

# EDA\_Tommy Lee

Tommy Lee

2024-02-23

## 1: Introduction

What is depression? Depression is a mood disorder that affects roughly over 20 million adults in the United States alone. Depression is also called major depressive disorder or clinical depression. This disorder can affect how you feel, behave and think, which could possibly lead to a variety of problems. Someone suffering from depression may have trouble doing the most basic day to day activities and may even have suicidal tendencies.

This data set represents the depression levels of adults, which includes 294 observations and 37 variables. In this final report, we will be focused on two variables, the age of the person and the sex of the person to determine which age and sex have the highest reported cases of depression.

```
depression <-read.table("C:/Math 130/Data/Depress.txt", header=TRUE, sep="\t")
str(depression)
```

```
## 'data.frame': 294 obs. of 37 variables:
## $ ID : int 1 2 3 4 5 6 7 8 9 10 ...
## $ SEX : int 2 1 2 2 2 1 2 1 2 1 ...
## $ AGE : int 68 58 45 50 33 24 58 22 47 30 ...
## $ MARITAL : int 5 3 2 3 4 2 2 1 2 2 ...
## $ EDUCAT : int 2 4 3 3 3 3 2 3 3 2 ...
## $ EMPLOY : int 4 1 1 3 1 1 5 1 4 1 ...
## $ INCOME : int 4 15 28 9 35 11 11 9 23 35 ...
## $ RELIG : int 1 1 1 1 1 1 1 1 2 4 ...
## $ C1 : int 0 0 0 0 0 0 2 0 0 0 ...
## $ C2 : int 0 0 0 0 0 0 1 1 1 0 ...
## $ C3 : int 0 1 0 0 0 0 1 2 1 0 ...
## $ C4 : int 0 0 0 0 0 0 2 0 0 0 ...
## $ C5 : int 0 0 1 1 0 0 1 2 0 0 ...
## $ C6 : int 0 0 0 1 0 0 0 1 3 0 ...
## $ C7 : int 0 0 0 0 0 0 0 0 0 0 ...
## $ C8 : int 0 0 0 3 3 0 2 0 0 0 ...
## $ C9 : int 0 0 0 0 3 1 2 0 0 0 ...
## $ C10 : int 0 0 0 0 0 0 0 0 0 0 ...
## $ C11 : int 0 0 0 0 0 0 0 0 0 0 ...
## $ C12 : int 0 1 0 0 0 1 0 0 3 0 ...
## $ C13 : int 0 0 0 0 0 2 0 0 0 0 ...
## $ C14 : int 0 0 1 0 0 0 0 0 3 0 ...
## $ C15 : int 0 1 1 0 0 0 3 0 2 0 ...
## $ C16 : int 0 0 1 0 0 2 0 1 3 0 ...
## $ C17 : int 0 1 0 0 0 1 0 1 0 0 ...
## $ C18 : int 0 0 0 0 0 0 0 1 0 0 ...
```

```
## $ C19      : int  0 0 0 0 0 0 0 1 0 0 ...
## $ C20      : int  0 0 0 0 0 0 1 0 0 0 ...
## $ CESD     : int  0 4 4 5 6 7 15 10 16 0 ...
## $ CASES    : int  0 0 0 0 0 0 0 0 1 0 ...
## $ DRINK    : int  2 1 1 2 1 1 2 2 1 1 ...
## $ HEALTH   : int  2 1 2 1 1 1 3 1 4 1 ...
## $ REGDOC   : int  1 1 1 1 1 1 1 2 1 1 ...
## $ TREAT    : int  1 1 1 2 1 1 1 2 1 2 ...
## $ BEDDAYS  : int  0 0 0 0 1 0 0 0 1 0 ...
## $ ACUTEILL: int  0 0 0 0 1 1 1 1 0 0 ...
## $ CHRONILL: int  1 1 0 1 0 1 1 0 1 0 ...
```

```
head(depression)
```

```
##   ID SEX AGE MARITAL EDUCAT EMPLOY INCOME RELIG C1 C2 C3 C4 C5 C6 C7 C8 C9 C10
## 1  1  2  68      5      2      4      4      1  0  0  0  0  0  0  0  0  0  0
## 2  2  1  58      3      4      1     15      1  0  0  1  0  0  0  0  0  0
## 3  3  2  45      2      3      1     28      1  0  0  0  0  1  0  0  0  0
## 4  4  2  50      3      3      3      9      1  0  0  0  0  1  1  0  3  0
## 5  5  2  33      4      3      1     35      1  0  0  0  0  0  0  0  3  3
## 6  6  1  24      2      3      1     11      1  0  0  0  0  0  0  0  0  1
##   C11 C12 C13 C14 C15 C16 C17 C18 C19 C20 CESD CASES DRINK HEALTH REGDOC TREAT
## 1  0  0  0  0  0  0  0  0  0  0  0  0  2  2  1  1
## 2  0  1  0  0  1  0  1  0  0  0  4  0  1  1  1  1
## 3  0  0  0  1  1  1  0  0  0  0  4  0  1  2  1  1
## 4  0  0  0  0  0  0  0  0  0  0  5  0  2  1  1  2
## 5  0  0  0  0  0  0  0  0  0  0  6  0  1  1  1  1
## 6  0  1  2  0  0  2  1  0  0  0  7  0  1  1  1  1
##   BEDDAYS ACUTEILL CHRONILL
## 1      0      0      1
## 2      0      0      1
## 3      0      0      0
## 4      0      0      1
## 5      1      1      0
## 6      0      1      1
```

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

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

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

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

## 2: Univariate Exploration

First we will look at their AGE:

```
summary(depression$AGE)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  18.00   28.00   42.50   44.41   59.00   89.00
```

```
mean(depression$AGE)
```

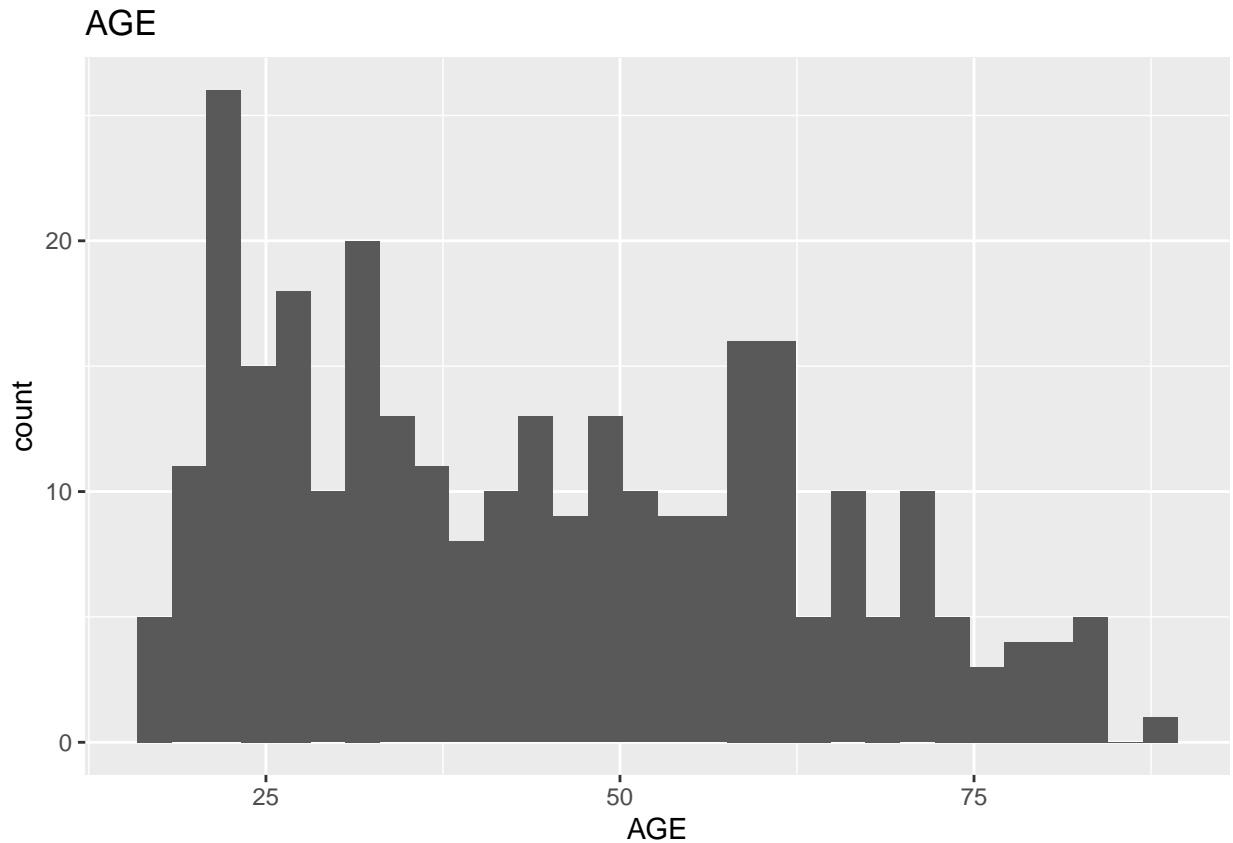
```
## [1] 44.41497
```

```
table(depression$AGE)
```

```
##
## 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43
##  5  5  6  6  9 11  9  6  9  4  5  4  6  5 10  5  9  4  6  5  2  1  5  1  9  7
## 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69
##  2  4  3  6  4  4  5  6  4  2  3  4  3  6  7  9  7  5  4  2  3  5  3  2  4  1
## 70 71 72 73 74 75 77 78 79 80 81 82 83 89
##  5  3  2  2  3  1  2  2  2  1  2  1  5  1
```

```
ggplot(depression, aes(x=AGE)) + geom_histogram() + ggtitle("AGE")
```

```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



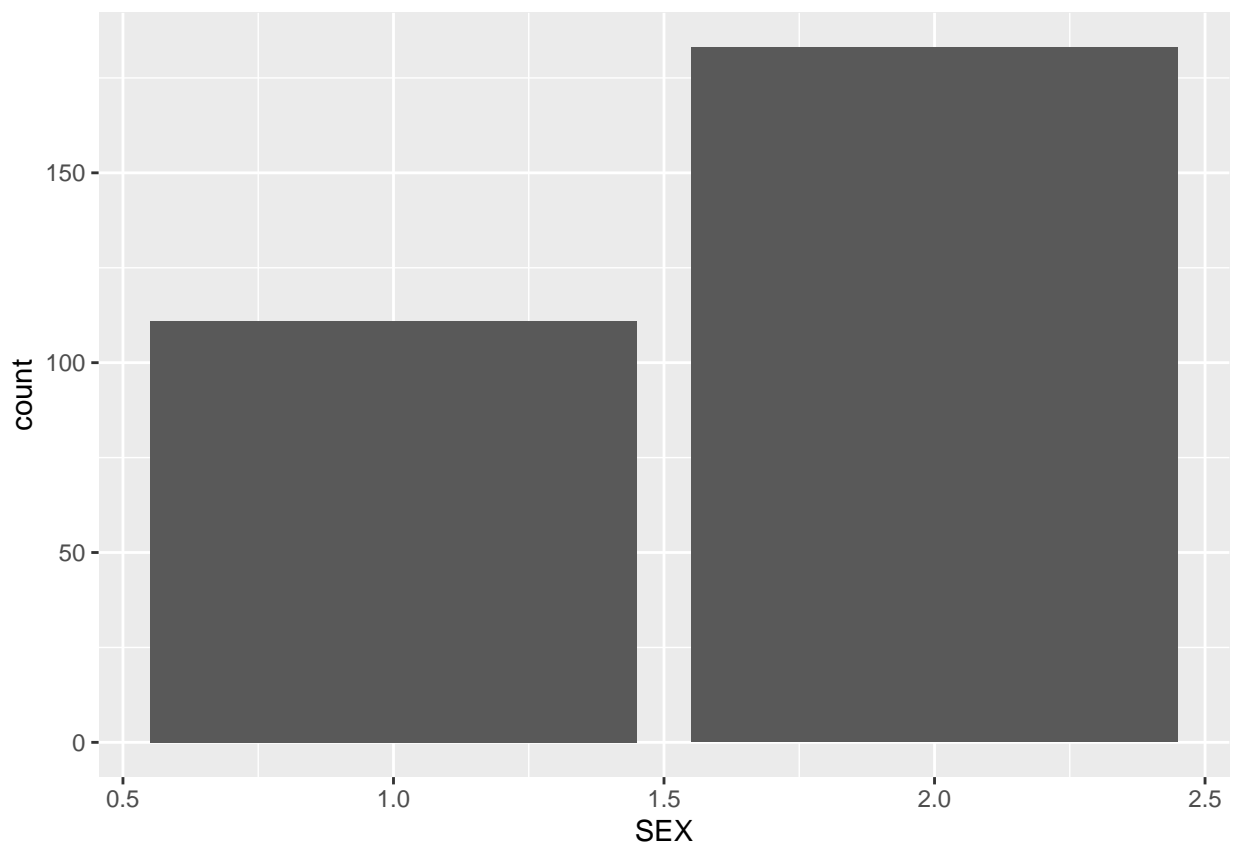
Why is this histogram important? This relays information about the level of depression indicated of the adults by age. This helps us get a better visualization at what the highest depressed age group. We can see that the amount of people in the age gap between 20 to 30 is more abundant than the rest of the age groups.

Next, we will look at their SEX (1 is male and 2 is female):

```
table(depression$SEX)
```

```
##
##  1  2
## 111 183
```

```
ggplot(depression, aes(x=SEX)) + geom_bar()
```



This bar chart is used to show the distinct difference between the male and female data sets. This indicates that there are more females in this study than there are males.

### 3: Bivariate Exploration

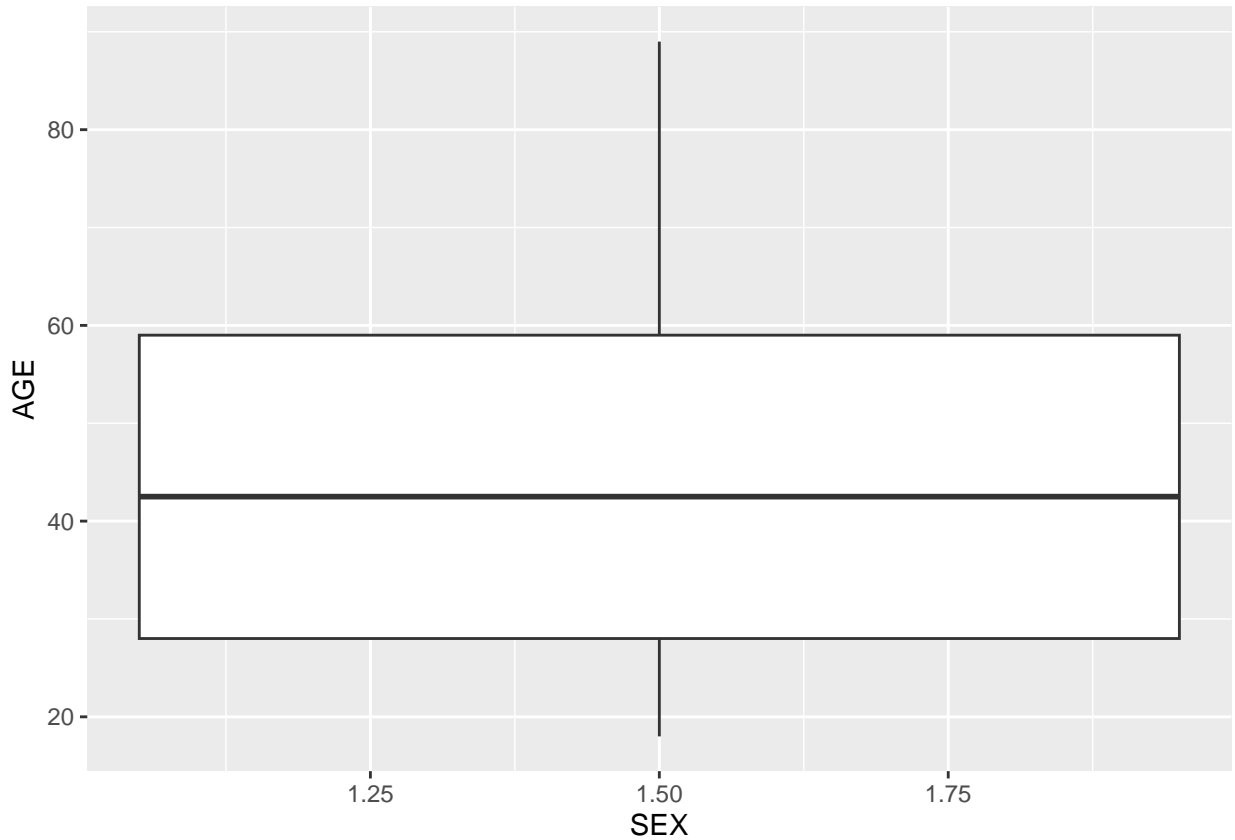
```
table(depression$SEX, depression$AGE) %>% prop.table(margin =2) %>% round(3)
```

```
##
##      18      19      20      21      22      23      24      25      26      27      28      29
##  1 0.400 0.400 0.667 0.333 0.556 0.182 0.111 0.333 0.556 0.500 0.200 0.500
##  2 0.600 0.600 0.333 0.667 0.444 0.818 0.889 0.667 0.444 0.500 0.800 0.500
##
```

```
##      30   31   32   33   34   35   36   37   38   39   40   41
##  1 0.333 0.600 0.600 0.000 0.111 0.250 0.500 0.600 0.500 1.000 0.800 0.000
##  2 0.667 0.400 0.400 1.000 0.889 0.750 0.500 0.400 0.500 0.000 0.200 1.000
##
##      42   43   44   45   46   47   48   49   50   51   52   53
##  1 0.333 0.143 0.500 0.500 0.667 0.333 0.750 0.750 0.400 0.167 0.250 0.500
##  2 0.667 0.857 0.500 0.500 0.333 0.667 0.250 0.250 0.600 0.833 0.750 0.500
##
##      54   55   56   57   58   59   60   61   62   63   64   65
##  1 0.000 0.250 0.333 0.167 0.429 0.444 0.429 0.400 0.750 0.500 1.000 0.200
##  2 1.000 0.750 0.667 0.833 0.571 0.556 0.571 0.600 0.250 0.500 0.000 0.800
##
##      66   67   68   69   70   71   72   73   74   75   77   78
##  1 0.000 0.000 0.000 0.000 0.400 0.000 0.500 1.000 0.333 0.000 0.500 0.500
##  2 1.000 1.000 1.000 1.000 0.600 1.000 0.500 0.000 0.667 1.000 0.500 0.500
##
##      79   80   81   82   83   89
##  1 0.500 0.000 0.500 0.000 0.200 0.000
##  2 0.500 1.000 0.500 1.000 0.800 1.000
```

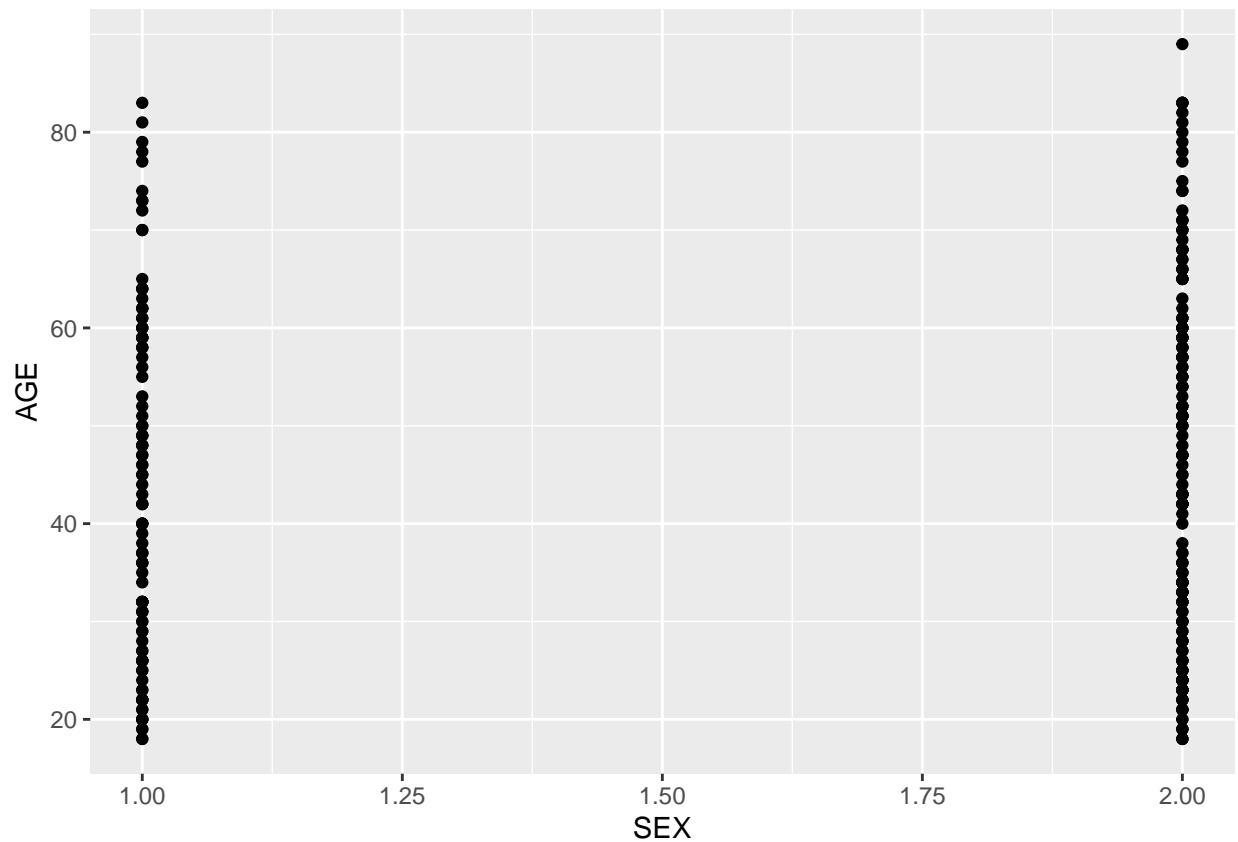
```
ggplot(depression, aes(x=SEX, y=AGE)) + geom_boxplot()
```

```
## Warning: Continuous x aesthetic
## i did you forget 'aes(group = ...)'?
```



This graph shows a combined data minimum, maximum, and upper and lower quartile of the male and female depression data sets.

```
ggplot(depression, aes(x=SEX, y=AGE)) + geom_point()
```



This point figure shows a better comparison to male and female age group that experience depression. From this we can see that females have a higher age that experience depression while the males have a lower age that experience depression. It is hard to tell from this figure but there are more females than males. **4: Conclusion**

The data presented shows the age group along with the sex of the people that took part in this data set. While looking at the data set gathered from the depression data, I found that there is a significant amount of more females depressed than males. This would also make me think how they conducted this experiment, if they randomly sampled or if they had a set number of people for each male and female. All in all the majority of those that are depressed are on the younger age group, the 18-30 group. I hypothesize that this would be the general time where people would start to think about their future or their careers or that they are simply lost in life.