

Depression Data Analysis Project,

Katie Curiel

3/8/2024

```
knitr::opts_chunk$set(echo = TRUE, message = FALSE, warning = FALSE)
library(dplyr)
library(ggplot2)
depress <- read.delim("/Users/katiecuriel/Desktop/MATH130/DATA/Depress.txt",
                      header = TRUE, sep="\t")
dim(depress)
```

```
## [1] 294 37
```

Introduction

The data set being explored in this project is the Depression data set which surveyed depression in 294 adults from Los Angeles County. The survey asked them their age, marital status, and their sex. I'm interested in finding out if more women or men experience depression and at what ages it's the most common. I want to see what age is the most common for depression.

Univariate Exploration

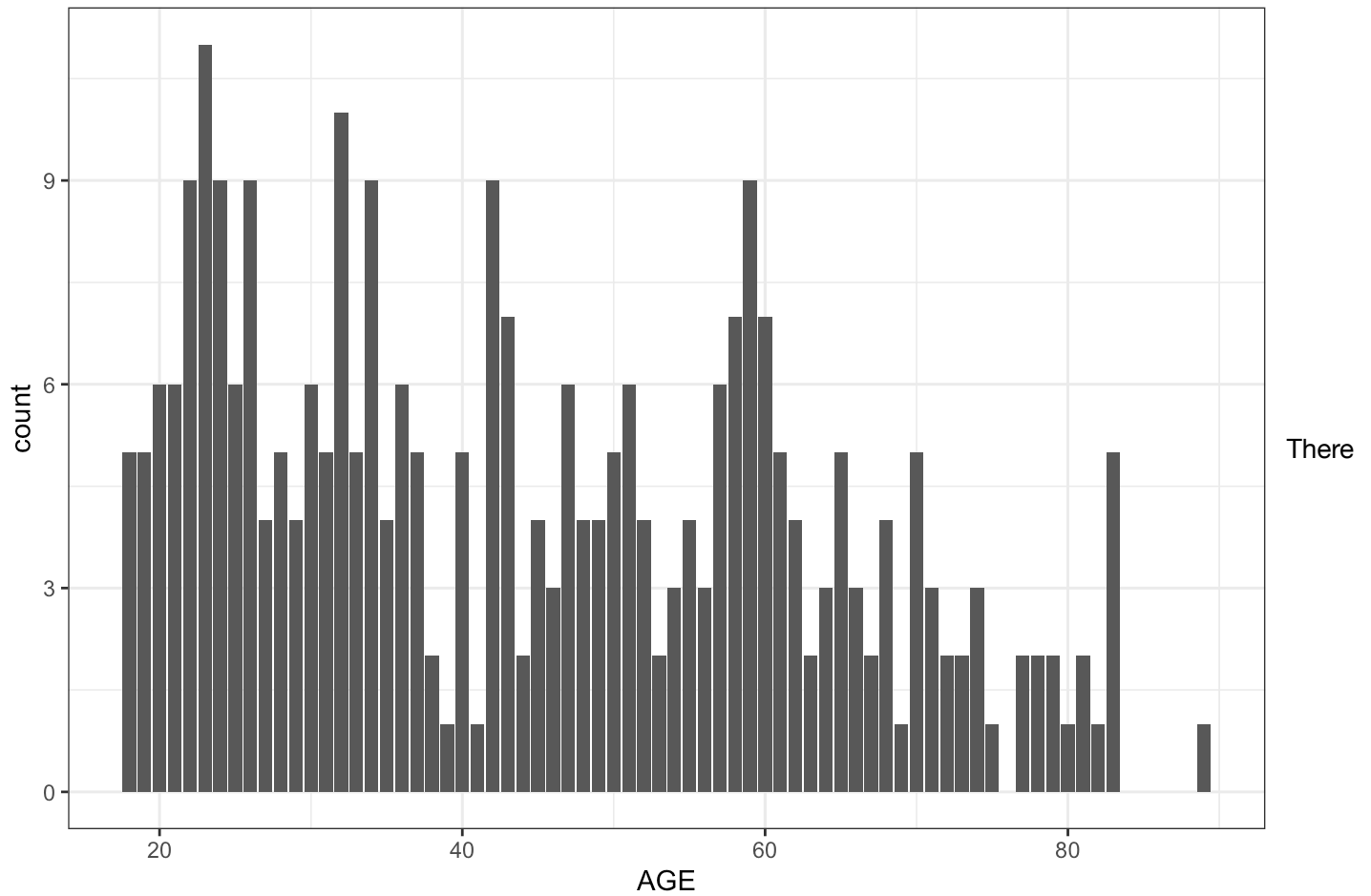
Age 1.

```
table(depress$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
```

2.

```
ggplot(depress, aes(x = AGE, fill = AGE)) + theme_bw() + geom_bar()
```

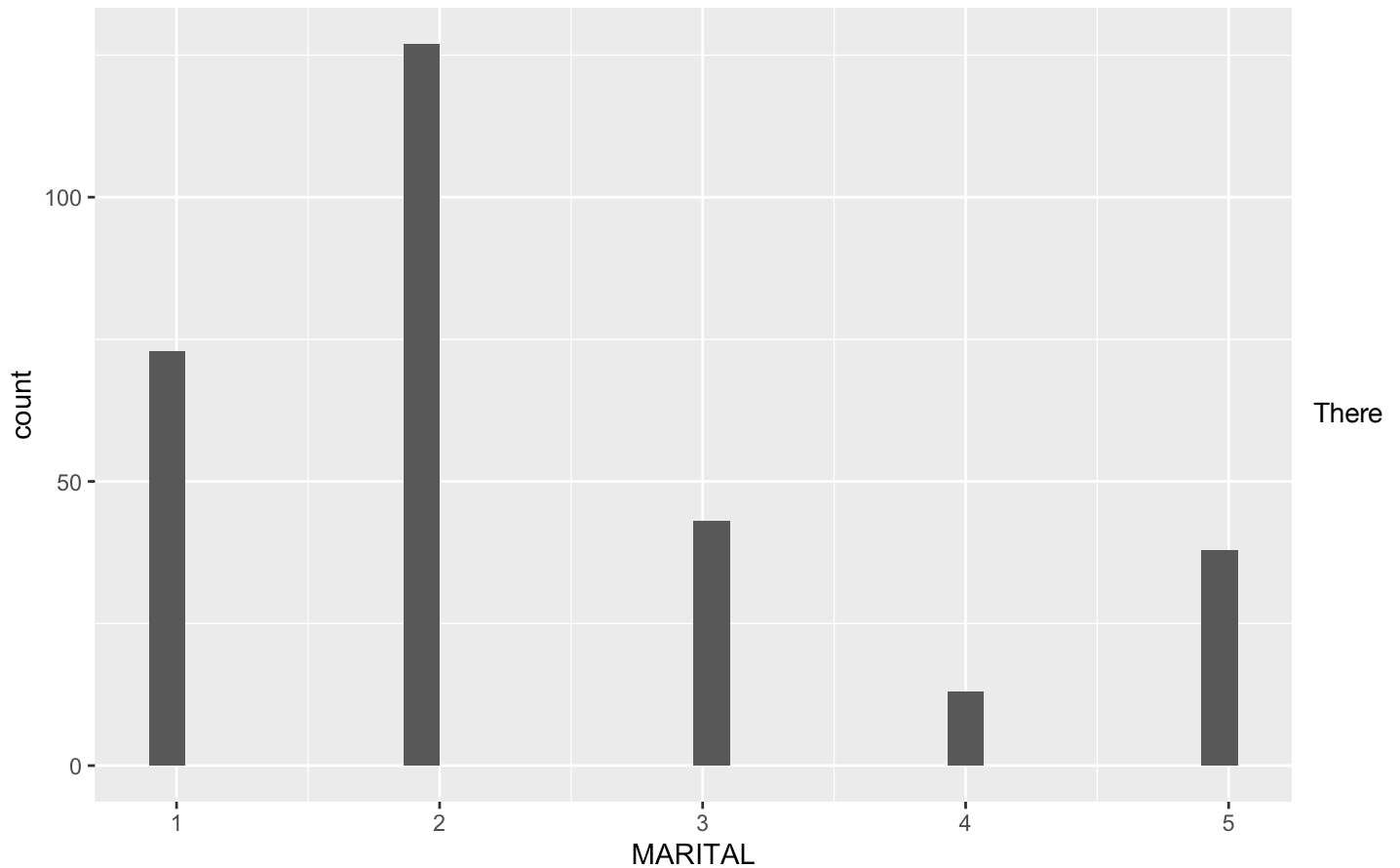


are are higher amounts of people experiencing depression through the mid to late 20s but there are a few spikes in between the ages 55- 73.

Marital 3.

```
ggplot(depress, aes(x= MARITAL)) + geom_histogram() + ggtitle("Depression in different marital status")
```

Depression in different marital status



are more people experiencing depression who are in a relationship but not married than those who are married.

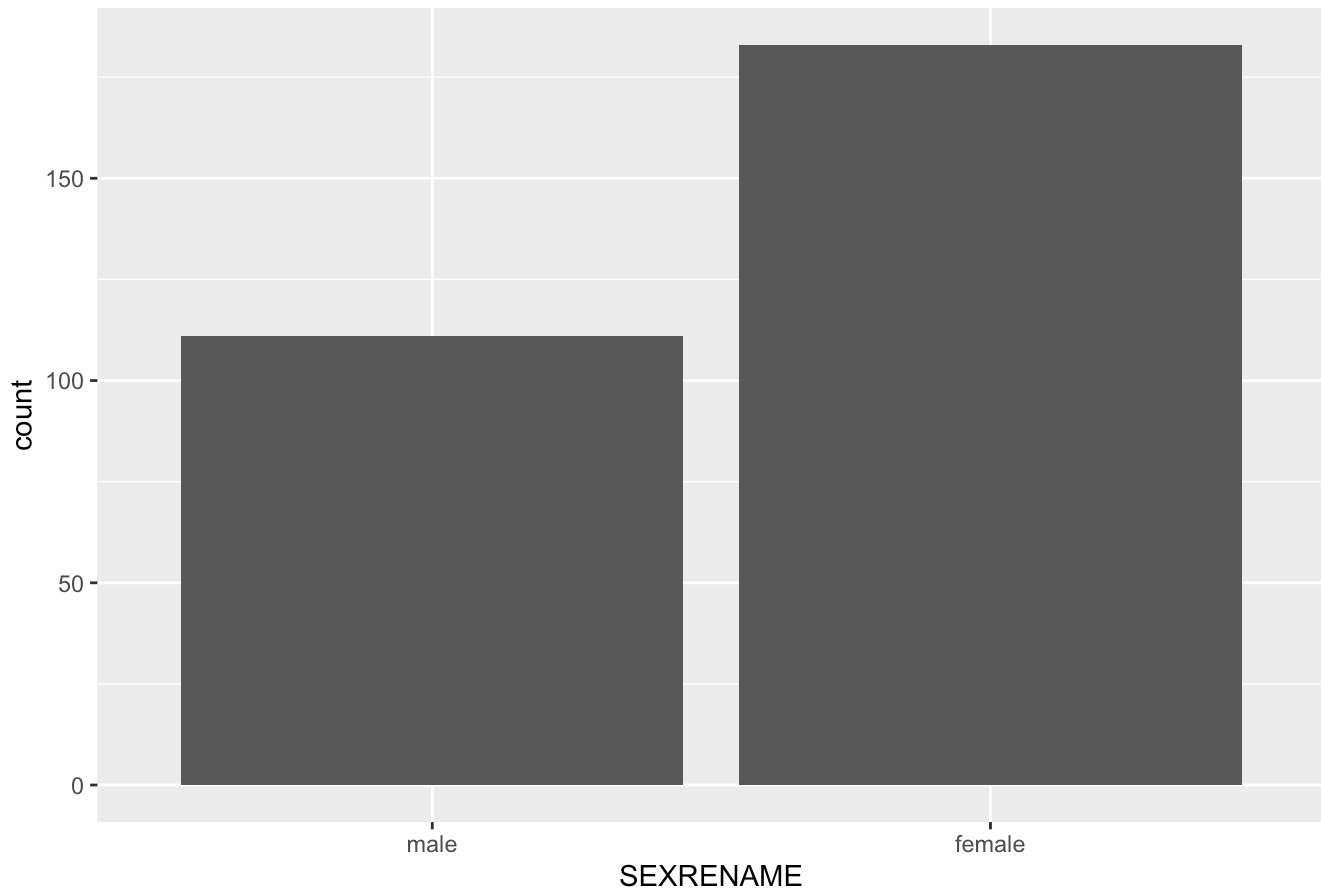
Sex 4.

```
depress$SEXRENAME <- factor(depress$SEX, labels = c("male", "female"))
summary(depress$SEX)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      1.000   1.000   2.000   1.622   2.000   2.000
```

5.

```
ggplot(depress, aes(SEXRENAME)) + geom_bar() + ggtitle("")
```



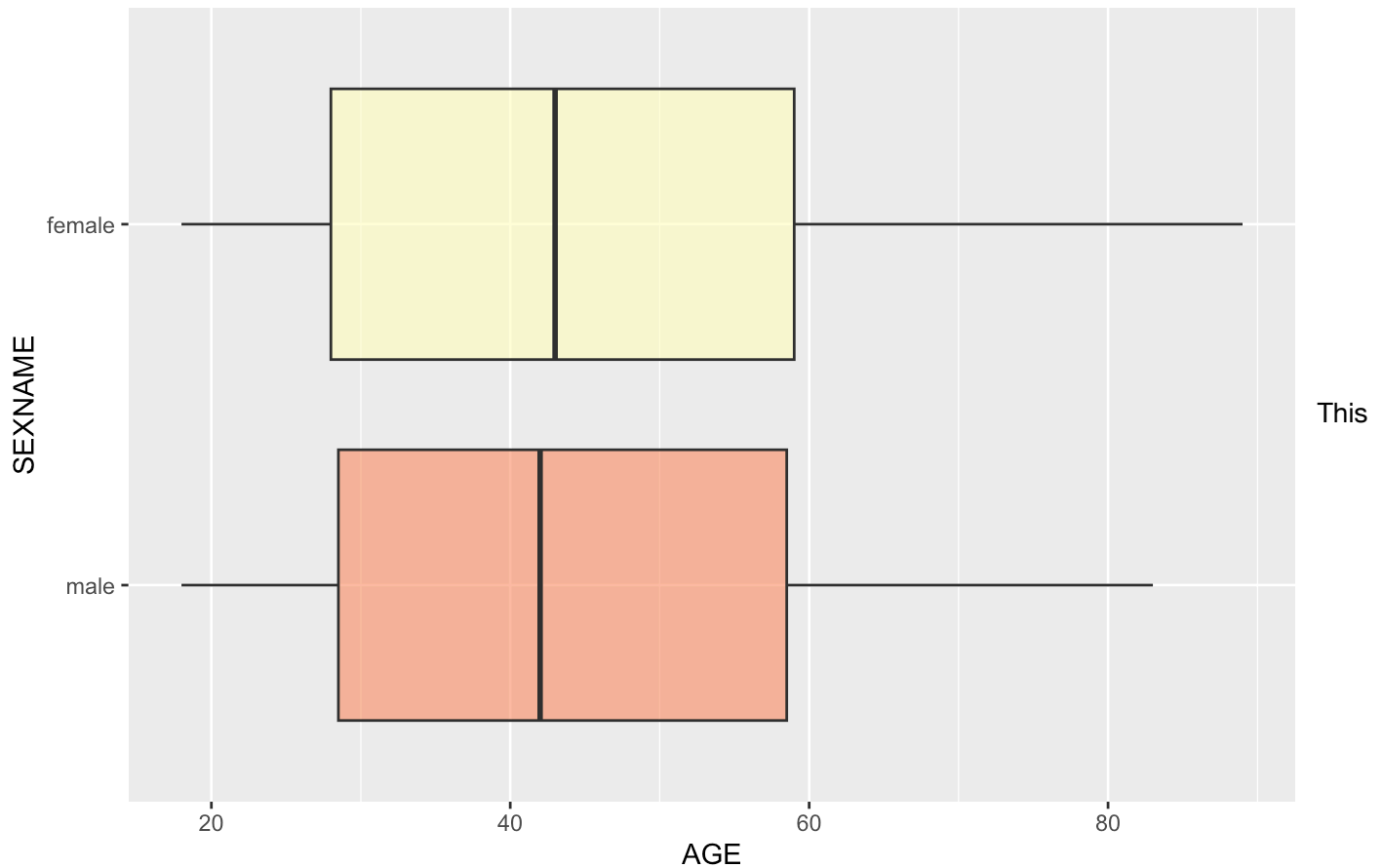
According to the plot females in this survey seem to suffer more from depression than males. There is around 100 more women in the survey saying they have depression than men.

Bivariate Comparison

1. Sex vs. Age

```
ggplot(depress, aes(x = SEXRENAME, y = AGE, fill = SEXRENAME, )) + geom_boxplot(alpha = 0.6) +  
  theme(legend.position = "none") + scale_fill_brewer(palette = "Spectral") + ylab("AGE") +  
  xlab("SEXNAME") + ggtitle("Sex vs. Age ") + coord_flip()
```

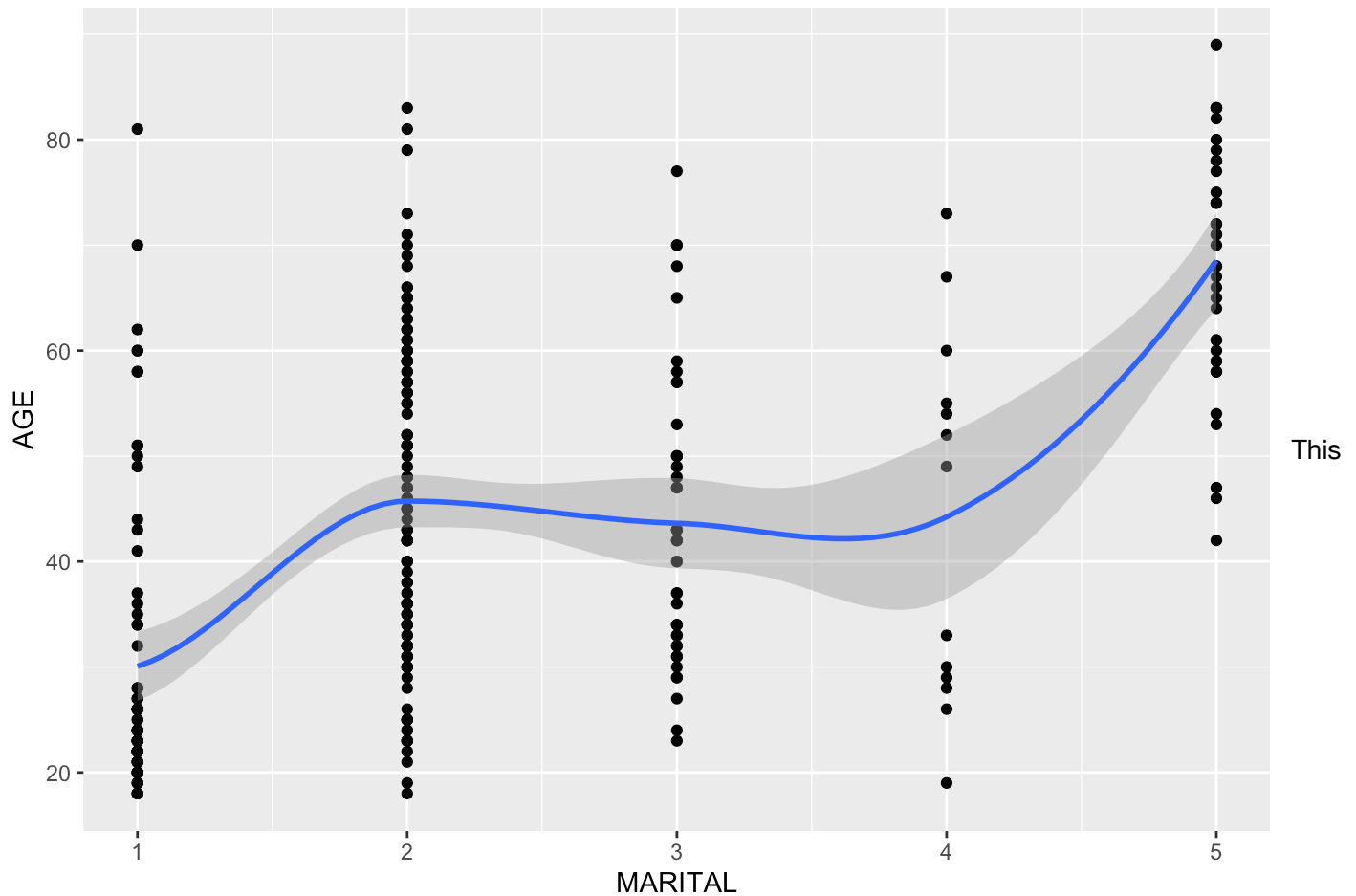
Sex vs. Age



box plot compares sex and age so with this box plot it shows that the average of females and males, from this survey, have depression between the ages 40-50. It also shows that there are some women that have depression well into their 80s and the men in this survey only go up to about 82.

2. Marital status vs. Age

```
ggplot(depress, aes(x = MARITAL , y = AGE)) + geom_point() + geom_smooth()
```



plot compares Age and Marital status showing that the highest points are at about 45 yrs and in the 2 marital status. The other high point is at 65-70 yrs and in the 5th marital status. The best fit line shows the difference between the two variables which I thought would be more drastic which is why I chose these two variables to compare but they do have some similarities.

Conclusion

In this project I learned that more women than men have depression but the ages that the average of both sexes dealing with depression is around the same age. I also learned that marital status and age are relatively similar in this survey since. This proves my hypothesis that women most often have more depression than men.