

Exploratory Data Analysis Project

Lanie Blomquist

09/19/2022

```
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
```

```
library(ggplot2)
```

Introduction

I chose to analyze the Depression data set taken from a study done on Los Angeles County adults. The data set includes 294 observations and 37 variables and I will be studying sex, marital status and level of depression. Sex is simply categorized by male or female, marital status is categorized into never married, married, divorced, separated, and widowed, and level of depression is ranked on a scale from 0-60, 0 being the lowest level of depression, and 60 the highest. I am interest to see whether there a correlations between gender and levels of depression, marital status and levels of depression, and relationships between all three variables. I anticipate that there will be a slight correlation between gender and levels of depression and an even more prominent relationship between marital status and levels of depression.

```
depress <- read.delim("/Users/lanieblomquist/Desktop/math130/data/depress_081217.txt", header=TRUE, sep=  
dim(depress)
```

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

Univariate Exploration

Levels of Depression

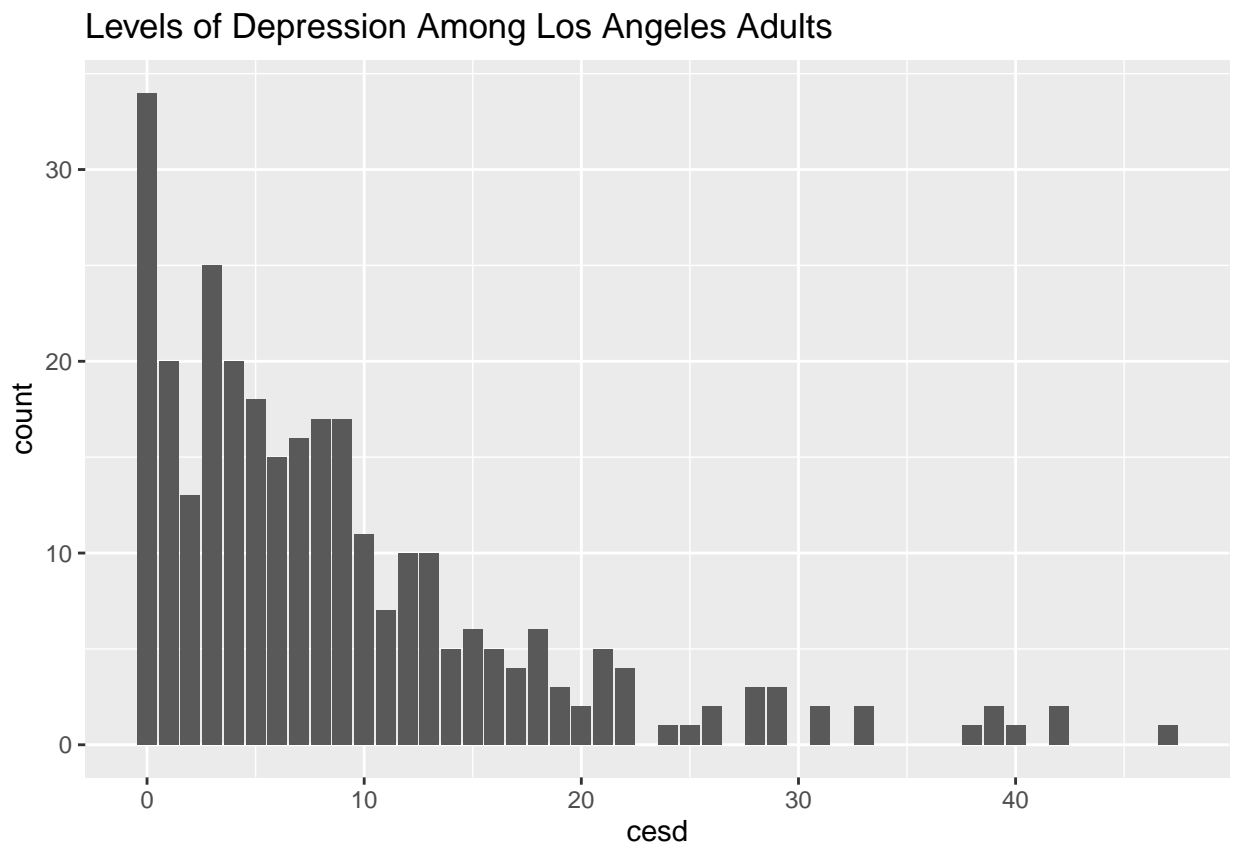
```
summary(depress$cesd)
```

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

```
table(depress$cesd)
```

```
##
##  0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 24 25 26
## 34 20 13 25 20 18 15 16 17 17 11  7 10 10  5  6  5  4  6  3  2  5  4  1  1  2
## 28 29 31 33 38 39 40 42 47
##  3  3  2  2  1  2  1  2  1
```

```
ggplot(depress,aes(x=cesd))+geom_bar()+ggtitle("Levels of Depression Among Los Angeles Adults")
```



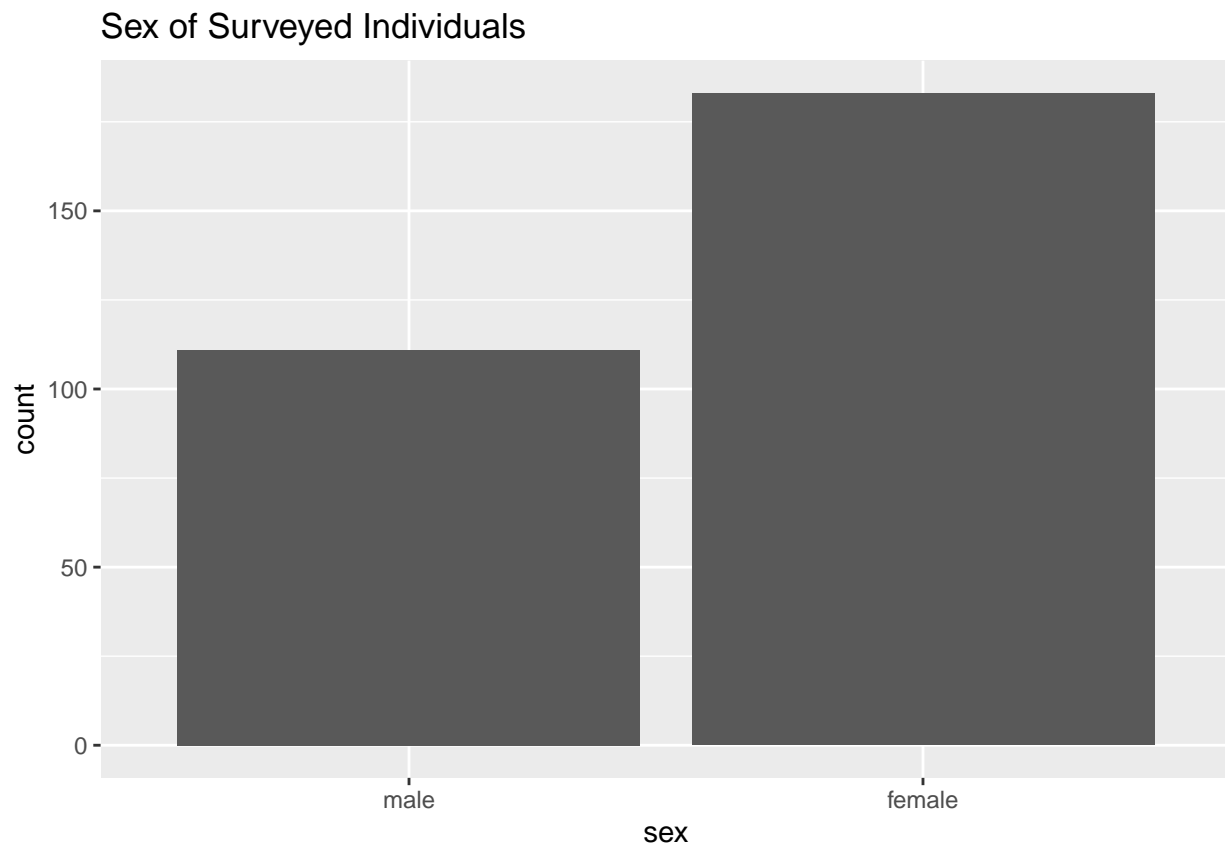
This data shows the spread of levels of depression among Los Angeles adults. The summary stats and table display the relative frequency statistics while the bar chart represents the abundance of each depression level. We can see the chart is skewed right, meaning that the majority of Los Angeles adults feel they share a lower depression level. Some individuals do fall into higher depression level categories, but many individuals reported a depression level of over 40 can relatively be considered outliers.

Sex

```
depress$sex<-factor(depress$sex, label=c("male","female"))  
summary(depress$sex)
```

```
##  male female  
##   111   183
```

```
ggplot(depress,aes(x=sex))+geom_bar()+ggtitle("Sex of Surveyed Individuals")
```



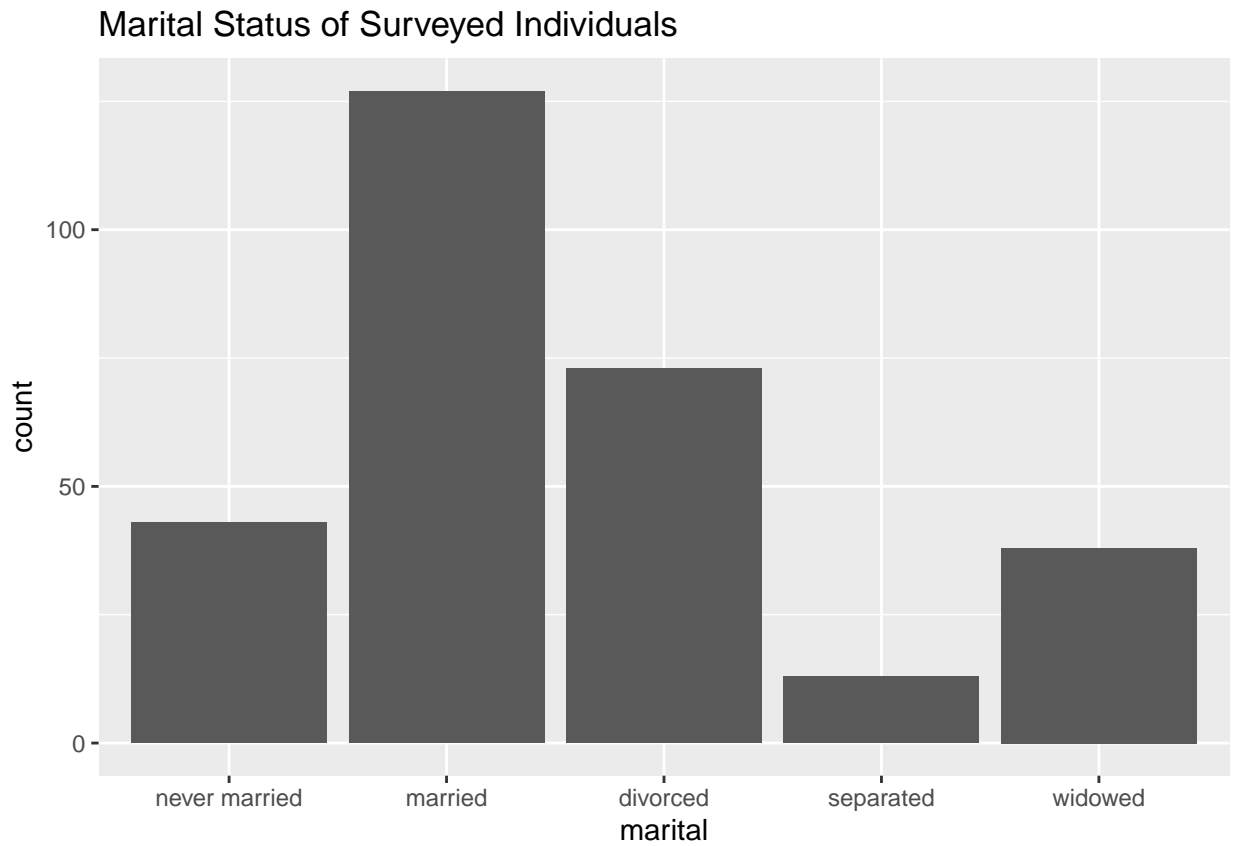
This data and bar chart show the sex of each of the surveyed individuals. We can see that the majority of surveyed individuals were female, having roughly 70 more participants than males.

Marital Status

```
depress$marital<-factor(depress$marital, label=c("never married","married","divorced","separated","widowed"))  
summary(depress$marital)
```

```
##  never married    married    divorced    separated    widowed  
##           43         127         73         13         38
```

```
ggplot(depress, aes(x=marital))+geom_bar()+ggtitle("Marital Status of Surveyed Individuals")
```



This data shows the marital status of each participant. It categorizes each individual and shows us that the majority of the surveyed individuals were married and very few were separated.

Bivariate Exploration

```
table(depress$cesd, depress$sex) %>%prop.table(margin=2) %>%round(3)
```

```
##  
##      male female  
##  0  0.153  0.093  
##  1  0.045  0.082  
##  2  0.045  0.044  
##  3  0.081  0.087  
##  4  0.063  0.071  
##  5  0.090  0.044  
##  6  0.045  0.055  
##  7  0.063  0.049  
##  8  0.072  0.049  
##  9  0.072  0.049  
## 10  0.036  0.038  
## 11  0.027  0.022
```

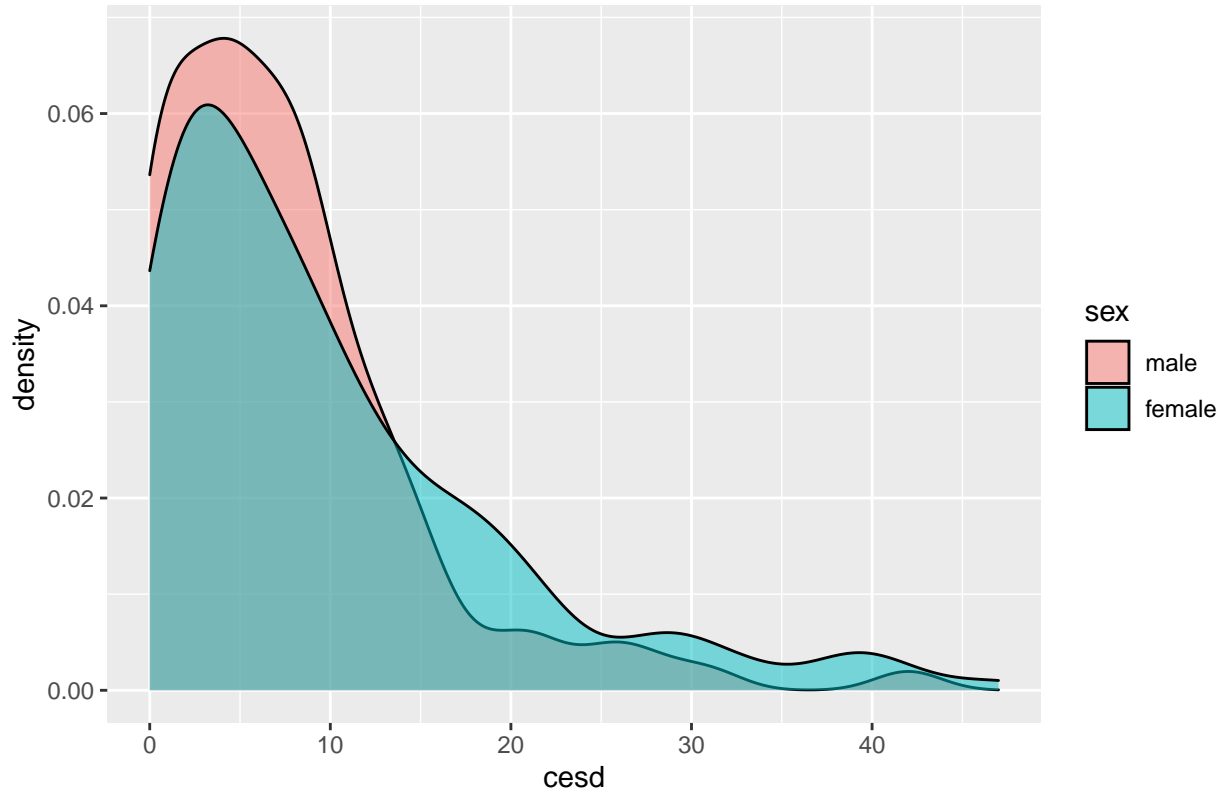
```
## 12 0.036 0.033
## 13 0.036 0.033
## 14 0.009 0.022
## 15 0.036 0.011
## 16 0.009 0.022
## 17 0.000 0.022
## 18 0.009 0.027
## 19 0.000 0.016
## 20 0.009 0.005
## 21 0.009 0.022
## 22 0.009 0.016
## 24 0.000 0.005
## 25 0.009 0.000
## 26 0.009 0.005
## 28 0.009 0.011
## 29 0.000 0.016
## 31 0.009 0.005
## 33 0.000 0.011
## 38 0.000 0.005
## 39 0.000 0.011
## 40 0.000 0.005
## 42 0.009 0.005
## 47 0.000 0.005
```

```
depress%>% group_by(sex)%>%summarize(avg=mean(cesd))
```

```
## # A tibble: 2 x 2
##   sex      avg
##   <fct> <dbl>
## 1 male    7.49
## 2 female  9.73
```

```
ggplot(depress,aes(x=cesd,fill=sex))+geom_density(alpha=.5)+ggtitle("levels of depression compared to s
```

levels of depression compared to sex



This data shows the relationship of gender to level of depression. It demonstrates that, on average, females record higher levels of depression than males.

```
table(depress$cesd, depress$marital) %>%prop.table(margin=2) %>%round(3)
```

```
##
##      never married married divorced separated widowed
##  0      0.070    0.150    0.055    0.077    0.184
##  1      0.047    0.102    0.027    0.000    0.079
##  2      0.140    0.016    0.014    0.000    0.105
##  3      0.047    0.102    0.055    0.077    0.132
##  4      0.070    0.063    0.096    0.000    0.053
##  5      0.093    0.047    0.096    0.077    0.000
##  6      0.023    0.047    0.068    0.154    0.026
##  7      0.070    0.039    0.068    0.000    0.079
##  8      0.070    0.055    0.068    0.077    0.026
##  9      0.047    0.071    0.041    0.077    0.053
## 10      0.047    0.016    0.068    0.000    0.053
## 11      0.023    0.031    0.027    0.000    0.000
## 12      0.047    0.024    0.041    0.077    0.026
## 13      0.023    0.047    0.027    0.000    0.026
## 14      0.000    0.016    0.027    0.077    0.000
## 15      0.023    0.039    0.000    0.000    0.000
## 16      0.047    0.008    0.014    0.000    0.026
## 17      0.000    0.016    0.014    0.000    0.026
## 18      0.000    0.024    0.027    0.077    0.000
```

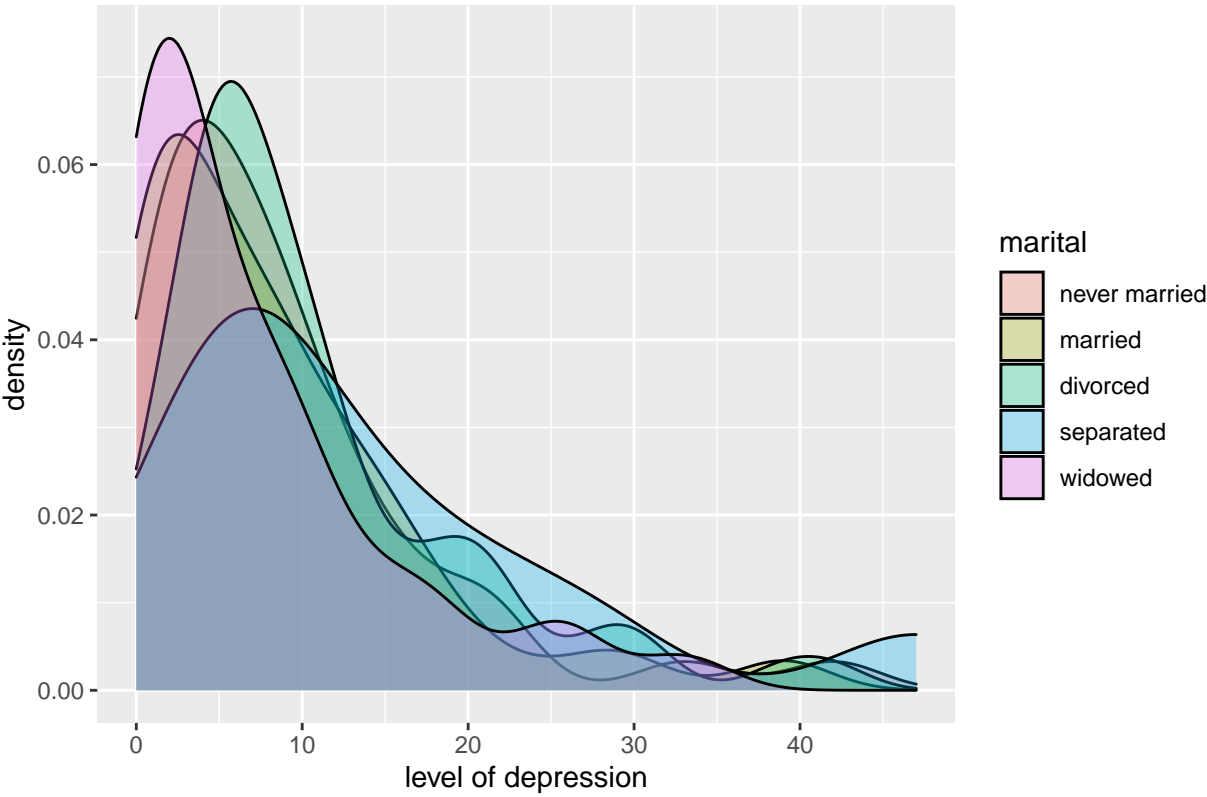
```
## 19      0.000  0.008  0.014  0.000  0.026
## 20      0.000  0.008  0.014  0.000  0.000
## 21      0.070  0.000  0.027  0.000  0.000
## 22      0.000  0.008  0.027  0.077  0.000
## 24      0.000  0.008  0.000  0.000  0.000
## 25      0.000  0.000  0.000  0.000  0.026
## 26      0.000  0.000  0.014  0.000  0.026
## 28      0.000  0.016  0.000  0.077  0.000
## 29      0.000  0.008  0.027  0.000  0.000
## 31      0.000  0.008  0.014  0.000  0.000
## 33      0.023  0.000  0.000  0.000  0.026
## 38      0.000  0.008  0.000  0.000  0.000
## 39      0.000  0.008  0.014  0.000  0.000
## 40      0.000  0.008  0.000  0.000  0.000
## 42      0.023  0.000  0.014  0.000  0.000
## 47      0.000  0.000  0.000  0.077  0.000
```

```
depress%>% group_by(marital)%>%summarize(avg=mean(cesd))
```

```
## # A tibble: 5 x 2
##   marital      avg
##   <fct>      <dbl>
## 1 never married  8.70
## 2 married       8.10
## 3 divorced     10.5
## 4 separated    13.7
## 5 widowed      7.05
```

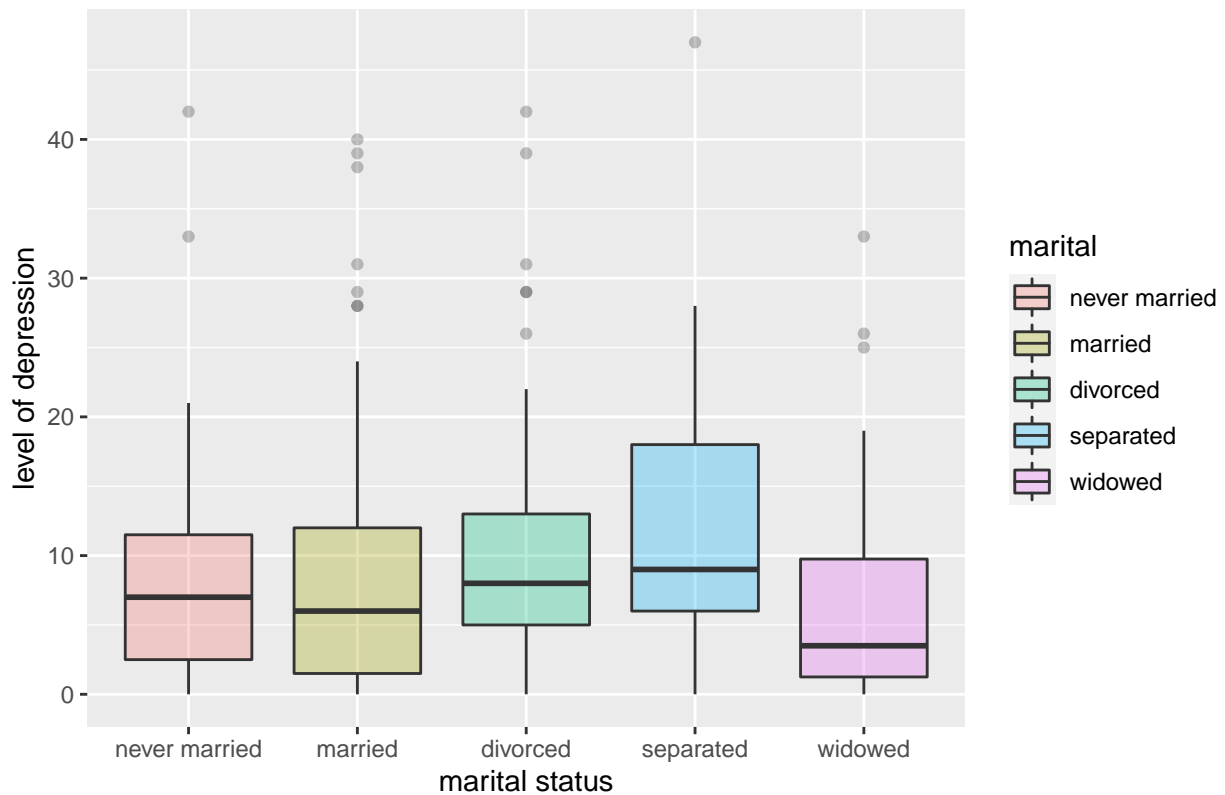
```
ggplot(depress,aes(x=cesd,fill=marital))+geom_density(alpha=.3)+xlab("level of depression")+ggtitle("depression by marital status")
```

density of depression levels according to marital status



```
ggplot(depress, aes(x=marital, y=cesd, fill=marital))+geom_boxplot(alpha=.3)+xlab("marital status")+ylab("density")
```


level of depression based on marital status



This data shows the relationship between levels of depression and marital status. It demonstrates that individuals who are separated record higher levels of depression, on average, than those with another marital status and those who are widowed demonstrate the lowest levels of depression, on average.

```
table(depress$cesd, depress$sex, depress$marital) %>%prop.table(margin=2) %>%round(3)
```

```
## , , = never married
##
##
##      male female
## 0 0.000 0.016
## 1 0.000 0.011
## 2 0.036 0.011
## 3 0.018 0.000
## 4 0.009 0.011
## 5 0.018 0.011
## 6 0.000 0.005
## 7 0.009 0.011
## 8 0.009 0.011
## 9 0.009 0.005
## 10 0.009 0.005
## 11 0.009 0.000
## 12 0.009 0.005
## 13 0.000 0.005
## 14 0.000 0.000
## 15 0.009 0.000
```

```

## 16 0.000 0.011
## 17 0.000 0.000
## 18 0.000 0.000
## 19 0.000 0.000
## 20 0.000 0.000
## 21 0.009 0.011
## 22 0.000 0.000
## 24 0.000 0.000
## 25 0.000 0.000
## 26 0.000 0.000
## 28 0.000 0.000
## 29 0.000 0.000
## 31 0.000 0.000
## 33 0.000 0.005
## 38 0.000 0.000
## 39 0.000 0.000
## 40 0.000 0.000
## 42 0.009 0.000
## 47 0.000 0.000
##
## , , = married
##
##
##      male female
## 0 0.117 0.033
## 1 0.036 0.049
## 2 0.009 0.005
## 3 0.045 0.044
## 4 0.027 0.027
## 5 0.036 0.011
## 6 0.018 0.022
## 7 0.027 0.011
## 8 0.027 0.022
## 9 0.054 0.016
## 10 0.000 0.011
## 11 0.009 0.016
## 12 0.009 0.011
## 13 0.027 0.016
## 14 0.000 0.011
## 15 0.027 0.011
## 16 0.000 0.005
## 17 0.000 0.011
## 18 0.000 0.016
## 19 0.000 0.005
## 20 0.009 0.000
## 21 0.000 0.000
## 22 0.000 0.005
## 24 0.000 0.005
## 25 0.000 0.000
## 26 0.000 0.000
## 28 0.009 0.005
## 29 0.000 0.005
## 31 0.009 0.000
## 33 0.000 0.000

```

```

## 38 0.000 0.005
## 39 0.000 0.005
## 40 0.000 0.005
## 42 0.000 0.000
## 47 0.000 0.000
##
## , , = divorced
##
##
##      male female
## 0 0.009 0.016
## 1 0.009 0.005
## 2 0.000 0.005
## 3 0.000 0.022
## 4 0.027 0.022
## 5 0.036 0.016
## 6 0.027 0.011
## 7 0.018 0.016
## 8 0.036 0.005
## 9 0.000 0.016
## 10 0.027 0.011
## 11 0.009 0.005
## 12 0.018 0.005
## 13 0.009 0.005
## 14 0.009 0.005
## 15 0.000 0.000
## 16 0.000 0.005
## 17 0.000 0.005
## 18 0.009 0.005
## 19 0.000 0.005
## 20 0.000 0.005
## 21 0.000 0.011
## 22 0.009 0.005
## 24 0.000 0.000
## 25 0.000 0.000
## 26 0.009 0.000
## 28 0.000 0.000
## 29 0.000 0.011
## 31 0.000 0.005
## 33 0.000 0.000
## 38 0.000 0.000
## 39 0.000 0.005
## 40 0.000 0.000
## 42 0.000 0.005
## 47 0.000 0.000
##
## , , = separated
##
##
##      male female
## 0 0.009 0.000
## 1 0.000 0.000
## 2 0.000 0.000
## 3 0.000 0.005

```

```

## 4 0.000 0.000
## 5 0.000 0.005
## 6 0.000 0.011
## 7 0.000 0.000
## 8 0.000 0.005
## 9 0.009 0.000
## 10 0.000 0.000
## 11 0.000 0.000
## 12 0.000 0.005
## 13 0.000 0.000
## 14 0.000 0.005
## 15 0.000 0.000
## 16 0.000 0.000
## 17 0.000 0.000
## 18 0.000 0.005
## 19 0.000 0.000
## 20 0.000 0.000
## 21 0.000 0.000
## 22 0.000 0.005
## 24 0.000 0.000
## 25 0.000 0.000
## 26 0.000 0.000
## 28 0.000 0.005
## 29 0.000 0.000
## 31 0.000 0.000
## 33 0.000 0.000
## 38 0.000 0.000
## 39 0.000 0.000
## 40 0.000 0.000
## 42 0.000 0.000
## 47 0.000 0.005
##
## , , = widowed
##
##
##      male female
## 0 0.018 0.027
## 1 0.000 0.016
## 2 0.000 0.022
## 3 0.018 0.016
## 4 0.000 0.011
## 5 0.000 0.000
## 6 0.000 0.005
## 7 0.009 0.011
## 8 0.000 0.005
## 9 0.000 0.011
## 10 0.000 0.011
## 11 0.000 0.000
## 12 0.000 0.005
## 13 0.000 0.005
## 14 0.000 0.000
## 15 0.000 0.000
## 16 0.009 0.000
## 17 0.000 0.005

```

```
## 18 0.000 0.000
## 19 0.000 0.005
## 20 0.000 0.000
## 21 0.000 0.000
## 22 0.000 0.000
## 24 0.000 0.000
## 25 0.009 0.000
## 26 0.000 0.005
## 28 0.000 0.000
## 29 0.000 0.000
## 31 0.000 0.000
## 33 0.000 0.005
## 38 0.000 0.000
## 39 0.000 0.000
## 40 0.000 0.000
## 42 0.000 0.000
## 47 0.000 0.000
```

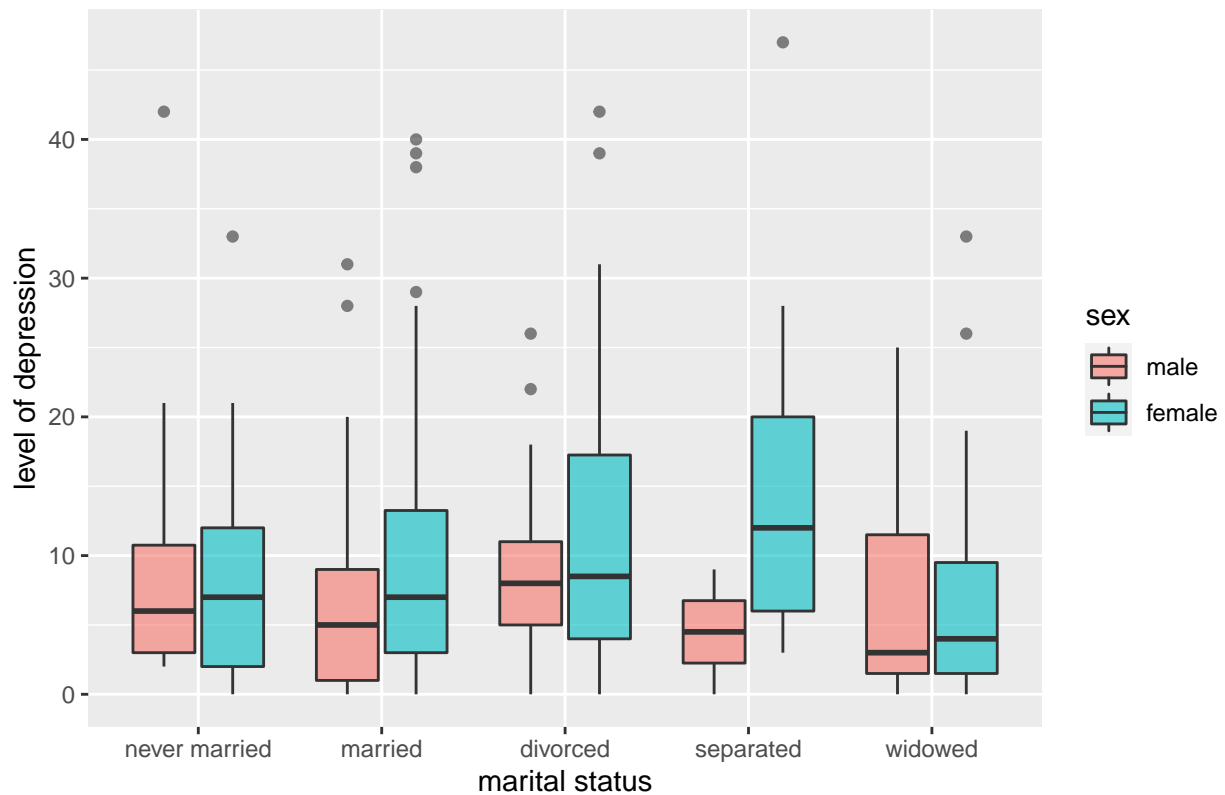
```
depress%>% group_by(sex,marital)%>%summarize(avg=mean(cesd))
```

```
## 'summarise()' has grouped output by 'sex'. You can override using the '.groups'
## argument.
```

```
## # A tibble: 10 x 3
## # Groups:   sex [2]
##   sex    marital      avg
##   <fct> <fct>      <dbl>
## 1 male  never married  9.06
## 2 male  married       6.36
## 3 male  divorced      8.79
## 4 male  separated     4.5
## 5 male  widowed      7.71
## 6 female never married  8.44
## 7 female married     9.43
## 8 female divorced   11.5
## 9 female separated  15.4
## 10 female widowed   6.90
```

```
ggplot(depress,aes(x=marital,y=cesd,fill=sex))+geom_boxplot(alpha=.6)+xlab("marital status")+ylab("level")
```

level of depression of males and females based on marital status



This data compared the relationships of all three variables to one another. It shows across the boxplot that males demonstrate lower levels of depression and females show higher levels of depression, while also categorizing the results by marital status. here we are able to compare all three variables at once by plotting marital status and sex on a single axis and level of depression against them.

Conclusion

I found that my initial hypothesis was somewhat correct as the variable did correspond with one another, but the variables I expected to react more with one another actually reacted less. I anticipated the marital status would have more of a relationship with levels of depression than sex would. The exact opposite actually occurred, with sex displaying a strong relationship with levels of depression, and marital status being less affected. I learned that females are more likely than males to display higher levels of depression, and those who are separated are the most depressed, while those who are widowed are the least depressed compared to all measured marital statuses.