Mental Health in Tech Survey - Part 1

Oscar Monroy

5/19/2022

Timestamp

Age

Gender

Country

state: If you live in the United States, which state or territory do you live in?

self_employed: Are you self-employed?

family_history: Do you have a family history of mental illness?

treatment: Have you sought treatment for a mental health condition?

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

no_employees: How many employees does your company or organization have?

remote_work: Do you work remotely (outside of an office) at least 50% of the time?

tech_company: Is your employer primarily a tech company/organization?

benefits: Does your employer provide mental health benefits?

care_options: Do you know the options for mental health care your employer provides?

wellness_program: Has your employer ever discussed mental health as part of an
employee wellness program?

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

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

leave: How easy is it for you to take medical leave for a mental health condition?

mentalhealthconsequence: Do you think that discussing a mental health issue with
your employer would have negative consequences?

physhealthconsequence: Do you think that discussing a physical health issue with your employer would have negative consequences?

coworkers: Would you be willing to discuss a mental health issue with your coworkers?

supervisor: Would you be willing to discuss a mental health issue with your direct supervisor(s)?

mentalhealthinterview: Would you bring up a mental health issue with a potential employer in an interview?

physhealthinterview: Would you bring up a physical health issue with a potential employer in an interview?

mentalvsphysical: Do you feel that your employer takes mental health as seriously as p. health?

obs_consequence: Have you heard of or observed negative consequences for coworkers with mental health conditions in your workplace?

```
mht <- read.csv("survey.csv")
summary(mht) # The data seems to be mostly error free, except Gender and age
```

```
##                Timestamp         Age                 Gender
##  2014-08-27 12:31:41:   2   Min.   :-1.726e+03   Male   :615
##  2014-08-27 12:37:50:   2   1st Qu.: 2.700e+01   male   :206
##  2014-08-27 12:43:28:   2   Median : 3.100e+01   Female :121
##  2014-08-27 12:44:51:   2   Mean   : 7.943e+07   M      :116
##  2014-08-27 12:54:11:   2   3rd Qu.: 3.600e+01   female : 62
##  2014-08-27 14:22:43:   2   Max.   : 1.000e+11   F      : 38
##  (Other)            :1247                        (Other):101
##            Country        state     self_employed family_history treatment
##  United States :751   CA     :138   No  :1095     No :767        No :622
##  United Kingdom:185   WA     : 70   Yes : 146     Yes:492        Yes:637
##  Canada        : 72   NY     : 57   NA's:  18
##  Germany       : 45   TN     : 45
##  Ireland       : 27   TX     : 44
##  Netherlands   : 27   (Other):390
##  (Other)       :152   NA's   :515
##    work_interfere         no_employees remote_work tech_company
##  Never    :213    1-5           :162   No :883     No : 228
##  Often    :144    100-500       :176   Yes:376     Yes:1031
##  Rarely   :173    26-100        :289
##  Sometimes:465    500-1000      : 60
##  NA's     :264    6-25          :290
##                   More than 1000:282
##
##       benefits     care_options   wellness_program      seek_help
##  Don't know:408   No      :501   Don't know:188     Don't know:363
##  No        :374   Not sure:314   No        :842     No        :646
##  Yes       :477   Yes     :444   Yes       :229     Yes       :250
##
##
##
##
##       anonymity                  leave     mental_health_consequence
```

```
##  Don't know:819   Don't know        :563   Maybe:477
##  No        : 65   Somewhat difficult:126   No   :490
##  Yes       :375   Somewhat easy     :266   Yes  :292
##                   Very difficult    : 98
##                   Very easy         :206
##
##
##  phys_health_consequence        coworkers          supervisor
##  Maybe:273               No          :260   No          :393
##  No   :925               Some of them:774   Some of them:350
##  Yes  : 61               Yes         :225   Yes         :516
##
##
##
##
##  mental_health_interview phys_health_interview  mental_vs_physical
##  Maybe: 207              Maybe:557              Don't know:576
##  No   :1008              No   :500              No        :340
##  Yes  :  44              Yes  :202              Yes       :343
##
##
##
##
##  obs_consequence
##  No :1075
##  Yes: 184
##
##
##
##
##
##
##  * Small family business - YMMV.
##  -
##
##  (yes but the situation was unusual and involved a change in leadership at a very high level in the c
##  A close family member of mine struggles with mental health so I try not to stigmatize it. My employe
##  (Other)
##  NA's
```

```
table(mht$Gender) # There's a lot of misspellings here...
```

```
##
##                            A little about you
##                                             1
##                                      Agender
##                                             1
##                                         All
##                                             1
##                                    Androgyne
##                                             1
##                              cis-female/femme
##                                             1
##                                   Cis Female
```

```
##                                     1
##                               cis male
##                                     1
##                               Cis Male
##                                     2
##                                Cis Man
##                                     1
##                                   Enby
##                                     1
##                                     f
##                                    15
##                                     F
##                                    38
##                                 femail
##                                     1
##                                Femake
##                                     1
##                                female
##                                    62
##                                Female
##                                   121
##                               Female
##                                     2
##                           Female (cis)
##                                     1
##                         Female (trans)
##                                     2
##                                 fluid
##                                     1
##                            Genderqueer
##                                     1
##                         Guy (-ish) ^_^
##                                     1
##                                     m
##                                    34
##                                     M
##                                   116
##                                  Mail
##                                     1
##                                 maile
##                                     1
##                                  Make
##                                     4
##                                   Mal
##                                     1
##                                  male
##                                   206
##                                  Male
##                                   615
##                                Male-ish
##                                     1
##                                  Male
##                                     3
##                             Male (CIS)
```

```
##                                           1
##                         male leaning androgynous
##                                           1
##                                        Malr
##                                           1
##                                         Man
##                                           2
##                                        msle
##                                           1
##                                         Nah
##                                           1
##                                      Neuter
##                                           1
##                                  non-binary
##                                           1
## ostensibly male, unsure what that really means
##                                           1
##                                           p
##                                           1
##                                       queer
##                                           1
##                                  queer/she/they
##                                           1
##                           something kinda male?
##                                           1
##                                  Trans-female
##                                           1
##                                   Trans woman
##                                           1
##                                       woman
##                                           1
##                                       Woman
##                                           3
```

```
# Normally, I'd do some regex to change the misspelled levels into
# correctly spelled form, but using the indices for the levels would
# make this job a lot faster. I'll also be making an "Other" level
# to fit all the people who don't fit within the definition of
# cis-gendered. Also because there is a very small smaple size for them.
# Trans women will also be fit into the "Female" category as they choose
# to identify as female. Ultimately, there will be some subjectivity at
# play here; for example, "male leaning androgynous" will be fit in the
# "Male" category as they still identify as male, but "something kinda male?"
# will be put in the "other" category as they are unsure of their status.
levels(mht$Gender)[c(1,2,3,4,10,20,21,22,38,39,40,42,43,44,45)] <- "Other"
levels(mht$Gender)[c(2,3,7:15,32:35)] <- "Female"
levels(mht$Gender)[-c(1,2)] <- "Male"
table(mht$Gender) # Much better
```

```
##
##  Other Female   Male
##     15    251    993
```

```r
# Time to fix the age variable.
t10 <- sort(mht$Age)
head(t10, 10) # These are some VERY young people...
```

```
##  [1] -1726   -29    -1     5     8    11    18    18    18    18
```

```r
tail(t10, 10) # We also got some aged 329 years old and 100 billion years old. Seems normal
```

```
##  [1] 5.70e+01 5.80e+01 6.00e+01 6.00e+01 6.10e+01 6.20e+01 6.50e+01 7.20e+01
##  [9] 3.29e+02 1.00e+11
```

```r
error_num <- c(t10[c(1:6, length(t10) - 1, length(t10))])
mht <- mht[-which(mht$Age %in% error_num), ]
summary(mht$Age)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   18.00   27.00   31.00   32.08   36.00   72.00
```

```r
# Now we look for duplicates
sum(duplicated(mht)) # None, luckily
```

```
## [1] 0
```

The Demographic

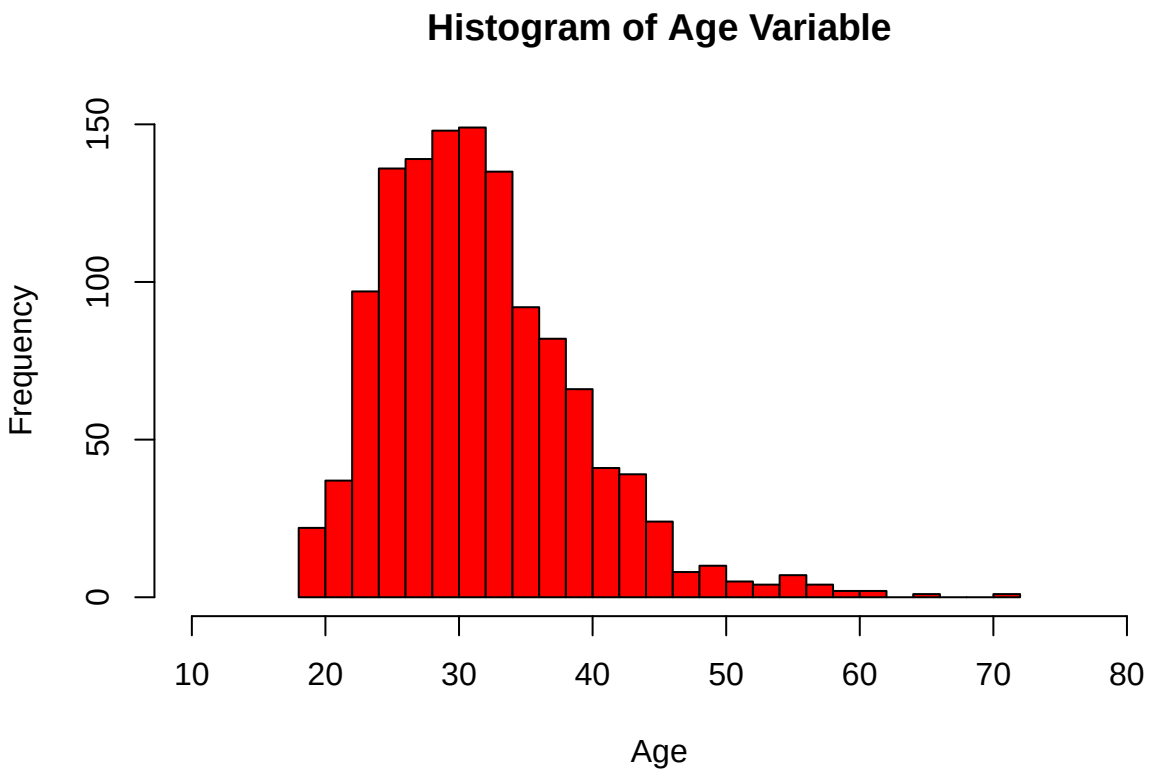
```r
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.6.3
```

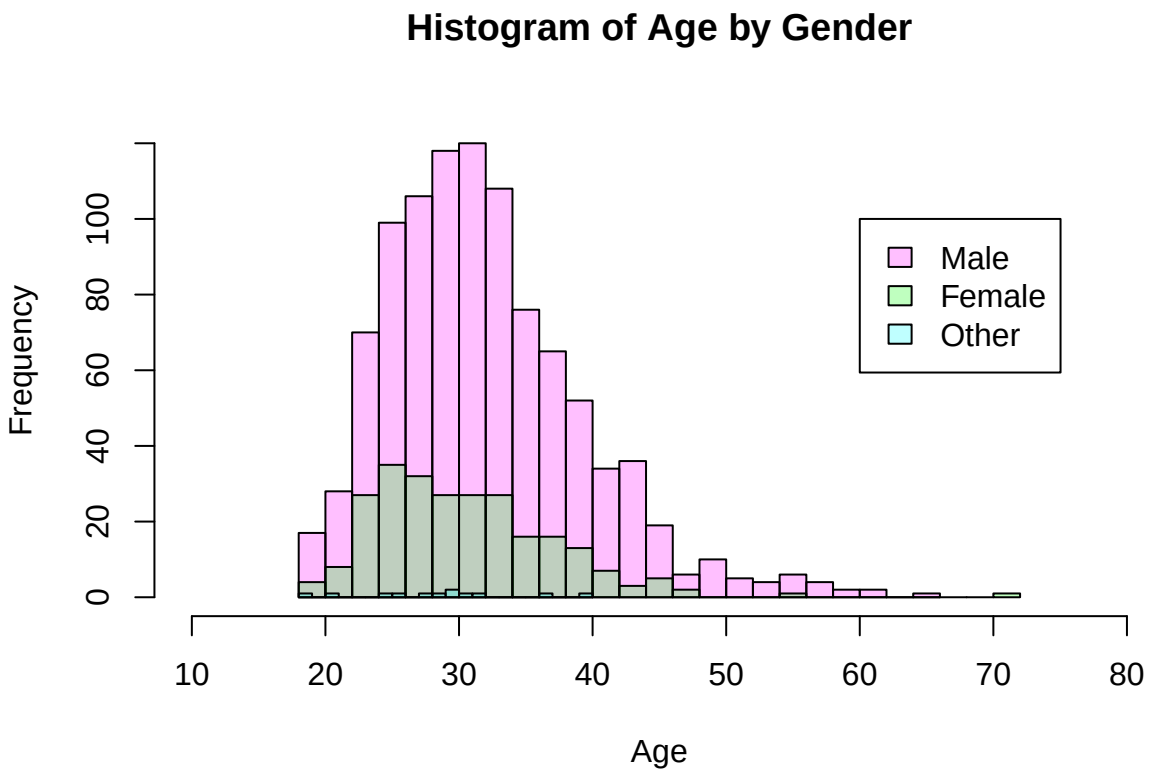
```r
library(dplyr)
```

```
## Warning: package 'dplyr' was built under R version 3.6.3
```

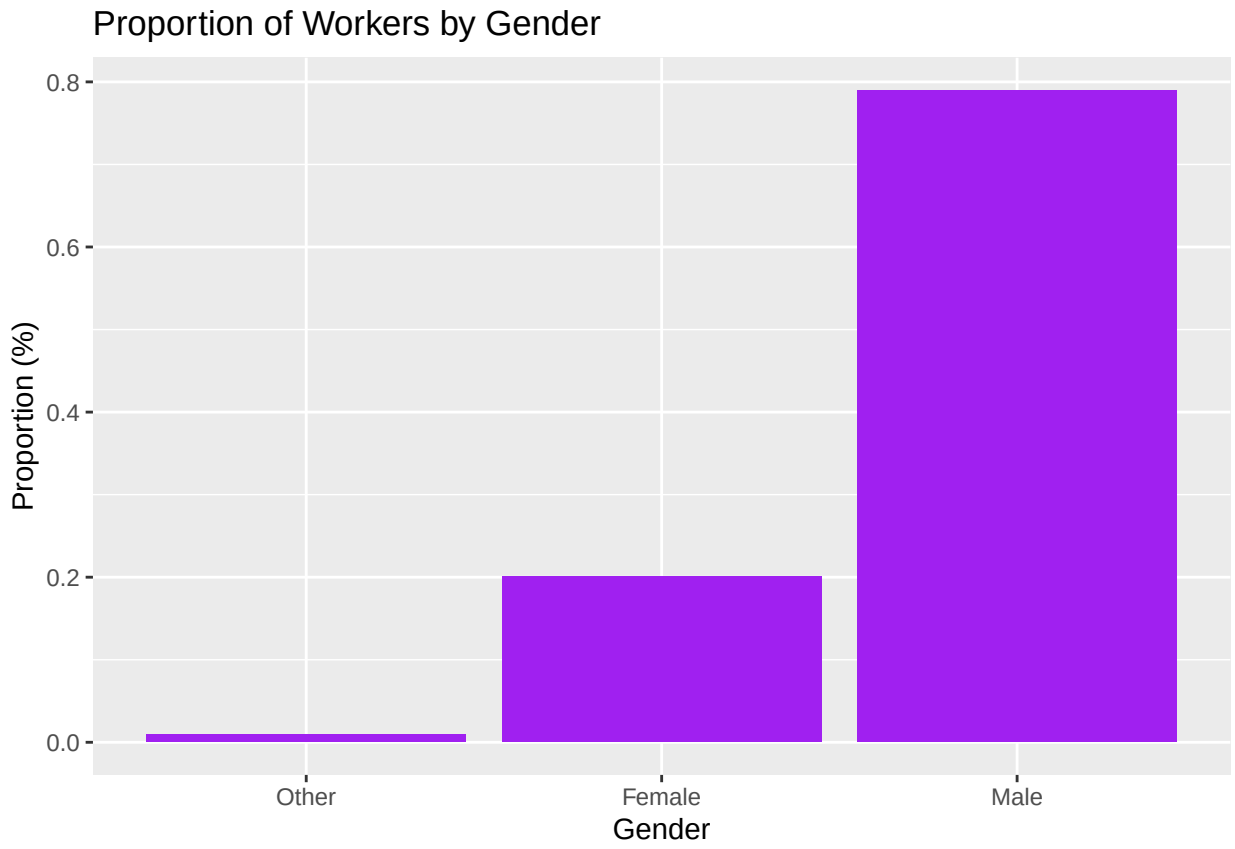
```r
# Age groups defined here:
# https://www.pewresearch.org/fact-tank/2019/01/17/where-millennials-end-and-generation-z-begins/
# Millenial -> 1981-1996
# Generation X -> 1965-1980
# Boomer & Silent Gen. -> 1928-1964

gen_groups <- cut(mht$Age, c(17,34,50,75),
                  labels=c("Millenial", "Generation X", "Boomer/Silent Gen"))
mht2 <- cbind(mht, "Gen" = gen_groups)

hist(mht2$Age, col = "Red", xlab = "Age", main = "Histogram of Age Variable",
     ylim = c(0, 150), xlim = c(10, 80), breaks = 20)
```




```
a_m <- mht2 %>%
  filter(Gender == "Male")
a_f <- mht2 %>%
  filter(Gender == "Female")
a_o <- mht2 %>%
  filter(Gender == "Other")

hist(a_m$Age, col = rgb(1,0,1,1/4), xlab = "Age", main = "Histogram of Age by Gender",
     ylim = c(0, 125), xlim = c(10, 80), breaks = 20)
hist(a_f$Age, col = rgb(0,1,0,1/4), breaks = 20, add = T)
hist(a_o$Age, col = rgb(0,1,1,1/4), breaks = 20, add = T)
legend(60, 100, legend = c("Male", "Female", "Other"),
       fill = c(rgb(1,0,1,1/4), rgb(0,1,0,1/4), rgb(0,1,1,1/4)))
```



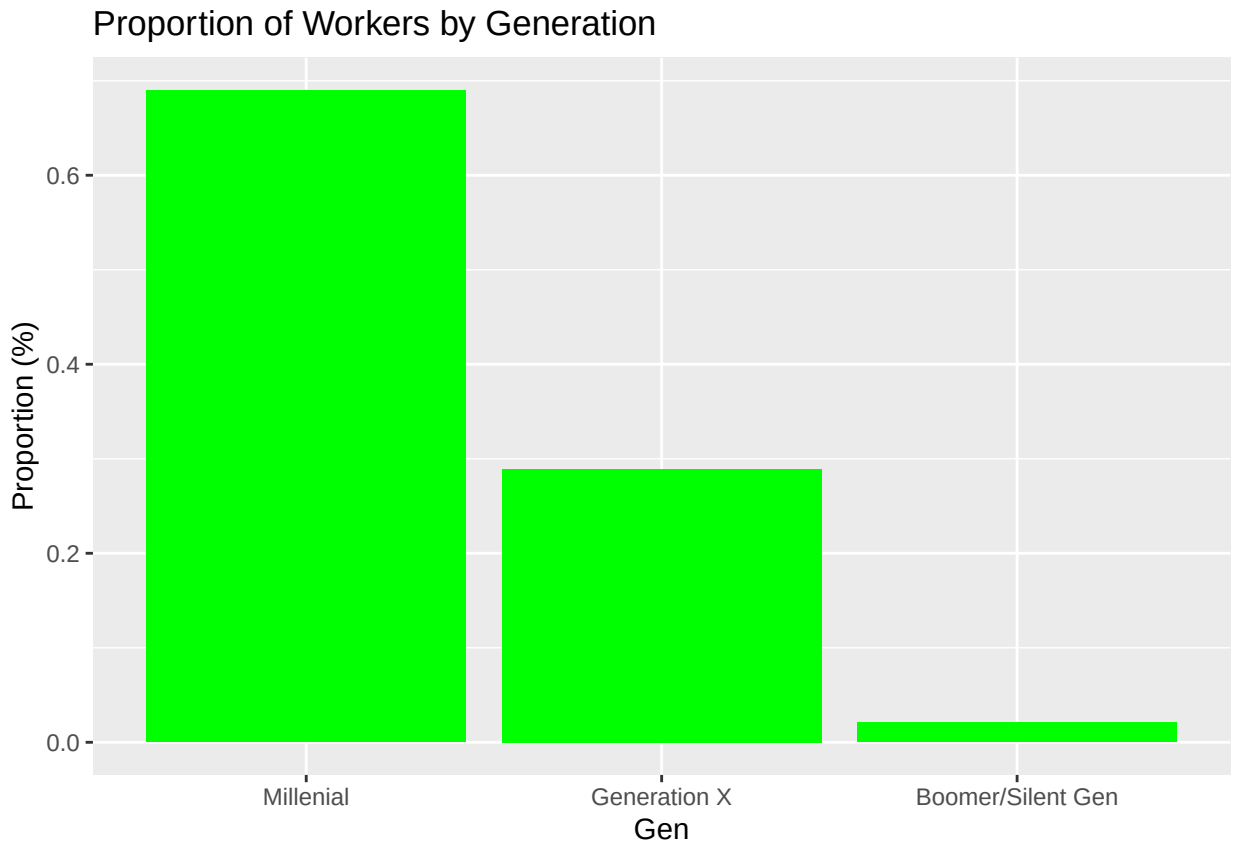

```
# We see that the majority of people that answered the survey
# are in their late 20's (25-29).

ggplot(mht2, aes(x = Gender)) +
  geom_bar(aes(y = (..count..)/sum(..count..)), fill = "purple") +
  ylab("Proportion (%)") +
  ggtitle("Proportion of Workers by Gender")
```




```
prop.table(table(mht2$Age))
```

```
##
##           18           19           20           21           22           23
## 0.0055955236 0.0071942446 0.0047961631 0.0127897682 0.0167865707 0.0407673861
##           24           25           26           27           28           29
## 0.0367705835 0.0487609912 0.0599520384 0.0567545963 0.0543565148 0.0679456435
##           30           31           32           33           34           35
## 0.0503597122 0.0535571543 0.0655475620 0.0559552358 0.0519584333 0.0439648281
##           36           37           38           39           40           41
## 0.0295763389 0.0343725020 0.0311750600 0.0263788969 0.0263788969 0.0167865707
##           42           43           44           45           46           47
## 0.0159872102 0.0223820943 0.0087929656 0.0095923261 0.0095923261 0.0015987210
##           48           49           50           51           53           54
## 0.0047961631 0.0031974420 0.0047961631 0.0039968026 0.0007993605 0.0023980815
##           55           56           57           58           60           61
## 0.0023980815 0.0031974420 0.0023980815 0.0007993605 0.0015987210 0.0007993605
##           62           65           72
## 0.0007993605 0.0007993605 0.0007993605
```

```
ggplot(mht2, aes(x = Gen)) +
  geom_bar(aes(y = (..count..)/sum(..count..)), fill = "green") +
  ylab("Proportion (%)") +
  ggtitle("Proportion of Workers by Generation")
```




```
prop.table(table(mht2$Gen))
```

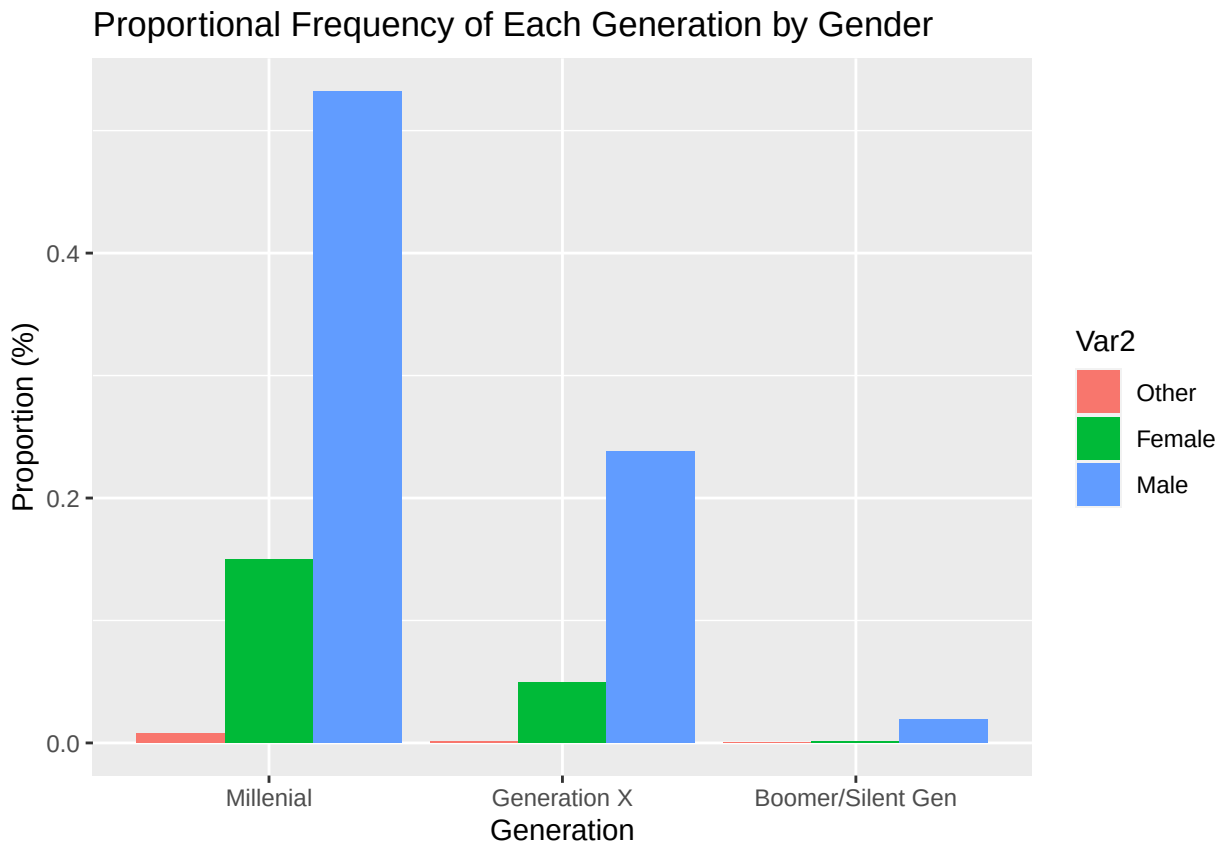
```
##
##         Millenial      Generation X Boomer/Silent Gen
##        0.68984812        0.28936851        0.02078337
```

```
p <- prop.table(table(mht2$Gen, mht2$Gender))
p_df <- as.data.frame(p)
p_df
```

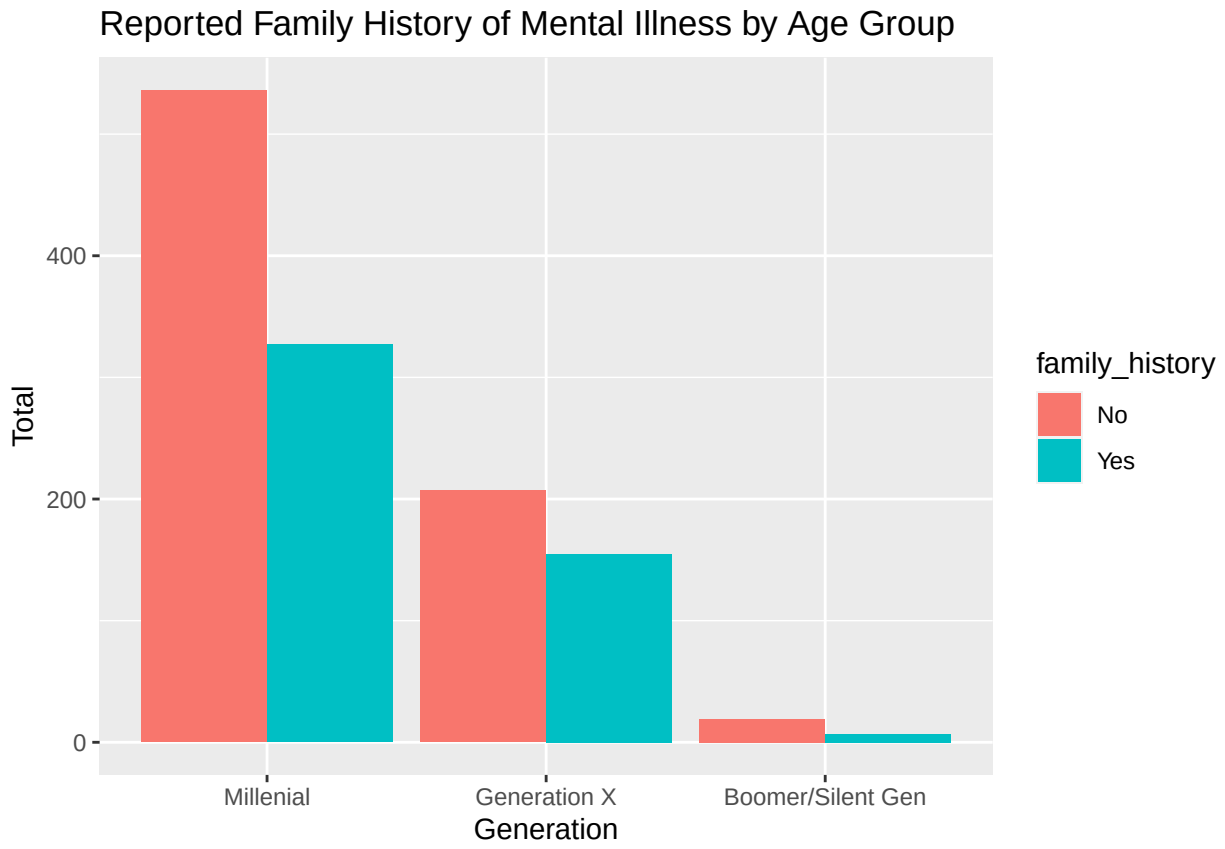
```
##                Var1   Var2        Freq
## 1         Millenial  Other 0.007993605
## 2      Generation X  Other 0.001598721
## 3 Boomer/Silent Gen  Other 0.000000000
## 4         Millenial Female 0.149480416
## 5      Generation X Female 0.049560352
## 6 Boomer/Silent Gen Female 0.001598721
## 7         Millenial   Male 0.532374101
## 8      Generation X   Male 0.238209432
## 9 Boomer/Silent Gen   Male 0.019184652
```

```
ggplot(p_df, aes(x = Var1, y = Freq, fill = Var2)) +
  geom_bar(stat="identity", position = "dodge") +
  xlab("Generation") +
```

```
  ylab("Proportion (%)") +
  ggtitle("Proportional Frequency of Each Generation by Gender")
```




```
fh_gen <- mht2 %>%
  select(Gen, family_history) %>%
  group_by(Gen) %>%
  count(family_history)

ggplot(fh_gen, aes(x = Gen, y = n, fill = family_history)) +
  geom_bar(stat="identity", position = "dodge") +
  xlab("Generation") +
  ylab("Total") +
  ggtitle("Reported Family History of Mental Illness by Age Group")
```




```
table(mht2$Country) # Countries that people from this survey are in
```

```
##
##              Australia                Austria         Bahamas, The
##                     21                      3                    0
##                Belgium Bosnia and Herzegovina               Brazil
##                      6                      1                    6
##               Bulgaria                 Canada                China
##                      4                     72                    1
##               Colombia             Costa Rica              Croatia
##                      2                      1                    2
##         Czech Republic                Denmark              Finland
##                      1                      2                    3
##                 France                Georgia              Germany
##                     13                      1                   45
##                 Greece                Hungary                India
##                      2                      1                   10
##                Ireland                 Israel                Italy
##                     27                      5                    7
##                  Japan                 Latvia               Mexico
##                      1                      1                    3
##                Moldova            Netherlands          New Zealand
##                      1                     27                    8
##                Nigeria                 Norway          Philippines
##                      1                      1                    1
```

```
##                 Poland               Portugal                Romania
##                      7                      2                      1
##                 Russia              Singapore               Slovenia
##                      3                      4                      1
##           South Africa                  Spain                 Sweden
##                      6                      1                      7
##            Switzerland               Thailand         United Kingdom
##                      7                      1                    184
##          United States                Uruguay               Zimbabwe
##                    746                      1                      0
```

```
table(mht2$state) # US states that people from this survey are in
```

```
##
##  AL  AZ  CA  CO  CT  DC  FL  GA  IA  ID  IL  IN  KS  KY  LA  MA  MD  ME  MI  MN
##   7   7 138   9   4   4  15  12   4   1  28  27   3   5   1  20   8   1  22  20
##  MO  MS  NC  NE  NH  NJ  NM  NV  NY  OH  OK  OR  PA  RI  SC  SD  TN  TX  UT  VA
##  12   1  14   2   3   6   2   3  57  27   6  29  29   1   5   3  45  44  11  14
##  VT  WA  WI  WV  WY
##   3  70  12   1   2
```

Survey Results

MH = Mental Health

PH = Physical Health

```
library(ggpubr)
```

```
## Warning: package 'ggpubr' was built under R version 3.6.3
```

```
summary(mht2[, -c(1:5, 27)])
```

```
##  self_employed family_history treatment   work_interfere         no_employees
##  No  :1091     No :762        No :619   Never    :212    1-5           :158
##  Yes : 142     Yes:489        Yes:632   Often    :140    100-500       :175
##  NA's:  18                              Rarely   :173    26-100        :288
##                                         Sometimes:464    500-1000      : 60
##                                         NA's     :262    6-25          :289
##                                                          More than 1000:281
##  remote_work tech_company       benefits     care_options   wellness_program
##  No :880     No : 226     Don't know:407   No      :499   Don't know:187
##  Yes:371     Yes:1025     No        :371   Not sure:313   No        :837
##                           Yes       :473   Yes     :439   Yes       :227
##
##
##
##       seek_help        anonymity                  leave
##  Don't know:363   Don't know:815   Don't know        :561
```

```
##  No        :641   No        : 64   Somewhat difficult:125
##  Yes       :247   Yes       :372   Somewhat easy     :265
##                                     Very difficult    : 97
##                                     Very easy         :203
##
##  mental_health_consequence phys_health_consequence        coworkers
##  Maybe:476                 Maybe:273                No          :258
##  No   :487                 No   :920                Some of them:771
##  Yes  :288                 Yes  : 58                Yes         :222
##
##
##
##         supervisor  mental_health_interview phys_health_interview
##  No          :390   Maybe: 207              Maybe:555
##  Some of them:349   No   :1003              No   :496
##  Yes         :512   Yes  :  41              Yes  :200
##
##
##
##   mental_vs_physical obs_consequence                Gen
##  Don't know:574      No :1070        Millenial        :863
##  No        :338      Yes: 181        Generation X     :362
##  Yes       :339                      Boomer/Silent Gen: 26
##
##
##
```

```
se <- mht2 %>%
  count(self_employed)

g1 <- ggplot(se[-3, ], aes(x = self_employed, y = n)) +
  geom_bar(stat="identity", position = "dodge", fill = c("red", "green")) +
  xlab("Answer") +
  ylab("Total") +
  ggtitle("Reported as Self-Employed") +
  theme(plot.title = element_text(size = 10))

fh <- mht2 %>%
  count(family_history)

g2 <- ggplot(fh, aes(x = family_history, y = n)) +
  geom_bar(stat="identity", position = "dodge", fill = c("red", "green")) +
  xlab("Answer") +
  ylab("Total") +
  ggtitle("Family History of Mental Illness?") +
  theme(plot.title = element_text(size = 10))

tr <- mht2 %>%
  count(treatment)

g3 <- ggplot(tr, aes(x = treatment, y = n)) +
  geom_bar(stat="identity", position = "dodge", fill = c("red", "green")) +
  xlab("Answer") +
  ylab("Total") +
```

```
  ggtitle("Sought Treatment for Mental Health?") +
  theme(plot.title = element_text(size = 10))

wi <- mht2 %>%
  count(work_interfere)
wi2 <- wi[-5, ]
wi2 <- wi2[c(1, 3, 4, 2), ]
wi2$work_interfere <- factor(wi2$work_interfere, levels = wi2$work_interfere)

g4 <- ggplot(wi2, aes(x = work_interfere, y = n)) +
  geom_bar(stat="identity", position = "dodge", fill = 1:4) +
  xlab("Answer") +
  ylab("Total") +
  ggtitle("Mental Health Interferes w/ Work?") +
  theme(plot.title = element_text(size = 10))

ne <- mht2 %>%
  count(no_employees)
ne2 <- ne[c(1, 5, 3, 2, 4, 6), ]
ne2$no_employees <- factor(ne2$no_employees, levels = ne2$no_employees)

g5 <- ggplot(ne2, aes(x = no_employees, y = n)) +
  geom_bar(stat="identity", position = "dodge", fill = 1:6) +
  xlab("Answer") +
  ylab("Total") +
  ggtitle("How Many Employees in the Company?") +
  theme(plot.title = element_text(size = 20))

rw <- mht2 %>%
  count(remote_work)

g6 <- ggplot(rw, aes(x = remote_work, y = n)) +
  geom_bar(stat="identity", position = "dodge", fill = c("red", "green")) +
  xlab("Answer") +
  ylab("Total") +
  ggtitle("Work Remotely?") +
  theme(plot.title = element_text(size = 10))

tc <- mht2 %>%
  count(tech_company)

g7 <- ggplot(tc, aes(x = tech_company, y = n)) +
  geom_bar(stat="identity", position = "dodge", fill = c("red", "green")) +
  xlab("Answer") +
  ylab("Total") +
  ggtitle("Work for a Tech Company?") +
  theme(plot.title = element_text(size = 10))

bn <- mht2 %>%
  count(benefits)

g8 <- ggplot(bn, aes(x = benefits, y = n)) +
  geom_bar(stat="identity", position = "dodge", fill = c("purple", "red", "green")) +
```

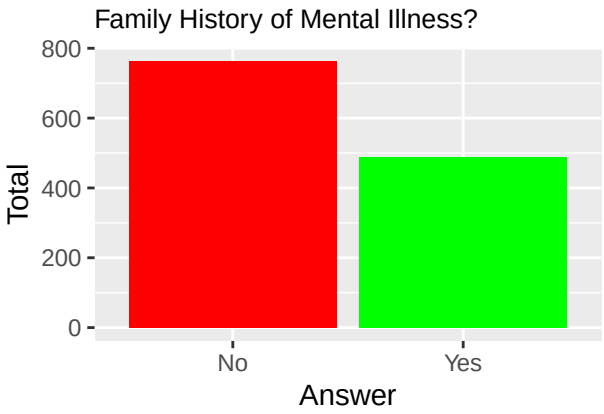
```
  xlab("Answer") +
  ylab("Total") +
  ggtitle("Employer Provide MH Benefits?") +
  theme(plot.title = element_text(size = 10))

co <- mht2 %>%
  count(care_options)

g9 <- ggplot(co, aes(x = care_options, y = n)) +
  geom_bar(stat="identity", position = "dodge", fill = c("red", "purple", "green")) +
  xlab("Answer") +
  ylab("Total") +
  ggtitle("Knowledge of MH Care Employer Have") +
  theme(plot.title = element_text(size = 10))

wp <- mht2 %>%
  count(wellness_program)

g10 <- ggplot(wp, aes(x = wellness_program, y = n)) +
  geom_bar(stat="identity", position = "dodge", fill = c("purple", "red", "green")) +
  xlab("Answer") +
  ylab("Total") +
  ggtitle("MH Part of Employee Wellness Program?") +
  theme(plot.title = element_text(size = 10))

sh <- mht2 %>%
  count(seek_help)

g11 <- ggplot(sh, aes(x = seek_help, y = n)) +
  geom_bar(stat="identity", position = "dodge", fill = c("purple", "red", "green")) +
  xlab("Answer") +
  ylab("Total") +
  ggtitle("Resources Provided for MH Issues?") +
  theme(plot.title = element_text(size = 10))

an <- mht2 %>%
  count(anonymity)

g12 <- ggplot(an, aes(x = anonymity, y = n)) +
  geom_bar(stat="identity", position = "dodge", fill = c("purple", "red", "green")) +
  xlab("Answer") +
  ylab("Total") +
  ggtitle("Anonymity Kept for MH Issues?") +
  theme(plot.title = element_text(size = 10))

lv <- mht2 %>%
  count(leave)
lv2 <- lv[c(4, 2, 1, 3, 5), ]
lv2$leave <- factor(lv2$leave, levels = lv2$leave)

g13 <- ggplot(lv2, aes(x = leave, y = n)) +
  geom_bar(stat="identity", position = "dodge", fill = 1:5) +
  xlab("Answer") +
```

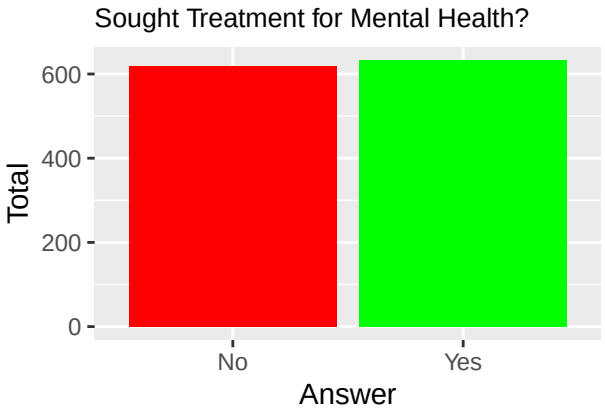
```
  ylab("Total") +
  ggtitle("Ease of Taking Leave for MH Issues?") +
  theme(plot.title = element_text(size = 20))

mc <- mht2 %>%
  count(mental_health_consequence)

g14 <- ggplot(mc, aes(x = mental_health_consequence, y = n)) +
  geom_bar(stat="identity", position = "dodge", fill = c("purple", "red", "green")) +
  xlab("Answer") +
  ylab("Total") +
  ggtitle("Discussing MH Issues have Consequences?") +
  theme(plot.title = element_text(size = 10))

pc <- mht2 %>%
  count(phys_health_consequence)

g15 <- ggplot(pc, aes(x = phys_health_consequence, y = n)) +
  geom_bar(stat="identity", position = "dodge", fill = c("purple", "red", "green")) +
  xlab("Answer") +
  ylab("Total") +
  ggtitle("Discussing PH Issues have Consequences?") +
  theme(plot.title = element_text(size = 10))

cw <- mht2 %>%
  count(coworkers)

g16 <- ggplot(cw, aes(x = coworkers, y = n)) +
  geom_bar(stat="identity", position = "dodge", fill = c("red", "purple", "green")) +
  xlab("Answer") +
  ylab("Total") +
  ggtitle("Willing to Discuss MH Issues w/ Coworkers?") +
  theme(plot.title = element_text(size = 10))

su <- mht2 %>%
  count(supervisor)

g17 <- ggplot(su, aes(x = supervisor, y = n)) +
  geom_bar(stat="identity", position = "dodge", fill = c("red", "purple", "green")) +
  xlab("Answer") +
  ylab("Total") +
  ggtitle("Willing to Discuss MH Issues w/ Supervisor(s)?") +
  theme(plot.title = element_text(size = 9))

mi <- mht2 %>%
  count(mental_health_interview)

g18 <- ggplot(mi, aes(x = mental_health_interview, y = n)) +
  geom_bar(stat="identity", position = "dodge", fill = c("purple", "red", "green")) +
  xlab("Answer") +
  ylab("Total") +
  ggtitle("Willing to Bring Up MH Issues in an Interview?") +
  theme(plot.title = element_text(size = 9))
```

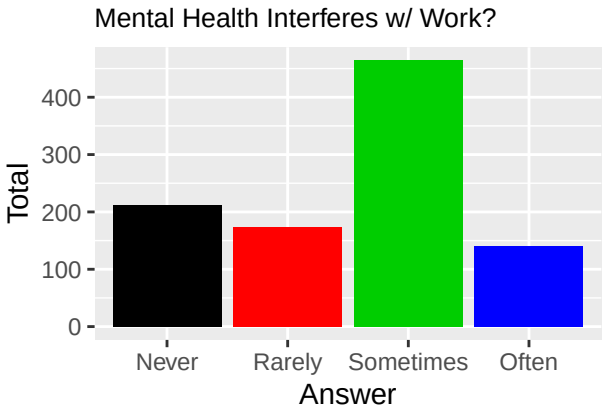
```r
pi <- mht2 %>%
  count(phys_health_interview)

g19 <- ggplot(pi, aes(x = phys_health_interview, y = n)) +
  geom_bar(stat="identity", position = "dodge", fill = c("purple", "red", "green")) +
  xlab("Answer") +
  ylab("Total") +
  ggtitle("Willing to Bring Up PH Issues in an Interview?") +
  theme(plot.title = element_text(size = 10))

mp <- mht2 %>%
  count(mental_vs_physical)

g20 <- ggplot(mp, aes(x = mental_vs_physical, y = n)) +
  geom_bar(stat="identity", position = "dodge", fill = c("purple", "red", "green")) +
  xlab("Answer") +
  ylab("Total") +
  ggtitle("Feel that Employer Takes MH as Seriously as PH?") +
  theme(plot.title = element_text(size = 8.5))

oc <- mht2 %>%
  count(obs_consequence)

g21 <- ggplot(oc, aes(x = obs_consequence, y = n)) +
  geom_bar(stat="identity", position = "dodge", fill = c("red", "green")) +
  xlab("Answer") +
  ylab("Total") +
  ggtitle("Heard/Observed Consequences for Coworkers w/ MH Issues?") +
  theme(plot.title = element_text(size = 10))

g_all <- c(g1, g2, g3, g4, g5, g6, g7,
           g8, g9, g10, g11, g12, g13, g14,
           g15, g16, g17, g18, g19, g20, g21)
ggarrange(g1, g2, g3, g4, g6, g7,
           g8, g9, g10, g11, g12, g14, g15,
            g16, g17, g18, g19, g20, g21, ncol = 2, nrow = 2)
```

```
## $`1`
```

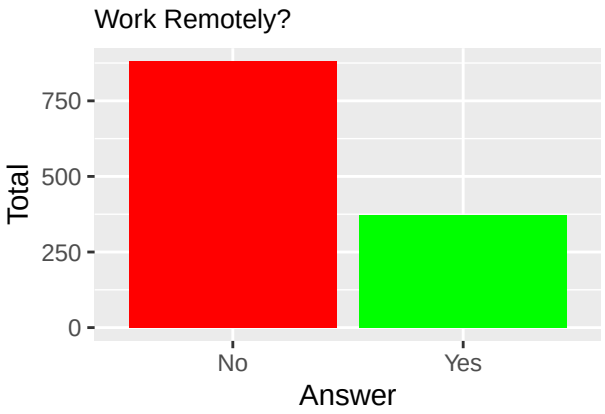

Reported as Self−Employed
Total
900
600
300
0
No
Yes
Answer

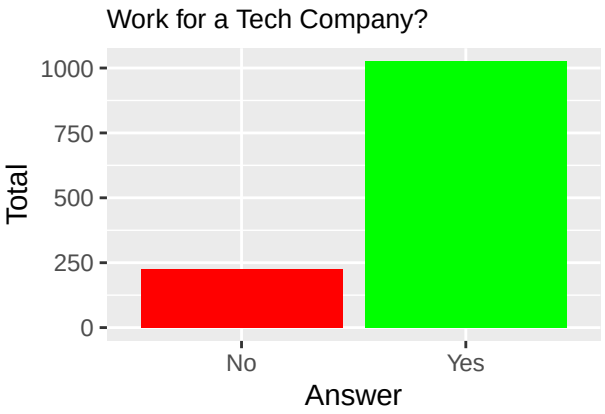

Family History of Mental Illness?
Total
800
600
400
200
0
No
Yes
Answer

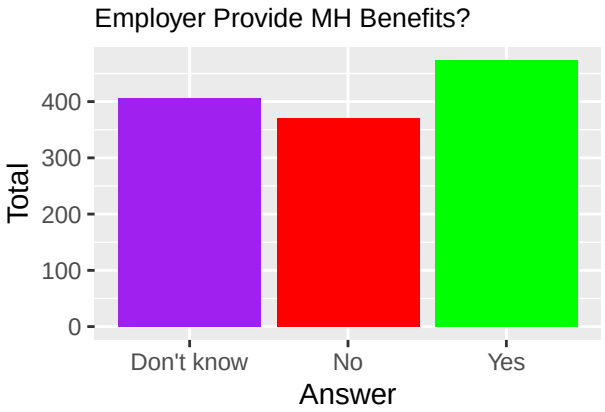

Sought Treatment for Mental Health?
Total
600
400
200
0
No
Yes
Answer

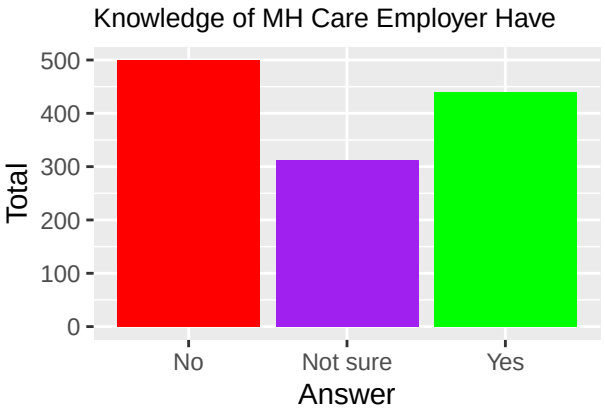

Mental Health Interferes w/ Work?
Total
400
300
200
100
0
Never
Rarely
Sometimes
Often
Answer

```
##
## $`2`
```


Work Remotely?
Total
750
500
250
0
No
Yes
Answer

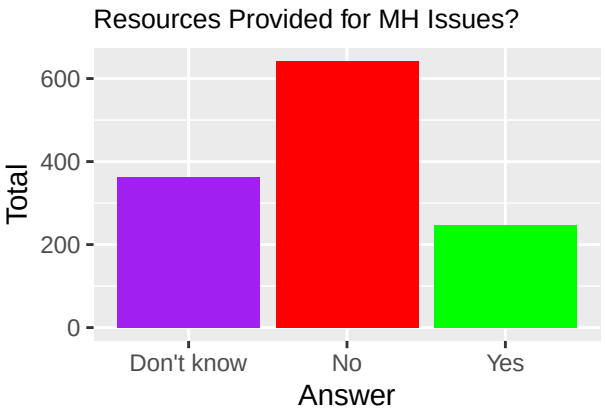

Work for a Tech Company?
Total
1000
750
500
250
0
No
Yes
Answer

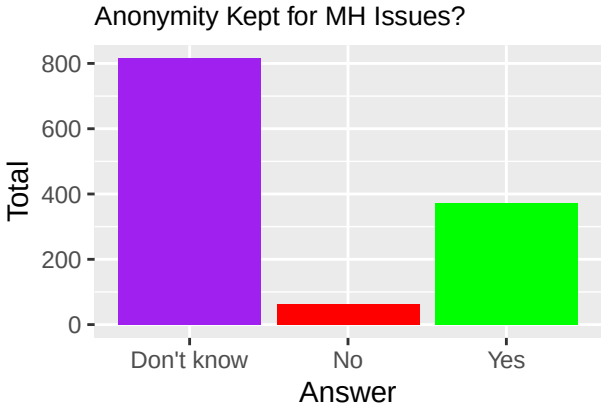

Employer Provide MH Benefits?
Total
400
300
200
100
0
Don't know
No
Yes
Answer

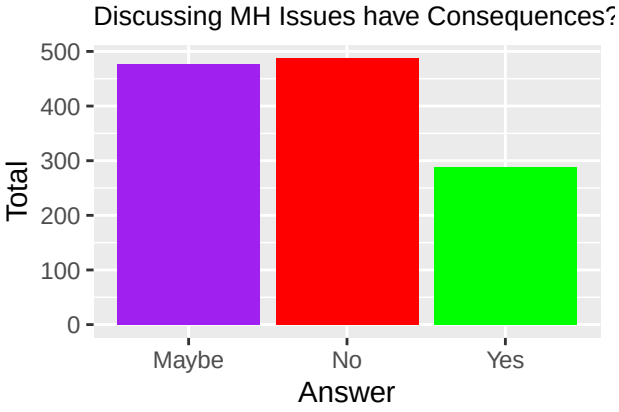

Knowledge of MH Care Employer Have
Total
500
400
300
200
100
0
No
Not sure
Yes
Answer

```
##
## $`3`
```

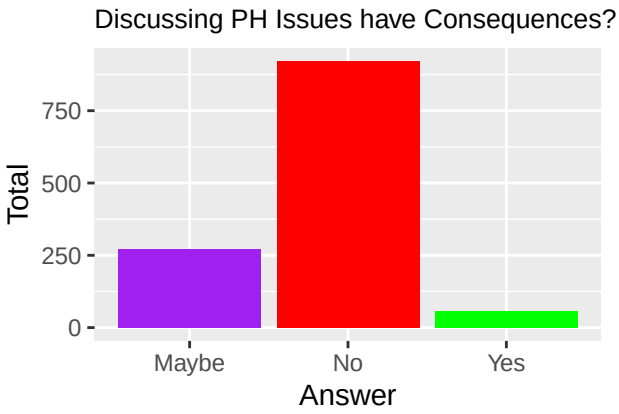

MH Part of Employee Wellness Program?
Total
800
600
400
200
0
Don't know
No
Yes
Answer

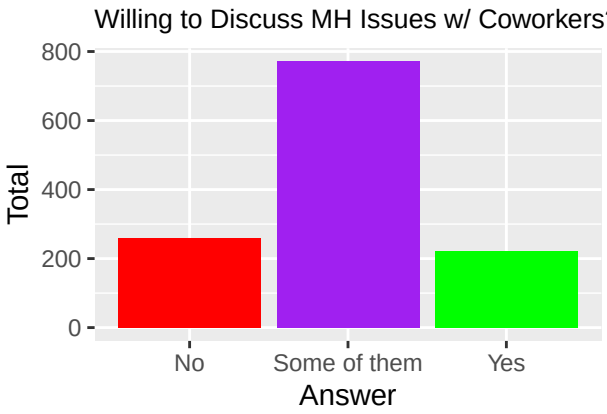

Resources Provided for MH Issues?
Total
600
400
200
0
Don't know
No
Yes
Answer

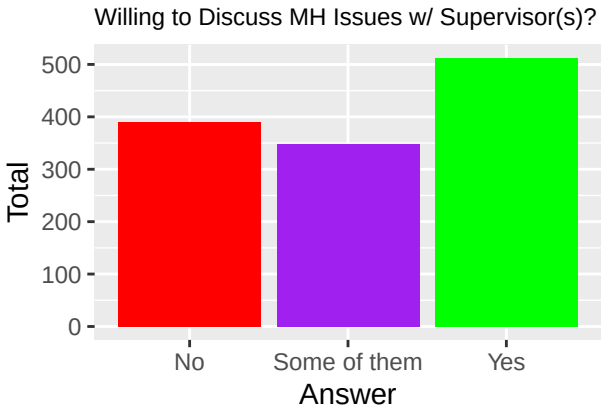

Anonymity Kept for MH Issues?
Total
800
600
400
200
0
Don't know
No
Yes
Answer

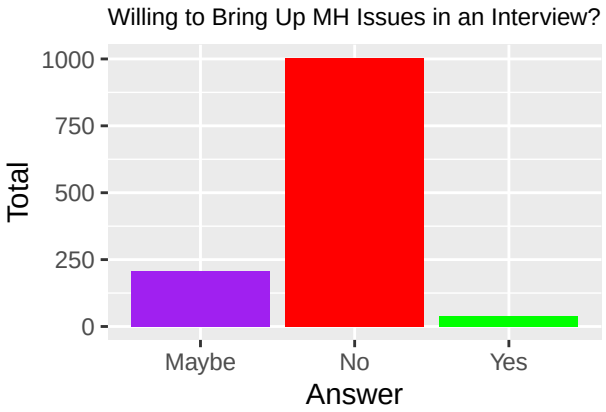

Discussing MH Issues have Consequences?
Total
500
400
300
200
100
0
Maybe
No
Yes
Answer

```
##
## $`4`
```

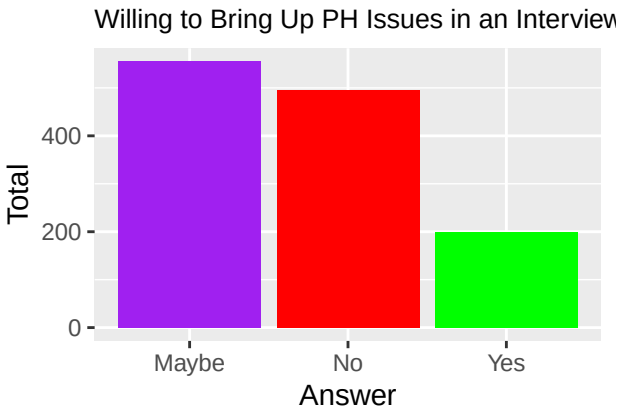

Discussing PH Issues have Consequences?
Total
750
500
250
0
Maybe
No
Yes
Answer

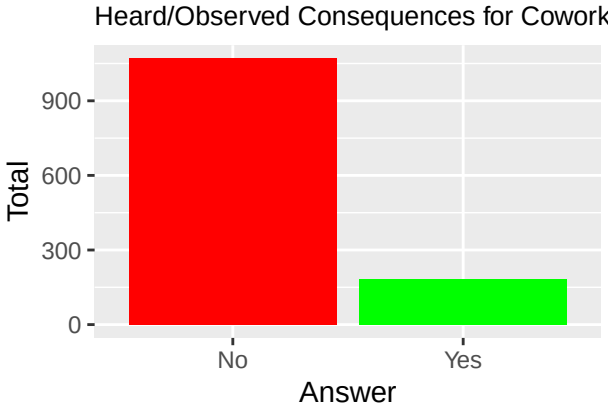

Willing to Discuss MH Issues w/ Coworkers?
Total
800
600
400
200
0
No
Some of them
Yes
Answer


Willing to Discuss MH Issues w/ Supervisor(s)?
Total
500
400
300
200
100
0
No
Some of them
Yes
Answer


Willing to Bring Up MH Issues in an Interview?
Total
1000
750
500
250
0
Maybe
No
Yes
Answer

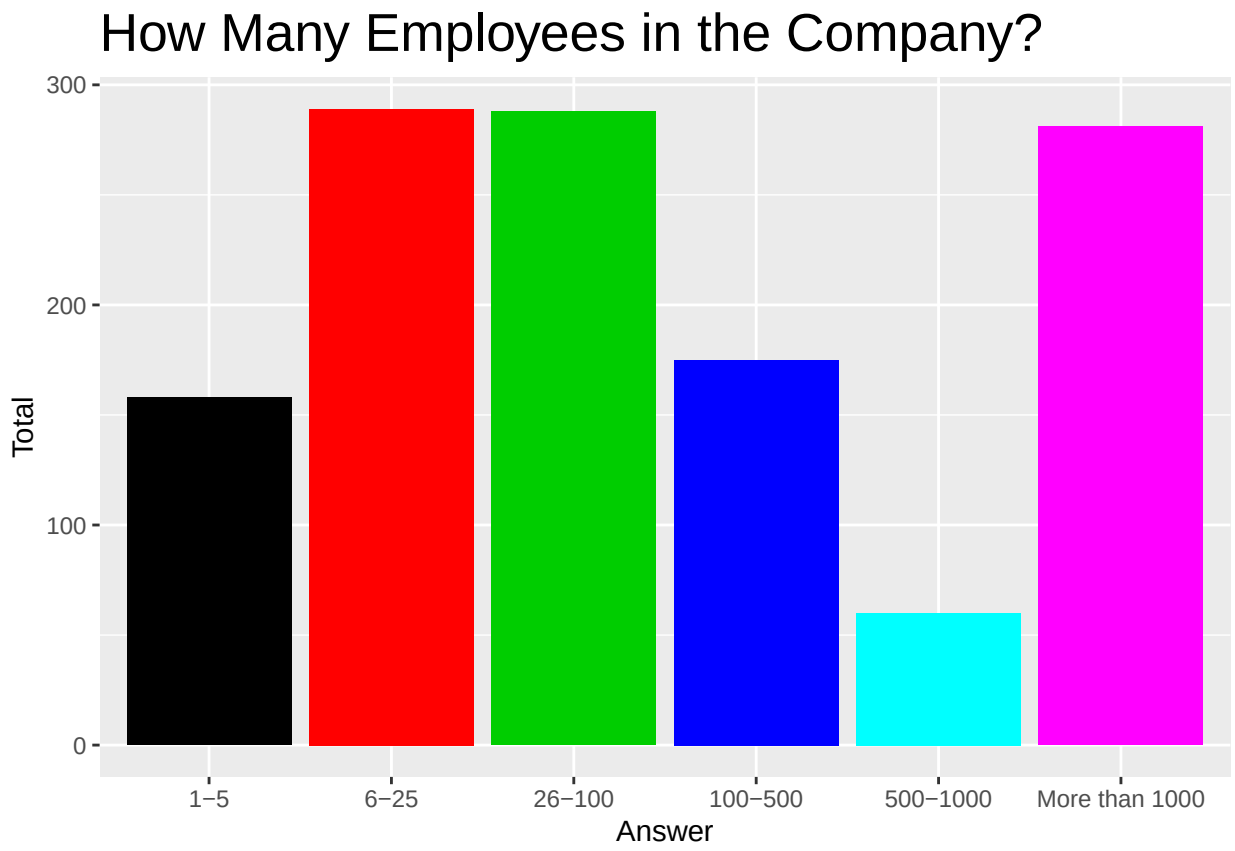
```
##
## $`5`
```



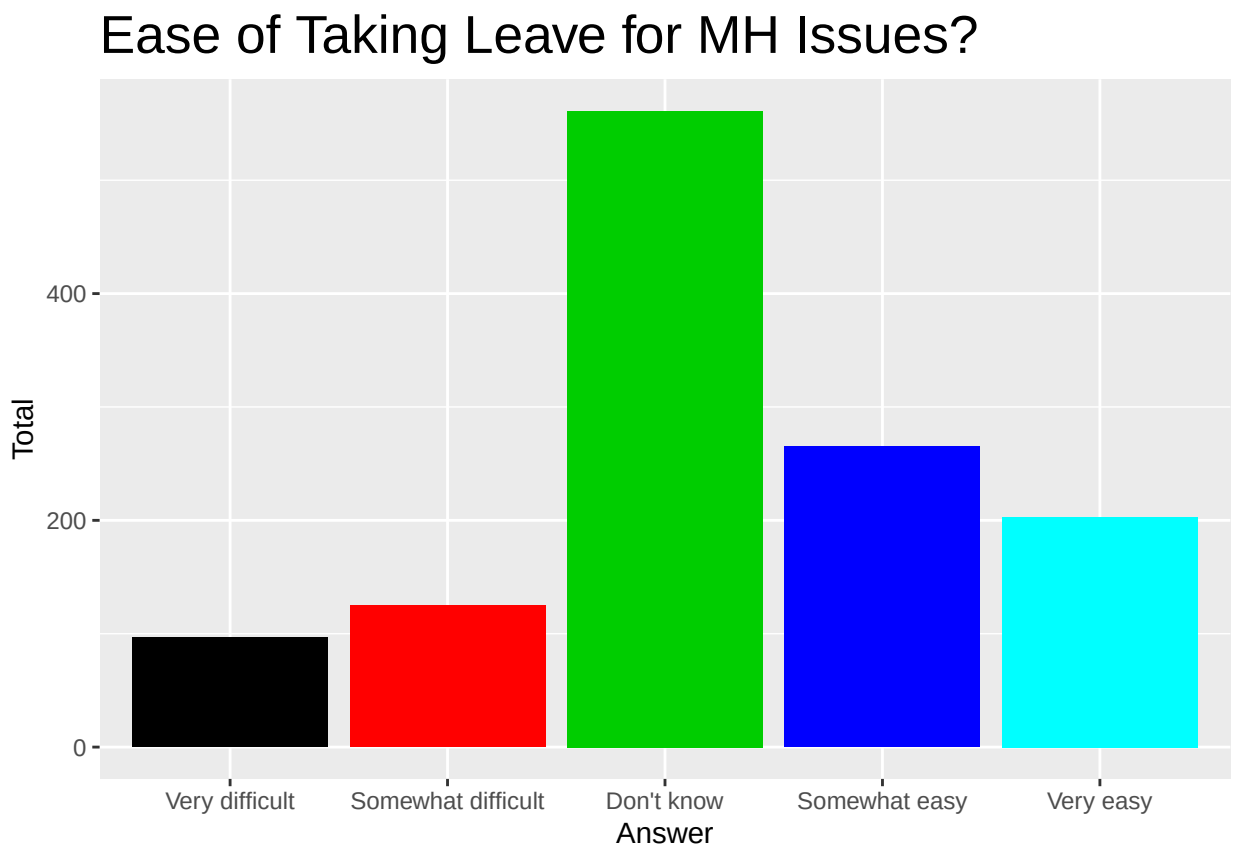



```
##
## attr(,"class")
## [1] "list"      "ggarrange"
```

```
g5
```


How Many Employees in the Company?
300
200
100
0
Total
1−5
6−25
26−100
100−500
500−1000
More than 1000
Answer

```
g13
```



Does Gender Make a Difference in the Survey Answers?

Since the "other" category of the Gender variable only has a sample size of 15, we won't include it as the amount is too small to analyze.

```
library(rcompanion)
```

```
## Warning: package 'rcompanion' was built under R version 3.6.3
```

```
f1 <- table("Gender" = mht2$Gender, "self_employed" = mht2$self_employed)

f2 <- table("Gender" = mht2$Gender, "family_history" = mht2$family_history)

f3 <- table("Gender" = mht2$Gender, "treatment" = mht2$treatment)

f4 <- table("Gender" = mht2$Gender, "work_interfere" = mht2$work_interfere)

f5 <- table("Gender" = mht2$Gender, "no_employees" = mht2$no_employees)

f6 <- table("Gender" = mht2$Gender, "remote_work" = mht2$remote_work)
```

```r
f7 <- table("Gender" = mht2$Gender, "tech_company" = mht2$tech_company)

f8 <- table("Gender" = mht2$Gender, "benefits" = mht2$benefits)

f9 <- table("Gender" = mht2$Gender, "care_options" = mht2$care_options)

f10 <- table("Gender" = mht2$Gender, "wellness_program" = mht2$wellness_program)

f11 <- table("Gender" = mht2$Gender, "seek_help" = mht2$seek_help)

f12 <- table("Gender" = mht2$Gender, "anonymity" = mht2$anonymity)

f13 <- table("Gender" = mht2$Gender, "leave" = mht2$leave)

f14 <- table("Gender" = mht2$Gender, "mh_consequence" = mht2$mental_health_consequence)

f15 <- table("Gender" = mht2$Gender, "ph_consequence" = mht2$phys_health_consequence)

f16 <- table("Gender" = mht2$Gender, "coworkers" = mht2$coworkers)

f17 <- table("Gender" = mht2$Gender, "supervisor" = mht2$supervisor)

f18 <- table("Gender" = mht2$Gender, "mh_interview" = mht2$mental_health_interview)

f19 <- table("Gender" = mht2$Gender, "ph_interview" = mht2$self_employed)

f20 <- table("Gender" = mht2$Gender, "mental_vs_physical" = mht2$mental_vs_physical)

f21 <- table("Gender" = mht2$Gender, "obs_consequence" = mht2$obs_consequence)
```

We'll first use Fisher's Test to see if there are some variables where there are differences in the survey results in respect to the survey takers' gender.

```r
pairwiseNominalIndependence(f1, compare = "row", fisher = T,
gtest = F, chisq = F, digits = 3)[3, ]
```

```
##      Comparison p.Fisher p.adj.Fisher
## 3 Female : Male     0.18        0.464
```

```r
pairwiseNominalIndependence(f2, compare = "row", fisher = T,
gtest = F, chisq = F, digits = 3)[3, ]
```

```
##      Comparison p.Fisher p.adj.Fisher
## 3 Female : Male 3.48e-07     1.04e-06
```

```
pairwiseNominalIndependence(f3, compare = "row", fisher = T,
gtest = F, chisq = F, digits = 3)[3, ]
```

```
##      Comparison p.Fisher p.adj.Fisher
## 3 Female : Male 2.19e-11     6.57e-11
```

```
pairwiseNominalIndependence(f4, compare = "row", fisher = T,
gtest = F, chisq = F, digits = 3)[3, ]
```

```
##      Comparison p.Fisher p.adj.Fisher
## 3 Female : Male 0.000389      0.00117
```

```
pairwiseNominalIndependence(f5, compare = "row", fisher = T,
gtest = F, chisq = F, digits = 3, simulate.p.value = T)[3, ]
```

```
##      Comparison p.Fisher p.adj.Fisher
## 3 Female : Male    0.001        0.003
```

```
pairwiseNominalIndependence(f6, compare = "row", fisher = T,
gtest = F, chisq = F, digits = 3)[3, ]
```

```
##      Comparison p.Fisher p.adj.Fisher
## 3 Female : Male    0.757            1
```

```
pairwiseNominalIndependence(f7, compare = "row", fisher = T,
gtest = F, chisq = F, digits = 3)[3, ]
```

```
##      Comparison p.Fisher p.adj.Fisher
## 3 Female : Male  0.00572       0.0172
```

```
pairwiseNominalIndependence(f8, compare = "row", fisher = T,
gtest = F, chisq = F, digits = 3)[3, ]
```

```
##      Comparison p.Fisher p.adj.Fisher
## 3 Female : Male 5.51e-07     1.65e-06
```

```
pairwiseNominalIndependence(f9, compare = "row", fisher = T,
gtest = F, chisq = F, digits = 3)[3, ]
```

```
##      Comparison p.Fisher p.adj.Fisher
## 3 Female : Male 5.97e-05     0.000179
```

```
pairwiseNominalIndependence(f10, compare = "row", fisher = T,
gtest = F, chisq = F, digits = 3)[3, ]
```

```
##      Comparison p.Fisher p.adj.Fisher
## 3 Female : Male   0.0879        0.264
```

```
pairwiseNominalIndependence(f11, compare = "row", fisher = T,
gtest = F, chisq = F, digits = 3)[3, ]
```

```
##      Comparison p.Fisher p.adj.Fisher
## 3 Female : Male    0.239        0.638
```

```
pairwiseNominalIndependence(f12, compare = "row", fisher = T,
gtest = F, chisq = F, digits = 3)[3, ]
```

```
##      Comparison p.Fisher p.adj.Fisher
## 3 Female : Male    0.392        0.791
```

```
pairwiseNominalIndependence(f13, compare = "row", fisher = T,
gtest = F, chisq = F, digits = 3)[3, ]
```

```
##      Comparison p.Fisher p.adj.Fisher
## 3 Female : Male    0.232        0.524
```

```
pairwiseNominalIndependence(f14, compare = "row", fisher = T,
gtest = F, chisq = F, digits = 3)[3, ]
```

```
##      Comparison p.Fisher p.adj.Fisher
## 3 Female : Male  0.00121      0.00363
```

```
pairwiseNominalIndependence(f15, compare = "row", fisher = T,
gtest = F, chisq = F, digits = 3)[3, ]
```

```
##      Comparison p.Fisher p.adj.Fisher
## 3 Female : Male   0.0101       0.0303
```

```
pairwiseNominalIndependence(f16, compare = "row", fisher = T,
gtest = F, chisq = F, digits = 3)[3, ]
```

```
##      Comparison p.Fisher p.adj.Fisher
## 3 Female : Male    0.149        0.447
```

```
pairwiseNominalIndependence(f17, compare = "row", fisher = T,
gtest = F, chisq = F, digits = 3)[3, ]
```

```
##      Comparison p.Fisher p.adj.Fisher
## 3 Female : Male 0.000456      0.00137
```

```
pairwiseNominalIndependence(f18, compare = "row", fisher = T,
gtest = F, chisq = F, digits = 3)[3, ]
```

```
##      Comparison p.Fisher p.adj.Fisher
## 3 Female : Male 3.73e-06     1.12e-05
```

```r
pairwiseNominalIndependence(f19, compare = "row", fisher = T,
gtest = F, chisq = F, digits = 3)[3, ]
```

```
##      Comparison p.Fisher p.adj.Fisher
## 3 Female : Male     0.18        0.464
```

```r
pairwiseNominalIndependence(f20, compare = "row", fisher = T,
gtest = F, chisq = F, digits = 3)[3, ]
```

```
##      Comparison p.Fisher p.adj.Fisher
## 3 Female : Male    0.564        0.564
```

```r
pairwiseNominalIndependence(f21, compare = "row", fisher = T,
gtest = F, chisq = F, digits = 3)[3, ]
```

```
##      Comparison p.Fisher p.adj.Fisher
## 3 Female : Male   0.0121       0.0363
```

Using the adjusted P-value of the results, we can see that the most significant variables (using a threshold of 0.05) are: family_history, treatment, work_interfere, no_employees, tech_company, benefits, care_options, mental_health_consequence, phys_health_consequence, supervisor, mental_health_interview, and obs_consequence.

Let's see the confusion matrices for those that are significant.

```r
# We'll use proportion tables where the total amount of
# the respective gender is used instead of the grand total
# to calculate the proportions.

list("family_history" = prop.table(f2[-1, ], margin = 1),
     "Treatment" = prop.table(f3[-1, ],margin = 1),
     "work_interfere" = prop.table(f4[-1, ], margin = 1),
     "no_employees" = prop.table(f5[-1, ], margin = 1),
     "tech_company" = prop.table(f7[-1, ], margin = 1),
     "benefits" = prop.table(f8[-1, ], margin = 1),
     "care_options" = prop.table(f9[-1, ], margin = 1),
     "mental_health_consequence" = prop.table(f14[-1, ], margin = 1),
     "phys_health_consequence" = prop.table(f15[-1, ], margin = 1),
     "supervisor" = prop.table(f17[-1, ], margin = 1),
     "mental_health_interview" = prop.table(f18[-1, ], margin = 1),
     "obs_consequence" = prop.table(f21[-1, ], margin = 1))
```

```
## $family_history
##         family_history
## Gender          No       Yes
##   Female 0.4701195 0.5298805
##   Male   0.6477733 0.3522267
##
```

```
## $Treatment
##         treatment
## Gender          No       Yes
##   Female 0.3107570 0.6892430
##   Male   0.5455466 0.4544534
##
## $work_interfere
##         work_interfere
## Gender       Never     Often    Rarely Sometimes
##   Female 0.1162791 0.1674419 0.2093023 0.5069767
##   Male   0.2437746 0.1363041 0.1651376 0.4547837
##
## $no_employees
##         no_employees
## Gender          1-5    100-500     26-100   500-1000       6-25 More than 1000
##   Female 0.11553785 0.17928287 0.21912351 0.08764940 0.15537849     0.24302789
##   Male   0.12955466 0.12854251 0.23279352 0.03846154 0.25202429     0.21862348
##
## $tech_company
##         tech_company
## Gender          No       Yes
##   Female 0.2430279 0.7569721
##   Male   0.1649798 0.8350202
##
## $benefits
##         benefits
## Gender   Don't know        No       Yes
##   Female  0.2868526 0.1952191 0.5179283
##   Male    0.3360324 0.3228745 0.3410931
##
## $care_options
##         care_options
## Gender          No  Not sure       Yes
##   Female 0.2868526 0.2669323 0.4462151
##   Male   0.4301619 0.2449393 0.3248988
##
## $mental_health_consequence
##         mh_consequence
## Gender       Maybe        No       Yes
##   Female 0.4382470 0.2908367 0.2709163
##   Male   0.3674089 0.4149798 0.2176113
##
## $phys_health_consequence
##         ph_consequence
## Gender        Maybe         No        Yes
##   Female 0.27490040 0.66135458 0.06374502
##   Male   0.20344130 0.75506073 0.04149798
##
## $supervisor
##         supervisor
## Gender          No Some of them       Yes
##   Female 0.3466135    0.3466135 0.3067729
##   Male   0.3046559    0.2580972 0.4372470
##
```

```
## $mental_health_interview
##         mh_interview
## Gender         Maybe          No         Yes
##   Female 0.083665339 0.908366534 0.007968127
##   Male   0.186234818 0.777327935 0.036437247
##
## $obs_consequence
##         obs_consequence
## Gender          No       Yes
##   Female 0.8047809 0.1952191
##   Male   0.8684211 0.1315789
```

We'll make some simple interpretations of the results we see above, though there'll be some assumptions that'll be made:

family_history - Women are more likely to have/admit that they have a family history of mental illness than men.

treatment: Women are more likely to admit/seek out treatments for their mental health than men are.

work_interfere: More men claim to never have mental health issues interfere with their work than women.However, the other answers where there are some claims of interference have seemingly similar rates between both genders, though women still are a bit more likely to admit it.

no_employees: More men appear to work in smaller companies while women appear to work in larger companies.

tech_company: More men in this survey work in tech companies than women, although the gap seems to be narrowing.

benefits: More women seem to understand the mental health benefits their employers offer than men. Perhaps women are more likely to care about/seek out information relating to benefits than men are.

care_options: Similarly to "benefits", women are more likely to know and undersrtand their mental health care options their workplace offers than men are.

mental_health_consequence: More women believe that there's a possibility that discussing mental health issues with their employers will lead to negative consequences than men. However, more men are confident that talking about mental health issues won't lead to consequences than women. It's unknown if this is because men just don't seem to care much for mental health issues, are confident in their abilities to get their employers on their side, or something else.

phys_health_consequence: Very similar to the results and interpretation found in "mental_health_consequence".

supervisor: More men are likely to discuss mental health issues with their direct supervisors tahn women are. However, more women are likely to be

more selective about which supervisor to talk to compared to men acccording to percent of answers for "Some of them".

mental_health_interview: Women are both far less unlikely to outright discuss mental health issues in an interview and less likely to consider doing so than ment are. While 8% of women would consider bringing up the subject in an interview, nearly 19% would consider doing so. It is unclear if this is because men are more confident in bringing up mental health issues in an interview, women believe they'll be taken less seriously as a candidate, both, or something else entirely.

obs_consequence: Slighltly more women claimed to have heard of or observed negative consequences for coworkers with mental health issues than men have.

Is it possible to predict whether someone is male or female based on the results of the survey answers?

We'll use Random Forest to see if this is possible.

```
library(randomForest)
```

```
## Warning: package 'randomForest' was built under R version 3.6.3
```

```
## randomForest 4.6-14
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
##
## Attaching package: 'randomForest'
```

```
## The following object is masked from 'package:dplyr':
##
##     combine
```

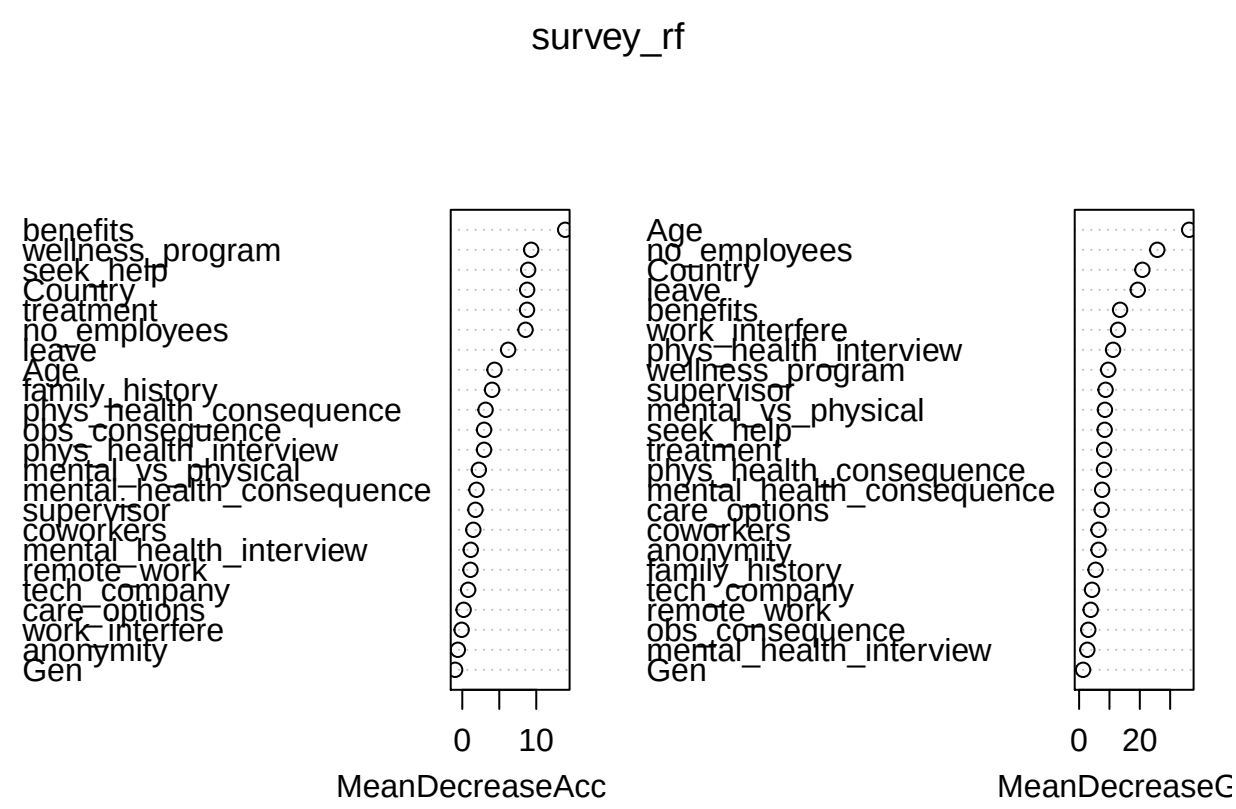
```
## The following object is masked from 'package:ggplot2':
##
##     margin
```

```
library(caret)
```

```
## Warning: package 'caret' was built under R version 3.6.3
```

```
## Loading required package: lattice
```

```r
set.seed(11)
mht_mf <- mht2[-which(mht2$Gender == "Other"), ]
mht_mf$Gender <- droplevels(mht_mf$Gender, exclude = "Other")
mht_mf <- mht_mf[, -c(1, 5, 6, 27)] # Exclude filler variables

# We'll split 70/30
sbst <- createDataPartition(mht_mf$Gender, p = 0.7, list = F)
train1 <- mht_mf[sbst, ]
test1 <- mht_mf[-sbst, ]

# We now begin the random forest modeling
survey_rf <- randomForest(Gender ~ ., train1, mtry = 23,
                          importance = T, na.action = na.omit)
# Using the model, we'll the testing eubset to make predictions.
survey_pred1 <- predict(survey_rf, test1)
varImpPlot(survey_rf)
```




```r
# Oddly enough, state and age seems to have the most influence on
# predicting the genders of the survey takers.

table_survey1 <- table("original" = test1$Gender, "prediction" = survey_pred1)
table_survey1
```

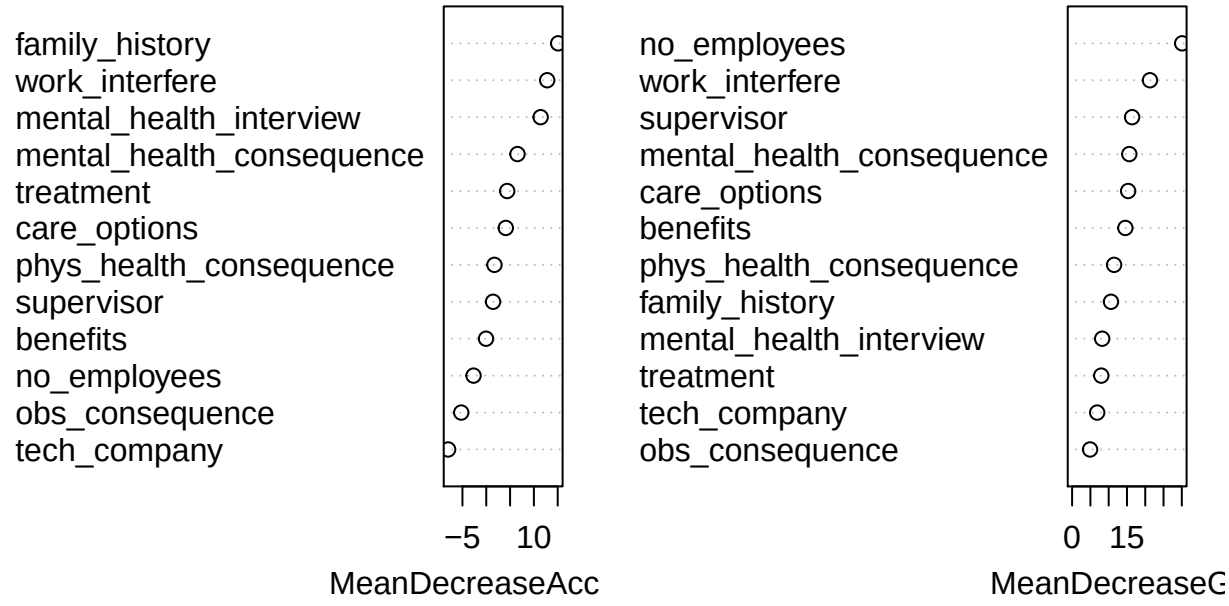
```
##          prediction
## original Female Male
```

```
##   Female      3   57
##   Male       10  222
```

```r
accuracy <- sum(diag(table_survey1)) / sum(table_survey1)
accuracy # Calculation of the prediction accuracy.
```

```
## [1] 0.7705479
```

```r
# While the accuracy percentage itself looks impressive, looking
# at the table does not as the model has a poor time predicting
# which of the survey takers are female whereas it has an easier
# time predicting male survey takers.

# Let's see what happens when we only use the variables that
# were found to be significant in the Fisher's Tests.
mht_mf2 <- mht_mf[, -c(1, 3, 8, 12, 13, 14, 15, 18,
                       21, 22, 24)]

# We'll split 70/30
sbst2 <- createDataPartition(mht_mf2$Gender, p = 0.5, list = F)
train2 <- mht_mf2[sbst2, ]
test2 <- mht_mf2[-sbst2, ]

# We now begin the random forest modeling
survey_rf2 <- randomForest(Gender ~ ., train2, mtry = 12,
                       importance = T, na.action = na.omit)
# Using the model, we'll the testing eubset to make predictions.
survey_pred2 <- predict(survey_rf2, test2)
varImpPlot(survey_rf2)
```




```
table_survey2 <- table("original" = test2$Gender, "prediction" = survey_pred2)
table_survey2
```

```
##          prediction
## original Female Male
##   Female     19   87
##   Male       53  335
```

```
accuracy2 <- sum(diag(table_survey2)) / sum(table_survey2)
accuracy2 # Calculation of the prediction accuracy.
```

```
## [1] 0.7165992
```

```
# The second model gains slightly more accuracy predicting which
# survey takers are female but loses some with predicting males.
# Overall, it's not very feasible to predict and distinguish between
# men and women using a model on the survey answers. However,
# this shouldn't discount the results gotten from the Fisher's Tests.
```

It's possible I may come back and continue analyzing this data set. Maybe next time, I'll see if geography produces differences in survey results. Maybe I'll try to optimize the random forest model and see if I squeeze out some more accuracy out of the prediction rates for women. But for now. Ultimately, I just think that there wasn't enough of a sample size compared to men and secondly, the answers (yes even some

of those deemed significant by the Fisher's Tests) were mostly similar for both genders. Granted, there are a few questions/variables where the answers were night and day but I don't even think those variables alone could've helped out the model. I'll definitely be moving onto a different data set for the meantime. Until next time...