EDA_grhamilton

Grant Hamilton

2023-09-15

```
##
## 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
depress <- read.table("C:/Users/bevog/Desktop/Math130/Data/depress.txt",
    header = TRUE, sep = "\t")
```

I will be analyzing the education level of each individual ('educat'), by their level of depression ('cesd'). I am interested in finding out if there is a correlation between people who have a higher education level struggling with more severe depression.

Research question: Do people with higher education deal with more severe depression? T

This data represents the number that were tested with the respective education levels.

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

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

This data represents the levels of depression that were recorded throughout the study.

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

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   3.000   7.000   8.884  12.000  47.000
```
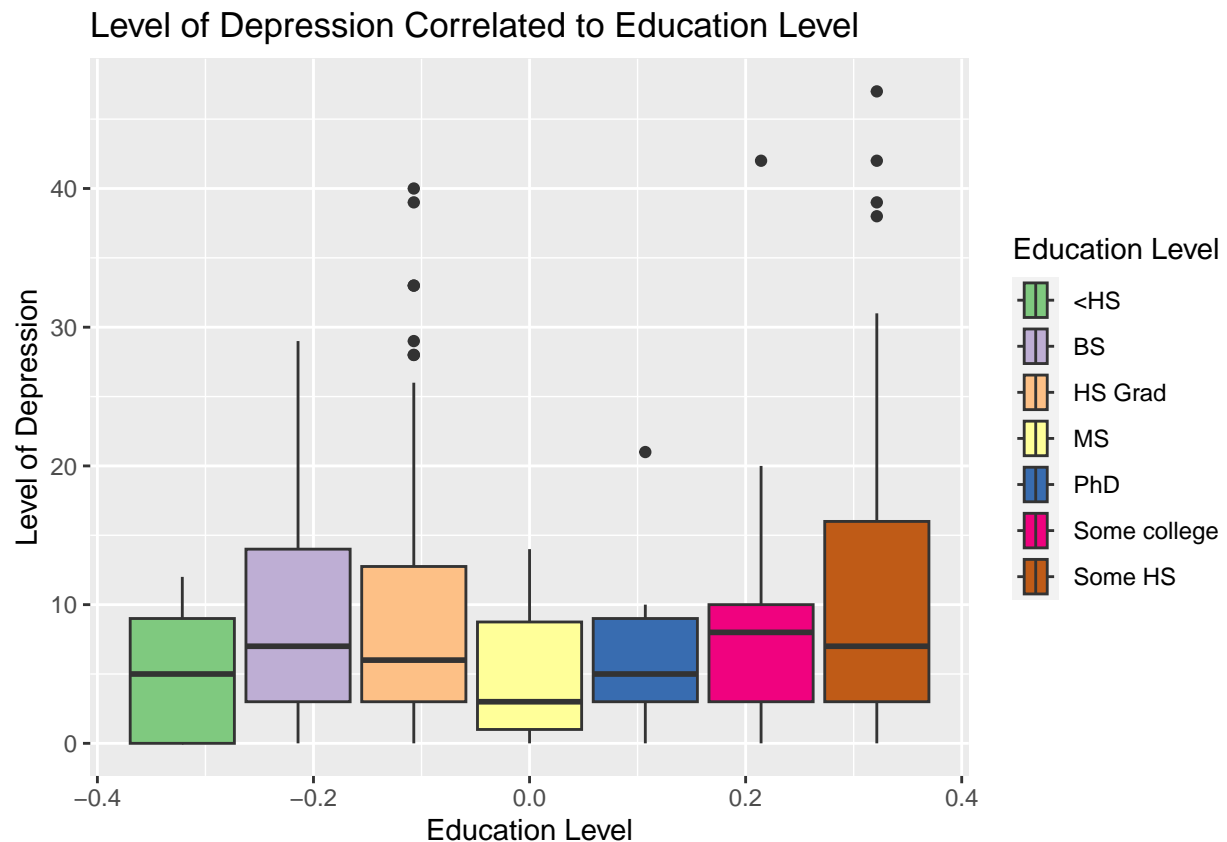
This data represents the average (mean) level of depression that was recorded from people with the respective educaton levels.

```
depress %>%
    group_by(educat) %>%
    summarise(sad = mean(cesd, na.rm = TRUE) %>%
        round(2))
```

```
## # A tibble: 7 x 2
##   educat         sad
##   <chr>        <dbl>
## 1 <HS           5.2
## 2 BS            9.33
## 3 HS Grad       8.97
## 4 MS            4.71
## 5 PhD           6.78
## 6 Some HS      10.8
## 7 Some college  7.83
```

This graph visualizes the relationship between educaton level and the level of depression recorded from people with different education levels.

```
library(RColorBrewer)
ggplot(depress, aes(x = cesd, fill = educat)) + geom_boxplot() +
    scale_fill_brewer(palette = "Accent", name = "Education Level") +
    coord_flip() + ggtitle("Level of Depression Correlated to Education Level") +
    xlab("Level of Depression") + ylab("Education Level")
```

In conclusion, I found that the data actually had the opposite trend of what I expected. I predicted that people who had achieved higher education levels would experience more severe depression. However, the data shows (excluding outlires) that people with some high school recorded the most severe depression.