MENTAL HEALTH IN UNIVERSITY AND INDUSTRY

MENTAL HEALTH MANIACS

Ann Biju, Cindy Cho, Ally de Vera, Amulya Rayasam

WHY DOES IT MATTER?

According to LinkedIn, 51% of tech professionals have been diagnosed with a mental health condition

57% of tech employees report experiencing constant burnout.

The four of us as well as many of you will be choosing careers in the technological industry, so we believe it is important for all of us to remain informed.

71% of tech professionals say their productivity is directly correlated to a mental health issue.

RESEARCH OBJECTIVES

How does getting treatment affect mental health in academia?

What is the relation between GPA and mental health disorders?

Do both males and females have the same percentage of mental health distress in the industry?

How do people perceive their mental health to interfere with their work?

DATA ANALYZED

Student Mental Health Dataset

•Conducted by a Google forms by an unknown university to examine their current academic situation and mental health

•11 columns with over 100 entries

Mental Health in Tech Dataset

• From a 2014 survey that measures attitudes towards mental health and frequency of mental health disorders in the tech workplace

•27 columns with 1259 entries

UNIVERSITY FINDINGS

Cleaning

- Removed columns that we were not interested in such as, ‘timestamp’, ‘age’, and ‘marital status’
- Removed all entries that did not have students majoring in 1 of the following areas of study: Engineering, BIT, Computer Science, or Biotechnology
- The values in the mental health issue columns (depression, panic attack, and anxiety) had to be changed to binary values of ‘0’ and ‘1’, so that regression could be performed on the data

As a result of step 3, **logistic regression** was performed!

Discoveries

Getting treatment has a significant effect on mental health issues seen in students - particularly depression

Lowest AIC score = best model = 22.55

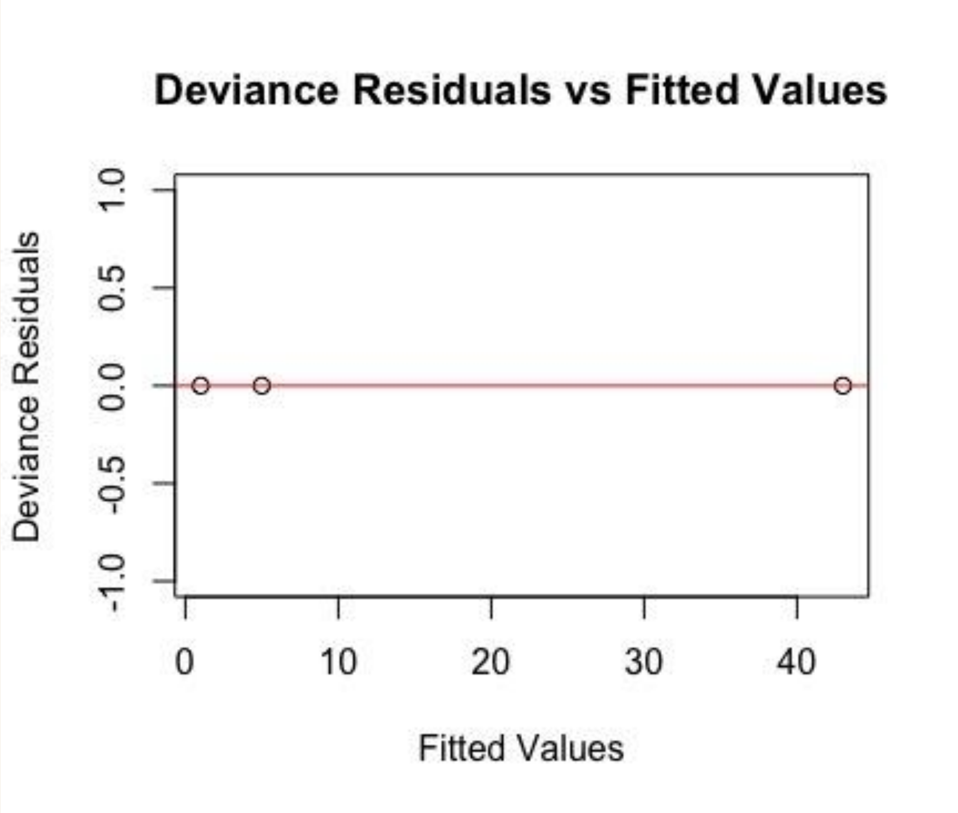
P value = 0.995 = very strong correlation between getting treatment and mental health issues

```
> step.mod <- stepAIC(model, direction = "both")
Start:  AIC=24.71
treatment ~ depression + panic.attack + anxiety

               Df Deviance    AIC
- anxiety       1   16.864 22.864
- panic.attack  1   18.532 24.532
<none>              16.709 24.709
- depression    1   23.307 29.307

Step:  AIC=22.86
treatment ~ depression + panic.attack

               Df Deviance    AIC
- panic.attack  1   18.550 22.550
<none>              16.864 22.864
+ anxiety       1   16.709 24.709
- depression    1   23.448 27.448

Step:  AIC=22.55
treatment ~ depression

               Df Deviance    AIC
<none>              18.550 22.550
+ panic.attack  1   16.864 22.864
+ anxiety       1   18.532 24.532
- depression    1   27.180 29.180
```

```
> summary(step.mod)

Call:
glm(formula = treatment ~ depression, family = "binomial", data = clean.students)

Deviance Residuals: 
     Min        1Q    Median        3Q       Max  
-0.73248  -0.73248  -0.00005  -0.00005   1.70113  

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   -20.57    3292.45  -0.006    0.995
depression     19.39    3292.45   0.006    0.995

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 27.18  on 45  degrees of freedom
Residual deviance: 18.55  on 44  degrees of freedom
AIC: 22.55

Number of Fisher Scoring iterations: 19
```

GPA and Mental Health Issues

```
> summary(model)

Call:
glm(formula = total_issues ~ gpa_group, family = poisson(), data = gpa_issues)

Deviance Residuals:
[1]  0  0  0

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)
(Intercept)     4.676e-11  1.000e+00   0.000 1.000000
gpa_groupmedium 1.609e+00  1.095e+00   1.469 0.141774
gpa_grouphigh   3.761e+00  1.012e+00   3.718 0.000201 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance:  6.5823e+01  on 2  degrees of freedom
Residual deviance: -6.6614e-16  on 0  degrees of freedom
AIC: 17.084

Number of Fisher Scoring iterations: 3
```



GPA and Mental Health Issues

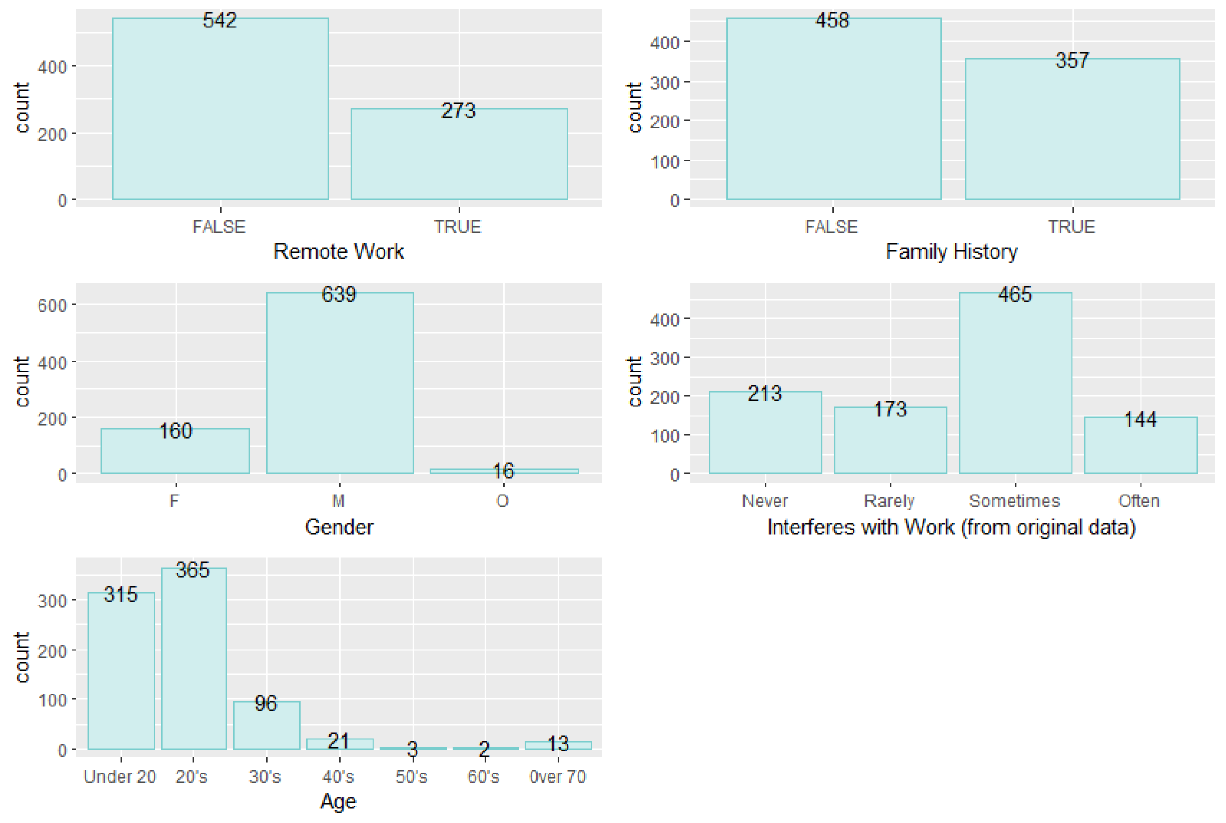
```
> stepAIC(model, direction = "both")
Start:  AIC=17.08
total_issues ~ gpa_group

            Df Deviance    AIC
<none>            0.000 17.084
- gpa_group  2   65.823 78.907

Call:  glm(formula = total_issues ~ gpa_group, family = poisson(), data = gpa_issues)

Coefficients:
    (Intercept)  gpa_groupmedium    gpa_grouphigh
      4.676e-11        1.609e+00        3.761e+00

Degrees of Freedom: 2 Total (i.e. Null);  0 Residual
Null Deviance:      65.82
Residual Deviance: -6.661e-16   AIC: 17.08
> exp(coef(stepwise_model))
    (Intercept) gpa_groupmedium   gpa_grouphigh
              1               5              43
```

INDUSTRY FINDINGS

Cleaning

- Out of the 27 variables, we deemed certain variables insignificant based on whether we thought they could accurately predict the existence of mental illness in an individual working in tech
- Examples include :
  - Observed consequences of speaking about mental health in the workplace
  - Attitudes towards informing coworkers of mental illness
  - Does one's mental illness affect their work
- Chose variables in our data set that we deemed valid predictors of mental illness in an individual working in tech, and encoded these variables numerically
- In our later analysis, we look at variables we removed for this predictive model
- Age, Gender, Country, Family history, Remote work, Tech company. Response variable identified as "Treatment"
- Further cleaned data by only including data entries where individual responded TRUE for working at a tech company -> 815 objects of 7 variables
- All variables were either logical or categorical variables, so we chose to use a logistic regression model

Our Variables



Variables considered for logistic regression model

- Remote work
- Gender
- Age
- Family History

Other:

- Perceived interference with work

Variable Selection

Fitting the Largest Possible Model to the Data

```
Call:
glm(formula = treatment ~ Age + Gender + Country + family_history +
    remote_work, data = mh_valid_tech)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.0191  -0.4577   0.1658   0.3517   0.7947

Coefficients:
                               Estimate Std. Error t value Pr(>|t|)
(Intercept)                   5.827e-01  1.689e-01   3.450  0.00059 ***
Age                           2.143e-03  1.266e-03   1.693  0.09086 .
GenderF                       4.391e-02  1.252e-01   0.351  0.72595
GenderM                      -1.279e-01  1.215e-01  -1.052  0.29308
CountryAustria               -5.062e-01  4.601e-01  -1.100  0.27156
CountryBahamas, The           6.372e-02  4.763e-01   0.134  0.89362
CountryBelgium               -5.105e-01  4.601e-01  -1.110  0.26749
CountryBosnia and Herzegovina -8.449e-01  4.605e-01  -1.835  0.06692 .
CountryBrazil                -4.255e-01  2.494e-01  -1.706  0.08841 .
CountryBulgaria               7.739e-02  2.817e-01   0.275  0.78363
CountryCanada                -1.766e-01  1.288e-01  -1.371  0.17074
CountryChina                 -5.556e-01  4.603e-01  -1.207  0.22780
CountryColombia              -5.127e-01  3.351e-01  -1.530  0.12641
CountryCroatia                2.880e-01  3.351e-01   0.859  0.39034
CountryCzech Republic        -8.469e-01  4.603e-01  -1.840  0.06615 .
CountryDenmark                1.488e-01  3.354e-01   0.444  0.65743
CountryFinland               -2.980e-02  3.351e-01  -0.089  0.92916
CountryFrance                -2.982e-02  2.498e-01  -0.119  0.90499
CountryGermany               -8.144e-02  1.400e-01  -0.582  0.56101
CountryHungary               -1.021e+00  4.615e-01  -2.212  0.02727 *
CountryIndia                  1.210e-01  2.499e-01   0.484  0.62835
CountryIreland               -1.081e-01  1.468e-01  -0.737  0.46164
CountryIsrael                -7.571e-01  3.351e-01  -2.260  0.02413 *
CountryItaly                 -4.400e-01  2.499e-01  -1.761  0.07867 .
```

```
CountryJapan                  1.037e-01  4.609e-01   0.225  0.82201
CountryMexico                -1.873e-01  3.344e-01  -0.560  0.57557
CountryMoldova                4.895e-01  4.601e-01   1.064  0.28769
CountryNetherlands           -2.354e-01  1.514e-01  -1.555  0.12040
CountryNew Zealand           -7.784e-02  2.022e-01  -0.385  0.70033
CountryPhilippines           -5.212e-01  4.601e-01  -1.133  0.25757
CountryPoland                -7.356e-02  2.022e-01  -0.364  0.71605
CountryPortugal              -5.127e-01  4.601e-01  -1.114  0.26548
CountryRussia                -5.127e-01  3.351e-01  -1.530  0.12641
CountrySingapore             -3.661e-01  2.494e-01  -1.468  0.14255
CountrySlovenia               5.045e-01  4.602e-01   1.096  0.27337
CountrySouth Africa          -6.046e-02  2.818e-01  -0.215  0.83016
CountrySweden                -3.116e-01  2.292e-01  -1.360  0.17432
CountrySwitzerland           -8.372e-02  2.290e-01  -0.366  0.71475
CountryThailand              -5.405e-01  4.602e-01  -1.174  0.24059
CountryUnited Kingdom        -2.750e-02  1.199e-01  -0.229  0.81869
CountryUnited States         -5.986e-02  1.136e-01  -0.527  0.59826
CountryZimbabwe              -2.143e+08  1.266e+08  -1.693  0.09086 .
family_historyTRUE            3.214e-01  3.265e-02   9.844  < 2e-16 ***
remote_workTRUE               1.508e-02  3.446e-02   0.438  0.66182
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.1987842)

    Null deviance: 190.86  on 814  degrees of freedom
Residual deviance: 153.26  on 771  degrees of freedom
AIC: 1041

Number of Fisher Scoring iterations: 2
```

Variable Selection

New Best Model after performing variable selection

```
> summary(transformed)

Call:
glm(formula = treatment ~ Gender + family_history, data = mh_valid_tech)

Deviance Residuals:
    Min       1Q   Median       3Q      Max  
-0.9410  -0.4484   0.2294   0.3812   0.5516  

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)         0.59099    0.11442   5.165 3.03e-07 ***
GenderF             0.02777    0.11783   0.236    0.814    
GenderM            -0.14261    0.11405  -1.250    0.212    
family_historyTRUE  0.32220    0.03206  10.050  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for gaussian family taken to be 0.2017017)

    Null deviance: 190.86  on 814  degrees of freedom
Residual deviance: 163.58  on 811  degrees of freedom
AIC: 1014.1

Number of Fisher Scoring iterations: 2
```

The variable selection process removed age, country, and remote work as significant predictors of mental illness within an individual

Variable Selection

Variance Inflation Factor

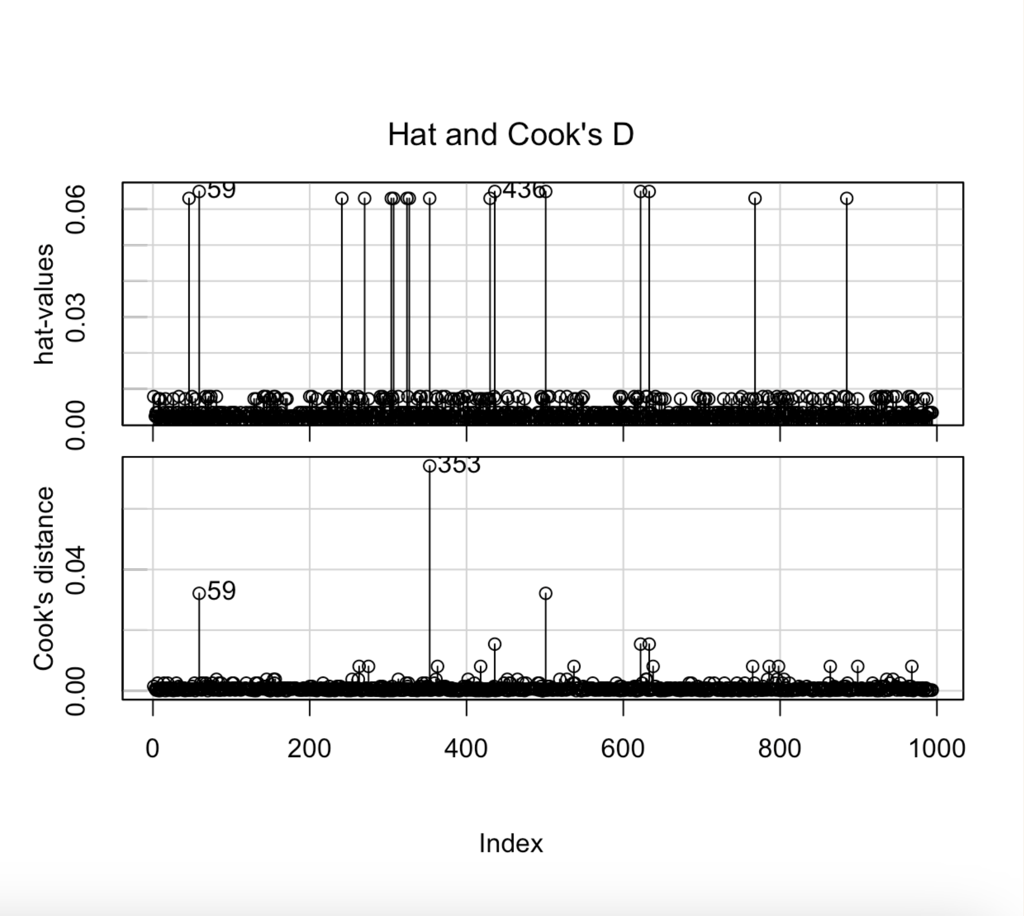
```
> vif(transformed)
                  GVIF Df GVIF^(1/(2*Df))
Gender         1.02226  2        1.005519
family_history 1.02226  1        1.011069
```

After model transformation, we see that VIF values for Gender or family history exceed 10, so multicollinearity does not exist within the transformed model.

Influence Analysis

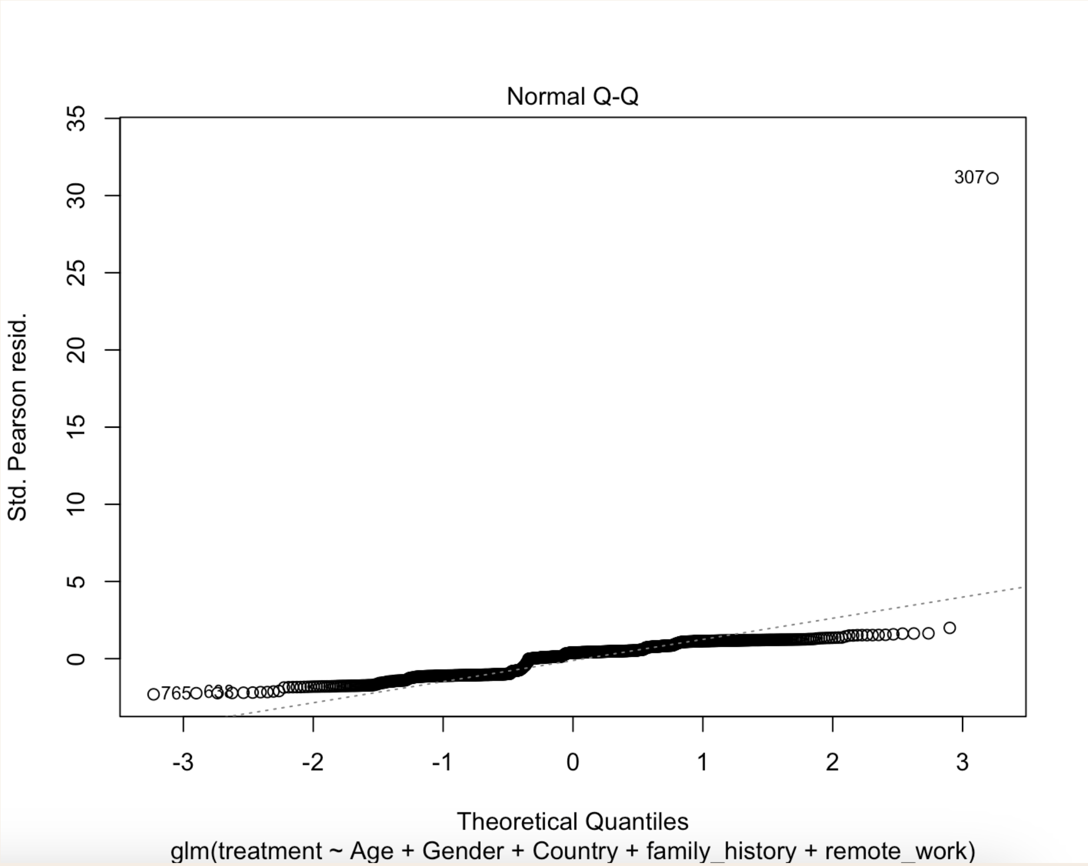
Identified 16 outliers based on hat values

```
> summary(influence.measures(transformed))
Potentially influential observations of
         glm(formula = treatment ~ Gender + family_history, data = mh_valid_tech) :

    dfb.1_ dfb.GndF dfb.GndM dfb.f_TR dffit   cov.r   cook.d hat
46   0.05  -0.05    -0.05     0.00     0.05    1.07_*  0.00   0.06_*
59  -0.36   0.34     0.35     0.07    -0.36_*  1.06_*  0.03   0.06_*
241  0.05  -0.05    -0.05     0.00     0.05    1.07_*  0.00   0.06_*
270  0.05  -0.05    -0.05     0.00     0.05    1.07_*  0.00   0.06_*
304  0.05  -0.05    -0.05     0.00     0.05    1.07_*  0.00   0.06_*
307  0.05  -0.05    -0.05     0.00     0.05    1.07_*  0.00   0.06_*
324  0.05  -0.05    -0.05     0.00     0.05    1.07_*  0.00   0.06_*
327  0.05  -0.05    -0.05     0.00     0.05    1.07_*  0.00   0.06_*
353 -0.52   0.52     0.53    -0.05    -0.55_*  1.05_*  0.07   0.06_*
430  0.05  -0.05    -0.05     0.00     0.05    1.07_*  0.00   0.06_*
436  0.25  -0.23    -0.24    -0.05     0.25_*  1.07_*  0.02   0.06_*
501 -0.36   0.34     0.35     0.07    -0.36_*  1.06_*  0.03   0.06_*
622  0.25  -0.23    -0.24    -0.05     0.25_*  1.07_*  0.02   0.06_*
633  0.25  -0.23    -0.24    -0.05     0.25_*  1.07_*  0.02   0.06_*
768  0.05  -0.05    -0.05     0.00     0.05    1.07_*  0.00   0.06_*
885  0.05  -0.05    -0.05     0.00     0.05    1.07_*  0.00   0.06_*
```

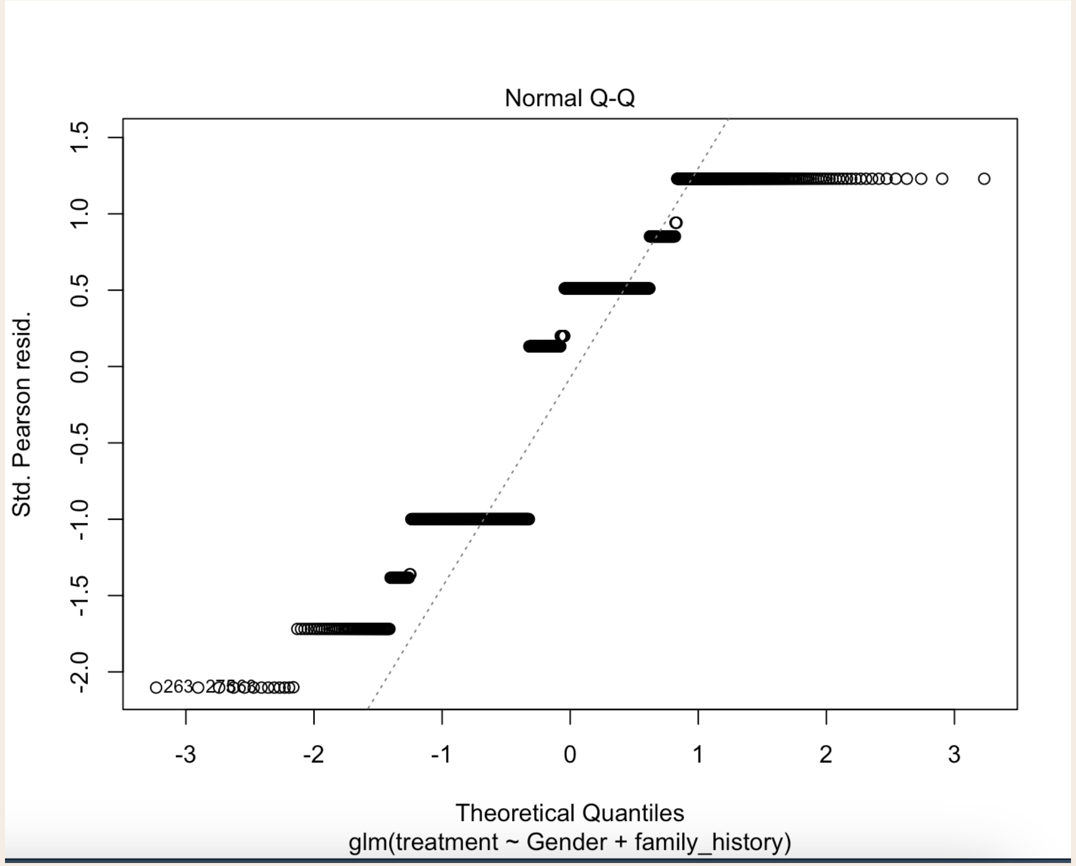


QQ Plots

Before Model Selection

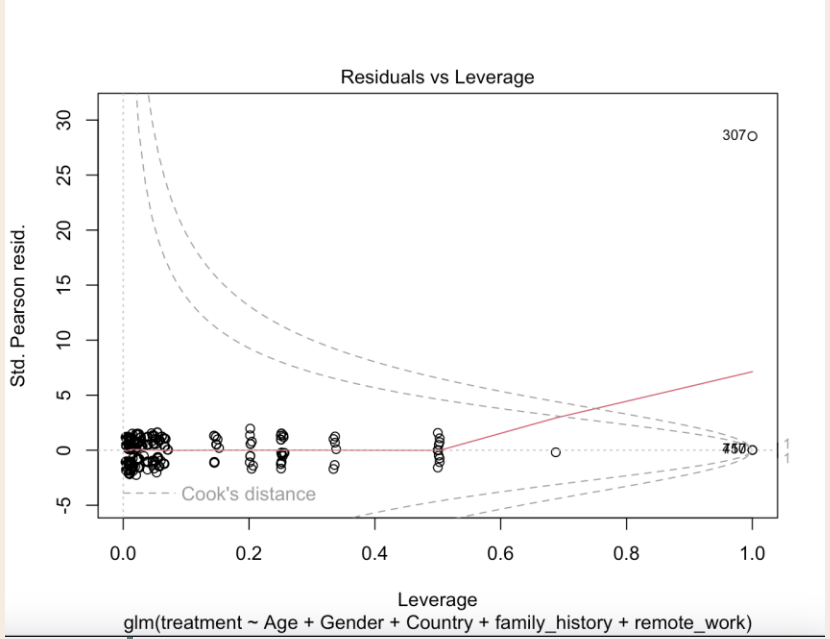


After Model Selection

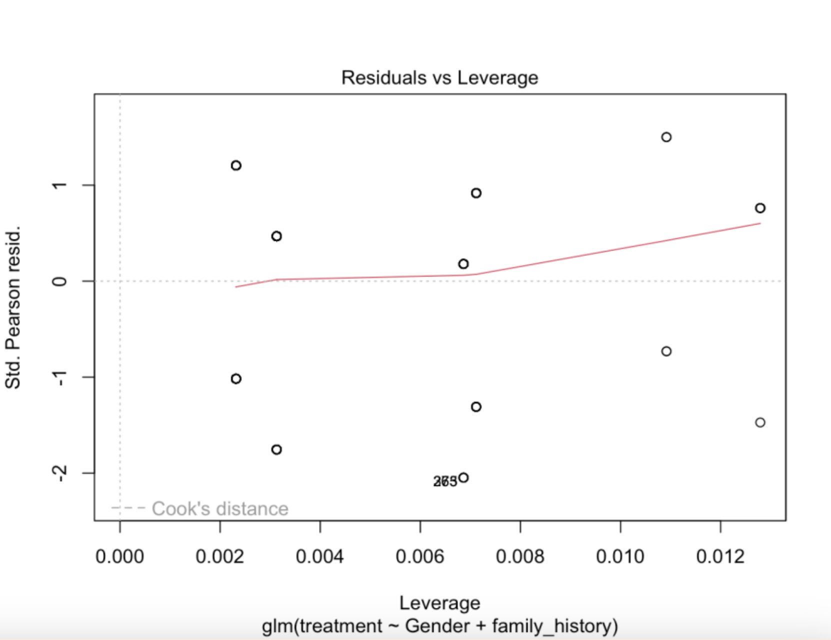


Residuals vs Leverage

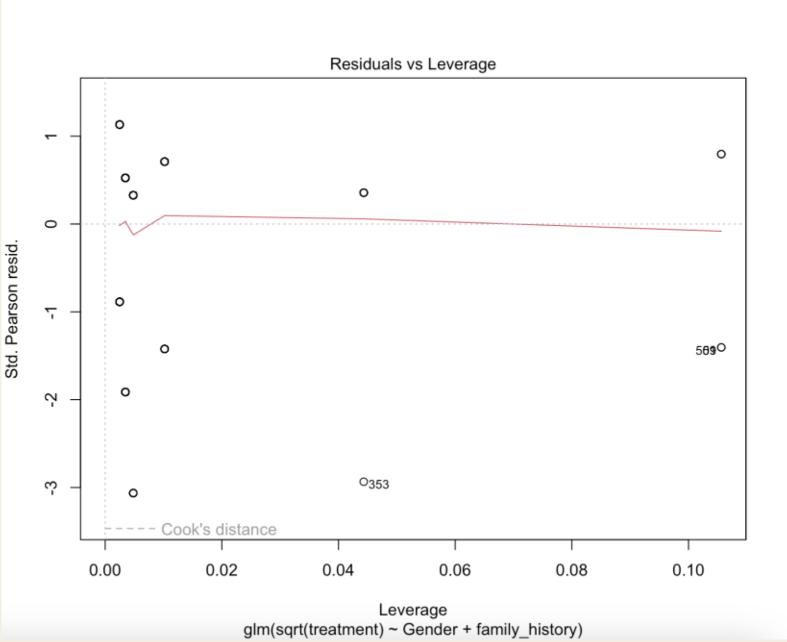
Before Model Selection



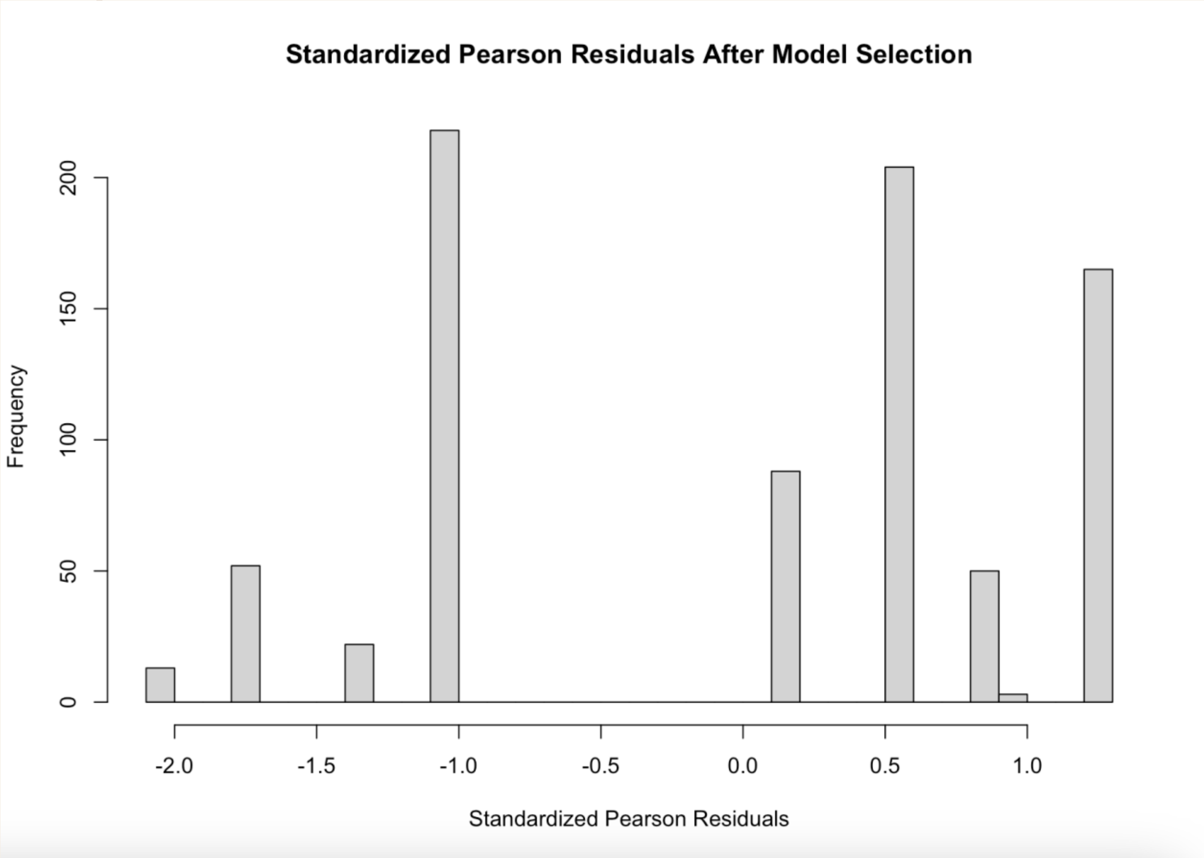
After Model Selection

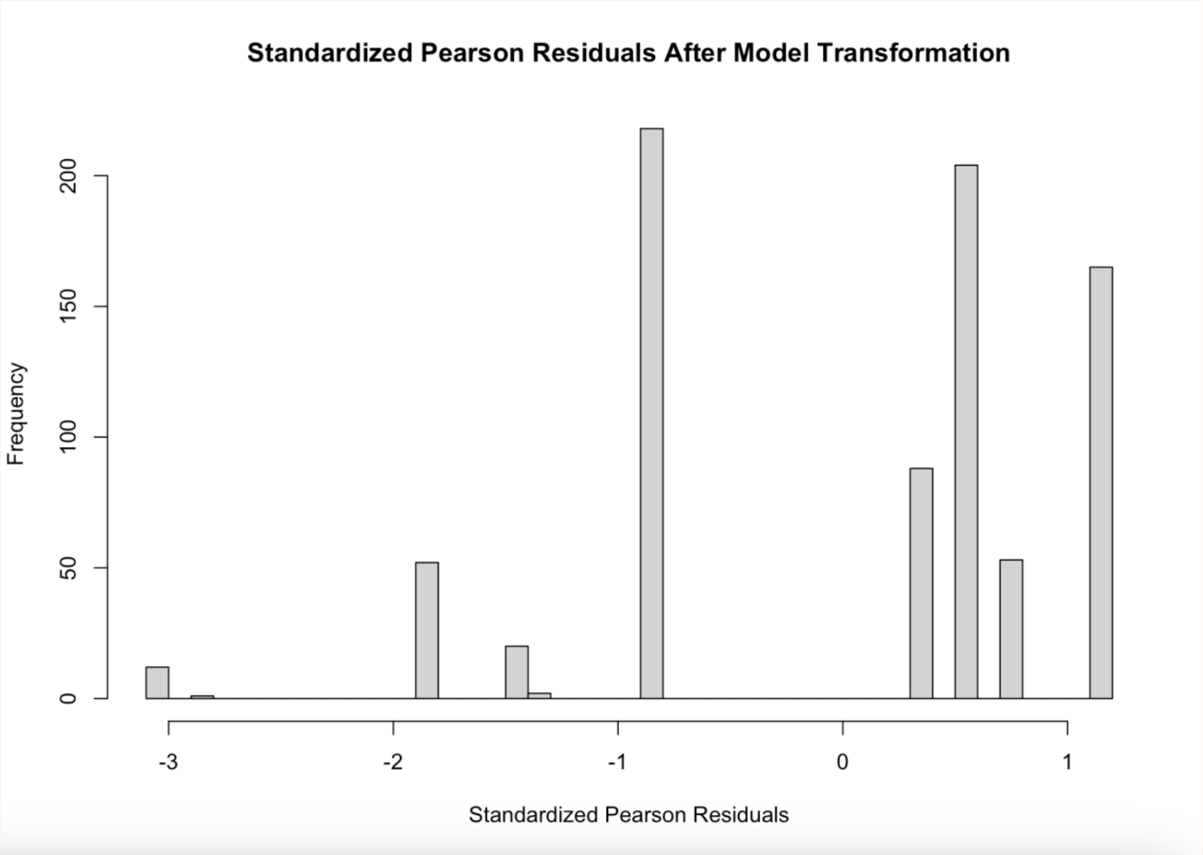


After Model Transformation

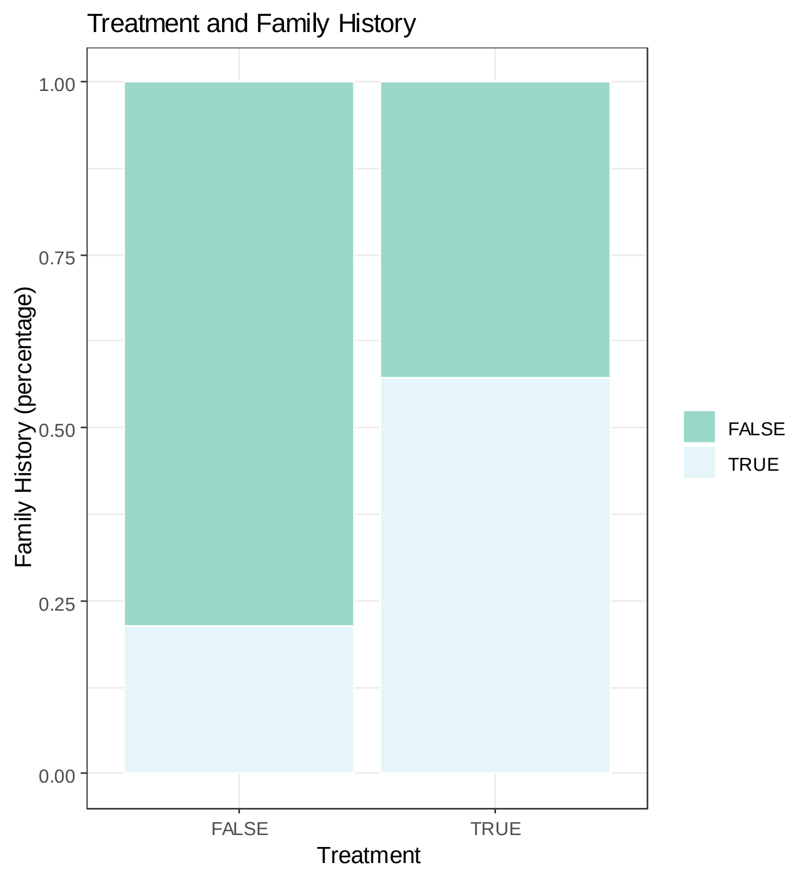


Residual Analysis of Model Transformation Using Pearson Method



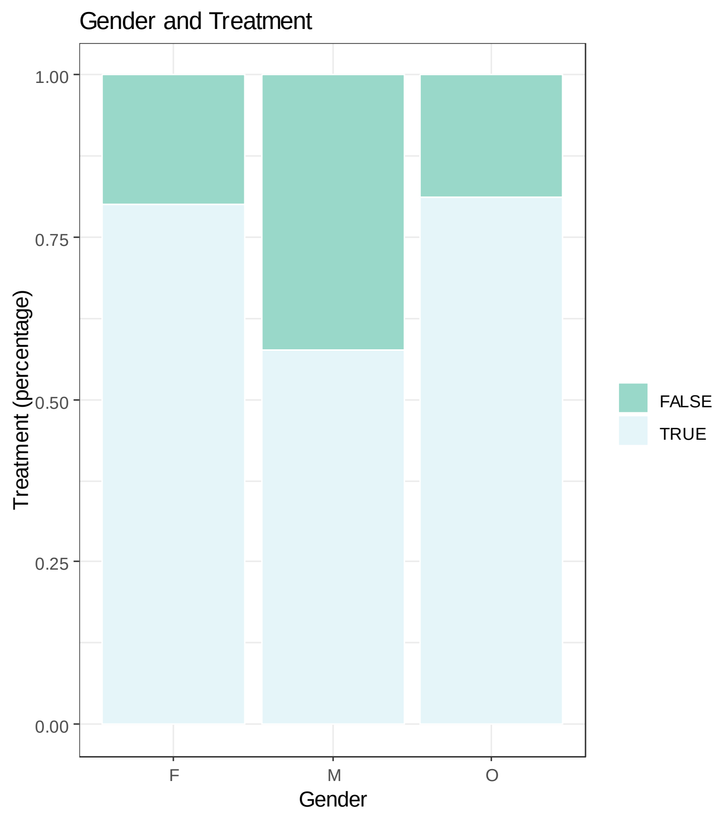


Family History



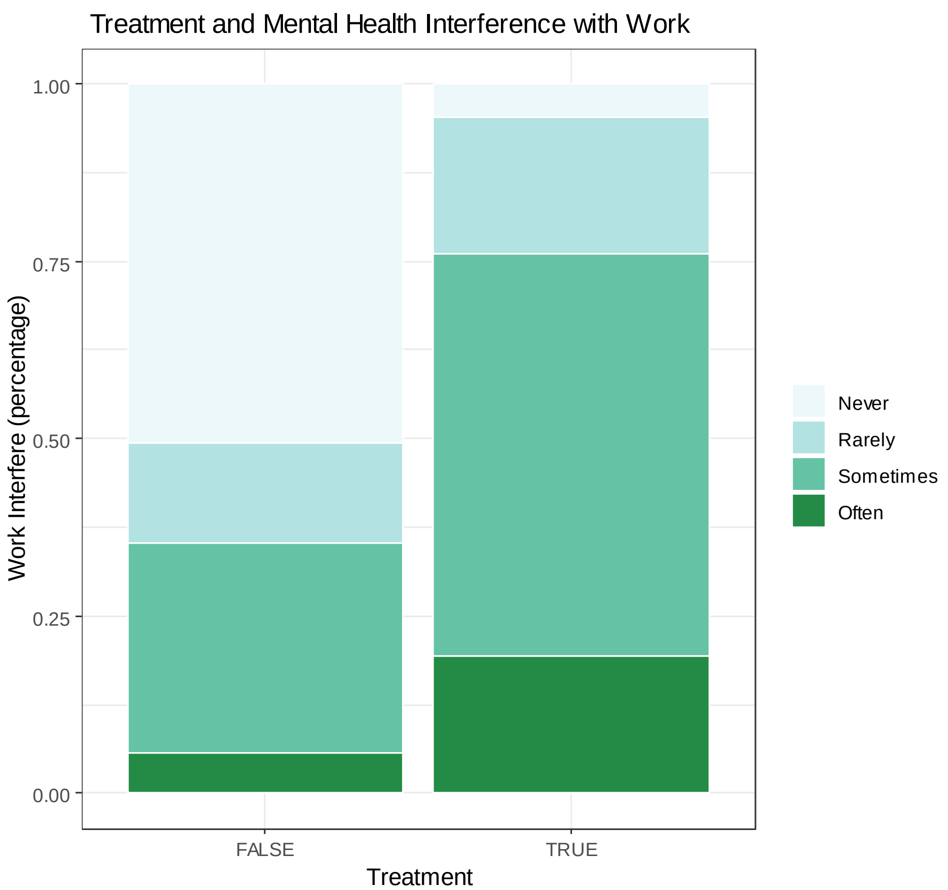

```
Call:
glm(formula = treatment ~ family_history, family = binomial,
    data = mental_health_transf)

Deviance Residuals:
   Min       1Q   Median       3Q      Max
-1.846  -1.137   0.634    1.218    1.218

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)        -0.09614    0.09356  -1.028    0.304
family_historyTRUE  1.59851    0.16602   9.628   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1077.71  on 814  degrees of freedom
Residual deviance:  972.67  on 813  degrees of freedom
AIC: 976.67

Number of Fisher Scoring iterations: 4
```

Gender




```
Call:
glm(formula = treatment ~ Gender, family = binomial, data = mental_health_transf)

Deviance Residuals: 
    Min       1Q   Median       3Q      Max  
-1.8297  -1.3126   0.6681   1.0480   1.0480  

Coefficients:
            Estimate Std. Error z value Pr(>|z|)    
(Intercept)  1.38629    0.19764   7.014 2.31e-12 ***
GenderM     -1.07392    0.21325  -5.036 4.76e-07 ***
GenderO      0.08004    0.67031   0.119    0.905    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1077.7  on 814  degrees of freedom
Residual deviance: 1046.0  on 812  degrees of freedom
AIC: 1052

Number of Fisher Scoring iterations: 4
```

Interference with Work




```
Call:
glm(formula = treatment ~ work_interfere, family = binomial,
    data = mental_health)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9623  -0.5510   0.7232   0.7232   1.9800

Coefficients:
                        Estimate Std. Error z value Pr(>|z|)
(Intercept)              -1.8083     0.1970   -9.18   <2e-16 ***
work_interfereRarely      2.6805     0.2581   10.39   <2e-16 ***
work_interfereSometimes   3.0160     0.2257   13.36   <2e-16 ***
work_interfereOften       3.5760     0.3075   11.63   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1304.6  on 994  degrees of freedom
Residual deviance: 1004.3  on 991  degrees of freedom
AIC: 1012.3

Number of Fisher Scoring iterations: 4
```

KEY FINDINGS

Treatment has a positive effect in students with depression

GPA and mental health are highly correlated, as GPA increased, the probability of experiencing mental health disorders increased

Male had the lowest proportion of those with a mental health disorder and female and other were about equal

Almost all employee's who have a mental health disorder feel that it in some way impedes with their work

FUTURE RESEARCH

- Find more explicit data such as mental health days taken, number of overtime hours, credit hours, vacation days taken, etc.
- Obtain a more balanced distribution of genders
- Consider a larger, continuous data set to observe trends over time
- Compare among a wider variety of majors and professional fields
- Narrow the scope of where data was pulled from (less countries)

THANK YOU!

QUESTIONS?

SOURCES

- https://www.linkedin.com/pulse/mental-health-major-concern-tech-industry-mentortribes?trk=public_post-content_share-article#:~:text=Mental%20Health%20Facts&text=51%25%20of%20tech%20professionals%20have,tech%20industry%20employees%20reported%20burnout
- https://www.kaggle.com/datasets/shariful07/student-mental-health
- https://www.kaggle.com/datasets/osmi/mental-health-in-tech-survey