

Final Project- Depression Data

Daisy Lopez Rojas

9/22/2021

Depression

Introduction

The following data, Depression, is taken from a study Los Angeles County residents in regards to depressions in adults. This data includes 294 observations from the first group of interviews. From this data set, the variables I will be focusing on are “Age”, “Education”, and “Health”. My outcome is to see how education and age are related to over all health of an individual.

```
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(ggplot2)  
library(forcats)  
library(knitr)  
Depress<- read.delim("C:/Users/daisy/OneDrive/Desktop/Math 130 (Rstudio)/Data/depress_081217.txt", head  
dim(Depress)
```

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

```
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 ...

```

Age

```
summary(Depress$age)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##  18.00   28.00   42.50   44.41   59.00   89.00
```

```
mean(Depress$age)
```

```
## [1] 44.41497
```

```
sd(Depress$age)
```

```
## [1] 18.08544
```

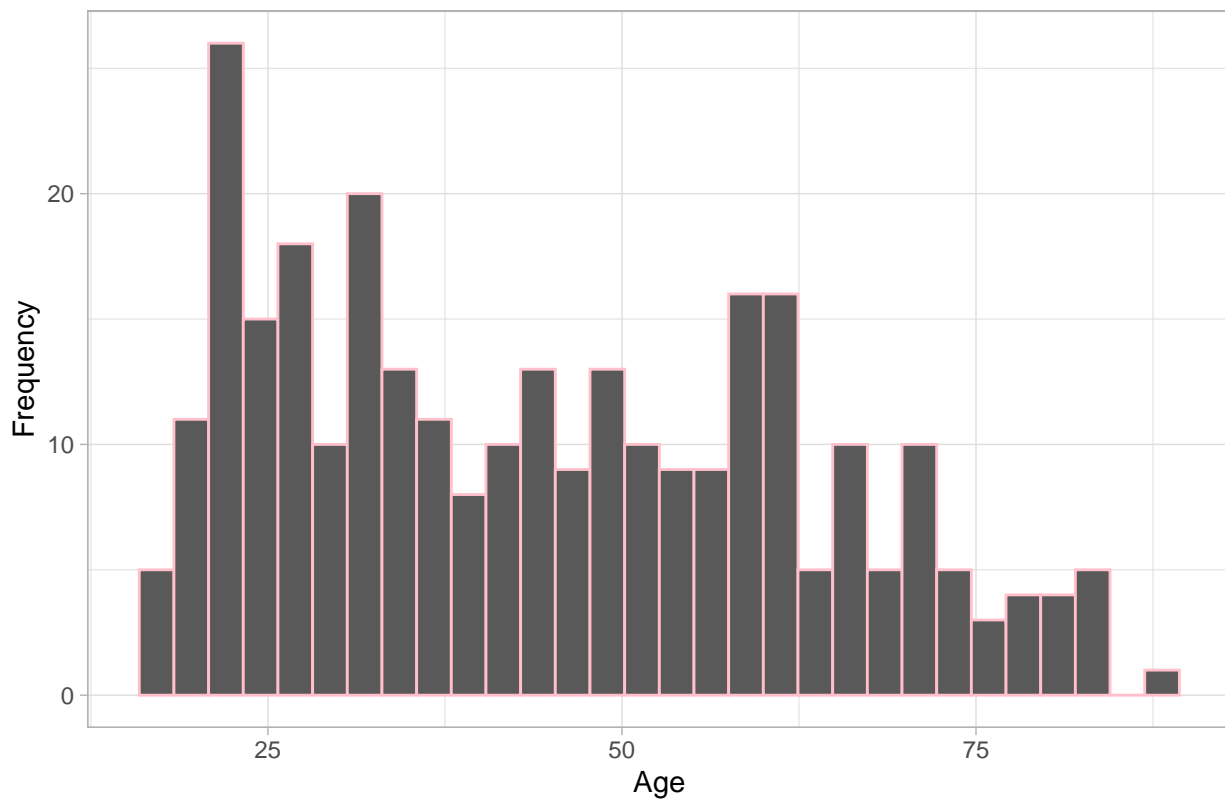
```
table(Depress$age)
```

```
##  
## 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43  
## 5 5 6 6 9 11 9 6 9 4 5 4 6 5 10 5 9 4 6 5 2 1 5 1 9 7  
## 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69  
## 2 4 3 6 4 4 5 6 4 2 3 4 3 6 7 9 7 5 4 2 3 5 3 2 4 1  
## 70 71 72 73 74 75 77 78 79 80 81 82 83 89  
## 5 3 2 2 3 1 2 2 2 1 2 1 5 1
```

```
ggplot(Depress, aes(x=age, fill=age))+geom_histogram(col="pink")+theme_light()+ggtitle("Distribution of Age in Study")
```

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

Distribution of Age in Study



The histogram graph above represents the age range and distribution of the interviewees in the study. The majority of participants of the study are from the age range of 20 to 60, with an increase in 23 years of age.

Education

```
summary(Depress$educat)
```

```
## Length Class Mode  
## 294 character character
```

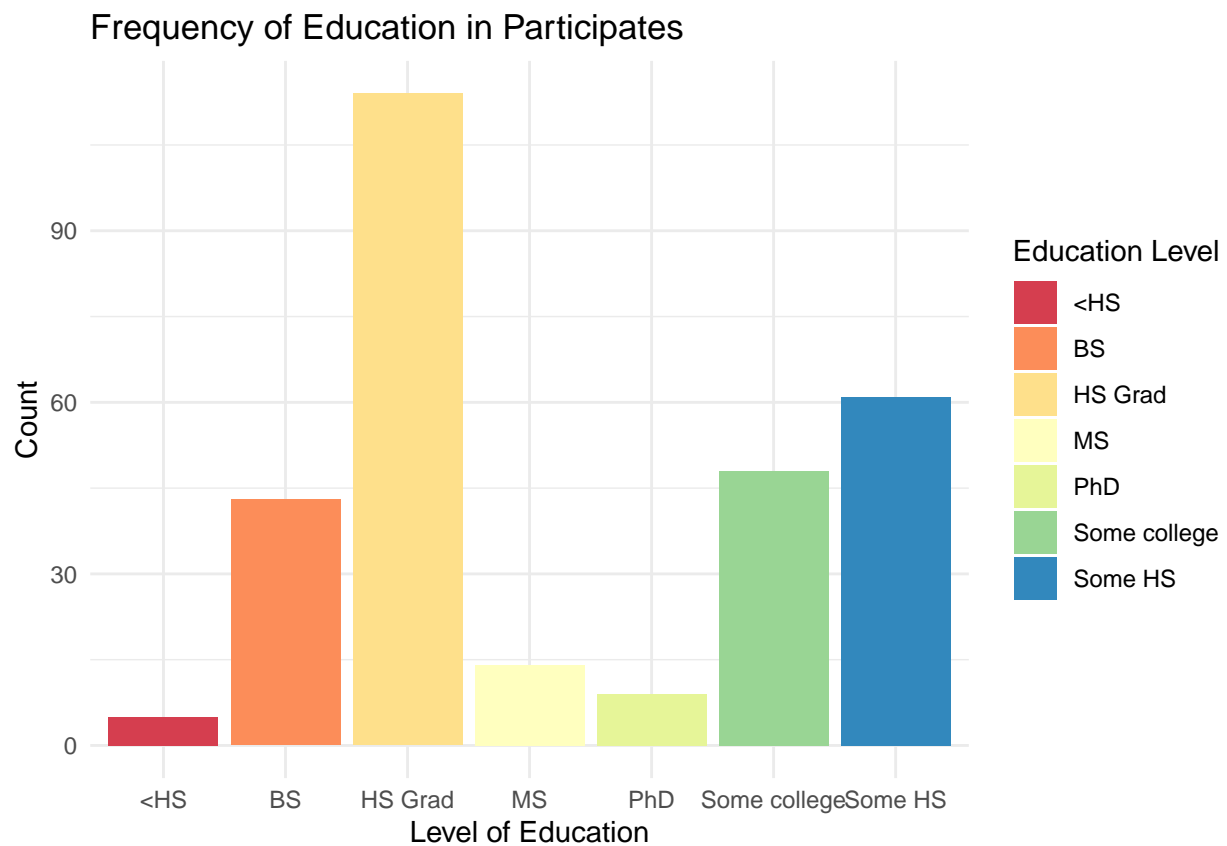
```
class(Depress$educat)
```

```
## [1] "character"
```

```
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))+geom_bar()+theme_minimal()+ggtitle("Frequency of Education in Participates")
```



The bar graph represents the count of interviewees' educational background. The majority of the interviewees have their high school diploma.

Health

```
summary(Depress$health)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.  
##      1.000  1.000   2.000  1.772  2.000   4.000
```

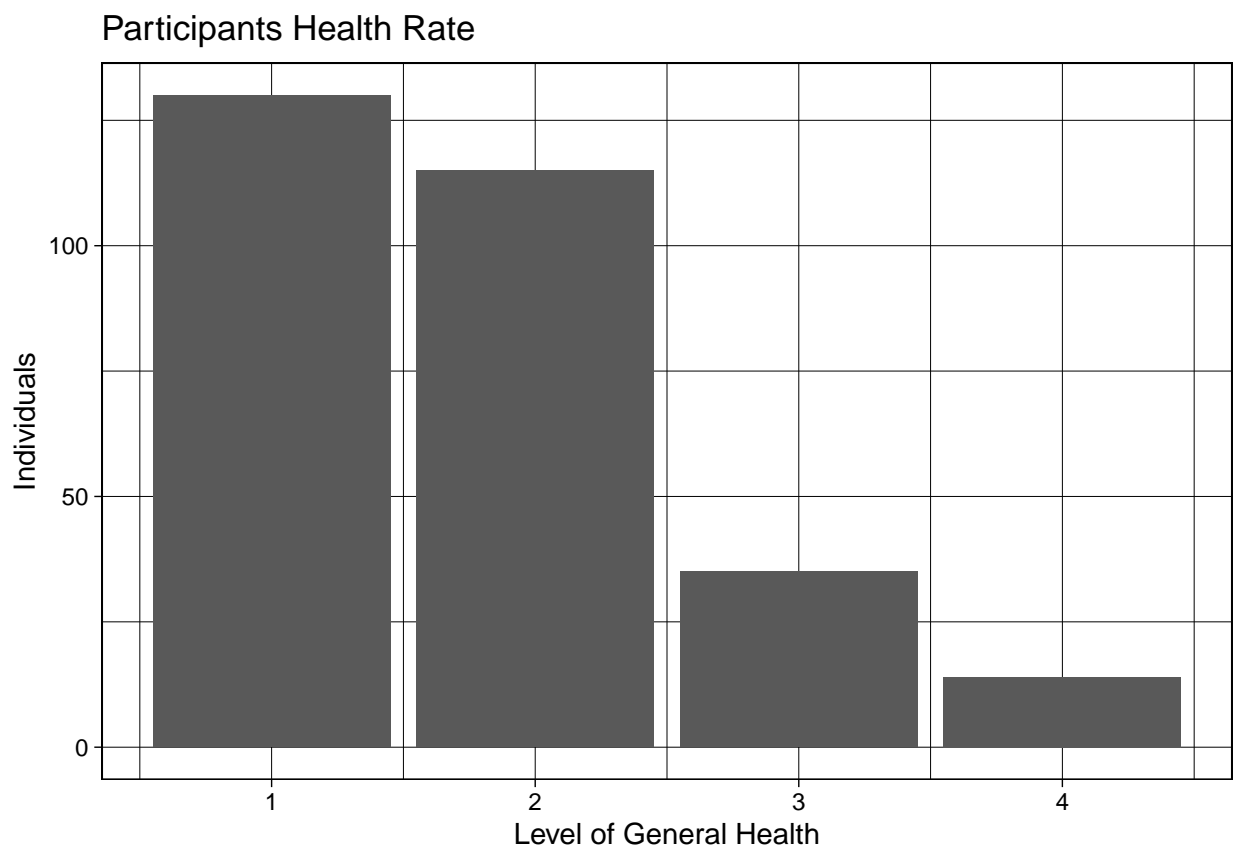
```
table(Depress$health)
```

```
##  
##  1  2  3  4  
## 130 115 35 14
```

```
sd(Depress$health)
```

```
## [1] 0.8379466
```

```
ggplot(Depress,aes(x=health, fill=health))+geom_bar()+theme_linedraw()+scale_fill_discrete(name="Rate o
```

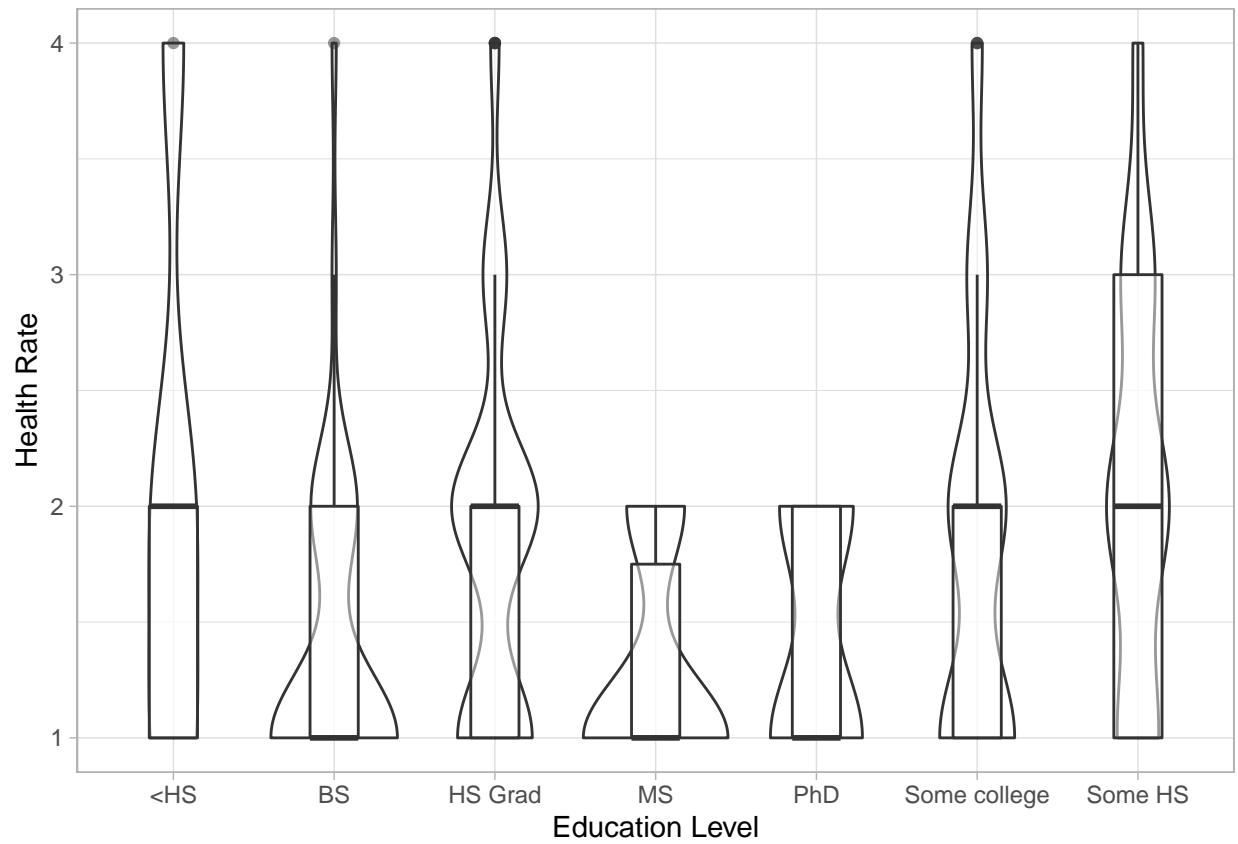


The bar graph, “Participants Health Rate”, illustrates the rates that the interviewee answer to the question regarding their general health. The range at which the participants of the study rated there overall health was from 1 to 4; 1 being excellent, 2 being good, 3 being fair, and 4 being poor. The majority rated their health between a 1 and 2, as seen in the graph above.

Bivariate Analysis

Health vs. Age vs. Education

```
ggplot(Depress, aes(x=educat, y= health, fill=age))+geom_violin(alpha=.5)+geom_boxplot(alpha=.5, width=.3)
```



The violin plot incorporates all three variables, Education, health, and age. From the graph, it can be inferred that the interviewees with lots of education, Master or PhD, rate their health lower.