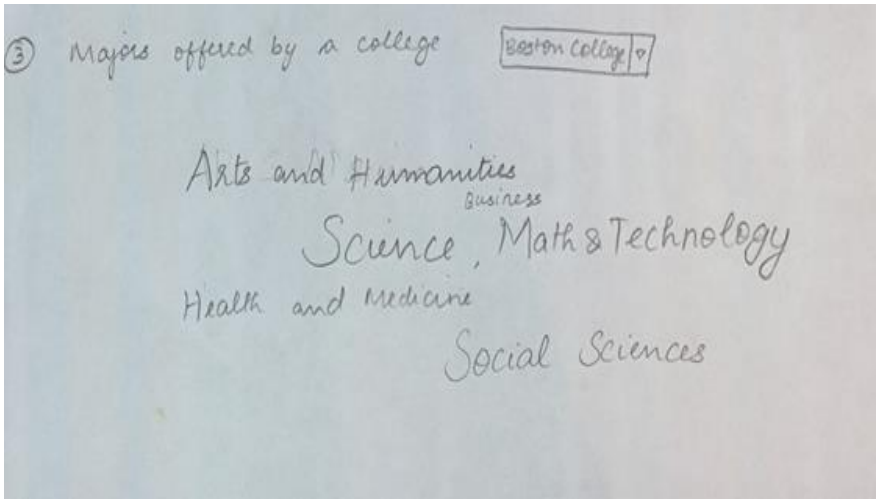
2. **What are the undergraduate colleges available in a particular state/city?**



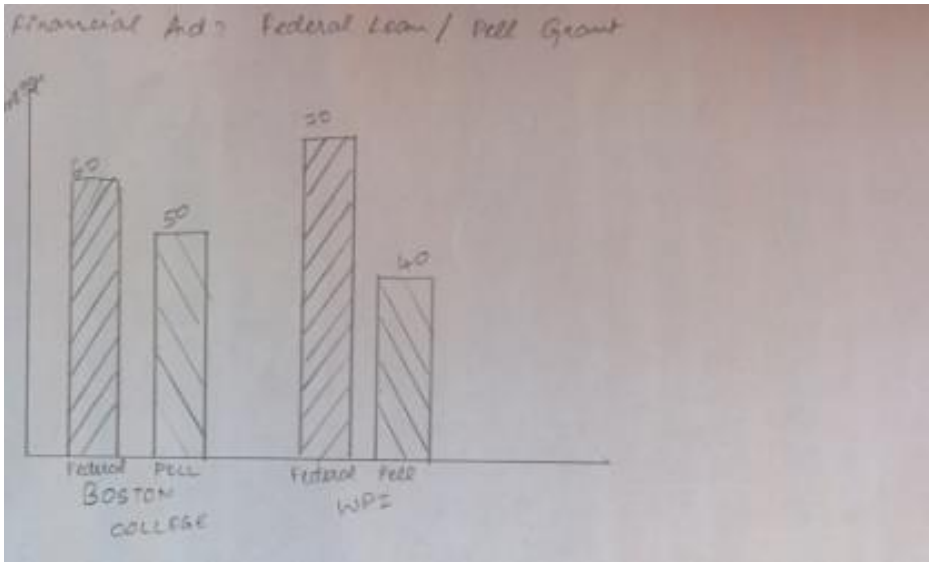
3. **Instate/ Outstate tuition fees**



4. **Majors offered by a college**



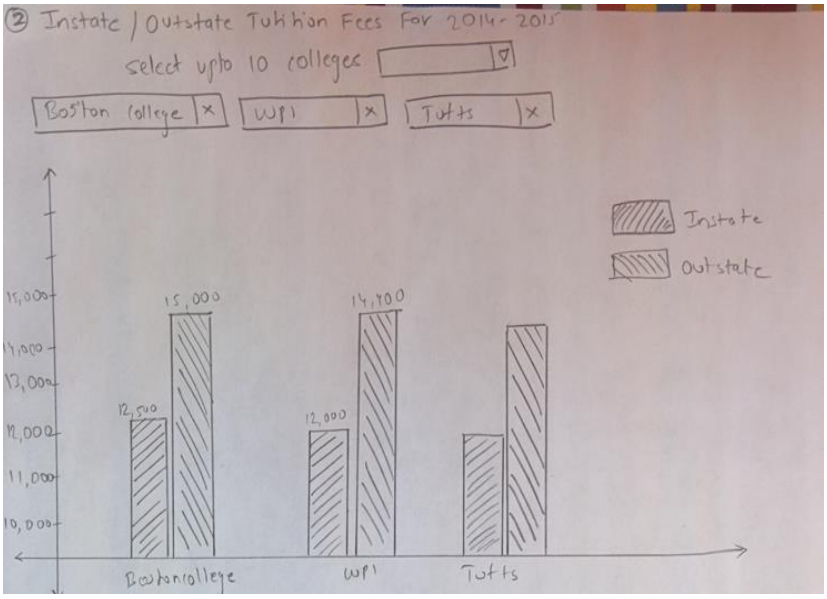
5. **Percentage of students who got Federal loan/ Pell grant**



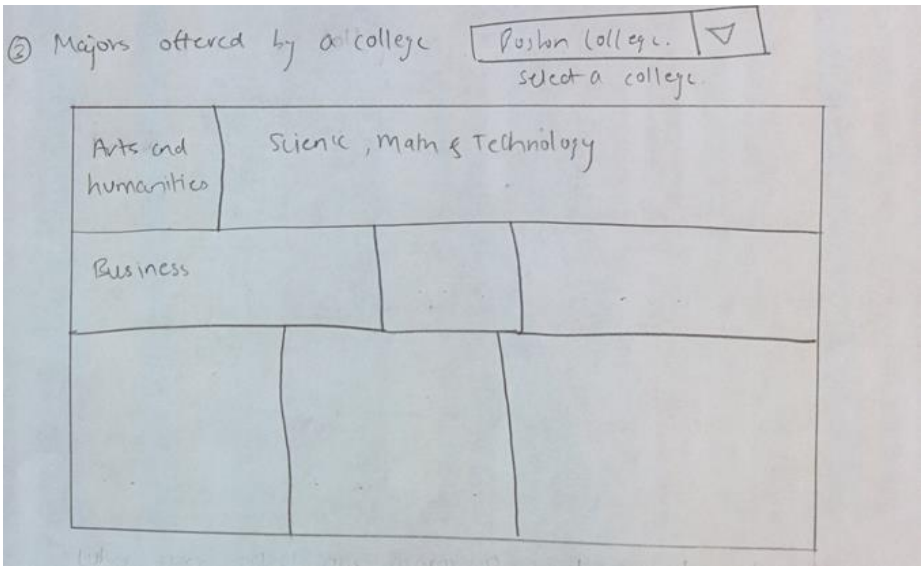
In the first idea we have separate charts for each type of question. The first query is visualized as a choropleth map in which each state will be color coded as per the total number of undergraduate colleges. A collapsible tree map that shows the list of colleges in a state categorized by city name, also categorized by private (non- profit or profit) and public. To visualize the third query we used slope chart in which we have parallel co-ordinate for each college. Fourth query, which is the majors offered in a particular college is visualized as 'Word cloud' in which the font size for a major depends on the number of students graduated in that major. For the last query we used multi-series bar chart with series1 as Percentage of students who received Federal Loan and series 2 percentage of students who received Pell Grant.

Idea 2:

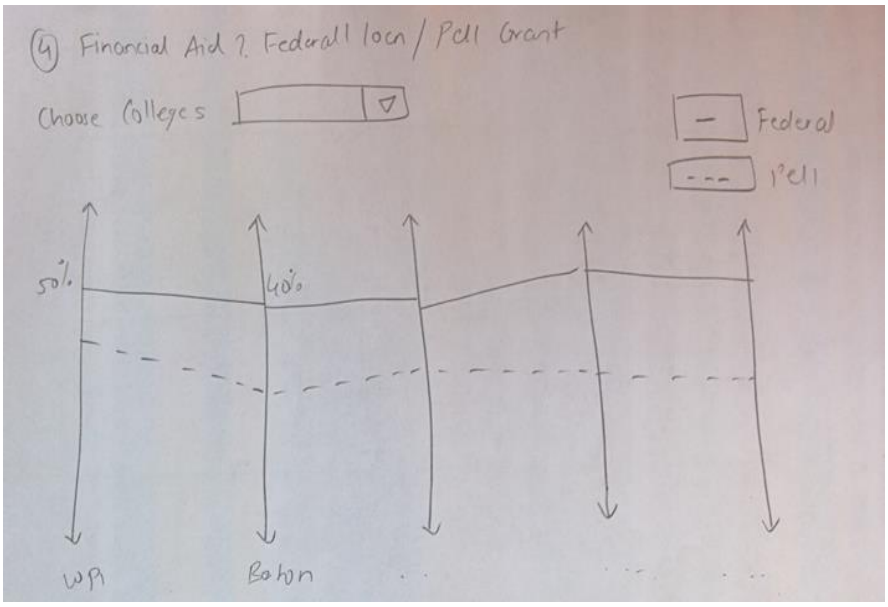
1. **Instate/ Outstate tuition fees**



2. **Majors offered by a college**



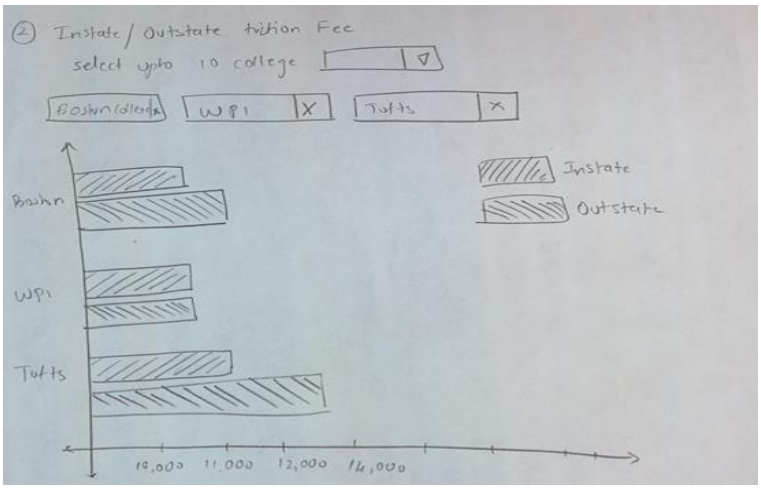
3. **Percentage of students who got Federal loan/ Pell grant**



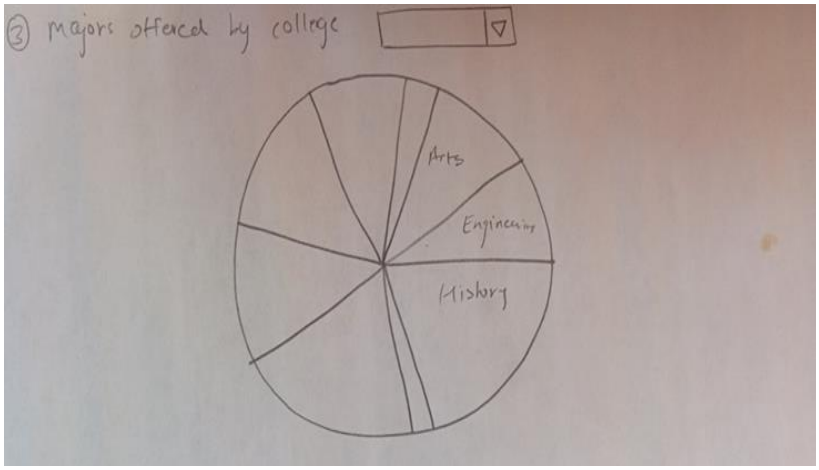
To visualize first and second query, we used the same visualizations from Idea 1. To visualize the third query we used multi-series bar chart with series 1 as instate tuition fees and series 2 as out state tuition fees. Fourth query, which is the majors offered in a particular college, is visualized as a 'tree map' block size for a major depends on the number of students graduated in that major. For the last query we used slope chart with parallel co-ordinates as colleges and two series. Here, series1 is percentage of students who received Federal Loan and series 2 is percentage of students who received Pell Grant.

Idea 3:

1. **Instate/ Outstate tuition fees**



2. **Majors offered by a college**



To visualize the third query we used horizontal multi-series bar chart with series 1 as instate tuition fees and series 2 as out state tuition fees. Fourth query, which is the majors offered in a particular college, is visualized as a ‘pie chart’ in which sector size for a major depends on the number of students graduated in that major.

<u>Final Design</u>

For our final design we chose the following visualizations for each query.

1. **What is the distribution of undergraduate colleges across the US?**
   We chose to visualize this data in a choropleth map because we want to represent total number of undergraduate colleges in all the states of the US. This allows the user to understand the distribution of colleges across the country.
2. **What are the undergraduate colleges available in a particular state/city?**
   For this query we plan to use collapsible tree map, in which the root will be the state name followed by city names and for each city we will have three categories such as Private(Non-Profit), Private(Profit) and Public. Tree maps are good to perceive hierarchical data.
3. **What is the annual cost of attendance?**
   For this query, we chose to go with multi-series vertical bar chart instead of slope chart because, some private colleges have same tuition for both in-state and out-state there will be overlapping in the slope chart for those cases. Whereas bar chart will not have any overlapping, making it easier to do comparison.
4. **What degrees are offered in a college?**

   We chose to use 'tree map' which is good way to do comparison between percentages of students graduated from a major for multiple majors just in a single chart. Whereas, word cloud is difficult to do comparison between majors and also it is difficult to perceive. Also pie charts will not be useful here because the sector size can become so small that some major's sectors can be overlapping.

WEEK 2

Questions

After receiving feedback about Proposal, we wanted to make minor changes to our questions. We merged distribution of colleges and what are the public, private (profit/non-profit) colleges into a single query.

So the finalized list of questions that we wanted to address is

1. How the undergraduate colleges in the US are distributed across states?
2. What is annual cost of attendance in terms of in-state and out state tuition fees?
3. What is the average income of students from a college after graduating?
4. What are the popular degrees in a college?
5. What is the percentage of students getting financial assistance?

Data

After talking with few friends, another problem we could define was that the salary a student may earn after the graduation also plays an important role to make a college decision. We decided to include another csv file which provided the data related to median earnings for a student after graduation and combined it with the original master csv file using Java.

Implementation

After submitting the final project proposal, since we had an idea what charts would be required we started working on designing the charts in d3.js.All the charts were developed as individual problems and data was subdivided.

WEEK 3

Design Evolution:

After getting feedback from proposal submission, instead of having a list of all colleges implement some filters that will help the student to narrow down the list of colleges and remove non-user friendly visualizations like collapsible tree map, we had a new design decision to make.

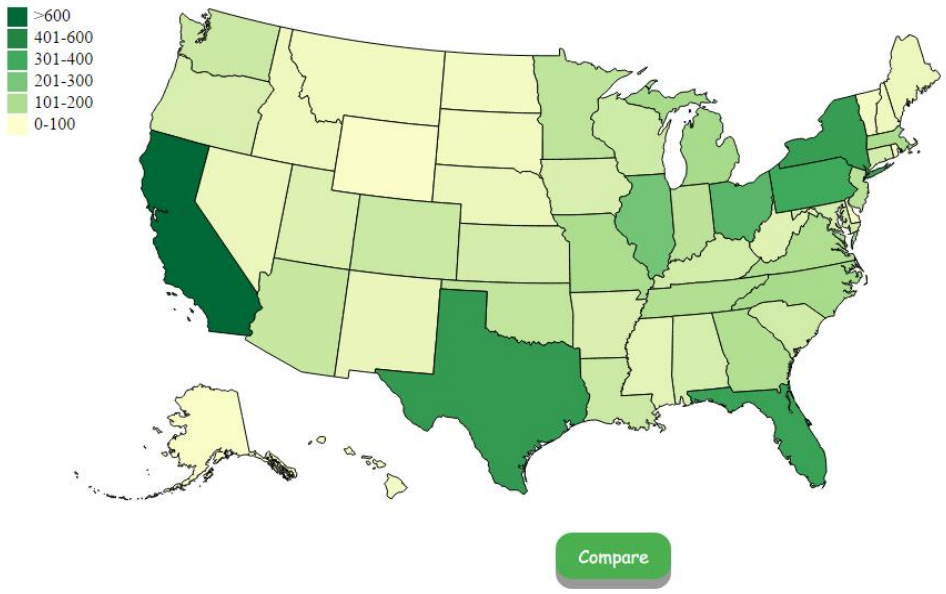
Also since the project was not just about data visualizations but as a whole a website which can tell a story, we made some new changes by adding filtering and tabs to view different charts for each problem type. The new design was implemented using jQuery and d3.js.

Implementation:

The new website design has a landing page where the user can see the distribution of colleges across states. When the user is ready to compare he/she is directed to a new page where the user can select up to 10 colleges for comparison. Filters like state, student population and type of school are provided to narrow down the list of colleges. Each of the charts is implemented as separate tab, making the data between the charts independent and hence making it more scalable. In future adding a new data and a new visualization can be done easily.

The below image is the landing page where each state is color coded according to the distribution of colleges.

Visualization 1

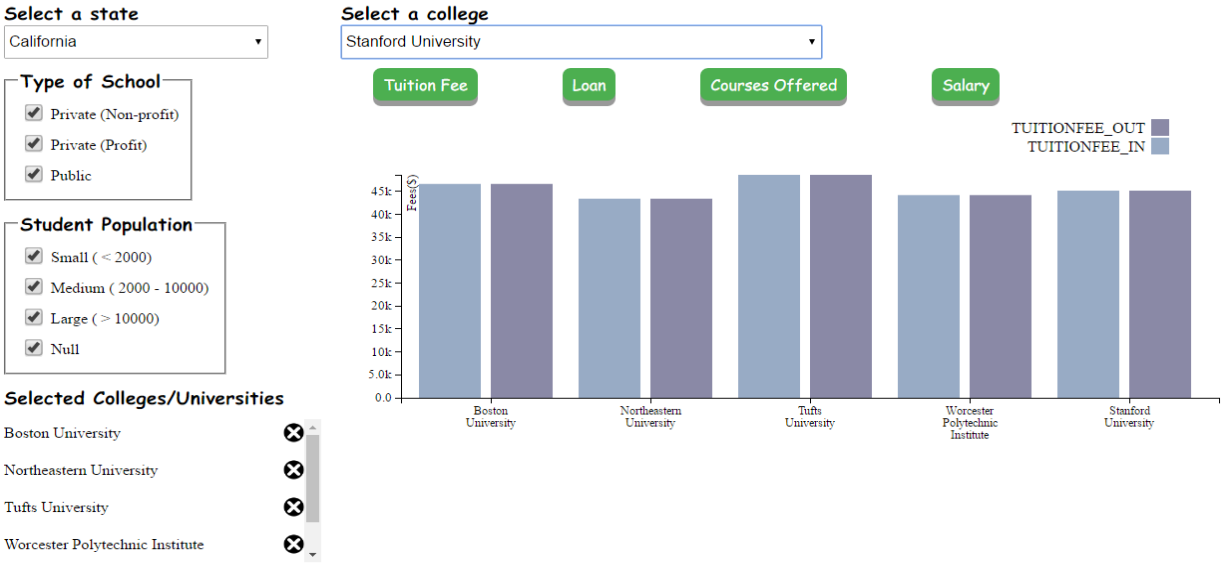


Once the user clicks on the compare page, the user is directed to a new page where he/she can select 10 colleges. The college list updates when you apply filters like State, Population distribution and type of college. Population distribution and type of college are implemented as checkboxes rather than a radio box since multiple values can be selected for each of them.

By default, when the page loads, the state is set to “Massachusetts” and all the checkboxes are selected. The default chart that is loaded is the tuition fees and user can select other tabs to view the data accordingly.

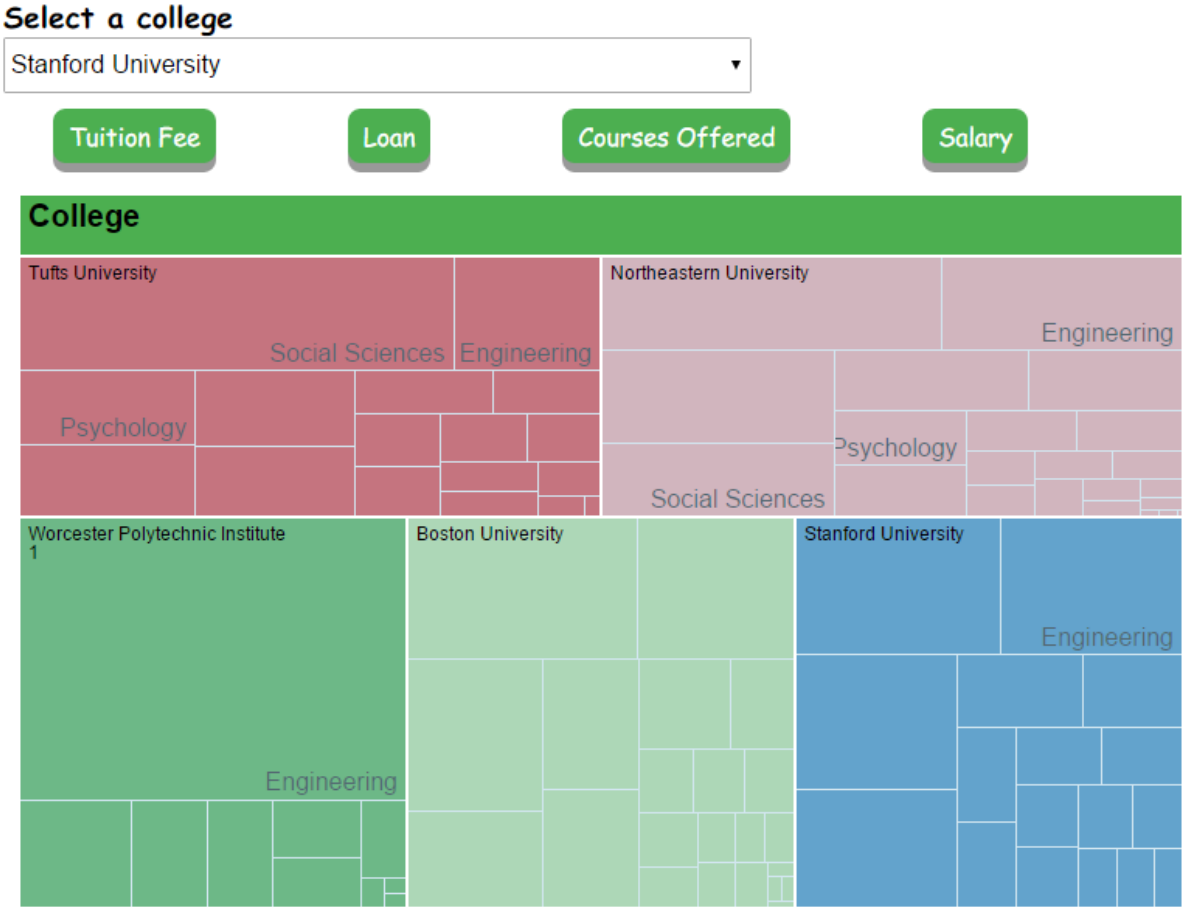
Below are the charts for each factor (Tuition fees, Loan, Courses, Earnings)

Visualization 2

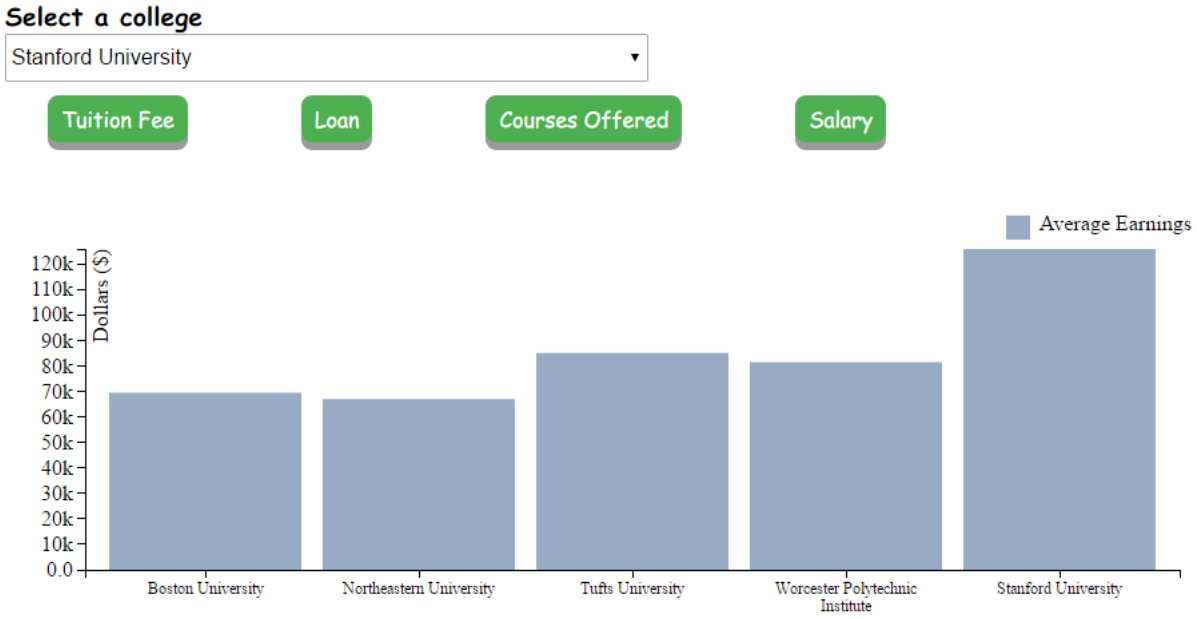


Visualization 3

Visualization 4



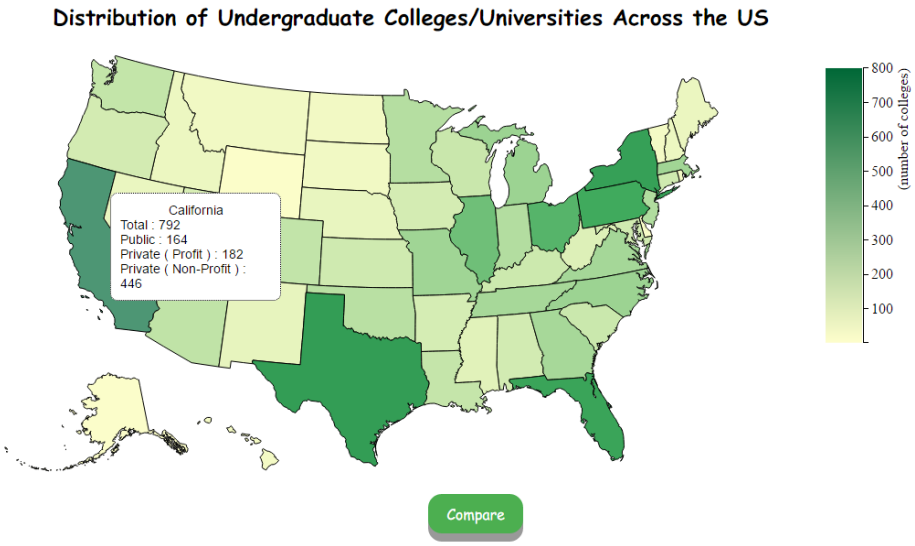
Visualization 5



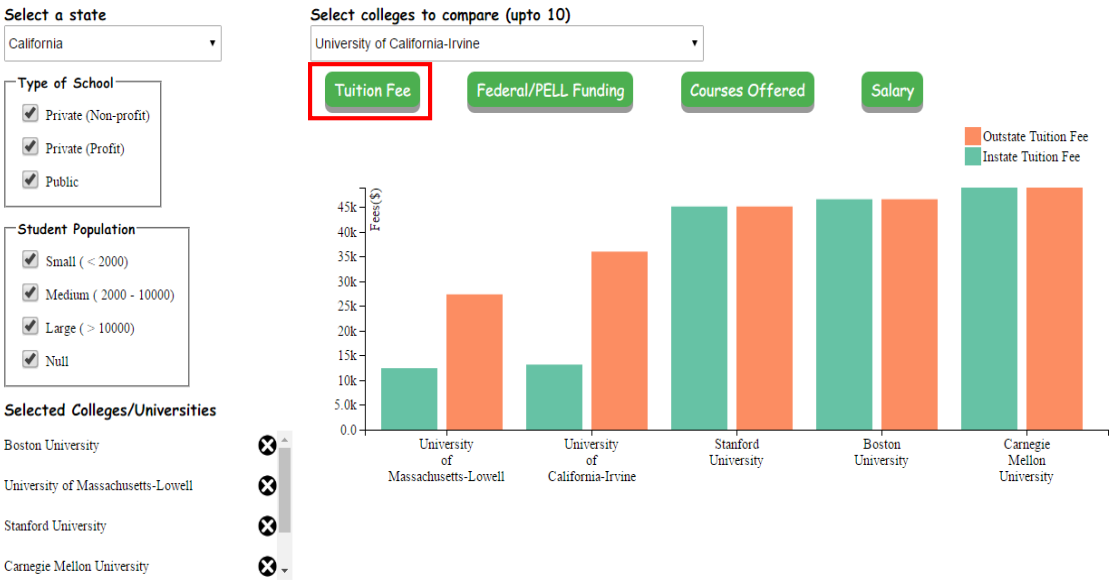
WEEK 4

This week, we added interactions to the already built visualizations. Few other changes were also made to the visualizations in areas such as color encoding, displaying the bar charts in sorted order. The added interactions display relevant information of college/university with which user can understand the visualizations better. The final visualizations are as follows,

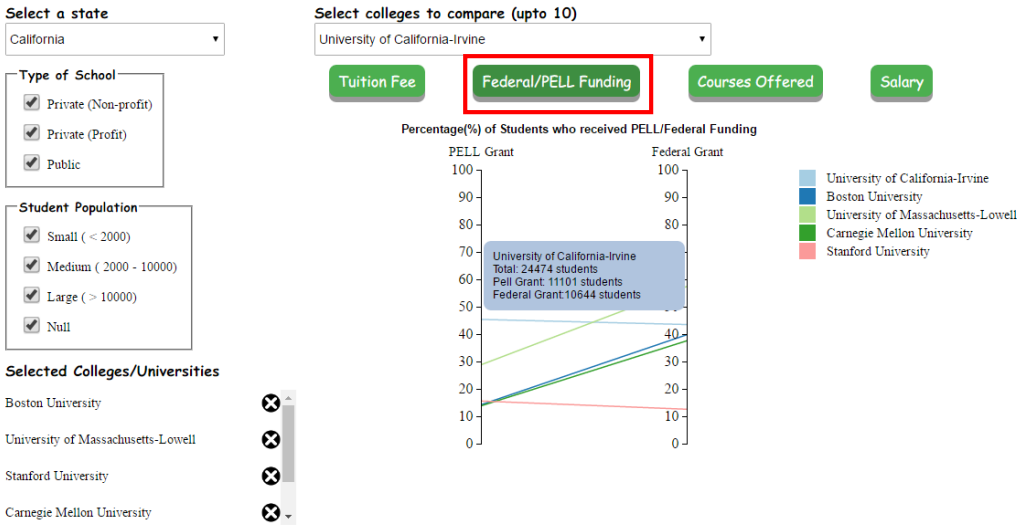
Visualization 1



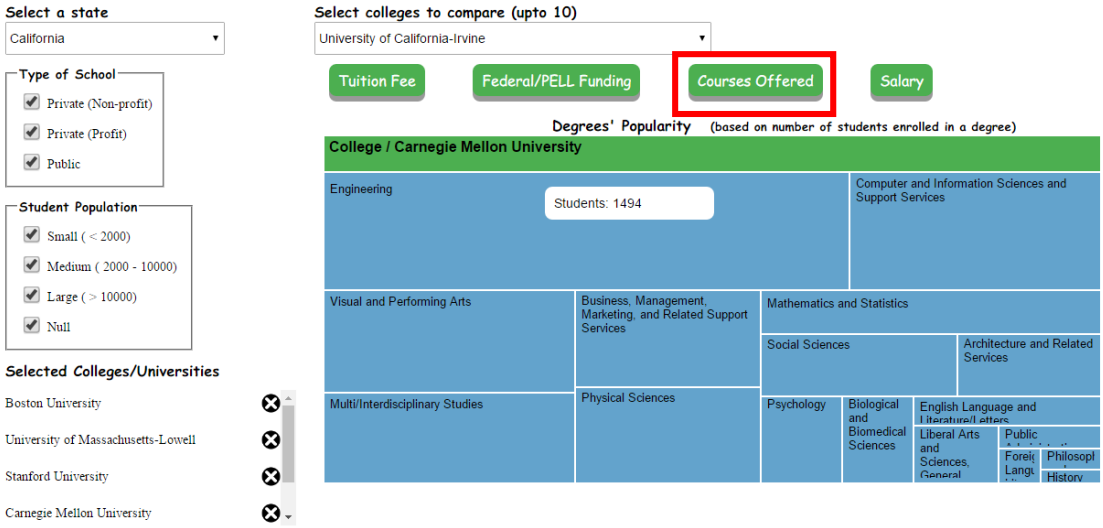
Visualization 2



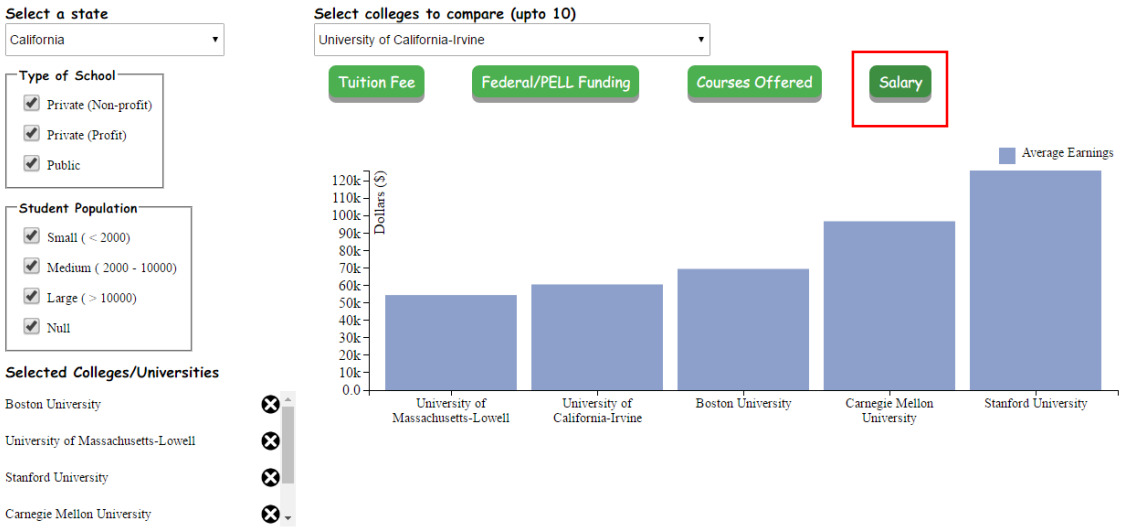
Visualization 3



Visualization 4



Visualization 5



WEEK 5

After the prototype presentation and the feedback received from the classmates and professor we decided to do 2 major changes in the websites and changed the problem statements.

What was suggested in the feedback?

1. Instead of having buttons to view charts for comparison between colleges, display all the charts one below the other so that student can see all data at same time.
2. Instead of just having comparison between 10 colleges, can data be shown for all colleges to allow comparison within a state or country?
3. Remove the college selection dropdown.
4. The data related to the popularity of the courses offered in a college was not accurate.

The final questions that our website will provide answers to are as follows:

1. How the undergraduate colleges in the US are distributed across states?
2. What are the public and private (profit/non-profit) undergraduate colleges available in a particular city?
3. What is the annual cost of attendance for a particular college?
4. What is the average salary a student might expect after graduation?
5. What is the percentage of students getting financial assistance?
6. How the colleges salary, admission rate, in state and outstate tuition fee's changes when compared to national average?

We also made UI changes as per the feedback and instead of having a dropdown to select a college we displayed data for all colleges in a tabular format and to show how salary, admission rate, in state and outstate tuition fee's changes when compared to national average we using in-line charts in the table column. The student can also sort the columns on name, salary, population, instate/outstate tuition fee and use school type checkboxes to allow filtering of

tabular data. We changed the fonts and made them more user friendly. Introduced a landing page so that user can navigate to the main website, screen cast, data source and process book

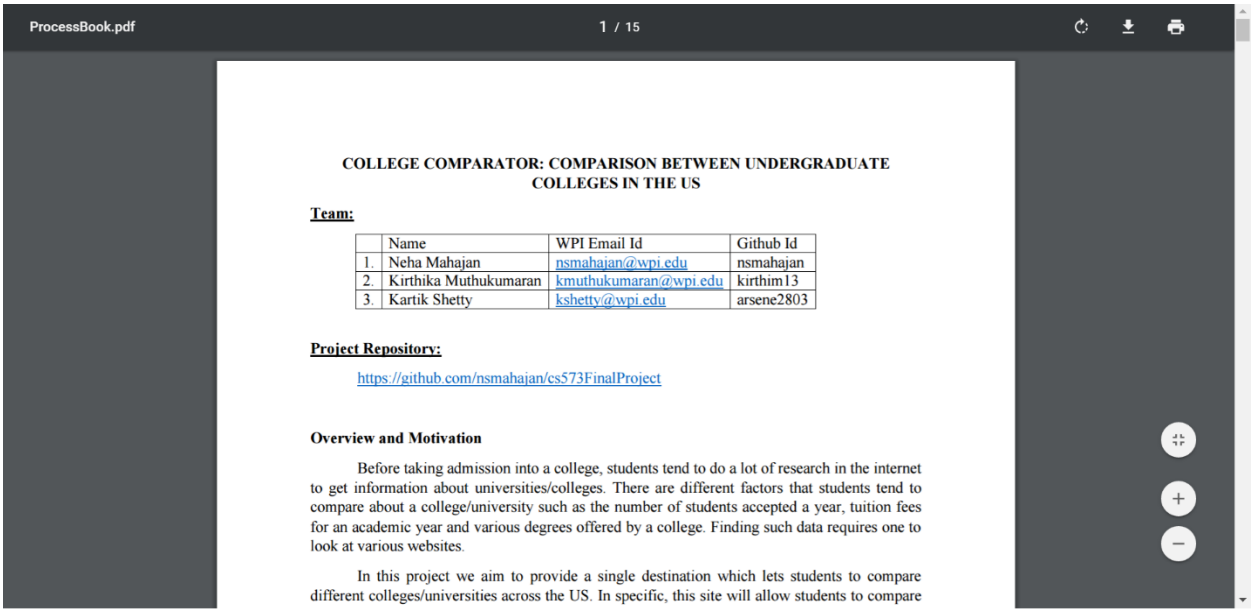
Visualization 1

Across US there are more than 7000 undergraduate colleges and in 2014, the total undergraduate enrollment in degree-granting postsecondary institutions was 17.3 million students. Undergraduate enrollment is projected to increase 14 percent from 17.3 million to 19.8 million students between 2014 and 2025.

Many factors affect the student's decision and this decision making process can be very hectic as it requires to visit various websites to get information on interested colleges. Our goal is to present a student a comparison tool which will help him/her to compare colleges based on factors like tuition fees, popular courses offered, percentage of students receiving loan, salary after graduation etc.

Process Book Screencast College Comparator Data Source

Visualization 2


ProcessBook.pdf 1 / 15

COLLEGE COMPARATOR: COMPARISON BETWEEN UNDERGRADUATE COLLEGES IN THE US

Team:

	Name	WPI Email Id	Github Id
1.	Neha Mahajan	nsmahajan@wpi.edu	nsmahajan
2.	Kirthika Muthukumaran	kmuthukumaran@wpi.edu	kirthim13
3.	Kartik Shetty	kshetty@wpi.edu	arsene2803

Project Repository:

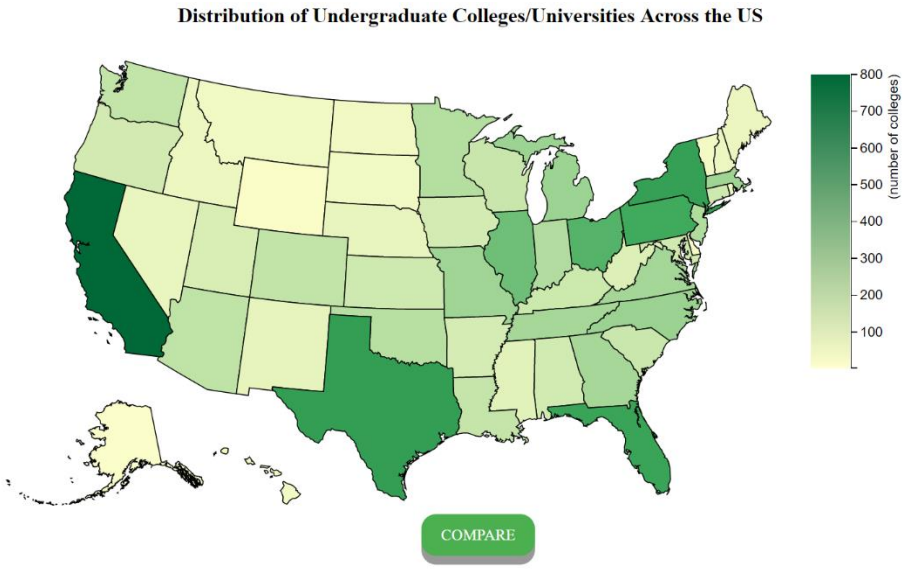
https://github.com/nsmahajan/cs573FinalProject

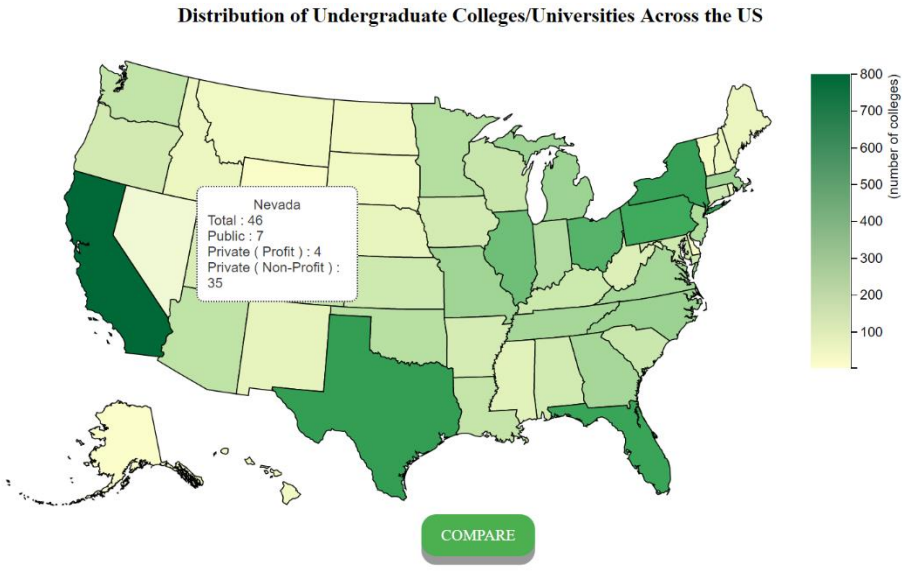
Overview and Motivation

Before taking admission into a college, students tend to do a lot of research in the internet to get information about universities/colleges. There are different factors that students tend to compare about a college/university such as the number of students accepted a year, tuition fees for an academic year and various degrees offered by a college. Finding such data requires one to look at various websites.

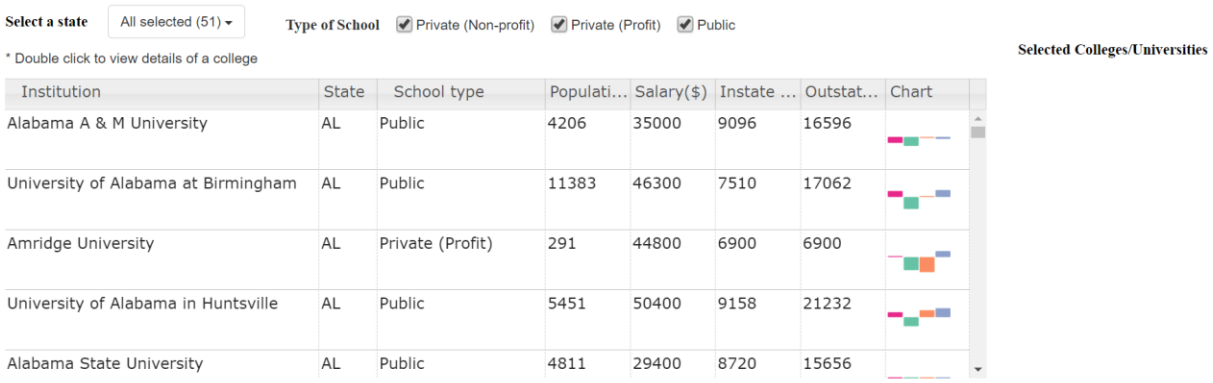
In this project we aim to provide a single destination which lets students to compare different colleges/universities across the US. In specific, this site will allow students to compare

Visualization 3

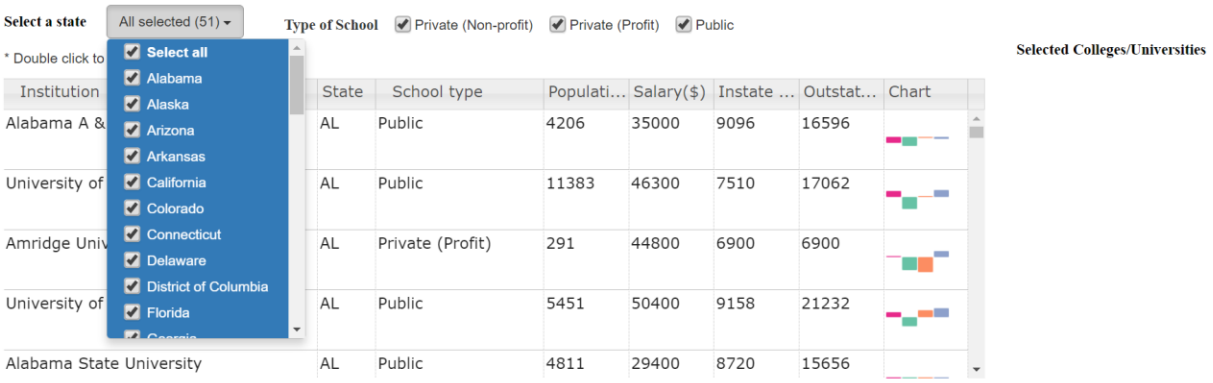




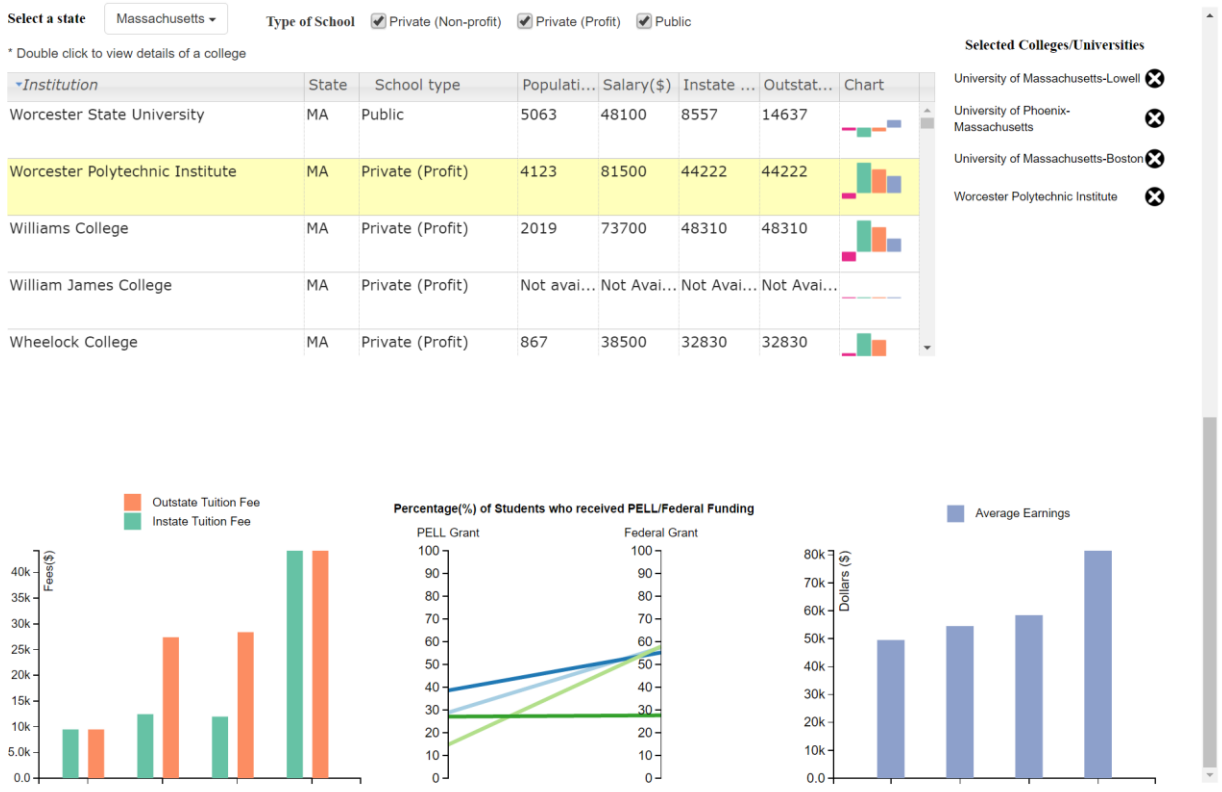
Visualization 4



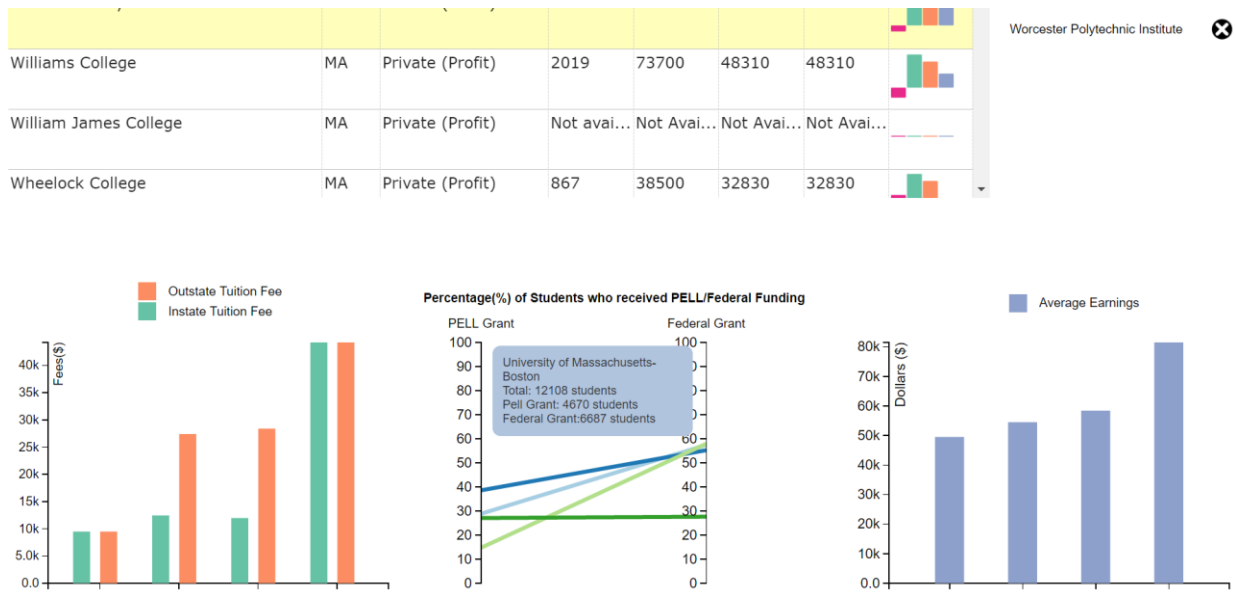
Visualization 5



Visualization 6



Visualization 7



References:

1. http://omnipotent.net/jquery.sparkline/#s-about
2. https://mleibman.github.io/SlickGrid/examples/example10-async-post-render.html
3. https://bl.ocks.org/mbostock/3887051
4. https://bl.ocks.org/jasondavies/1341281