

Part 2: Dataset Summary

Keith Lim

2025-11-03

```
library(tidyverse)
library(dplyr)
library(ggplot2)
library(readr)
library(tibble)

sdd <- read_csv("student_depression_dataset.csv")
```

```
## Warning: One or more parsing issues, call 'problems()' on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)
```

```
str(sdd) #check the data types
```

```
## spc_tbl_ [27,901 x 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 ...
##   $ 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'" ...
##   $ 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>
```

#Mean of entire dataset ##data cleaning: remove certain variables id is excluded since it is just a key for each row. 'Gender', 'City', 'Profession', 'Sleep Duration', 'Dietary Habits', 'Degree', 'Have you ever had suicidal thoughts', and 'Family History of Mental Illness' are excluded as they are categorical variables. 'Job satisfaction' and 'Work Pressure' contain only no other values except for 0 in the original dataset. They will be excluded. 'Depression' is not included; it is a binary variable.

*'Financial Stress' is labelled a Character data type but it contains numerical values, and hence will be included in the code chunk below.

```
#convert Financial Stress from Chr to Num

sdd <- sdd %>%
  mutate(`Financial Stress` = as.numeric(`Financial Stress`)) #<- use backticks if not R will see it as

#Find which columns are numeric
numeric_columns <- sapply(sdd, is.numeric)

#keep only columns with numeric values
numeric_data <- sdd[numeric_columns]

#Calculate means for those columns
means <- sapply(numeric_data, mean, na.rm = TRUE) #<-ignore missing values

options(scipen = 999) #<-turn off scientific formatting to prevent mean from showing e... (disable scie

means_sdd <- data.frame(
  Variable = names(means), #<- one column called variable, containing variable names (extract names of
  Mean = round(as.numeric(means), 2) #<-round to 2 decimals for clarity
) #one column called Mean, containing mean values

means_sdd <- means_sdd %>%
  filter(!Variable %in% c("id", "Job Satisfaction", "Work Pressure", "Depression"))
#filter out id, job satisfaction, work pressure, Depression

print(means_sdd)
```

```
##           Variable Mean
## 1           Age 25.82
```

```
## 2 Academic Pressure 3.14
## 3 CGPA 7.66
## 4 Study Satisfaction 2.94
## 5 Work/Study Hours 7.16
## 6 Financial Stress 3.14
```

#Finding median

Use the same code as seen above, but find median

```
#Calculate median for those columns
medians <- sapply(numeric_data, median, na.rm = TRUE) #<-ignore missing values

options(scipen = 999) #<-turn off scientific formatting to prevent mean from showing e... (disable sci

medians_sdd <- data.frame(
  Variable = names(medians), #<- one column called variable, containing variable names (extract names of
  Median = round(as.numeric(medians), 2) #<-round to 2 decimals for clarity
) #one column called median, containing median values

medians_sdd <- medians_sdd %>%
  filter(!Variable %in% c("id", "Job Satisfaction", "Work Pressure", "Depression"))
#filter out id, job satisfaction, work pressure, Depression

print(medians_sdd)
```

```
##          Variable Median
## 1           Age 25.00
## 2 Academic Pressure 3.00
## 3           CGPA 7.77
## 4 Study Satisfaction 3.00
## 5 Work/Study Hours 8.00
## 6 Financial Stress 3.00
```

#Calculate Standard Deviation

Use same code as above, but find standard deviation

```
#Calculate standard deviation (stand) for those columns
stand <- sapply(numeric_data, sd, na.rm = TRUE) #<-ignore missing values

options(scipen = 999) #<-turn off scientific formatting to prevent mean from showing e... (disable sci

stand_sdd <- data.frame(
  Variable = names(stand), #<- one column called variable, containing variable names (extract names of
  standard_deviation = round(as.numeric(stand), 2) #<-round to 2 decimals for clarity
) #one column called stand, containing standard deviation values

stand_sdd <- stand_sdd %>%
  filter(!Variable %in% c("id", "Job Satisfaction", "Work Pressure", "Depression"))
#filter out id, job satisfaction, work pressure, Depression

print(stand_sdd)
```

```
##           Variable standard_deviation
## 1           Age                4.91
## 2 Academic Pressure            1.38
## 3           CGPA                1.47
## 4 Study Satisfaction            1.36
## 5 Work/Study Hours             3.71
## 6 Financial Stress             1.44
```

#Find Maximum and Minimum

Use code from above, but to find max and min instead.

```
#Find min and max for numerical variables
mins <- sapply(numeric_data, min, na.rm = TRUE)
maxs <- sapply(numeric_data, max, na.rm = TRUE)

max_min_sdd <- data.frame(Variable = names(mins), #<- one column called variable, containing variable names
  Min = round(as.numeric(mins), 2),
  Max = round(as.numeric(maxs), 2)
) #<-round to 2 decimals for clarity
#one column called min, one called max

max_min_sdd <- max_min_sdd %>%
  filter(!Variable %in% c("id", "Job Satisfaction", "Work Pressure", "Depression"))
#filter out id, job satisfaction, work pressure, Depression

print(max_min_sdd)
```

```
##           Variable Min Max
## 1           Age    18  59
## 2 Academic Pressure    0   5
## 3           CGPA     0  10
## 4 Study Satisfaction    0   5
## 5 Work/Study Hours     0  12
## 6 Financial Stress     1   5
```

#Finding means, medians, standard deviations, and min/max for subgroups within the dataset

Subgroups include gender, city, degree, sleep duration, dietary habits, Have you ever had suicidal thoughts?, and Family history of mental illness *Profession is excluded as profession is the same throughout the dataset (students)

```
#try for gender subgroup

library(dplyr)

by_gender_stats <- sdd %>%
  group_by(Gender) %>%
  summarise(
    #Age
    mean_age = mean(Age, na.rm = TRUE),
    median_age = median(Age, na.rm = TRUE),
    sd_age = sd(Age, na.rm = TRUE),
    min_age = min(Age, na.rm = TRUE),
```

```

max_age = max(Age, na.rm = TRUE),

#Academic Pressure
mean_acad = mean(`Academic Pressure`, na.rm = TRUE),
median_acad = median(`Academic Pressure`, na.rm = TRUE),
sd_acad = sd(`Academic Pressure`, na.rm = TRUE),
min_acad = min(`Academic Pressure`, na.rm = TRUE),
max_acad = max(`Academic Pressure`, na.rm = TRUE),

#CGPA
mean_cgpa = mean(CGPA, na.rm = TRUE),
median_cgpa = median(CGPA, na.rm = TRUE),
sd_cgpa = sd(CGPA, na.rm = TRUE),
min_cgpa = min(CGPA, na.rm = TRUE),
max_cgpa = max(CGPA, na.rm = TRUE),

#Study Satisfaction
mean_study = mean(`Study Satisfaction`, na.rm = TRUE),
median_study = median(`Study Satisfaction`, na.rm = TRUE),
sd_study = sd(`Study Satisfaction`, na.rm = TRUE),
min_study = min(`Study Satisfaction`, na.rm = TRUE),
max_study = max(`Study Satisfaction`, na.rm = TRUE),

#Work/Study Hours
mean_hours = mean(`Work/Study Hours`, na.rm = TRUE),
median_hours = median(`Work/Study Hours`, na.rm = TRUE),
sd_hours = sd(`Work/Study Hours`, na.rm = TRUE),
min_hours = min(`Work/Study Hours`, na.rm = TRUE),
max_hours = max(`Work/Study Hours`, na.rm = TRUE),

#Financial Stress
mean_stress = mean(`Financial Stress`, na.rm = TRUE),
median_stress = median(`Financial Stress`, na.rm = TRUE),
sd_stress = sd(`Financial Stress`, na.rm = TRUE),
min_stress = min(`Financial Stress`, na.rm = TRUE),
max_stress = max(`Financial Stress`, na.rm = TRUE)
)

```

```
by_gender_stats
```

```

## # A tibble: 2 x 31
##   Gender mean_age median_age sd_age min_age max_age mean_acad median_acad
##   <chr>      <dbl>      <dbl> <dbl>  <dbl>  <dbl>      <dbl>      <dbl>
## 1 Female    25.8         25  4.90    18    58      3.18         3
## 2 Male     25.9         26  4.91    18    59      3.11         3
## # i 23 more variables: sd_acad <dbl>, min_acad <dbl>, max_acad <dbl>,
## #   mean_cgpa <dbl>, median_cgpa <dbl>, sd_cgpa <dbl>, min_cgpa <dbl>,
## #   max_cgpa <dbl>, mean_study <dbl>, median_study <dbl>, sd_study <dbl>,
## #   min_study <dbl>, max_study <dbl>, mean_hours <dbl>, median_hours <dbl>,
## #   sd_hours <dbl>, min_hours <dbl>, max_hours <dbl>, mean_stress <dbl>,
## #   median_stress <dbl>, sd_stress <dbl>, min_stress <dbl>, max_stress <dbl>

```

```
print(by_gender_stats)
```

```
## # A tibble: 2 x 31
##   Gender mean_age median_age sd_age min_age max_age mean_acad median_acad
##   <chr>      <dbl>      <dbl> <dbl>  <dbl>  <dbl>      <dbl>      <dbl>
## 1 Female    25.8        25   4.90    18    58      3.18        3
## 2 Male     25.9        26   4.91    18    59      3.11        3
## # i 23 more variables: sd_acad <dbl>, min_acad <dbl>, max_acad <dbl>,
## #   mean_cgpa <dbl>, median_cgpa <dbl>, sd_cgpa <dbl>, min_cgpa <dbl>,
## #   max_cgpa <dbl>, mean_study <dbl>, median_study <dbl>, sd_study <dbl>,
## #   min_study <dbl>, max_study <dbl>, mean_hours <dbl>, median_hours <dbl>,
## #   sd_hours <dbl>, min_hours <dbl>, max_hours <dbl>, mean_stress <dbl>,
## #   median_stress <dbl>, sd_stress <dbl>, min_stress <dbl>, max_stress <dbl>
```

```
#for city subgroup
```

```
by_city_stats <- sdd %>%
  group_by(City) %>%
  summarise(
    #Age
    mean_age = mean(Age, na.rm = TRUE),
    median_age = median(Age, na.rm = TRUE),
    sd_age = sd(Age, na.rm = TRUE),
    min_age = min(Age, na.rm = TRUE),
    max_age = max(Age, na.rm = TRUE),

    #Academic Pressure
    mean_acad = mean(`Academic Pressure`, na.rm = TRUE),
    median_acad = median(`Academic Pressure`, na.rm = TRUE),
    sd_acad = sd(`Academic Pressure`, na.rm = TRUE),
    min_acad = min(`Academic Pressure`, na.rm = TRUE),
    max_acad = max(`Academic Pressure`, na.rm = TRUE),

    #CGPA
    mean_cgpa = mean(CGPA, na.rm = TRUE),
    median_cgpa = median(CGPA, na.rm = TRUE),
    sd_cgpa = sd(CGPA, na.rm = TRUE),
    min_cgpa = min(CGPA, na.rm = TRUE),
    max_cgpa = max(CGPA, na.rm = TRUE),

    #Study Satisfaction
    mean_study = mean(`Study Satisfaction`, na.rm = TRUE),
    median_study = median(`Study Satisfaction`, na.rm = TRUE),
    sd_study = sd(`Study Satisfaction`, na.rm = TRUE),
    min_study = min(`Study Satisfaction`, na.rm = TRUE),
    max_study = max(`Study Satisfaction`, na.rm = TRUE),

    #Work/Study Hours
    mean_hours = mean(`Work/Study Hours`, na.rm = TRUE),
    median_hours = median(`Work/Study Hours`, na.rm = TRUE),
    sd_hours = sd(`Work/Study Hours`, na.rm = TRUE),
    min_hours = min(`Work/Study Hours`, na.rm = TRUE),
    max_hours = max(`Work/Study Hours`, na.rm = TRUE),
```

```

#Financial Stress
mean_stress = mean(`Financial Stress`, na.rm = TRUE),
median_stress = median(`Financial Stress`, na.rm = TRUE),
sd_stress = sd(`Financial Stress`, na.rm = TRUE),
min_stress = min(`Financial Stress`, na.rm = TRUE),
max_stress = max(`Financial Stress`, na.rm = TRUE)
)

by_city_stats

## # A tibble: 52 x 31
##   City          mean_age median_age sd_age min_age max_age mean_acad median_acad
##   <chr>          <dbl>      <dbl> <dbl> <dbl> <dbl>      <dbl>      <dbl>
## 1 'Less Delhi'    29         29  NA    29    29         2         2
## 2 'Less than ~    31         31  NA    31    31         1         1
## 3 3.0            21         21  NA    21    21         3         3
## 4 Agra           25.5        25   4.85   18    54        2.98        3
## 5 Ahmedabad      24.8        24   5.04   18    34        3.25        3
## 6 Bangalore      24.8        24   5.12   18    37        3.19        3
## 7 Bhavna         24         24   8.49   18    30        3.5         3.5
## 8 Bhopal          25.5        25   4.96   18    51        3.11        3
## 9 Chennai        25.1        25   5.14   18    58        3.30        3
## 10 City          28.5        28.5 0.707   28    29         3         3
## # i 42 more rows
## # i 23 more variables: sd_acad <dbl>, min_acad <dbl>, max_acad <dbl>,
## #   mean_cgpa <dbl>, median_cgpa <dbl>, sd_cgpa <dbl>, min_cgpa <dbl>,
## #   max_cgpa <dbl>, mean_study <dbl>, median_study <dbl>, sd_study <dbl>,
## #   min_study <dbl>, max_study <dbl>, mean_hours <dbl>, median_hours <dbl>,
## #   sd_hours <dbl>, min_hours <dbl>, max_hours <dbl>, mean_stress <dbl>,
## #   median_stress <dbl>, sd_stress <dbl>, min_stress <dbl>, ...

print(by_city_stats)

## # A tibble: 52 x 31
##   City          mean_age median_age sd_age min_age max_age mean_acad median_acad
##   <chr>          <dbl>      <dbl> <dbl> <dbl> <dbl>      <dbl>      <dbl>
## 1 'Less Delhi'    29         29  NA    29    29         2         2
## 2 'Less than ~    31         31  NA    31    31         1         1
## 3 3.0            21         21  NA    21    21         3         3
## 4 Agra           25.5        25   4.85   18    54        2.98        3
## 5 Ahmedabad      24.8        24   5.04   18    34        3.25        3
## 6 Bangalore      24.8        24   5.12   18    37        3.19        3
## 7 Bhavna         24         24   8.49   18    30        3.5         3.5
## 8 Bhopal          25.5        25   4.96   18    51        3.11        3
## 9 Chennai        25.1        25   5.14   18    58        3.30        3
## 10 City          28.5        28.5 0.707   28    29         3         3
## # i 42 more rows
## # i 23 more variables: sd_acad <dbl>, min_acad <dbl>, max_acad <dbl>,
## #   mean_cgpa <dbl>, median_cgpa <dbl>, sd_cgpa <dbl>, min_cgpa <dbl>,
## #   max_cgpa <dbl>, mean_study <dbl>, median_study <dbl>, sd_study <dbl>,
## #   min_study <dbl>, max_study <dbl>, mean_hours <dbl>, median_hours <dbl>,
## #   sd_hours <dbl>, min_hours <dbl>, max_hours <dbl>, mean_stress <dbl>,
## #   median_stress <dbl>, sd_stress <dbl>, min_stress <dbl>, ...

```

```

#for degree

by_degree_stats <- sdd %>%
  group_by(Degree) %>%
  summarise(
    #Age
    mean_age = mean(Age, na.rm = TRUE),
    median_age = median(Age, na.rm = TRUE),
    sd_age = sd(Age, na.rm = TRUE),
    min_age = min(Age, na.rm = TRUE),
    max_age = max(Age, na.rm = TRUE),

    #Academic Pressure
    mean_acad = mean(`Academic Pressure`, na.rm = TRUE),
    median_acad = median(`Academic Pressure`, na.rm = TRUE),
    sd_acad = sd(`Academic Pressure`, na.rm = TRUE),
    min_acad = min(`Academic Pressure`, na.rm = TRUE),
    max_acad = max(`Academic Pressure`, na.rm = TRUE),

    #CGPA
    mean_cgpa = mean(CGPA, na.rm = TRUE),
    median_cgpa = median(CGPA, na.rm = TRUE),
    sd_cgpa = sd(CGPA, na.rm = TRUE),
    min_cgpa = min(CGPA, na.rm = TRUE),
    max_cgpa = max(CGPA, na.rm = TRUE),

    #Study Satisfaction
    mean_study = mean(`Study Satisfaction`, na.rm = TRUE),
    median_study = median(`Study Satisfaction`, na.rm = TRUE),
    sd_study = sd(`Study Satisfaction`, na.rm = TRUE),
    min_study = min(`Study Satisfaction`, na.rm = TRUE),
    max_study = max(`Study Satisfaction`, na.rm = TRUE),

    #Work/Study Hours
    mean_hours = mean(`Work/Study Hours`, na.rm = TRUE),
    median_hours = median(`Work/Study Hours`, na.rm = TRUE),
    sd_hours = sd(`Work/Study Hours`, na.rm = TRUE),
    min_hours = min(`Work/Study Hours`, na.rm = TRUE),
    max_hours = max(`Work/Study Hours`, na.rm = TRUE),

    #Financial Stress
    mean_stress = mean(`Financial Stress`, na.rm = TRUE),
    median_stress = median(`Financial Stress`, na.rm = TRUE),
    sd_stress = sd(`Financial Stress`, na.rm = TRUE),
    min_stress = min(`Financial Stress`, na.rm = TRUE),
    max_stress = max(`Financial Stress`, na.rm = TRUE)
  )

by_degree_stats

```

```

## # A tibble: 28 x 31
##   Degree      mean_age median_age sd_age min_age max_age mean_acad median_acad
##   <chr>         <dbl>     <dbl> <dbl>  <dbl>  <dbl>     <dbl>     <dbl>

```

```
## 1 'Class 12'      20.1      19  3.24      18      58      3.36      3
## 2 B.Arch          25.6      25  3.81      19      38      3.06      3
## 3 B.Com           26.2      26  3.92      20      36      3.08      3
## 4 B.Ed            28.0      28  3.90      18      54      3.12      3
## 5 B.Pharm         27.3      27  3.27      21      34      3.19      3
## 6 B.Tech          27.1      27  4.07      18      34      3.08      3
## 7 BA              26.0      25  3.85      18      34      3.11      3
## 8 BBA             26.3      26  3.70      21      34      3.07      3
## 9 BCA             27.0      27  4.25      18      34      3.13      3
## 10 BE            28.2      28  3.80      20      34      3.09      3
## # i 18 more rows
## # i 23 more variables: sd_acad <dbl>, min_acad <dbl>, max_acad <dbl>,
## #   mean_cgpa <dbl>, median_cgpa <dbl>, sd_cgpa <dbl>, min_cgpa <dbl>,
## #   max_cgpa <dbl>, mean_study <dbl>, median_study <dbl>, sd_study <dbl>,
## #   min_study <dbl>, max_study <dbl>, mean_hours <dbl>, median_hours <dbl>,
## #   sd_hours <dbl>, min_hours <dbl>, max_hours <dbl>, mean_stress <dbl>,
## #   median_stress <dbl>, sd_stress <dbl>, min_stress <dbl>, ...
```

```
print(by_degree_stats)
```

```
## # A tibble: 28 x 31
##   Degree      mean_age median_age sd_age min_age max_age mean_acad median_acad
##   <chr>      <dbl>      <dbl> <dbl> <dbl> <dbl>      <dbl>      <dbl>
## 1 'Class 12'    20.1        19  3.24    18    58      3.36        3
## 2 B.Arch       25.6        25  3.81    19    38      3.06        3
## 3 B.Com        26.2        26  3.92    20    36      3.08        3
## 4 B.Ed         28.0        28  3.90    18    54      3.12        3
## 5 B.Pharm      27.3        27  3.27    21    34      3.19        3
## 6 B.Tech       27.1        27  4.07    18    34      3.08        3
## 7 BA           26.0        25  3.85    18    34      3.11        3
## 8 BBA          26.3        26  3.70    21    34      3.07        3
## 9 BCA          27.0        27  4.25    18    34      3.13        3
## 10 BE         28.2        28  3.80    20    34      3.09        3
## # i 18 more rows
## # i 23 more variables: sd_acad <dbl>, min_acad <dbl>, max_acad <dbl>,
## #   mean_cgpa <dbl>, median_cgpa <dbl>, sd_cgpa <dbl>, min_cgpa <dbl>,
## #   max_cgpa <dbl>, mean_study <dbl>, median_study <dbl>, sd_study <dbl>,
## #   min_study <dbl>, max_study <dbl>, mean_hours <dbl>, median_hours <dbl>,
## #   sd_hours <dbl>, min_hours <dbl>, max_hours <dbl>, mean_stress <dbl>,
## #   median_stress <dbl>, sd_stress <dbl>, min_stress <dbl>, ...
```

```
#for sleep duration subgroup
```

```
by_sleep_stats <- sdd %>%
  group_by(`Sleep Duration`) %>%
  summarise(
    #Age
    mean_age = mean(Age, na.rm = TRUE),
    median_age = median(Age, na.rm = TRUE),
    sd_age = sd(Age, na.rm = TRUE),
    min_age = min(Age, na.rm = TRUE),
    max_age = max(Age, na.rm = TRUE),
```

```

#Academic Pressure
mean_acad = mean(`Academic Pressure`, na.rm = TRUE),
median_acad = median(`Academic Pressure`, na.rm = TRUE),
sd_acad = sd(`Academic Pressure`, na.rm = TRUE),
min_acad = min(`Academic Pressure`, na.rm = TRUE),
max_acad = max(`Academic Pressure`, na.rm = TRUE),

#CGPA
mean_cgpa = mean(CGPA, na.rm = TRUE),
median_cgpa = median(CGPA, na.rm = TRUE),
sd_cgpa = sd(CGPA, na.rm = TRUE),
min_cgpa = min(CGPA, na.rm = TRUE),
max_cgpa = max(CGPA, na.rm = TRUE),

#Study Satisfaction
mean_study = mean(`Study Satisfaction`, na.rm = TRUE),
median_study = median(`Study Satisfaction`, na.rm = TRUE),
sd_study = sd(`Study Satisfaction`, na.rm = TRUE),
min_study = min(`Study Satisfaction`, na.rm = TRUE),
max_study = max(`Study Satisfaction`, na.rm = TRUE),

#Work/Study Hours
mean_hours = mean(`Work/Study Hours`, na.rm = TRUE),
median_hours = median(`Work/Study Hours`, na.rm = TRUE),
sd_hours = sd(`Work/Study Hours`, na.rm = TRUE),
min_hours = min(`Work/Study Hours`, na.rm = TRUE),
max_hours = max(`Work/Study Hours`, na.rm = TRUE),

#Financial Stress
mean_stress = mean(`Financial Stress`, na.rm = TRUE),
median_stress = median(`Financial Stress`, na.rm = TRUE),
sd_stress = sd(`Financial Stress`, na.rm = TRUE),
min_stress = min(`Financial Stress`, na.rm = TRUE),
max_stress = max(`Financial Stress`, na.rm = TRUE)
)

```

by_sleep_stats

```

## # A tibble: 5 x 31
##   'Sleep Duration'    mean_age median_age sd_age min_age max_age mean_acad
##   <chr>              <dbl>      <dbl> <dbl> <dbl>  <dbl>    <dbl>
## 1 '5-6 hours'        25.9        26  4.93   18    59     3.12
## 2 '7-8 hours'        25.8        25  4.90   18    58     3.14
## 3 'Less than 5 hours' 25.8        26  4.86   18    51     3.23
## 4 'More than 8 hours' 25.8        25  4.95   18    54     3.05
## 5 Others            27.3        28  4.65   18    34     3.17
## # i 24 more variables: median_acad <dbl>, sd_acad <dbl>, min_acad <dbl>,
## #   max_acad <dbl>, mean_cgpa <dbl>, median_cgpa <dbl>, sd_cgpa <dbl>,
## #   min_cgpa <dbl>, max_cgpa <dbl>, mean_study <dbl>, median_study <dbl>,
## #   sd_study <dbl>, min_study <dbl>, max_study <dbl>, mean_hours <dbl>,
## #   median_hours <dbl>, sd_hours <dbl>, min_hours <dbl>, max_hours <dbl>,
## #   mean_stress <dbl>, median_stress <dbl>, sd_stress <dbl>, min_stress <dbl>,
## #   max_stress <dbl>

```

```
print(by_sleep_stats)
```

```
## # A tibble: 5 x 31
##   'Sleep Duration'    mean_age median_age sd_age min_age max_age mean_acad
##   <chr>              <dbl>      <dbl> <dbl> <dbl> <dbl>    <dbl>
## 1 '5-6 hours'        25.9        26  4.93   18    59     3.12
## 2 '7-8 hours'        25.8        25  4.90   18    58     3.14
## 3 'Less than 5 hours' 25.8        26  4.86   18    51     3.23
## 4 'More than 8 hours' 25.8        25  4.95   18    54     3.05
## 5 Others            27.3        28  4.65   18    34     3.17
## # i 24 more variables: median_acad <dbl>, sd_acad <dbl>, min_acad <dbl>,
## #   max_acad <dbl>, mean_cgpa <dbl>, median_cgpa <dbl>, sd_cgpa <dbl>,
## #   min_cgpa <dbl>, max_cgpa <dbl>, mean_study <dbl>, median_study <dbl>,
## #   sd_study <dbl>, min_study <dbl>, max_study <dbl>, mean_hours <dbl>,
## #   median_hours <dbl>, sd_hours <dbl>, min_hours <dbl>, max_hours <dbl>,
## #   mean_stress <dbl>, median_stress <dbl>, sd_stress <dbl>, min_stress <dbl>,
## #   max_stress <dbl>
```

```
#for dietary habits subgroup
```

```
by_diet_stats <- sdd %>%
  group_by(`Dietary Habits`) %>%
  summarise(
    #Age
    mean_age = mean(Age, na.rm = TRUE),
    median_age = median(Age, na.rm = TRUE),
    sd_age = sd(Age, na.rm = TRUE),
    min_age = min(Age, na.rm = TRUE),
    max_age = max(Age, na.rm = TRUE),

    #Academic Pressure
    mean_acad = mean(`Academic Pressure`, na.rm = TRUE),
    median_acad = median(`Academic Pressure`, na.rm = TRUE),
    sd_acad = sd(`Academic Pressure`, na.rm = TRUE),
    min_acad = min(`Academic Pressure`, na.rm = TRUE),
    max_acad = max(`Academic Pressure`, na.rm = TRUE),

    #CGPA
    mean_cgpa = mean(CGPA, na.rm = TRUE),
    median_cgpa = median(CGPA, na.rm = TRUE),
    sd_cgpa = sd(CGPA, na.rm = TRUE),
    min_cgpa = min(CGPA, na.rm = TRUE),
    max_cgpa = max(CGPA, na.rm = TRUE),

    #Study Satisfaction
    mean_study = mean(`Study Satisfaction`, na.rm = TRUE),
    median_study = median(`Study Satisfaction`, na.rm = TRUE),
    sd_study = sd(`Study Satisfaction`, na.rm = TRUE),
    min_study = min(`Study Satisfaction`, na.rm = TRUE),
    max_study = max(`Study Satisfaction`, na.rm = TRUE),

    #Work/Study Hours
    mean_hours = mean(`Work/Study Hours`, na.rm = TRUE),
```

```

median_hours = median(`Work/Study Hours`, na.rm = TRUE),
sd_hours = sd(`Work/Study Hours`, na.rm = TRUE),
min_hours = min(`Work/Study Hours`, na.rm = TRUE),
max_hours = max(`Work/Study Hours`, na.rm = TRUE),

#Financial Stress
mean_stress = mean(`Financial Stress`, na.rm = TRUE),
median_stress = median(`Financial Stress`, na.rm = TRUE),
sd_stress = sd(`Financial Stress`, na.rm = TRUE),
min_stress = min(`Financial Stress`, na.rm = TRUE),
max_stress = max(`Financial Stress`, na.rm = TRUE)
)

by_diet_stats

## # A tibble: 4 x 31
##   'Dietary Habits' mean_age median_age sd_age min_age max_age mean_acad
##   <chr>           <dbl>      <dbl> <dbl>   <dbl>   <dbl>   <dbl>
## 1 Healthy          26.1        26    4.96    18     58     2.99
## 2 Moderate          26.0        26    4.94    18     51     3.09
## 3 Others            26.4       25.5    5.14    18     34     3.08
## 4 Unhealthy         25.4        25    4.81    18     59     3.30
## # i 24 more variables: median_acad <dbl>, sd_acad <dbl>, min_acad <dbl>,
## #   max_acad <dbl>, mean_cgpa <dbl>, median_cgpa <dbl>, sd_cgpa <dbl>,
## #   min_cgpa <dbl>, max_cgpa <dbl>, mean_study <dbl>, median_study <dbl>,
## #   sd_study <dbl>, min_study <dbl>, max_study <dbl>, mean_hours <dbl>,
## #   median_hours <dbl>, sd_hours <dbl>, min_hours <dbl>, max_hours <dbl>,
## #   mean_stress <dbl>, median_stress <dbl>, sd_stress <dbl>, min_stress <dbl>,
## #   max_stress <dbl>

print(by_diet_stats)

## # A tibble: 4 x 31
##   'Dietary Habits' mean_age median_age sd_age min_age max_age mean_acad
##   <chr>           <dbl>      <dbl> <dbl>   <dbl>   <dbl>   <dbl>
## 1 Healthy          26.1        26    4.96    18     58     2.99
## 2 Moderate          26.0        26    4.94    18     51     3.09
## 3 Others            26.4       25.5    5.14    18     34     3.08
## 4 Unhealthy         25.4        25    4.81    18     59     3.30
## # i 24 more variables: median_acad <dbl>, sd_acad <dbl>, min_acad <dbl>,
## #   max_acad <dbl>, mean_cgpa <dbl>, median_cgpa <dbl>, sd_cgpa <dbl>,
## #   min_cgpa <dbl>, max_cgpa <dbl>, mean_study <dbl>, median_study <dbl>,
## #   sd_study <dbl>, min_study <dbl>, max_study <dbl>, mean_hours <dbl>,
## #   median_hours <dbl>, sd_hours <dbl>, min_hours <dbl>, max_hours <dbl>,
## #   mean_stress <dbl>, median_stress <dbl>, sd_stress <dbl>, min_stress <dbl>,
## #   max_stress <dbl>

#for suicidal thoughts subgroup

by_suicide_stats <- sdd %>%
  group_by(`Have you ever had suicidal thoughts ?`) %>%
  summarise(

```

```

#Age
mean_age = mean(Age, na.rm = TRUE),
median_age = median(Age, na.rm = TRUE),
sd_age = sd(Age, na.rm = TRUE),
min_age = min(Age, na.rm = TRUE),
max_age = max(Age, na.rm = TRUE),

#Academic Pressure
mean_acad = mean(`Academic Pressure`, na.rm = TRUE),
median_acad = median(`Academic Pressure`, na.rm = TRUE),
sd_acad = sd(`Academic Pressure`, na.rm = TRUE),
min_acad = min(`Academic Pressure`, na.rm = TRUE),
max_acad = max(`Academic Pressure`, na.rm = TRUE),

#CGPA
mean_cgpa = mean(CGPA, na.rm = TRUE),
median_cgpa = median(CGPA, na.rm = TRUE),
sd_cgpa = sd(CGPA, na.rm = TRUE),
min_cgpa = min(CGPA, na.rm = TRUE),
max_cgpa = max(CGPA, na.rm = TRUE),

#Study Satisfaction
mean_study = mean(`Study Satisfaction`, na.rm = TRUE),
median_study = median(`Study Satisfaction`, na.rm = TRUE),
sd_study = sd(`Study Satisfaction`, na.rm = TRUE),
min_study = min(`Study Satisfaction`, na.rm = TRUE),
max_study = max(`Study Satisfaction`, na.rm = TRUE),

#Work/Study Hours
mean_hours = mean(`Work/Study Hours`, na.rm = TRUE),
median_hours = median(`Work/Study Hours`, na.rm = TRUE),
sd_hours = sd(`Work/Study Hours`, na.rm = TRUE),
min_hours = min(`Work/Study Hours`, na.rm = TRUE),
max_hours = max(`Work/Study Hours`, na.rm = TRUE),

#Financial Stress
mean_stress = mean(`Financial Stress`, na.rm = TRUE),
median_stress = median(`Financial Stress`, na.rm = TRUE),
sd_stress = sd(`Financial Stress`, na.rm = TRUE),
min_stress = min(`Financial Stress`, na.rm = TRUE),
max_stress = max(`Financial Stress`, na.rm = TRUE)
)

by_suicide_stats

## # A tibble: 2 x 31
##   Have you ever had suici~1 mean_age median_age sd_age min_age max_age mean_acad
##   <chr>                <dbl>      <dbl> <dbl>   <dbl>   <dbl>      <dbl>
## 1 No                    26.6         27  4.99     18     58        2.67
## 2 Yes                   25.4         25  4.80     18     59        3.42
## # i abbreviated name: 1: 'Have you ever had suicidal thoughts ?'
## # i 24 more variables: median_acad <dbl>, sd_acad <dbl>, min_acad <dbl>,
## #   max_acad <dbl>, mean_cgpa <dbl>, median_cgpa <dbl>, sd_cgpa <dbl>,

```

```
## # min_cgpa <dbl>, max_cgpa <dbl>, mean_study <dbl>, median_study <dbl>,
## # sd_study <dbl>, min_study <dbl>, max_study <dbl>, mean_hours <dbl>,
## # median_hours <dbl>, sd_hours <dbl>, min_hours <dbl>, max_hours <dbl>,
## # mean_stress <dbl>, median_stress <dbl>, sd_stress <dbl>, ...
```

```
print(by_suicide_stats)
```

```
## # A tibble: 2 x 31
##   Have you ever had suici~1 mean_age median_age sd_age min_age max_age mean_acad
##   <chr>                <dbl>      <dbl>  <dbl>  <dbl>  <dbl>      <dbl>
## 1 No                    26.6        27   4.99    18    58        2.67
## 2 Yes                   25.4        25   4.80    18    59        3.42
## # i abbreviated name: 1: 'Have you ever had suicidal thoughts ?'
## # i 24 more variables: median_acad <dbl>, sd_acad <dbl>, min_acad <dbl>,
## # max_acad <dbl>, mean_cgpa <dbl>, median_cgpa <dbl>, sd_cgpa <dbl>,
## # min_cgpa <dbl>, max_cgpa <dbl>, mean_study <dbl>, median_study <dbl>,
## # sd_study <dbl>, min_study <dbl>, max_study <dbl>, mean_hours <dbl>,
## # median_hours <dbl>, sd_hours <dbl>, min_hours <dbl>, max_hours <dbl>,
## # mean_stress <dbl>, median_stress <dbl>, sd_stress <dbl>, ...
```

```
#for family history of mental illness
```

```
by_history_stats <- sdd %>%
  group_by(`Family History of Mental Illness`) %>%
  summarise(
    #Age
    mean_age = mean(Age, na.rm = TRUE),
    median_age = median(Age, na.rm = TRUE),
    sd_age = sd(Age, na.rm = TRUE),
    min_age = min(Age, na.rm = TRUE),
    max_age = max(Age, na.rm = TRUE),

    #Academic Pressure
    mean_acad = mean(`Academic Pressure`, na.rm = TRUE),
    median_acad = median(`Academic Pressure`, na.rm = TRUE),
    sd_acad = sd(`Academic Pressure`, na.rm = TRUE),
    min_acad = min(`Academic Pressure`, na.rm = TRUE),
    max_acad = max(`Academic Pressure`, na.rm = TRUE),

    #CGPA
    mean_cgpa = mean(CGPA, na.rm = TRUE),
    median_cgpa = median(CGPA, na.rm = TRUE),
    sd_cgpa = sd(CGPA, na.rm = TRUE),
    min_cgpa = min(CGPA, na.rm = TRUE),
    max_cgpa = max(CGPA, na.rm = TRUE),

    #Study Satisfaction
    mean_study = mean(`Study Satisfaction`, na.rm = TRUE),
    median_study = median(`Study Satisfaction`, na.rm = TRUE),
    sd_study = sd(`Study Satisfaction`, na.rm = TRUE),
    min_study = min(`Study Satisfaction`, na.rm = TRUE),
    max_study = max(`Study Satisfaction`, na.rm = TRUE),
```

```

#Work/Study Hours
mean_hours = mean(`Work/Study Hours`, na.rm = TRUE),
median_hours = median(`Work/Study Hours`, na.rm = TRUE),
sd_hours = sd(`Work/Study Hours`, na.rm = TRUE),
min_hours = min(`Work/Study Hours`, na.rm = TRUE),
max_hours = max(`Work/Study Hours`, na.rm = TRUE),

#Financial Stress
mean_stress = mean(`Financial Stress`, na.rm = TRUE),
median_stress = median(`Financial Stress`, na.rm = TRUE),
sd_stress = sd(`Financial Stress`, na.rm = TRUE),
min_stress = min(`Financial Stress`, na.rm = TRUE),
max_stress = max(`Financial Stress`, na.rm = TRUE)
)

by_history_stats

## # A tibble: 2 x 31
##   Family History of Menta~1 mean_age median_age sd_age min_age max_age mean_acad
##   <chr>                <dbl>      <dbl> <dbl> <dbl> <dbl>      <dbl>
## 1 No                    25.8         26  4.88    18    58        3.10
## 2 Yes                   25.8         25  4.94    18    59        3.18
## # i abbreviated name: 1: 'Family History of Mental Illness'
## # i 24 more variables: median_acad <dbl>, sd_acad <dbl>, min_acad <dbl>,
## #   max_acad <dbl>, mean_cgpa <dbl>, median_cgpa <dbl>, sd_cgpa <dbl>,
## #   min_cgpa <dbl>, max_cgpa <dbl>, mean_study <dbl>, median_study <dbl>,
## #   sd_study <dbl>, min_study <dbl>, max_study <dbl>, mean_hours <dbl>,
## #   median_hours <dbl>, sd_hours <dbl>, min_hours <dbl>, max_hours <dbl>,
## #   mean_stress <dbl>, median_stress <dbl>, sd_stress <dbl>, ...

print(by_history_stats)

```

```

## # A tibble: 2 x 31
##   Family History of Menta~1 mean_age median_age sd_age min_age max_age mean_acad
##   <chr>                <dbl>      <dbl> <dbl> <dbl> <dbl>      <dbl>
## 1 No                    25.8         26  4.88    18    58        3.10
## 2 Yes                   25.8         25  4.94    18    59        3.18
## # i abbreviated name: 1: 'Family History of Mental Illness'
## # i 24 more variables: median_acad <dbl>, sd_acad <dbl>, min_acad <dbl>,
## #   max_acad <dbl>, mean_cgpa <dbl>, median_cgpa <dbl>, sd_cgpa <dbl>,
## #   min_cgpa <dbl>, max_cgpa <dbl>, mean_study <dbl>, median_study <dbl>,
## #   sd_study <dbl>, min_study <dbl>, max_study <dbl>, mean_hours <dbl>,
## #   median_hours <dbl>, sd_hours <dbl>, min_hours <dbl>, max_hours <dbl>,
## #   mean_stress <dbl>, median_stress <dbl>, sd_stress <dbl>, ...

```