

Math 130 Final Project- Depression Data Set

Gerardo Tapia

2024-09-30

Introduction

The Depression data set helps us understand people from all backgrounds and their level of depression based on survey questions. Exploring this data set could give a better understanding of what type of people deal with depression at higher rates than others. In this data set I will be exploring the variables: gender, CESD, employment. I would like to gain further understanding if women or men deal with depression at higher rates depending on their employment status. Without any given background information, men may deal with higher rates of depression than woman based off of their employment status.

Univariate variable

SEX

```
Depress <- Depress %>%  
  mutate(SEX = recode(SEX, `1` = "Male", `2` = "Female"))  
table(Depress$SEX)
```

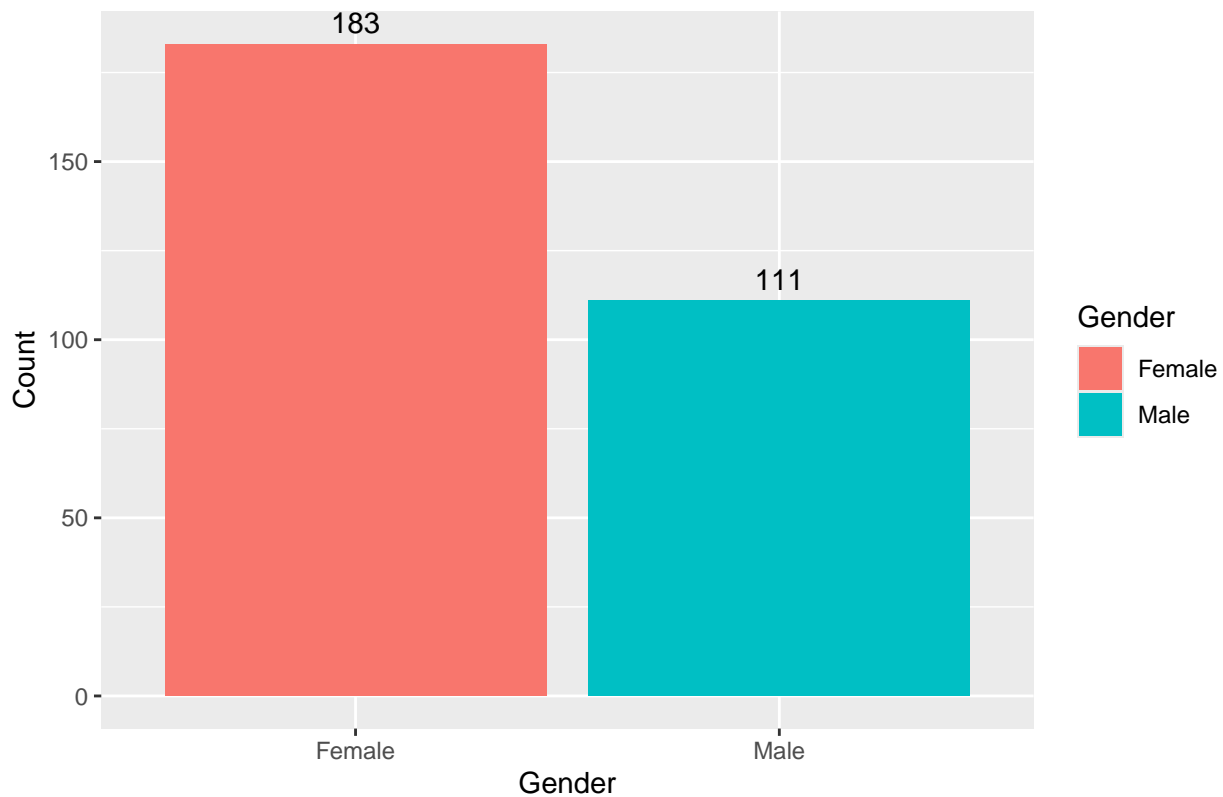
```
##  
## Female    Male  
##    183     111
```

The bar graph below shows the number of males and females included in this study. It is important to note that the variable observations names were recoded to help further understand the graph and its details. In this study, of those surveyed, 183 were female, and 111 were male.

```
ggplot(Depress, aes(x = SEX, fill = SEX)) +  
  geom_bar() +  
  labs(title = "Distribution of Gender in the Dataset", x = "Gender",  
        y = "Count", fill = "Gender") + geom_text(stat = "count",  
                                                  aes(label = ..count..),  
                                                  vjust= -0.5)
```

```
## Warning: The dot-dot notation (`..count..`) was deprecated in ggplot2 3.4.0.  
## i Please use `after_stat(count)` instead.  
## This warning is displayed once every 8 hours.  
## Call `lifecycle::last_lifecycle_warnings()` to see where this warning was  
## generated.
```

Distribution of Gender in the Dataset



Employment(EMPLOY)

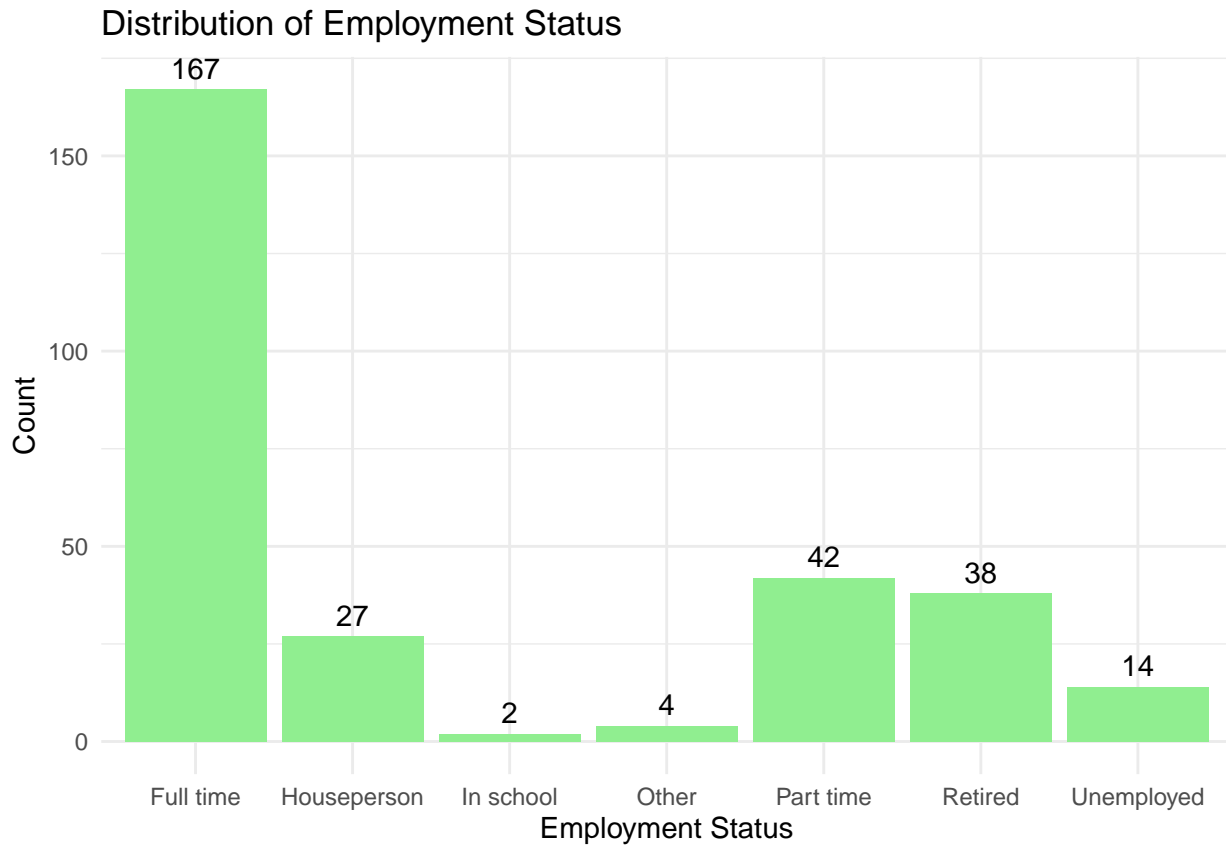
```
Depress <- Depress %>%  
  mutate(EMPLOY = recode(EMPLOY,  
    `1` = "Full time",  
    `2` = "Part time",  
    `3` = "Unemployed",  
    `4` = "Retired",  
    `5` = "Houseperson",  
    `6` = "In school",  
    `7` = "Other"))  
table(Depress$EMPLOY)
```

```
##  
##   Full time Houseperson   In school   Other   Part time   Retired  
##     167           27         2         4         42         38  
## Unemployed  
##     14
```

The table and graph show the distribution of employment status amongst those surveyed in this study. The study gave integers to represent each of the observations. I recoded the number integers to represent actual employment status.

```
ggplot(Depress, aes(x = factor(EMPLOY))) +  
  geom_bar(fill = "lightgreen") +  
  labs(title = "Distribution of Employment Status", x = "Employment Status",  
    y = "Count") + geom_text(stat = "count", aes(label = ..count..),
```

```
vjust= -0.5) + theme_minimal()
```

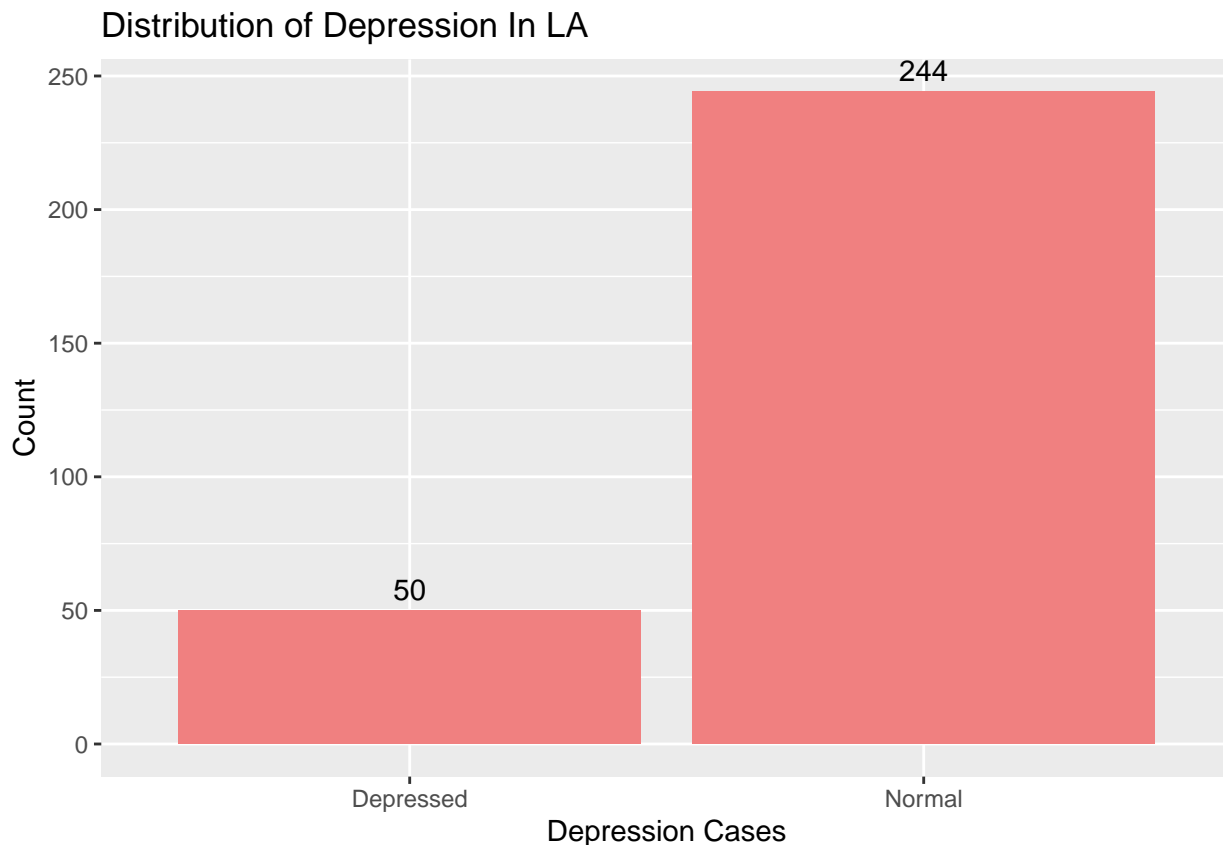


Cases

```
Depress <- Depress %>% mutate(CASES = recode(CASES, '0' = "Normal", '1' = "Depressed"))  
table(Depress$CASES)
```

```
##  
## Depressed   Normal  
##          50     244
```

```
ggplot(Depress, aes(x = factor(CASES))) +  
  geom_bar(fill = "lightcoral") +  
  labs(title = "Distribution of Depression In LA", x = "Depression Cases",  
        y = "Count") + geom_text(stat = "count", aes(label = ..count..),  
                                 vjust= -0.5)
```



It is important to note that in this study those who were deemed “depressed” had a score CSED= ≤ 16 . The CSED variable showcases the total sum from questions asked where the observations are labeled as followed: 1 = Some or a little often you felt or behaved this way during the past (1 to 2 days) 2 = Occasionally or a moderate amount of time (3 to 4 days) 3 = Most or all of the time (5 to 7 days) Overall, the study shows the distribution of people who were labeled as Normal or Depressed. Only 50 people resulted in being ‘depressed’ and 244 persons were labeled as ‘Normal’

Bivariate Variable

```
Depress %>% select(SEX,EMPLOY, CASES) %>%
tbl_summary(by = "EMPLOY")
```

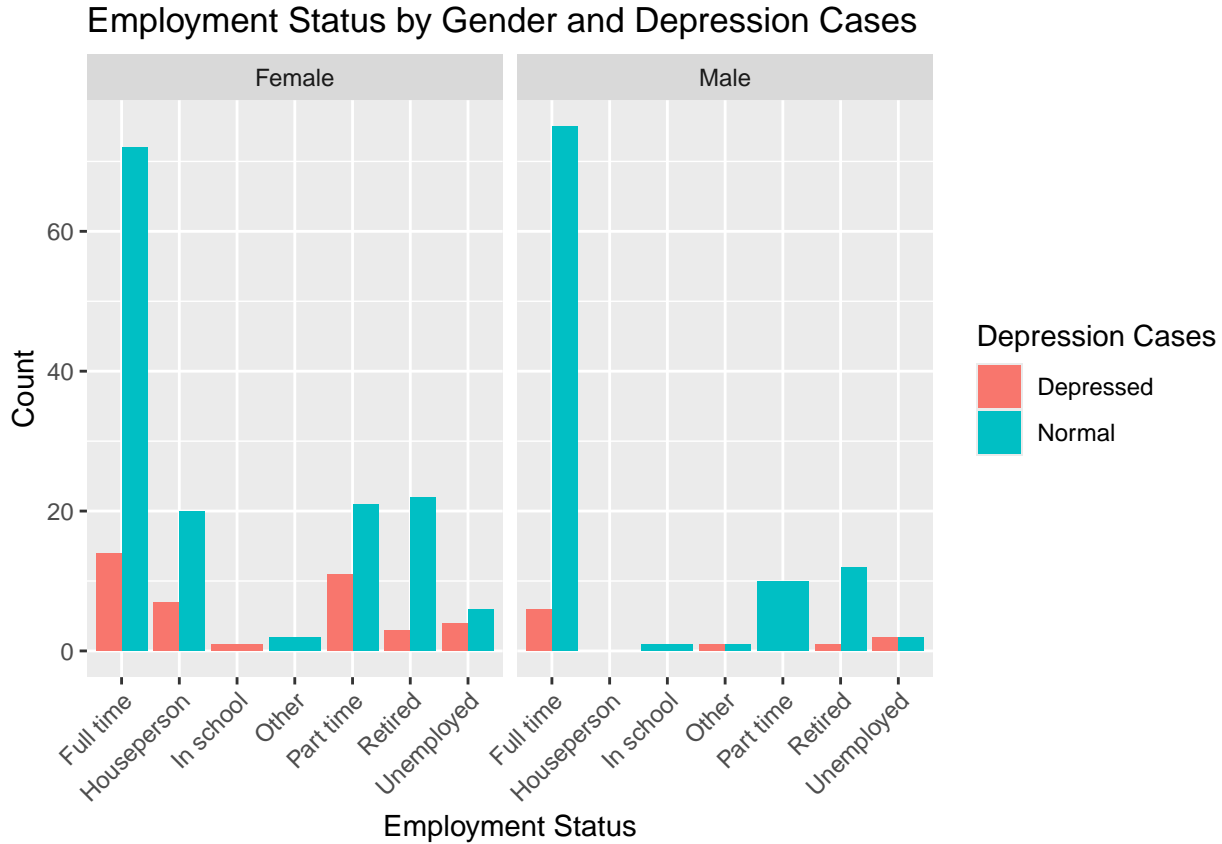
Characteristic	Full time N = 167 ¹	Houseperson N = 27 ¹	In school N = 2 ¹	Other N = 4 ¹
SEX				
Female	86 (51%)	27 (100%)	1 (50%)	2 (50%)
Male	81 (49%)	0 (0%)	1 (50%)	2 (50%)
CASES				
Depressed	20 (12%)	7 (26%)	1 (50%)	1 (25%)
Normal	147 (88%)	20 (74%)	1 (50%)	3 (75%)

¹n (%)

The table represents percentages of cases were people were depressed based on their employment status. The

table also shows number and percentages of male and female who where labeled as depressed or normal.

```
ggplot(Depress, aes(x = EMPLOY, fill = factor(CASES))) +
  geom_bar(position = "dodge") +
  facet_wrap(~ SEX) +
  labs(title = "Employment Status by Gender and Depression Cases",
        x = "Employment Status", y = "Count", fill = "Depression Cases") + theme(axis.text.x = element_t
```



Conclusion

In conclusion, through graphing and getting to know the variables, the hypothesis is rejected. In fact, females who worked full-time showed higher rates of depression, than males who worked full-time. In all types of employment status, women showed higher rates of depression cases based on this study. Lastly, you could also argue that the rate of people whose results were ‘normal’ where substantially higher; which could have skewed the data.