

Mental Health in the Tech Industry

Silei Li and Mariel Pacada

12/8/2019

```
# Download the raw data needed for this project from these github links :)
# https://github.com/marielpacada/mental-health-in-tech/blob/master/mental-health-in-tech.csv
# https://github.com/marielpacada/mental-health-in-tech/blob/master/countries.csv
```

Dataset and Goals

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

Data Cleaning Choices

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

Part 1: Initial Exploratory Data Analysis

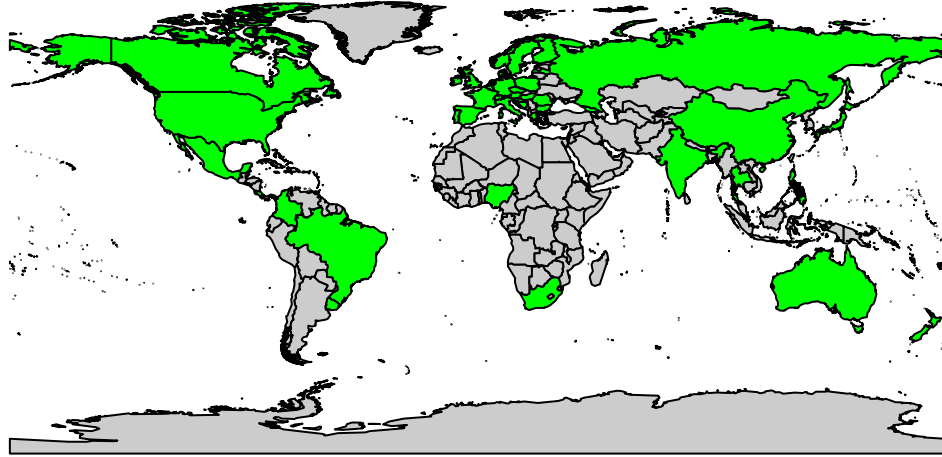
First, let's get to know our data!

Below we see a map in which all the countries represented in the survey are shaded in.

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

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

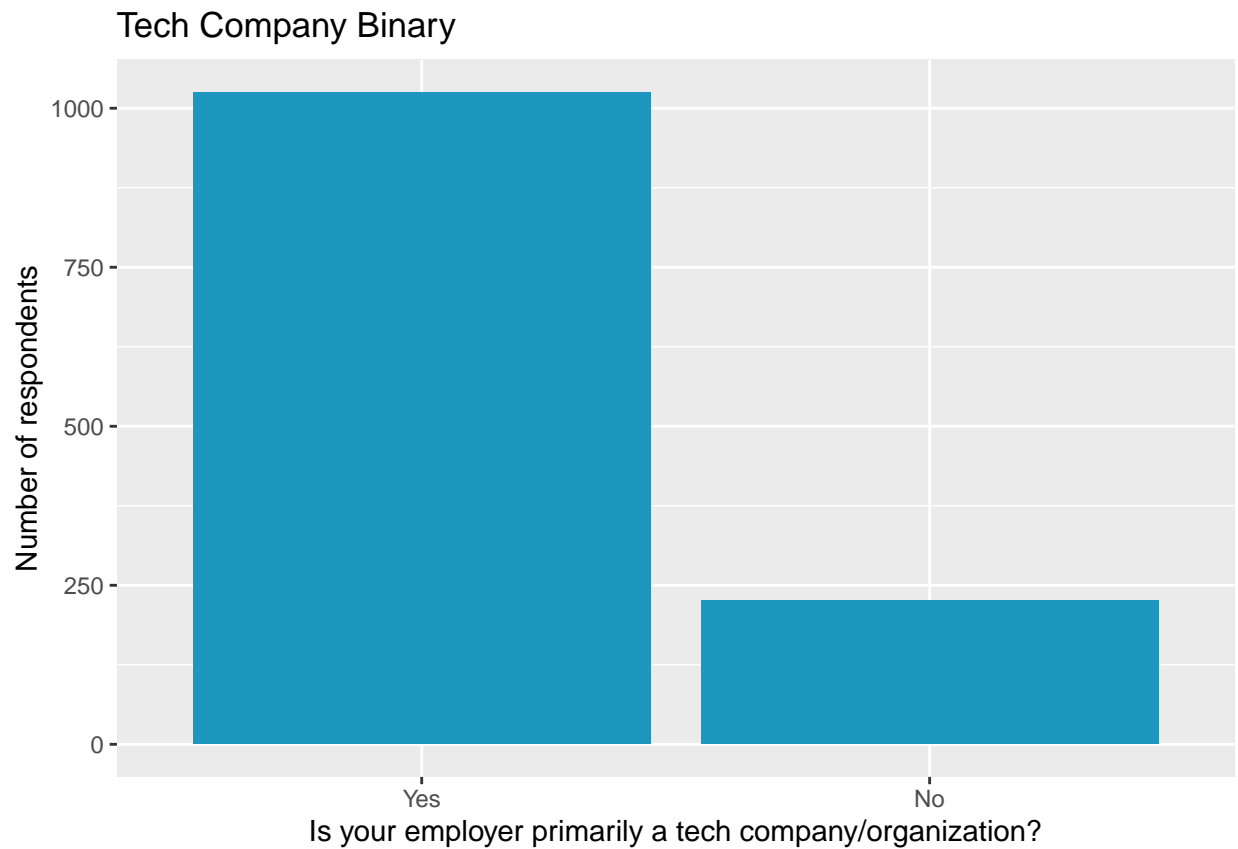
Countries represented in the dataset



Next, we see that most, but not all, of the employees in the dataset work for companies that are primarily a tech company or organization.

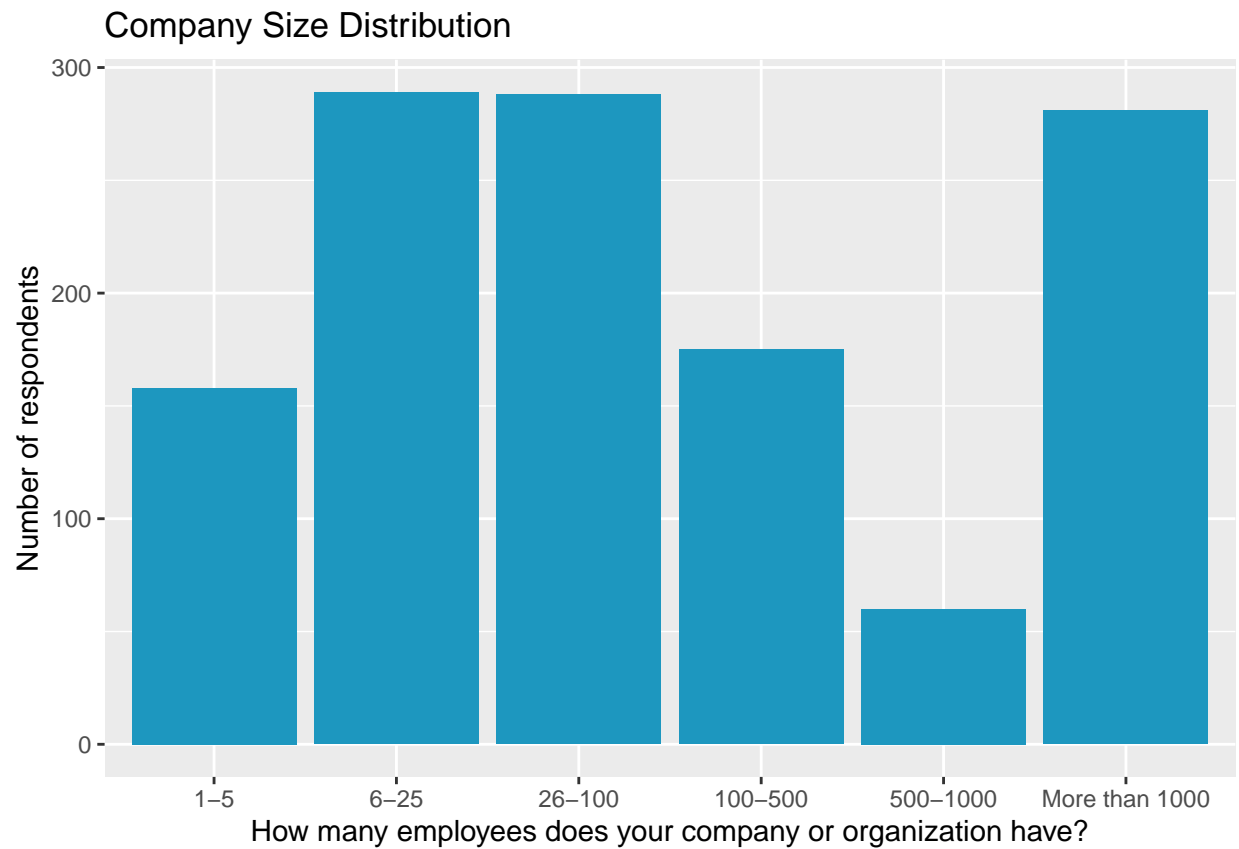
```
tech_binary <- survey %>%
  filter(!is.na(tech_company)) %>%
  dplyr::select(tech_company) %>%
  mutate(tech_company = factor(tech_company, levels = c("Yes", "No")))

ggplot(tech_binary, aes(x = tech_company)) + geom_bar(fill = "#1D97BF") + labs(x = "Is your employer pr
```



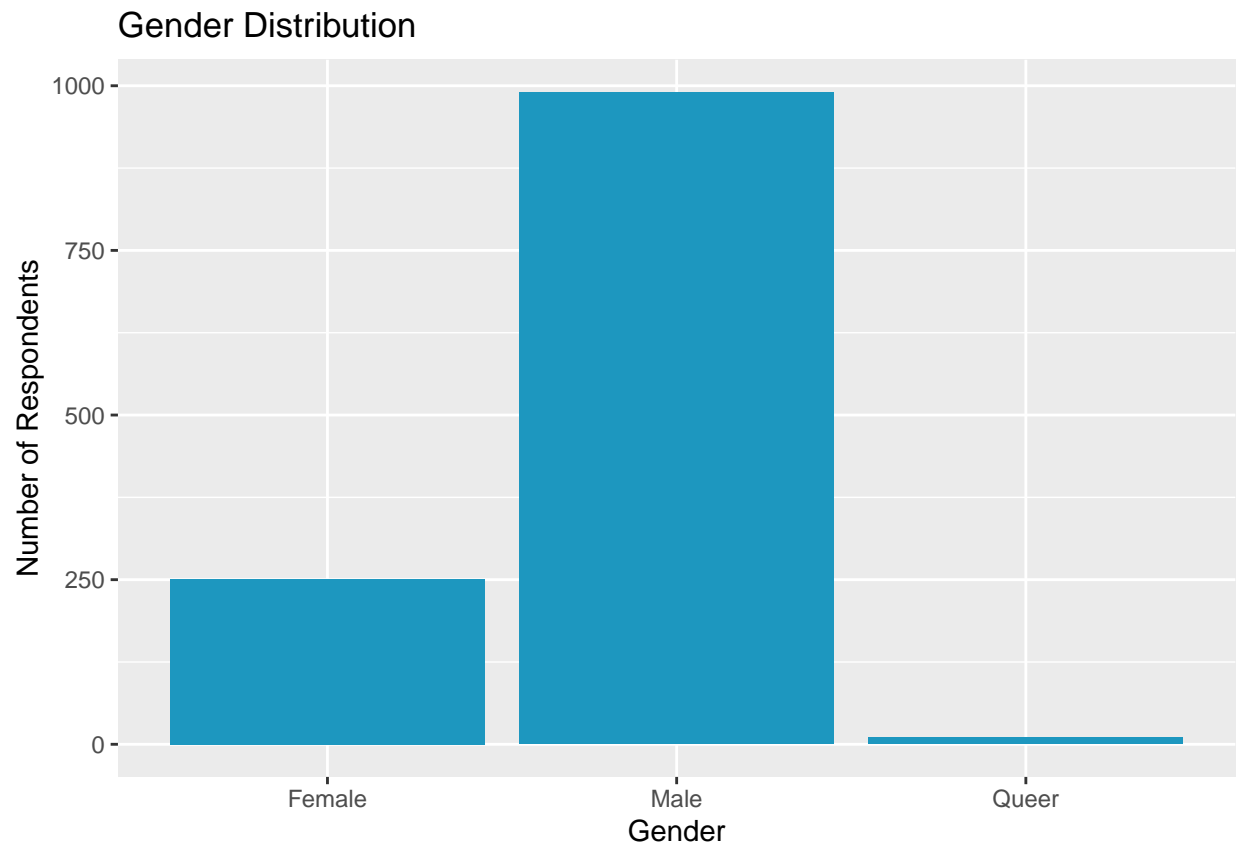
Here, we see the distribution of company size.

```
num_employee <- survey %>%  
  dplyr::select(no_employees) %>%  
  mutate(no_employees = factor(no_employees, levels = c("1-5", "6-25", "26-100", "100-500",  
    "500-1000", "More than 1000")))  
  
ggplot(num_employee, aes(x = no_employees)) + geom_bar(fill = "#1D97BF") + labs(x = "How many employees")
```



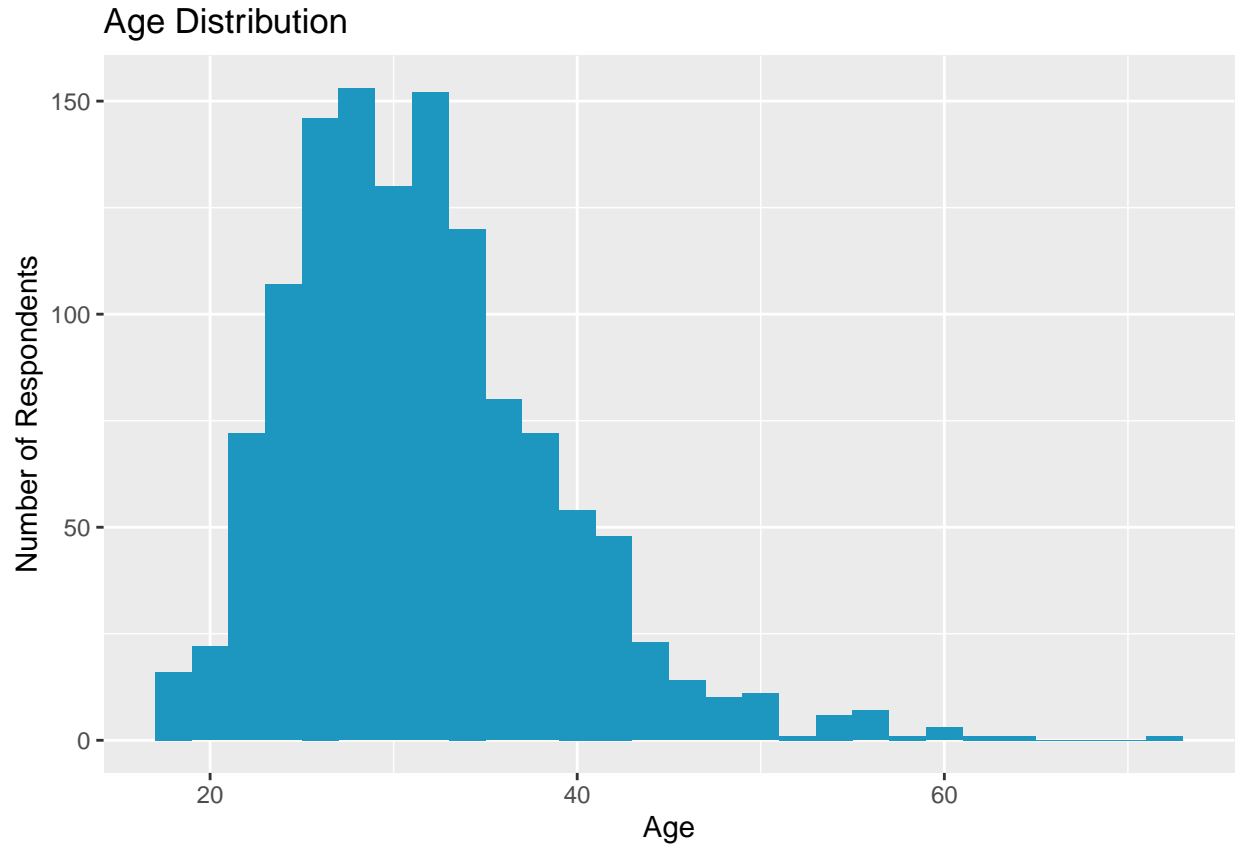
This is the distribution of gender. Surprised?

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



Finally, this is the distribution of age.

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



Part 2: Correlational Analysis

For this part, we introduce the hypotheses we aim to explore.

General Hypotheses

1. Company size is related to an employee's willingness to discuss a mental health issue.

Bigger companies may have more competitive environments; thus, people may be less likely to discuss their mental health situations and seek help. Smaller companies may be more private and personal, where employees and employers or supervisors may know each other better and have more time and intimacy to communicate.

2. Bigger companies provide better mental health care benefits and resources.

Bigger companies may be richer and more willing to spend on the mental wellness of their employees.

3. We expect some geographical difference among mental health care benefits.

Different countries or different states in the United States have different legislations regarding how companies should handle employee's mental health conditions, and this may be reflected in the respondent's answers to the this survey.

Methodologies

We will use a number of methods to illustrate and attempt to answer the questions we have. We have lots of categorical data that cannot be very well analyzed with continuous graphs. We are therefore using **scatterplots** to show the clustering patterns, and **barplots** to show data distribution. We will also make **maps** to demonstrate how mental health benefits vary depending on where a person is employed. In order to test if larger companies provide better mental health care resources, we will use the mental health benefit features to predict company sizes. Because most data is categorical, we will build **decision trees** and **Naive Bayes models**. Lastly, we will use a **t-test** to discern the significance of our data.

If you have a mental health condition, do you feel that it interferes with your work?

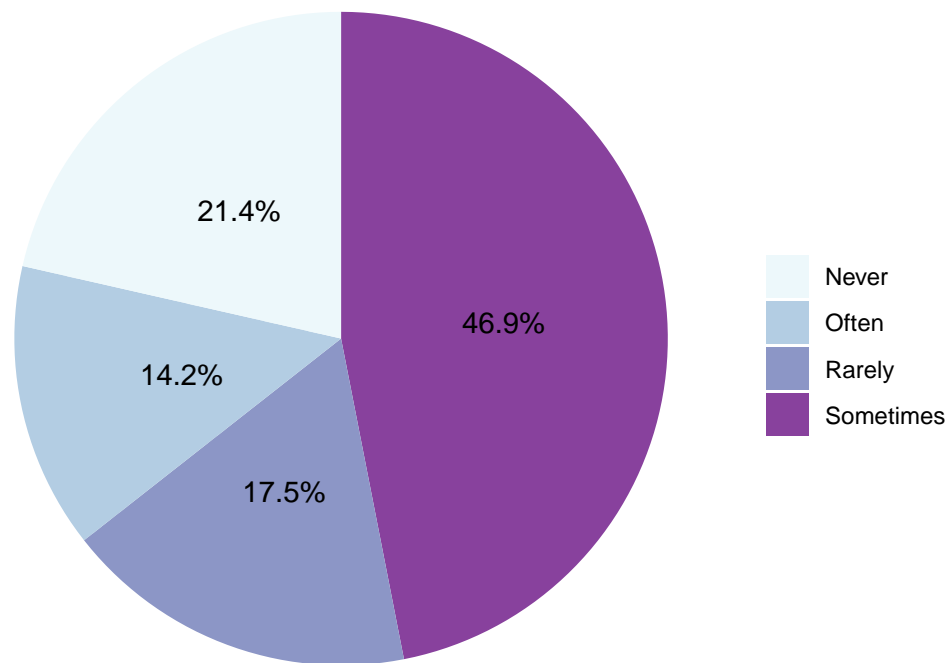
Most respondents do feel so!

Main takeaway: For those who have a mental health condition, there is noticeable interference in their work. The interference is significant especially if the person has sought treatment.

Among everyone:

```
pie_all <- data.frame(table(survey$work_interfere))
#this table() function automatically excludes (not counting) the NAs
ggplot(pie_all, aes(fill=Var1,x="",y=Freq))+
  geom_bar(stat="identity")+
  labs(x="",y="percentage of people",title="Does your mental health confition interfere with your work?")
  coord_polar("y", start=0)+
  theme_void()+
  geom_text(aes(label = percent(Freq/(sum(Freq)))), position = position_stack(vjust = 0.5)) +
  theme(legend.title = element_blank())+
  scale_fill_brewer(palette="BuPu")
```

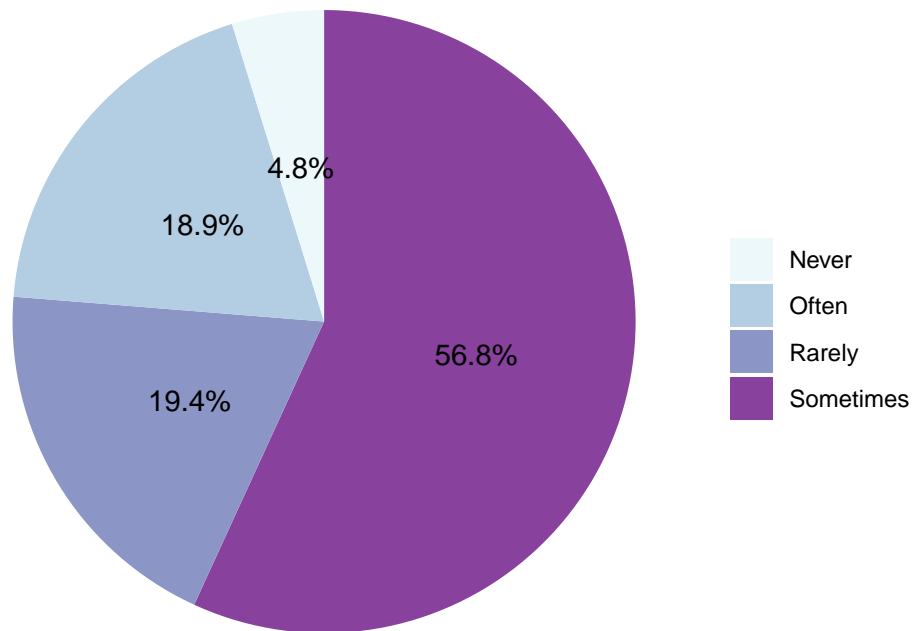
Does your mental health condition interfere with your work?



Among those who have sought treatment:

```
pie_treat <- data.frame(table(survey %>%  
  filter(work_interfere != "NA" & treatment=="Yes") %>%  
  dplyr::select(work_interfere)))  
  
ggplot(pie_treat, aes(fill=Var1,x="",y=Freq)) +  
  geom_bar(stat="identity") +  
  labs(x = "", y = "percentage of people", title = "Does your mental health condition interfere with you  
    (Given the person has sought treatment)") +  
  coord_polar("y", start=0) +  
  theme_void() +  
  geom_text(aes(label = percent(Freq/(sum(Freq)))), position = position_stack(vjust = 0.5)) +  
  theme(legend.title = element_blank()) +  
  scale_fill_brewer(palette="BuPu")
```

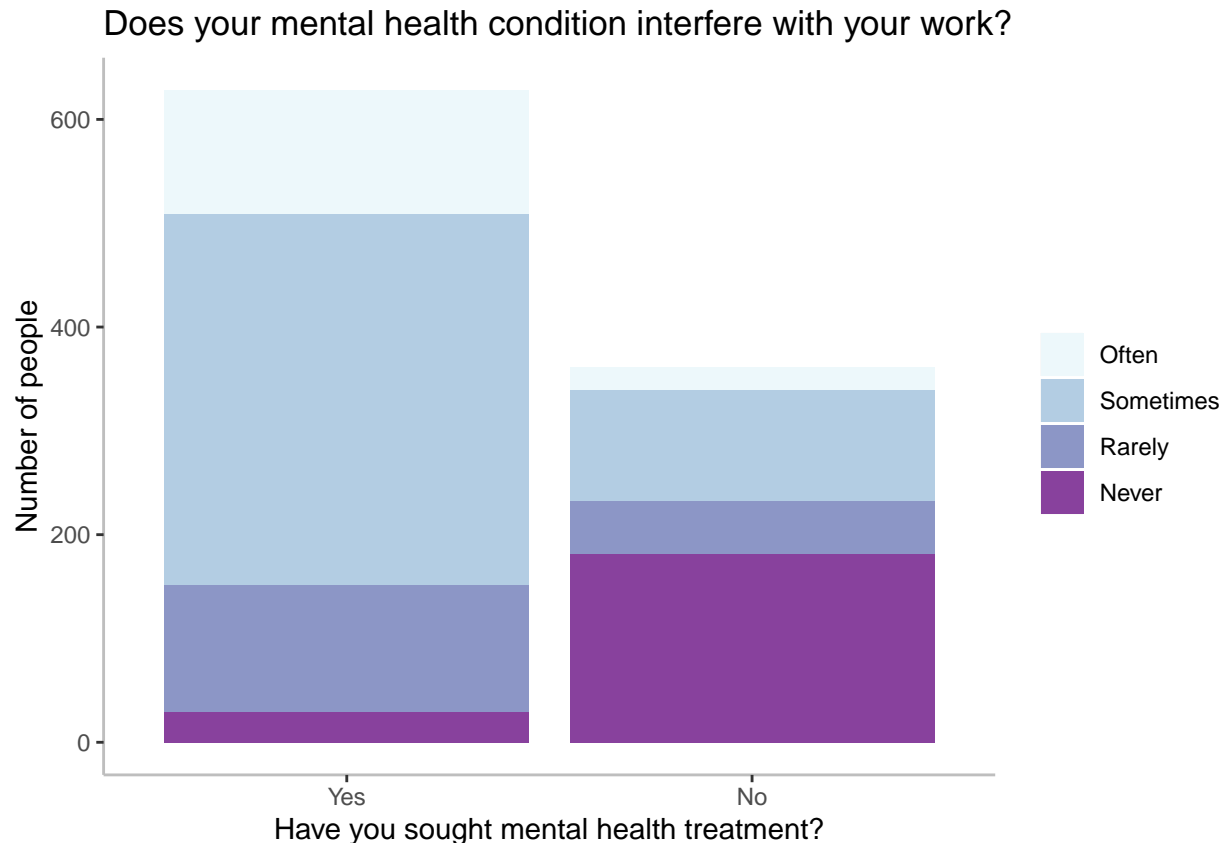

Does your mental health condition interfere with your work?
(Given the person has sought treatment)



Compare the work interference from mental health condition, between those who have sought treatment vs. those who have not:

We observe that people who have sought treatment are more likely and more frequently affected by their mental health conditions (It may be that the conditions are interfering with their work, that led them to seek treatment).

```
survey%>%
  filter(work_interfere!="NA")%>%
  mutate(work_interfere = factor(work_interfere, levels = c("Often", "Sometimes", "Rarely", "Never")))
  ggplot(aes(fill=work_interfere,x=factor(treatment,levels=c("Yes","No")),y=1))+geom_bar(position="stack")
  labs(x="Have you sought mental health treatment?",y="Number of people",title="Does your mental health")
  theme(legend.title = element_blank(),panel.grid.major = element_blank(), panel.grid.minor = element_b
panel.background = element_blank(), axis.line = element_line(colour = "grey"))+
  scale_fill_brewer(palette="BuPu")
```



Communicating the problem: If you have a problem, say it out!

People are often reluctant to talk about a mental health issue they might have. Why is this? Are they worried about facing consequences if the employer knows? If it's the employer that they are afraid of, are people more willing to talk to coworkers about it? Is it easier to speak of a physical health issue compared to mental health? Does the situation vary in companies of different sizes - do bigger companies do a better job in promoting conversations, or do they exert worse stress?

Let's explore the data further!

How does observation of negative consequences for coworkers with mental health conditions in workplace affect discussion of mental health conditions?

Surprisingly, it seems that the expectation of negative consequence is not highly related to whether one has seen a coworker facing the situation.

Main takeaway: There is some relationship between the two.

For example, those who said that they have observed coworkers facing consequences are more likely to expect that discussing mental health issue will lead to consequence than not; however, some of them still believe it won't necessarily lead to negative result. Similarly, for those who have not heard of or seen coworkers facing negative consequences, they are more likely to expect that discussing such issue will not bring a consequence. Many of them still expect some negative consequences - not seeing similar situations does not free everyone from the worries.

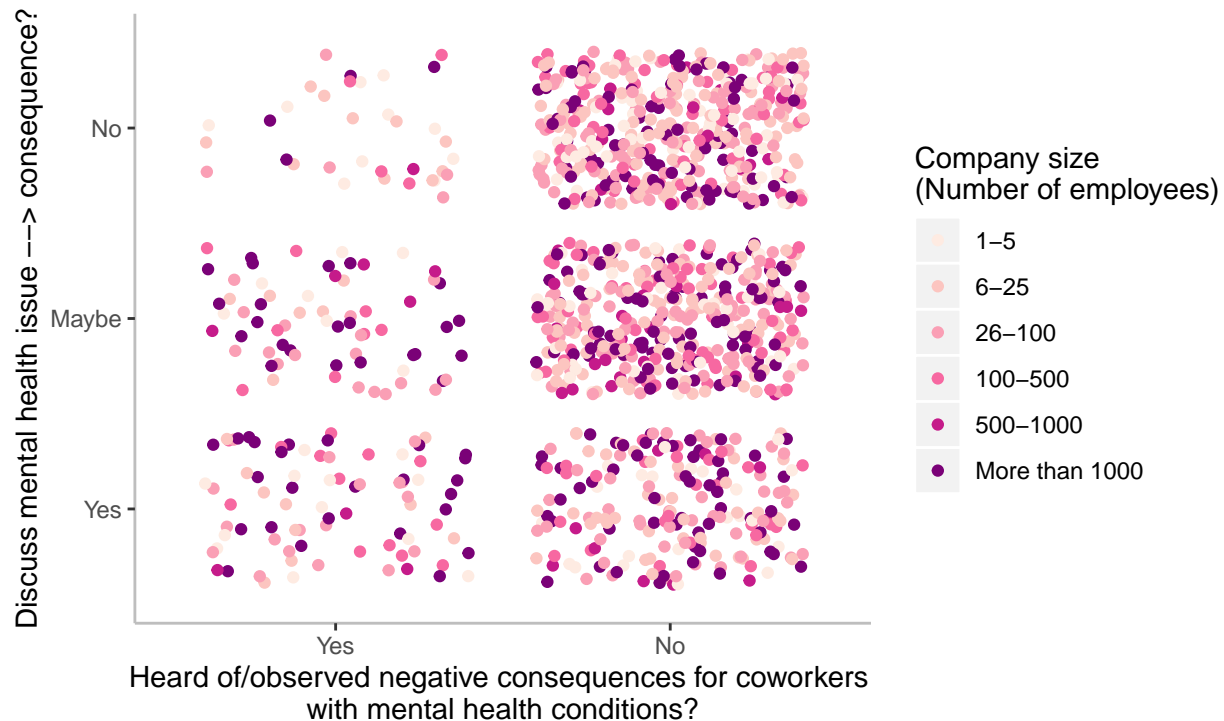
```
#survey%>% filter(treatment=="YES")
ggplot(survey, aes(x=factor(obs_consequence,levels = c("Yes","No")),
                  y=factor(mental_health_consequence, levels = c("Yes", "Maybe","No")),
                  col=factor(no_employees,levels=c("1-5","6-25","26-100","100-500","500-1000","More than 1000"))),
  geom_point(position="jitter")+
  labs(title="Is seeing a coworker with mental health condition face consequence \nrelated to people's expectation that discussing such an issue leads\n\n",
       x="Heard of/observed negative consequences for coworkers \nwith mental health conditions?",
       y="Discuss mental health issue --> consequence?",
       col = "Company size \n(Number of employees)"+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank(), axis.line = element_line(colour = "grey"))+
  scale_color_brewer(palette="RdPu")
```

Is seeing a coworker with mental health condition face consequence related to people's expectation that discussing such an issue leads to a similar consequence?



```
#survey%>% filter(treatment=="YES")
ggplot(survey, aes(x=factor(obs_consequence,levels = c("Yes","No")),
                  y=factor(mental_health_consequence, levels = c("Yes", "Maybe","No")),
                  col=factor(no_employees,levels=c("1-5","6-25","26-100","100-500","500-1000","More than 1000"))),
  geom_point(position="jitter")+
  labs(title="Is seeing a coworker with mental health condition face consequence \nrelated to people's expectation that discussing such an issue leads\n\n",
       x="Heard of/observed negative consequences for coworkers \nwith mental health conditions?",
       y="Discuss mental health issue --> consequence?",
       col = "Company size \n(Number of employees)"+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
        panel.background = element_blank(), axis.line = element_line(colour = "grey"))+
  scale_color_brewer(palette="RdPu")
```

Is seeing a coworker with mental health condition face consequence related to people's expectation that discussing such an issue leads to a similar consequence?



Why do people not talk to supervisors about mental health problems? Is it that they are afraid of negative consequences?

We notice many people are inclined not to talk about mental health problems with their supervisors.

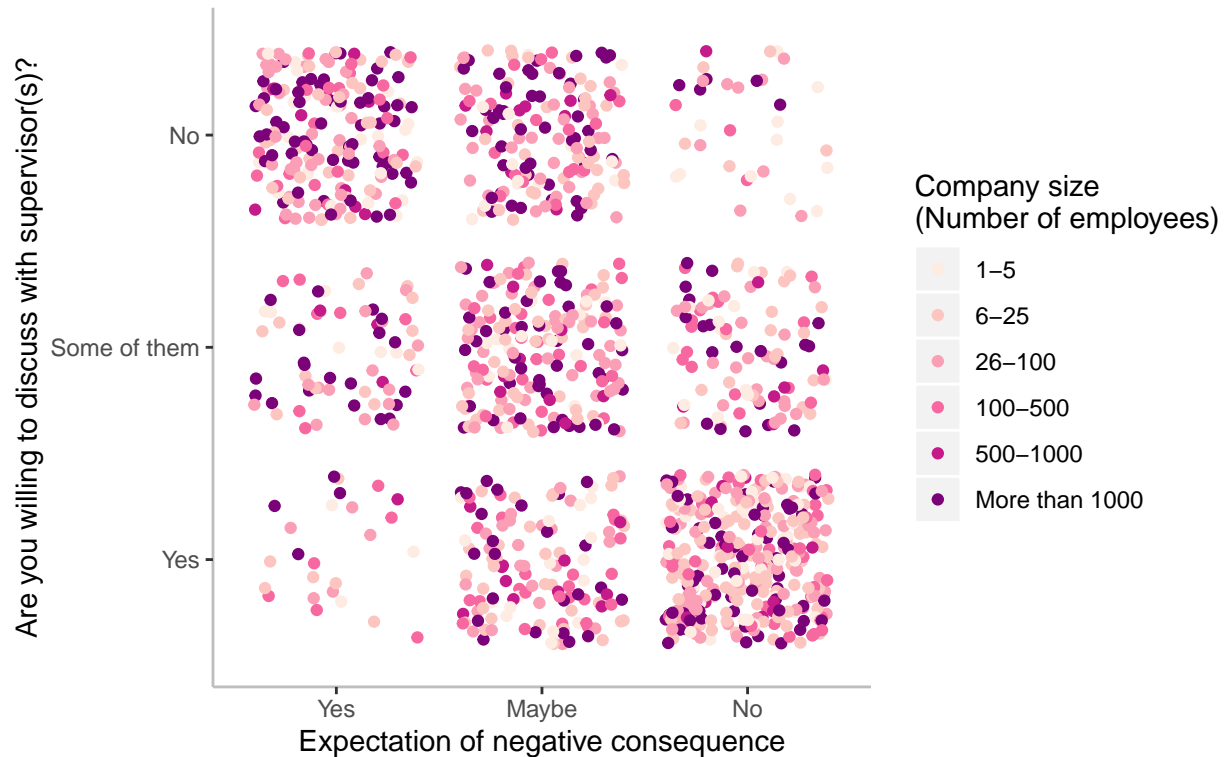
The graph here shows that if an employee thinks discussing a mental health issue with the supervisor will not likely lead to some negative consequence, they will be more willing to talk about it. Vice versa, if people expect the discussion to have negative impact, they are less likely to communicate the issue with supervisors.

However, we can also observe that it's not just about expectation of negative consequence. It seems that people are reluctant to discuss mental health issues with others in general, whether it's a supervisor or a coworker.

We still do not observe a significant effect from the size of the companies.

```
ggplot(survey, aes(x=factor(mental_health_consequence, levels = c("Yes", "Maybe", "No")),
  y=factor(supervisor, levels = c("Yes", "Some of them", "No")),
  col=factor(survey$no_employees, levels=c("1-5", "6-25", "26-100", "100-500", "500-1000", "More than 1000")),
  geom_point(position="jitter")+
  labs(x="Expectation of negative consequence",
    y="Are you willing to discuss with supervisor(s)?",
    title="Do people avoid talking to supervisors about mental health issues \nbecause they think it
    col = "Company size \n(Number of employees)"+
  theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
    panel.background = element_blank(), axis.line = element_line(colour = "grey"))+
  scale_color_brewer(palette="RdPu")
```

Do people avoid talking to supervisors about mental health issues because they think it brings negative consequences?



Are people more willing to discuss mental health issues with coworkers compared to supervisors?

It seems like most people are unwilling to discuss mental health issues with supervisors in general, regardless if they think there will be a negative consequence. We are curious whether the communication is better with coworkers.

If there is a negative correlation, maybe people choose to discuss only with coworkers, and feel it's enough expressing their concerns to coworkers, and they don't bother talking about it to supervisors.

If there is a positive relationship, maybe people are afraid that the coworkers will tell the supervisors, or they just don't like other people to know their mental health conditions in general.

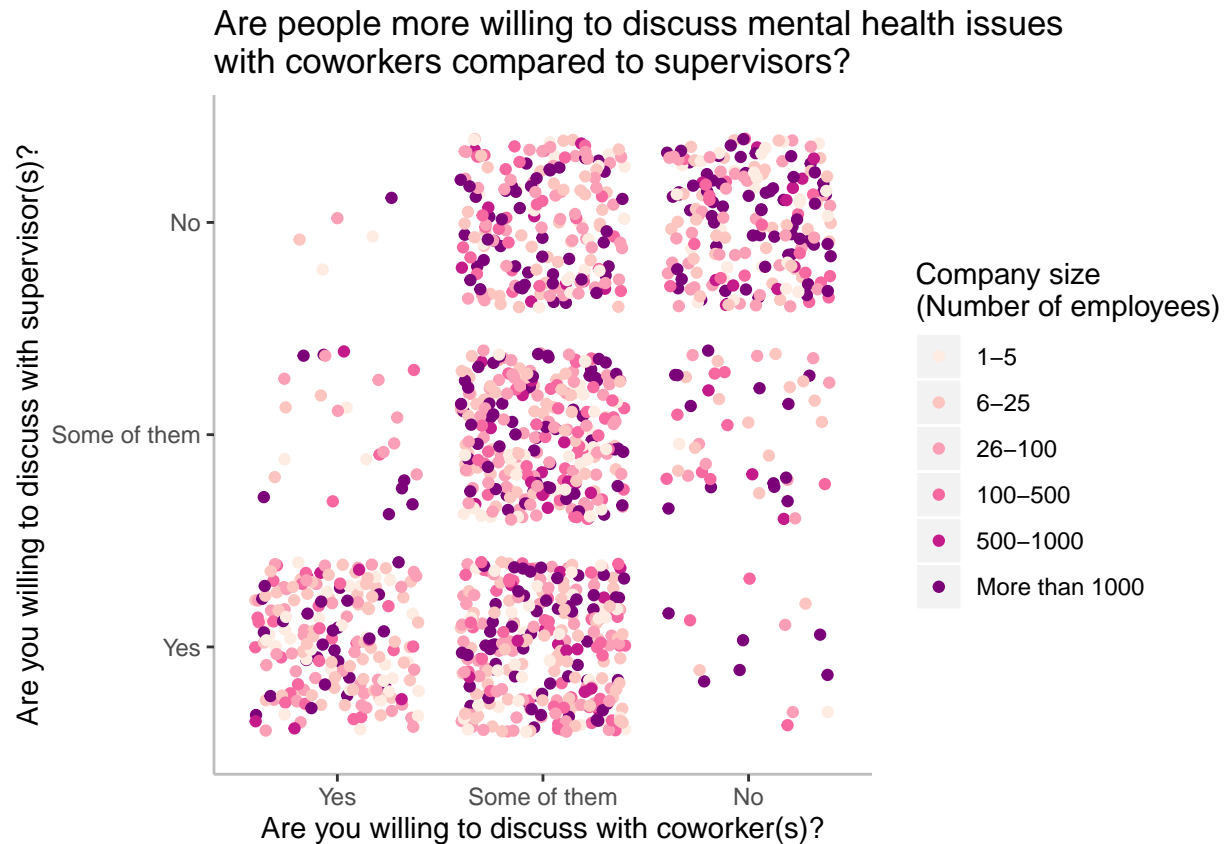
After making the plot below, we observe that **most people are willing to discuss with "some of" the coworkers**, which is quite understandable; in comparison, they are **less likely to discuss with supervisors in general**. However, we can see a positive correlation between the two: **People who are willing to discuss mental health issues with coworkers (who answered "Yes" and we assume they are open to discuss with all or at least most coworkers), are also likely to be willing to discuss it with supervisors**. Similarly, **people who are unwilling to discuss mental health issue with coworkers, are also likely unwilling to discuss it with supervisors**.

```
ggplot(survey, aes(x=factor(coworkers,levels = c("Yes", "Some of them","No")),
  y=factor(supervisor,levels = c("Yes", "Some of them","No")),
  col=factor(survey$no_employees,levels=c("1-5", "6-25", "26-100", "100-500", "500-1000", "More than 1000")),
  geom_point(position="jitter")+
  labs(x="Are you willing to discuss with coworker(s)?",
    y="Are you willing to discuss with supervisor(s)?",
```

```

title="Are people more willing to discuss mental health issues \nwith coworkers compared to super
col = "Company size \n(Number of employees)" +
theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
panel.background = element_blank(), axis.line = element_line(colour = "grey")) +
scale_color_brewer(palette="RdPu")

```



Is it just mental health conditions that people don't want to talk about? Or is discussing physical health conditions also facing barriers?

Physical health is also important to ensure life quality and working efficiency. Here, we compare people's willingness to bring up physical vs mental health issues before employment (during interviews) and during employment, with supervisors.

We observe from below: It seems that **some people just don't want to discuss any health issue they have with the supervisors (or potential employers), whether it's physical or mental. It also seems to be independent of company sizes.**

Overall, we should advocate for a working environment in these tech companies that empowers people to better communicate their health issues, whether it's physical and mental. Both employers and employees should better recognize the issues as solvable challenges, rather than barrier to work.

```

ggplot(survey, aes(x=factor(phys_health_interview,levels = c("No", "Maybe","Yes")),
y=factor(mental_health_interview,levels = c("No", "Maybe","Yes")),
col=factor(survey$no_employees,levels=c("1-5", "6-25", "26-100", "100-500", "500-1000", "More than 1000")),
geom_point(position="jitter") +
labs(title="Would you bring up a physical/mental issue during interview \nwith a potential employer?"

```

```

x="Bring up a PHYSICAL health issue?",
y="Bring up MENTAL health issue?",
col = "Company size \n(Number of employees)" +
theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(), panel.background = element
scale_color_brewer(palette="RdPu")

```

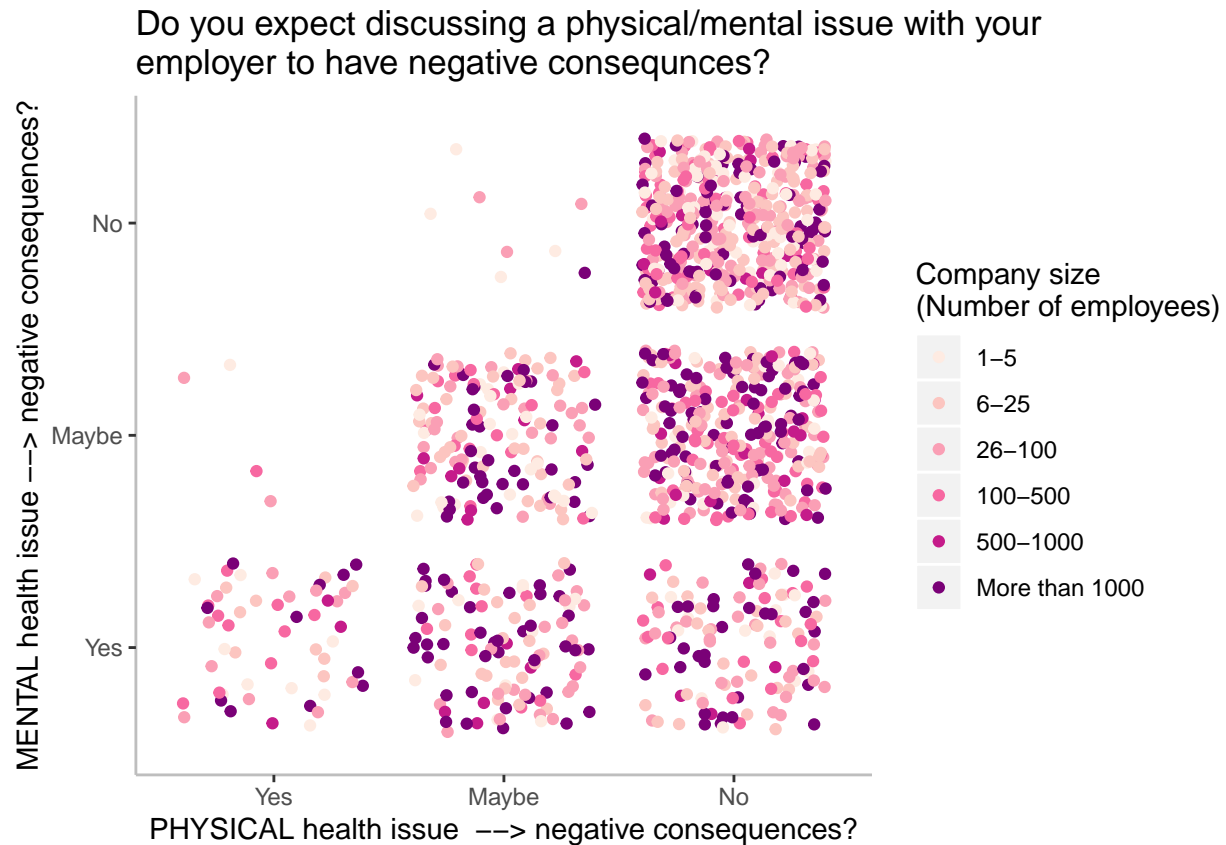
Would you bring up a physical/mental issue during interview with a potential employer?



```

ggplot(survey, aes(x=factor(phys_health_consequence, levels = c("Yes", "Maybe", "No")),
y=factor(mental_health_consequence, levels = c("Yes", "Maybe", "No")),
col=factor(survey$no_employees, levels=c("1-5", "6-25", "26-100", "100-500", "500-1000", "More than 1000")),
geom_point(position="jitter")+
labs(title="Do you expect discussing a physical/mental issue with your \nemployer to have negative consequences?",
x="PHYSICAL health issue --> negative consequences?",
y="MENTAL health issue --> negative consequences?",
col = "Company size \n(Number of employees)" +
theme(panel.grid.major = element_blank(), panel.grid.minor = element_blank(), panel.background = element
scale_color_brewer(palette="RdPu")

```



Now, let's explore some correlations between the size of a company and its mental health resources. When viewing some of these graphs, we advise that you see "Don't know" as part of "No", as it suggests that the company might have failed to inform the employees about their care options.

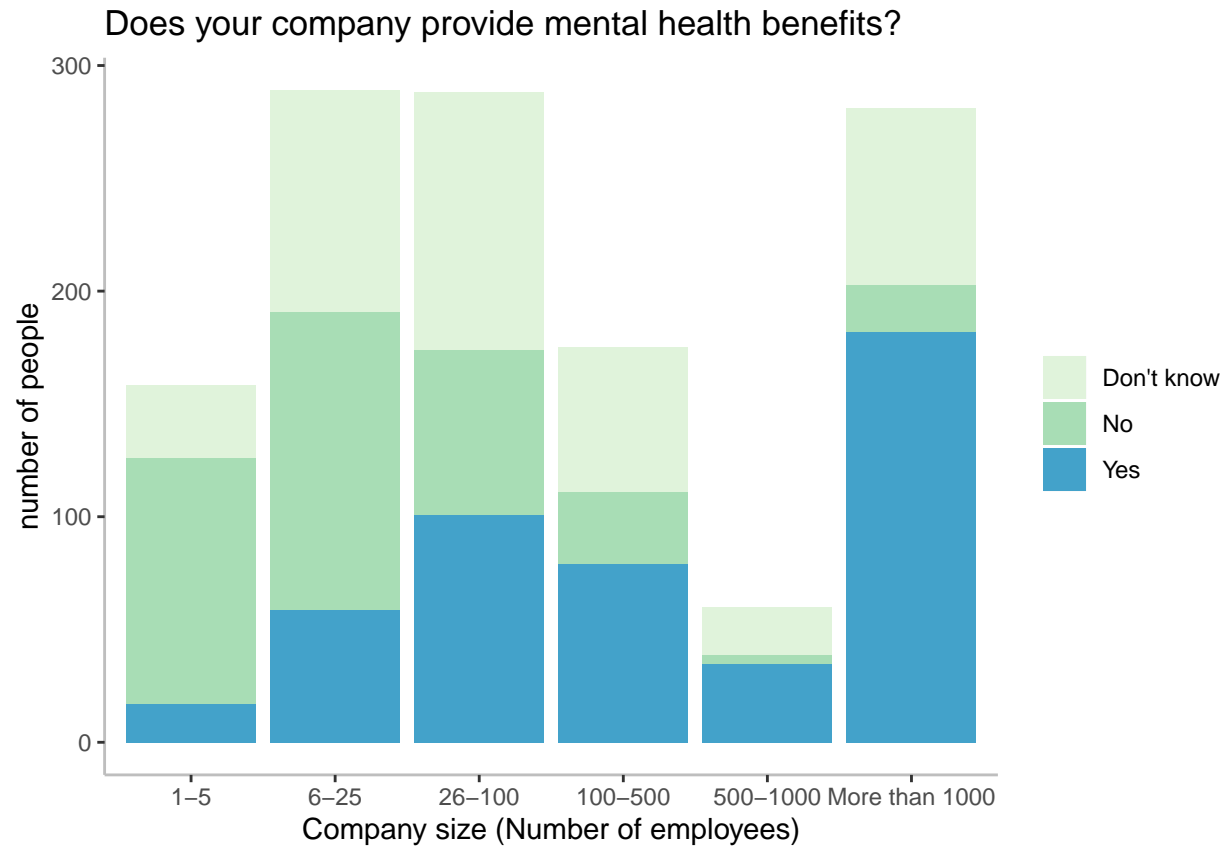
Are larger companies more likely to provide mental health benefits?

Generally speaking, yes!

Main takeaway: While bigger companies seem to provide more benefits, they don't necessarily do a better job at informing employees about their care options.

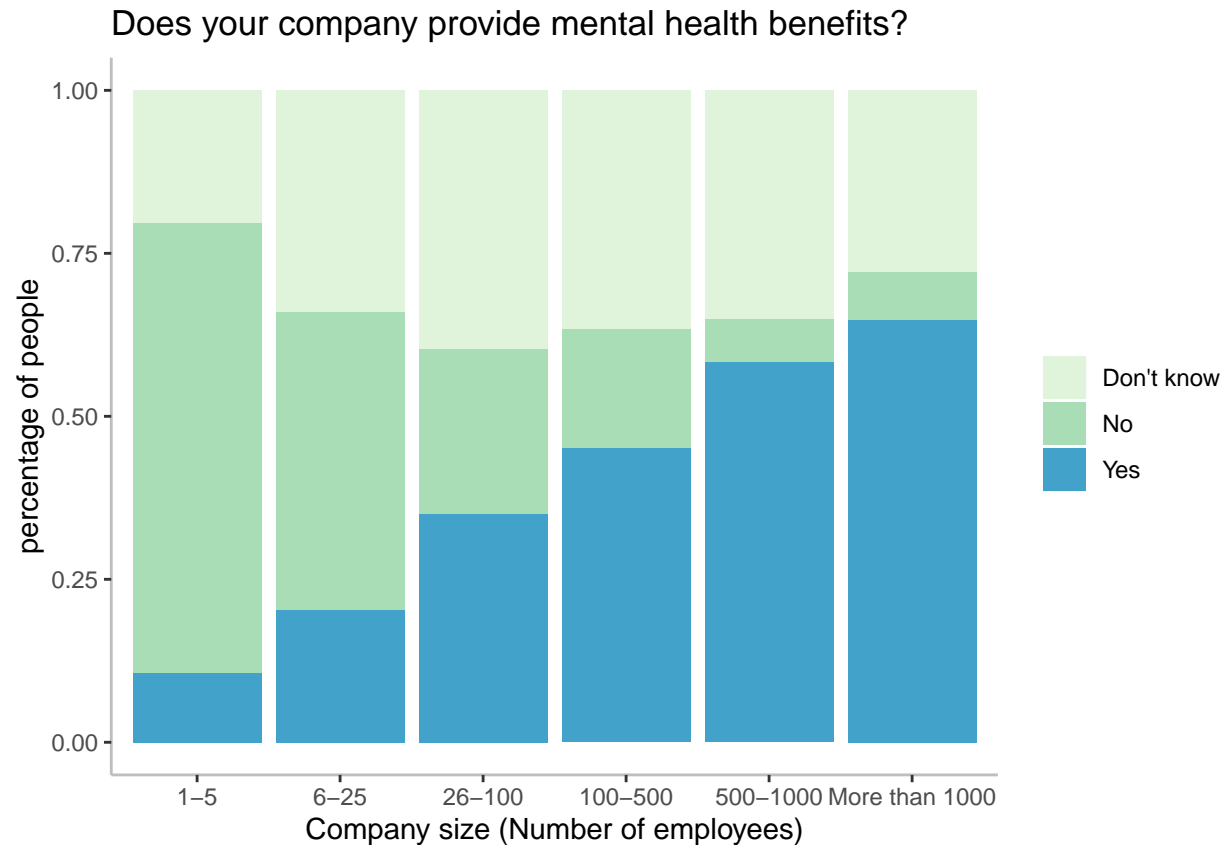
Raw data:

```
survey%>%
  mutate(no_employees = factor(no_employees, levels = c("1-5", "6-25", "26-100", "100-500", "500-1000", "More than 1000")))
  ggplot(aes(fill=benefits, x=no_employees, y=1)) + geom_bar(position="stack", stat="identity") +
  labs(x="Company size (Number of employees)", y="number of people", title="Does your company provide mental health benefits?") +
  theme(legend.title = element_blank(), panel.grid.major = element_blank(), panel.grid.minor = element_blank()) +
  scale_fill_brewer(palette="GnBu")
```

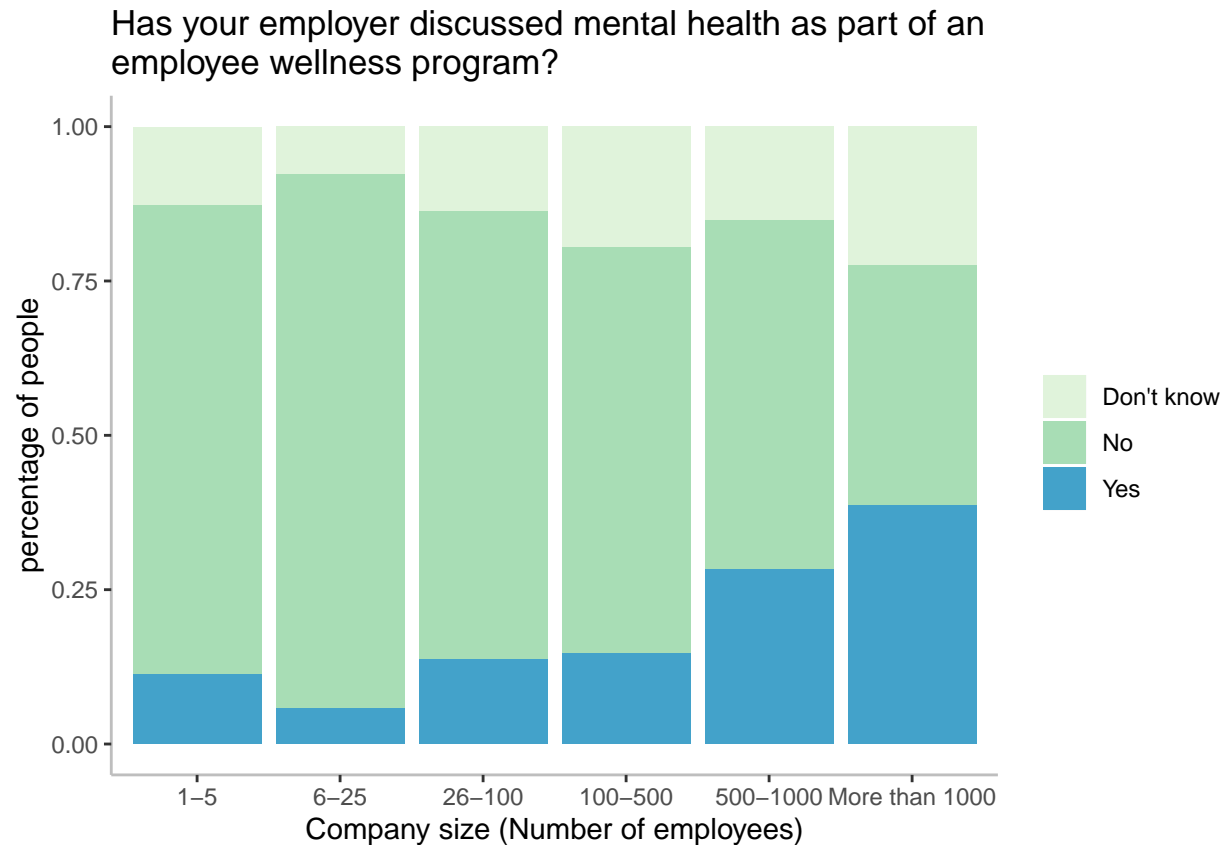
Normalized data:

```
survey%>%
  mutate(no_employees = factor(no_employees, levels = c("1-5", "6-25", "26-100", "100-500", "500-1000", "More than 1000")))
  ggplot(aes(fill=benefits, x=no_employees, y=1)) + geom_bar(position="fill", stat="identity") +
  labs(x="Company size (Number of employees)", y="percentage of people", title="Does your company provide mental health benefits?") +
  theme(legend.title = element_blank(), panel.grid.major = element_blank(), panel.grid.minor = element_blank()) +
  scale_fill_brewer(palette="GnBu")
```



We then examined whether the companies provide mental health care as part of their employee wellness program. It seems that in general, larger companies are more likely to introduce mental health benefits as part of a wellness program, which might provide more systematic service.

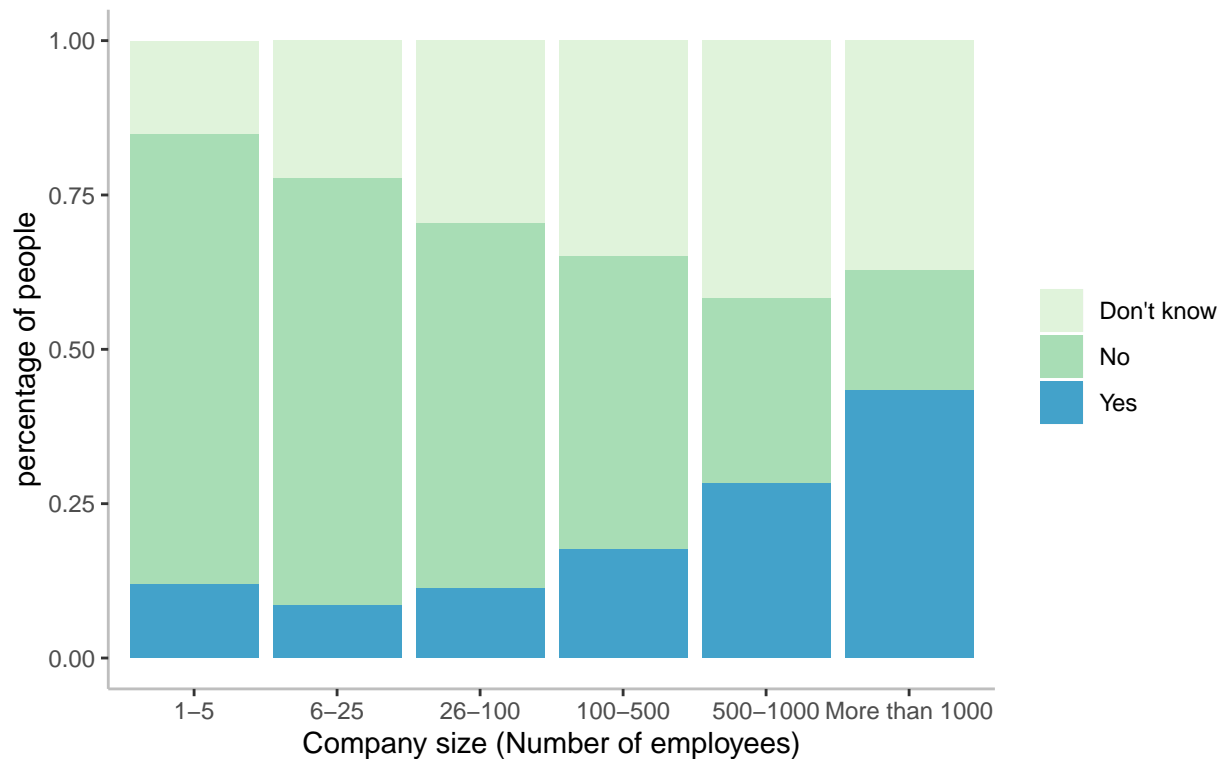
```
survey%>%
  mutate(no_employees = factor(no_employees, levels = c("1-5", "6-25", "26-100", "100-500", "500-1000", "More than 1000")))
  ggplot(aes(fill=wellness_program, x=no_employees, y=1)) +
  geom_bar(position="fill", stat="identity") +
  labs(x="Company size (Number of employees)",
       y="percentage of people",
       title="Has your employer discussed mental health as part of an \nemployee wellness program?") +
  theme(legend.title = element_blank(), panel.grid.major = element_blank(), panel.grid.minor = element_blank()) +
  scale_fill_brewer(palette="GnBu")
```



Again, there is a general trend that larger companies are more likely to provide resources where employees can acquire more knowledge about their mental health concerns and how to find the help they need.

```
survey%>%
  mutate(no_employees = factor(no_employees, levels = c("1-5", "6-25", "26-100", "100-500", "500-1000", "More than 1000")))
ggplot(aes(fill=seek_help, x=no_employees, y=1)) +
  geom_bar(position="fill", stat="identity") +
  labs(x="Company size (Number of employees)",
       y="percentage of people",
       title="Does your employer provide resources to learn more about \nmental health issues and how to find the help they need?",
       theme(legend.title = element_blank(), panel.grid.major = element_blank(), panel.grid.minor = element_blank(),
       scale_fill_brewer(palette="GnBu"))
```

Does your employer provide resources to learn more about mental health issues and how to seek help?

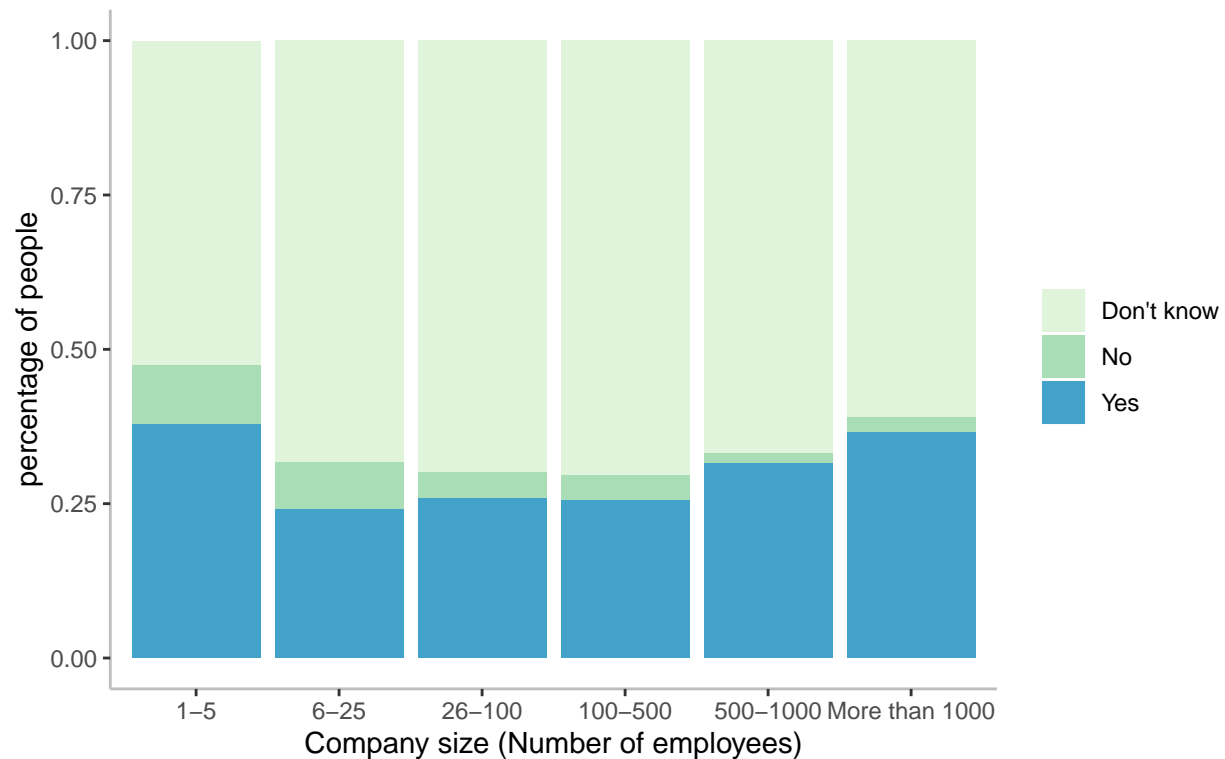


Another important factor that may explain the relationship between company size and mental health is anonymity.

Privacy is an important right we should respect, especially when it comes to physical and mental health issues, which might lead to potential disadvantages. Also, people with mental health conditions may be more sensitive about their privacy in their first place. **Here, it seems that the level of protection of anonymity is similar across companies of different sizes.** This may not be all dependent on the companies - it often relates to what services they collaborate with. Whether anonymity is well protected often has to rely on the care provider as well. Also, it is difficult to know for if your confidentials were leaked as they do not always appear “symptomatic.”

```
survey%>%
  mutate(no_employees = factor(no_employees, levels = c("1-5", "6-25", "26-100", "100-500", "500-1000", "More than 1000")))
  ggplot(aes(fill=anonymity, x=no_employees, y=1)) +
  geom_bar(position="fill", stat="identity") +
  labs(x="Company size (Number of employees)",
       y="percentage of people",
       title="Is your anonymity protected if you choose to take advantage of mental health \nor substance use services?")
  theme(legend.title = element_blank(), panel.grid.major = element_blank(), panel.grid.minor = element_blank())
  scale_fill_brewer(palette="GnBu")
```

Is your anonymity protected if you choose to take advantage of mental health or substance abuse treatment resources?



Part 3: Classification

In this part, we will use classification models – more specifically, decision trees and Naive Bayes models – to draw some conclusions for our hypotheses.

```
set.seed(1)
shuffled <- sample_n(survey, nrow(survey))
split <- 0.8 * nrow(shuffled)
train <- shuffled[1 : split, ]
test <- shuffled[(split + 1) : nrow(shuffled), ]

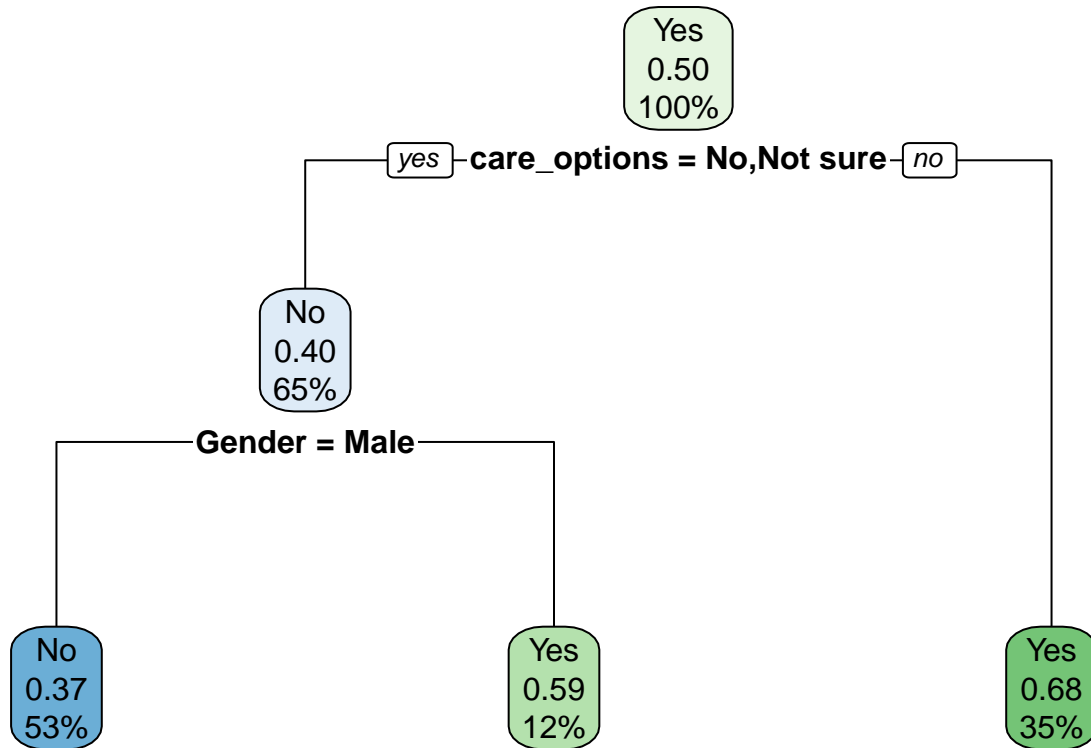
accuracy <- function(ground_truth, predictions) {
  mean(ground_truth == predictions)
}

# function to calculate accuracy of the model
accuracy <- function(ground_truth, predictions) {
  mean(ground_truth == predictions)
}
```

What are some variables that predict whether an employee has sought treatment?

Gender and Knowledge of options for mental health care?

```
treatment_tree <- rpart(treatment ~ Gender + care_options, data = train, method = "class")
rpart.plot(treatment_tree)
```



```
predict_treatment <- predict(treatment_tree, test, type = 'class')
accuracy(predict_treatment, survey$treatment)
```

```
## [1] 0.4828137
```

Perceiving negative consequences if discuss mental health with employer?

```
treatment_tree2 <- rpart(treatment ~ mental_health_consequence + Age, data = train, method = "class")
rpart.plot(treatment_tree2)
```

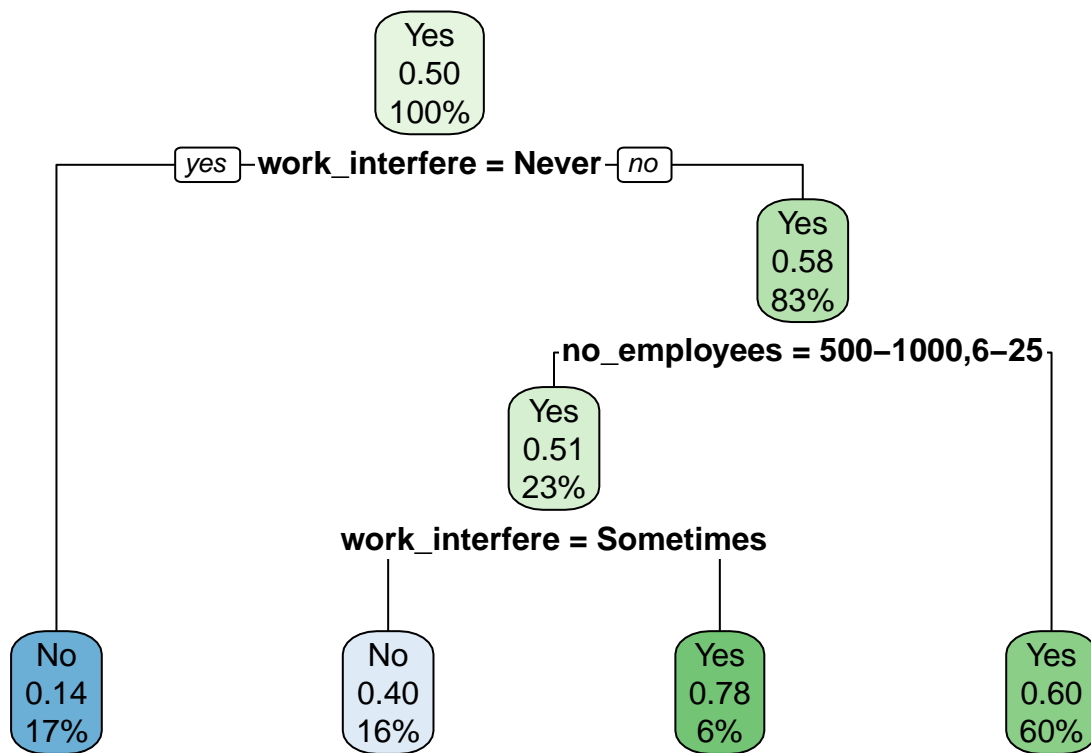


```
predict_treatment2 <- predict(treatment_tree2, test, type = 'class')
accuracy(predict_treatment2, survey$treatment)
```

```
## [1] 0.5091926
```

Work inteference and company size?

```
treatment_tree3 <- rpart(treatment ~ work_interfere + no_employees, data = train, method = "class")
rpart.plot(treatment_tree3)
```



```
predict_treatment3 <- predict(treatment_tree3, test, type = 'class')
accuracy(predict_treatment3, survey$treatment)
```

```
## [1] 0.4980016
```

As we can see from the results, the decision tree that predicts whether an employee has sought treatment based on gender and knowledge of care options gives us the best accuracy of about 52 percent.

Similarly we can build a Naive Bayes model. This time, we try to predict company size based on some variables regarding mental health care options and social taboos.

```
my_nb <- NaiveBayes(no_employees ~ benefits+
  care_options+
  wellness_program+
  seek_help+
  anonymity, data = train)
predict <- predict(my_nb, test)$class
mean(predict == test$no_employees) # Testing accuracy
```

```
## [1] 0.408
```

We see that the Naive Bayes model does not provide strong predictions, only yielding about a 40% accuracy.

As above, we can predict whether an employee has sought treatment with the variables in the dataset with about 50% accuracy.

Part 4: Maps

The last of our visualizations will feature maps as we explore geographical differences in mental health issues.

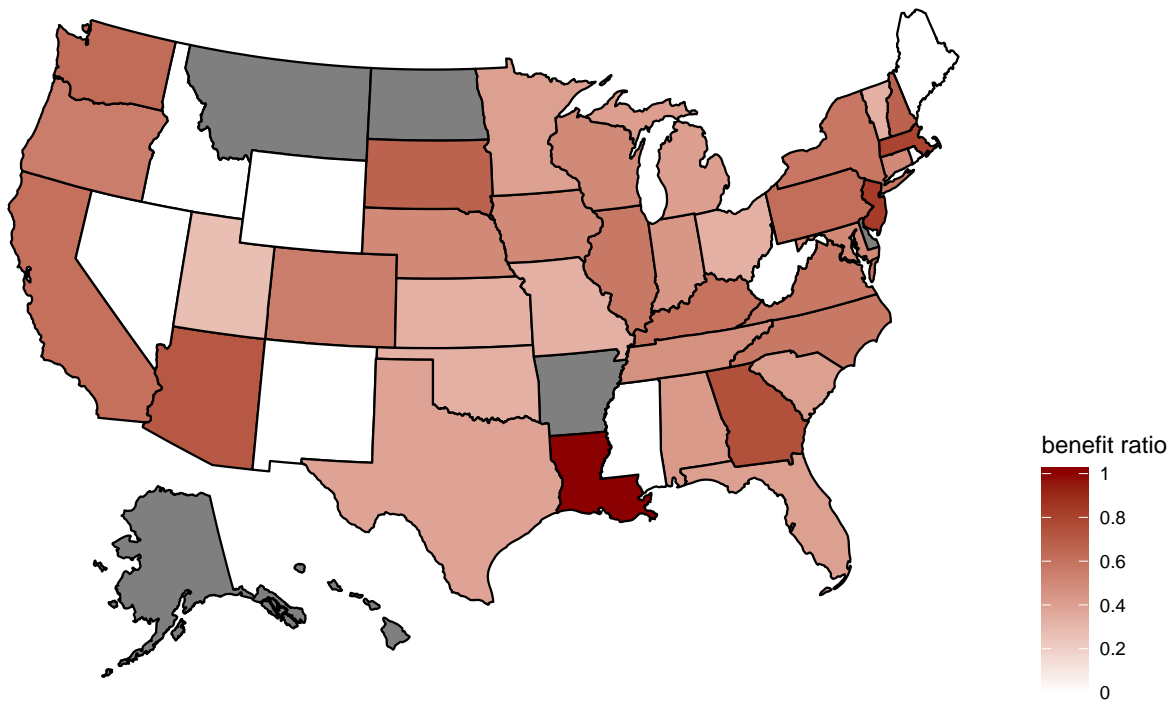
```
statebenefits <- survey %>% dplyr::select(state, benefits) %>% na.omit()
states <- unique (statebenefits$state)
benefitratio <- c()

for (i in 1:length(states)){
  totalstates <- nrow(statebenefits %>% filter(state==states[i]))
  totalyes <- nrow(statebenefits %>% filter(state==states[i] & benefits == "Yes"))
  newratio <- totalyes/totalstates
  benefitratio <- append(benefitratio, newratio)
}

states <- data.frame(states)
benefitratio <- data.frame(benefitratio)
states <- cbind(states, benefitratio)
states <- states %>%
  rename(state=states)

plot_usmap(data = states, values = "benefitratio") +
  scale_fill_continuous(name = "benefit ratio", low="white", high="darkred", breaks = c(0.0,0.2,0.4,0.6,0.8,1.0)) +
  theme(legend.position = "right") +
  labs(title="Does your company provide benefits? (based on ratio)"
```

Does your company provide benefits? (based on ratio)



Noticeably, benefit ratio is highest in LA, NJ, and MA. We have 1 person from LA, 6 from NJ, and 20 from MA included in the survey (not a huge sample size).

```
nrow(statebenefits %>% filter(state=="LA"))
```

```
## [1] 1
```

```
nrow(statebenefits %>% filter(state=="NJ"))
```

```
## [1] 6
```

```
nrow(statebenefits %>% filter(state=="MA"))
```

```
## [1] 20
```

Now, we explore how employees responded across the globe.

```
countrybenefits <- survey %>% dplyr::select(Country, benefits) %>% na.omit()
countries <- unique (countrybenefits$Country)
benefitratio2 <- c()

for (i in 1:length(countries)){
  totalcountries <- nrow(countrybenefits %>% filter(Country==countries[i]))
  totalyes <- nrow(countrybenefits %>% filter(Country==countries[i] & benefits == "Yes"))
  newratio <- totalyes/totalcountries
  benefitratio2 <- append(benefitratio2, newratio)
}

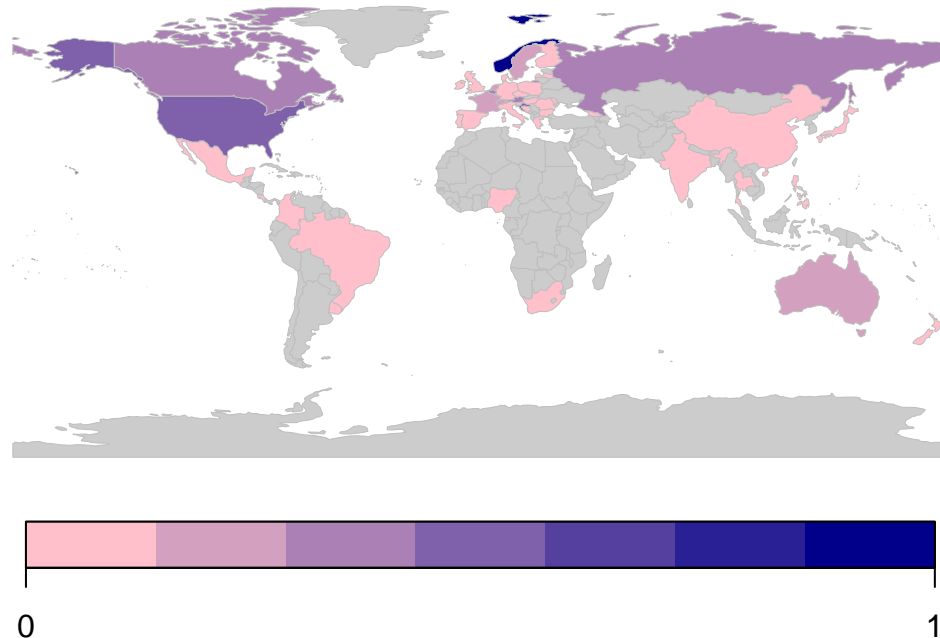
countries <- data.frame(countries)
benefitratio2 <- data.frame(benefitratio2)
countries <- countries %>%
  rename(country=countries)
countries <- cbind(countries, benefitratio2)

datamap <- joinCountryData2Map(countries, joinCode = "NAME",
  nameJoinColumn = "country")
```

```
## 46 codes from your data successfully matched countries in the map
## 0 codes from your data failed to match with a country code in the map
## 197 codes from the map weren't represented in your data
```

```
mapCountryData(datamap,
  nameColumnToPlot="benefitratio2",
  catMethod = "numerical",
  missingCountryCol = gray(.8),
  colourPalette=c("pink","darkblue"),
  mapTitle="Benefit ratio by country",
  addLegend = TRUE)
```

Benefit ratio by country



The ratio of which companies provide mental health benefits **vary a lot across countries**. Canada and some Northern European countries seem to be countries where companies are most likely to provide mental health benefits. Countries in Latin America, Africa, and parts of Asia and Europe are not doing as a good job in providing mental health benefits.

But still, keep in mind that our data set is relatively small and limited (it's about 1200 people, but they represent about 50 countries, which means not many people, and no many companies from each country).

Part 5: Hypothesis Testing

Finally, our last analytical method is hypothesis testing using t-tests! This will allow us to discern just how significant the differences are in our data.

Gender vs. Treatment

```
gender_vs_treatment <- survey %>%
  dplyr::select(Gender, treatment) %>%
  filter(Gender != "Queer")

gender_table <- with(gender_vs_treatment, table(Gender, treatment))
gender_table <- gender_table[c(16, 30), 1:2]
gender_table
```

```
##           treatment
## Gender    No Yes
##  Female   78 173
##   Male   540 450
```

```
gender_t_test <- fisher.test(gender_table)
gender_t_test
```

```
##
## Fisher's Exact Test for Count Data
##
## data:  gender_table
## p-value = 2.208e-11
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.2759170 0.5091636
## sample estimates:
## odds ratio
##  0.3760166
```

In this hypothesis test, we study the significance of the difference between males and females when it comes to seeking treatment for a mental health condition. Our null hypothesis states that the ratio of men seeking treatment is the same as the ratio of women seeking treatment. As we can see, our p-value has a magnitude of -11 , which means that we can reject the null hypothesis. This result suggests that there is a difference between the ratio of women seeking treatment and that of men.

Country vs. Benefits

```
country_table <- with(survey, table(Country, benefits))
country_table <- addmargins(country_table, FUN = list(Total = sum), quiet = TRUE)
country_table <- country_table[, 2:3]

north_america <- country_table[8,] + country_table[46,]
not_north_america <- country_table[49,] - north_america

country_table <- matrix(c(north_america, not_north_america), ncol = 2, byrow = TRUE)
colnames(country_table) <- c("No", "Yes")
rownames(country_table) <- c("North America", "Other")
country_table <- as.table(country_table)
country_table
```

```
##              No Yes
## North America 137 424
## Other         234  49
```

```
country_t_test <- fisher.test(country_table)
country_t_test
```

```
##
## Fisher's Exact Test for Count Data
##
## data:  country_table
## p-value < 2.2e-16
## alternative hypothesis: true odds ratio is not equal to 1
## 95 percent confidence interval:
##  0.04613097 0.09860126
## sample estimates:
## odds ratio
## 0.06793444
```

This hypothesis test looks at the difference between North America (which we consider here as only United States and Canada) and all other countries outside of North America when it comes to companies providing benefits. Our null hypothesis states that the ratio of North America providing benefits for its employees is equal to the ratio of other countries providing benefits for its employees. As with above, we see that the p-value is very low, which would suggest that we have significant statistical evidence to reject this null. Thus, this result suggests that there is a difference between the two ratios.

Conclusion

Back to Our General Hypotheses - How valid were they? What did we learn?

1. Company size is related to employee's willingness to discuss a mental health issue.

Answer: Not necessarily. We notice that in companies of various sizes, people generally find it difficult to discuss their mental health issues with supervisors. Sometimes it may be partly attributed to people's worry that such discussion will lead to negative consequences, or that they have seen a coworker with mental health conditions face some consequences.

However, people who don't worry about, or haven't observed these negative consequences, are still usually inclined not to discuss their mental health issues. In our understanding, it may be because as a society we often recognize mental health conditions as long-term, unsolvable barrier to work, and revealing our mental health issues to others will lead to disadvantages at work. Additionally, we also found that many people in the tech industry, are as unwilling to discuss their physical health issues as the mental health issues. We really should advocate for an environment that's more open to discussion around health conditions, and more inclusive to those who have a problem that needs medical interference.

2. Bigger companies provide better mental health care benefits/resources

Answer: Yes! In general, we found out that companies of larger sizes do better at providing benefits, including the benefits as part of their wellness program, and providing resources to learn more about mental health issues and how to seek help.

However, larger companies do not have any outstanding performance in terms of informing employees of their options for mental health care and protecting anonymity when an employee seeks mental health or substance abuse treatment resources.

3. We expect some geographical difference among mental health care benefits.

Answer: Yes, we have demonstrated how mental health benefits differ across countries and states. However, we cannot define any clear geographical pattern (e.g. companies in East coast states and West coast states don't differ significantly; neither do companies in European and Asian countries). Also, we do not have a dataset big enough to be truly representative of global trends.