Final Project: Depression Data

Kaitlin Hensley

2/22/2022

#Introduction In this project, I will be exploring the "depression" data set. This data set comes from research taken from adults living in LA County. It includes 294 observations of 37 variables. I will be exploring the marital status, age, and education variables. My goal is to get a better understanding of what categories of people are most prevalent in this area; as well as to see if people of different ages are more likely to have a certain marital status and education level.

```
depress <- read.delim("/Users/kaitl/OneDrive/Desktop/Math130/Data/depress_081217.txt", header=TRUE,sep=
dim(depress)
```
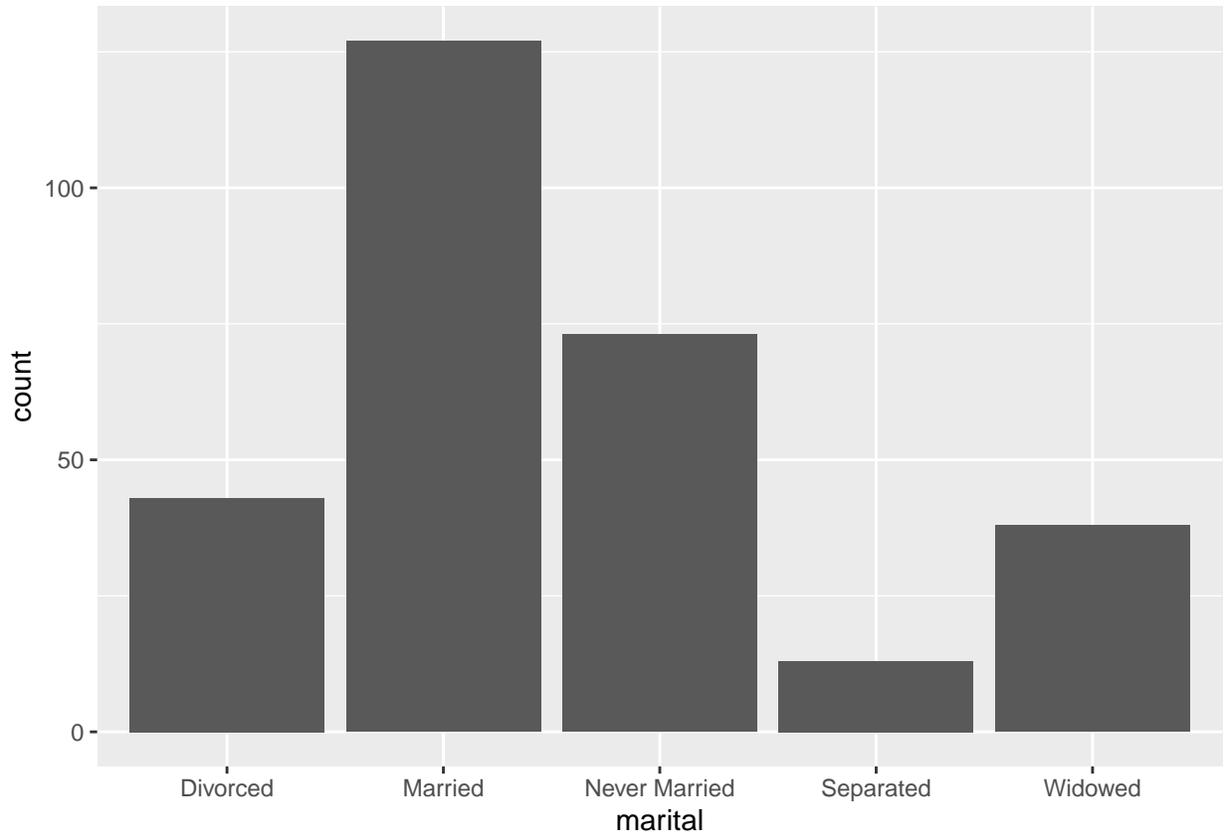
```
## [1] 294  37
```

```
library(ggplot2)
```

```
## Warning in register(): Can't find generic 'scale_type' in package ggplot2 to
## register S3 method.
```

##Univariate Exploration 1.Marital status

```
ggplot(depress, aes(x=marital)) + geom_bar()
```
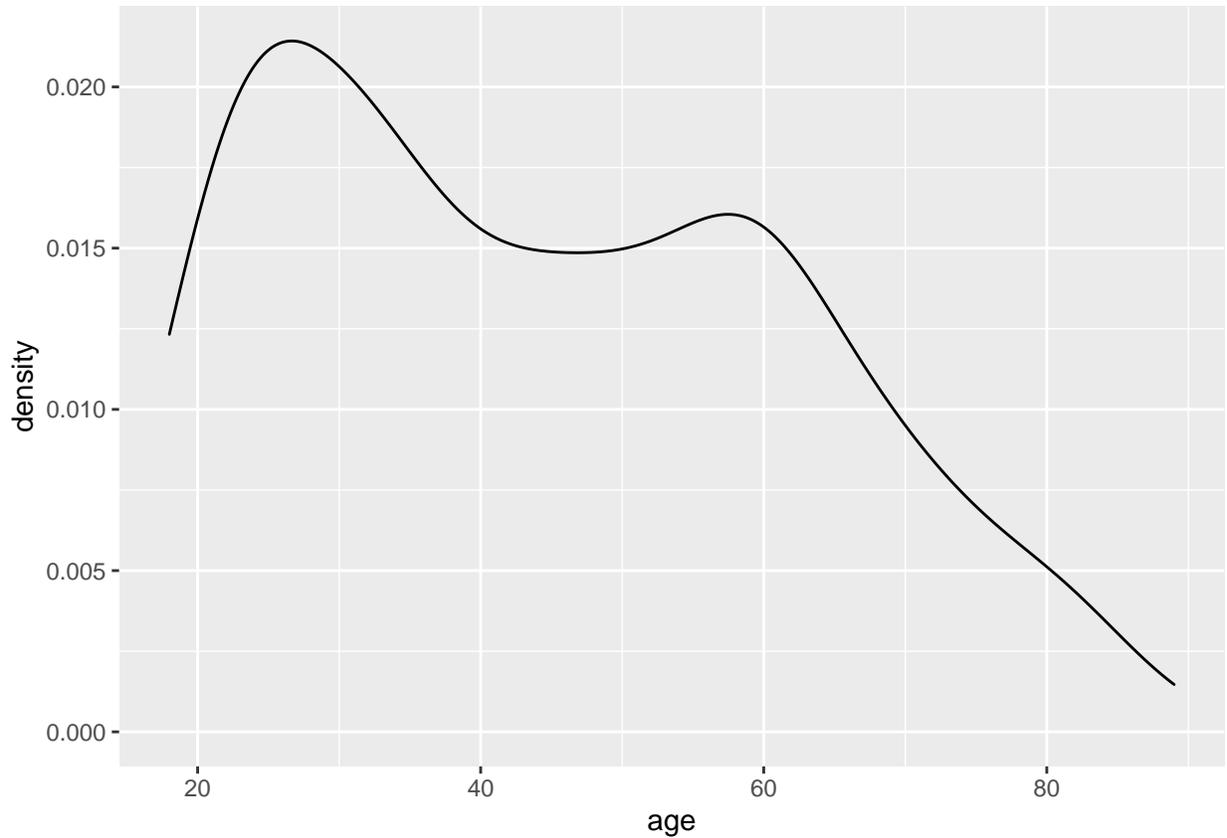
```
table(depress$marital)
```

```
##
##      Divorced     Married Never Married     Separated      Widowed
##            43         127            73            13           38
```

Here we examine the marital status of the participants. The majority of participants are married.

2.Age

```
ggplot(depress, aes(x=age)) + geom_density()
```

Here we take a look at the density plot of the distribution of ages in the participants. We can see that the majority of people that took part in this study were in their late 20s.
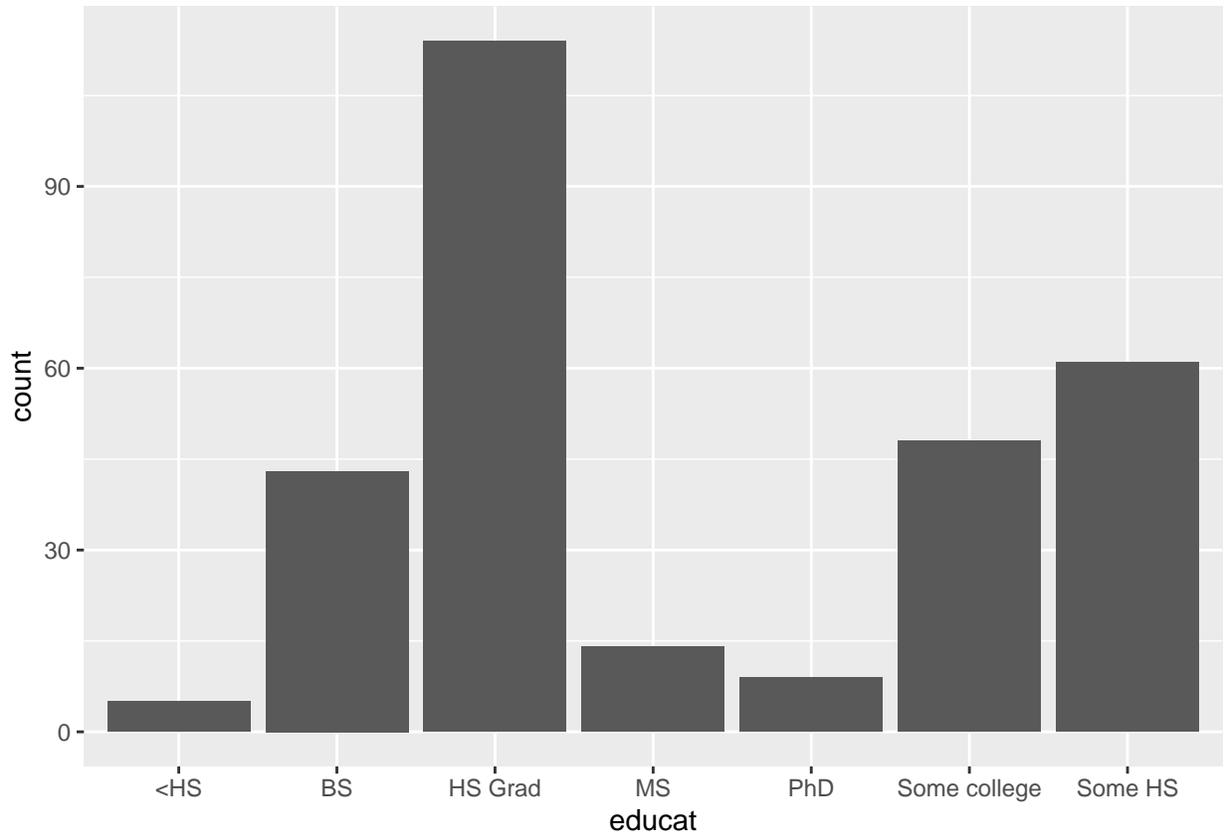
```
summary(depress$age,digits=3)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    18.0    28.0    42.5    44.4    59.0    89.0
```

Given these summary statistics we can conclude that 25% of participants are 28 or under and 75% of participants were 59 and under.

3.Education

```
ggplot(depress, aes(x=educat)) + geom_bar()
```

```
table(depress$educat)
```

```
##
##          <HS          BS      HS Grad           MS          PhD Some college
##            5          43          114           14            9           48
##      Some HS
##           61
```

This table summarizes the bar graph above. These show the exact number of participants in each category. We can see that the majority of participants are HS Graduates.

###Bivariate Exploration First we will explore age vs marital status,

```
ggplot(depress, aes(x=marital, y=age)) + geom_point() + ggtitle("Distribution of Age in different Marita
```

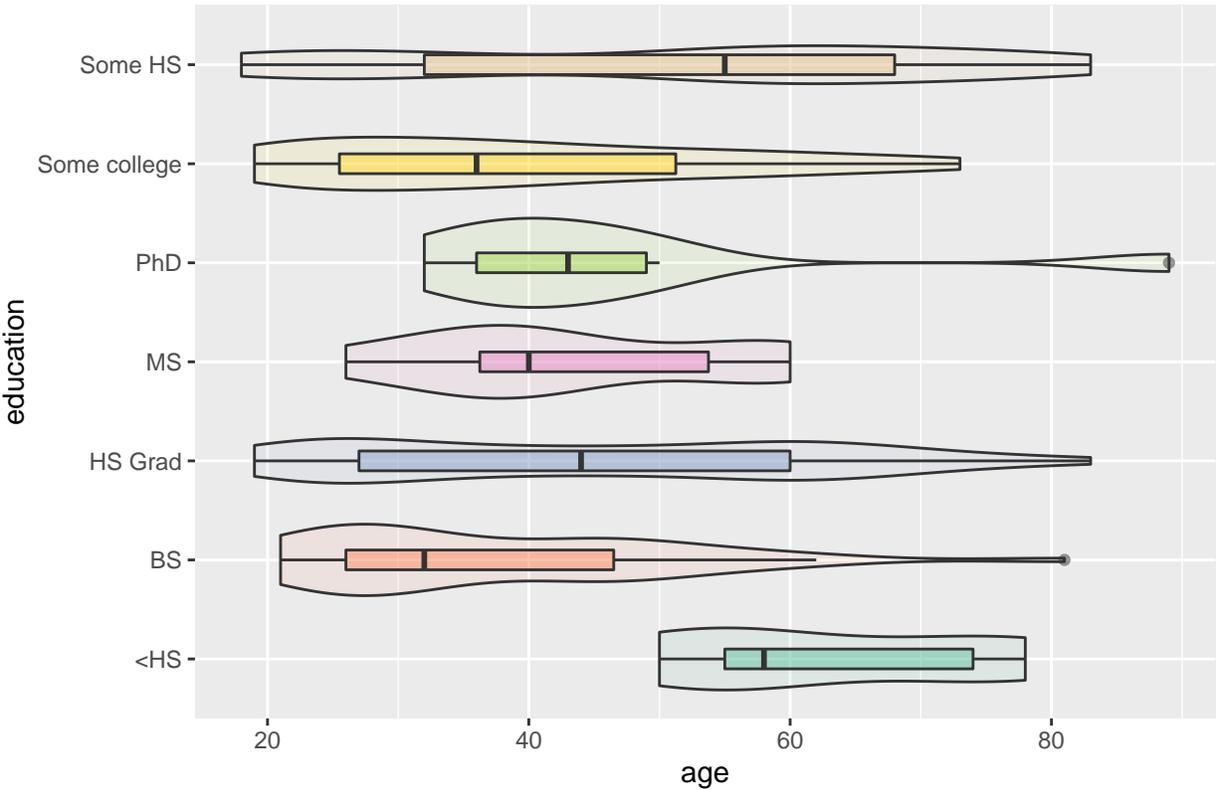## Distribution of Age in different Marital Statuses
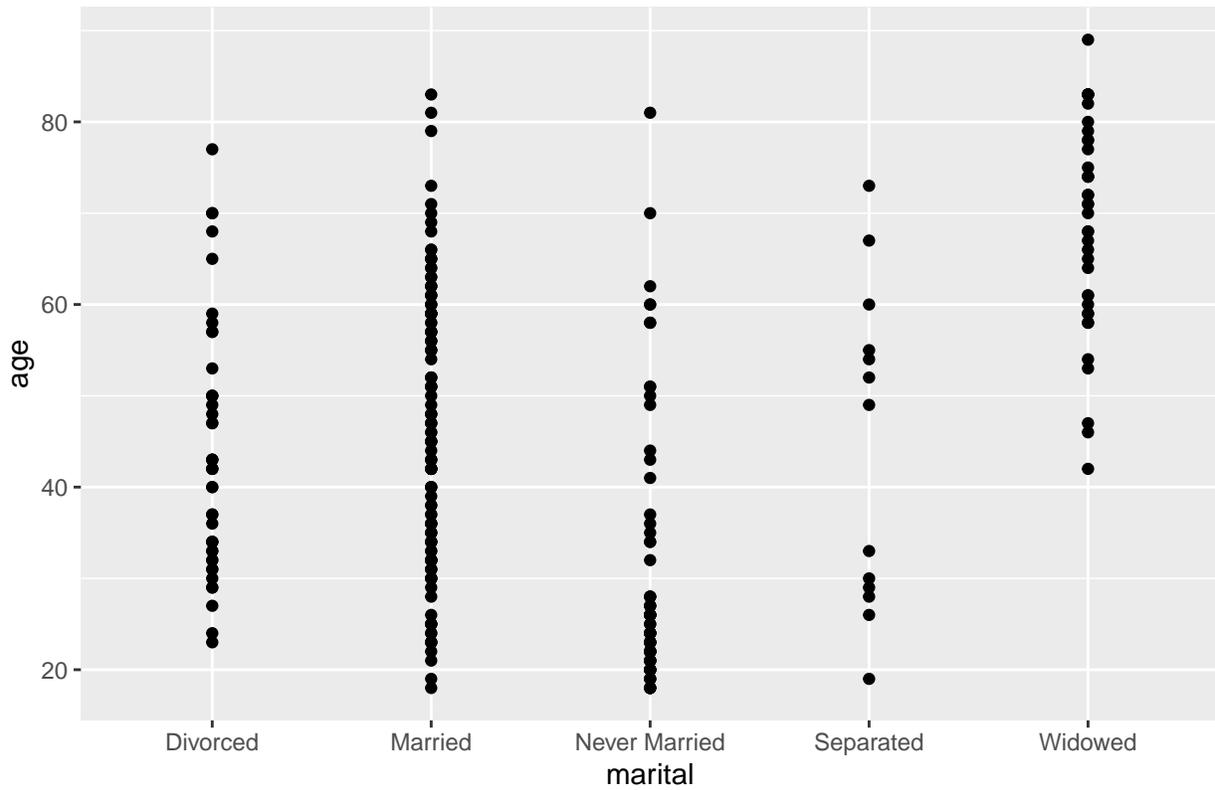
This graph shows that in each category the age range is quite wide.

Next, we will explore age vs. education level

```
ggplot(depress, aes(x=educat, y=age, fill=educat))+scale_fill_brewer( palette="Set2", guide='none')+ xla
geom_violin(alpha=.1) +
geom_boxplot(alpha=.5, width=.2) + coord_flip()
```

Distribution of Age in Education level

Here we can see that the majority of people with less than a highschool education are between 50 and 80 years old. The widest distribution of age is in people with some high school education.

###Conclusion In conclusion, we can see that there is a wide distribution of ages in each marital category, with the exception of the widowed category that mainly included people over the age of 40. The majority of people with less than a high school education are between the ages of 50 and 80. People in the categories "some HS", "Highschool graduate", and " BS" have a similarly very wide distribution. People with PhD's are mainly between the ages of 30 and 60 with a small outlier group going up to age 89. People with their MS degree are between ages 25 and 60.