

Math 130 Project

Jacob Madrid
2/24/2022

```
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
```

```
library(sjPlot)
```

```
## Install package "strengjacked" from GitHub ("devtools::install_github("strengjacked/strengjacked")") to load a
## ll sj-packages at once!
```

The data we have chosen is depression, this is about 294 people being observed in LA county for a study of depression with about 19 variables. The variables we have chosen are employed, drink, and marital. We will test to see who is more likely to get depression. We would like to see the correlation of different variables that are common in depressive people.

```
depress <- read.delim("https://norcalbiostat.netlify.app/data/depress_081217.txt", header=TRUE, sep="\t")
dim(depress)
```

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

This tells us that we have 294 different observations and 37 variables

UNIVARIATE EXPLORATION

Variables being observed

For the first variable, we decided to look at the employment status of those who are depressed, to see if there is a direct correlation between the two.

Employment Status:

```
table(depress$employ)
```

```
##
##      FT Houseperson  In School    Other    PT    Retired
##      167           27           2         4    42         38
##      Unemp
##      14
```

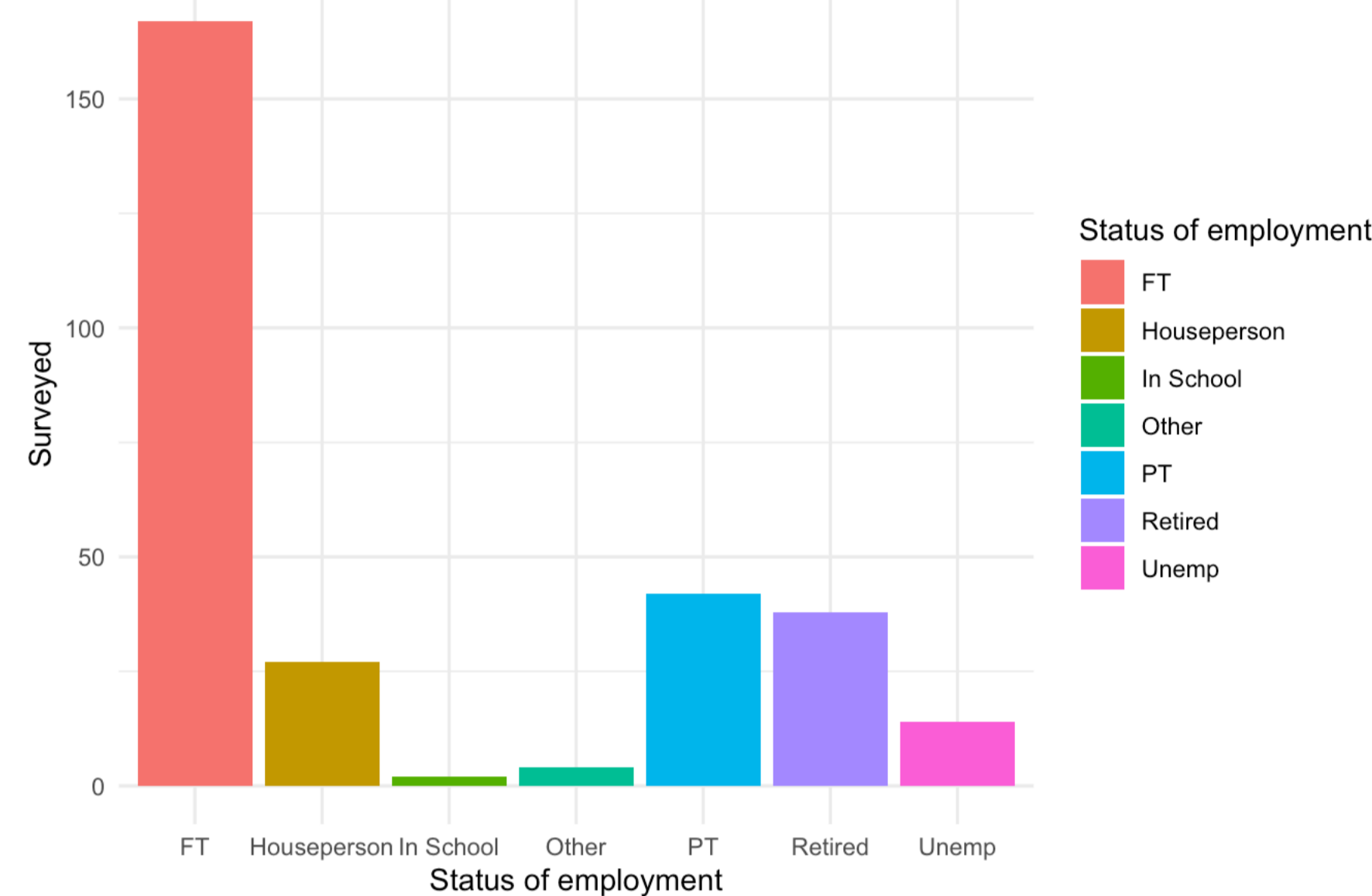
```
summary(depress$employrename)
```

```
## Length Class Mode
##      0 NULL NULL
```

This data point shows how the employment status is for the 294 residents, with the majority of the residents being full time workers

```
ggplot(depress, aes(x=employ, fill=employ)) + geom_bar() + xlab("Status of employment") + ylab("Surveyed") + ggtitle("Depression Rates between People with Different Types of Employment") + scale_fill_discrete(name="Status of employment") + theme_minimal()
```

Depression Rates between People with Different Types of Employment

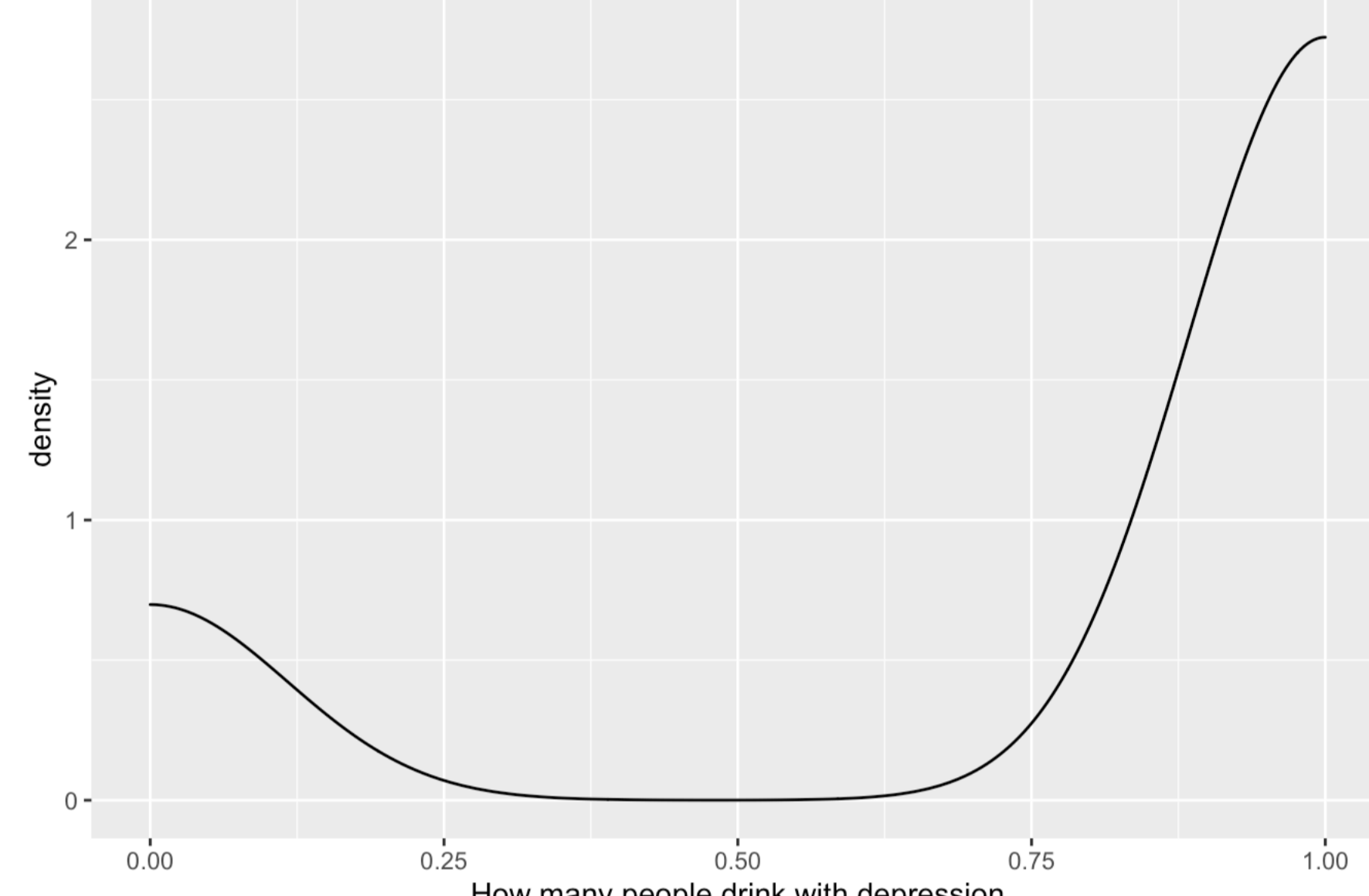


For the second variable, we looked at the drinking habits of the 294 residents to see if drinking is a common trend among those suffering from depression

Drink:

```
ggplot(depress, aes(x=drink)) + geom_density() + xlab("How many people drink with depression") + ggtitle("The correlation between drinks and depression")
```

The correlation between drinks and depression



From the results, it looks as though

people who suffer from depression drink more

For the last variable, we see how the marital statuses of the 294 residents of LA County suffering from depression

Marital:

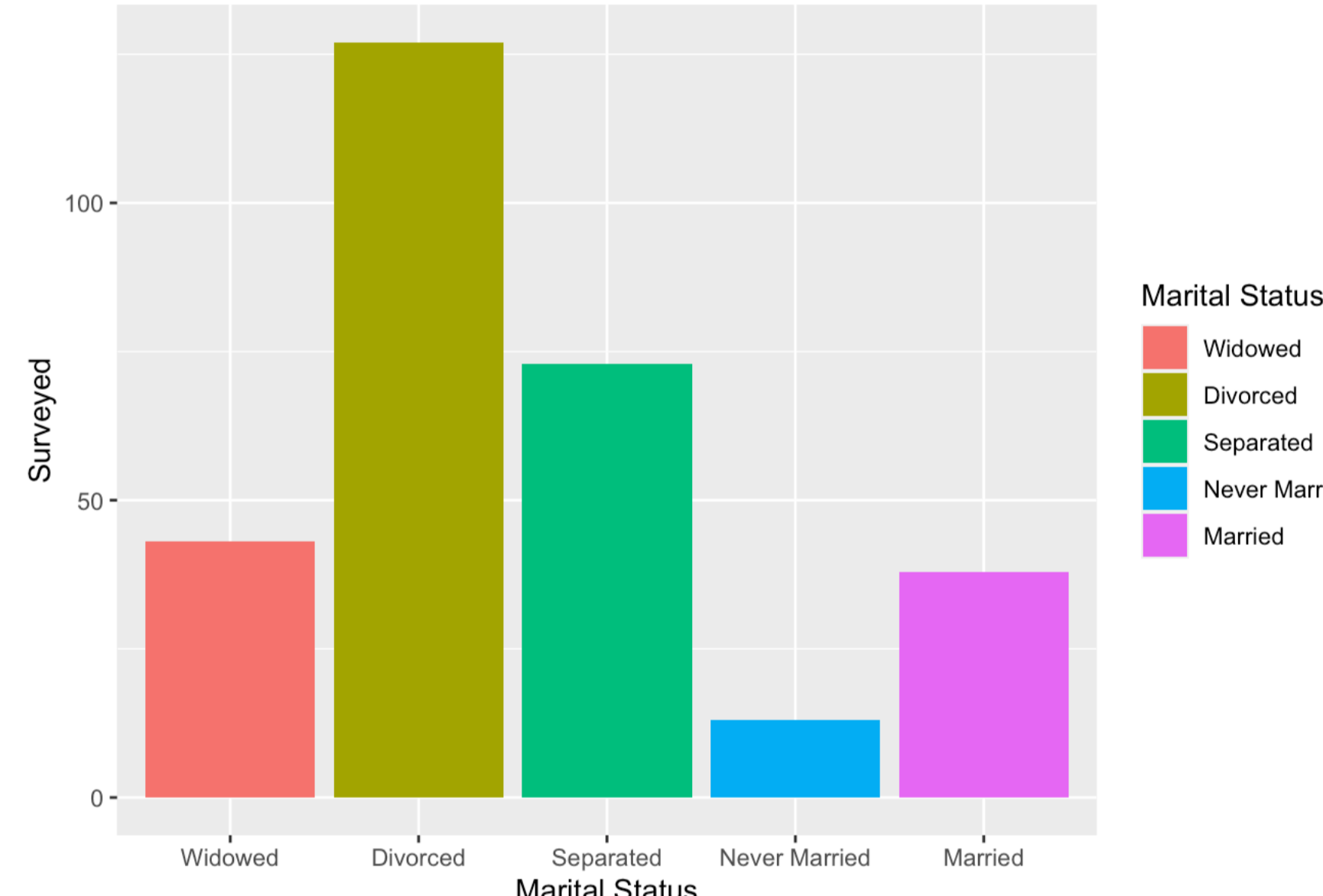
```
depress$maritalrename <- factor(depress$marital, labels=c("Widowed", "Divorced", "Separated", "Never Married", "Married"))
summary(depress$maritalrename)
```

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

As shown in the table, divorced residents are more susceptible to depression

```
ggplot(depress, aes(x=maritalrename, fill=maritalrename)) + geom_bar() + xlab("Marital Status") + ylab("Surveyed") + ggtitle("Depression Rates between Marital Status") + scale_fill_discrete(name="Marital Status")
```

Depression Rates between Marital Status



As seen from the graph, the highest

among depressed residents are those who are divorcees, with separated being the next closest. Never married is the lowest group.

CESD Depression Scores

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

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

As shown here, these are the depression scores from all groups, with all the 294 residents having a mean of 8.884 and a median of 7. The lowest score was 0, with the highest being 47.

BIVARIATE EXPLORATION

Bivariate: two variables together

For our first bivariate variable, we will look at the correlation between employment and the CESD scores

Employment and CESD

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

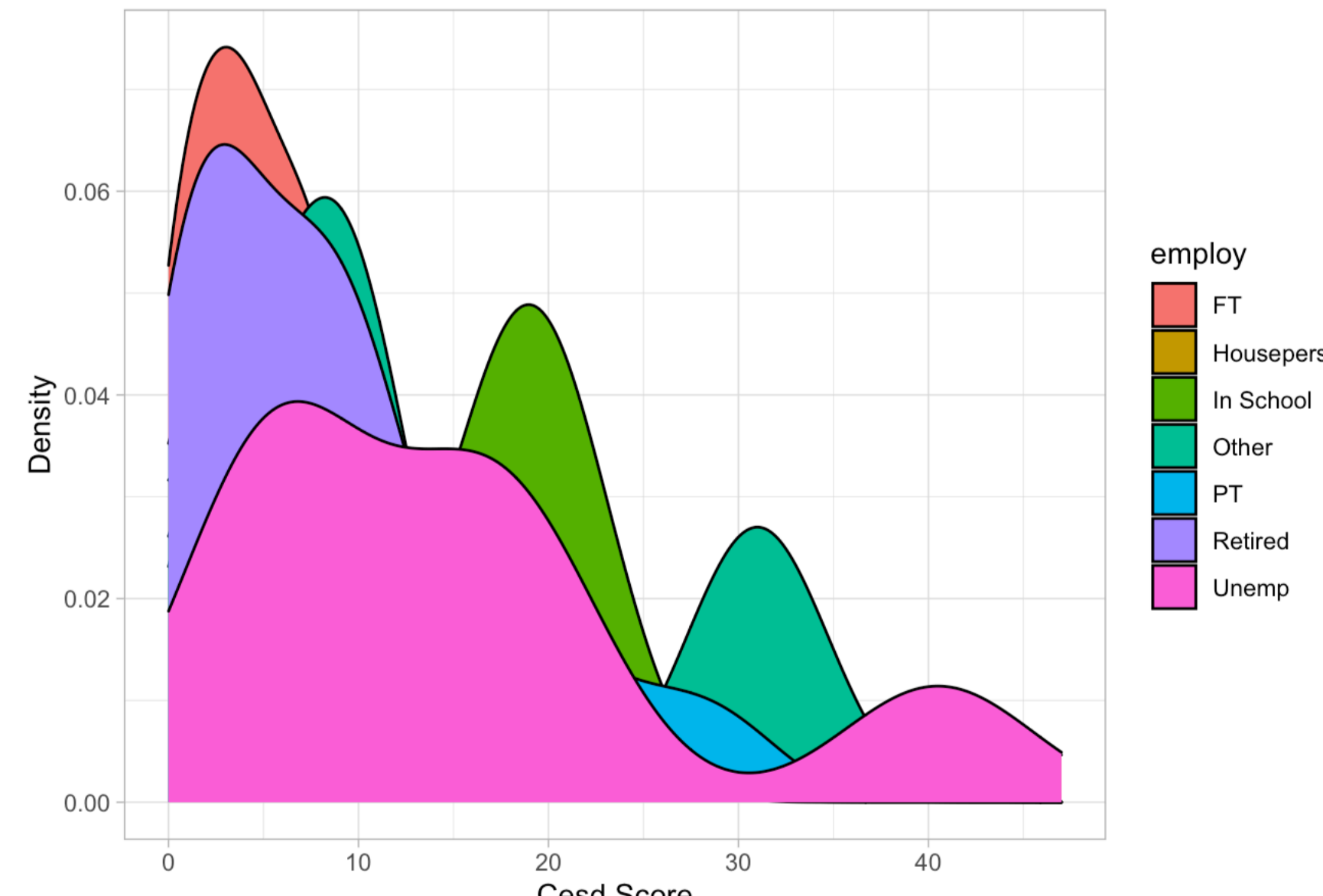
```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.000   3.000   7.000   8.884  12.000  47.000
```

```
summary(depress$employrename)
```

```
## Length Class Mode
##      0 NULL NULL
```

```
ggplot(depress, aes(x=cesd, fill=employ)) + geom_density() + scale_fill_discrete(name="employ") + xlab("Cesd Score") + ylab("Density") + ggtitle("Relationship between Employment and Cesd scores") + theme_light()
```

Relationship between Employment and Cesd scores



As shown in the graph, fulltime

workers usually have a higher CESD score compared to an unemployed person.

Now we will look at drinking and CESD scores

Drink and CESD Scores

```
summary(depress$drink)
```

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.000   1.000   1.000   0.7959  1.0000  1.0000
```

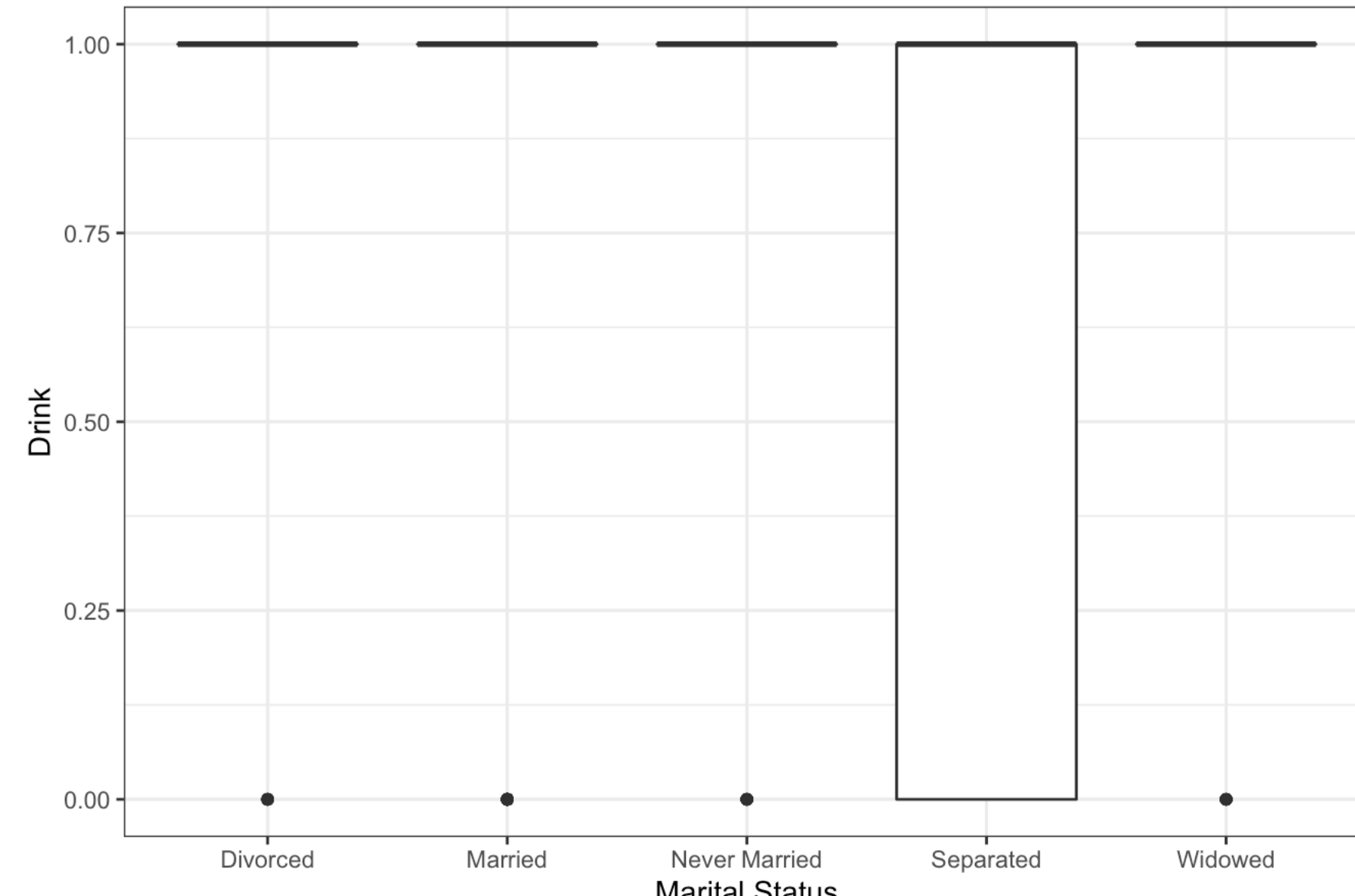
Here are the correlating stats of depression and marital status

```
summary(depress$marital)
```

```
##      Length Class Mode
##      294 character character
```

```
ggplot(depress, aes(y=drink, x=marital)) + geom_boxplot() + theme_bw() + xlab("Marital Status") + ylab("Drink") + ggtitle("The Distribution of Drink Based on Marital Status")
```

The Distribution of Drink Based on Marital Status



As shown in the graph, separated

people are more likely to be heavy drinkers.