Project Proposal

due October 11, 2021 by 11:59 PM

Gi Chun, Jay Nolt, Ashley Bae - Team Jackie Fan Club

10/15/2021

Load Packages

```
library(tidyverse)
library(readxl)
```

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

Introduction and Data, including Research Questions

Our general research question is: "What is the correlation between specific identity factors and health access?" This is the question we're posing to start our project, though as we proceed, we would have more clarity on what specific variables to analyze statistically significant relationships. We chose this question because we want to highlight interesting trends that may arise, especially in examining intersectional inequality between historically marginalized identities within race, sexuality, and class. We are motivated to ask our question from backgrounds of interest in health, racial justice, and wanting to examine US-based data. 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 negative health outcomes as opposed to someone who's either black or queer.

We recognize that those in LGBT communities face worse health outcomes than their straight/cis counterparts, which the CDC recognizes as well as American Progress, who cites this outcome as affected by "particularly their reduced access to employer-provided health insurance, the social stigma that exists against LGBT people, and a lack of cultural competence in the health care system" (https://www.americanprogress.org/issues/lgbtq-rights/reports/2009/12/21/7048/how-to-close-the-lgbt-health-disparities-gap/). There are also many studies that examine disparity in POC's access to healthcare, so we wanted to examine possibilities in the intersection of race and sexuality. 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 dataset 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 dataset particularly focuses on questions relating to LGBT communities and people of color.

Glimpse

(Please use `glimpse` for your data, uploaded into the data folder, here.)

```
dataFiltered <- data %>%
  select(Q15B, Q15C, Q17A, Q17B, Q18A1: Q18A5, Q18C, Q18G, Q18J, Q19A1: Q19A7, Q22A, Q22B, Q25)
  glimpse(dataFiltered)
```

```
## Rows: 4,953
## Columns: 22
## $ Q15B  <fct> (4) ., (2) ., (3) ., (1) Never, (3) ., (4) ., (4) ., NA, NA, (3)~
## $ Q15C  <fct> (4) ., (5) ., (4) ., (2) ., (2) ., (6) Always, (5) ., (6) Always~
## $ Q17A  <fct> (0) No, (1) Yes, (1) Yes, (1) Yes, (1) Yes, (0) No, (1) Yes, (1)~
## $ Q17B  <fct> (0) No, (0) No, (1) Yes, (1) Yes, (1) Yes, (0) No, (1) Yes, (1) ~
## $ Q18A1 <fct> NA, (1) Yes, (1) Yes, NA, (0) No, (0) No, (0) No, (1) Yes, NA, N~
## $ Q18A2 <fct> (1) Yes, (1) Yes, NA, (1) Yes, (1) Yes, (0) No, (0) No, NA, (1) ~
## $ Q18A3 <fct> NA, (0) No, NA, NA, (0) No, (1) Yes, (0) No, NA, NA, NA, NA, (0)~
## $ Q18A4 <fct> NA, (0) No, NA, NA, (0) No, (0) No, (0) No, NA, NA, NA, NA, (0) ~
## $ Q18A5 <fct> NA, (0) No, NA, NA, (0) No, (0) No, (1) Yes, NA, NA, NA, NA, (1)~
## $ Q18C  <fct> (05) Queer, (03) Bisexual, (01) Gay, (02) Lesbian, (02) Lesbian,~
## $ Q18G  <dbl> 1988, 1987, 1959, 1988, 1990, 1971, 1987, 1985, 1990, 1975, 1989~
## $ Q18J  <fct> Blanked for Confidentiality - ICPSR ADDED                    , Bl~
## $ Q19A1 <fct> (1) Yes, (1) Yes, (1) Yes, (1) Yes, (1) Yes, (1) Yes, (1) Yes, (~
## $ Q19A2 <fct> NA, (0) No, NA, NA, (0) No, (0) No, (0) No, (0) No, NA, NA, NA, ~
## $ Q19A3 <fct> NA, (0) No, NA, NA, (0) No, (0) No, (0) No, (0) No, NA, NA, NA, ~
## $ Q19A4 <fct> NA, (0) No, NA, NA, (0) No, (0) No, (0) No, (0) No, NA, NA, NA, ~
## $ Q19A5 <fct> NA, (0) No, NA, NA, (0) No, (0) No, (0) No, (0) No, NA, NA, NA, ~
## $ Q19A6 <fct> NA, (0) No, NA, NA, (0) No, (0) No, (0) No, (0) No, NA, NA, NA, ~
## $ Q19A7 <fct> NA, (0) No, NA, NA, (0) No, (0) No, (0) No, (0) No, NA, NA, NA, ~
## $ Q22A  <fct> "(5) Bachelor's Degree", "(3) Some College, no degree", "(7) Gra~
## $ Q22B  <fct> "(01) under $8,500", "(03) $11,000-$13,499", "(10) $50,000-74,99~
## $ Q25   <fct> (3) Good, (5) Excellent, NA, (3) Good, (4) Very good, (3) Good, ~
```

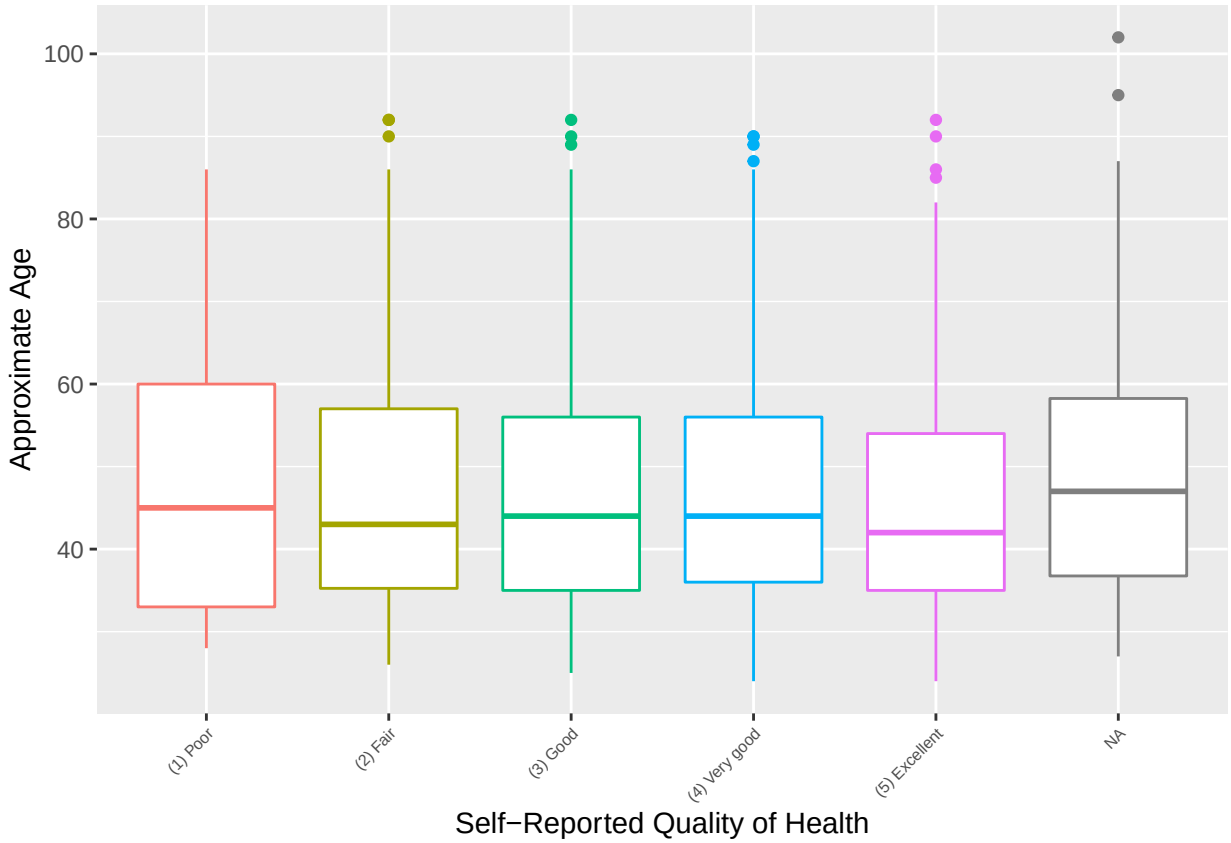
Data Analysis Plan

We want to look at health access outcomes for minority identity groups. Specifically, we want to identify if intersectional groups experience worse health outcomes than those who have one minority status. Thus, our outcome variables will include access to health insurance, visits to a regular doctor, and personal health assessment. Our predictors will be race, gender, sexuality. Our comparison groups will be age, education level, and household income. We may choose to specify our scope of analysis to cover fewer variables as we begin to analyze data since we want to focus on highlighting interesting trends.

Included below is a preliminary data visualization illustrating the types of data that we was collected. This is a series of boxplots, each one representing a group of respondents who reported their personal assessment of their own health. The boxplot summarizes the spread of approximate age. In this case, the general similarities between these groups means this dataset doesn't show evidence of discrimination of health outcomes by age. Our later analyses will look further into unequal access to medical care. This plot also includes the missing data which skews slightly more to older respondents which is a factor to consider when doing any sort of data

analysis or drawing conclusions.

```
## Warning: Removed 206 rows containing non-finite values (stat_boxplot).
```



The essential goal for our project is to understand the role of intersectional identities to healthcare access. For example, how does a white, cis, straight man's access to healthcare differ from a white, cis, homosexual man's access, a black, cis, straight, man, and so on. We may construct a segmented bar plot of access to health insurance by sexuality, faceted by race. In this way, we can see the trends within different queer minority groups as well as how those trends differ for different racial groups. We can also introduce a socioeconomic component such as income level to show how health access is affected by household income in addition to the identity factors mentioned above. As we run statistical tests and data visualizations, we hope to narrow the scope of our focus to key intersectional groups which help defend or refute our central hypothesis.
Statistical methods that we have learned so far that will be useful to answer our questions are t-tests and probability calculations. We expect to apply more statistical methods to our project as we learn more throughout class. We'd also conduct probability tests that would indicate non-independence of identity variables to prove our hypothesis to be true. We are likely to also use a Chi Square and two-way ANOVA test (as we understand them now) as more applicable statistical methods to the kind of data available in our dataset in comparison to the t-test. Overall, to support our hypothesis, the results that we would need include: significant p-values from t-tests and non-independence from our probability tests. We can use these tests to understand the variance in means across the different populations we examine. In our results, a lower p-value than the alpha that we set would deem statistical significance. For an ANOVA test, we would be able to deem the difference as statistically significant if the F statistic is higher than the critical value, which corresponds to the alpha, which is often .05.

Notes of findings: Questions we chose to tentatively analyze: 15b - feeling uncomfortable in queer space bc of race 15c feeling uncomfortable in racial space bc of queer identity 17a - health insurance access 17b - healthcare provider access 18a - gender identity 18c - sexuaity 18g - age 18j - country of birth 19 - race 22a - educational attainment 22b - household income 25 - personal assessment of health