Data Visualization Project Proposal

Team Members: Yuting Mei (ym2741), Renyuan Chen(rc3373), Xingchen Li(xl3005), Shanxiao Wan(sw3510)

Title: Mental Health in Tech Workplace

Abstract:

- In this project, we want to explore what factors may be related to mental health. Specifically, we want to explore the following questions:

- How does the number of people who have mental health challenges differs by state? Maybe states with higher workload would have more people with psychological problems.

- How personal attributes are related to mental health condition? Maybe the elder has more stress than the young and thus they need more psychological treatment.

- What are the strongest predictors of mental health illness in the workplace? Maybe family history of mental health illness could be one of the strong predictors.

- We plan to further explore participants' response to the question Would you bring up a mental health issue with a potential employer in an interview. What would influence respondents' answer? What are their concerns?

Visualization:

1. Bar chart and Box Plot. We want to compare how different groups of people(by age, sex) perform differently in mental health condition and plan to use ggplot2.

2. Bubble chart. We would like to include company size when studying the relationship between mental health condition and whether employer provide mental health benefits. The larger the company size, the larger the bubble size.

3. Map. We plan to use tmap to draw a choropleth map of treatment, which shows how the number of people recieving psychological treatment differs by state. In addition, We plan to build an interactive map and provide at least three pieces of information on the company info and mental health condition in a popup. We

also would like to add marker clustering, so that zooming in will reveal the individual locations but ta zehe zoomed out map only shows the clusters.

4. Radar Chart. We would like to draw a portrait of those with mental health illness from the dimension of several factors such as family_history, remote_work, benefits, tech_company, seek_help, etc.

5. Pyramid Plot. We plan to provide a pyramid plot to show how the words of different answers differ in frequency.

6. Point plot. We would like to show the relationship of different sentiment and answers.

Data Description:

- This dataset is from 2016 OSMI Mental Health in Tech Survey (https://osmihelp.org/research). Open Sourcing Mental Illness (OSMI) is a non-profit corporation focusing on supporting mental wellness in the tech and open source communities. The dataset is downloaded from: https://www.kaggle.com/osmi/mental-health-in-tech-2016?select=mental-heath-in-tech-2016_20161114.csv

- The Dataset topic is the attitudes towards mental health and mental health disorders in the tech workplace. Above 1400 people participated in the survey.

- There are 63 variables in the dataset, including demographic variables, spatial data, text data， survey results for mental health disorder measurement and support for mental health from the employer.