Depression

Ger Xiong

9/18/23

```
##
## 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

## [1] 294  37
```

Introduction

The data set I'll be analyzing is in regards to depression. This data set consists of 294 observations of 37 variables. Through this data I want to see how someone's martial status and employment may affect their well-being. Does one's marital status affect and contribute to an increase in depression from one's employment and vice-versa. I hypothesize that being married or divorced and working full time will have the highest depression rates.

Univariate Exploration

Marital

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

```
##
##       Widowed      Divorced       Married Never Married     Separated
##            43           127            73            13            38
```
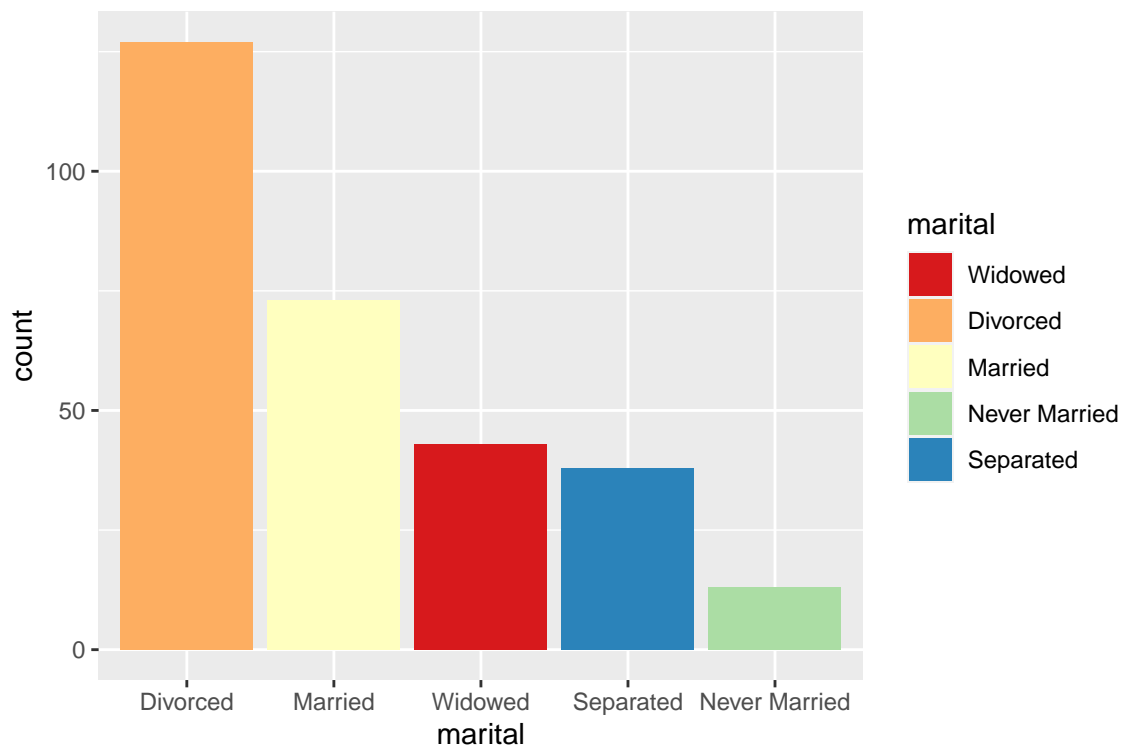
```r
(127/294) * 100
```

```
## [1] 43.19728
```

```
(13/294) * 100
```

```
## [1] 4.421769
```

First, I evaluated with a table to find out how many people are widowed, divorced, married, never married and separated. Secondly, I calculated the percentage for the highest and lowest depression rates.

```
ggplot(depress, aes(x = forcats::fct_infreq(marital), fill = marital)) + geom_bar() +
    xlab("marital") + scale_fill_brewer(palette = "Spectral")
```


Then, from my table and bar graph I was able to establish that those whose marital status as divorced has a higher depression rate coming in at 43.2% and those with the least are individuals who've never been married before at a rate of 4.4%.

Employment

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

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

```
167 + 27 + 2 + 4 + 42 + 38 + 14
```

```
## [1] 294
```
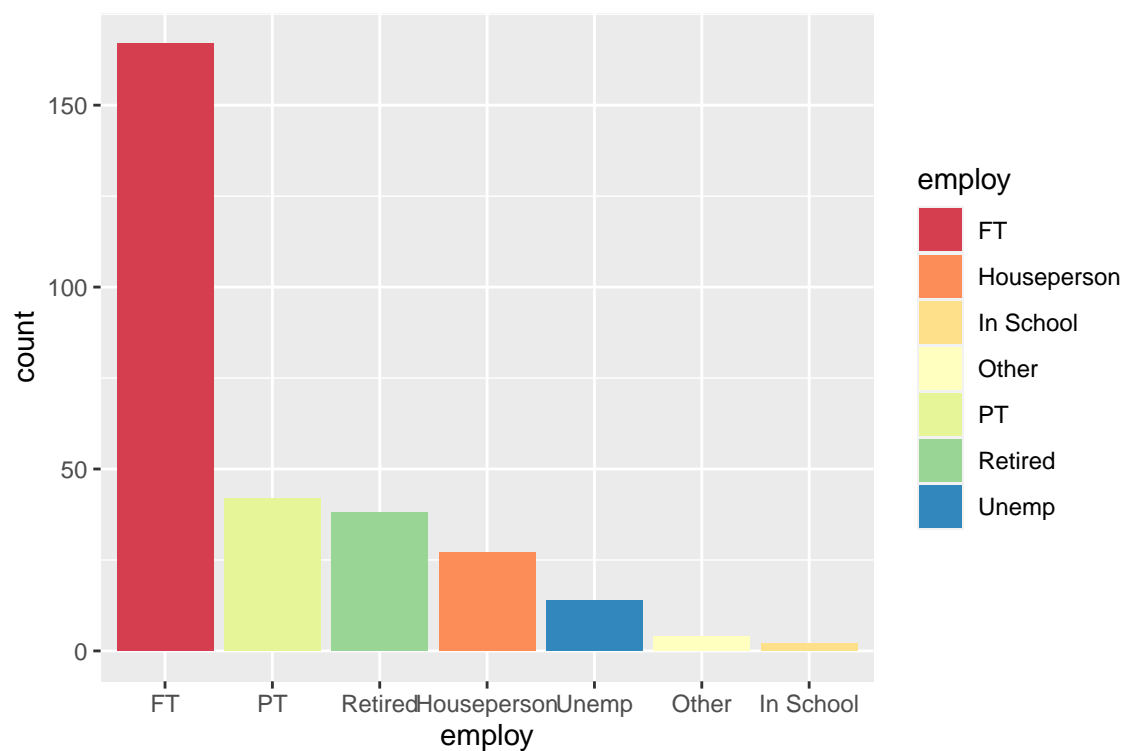
```
(167/294) * 100
```

```
## [1] 56.80272
```

```
(2/294) * 100
```

```
## [1] 0.6802721
```

My second set of table consists of employment status whether an individual is full time, part time, retired, house-person, unemployed, other or in school.

```
ggplot(depress, aes(x = forcats::fct_infreq(employ), fill = employ)) + geom_bar() +
    xlab("employ") + scale_fill_brewer(palette = "Spectral")
```


From the data sets it can be established that full time employees have a higher depression rate at 56.8% and those who are in school have the least depression rate at about 0.7%.

Bivariate Exploration

```r
table(depress$marital, depress$employ)
```
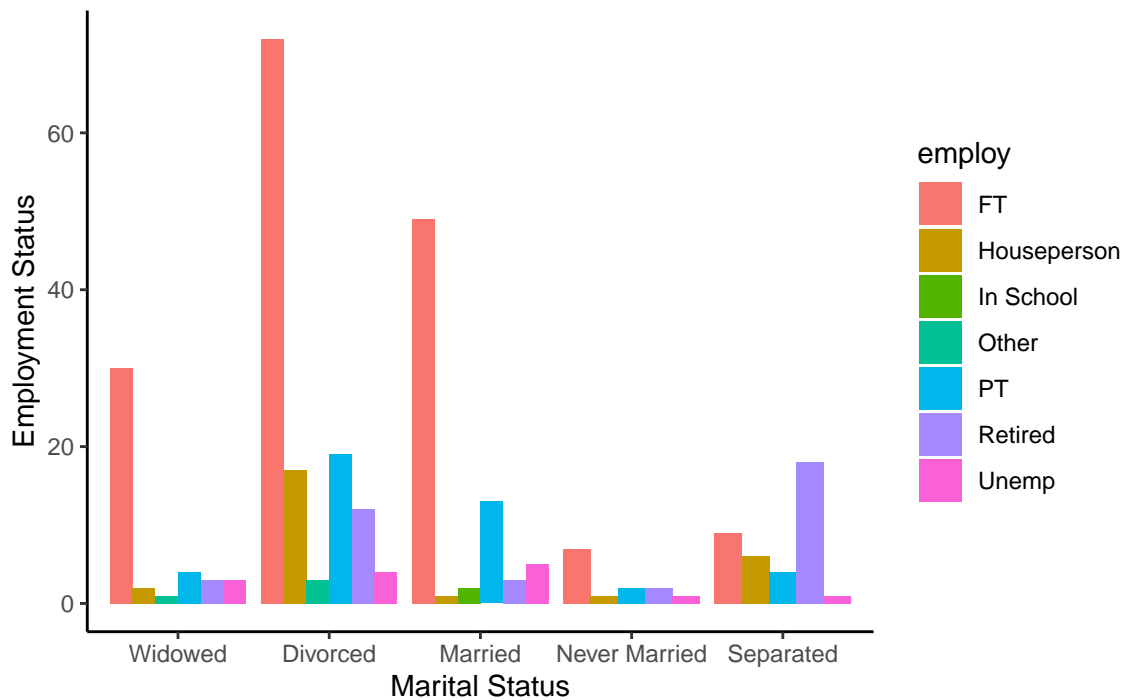
```
##
##                 FT Houseperson In School Other PT Retired Unemp
##   Widowed       30           2         0     1  4       3     3
##   Divorced      72          17         0     3 19      12     4
##   Married       49           1         2     0 13       3     5
##   Never Married  7           1         0     0  2       2     1
##   Separated      9           6         0     0  4      18     1
```

This table identifies that the marital status divorce makes up the majority of individuals who are employed full time. From this information we can theorize that the highest depression rate will be coming from those who are divorced and employed full time.

```r
ggplot(depress, aes(x = marital, fill = employ)) + geom_bar(position = "dodge") +
    theme_classic() + ggtitle("Depression Rate Between Marital Status & Employment") +
    ylab("Employment Status") + xlab("Marital Status") + scale_color_brewer("Set1")
```



```r
table(depress$marital, depress$employ) %>%
    prop.table(margin = 1) %>%
    round(2)
```

```
##
##                 FT Houseperson In School Other   PT Retired Unemp
##   Widowed       0.70        0.05      0.00  0.02 0.09    0.07  0.07
##   Divorced      0.57        0.13      0.00  0.02 0.15    0.09  0.03
```

```
##   Married       0.67        0.01        0.03  0.00 0.18    0.04  0.07
##   Never Married 0.54        0.08        0.00  0.00 0.15    0.15  0.08
##   Separated     0.24        0.16        0.00  0.00 0.11    0.47  0.03
```

Conclusion

Overall, upon comparing the two data sets it can be concluded that those who were widowed and employed full time actually make up the highest depression rate. Surprisingly, those who were divorced and employed full time did not have the highest depression rate even though they had the highest rate on a univariate level. That's because even though there were a greater number of divorcees they were outnumbered by the smaller amount of widows. Even though there weren't as much widows the group had a greater amount of depressed individuals compared to the others (therefore they had a higher ratio). From the table the highest depression rate came from those who were employed full time only differing based on marital status: widows came in at 70%, married at 67%, divorced at 57%, never married at 54% and separated at 24%. All in all I found this data to be very interesting. I had hypothesized that divorced or married individuals who worked full time would have the highest depression rates. But after running my data I was incorrect because widows employed at full time made the top percent.