Math 130 Exploratory Data Analysis Project

Kyle Dobson

2/25/2022

```
library(ggplot2)
depression <-read.table("C:/Users/kyled/OneDrive/Documents/MATH130/data/depress_081217.txt", header=TRU
```

Introduction

The depression data set is from the first set of interviews of a prospective study of depression in the adult residents of Los Angeles County and includes 294 observations. This survey contains questions based simply on several demographics along with lifestyle choices.The analysis will consist of three main varibales of interest, they include: CESD and INCOME. The goal is to determine if there is a correlation between Income status and its effect on someone's level of depression.
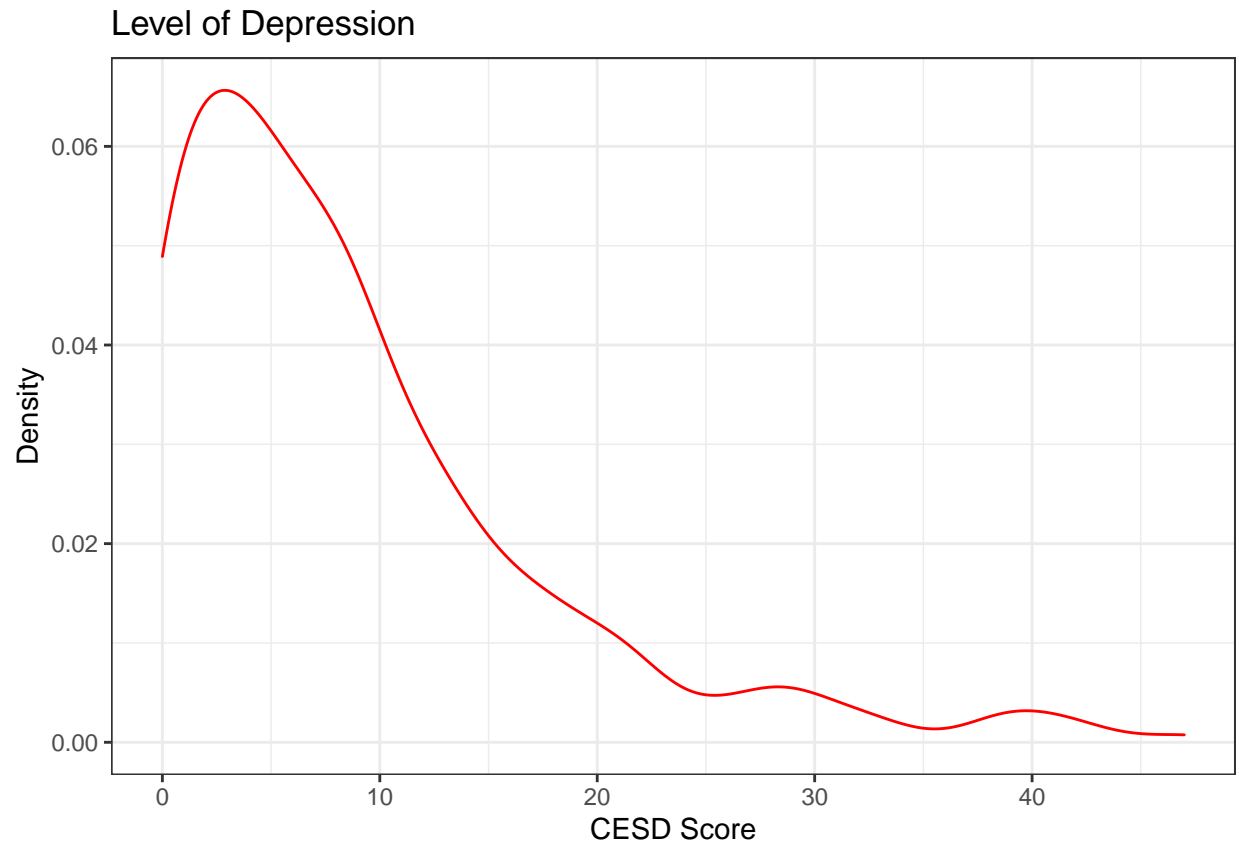
Univariate Explanation

Variable One: CESD

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

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

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

Level of Depression

The variable CESD describes the level of depression for each participant, 0 being the lowest and 60 being the highest.
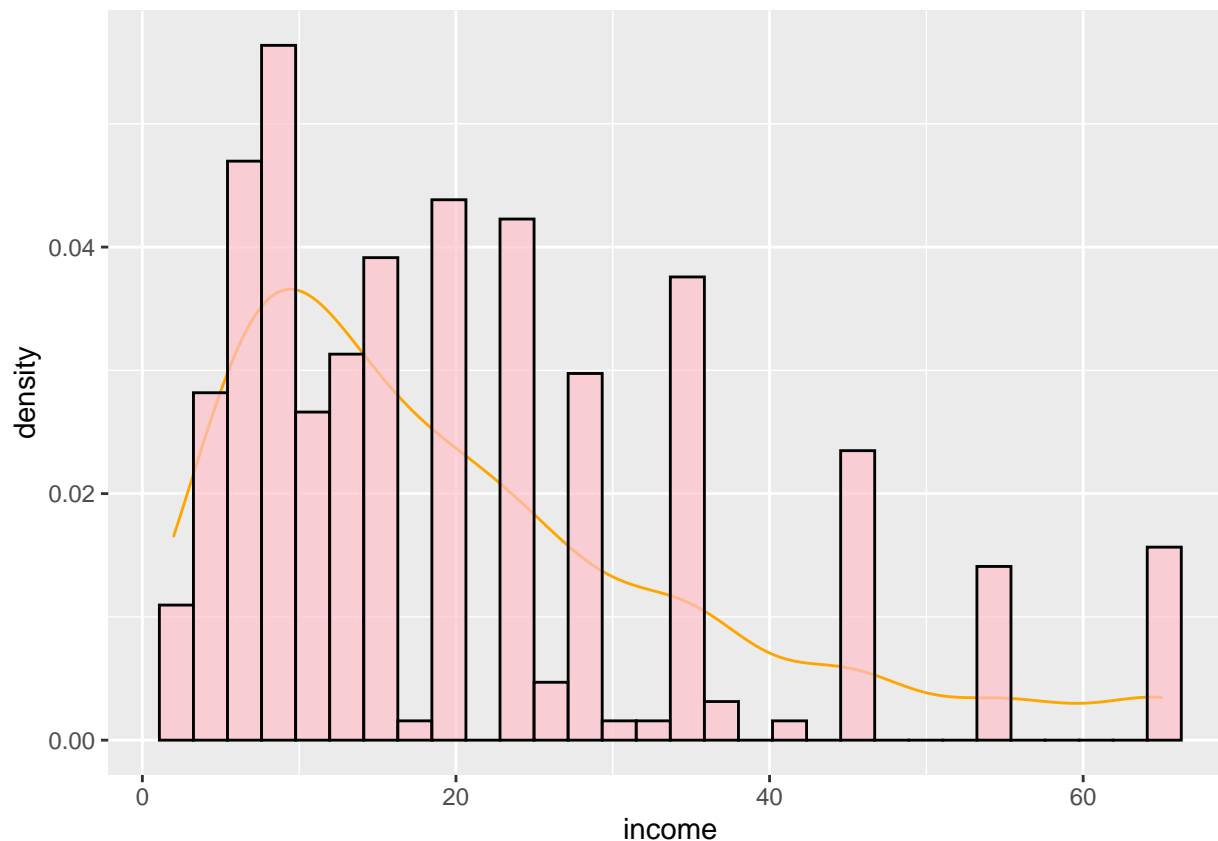
Variable 2: Income

```
summary(depression$income)
```

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

```
ggplot(depression, aes(x=income)) + geom_density(col="orange") +
geom_histogram(aes(y=..density..), colour="black", fill="pink", alpha = 0.7)
```

```
## `stat_bin()` using `bins = 30`. Pick better value with `binwidth`.
```

This variable describes the income of each participant in the survey.The income variable has a range from 2 to 65. Meaning the lowest income is $2,000 and the highest income is $65,000 and above.
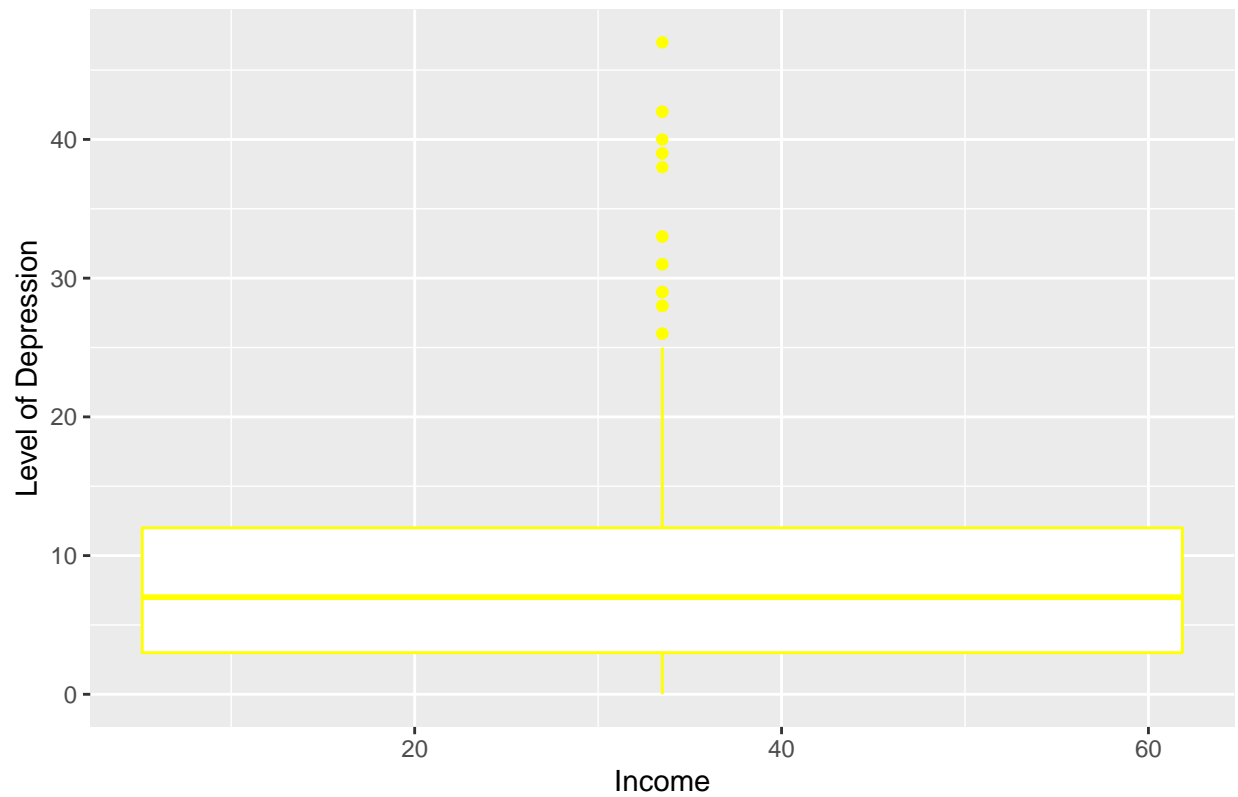
Bivariate Exploration

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

```
## Warning: Continuous x aesthetic -- did you forget aes(group=...)?
```

```
## Warning: It is deprecated to specify 'guide = FALSE' to remove a guide. Please
## use 'guide = "none"' instead.
```

Relationship Between Income and Depression Level


```
table(depression$income,depression$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 26 28 29
##   2  0 0 1 0 0 0 0 0 0 0  0  0  0  1  0  0  0  0  0  0  0  2  0  0  0  0  0
##   4  2 0 0 2 0 2 0 1 0 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  1
##   5  2 0 0 1 0 1 0 0 2 2  1  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   6  2 0 0 2 0 0 0 0 1 0  1  1  0  2  0  1  0  0  0  0  0  2  0  0  0  0  0
##   7  3 0 3 0 0 1 0 0 0 1  0  0  0  2  2  0  2  0  2  0  0  1  0  0  0  1  0  0
##   8  1 0 0 1 2 0 4 1 0 1  1  1  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   9  2 2 1 4 1 3 0 2 1 1  2  0  1  0  0  0  0  0  0  2  0  0  0  0  0  0  0
##   11 1 1 2 1 0 0 0 3 1 1  1  0  0  0  0  1  0  1  0  1  0  0  0  1  0  1  1
##   12 0 1 0 0 0 0 0 0 0 0  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0  0
##   13 0 1 0 1 2 2 1 1 2 0  1  2  0  0  0  1  1  0  0  0  1  0  0  0  0  0  0
##   15 1 2 0 2 4 0 2 0 4 0  1  0  2  0  1  1  0  0  2  0  0  0  2  0  0  0  0
##   16 0 0 0 0 1 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   18 0 0 0 0 0 0 0 0 0 0  0  0  0  1  0  0  0  0  0  0  0  0  0  0  0  0  0
##   19 2 2 1 2 2 2 4 2 0 1  0  0  1  1  0  0  1  0  0  0  1  1  0  0  0  1  0
##   20 1 1 0 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   23 3 1 1 0 1 2 0 4 3 0  2  0  0  1  1  2  1  1  0  0  0  0  0  0  0  0  0
##   24 0 0 0 0 1 1 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   25 0 0 0 0 0 0 0 0 0 0  0  0  0  0  0  0  0  1  0  0  0  0  0  0  0  0  0
##   26 1 0 0 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   27 0 0 0 0 0 0 1 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##   28 3 1 2 1 1 1 0 0 1 1  1  0  0  2  0  0  0  0  1  0  0  0  0  0  1  1  1
```

```
##    31 0 0 0 0 0 0 0 0 0 0 1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##    32 0 1 0 0 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##    35 4 3 2 3 1 1 2 0 2 1  0  0  1  1  0  1  0  1  0  0  0  0  0  1  0  0  0  0
##    36 1 0 0 0 0 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##    37 0 0 0 0 1 0 0 0 0 0 0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##    42 0 0 0 0 0 0 0 0 0 0 1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##    45 3 3 0 3 1 0 0 0 0 1  0  2  0  1  0  0  0  0  0  0  1  0  0  0  0  0  0
##    55 1 0 0 0 2 0 0 1 0 2  0  2  1  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##    65 1 1 0 2 0 2 1 0 1 2  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0  0
##
##       31 33 38 39 40 42 47
## 2     1  0  0  1  0  1  0
## 4     0  0  0  0  0  0  0
## 5     0  0  0  0  0  0  0
## 6     0  0  0  0  0  0  0
## 7     0  0  0  0  0  0  0
## 8     0  0  0  0  0  1  0
## 9     0  0  0  0  0  0  0
## 11    0  0  0  0  0  0  0
## 12    0  0  0  0  0  0  0
## 13    0  1  0  1  0  0  0
## 15    0  0  0  0  0  0  0
## 16    0  0  0  0  0  0  0
## 18    0  0  0  0  0  0  0
## 19    0  1  0  0  0  0  0
## 20    0  0  1  0  0  0  0
## 23    1  0  0  0  1  0  0
## 24    0  0  0  0  0  0  0
## 25    0  0  0  0  0  0  0
## 26    0  0  0  0  0  0  0
## 27    0  0  0  0  0  0  0
## 28    0  0  0  0  0  0  1
## 31    0  0  0  0  0  0  0
## 32    0  0  0  0  0  0  0
## 35    0  0  0  0  0  0  0
## 36    0  0  0  0  0  0  0
## 37    0  0  0  0  0  0  0
## 42    0  0  0  0  0  0  0
## 45    0  0  0  0  0  0  0
## 55    0  0  0  0  0  0  0
## 65    0  0  0  0  0  0  0
```

From this boxplot and the table it can determined that majority of the participants gave a depression level rating between 0 and 15 showing intially that majority of the group was not heavily depressed. Furthermore, the income axis shows a large spike of higher depression levels between the incomes 30 and 40(translating to $30,000 and $40,000 in real life) showing a higher majority of people in this income specifically suffer more from depression.The less money you make a year does not show to have an effect on a participants level of depression.

Conclusion

My inital thought that there may be a correlation between income and level of depression has not been seen in this data as the participants that answered the highest rating of depression all trended towards the same

yearly income, leading me to believe there was is no direct correlation from this data.