Wellness in the Workplace:

An Exploration of Mental Health in US Tech Workers

Data Science For All / Women 2021/09 - 2021/10

Team 18

Carey Huh, Huiwen Goy, Elizabeth Ortega, Xiaoer Hu,

Min Haeng Cho, Ofure Ebhomielen & Saphonia Foster

1. Introduction

Problem Overview

As technology evolves, the tech industry and the number of people working within it are growing. Mental health issues are common and have an impact on our personal and professional lives, but we don't know much about the mental health issues that specifically affect people in the tech industry. Mental health problems can be costly, as they can impact quality of life and work performance, and incur costs for employees and employers alike. We want to better understand these issues so that we can help improve support and services. For example, we want to understand how mental health issues and openness around their discussion have evolved recently, particularly with the start of the pandemic. We also aim to discover the services and solutions that are currently used by tech companies and evaluate how well they work, to improve awareness and support for mental health well-being in the tech industry.

According to Open Sourcing Mental Illness (OSMI), almost 50% of professionals within the tech industry who took the OSMI survey reported that they have a mental health disorder. This is striking compared to the 20.6% of adults in the US who experienced mental illness in 2019. We were surprised by this but also mindful of the fact that the OSMI dataset may be highly biased because of the opt-in nature of the surveys. These issues motivated us to conduct an in-depth evaluation of the OSMI dataset. This includes an exploration of the prevalence of disorders, the culture around discussing mental health, and resources provided within the tech industry. In addition, the surprising finding stimulated team discussions on how best to interpret and report on findings from data with inherent biases such as survey data.

There are several significant events that could also impact mental health care accessibility and awareness of mental health issues, such as legislation changes around health care and the beginning of the Covid-19 pandemic. We will consider these events when interpreting trends across 2014-2020.

Problem Impact

As professionals working in or transitioning into data science roles, we are especially interested in the work culture of companies in the tech space. It is important for us to know what the culture is actually like in the tech industry around important topics like mental health. Mental health is increasingly brought up in public discourse and stigma around it appears to be decreasing. We care whether tech companies are creating environments where employees feel safe and open to discuss their mental health issues. Tech companies are known for their progressive nature and taking the lead on advancing benefits for their employees like allowing for remote work and mental health days off. Mental health benefits are often advertised on job postings and company websites to show that they prioritize the mental health of their employees. With this project, we can dig into the actual results of these efforts by seeing what tech employees have experienced in their own lives in terms of mental health issues and what they perceive as their company culture around mental health topics.

Our project can also help inform tech companies on this issue. For example, we want to know how tech workers are affected by problems with mental health and how these issues can directly impact their productivity, and which company-provided services are helpful. Healthcare costs are often the largest expense for a company after salaries, and hiring new employees is expensive. Increasing productivity, decreasing turnover and attracting the best talent through a greater focus on supporting employee mental health can greatly impact the bottom line and help companies advance towards their goals.

This discussion is particularly timely, considering current events (e.g., Facebook whistleblower testimony, Netflix employees walk-out) that showcase instances of tech employees speaking out and taking action in opposition to their companies' stance and culture around certain issues like public safety, hate speech and diversity/inclusion. Looking into the future, we anticipate that tech companies may be expected to take the lead in helping to solve the complex but increasingly important issues of mental health in the workplace.

Key Questions

Key Q1. What does the mental health landscape look like in the US tech sector?

- What is the prevalence and what are the most common types of disorders in US tech workers? How are these changing over time?
- Are certain groups more susceptible to problems with mental health?
- Is the prevalence of mental health disorder higher in the tech sector?

Key Q2. What lessons on mental health care service delivery and workplace culture can we glean from tech sector data?

- Which factors are associated with greater openness about mental health issues in the workplace?
- For tech workers who are experiencing mental health problems, which types of workplace support are associated with greater productivity?
- How important is effective treatment of mental health disorders to tech workers' productivity?

In Part C of this report (Statistical Analysis & Machine Learning), sections have been color-coded according to which key question is being addressed (orange: Q1, green: Q2).

2. Data Analysis & Computation

A. Datasets, Data Wrangling & Cleaning

I. OSMI survey

The main data are responses to an annual mental health survey conducted by Open Sourcing Mental Illness (OSMI), a US-based non-profit run by mental health advocates who are in or connected to the tech industry. Survey questions are related to employment type, healthcare coverage, workplace support for mental health, personal history of mental health, effects of mental health on productivity, and workplace culture surrounding mental health.

Respondents were recruited by OSMI via Twitter and Facebook from various web development communities among others, and from conferences that OSMI members spoke at. In total, 4398

anonymous responses were collected from 2014 to 2020. The table below shows the number of survey questions by year and number of respondents. Note that no data was available for 2015.

Year	Number of survey questions	Number of respondents
2014	27	1260
2016	63	1433
2017	123	756
2018	123	417
2019	82	352
2020	120	180

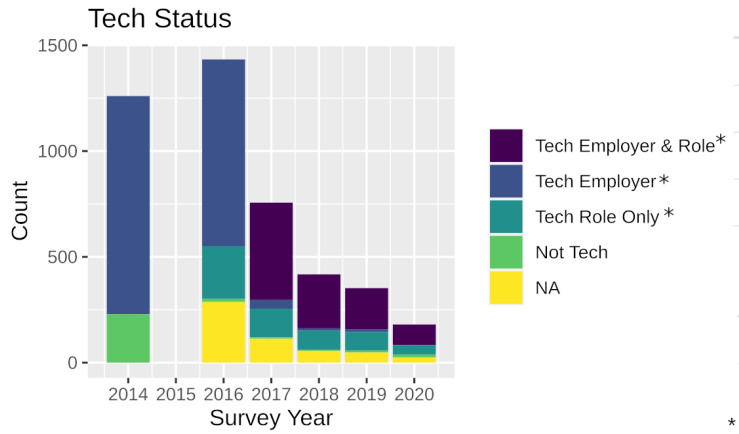
Merging and cleaning OSMI data

Survey data were downloaded as .csv files from the OSMI website (one file per year). In the data files, survey questions were listed as column names, with each respondent as one row, thus the data was in wide format. Not all survey questions existed across years. Also, some survey questions were identical between years but had different categories of responses (e.g. Yes/No/Possibly in some years and Yes/No/Maybe in others). We divided the ~120 variables among 6 team members and produced a Google Colab jupyter notebook that aggregated and cleaned the data, using primarily pandas and numpy libraries in Python.

Data cleaning & final output:

- Survey questions (variable names) were made consistent across the years (if discrepancies were found, 2020 version's wording was used)
- Response categories were made consistent on similar questions
- "Yes" / "No" binary answers were coded as numeric "1" / "0" in order to model data
- Responses with more than 2 categories (e.g., "Yes" / "Maybe" / "No" / "Don't know") were left as they were but their order was considered during analysis
- 'Gender' was a free-response item; 180 different responses were re-coded to 3 (female, male, non-binary) and missing values
- 'Age' was a free-response item; values <18 (including negative values) were re-coded to 18, as we assumed that respondents are working adults as per the target of the survey, while values >74 (e.g., 999) were clear outliers based on the histogram and were treated as missing values
- Any question that was missing in a particular year was filled with missing values
- Survey questions that had no data across all years were dropped from the final dataset
- The final OSMI dataset was created by stacking datasets from different years, with one respondent per row, survey questions as columns, and an added column for the year of survey, resulting in a dataframe with 4398 rows x 105 columns
- Final dataset was exported as a .csv file, so that anyone can explore, visualize and run models on the data using any platform, including Python, R or Tableau

Inclusion criteria



Tech Status	# of Responses
Tech Employer & Role*	1005
Tech Employer*	1980
Tech Role Only*	607
Not Tech	277
NA	529

Tech Status	# of Responses
Tech*	3592
Not Tech	806

* Respondents that were considered a "tech worker"

For most analyses, we considered data from respondents that met both of the following criteria:

1. The respondent was a “tech worker”:
    - Respondents had to answer “Yes” to at least one of the questions: “Is your employer primarily a tech company/organization?” and “Is your primary role within your company related to tech/IT?”. The rest were considered to be “non-tech” (answered “no” to both or did not answer).
2. The respondent lived in the US:
    - We reasoned that healthcare services depend on where people live. In addition, healthcare is tied to employment in the US, but not in some other countries.

II. BRFSS survey

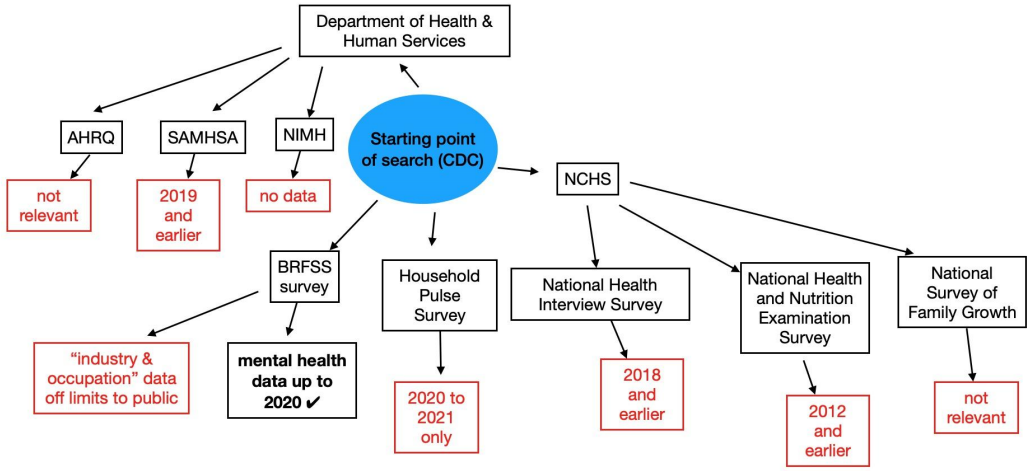
Finding representative data on mental health trends in the US

Besides analyzing data from the opt-in OSMI survey, we also wanted to find data on mental health that was more representative of the US population than the OSMI dataset. We decided that the dataset should meet three criteria:

1. There should be demographic variables on age, sex, race, education, income, and employment (ideally, type of occupation);
2. There should be mental health-related variables such as frequency of anxiety and depression;
3. The data should cover at least 2016 through 2020, to align with the timing of OSMI data collection.

We searched for relevant datasets in public repositories hosted by the Centers for Disease Control and Prevention (CDC), National Center for Health Statistics (NCHS), National Institute of Mental Health (NIMH), Substance Abuse and Mental Health Services Administration (SAMHSA), and the Agency for Healthcare Research and Quality (AHRQ). The figure below

summarizes the process of arriving at the BRFSS dataset as the best match for our criteria, with red boxes denoting “dead ends” due to datasets not meeting one of the criteria or not being publicly available.



About the BRFSS survey

The Behavioral Risk Factor Surveillance System (BRFSS) survey is an annual survey that uses telephone/cellphone sampling to collect data from non-institutionalized adults in the US. Besides demographic questions, the BRFSS core questionnaire covers everyday behaviors related to health outcomes, including exercise, sleep, smoking and alcohol consumption, with optional modules that cover more specific topics such as diabetes. An optional module on industry and occupation was used by 24 states in 2020 and in earlier years, but these data are considered sensitive and are not publicly available. Due to this, we could not consider whether BRFSS survey respondents worked in the tech industry.

Compiling data from the BRFSS survey

Datafiles for each BRFSS year from 2016 to 2020 were downloaded from the website in SAS xport format. Files were read into R using the function sasxport.get() from the Hmisc library. A total of 2,193,981 records were available from the five years. Respondents within each year were assigned individual weights, which made the sample more representative of the entire population by taking into account different probabilities of being selected (e.g. due to area code), and taking into account the respondent’s age group, gender and ethnicity. To analyze multiple years’ datasets as one combined dataset, individual weights had to be further adjusted by multiplying the original weight by the proportion of that year’s data in the combined dataset. For instance, 2020 had 401,958 respondents, which was 0.1832 of the total sample of 2,193,981, so 2020's individual weights were adjusted by [original weight * 0.1832]. To make the dataset a more manageable size, 16 variables of interest were extracted from ~300 available variables, covering age, sex, race, education, geographic location (state), physical and

mental health, employment status, income, and healthcare coverage, plus 5 additional survey-related variables (individual weights, strata, primary sampling unit, date of interview, and complete/partial interview status).

BRFSS versus OSMI variables

The question of interest in the BRFSS core questionnaire was: "Thinking about your mental health, which includes stress, depression, and problems with emotions, for how many days during the past 30 days was your mental health not good?" Respondents gave a number from 1 to 30, or reported “none”, or “refused”/“don’t know”. A calculated survey variable was available, collapsing responses into None, 1-13 days, 14+ days, and refused/don’t know. The OSMI survey did not contain a similar question, but asked many other questions on mental health (see exploratory data analysis).

	OSMI survey	BRFSS (population health)
Age	Numeric	Categories from 18-24 to 80+
Sex	Male / Female / Non-binary	Male / Female
Race	8 categories	8 categories
Employment	Tech/non-tech employer Tech/non-tech role	Employment status only
Country	61% US	US residents only

B. Exploratory Data Analysis

The table below shows some characteristics of the respondents in each survey.

	OSMI survey	BRFSS (population health)
Sample size	4,398	2,193,981
Age	Mean = 33.8, SD = 8.2	30% 18-34, 33% 35-54, 38% 55+
Sex	Male 73%, Female 25%, Non-binary 1.7%	Male 48.7%, Female 51.3%
Race	86% white	61% white
Country	61% US	100% US

The responses in the OSMI survey were almost all categorical (e.g. Yes/Maybe/No/Don’t know), except for age, which we considered a continuous variable. There were also a few free-response text items (e.g. “Describe a time when…”).

We decided to categorize the questions in the OSMI survey to form an overview of the topics covered by this survey. The table below shows a sample of key questions by category, and the proportion of missing data per variable, by year.

		Proportion of Missing Data					
Category	**OSMI question**	**2014**	**2016**	**2017**	**2018**	**2019**	**2020**
Demographic	What is your age?	0.16	0.14	0.26	0.00	0.00	0.00
	What is your gender?	0.24	0.63	2.38	1.20	3.13	1.11
	What country do you live in?	0.00	0.00	0.26	0.00	0.00	0.00
	What is your race?	100.00	100.00	34.26	25.42	42.05	63.89
Employment	How many employees does your organization have?	0.00	20.03	14.95	13.43	13.64	13.89
	Is your employer primarily a tech organization?	0.00	20.03	14.95	13.43	13.64	13.89
	Is your primary role related to tech/IT?	100.00	81.65	14.95	13.43	13.64	13.89
Disorder	Do you currently have a mental health disorder?	100.00	0.00	0.00	0.00	0.00	0.00
	Have you ever been diagnosed with a mental health disorder?	100.00	0.00	57.14	54.20	58.24	71.67
	Have you ever sought treatment for a mental health disorder from a mental health professional?	100.00	0.00	0.00	0.00	0.00	0.00
	Do you have a family history of mental illness?	0.00	0.00	0.00	0.00	0.00	0.00
Productivity	If you have a mental health disorder, how often do you feel that it interferes with your work when being treated effectively?	100.00	100.00	0.00	0.00	0.00	0.00
	If you have a mental health disorder, how often do you feel that it interferes with your work when NOT being treated effectively (i.e., when you are experiencing symptoms)?	100.00	100.00	0.00	0.00	0.00	0.00
	Do you believe your productivity is ever affected by a mental health issue?	100.00	79.97	85.05	86.57	86.36	86.11
	If yes, what percentage of your work time (time performing primary or secondary job functions) is affected by a mental health issue?	100.00	85.76	88.76	90.17	89.49	88.89
Healthcare	Do you know the options for mental health care available under your employer-provided health coverage?	0.00	29.31	23.81	22.30	21.02	26.11
	Do you know local or online resources to seek help for a mental health issue?	100.00	79.97	85.05	86.57	86.36	86.11
	Does your employer provide mental health benefits as part of healthcare coverage?	0.00	20.03	14.95	13.43	13.64	13.89
	Does your employer offer resources to learn more about mental health disorders and options for seeking help?	100.00	20.03	14.95	13.43	13.64	13.89
	Do you have medical coverage that includes treatment of mental health disorders?	100.00	79.97	85.05	86.57	86.36	86.11

Openness & Culture	Has your employer ever formally discussed mental health (for example, as part of a wellness campaign or other official communication)?	100.00	20.03	14.95	13.43	13.64	13.89
	Is your anonymity protected if you choose to take advantage of mental health or substance abuse treatment resources provided by your employer?	100.00	20.03	14.95	13.43	13.64	13.89
	If a mental health issue prompted you to request a medical leave from work, how easy or difficult would it be to ask for that leave?	100.00	20.03	14.95	13.43	13.64	13.89
	Would you feel more comfortable talking to your coworkers about your physical health or your mental health?	100.00	100.00	14.95	13.43	13.64	13.89
	Would you feel comfortable discussing a mental health issue with your direct supervisor(s)?	100.00	100.00	14.95	13.43	13.64	13.89
	Have you ever discussed your mental health with your employer?	100.00	100.00	14.95	13.43	13.64	13.89
	Would you feel comfortable discussing a mental health issue with your coworkers?	100.00	100.00	14.95	13.43	13.64	13.89
	Have you ever discussed your mental health with coworkers?	100.00	100.00	14.95	13.67	14.49	13.89
	Overall, how much importance does your employer place on physical health?	100.00	100.00	14.95	13.43	13.64	13.89
	Overall, how much importance does your employer place on mental health?	100.00	100.00	14.95	13.43	13.64	13.89
	If you have been diagnosed or treated for a mental health disorder, do you ever reveal this to coworkers or employees?	100.00	79.97	85.05	86.57	86.36	86.11
	If you have revealed a mental health disorder to a coworker or employee, how has this impacted you or the relationship?	100.00	79.97	85.05	86.57	86.36	86.11
	Have your observations of how another individual who discussed a mental health issue made you less likely to reveal a mental health issue yourself in your current workplace?	100.00	54.15	21.16	23.26	19.89	20.00
	How willing would you be to share with friends and family that you have a mental illness?	100.00	0.00	0.00	0.00	0.00	0.00
	Are you openly identified at work as a person with a mental health issue?	100.00	100.00	0.26	0.00	0.00	0.00
	Has being identified as a person with a mental health issue affected your career?	100.00	100.00	88.89	87.77	87.22	86.11
	If they knew you suffered from a mental health disorder, how do you think that your team members/co-workers would react?	100.00	100.00	0.26	0.00	0.00	0.00
	Have you observed or experienced an unsupportive or badly handled response to a mental health issue in your current or previous workplace?	100.00	6.21	0.26	0.00	0.00	0.00

	Have you observed or experienced a supportive or well handled response to a mental health issue in your current or previous workplace?	100.00	100.00	0.26	0.00	0.00	0.00
	Overall, how well do you think the tech industry supports employees with mental health issues?	100.00	100.00	0.26	0.00	0.00	0.00

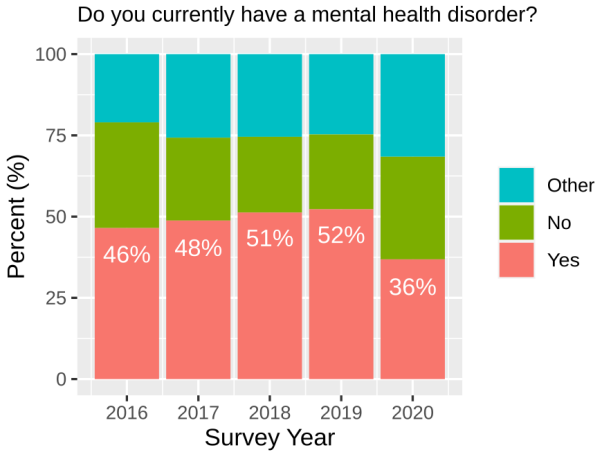
From the above table, it is clear that there are many questions where data are completely missing (red highlighted cells) in 2014 and 2016. This is due to many questions being added to the survey in 2017-2020; there were 27 questions in 2014 and 120 questions in 2020. However, there were more respondents participating in earlier surveys compared to later (1260 in 2014; 180 in 2020). Thus, we decided to include data from all 6 years in our analysis.

C. Statistical Analysis & Machine Learning

Key Q1. What does the mental health landscape look like in the US tech sector?

What is the prevalence and what are the most common types of disorders in US tech workers? How are these changing over time?

The overall prevalence of having a "current mental health disorder" was 48% among US tech workers. The remaining respondents said "No", "Maybe" or "Don't know", with the latter two recategorized as "Other" in the figure below. There was a mild trend of prevalence increasing over time, except for 2020 (but note that 2020 had relatively few responses).



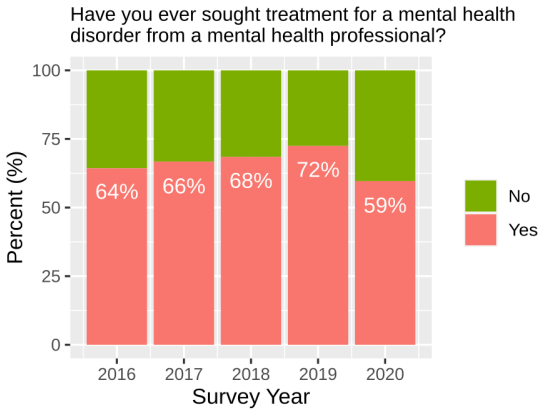
Year	# of Yes Responses	Total # of Responses	Prevalence (%)
2016	323	695	46
2017	220	451	48
2018	143	279	51
2019	93	178	52
2020	21	57	36

# of Yes Responses	Total # of Responses	Prevalence (%)
800	1660	48

Three questions dealt with past history of mental health disorders: "Have you ever sought treatment for a mental health disorder from a mental health professional?", "Have you had a mental health disorder in the past?" and "Do you have a family history of mental illness?". Percentages of "Yes" were 66%, 53%, 50%, respectively. Considering that current prevalence is

48%, these values of past history are similar to the current prevalence and provide internal consistency.

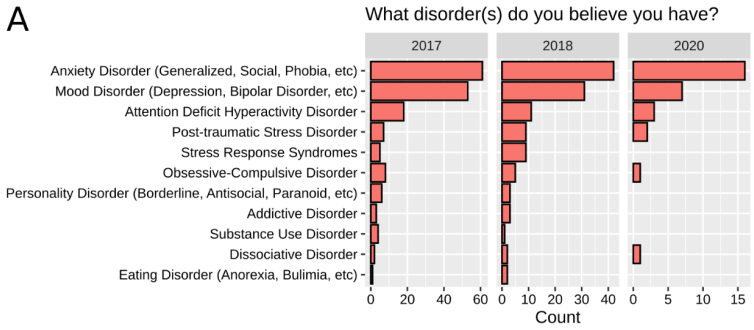
The answers to whether one ever sought treatment consisted of “Yes”, “No” only, which is different from the other prevalence questions. There again appears to be a trend that the percentage of people seeking treatment for mental health issues is growing with time. We again see that 2020 is different from the other years (possibly due to low sample size). Notably, the percentage of respondents having sought treatment is higher (66%) than mental health disorder prevalence values, suggesting that even those respondents that do not consider themselves to have had a mental health disorder sought treatment from a mental health professional at some point in their lives.



# of Yes Responses	Total # of Responses	Prevalence (%)
1102	1660	66

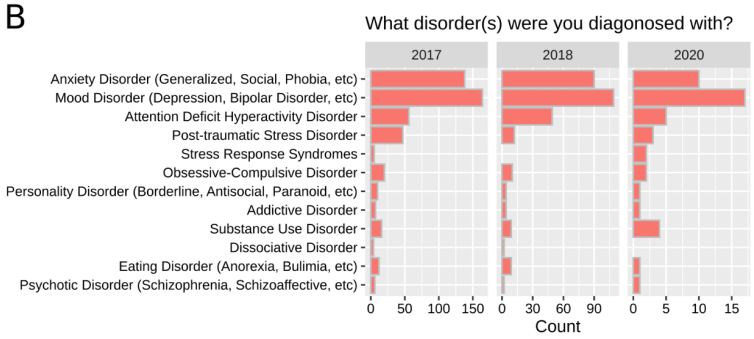
In 2017, 2018 and 2020, respondents were asked which mental health disorder(s) they believe they have (A) and which disorder(s) they were diagnosed with (B). For both questions, anxiety, mood, attention deficit hyperactivity and post-traumatic stress disorders were among the most common types reported by US tech workers. Substance use, eating and psychotic disorders were among the least common types.

A



Top 6 Mental Health Disorder Types	% of Responses
Anxiety Disorder (Generalized, Social, Phobia, etc)	37.66
Mood Disorder (Depression, Bipolar Disorder, etc)	28.80
Attention Deficit Hyperactivity Disorder	10.13
Post-traumatic Stress Disorder	5.70
Obsessive-Compulsive Disorder	4.43
Stress Response Syndromes	4.43

B



Top 6 Mental Health Disorder Types	% of Responses
Mood Disorder (Depression, Bipolar Disorder, etc)	34.86
Anxiety Disorder (Generalized, Social, Phobia, etc)	28.61
Attention Deficit Hyperactivity Disorder	13.22
Post-traumatic Stress Disorder	7.45
Obsessive-Compulsive Disorder	3.85
Substance Use Disorder	3.49

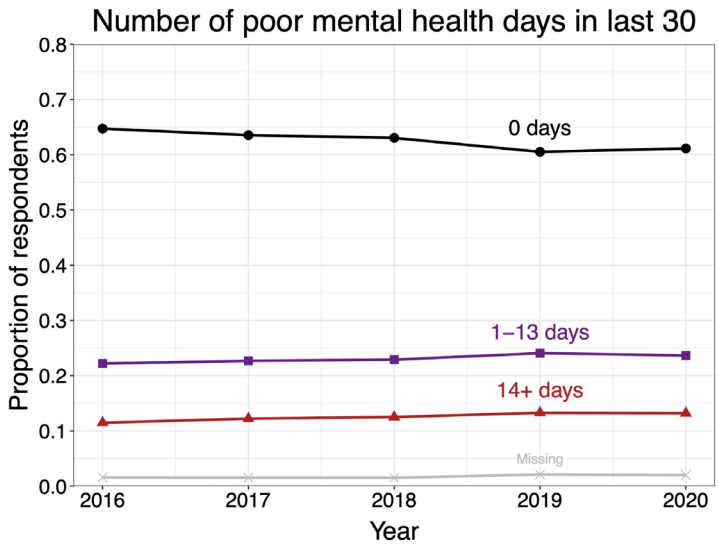
appears stable over this span of time. Note that the ordering of disorder types was kept consistent between A and B. Interestingly, anxiety disorder was the most common disorder that respondents believed that they had; however, mood disorder was the most commonly diagnosed disorder.

Are certain groups more susceptible to problems with mental health?

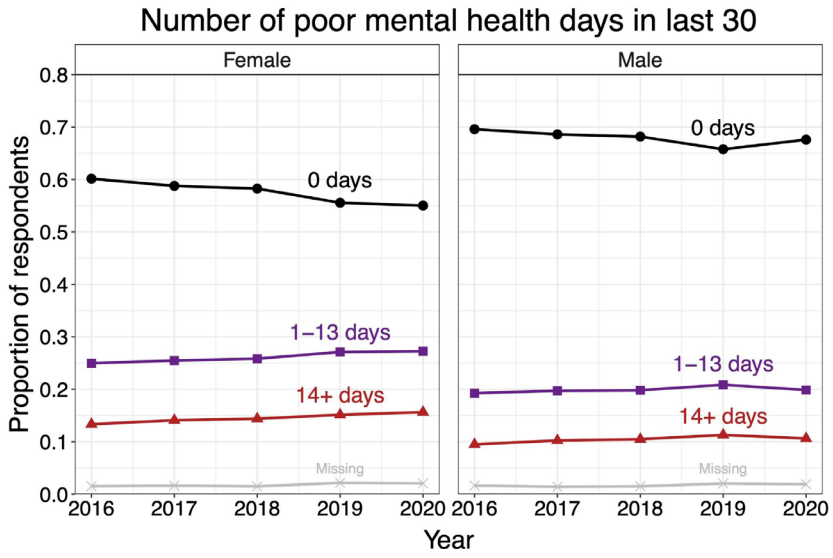
General US adult population (BRFSS survey data)

In response to the BRFSS question “Thinking about your mental health, which includes stress, depression, and problems with emotions, for how many days during the past 30 days was your mental health not good?”, respondents provided a number, which was re-coded in the official survey dataset to four levels: 0 days, 1 to 13 days, 14+ days, or Refused/Don’t know.

The proportions of respondents in each category (weighted by survey case weights) showed that the number of poor mental health days appeared fairly stable over time, perhaps with a slight trend of increasing poor mental health days.



Splitting the data by sex (see below), the analysis showed that females had more poor mental health days than males, and that females also had an increasing number of such days over time.



To examine which demographic variables affected mental health, a **logistic regression model** was conducted using the svyglm() function in the R package 'survey', which incorporates a survey design into a model, including case weights, strata and primary sampling units. The outcome measure of "poor mental health days" was recoded into a binary variable, with "1-13 days" and "14+ days" collapsed into "some" poor mental days, while the largest category of "0 days" was kept as "none". For easier interpretation, most predictors were recoded from multiple levels into the two levels shown in the table below. All rows with missing data were excluded.

Predictor	**Coded 0 (% of sample)**	**Coded 1 (% of sample)**
Age group	Older, 45+ years (53.5%)	Younger, 18-44 years (46.5%)
Sex	Male (49.7%)	Female (50.2%)
Race	White (63.3%)	Non-white (36.7%)
Education	Less than college (70.4%)	College (29.6%)
Household income	<$75,000 (64.4%)	$75,000+ (35.6%)
Employment status	Unemployed/ Unable to work (11.9%)	Employed/ student/ homemaker/ retired (88.1%)

The table below shows the coefficients for the model, with the coefficients exponentiated to odds ratios and 95% confidence intervals for odds ratios. The model showed being younger, female, white, unemployed, and having lower household income increased the risk of having poor mental health days, while the effect of formal education was relatively small.

Predictor (coded 1)	**Estimate**	**S.E.**	***t***	***p***	**Odds ratio**	**95% C.I. lower bound**	**95% C.I. higher bound**
Age (younger)	0.770	0.0073	105.7	<0.001	2.16	2.13	2.19
Sex (female)	0.519	0.0072	72.6	<0.001	1.68	1.66	1.70
Race (non-	-0.312	0.0085	-36.7	<0.001	0.73	0.72	0.74

white)							
Edu (college)	0.036	0.0076	4.80	<0.001	1.04	1.02	1.05
Income (higher)	-0.183	0.0082	-22.3	<0.001	0.83	0.82	0.85
Employed (yes)	-0.918	0.0115	-80.1	<0.001	0.40	0.39	0.41

Are certain groups more susceptible to problems with mental health?

US tech workers (OSMI survey data)

The goal of this analysis was to examine if demographic characteristics (age, gender, race) affected the chances of a respondent having a current mental health disorder. To this aim, we employed a **logistic regression model**. The analysis only included US tech workers who responded "Yes" or "No" to the question: "Do you currently have a mental health disorder?" (n=706; 96% of those that responded “Yes” had also been formally diagnosed with a mental health disorder). Rows were dropped if they had missing values for the predictors ‘age’, ‘gender’, and ‘race’. Gender had three categories (male, female, non-binary) and race had 8 categories; gender and race were dummy-coded. The respondents were predominantly male (60.9%, with 3.0% who were non-binary), and predominantly white (86.8%). The final variables used in the model were: age (continuous; standardized), gender_Female (1 or 0), race_White (1 or 0). The outcome measure was whether or not the respondent had a current mental health disorder (“Yes” = 1, “No” = 0).

Using the LogisticRegression function from scikit-learn (with class weights added, as 66% of cases were “Yes”), the resulting logistic regression model had 57.5% accuracy (55.7% sensitivity, 38.8% false positive rate; 73.9% precision; see confusion matrix below). The model’s AUC was 0.58. These metrics indicated that age, gender and race were not strong predictors of which respondents had a current mental health disorder.

	Predicted current mental health disorder (number of cases)	
Actual current disorder	**No**	**Yes**
No	**145**	92
Yes	208	**261**

The model’s coefficients were exponentiated to odds ratios (table below). Older respondents had a lower risk of having a current mental health disorder. Being female was associated with a 91% greater risk. While there is an apparent relationship between a respondent’s race and the risk of a mental health disorder, it should be kept in mind that nearly all respondents reported their race as “white”, which makes race a poor predictor in this model.

	Odds ratio
Age	0.768
Gender (Female)	1.91
Race (White)	2.29

Is the prevalence of mental health disorder higher in the tech sector?

Occupation data in the population survey were considered sensitive and were not publicly available. However, there was a mix of tech and non-tech workers in the OSMI survey. We examined the proportions of those who answered “Yes” to “Do you currently have a mental health disorder?”, split by their country of residence and whether they were a tech worker.

		Current mental health disorder?	
		Yes (count)	No, Maybe, Don’t know, NA (count)
Country	**Work**		
US	**Tech**	1471	800
	Non-tech	11	15
Other	**Tech**	1079	242
	Non-tech	16	6

The proportions suggested that the rate of mental health disorders was higher in US tech workers than in US non-tech workers, and that the difference between tech and non-tech was less dramatic outside the US. However, note that there were generally few non-tech respondents, so this apparent difference may simply be due to a highly biased sample.

	Prevalence (%)	
	Tech	**Non-tech**
US	64.8	42.3
Other countries	81.7	72.7

Key Q2. What lessons on mental health care service delivery and workplace culture can we glean from tech sector data?

Which factors are associated with greater openness about mental health issues in the workplace?

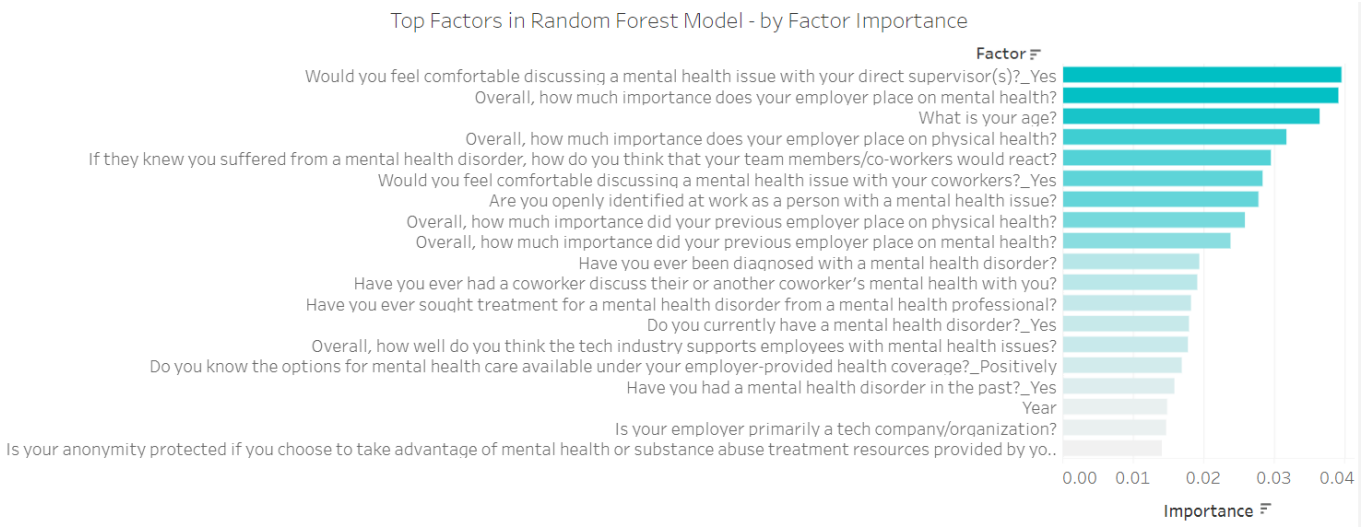
Respondents showed different levels of openness around issues surrounding mental health. In order to investigate what might explain these differences, we decided to create a predictive

model for the question "Have you ever discussed mental health with your employer?". This question was asked in survey years 2017-2020 and had the least amount of missing responses within the category of questions surrounding how comfortable respondents felt about speaking about mental health issues at work. The responses to this question were either Missing/Null, Yes or No.

In order to model this binary response, we first began with a simple decision tree to classify the response to this question. When considering the low accuracy scores obtained by the decision tree, we moved instead to using a **random forest algorithm** as our classifier. This non-parametric technique is flexible and is more accurate and robust than a single decision tree since it averages the responses of many decision trees. We tuned the random forest classifier optimizing it for several different factors like the number of trees used, using average accuracy as the target. In order to evaluate accuracy, we used a repeated stratified k-fold split of the data with 10 different splits and 3 repetitions instead of using a simple train/test split, since we had unequal samples of Yes/No respondents and a smaller dataset due to only 2017-2020 data being available. We also used an option of the random forest classifier to adjust the class weight for this unbalanced target issue.

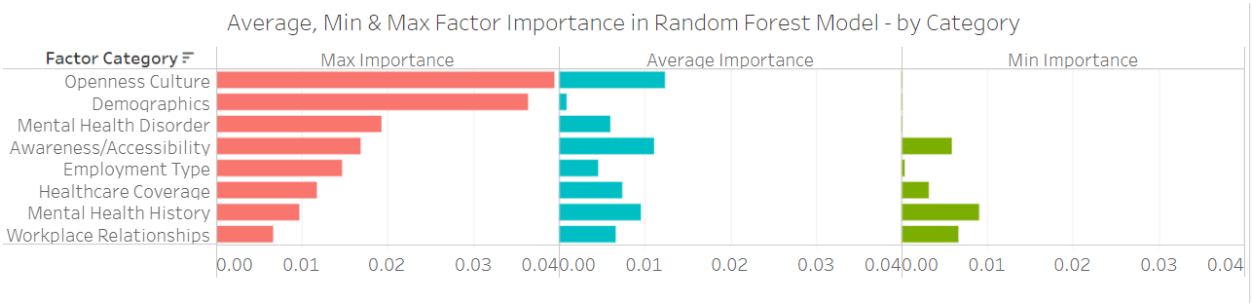
Overall, using the repeated stratified k-fold split to assess accuracy, our final model with 100 trees resulted in an average accuracy score of **82%** when predicting where a respondent had ever discussed mental health with their employer. We felt that this model was accurate enough that we could use the results to find some useful information for employers when trying to create a more open and inclusive environment surrounding mental health topics.

A useful feature of decision trees is that we can visualize the importance of each feature with an importance score that is comparable across the different factors in the model. Overall, the random forest classifier used our entire list of variables available - 289 when including all indicator variables. Below is a visual showing the most important factors by feature importance and how much of an effect each had on the random forest classifier algorithm.



Many of these top factors identified by the random forest classifier are factors within the control of an employer when setting the tone for an inclusive company culture. Factors such as how much importance does your employer place on mental health or physical health is controllable through company culture. Culture can also dictate important factors like whether respondents felt comfortable discussing mental health with their coworkers or supervisors. Companies can also ensure that another important factor is in place by making sure that employees are aware of the mental health options offered to them.

We gave each of the survey questions one or multiple categories to try to get at what type of question was being asked of respondents and whether it fell into a measure of the company or respondent's "Openness Culture" towards mental health or a few other categories: Demographics, Mental Health Disorder Information, Awareness/Accessibility of Mental Health Options, Employment Type, Healthcare Coverage, Respondent's Mental Health History and Workplace Relationships (see also table on pages 8-9).



Above is a visualization showing the maximum, minimum, and average feature importance for each question category. There are some categories that include more questions than others. In analyzing the feature importance in this way, we can see that openness culture which measures both employer openness and respondent openness towards mental health has the largest average feature importance as well as the largest maximum importance which makes sense when assessing this question. It also makes sense that respondent demographic factors like age and gender are also high when it comes to max importance but aren't as high when it comes to average feature importance which is hopeful for employers since they can't change employee demographics but they can change the openness of their culture.

Another important category is Awareness/Accessibility of mental health coverage and available mental health options. Employers can control this aspect for their employees as well by ensuring that their employees are often reminded of the different mental health services provided by both their health coverage as well as other options for mental health resources like optional services and community based resources. It makes sense that Mental Health Disorder is also an important category, given that those that have experienced mental health disorders either currently or in the past are more likely to discuss their mental health issues with their employer.

For tech workers who are experiencing mental health problems, which types of workplace support are associated with greater productivity?

The goal of this analysis was to examine which workplace and help-seeking factors buffered the effects of mental health issues on productivity.

There were 529 individuals with data for the question "Do you believe your productivity is ever affected by a mental health issue?". (These cases had missing data on whether individuals were tech workers, but they were likely tech workers as most respondents were tech workers.) From respondents that answered the above question, data were only available for 3 questions related to medical coverage, knowledge of resources, and willingness-to-disclose:

1. Do you have medical coverage that includes treatment of mental health disorders?
2. Do you know local or online resources to seek help for a mental health issue?
3. If you have been diagnosed or treated for a mental health disorder, do you ever reveal this to coworkers or employees?

The table below shows the numbers and percentages of respondents reporting whether work productivity was ever affected by a mental health issue (n=529), split by whether respondents had a current mental health disorder. The vast majority of employees who have a current mental health disorder reported that their work productivity was affected (bolded cells).

	Number reporting productivity affected at work (%)			
Current mental health disorder	**Yes**	**Unsure**	**No**	**Not applicable**
Yes	**204 (90.7)**	13 (5.8)	8 (3.6)	0 (0.0)
Maybe	**97 (84.3)**	11 (9.6)	3 (2.6)	4 (3.5)
No	69 (42.3)	35 (21.5)	14 (8.6)	45 (27.6)
Don't know	26 (65.4)	5 (19.2)	1 (3.8)	3 (11.5)

A **multinomial logistic regression** was conducted on the subset of respondents who currently had or thought they might have a mental health disorder, *and* who reported that their productivity was affected by a mental health issue (n=301; bolded cells).

Three predictors were recoded to binary variables: 1) Having or not having medical coverage for mental health treatment, 2) knowing or not knowing resources for help, and 3) being willing or unwilling to reveal a diagnosis to coworkers. The outcome measure was the respondents' answer to the question "If [you believe your productivity is ever affected by a mental health issue], what percentage of your work time is affected by a mental health issue?", with four options for responses: 1-25%, 26-50%, 51-75%, 76-100%. Rows with missing data in any of the predictors or outcome measure were dropped, resulting in a small remaining sample (n=135).

Using the LogisticRegression function from scikit-learn, with parameters: class = 'multinomial', solver = 'lbfgs', fit_intercept = True, penalty = 'None', class_weight = 'balanced', the resulting model had poor accuracy, correctly predicting only 27% of the 135 cases (see confusion matrix below). In other words, medical coverage, knowledge of resources, and willingness-to-disclose did not predict how much respondents' mental health issues affected their productivity at work. However, the small amount of data likely affected the quality of this model.

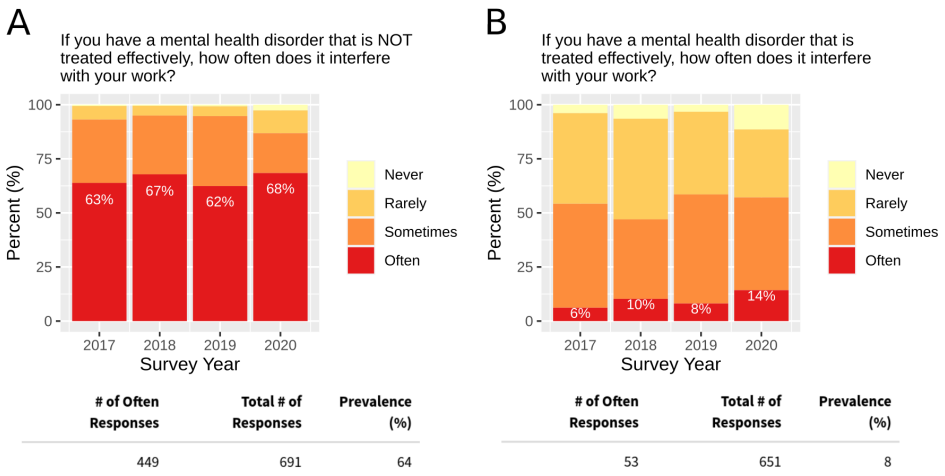
		Predicted responses about percentage of time affected			
		1-25%	**26-50%**	**51-75%**	**76-100%**
Actual responses	**1-25%**	**11**	2	27	13
	26-50%	4	**6**	21	13
	51-75%	4	2	**14**	5
	76-100%	2	0	5	**6**

How important is effective treatment of mental health disorders to tech workers' productivity?

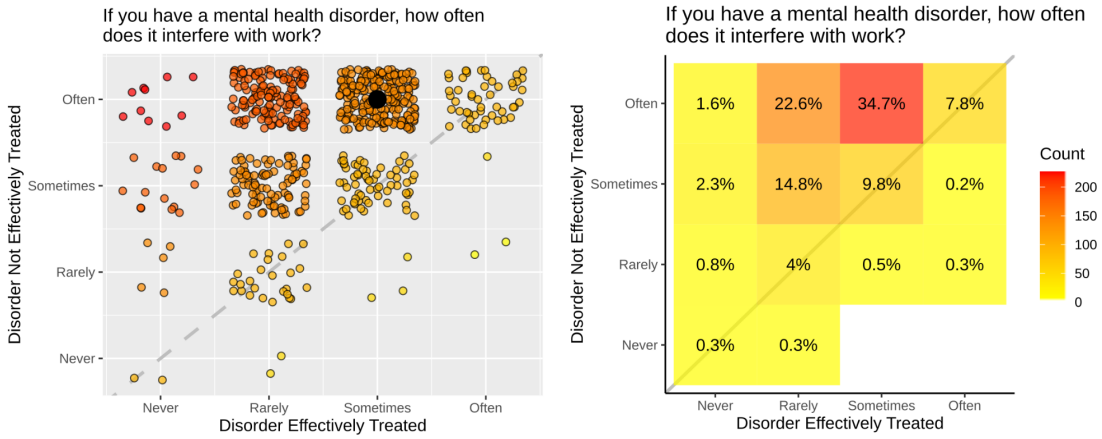
Two questions dealt with the effect of treatment of mental health disorder on work productivity:

1. If you have a mental health disorder, how often do you feel that it interferes with your work when NOT being treated effectively (i.e., when you are experiencing symptoms)?
2. If you have a mental health disorder, how often do you feel that it interferes with your work when being treated effectively?

These questions were only asked in 2017-2020 surveys. The figure below shows the percentage of responses, colored according to severity: with "Often" being red and "Never" being light yellow ("Often" means mental health issues interfere with work often). A shows responses when a mental health disorder is NOT being treated effectively, while B shows responses when their mental health issues are being treated effectively, with the percentage of respondents saying "Often" moving from 64% to 8% with effective treatment of disorder/symptoms. This shows that for those with mental health disorders, effective treatment may be key to boosting productivity.



Following the robust trend we identified in the figure above, we sought to test whether there was an effect of treatment on how often mental health disorders interfered with the respondent's work within the same individuals. This is possible since the same respondents answered both questions in most cases.



We found that ~77% of US tech respondents said their mental health disorder interfered with their work more often when not treated effectively compared to when treated effectively (data points to the left and above of the unity line). More than a third of US tech respondents said that their mental health disorder interfered with their work, sometimes when effectively treated and often when not effectively treated (indicated by the black dot which represents the median of the dataset). Albeit few, particularly concerning are those that responded that when their disorder is being well treated, it never interferes with their work, however when it is not well treated, their work is often impacted (left panel color-coding: redder colors indicates more work interference when disorder is not effectively treated). The analysis shows that effective treatment of mental health disorders is a key factor in reducing interference from mental health issues, thus boosting tech workers' productivity.

3. Conclusions and Take-Aways

From the opt-in OSMI survey, we found that the prevalence of mental health disorders among US tech workers was 48%. While this number is likely inflated due to self-selection of respondents, it suggests that mental health disorders are common in tech workers. Seeking professional help for mental health issues is also a common behavior, with 66% of US tech workers doing so. Anxiety, mood and attention deficit hyperactivity disorders were among the most common mental health disorders reported by US tech workers.

It is difficult to predict who is at risk for mental health issues just by looking at basic demographic characteristics (age, gender, race). Large population surveys suggest that females are at higher risk for poor mental health, but we do not know the underlying causes.

Most tech workers who have a mental health disorder report that it affects their productivity to some extent. Having access to effective mental health treatment is key, as tech workers who have mental health issues experience less interference with their work when their mental health issues are being treated effectively.

There are many ways for employers to create a culture that is more open to speaking about mental health, including providing mental healthcare options, laying out these options clearly so that employees are aware of them, and incorporating mental health as a standard part of communication about wellness.

4. Future directions

Greater mental health burden on females

One direction for further research could be to better understand why people who identify as female have a greater burden on their mental health than those who identify as male. In the population health survey, females had more poor mental health days than males. In the OSMI, the risk of having a mental health disorder was higher for females (though gender was a weak predictor). It is well known that the bulk of household labor and caregiving falls on women rather than men, adding to their workload. In the tech world, women tend to be in the minority and are subject to various biases, which may lead to greater stress and lower work satisfaction.

Relationship between mental health support and worklife satisfaction

We felt that the OSMI survey lacked one key category of questions that could have drawn a direct connection between mental health support and a company’s bottom line: there were no questions in the OSMI survey related to job satisfaction, only to productivity (and only for those affected by mental health issues). We expect that tech workers would be happier and more likely to remain at companies with better mental health support and a more open culture around talking about mental health.

5. Acknowledgement

We would like to acknowledge the amazing teaching assistant Kessie Zhang and mentor Nisha Kumaraswamy for their helpful input and discussion on this project. We are grateful to the staff at Correlation One and all others for working tirelessly to organize, teach, mentor and actively participate in this DS4A / Women summit.

6. References

Open Sourcing Mental Illness (OSMI) (2014-2020). https://osmihelp.org/research

Centers for Disease Control and Intervention (CDC) Behavioral Risk Factor Surveillance System (BRFSS) Behavioral Risk Factor Surveillance System (1984-2021). https://www.cdc.gov/brfss/about/index.htm

Facebook whistleblower testimony (2021). https://www.npr.org/2021/10/05/1043377310/facebook-whistleblower-frances-haugen-congress

Netflix employees walk-out (2021). https://www.cnn.com/2021/10/20/entertainment/netflix-employees-walk-out/index.html