Exploratory Data Analysis Project

Ryan S

2024-03-01

1) Introduction In this analysis we will explore the relationships between Religious Affiliation, Education Level, and status of Depression.

This dataset contains information on individuals in Los Angeles who participated in a prospective study on depression, including their religious affiliation, education level, and status of depression, among others. I will go into further depth on how some of these variables were calculated.

My research question is: Is there an association between the religious affiliation, education level, and risk of depression in individuals living in Los Angeles?

I hypothesize that the highest levels of depression will be from those associated with no religion, and that those with higher education levels will tend to be not religious.

2) Univariate Exploration

First let's examine the spread and distributions of the variables I've chosen to analyze. These will give us an idea of each variable on its own and will help the eventual comparison of the variables with each other.

Religious Affiliation

```r
depress$RELIG <- factor(depress$RELIG, labels=c("Protestant", "Catholic", "Jewish", "None", "Other"))

relig_dist <- table(depress$RELIG)

barplot(relig_dist, main="DISTRIBUTION OF RELIGIOUS AFFILIATIONS", xlab="Religion", ylab="Frequency", co
```

DISTRIBUTION OF RELIGIOUS AFFILIATIONS
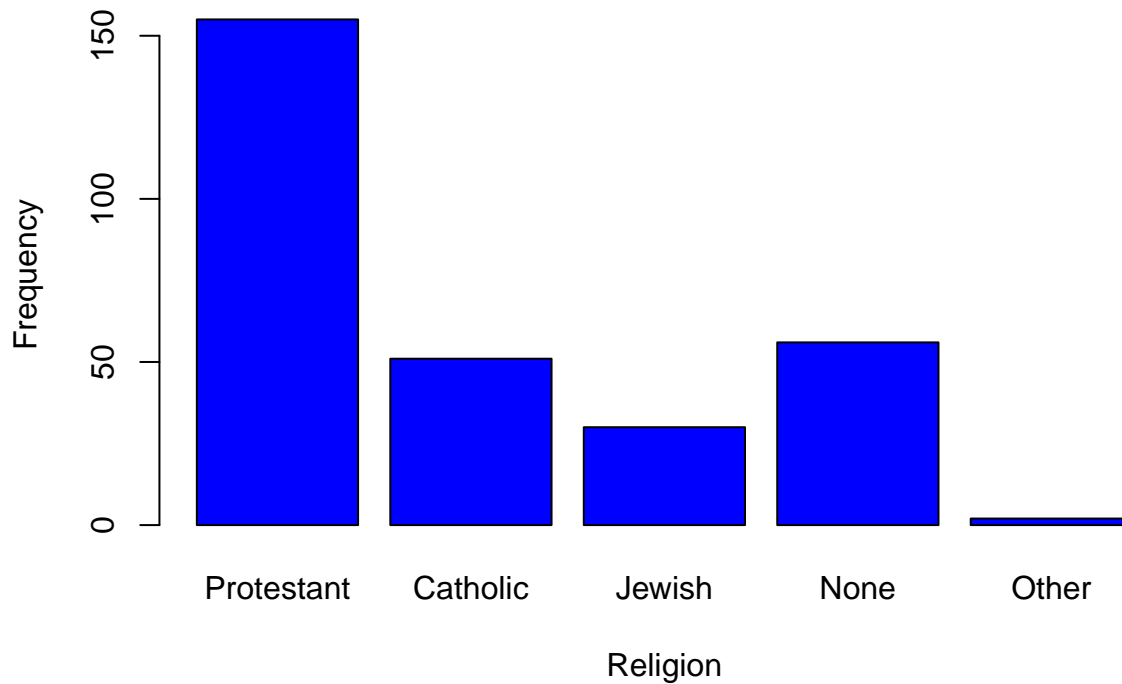
The highest frequency of religious affiliants were clearly Christian, with the subsect of protestant having the highest overall frequency. Following protestant in frequency was no religion, then Catholic, Jewish, and all other religions.

```r
depress$EDUCAT <- factor(depress$EDUCAT, labels = c("< HS", "Some HS", "HS diploma", "Some College", "Ba

edu_lev <- table(depress$EDUCAT)

barplot(edu_lev, main="DISTRIBUTION OF EDUCATION LEVELS", xlab="Education Level", ylab="Frequency", col=
```

DISTRIBUTION OF EDUCATION LEVELS
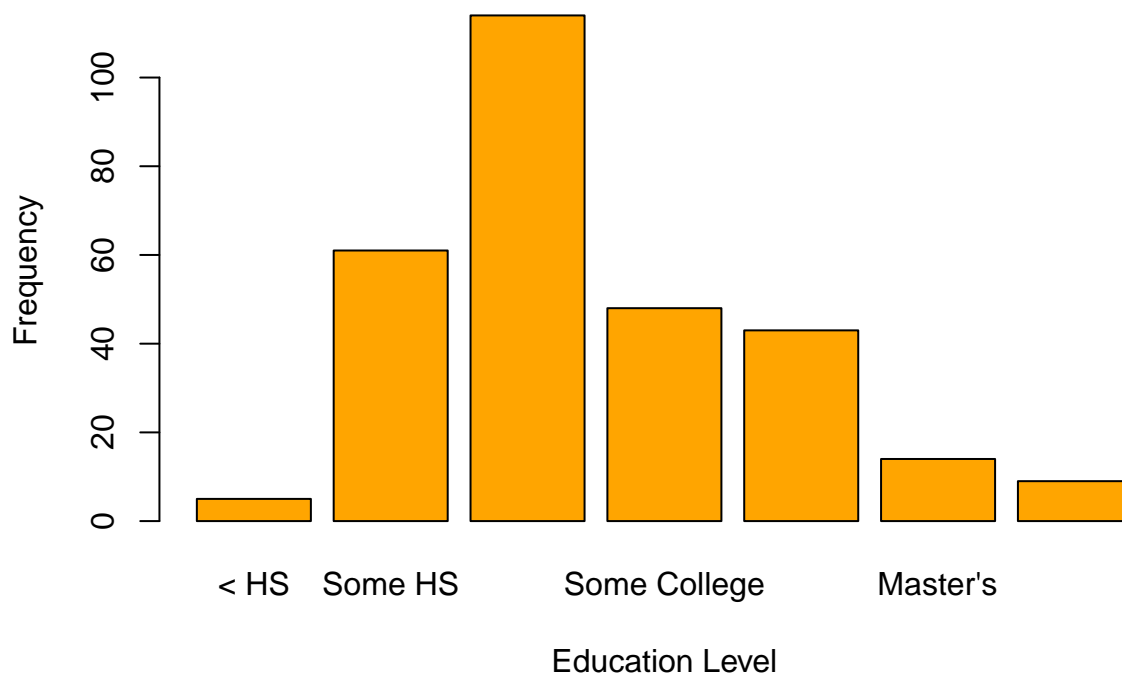
In ascending order, the bins represent increasing education levels, starting at less than a high school education, then some high school education, a high school diploma, some college, a bachelor's degree, master's degree, and finally doctorate. From this study, the large majority of people sampled had a high school diploma, but did not go one to any further education.

```
depress.score <- table(depress$CESD)

barplot(depress.score, main="DISTRIBUTION OF DEPRESSION SCORES", xlab="Depression Score", ylab="Frequen
```

DISTRIBUTION OF DEPRESSION SCORES
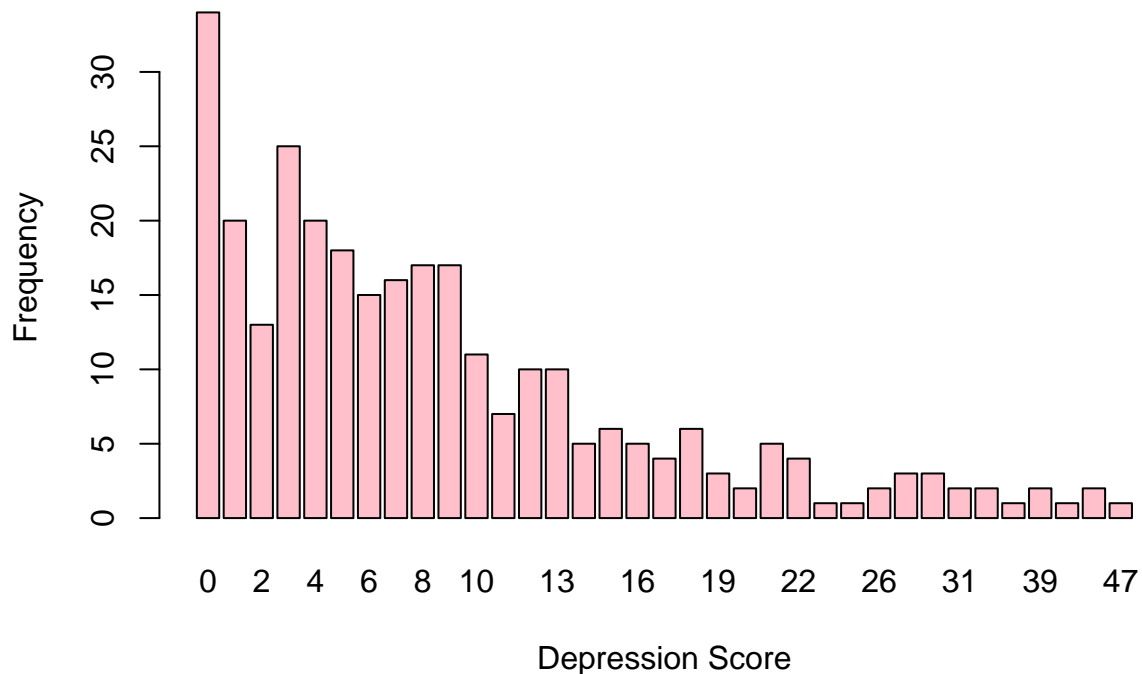
Depression scores were collected by asking participants to identify the number of times during the past week they had thought, felt, or acted a certain way based off of an established depression scale questionaire. A total score of 16 or higher was identified as "Depressed" while a total score lower than 16 was identified as "Normal".

```
table(depress$CASES)
```
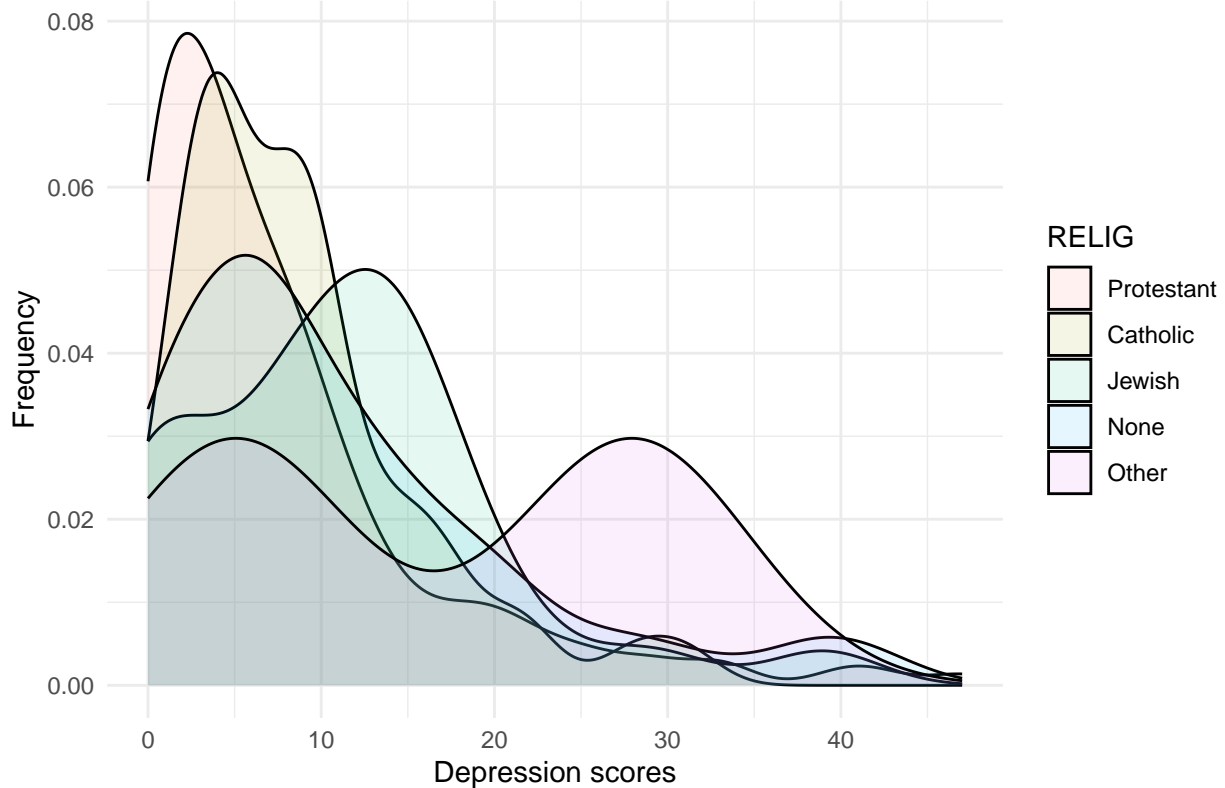
```
##
##   0   1
## 244  50
```

3) Bivariate Exploration:

```
ggplot(depress, aes(x = CESD, fill = RELIG)) + geom_density(alpha=0.1, position = "dodge") + theme_minir
ylab("Frequency") + xlab("Depression scores") + scale_color_brewer("Set12")
```

```
## Warning: Width not defined
## i Set with `position_dodge(width = ...)`
```

## Depression Rates Among Religious Affiliations

From this chart, we can see that the highest depression scores were those from non-Christian and non-Jewish religious affiliations. Since "other" is ambiguous here, it is hard to draw many conclusions from this chart alone, as it could include a variety of religions. The people sampled who identified as Jewish also tended to have slightly higher scores than the other groups sampled.

```
ggplot(depress, aes(x = CESD, fill = EDUCAT)) + geom_histogram(position = "dodge") +
theme_classic() + ggtitle("Depression Rates Among Education Levels") +
ylab("Frequency") + xlab("Depression scores") + scale_color_brewer("Set1")
```

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

Depression Rates Among Education Levels
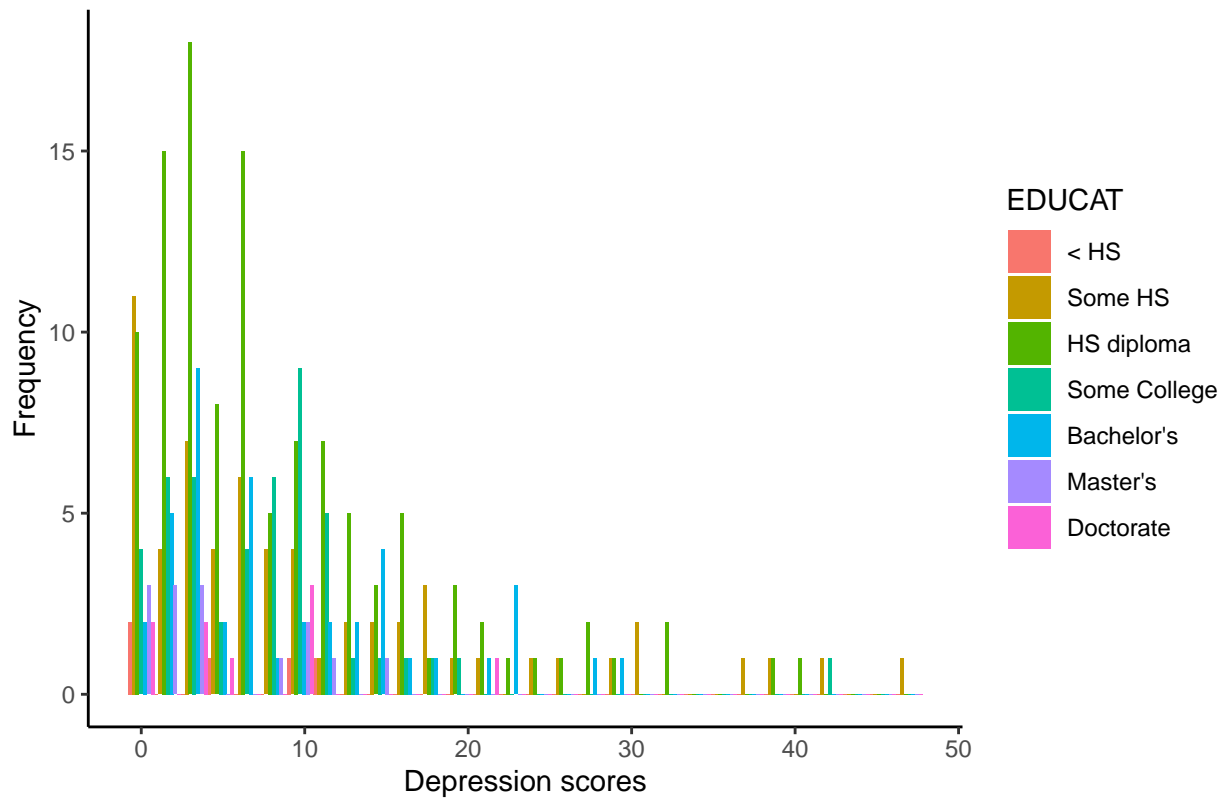
From this histogram, we can see that there were almost no high depression scores associated with higher level education (Bachelor's Degree and beyond). Most of the highest depression scores had lower than a high school education.

```
ggplot(depress, aes(x = EDUCAT, fill = RELIG)) + geom_bar(position = "dodge") + theme_minimal() + ggtitl
ylab("Frequency") + xlab("Education Level") + scale_color_brewer("Set12")
```

Religious Affiliations by Education Level
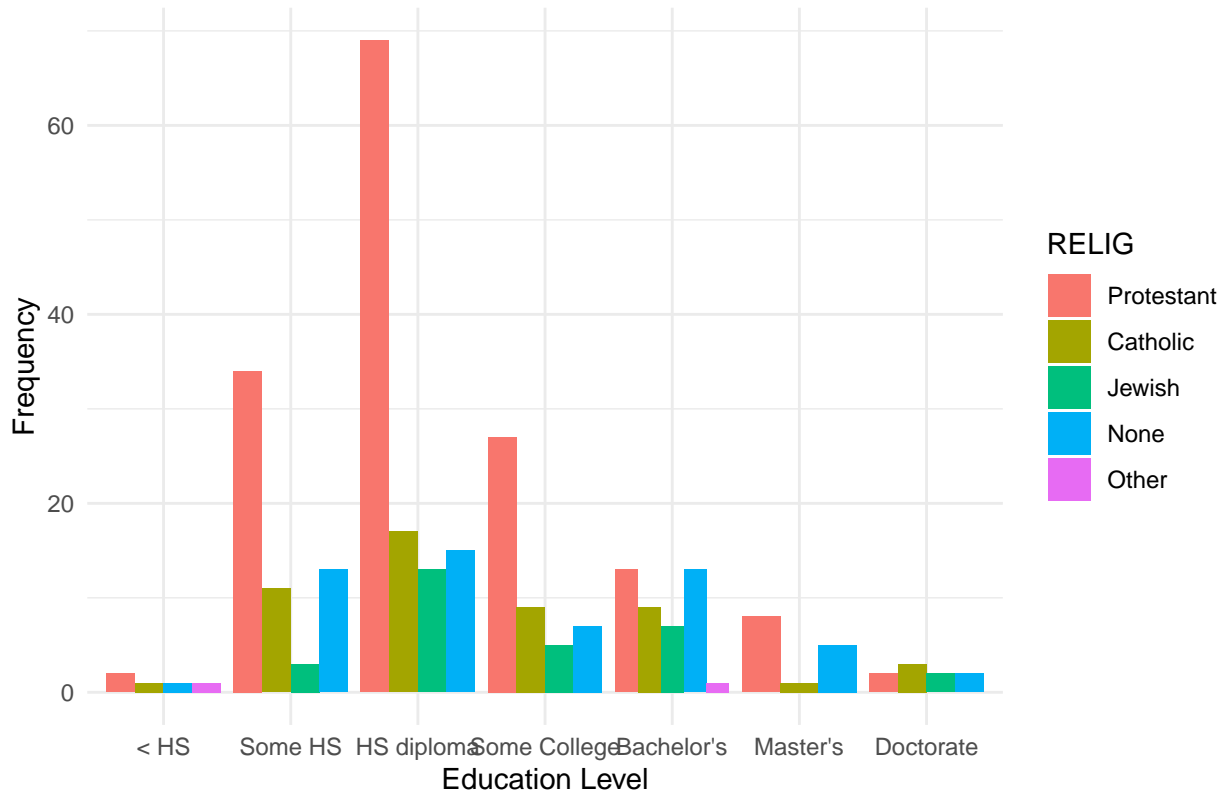
With this bar chart, we can see the distribution of religious affiliations among the various education levels. Likely due to a small sample size, our chart does not tell us much about the spread, since its shape is very similar to the shape of the education level bar chart on its own. It is interesting to note that most of the sampled Protestants did not go on to pursue education past high school, while the distribution of other religions was more evenly distributed around education levels.

4) Conclusions

Due to the small sample size and the variables I chose to explore, it is hard to come to any concrete conclusions regarding the relationships between religion, education, and depression. Most of the people sampled in the study were not recognized as being depressed from their scores, so the small number that were is too prone to association/causation errors. My hypothesis that the group that was not religious would include the highest overall depression scores was false. In fact, the data show that it was actually the group of all other religions that accounted for the highest depression scores. My hypothesis that the non-religious group would have the highest levels of education was also false. The highest amount of people with Master's Degrees came from the Protestant group and the highest amount of people with Doctorate's came from the Catholic group.