**Mental Health in the Workplace**



# Section I

The data included in this study is provided by Open Sourcing Mental Illness, LTD. The survey was conducted to determine attitudes towards mental health disorders with tech workers. Survey data can be found [here](https://www.kaggle.com/osmi/mental-health-in-tech-2016). The data contains 1,433 observations and 63 variables relating to current employer and previous employers.

As mental health starts to move from the shadows of everyday life to the forefront of social media, traditional news outlets, and even at the dinner table, the team decided to put several questions to analysis to better understand issues around mental health.

# Section II

The team set out to answer several questions:

1. Do certain technology job roles line up with specific mental health diagnoses?
2. Is mental healthcare coverage higher in certain states?
3. Are there higher reports of mental health issues for remote workers?
4. Do workes with substance abuse disorder feel more or less willing to discuss their issue with their supervisor when compared to other disorders?
5. Do workers with substance abuse disorder report that their anonymity was more or less protected when compmared to other mental health issues?
6. Do workers with substance abuse disorder report that they have more experience with unsuportive response when compared to other mental health issues?
7. Do workers with substance abuse disorder report that they experience negative consequences for divulging mental health issues?

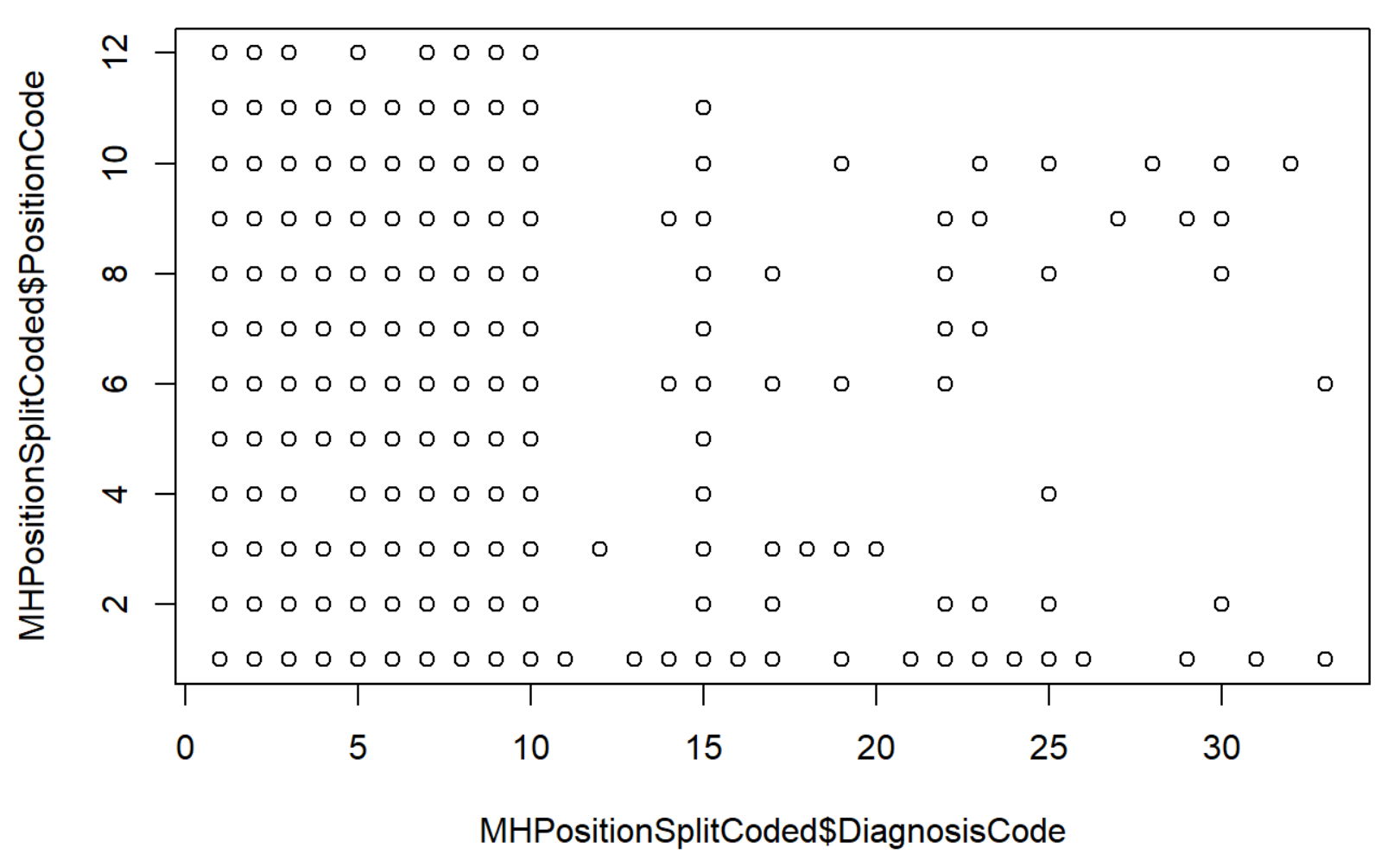
Please see appendix A for R code

# Job Analysis

1. Do certain technology job roles line up with specific mental health diagnoses?

Our team was interested in determining if there is a correlation between mental health diagnoses and technology job roles. For example, is there a higher prevalence of anxiety disorder attributed to back-end developers? If we can find a correlation, perhaps we can better target treatment based on a person’s job role.

By analyzing the data, we do not find any significant correlation between job role and mental health diagnoses. We will need to determine other avenues for targeting treatment.



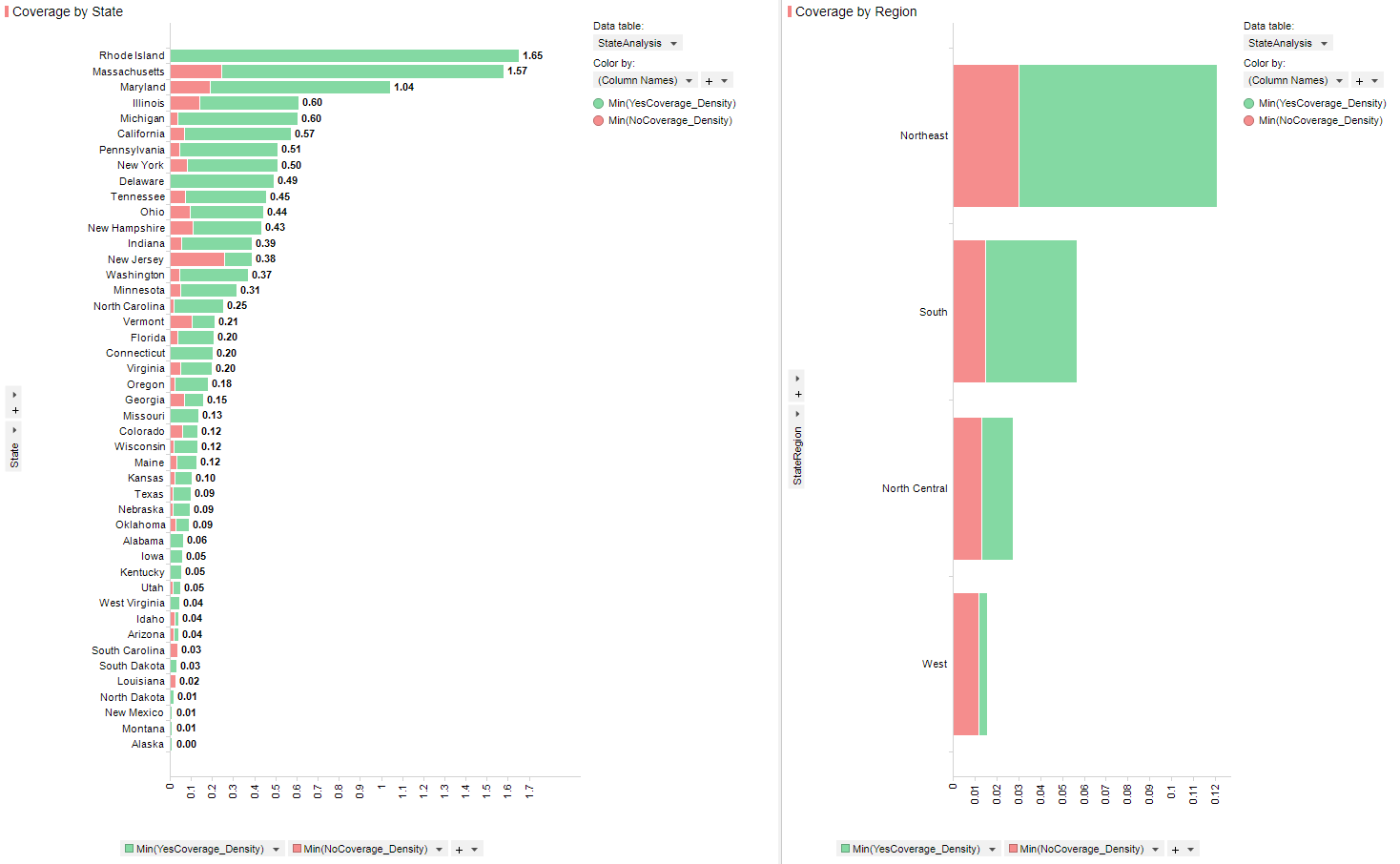
# State Analysis

1. Is mental healthcare coverage higher in certain states?

The team subsetted the data to only rows that had state names and for rows that had identified if the company provided mental health coverage. From this initial exclusion of data the data was subsetted from 1,433 rows of data to 515 rows of data which constitutes 36% of useable data.

From the limited data set we can see that most companies offer mental health insurance. When normalizing the data to the per 1000 square miles of each state, we see that Rhode Island, Massachusetts, and Maryland have thie highest coverage of mental health insurance. This coincides when charting the coverage by region, which shows the Northeast region of the US having the highest coverage.

Though the analysis shows the Northeast having the most insured individuals with mental health coverage, the results should be used cautiously as the bulk of the observations are unusable due to a lack of response.

*District of Columbia was removed as it is not a state.*

# Remote Working Analysis

1. Are there higher reports of mental health issues for remote workers?

Table of Contents

[Section I 2](#_Toc522445262)

[Section II 2](#_Toc522445263)

[Job Analysis 2](#_Toc522445264)

[State Analysis 3](#_Toc522445265)

[Remote Working Analysis 3](#_Toc522445266)

[Substance Abuse Analysis 5](#_Toc522445267)

[Discussing issues with supervisors 5](#_Toc522445268)

[Anonymity Protected 6](#_Toc522445269)

[Experience unsupportive responses 7](#_Toc522445270)

[Negative consequences 8](#_Toc522445271)

[Conclusion 9](#_Toc522445272)

[GitHub Location 11](#_Toc522445273)

[R Code 11](#_Toc522445274)

The team subsetted the data to include the columns of ‘WorkRemotely’ and ‘NUM\_MHCurrently’. These two columns are used to indicate if the observation has a mental health issue or not. The team provided multiple ways to cross check the results to ensure that the results were verified.

A table is produced that aggregates the groups by the categorical variable ‘WorkRemotely’

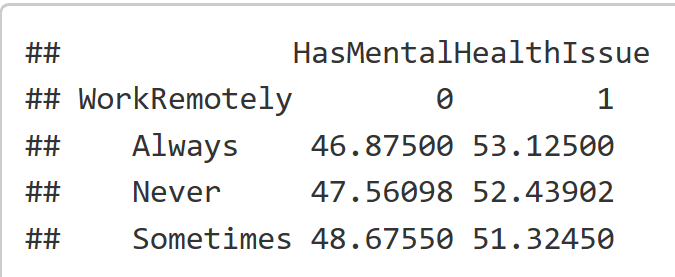
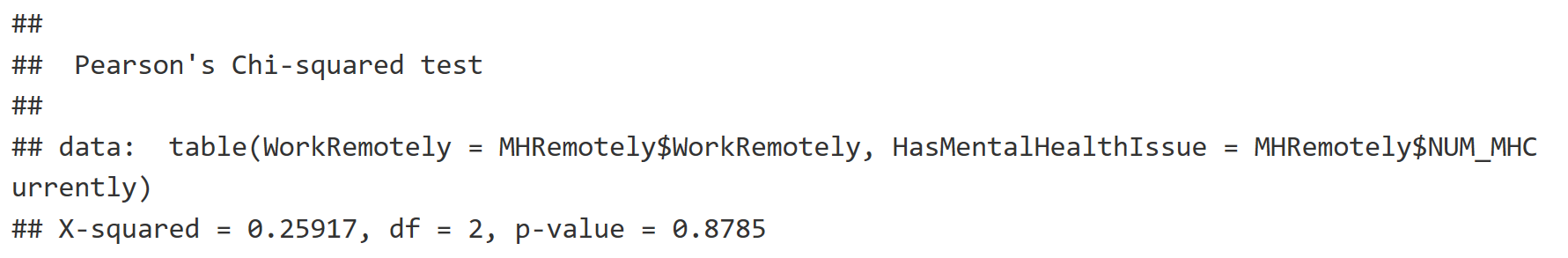
1. Always
2. Never
3. Sometimes

The binary responses of ‘NUM\_MHCurrently’ are:

1. No = 0
2. Yes = 1

A proportion table is provided that shows the percentage of receiving a mental health issue by category. From first look it seems that there is no significant chance of receiving a mental health issue from any of the categorical variables. To test and confirm this hypothesis the team decided to do a chi square test, a regression test, and a simple ANOVA test. For each of these tests the p-value is greater than alpha of 0.05 indicating that there is no statistically significant chance of receiving a mental health issue from working at home, working at the office and at home, or working in the office full time.

From the data set provided the team concluded that there is no significant chance of being diagnosed with a mental health issue from individuals that work at the office, home, or both.

# Substance Abuse Analysis

In the next section the questions will pertain to individuals who have confirmed a mental health issue in the survey.

## Discussing issues with supervisors

Do workers with substance abuse disorder feel more or less willing to discuss their issue with their supervisor when compared to other disorders?

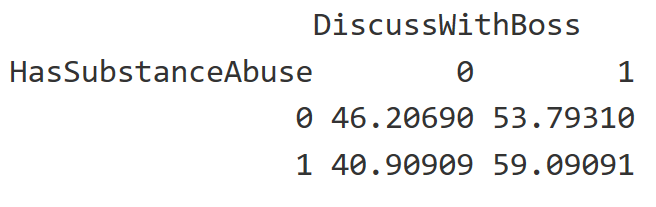
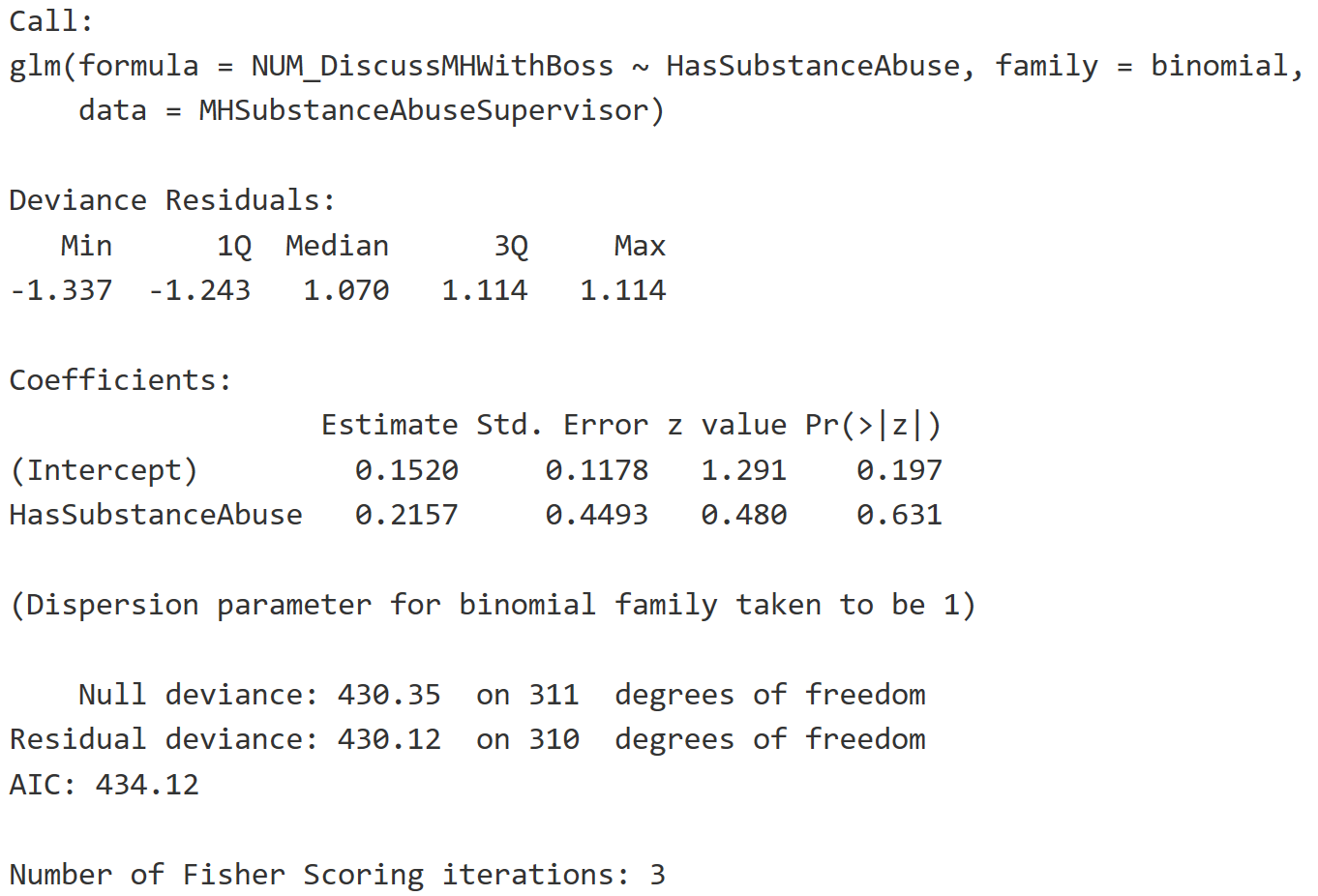
The team subsetted the data to include the columns of ‘HasSubstanceAbuse’ and ‘NUM\_DiscussMHWithBoss’. The NUM\_DiscussMHWithBoss column is used to indicate if the observation is willing to discuss their mental health issue with their direct supervisor. The HasSubstanceAbuse column indicates if the observation’s mental health issue relates to substance abuse.

A table is produced that aggregates the groups by the binary variable ‘HasSubstanceAbuse’

The binary responses of ‘NUM\_DiscussMHWithBoss’ are:

1. No = 0
2. Yes = 1

A proportion table is provided that shows the percentage of workers with mental health issues that would discuss the issue with their direct supervisor. It appears that workers with a substance abuse disorder are more likely to discuss the situation with their boss when compared to workers with other mental health issues. 59.1% of workers with a substance abuse problem would discuss the issue with their supervisor compared to 53.8% of workers with other mental health issues.

## Anonymity Protected

1. Do workers with substance abuse disorders report that their anonymity was more or less protected when compared to other mental health issues?

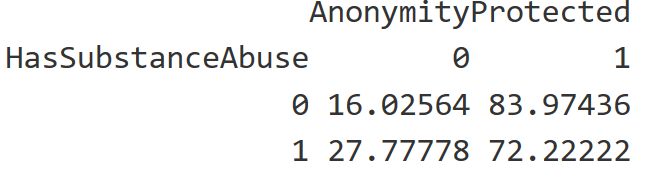
The team subsetted the data to include the columns of ‘HasSubstanceAbuse’ and ‘NUM\_AnonymityProtected’. The NUM\_AnonymityProtected column is used to indicate if the observation reports their anonymity was protected after divulging their mental health issue. The HasSubstanceAbuse column indicates if the observation’s mental health issue relates to substance abuse.

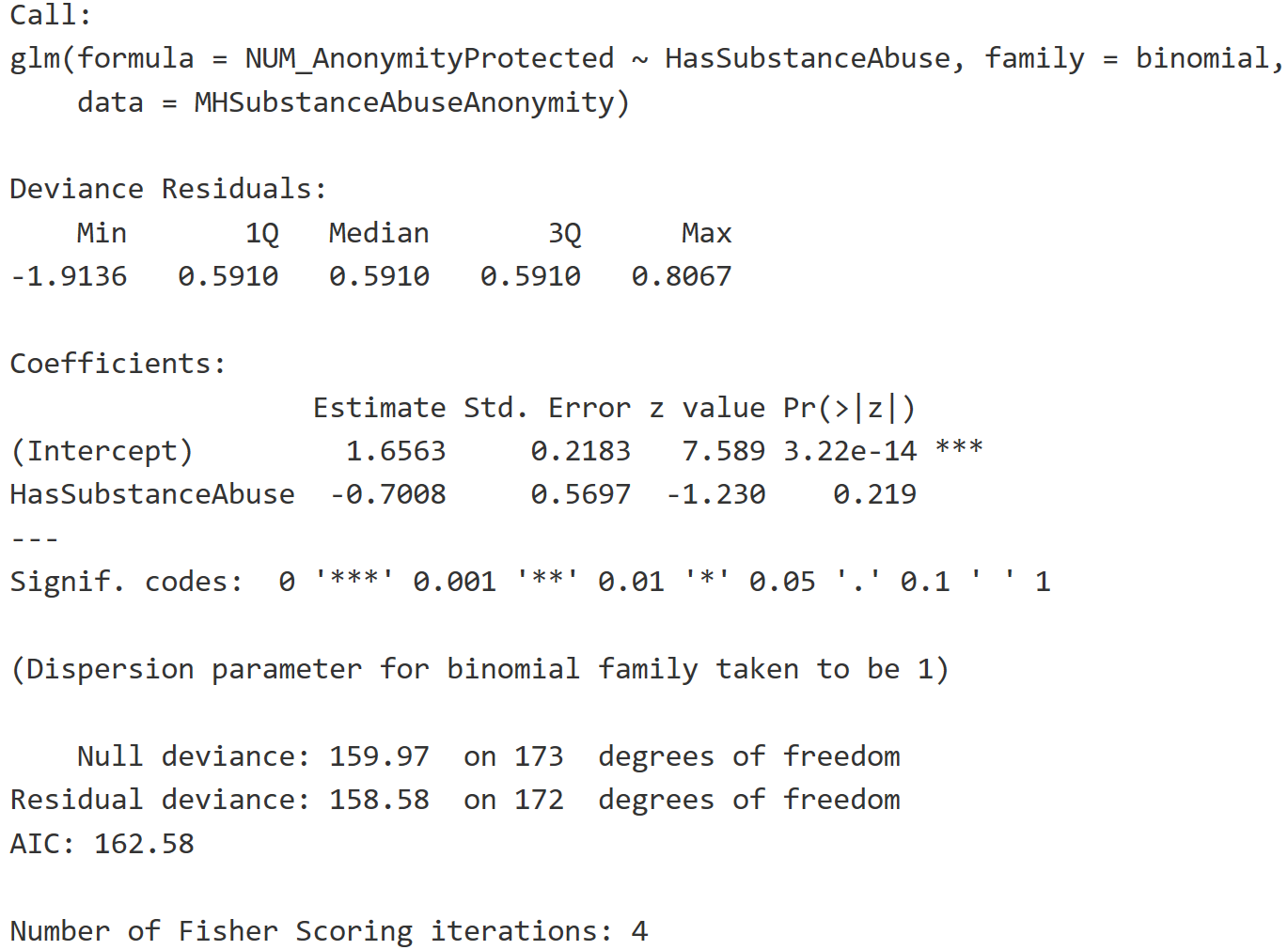
A table is produced that aggregates the groups by the binary variable ‘HasSubstanceAbuse’

The binary responses of ‘NUM\_AnonymityProtected’ are:

1. No = 0
2. Yes = 1

A proportion table is provided that shows the percentage of workers with mental health issues that believe their anonymity was protected after divulging a mental health issue. It appears that workers with a substance abuse disorder are less likely to report their anonymity was protected when compared to other mental health issues. 72.22% of workers with substance abuse disorder report their anonymity was protected compared to 83.97% of workers with other mental health issues. It is encouraging to see that workers with both substance abuse disorders and other mental health issues report a high degree of anonymity. However, further analysis is warranted to determine why workers with a substance abuse disorder feel their anonymity was less protected when compared to other mental health issues.





## Experience unsupportive responses

1. Do workers with substance abuse disorder report that they have experience with unsupportive responses when compared to other mental health issues?

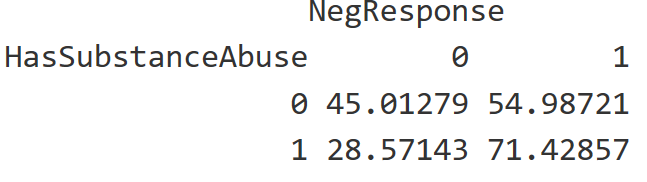
The team subsetted the data to include the columns of ‘HasSubstanceAbuse’ and ‘NUM\_NegResponseWithMH’. The NUM\_NegResponseWithMH column is used to indicate if the observation reports observing or experiencing an unsupportive response with mental health issues in the workplace. The HasSubstanceAbuse column indicates if the observation’s mental health issue relates to substance abuse.

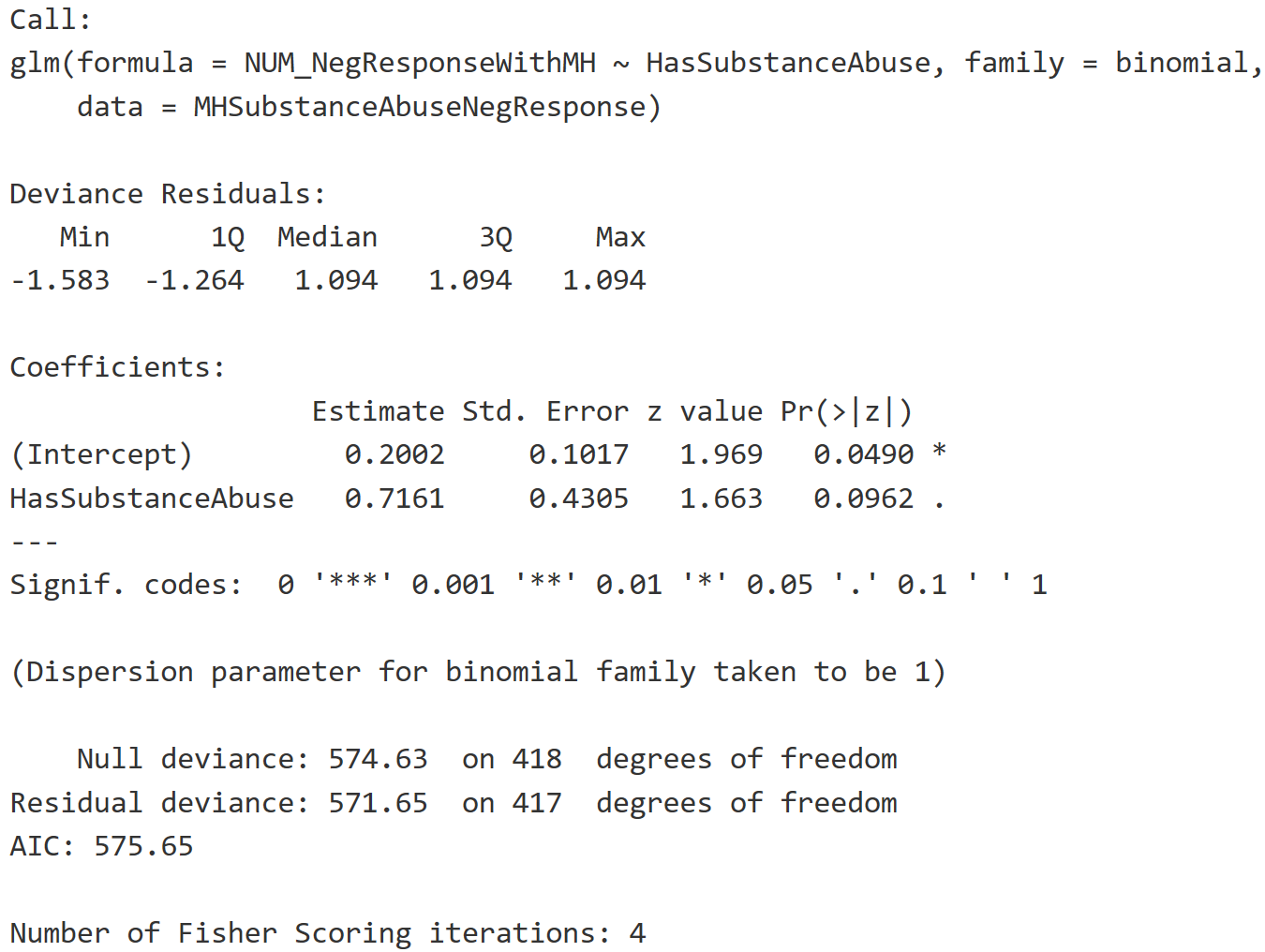
A table is produced that aggregates the groups by the binary variable ‘HasSubstanceAbuse’

The binary responses of ‘NUM\_NegResponseWithMH’ are:

1. No = 0
2. Yes = 1

A proportion table is provided that shows the percentage of workers with mental health issues that report observing or experiencing an unsuportive response to a mental health issue. It appears that workers with a substance abuse disorder are more likely to observe or experience an unsuportive reponse when compared to other mental health issues. 71.43% of workers with a substance abuse disorder report an unsuportive response compared to 54.99% of workers with other mental health issues. It is disturbing to see such a high report of unsuportive responses for workers with substance abuse disorders.





## Negative consequences

1. Do workers with substance abuse disorder report that they experience more negative consequences when divulging a mental health compared to other mental health issues?

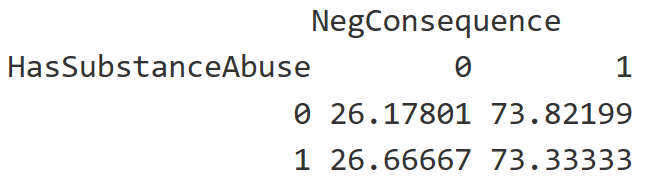
The team subsetted the data to include the columns of ‘HasSubstanceAbuse’, ‘CAT\_MHHurtCareer’, ‘NUM\_ObsNegOpenWithMH’, and ‘NUM\_CoWorkersViewYouNegKnewMH.’ We utilized 3 data points (‘CAT\_MHHurtCareer’, ‘NUM\_ObsNegOpenWithMH’, and ‘NUM\_CoWorkersViewYouNegKnewMH’) as an aggregation to generate the “HasNegativeConsequence” column. If any of these questions were answered positively, we set the HasNegativeConsequence column to 1. Since many factors could be interpreted as negative consequences, we felt this was the best way to gauge the situation. The HasSubstanceAbuse column indicates if the observation’s mental health issue relates to substance abuse.

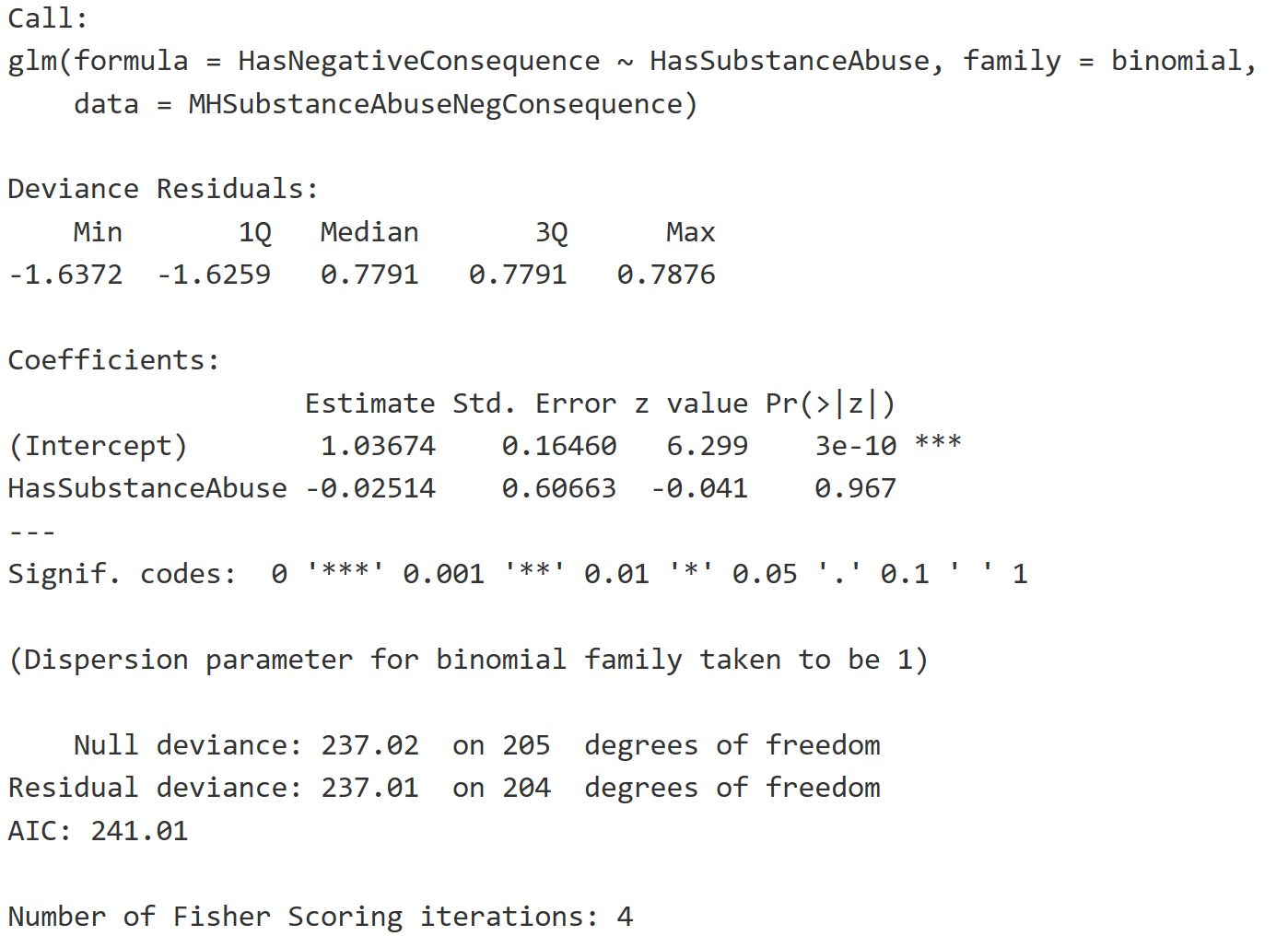
A table is produced that aggregates the groups by the binary variable ‘HasSubstanceAbuse’

The binary responses of ‘HasNegativeConsequence’ are:

1. No = 0
2. Yes = 1

A proportion table is provided that shows the percentage of workers with mental health issues that report experiencing negative consequences after divulging a mental health issue. Interestingly enough, both employees with substance abuse disorder and other mental health issues experienced similar rates of negative consequences. 73.33% of workers with substance abuse disorder reported experiencing negative consequences compared to 73.82% of workers with other mental health issues. It is unfortunate so see such a high percentage of workers reporting experiencing negative consequences after divulging a mental health issue. Further investigation is warranted to determine the high rate of negative consequences.





# Conclusion

The team has done initial analysis around the questions originally poised. The data set has lead into several interesting outcomes, however due to the lack of information around the survey the team is unable to apply generalizations to a wider population. The Open Source Mental Illness (OSMI) group does not provide information on how the survey was distributed and how the responses were recorded. Without knowing the full-size population or if the survey was randomly sent to individuals, the team is unable to apply any generalization to a population that is outside of this study.

Though we cannot make inferences to the population, this study does provide insight on additional aspects of mental health that should be investigated further. Items such as why workers with substance abuse disorders feel that their privacy was less protected is an aspect that could use additional investigation. The study has limited use but does provide a baseline to compare against other studies that could provide a longitudinal comparison.

# GitHub Location

https://github.com/djserna/6306-case-study-2

# R Code

#load necessary libraries for this analysis.

if (!require(plyr)) install.packages("plyr")

if (!require(DataExplorer)) install.packages("DataExplorer")

if (!require(sqldf)) install.packages("sqldf")

if (!require(Hmisc)) install.packages("Hmisc")

if (!require(tidyr)) install.packages("tidyr")

if (!require(dplyr)) install.packages("dplyr")

if (!require(car)) install.packages("car")

#Read in File

MHData <- read.csv("mental-heath-in-tech-2016\_20161114.csv", header=TRUE, sep=",")

#Update Column Names

colnames(MHData)[colnames(MHData)=='Are.you.self.employed.'] <- 'SelfEmployed'

colnames(MHData)[colnames(MHData)=='How.many.employees.does.your.company.or.organization.have.'] <- 'CompanySize'

colnames(MHData)[colnames(MHData)=='Is.your.employer.primarily.a.tech.company.organization.'] <- 'TechCompany'

colnames(MHData)[colnames(MHData)=='Is.your.primary.role.within.your.company.related.to.tech.IT.'] <- 'RoleIT'

colnames(MHData)[colnames(MHData)=='Does.your.employer.provide.mental.health.benefits.as.part.of.healthcare.coverage.'] <- 'ProvideMHCoverage'

colnames(MHData)[colnames(MHData)=='Do.you.know.the.options.for.mental.health.care.available.under.your.employer.provided.coverage.'] <- 'AwareOfCoverage'

colnames(MHData)[colnames(MHData)=='Has.your.employer.ever.formally.discussed.mental.health..for.example..as.part.of.a.wellness.campaign.or.other.official.communication..'] <- 'CompanyDiscussMH'

colnames(MHData)[colnames(MHData)=='Does.your.employer.offer.resources.to.learn.more.about.mental.health.concerns.and.options.for.seeking.help.'] <- 'CompanyOfferResources'

colnames(MHData)[colnames(MHData)=='Is.your.anonymity.protected.if.you.choose.to.take.advantage.of.mental.health.or.substance.abuse.treatment.resources.provided.by.your.employer.'] <- 'AnonymityProtected'

colnames(MHData)[colnames(MHData)=='If.a.mental.health.issue.prompted.you.to.request.a.medical.leave.from.work..asking.for.that.leave.would.be.'] <- 'MHPromptedMedLeave'

colnames(MHData)[colnames(MHData)=='Do.you.think.that.discussing.a.mental.health.disorder.with.your.employer.would.have.negative.consequences.'] <- 'DiscussMHCompanyNegative'

colnames(MHData)[colnames(MHData)=='Do.you.think.that.discussing.a.physical.health.issue.with.your.employer.would.have.negative.consequences.'] <- 'DiscussPHCompanyNegative'

colnames(MHData)[colnames(MHData)=='Would.you.feel.comfortable.discussing.a.mental.health.disorder.with.your.coworkers.'] <- 'DiscussMHWithCoWorkers'

colnames(MHData)[colnames(MHData)=='Would.you.feel.comfortable.discussing.a.mental.health.disorder.with.your.direct.supervisor.s..'] <- 'DiscussMHWithBoss'

colnames(MHData)[colnames(MHData)=='Do.you.feel.that.your.employer.takes.mental.health.as.seriously.as.physical.health.'] <- 'CompanySeriousMHasPH'

colnames(MHData)[colnames(MHData)=='Have.you.heard.of.or.observed.negative.consequences.for.co.workers.who.have.been.open.about.mental.health.issues.in.your.workplace.'] <- 'ObsNegOpenWithMH'

colnames(MHData)[colnames(MHData)=='Do.you.have.medical.coverage..private.insurance.or.state.provided..which.includes.treatment.of.Ã.mental.health.issues.'] <- 'MedicalCoverage'

colnames(MHData)[colnames(MHData)=='Do.you.know.local.or.online.resources.to.seek.help.for.a.mental.health.disorder.'] <- 'KnowMHResources'

colnames(MHData)[colnames(MHData)=='If.you.have.been.diagnosed.or.treated.for.a.mental.health.disorder..do.you.ever.reveal.this.to.clients.or.business.contacts.'] <- 'MHDisclosedToClients'

colnames(MHData)[colnames(MHData)=='If.you.have.revealed.a.mental.health.issue.to.a.client.or.business.contact..do.you.believe.this.has.impacted.you.negatively.'] <- 'MHDisclosedToClientsWithNegActions'

colnames(MHData)[colnames(MHData)=='If.you.have.been.diagnosed.or.treated.for.a.mental.health.disorder..do.you.ever.reveal.this.to.coworkers.or.employees.'] <- 'MHDisclosedToCoWorkers'

colnames(MHData)[colnames(MHData)=='If.you.have.revealed.a.mental.health.issue.to.a.coworker.or.employee..do.you.believe.this.has.impacted.you.negatively.'] <- 'MHDisclosedToCoWorkersWithNegActions'

colnames(MHData)[colnames(MHData)=='Do.you.believe.your.productivity.is.ever.affected.by.a.mental.health.issue.'] <- 'MHAffectProductivity'

colnames(MHData)[colnames(MHData)=='If.yes..what.percentage.of.your.work.time..time.performing.primary.or.secondary.job.functions..is.affected.by.a.mental.health.issue.'] <- 'PercWorkAffectedByMH'

colnames(MHData)[colnames(MHData)=='Do.you.have.previous.employers.'] <- 'PrevCo'

colnames(MHData)[colnames(MHData)=='Have.your.previous.employers.provided.mental.health.benefits.'] <- 'PrevCoProvideMH'

colnames(MHData)[colnames(MHData)=='Were.you.aware.of.the.options.for.mental.health.care.provided.by.your.previous.employers.'] <- 'PrevCoAwareMHCoverage'

colnames(MHData)[colnames(MHData)=='Did.your.previous.employers.ever.formally.discuss.mental.health..as.part.of.a.wellness.campaign.or.other.official.communication..'] <- 'PrevCoDiscussMH'

colnames(MHData)[colnames(MHData)=='Did.your.previous.employers.provide.resources.to.learn.more.about.mental.health.issues.and.how.to.seek.help.'] <- 'PrevCoOfferResources'

colnames(MHData)[colnames(MHData)=='Was.your.anonymity.protected.if.you.chose.to.take.advantage.of.mental.health.or.substance.abuse.treatment.resources.with.previous.employers.'] <- 'PrevCoAnonymityProtected'

colnames(MHData)[colnames(MHData)=='Do.you.think.that.discussing.a.mental.health.disorder.with.previous.employers.would.have.negative.consequences.'] <- 'PrevCoDiscussMHCompanyNegative'

colnames(MHData)[colnames(MHData)=='Do.you.think.that.discussing.a.physical.health.issue.with.previous.employers.would.have.negative.consequences.'] <- 'PrevCoDiscussPHCompanyNegative'

colnames(MHData)[colnames(MHData)=='Would.you.have.been.willing.to.discuss.a.mental.health.issue.with.your.previous.co.workers.'] <- 'PrevCoDiscussMHWithCoWorkers'

colnames(MHData)[colnames(MHData)=='Would.you.have.been.willing.to.discuss.a.mental.health.issue.with.your.direct.supervisor.s..'] <- 'PrevCoDiscussMHWithBoss'

colnames(MHData)[colnames(MHData)=='Did.you.feel.that.your.previous.employers.took.mental.health.as.seriously.as.physical.health.'] <- 'PrevCoCompanySeriousMHasPH'

colnames(MHData)[colnames(MHData)=='Did.you.hear.of.or.observe.negative.consequences.for.co.workers.with.mental.health.issues.in.your.previous.workplaces.'] <- 'PrevCoObsNegOpenWithMH'

colnames(MHData)[colnames(MHData)=='Would.you.be.willing.to.bring.up.a.physical.health.issue.with.a.potential.employer.in.an.interview.'] <- 'PotDiscussPH'

colnames(MHData)[colnames(MHData)=='Why.or.why.not.'] <- 'PotDiscussPH\_Why'

colnames(MHData)[colnames(MHData)=='Would.you.bring.up.a.mental.health.issue.with.a.potential.employer.in.an.interview.'] <- 'PotDiscussMH'

colnames(MHData)[colnames(MHData)=='Why.or.why.not..1'] <- 'PotDiscussMH\_Why'

colnames(MHData)[colnames(MHData)=='Do.you.feel.that.being.identified.as.a.person.with.a.mental.health.issue.would.hurt.your.career.'] <- 'MHHurtCareer'

colnames(MHData)[colnames(MHData)=='Do.you.think.that.team.members.co.workers.would.view.you.more.negatively.if.they.knew.you.suffered.from.a.mental.health.issue.'] <- 'CoWorkersViewYouNegKnewMH'

colnames(MHData)[colnames(MHData)=='How.willing.would.you.be.to.share.with.friends.and.family.that.you.have.a.mental.illness.'] <- 'DiscloseMHFamilyFriends'

colnames(MHData)[colnames(MHData)=='Have.you.observed.or.experienced.an.unsupportive.or.badly.handled.response.to.a.mental.health.issue.in.your.current.or.previous.workplace.'] <- 'NegResponseWithMH'

colnames(MHData)[colnames(MHData)=='Have.your.observations.of.how.another.individual.who.discussed.a.mental.health.disorder.made.you.less.likely.to.reveal.a.mental.health.issue.yourself.in.your.current.workplace.'] <- 'LessLikelyDiscloseMHPreviousEncounter'

colnames(MHData)[colnames(MHData)=='Do.you.have.a.family.history.of.mental.illness.'] <- 'FamilyHistoryMH'

colnames(MHData)[colnames(MHData)=='Have.you.had.a.mental.health.disorder.in.the.past.'] <- 'MHDisorderPast'

colnames(MHData)[colnames(MHData)=='Do.you.currently.have.a.mental.health.disorder.'] <- 'MHCurrently'

colnames(MHData)[colnames(MHData)=='If.yes..what.condition.s..have.you.been.diagnosed.with.'] <- 'MHCurrentlyDiagnosed'

colnames(MHData)[colnames(MHData)=='If.maybe..what.condition.s..do.you.believe.you.have.'] <- 'MHCurrentlyDiagnosedConditions'

colnames(MHData)[colnames(MHData)=='Have.you.been.diagnosed.with.a.mental.health.condition.by.a.medical.professional.'] <- 'MHDiagnosedByDoc'

colnames(MHData)[colnames(MHData)=='If.so..what.condition.s..were.you.diagnosed.with.'] <- 'MHDiagnosedByDoc\_Diagnosis'

colnames(MHData)[colnames(MHData)=='Have.you.ever.sought.treatment.for.a.mental.health.issue.from.a.mental.health.professional.'] <- 'MHSoughtTreatment'

colnames(MHData)[colnames(MHData)=='If.you.have.a.mental.health.issue..do.you.feel.that.it.interferes.with.your.work.when.being.treated.effectively.'] <- 'MHInteferesWhenTreated'

colnames(MHData)[colnames(MHData)=='If.you.have.a.mental.health.issue..do.you.feel.that.it.interferes.with.your.work.when.NOT.being.treated.effectively.'] <- 'MHInteferesWhenNotTreated'

colnames(MHData)[colnames(MHData)=='What.is.your.age.'] <- 'Age'

colnames(MHData)[colnames(MHData)=='What.is.your.gender.'] <- 'Gender'

colnames(MHData)[colnames(MHData)=='What.country.do.you.live.in.'] <- 'Country'

colnames(MHData)[colnames(MHData)=='What.US.state.or.territory.do.you.live.in.'] <- 'State'

colnames(MHData)[colnames(MHData)=='What.country.do.you.work.in.'] <- 'CountryWorkIn'

colnames(MHData)[colnames(MHData)=='What.US.state.or.territory.do.you.work.in.'] <- 'StateWorkIn'

colnames(MHData)[colnames(MHData)=='Which.of.the.following.best.describes.your.work.position.'] <- 'CurrentPosition'

colnames(MHData)[colnames(MHData)=='Do.you.work.remotely.'] <- 'WorkRemotely'

#List Column names

colnames(MHData)

#Graphical representation of missing vaules using 'DataExporer' library

plot\_missing(MHData, title = "Percent of Missing Values")

#Function to count all NA's in columns

propmiss <- function(dataframe) {

m <- sapply(dataframe, function(x) {

data.frame(

na\_count=sum(is.na(x)),

Obs=length(x),

perc\_missing=sum(is.na(x))/length(x)\*100

)

})

d <- data.frame(t(m))

d <- sapply(d, unlist)

d <- as.data.frame(d)

d$variable <- row.names(d)

row.names(d) <- NULL

d <- cbind(d[ncol(d)],d[-ncol(d)])

return(d[order(d$na\_count, decreasing=TRUE), ])

}

#show results of NA's counted

propmiss(MHData)

#remove single quotes

MHData$ProvideMHCoverage <- gsub("\'","", MHData$ProvideMHCoverage)

MHData$AnonymityProtected <- gsub("\'","", MHData$AnonymityProtected)

MHData$MHHurtCareer <- gsub("\'","", MHData$MHHurtCareer)

MHData$CoWorkersViewYouNegKnewMH <- gsub("\'","", MHData$CoWorkersViewYouNegKnewMH)

MHSubset <- sqldf("SELECT

SelfEmployed

,ProvideMHCoverage

,AnonymityProtected

,DiscussMHCompanyNegative

,DiscussPHCompanyNegative

,DiscussMHWithBoss

,ObsNegOpenWithMH

,MedicalCoverage

,PercWorkAffectedByMH

,PrevCoAwareMHCoverage

,MHHurtCareer

,CoWorkersViewYouNegKnewMH

,NegResponseWithMH

,MHCurrently

,MHCurrentlyDiagnosed

,Age

,Gender

,State

,CurrentPosition

,WorkRemotely

,case

when WorkRemotely = 'Never' then 0

when WorkRemotely = 'Always' then 1

when WorkRemotely = 'Sometimes' then 2

else NULL

end as NUM\_WorkRemotely

,case

when ProvideMHCoverage = 'Not eligible for coverage / N/A' then 0

when ProvideMHCoverage = 'No' then 0

when ProvideMHCoverage = 'Yes' then 1

when ProvideMHCoverage = 'I dont know' then 2

else NULL

end as NUM\_ProvideMHCoverage

,case

when AnonymityProtected = 'No' then 0

when AnonymityProtected = 'Yes' then 1

when AnonymityProtected = 'I dont know' then 2

else NULL

end as NUM\_AnonymityProtected

,case

when DiscussMHCompanyNegative = 'No' then 0

when DiscussMHCompanyNegative = 'Yes' then 1

when DiscussMHCompanyNegative = 'Maybe' then 2

else NULL

end as NUM\_DiscussMHCompanyNegative

,case

when DiscussPHCompanyNegative = 'No' then 0

when DiscussPHCompanyNegative = 'Yes' then 1

when DiscussPHCompanyNegative = 'Maybe' then 2

else NULL

end as NUM\_DiscussPHCompanyNegative

,case

when DiscussMHWithBoss = 'No' then 0

when DiscussMHWithBoss = 'Yes' then 1

when DiscussMHWithBoss = 'Maybe' then 2

else NULL

end as NUM\_DiscussMHWithBoss

,case

when ObsNegOpenWithMH = 'No' then 0

when ObsNegOpenWithMH = 'Yes' then 1

else NULL

end as NUM\_ObsNegOpenWithMH

,case

when MedicalCoverage = 0 then 0

when MedicalCoverage = 1 then 1

when MedicalCoverage = 'NA' then NULL

when MedicalCoverage is NULL then NULL

end as NUM\_MedicalCoverage

,case

when PercWorkAffectedByMH = '1-25%' then 'Low'

when PercWorkAffectedByMH = '26-50%' then 'Low\_to\_Medium'

when PercWorkAffectedByMH = '51-75%' then 'Medium'

when PercWorkAffectedByMH = '76-100%' then 'High'

else NULL

end as CAT\_PercWorkAffectedByMH

,case

when MHHurtCareer = 'Maybe' then 2

when MHHurtCareer = 'No, I dont think it would' then 0

when MHHurtCareer = 'No, it has not' then 0

when MHHurtCareer = 'Yes, ' then 1

when MHHurtCareer = 'Yes, I think it would' then 1

when MHHurtCareer = 'Yes, it has' then 1

else NULL

end as CAT\_MHHurtCareer

,case

when CoWorkersViewYouNegKnewMH = 'Maybe' then 2

when CoWorkersViewYouNegKnewMH = 'No, I dont think they would' then 0

when CoWorkersViewYouNegKnewMH = 'No, they do not' then 0

when CoWorkersViewYouNegKnewMH = 'Yes, ' then 1

when CoWorkersViewYouNegKnewMH = 'Yes, I think they would' then 1

when CoWorkersViewYouNegKnewMH = 'Yes, they do' then 1

else NULL

end as NUM\_CoWorkersViewYouNegKnewMH

,case

when NegResponseWithMH = 'No' then 0

when NegResponseWithMH = 'Maybe/Not sure' then 2

when NegResponseWithMH = 'Yes, I experienced' then 1

when NegResponseWithMH = 'Yes, I observed' then 1

when NegResponseWithMH = 'N/A' then NULL

when NegResponseWithMH is NULL then NULL

else NULL end as NUM\_NegResponseWithMH

,case

when MHCurrently = 'Maybe' then 2

when MHCurrently = 'No' then 0

when MHCurrently = 'Yes' then 1

else NULL

end as NUM\_MHCurrently

,case

when MHCurrentlyDiagnosed like 'Addictive Disorder' then 'Addiction'

when MHCurrentlyDiagnosed like 'Anxiety' then 'Anxiety'

when MHCurrentlyDiagnosed like 'Attention Def' then 'Attention'

when MHCurrentlyDiagnosed like 'Autism' then 'Autism'

when MHCurrentlyDiagnosed like 'Burn' then 'Burnout'

when MHCurrentlyDiagnosed like 'Combination of physical' then 'Attention'

when MHCurrentlyDiagnosed like 'Depression' then 'Depression'

when MHCurrentlyDiagnosed like 'Eating Disorder' then 'Eating'

when MHCurrentlyDiagnosed like 'I haven' then 'Unknown'

when MHCurrentlyDiagnosed like 'Mood Disorder' then 'Mood'

when MHCurrentlyDiagnosed like 'Obsessive-Compulsive Disorder' then 'Obsessive-Compulsive'

when MHCurrentlyDiagnosed like 'PDD-NOS' then 'PDD-NOS'

when MHCurrentlyDiagnosed like 'Personality Disorder' then 'Personality'

when MHCurrentlyDiagnosed like 'Post-traumatic Stress Disorder' then 'PTSD'

when MHCurrentlyDiagnosed like 'Schizotypal Personality Disorder' then 'Schizotypal'

when MHCurrentlyDiagnosed like 'Seasonal Affective Disorder' then 'Seasonal\_Affective'

when MHCurrentlyDiagnosed like 'Sexual addiction' then 'Sexual\_Addiction'

when MHCurrentlyDiagnosed like 'Stress Response Syndromes' then 'Stress\_Response'

when MHCurrentlyDiagnosed like 'Transgender' then 'Transgender/Mood/Anxiety'

when MHCurrentlyDiagnosed like 'Traumatic Brain Injury' then 'Traumatic\_Brain\_Injury'

else NULL

end as CAT\_MHCurrentlyDiagnosed

,case

when CurrentPosition = 'Designer' then 'Designer'

when CurrentPosition = 'Support|Designer' then 'Designer'

when CurrentPosition = 'Support|Designer|Front-end Developer' then 'Designer'

when CurrentPosition = 'Back-end Developer' then 'Developer'

when CurrentPosition = 'Other|Back-end Developer' then 'Developer'

when CurrentPosition = 'Support|Back-end Developer' then 'Developer'

when CurrentPosition = 'Support|Front-end Developer|Back-end Developer' then 'Developer'

when CurrentPosition = 'Front-end Developer' then 'Developer'

when CurrentPosition = 'Other|Designer|Front-end Developer' then 'Developer'

when CurrentPosition = 'Other|Front-end Developer' then 'Developer'

when CurrentPosition = 'Dev Evangelist/Advocate' then 'DevEvangelist'

when CurrentPosition = 'Other|Dev Evangelist/Advocate' then 'DevEvangelist'

when CurrentPosition = 'DevOps/SysAdmin' then 'DevOps'

when CurrentPosition = 'Other|DevOps/SysAdmin|Back-end Developer' then 'DevOps'

when CurrentPosition = 'Support|DevOps/SysAdmin' then 'DevOps'

when CurrentPosition = 'Executive Leadership' then 'Exec\_Leadership'

when CurrentPosition = 'Other|Executive Leadership' then 'Exec\_Leadership'

when CurrentPosition = 'HR' then 'HR'

when CurrentPosition = 'Other|HR' then 'HR'

when CurrentPosition = 'One-person shop' then 'OnePerson'

when CurrentPosition = 'Other|Front-end Developer|Designer|One-person shop' then 'OnePerson'

when CurrentPosition = 'Other|One-person shop' then 'OnePerson'

when CurrentPosition = 'Sales|Support|DevOps/SysAdmin|Executive Leadership' then 'OnePerson'

when CurrentPosition = 'Support|Sales|Back-end Developer|Front-end Developer|Designer|One-person shop' then 'OnePerson'

when CurrentPosition = 'Other' then 'Other'

when CurrentPosition = 'Support|Other' then 'Other'

when CurrentPosition = 'Sales' then 'Sales'

when CurrentPosition = 'Other|Supervisor/Team Lead' then 'Supervisor'

when CurrentPosition = 'Supervisor/Team Lead' then 'Supervisor'

when CurrentPosition = 'Support|HR|Supervisor/Team Lead|Executive Leadership' then 'Supervisor'

when CurrentPosition = 'Other|Support' then 'Support'

when CurrentPosition = 'Support' then 'Support'

end as CAT\_CurrentPosition

,case

WHEN MHCurrentlyDiagnosed LIKE '%Substance Use%' then 1

else 0

end AS HasSubstanceAbuse

FROM MHData")

#Subset of data for analysis

MHSubsetAnalysis <- sqldf("SELECT

SelfEmployed

,NUM\_ProvideMHCoverage

,NUM\_AnonymityProtected

,NUM\_DiscussMHCompanyNegative

,NUM\_DiscussPHCompanyNegative

,NUM\_DiscussMHWithBoss

,NUM\_ObsNegOpenWithMH

,NUM\_MedicalCoverage

,CAT\_PercWorkAffectedByMH

,CAT\_MHHurtCareer

,NUM\_CoWorkersViewYouNegKnewMH

,NUM\_NegResponseWithMH

,NUM\_MHCurrently

,CAT\_MHCurrentlyDiagnosed

,Age

,State

,CAT\_CurrentPosition

,WorkRemotely

,NUM\_WorkRemotely

,MHCurrentlyDiagnosed

,HasSubstanceAbuse

FROM MHSubset")

#Create a function that will massage our data into a consumable format.

explodeDataFrameColumn <- function(dataframe, columnName){

columnIndex <- which(names(dataframe) == columnName)

#remove rows with blank values

dataframe <- dataframe[!(is.na(dataframe[,columnIndex]) | dataframe[,columnIndex] == ""),]

#Split pipe delimitted values into rows.

returnValue <- separate\_rows(dataframe, columnName, sep = "\\|")

}

#Create a subset data frame to hold diagnoses and job role data.

MHPosition <- data.frame(CurrentDiagnoses = MHSubset$MHCurrentlyDiagnosed, CurrentPosition = MHSubset$CurrentPosition)

#break up the pipe delimitted values into rows.

MHPositionSplit <- explodeDataFrameColumn(MHPosition, "CurrentDiagnoses")

MHPositionSplit <- explodeDataFrameColumn(MHPositionSplit, "CurrentPosition")

#create lookup tables so we can run correlation analysis

diagnoses <- data.frame(Diagnosis = unique(MHPositionSplit$CurrentDiagnoses), DiagnosisCode = 1:length(unique(MHPositionSplit$CurrentDiagnoses)))

positions <- data.frame(Position = unique(MHPositionSplit$CurrentPosition), PositionCode = 1:length(unique(MHPositionSplit$CurrentPosition)))

#rename columns so we can use merge function.

colnames(MHPositionSplit) <- c("Diagnosis", "Position")

MHPositionSplitCoded <- merge(MHPositionSplit, diagnoses, by = "Diagnosis")

MHPositionSplitCoded <- merge(MHPositionSplitCoded, positions, by = "Position")

plot(MHPositionSplitCoded$DiagnosisCode, MHPositionSplitCoded$PositionCode)

#correlation value

cor(MHPositionSplitCoded$DiagnosisCode, MHPositionSplitCoded$PositionCode)

##############################

###

### Data Frame of NO coverage

###

##############################

#ProvideMHCoverage 0 = No

#ProvideMHCoverage 1 = yes

MHSubset\_StateAnalysisNOs <- sqldf("Select

State

,NUM\_ProvideMHCoverage

FROM MHSubsetAnalysis

WHERE State <> ''

AND NUM\_ProvideMHCoverage = 0")

#count of medical centers by state

MHSubset\_StateAnalysisNOs <- sqldf("select State

,count(\*) as CountOfNoCoverage

from MHSubsetAnalysis

WHERE State <> ''

AND NUM\_ProvideMHCoverage = 0

group by State

Order by 2 desc")

#Add sum column

MHSubset\_StateAnalysisNOs$sum <- ave(MHSubset\_StateAnalysisNOs$CountOfNoCoverage, FUN=sum)

colnames(MHSubset\_StateAnalysisNOs)[colnames(MHSubset\_StateAnalysisNOs)=='sum'] <- 'TotalNoCoverage'

##############################

###

### Data Frame of Yes coverage

###

##############################

MHSubset\_StateAnalysisYES <- sqldf("Select

State

,NUM\_ProvideMHCoverage

FROM MHSubsetAnalysis

WHERE State <> ''

AND NUM\_ProvideMHCoverage = 1")

#count of medical centers by state

MHSubset\_StateAnalysisYES <- sqldf("select State

,count(\*) as CountOfYesCoverage

from MHSubsetAnalysis

WHERE State <> ''

AND NUM\_ProvideMHCoverage = 1

group by State

Order by 2 desc")

#Add sum column

MHSubset\_StateAnalysisYES$sum <- ave(MHSubset\_StateAnalysisYES$CountOfYesCoverage, FUN=sum)

colnames(MHSubset\_StateAnalysisYES)[colnames(MHSubset\_StateAnalysisYES)=='sum'] <- 'TotalYesCoverage'

#################################

#####

##### Make State DB

#####

#################################

#Create State DB data frame

StateDB <- data.frame(state.name, state.area, state.region)

colnames(StateDB)[colnames(StateDB)=='state.name'] <- 'State'

colnames(StateDB)[colnames(StateDB)=='state.area'] <- 'StateSize\_mi2' #Data is in Square miles

colnames(StateDB)[colnames(StateDB)=='state.region'] <- 'StateRegion'

#Add district of Columbia to StateDB Data Frame

DistrictColumbia <- data.frame("District of Columbia", "68.34", "South")

names(DistrictColumbia) <- c("State", "StateSize\_mi2", "StateRegion")

StateDB <- rbind(StateDB, DistrictColumbia)

##############################

###

### Merge yes and no data frames

### Merge StateDB

###

##############################

#Merge Yes and No data frames

StateAnalysis <- join(MHSubset\_StateAnalysisYES, MHSubset\_StateAnalysisNOs, by="State", type="full")

#Merge StateDB

StateAnalysis <- join(StateAnalysis, StateDB, by="State", type="inner")

#######################################################

####

#### Calculate new variable Yes/No coverage density

####

#######################################################

#Create density calculations

StateAnalysis <- sqldf("SELECT \*

,CountOfYesCoverage/(StateSize\_mi2/1000.0) as YesCoverage\_Density

,CountOfNoCoverage/(StateSize\_mi2/1000.0) as NoCoverage\_Density

FROM StateAnalysis")

#Remove district of columbia -- outlier

StateAnalysis <- StateAnalysis[!(StateAnalysis$State=="District of Columbia"),]

##############################

###

### Location Demographics

### MH Issues workers who work remotely

###

##############################

#Are there higher reports of mental health issues for remote workers?

#No = 0

#Yes = 1

MHRemotely <- sqldf("SELECT

NUM\_MHCurrently

,WorkRemotely

FROM MHSubsetAnalysis

WHERE NUM\_MHCurrently <> 2")

#Number obs in each category

table(WorkRemotely = MHRemotely$WorkRemotely, HasMentalHealthIssue = MHRemotely$NUM\_MHCurrently)

#Proportion of recieveing MH issue

prop.table(table(WorkRemotely = MHRemotely$WorkRemotely, HasMentalHealthIssue = MHRemotely$NUM\_MHCurrently), margin = 1)\*100

#Chi square test of association

chisq.test(table(WorkRemotely = MHRemotely$WorkRemotely, HasMentalHealthIssue = MHRemotely$NUM\_MHCurrently))

fitRW <- glm(NUM\_MHCurrently ~ WorkRemotely, data=MHRemotely, family = binomial )

summary(fitRW)

#Get a subset of workers with mental health issues for analysis

MHSubstanceAbuseSupervisor <- sqldf("SELECT

NUM\_DiscussMHWithBoss

,HasSubstanceAbuse

FROM MHSubsetAnalysis

WHERE NUM\_DiscussMHWithBoss <> 2

AND NUM\_MHCurrently=1")

#Number obs in each category

table(HasSubstanceAbuse=MHSubstanceAbuseSupervisor$HasSubstanceAbuse, DiscussWithBoss=MHSubstanceAbuseSupervisor$NUM\_DiscussMHWithBoss)

#Proportion of those with substance abuset that would discuss the issue with their supervisor

prop.table(table(HasSubstanceAbuse=MHSubstanceAbuseSupervisor$HasSubstanceAbuse, DiscussWithBoss=MHSubstanceAbuseSupervisor$NUM\_DiscussMHWithBoss), margin = 1)\*100

#Chi square test of association

chisq.test(table(HasSubstanceAbuse=MHSubstanceAbuseSupervisor$HasSubstanceAbuse, DiscussWithBoss=MHSubstanceAbuseSupervisor$NUM\_DiscussMHWithBoss))

fitSA <- glm(NUM\_DiscussMHWithBoss ~ HasSubstanceAbuse, data=MHSubstanceAbuseSupervisor, family = binomial )

summary(fitSA)

#Get a subset of workers with mental health issues for analysis

MHSubstanceAbuseAnonymity <- sqldf("SELECT

NUM\_AnonymityProtected

,HasSubstanceAbuse

FROM MHSubsetAnalysis

WHERE NUM\_AnonymityProtected <> 2

AND NUM\_MHCurrently=1")

#Number obs in each category

table(HasSubstanceAbuse=MHSubstanceAbuseAnonymity$HasSubstanceAbuse, AnonymityProtected=MHSubstanceAbuseAnonymity$NUM\_AnonymityProtected)

#Proportion of those with substance abuset that would discuss the issue with their supervisor

prop.table(table(HasSubstanceAbuse=MHSubstanceAbuseAnonymity$HasSubstanceAbuse, AnonymityProtected=MHSubstanceAbuseAnonymity$NUM\_AnonymityProtected), margin = 1)\*100

#Chi square test of association

chisq.test(table(HasSubstanceAbuse=MHSubstanceAbuseAnonymity$HasSubstanceAbuse, AnonymityProtected=MHSubstanceAbuseAnonymity$NUM\_AnonymityProtected))

fitSAA <- glm(NUM\_AnonymityProtected ~ HasSubstanceAbuse, data=MHSubstanceAbuseAnonymity, family = binomial )

summary(fitSAA)

#Get a subset of workers with mental health issues for analysis

MHSubstanceAbuseNegResponse <- sqldf("SELECT

NUM\_NegResponseWithMH

,HasSubstanceAbuse

FROM MHSubsetAnalysis

WHERE NUM\_NegResponseWithMH <> 2

AND NUM\_MHCurrently=1")

#Number obs in each category

table(HasSubstanceAbuse=MHSubstanceAbuseNegResponse$HasSubstanceAbuse, NegResponse=MHSubstanceAbuseNegResponse$NUM\_NegResponseWithMH)

#Proportion of those with substance abuset that would discuss the issue with their supervisor

prop.table(table(HasSubstanceAbuse=MHSubstanceAbuseNegResponse$HasSubstanceAbuse, NegResponse=MHSubstanceAbuseNegResponse$NUM\_NegResponseWithMH), margin = 1)\*100

#Chi square test of association

chisq.test(table(HasSubstanceAbuse=MHSubstanceAbuseNegResponse$HasSubstanceAbuse, NegResponse=MHSubstanceAbuseNegResponse$NUM\_NegResponseWithMH))

fitSAR <- glm(NUM\_NegResponseWithMH ~ HasSubstanceAbuse, data=MHSubstanceAbuseNegResponse, family = binomial )

summary(fitSAR)

#Get a subset of workers with mental health issues for analysis

MHSubstanceAbuseNegConsequence <- sqldf("SELECT

CAT\_MHHurtCareer

,NUM\_ObsNegOpenWithMH

,NUM\_CoWorkersViewYouNegKnewMH

,CASE

WHEN CAT\_MHHurtCareer = 1 OR NUM\_ObsNegOpenWithMH = 1 OR NUM\_CoWorkersViewYouNegKnewMH = 1 then 1

else 0

END as HasNegativeConsequence

,HasSubstanceAbuse

FROM MHSubsetAnalysis

WHERE NUM\_ObsNegOpenWithMH <> 2

AND CAT\_MHHurtCareer <> 2

AND NUM\_CoWorkersViewYouNegKnewMH <> 2

AND NUM\_MHCurrently=1")

#Number obs in each category

table(HasSubstanceAbuse=MHSubstanceAbuseNegConsequence$HasSubstanceAbuse, NegConsequence=MHSubstanceAbuseNegConsequence$HasNegativeConsequence)

#Proportion of those with substance abuset that would discuss the issue with their supervisor

prop.table(table(HasSubstanceAbuse=MHSubstanceAbuseNegConsequence$HasSubstanceAbuse, NegConsequence=MHSubstanceAbuseNegConsequence$HasNegativeConsequence), margin = 1)\*100

#Chi square test of association

chisq.test(table(HasSubstanceAbuse=MHSubstanceAbuseNegConsequence$HasSubstanceAbuse, NegConsequence=MHSubstanceAbuseNegConsequence$HasNegativeConsequence))

fitSAC <- glm(HasNegativeConsequence ~ HasSubstanceAbuse, data=MHSubstanceAbuseNegConsequence, family = binomial )

summary(fitSAC)