

Final Report

due November 16, 2021 by 11:59 PM

Ashley Bae, Gi Chun, Jay Nolt

November 2021

Load Packages

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

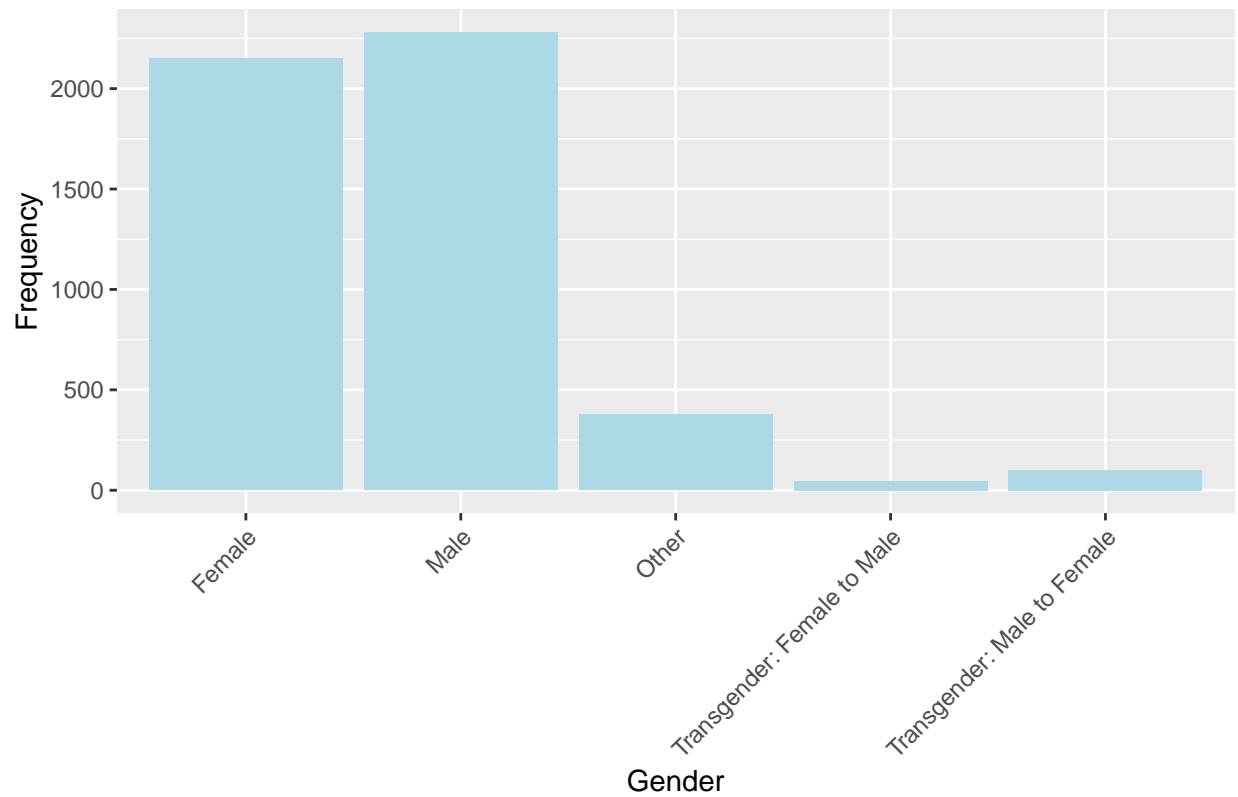
Load Data

```
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>

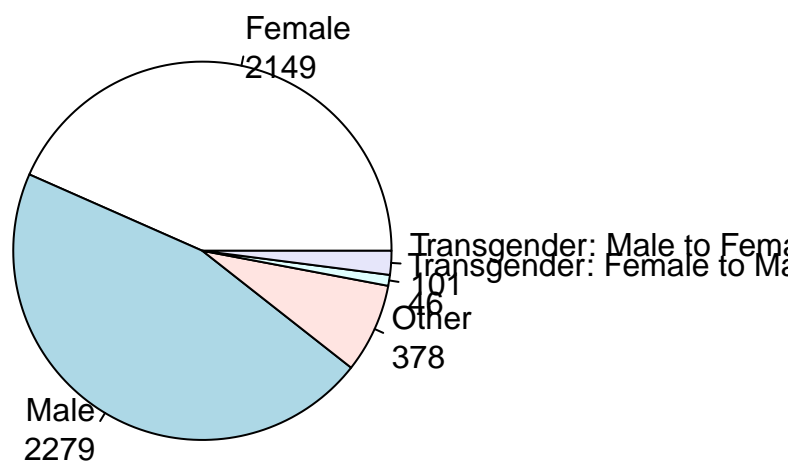
```
ggplot(data, aes(x = gender)) +
  geom_bar(fill = "light blue") +
  labs(x = "Gender",
       y = "Frequency",
       title = "Distribution of Respondant Gender") +
  theme(axis.text.x = element_text(angle = 45,
                                    hjust = 1))
```

Distribution of Respondant Gender

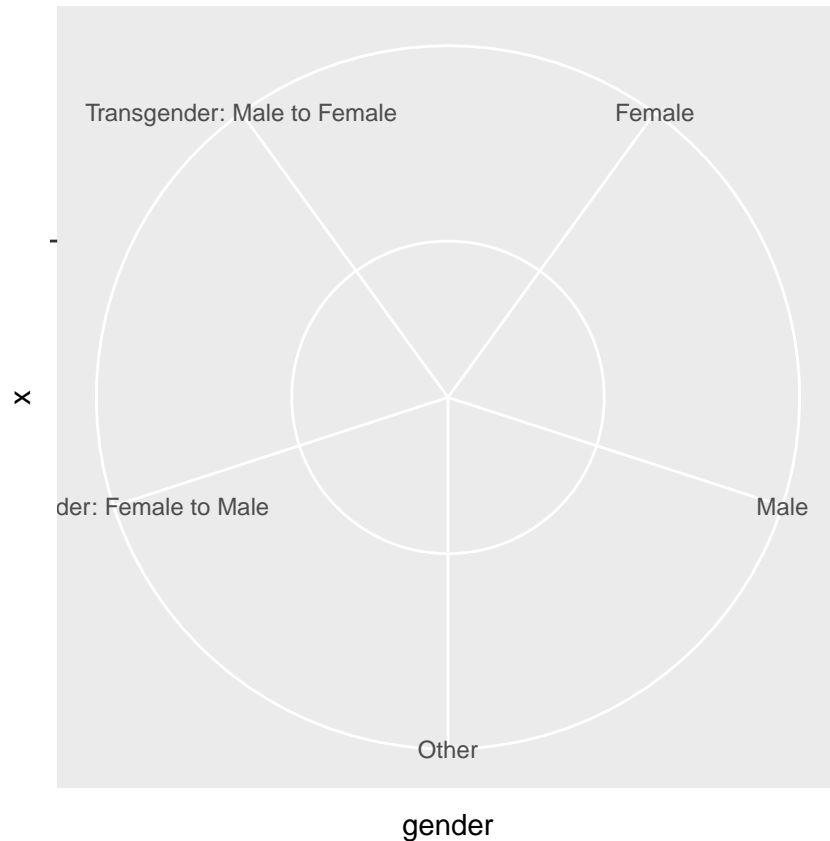


```
# Pie Chart with Percentages
gendertable <- table(data$gender)
lbls <- paste(names(gendertable), "\n", gendertable, sep="")
pie(gendertable, labels = lbls,
    main="Pie Chart of Gender Distribution of Respondents")
```

Pie Chart of Gender Distribution of Respondents



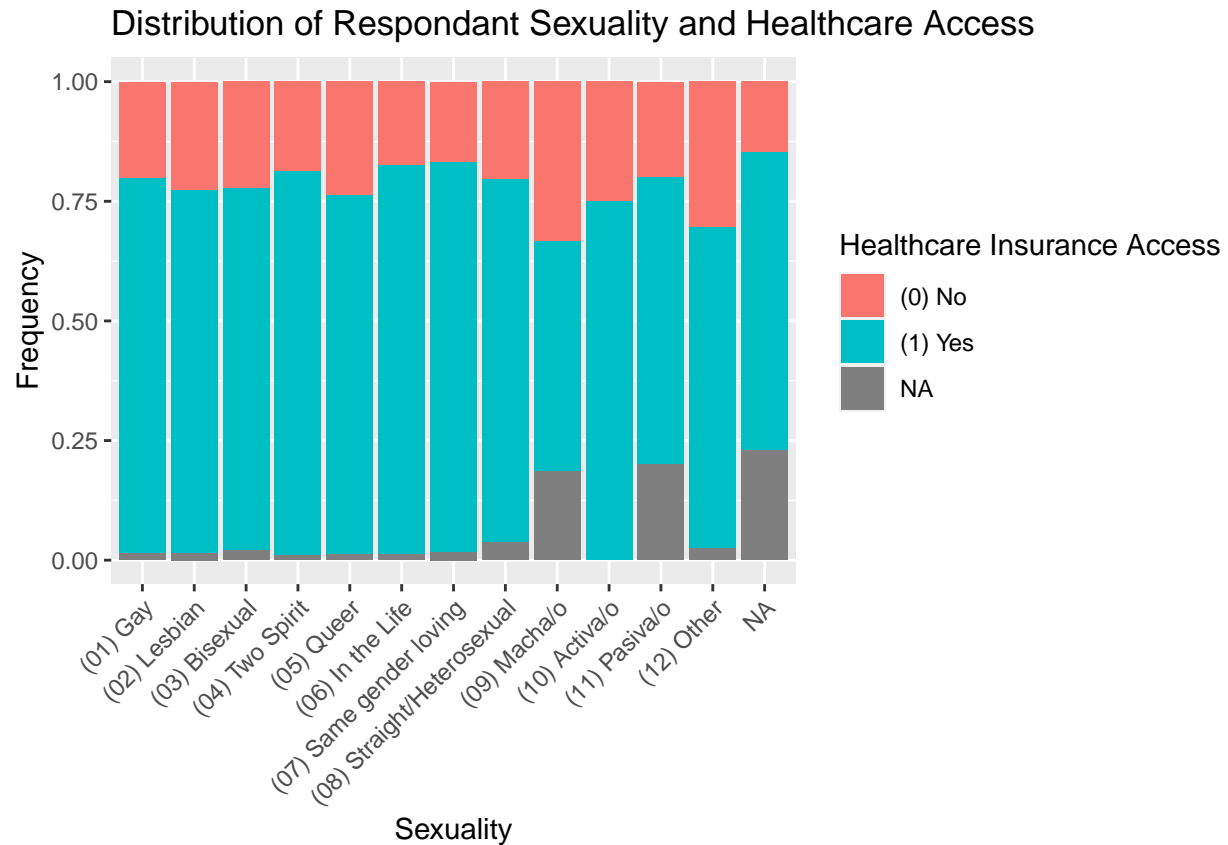
```
# attempt at ggplot2 way...  
ggplot(data, aes(x = "", y = gender, fill = "gender")) +  
  coord_polar(theta = "y")
```



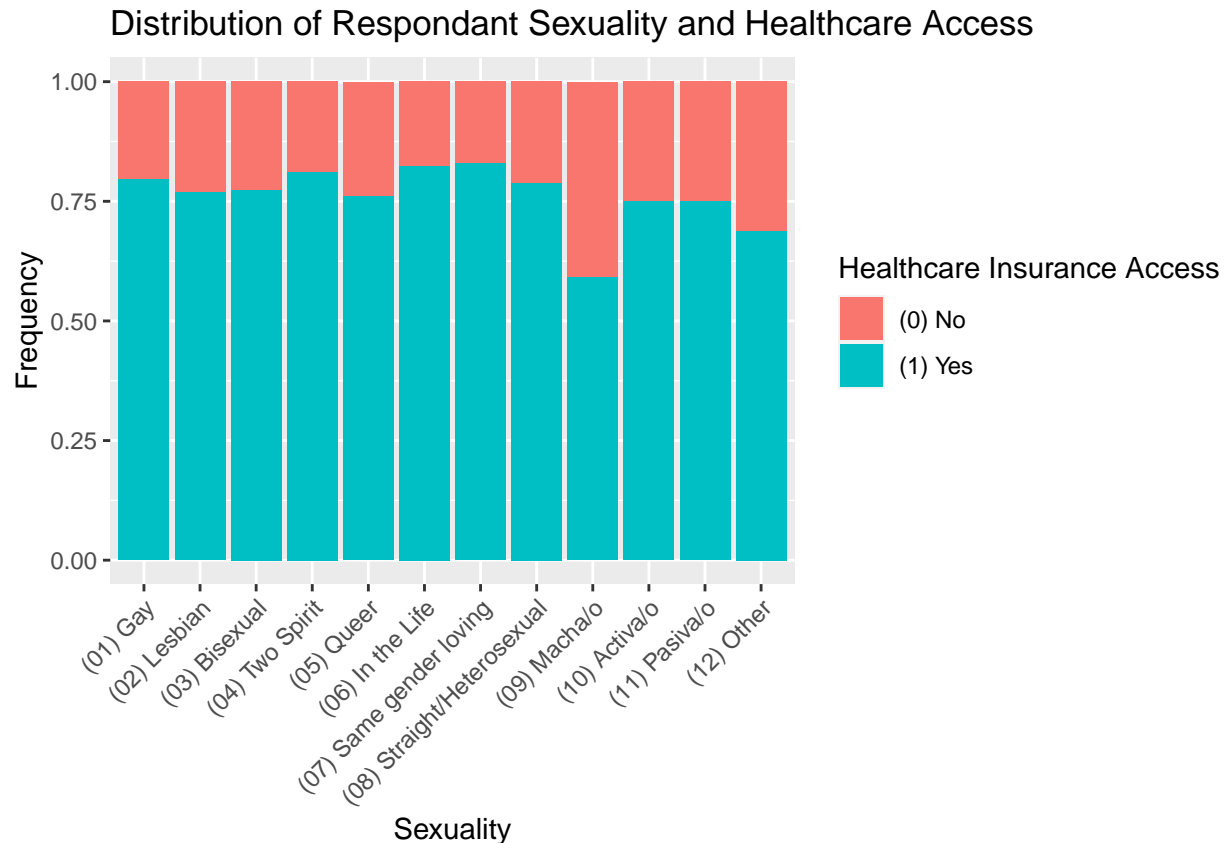
```
#geom_label_repel(nudge_y = 1, direction = "y", hjust = 1) #for if label needs adjusting?
```

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))
```



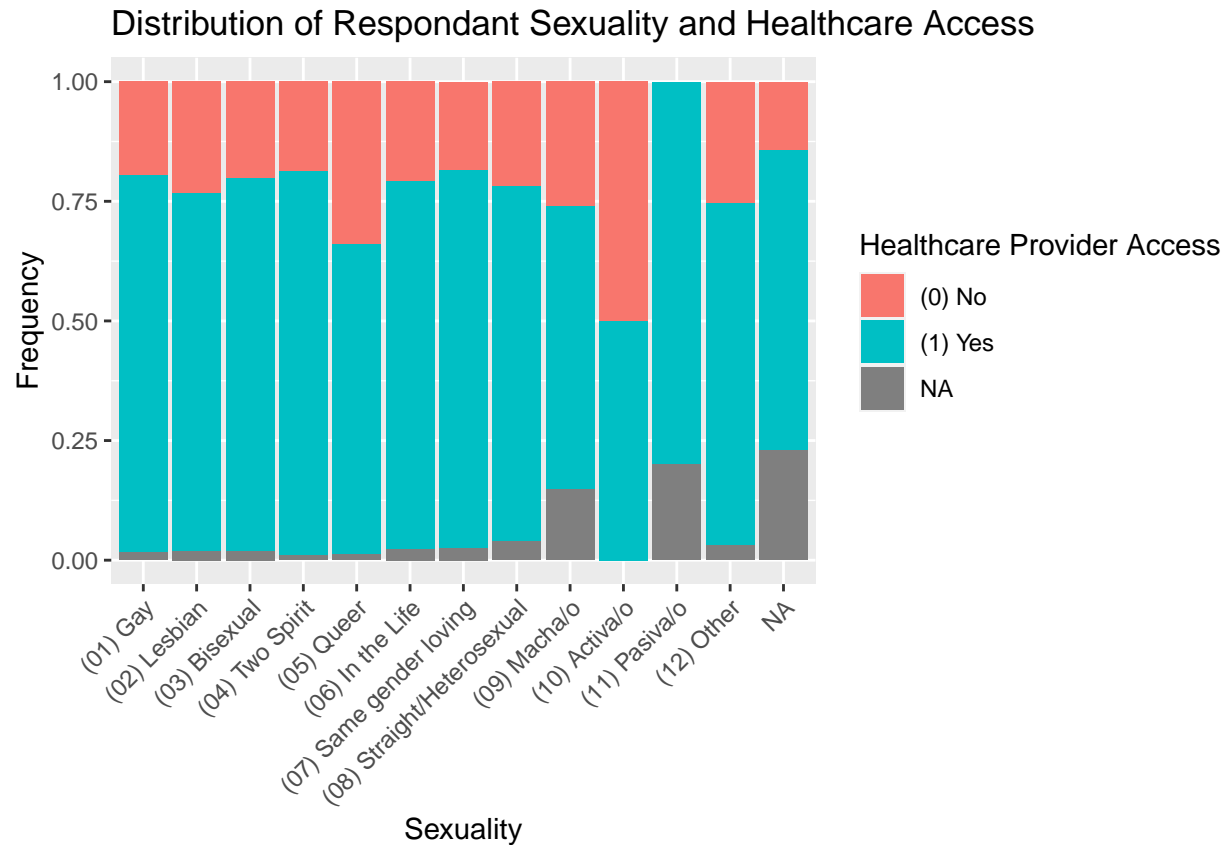
```
data %>%
  filter(!is.na(healthInsureAcc),
         !is.na(sexuality)
  ) %>%
  ggplot(aes(x = sexuality,
             fill = healthInsureAcc)) +
  geom_bar(position = "fill") +
  labs (x = "Sexuality",
        y = "Frequency",
        fill = "Healthcare Insurance Access",
        title = "Distribution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



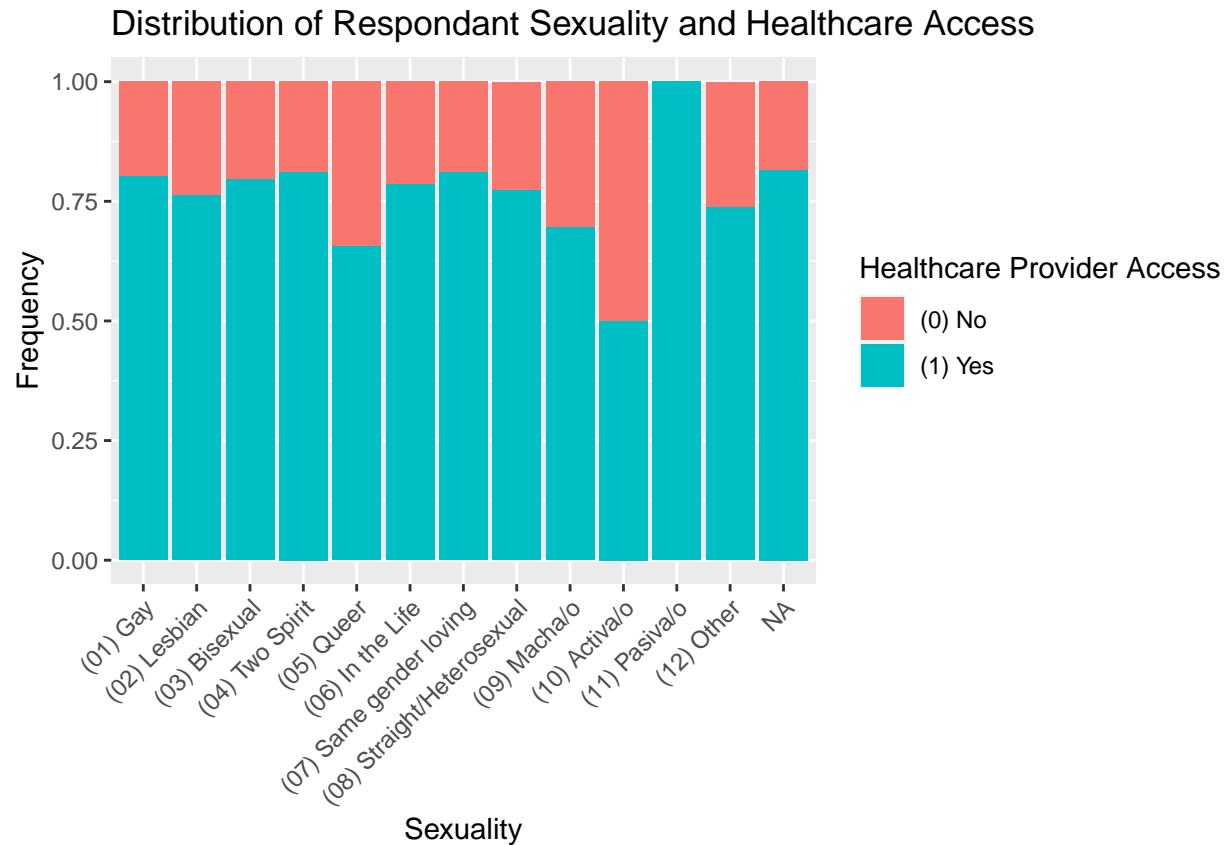
We used a segmented bar graph to visualize sexuality of respondents and their access to healthcare. The differences are not so visually distinct, especially when attempting to compare straight/heterosexual individuals to other historically and presently marginalized sexual identities. This is especially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

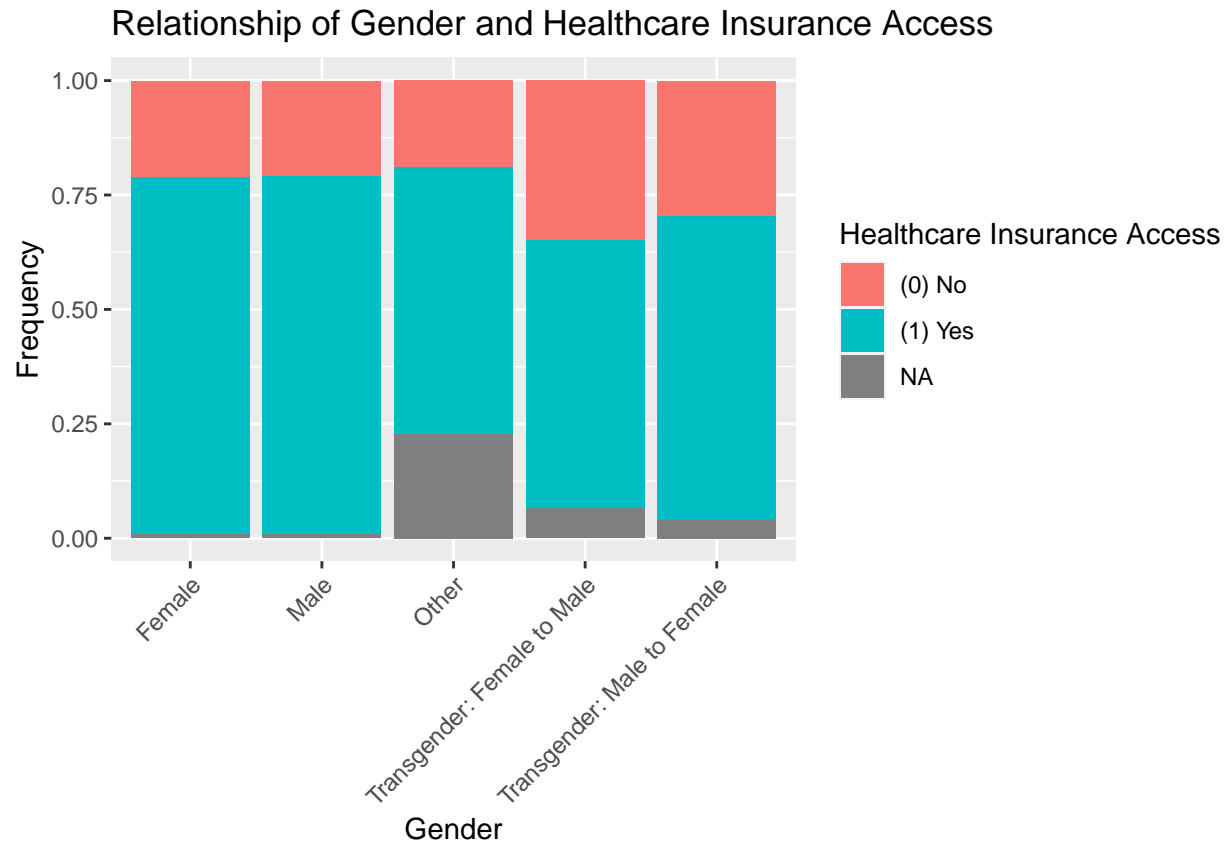
```
data %>%
  ggplot(aes(x = sexuality,
              fill = healthProvideAcc)) +
  geom_bar(position = "fill") +
  labs (x = "Sexuality",
        y = "Frequency",
        fill = "Healthcare Provider Access",
        title = "Distribution of Respondant Sexuality and Healthcare Access") +
  theme(axis.text.x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



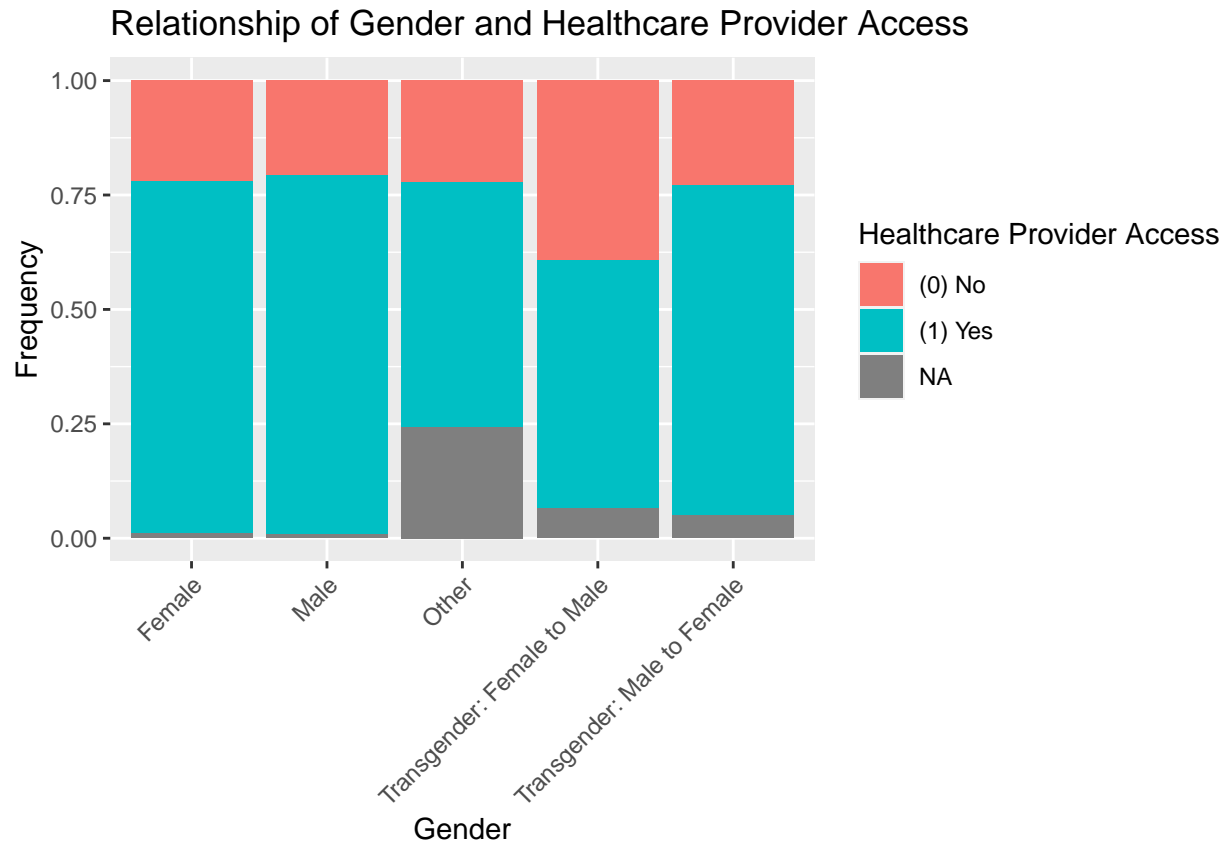
```
data %>%
  filter(is.na(healthProvideAcc) == FALSE) %>%
  ggplot(aes(x = sexuality,
             fill = healthProvideAcc)) +
  geom_bar(position = "fill") +
  labs (x = "Sexuality",
        y = "Frequency",
        fill = "Healthcare Provider Access",
        title = "Distribution of Respondant Sexuality 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 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 statistically significant impacts on health outcomes due to sexuality, gender, race, and intersectional identities, we ran a few statistical tests.

we may not need this?? To first observe statistical significance of variables

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

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

We recognize the limitations of not only the collection method, e.g. allowing for only one answer to a question with many options that are not necessarily mutually exclusive, but also fitting identity, neither quantifiable

nor simple, for the sake of data analysis.

Sexuality took the value of 0 if straight/heterosexual, and 1 if of another sexuality in response to the question “Which one label comes closest to how you describe your sexual identity?”.

Gender took the value of 1 if transgender (F2M or M2F), cisgender (female or male). For race, we kept each group to recognize the different experiences of the socially constructed categories.

```
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)
#
# 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)

insurancefit <- logistic_reg() %>%
  set_engine("glm") %>%
  fit(healthInsureAcc ~ race + gender,
```

```

data = data, family = "binomial")

tidy(insurancefit, conf.int = TRUE, exponentiate = TRUE)

## # A tibble: 11 x 7
##   term                estimate std.error statistic  p.value conf.low conf.high
##   <chr>                <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept)          5.39      0.168     10.0  9.36e-24    3.92     7.57
## 2 raceBlack            0.825     0.175     -1.10  2.70e- 1    0.579     1.15
## 3 raceHispanic         0.488     0.184     -3.90  9.60e- 5    0.337     0.694
## 4 raceMultiracial      0.612     0.177     -2.77  5.55e- 3    0.428     0.858
## 5 raceNative           0.435     0.367     -2.27  2.34e- 2    0.215     0.913
## 6 raceOther            0.518     0.268     -2.45  1.42e- 2    0.307     0.881
## 7 raceWhite            0.757     0.182     -1.53  1.26e- 1    0.525     1.07
## 8 genderMale           0.998     0.0743    -0.0286 9.77e- 1    0.863     1.15
## 9 genderOther          0.832     0.147     -1.25  2.10e- 1    0.626     1.12
## 10 genderTransgender: ~ 0.472     0.323     -2.33  1.98e- 2    0.253     0.906
## 11 genderTransgender: ~ 0.612     0.228     -2.16  3.10e- 2    0.395     0.968

```

```

providerfit <- logistic_reg() %>%
  set_engine("glm") %>%
  fit(healthProvideAcc ~ race + gender,
      data = data, family = "binomial")

tidy(insurancefit, conf.int = TRUE, exponentiate = TRUE)

```

```

## # A tibble: 11 x 7
##   term                estimate std.error statistic  p.value conf.low conf.high
##   <chr>                <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
## 1 (Intercept)          5.39      0.168     10.0  9.36e-24    3.92     7.57
## 2 raceBlack            0.825     0.175     -1.10  2.70e- 1    0.579     1.15
## 3 raceHispanic         0.488     0.184     -3.90  9.60e- 5    0.337     0.694
## 4 raceMultiracial      0.612     0.177     -2.77  5.55e- 3    0.428     0.858
## 5 raceNative           0.435     0.367     -2.27  2.34e- 2    0.215     0.913
## 6 raceOther            0.518     0.268     -2.45  1.42e- 2    0.307     0.881
## 7 raceWhite            0.757     0.182     -1.53  1.26e- 1    0.525     1.07
## 8 genderMale           0.998     0.0743    -0.0286 9.77e- 1    0.863     1.15
## 9 genderOther          0.832     0.147     -1.25  2.10e- 1    0.626     1.12
## 10 genderTransgender: ~ 0.472     0.323     -2.33  1.98e- 2    0.253     0.906
## 11 genderTransgender: ~ 0.612     0.228     -2.16  3.10e- 2    0.395     0.968

```

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.

```

summary(aov(Q18C~Q25,data=data_filtered2))

##           Df Sum Sq Mean Sq F value   Pr(>F)
## Q25         1    112   111.57   13.73 0.000213 ***
## Residuals 4552   36974     8.12
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## 399 observations deleted due to missingness

```

```
sexpair <- pairwise.t.test(data_filtered2$Q18C,data_filtered2$Q25, p.adj = "holm")
sigpairs <- broom::tidy(sexpair) %>%
  filter(p.value<0.05) %>%
  arrange(group1,group2)
nrow(sigpairs)
```

```
## [1] 2
```

```
print(sigpairs)
```

```
## # A tibble: 2 x 3
##   group1 group2 p.value
##   <chr>  <chr>   <dbl>
## 1 4      2      0.0217
## 2 5      2      0.0186
```

4: 2-spirit 5: queer