Data Anaylsis Project

Janette Munguia Math-130

3/8/2022

1. Introduction

The data that I will be analyzing in this report is the depression data set that contains 294 observations and 37 variables. This data set provides information on adult residents from Los Angeles County in regards to depression. The questions I will focus on is the marital status, the education level, and the participants income in which depression has effected the 294 participants.

2. Univivariate Descriptions
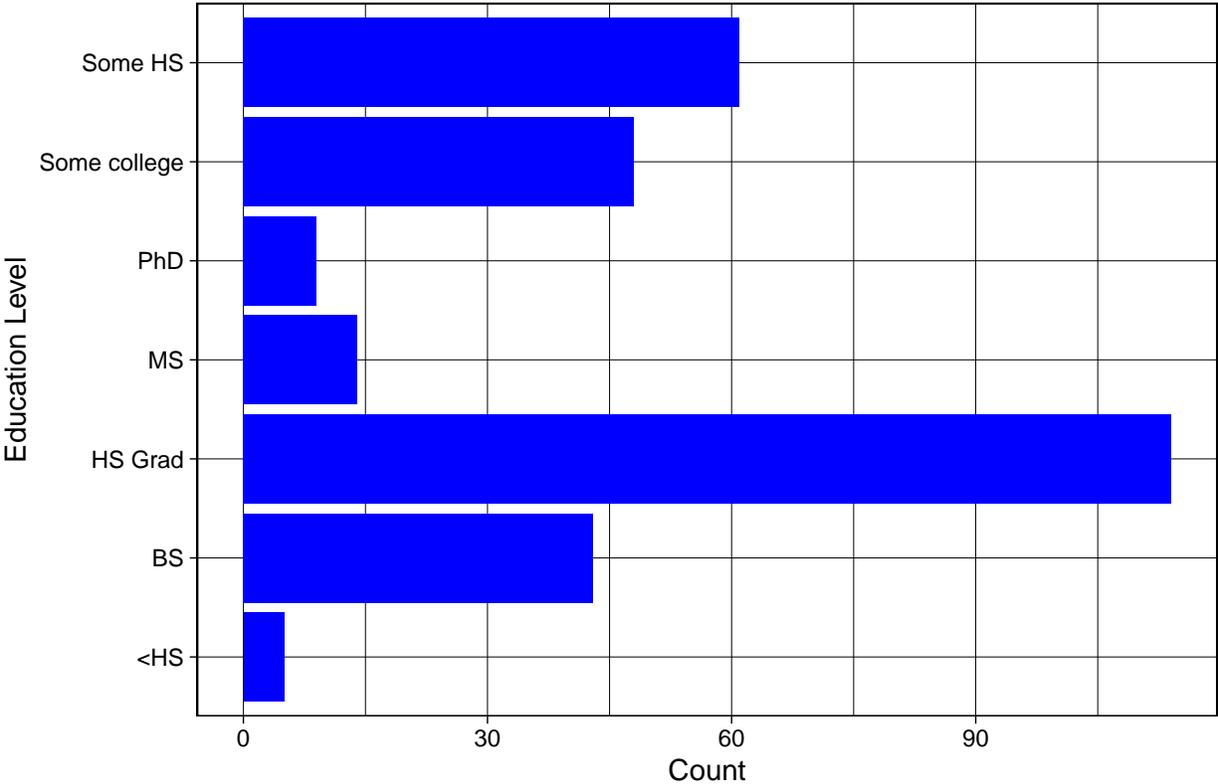
The information below shows the summary statistics for the education variable.

```
table(depression$educat)
```

```
##
##          <HS          BS      HS Grad          MS          PhD Some college
##            5          43          114          14            9           48
##      Some HS
##           61
```

```
ggplot(depression, aes(x=educat)) + geom_bar(fill="blue") + coord_flip() +
ylab("Count") + xlab("Education Level") +
ggtitle("Education Levels of Participants") + theme_linedraw()
```
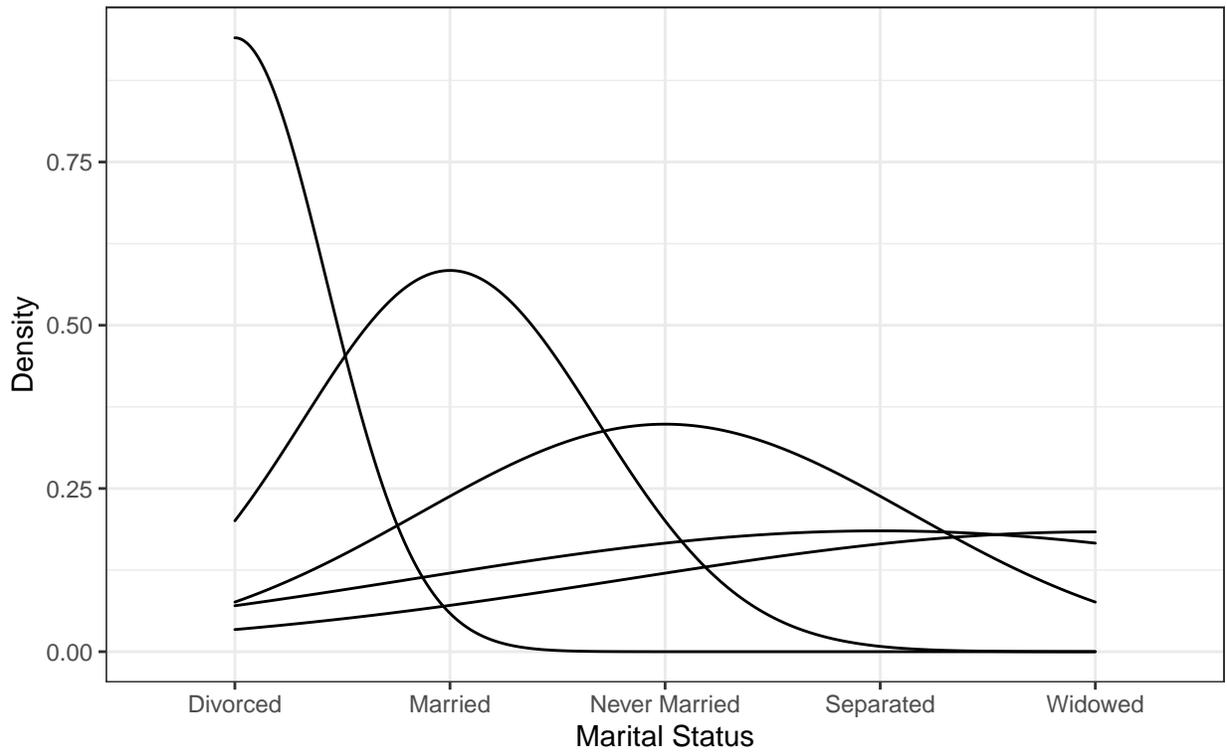
Education Levels of Participants

This graph displays the different education levels that the participants could potentially have. According to the graph, most of the participants were at "least a high school graduate". The education levels ranged from "less than high school" to having a "PHD".

```
ggplot(depression, aes(x=marital)) + geom_density() + ggtitle("Depression Level in relation to
    the Participants Marital Status") + xlab("Marital Status") + ylab("Density") +theme_bw()
```

Depression Level in relation to the Participants Marital Status

This graph shows the depression levels in relation to the participants marital status. The participants who were "separated" or "widowed" seemed to have a lower density than the participants who were still in a relationship. The "divorced" and "married" participants had a higher density out of all the groups.

```r
boxplot(depression$income, horizontal= TRUE, main= "Distribution of Income Status",
xlab="Income (In thousands of dollars)")
```

Distribution of Income Status
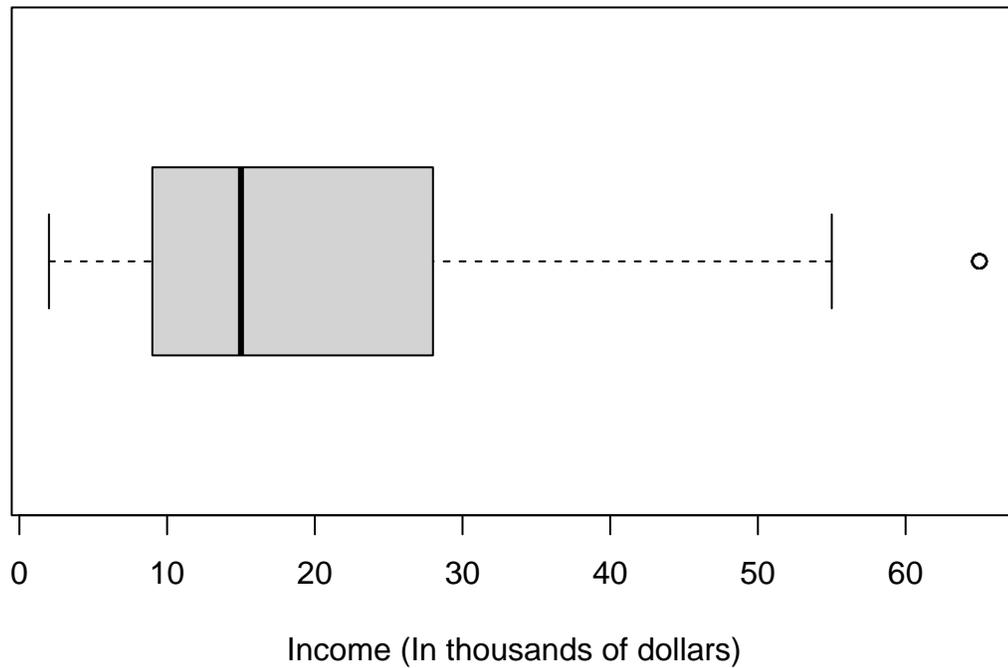

```
summary(depression$income)
```

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

The box plot shows the range of the income of the participants from $2,000 to $65,000 with an outlier at $65,000. The average income that the particpants made was 20.57 ($20,570) with a median at 15 ($15,000).

3. Bivariate Descriptions

```
ggplot(depression, aes(x=educat, y=income, fill=educat,)) + geom_boxplot(alpha=0.6) +
theme(legend.position="none") +
scale_fill_brewer(palette="Spectral") +ylab ("Income (In thousands of dollars)") +
xlab ("Education Status") +ggtitle("Income vs Education Status") +coord_flip()
```

Income vs Education Status
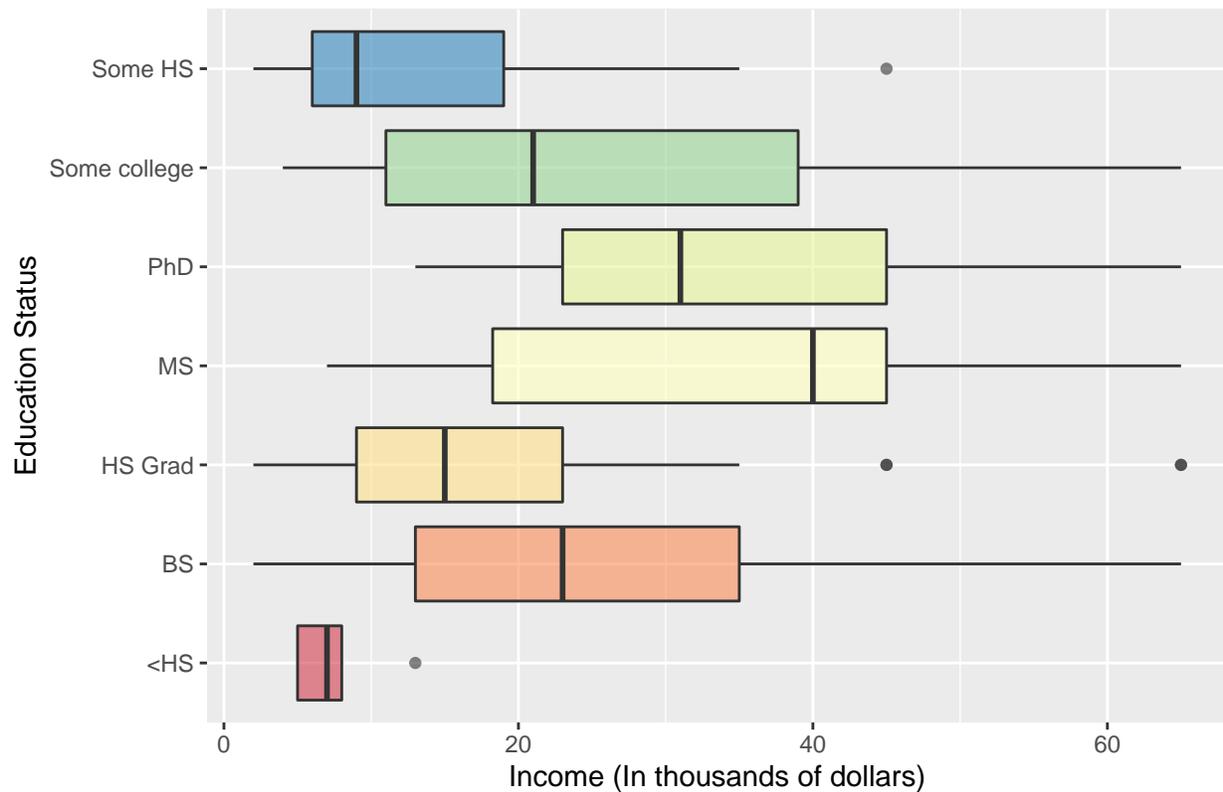
The grouped box plot shows the comparison between the education status and the income status which is represented in thousands of dollars. The participants who had a higher degree in education level made more money in comparison to those who had a lower level of education. For example, the participants who had a "Masters Degree" earned roughly $40,000 compared to a "high school graduate" who earned less than $20,000. The lowest income can be seen by those whose education status was lower than a high school status.

```
ggplot(depression, aes(x=marital, fill=educat)) +
geom_bar(position = "dodge", color="black") + xlab("Marital") + ylab("Count") +
scale_fill_discrete(name="Education Level") +
ggtitle("Marital Status by Education Levels") + theme_bw()
```

Marital Status by Education Levels
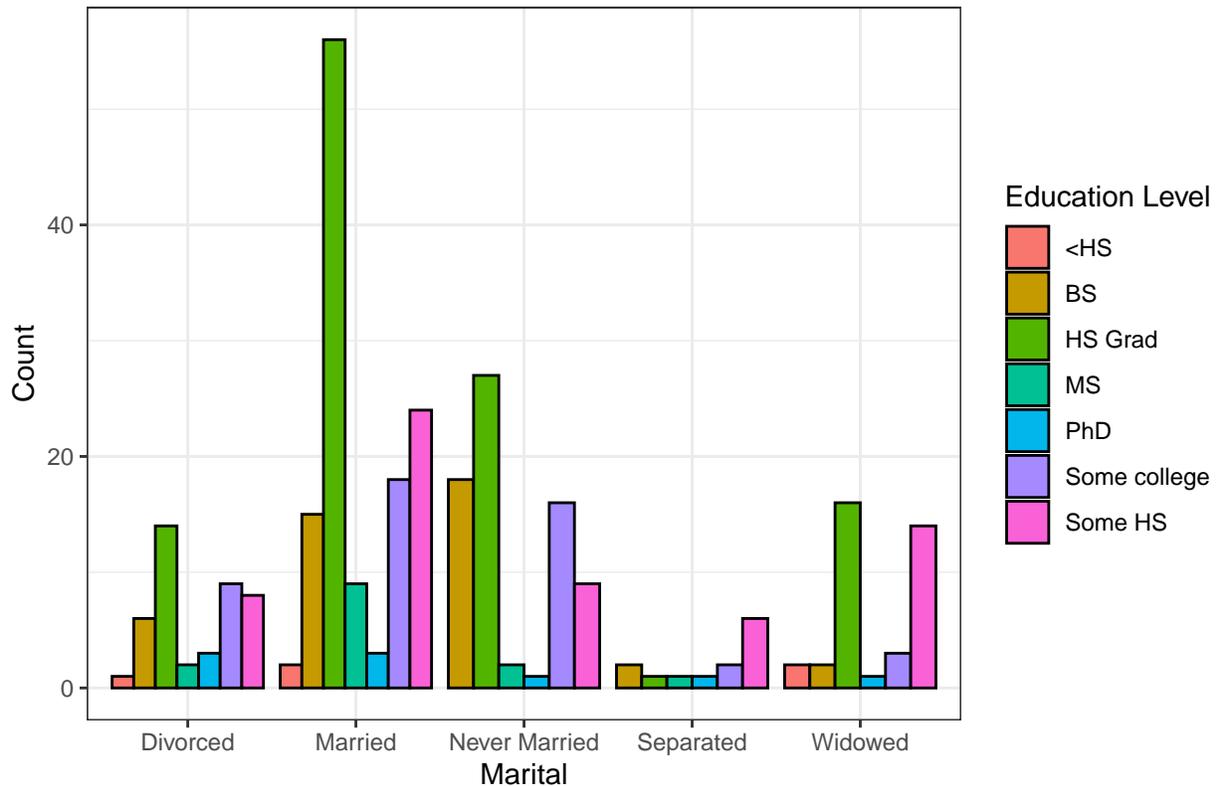

```
table(depression$marital)
```

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

The graph shows the Marital Status by the Education Level. The table shows the number of participants in each category. According to the graph, the participants who are "separated" have the lowest degree of education and those who are "married" have a higher education level. The count for having an education level that was lower than a high school level was dropped form "never married" and "separated" for having no value.

```
ggplot(depression, aes(x=marital, fill=educat)) + geom_density(alpha=0.3) +
scale_color_discrete(name="Education Level") + ylab("Density") +
xlab("Marital Status") + ggtitle("Marital Status by Education Levels") + theme_bw()
```

```
## Warning: Groups with fewer than two data points have been dropped.

## Warning: Groups with fewer than two data points have been dropped.

## Warning: Groups with fewer than two data points have been dropped.

## Warning: Groups with fewer than two data points have been dropped.
```

```
## Warning: Groups with fewer than two data points have been dropped.

## Warning: Groups with fewer than two data points have been dropped.

## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf

## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf

## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf

## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf

## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf

## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning
## -Inf
```


Marital Status by Education Levels
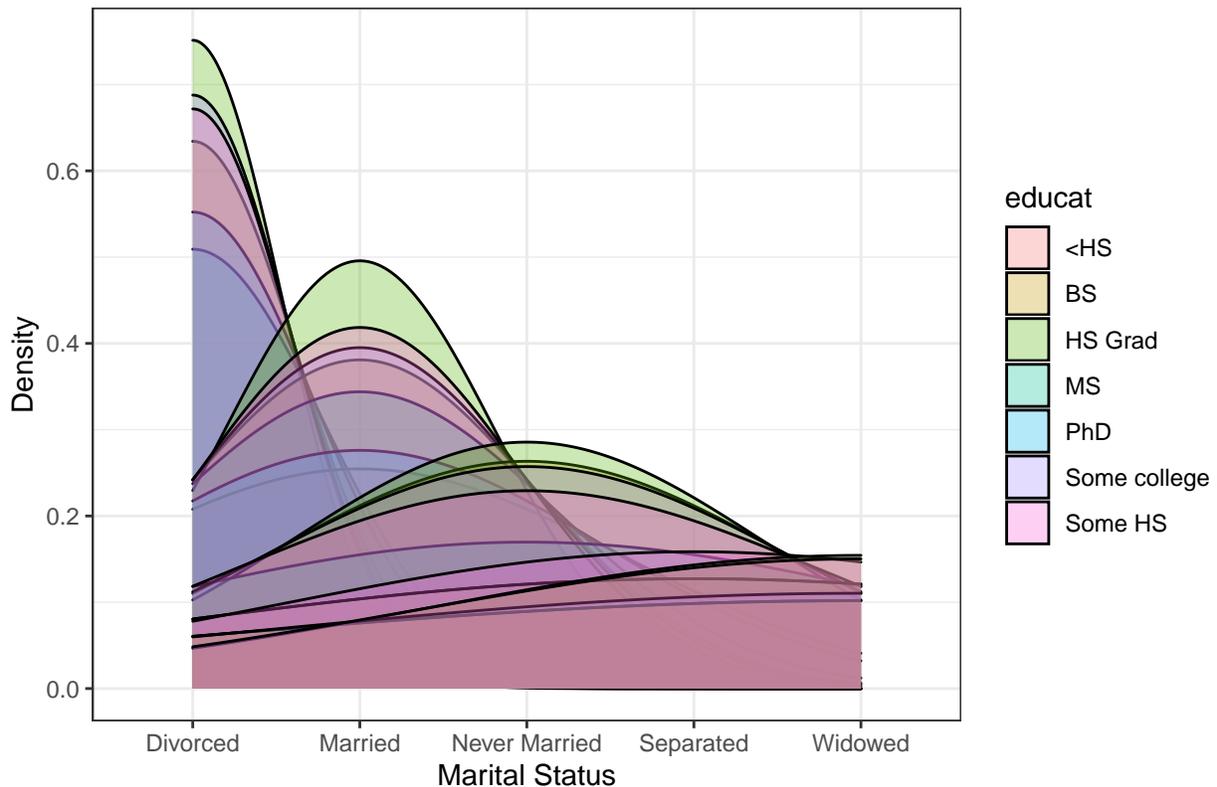
This overlaid density plot shows us that being a "high school graduate" was more commonly found in most of the participants marital status except in those who were "separated". There is a large difference in education status in those who were "divorced" and "married" than the individuals who were never "married", "separated", or "widowed". The density plot depicts a multimodal distribution of the education status.

Conclusion

The participants that were tested showed an education level that ranged from lower than high school to having a PHD. We were also able to see that for the employment status, the average income was approximately \$20,000 and ranged from \$2,000 to \$65,000. The relationship between the education level and the income showed that the participants who had a higher degree made more money compared to those who had a lower level of education.