Diversity in Tech Companies
Stat 346
Emili Robles
Oct. 21 2024

Dataset **GithubCode** **Shiny App**

Introduction

For this shiny portfolio project, I chose to use the data based on diversity in tech companies found on Kaggle. This dataset provides demographic information on employees at various tech companies, specifically focusing on their cultural and gender diversity from 2014 to 2018. I wanted to work with this dataset because of my personal background—my ethnicity and culture are a significant part of my identity. As a first-generation Latina pursuing a career in tech, I recognize the importance of advocating for diversity within this space.

By analyzing this dataset, I want to understand not just the numbers, but how these numbers reflect the realities of cultural and gender diversity across different companies. Projects like this are a way to raise awareness and remind ourselves that there's still much work to be done to create an inclusive tech industry.

Some key questions I want to explore include: “How do gender disparities manifest within the Hispanic and Black demographics?” “Which companies are leading the way in fostering inclusivity?” and “What steps can we take to make tech companies more inclusive going forward?” My goal is to shed light on where representation is lacking and help drive efforts toward a more inclusive space.

Interesting Facts

The biggest takeaway from this dataset was the overrepresentation of White employees across most tech companies, with some of them showing as high as 70% - 80% White employees. An unexpected finding was the fact that Asian representation looks to be somewhat of a competition when going against the percentage of White employees. But there are some companies like Yelp and HP who fall low on Asian representation for some reason. Overall, looking at the data we can almost right away see that ethnic groups like Black and Latino communities are extremely low.

Creating the Interface

The interface was made using Shiny with a layout consisting of a sidebar panel for user input and a main panel to display the bar chart and the data tables. The user is able to select a range of

years or just look at one year, and the graph and data dynamically update based on the selection made.

I first had to clean the data though because I had to convert it to numeric data types to be able to plot the percentages properly. I also had to rename the columns because the csv was messy with reading '%', as well as reshaping the data with pivot_longer() to plot the ethnicity percentages for each company.

When it came to the style customization, I used theme_minimal() to help keep the graph uncluttered as well as using a color palette set3 because it provided a soft look to the graph. The company names were also overlapping each other so I used the rotated x-axis labels to prevent this.

Reactive Graph