

Final Report

due November 16, 2021 by 11:59 PM

Jackie Fan Club: Ashley Bae, Gi Chun, Jay Nolt

November 2021

Load Packages

```
library(tidyverse)
library(tidymodels)
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 = ""
)

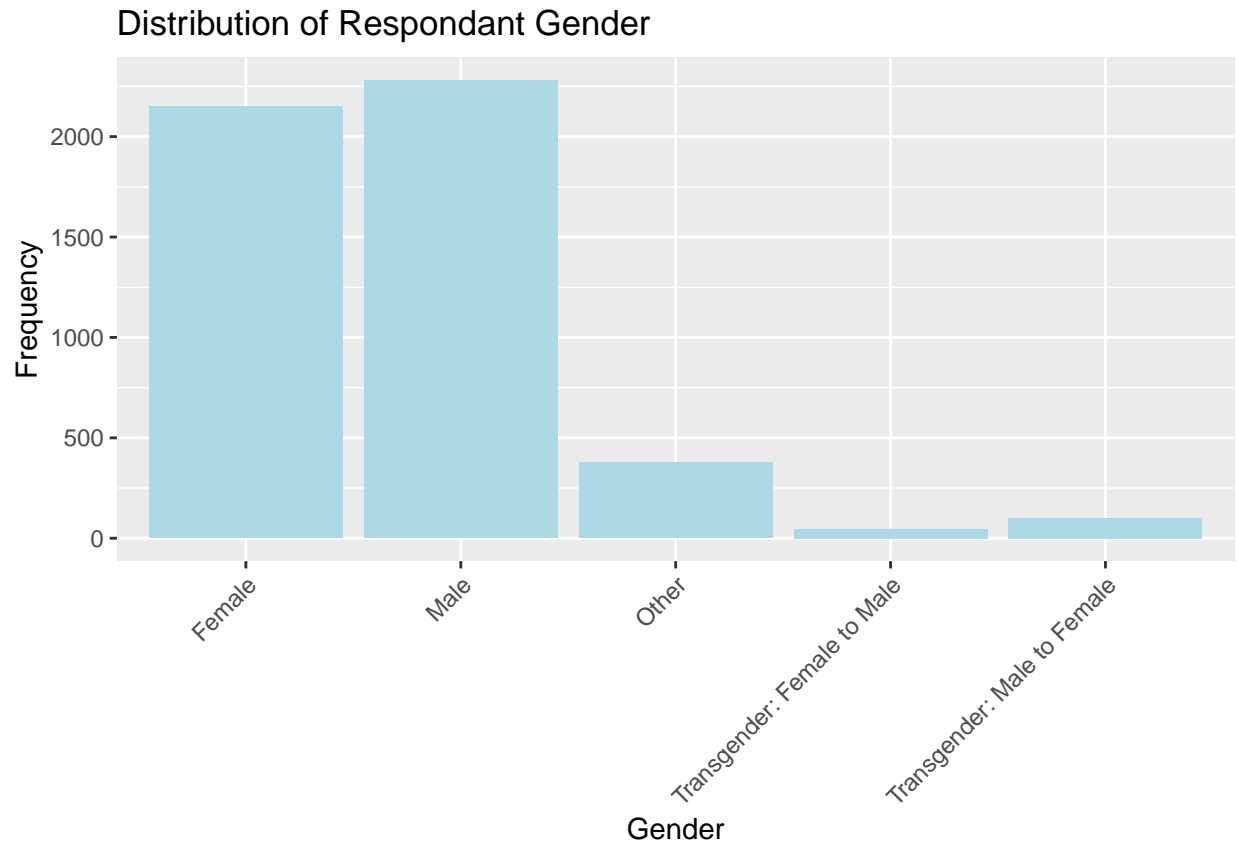
data$gender <- paste(data$male, data$female)
data$gender <- paste(data$gender, data$m2f)
data$gender <- paste(data$gender, data$f2m)
data$gender <- paste(data$gender, data$genderOther)

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")
)

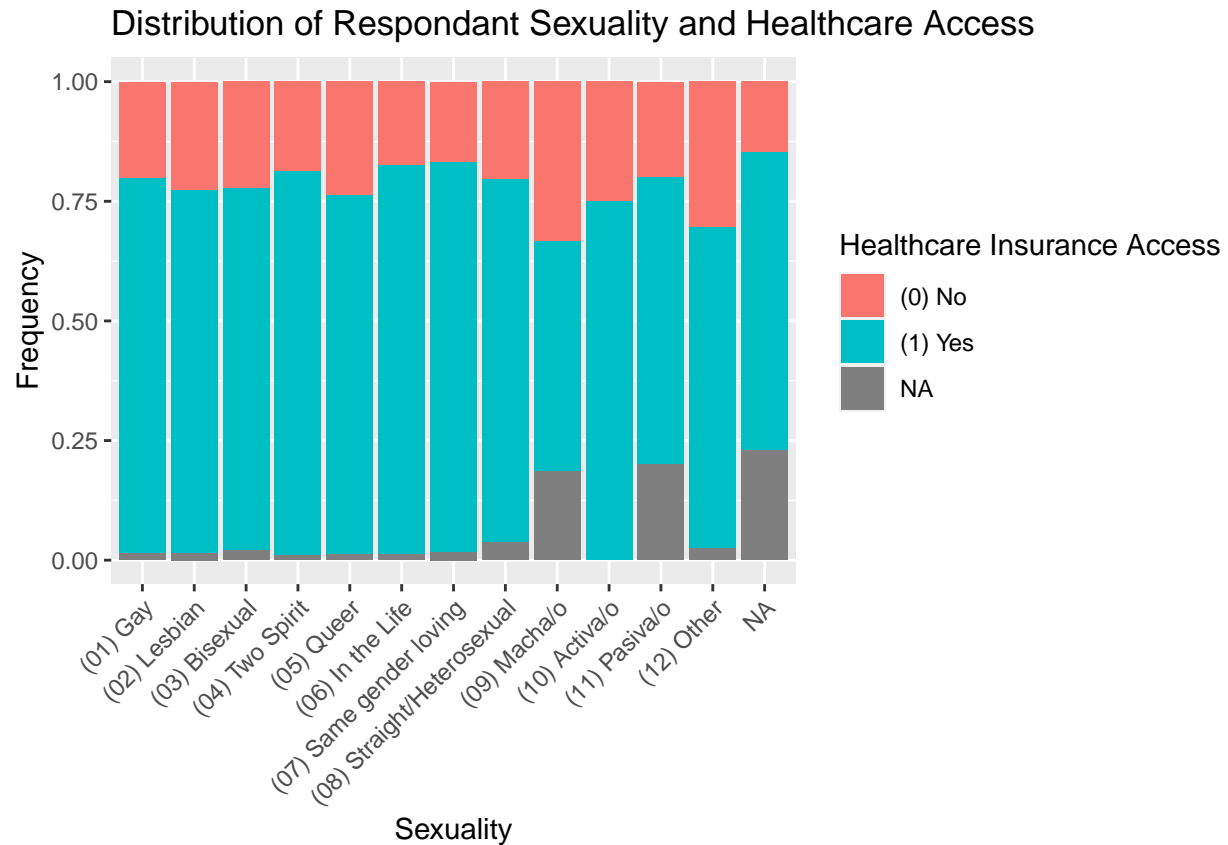
ggplot(data, aes(x = gender)) +
  geom_bar(fill = "light blue") +
  labs (x = "Gender",
        y = "Frequency",
        title = "Distribution of Respondant Gender") +
  theme(axis.text.x = element_text(angle = 45,
                                     hjust = 1))

```

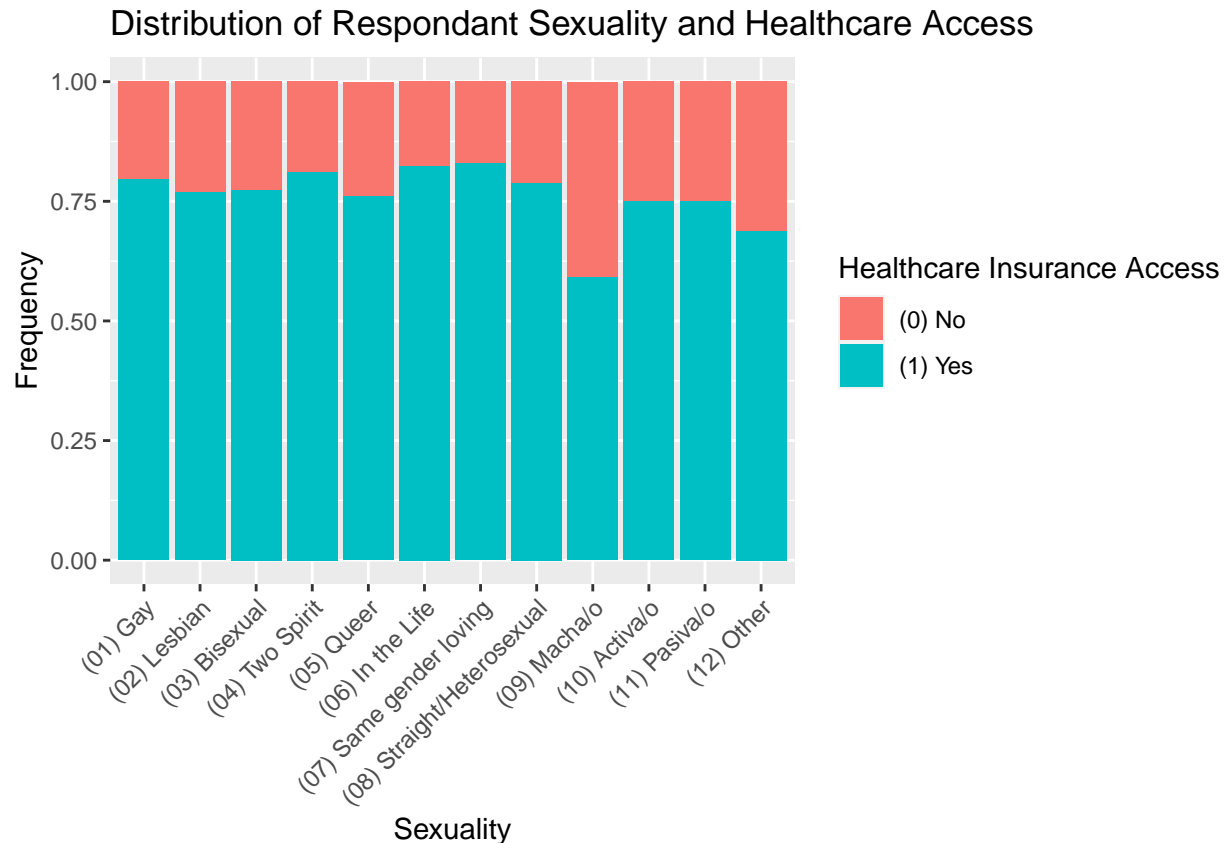


We visualized gender identities of participants with a pie chart. The biggest group of respondents are male, followed by female, followed by “other.” There are more M2F transgender individuals than F2M transgender individuals captured in this study.

```
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))
```



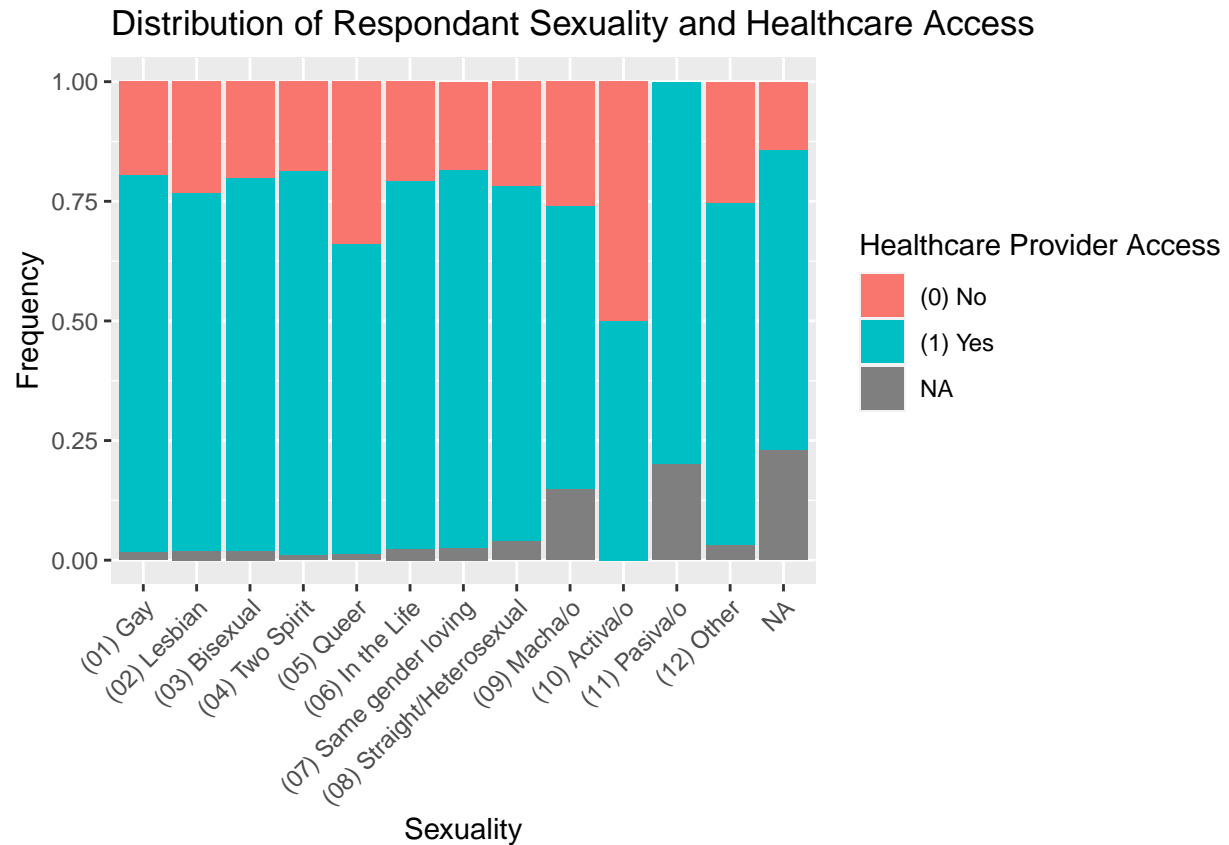
```
data %>%
  filter(!is.na(healthInsureAcc),
         !is.na(sexuality)
  ) %>%
  ggplot(aes(x = sexuality,
             fill = healthInsureAcc)) +
  geom_bar(position = "fill") +
  labs (x = "Sexuality",
        y = "Frequency",
        fill = "Healthcare Insurance Access",
        title = "Distribution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



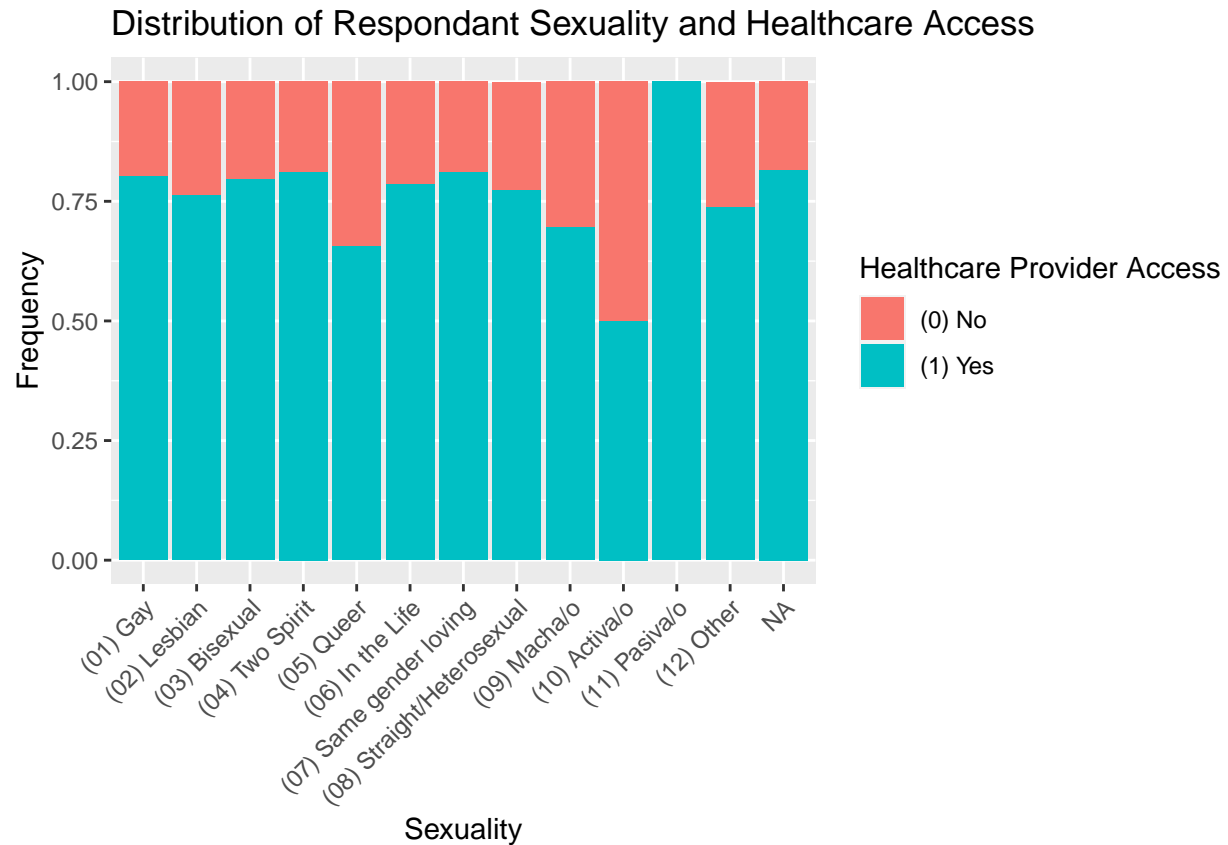
We used a segmented bar graph to visualize sexuality of respondents and their access to healthcare. The differences are not so visually distinct, especially when attempting to compare straight/heterosexual individuals to other historically and presently marginalized sexual identities. This is especially 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 respondents 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 included 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))
```



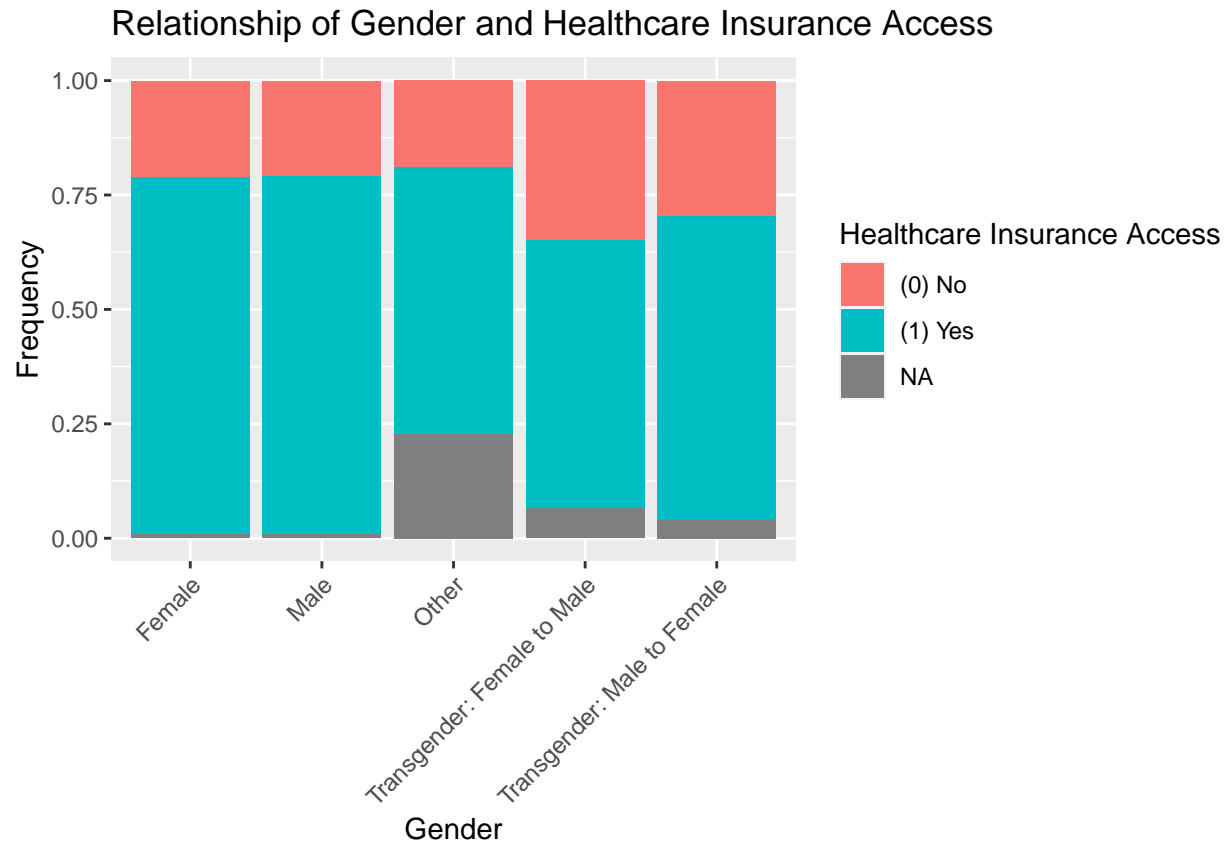
```
data %>%
  filter(is.na(healthProvideAcc) == FALSE) %>%
  ggplot(aes(x = sexuality,
             fill = healthProvideAcc)) +
  geom_bar(position = "fill") +
  labs (x = "Sexuality",
        y = "Frequency",
        fill = "Healthcare Provider Access",
        title = "Distribution of Respondant Sexuality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



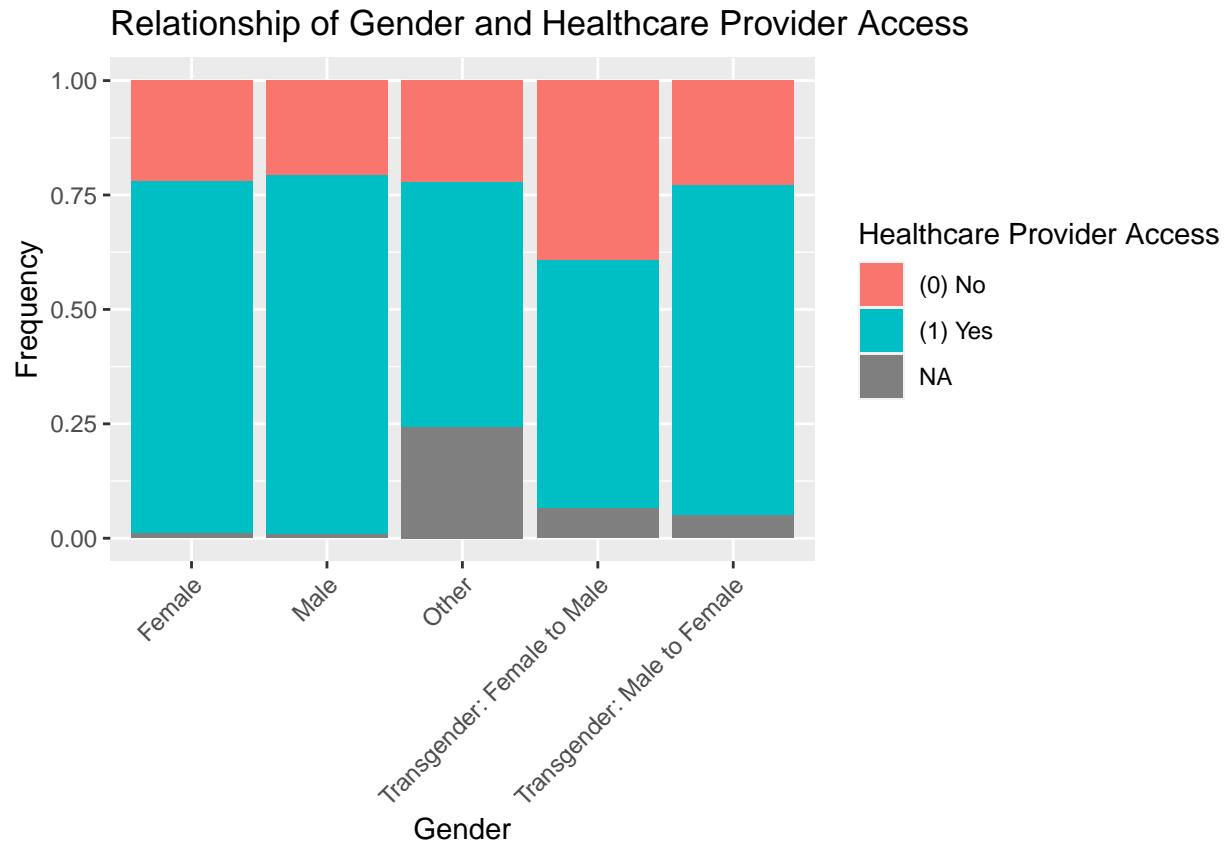
```
table(data$sexuality, data$healthProvideAcc)
```

```
##
##              (0) No (1) Yes
## (01) Gay          347   1399
## (02) Lesbian      258    829
## (03) Bisexual     106    411
## (04) Two Spirit    20     86
## (05) Queer        109    208
## (06) In the Life   18     66
## (07) Same gender loving 44   187
## (08) Straight/Heterosexual 90  306
## (09) Macha/o        7    16
## (10) Activa/o        2     2
## (11) Pasiva/o        0     4
## (12) Other         41   115
```

```
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 transgender 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.

Data Analysis

To evaluate impacts of intersectional identities on health outcomes, we ran a few statistical tests.

Sexuality: strait/heterosexual vs queer (all other sexualities.. hard to distinguish) Gender: cisgender vs transgender Race: keep each group

```
data_fit <- data %>%
  mutate(
    SEXUALITY = if_else(sexuality == "(08) Straight/Heterosexual", 0, 1),
    GENDER = if_else((gender == "Male") | (gender == "Female"), 0, 1),
    BLACK = if_else(black == "(1) Yes", 1, 0),
    HISPANIC = if_else(hispanic == "(1) Yes", 1, 0),
    ASIAN = if_else(asian == "(1) Yes", 1, 0),
    NATIVE = if_else(native == "(1) Yes", 1, 0),
    MULTIRACE = if_else(multi == "(1) Yes", 1, 0),
    RACEOTHER = if_else(raceOther == "(1) Yes", 1, 0)
  )
```

```
# insurance_fit_multi <- logistic_reg() %>%
#   set_engine("glm") %>%
#   fit(
#     healthInsureAcc ~ SEXUALITY + GENDER + BLACK + HISPANIC + ASIAN + NATIVE + MULTIRACE + RACEOTHER,
#     data = data_fit,
#     family = "binomial"
#   )
# tidy(insurance_fit_multi,
#       conf.int = TRUE,
#       exponentiate = TRUE)
```

```
insurance_fit_multi_select <- logistic_reg() %>%
  set_engine("glm") %>%
  fit(
    healthInsureAcc ~ GENDER + BLACK + HISPANIC + ASIAN + NATIVE,
    data = data_fit,
    family = "binomial"
  )
tidy(insurance_fit_multi_select,
     conf.int = TRUE,
     exponentiate = TRUE)
```

```
## # A tibble: 6 x 7
##   term          estimate std.error statistic  p.value conf.low conf.high
##   <chr>          <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept)    3.87      0.0824    16.4  1.43e-60    3.30    4.55
## 2 GENDER         0.854     0.164    -0.963 3.36e- 1    0.623    1.19
## 3 BLACK          1.13     0.109     1.11 2.69e- 1    0.912    1.40
## 4 HISPANIC       0.644     0.112    -3.93 8.60e- 5    0.518    0.803
## 5 ASIAN          1.23     0.170     1.23 2.19e- 1    0.890    1.73
## 6 NATIVE         0.624     0.188    -2.51 1.21e- 2    0.434    0.909
```

```
insurance_fit_multi_interact <- logistic_reg() %>%
  set_engine("glm") %>%
  fit(
    healthInsureAcc ~ GENDER*BLACK + GENDER*HISPANIC + GENDER*ASIAN + GENDER*NATIVE,
    data = data_fit,
    family = "binomial"
  )
tidy(insurance_fit_multi_interact,
     conf.int = TRUE,
     exponentiate = TRUE)
```

```
## # A tibble: 10 x 7
##   term          estimate std.error statistic  p.value conf.low conf.high
##   <chr>          <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept)    3.90      0.0863    15.8  3.94e-56    3.30    4.63
## 2 GENDER         0.804     0.253    -0.866 3.86e- 1    0.495    1.34
## 3 BLACK          1.13     0.116     1.04 3.00e- 1    0.899    1.42
## 4 HISPANIC       0.631     0.118    -3.90 9.62e- 5    0.501    0.796
## 5 ASIAN          1.21     0.180     1.05 2.94e- 1    0.856    1.73
## 6 NATIVE         0.619     0.205    -2.34 1.94e- 2    0.417    0.934
## 7 GENDER:BLACK    0.969     0.343    -0.0908 9.28e- 1    0.496    1.91
## 8 GENDER:HISPANIC 1.23      0.383     0.545 5.86e- 1    0.588    2.65
```

```
## 9 GENDER:ASIAN      1.16      0.565      0.261 7.94e- 1      0.405      3.86
## 10 GENDER:NATIVE    1.03      0.514      0.0529 9.58e- 1      0.382      2.92
```

We removed variables that are statistically insignificant when evaluating p values, which were sexuality (whether someone identified as straight/heterosexual or other sexualities), being multiracial, and being of “other” racial categories than options given on the survey.

We conducted another multiple regression with interaction terms, in order to evaluate intersection of gender and race groups.

```
observed_chisq_statistic <- data %>%
  specify(healthInsureAcc ~ sexuality) %>%
  calculate(stat = "Chisq")
```

```
## Warning: Removed 270 rows containing missing values.
```

```
#fisher.test(data$healthInsureAcc,data$sexuality, simulate.p.value = TRUE)
```