MATH 130 Exploratory Data Analysis Project

Brooke Rosenow

2.25.2022

```
knitr::opts_chunk$set(fig.width=6, fig.height=4) # This sets all figure sizes in the document un
less otherwise specified.
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)
```

```
depress <- read.delim("C:/Users/Brooke/Documents/Spring 2022/Math 130/EDA/depress_081217.txt", h
eader=TRUE,sep="\t")
```

Introduction

This data set is on a study of depression in adult residents living in Los Angeles County. The data set includes 294 observations and 37 variables. This project will specifically look at two variables in the data set: marital status and cesd. There are 5 variables with marital status: never married, married, divorced, separated, and widowed. The cesd variable is a continuous variable summarizing the answers of indiviuals surveyed when asked how they feel. A value of 0 = lowest level possible and 60 = highest level possible. Based off of these variables, one question is asked: (1) Does a certain marital status have a high amount of individuals that experience depression?

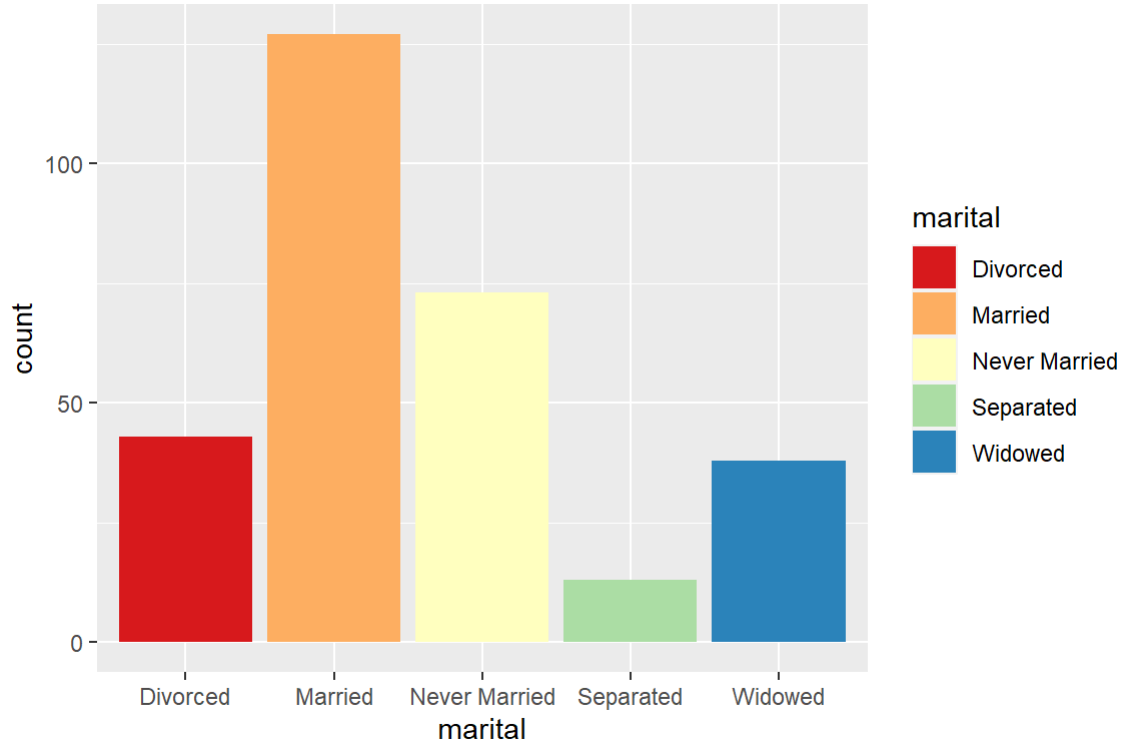
Univariate Exploration

Marital

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

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

```
ggplot(depress, aes(x=marital, fill=marital)) + geom_bar() + scale_fill_brewer(palette="Spectra
l")
```



CESD

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

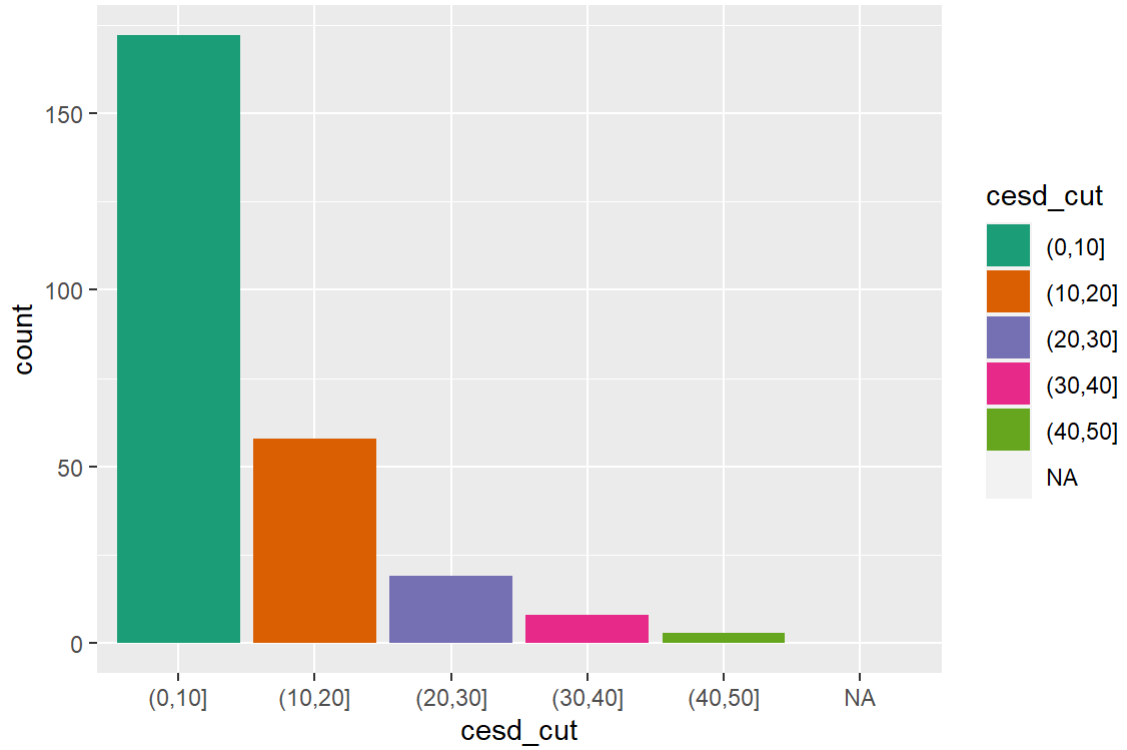
```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##   0.000   3.000   7.000   8.884  12.000  47.000
```

There was a max score of 47 out of 60 and a minum score of 0. The average score was 8.88.

```
depress$cesd_cut <- cut(depress$cesd, seq(0,60, 10))
table(depress$cesd_cut)
```

```
##
##  (0,10] (10,20] (20,30] (30,40] (40,50] (50,60]
##     172      58      19       8       3       0
```

```
ggplot(depress, aes(x=cesd_cut, fill=cesd_cut)) + geom_bar() + scale_fill_brewer(palette="Dark2"
)
```



Since this variable had a large data set, I grouped the variable into smaller subdivisions to have a clearer visualization of the distribution between the scores.

Bivariate Exploration

Marital vs CESD

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

```
##
##                  0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20
##   Divorced       3  2  6  2  3  4  1  3  3  2  2  1  2  1  0  1  2  0  0  0  0
##   Married       19 13  2 13  8  6  6  5  7  9  2  4  3  6  2  5  1  2  3  1  1
##   Never Married  4  2  1  4  7  7  5  5  5  3  5  2  3  2  2  0  1  1  2  1  1
##   Separated      1  0  0  1  0  1  2  0  1  1  0  0  1  0  1  0  0  0  1  0  0
##   Widowed        7  3  4  5  2  0  1  3  1  2  2  0  1  1  0  0  1  1  0  1  0
##
##                 21 22 24 25 26 28 29 31 33 38 39 40 42 47
##   Divorced       3  0  0  0  0  0  0  0  1  0  0  0  1  0
##   Married        0  1  1  0  0  2  1  1  0  1  1  1  0  0
##   Never Married  2  2  0  0  1  0  2  1  0  0  1  0  1  0
##   Separated      0  1  0  0  0  1  0  0  0  0  0  0  0  1
##   Widowed        0  0  0  1  1  0  0  0  1  0  0  0  0  0
```

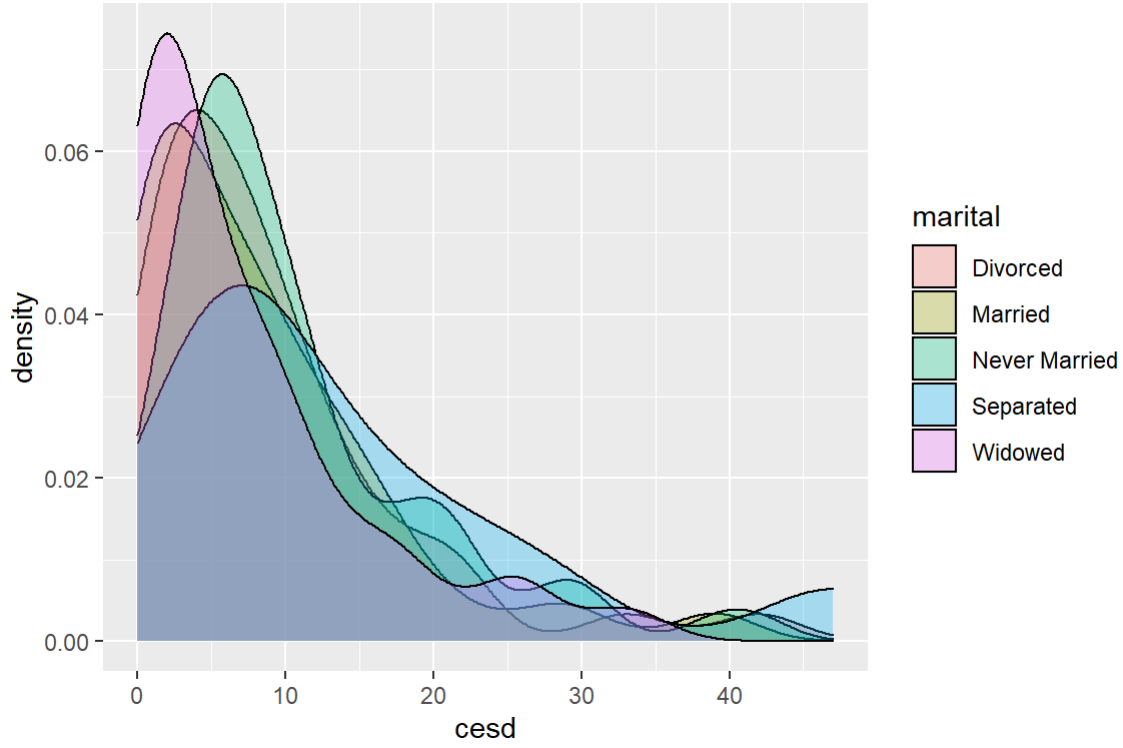
This table shows the individual scores matched with the marital status.

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

```
##
##                (0,10] (10,20] (20,30] (30,40] (40,50] (50,60]
##   Divorced         28       7       3       1       1       0
##   Married          71      28       5       4       0       0
##   Never Married    44      15       7       2       1       0
##   Separated         6       3       2       0       1       0
##   Widowed          23       5       2       1       0       0
```

This table shows the grouped scores matched with marital status.

```
ggplot(depress, aes(x=cesd, fill=marital)) + geom_density(alpha=0.3)
```



The density plot shows that widowed adults have a lower score relating to depression, while those that have never been married score a bit higher than other individuals with different marital statues. Separated individuals appear to have higher scores than other marital statuses as well.