CSC 465

Mental Health Survey

Data Visualization | Team 3 | Winter 2016

Group Technical Report

Introduction

According to the National Alliance on Mental Illness (NAMI), a mental illness is a condition that affects a person's thinking, feeling, or mood [1]. These conditions may affect someone's ability to function or relate to others. One in five adults experience a mental health condition every year. There are several types of mental health conditions and the severity of these conditions differs among each individual.

Early engagement and support to individuals with a mental health condition is crucial to improving outcomes and recovery. According to NAMI, a mental health condition is usually caused because of multiple factors [1]. A stressful work environment can be one of the causes to a mental health condition.

We looked at a survey dataset on Kaggle.com titled 'Mental Health in Tech Survey'. The researchers asked 1260 participants at Technology companies about their history and opinions on mental health, and also collected demographic data.

Data Fields

The demographic information collected in the survey included age, gender, country and state. Information about general aspects regarding mental health such as treatment and family history was collected. The researchers also collected information specific to employers that could have an effect on mental health, such as company-provided health benefits, wellness programs at work, and ease of taking leaves. Work environment related information such as likelihood of discussing mental health issues with co-workers and supervisors, employer perspective of mental health, and observation of negative consequences due to discussing mental health issues at work were also collected. The survey concluded with a free text entry of participant comments.

By visualizing the data variables in the survey, we aim to present people's opinions towards their workplaces, and how supportive workplaces are towards mental health conditions. We also aim to provide suggestions to improve workplace conditions in order to provide a better environment for people with mental health disorders.

Exploratory Analysis

We started by creating an exploratory graph **(Figure 1)** to see the relationship between family history and treatment. We had size show the number of responses. We can tell respondents who do not have family history of mental illness are not likely to be taking treatment, and vice versa. This implies that there is a positive correlation between the two variables. We further enhanced this graph in our visualizations to show this distribution across states.

We then wanted to know if someone who was taking treatment had benefits to cover it. We explored this by creating a bar graph **(Figure 2)** to compare the number of respondents who had benefits to the number of respondents who required treatment. While it showed that the majority who required treatment had benefits, it also showed that a surprisingly large number of respondents who required treatment did not have benefits, or did not know if they their employer provided benefits. This prompted us to improve the visualization to show this in a better way and see how gender plays a role in it.

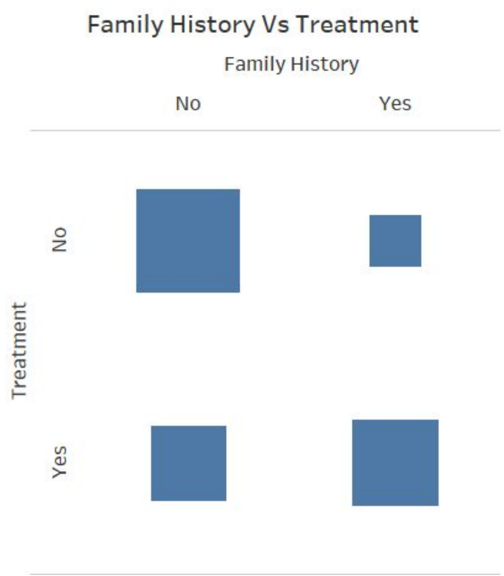


Figure 1: Graph showing relationship between respondents' treatment and family history

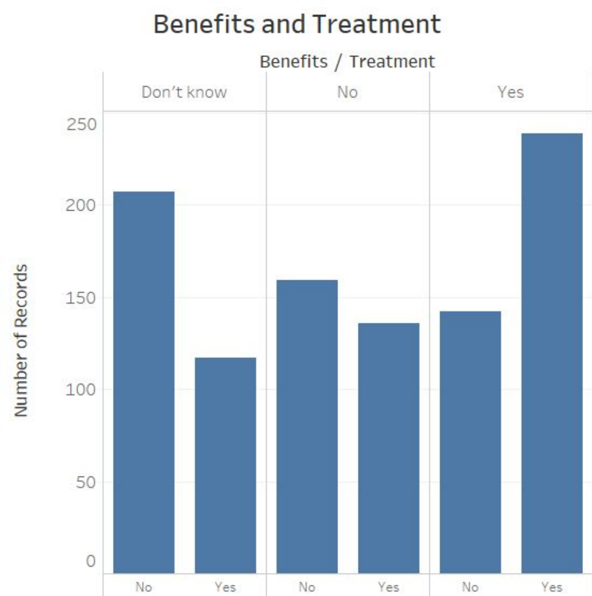


Figure 2: Bar graph showing distribution of benefits among respondents

Visualizations

Family History and Treatment

In our exploratory analysis we saw there was a positive correlation between treatment and family history. We wanted to show the distribution of this relationship across states in the US in order to see which states this was most prevalent in. We started by converting the categorical data into numerical data - The 'Yes' responses were converted to 1's and the 'No' responses were converted to 0's. We made two choropleths to show the distribution of family history **(Figure 3)** and treatment **(Figure 4)**.

When looking at the choropleths, we can see that the first two choropleths look almost identical. The states that have the highest concentration of individuals taking treatment also had the highest concentration of family history, and vice versa. A third choropleth **(Figure 5)** was created to make the similarity between **Figure 3** and **Figure 4** more profound. **Figure 5** shows that the relationship between remote work benefits and treatment isn't as strong across the states as it is with family history and treatment.

While making these choropleths we changed the colors in order to make clear the difference between states that had a low number of respondents and states where no data was collected. We also changed the colors so you could tell more clearly which states had a larger number of 'Yes' responses compared to states that had a lower number.

Participants with family history of mental health disorder

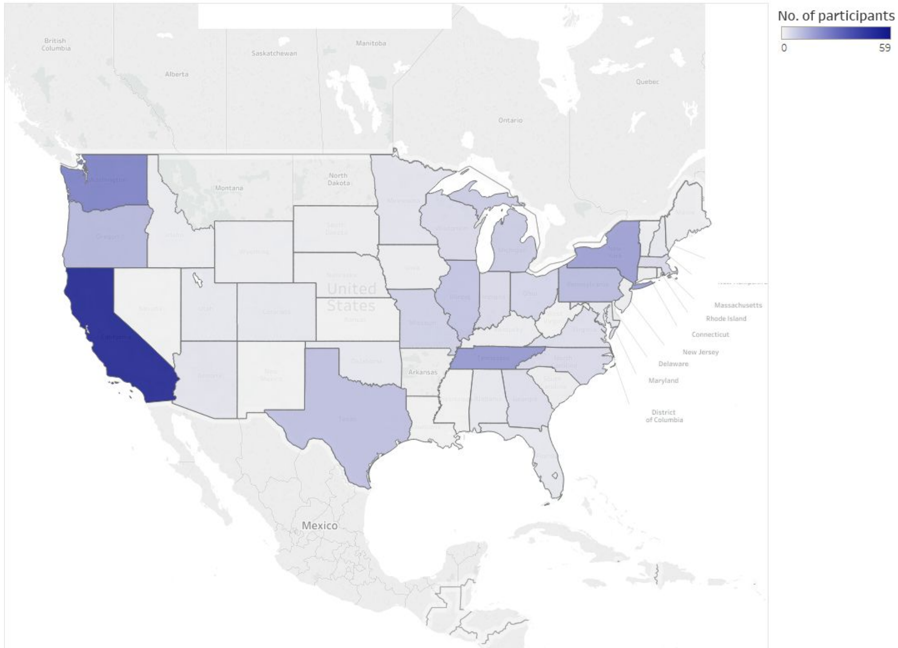


Figure 3: Choropleth showing distribution of participants across US states that have family history of mental health disorder

Participants who took treatment for mental health disorder

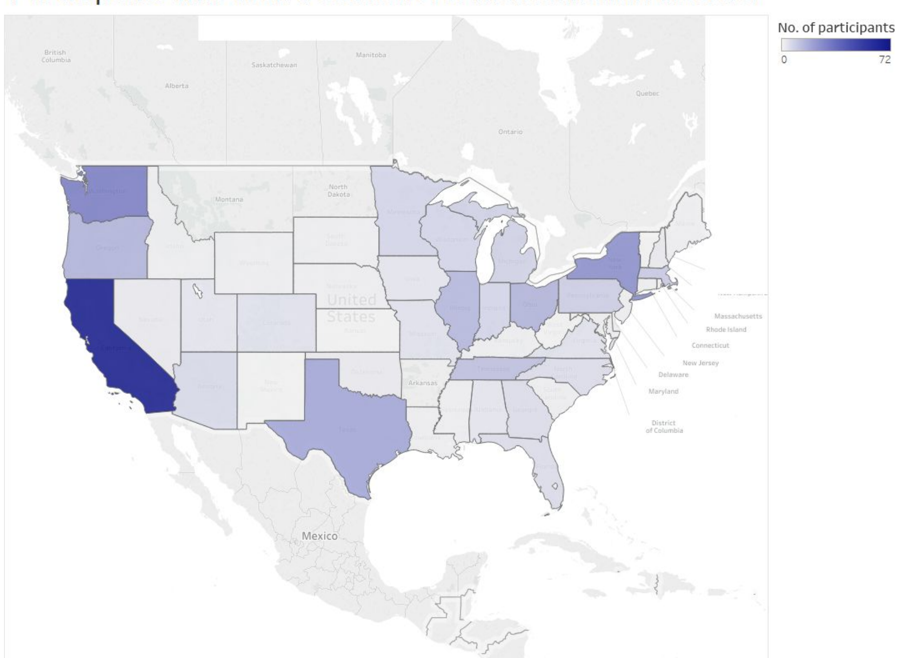


Figure 4: Choropleth showing distribution of participants across US states that took treatment for mental health disorder

Participants who had remote work benefits

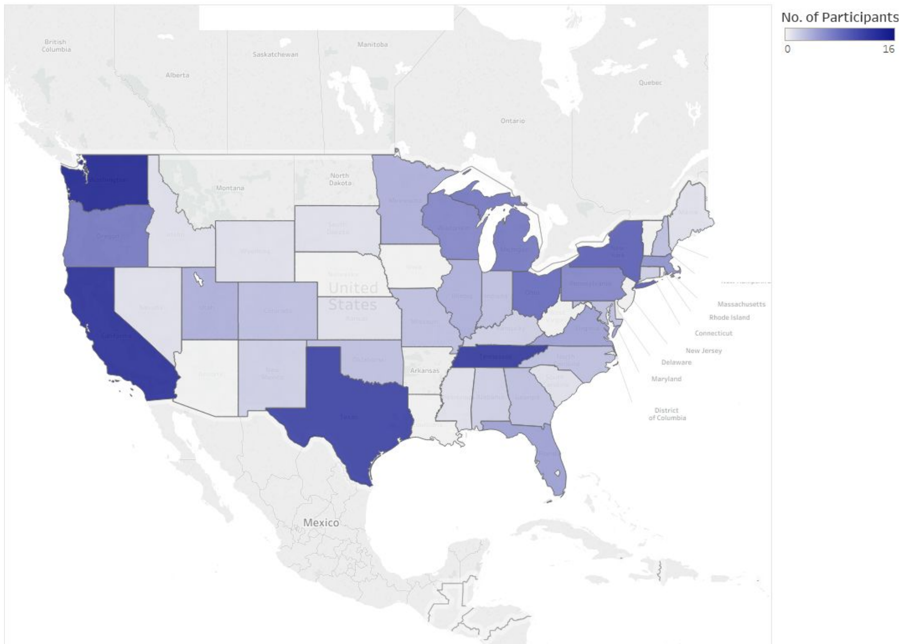


Figure 5: Choropleth showing distribution of participants across US states that had remote work benefits

The choropleth shows that states like California, Illinois, Texas, and New York have the most amount of individuals who have both family history of mental illness and have been taking treatment. This shows that people in these states are more likely to be taking treatment or soon going to be taking treatment if they have a family history already. By knowing this we can start intervention programs for people who are at risk, and get them the help that they need in the workplace to make their lives easier.

Work Environment

We made a mosaic plot **(Figure 6)** showing the likelihood of participants to discuss mental health issues with their supervisor and/or coworkers. Since we had three variables, a mosaic plot shows the relationship best since it is able to plot the proportion for each variable with width and height, and number of responses with area. The mosaic plot is also able to use color to distinguish answers from each other, making it easy to read.

The width of each rectangle shows the proportion of respondents that put down that answer for if they would talk to their supervisor about mental health. The height of each rectangle shows the proportion of respondents that put down that answer for if they would talk to their coworkers about mental health.

From the mosaic plot, we can see how respondents felt about discussing mental health issues at work. This is important to know, as it helps to be able to comfortably talk about a mental health issue at work, especially with a supervisor. The supervisor could help them lower their stress level if needed and allow them to take time off.

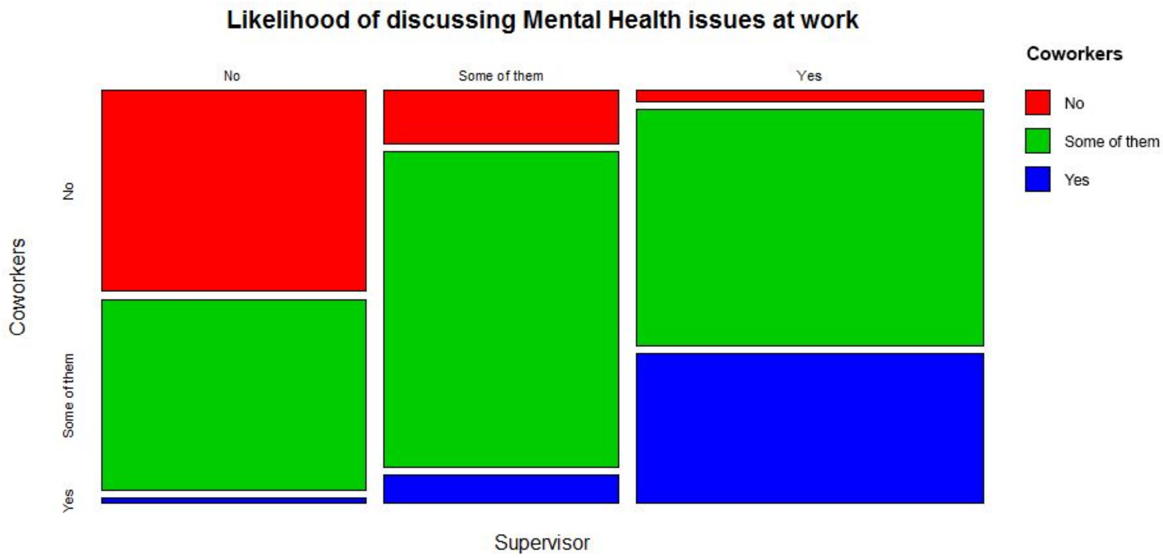


Figure 6: Mosaic plot showing how likely respondents are to discuss about mental health issues with coworkers and/or supervisors

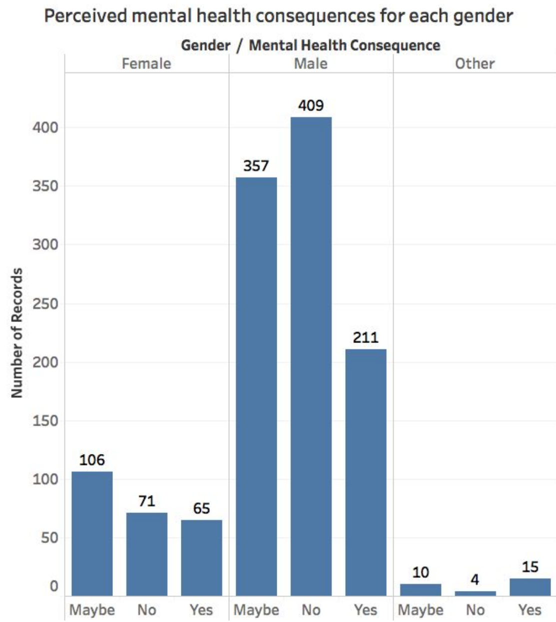
The plot shows that if you are less likely to be able to talk to your coworker, then you are less likely to be able to talk to your supervisor. Conversely, if you are more likely to be able to talk to your coworkers, then you are more likely to talk to your supervisor. Very few respondents said that they would be able to talk to their coworkers but not their supervisor, or vice versa. This shows that respondents felt similar comfortability with their coworkers and supervisors to talk about mental health conditions.

Looking across fields, we can see that there are a lot of respondents that said they would talk to some of their coworkers. It is a good sign that people feel comfortable in their workplace to talk to at least someone even if it is just their friends. A lot of respondents also said they would talk to some of their supervisors. It is important for workers to feel comfortable with all their supervisors, but feeling comfortable with at least some of them is an indication of a good work environment.

Gender

We did bar graphs and boxplots to see how gender plays a role and how different genders have different opinions on mental health. The most number of responses came from males, followed by females, and then from individuals that didn't identify as either female or male. We went through the data to make sure responses that were marked 'M' or 'F', or ones with spelling mistakes, got categorized correctly. We compared how males, females, and others differed in their responses the survey.

In **Figure 7**, we plotted a variable that showed whether respondents thought that discussing mental health issues at work would have negative consequences. The bar graph shows that most females and respondents who identified as 'Other' thought there might be negative consequences, and most males felt there wouldn't be negative consequences.



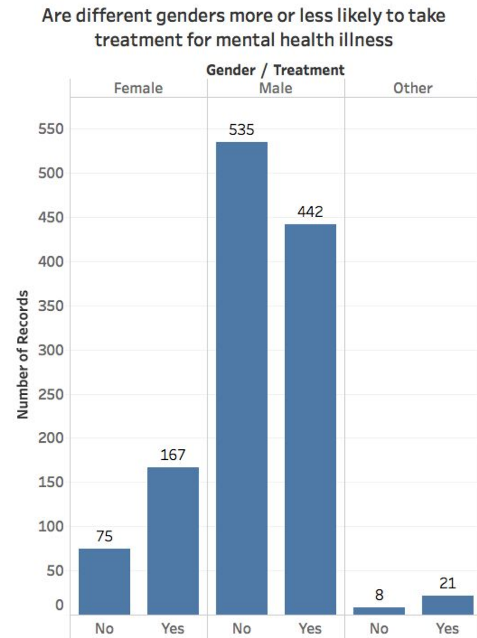
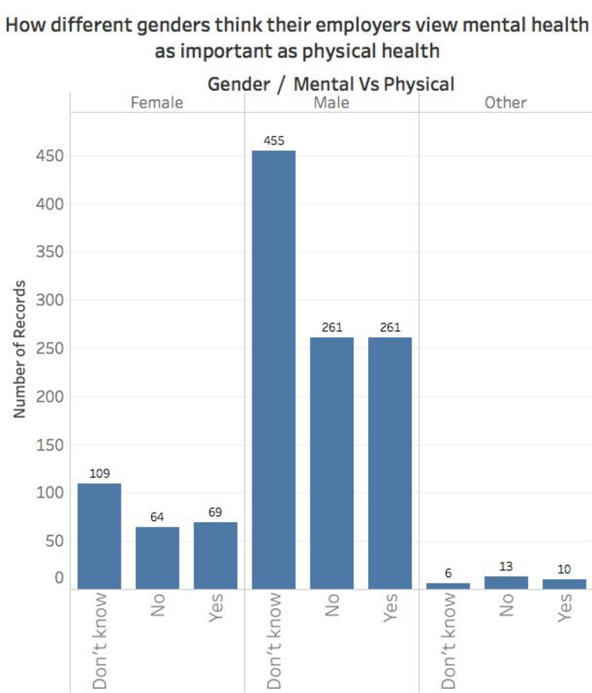


Figure 7: Bar graph showing perceived consequences to discussing mental health issues for each gender

Figure 8: Bar graph showing how likely each gender group are to take treatment for mental health illness



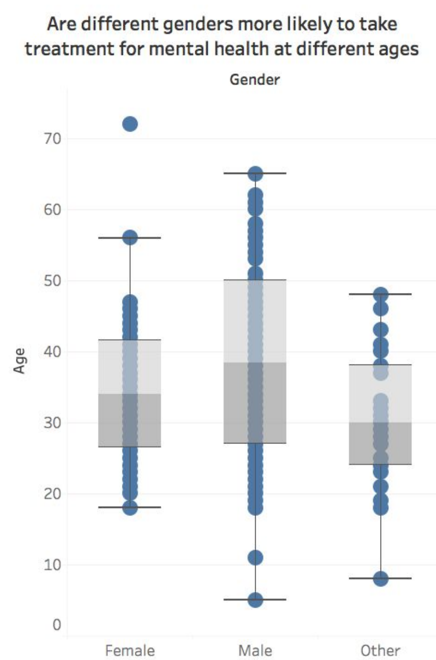


Figure 9: Bar graph showing how different gender groups' opinions about whether their employer values mental health as much as physical health

Figure 10: Boxplot showing age and gender distribution of respondents who are likely to take treatment for mental health illness

Figure 8 shows that females and respondents that identified as 'Other' were more likely to take treatment for mental health, whereas males were less likely to take treatment.

In **Figure 9**, we can see the distribution of different genders' opinions about whether their employer viewed mental health as important as physical health. Females and males had a mostly similar distribution with majority of each gender responding that they did not know, and around an equal amount agreeing and disagreeing. Respondents who identified as 'Other' seemed mostly indifferent towards the subject, with the majority of them disagreeing that employers viewed mental health as important as physical health.

Figure 10 shows the distribution of ages for different genders who are getting treatment. It shows that women are more likely to take treatment at an earlier age than males, and respondents who identified as 'Other' took treatment at an earlier age than males and females.

Based on all four graphs we can say that males are more likely to develop a mental health condition at a higher age than females, and are also less likely to take treatment. Males are also less likely to think that there would be negative consequences at the workplace for discussing mental health issues. Respondents that identify as 'Other' are the most likely to take treatment for a mental health condition, and at a younger age than males and females. These respondents also felt there would be a negative consequence to discussing mental health issues. This information is important to know to enable workplaces to improve mental health conditions for all genders, and implement training workshops to employees on how to show this to different genders. These could be geared toward individuals that identify as 'Other' or to younger females because they are the most susceptible.

Treatment and Benefits

To improve upon our exploratory visualizations regarding the correlation between treatment and benefits, we created a contingency plot **(Figure 11)** to clarify the connection between respondents who have benefits and respondents who need treatment, and to see how gender plays a role.

The plot was created using Tableau, with a sequential color scheme showing the number of responses. We changed the palette to show the difference in responses better. The Y-axis labels were rotated to avoid confusion with X-axis labels. Mark labels for each square in the contingency plot were enabled to further clarify the number of records in each category, so it is easy to compare the squares with each other. We had it broken down by different genders to show how males, females, and others differed in their answers if they took treatment.

The plot shows that while majority of respondents who need treatment have benefits, there is still a large number of respondents who do not have benefits, or do not know if they have benefits. This number is actually larger than the respondents who have benefits. This is a cause of concern, because people who need treatment should be provided benefits by the company. It is also worrying that such a large number of respondents are unaware of their benefits package.

There are many respondents who were taking treatment that said they didn't know if they had mental health benefits. It is important for employers to spread awareness about their benefits so people can take advantage of it and get the proper treatment they need, instead of having to pay more for treatment and medication when they might not have to.

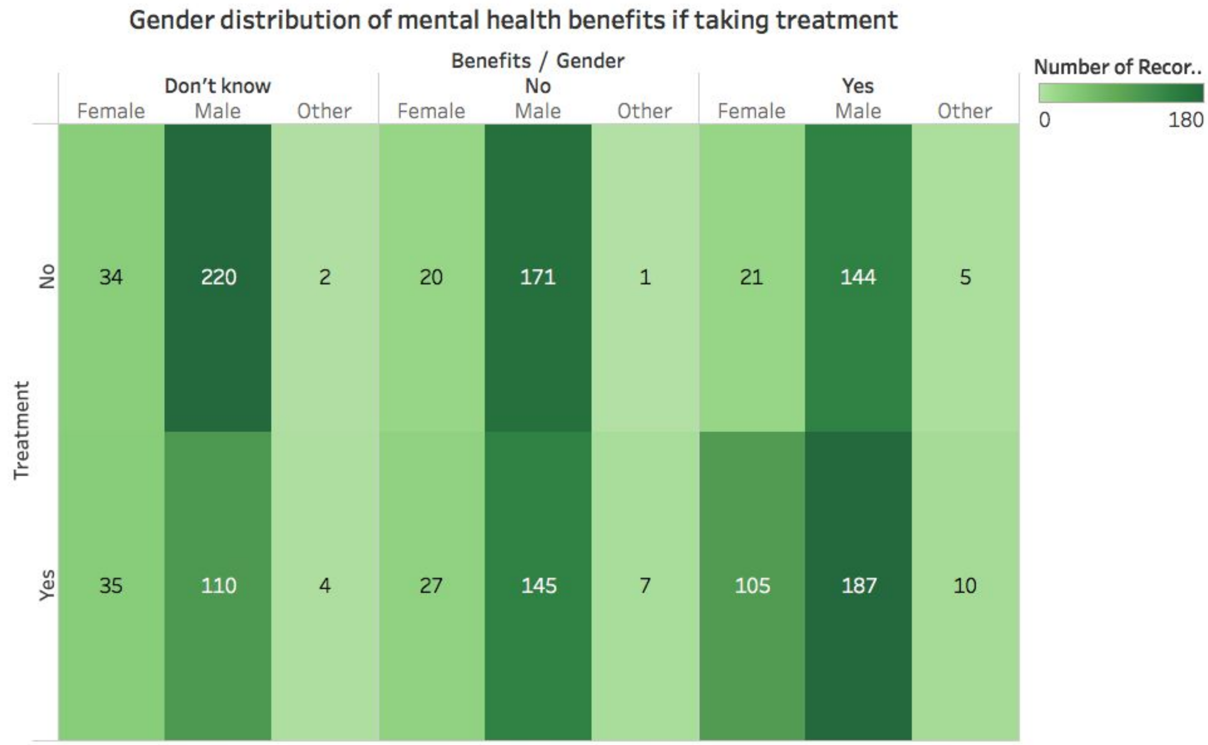


Figure 11: Contingency plot showing how genders differ in receiving mental health benefits if they are taking treatment or not

When looking at the overall breakdown of gender, females and respondents who identified as 'Other' had more of a similar distribution than males. The most clear conclusion from this graph is that if females were taking treatment, they were more likely to have benefits for mental health. This could be because they seek out jobs that have benefits.

Analysis and Discussion

Though the data was from a global survey, we had very limited information from countries besides the USA. We tried to compare treatments for respondents in the US to countries across the globe, but since there was insufficient data for other countries, we felt the comparison would not be valid. As a result, we decided to work only with survey data that was US-specific.

There were timestamps from when the participant took the survey. We wanted to see if there was difference in responses depending on the season. However, most of the responses were from August, and there were a few for later months, which we think were from respondents that answered the survey later than when it was given out. A future direction could be to see if there is a difference in responses depending on seasons and see how seasonal depression plays a role.

We also wanted to see what mental health disorders respondents reported the most. Since there were so many terms, we felt the best way to visualize these would be using a word cloud. We also wanted to visualize respondent comments from the survey to get a glimpse of what they said.

Figure 12 shows the distribution of mental health disorders that participants reported. The most commonly reported ones were depression, bipolar disorder, social disorder and mood disorder. **Figure 13** shows the comments that participants wrote in the survey. Most respondents mentioned their employer, company, work, and issues. This could show that people are affected by mental health issues at work and with their employer. Based on these two word clouds, we hope to visualize for employers why it is important to be aware of mental health conditions, and to provide a safe environment for everyone.

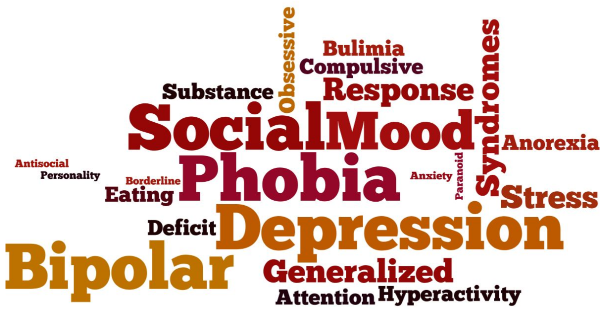


Figure 12: Word cloud of mental health disorders

Figure 13: Word cloud of participant comments

Future Directions and Suggestions for Employers

If we had more time, we would have liked to see how respondents' comments differed depending on if they were taking treatment or not. It would also be interesting to see if the survey results differed for companies in a different field, such as healthcare or education. When looking at the data, we don't know the people who aren't aware if they need treatment or if they have a family history for mental health disorders and could have answered incorrectly. A next step for future analysis could be more studies on people who aren't aware of their family history and/or if they need treatment.

Based on all of our graphs, some suggestions to improve workplace environments would be to improve awareness of issues that exacerbate employees' mental health, and to have pro-active measures to ensure employees' best health interests. Employers should make their health benefits more transparent and provide remote work opportunities for employees with mental health conditions. Discussion should be encouraged and workplaces should try to improve gender equality, and attend to requirements of employees who identify as other genders than male or female.

Appendix

Sources

R Code for Mosaic Plot

```
library(readr)
#Importing data
mydata <- read_csv("D:/survey_cleaned.csv")
View(mydata)

# Selecting variables - supervisor, coworkers, mental_health_interview
variable <- c("supervisor","coworkers","mental_health_interview")
newdata <- mydata[variable]

#Making table for the columns we required
totals = table(newdata$supervisor,newdata$coworkers)
totals

#Plotting mosaic graph
mosaicplot(totals, main = "Likelihood of discussing Mental Health issues at
work", col = c(3,4,2), xlab = "Supervisor", ylab = "Coworkers")
```

Tools

1. Tableau
2. Wordle (Used to create word clouds)

Links

[1] NAMI: National Alliance on Mental Illness
http://www.nami.org/Learn-More/Mental-Health-Conditions

Data Set:
https://www.kaggle.com/osmi/mental-health-in-tech-survey

Individual Report

Anika

For our, I chose our dataset from kaggle and was also responsible for doing the background research on mental health in order to see how our visualization tells the story and what are some insights we can gain. I also helped in seeing how our visualizations related to the story we wanted to tell and I helped in coming up with suggestions for the workplace based on what we saw in the visualizations. For example, there was a difference in the genders on how they think their workplace views mental health, and there's a concentration of people taking treatment in bigger cities. Can companies use this information to change their work environment? I wrote the draft of the introduction in order to introduce our dataset and the relevance it has in our society.

I also worked on the gender explanatory visualizations using a combination of exploratory visualizations that we had done before. I made sure that the visualizations we chose actually showed a difference between the genders, and contributed to the story. I did the exploratory visualization between family history and treatment and thought of how the do our explanatory visualization showing the correlation between the two using maps. I wrote the analysis for the gender explanatory visualization and the map visualization. I included the conclusions we got from the visualizations and then how that related to the story we wanted to tell.

I helped refine some of the visualizations that other people in my group had done. This includes changing the color so the difference is more obvious between variables. I also helped in identifying problems with visualizations, like if there were too many variables, if the axis and title needed to bigger/centered, or if we needed values to show difference like in contingency plots or there's no needs based on the conclusions we wanted to draw like in the maps.

During this course I learned a lot about visualizations. I was able to make visualizations I had seen in journal articles that I had never seen before. Sometimes it's hard to put findings into words but with visualizations it's a lot easy to interpret. We learned a lot of different ways to show the same data, and it was cool to see what were wrong with certain graphs and how could we change it to make it better. In our data, we had a lot of categorical variables and not as much numerical data. I thought this would be very hard to show in graphs as we dug more into it. However, after a lot of revisions we were able to show more complex graphs like contingency plots with multiple variables, and even learned how to do a mosaic plot on R. I had always seen word clouds on reports, and it was interesting to see how to make one, and change the data to get the most out of it. I will definitely be using what I learned in this class for reports in the future. I have already started in my internship, especially with refining simple bar graphs to make the most clear conclusions to my team.

Deepak

Contributions

I helped schedule and coordinate team meetings by creating a team agreement document with each team member's availability timings. As the team liaison, I also gathered questions on behalf of the team and presented them to the professor.

I created the contingency plot visualizations showing likelihood of discussing mental health issues at work and whether respondents requiring treatment have benefits. These were later converted to a mosaic plot and other more explanatory visualizations. I also created exploratory univariate scatterplot visualizations to show correlations between age, gender and treatment.

I drafted the final presentation and report template, and edited content to make the writing more succinct and clear. I also formatted the entire report to make it more scannable. For the final presentation, I created the animated GIF that showed the relationship between family history of mental illness and requirement of treatment. I did this by adding the two choropleths in Photoshop and using the Timeline feature to do motion tweening between two frames.

Reflection

I learned that sourcing data depending on certain requirements is very challenging and an important skill for data projects. Initially, I was interested in looking for data in the sports or health field, or a combination of both; specifically data related to soccer player injuries. I thought it was very difficult to find something as specific as that, unless it was publicly available on sites like Kaggle.com.

Looking back at our own project, I think we did a great job despite having multiple categorical data, which limited our ability to do complex graphing methods such as network graphs or other time-series plots.

I am glad to have had hands-on experience with both Tableau and R in this class, although I wish I explored more into D3 or Shiny for interactive visualizations. I think it would be useful if we were taught how to rig some of the D3 gallery examples to make our own interactive visualizations.

Jaymin

As part of our project, after we selected data set from Kaggle, I cleaned the data set. I saw some missing values, some unstructured values (For example, for males there were, M, male, Male etc) in the data, therefore I cleaned the data. I also created hash values for some of our variables which are useful to find number of variables in some of the visualization.

I started to explore data by creating boxplot for Treatment per age and gender. I created two choropleths with, if remote work affect on treatment and Increase of leave with increase in mental health consequences. I created bar chart which can describes in which conditions a person is required to take treatment. I also created two stacked bar graphs, how easy for people to take leave who takes treatment or not and size of company for people who are taking treatment. And also created Pie chart as a part of exploratory just to see the proportion of people getting treatment per gender.

I also created word cloud for mental health disorders that people are facing. Lastly, I created mosaic plot using R for likelihood of discussing mental health issues at work, with coworkers and supervisors.

Reflection:

I did not have much knowledge about visualization when I took this class, I heard of Tableau but never used it before. After taking this class, I get better idea of how a graph looks like and what are the possible chunks in a graph. Now whenever I see a graph I tried to find chunks in it and try to think of other possible ways to make it look better. Tableau became handy with this subject which I never used, especially some typical homework problems make it somewhat easy to use.

Project was the best way to apply the knowledge whatever we gain from the class, we can create different types of graphs we learnt in the class. At some point while doing project, I felt that as we are having more categorical variables we are lacking to get experience of making some graphs like heat maps and network graphs.But overall I am satisfied with whatever we come up with was useful and in real world data we can have this kind of categorical data.

Kripa

Initially, as the part of the project all of us differently researched for what we should use as our data source. Finally, we selected the dataset which Anika suggested. I contributed in brainstorming about the data cleaning process, how to go about the exploratory and explanatory visualizations. I had made couple of exploratory visualizations. I helped editing and reviewing the project report and the presentation. I worked with Anika on the exploratory visualization for the family history, treatment and wellness program and helped writing the explanation for the same. I reviewed the report making sure that we answered and covered all the important aspects suggested by the professor that should be there in the project.

This was one of the best courses that I have had so far. I really had fun working with my group and implementing the knowledge accumulated during the quarter in the class. This project helped me understand the importance of the research related to data and data cleaning. It gave me basic insights on how can we use data visualization for predicting and implementing awareness regarding a cause and give people more insights about what the predictions show from the visualizations. It was a great course , which can be a basic building block for my career in data science.