Exploratory Data Analysis

Yvette Barragan

2/25/2022

Introduction

The data on depression is the data set I'll be investigating for this exploratory data project. This data is based on a study of Los Angeles County residents who are depressed. There are 294 observations with factors such as gender, age, marital status, education, employment status, income, religion, and so on. Age and marital status, though, are the two variables I'll be focused on. What I'd like to know is whether someone's marital status has anything to do with whether or not they are depressed.

```
depression <-read.table("/Users/ybarr/OneDrive/Documents/math130/data/depress_081217.txt",
header=TRUE, sep="\t")

library(ggplot2)
```

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

Univariate Exploration

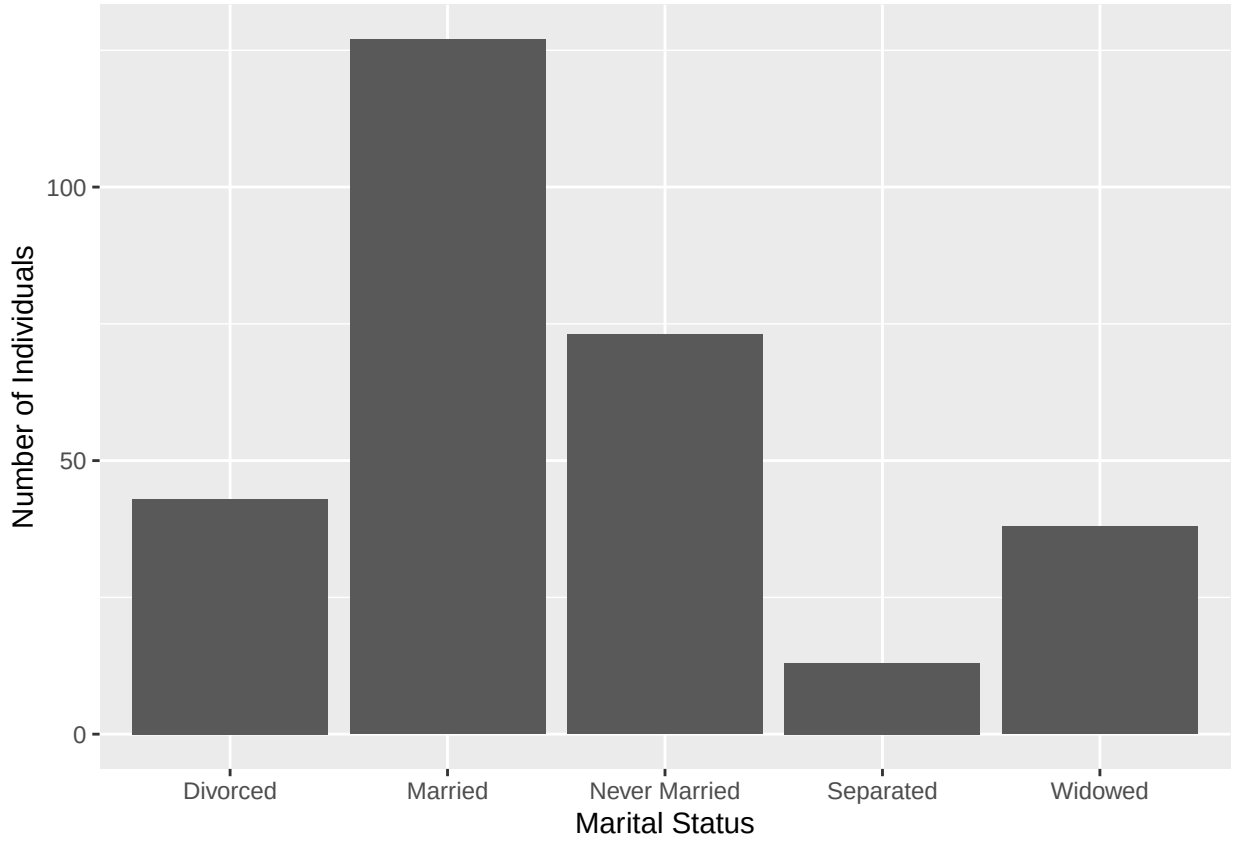
The first variable I'll look at is marital status, because I want to discover if partnerships have an impact on someone who is depressed. People who are in a relationship may appear to be less depressed because they have someone to enjoy life with, but there is also the possibility that it will have an adverse effect on their mental health because there are two people and their feelings to deal with instead of just one, which could take a toll on someone.

```
table(depression$marital)
```

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

127 of the people interviewed were married, whereas 73 had never been married. It's interesting to see that individuals who were divorced or widowed scored significantly lower than those who had never married, because I would have believed that those situations would be more difficult to deal with and eventually lead to depression.

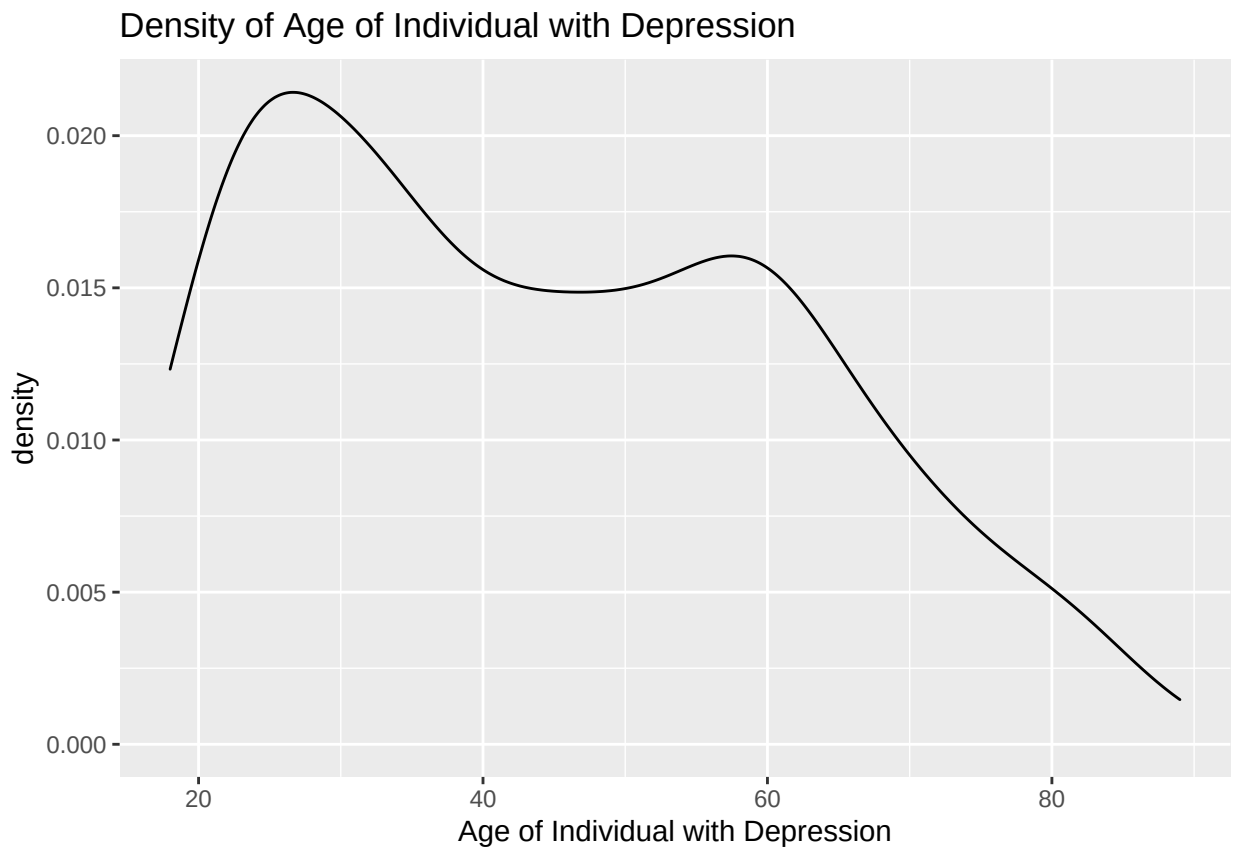
```
ggplot(depression, aes(x=marital)) + geom_bar() + ylab("Number of Individuals")+
  xlab("Marital Status")
```



This ggplot clearly illustrates that married people are more than twice as likely to be depressed as those who have never been married. While those were were separated were surprisinly the lowest group.

Age is the other variable I'll be focused on. I was curious about the average years of depression, because I'd suppose that those between the ages of 20 and 40 are more prone to be depressed because those are the years when life really gets moving.

```
ggplot(depression, aes(x=age)) + geom_density() +xlab("Age of Individual with Depression") +
  ggtitle("Density of Age of Individual with Depression")
```



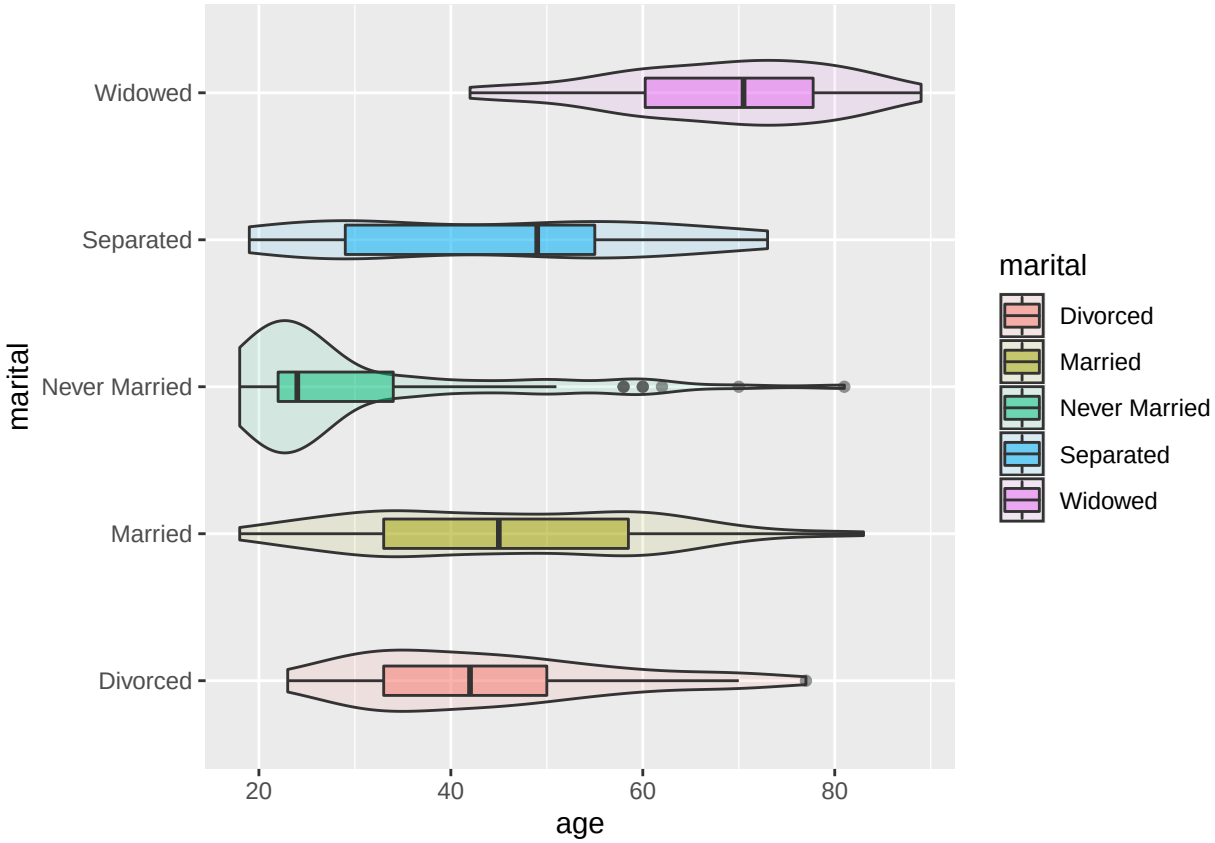

```
summary(depression$age)
```

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

This plot indicates that the youngest person with depression was 18 years old, while the oldest was 89 years old. However, the summary reveals that, while the maximum age for someone who is depressed is 89, the average age range of people are in their 40s. Which, to me, makes the most sense, because by that time, the kids have most likely left for college, there is likely to be some career activities going on, and perhaps even more things going on in the marriage.

Bivariate Exploration

```
ggplot(depression, aes(x=age, y=marital, fill=marital)) + geom_violin(alpha=.1) +
  geom_boxplot(alpha=.5, width=.2)
```



Both variables, age and marital status, are shown in this violet plot. The majority of individuals who were married were in their 30s to 40s, while those who had never married were in their 20s to 30s. This also demonstrates that the majority of individuals who were depressed were in their 30s to 40s, regardless of their marital status.

Conclusion

When we look at the connection between marital status and age, we can see that individuals who were married, divorced, or separated were in the same age range as the majority of those interviewed. Although a person's marital status isn't the cause of their sadness, it does indicate that their marital situation may contribute to depression.