The final project

AUTHOR
Nano Omine

Introduction

For my final project, I chose Depression data set. This data set is a study on
depression by interviewing 294 adult residents in Los Angels county. It has 37
variables and I analyzed 2 variables of them; "cesd" and "sex". By analyzing them, I
observed the different depression level in sex.

```
depression <-read.table( "/Users/ominenano/Downloads/Chico State/222 FA
library(ggplot2)
```

Univariate Exploration

"cesd" variable

```
summary(depression$"cesd")
```

```
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  0.000   3.000   7.000   8.884  12.000  47.000
```

This is a summary statistics of "cesd". "cesd" tells the depression level of each
participant. In the data set, the maximum score is 60 and the minimum score is 0. The
mean is 8.9 and the median is 7. The highest score is 47 and the lowest score 0. The
25th percentile of this data has a score of 3 or less, and the 75th percentile has a
score of 12 or less. This data is skewed to right because the mean is bigger than the
median.

```
(p<- ggplot(data = depression, aes(cesd)) +
     geom_histogram(bins = 21) +
     labs(x = "Level of Depression"))
```
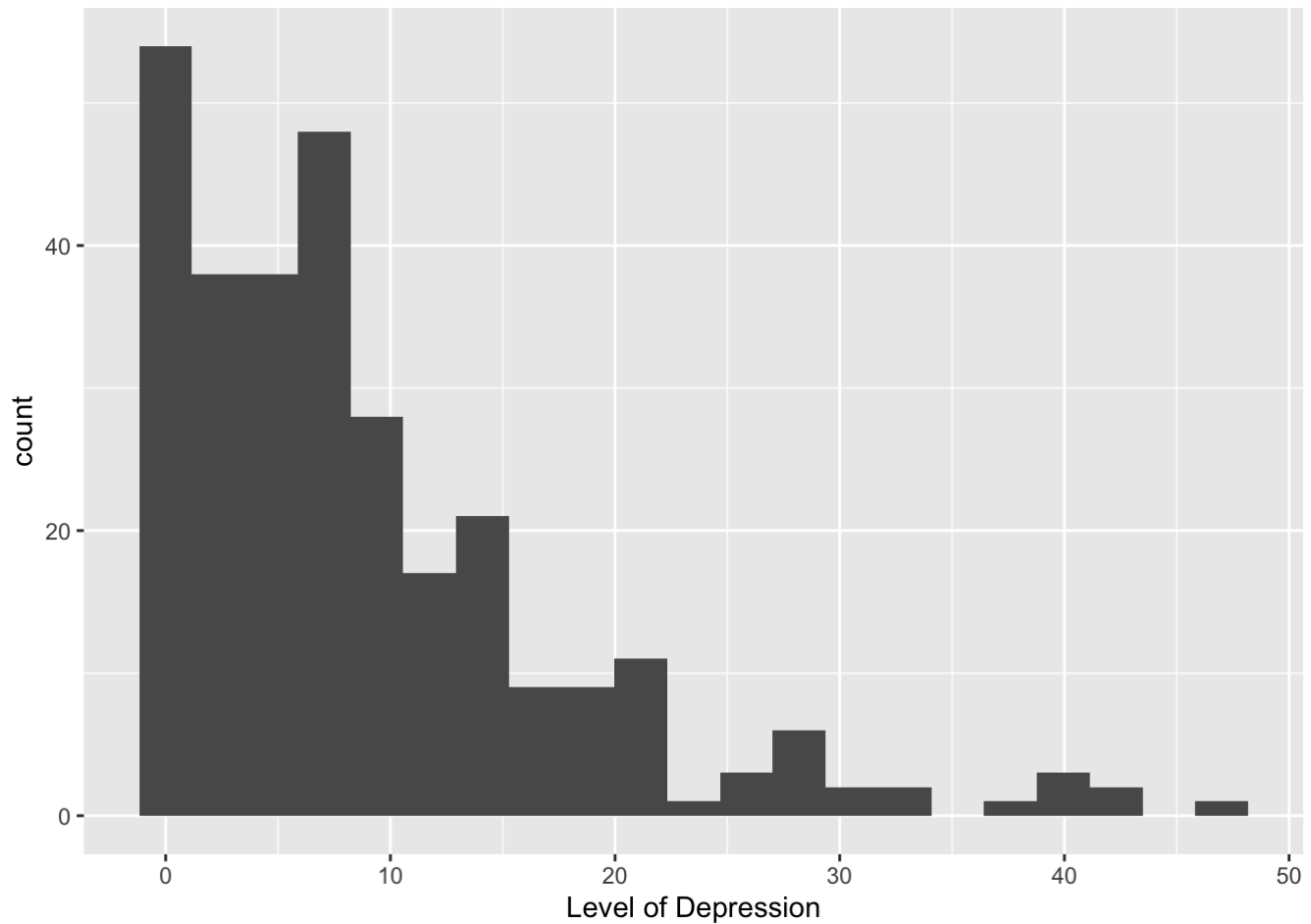
This data is skewed to the right. Most of participants' depression levels are lower than 30.

"sex" Variable

```
table(depression$sex)
```
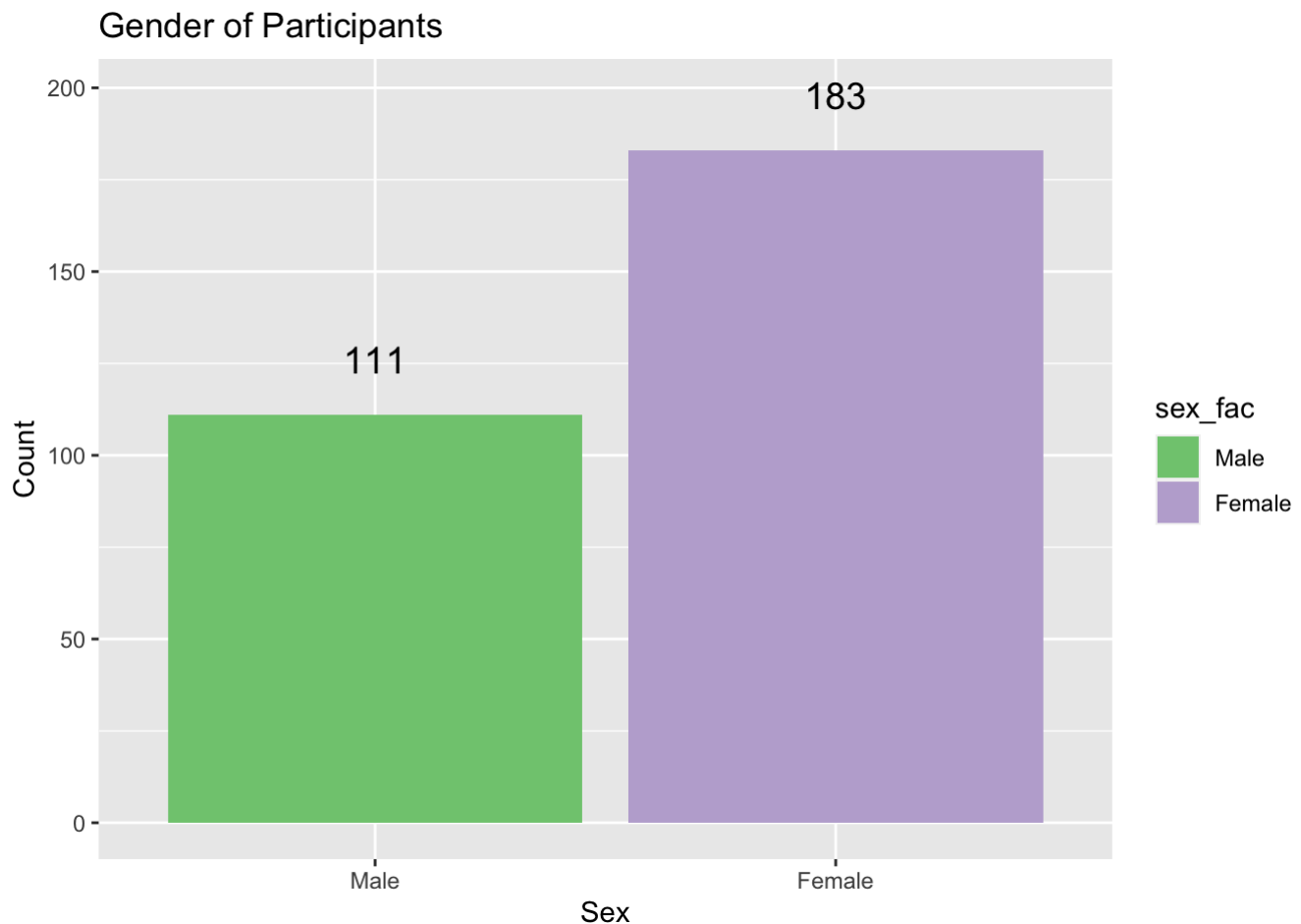
```
  0   1
111 183
```

```
class(depression$sex)
```

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

```
depression$sex_fac <- factor(depression$sex, labels=c("Male", "Female"))
table(depression$sex_fac)
```

```
 Male Female
  111    183
```

```
ggplot(depression, aes(x=sex_fac, fill= sex_fac)) +
geom_bar(aes(y=..count..)) + scale_fill_brewer(palette= "Accent") +
ggtitle("Gender of Participants") + ylab("Count") + xlab("Sex") + geom_
stat='count', size = 5)
```

Gender of Participants

The population of this study includes 111 males and 183 females.

Bivariate Exploration:

```
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

```
summarise(group_by(depression,sex_fac), Mean=mean(cesd), median(cesd))
```

```
# A tibble: 2 × 3
  sex_fac  Mean `median(cesd)`
  <fct>    <dbl>         <int>
1 Male      7.49             6
2 Female    9.73             7
```

This is a summary statistics between two variables; "sex" and "cesd". For depression level in males, the mean is 7.5 and the median is 6. For depression level in female, the mean is 9.7 while the median is 7. Female has a higher rate of depression.

```
ggplot(depression, aes(x=cesd, fill=sex_fac)) + geom_density(alpha=.3)
scale_fill_discrete(name="Sex") + ggtitle("
Relationship Between Sex and Depression Level") + xlab("CESD Score") +
ylab("Density") + theme_bw()
```

Relationship Between Sex and Depression Level
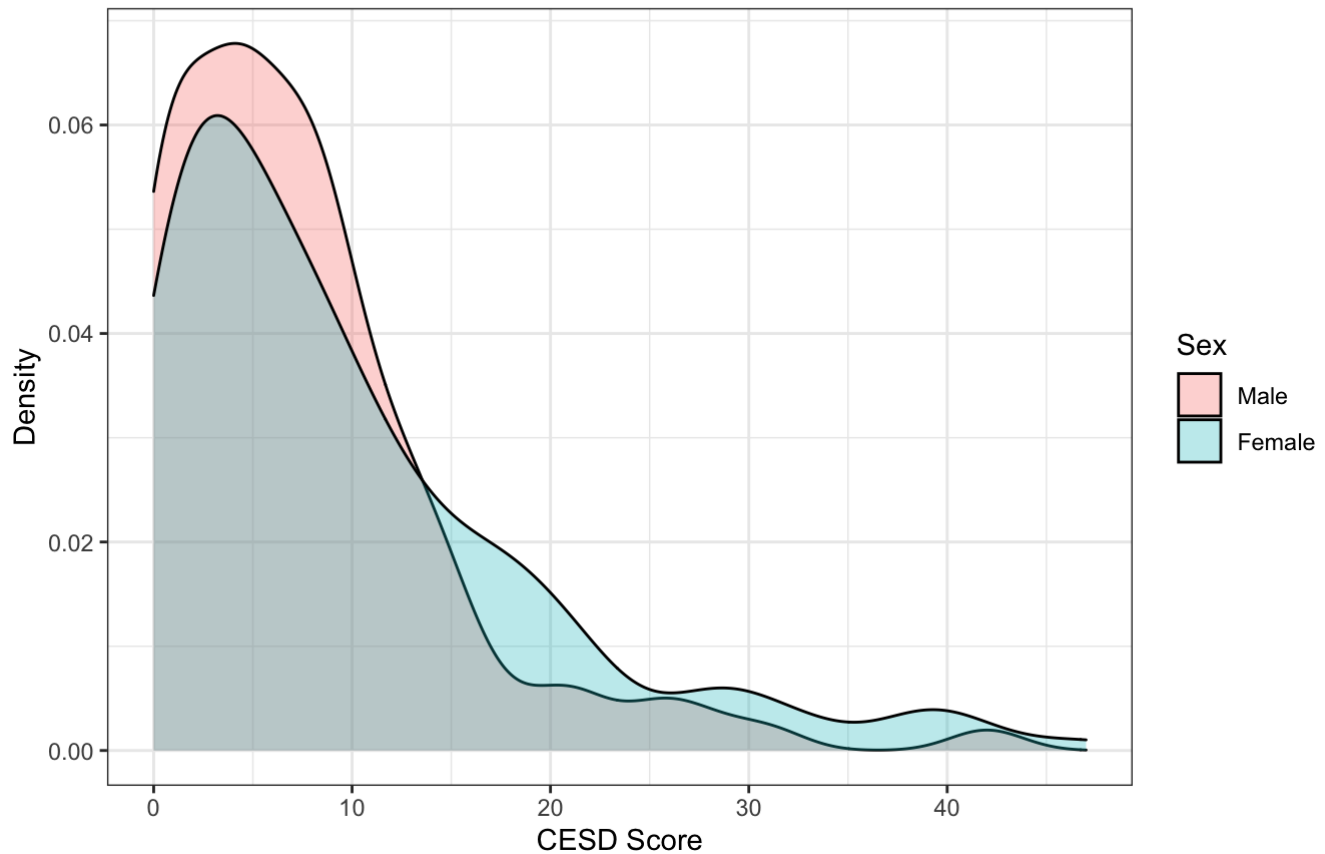
This is a density plot of the relationship of depression level in males and females. Pastel blue shows CESD score density in females and pastel pink shows CESD score density in males. Females' CESD score density is higher when CESD score is higher than males' one. On the other hand, when CESD is lower, males' CESD density is higher than females'. That explains females tend to feel more depressed than males.

Conclution

By comparing two variables, depression level is observed and the different rate of depression in sex is identified. In general, participants didn't feel extremely depressed. Among females and males, females tend to have higher depression rate than males.