Exploratory Data Analysis

Rosemary Linares

2022-09-21

Selected data set: Depression Intro: The data set chosen will provide collected data based off of observations from males and females who have been feeling depressed. The data set has the variables in this analysis focus on each the 294 participants age, education, and if they drink.

#1. Univariate Exploration:

```r
depress$drank <- factor(depress$drink, labels=c( "No", "yes"))
```

```r
class(depress$drink)
```

```
## [1] "integer"
```

```r
depress$spam_fac <- factor(depress$drink, labels=c("no", "yes"))
```

```r
class(depress$sex)
```

```
## [1] "integer"
```

```r
depress$spam_fac <- factor(depress$sex, labels=c("male", "female"))
```
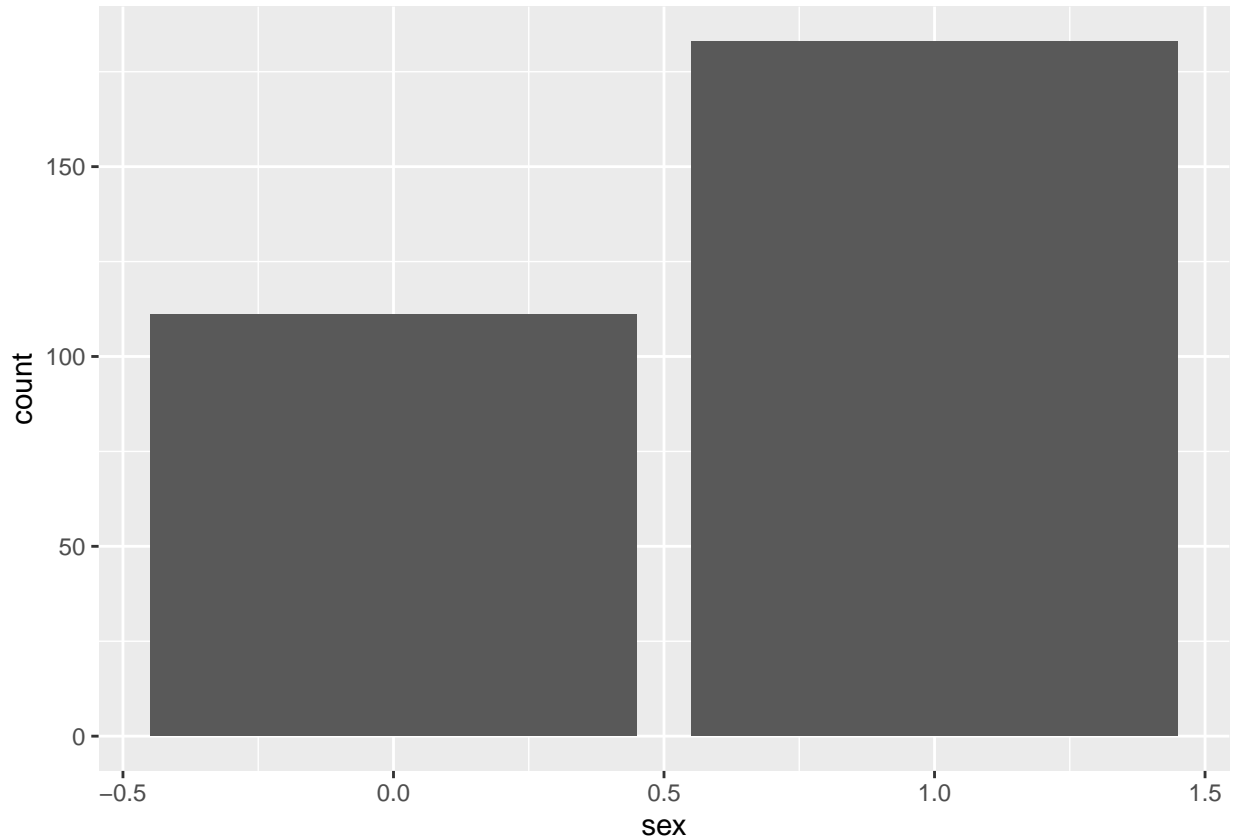
```r
wer<- select(depress, sex, educat, drink)
```

Below are the summary statitics for how many paricipants of the observations were male or female.

```r
table(depress$sex)
```

```
##
##   0   1
## 111 183
```

As stated above the table displays the gender for each participant. This shows us that of the 294 observations 111 were Male and 183 were Female.

```r
ggplot(depress,aes(x=sex))+geom_bar()
```

After looking at the bar graph we see there is a 75 observant difference.

Education table

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

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

The data displays the education background of each person that partcipated. We can see that the majority of those who have experienced depression had higher rates when they were around teen years and early adulthood where they would attend highschool or college.

Education graph

```
ggplot(depress, aes(x=educat, fill=educat)) + geom_bar() +
    scale_fill_manual(values=c("salmon", "slategray4", "violet","lightpink", "mediumpurple2", "thistle2"
```
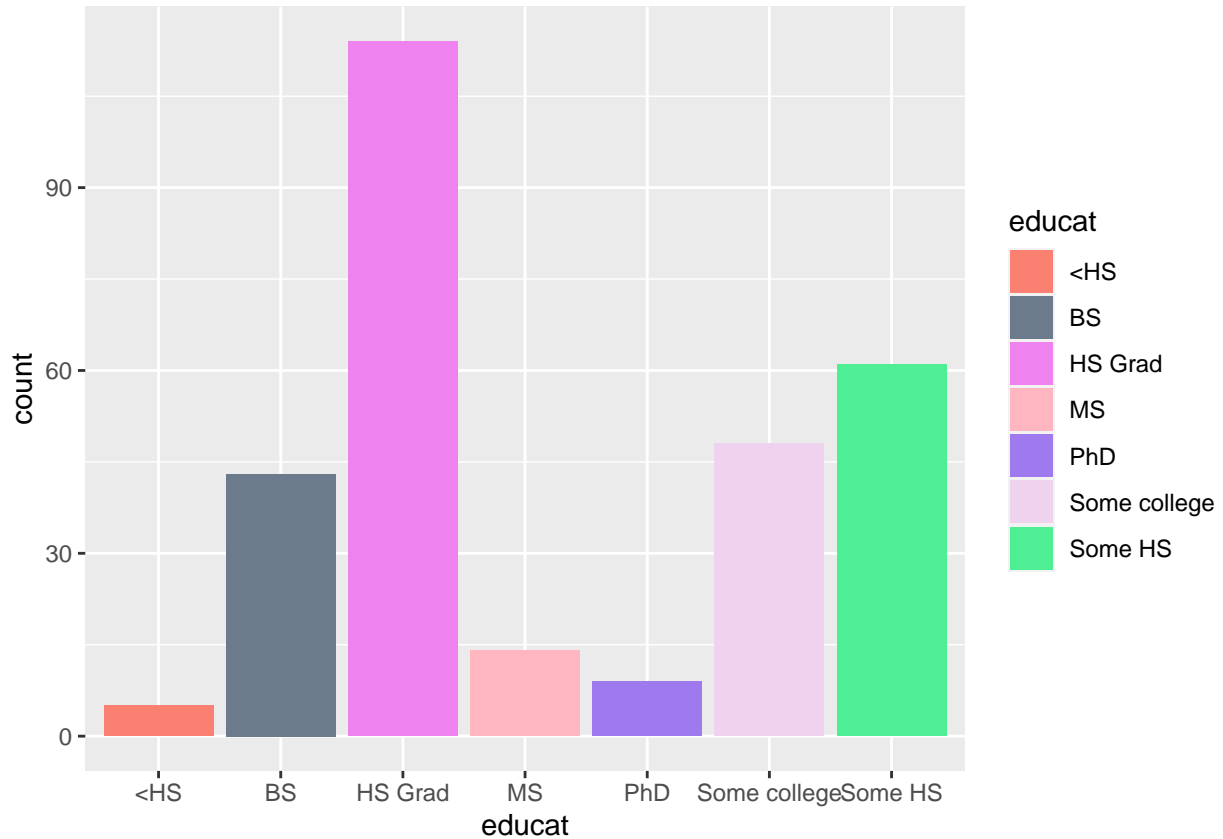
Now that we can visualize the values by color you can see the higher rates around particpants who attended education and further that were observed with depression.
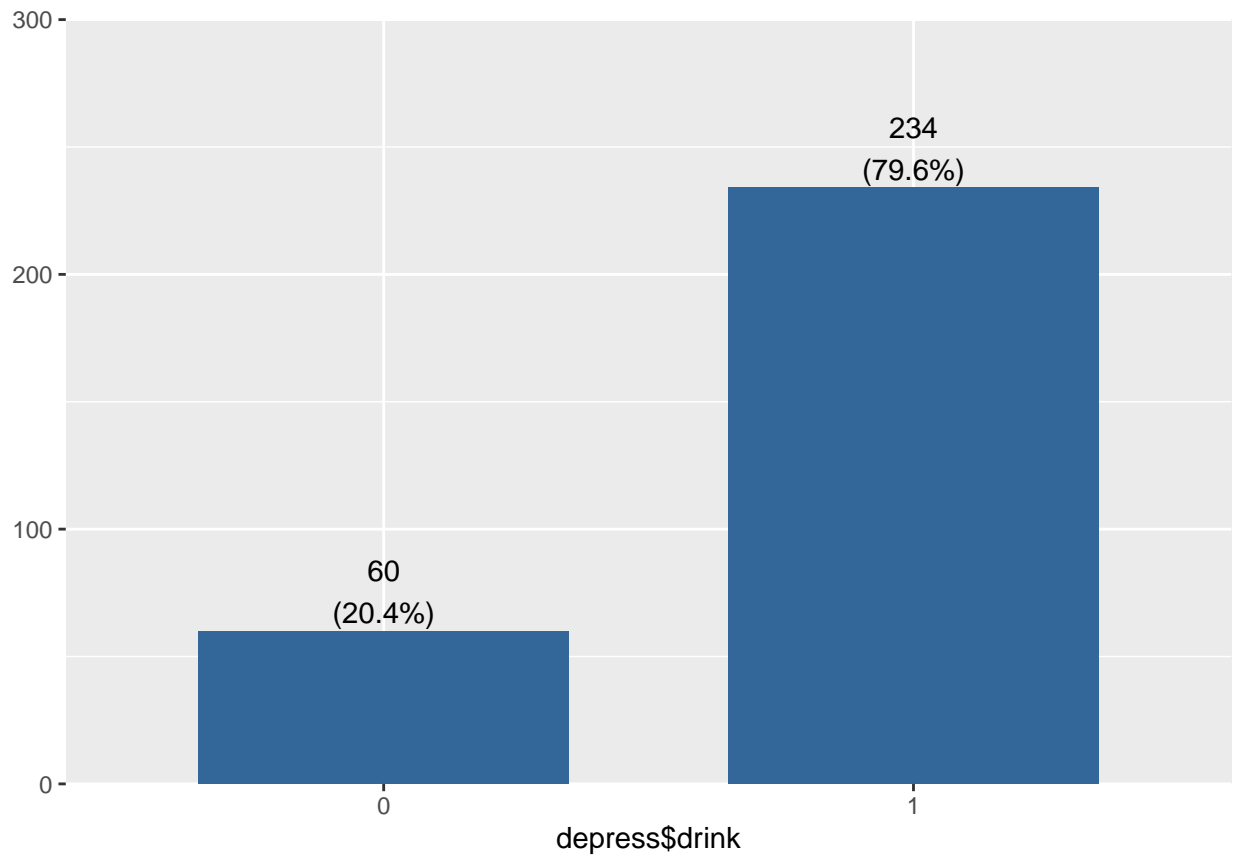
If They Drink

```
table(depress$drink)
```

```
##
##   0   1
##  60 234
```

The table shows how many particpants have answered if they drink or not.
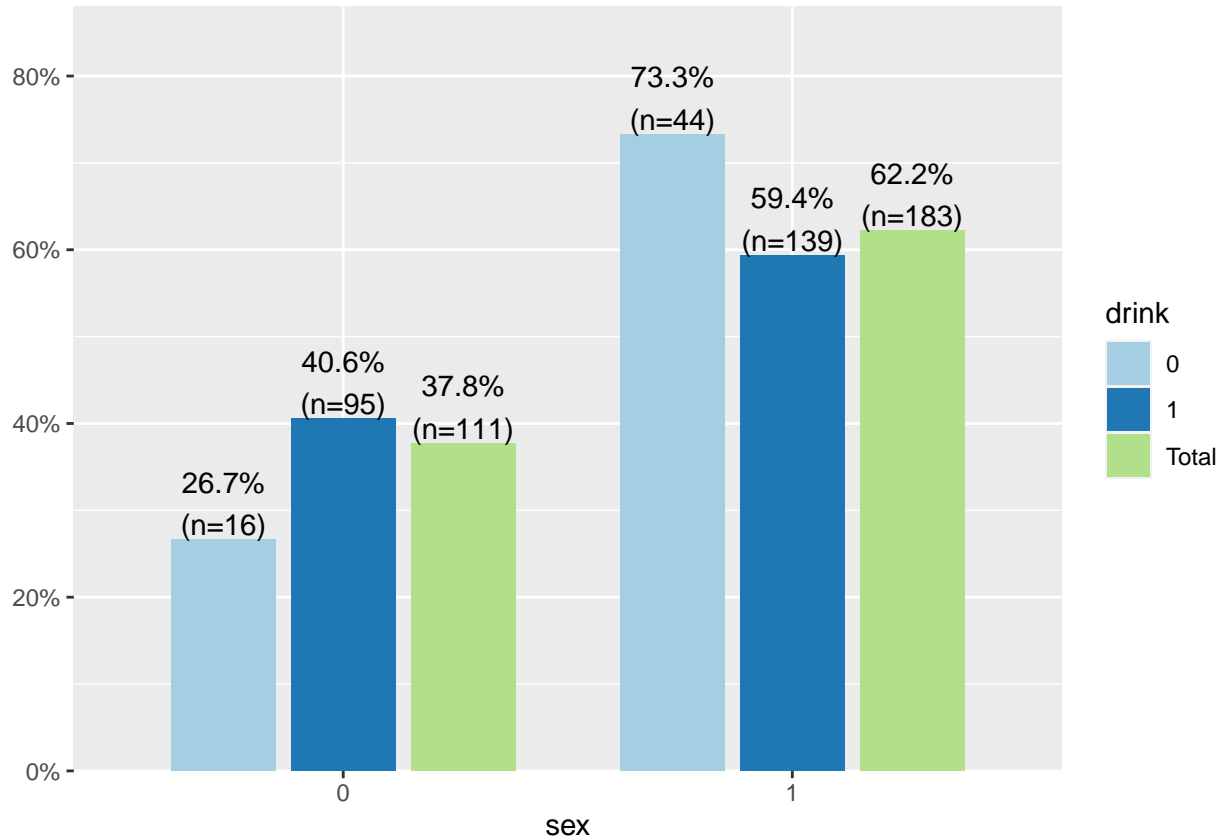
```
library(sjPlot)
```

```
## #refugeeswelcome
```

```
plot_frq(depress$drink)
```

The graph above shows the percenatge of the total 297 observations who answered if they drink.

Grouped table

```
plot_xtab(depress$sex, depress$drink)
```

Between Males and Females this graph is put on view of each Sex how many drink or not and its total.

#2. Bivariate Exploration:

You can also embed plots, for example:

```
table(wer$drink,wer$educat, wer$sex)
```

```
## , ,  = 0
##
##
##      <HS BS HS Grad MS PhD Some college Some HS
##   0   0  1       5  3   0            3       4
##   1   4 16      34  5   6           15      15
##
## , ,  = 1
##
##
##      <HS BS HS Grad MS PhD Some college Some HS
##   0   0  5      17  1   0            6      15
##   1   1 21      58  5   3           24      27
```

As sorted by the chart we can notice how many mentioning if they have had a drink or not and what was their education background.

Education background against If they drink

When looking at this ggplot you can see the the participants have a high rate of drinking. You notice it may be in the teen years and early adulthood of attending highschool and college or facing the life change soon after.

```
ggplot(depress, aes(x=educat)) + geom_density() + ggtitle("Drinking & Education Background") +
    xlab("Education") + ylab("Drinking levels") +theme_classic()
```
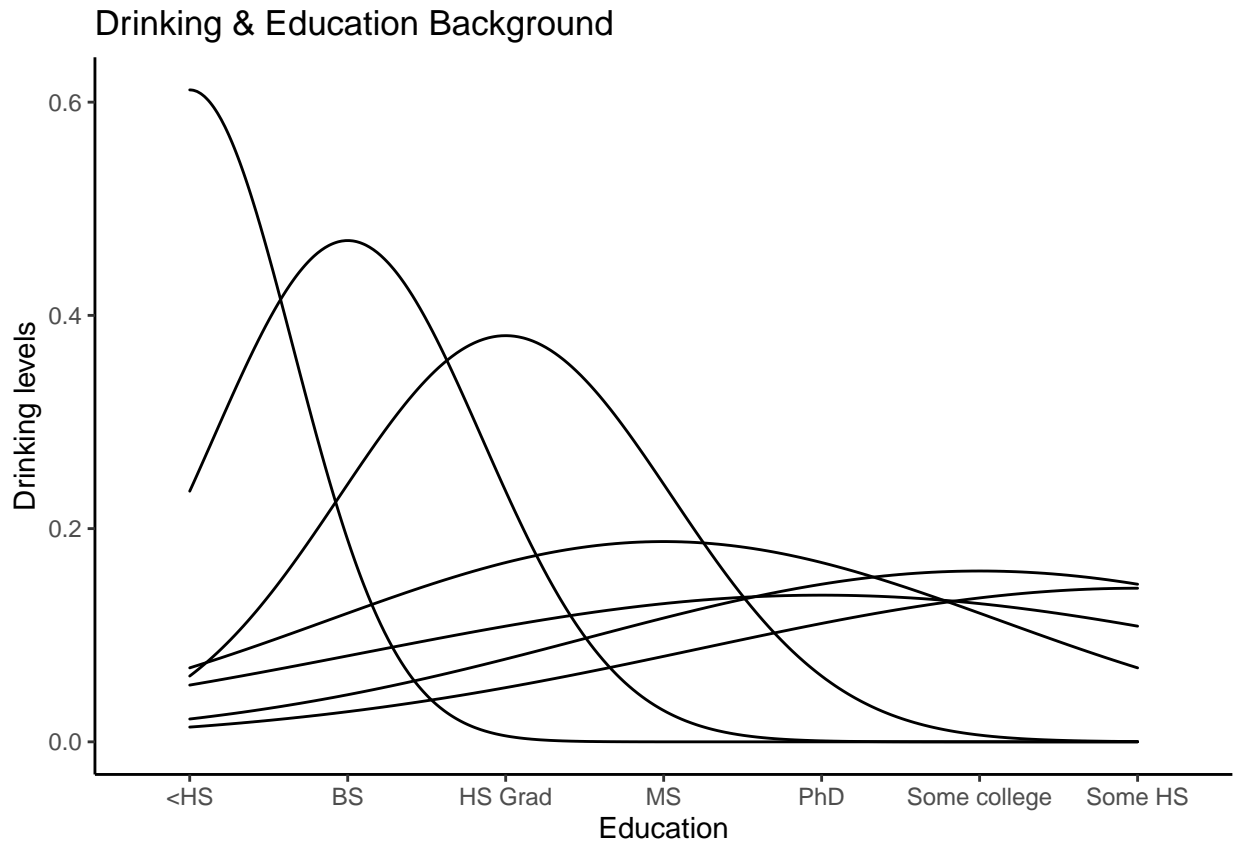

The reports of drinking levels by the participants background is shown and you can see low levels but also the peak for how many reported yes.

Conclusion:

```
head(depress[,c('educat', 'drink', 'sex')])
```

```
##             educat drink sex
## 1         Some HS     0   1
## 2 Some college       1   0
## 3         HS Grad     1   1
## 4         HS Grad     0   1
## 5         HS Grad     1   1
## 6         HS Grad     1   0
```

As stated above the average we see is more female who are depressed and have drinking tendencies.

After exploring the data and noticing that those who were observed depressed have a background of drinking through highschool and post college. You can clearly see that females had a higher drinking rate over men who stayed on the low side in this this study.