Phyllis Chen

Introduction

My interests in researching the field of disability stems from my considerations during job search and thinking about the accessibility of job search. I noticed how often when applying to jobs, there are additional questions about disability status in the initial applications. This led me to think about how diverse experiences of individuals with disabilities and societal factors could influence their education levels, salaries, or chances of employment. Thus, I decided to search for references online that could help guide these curiosities. In finding relevant research prior to starting my investigation, I searched for information on ways that disabilities shape all aspects of people's lives.

This includes focusing on the intricate relationship between disability, economic conditions, and healthcare access, and by drawing insights from reputable sources. Two key studies that helped shape my initial understanding of these dynamics. The first, "Economic Conditions and Disability of Older Adults," available on PubMed Central, discusses the pivotal role of economic conditions, particularly income, in influencing disability among older adults. The second study highlights disability-specific disparities in healthcare access, emphasizing the multifaceted challenges faced by individuals with disabilities. Specifically, they concluded that "economic conditions (income) plays dominant role" in disability of old adults (NIH) and how there are many "disability-specific disparities in health care access" (CDC). Finding inspiration from this available research, I wanted to explore different correlations based on different disabilities individually instead of altogether and how they weigh into results of different demographic factors.

I decided to focus on identifying potential correlations between the occurrence of various disabilities and specific demographic characteristics. By analyzing data on disability prevalence alongside these factors, I can identify any patterns or draw conclusions based on the data presented. My first dataset is drawn from the CDC, which covers a summary of Disability and Health Data System. This includes self-reported functional limitations in six categories of disabilities and focuses on adults over the age of 18. My next dataset is a Household Pulse Survey collected by healthdata.gov by the U.S. Census Bureau, with self-reported limitations and experiences on vaccination along with disability data. Although vaccination status is not a common demographic factor usually observed with disabilities, I decided that using the first dataset can focus on disabilities and demographics, while the second dataset can offer insight on how the COVID-19 pandemic possible impacted people with disabilities based on vaccination rates.

Thus, I proposed the research question: *How does the prevalence of disabilities correspond to demographic factors such as employment status, vaccination rates, educational attainment, and socioeconomic backgrounds within a given population?*

Methods

To begin my analysis, I first read through my datasets and skimmed through the information given, while also using the given CDC resource with explanations for the columns (DHDS - Prevalence of Disability Status and Types | Data | Centers for Disease Control and Prevention (cdc.gov)). Prior to starting, I loaded in the two packages that I would use throughout this exploration, which were `tidyverse` and `ggplot2`. My first dataset from the CDC had a lot of content, and I started off with this one. By reading in the column names, I determined that to clean my data, it was necessary to take out variables that were not relevant. I felt the overflow of information would hinder my efficiency to analyze this dataset. I decided to remove columns such as row ID, the location, the data source, category IDs, geolocation, and etc. These columns, while useful, were not helpful to my analysis focusing on the disabilities listed and demographic factors. After this, I decided to remove rows with NA values in the cleaned data, so I can maintain integrity of the analysis later. After, I selected relevant columns that will have relevant information on either demographic information or disability prevalences. I decided these would be my key variables: `Category`, `Indicator`, `Response`, `Data_Value_Alt`, `Number`, `WeightedNumber`, `Stratification1`, and `StratificationCategoryID1`. From searching through the key online, I discovered that the `Data_Value_Alt` is the percentage of individuals with a particular disability after adjusting for age. It provides a standardized measure that considers differences in age distributions across different groups or categories. The `WeightedNumber` variable is the count of individuals with a specific condition, considering the survey weights. It provides an estimate of the total number of affected individuals while considering the sampling design and population. These two columns stood out to me because `Data_Value_Alt` is a percentage that accounts for age differences and proportions of across different groups of disabilities, while `WeightedNumber` is an estimated count that considers survey weights for accurate representation in the survey data. These would be two different columns that I focus on during visualizations and analysis to draw meaningful conclusions with different disabilities.

I then decided to focus only on the demographics category. I filtered the dataset to include only "Demographics" information. By only looking at the unique responses, I recognized that the unique responses for observation were variables such as education level, employment status, salary, and marital status. So, I created a new data table to be used throughout for visualization, including key demographics such as salary levels, employment status, and education levels for comparing disability prevalence, considering their relevance and potential impact on disability status. I believed that marital status would not give me a lot of insight on disability prevalence, so I removed it. Next, to perform grouping and stratification, I filtered and grouped data based on disability types of interest, including "Self-care Disability," "Mobility Disability," "Independent Living Disability," "Vision Disability," "Hearing Disability," "Cognitive Disability," "No Disability," and "Any Disability."

After completing a selection of key variables and grouping of data, I made visualizations to find any trends and patterns in disability prevalence across different groups and different disabilities. I did general variable grouping and factorization to then begin my analysis. I made bar plots with different types of disabilities and their employment statuses, salary levels, and highest levels of education received for the 6 different disabilities. I also made visualizations of all disabilities compared to no disabilities, so that I can see a general overall comparison of the impacts of

disabilities in addition to looking at disability types individually. From practices during class, I discovered that stacked bar plots are great to visualize the relationship between disability types with the demographic factors because they offer a comprehensive understanding of these relationships, and seeing the disabilities compared side by side and stacked together gives a lot of insight for what stands out or what patterns exist.

After visualizations, I decided to use hypothesis testing using t-tests and ANOVA tests for comparing prevalence between the groups of employment status, income, and education levels. Using t-tests allowed me to compare the means of two groups, and I focused on performing t-tests on the `Data_Value_Alt` variable to see if there is a difference between "Any Disability" and "No Disability" groups in their means. From these t-tests, I could look at education level comparisons, salary level comparisons, and employment status comparisons. I also decided to perform AONVA testing on specifically the salary levels, by looking at the interaction effects of `Stratification1` and `Response`, which will allow me to assess whether the relationship between disability types and prevalence on different salary ranges. I chose to do this because salary ranges gave me sufficient data and different salaries to perform this testing well. If I looked at the interaction of the two variables and the dependent variable, which was `Data_Value_Alt`, I could decide if the relationship between disability types and their prevalence is consistent or not with all salary ranges. So, I multiplied the two predictors or independent variables for the interaction.

Next, I will be discussing methods for my next dataset, (https://healthdata.gov/dataset/Household-Pulse-Survey-HPS-COVID-19-Vaccination-am/a8aa-yctt/about_data). There was a lot less data cleaning and wrangling involved because this dataset was a lot more limited and simpler. At first, I wanted to combine the vaccination dataset with the previous disability prevalence dataset from the CDC as they both included specific disabilities, but I was not sure if these disabilities were defined the same way, as the vaccination dataset had only five groups of disabilities instead of eight. I determined that if I combined them, it would disregard the disabilities that were not in another dataset, which would make my results more biased. So, I decided to look at this data separately and try to see if I can draw the same conclusions.

In the vaccination dataset, I cleaned the data to only include various vaccination levels and disability status for visualizations. I categorized those with and without disability to make grouped bar charts for Vaccination statuses of the surveyed individuals. This lets me compare vaccination status with the visualizations I made with the first dataset to see if there were similar conclusions for disability vs. no disabilities. I then performed t-tests to compare the means of vaccination rates between individuals with and without a disability. I repeated this for the 5 disabilities with data available in this dataset. I can draw conclusions based on the significance levels of these t tests to see identify patterns in vaccination rates and disability prevalences for various disabilities and see if they can compare to any of my previous results with demographic factors. Lastly, I decided to do a logistic model for the vaccination rates. I created a new table that mutates vaccination values so that they are binaries. This means that vaccinated and boosted individuals will be viewed as "Vaccinated", and the remaining individuals surveyed would have the value of "Not Vaccinated". Using this table, I used the formular to create a logistic regression

model to see the binary outcome variable with respect to the five different disability types of vaccinated or not vaccinated.

Results

Figure 1 and 2 Comparing Salary amounts with age adjusted prevalence % and Weighted Numbers

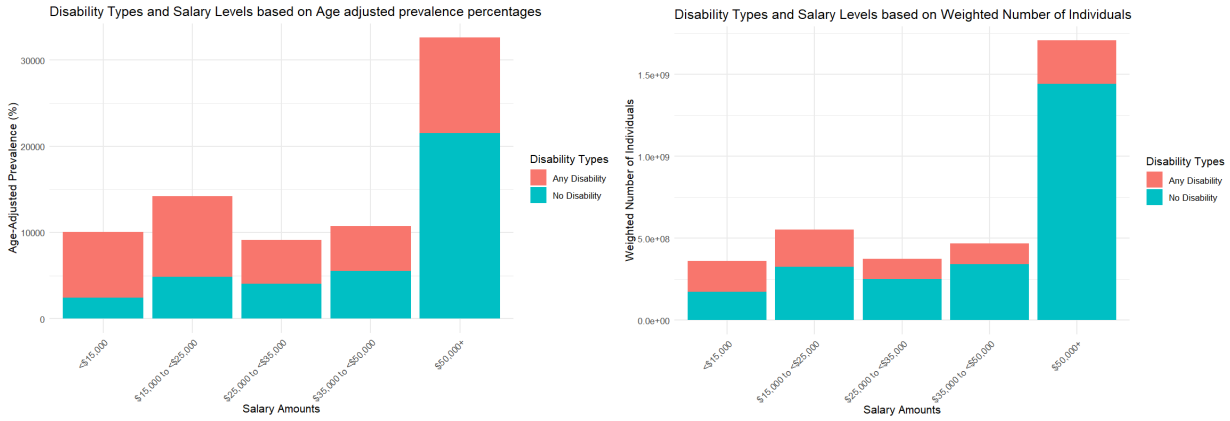


Figure 3 and 4 Comparing Education levels with age adjusted prevalence % and Weighted Numbers:

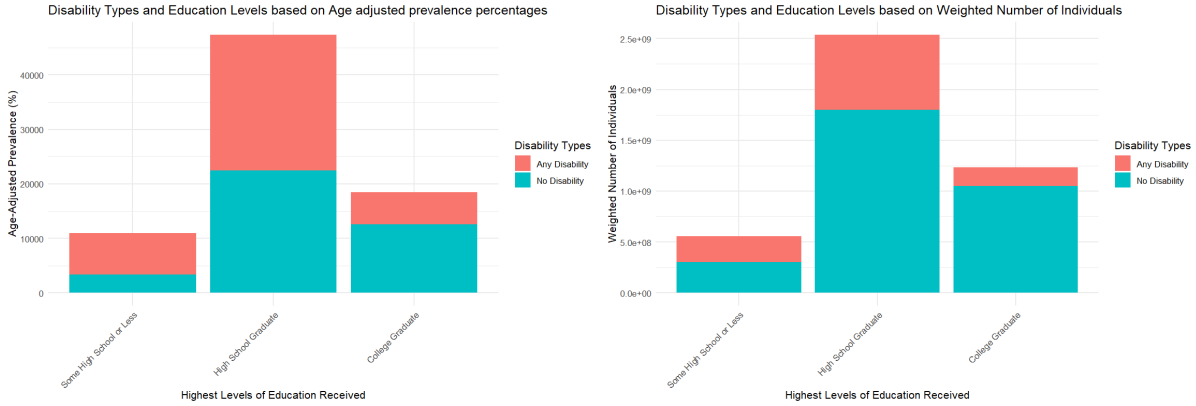
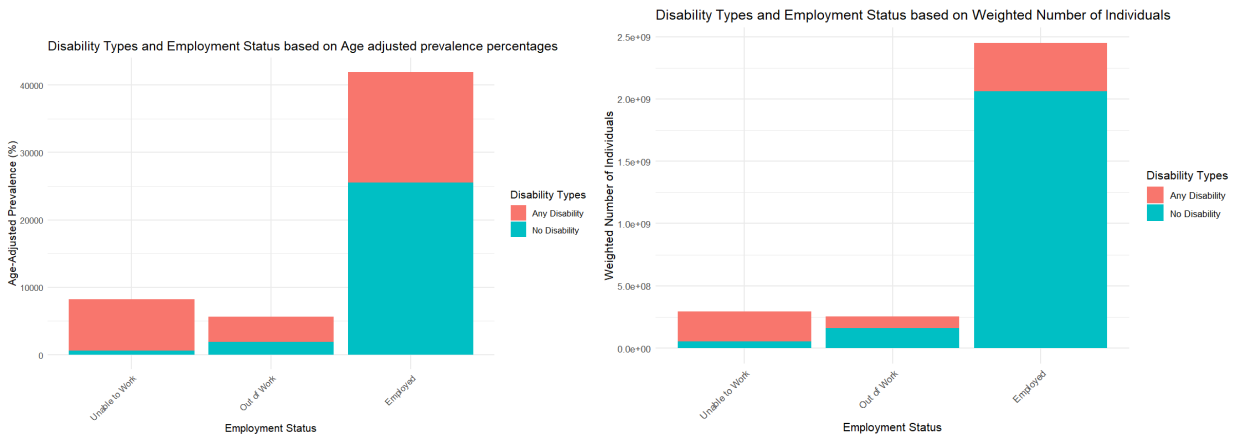


Figure 5 and 6 Comparing employment statuses with age adjusted prevalence % and Weighted Numbers:



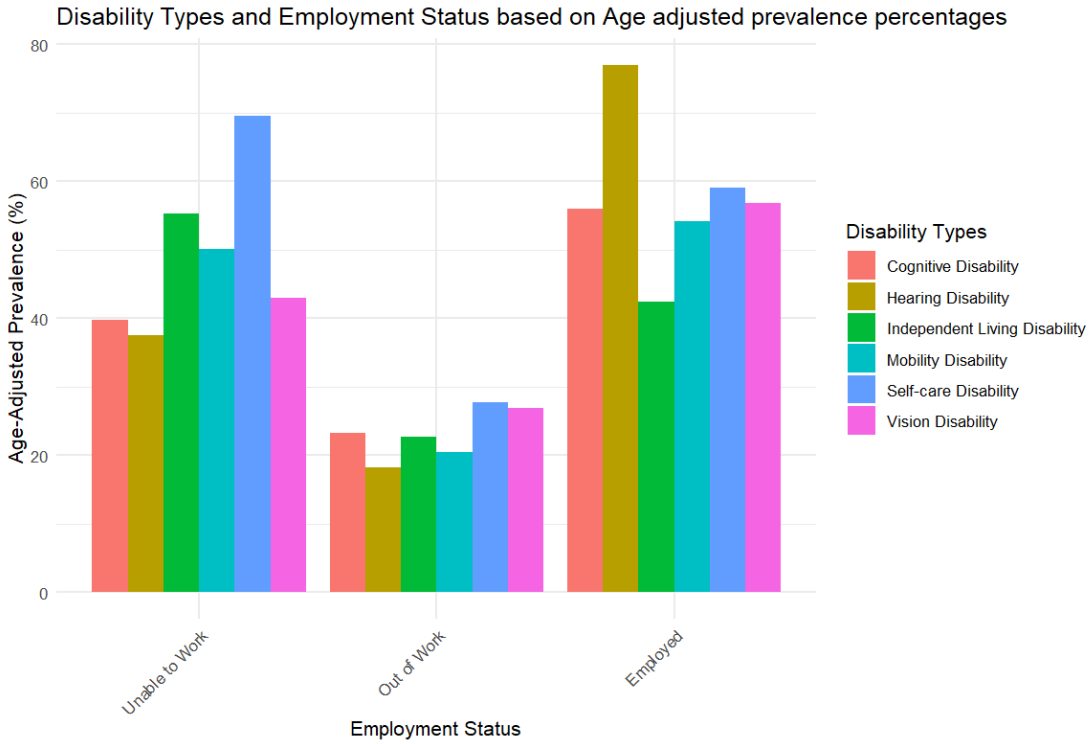
Based on Figures 1-6:

By looking at figures 1-6, I noticed that there were not a lot of differences in the graphical representations of age adjusted prevalence % and Weighted Numbers, so I continued to only use age adjusted prevalence % in my next visualizations. I see that for salary amounts in Figures 1 and 2, in the highest salary category, individuals with no disabilities have a lot higher salaries than those with any disabilities. For the lowest salary category, there were way more individuals with disabilities having lower salaries than those with no disabilities.

In Figures 3 and 4, in levels of education received, there were a lot more individuals with no disabilities being college graduates than those with any disabilities. For the level of education being high school graduates, there was generally more individuals without disabilities than those with disabilities. However, for the category of some high school or less, there was fairly even distribution of individuals with and without disabilities.

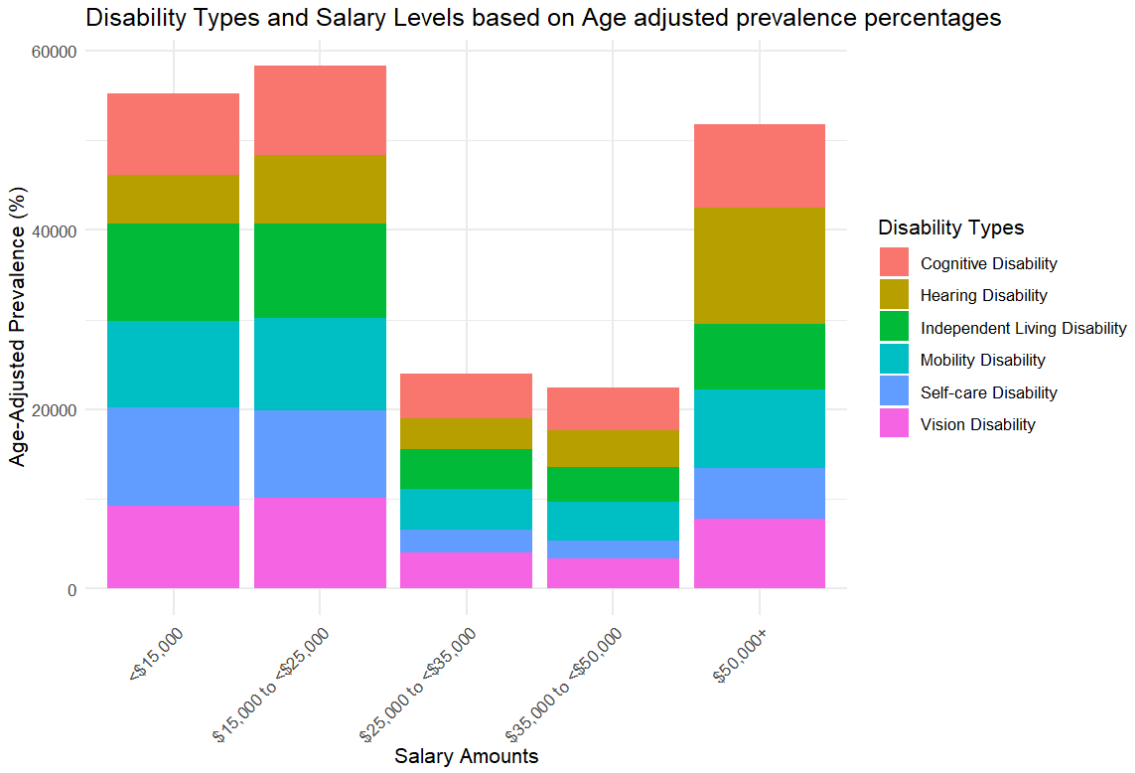
In Figures 5 and 6, in employment statuses, many more individuals with disabilities are unable to work compared to individuals without disabilities. For those currently employed, there are a lot more individuals employed without disabilities than with disabilities.

Figure 7 Only using age adjusted prevalence % for specific disability types with employment status:



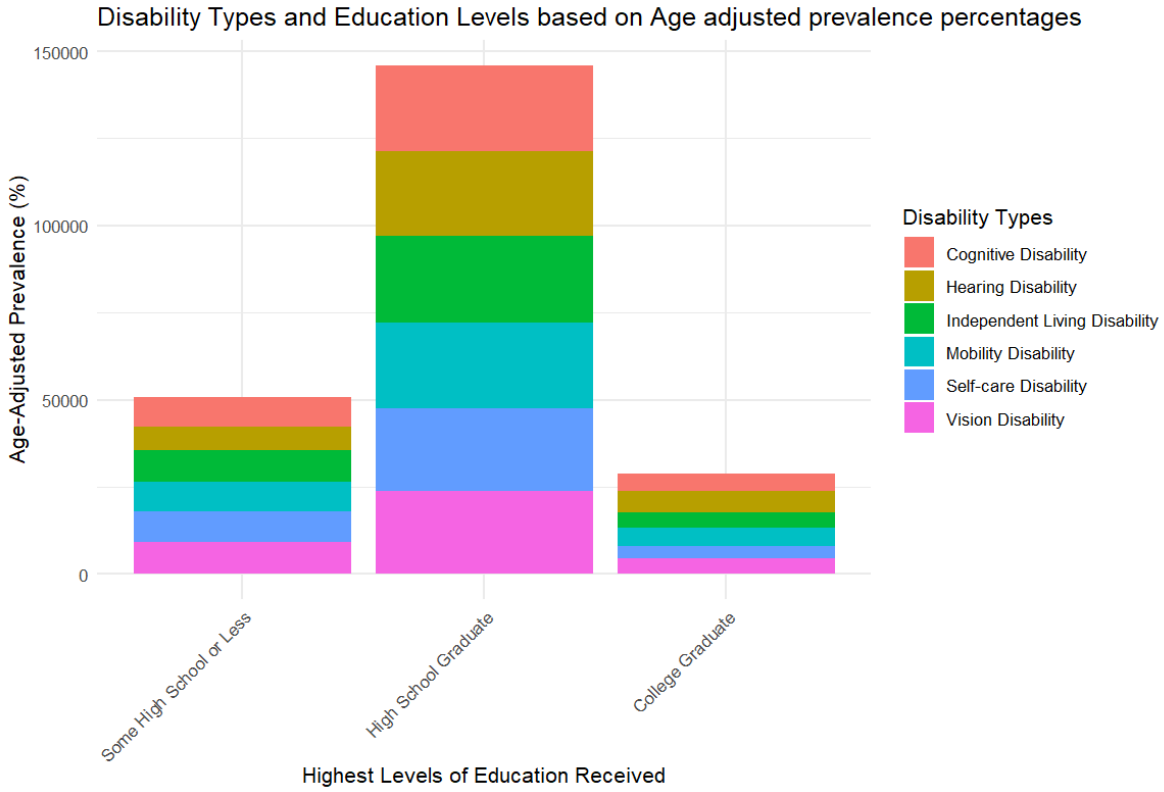
In Figure 7, I observed that there were more individuals with self-care disabilities that are unable to work, and the least amount of disability type employed were individuals with an independent living disability.

Figure 8 Only using age adjusted prevalence % for specific disability types with salary amounts:



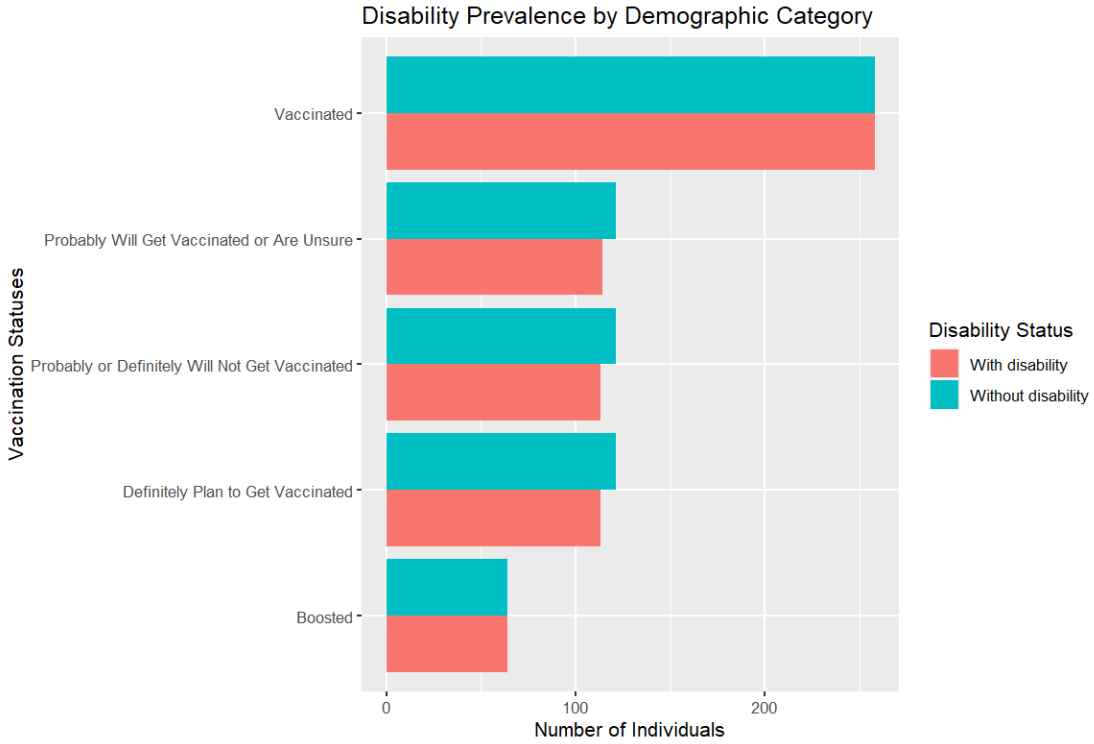
In Figure 8, I observed that there was a great disparity between the two middle salary amounts. Those with disabilities had higher prevalences in very low salaries or very high salaries, and there were less in the mid-range. However, it seems that hearing disabilities in both salary amounts and employment were the highest in having employment as well as higher salaries. It is possible that hearing disabilities impact these factors the least, while independent living or selfcare disabilities affect employment and salary amounts the most.

Figure 9 Only using age adjusted prevalence % for specific disability types with education levels:



In Figure 9, I observed there were a great amount of individuals experiencing disabilities that did not receive higher education past high school. For all disability cateogires, there were very low numbers in prevalence for college graduates. Most of this data was concentrated at high school graduate education levels, and some in some high school or less, but very little for each type in college graduate categories.

Figure 10 Using vaccination dataset to compare vaccination statuses with disability prevalence by disability status



In Figure 10, I observed that for people that were vaccinated, there was fairly equal amounts of people vaccinated whether or not they were individuals with disability statuses. However, for those that are not currently vaccinated, there were generally more people without disabilities than those with disabilities.

Discussion

By looking at the visualizations overall, I could determine that there are clear impacts on disability prevalence with socioeconomic outcomes and demographics factors. A significant disparity exists with higher concentrations of people with disabilities occupying lower income brackets. This could mean a potential link between disability and limited earning potential. Furthermore, the visualizations indicate a correlation between disability and educational attainment. Individuals with disabilities are less likely to hold college degrees compared to the general population. This disparity could be attributed to various factors, such as accessibility challenges in educational institutions or the impact of specific disabilities on learning to pursue higher education. The visualizations also allow me to draw conclusions regarding employment. The data shows a considerably higher unemployment rate among individuals with disabilities compared to those without. This could mean potential workplace accessibility issues or limitations imposed by specific disabilities that hinder employment opportunities.

When looking at specific disability types, I was able to identify a few trends. Individuals with self-care and independent living limitations appear to face the more significant challenges, reflected in their highest unemployment rates and lowest salaries. This suggests a substantial

impact of these specific disabilities on work capabilities and earning potential. However, the data also suggests that individuals with hearing impairments might fare slightly better. Compared to other disability types, they seem to have higher employment rates and potentially higher salaries. This observation might indicate that hearing disabilities have a lesser impact on work capabilities compared to other types.

Given my observations and analysis, I can make connections between disability and various socio-economic and demographic factors. I do see that by only looking at economic outcomes based on my visualizations created, there were patterns in salaries for those experiencing disabilities. This aligns with my initial literature review, which states that "disability is associated with bad economic outcomes…declines in income and consumption for the chronically and severely disabled group are more than twice as large as those experienced by average disabled individuals" (NBER).

However, when looking at the results of my t-tests, I realized that there is no significant evidence to suggest that the mean of the `Data_Value_Alt` variable (the age-adjusted prevalence) is different between the "Any Disability" and "No Disability" groups for employment statuses. The p-value is higher than the typical significance level of 0.05. This means that the observed differences in means are likely due to random variation, and there's no strong evidence to suggest a true difference in the age-adjusted prevalence of the measured condition between these two groups. I came to the same conclusion for the education level comparisons, where the p-value is 0.9991. This means there is no significant difference in age-adjusted prevalence between individuals with any disability and those without a disability, based on education levels. Given a p-value of 0.9606 for education level comparisons, I was able to reach the same conclusion.

Next, I observed the results of the ANOVA formula, where I test whether the effect of one variable (`Stratification1`) on the dependent variable (`Data_Value_Alt`) is different depending on the level of the other variable (Response), and vice versa. In other words, I look at whether the relationship between disability type and prevalence varies across different salary ranges. I observed from the results of the ANOVA testing that the interactions between stratification and the data values are large, and the associated p-values are very close to zero. This means that there both disability types and salary ranges play a significant role in determining the outcome (Data_Value_Alt). Additionally, the relationship between disability types and the outcome depends on the salary range, and vice versa.

The results from the t-tests were different from the graphics I created, and I believe it may be because the categories of variables have different numbers of observations. The absence of significant differences in age-adjusted prevalence implies I should not draw definitive conclusions based on mean comparisons alone.

I was drawing different conclusions based on the visualizations compared to the t-tests. This helped me realize that while visualizations provided valuable insights and trends, t-tests emphasize the importance of statistical significance. There are many factors that could contribute to the relationship between disability and demographic factors, and to be able to reach a conclusive answer about the connection between the two, I would need similar results in visual patterns as well as significance in statistical calculations. Yet, I believe that if I looked at

disability types individually such as when looking at the ANOVA test results, I can confirm that salary ranges are significant for disability types. While t-tests provided me with initial comparisons, ANOVA results reinforce the dynamics between disability types and socioeconomic factors and help validate the conclusions from my visualizations.

This helps me answer my research question partially, yet I am unable to fully answer the impact of disability prevalence on the factor of vaccination rate. In only looking at visualizations, vaccination rates amongst individuals with and without disabilities seem relatively similar for those who are vaccinated. However, a higher proportion of unvaccinated individuals appear to be without disabilities. This was an interesting observation, and I would need to perform more research. I am unsure of the reasons behind correspondence or possible links between vaccination rate and disability prevalence. However, I was able to identify these trends in the data and this is a possible limitation that needs further observation.

Another limitation overall is that these observations highlight correlations, not necessarily causal relationships. I would need to perform further analysis to explore the factors driving the observed disparities for disabilities. Moving forward, I would like to do further research that dives into the reasons behind the observed disparities in employment, salary, and education based on disability status. I made a few assumptions based on the observations I saw and the references in articles I discovered from NBER. However, perhaps I could be investigating existing policies and programs aimed at addressing the challenges faced by individuals with disabilities in the workplace and education system for better contexts. I was not able to find any specific studies like this, which was why I made the assumptions based on my analysis of the data and visualizations. Additionally, I would want to explore potential explanations for the observed vaccination patterns among individuals with and without disabilities. I have not been able to find many articles between this factor and vaccination rates, and the dataset is very limiting and lacks room for me to do further analysis. Recognizing the multifaceted nature of the challenges faced by individuals with disabilities, I am motivated to continue more detailed examination of different disabilities individually. I look forward for further investigation, which will help pave the way for developing effective solutions to address the existing inequalities.

References

Data.cdc.gov. “Household Pulse Survey (HPS): Covid-19 Vaccination among People with Disabilities.” *HealthData.Gov*, 6 Aug. 2021, healthdata.gov/dataset/Household-Pulse-Survey-HPS-COVID-19-Vaccination-am/a8aa-yctt/about_data. Accessed 11 Mar. 2024.

“Disability and Health Data System (DHDS).” *Centers for Disease Control and Prevention*, data.cdc.gov/Disability-Health/Disability-and-Health-Data-System-DHDS-/k62p-6esq/about_data. Accessed 11 Mar. 2024.

Liu, Han, and Meng Wang. “Socioeconomic Status and ADL Disability of the Older Adults: Cumulative Health Effects, Social Outcomes and Impact Mechanisms.” PubMed Central, 10 Feb. 2022, www.ncbi.nlm.nih.gov/pmc/articles/PMC8830695. Accessed 11 Mar. 2024.

“The Prevalence and Economic Consequences of Disability.” *NBER*, www.nber.org/bah/2013no1/prevalence-and-economic-consequences-disability. Accessed 11 Mar. 2024.

“Prevalence of Disabilities and Health Care Access by Disability Status and Type among Adults - United States, 2016.” *Centers for Disease Control and Prevention*, 16 Aug. 2018, www.cdc.gov/mmwr/volumes/67/wr/mm6732a3.htm. Accessed 11 Mar. 2024.