

EDA_JHerbst

Jacob Herbst

9/21/2022

Introduction

```
depress <- read.table("/Users/jacobherbst/Desktop/math130/data/depress_081217.txt", header = TRUE, sep = ",")
```

```
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)
```

The data set that I am choosing to use for this project is the depression data set. The depression data set is from the first set of interviews of a prospective study of depression in the adult residents of Los Angeles County and includes 294 observations. I feel as though there are a lot of different variables that can contribute to a person dealing with depression, even if they may not be obvious. The variables I will be using are employment and marital status to try and find out if there is a trend within these variables, having a visible relationship with depression, then age and marital status. My research question and what I am interested in finding out is is there is a visible link between the variables of age and marital status when it comes to depression.

Univariate Exploration

```
summary(depress$maritalrename)
```

```
## Length Class Mode
```

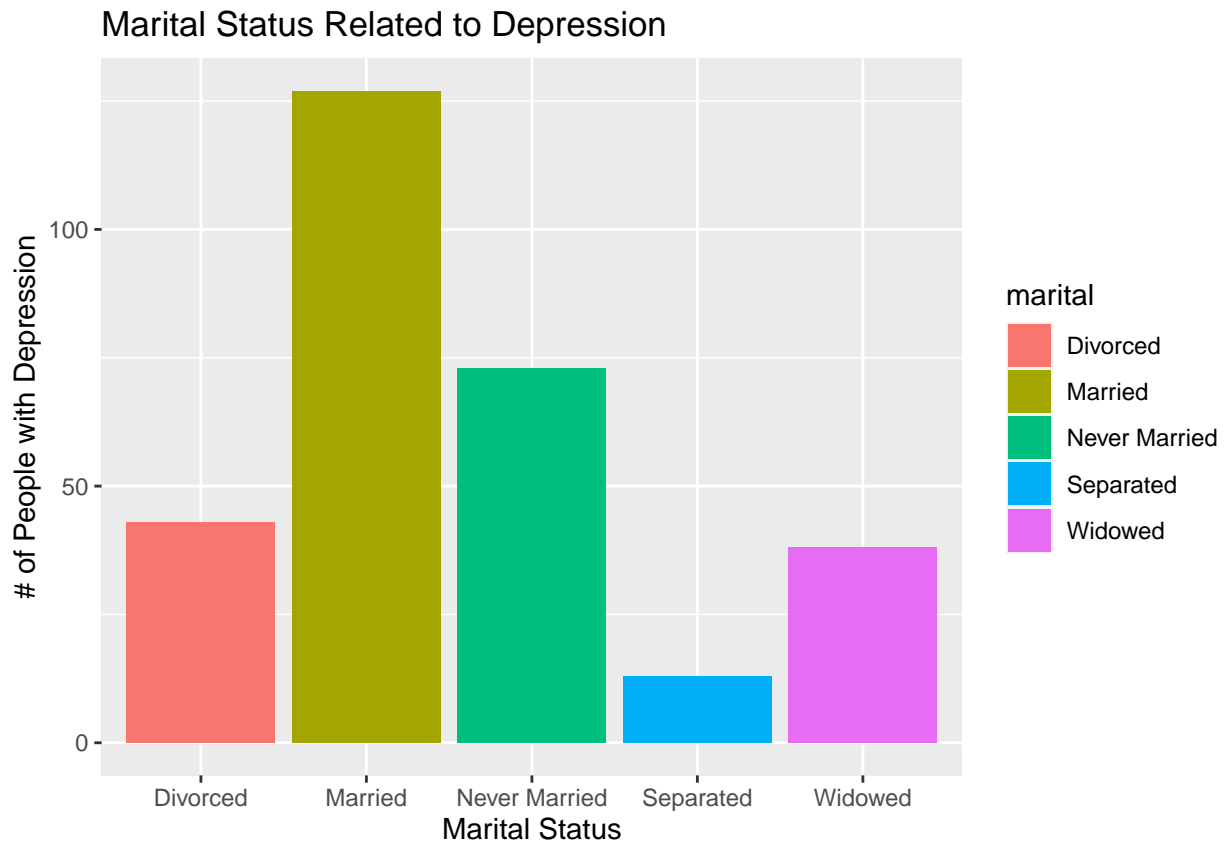
```
##      0  NULL  NULL
```

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

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

The results of this table are fairly shocking to me, as I would assume people that are divorced, or never married would be more depressed.

```
ggplot(depress, aes(x=marital, fill=marital))+geom_bar()+xlab("Marital Status") + ylab("# of People with Depression")  
ggtitle("Marital Status Related to Depression")
```



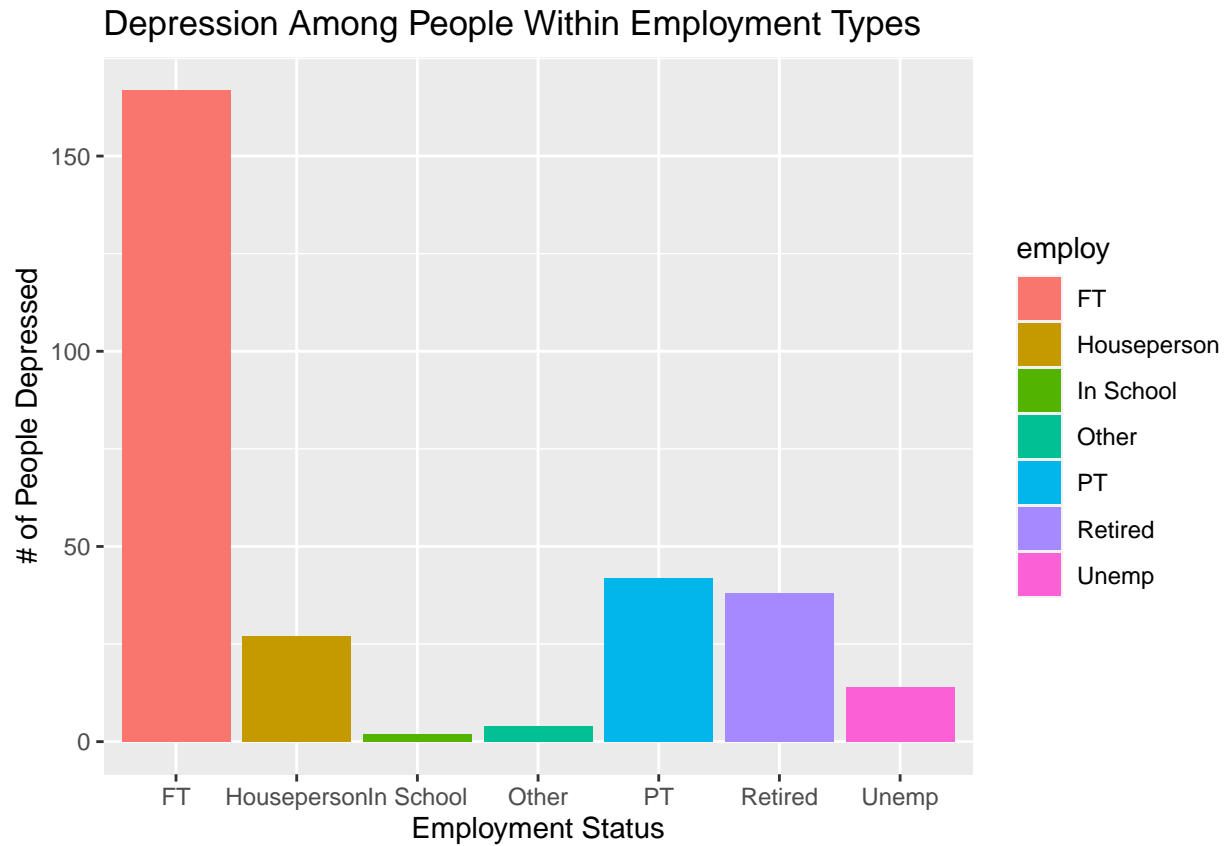
The bar chart displays a more clear view of people with depression, based on their marital status, showing the greatest number of people with depression based on marital status, are those that are married.

```
table(depress$employ)
```

```
##  
##      FT Houseperson  In School      Other      PT      Retired  
##          167          27           2           4          42          38  
##      Unemp  
##          14
```

Based on this chart, people without jobs seem to be more depressed than others.

```
ggplot(depress, aes(x=employ, fill=employ)) + geom_bar()+xlab("Employment Status") + ylab("# of People Depressed")
```



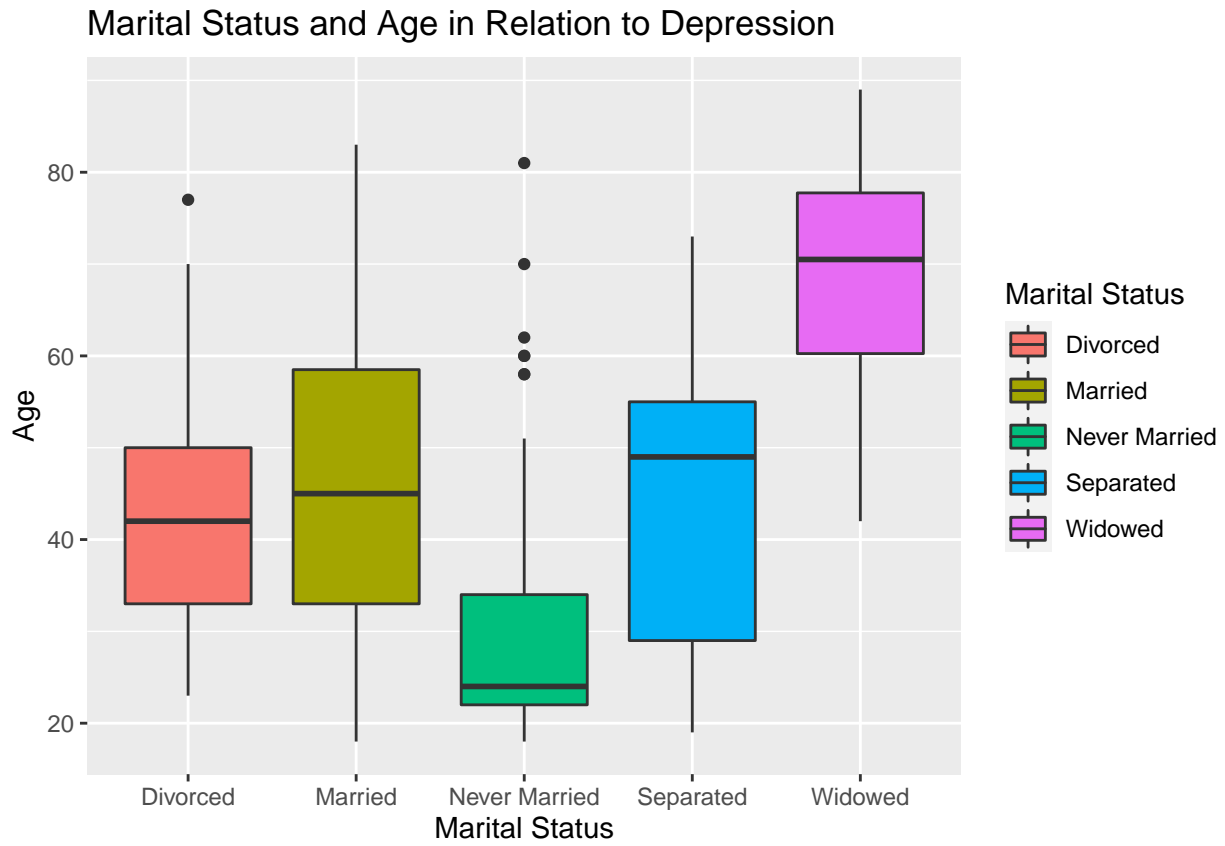
This bar chart gives a better view of the comparison between employment types in relation to depression.
 ## Bivariate Exploration

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

```
##
##           18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38
## Divorced    0 0 0 0 0 1 1 0 0 1 0 2 1 2 2 2 3 0 1 2 0
## Married     1 1 0 1 1 4 2 4 1 0 1 1 4 3 7 2 4 3 4 2 2
## Never Married 4 3 6 5 8 6 6 2 7 3 3 0 0 0 1 0 2 1 1 1 0
## Separated   0 1 0 0 0 0 0 0 1 0 1 1 1 0 0 1 0 0 0 0 0
## Widowed     0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
##
##           39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59
## Divorced    0 2 0 3 3 0 0 0 2 1 1 3 0 0 1 0 0 0 2 1 1
## Married     1 3 0 5 3 1 4 2 3 3 1 1 4 3 0 1 3 3 4 2 6
## Never Married 0 0 1 0 1 1 0 0 0 0 1 1 2 0 0 0 0 0 0 2 0
## Separated   0 0 0 0 0 0 0 0 0 0 1 0 0 1 0 1 1 0 0 0 0
## Widowed     0 0 0 1 0 0 0 1 1 0 0 0 0 0 1 1 0 0 0 2 2
##
##           60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 77 78 79 80 81
## Divorced    0 0 0 0 0 1 0 0 1 0 2 0 0 0 0 0 1 0 0 0 0
## Married     3 3 3 2 2 3 2 0 1 1 1 1 0 1 0 0 0 0 1 0 1
## Never Married 2 0 1 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1
```

```
## Separated      1 0 0 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0 0
## Widowed        1 2 0 0 1 1 1 1 2 0 1 2 2 0 3 1 1 2 1 1 0
##
##              82 83 89
## Divorced      0 0 0
## Married       0 1 0
## Never Married 0 0 0
## Separated     0 0 0
## Widowed       1 4 1
```

```
ggplot(depress, aes(x=marital, y=age, fill=marital)) + geom_boxplot() + scale_fill_discrete(name="Marital Status")
```



Conclusion Based off of the box plot and table, there is some, relation to depression in terms of age, and marital status. There are some outliers in the never married and divorced, but other than that it is very sound.