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.

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.

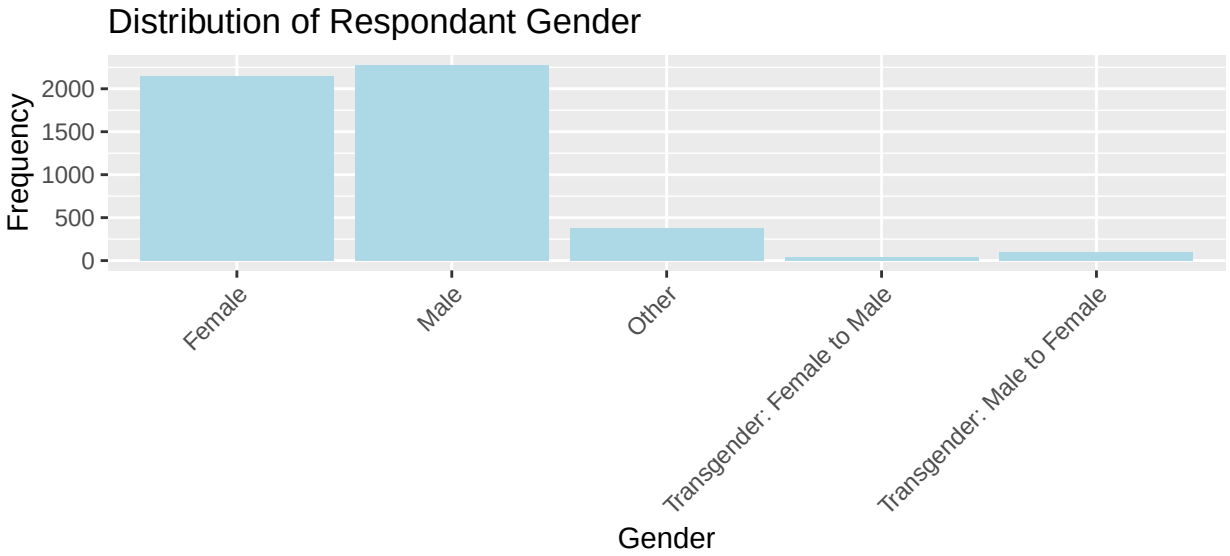
Since respondents could pick multiple selections for racial identity, we synthesized multiple selections into a single variable. If respondents selected multiple races, they would be listed as multiracial. In this way, we could create data visualizations showing an identifiable number of groups to identify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

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.

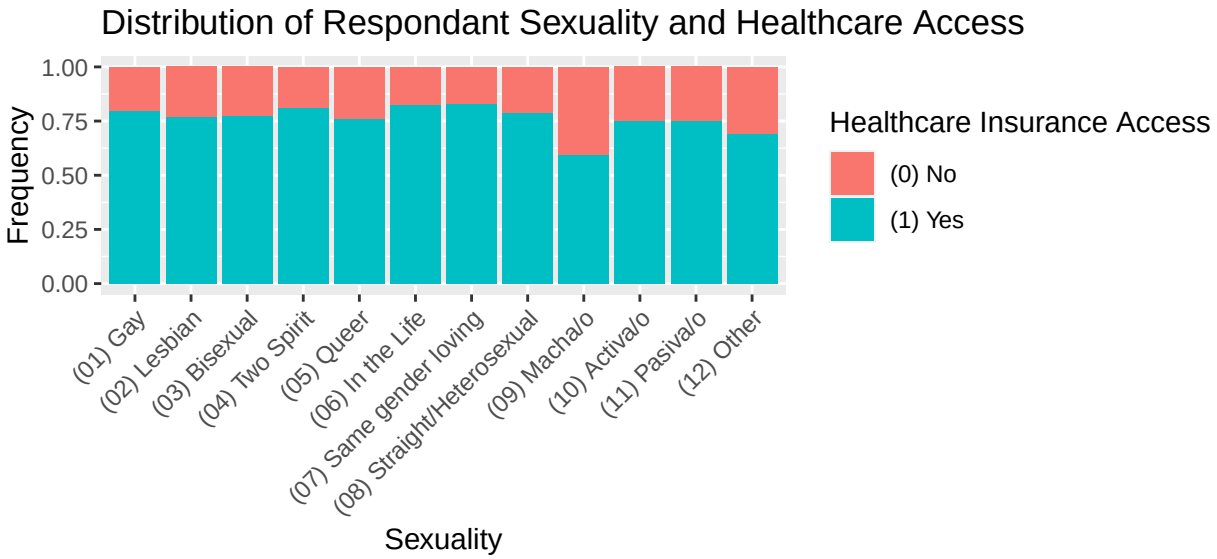
Preliminary Data Visualizations

We visualized distribution of respondant gender identities of participants with a bar graph. 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.



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?"

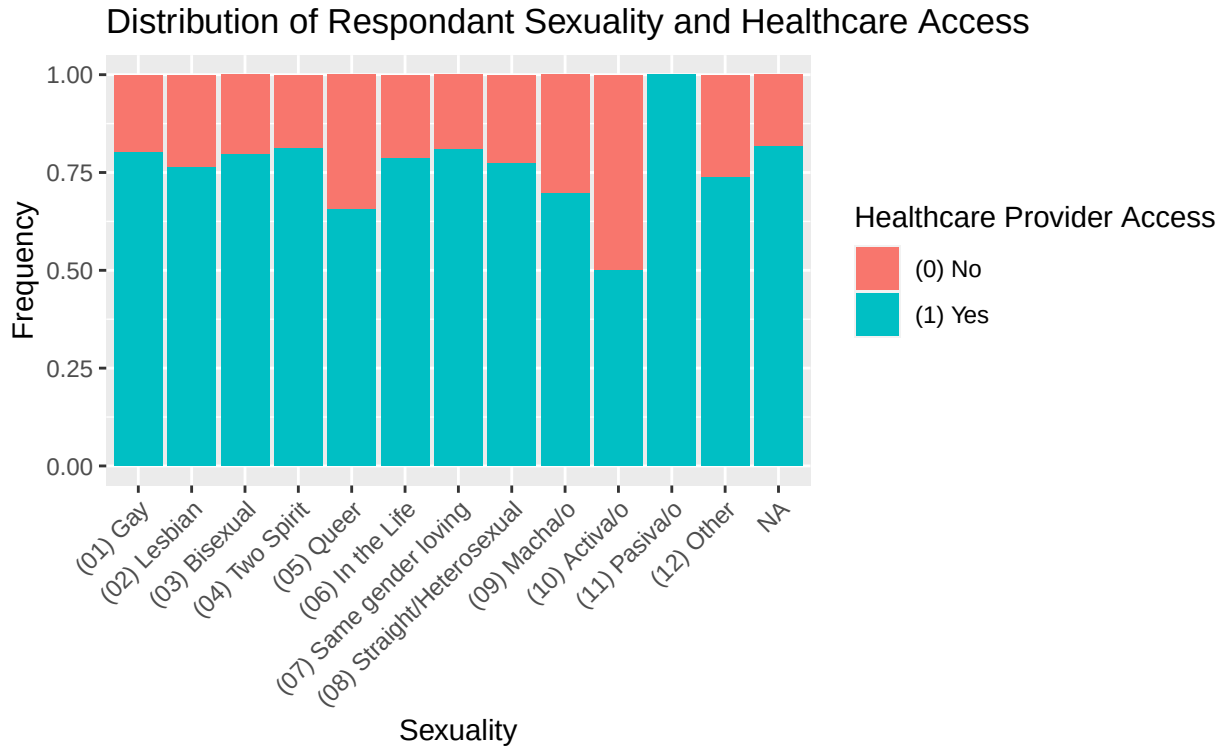
In the second visualization of sexuality and access to healthcare (pictured below), we visualized the with NA values removed. This helps us more clearly visualize the data. We can see more clearly that the group with least health insurance access is those identifying as macha/o.

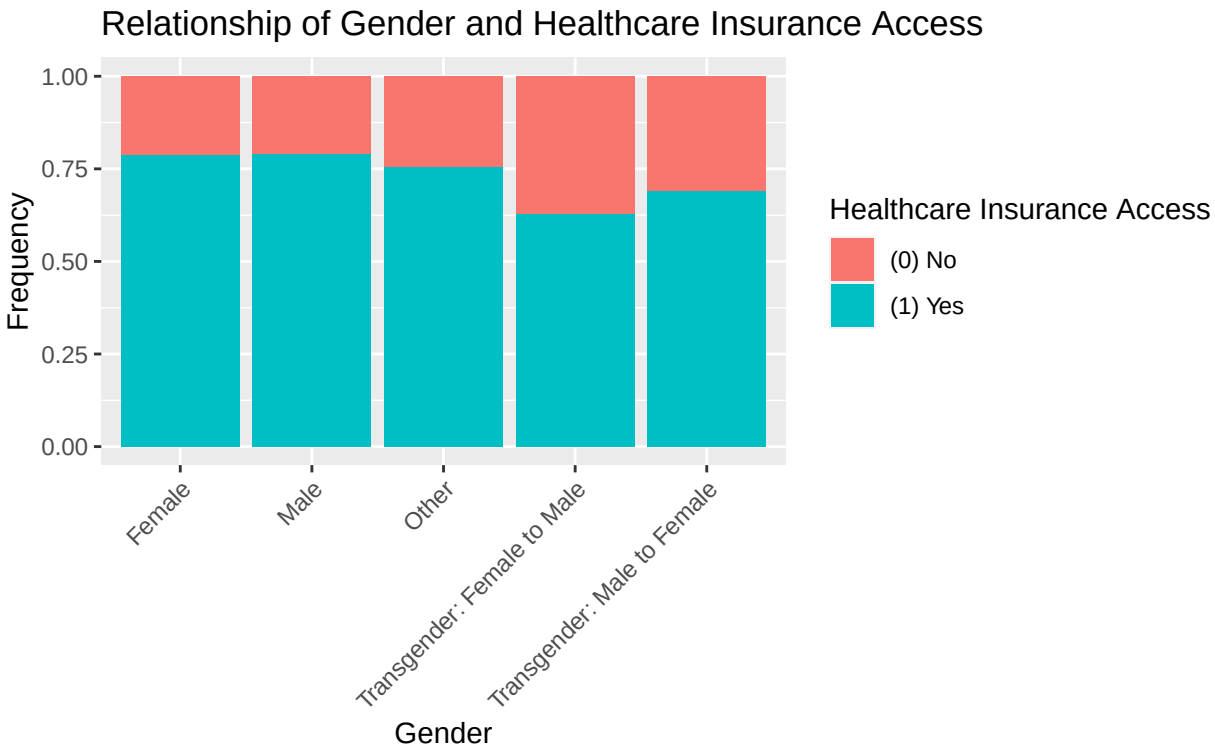


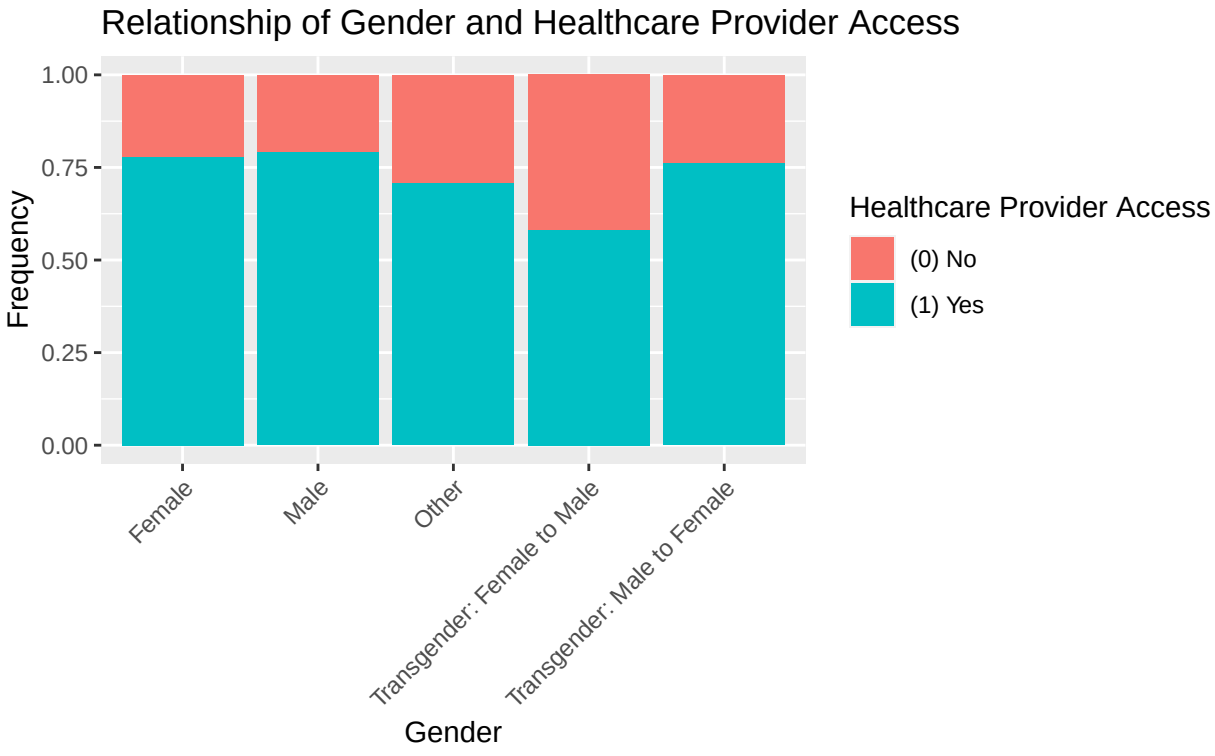
Further Data Visualizations

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.




Relationship of Gender and Healthcare Insurance Access
1.00
0.75
0.50
0.25
0.00
Frequency
Female
Male
Other
Transgender: Female to Male
Transgender: Male to Female
Gender
Healthcare Insurance Access
(0) No
(1) Yes


Relationship of Gender and Healthcare Provider Access
1.00
0.75
0.50
0.25
0.00
Frequency
Female
Male
Other
Transgender: Female to Male
Transgender: Male to Female
Gender
Healthcare Provider Access
(0) No
(1) Yes

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 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
```

According to the overall ANOVA test conducted between sexuality and self-reported health outcomes, there are 2 groups that are shown to be statistically different than the other sexuality groups. A stepdown ANOVA was conducted to see which groups were statistically different 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 a statistically different self-reported health outcome score from other sexuality group. T-tests were conducted in order to determine how Two-Spirit and Queer groups differ in comparison to Straight groups in self-reported healthcare outcomes.

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

```
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")
```

```
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>
```

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

```
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
```

```
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
```