Final Project

Mia Moilanen

9/27/2021

```
library(ggplot2)
depression <- read.table("C:/Users/moilm/Downloads/MATH130/depress_081217.txt",
header=TRUE, sep="\t")
head(depression)
```

```
##   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
```

Introduction

This analysis is based on the depression data set, which includes results from a survey taken in Los Angeles County with 294 adult participants. The survey contains questions focused on demographics, and the lifestyle and mental health of the participants. This analysis will explore three variables of interest: SEX, CESD, and DRINK. The variable "SEX" describes the gender of the participants, with the options "Male" or "Female". "CESD" describes the level of depression experienced by the participants, with the lowest possible level being "0" and the highest being "60". These numbers are obtained from the sum of previous questions in which the participants ranked how often they had experienced specific symptoms of depression in the last week. "DRINK" describes whether the participant is a regular drinker, with the options being "Yes" or "No". The relationship between level of depression and gender will be explored, alongside level of depression and drinking habits, to determine if there is any correlation.

Univariate Description

VARIABLE: SEX

```
depression$sex_fac <- factor(depression$sex, labels=c("Male", "Female"))
table(depression$sex_fac, useNA="always")
```

```
##
##   Male Female   <NA>
##    111    183      0
```

Above is a table describing the variable "SEX", which, as stated previously, includes two options: "Male" or "Female". Of the 294 participants in this study, 111 are male and 183 are female.

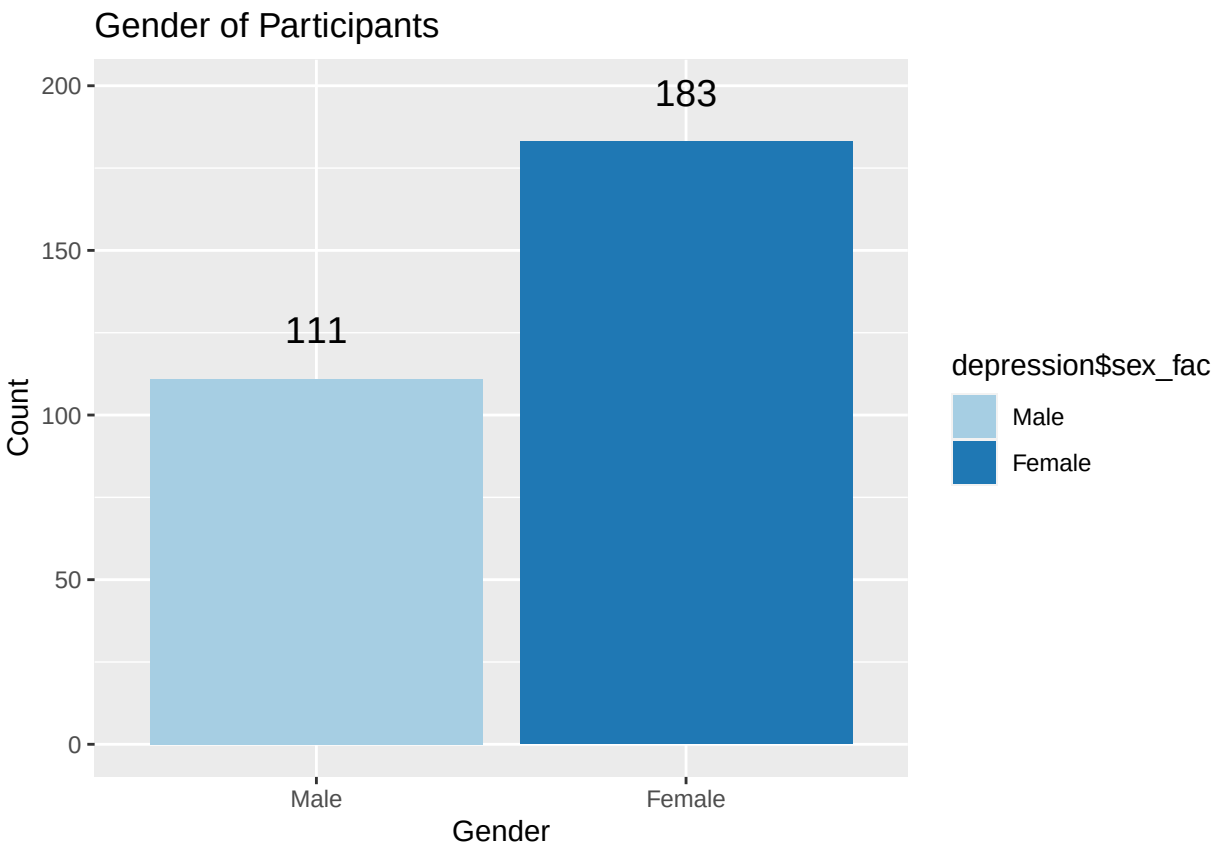
```
ggplot(depression, aes(x= depression$sex_fac, fill= depression$sex_fac)) +
geom_bar(aes(y=..count..)) + scale_fill_brewer(palette= "Paired") +
ggtitle("Gender of Participants") + ylab("Count") + xlab("Gender") +
  geom_text(aes(y=..count.. + 15, label=..count..),
            stat='count', size = 5)
```

```
## Warning: Use of ‘depression$sex_fac‘ is discouraged. Use ‘sex_fac‘ instead.

## Warning: Use of ‘depression$sex_fac‘ is discouraged. Use ‘sex_fac‘ instead.

## Warning: Use of ‘depression$sex_fac‘ is discouraged. Use ‘sex_fac‘ instead.

## Warning: Use of ‘depression$sex_fac‘ is discouraged. Use ‘sex_fac‘ instead.
```



Above is a barchart representing the variable "SEX", with the marked gender of the participants as the x-axis and the number of participants as the y-axis. This chart shows more females participated in this study than males.

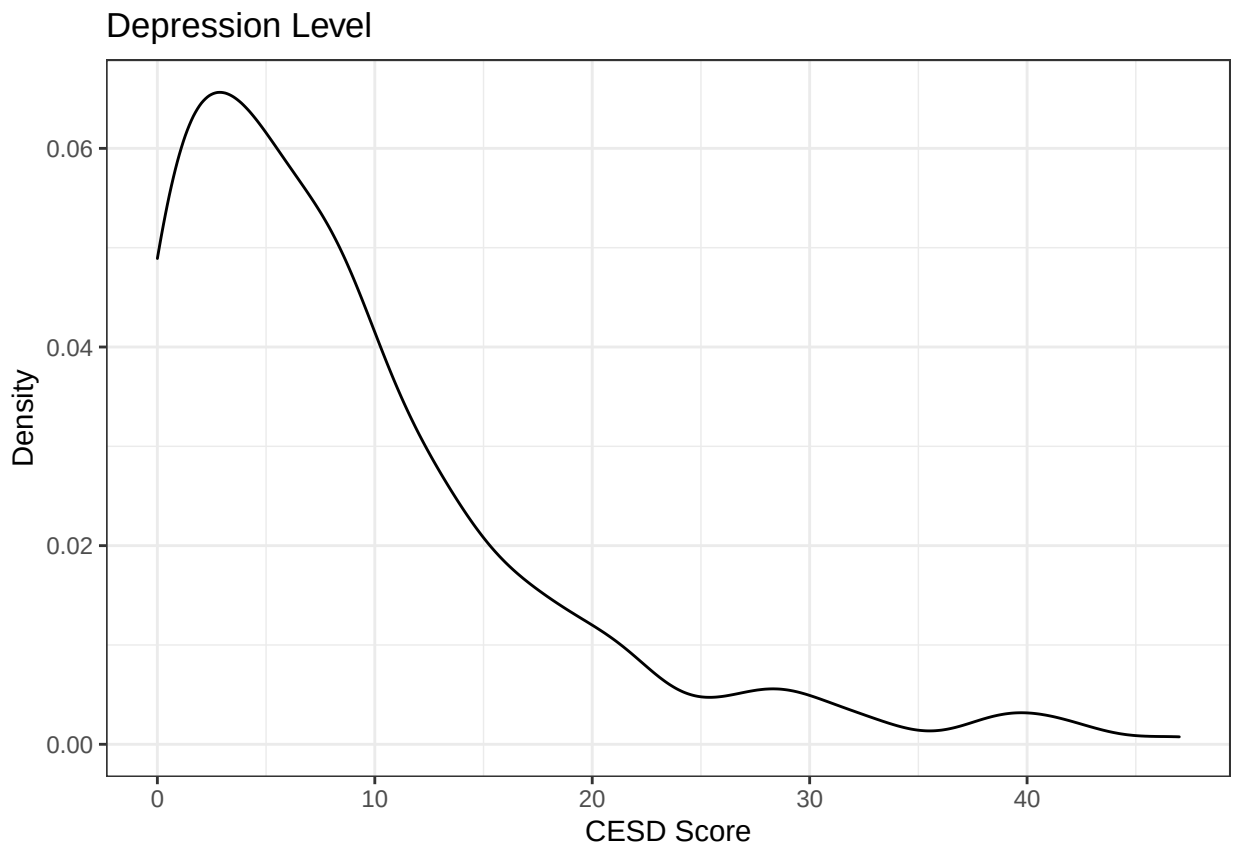
VARIABLE: CESD

```
summary(depression$cesd)
```

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

Above is a summary of the variable "CESD", which evaluates an individual's depression level based on the sum of their answers to previous questions. The highest level an individual can get is 60, and the lowest is 0. The maximum score of depression participants received in this study is 47, and the minimum is 0. The mean score is 8.884.

```
ggplot(depression, aes(x=cesd)) + geom_density() + ggtitle("Depression Level") +
  xlab("CESD Score") + ylab("Density") +theme_bw()
```



Above is a density graph showing the levels of depression in the participants, with depression level as the x-axis and density as the y-axis. As shown in the graph, a majority of the data is congregated towards the left, indicating most participants received scores of lower levels of depression.

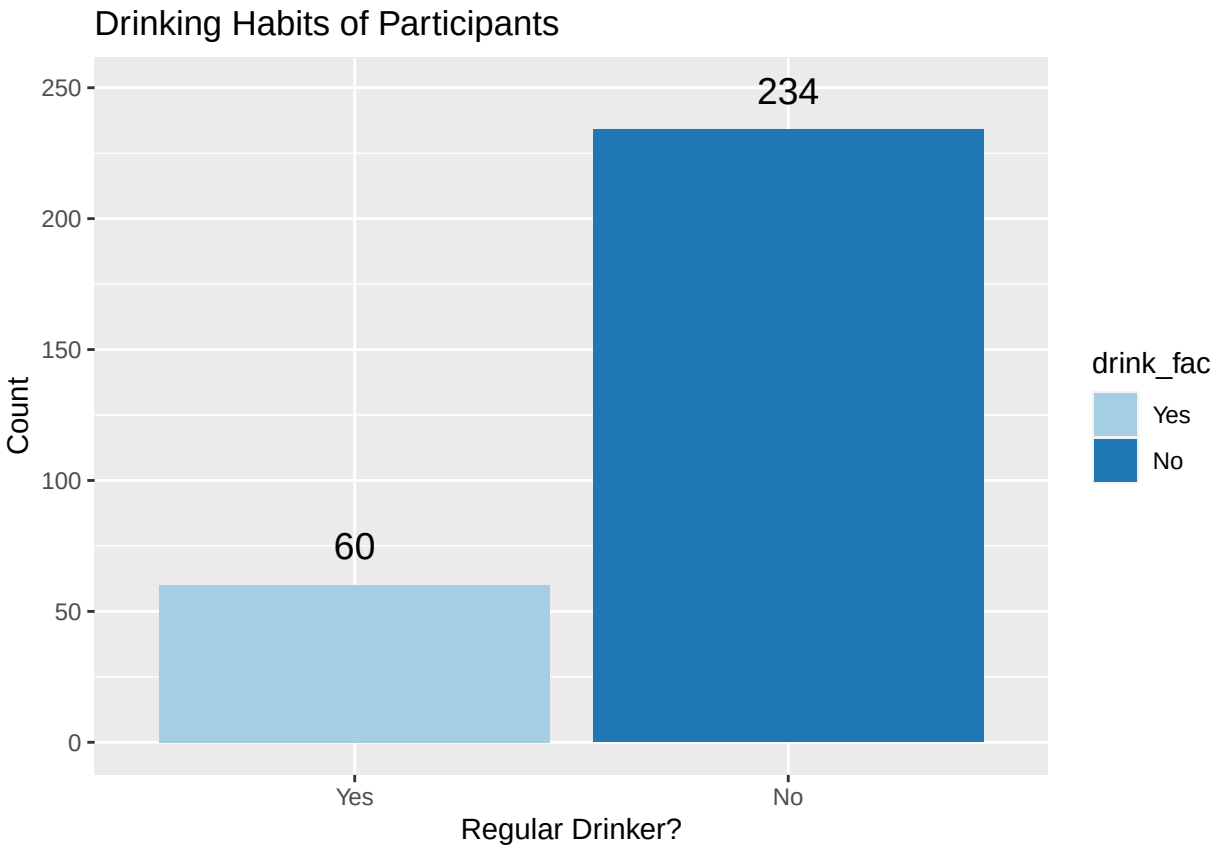
VARIABLE: DRINK

```
depression$drink_fac <- factor(depression$drink, labels=c("Yes", "No"))
table(depression$drink_fac)
```

```
##
## Yes  No
##  60 234
```

The above table describes the variable "DRINK", in which participants stated whether or not they were regular drinkers. As shown in the table, 234 of the 294 participants said they were not regular drinkers, while 60 said they were.

```
ggplot(depression, aes(x=drink_fac, fill=drink_fac)) +
  geom_bar(aes(y=..count..)) + scale_fill_brewer(palette= "Paired") +
  ggtitle("Drinking Habits of Participants") + ylab("Count") +
  xlab("Regular Drinker?") + geom_text(aes(y=..count.. + 15, label=..count..),
                                       stat='count', size = 5)
```

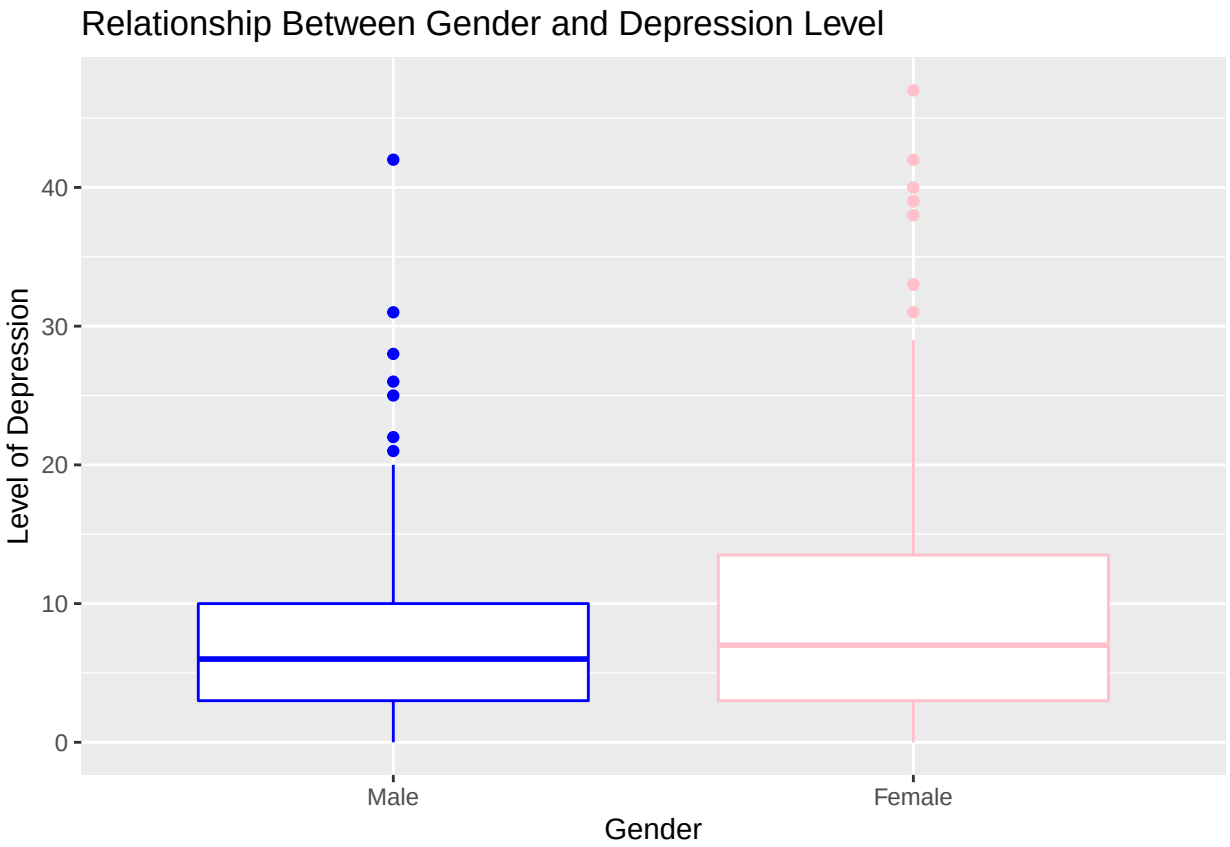


Above is a barchart depicting the results of the variable "DRINK". As shown, a majority of participants stated they were not regular drinkers.

Bivariate Comparison

SEX vs CESD

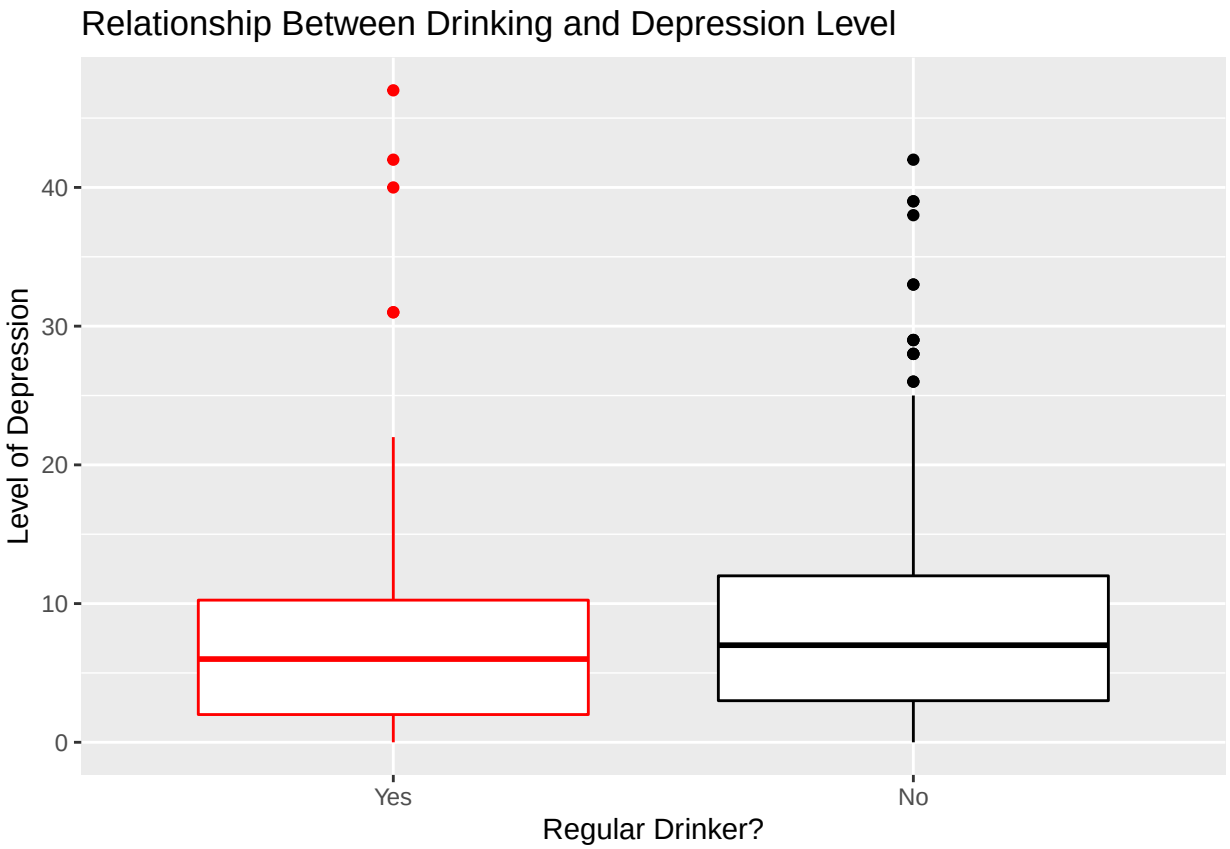
```
ggplot(depression, aes(x=sex_fac, y=cesd, col=sex_fac)) + geom_boxplot()+
  xlab("Gender")+ ylab("Level of Depression")+
  ggtitle("Relationship Between Gender and Depression Level")+
  scale_color_manual(values = c("blue", "pink"), guide= FALSE)
```



This boxplot demonstrates the potential relationship between gender and depression level, with the x-axis representing gender and the y-axis representing the participants' depression scores. A majority of men in this study experience depression levels below 10, with the highest outlier receiving a score of about 43. A majority of females in this study have depression levels lying under 15, with the highest outlier appearing to be the maximum score of 47. The male and female groups have the same minimum lying at 0, and similar medians, with the females' being just above the males'. Women have a greater IQR, but this may be due to there being a more female participants. Men and women have the same number of outliers, but the women's lie higher than the men's. The information from this boxplot indicates a potential relationship between gender and depression, with women experiencing greater symptoms.

DRINK VS CESD

```
ggplot(depression, aes(x=drink_fac, y=cesd, col=drink_fac)) + geom_boxplot()+
  xlab("Regular Drinker?")+ ylab("Level of Depression")+
  ggtitle("Relationship Between Drinking and Depression Level")+
  scale_color_manual(values = c("red", "black"), guide= FALSE)
```



The above boxplot shows the potential relationship between drinking and depression level. A majority of the participants who are regular drinkers experience levels of depression under 10, while a majority of the participants who are not regular drinkers experience levels of depression under 15. Non-regular drinkers have a greater IQR, but this may be due to their greater abundance. While the majority of regular drinkers appeared to receive slightly lower scores of depression, this group has its outliers sitting higher than those of the non-regular drinkers, with the highest being the maximum 47. However, due to the inconsistency of the correlation, there does not appear to be a significant relationship between depression level and drinking habits.