---
title: “Project Proposal”
subtitle: “due October 11, 2021 by 11:59 PM”
author: “Grace Lee and Ji Yun Hyo, TBD3”
date: “October 11th, 2021”
output: pdf_document

Load Packages

Load Data

```
data <- read.csv(file = '../data/COVID_raw_12.8.csv')
tidy_data <- select(data, c('Participant_ID', 'age',"usres", "state", "race", "sex", "localsip", "locals
```

Introduction and Data, including Research Questions

In response to the COVID-19 pandemic, 42 states and territories issued mandatory stay-at-home orders between March 1 to May 31, 2020, affecting 2,355 out of 3,233 (73%) U.S. counties (Moreland et al., 2020). These stay-at-home orders and social distancing policies reduced both population movement as well as person-to-person contact, which slowed the spread of COVID-19 and prevented an even more significant loss of life. In a study published by Cambridge University Press in May 2020, the total number of infections was projected to reach 287 million in the absence of stay-at-home and social distancing policies and 188 million with the enforcement of these policies, translating to 1.24 million lives saved (Thunström et al., 2020).

Due to the importance of stay-at-home orders in slowing the spread of COVID in the United States, we want to examine the extent to which different populations obeyed stay-at-home orders or stayed home even in the absence of shelter-in-place policies. To do so, we used the dataset, “Associations of Urbanicity and Sociodemographic Characteristics with Protective Health Behaviors and Reasons for Leaving the Home during COVID-19,” found on the Harvard Dataverse. The data was collected through a 15-minute online questionnaire of U.S. adults (N = 2,441). Participants were recruited through social media platforms such as Twitter, Instagram, and Facebook. The dataset had 66 variables corresponding to the questions on the online questionnaire, and we chose to focus on the survey responses pertaining to (1) age, (2) country of residence, (3) state of residence, (4) race, (5) sex, (6) if local stay-at-home orders existed, (7) if the participant stayed home as much as possible even if no stay-at-home order existed, (8) if the participant stayed home as much as possible even if they didn’t know if a stay-at-home order existed, (9) if/how the participant protected themselves when leaving home, (10) reasons for leaving home during the stay-at-home order, (11) average hours per day spent at home during pandemic, (12) if the participant had ever contracted COVID, (13) if anyone in the participant’s household had ever contracted COVID, (14) if any close friends had contracted COVID, (15) if the participant lived in an urban, suburban, or rural area, (16) whether the participant had been tested for COVID, (17) educational attainment, and (18) annual income. Each participant/observation was identified by a unique participant ID.

References: Centers for Disease Control and Prevention. (2020, September 3). Timing of state and territorial COVID-19 stay-at-home orders and changes in population movement - United States, March 1–May 31, 2020. Centers for Disease Control and Prevention.

Thunström, L., Newbold, S. C., Finnoff, D., Ashworth, M., & Shogren, J. F. (2020, May 21). The benefits and costs of using social distancing to flatten the curve for covid-19: Journal of Benefit-Cost Analysis. Cambridge Core.

Glimpse

```
glimpse(tidy_data)
```

```
## Rows: 2,441
## Columns: 31
## $ Participant_ID      <int> 1, 2, 3, 4, 5, 6, 7, 9, 10, 11, 12, 13, 14, 15, 16~
## $ age                 <int> 27, 26, 27, 23, 24, 40, 36, 35, 28, 36, 31, 31, 55~
## $ usres               <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
## $ state               <int> 44, 44, 44, 38, 44, 34, 44, 7, 44, 26, 48, 44, 44,~
## $ race                <int> 5, 4, 4, 5, 1, 4, 5, 4, 4, 4, 4, 4, 4, 1, 6, 4, 4,~
## $ sex                 <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 1, 2, 1,~
## $ localsip            <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1,~
## $ localsip2           <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
## $ localsip3           <int> NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA, NA~
## $ leavehomeact___1    <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 1, 1,~
## $ leavehomeact___2    <int> 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 0, 1, 1, 0, 0, 1, 0,~
## $ leavehomeact___3    <int> 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ leavehomeact___4    <int> 1, 1, 1, 0, 1, 1, 0, 0, 1, 1, 0, 0, 0, 1, 1, 1, 1,~
## $ leavehomeact___5    <int> 1, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0, 0,~
## $ leavehomeact___6    <int> 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1,~
## $ leavehomeact___7    <int> 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,~
## $ leavehomereason___1 <int> 1, 0, 0, 1, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 1, 0, 0,~
## $ leavehomereason___2 <int> 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0, 0,~
## $ leavehomereason___3 <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1, 0, 1,~
## $ leavehomereason___4 <int> 0, 1, 1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1,~
## $ leavehomereason___5 <int> 0, 0, 0, 1, 0, 1, 1, 0, 1, 0, 0, 1, 0, 0, 0, 1, 0,~
## $ leavehomereason___6 <int> 0, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0, 1,~
## $ leavehomereason___7 <int> 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1, 0,~
## $ localsiphours       <int> 14, 23, 24, 14, 24, 24, 23, 24, 24, 22, 24, 20, 22~
## $ covidsick           <int> 2, 2, 2, 2, 2, 2, 3, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
## $ hhcovidsick         <int> 2, 2, 2, 2, 2, 2, 3, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
## $ ffcovidsick         <int> 2, 3, 1, 2, 3, 1, 1, 4, 3, 4, 2, 2, 4, 1, 4, 2, 2,~
## $ Classification      <fct> Urban, Urban, Suburban, Rural, Urban, Rural, Urban~
## $ covidtest           <int> 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,~
## $ educ                <int> 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 6, 4, 6, 6, 6, 6,~
## $ hhincome            <int> 12, 11, 11, 5, 3, 7, 3, 6, 12, 12, 12, 12, 12, 12,~
```

Data Analysis Plan

(Specify the outcome (response, Y) and predictor (explanatory, X) variables you will use to answer your question, as well as the comparison groups you will use, if applicable. You may include very preliminary exploratory data analysis, including some summary statistics and visualizations, along with some explanation on how they help you learn more about your data. Note the statistical method(s) that you believe will be useful in answering your question(s). What results from these specific statistical methods are needed to support your hypothesized answer?)

```
number_of_hours <- tidy_data %>%
  group_by(race) %>%
 # summarize(localsiphours) %>%
  summarise_at(vars(localsiphours), list(hours = mean),na.rm = TRUE) #to summarize count

number_of_hours_two <- tidy_data %>%
  group_by(Classification) %>%
```

```
 # summarize(localsiphours) %>%
  summarise_at(vars(localsiphours), list(hours = mean),na.rm = TRUE) %>% #to summarize count
  print()
```

```
## # A tibble: 4 x 2
##   Classification hours
##   <fct>          <dbl>
## 1 Rural           21.5
## 2 Suburban        21.3
## 3 Urban           21.1
## 4 <NA>            16.6
```

```
tidy_data$race[is.na(tidy_data$race) == TRUE] <- "6"
tidy_data$Classification[is.na(tidy_data$Classification)== TRUE] <- "Urban"
number_of_hours$Classification[number_of_hours$race == 0] <- "American Indian"
```

```
## Warning: Unknown or uninitialised column: `Classification`.
```

```
number_of_hours$race[number_of_hours$race == 1] <- "Asian"
number_of_hours$race[number_of_hours$race == 2] <- "Native Hawaiian"
number_of_hours$race[number_of_hours$race == 3] <- "Black"
number_of_hours$race[number_of_hours$race == 4] <- "White"
number_of_hours$race[number_of_hours$race == 5] <- "Mixed"
number_of_hours$race[number_of_hours$race == 6] <- "Unknown"

ggplot(data=number_of_hours, aes(x=race, y=hours)) +
  geom_bar(stat="identity") +
  labs (
    y = "Number of Hours Remained at Home",
    x = "Race",
    title = "Number of Hours Remained at Home by Race",
    )
```