EDA Depression Project

TreasureKing

6/27/25

Introduction

I will be using the Depression dataset, which includes information on mental health and demographics like depression scores (CESD), sex (SEX), and religion (RELIGION). My research question is: How do depression levels vary across religion and sexes?

I will explore: -cesd: continuous depression score (0 to 60) -religion: categorical variable (protestant, catholic, jewish, none, other) -sex: categorical variable (male/female) — # Univariate Exploration

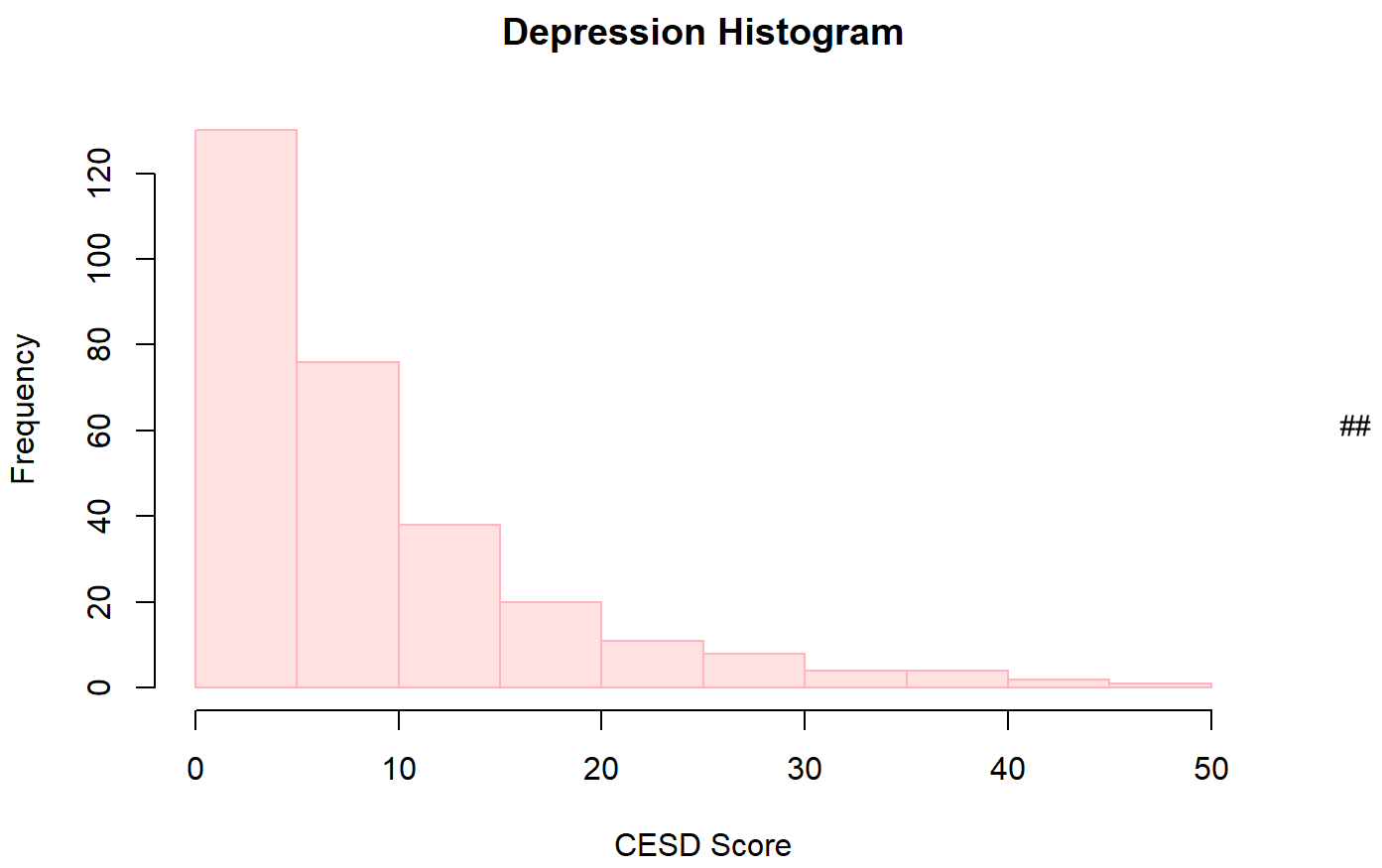
```
depression <- read.table("C:\\Users\\treas\\OneDrive\\Desktop\\MATH130\\data\\Depress.txt", head
er = TRUE, sep = "\t", quote = "")

names(depression)[names(depression) == "X.CESD."] <- "CESD"
names(depression)[names(depression) == "X.SEX."] <- "SEX"
names(depression)[names(depression) == "X.RELIG."] <- "RELIGION"

depression$RELIGION <- factor(depression$RELIGION, levels=c(1,2,3,4,5), labels=c("Protestant",
"Catholic", "Jewish", "None", "Other"))
depression$SEX <- factor(depression$SEX, levels = c(1, 2), labels = c("Male", "Female"))
```

Depression Score (CESD)

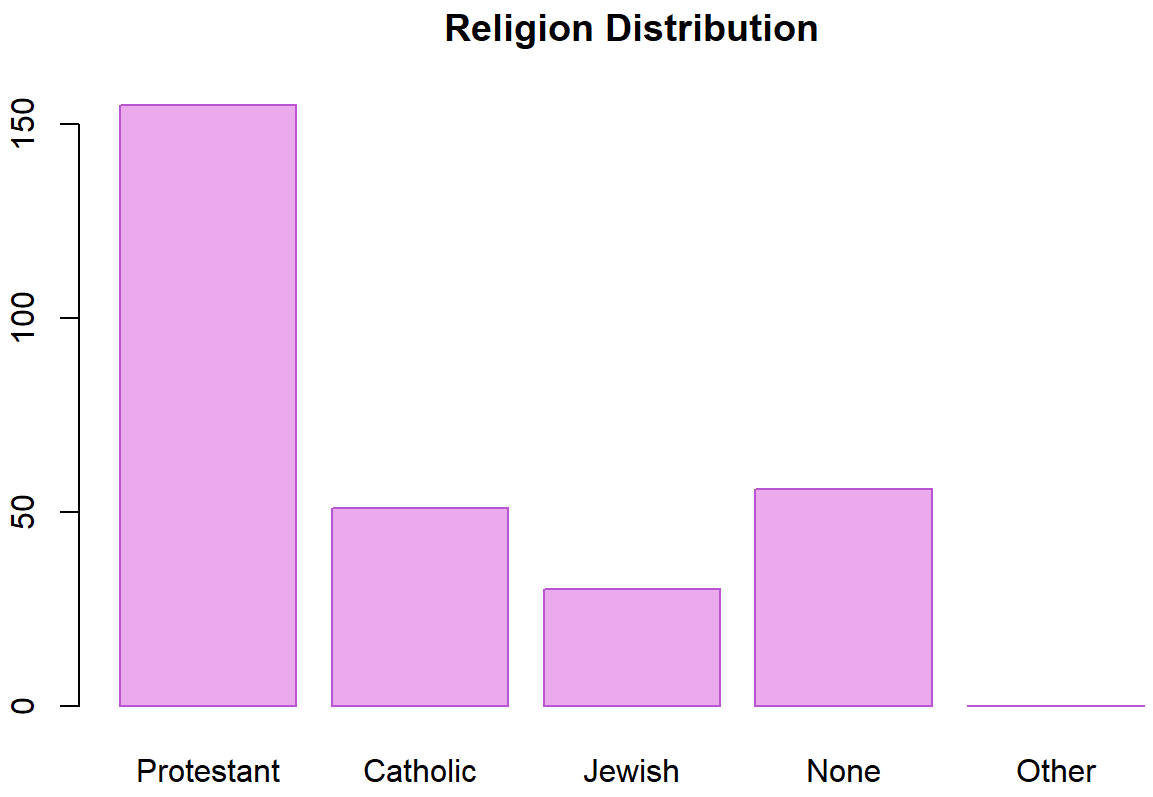
```
hist(depression$CESD,
     main = "Depression Histogram",
     col = "mistyrose",
     xlab = "CESD Score",
     border = "lightpink")
```



##

Religion

```
barplot(table(depression$RELIGION), col="plum2", main="Religion Distribution", border="mediumorc
hid")
```

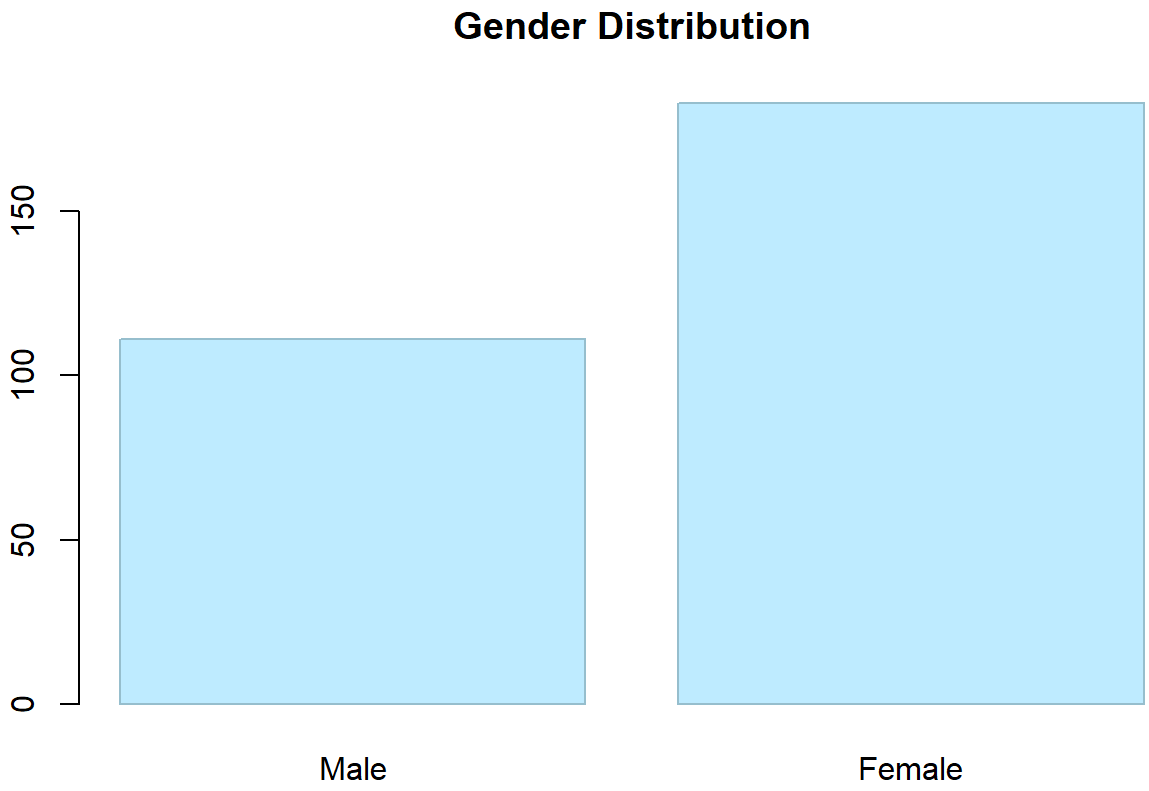


Sex

```
table(depression$SEX)
```

```
##
##   Male Female
##    111    183
```

```
barplot(table(depression$SEX), col="lightblue1", main="Gender Distribution", border="lightblue
3")
```

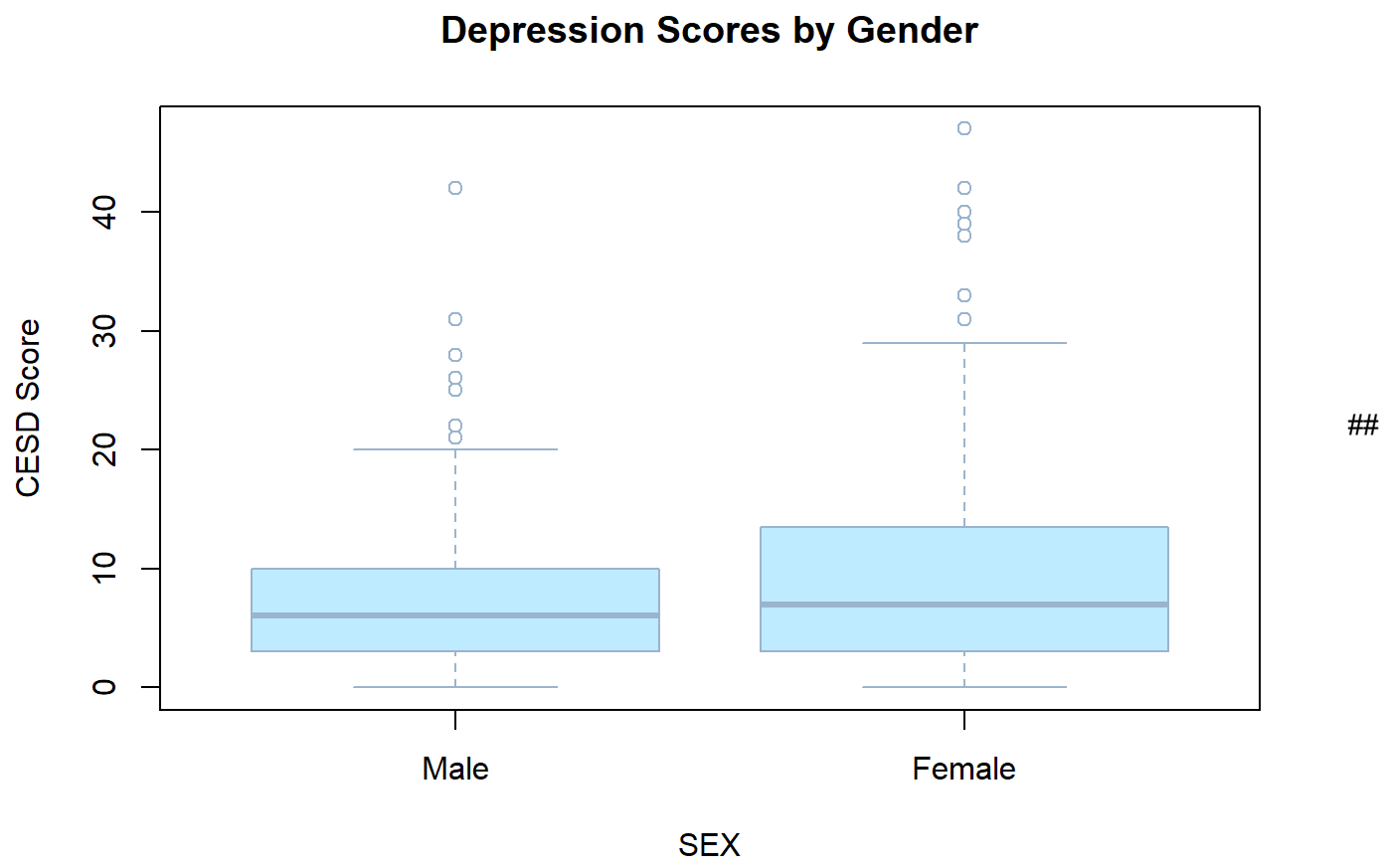


Bivariate Exploration

CESD by Gender

```
mean_male <- mean(depression$CESD[depression$SEX == "Male"], na.rm = TRUE)
mean_female <- mean(depression$CESD[depression$SEX == "Female"], na.rm = TRUE)

boxplot(CESD ~ SEX,
        data = depression,
        col = "lightblue1",
        main = "Depression Scores by Gender",
        border = "slategray3",
        ylab = "CESD Score")
```



##

CESD by Religion

```
mean(depression$CESD[depression$RELIGION == "Protestant"], na.rm=TRUE)
```

```
## [1] 7.677419
```

```
mean(depression$CESD[depression$RELIGION == "Catholic" ], na.rm=TRUE)
```

```
## [1] 8.666667
```

```
mean(depression$CESD[depression$RELIGION == "Jewish"], na.rm=TRUE)
```

```
## [1] 11.26667
```

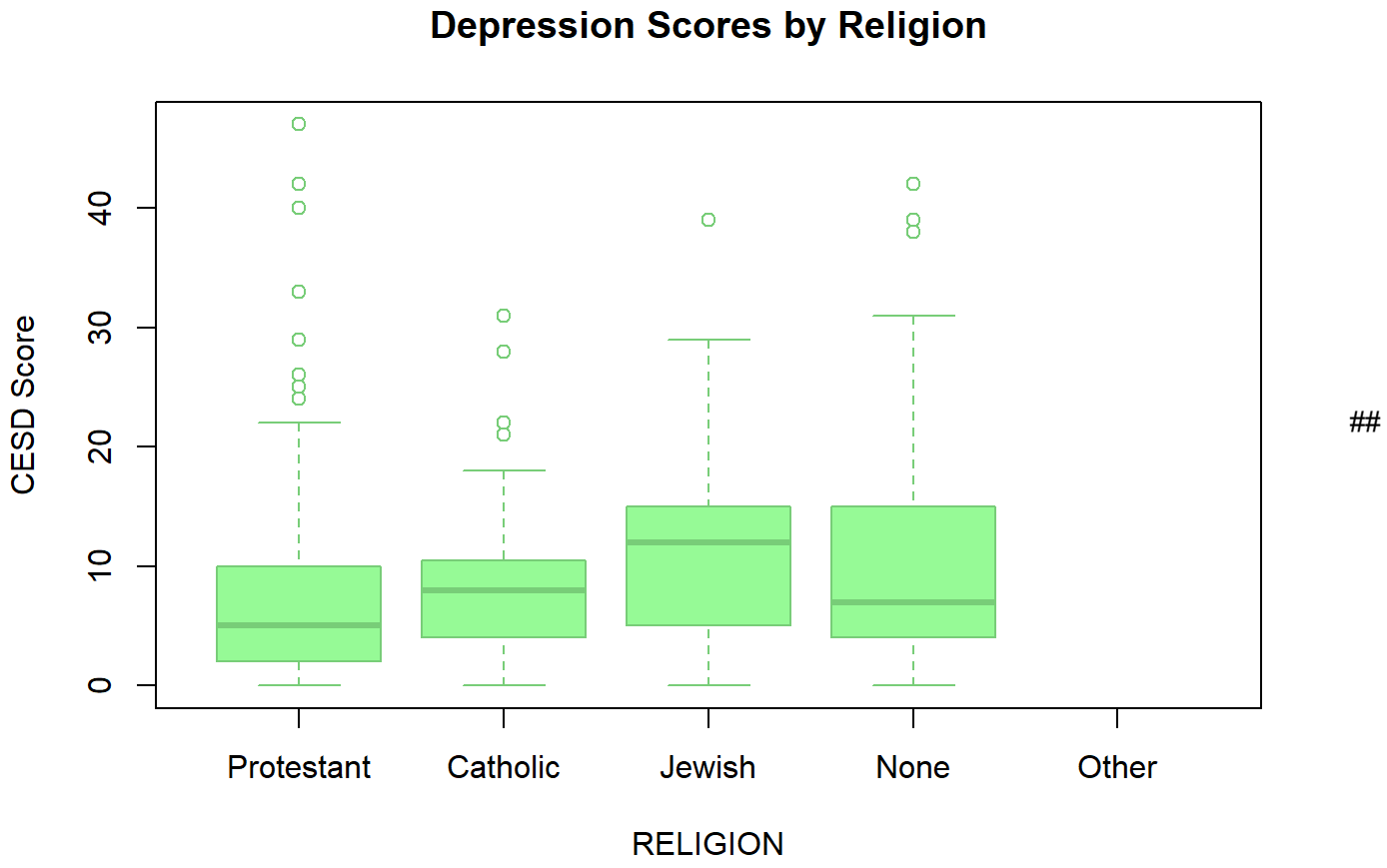
```
mean(depression$CESD[depression$RELIGION == "None"], na.rm=TRUE)
```

```
## [1] 10.875
```

```
mean(depression$CESD[depression$RELIGION == "Other"], na.rm=TRUE)
```

```
## [1] NaN
```

```
boxplot(CESD ~ RELIGION, data=depression, col="palegreen", main="Depression Scores by Religion",
border="palegreen3", ylab="CESD Score")
```

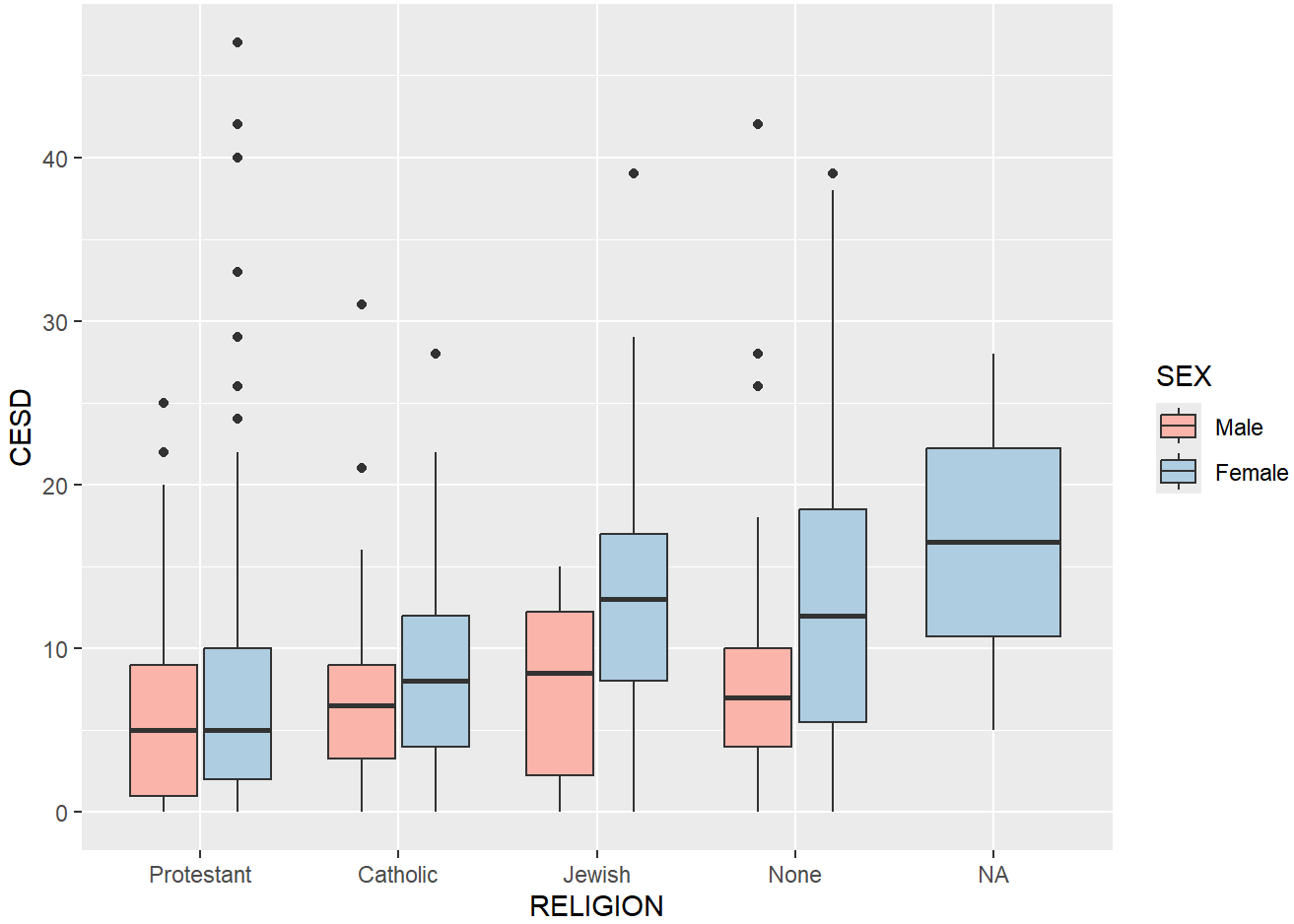


##

CESD by Gender and Religion

```
library(ggplot2)
library(RColorBrewer)

ggplot(depression, aes(x = RELIGION, y = CESD, fill = SEX)) +
  geom_boxplot() +
  scale_fill_brewer(palette="Pastel1")
```




```
  labs(title = "CESD Scores by Religion and Gender", x = "Religion", y = "CESD Score") +
  theme_minimal()
```

```
## NULL
```

Conclusion

Based on this exploratory analysis, females had higher CESD scores, on average, than males across every religion. When looking at the pattern from left to right in the religion categories (Protestant to Other), there is a general increase in depression scores. This pattern is prominent in the female category. With the male category staggers a bit, and Jewish is the highest amount while Protestant and Catholic are equal being the least.