Jacob Lee & Henrik Nyhus

Professor Radhika Jain

CIS 4170

Final Project

Worrisome Picture of the Student Debt Crisis in the United States

Abstract

This project explores the development of the student debt crisis in the past decade using various visualization techniques with the powerful explanatory tool, D3. The key data used to solidify our initial speculation that it might lead to another economic crisis were the composition of the U.S. household debt and credit by year, cumulative growth, delinquent balance, premium on interest rate in comparison and with respect to the student loan. The key findings were that now the total loan is the second largest national debt category; its cumulative growth is over 180% since Q1 2006; more borrowers are becoming delinquent with student loans; and the premium on the student debt is more than half as much as that of the credit card. Designing these visualizations went through many iterations for optimization both in terms of storytelling and our capabilities with D3. Moving forward, more interactivity and visualization techniques will be applied to further develop this product.

Motivation

Education is undoubtedly the foundation of a nation's economy and the driving force of innovation. This results in creating immense pressure to educate the succeeding generations, and

this is especially true in our society, as higher education has become an inevitable rite of passage for all to integrate to the workforce as a healthy member. However, with the rapid increase in tuition costs and economic pressure, most high school graduates can't avoid taking out significant loans to attend college. Even after earning a college degree, recent college graduates are met with job opportunities that are not on par with the promising prospect with which they had been promised when making what's generally the biggest purchase in their lives. Hence, an increasing number of students fail to make repayments in time, putting a considerable portion of our nation's debt at great risk. This cycle is particularly worrisome as similar patterns were observed in the lead-up of the subprime mortgage crisis. Having experienced such devastating effects of the financial crisis, we wanted to further evaluate what is one of the most pressing economic issues in our time that directly affects not only us as individuals but also the global economy. In addition to being millennials, we felt personal connection to the problem as we observe our peers currently struggling.

Design Process

We began our journey with a question: how has the student debt crisis been developing and should we be worried? And we knew that our visualizations can only be as interesting as their data. In order to answer the question and establish our story, we looked for suitable data for exploring the issue and backing our speculation that it might be another financial crisis. By combing through numerous articles, we were able to gain general understanding of the gravity and development of the issue. Generally, they pointed to the same data to support their theses: the Federal Reserve's economic research data. The key indicators were the overall composition of the U.S. household debt and credit, the student loan delinquency rate, and the student loan

growth rate. Fortunately, the Fed has visualized the data elegantly (https://www.newyorkfed.org/microeconomics/hhdc.html), so it was easy to understand the patterns at a glance. Yet, the challenge was that the numerical data weren't available in an organized format, such as csv and excel. We also looked at the quarterly reports to see if they contained any data table or list, we realized that the only way to get the historical data was to aggregate them ourselves by manually recording them from the available charts. In addition to the aforementioned indicators, we wanted to find out how much the federal government and the student loan issuers are taking advantage of students in terms of interest rates. This required independent fact finding on our part as the comparison between the interest rates on various kinds of loan wasn't available via articles or reports. As the interest rates are calculated based on their respective underlying rates, namely the prime rate and the 10-year treasury note interest rate, we sourced the historical data for the underlying rates in conjunction with the market interest rates to which the borrowers are subject. Following is the list of our sources:

- https://www.newyorkfed.org/microeconomics/hhdc.html
- http://www.freddiemac.com/pmms/pmms30.htm
- https://research.stlouisfed.org/fred2/series/TERMCBAUTO48NS
- https://research.stlouisfed.org/fred2/series/TERMCBCCALLNS
- https://www.edvisors.com/college-loans/federal/stafford/interest-rates/
- http://www.fedprimerate.com/wall_street_journal_prime_rate_history.htm
- http://www.multpl.com/10-year-treasury-rate/table/by-year

Following the sourcing, we had to tailor the data. In terms of difficulty, the manually aggregated data were easier to work with because the size was relatively small and we dictated

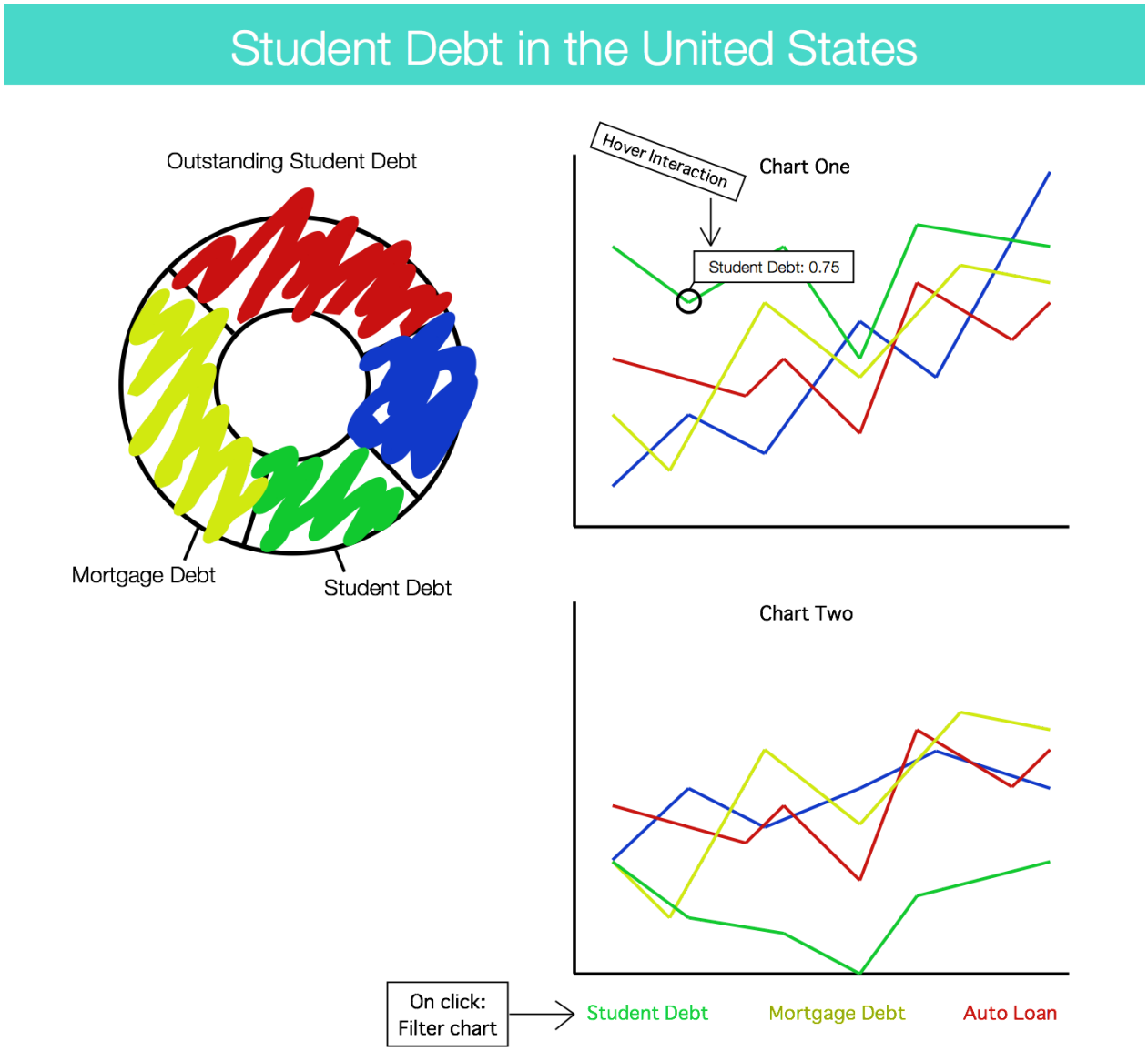
the sorting and its format from the beginning. For example, the overall household debt and credit data only required rounding up the numbers to the same decimal place and excluding data other than Q4 for each year. However, other datasets required extra work. For the cumulative growth data, we repeated the same process and calculated each quarter's growth by comparing it to Q1 2006. While doing this, we ran into a minor issue of structuring the data so that the lines have the same labeling for each category of loan. This was important because we wanted them to act in unison and have matching colors in the final version. Different code samples had conflicting ways of organizing data in terms of key-value relations. Some used column names as keys; some used columns dedicated to names that are intended for keys; And some used internal object arrays. A minor difference in formatting entailed a major change in the source code and added exponentially more complexity, so we carefully analyzed the different methods and decided to create a separate column with category names in csv. We then noticed that a specific order has to be applied to each dataset for color-matching purposes: housing, auto, credit card, student and other. Furthermore, the dates needed to be matched for different categories. So we deleted the rows with null values and unwanted dates. This process became particularly tedious for the premium data as there were three datasets that needed to be connected. Although the process sounds like a quick and easy process in retrospect, it took us numerous trial and errors to finally get them in the right format.

One weakness we discovered in the coverage of the issue in the articles was that the visualizations were segmented due to their verbal explanations and distributed across different articles. What we wanted to do was first visualize some of the data in a different way then put them together so that our thesis could speak for itself with less texts and more data. As the name

of our visualization tool suggests, we wanted this to be a data-driven document. In terms of visualization techniques, we were fascinated and inspired by various examples on the D3 website, especially the ones on NYTimes by Mike Bostock. From force graphs to voronoi diagrams, D3 is extremely versatile and flexible and allows for immense amount of control over creating visualizations. This, however, comes with a cost because more flexibility means more coding and more complexity. Neither of us had had any experience with using the library, so the scope of our visualizations became quite limited to what techniques were readily accessible and relatively easy to produce and simultaneously suitable to visualize our data.

For the composition of the household debt and credit, what we found during our research was a time series line graph with multiple lines representing different loan categories. Although this is great to visualize the change over time, one can use a pie chart to put their sizes in perspective, and by utilizing animation interactive, their growth can also be conveyed as well. For cumulative growth, delinquency and premium on interest rate, the multi-line graph was the best as the change over time was the key finding, but we wanted to take a step further by linking

them together and allow the user to filter based on categories. The below figure represents how



we envisioned the product.

Figure 1, Mockup

To do this we needed to learn the basics of D3: svg, data binding, scale and so much more. An online book and many youtube videos later, we started with the multi-line graphs.

There were many nuances that prevented us from even creating the grid lines, let alone the browser compatibility issues. We sourced bits and pieces from different example codes. Each small error we had caused hours of debugging and doing extensive research on stackoverflow and google. For example, all visualizations in D3 requires creating a svg canvas and appending a group element to contain the lines including the x and y axes. Simply missing to append a group element for one group of lines prevents the complete rendering. So, from the very minute details like the ticks on the scale to animation on the donut chart, it was a strenuous process.

Interactives proved to be the toughest challenge as the level of complexity wasn't surmountable with the time constraint. Our ambitious goal was to incorporate the user input to be shown and manipulate the visualizations to a great extent, but the more realistic goal was to amplify the static visualizations we had with what we learned how to do. For the donut chart, we created buttons on the bottom so the user can adjust the composition by year. Further, the user can hover on the sections to see the exact dollar amounts. For the multi-line graphs, we created labels on them bottom that allow the user to filter out different lines. Also, the user can hover on the chart to see the mouse-following vertical line. Our intention here was to show exact values for each line at a certain x value and zoom on demand, however incorporating these tooltips turned out to be very challenging and required more learning. After days of relentless trial and error, our final is as shown below:

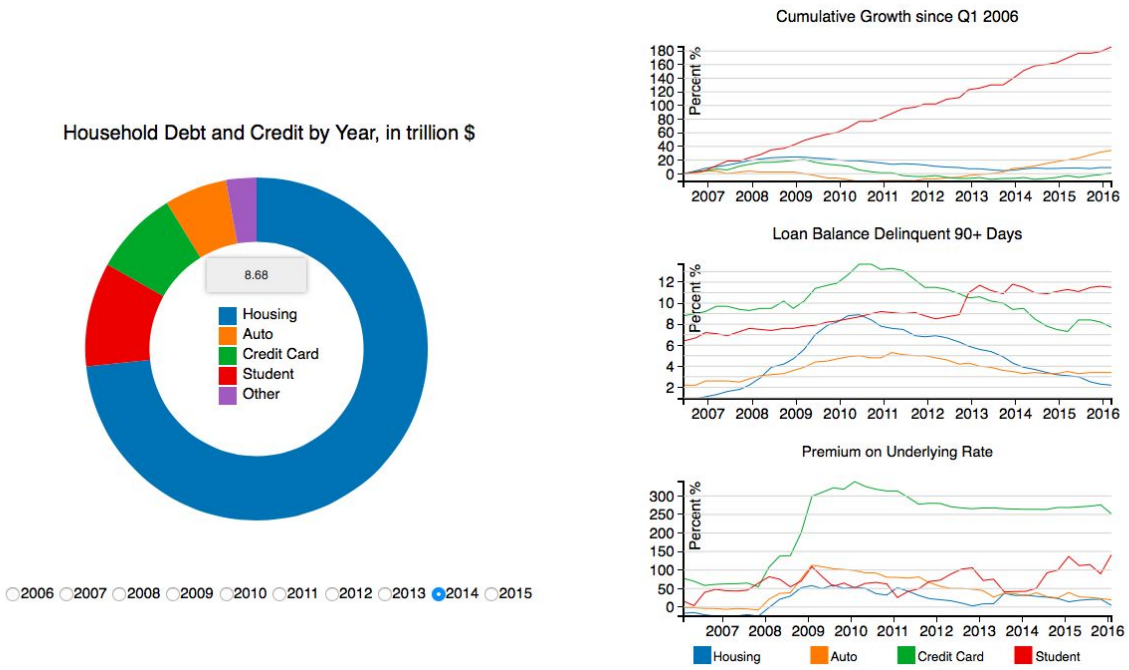


Figure 2, Final Product

User Testing / Feedback

Our feedback was generally positive with emphasis on the animation and filtering. Further, the consistency in colors and alignment generated positive impressions and user experience. On the other hand, an user raised a concern about not being able to hover over the different lines and get specific values, as having to use the horizontal grid lines as reference points is less user-friendly. The user was also concerned about the difficulty with finding the filter feature and the small font sizes: “I like it! One thing I would check is if you can autosize in html to the window, in FF there is a lot of white space I can scroll to, and the word percent is kinda hard to see on the graph cause of the lines.” We could not implement the hover feature on the multi-line graphs due to its complexity, but we took the feedback into consideration and attached small squares next to the labels and increased the font sizes.

Strengths and Weaknesses

The biggest strength of our visualizations is that it tells us an extensive story about what has happened with student loans in the past decade. We wanted to visualize the worrisome trends which is in many respects are similar to the subprime mortgage crisis. Another strength of our visualization is the clean design and minimal use of ink, which allows the user to focus on the data and not irrelevant chart junk. To make the data clearer, the user can filter out the different debt variables so they can study each variable by itself. The biggest weakness of our design is the relative lack of interactivity that can enhance independent exploration on the user's part and that a verbal explanation still has to accompany to provide more background. It would be helpful if we could incorporate the story either visually or auditory so we could spread the word to our users more easily and efficiently. Another weakness is that we cannot get more specific information that can be applied to different users' backgrounds. It would also be helpful if the user could choose their college, major and home state and we could make a recommendation about their choice or give them a clear picture of the financial consequences of their decisions.

Further work

Going forward with this project we would continue to work on the weaknesses by adding more user interactivity to our visualization. To add a visual or auditory storyline where we can talk about specific policies the federal government made through time would be helpful in conveying the message. We would also create a more personalized experience where millennials can get

specific advice about their situation using tuition cost, underemployment, unemployment and income data.

Acknowledgements

- Interactive Data Visualization for the Web - http://chimera.labs.oreilly.com/books/1230000000345/index.html
- D3Vienno - https://www.youtube.com/user/d3Vienno/playlists
- Mike Bostock's Blocks - https://bl.ocks.org/mbostock
- The Center for Microeconomic Data - https://www.newyorkfed.org/microeconomics/hhdc.html