Final Report Semester Project

# Loading the Data and Necessary Packages

knitr::opts\_chunk$set(comment = NA)  
library(dplyr)  
library(tidyr)  
library(ggplot2)  
library(grid)  
library(gridExtra)  
library(knitr)  
health <- read.csv("mentalsurvey\_clean.csv")

# Introduction

This document contains the fully commented analysis script and outputs, written summary of the project, and summary of contributorship to each part of the project.

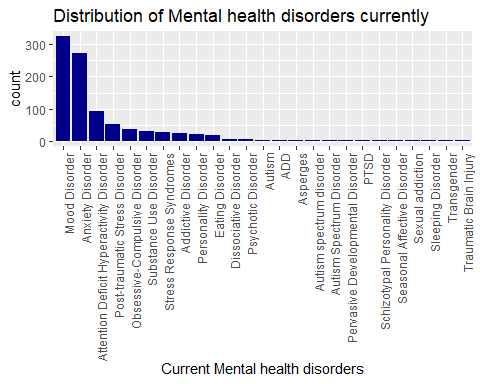
# Section 1: Distribution of mental health disorders based on different demographics

### Distribution of Mental health disorders

#since some people have multiple disorders, we have to split them   
tmp = health %>% separate(mhd,   
 sep = '\\|',  
 c('mhd\_1', 'mhd\_2', 'mhd\_3', 'mhd\_4', 'mhd\_5',  
 'mhd\_6', 'mhd\_7', 'mhd\_8', 'mhd\_9'),  
 fill = 'right')  
  
  
chk = tmp %>% select(matches('mhd\_[1-9]'))  
  
comb = as.vector(t(chk))  
comb1 = comb[comb != ""&!is.na(comb)]  
  
#Shortening the factor levels by splitting  
disorders=sapply(strsplit(comb1, split = " \\("), `[`, 1)  
dataf <- as.data.frame(matrix(disorders))  
colnames(dataf)<- "Disorder"  
#Removing inconsistent response  
subdataf <- subset(dataf, disorders != "I haven\'t been formally diagnosed, so I felt uncomfortable answering, but Social Anxiety and Depression.")  
count\_disorder <-data.frame(count(subdataf, Disorder, sort = TRUE))  
kable(count\_disorder)

|  |  |
| --- | --- |
| Disorder | n |
| Mood Disorder | 325 |
| Anxiety Disorder | 272 |
| Attention Deficit Hyperactivity Disorder | 93 |
| Post-traumatic Stress Disorder | 51 |
| Obsessive-Compulsive Disorder | 36 |
| Substance Use Disorder | 31 |
| Stress Response Syndromes | 28 |
| Addictive Disorder | 24 |
| Personality Disorder | 22 |
| Eating Disorder | 19 |
| Dissociative Disorder | 5 |
| Psychotic Disorder | 4 |
| Autism | 2 |
| ADD | 1 |
| Asperges | 1 |
| Autism spectrum disorder | 1 |
| Autism Spectrum Disorder | 1 |
| Pervasive Developmental Disorder | 1 |
| PTSD | 1 |
| Schizotypal Personality Disorder | 1 |
| Seasonal Affective Disorder | 1 |
| Sexual addiction | 1 |
| Sleeping Disorder | 1 |
| Transgender | 1 |
| Traumatic Brain Injury | 1 |

reorder\_size <- function(x) {  
 factor(x, levels = names(sort(table(x), decreasing = TRUE)))  
}  
#Plot to see the distribution of current mental health illness  
p <- ggplot(data=subdataf, aes(x = reorder\_size(Disorder)))+  
 geom\_bar(fill='dark blue')+  
 theme(axis.text.x = element\_text(angle = 90, hjust = 1))  
  
p+xlab("Current Mental health disorders")+ggtitle("Distribution of Mental health disorders currently")



#### Interpretation of Results

The most common mental health disorder appears to be mood disorder with a count of 325, followed by anxiety disorder with a count of 272.

### Distribution of mental health status among gender

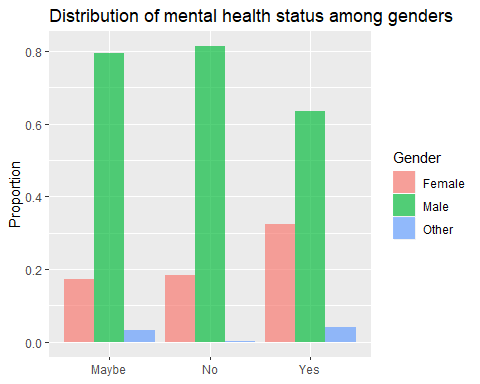
gender\_disorder <- table(health$Do.you.currently.have.a.mental.health.disorder., health$gender)  
gender\_disorder

Female Male Other  
 Maybe 44 202 8  
 No 81 359 1  
 Yes 146 286 19

#Table: % of who may, don't or have mental disorder among genders  
rowprop <- prop.table(gender\_disorder,1)   
rowprop

Female Male Other  
 Maybe 0.173228346 0.795275591 0.031496063  
 No 0.183673469 0.814058957 0.002267574  
 Yes 0.323725055 0.634146341 0.042128603

dr1 <-as.data.frame(rowprop)  
# Distribution of respondents among gender provided they may/have/donot have a mental diorder   
ggplot(data = dr1, aes(x =Var1 , y = Freq, fill = Var2)) +   
 geom\_bar(stat = 'identity', position = 'dodge', alpha = 2/3) +   
 labs(x = NULL, y = 'Proportion', fill = 'Gender')+ggtitle('Distribution of mental health status among genders')



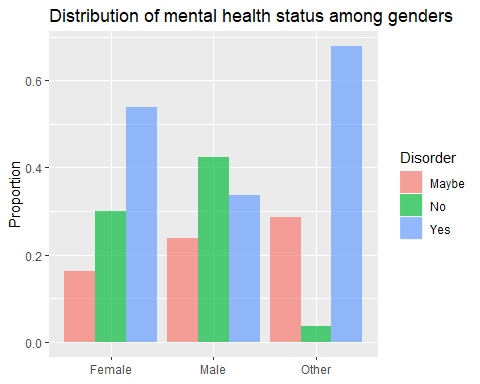
#### Interpretation of results

The table and figure above shows that among the people that said Yes, 32.3% are females, 63.4% are males and 4.2% are otherss. Among the people that said No, 81.4% are males, 18.3% are females and 0.2% are others.Among the people that said Maybe, 79.5% are males, 17.3% are females and 3.1% are others.

#Table : % of each level gender who may, don't or have mental disorder  
colprop <- prop.table(gender\_disorder,2)  
colprop

Female Male Other  
 Maybe 0.16236162 0.23848878 0.28571429  
 No 0.29889299 0.42384888 0.03571429  
 Yes 0.53874539 0.33766234 0.67857143

# Distribution of respondents who may/have/donot have a mental diorder over the gender  
dr2 <-as.data.frame(colprop)  
ggplot(data = dr2, aes(x =Var2 , y = Freq, fill = Var1)) +   
 geom\_bar(stat = 'identity', position = 'dodge', alpha = 2/3) +   
 labs(x = NULL, y = 'Proportion', fill = 'Disorder')+ggtitle('Distribution of mental health status among genders')



#### Interpretation of results

The table and figure above show that the proportion of females who said Yes (0.53) are more than the proportion of females that said no (0.29) amd maybe (0.16). The proportion of males that said No (0.42) are more than the proportion of males who said Yes (0.33) and Maybe (0.23). In the gender Other, the proportion that said Yes (0.67) are more than the proportion that said No (0.03) and Maybe (0.28).

### Distribution of mental health disorder based on age groups

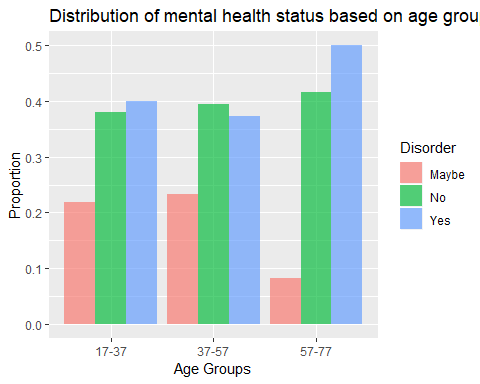
summary(health$age)

Min. 1st Qu. Median Mean 3rd Qu. Max.   
 17.00 28.00 32.00 33.37 38.00 74.00

#Creating labels for age groups  
labs <- c(paste(seq(17, 74, by = 20), seq(37, 80, by = 20),  
 sep = "-"))  
health$AgeGroup <- cut(health$age, breaks = c(seq(17, 74, by = 20), Inf), labels = labs, right = FALSE)  
  
agegr <- table(health$Do.you.currently.have.a.mental.health.disorder.,health$AgeGroup)  
ageprop <- prop.table(agegr,2)  
ageprop

17-37 37-57 57-77  
 Maybe 0.21933086 0.23241590 0.08333333  
 No 0.38042131 0.39449541 0.41666667  
 Yes 0.40024783 0.37308869 0.50000000

dfageprop <- as.data.frame(ageprop)  
  
#Distribution of respondents reporting a mental disorder by age groups  
ggplot(data = dfageprop, aes(x =Var2 , y = Freq, fill = Var1)) +   
 geom\_bar(stat = 'identity', position = 'dodge', alpha = 2/3) +   
 labs(x = "Age Groups", y = 'Proportion', fill = 'Disorder')+ggtitle('Distribution of mental health status based on age groups')



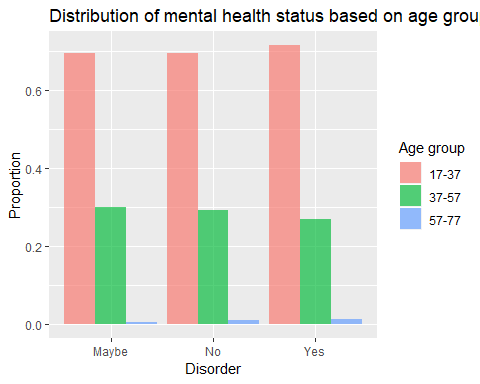
#### Interpretation of results

The table and plot above show that in the age group of 17-37, 40% said they currently have a mental disorder, 38% said they donot have a mental disorder currently and 21.9% said are not sure. In the 37-57 age group, 39.4% said they donot have a mental disorder currently, 37.3% said Yes and 23.2% said maybe or are not sure. In the 57-77 agegroup, 50% said Yes, 41.6% said No and 8.3% said maybe.

ageprop1 <- prop.table(agegr,1)  
ageprop1

17-37 37-57 57-77  
 Maybe 0.696850394 0.299212598 0.003937008  
 No 0.696145125 0.292517007 0.011337868  
 Yes 0.716186253 0.270509978 0.013303769

dfageprop1 <- as.data.frame(ageprop1)  
ggplot(data = dfageprop1, aes(x =Var1 , y = Freq, fill = Var2)) +   
 geom\_bar(stat = 'identity', position = 'dodge', alpha = 2/3) +   
 labs(x = "Disorder", y = 'Proportion', fill = 'Age group')+ggtitle('Distribution of mental health status based on age groups')



#### Interepretion of results

The table and plot above show that among the people that said they have mental disorder currently, 71.6% lie in the 17-37 age group, 27% lie in the 37-57 age group and 1.3% lie in the age group of 57-77. Among the people that said No, 69.6% lie in the 17-37 age group, 29.2% lie in the 37-57 age group and 1.1% lie in the 57-77 group. Among the people that said Maybe, 69.6% liein the 17-37 age group, 29.9% lie in the 37-57 age group and .3% lie in the 57-77 age group.

### Distribution of mental health disorder based on geography

table(health$What.US.state.or.territory.do.you.live.in. , health$Do.you.currently.have.a.mental.health.disorder.)

Maybe No Yes  
 105 209 123  
 Alabama 0 1 3  
 Alaska 1 1 0  
 Arizona 0 4 0  
 California 22 35 53  
 Colorado 7 6 8  
 Connecticut 0 2 1  
 Delaware 0 0 1  
 District of Columbia 0 0 2  
 Florida 4 7 6  
 Georgia 5 3 3  
 Idaho 0 0 3  
 Illinois 5 18 24  
 Indiana 3 7 11  
 Iowa 0 1 4  
 Kansas 3 7 4  
 Kentucky 1 0 1  
 Louisiana 0 0 2  
 Maine 1 1 3  
 Maryland 4 6 3  
 Massachusetts 5 8 6  
 Michigan 11 15 17  
 Minnesota 6 13 19  
 Missouri 1 6 3  
 Montana 0 1 0  
 Nebraska 1 5 4  
 Nevada 1 1 0  
 New Hampshire 1 2 2  
 New Jersey 1 0 4  
 New Mexico 0 0 1  
 New York 7 9 19  
 North Carolina 4 5 11  
 North Dakota 2 0 1  
 Ohio 6 5 10  
 Oklahoma 1 7 4  
 Oregon 4 11 13  
 Pennsylvania 4 8 18  
 Rhode Island 0 2 1  
 South Carolina 0 0 1  
 South Dakota 1 1 1  
 Tennessee 5 9 13  
 Texas 12 7 18  
 Utah 2 0 4  
 Vermont 0 3 0  
 Virginia 6 3 6  
 Washington 10 11 14  
 West Virginia 1 0 0  
 Wisconsin 1 1 6

#### Interpretation of results

From the above table output, we can see that the US state with the majority of mental health illness cases is California, followed by Illoinois and New York

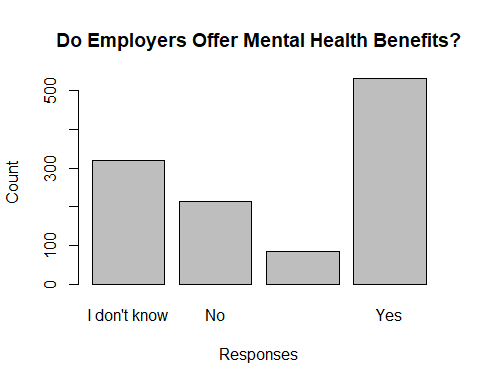
# Section 2: Analysis of distribution of mental health benefits offered by companies in the dataset

### Distribution of employers offering mental healh benefits as part of healthcare coverage

#Creating a table for the variable "Does your employer provide mental health benefits as part of healthcare coverage?""  
  
table(health$Does.your.employer.provide.mental.health.benefits.as.part.of.healthcare.coverage.)

I don't know No   
 319 213   
Not eligible for coverage / N/A Yes   
 83 531

barplot(table(health$Does.your.employer.provide.mental.health.benefits.as.part.of.healthcare.coverage.), xlab = "Responses", ylab = "Count", main = "Do Employers Offer Mental Health Benefits?")



#### Interpretation of results

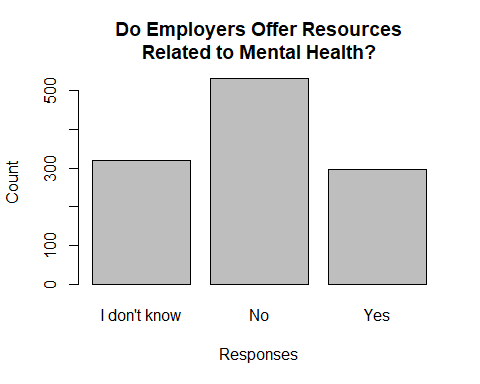
From the above output, we can see that out of a total 1433 companies, only 531 provide mental health benefits as part of healthcare coverage, which is approximately one-third of the responses.

### Distribution of employers offering resources to learn more about mental health concerns and options for seeking help

table(health$Does.your.employer.offer.resources.to.learn.more.about.mental.health.concerns.and.options.for.seeking.help.)

I don't know No Yes   
 320 531 295

barplot(table(health$Does.your.employer.offer.resources.to.learn.more.about.mental.health.concerns.and.options.for.seeking.help.), xlab = "Responses", ylab = "Count", main = "Do Employers Offer Resources\nRelated to Mental Health?")



#### Interpretation of results

From the above output, we can see that a very small number of companies from the dataset of 1433 entries offer resources to employees to learn more about mental health concerns and options for seeking help.i.e., only 295 employers. For the dataset under study, this implies that a very small number of employers are taking steps to aquaint their employees with mental health benefits in the respective companies. Also, we can see that a significant number of employees, i.e., 320 out of 1433, do not know if their employers offer resporces to seek help regading mental health concerns. This also implies lack of initiative in generating awareness about mental health-related policies.

### Do Companies in the survey dataset have formal discussions to Assist workers with Mental and Behavioral Health Policies

For companies with mental health care options available, let us find out what percentage have a formal discussion (as part of a campaign/official communication) with employees to generate awareness about the mental health care policies. Here, we can create subset for companies with mental health care options available. This dataset(subset) will contain only those companies that provide mental health benefits as part of healthcare coverage

Company<-subset(health, Does.your.employer.provide.mental.health.benefits.as.part.of.healthcare.coverage.=="Yes")

#### Percentage of employers who have a formal discussion about mental health with their employees

#Creating a table for for the variable related to employers formally discussing mental health with employees  
table(Company$Has.your.employer.ever.formally.discussed.mental.health..for.example..as.part.of.a.wellness.campaign.or.other.official.communication..)

I don't know No Yes   
 59 331 141

#Computing the percentage  
percent<-table(Company$Has.your.employer.ever.formally.discussed.mental.health..for.example..as.part.of.a.wellness.campaign.or.other.official.communication..)/length(Company$Has.your.employer.ever.formally.discussed.mental.health..for.example..as.part.of.a.wellness.campaign.or.other.official.communication..)  
#The parameter percent gives us the value of the percenntage needed  
percent

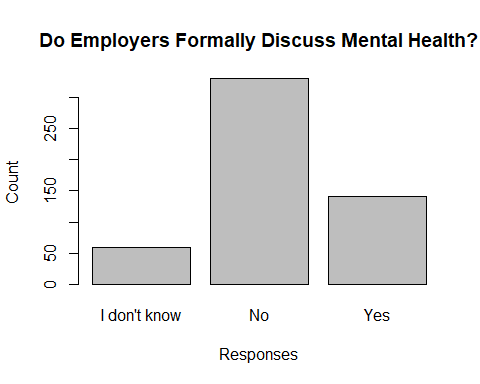
I don't know No Yes   
 0.1111111 0.6233522 0.2655367

#### Interpretation of results

The value of the parameter percent in the above output for the column “Yes” gives us the percentage of employers who have formally discussed mental health with their employees (considering only those companies who provide mental health benefits to their employees). Thus, about 26.5 percent of employers hve a formal discussion about mental health and related policies with their employees. For the dataset under study, this is a very small number. In order to make the employees familiar with the mental health policies and benefits, it is crucial that the employers have a formal discussion with them.

#### Visualizing the proportion of responses for employers formally discussing mental health with theri employees

Company<-droplevels(Company)  
barplot(table(Company$Has.your.employer.ever.formally.discussed.mental.health..for.example..as.part.of.a.wellness.campaign.or.other.official.communication..), xlab = "Responses", ylab = "Count", main = "Do Employers Formally Discuss Mental Health?")



# Section 3: What are Factors that Might Limit the Impact of Behavioral Health Outreach

### Potential Negative Consequences for discussing Health Issues with Employer: Physical vs. Mental

If employees feel that discussing physical or mental health issues would have negative consequences. If so, they may be less likely to reach out for help.

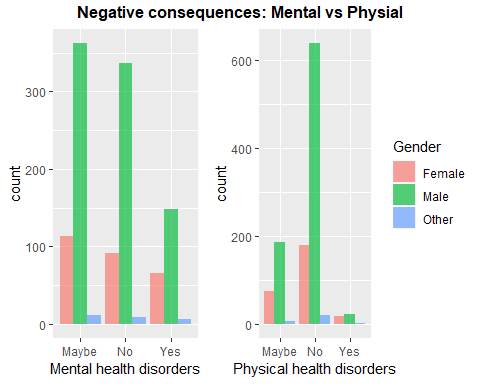
knitr::opts\_chunk$set(echo = TRUE)  
#Creating a table with gender and mental health responses  
tab <- table(health$Do.you.think.that.discussing.a.mental.health.disorder.with.your.employer.would.have.negative.consequences., health$gender)  
tab

Female Male Other  
 Maybe 113 362 12  
 No 92 337 9  
 Yes 66 148 7

#Creating a table with gender and physical health responses  
tab1 <- table(health$Do.you.think.that.discussing.a.physical.health.issue.with.your.employer.would.have.negative.consequences., health$gender)  
tab1

Female Male Other  
 Maybe 75 187 6  
 No 179 638 20  
 Yes 17 22 2

#creating the combined plot  
tabdf <- as.data.frame(tab)  
tab1df <- as.data.frame(tab1)  
  
plot1 <- ggplot(data = tabdf, aes(x = Var1, y =Freq, fill = Var2)) +   
 geom\_bar(stat = 'identity', position = 'dodge', alpha = 2/3) +   
 labs(x = "Mental health disorders", y = 'count')+ theme(legend.position='none')  
  
plot2 <- ggplot(data = tab1df, aes(x = Var1, y =Freq, fill = Var2)) +   
 geom\_bar(stat = 'identity', position = 'dodge', alpha = 2/3) +   
 labs(x = "Physical health disorders", y = 'count', fill = 'Gender')  
title1=textGrob("Negative consequences: Mental vs Physial", gp=gpar(fontface="bold"))  
grid.arrange(plot1, plot2,ncol=2,top = title1, widths = c(3/4,1))



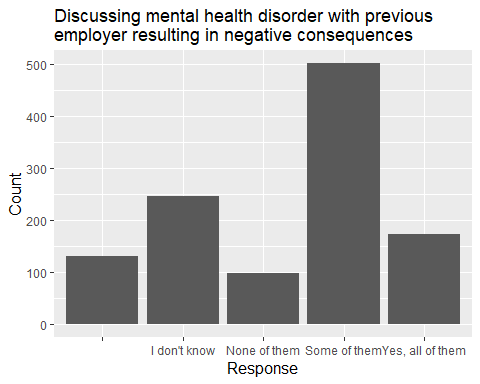
#### Interpretation of Results

The figure above shows that majority of people are certain that discussing physical health issues with their employers will not result in negative consequences. Very few believe that discussing physical health issues will have negative consequences. The responses to whether discussing mental health disorders with employers will result in negative consequences is more divided. Some people believe that discussing mental health disorders will have a negative consequence. This is probably due to the stigma at workplace which may make them believe that they might face negative consequences if their mental issues are brought infront of a employer. This will make them less likely to reach out for help, discuss their issues with their employers, take time off work. However, there are still more people who believe there won’t be any negative consequences. This could be because they have already developed a smooth employer-employee relationship. There are comparatively more number of people who are unsure of whether or not there would be any negative consequences if mental health disorders are discussed with their employers.

### Analzing the relationship between perceiving the negative impact of mental illness at previous workplace and the impact at current workplace

It is likely that if employees perceived a negative impact from discussing mental illness at a previous workplace, they may have similar perception at the current workplace as well. We can plot the ggplot for the independent variable “Do you think that discussing a mental health disorder with previous employers would have negative consequences?”

ggplot(health, aes(x = health$Do.you.think.that.discussing.a.mental.health.disorder.with.previous.employers.would.have.negative.consequences.)) + geom\_bar() + ggtitle("Discussing mental health disorder with previous\nemployer resulting in negative consequences") + xlab("Response") + ylab("Count") + theme(axis.title=element\_text(size=12))



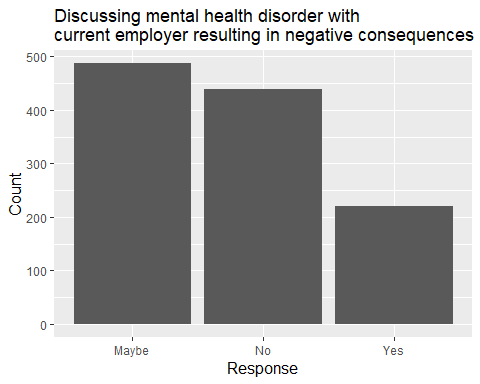
summary(health$Do.you.think.that.discussing.a.mental.health.disorder.with.previous.employers.would.have.negative.consequences.)

I don't know None of them Some of them   
 131 245 97 501   
Yes, all of them   
 172

From the above ggplot, we can observe that a majority of respondents(615) feel that discussing mental health disorders with some of their previous employers resulted in negative consequences for them. While a significant number of respondents responded that they do not know, we can conclude from the graph that the majority did experience negative consequences after discussing mental health disorders with their previous employers

Plotting the ggplot for the dependent variable “Do you think that discussing a mental health disorder with your employer would have negative consequences?”

ggplot(health, aes(x = health$Do.you.think.that.discussing.a.mental.health.disorder.with.your.employer.would.have.negative.consequences.)) + geom\_bar() + ggtitle("Discussing mental health disorder with\ncurrent employer resulting in negative consequences") + xlab("Response") + ylab("Count") + theme(axis.title=element\_text(size=12))



summary(health$Do.you.think.that.discussing.a.mental.health.disorder.with.your.employer.would.have.negative.consequences.)

Maybe No Yes   
 487 438 221

From the above ggplot, we can see that a vast majority of respondents(487) are not sure if discussing mental health disorders with their current employers would have negative consequences; followed by about 438 respondents who do not think that it could lead to negative consequences.

#### Chi-squared test

m1<-table(health$Do.you.think.that.discussing.a.mental.health.disorder.with.previous.employers.would.have.negative.consequences., health$Do.you.think.that.discussing.a.mental.health.disorder.with.your.employer.would.have.negative.consequences.)  
chisq.test(m1)

Pearson's Chi-squared test  
  
data: m1  
X-squared = 158.59, df = 8, p-value < 2.2e-16

#### Interpretation of results

We can see from the above chi-squared test that the p-value is very small, meaning we can reject the hypothesis of independence. Thus there is a relationship between perceiving the negative impact of mental illness at previous workplace and perceiving the same at the current workplace. Employees who have perceived or experienced a negative impact of discussing a mental illness with their previous employer feel the same way about discussing it with their current employer. This type of attitude may be the result of some form of negative impact such as bias, discrimination or lack of oppurtunities that they faced at the previous workplace. This might have influenced their decision to discusstheir mental illnesses with their current employer as well.

## Anonymity Policies

This part of the analysis looks at answers to the question “Is your anonymity protected if you choose to take advantage of mental health or substance abuse treatment resources provided by your employer?” by reported company size and whether the respondent’s company is primarily a tech company. This is is important because, if employees feel that their anonymity is not protected, they may be less likely to reach out for help.

### Anonymity Policies by Company Size

#Put company size in order  
health$CompanySize = factor(health$How.many.employees.does.your.company.or.organization.have, levels=c("1-5", "6-25", "26-100", "100-500", "500-1000", "More than 1000"))  
#Modify Labels for Anonymity Policies  
health$AnonPolicies2 <- factor(health$Is.your.anonymity.protected.if.you.choose.to.take.advantage.of.mental.health.or.substance.abuse.treatment.resources.provided.by.your.employer.,  
levels = c("I don't know", "No", "Yes"),  
labels = c("Don't Know", "No", "Yes"))  
#Create subset  
AnoymityPolicies\_By\_CompanySize <- table(health$AnonPolicies2, health$CompanySize)

#### Crosstab Table

#Print Cross-tab table  
kable(AnoymityPolicies\_By\_CompanySize)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 1-5 | 6-25 | 26-100 | 100-500 | 500-1000 | More than 1000 |
| Don’t Know | 31 | 133 | 196 | 166 | 56 | 160 |
| No | 20 | 24 | 18 | 8 | 2 | 12 |
| Yes | 9 | 53 | 78 | 74 | 22 | 84 |

#### Proportions Table

#Create a proportions table by Row  
kable(prop.table(AnoymityPolicies\_By\_CompanySize, 2))

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 1-5 | 6-25 | 26-100 | 100-500 | 500-1000 | More than 1000 |
| Don’t Know | 0.5166667 | 0.6333333 | 0.6712329 | 0.6693548 | 0.700 | 0.625000 |
| No | 0.3333333 | 0.1142857 | 0.0616438 | 0.0322581 | 0.025 | 0.046875 |
| Yes | 0.1500000 | 0.2523810 | 0.2671233 | 0.2983871 | 0.275 | 0.328125 |

#### Chi-Squared Test

#Do a chi-squared test  
chisq.test(AnoymityPolicies\_By\_CompanySize)

Warning in chisq.test(AnoymityPolicies\_By\_CompanySize): Chi-squared  
approximation may be incorrect

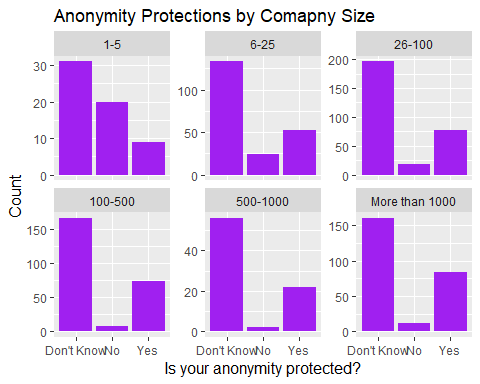
Pearson's Chi-squared test  
  
data: AnoymityPolicies\_By\_CompanySize  
X-squared = 80.824, df = 10, p-value = 3.461e-13

Our null hypothesis is that there is that company size and anonymity policies are independent. Our alternative hypothesis is that there is a relationship between these two variables. Because the sensitivity of this test to larger sample sizes we are setting the p value to .01 (Townsley, 2002)

We get a warning that the approximation may be incorrect, likely because some of the values from the table are less than 5 (Townsley, 2002). Therefore, we cannot draw any conclusions for this test about whether these two variables are related.

#### Visualizing the Data

#Create visualization  
AnoymityPolicies\_By\_CompanySize\_Viz <- ggplot(health, aes(x = health$AnonPolicies2))+geom\_bar(fill="purple")+facet\_wrap(~health$CompanySize, scales="free\_y") + ggtitle("Anonymity Protections by Comapny Size") + xlab("Is your anonymity protected?") + ylab("Count")+ theme(axis.title=element\_text(size=12))  
AnoymityPolicies\_By\_CompanySize\_Viz



#### Interpretation of Results

The table and chart above show that across all company sizes, most respondents indicate that they do not know what their current company’s anonymity policy is towards those using mental health services. For companies of 1-5 employees, the next most common answer is “No” (anonymity is not protected). However, for companies larger than 5 employees, the results appear fairly consistent, with the majority of respondents (60-70 percent) indicating that they do not know anonymity policies, a sizable minority of respondents saying that anonymity is protected (25-32 percent), and less than 10 percent of respondents in each group saying anonymity is not protected. Based off of the chi-squared test, it is not clear if there is a statistically significant difference in anonymity policies of companies of different sizes. However, the fact that the majority of respondents across company sizes do not know their company’s anonymity policy is something that has to be addressed in any outreach program, as a study by Milne, et al (1994) has found that participants tended to be more confident in a company’s Employee Assistance Program if they felt it was confidential (p. 141). Therefore, any outreach must make the company’s anonymity policy clear.

### Anonymity Policies by Whether it is a Tech Company

#Add labels to tech company org question  
health$TechCompany2 <- factor(health$Is.your.employer.primarily.a.tech.company.organization.,  
levels = c(0, 1),  
labels = c("Not primarily tech", "Primarily Tech"))  
#Create a subset  
AnonymityPolicies\_by\_TechCompany <- table(health$Is.your.anonymity.protected.if.you.choose.to.take.advantage.of.mental.health.or.substance.abuse.treatment.resources.provided.by.your.employer., health$TechCompany2)

#### Crosstab Table

#Print Cross-tab table  
kable(AnonymityPolicies\_by\_TechCompany)

|  |  |  |
| --- | --- | --- |
|  | Not primarily tech | Primarily Tech |
| I don’t know | 169 | 573 |
| No | 15 | 69 |
| Yes | 79 | 241 |

#### Proportions table

#Create a proportions table by Row  
kable(prop.table(AnonymityPolicies\_by\_TechCompany, 2))

|  |  |  |
| --- | --- | --- |
|  | Not primarily tech | Primarily Tech |
| I don’t know | 0.6425856 | 0.6489241 |
| No | 0.0570342 | 0.0781427 |
| Yes | 0.3003802 | 0.2729332 |

#### Chi-Squared Test

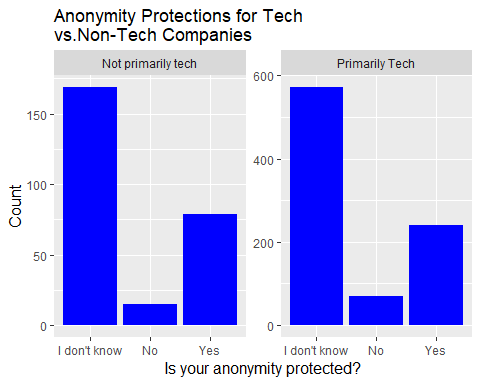
#Do a chi-squared test  
chisq.test(AnonymityPolicies\_by\_TechCompany)

Pearson's Chi-squared test  
  
data: AnonymityPolicies\_by\_TechCompany  
X-squared = 1.7911, df = 2, p-value = 0.4084

Our null hypothesis is that there is that company size and whether it is primarily a tech company are independent. Our alternative hypothesis is that there is a relationship between these two variables. Because the sensitivity of this test to larger sample sizes we are setting the p value to .01 (Townsley, 2002). This chi-squared test’s p value is very large. Therefore, we fail to reject the null.

#### Visualizing the Data

#Create Visualization  
AnonymityPolicies\_by\_TechCompany\_viz <- ggplot(health, aes(x = health$Is.your.anonymity.protected.if.you.choose.to.take.advantage.of.mental.health.or.substance.abuse.treatment.resources.provided.by.your.employer.))+geom\_bar(fill="blue")+facet\_wrap(health$TechCompany2 ~ ., scales="free\_y") + ggtitle("Anonymity Protections for Tech\nvs.Non-Tech Companies") + xlab("Is your anonymity protected?") + ylab("Count") + theme(axis.title=element\_text(size=12))  
AnonymityPolicies\_by\_TechCompany\_viz



#### Interpretation of Results

The difference between primarily tech and non-primarily tech companies is not statistically significant in regards to anonymity policies for those seeking mental health services. However, this may be influenced by the fact that the majority of those surveyed work for primiarily tech-oriented companies, so we do not have a large sample of those working for non-tech companies. Furthermore, we do not know the industry that these non-tech companies are in, which means we cannot draw any firm conclusions here. Nevertheless, in both cases, the clear majority of those surveyed indicated that they do not know their current company’s anonymity policy is towards those using mental health services. In both cases, the next most common response to the question is that anonymity is protected for those using mental health services.

## Ease of Asking for Leave For A Mental Health Issue

This part of the analysis looks at answers to the question “If a mental health issue prompted you to request a medical leave from work asking for that leave would be” by respondent demographics as well as their reports of several company characteristics. This is important because if employees feel that asking for time off from work for medical leave, they may be less likely to seek the help they need.

### Answers by Gender

#Modify Labels for Perceived Leave Difficulty QUestion (ideal if we can get it to fit)  
health$perceivedleavediff1 <- factor(health$If.a.mental.health.issue.prompted.you.to.request.a.medical.leave.from.work..asking.for.that.leave.would.be.,  
levels = c("Very easy", "Somewhat easy", "Neither easy nor difficult","Somewhat difficult", "Very difficult", "I don't know"),  
labels = c("Very easy (VE)", "Somewhat Easy (SE)", "Neutral (N)", "Somewhat Hard (SH)", "Very Hard (VH)", "Don't Know (DK)"))  
#Modify Labels for Perceived Leave Difficulty Question (if health$perceivedleavediff1 does not fit, use this as an alternate)  
health$perceivedleavediff2 <- factor(health$If.a.mental.health.issue.prompted.you.to.request.a.medical.leave.from.work..asking.for.that.leave.would.be.,  
levels = c("Very easy", "Somewhat easy", "Neither easy nor difficult","Somewhat difficult", "Very difficult", "I don't know"),  
labels = c("VE", "SE", "N", "SH", "VH", "DK"))  
#Create a subset  
LeaveDiffbyGender <- table(health$perceivedleavediff1, health$gender)

#### Crosstab Table

#Print Crosstab  
kable(LeaveDiffbyGender)

|  |  |  |  |
| --- | --- | --- | --- |
|  | Female | Male | Other |
| Very easy (VE) | 48 | 168 | 4 |
| Somewhat Easy (SE) | 60 | 213 | 8 |
| Neutral (N) | 38 | 137 | 3 |
| Somewhat Hard (SH) | 58 | 136 | 5 |
| Very Hard (VH) | 39 | 76 | 3 |
| Don’t Know (DK) | 28 | 117 | 5 |

#### Proportions Table

#Create Proportions table  
kable(prop.table(LeaveDiffbyGender, 2))

|  |  |  |  |
| --- | --- | --- | --- |
|  | Female | Male | Other |
| Very easy (VE) | 0.1771218 | 0.1983471 | 0.1428571 |
| Somewhat Easy (SE) | 0.2214022 | 0.2514758 | 0.2857143 |
| Neutral (N) | 0.1402214 | 0.1617473 | 0.1071429 |
| Somewhat Hard (SH) | 0.2140221 | 0.1605667 | 0.1785714 |
| Very Hard (VH) | 0.1439114 | 0.0897285 | 0.1071429 |
| Don’t Know (DK) | 0.1033210 | 0.1381346 | 0.1785714 |

#### Chi-Squared

#Run chi-squared test  
chisq.test(LeaveDiffbyGender)

Warning in chisq.test(LeaveDiffbyGender): Chi-squared approximation may be  
incorrect

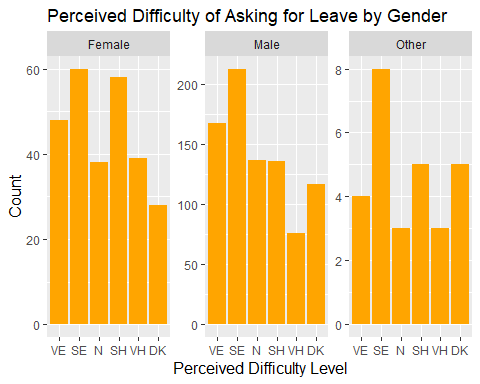
Pearson's Chi-squared test  
  
data: LeaveDiffbyGender  
X-squared = 14.474, df = 10, p-value = 0.1524

Our null hypothesis is that there is that gender and perceived difficulty of requesting leave are independent. Our alternative hypothesis is that there is a relationship between these two variables. Because the sensitivity of this test to larger sample sizes we are setting the p value to .01 (Townsley, 2002)

We get a warning that the approximation may be incorrect, likely because some of the values from the table are less than 5 (Townsley, 2002). However, the large p value potentially indicates that that the differences between genders may not be statistically significant (or we do not have enough information because of the small number of people of other genders in our dataset).

#### Visualizing the Data

#Create Visualization  
LeaveDiffbyGender\_Viz <- ggplot(health, aes(x = health$perceivedleavediff2))+geom\_bar(fill="orange")+facet\_wrap(health$gender ~ ., scales="free\_y") + ggtitle("Perceived Difficulty of Asking for Leave by Gender") + xlab("Perceived Difficulty Level") + ylab("Count") + theme(axis.title=element\_text(size=12))  
LeaveDiffbyGender\_Viz



NOTE: VE = Very Easy, SE = Somewhat Easy, N = Neutral (or Neither Easy nor Difficult), SH = Somewhat Hard, VH = Very Hard, DK = Don’t Know.

#### Interpretation of Results

Across all genders, the most common answer is “Somewhat Easy”. Interestingly, for women the second most common answer is “Somewhat Difficult”, while for men the second most common answer is “Very Easy.” For those indicating their gender falls into another category, an equal amount indicated that they felt asking for leave was very hard or provided a neutral response (“neither easy or hard”). Across all genders there is a sizable group of people indicating they find it somewhat to very hard to ask for leave, which indicates that perceived difficulty may vary by company. That being said, there are fewer women and those from other genders participating in this survey (compared to men). This means that we may not have a complete picture of the opinions of women and those who state their gender as other than male or female.

### Answers by Age Group

#Do the AgeGroup Variable  
labs <- c(paste(seq(17, 74, by = 20), seq(37, 80, by = 20),  
 sep = "-"))  
health$AgeGroup <- cut(health$age, breaks = c(seq(17, 74, by = 20), Inf), labels = labs, right = FALSE)  
#Create dataframe  
LeaveDiffbyAge <- table(health$perceivedleavediff1, health$AgeGroup)

#### Crosstab Table

#Print Crosstab table  
kable(LeaveDiffbyAge)

|  |  |  |  |
| --- | --- | --- | --- |
|  | 17-37 | 37-57 | 57-77 |
| Very easy (VE) | 157 | 62 | 1 |
| Somewhat Easy (SE) | 203 | 77 | 1 |
| Neutral (N) | 126 | 48 | 4 |
| Somewhat Hard (SH) | 134 | 63 | 2 |
| Very Hard (VH) | 82 | 33 | 3 |
| Don’t Know (DK) | 105 | 44 | 1 |

#### Proportions Table

#Create Proportions table  
kable(prop.table(LeaveDiffbyAge, 2))

|  |  |  |  |
| --- | --- | --- | --- |
|  | 17-37 | 37-57 | 57-77 |
| Very easy (VE) | 0.1945477 | 0.1896024 | 0.0833333 |
| Somewhat Easy (SE) | 0.2515489 | 0.2354740 | 0.0833333 |
| Neutral (N) | 0.1561338 | 0.1467890 | 0.3333333 |
| Somewhat Hard (SH) | 0.1660471 | 0.1926606 | 0.1666667 |
| Very Hard (VH) | 0.1016109 | 0.1009174 | 0.2500000 |
| Don’t Know (DK) | 0.1301115 | 0.1345566 | 0.0833333 |

#### Chi-Squared

chisq.test(LeaveDiffbyAge)

Warning in chisq.test(LeaveDiffbyAge): Chi-squared approximation may be  
incorrect

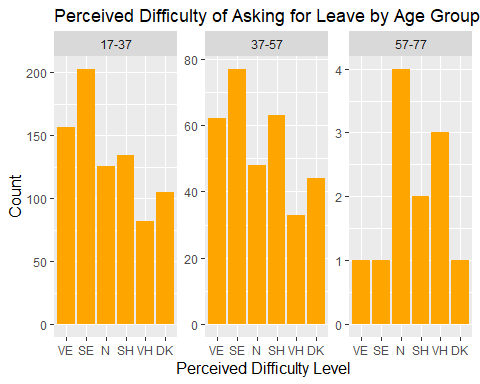
Pearson's Chi-squared test  
  
data: LeaveDiffbyAge  
X-squared = 8.6657, df = 10, p-value = 0.5641

Our null hypothesis is that there is that age group and perceived difficulty of requesting leave are independent. Our alternative hypothesis is that there is a relationship between these two variables. Because the sensitivity of this test to larger sample sizes we are setting the p value to .01 (Townsley, 2002)

We get a warning that the approximation may be incorrect, likely because because of the small number of respondents in our survey in the 57-77 age group (Townsley, 2002). However, the large p value potentially indicates that that the differences in perceived ease of requesting leave between age may not be statistically significant (or we do not have enough information because of the small number of people in the 57-77 age group in our dataset).

#### Visualizing the Data

#Create Visualization  
LeaveDiffbyAge\_Viz <- ggplot(health, aes(x = health$perceivedleavediff2))+geom\_bar(fill = "orange")+facet\_wrap(health$AgeGroup ~ ., scales="free\_y") + ggtitle("Perceived Difficulty of Asking for Leave by Age Group") + xlab("Perceived Difficulty Level") + ylab("Count") + theme(axis.title=element\_text(size=12))  
LeaveDiffbyAge\_Viz



NOTE: VE = Very Easy, SE = Somewhat Easy, N = Neutral (or Neither Easy nor Difficult), SH = Somewhat Hard, VH = Very Hard, DK = Don’t Know.

#### Interpretation of Results

The most common response for those in the 17-27 and 37-57 age ranges is that asking for leave for a mental health issue was “Somewhat Easy” and the second most common isthat asking for leave “Very Easy”. However, there is a significant number in both age groups who provided a neutral answer (e.g., neither easy or difficult) or indicated that they felt asking for leave was hard/very hard. That being said, there are few people who are 57-77 participating in this survey and there are fewer people in the 37-57 age group than the 17-27 age group. Therefore, we may not have a complete picture of how perceived ease of requesting leave for a mental health issue may change as people age.

### Answers by Company Size

#Create subset  
LeaveDiffbyCompanySize <- table(health$perceivedleavediff1, health$CompanySize)

#### Crosstab Table

#Print Crosstab Table  
kable(LeaveDiffbyCompanySize)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 1-5 | 6-25 | 26-100 | 100-500 | 500-1000 | More than 1000 |
| Very easy (VE) | 16 | 42 | 59 | 48 | 15 | 40 |
| Somewhat Easy (SE) | 12 | 41 | 74 | 62 | 30 | 62 |
| Neutral (N) | 6 | 40 | 43 | 30 | 14 | 45 |
| Somewhat Hard (SH) | 11 | 45 | 46 | 48 | 6 | 43 |
| Very Hard (VH) | 13 | 16 | 36 | 15 | 5 | 33 |
| Don’t Know (DK) | 2 | 26 | 34 | 45 | 10 | 33 |

#### Proportions Table

kable(prop.table(LeaveDiffbyCompanySize, 2))

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 1-5 | 6-25 | 26-100 | 100-500 | 500-1000 | More than 1000 |
| Very easy (VE) | 0.2666667 | 0.2000000 | 0.2020548 | 0.1935484 | 0.1875 | 0.1562500 |
| Somewhat Easy (SE) | 0.2000000 | 0.1952381 | 0.2534247 | 0.2500000 | 0.3750 | 0.2421875 |
| Neutral (N) | 0.1000000 | 0.1904762 | 0.1472603 | 0.1209677 | 0.1750 | 0.1757812 |
| Somewhat Hard (SH) | 0.1833333 | 0.2142857 | 0.1575342 | 0.1935484 | 0.0750 | 0.1679688 |
| Very Hard (VH) | 0.2166667 | 0.0761905 | 0.1232877 | 0.0604839 | 0.0625 | 0.1289062 |
| Don’t Know (DK) | 0.0333333 | 0.1238095 | 0.1164384 | 0.1814516 | 0.1250 | 0.1289062 |

#### Chi-Squared

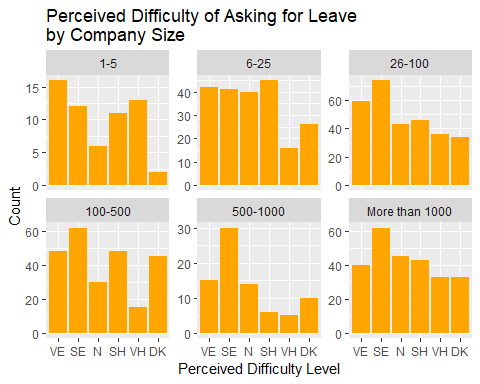
chisq.test(LeaveDiffbyCompanySize)

Pearson's Chi-squared test  
  
data: LeaveDiffbyCompanySize  
X-squared = 52.482, df = 25, p-value = 0.001041

Our null hypothesis is that there is that company size and perceived difficulty of requesting leave are independent. Our alternative hypothesis is that there is a relationship between these two variables. Because the sensitivity of this test to larger sample sizes we are setting the p value to .01 (Townsley, 2002). Because the p value is less than .01 we can reject the null and say that we think that there is a relationship between these two variables.

#### Visualizing the Data

#Create Visualization  
LeaveDiffbyCompanySize\_viz <- ggplot(health, aes(x = health$perceivedleavediff2))+geom\_bar(fill = "orange")+facet\_wrap(health$CompanySize~ ., scales="free\_y") + ggtitle("Perceived Difficulty of Asking for Leave\nby Company Size") + xlab("Perceived Difficulty Level") + ylab("Count")  
LeaveDiffbyCompanySize\_viz



NOTE: VE = Very Easy, SE = Somewhat Easy, N = Neutral (or Neither Easy nor Difficult), SH = Somewhat Hard, VH = Very Hard, DK = Don’t Know.

#### Interpretation of Results

For very small companies (1-5 employees), there seems to be a fairly even split between those who perceive taking leave for mental health to be easy and those who perceive it to be hard. For other small companies (6-25 employees), approximately the same number of people say that it is easy/somewhat easy or provide a neutral response. However, in this case, the most common answer is that they would perceive it as somehwat hard to ask for leave. For larger companies, the most common answer is that they perceive it to be somewhat easy to request leave for a mental illness. Therefore it appears that as the company size gets larger, it may be perceived as easier in most companies to request leave related to a mental illness. However, there is still sizable minority of employees who work for larger companies (over 25 employees) who do perceive it as hard or very hard to ask for leave. Therefore, a lot may vary by the company.

### Answers by Whether Current Company Provides Mental Health Benefits

#Modify Labels for mental health benefit provision question  
health$MHBenefit2 <- factor(health$Does.your.employer.provide.mental.health.benefits.as.part.of.healthcare.coverage.,  
levels = c("I don't know", "No", "Not eligible for coverage / N/A","Yes"),  
labels = c("Don't Know", "No", "Ineligible/NA", "Yes"))  
#Create the subset  
LeaveDiffbyMHBenefit <- table(health$perceivedleavediff1, health$MHBenefit2)

#### Crosstab Table

#Print the Crosstab Table  
kable(LeaveDiffbyMHBenefit)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Don’t Know | No | Ineligible/NA | Yes |
| Very easy (VE) | 55 | 36 | 18 | 111 |
| Somewhat Easy (SE) | 81 | 46 | 17 | 137 |
| Neutral (N) | 52 | 29 | 16 | 81 |
| Somewhat Hard (SH) | 51 | 49 | 17 | 82 |
| Very Hard (VH) | 25 | 31 | 10 | 52 |
| Don’t Know (DK) | 55 | 22 | 5 | 68 |

#### Proportions Table

#Print the Crosstab Table  
kable(prop.table(LeaveDiffbyMHBenefit, 2))

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Don’t Know | No | Ineligible/NA | Yes |
| Very easy (VE) | 0.1724138 | 0.1690141 | 0.2168675 | 0.2090395 |
| Somewhat Easy (SE) | 0.2539185 | 0.2159624 | 0.2048193 | 0.2580038 |
| Neutral (N) | 0.1630094 | 0.1361502 | 0.1927711 | 0.1525424 |
| Somewhat Hard (SH) | 0.1598746 | 0.2300469 | 0.2048193 | 0.1544256 |
| Very Hard (VH) | 0.0783699 | 0.1455399 | 0.1204819 | 0.0979284 |
| Don’t Know (DK) | 0.1724138 | 0.1032864 | 0.0602410 | 0.1280603 |

#### Chi-Squared

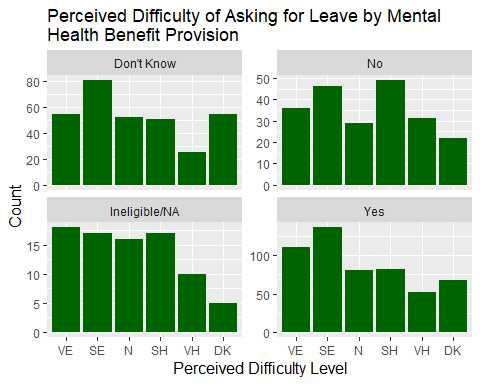
#Run Chi-squared test  
chisq.test(LeaveDiffbyMHBenefit)

Pearson's Chi-squared test  
  
data: LeaveDiffbyMHBenefit  
X-squared = 25.93, df = 15, p-value = 0.03877

Our null hypothesis is that there is that reported mental health benefit provision and perceived difficulty of requesting leave are independent. Our alternative hypothesis is that there is a relationship between these two variables. Because the sensitivity of this test to larger sample sizes we are setting the p value to .01 (Townsley, 2002). Because the p value is above .01, we fail to reject the null. There does not appear to be a statistically significant relationship between the two variables.

#### Visualizing the Data

#Create Visualization  
LeaveDiffbyMHBenefit\_viz <- ggplot(health, aes(x = health$perceivedleavediff2))+geom\_bar(fill = "dark green")+facet\_wrap(~ health$MHBenefit2, scales="free\_y") + ggtitle("Perceived Difficulty of Asking for Leave by Mental\nHealth Benefit Provision") + xlab("Perceived Difficulty Level") + ylab("Count") + theme(axis.title=element\_text(size=12))  
LeaveDiffbyMHBenefit\_viz



NOTE: VE = Very Easy, SE = Somewhat Easy, N = Neutral (or Neither Easy nor Difficult), SH = Somewhat Hard, VH = Very Hard, DK = Don’t Know.

#### Interpretation of Results

For those who do not get mental health benefits, more participants responded that they perceive it as somewhat difficult to request leave for a mental health issue than any other response (with somewhat easy being the second most common). Among those who do get mental health benefits, most indicated they perceive it as very easy or somwhat easy to request leave for a mental health issue. Most of those who did not know if mental health benefits are offered inidicated they perceived it to be somewhat easy to request leave for a mental health issue. Those who are ineligible (or for whom it was not applicable) were evently split among the responses, although relatively few people selected this response compared to the others. However, the difference between the two variables does not appear to be statistically significant.

### Answers by Whether Current Company Has Ever Formally Discussed Mental Health

#Create the subset  
LeaveDiffbyMHDiscussion <- table(health$perceivedleavediff1, health$Has.your.employer.ever.formally.discussed.mental.health..for.example..as.part.of.a.wellness.campaign.or.other.official.communication..)

#### Crosstab Table

#Print crosstab table  
kable(LeaveDiffbyMHDiscussion)

|  |  |  |  |
| --- | --- | --- | --- |
|  | I don’t know | No | Yes |
| Very easy (VE) | 23 | 129 | 68 |
| Somewhat Easy (SE) | 25 | 183 | 73 |
| Neutral (N) | 20 | 126 | 32 |
| Somewhat Hard (SH) | 13 | 161 | 25 |
| Very Hard (VH) | 3 | 105 | 10 |
| Don’t Know (DK) | 19 | 109 | 22 |

#### Proportions Table

#Create proportions table  
kable(prop.table(LeaveDiffbyMHDiscussion, 2))

|  |  |  |  |
| --- | --- | --- | --- |
|  | I don’t know | No | Yes |
| Very easy (VE) | 0.2233010 | 0.1586716 | 0.2956522 |
| Somewhat Easy (SE) | 0.2427184 | 0.2250923 | 0.3173913 |
| Neutral (N) | 0.1941748 | 0.1549815 | 0.1391304 |
| Somewhat Hard (SH) | 0.1262136 | 0.1980320 | 0.1086957 |
| Very Hard (VH) | 0.0291262 | 0.1291513 | 0.0434783 |
| Don’t Know (DK) | 0.1844660 | 0.1340713 | 0.0956522 |

#### Chi-Squared

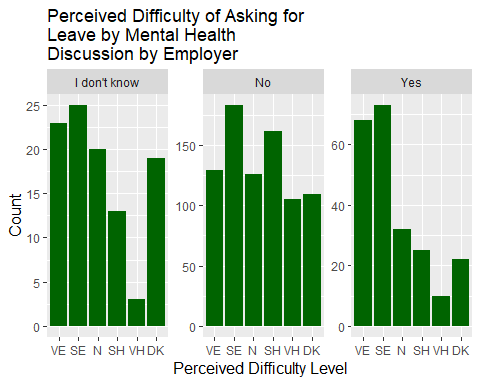
#Run chi-squared test  
chisq.test(LeaveDiffbyMHDiscussion)

Pearson's Chi-squared test  
  
data: LeaveDiffbyMHDiscussion  
X-squared = 58.706, df = 10, p-value = 6.364e-09

Our null hypothesis is that there is that reported mental health discussion by current employer and perceived difficulty of requesting leave are independent. Our alternative hypothesis is that there is a relationship between these two variables. Because the sensitivity of this test to larger sample sizes we are setting the p value to .01 (Townsley, 2002). Because the p value is below .01, we reject the null. There appears to be a statistically significant relationship between the two variables.

#### Visualizing the Data

#Create Visualization  
LeaveDiffbyMHDiscussion\_viz <- ggplot(health, aes(x = health$perceivedleavediff2))+geom\_bar(fill = "dark green")+facet\_wrap( health$Has.your.employer.ever.formally.discussed.mental.health..for.example..as.part.of.a.wellness.campaign.or.other.official.communication.. ~., scales="free\_y") + ggtitle("Perceived Difficulty of Asking for \nLeave by Mental Health\nDiscussion by Employer") + xlab("Perceived Difficulty Level") + ylab("Count") + theme(axis.title=element\_text(size=12))  
LeaveDiffbyMHDiscussion\_viz



NOTE: VE = Very Easy, SE = Somewhat Easy, N = Neutral (or Neither Easy nor Difficult), SH = Somewhat Hard, VH = Very Hard, DK = Don’t Know.

#### Interpretation of Results

Across all categories (e.g., yes they have discussed, no they haven’t, don’t know), the most common response is that most recipients think it would be somewhat easy to request leave. Interestingly, for those who did not know if there has been a discussion, a large contingent also indicated they don’t know how hard it would be to request leave. For those where there has been a discussion, far fewer people say it would be hard to request leave to handle a mental illness than those who say it would be easy to very easy. For those where there has not been a discussion (which is the majority of respondents), the second most common answer is that it would be somewhat hard to request leave to handle a mental illness. In addition, the proportion of respondents who indicated that it would be very hard to ask for leave is a lot lower in companies where it is reported that there was a mental health discussion or that it is unknown that there has been a mental health discussion compared to companies where there has not been a discussion.

### Answers by Whether Current Company Offers Other Resources

#Creating the subset  
LeaveDiffbyOtherMHResources <- table(health$perceivedleavediff1, health$Does.your.employer.offer.resources.to.learn.more.about.mental.health.concerns.and.options.for.seeking.help.)

#### Crosstab Table

#Print the crosstab  
kable(LeaveDiffbyOtherMHResources)

|  |  |  |  |
| --- | --- | --- | --- |
|  | I don’t know | No | Yes |
| Very easy (VE) | 59 | 84 | 77 |
| Somewhat Easy (SE) | 80 | 123 | 78 |
| Neutral (N) | 49 | 80 | 49 |
| Somewhat Hard (SH) | 52 | 110 | 37 |
| Very Hard (VH) | 30 | 74 | 14 |
| Don’t Know (DK) | 50 | 60 | 40 |

#### Proportions Table

#Create the proportions table  
kable(prop.table(LeaveDiffbyOtherMHResources, 2))

|  |  |  |  |
| --- | --- | --- | --- |
|  | I don’t know | No | Yes |
| Very easy (VE) | 0.184375 | 0.1581921 | 0.2610169 |
| Somewhat Easy (SE) | 0.250000 | 0.2316384 | 0.2644068 |
| Neutral (N) | 0.153125 | 0.1506591 | 0.1661017 |
| Somewhat Hard (SH) | 0.162500 | 0.2071563 | 0.1254237 |
| Very Hard (VH) | 0.093750 | 0.1393597 | 0.0474576 |
| Don’t Know (DK) | 0.156250 | 0.1129944 | 0.1355932 |

#### Chi-Squared

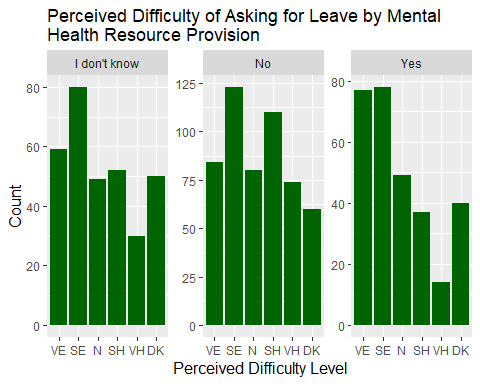
#Run chi-squared test  
chisq.test(LeaveDiffbyOtherMHResources)

Pearson's Chi-squared test  
  
data: LeaveDiffbyOtherMHResources  
X-squared = 38.212, df = 10, p-value = 3.486e-05

Our null hypothesis is that there is that reported other mental health resources by current employer and perceived difficulty of requesting leave are independent. Our alternative hypothesis is that there is a relationship between these two variables. Because the sensitivity of this test to larger sample sizes we are setting the p value to .01 (Townsley, 2002). Because the p value is below .01, we reject the null. There appears to be a statistically significant relationship between the two variables.

#### Visualizing the Data

#Create visualization  
LeaveDiffbyOtherMHResources\_viz <- ggplot(health, aes(x = health$perceivedleavediff2))+geom\_bar(fill = "dark green")+facet\_wrap( health$Does.your.employer.offer.resources.to.learn.more.about.mental.health.concerns.and.options.for.seeking.help. ~., scales="free\_y") + ggtitle("Perceived Difficulty of Asking for Leave by Mental\nHealth Resource Provision") + xlab("Perceived Difficulty Level") + ylab("Count") + theme(axis.title=element\_text(size=12))  
LeaveDiffbyOtherMHResources\_viz



NOTE: VE = Very Easy, SE = Somewhat Easy, N = Neutral (or Neither Easy nor Difficult), SH = Somewhat Hard, VH = Very Hard, DK = Don’t Know.

#### Interpretation of Results

Across all categories (e.g., yes there are resources, no they aren’t, don’t know), the most common response is that most recipients think it would be somewhat easy to request leave. Similarly to the results provided above, in companies where there are other resources provided, fewer people say it would be hard to request leave to handle a mental illness than those who say it would be easy to very easy. For those where there are not additional resources offered related to mental health (the majority of respondents), the second most common answer is that it would be somewhat hard to request leave to hanlde a mental illness.

### Answers by If They Ever Saw A Poor Response to Mental Health Issue

#Modify Labels for poor response to mental health issue question to make them shorter  
health$expwpoorMHresponse <- factor(health$Have.you.observed.or.experienced.an.unsupportive.or.badly.handled.response.to.a.mental.health.issue.in.your.current.or.previous.workplace.,  
levels = c("Maybe/Not sure", "N/A", "No","Yes, I experienced", "Yes, I observed"),  
labels = c("Don't Know", "N/A", "No", "Yes, experienced", "Yes, observed"))  
#Create subset  
LeaveDiffbyPoorResponse <- table(health$perceivedleavediff1, health$expwpoorMHresponse)

#### Crosstab Table

#Create table  
kable(LeaveDiffbyPoorResponse)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Don’t Know | N/A | No | Yes, experienced | Yes, observed |
| Very easy (VE) | 34 | 7 | 136 | 17 | 26 |
| Somewhat Easy (SE) | 58 | 14 | 126 | 35 | 48 |
| Neutral (N) | 46 | 10 | 70 | 18 | 34 |
| Somewhat Hard (SH) | 60 | 10 | 66 | 32 | 31 |
| Very Hard (VH) | 38 | 3 | 24 | 20 | 33 |
| Don’t Know (DK) | 42 | 8 | 69 | 10 | 21 |

#### Proportions Table

#Create proportions table  
kable(prop.table(LeaveDiffbyPoorResponse, 2))

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Don’t Know | N/A | No | Yes, experienced | Yes, observed |
| Very easy (VE) | 0.1223022 | 0.1346154 | 0.2769857 | 0.1287879 | 0.1347150 |
| Somewhat Easy (SE) | 0.2086331 | 0.2692308 | 0.2566191 | 0.2651515 | 0.2487047 |
| Neutral (N) | 0.1654676 | 0.1923077 | 0.1425662 | 0.1363636 | 0.1761658 |
| Somewhat Hard (SH) | 0.2158273 | 0.1923077 | 0.1344196 | 0.2424242 | 0.1606218 |
| Very Hard (VH) | 0.1366906 | 0.0576923 | 0.0488798 | 0.1515152 | 0.1709845 |
| Don’t Know (DK) | 0.1510791 | 0.1538462 | 0.1405295 | 0.0757576 | 0.1088083 |

#### Chi-Squared

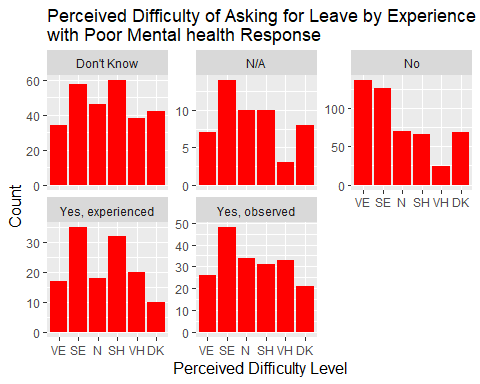
chisq.test(LeaveDiffbyPoorResponse)

Pearson's Chi-squared test  
  
data: LeaveDiffbyPoorResponse  
X-squared = 82.616, df = 20, p-value = 1.406e-09

Our null hypothesis is that there is that experience with a poor response to a mental health issue (current or former workplace) and perceived difficulty of requesting leave are independent. Our alternative hypothesis is that there is a relationship between these two variables. Because the sensitivity of this test to larger sample sizes we are setting the p value to .01 (Townsley, 2002). Because the p value is below .01, we reject the null. There appears to be a statistically significant relationship between the two variables.

#### Visualizing the Data

#Create Visualization  
LeaveDiffbyPoorResponse\_viz <- ggplot(health, aes(x = health$perceivedleavediff2))+geom\_bar(fill = "red")+facet\_wrap(health$expwpoorMHresponse ~., scales="free\_y") + ggtitle("Perceived Difficulty of Asking for Leave by Experience\nwith Poor Mental health Response") + xlab("Perceived Difficulty Level") + ylab("Count") + theme(axis.title=element\_text(size=12))  
LeaveDiffbyPoorResponse\_viz



NOTE: E = Very Easy, S = Somewhat Easy, N = Neutral (or Neither Easy nor Difficult), SH = Somewhat Hard, H = Very Hard, D = Don’t Know.

#### Interpretation of Results

Most people surveyed indicate that they have never had an experience with an unsupportive or badly handled response to mental health in the workplace. Among those who selected this response, the vast majority indicate they think that it would be very or somewhat easy to request leave to handle a mental illness. For those who have observed a badly handled response, the most common response is that they would find it somewhat easy to request leave, but the second most common is that it they would find it somewhat difficult. For those who have experienced a poorly handled response, the most common response is that they would find it somewhat easy to request leave, but a significant amount said they would find it hard or very hard to request leave (or provided a neutral response saying it would be neither easy or difficult). Therefore, it is possible that experience with a poorly handled response may have an impact on perceived ease of requesting leave.

## Comfort with talking to Supervisor About Mental Health

This part of the analysis looks at answers to the question “Would you feel comfortable discussing a mental health disorder with your direct supervisor” by respondent demographics as well as their reports of several company characteristics. This is important because if an employee has a mental health issue, they should ideally be comfortable discussing it with a supervisor in case they need additional assistance.

### Answers by Gender

#Create subset  
DiscMHBossbyGender <- table(health$Would.you.feel.comfortable.discussing.a.mental.health.disorder.with.your.direct.supervisor.s.., health$gender)

#### Crosstab Table

#Create crosstab  
kable(DiscMHBossbyGender)

|  |  |  |  |
| --- | --- | --- | --- |
|  | Female | Male | Other |
| Maybe | 87 | 286 | 9 |
| No | 84 | 245 | 7 |
| Yes | 100 | 316 | 12 |

#### Proportions Table

#Create proportion table  
kable(prop.table(DiscMHBossbyGender, 2))

|  |  |  |  |
| --- | --- | --- | --- |
|  | Female | Male | Other |
| Maybe | 0.3210332 | 0.3376623 | 0.3214286 |
| No | 0.3099631 | 0.2892562 | 0.2500000 |
| Yes | 0.3690037 | 0.3730815 | 0.4285714 |

#### Chi-Squared

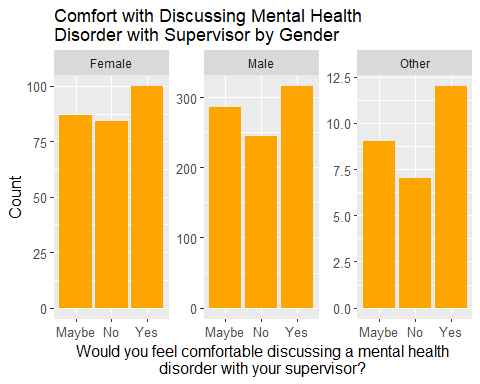
#Run chi-squared test  
chisq.test(DiscMHBossbyGender)

Pearson's Chi-squared test  
  
data: DiscMHBossbyGender  
X-squared = 0.90786, df = 4, p-value = 0.9234

Our null hypothesis is that there is that reported comfort discussing mental health with supervisor and gender are independent. Our alternative hypothesis is that there is a relationship between these two variables. Because the sensitivity of this test to larger sample sizes we are setting the p value to .01 (Townsley, 2002). Because the p value is very large, we fail to reject the null. There does not appear to be a statistically significant relationship between the two variables. However, this might be influenced by the small frequencies among those who report having an other gender (Townsley, 2002).

#### Visualizing the Data

#Create visualization  
DiscMHBossbyGender\_viz <- ggplot(health, aes(x = health$Would.you.feel.comfortable.discussing.a.mental.health.disorder.with.your.direct.supervisor.s..))+geom\_bar(fill = "orange")+facet\_wrap(~ health$gender, scales="free\_y") + ggtitle("Comfort with Discussing Mental Health\nDisorder with Supervisor by Gender") + xlab("Would you feel comfortable discussing a mental health\ndisorder with your supervisor?") + ylab("Count") + theme(axis.title=element\_text(size=12), axis.text=element\_text(size=10))  
DiscMHBossbyGender\_viz



#### Interpretation of Results

Most respondents across genders indicate that they would be comfortable discussing a mental health disorder with their supervisors. With male employees, a lower proportion indicate that they would not be comfortable discussing a mental health disorder with their supervisors. Meanwhile, with female and employees of other genders, the results are slightly more evenly split among the three options. However, the difference is not statistically significant.

### Answers by Age Group

#Create subset  
DiscMHBossbyAge <- table(health$Would.you.feel.comfortable.discussing.a.mental.health.disorder.with.your.direct.supervisor.s.., health$AgeGroup)

#### Crosstab Table

#Run crosstab table  
kable(DiscMHBossbyAge)

|  |  |  |  |
| --- | --- | --- | --- |
|  | 17-37 | 37-57 | 57-77 |
| Maybe | 265 | 114 | 3 |
| No | 236 | 96 | 4 |
| Yes | 306 | 117 | 5 |

#### Proportions Table

kable(prop.table(DiscMHBossbyAge, 2))

|  |  |  |  |
| --- | --- | --- | --- |
|  | 17-37 | 37-57 | 57-77 |
| Maybe | 0.3283767 | 0.3486239 | 0.2500000 |
| No | 0.2924411 | 0.2935780 | 0.3333333 |
| Yes | 0.3791822 | 0.3577982 | 0.4166667 |

#### Chi-Squared

chisq.test(DiscMHBossbyAge)

Warning in chisq.test(DiscMHBossbyAge): Chi-squared approximation may be  
incorrect

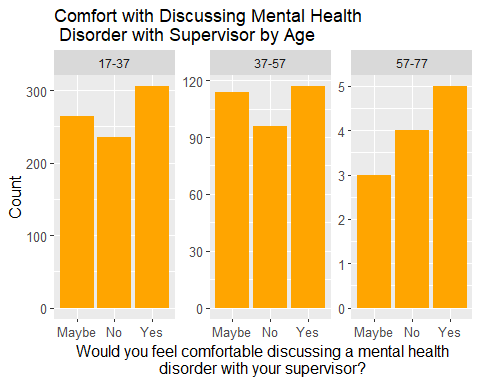
Pearson's Chi-squared test  
  
data: DiscMHBossbyAge  
X-squared = 0.95201, df = 4, p-value = 0.917

Our null hypothesis is that there is that age group and perceived difficulty of requesting leave are independent. Our alternative hypothesis is that there is a relationship between these two variables. Because the sensitivity of this test to larger sample sizes we are setting the p value to .01 (Townsley, 2002)

We get a warning that the approximation may be incorrect, likely because because of the small number of respondents in our survey in the 57-77 age group (Townsley, 2002). However, the large p value indicates that that the differences in perceived ease of requesting leave between age may not be statistically significant (or we do not have enough information because of the small number of people in the 57-77 age group in our dataset).

#### Visualizing the Data

DiscMHBossbyAge\_viz <- ggplot(health, aes(x = health$Would.you.feel.comfortable.discussing.a.mental.health.disorder.with.your.direct.supervisor.s..))+geom\_bar(fill = "orange")+facet\_wrap(health$AgeGroup ~., scales="free\_y") + ggtitle("Comfort with Discussing Mental Health\n Disorder with Supervisor by Age") + xlab("Would you feel comfortable discussing a mental health\ndisorder with your supervisor?") + ylab("Count") + theme(axis.title=element\_text(size=12), axis.text=element\_text(size=10))  
DiscMHBossbyAge\_viz



#### Interpretation of Results

The majority of respondents in the 17-37 and 37-57 age groups responded that they would feel comfortable discussing a mental health disorder with their supervisor. The next most common response is that they might, and no received the least amount of responses from the 17-37 and 37-57 age groups. However, in both cases over a quarter of participants in both age groups answered that they would not feel comfortable. The 57-77 range has a low number of responses, although in that case “No” is the second most common reponse. Given the chi-squared test results and the visualization, it does not appear that age has much of an impact on comfort level in discussing a mental health disorder with a supervisor. However, the fact that there are fewer responses from the 37-57 and especially the 57-77 age groups may impact our results.

### Answers by Company Size

#Creating subset  
DiscMHBossbyCompSize <- table(health$Would.you.feel.comfortable.discussing.a.mental.health.disorder.with.your.direct.supervisor.s.., health$CompanySize)

#### Crosstab Table

#Creating the crosstab  
kable(DiscMHBossbyCompSize)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 1-5 | 6-25 | 26-100 | 100-500 | 500-1000 | More than 1000 |
| Maybe | 14 | 69 | 105 | 87 | 32 | 75 |
| No | 21 | 61 | 74 | 66 | 17 | 97 |
| Yes | 25 | 80 | 113 | 95 | 31 | 84 |

#### Proportions Table

#Create the proportions table  
kable(prop.table(DiscMHBossbyCompSize, 2))

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 1-5 | 6-25 | 26-100 | 100-500 | 500-1000 | More than 1000 |
| Maybe | 0.2333333 | 0.3285714 | 0.3595890 | 0.3508065 | 0.4000 | 0.2929688 |
| No | 0.3500000 | 0.2904762 | 0.2534247 | 0.2661290 | 0.2125 | 0.3789062 |
| Yes | 0.4166667 | 0.3809524 | 0.3869863 | 0.3830645 | 0.3875 | 0.3281250 |

#### Chi-Squared

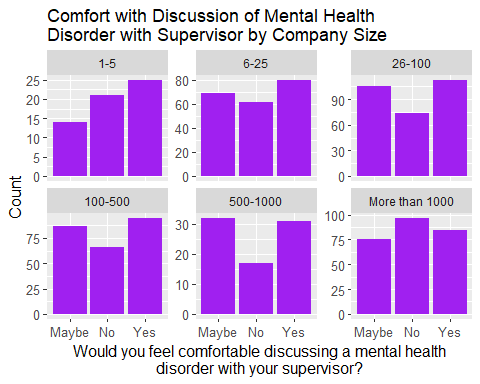
#Run chi-squared test  
chisq.test(DiscMHBossbyCompSize)

Pearson's Chi-squared test  
  
data: DiscMHBossbyCompSize  
X-squared = 18.002, df = 10, p-value = 0.05493

Our null hypothesis is that there is that reported comfort discussing mental health with supervisor and reported company size are independent. Our alternative hypothesis is that there is a relationship between these two variables. Because the sensitivity of this test to larger sample sizes we are setting the p value to .01 (Townsley, 2002). Because the p value is not less than .01 we cannot reject the null.

#### Visualizing the Data

#Create visualization  
DiscMHBossbyCompSize\_viz <- ggplot(health, aes(x = health$Would.you.feel.comfortable.discussing.a.mental.health.disorder.with.your.direct.supervisor.s..))+geom\_bar(fill = "purple")+facet\_wrap(health$CompanySize ~., scales="free\_y") + ggtitle("Comfort with Discussion of Mental Health\nDisorder with Supervisor by Company Size") + xlab("Would you feel comfortable discussing a mental health\ndisorder with your supervisor?") + ylab("Count") + theme(axis.title=element\_text(size=12), axis.text=element\_text(size=10))  
DiscMHBossbyCompSize\_viz



#### Interpretation of Results

The pattern that we see in the data visualization is interesting. For companies in the 26-100, 100-500, and 500-1000 groups (and to a lesser extent in the 6-25 group), fewer respondents indicate they would not feel comfortable discussing a mental health disorder with their supervisor than selecting the “Maybe” or “Yes” options. However, in the “More than 1000” group, most respondents indicated they would not feel comfortable (although a significant amount answered “Maybe” or “Yes”). It is possible that for larger companies people may feel less comfortable discussing these issues with supervisors because it the organization is more bureaucratic or there are employee/supervisor relationships that are less personal. However, it is also possible that these results indicate that the results are depedendent on the company as well as other factors. Because the null is larger than .01, we fail to reject the null that there is a significant relationship between the two variables.

### Answers by If They Ever Saw A Poor Response to Mental Health Issue

#Create subset  
DiscMHBossbyPoorResponse1 <- table(health$Would.you.feel.comfortable.discussing.a.mental.health.disorder.with.your.direct.supervisor.s.., health$Have.you.observed.or.experienced.an.unsupportive.or.badly.handled.response.to.a.mental.health.issue.in.your.current.or.previous.workplace.)

#### Crosstab Table

#Create crosstab table  
kable(DiscMHBossbyPoorResponse1)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Maybe/Not sure | N/A | No | Yes, I experienced | Yes, I observed |
| Maybe | 97 | 12 | 161 | 48 | 64 |
| No | 101 | 24 | 109 | 35 | 67 |
| Yes | 80 | 16 | 221 | 49 | 62 |

#### Proportions Table

#Create proportions table  
kable(prop.table(DiscMHBossbyPoorResponse1, 2))

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Maybe/Not sure | N/A | No | Yes, I experienced | Yes, I observed |
| Maybe | 0.3489209 | 0.2307692 | 0.3279022 | 0.3636364 | 0.3316062 |
| No | 0.3633094 | 0.4615385 | 0.2219959 | 0.2651515 | 0.3471503 |
| Yes | 0.2877698 | 0.3076923 | 0.4501018 | 0.3712121 | 0.3212435 |

#### Chi-Squared

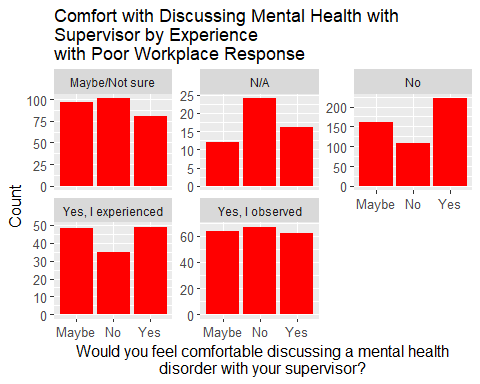
#Run chi-squared test  
chisq.test(DiscMHBossbyPoorResponse1)

Pearson's Chi-squared test  
  
data: DiscMHBossbyPoorResponse1  
X-squared = 37.901, df = 8, p-value = 7.852e-06

Our null hypothesis is that there is that reported comfort discussing mental health with supervisor and experience with a poor response to a mental health issue are independent. Our alternative hypothesis is that there is a relationship between these two variables. Because the sensitivity of this test to larger sample sizes we are setting the p value to .01 (Townsley, 2002). Because the p value is less than .01 we can reject the null and state that we suspect that there is a relationship between these two variables.

#### Visualizing the Data

#Create visualization  
DiscMHBossbyPoorResponse1\_viz <- ggplot(health, aes(x = health$Would.you.feel.comfortable.discussing.a.mental.health.disorder.with.your.direct.supervisor.s..))+geom\_bar(fill = "red")+facet\_wrap(health$Have.you.observed.or.experienced.an.unsupportive.or.badly.handled.response.to.a.mental.health.issue.in.your.current.or.previous.workplace. ~., scales="free\_y") + ggtitle("Comfort with Discussing Mental Health with\nSupervisor by Experience \nwith Poor Workplace Response") + xlab("Would you feel comfortable discussing a mental health\ndisorder with your supervisor?") + ylab("Count") + theme(axis.title=element\_text(size=12), axis.text=element\_text(size=10))  
DiscMHBossbyPoorResponse1\_viz



#### Interpret Results

When asked about experience with a poorly handled response to mental health, respondents who were not sure if they expereinced it were evenly split between answering that they maybe comfortable discussing a mental health disorder with supervisors or not comfortable (there was also a large group that said they would be comfortable). For those who answered that they did not have experience with a poor response, they also answered yes more often than maybe and no when also asked about their comfort with discussing a mental health disorder with a supervisor. Interestingly, those who stated “Yes, I observed” a poor response to mental health in the workplace were evenly split selected maybe the most often when also asked if they would feel comfortable discussing a mental health disorder with coworkers (and “yes” was the second most common response). However, those who stated that they experienced a poorly handled response were evenly split between the yes, no, and maybe answers when asked about comfort level discussing mental health with a supervisor.

### Answers by If They Think Discussing a Mental Health Disorder with Employer Would Have Negative Consequences

#Modify Labels for perception of mental health disclosure impact  
health$negimp <- factor(health$Do.you.think.that.discussing.a.mental.health.disorder.with.your.employer.would.have.negative.consequences.,  
levels = c("Maybe", "No", "Yes"),  
labels = c("Maybe Expect Impact", "Do Not Expect Impact", "Expect Impact"))  
#create subset  
DiscMHBossbyNegImpact <- table(health$Would.you.feel.comfortable.discussing.a.mental.health.disorder.with.your.direct.supervisor.s.., health$negimp)

#### Crosstab Table

kable(DiscMHBossbyNegImpact)

|  |  |  |  |
| --- | --- | --- | --- |
|  | Maybe Expect Impact | Do Not Expect Impact | Expect Impact |
| Maybe | 218 | 115 | 49 |
| No | 164 | 24 | 148 |
| Yes | 105 | 299 | 24 |

#### Proportions Table

kable(prop.table(DiscMHBossbyNegImpact, 2))

|  |  |  |  |
| --- | --- | --- | --- |
|  | Maybe Expect Impact | Do Not Expect Impact | Expect Impact |
| Maybe | 0.4476386 | 0.2625571 | 0.2217195 |
| No | 0.3367556 | 0.0547945 | 0.6696833 |
| Yes | 0.2156057 | 0.6826484 | 0.1085973 |

#### Chi-Squared

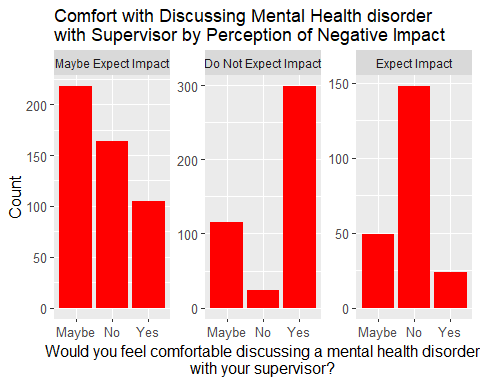
chisq.test(DiscMHBossbyNegImpact)

Pearson's Chi-squared test  
  
data: DiscMHBossbyNegImpact  
X-squared = 414.95, df = 4, p-value < 2.2e-16

Our null hypothesis is that there is that reported comfort discussing mental health with supervisor and perception that disclosing a mental health disorder to employer would have a negative impact on career are independent. Our alternative hypothesis is that there is a relationship between these two variables. Because the sensitivity of this test to larger sample sizes we are setting the p value to .01 (Townsley, 2002). Because the p value is less than .01 we can reject the null and state that we suspect that there is a relationship between these two variables.

#### Visualizing the Data

DiscMHBossbyNegImpact\_viz <- ggplot(health, aes(x = health$Would.you.feel.comfortable.discussing.a.mental.health.disorder.with.your.direct.supervisor.s..))+geom\_bar(fill = "red")+facet\_wrap(health$negimp ~., scales="free\_y") + ggtitle("Comfort with Discussing Mental Health disorder\nwith Supervisor by Perception of Negative Impact") + xlab("Would you feel comfortable discussing a mental health disorder\nwith your supervisor?") + ylab("Count") + theme(axis.title=element\_text(size=12), axis.text=element\_text(size=10))  
DiscMHBossbyNegImpact\_viz



#### Interpret Results

The results here are fairly clear. Most people who responded that they feel that discussing a mental health disorder would have a negative impact also are not comfortable discussing a mental health issue with a supervisor. The reverse is also true, most people who responded that they do not feel that discussing a mental health disorder would have a negative impact also indicate that they would be comfortable discussing a mental health issue with their supervisor. For those who responded that there maybe a negative impact, most also answered that they may be comfortable discussing a mental health disorder with their supervisor (with “no” being the second most common response for that group). Therefore, any outreach that we do or program that we plan needs to keep this in mind.

## Comfort with taking to Coworkers About Mental Health

This part of the analysis looks at answers to the question “Would you feel comfortable discussing a mental health disorder with your coworkers” by respondent demographics as well as their reports of several company characteristics. This is important because if an employee has a mental health issue, they should ideally be comfortable discussing it with a coworkers in case they need additional assistance. In addition, we may want to consider training employees on how to handle instances where coworkers reach out to them about mental health.

### Answers by Gender

#Create subset  
DiscMHCoworkerbyGender <- table(health$Would.you.feel.comfortable.discussing.a.mental.health.disorder.with.your.coworkers., health$gender)

#### Crosstab Table

kable(DiscMHCoworkerbyGender)

|  |  |  |  |
| --- | --- | --- | --- |
|  | Female | Male | Other |
| Maybe | 107 | 360 | 12 |
| No | 99 | 287 | 6 |
| Yes | 65 | 200 | 10 |

#### Proportions Table

kable(prop.table(DiscMHCoworkerbyGender, 2))

|  |  |  |  |
| --- | --- | --- | --- |
|  | Female | Male | Other |
| Maybe | 0.3948339 | 0.4250295 | 0.4285714 |
| No | 0.3653137 | 0.3388430 | 0.2142857 |
| Yes | 0.2398524 | 0.2361275 | 0.3571429 |

#### Chi-Squared

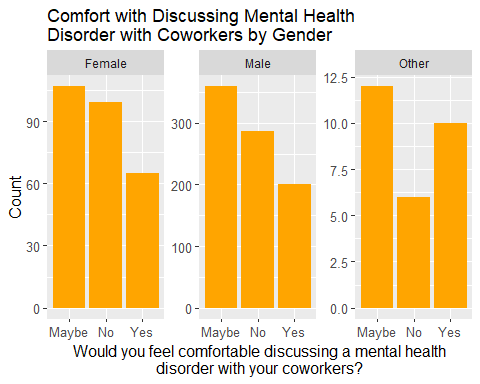
chisq.test(DiscMHCoworkerbyGender)

Pearson's Chi-squared test  
  
data: DiscMHCoworkerbyGender  
X-squared = 3.9002, df = 4, p-value = 0.4197

Our null hypothesis is that there is that reported comfort discussing mental health with coworkers and gender are independent. Our alternative hypothesis is that there is a relationship between these two variables. Because the sensitivity of this test to larger sample sizes we are setting the p value to .01 (Townsley, 2002). The chi-squared is very large. Therefore, we cannot reject the null hypothesis that these categories are independent.

#### Visualizing the Data

DiscMHCoworkerbyGender\_viz <- ggplot(health, aes(x = health$Would.you.feel.comfortable.discussing.a.mental.health.disorder.with.your.coworkers.))+geom\_bar(fill = "orange")+facet\_wrap(health$gender ~., scales="free\_y") + ggtitle("Comfort with Discussing Mental Health\nDisorder with Coworkers by Gender") + xlab("Would you feel comfortable discussing a mental health\ndisorder with your coworkers?") + ylab("Count") + theme(axis.title=element\_text(size=12), axis.text=element\_text(size=10))  
DiscMHCoworkerbyGender\_viz



#### Interpret Results

The chi-squared test appears to indicate that there is not a significant relationship between gender and comfort with discussing mental health with coworkers. However, these results may be impacted by the fact that there are fewer respondents who are female or another gender than men. However, a higher proportion of female respondents indicate that they do not feel comfortable discussing a mental health disorder with coworkers. However, across all genders the most common response is “maybe.” This indicates that a lot may depend on their relationships with individual coworkers (e.g., they may feel comfortable with discussing with a specific coworker). Interestingly, as stated above, most respondents across genders indicated that they would be comfortable discussing a mental health disorder with their supervisors.

### Answers by Age Group

#Create subset  
DiscMHCoworkerbyAge <- table(health$Would.you.feel.comfortable.discussing.a.mental.health.disorder.with.your.coworkers., health$AgeGroup)

#### Crosstab Table

kable(DiscMHCoworkerbyAge)

|  |  |  |  |
| --- | --- | --- | --- |
|  | 17-37 | 37-57 | 57-77 |
| Maybe | 338 | 139 | 2 |
| No | 272 | 116 | 4 |
| Yes | 197 | 72 | 6 |

#### Proportions Table

kable(prop.table(DiscMHCoworkerbyAge, 2))

|  |  |  |  |
| --- | --- | --- | --- |
|  | 17-37 | 37-57 | 57-77 |
| Maybe | 0.4188352 | 0.4250765 | 0.1666667 |
| No | 0.3370508 | 0.3547401 | 0.3333333 |
| Yes | 0.2441140 | 0.2201835 | 0.5000000 |

#### Chi-Squared

chisq.test(DiscMHCoworkerbyAge)

Warning in chisq.test(DiscMHCoworkerbyAge): Chi-squared approximation may  
be incorrect

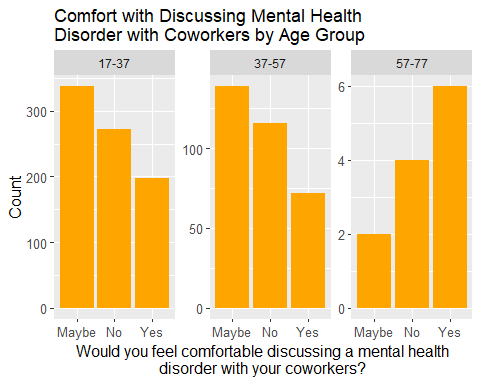
Pearson's Chi-squared test  
  
data: DiscMHCoworkerbyAge  
X-squared = 6.0422, df = 4, p-value = 0.196

Our null hypothesis is that there is that reported comfort discussing mental health with coworkers and age are independent. Our alternative hypothesis is that there is a relationship between these two variables. Because the sensitivity of this test to larger sample sizes we are setting the p value to .01 (Townsley, 2002)

We get a warning that the approximation may be incorrect, likely because because of the small number of respondents in our survey in the 57-77 age group (Townsley, 2002). However, the large p value indicates that that the differences in perceived ease of requesting leave between age may not be statistically significant (or we do not have enough information because of the small number of people in the 57-77 age group in our dataset).

#### Visualizing the Data

DiscMHCoworkerbyAge\_viz <- ggplot(health, aes(x = health$Would.you.feel.comfortable.discussing.a.mental.health.disorder.with.your.coworkers.))+geom\_bar(fill = "orange")+facet\_wrap(health$AgeGroup ~., scales="free\_y") + ggtitle("Comfort with Discussing Mental Health\nDisorder with Coworkers by Age Group") + xlab("Would you feel comfortable discussing a mental health\ndisorder with your coworkers?") + ylab("Count") + theme(axis.title=element\_text(size=12), axis.text=element\_text(size=10))  
DiscMHCoworkerbyAge\_viz



#### Interpret Results

Members of the 17-37 and 37-57 age group have a similar pattern. The majority of respondents indicated that they might feel comfortable discussing a mental health disorder with coworkers, with the second most common response being that they would not. When I looked at the comfort level of discussing a mental health disorder with supervisors, the results were different. Interestingly, as stated above, the majority of respondents in the 17-37 and 37-57 age groups responded that they would feel comfortable discussing a mental health disorder with their supervisor. Therefore, there may be a higher comfort level discussing mental health disorders with supervisors than other coworkers (or respondents’ comfort level may vary by the coworker).

### Answers by Company Size

#Create subset  
DiscMHCoworkerbyCompSize <- table(health$Would.you.feel.comfortable.discussing.a.mental.health.disorder.with.your.coworkers., health$CompanySize)

#### Crosstab Table

#Create crosstab table  
kable(DiscMHCoworkerbyCompSize)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 1-5 | 6-25 | 26-100 | 100-500 | 500-1000 | More than 1000 |
| Maybe | 20 | 83 | 140 | 110 | 32 | 94 |
| No | 19 | 75 | 84 | 79 | 21 | 114 |
| Yes | 21 | 52 | 68 | 59 | 27 | 48 |

#### Proportions Table

#Create proportions table  
kable(prop.table(DiscMHCoworkerbyCompSize, 2))

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | 1-5 | 6-25 | 26-100 | 100-500 | 500-1000 | More than 1000 |
| Maybe | 0.3333333 | 0.3952381 | 0.4794521 | 0.4435484 | 0.4000 | 0.3671875 |
| No | 0.3166667 | 0.3571429 | 0.2876712 | 0.3185484 | 0.2625 | 0.4453125 |
| Yes | 0.3500000 | 0.2476190 | 0.2328767 | 0.2379032 | 0.3375 | 0.1875000 |

#### Chi-Squared

#Chi-Squared (the p value is small)  
chisq.test(DiscMHCoworkerbyCompSize)

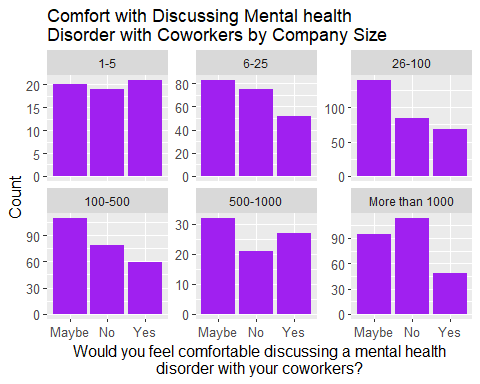
Pearson's Chi-squared test  
  
data: DiscMHCoworkerbyCompSize  
X-squared = 27.848, df = 10, p-value = 0.001909

Our null hypothesis is that there is that reported comfort discussing mental health with coworkers and age are independent. Our alternative hypothesis is that there is a relationship between these two variables. Because the sensitivity of this test to larger sample sizes we are setting the p value to .01 (Townsley, 2002)

The chi-squared test shows a p value that is less than 0.01. Therefore, we can reject the null hypothesis (that there is no relationship between these variables). This indicates that there may be a relationship between company size and comfort level with talking to coworkers about a mental health issue.

#### Visualizing the Data

DiscMHCoworkerbyCompSize\_viz <- ggplot(health, aes(x = health$Would.you.feel.comfortable.discussing.a.mental.health.disorder.with.your.coworkers.))+geom\_bar(fill = "purple")+facet\_wrap(health$CompanySize ~., scales="free\_y") + ggtitle("Comfort with Discussing Mental health\nDisorder with Coworkers by Company Size") + xlab("Would you feel comfortable discussing a mental health\ndisorder with your coworkers?") + ylab("Count") + theme(axis.title=element\_text(size=12), axis.text=element\_text(size=10))  
DiscMHCoworkerbyCompSize\_viz



#### Interpret Results

Respondents from companies with 1-5 employees are roughly evenly split among the three responses (yes, no, maybe) when asked if they would feel comfortable discussing a mental health disorder with coworkers. Results are similar with companies for 6-25 employees, but a lower proportion of employees in that category indicate they would be comfortable. For the 26-100 and 100-500 employee companies, most indicate that they may be comfortable with “no” being the second most common response. Respondents from companies with 500-1000 employees also select maybe the most often, but “yes” is the second most common response. Finally, respondents from companies with more than 1000 employees select the “no” answer the most often, with maybe being the second most common. The results indicate that there may be a relationship between company size and willingness to talk to coworkers about mental illness. However, the variation in responses also indicates that the reasons for the comfort, discomfort, or uncertainty may vary by company.

### Answers by If They Ever Saw A Poor Response to Mental Health Issue

#create subset  
DiscMHCoworkerbyPoorResponse <- table(health$Would.you.feel.comfortable.discussing.a.mental.health.disorder.with.your.coworkers., health$Have.you.observed.or.experienced.an.unsupportive.or.badly.handled.response.to.a.mental.health.issue.in.your.current.or.previous.workplace.)

#### Crosstab Table

kable(DiscMHCoworkerbyPoorResponse)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Maybe/Not sure | N/A | No | Yes, I experienced | Yes, I observed |
| Maybe | 132 | 19 | 195 | 58 | 75 |
| No | 105 | 22 | 150 | 39 | 76 |
| Yes | 41 | 11 | 146 | 35 | 42 |

#### Proportions Table

kable(prop.table(DiscMHCoworkerbyPoorResponse, 2))

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Maybe/Not sure | N/A | No | Yes, I experienced | Yes, I observed |
| Maybe | 0.4748201 | 0.3653846 | 0.3971487 | 0.4393939 | 0.3886010 |
| No | 0.3776978 | 0.4230769 | 0.3054990 | 0.2954545 | 0.3937824 |
| Yes | 0.1474820 | 0.2115385 | 0.2973523 | 0.2651515 | 0.2176166 |

#### Chi-Squared

chisq.test(DiscMHCoworkerbyPoorResponse)

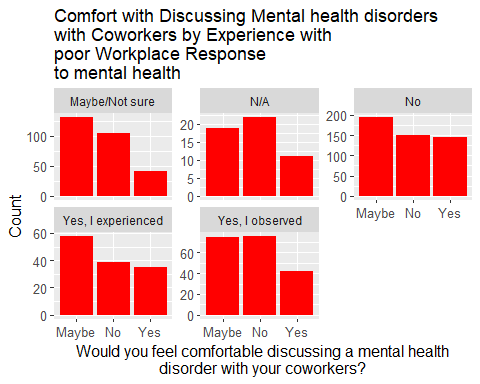
Pearson's Chi-squared test  
  
data: DiscMHCoworkerbyPoorResponse  
X-squared = 27.416, df = 8, p-value = 0.0005991

Our null hypothesis is that there is that reported comfort discussing mental health with coworkers and experience with a poor response to a mental health issue (current or prior employer) are independent. Our alternative hypothesis is that there is a relationship between these two variables. Because the sensitivity of this test to larger sample sizes we are setting the p value to .01 (Townsley, 2002)

The chi-squared test shows a p value that is less than .01. Therefore, we can reject the null hypothesis (that there is no relationship between these variables). This indicates that there may be a relationship between experience with a poor response to a mental health issue with talking to coworkers about a mental health issue.

#### Visualizing the Data

DiscMHCoworkerbyPoorResponse\_viz <- ggplot(health, aes(x = health$Would.you.feel.comfortable.discussing.a.mental.health.disorder.with.your.coworkers.))+geom\_bar(fill="red")+facet\_wrap(health$Have.you.observed.or.experienced.an.unsupportive.or.badly.handled.response.to.a.mental.health.issue.in.your.current.or.previous.workplace. ~., scales="free\_y") + ggtitle("Comfort with Discussing Mental health disorders\nwith Coworkers by Experience with\npoor Workplace Response\nto mental health") + xlab("Would you feel comfortable discussing a mental health\ndisorder with your coworkers?") + ylab("Count") + theme(axis.title=element\_text(size=12), axis.text=element\_text(size=10))  
DiscMHCoworkerbyPoorResponse\_viz



#### Interpret Results

When asked about experience with a poorly handled response to mental health, respondents across all categories except for “N/A” (which had relatively few respondents), also answered that they maybe comfortable discussing a mental health disorder with coworkers. Interestingly, those who stated “Yes, I observed” a poor response to mental health in the workplace were evenly split between the maybe and no answers when also asked if they would feel comfortable discussing a mental health disorder with coworkers. However, those who stated that they experienced a poorly handled response selected “Maybe” most often when also asked if they would feel comfortable discussing a mental health disorder with coworkers. The remainder were about evenly split between the “no” and “yes” answers. This may indicate that observing a poorly handled response to a mental health issue in the workplace may have a greater negative effect on willingness to discuss with coworkers than those experiencing it.

### Answers by If They Think Coworkers Would View Them Negatively If They Knew You Suffered From a Mental Health Issue

#Modify Labels for perception of mental health disclosure impact by coworkers  
health$negimpco <- factor(health$Do.you.think.that.team.members.co.workers.would.view.you.more.negatively.if.they.knew.you.suffered.from.a.mental.health.issue.,  
levels = c("Maybe", "No, I don't think they would", "No, they do not", "Yes, I think they would", "Yes, they do"),  
labels = c("Maybe", "No, don't think so", "No, they don't", "Yes, think so", "Yes, they do"))  
#create subset  
DiscMHCoworkerbyNegView <- table(health$Would.you.feel.comfortable.discussing.a.mental.health.disorder.with.your.coworkers., health$Do.you.think.that.team.members.co.workers.would.view.you.more.negatively.if.they.knew.you.suffered.from.a.mental.health.issue.)

#### Crosstab Table

kable(DiscMHCoworkerbyNegView)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Maybe | No, I don’t think they would | No, they do not | Yes, I think they would | Yes, they do |
| Maybe | 234 | 137 | 7 | 87 | 14 |
| No | 127 | 51 | 1 | 201 | 12 |
| Yes | 100 | 115 | 36 | 20 | 4 |

#### Proportions Tabale

kable(prop.table(DiscMHCoworkerbyNegView, 2))

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Maybe | No, I don’t think they would | No, they do not | Yes, I think they would | Yes, they do |
| Maybe | 0.5075922 | 0.4521452 | 0.1590909 | 0.2824675 | 0.4666667 |
| No | 0.2754881 | 0.1683168 | 0.0227273 | 0.6525974 | 0.4000000 |
| Yes | 0.2169197 | 0.3795380 | 0.8181818 | 0.0649351 | 0.1333333 |

#### Chi-Squared

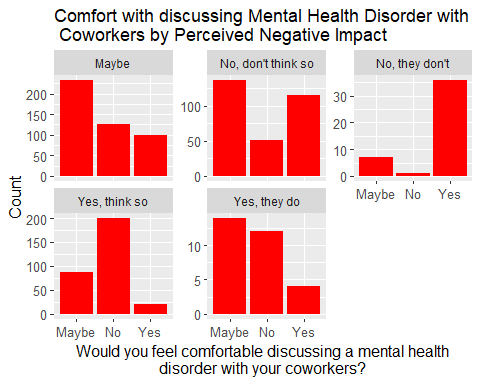
chisq.test(DiscMHCoworkerbyNegView)

Pearson's Chi-squared test  
  
data: DiscMHCoworkerbyNegView  
X-squared = 291.08, df = 8, p-value < 2.2e-16

Our null hypothesis is that there is that reported comfort discussing mental health with coworkers and experience with a poor response to a mental health issue (current or prior employer) are independent. Our alternative hypothesis is that there is a relationship between these two variables. Because the sensitivity of this test to larger sample sizes we are setting the p value to .01 (Townsley, 2002). The chi-squared is less than .01. Therefore, we can reject the null and say that there is a relationship between these two variables.

#### Creating the Visualization

DiscMHCoworkerbyNegView\_viz <- ggplot(health, aes(x = health$Would.you.feel.comfortable.discussing.a.mental.health.disorder.with.your.coworkers.))+geom\_bar(fill = "red")+facet\_wrap(health$negimpco ~., scales="free\_y") + ggtitle("Comfort with discussing Mental Health Disorder with\n Coworkers by Perceived Negative Impact") + xlab("Would you feel comfortable discussing a mental health\ndisorder with your coworkers?") + ylab("Count") + theme(axis.title=element\_text(size=12), axis.text=element\_text(size=10))  
DiscMHCoworkerbyNegView\_viz



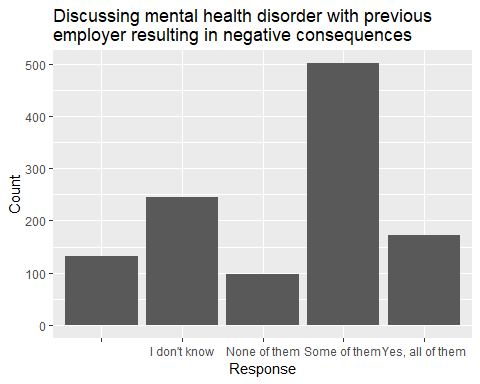
#### Interpretation of Results

In this visualization we see that perceived impact of discussing a mental health disorder with coworkers has a potentially significant impact on comfort level with discussing a mental health disorder with coworkers. For those who feel that they would be viewed negatively by coworkers if they knew the respondent had a mental health issue, they frequently answer that they would not feel comfortable discussing a mental health disorder with coworkers. For those who feel like they are negatively viewed by coworkers because of a mental health issue, appear to be evenly split between maybe being comfortable and not being comfortable. Meanwhile, for those who feel that they are not viewed more negatively because of a mental health disorder, they most frequently say they would be comfortable discussing mental health with coworkers. For those who did not think they would be viewed negatively, they most frequently select that they may be comfortable discussing a mental health disorder with other coworkers, with yes being the second most common response from that group. This indicates that we may want to consider providing training to employees about mental illness in the workplace.

### Analzing the relationship between perceiving the negative impact of mental illness at previous workplace and the impact at current workplace

#### Plotting the ggplot for the independent variable “Do you think that discussing a mental health disorder with previous employers would have negative consequences?”

ggplot(health, aes(x = health$Do.you.think.that.discussing.a.mental.health.disorder.with.previous.employers.would.have.negative.consequences.)) + geom\_bar() + ggtitle("Discussing mental health disorder with previous\nemployer resulting in negative consequences") + xlab("Response") + ylab("Count")



summary(health$Do.you.think.that.discussing.a.mental.health.disorder.with.previous.employers.would.have.negative.consequences.)

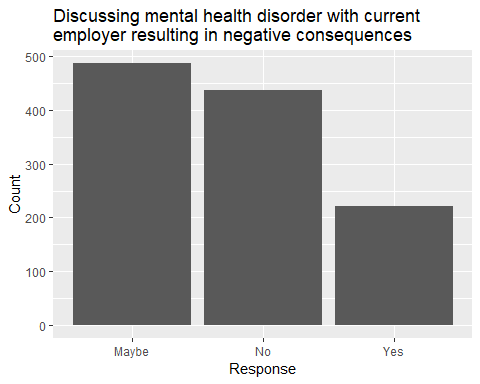
I don't know None of them Some of them   
 131 245 97 501   
Yes, all of them   
 172

#### Interpreting the results

From the above ggplot, we can observe that a majority of respondents(615) feel that discussing mental health disorders with some of their previous employers resulted in negative consequences for them. While a significant number of respondents responded that they do not know, we can conclude from the graph that the majority did experience negative consequences after discussing mental health disorders with their previous employers

### Plotting the ggplot for the dependent variable “Do you think that discussing a mental health disorder with your employer would have negative consequences?”

ggplot(health, aes(x = health$Do.you.think.that.discussing.a.mental.health.disorder.with.your.employer.would.have.negative.consequences.)) + geom\_bar() + ggtitle("Discussing mental health disorder with current\nemployer resulting in negative consequences") + xlab("Response") + ylab("Count")



summary(health$Do.you.think.that.discussing.a.mental.health.disorder.with.your.employer.would.have.negative.consequences.)

Maybe No Yes   
 487 438 221

#### Interpretation of results

From the above ggplot, we can see that a vast majority of respondents(487) are not sure if discussing mental health disorders with their current employers would have negative consequences; followed by about 438 respondents who do not think that it could lead to negative consequences

#### Chi-squared test

t1<-table(health$Do.you.think.that.discussing.a.mental.health.disorder.with.previous.employers.would.have.negative.consequences., health$Do.you.think.that.discussing.a.mental.health.disorder.with.your.employer.would.have.negative.consequences.)  
chisq.test(t1)

Pearson's Chi-squared test  
  
data: t1  
X-squared = 158.59, df = 8, p-value < 2.2e-16

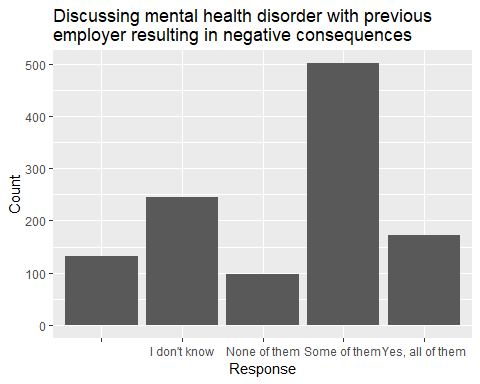
#### Interpret results

We can see that the p-value of the Chi-squared test is very small, meaning we can reject the hypothesis of independence. Thus there is a relationship between perceiving the negative impact of mental illness at previous workplace and the impact at current workplace

### Analzing the relationship between perceiving the negative impact of mental illness at previous workplace and the impact at current workplace

#### Plotting the ggplot for the independent variable “Do you think that discussing a mental health disorder with previous employers would have negative consequences?”

ggplot(health, aes(x = health$Do.you.think.that.discussing.a.mental.health.disorder.with.previous.employers.would.have.negative.consequences.)) + geom\_bar() + ggtitle("Discussing mental health disorder with previous\nemployer resulting in negative consequences") + xlab("Response") + ylab("Count")



summary(health$Do.you.think.that.discussing.a.mental.health.disorder.with.previous.employers.would.have.negative.consequences.)

I don't know None of them Some of them   
 131 245 97 501   
Yes, all of them   
 172

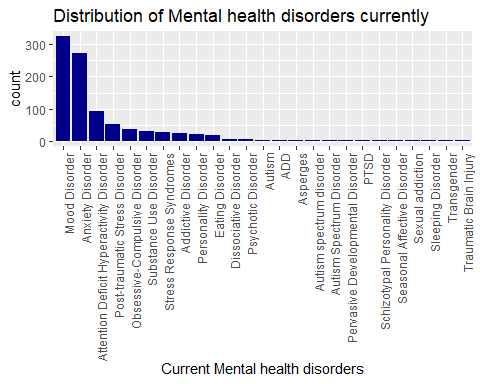
# Summary

According to the 2010 U.S. National Comorbidity Survey, about 18 percent of employed adults have had symptoms of a mental health disorder in the previous month (Harvard Medical School, 2010, para. 1). However, mental health stigma can inhibit people’s willingness to seek help (Harvard Medical School, 2010, para. 2). Research on employee resource programs in the workplace targeting mental health/substance abuse issues indicates that conducting outreach to workers on mental health as well as expanding benefits may result in more seeking help for mental health and/or substance abuse issues (Lindrooth, et al, 2005, pp. 1092, 1102; Lo Sasso, et al, 2006, pp. 366, 370; Deitz, et al, 2005, p. 313). Benefits could include improved employee mood/health, increased productivity, and lower absenteeism in the workplace (Harvard Medical School, 2010, para. 28-29; Lo Sasso, Rost, & Beck, 2006, p. 352).

Our target audience are Human Resources professionals who work for technology companies (or employ technology-focused workers) who are interested in learning more about mental illness in the workplace and are considering expanding wellness programs. For our analysis, we used the Open Sourcing Mental Illness dataset. The dataset is a 2016 voluntary survey done on mental health in the technology industry (Open Sourcing Mental Illness, LTD, 2016b). This dataset allows us to learn more about what mental health disorders are common, what other companies are doing, and how workers view mental health in the workplace. For our analysis we only included those who worked for a company (i.e., not self-employed). Out of 1,433 total entries, 1,146 were included. We cleaned the dataset by standardizing data, as well as correcting typos and formatting errors. During our analysis, we shortened some of the labels to make it easier to fit on visualizations and tables.

One plot we used in our analysis is the Distribution of Mental health disorders that respondents to the survey had. The visualization and the description of results can also be found in the “Distribution of Mental health disorders” subheading of Section 1.

p+xlab("Current Mental health disorders")+ggtitle("Distribution of Mental health disorders currently")



As indicated above, mood disorder and anxiety disorder are the most common in our analysis. Furthermore, it is clear that a significant number of people who responded to the survey have at least one mental health disorder. This analysis makes it clear that Human Resource professionals who are interested in expanding outreach need to review their policies regarding requesting leave and protection of employee anonymity when they are handling a mental health issue. Furthermore, the analysis shows that there is unease among many employees regarding the impact of disclosing a mental health disorder on their career. Therefore, it is likely necessary to provide training to employees and supervisors with the goal of reducing stigma around mental health.

# Who did what in the project

Monalisa -

Laura - For the work plan, I worked with the group to draft the sections on dataset used, audience, and research questions. I also reviewed a list of studies on mental health in the workplace and drafted summaries that can be used as a reference by the group. For the data cleaning documentation, I examined nine variables and noted if there were any issues. I worked with other group members to determine how to handle the variables. For the R Script, I did frequency tables related to my analyses and the variables I examined for data cleaning. For the R Plot Draft, I created visualizations and interpretations for my analyses (in section 3, the portions on anonymity policies, ease of taking leave, comfort discussing with coworkers and supervisors). For the presentation, I created a template for the presentation that included what was needed and placeholder slides for the analyses the group did. I also did slides related to my analysis (anonymity policies, ease of taking leave, comfort discussing with coworkers and supervisors) and used the findings to draft some recommendations.For the final package, I did the first draft of the summary as well as revised and cleaned up the analyses that I worked on.

Shruti - Collaborated with the team members to decide upon work plan and the tasks. For data cleaning, I examined nine variables for any issues. Worked on the age variable where I replaced the meaningless ages like 3, 110 with the average age. I exmained all the answers to the gender variable as they were subjective responses and divided them into Male, Female and others accordingly. For the descriptive analysis, I worked with the nine variables using summary and tables. For the visualization and plotting, I worked on the distribution of mental disorder. I analysed the responses to come up with a way to split it. For instance, the responses were of the format - Anxiety Disorder (Generalized, Social, Phobia, etc)|Mood Disorder (Depression, Bipolar Disorder, etc), I split them and separated them using R code to just the higher categories like anxiety Disorder, Mood Disorder to visualise it. For ease of visualisations using age, I created age groups (17-37, 37-57 etc). I visualised the distribution of mental illness across gender and age. I also did the visualisation and interpretations on negative consequences on discussing physical vs mental disorders. I documented my parts in the R markdown file. I worked on my parts for the presentation. For final package, I cleaned up the analaysis and visualizations that I worked on as well.

# Works Cited

Deitz, D., Cook, R., & Hersch, R. (2005). Workplace Health Promotion and Utilization of Health Services: Follow-up Data Findings. The Journal of Behavioral Health Services & Research, 32(3), 306-319. Retrieved October 21, 2018 from <https://doi-org.proxy-um.researchport.umd.edu/10.1007/BF02291830>

Harvard Medical School (2010). Mental health problems in the workplace [Newsletter]. Retrieved December 8, 2018, from <https://www.health.harvard.edu/newsletter_article/mental-health-problems-in-the-workplace>.

Lo Sasso, A. T., Lindrooth, R. C., Lurie, I. Z., & Lyons, J. S. (2006). Expanded Mental Health Benefits and Outpatient Depression Treatment Intensity. Medical Care, 44(4), 366-372. Retrieved October 21, 2018 from <https://doi-org.proxy-um.researchport.umd.edu/10.1097/01.mlr.0000204083.55544.f8>

Lindrooth, R. C., Lo Sasso, A. T., & Lurie, I. Z. (2005). The Effect of Expanded Mental Health Benefits on Treatment Initiation and Specialist Utilization. Health Services Research, 40(4), 1092-1107. Retrieved October 21, 2018 from <https://doi-org.proxy-um.researchport.umd.edu/10.1111/j.1475-6773.2005.00406.x>

Lo Sasso, A. T., Rost, K., & Beck, A. (2006). Modeling the Impact of Enhanced Depression Treatment on Workplace Functioning and Costs: A Cost-Benefit Approach. Medical Care, 44(4), 352-358. Retrieved October 21, 2018 from <https://doi-org.proxy-um.researchport.umd.edu/10.1097/01.mlr.0000204049.30620.1e>

Milne, S. H., Blum, T. C., & Roman, P. M. (1994). Factors Influencing Employees Propensity to Use an Employee Assistance Program. Personnel Psychology, 47(1), 123-145. Retrieved October 24, 2018 from <http://search.ebscohost.com.proxy-um.researchport.umd.edu/login.aspx?direct=true&db=bth&AN=9411113184&site=ehost-live>

Open Sourcing mental Illness, LTD (2016a). OSMI (Version 1) [Data file]. Retrieved September 7, 2018, from <https://osmihelp.org/research>

Open Sourcing Mental Illness, LTD (2016b). osmi-survey-2016\_1479139902 [Metadata File]. Retrieved September 19, 2018, from <https://osmihelp.org/research>.

Townsley, E.R. (2002). Some limitations of chi-square. Retrieved December 6, 2018 from <https://www.mtholyoke.edu/courses/etownsle/qr/Chi%20square%20limitations.htm>.