

Exploratory Data Analysis

Jennell Hall

2023-09-20

```
## 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

```
library(forcats)
library(knitr)
library(sjPlot)
```

```
depress <- read.delim("/Users/nellz/Desktop/Math 130/homework/Notes /data/depress_081217.txt", header=T)
dim(depress)
```

```
## [1] 294 37
```

```
library(ggplot2)
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
```

1. Introduction This data set is adult residents in Los Angeles County that are being interviewed about depression. It has a data set of observations and 37 variables. If you want to get more 'details about the study it can be found in Practical Multivariate Analysis, 5th edition by Afifi, May and Clark. I will be examining the relationship between sex, depression and age those are my two variables.

2. Univariate Descriptions Females and Males: The table shows how many males and females were interviewed and volunteered to be in the study

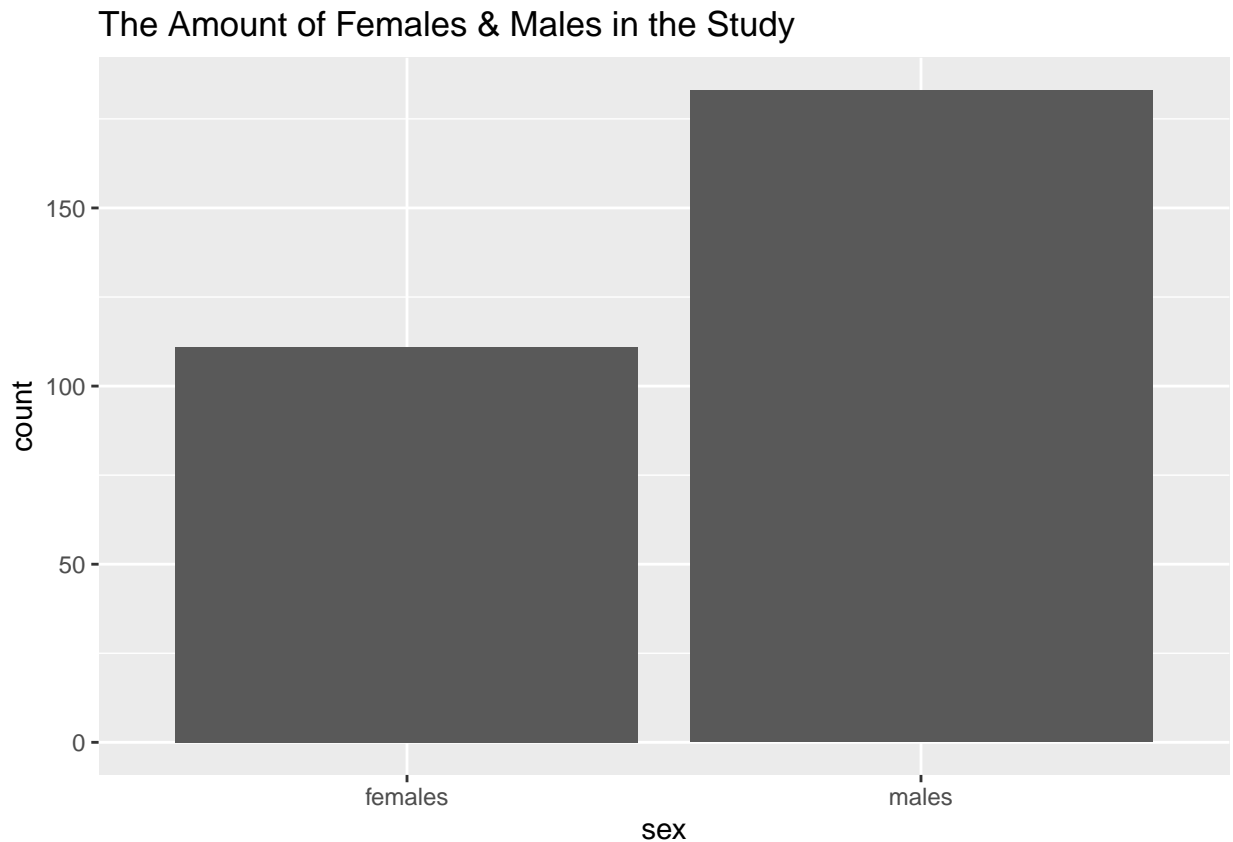
```
summary(depress$sex)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
## 0.0000  0.0000   1.0000  0.6224  1.0000  1.0000
```

```
depress$sex <- factor(depress$sex, labels = c("females", "males"))
table(depress$sex)
```

```
##
## females  males
##    111    183
```

```
ggplot(depress, aes(x=sex)) + geom_bar() + ggtitle("The Amount of Females & Males in the Study ")
```



The bar shows how many more males there were in the study and how few of the females actually participate in the study. Maybe that will make a difference in my research question.

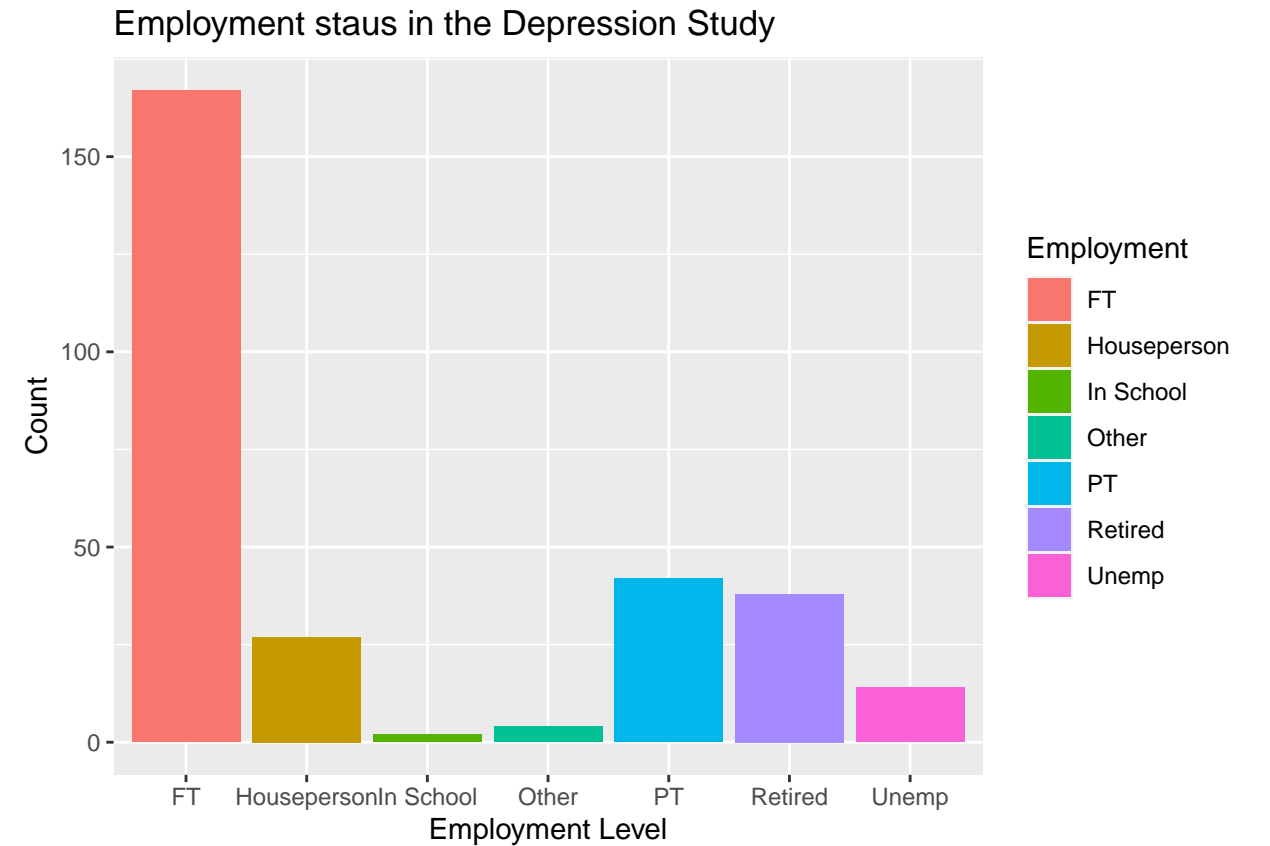
```
table(depress$employ)
```

```
##
##      FT Houseperson  In School    Other    PT    Retired
```

```
##          167          27          2          4          42          38
##      Unemp
##          14
```

This table shows how many people in th study that are retired,in school, unemployment, full time, house-
personand part time

```
ggplot(depress, aes(x=employ, fill=employ))+ geom_bar()+ ggtitle("Employment staus in the Depression S
```



The bar graph shows the individuals levels of employment.As you can see in the graph most people in the study are full time. A lower percent of them are in school or unemployment.

3.Bivariate Comparison

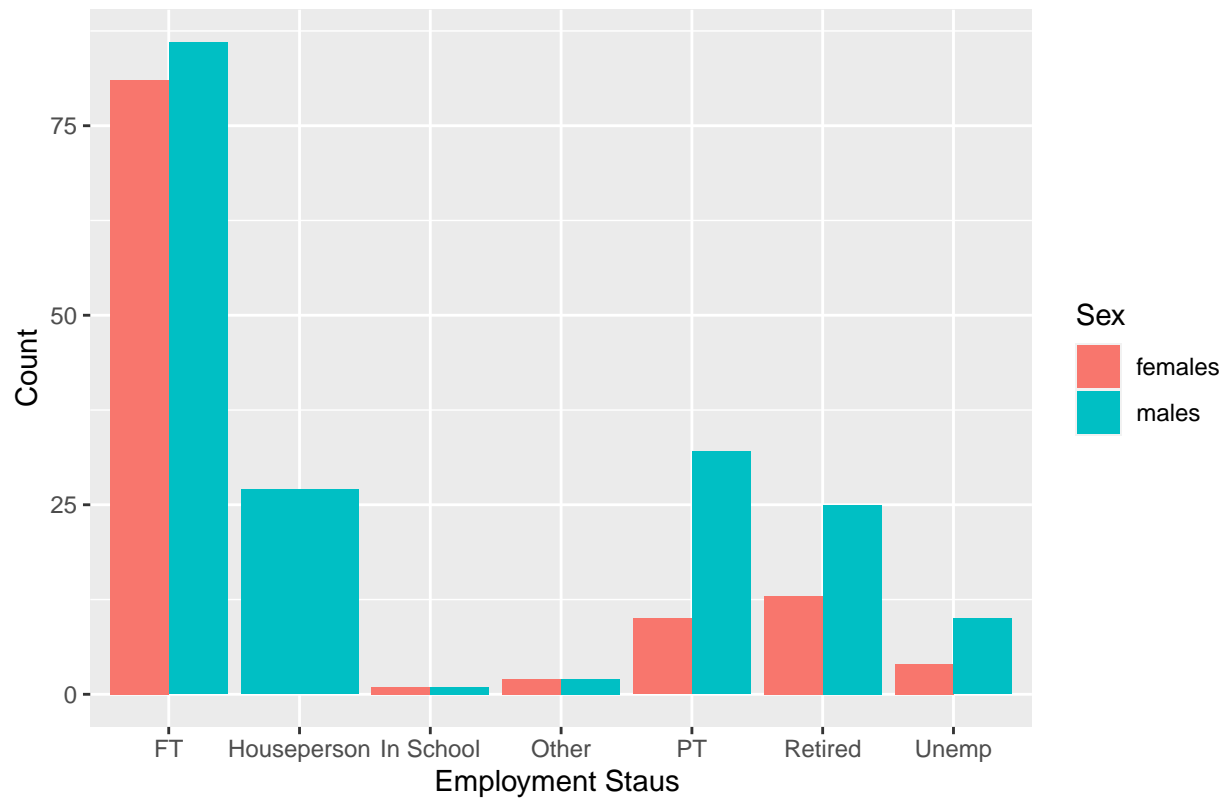
```
table(depress$sex, depress$employ)
```

```
##
##      FT Houseperson In School Other PT Retired Unemp
## females 81          0          1  2 10      13      4
## males  86          27          1  2 32      25     10
```

The table above shows the different levels of employment status from females and males

```
ggplot(depress, aes(x=employ, fill=sex)) + geom_bar(position = "dodge")+ ggtitle("Employment Level of I
```

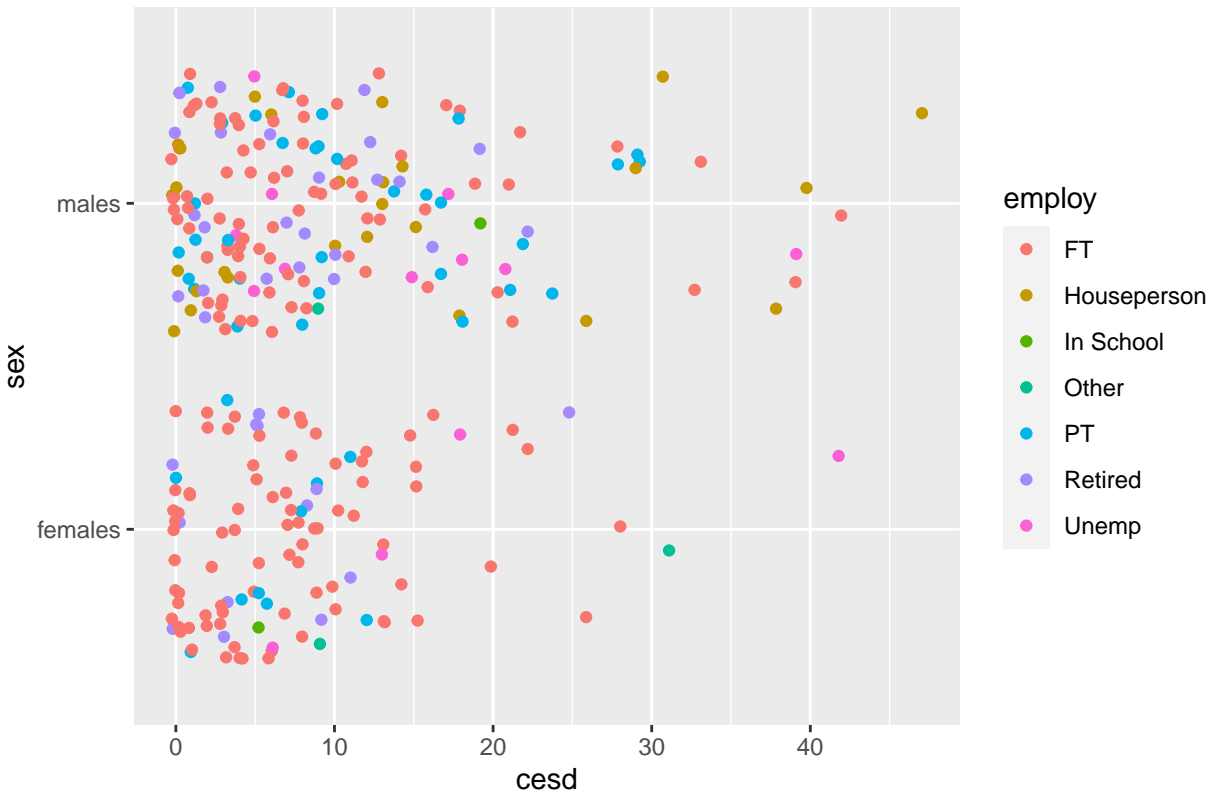
Employment Level of Males and Females in the Depression Study



Shown in the graph you can see the more males that have a full time job were more depressed then females who had a full time job

```
ggplot(depress, aes(x =cesd, y =sex , color = employ)) +  
  geom_jitter(width = .3) + ggtitle("Sex & Employment vs.CESD Scores")
```

Sex & Employment vs.CESD Scores



The Females that a full time job had an higher CESD score.

4.Conclusion The average female who had a full-time job were the ones who were more depressed and a high CESD score. The males were more depressed then the females.