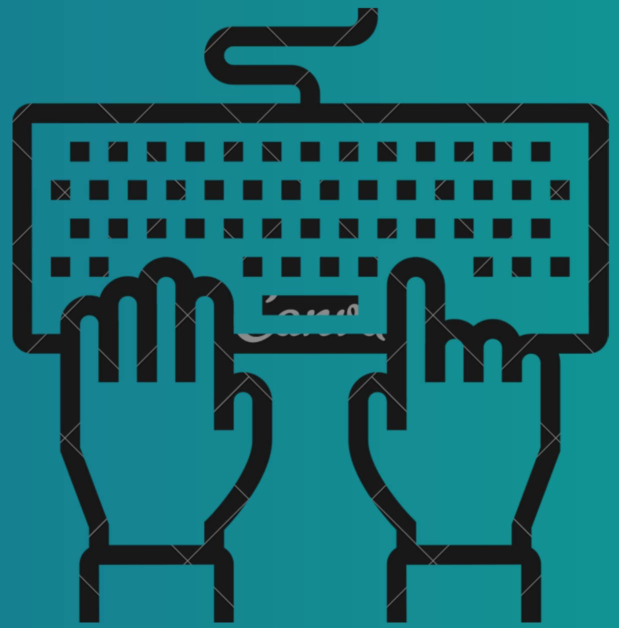
Mind Matters: Mental Health in Tech



CSP571: Project Presentation
Spring 2023

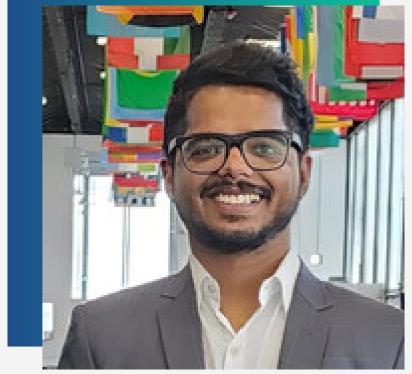
- Zainab Hasnain
- Aditya Sharma
- Abhishek Jaiswal
- Mohammad Ans Qadri

OUR TEAM MEMBERS

ADITYA SHARMA

A20516668

asharma55@hawk.iit.edi



ABHISHEK JAISWAL

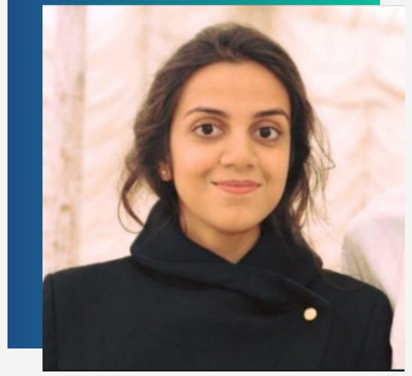
A20380004

ajaiswal1@hawk.iit,edu

MOHAMMAD ANS QADRI

A20521573

mqadri@hawk.iit.edu



ZAINAB HASNAIN

A20516879

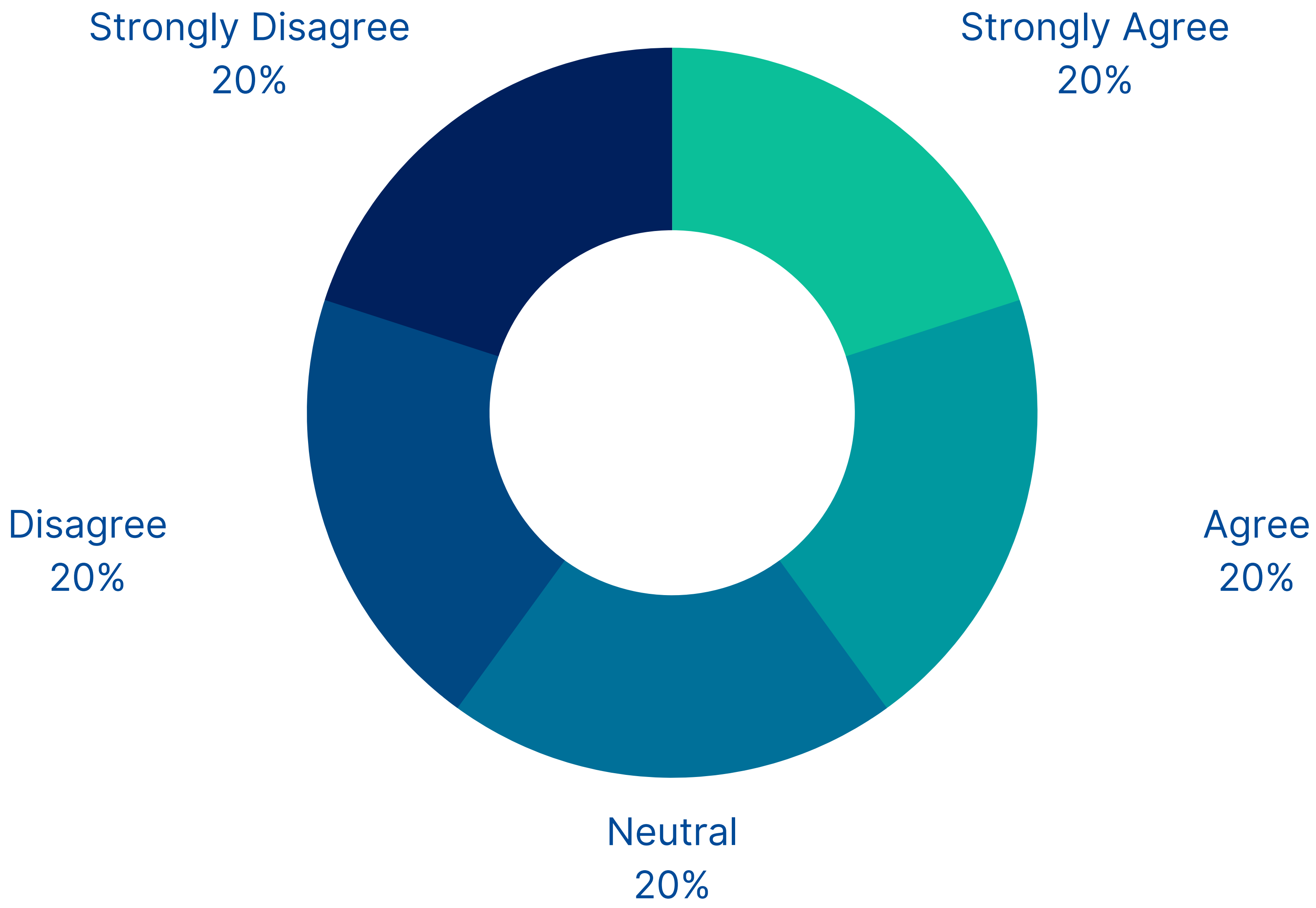
zhasnain1@hawk.iit.edu

Presentation Summary

- Dataset and Problem Statement
- Methodology and Timeline
- Insights from Data
- Predictive Modelling
- Conclusion

Our Dataset

OSMI Survey Responses



According to the 2021 OSMI Mental Health in Tech Survey, over 90% of workers were diagnosed with a mental health disorder.

The survey contains questions related to employee's personal data like **age**, **gender**, **country, family history**, etc, along with details of workplace environment, such as **company size, easy to take leave, supportive employer**, etc.

Executive Summary

Poor mental health in the tech industry is a growing concern, but through data-driven analysis and predictive modeling, we can uncover the factors that contribute to employees seeking treatment.

Key Research Issues

- What is the **distribution of data** among different age groups, genders, and nationality?
- How can we use the **yearly survey responses** data to analyze the general perception of mental health in the tech industry?
- Do the trends in data change **pre-covid** and **post-covid**?
- What is the **most significant factor** which leads to an employee seeking for mental health?
- How can we model and **evaluate a predictive mode**l on our dataset for future prediction?
- How **accurate** are our models according to the chosen evaluation metrics?
- What is the **scope** and suggested **improvements** for this project?

Key Findings

- Most of the responses are from **men** aged between **20 to 40** from the **United States**
- We have used the years 2014, 2015, and 2016 as the **pre-covid data** and 2019, 2020, 2021 for **post-covid** analysis
- **Mental health interference with work, family history, employer's provided benefits,** and **gender** are all strong predictor variables
- **Naive Bayes** model most accurately classifies whether an employee will seek treatment based on the survey responses
- Our data is **heavily unbalanced**, nonetheless provides some valuable insights

Project Timeline

- **WEEK 1** — Ideation: Data Gathering
- **WEEK 2** — Data Pre-Processing and Cleaning
- **WEEK 3** — Exploratory Data Analysis and Visualizations
- **WEEK 4** — Model Selection, Implementation, and Evaluation
- **WEEK 5** — Project Conclusion and Documentation
- **PRESENT** — Project Presentation

Methodology

01

Data Gathering

- Gathering survey responses for each year
- Analyzing different columns for each dataset
- Removing unnecessary columns
- Combining pre-covid and post-covid datasets
- Combining one consolidated dataset for prediction purposes
- Define project scope and plan project pipeline

02

Data Pre-processing, EDA & Visualization

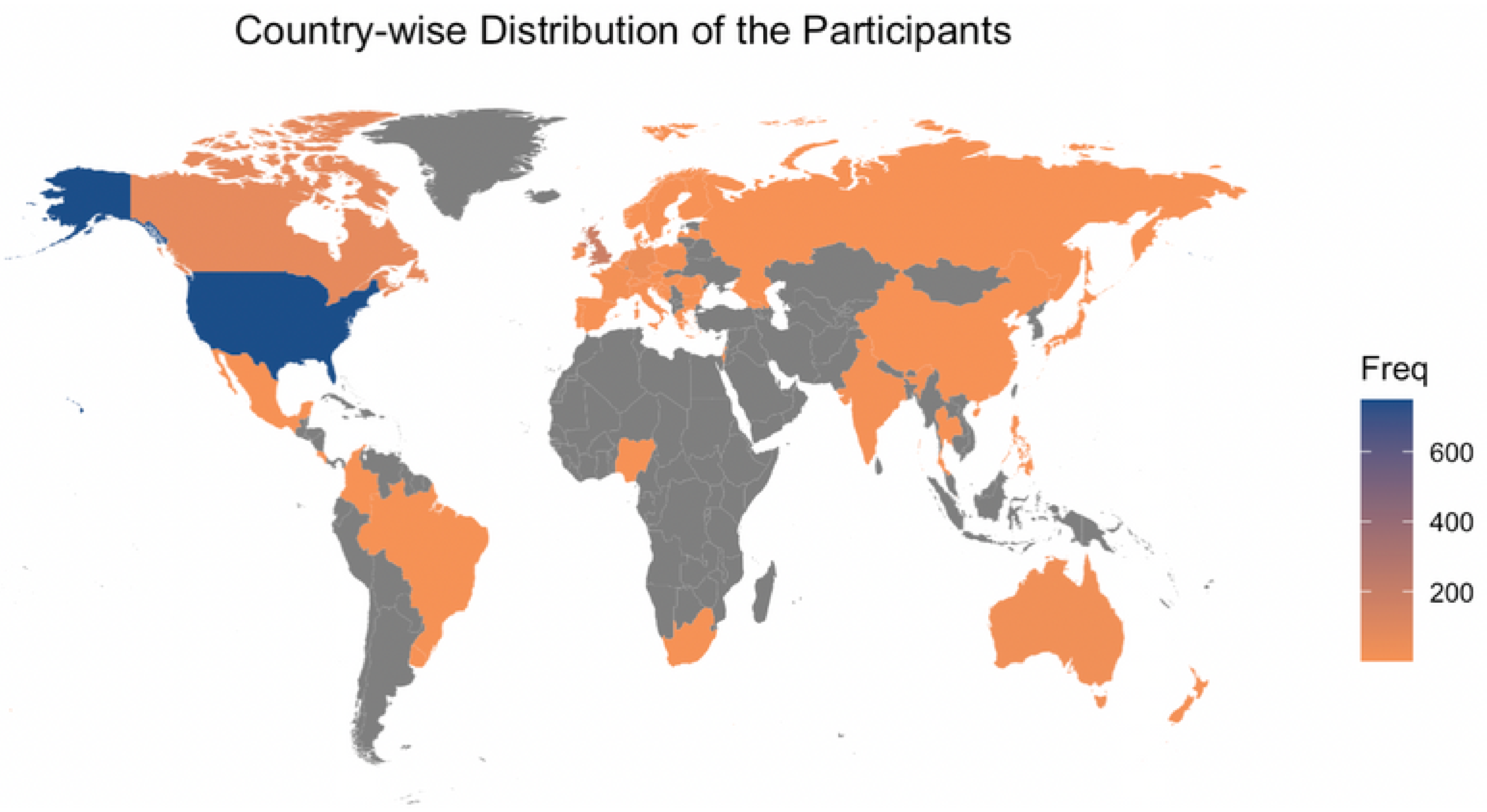
- Changing datatype of columns
- Handling any null values/outliers
- Cleaning each column data manually
- Example: 'M', 'male', 'Male', 'm', 'man', --------> 'M'
- Creating visuals to understand data distribution among different groups
- Plotting significant variables with response variable

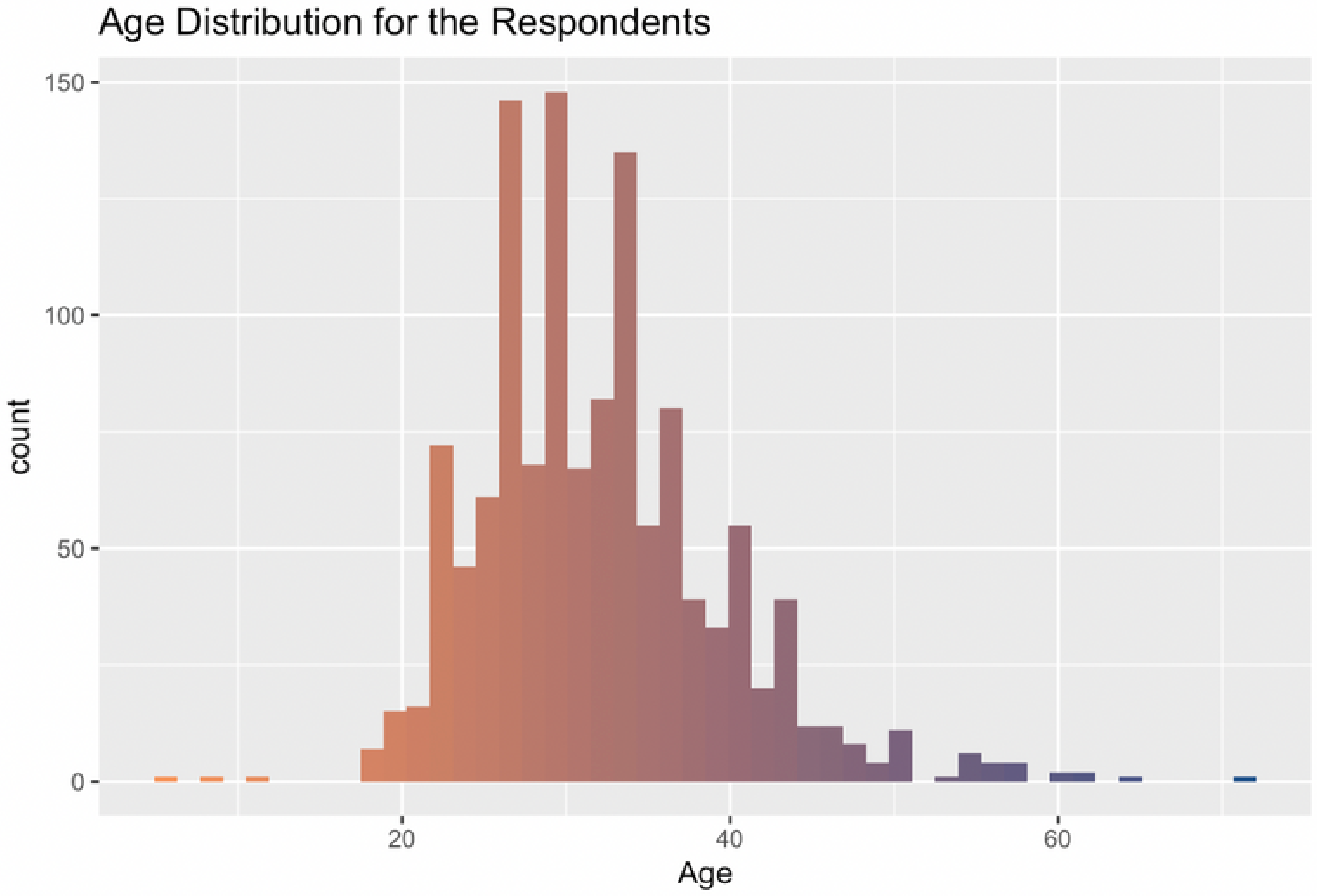
03

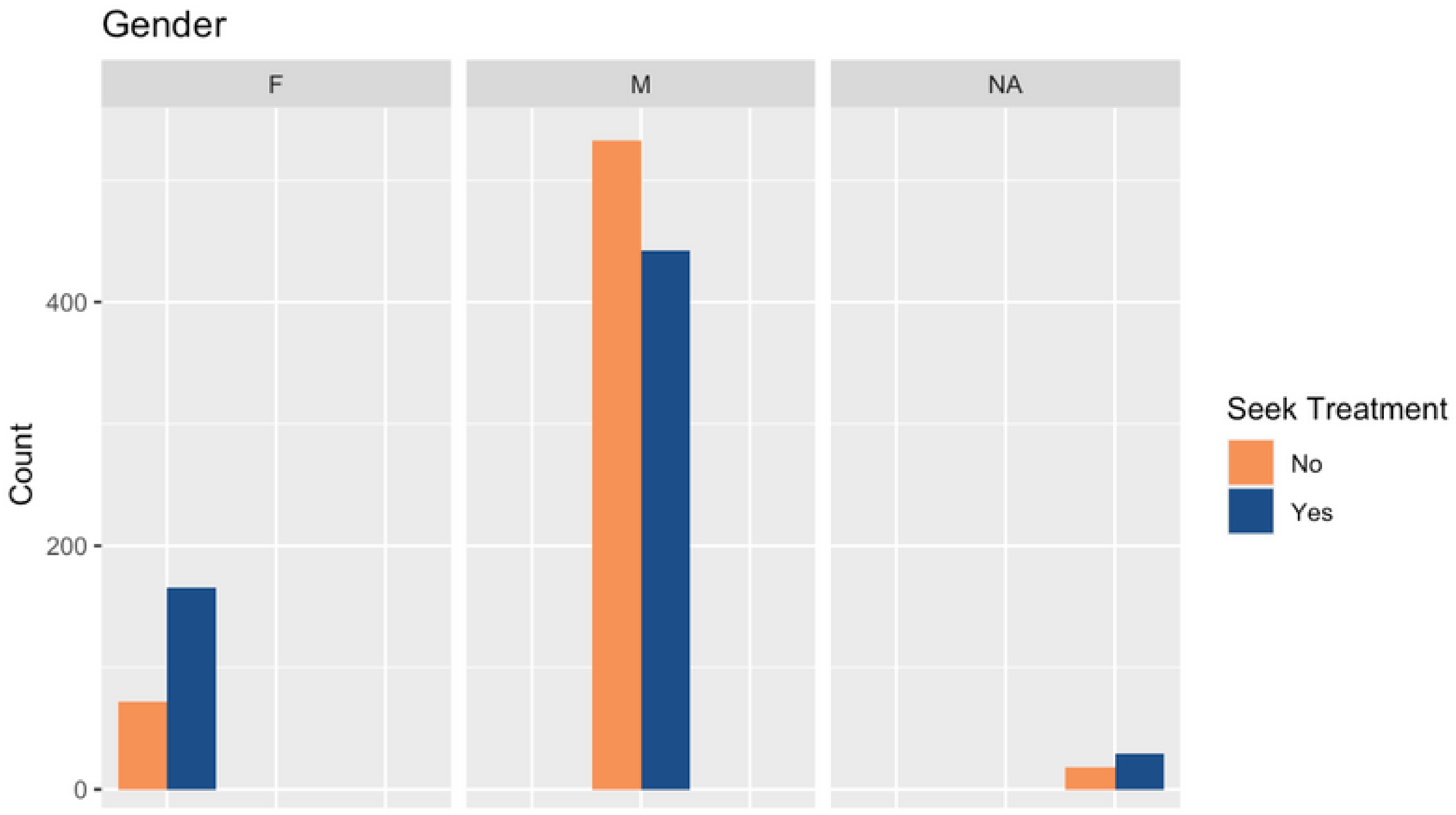
Predictive Modelling

- Perform 80/20 train-test split
- Use different classification algorithms to model the data
- We selected Random Forest, Decision Trees, Naive Bayes, and Linear Support Vector Machines
- Evaluate the performance of the models using various metrics
- Conclude by discussing project's scope, limitations and potential future work

Pre-Covid Distribution of Data

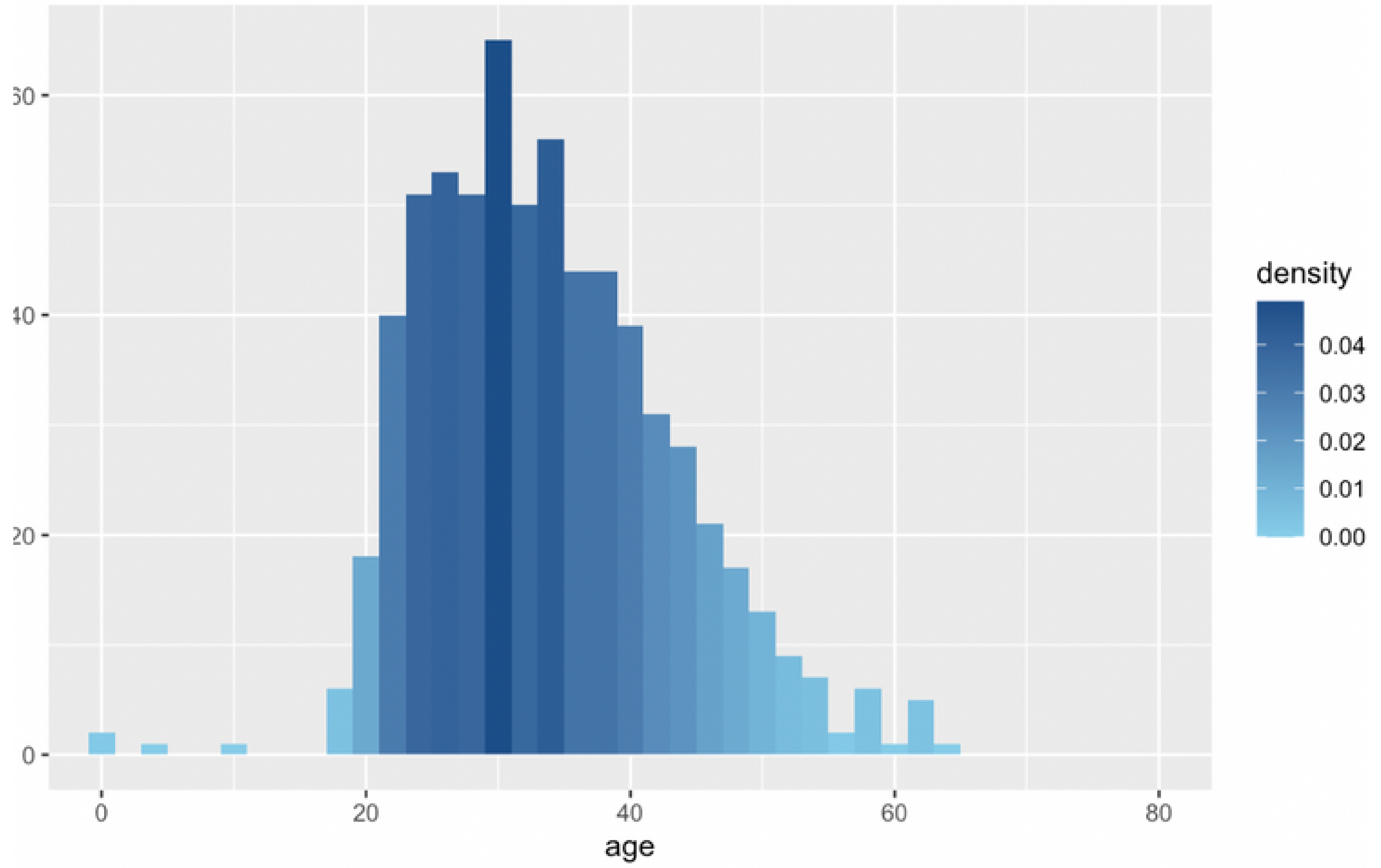




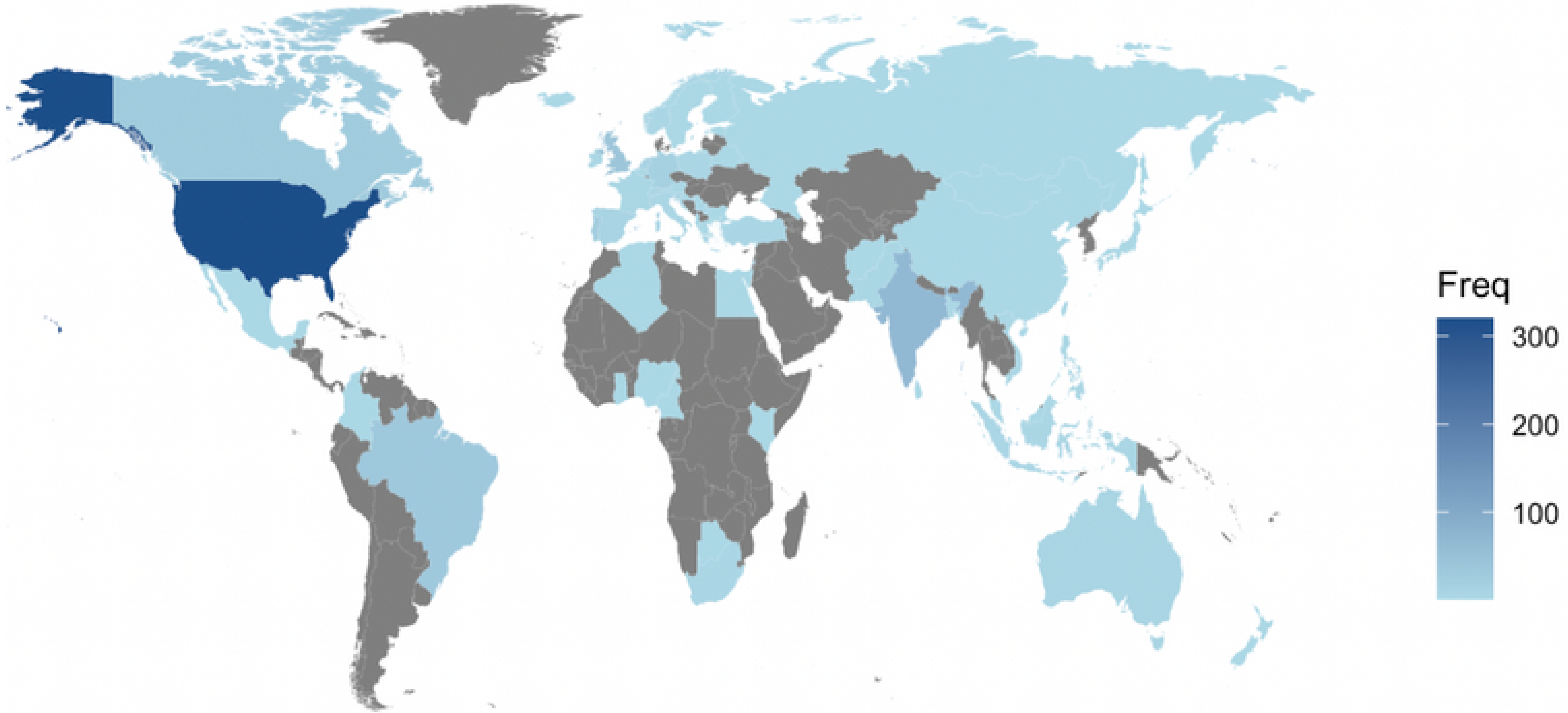


Post-Covid Distribution of Data

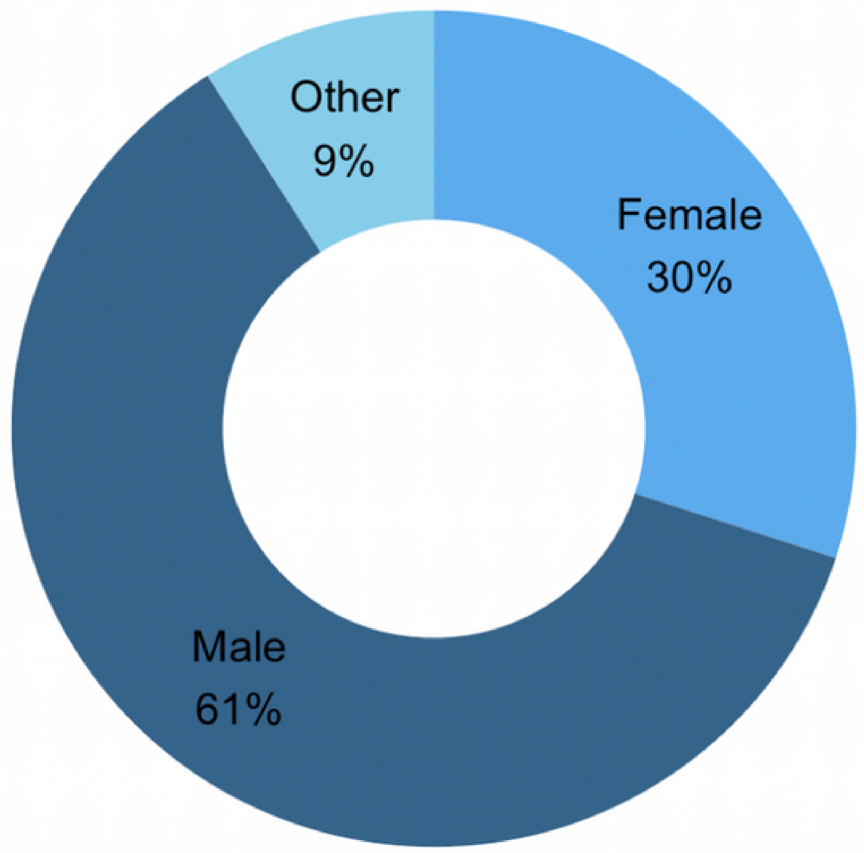
Age Distribution for the Respondents



Country-wise Distribution of the Participants

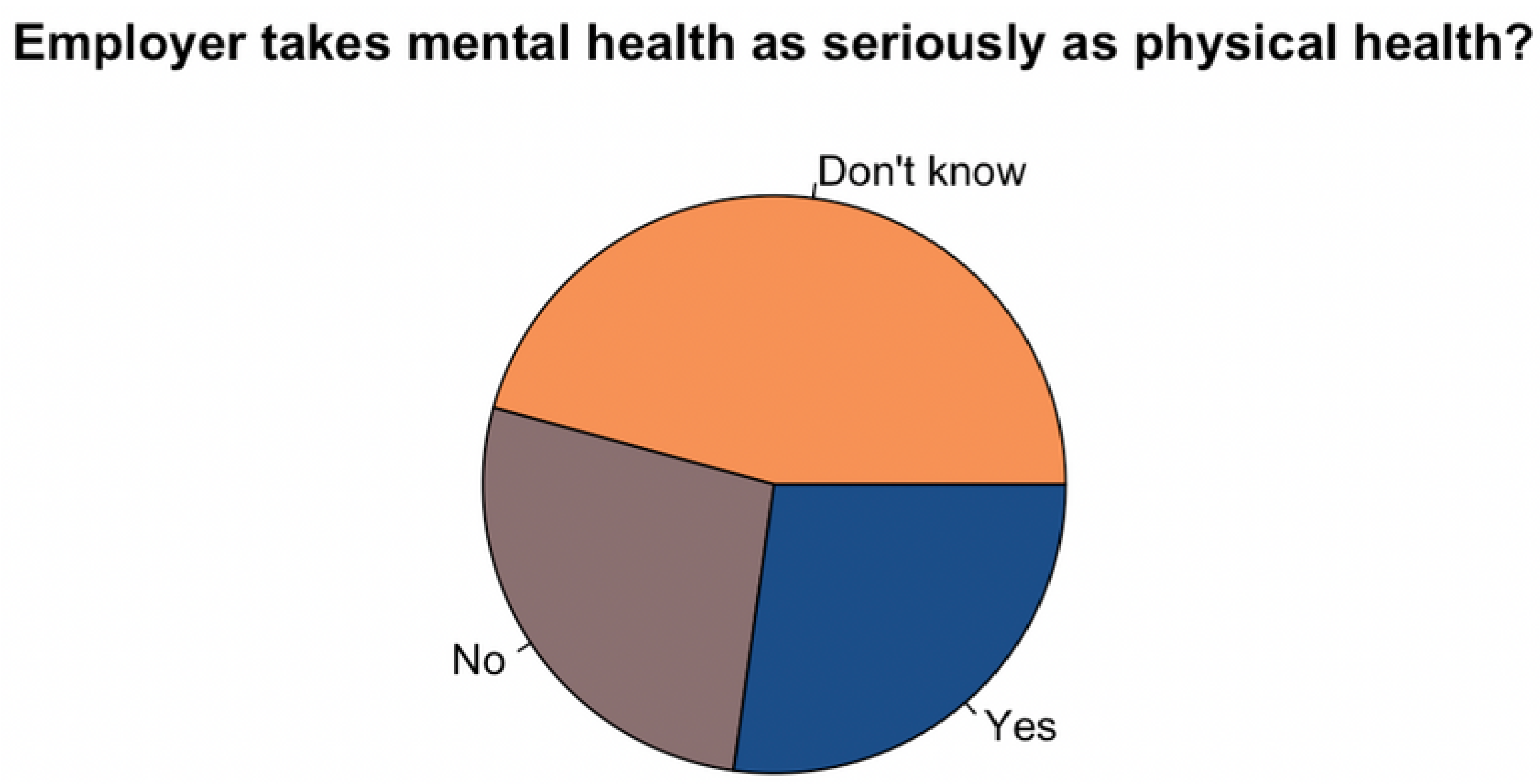


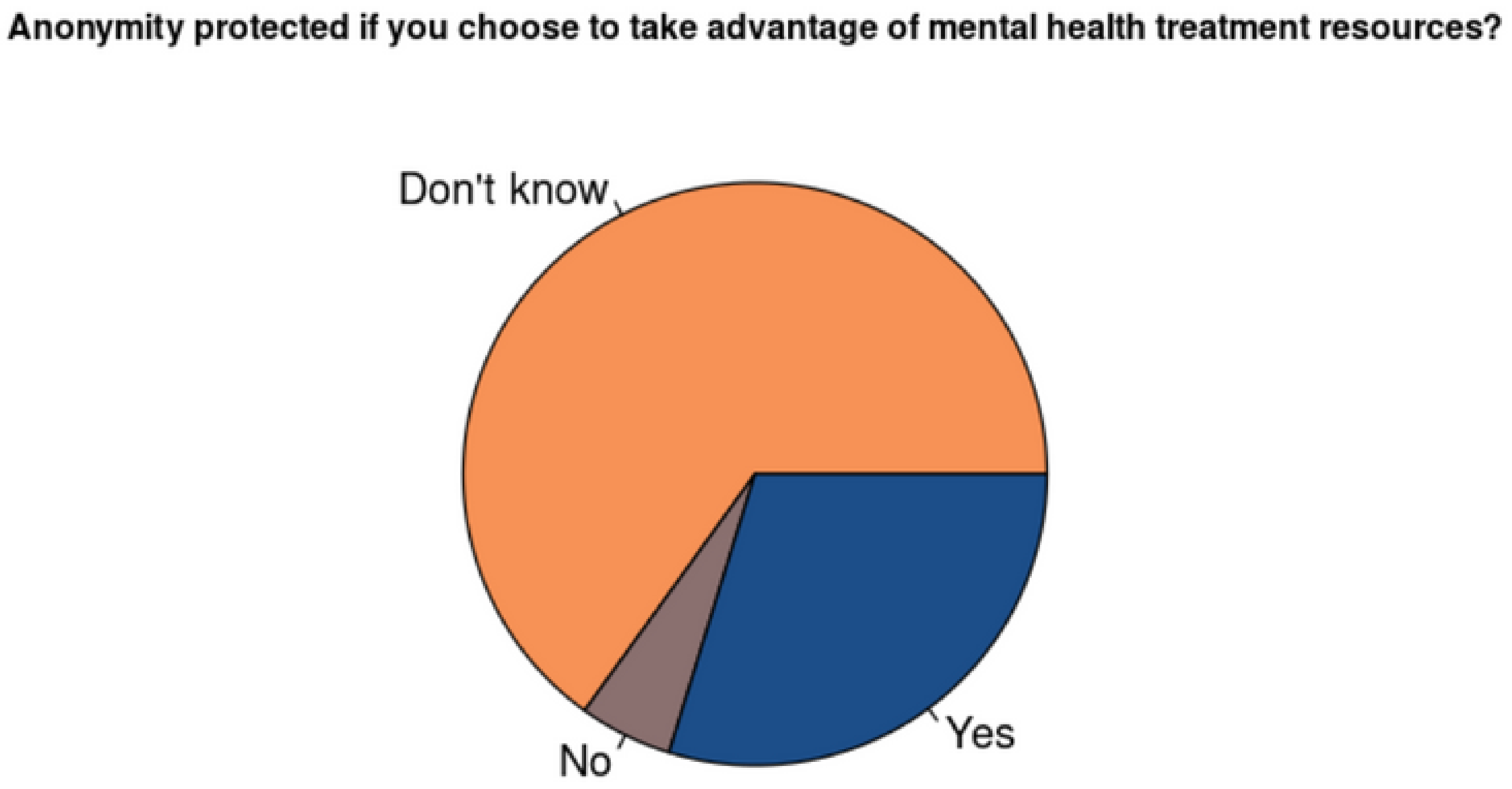
Gender of Survey Participants

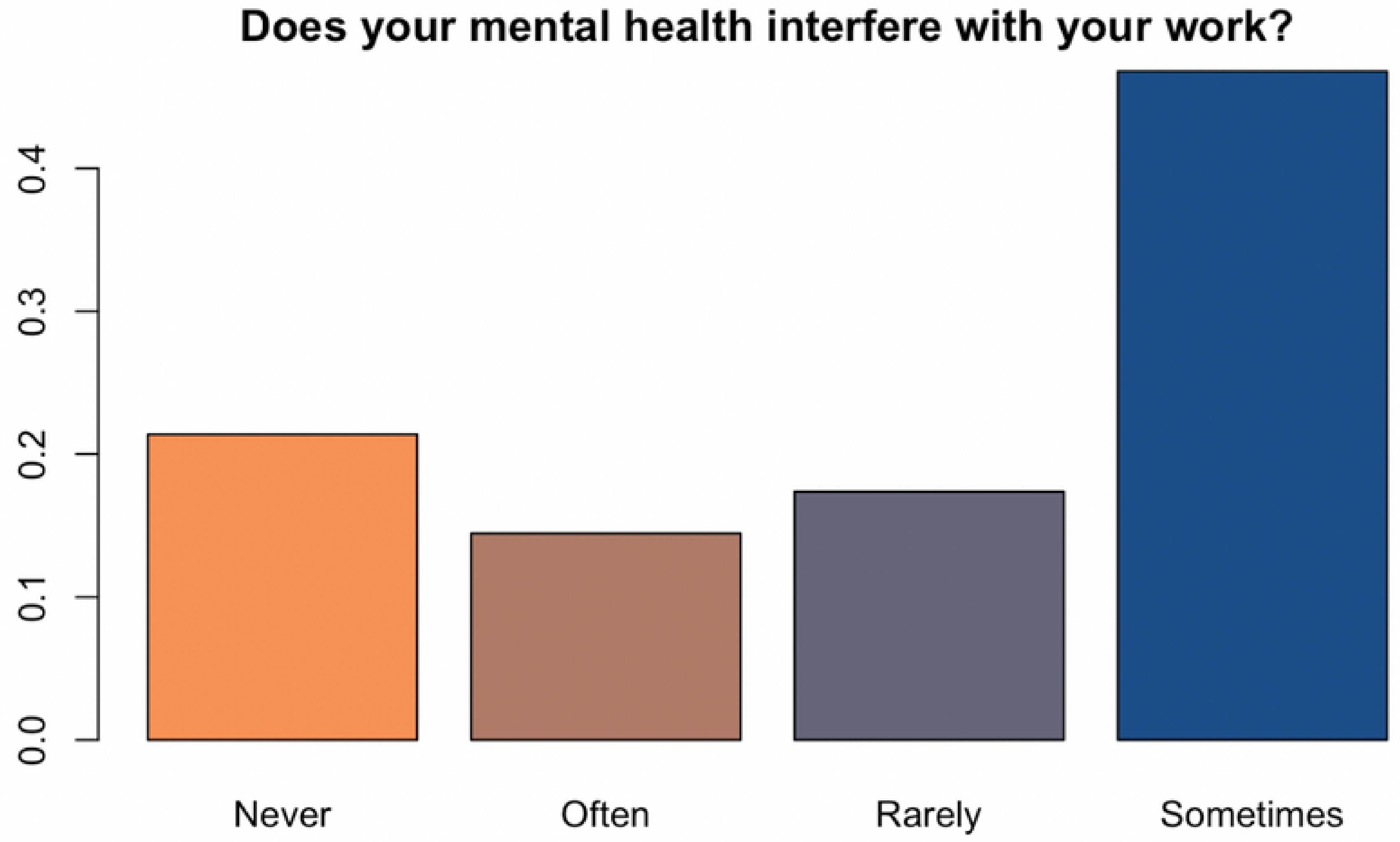


Attitudes Towards Mental Health

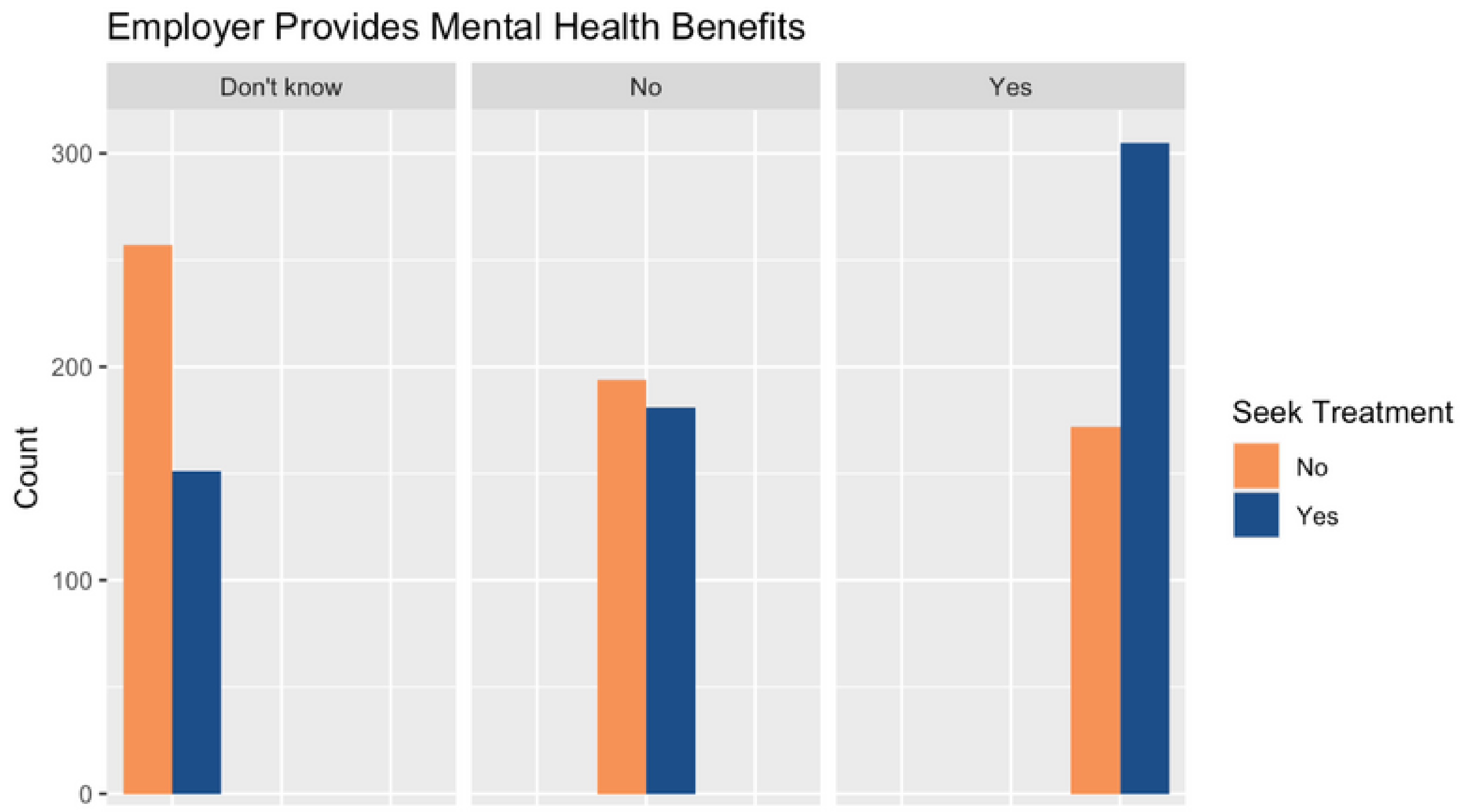
Over 50% of respondents sought for help if their mental health interfered with work

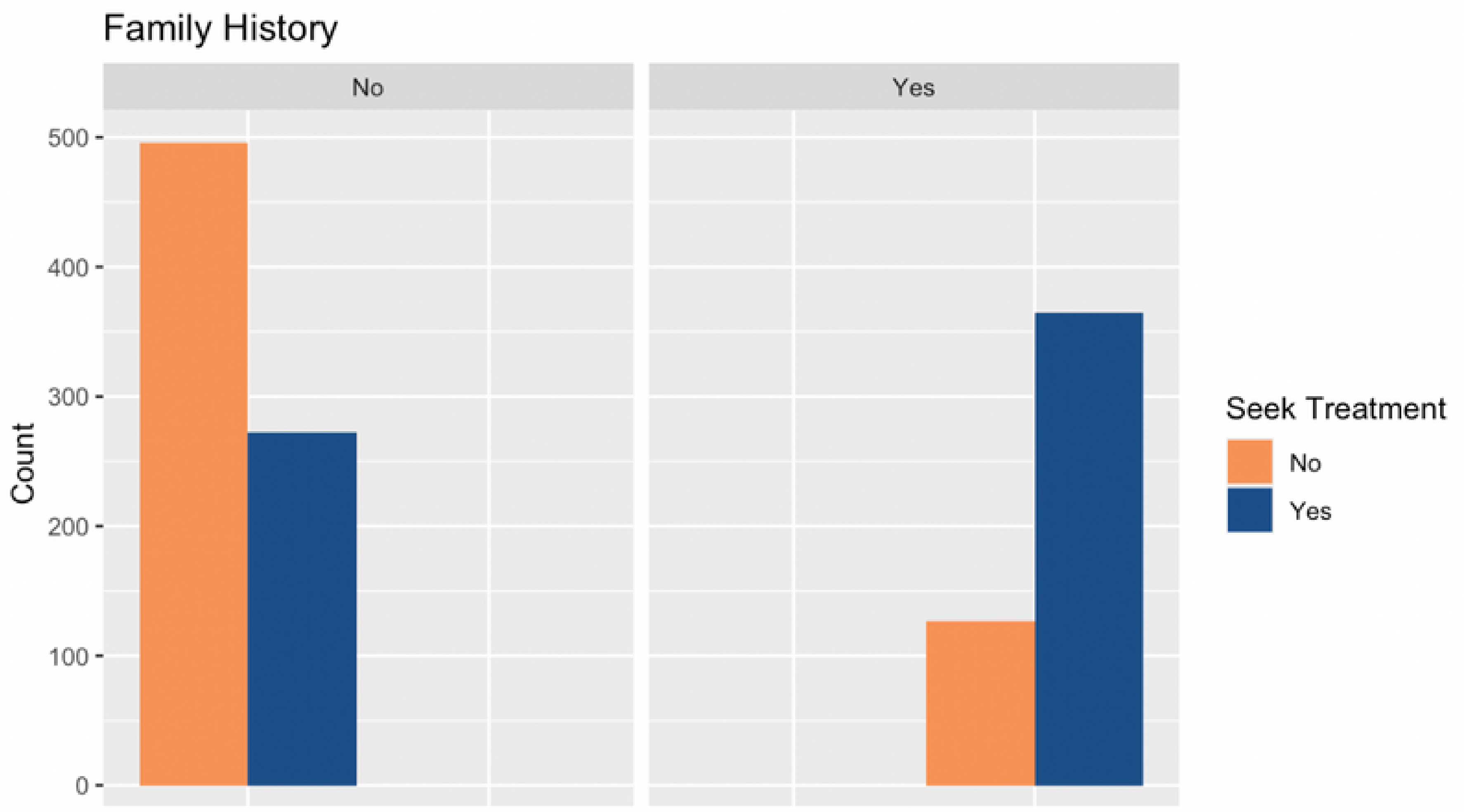


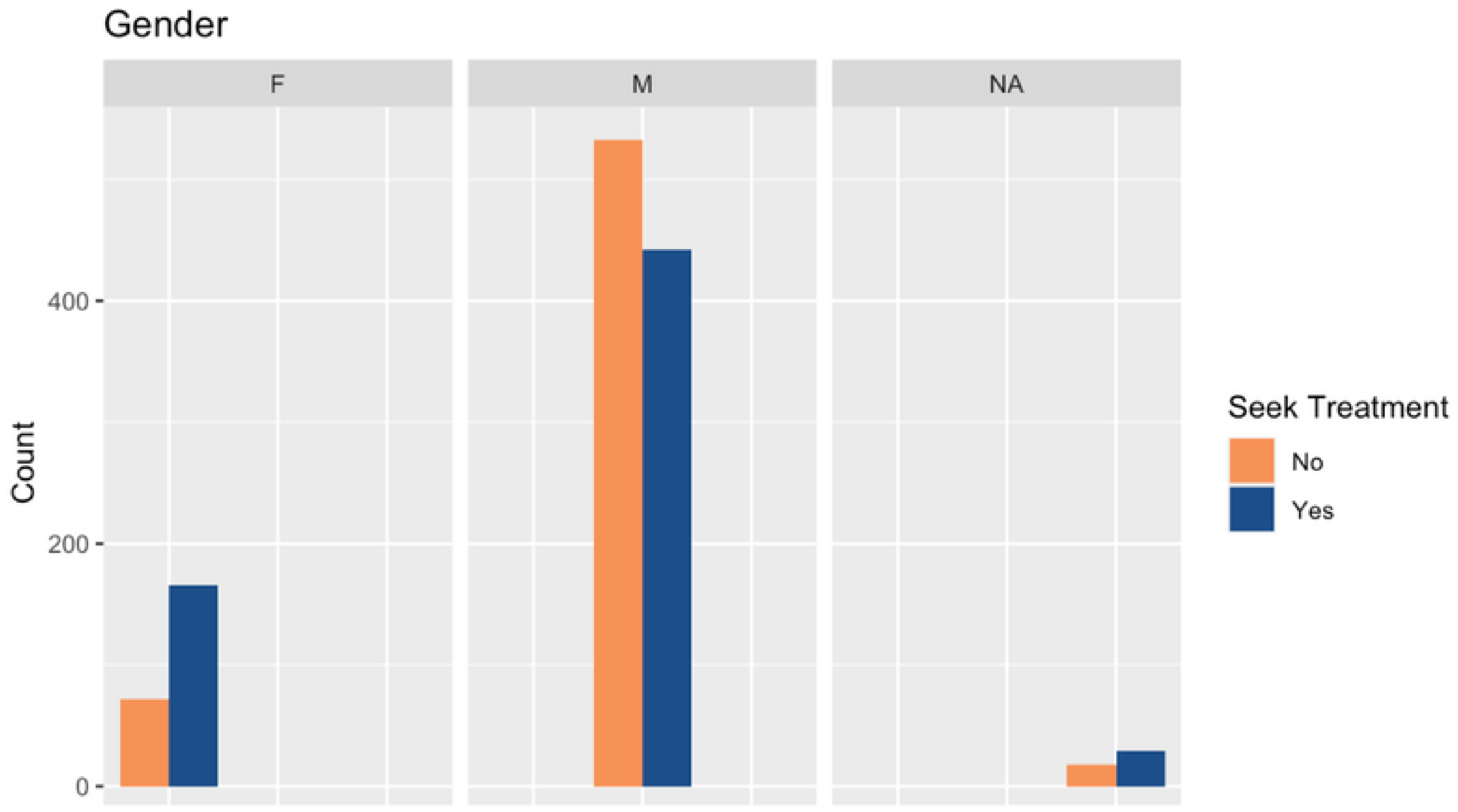




Relationship Between Predictors and Response



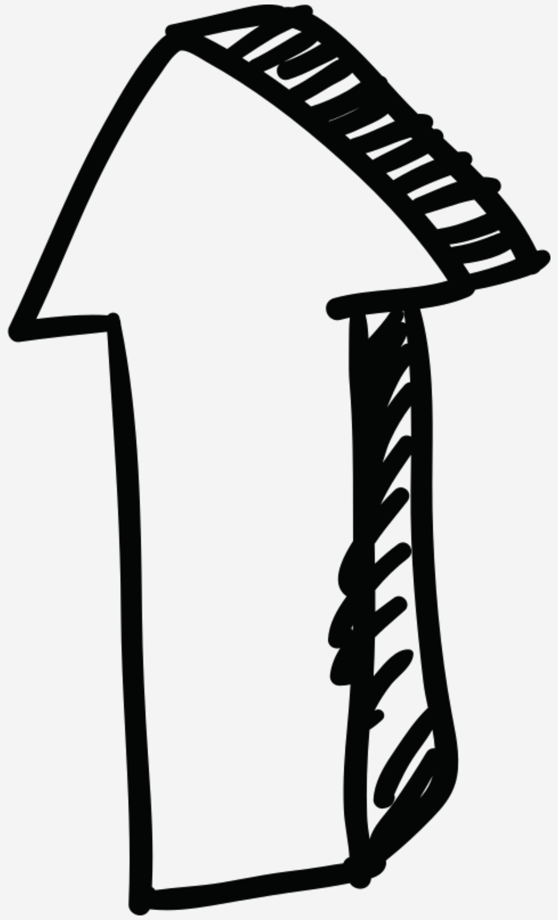




What are some of the strongest predictors?

Response Variable: Seek Treatment

Strong Predictors



- Interference with Work
- Family History of Mental Health
- Employer Provides Mental Health Benefits
- Gender
- Anonymity Protected
- Unsupportive/Supportive Response at Work

Models and Performance

Models	Testing Accuracy (%)
Logistic Regression	50.9
Decision Trees	71.96
Random Forest	77.28
Naive Bayes	71.97
Support Vector Machines (Linear)	51.22

Evaluation Metric: Test-Set Accuracy
Best Model: Random Forest

Challenges and Limitations

- Unbalanced dataset
- Demographic Bias ('USA' Dominant)
- Gender Bias ('Male' Dominant)
- Empty and Unsure Responses
- Limited Responses over the Years

Future Work

01

Gather more data from different sources for a more in-depth study of mental health issues in the tech industry

02

Use of ensemble methods and **other classification algorithms** for more accurate prediction

03

Expand our research by examining attitudes towards mental health in **other industries such as healthcare, hospitality, transportation**, etc

Send us a message at zhasnain1@hawk.iit.edu for feedback and questions.