

MENTAL HEALTHCARE

ASSIGNMENT 11.3

Siddhartha Bhaumik

2022-06-04

Introduction

- Mental health disorders are one of the leading health issues in the United States and it affects almost 10% of the population. This has significantly increased since the beginning of Covid19 pandemic and is not just seen in adults but children and young adults as well.
- Mental health awareness is another significant problem as lot of people are unaware of their problems until those turn into severe health issues like anxiety, depression, and other life-threatening risks.
- Also, many people don't know how to or from where to get help. Or they feel ashamed talking about their condition as that may impact their relationship both personally and professionally as well as their growth in life and at workplace.
- Another important thing to note is that in U.S there is a shortage in certified mental health professionals.
- Data Science/Artificial Intelligence can play an important part here by bridging some of the current gaps in Mental Healthcare sector.

Problem Statement

- Mental health is often overlooked which many times leads up to serious health issues. This is not specific to any age group as all are vulnerable. Based on the statistics, around 10% of U.S population has some form of mental health issue out of which up to 20% are children and young adults.
- Awareness and acceptance is the key to address this problem.

Steps taken To Address This Problem

- Easy access to counseling and proper guidance at every stage of life. Be it at home by parents, at school by teachers/professionals, at workplace by professionals, etc.
- Bring awareness and acceptance so that people who need help or people around you can identify this at an early stage and find support from friends/family and professionals. This can be done through wellness programs and seminars, etc.
- Majority of people have smartphones. AI mobile apps can be a great solution for self-assessment as well as therapy and other needs.
- Several fitness gadgets are in market which can track your sleep, heart rate, blood pressure, etc. and can share that information through apps which can further evaluate and predict your overall health.
- With so much patient data available digitally like health reports, lab reports, social media interactions, etc., different AI tools and techniques can analyze patient's data and flag physical and mental states. This can help in early detection and remedies.

Analysis

```
## Load required package
```

```
library(ggplot2)
```

```
library(pastecs)
```

```
library(dplyr)
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:pastecs':
```

```
##
```

```
## first, last
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
## filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## intersect, setdiff, setequal, union
```

```
library(purrr)
```

```
library(stringr)
```

```
library(factoextra) # clustering algorithms & visualization
```

```
## Welcome! Want to learn more? See two factoextra-related books at https://goo.gl/ve3WBa
```

```
library(cluster) # clustering algorithms
```

```
theme_set(theme_minimal())
```

```
## Load 'Any Mental Illness in the Past Year Data' from 51 US states to
```

```
state_any_mental_df <- read.csv("/Users/siddharthabhaumik/Documents/GitHub/dsc520/map_data.csv")
```

```
## Load 'Major Depressive Episode in the Past Year Data' from 51 US states to
```

```
state_dep_mental_df <- read.csv("/Users/siddharthabhaumik/Documents/GitHub/dsc520/data/map_data_rcrd.csv")
```

```
## Load 'Received Mentalhealth services in Past Year Data' from 51 US states to
```

```
state_rcvd_mental_df <- read.csv("/Users/siddharthabhaumik/Documents/GitHub/dsc520/data/map_data_dep.csv")
```

```
### Covid19 US dataset
```

```
covid19_us_df <- readxl::read_excel("/Users/siddharthabhaumik/Documents/GitHub/dsc520/data/search_term_1.csv")
```

```
## Load the 'Mental Health in Tech Survey' to
```

```
tech_survey_df <- read.csv("/Users/siddharthabhaumik/Documents/GitHub/dsc520/survey.csv")
```

```
## Viewing US Sample data related to 'Covid19 & Mental health effect/awareness'
```

```
head(covid19_us_df)
```

```
## # A tibble: 6 x 10
```

```
## Week depression anxiety `obsessive compulsive ~` ocd insomnia
```

```
## <dtm> <dbl> <dbl> <dbl> <dbl> <dbl>
```

```
## 1 2019-06-16 00:00:00 70 89 37 69 77
```

```
## 2 2019-06-23 00:00:00 70 91 51 73 83
```

```
## 3 2019-06-30 00:00:00      63      87      41      70      74
## 4 2019-07-07 00:00:00      74      92      60      74      84
## 5 2019-07-14 00:00:00      70      92      70      77      81
## 6 2019-07-21 00:00:00      75      93      42      72      82
## # ... with 4 more variables: `panic attack` <dbl>, `mental health` <dbl>,
## #   counseling <dbl>, psychiatrist <dbl>
```

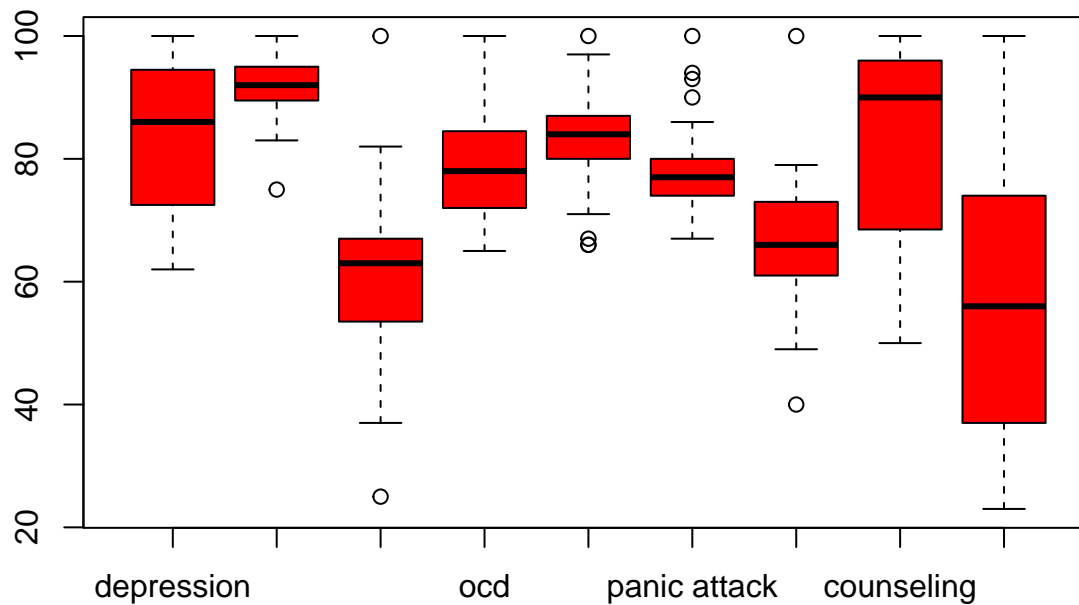
```
## summary
```

```
summary(covid19_us_df)
```

```
##      Week      depression      anxiety
## Min.   :2019-06-16 00:00:00 Min.   : 62.00 Min.   : 75.00
## 1st Qu.:2019-09-11 12:00:00 1st Qu.: 72.50 1st Qu.: 89.50
## Median :2019-12-08 00:00:00 Median : 86.00 Median : 92.00
## Mean   :2019-12-08 00:00:00 Mean   : 84.29 Mean   : 91.78
## 3rd Qu.:2020-03-04 12:00:00 3rd Qu.: 94.50 3rd Qu.: 95.00
## Max.   :2020-05-31 00:00:00 Max.   :100.00 Max.   :100.00
## obsessive compulsive disorder      ocd      insomnia      panic attack
## Min.   : 25.00      Min.   : 65.00 Min.   : 66.00 Min.   : 67
## 1st Qu.: 53.50      1st Qu.: 72.00 1st Qu.: 80.00 1st Qu.: 74
## Median : 63.00      Median : 78.00 Median : 84.00 Median : 77
## Mean   : 60.47      Mean   : 78.55 Mean   : 83.45 Mean   : 78
## 3rd Qu.: 67.00      3rd Qu.: 84.50 3rd Qu.: 87.00 3rd Qu.: 80
## Max.   :100.00      Max.   :100.00 Max.   :100.00 Max.   :100
## mental health      counseling      psychiatrist
## Min.   : 40.00 Min.   : 50.00 Min.   : 23.00
## 1st Qu.: 61.00 1st Qu.: 68.50 1st Qu.: 37.00
## Median : 66.00 Median : 90.00 Median : 56.00
## Mean   : 66.55 Mean   : 83.82 Mean   : 56.08
## 3rd Qu.: 73.00 3rd Qu.: 96.00 3rd Qu.: 74.00
## Max.   :100.00 Max.   :100.00 Max.   :100.00
```

```
## plot
```

```
boxplot(covid19_us_df[-1], col = "red", varwidth = TRUE, xlab = "Mental Health Disorders")
```



Mental Health Disorders

*# During Covid19 pandemic, The mental health related search terms like "depression",
"anxiety", "ocd", etc. increased significantly on google search engine.*

*# This shows people either impacted directly or hearing from someone from their family
and friends. But, it shows increased awareness and looking for options on the internet
to address this problem.*

*# From this dataset I can see Depression, Anxiety and counseling are the top 3 search
items but other than that not much information can be extracted from this dataset.*

US States Mental Heath Data

Viewing Sample data from 'Any mental illness/past year' dataset

`head(state_any_mental_df)`

```
##           outcome  age_group year_pair    state estimate
## 1 Any Mental Illness in the Past Year 18 or Older 2018-19    Alabama 0.212901
## 2 Any Mental Illness in the Past Year 18 or Older 2018-19     Alaska 0.214692
## 3 Any Mental Illness in the Past Year 18 or Older 2018-19    Arizona 0.200635
## 4 Any Mental Illness in the Past Year 18 or Older 2018-19    Arkansas 0.203352
## 5 Any Mental Illness in the Past Year 18 or Older 2018-19  California 0.194866
## 6 Any Mental Illness in the Past Year 18 or Older 2018-19    Colorado 0.231950
##   ci_lower ci_upper
## 1 0.190994 0.236585
## 2 0.194543 0.236314
## 3 0.179126 0.224021
## 4 0.182782 0.225598
## 5 0.184540 0.205624
## 6 0.209919 0.255546
```

`str(state_any_mental_df)`

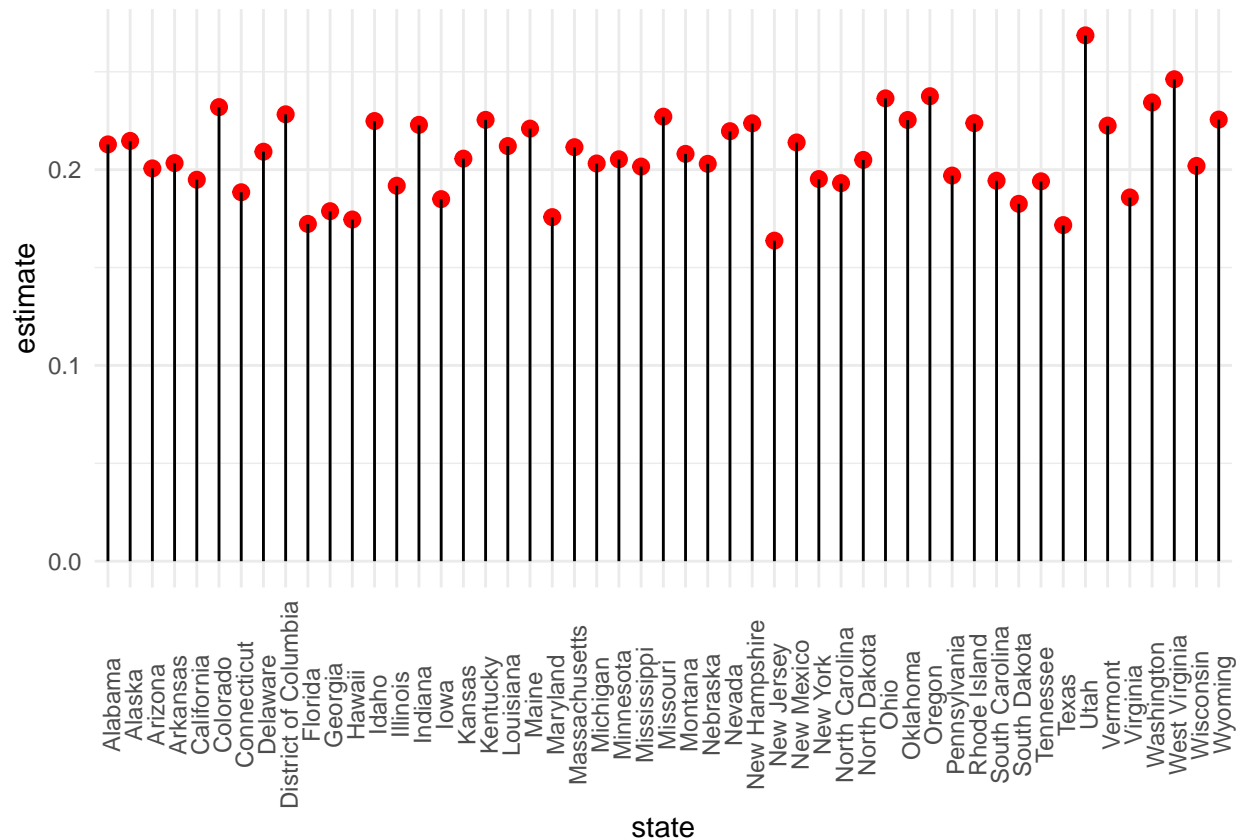
`## 'data.frame': 51 obs. of 7 variables:`

`## $ outcome : chr "Any Mental Illness in the Past Year" "Any Mental Illness in the Past Year" "Any Mental Illness in the Past Year"`

```
## $ age_group: chr "18 or Older" "18 or Older" "18 or Older" "18 or Older" ...
## $ year_pair: chr "2018-19" "2018-19" "2018-19" "2018-19" ...
## $ state : chr "Alabama" "Alaska" "Arizona" "Arkansas" ...
## $ estimate : num 0.213 0.215 0.201 0.203 0.195 ...
## $ ci_lower : num 0.191 0.195 0.179 0.183 0.185 ...
## $ ci_upper : num 0.237 0.236 0.224 0.226 0.206 ...
```

```
## Plot
```

```
ggplot(state_any_mental_df, aes(x = state, y = estimate)) +
  geom_point(size = 2.5, color = "red") +
  geom_segment(aes(x = state, xend = state, y = 0, yend = estimate)) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.7)) # Rotate axis label
```



```
# The top 3 states with any mental illness in year 2018-19 are Utah, West Virginia & Oregon.
```

```
## Viewing Sample data from 'Major depressive episode/past year' dataset
```

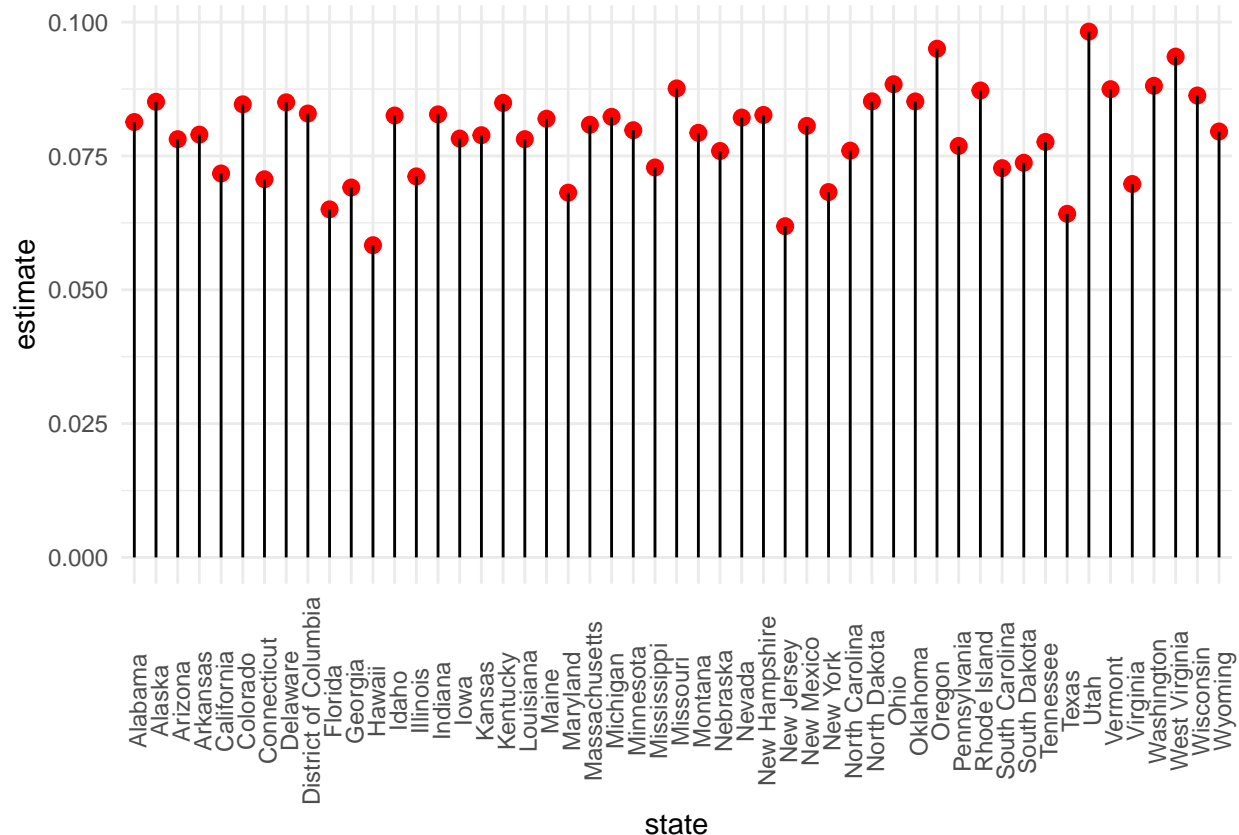
```
head(state_dep_mental_df)
```

```
##
## 1 Major Depressive Episode in the Past Year 18 or Older 2018-19 Alabama
## 2 Major Depressive Episode in the Past Year 18 or Older 2018-19 Alaska
## 3 Major Depressive Episode in the Past Year 18 or Older 2018-19 Arizona
## 4 Major Depressive Episode in the Past Year 18 or Older 2018-19 Arkansas
## 5 Major Depressive Episode in the Past Year 18 or Older 2018-19 California
## 6 Major Depressive Episode in the Past Year 18 or Older 2018-19 Colorado
## estimate ci_lower ci_upper
## 1 0.081327 0.069224 0.095330
## 2 0.085115 0.073037 0.098976
```

```
## 3 0.078124 0.066463 0.091629
## 4 0.078977 0.067765 0.091862
## 5 0.071717 0.065501 0.078473
## 6 0.084633 0.072943 0.097998
```

Plot

```
ggplot(state_dep_mental_df, aes(x = state, y = estimate)) +
  geom_point(size = 2.5, color = "red") +
  geom_segment(aes(x = state, xend = state, y = 0, yend = estimate)) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.7)) # Rotate axis label
```



#The top 3 states with major depressive episode in year 2018-19 are Utah, Oregon & West Virginia.

Viewing Sample data from 'Received Mental health services/past year' dataset

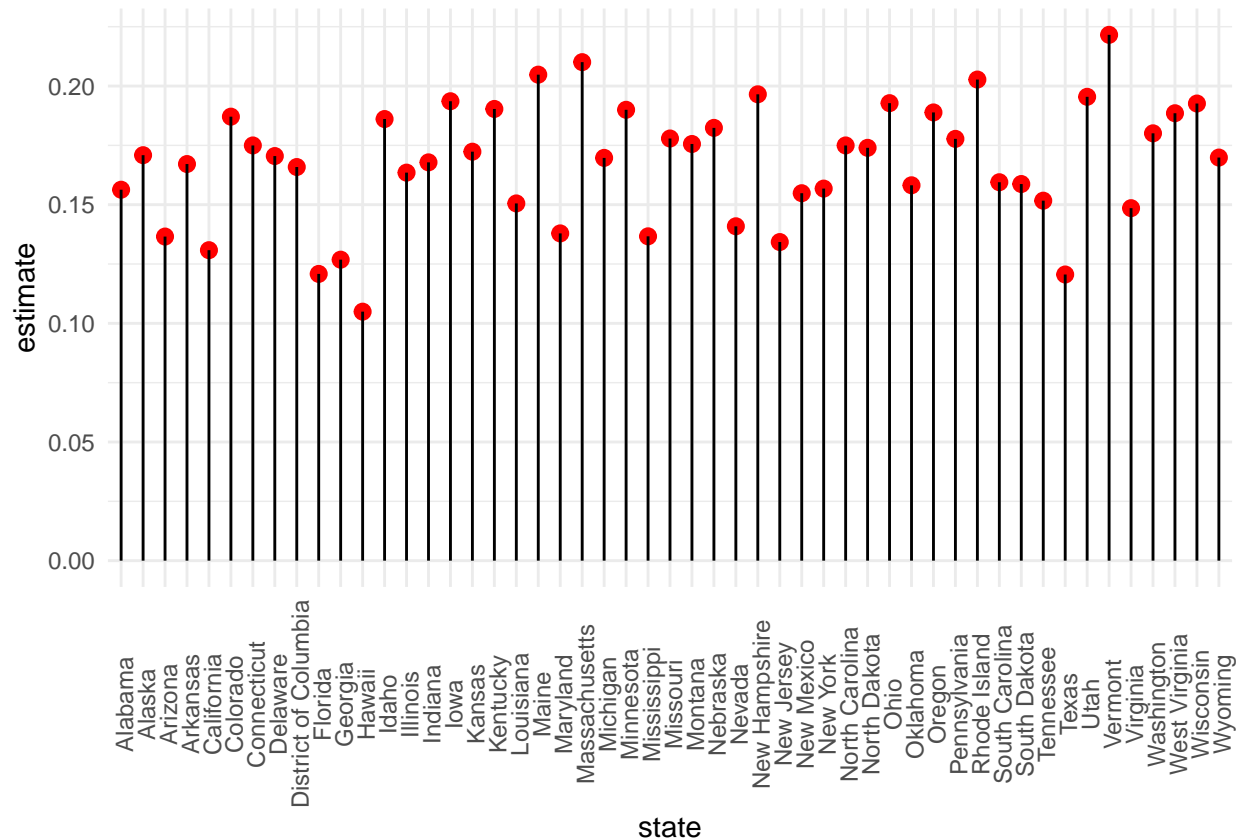
```
head(state_rcvd_mental_df)
```

```
##               outcome  age_group year_pair
## 1 Received Mental Health Services in the Past Year 18 or Older 2018-19
## 2 Received Mental Health Services in the Past Year 18 or Older 2018-19
## 3 Received Mental Health Services in the Past Year 18 or Older 2018-19
## 4 Received Mental Health Services in the Past Year 18 or Older 2018-19
## 5 Received Mental Health Services in the Past Year 18 or Older 2018-19
## 6 Received Mental Health Services in the Past Year 18 or Older 2018-19
##      state estimate ci_lower ci_upper
## 1   Alabama 0.156306 0.137093 0.177657
## 2    Alaska 0.170857 0.151458 0.192176
## 3   Arizona 0.136570 0.118184 0.157306
## 4  Arkansas 0.167128 0.148040 0.188134
```

```
## 5 California 0.130828 0.121788 0.140431
## 6 Colorado 0.187090 0.167010 0.208979
```

#Plot

```
ggplot(state_rcvd_mental_df, aes(x = state, y = estimate)) +
  geom_point(size = 2.5, color = "red") +
  geom_segment(aes(x = state, xend = state, y = 0, yend = estimate)) +
  theme(axis.text.x = element_text(angle = 90, vjust = 0.7)) # Rotate axis label
```



*#The top 3 states with people receiving mental health related services in year 2018-19
are Vermont, Massachusetts & Maine.*

*# So, based on the plots, we can see that Utah, Oregon and West Virginia are the
top 3 states for any mental health illness as well as major depressive episodes.*

*# But, from treatment perspective, none of the 3 are in the top 3 list.
This shows that even though lot of cases are reported, still not everyone is
availing any kind of treatments.*

Mental health at Work place

```
tech_survey_upd_df <- tech_survey_df %>% filter(Country == "United States") %>% filter(Age > 12) %>% se
head(tech_survey_upd_df)
```

```
##   Age Gender family_history treatment remote_work work_interfere  benefits
## 1  37 Female           No        Yes           No           Often      Yes
## 2  44      M           No        No           No           Rarely Don't know
```

```
## 3 31 Male No No Yes Never Yes
## 4 33 Male Yes No No Sometimes Yes
## 5 35 Female Yes Yes Yes Sometimes No
## 6 42 Female Yes Yes No Sometimes Yes
## wellness_program seek_help anonymity mental_health_consequence
## 1 No Yes Yes No
## 2 Don't know Don't know Don't know Maybe
## 3 Don't know Don't know Don't know No
## 4 No Don't know Don't know No
## 5 No No No Maybe
## 6 No No No Maybe
## obs_consequence
## 1 No
## 2 No
## 3 No
## 4 No
## 5 No
## 6 No
```

Standardize Gender with Male, Female, Other

```
tech_survey_upd_df["Gender"][tech_survey_upd_df["Gender"] == "M" | tech_survey_upd_df["Gender"] == "m" |
  tech_survey_upd_df["Gender"] == "male" | tech_survey_upd_df["Gender"] == "Male" |
  tech_survey_upd_df["Gender"] == "Cis male" | tech_survey_upd_df["Gender"] == "Cis Male" |
  tech_survey_upd_df["Gender"] == "Male-ish" | tech_survey_upd_df["Gender"] == "Male-ish" |
  tech_survey_upd_df["Gender"] == "Man" | tech_survey_upd_df["Gender"] == "Man" |
  tech_survey_upd_df["Gender"] == "Malr" | tech_survey_upd_df["Gender"] == "Malr" |
  tech_survey_upd_df["Gender"] == "Mal" | tech_survey_upd_df["Gender"] == "Mal" |
  tech_survey_upd_df["Gender"] == "maile" | tech_survey_upd_df["Gender"] == "maile"] <- "Male"

tech_survey_upd_df["Gender"][tech_survey_upd_df["Gender"] == "F" | tech_survey_upd_df["Gender"] == "f" |
  tech_survey_upd_df["Gender"] == "female" | tech_survey_upd_df["Gender"] == "female" |
  tech_survey_upd_df["Gender"] == "Cis Female" | tech_survey_upd_df["Gender"] == "Cis Female" |
  tech_survey_upd_df["Gender"] == "Woman" | tech_survey_upd_df["Gender"] == "Woman" |
  tech_survey_upd_df["Gender"] == "Femake" | tech_survey_upd_df["Gender"] == "Femake" |
  tech_survey_upd_df["Gender"] == "Female (trans)"] <- "Female"

tech_survey_upd_df["Gender"][tech_survey_upd_df["Gender"] == "Female (trans)" | tech_survey_upd_df["Gender"] == "non-binary" |
  tech_survey_upd_df["Gender"] == "non-binary" | tech_survey_upd_df["Gender"] == "Genderqueer" | tech_survey_upd_df["Gender"] == "Genderqueer" |
  tech_survey_upd_df["Gender"] == "Trans woman" ] <- "Others"

head(tech_survey_upd_df)
```

```
## Age Gender family_history treatment remote_work work_interfere benefits
## 1 37 Female No Yes No Often Yes
## 2 44 Male No No No Rarely Don't know
## 3 31 Male No No Yes Never Yes
## 4 33 Male Yes No No Sometimes Yes
## 5 35 Female Yes Yes Yes Sometimes No
## 6 42 Female Yes Yes No Sometimes Yes
## wellness_program seek_help anonymity mental_health_consequence
## 1 No Yes Yes No
## 2 Don't know Don't know Don't know Maybe
## 3 Don't know Don't know Don't know No
```



```
## 4          No Don't know Don't know          No
## 5          No          No          No        Maybe
## 6          No          No          No        Maybe
## obs_consequence
## 1          No
## 2          No
## 3          No
## 4          No
## 5          No
## 6          No
```

*# Only considering survey results from United States as its the majority.
 # Noticed some negative numbers under 'Age' column which I will filter out.
 # Under 'Gender' column, I see lot of variation and spelling error like Male,Mail,maile,
 # M, Cis Male, Female, Cis Female, etc. I will make it consistent as Male, Female, Other.
 # Dropped some columns like State, No of Employee, Tech company, etc. as I don't think
 # they add much value.*

Basically I am looking for how many people opted for 'Treatment'.

Viewing Mental health in US Tech industry

```
head(tech_survey_upd_df)
```

```
## Age Gender family_history treatment remote_work work_interfere benefits
## 1  37 Female          No      Yes          No      Often      Yes
## 2  44  Male          No      No           No      Rarely Don't know
## 3  31  Male          No      No           Yes      Never      Yes
## 4  33  Male          Yes      No           No      Sometimes Yes
## 5  35 Female          Yes      Yes          Yes      Sometimes No
## 6  42 Female          Yes      Yes          No      Sometimes Yes
## wellness_program seek_help anonymity mental_health_consequence
## 1          No      Yes      Yes          No
## 2      Don't know Don't know Don't know      Maybe
## 3      Don't know Don't know Don't know      No
## 4          No Don't know Don't know      No
## 5          No      No      No      Maybe
## 6          No      No      No      Maybe
## obs_consequence
## 1          No
## 2          No
## 3          No
## 4          No
## 5          No
## 6          No
```

#summary

```
summary(tech_survey_upd_df)
```

```
##      Age      Gender      family_history      treatment
## Min.   : 18.00  Length:747      Length:747      Length:747
## 1st Qu.: 28.00  Class :character  Class :character  Class :character
## Median : 32.00  Mode  :character  Mode  :character  Mode  :character
## Mean   : 33.53
## 3rd Qu.: 38.00
## Max.   :329.00
## remote_work      work_interfere      benefits      wellness_program
```

```
## Length:747          Length:747          Length:747          Length:747
## Class :character    Class :character    Class :character    Class :character
## Mode :character     Mode :character     Mode :character     Mode :character
##
##
##
## seek_help          anonymity          mental_health_consequence
## Length:747          Length:747          Length:747
## Class :character    Class :character    Class :character
## Mode :character     Mode :character     Mode :character
##
##
##
## obs_consequence
## Length:747
## Class :character
## Mode :character
##
##
##
```

```
str(tech_survey_upd_df)
```

```
## 'data.frame':    747 obs. of  12 variables:
## $ Age              : num  37 44 31 33 35 42 31 42 36 29 ...
## $ Gender            : chr   "Female" "Male" "Male" "Male" ...
## $ family_history    : chr   "No" "No" "No" "Yes" ...
## $ treatment         : chr   "Yes" "No" "No" "No" ...
## $ remote_work       : chr   "No" "No" "Yes" "No" ...
## $ work_interfere    : chr   "Often" "Rarely" "Never" "Sometimes" ...
## $ benefits          : chr   "Yes" "Don't know" "Yes" "Yes" ...
## $ wellness_program  : chr   "No" "Don't know" "Don't know" "No" ...
## $ seek_help         : chr   "Yes" "Don't know" "Don't know" "Don't know" ...
## $ anonymity         : chr   "Yes" "Don't know" "Don't know" "Don't know" ...
## $ mental_health_consequence: chr   "No" "Maybe" "No" "No" ...
## $ obs_consequence   : chr   "No" "No" "No" "No" ...
```

```
# Replace Gender, Wellness program & Seek help column values to numeric
```

```
tech_survey_upd_df["Gender"][tech_survey_upd_df["Gender"] == "Male "] <- 1
```

```
tech_survey_upd_df["Gender"][tech_survey_upd_df["Gender"] == "Female "] <- 2
```

```
tech_survey_upd_df["Gender"][tech_survey_upd_df["Gender"] == "Others"] <- 3
```

```
tech_survey_upd_df["family_history"][tech_survey_upd_df["family_history"] == "Yes"] <- 1
```

```
tech_survey_upd_df["family_history"][tech_survey_upd_df["family_history"] == "No"] <- 2
```

```
tech_survey_upd_df["wellness_program"][tech_survey_upd_df["wellness_program"] == "Yes"] <- 1
```

```
tech_survey_upd_df["wellness_program"][tech_survey_upd_df["wellness_program"] == "No"] <- 2
```

```
tech_survey_upd_df["wellness_program"][tech_survey_upd_df["wellness_program"] == "Don't know"] <- 3
```

```
tech_survey_upd_df["seek_help"][tech_survey_upd_df["seek_help"] == "Yes"] <- 1
```

```
tech_survey_upd_df["seek_help"][tech_survey_upd_df["seek_help"] == "No"] <- 2
```

```
tech_survey_upd_df["seek_help"][tech_survey_upd_df["seek_help"] == "Don't know"] <- 3
```

```
# Filter out age above 100
```

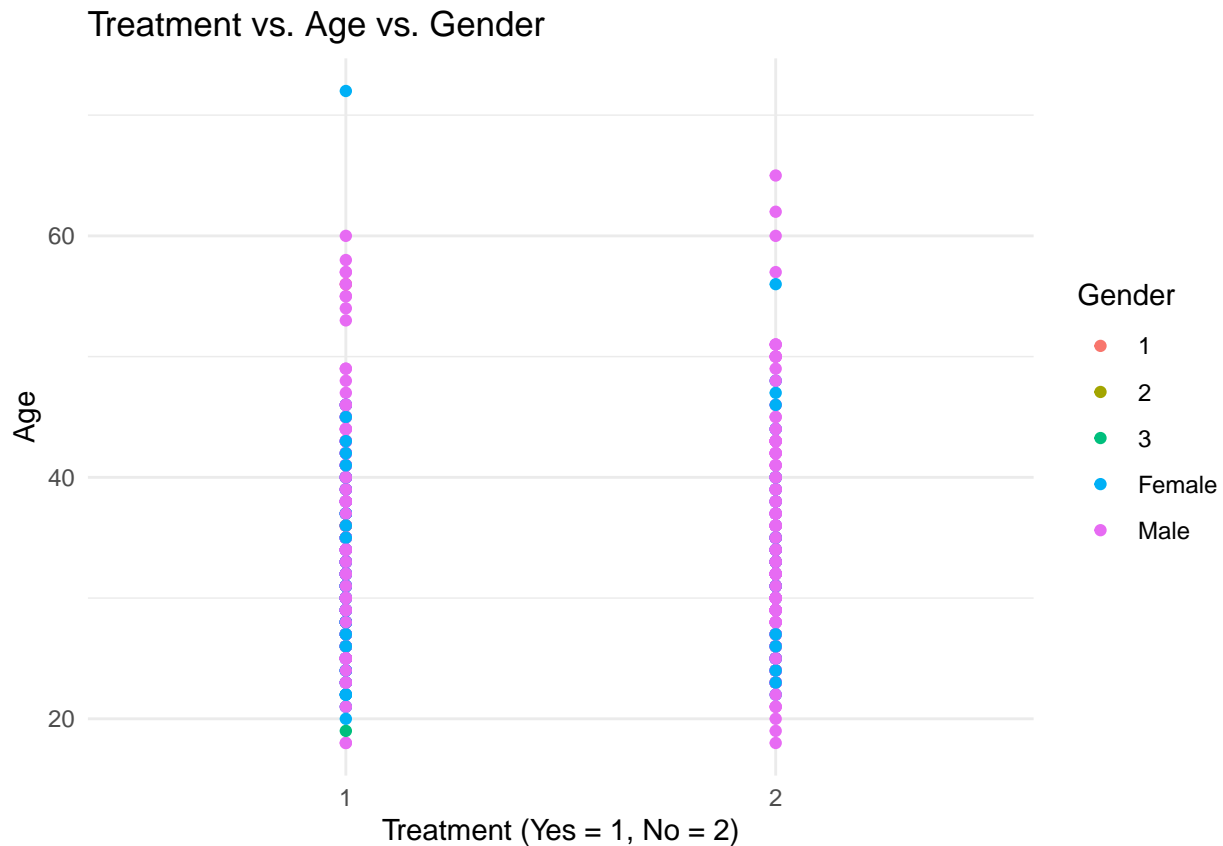
```
tech_survey_upd_df2 <- tech_survey_upd_df %>% filter(Age < 100) %>% select(Age,Gender,family_history,treatment)
```

```
tech_survey_upd_df2["treatment"][tech_survey_upd_df2["treatment"] == "Yes"] <- 1
```

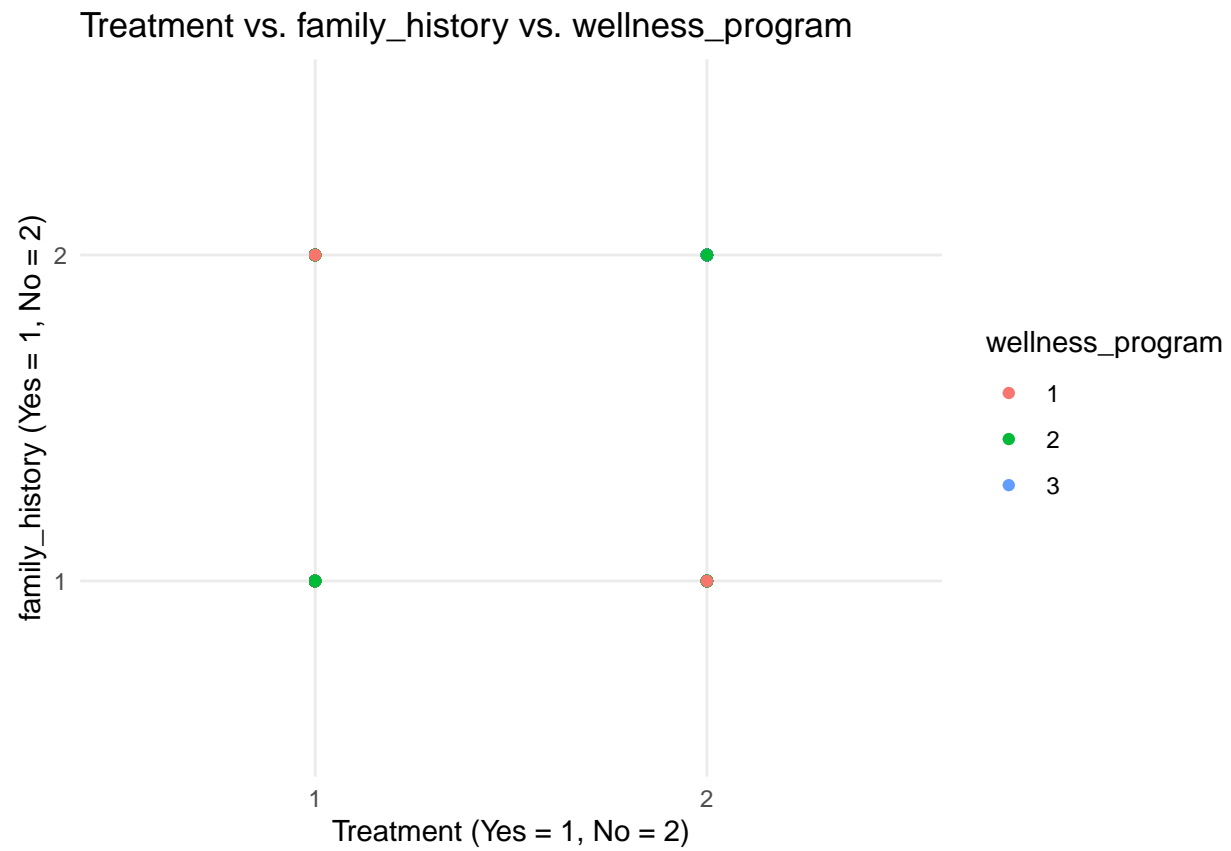
```
tech_survey_upd_df2["treatment"][tech_survey_upd_df2["treatment"] == "No"] <- 2
```

```
# Plot
```

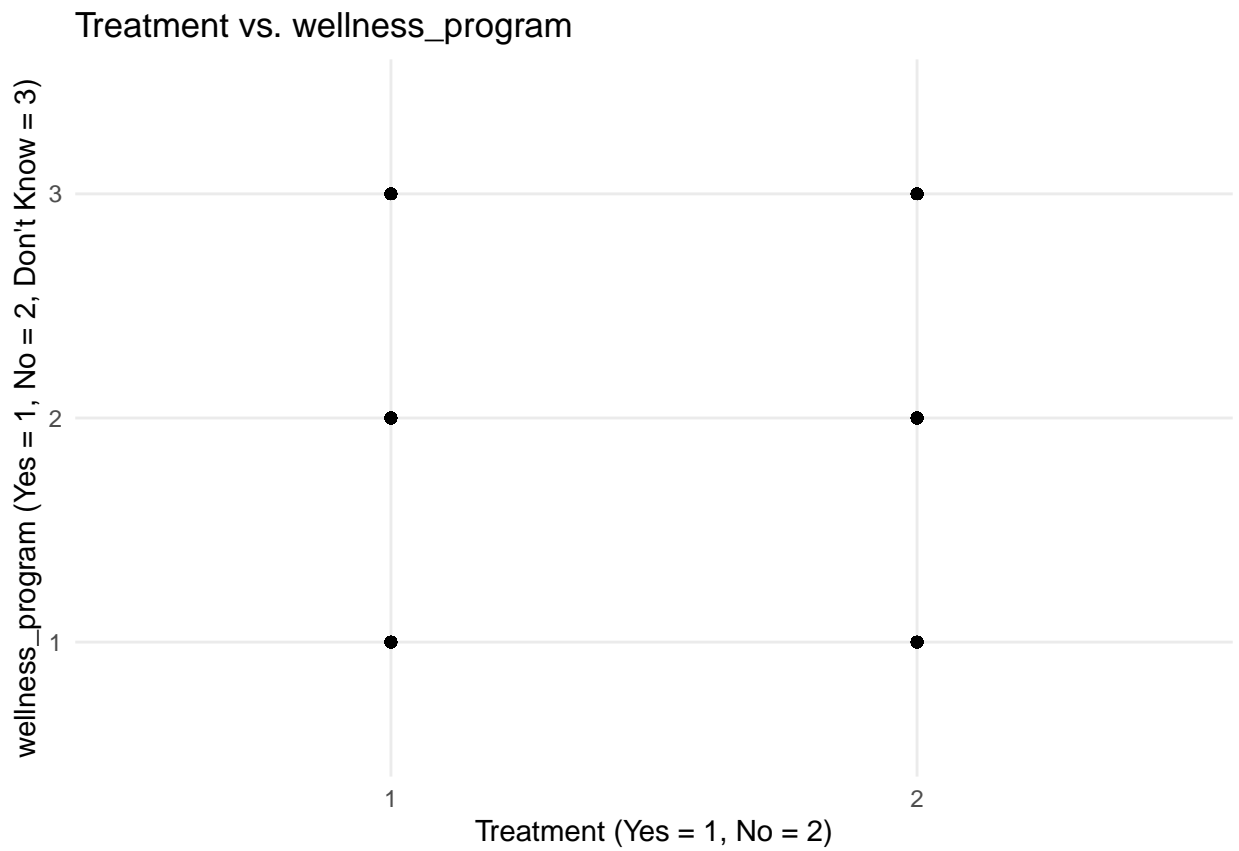
```
ggplot(tech_survey_upd_df2, aes(x=treatment, y=Age, col=Gender)) + geom_point() + ggtitle("Treatment vs. Age vs. Gender")
```



```
ggplot(tech_survey_upd_df2, aes(x=treatment, y=family_history, col=wellness_program)) + geom_point() +
```

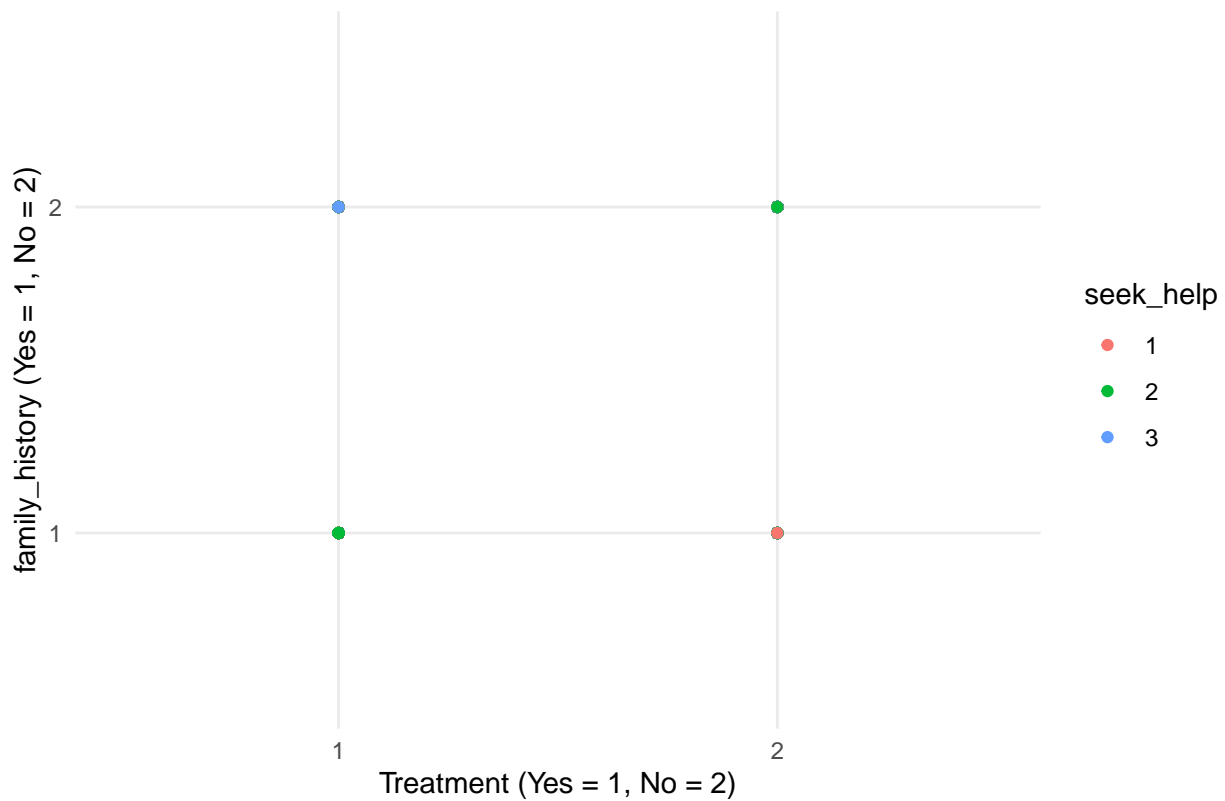


```
ggplot(tech_survey_upd_df2, aes(x=treatment, y=wellness_program)) + geom_point() + ggtitle("Treatment vs. family_history vs. wellness_program")
```

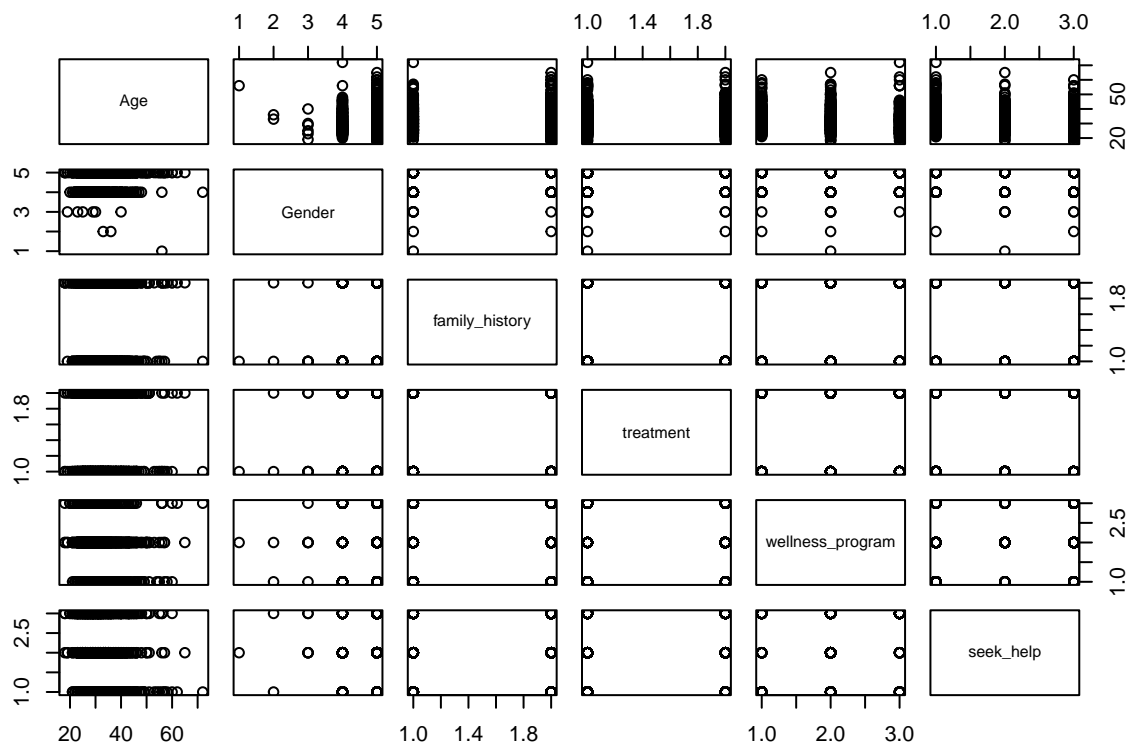


```
ggplot(tech_survey_upd_df2, aes(x=treatment, y=family_history, col=seek_help)) + geom_point() + ggtitle
```

Treatment vs. family_history vs. Seek_help



```
plot(tech_survey_upd_df2)
```



My target variable here is 'treatment'. Workplaces which promotes mental health and support their employees through different benefits and wellness programs will see more

people opting for treatments and other kind of help if needed.

*# I see more Male population tied to mental health but I see the ratio of male population
much higher than female and others in the dataset.*

This can be due to more Male working in the tech industry or the survey population is biased.

*# I also looked at family history to see if those people tend to be more aware qnd
seek help if needed but didnt see much correlation.*

Limitations

- The survey datasets picked up for Covid19 doesn't have much details to link it with other US State mental health datasets. Also, I was looking for overall workplace dataset and not specifically tech industry dataset.

Concluding Remarks

- AI is going to revolutionize the mental health care system due to its accessibility, affordability, efficiency, accuracy, and support. Despite some of the drawbacks related to overall data privacy, I feel AI is the key to control this problem through smartphones and other smart devices accessible to people and wellness/healthcare professionals in near future.

Datasets/Citations

- "COVID-19 and Mental Health Search Terms" dataset from Kaggle. <https://www.kaggle.com/datasets/luckybro/mental-health-search-term> The search interest of mental health related terms on Google before and after the outbreak of COVID-19 pandemic reveals how public's concern is affected by the pandemic, and its impact to mental health of people around the world.
- "Mental Health in Tech Survey" dataset from Kaggle <https://www.kaggle.com/datasets/osmi/mental-health-in-tech-survey> This dataset is from a 2014 survey that measures attitudes towards mental health and frequency of mental health disorders in the tech workplace.
- "Any Mental Illness in the Past Year among Adults Aged 18 or Older, by State: 2018-2019" dataset from SAMHDA.gov <https://pdas.samhsa.gov/saes/state> This dataset is maintained by 'Substance Abuse & Mental Health Data Archive' government agency and contains any type of mental health related issues in adults aged 18 and older for the year 2018-2019.