MATH 130 Exploratory Data Analysis Project

Kendra Miriah Cote

9/24/2021

Selected Data Set: Depression

Introduction

For this exploratory data analysis project, I decided to use the Depression data set. This data set is from the first set of interviews focusing on the prospective study of depression in adult residents of Los Angeles County which includes 294 observations and 37 variables. If more research is desired, you can find the information in the Practical Multivariate Analysis, 5th edition by May and Clark. The variables I will be researching will be sex and marital status. I wish to find out which sex, female or male, is more depressed and in which marital status are they more depressed.

These are the packages and operations I performed to start the project.

```
library(ggplot2)
library(forcats)
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(RColorBrewer)
library(sjPlot)
```

```
## Registered S3 methods overwritten by 'parameters':
##   method                           from
##   as.double.parameters_kurtosis    datawizard
##   as.double.parameters_skewness    datawizard
##   as.double.parameters_smoothness  datawizard
##   as.numeric.parameters_kurtosis   datawizard
##   as.numeric.parameters_skewness   datawizard
##   as.numeric.parameters_smoothness datawizard
```

```
##    print.parameters_distribution    datawizard
##    print.parameters_kurtosis        datawizard
##    print.parameters_skewness        datawizard
##    summary.parameters_kurtosis      datawizard
##    summary.parameters_skewness      datawizard
```

```
depress <- read.delim("/Users/kendr/Documents/R Easy Access/depress_081217.txt", header=TRUE,sep="\t")
dim(depress)
```

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

294 Observations and 37 Variables

Uni-variate Descriptions

Sex

During this first section I wish to explore which sex (female or male) is more likely to experience depression than the other based off the data provided.

```
class(depress$sex)
```

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

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

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

The first code chunk shows that my data set is a series of integers. The second code chunk shows a table of depression comparisons for females and males. This table is showing that more males interviewed experienced depression than females who were interviewed. Compared to females, **72** more males experience depression.

```
ggplot(depress, aes(x=sex, fill=sex))+ geom_bar()+ theme_light()+
scale_fill_manual(values=c("pink", "lightblue"))+
ggtitle("Females and Males with Depression")+
ylab("Number of Individuals")+ xlab("Individuals Sex")
```

Females and Males with Depression
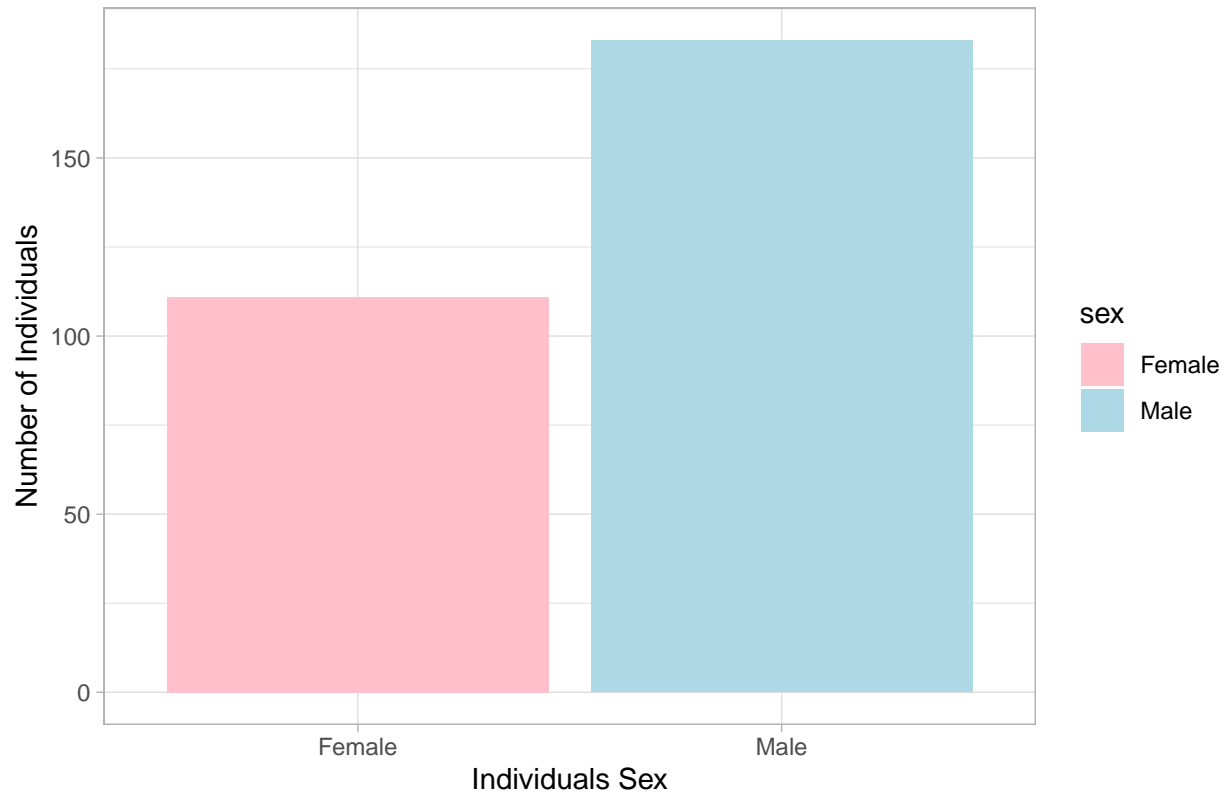
The above graph shows that males interviewed were more depressed than the females who were interviewed.

Marital

Now I will focus on marital status and whether being married, divorced, widowed, separated, or never married has an effect on an individuals depression.

```
class(depress$marital)
```

```
## [1] "character"
```

```
table(depress$marital)
```

```
##
##      Divorced      Married Never Married    Separated      Widowed
##            43          127            73           13           38
```

Out of the individuals that were interviewed **43** of them were depressed when divorced, **127** of them were depressed when married, **73** were depressed having never been married, **13** of them were depressed when separated from their partner, and **38** were depressed when widowed from their partner.

This was a bit shocking because I expected to see that those individuals whose partners had died would have more depression due to their absence. You could infer that being in a marriage can cause depression among the individuals interviewed.

```
ggplot(depress, aes(x=marital,fill=marital)) + geom_bar()+theme_light()+
scale_color_brewer("Set1")+ggtitle("Marital Status and Depression")+
ylab("Number of Individuals")+ xlab("Marital Status")
```

Marital Status and Depression
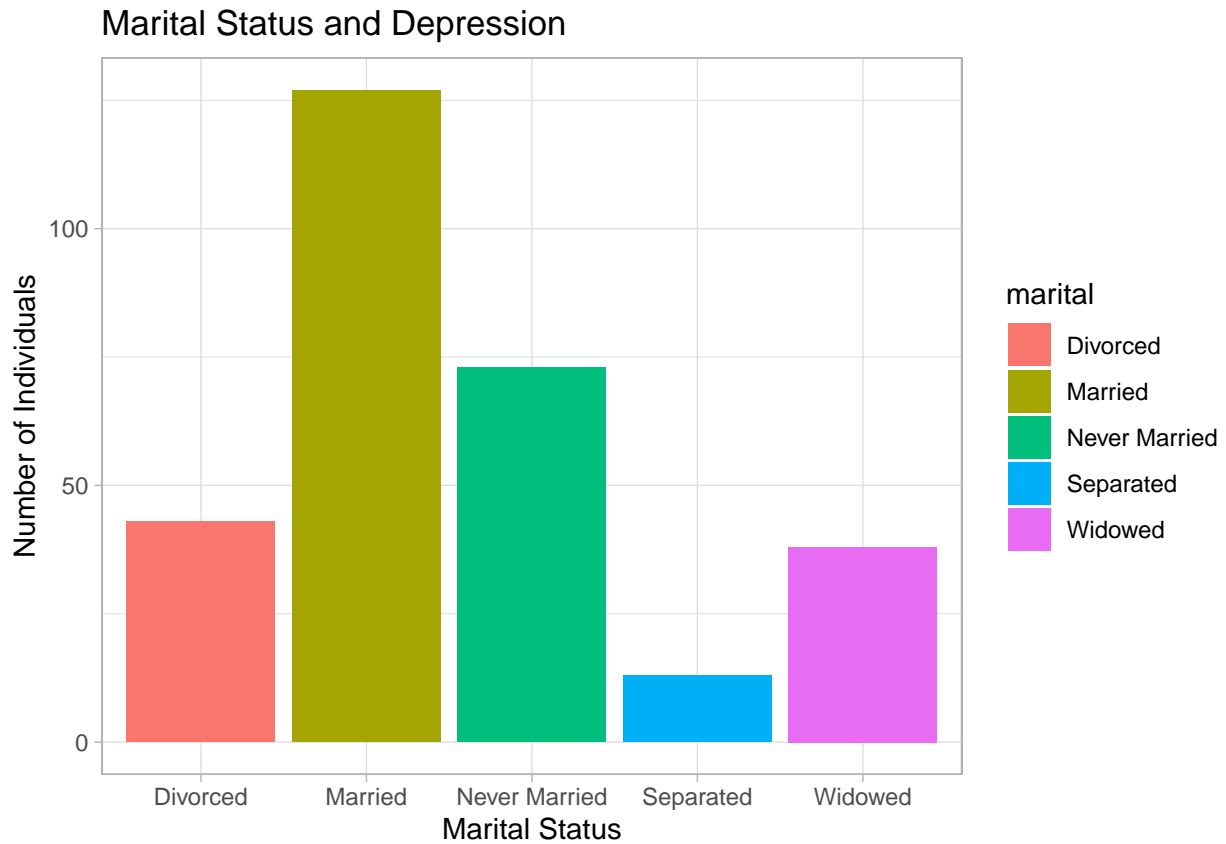
This graph shows that the number of individuals depressed while married was twice as much as those who were never married and definitely greater than those divorced, separated, or widowed.

Bi-variate Comparison

This next section will be focusing which individual's sex was more depressed depending on their marital status. This is a two-way table showing the comparison between males and females with depression when they were married, divorced, separated, never married, or widowed.

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

```
##
##          Divorced Married Never Married Separated Widowed
##   Female       18      55            29         2       7
##   Male         25      72            44        11      31
```

This table shows that males in every category were more depressed than females no matter what marital status they held. **25** males and **18** females were depressed when divorced, **72** males and **55** females were depressed when married, **44** males and **29** females were depressed if they were never married, **11** males and **2** females were depressed when separated, and **31** males and **7** females were depressed when their partner passed away.

```
ggplot(depress, aes(x=marital, fill=sex))+geom_bar(position="dodge")+theme_classic()+
ggtitle("Marital Status and Depression With Females and Males")+
ylab("Number of Individuals")+ xlab("Marital Status")+scale_color_brewer("Set1")
```


Marital Status and Depression With Females and Males
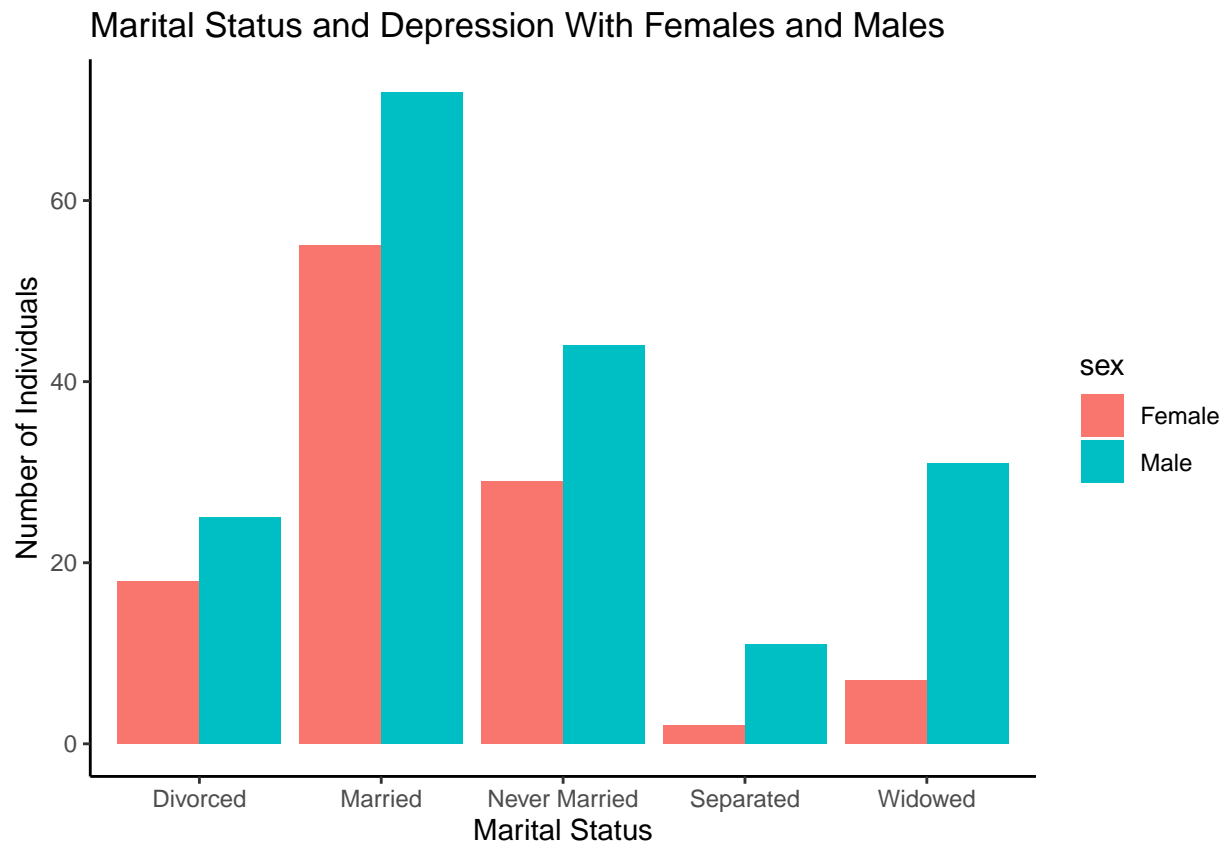
```
plot_xtab(depress$sex,depress$marital, margin="row",show.total=FALSE)
```

```
## Warning: 'guides(<scale> = FALSE)' is deprecated. Please use 'guides(<scale> =
## "none")' instead.
```
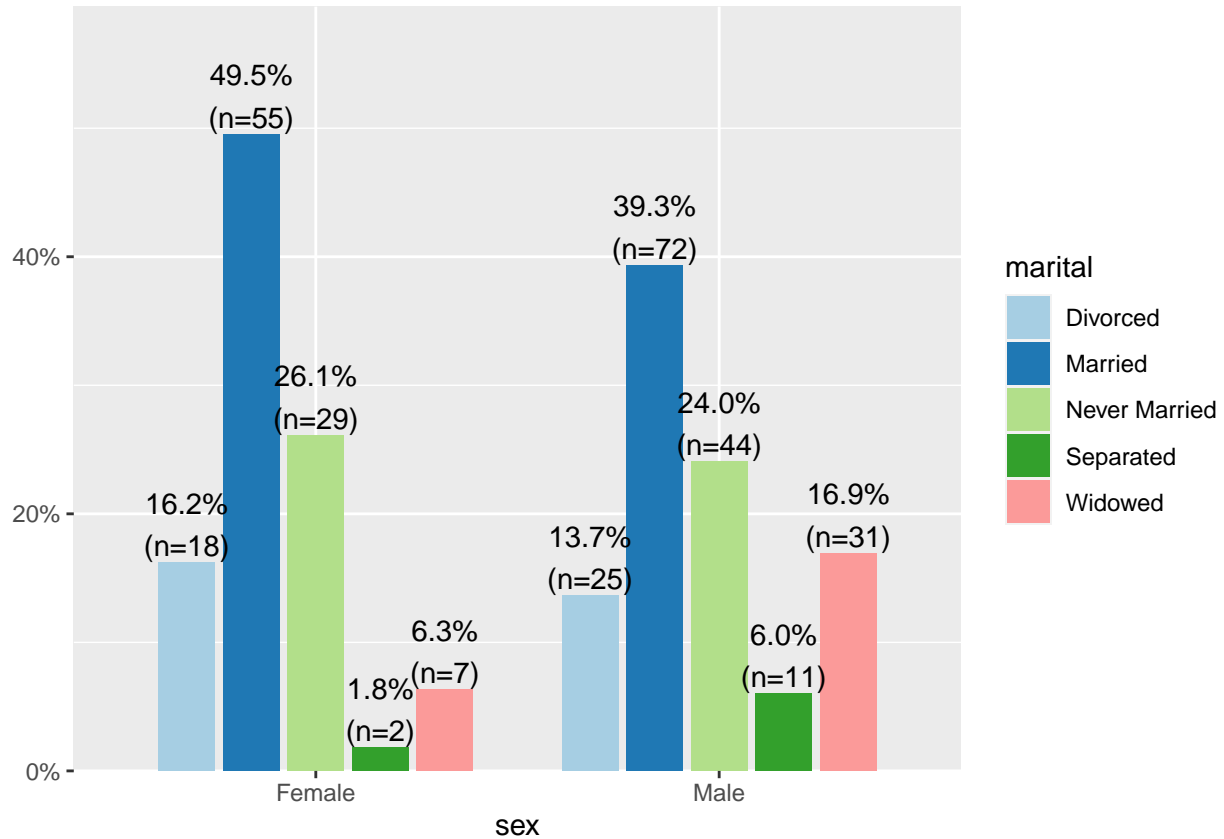
```
round(prop.table(table(depress$sex, depress$marital), margin=1),3)
```

```
##
##          Divorced Married Never Married Separated Widowed
##   Female    0.162   0.495         0.261     0.018   0.063
##   Male      0.137   0.393         0.240     0.060   0.169
```

The first graph shows the number of individuals of each sex and the level of depression they had depending on their marital status. As we can see the males interviewed were more depressed in both categories that were comparing them to the opposite sex and in every marital status that was questioned. I found it interesting that men were a great deal more depressed when their partner passed away than women in the same category. It leads you to wonder what each of these interviewed individual's home life was like.

Both the second graph and table above are giving the percentages of the data displayed. Even though individually men were more depressed than women these percentages show that out of women and men category it shows that **16%** of women were depressed after divorce while **13%** of men were depressed after divorce, **49%** of females and **39%** of males were married and depressed, **26%** of females and **24%** of males were depressed because they were never married, **1.8%** of females and **6%** of males were depressed after separation and finally **6.3%** of females and **16%** of males were depressed after their partner passed away.

These statistics are very intriguing and I wish to see more data from different locations around the world. This study was done in LA so it strikes the question of their location having anything to do with their level of depression so continuing the studying in different locations would be a fascinating opportunity.