Course: Data Exploration and Preparation

Course Code: CAP482
CA 2
Dated: - 7/April/2025

Submitted By

Name: Aditya Kumar Yadav
Roll No: 42
Reg: 12307896
Section: De225, Group: 2

Submitted to
Ms. Ranjit Kaur Walia
UID: 28632
Assistant Professor
SCA, LPU

Lovely Faculty of Technology & Sciences

School of Computer Applications

Lovely Professional University

Punjab

Analyzing Student Depression Using R

Aditya Kumar Yadav

2025-04-28

RPubs Report: rpubs/adityadav_01/AnalyzingStudentDepressionUsingR

Project Overview

This project analyzes student depression using a dataset. It aims to find major reasons behind depression, identify common patterns, and understand how depression levels vary among students to help improve mental health.

Dataset Used

The dataset includes student details like age, gender, academic performance, habits, sleep, and stress levels. It was downloaded from Kaggle (Student Depression Dataset). "Work Pressure" and "Job Satisfaction" columns were removed.

Objectives

1. Understand the dataset (columns, missing values, data types). 2. Perform basic analysis with filtering, grouping, and summarizing. 3. Find key patterns in depression and stress factors. 4. Gain insights to support student mental health.

Level 1: Basic Exploration

```
# Load required libraries
library(readr)
library(dplyr)
```

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

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

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
library(tidyr)
```

0: Load data set

```
data <- read_csv("C:/Users/Aditya Yadav/Downloads/student depression.csv")
```

```
## Rows: 27901 Columns: 18
## ── Column specification
────────────────────────────────────────────────────────
## Delimiter: ","
## chr  (8): Gender, City, Profession, Sleep Duration, Dietary Habits,
Degree, ...
## dbl (10): id, Age, Academic Pressure, Work Pressure, CGPA, Study
Satisfactio...
##
## ℹ Use `spec()` to retrieve the full column specification for this data.
## ℹ Specify the column types or set `show_col_types = FALSE` to quiet this
message.
```

```r
head(data)
```

```
## # A tibble: 6 × 18
##      id Gender   Age City   Profession `Academic Pressure` `Work Pressure`
CGPA
##   <dbl> <chr>  <dbl> <chr>  <chr>                    <dbl>           <dbl>
<dbl>
## 1     2 Male      33 Visak… Student                      5               0
8.97
## 2     8 Female    24 Banga… Student                      2               0
5.9
## 3    26 Male      31 Srina… Student                      3               0
7.03
## 4    30 Female    28 Varan… Student                      3               0
5.59
## 5    32 Female    25 Jaipur Student                      4               0
8.13
## 6    33 Male      29 Pune   Student                      2               0
5.7
## # ℹ 10 more variables: `Study Satisfaction` <dbl>, `Job Satisfaction`
<dbl>,
## #   `Sleep Duration` <chr>, `Dietary Habits` <chr>, Degree <chr>,
## #   `Have you ever had suicidal thoughts ?` <chr>, `Work/Study Hours`
<dbl>,
## #   `Financial Stress` <dbl>, `Family History of Mental Illness` <chr>,
## #   Depression <dbl>
```

1: Data understanding

```r
# Check structure of dataset (data types of each column)
str(data)
```

```
## spc_tbl_ [27,901 × 18] (S3: spec_tbl_df/tbl_df/tbl/data.frame)
##  $ id                                   : num [1:27901] 2 8 26 30 32 33 52
56 59 62 ...
##  $ Gender                               : chr [1:27901] "Male" "Female"
"Male" "Female" ...
```

```
##  $ Age                                  : num [1:27901] 33 24 31 28 25 29
30 30 28 31 ...
##  $ City                                 : chr [1:27901] "Visakhapatnam"
"Bangalore" "Srinagar" "Varanasi" ...
##  $ Profession                           : chr [1:27901] "Student"
"Student" "Student" "Student" ...
##  $ Academic Pressure                    : num [1:27901] 5 2 3 3 4 2 3 2 3
2 ...
##  $ Work Pressure                        : num [1:27901] 0 0 0 0 0 0 0 0 0
0 ...
##  $ CGPA                                 : num [1:27901] 8.97 5.9 7.03 5.59
8.13 5.7 9.54 8.04 9.79 8.38 ...
##  $ Study Satisfaction                   : num [1:27901] 2 5 5 2 3 3 4 4 1
3 ...
##  $ Job Satisfaction                     : num [1:27901] 0 0 0 0 0 0 0 0 0
0 ...
##  $ Sleep Duration                       : chr [1:27901] "5-6 hours" "5-6
hours" "Less than 5 hours" "7-8 hours" ...
##  $ Dietary Habits                       : chr [1:27901] "Healthy"
"Moderate" "Healthy" "Moderate" ...
##  $ Degree                               : chr [1:27901] "B.Pharm" "BSc"
"BA" "BCA" ...
##  $ Have you ever had suicidal thoughts ?: chr [1:27901] "Yes" "No" "No"
"Yes" ...
##  $ Work/Study Hours                     : num [1:27901] 3 3 9 4 1 4 1 0 12
2 ...
##  $ Financial Stress                     : num [1:27901] 1 2 1 5 1 1 2 1 3
5 ...
##  $ Family History of Mental Illness     : chr [1:27901] "No" "Yes" "Yes"
"Yes" ...
##  $ Depression                           : num [1:27901] 1 0 0 1 0 0 0 0 1
1 ...
##  - attr(*, "spec")=
##   .. cols(
##   ..   id = col_double(),
##   ..   Gender = col_character(),
##   ..   Age = col_double(),
##   ..   City = col_character(),
##   ..   Profession = col_character(),
##   ..   `Academic Pressure` = col_double(),
##   ..   `Work Pressure` = col_double(),
##   ..   CGPA = col_double(),
##   ..   `Study Satisfaction` = col_double(),
##   ..   `Job Satisfaction` = col_double(),
##   ..   `Sleep Duration` = col_character(),
##   ..   `Dietary Habits` = col_character(),
##   ..   Degree = col_character(),
##   ..   `Have you ever had suicidal thoughts ?` = col_character(),
##   ..   `Work/Study Hours` = col_double(),
##   ..   `Financial Stress` = col_double(),
##   ..   `Family History of Mental Illness` = col_character(),
##   ..   Depression = col_double()
##   .. )
##  - attr(*, "problems")=<externalptr>
```

Dataset has 27,901 rows and 18 columns initially. Columns are a mix of numeric and character types.

```
# Get summary statistics (min, max, mean, etc.)
summary(data)
```

```
##        id            Gender               Age            City
##  Min.   :     2   Length:27901       Min.   :18.00   Length:27901
##  1st Qu.: 35039   Class :character   1st Qu.:21.00   Class :character
##  Median : 70684   Mode  :character   Median :25.00   Mode  :character
##  Mean   : 70442                      Mean   :25.82
##  3rd Qu.:105818                      3rd Qu.:30.00
##  Max.   :140699                      Max.   :59.00
##
##   Profession        Academic Pressure Work Pressure          CGPA
##  Length:27901       Min.   :0.000     Min.   :0.00000   Min.   : 0.000
##  Class :character   1st Qu.:2.000     1st Qu.:0.00000   1st Qu.: 6.290
##  Mode  :character   Median :3.000     Median :0.00000   Median : 7.770
##                     Mean   :3.141     Mean   :0.00043   Mean   : 7.656
##                     3rd Qu.:4.000     3rd Qu.:0.00000   3rd Qu.: 8.920
##                     Max.   :5.000     Max.   :5.00000   Max.   :10.000
##
##  Study Satisfaction Job Satisfaction   Sleep Duration     Dietary Habits
##  Min.   :0.000      Min.   :0.000000   Length:27901       Length:27901
##  1st Qu.:2.000      1st Qu.:0.000000   Class :character   Class :character
##  Median :3.000      Median :0.000000   Mode  :character   Mode  :character
##  Mean   :2.944      Mean   :0.000681
##  3rd Qu.:4.000      3rd Qu.:0.000000
##  Max.   :5.000      Max.   :4.000000
##
##     Degree          Have you ever had suicidal thoughts ? Work/Study Hours
##  Length:27901       Length:27901                          Min.   : 0.000
##  Class :character   Class :character                      1st Qu.: 4.000
##  Mode  :character   Mode  :character                      Median : 8.000
##                                                           Mean   : 7.157
##                                                           3rd Qu.:10.000
##                                                           Max.   :12.000
##
##  Financial Stress Family History of Mental Illness   Depression
##  Min.   :1.00     Length:27901                     Min.   :0.0000
##  1st Qu.:2.00     Class :character                 1st Qu.:0.0000
##  Median :3.00     Mode  :character                 Median :1.0000
##  Mean   :3.14                                      Mean   :0.5855
##  3rd Qu.:4.00                                      3rd Qu.:1.0000
##  Max.   :5.00                                      Max.   :1.0000
##  NA's   :3
```

Depression column is numeric (0 = No Depression, 1 = Depression).

```
# Get data set dimensions (total rows and columns)
dim(data)

## [1] 27901    18
```

2: Missing Values

```
# Count total missing values
sum(is.na(data))

## [1] 3

# Column-wise count of missing values
colSums(is.na(data))

##                                     id
Gender
##                                      0
0
##                                    Age
City
##                                      0
0
##                             Profession                     Academic
Pressure
##                                      0
0
##                          Work Pressure
CGPA
##                                      0
0
##                     Study Satisfaction                      Job
Satisfaction
##                                      0
0
##                         Sleep Duration                        Dietary
Habits
##                                      0
0
##                                 Degree Have you ever had suicidal thoughts
?
##                                      0
0
##                       Work/Study Hours                      Financial
Stress
```

```
##                                     0
3
##       Family History of Mental Illness
Depression
##                                     0
0
```

3: Clean Data

```r
# Remove unnecessary columns
data <- data %>% select(-`Work Pressure`, -`Job Satisfaction`)
#Remove rows with NA
data<-na.omit(data)
```

Work Pressure and Job Satisfaction columns were removed because they were not useful. Rows with NA values were also removed (now dataset has slightly fewer rows). Now the data is clean and ready for further analysis.

4: Calculate the percentage of students with depression

```r
percentage_depressed <- mean(data$Depression) * 100
print(paste("Percentage of students with depression:",
round(percentage_depressed, 2), "%"))

## [1] "Percentage of students with depression: 58.55 %"
```

Level 2: Identifying Patterns

5: Find the most common stress factors

```r
# Mean calculation with NA handling
financial_stress_mean <- mean(data$`Financial Stress`, na.rm = TRUE)
academic_pressure_mean <- mean(data$`Academic Pressure`, na.rm = TRUE)

# Comparison using if-else
if (academic_pressure_mean > financial_stress_mean) {
  print("Academic pressure is higher than financial stress for depression")
} else if (financial_stress_mean > academic_pressure_mean) {
  print("Financial stress is higher than academic pressure for depression")
} else {
  print("Both academic pressure and financial stress are equal for
depression")
}

## [1] "Academic pressure is higher than financial stress for depression"
```

6: List students with depression

```
students_with_depression <- data %>% filter(Depression == 1)
head(students_with_depression)

## # A tibble: 6 × 16
##      id Gender   Age City          Profession `Academic Pressure`  CGPA
##   <dbl> <chr>  <dbl> <chr>         <chr>                    <dbl> <dbl>
## 1     2 Male      33 Visakhapatnam Student                      5  8.97
## 2    30 Female    28 Varanasi      Student                      3  5.59
## 3    59 Male      28 Nagpur        Student                      3  9.79
## 4    62 Male      31 Nashik        Student                      2  8.38
## 5    83 Male      24 Nagpur        Student                      3  6.1
## 6    94 Male      27 Kalyan        Student                      5  7.04
## # ℹ 9 more variables: `Study Satisfaction` <dbl>, `Sleep Duration` <chr>,
## #   `Dietary Habits` <chr>, Degree <chr>,
## #   `Have you ever had suicidal thoughts ?` <chr>, `Work/Study Hours`
<dbl>,
## #   `Financial Stress` <dbl>, `Family History of Mental Illness` <chr>,
## #   Depression <dbl>
```

7: Identify students with low CGPA and depression

```
# Total depressed students with CGPA below average
low_performance_depressed <- data %>%
  filter(CGPA < mean(CGPA) & Depression == 1)%>%
  nrow()

# Total students with CGPA below average
low_cgpa_students <- data %>%
  filter(CGPA < mean(CGPA)) %>%
  nrow()

# Calculate percentage
percentage_low_cgpa_depressed <- (low_performance_depressed /
low_cgpa_students) * 100

# Print result
print(paste("Percentage of students with low CGPA who are depressed:",
round(percentage_low_cgpa_depressed, 2), "%"))

## [1] "Percentage of students with low CGPA who are depressed: 56.84 %"
```

INTERPRETATION:

Academic pressure is found to be higher than financial stress among students with depression. A list of students suffering from depression was successfully filtered. Around

56.84% of students who have low CGPA are also suffering from depression. This shows a strong link between academic performance and depression levels.

Level 3: Grouping & Summarization

8: Group data by Age and calculate percentage of depressed students in each group

```
age_group_depression <- data %>%
  group_by(Age) %>%
  summarise(
    total_students = n(),
    depressed_students = sum(Depression),
    percentage_depressed = (depressed_students / total_students) * 100
  )

# Print result
print(age_group_depression)
```

```
## # A tibble: 34 × 4
##      Age total_students depressed_students percentage_depressed
##    <dbl>          <int>              <dbl>                <dbl>
##  1    18           1587               1216                 76.6
##  2    19           1560               1100                 70.5
##  3    20           2236               1579                 70.6
##  4    21           1726               1169                 67.7
##  5    22           1160                701                 60.4
##  6    23           1645               1051                 63.9
##  7    24           2258               1509                 66.8
##  8    25           1784               1082                 60.7
##  9    26           1155                663                 57.4
## 10    27           1462                887                 60.7
## # ℹ 24 more rows
```

INTERPRETATION:

Age group 18 has the highest depression rate: 76.6% students are depressed.

9: Relationship between CGPA and depression by grouping students into CGPA

```
# Create CGPA Groups
data <- data %>%
  mutate(CGPA_Group = cut(CGPA,
                          breaks = c(-1, 4, 7, 10),
                          labels = c("0-4", "5-7", "8-10"),
                          right = TRUE))

# Calculate total students and depressed students per CGPA group
cgpa_depression_summary <- data %>%
  group_by(CGPA_Group) %>%
  summarise(
    Total_Students = n(),
    Total_Depressed = sum(Depression == 1, na.rm = TRUE),
    Percentage_Depressed = (Total_Depressed / Total_Students) * 100
  )

# Print result
print(cgpa_depression_summary)

## # A tibble: 3 × 4
##   CGPA_Group Total_Students Total_Depressed Percentage_Depressed
##   <fct>               <int>           <int>                <dbl>
## 1 0-4                     9               4                 44.4
## 2 5-7                  9730            5589                 57.4
## 3 8-10                18159           10742                 59.2
```

INTERPRETATION:

Surprisingly, even students with good CGPA (8-10) are facing high depression rates, showing good marks ≠ mental peace.

10: Which degree program has the highest number of depressed students

```
# Group by Degree and calculate depression stats
degree_depression <- data %>%
  group_by(Degree) %>%
  summarise(
    Total_Students = n(),
    Depressed_Students = sum(Depression == 1, na.rm = TRUE),
    Percentage_Depressed = (Depressed_Students / Total_Students) * 100
  ) %>%
  arrange(desc(Percentage_Depressed))  # Sort by highest depression
percentage

# Show the result
head(degree_depression)

## # A tibble: 6 × 4
##   Degree   Total_Students Depressed_Students Percentage_Depressed
##   <chr>             <int>              <int>                <dbl>
## 1 Class 12           6080               4303                 70.8
## 2 Others               35                 21                 60
## 3 B.Arch             1478                871                 58.9
## 4 BSc                 888                523                 58.9
## 5 BBA                 696                407                 58.5
## 6 MBBS                695                404                 58.1
```

INTERPRETATION :

The trend across degree programs reveals that higher education doesn’t necessarily equate to better mental health, and Class 12 students may be struggling the most with mental health issues.

Level 4: Ranking & Comparison

11: Rank students based on CGPA and Depression levels

```
ranked_data <- data %>%
  # Rank based on CGPA (higher CGPA = better rank)
  mutate(CGPA_Rank = dense_rank(desc(CGPA))) %>%
  # Arrange by Depression (1 first) and CGPA_Rank
  arrange(desc(Depression), CGPA_Rank) %>%
  # Assign final ranking
  mutate(Final_Rank = row_number())

# Select relevant columns
ranked_students <- ranked_data %>%
  select(id, CGPA, CGPA_Rank, Depression, Final_Rank)

# Print top 10 ranked students
print(head(ranked_students, 10))

## # A tibble: 10 × 5
##        id  CGPA CGPA_Rank Depression Final_Rank
##     <dbl> <dbl>     <int>      <dbl>      <int>
##   1 13170    10         1          1          1
##   2 15800    10         1          1          2
##   3 22499    10         1          1          3
##   4 24975    10         1          1          4
##   5 25353    10         1          1          5
##   6 25482    10         1          1          6
##   7 26892    10         1          1          7
##   8 32697    10         1          1          8
##   9 34831    10         1          1          9
## 10 38077    10         1          1         10
```

INTERPRETATION :

Top students ranked by CGPA and depression levels show that those with high CGPA (10) are consistently ranked highly, even if they are depressed (Depression = 1).

12: What is the count and percentage of depression cases among males and females

```
# Count of depression cases by gender
table(data$Gender, data$Depression)

## 
##             0    1
##   Female 5132 7220
##   Male   6431 9115

# Percentage of depression in each gender
prop.table(table(data$Gender, data$Depression)) * 100

## 
##                 0        1
##   Female 18.39558 25.87999
##   Male   23.05183 32.67259
```

#Interpretation: *Males have a higher percentage of depression compared to females.*

13: Count and percentage of depressed students by dietary habit

```
dietary_depression <- data %>%
  group_by(`Dietary Habits`, Depression) %>%
  summarise(count = n(), .groups = "drop") %>%
  mutate(Percentage = (count / sum(count)) * 100)

# Print result
print(dietary_depression)

## # A tibble: 8 × 4
##   `Dietary Habits` Depression count Percentage
##   <chr>                 <dbl> <int>      <dbl>
## 1 Healthy                   0  4177    15.0   
## 2 Healthy                   1  3472    12.4   
## 3 Moderate                  0  4363    15.6   
## 4 Moderate                  1  5558    19.9   
## 5 Others                    0     4     0.0143
## 6 Others                    1     8     0.0287
## 7 Unhealthy                 0  3019    10.8
## 8 Unhealthy                 1  7297    26.2
```

#Interpretation: *Unhealthy dietary habits correlate with a higher percentage of depression.*

14: Who are depressed have also had suicidal thoughts

```
suicidal_depressed <- data %>%
  filter(Depression == 1, `Have you ever had suicidal thoughts ?` == "Yes")
%>%
  summarise(count = n())

print(suicidal_depressed)
```

```
## # A tibble: 1 × 1
##   count
##   <int>
## 1 13957
```

## Interpretation:

A total of 13,957 students (around 46% of depressed students) have reported having suicidal thoughts. A significant number of depressed students are experiencing suicidal thoughts, highlighting the critical need for mental health support.

Level 5: Creating New Insights

15: Add a new column “Depression_Status”

```
data <- data %>%
  mutate(Depression_Status = ifelse(Depression == 1, "Depressed", "No
Depression"))
head(data)
```

```
## # A tibble: 6 × 18
##      id Gender   Age City          Profession `Academic Pressure`  CGPA
##   <dbl> <chr>  <dbl> <chr>         <chr>                     <dbl> <dbl>
## 1     2 Male      33 Visakhapatnam Student                       5  8.97
## 2     8 Female    24 Bangalore     Student                       2  5.9
## 3    26 Male      31 Srinagar      Student                       3  7.03
## 4    30 Female    28 Varanasi      Student                       3  5.59
## 5    32 Female    25 Jaipur        Student                       4  8.13
## 6    33 Male      29 Pune          Student                       2  5.7
## # ℹ 11 more variables: `Study Satisfaction` <dbl>, `Sleep Duration` <chr>,
## #   `Dietary Habits` <chr>, Degree <chr>,
## #   `Have you ever had suicidal thoughts ?` <chr>, `Work/Study Hours`
<dbl>,
## #   `Financial Stress` <dbl>, `Family History of Mental Illness` <chr>,
## #   Depression <dbl>, CGPA_Group <fct>, Depression_Status <chr>
```

16: Create a “Total_Stress” column

```
data <- data %>%
  mutate(Total_Stress = `Academic Pressure` + `Financial Stress`)
head(data)
```

```
## # A tibble: 6 × 19
##      id Gender   Age City          Profession `Academic Pressure`  CGPA
##   <dbl> <chr>  <dbl> <chr>         <chr>                    <dbl> <dbl>
## 1     2 Male      33 Visakhapatnam Student                      5  8.97
## 2     8 Female    24 Bangalore     Student                      2  5.9
## 3    26 Male      31 Srinagar      Student                      3  7.03
## 4    30 Female    28 Varanasi      Student                      3  5.59
## 5    32 Female    25 Jaipur        Student                      4  8.13
## 6    33 Male      29 Pune          Student                      2  5.7
## # ℹ 12 more variables: `Study Satisfaction` <dbl>, `Sleep Duration` <chr>,
## #   `Dietary Habits` <chr>, Degree <chr>,
## #   `Have you ever had suicidal thoughts ?` <chr>, `Work/Study Hours`
<dbl>,
## #   `Financial Stress` <dbl>, `Family History of Mental Illness` <chr>,
## #   Depression <dbl>, CGPA_Group <fct>, Depression_Status <chr>,
## #   Total_Stress <dbl>
```

Rename the columns for easy-to-use

```
colnames(data) <- c("id", "gender", "age", "city", "profession",
                    "academic_pressure", "cgpa", "study_satisfaction",
                    "sleep_duration", "dietary_habits", "degree",
                    "suicidal_thoughts", "work_study_hours",
                    "financial_stress", "mental_illness_history",
                    "depression", "cgpa_group", "depression_status",
                    "total_stress")
colnames(data)
```

```
##  [1] "id"                     "gender"                 "age"
##  [4] "city"                   "profession"             "academic_pressure"
##  [7] "cgpa"                   "study_satisfaction"     "sleep_duration"
## [10] "dietary_habits"         "degree"                 "suicidal_thoughts"
## [13] "work_study_hours"       "financial_stress"
"mental_illness_history"
## [16] "depression"             "cgpa_group"             "depression_status"
## [19] "total_stress"
```

Interpretation:

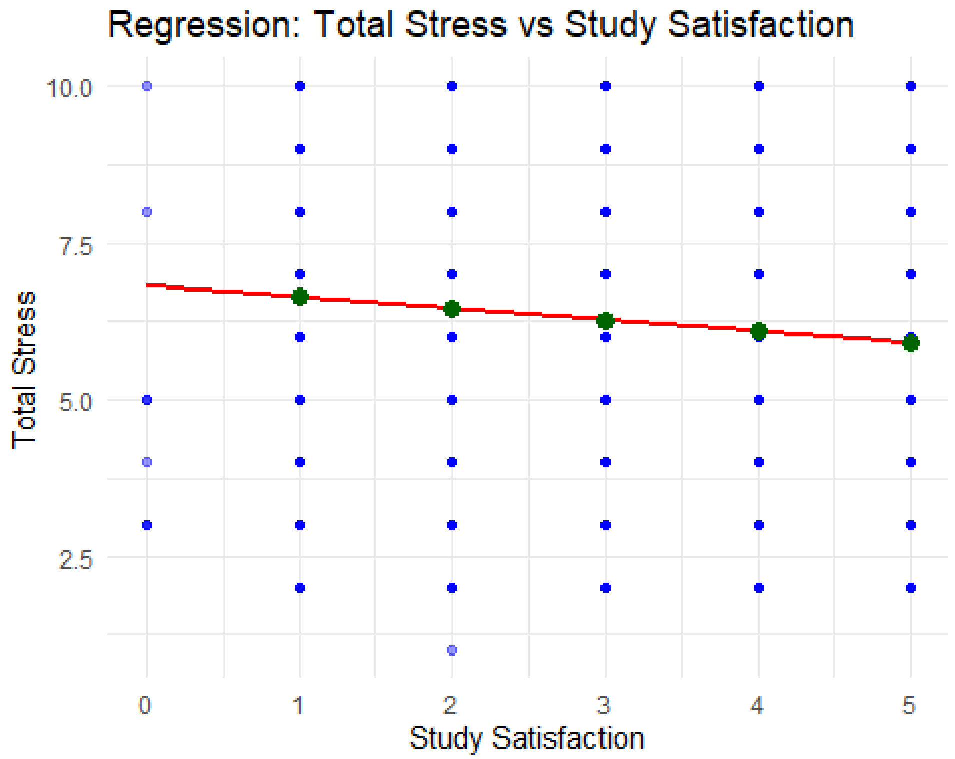
Renamed columns for better clarity (e.g., Academic Pressure to academic_pressure, CGPA to cgpa, etc.).

Level 6: Regression

Q1: Simple Linear Regression : Total_Stress based on Study Satisfaction

```
# Load required library for plotting
library(ggplot2)

# Step 1: Build the regression model: Total_Stress ~ Study_Satisfaction
stress_model <- lm(total_stress ~ study_satisfaction, data = data)

# Step 2: View model summary
summary(stress_model)

##
## Call:
## lm(formula = total_stress ~ study_satisfaction, data = data)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -5.452 -1.634  0.092  1.548  4.092
##
## Coefficients:
##                     Estimate Std. Error t value Pr(>|t|)
## (Intercept)         6.815531   0.030318  224.80   <2e-16 ***
## study_satisfaction -0.181499   0.009348  -19.42   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.125 on 27896 degrees of freedom
## Multiple R-squared:  0.01333,    Adjusted R-squared:  0.0133
## F-statistic:   377 on 1 and 27896 DF,  p-value: < 2.2e-16

# Step 3: Create new Study Satisfaction values for prediction
new_data_stress <- data.frame(study_satisfaction = c(1, 2, 3, 4, 5))

# Step 4: Predict Total_Stress using the model
new_data_stress$predicted_stress <- predict(stress_model, newdata =
new_data_stress)

# Step 5: Print predictions
print("Predicted Total Stress:")
```

```
## [1] "Predicted Total Stress:"

print(new_data_stress)

##   study_satisfaction predicted_stress
## 1                  1         6.634032
## 2                  2         6.452533
## 3                  3         6.271034
## 4                  4         6.089536
## 5                  5         5.908037

# Step 6: Visualization - Actual data, regression line, and predicted points
ggplot(data, aes(x = study_satisfaction, y = total_stress)) +
  geom_point(alpha = 0.4, color = "blue") +  # Blue points for actual data
  geom_smooth(method = "lm", se = FALSE, color = "red") +  # Red regression
line
  geom_point(data = new_data_stress, aes(x = study_satisfaction, y =
predicted_stress),
             color = "darkgreen", size = 3) +  # Dark green points for
predicted data
  labs(title = "Regression: Total Stress vs Study Satisfaction",
       x = "Study Satisfaction",
       y = "Total Stress") +
  theme_minimal()  # Apply minimal theme for clean look

## `geom_smooth()` using formula = 'y ~ x'
```



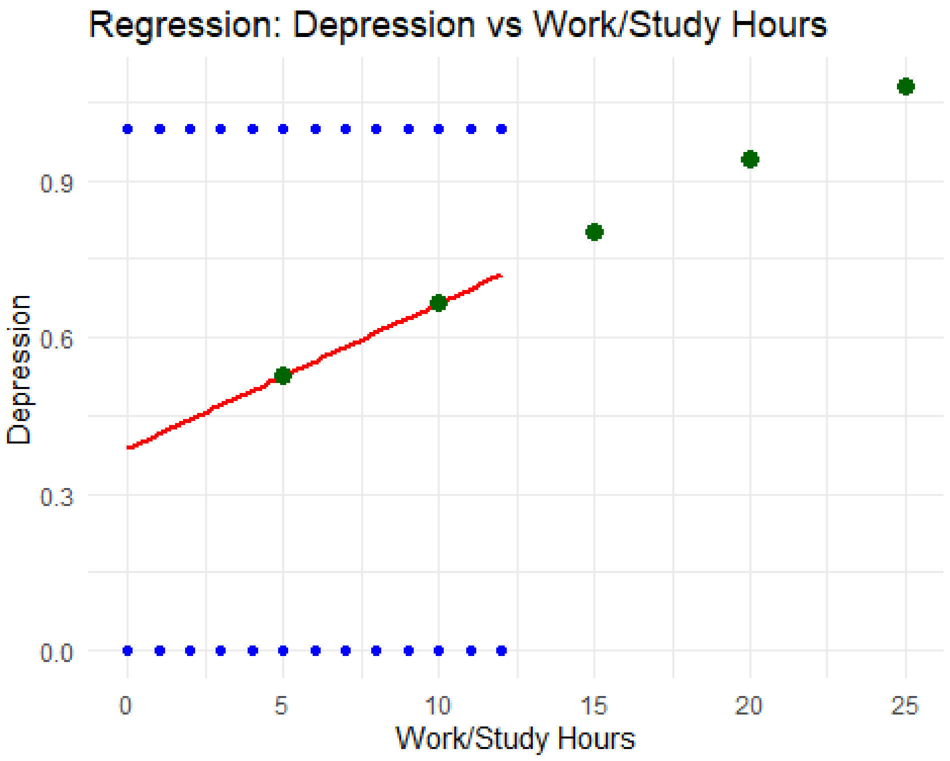
Interpretation:

The simple linear regression model shows a weak negative relationship between study satisfaction and total stress, with study satisfaction explaining only 1.33% of the variation in total stress. As study satisfaction increases, total stress tends to decrease, but the low R-squared value suggests that study satisfaction isn't a strong predictor of total stress.

Q2: Simple Linear Regression : Predicting Depression Score using Work/Study Hours

```
# Load required library for plotting
library(ggplot2)

# Step 1: Build the regression model: Depression ~ Work/Study Hours
depression_model <- lm(depression ~ work_study_hours, data = data)

# Step 2: View model summary
summary(depression_model)

##
## Call:
## lm(formula = depression ~ work_study_hours, data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.7198 -0.5257  0.3079  0.3911  0.6129
##
## Coefficients:
##                  Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.387139   0.006271   61.73   <2e-16 ***
## work_study_hours 0.027721   0.000778   35.63   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4818 on 27896 degrees of freedom
## Multiple R-squared:  0.04353,    Adjusted R-squared:  0.04349
## F-statistic:  1269 on 1 and 27896 DF,  p-value: < 2.2e-16

# Step 3: Create new Work/Study Hours values for prediction
new_data_depression <- data.frame(work_study_hours = c(5, 10, 15, 20, 25))

# Step 4: Predict Depression using the model
new_data_depression$predicted_depression <- predict(depression_model, newdata
= new_data_depression)

# Step 5: Print predictions
print("Predicted Depression based on Work/Study Hours:")
```

```
## [1] "Predicted Depression based on Work/Study Hours:"

print(new_data_depression)

##   work_study_hours predicted_depression
## 1                5            0.5257437
## 2               10            0.6643483
## 3               15            0.8029530
## 4               20            0.9415576
## 5               25            1.0801622

# Step 6: Visualization - Actual data, regression line, and predicted points
ggplot(data, aes(x = work_study_hours, y = depression)) +
  geom_point(alpha = 0.4, color = "blue") +  # Blue points for actual data
  geom_smooth(method = "lm", se = FALSE, color = "red") +  # Red regression
line
  geom_point(data = new_data_depression, aes(x = work_study_hours, y =
predicted_depression),
             color = "darkgreen", size = 3) +  # Dark green points for
predicted data
  labs(title = "Regression: Depression vs Work/Study Hours",
       x = "Work/Study Hours",
       y = "Depression") +
  theme_minimal()  # Apply minimal theme for clean look

## `geom_smooth()` using formula = 'y ~ x'
```



Interpretation:

The simple linear regression model shows a weak positive relationship between work/study hours and depression, with work/study hours explaining only 4.35% of the variation in depression levels. As work/study hours increase, depression scores also rise, but the low R-squared value suggests work/study hours are not a strong predictor of depression.

Q3: Multiple Linear Regression : Depression ~ Academic Pressure + Financial Stress

```
# Step 1: Corrected model
depression_model_multi <- lm(depression ~ academic_pressure +
financial_stress, data = data)

# Step 2: Summary
summary(depression_model_multi)

##
## Call:
## lm(formula = depression ~ academic_pressure + financial_stress,
##     data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -1.0605 -0.3450  0.0418  0.3481  0.9613
##
## Coefficients:
##                    Estimate Std. Error t value Pr(>|t|)
## (Intercept)       -0.216793   0.007592  -28.56   <2e-16 ***
## academic_pressure  0.153170   0.001791   85.51   <2e-16 ***
## financial_stress   0.102285   0.001722   59.41   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.4085 on 27895 degrees of freedom
## Multiple R-squared:  0.3124, Adjusted R-squared:  0.3124
## F-statistic:  6338 on 2 and 27895 DF,  p-value: < 2.2e-16

# Step 3: New data for prediction (total_stress removed)
new_data_depression_multi <- data.frame(
  academic_pressure = c(6, 8, 10, 12, 14),
  financial_stress = c(5, 6, 7, 8, 9)
)

# Step 4: Predict Depression
new_data_depression_multi$predicted_depression <-
```

```
predict(depression_model_multi, newdata = new_data_depression_multi)

# Step 5: Print predictions
print("Predicted Depression based on Academic Pressure and Financial
Stress:")
```

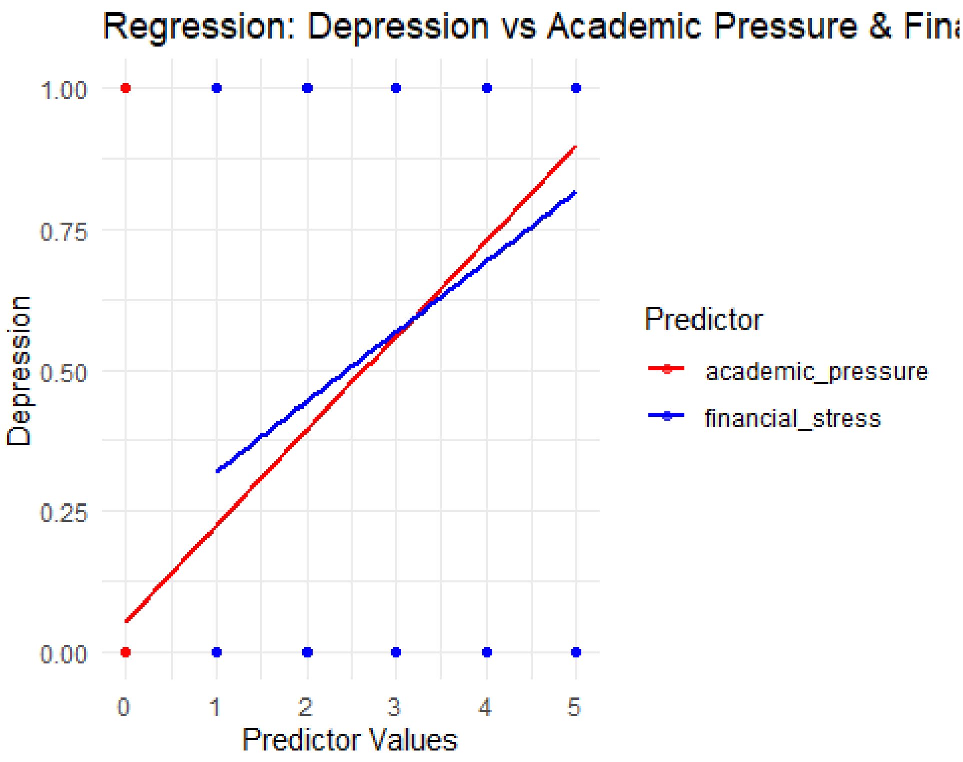
```
## [1] "Predicted Depression based on Academic Pressure and Financial
Stress:"
```

```
print(new_data_depression_multi)
```

```
##   academic_pressure financial_stress predicted_depression
## 1                 6                5             1.213651
## 2                 8                6             1.622275
## 3                10                7             2.030900
## 4                12                8             2.439525
## 5                14                9             2.848149
```

```
# Step 6: Visualization (Separate lines for academic_pressure and
financial_stress)
library(tidyr)

# Convert for plotting
data_long <- data %>%
  gather(key = "Predictor", value = "Value", academic_pressure,
financial_stress)

ggplot(data_long, aes(x = Value, y = depression, color = Predictor)) +
  geom_point(alpha = 0.6) +
  geom_smooth(method = "lm", se = FALSE) +
  labs(title = "Regression: Depression vs Academic Pressure & Financial
Stress",
       x = "Predictor Values",
       y = "Depression") +
  theme_minimal() +
  scale_color_manual(values = c("red", "blue"))
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



Interpretation:

The multiple linear regression model predicts depression based on academic pressure and financial stress, showing a moderate correlation (R-squared = 31.24%). Both factors significantly increase depression, with predictions rising from 1.21 to 2.85 as academic pressure and financial stress increase. The visualization shows positive trends for both predictors.

Q4: Polynomial Regression : Study Satisfaction vs Total Stress

```
# Load required libraries
library(ggplot2)

# Step 1: Build the polynomial regression model (2nd degree polynomial)
polynomial_model <- lm(total_stress ~ poly(study_satisfaction, 2), data =
data)

# Step 2: View model summary
summary(polynomial_model)

##
## Call:
```

```
## lm(formula = total_stress ~ poly(study_satisfaction, 2), data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -5.4233 -1.6943  0.0197  1.5767  4.0197
##
## Coefficients:
##                               Estimate Std. Error t value Pr(>|t|)
## (Intercept)                    6.28120    0.01272 493.817  < 2e-16 ***
## poly(study_satisfaction, 2)1 -41.26191    2.12453 -19.422  < 2e-16 ***
## poly(study_satisfaction, 2)2   8.40670    2.12453   3.957 7.61e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.125 on 27895 degrees of freedom
## Multiple R-squared:  0.01389,    Adjusted R-squared:  0.01382
## F-statistic: 196.4 on 2 and 27895 DF,  p-value: < 2.2e-16
```

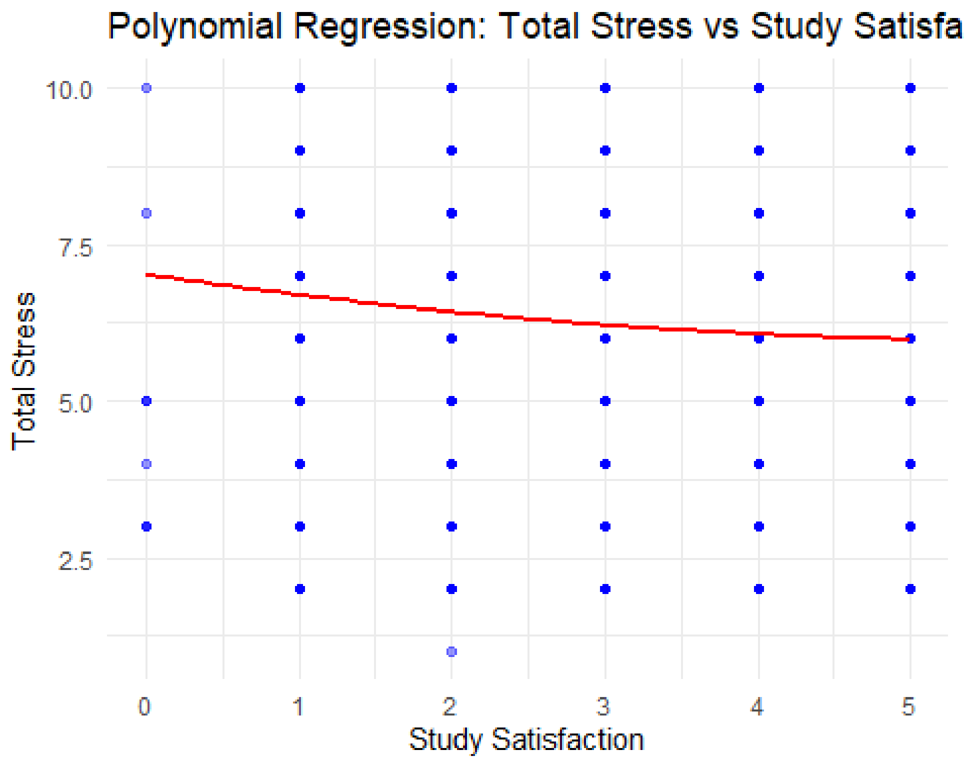
```
# Step 3: Create new data for prediction (you can customize the values)
new_data_polynomial <- data.frame(study_satisfaction = seq(1, 5, by = 0.1))

# Step 4: Predict Total Stress using the polynomial model
new_data_polynomial$predicted_stress <- predict(polynomial_model, newdata =
new_data_polynomial)

# Step 5: Print predictions
print("Predicted Total Stress (Polynomial Regression):")
```

```
## [1] "Predicted Total Stress (Polynomial Regression):"
```

```
print(new_data_polynomial)
```

```
##    study_satisfaction predicted_stress
## 1                 1.0         6.694259
## 2                 1.1         6.664389
## 3                 1.2         6.635136
## 4                 1.3         6.606499
## 5                 1.4         6.578479
## 6                 1.5         6.551075
## 7                 1.6         6.524287
## 8                 1.7         6.498117
## 9                 1.8         6.472562
## 10                1.9         6.447625
## 11                2.0         6.423303
## 12                2.1         6.399599
## 13                2.2         6.376510
## 14                2.3         6.354039
## 15                2.4         6.332184
## 16                2.5         6.310945
## 17                2.6         6.290323
## 18                2.7         6.270317
## 19                2.8         6.250928
## 20                2.9         6.232155
## 21                3.0         6.213999
## 22                3.1         6.196460
## 23                3.2         6.179537
## 24                3.3         6.163230
## 25                3.4         6.147540
## 26                3.5         6.132467
## 27                3.6         6.118010
## 28                3.7         6.104169
## 29                3.8         6.090945
## 30                3.9         6.078338
## 31                4.0         6.066347
## 32                4.1         6.054973
## 33                4.2         6.044215
## 34                4.3         6.034073
## 35                4.4         6.024549
## 36                4.5         6.015640
## 37                4.6         6.007348
## 38                4.7         5.999673
## 39                4.8         5.992614
## 40                4.9         5.986172
## 41                5.0         5.980346
```

```
# Step 6: Visualization - Actual data, polynomial regression curve, and
predicted points
ggplot(data, aes(x = study_satisfaction, y = total_stress)) +
  geom_point(alpha = 0.4, color = "blue") +  # Blue points for actual data
  geom_smooth(method = "lm", formula = y ~ poly(x, 2), se = FALSE, color =
"red") +  # Polynomial regression line
  labs(title = "Polynomial Regression: Total Stress vs Study Satisfaction",
       x = "Study Satisfaction",
       y = "Total Stress") +
  theme_minimal()  # Apply minimal theme for clean look
```



Interpretation:

The polynomial regression model shows that study satisfaction and total stress have a non-linear relationship. Initially, as study satisfaction increases, total stress decreases, but after reaching a certain level of satisfaction, stress begins to rise again. The model explains only 1.39% of the total stress variation, indicating other factors may also influence stress. Both study satisfaction terms in the model are statistically significant, and the model overall is highly significant.

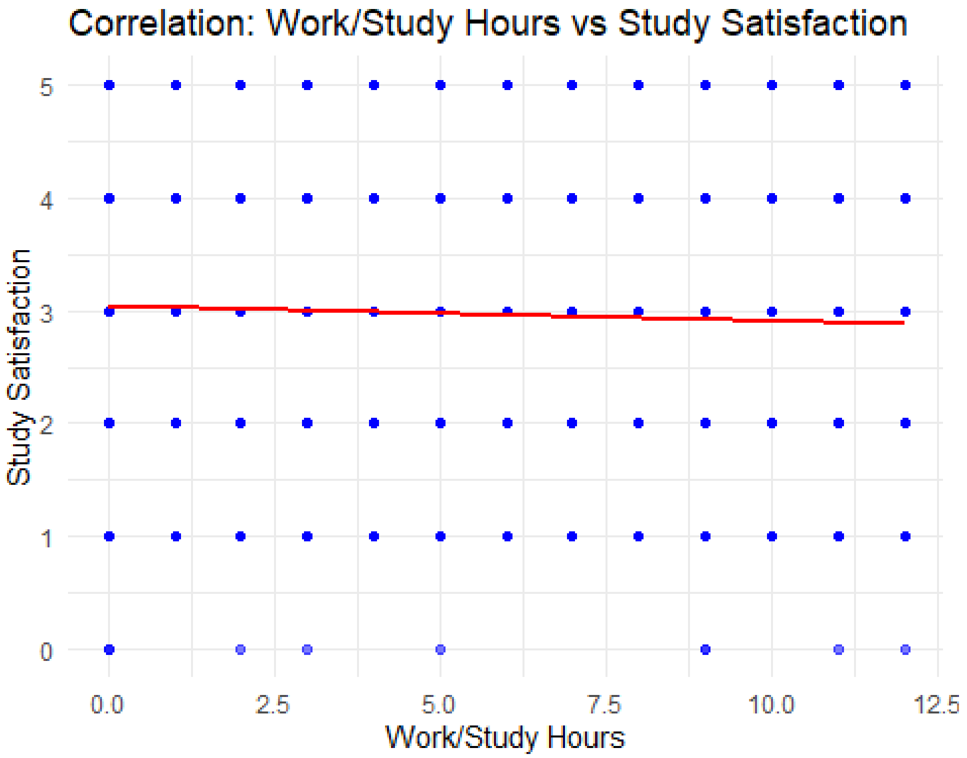
Level 7: Correlation

Q1: Correlation: Work/Study Hours vs Study Satisfaction

```
cor_work_study_hours_study_satisfaction <- cor(data$work_study_hours,
data$study_satisfaction, method = "pearson")
print(paste("Correlation between Work/Study Hours and Study Satisfaction: ",
cor_work_study_hours_study_satisfaction))

## [1] "Correlation between Work/Study Hours and Study Satisfaction:  -
0.0363560754182014"

# Visualization: Correlation Plot
library(ggplot2)
```

```r
# Scatter Plot with Correlation Line - Visualizing linear relationship
between Work/Study Hours and Study Satisfaction
ggplot(data, aes(x = work_study_hours, y = study_satisfaction)) +
  geom_point(color = "blue", alpha = 0.5) +
  geom_smooth(method = "lm", color = "red", se = FALSE) +
  labs(title = "Correlation: Work/Study Hours vs Study Satisfaction", x =
"Work/Study Hours", y = "Study Satisfaction") +
  theme_minimal()
```

```
## `geom_smooth()` using formula = 'y ~ x'
```



Interpretation:

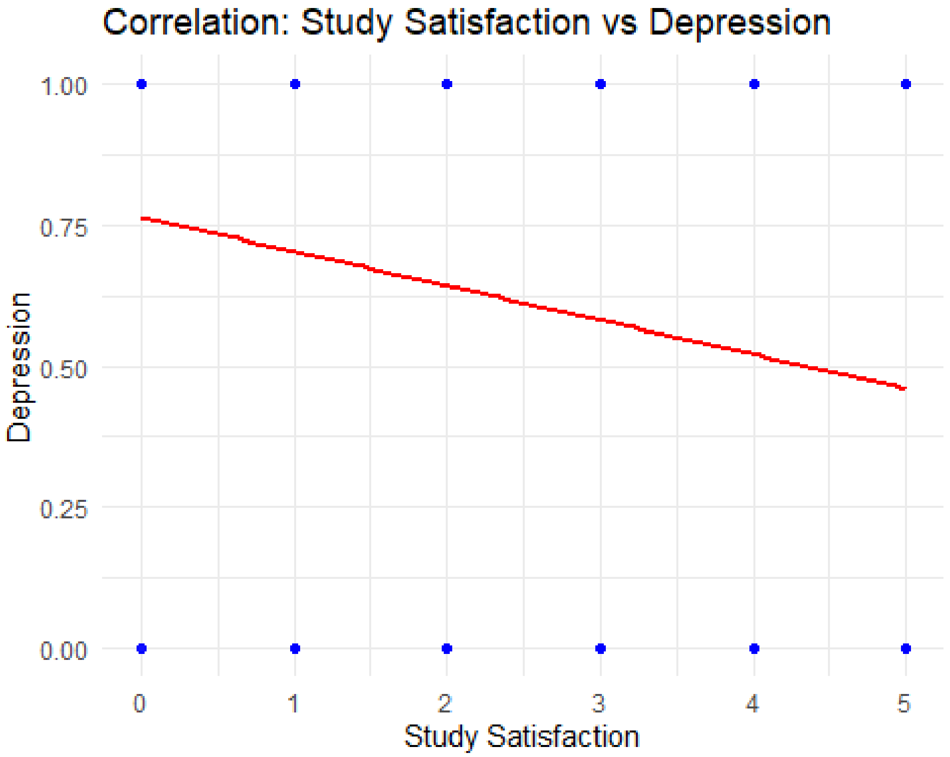
The correlation between Work/Study Hours and Study Satisfaction is very weak, with a value of -0.036. This indicates that there is virtually no linear relationship between the two variables.

The scatter plot with the red line represents a linear regression fit, but given the low correlation, the line doesn't show any strong trend or pattern. The data points are widely spread, suggesting that work/study hours do not significantly affect study satisfaction.

Q2: Correlation: Study Satisfaction vs Depression

```r
cor_study_satisfaction_depression <- cor(data$study_satisfaction,
data$depression, method = "pearson")
```

```r
print(paste("Correlation between Study Satisfaction and Depression: ",
cor_study_satisfaction_depression))

## [1] "Correlation between Study Satisfaction and Depression:  -
0.168010323159483"

# Scatter Plot with Correlation Line
ggplot(data, aes(x = study_satisfaction, y = depression)) +
  geom_point(color = "blue", alpha = 0.5) +
  geom_smooth(method = "lm", color = "red", se = FALSE) +
  labs(title = "Correlation: Study Satisfaction vs Depression", x = "Study
Satisfaction", y = "Depression") +
  theme_minimal()

## `geom_smooth()` using formula = 'y ~ x'
```



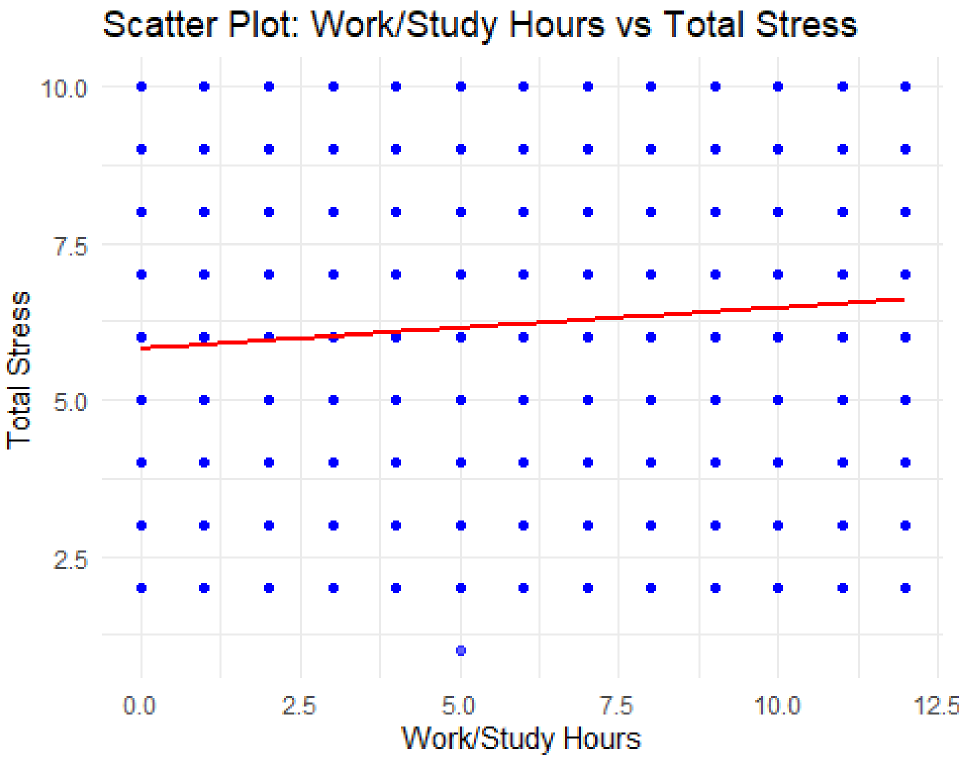
Interpretation:

The correlation between Study Satisfaction and Depression is -0.168, which is a weak negative correlation. This suggests that as study satisfaction decreases, depression tends to increase, but the relationship is not strong or significant.

The scatter plot with the red regression line also shows a weak downward trend, but the spread of the data points indicates that study satisfaction doesn't have a strong effect on depression.

Q3: Correlation: Work/Study Hours vs Total Stress

```
# Step 1: Calculate Pearson Correlation
cor_work_study_hours_total_stress <- cor(data$work_study_hours,
data$total_stress, method = "pearson")

# Step 2: Print Correlation
print(paste("Correlation between Work/Study Hours and Total Stress: ",
cor_work_study_hours_total_stress))

## [1] "Correlation between Work/Study Hours and Total Stress:
0.112604598981649"

# Step 3: Visualization with Scatter Plot and Regression Line
library(ggplot2)
ggplot(data, aes(x = work_study_hours, y = total_stress)) +
  geom_point(color = "blue", alpha = 0.6) +   # Scatter points
  geom_smooth(method = "lm", se = FALSE, color = "red") +  # Linear
regression line
  labs(title = "Scatter Plot: Work/Study Hours vs Total Stress",
       x = "Work/Study Hours",
       y = "Total Stress") +
  theme_minimal()

## `geom_smooth()` using formula = 'y ~ x'
```



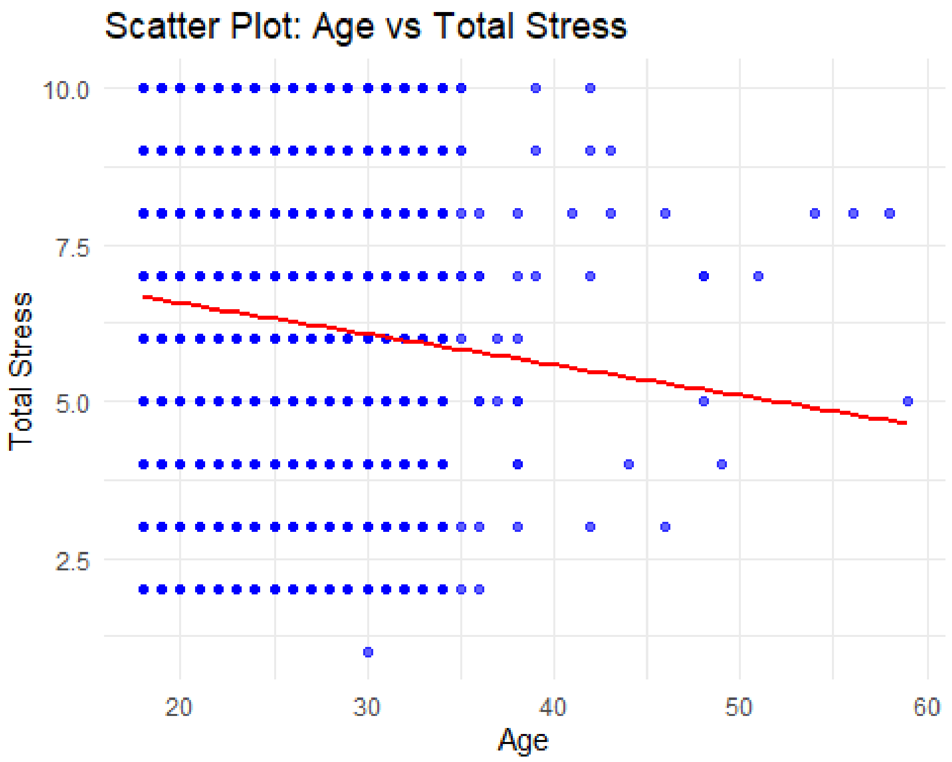
Interpretation:

The correlation between Work/Study Hours and Total Stress is 0.113, indicating a very weak positive correlation. This suggests that as work/study hours increase, total stress tends to increase slightly, but the relationship is not significant.

The scatter plot with the red regression line shows a slight upward trend, but the data points are widely spread, reinforcing the weak relationship between the two variables.

Q4: Correlation: Age vs Total Stress

```
# Step 1: Calculate Pearson Correlation
correlation_age_total_stress <- cor(data$age, data$total_stress, method =
"pearson")

# Step 2: Print Correlation
print(paste("Correlation between Age and Total Stress:",
correlation_age_total_stress))

## [1] "Correlation between Age and Total Stress: -0.11282592697228"

# Step 3: Visualization with Scatter Plot and Regression Line
ggplot(data, aes(x = age, y = total_stress)) +
  geom_point(color = "blue", alpha = 0.6) +   # Scatter points
  geom_smooth(method = "lm", se = FALSE, color = "red") +  # Linear
regression line
  labs(title = "Scatter Plot: Age vs Total Stress",
       x = "Age",
       y = "Total Stress") +
  theme_minimal()

## `geom_smooth()` using formula = 'y ~ x'
```



Interpretation:

The correlation between Age and Total Stress is -0.113, indicating a very weak negative correlation. This suggests that as age increases, total stress tends to decrease slightly, but the relationship is not significant.

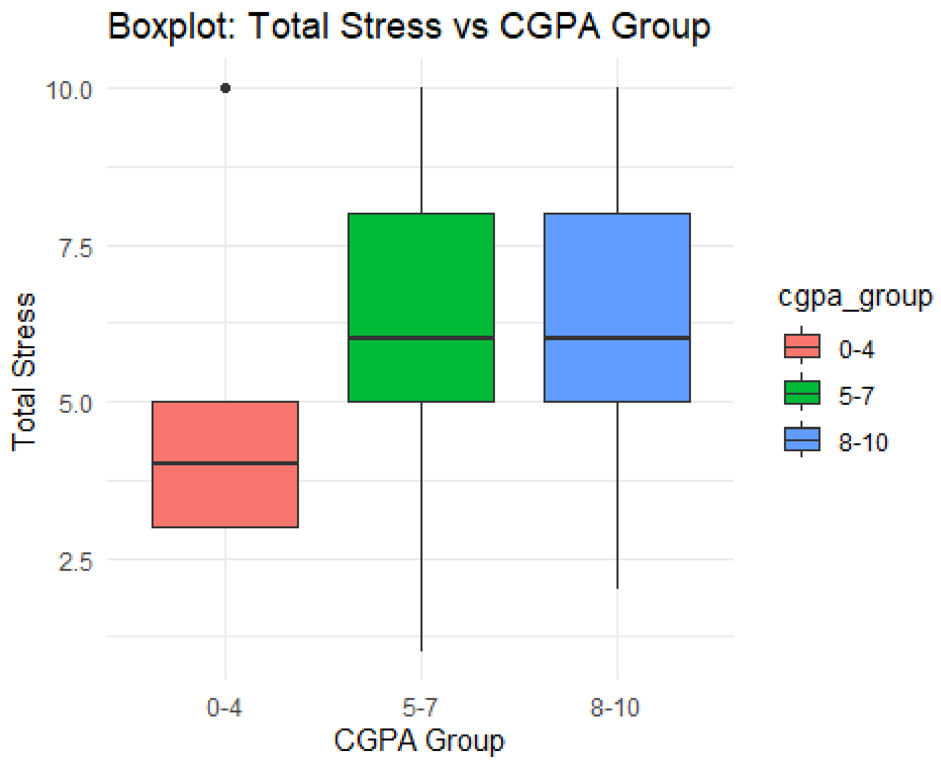
The scatter plot with the red regression line shows a slight downward trend, but the data points are scattered, indicating a very weak relationship between the two variables.

Level 8: ANOVA

Q1: ANOVA : Is there a significant difference in Total Stress across different CGPA Groups

```
# Step 1: Perform ANOVA to check the relationship between CGPA Groups and
Total Stress
anova_result <- aov(total_stress ~ cgpa_group, data = data)

# Step 2: Print the ANOVA summary
summary(anova_result)

##                  Df Sum Sq Mean Sq F value Pr(>F)
## cgpa_group        2     37  18.734   4.094 0.0167 *
## Residuals   27895 127643   4.576
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Step 3: Visualization - Boxplot of Total Stress by CGPA Groups
library(ggplot2)
ggplot(data, aes(x = cgpa_group, y = total_stress, fill = cgpa_group)) +
  geom_boxplot() +
  labs(title = "Boxplot: Total Stress vs CGPA Group", x = "CGPA Group", y =
"Total Stress") +
  theme_minimal()
```



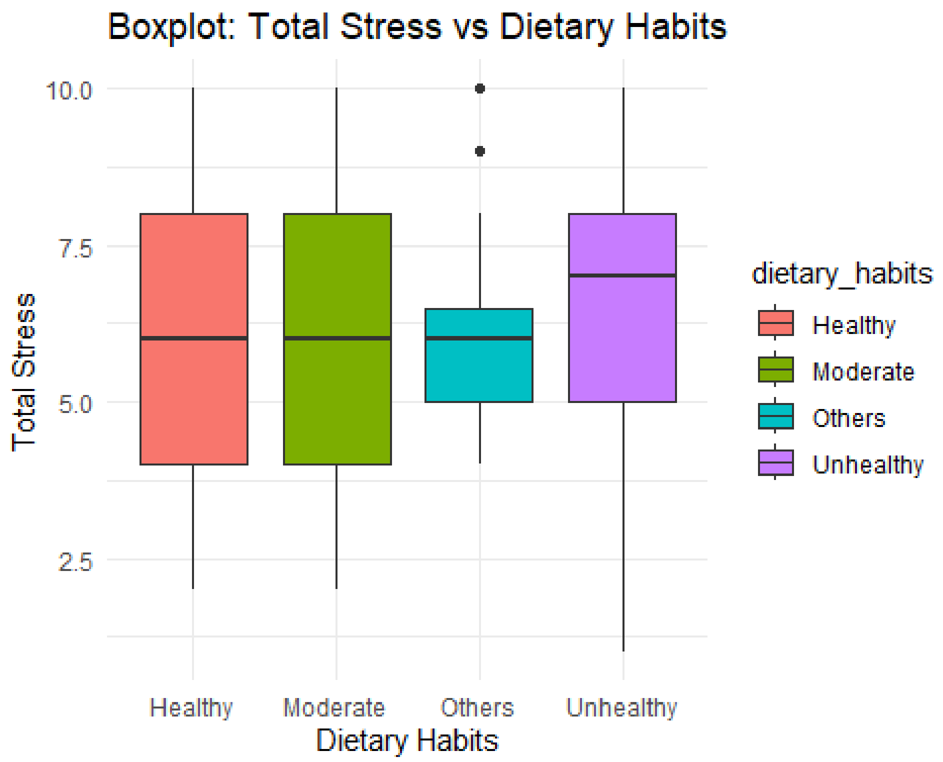
Interpretation:

The results of the ANOVA test show that there is a statistically significant difference in Total Stress across different CGPA Groups. The p-value is 0.0167, which is less than the significance level of 0.05, indicating that at least one group is significantly different from the others in terms of total stress.

The boxplot visualizes this difference, showing how total stress varies within each CGPA group.

Q2: ANOVA: Is there a significant difference in Total Stress across different Dietary Habits

```
# Step 1: Perform ANOVA
anova_dietary_stress <- aov(total_stress ~ dietary_habits, data = data)

# Step 2: Print ANOVA Summary
summary(anova_dietary_stress)

##                   Df Sum Sq Mean Sq F value Pr(>F)    
## dietary_habits     3   1832   610.5   135.3 <2e-16 ***
## Residuals      27894 125849     4.5                   
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Step 3: Visualization - Boxplot of Total Stress by Dietary Habits
library(ggplot2)
ggplot(data, aes(x = dietary_habits, y = total_stress, fill =
dietary_habits)) +
  geom_boxplot() +
  labs(title = "Boxplot: Total Stress vs Dietary Habits",
       x = "Dietary Habits",
       y = "Total Stress") +
  theme_minimal()
```



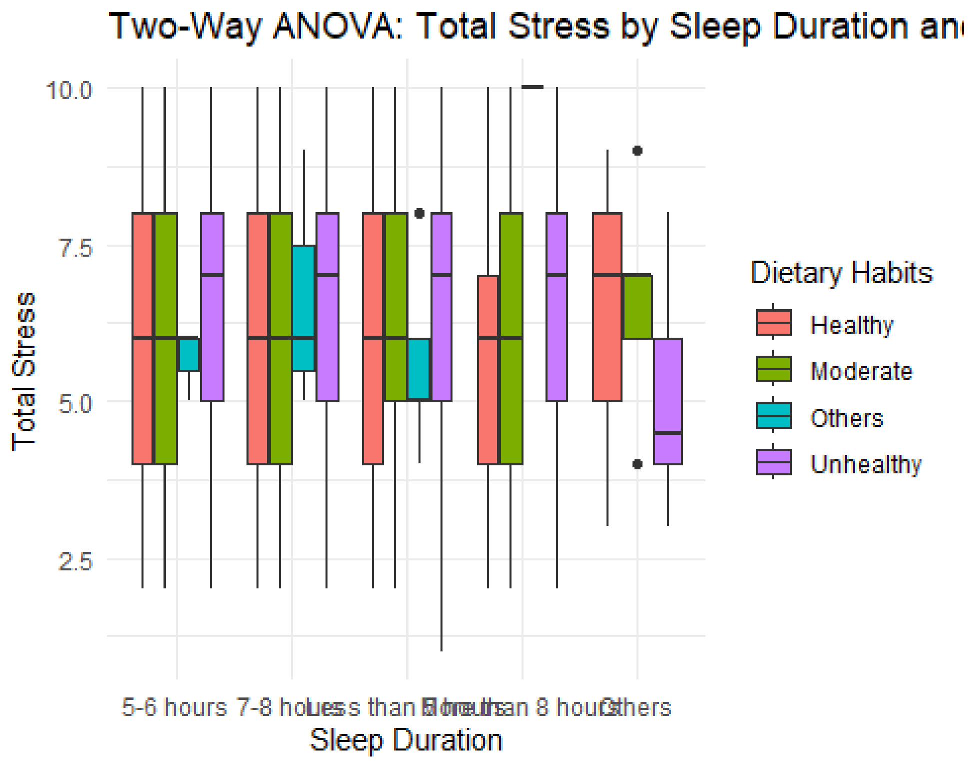
Interpretation:

The ANOVA test reveals a highly significant difference in Total Stress across different Dietary Habits, with a p-value of <2e-16. This p-value is much smaller than the significance level of 0.05, suggesting that dietary habits have a significant impact on total stress.

The boxplot visualizes this difference, showing how total stress varies across the different categories of dietary habits.

Q3: Two-Way ANOVA: Does Total Stress depend on Sleep Duration and Dietary Habits?

```
# Step 1: Perform Two-Way ANOVA
two_way_anova <- aov(total_stress ~ sleep_duration * dietary_habits, data =
data)

# Step 2: Print ANOVA Summary
summary(two_way_anova)

##                                  Df Sum Sq Mean Sq F value   Pr(>F)    
## sleep_duration                    4    193    48.1  10.687 1.18e-08 ***
## dietary_habits                    3   1823   607.7 134.903  < 2e-16 ***
## sleep_duration:dietary_habits    11     73     6.7   1.483     0.13    
## Residuals                     27879 125592     4.5                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

# Step 3: Visualization - Interaction Boxplot
library(ggplot2)
ggplot(data, aes(x = sleep_duration, y = total_stress, fill =
dietary_habits)) +
  geom_boxplot(position = position_dodge(0.8)) +
  labs(title = "Two-Way ANOVA: Total Stress by Sleep Duration and Dietary
Habits",
       x = "Sleep Duration",
       y = "Total Stress",
       fill = "Dietary Habits") +
  theme_minimal()
```



Interpretation:

The results of the Two-Way ANOVA show:

1. Sleep Duration significantly affects Total Stress with a p-value of 1.18e-08.

2. Dietary Habits also have a significant impact on Total Stress, with a p-value of <2e-16.

3. The interaction between Sleep Duration and Dietary Habits does not have a significant effect on Total Stress, as the p-value is 0.13 (greater than 0.05).

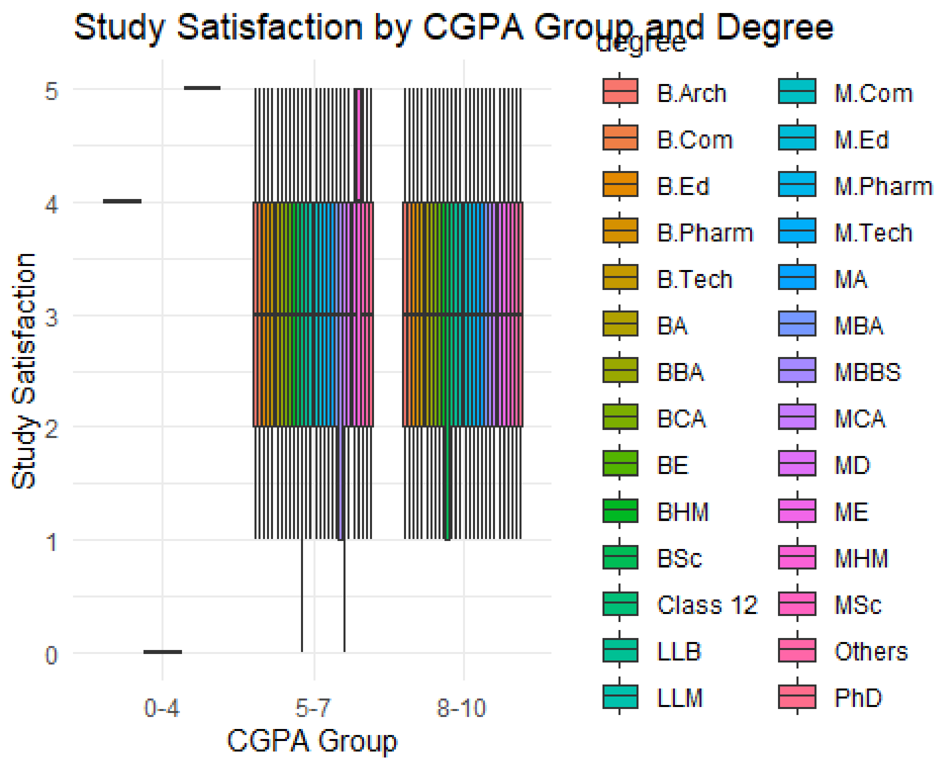
The interaction boxplot provides a visual representation of how the combination of sleep duration and dietary habits influences total stress.

Q4: Two-Way ANOVA: Does Study Satisfaction depend on CGPA Group and Degree?

```
# Step 1: Perform Two-Way ANOVA
two_way_anova <- aov(study_satisfaction ~ cgpa_group * degree, data = data)

# Step 2: Print ANOVA Summary
summary(two_way_anova)
```

```
##                       Df Sum Sq Mean Sq F value   Pr(>F)    
## cgpa_group             2     73   36.51  19.848 2.43e-09 ***
## degree                27    288   10.65   5.790  < 2e-16 ***
## cgpa_group:degree     29    116    4.01   2.181 0.000242 ***
## Residuals          27839  51206    1.84                     
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Step 3: Visualization - Interaction Boxplot
ggplot(data, aes(x = cgpa_group, y = study_satisfaction, fill = degree)) +
  geom_boxplot(position = position_dodge(0.8)) +
  labs(title = "Study Satisfaction by CGPA Group and Degree", x = "CGPA
Group", y = "Study Satisfaction") +
  theme_minimal()
```



Interpretation:

The results of the Two-Way ANOVA show:

1. CGPA Group significantly affects Study Satisfaction with a p-value of 2.43e-09.

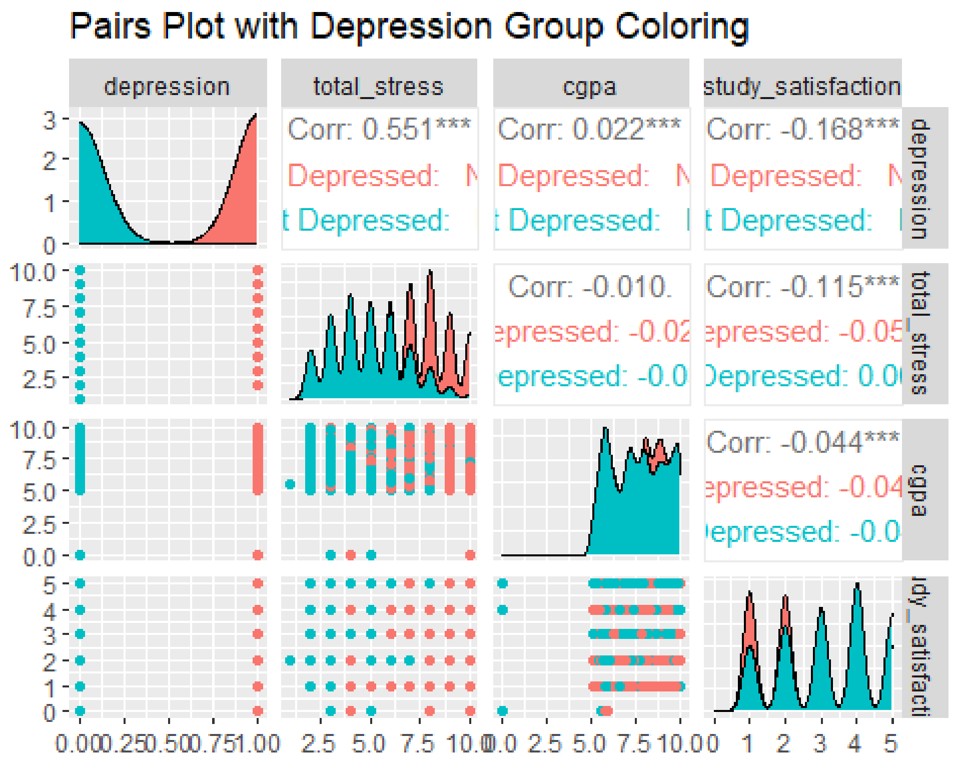
2. Degree also significantly impacts Study Satisfaction with a p-value of <2e-16.

3. The interaction between CGPA Group and Degree has a significant effect on Study Satisfaction, with a p-value of 0.000242, indicating that the impact of CGPA on satisfaction varies across different degrees.

The interaction boxplot visualizes how the combination of CGPA group and degree influences study satisfaction.

Pair Plot

```
library(GGally)

## Registered S3 method overwritten by 'GGally':
##   method from
##   +.gg   ggplot2

# Step 1: Create grouping
data$depression_group <- ifelse(data$depression == 1, "Depressed", "Not
Depressed")

# Step 2: Plot
ggpairs(
  data,
  columns = c("depression", "total_stress", "cgpa", "study_satisfaction"),
  aes(color = depression_group),
  title = "Pairs Plot with Depression Group Coloring"
)

## Warning in cor(x, y): the standard deviation is zero
## Warning in cor(x, y): the standard deviation is zero
## Warning in cor(x, y): the standard deviation is zero
## Warning in cor(x, y): the standard deviation is zero
## Warning in cor(x, y): the standard deviation is zero
## Warning in cor(x, y): the standard deviation is zero
```



#Project Overview: *After completing the project, we were able to gather valuable insights into the factors contributing to student depression. By analyzing the dataset and performing correlation and statistical tests, the following conclusions were made:*

#Conclusion #1: Key Insights from the Analysis: #Biggest Factor Influencing Depression:

Study Satisfaction significantly influences depression. A negative correlation shows that lower study satisfaction leads to higher depression levels.

#Correlation Highlights:

#Work/Study Hours and Study Satisfaction: *A very weak negative correlation (-0.036) suggests that work/study hours do not greatly affect satisfaction.*

#Study Satisfaction and Depression: *A weak negative correlation (-0.168) means lower study satisfaction is slightly linked to higher depression.*

#Work/Study Hours and Total Stress: *A weak positive correlation (0.11) indicates that more work/study hours slightly increase stress levels*

#Age and Total Stress: *A small negative correlation (-0.11) suggests younger students may experience more stress*

#2: Statistical Results (ANOVA and Two-Way Analysis): #ANOVA Results for Total Stress:

#CGPA Groups: *Significant difference in total stress across CGPA groups (p = 0.0167), with lower CGPA students experiencing higher stress.*

#Dietary Habits: *Strong effect on total stress (p < 2e-16), indicating that eating habits significantly affect stress.*

#Sleep Duration and Dietary Habits: *Both factors significantly impact stress, but their interaction is not significant (p = 0.13).*

#Study Satisfaction and CGPA Group/Degree: *The two-way ANOVA shows that CGPA Group (p = 2.43e-09) and Degree (p < 2e-16) significantly affect Study Satisfaction, with their interaction also being significant (p = 0.000242). This means that students' satisfaction with their studies differs based on their CGPA and degree type.*

#3: Summary of Results: *The analysis highlights that study satisfaction and dietary habits play the most critical roles in influencing student depression and stress. The lower the study satisfaction, the higher the depression.*

CGPA and degree influence study satisfaction, while dietary habits and sleep duration significantly affect total stress.

4: Key Factors contributing to higher stress and depression:

#Age Group (18 years): *76% of students in this group are experiencing significant depression/stress.*

#CGPA (8-10): *60% of students with a CGPA between 8-10 report higher stress.*

#12th Grade Students: *70% report higher stress.*

#Male Students: *60% report higher depression and stress.*

#Financial and Academic Pressure: *Students facing both pressures report higher stress levels.*

#Sleep Duration: *Students who sleep less experience more stress.*

#Unhealthy Eating Habits: *Poor diet leads to higher stress levels.*

#Over-Studying: *Excessive studying increases stress and depression.*

#Low Study Satisfaction: *Students with low study satisfaction report higher depression levels.*

#Suicidal Thoughts: *13,957 students report having suicidal thoughts, indicating the severity of the issue*