

Exploratory Data Analysis (EDA)

Student Depression



Group Assignment

Statistical Computing with R - CM2062

Group Members

D/DBA/24/0024 – Visura Rodrigo

D/DBA/24/0028 – Chamathka Gaurammanna

D/DBA/24/0030 – Kaushal Senavirathna

D/DBA/24/0034 – Sachila Hansana

Table of Contents

INTRODUCTION	3
PROBLEM STATEMENT	4
WHAT IS EXPLORATORY DATA ANALYSIS?	5
EXPLORATORY DATA ANALYSIS	6
1. Basic Summary	6
2. Gender vs Depression	7
3. Academic Pressure vs Depression	8
4. CGPA (Academic Performance) vs Depression	9
5. Study Satisfaction vs Depression	10
6. Sleep Duration vs Depression	11
7. Dietary Habits vs Depression	12
8. Suicidal Thoughts vs Depression	13
9. Study Hours vs Depression	14
10. Financial Stress vs Depression	15
11. Family History of Mental illness vs Depression	16
CONCLUSION	17
FULL R CODE	19
DATA SET	25

INTRODUCTION

Mental health can significantly impact students' academic performance, social life and overall wellbeing. We chose to analyze dataset title "Student depression dataset: Analyzing mental health trends and predictors among students" got from Kaggle.com website. The latest set offers valuable factors influencing depression among students.

The original dataset consists of 18 variables and covers a broad spectrum of student related information including,

- Demographic data
- Academic indicators
- lifestyle habits
- Work-related stressors
- Mental health history

For a more accurate and better comparison we concluded to consider only students within a specific age range. So, we filtered the dataset including only those students enrolled in "Class 12".

The main goal of our analysis is to identify the key patterns and potential predictors associated with depression among Class 12 students. By utilizing the statistical and visualization capabilities of the R programming language, we expect to derive meaningful insights that could contribute to early intervention strategies and better mental health support systems within educational institutions.

This report will cover detailed analyses of variables such as,

- Academic pressure
- Sleep duration
- Financial stress
- Family history of mental illness

to understand their potential association with the presence or absence of depression, as denoted by the "Depression_Status" variable. Our findings aim to support data-driven decision-making in both academic and psychological contexts, emphasizing the importance of student well-being as a fundamental pillar of academic success.

PROBLEM STATEMENT

In today's fast-moving and competitive academic world, **mental health issues among students** are becoming increasingly common. One of the most concerning mental health problems faced by students is **depression**.

In our project, we focused on analyzing the "Student Depression Dataset" to understand the key factors that might be contributing to depression among **Class 12 students**.

We noticed that:

- Students often hide their mental struggles due to fear of judgment.
- External pressures like grades, expectations, and financial problems can heavily impact mental health.
- Lifestyle choices such as poor sleep habits and unhealthy diets also play a major role.

By using **Exploratory Data Analysis (EDA)** with **R programming**, we aim to **find patterns** and **predictors** related to depression.

WHAT IS EXPLORATORY DATA ANALYSIS?

Exploratory Data Analysis (EDA) is the process of **investigating datasets** to **discover patterns, spot anomalies, test hypotheses, and check assumptions** using summary statistics and graphical representations.

EDA helps us to **make sense of the data** by asking questions like:

- What does the data tell us?
- Are there any strange patterns or unexpected results?
- Which variables seem related to each other?
- Are there missing values or errors?

In our project, EDA helped us:

- **Summarize the data**
- **Visualize patterns**
- **Identify important factors**

What We Actually Did During Our EDA:

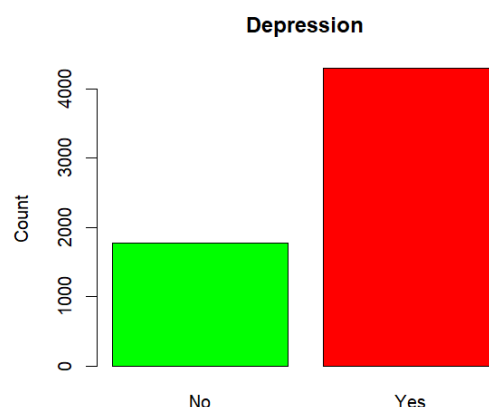
- Cleaned and filtered the data to focus only on **Class 12** students.
- Looked at how depression varies by **gender, sleep habits, financial stress, and diet**.
- Used **plots and summaries** to make it easier to spot important patterns.
- Interpreted the graphs in a simple way to explain **how different factors may increase depression risk**.

EXPLORATORY DATA ANALYSIS

Basic Summary

```
> summary(class12$Age)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 18.00  18.00  19.00  20.13  20.00   58.00
> summary(class12$`Academic Pressure`)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.000  3.000  3.000  3.359  4.000   5.000
> summary(class12$`Work Pressure`)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.000000 0.000000 0.000000 0.001974 0.000000 5.000000
> summary(class12$CGPA)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.000  6.260  7.750  7.594  8.750   10.000
> summary(class12$`Study Satisfaction`)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.000  2.000  3.000  3.078  4.000   5.000
> summary(class12$`Work/Study Hours`)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.000  5.000  8.000  7.278  10.000   12.000
> summary(class12$`Financial Stress`)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 1.000  2.000  3.000  3.279  5.000   5.000
```

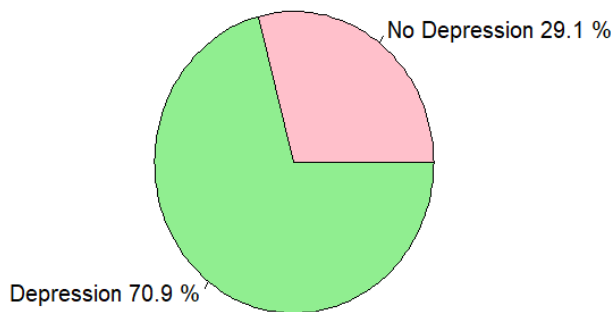
This summary will help us with the comparisons we make later in this report.



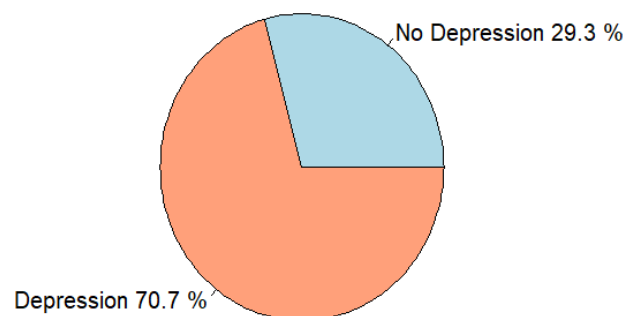
This displays the number of students with and without depression. We can see that the number of students suffering from depression is high. Let's see what the reasons are for that.

Gender vs Depression

Depression among Female Students



Depression among Male Students



Compared to male students, female students show higher rates of depression.

Female Depression 70.9%

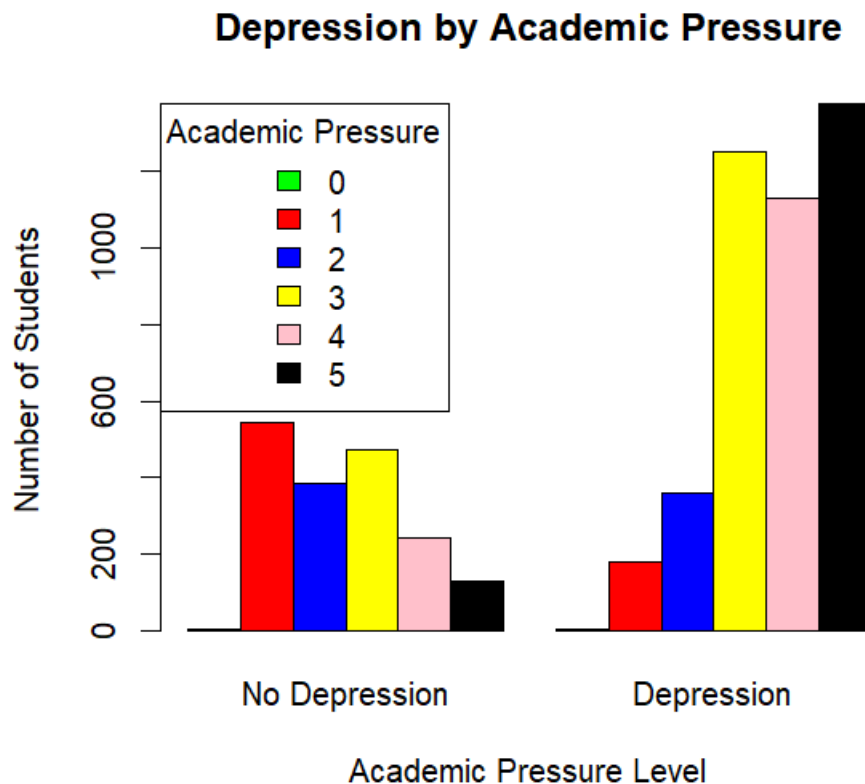
Male Depression 70.7%

The data clearly shows that female students experience a higher rate of depression (70.9%) compared to male students (70.7%). While the difference might seem small but in a larger student population, this could represent a significant number of individuals.

This data highlights a gender difference, it doesn't explain why this difference exists. Some factors could be

- Societal expectations and gender roles.
- Experiences of discrimination or inequality.
- Biological or hormonal factors.

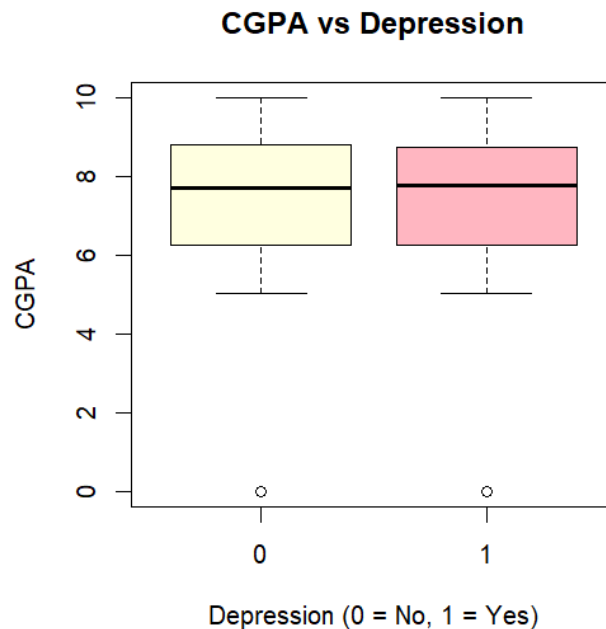
Academic Pressure vs Depression



At lower academic pressure levels, a larger number of students report "No Depression" compared to those with higher pressure. But at higher academic pressure levels, we see a substantial increase in the number of students reporting "Depression."

This data shows that high levels of academic pressure make students more likely to suffer from depression.

CGPA (Academic Performance) vs Depression

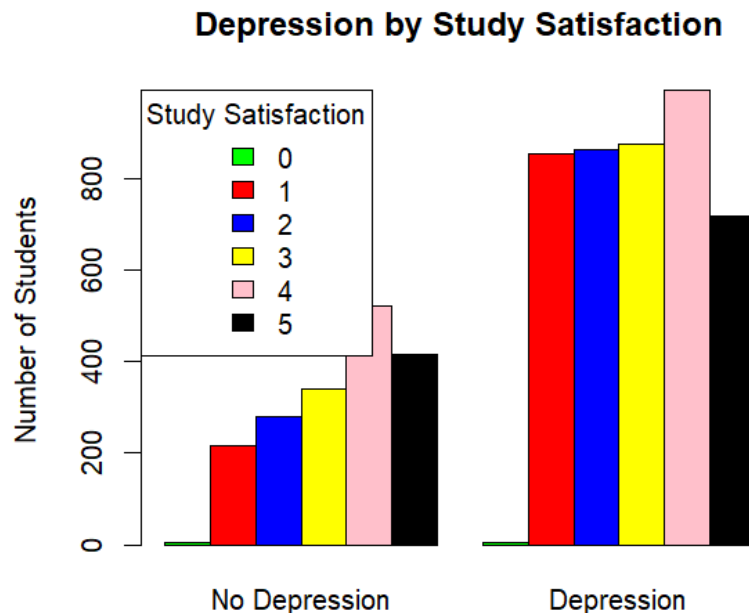


```
> boxplot_stats
class12$Depression: 0
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.000  6.250   7.700   7.597  8.810  10.000
-----
class12$Depression: 1
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.000  6.270   7.770   7.593  8.740  10.000
```

This compares the distribution of CGPA for students who do not report depression and those who do.

CGPA scores are normally distributed but slightly left-skewed. Looking at the mean values, there are no major differences in them. But the sample size is large. So, Depressed students have a lower CGPA on mean compared to non-depressed students. So, we can say that depression affects academic performance.

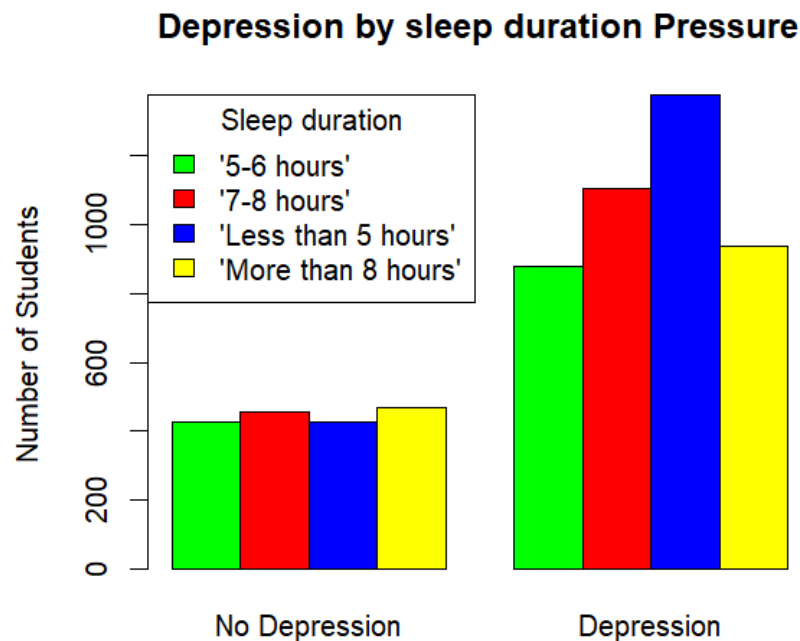
Study Satisfaction vs Depression



This displays the count of students at different levels of study satisfaction (0 to 5) for both those without reported depression and those with reported depression.

It appears that higher academic satisfaction is associated with a greater number of students reporting "no depression," whereas lower academic satisfaction is associated with a greater number of students reporting "depression."

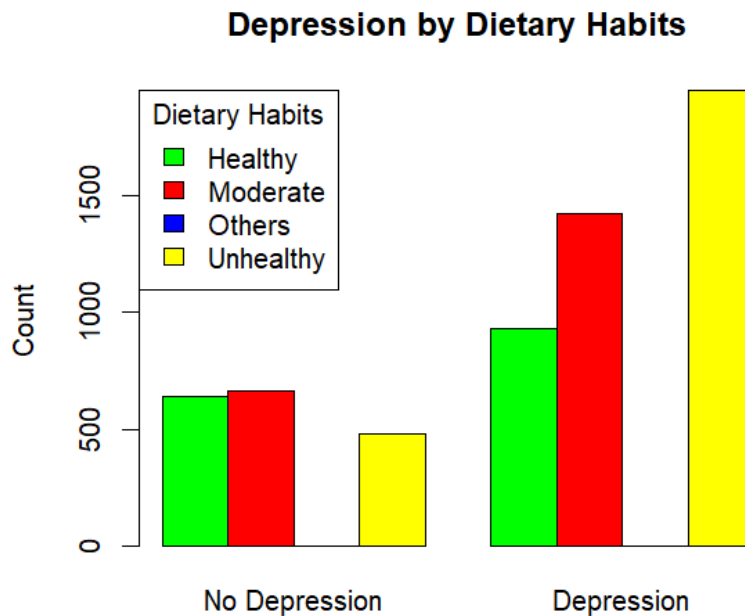
Sleep Duration vs Depression



This compares the distribution of reported sleep hours for students who do not report depression and those who do.

The analysis highlights the importance of adequate sleep time (about 7-8 hours) for potential mental well-being. Both very short and very long sleep durations appeared to be associated with an increased number of students reporting depression.

Dietary Habits vs Depression

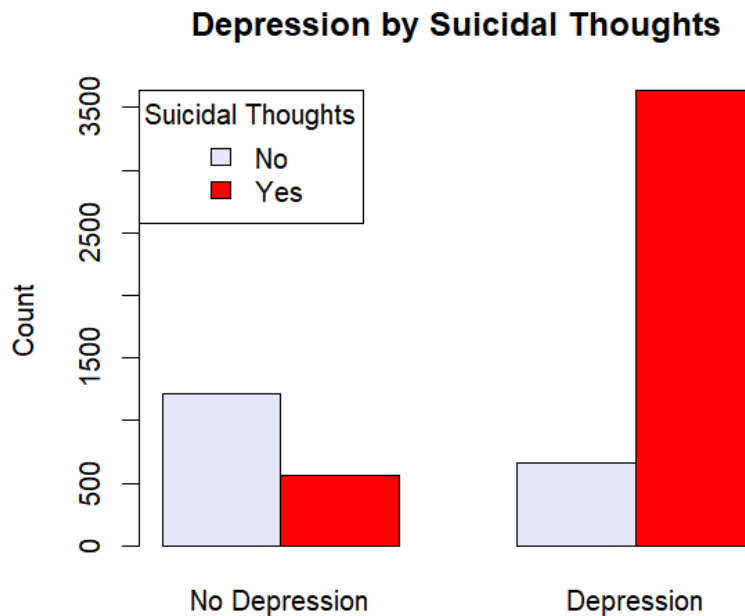


This compares the number of students with different dietary habits for those who do not report depression and those who do.

There is a strong link between "unhealthy" eating habits and a high prevalence of depression. The number of students following an unhealthy diet is dramatically higher in the "depression" group compared to the "non-depression" group.

Students with "healthy" eating habits had a higher representation in the "no depression" category compared to the "depression" category, indicating a potential protective effect of a healthy diet.

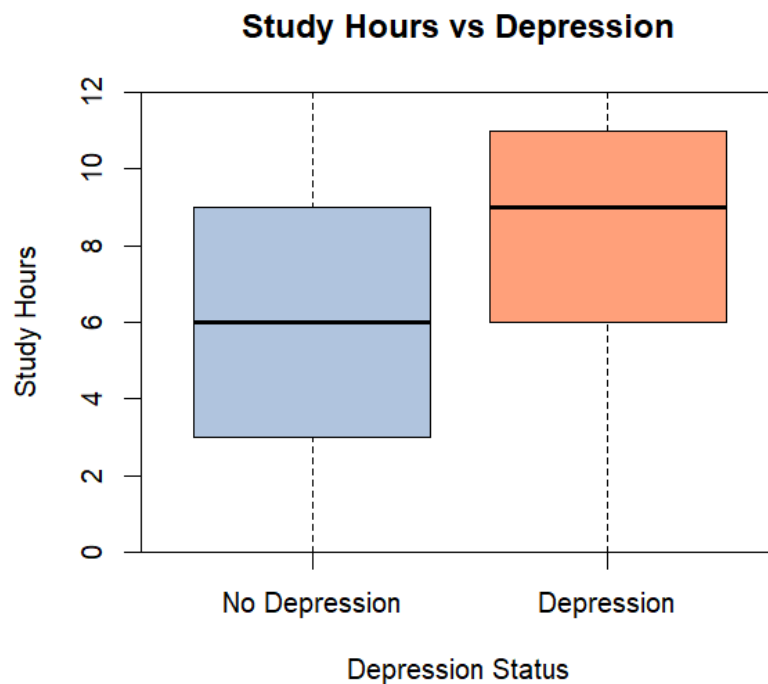
Suicidal Thoughts vs Depression



This compares the count of students who have and have not experienced suicidal thoughts, categorized by their reported depression status.

There is a very strong positive association between depression and suicidal ideation. Students reporting depression were significantly more likely to experience suicidal thoughts.

Study Hours vs Depression



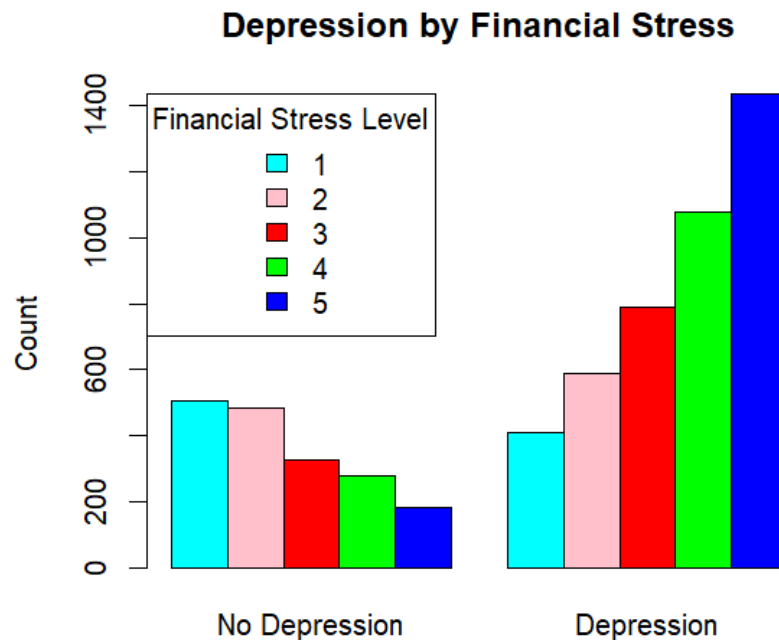
```
> by(class12$`Work/Study Hours`, class12$Depression, summary)
class12$Depression: 0
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.000  3.000   6.000   6.108  9.000  12.000
-----
class12$Depression: 1
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.000  6.000   9.000   7.762 11.000  12.000
```

This compares the distribution of reported study hours per week for students who do not report depression and those who do.

Provided statistics suggest a tendency for students who report depression to study more hours per week compared to those who do not report depression.

Depression can sometimes lead to difficulty concentrating and paying attention, causing students to spend more time studying to engage in academic activities. Students experiencing depression might feel increased pressure to perform academically, leading them to study more.

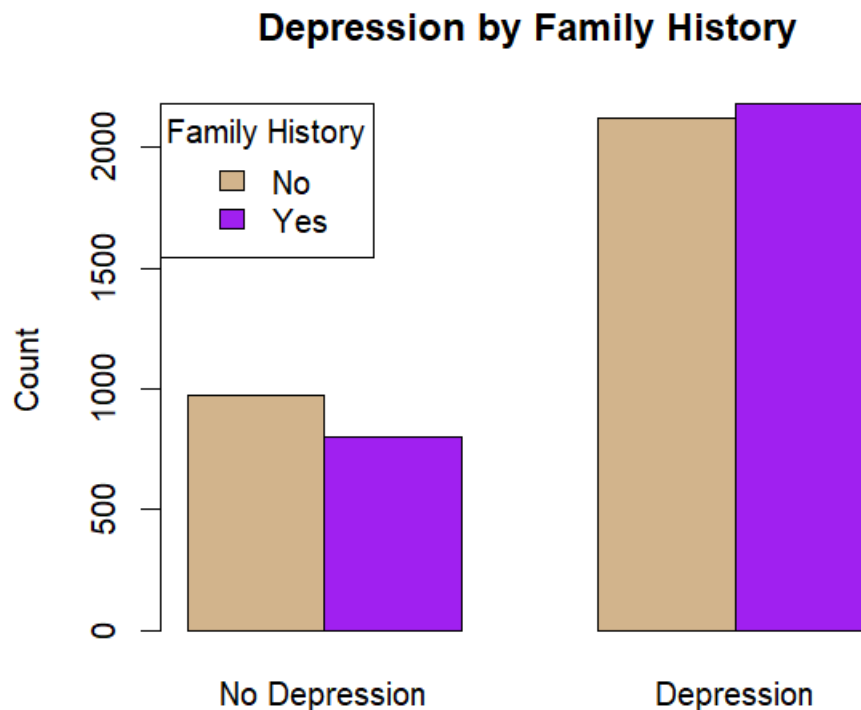
Financial Stress vs Depression



This displays the count of students at different levels of reported financial stress for both those without reported depression and those with reported depression.

There is a strong positive correlation between financial stress levels and the likelihood of experiencing depression. As the level of financial stress increases, so does the number of students reporting depression.

Family History of Mental illness vs Depression



This displays the count of students' family history of mental illness for both those without reported depression and those with reported depression.

There is a positive association between a family history of mental illness and the likelihood of experiencing depression. Students with a family history of mental illness are more likely to report depression compared to those who do not report depression.

CONCLUSION

This exploratory data analysis has indicated several important factors influencing depression in students. These findings will help create better support for students. Here's a summary of what we found in the Student Depression dataset:

- **How common is depression?**
 - We found that many students have depression. This means taking action and getting help is critical.
- **Gender and Depression**
 - Slightly more female students (70.9%) have depression compared to male students (70.7%). Although the difference is small, it tells us that gender may play a role in student mental health.
- **Pressure from School**
 - When students feel a lot of pressure from school, they are more likely to suffer from depression. School stress is a big problem.
- **Grades and Depression**
 - Depressed students have slightly lower grades. This suggests that depression may make it difficult to do well in school.
- **Enjoying Studies**
 - If students are not happy with their studies, they are likely to become depressed. Feeling good about studying is important for mental health.
- **Importance of sleep**
 - Not getting enough sleep (less than 5 hours) or sleeping too much (more than 8 hours) is linked to more depression. Getting 7-8 hours of sleep seems best.

- **Diet and Mood**
 - Eating unhealthy foods is linked to higher rates of depression. Eating healthy can help protect against depression.
- **Suicidal Thoughts**
 - There is a strong connection between depression and suicidal thoughts. Depressed students are more likely to have suicidal thoughts.
- **Hours spent studying**
 - Depressed students often study more. This may be because they have trouble concentrating or feel pressured to do well.
- **Worry about money**
 - When students worry a lot about money, they are more likely to become depressed. Financial pressure is a big factor.
- **Family History**
 - Students with family members suffering from mental illness are also more likely to suffer from depression.

Overall, this study suggests that many factors can contribute to depression in students. It is important for schools and others to support students' mental health by looking at their schoolwork, lifestyle and financial circumstances!

FULL R CODE

```
library(readr)

student_depression_dataset <- read_csv("C:/Users/a12u/Downloads/archive
(1)/student_depression_dataset.csv")

View(student_depression_dataset)


class12 <- subset(student_depression_dataset, Degree == "Class 12")


# Basic summary

summary(class12$Age)

summary(class12$`Academic Pressure` )

summary(class12$`Work Pressure` )

summary(class12$CGPA)

summary(class12$`Study Satisfaction` )

summary(class12$`Sleep Duration` )

summary(class12$`Work/Study Hours` )

summary(class12$`Financial Stress` )


#Depression

depress <- table(class12$Depression)

depresslable = c("No","Yes")

barplot(depress,beside = TRUE, col = c("green", "red" ),main = "Depression", ylab =
"Count", names.arg = depresslable)
```

```
# Gender vs Depression
```

```
gender_depression <- table(class12$Gender, class12$Depression)
```

```
y=c("No Depression", "Depression")
```

```
# For Males
```

```
male_data <- gender_depression["Male", ]
```

```
male_labels <- paste(y, round(100 * male_data/sum(male_data), 1), "%")
```

```
pie(male_data,
```

```
  labels = male_labels,
```

```
  col = c("lightblue", "lightsalmon"),
```

```
  main = "Depression among Male Students")
```

```
# For Females
```

```
female_data <- gender_depression["Female", ]
```

```
female_labels <- paste(y, round(100 * female_data/sum(female_data), 1), "%")
```

```
pie(female_data,
```

```
  labels = female_labels,
```

```
  col = c("pink", "lightgreen"),
```

```
  main = "Depression among Female Students")
```

```
# Academic pressure vs Depression
```

```
academic_depression <- table(class12$`Academic Pressure`, class12$Depression)
```

```
barplot(academic_depression, beside = TRUE,
```

```
  col = c("green", "red", "blue", "yellow", "pink", "black"),
```

```
  main = "Depression by Academic Pressure",
```

```
  xlab = "Academic Pressure Level",
```

```
ylab = "Number of Students",  
names.arg = y,  
legend = rownames(academic_depression),  
args.legend = list(title = "Academic Pressure", x = "topleft"))
```

```
# CGPA vs Depression
```

```
boxplot(CGPA ~ Depression, data = class12, col = c("yellow", "pink"),  
        main = "CGPA vs Depression", xlab = "Depression (0 = No, 1 = Yes)", ylab = "CGPA")
```

```
boxplot_stats <- by(class12$CGPA, class12$Depression, summary)  
boxplot_stats
```

```
# Study Satisfaction vs Depression
```

```
study_sat_depression <- table(class12$'Study Satisfaction', class12$Depression)
```

```
barplot(study_sat_depression, beside = TRUE,  
        col = c("green", "red", "blue", "yellow", "pink", "black"),  
        main = "Depression by Study Satisfaction",  
        ylab = "Number of Students",  
        names.arg = c("No Depression", "Depression"),  
        legend.text = rownames(study_sat_depression),  
        args.legend = list(title = "Study Satisfaction", x = "topleft"))
```

```
# Sleep Duration vs Depression
```

```
sleep_depression <- table(class12$`Sleep Duration`, class12$Depression)
```

```
barplot(sleep_depression, beside = TRUE,  
        col = c("green", "red","blue","yellow"),  
        main = "Depression by sleep duration Pressure",  
        ylab = "Number of Students",  
        names.arg = y,  
        legend = rownames(sleep_depression),  
        args.legend = list(title = "Sleep duration", x = "topleft"))
```

```
# Dietary Habits vs Depression
```

```
diet_depression <- table(class12$'Dietary Habits', class12$Depression)
```

```
barplot(diet_depression,  
        beside = TRUE,  
        col = c("green", "red","blue","yellow"),  
        main = "Depression by Dietary Habits",  
        ylab = "Count",  
        names.arg = y,  
        legend.text = rownames(diet_depression),  
        args.legend = list(title = "Dietary Habits", x = "topleft"))
```

```
# Suicidal Thoughts vs Depression
```

```
suicide_depression <- table(class12$'Have you ever had suicidal thoughts ?',  
                             class12$Depression)
```

```
barplot(suicide_depression,  
        beside = TRUE,
```

```
col = c("lavender", "red"),  
main = "Depression by Suicidal Thoughts",  
ylab = "Count",  
names.arg = y,  
legend.text = rownames(suicide_depression),  
args.legend = list(title = "Suicidal Thoughts", x = "topleft"))
```

Study Hours vs Depression

```
boxplot(class12$`Work/Study Hours` ~ class12$Depression,  
data = class12,  
col = c("lightsteelblue", "lightsalmon"),  
main = "Study Hours vs Depression",  
xlab = "Depression Status",  
ylab = "Study Hours",  
names = c("No Depression", "Depression"))
```

```
by(class12$`Work/Study Hours`, class12$Depression, summary)
```

Financial Stress vs Depression

```
fin_stress_depression <- table(class12$'Financial Stress', class12$Depression)  
barplot(fin_stress_depression,  
beside = TRUE,  
col = c("cyan", "pink", "red", "green", "blue"),  
main = "Depression by Financial Stress",  
ylab = "Count",  
names.arg = y,
```

```
legend.text = rownames(fin_stress_depression),  
args.legend = list(title = "Financial Stress Level", x = "topleft"))
```

```
# Family History vs Depression
```

```
family_history_depression <- table(class12$'Family History of Mental Illness',  
class12$Depression)
```

```
barplot(family_history_depression,  
        beside = TRUE,  
        col = c("tan", "purple"),  
        main = "Depression by Family History",  
        ylab = "Count",  
        names.arg = y,  
        legend.text = rownames(family_history_depression),  
        args.legend = list(title = "Family History", x = "topleft"))
```


DATA SET

Name - Student Depression Dataset

Download Link -

<https://www.kaggle.com/datasets/adilshamim8/student-depression-dataset/data>

~ THANK YOU ~