Exploratory Data Analysis Project

Depression

Inroduction

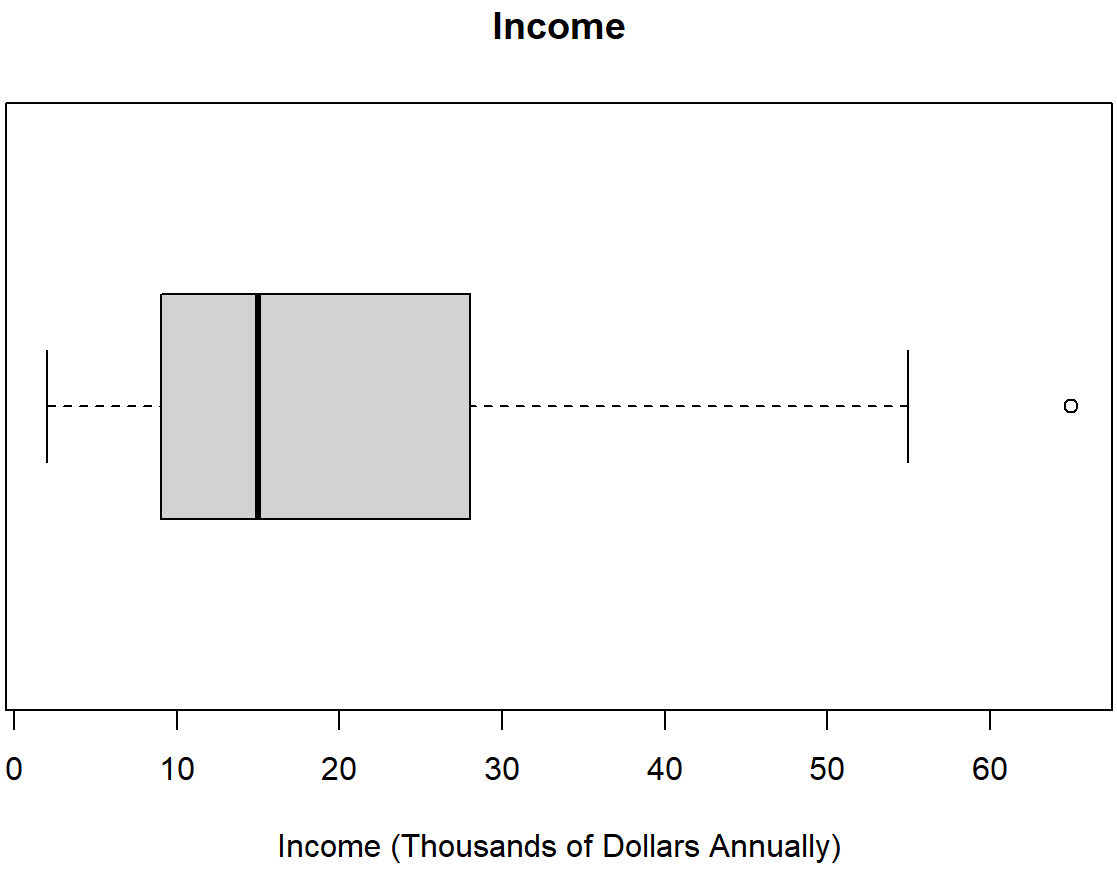
I am going to explore the Depression data set. This data set looks at depression in the adult residents of Los Angeles County and includes 294 observations. The variables I am interested in analyzing are: INCOME, DRINK, EDUCAT, BEDDAYS and TREAT. INCOME is a continuous variable reported in thousands of dollars per year. DRINK shows whether the person is a regular drinker– 1 is yes, 2 is no. EDUCAT is what level of education the individual has. BEDDAYS is whether the individual has spent entire day(s) in bed within the last 2 months before the survey. TREAT shows whether a doctor has prescribed treatments for the depression either through medications or lifestyle changes– 1 is yes they have been treated, and 2 is no they have not. I am interested to see how treatment varies between different income levels. I am also interested to see how level of education factors into treatment and beddays.

Univariate Exploration

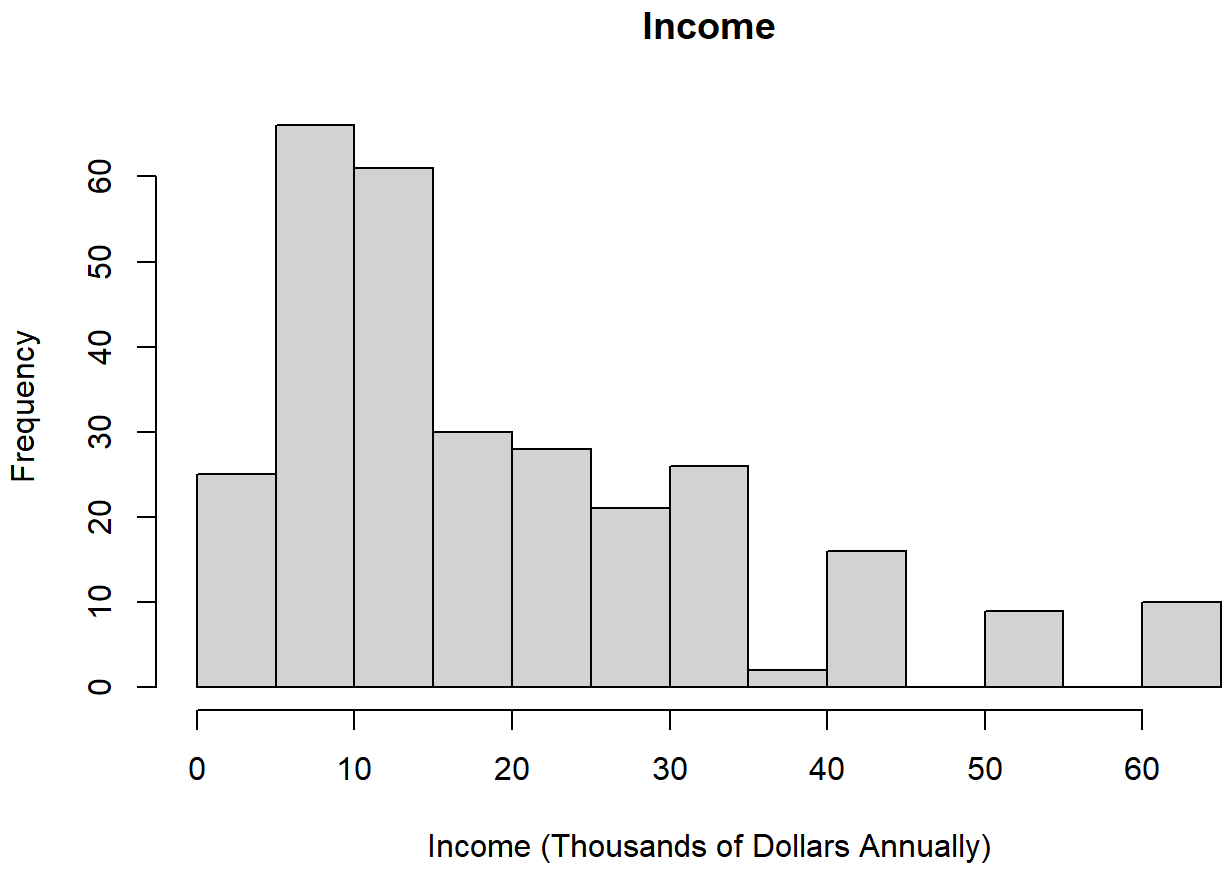
```
summary(depress$INCOME)
```

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    2.00    9.00   15.00   20.57   28.00   65.00
```

```
boxplot(depress$INCOME, horizontal = TRUE, main="Income", xlab="Income (Thousands of Dollars Ann
ually)")
```



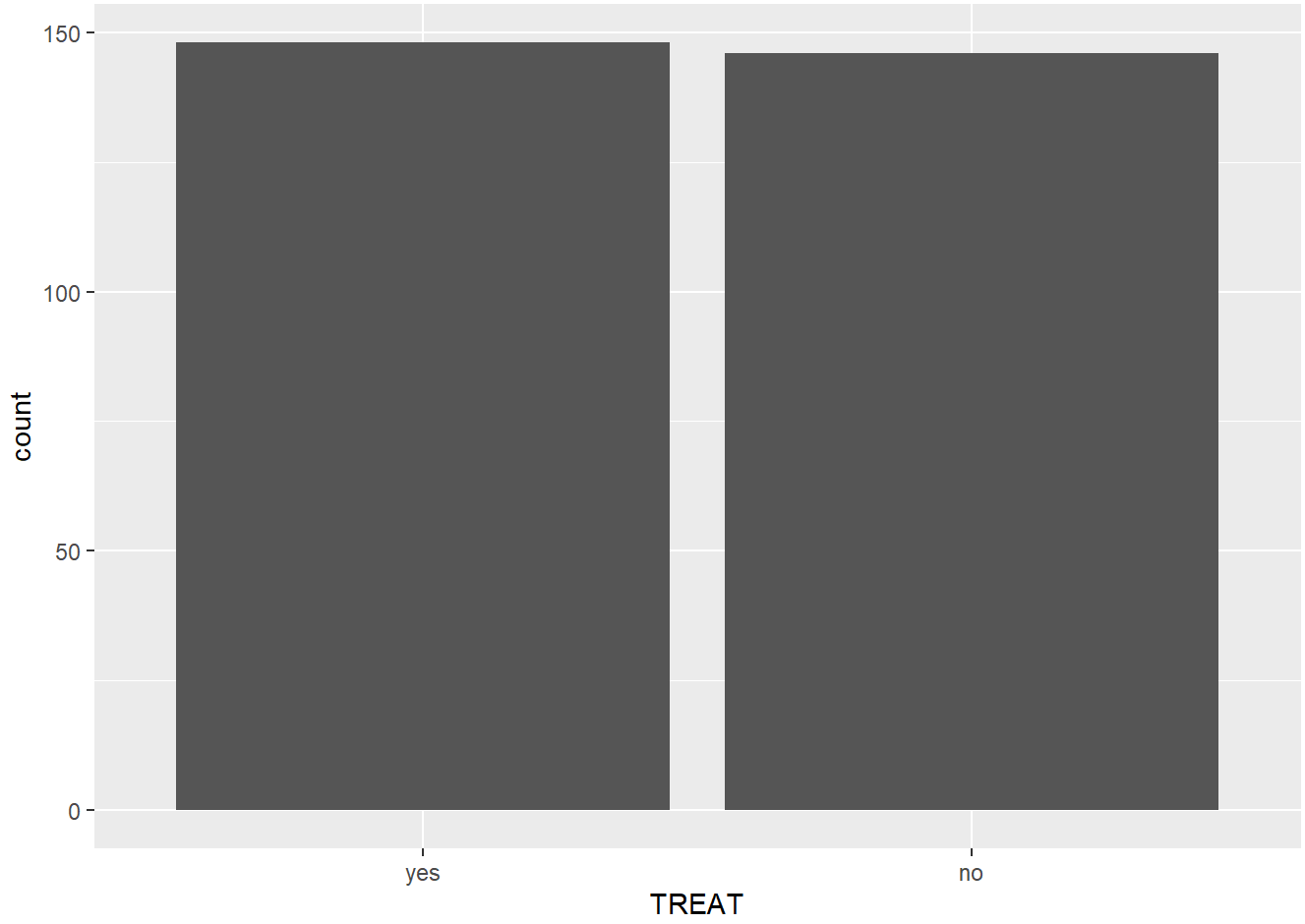

```
hist(depress$INCOME, main= "Income", xlab = "Income (Thousands of Dollars Annually)")
```




```
depress$TREAT <- factor(depress$TREAT, labels=c("yes","no"))
table(depress$TREAT)
```

```
##
## yes  no
## 148 146
```

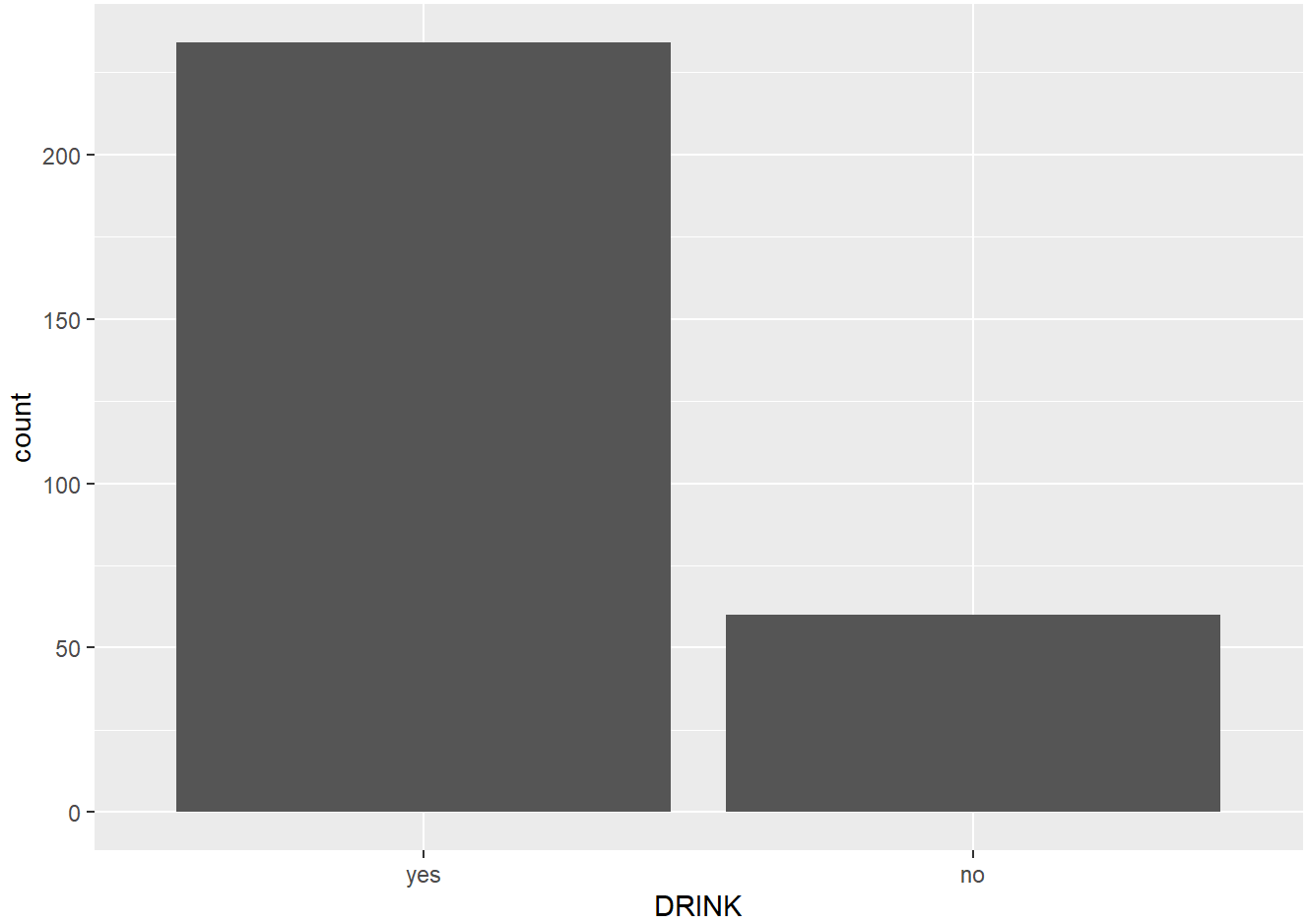
```
ggplot(depress, aes(x=TREAT)) + geom_bar()
```




```
depress$DRINK <- factor(depress$DRINK, labels=c("yes","no"))
table(depress$DRINK)
```

```
##
## yes  no
## 234  60
```

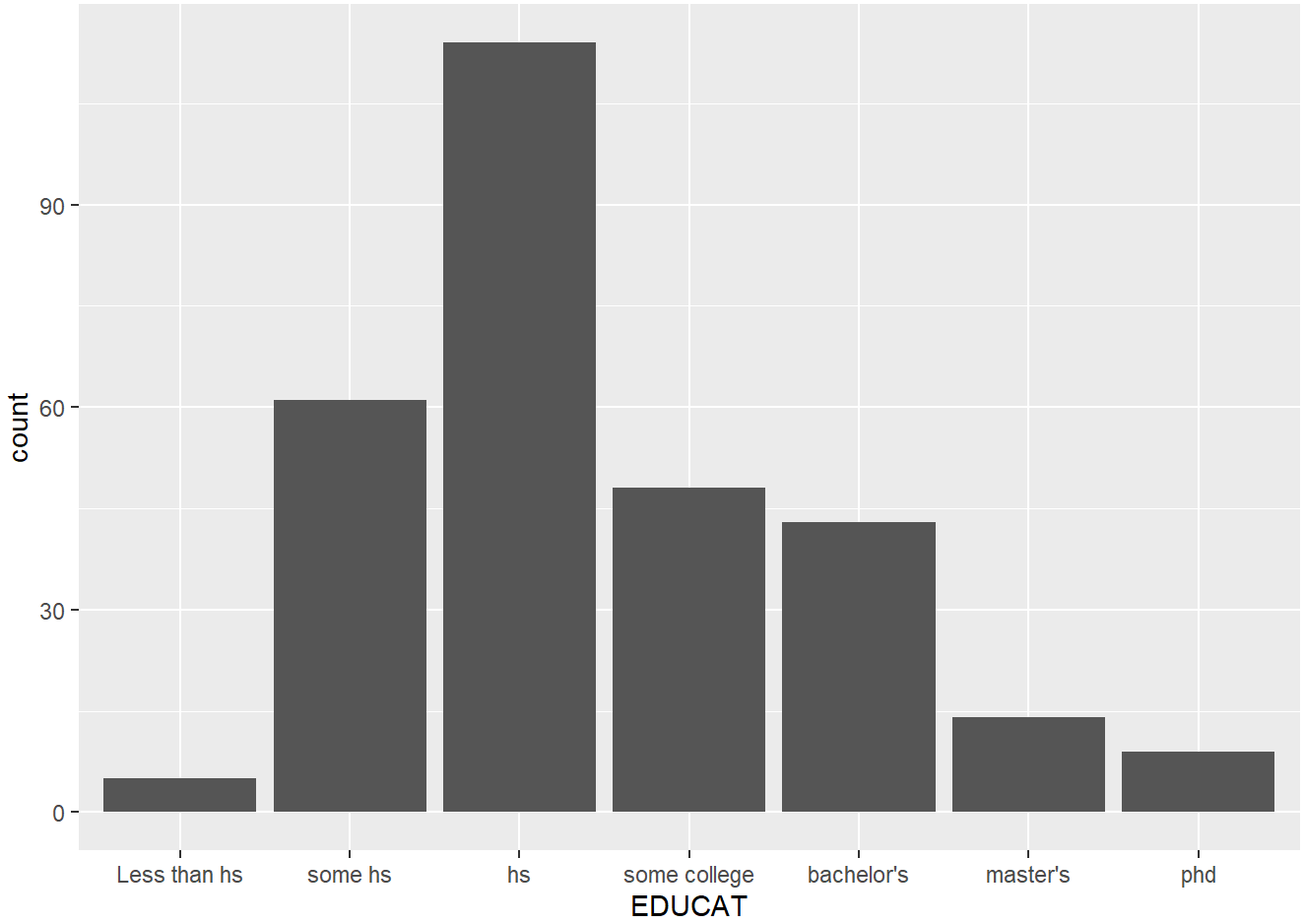
```
ggplot(depress, aes(x=DRINK)) + geom_bar()
```




```
depress$EDUCAT <- factor(depress$EDUCAT, labels=c("Less than hs","some hs", "hs","some colleg
e","bachelor's","master's","phd"))
table(depress$EDUCAT)
```

```
##
## Less than hs      some hs           hs some college   bachelor's     master's
##            5           61          114           48           43           14
##          phd
##            9
```

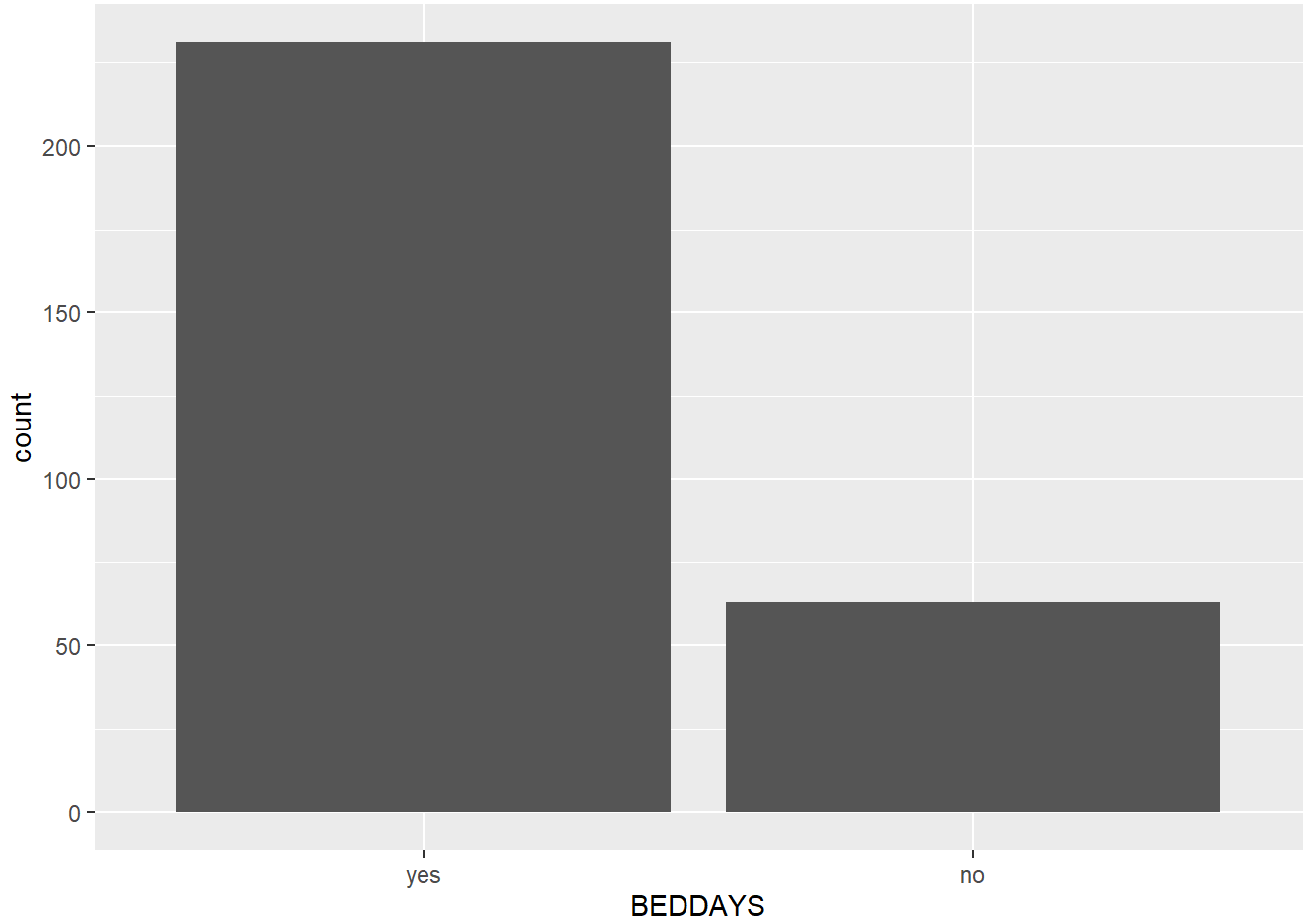
```
ggplot(depress, aes(x=EDUCAT)) + geom_bar()
```




```
depress$BEDDAYS <- factor(depress$BEDDAYS, labels=c("yes","no"))
table(depress$BEDDAYS)
```

```
##
## yes  no
## 231  63
```

```
ggplot(depress, aes(x=BEDDAYS)) + geom_bar()
```

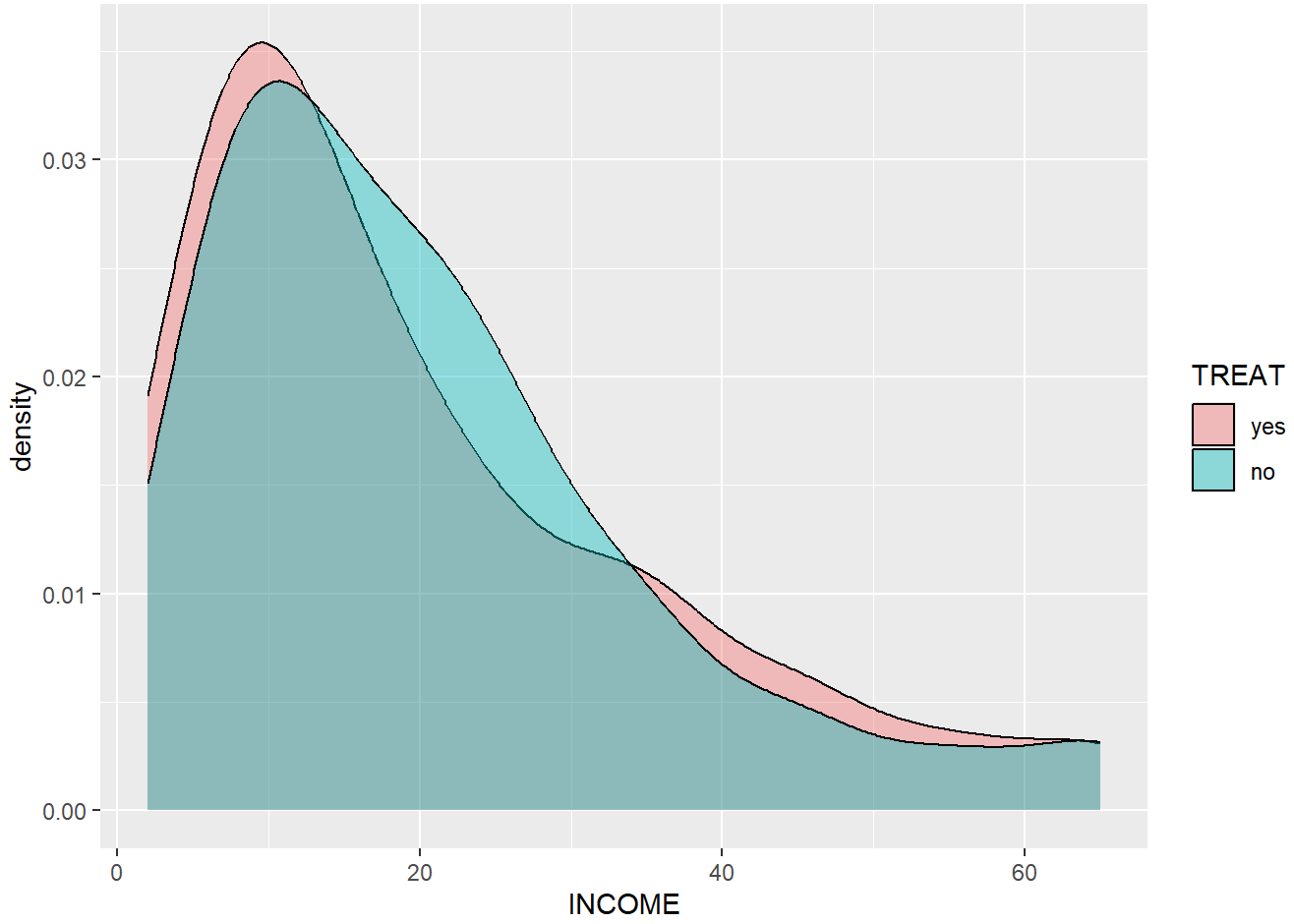


Bivariate Exploration

```
table(depress$INCOME, depress$TREAT)
```

```
##
##      yes no
##   2    5  2
##   4    5  3
##   5    3  7
##   6    9  3
##   7   11  7
##   8    7  7
##   9    9 13
##   11  11  6
##   12   1  1
##   13   7 11
##   15  13 11
##   16   1  0
##   18   0  1
##   19  11 14
##   20   1  2
##   23  10 15
##   24   1  1
##   25   0  1
##   26   0  1
##   27   0  1
##   28   9 10
##   31   0  1
##   32   0  1
##   35  14 10
##   36   0  1
##   37   1  0
##   42   0  1
##   45   9  6
##   55   5  4
##   65   5  5
```

```
ggplot(depress, aes(x = INCOME, fill = TREAT)) + geom_density(alpha=.4)
```



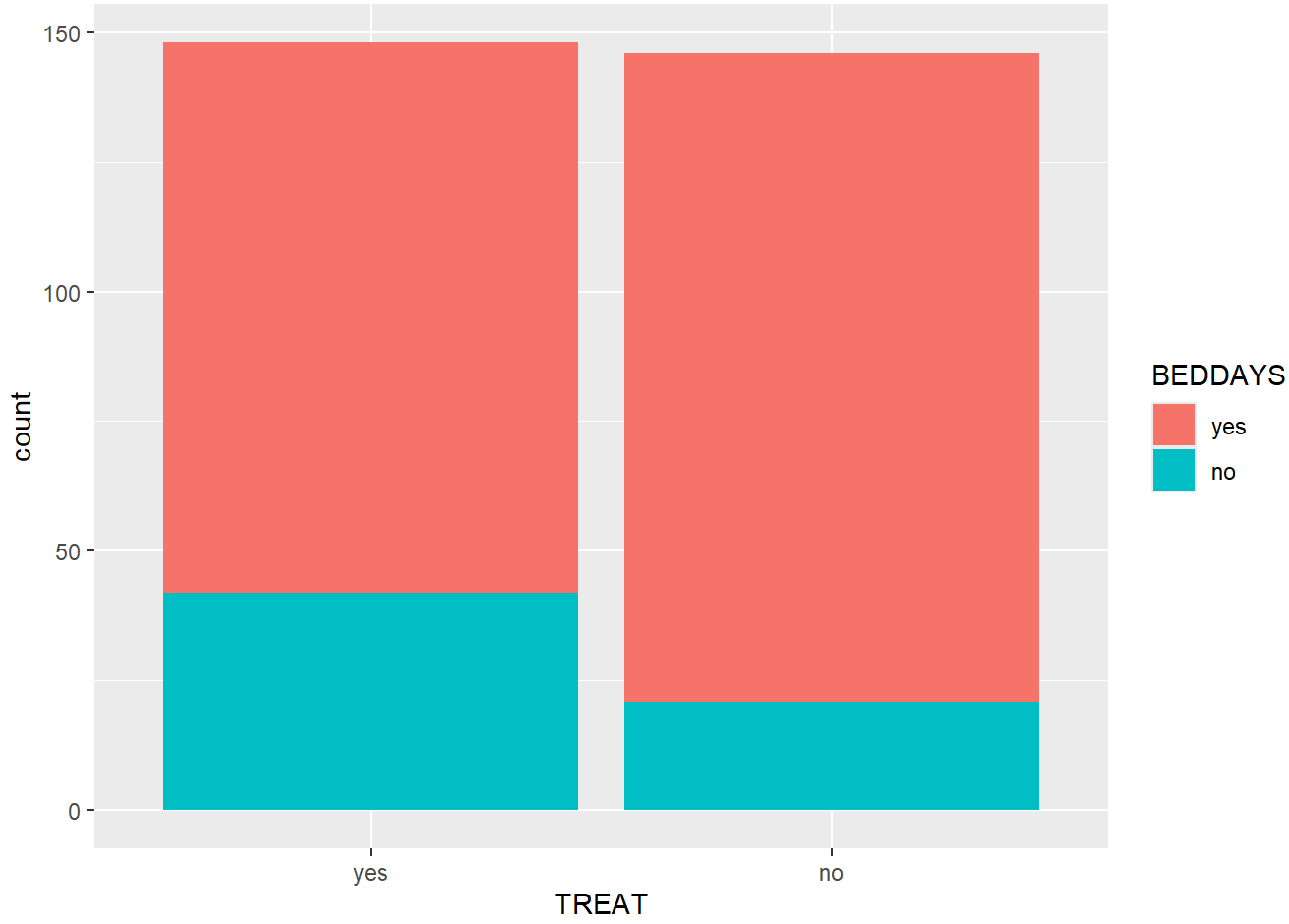

```
table(depress$BEDDAYS, depress$TREAT)
```

```
##
##       yes  no
##   yes 106 125
##   no   42  21
```

```
table(depress$BEDDAYS, depress$TREAT) %>% prop.table() %>% round(3)
```

```
##
##         yes    no
##   yes 0.361 0.425
##   no  0.143 0.071
```

```
ggplot(depress, aes(x=TREAT, fill=BEDDAYS)) + geom_bar()
```



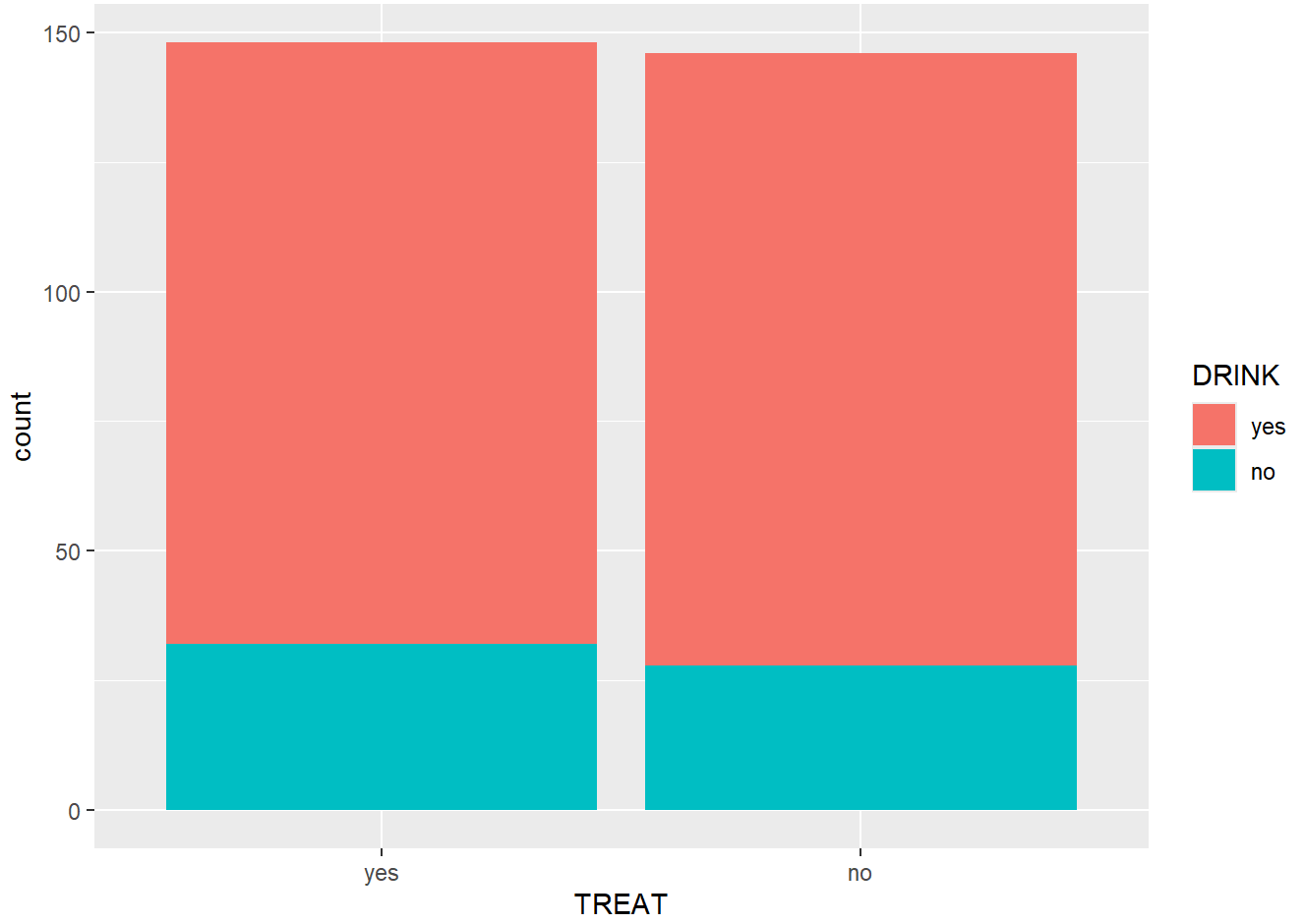

```
table(depress$BEDDAYS, depress$DRINK)
```

```
##
##       yes  no
##   yes 186  45
##   no   48  15
```

```
table(depress$BEDDAYS, depress$DRINK) %>% prop.table() %>% round(3)
```

```
##
##         yes    no
##   yes 0.633 0.153
##   no  0.163 0.051
```

```
ggplot(depress, aes(x=TREAT, fill=DRINK)) + geom_bar()
```



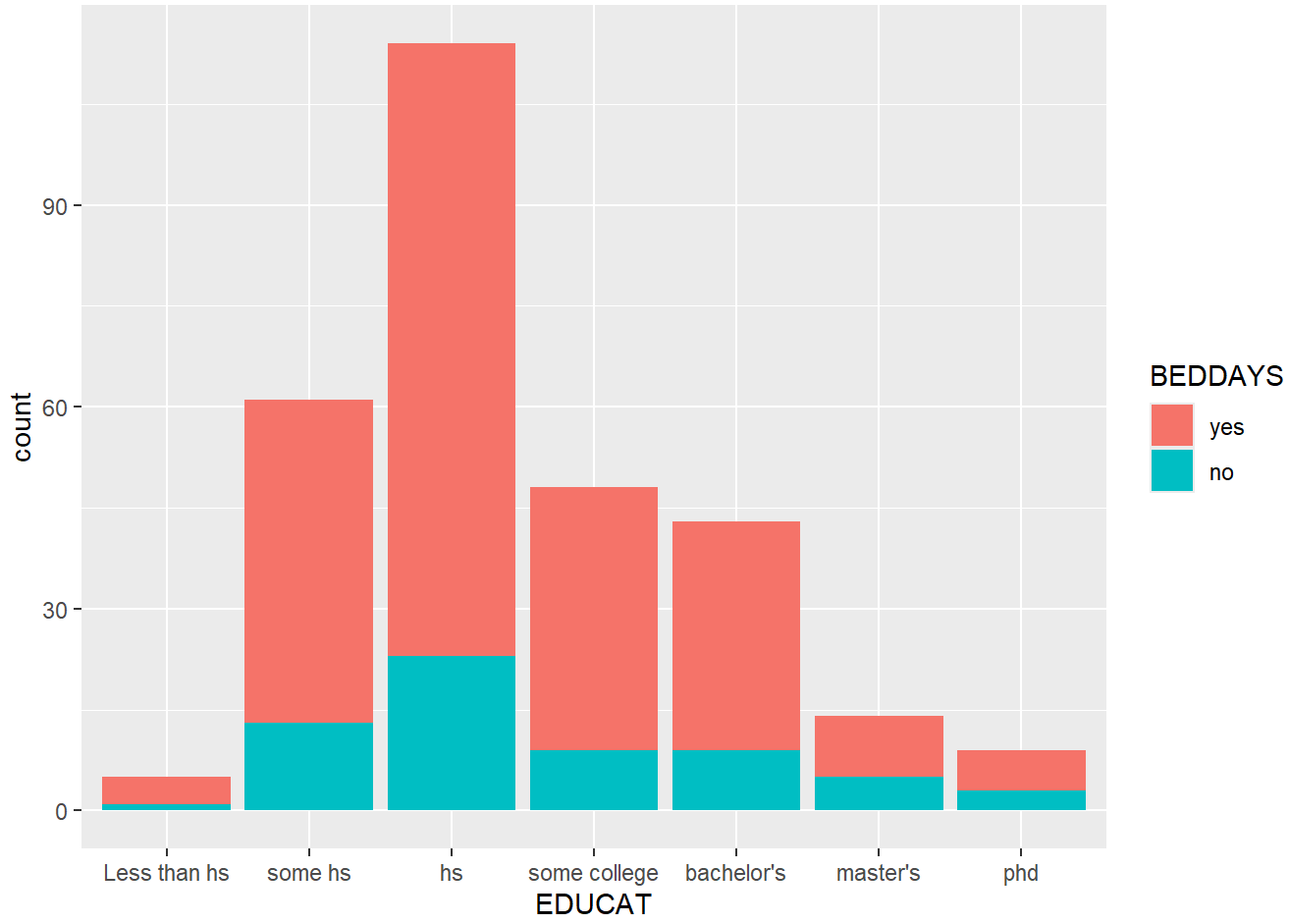

```
table(depress$EDUCAT, depress$BEDDAYS)
```

```
##
##                yes no
##   Less than hs   4  1
##   some hs       48 13
##   hs            91 23
##   some college  39  9
##   bachelor's    34  9
##   master's       9  5
##   phd            6  3
```

```
table(depress$EDUCAT, depress$BEDDAYS) %>% prop.table() %>% round(3)
```

```
##
##                  yes    no
##   Less than hs 0.014 0.003
##   some hs      0.163 0.044
##   hs           0.310 0.078
##   some college 0.133 0.031
##   bachelor's   0.116 0.031
##   master's     0.031 0.017
##   phd          0.020 0.010
```

```
ggplot(depress, aes(x=EDUCAT, fill=BEDDAYS)) + geom_bar()
```



Conclusion

From my bivariate exploration, I found that fewer people who receive treatment have beddays. There are 42 individuals that don't have beddays and are receiving treatment for their depression. There are 21 individuals who don't have beddays and also don't receive treatment. The treatment group has twice as many individuals that don't have beddays.