Mental Health

Capstone Project

Cohort B: Alvaro Chinchayan, Leighton Li, Andrey Lifar, Yoki Liu, Yue Ping, Sherry Zuo

2/27/2020

Our main goal is to answer questions for:
What's the overview of people's mental health problem?
Can you predict whether a patient should be treated of his/her mental illness or not according to the value obtained based on our data?
Is there a correlation between mental health and whether job lets them take them break/ alow them to work from home/remote worker?
What can we get for the tweets data?

Our first part will focus on the OSMI Mental Health in Tech Surveys data
We merged 2016-2019 data into a new dataset by python, more code could see in the Mental_Health.ipynb file. We will use the merged dataset "mental_health" to do an overview of the mental illness
We will do supervised machine learning focus on the year 2014 data

Data Loading

```
#import data
data<-read_csv("~/Desktop/mental_health.csv")
```

```
## Warning: Missing column names filled in: 'X1' [1]
```

```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   X1 = col_double(),
##   self_empl_flag = col_double(),
##   tech_comp_flag = col_double(),
##   tech_role_flag = col_double(),
##   prev_employers_flag = col_double(),
##   age = col_double(),
##   sex = col_double()
## )
```

```
## See spec(...) for full column specifications.
```

```
## Warning: 608 parsing failures.
##  row            col expected actual                          file
## 2607 tech_comp_flag a double   True '~/Desktop/mental_health.csv'
## 2607 tech_role_flag a double   True '~/Desktop/mental_health.csv'
## 2608 tech_comp_flag a double   True '~/Desktop/mental_health.csv'
## 2608 tech_role_flag a double   True '~/Desktop/mental_health.csv'
## 2609 tech_comp_flag a double   True '~/Desktop/mental_health.csv'
## .... ............. ........ ...... ..............................
## See problems(...) for more details.
```

```
data2014<-read_csv("~/Desktop/2014.csv")
```

```
## Parsed with column specification:
## cols(
##   .default = col_character(),
##   Age = col_double()
## )
## See spec(...) for full column specifications.
```

Data Cleaning

```
#remove first column since it's just the numbers of row
data<-data[,2:29]
```

```
skim(data)
```

```
## Skim summary statistics
##  n obs: 2958
##  n variables: 28
##
## -- Variable type:character ---------------------------
##                          variable missing complete    n min  max empty
##                      comp_no_empl     504     2454 2958   3    8     0
##                      country_live       2     2956 2958   2   22     0
##                      country_work       2     2956 2958   2   22     0
##           future_ph_specification       0     2958 2958   2    5     0
##                  live_us_teritory    1106     1852 2958   4   20     0
##                 mh_anonimity_flag     504     2454 2958   2   12     0
##         mh_bad_response_workplace      91     2867 2958   2   30     0
##  mh_diagnosed&reveal_clients_flag    2454      504 2958  11   41     0
##   mh_diagnosed&reveal_cowork_flag    2454      504 2958  11   41     0
##               mh_disorder_current       0     2958 2958   2   10     0
##                  mh_disorder_past      15     2943 2958   2   10     0
##            mh_employer_discussion     504     2454 2958   2   12     0
##                    mh_family_hist       0     2958 2958   2   12     0
##                    mh_prod_impact    2454      504 2958   2   20     0
##                mh_prod_impact_perc    2591      367 2958   5    7     0
##       mh_sharing_friends/fam_flag       0     2958 2958   1   53     0
##                  prev_mh_anonimity     368     2590 2958   2   12     0
##                   prev_mh_benefits     368     2590 2958   8   17     0
##        prev_mh_benefits_awareness     368     2590 2958  18   31     0
##                 prev_mh_discussion     368     2590 2958   8   17     0
##                        why/why_not     696     2262 2958   1 2475     0
##                   work_us_teritory    1086     1872 2958   4   20     0
##  n_unique
##         6
##        73
##        76
##         3
##        48
##         3
##         5
```

```
##         5
##         5
##         5
##         5
##         3
##         3
##         4
##         4
##        17
##         4
##         4
##         6
##         4
##      2215
##        50
##
## -- Variable type:numeric ------------------------------
##             variable missing complete    n  mean   sd p0 p25 p50 p75 p100
##                  age       2     2956 2958 34.61 8.25 18  29  33  39   74
##  prev_employers_flag       0     2958 2958  0.88 0.33  0   1   1   1    1
##       self_empl_flag       0     2958 2958  0.17 0.38  0   0   0   0    1
##                  sex      24     2934 2958  1.32 0.52  1   1   1   2    3
##       tech_comp_flag     808     2150 2958  0.77 0.42  0   1   1   1    1
##       tech_role_flag    1691     1267 2958  0.94 0.24  0   1   1   1    1
##      hist
##
##
##
##
##
##
```

```
#rename columns
names(data2014)[5:27]<-c("state","self_employed","family_history","mh_treatment", "interfere","company_
```

```
skim(data2014)
```

```
## Skim summary statistics
##  n obs: 1260
##  n variables: 27
##
## -- Variable type:character ---------------------------
##                      variable missing complete    n min  max empty
##           anonymity_protected       0     1260 1260   2   10     0
##         awareness_mh_benefits       0     1260 1260   2    8     0
##                      comments    1097      163 1260   1 3589     0
##                  company_size       0     1260 1260   3   14     0
##                       Country       0     1260 1260   5   22     0
##                family_history       0     1260 1260   2    3     0
##                        Gender       0     1260 1260   1   46     0
##                     interfere     264      996 1260   5    9     0
##          interview_mh_bringup       0     1260 1260   2    5     0
##          interview_ph_bringup       0     1260 1260   2    5     0
```

```
##            medical_leave_easy       0     1260 1260   9   18     0
##                  mh_benefits       0     1260 1260   2   10     0
##                   mh_discuss       0     1260 1260   2   10     0
##          mh_disscuss_coworker       0     1260 1260   2   12     0
##        mh_disscuss_supervisor       0     1260 1260   2   12     0
##  mh_negative_consequence_flag       0     1260 1260   2    5     0
##                 mh_resources       0     1260 1260   2   10     0
##                mh_serious_ph       0     1260 1260   2   10     0
##                 mh_treatment       0     1260 1260   2    3     0
##  ph_negative_consequence_flag       0     1260 1260   2    5     0
##                       remote       0     1260 1260   2    3     0
##                self_employed      18     1242 1260   2    3     0
##                        state     516      744 1260   2    2     0
##                 tech_company       0     1260 1260   2    3     0
##                    Timestamp       0     1260 1260  16   19     0
##                witness_mh_nc       0     1260 1260   2    3     0
##  n_unique
##         3
##         3
##       159
##         6
##        48
##         2
##        47
##         4
##         3
##         3
##         5
##         3
##         3
##         3
##         3
##         3
##         3
##         3
##         3
##         2
##         3
##         2
##         2
##        45
##         2
##      1249
##         2
##
## -- Variable type:numeric ------------------------------
##  variable missing complete    n    mean      sd    p0 p25 p50 p75  p100
##       Age       0     1260 1260 7.9e+07 2.8e+09 -1726  27  31  36 1e+11
##      hist
##
```

```
#make a copy of our data
maindata<-data2014
#drop variables "Timestamp", "Country", "comments" just to make our lives easier.
maindata<-subset(maindata, select = -c(1,4,27))
```

```r
names(maindata)[1:2]<-c("age","gender")
```

```r
data.table(maindata)
```

```
##       age gender state self_employed family_history mh_treatment interfere
##    1:  37 Female    IL          <NA>             No          Yes     Often
##    2:  44      M    IN          <NA>             No           No    Rarely
##    3:  32   Male  <NA>          <NA>             No           No    Rarely
##    4:  31   Male  <NA>          <NA>            Yes          Yes     Often
##    5:  31   Male    TX          <NA>             No           No     Never
##   ---
## 1256:  32   Male    IL            No            Yes          Yes     Often
## 1257:  34   male    CA            No            Yes          Yes Sometimes
## 1258:  46      f    NC            No             No           No      <NA>
## 1259:  25   Male    IL            No            Yes          Yes Sometimes
## 1260:  41   Male  <NA>            No             No           No Sometimes
##         company_size remote tech_company mh_benefits awareness_mh_benefits
##    1:          6-25     No          Yes         Yes               Not sure
##    2: More than 1000     No           No  Don't know                    No
##    3:          6-25     No          Yes          No                    No
##    4:         26-100     No          Yes          No                   Yes
##    5:        100-500    Yes          Yes         Yes                    No
##   ---
## 1256:         26-100    Yes          Yes         Yes                   Yes
## 1257: More than 1000     No          Yes         Yes                   Yes
## 1258:        100-500    Yes          Yes          No                   Yes
## 1259:         26-100     No           No         Yes                   Yes
## 1260:       500-1000     No           No          No                    No
##       mh_discuss mh_resources anonymity_protected medical_leave_easy
##    1:         No          Yes                Yes      Somewhat easy
##    2: Don't know   Don't know         Don't know         Don't know
##    3:         No           No         Don't know Somewhat difficult
##    4:         No           No                 No Somewhat difficult
##    5: Don't know   Don't know         Don't know         Don't know
##   ---
## 1256:         No           No                Yes Somewhat difficult
## 1257:         No           No         Don't know Somewhat difficult
## 1258:         No           No         Don't know         Don't know
## 1259:         No           No                Yes         Don't know
## 1260:         No           No         Don't know         Don't know
##       mh_negative_consequence_flag ph_negative_consequence_flag
##    1:                           No                           No
##    2:                        Maybe                           No
##    3:                           No                           No
##    4:                          Yes                          Yes
##    5:                           No                           No
##   ---
## 1256:                           No                           No
## 1257:                          Yes                          Yes
## 1258:                          Yes                           No
## 1259:                        Maybe                           No
## 1260:                          Yes                           No
##       mh_disscuss_coworker mh_disscuss_supervisor interview_mh_bringup
```

```
##    1:         Some of them                   Yes                   No
##    2:                   No                    No                   No
##    3:                  Yes                   Yes                  Yes
##    4:         Some of them                    No                Maybe
##    5:         Some of them                   Yes                  Yes
##   ---
## 1256:         Some of them                   Yes                   No
## 1257:                   No                    No                   No
## 1258:                   No                    No                   No
## 1259:         Some of them                    No                   No
## 1260:         Some of them                    No                   No
##       interview_ph_bringup mh_serious_ph witness_mh_nc
##    1:                Maybe           Yes            No
##    2:                   No    Don't know            No
##    3:                  Yes            No            No
##    4:                Maybe            No           Yes
##    5:                  Yes    Don't know            No
##   ---
## 1256:                   No           Yes            No
## 1257:                   No            No            No
## 1258:                   No            No            No
## 1259:                   No    Don't know            No
## 1260:                Maybe            No           Yes
```

```
skim(maindata)
```

```
## Skim summary statistics
##  n obs: 1260
##  n variables: 24
##
## -- Variable type:character ---------------------------
##                      variable missing complete    n min max empty n_unique
##           anonymity_protected       0     1260 1260   2  10     0        3
##         awareness_mh_benefits       0     1260 1260   2   8     0        3
##                  company_size       0     1260 1260   3  14     0        6
##                family_history       0     1260 1260   2   3     0        2
##                        gender       0     1260 1260   1  46     0       47
##                      interfere     264      996 1260   5   9     0        4
##          interview_mh_bringup       0     1260 1260   2   5     0        3
##          interview_ph_bringup       0     1260 1260   2   5     0        3
##            medical_leave_easy       0     1260 1260   9  18     0        5
##                   mh_benefits       0     1260 1260   2  10     0        3
##                    mh_discuss       0     1260 1260   2  10     0        3
##          mh_disscuss_coworker       0     1260 1260   2  12     0        3
##        mh_disscuss_supervisor       0     1260 1260   2  12     0        3
##  mh_negative_consequence_flag       0     1260 1260   2   5     0        3
##                  mh_resources       0     1260 1260   2  10     0        3
##                 mh_serious_ph       0     1260 1260   2  10     0        3
##                  mh_treatment       0     1260 1260   2   3     0        2
##  ph_negative_consequence_flag       0     1260 1260   2   5     0        3
##                        remote       0     1260 1260   2   3     0        2
##                 self_employed      18     1242 1260   2   3     0        2
##                         state     516      744 1260   2   2     0       45
##                  tech_company       0     1260 1260   2   3     0        2
```

```
##                witness_mh_nc       0     1260 1260   2   3     0        2
##
## -- Variable type:numeric ----------------------------
##  variable missing complete    n    mean      sd    p0 p25 p50 p75  p100
##       age       0     1260 1260 7.9e+07 2.8e+09 -1726  27  31  36 1e+11
##      hist
##
```

```
#make word into lowercases in column gender
maindata$gender<-str_to_lower(maindata$gender)
#build gender group
male_str = c("male", "m", "male-ish", "maile", "mal", "male (cis)", "make", "male ", "man","msle", "mail
trans_str = c("trans-female", "something kinda male?", "queer/she/they", "non-binary","nah", "all", "enb
female_str = c("cis female", "f", "female", "woman",  "femake", "female ","cis-female/femme", "female (
#replace details
maindata$gender <- ifelse(maindata$gender %in% male_str, "M",ifelse(maindata$gender %in% female_str,"F"
```

```
#we want to focus on the US
maindata<-maindata[!is.na(maindata$state),]
```

```
#remove other NAs
maindata<-na.omit(maindata)
```

```
skim(maindata)
```

```
## Skim summary statistics
##  n obs: 589
##  n variables: 24
##
## -- Variable type:character ---------------------------
##                      variable missing complete   n min max empty n_unique
##           anonymity_protected       0      589 589   2  10     0        3
##         awareness_mh_benefits       0      589 589   2   8     0        3
##                  company_size       0      589 589   3  14     0        6
##                family_history       0      589 589   2   3     0        2
##                        gender       0      589 589   1   1     0        3
##                      interfere       0      589 589   5   9     0        4
##          interview_mh_bringup       0      589 589   2   5     0        3
##          interview_ph_bringup       0      589 589   2   5     0        3
##            medical_leave_easy       0      589 589   9  18     0        5
##                   mh_benefits       0      589 589   2  10     0        3
##                    mh_discuss       0      589 589   2  10     0        3
##          mh_disscuss_coworker       0      589 589   2  12     0        3
##        mh_disscuss_supervisor       0      589 589   2  12     0        3
##  mh_negative_consequence_flag       0      589 589   2   5     0        3
##                  mh_resources       0      589 589   2  10     0        3
##                 mh_serious_ph       0      589 589   2  10     0        3
##                  mh_treatment       0      589 589   2   3     0        2
##  ph_negative_consequence_flag       0      589 589   2   5     0        3
##                        remote       0      589 589   2   3     0        2
##                 self_employed       0      589 589   2   3     0        2
##                         state       0      589 589   2   2     0       43
##                  tech_company       0      589 589   2   3     0        2
```

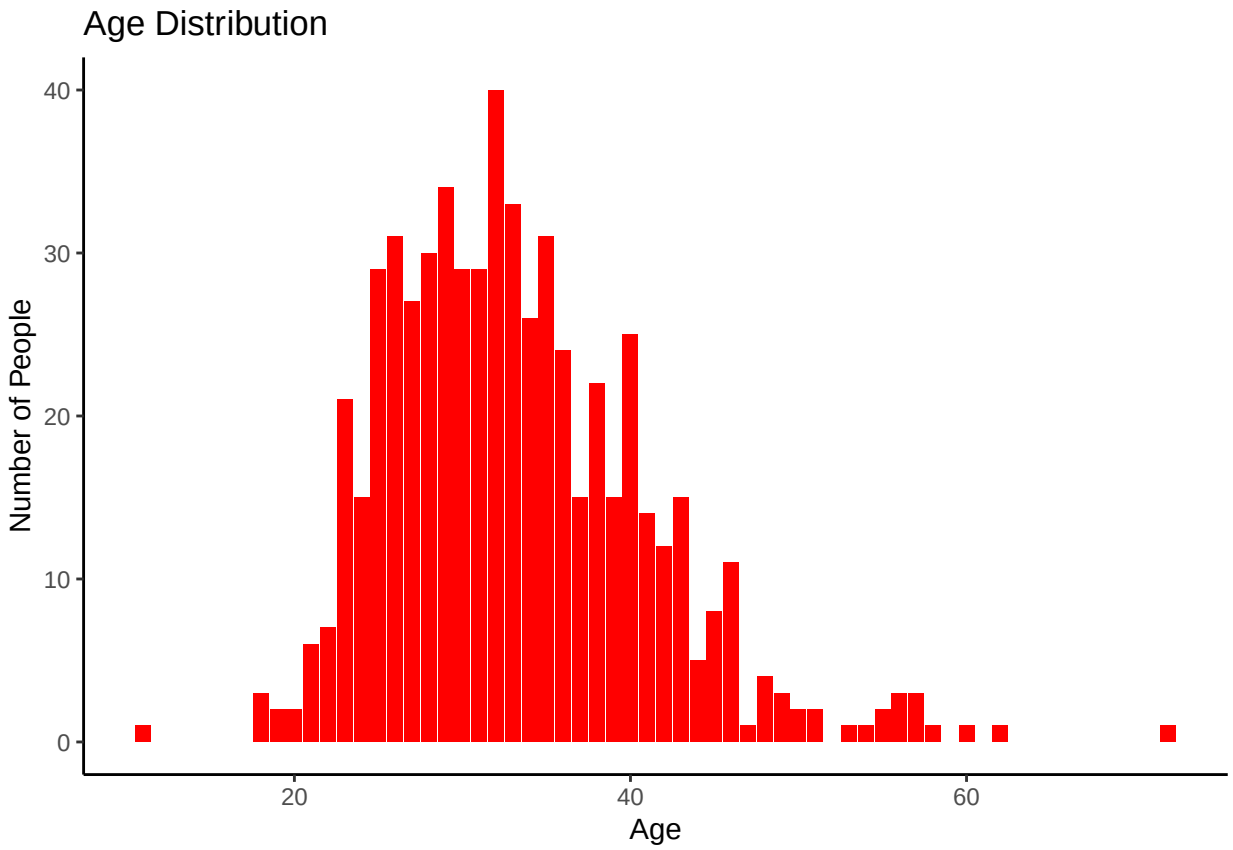
```
##                witness_mh_nc       0      589 589   2   3     0        2
##
## -- Variable type:numeric ----------------------------
##  variable missing complete   n  mean    sd p0 p25 p50 p75 p100     hist
##       age       0      589 589 33.67 14.41 11  28  32  38  329
```

```
#there is one person list 329, so delete this row
maindata<-maindata[maindata$age!=329,]
```

```
names(maindata)
```

```
##  [1] "age"                           "gender"
##  [3] "state"                         "self_employed"
##  [5] "family_history"                "mh_treatment"
##  [7] "interfere"                     "company_size"
##  [9] "remote"                        "tech_company"
## [11] "mh_benefits"                   "awareness_mh_benefits"
## [13] "mh_discuss"                    "mh_resources"
## [15] "anonymity_protected"           "medical_leave_easy"
## [17] "mh_negative_consequence_flag" "ph_negative_consequence_flag"
## [19] "mh_disscuss_coworker"          "mh_disscuss_supervisor"
## [21] "interview_mh_bringup"          "interview_ph_bringup"
## [23] "mh_serious_ph"                 "witness_mh_nc"
```

```
maindata%>%
  ggplot(aes(x=age))+
  geom_bar(fill="red")+
  labs(title="Age Distribution", x="Age", y="Number of People")+
  theme(panel.background = element_rect(fill="transparent"),
        axis.line = element_line(colour = "black"))
```




```
# maindata%>%
#   group_by(age)%>%
#   tally() %>%
#   ggplot(aes(x=age,y = n))+
#   geom_bar(stat = 'identity', fill="red")+
#   labs(title="Age Distribution", x="Age", y="Number of People")+
#   theme(panel.background = element_rect(fill="transparent"),
#         axis.line = element_line(colour = "black")) +
#   geom_text(aes(label = n))
```

```
a=table(maindata$gender)
a
```

```
##
##   F   M   T
## 147 433   8
```

```
str(a)
```

```
##  'table' int [1:3(1d)] 147 433 8
##  - attr(*, "dimnames")=List of 1
##   ..$ : chr [1:3] "F" "M" "T"
```

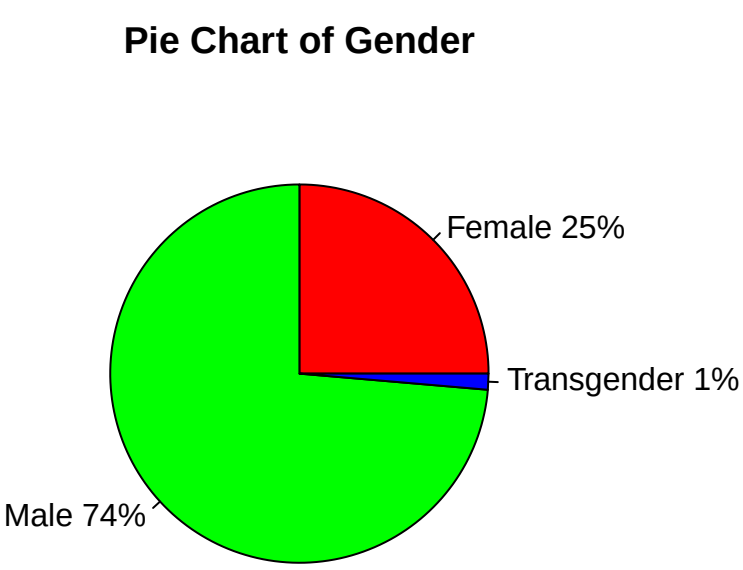
```
gender_count<-as.vector(a)
gender_percentage<-c("Female","Male","Transgender")
gender_count
```

```
## [1] 147 433   8
```

```
gender_percentage
```

```
## [1] "Female"      "Male"        "Transgender"
```

```
pct1 <- round(gender_count/sum(gender_count)*100)
gender_percentage <- paste(gender_percentage, pct1) # add percents to labels
gender_percentage <- paste(gender_percentage,"%",sep="") # ad % to labels
pie(gender_count,labels = gender_percentage, col=rainbow(length(gender_percentage)),
   main="Pie Chart of Gender")
```




```
#state could make a map, still considering
```

```
b=table(maindata$self_employed)
b
```

```
##
##  No Yes
## 540  48
```

```
str(b)
```

```
##  'table' int [1:2(1d)] 540 48
##  - attr(*, "dimnames")=List of 1
##   ..$ : chr [1:2] "No" "Yes"
```

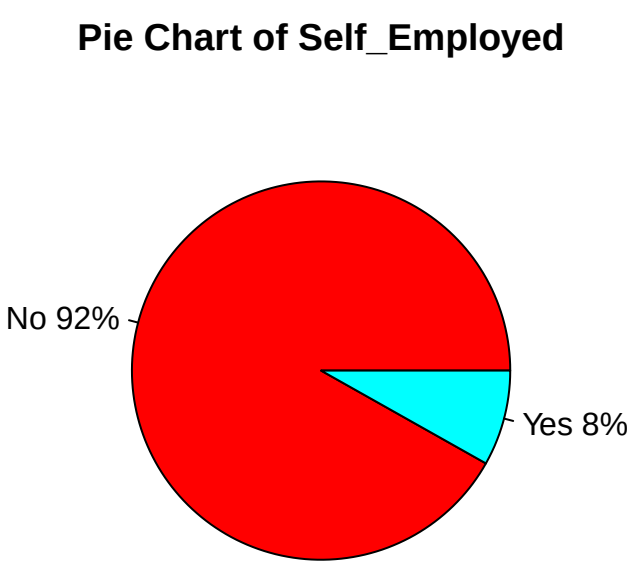
```
self_employed_count<-as.vector(b)
self_employed_percentage<-names(b)
self_employed_count
```

```
## [1] 540  48
```

```
self_employed_percentage
```

```
## [1] "No"  "Yes"
```

```
pct2 <- round(self_employed_count/sum(self_employed_count)*100)
self_employed_percentage <- paste(self_employed_percentage, pct2) # add percents to labels
self_employed_percentage <- paste(self_employed_percentage,"%",sep="") # ad % to labels
pie(self_employed_count,labels = self_employed_percentage, col=rainbow(length(self_employed_percentage)),
   main="Pie Chart of Self_Employed")
```




```
c=table(maindata$family_history)
c
```

```
##
##  No Yes
## 294 294
```

```
str(c)
```

```
##  'table' int [1:2(1d)] 294 294
##  - attr(*, "dimnames")=List of 1
##   ..$ : chr [1:2] "No" "Yes"
```

```
family_history_count<-as.vector(c)
family_history_percentage<-names(c)
family_history_count
```

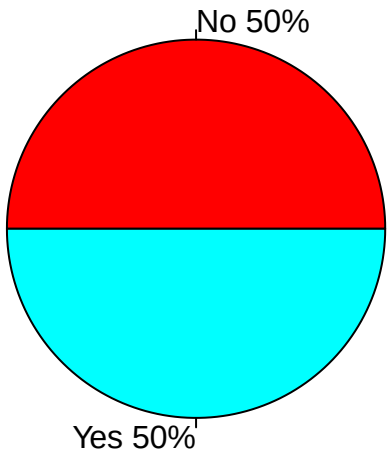
```
## [1] 294 294
```

```
family_history_percentage
```

```
## [1] "No"  "Yes"
```

```
pct3 <- round(family_history_count/sum(family_history_count)*100)
family_history_percentage <- paste(family_history_percentage, pct3) # add percents to labels
family_history_percentage <- paste(family_history_percentage,"%",sep="") # ad % to labels
pie(family_history_count,labels = family_history_percentage, col=rainbow(length(family_history_percentag
   main="Pie Chart of Family History of Mental Health")
```

Pie Chart of Family History of Mental Health




```
#Our outcome varaible
d=table(maindata$mh_treatment)
d
```

```
##
##  No Yes
## 192 396
```

```
str(d)
```

```
##  'table' int [1:2(1d)] 192 396
##  - attr(*, "dimnames")=List of 1
##   ..$ : chr [1:2] "No" "Yes"
```

```
mh_treatment_count<-as.vector(d)
mh_treatment_percentage<-names(d)
mh_treatment_count
```

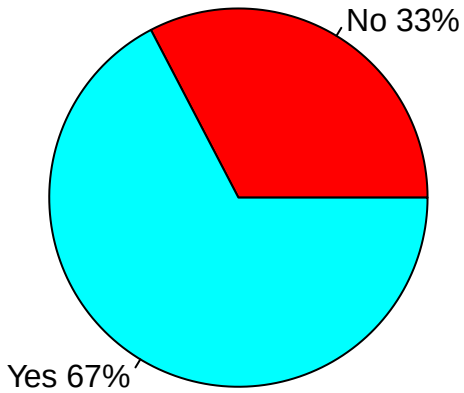
```
## [1] 192 396
```

```
mh_treatment_percentage
```

```
## [1] "No"  "Yes"
```

```
pct4 <- round(mh_treatment_count/sum(mh_treatment_count)*100)
mh_treatment_percentage <- paste(mh_treatment_percentage, pct4) # add percents to labels
mh_treatment_percentage <- paste(mh_treatment_percentage,"%",sep="") # ad % to labels
pie(mh_treatment_count,labels = mh_treatment_percentage, col=rainbow(length(mh_treatment_percentage)),
   main="Pie Chart of Whether seeking Mental Health Treatment")
```

Pie Chart of Whether seeking Mental Health Treatment




```
e=table(maindata$interfere)
e
```

```
##
##     Never     Often    Rarely Sometimes
##       121        78       107       282
```

```
str(e)
```

```
##  'table' int [1:4(1d)] 121 78 107 282
##  - attr(*, "dimnames")=List of 1
##   ..$ : chr [1:4] "Never" "Often" "Rarely" "Sometimes"
```

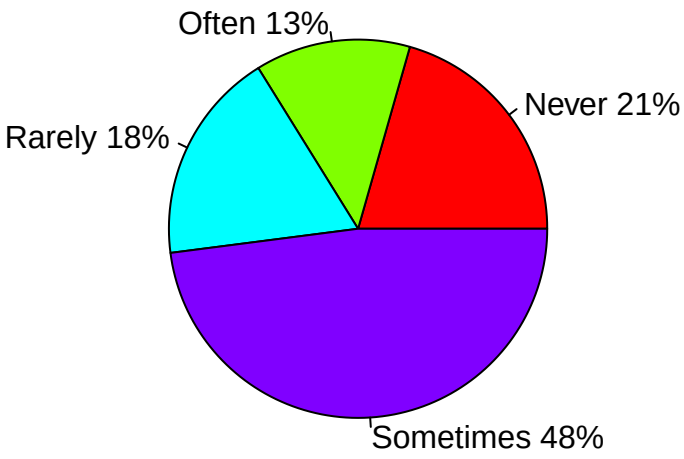
```
interfere_count<-as.vector(e)
interfere_percentage<-names(e)
interfere_count
```

```
## [1] 121  78 107 282
```

```
interfere_percentage
```

```
## [1] "Never"     "Often"     "Rarely"    "Sometimes"
```

```
pct5 <- round(interfere_count/sum(interfere_count)*100)
interfere_percentage <- paste(interfere_percentage, pct5) # add percents to labels
interfere_percentage <- paste(interfere_percentage,"%",sep="") # ad % to labels
pie(interfere_count,labels = interfere_percentage, col=rainbow(length(interfere_percentage)),
   main="Pie Chart of Interfere")
```




```
f=table(maindata$company_size)
f
```

```
##
##            1-5        100-500         26-100       500-1000           6-25
##             67             90            136             30            100
## More than 1000
##            165
```

```
str(f)
```

```
##  'table' int [1:6(1d)] 67 90 136 30 100 165
##  - attr(*, "dimnames")=List of 1
##   ..$ : chr [1:6] "1-5" "100-500" "26-100" "500-1000" ...
```

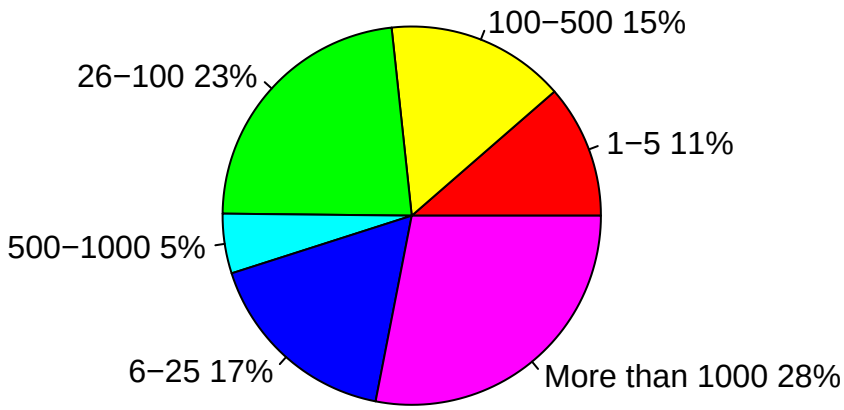
```
company_size_count<-as.vector(f)
company_size_percentage<-names(f)
company_size_count
```

```
## [1]  67  90 136  30 100 165
```

```
company_size_percentage
```

```
## [1] "1-5"           "100-500"       "26-100"        "500-1000"
## [5] "6-25"          "More than 1000"
```

```
pct6 <- round(company_size_count/sum(company_size_count)*100)
company_size_percentage <- paste(company_size_percentage, pct6) # add percents to labels
company_size_percentage <- paste(company_size_percentage,"%",sep="") # ad % to labels
pie(company_size_count,labels = company_size_percentage, col=rainbow(length(company_size_percentage)),
   main="Pie Chart of Company Size")
```




```
g=table(maindata$remote)
g
```

```
##
##  No Yes
## 403 185
```

```
str(g)
```

```
##  'table' int [1:2(1d)] 403 185
##  - attr(*, "dimnames")=List of 1
##   ..$ : chr [1:2] "No" "Yes"
```

```
remote_count<-as.vector(g)
remote_percentage<-names(g)
remote_count
```

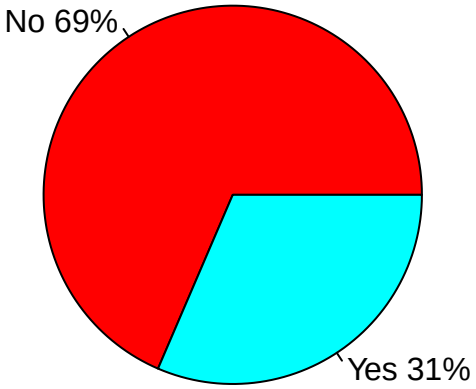
```
## [1] 403 185
```

```
remote_percentage
```

```
## [1] "No"  "Yes"
```

```
pct7 <- round(remote_count/sum(remote_count)*100)
remote_percentage <- paste(remote_percentage, pct7) # add percents to labels
remote_percentage <- paste(remote_percentage,"%",sep="") # ad % to labels
pie(remote_count,labels = remote_percentage, col=rainbow(length(remote_percentage)),
   main="Pie Chart of Remote")
```

Pie Chart of Remote




```
h=table(maindata$tech_company)
h
```

```
##
##  No Yes
## 105 483
```

```
str(h)
```

```
##  'table' int [1:2(1d)] 105 483
##  - attr(*, "dimnames")=List of 1
##   ..$ : chr [1:2] "No" "Yes"
```

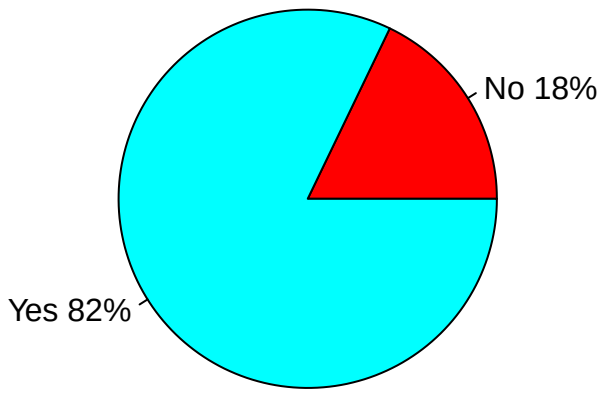
```
tech_company_count<-as.vector(h)
tech_company_percentage<-names(h)
tech_company_count
```

```
## [1] 105 483
```

```
tech_company_percentage
```

```
## [1] "No"  "Yes"
```

```
pct8 <- round(tech_company_count/sum(tech_company_count)*100)
tech_company_percentage <- paste(tech_company_percentage, pct8) # add percents to labels
tech_company_percentage <- paste(tech_company_percentage,"%",sep="") # ad % to labels
pie(tech_company_count,labels = tech_company_percentage, col=rainbow(length(tech_company_percentage)),
   main="Pie Chart of Tech Company")
```




```
i=table(maindata$mh_benefits)
i
```

```
##
## Don't know         No        Yes
##        163         97        328
```

```
str(i)
```

```
##  'table' int [1:3(1d)] 163 97 328
##  - attr(*, "dimnames")=List of 1
##   ..$ : chr [1:3] "Don't know" "No" "Yes"
```

```
mh_benefits_count<-as.vector(i)
mh_benefits_percentage<-names(i)
mh_benefits_count
```

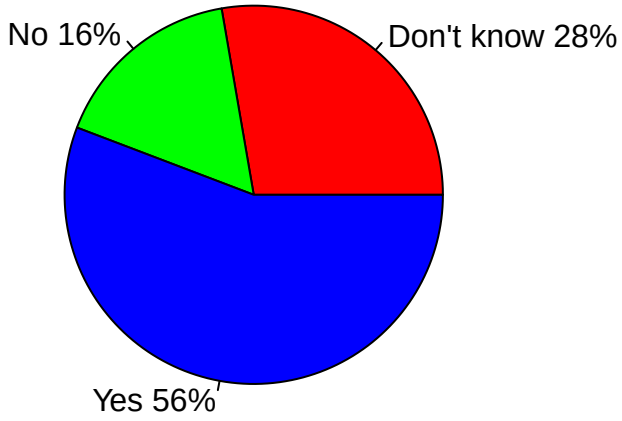
```
## [1] 163  97 328
```

```
mh_benefits_percentage
```

```
## [1] "Don't know" "No"         "Yes"
```

```
pct9 <- round(mh_benefits_count/sum(mh_benefits_count)*100)
mh_benefits_percentage <- paste(mh_benefits_percentage, pct9) # add percents to labels
mh_benefits_percentage <- paste(mh_benefits_percentage,"%",sep="") # ad % to labels
pie(mh_benefits_count,labels = mh_benefits_percentage, col=rainbow(length(mh_benefits_percentage)),
   main="Pie Chart of Mental Health Benefits")
```

Pie Chart of Mental Health Benefits




```
j=table(maindata$awareness_mh_benefits)
j
```

```
##
##       No Not sure      Yes
##      173      141      274
```

```
str(j)
```

```
##  'table' int [1:3(1d)] 173 141 274
##  - attr(*, "dimnames")=List of 1
##   ..$ : chr [1:3] "No" "Not sure" "Yes"
```

```
awareness_mh_benefits_count<-as.vector(j)
awareness_mh_benefits_percentage<-names(j)
awareness_mh_benefits_count
```

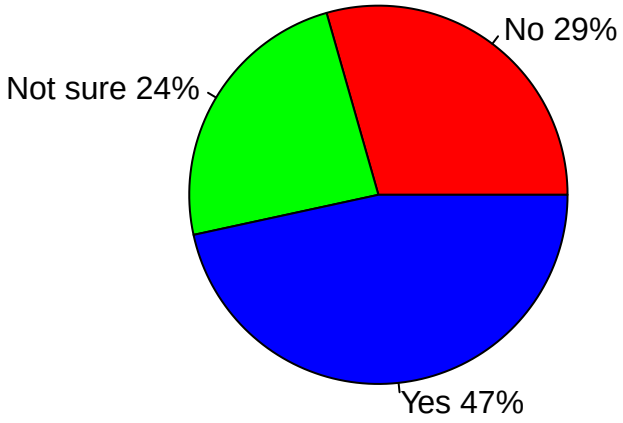
```
## [1] 173 141 274
```

```
awareness_mh_benefits_percentage
```

```
## [1] "No"       "Not sure" "Yes"
```

```
pct10 <- round(awareness_mh_benefits_count/sum(awareness_mh_benefits_count)*100)
awareness_mh_benefits_percentage <- paste(awareness_mh_benefits_percentage, pct10) # add percents to la
awareness_mh_benefits_percentage <- paste(awareness_mh_benefits_percentage,"%",sep="") # ad % to labels
pie(awareness_mh_benefits_count,labels = awareness_mh_benefits_percentage, col=rainbow(length(awareness_
   main="Pie Chart of Awarenessof Mental Health Benefits")
```

Pie Chart of Awarenessof Mental Health Benefits




```
k=table(maindata$mh_discuss)
k
```

```
##
## Don't know         No        Yes
##         91        358        139
```

```
str(k)
```

```
##  'table' int [1:3(1d)] 91 358 139
##  - attr(*, "dimnames")=List of 1
##   ..$ : chr [1:3] "Don't know" "No" "Yes"
```

```
mh_discuss_count<-as.vector(k)
mh_discuss_percentage<-names(k)
mh_discuss_count
```

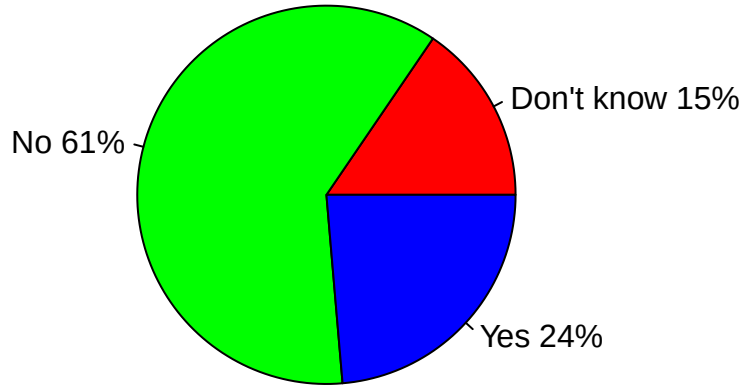
```
## [1]  91 358 139
```

```
mh_discuss_percentage
```

```
## [1] "Don't know" "No"         "Yes"
```

```
pct11 <- round(mh_discuss_count/sum(mh_discuss_count)*100)
mh_discuss_percentage <- paste(mh_discuss_percentage, pct11) # add percents to labels
mh_discuss_percentage <- paste(mh_discuss_percentage,"%",sep="") # ad % to labels
pie(mh_discuss_count,labels = mh_discuss_percentage, col=rainbow(length(mh_discuss_percentage)),
   main="Pie Chart of Mental Health Discussion")
```

Pie Chart of Mental Health Discussion




```
l=table(maindata$mh_resources)
l
```

```
##
## Don't know         No        Yes
##        194        236        158
```

```
str(l)
```

```
##  'table' int [1:3(1d)] 194 236 158
##  - attr(*, "dimnames")=List of 1
##   ..$ : chr [1:3] "Don't know" "No" "Yes"
```

```
mh_resources_count<-as.vector(l)
mh_resources_percentage<-names(l)
mh_resources_count
```

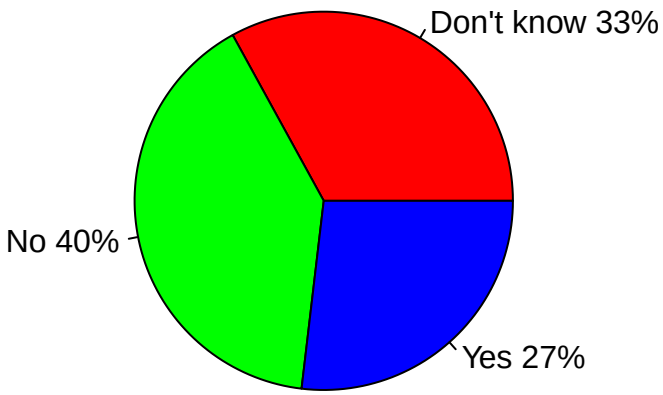
```
## [1] 194 236 158
```

```
mh_resources_percentage
```

```
## [1] "Don't know" "No"         "Yes"
```

```
pct12 <- round(mh_resources_count/sum(mh_resources_count)*100)
mh_resources_percentage <- paste(mh_resources_percentage, pct12) # add percents to labels
mh_resources_percentage <- paste(mh_resources_percentage,"%",sep="") # ad % to labels
pie(mh_resources_count,labels = mh_resources_percentage, col=rainbow(length(mh_resources_percentage)),
   main="Pie Chart of Mental Health Resources")
```

Pie Chart of Mental Health Resources




```
m=table(maindata$anonymity_protected)
m
```

```
##
## Don't know         No        Yes
##        379         15        194
```

```
str(m)
```

```
##  'table' int [1:3(1d)] 379 15 194
##  - attr(*, "dimnames")=List of 1
##   ..$ : chr [1:3] "Don't know" "No" "Yes"
```

```
anonymity_protected_count<-as.vector(m)
anonymity_protected_percentage<-names(m)
anonymity_protected_count
```

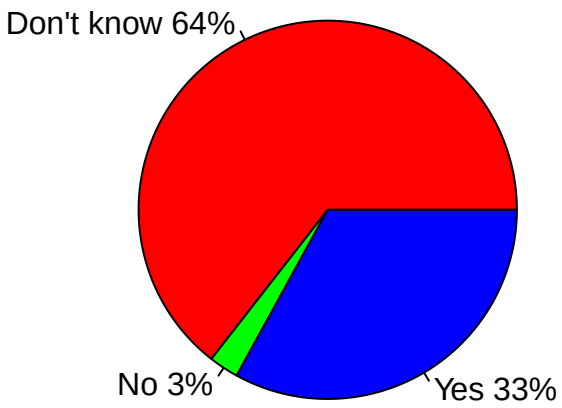
```
## [1] 379  15 194
```

```
anonymity_protected_percentage
```

```
## [1] "Don't know" "No"         "Yes"
```

```
pct13 <- round(anonymity_protected_count/sum(anonymity_protected_count)*100)
anonymity_protected_percentage <- paste(anonymity_protected_percentage, pct13) # add percents to labels
anonymity_protected_percentage <- paste(anonymity_protected_percentage,"%",sep="") # ad % to labels
pie(anonymity_protected_count,labels = anonymity_protected_percentage, col=rainbow(length(anonymity_pro
   main="Pie Chart of Anonymity Mental Health Protection")
```

Pie Chart of Anonymity Mental Health Protection




```
n=table(maindata$medical_leave_easy)
n
```

```
##
##         Don't know Somewhat difficult      Somewhat easy
##                286                  59                114
##     Very difficult          Very easy
##                 47                  82
```

```
str(n)
```

```
##  'table' int [1:5(1d)] 286 59 114 47 82
##  - attr(*, "dimnames")=List of 1
##   ..$ : chr [1:5] "Don't know" "Somewhat difficult" "Somewhat easy" "Very difficult" ...
```

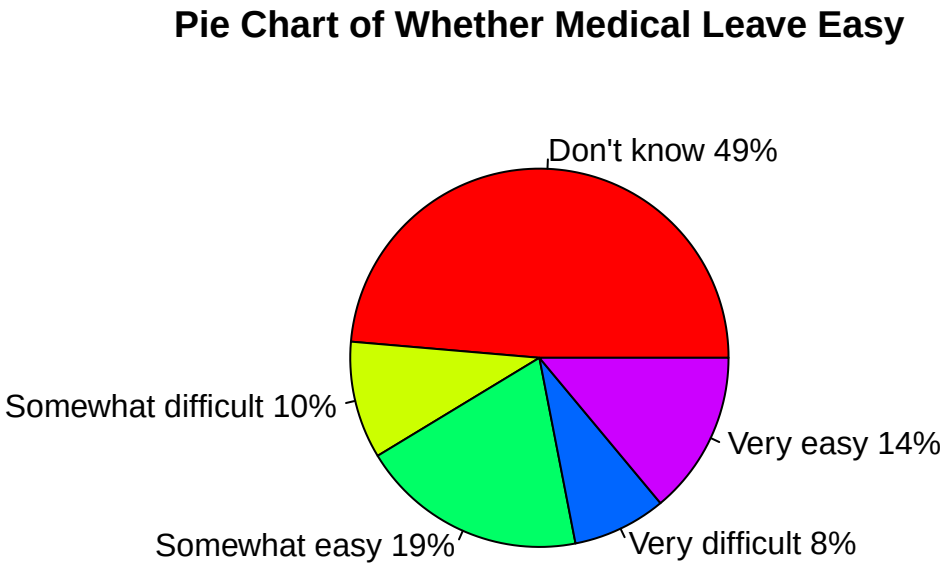
```
medical_leave_easy_count<-as.vector(n)
medical_leave_easy_percentage<-names(n)
medical_leave_easy_count
```

```
## [1] 286  59 114  47  82
```

```
medical_leave_easy_percentage
```

```
## [1] "Don't know"         "Somewhat difficult" "Somewhat easy"
## [4] "Very difficult"     "Very easy"
```

```
pct14 <- round(medical_leave_easy_count/sum(medical_leave_easy_count)*100)
medical_leave_easy_percentage <- paste(medical_leave_easy_percentage, pct14) # add percents to labels
medical_leave_easy_percentage <- paste(medical_leave_easy_percentage,"%",sep="") # ad % to labels
pie(medical_leave_easy_count,labels = medical_leave_easy_percentage, col=rainbow(length(medical_leave_ea
   main="Pie Chart of Whether Medical Leave Easy")
```




```
o=table(maindata$mh_negative_consequence_flag)
o
```

```
##
## Maybe    No   Yes
##   244   195   149
```

```
str(o)
```

```
##  'table' int [1:3(1d)] 244 195 149
##  - attr(*, "dimnames")=List of 1
##   ..$ : chr [1:3] "Maybe" "No" "Yes"
```

```
mh_negative_consequence_count<-as.vector(o)
mh_negative_consequence_percentage<-names(o)
mh_negative_consequence_count
```

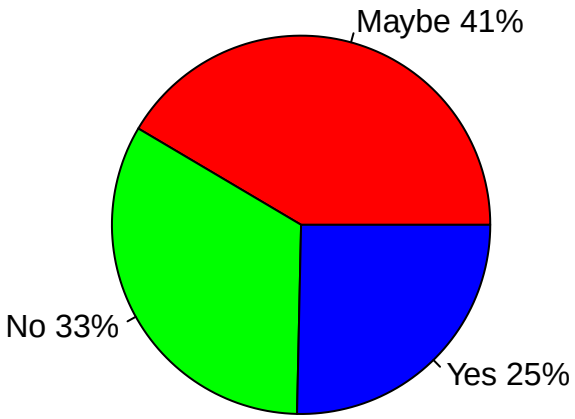
```
## [1] 244 195 149
```

```
mh_negative_consequence_percentage
```

```
## [1] "Maybe" "No"    "Yes"
```

```
pct15 <- round(mh_negative_consequence_count/sum(mh_negative_consequence_count)*100)
mh_negative_consequence_percentage <- paste(mh_negative_consequence_percentage, pct15) # add percents t
mh_negative_consequence_percentage <- paste(mh_negative_consequence_percentage,"%",sep="") # ad % to la
pie(mh_negative_consequence_count,labels = mh_negative_consequence_percentage, col=rainbow(length(mh_neg
   main="Whether Mental Health have Negative Consequence for Interview")
```

Whether Mental Health have Negative Consequence for Interview




```
p=table(maindata$ph_negative_consequence_flag)
p
```

```
##
## Maybe    No   Yes
##   126   437    25
```

```
str(p)
```

```
##  'table' int [1:3(1d)] 126 437 25
##  - attr(*, "dimnames")=List of 1
##   ..$ : chr [1:3] "Maybe" "No" "Yes"
```

```
ph_negative_consequence_flag_count<-as.vector(p)
ph_negative_consequence_flag_percentage<-names(p)
ph_negative_consequence_flag_count
```

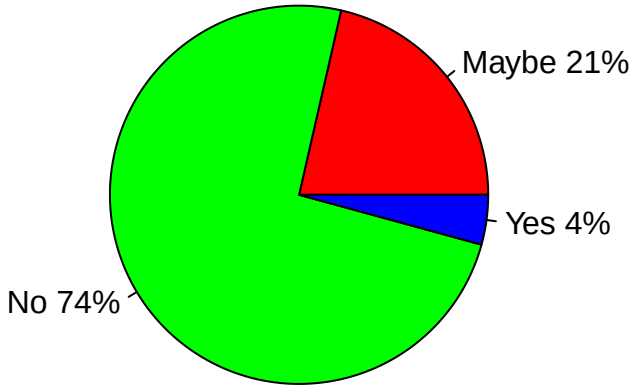
```
## [1] 126 437  25
```

```
ph_negative_consequence_flag_percentage
```

```
## [1] "Maybe" "No"    "Yes"
```

```
pct16 <- round(ph_negative_consequence_flag_count/sum(ph_negative_consequence_flag_count)*100)
ph_negative_consequence_flag_percentage <- paste(ph_negative_consequence_flag_percentage, pct16) # add
ph_negative_consequence_flag_percentage <- paste(ph_negative_consequence_flag_percentage,"%",sep="") #
pie(ph_negative_consequence_flag_count,labels = ph_negative_consequence_flag_percentage, col=rainbow(le
   main="Whether Physical Health have Negative Consequence for Interview")
```

Whether Physical Health have Negative Consequence for Interview




```
q=table(maindata$mh_disscuss_coworker)
q
```

```
##
##           No Some of them          Yes
##          123          373           92
```

```
str(q)
```

```
##  'table' int [1:3(1d)] 123 373 92
##  - attr(*, "dimnames")=List of 1
##   ..$ : chr [1:3] "No" "Some of them" "Yes"
```

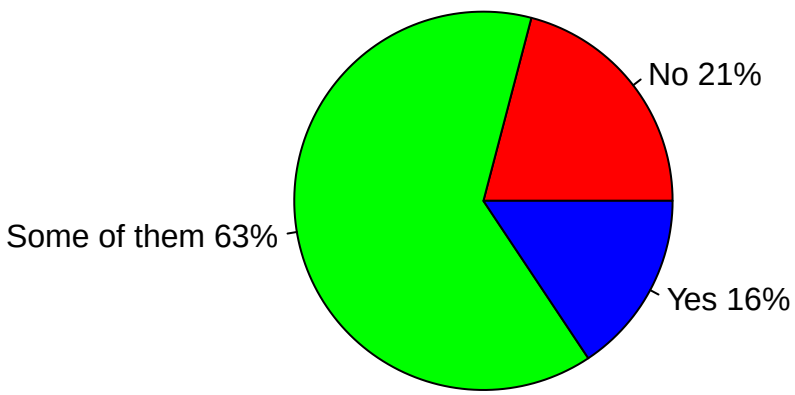
```
mh_disscuss_coworker_count<-as.vector(q)
mh_disscuss_coworker_percentage<-names(q)
mh_disscuss_coworker_count
```

```
## [1] 123 373  92
```

```
mh_disscuss_coworker_percentage
```

```
## [1] "No"           "Some of them" "Yes"
```

```
pct17 <- round(mh_disscuss_coworker_count/sum(mh_disscuss_coworker_count)*100)
mh_disscuss_coworker_percentage <- paste(mh_disscuss_coworker_percentage, pct17) # add percents to labe
mh_disscuss_coworker_percentage <- paste(mh_disscuss_coworker_percentage,"%",sep="") # ad % to labels
pie(mh_disscuss_coworker_count,labels = mh_disscuss_coworker_percentage, col=rainbow(length(mh_disscuss
   main="Whether Disscuss Mental Health with Coworker")
```




```
r=table(maindata$mh_disscuss_supervisor)
r
```

```
##
##           No Some of them          Yes
##          190          176          222
```

```
str(r)
```

```
##  'table' int [1:3(1d)] 190 176 222
##  - attr(*, "dimnames")=List of 1
##   ..$ : chr [1:3] "No" "Some of them" "Yes"
```

```
mh_disscuss_supervisor_count<-as.vector(r)
mh_disscuss_supervisor_percentage<-names(r)
mh_disscuss_supervisor_count
```

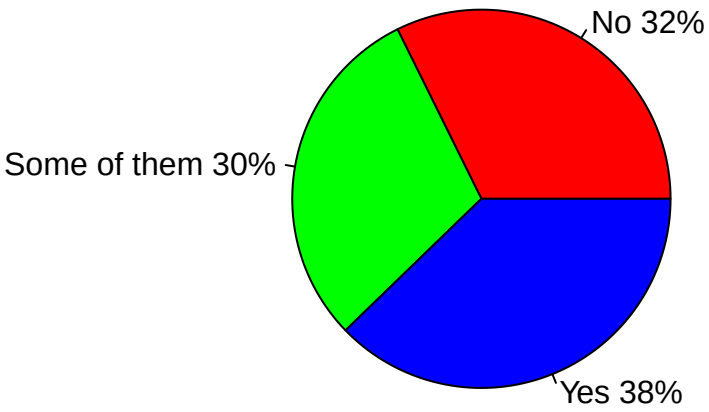
```
## [1] 190 176 222
```

```
mh_disscuss_supervisor_percentage
```

```
## [1] "No"           "Some of them" "Yes"
```

```
pct18 <- round(mh_disscuss_supervisor_count/sum(mh_disscuss_supervisor_count)*100)
mh_disscuss_supervisor_percentage <- paste(mh_disscuss_supervisor_percentage, pct18) # add percents to
mh_disscuss_supervisor_percentage <- paste(mh_disscuss_supervisor_percentage,"%",sep="") # ad % to labe
pie(mh_disscuss_supervisor_count,labels = mh_disscuss_supervisor_percentage, col=rainbow(length(mh_disso
   main="Whether Disscuss Mental Health with Supervisor")
```

Whether Disscuss Mental Health with Supervisor




```
s=table(maindata$interview_mh_bringup)
s
```

```
##
## Maybe    No   Yes
##    74   502    12
```

```
str(s)
```

```
##  'table' int [1:3(1d)] 74 502 12
##  - attr(*, "dimnames")=List of 1
##   ..$ : chr [1:3] "Maybe" "No" "Yes"
```

```
interview_mh_bringup_count<-as.vector(s)
interview_mh_bringup_percentage<-names(s)
interview_mh_bringup_count
```

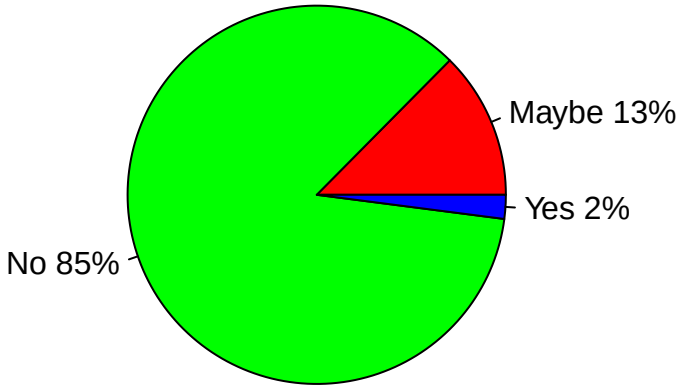
```
## [1]  74 502  12
```

```
interview_mh_bringup_percentage
```

```
## [1] "Maybe" "No"    "Yes"
```

```
pct19 <- round(interview_mh_bringup_count/sum(interview_mh_bringup_count)*100)
interview_mh_bringup_percentage <- paste(interview_mh_bringup_percentage, pct19) # add percents to labe
interview_mh_bringup_percentage <- paste(interview_mh_bringup_percentage,"%",sep="") # ad % to labels
pie(interview_mh_bringup_count,labels = interview_mh_bringup_percentage, col=rainbow(length(interview_mh
   main="Whether Bring Up Mental Health into Interview")
```

Whether Bring Up Mental Health into Interview




```
t=table(maindata$interview_ph_bringup)
t
```

```
##
## Maybe    No   Yes
##   251   262    75
```

```
str(t)
```

```
##  'table' int [1:3(1d)] 251 262 75
##  - attr(*, "dimnames")=List of 1
##   ..$ : chr [1:3] "Maybe" "No" "Yes"
```

```
interview_ph_bringup_count<-as.vector(t)
interview_ph_bringup_percentage<-names(t)
interview_ph_bringup_count
```

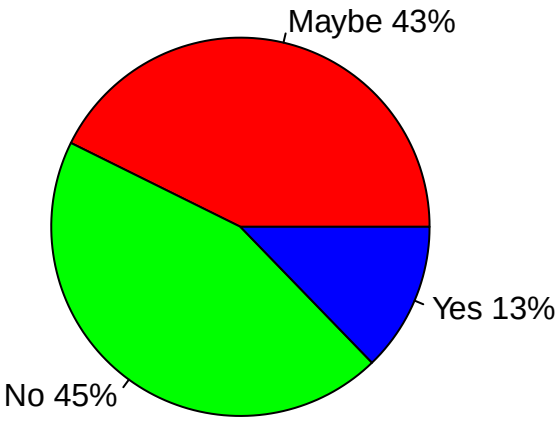
```
## [1] 251 262  75
```

```
interview_ph_bringup_percentage
```

```
## [1] "Maybe" "No"    "Yes"
```

```
pct20 <- round(interview_ph_bringup_count/sum(interview_ph_bringup_count)*100)
interview_ph_bringup_percentage <- paste(interview_ph_bringup_percentage, pct20) # add percents to labe
interview_ph_bringup_percentage <- paste(interview_ph_bringup_percentage,"%",sep="") # ad % to labels
pie(interview_ph_bringup_count,labels = interview_ph_bringup_percentage, col=rainbow(length(interview_pl
   main="Whether Disscuss Physical Health with Coworker")
```

Whether Disscuss Physical Health with Coworker




```
u=table(maindata$mh_serious_ph)
u
```

```
##
## Don't know         No        Yes
##        275        158        155
```

```
str(u)
```

```
##  'table' int [1:3(1d)] 275 158 155
##  - attr(*, "dimnames")=List of 1
##   ..$ : chr [1:3] "Don't know" "No" "Yes"
```

```
mh_serious_ph_count<-as.vector(u)
mh_serious_ph_percentage<-names(u)
mh_serious_ph_count
```

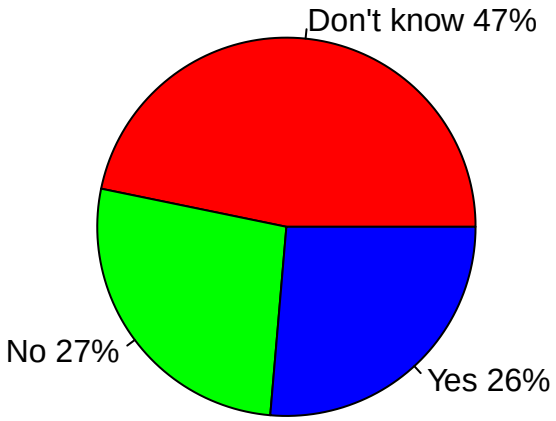
```
## [1] 275 158 155
```

```
mh_serious_ph_percentage
```

```
## [1] "Don't know" "No"         "Yes"
```

```
pct21 <- round(mh_serious_ph_count/sum(mh_serious_ph_count)*100)
mh_serious_ph_percentage <- paste(mh_serious_ph_percentage, pct21) # add percents to labels
mh_serious_ph_percentage <- paste(mh_serious_ph_percentage,"%",sep="") # ad % to labels
pie(mh_serious_ph_count,labels = mh_serious_ph_percentage, col=rainbow(length(mh_serious_ph_percentage)),
   main="Mental Health vs. Physical Health")
```

Mental Health vs. Physical Health




```
v=table(maindata$witness_mh_nc)
v
```

```
##
##  No Yes
## 509  79
```

```
str(v)
```

```
##  'table' int [1:2(1d)] 509 79
##  - attr(*, "dimnames")=List of 1
##   ..$ : chr [1:2] "No" "Yes"
```

```
witness_mh_nc_count<-as.vector(v)
witness_mh_nc_percentage<-names(v)
witness_mh_nc_count
```

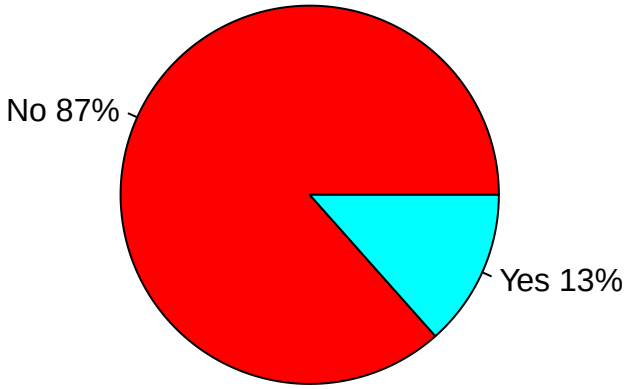
```
## [1] 509  79
```

```
witness_mh_nc_percentage
```

```
## [1] "No"  "Yes"
```

```
pct22 <- round(witness_mh_nc_count/sum(witness_mh_nc_count)*100)
witness_mh_nc_percentage <- paste(witness_mh_nc_percentage, pct22) # add percents to labels
witness_mh_nc_percentage <- paste(witness_mh_nc_percentage,"%",sep="") # ad % to labels
pie(witness_mh_nc_count,labels = witness_mh_nc_percentage, col=rainbow(length(witness_mh_nc_percentage)),
   main="Whether Witness the Mental Health Negative Consequence")
```

Whether Witness the Mental Health Negative Consequence



Encoding Data

```
#Make a copy to make us life easier and could use the categorical dataset when we need
dummydata<-maindata
```

```
#Give numbers for those variable could have a order
dummydata$age=
  ifelse(dummydata$age<=20,1,
         ifelse(dummydata$age>20 & dummydata$age<=40 ,2,
                ifelse(dummydata$age>40 & dummydata$age<=60,3,4)))
dummydata$self_employed<-ifelse(dummydata$self_employed=="Yes", 1,0)
dummydata$family_history<-ifelse(dummydata$family_history=="Yes", 1,0)
dummydata$mh_treatment<-ifelse(dummydata$mh_treatment=="Yes", 1,0)
dummydata$interfere=
  ifelse(dummydata$interfere=='Never',1,
         ifelse(dummydata$interfere=='Rarely',2,
                ifelse(dummydata$interfere=="Sometimes",3,4)))
dummydata$company_size=
  ifelse(dummydata$company_size=='1-5',1,
         ifelse(dummydata$company_size=='6-25',2,
                ifelse(dummydata$company_size=="26-100",3,
                       ifelse(dummydata$company_size=="100-500",4,
                              ifelse(dummydata$company_size=="500-1000",5,6)))))
dummydata$remote<-ifelse(dummydata$remote=="Yes", 1,0)
dummydata$tech_company<-ifelse(dummydata$tech_company=="Yes", 1,0)
dummydata$mh_negative_consequence_flag=
  ifelse(dummydata$mh_negative_consequence_flag=='No',0,
         ifelse(dummydata$mh_negative_consequence_flag=='Maybe',1,2))
dummydata$ph_negative_consequence_flag=
  ifelse(dummydata$ph_negative_consequence_flag=='No',0,
         ifelse(dummydata$ph_negative_consequence_flag=='Maybe',1,2))
dummydata$mh_disscuss_coworker=
  ifelse(dummydata$mh_disscuss_coworker=='No',0,
         ifelse(dummydata$mh_disscuss_coworker=='Some of them',1,2))
dummydata$mh_disscuss_supervisor=
  ifelse(dummydata$mh_disscuss_supervisor=='No',0,
```

```
        ifelse(dummydata$mh_disscuss_supervisor=='Some of them',1,2))
dummydata$interview_mh_bringup=
  ifelse(dummydata$interview_mh_bringup=='No',0,
        ifelse(dummydata$interview_mh_bringup=='Maybe',1,2))
dummydata$interview_ph_bringup=
  ifelse(dummydata$interview_ph_bringup=='No',0,
        ifelse(dummydata$interview_ph_bringup=='Maybe',1,2))
dummydata$witness_mh_nc<-ifelse(dummydata$witness_mh_nc=="Yes", 1,0)
```

```
skim(dummydata)
```

```
## Skim summary statistics
##  n obs: 588
##  n variables: 24
##
## -- Variable type:character ---------------------------
##               variable missing complete   n min max empty n_unique
##    anonymity_protected       0      588 588   2  10     0        3
##  awareness_mh_benefits       0      588 588   2   8     0        3
##                 gender       0      588 588   1   1     0        3
##     medical_leave_easy       0      588 588   9  18     0        5
##            mh_benefits       0      588 588   2  10     0        3
##             mh_discuss       0      588 588   2  10     0        3
##          mh_resources       0      588 588   2  10     0        3
##         mh_serious_ph       0      588 588   2  10     0        3
##                 state       0      588 588   2   2     0       43
##
## -- Variable type:numeric ----------------------------
##                      variable missing complete   n  mean   sd p0 p25 p50
##                           age       0      588 588 2.14  0.4   1   2 2
##                  company_size       0      588 588 3.7   1.74  1   2 3
##                family_history       0      588 588 0.5   0.5   0   0 0.5
##                     interfere       0      588 588 2.54  0.96  1   2 3
##          interview_mh_bringup       0      588 588 0.17  0.42  0   0 0
##          interview_ph_bringup       0      588 588 0.68  0.69  0   0 1
##          mh_disscuss_coworker       0      588 588 0.95  0.6   0   1 1
##        mh_disscuss_supervisor       0      588 588 1.05  0.84  0   0 1
##  mh_negative_consequence_flag       0      588 588 0.92  0.76  0   0 1
##                 mh_treatment       0      588 588 0.67  0.47  0   0 1
##  ph_negative_consequence_flag       0      588 588 0.3   0.54  0   0 0
##                        remote       0      588 588 0.31  0.46  0   0 0
##                self_employed       0      588 588 0.082 0.27  0   0 0
##                 tech_company       0      588 588 0.82  0.38  0   1 1
##                witness_mh_nc       0      588 588 0.13  0.34  0   0 0
##  p75 p100     hist
##    2    4
##    6    6
##    1    1
##    3    4
##    0    2
##    1    2
##    1    2
##    2    2
```

```
##    2    2
##    1    1
##    1    2
##    1    1
##    0    1
##    1    1
##    0    1
```

```
#use this for some variable include "I don't know" or "Not sure" and for variable "gender"
dummydata<-dummy_cols(dummydata, select_columns = c("anonymity_protected", "awareness_mh_benefits","gen
```

```
#after adding dummies, remove original categorical variables
dummydata <- dummydata[-c(2,11:17,23)]
```

```
#We would try with state dummys or delete state when we do linear regression
dummydata<-dummy_cols(dummydata, select_columns = c("state"))
```

```
#leave numeric variables and all dummies with state as dummies
dummydata<-dummydata[-2]
dim(dummydata)
```

```
## [1] 588  83
```

```
names(dummydata)
```

```
##  [1] "age"
##  [2] "self_employed"
##  [3] "family_history"
##  [4] "mh_treatment"
##  [5] "interfere"
##  [6] "company_size"
##  [7] "remote"
##  [8] "tech_company"
##  [9] "ph_negative_consequence_flag"
## [10] "mh_disscuss_coworker"
## [11] "mh_disscuss_supervisor"
## [12] "interview_mh_bringup"
## [13] "interview_ph_bringup"
## [14] "witness_mh_nc"
## [15] "anonymity_protected_Yes"
## [16] "anonymity_protected_No"
## [17] "anonymity_protected_Don't know"
## [18] "awareness_mh_benefits_Not sure"
## [19] "awareness_mh_benefits_Yes"
## [20] "awareness_mh_benefits_No"
## [21] "gender_M"
## [22] "gender_F"
## [23] "gender_T"
## [24] "medical_leave_easy_Very easy"
## [25] "medical_leave_easy_Somewhat difficult"
## [26] "medical_leave_easy_Don't know"
## [27] "medical_leave_easy_Somewhat easy"
```

```
## [28] "medical_leave_easy_Very difficult"
## [29] "mh_benefits_Yes"
## [30] "mh_benefits_No"
## [31] "mh_benefits_Don't know"
## [32] "mh_discuss_Yes"
## [33] "mh_discuss_No"
## [34] "mh_discuss_Don't know"
## [35] "mh_resources_Don't know"
## [36] "mh_resources_No"
## [37] "mh_resources_Yes"
## [38] "mh_serious_ph_Yes"
## [39] "mh_serious_ph_No"
## [40] "mh_serious_ph_Don't know"
## [41] "state_MD"
## [42] "state_NY"
## [43] "state_NC"
## [44] "state_MA"
## [45] "state_IA"
## [46] "state_CA"
## [47] "state_TN"
## [48] "state_OH"
## [49] "state_PA"
## [50] "state_WA"
## [51] "state_WI"
## [52] "state_TX"
## [53] "state_IN"
## [54] "state_IL"
## [55] "state_UT"
## [56] "state_OR"
## [57] "state_FL"
## [58] "state_MN"
## [59] "state_MO"
## [60] "state_AZ"
## [61] "state_CT"
## [62] "state_GA"
## [63] "state_NE"
## [64] "state_WV"
## [65] "state_OK"
## [66] "state_MI"
## [67] "state_KS"
## [68] "state_NH"
## [69] "state_KY"
## [70] "state_AL"
## [71] "state_VA"
## [72] "state_CO"
## [73] "state_SC"
## [74] "state_SD"
## [75] "state_NV"
## [76] "state_NJ"
## [77] "state_ID"
## [78] "state_MS"
## [79] "state_WY"
## [80] "state_LA"
## [81] "state_DC"
```

```
## [82] "state_ME"
## [83] "state_VT"
```

```
#leave numeric variables and all dummies with state as dummies
dummydata2<-dummydata[-c(41:83)]
dim(dummydata2)
```

```
## [1] 588  40
```

```
names(dummydata2)
```

```
##  [1] "age"
##  [2] "self_employed"
##  [3] "family_history"
##  [4] "mh_treatment"
##  [5] "interfere"
##  [6] "company_size"
##  [7] "remote"
##  [8] "tech_company"
##  [9] "ph_negative_consequence_flag"
## [10] "mh_disscuss_coworker"
## [11] "mh_disscuss_supervisor"
## [12] "interview_mh_bringup"
## [13] "interview_ph_bringup"
## [14] "witness_mh_nc"
## [15] "anonymity_protected_Yes"
## [16] "anonymity_protected_No"
## [17] "anonymity_protected_Don't know"
## [18] "awareness_mh_benefits_Not sure"
## [19] "awareness_mh_benefits_Yes"
## [20] "awareness_mh_benefits_No"
## [21] "gender_M"
## [22] "gender_F"
## [23] "gender_T"
## [24] "medical_leave_easy_Very easy"
## [25] "medical_leave_easy_Somewhat difficult"
## [26] "medical_leave_easy_Don't know"
## [27] "medical_leave_easy_Somewhat easy"
## [28] "medical_leave_easy_Very difficult"
## [29] "mh_benefits_Yes"
## [30] "mh_benefits_No"
## [31] "mh_benefits_Don't know"
## [32] "mh_discuss_Yes"
## [33] "mh_discuss_No"
## [34] "mh_discuss_Don't know"
## [35] "mh_resources_Don't know"
## [36] "mh_resources_No"
## [37] "mh_resources_Yes"
## [38] "mh_serious_ph_Yes"
## [39] "mh_serious_ph_No"
## [40] "mh_serious_ph_Don't know"
```