Depression_EDA: Employment & Education Status

Gloria Jimenez

2/21/2022

Introduction

The data set I will be exploring is the Depression data set. This data was obtained by a first set of interviews of adult residents of the Los Angeles County as a prospective depression study. My research question is what is the impact of education and employment status on depression rate.

Data

```
Depression <- read.table("C:/Users/Gloria/Desktop/MATH130/data/depress_081217.txt", header=TRUE, sep="\t
table(Depression$employ, Depression$educat)
```

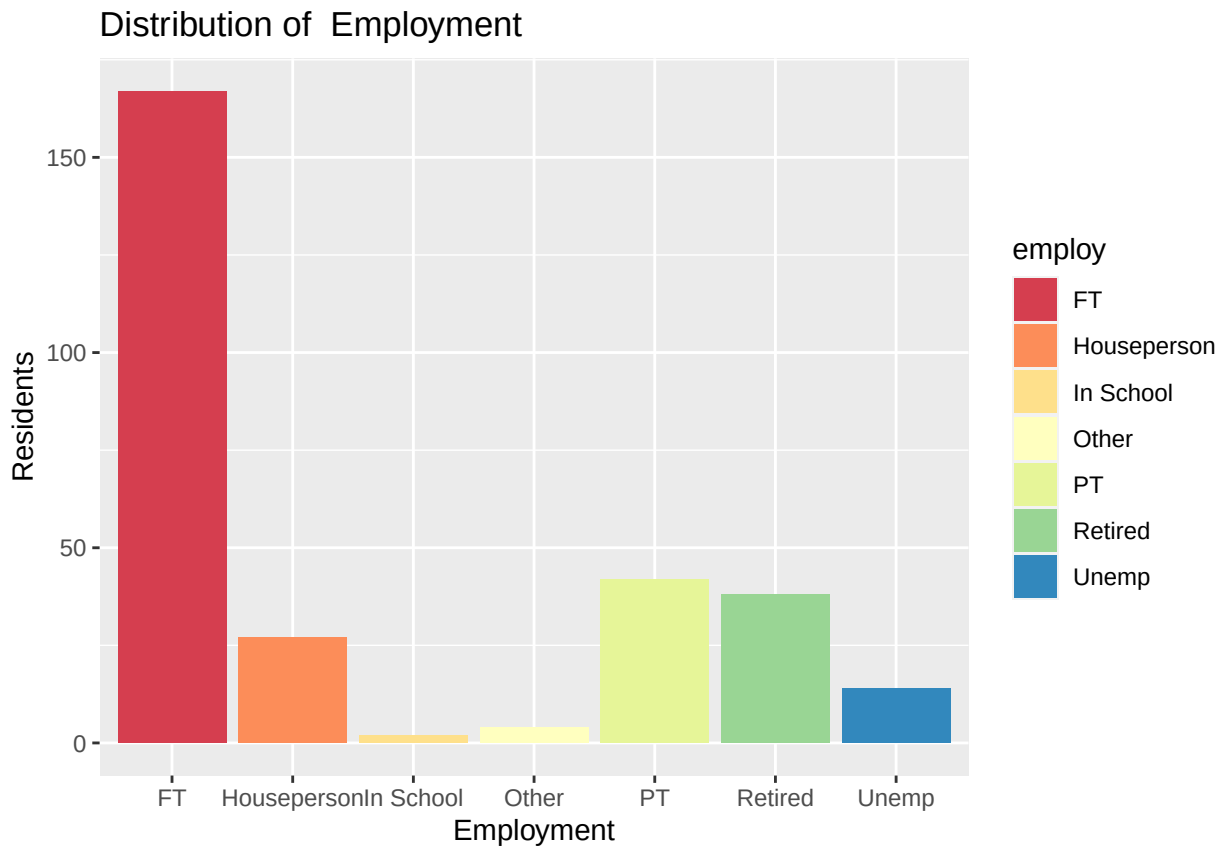
```
##
##               <HS BS HS Grad MS PhD Some college Some HS
##   FT            1 33      67 13   7           29      17
##   Houseperson   0  1       8  0   0            4      14
##   In School     0  0       1  0   0            0       1
##   Other         0  0       1  0   0            1       2
##   PT            1  5      17  1   2           11       5
##   Retired       2  1      16  0   0            3      16
##   Unemp         1  3       4  0   0            0       6
```

Univariate Exploration
Employment

```r
library(RColorBrewer)
table(Depression$employ)
```

```
##
##          FT Houseperson   In School       Other          PT     Retired
##         167          27           2           4          42          38
##       Unemp
##          14
```

```
ggplot(Depression, aes(x=employ,fill=employ)) + geom_bar() +ggtitle("Distribution of  Employment")+
  ylab("Residents") + xlab("Employment") +scale_fill_brewer(palette="Spectral")
```



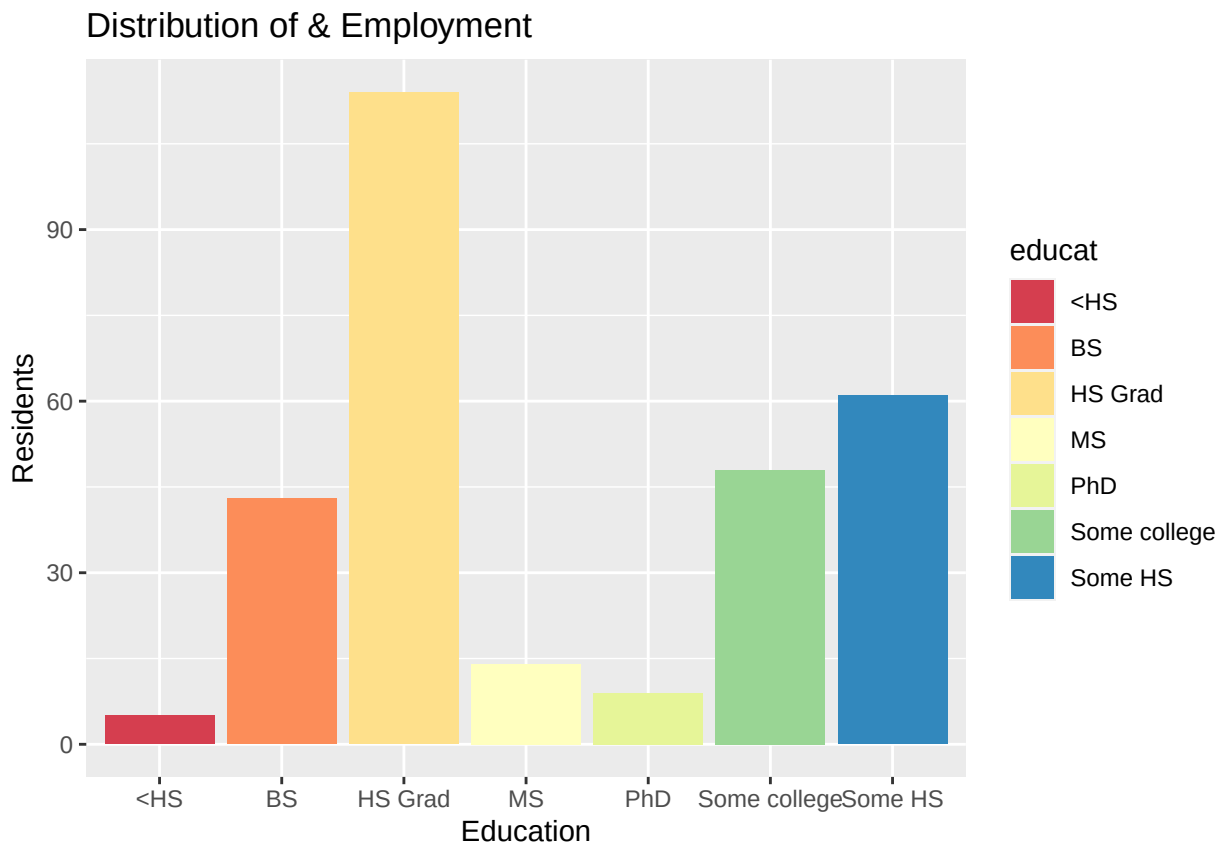

```
This is the Employment Status data of the Residents
ranging from being full time(167), part time(42), unemployed(14) and being a houseperson(27), in school
```

Education

```
library(RColorBrewer)
table(Depression$educat)
```

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

```
ggplot(Depression, aes(x=educat,fill=educat)) + geom_bar() +ggtitle("Distribution of & Employment")+
  ylab("Residents") + xlab("Education") +scale_fill_brewer(palette="Spectral")
```



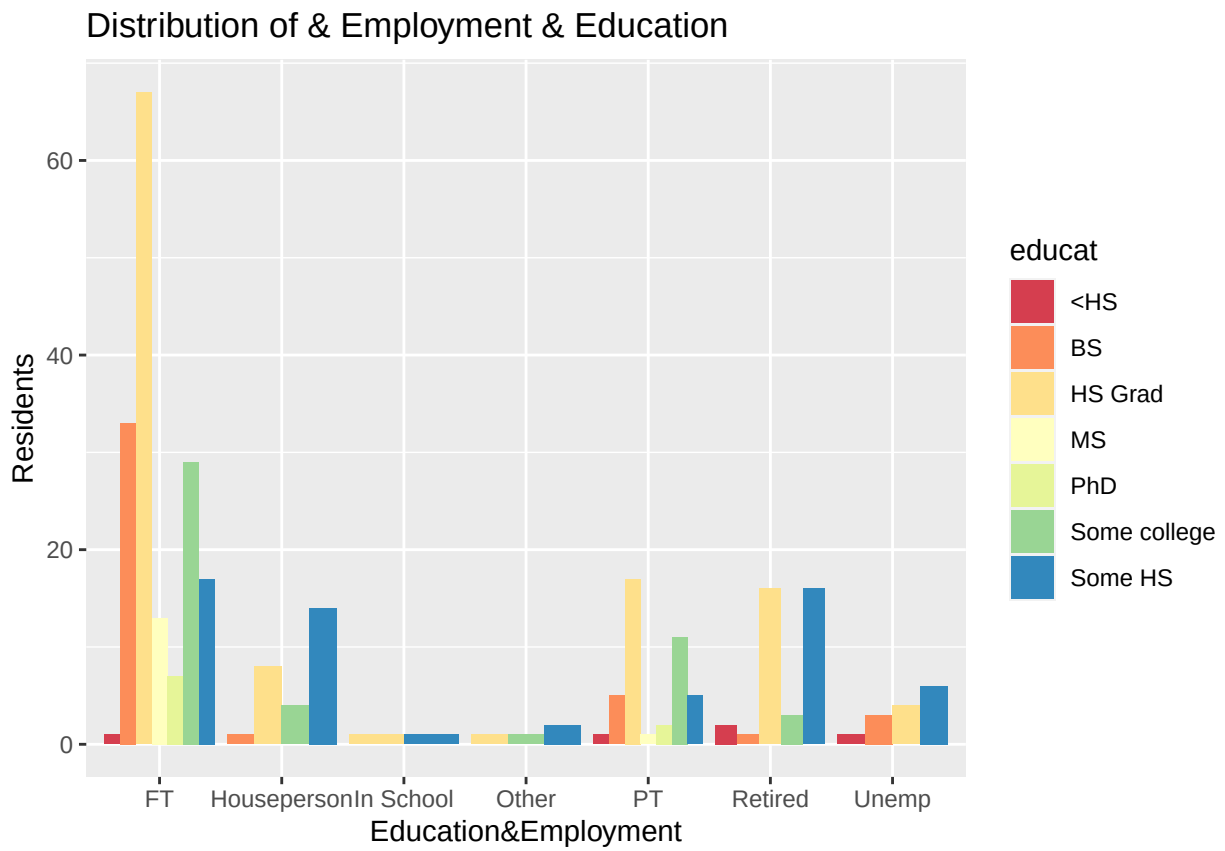
This data above is the Education level of the residents ranging from some high school(5), high school grad(114), some college(48), Bachelor's(43),PhD(9), MS(14).

Bivaritate

```
table(Depression$employ, Depression$educat)
```

```
##
##                <HS BS HS Grad MS PhD Some college Some HS
##   FT             1 33       67 13   7           29       17
##   Houseperson    0  1        8  0   0            4       14
##   In School      0  0        1  0   0            0        1
##   Other          0  0        1  0   0            1        2
##   PT             1  5       17  1   2           11        5
##   Retired        2  1       16  0   0            3       16
##   Unemp          1  3        4  0   0            0        6
```

```
ggplot(Depression, aes(x=employ, fill=educat)) + geom_bar(position = "dodge")+ggtitle("Distribution of &
  ylab("Residents") + xlab("Education&Employment") +scale_fill_brewer(palette="Spectral")
```



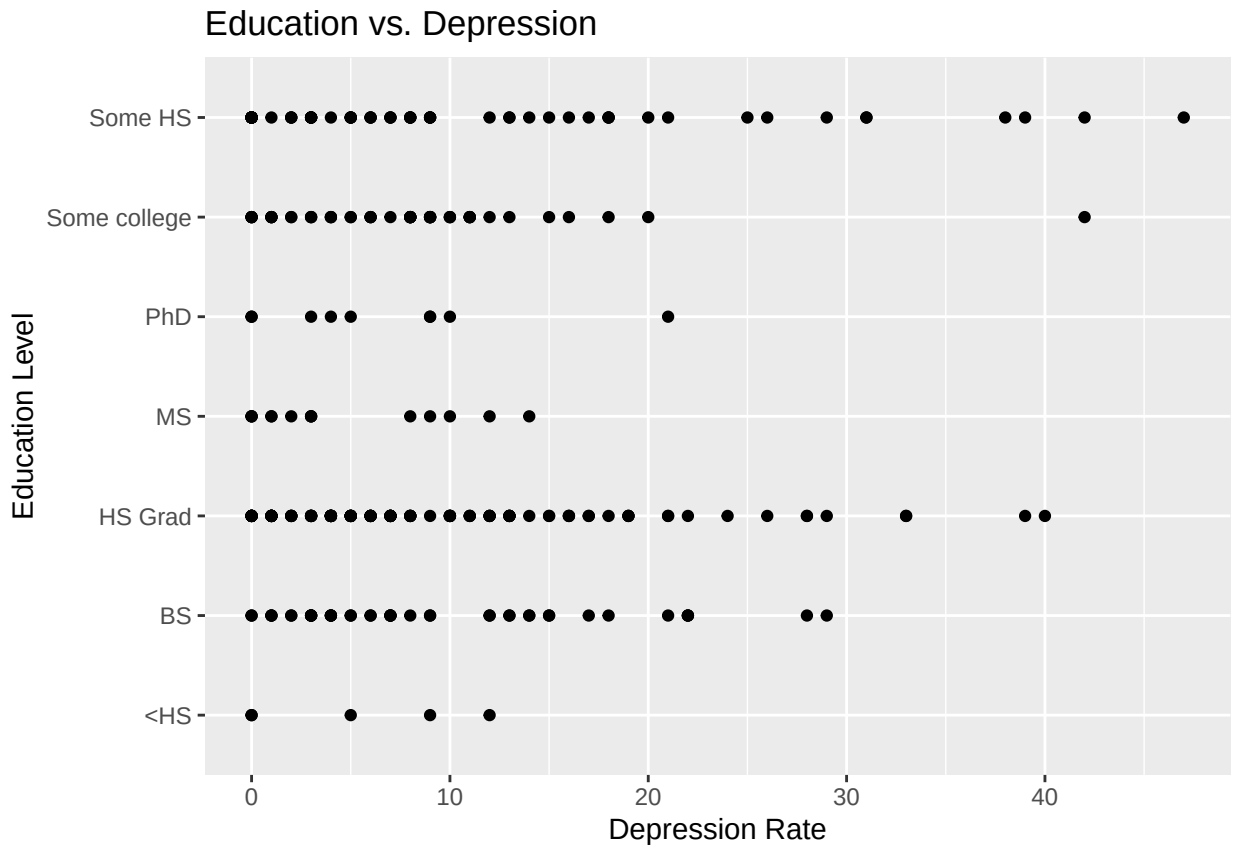

```
table(Depression$employ, Depression$educat) %>% prop.table(margin=1) %>% round(3)
```

```
##
##                 <HS    BS HS Grad    MS   PhD Some college Some HS
##   FT          0.006 0.198   0.401 0.078 0.042        0.174   0.102
##   Houseperson 0.000 0.037   0.296 0.000 0.000        0.148   0.519
##   In School   0.000 0.000   0.500 0.000 0.000        0.000   0.500
##   Other       0.000 0.000   0.250 0.000 0.000        0.250   0.500
##   PT          0.024 0.119   0.405 0.024 0.048        0.262   0.119
##   Retired     0.053 0.026   0.421 0.000 0.000        0.079   0.421
##   Unemp       0.071 0.214   0.286 0.000 0.000        0.000   0.429
```

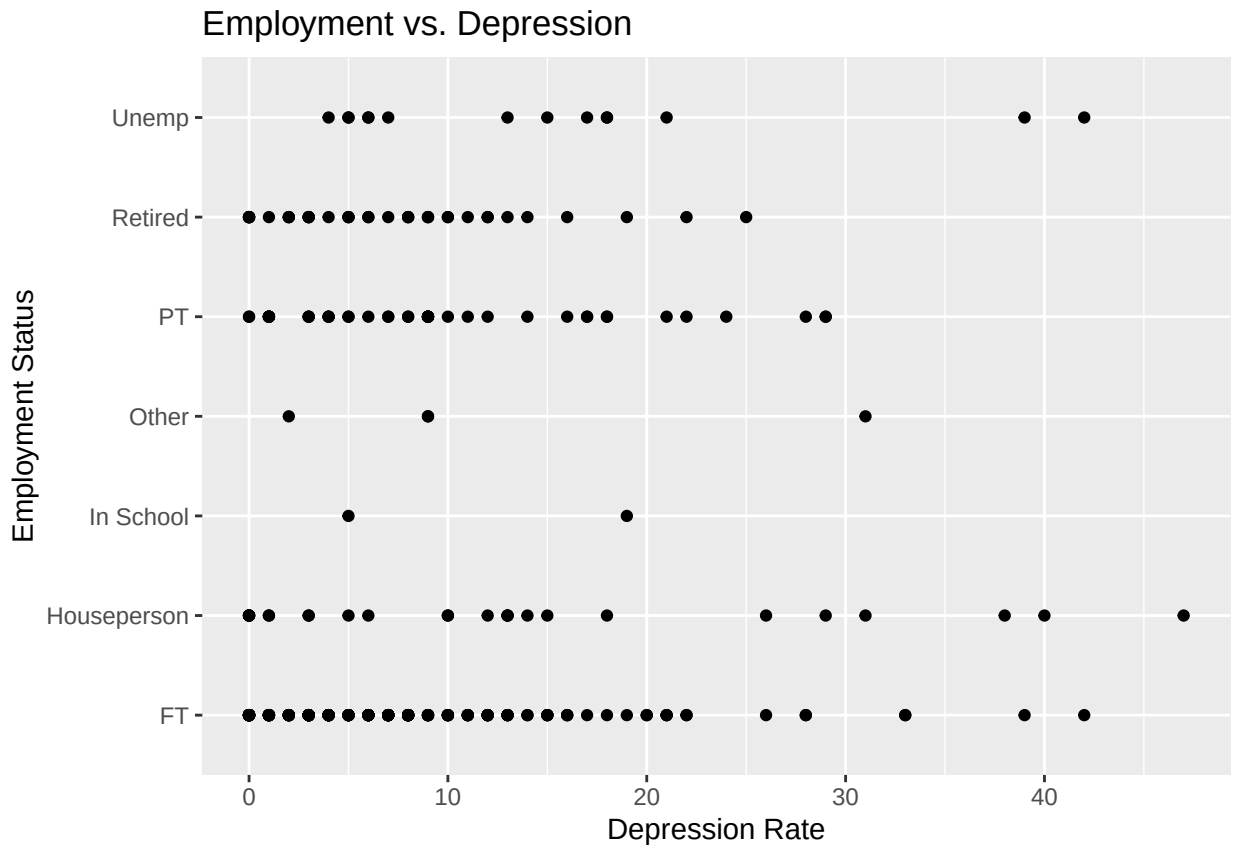
The data above is the two statistics, Employment & Education side by side and finished with a statistical table of the two.

Conclusion

```
ggplot(Depression, aes(x=cesd, y=educat)) + geom_point()+ggtitle("Education vs. Depression")+
  ylab("Education Level") + xlab("Depression Rate")
```




```
ggplot(Depression, aes(x=cesd, y=employ)) + geom_point()+ggtitle("Employment vs. Depression")+
  ylab("Employment Status") + xlab("Depression Rate")
```



In conclusion I expected to see higher rates with depression and low education and employment rates. These last two sets compare the data of employment vs. depression as well as education vs. depression.