John Perez and Sydney Trout
Dr. Iuricich
CPSC 4030
15 December 2022

Final Delivery Process Book

Important Information:

Title: Clemson Grade Distribution Visualization

Team Members:

- John Perez C11045966 jjperez@clemson.edu
- Sydney Trout C20805591 sltrout@clemson.edu

Project Repository: https://github.com/sydneytrout/DataVisProjectSite

Project Site: https://sydneytrout.github.io/DataVisProjectSite/

Video Link: https://youtu.be/-w3IfmHXBkM

Overview and Motivation:

Every college student is familiar with registration season. A face off between you and other students with better registration times, all fighting to get a seat in your desired course. For many, this game involves immense research, diving deep into RateMyProfessor, reaching out to older students, and looking through previous semester's grade distributions. Clemson University grants their students and faculty access to grade distributions from every semester dating back to Fall of 2013. However, these PDFs are very dense, all falling around 150 pages in length.

As we approach our graduation in December, we both have a great deal of experience navigating these files. Aside from trying to coordinate our own schedules, we have both been captivated by some of the data that lies within them. You hear rumors of how difficult Mechanical Engineering Statics and Dynamics is, but until you see the 1% A rate, you don't really believe it. So when we were presented with the guidelines for this project, this dataset seemed like the perfect choice for us, as it is something we were both very familiar with and also intrigued by. We hoped that through our project we would be able to turn these dense PDFs into something visually aiding to students in our same boat. Rather than flipping between twenty different files, our goal was to have students use our site as a guide to crafting their perfect semester.

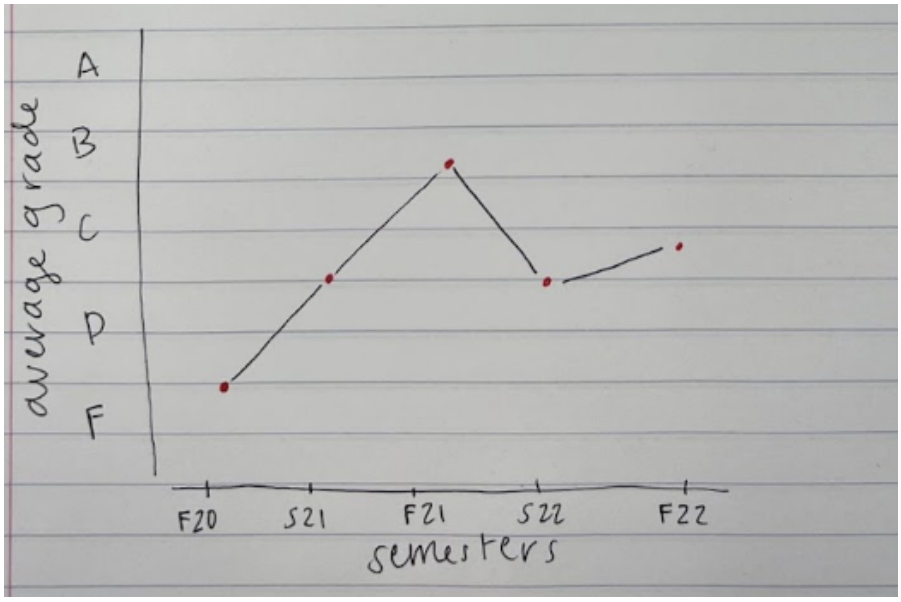
Related Work:

The inspiration for our dataset stemmed from our personal usage of Clemson's Grade Distribution Reports. It was something that we both relied on throughout college registration season, and something that sparked questions for both of us. During our initial prototype creation process, we drew a lot of our inspiration from examples that we examined in class. Specifically, there was a period of time where we were going to use a visualization similar to the Game of Thrones one we analyzed at the beginning of the semester. While we did not end up using this type of chart, it was material like this that impacted our final design choice. Many of the interaction elements also stemmed from those similar to what was shown in class.

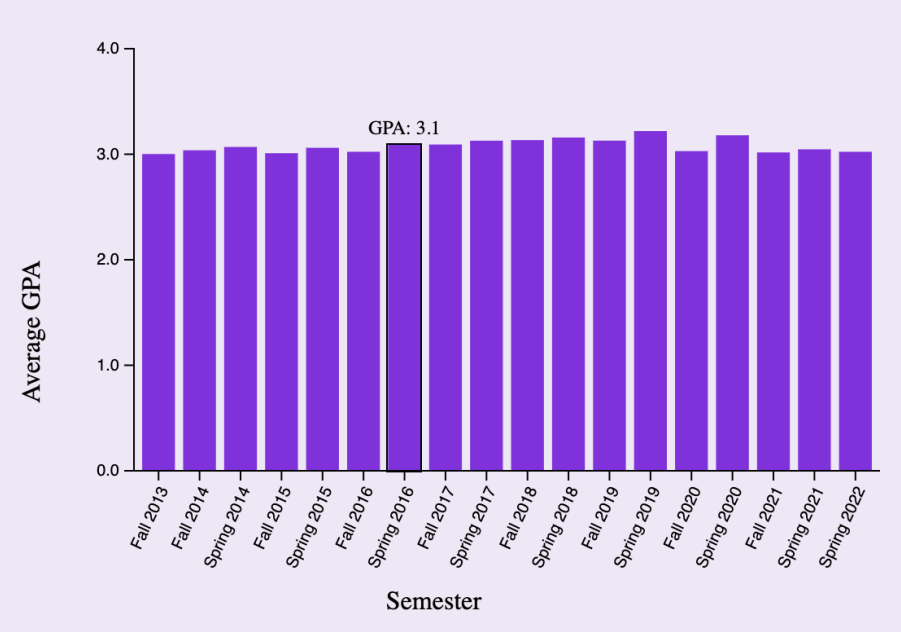
Questions:

When we were first tasked with this assignment, we brainstormed a series of questions that our data could possibly answer through various visualizations. The first of these questions was has grade inflation occurred over the last decade at Clemson University? As our design progressed, we ended up implementing a bar chart that displayed the change in average GPA over the twenty semesters of data we had collected. This differed from our original design plan, which was to create a line graph with each semester as the data points.

Original Design:

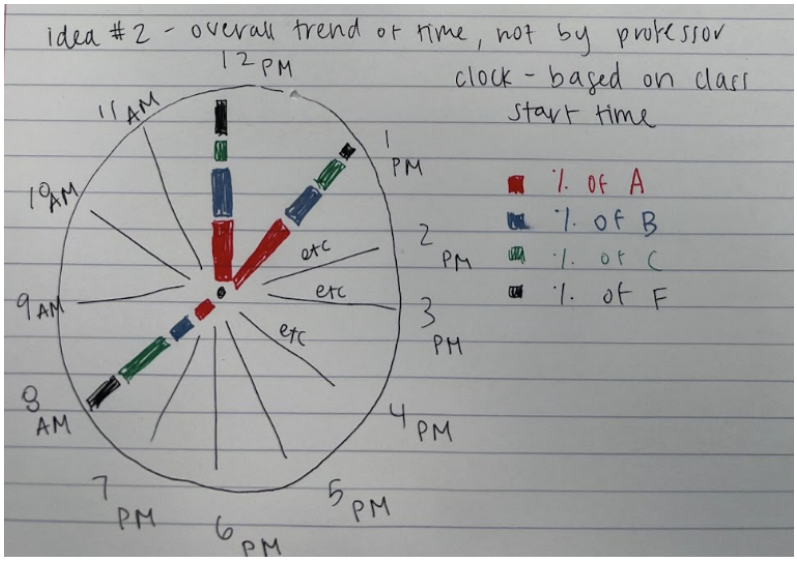


Final Design:



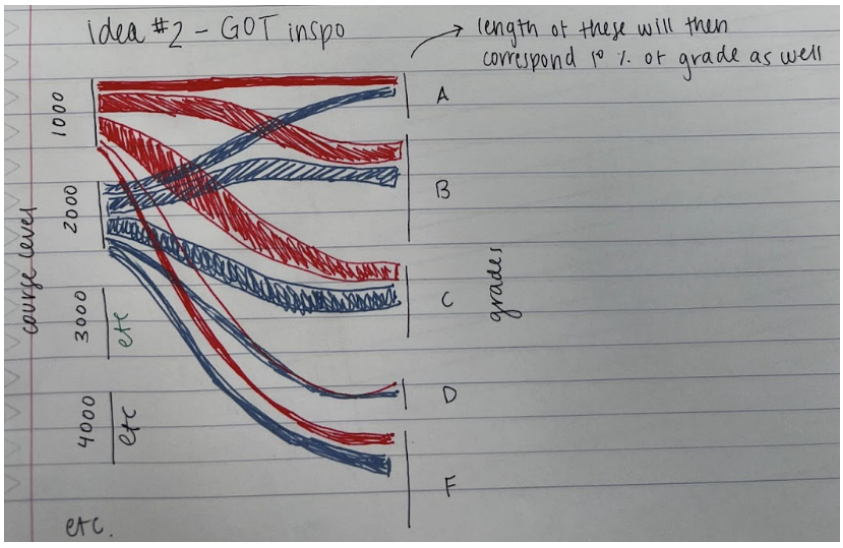
The next question we planned to answer with our visualizations was does course section time have an impact on class performance in CPSC? This question sprouted from data that Dr. Dean graciously provided for us, which contained specific course times for all CPSC classes from the last two semesters. We created a clock-like prototype for this data that would use stacked bar charts pointing in the corresponding time position. Unfortunately, we never got around to using this data, but were we to continue developing this project, we definitely would look into implementing.

Original Design:



Another question we had proposed was whether course level had any correlation to higher or lower grades. During our brainstorming phase, we had thought about using a visualization similar to the Game of Thrones one that we had analyzed at the beginning of the semester. However, we once again determined that a bar chart would work more coherently with the rest of our site.

Original Design:



Final Design:

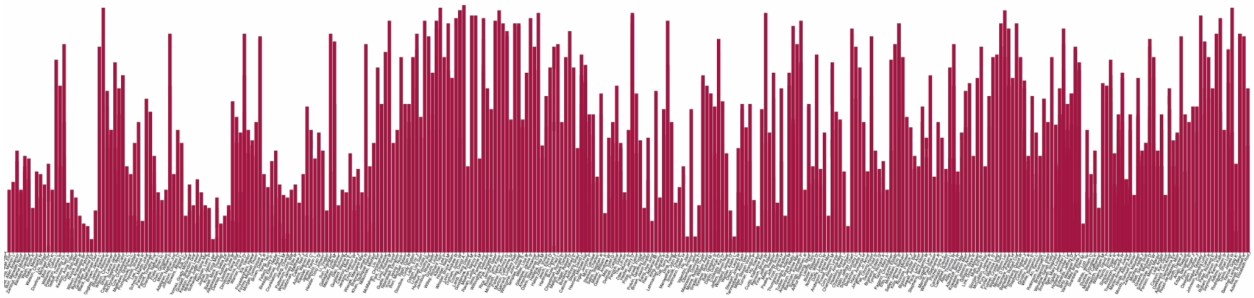
Some other questions that we brainstormed included which classes have the highest variation in grade distribution, is there a correlation between higher average grades and department, and in classes that are shared between undergraduate and graduate students, who generally performs better? While answers to these questions can probably be derived from our visualizations, the main things we focused on were showing comparisons between semesters, departments, course levels, and professors/specific courses.

Data:

As mentioned before, our data was pulled from Clemson University's publicly available grade distributions that range from Fall 2013 to Spring 2022. These PDFs contain final grade percentages as well as the professor who taught the course for every course available at Clemson for a given semester. After downloading these twenty files as PDFs, we converted them into CSV files using an online converter. We then wrote a script that removed the flaws that the online converter created, as there were a handful of mistakes that came from page breaks and such. This left us with refined CSV files that we were then able to pull course names, codes, A-F percentages, and professors from, which all worked hand in hand as the attributes to our visualizations.

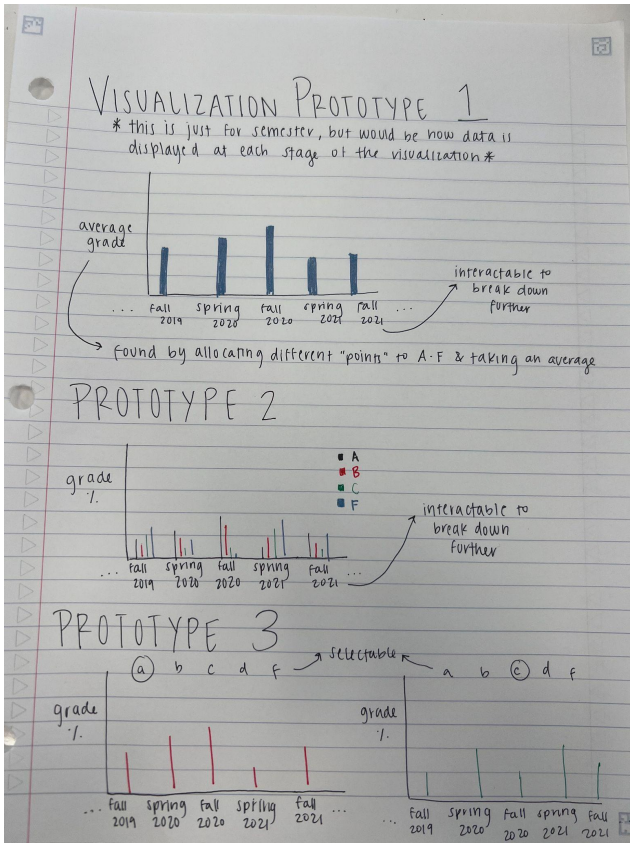
Exploratory Data Analysis:

After initially downloading the data and converting from the PDFs to CSVs, we immediately took our data into a previous d3 assignment to play around with it. Using one of the files, we plotted all of the professors along the x-axis of a bar chart, ranking them from best to worst based on their A rates. This was our first glance into our dataset. It laid the foundations for the different questions that we would eventually try to answer with our dataset, and is what gave us the original barchart idea that we would eventually implement better in our final design.

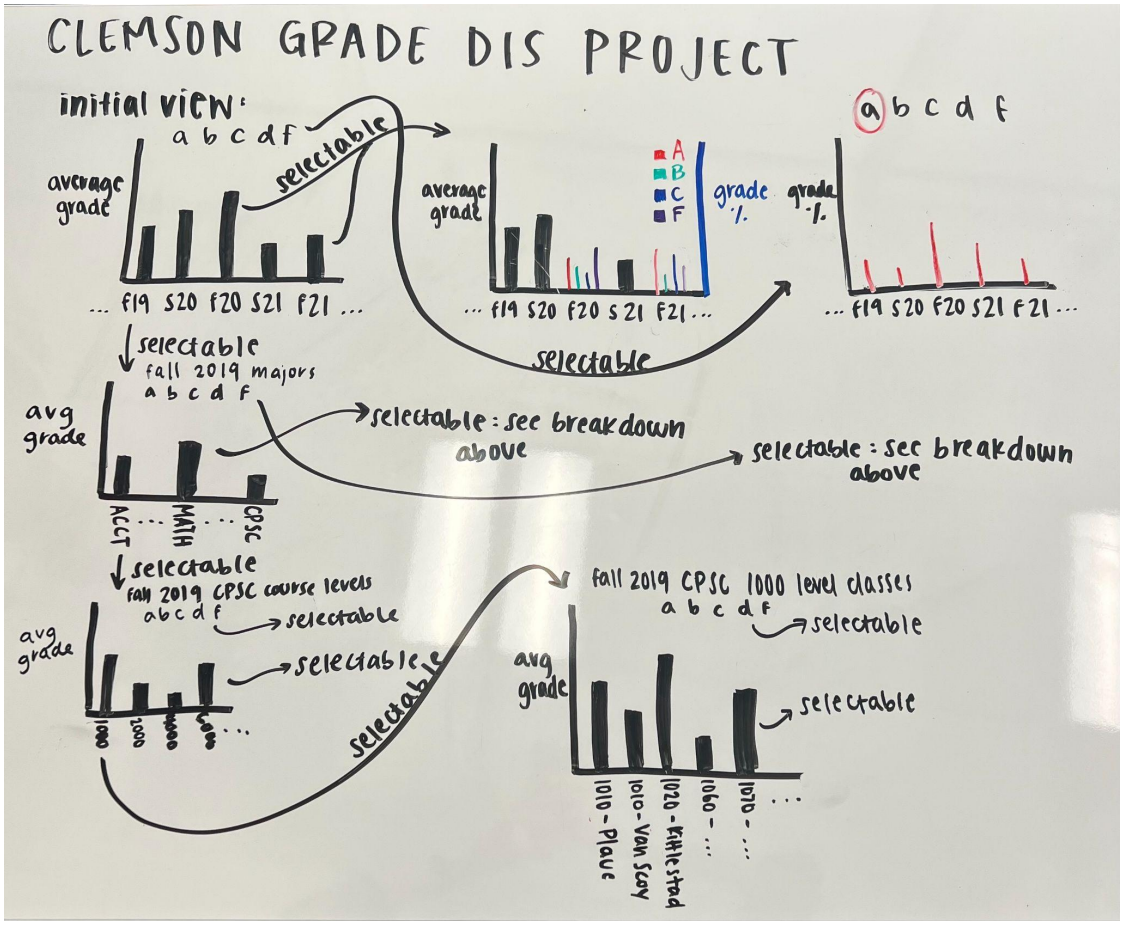


Design Evolution:

The first stage to creating our final design involved asking a variety of questions about our data. These questions were touched on earlier, and we created possible solutions ranging from line graphs to clocks as seen in the description above. We then narrowed down our questions and created three different prototypes for our page. The first prototype utilized a regular bar chart with average grade along the y-axis with all of the bars being interactable to break down from semester into department, course level, and courses/professors. The second prototype had five different colored bars associated with each semester/department/course level/course that reflected the percentages of A-F for each. The final prototype was similar to prototype two, but instead only showed one grade letter at a time, with that input being selectable at the top.



Each of these three prototypes contained elements that we wanted to incorporate into our final design, so rather than selecting one to proceed with, we decided to make our final prototype combination of the three. In this final prototype, we would start with an initial view displaying the data based on semester, with the average grade on a 4.0 scale on the y-axis. These bars would all be selectable, and when clicked, it would break the specific one down into five bars side by side that showed A-F percentages, which would also create an additional y-axis on the right side that ran from 0 to 100. There would also be buttons on the top of the graph, labeled A-F, that would allow you to see all the specified letter grade's percentages for all semesters at once. The final feature of this prototype would come from selecting the semester labels on the x-axis. This would refresh the chart to show departments in that semester on the x-axis, with the same features of being able to select the bars and grades at the top. Similarly, if you were to select the department labels, it would narrow down into the levels for that department, and then once again into specific classes in that level.



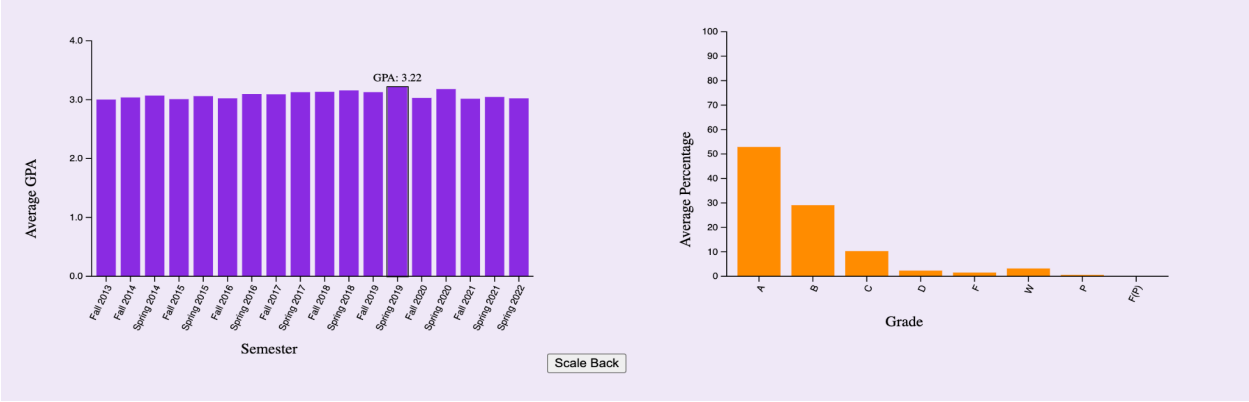
When we first started implementing our designs into code, we planned on using React. However, we quickly learned that incorporating d3 with React was not the easiest process. On our first attempt, we were only able to create a small axis line that we failed to find a way to increase the size of. This led us back to the original implementation plan of just using d3, so we started over.

Our first step on this new implementation was to create the initial semester barchart. We took inspiration from the other bar charts that we had created in previous assignments, so the actual graph configuring did not take long. However, we had to filter our data by semester and then come up with the proper algorithm to calculate the average GPA. In addition to this, we created a zoomed in view button, since all of the GPAs were roughly the same, that would focus around the 3.0 area. We then moved onto creating the department bar chart. This involved manipulating the data differently than we had for the semester chart, which proved to be a little more challenging. For this bar chart we also decided to add a color gradient that would be used to display the density of courses that were within that department. We repeated this process with the level and professor charts, each being a little different than the one before it. This resulted in four charts that would change based on the one before it, but we lacked connection if something further up the line was adjusted. After figuring out a solution to this problem, we moved onto the on hover that would display the grade breakdowns. This step was much easier, and we ended up having eight working graphs when we submitted our initial prototype.

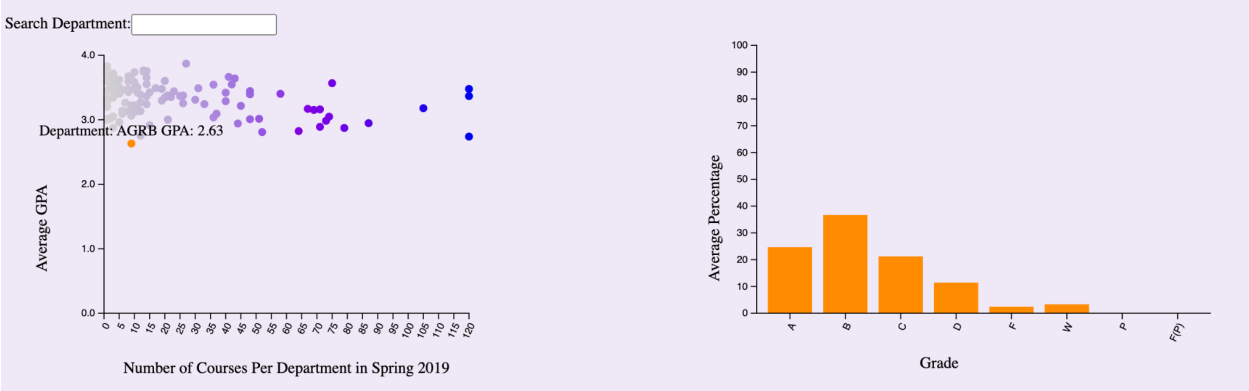
Based upon the feedback we received from our initial submission, there was a lot of work to be done. The first step we took was to merge the eight different bar charts into two that would refresh on selection. We then completely did away with the bar chart for the department visualization, as there was too much data for it to reasonably be shown in that manner. It took a bit of effort to convert it into a scatter plot, and an additional bit of research to implement the search bar, but we were very pleased with this adjustment. We also changed the color gradient that we had initially created to be purely on a light gray purple scale. The other major change that we made was within the professor chart. We realized that for lab type classes where one professor was technically in charge of a surplus of sections, it expanded our bar chart to an unreasonable degree. To resolve this we implemented a condition that averaged GPAs for professors that were teaching the same course. The combination of these, along with some formatting and sizing readjustments, led us to our final design.

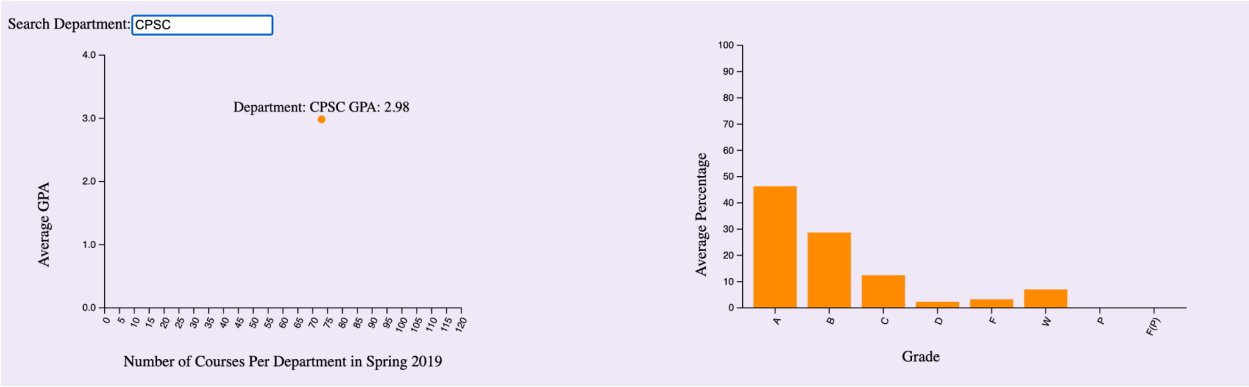
Implementation:

Upon loading our site, you are presented with a bar chart that outlines the different semesters on the x-axis with their average grade on a 4.0 scale along the y-axis. When you hover over each of the bars, the respective A-F percentages appear on a separate bar chart on the right.

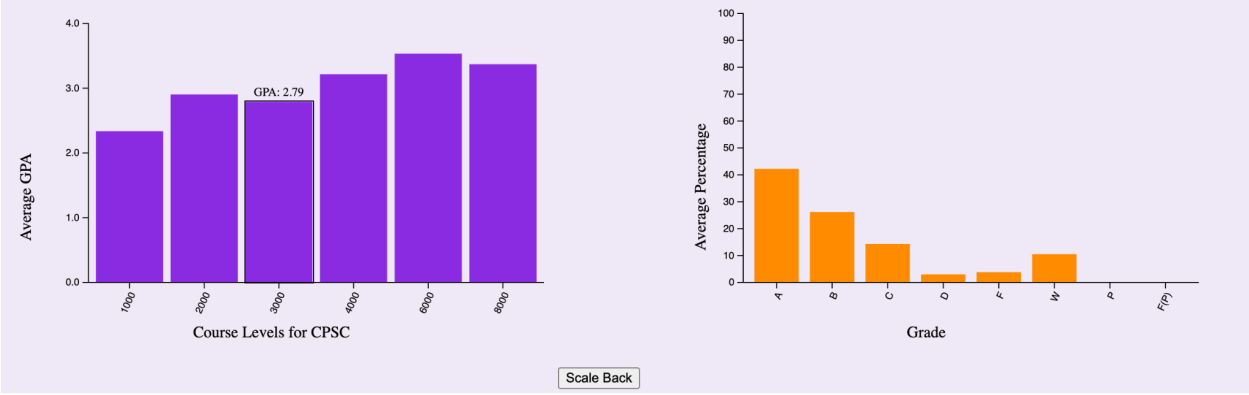


If you were to then select the bars of the initial chart, it would break down into a visualization of that respective semester's departments. However, rather than this being a bar chart, we transitioned it into a scatter plot. We made this adjustment based on the surplus of departments that reside within each semester, as it creates a very lengthy bar chart that would be unsuitable for users. This scatter plot is sorted based on the number of classes that lie within that department, and are colored on a light gray to purple scale to display that as well. There is a search bar on the top of this chart that allows for users to find their respective departments easier, or they can simply hover over the nodes to find them.

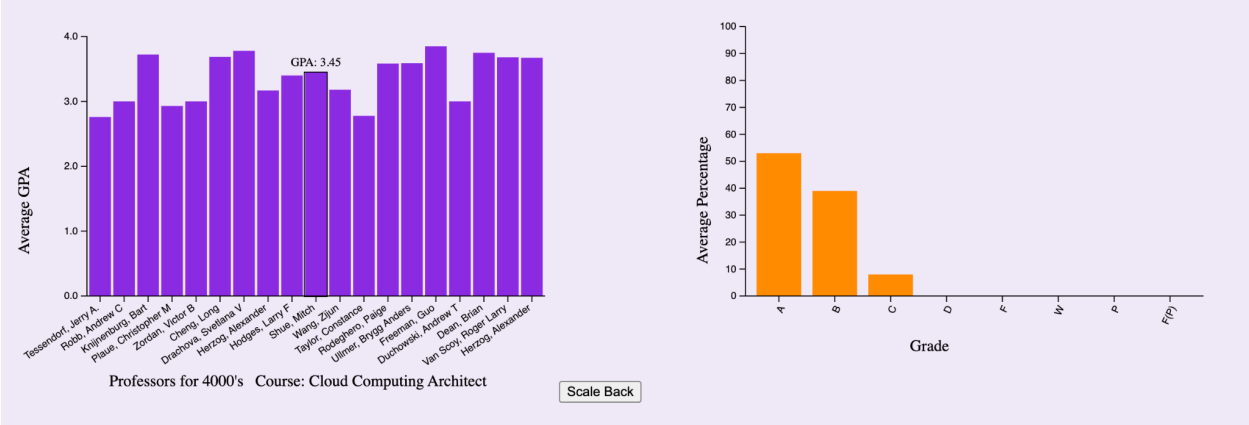




Similarly to how the semester barchart would display grade percentages on the right when hovered over, the scatterplot does the same. When the scatter plot dots are selected, that department's levels are displayed along the x-axis and a barchart similar to the initial one is displayed.



This remains the same when you select the level bars, with specific professors lining the bottom of this bar chart. In this final barchart, not only does the graph along the right change, but the class name is also displayed at the bottom of the graph.



Finally, a scale back button lies within the center of the page, which enables the user to zoom out of the data at any point. This summarizes all of the available interaction that lies within our visualizations.

Evaluation:

The questions we tried to prove with our visualizations surrounded the semester, department, level, and professor data that lies within Clemson's Grade Distributions. We were able to answer some of the questions we had initially outlined when we started the project. The first was has grade inflation occurred over the last decade at Clemson University? We had originally asked this question thinking that the answer would be yes, but based upon the first bar chart we displayed, there is not much fluctuation that can be seen between the twenty semesters, aside from a few spikes surrounding the semesters of COVID. We also asked if course level had any correlation to higher or lower grades. The results that the level bar charts display do not outline a very clear answer to this question, but in many departments the higher level classes seem to have higher GPAs than the lower, which could be due to the size of the classes or the fact that the lower level classes are often the weed out courses. Aside from these questions, we also wanted the main purpose of our site to be for users to navigate to the specific course they were looking for. Overall, we believe that our visualizations properly answered these questions. However, there are some small flaws when it comes to the analysis of our data due to a lack of attributes in our data set. We did not have access to things such as class size, which means comparing GPA can be a little skewed, especially when analyzing a class of twelve compared to a hundred. Ideally, our data set would have contained class size and course section time, as this would have enhanced the visualizations we created. With more time we also would have liked to implement a universal search feature that would take you directly to the professor you are looking for. However, we are pleased with our end result and it confirms our suspicions that CPSC 4030 with Dr. Iuricich is the best course at Clemson.

☺