

fp_step2_100

Zeyu Hou

2025-05-11

Data Cleaning:

```
summary(data)
```

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

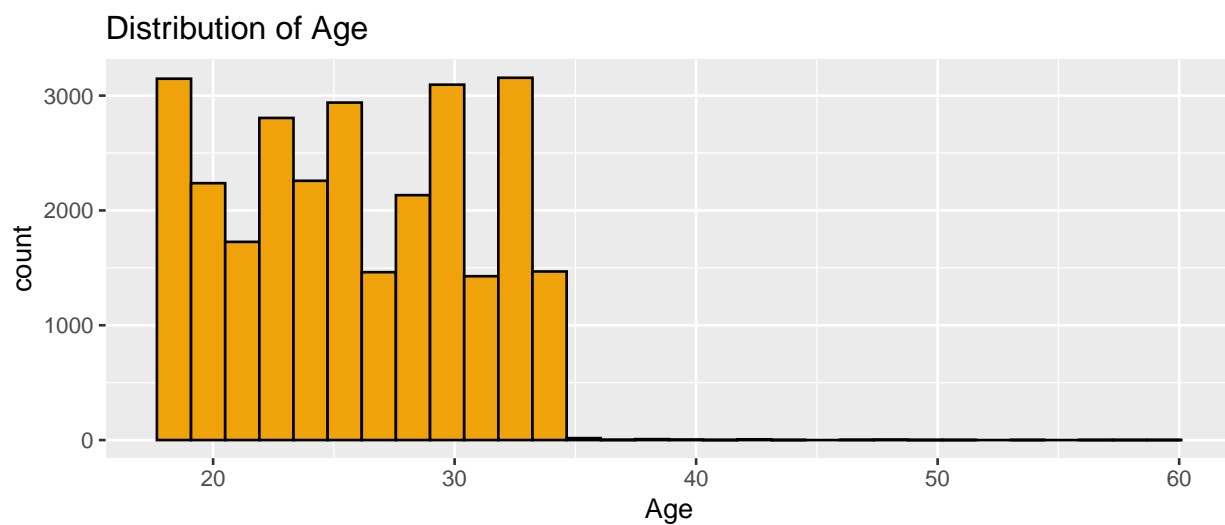
```
colSums(is.na(data))
```

```
##           id           Gender
##           0           0
##          Age           City
##           0           0
##      Profession Academic.Pressure
##           0           0
##      Work.Pressure           CGPA
##           0           0
##      Study.Satisfaction Job.Satisfaction
##           0           0
##      Sleep.Duration      Dietary.Habits
##           0           0
##      Degree Have.you.ever.had.suicidal.thoughts..
##           0           0
##      Work.Study.Hours      Financial.Stress
##           0           0
##      Family.History.of.Mental.Illness      Depression
##           0           0
```

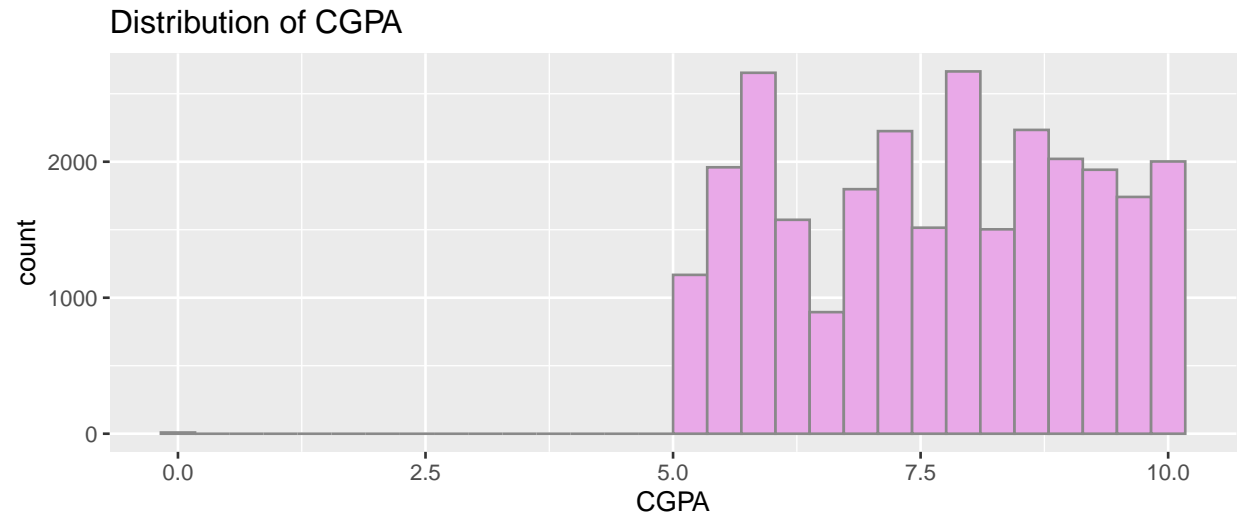
since no na, no need for data cleaning :)

Visualization:

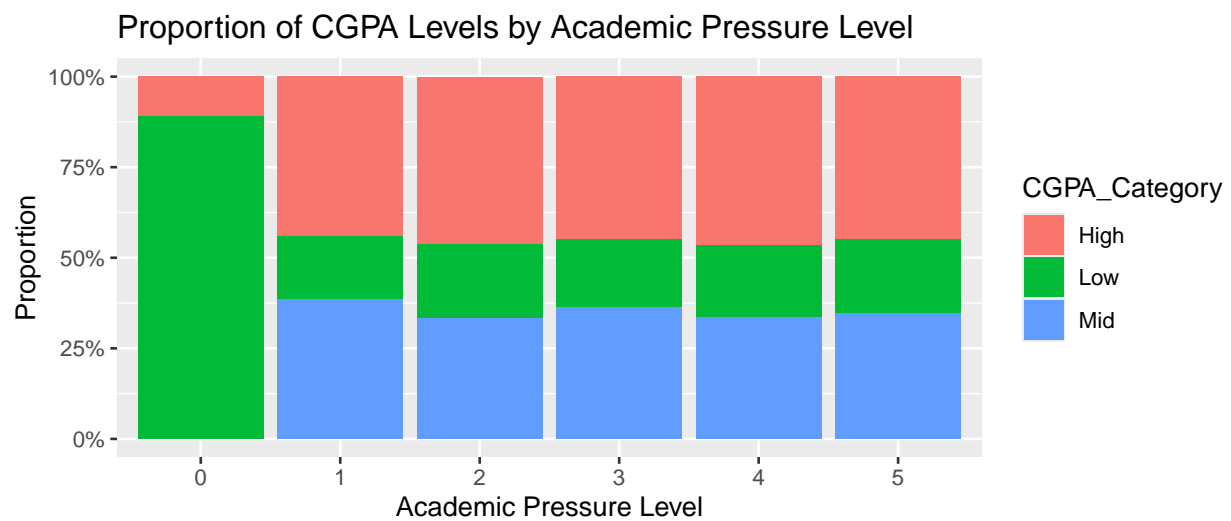
```
# Distribution of Age
p1
```



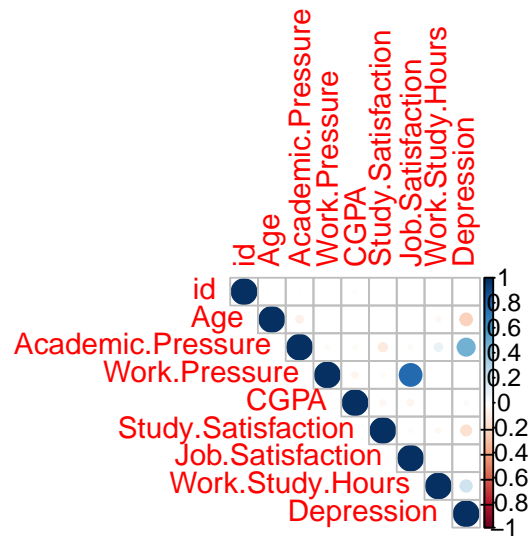
```
# Distribution of CGPA
p2
```



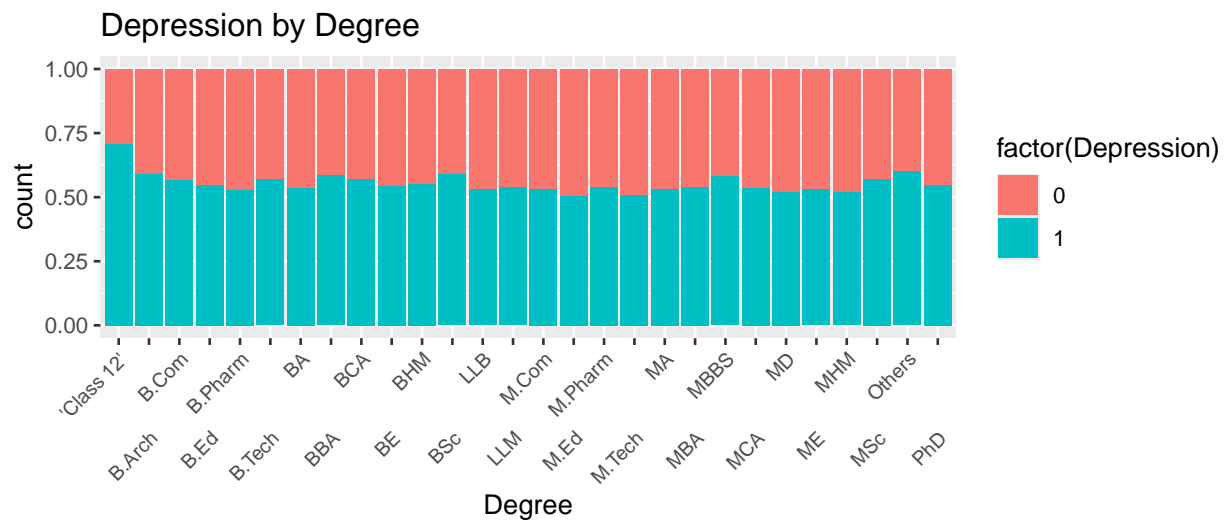
```
# CGPA vs Academic Pressure
p3
```



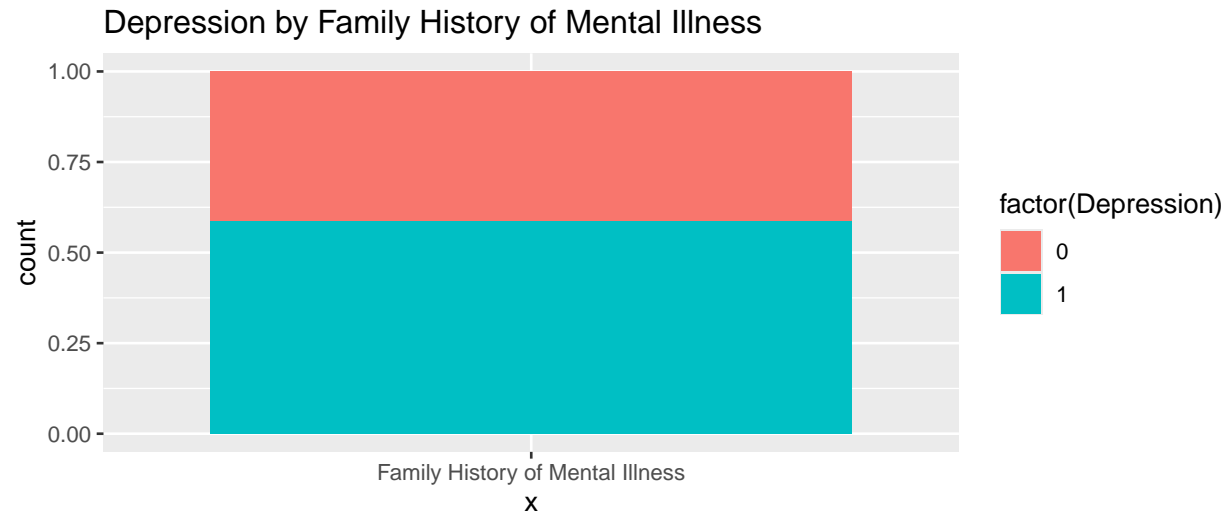
```
# Correlation Matrix
numeric_cols <- sapply(data, is.numeric)
cor_matrix <- cor(data[, numeric_cols], use = 'complete.obs')
corrplot(cor_matrix, method = 'circle', type = 'upper')
```



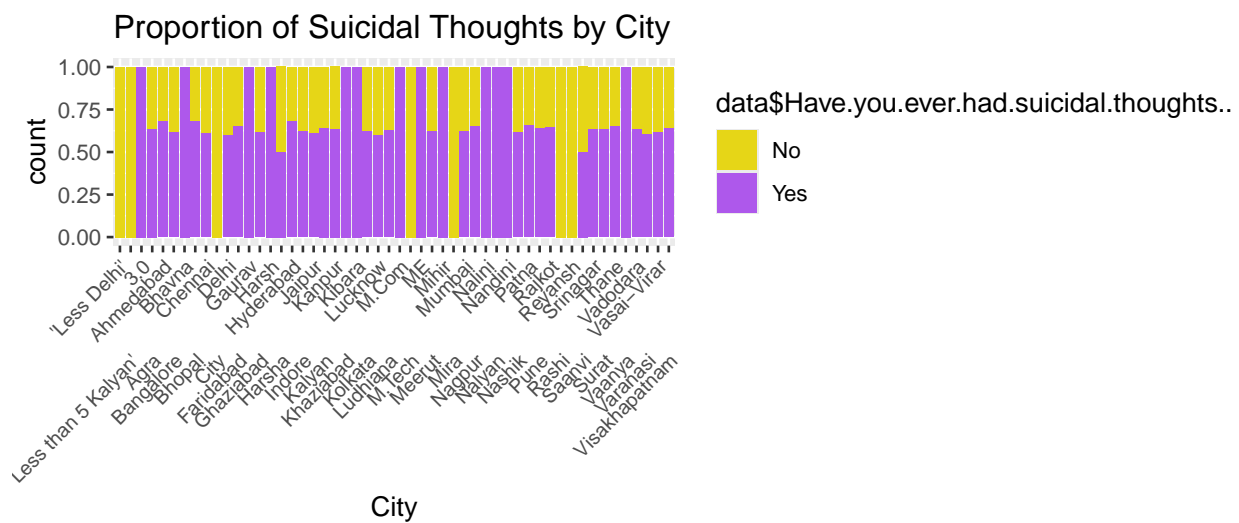
Depression by Degree
p5



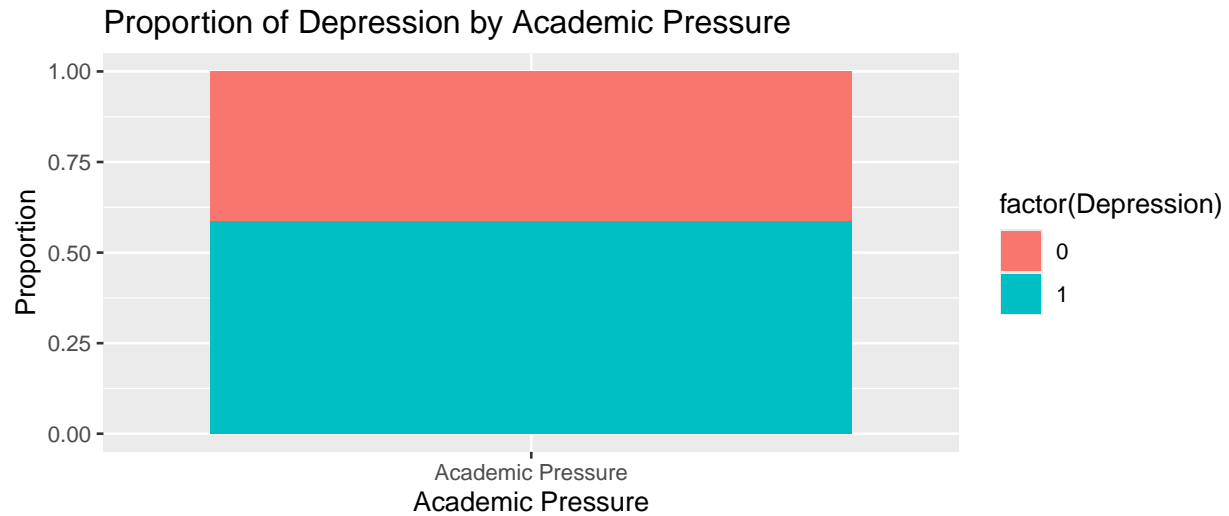
Family History of Mental Illness vs Depression
p6



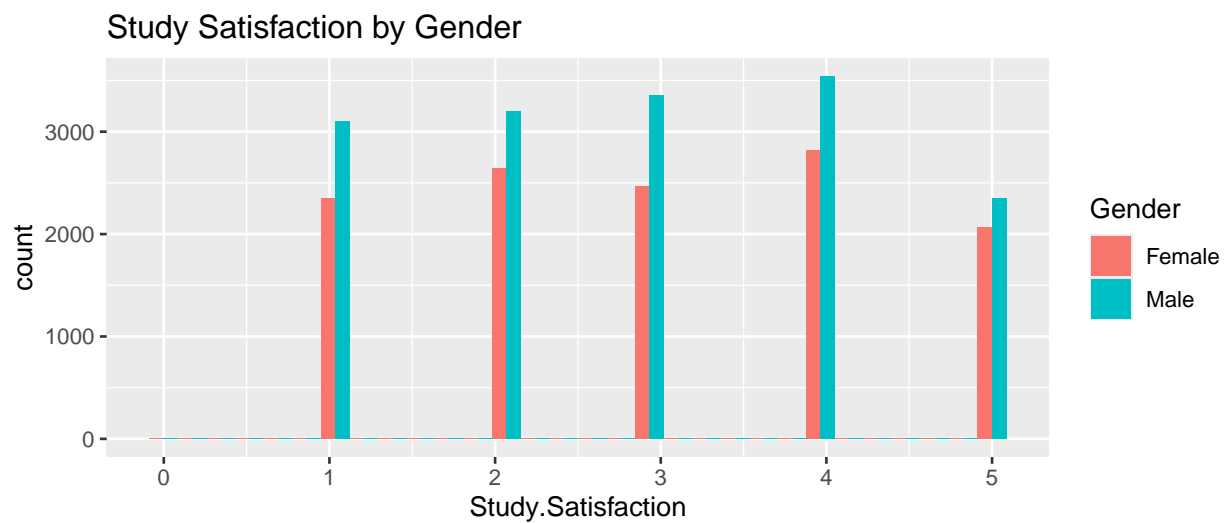
Proportion of Suicidal Thoughts by City
p7



Depression vs Academic Pressure
p8



Distribution Comparisons - Study Satisfaction by Gender
p9



```
# Skewness and Kurtosis
age_skewness <- skewness(data$Age, na.rm = TRUE)
age_kurtosis <- kurtosis(data$Age, na.rm = TRUE)
cgpa_skewness <- skewness(data$CGPA, na.rm = TRUE)
cgpa_kurtosis <- kurtosis(data$CGPA, na.rm = TRUE)
age_skewness
```

```
## [1] 0.1322247
```

```
age_kurtosis
```

```
## [1] -0.8464239
```

```
cgpa_skewness
```

```
## [1] -0.1130511
```

```
cgpa_kurtosis
```

```
## [1] -1.023317
```