EDA Project

Ashley Meade

2/21/2022

Introduction to Depression Data Set

I will be analyzing the depression data set. This data set has 294 observations and 37 variables. This data set is from interviews of adults in Los Angeles County with depression. The two variables I will be using are marital status and income. Is there a connection between marital status and income?

```
depress <- read.delim("depress_081217.txt", header = TRUE, sep = "\t")
```

```
head(depress)
```

```
##   id sex age   marital       educat  employ income relig c1 c2 c3 c4 c5 c6 c7
## 1  1   1  68   Widowed      Some HS Retired      4     1  0  0  0  0  0  0  0
## 2  2   0  58  Divorced Some college      FT     15     1  0  0  1  0  0  0  0
## 3  3   1  45   Married      HS Grad      FT     28     1  0  0  0  0  1  0  0
## 4  4   1  50  Divorced      HS Grad   Unemp      9     1  0  0  0  0  1  1  0
## 5  5   1  33 Separated      HS Grad      FT     35     1  0  0  0  0  0  0  0
## 6  6   0  24   Married      HS Grad      FT     11     1  0  0  0  0  0  0  0
##   c8 c9 c10 c11 c12 c13 c14 c15 c16 c17 c18 c19 c20 cesd cases drink health
## 1  0  0   0   0   0   0   0   0   0   0   0   0   0    0     0     0      2
## 2  0  0   0   0   1   0   0   1   0   1   0   0   0    4     0     1      1
## 3  0  0   0   0   0   0   1   1   1   0   0   0   0    4     0     1      2
## 4  3  0   0   0   0   0   0   0   0   0   0   0   0    5     0     0      1
## 5  3  3   0   0   0   0   0   0   0   0   0   0   0    6     0     1      1
## 6  0  1   0   0   1   2   0   0   2   1   0   0   0    7     0     1      1
##   regdoc treat beddays acuteill chronill
## 1      1     1       0        0        1
## 2      1     1       0        0        1
## 3      1     1       0        0        0
## 4      1     0       0        0        1
## 5      1     1       1        1        0
## 6      1     1       0        1        1
```

1. Univariate Variables

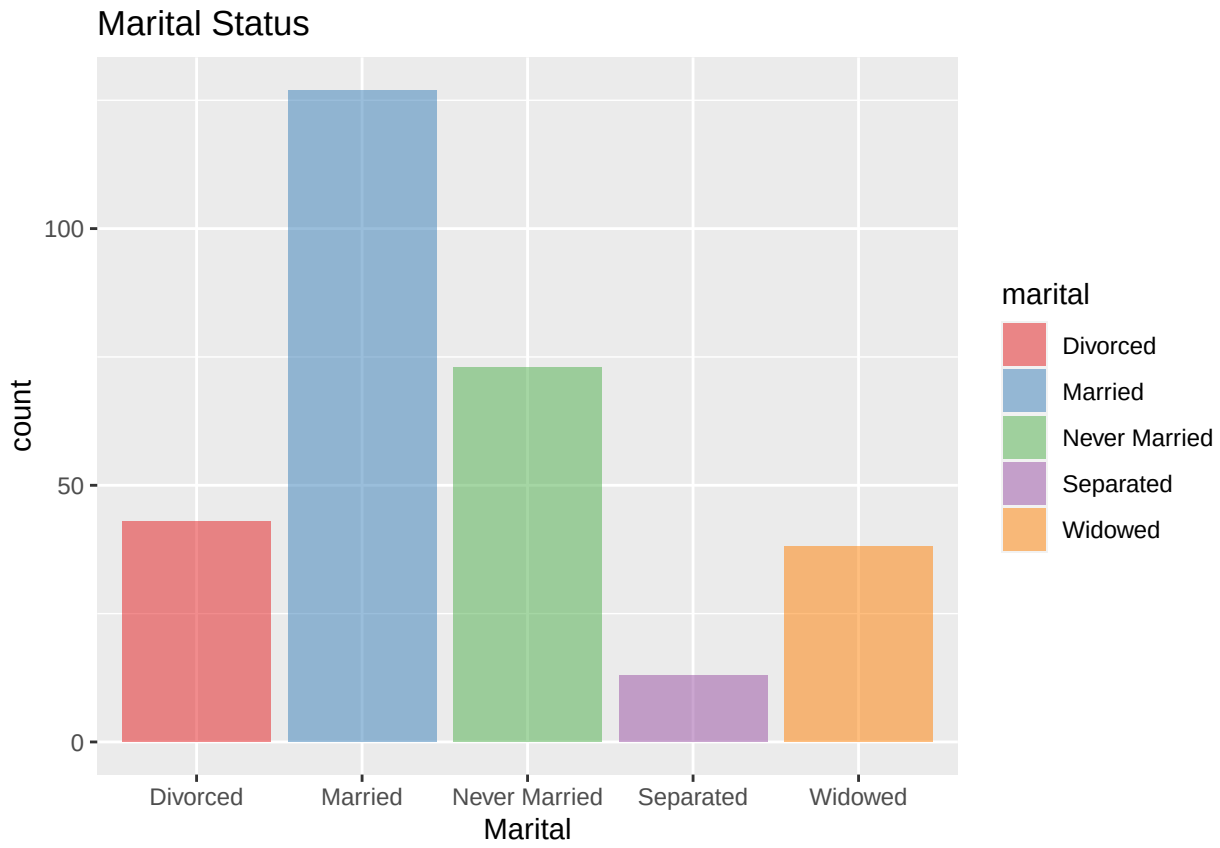
The first variable is the marital status. This table shows the marital status of the respondent.

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

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

This bar chart is expressing the different marital status of the respondents.

```
ggplot(depress, aes(x=marital, fill=marital)) + geom_bar(alpha=.5) + xlab("Marital") + scale_fill_brewe
```



The second variable is the respondents income.

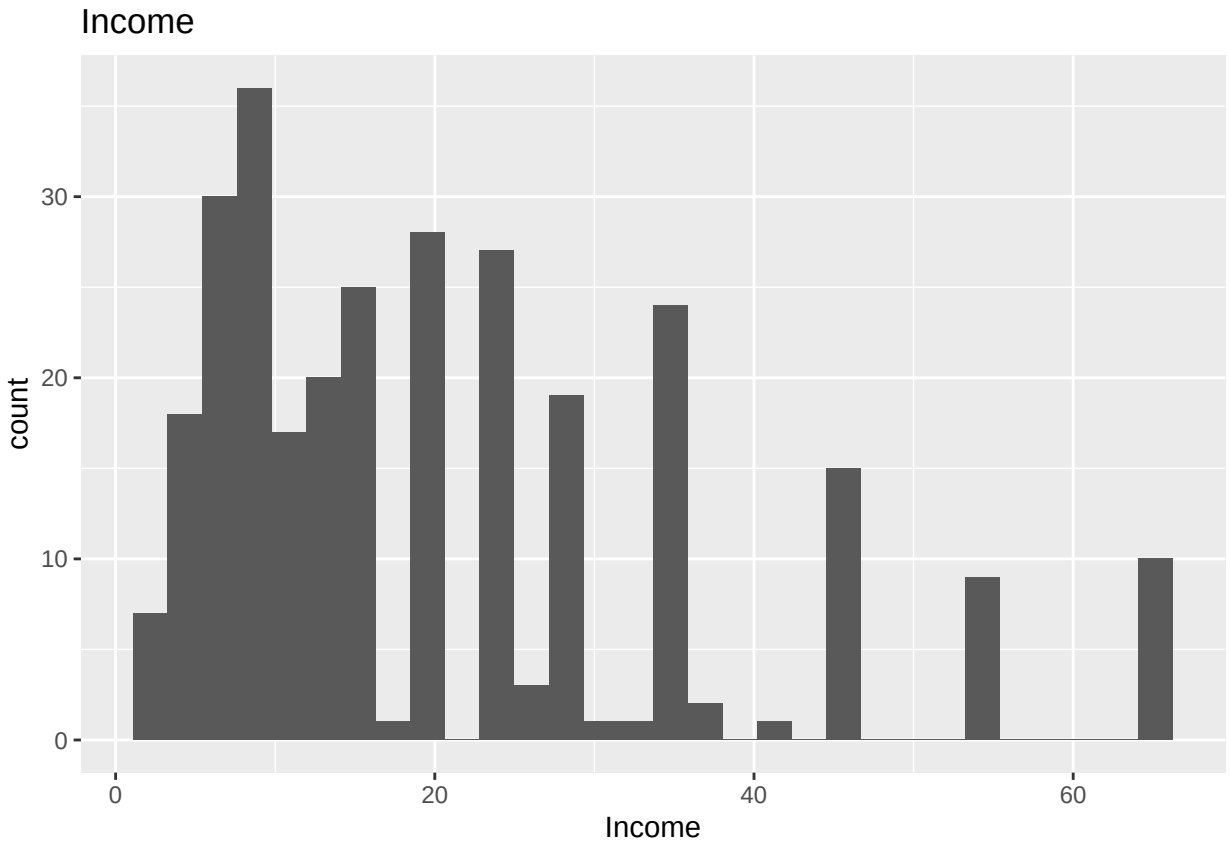
```
summary(depress$income)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    2.00    9.00   15.00   20.57   28.00   65.00
```

This histogram shows the variety of income the respondents receive.

```
ggplot(depress, aes(x=income)) + geom_histogram() + xlab("Income") + ggtitle("Income")
```

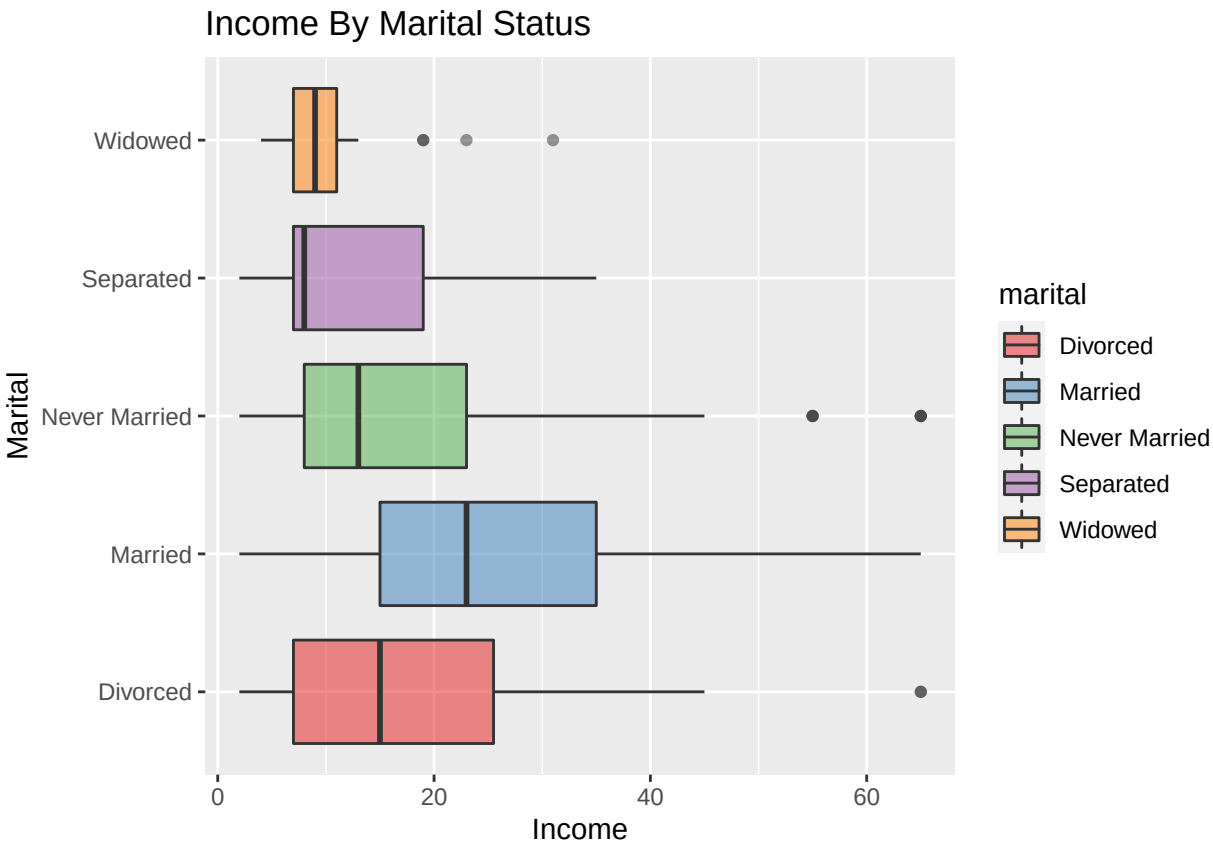
```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```



Bivariate Exploration

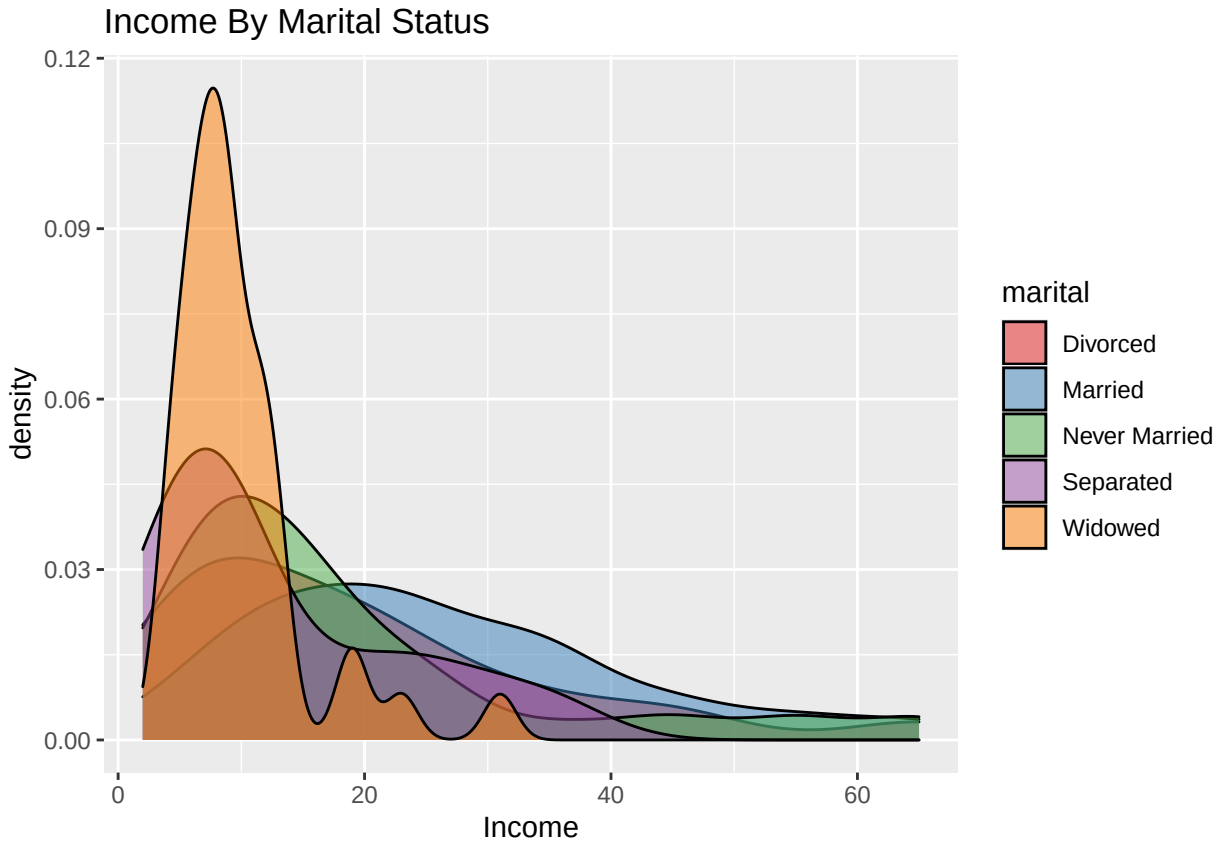
This box plot shows a summary of the distribution of income for each of the five categories of marital status. It gives the five-number summary of the distributions of the data. This shows the summary table for the income graphically separated by marital status.

```
ggplot(depress, aes(x=income, y=marital, fill=marital)) + geom_boxplot(alpha=.5) + ggtitle("Income By Ma
```



This density plot shows the distribution of income for each category of marital status in more detail than the box plot, but is lacking the numerical summary statistics of the box plot.

```
ggplot(depress, aes(x=income, fill=marital)) + geom_density(alpha=.5) + ggtitle("Income By Marital Statu
```



Conclusion

The boxplot shows the income distributions for each marital status is rather different. The income histogram is skewed right which shows that most respondents have a much lower income than the other respondents. The marital status bar chart shows that the largest percentage are married the smallest percentage are separated. The income by marital status shows that the widowed are the most skewed to the right, while the married are the least skewed to the right. Married has the most income, while widowed has the least income.