

Final Project

2022-09-19

Rates of Depression on Individuals of Various Religious Denominations

Introduction

The data set I will be working from in my project is the “depression” dataset. It is from an interview study of depression on the adults whom resided in Los Angeles. It includes 294 observations and looks at things like drinking habits, education, religious affiliation and much more. For this analysis however, I will be looking to see if there is any relationship between specific religious practices and rates of depression using the “relig” and “cesd” data sets. First, we will look at both of these variables individually then we will analyze them together.

Univariate Exploration

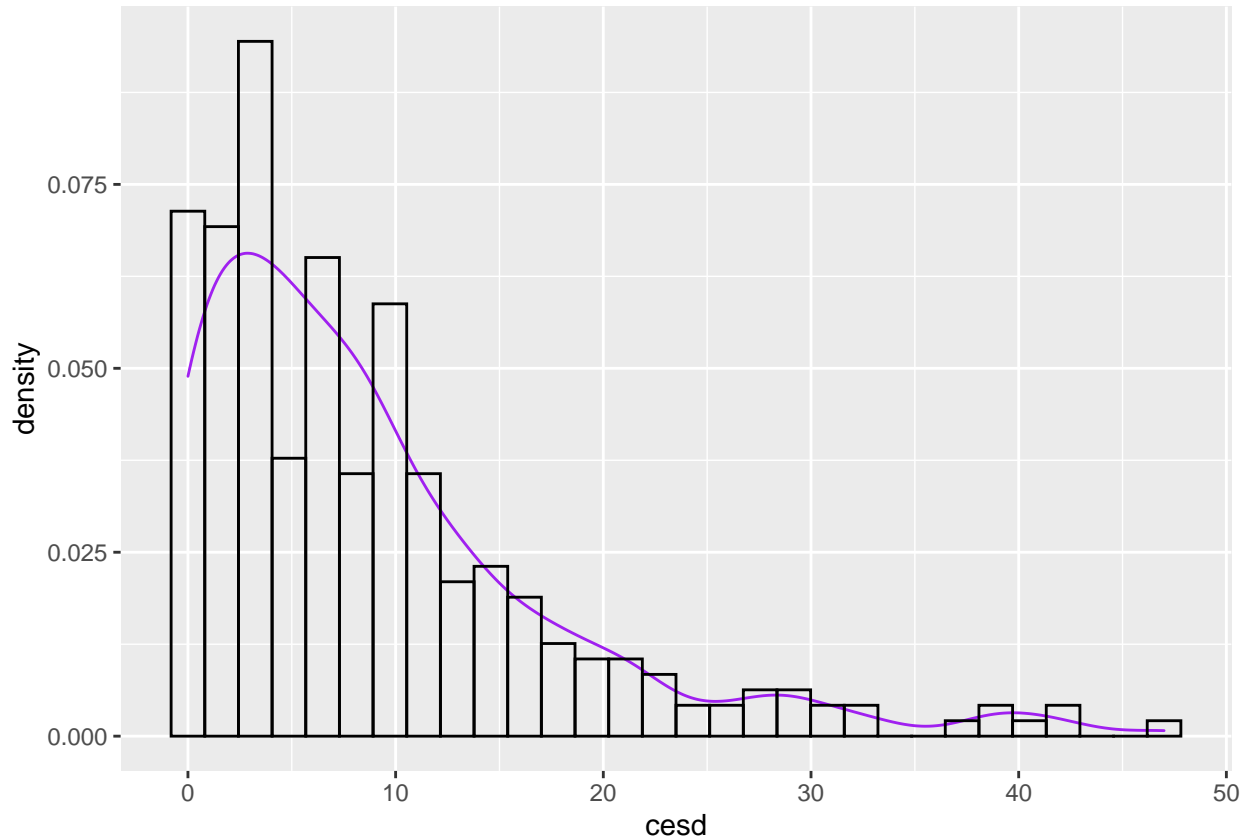
For the “cesd” dataset, they were rating people symptoms of depression on a scale from 0 (no depressive symptoms) to 60 (severely depressed). Per the study, participants who had scores that were above 16 were considered to be depressed. We can see that our median value is 7.0, our third quartile is only 12 so I would imagine from these calculations that we will have far more people with mild depressive symptoms but not clinically depressed versus those that are. We will look at it in a density plot to get a better idea of the spread of our data.

```
summary(depression$cesd)
```

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

```
ggplot(depression, aes(x=cesd))+ geom_density(col="purple") + geom_histogram(aes(y=..density..), colour="yellow")
```

```
## ‘stat_bin()’ using ‘bins = 30’. Pick better value with ‘binwidth’.
```



Looking at the histogram with density plot overlaid, we can see the distribution of our data. It looks like the larger portion of the population sampled were lower on the cesd scale than those that scored high.

Now, we will look at the religion data:

```
table(depression$relig, useNA = "always")
```

```
##
##  1  2  3  4 <NA>
## 155 51 30 56  2
```

These numbers don't mean much to us yet. It will make more sense after reformatting the numbers to represent the categories of religions they represent.

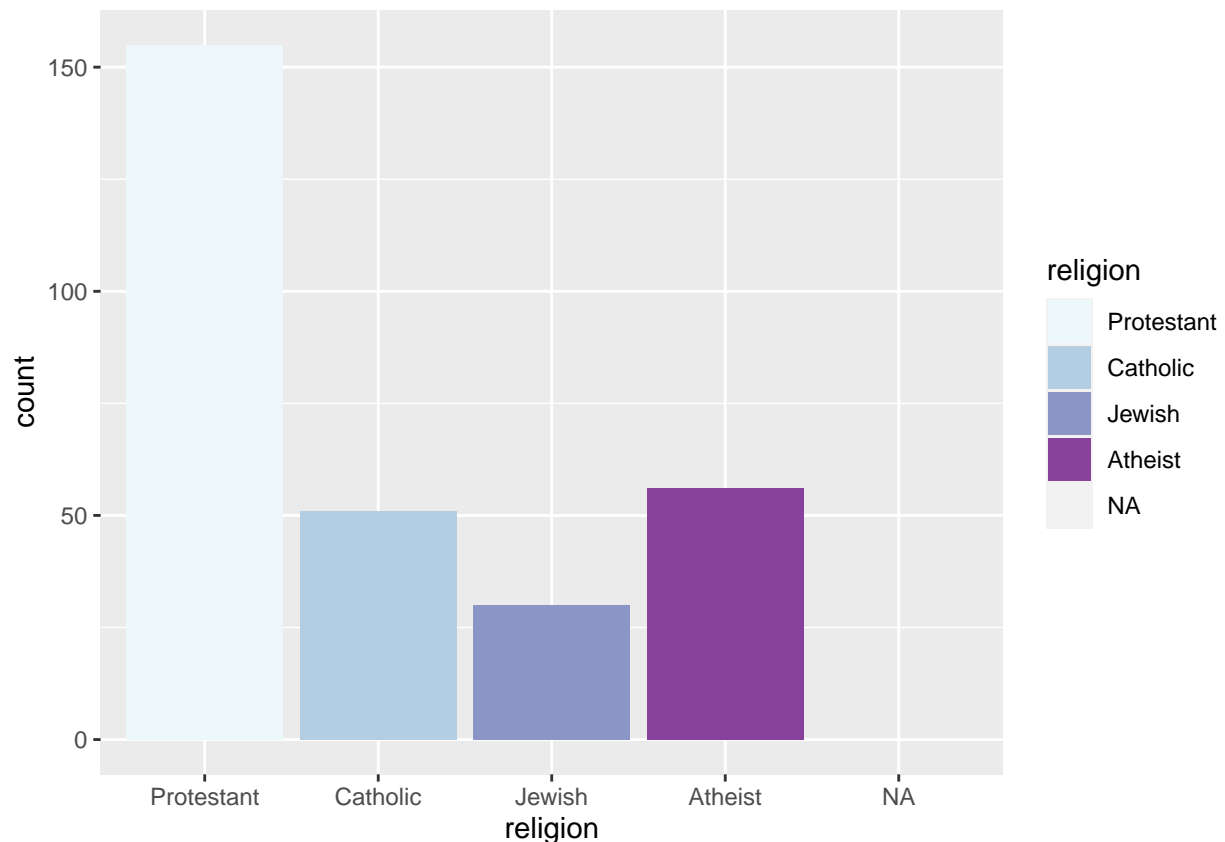
```
depression$religion <- factor(depression$relig, labels = c("Protestant", "Catholic", "Jewish", "Atheist"))
table(depression$relig, depression$religion, useNA="always")
```

```
##
## Protestant Catholic Jewish Atheist <NA>
## 1 155 0 0 0 0
```

```
## 2      0      51      0      0      0
## 3      0      0      30     0      0
## 4      0      0      0      56     0
## <NA>    0      0      0      0      2
```

Awesome! Now we can see that in our sample, we have 155 Protestants, 51 Catholics, 30 Jewish, 56 Atheist and two participants who did not answer. Let's throw a bar chart up to get a visual representation of the sizes of the various groups.

```
ggplot(depression, aes(x=religion, fill=religion)) + geom_bar() + scale_fill_brewer(palette="BuPu")
```



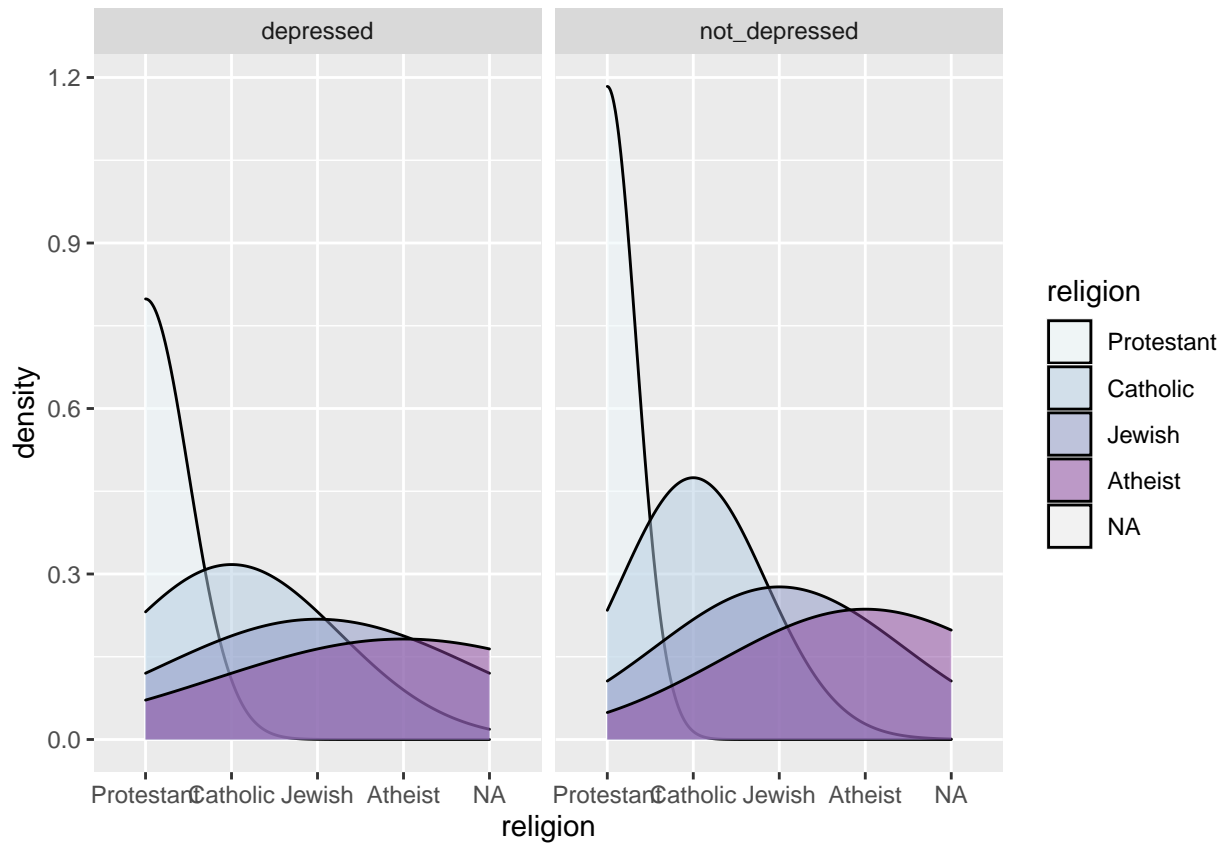
Bivariate Exploration ### Okay, so now that we feel that we have a good grasp on our individual data points, let's take a look at combining the two. Let's explore what relationship religious views have on rates of depression. For our purposes here, we will say that those with a score over 16 on the cesd variable chart have depression and those who are under do not.

```
depression$depressed = ifelse(depression$cesd > 16, "depressed", "not_depressed")
table(depression$depressed)
```

```
##
## depressed not_depressed
##          45          249
```

```
ggplot(depression, aes(x=religion, fill=religion)) + geom_density(alpha=.5) + facet_wrap(~depressed)+ s
```

```
## Warning: Groups with fewer than two data points have been dropped.  
## Groups with fewer than two data points have been dropped.  
  
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning  
## -Inf  
  
## Warning in max(ids, na.rm = TRUE): no non-missing arguments to max; returning  
## -Inf
```



Conclusion

Looking at the two density plots comparing religious denominations to the cesd score of over 16, we see that the plots look very similar in their distribution. It looks like the Catholic and Protestant comparisons of density of population with and without depression have the most variation. There was a tripled over-representation of Protestants to the other religious denominations so that is likely why we are getting much more effect in our chart. Besides that, they all seem to follow the same patterns, there is no outlier that seems to be experiencing more depression in relation to others when taking population sampled into account.