

# Mental Health in the Tech Industry

*Silei Li and Mariel Pacada*

*12/8/2019*

## Dataset and Goals

This project will explore the relationship between mental health and the conditions in a workplace within the tech industry, which is relevant to the course as quite a few fellow students are aiming to work in this field. The dataset we will use includes data from a 2014-2015 survey that measures attitudes towards mental health and frequency of mental health disorders in the tech workplace. We will explore whether the presence of mental health care options in one's company predicts the mental health of an employee. We will also explore geographical differences, asking whether North America or Europe has better options and less social taboos on mental health.

## Data Cleaning Choices

The bulk of our data cleaning came from the gender column. The response for this category in the survey seems to be freeform, as there were many different answers that meant the same answer (e.g. "male" and "mail" would be the same as "Male"). To avoid mis- and over-classification, we have decided to clean the ones to be decidedly "Male" and "Female", respectively, and categorize others as simply "Queer".

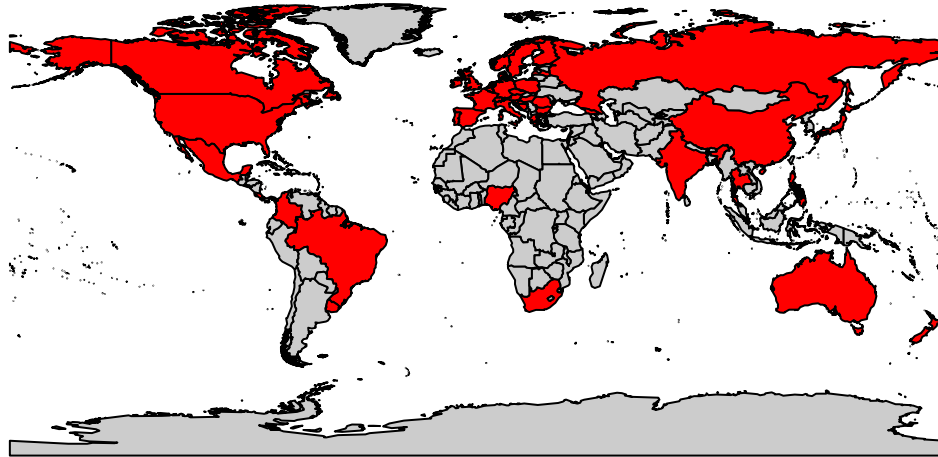
## Part 1: Initial Exploratory Data Analysis

First, let's get to know our data!

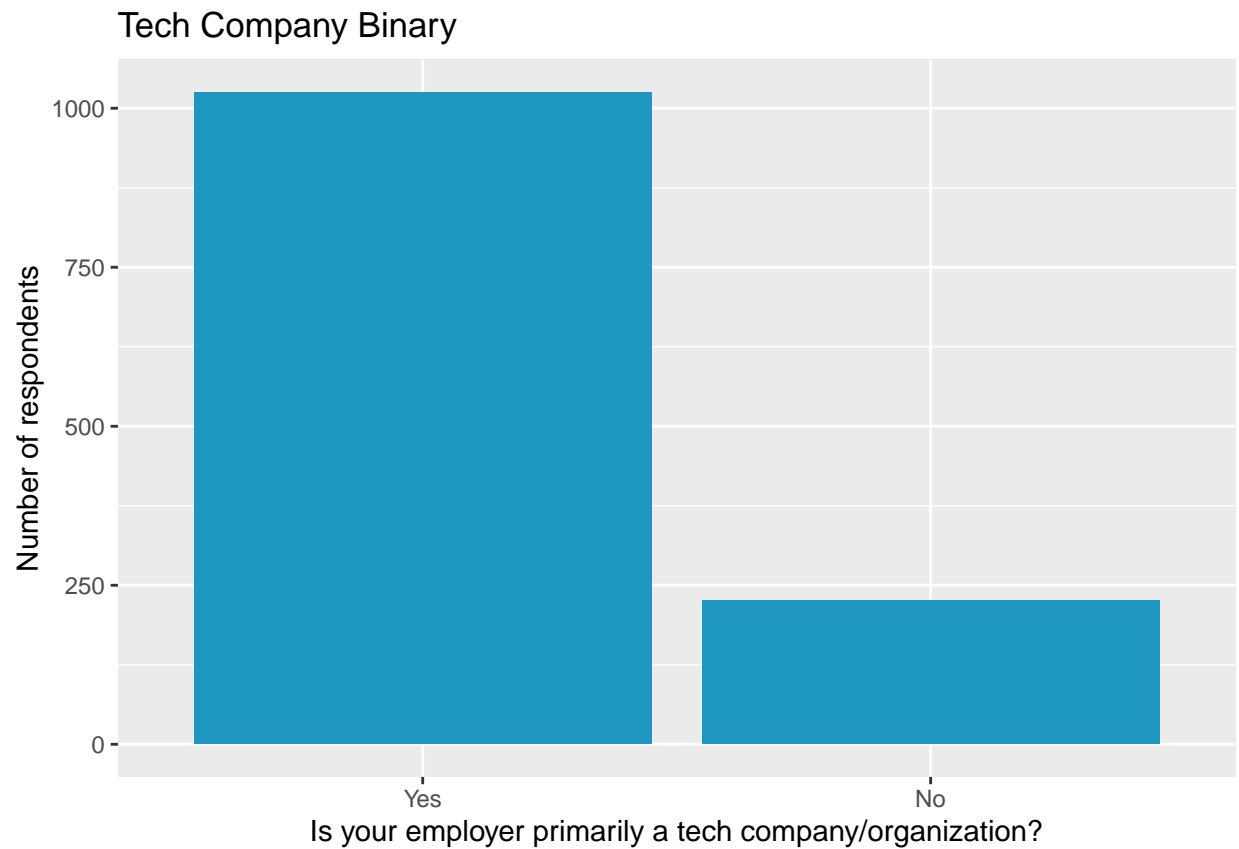
```
data(wrld_simpl)
map_countries = wrld_simpl@data$NAME %in% survey_countries$Country

plot(wrld_simpl, col = c(gray(.80), "red")[map_countries+1], main = "Countries represented in the dataset")
```

## Countries represented in the dataset

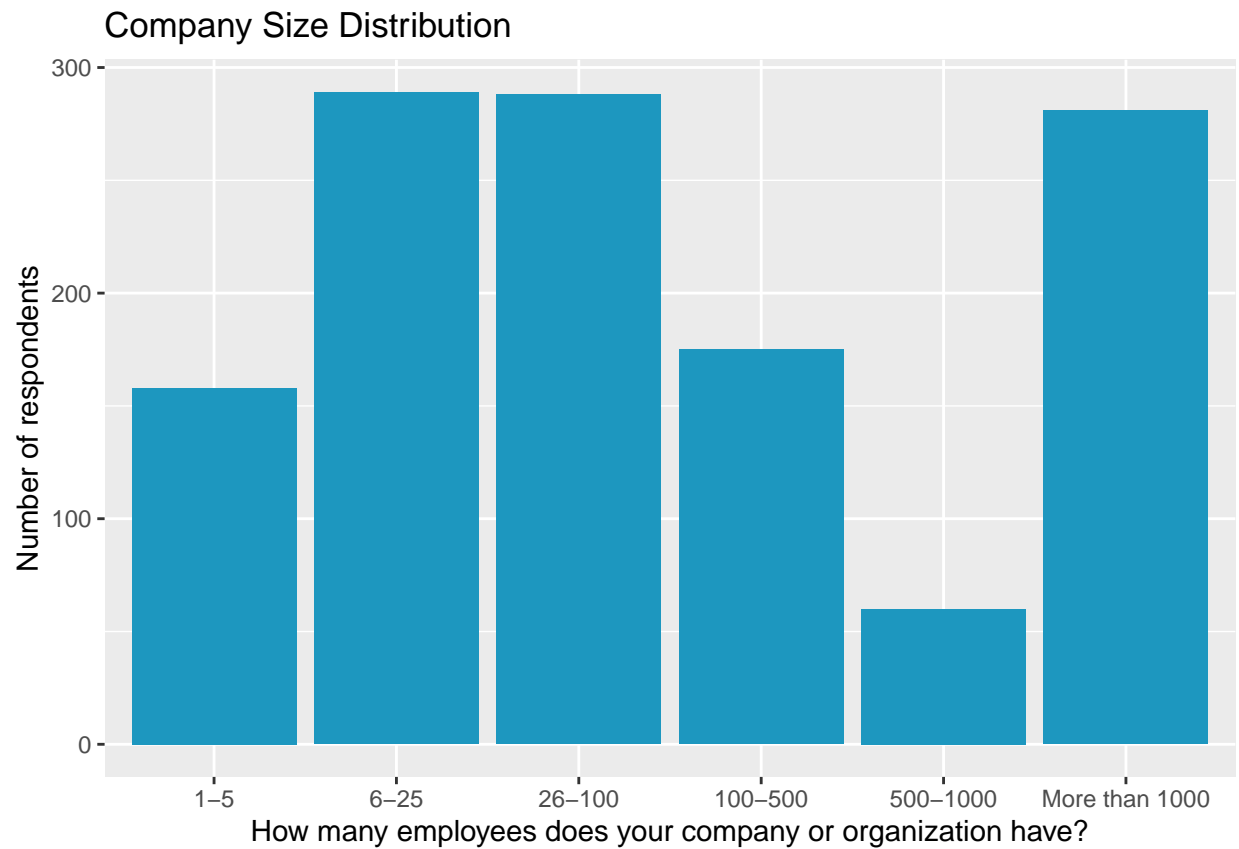


```
tech_binary <- survey %>%  
  filter(!is.na(tech_company)) %>%  
  dplyr::select(tech_company) %>%  
  mutate(tech_company = factor(tech_company, levels = c("Yes", "No")))  
  
ggplot(tech_binary, aes(x = tech_company)) + geom_bar(fill = "#1D97BF") + labs(x = "Is your employer pr
```

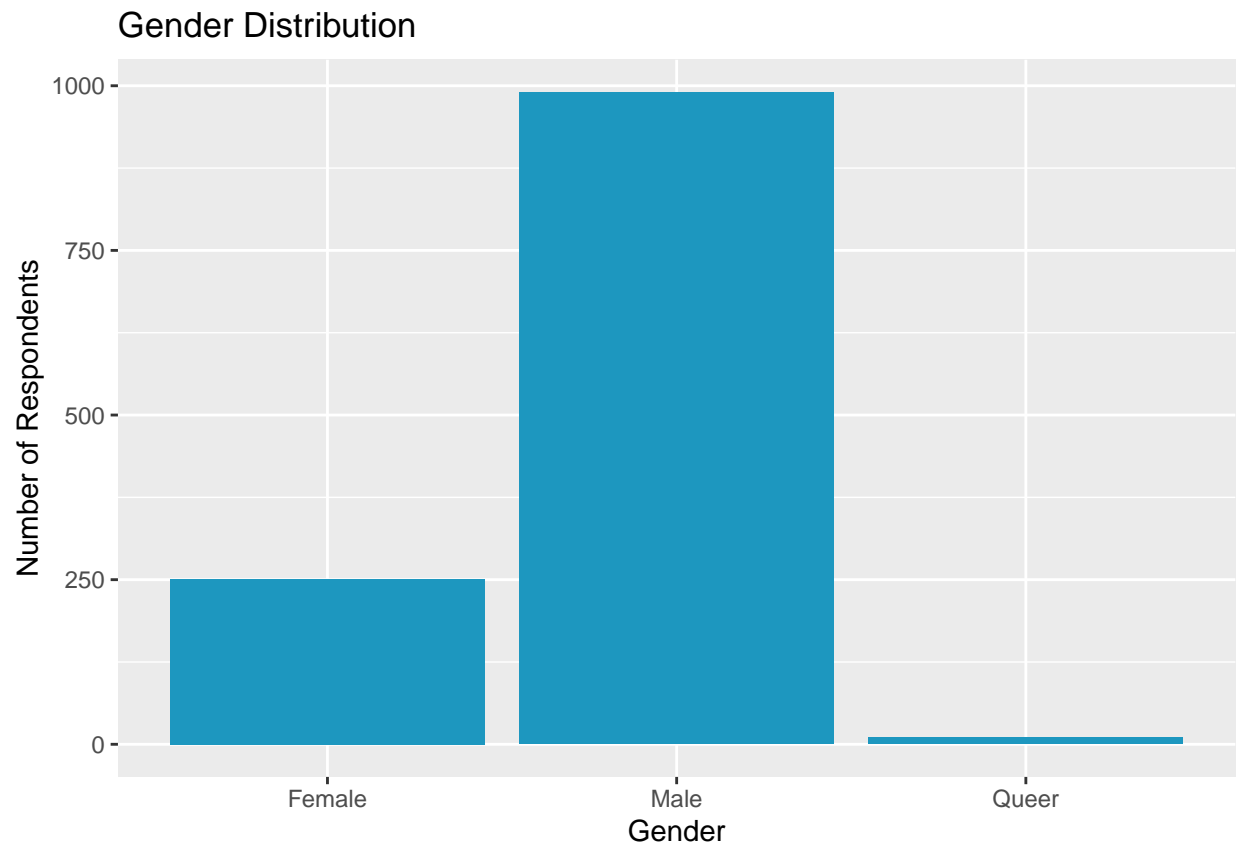


```
num_employee <- survey %>%
  dplyr::select(no_employees) %>%
  mutate(no_employees = factor(no_employees, levels = c("1-5", "6-25", "26-100", "100-500",
    "500-1000", "More than 1000")))

ggplot(num_employee, aes(x = no_employees)) + geom_bar(fill = "#1D97BF") + labs(x = "How many employees")
```



```
ggplot(survey, aes(x = Gender)) + geom_bar(fill = "#1D97BF") + labs(y = "Number of Respondents", title = "Company Size Distribution")
```



```
ggplot(survey, aes(x = Age) )+ geom_histogram(binwidth = 2, fill = "#1D97BF") + labs(y = "Number of Resps
```

