# Survey on Mental Health Across Programs of Study in University

Mental Health Maniacs - Deanna King, Jim Moroney, Pooja Patel, Christopher Wilhite

2022-11-23

#### **Packages**

In order to better interpret the dataset, we utilize the pander package for table creation and manipulation.

library(pander)

#### Student Mental Health Data

All data required for this interpretation was obtained from the International Islamic University in Malaysia. This dataset is publicly available via Kaggle, and contains the following features:

- Timestamp time at which the survey was completed
- Choose your gender gender (male or female) of the participant
- Age age of the participant at the time of survey completion
- What is your course? program in which the participant is majoring
- ullet Your current year of Study how many years the participant has attended university
- What is your CGPA? current grade point average (or the ratio of grade points earned to grade points attempted), calculated on a 0.0-4.0 scale
- Marital Status describes whether or not the participant is married
- Do you have Depression? states whether or not the participant has depression
- Do you have Anxiety? states whether or not the participant has anxiety
- Do you have Panic attacks? states whether or not the participant experiences panic attacks
- Did you seek any specialist for a treatment? states whether or not the participant sought professional treatment for any mental health concerns

```
studentData_df <- read.csv(file="./StudentMentalHealth.csv")
#str(studentData_df)
summary(studentData_df)</pre>
```

```
##
     Timestamp
                       Choose.your.gender
                                                           What.is.your.course.
##
   Length: 101
                       Length: 101
                                          Min.
                                                 :18.00
                                                           Length: 101
   Class : character
                       Class : character
                                          1st Qu.:18.00
                                                           Class : character
   Mode :character Mode :character
                                          Median :19.00
                                                           Mode : character
                                                  :20.53
##
                                          Mean
##
                                           3rd Qu.:23.00
##
                                          Max.
                                                  :24.00
##
                                          NA's
## Your.current.year.of.Study What.is.your.CGPA. Marital.status
  Length: 101
                               Length: 101
                                                  Length: 101
```

```
##
    Mode :character
                                Mode :character
                                                    Mode :character
##
##
##
##
##
   Do.you.have.Depression. Do.you.have.Anxiety. Do.you.have.Panic.attack.
##
    Length: 101
                             Length: 101
                                                   Length: 101
##
    Class : character
                             Class : character
                                                   Class : character
##
    Mode :character
                             Mode :character
                                                   Mode :character
##
##
##
##
##
    Did.you.seek.any.specialist.for.a.treatment.
##
    Length: 101
    Class :character
##
##
    Mode :character
##
##
##
##
#pander(studentData_df)
```

Class : character

Class :character

### **Data Cleaning**

Class :character

We thoroughly examined the data to ensure that no noisy or missing data values were present. More specifically, we ensured that no negative values existed in our numerical attributes (age, year of study, CGPA), and that no missing values were present in any tuple. Upon examination, only one column was found to have any missing data: Age. Though it is not particularly robust, we opted to fill in this missing data with the mean of the column.

```
studentData_df$Age[is.na(studentData_df$Age)] <- floor(mean(studentData_df$Age, na.rm=TRUE))
```

In order to further facilitate our analysis of this data, we deemed it appropriate to drop the Timestamp column, as it provided no relevant information. It seemed to be more of a vanity metric for the circumstances in which the data was originally acquired.

```
studentData_df = subset(studentData_df, select = -c(1))
summary(studentData_df)
```

```
##
    Choose.your.gender
                                        What.is.your.course.
                             Age
##
   Length: 101
                       Min.
                             :18.00
                                        Length: 101
##
    Class :character
                       1st Qu.:18.00
                                        Class : character
##
    Mode :character
                       Median :19.00
                                        Mode :character
##
                        Mean
                               :20.52
                       3rd Qu.:23.00
##
                       Max.
                               :24.00
##
##
  Your.current.year.of.Study What.is.your.CGPA. Marital.status
##
  Length: 101
                                Length: 101
                                                   Length: 101
   Class : character
                                Class : character
                                                   Class : character
##
```

```
Mode
         :character
                                     :character
##
                                Mode
                                                    Mode : character
##
##
##
##
    Do.you.have.Depression. Do.you.have.Anxiety. Do.you.have.Panic.attack.
##
   Length: 101
                                                   Length: 101
                             Length: 101
    Class : character
                             Class : character
##
                                                   Class : character
    Mode :character
                             Mode :character
##
                                                   Mode : character
##
##
##
##
    Did.you.seek.any.specialist.for.a.treatment.
##
   Length: 101
##
   Class : character
   Mode :character
##
##
##
##
```

Due to the method in which this survey was conducted, some features of the data were able to be entered in an non-deterministic manner. The column Your.current.year.of.Study exhibits this the most, as random capitalization in the responses creates several different "bins" of responses for data that is otherwise meant to be the same. To fix this, we elected to simple cast all characters in this column to an uppercase state to remove any ambiguity.

studentData\_df\$Your.current.year.of.Study <- toupper(studentData\_df\$Your.current.year.of.Study)

#### **Data Wrangling**

#### Renaming Columns

For both exploratory analysis and display purposes, data wrangling consisted of making the data look more presentable and easier to parse. Through some minor idiosyncrasies of the method through which this data was obtained, lengthy and oddly formatted names currently index most of our columns. To resolve this, we gave each column a less-verbose name that still unambiguously indicates what data said column held.

- Choose.your.gender becomes simply Gender
- Age age of the participant at the time of survey completion
- What.is.your.course. is simplified into Major
- Your.current.year.of.Study is shortened to Year
- What.is.your.CGPA. is shortened to GPA

The following attributes have been shortened to just their respective affects. It is assumed that the names are preceded by either "is" or "has" before each condition (i.e. "has Anxiety).

- Marital.status becomes Married
- Do.you.have.Depression. becomes Depressed
- Do.you.have.Anxiety. becomes Anxiety
- Do.you.have.Panic.attack. becomes Panic
- Did.you.seek.any.specialist.for.a.treatment. becomes Treatment

With all of these, we were seeking the simplicity of single word names.

```
colnames(studentData df)[colnames(studentData df) ==
                           'Choose.your.gender'] <- 'Gender'
colnames(studentData df)[colnames(studentData df) ==
                           'What.is.your.course.'] <- 'Major'
colnames(studentData df)[colnames(studentData df) ==
                           'Your.current.year.of.Study'] <- 'Year'
colnames(studentData_df)[colnames(studentData_df) ==
                           'What.is.your.CGPA.'] <- 'GPA'
colnames(studentData_df)[colnames(studentData_df) ==
                           'Marital.status'] <- 'Married'
colnames(studentData_df)[colnames(studentData_df) ==
                           'Do.you.have.Depression.'] <- 'Depressed'
colnames(studentData_df)[colnames(studentData_df) ==
                           'Do.you.have.Anxiety.'] <- 'Anxiety'
colnames(studentData_df)[colnames(studentData_df) ==
                           'Do.you.have.Panic.attack.'] <- 'Panic'
colnames(studentData_df)[colnames(studentData_df) ==
                           'Did.you.seek.any.specialist.for.a.treatment.'] <- 'Treatment'
```

#### Categorizing Ambiguous Data

Given that we are looking for various correlations between Science, Technology, Engineering, and Mathematics (STEM) majors, and the various mental health issues that might affect them, the detailed knowledge of what major a student is enrolled in doesn't interest us as data - we only care whether or not it's considered STEM. Unfortunately, there is no easy algorithmic way to do this; we had considered using various "sounds-like" libraries and methods, but decided for a dataset this small that it would be best to just manually build a new column by hand. This included looking for what the responses from the Major column in our dataset correlated to and simply populating a new column with "Yes" or "No" before adding it to our dataset. We added this column as STEM. We have included the first 5 rows as a sample of the data set, and the full data set is listed in Appendix A at the end of this document.

Table 1: Table continues below

Gender	Age	Major	Year	GPA	Married	Depressed
Female	18	Engineering	YEAR 1	3.00 - 3.49	No	Yes
Male	21	Islamic education	YEAR $2$	3.00 - 3.49	No	No
Male	19	$\operatorname{BIT}$	YEAR 1	3.00 - 3.49	No	Yes

Gender	Age	Major	Year	GPA	Married	Depressed
Female	22	Laws	YEAR 3	3.00 - 3.49	Yes	Yes
Male	23	Mathemathics	YEAR 4	3.00 - 3.49	No	No

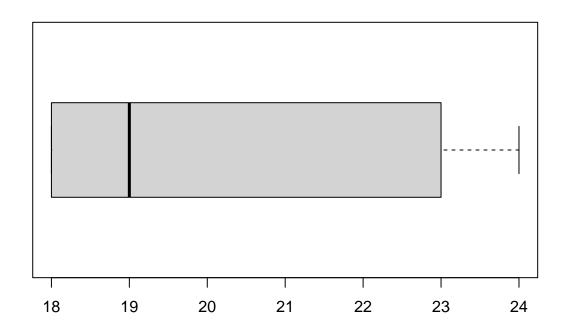
Anxiety	Panic	Treatment	STEM
No	Yes	No	Yes
Yes	No	No	No
Yes	Yes	No	Yes
No	No	No	No
No	No	No	Yes

# **Exploratory Data Analysis**

## Age Distribution

The following boxplot shows the distribution of each participant's age. The youngest participants are 18 years old, and the oldest are 24 years old. Given that our mean is 20.52, we can be assured that our data is fairly representative of the average college student.

boxplot(studentData\_df\$Age, horizontal = TRUE)



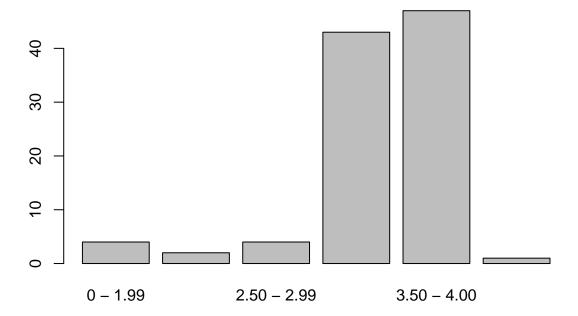
#### summary(studentData\_df\$Age)

```
## Min. 1st Qu. Median Mean 3rd Qu. Max.
## 18.00 18.00 19.00 20.52 23.00 24.00
```

### **CGPA** Distribution

The following histogram shows the frequency of each CGPA interval for participants. Most participants have a CGPA within the range of 3.00 - 3.49.

```
gpaTable <- table(studentData_df$GPA)
barplot(gpaTable)</pre>
```



## Analysis of Key Data Characteristics

#### Distribution of Gender and Major

Below we can see two key aspects of our data set displayed in a contingency table. Represented in the rows is the distribution of male and female students, with the columns showing how many of each are majoring in a STEM field or not. With the "Sum" features of this table, we can see that nearly 75 percent of our student sample is female, with just over two thirds of all students majoring in some STEM field. This could indicate that we are not accurately representing both males and females with this data, and this will be kept in mind moving forward.

Table 3: STEM Majors by Gender

	No	Yes	Sum
Female	28	47	75
Male	5	21	26
Sum	33	68	101

The data we're working with indicates that 62.67% of all female students are in STEM, while 80.77% of all male students represented are majoring in some STEM field.

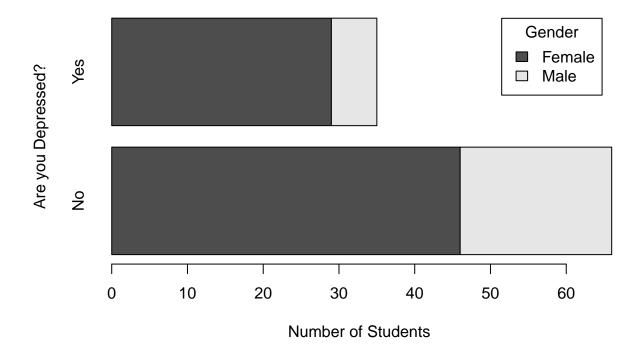
#### Mental Health by Gender and Major

Given that our data is so heavily skewed in samples toward females, we decided to analyze how self-reported depression stacks up by both gender and major. Below is a contingency table that represents gender by row, and whether or not the student reported as depressed by column.

Table 4: Depressed Students by Gender

	No	Yes	Sum
Female	46	29	75
Male	20	6	26
Sum	66	35	101

The data we're working with indicates that 38.67% of all female students reported as depressed, while 23.08% of all male students indicated in the survey that they are depressed. This data is graphically represented in a bar chart below.

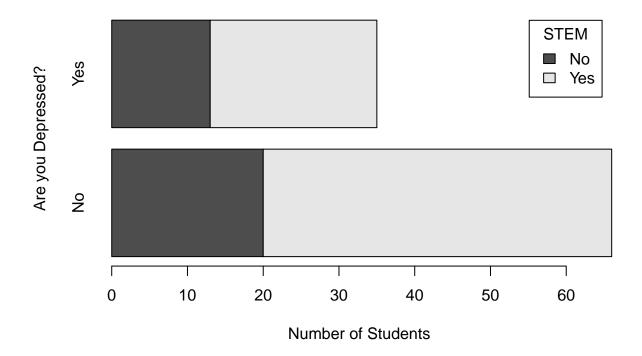


Perhaps a more relevant exploration of this data, especially for the purposes of this survey, would be the relationship between being a STEM major and reporting as Depressed. Our table below once again represents whether a student is depressed or not by column, but this time the rows give us the quality of being in a STEM related major.

Table 5: Depressed Students by Major

	No	Yes	Sum
No	20	13	33
Yes	46	22	68
Sum	66	35	101

Of all students surveyed, we can ascertain that 34.65% of them report having depression. In finer granularity, depression among STEM majors is at 32.35%, and 39.39% in non-STEM majors.

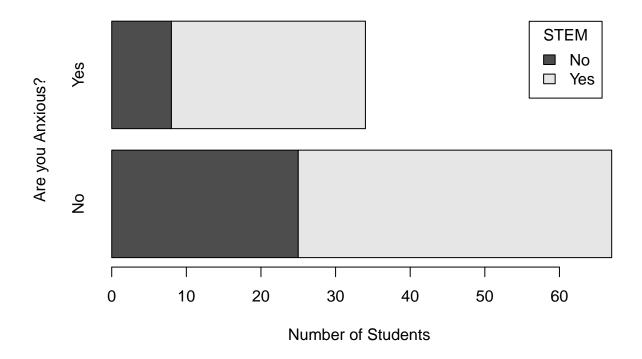


We can also look at mental health affects outside of depression. For instance, this table gives us STEM major by row and whether a student reported having anxiety attacks or not by column.

Table 6: Anxious Students by Major

	No	Yes	Sum
No	25	8	33
Yes	42	26	68
Sum	67	34	101

Anxiety is less prevalent among the students who were polled for this data set, but only just at 33.66% Though anxiety among STEM majors is significantly higher at 38.24%, and much lower at 24.24% in non-STEM majors.



In all of these fields, the bar plots serve to give a sense of scale in the sampled populations, and gives a rough idea of what ratio one could expect from a given sample analysis.

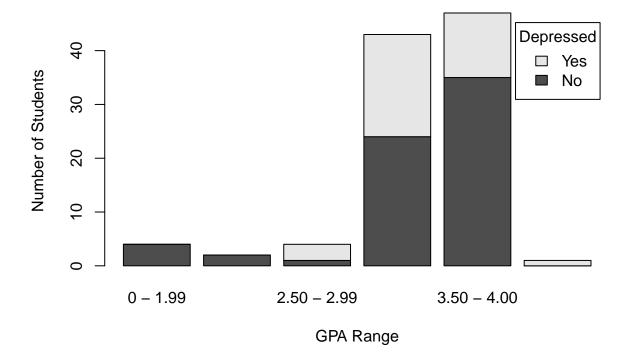
#### Relating GPA to Depression

The rows of the following table contain several ranges of GPA that represents the sample, and the columns once again show whether or not a student is depressed.

Table 7: Depressed Students by GPA

	No	Yes	Sum
0 - 1.99	4	0	4
2.00 - 2.49	2	0	2
2.50 - 2.99	1	3	4
3.00 - 3.49	24	19	43
3.50 - 4.00	35	12	47
3.50 - 4.00	0	1	1
**Sum**	66	35	101

```
barplot (t(gpaDepressiontable),
    horiz=FALSE, ylab = "Number of Students", xlab = "GPA Range",
    legend.text = TRUE, args.legend = list(title = "Depressed"))
```

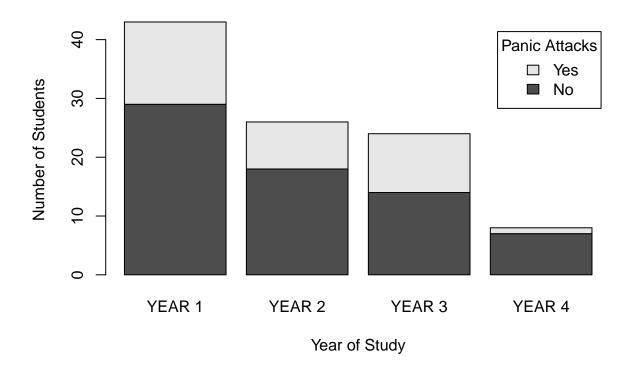


#### Progression in College and Panic Attacks

To determine if a relationship exists between the features Year and Panic, we analyzed the following materials:

Table 8: Panic in Students by Year

	No	Yes	Sum
YEAR 1	29	14	43
YEAR 2	18	8	26
YEAR 3	14	10	24
YEAR 4	7	1	8
Sum	68	33	101



## Conclusions and Objectives

Upon extensive analysis of the dataset, we determined that the features GPA, Major, and Year will be most important in predicting the presence of mental health disorders. There are no obvious outliers or unexpected patterns in the dataset, but our analysis proved that a disproportionate amount of STEM majors and Females exist in the sample. The following hypotheses were drawn as a result:

- 1. A cause/effect relationship exists between a student's GPA and the prevalence of Depression
- 2. Students majoring in STEM fields are more likely to have anxiety than depression
- 3. The number of years in which a student has attended college does not have an effect on the prevalence of panic attacks

During the next phase of analysis, we plan on using statistical methods to determine the accuracy of our hypotheses. Machine learning algorithms do not appear necessary at the moment; we may decide to integrate machine learning techniques as we proceed.

#### Appendix A - Data

Table 9: Mental Health in University Students (continued below)

Gender	Age	Major	Year	GPA	Married
Female	18	Engineering	YEAR 1	3.00 - 3.49	No
Male	21	Islamic education	YEAR $2$	3.00 - 3.49	No
Male	19	BIT	YEAR 1	3.00 - 3.49	No
Female	22	Laws	YEAR 3	3.00 - 3.49	Yes
Male	23	Mathemathics	YEAR $4$	3.00 - 3.49	No
Male	19	Engineering	YEAR $2$	3.50 - 4.00	No
Female	23	Pendidikan islam	YEAR $2$	3.50 - 4.00	Yes
Female	18	BCS	YEAR $1$	3.50 - 4.00	No
Female	19	Human Resources	YEAR $2$	2.50 - 2.99	No
Male	18	$\operatorname{Irkhs}$	YEAR 1	3.50 - 4.00	No
Female	20	Psychology	YEAR 1	3.50 - 4.00	No
Female	24	Engineering	YEAR $3$	3.50 - 4.00	Yes
Female	18	BCS	YEAR $1$	3.00 - 3.49	No
Male	19	Engineering	YEAR $1$	3.00 - 3.49	No
Female	18	KENMS	YEAR $2$	3.50 - 4.00	No
Male	24	BCS	YEAR $3$	3.50 - 4.00	No
Female	24	Accounting	YEAR 3	3.00 - 3.49	No
Female	24	$\operatorname{ENM}$	YEAR $4$	3.00 - 3.49	Yes
Female	20	BIT	YEAR $2$	3.50 - 4.00	No
Female	18	Marine science	YEAR $2$	3.50 - 4.00	Yes
Female	19	Engineering	YEAR 1	3.00 - 3.49	No
Female	18	KOE	YEAR $2$	3.00 - 3.49	No
Female	24	BCS	YEAR $1$	3.50 - 4.00	No
Female	24	Engineering	YEAR 1	3.00 - 3.49	No
Female	23	BCS	YEAR 3	3.50 - 4.00	No
Female	18	Banking Studies	YEAR 1	3.50 - 4.00	No
Female	19	Engineering	YEAR 1	3.50 - 4.00	No
Male	18	Engineering	YEAR $2$	3.00 - 3.49	Yes
Female	24	BIT	YEAR 3	3.50 - 4.00	Yes
Female	24	$\operatorname{BCS}$	YEAR $4$	3.50 - 4.00	No
Female	23	Business Administration	YEAR $2$	3.00 - 3.49	No
Male	18	$\operatorname{BCS}$	YEAR $2$	3.00 - 3.49	No
Male	19	$\operatorname{BCS}$	YEAR 1	3.50 - 4.00	No
Male	18	$\operatorname{BCS}$	YEAR $2$	3.50 - 4.00	Yes
Female	19	BIT	YEAR 1	3.00 - 3.49	No
Female	18	Engineering	YEAR 1	2.00 - 2.49	No
Female	18	Law	YEAR 3	3.00 - 3.49	No
Female	19	BIT	YEAR 1	2.50 - 2.99	No
Female	18	KIRKHS	YEAR 1	3.50 - 4.00	No
Female	24	Engineering	YEAR $2$	2.50 - 2.99	Yes
Female	24	BIT	YEAR 3	3.00 - 3.49	No
Female	22	Engineering	YEAR $4$	3.50 - 4.00	No
Female	20	Usuluddin	YEAR $2$	3.00 - 3.49	No
Male	20	BIT	YEAR 1	0 - 1.99	No
Male	23	$\mathrm{TAASL}$	YEAR $2$	3.50 - 4.00	No
Male	18	BCS	YEAR 1	3.50 - 4.00	No
Female	19	Engineering	YEAR 1	3.50 - 4.00	No

Gender	Age	Major	Year	GPA	Married
Female	18	Engine	YEAR $4$	3.50 - 4.00	No
Male	24	BCS	YEAR $2$	3.00 - 3.49	No
Female	24	BCS	YEAR 3	3.50 - 4.00	No
Female	23	ALA	YEAR 1	2.50 - 2.99	Yes
Female	18	BCS	YEAR $2$	3.50 - 4.00	No
Female	19	Biomedical science	YEAR 3	3.00 - 3.49	No
Female	20	koe	YEAR 3	3.00 - 3.49	Yes
Female	19	BCS	YEAR 1	3.50 - 4.00	No
Male	21	$\operatorname{BCS}$	YEAR 1	3.00 - 3.49	No
Male	23	Kirkhs	YEAR 3	3.50 - 4.00	No
Female	20	$\operatorname{BENL}$	YEAR 3	3.00 - 3.49	No
Female	18	BCS	YEAR 1	3.50 - 4.00	No
Female	$\overline{23}$	Benl	YEAR 1	3.00 - 3.49	No
Female	18	IT	YEAR 3	3.00 - 3.49	No
Female	19	BCS	YEAR 1	3.50 - 4.00	No
Female	18	CTS	YEAR 1	3.50 - 4.00	No
Female	$\frac{16}{24}$	engin	YEAR 1	3.50 - 4.00	No
Female	$\frac{24}{24}$	Engine	YEAR 1	3.50 - 4.00	No
Female	23	Econs	YEAR 1	3.50 - 4.00	No
Female	18	KOE	YEAR 3	3.00 - 3.49	No
Male		MHSC			
	19		YEAR 3	3.00 - 3.49	Yes
Female	18	Malcom	YEAR 1	3.50 - 4.00	No N-
Female	24	Kop	YEAR 4	3.00 - 3.49	No
Female	24	Biomedical science	YEAR 1	3.00 - 3.49	No
Female	18	Laws	YEAR 3	3.50 - 4.00	No
Female	19	BIT	YEAR 3	3.00 - 3.49	Yes
Male	18	Biomedical science	YEAR 1	0 - 1.99	No
Male	24	BIT	YEAR 3	3.50 - 4.00	No
Female	24	KOE	YEAR 1	3.50 - 4.00	No
Female	23	Engineering	YEAR 1	3.00 - 3.49	No
Female	18	Human Sciences	YEAR $2$	3.00 - 3.49	No
Female	19	Biotechnology	YEAR 3	0 - 1.99	No
Female	18	Engineering	YEAR $4$	3.50 - 4.00	No
Female	24	Communication	YEAR $2$	3.50 - 4.00	Yes
Female	24	Diploma Nursing	YEAR $2$	3.50 - 4.00	No
Female	19	Engineering	YEAR 1	3.00 - 3.49	No
Female	19	Pendidikan Islam	YEAR $2$	3.00 - 3.49	No
Male	23	Radiography	YEAR 1	3.00 - 3.49	No
Female	18	psychology	YEAR 1	3.50 - 4.00	No
Female	19	Figh fatwa	YEAR 3	3.00 - 3.49	No
Female	18	psychology	YEAR 1	3.50 - 4.00	No
Male	24	BIT	YEAR 1	3.00 - 3.49	No
Male	24	Engineering	YEAR 2	2.00 - 2.49	No
Female	23	DIPLOMA TESL	YEAR 3	3.50 - 4.00	No
Male	18	Koe	YEAR 2	3.00 - 3.49	No
Female	19	KOE	YEAR 2	3.00 - 3.49	Yes
Female	18	BENL	YEAR 1	3.00 - 3.49	No
Female	$\frac{16}{24}$	Figh	YEAR 3	0 - 1.99	No
Female	18	Islamic Education	YEAR 1	3.50 - 4.00	No
Female	21	BCS		3.50 - 4.00	No
			YEAR 1		
Male	18	Engineering	YEAR 2	3.00 - 3.49	No V
Female	19	Nursing	YEAR 3	3.50 - 4.00	Yes

Gender	Age	Major	Year	GPA	Married
Female	23	Pendidikan Islam	YEAR 4	3.50 - 4.00	No
Male	20	Biomedical science	YEAR $2$	3.00 - 3.49	No

Depressed	Anxiety	Panic	Treatment	STEM
Yes	No	Yes	No	Yes
No	Yes	No	No	No
Yes	Yes	Yes	No	Yes
Yes	No	No	No	No
No	No	No	No	Yes
No	No	Yes	No	Yes
Yes	No	Yes	No	No
No	Yes	No	No	Yes
No	No	No	No	No
No	Yes	Yes	No	No
No	No	No	No	Yes
Yes	No	No	No	Yes
Yes	No	No	No	Yes
No	No	No	No	Yes
No	Yes	No	No	No
No	No	No	No	Yes
No	No	No	No	No
Yes	Yes	Yes	No	No
No	Yes	No	No	Yes
Yes	Yes	Yes	No	Yes
No	No	Yes	No	Yes
No	No	No	No	Yes
No	No	No	No	Yes
No	No	No	No	Yes
Yes	Yes	Yes	No	Yes
No	No	No	No	No
No	No	No	No	Yes
Yes	Yes	No	No	Yes
Yes	Yes	Yes	Yes	Yes
No	No	No	No	Yes
No	No	No	No	No
No	No	No	No	Yes
No	No	Yes	No	Yes
Yes	Yes	No	Yes	Yes
Yes	Yes	Yes	No	Yes
No	No	No	No	Yes
Yes	Yes	No	No	No
Yes	Yes	Yes	No	Yes
No	No	No	No	No
Yes	No	Yes	Yes	Yes
No	Yes	No	No	Yes
No	No	No	No	Yes
Yes	No	No	No	No
No	No	No	No	Yes
No	No	Yes	No	No
No	Yes	Yes	No	Yes

Depressed	Anxiety	Panic	Treatment	STEM
No	Yes	No	No	Yes
No	No	No	No	Yes
Yes	No	No	No	Yes
No	No	Yes	No	Yes
Yes	No	Yes	Yes	No
No	Yes	No	No	Yes
No	No	No	No	Yes
Yes	Yes	Yes	No	Yes
Yes	No	Yes	Yes	Yes
No	No	No	No	Yes
No	No	No	No	No
Yes	Yes	No	No	No
No	No	No	No	Yes
No	No	No	No	No
No	No	Yes	No	Yes
No	No	No	No	Yes
No	No	Yes	No	Yes
No	No	Yes	No	Yes
No	No	No	No	Yes
Yes	Yes	No	No	No
No	Yes	No	No	Yes
Yes	No	Yes	No	No
Yes	No	No	No	No
No	Yes	No	No	Yes
No	No	No	No	Yes
No	No	Yes	No	No
Yes	No N-	No N-	No No	Yes
No N-	No V	No N-	No No	Yes
No	Yes	No	No	Yes
No	Yes	Yes	No	Yes
Yes	No	No	No	Yes
No	No	Yes	No	No
No	No	No	No	Yes
No	No	No	No	Yes
Yes	Yes	Yes	No	No
No	No	No	No	No
Yes	Yes	No	No	Yes
No	No	No	No	No
No	No	No	No	Yes
Yes	Yes	No	Yes	Yes
No	No	No	No	No
Yes	Yes	Yes	No	Yes
No	Yes	No	No	Yes
No	No	Yes	No	Yes
No	No	Yes	No	No
No	Yes	No	No	Yes
Yes	No	No	No	Yes
Yes	No	No	No	No
No	No	Yes	No	No
No	No	No	No	No
No	Yes	No	No	Yes
	Yes	No	No	Yes

Depressed	Anxiety	Panic	Treatment	STEM
Yes	No	Yes	No	No
No	No	No	No	No
No	No	No	No	Yes