Final Project

INFO 5100
May 13, 2019
Team: Andrea Gonzalez-Del-Valle (avg37), Xu Jing (xj88), Nianyi Wang (nw344)
Final Project

Data

For this project, we used two datasets that were found from Kaggle.

The first dataset is from young people survey which consists of responses from young people aged 15 to 30. Each variable can be split into the following categories: music preferences, movie preferences, hobbies & interests, phobias, health habits, personality traits & views on life & opinions, spending habits and demographics. Participants will choose Strongly disagree 1-2-3-4-5 Strongly agree for each question. Since there are so many different questions for each category, we decided only to use ten of them just to simplify the data and also make the graph more clear. The variables that we chose are very typical in that category. Here are the following variables that were used in our project:

Music Preferences: dance, country, classical, pop, rock, opera, hip-hop, musical, jazz, punk
Movie Preferences: horror, thriller, comedy, romantic, sci-fi, war, tale, cartoon, document, action
Hobbies: outdoor, dancing, instrument, sport, gardening, shopping, pet, social, poetry writing, cars
Phobias: flying, thunder/lightning, darkness, heights, spiders, snakes, rats/mice, aging, dangerous dogs, public speaking
Health: healthy lifestyle, non-healthy lifestyle
Spending Habits: save money, enjoy large shopping mall, prefer branded clothing, spend money on entertainment, spend money on appearance, spend money on tech, spend money on healthy food

We didn't include personality traits & views on life & opinions because there are about 60 questions under this category and it's hard for us to determine which ten of them should be used.

We first preprocessed this dataset before use. For each variable, we calculated the average points, grouped them by age and gender and stored into JSON file. Since this dataset will also be used for the pie

chart, we captured the percentage of each answer, i.e., the percentage of people who choose 5 and stored into another JSON file.

The second dataset is from college salaries which contains data of salary distributions by major. For each major, we have starting median salary, mid-career median salary, and percentiles of mid-career salary. Variables are percentiles and major. We decided to plot the distribution of mid-career salaries. We first processed the dataset as a CSV and simplified some of the titles and text. I returned the CSV as an array of objects containing the major labels and mid-career salaries values. We decided to not include starting salary in the processing because we felt like it may be too much information in one graph for the user. Returning the data into an array of objects made the data simple to work with in the d3 library.

Visual Design Rationale

Radius Chart Showing Survey Averages for Age & Gender

We used the Radius Chart to show the overview of the participants' quantity results regarding different questions. Three white circles represent rating 1, rating 3, and rating 5. For different categories, ratings have different meanings. For instance, rating 1 means "Not afraid of all" for the Phobias, while rating means "Strongly disagree" for the Music Preferences.

Position: Since the dropdown might overlap with the radius chart when users are selecting the items from the Dropdown menu, and thus we separated the dropdowns with the radius chart —— we placed two Dropdown at the left of the screen while the radius chart is placed at the right side of the screen.

Position for texts: Texts on the Radius Charts are rotated differently so that it would be more intuitive for users to read the texts.

Color: We adopted 6 different colors to differentiate 6 categories and assigned the colors to different categories by using the colorScale d3.scaleOrdinal().

Transformation: we used arc() to build the Radius Charts, which contains the xScale and yScale. xScale decides angels while yScale decides Radius.

Donut Chart Showing Category Item Breakdown

On the center of the Radius Chart, we created a donut chart to show the breakdown of the answers for each question (1 is the answer Strongly Disagree to 5 is the answer for Strongly Agree). We decided to use the donut chart instead of a pie chart because we wanted to include the category that the user selected in the middle of the graph. When calculating the average points for each question, we also calculated the

percentage for each answer. The circular barplot already shows the average points, we think some users might also be curious about the breakdown. The larger area represented a higher percentage. We decided to use very different colors that users would be able to distinguish.

Salaries by Major

For the salaries graph, we decided to create a line chart that shows the percentiles of mid-year salaries. Because we collected data on percentiles, we decided the best visualization would be to create a distribution chart. We are displaying an approximate visualization of the distribution. Because the y-axis shows percentages, the 10th percentile and 90th percentile will both plot at 10%, and 25th percentile and 75th percentile will both plot at 25%.

A large tradeoff, was that data was not available for minimum and maximum mid-career salaries so we could not plot the entire distribution. However, because we make the percentiles clear in the hover interactions, we believe we mitigated against the viewing risk. We also decided to plot they scale stating at 10%, rather than 0, so the line chart wouldn't just abruptly curtail when y was less than 10%.

We decided to create a natural curve connecting the data points and used d3 curveCatmullRom when connecting the data points. We added the text label to the median salary because this data seemed most relevant for the average viewer. We played around with the thickness and font-size of the median salary text to add to the visibility and accessibility of the viewer.

Interactive Design Rationale

Dropdown Menu

Users can select their ages, genders and (potential) majors in order to see the statistic results. We chose dropdown menu instead of input box because this would force the user to input data in the format that we needed. Especially for the major, there are so many different ways to say one major which will be hard for us really use the input without any restrictions. The down arrow of each box indicates that users could click on it to see the expansion.

Radius Chart Showing Survey Averages for Age & Gender

We added the transition() function and delay() function to the Radius Chart so that each path would show up one by one. Especially, we used the ease function to make the chart more dynamic. When users select

a value from the dropdown menu, the Radius Chart will clean all the old graphics and redraw the new diagrams. The Radius Chart is clickable. As the users click on a certain bar of the Radius Chart, the Donut Chart would show up, and present more detailed information about the bar they chose.

Donut Chart Showing Category Item Breakdown

When selecting a certain bar from the circular barplot, the donut chart will be updated to the corresponding category. Users can hover over the donut chart to see the actual data of the percentage. The opacity of the selected part will go down to indicate the selection. We decided to add this interaction because using the area to show portion might be a little confusing for some people who are not good at determining which area is larger. Thus, showing actual data will help them understand which answer has been selected most.

Salaries by Major

For the salaries graph, the user first discovers the interaction by selecting their major from the dropdown box. Because the user will first see a blank screen, they know they will have to interact with the dropdowns. After the user selects their major, they then see the graph with the distribution of salaries of people who majored in their selected major. They can choose between 50 majors and can switch the graphs data to suit their chosen major. We believe the customizability adds to a personalized interaction.

We first had a static line graph, but then added additional interactivity with the hoverable datapoint circles. The user can hover over the circles to see a tooltip with the percentile (10th, 25th, 75th, 90th) and to see the specific mid-career salary at that percentile. We felt like adding tooltips was an effective user experience decision because, otherwise, it would be difficult to discern the data points simply looking at the line.

The Story

We want our tool to ultimately get viewers thinking about how others live. The data itself explores young people's common habits, phobias, job prospects, and views about the world. It uses over 1000 responses of people between the ages of 15 to 30 and is segmented by different demographic factors. We thought it would be interesting to make the data interactive and let users gain insight into how people similar to them (in age, gender, and college major) live.

We focused on displaying averages or medians in our visualization. Average response in the survey is displayed in the circular barplot. Median mid-career salary is predominantly displayed in our salaries chart. Averages and medians indicate standard. People often like to compare themselves to averages/medians because it indicates a sense of "normalcy." Especially in some classes at Cornell, knowing how you stack up against the median can be essential because it indicates your performance relative to your peers and reflects the final grade. In terms of grades or salaries, mostly all people would rather be at or above the average/median. However, when it comes to more qualitative qualities like movie preferences or hobbies, it is hard to say whether people would want to be "average." For these qualitative studies, there is also no objective metric on what is a "good" answer or a "bad" answer. By observing how users interact with our visualization and asking them their opinions on being "normal", we would get surprising results on the tradeoffs of our tool.

The visualization overall shows viewers the preferences, hobbies, fears, and financial prospects of people of different backgrounds. We hope this creates a sense of empathy and intrigue on how other people live.

Outline of team contributions

Our project contains three graphs and each group member is responsible for one of them. Xu made the circular barplot, Andrea made the line graph and Nianyi made the donut chart. Since the dataset of young people survey is very large and consists of many responses that we don't need, Nianyi preprocessed this dataset into JSON files. We first made sure that we had all the graphs ready for demo. Then Andrea and Nianyi worked on interaction elements that could combine three graphs together with the dropdown menu while Xu was working on the dropdown bar polishing, the style of the whole visualization adjustment, and helping with debugs. We all worked on this document. We roughly spent 3-5 hours on our own and 1h meeting together per week. The most time-consuming part was cooperating all three graphs with correct interactions.