Survery on Mental Health Across Programs of Study in University

Deanna King, Jim Moroney, Pooja Patel, Christopher Wilhite

2022-11-19

Packages

In order to better interpret the dataset, we utilize the `pander` package for table creation and manipulation. Likewise, we use `insert other packages` for `insert reasoning`.

```
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
- `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)
```

```
##    Timestamp         Choose.your.gender      Age         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
##                                        Mean   :20.53
##                                        3rd Qu.:23.00
##                                        Max.   :24.00
##                                        NA's   :1
##  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            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)
```

Data Cleaning

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 a simple 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 to what we were looking for and 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       Age           What.is.your.course.
##   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           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
##
##
##
```

Data Wrangling

Renaming Columns

Wrangling for the most part consisted of making the data look more presentable and easier to parse for our exploratory analysis and display purposes. Through some minor idiosyncrasies of the method through which this data was obtained, lengthy and oddly formatted names currently index most of our columns; for these reasons we gave the each column a less verbose name that still unambiguously indicated 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` - has been summarily shortened to `Year`
- `What.is.your.CGPA.` - similarly shortened to just `GPA`

The following attributes have been shortened to just their respective affects. It is assumed that the names are preceeded by, "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.'] <-
```

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'd 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`.

```
studentData_df['STEM'] <- c('Yes','No','Yes','No','Yes','Yes','No','Yes','No','No',
                            'Yes','Yes','Yes','Yes','No','Yes','No','No','Yes','Yes',
                            'Yes','Yes','Yes','Yes','Yes','No','Yes','Yes','Yes','Yes',
                            'No','Yes','Yes','Yes','Yes','Yes','No','Yes','No','Yes',
                            'Yes','Yes','No','Yes','No','Yes','Yes','Yes','Yes','Yes',
                            'No','Yes','Yes','Yes','Yes','Yes','No','No','Yes','No',
                            'Yes','Yes','Yes','Yes','Yes','No','Yes','No','No','Yes',
                            'Yes','No','Yes','Yes','Yes','Yes','Yes','No','Yes','Yes',
                            'No','No','Yes','No','Yes','Yes','No','Yes','Yes','Yes',
                            'No','Yes','Yes','No','No','No','Yes','Yes','No','No',
                            'Yes')
pander(studentData_df)
```

Table 1: Table continues 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	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	ENM	year 4	3.00 - 3.49	Yes
Female	20	BIT	Year 2	3.50 - 4.00	No

Gender	Age	Major	Year	GPA	Married
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	BCS	year 4	3.50 - 4.00	No
Female	23	Business Administration	Year 2	3.00 - 3.49	No
Male	18	BCS	year 2	3.00 - 3.49	No
Male	19	BCS	year 1	3.50 - 4.00	No
Male	18	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	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
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	BCS	year 1	3.00 - 3.49	No
Male	23	Kirkhs	Year 3	3.50 - 4.00	No
Female	20	BENL	Year 3	3.00 - 3.49	No
Female	18	BCS	year 1	3.50 - 4.00	No
Female	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	24	engin	year 1	3.50 - 4.00	No
Female	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	19	MHSC	Year 3	3.00 - 3.49	Yes
Female	18	Malcom	year 1	3.50 - 4.00	No
Female	24	Kop	year 4	3.00 - 3.49	No
Female	24	Biomedical science	year 1	3.00 - 3.49	No

Gender	Age	Major	Year	GPA	Married
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	Fiqh 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	24	Fiqh	Year 3	0 - 1.99	No
Female	18	Islamic Education	year 1	3.50 - 4.00	No
Female	21	BCS	year 1	3.50 - 4.00	No
Male	18	Engineering	Year 2	3.00 - 3.49	No
Female	19	Nursing	Year 3	3.50 - 4.00	Yes
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

Depressed	Anxiety	Panic	Treatment	STEM
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
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

Depressed	Anxiety	Panic	Treatment	STEM
No	No	No	No	Yes
No	No	Yes	No	No
Yes	No	No	No	Yes
No	No	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	Yes	No	No	Yes
Yes	No	Yes	No	No
No	No	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. 20.52.

```
boxplot(studentData_df$Age, horizontal = TRUE)
```
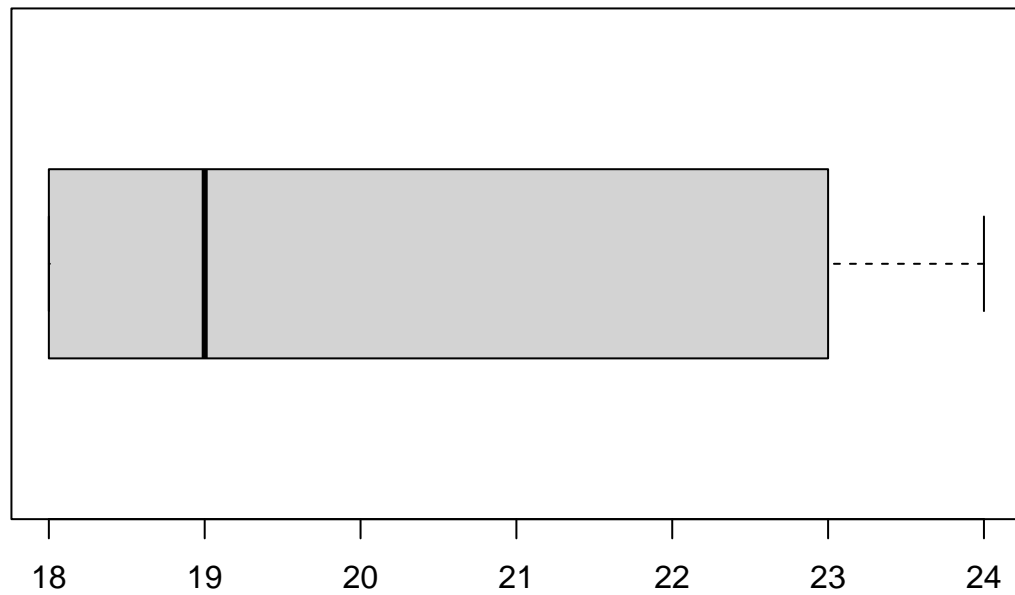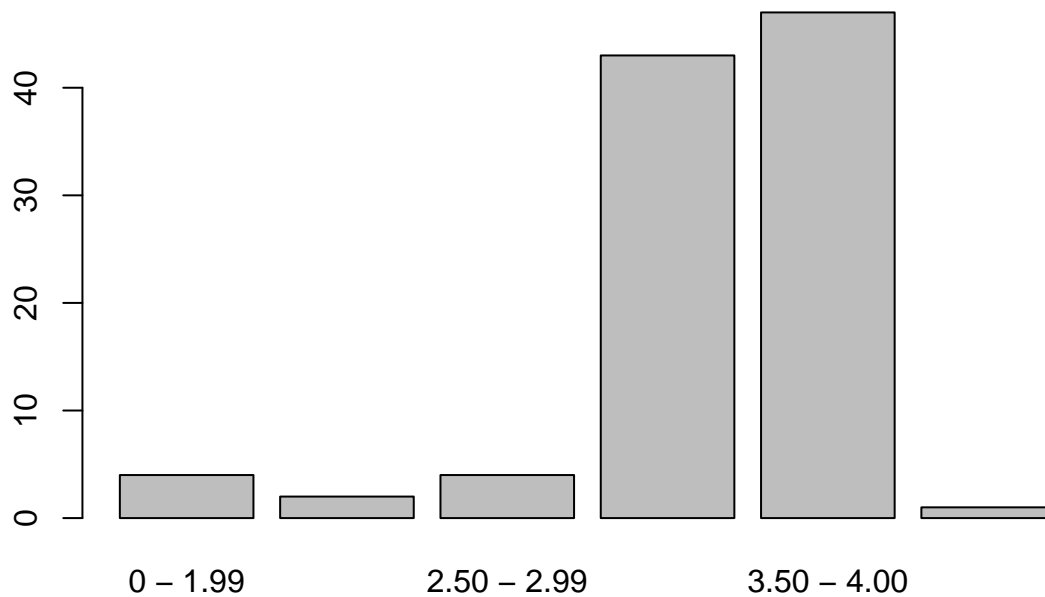
```
summary(studentData_df$Age)
```

```
##    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 (add mode!).

```
gpaTable <- table(studentData_df$GPA)
barplot(gpaTable)
```

Something with the STEM major vs mental health (depression and anxiety and panic)

Jim- i can't do this until data has been wrangled. i think one of those matrix things could work for this one

```
addmargins(with(studentData_df, table(studentData_df$Gender, studentData_df$STEM)))
```

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

```
addmargins(with(studentData_df, table(studentData_df$Depressed, studentData_df$STEM)))
```

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

Something to do with gpa vs depression

Jim - same thing as above, maybe a normal plot? stacked box plot, y depressed x gpa range

Something to do with year in school vs panic attacks

Jim - same thing again, i definitely think a normal plot for this one

Conclusions

Should answer the following questions:

Are there any unexpected patterns or relationships in your data? Does there appear to be any cause/effect phenomena? Can you suggest hypotheses for these relationships? Which variables are important? Does the data contain any anomalies or outliers? What assumptions are you making about the data, and can you verify these speculations?

Jim - i can write this after all of the data has been wrangled and charts have been made :) thank you!! <3