```
---
title: "Project"
output:
  html_document: default
  pdf_document: default
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
library(ggplot2)
library(dplyr)
library(sjPlot)
```

`INTRODUCTION`

The data set that is being used here was gathered from Los Angeles County, and focuses on the relationship between several variables and depression. This project will focus on two variables, those being religion and depression scores. The research question of this project is, "What is the correlation between religion and depression rates, and are there differences in each variable?". I will be interested in what the analysis may suggest about religion's affect on long term happiness.

```{r}
Depression <- read.table("/Users/grantheath/Desktop/MATH130/DATA/
depress_081217.txt",header=TRUE, sep="\t")
Depression$relig[Depression$relig==1]<-"Protestant"
Depression$relig[Depression$relig==2]<-"Catholic"
Depression$relig[Depression$relig==3]<-"Jewish"
Depression$relig[Depression$relig==4]<-"None"
Depression$relig[Depression$relig==5]<-"Other"
Depression$relig[is.na(Depression$relig)]<-"Other"
```

`UNIVARIATE EXPLORATION`

The data being explored here consists of five variables. It observes the rates of depression and their frequency in Catholic, Jewish, Protestant, and Other religions. There is also a variable for no religion to compare them to. Depression is a quantitative variable that was determined in the data set as the sum of the scores from 20 different questions about frequency of negative feelings. The higher the score the more negative feelings.

```{r}
Pcesd<-filter(Depression, relig=="Protestant")
mean(Pcesd$cesd)
Ccesd<-filter(Depression, relig=="Catholic")
mean(Ccesd$cesd)
```
```

```
Jcesd<-filter(Depression, relig=="Jewish")
mean(Jcesd$cesd)
Ncesd<-filter(Depression, relig=="None")
mean(Ncesd$cesd)
Ocesd<-filter(Depression, relig=="Other")
mean(Ocesd$cesd)
Acesd<-filter(Depression,relig=="Protestant"|relig=="Catholic"|
relig=="Jewish"|relig=="None"|relig=="Other")
mean(Acesd$cesd)
sd(Acesd$cesd)
```

```{r}
ggplot(Acesd,aes(fill=relig, x=cesd))+geom_histogram()+xlab("Depression
Scores")+ylab("Frequency")
```

These values show the average depression scores of each religion. The results show that other religions have the highest depression scores on average, while protestants have the lowest. This could suggest that other religions have significantly higher rates of depression, but the small sample size in this data set might cause this to be a misrepresentation of it. The graph shows the frequency of different ranges of depression scores.

'BIVARIATE EXPLORATION'

```{r}
ggplot(Depression,aes(x=cesd, color=relig))+geom_density(alpha=1)
+xlab("Depression Score")+ylab("Frequency")+scale_fill_brewer(palette="Set1",
name="Religeon")
```

```{r}
ggplot(Depression,aes(x=cesd, color=relig))+geom_density(alpha=1)
+facet_wrap(~relig,scales = "free_x")+xlab("Depression Score")
+ylab("Frequency")
```

These density plots show us a basic idea of depression score frequencies between the religions. Catholic and Protestants have a high amount of low depression scores. Other religions show a oddly shaped curve that shows unusually high scores, but it can likely be attributed to sample size. Non religious and Jewish people show similarly shaped plots, but the Jewish plot shows a small lean towards higher depression scores.

```{r}
```

```
ggplot(Depression,aes(y=cesd, x=relig, fill=relig))+geom_boxplot(alpha=1)
+xlab("Religion")+ylab("Depression Score")+scale_fill_brewer(palette="Set1",
guide="none")
```

```{r}
ggplot(Depression,aes(y=cesd, x=relig, fill=relig))+geom_boxplot(alpha=1)
+xlab("Religion")+ylab("Depression Score")+scale_fill_brewer(palette="Set1",
guide="none")+facet_wrap(~relig,scales = "free_x")
```

This pair of boxplots shows the frequencies of each group and their respective data points. It shows that other religions have the highest mean depression score while Protestants have the lowest. The data set also shows that other religions have the lowest amount of outlines while protestants have the highest. This may be due to their sample size, as other religions have the lowest while Protestants have the highest. Protestants also have the highest scores involving outlines, while people with no religion have the highest maximum value. Both Jewish and non religious people have similar maximums, but have mean values that suggest non religious people have more people with higher depression scores.

`CONCLUSION`

These graphs show some minor differences in the happiness of different sampled religions. My expectations didn't meet the data's results, as I expected there to be similar depression scores among the religions, while also having large differences in the four religions compared to the non religious group. But the data shows that non religious is one of the happier groups while differences between religious groups show significant differences when comparing Protestants to Jewish and other groups. The low sample size of this dataset might limit the usefulness of this data, but it is likely that at least a vague trend matching this could be produced with more surveying.