Exploratory Data Analysis

Angelica Gallegos

2/25/2022

Introduction:

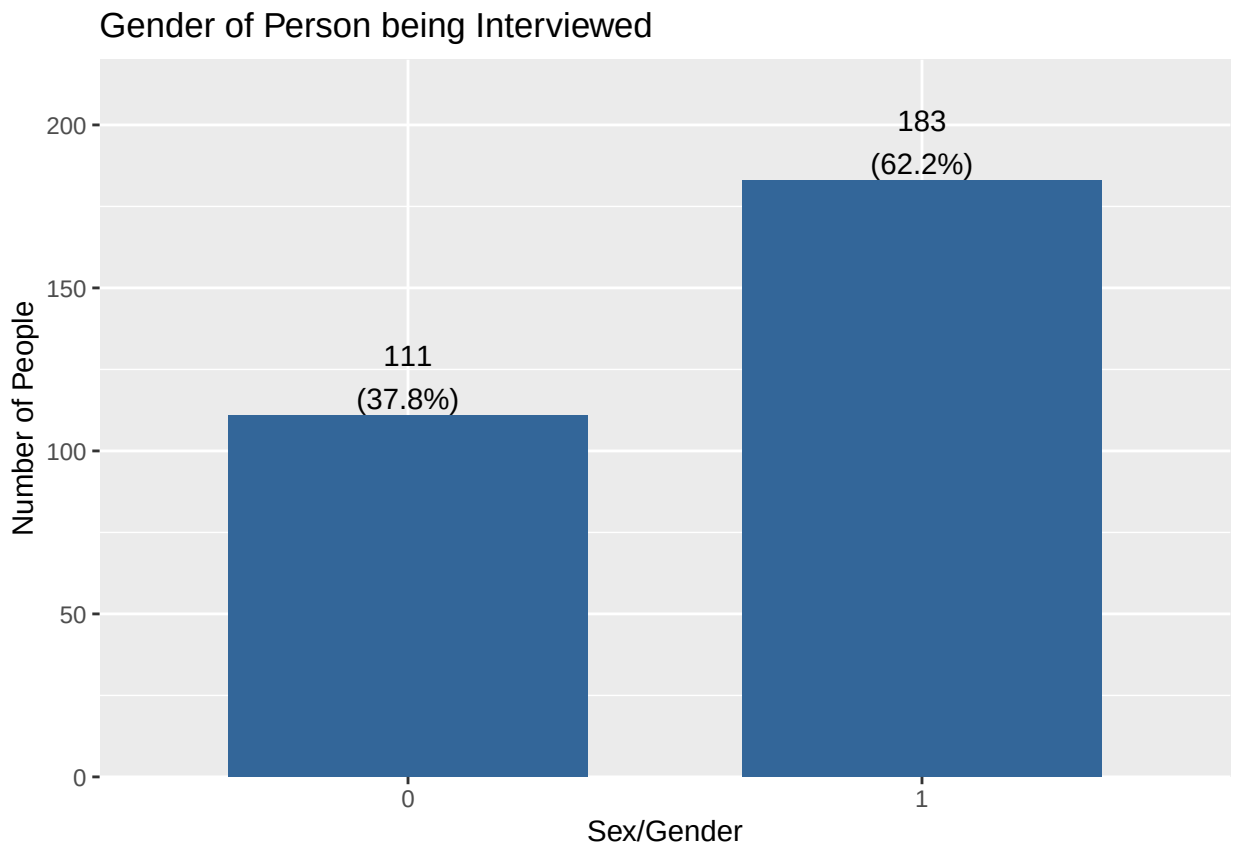
In this data analysis project I will be analyzing a data set named "Depression". This data set is of 294 interviews done on Los Angeles adult residents about depression.There are many variables in this ddta set but I will be specifically looking at three different variables,. First one is "sex", this is the gender of the person being interviewed, a males is shown as the number 0, a female is shown as the number 1. The next variable is "educat", this is the level of education the person being interviewed has received, there are 7 different choices in the level of education ranging from less than high school to finished doctorate.The last variable I will be looking at is "cesd", this is the level of depression based on the questions that were asked and their response to them with values ranging from as low as 0 to as high as 60. I am interested to look at the relationship between depression level(cesd) to the gender of the person being interviewed(sex) as well as the relationship between depression level(cesd) to the education the person being interviewed has received(educat).

Univariate Exploration:

Here, I will be looking in depth into each variable and any common or average occurrences we see within the variable.

Variable:sex

```
plot_frq(depress$sex)+ggtitle("Gender of Person being Interviewed")+xlab("Sex/Gender")+ylab("Number of
```



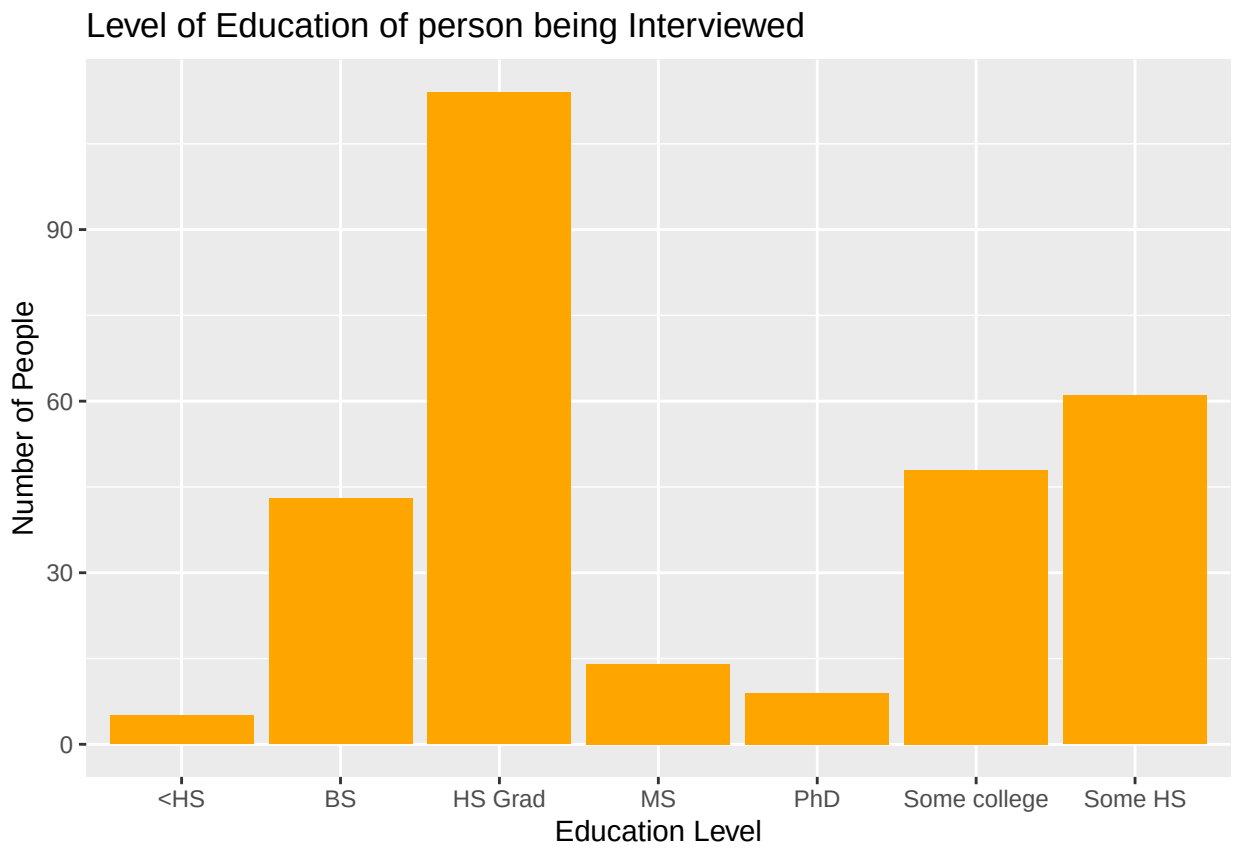

```
table(depress$sex)
```

```
##
##   0   1
## 111 183
```

The Frequency Distribution above shows the number of people being interviewed of each gender, we see a majority of people are female having about 1.5x as many participants than males. We can see the accurate number of 111 of males(0) and 183 of females(1) by using the table above and the frequency distribution as well.

Variable:educat

```
ggplot(depress,aes(x=educat))+geom_bar(fill='orange')+ggtitle("Level of Education of person being Inter
```



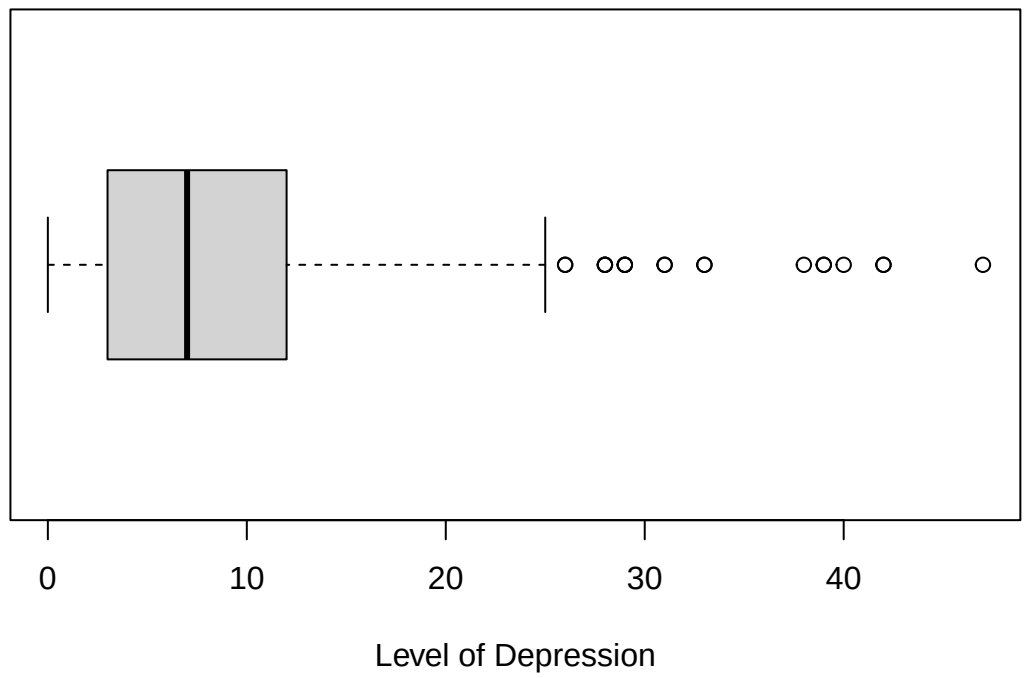

```
table(depress$educat)
```

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

The Bar Graph above shows the level of education the person being interviewed has received. We can see that the level of education that has more people than any other level was that of high school grad.The table above gives us the specific quantity of people in each level of education that we couldn't accurately see in the bar graph.

Variable:cesd

```
boxplot(depress$cesd,horizontal = TRUE,main="Distribution of Depression Level",xlab=("Level of Depressi
```




```
summary(depress$cesd)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   3.000   7.000   8.884  12.000  47.000
```

The boxplot above shows us the Distribution of Depression level of people being interviewed. We can see that the average level is fairly low with 75% of people being under about 12. There is also some very high values ranging up to nearly 50. We can precisely measure the distribution with the summary table, which accurately depicts the quarters and median of the data.

Bivariate Exploration: Here I will be comparing (1) the sex variable with depression level and (2) the education level with depression level.

Comparison Between Sex vs. Depression Level:

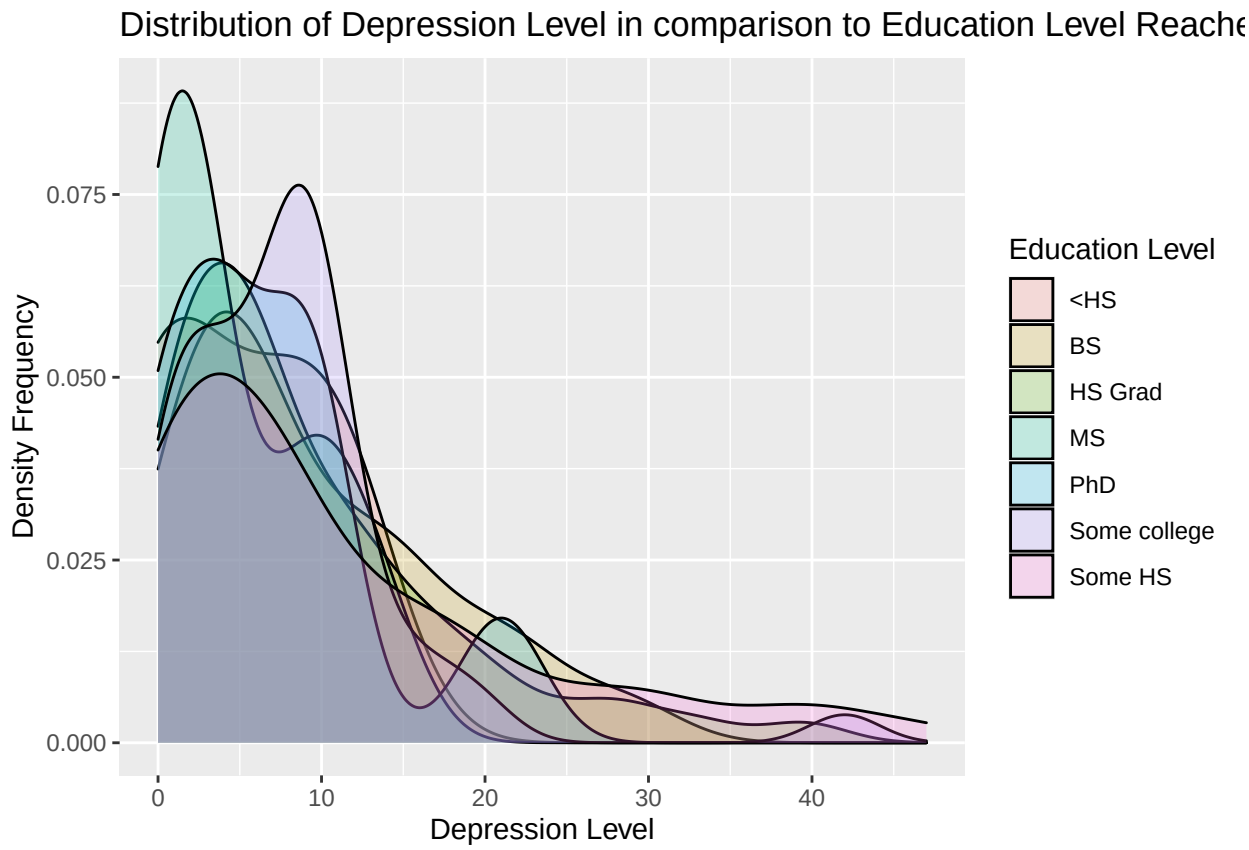
```
table(depress$sex,depress$cesd)
```

```
##
##      0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 24 25
##   0 17  5  5  9  7 10  5  7  8  8  4  3  4  4  1  4  1  0  1  0  1  1  1  0  1
##   1 17 15  8 16 13  8 10  9  9  9  7  4  6  6  4  2  4  4  5  3  1  4  3  1  0
##
##     26 28 29 31 33 38 39 40 42 47
##   0  1  1  0  1  0  0  0  0  1  0
##   1  1  2  3  1  2  1  2  1  1  1
```

The Frequency Table above shows us the number of both males and females seperately in regards to their depression level. With this table we are able to see that males and females had the same number of people in the same depression level in the 0, 20, 26, and 42 levels.

Comparison Between Education Level vs. Depression level:

```
ggplot(depress,aes(x=cesd,fill=educat))+geom_density(alpha=.2,color='black')+ggtitle("Distribution of D
```



The density plot above shows us the relationship and distribution of depression level in regards to the educational level received. We can see that there is a peak in MS education level right at the beginning which happens to be a very low depression rate as well as it is the highest peak overall throughout the density plot.

Conclusion: I found that in the relationship between Sex and Depression Level there is lower depression level overall in both genders. There are some outlier depression levels for each but overall, both genders in this study tend to have a depression level under around 15. In the relationship between Education Level and Depression Level, there is peaks in depression levels under about 15 for all levels of education. As depression level goes higher, the frequency of people in that education level is lower.