

Exploratory Data Analysis

Kenzie McDonald

2024-09-18

Introduction

The selected data set: Depression. The data set provides a prospective study of adult men and women in Los Angeles County that feel depressed. This set includes 294 observations but the focus variables will include age, education, and if they drink. Does education level and amount of alcohol consumed play a role in depression rates? I'm interested in finding out what sex experiences more depression and what factors attribute to it.

Univariate Exploration

```
depress <- read.delim("Depress.txt")
class(depress$DRINK)
```

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

```
depress$drank <- factor(depress$DRINK, labels=c("Yes", "No"))
table(depress$drank)
```

```
##
## Yes No
## 234 60
```

Out of the 294 observations 234 people state that they're a regular drinker and 60 people said they do not drink regularly.

```
class(depress$SEX)
```

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

```
depress$s <- factor(depress$SEX, labels=c("Male", "Female"))
table(depress$s)
```

```
##
## Male Female
## 111 183
```

183 of the 294 observations in this data set were female and 111 were males in this observation. There is an observed difference of 72 between the two sexes.

```
class(depress$EDUCAT)
```

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

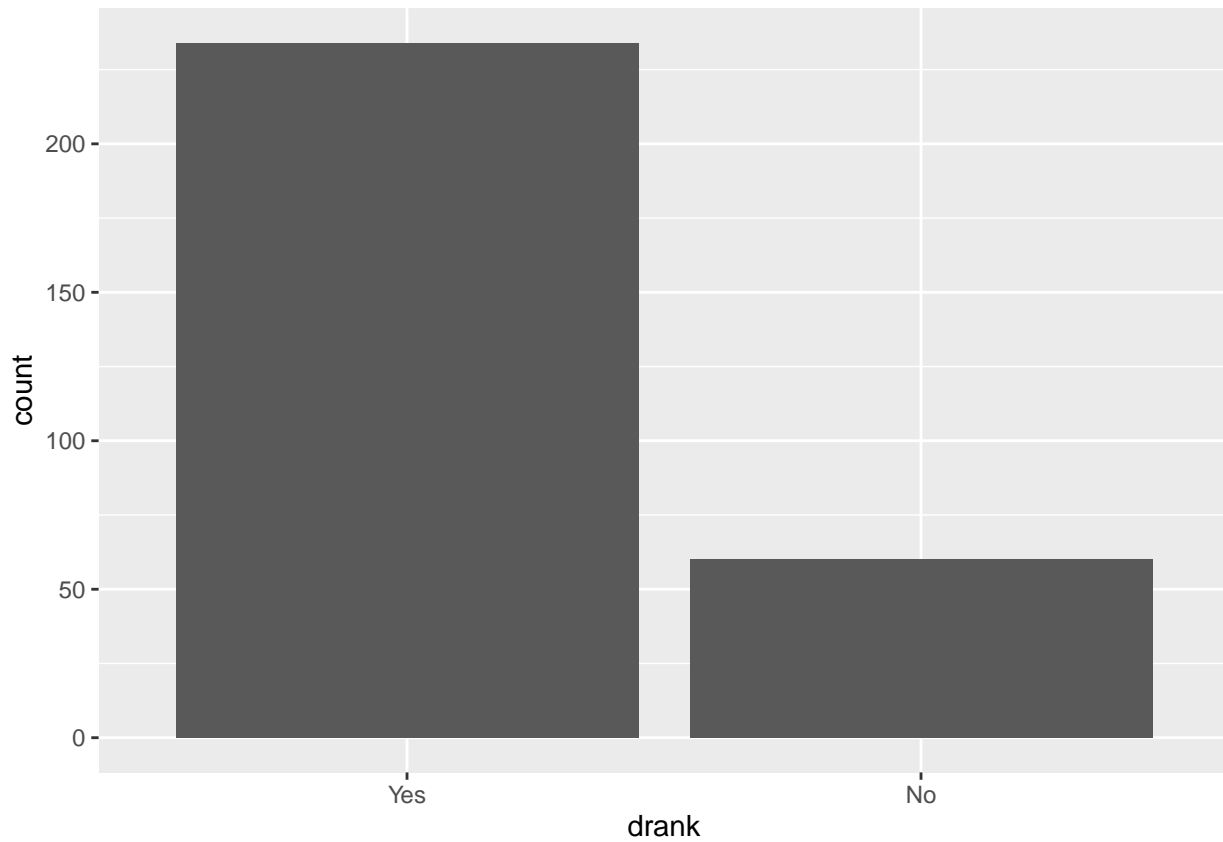
```
depress$education <- factor(depress$EDUCAT, labels=c("<HS", "Some HS", "Finished HS", "Some College", "B", "PHD"))
table(depress$education)
```

```
##
## <HS Some HS Finished HS Some College BS MS
## 5 61 114 48 43 14
## PHD
```

```
##          9
```

This table displays the level of education from each person that participated. From looking at this data most participants have a High School diploma. There are approximately equal amounts of participants that have some college experience or a Bachelor's degree.

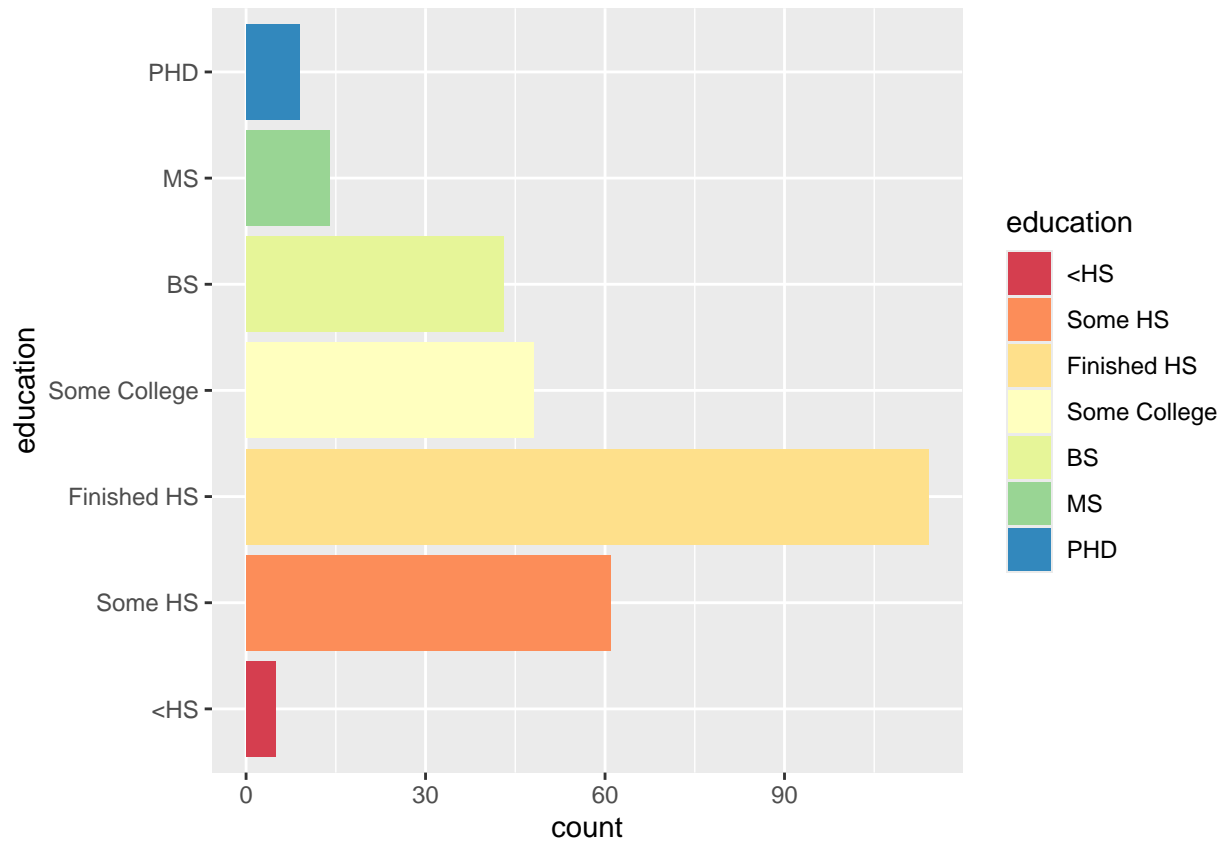
```
library(ggplot2)
ggplot(depress, aes(x=drank)) + geom_bar()
```



From looking at the graph you can see that there is a difference of about 175 of people who don't drink compared to people who do drink.

Educational Graph

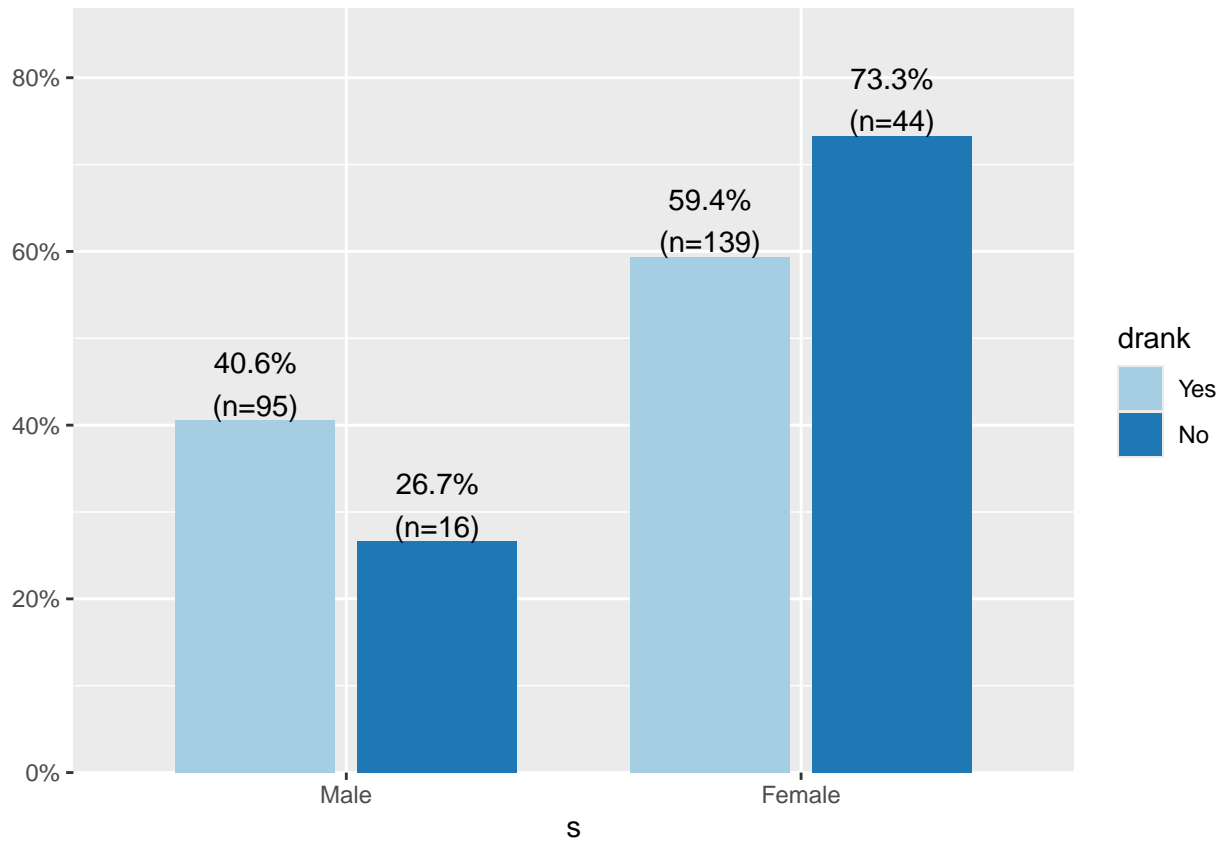
```
ggplot(depress, aes(x=education, fill=education)) + geom_bar() + scale_fill_brewer(palette="Spectral")
```



Now we can visualize the values by color and we can see that the majority of the observations have a high school diploma. In addition, there is about an equal amount of observations that have some college experience or have their bachelor's degree.

Bivariate Exploration

```
library(sjPlot)
plot_xtab(depress$s, depress$drank, show.total = FALSE)
```



Based on this graph we can see that 73.3% of females observed state that they drink regularly and 40.6% of the men observed state they drink regularly.

```
library(magrittr)
table(depress$drank, depress$s, depress$education)
```

```
## , , = <HS
##
##
##      Male Female
## Yes    4      1
## No     0      0
##
## , , = Some HS
##
##      Male Female
## Yes   15     27
## No    4     15
##
## , , = Finished HS
##
##      Male Female
## Yes   34     58
## No    5     17
##
## , , = Some College
```

```
##
##
##      Male Female
## Yes   15     24
## No    3      6
##
## , , = BS
##
##      Male Female
## Yes   16     21
## No    1      5
##
## , , = MS
##
##      Male Female
## Yes    5      5
## No     3      1
##
## , , = PHD
##
##      Male Female
## Yes    6      3
## No     0      0
```

This is an embedded table that shows the education levels of each participant along with their sex and whether or not they drink on a regular bases. You can see from this table that participants with a PHD or less than a High School Diploma 100% of the observed state they drink regularly. For the rest of the data set you can see that the majority of people drink regularly.

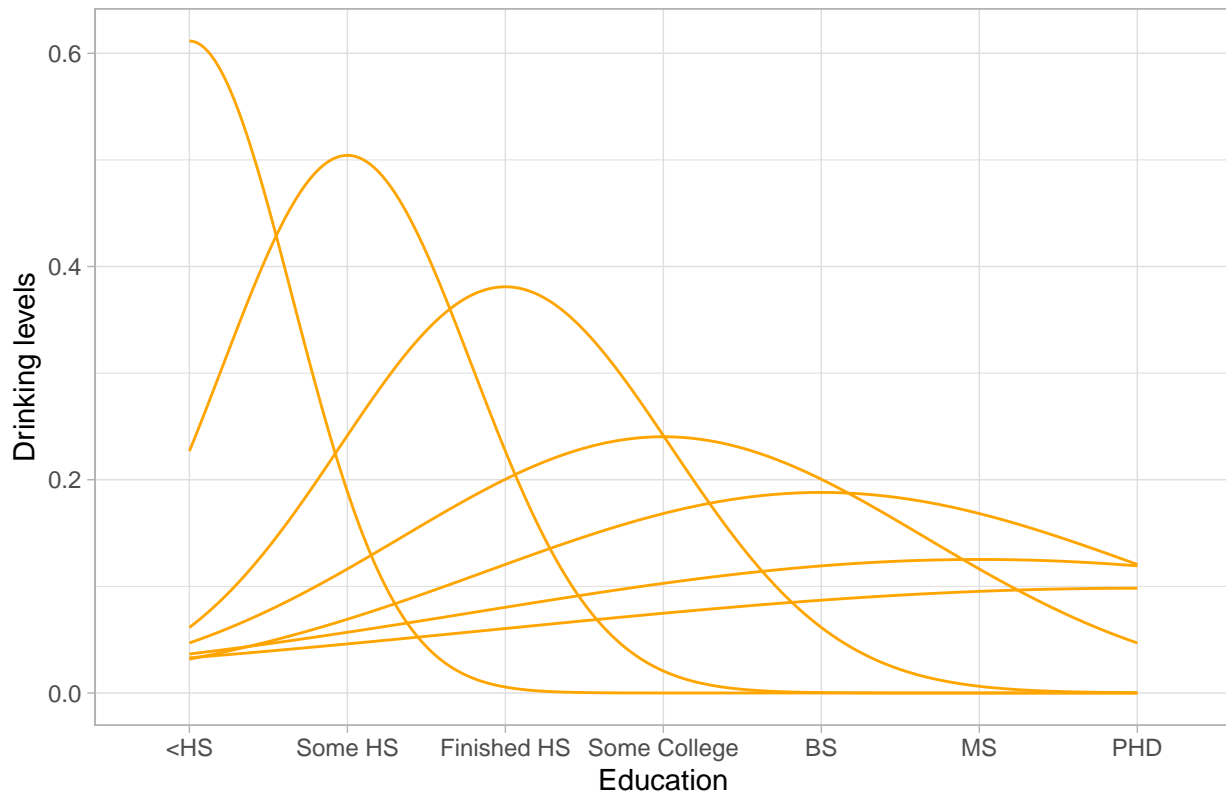
```
table(depress$education, depress$drank) %>% prop.table(margin=1) %>% round(4)
```

```
##
##              Yes    No
## <HS           1.0000 0.0000
## Some HS       0.6885 0.3115
## Finished HS   0.8070 0.1930
## Some College  0.8125 0.1875
## BS            0.8605 0.1395
## MS            0.7143 0.2857
## PHD           1.0000 0.0000
```

This table is a proportions table of education level against drinking provided to further prove the statement that the majority of the people observed drink regularly no matter.

```
ggplot(depress, aes(x=education)) + geom_density(col="orange") + ggtitle("Education Background + Drinking")
```

Education Background + Drinking



From looking at this graph you can see the background of the participants and in addition the low levels of drinking. The peaks represent the percentage of people who reported yes to drinking regularly.

Conclusion

```
head(depress[c('education', 'DRINK', 'SEX')])
```

```
##      education DRINK SEX
## 1      Some HS      2   2
## 2 Some College      1   1
## 3 Finished HS      1   2
## 4 Finished HS      2   2
## 5 Finished HS      1   2
## 6 Finished HS      1   1
```

For the “DRINK” column 1 represents yes and 2 = no. For the column “SEX” 1 = Male and 2 = Female. You can see that the average of this observation is Females who are depressed have a tendency to drink. After observing the data and noticing that those who were observed to be depressed have a drinking background. In addition it is clear that females had a higher drinking tendency then the males.