Final Report

due November 16, 2021 by 11:59 PM

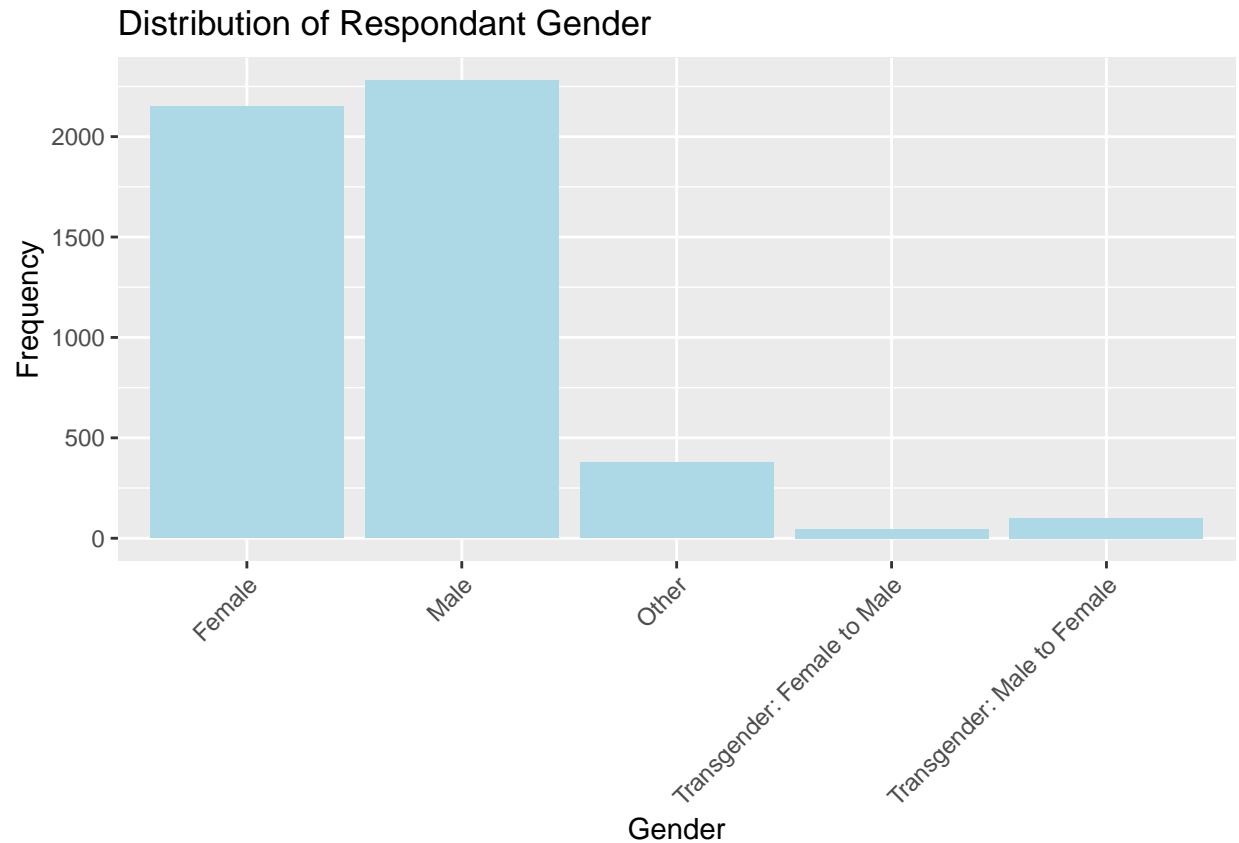Ashley Bae, Gi Chun, Jay Nolt

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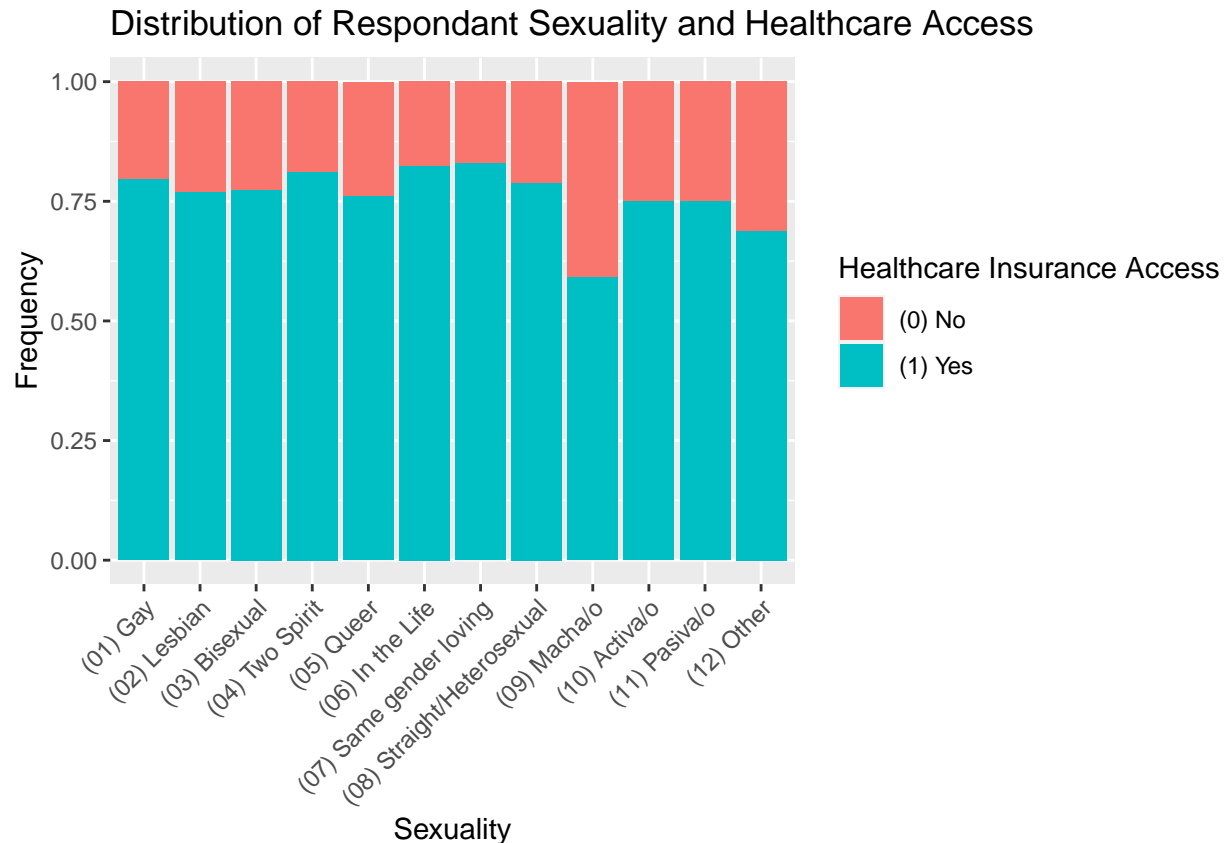
Data Citation: Battle, Juan, Pastrana, Antonio Jay, and Daniels, Jessie. Social Justice Sexuality Project: 2010 National Survey, including Puerto Rico. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor], 2013-08-09. https://doi.org/10.3886/ICPSR34363.v1

```
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

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.

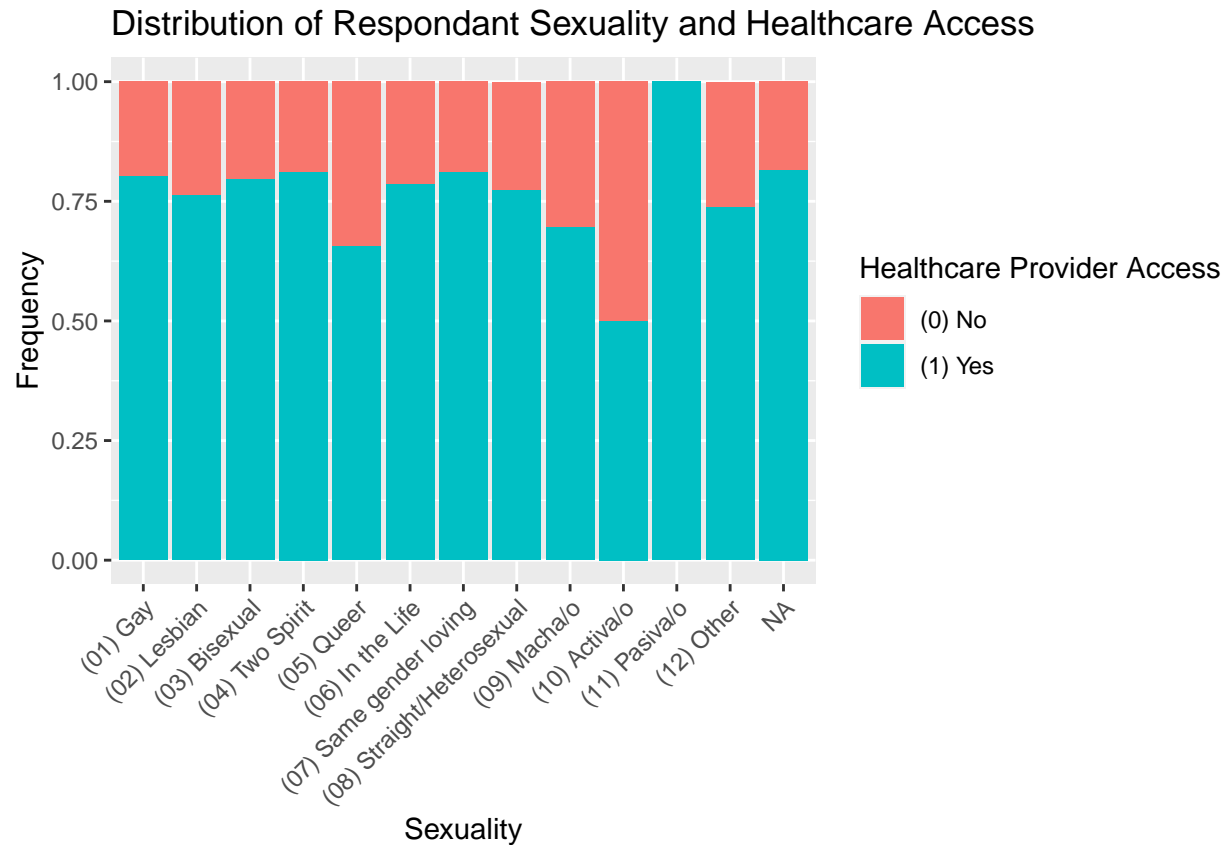```
data %>%
  filter(!is.na(healthInsureAcc),
         !is.na(sexuality)
) %>%
  ggplot(aes(x = sexuality,
          fill = healthInsureAcc)) +
  geom_bar(position = "fill") +
  labs (x = "Sexuality",
        y = "Frequency",
        fill = "Healthcare Insurance Access",
        title = "Distribution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

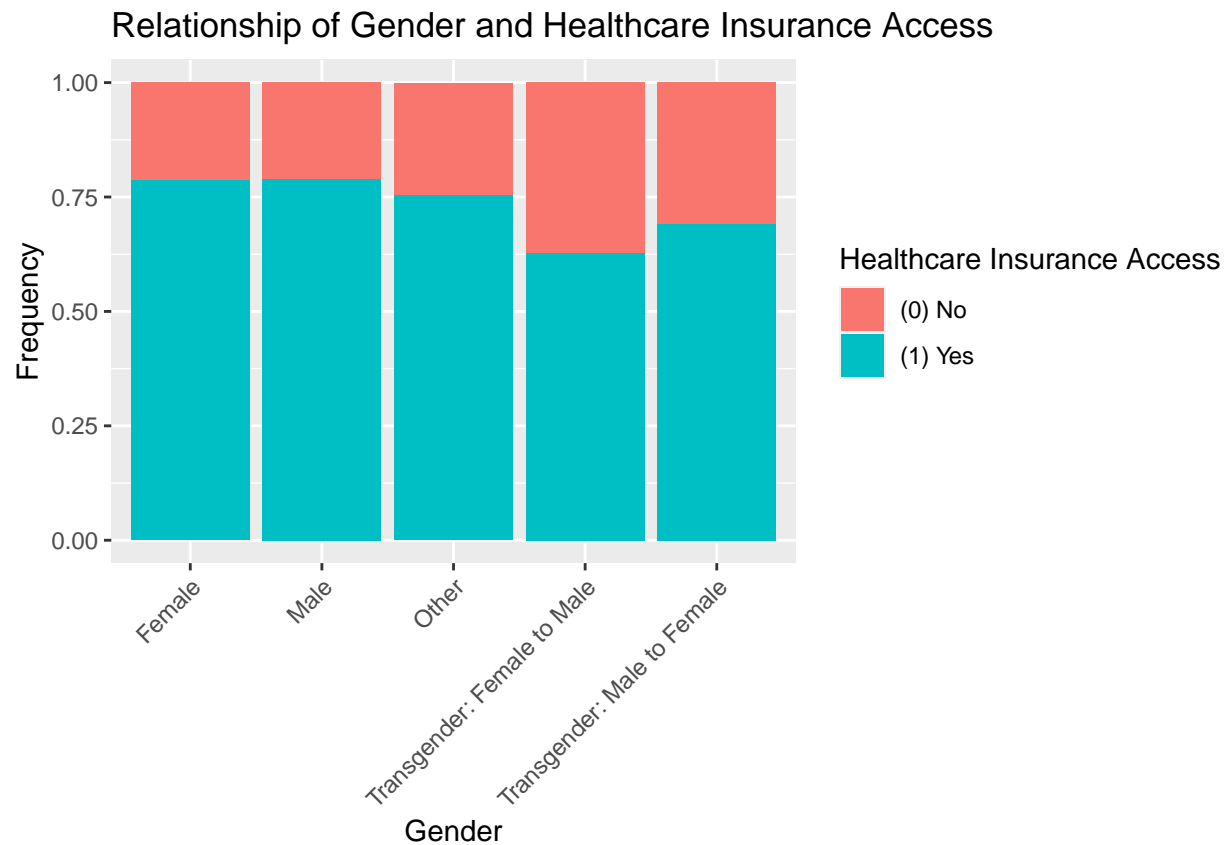Distribution of Respondant Sexuality and Healthcare Access

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 %>%
  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))
```

Distribution of Respondant Sexuality and Healthcare Access


```
data %>%
  filter(is.na(healthInsureAcc) == FALSE) %>%
  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))
```

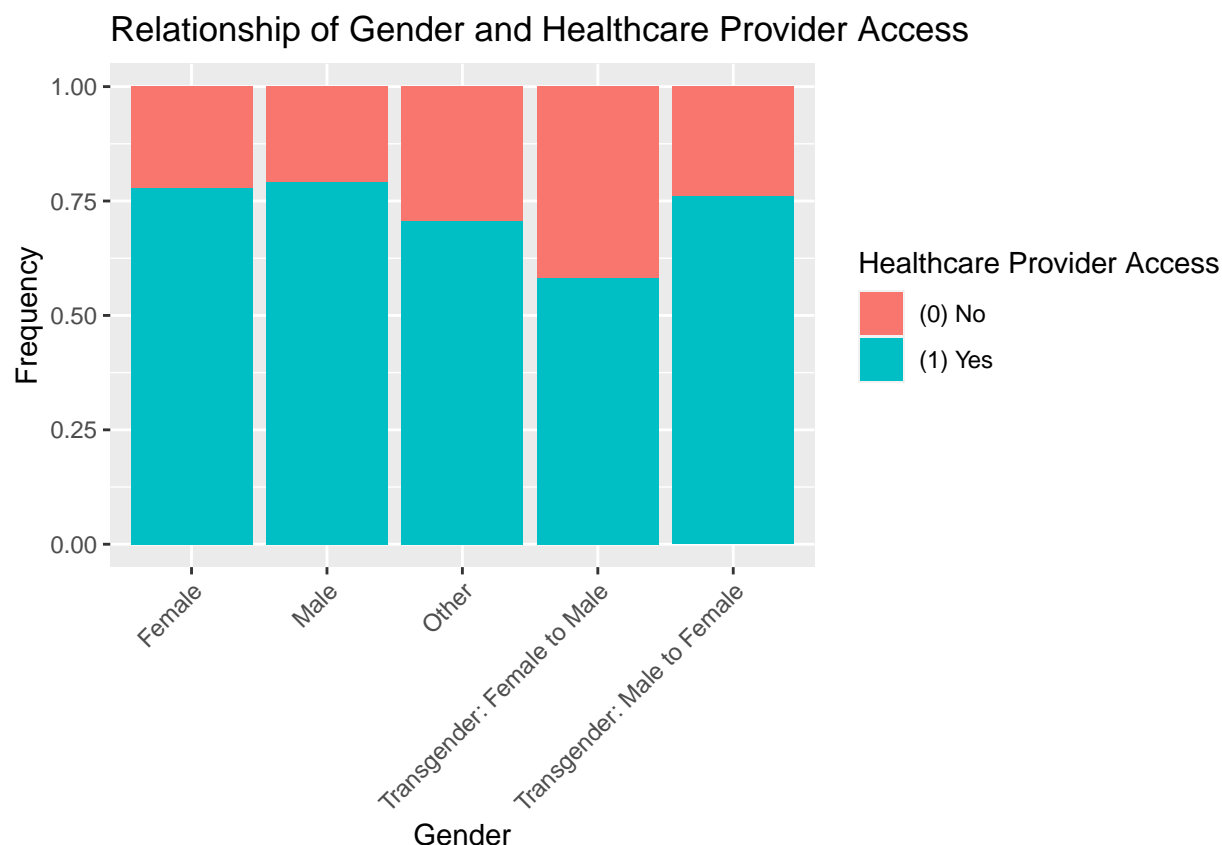# Relationship of Gender and Healthcare Insurance Access


```
data %>%
filter(is.na(healthProvideAcc) == FALSE) %>%
ggplot(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))
```

Relationship of Gender and Healthcare Provider Access

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.

Data Analysis

To evaluate statistically significant impacts on health outcomes due to sexuality, gender, race, and intersectional identities, we ran a few statistical tests.

```
##
## Call:
## glm(formula = healthInsureAcc ~ race + gender_trans + sexuality_queer,
##     family = "binomial", data = data_relevel)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9549   0.5782   0.6400   0.7339   0.9823
##
## Coefficients:
##               Estimate Std. Error z value Pr(>|z|)
## (Intercept)    1.45824    0.14642   9.959  < 2e-16 ***
## raceAsian      0.29236    0.18593   1.572  0.11586
```

```
## raceBlack          0.06985    0.10489    0.666   0.50548
## raceHispanic      -0.47353    0.11962   -3.959 7.54e-05 ***
## raceMultiracial   -0.23755    0.10799   -2.200   0.02782 *
## raceNative         -0.59574    0.33724   -1.767   0.07731 .
## raceOther          -0.34082    0.25905   -1.316   0.18828
## gender_trans1      -0.33816    0.11871   -2.849   0.00439 **
## sexuality_queer1   -0.04648    0.12939   -0.359   0.71941
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 4916.6  on 4682  degrees of freedom
## Residual deviance: 4868.7  on 4674  degrees of freedom
##   (270 observations deleted due to missingness)
## AIC: 4886.7
##
## Number of Fisher Scoring iterations: 4

##
## Call:
## glm(formula = healthProvideAcc ~ race + gender_trans + sexuality_queer,
##     family = "binomial", data = data_relevel)
##
## Deviance Residuals:
##     Min       1Q    Median       3Q       Max
## -1.9047   0.6097    0.7018   0.7359    0.9721
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)       1.275621   0.142233   8.969  < 2e-16 ***
## raceAsian        -0.115076   0.161243  -0.714 0.475425
## raceBlack         0.320888   0.103583   3.098 0.001949 **
## raceHispanic     -0.323065   0.117352  -2.753 0.005906 **
## raceMultiracial  -0.099533   0.105247  -0.946 0.344294
## raceNative        0.368565   0.400993   0.919 0.358027
## raceOther         0.064777   0.274042   0.236 0.813140
## gender_trans1    -0.440056   0.117257  -3.753 0.000175 ***
## sexuality_queer1 -0.008221   0.126545  -0.065 0.948199
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 4958.5  on 4670  degrees of freedom
## Residual deviance: 4904.5  on 4662  degrees of freedom
##   (282 observations deleted due to missingness)
## AIC: 4922.5
##
## Number of Fisher Scoring iterations: 4

##
## Call:
## glm(formula = healthInsureAcc ~ race * gender_trans + gender_trans *
##     sexuality_queer + sexuality_queer * race, family = "binomial",
```

```
##     data = data_relevel)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9988   0.5775   0.6460   0.7459   1.1871
##
## Coefficients:
##                                 Estimate Std. Error z value Pr(>|z|)
## (Intercept)                      0.80426    0.26626   3.021  0.00252 **
## raceAsian                        1.04763    0.51592   2.031  0.04230 *
## raceBlack                        0.90561    0.36293   2.495  0.01259 *
## raceHispanic                     0.44759    0.41109   1.089  0.27625
## raceMultiracial                  0.54466    0.36802   1.480  0.13888
## raceNative                      -0.31644    1.67292  -0.189  0.84997
## raceOther                       -0.50990    0.73280  -0.696  0.48654
## gender_trans1                   -0.33030    0.44236  -0.747  0.45526
## sexuality_queer1                 0.69278    0.27897   2.483  0.01302 *
## raceAsian:gender_trans1          0.01243    0.60269   0.021  0.98355
## raceBlack:gender_trans1          0.26002    0.35185   0.739  0.45990
## raceHispanic:gender_trans1       0.05596    0.44613   0.125  0.90018
## raceMultiracial:gender_trans1    0.40846    0.34952   1.169  0.24256
## raceNative:gender_trans1        -0.15752    0.82461  -0.191  0.84851
## raceOther:gender_trans1          0.01304    0.61666   0.021  0.98313
## gender_trans1:sexuality_queer1  -0.22731    0.40448  -0.562  0.57414
## raceAsian:sexuality_queer1      -0.83778    0.54873  -1.527  0.12682
## raceBlack:sexuality_queer1      -0.94187    0.37634  -2.503  0.01232 *
## raceHispanic:sexuality_queer1   -1.01951    0.42852  -2.379  0.01735 *
## raceMultiracial:sexuality_queer1 -0.90439   0.38226  -2.366  0.01799 *
## raceNative:sexuality_queer1     -0.24233    1.62682  -0.149  0.88158
## raceOther:sexuality_queer1       0.25343    0.77473   0.327  0.74358
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 4916.6  on 4682  degrees of freedom
## Residual deviance: 4855.8  on 4661  degrees of freedom
##   (270 observations deleted due to missingness)
## AIC: 4899.8
##
## Number of Fisher Scoring iterations: 4

##
## Call:
## glm(formula = healthProvideAcc ~ race * gender_trans + gender_trans *
##     sexuality_queer + sexuality_queer * race, family = "binomial",
##     data = data_relevel)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9876   0.6034   0.6661   0.7341   1.1118
##
## Coefficients:
##                                 Estimate Std. Error z value Pr(>|z|)
```

```
## (Intercept)                              0.82666    0.26830   3.081  0.00206 **
## raceAsian                                0.56621    0.46303   1.223  0.22139
## raceBlack                                0.78447    0.36138   2.171  0.02995 *
## raceHispanic                             0.24256    0.40080   0.605  0.54504
## raceMultiracial                          0.34510    0.36168   0.954  0.34000
## raceNative                              12.15323  229.63045   0.053  0.95779
## raceOther                                0.03062    0.77701   0.039  0.96857
## gender_trans1                           -0.17389    0.44418  -0.391  0.69544
## sexuality_queer1                         0.49386    0.27989   1.764  0.07765 .
## raceAsian:gender_trans1                 -0.41998    0.52078  -0.806  0.42000
## raceBlack:gender_trans1                  0.38869    0.35516   1.094  0.27378
## raceHispanic:gender_trans1               0.38897    0.46113   0.843  0.39895
## raceMultiracial:gender_trans1            0.07483    0.33657   0.222  0.82405
## raceNative:gender_trans1                -0.23994    0.88561  -0.271  0.78644
## raceOther:gender_trans1                 -0.51226    0.63813  -0.803  0.42212
## gender_trans1:sexuality_queer1          -0.42527    0.40179  -1.058  0.28986
## raceAsian:sexuality_queer1              -0.71129    0.49050  -1.450  0.14703
## raceBlack:sexuality_queer1              -0.54599    0.37493  -1.456  0.14532
## raceHispanic:sexuality_queer1           -0.64560    0.41821  -1.544  0.12265
## raceMultiracial:sexuality_queer1        -0.49195    0.37553  -1.310  0.19018
## raceNative:sexuality_queer1            -11.78735  229.62994  -0.051  0.95906
## raceOther:sexuality_queer1               0.24841    0.82073   0.303  0.76214
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 4958.5  on 4670  degrees of freedom
## Residual deviance: 4893.5  on 4649  degrees of freedom
##   (282 observations deleted due to missingness)
## AIC: 4937.5
##
## Number of Fisher Scoring iterations: 11

##    Resid. Df      Resid. Dev          Df        Deviance        Pr(>Chi)
## Min.   :4661    Min.   :4856    Min.   :13    Min.   :12.93    Min.   :0.4535
## 1st Qu.:4664    1st Qu.:4859    1st Qu.:13    1st Qu.:12.93    1st Qu.:0.4535
## Median :4668    Median :4862    Median :13    Median :12.93    Median :0.4535
## Mean   :4668    Mean   :4862    Mean   :13    Mean   :12.93    Mean   :0.4535
## 3rd Qu.:4671    3rd Qu.:4865    3rd Qu.:13    3rd Qu.:12.93    3rd Qu.:0.4535
## Max.   :4674    Max.   :4869    Max.   :13    Max.   :12.93    Max.   :0.4535
##                                 NA's   :1     NA's   :1        NA's   :1

##    Resid. Df      Resid. Dev          Df        Deviance        Pr(>Chi)
## Min.   :4649    Min.   :4894    Min.   :13    Min.   :11.03    Min.   :0.6082
## 1st Qu.:4652    1st Qu.:4896    1st Qu.:13    1st Qu.:11.03    1st Qu.:0.6082
## Median :4656    Median :4899    Median :13    Median :11.03    Median :0.6082
## Mean   :4656    Mean   :4899    Mean   :13    Mean   :11.03    Mean   :0.6082
## 3rd Qu.:4659    3rd Qu.:4902    3rd Qu.:13    3rd Qu.:11.03    3rd Qu.:0.6082
## Max.   :4662    Max.   :4905    Max.   :13    Max.   :11.03    Max.   :0.6082
##                                 NA's   :1     NA's   :1        NA's   :1
```

We ran a multiple regression for each health insurance access and health provider access conditional on gender, race, and sexuality. We saw that our referent group of white, cis, heterosexual individuals are 1.47 as likely to have health insurance than non-white, trans, non-heterosexual counterparts and 1.32 times as likely

for access to a regular healthcare provider.

Sexuality is a complicated variable to work with, given that the question posed in the survey offers 12 options that are not all mutually exclusive, yet the answer given must be only one of the above. Also, to retain all options given would overcomplicate and weaken the model. Therefore, we decided to distinguish between heterosexual and all other sexuality answers in our subsequent models.

In healthcare access, we found that the most negatively affected race group was Hispanic and the most negatively affected gender group was transgender individuals.

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

Statistically, 2-spirit and queer are different from the rest