eda.dabeaszepeda

Diego Beas Zepeda

2022-09-28

```
depress <- read.delim("C:/Users/camin/OneDrive/Desktop/Math 130/Data/depress_081217.txt",
                      header=TRUE,sep="\t")
library(ggplot2)
library(dplyr)
library(forcats)
library(knitr)
library(sjPlot)
```

```
## Warning: package 'ggplot2' was built under R version 4.1.2
```

```
## Warning: package 'dplyr' was built under R version 4.1.2
```

```
##
## 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
```

```
## Warning: package 'knitr' was built under R version 4.1.2
```

I will be looking at the Depression data set this study was done interviewing Los Angeles County adult residents were conducted. There are 294 observations and 37 variables in the data set. I will be investigating the link between depression, education and income. I anticipate a poor link between despression and education. and a fair link between depression and income I know that a person's mental health can be impacted by a variety of factors but i will be look at these three

Univariate Exploration

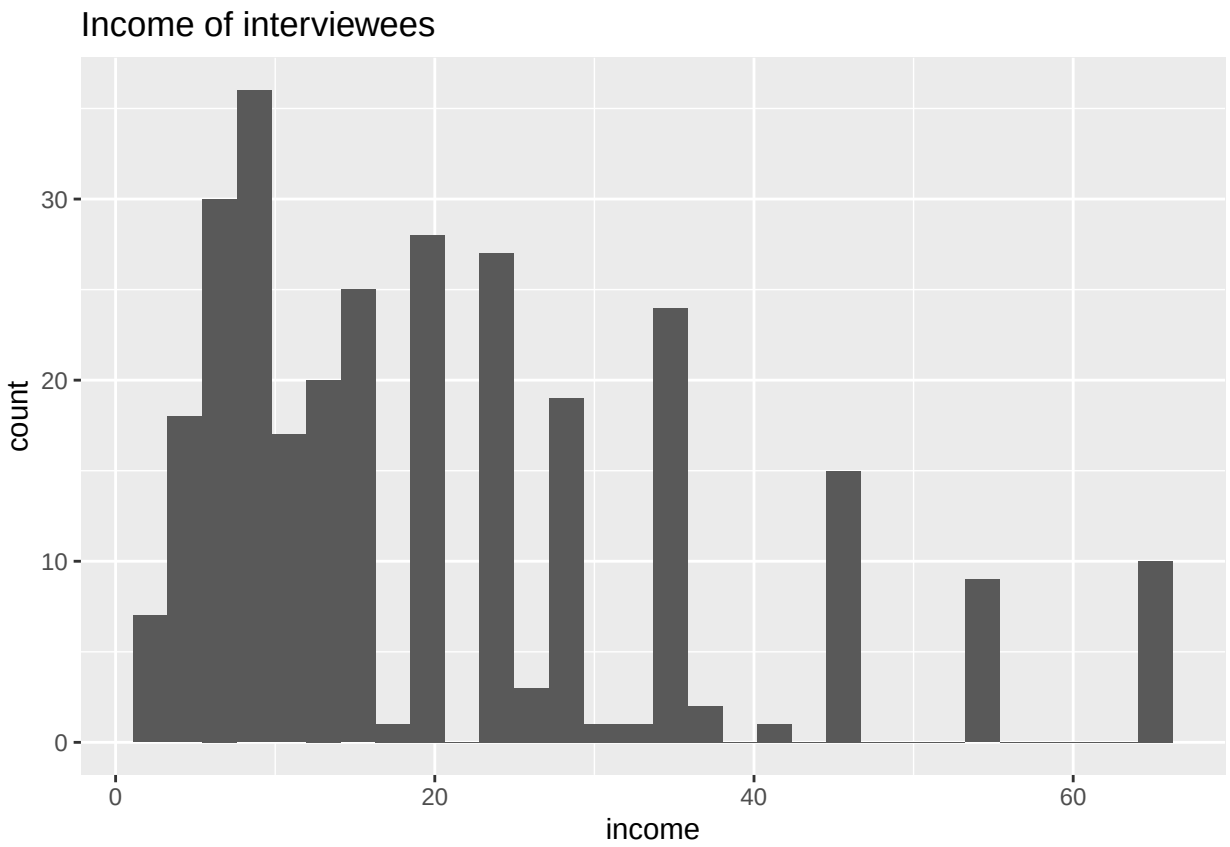
Income

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

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

```
ggplot(depress, aes(x=income)) + geom_histogram() + ggtitle("Income of interviewees")
```

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



(note i do not know how much each units standsfor) This summary shows the distribution of average income among the inerviewee in this study. Average income being 20 units why the highest was 65 and lowest being 2 units

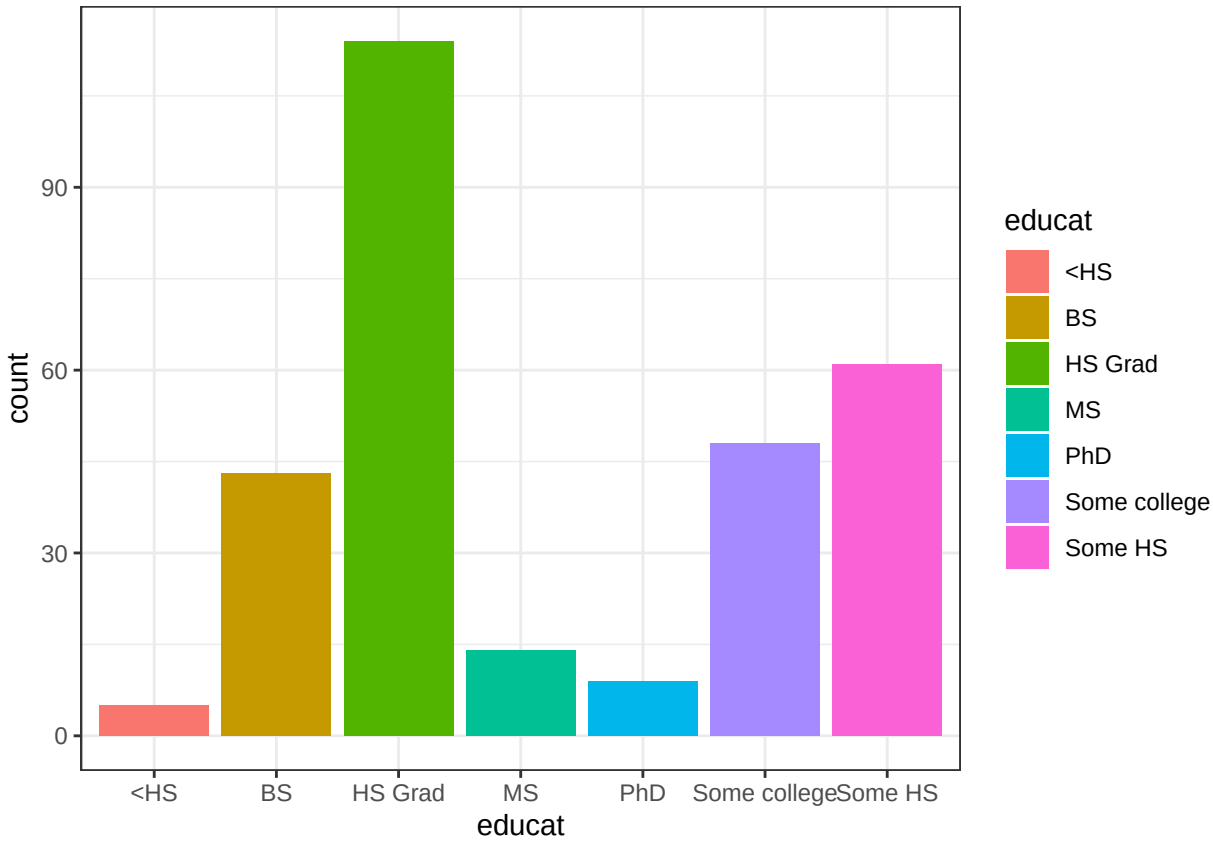
The histogram exhibits the income distribution of the participants in the study. The graph shows its skewed to the right with a median of 15 and a average of 20 it also shows a much more dense coverage in lower incomes.

The table shows the number of participants and their education

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

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

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



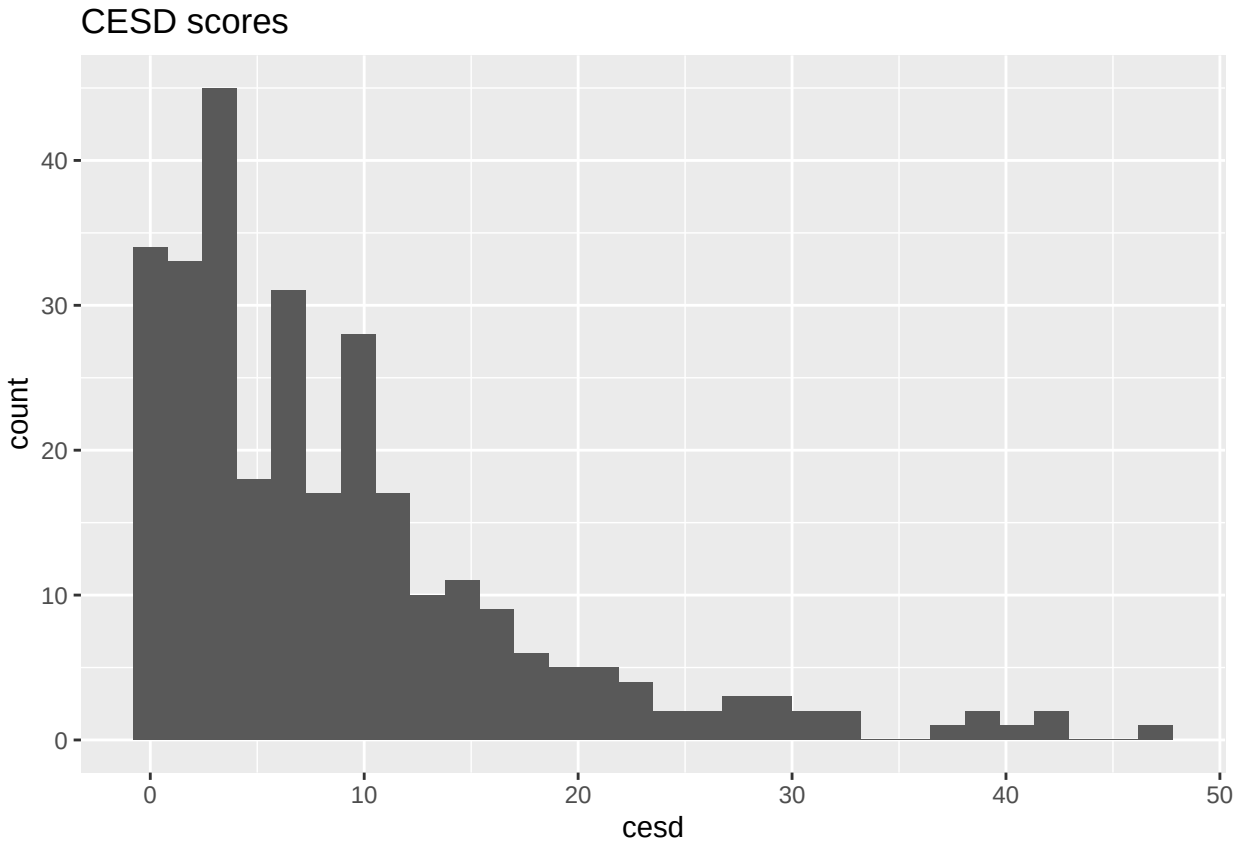
THis bar graph shows whats the education level of the interviewees with the biggest group atleast finishing high school with while other groups like BS, some colleg and some HS not being so far behind.

```
summary(depress$cesd)
```

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

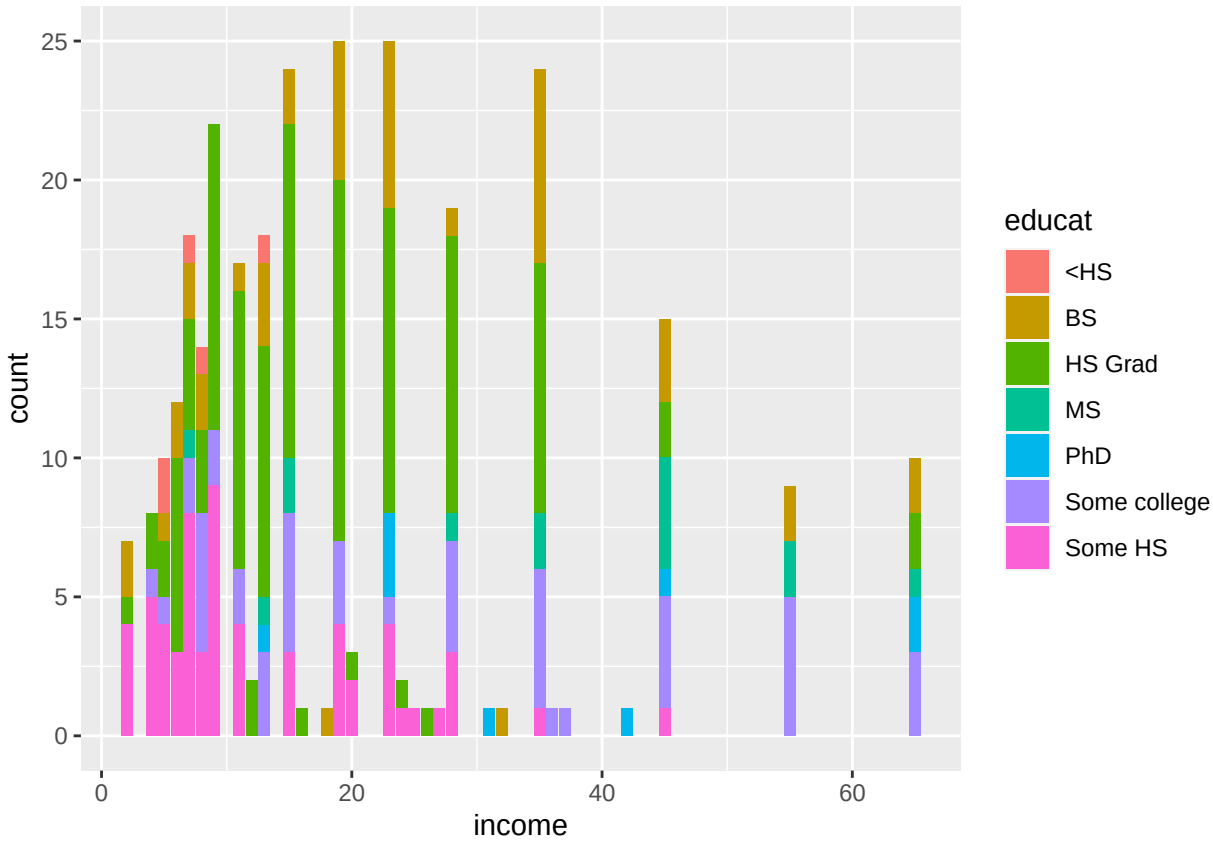
```
ggplot(depress, aes(x=cesd)) + geom_histogram() + ggtitle("CESD scores")
```

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



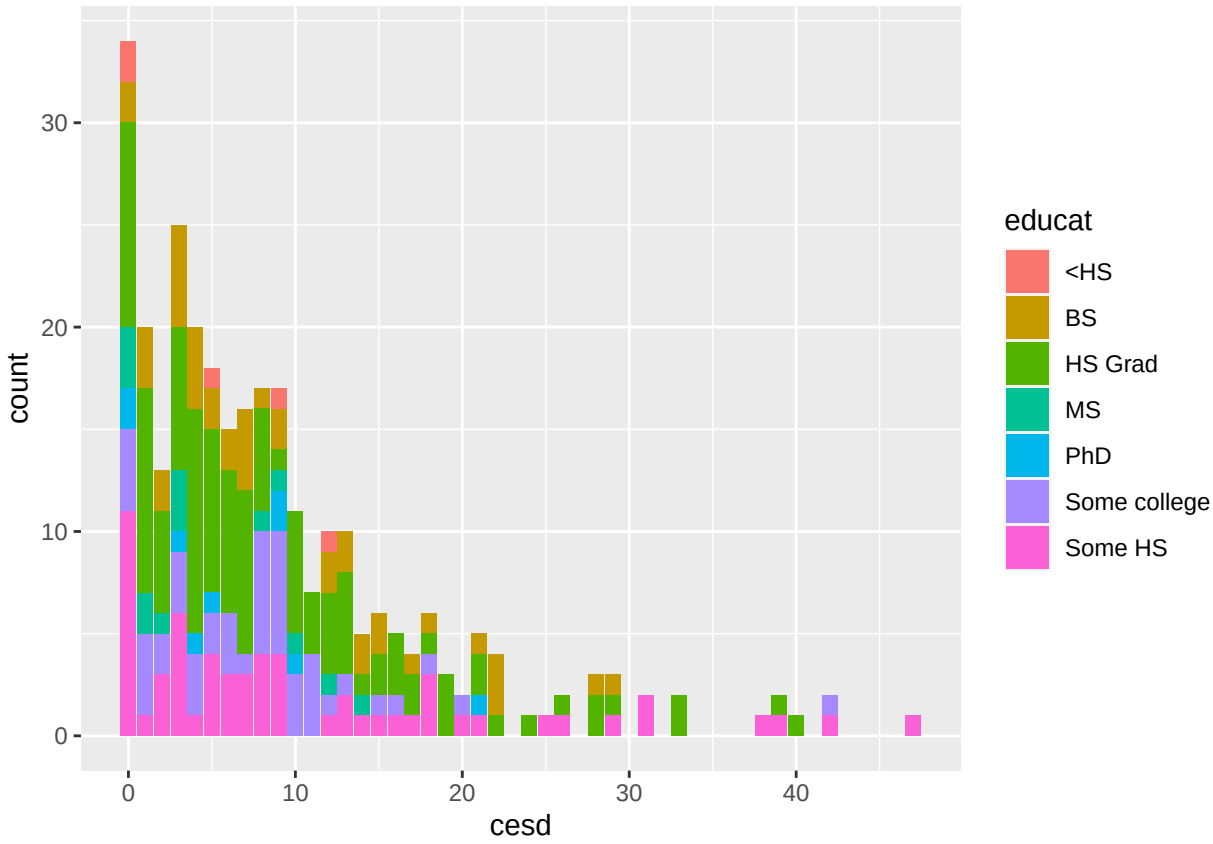
This table shows how many of the interviewees CESD scores which measures level of depression 0 being the lowesr and 60 being the highest

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



This graph is very similar to the firat graph i showed this does a much better job at at spliting up the different incomes to also show which level of education tend to have a higher levels of income compared to the other

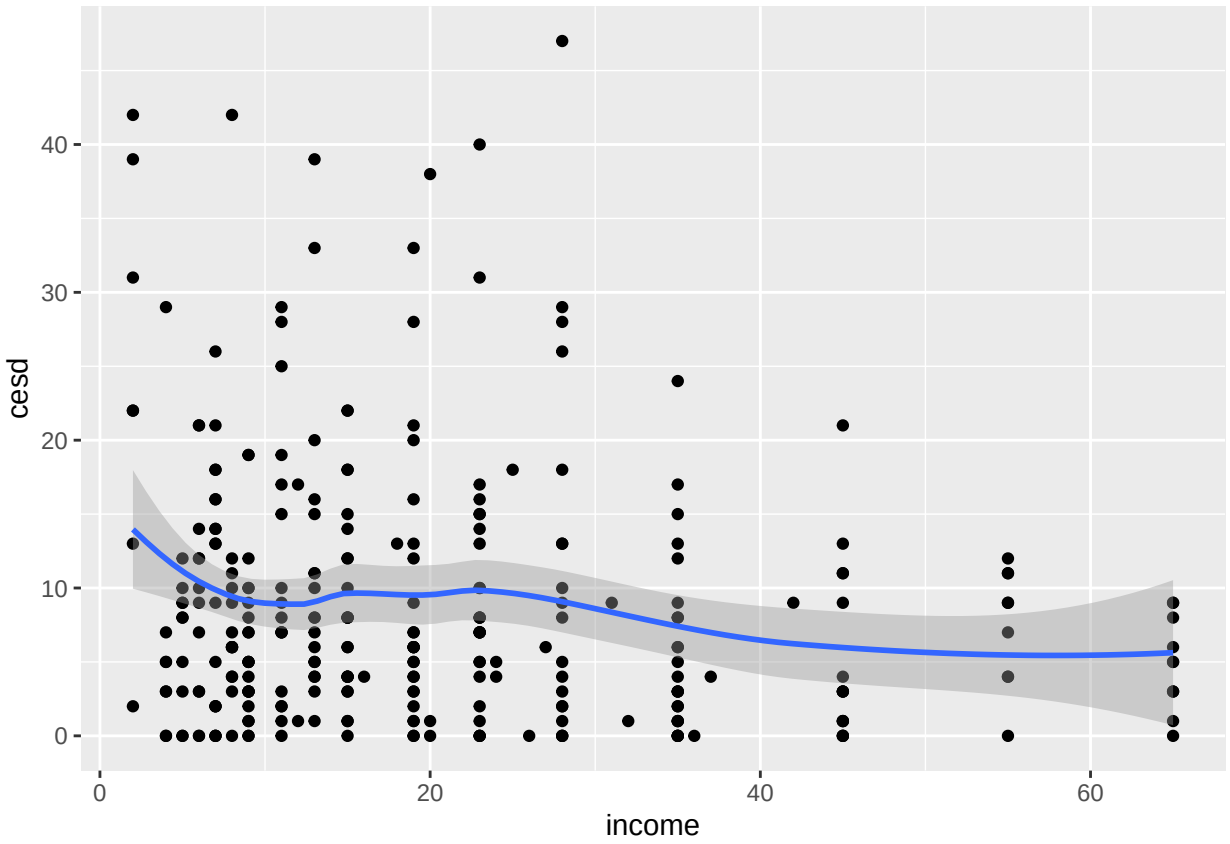
```
ggplot(depress, aes(x=cesd , fill=educat)) + geom_bar()
```



This Bar graph shows Cesd scores which measure level of depression and education level though we can't get much information we can make very interesting observations such as most of the interviews with high CESD scores never seemed to have finish college and have either had some college, were a HS Graduate, or only had some high school eduction

```
ggplot(depress, aes(x=income, y=cesd)) + geom_point()+ geom_smooth()
```

```
## `geom_smooth()` using method = 'loess' and formula 'y ~ x'
```



This graph shoes the relationship between income levels and CESD scores. Though there doesn't seem to be a very strong coraltion between the two it does seem that higher incomes have lower CESD scores than lower incomes

Conclusion I wanted to compare how different educational and income levels related to CESD scores or level of depression I would have to say my hypothesis was right as the coralation with depression and education was barely noticible as most of the higher CESD scores were people with lower education but overall hard to tell. While the income to CESD graph was much easier to see a relationship though it was still a very weak one as the the slope of the line didnt go down much