Final Report

due November 16, 2021 by 11:59 PM

Ashley Bae, Gi Chun, Jay Nolt

November 2021

Data

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

Our data came from the Social Justice Sexuality Project (SJS) through the University of Michigan's Resource Center for Minority Data. From January of 2010 to December of 2010, this project conducted surveys with over 5,000 participants who were Black, Latina/o, Asian and Pacific Islander, and multiracial lesbian, gay, bisexual, and transgender (LGBT) people in 50 US states and Puerto Rico.

The survey was given via mail questionnaire, on-site questionnaire, and web-based survey in Spanish and English depending on the respondent. In the data set itself, each case is an individual respondent who filled out the survey. Each question is a variable with the value determined by an individual's survey response. Since there were hundreds of variables, we chose to remove some columns since we were looking at a more narrow scope of variables. Although the variables touch on very broad topics, this data set particularly focuses on questions relating to LGBT communities and people of color.

We recognize the limitations of not only the collection method, e.g. allowing for only one answer of sexuality to a question with many options that are not necessarily mutually exclusive, convenience sampling, but also fitting identity, neither quantifiable nor simple, for the sake of data analysis. Additionally, some variables are not necessarily independent. For example, if someone were to say they were lesbian, it's most likely that they're female or M2F transgender.

Our hypothesis is that those who exist at the intersection of marginalized minority groups in race and sexuality would have compounded impacts of inequality than if they were only representing one of the marginalized identities. For example, someone who is black and queer would face more negative health outcomes compared to someone who's either black or queer.

Data Wrangling

We had data that was not in an ideal form to be visualized or analyzed. Thus, we used data wrangling to synthesize many responses into a single variable.

Since respondents could pick multiple selections for gender identity or specify other, we collapsed all the answers into a single variable. If respondents selected multiple genders or listed their own, they would be categorized as other. When conducting analyses, we differentiated between cis and transgender individuals for ease of analysis.

When conducting data analyses involving sexuality, we differentiated between queer and straight individuals for ease of analysis.

Since respondents could pick multiple selections for racial identity, we synthesized multiple selections into a single variable. If respondents selected multiple races, they would be listed as multiracial. In this way, we could create data visualizations showing an identifiable number of groups to identify the general distribution. However, this also means that the multiracial category encompass many identities, so we must proceed with caution when analyzing its significance. The following is an example of how we synthesized our race variable based on many individual questions.

```r
data <- data %>%
  mutate(black = case_when(
  black == "(1) Yes" ~ "black",
  is.na(male) ~ "",
  TRUE ~ ""),
  hispanic = case_when(
  hispanic == "(1) Yes" ~ "hispanic",
  is.na(hispanic) ~ "",   TRUE ~ ""),
  asian = case_when(
  asian == "(1) Yes" ~ "asian",
  is.na(asian) ~ "",
  TRUE ~ ""),
  native = case_when(
  native == "(1) Yes" ~ "native",
  is.na(native) ~ "", TRUE ~ ""),
  white = case_when(
  white == "(1) Yes" ~ "white",
  is.na(white) ~ "",
  TRUE ~ ""),
  multi = case_when(
  multi == "(1) Yes" ~ "multi",
  is.na(multi) ~ "", TRUE ~ ""),
  raceOther = case_when(
  raceOther == "(1) Yes" ~ "Other",
  is.na(raceOther) ~ "",
  TRUE ~ ""),
  race = ""
)
data$race <- paste(data$black, data$hispanic)
data$race <- paste(data$race, data$asian)
data$race <- paste(data$race, data$native)
data$race <- paste(data$race, data$white)
data$race <- paste(data$race, data$multi)
data$race <- paste(data$race, data$raceOther)
data$race <- trimws(data$race)
data <- data %>%
mutate(race = case_when(
  race == "black" ~ "Black",
  race == "hispanic" ~ "Hispanic",
  race == "asian" ~ "Asian",
  race == "native" ~ "Native",
  race == "white" ~ "White",
  race == "multi" ~ "Multiracial",
  race == "other" ~ "Other",
  race == "" ~ "Other",
  TRUE ~ "Multiracial"))
```

Preliminary Data Visualizations

We visualized the distribution of respondent gender identities of participants with a bar graph. The biggest group of respondents are male, followed by female, followed by "other." There are more female transgender individuals than male transgender individuals captured in this study.

Distribution of Respondant Gender
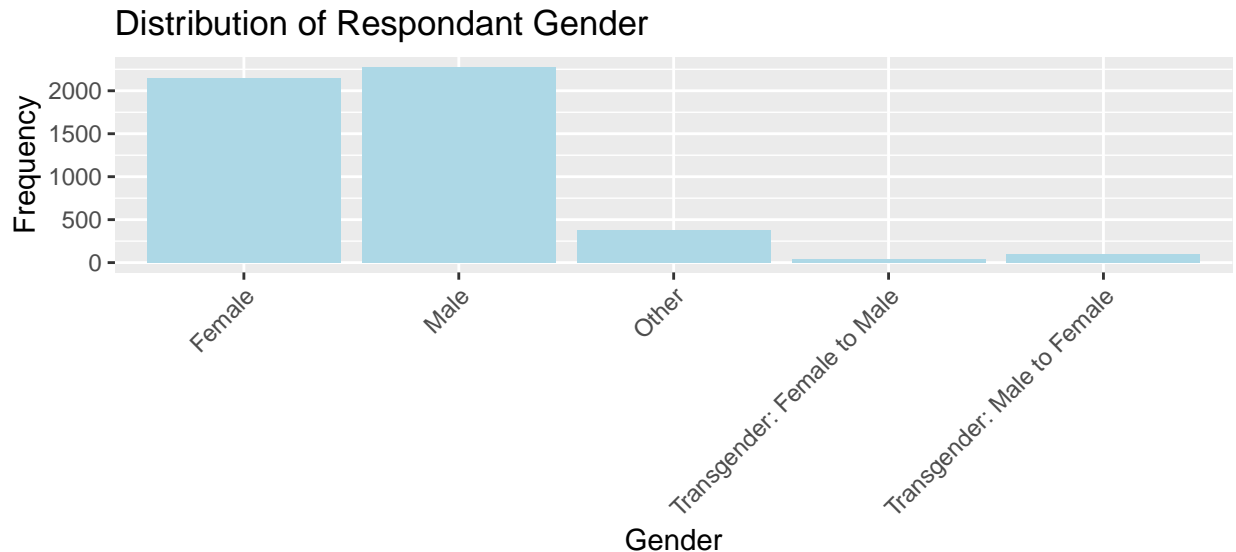
We used a segmented bar graph to visualize sexuality of respondents and their access to healthcare. The differences are not so visually distinct, especially when attempting to compare straight/heterosexual individuals to other historically and presently marginalized sexual identities. This is especially 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?"
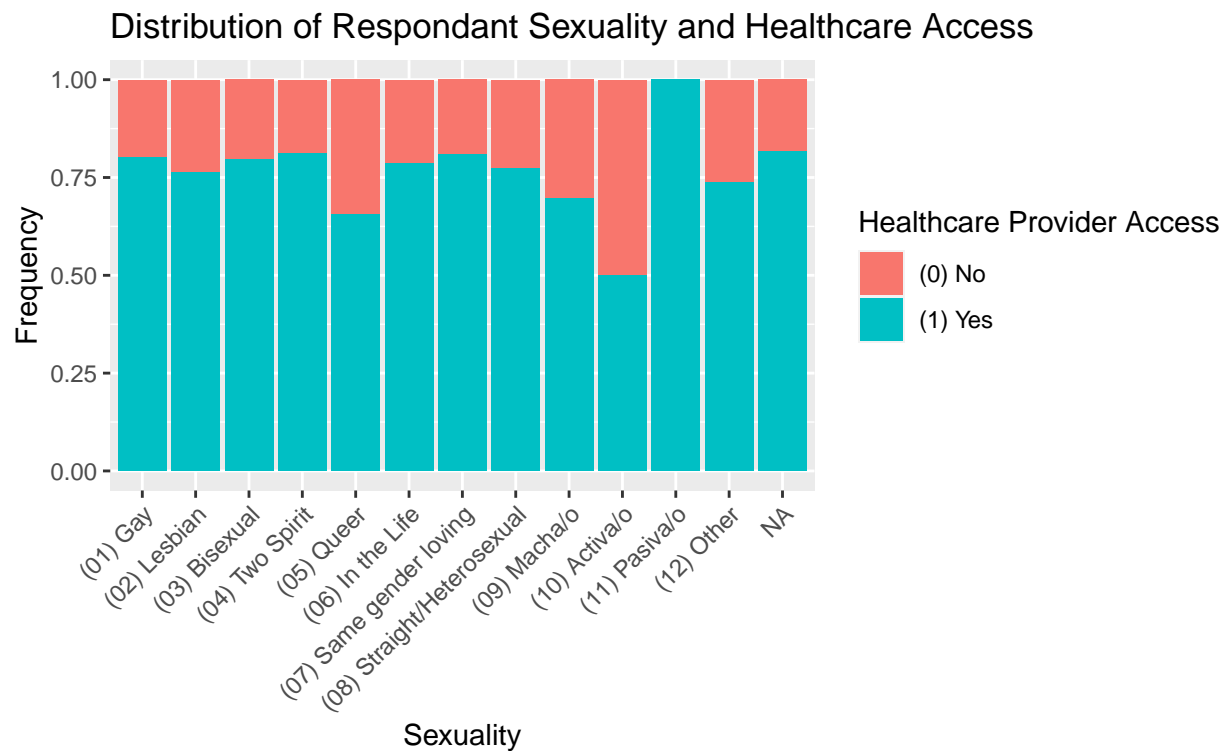
In the second visualization of sexuality and access to healthcare (pictured below), we visualized the with NA values removed. This helps us more clearly visualize the data. We can see more clearly that the group with least health insurance access is those identifying as macha/o.

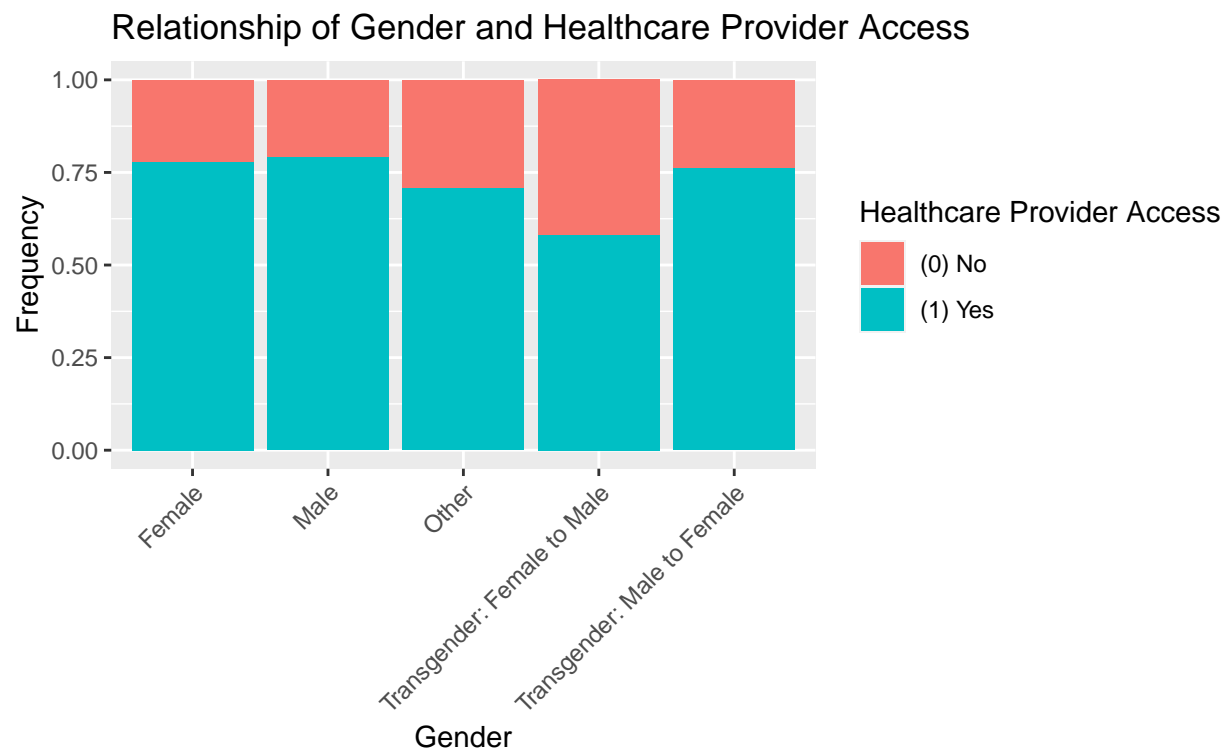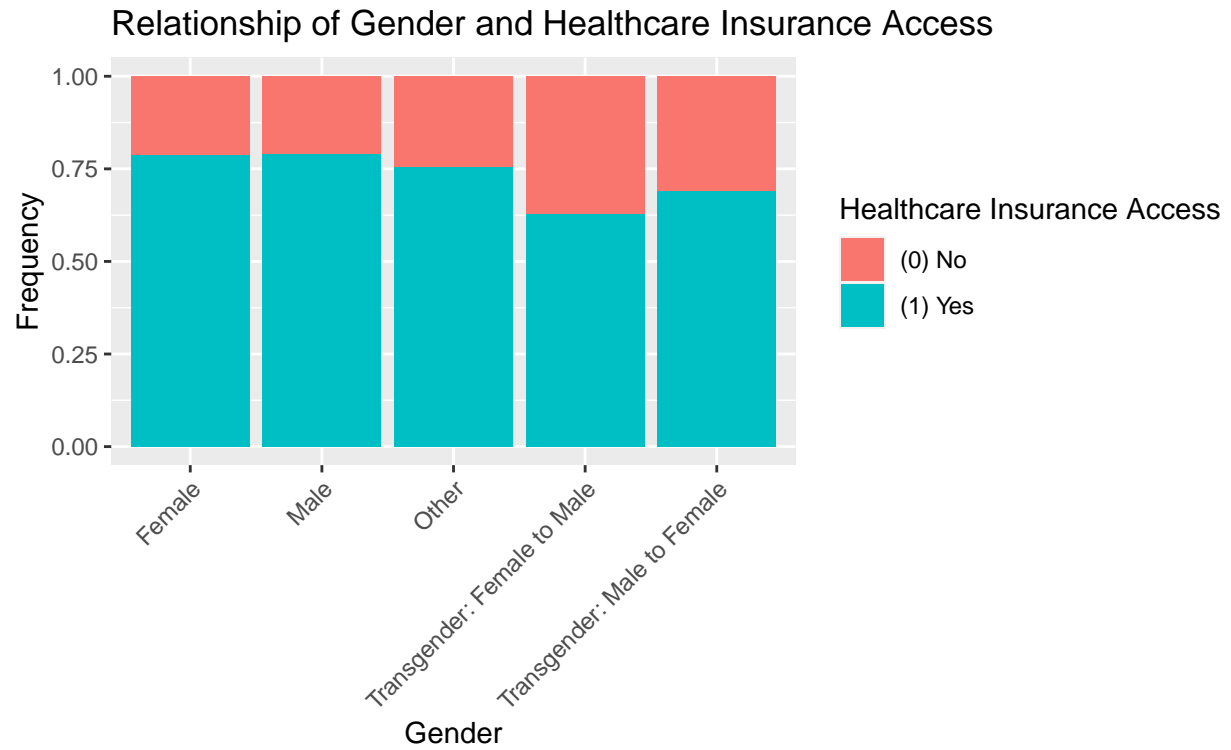## Distribution of Respondant Sexuality and Healthcare Access

3
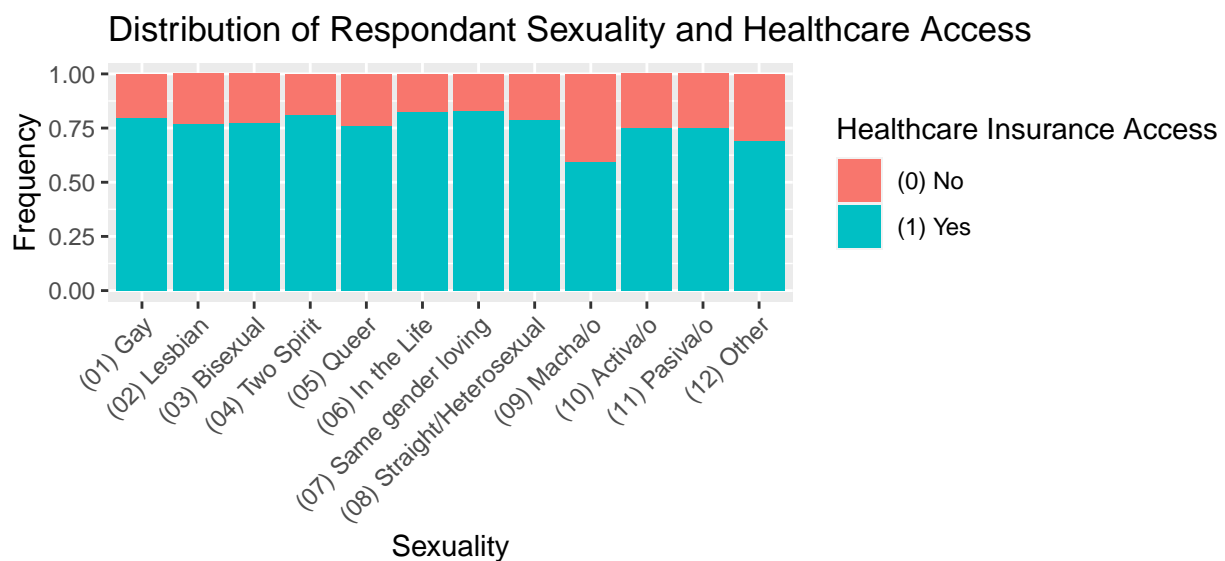
Further Data Visualizations

For the rest of our data visualizations, we filtered out NA values in order to better highlight the distribution of the information that was provided. We did this after identifying trends in NA values to ensure we weren't misrepresenting our data by removing NA values.

In the following 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. Additionally, when comparing the gender groups, we noted 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 into the apparent disparity. It appears that one's gender identity may have a relationship to access to healthcare.

Distribution of Respondent Sexuality and Healthcare Access

Relationship of Gender and Healthcare Insurance Access

Relationship of Gender and Healthcare Provider Access

Data Analysis

Logistic Regression

To evaluate statistically significant impacts on health outcomes due to sexuality, gender, race, and intersectional identities, we ran a few statistical tests. We selected white, cisgendered, heterosexual individuals as our referent group, to compare with other marginalized identities.

We ran a multiple regression for each health insurance access and health provider access conditional on gender, race, and sexuality. 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 models.

We conducted another set of multiple regression with interaction terms and attempted to better fit with specific interactions, in order to evaluate intersection of gender and race groups. Although found significant differences due to intersections of race and gender, we found that the models themselves fitted with interaction terms were not significantly better and fared worse than without interaction. To determine this, we used Akaike information criterion (AIC), which is a tool for comparing different models and a smaller number indicates a better fit, to conclude that it prefers the models without interactions. Thus, we proceeded with multiple regressions without interaction terms but variable on race, gender, and sexuality to predict each insurance and healthcare provider access.

```
# insurance fit -----------------------------------------------------------------
insurancefit <- glm(healthInsureAcc ~ race + gender_trans + sexuality_queer,
         data = data_relevel, family = "binomial")
summary(insurancefit, conf.int = TRUE, exponentiate = TRUE)
```

```
##
## 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
##   (87 observations deleted due to missingness)
```

```
## AIC: 4886.7
##
## Number of Fisher Scoring iterations: 4
# access fit --------------------------------------------------------------------
providerfit <- glm(healthProvideAcc ~ race + gender_trans + sexuality_queer,
          data = data_relevel, family = "binomial")
summary(providerfit, conf.int = TRUE, exponentiate = TRUE)


##
## 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
##   (99 observations deleted due to missingness)
## AIC: 4922.5
##
## Number of Fisher Scoring iterations: 4
exp(insurancefit$coefficients)


##      (Intercept)         raceAsian         raceBlack      raceHispanic
##        4.2983934         1.3395812         1.0723436         0.6228014
##  raceMultiracial        raceNative         raceOther      gender_trans1
##        0.7885551         0.5511542         0.7111845         0.7130832
## sexuality_queer1
##        0.9545807

exp(providerfit$coefficients)


##      (Intercept)         raceAsian         raceBlack      raceHispanic
##        3.5809252         0.8912982         1.3783509         0.7239271
##  raceMultiracial        raceNative         raceOther      gender_trans1
##        0.9052599         1.4456581         1.0669213         0.6440002
```

```
## sexuality_queer1
##        0.9918123
```

When we looked at the multiple logistical regression conditional on race, sexuality, and gender, we saw that our referent group of white, cis, heterosexual individuals are 4.29 times more likely to have health insurance than non-white, trans, non-heterosexual counterparts with a highly significant p value of less than 2e-16. White, cis, heterosexuals are also predicted to be 3.58 times more likely to have access to a regular healthcare provider than marginalized identities, with a significant p value of less than 2e-16.

To talk about statistically significant predictors to health insurance access, we found that the most negatively affected race group was Hispanic (with a p value of 7.54e-05), followed by transgender (with a p value of 0.00439), followed by multiracial individuals (with a p value of 0.02782), and followed by native individuals (with a p value of 0.07731).

For statistically significant predictors in whether or not someone has access to a regular healthcare provider or not, we found that the most negatively affected race group was transgender (with a p value of 0.000175), followed by Hispanic (with a p value of 0.005906), and followed by black individuals (with a p value of 0.001949).

Both Hispanic and transgender groups are significantly disadvantaged to access to both health care and a healthcare provider.

ANOVA test, Step Down Test, and t-tests

```
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
```

The overall ANOVA test resulted in a significant p-value. According to the overall ANOVA test conducted between sexuality and self-reported health assessment score, there are 2 groups that are shown to be statistically different than the other sexuality groups in their self-reported health assessment score. A step down test was conducted to see which sexuality groups were statistically different their self-reported health assessment score than the rest.

```
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
```

According to the step-down ANOVA, Two-Spirit and Queer sexuality groups have statistically significant different self-reported health assessment score from the rest of the sexuality groups. T-tests were conducted in

order to determine how Two-Spirit and Queer groups differ in comparison to Straight groups in self-reported healthcare assessment.

```r
data_test <- data%>%
  mutate(id = seq.int(nrow(data))) %>%
  select(id, sexuality, assessHealth) %>%
  pivot_wider(names_from = sexuality, values_from = assessHealth)
```

```r
data_test <- rename(data_test,"Straight"="(08) Straight/Heterosexual")
data_test <- rename(data_test,"Queer"="(05) Queer")
data_test <- rename(data_test,"TwoSpirit"="(04) Two Spirit")
```

```r
data_test
```

```
## # A tibble: 4,953 x 14
##        id Queer        `(03) Bisexual` `(01) Gay` `(02) Lesbian` `(07) Same gend~
##     <int> <fct>        <fct>           <fct>      <fct>          <fct>
## 1      1 (3) Good      <NA>            <NA>       <NA>           <NA>
## 2      2 <NA>          (5) Excellent   <NA>       <NA>           <NA>
## 3      3 <NA>          <NA>            <NA>       <NA>           <NA>
## 4      4 <NA>          <NA>            <NA>       (3) Good       <NA>
## 5      5 <NA>          <NA>            <NA>       (4) Very good  <NA>
## 6      6 <NA>          <NA>            <NA>       <NA>           (3) Good
## 7      7 (4) Very good <NA>            <NA>       <NA>           <NA>
## 8      8 <NA>          (5) Excellent   <NA>       <NA>           <NA>
## 9      9 <NA>          <NA>            <NA>       <NA>           <NA>
## 10    10 (4) Very good <NA>            <NA>       <NA>           <NA>
## # ... with 4,943 more rows, and 8 more variables: Straight <fct>,
## #   (12) Other <fct>, TwoSpirit <fct>, (06) In the Life <fct>, NA <fct>,
## #   (09) Macha/o <fct>, (11) Pasiva/o <fct>, (10) Activa/o <fct>
```

```r
  data_test$Straight <- as.numeric(data_test$Straight)
  data_test$Queer <- as.numeric(data_test$Queer)
  data_test$TwoSpirit <- as.numeric(data_test$TwoSpirit)
```

```r
t.test(data_test$Queer, data_test$Straight)
```

```
##
##  Welch Two Sample t-test
##
## data:  data_test$Queer and data_test$Straight
## t = -2.4816, df = 672.53, p-value = 0.01332
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.31606532 -0.03684074
## sample estimates:
## mean of x mean of y
##  3.447284  3.623737
```

The calculated p-value is 0.01332, indicating that the null hypothesis stating that there is no statistical difference between queer and straight groups in their self-reported health assessment score should be rejected. The confidence interval indicates that we are 95% confident that individuals that identity as queer self-report lower health assessment scores by a difference ranging from -0.316 to -0.037 in comparison to straight individuals.

```r
t.test(data_test$TwoSpirit, data_test$Straight)
```

```
## 
##  Welch Two Sample t-test
## 
## data:  data_test$TwoSpirit and data_test$Straight
## t = -1.2487, df = 157.58, p-value = 0.2136
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.3442768  0.0775713
## sample estimates:
## mean of x mean of y
##  3.490385  3.623737
```

This t-test calculated an insignificant p-value, indicating that we fail to reject the null hypothesis stating that there is no statistically significant difference between two spirit and straight groups in their self-reported healthcare assessment. Based on the ANOVA and Step Down test, two spirit groups may be statistically different from other sexuality groups, but not necessarily different with straight groups. Therefore, this difference is not necessarily relevant to our research question, although this test was valuable in its clarification.

Conclusion

We sought to answer the question: How do sexuality, race, and gender affect healthcare access? Our research yields the conclusion that minority identity groups do tend to have worse health care access than the majority groups. In healthcare access, we found that the most negatively affected race group was Hispanic and the most negatively affected gender group was transgender individuals.

Our logistic regression models which included interaction variables did not fit the data better than our model which included single variables. As a result, our conclusions oppose our original hypothesis that those who exist at the intersection of marginalized minority groups in race, gender, and sexuality would have statistically significant compounded impacts of inequality than if they were only representing one of the marginalized identities. While our results were statistically significant, we did have to make multiple simplifications in our data analysis which could have affected our models' goodness of fit, due to limitations earlier explained.

From our logistic regression models, both Hispanic and transgender groups are significantly disadvantaged to access to both health care and a healthcare provider. For the Hispanic race category, it may be explained that they are often disadvantaged in legal status and therefore legal protection, income, and health access in this case. Out of 188 non-citizens in this study, Hispanics number 57, the biggest race group of non-citizens. There are also many studies conducted about transgenders' access to healthcare providers as determined by less knowledge of the provider on trans issues. To quote a study by Safer et al. called "Barriers to Health Care for Transgender Individuals," "transgender patients report that lack of providers with expertise in transgender medicine represents the single largest component inhibiting access."

We would also like to state that although our analyses may have shown that statistically these identities do not impact healthcare outcomes or experiences, it is essential to note that these statistics do not mean that discrimination on any of these identities in healthcare does not exist. Our project was not designed to discount individual experiences based on our findings, and more sophisticated statistical analyses may be a better representation of our data in how this problem exists in the real world.

(citation: Safer et al. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4802845/#R14)