

A STATISTICAL RESERACH ON THE EFFECT OF MENTAL HEALTH ON STUDENTS CGPA DATASET

Mrs. Owoade Lola 7/23/2022

About the Data

This Data set was collected by a survey conducted by Google forms from University student in order to examine their current academic situation and mental health.

Therefore the indicators include: - Gender (Categorical Variable) - Age (Non-categorical Variable) - Course of study (Categorical Variable) - Year of study (Categorical Variable) - CGPA (Categorical Variable) - Marital status (Categorical Variable) - Do you have Depression? (Categorical Variable) - Do you have Anxiety? (Categorical Variable) - Do you have Panic attack? (Categorical Variable) - Did you seek any specialist for a treatment? (Categorical Variable)

Dpendent Variable

- CGPA

Independent Variables

- Anxiety
- Depression
- Panic Attack

Research Questions:

- Does Anxiety, Depression and Panic Attack has effect on the students CGPA?
- Does Age and Marital status affect CGPA

Objectives:

- To describe and visualize the mental health of students with respect to their CGPA

- To determine the impact of Anxiety, Depression, and Panic attack on the students CGPA
- To investigate any relationship or association between gender, marital status and CGPA

Hypothesis

- H0: There is significance relationship between variables
- H1: Not H0

Importing the dataset

```
#reading the data
data <- read.csv(file.choose(), sep = ",", header = TRUE)
head(data)
```

```
##   gender Age      course year.of.Study      CGPA Marital.status
## 1 Female  18      Engineering      Year 1 3.00 - 3.49          No
## 2   Male  21 Islamic education      Year 2 3.00 - 3.49          No
## 3   Male  19              BIT      Year 1 3.00 - 3.49          No
## 4 Female  22              Laws      Year 3 3.00 - 3.49          Yes
## 5   Male  23      Mathematics      Year 4 3.00 - 3.49          No
## 6   Male  19      Engineering      Year 2 3.50 - 4.00          No
##   Do.you.have.Depression. Do.you.have.Anxiety. Do.you.have.Panic.attack.
## 1                      Yes                  No                      Yes
## 2                      No                  Yes                      No
## 3                      Yes                  Yes                      Yes
## 4                      Yes                  No                      No
## 5                      No                  No                      No
## 6                      No                  No                      Yes
##   Did.you.seek.any.specialist.for.a.treatment.
## 1                                           No
## 2                                           No
## 3                                           No
## 4                                           No
## 5                                           No
## 6                                           No
```

```
attach(data)
```

from the data set, only Age is seen to be integer and quantitative, where all other variables are qualitative and categorical in nature.

Data checking and Cleaning

```
summary(data)
```

```
##      gender              Age      course      year.of.Study
## Length:101      Min.    :18.00  Length:101      Length:101
## Class :character 1st Qu.:18.00  Class :character Class :character
## Mode  :character Median :19.00  Mode  :character Mode  :character
##                      Mean   :20.53
##                      3rd Qu.:23.00
##                      Max.   :24.00
##                      NA's   :1
##      CGPA      Marital.status      Do.you.have.Depression.
## Length:101      Length:101      Length:101
## Class :character Class :character Class :character
## Mode  :character Mode  :character Mode  :character
##
##
##
##
## Do.you.have.Anxiety. Do.you.have.Panic.attack.
## Length:101      Length:101
## Class :character Class :character
## Mode  :character Mode  :character
##
##
##
##
## Did.you.seek.any.specialist.for.a.treatment.
## Length:101
## Class :character
## Mode  :character
##
##
##
##
```

Exploratory Data Analysis (EDA)

Univariate Analysis and Graphical Representation

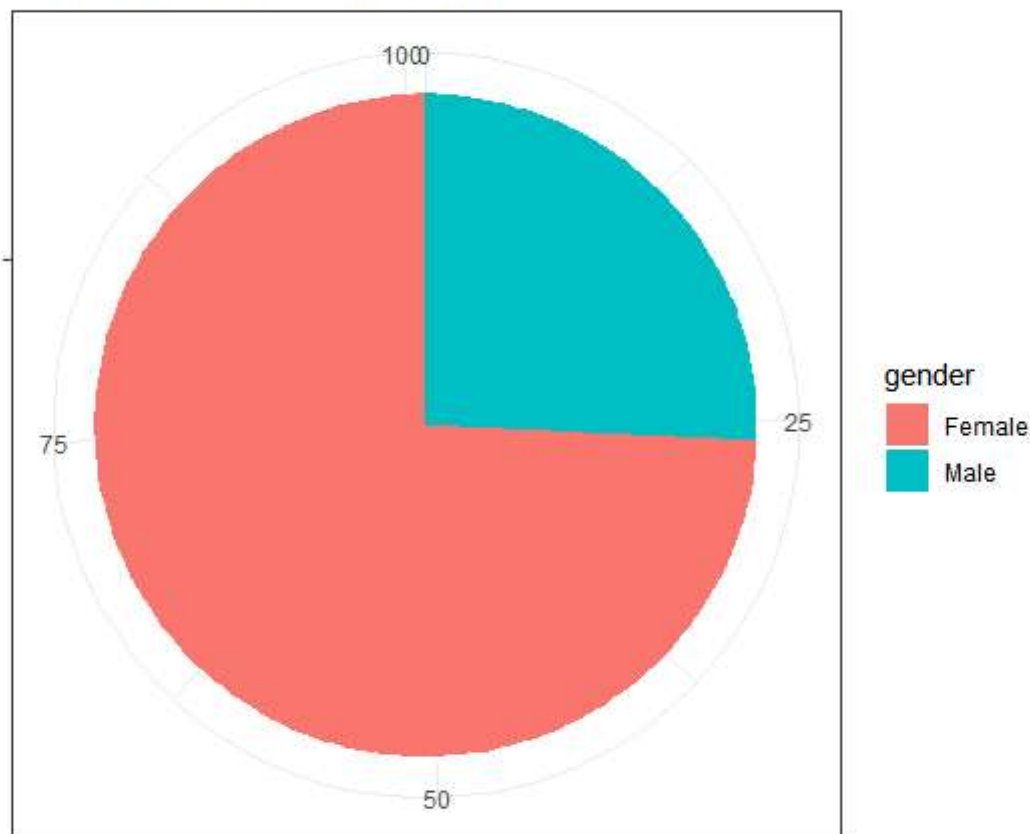
Importing required Packages

```
library(ggplot2)
library(dplyr)
```

Gender Distribution

```
pi = data.frame(table(gender))
ggplot(pi, aes(x = "", y = Freq, fill = gender))+
  geom_bar(stat = "identity", width = 1, )+
  coord_polar("y", start = 0)+
  theme_bw()+
  labs(title = "Pie Chart Gender Distribution", x = " ", y = " ")
```

Pie Chart Gender Distribution



```
#A <- data$gender
#AB = data.frame(table(A))
#colours <- c("red","blue")
#percentlabels <- round(100*AB$Freq/sum(AB$Freq), 1)
#pielabel <- paste(c("No", "Yes"), " ", percentlabels, "%", sep="")
#pie(AB$Freq, main="Pie Chart for Gender", col=colours, labels = pielabel, cex=0.8
# legend("topright", c("No", "Yes"),bty = "n",cex = 0.8, fill = colours)
```

Descriptive Statistics of Age

```
library(dplyr)
su = data %>%
  summarize(n = n(),
            Min = min(Age, na.rm = T),
            Mean = mean(Age, na.rm = T),
            Median = median(Age, na.rm = T),
            Max = max(Age, na.rm = T),
            Standard_Dev = sd(Age, na.rm = T),
            Standard_Error = Standard_Dev/sqrt(n))

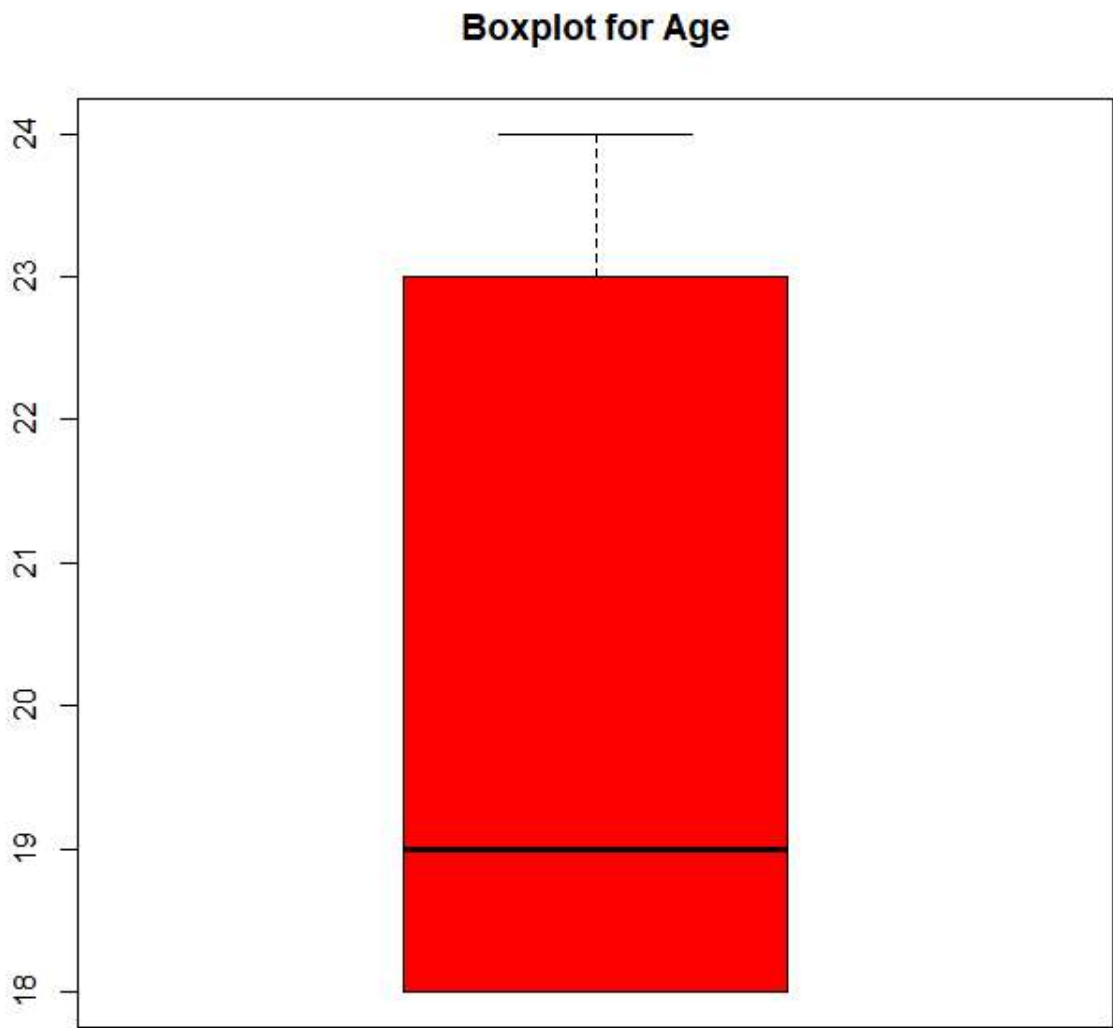
su
```

| ## | | n | Min | Mean | Median | Max | Standard_Dev | Standard_Error |
|----|---|-----|-----|-------|--------|-----|--------------|----------------|
| ## | 1 | 101 | 18 | 20.53 | 19 | 24 | 2.49628 | 0.2483891 |

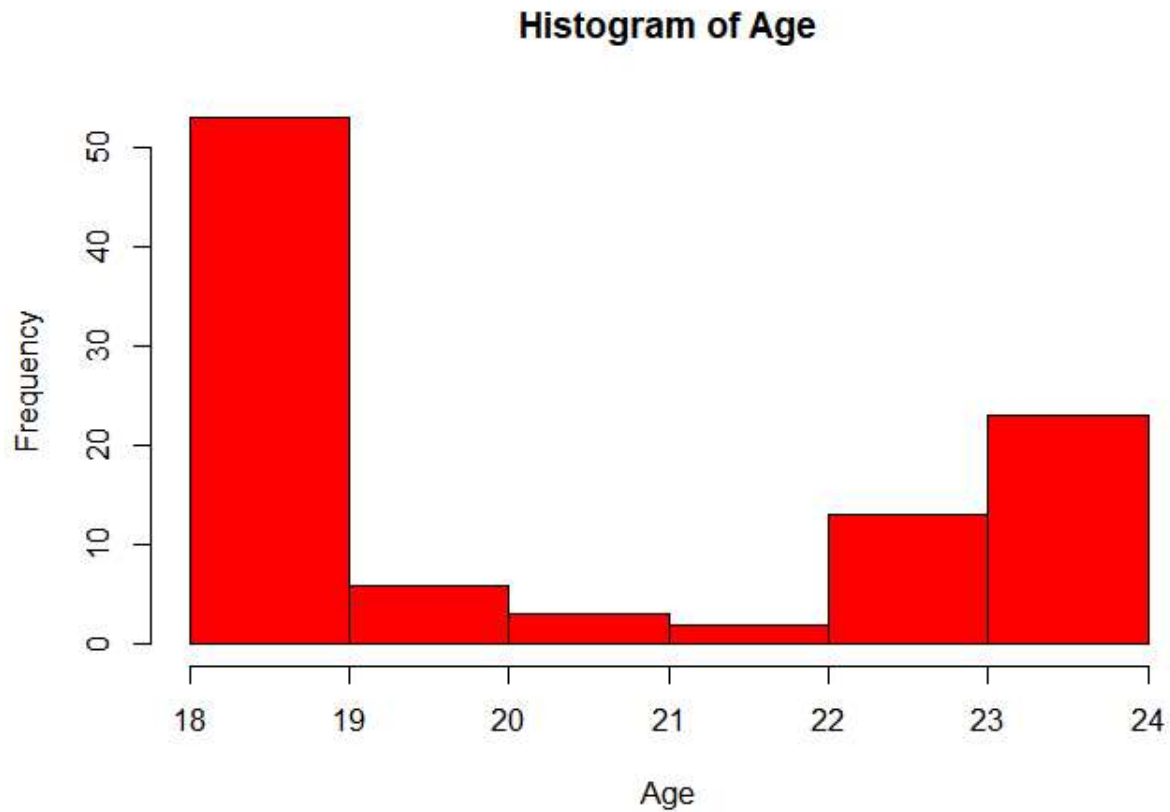
From the descriptive statistics of the students age in the table above, the average age of the students is 20, with the minimum and maximum age of 18 and 24. Also, the dispersion level of the studentd age is approximately 3, while the level at which the sample mean is likely to be, to the actual population mean is 0.25 respectively.

The Age is represented graphically with Box plot and Histogram Below ### Boxplot of Age

```
boxplot(Age, col = "red", main = "Boxplot for Age", na.rm = T)
```



```
hist(data$Age, col = "red", main = "Histogram of Age", xlab = "Age")
```

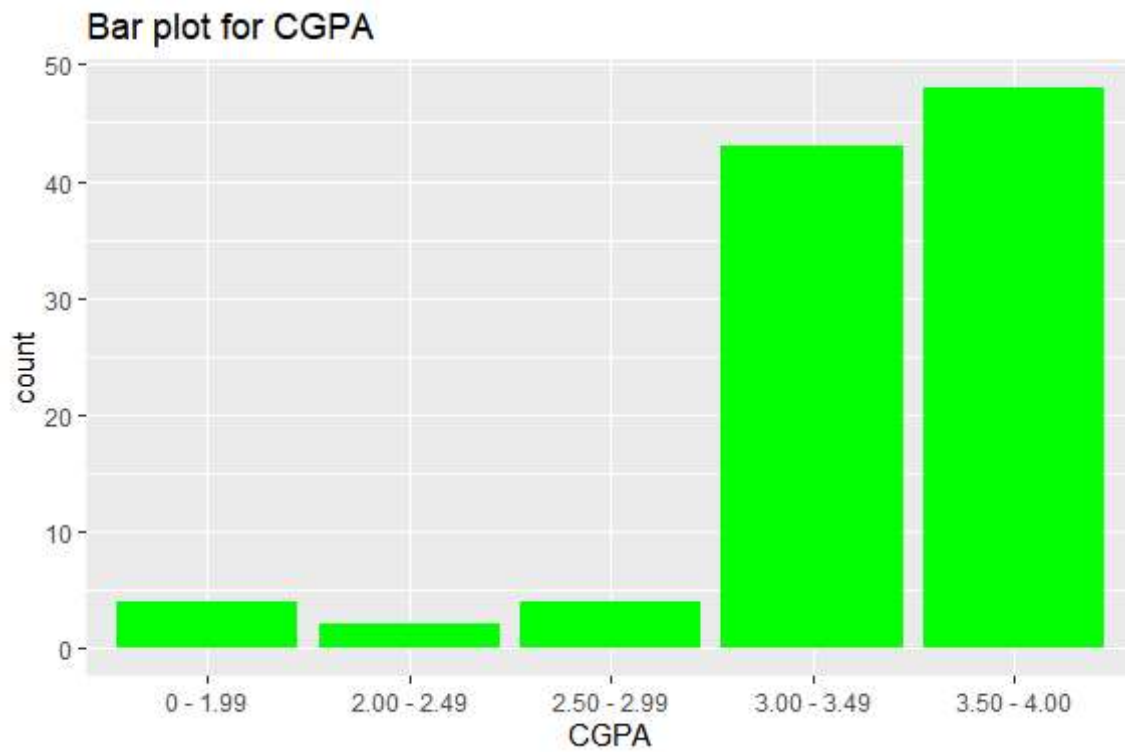


Frequencies and Percentage of students CGPA

```
d = data.frame(table(CGPA))
percentage = (d$Freq/sum(d$Freq) * 100)
Percentage = paste(percentage, "%")
CGP = data.frame(d,Percentage)
CGP
```

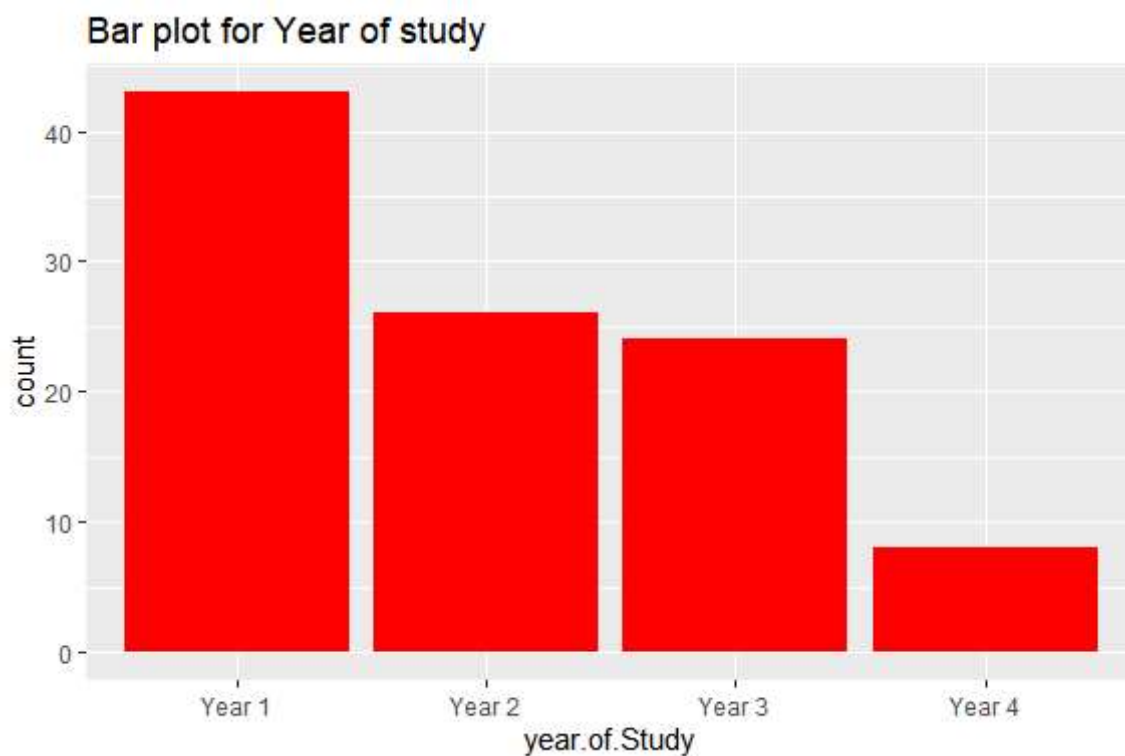
```
##          CGPA Freq      Percentage
## 1  0 - 1.99    4 3.96039603960396 %
## 2 2.00 - 2.49    2 1.98019801980198 %
## 3 2.50 - 2.99    4 3.96039603960396 %
## 4 3.00 - 3.49   43 42.5742574257426 %
## 5 3.50 - 4.00   48 47.5247524752475 %
```

```
library(ggplot2)
ggplot(data = data, aes(x=CGPA))+
  geom_bar(na.rm = TRUE, fill = "green")+
  theme_grey()+
  labs(title = "Bar plot for CGPA")
```



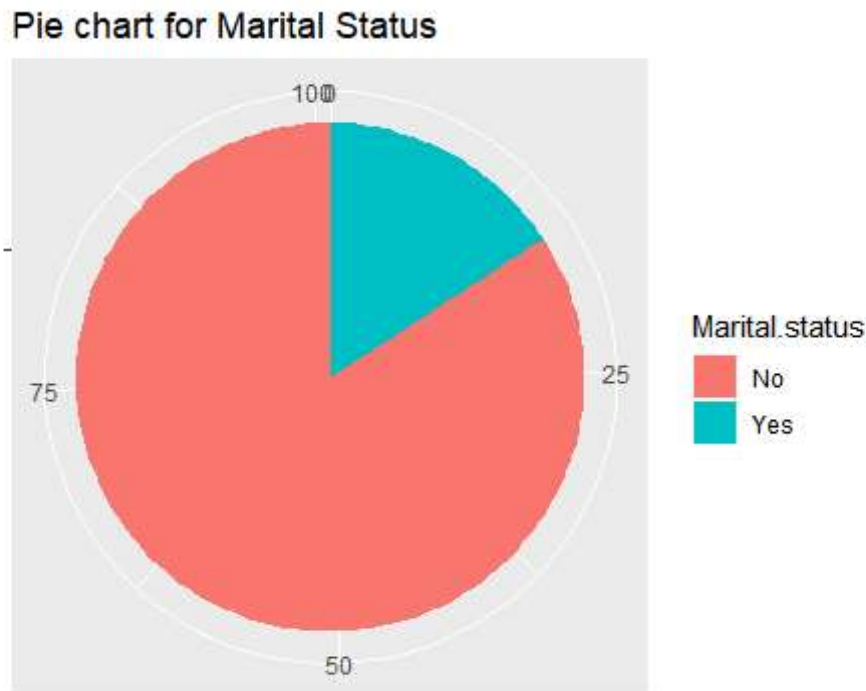
Year of Study Distribution

```
library(ggplot2)
ggplot(data = data, aes(x=year.of.Study))+
  geom_bar(na.rm = TRUE, fill = "red")+
  theme_grey()+
  labs(title = "Bar plot for Year of study")
```



Marital status

```
library(ggplot2)
pe = data.frame(table(Marital.status))
ggplot(pe, aes(x = "", y = Freq, fill = Marital.status))+
  geom_bar(stat = "identity", width = 1, )+
  coord_polar("y", start = 0)+
  theme_grey()+
  labs(title = "Pie chart for Marital Status", x = " ", y = " ")
```



Frequencies of Anxiety, Depression, Panic attack and Treatment status

```
Responses = c("No", "Yes")
summ = data %>%
  summarize(Anxiety = table(Do.you.have.Anxiety.),
            Depression = table(Do.you.have.Depression.),
            Panic.Attack = table(Do.you.have.Panic.attack.),
            Treatment = table(Did.you.seek.any.specialist.for.a.treatment.))
describe = data.frame(Responses, summ)
describe
```

```
## Responses Anxiety Depression Panic.Attack Treatment
## 1 No 67 66 68 95
## 2 Yes 34 35 33 6
```

Percentage of Anxiety, Depression, Panic attack and Treatment status

```
most = describe %>%
  summarize( n = n(),
             Anxiety = Anxiety/sum(Anxiety)*100,
             Depression = Depression/sum(Depression)*100,
             Panic.Attack = Panic.Attack/sum(Panic.Attack)*100,
             Treatment = Treatment/sum(Treatment)*100)

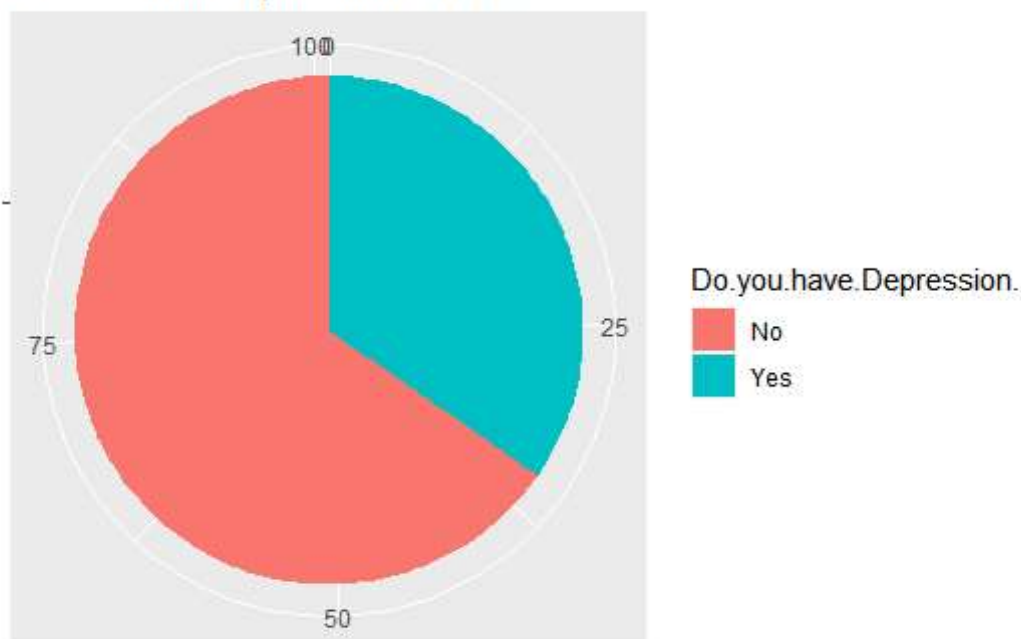
most
```

```
##   n Anxiety Depression Panic.Attack Treatment
## 1 2 66.33663   65.34653    67.32673 94.059406
## 2 2 33.66337   34.65347    32.67327  5.940594
```

Do you have Depression?

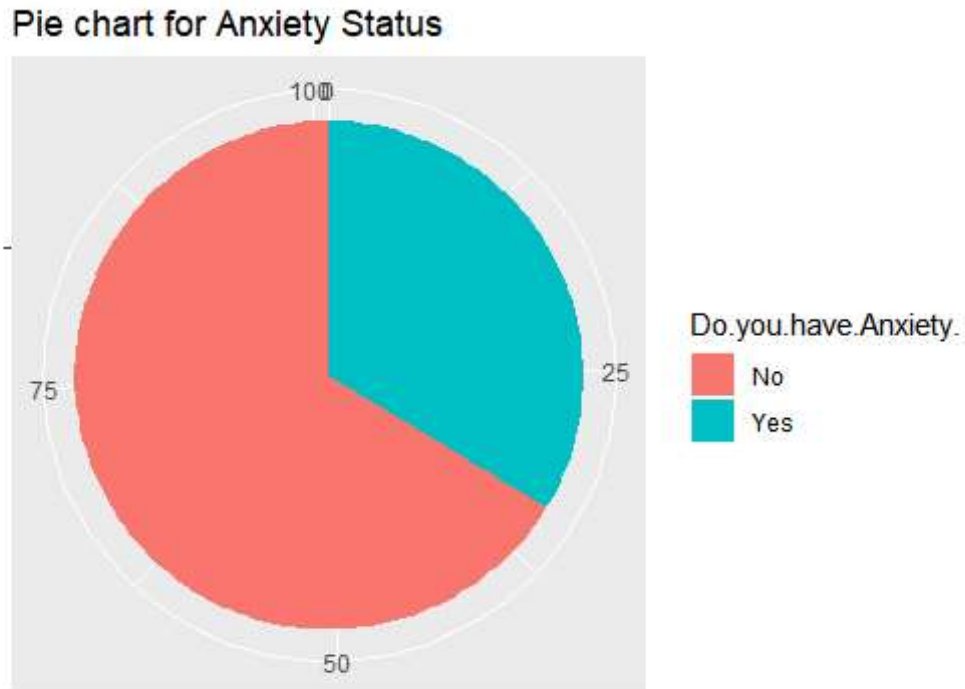
```
library(ggplot2)
pw = data.frame(table(Do.you.have.Depression.))
ggplot(pw, aes(x = "", y = Freq, fill = Do.you.have.Depression.))+
  geom_bar(stat = "identity", width = 1, )+
  coord_polar("y", start = 0)+
  theme_grey()+
  labs(title = "Pie chart for Depression Status", x = " ", y = " ")
```

Pie chart for Depression Status



Do you have Anxiety?

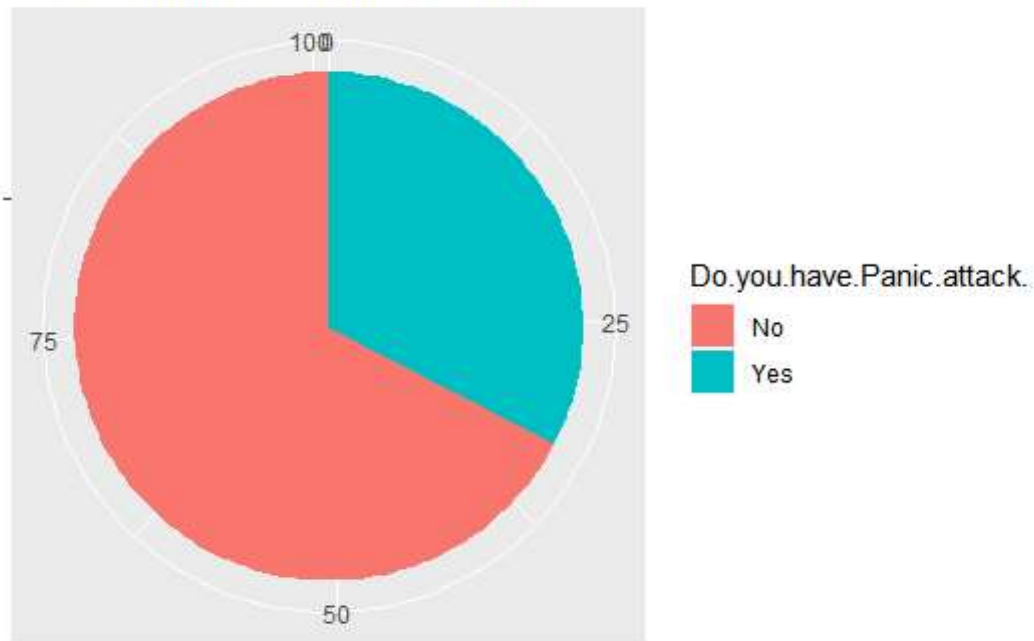
```
library(ggplot2)
pw = data.frame(table(Do.you.have.Anxiety.))
ggplot(pw, aes(x = "", y = Freq, fill = Do.you.have.Anxiety.))+
  geom_bar(stat = "identity", width = 1, )+
  coord_polar("y", start = 0)+
  theme_grey()+
  labs(title = "Pie chart for Anxiety Status", x = " ", y = " ")
```



Do you have Panic Attack?

```
library(ggplot2)
pw = data.frame(table(Do.you.have.Panic.attack.))
ggplot(pw, aes(x = "", y = Freq, fill = Do.you.have.Panic.attack.))+
  geom_bar(stat = "identity", width = 1, )+
  coord_polar("y", start = 0)+
  theme_grey()+
  labs(title = "Pie chart for Panic attack Status", x = " ", y = " ")
```

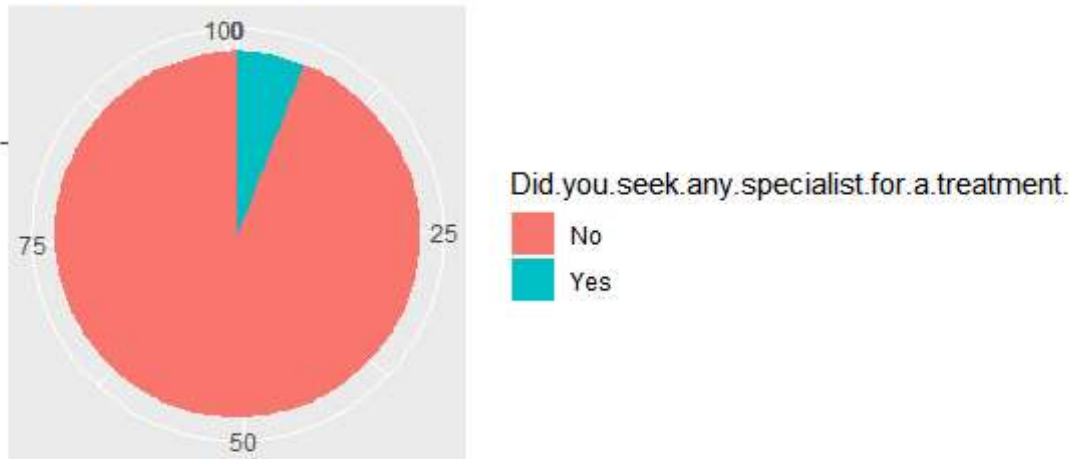
Pie chart for Panic attack Status



Did you seek any specialist for a treatment?

```
library(ggplot2)
pw = data.frame(table(Did.you.seek.any.specialist.for.a.treatment.))
ggplot(pw, aes(x = "", y = Freq, fill = Did.you.seek.any.specialist.for.a.treatment.))
  geom_bar(stat = "identity", width = 1, )+
  coord_polar("y", start = 0)+
  theme_grey()+
  labs(title = "Pie chart for specialist treatment status", x = " ", y = " ")
```

Pie chart for specialist treatment status



MULTIVARIATE DATA ANALYSIS AND VISUALIZATIONS

Comparative analysis of Average Age with respect to gender

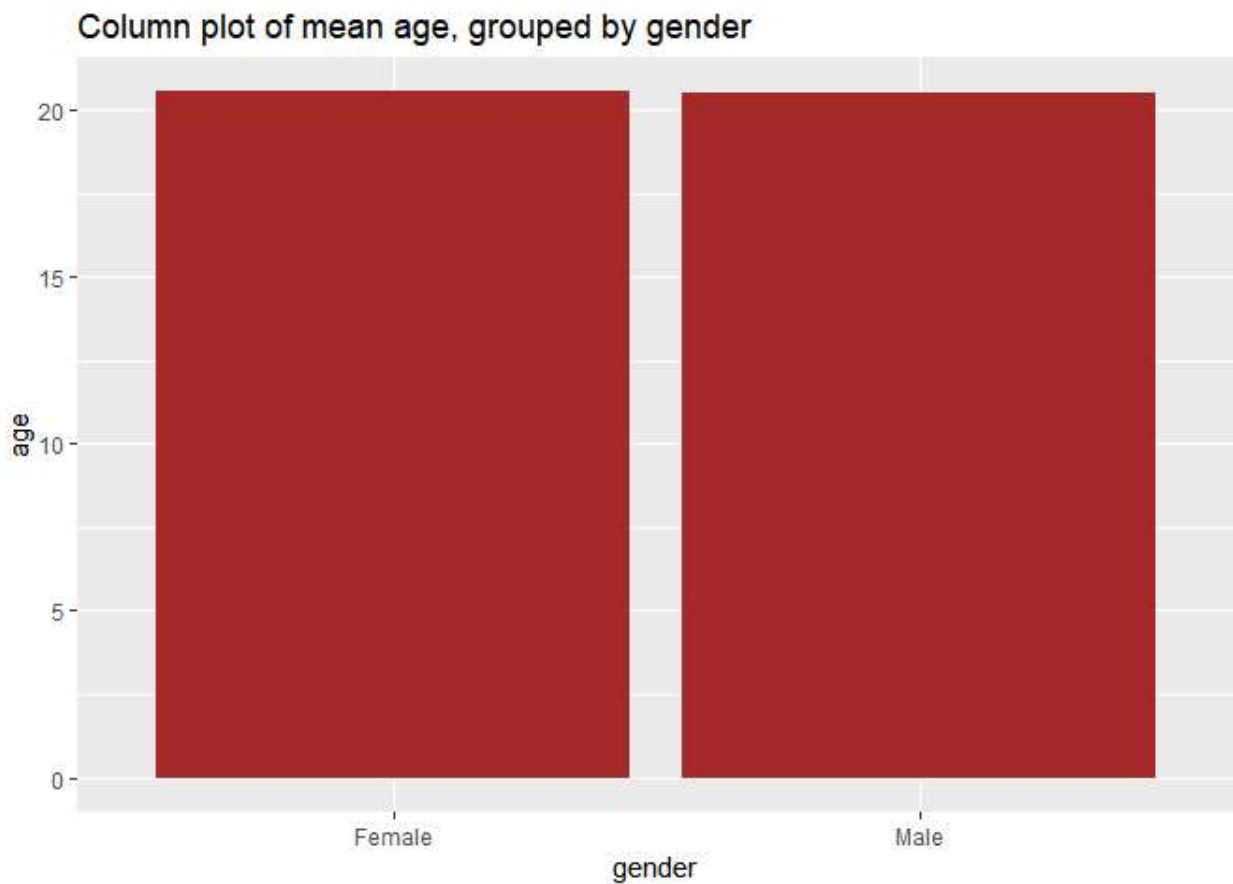
```
plotdata <- data %>%
  group_by(gender) %>%
  summarize(n = n(),
            mean = mean(Age, na.rm=T),
            sd = sd(Age, na.rm = T),
            se = sd/sqrt(n))
plotdata
```

```
## # A tibble: 2 x 5
##   gender      n mean    sd    se
##   <chr> <int> <dbl> <dbl> <dbl>
## 1 Female    75  20.5  2.52  0.291
## 2 Male     26  20.5  2.49  0.487
```

Comparative column plot of mean Age with respect to Gender

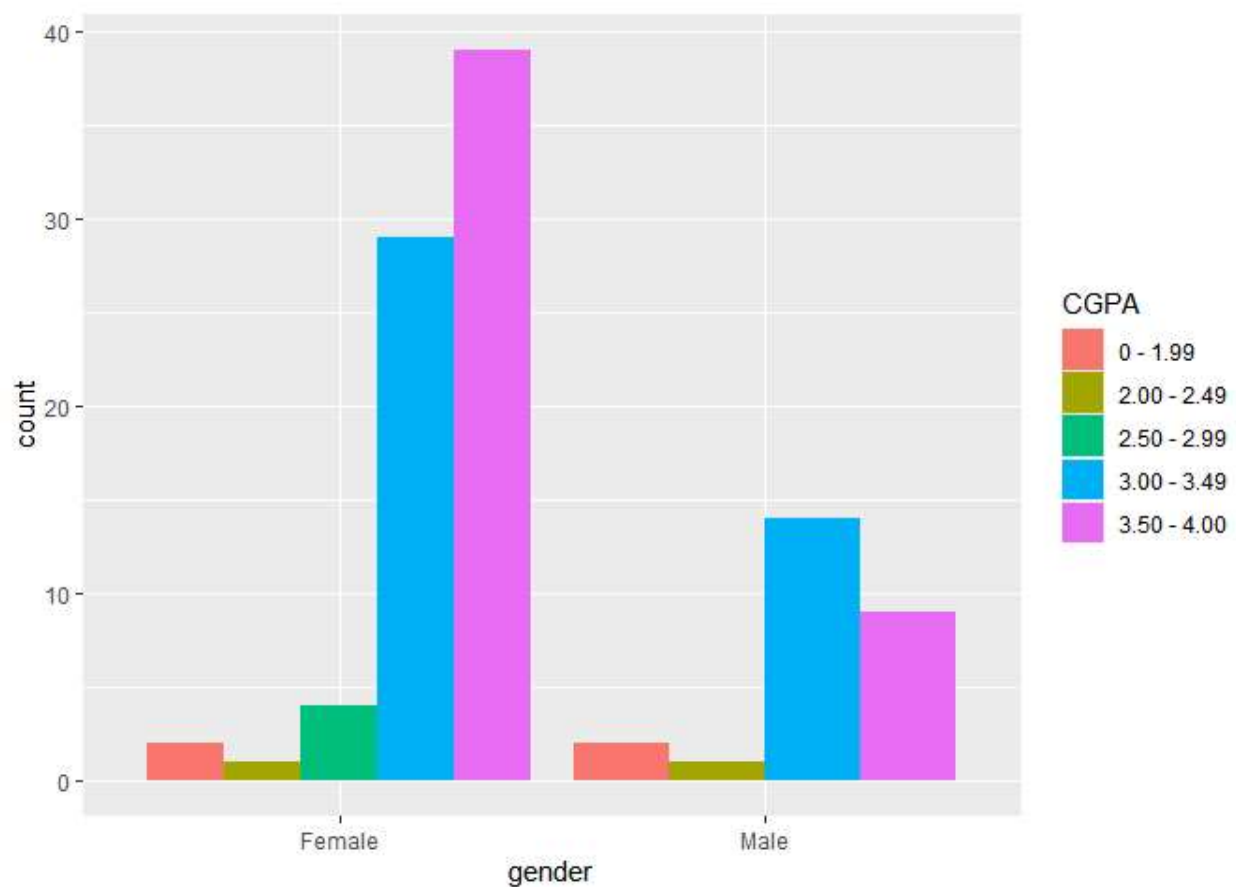
```
plotme <- data %>%  
  select(gender, Age)%>%  
  group_by(gender) %>%  
  summarize(age = mean(Age, na.rm = T)) %>%  
  ggplot(aes(x = gender, y = age))+  
  geom_col(fill = "brown")+  
  labs(title = "Column plot of mean age, grouped by gender")
```

plotme



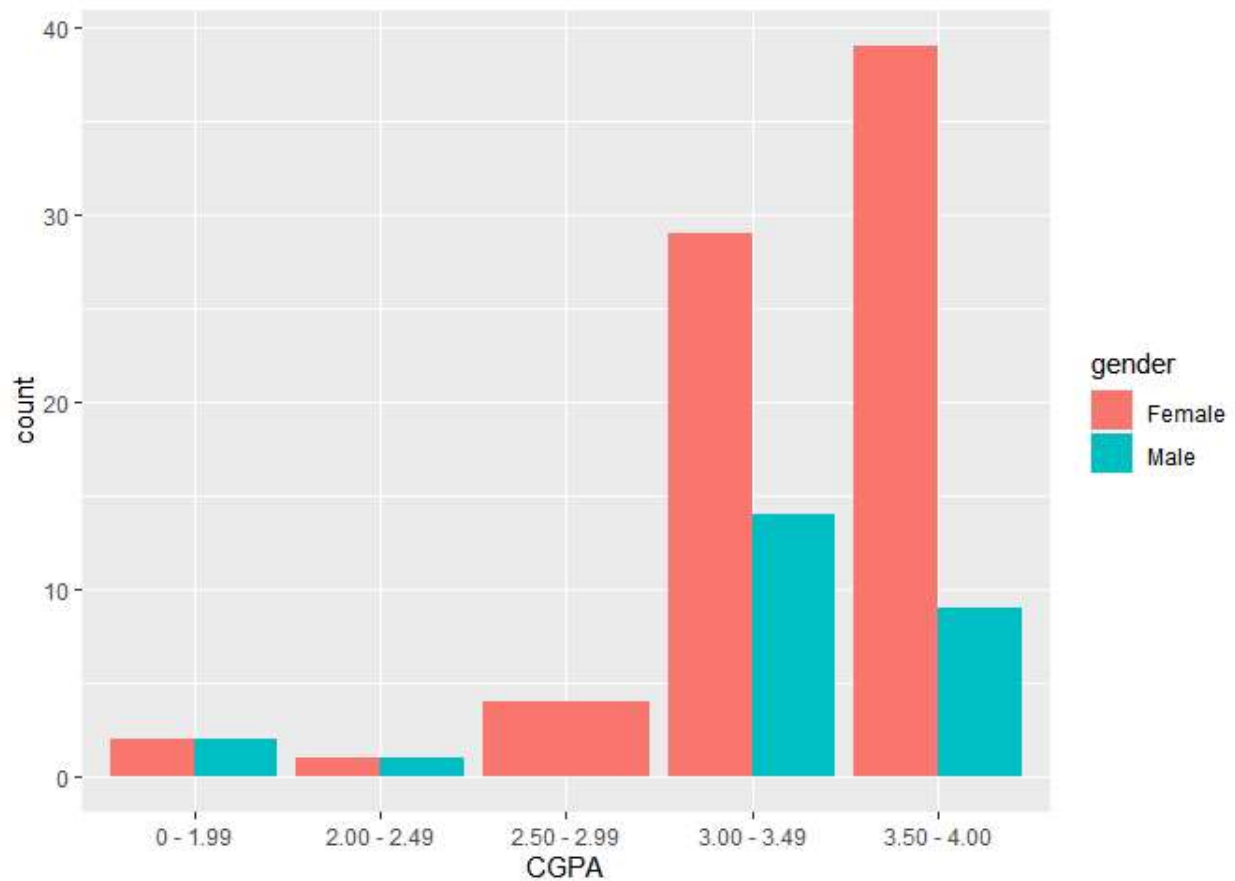
Comparative Bar plot of Gender with respect to CGPA

```
ggplot(data, aes(x = gender, fill = CGPA))+  
  geom_bar(position = "dodge")
```



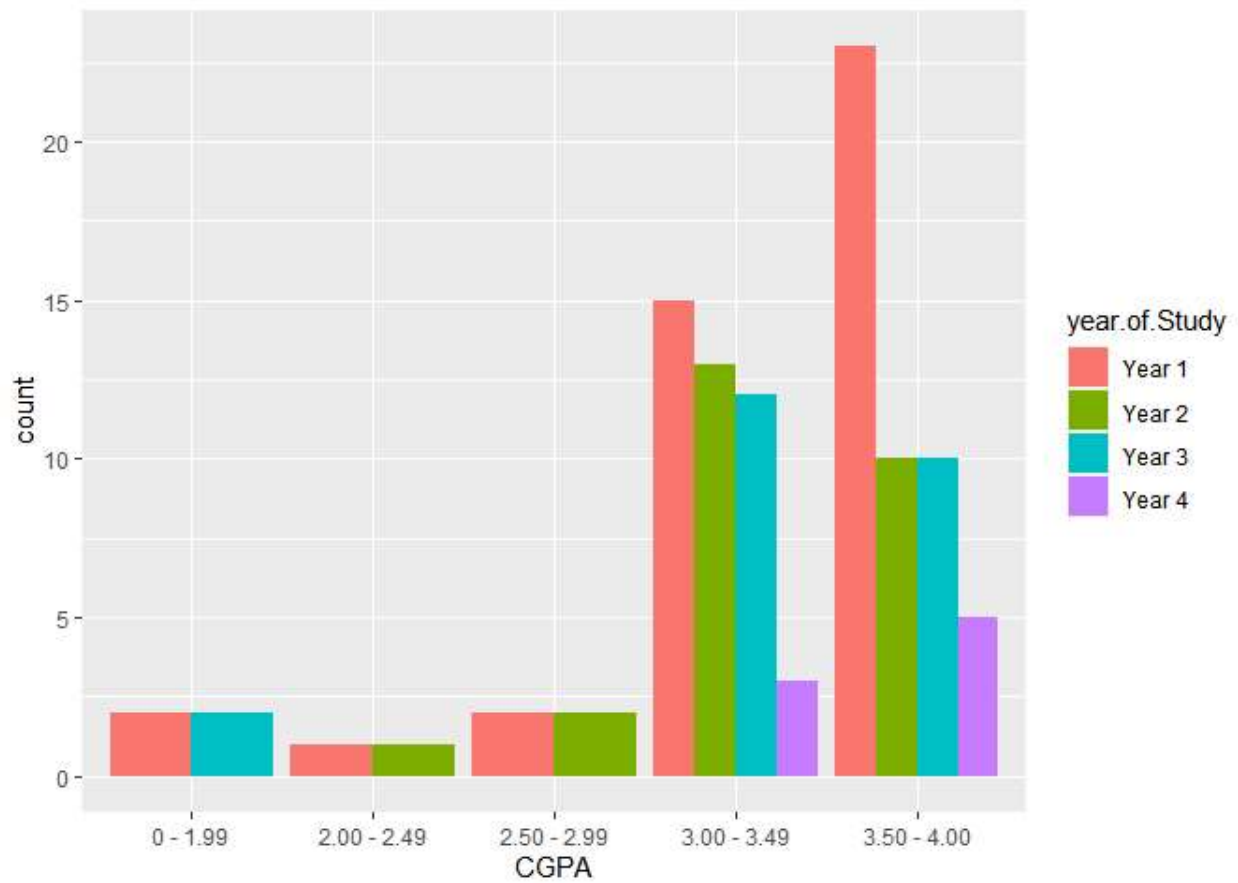
Comparative Bar plot of CGPA with respect to Gender

```
ggplot(data, aes(x = CGPA, fill = gender))+  
geom_bar(position = "dodge")
```



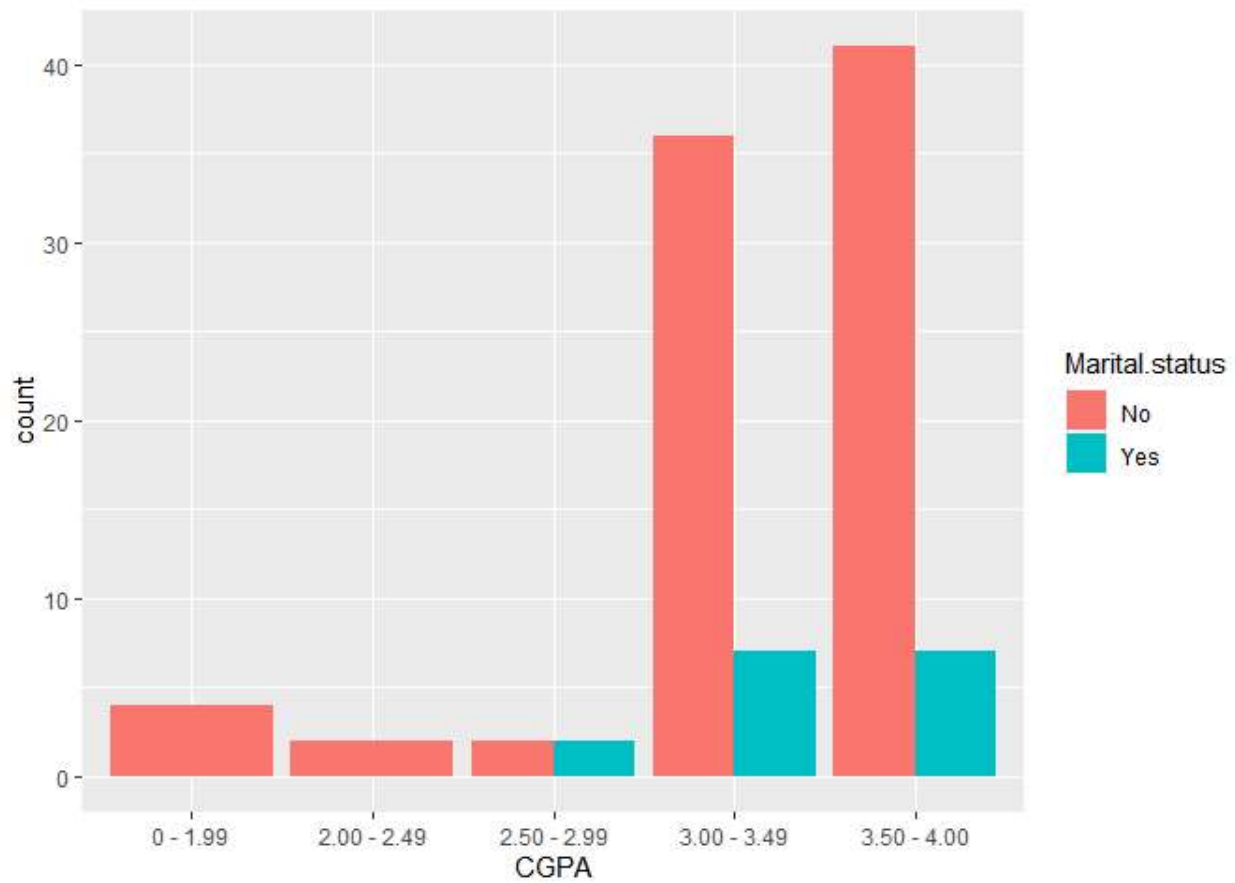
Comparative Bar plot of study year with respect to CGPA

```
ggplot(data, aes(x = CGPA, fill = year.of.Study))+  
geom_bar(position = "dodge")
```

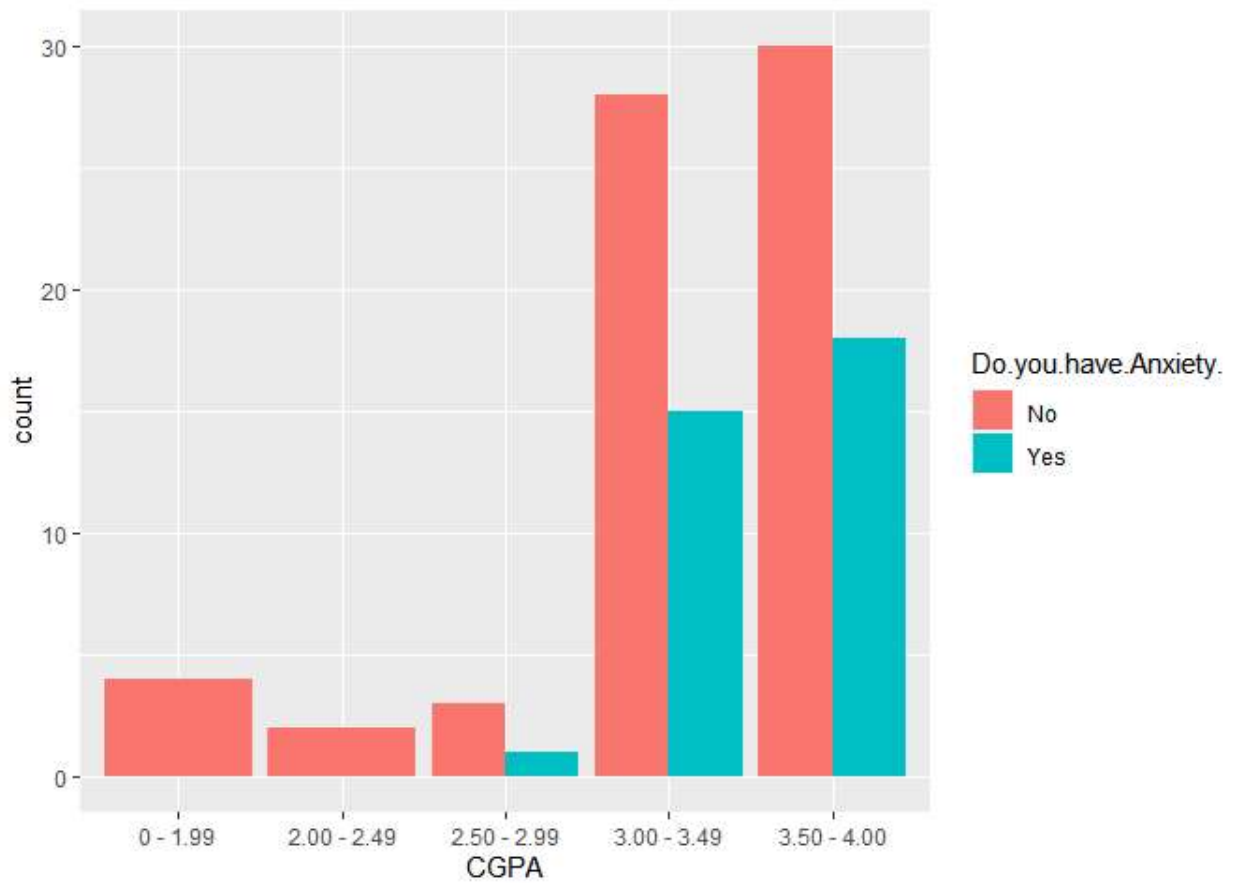
Comparative Bar plot of Marital status with respect to CGPA

```
ggplot(data, aes(x = CGPA, fill = Marital.status))+  
geom_bar(position = "dodge")
```



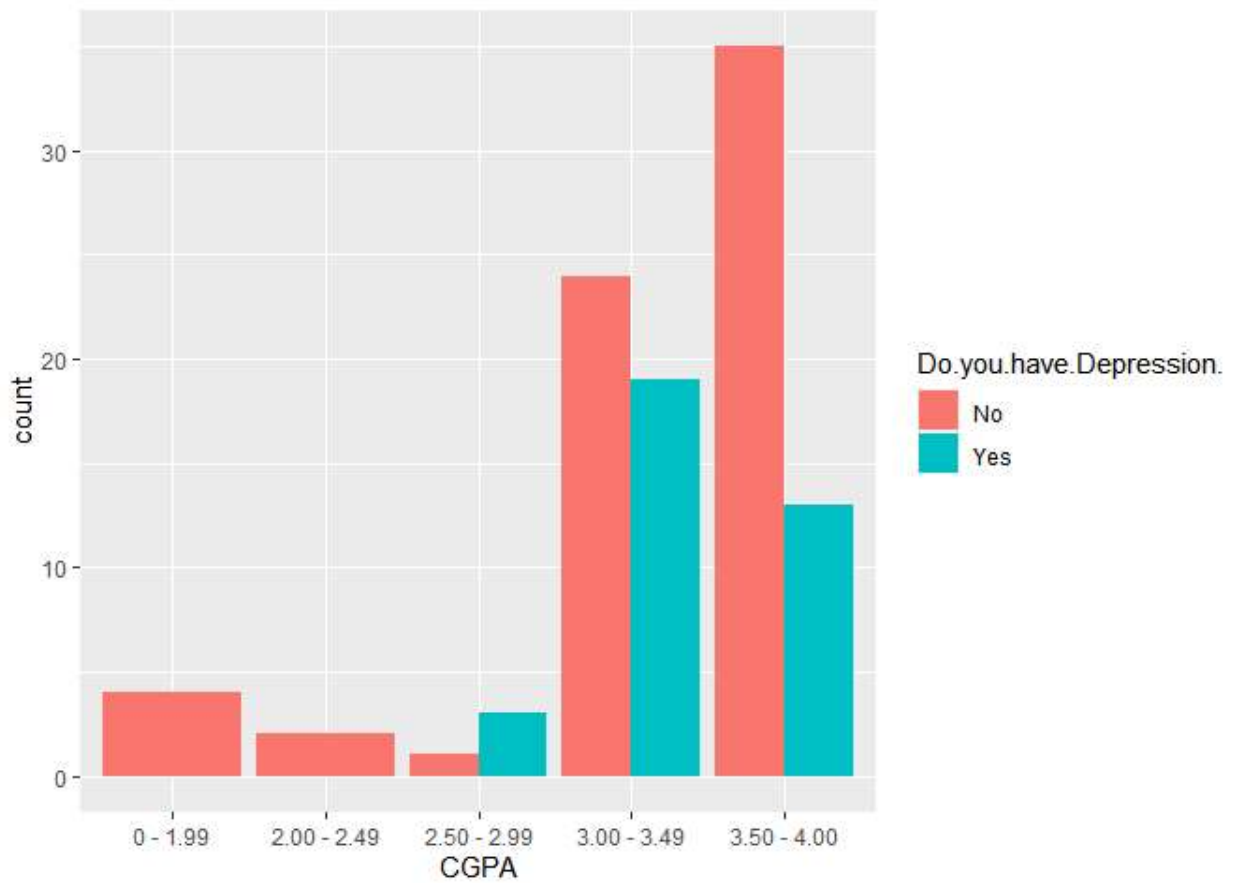
Comparative Bar plot of Anxiety status with respect to CGPA

```
ggplot(data, aes(x = CGPA, fill = Do.you.have.Anxiety.)) +  
  geom_bar(position = "dodge")
```



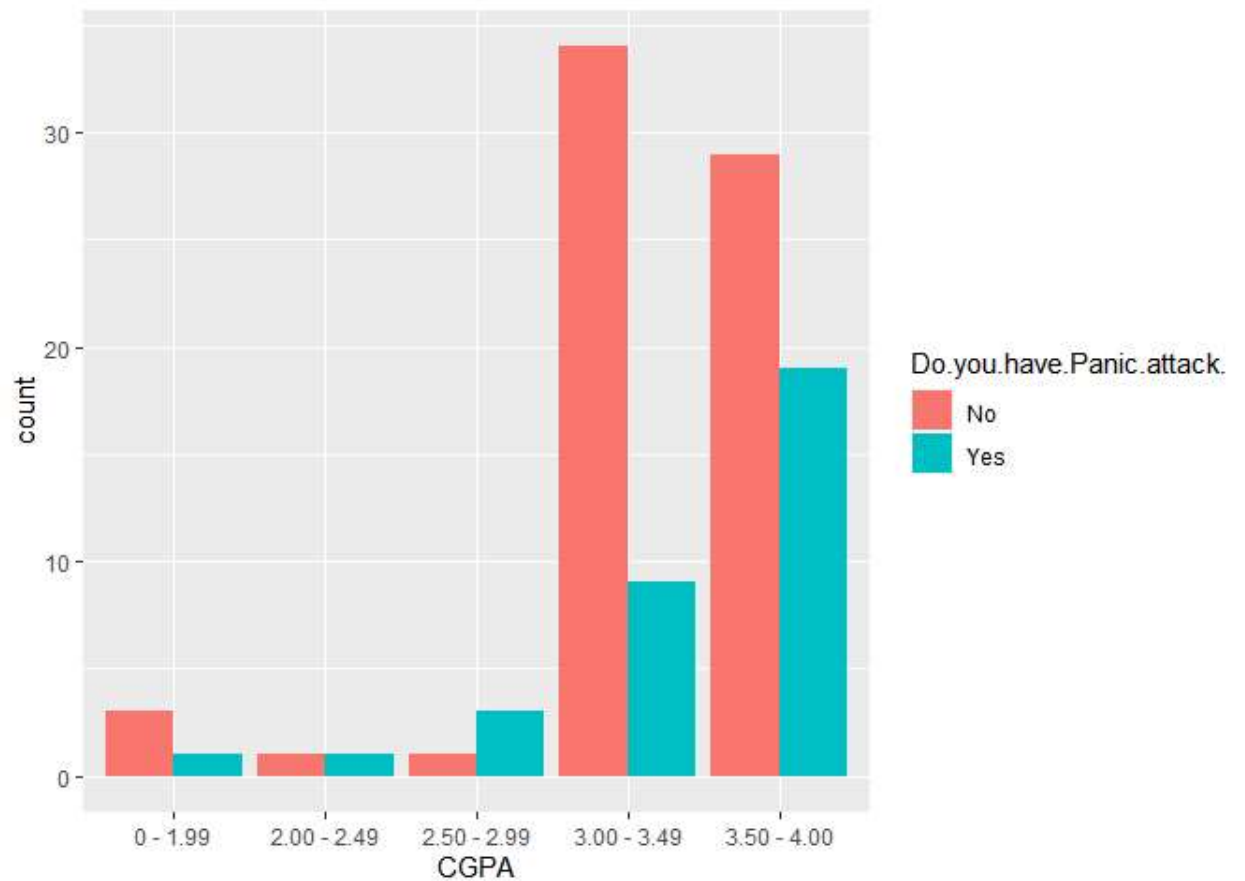
Comparative Bar plot of Depression with respect to CGPA

```
ggplot(data, aes(x = CGPA, fill = Do.you.have.Depression.))+  
geom_bar(position = "dodge")
```



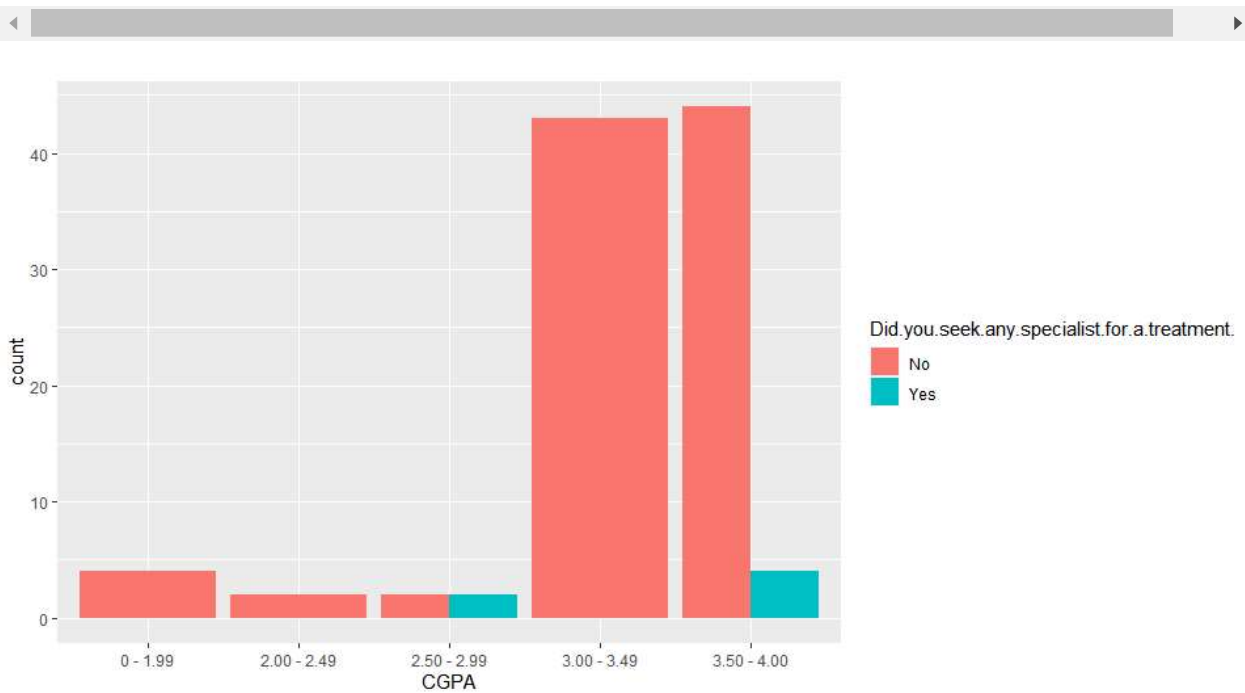
Comparative Bar plot of Panic attack with respect to CGPA

```
ggplot(data, aes(x = CGPA, fill = Do.you.have.Panic.attack.))+  
geom_bar(position = "dodge")
```



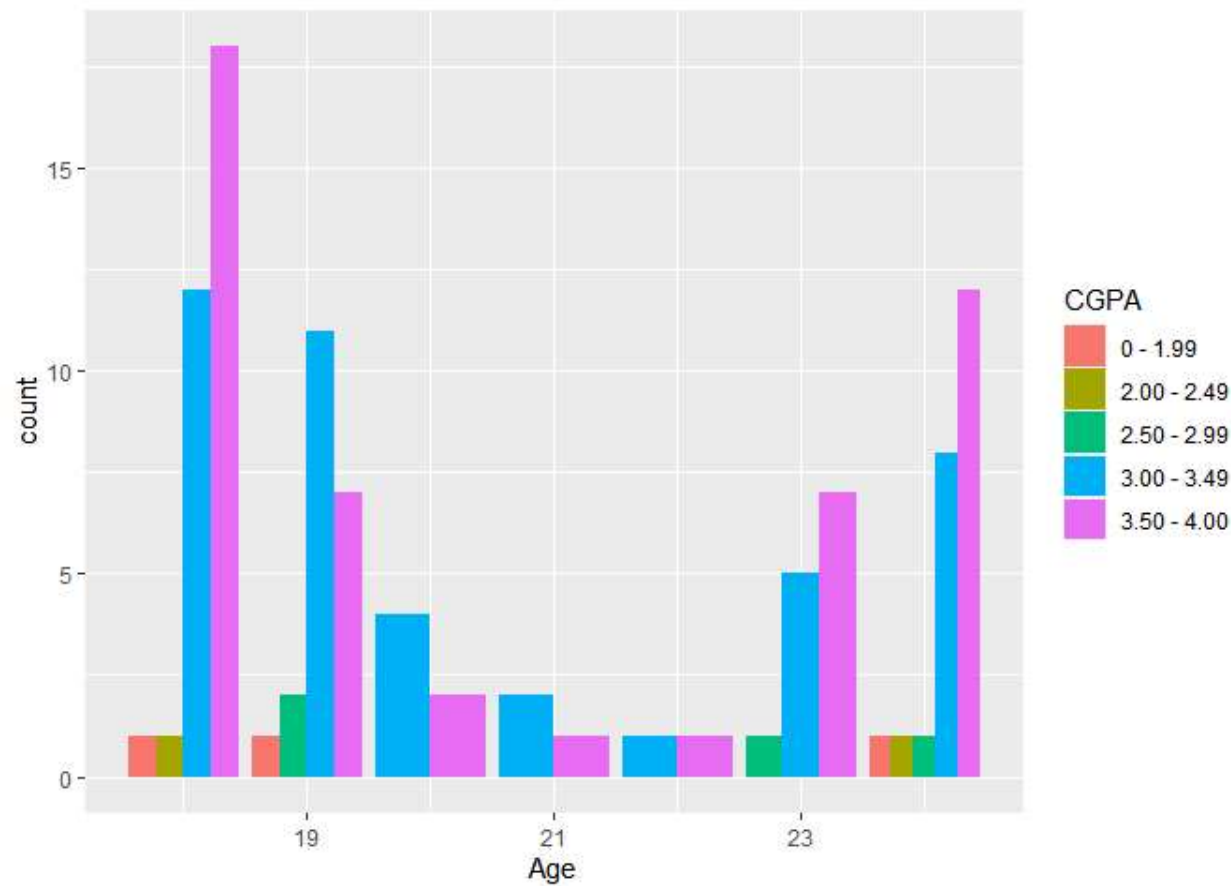
Comparative Bar plot of Treatment status with respect to CGPA

```
ggplot(data, aes(x = CGPA, fill = Did.you.seek.any.specialist.for.a.treatment.))  
geom_bar(position = "dodge")
```



Comparative Bar plot of Age with respect to CGPA

```
ggplot(data, aes(x = Age, fill = CGPA))+  
geom_bar(position = "dodge", na.rm = T)
```



```
#geom_text(aes(label=round( ,1)), hjust = -.1)
```

INFERENCIAL STATISTICS

Investigating the impact of Anxiety on students CGPA

```
a = table(CGPA, Do.you.have.Anxiety.)  
a
```

| ## | | Do.you.have.Anxiety. | |
|---------|----------|----------------------|-----|
| ## CGPA | | No | Yes |
| ## | 0 - 1.99 | 4 | 0 |

```
##      2.00 - 2.49  2   0
##      2.50 - 2.99  3   1
##      3.00 - 3.49 28  15
##      3.50 - 4.00 30  18
```

```
chisq.test(a)
```

```
##
##  Pearson's Chi-squared test
##
## data:  a
## X-squared = 3.5243, df = 4, p-value = 0.4742
```

```
b = table(CGPA, Do.you.have.Depression.)
b
```

```
##              Do.you.have.Depression.
## CGPA              No Yes
##  0 - 1.99          4   0
##  2.00 - 2.49       2   0
##  2.50 - 2.99       1   3
##  3.00 - 3.49      24  19
##  3.50 - 4.00      35  13
```

```
chisq.test(b)
```

```
##
##  Pearson's Chi-squared test
##
## data:  b
## X-squared = 8.9975, df = 4, p-value = 0.06116
```

```
c = table(CGPA, Do.you.have.Panic.attack.)
c
```

```
##              Do.you.have.Panic.attack.
## CGPA              No Yes
##  0 - 1.99          3   1
##  2.00 - 2.49       1   1
##  2.50 - 2.99       1   3
```

```
##      3.00 - 3.49 34    9
##      3.50 - 4.00 29   19
```

```
chisq.test(c)
```

```
##
##  Pearson's Chi-squared test
##
## data:  c
## X-squared = 7.3752, df = 4, p-value = 0.1173
```

To investigate any relationship or association between gender, marital status and CGPA

```
d = table(CGPA, gender)
d
```

```
##           gender
## CGPA      Female Male
##  0 - 1.99         2    2
##  2.00 - 2.49         1    1
##  2.50 - 2.99         4    0
##  3.00 - 3.49        29   14
##  3.50 - 4.00        39    9
```

```
chisq.test(d)
```

```
##
##  Pearson's Chi-squared test
##
## data:  d
## X-squared = 5.5063, df = 4, p-value = 0.2392
```

```
e = table(CGPA, Marital.status)
e
```

```
##           Marital.status
## CGPA      No Yes
##  0 - 1.99         4    0
```



```
## 2.00 - 2.49 2 0
## 2.50 - 2.99 2 2
## 3.00 - 3.49 36 7
## 3.50 - 4.00 41 7
```

```
chisq.test(e)
```

```
##
## Pearson's Chi-squared test
##
## data: e
## X-squared = 4.6933, df = 4, p-value = 0.3202
```

FINDINGS AND CONCLUSIONS

- From the result, female seems to be higher than male with frequency ratio of 75:26, and the students CGPA seems to be depending on age as female students are seen to have the highest CGPA in each CGPA grades, with same averages of ages (20years).
- Also, CGPA seems to be having some kind of relationship with students year of study, as students in first years shows to have the highest CGPA, while those in the forth level has the lowest counts of grade point, and the CGPA counts decreases over the study years especially in the Upper level and Distinction grade points.
- Marital staus also shows a positive impact on CGPA as single students dominated Upper credit and Distinction grape point, though few married students also dominated these points, but not even close to that of single students.
- Amxiety as well have an impact on CGPA as students with no anxiety are far more than those with anxiety in the grade points.
- Depression can also determine students CGPA as students with no depression cases also counted higher than those with depressions in the grade points, except in lower credit, and same thing goes for Panic attack as well.
- Also, Age proves to determine CGPA with the highest distinction and upper credit counts falling under the lowest age of 18.

Note: However, we cannot make much conclusions with these visualization results as the sample size is is small, and the gender distribution counts is too far from each other.

- From the chisquare test, we can conclude that anxiety has no significance relationship with CGPA, neither does Depression nor Panic attack has any significance relationship with CGPA.

- Gender and Marital status also tested to have no statistical significance with CGPA
- Therefore , we can conclude that none of depression, anxiety, panic attack, gender or marital status cannot determine CGPA, and students can pass and have good grades despite experiencing anxiety, depression, panic attack, being any of the gender or even being single or married.