I590 Data Visualization
Midterm Report

December 15, 2017

Team Members:

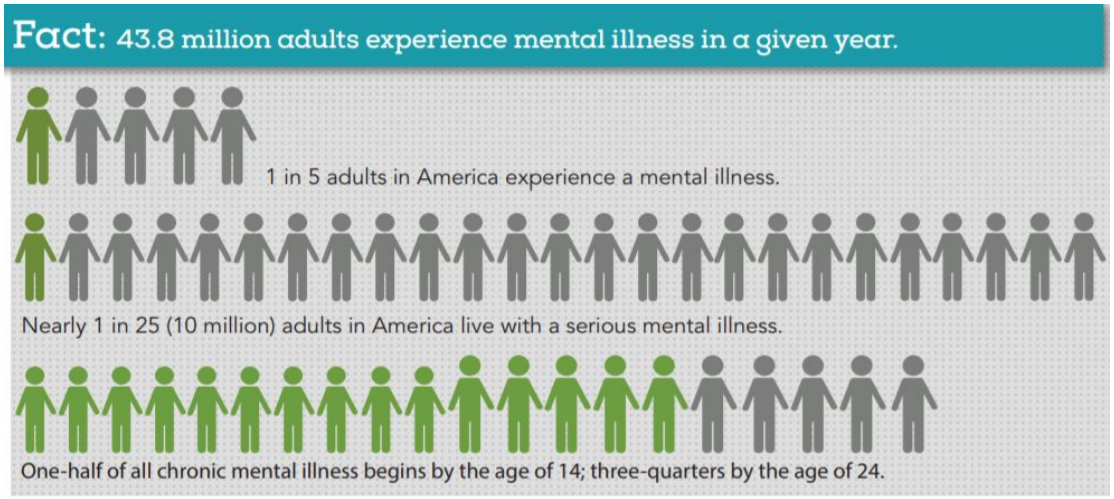
- Jessica Johnson
- Brent Matthews
- Saurabh Jain

1. INTRODUCTION

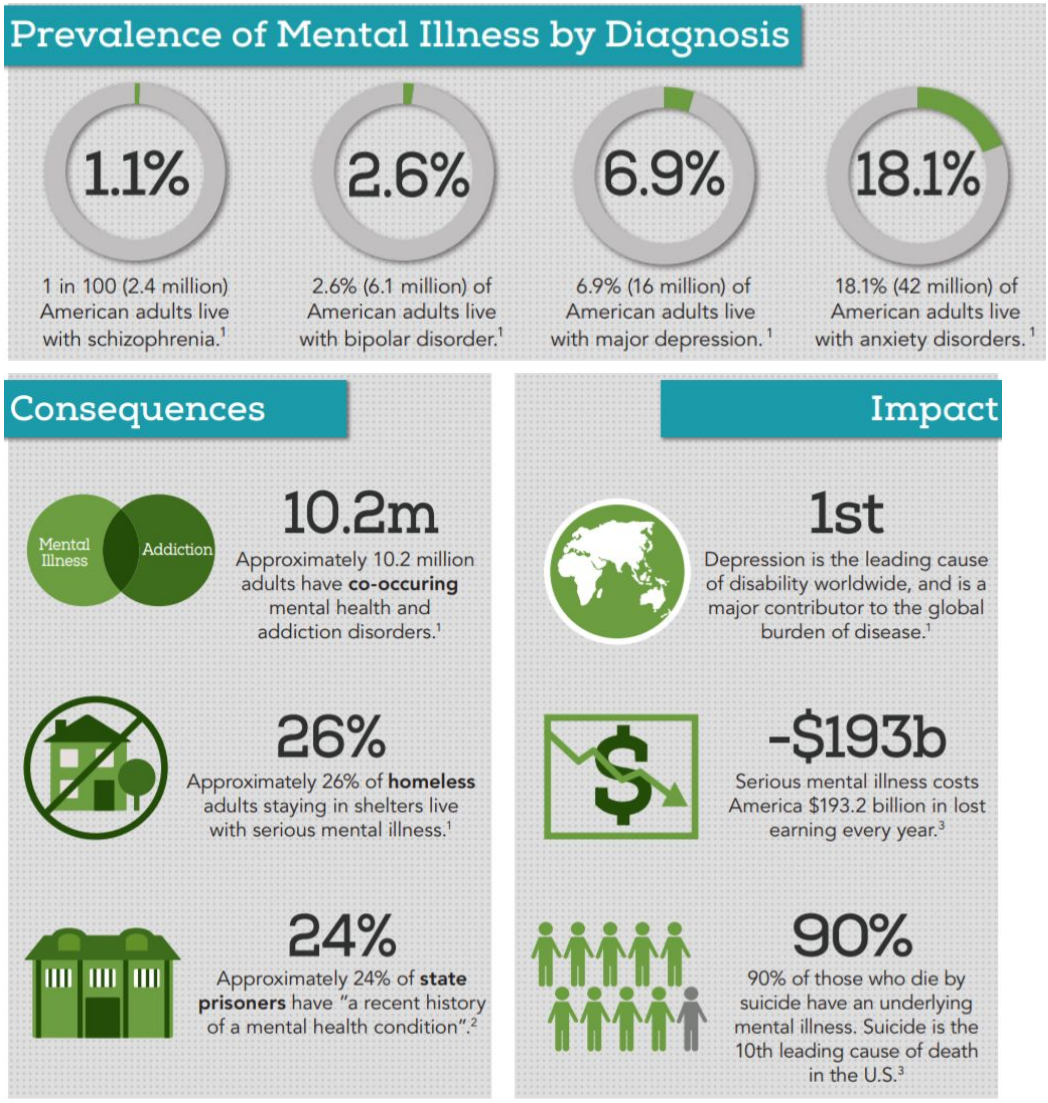
Motivation:

This project will look into mental health and attempt to bring light to the amount of people that are suffering from this disease. This information will be useful for employers in understanding the scope of the problem and give insight into dealing with their team members who suffer from this disease. Understanding how workers feel about sharing their mental illness with their co-workers and managers as well as how they are treated is necessary to improve workplace conditions and overall efficiency. We will also uncover valuable insight that may provide useful for the treatment of mental illness as well as reducing the stigma associated with it.

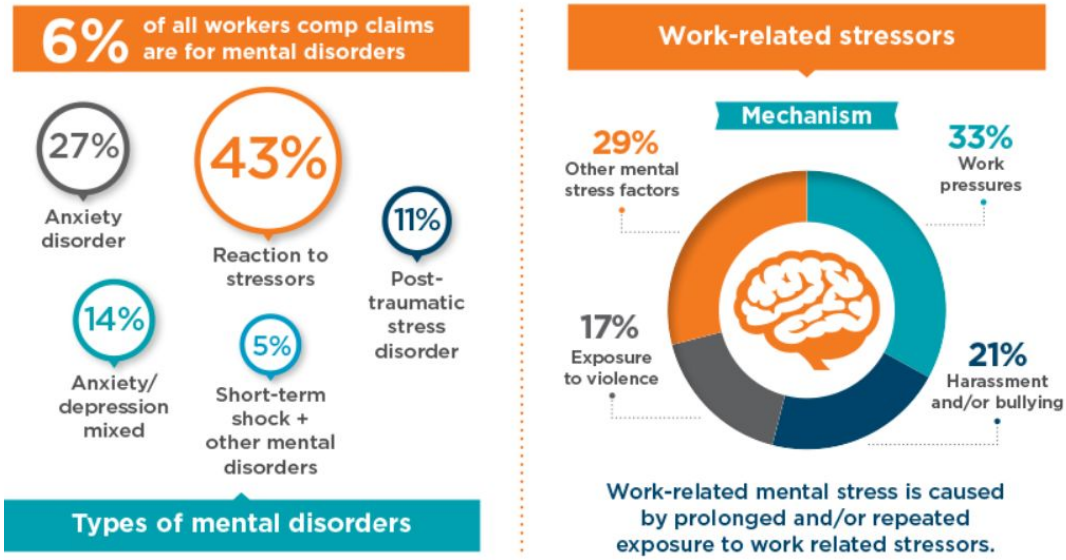
National Alliance on Mental Illness has released statistics regarding mental illness in the United States. One in five adults suffer from mental illness and it most likely starts before they become 25 years of age. As young people are completing their education and entering the workplace, those suffering from mental illness will face a greater challenge integrating into their new positions.



The types of mental illness suffered by the US population varies but are primarily anxiety and depression. Over 18 percent of the population struggle with anxiety and nearly 7 percent suffer from depression. The consequences due to these illnesses are staggering. Over a quarter of individuals with mental illness are homeless and and many suffer from addiction as well. Depression is a leading cause of disability, preventing many from holding a job. The financial implications are immense, with 193 billion dollars lost in earnings each year.



These issues are not only in the US. Safe Work Australia created a visual outlining work related to mental disorders showing that 6% of comp claims are due to mental disorders. 43 percent of these claims are due to the stress of the workplace, which can exacerbate existing conditions.



The Canadian Centre for Occupational Health and Safety made a list of ways employers can help to improve mental health in the Canadian workplace. This includes concepts such as improving work processes and tasks, clarify roles, prioritizing mental health within the workplace, conflict resolution as well as increased dialogue.

It is clear that depression affects individuals in many aspects of their lives. One such place is the workplace and studying the effects of how people feel in regard to their mental illness within the workplace can help improve working conditions and individuals well-being. Moreover, many companies are looking to support, and having more insights generated through data would be helpful in this endeavor

We have focused in on the tech industry as many employees work alone and face a great deal of pressure. Entrepreneurs, start-up employees, and workers in large corporations are face stressors as well as the need to perform well and deliver results. There is a sense of respect given to those who put in long hours and are able to create a working product or service that is superior. This competitive, solitary environment, where weakness is not respected, would be very difficult to navigate while enduring mental illness.

Existing Work:

We have found various data sets pertaining to mental illness. The main dataset we intend to use is from Open Sourcing Mental Illness (OSMI), a non-profit organization dedicated to improving conditions in the workplace for those suffering from mental illness. Several individuals have already explored the data and added their findings online. We hope to run our own analysis and possibly make a connection within the data that has not been established.

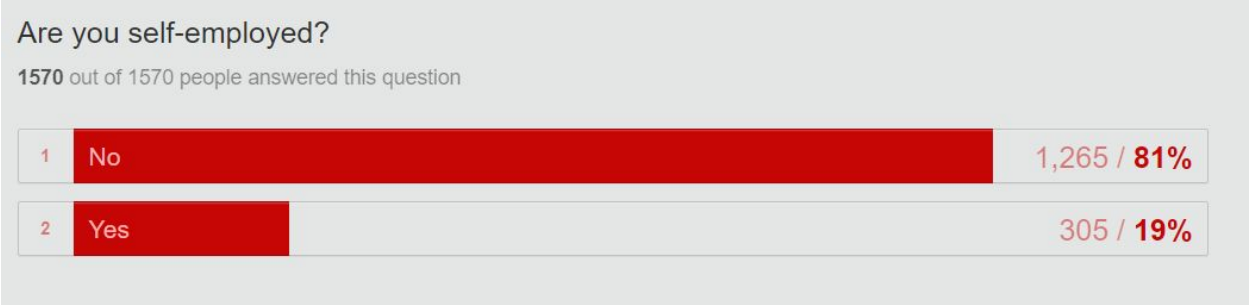
OSMI Data:

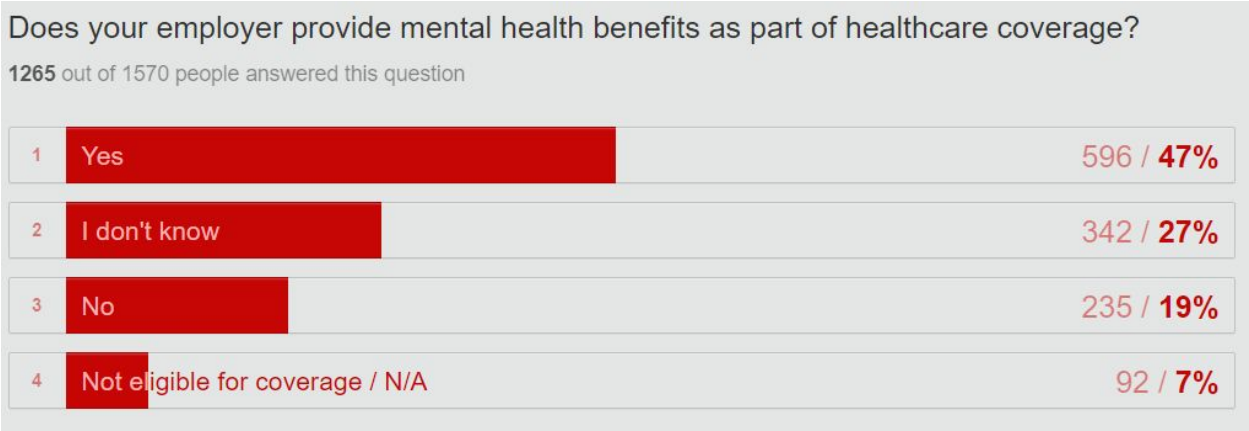
The 2016 survey currently has over 1400 responses to questions related to attitudes toward mental health and the frequency of mental health disorders in the tech sector. According to OSMI they are looking for understanding in “how mental health is viewed within the tech/IT workplace, and the prevalence of certain mental health disorders within the tech industry.”

This OSMI survey was also completed in 2014 and there were various different analyses done on the data. One such analysis is located at the NYC Data Science blog. They show various different aspects of the data that were reviewed. This information will help us to focus on areas that may not have been previously analyzed.

Some existing visualizations related to OSMI data and observations from the same:

1. OSMI Report 2016
The reports available in website for Open Sourcing Mental Illness contains very basic visualizations. It contains a bar graph for each question for just displaying the count of various answers. Although these graphs are useful for initial analysis but lack in comparing data of two columns to correlated dependency between them. We have tried to cover this by creating visualizations comparing multiple columns data and gathering more insights from them
Report for 2016 available in:
https://osmi.typeform.com/report/Ao6BTw/U76z
e.g.





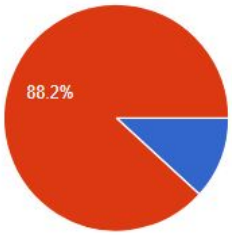
2. OSMI Report 2014

Another set of visualization for the year 2014 offer another set of similar basic visualization per column data. Some of the problems with this set of visualizations are:

- 1. It just shows the count of number of responses in a pie chart
- 2. Repeat of information first in pie chart and then again in text table
- 3. Colors chosen for pie chart are very bright, they could have been soften for easy visibility
- 4. A part of pie has percentage numbers, another pie does not
- 5. A legend along with color could have been provided for easy mapping
- 6. A better color palette could have been chosen where colors show some rhythm
- 7. In the third graph shown below it is almost impossible to recognize which pie belongs to which country

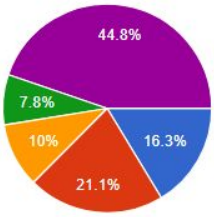
Report for 2014 available in:
https://docs.google.com/forms/d/1FqpxZXORRNDQ4niSeqLx0Qy2VfUod_qlAvpnOyTEB9k/viewanalytics

Are you self-employed?



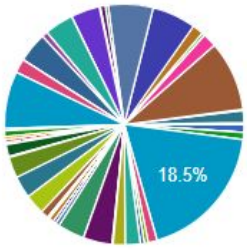
Yes	**146**	11.8%
No	**1096**	88.2%

How easy is it for you to take medical leave for a mental health condition?



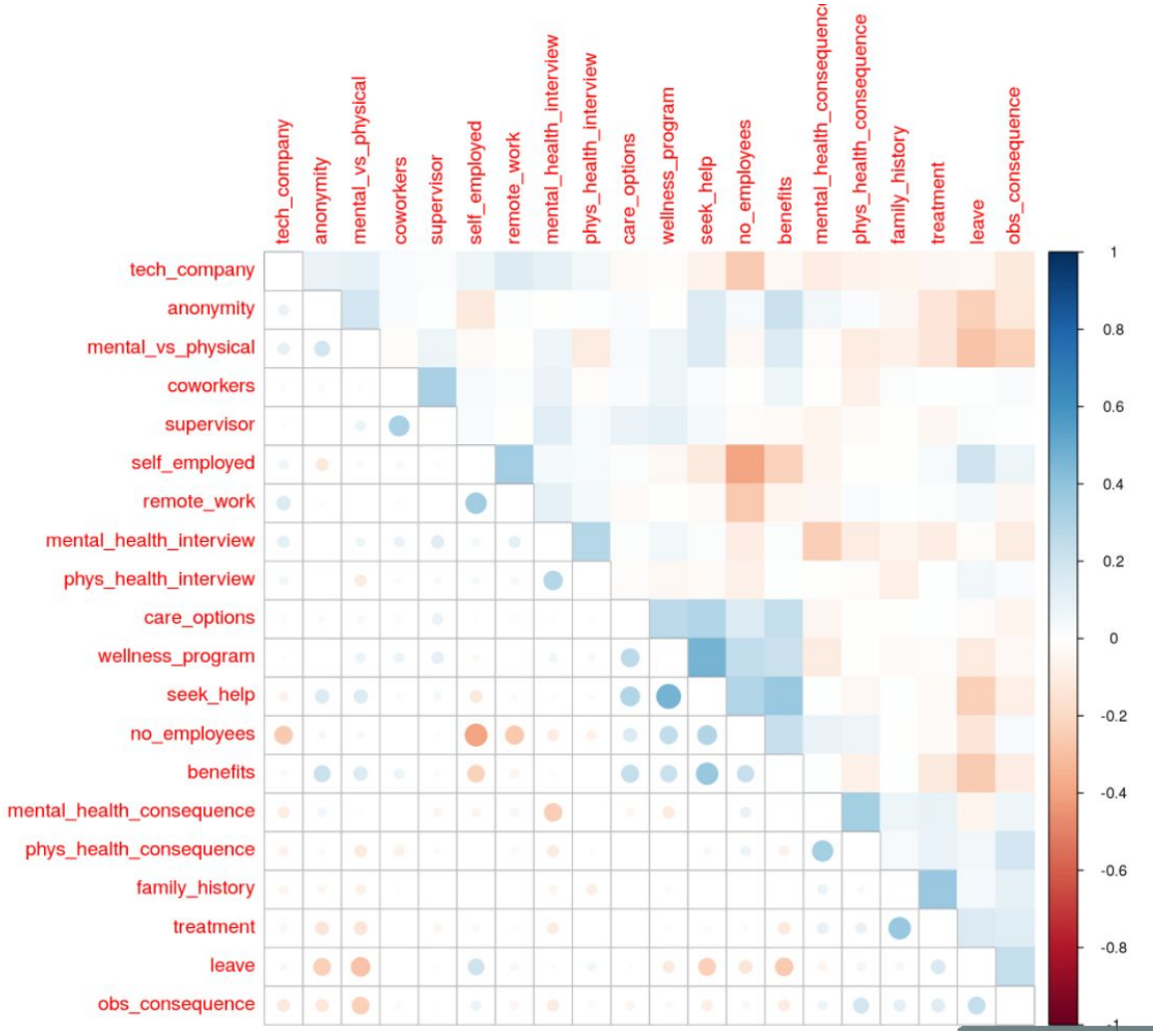
Very easy	**206**	16.3%
Somewhat easy	**266**	21.1%
Somewhat difficult	**126**	10%
Very difficult	**98**	7.8%
Don't know	**564**	44.8%

If you live in the United States, which state or territory do you live in?



AL	**8**	1.1%
AK	**0**	0%
AS	**0**	0%
AZ	**7**	0.9%
AR	**0**	0%
CA	**138**	18.5%
CO	**9**	1.2%
CT	**4**	0.5%
DE	**0**	0%

3. Kaggle submission: A version of visualization available as a submission in Kaggle for analysis related to OSMI data is available at:
https://www.kaggle.com/benjaminlott/machine-learning-for-mental-health-7-methods/notebook



Some positive features of this visualization:

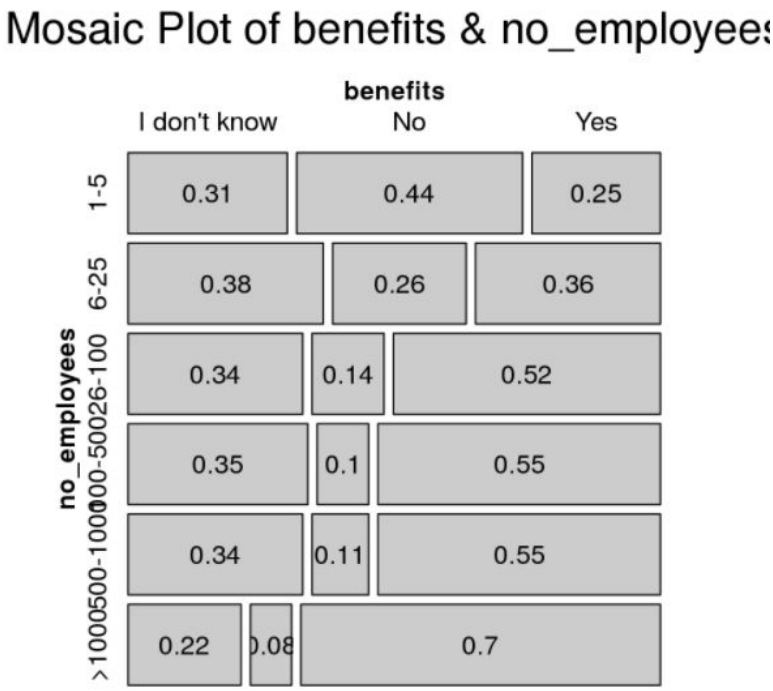
- 1. A good soft color palette has been used which allows easy viewing for eyes
- 2. The graph is able to convey the relationship and interdependence between all features using a single visualization

Some negative features of this visualization:

- 1. A single color with different intensity could also have been used to show dependency between two variables
- 2. Information is repeated as one half of diagonal values are sufficient to show details. Another half could have been used to show some more relevant information such as the actual correlation coefficient from 0 to 1.
- 3. In one half complete square is filled whereas in other half filled circles are being shown. They give an impression that different information is being conveyed, but looking closely we that same information is being conveyed

4. Kaggle submission: Another Kaggle submission gives relationship between two variables using Mosaic plot:
https://www.kaggle.com/c34klh123/survey-in-tech-workplace

Some observation on this graph:
Good points:



- 1. A new type of graph has been used to give info about relationship between two features
- 2. Width of each component clearly displays the importance of that feature combination
- Improvements that can be made
- 3. The title of graph has start and end text partially cut
- 4. Instead of using _ in axis title it would have been better to remove them
- 5. One of the number in small bars is not visible
- 6. The middle most bar belongs to No and No for two features but the head of the column in x axis is far away from that bar. This can cause some confusion or delay in understanding of graph
- 7. No color scheme has been used
- 8. A stacked bar chart also shows same data and uses less ink
- 9. In second graph the range for y-axis is not clear, some numbers have got so close it is difficult to read

Contribution:

We discovered various different interesting information while analyzing this data set. In particular one point that was specifically interesting was the that such a large number of respondents did not have health insurance or were not even aware if they did. This was a rather startling statistic and really needs to be addressed by the industry.

Along with that the attitudes towards mental health are somewhat alarming. Almost 50% of respondents did not believe that their company would treat mental health the same as physical health concerns.

This among other items of interest are discovered in the research we completed. We believe that this knowledge is useful and important to the understanding of mental health in the tech industry.

2. DATA AND METHODS

Ideas, sketches, prototypes:

Process: analysis of data, candidate visualization methods, failed experiments, why certain method did not work while others did, and so on

We did some data analysis and exploration using the dataworld website. This site allowed queries to be run based on your chosen parameters. This was used for some basic queries and initial understanding of the data.

Using the dataset provided from data world, we were able to import the data into a python jupyter notebook and complete analysis on all different aspects of the information. We utilized various different packages within python to complete a thorough analysis.

In the initial submittal we completed most of the python analysis using bar charts to describe the proportions of the respondents answers. In this final we utilized more pie charts to give better understanding of the breakdown by percentages. This provided us with much easier to understand information. Although we don't want to use pie charts when unnecessary, this case provided that they were useful.

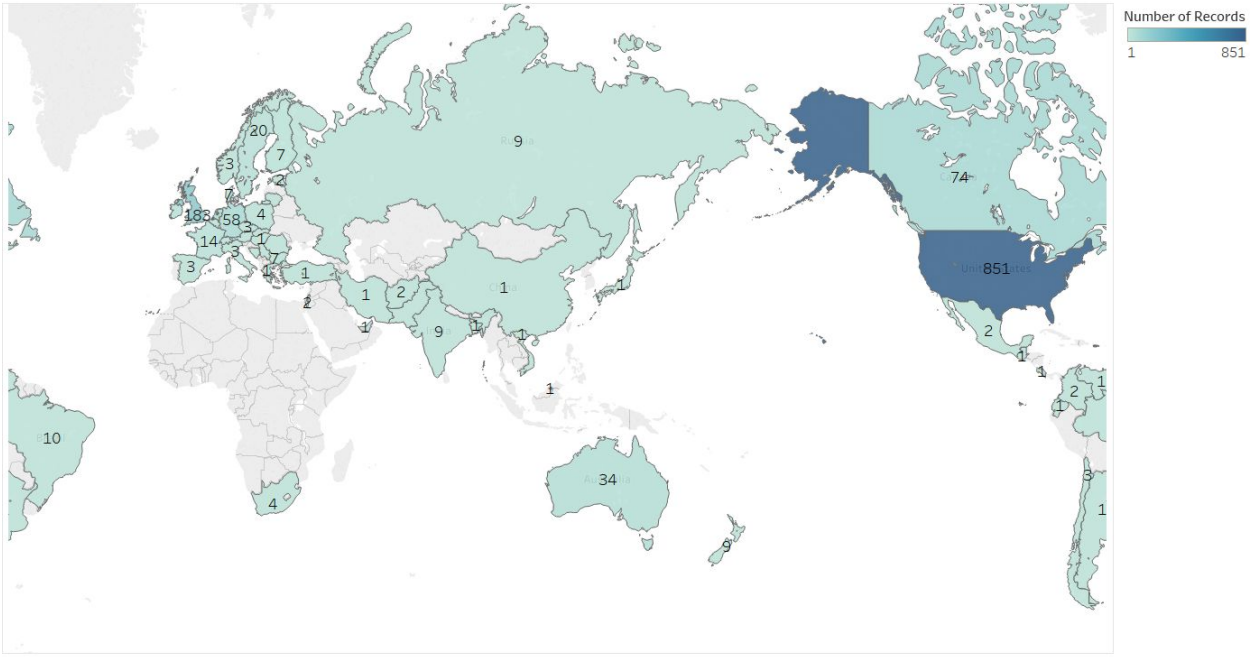
Visualization Method Selection:

Visualizations were considered within the context of the python packages as well as utilizing Tableau for geographical context. Python provided us with the basic visualizations for bar charts and pie charts. The information we analyzed did not give us a lot of different options for advanced visualizations with statistical analysis. Tableau gave us the ability to break down the plots by geographic area as well as providing us with interesting plot combinations.

3. RESULTS

In order to understand basic insights from the data we generated basic visualizations for some of the important features in the dataset.
We plotted this data on a geographic map of the world with color intensity signifying the number of participants. We clearly see that the maximum participants are from US, while the second highest count was in the UK, followed by Canada.

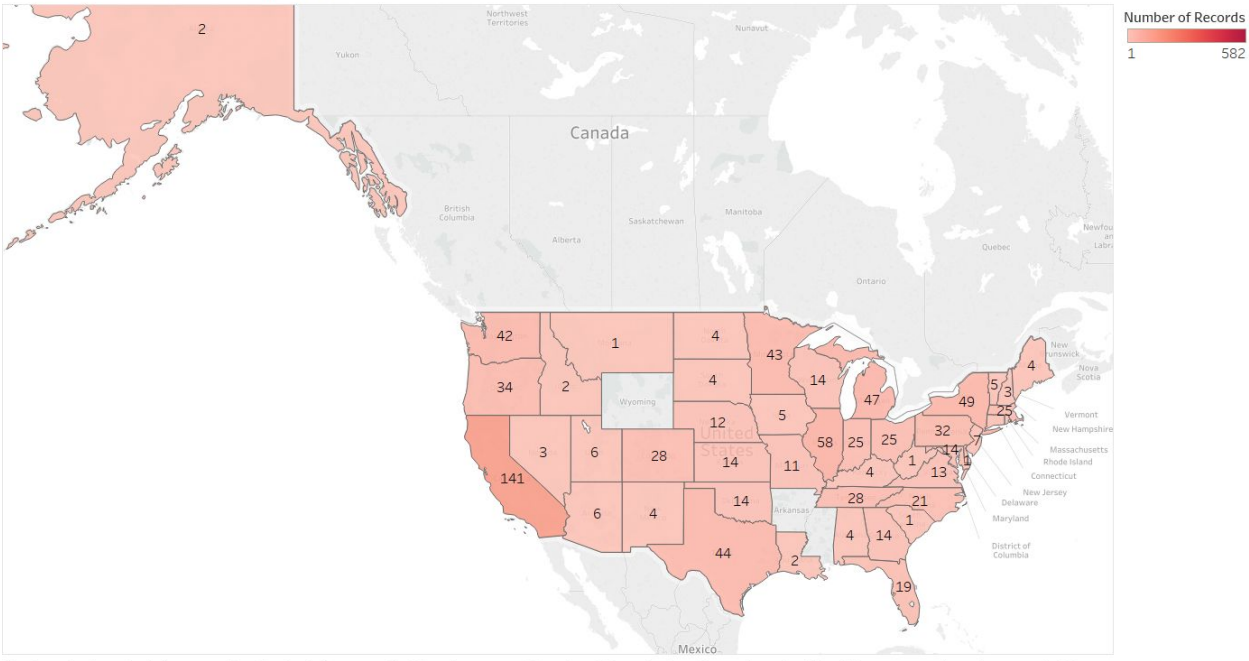
Countries in which the person works in



Map based on Longitude (generated) and Latitude (generated). Color shows sum of Number of Records. Details are shown for What country do you work in?.

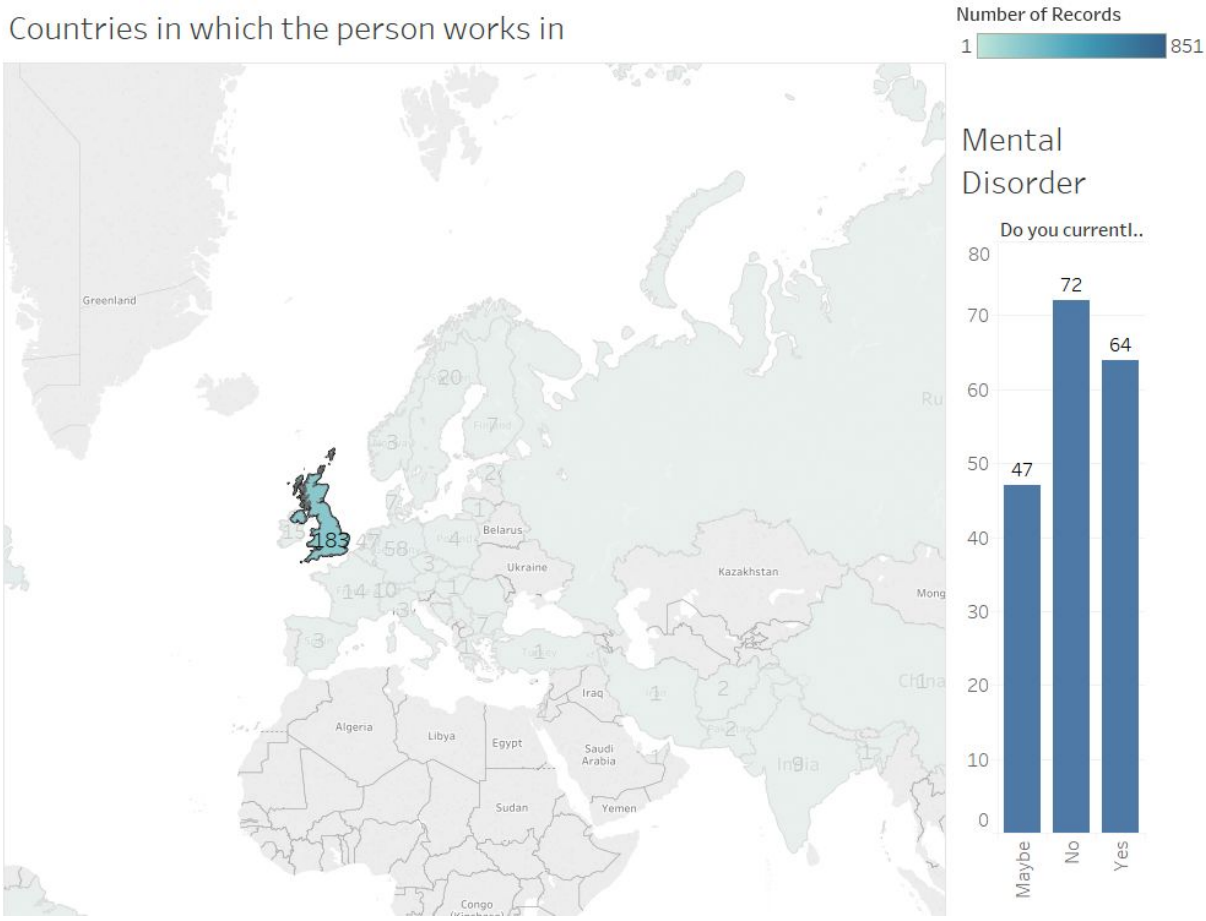
We looked at this information further to break out data for the United States. Out of the participants from US we plotted them by state to see how many participants belong to each state. This distribution is fairly evenly spread in various states with maximum participation from California which contains Silicon Valley where a significant number of a tech professionals reside. We looked at the geographical distribution of the tech industry by the number of people affected by mental disorder currently. We observe that more people are affected by mental disorder in the United States compared to other developed and developing countries. For this we prepared an interactive chart wherein selecting any country we get the number of people who currently have mental disorder.

US state in which the employee works in

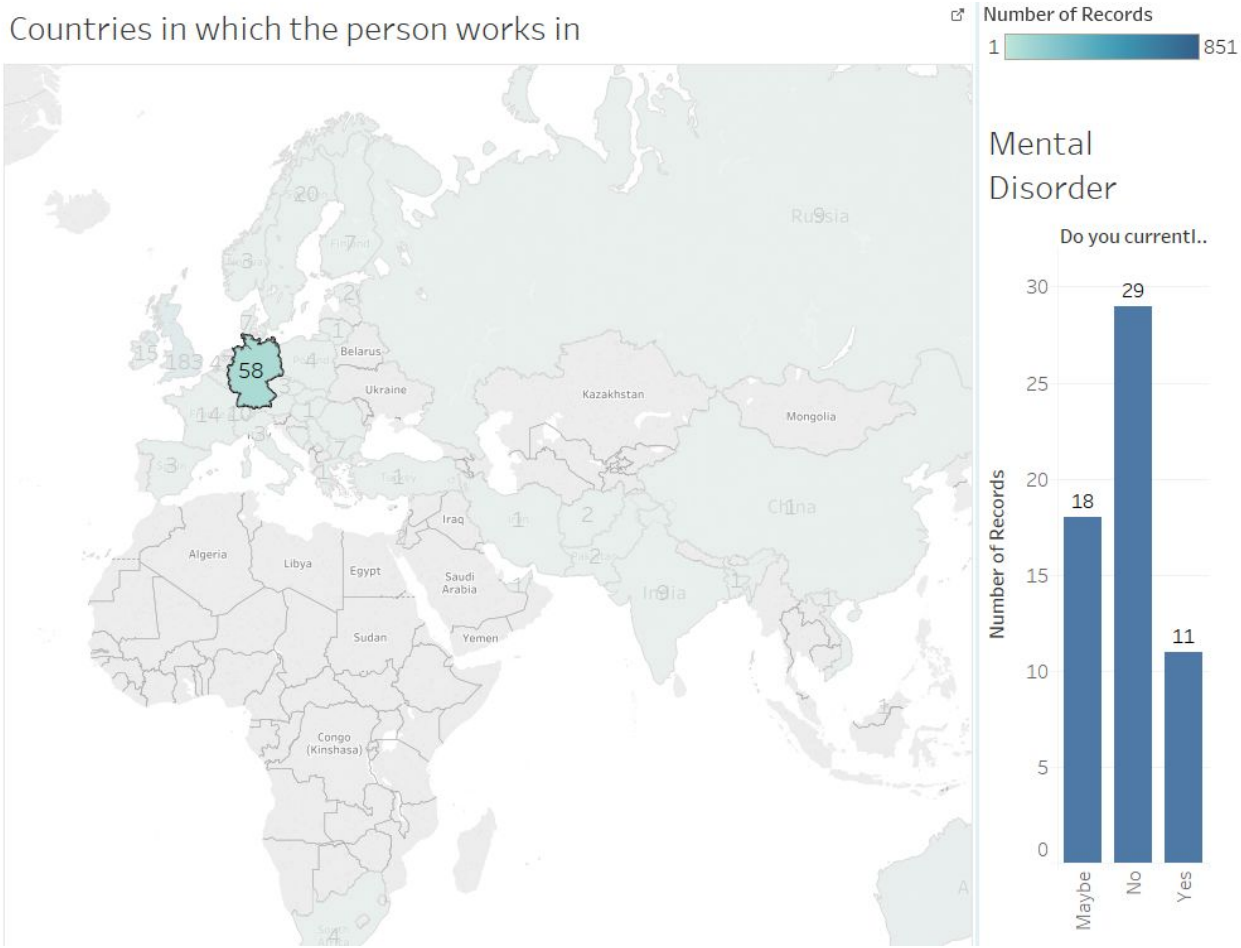


Map based on Longitude (generated) and Latitude (generated). Color shows sum of Number of Records. Details are shown for What US state or territory do you work in?.

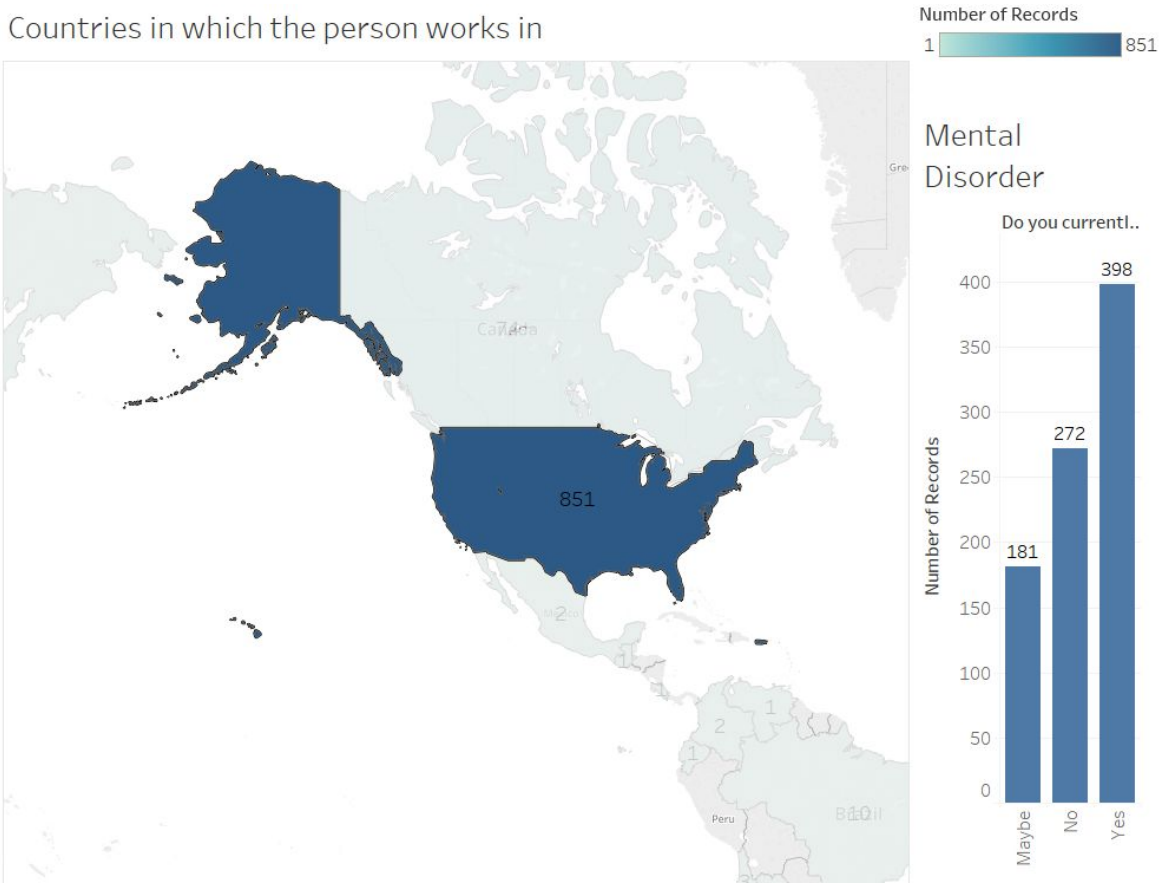
Breaking this down further we looked at the United Kingdom and found that 64 out of 183 respondents in survey having some sort of mental disorder



Looking at Germany, we found that 11 out of 58 participants have mental disorder of some sort.



Finally, looking at the United States: 398 out of 851 respondents agreed to having some sort of mental disorder

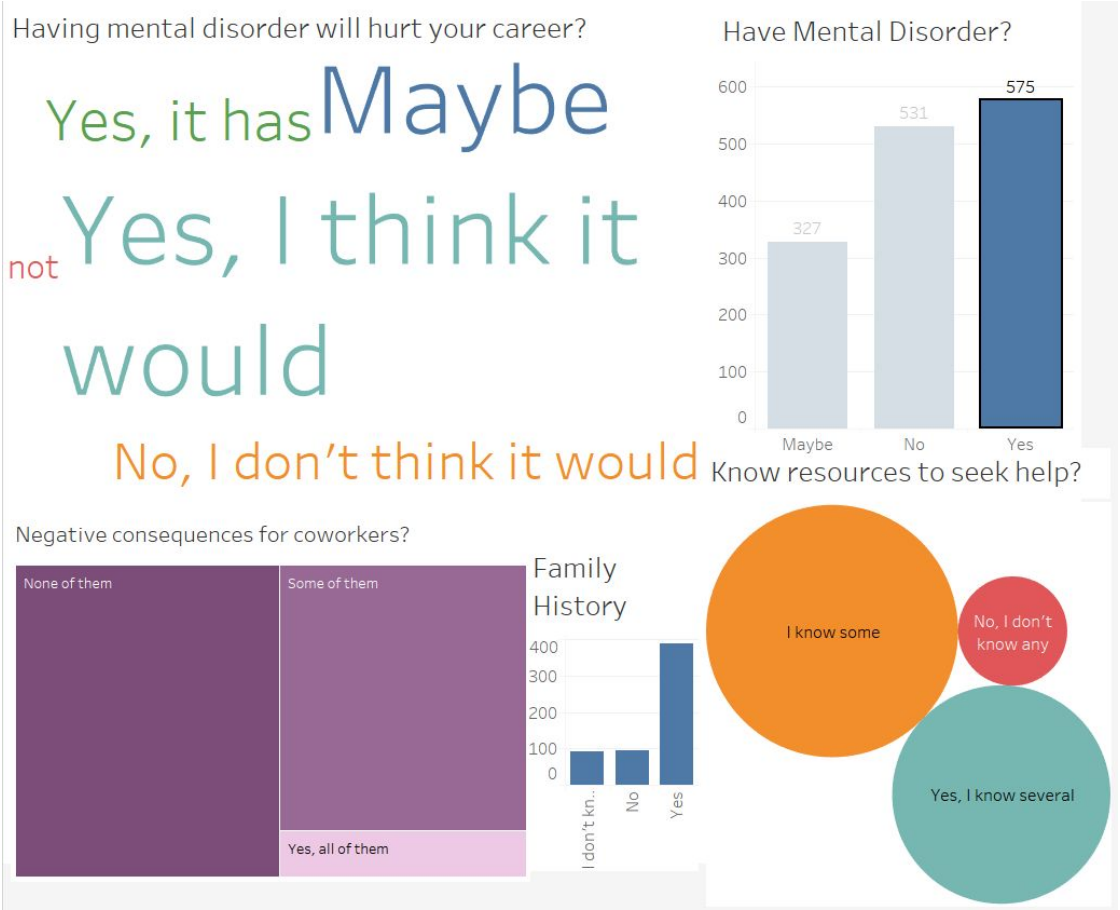


Getting insights from one feature / variable / question will help us understand the influence of that feature on results but if we can combine various features / questions at the same time we can get much better insights. In the current dataset we do this by creating an interactive visualization combining five questions/feature

1. Do you currently have mental disorder?
2. Do you know local and online resources to seek help for mental disorder?
3. Do you feel that being identified as a person with mental disorder will hurt your career
4. Do you have a family history of mental disorder
5. Do you hear of or observe negative consequences for coworkers having mental disorder

Since this is an interactive visualization, selecting any data automatically applies a filter on all other sub visualizations which are part of it. Thus to understand the interplay and get better insights we filter on the three answers for question “Having mental disorder” and observe the values of other questions.

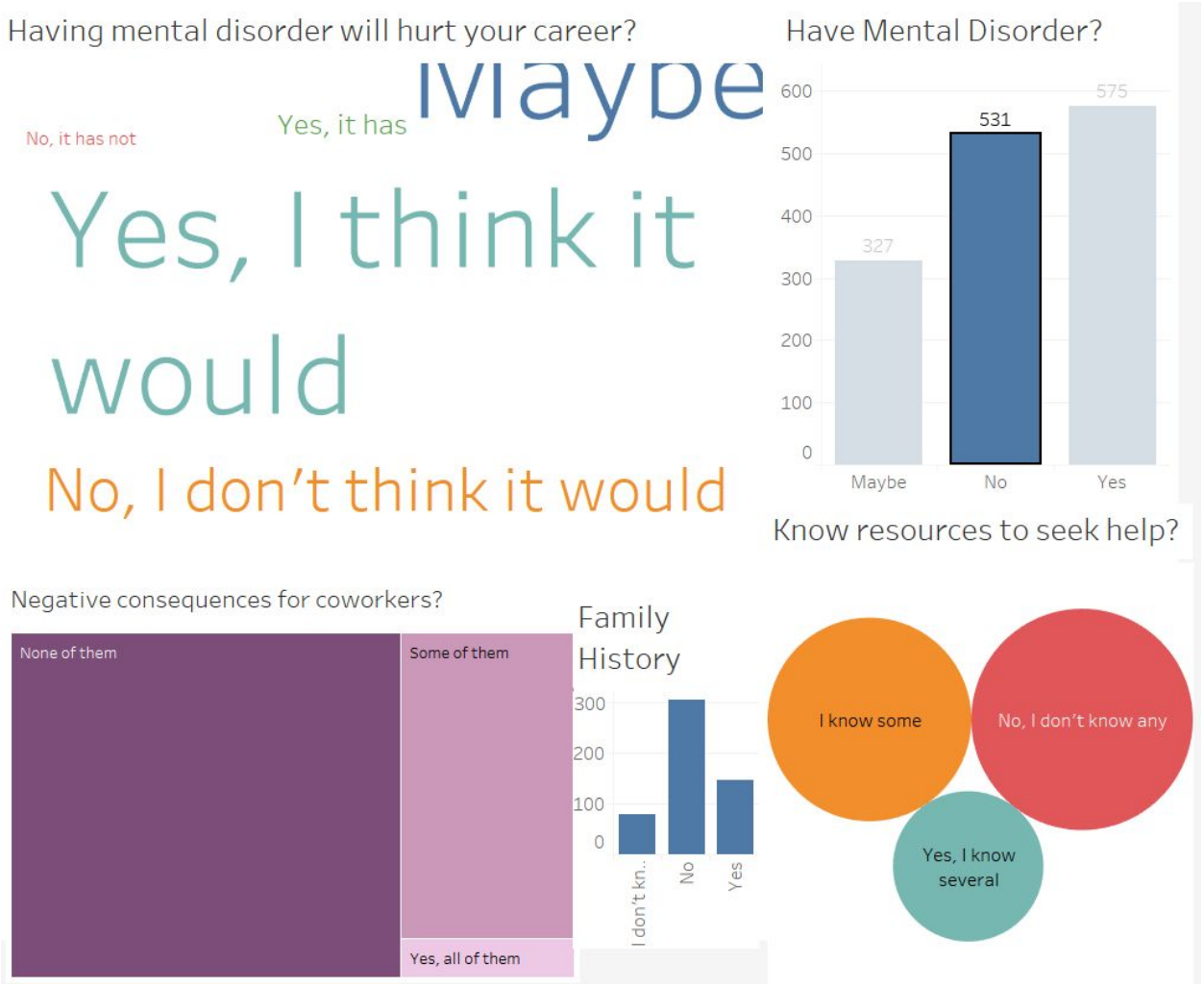
1. If the answer for any question “Do you currently have mental disorder?” Is Yes we have the following visualization.



Observations:

- 1. In people having mental disorder, very high percentage of people have family history of mental disorder
- 2. They know some or several resources to seek help, this is a good sign
- 3. They do not see negative consequences for coworkers having mental disorder
- 4. And they strongly believe that having mental disorder will negatively harm their career. This is a bad sign and needs to be addressed

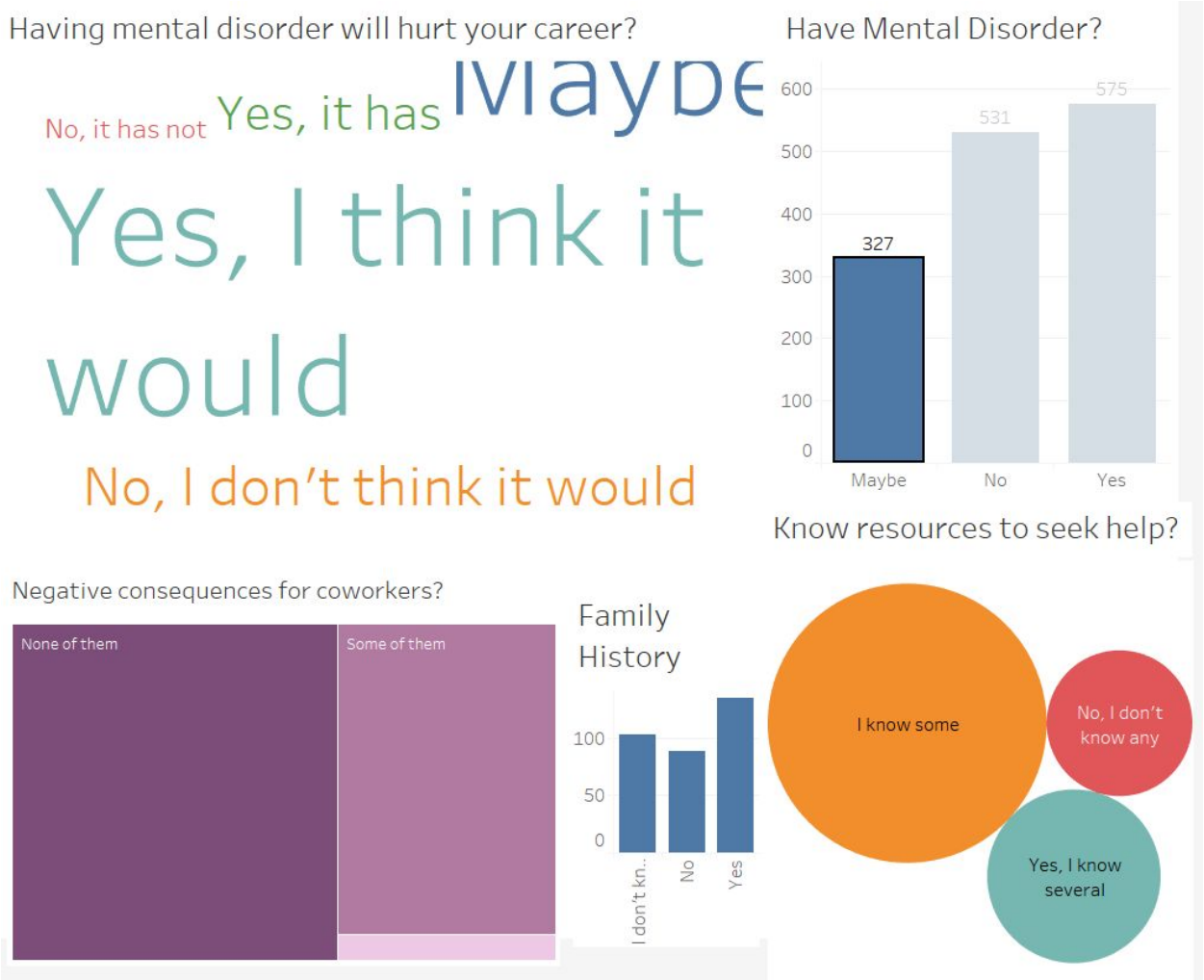
2. If the answer for the question “Do you currently have mental disorder?” Is No,



Observations:

- Most of the people who do not have mental disorder also don’t have a family history of the same. It seems the correlation between mental disorder and giving family for the same is very high
- This set of people also think that having mental disorder will affect their growth in career
- They don’t see negative consequences for coworkers having mental disorder. Here we see that people think their career will get affected negatively even though they don’t see the similar behavior in their coworkers
- A lot of people not having mental disorder also do not know about any resources to seek help

3. If answer for question “Do you currently have mental disorder?” Is “MayBe”



This is a very sensitive set of people who are unsure whether they have any mental disorder. A proactive step to get themselves diagnosed and taking professional help will certainly help avoiding problems and getting things corrected very early.

Observations:

- We see a mixed response for question family history of mental disorder here. A large number of people responded that they have a family history of the same. As seen in earlier two observations since this factor has high correlation with family disorder, these set of people have high chances of having mental disorder
- It is good that a large number of people in this et have some or several knowledge about resources to seek help
- This set of people also see that their coworkers do not have negative consequences of mental disorder. This observations is same for all three responses f people having mental disorder
- People who are unsure about their mental disorder also feel that having mental disorder will hurt their career growth. This could also be a reason why these people do not want to disclose their mental disorder condition.

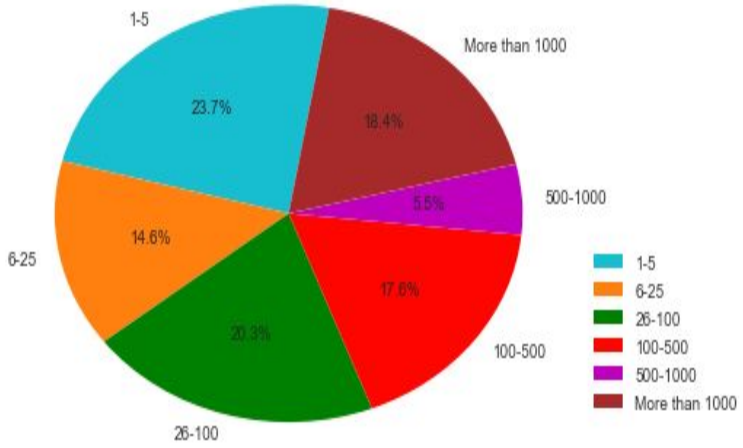
Discussion and Conclusion:

Python libraries used:
Matplotlib - used for basic working with dataframes and generating visualization
Seaborn - used for generating high quality and different variations of visualizations
Numpy, pandas - to work with dataset as a dataframe.

Along with this some of the geographic maps based visualizations were created using Tableau.

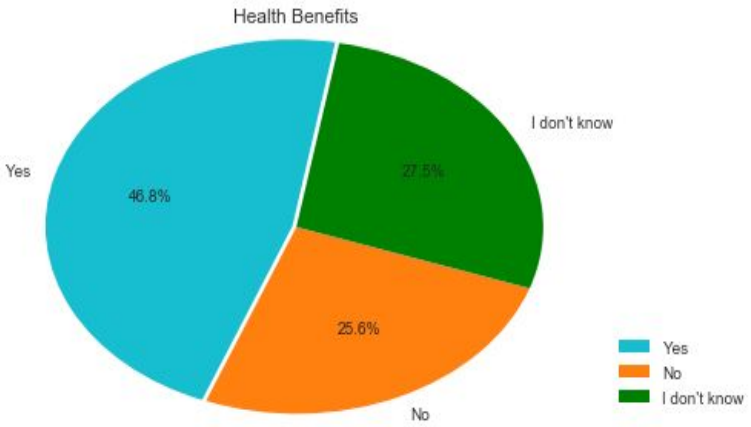
Results and Insights:

We started by looking into the various different questions we had listed.
To start the analysis we completed a breakdown of the different sized businesses that the respondents were a part of. Below is a pie chart showing the breakdown by number of employees in the company. We can see that there is quite a dispersion of employees with an almost equal distribution of groupings. With the exception of the group of 500-1000. More than 75% of all respondents were from companies with less than 500 employees.



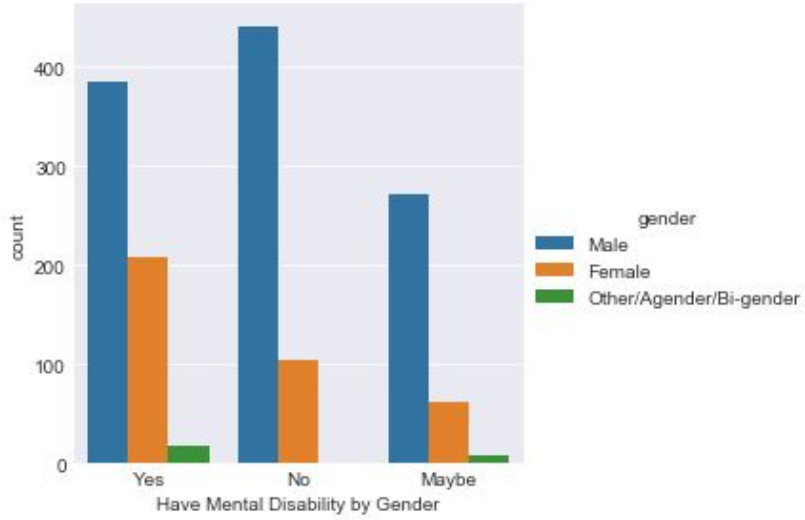
What percentage of employees have mental health coverage? How many have mental health disorder?

The results we found were that there are a lot of people in the industry who either don't have or are not aware of having mental health coverage. We see that only about 46.8% percent of the respondents say they have mental health coverage while another 27.5% do not know if they have the coverage. See figure.

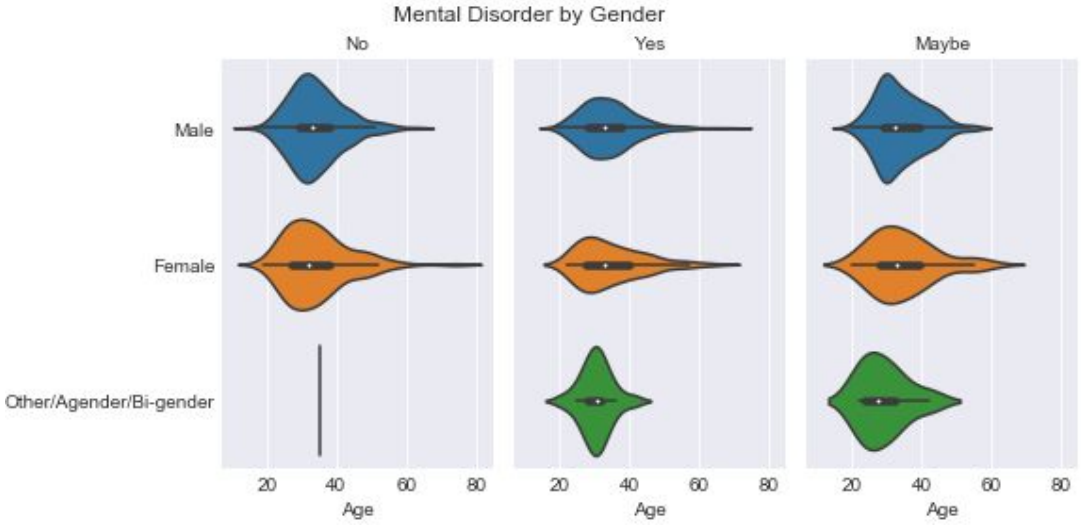


We also found that 79% of the respondents say they either have a mental health disorder or might have one. Utilizing

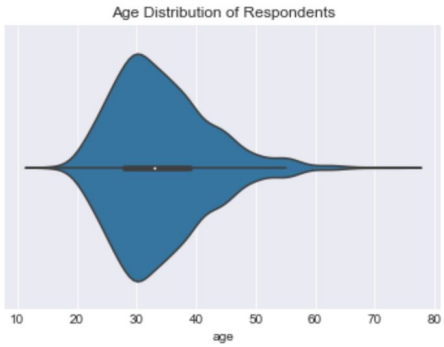
another bar chart we found that the breakdown by gender showed that we have a larger number of men with a current mental health disorder compared to women. However, the breakdown of male and female respondents seems to show that this is pretty much relative to the proportions by gender.



One interesting thing we see is that the females seem to be more confident of their diagnosis as you can see there are very few “maybes” listed. By contrast, you see that a much larger number (and proportion) of the male’s identified themselves as “maybe” having a mental health disorder. It may not good to generalize, but this could potentially be an effect of men not going to the doctor for help when they are having a problem. I think this would be an interesting aspect to explore further.



To get some further information on gender/age distributions, we developed some violin plots to help us visualize the breakdown of the age distribution of mental health disorders as well as a breakdown by gender.



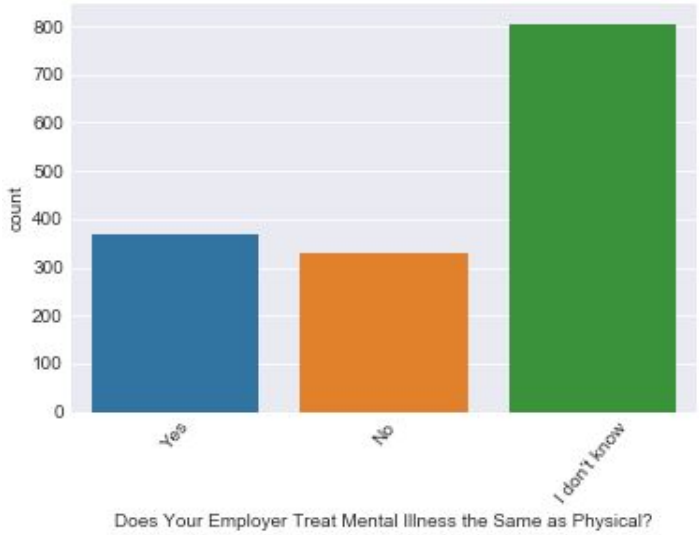
With the violin plots we can see that the most common age for mental disorders is right around age 30. However, each plot matches the general distribution (second plot below) for the ages of all respondents. There doesn't appear to be an age more likely of having a mental health disorder. We also see a very similar distribution between each gender.

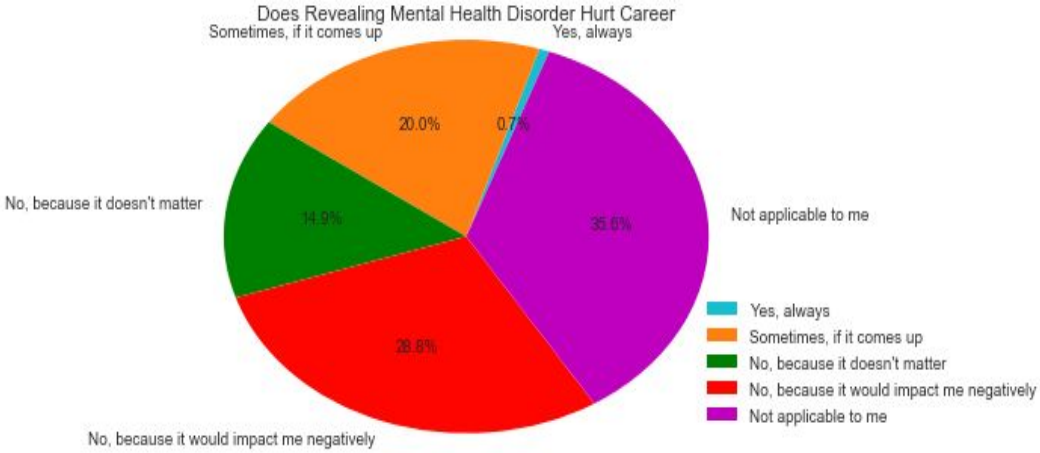
What is the attitude towards mental health?

The willingness of respondents to share mental health disorder among co-workers was explored by looking at a couple of different visualizations to help us get a good understanding of this information.

One of the questions asked of the respondents was related to fear of exposing their disorder. We found those who were afraid reveal their mental health disorder to a client because they were concerned it would negatively impact them, was roughly 29% (red slice - “No, because it would impact me negatively”). In addition to that, another 15% did not think it was important to share. Combined, approximately 44% would not reveal their mental health

disorder to their clients.



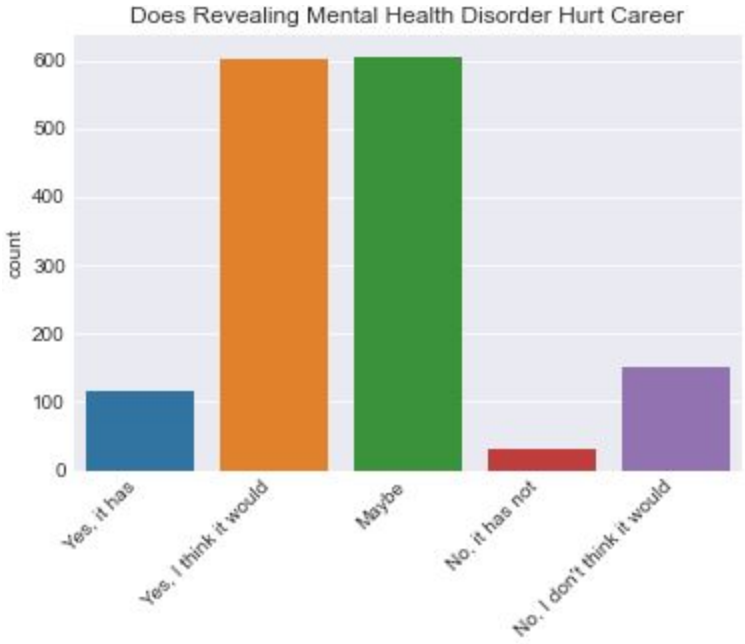


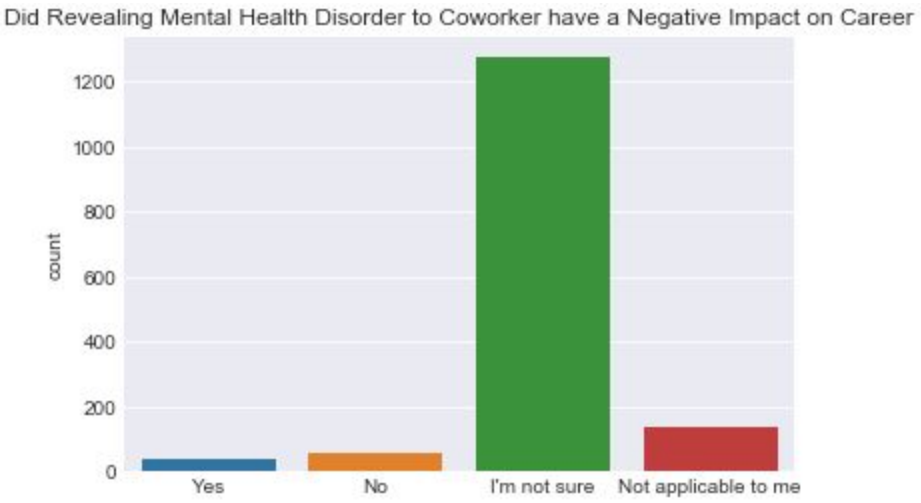
We also took a look at the breakdown of respondents who believed their employer treats mental health the same as physical health. We found that it is about 50/50 saying yes and no. We did not include the "I don't know" answers in that count. There were approximately the 50% of total respondents that answered “I don’t know”.

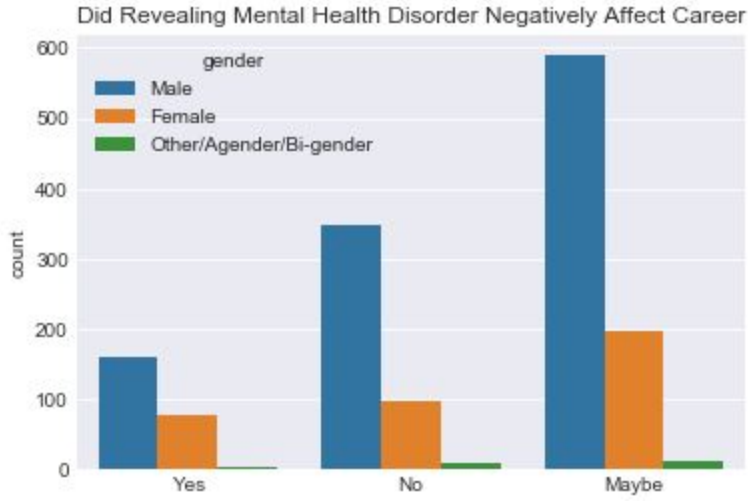
Another aspect of this career related question, was looking at a breakdown of beliefs that being identified as having a mental health disorder would hurt the respondents career. The results from the survey identified that many people believed that, if they were identified as having a mental health disorder, that would hurt their career. In fact, combined, those saying “yes, it has”, “yes, I think it would”, and “maybe”, the total makes up approximately 85% of the total. This clearly identifies a stigma in the industry.

The good news, however, is that the analysis of individuals who have previously revealed their disorder, very few identified that as conclusively affecting their career. Most were unsure if it had any effect at all.

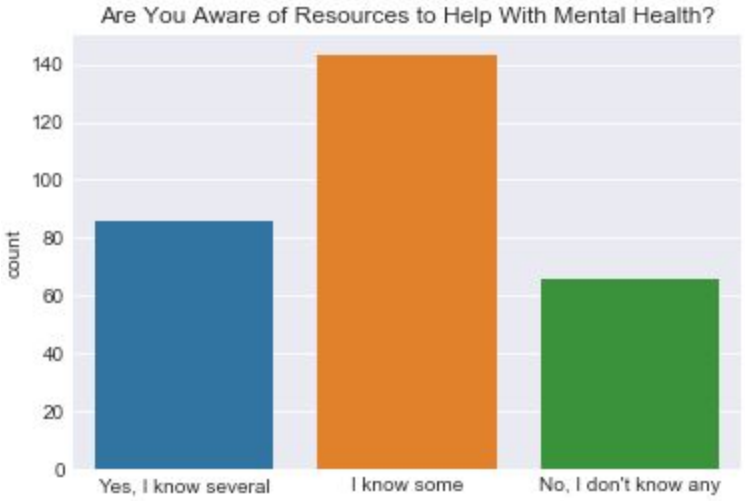
Further, we took a look at this information to determine if there was a difference in this perception by gender. We found that there is approximately equal percentages of respondents who felt that it would or would not impact them. This was very similar across all genders.







Another aspect of mental health in the workplace that we considered was the knowledge of the respondents to resources for help. We looked at the number of respondents who were aware of the mental health resources available to them. We found that a majority of the respondents were at least aware of some resources available. However, there is still a large number that did not know their resources. This should be addressed by the industry.

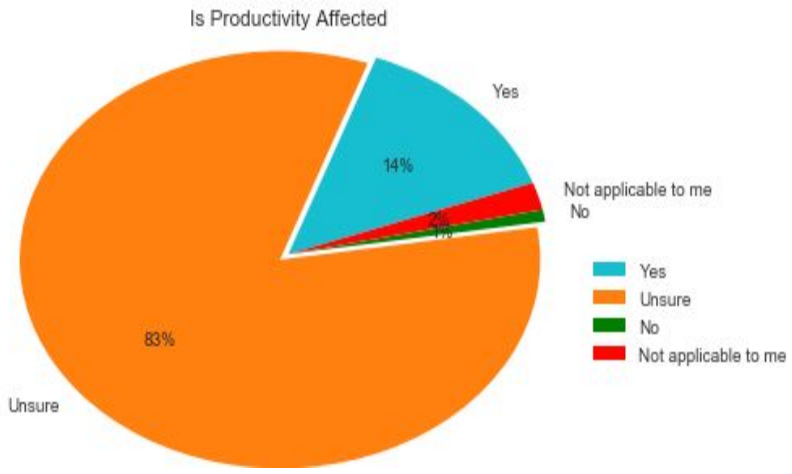


Looking more into this by gender, we see that, compared to men, many

females do not have information about help for mental health disorders. Again, this should be addressed to make the resources more readily accessible.

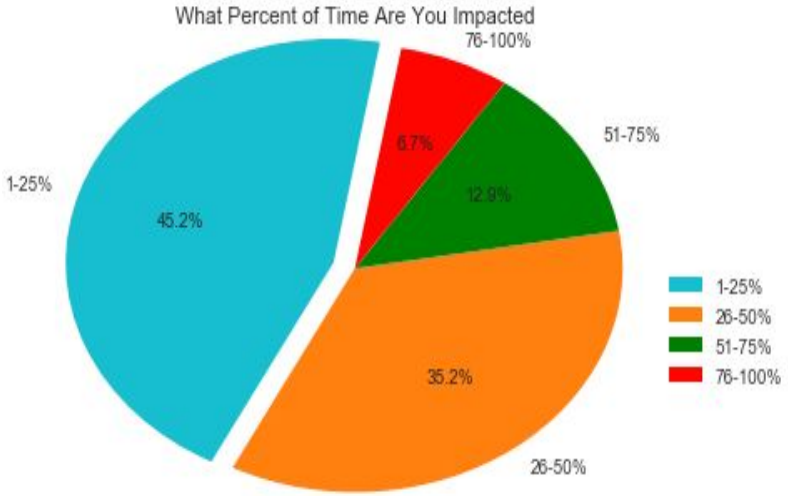
Is productivity affected? That was the next question we evaluated.

We found in our analysis that 83% of the respondents were unsure if their productivity was affected by their mental illness. This seems to be a concern, because it is hard to understand the need for help if you don't understand the effect it is taking on you.



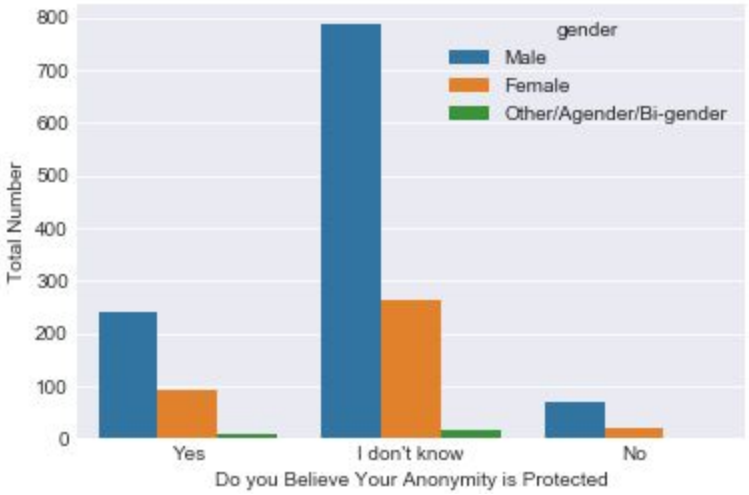
We wanted to explore this further. So we took a look at those who responded that their productivity was affected by their disorder. We found the following

- 45% think that it affects 1-25% of their working time.
- 35.2% think that it affects 26-50% of their working time.
- 12.9% think that it affects 51-75% of their working time.
- 6.7% think that it affects 76-100% of their working time.



Was anonymity protected if employee disclosed mental problem?

We explored this question and discovered that one of the reasons employee don't disclose their mental problems is due to lack of anonymity. This survey suggests that over 50% people are not sure if their anonymity will be protected. This can instill fear in the employees and prevent them from getting the help they need. We found that this response was similar across genders.



This should be addressed by having proper guidelines to deal with such issues organization wide.

Discussion Conclusion:

We discovered some interesting information in this research and found some items that should continue to be investigated in the future. One item of interest in particular is the difference in the Male and Female responses to the question of whether they had a mental health disorder. We made a possible connection to the fact that Male's more often identified themselves as "maybe" having a mental health disorder than Females. The connection could be the general lack of male acceptance in going to the doctor for help when they are having a problem.

We also found many things that need to be addressed by the industry. Namely they need to develop better culture to eliminate the stigma of mental health. To do this they should provide better resources to individuals and create a safe atmosphere where anonymity can be protected.

References:

osmihelp.org/

cdc.gov/mentalhealth/data-stats.htm

data.world/kittybot/osmi-mental-health-tech-2016

data.world/dataforacause/osmi-mental-health-clean

github.com/osmihelp

twitter.com/osmihelp

blog.nycdatascience.com/student-works/r-shiny/interesting-study-exploring-mental-health-conditions-tech-workplace/

osmi.typeform.com/report/Ao6BTw/U76z

medium.com/towards-data-science/data-and-mental-health-the-osmi-survey-2016-39a3d308ac2f

mentalhealthamerica.net/issues/2016-state-mental-health-america-report-overview-historical-data
seaborn.pydata.org/

http://www.acas.org.uk/index.aspx?articleid=1900

https://www.indiegogo.com/projects/open-sourcing-mental-illness-stronger-than-fear#/

https://www.nami.org/NAMI/media/NAMI-Media/Infographics/GeneralMHFacts.pdf

https://www.safeworkaustralia.gov.au/sites/swa/files/infographics-mental-disorders_0.jpg

https://www.tableau.com/