

EDA_tejaz

2024-09-23

Introduction

For the EDA project, I will be using the “Depression” data set. This data set is showing 294 observations in total but 37 variables. Specifically I will be looking into two variables which are marital status and sex. The goal I intend to explore is, if there is any correlation between the sex of someone and their marital status related to depression.

```
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
```

```
depress <- read.delim("/Users/taniaejaz/Desktop/Math130/data/Depress.txt", header
=TRUE, sep="\t")
head(depress)
```

```
## ID SEX AGE MARITAL EDUCAT EMPLOY INCOME RELIG C1 C2 C3 C4 C5 C6 C7 C8 C9 C10
## 1 1 2 68 5 2 4 4 1 0 0 0 0 0 0 0 0 0 0
## 2 2 1 58 3 4 1 15 1 0 0 1 0 0 0 0 0 0 0
## 3 3 2 45 2 3 1 28 1 0 0 0 0 1 0 0 0 0 0
## 4 4 2 50 3 3 3 9 1 0 0 0 0 1 1 0 3 0 0
## 5 5 2 33 4 3 1 35 1 0 0 0 0 0 0 0 3 3 0
## 6 6 1 24 2 3 1 11 1 0 0 0 0 0 0 0 0 1 0
## C11 C12 C13 C14 C15 C16 C17 C18 C19 C20 CESD CASES DRINK HEALTH REGDOC TREAT
## 1 0 0 0 0 0 0 0 0 0 0 0 2 2 1 1
## 2 0 1 0 0 1 0 1 0 0 4 0 1 1 1 1
## 3 0 0 0 1 1 1 0 0 0 4 0 1 2 1 1
## 4 0 0 0 0 0 0 0 0 0 5 0 2 1 1 2
## 5 0 0 0 0 0 0 0 0 0 6 0 1 1 1 1
## 6 0 1 2 0 0 2 1 0 0 7 0 1 1 1 1
## BEDDAYS ACUTEILL CHRONILL
## 1 0 0 1
## 2 0 0 1
## 3 0 0 0
## 4 0 0 1
## 5 1 1 0
## 6 0 1 1
```

Univariate Exploration

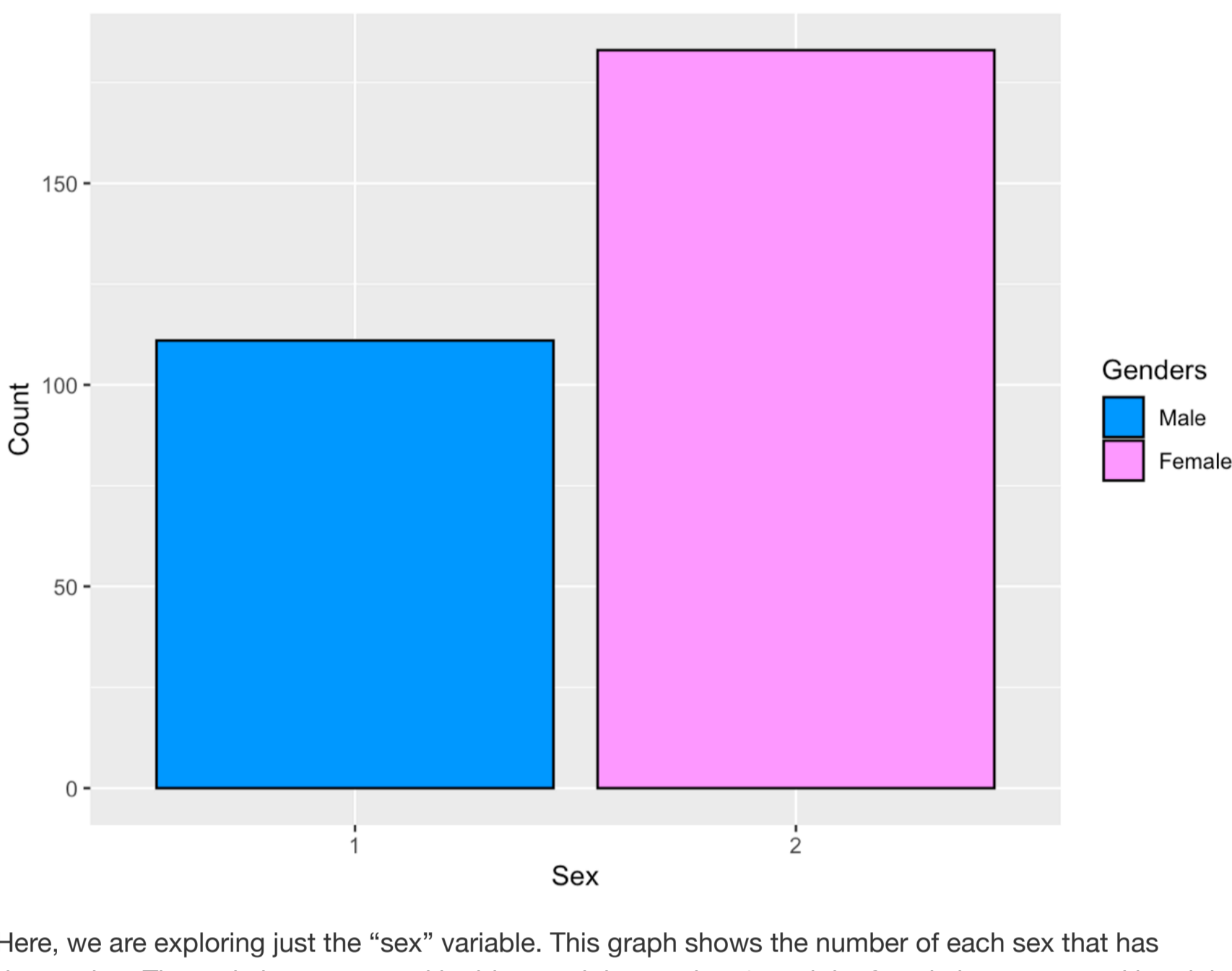
Variable: Sex

```
table(depress$SEX)
```

```
##
## 1 2
## 111 183
```

The table shows males being represented by number 1 and females being represented by number 2. It is hinting that 111 males are depressed while 183 females are depressed.

```
ggplot(depress, aes(as.factor(SEX), fill = as.factor(SEX))) +
geom_bar(colour="#000") +
scale_fill_manual(values = c("1" = "#0099FF", "2" = "#FF99FF"), labels = c("1" = "
Male", "2" = "Female")) +
xlab("Sex") + ylab("Count") + labs(fill = "Genders")
```



Here, we are exploring just the “sex” variable. This graph shows the number of each sex that has depression. The male is represented by blue, and the number 1, and the female is represented by pink and the number 2. As you can see, we can already predict that females have a higher depression rate.

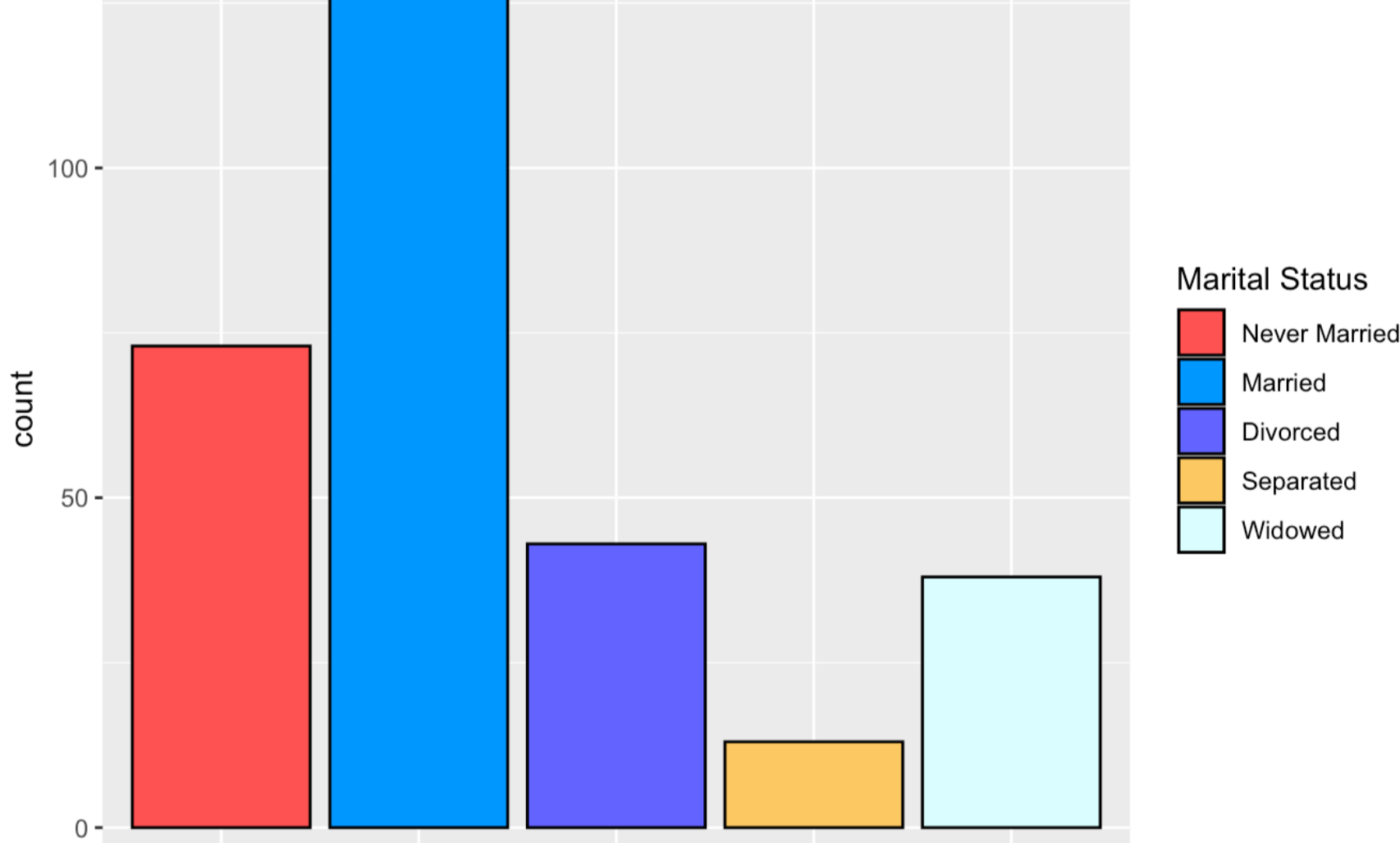
Variable: Marital

```
table(depress$MARITAL)
```

```
##
## 1 2 3 4 5
## 73 127 43 13 38
```

This table gives numeric data on just one variable, “Marital.” It shows one meaning “Never married,” two meaning “Married,” three meaning “Divorced,” four meaning “Separated,” and five meaning “Widowed.” As we can see, “Married” has the highest value while “Separated” has the lowest value.

```
ggplot(depress, aes(as.factor(MARITAL), fill = as.factor(MARITAL))) +
geom_bar(colour="#000") +
scale_fill_manual(values = c("1" = "#FF5555", "2" = "#0099FF", "3" = "#6666FF", "4"
= "#FFCC66", "5" = "#DDFFFF", "6" = "#DD9900"),
labels=c("1" = "Never Married", "2" = "Married", "3" = "Divorced", "4" = "Separat
ed", "5" = "Widowed", "6" = "Missing")) +
xlab("Marital Status")+labs(fill = "Marital Status")
```



Here, we are just exploring the “marital” variable. This graph displays the variable “marital statuses” of the depressed people in the forms of x values and the amount of people in each status is displayed in the y values. As you can the marital status for married people takes the lead while “separated” couples is the lowest.

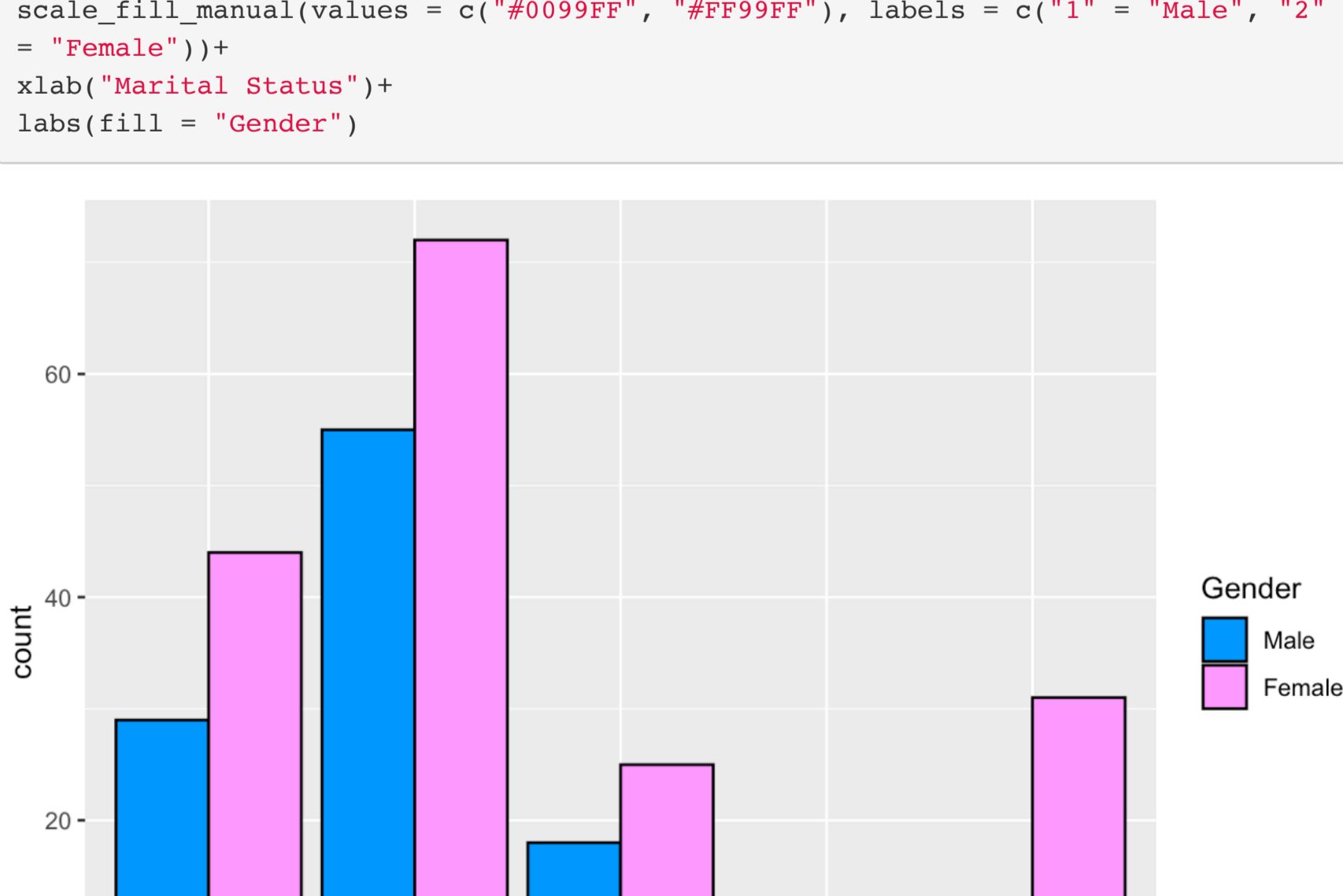
Bivariate Exploration

```
table(depress$SEX, depress$MARITAL)
```

```
##
## 1 2 3 4 5
## 1 29 55 18 2 7
## 2 44 72 25 11 31
```

This table gives numeric data. It shows the correlation between sex and marital status. We can see the exact values in each category of the marital status (1 through 6) in males and females represented (1 and 2).

```
ggplot(depress, aes(x = as.factor(MARITAL), fill = as.factor(SEX)))+
geom_bar(position = "dodge", colour = "#000")+
scale_fill_manual(values = c("#0099FF", "#FF99FF"), labels = c("1" = "Male", "2"
= "Female"))+
xlab("Marital Status")+
labs(fill = "Gender")
```



This graph explores both sex and marital status variables together. It reveals how the higher depression rates in females reflects onto their marriages. The graph keeps a constant pattern between male and female marital statuses, but the female’s higher depression rate causes more drastic effects.

Conclusion

In conclusion, sex correlates with marital status and depression. We can conclude this because, in previous graphs, we see that married females have a higher count above 60 while married males have a lower count at above 40 but below 60. Furthermore, we saw in numeric representation that females who are not married are more prone to depression than males. Also, females, in general, are more depressed than males at a count of 183. Therefore, in the data I used, marital status and sex played a factor in depression. Otherwise it may vary depending on different data given.