Report 1.1

What data set have you decided to use?

For this lab I decided to use the gss_cats dataset

Report 1.2

Which two variables from your data set will be analyzed?

The two variables I decided to look at are age and religious affiliation.

Report 1.3

What is your research question?

What is the relationship between age and being religious?

Report 1.4

What is your data analysis plan? Please be descriptive.

I plan to group the religions into Religious and not religious. Then I will compare the summary statistics between the two groups to see how they vary.

Report 1.5

What are some potential limitations for your analysis?

There is no scale of how religious people are through something like church attendance so it is not possible to look deeper at the religion participation levels by age.

Report 1.6

Does your data contain missing values? If so, how have you dealt with these values?

I removed the responses of Don't Know and No answer as they are not able to be grouped into religious or not religious. Additionally, there were 68 missing values that were removed before analyzing.

Report 1.7

Please include all code used to clean and manipulate the variables.

```
## Name: Robert James
## Assignment: Lab 2
## Date: 10/22/2023
## Purpose: Explore the Relationship between age and religous affiliation
getwd()
setwd("C:/Users/dell/Documents/Data202/stats-pt3")
# install packages
install.packages("tidyverse", repos = "http://cran.us.r-project.org")
library(tidyverse)
library(dplyr)
library(critstats)
library(forcats)
library(ggplot2)

df <- gss_cat #load dataset as df
```

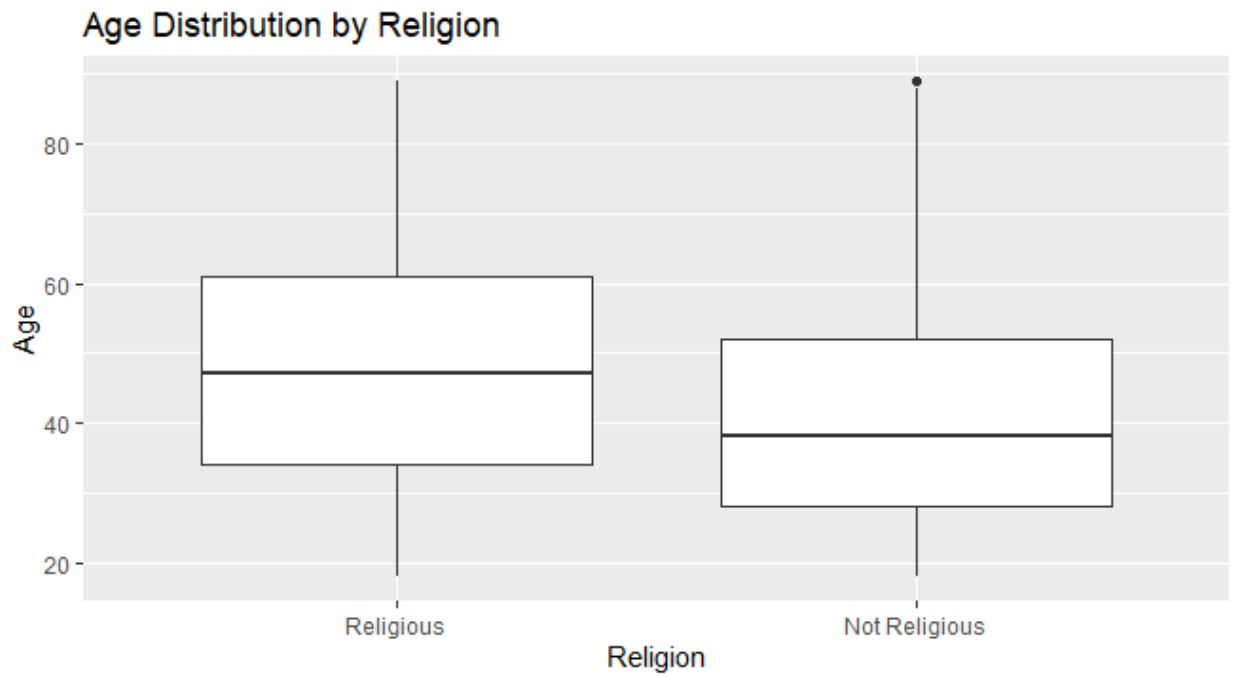
```
df <- df %>% select(age, relig) %>% rename(religion=relig) # Only select age and religion and rename relig as religion
table(df$religion) # view the number of answers in each group
df <- df %>% filter(religion != "No answer", religion != "Don't know") # remove No answer and Don't Know responses
unique(df$religion)
#Assign None to not religious and all others to religious
df <- df %>% mutate(religion = fct_recode(religion,
                          "Religious" = "Protestant",
                          "Religious" = "Native american",
                          "Religious" = "Inter-nondenominational",
                          "Religious" = "Catholic",
                          "Religious" = "Orthodox-christian",
                          "Religious" = "Christian",
                          "Religious" = "Jewish",
                          "Religious" = "Moslem/islam",
                          "Religious" = "Buddhism",
                          "Religious" = "Hinduism",
                          "Religious" = "Other eastern",
                          "Religious" = "Other",
                          "Not Religious" = "None",))
unique(df$religion)
summary(df)
# find the measures of center/variance of the ages by group
age_and_religion <- df %>% group_by(religion) %>%
  summarize(
    mean_age = mean(age, na.rm = TRUE),
    median_age = median(age, na.rm = TRUE),
    sd_age = sd(age, na.rm = TRUE)
  )
#box plot of the ages by religious group
age_and_religion
age_religion_plot <- ggplot(df, aes(x = religion, y = age)) +
  geom_boxplot() +
  labs(title = "Age Distribution by Religion", x = "Religion", y = "Age")

age_religion_plot
```

Report 1.8

What relationship, if any, exists between the two variables?

```
religion      mean_age median_age sd_age
```

	<fct>	<dbl>	<dbl>	<dbl>
1	Religious	48.4	47	17.4
2	Not Religious	41.2	38	15.6



Overall, the non-religious group is younger than the religious group. The median age of the non-religious group is 9 years younger than the religious group and the mean age is about 7 years younger. Both groups vary and all ages are present in each group but the non religious group trends slightly younger.

Report 1.9

How do these findings relate to your research question and theory?

These findings show that the relationship between age and religious affiliation is that those that identify as religious are typically older than those that identify as not religious

Report 1.10

What limitations exist as a result of the data analysis?

Outliers may skew the data towards one direction even if it does not represent the entire group accurately. Additionally, Other factors may have an influence on religious affiliation more than age such as race, income, or party affiliation which were not analyzed in this exploration.