Employment in STEM

Arielle Greenberg[*] Noreen Hossain[†] Anaelle Surprise[‡] Reela Taher[§]

Northeastern University
DS4200 Information Visualization

ABSTRACT

Our interactive visualization tool aims to promote diversity and inclusion in STEM-related fields by supporting several domain tasks. These tasks include exploring the employment situation for minority groups in STEM-related fields, identifying underrepresented groups and areas where additional resources and support are needed, and making informed decisions about entering their desired field based on the level of diversity to expect. Our visualization utilizes data sourced from the National Science Foundation (NSF) report to create an all-inclusive and adaptable representation of the employment metrics for individuals from various demographics in the STEM field. The tool provides users with access to data on the representation of different demographics in STEM-related fields, including sex, race/ethnic identity, and disability status. Users can explore the male/female distribution of people in STEM-related fields within each race/ethnic group and employment metrics across different STEM jobs, enabling them to gain insights into the current state of diversity and inclusion in these fields. It is important to support these tasks because diversity and inclusion are critical factors in promoting innovation and progress in the STEM field. By providing individuals with access to data and insights that highlight the under-representation of certain groups, organizations can create a more diverse and inclusive workforce that leverages the unique perspectives and experiences of individuals from different backgrounds, leading to more innovative solutions to complex problems. Our interactive visualization tool is a valuable resource for individuals and organizations alike, helping to promote diversity and inclusion in STEM-related fields and support the creation of a more equitable and just society. By empowering individuals from different backgrounds to explore, identify, and ultimately address the disproportionate representation of different demographics in STEM-related fields, the tool is making an important contribution to promoting diversity and inclusion in these fields.

1 INTRODUCTION

The under-representation of minorities in the STEM field is a persistent issue that requires attention and innovative solutions. According to the National Science Foundation [1], STEM occupations are 89% white and 72% male. Additionally, according to this report on Diversity and STEM: Women, Minorities, and Persons with Disabilities [3] within S&E occupations, the rates of those reporting at least one disability were sometimes as low as 8%. This lack of diversity and inclusion in STEM fields hinders innovation, competitiveness, and social equity. For STEM companies to create solutions to problems that affect all people, they must make sure that they are hearing from all kinds of people. Our interactive visualization aims to address this problem by providing a tool that empowers individuals from different backgrounds to explore, identify, and ultimately address the disparities between different groups within occupations in the STEM fields. The data used in our visualization is sourced from the National Center for Science and Engineering Statistics (NCSES) report titled ”Women, Minorities, and Persons with Disabilities in Science and Engineering,” which provides comprehensive and detailed data on employment metrics in various STEM fields. The report presents data on various demographics, including race/ethnicity, age, gender, and disability status. The employment metrics provided in the report are organized by industry, occupation group, and job title and includes information on the educational requirements, job outlook, and the median salary for each occupation.These data sources are based on a variety of sources, including the U.S. Census Bureau's American Community Survey and the Bureau of Labor Statistics' Current Population Survey. Our visualization combines various reports and enables users to interact with this data and explore the employment situation in STEM fields based on their own specific demographic group of interest. Users can filter the data by demographic characteristics, such as race/ethnicity, sex, and disability status. Overall, the data provided in these reports is a valuable resource for understanding the current employment situation in STEM fields, as well as identifying areas where additional support and resources may be needed to address the under-representation of certain groups within occupations. Our visualization aims to create an accurate image of the employment situation in STEM-related occupations for the user's demographic group of interest, by providing a hyper-specific view of user-selected demographic groups. By providing an approximate count for those within certain groups, our visualization enables users to gain a comprehensive understanding of the employment situation and level of representation for those in their selected groups. This allows them to make informed decisions and advocate for policies that address the under-representation of minorities in specific occupations. Supporting numerous domain tasks such as demographic analysis is critical for promoting diversity and inclusion in the STEM workforce. Underrepresentation of groups leads to gaps in understanding of problems that different STEM fields are trying to address, and these disparities in these occupations need to be addressed. By empowering minorities with the tools and information they need to succeed in STEM, our visualization can help create a more inclusive, vibrant, and sustainable STEM ecosystem.

1.1 Domain Tasks

Our visualization tool can support the domain tasks of filtering and comparing the information presented. The user can compare the groups at each stage of the visualization by evaluating the differences in the heights of the bars in the bar charts for each group. Next, the user can filter the total set of data by different demographic groups to find the specific group they are looking for. This task involves identifying, comparing, and summarizing the data for a combination of specific demographic groups. Both tasks are supported by the visualization's ability to filter and display data according to the user's selections, namely which bars the user decides to click on. The visualization allows the user to explore and compare employment data for different demographic groups, providing a clear and accurate representation of employment in STEM fields across different demo-

[*]e-mail: greenberg.ar@northeastern.edu
[†]e-mail: hossain.n@northeastern.edu
[‡]e-mail: surprise.a@northeastern.edu
[§]e-mail: taher.r@northeastern.edu

graphic groups within their selected state. By using the visualization to identify patterns and trends in employment, the user, which can be those seeking employment or recruiters, can make informed decisions and advocate for policies that address the under-representation of minority groups in the STEM field in the United States.

2 Related Work

Paper 1: Increasing gender diversity in STEM: A tool for raising awareness of the engineering profession [2]

This paper discusses a tool made by researchers to increase gender diversity in STEM by allowing high school individuals interested in engineering to filter through their personality, views, and expectations, and see how closely they relate to that of a professional engineer. The tool's name is ANNA, and its primary goal is not only to provide access to role models to the young adolescents using it but also to implement transparency and increased awareness behind the engineering profession, something we'd like to do with STEM-related fields as a whole in our project [2].

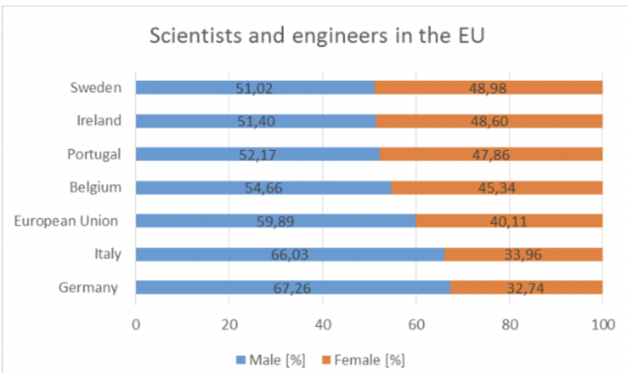

Figure 1: Graph taken from Paper 1 [2]

Figure 1 is taken from the paper and highlights the distribution of male and female scientists and engineers in the EU [2]. Even though we plan on creating a visualization that focuses on many more characteristics than simply sex, such as education, race, state, and disability status, this visualization does highlight the need for simplicity. In an effort to make our message as digestible as possible, we plan on mimicking the nature of this visualization by creating individual visualizations for each characteristic, and then one final interactive visualization that incorporates all factors.

The ANNA tool incorporates filtering, something we aim to emulate in our visualization. The tool allows users to filter based on characteristics like gender, graduate/undergraduate, field of study, age, interest, university, etc [2]. By analyzing the ANNA tool, we are informing our own work, as this has inspired us to incorporate filtering into our visualization. We decided to filter based on similar, but not identical, characteristics, as listed above, but nonetheless the filtering in both projects emphasizes the fact that those utilizing these tools/visualization are multi-layered individuals. We cannot be defined by one simple demographic. Rather, we are one of the infinite types of combinations of people, and that's something that needs to be accounted for.

While some ideas in the paper differ from ours as they highlight the expectations women face when entering the engineering field, the sentiment remains the same: there is an undeniable gender gap in STEM, and systemic discrimination is preventing women from maximizing their potential in STEM-related fields. We as advocates must increase transparency and understanding in order to make a change.

Paper 2: Using Facebook Ads Data to Assess Gender Balance in STEM: Evidence from Brazil [5]

This paper highlights the need for increased workforce diversification in STEM-related fields. The authors recognize the disappointing fact that men dominate the STEM workforce over women, and they claim that the lack of diversification is inhibiting productivity in our current world economy [5]. The authors discuss that in order to understand the overall priorities and values of different world regions, we must analyze the data regarding gender parity. They choose to analyze Facebook ads and expose the discrepancies between male and female interaction towards STEM-related ads [5]. This idea is the driving force of our project. Unlike the article "Increasing gender diversity in STEM: A tool for raising awareness of the engineering profession" [2] that we discussed above, the main goal of our project is not to offer solutions, or at least not in a technical sense. After working with our data and brainstorming for this project, we came to the conclusion that offering transparency and awareness of discrimination in STEM is, in some sense, a solution. Once people can truly visualize the comparison between employed and unemployed STEM workers of different demographics, it will expose the pressing need for change in our workforce.

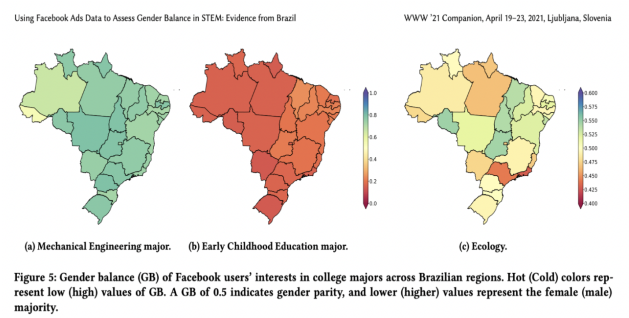

Figure 2: Graph taken from Paper 2 [5]

As we were analyzing the paper, we came across the visualization in **Figure 2**, which highlights the "Gender balance of Facebook users' interests in college majors across Brazilian Regions" [5]. In this visualization, the author uses color hue and saturation as channels to show the different extremes: female majority and male majority. This visualization inspired us to incorporate color and hue into our visualization. Since we have so many different demographics to compare against one another, we were struggling to find ways to implement different comparison channels without making the visualization too busy. Since the project does not restrict us to just one visualization, this inspired us to perhaps create a very similar visualization highlighting the male and female representation in STEM roles in each state.

3 Use Case

Jasmine is a black female with a physical disability who is interested in pursuing a career in the STEM field. She faces significant challenges due to the under-representation of minority groups in STEM-related fields, including her race and disability status. Jasmine uses our interactive visualization tool to gain insights into the employment situation for people in STEM fields who share her identity. Jasmine accesses the first chart in the tool, which shows the male/female distribution of people in STEM-related fields within each race/ethnic group. She is interested in exploring the representation of black females with disabilities in particular, so she clicks on the female bar. This takes her to the next chart, which shows the employment metrics for black females in STEM-related fields, broken down into their respective disability status counts for each STEM-field occupation. Jasmine is pleased to see that her field of interest, computer and information science, is represented in the chart. She hovers over the bars and notes that the representation of non-disabled individuals is a bit higher than that of disabled individuals. She clicks on the bars to view the proportions of disabled individuals to non-disabled individuals in the form of a pie chart. She is disappointed to see that the representation of her demographic is significantly low. Using the insights gained from the visualization, Jasmine can make informed decisions about pursuing a career in the STEM field. She can also advocate for policies that address the under-representation of minority groups, such as black females with disabilities, in STEM fields. By using the visualization to identify areas where additional resources and support are needed, Jasmine can create a path for herself in the STEM field and pave the way for others like her. The interactive visualization tool is a valuable resource for individuals like Jasmine, empowering them to explore, identify, and ultimately address the disproportionate representation of different demographics in STEM-related fields.

4 DATA

4.1 Where the Data Comes From

The data comes from the Bureau of Labor Statistics (BLS) and the National Center for Science and Engineering Statistics (NCSES). The data from the BLS was collected via survey in 2021 to investigate the gender split in the United States' tech sector. The data utilized from NCSES is from two different studies. The first study is titled "The STEM Labor Force of Today: Scientists, Engineers, and Skilled Technical Workers" [4] which utilized various education and workforce surveys to provide an analysis of the US STEM labor force's education levels, growth rates, salaries, and demographics. The second study is titled "Diversity and STEM: Women, Minorities, and Persons with Disabilities" [3] which utilized NCSES' workforce surveys as well as some census data.

4.2 Biases and Ethical Considerations

Much of the data collected through these two studies were compiled from United States Census data, as well as through surveys run by the NCSES to collect information about college graduates, doctorate candidates, and postsecondary education data for students in science and engineering. Census data has historically posed problems of undercounting certain populations, which the studies seem to try and mitigate by utilizing multiple sources of information. Because of this, we were unable to include certain groups in our analysis due to either data confidentiality reasons, or data reliability issues. In the original tables, these data points are noted by "D," for data suppressed for data confidentiality concerns, and "S" for data reliability concerns. In our copy of the data, we replaced these values with -1, and -2, respectively to note that there was no data for these groups, but allowing for our graphing methods to be able to compute them. These groups include Indigenous Americans (listed in the study as American Indian), Native Hawaiian as broad groups, and other more specific groups such as African American Psychologists.

Another consideration for the NCSES studies is that they were either called in, or web administered. Since most respondents chose to self-administer the survey, as it was web-based, there may be a slight bias towards those with more experience and access to technology. However, the surveys are given to graduate students, so presumably, there would be less variability in their access to these things than there would be in the general population.

4.3 Cleaning Process

Most of the data were in a table format, processed to be visually informative rather than in a format primed for manipulation. To clean the data, only columns and rows containing information relevant to sex were retained in order to remove the hierarchical structure of the tables. In addition to this, the numbers were cleaned to remove commas from all cells and the null values were standardized. Meaning, the various symbols representing a lack of data were replaced with -1 and -2 as described in the previous section. Next, we realized that the three separate data tables presented an issue for visualization, as they described the group as a whole, but through different lenses. We wanted to see if there was a way to combine the values, in order to create a graph that allows the user to filter through the "whole," and make the graph interactive. The way we chose to do so was to find approximate counts for the most specific possible groups, namely the unique combinations of race, sex, and disability status.

The data we narrowed down to encompasses 7 race/ethnic groups, 2 sexes, 9 S&E occupations, and 2 disability statuses, resulting in 252 unique combinations of the 4 groups. We wrote a short Python code snippet to calculate each unique combination. The code outputted a list of lists of each combination, (ex: [['Social Scientist', 'American Indian', 'With Disability', 'Male'], ['Social Scientist', 'American Indian', 'With Disability', 'Female'],...,['Industrial Engineer', 'More Than One Race', 'Without Disability', 'Male'], ['Industrial Engineer', 'More Than One Race', 'Without Disability', 'Female']]). We then used this list of lists and turned it into a DataFrame to create an empty CSV, which we would then populate manually using Excel. The calculations we used to approximate the counts of each group are as follows. The total number of people for the occupation, sex, disability status, and race of interest were written down, in order to perform the calculations. Since the counts of those in each occupation were consistent across the three tables from our original data (Tables 9-5, 9-6, 9-8.), we used these counts as the baseline for calculating the proportions of each group within the occupation. We'll look at one specific group as an example to understand the calculations, let's do Asian Female Computer Engineers Without a Disability. From all the tables, we can see that the total count of electrical or computer hardware engineers (we refer to this occupation as Computer Engineers) is 412,000. Then, from Table 9-5, we find that there are 43,000 Females that are Computer Engineers for all degree levels (Bachelors, Masters, and PhDs). In 9-6, we see that there are 110,000 Computer Engineers that are Asian. Finally, from Table 9-8, for all Computer Engineers using all degree levels, there are 374,000 people that do not report having a disability.

Now that we have our totals, we can find proportions using probabilities. Sex, race, and disability status are assumed to be independent for our calculations, for the data that we are presented. Then, by the law of independence, we can say that the probability of someone being of race A AND sex B AND disability status C is the probability of being of race A multiplied by the probability of being of sex B multiplied by the probability of being of disability status C. In other words: $P(A \wedge B \wedge C) = P(A) * P(B) * P(C)$.

Lastly, we get to our calculations. Using our example above, the probability of being Asian in the occupation of Computer Engineers is P(Asian) = 110,000/412,000. The probability of being Female in the occupation of computer engineers is P(Female) = 43,000/412,000. The probability of being in the group that does not report a disability is P(Without Disability) = 374,000/412,000. Multiplying these all together results in $P(Asian \wedge Female \wedge WithoutDisability) \approx 0.02529537$. Finally, multiplying this proportion against our original total count of computer engineers gives us the approximate count of people for these three groups within this occupation, namely Female Asian Computer Engineers Without a Disability, about 10421.69384. We used Excel formulas that required us to input the total counts for each group to calculate the amount for each combination of race, sex, occupation, and disability status. An example of this calculation is shown in the image below. To use this final Excel sheet to make our visualizations, we reorganized the values read into the JS file from the Excel sheet into a JSON or dictionary-like data structure, based on the source code for our visualization, listed in our Acknowledgements section, as well as at this link: Original Visualization.

Legend for counts in formula						
With Disability Count	Male Count		Total Count for Occupation (Computer Engineers)			
Without Disability Co	Female Count		Count for Race (Asian)			
	MATH FOR VALUE FINDING:					
	Occupation Total Count	Race	Disability S	Sex	Type these values in the big final table	
Occupation	412,000	110000	38000	369000	9086.742389	Male/Race/With Disability
			374000	43000	1058.888679	Female/Race/With Disability
	R/WD/M	R/WOD/M	R/WD/F	R/WOD/F	89432.67509	Male/Race/Without Disability
Race Proportion	0.266990291	0.2669903	0.26699	0.26699	10421.69384	Female/Race/Without Disability
Sex Proportion	0.895631068	0.8956311	0.104369	0.104369		
Disability Proportion	0.09223301	0.907767	0.092233	0.907767		
TOTAL PROPORTION	0.0220552	0.2170696	0.00257	0.025295		
Count	9086.742389	89432.675	1058.889	10421.69		

Figure 3: Excel Calculation

5 DESIGN PROCESS

Below are three rough drafts of our final visualization. We utilized ideas from all three drafts and incorporated them into our final draft, which we will discuss below.

5.1 Bar Charts

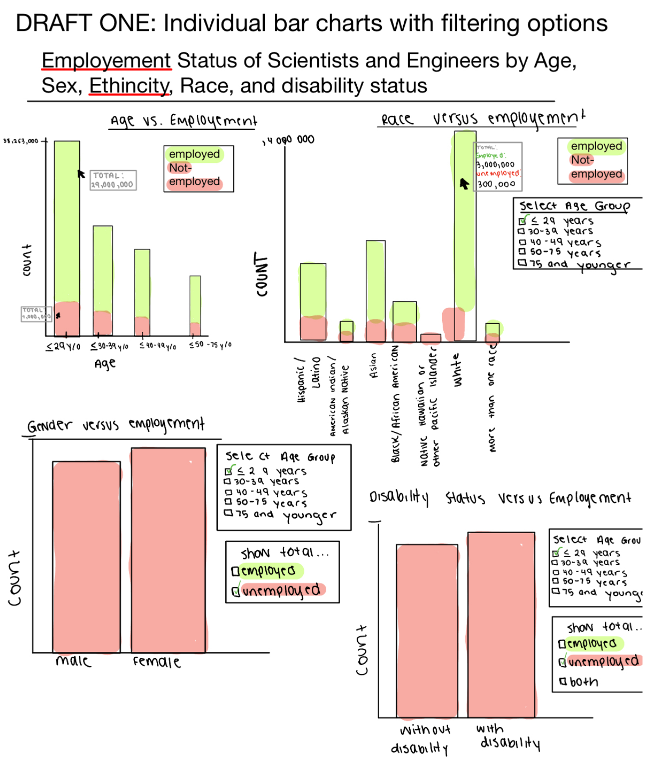

Figure 4: Rough Sketch 1

Our first draft included a series of individual bar charts, each representing a different characteristic, like age, race, gender, and disability status. The height of the bars represents the total items in our data-set within that respective demographic, and we include the counts of employed versus unemployed persons within that demographic as well. This draft allowed us to experiment with different filtering and hovering tools. Our main goal with this draft was to explore the best way to represent the amount of employed versus unemployed people in each group (represented by a bar). We began by experimenting with each bar separated into two groups: employed (represented by green) and unemployed (represented by red). We implemented toolboxes that appear when the user hovers over the bar, showing the count for each respecting group. We also experimented with filtering methods, like implementing a selection box allowing users to select which group, employed or unemployed, they would like to see the metrics for. Further, we added functionality that allows the user to visualize the metrics based on multiple characters, like gender and age. Ultimately, this draft introduced us to the idea of filtering based on multiple characteristics and utilizing bar graphs as a method of comparison, which are both things we apply to our final draft. We decided not to continue with this draft as we believed it to be too simple, and we wanted to explore other, more interactive visualization methods we have learned in class.

5.2 Tree Diagram

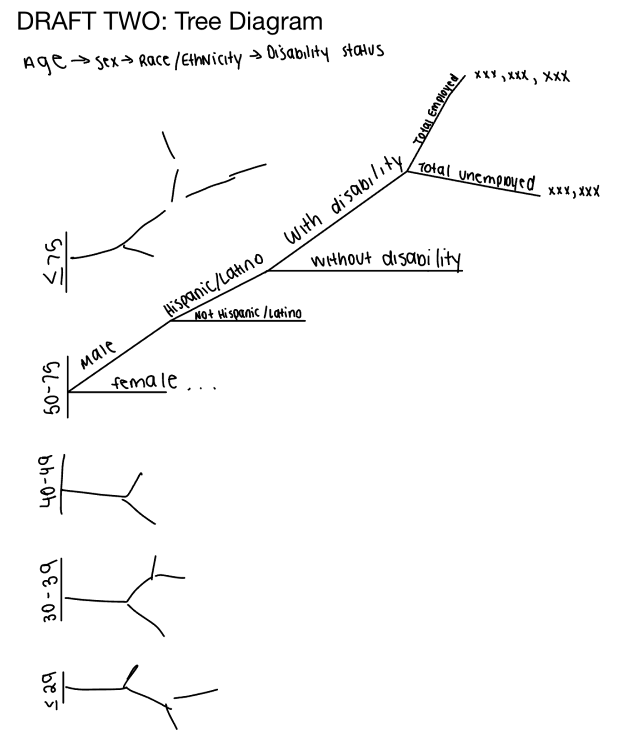

Figure 5: Rough Sketch 2

Our second draft (See Figure 4) was of a tree diagram, which allows users to follow the "branches" of groups they identify with or are interested in. Though it is not represented in the draft, we discussed the idea of implementing brushing into this visualization. Users would be able to click on a branch, and the visualization would focus on the group they selected. This draft really solidified our desire to implement conditional filtering, something that will be difficult, but not impossible, with the data we are working with. However, we felt the tree diagram to be a bit too restrictive and uninteresting. For instance, each branch is splitting off into two, but for categories like 'race' for instance, there are many more groups than just 'Hispanic/Latino' and 'Not Hispanic/Latino'.

5.3 Tree Map

Our third draft (See Figure 5) is the one we most closely emulated our final draft after. This idea is a tree map, implementing selection and brushing, two ideas we have discussed in class. The user starts by selecting their state. With each selection, a new category comes up, and the user can select which group they most identify with. Each

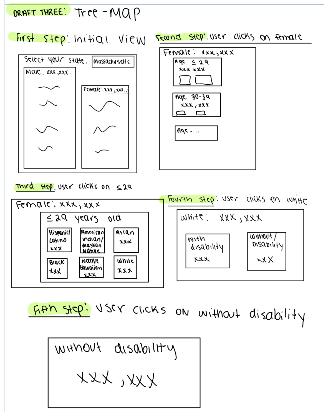

Figure 6: Tree Map

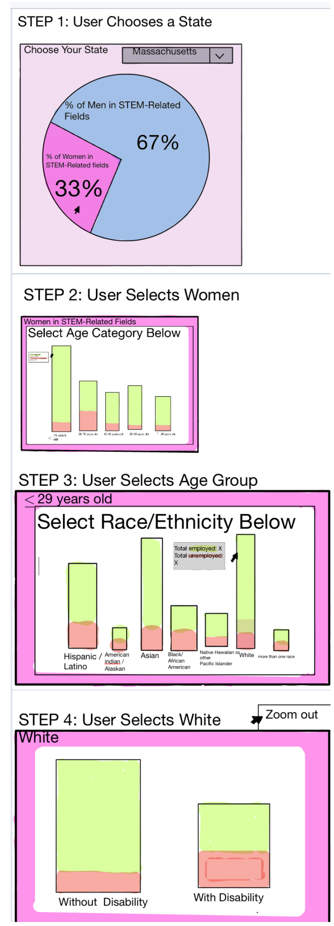

Figure 7: Polished Design

selection would show the number of employed people in STEM-related fields within that group.

We instantly knew this draft was our most promising one, but it still needed some tweaking. We felt that it lacked the element of comparison, something draft one so clearly laid out with the sizes of the bars. We also knew that we would be facing a challenge: conditional values. Despite these challenges, we decided to proceed with this draft and formulate our final design.

5.4 Polished Design

Our final draft (See Figure 6) is user-friendly, interactive, and informative. Like draft 3, it is a tree map, but we took our considerations and hesitations regarding draft three and made the changes necessary to make the most effective visualization. Our user begins by choosing their state from a drop-down menu. A pie chart showing the proportion of men and women in STEM-related fields in that state appears. The user then clicks on their group of interest, and a brushing tool zooms in on that group and displays separate bar graphs representing age categories. Using a similar idea from graph one, we decided to implement bar graphs to highlight comparison, something we failed to do in draft 3. We also utilized channels like color to show the comparison between unemployed and employed people in each group. We implement a toolbox that shows the total employed (represented by green) and total unemployed (represented by red) people in that group. As we proceed with this project, we will discuss if it is necessary to include a y-axis despite the toolbox. We will also consider adding a 'total' number in the toolbox as well. As the user moves along, they can select which age category interests them, and they are brought to another frame: a comparison of unemployed versus employed people among each race/ethnicity category. The marks in this visualization, just like the 'age' frame, are lines. We utilize channels like vertical position to signal comparison, and color to signal categories. As the user progresses, they once again choose their category of choice, and they are brought to disability status, which marks the last step of our draft. Users have the option to zoom out and start from the top.

We recognize the changes that need to be made to this visualization and the challenges that come with it. For instance, we want to find a way to implement education status, despite it not being in the same data set as the other categories. We also recognize that we will need to implement a statistical model that calculates the conditional values, and with this, we might need to make a shift from showing values to showing proportions. As we proceed, we plan to implement some more creative ways of highlighting comparisons instead of simple bar graphs. Overall, though, this final draft is a great start on our path to making the most informative, user-friendly, interactive visualization.

6 Evolving to the Final Design

We worked with several separate datasets, each exploring different aspects of employment rates in STEM fields. One focused on age, another on states, a third on sex (male and female), a fourth on different race and ethnic groups, and a fifth on disability status. We experimented with different visualization techniques, including bar charts, tree diagrams, and tree maps, and while our sketches varied somewhat, they shared similar priorities. Our primary objective was to interweave the various characteristics - age, state, sex, race/ethnicity, and disability status - in a way that would allow users to feel as accurately represented as possible. As humans, we each belong to numerous different groups, and reducing a person's identity to a single characteristic, such as age, state, sex, race/ethnicity, or disability status, would not fully convey the complexity of intersectional discrimination. All of our sketches shared certain common features, such as brushing, zooming, tooltips, and grouped bar graphs. Two major concerns arose for us: how could we best create conditional values from these four distinct datasets to accurately represent the user, and how could we best serve the user through our visualization? To better serve the user, we chose to shift our focus from overall employment rates in STEM fields to a more detailed analysis of how these rates vary across demographic groups and occupations. For example, is a Black woman without a disability better represented in one occupation over another? If a user is deciding between two fields, our visualization tool could potentially help them choose based on which field they feel more accurately represents them. This led to the creation of our final visualization, which emphasizes a personalized representation of specific individuals within various STEM-related fields.

7 Final Design

The final visualization evolved into a three-layer hierarchical bar and pie chart. The initial phase of the visualization displays the breakdown of people in STEM-related fields for each race/ethnic group by their sex. The races/ethnicities looked at include African-American, American Indian, Asian, Latino, More than One Race, Native Hawaiian, and white. The sexes are male or female. The sex of the group is denoted using color, with blue representing male, and pink representing female (Figure 8). Upon clicking the bar representing the sex-group for the race the user wants to view, the visualization shifts, now showing the employment metrics for people of the chosen race and sex. These employment metrics are broken down into their respective disability status counts for each STEM-field occupation. The occupations considered in the visualization are Social Scientists, Mathematical Scientists, Psychologists, Mechanical Engineers, Computer Engineers, Civil Engineers, Industrial Engineers, and Computer and Information Scientists. This final bar graph represents the counts of individuals in each occupation with the selected race and sex, comparing the number of those with disabilities to those without (Figure 9). The final bar graph gives way to a pie chart when the bars are clicked, providing an alternate view of the proportion of disabled and non-disabled status in the chosen occupation, race, and sex(Figure 10). To go back to the initial chart at any stage, you can click the "BACK BUTTON." The visualization also features a note explaining the lack of visible data for American Indian and Native Hawaiian people.

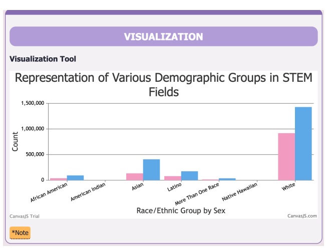

Figure 8: First Layer of Visualization

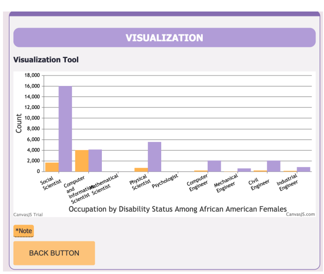

Figure 9: Second Layer of Visualization

Our target user is a minority exploring their options in the hopes of finding a STEM occupation where their demographic is adequately represented. This user will start by selecting the bar representing the data of their race and sex. Suppose the user is an African-American female, upon selecting the bar she will see the numbers of employed African-American females in 5 different occupations broken down by disability status. She may decide to pursue being a Social Scientist due to there being the highest amount of representation of her demographic in that field according to the visualization. She may also decide to pursue being an industrial engineer due to the lack

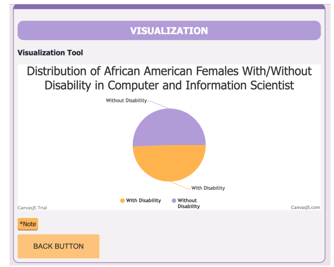

Figure 10: Third Layer of Visualization

of representation of her demographic. Suppose the user also has a disability, she may decide to pursue a career path within Computer and Information Science due to a seemingly high number of those employed in that field with a disability. The user selects Computer and Information Scientists and finds that the number of African American females with disabilities is relatively equal to the number of African-American females without disabilities in the field. The user decides to look into Computer and Information Science, intrigued by the representation of her demographic.

8 Usability Testing

Our team is conducting usability testing for our interactive visualization, which aims to address the under-representation of minorities in STEM-related fields. Prior to using our visualization, participants will be given a pre-survey that includes basic demographic questions to identify the diversity of our participants and potential areas where our visualization may lack coverage. The survey also aims to gather opinions and attitudes toward diversity and inclusion in STEM fields. Participants will be asked about their gender identity, ethnicity, field of study or area of expertise, and their opinions on the availability of resources and support for minority groups in STEM fields. They will also be asked to identify two barriers that are most evident for underrepresented groups in STEM fields, and to what extent they agree with statements about the barriers facing people with disabilities and the importance of diversity and inclusion in STEM fields. Participants then interacted with the visualization and provided feedback on its effectiveness in our post-survey. From our post-survey results, we learned that the visualization helped many participants better understand the employment situation for minority groups in the STEM field and provided insights that they were not previously aware of. However, some participants also indicated that more demographic groups should be included in the survey to gain a more comprehensive understanding of diversity and inclusion in STEM fields. Additionally, some participants suggested that we should fix the labels for each of the linked visual encodings to make it easier for the user. During this aspect of our usability testing, participants will be asked to perform specific tasks using our visualization, and we will collect data on their performance and experience. Additionally, we will ask participants to provide feedback on the visualization, including its ease of use, clarity of information, and effectiveness in addressing the under-representation of minorities in STEM fields. These tasks will be designed to test the effectiveness and usability of the visualization for the specific demographics and needs of the participants. This data will be used to improve and refine the visualization to better meet the needs of our target audience and promote diversity and inclusion in STEM fields. Overall, the usability testing helped us to identify areas where we can improve the visualization and better serve the needs of our users. By incorporating this feedback, we can create a more comprehensive and user-friendly tool to promote diversity and inclusion in STEM fields.

9 Discussion

Our final visualization tool provides a clear overview of the representation of different demographic groups in nine STEM-related fields, but we recognize that there is room for improvement. Our main goal was to be as representative as possible. Discussing the under-representation of different demographics while inadvertently excluding certain groups is counter intuitive. Unfortunately, certain demographic groups such as American Indians and Native Hawaiians are not fully represented in our graph due to data collection issues, resulting in missing bars that affect the accuracy of our analysis. Ideally, we would have had access to the data for these groups so we could avoid excluding any individuals. Further, we only chose nine STEM-related fields to represent in our visualization, even though the datasets we utilized had information on over 40. While we chose these nine occupations for their broad applicability, we recognize the need for greater specificity in our selection of occupations to better personalize the experience for the user. Moving forward, we would prioritize data that accurately represents the diversity of the population, including a wider range of gender identities, and differentiating between physical and cognitive disabilities. Finally, as reflected in our original drafts, we wanted to utilize state as a level of our visualization, but it was quite difficult with the data we were given. It's important to take into account location when discussing discrimination as different parts of the United States have nuanced issues with discrimination. Overall, we felt we did a great job with the data that was given to us. We creatively combined 3 separate datasets into one using conditional formatting, and ideally, there will be data available in the future that has this conditional formatting in place as is, and that may potentially include geographic data that we can use to extend our visualization further.

10 Conclusion

This project illuminates the reality of discrimination against underrepresented demographic groups in STEM-related fields. Our visualization offers a tangible depiction of the existing disparity we all know too well. As female STEM students at a predominantly white institution, we recognized the need to create a tool that enables individuals from diverse backgrounds to easily visualize their group's representation in various fields. The benefits of this tool extend beyond Northeastern to the entire United States. We hope that such tools encourage a society that is more aware of workplace discrimination and holds businesses accountable for creating more equitable and inclusive environments.

Acknowledgments

We would like to thank Professor Ab Mosca, and our TAs for Data Visualization.

References

[1] Science amp; engineering indicators, 2018.

[2] M. G. Ballatore, L. Barman, J. De Borger, J. Ehlermann, R. Fryers, K. Kelly, J. Misiewicz, I. Naimi-Akbar, and A. Tabacco. Increasing gender diversity in stem: A tool for raising awareness of the engineering profession. In *Proceedings of the Seventh International Conference on Technological Ecosystems for Enhancing Multiculturality*, TEEM'19, p. 216–222. Association for Computing Machinery, New York, NY, USA, 2019. doi: 10.1145/3362789.3362832

[3] E. G. Deitz and Steven. Diversity and stem: Women, minorities, and persons with disabilities, Jan 2023.

[4] A. Okrent and A. Burke. The stem labor force of today: Scientists, engineers, and skilled technical workers, Aug 2021.

[5] C. C. Vieira and M. Vasconcelos. Using facebook ads data to assess gender balance in stem: Evidence from brazil. In *Companion Proceedings of the Web Conference 2021*, WWW '21, p. 145–153. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3442442.3453456