Exploratory Data Analysis

Sajan Willbanks

2022-09-19

```
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(ggplot2)
library(RColorBrewer)
```

Data Set

Depression

```
depress <- read.delim("/Users/sajanwillbanks/Desktop/math130/data/depress_081217.txt", header = TRUE, se
dim(depress)
```

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

Introduction

In this project, I plan to explore the Depression data set, and find relationships between variables such as people's drinking habits and past marital status, and how they correlate to the illness.
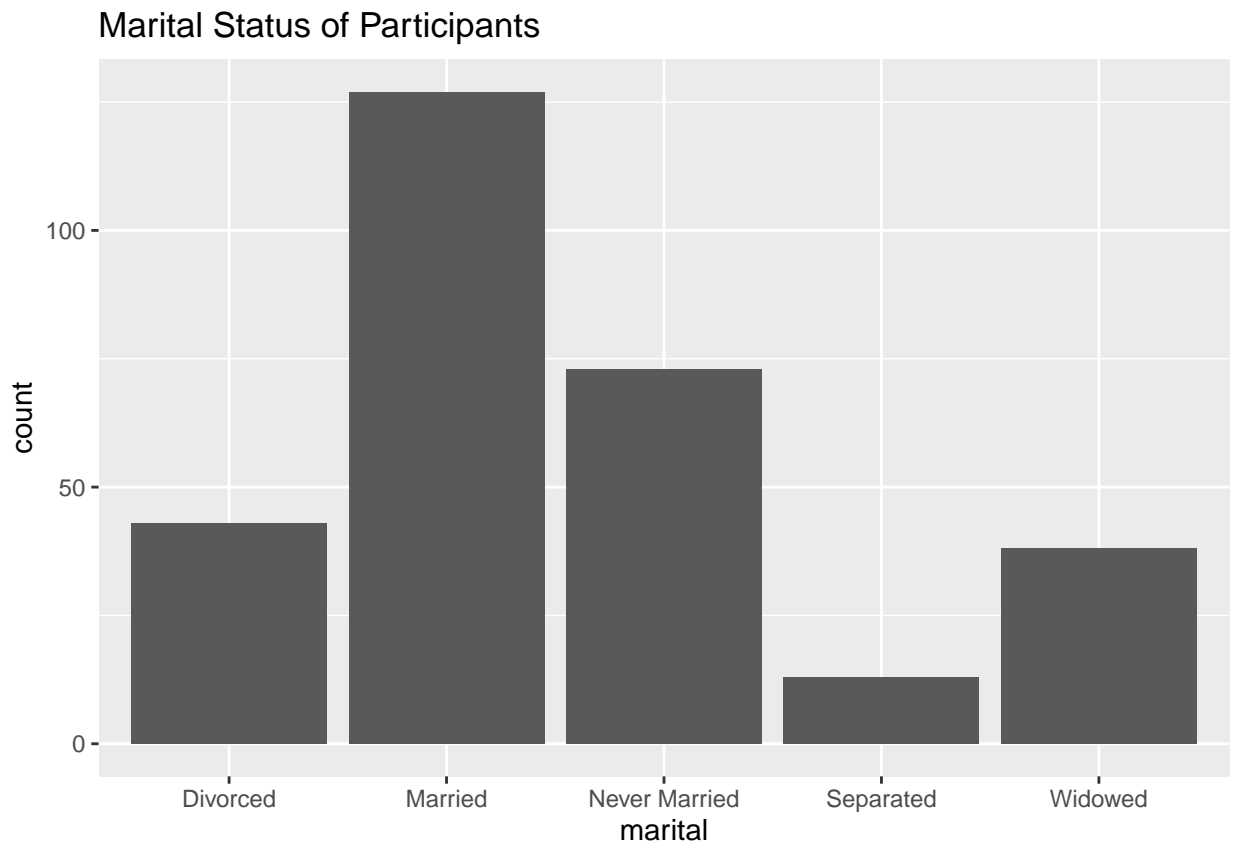
Univariate Analysis

Marital Status

The data below shows the marital status of each participant

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

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

```
ggplot(depress, aes(x= marital)) +
geom_bar() +
ggtitle("Marital Status of Participants")
```

Marital Status of Participants

Of the 294 participants, a majority of people reported being married. The next largest group reported never being married.
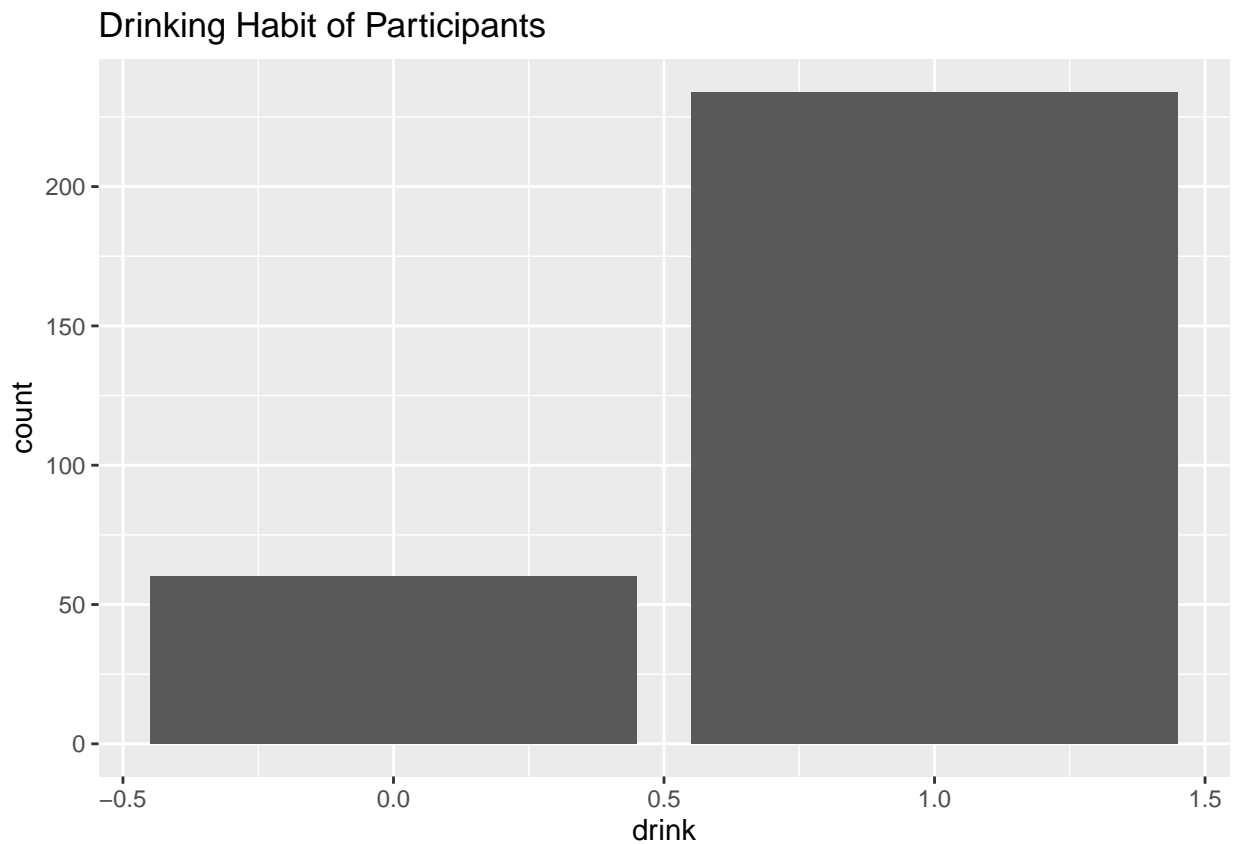
Drinking Habit

The table and graph below show weather or not the patient is a regular drinker.

```
table(depress$drink)
```

```
##
##    0   1
##   60 234
```

```
ggplot(depress, aes(x= drink)) +
geom_bar() +
ggtitle("Drinking Habit of Participants")
```

Drinking Habit of Participants

The overwhelming majority of participants reported having a regular drinking habit.

Bivariate Analysis

Drinking habit and Marital Status

The bar chart and two-way table below shows the relationship between drinking habits of the participants, and their marital status.

```
ggplot(depress, aes(x= drink, fill= marital)) + geom_bar() + ggtitle("Drinking Habit and Marital Status
```

Drinking Habit and Marital Status
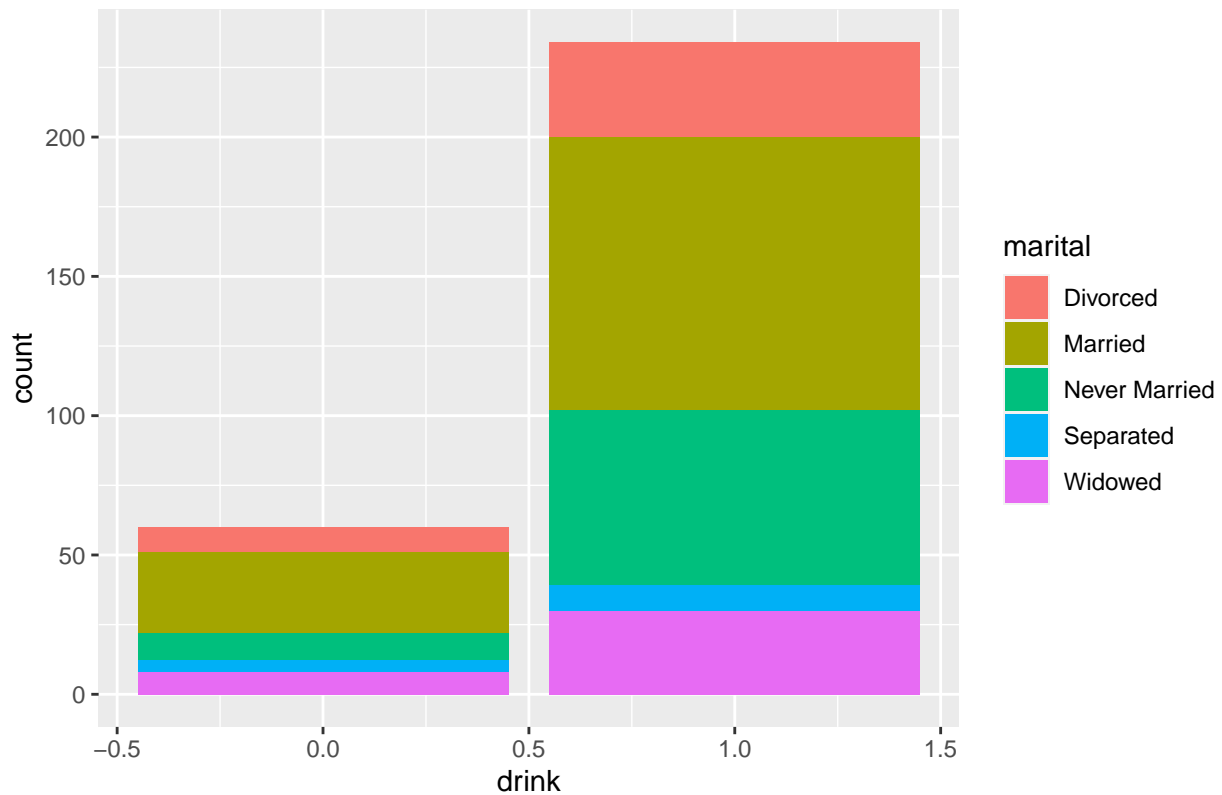

```
table(depress$marital, depress$drink) %>% prop.table(margin = 1) %>% round(3)
```

```
##
##                       0     1
##   Divorced        0.209 0.791
##   Married         0.228 0.772
##   Never Married   0.137 0.863
##   Separated       0.308 0.692
##   Widowed         0.211 0.789
```

These graphics show That of the participants that are regular drinkers, a majority of them are also married.

Marital Status and Depression

Using the cesd variable, where a cesd score greater than 16 means a participant feels depressed, This graph represents the relationship between marital status and how depressed a person feels.

```
ggplot(depress, aes(x=cesd, fill=marital)) + geom_density(alpha = .25) +
ggtitle("Marital Status and Depression") +
xlab("CESD Score") +
ylab("Marital Status")
```

Marital Status and Depression
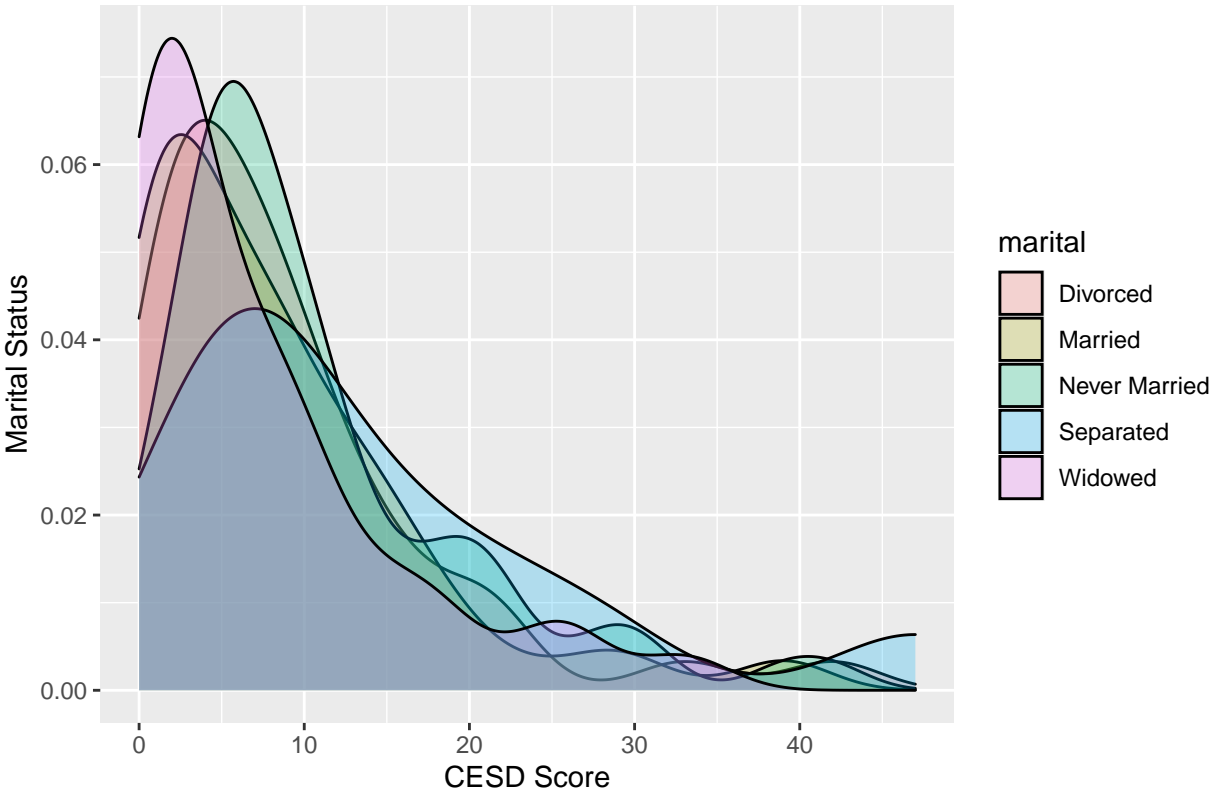
Conclusion

Going in to this project, I was interested in the relationship between their marital status and drinking habits, and how they relate to feelings of depression. In the end, I found out that a majority of the people who were married were also regular drinkers. I also found, as shown by the graph above, that a large number of people who reported feeling depressed are divorced.