Final Project - Depression

Matthew Solone

9/20/2020

Introduction

In this report I will be exploring the Depression data set which is a data set of 294 observations and 37 variables from various interviews from Los Angeles County. Specifically I will be looking into the relation between income, regular drinking, and chronic illness.

```
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(knitr)
depress <- read.table("~/documents/Math130/data/depression.txt", header=TRUE, sep="\t")
str(depress)
```

```
## 'data.frame':    294 obs. of  37 variables:
##  $ id      : int  1 2 3 4 5 6 7 8 9 10 ...
##  $ sex     : int  1 0 1 1 1 0 1 0 1 0 ...
##  $ age     : int  68 58 45 50 33 24 58 22 47 30 ...
##  $ marital : chr  "Widowed" "Divorced" "Married" "Divorced" ...
##  $ educat  : chr  "Some HS" "Some college" "HS Grad" "HS Grad" ...
##  $ employ  : chr  "Retired" "FT" "FT" "Unemp" ...
##  $ income  : int  4 15 28 9 35 11 11 9 23 35 ...
##  $ relig   : int  1 1 1 1 1 1 1 1 2 4 ...
##  $ c1      : int  0 0 0 0 0 0 2 0 0 0 ...
##  $ c2      : int  0 0 0 0 0 0 1 1 1 0 ...
##  $ c3      : int  0 1 0 0 0 0 1 2 1 0 ...
##  $ c4      : int  0 0 0 0 0 0 2 0 0 0 ...
##  $ c5      : int  0 0 1 1 0 0 1 2 0 0 ...
##  $ c6      : int  0 0 0 1 0 0 0 1 3 0 ...
##  $ c7      : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ c8      : int  0 0 0 3 3 0 2 0 0 0 ...
##  $ c9      : int  0 0 0 0 3 1 2 0 0 0 ...
##  $ c10     : int  0 0 0 0 0 0 0 0 0 0 ...
##  $ c11     : int  0 0 0 0 0 0 0 0 0 0 ...
```

```
##  $ c12     : int  0 1 0 0 0 1 0 0 3 0 ...
##  $ c13     : int  0 0 0 0 0 2 0 0 0 0 ...
##  $ c14     : int  0 0 1 0 0 0 0 0 3 0 ...
##  $ c15     : int  0 1 1 0 0 0 3 0 2 0 ...
##  $ c16     : int  0 0 1 0 0 2 0 1 3 0 ...
##  $ c17     : int  0 1 0 0 0 1 0 1 0 0 ...
##  $ c18     : int  0 0 0 0 0 0 0 1 0 0 ...
##  $ c19     : int  0 0 0 0 0 0 0 1 0 0 ...
##  $ c20     : int  0 0 0 0 0 0 1 0 0 0 ...
##  $ cesd    : int  0 4 4 5 6 7 15 10 16 0 ...
##  $ cases   : int  0 0 0 0 0 0 0 0 1 0 ...
##  $ drink   : int  0 1 1 0 1 1 0 0 1 1 ...
##  $ health  : int  2 1 2 1 1 1 3 1 4 1 ...
##  $ regdoc  : int  1 1 1 1 1 1 1 0 1 1 ...
##  $ treat   : int  1 1 1 0 1 1 1 0 1 0 ...
##  $ beddays : int  0 0 0 0 1 0 0 0 1 0 ...
##  $ acuteill: int  0 0 0 0 1 1 1 1 0 0 ...
##  $ chronill: int  1 1 0 1 0 1 1 0 1 0 ...
```

```
head(depress)
```

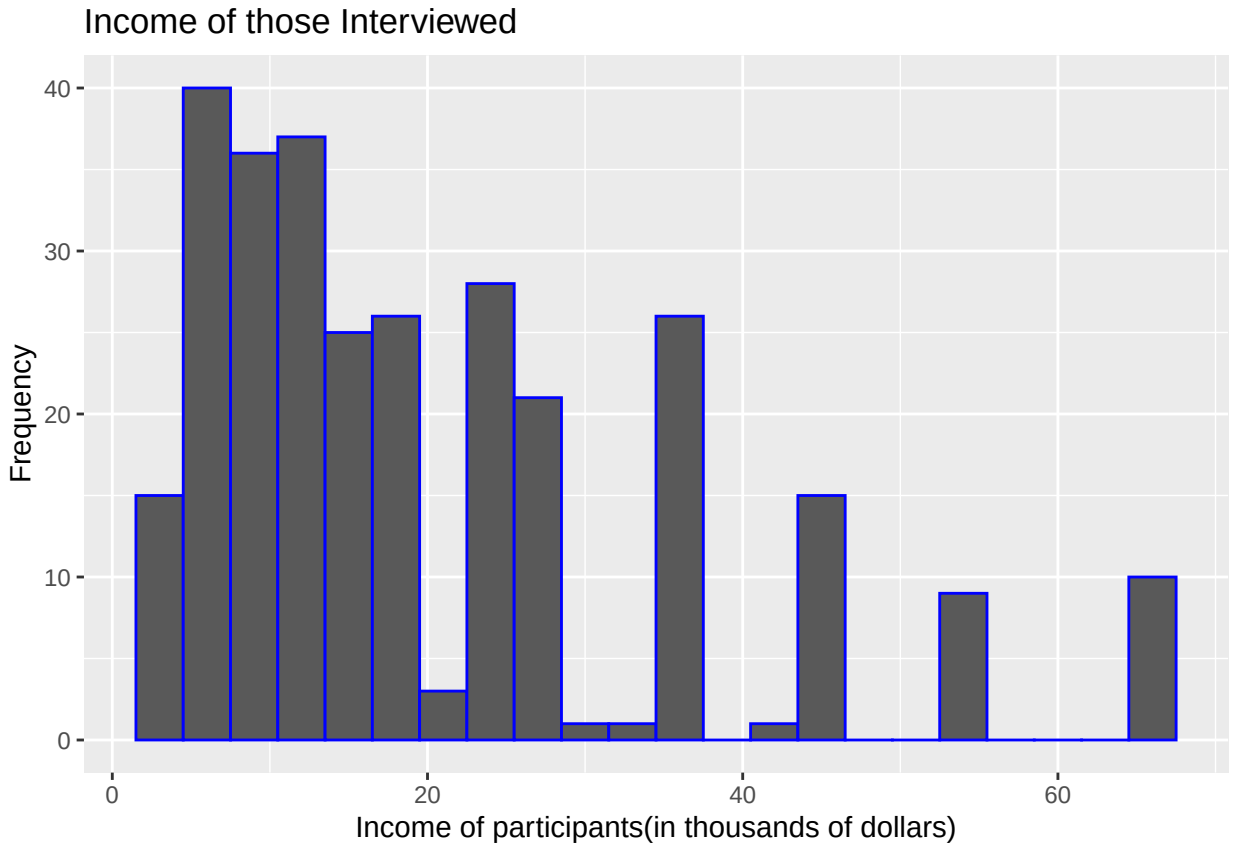
```
##   id sex age   marital       educat  employ income relig c1 c2 c3 c4 c5 c6 c7
## 1  1   1  68   Widowed      Some HS Retired      4     1  0  0  0  0  0  0  0
## 2  2   0  58  Divorced Some college      FT     15     1  0  0  1  0  0  0  0
## 3  3   1  45   Married      HS Grad      FT     28     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
```

Univariate description

Income

First I will explore the income of those interviewed.

```
ggplot(depress, aes(x=income))+geom_histogram(col = "blue", binwidth=3)+xlab("Income of participants(in
```




```
summary(depress$income)
```

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

From the data we can see that the majority(75%) of the participants have incomes less that $28,000, with 25% of the participants under the US poverty line. The data is also skewed to the right giving us the same idea from the histogram.

Regular Drinking

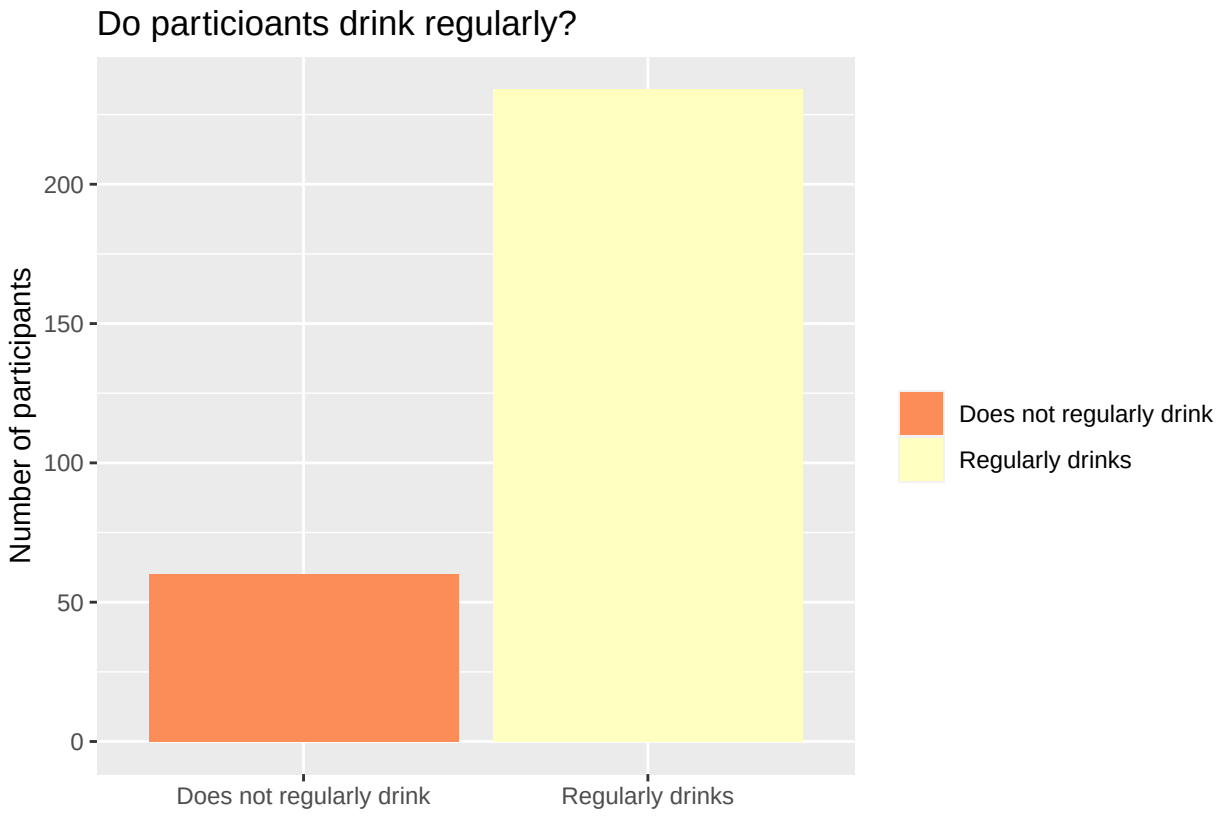
Now we will examine the data collected on whether or not our participants regularly drink. According to the code book: 1=Yes 2=No

```
depress$drink <- factor(depress$drink, labels=c("Does not regularly drink","Regularly drinks"))
depress$drink%>%table()%>%prop.table()
```

```
## .
## Does not regularly drink         Regularly drinks
##                0.2040816                0.7959184
```

From here we see that around 79.6% of the participant drink regularly.

```
ggplot(depress, aes(x=drink, fill=drink))+geom_bar()+scale_fill_brewer(palette = "Spectral",name=" ")+x
```



So as previously stated a vast majority of the people interviewed admit to drinking regularly. However does this have nay correlation with chronic illness or income?

Chronic Illness

Now we will be exploring the variable Chronic illness in which the participants answered if they have had a chronic illness in the last year.

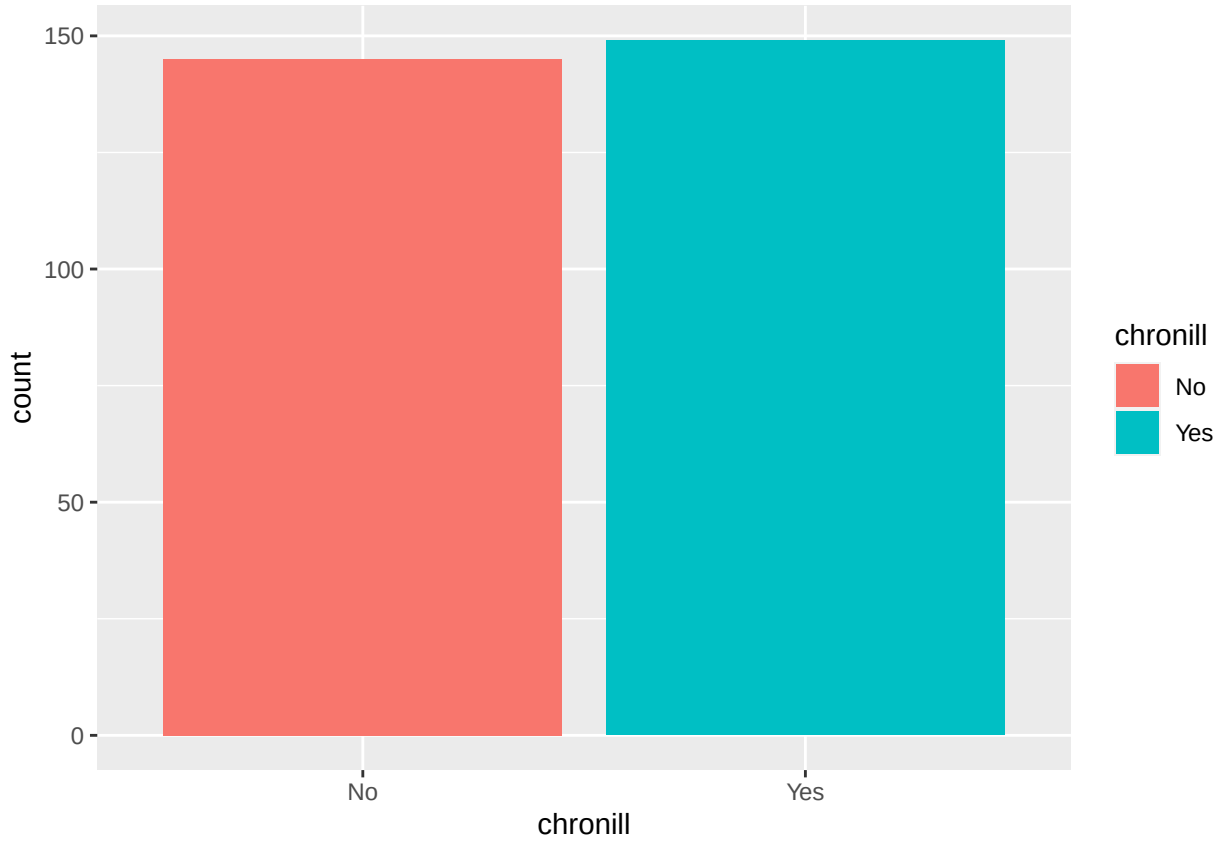
```
depress$chronill <- factor(depress$chronill, labels=c("No","Yes"))
table(depress$chronill)
```

```
##
##  No Yes
## 145 149
```

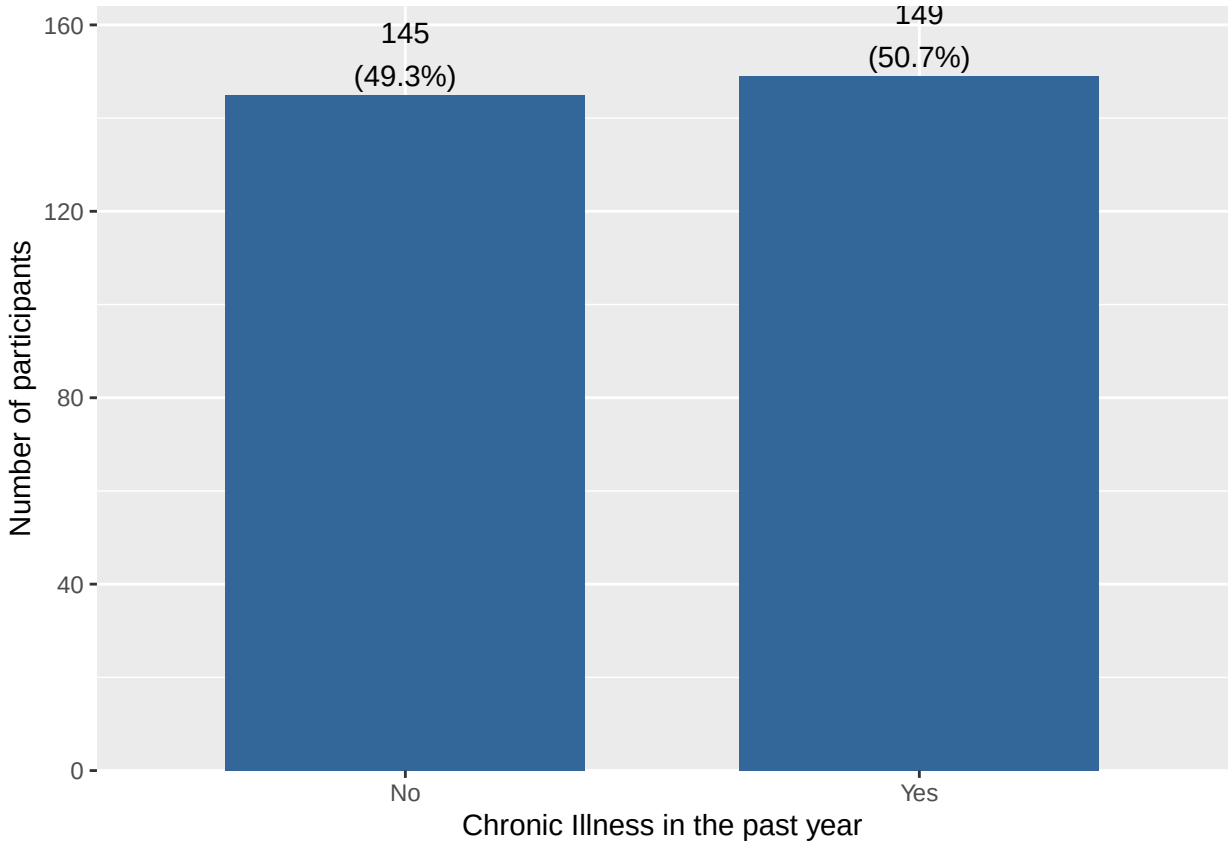
```
prop.table(table(depress$chronill))
```

```
##
##        No       Yes
## 0.4931973 0.5068027
```

```
ggplot(depress, aes(x=chronill,fill=chronill))+geom_bar()
```




```
library(sjPlot)
plot_frq(depress$chronill)+xlab("Chronic Illness in the past year")+ylab("Number of participants")
```

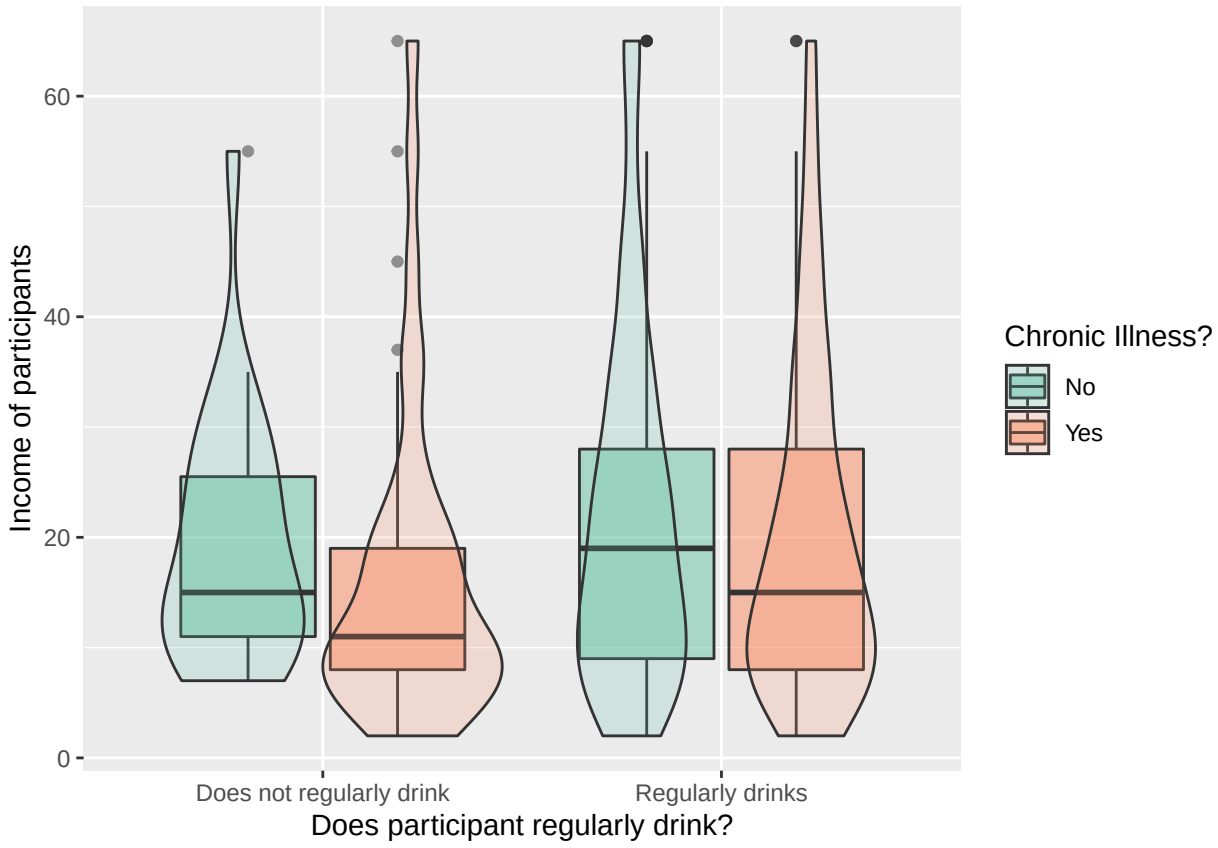


According to the bar plot and the proportion table it seems like the spread of the participants is equal. The slightly higher value is assigned to yes, they have had a chronic illness in the past year.

Multi Variate

Chronic Illness vs Regular drinker vs Income

```
ggplot(depress, aes(y=income, x=drink, fill=chronill)) + geom_boxplot(alpha=.5)+geom_violin(alpha=.2)+s
```



In conclusion, among the people who regularly drink and those who not drink regularly have very similar trends. The biggest difference is among those who do not regularly drink and have had a chronic illness in the past year. 75% of them with less than $20,000. So from these tests there seems to be some correlation between NOT drinking regularly, chronic illness, and lower income based on this data.