Final Report

due November 16, 2021 by 11:59 PM

Ashley Bae, Gi Chun, Jay Nolt

November 2021

Load Packages

```
library(tidyverse)
library(readxl)
library(dbplyr)
library(ggplot2)
```

Load Data

```
load(file = "~/sta198/Jackie-Fan-Club/data/ICPSR_34363/DS0001/34363-0001-Data.rda")
#rename file
data <- da34363.0001
```

Data Citation: Battle, Juan, Pastrana, Antonio Jay, and Daniels, Jessie. Social Justice Sexuality Project: 2010 National Survey, including Puerto Rico. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2013-08-09. https://doi.org/10.3886/ICPSR34363.v1

```
data_filtered <- data %>%
  select(Q15B,
         Q15C,
         Q17A,
         Q17B,
         Q18A1:Q18A5,
         Q18C,
         Q18G,
         Q19A1:Q19A7,
         Q22A,
         Q22B,
         Q25)

# remove descriptions and other attributes
data_filtered2 <- lapply(data, function(x) {attributes(x) <- NULL; x}) %>%
  as.data.frame() %>%
  select(Q15B,
         Q15C,
         Q17A,
         Q17B,
         Q18A1:Q18A5,
         Q18C,
```

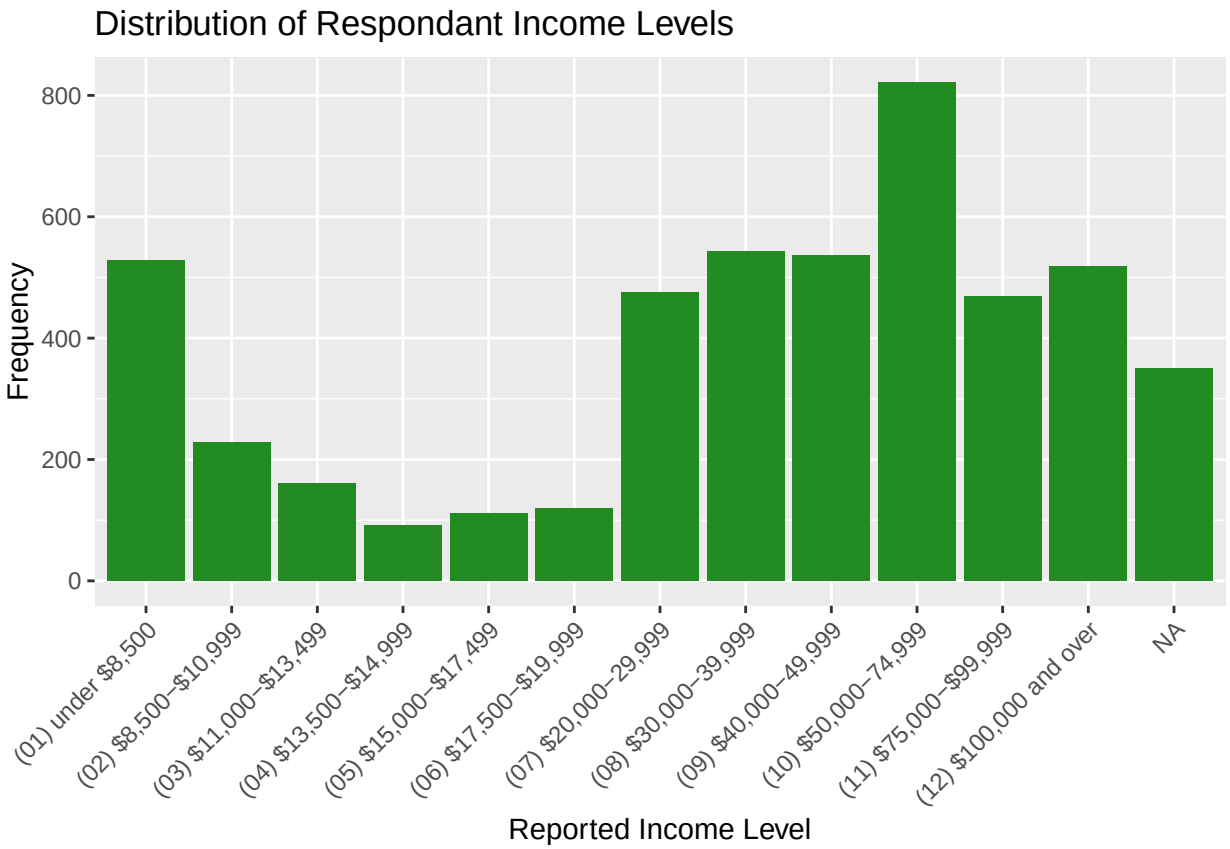
```
          Q18G,
          Q19A1:Q19A7,
          Q22A,
          Q22B,
          Q25)
```

```
data <- data_filtered %>%
    select(Q17A,
          Q17B,
          Q18A1:Q18A5,
          Q18C,
          Q18G,
          Q19A1:Q19A7,
          Q22A,
          Q22B,
          Q25) %>%
  rename(healthInsureAcc = Q17A) %>%
  rename(healthProvideAcc = Q17B) %>%
  rename(male = Q18A1) %>%
  rename(female = Q18A2) %>%
  rename(m2f = Q18A3) %>%
  rename(f2m = Q18A4) %>%
  rename(genderOther = Q18A5) %>%
  rename(sexuality = Q18C) %>%
  mutate(age = 2021 - Q18G) %>%
  rename(black = Q19A1) %>%
  rename(hispanic = Q19A2) %>%
  rename(asian = Q19A3) %>%
  rename(native = Q19A4) %>%
  rename(white = Q19A5) %>%
  rename(multi = Q19A6) %>%
  rename(raceOther = Q19A7) %>%
  rename(edu = Q22A) %>%
  rename(income = Q22B) %>%
  rename(assessHealth = Q25)
```

```
data <- data %>%
  mutate(male = case_when(
  male == "(1) Yes" ~ "male",
  is.na(male) ~ "",
  TRUE ~ ""),
  female = case_when(
  female == "(1) Yes" ~ "female",
  is.na(female) ~ "",
  TRUE ~ ""),
  m2f = case_when(
  m2f == "(1) Yes" ~ "m2f",
  is.na(m2f) ~ "",
  TRUE ~ ""),
  f2m = case_when(
  f2m == "(1) Yes" ~ "f2m",
  is.na(f2m) ~ "",
  TRUE ~ ""),
  genderOther = case_when(
```

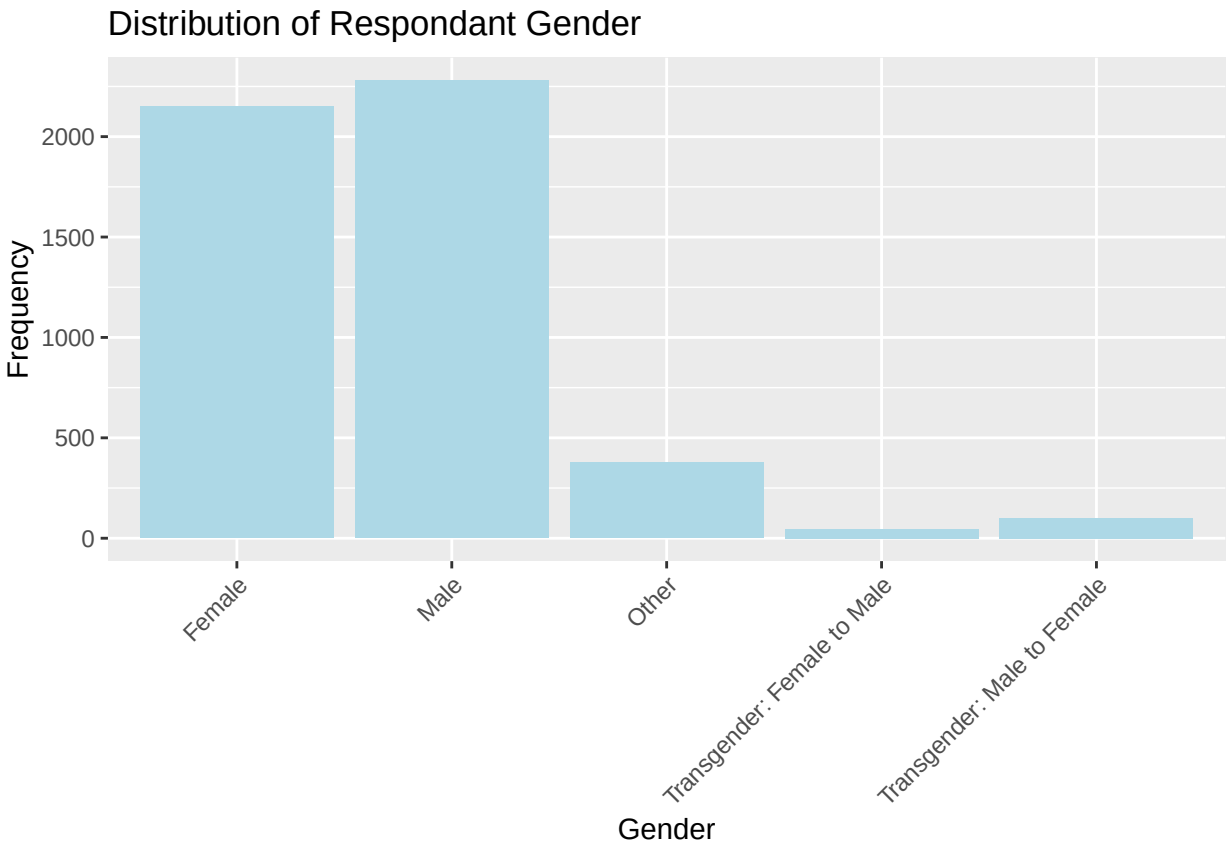
```
  genderOther == "(1) Yes" ~ "Other",
  is.na(genderOther) ~ "",
  TRUE ~ ""),
  gender = ""
)

data$gender <- paste(data$male, data$female)
data$gender <- paste(data$gender, data$m2f)
data$gender <- paste(data$gender, data$f2m)
data$gender <- paste(data$gender, data$genderOther)

data$gender <- trimws(data$gender)

data <- data %>%
mutate(gender = case_when(
  gender == "male" ~ "Male",
  gender == "female" ~ "Female",
  gender == "m2f" ~ "Transgender: Male to Female",
  gender == "f2m" ~ "Transgender: Female to Male",
  TRUE ~ "Other")
)
```

```
ggplot(data, aes(x = income)) +
  geom_bar(fill = "#228b22") +
  labs (x = "Reported Income Level",
        y = "Frequency",
        title = "Distribution of Respondant Income Levels") +
  theme(axis.text.x = element_text(angle = 45,
                                   hjust = 1))
```

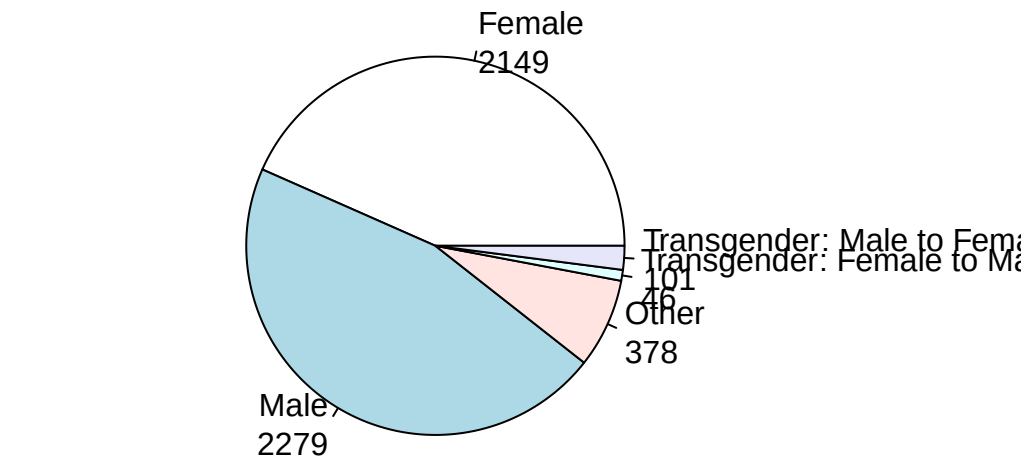


We visualized income distribution of respondents in this study. The mode income bracket, with the greatest number of people, is \$50,000-\$74,999. The distribution has a left skew with the majority of the data in the upper ranges \$20,000 and over. The bracket with the lowest count was \$13,500-\$14,999. Of the respondents, there were 4953 with income data, and there were 351 people for whom we didn't have income data.

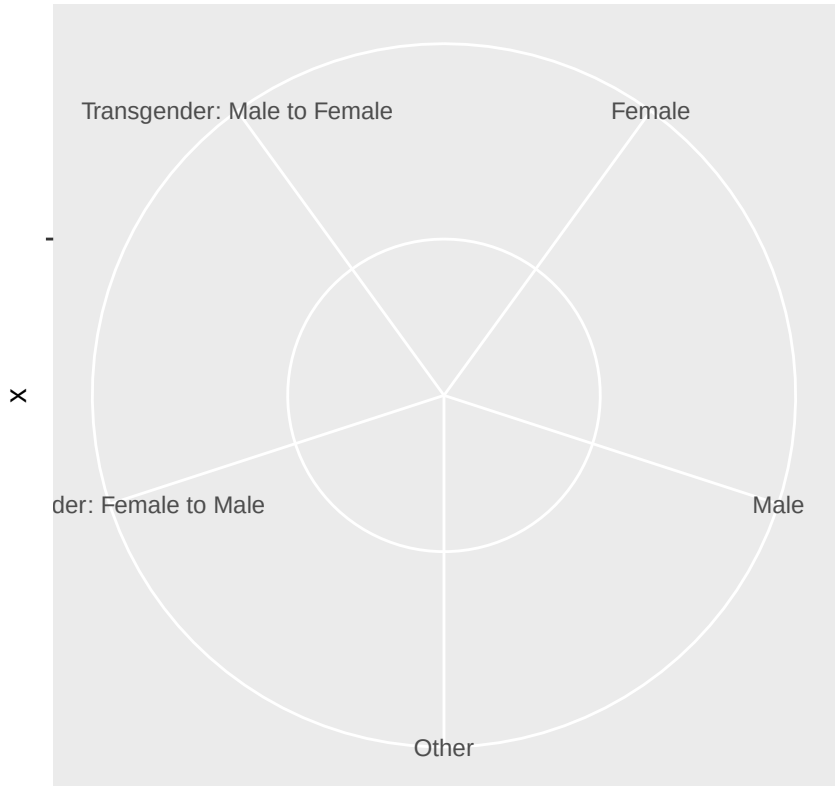
```
ggplot(data, aes(x = gender)) +
  geom_bar(fill = "light blue") +
  labs (x = "Gender",
        y = "Frequency",
        title = "Distribution of Respondant Gender") +
  theme(axis.text.x = element_text(angle = 45,
                                   hjust = 1))
```




```
# Pie Chart with Percentages
gendertable <- table(data$gender)
lbls <- paste(names(gendertable), "\n", gendertable, sep="")
pie(gendertable, labels = lbls,
   main="Pie Chart of Gender Distribution of Respondents")
```



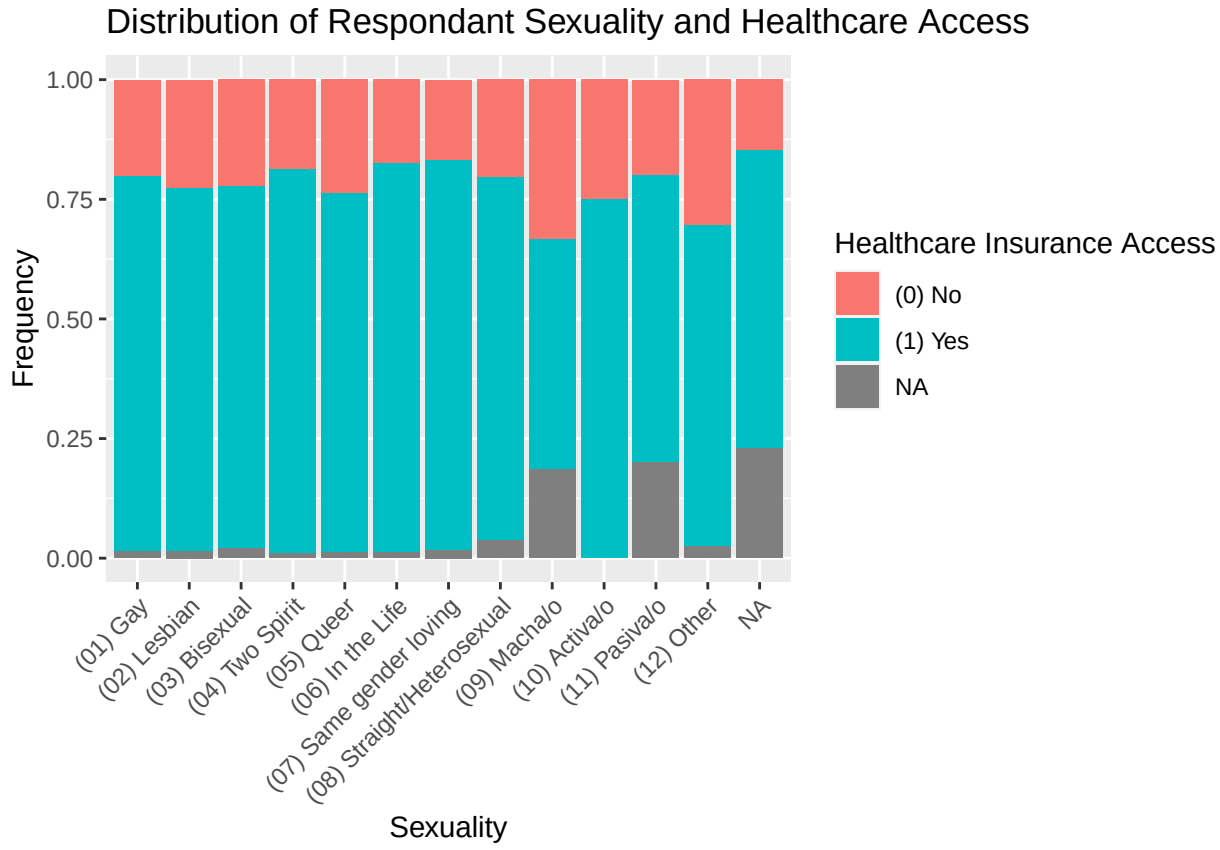

```
# attempt at ggplot2 way...
ggplot(data, aes(x = "", y = gender, fill = "gender")) +
  coord_polar(theta = "y")
```



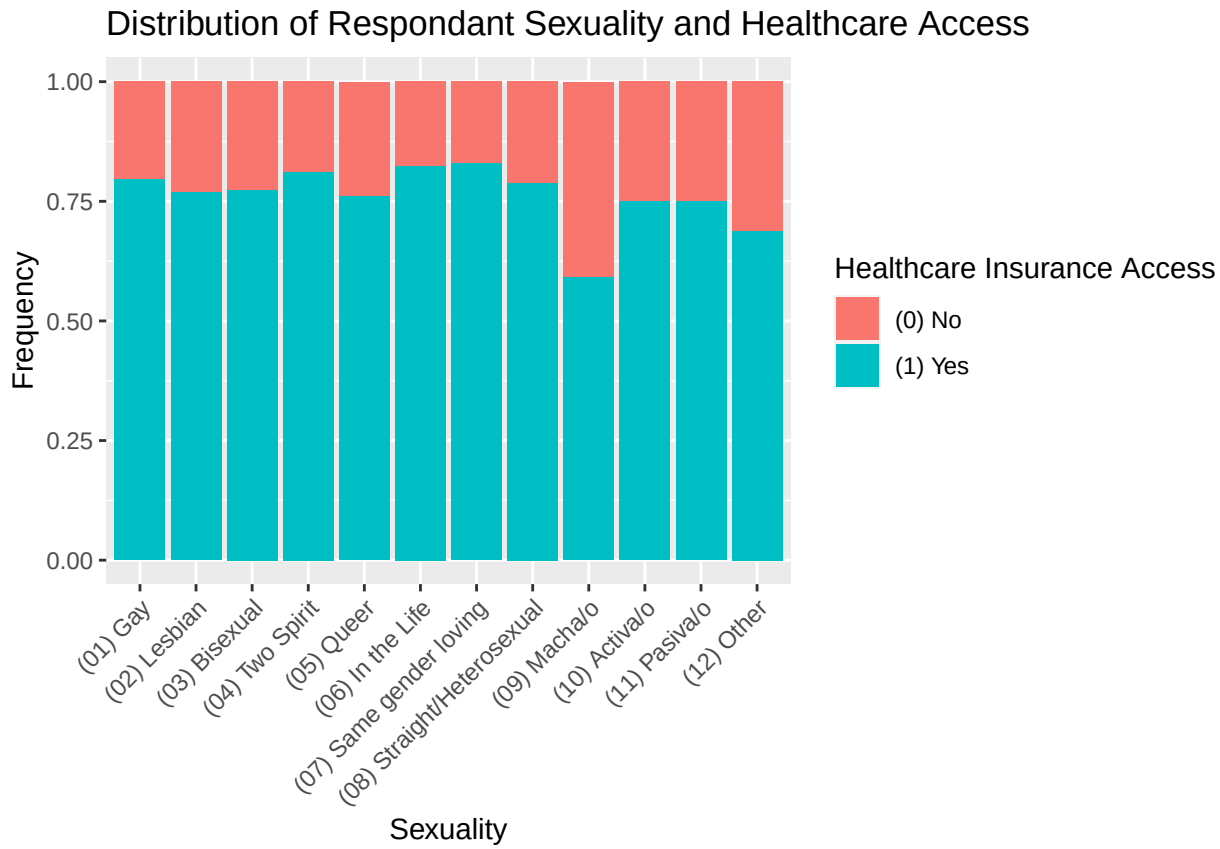

```
  #geom_label_repel(nudge_y = 1, direction = "y", hjust = 1) #for if label needs adjusting?
```

We visualized gender identities of participants with a pie chart. The biggest group of respondents are male, followed by female, followed by "other." There are more M2F transgender individuals than F2M transgender individuals captured in this study.

```
ggplot(data, aes(x = sexuality,
                 fill = healthInsureAcc)) +
  geom_bar(position = "fill") +
  labs (x = "Sexuality",
        y = "Frequency",
        fill = "Healthcare Insurance Access",
        title = "Distribution of Respondant Sexuality and Healthcare Access") +
  theme(axis.text.x = element_text(angle = 45,
                                   hjust = 1))
```



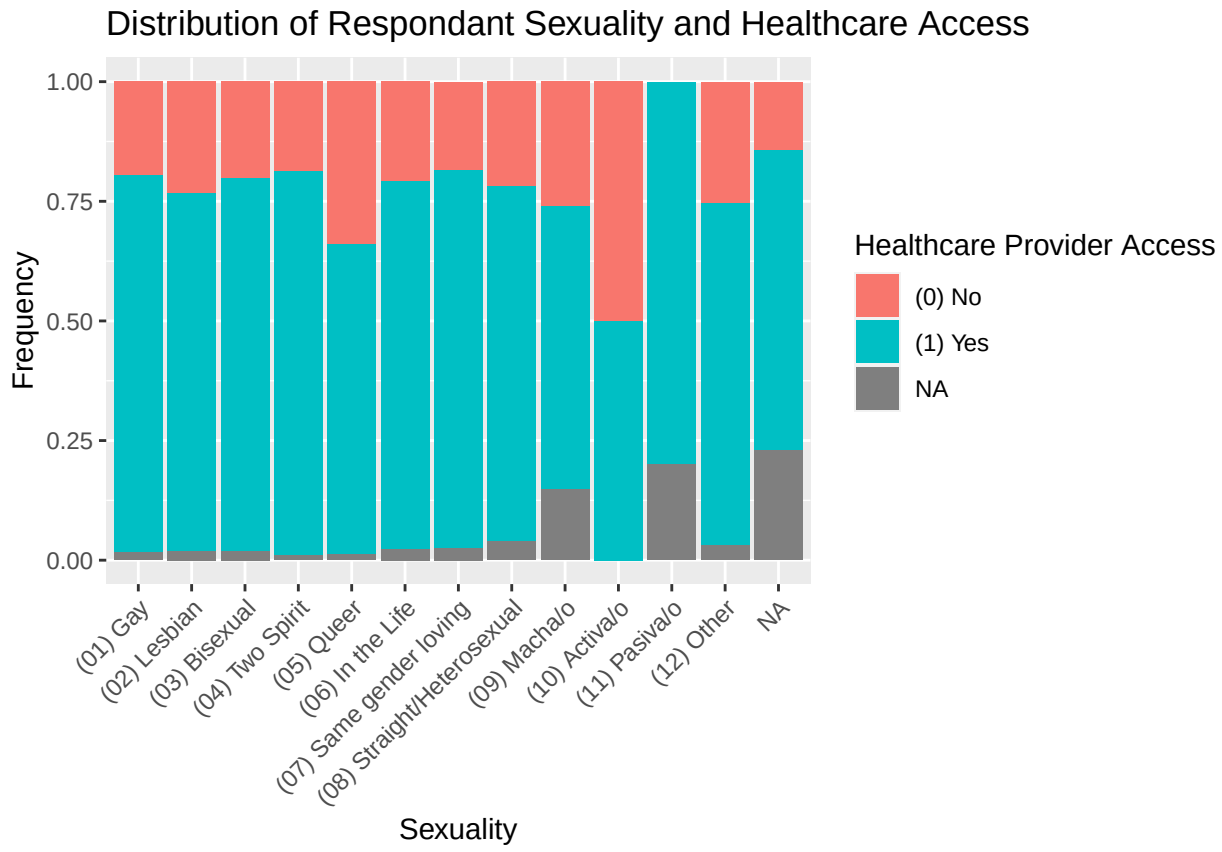

```
data %>%
  filter(!is.na(healthInsureAcc),
         !is.na(sexuality)
         ) %>%
  ggplot(aes(x = sexuality,
             fill = healthInsureAcc)) +
  geom_bar(position = "fill") +
  labs (x = "Sexuality",
        y = "Frequency",
        fill = "Healthcare Insurance Access",
        title = "Distribution of Respondant Sexuality and Healthcare Access") +
  theme(axis.text.x = element_text(angle = 45,
                                   hjust = 1))
```



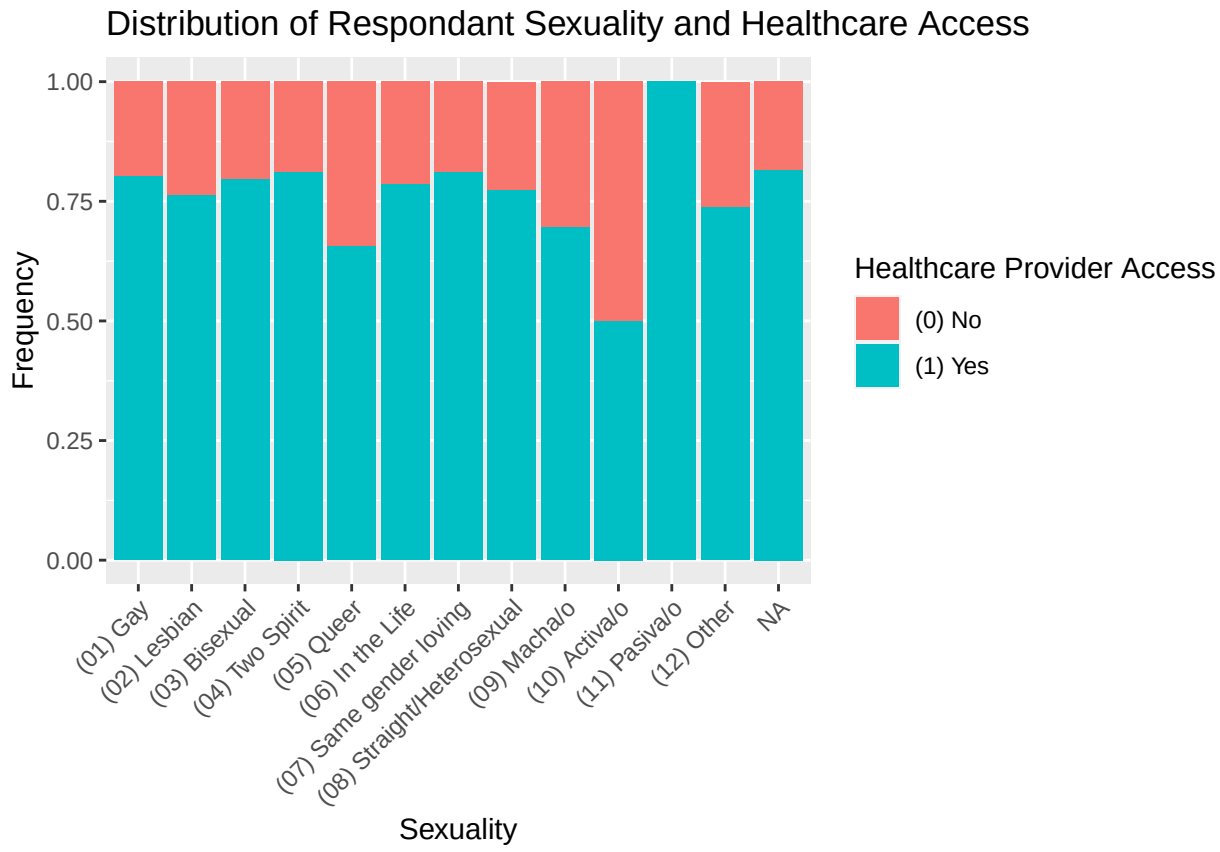
We used a segmented bar graph to visualize sexuality of respondents and their access to healthcare. The differences are not so visually distinct, especially when attempting to compare straight/heterosexual individuals to other historically and presently marginalized sexual identities. This is especially so in groups with many NA responses. In addition, the multitude of identities of sexuality which also may overlap or have culturally distinct attributes may complicate data analysis. The respondants are also prompted to choose only one identity, to the question "Which one label comes closest to how you describe your sexual identity?"

In the second visualization of sexuality and access to healthcare, we visualized the with NA values removed. This helps us more clearly visualize the data. We can see more clearly that the group with least health insurance access is those identifying as macha/o.

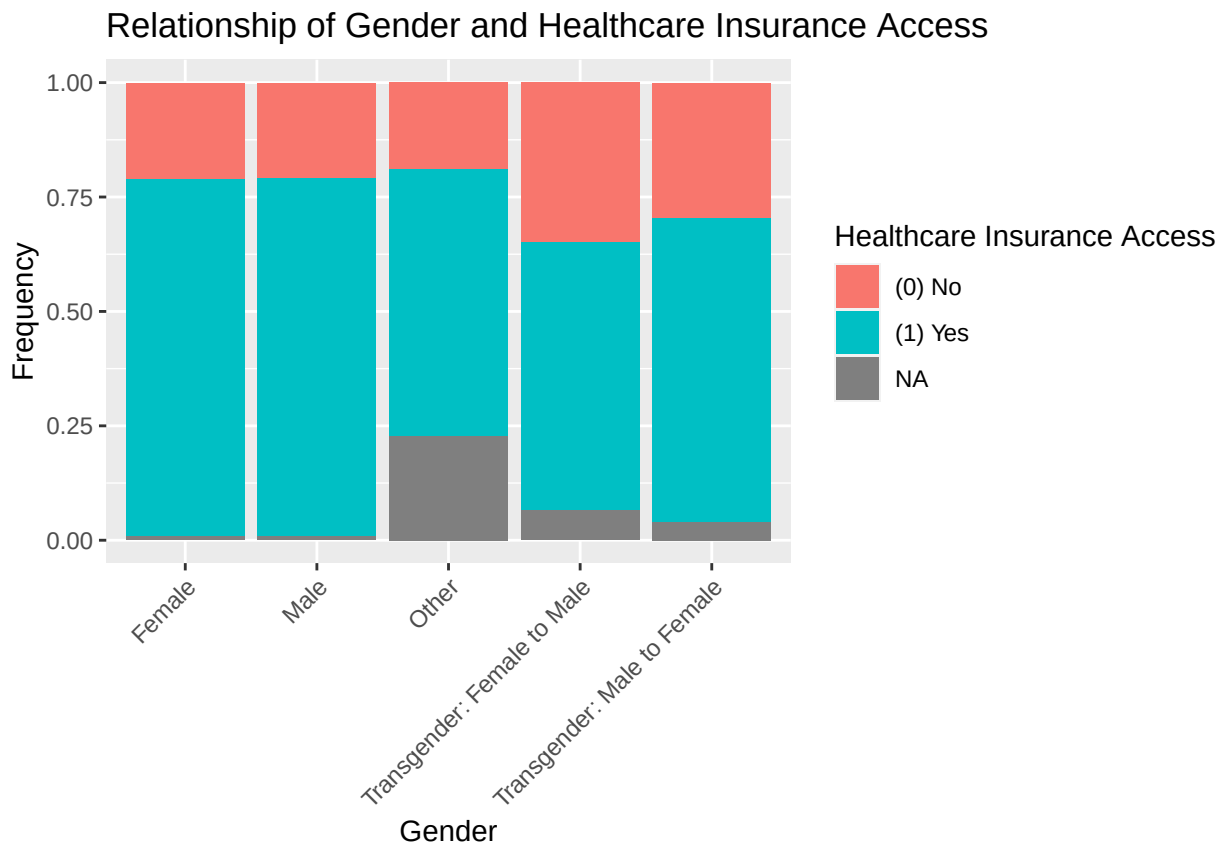
```
data %>%
ggplot(aes(x = sexuality,
                        fill = healthProvideAcc)) +
  geom_bar(position = "fill") +
  labs (x = "Sexuality",
        y = "Frequency",
        fill = "Healthcare Provider Access",
        title = "Distribution of Respondant Sexuality and Healthcare Access") +
  theme(axis.text.x = element_text(angle = 45,
                                   hjust = 1))
```



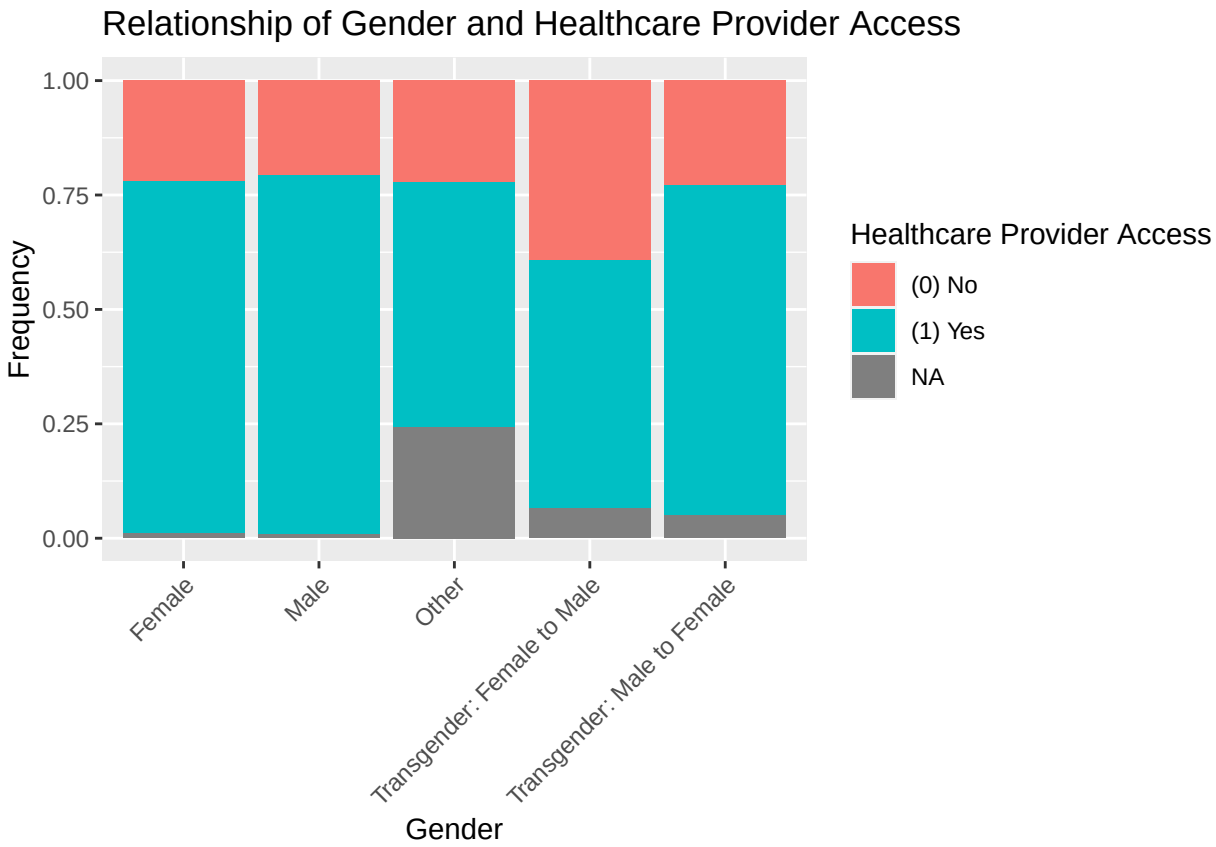

```
data %>%
  filter(is.na(healthProvideAcc) == FALSE) %>%
ggplot(aes(x = sexuality,
                        fill = healthProvideAcc)) +
  geom_bar(position = "fill") +
  labs (x = "Sexuality",
        y = "Frequency",
        fill = "Healthcare Provider Access",
        title = "Distribution of Respondant Sexuality and Healthcare Access") +
  theme(axis.text.x = element_text(angle = 45,
                                   hjust = 1))
```




```
data %>%
ggplot(aes(x = gender,
           fill = healthInsureAcc)) +
  geom_bar(position = "fill") +
  labs (x = "Gender",
        y = "Frequency",
        fill = "Healthcare Insurance Access",
        title = "Relationship of Gender and Healthcare Insurance Access") +
  theme(axis.text.x = element_text(angle = 45,
                                   hjust = 1))
```




```
ggplot(data, aes(x = gender,
                 fill = healthProvideAcc)) +
  geom_bar(position = "fill") +
  labs (x = "Gender",
        y = "Frequency",
        fill = "Healthcare Provider Access",
        title = "Relationship of Gender and Healthcare Provider Access") +
  theme(axis.text.x = element_text(angle = 45,
                                   hjust = 1))
```



In the above segmented bar graphs, respondents respond to "Do you have health insurance?" and "Do you have a regular doctor or health care provider?", respectively. The visualizations show that a greater percentage of transgender individuals answered "no" than cisgender individuals. The greatest percentage of "no" responses came from the group of F2M tansgender individuals. Also, when comparing the gender groups, note that there are more missing values for transgendered individuals than cisgendered respondents for both access to insurance and regular healthcare provider, which may provide more insight to the disparity. It appears that one's gender identity may have a relationship to access to healthcare.