**Methodology**

**Purpose**

The purpose of this project was to compare the results from two surveys, one in 2014 and the other in 2016, about the attitudes of mental health issues in the tech industry and where can people focus on if they want to make improvements in the future.

**Methodological Considerations**

The 2014 dataset looked at several independent factors to evaluate a few dependent variables that can explain the patterns and trends in mental health attitude: whether they get treatment for their concerns and how willing they are willing to talk to either their co-workers or supervisors to improve their working environment. To evaluate the variables listed above, the study used the following potential independent factors: Age, Gender, Country, State, if any known issues interfere with their work, checking if they were self-employed or have a family history of mental health issues, if their employers offer leave or support, number of colleagues, whether they work remotely at least half of the time or if it’s a predominant tech company or one where employees receive benefits.

**Data Collected**

Both datasets came from Kaggle.com. When the raw data was first looked, several irrelevant variables such as the time stamp as well as the any data points where the gender wasn’t male of female and any rows where a non-applicable value was presented was removed. This cut down the total number of data points from 1259 to 949. Once the sample size was reduced, all non-independent and dependent variables were removed, and the data was reorganized to make it simpler to read.

The next step to determine a range of scores for each factor since most of them were on a scale. Answers that were “Yes” were scored as 2, those that were “No” were changed to 0 and any answer that was “Not Sure”, “Some of them” or “I Don’t Know” was treated as 1. The scale for the question on whether mental health issues impacted ranged from never happening to often; the respective scale is Never (0), Rarely (1), Sometimes (2), and Often (3). The variable that stated whether their business allowed for leaved mental health related absences was scored from “Very Difficult” (1), “Somewhat Difficult” (2), “Don’t Know” (3), “Somewhat Easy” (4), “Very Easy” (5).

The studies initial results were questionable because the number of co-worker’s column had six possible answer choices: “Less then 5”, “5-25”, “26-100”, “100-500”, “500-1000”, and “More than 1000”. While it is clear to the audience that these are six choices, R studio took it as six subgroups within a group. To fix the potential irregularities, the researchers redid the analysis with six separate columns, each using a Yes/No or 1/0 as the values.

**Methodology**

This study split the data by everyone, men, and female. A general profile for each subset of data was created to better understand the conclusions and implications derived from the regression analysis. The profile consisted of the following elements: quantile data, country distribution, and distribution of people from the United States by coasts and number of co-workers, defined by a range, respectively.

The next step was to run a correlation analysis with the male and female dataset. After the correlation was done, the study designated four dependent variables: whether they seek help, treatment and if they disclose those issues to either their co-workers or supervisors. R studio was used to determine potential data models that would create predictions that would be used in the next step, where the objective would be to compare the average of a series of predictions from the 2014 data models to the actual 2016 one.