

Homework 5: Exploratory Data Analysis Project

Carissa Laramore

September 19, 2022

```
library(ggplot2)
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(sjPlot)
```

Introduction: Depression

In my Exploratory Data Analysis Project I am going to analyze the study of depression in adults living in Los Angeles County. The data set has 294 observations and the variables that I chose to explore in this project are sex, education, and religion. Other variables in the data set include age, marital status, and income amongst many others. The question that I want to explore is if males or females have a higher rate of depression and if their education and religion status has an affect on that result.

Read in the "Depression" data set.

```
depress <- read.delim("/Users/carissalaramore/Desktop/Math 130/depress_081217.txt", header=TRUE, sep="\t")
```

Univariate Exploration

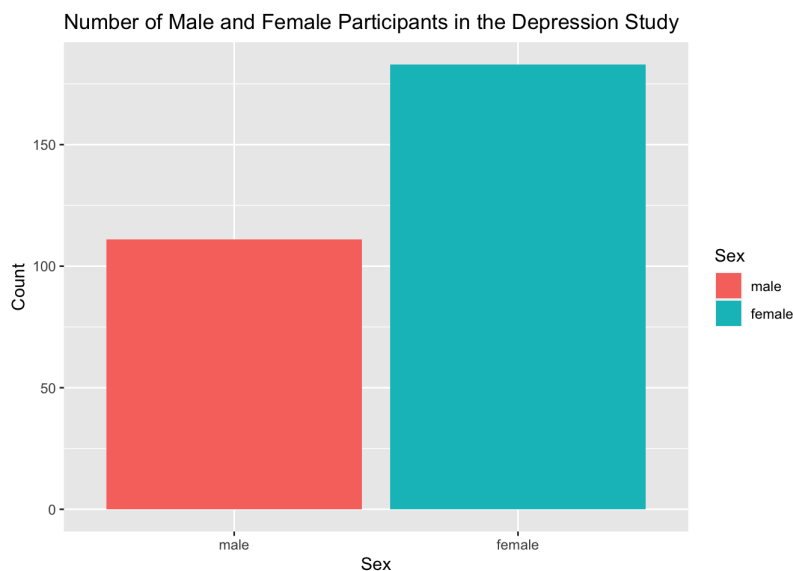
Sex of Individuals

```
depress$sex <- factor(depress$sex, labels=c("male", "female"))
table(depress$sex)
```

```
##
##  male female
##   111   183
```

This table shows that in the entire study there are 111 male participants and 183 female participants.

```
ggplot(depress, aes(x=sex, fill=sex))+ geom_bar()+ ggtitle("Number of Male and Female Participants in the Depression Study")+ xlab("Sex")+ ylab("Count")+ scale_fill_discrete("Sex")
```



This bar graph represents the amount of male individuals versus the amount of female individuals that were participating in this study. You can see that there was a significantly larger group of females that were examined in this study. I think this can affect the results that I conclude from the data but I think the group that was observed was too small to make any real inferences without any further data collection.

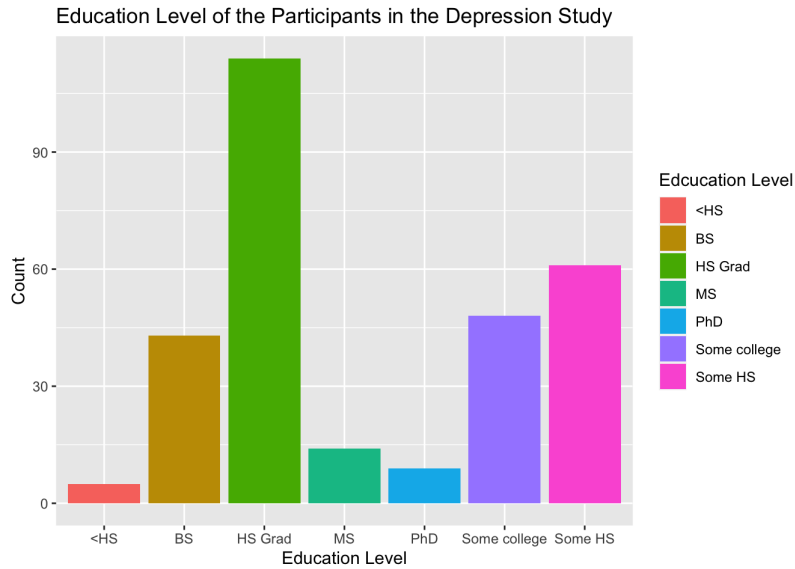
Education Level of Individuals

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

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

This table shows us the number of individuals in each level of education for the entirety of the population that was studied. The majority of the population were at most high school graduates with the smallest group being less than a high school education.

```
ggplot(depress, aes(x=educat, fill=educat))+ geom_bar()+ ggtitle("Education Level of the Participants in the Depression Study")+ xlab("Education Level")+ ylab("Count")+scale_fill_discrete("Education Level")
```



This bar graph represents the number of individuals that are in each group for levels of education for the entirety of the population that was studied. The study showed that less than 5 individuals achieved less than high school education, 61 had some high school education, and 114 graduated from high school. It also shows that 48 had some college-level education, 43 achieved their Bachelor's degree, 14 achieved their Master's degree, and 9 completed their PhD.

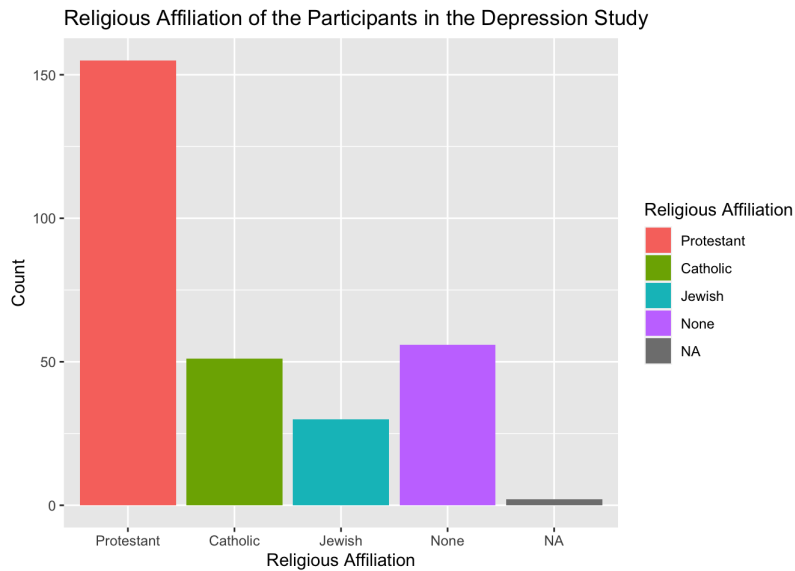
Religious Affiliation of Individuals

```
depress$relig <- ordered(depress$relig, levels = c(1,2,3,4,5), labels=c("Protestant", "Catholic", "Jewish", "None", "Other"))  
table(depress$relig)
```

```
##  
## Protestant Catholic Jewish None Other  
##      155      51      30      56      0
```

This table shows us the number of individuals in each religious affiliation for the entirety of the population that was studied. The majority of the population claim that they are Protestant with the smallest group being "Other" having 0.

```
ggplot(depress, aes(x=relig, fill=relig))+ geom_bar()+ ggtitle("Religious Affiliation of the Participants in the Depression Study")+ xlab("Religious Affiliation")+ ylab("Count")+scale_fill_discrete("Religious Affiliation")
```



This bar graph represents the number of individuals that are in each group for religious affiliation for the entirety of the population that was studied. The group with the highest with 155 Protestants, 56 with no religious affiliation, 51 Catholic, 30 Jewish, and 0 that have other religious affiliation.

Bivariate Exploration

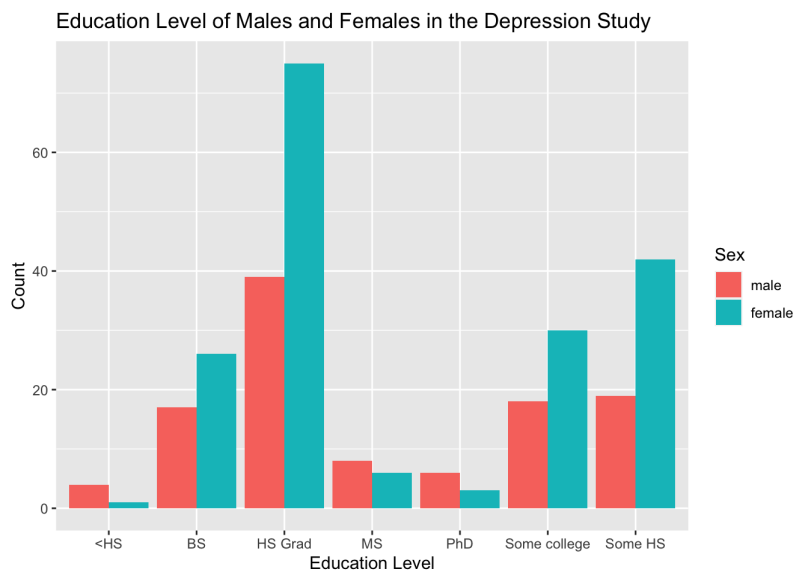
Sex and Education Level

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

```
##
##      <HS BS HS Grad MS PhD Some college Some HS
## male   4 17  39  8  6           18    19
## female  1 26  75  6  3           30    42
```

This table is going to show us the sex of the participant that is being studied and their level of education that they achieved.

```
ggplot(depress, aes(x=educat, fill=sex)) + geom_bar(position = "dodge")+ ggtitle("Education Level of Males and Females in the Depression Study")+ xlab("Education Level")+ ylab("Count")+scale_fill_discrete("Sex")
```



We can see on the graph that the group with the highest rate of depression would be females that have a high school graduate education level. In a little over half of the levels of education the number of females is greater than the number of males, this could possibly mean that females have a higher rate of depression but since the data is so staggered I don't think we can conclude that education level has an affect on rates of depression in males and females.

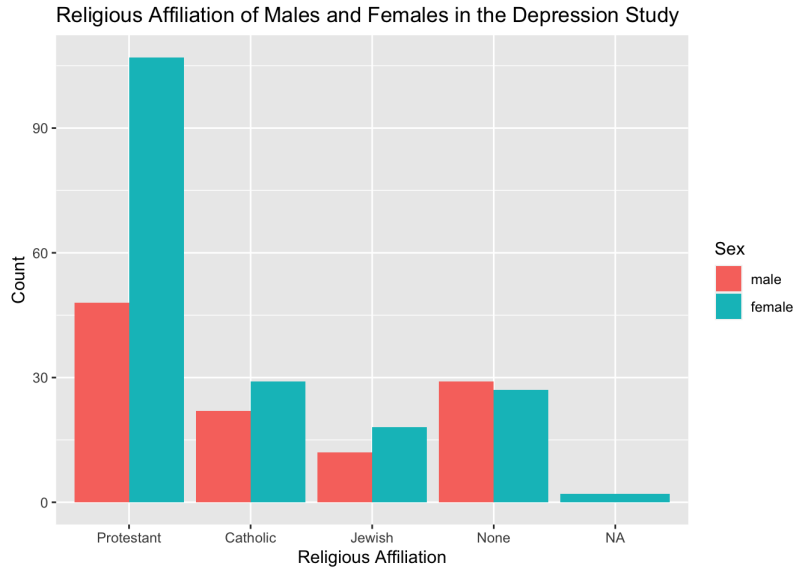
Sex and Religious Affiliation

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

```
##
##      <HS BS HS Grad MS PhD Some college Some HS
## male   4 17  39  8  6           18    19
## female  1 26  75  6  3           30    42
```

This table is going to show us the sex of the participant that is being studied and their religion that they identify with.

```
ggplot(depress, aes(x=relig, fill=sex)) + geom_bar(position = "dodge")+ ggtitle("Religious Affiliation of Males and Females in the Depression Study")+ xlab("Religious Affiliation")+ ylab("Count")+scale_fill_discrete("Sex")
```



This graph shows us that female Protestants have the highest rate of depression in this study. I also believe that we can't make a definite conclusion about religious affiliation affecting the rate of depression in males and females because the sample size is small and the proportions of males to females is off.

Conclusion

What the data in this study shows is that females have a higher rate of depression. I don't think the sample size is big enough to make conclusions about the entire population and the proportion of males and females is not equal which would make it easier to conclude which sex has a higher rate of depression. I can also conclude that education level and religious affiliation does not seem to have an affect on depression rate but this is something that I could investigate further with further tests and a different population.