

Final Project

Chloe Taylor

9/22/2020

```
library(ggplot2)
library(forcats)
library(RColorBrewer)

Depression <-
read.table("C:/Users/chloe/OneDrive/Desktop/Math130/depress_081217.txt",
header=TRUE, sep="\t")
```

Data

For this project I will be analyzing the Depression data set. This data set explores interview results of adults living in Los Angeles County. The questions focus on the lifestyle, demographics, and mental health of the 294 participants.

First I will be analyzing the variables 'drink' and 'cesd' to see if there is a correlation between drinking regularly and depression. The 'cesd' variable is a numerical value that gives participants a score (1-60) based on the questions relating to depression scale. I expect to see a correlation between drinking and depression after analyzing these variables.

Next I will be looking at the variables 'income' and 'cesd' to see if there is any connection between the incomes of the people interviewed and their depression scores. It seems likely that decreased economic stability could be a factor linked to depression.

Finally, I will be examining the connection between chronic illness and depression using the variables 'chronill' and 'cesd'.

Univariate Description

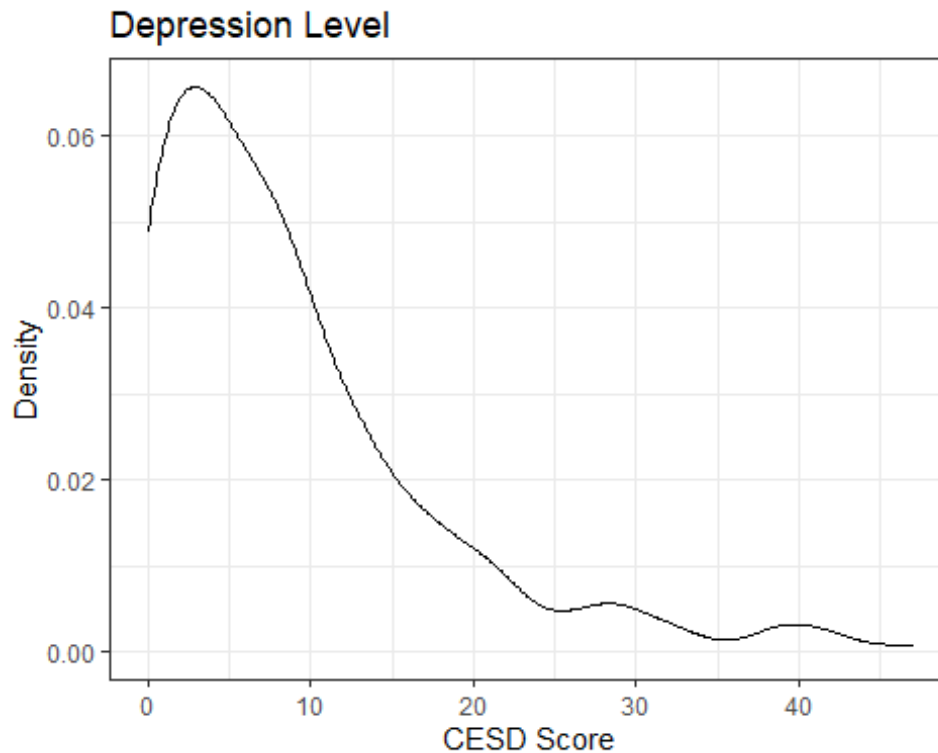
CESD VARIABLE The information below shows the summary statistics for the CESD variable. As I mentioned above, the CESD variable is a numerical value that gives participants a score (1-60) based on the questions relating to depression scale. The maximum depression score that anyone received in this study was a 47. The mean score was an 8.884.

```
summary(Depression$cesd)
```

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

This is a density plot of the CESD variable. As you can see, most of the participants scores are congregated near the left side of the plot meaning they received lower scores and likely don't have clinical depression.

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

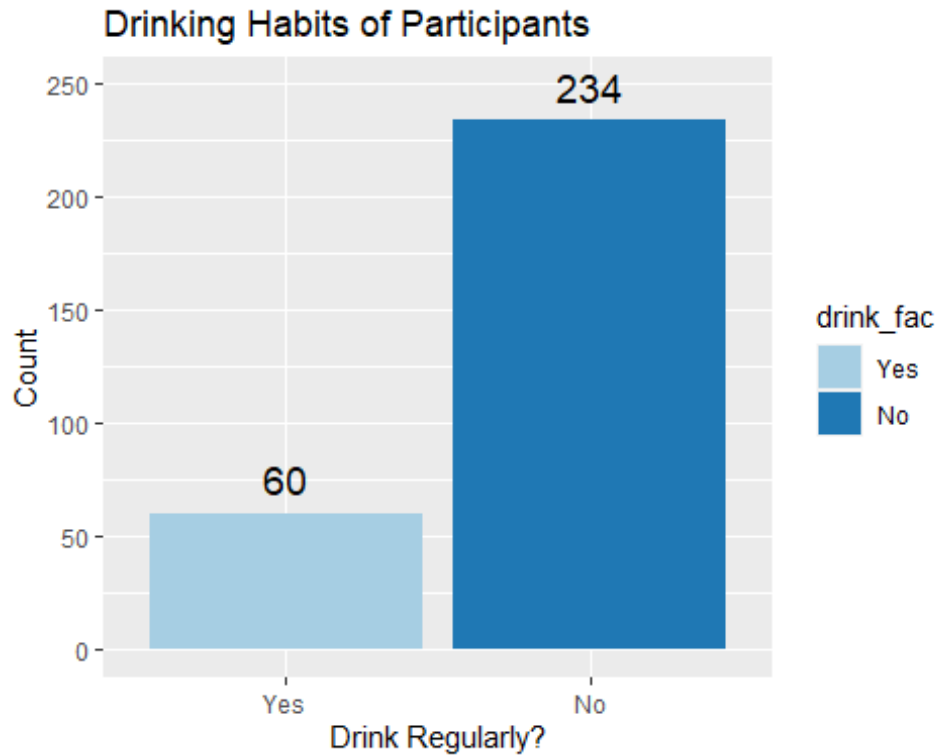


DRINK VARIABLE The table and bar graph below show the number of participants who answered 'Yes' or 'No' when asked if they were a regular drinker. The majority of participants answered 'No' to this question.

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

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

```
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("Drink Regularly?") +  
geom_text(aes(y=..count.. + 15, label=..count..),  
stat='count', size = 5)
```



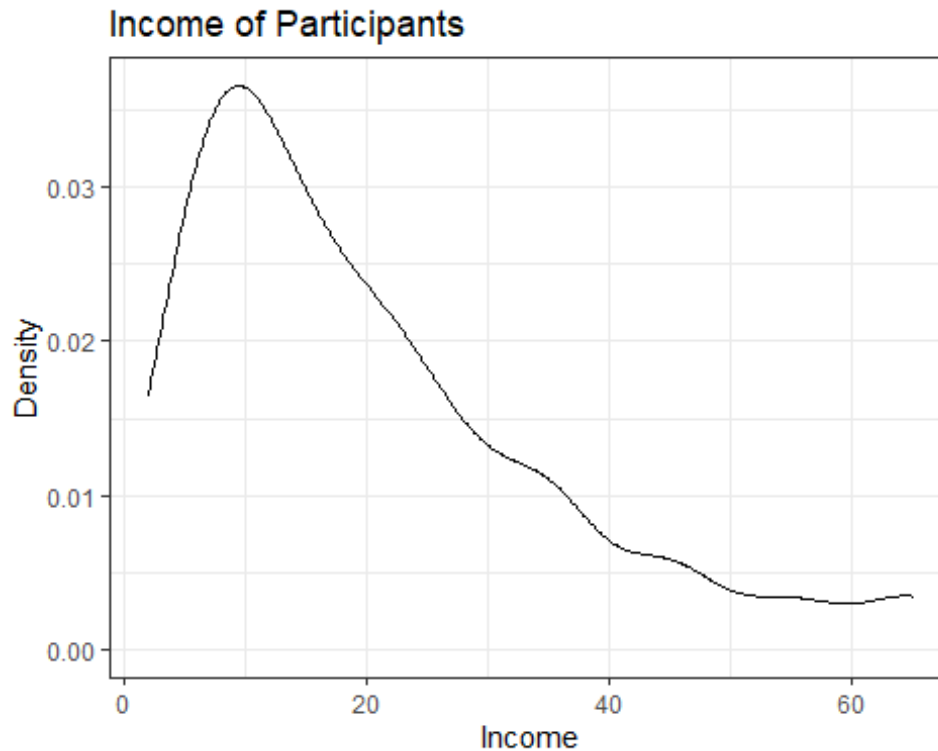
INCOME VARIABLE The summary statistics for the Income variable are shown below. The numbers represent the participants' income in thousands of dollars per year. The Maximum income was \$65,000 and the minimum income was \$2,000. The average income of the participants in this study was \$20,570 per year.

```
summary(Depression$income)
```

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

The density plot below shows the distribution of incomes of the participants in this study. The majority of people made below \$20,000 per year.

```
ggplot(Depression, aes(x= income)) + geom_density() + ggtitle("Income of
Participants") + ylab("Density") + xlab("Income") + theme_bw()
```



CHRONIC ILLNESS VARIABLE The table and bar graph below show the number of participants that answered 'Yes' or 'No' when asked if they had experienced any chronic illness in the past year.

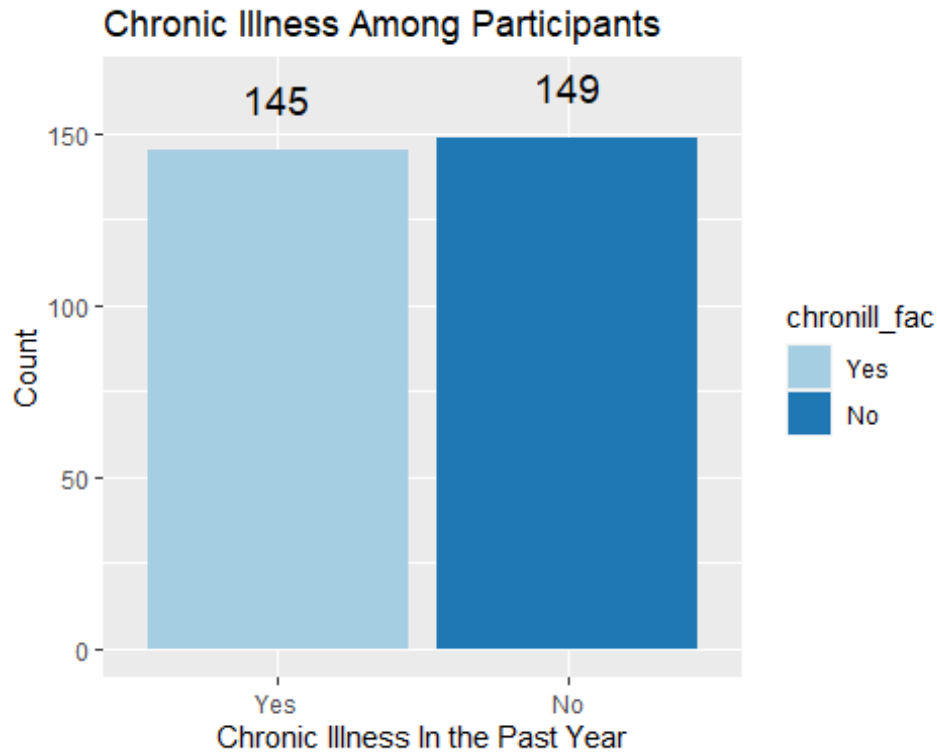
```

Depression$chronill_fac <- factor(Depression$chronill, labels=c("Yes", "No"))
table(Depression$chronill_fac)

##
## Yes No
## 145 149

ggplot(Depression, aes(x= chronill_fac, fill=chronill_fac)) + geom_bar(aes(y
= ..count..)) + ggtitle("Chronic Illness Among Participants") + xlab("Chronic
Illness In the Past Year") + ylab("Count") + geom_text(aes(y=..count.. + 15,
label=..count..), stat='count', size = 5) + scale_fill_brewer(palette=
"Paired")

```

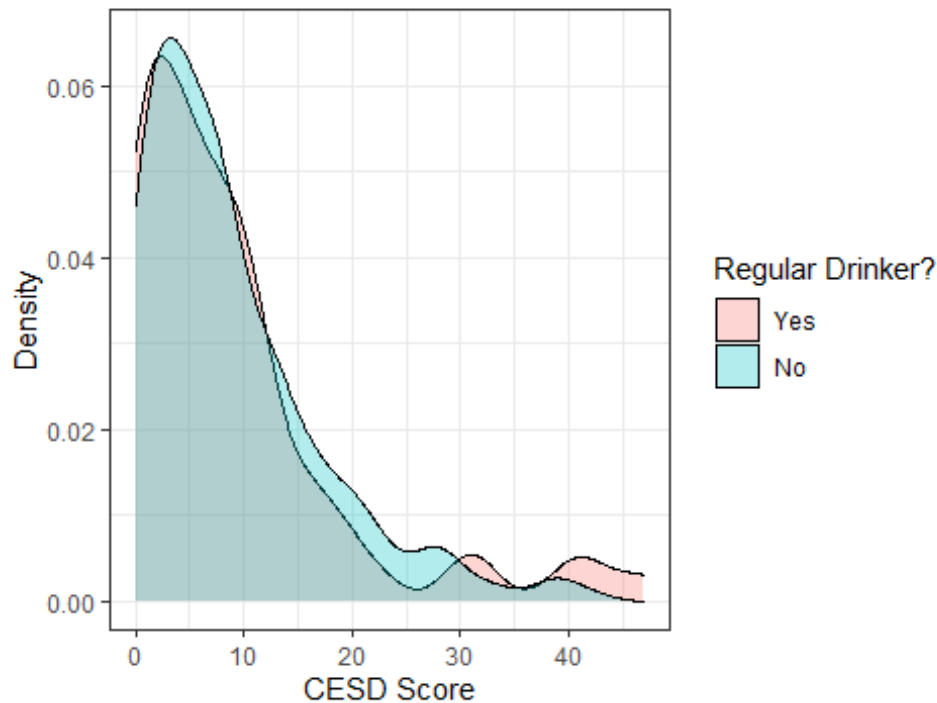


##Bivariate Comparison

CESD VS. DRINK First I will be comparing the variables 'cesd' and 'drink' using a density plot.

```
ggplot(Depression, aes(x=cesd, fill=drink_fac)) + geom_density(alpha=.3) +
scale_fill_discrete(name="Regular Drinker?") + ggtitle("Is There a
Relationship Between Drinking and Depression?") + xlab("CESD Score") +
ylab("Density") + theme_bw()
```

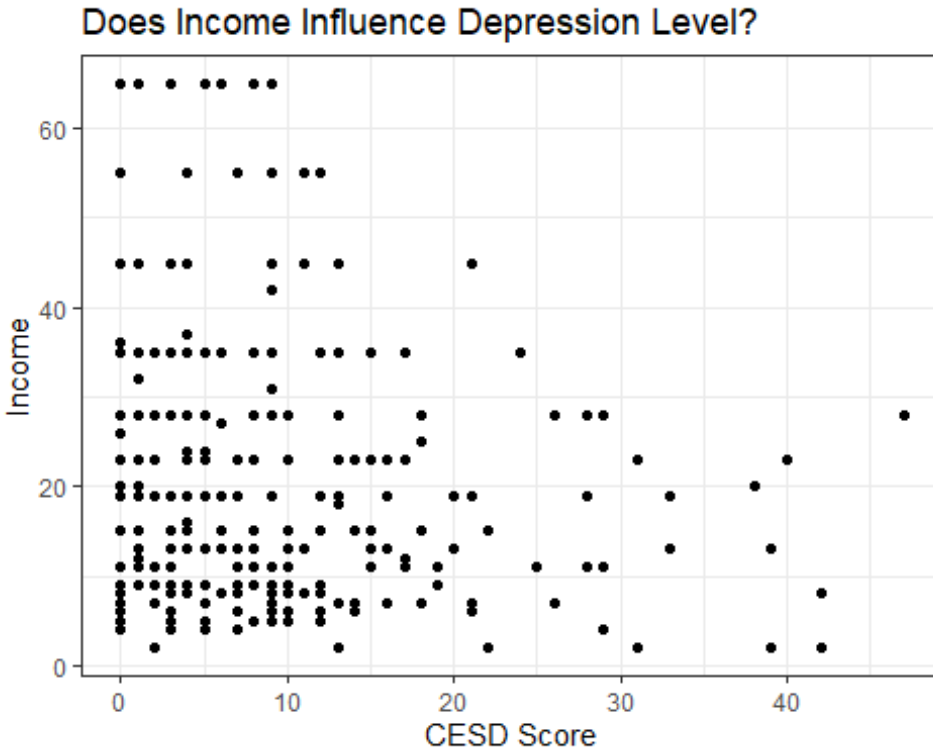
Is There a Relationship Between Drinking and Depression?



This density plot tells us that the study found no significant difference between the depression levels of those who drink regularly and those who don't drink regularly. The curves overlap very closely. The results are different than what I expected to see.

CESD VS. INCOME Next I will look at the relationship between income and depression score.

```
ggplot(Depression, aes(x=cesd, y=income)) + geom_point() + theme_bw() +  
ggtitle("Does Income Influence Depression Level?") + xlab("CESD Score") +  
ylab("Income")
```

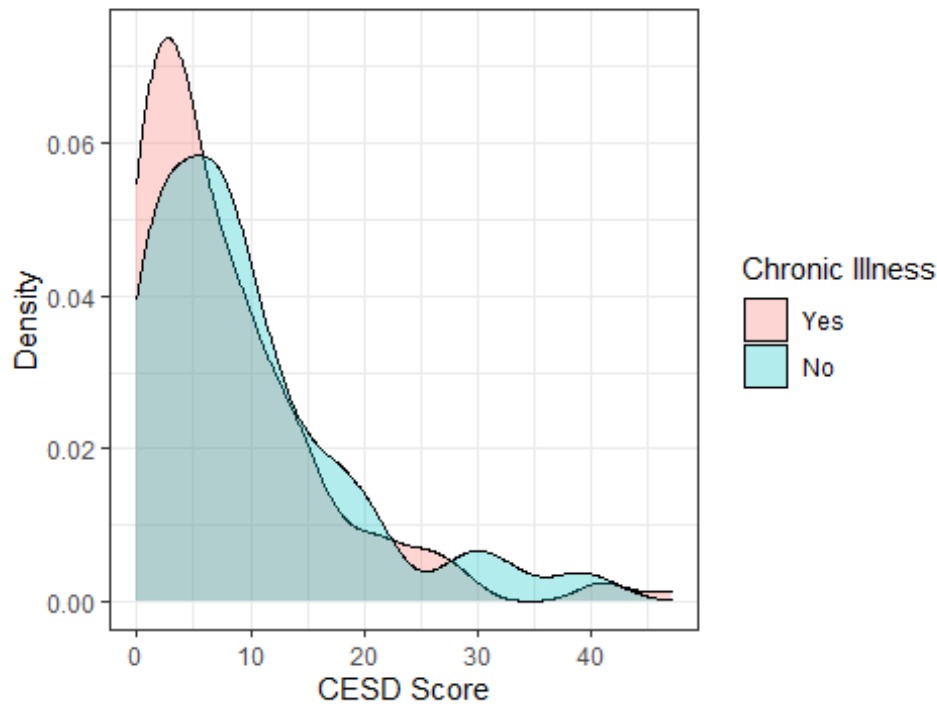


This dot plot shows us that people who scored the highest on the CESD scale do seem to have lower incomes. From this we can infer that lower income does play a role in depression.

CESD VS. CHRONIC ILLNESS

```
ggplot(Depression, aes(x=cesd, fill=chronill_fac)) + geom_density(alpha=.3) +
ggtitle("Does Chronic Illness Influence Depression?") + xlab("CESD Score") +
ylab("Density") + theme_bw() + scale_fill_discrete("Chronic Illness")
```

Does Chronic Illness Influence Depression?



This density plot tells us that there doesn't seem to be much of a difference in depression levels between those dealing with a chronic illness and those who are not. Chronic illness does not seem to play a significant role in influencing depression, at least from the results of this study.