

STUDENT DEPRESSION ANALISYS ON R STUDIO

LIBRARIES AND DATA SET IMPORT

```
library(tidyverse)
```

```
## — Attaching core tidyverse packages — tidyverse 2.0.0 —
## ✓ dplyr      1.1.4      ✓ readr      2.1.5
## ✓ forcats    1.0.0      ✓ stringr   1.5.1
## ✓ ggplot2    3.5.1      ✓ tibble    3.2.1
## ✓ lubridate  1.9.4      ✓ tidyr     1.3.1
## ✓ purrr      1.0.2
## — Conflicts — tidyverse_conflicts() —
## ✖ dplyr::filter() masks stats::filter()
## ✖ dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(readxl)
library(janitor)
```

```
##
## Adjuntando el paquete: 'janitor'
##
## The following objects are masked from 'package:stats':
##
##      chisq.test, fisher.test
```

```
library(ggplot2)
library(skimr)
```

```
Student_Depression_Dataset <- read_csv("C:/Users/Daniela/Desktop/Student Depression Proyect/Student Depression Dataset.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.
```

EXPLORING AND UNDERSTANDING THE DATA CONTAINED IN THE DATA SET

```
#To have a preview of the data set
head(Student_Depression_Dataset)
```

	id <div><dbl></div>
	2
	8
	26
	30
	32
	33

6 rows | 1-1 of 18 columns

```
#To see how many columns the data set has and there names.
colnames(Student_Depression_Dataset)
```

```
## [1] "id"
## [2] "Gender"
## [3] "Age"
## [4] "City"
## [5] "Profession"
## [6] "Academic Pressure"
## [7] "Work Pressure"
## [8] "CGPA"
## [9] "Study Satisfaction"
## [10] "Job Satisfaction"
## [11] "Sleep Duration"
## [12] "Dietary Habits"
## [13] "Degree"
## [14] "Have you ever had suicidal thoughts ?"
## [15] "Work/Study Hours"
## [16] "Financial Stress"
## [17] "Family History of Mental Illness"
## [18] "Depression"
```

```
#This provides us with more in dept information about data stored on each column, as well as vectors types and va
lues amounts.
str(Student_Depression_Dataset)
```

```
## 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>
```

DATA CLEANING

```
#Let's find out if there are any missing values, and also, considering the information that should be contained in each column, we'll check if the unique values displayed are consistent with the possible values for each column
skim_without_charts(Student_Depression_Dataset)
```

Data summary

Name	Student_Depression_Dataset
Number of rows	27901
Number of columns	18
Column type frequency:	
character	8
numeric	10
Group variables	
None	

Variable type: character

skim_variable	n_missing	complete_rate	min	max	empty	n_unique	whitespace
Gender	0	1	4	6	0	2	0
City	0	1	2	18	0	52	0
Profession	0	1	4	22	0	14	0
Sleep Duration	0	1	6	17	0	5	0
Dietary Habits	0	1	6	9	0	4	0
Degree	0	1	2	8	0	28	0
Have you ever had suicidal thoughts ?	0	1	2	3	0	2	0
Family History of Mental Illness	0	1	2	3	0	2	0

Variable type: numeric

skim_variable	n_missing	complete_rate	mean	sd	p0	p25	p50	p75	p100
id	0	1	70442.15	40641.18	2	35039.00	70684.00	105818.00	140699
Age	0	1	25.82	4.91	18	21.00	25.00	30.00	59
Academic Pressure	0	1	3.14	1.38	0	2.00	3.00	4.00	5
Work Pressure	0	1	0.00	0.04	0	0.00	0.00	0.00	5
CGPA	0	1	7.66	1.47	0	6.29	7.77	8.92	10
Study Satisfaction	0	1	2.94	1.36	0	2.00	3.00	4.00	5
Job Satisfaction	0	1	0.00	0.04	0	0.00	0.00	0.00	4
Work/Study Hours	0	1	7.16	3.71	0	4.00	8.00	10.00	12
Financial Stress	3	1	3.14	1.44	1	2.00	3.00	4.00	5
Depression	0	1	0.59	0.49	0	0.00	1.00	1.00	1

```
#In the financial stress column there are 3 missing values, as we can't have access to the information to fill in the blanks, we're going to drop the rows with missing values.
```

```
Student_Depression_Clean <- drop_na(Student_Depression_Dataset)
```

```
#This data is pretty clean already, but there's some work we can still apply on it to make it even more tidy and manageable.
```

```
#First, let's get rid of the "id" column, which provides no relevant information for our analysis.
```

```
Student_Depression_Clean <- Student_Depression_Dataset %>%
  select(-"id")
```

#Now lets make sure there are only numbers, letters and underscores on the data set
clean_names(Student_Depression_Clean)

gender										
<chr>										
Male										
Female										
Male										
Female										
Female										
Male										
Male										
Female										
Male										
Male										
1-10 of 10,000 rows 1-1 of 17 columns										
Previous 1 2 Next										

#Next, let's have a peak into the unique values contained in each one of the columns, that way we can make sure there are not values misspelled or out of context.
valores_unicos <- lapply(Student_Depression_Clean, unique)

valores_unicos

```
## $Gender
## [1] "Male" "Female"
##
## $Age
## [1] 33 24 31 28 25 29 30 27 19 20 23 18 21 22 34 32 26 39 35 42 36 58 49 38 51
## [26] 44 43 46 59 54 48 56 37 41
##
## $City
## [1] "Visakhapatnam" "Bangalore" "Srinagar"
## [4] "Varanasi" "Jaipur" "Pune"
## [7] "Thane" "Chennai" "Nagpur"
## [10] "Nashik" "Vadodara" "Kalyan"
## [13] "Rajkot" "Ahmedabad" "Kolkata"
## [16] "Mumbai" "Lucknow" "Indore"
## [19] "Surat" "Ludhiana" "Bhopal"
## [22] "Meerut" "Agra" "Ghaziabad"
## [25] "Hyderabad" "Vasai-Virar" "Kanpur"
## [28] "Patna" "Faridabad" "Delhi"
## [31] "Saanvi" "M.Tech" "Bhavna"
## [34] "Less Delhi" "City" "3.0"
## [37] "Less than 5 Kalyan" "Mira" "Harsha"
## [40] "Vaanya" "Gaurav" "Harsh"
## [43] "Reyansh" "Kibara" "Rashi"
## [46] "ME" "M.Com" "Nalyan"
## [49] "Mihir" "Nalini" "Nandini"
## [52] "Khaziabad"
##
## $Profession
## [1] "Student" "Civil Engineer" "Architect"
## [4] "UX/UI Designer" "Digital Marketer" "Content Writer"
## [7] "Educational Consultant" "Teacher" "Manager"
## [10] "Chef" "Doctor" "Lawyer"
## [13] "Entrepreneur" "Pharmacist"
##
## $`Academic Pressure`
## [1] 5 2 3 4 1 0
##
## $`Work Pressure`
## [1] 0 5 2
##
## $CGPA
## [1] 8.9700 5.9000 7.0300 5.5900 8.1300 5.7000 9.5400 8.0400 9.7900
## [10] 8.3800 6.1000 7.0400 8.5200 5.6400 8.5800 6.5100 7.2500 7.8300
## [19] 9.9300 8.7400 6.7300 5.5700 8.5900 7.1000 6.0800 5.7400 9.8600
## [28] 6.7000 6.2100 5.8700 6.3700 9.7200 5.8800 9.5600 6.9900 5.2400
```

```

## [37] 9.2100 7.8500 6.9500 5.8600 7.9200 9.6600 8.9400 9.7100 7.8700
## [46] 5.6000 7.9000 5.4600 6.7900 8.7000 7.3800 8.5000 7.0900 9.8200
## [55] 8.8900 7.9400 9.1100 6.7500 7.5300 9.4900 9.0100 7.6400 5.2700
## [64] 6.0000 9.4400 5.7500 7.5100 9.0500 6.3800 8.9500 9.8800 5.3200
## [73] 6.2700 7.7000 8.1000 9.5900 8.9600 5.5100 7.4300 8.7900 9.9500
## [82] 5.3700 6.8600 8.3200 9.7400 5.6600 7.4800 8.2300 8.8100 6.0300
## [91] 5.5600 5.6800 5.1400 7.6100 6.1700 8.1700 9.8700 8.7500 6.1600
## [100] 9.5000 7.9900 5.6700 8.9200 6.1900 5.7600 6.2500 5.1100 5.5800
## [109] 5.6500 9.8900 8.0300 6.6100 9.4100 8.6400 7.2100 8.2800 6.0400
## [118] 9.1300 8.0800 9.9600 5.1200 8.3500 7.0700 9.6000 9.2400 8.5400
## [127] 8.7800 8.9300 8.9100 9.0400 6.8300 5.8500 7.7400 6.4100 8.9000
## [136] 7.7500 7.8800 5.4200 7.5200 7.6800 8.4000 9.3900 6.8400 5.9900
## [145] 8.6200 8.5300 7.4700 6.7800 6.4200 9.9200 8.3900 5.8900 7.2200
## [154] 6.8100 9.0200 9.9700 9.6300 9.6700 5.4100 7.2700 6.0500 6.8500
## [163] 9.3300 5.8100 6.5300 5.9800 6.0200 6.7400 5.2600 7.7200 7.3900
## [172] 8.4300 9.3400 5.4400 5.8200 5.7200 8.1900 8.4400 8.9800 9.3700
## [181] 5.8000 7.2800 7.6000 7.9100 9.1700 7.4600 9.4300 9.9100 9.3600
## [190] 5.1600 7.0800 9.2600 8.8300 10.0000 7.8000 9.4600 6.6300 7.2400
## [199] 6.4700 7.7700 5.0600 7.1700 8.2400 6.8800 9.0300 5.0800 5.4500
## [208] 8.4600 9.1900 6.3600 8.7300 7.1100 9.1200 9.4000 8.1100 9.9800
## [217] 5.5500 8.6100 8.1400 6.8900 9.8400 5.4800 8.2100 7.8200 8.5500
## [226] 5.7900 8.7700 8.2900 6.9200 7.3700 9.7000 6.2600 7.2600 7.5000
## [235] 6.8200 7.1500 5.7700 5.9100 5.1000 7.7100 9.0600 5.7100 5.8400
## [244] 9.4200 6.2300 6.2900 5.2500 9.6900 9.9000 6.3900 8.0900 5.8300
## [253] 5.4700 6.5600 8.7100 9.9400 6.6900 5.5200 7.3000 7.0200 6.3300
## [262] 8.0700 8.3700 8.0000 7.7900 8.6500 6.2800 7.3500 8.6900 7.1200
## [271] 7.3200 7.1300 5.9700 5.0900 6.9100 6.7600 6.5200 7.4500 8.5600
## [280] 6.5000 8.6300 8.2700 8.4900 6.5900 9.2900 5.3000 7.0600 5.3800
## [289] 6.6500 9.1600 8.0100 8.2500 8.0200 8.4700 7.3400 8.8800 7.1400
## [298] 8.4200 5.1700 9.1000 7.4900 9.8500 7.4200 9.3100 6.3500 7.0000
## [307] 5.3900 5.6100 9.7800 9.2500 5.6900 9.4700 8.1600 7.2300 6.4600
## [316] 0.0000 8.2600 6.3200 6.7700 8.8500 5.0300 7.6500 5.7800 6.2400
## [325] 5.3500 6.0600 7.7800 6.6400 7.0625 6.9800 6.4400 6.0900
##
## $`Study Satisfaction`
## [1] 2 5 3 4 1 0
##
## $`Job Satisfaction`
## [1] 0 3 4 2 1
##
## $`Sleep Duration`
## [1] "5-6 hours" "Less than 5 hours" "7-8 hours"
## [4] "More than 8 hours" "Others"
##
## $`Dietary Habits`
## [1] "Healthy" "Moderate" "Unhealthy" "Others"
##
## $Degree
## [1] "B.Pharm" "BSc" "BA" "BCA" "M.Tech" "PhD"
## [7] "Class 12" "B.Ed" "LLB" "BE" "M.Ed" "MSc"
## [13] "BHM" "M.Pharm" "MCA" "MA" "B.Com" "MD"
## [19] "MBA" "MBBS" "M.Com" "B.Arch" "LLM" "B.Tech"
## [25] "BBA" "ME" "MHM" "Others"
##
## $`Have you ever had suicidal thoughts ?`
## [1] "Yes" "No"
##
## $`Work/Study Hours`
## [1] 3 9 4 1 0 12 2 11 10 6 8 5 7
##
## $`Financial Stress`
## [1] 1 2 5 3 4 NA
##
## $`Family History of Mental Illness`
## [1] "No" "Yes"
##
## $Depression
## [1] 1 0

```

#After checking the unique values on City attribute, we can spot an odd value "3.0", witch is completely out of context considering there should only be city names. Let's get rid of the row containing that value. And also other rows with names that not correlate to city names

```
Student_Depression_Clean <- Student_Depression_Clean[Student_Depression_Clean$City != "3.0",]  
  
Student_Depression_Clean <- Student_Depression_Clean[Student_Depression_Clean$City != "Less than 5 Kalyan",]  
  
Student_Depression_Clean <- Student_Depression_Clean[Student_Depression_Clean$City != "Less Delhi",]  
  
Student_Depression_Clean <- Student_Depression_Clean[Student_Depression_Clean$City != "ME",]  
  
Student_Depression_Clean <- Student_Depression_Clean[Student_Depression_Clean$City != "City",]  
  
Student_Depression_Clean <- Student_Depression_Clean[Student_Depression_Clean$City != "M.Tech",]  
  
Student_Depression_Clean <- Student_Depression_Clean[Student_Depression_Clean$City != "M.Com",]
```

EDA (Exploratory Data Analysis)

GENDER ANALYSIS

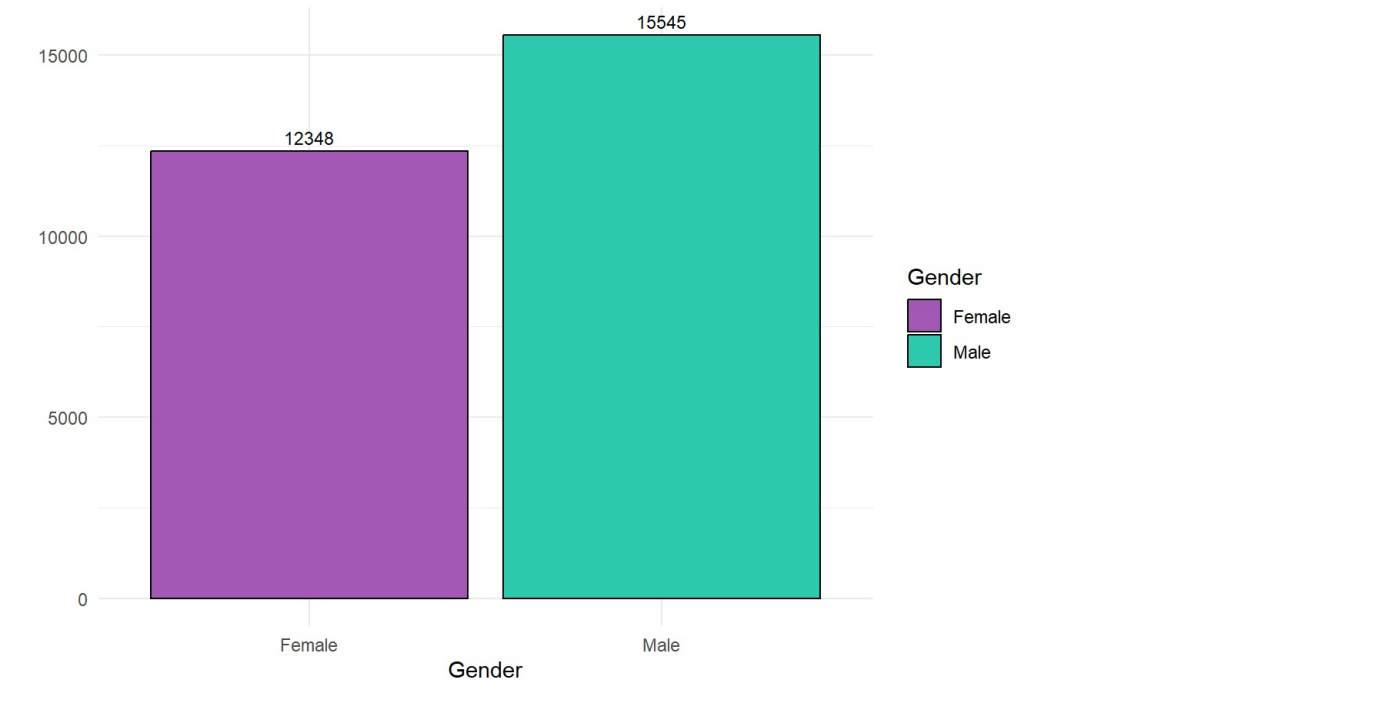
#Creating a table containing only gender data, and counting the amount of values per category.

```
gender_table <- table(Student_Depression_Clean$Gender)  
gender_table_df <- as.data.frame(gender_table)  
colnames(gender_table_df) <- c(x="GENDER", y="COUNT")  
print(gender_table_df)
```

```
##  GENDER COUNT  
## 1 Female 12348  
## 2  Male 15545
```

```
#GENDER BAR GRAPH  
ggplot(data = Student_Depression_Clean)+  
  geom_bar(mapping = aes(x=Gender, fill = Gender), color ="Black")+  
  geom_text(  
    stat = "count",  
    aes(x = Gender, label = after_stat(count)),  
    vjust = -0.5,  
    size = 3  
  ) +  
  labs(  
    title = "GENDER DISTRIBUTION FEMALE VS MALE",  
    x= "Gender",  
    y= ""  
  )+  
  theme_minimal()+  
  scale_fill_manual(values = c("#A358B6", "#2DCAB0"))+  
  theme(  
    plot.title = element_text(  
      family = "serif",  
      face = "bold",  
      size = 16,  
      hjust = 0.5,  
      vjust = 1  
    )  
  )  
)
```

GENDER DISTRIBUTION FEMALE VS MALE



AGE ANALYSIS

```
#Making a table to display age distribution.

Age_table <- table(Student_Depression_Clean$Age)

Age_table_df <- as.data.frame(Age_table)

colnames(Age_table_df) <- c(x="AGE", y="COUNT")

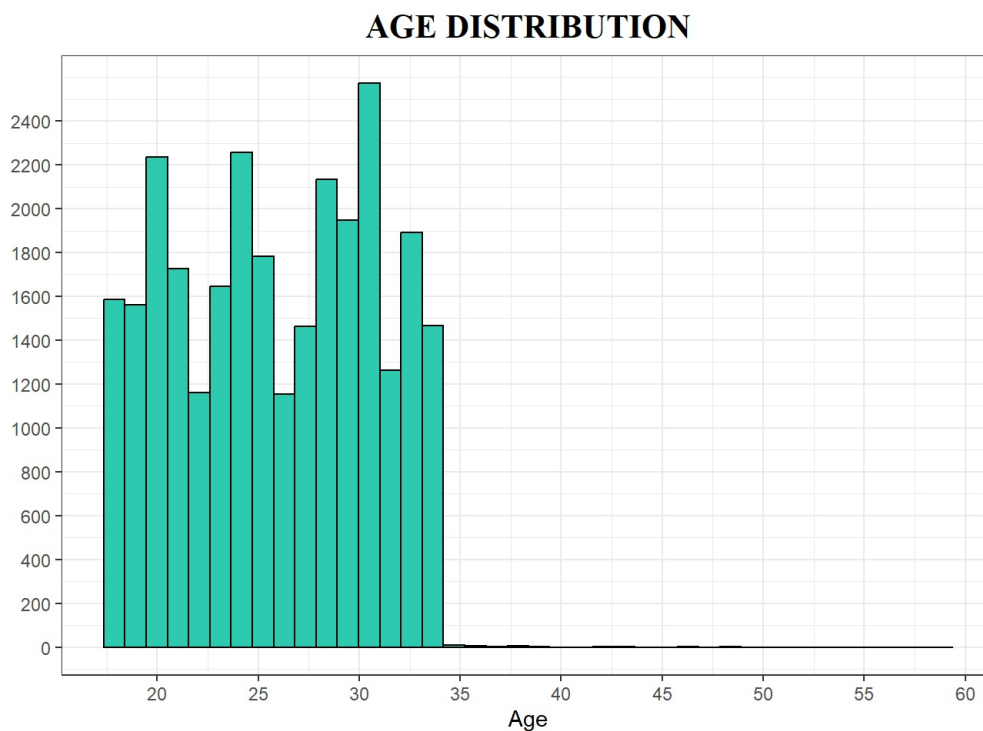
Age_table_df
```

AGE
<fct>
18
19
20
21
22
23
24
25
26
27

1-10 of 34 rows | 1-1 of 2 columns

Previous12Next

```
#AGE HISTOGRAM
ggplot(data = Student_Depression_Clean)+
  geom_histogram(
    aes(x=Age),
    bins = 40,
    fill = "#2DCAB0",
    color = "black"
  )+
  labs(
    title = "AGE DISTRIBUTION",
    X= "Age",
    y= ""
  )+
  scale_x_continuous(
    breaks = seq(from= 15, to = 60, by = 5)
  ) +
  scale_y_continuous(
    breaks = seq(from= 0, to = 2400, by = 200)
  ) +
  theme_bw()+
  scale_fill_manual(values = c("", "#2DCAB0"))+
  theme(
    plot.title = element_text(
      family = "serif",
      face = "bold",
      size = 16,
      hjust = 0.5,
      vjust = 1
    )
  )
)
```



Most of the students are between the ages of 18 and 34, and a really small portion of them are between ages 35 to 59, a normal pattern for university scenarios.

CITY ANALYSIS

```
city_count <- table(Student_Depression_Clean$City)
city_count_df <- as.data.frame(city_count)
colnames(city_count_df) <- c(x="CITY", y= "COUNT")
city_count_df <- city_count_df %>%
  arrange(desc(COUNT))
city_count_df
```

CITY

<fct>



Kalyan
Srinagar
Hyderabad
Vasai-Virar
Lucknow
Thane
Ludhiana
Agra
Surat
Kolkata

1-10 of 45 rows | 1-1 of 2 columns

Previous12Next

PROFESION ANALYSIS

```
profession_table <- table(Student_Depression_Clean$Profession)
profession_table_df <- as.data.frame(profession_table)
colnames(profession_table_df ) <- c(x="PROFESSION", y="COUNT")
profession_table_df <- profession_table_df %>%
  arrange(desc(COUNT))
profession_table_df
```

PROFESSION
<fct>
Student
Architect
Teacher
Digital Marketer
Chef
Content Writer
Doctor
Pharmacist
Civil Engineer
Educational Consultant

1-10 of 14 rows | 1-1 of 2 columns

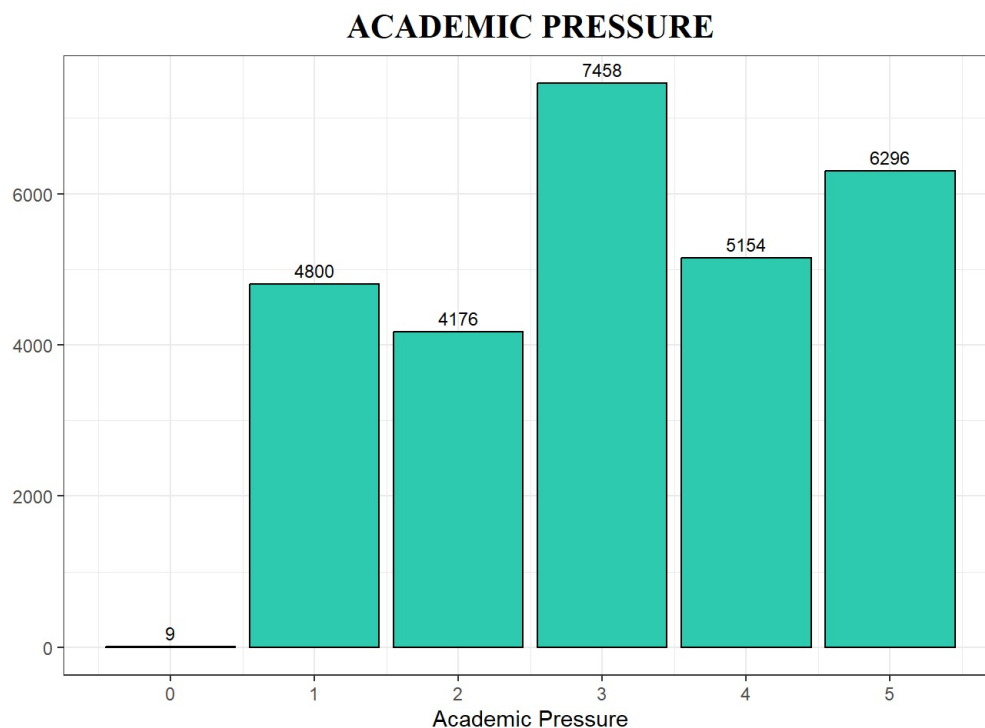
Previous12Next

ACADEMIC PRESSURE ANALYSIS

```
academic_pressure_table <- table(Student_Depression_Clean`Academic Pressure`)
academic_pressure_table_df <- as.data.frame(academic_pressure_table)
colnames(academic_pressure_table_df) <- c(x="Academic Pressure", y="Count")
academic_pressure_table_df
```

Academic Pressure
<fct>
0
1
2
3
4
5

```
#ACADEMIC PRESSURE BAR GRAPH
ggplot(data = Student_Depression_Clean)+
  geom_bar(
    aes(x=`Academic Pressure`),
    fill = "#2DCAB0",
    color = "black"
  )+
  geom_text(
    stat = "count",
    aes(x = `Academic Pressure`, label = after_stat(count)),
    vjust = -0.5,
    size = 3
  )+
  scale_x_continuous(
    breaks = seq(from=0, to=5, by=1)
  )+
  labs(
    title = "ACADEMIC PRESSURE",
    y=""
  )+
  theme_bw()+
  theme(
    plot.title = element_text(
      family = "serif",
      face = "bold",
      size = 16,
      hjust = 0.5,
      vjust = 1
    )
  )
)
```



WORK PRESSURE ANALYSIS

```
#WORK PREASSURE TABLE
work_pressure_table <- table(Student_Depression_Clean$`Work Pressure`)

work_pressure_table_df <- as.data.frame(work_pressure_table)

colnames(work_pressure_table_df) <- c(x= "Work Pressure", y= "Count")

work_pressure_table_df
```

Work Pressure

<fct>



0	
2	
5	
3 rows 1-1 of 2 columns	

All of the students (except for 3), gave the same answer: "0". This gives us no insight about work related issues take part in depression among the students.This is why we'll take this columns from the database.

```
Student_Depression_Clean <- Student_Depression_Clean %>%
  select(-"Work Pressure")
```

Same thing happens with column named "Job Satisfaction" the amount of answers outside 0 is minimal, witch does not give us a representative sample for further analysis.

JOB SATISFACTION ANALYSIS

```
job_satisfaction_table <- table(Student_Depression_Clean$`Job Satisfaction`)
job_satisfaction_table_df <- as.data.frame(job_satisfaction_table)
colnames(job_satisfaction_table_df) <- c(x="Job Satis", y="Count")
job_satisfaction_table_df
```

Job Satis	
<fct>	
0	
1	
2	
3	
4	
5 rows 1-1 of 2 columns	

We'll drop Job Satisfaction column from our data base as well.

```
Student_Depression_Clean <- Student_Depression_Clean %>%
  select(-"Job Satisfaction")
```

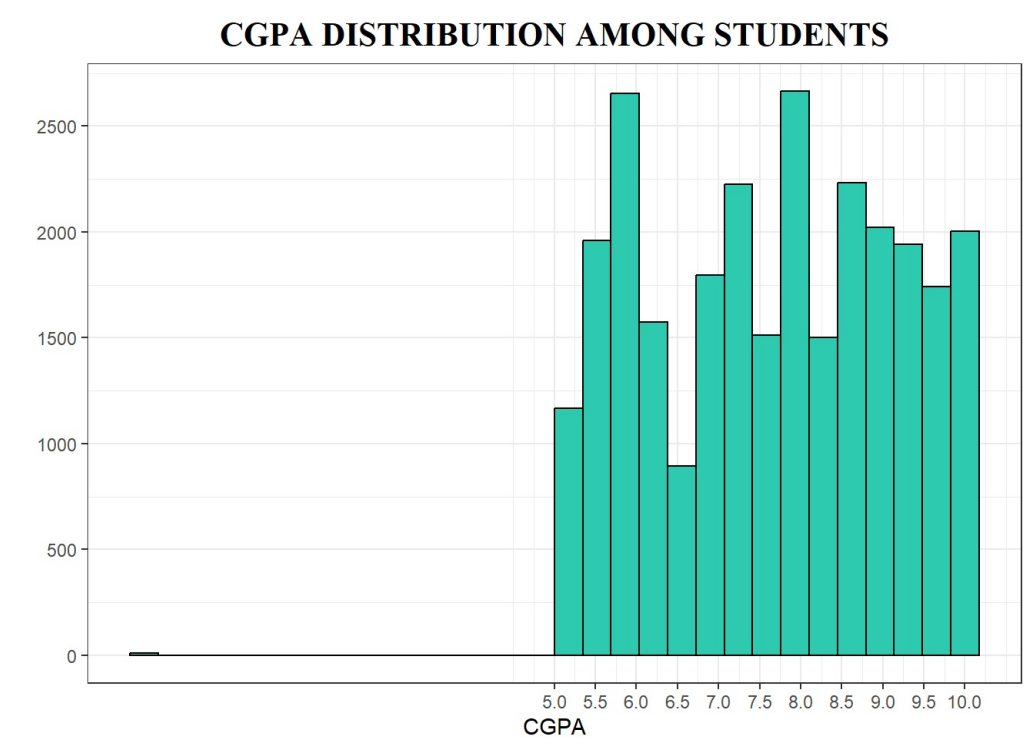
CGPA SCORE ANALYSIS

```
CGPA_table <- table(Student_Depression_Clean$CGPA)
CGPA_table_df <- as.data.frame(CGPA_table)
colnames(CGPA_table_df) <- c(x="CGPA SCORE", y="Count")
CGPA_table_df <- CGPA_table_df %>%           #drop not valid values
  filter(`CGPA SCORE` != 0)
CGPA_table_df
```

CGPA SCORE	
<fct>	
5.03	
5.06	

5.08	
5.09	
5.1	
5.11	
5.12	
5.14	
5.16	
5.17	
1-10 of 331 rows 1-1 of 2 columns	
Previous 1 2 Next	

```
#CGPA HISTOGRAM
ggplot(data = Student_Depression_Clean) +
  geom_histogram(
    aes(x = CGPA),
    bins = 30,
    fill = "#2DCAB0",
    color = "black"
  )+
  scale_x_continuous(
    breaks= seq(from=5, to=(max(Student_Depression_Clean$CGPA)), by=0.5))+
  scale_y_continuous(
    breaks = seq(from=0, to=3000, by=500)
  )+
  labs(
    title = "CGPA DISTRIBUTION AMONG STUDENTS",
    y=""
  )+
  theme_bw()+
  theme(
    plot.title = element_text(
      family = "serif",
      face = "bold",
      size = 16,
      hjust = 0.5,
      vjust = 1))
```



```
study_satisfaction_table <- table(Student_Depression_Clean$`Study Satisfaction`)

study_satisfaction_table_df <- as.data.frame(study_satisfaction_table)

colnames(study_satisfaction_table_df) <- c(x="STUDY SATISFACTION", y="COUNT")

study_satisfaction_table_df
```

STUDY SATISFACTION

<fct>

0

1

2

3

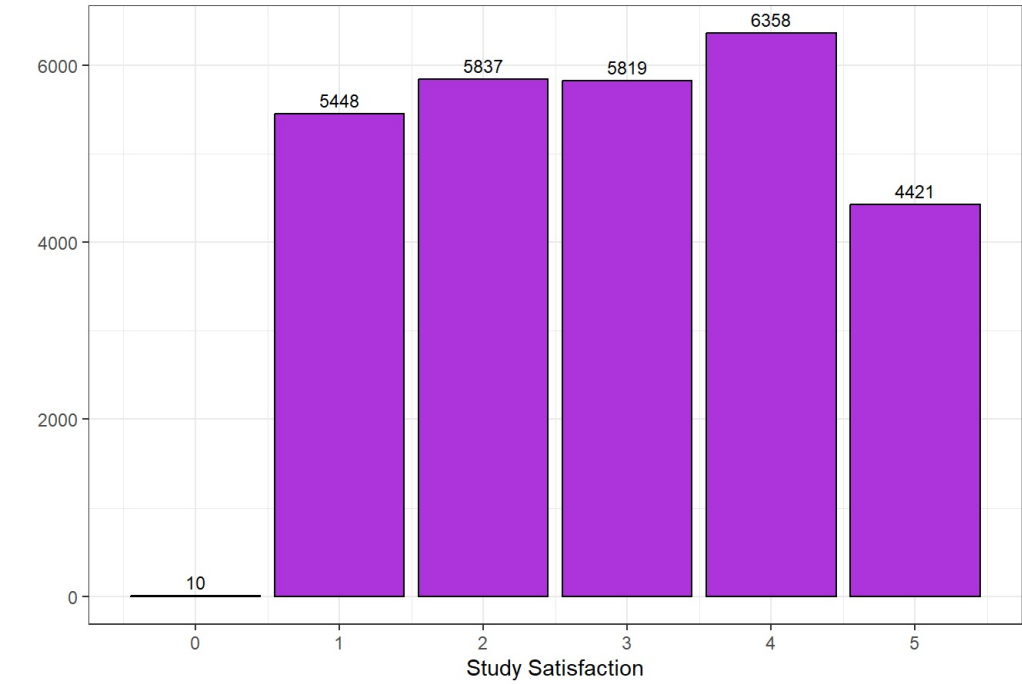
4

5

6 rows | 1-1 of 2 columns

```
#STUDY SATISFACTION BAR GRAPH
ggplot(data = Student_Depression_Clean)+
  geom_bar(
    aes(x=`Study Satisfaction`),
    fill = "#AD34DA",
    color = "black"
  )+
  geom_text(
    stat = "count",
    aes(x = `Study Satisfaction`, label = after_stat(count)),
    vjust = -0.5,
    size = 3
  )+
  scale_x_continuous(
    breaks = seq(from=0, to=5, by=1)
  )+
  labs(
    title = "STUDY SATISFACTION",
    y=""
  )+
  theme_bw()+
  theme(
    plot.title = element_text(
      family = "serif",
      face = "bold",
      size = 16,
      hjust = 0.5,
      vjust = 1
    )
  )
)
```

STUDY SATISFACTION

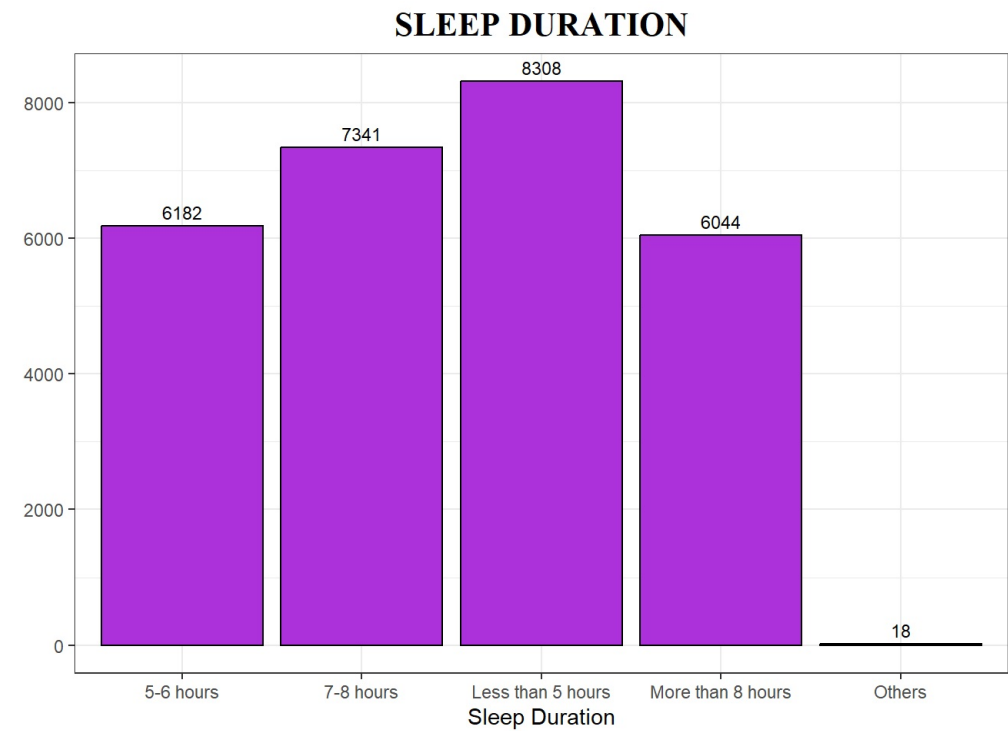


SLEEP DURATION ANALYSIS

```
sleep_duration_table <- table(Student_Depression_Clean$`Sleep Duration`)\nsleep_duration_table_df <- as.data.frame(sleep_duration_table)\ncolnames(sleep_duration_table_df) <- c(x="SLEEP DURATION", y="COUNT")\nsleep_duration_table_df
```

SLEEP DURATION
<fct>
5-6 hours
7-8 hours
Less than 5 hours
More than 8 hours
Others
5 rows 1-1 of 2 columns

```
#SLEEP_DURATION BAR GRAPH
ggplot(data = Student_Depression_Clean)+
  geom_bar(
    aes(x=`Sleep Duration`),
    fill = "#AD31DA",
    color = "black"
  )+
  geom_text(
    stat = "count",
    aes(x = `Sleep Duration`, label = after_stat(count)),
    vjust = -0.5,
    size = 3
  )+
  labs(
    title = "SLEEP DURATION",
    y=""
  )+
  theme_bw()+
  theme(
    plot.title = element_text(
      family = "serif",
      face = "bold",
      size = 16,
      hjust = 0.5,
      vjust = 1
    )
  )
)
```



DIETARY HABITS ANALYSIS

```
dietary_habits_table <- table(Student_Depression_Clean$`Dietary Habits`)

dietary_habits_table_df <- as.data.frame(dietary_habits_table)

colnames(dietary_habits_table_df) <- c(x="DIETARY HABITS", y="COUNT")

dietary_habits_table_df
```

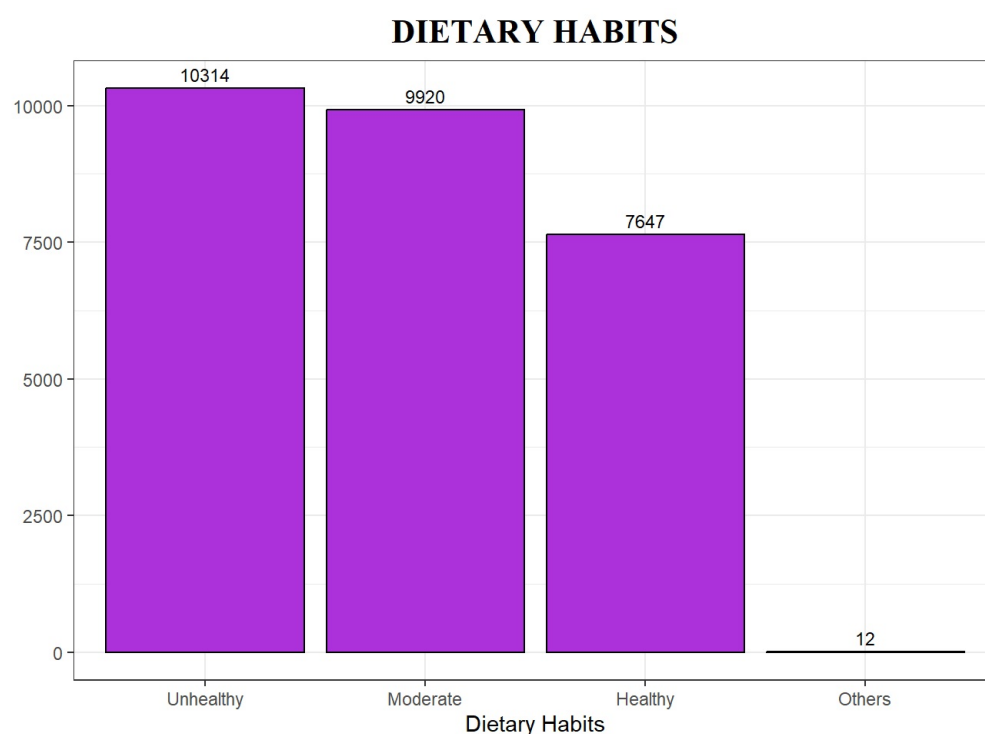
DIETARY HABITS	
<fct>	
Healthy	
Moderate	
Others	

Unhealthy

4 rows | 1-1 of 2 columns

```
#Reorganizing the categories to be displayed how I prefer
Student_Depression_Clean$`Dietary Habits` <- factor(
  Student_Depression_Clean$`Dietary Habits`,
  levels = c("Unhealthy", "Moderate", "Healthy", "Others")
)
```

```
#DIETARY HABITS BAR GRAPH
ggplot(data = Student_Depression_Clean)+
  geom_bar(
    aes(x=`Dietary Habits`),
    fill = "#AD31DA",
    color = "black"
  )+
  geom_text(
    stat = "count",
    aes(x = `Dietary Habits`, label = after_stat(count)),
    vjust = -0.5,
    size = 3
  )+
  labs(
    title = "DIETARY HABITS",
    y=""
  )+
  theme_bw()+
  theme(
    plot.title = element_text(
      family = "serif",
      face = "bold",
      size = 16,
      hjust = 0.5,
      vjust = 1
    )
  )
)
```



STUDENT DEGREE ANALYSIS


```
degree_table <- table(Student_Depression_Clean$Degree)

degree_table_df <- as.data.frame(degree_table)

colnames(degree_table_df) <- c(x="DEGREE", y="COUNT")

degree_table_df
```

DEGREE
<fct>
B.Arch
B.Com
B.Ed
B.Pharm
B.Tech
BA
BBA
BCA
BE
BHM

1-10 of 28 rows | 1-1 of 2 columns

Previous12Next

SUICIDAL THOUGHTS ANALYSIS

```
suicidal_thoughts_table <- table(Student_Depression_Clean$`Have you ever had suicidal thoughts ?`)

suicidal_thoughts_table_df <- as.data.frame(suicidal_thoughts_table)

colnames(suicidal_thoughts_table_df) <- c(x="Have you ever had suicidal thoughts ?", y="COUNT")

suicidal_thoughts_table_df
```

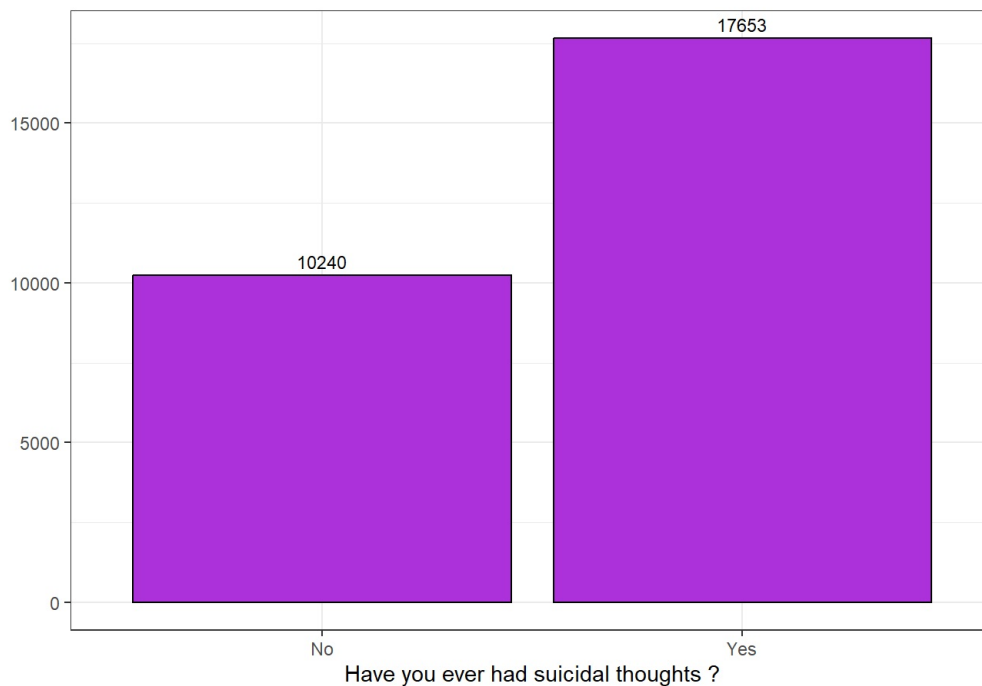
Have you ever had suicidal thoughts ?
<fct>
No
Yes

2 rows | 1-1 of 2 columns

```
#Have you ever had suicidal thoughts ? BAR CHART
```

```
ggplot(data = Student_Depression_Clean)+  
  geom_bar(mapping = aes(x=`Have you ever had suicidal thoughts ?`),  
    fill = "#AD31DA",  
    color = "black"  
  )+  
  geom_text(  
    stat = "count",  
    aes(x = `Have you ever had suicidal thoughts ?`, label = after_stat(count)),  
    vjust = -0.5,  
    size = 3  
  )+  
  labs(  
    title = "SUICIDAL THOUGHTS ON STUDENTS",  
    y=""  
  )+  
  theme_bw()+  
  theme(  
    plot.title = element_text(  
      family = "serif",  
      face = "bold",  
      size = 16,  
      hjust = 0.5,  
      vjust = 1  
    )  
  )  
)
```

SUICIDAL THOUGHTS ON STUDENTS



WORK/STUDY HOURS ANALYSIS

```
work_study_hours_table <- table(Student_Depression_Clean$`Work/Study Hours`)  
work_study_hours_table_df <- as.data.frame(work_study_hours_table)  
colnames(work_study_hours_table_df) <- c(x="Work/Study Hours", y="COUNT")  
work_study_hours_table_df
```

Work/Study Hours

<fct>

0

1

2

3

4

5

6

7

8

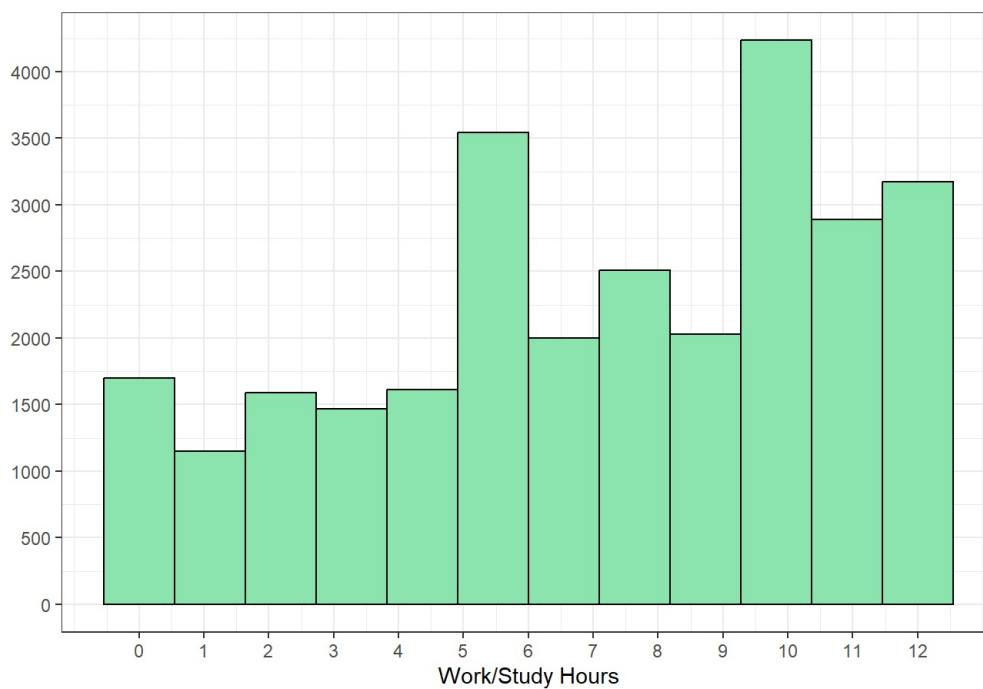
9

1-10 of 13 rows | 1-1 of 2 columns

Previous 1 2 Next

```
#WORK/STUDY HOURS BAR GRAPH
ggplot(data = Student_Depression_Clean) +
  geom_histogram(
    aes(x = `Work/Study Hours`),
    bins = 12,
    fill = "#8BE4AE",
    color = "black"
  )+
  scale_x_continuous(
    breaks= seq(from=0, to=12, by=1))+
  scale_y_continuous(
    breaks = seq(from=0, to=4500, by=500)
  )+
  labs(
    title = "STUDY/WORK HOURS SPENT BY STUDENTS",
    y=""
  )+
  theme_bw()+
  theme(
    plot.title = element_text(
      family = "serif",
      face = "bold",
      size = 16,
      hjust = 0.5,
      vjust = 1))
```

STUDY/WORK HOURS SPENT BY STUDENTS



```
financial_stress_table <- table(Student_Depression_Clean`Financial Stress`)

financial_stress_table_df <- as.data.frame(financial_stress_table)

colnames(financial_stress_table_df) <- c(x="Financial Stress", y="COUNT")

financial_stress_table_df
```

Financial Stress

<fct>

1

2

3

4

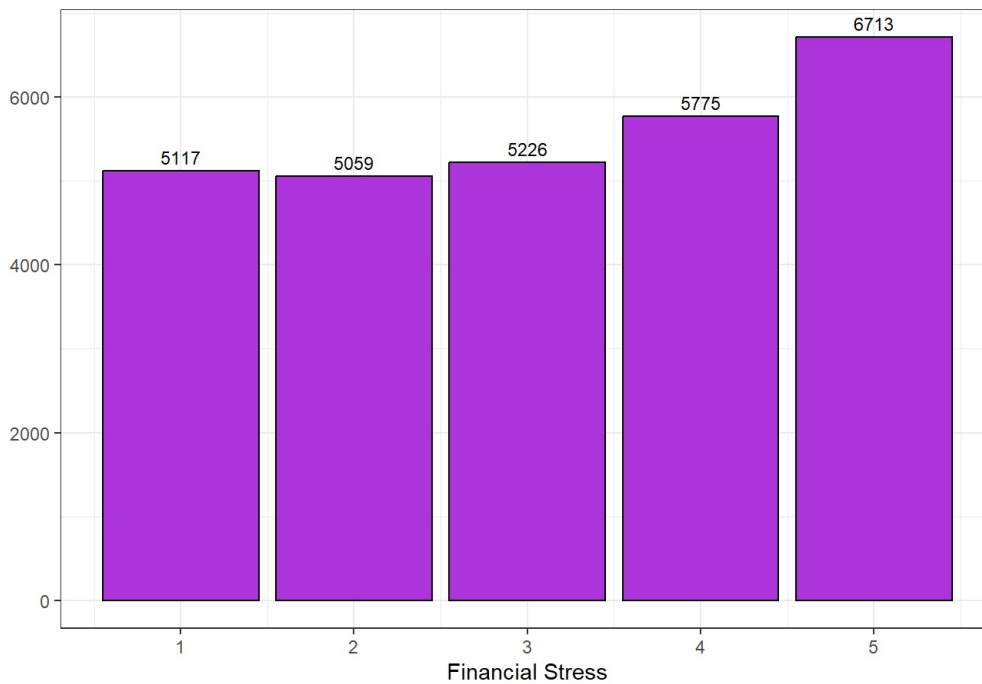
5

5 rows | 1-1 of 2 columns

```
#FINANCIAL STRESS BAR GRAPH
ggplot(data = Student_Depression_Clean)+
  geom_bar(
    aes(x=`Financial Stress`),
    fill = "#AD34DA",
    color = "black"
  )+
  geom_text(
    stat = "count",
    aes(x = `Financial Stress`, label = after_stat(count)),
    vjust = -0.5,
    size = 3
  )+
  scale_x_continuous(
    breaks = seq(from=1, to=5, by=1)
  )+
  labs(
    title = "FINANCIAL STRESS",
    y=""
  )+
  theme_bw()+
  theme(
    plot.title = element_text(
      family = "serif",
      face = "bold",
      size = 16,
      hjust = 0.5,
      vjust = 1
    )
  )
)
```

```
## Warning: Removed 3 rows containing non-finite outside the scale range (`stat_count()`).
## Removed 3 rows containing non-finite outside the scale range (`stat_count()`).
```

FINANCIAL STRESS



FAMILY HISTORY OF MENTAL ILLNES ANALYSIS

```
family_history_of_mental_illness_table <- table(Student_Depression_Clean$`Family History of Mental Illness`)
family_history_of_mental_illness_table_df <- as.data.frame(family_history_of_mental_illness_table)
colnames(family_history_of_mental_illness_table_df) <- c(x="Family History of Mental Illness", y="COUNT")
family_history_of_mental_illness_table_df
```

Family History of Mental Illness

<fct>

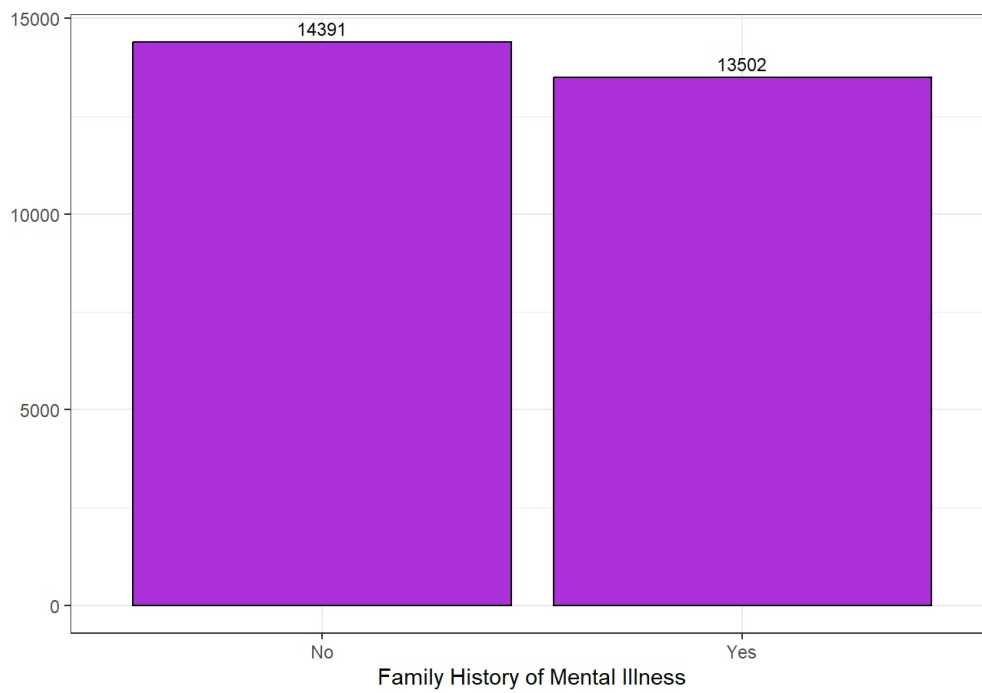
No

Yes

2 rows | 1-1 of 2 columns

```
#FAMILY HISTORY OF MENTAL ILLNES BAR GRAPH
ggplot(data = Student_Depression_Clean)+
  geom_bar(mapping = aes(x=`Family History of Mental Illness`,
    fill = "#AD31DA",
    color = "black"
  )+
  geom_text(
    stat = "count",
    aes(x = `Family History of Mental Illness`, label = after_stat(count)),
    vjust = -0.5,
    size = 3
  )+
  labs(
    title = "FAMILY HISTORY OF MENTAL ILLNES ",
    y=""
  )+
  theme_bw()+
  theme(
    plot.title = element_text(
      family = "serif",
      face = "bold",
      size = 16,
      hjust = 0.5,
      vjust = 1
    )
  )
)
```

FAMILY HISTORY OF MENTAL ILLNES

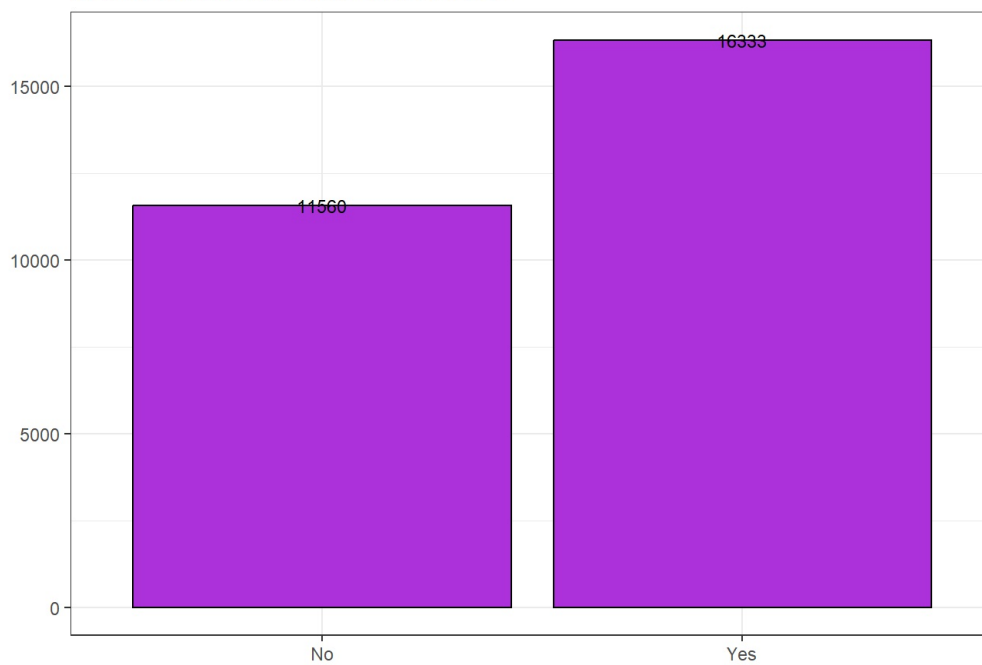


DEPRESSION ANALYSIS

```
#Transforming 0 and 1 values into "Yes" and "No"
Student_Depression_Clean$Depression <- factor(Student_Depression_Clean$Depression,
                                              levels = c(0, 1),
                                              labels = c("No", "Yes"))

#Student depression Bar chart
ggplot(data = Student_Depression_Clean)+
  geom_bar(aes(x=Depression), fill = "#AD31DA",color = "black")+
  geom_text(
    stat = "count",
    aes(x=Depression, label = after_stat(count)),
    vjust = 0.5,
    size = 3
  )+
  labs(
    title = "DEPRESSION AMONG STUDENTS",
    x = "",
    y=""
  )+
  theme_bw()+
  scale_x_discrete()
```

DEPRESSION AMONG STUDENTS

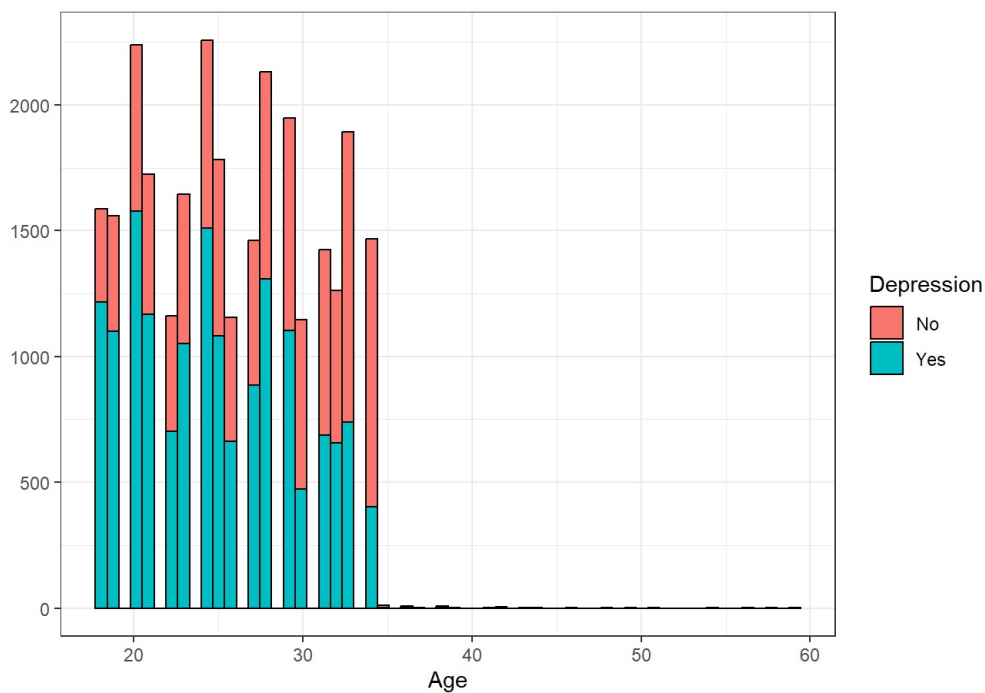


IN DEPTH ANALISYS

#AGE AGAINST DEPRESSION

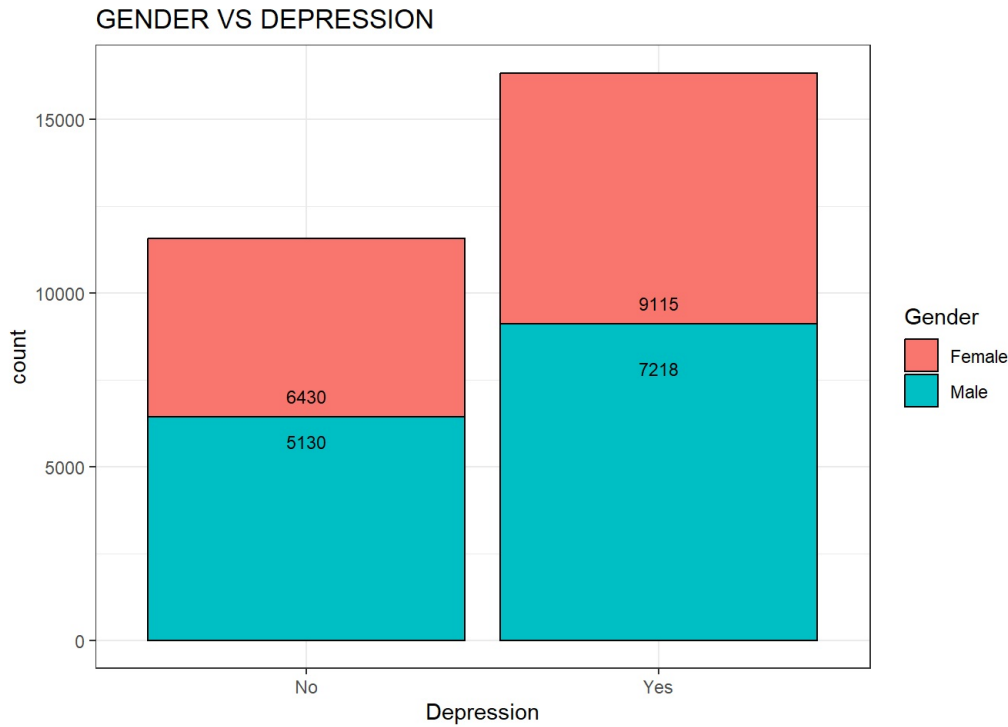
```
ggplot(data = Student_Depression_Clean)+  
  geom_histogram(aes(x=Age, fill = Depression),  
                 bins = 60,  
                 color = "black")+  
  labs(  
    title = "STUDENT DEPRESSION BY AGE RANGE",  
    y=""  
  )+  
  theme_bw()
```

STUDENT DEPRESSION BY AGE RANGE



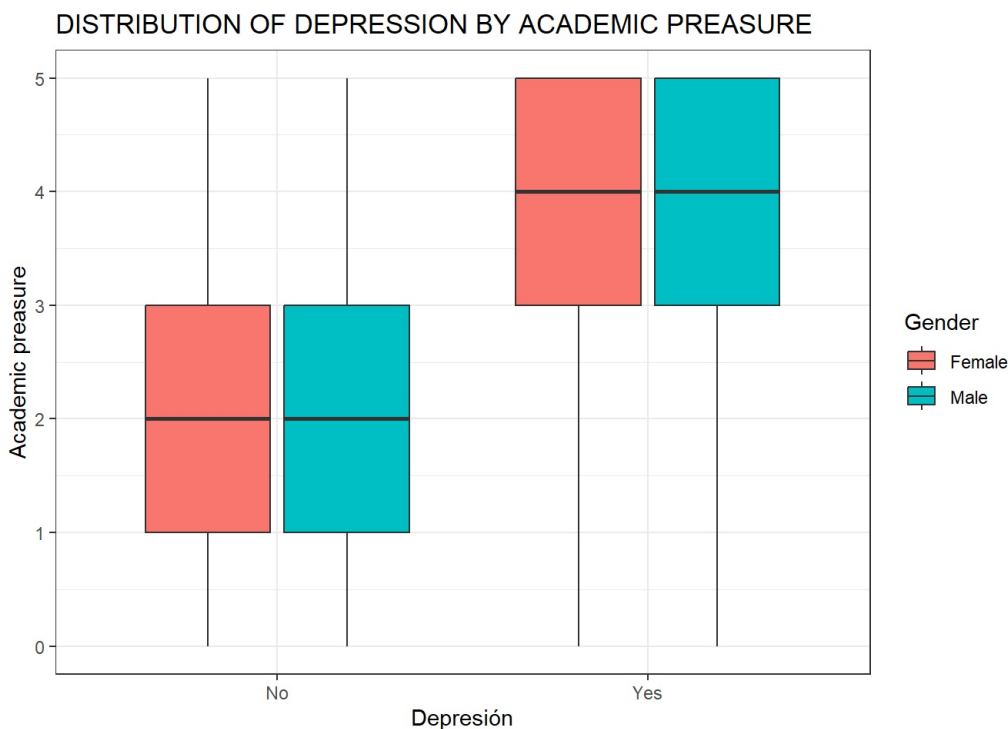
#DEPRESSION AGAINST GENDER

```
ggplot(data = Student_Depression_Clean)+  
  geom_bar(aes(x=Depression, fill = Gender),  
    color = "black")+  
  geom_text(stat = "count",  
    aes(x=Depression, label = after_stat(count), group = Gender),  
    vjust = -1,  
    size =3  
  )+  
  labs(  
    title = "GENDER VS DEPRESSION"  
  )+  
  theme_bw()
```



#DISTRIBUTION OF DEPRESSION BY ACADEMIC PREASURE

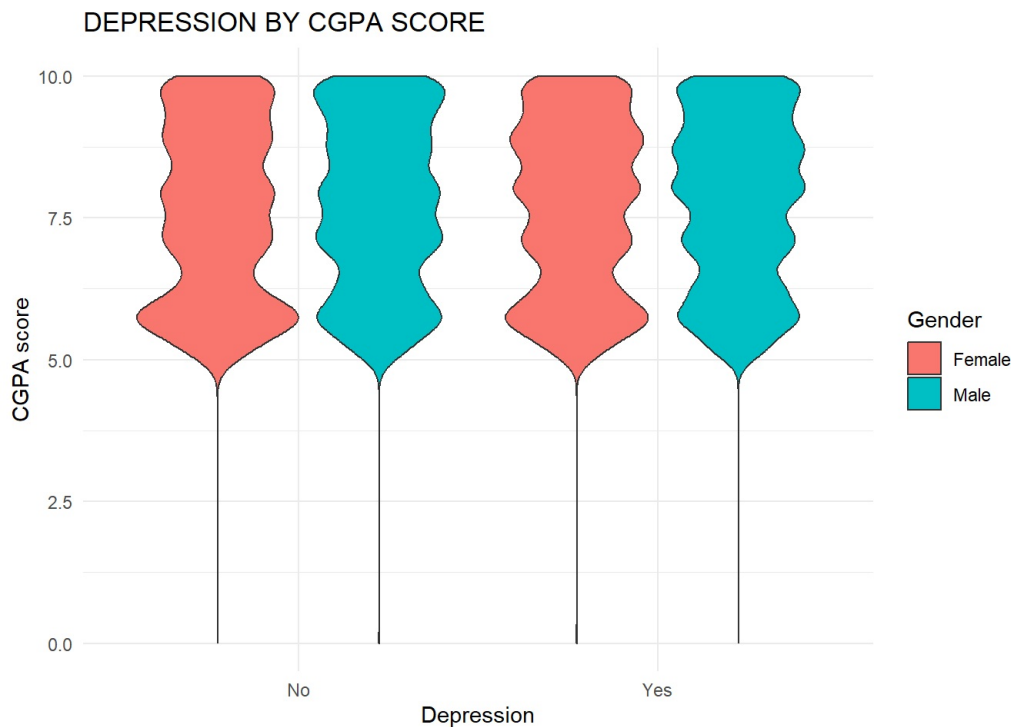
```
ggplot(data = Student_Depression_Clean) +  
  geom_boxplot(aes(x = Depression, y = `Academic Pressure`, fill = Gender)) +  
  labs(title = "DISTRIBUTION OF DEPRESSION BY ACADEMIC PREASURE", x = "Depresión", y = "Academic preasure") +  
  theme_bw()
```



Here we have an interesting insight, both female and male depression are spotted on academic pressures between 3 and 5. And likewise, most of the students who don't report on having depression, are the ones with the lowest scores for academic pressure.

#DEPRESSION BY CGPA SCORE

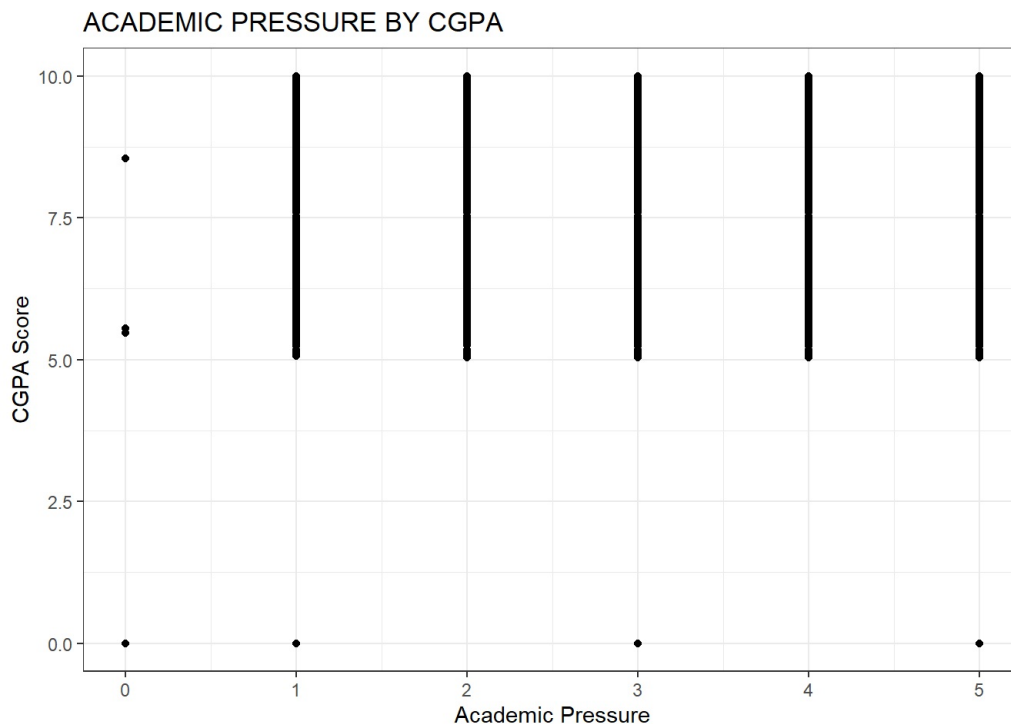
```
ggplot(data = Student_Depression_Clean) +  
  geom_violin(aes(x = Depression, y = CGPA, fill = Gender)) +  
  labs(title = "DEPRESSION BY CGPA SCORE", x = "Depression", y = "CGPA score") +  
  theme_minimal()
```



In this case, CGPA score, seems to have no big impact in the way depression behaves, also, the graph shows us really similar behavior between male and females.

#ACADEMIC PRESSURE BY CGPA

```
ggplot(data = Student_Depression_Clean) +  
  geom_point(aes(x = `Academic Pressure`, y = CGPA)) +  
  labs(title = "ACADEMIC PRESSURE BY CGPA", x = "Academic Pressure", y = "CGPA Score") +  
  theme_bw()
```

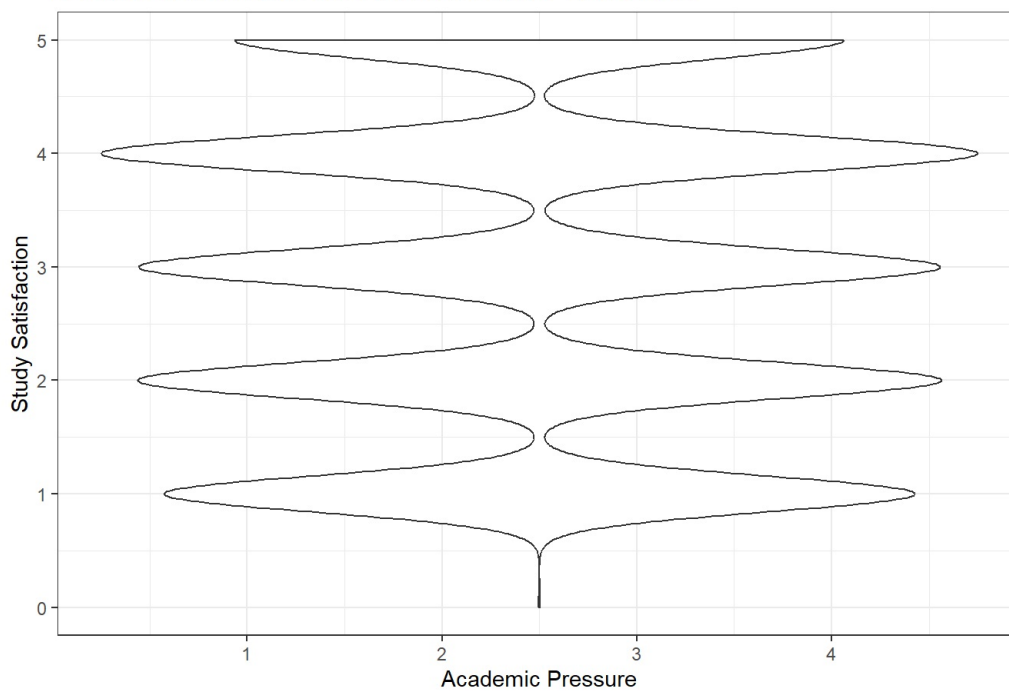


Academic pressure seems to have no correlation with CGPA scores. CGPA score is stable through different levels of academic pressure.

#STUDY SATISFACTION BY ACADEMIC PRESSURE

```
ggplot(data = Student_Depression_Clean) +  
  geom_violin(aes(x = `Academic Pressure`, y = `Study Satisfaction`)) +  
  labs(title = "STUDY SATISFACTION BY ACADEMIC PRESSURE", x = "Academic Pressure", y = "Study Satisfaction") +  
  theme_bw()
```

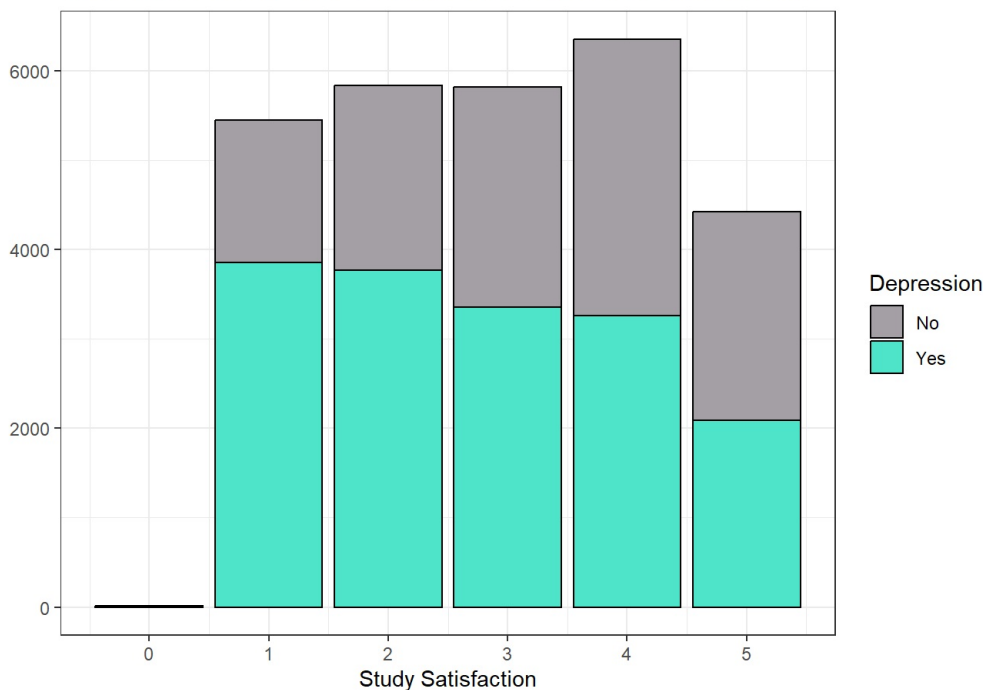
STUDY SATISFACTION BY ACADEMIC PRESSURE



In this case, the violin bar shows symmetric behavior between academic pressure and study satisfaction.

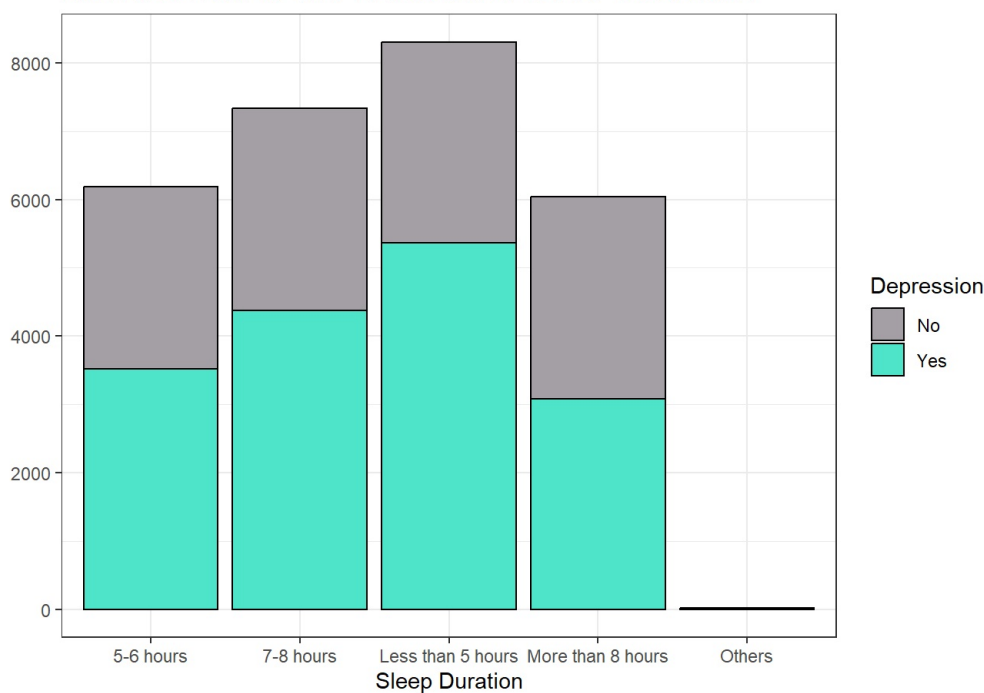
```
#DISTRIBUTION OF DEPRESSION BY STUDY SATISFACTION
ggplot(data = Student_Depression_Clean) +
  geom_bar(aes(x = `Study Satisfaction`, fill = Depression),
    color = "black") +
  scale_fill_manual(values = c("No" = "#A39FA4", "Yes" = "#4DE4CA")) +
  scale_x_continuous(breaks = seq(from = 0, to = 5, by = 1))+
  labs(title = "DISTRIBUTION OF DEPRESSION BY STUDY SATISFACTION", x = "Study Satisfaction", y = "") +
  theme_bw()
```

DISTRIBUTION OF DEPRESSION BY STUDY SATISFACTION



```
#DISTRIBUTION OF DEPRESSION BY SLEEP DURATION
ggplot(data = Student_Depression_Clean) +
  geom_bar(aes(x = `Sleep Duration`, fill = Depression),
    color = "black") +
  scale_fill_manual(values = c("No" = "#A39FA4", "Yes" = "#4DE4CA")) +
  labs(title = "DISTRIBUTION OF DEPRESSION BY SLEEP DURATION", x = "Sleep Duration", y = "") +
  theme_bw()
```

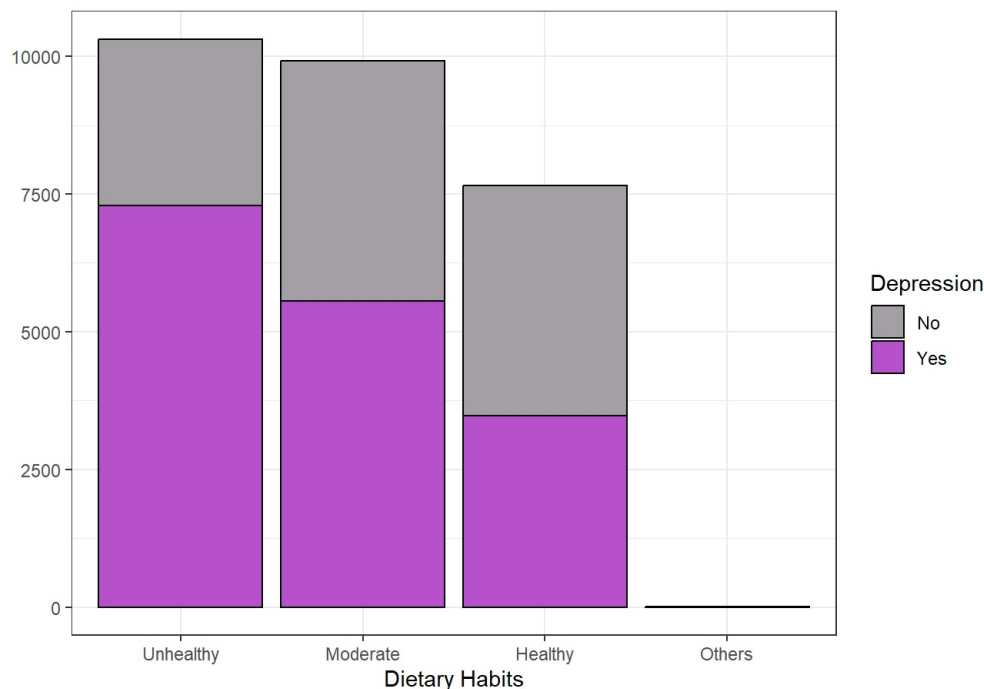
DISTRIBUTION OF DEPRESSION BY SLEEP DURATION



The biggest portion of students reporting depression are getting 6 hours or less of sleep. Also, students that are sleeping more than 8 hours report less depression.

```
#DISTRIBUTION OF DEPRESSION BY DIETARY HABITS
ggplot(data = Student_Depression_Clean) +
  geom_bar(aes(x = `Dietary Habits`, fill = Depression),
    color = "black") +
  scale_fill_manual(values = c("No" = "#A39FA4", "Yes" = "#B750CB")) +
  labs(title = "DISTRIBUTION OF DEPRESSION BY DIETARY HABITS", x = "Dietary Habits", y = "") +
  theme_bw()
```

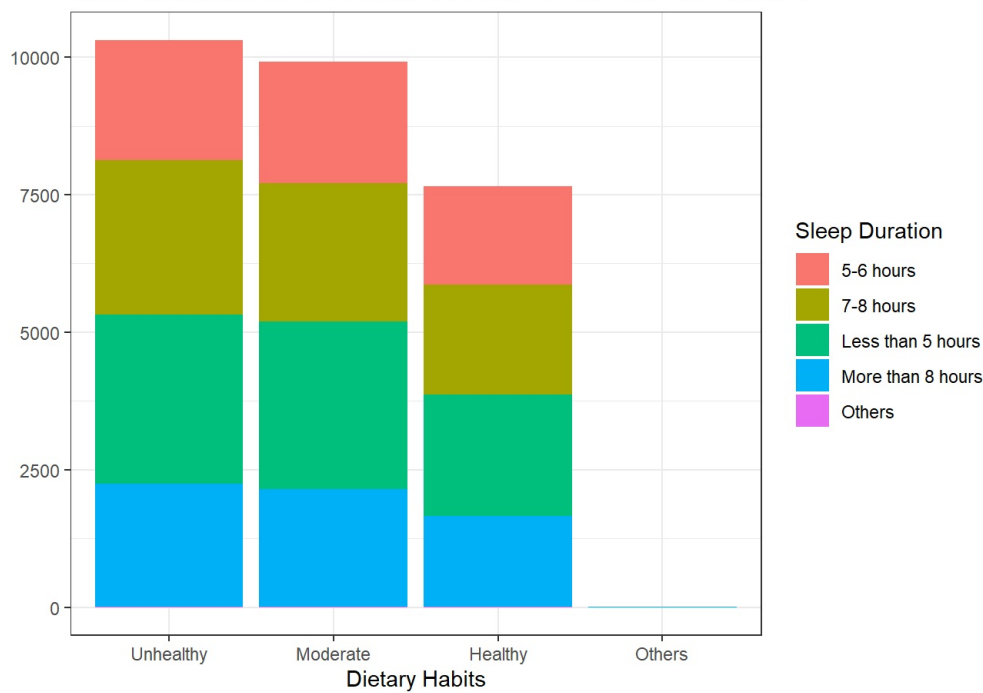
DISTRIBUTION OF DEPRESSION BY DIETARY HABITS



As expected, people with healthy dietary habits report the less amount of depression, while, students with moderate and specially unhealthy habits, are the ones with the most amount of cases of depression. This combined with sleeping habits, could give us some insight about habits and it correlation with depression.

```
#DISTRIBUTION OF DIETARY HABITS BY SLEEP DURATION
ggplot(data = Student_Depression_Clean) +
  geom_bar(aes(x = `Dietary Habits`, fill = `Sleep Duration`)) +
  labs(title = "DISTRIBUTION OF DIETARY HABITS BY SLEEP DURATION", x = "Dietary Habits", y = "") +
  theme_bw()
```

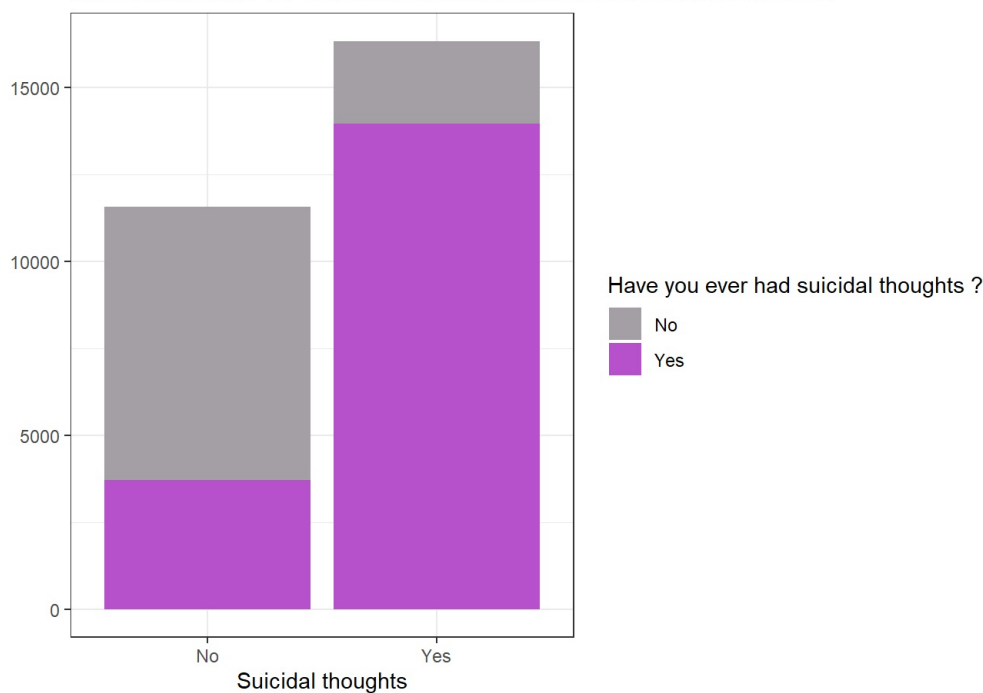
DISTRIBUTION OF DIETARY HABITS BY SLEEP DURATION



#DISTRIBUTION OF SUICIDAL THOUGHTS AND DEPRESSION

```
ggplot(data = Student_Depression_Clean) +
  geom_bar(aes(x = Depression, fill = `Have you ever had suicidal thoughts ?`)) +
  scale_fill_manual(values = c("No" = "#A39FA4", "Yes" = "#B750CB")) +
  labs(title = "DISTRIBUTION OF SUICIDAL THOUGHTS AND DEPRESSION", x = "Suicidal thoughts", y = "") +
  theme_bw()
```

DISTRIBUTION OF SUICIDAL THOUGHTS AND DEPRESSION

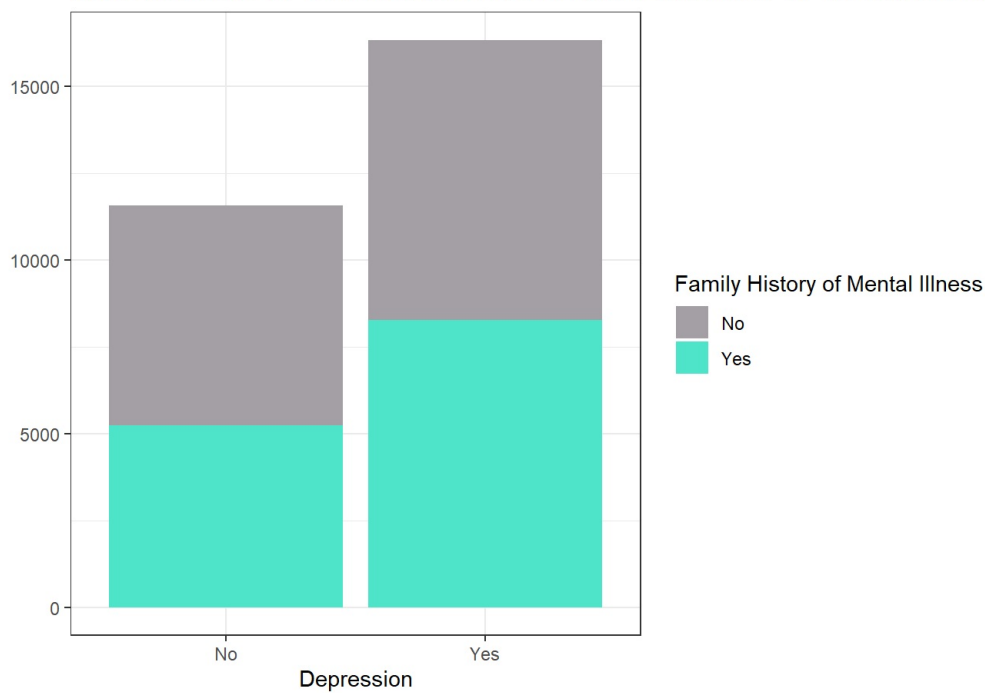


Most of the students who speak about having depression admits on having also suicidal thoughts.

#DISTRIBUTION OF DEPRESSION AND FAMILY HISTORY OF MENTAL ILLNESS

```
ggplot(data = Student_Depression_Clean) +
  geom_bar(aes(x = Depression, fill = `Family History of Mental Illness`)) +
  scale_fill_manual(values = c("No" = "#A39FA4", "Yes" = "#4DE4CA")) +
  labs(title = "DISTRIBUTION OF DEPRESSION AND FAMILY HISTORY OF MENTAL ILLNESS", x = "Depression", y = "") +
  theme_bw()
```

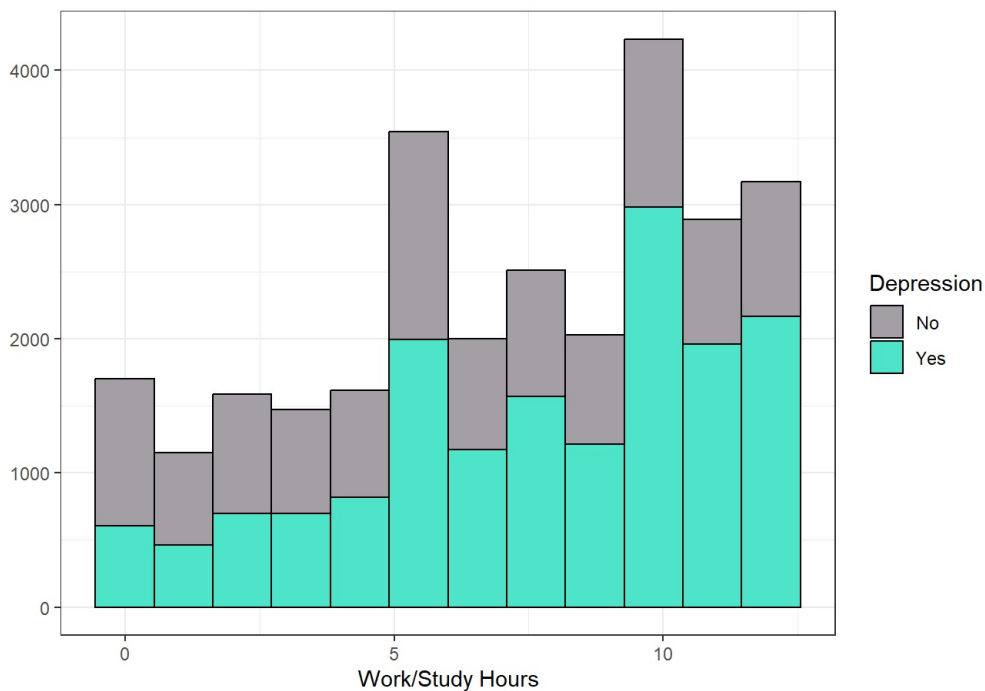
DISTRIBUTION OF DEPRESSION AND FAMILY HISTORY OF MENTAL ILLNESS



Both distributions between students with and without depression show similar proportions of family history of mental illness.

```
#WORK/STUDY HOURS AND DEPRESSION
ggplot(data = Student_Depression_Clean)+
  geom_histogram(aes(x=`Work/Study Hours`, fill = Depression),
    bins = 12,
    color = "black")+
  scale_fill_manual(values = c("No" = "#A39FA4", "Yes" = "#4DE4CA")) +
  labs(
    title = "WORK/STUDY HOURS AND DEPRESSION",
    y=""
  )+
  theme_bw()
```

WORK/STUDY HOURS AND DEPRESSION



```
#DISTRIBUTION OF FINANCIAL STRESS AND FAMILY HISTORY OF MENTAL ILLNESS"
ggplot(data = Student_Depression_Clean) +
  geom_bar(aes(x = `Financial Stress`, fill = Depression)) +
  scale_fill_manual(values = c("No" = "#A39FA4", "Yes" = "#B750CB")) +
  labs(title = "DISTRIBUTION OF FINANCIAL STRESS AND FAMILY HISTORY OF MENTAL ILLNESS", x = "Depression", y = "")
+
  theme_bw()
```

Warning: Removed 3 rows containing non-finite outside the scale range
(`stat_count()`).

DISTRIBUTION OF FINANCIAL STRESS AND FAMILY HISTORY OF MENTAL

