FInal Project for 130

Adrian Sanchez

2023-09-15

```
## Install package "strengejacke" from GitHub ('devtools::install_github("strengejacke/strengejacke")')
```

```
##
## 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
```

R Markdown

```
depress <- read.delim("C:/Users/Adrian Sanchez/Desktop/MATH130/NOTES/depress_081217.txt",
    header = TRUE, sep = "\t")
dim(depress)
```

```
## [1] 294  37
```

Introduction

Hello, today I am going to do some cross-analyzes with this data set about depression This data set comes from Los Angeles County and it inculdes 294 observations of people who where interviewed about depression. I am going to be using the variables,Age,Education,cesd and Income to show if there is any correlation bewtween them that can be linked to depression status. I want to infer if I can see any type of relationship bewteen these varibles and draw a conclusion that depression has a part to play with those variables

Univariate Exploration

I first want too examine the variable Educaiton that shows what level of schooling that they completed. It ranges from the subjects havingLess than highschool completed too completeing a doctoral degree. There are 294 subjects in total and most of the seem to have at least completed at least highschool.

```
table(depress$Educational)
```

```
## < table of extent 0 >
```

```
depress$Educational <- factor(depress$educat, levels = c("<HS", "Some HS", "HS Grad",
    "Some college", "BS", "MS", "PhD"))
plot_frq(depress$Educational, geom.colors = "blue") + xlab("Education Level")
```
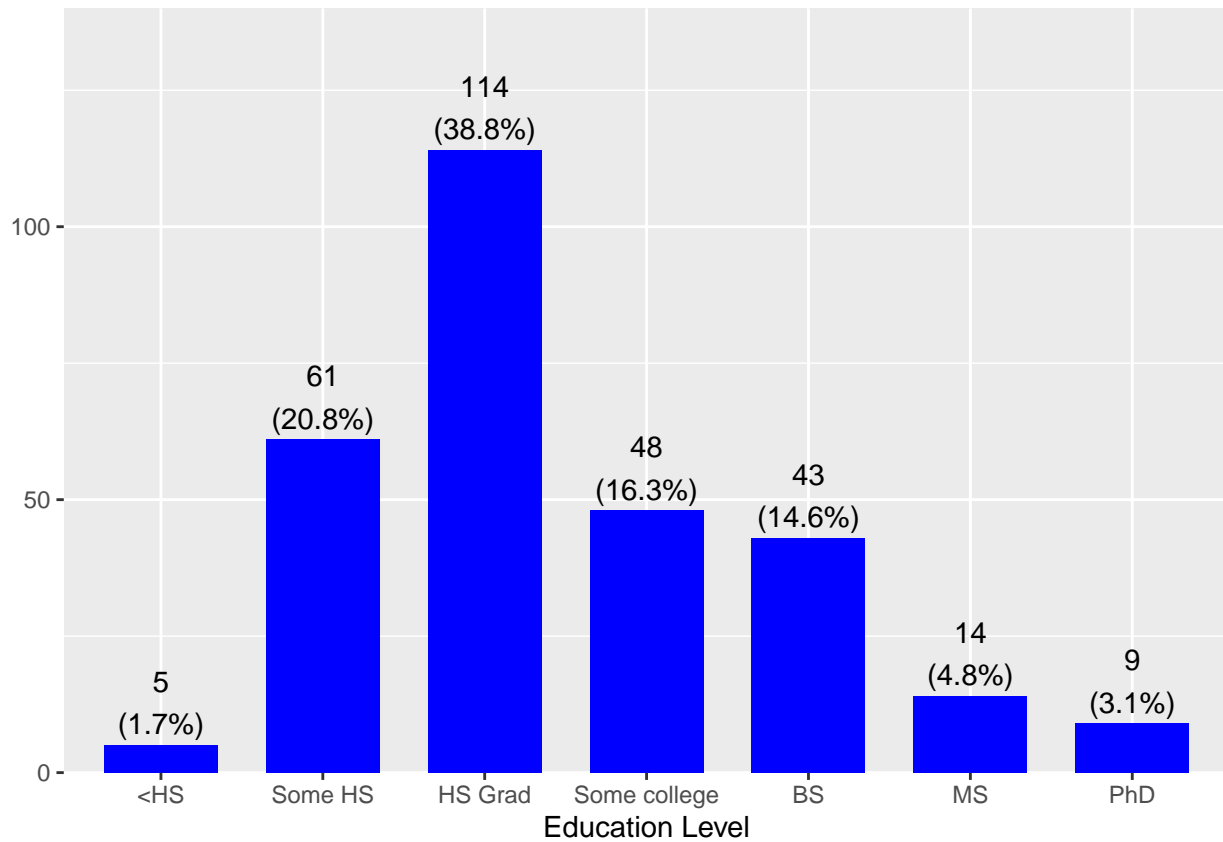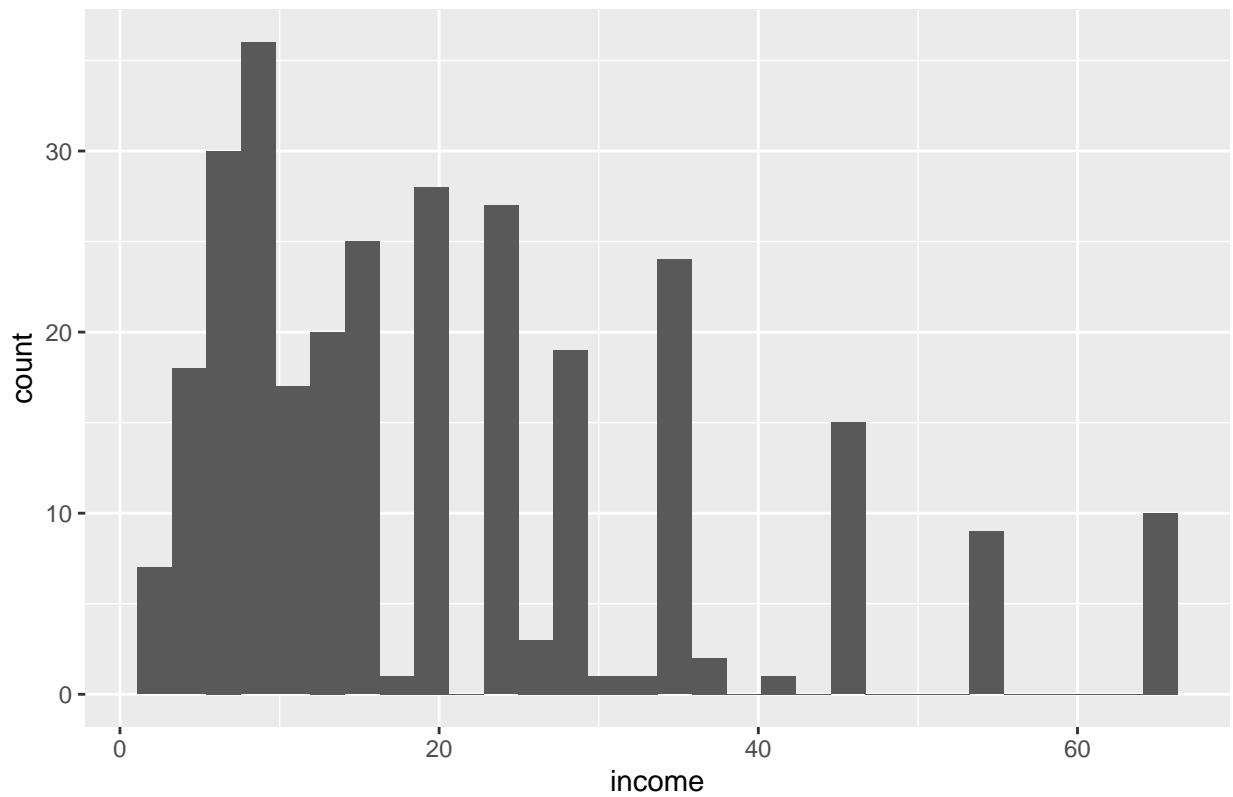


```
ggplot(depress, aes(x = income)) + geom_histogram() + ggtitle("Income in thousands for particpants")
```

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

Income in thousands for particpants

According to my frequencey table it appears that most of the people who were interviewed were Hs grads

```
ggplot(depress, aes(x = income)) + geom_bar() + ggtitle("Income of particpants in  thousands")
```

Income of particpants in thousands
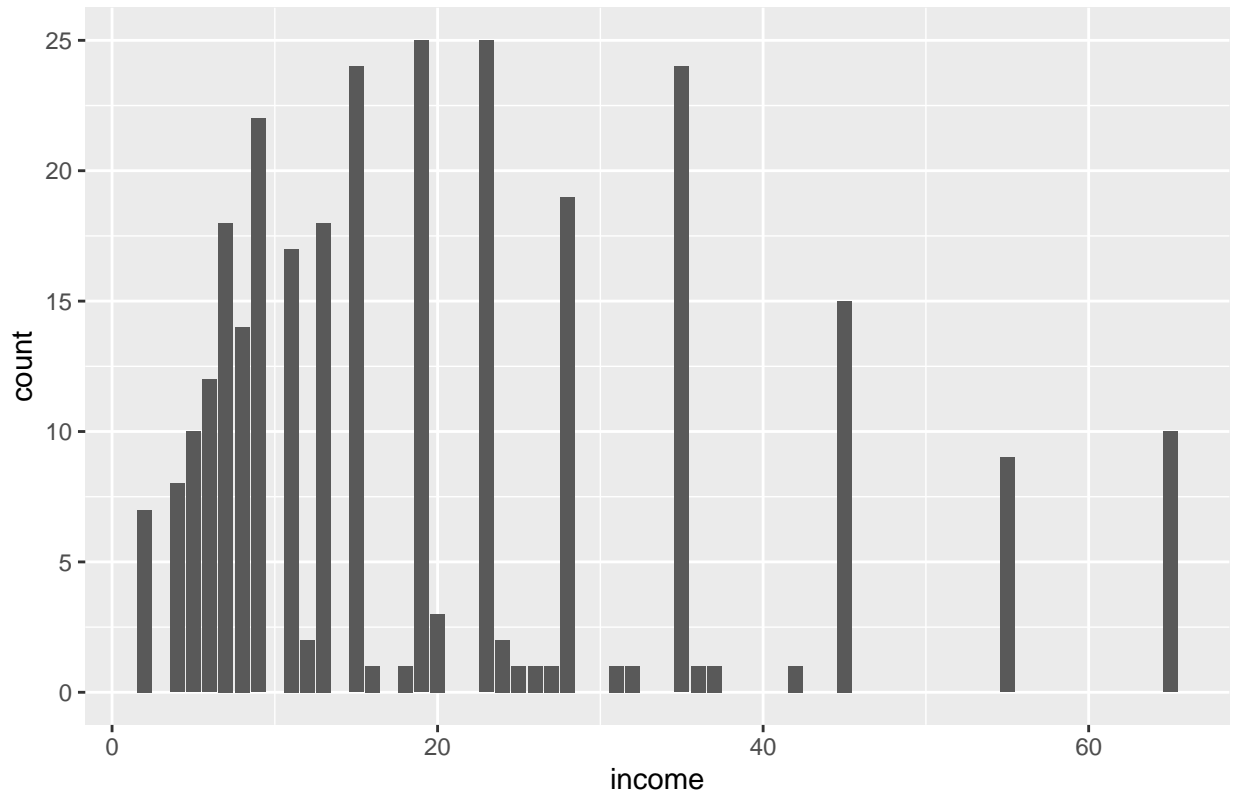
Income of particpants in the Survey and it seems most of them make less 20 thousands a year

```
ggplot(depress, aes(x = age)) + geom_bar() + ggtitle("Age of people in Survey")
```

Age of people in Survey

Age of the particpants in the survey and seems most of them are between the 20 to 30 year old range.

```
ggplot(depress, aes(cesd)) + geom_bar() + ggtitle("")
```
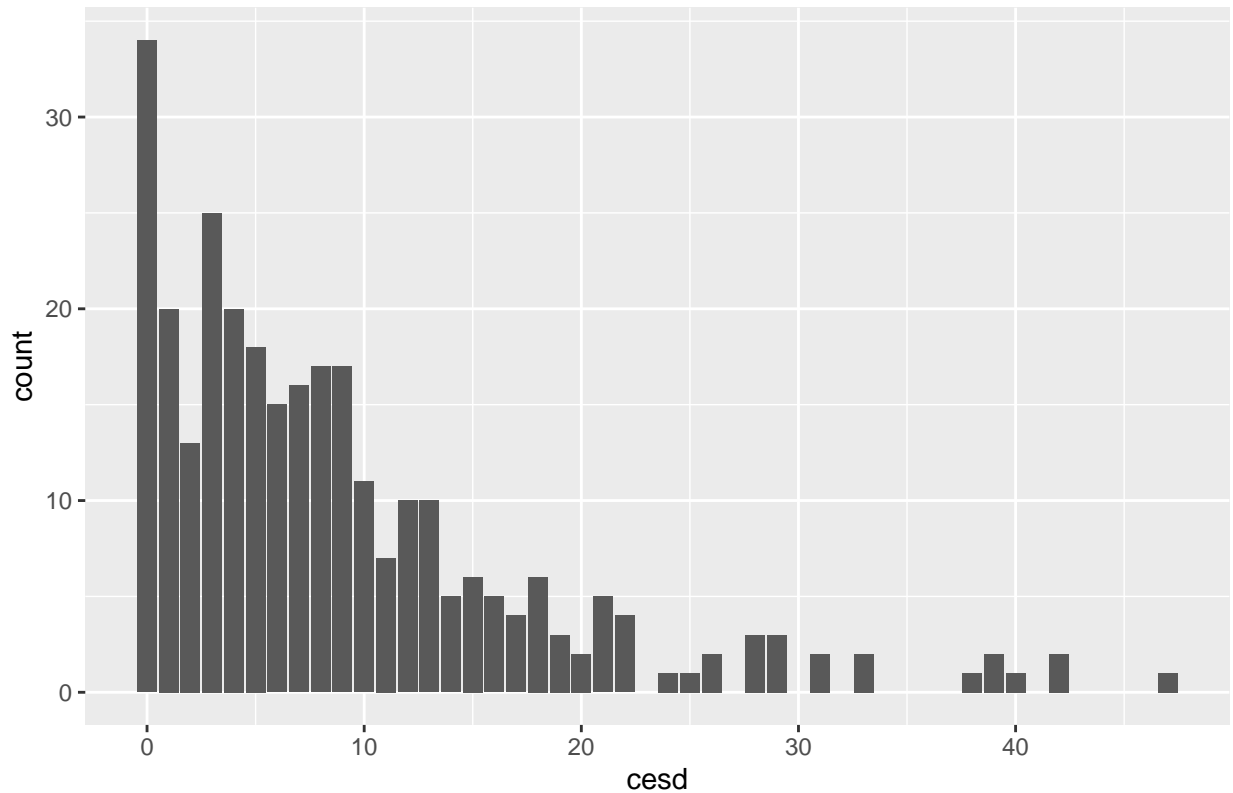
This varible cesd is a used to determine how frequently they have depressing thoughts throughout the day and it ranges from 0 to 60. 0 is the lowest possible and 60 is the hightest level possible. Thie vailbe is actually the sum of another varible in this data set called C1-C20. Which ranges from0-3 0= Rarley or none times a day, and 3= being Most or all of the day.
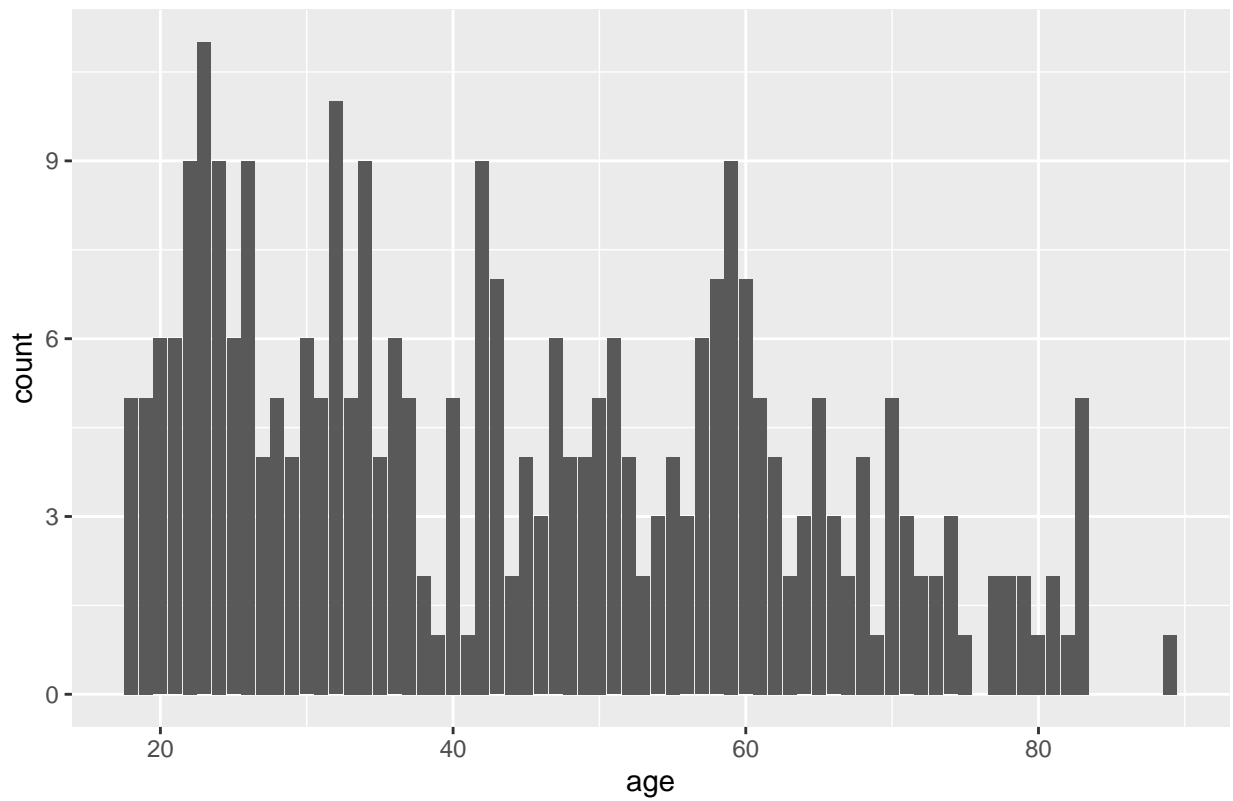
Bivarite Comparison

```
a <- mutate(depress, Income_age_disparity = age - income)
select(a, Income_age_disparity)
```

```
##      Income_age_disparity
## 1                      64
## 2                      43
## 3                      17
## 4                      41
## 5                      -2
## 6                      13
## 7                      47
## 8                      13
## 9                      24
## 10                     -5
## 11                     -5
## 12                     33
## 13                     11
```

```
## 14             48
## 15              8
## 16             15
## 17             15
## 18             36
## 19             11
## 20             55
## 21             38
## 22            -10
## 23             65
## 24             42
## 25             37
## 26             33
## 27             10
## 28             68
## 29             15
## 30             14
## 31             28
## 32            -14
## 33             25
## 34             12
## 35             33
## 36             25
## 37             14
## 38             20
## 39             36
## 40             19
## 41              8
## 42              9
## 43             -8
## 44             23
## 45             43
## 46             44
## 47             -3
## 48             20
## 49              1
## 50             19
## 51             72
## 52             -4
## 53             32
## 54             74
## 55             43
## 56             20
## 57             28
## 58             32
## 59             44
## 60             -8
## 61             62
## 62             30
## 63             11
## 64             50
## 65             51
## 66              6
## 67             57
```

```
## 68             -5
## 69             27
## 70             66
## 71            -11
## 72             11
## 73             30
## 74             19
## 75             53
## 76             59
## 77            -45
## 78             63
## 79             72
## 80             68
## 81             62
## 82             20
## 83             64
## 84             53
## 85             24
## 86             13
## 87             54
## 88             30
## 89             10
## 90              3
## 91             20
## 92             15
## 93             44
## 94             46
## 95            -44
## 96             75
## 97             46
## 98             60
## 99             61
## 100             7
## 101             3
## 102            13
## 103           -28
## 104            25
## 105            -3
## 106            14
## 107            21
## 108             6
## 109            -3
## 110            -9
## 111            15
## 112            64
## 113            27
## 114            49
## 115            27
## 116            43
## 117            26
## 118            13
## 119            73
## 120            61
## 121            -3
```

```
## 122            74
## 123            27
## 124            16
## 125            21
## 126            69
## 127            64
## 128            64
## 129            45
## 130            45
## 131            23
## 132            40
## 133            71
## 134            26
## 135            20
## 136            13
## 137             7
## 138           -11
## 139           -16
## 140            24
## 141            15
## 142            14
## 143            -5
## 144            13
## 145             3
## 146            18
## 147             8
## 148            14
## 149             8
## 150            29
## 151            19
## 152            18
## 153            69
## 154            16
## 155            16
## 156            11
## 157             9
## 158            33
## 159            12
## 160             0
## 161             1
## 162           -20
## 163             1
## 164            17
## 165           -16
## 166           -29
## 167            14
## 168             2
## 169             4
## 170            66
## 171            10
## 172            56
## 173            45
## 174            12
## 175            51
```

```
## 176                   66
## 177                   19
## 178                  -10
## 179                   13
## 180                    2
## 181                   43
## 182                   46
## 183                   17
## 184                    2
## 185                   77
## 186                   27
## 187                   19
## 188                   35
## 189                   11
## 190                   75
## 191                    3
## 192                   49
## 193                   52
## 194                    2
## 195                   57
## 196                   -6
## 197                    6
## 198                   19
## 199                   18
## 200                   27
## 201                   12
## 202                   52
## 203                   14
## 204                    7
## 205                   52
## 206                   59
## 207                   21
## 208                   14
## 209                   38
## 210                    0
## 211                   24
## 212                    4
## 213                   53
## 214                   17
## 215                   69
## 216                   15
## 217                   48
## 218                   27
## 219                   16
## 220                   14
## 221                  -27
## 222                  -13
## 223                  -28
## 224                  -13
## 225                   15
## 226                   50
## 227                   58
## 228                   16
## 229                   28
```

```
## 230          52
## 231         -33
## 232          -5
## 233           7
## 234         -46
## 235          19
## 236         -36
## 237          61
## 238          50
## 239          13
## 240          12
## 241          -9
## 242         -14
## 243          13
## 244          21
## 245          47
## 246          12
## 247          34
## 248          60
## 249           7
## 250          52
## 251          63
## 252          74
## 253          70
## 254         -27
## 255          14
## 256          27
## 257          40
## 258          10
## 259           8
## 260          -5
## 261          42
## 262          24
## 263          42
## 264          25
## 265          25
## 266          13
## 267          47
## 268          28
## 269           0
## 270         -29
## 271          -3
## 272          13
## 273           5
## 274          31
## 275          24
## 276          44
## 277          71
## 278           9
## 279          20
## 280          27
## 281         -22
## 282          -3
## 283           2
```
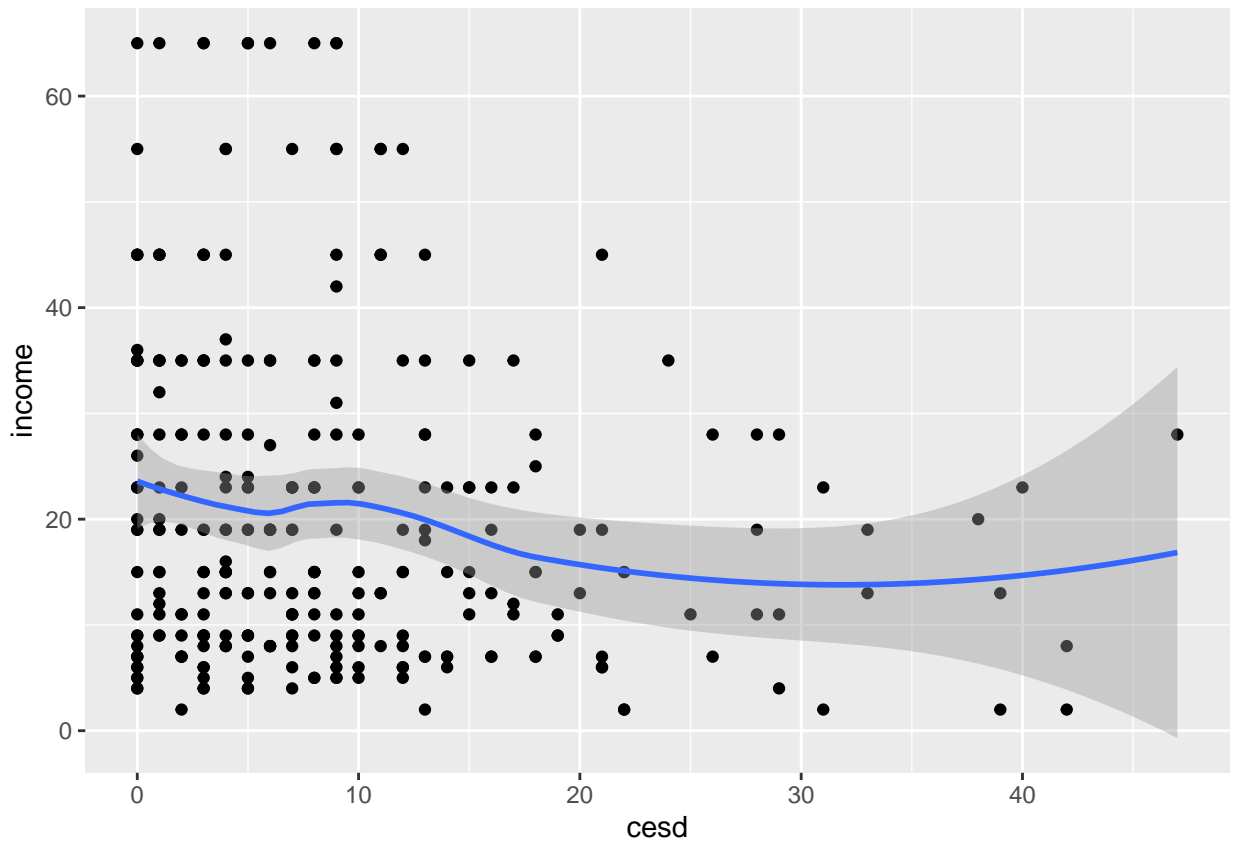
```
## 284                 79
## 285                 24
## 286                  7
## 287                 21
## 288                 33
## 289                 -9
## 290                 26
## 291                 10
## 292                  9
## 293                 15
## 294                 49
```

I am creating this variable to try and describe the disparity of income and age between all my observati

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

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


Their seems to be no strong correlation bewtween my vairbles cesd and income most of my data does not fit with my line of best fit and their are many outliers in my graph. I choose these two varibles because i really though that a person income would have an effect from their depress thoughts but it seems that is not the case.

```
ggplot(depress, aes(x = cesd, y = age, color = Educational)) + geom_jitter(width = 0.5) +
    ggtitle("Age and Education Level vs. CESD Scores")
```

Age and Education Level vs. CESD Scores
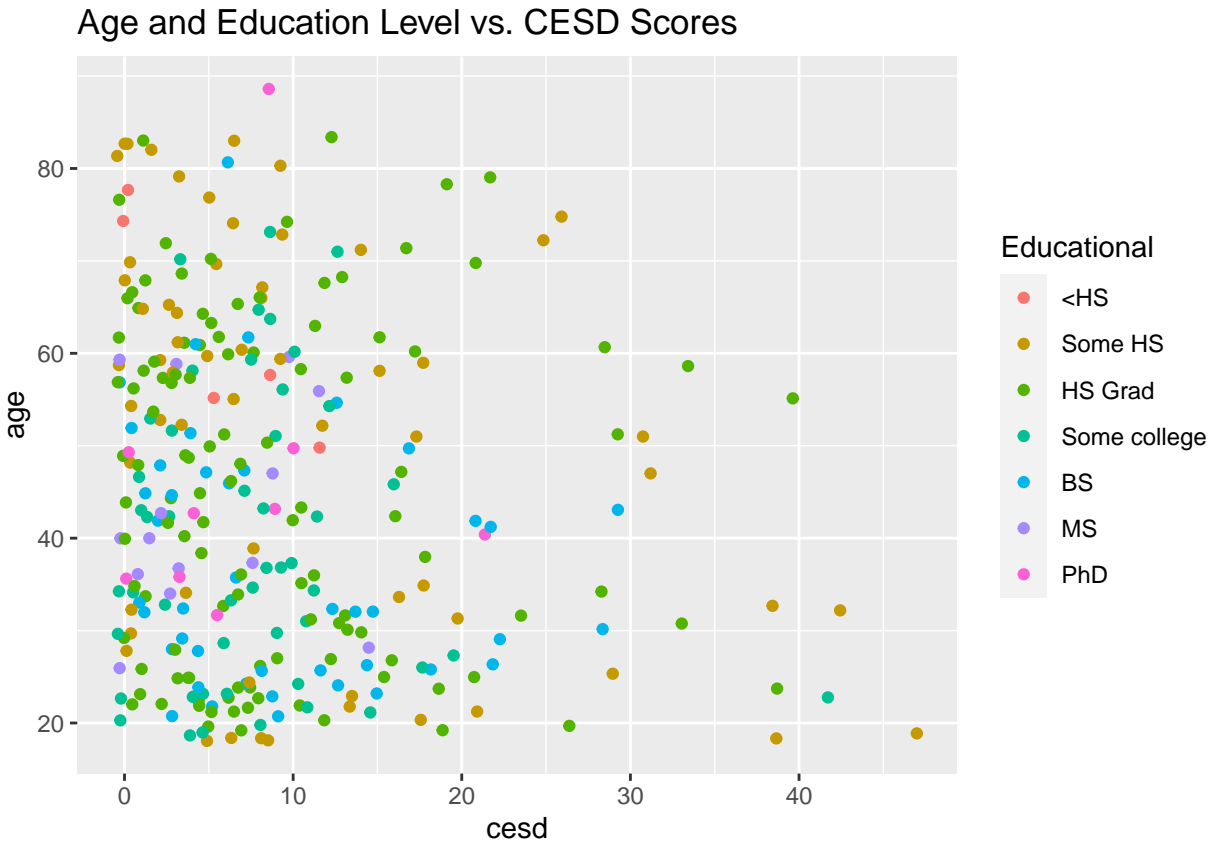
From my graph comparison with age education and their cesd score it seems that all education level particpants seem to fall within the 20 cesd range no matter what their background in education is. However I will say that most of the outliers are made from people wih Some HS and Hs grad education backgroud. This could be just because we had more observations of some HS and HS Grad.

Conclusion

It seems that there are no strong correlation between the varibles I choose to explain a reason why some people have more depressing thoughts than others.From the results I can tell niether education, income or age has a direct correlation between depression.Perhaps, I need to include more vaiblres into my model to see if there is any correlation for depression in this data set.