

# Exploratory Data Analysis

Amanda Gutierrez

2022-09-19

```
depress <- read.delim("C:/Users/amgut/OneDrive/Desktop/MATH130/Data/depress_081217.txt", header=TRUE, sep=";", as.is=TRUE)
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
```

```
library(forcats)
library(knitr)
library(sjPlot)
```

```
## Learn more about sjPlot with 'browseVignettes("sjPlot")'.
```

## 1. Introduction

This data set allows me to analyze the first set of interviews of a prospective study of depression. Adult resident of Los Angeles County were interviewed. The data set consists of 294 observations and 37 variables. More information on the study can be found in Practical Multivariate Analysis, 5th edition by Afifi, May and Clark. Although there are a number of variables that can affect a person's mental health, I will be exploring the the relationship between age and depression. I expect to see a negative correlation between age and depression.

## 2. Univariate Descriptions

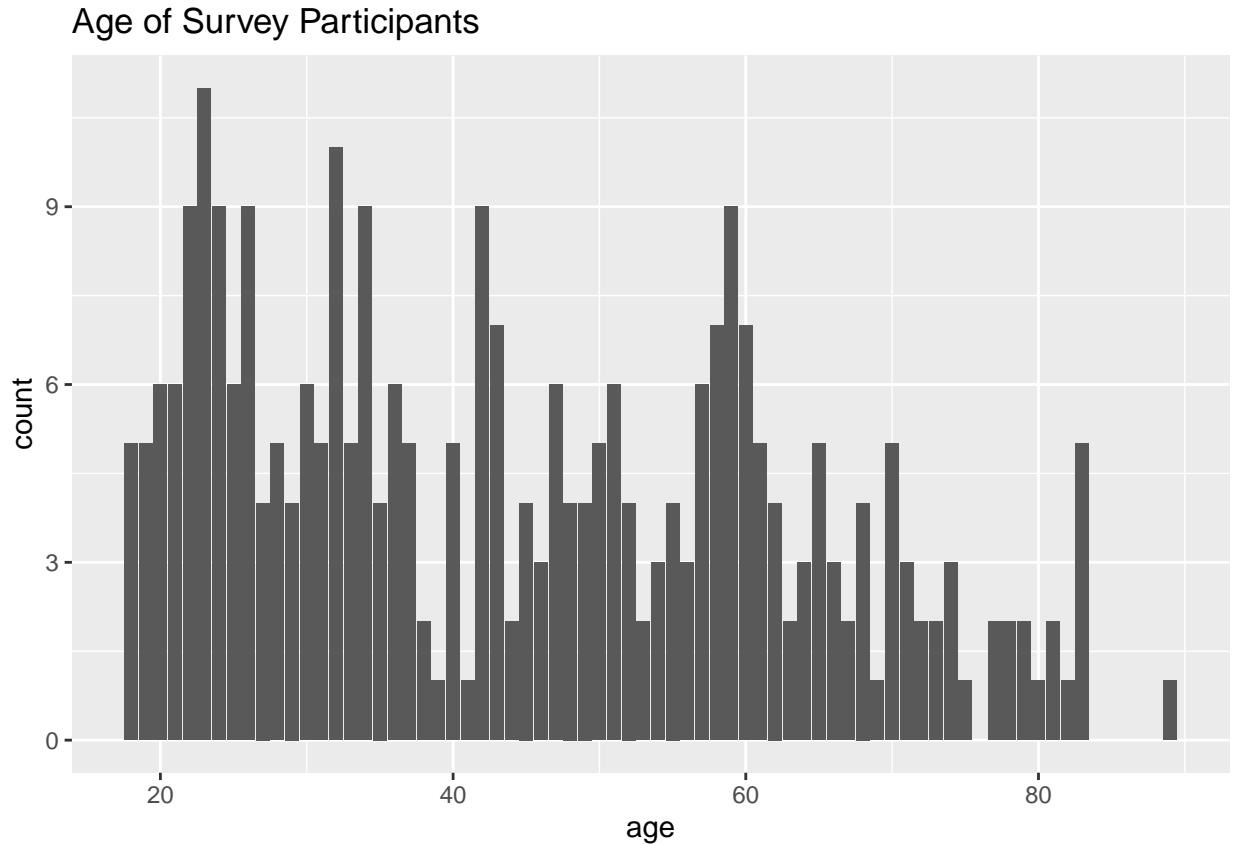
### Age:

This summary shows the distribution of average age among the participants in this study. The youngest participants were 18 years old, and the oldest participant was 89 years old. The average age of the participants in the study was around 44 years old.

```
summary(depress$age)
```

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

```
ggplot(depress, aes(x=age)) + geom_bar() + ggtitle("Age of Survey Participants")
```



The bar graph exhibits the general age distribution of the participants in the study. The bar graph shows that there are more participants below 45 years old. Surprisingly, the study still has a significant amount of participants above 60, and the study has a very broad range in age.

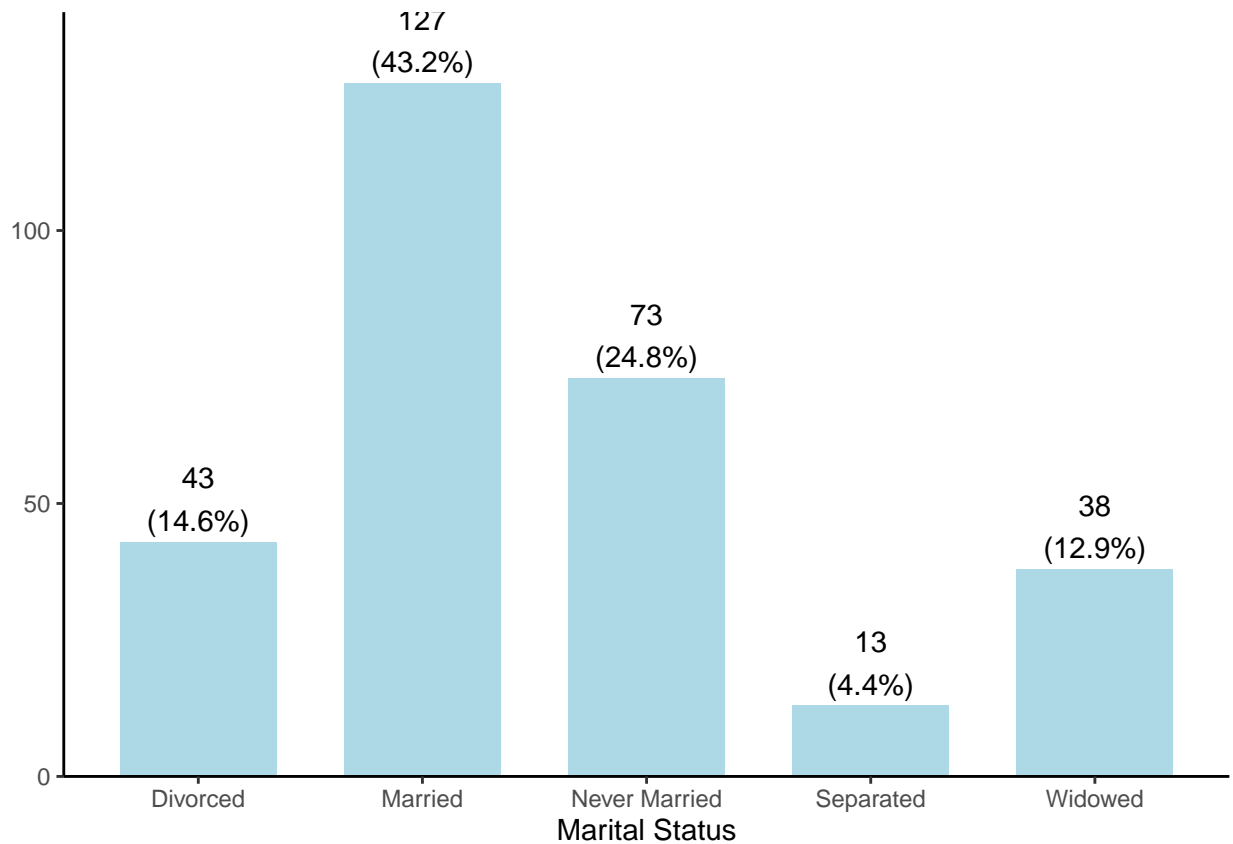
### Marital Status:

The table shows the number of participants and their marital status.

```
table(depress$marital)
```

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

```
depress$marital <- factor(depress$marital, levels=c("Divorced", "Married", "Never Married",
"Separated", "Widowed"))
plot_frq(depress$marital, geom.colors = "light blue") + xlab("Marital Status") + theme_classic()
```

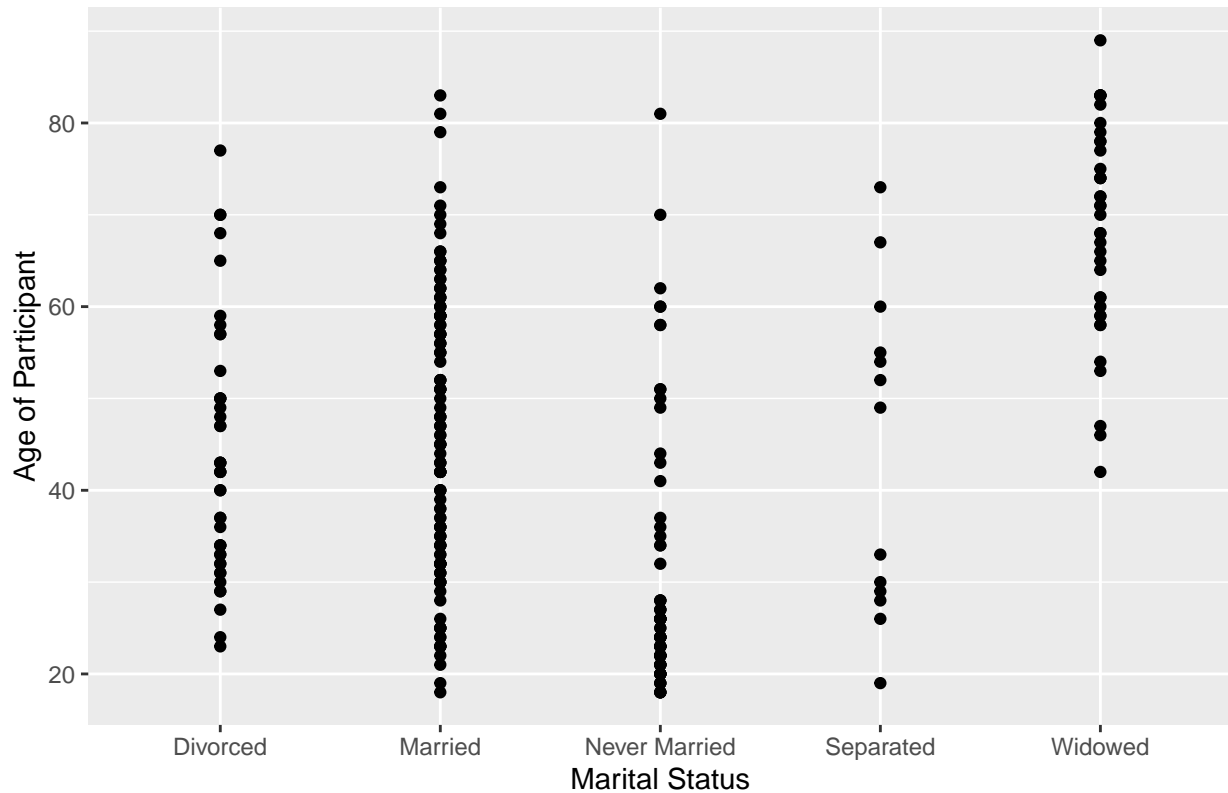


The total, 294, number of participants or observations is represented by this bar graph. It exhibits the marital status and the percent of participants within each category. 43.2% of the observations studied were married, and they made up most of the participants.

### 3. Bivariate Comparison

```
ggplot(depress, aes(x=marital, y=age)) + geom_point() + ggtitle("Marital Status and Age") + xlab("Marital Status")
```

## Marital Status and Age



A majority of the participants in this study are married within the 20 to 60 years old. The summary data set below, shows the age and marital status for every observation in the study.

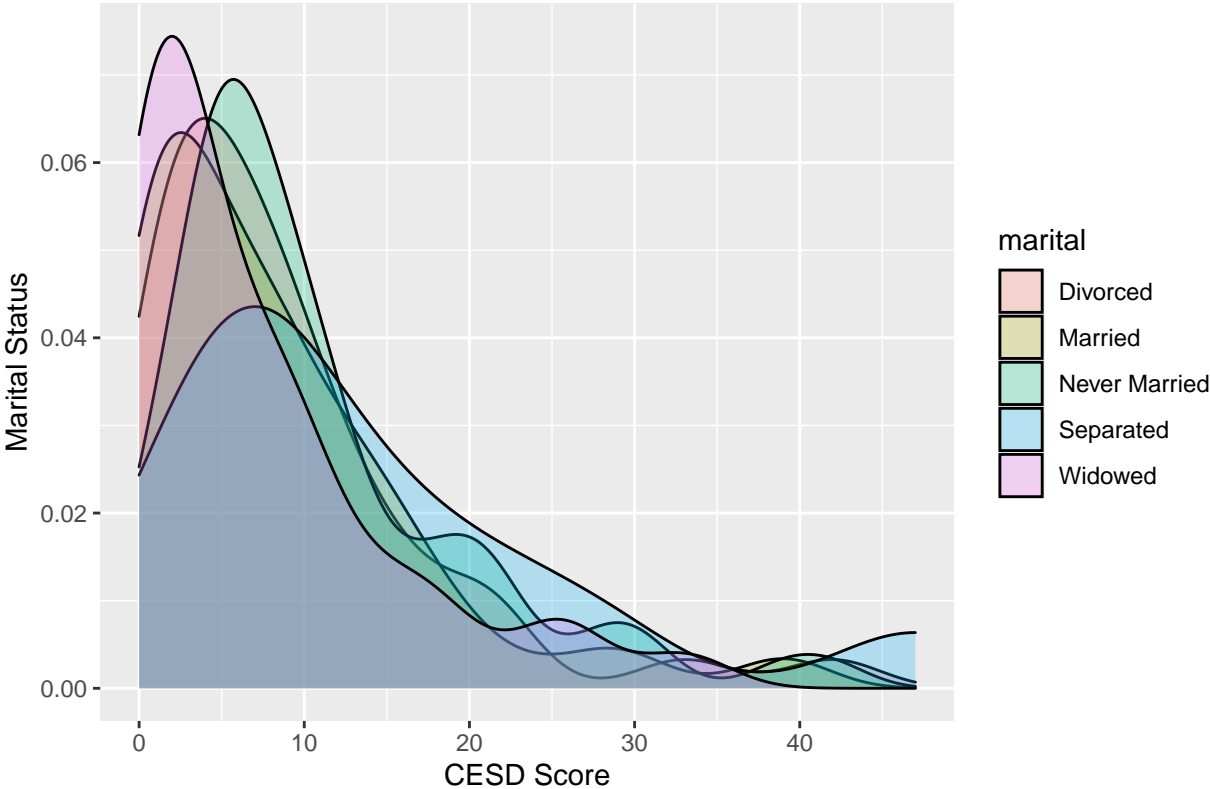
```
table( depress$age, depress$marital)
```

```
##
##      Divorced Married Never Married Separated Widowed
## 18          0      1           4          0          0
## 19          0      1           3          1          0
## 20          0      0           6          0          0
## 21          0      1           5          0          0
## 22          0      1           8          0          0
## 23          1      4           6          0          0
## 24          1      2           6          0          0
## 25          0      4           2          0          0
## 26          0      1           7          1          0
## 27          1      0           3          0          0
## 28          0      1           3          1          0
## 29          2      1           0          1          0
## 30          1      4           0          1          0
## 31          2      3           0          0          0
## 32          2      7           1          0          0
## 33          2      2           0          1          0
## 34          3      4           2          0          0
## 35          0      3           1          0          0
## 36          1      4           1          0          0
```

##	37	2	2	1	0	0
##	38	0	2	0	0	0
##	39	0	1	0	0	0
##	40	2	3	0	0	0
##	41	0	0	1	0	0
##	42	3	5	0	0	1
##	43	3	3	1	0	0
##	44	0	1	1	0	0
##	45	0	4	0	0	0
##	46	0	2	0	0	1
##	47	2	3	0	0	1
##	48	1	3	0	0	0
##	49	1	1	1	1	0
##	50	3	1	1	0	0
##	51	0	4	2	0	0
##	52	0	3	0	1	0
##	53	1	0	0	0	1
##	54	0	1	0	1	1
##	55	0	3	0	1	0
##	56	0	3	0	0	0
##	57	2	4	0	0	0
##	58	1	2	2	0	2
##	59	1	6	0	0	2
##	60	0	3	2	1	1
##	61	0	3	0	0	2
##	62	0	3	1	0	0
##	63	0	2	0	0	0
##	64	0	2	0	0	1
##	65	1	3	0	0	1
##	66	0	2	0	0	1
##	67	0	0	0	1	1
##	68	1	1	0	0	2
##	69	0	1	0	0	0
##	70	2	1	1	0	1
##	71	0	1	0	0	2
##	72	0	0	0	0	2
##	73	0	1	0	1	0
##	74	0	0	0	0	3
##	75	0	0	0	0	1
##	77	1	0	0	0	1
##	78	0	0	0	0	2
##	79	0	1	0	0	1
##	80	0	0	0	0	1
##	81	0	1	1	0	0
##	82	0	0	0	0	1
##	83	0	1	0	0	4
##	89	0	0	0	0	1

```
ggplot(depress, aes(x=cesd, fill=marital))+ geom_density(alpha= .25) + ggtitle("Marital Status and Depress")
```

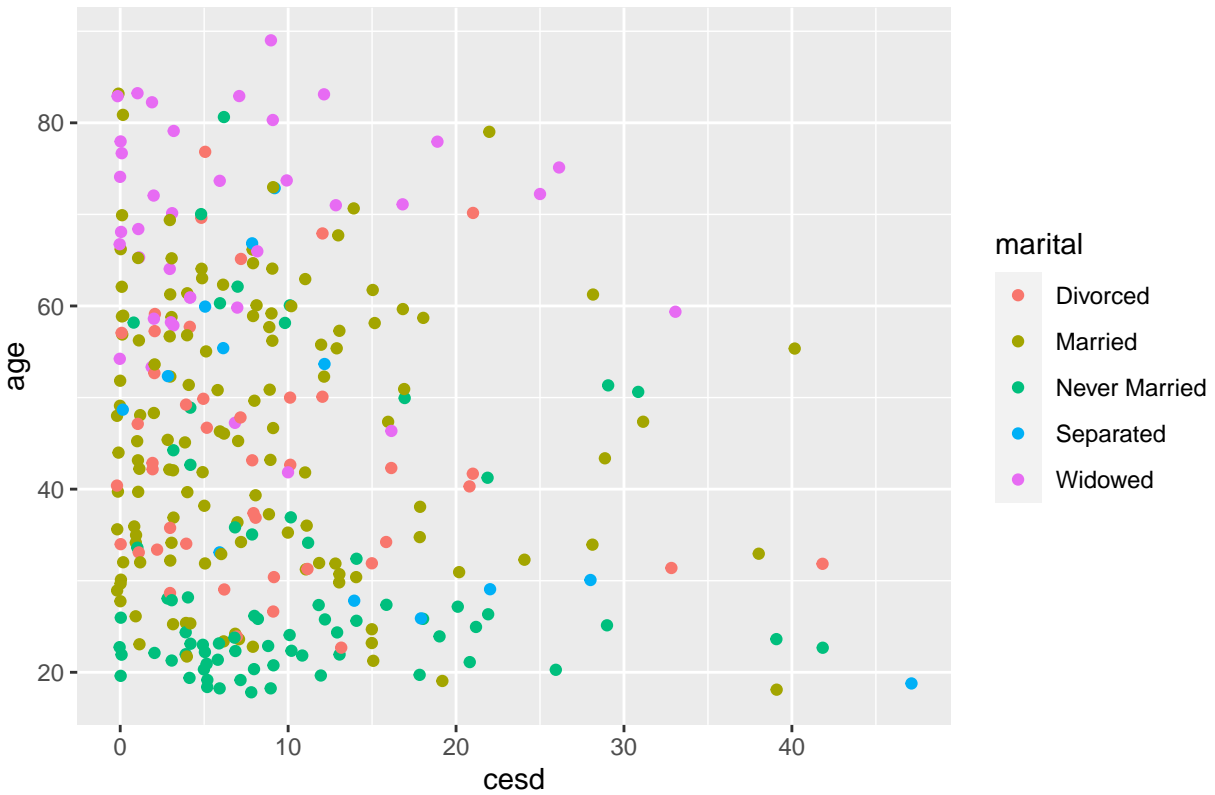
# Marital Status and Depression



The average age of the participants being observed in this study is 44 years old, and the graph above shows that participants who have separated with their significant other have higher CESD scores.

```
ggplot(depress, aes(x =cesd, y =age , color = marital)) +  
  geom_jitter(width = .2) + ggtitle("Age and Marital Status vs. CESD Scores")
```

## Age and Marital Status vs. CESD Scores



This scatter plot shows the general distribution of the age and marital status vs. their CESD score.

## 4. Conclusion

The average age of the participants being observed in this study was 44 years old and 43.2% of those participants were married. There was a larger concentration of participants with an age higher than 20 years old and being married. Married participants with an age higher than 20 years old have a higher correlation to have an increased CESD score.