

Project Proposal

due October 11, 2021 by 11:59 PM

Grace Lee and Ji Yun Hyo, TBD3

October 11th, 2021

Load Packages

```
library(dplyr)
library(tidyverse)
library(sf)
library(viridis)
library(ggspatial) #for scale annotation
```

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', 'local...
```

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 U.S. counties (CDC, 2020). These stay-at-home policies reduced both population movement and person-to-person contact, which slowed the spread of COVID-19. 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 ask if the average number of hours spent at home differed between different populations. 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) recruited through social media platforms such as Twitter, Instagram, and Facebook. The dataset had 66 variables corresponding to the questionnaire questions, 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 even if no order existed, (8) if the participant stayed home even if they didn’t know if an order existed, (9) how the participant protected themselves when leaving home, (10) reasons for leaving home during the order, (11) average hours per day spent at home during the pandemic, (12) if the participant had contracted COVID, (13) if anyone in the household had 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.

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,~
## $ 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, 1, 6, 4, 4,~
## $ sex                  <int> 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,~
## $ 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,~
## $ leavehomeact___2     <int> 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 0, 1, 1, 0, 0, 1,~
## $ leavehomeact___3     <int> 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,~
## $ leavehomeact___5     <int> 1, 1, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 0, 1, 1, 0,~
## $ leavehomeact___6     <int> 1, 1, 1, 0, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0,~
## $ leavehomeact___7     <int> 0, 0, 0, 0, 1, 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,~
## $ leavehomereason___2  <int> 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 1, 0,~
## $ leavehomereason___3  <int> 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 1, 0, 1, 1, 0, 1,~
## $ leavehomereason___4  <int> 0, 1, 1, 0, 0, 0, 0, 0, 1, 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,~
## $ leavehomereason___6  <int> 0, 1, 0, 0, 0, 0, 0, 1, 0, 1, 0, 0, 1, 0, 0, 0,~
## $ leavehomereason___7  <int> 0, 0, 0, 0, 0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 1,~
## $ 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,~
## $ hhcovidsick          <int> 2, 2, 2, 2, 2, 2, 3, 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,~
## $ Classification       <chr> "Urban", "Urban", "Suburban", "Rural", "Urban", "R~
## $ covidtest            <int> 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,~
## $ hhincome             <int> 12, 11, 11, 5, 3, 7, 3, 6, 12, 12, 12, 12, 12, 12,~
```

Data Analysis Plan

In order to explore the relationships between certain demographic characteristics and hours stayed at home during the pandemic, we will conduct multiple two-sample t-tests comparing the mean number of hours spent at home during the pandemic between different sexes, races (e.g. Asian vs. White), levels of education, and income levels, among others. We also plan on constructing confidence intervals regarding the number of hours spent at home for populations with and without formal stay-at-home orders. In addition, we plan on examining the most frequently cited methods of protection from COVID used when in public as well as reasons for leaving home during a stay-in-place order.

At present, we hypothesize that Asian people will have significantly greater average hours spent at home than White people as they had 0.7 times the hospitalization rate of white people (CDC, 2020). We also hypothesize that groups with income over \$150,000 will have greater average hours spent at home due to essential workers, whose incomes are lower on average, being required to leave home for work. To achieve these results, we would need a significant p-value of under 0.05 from our t-test. The table and graph below give us a preliminary idea of the differences in average hours spent home among different populations such as

urban vs. rural and between different races.

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.

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

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

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

```
tidy_data$Classification[is.na(tidy_data$Classification)== TRUE] <- "Urban"
```

```
reasons <-c("Work", "Provide Care for Others", "Grocery Shopping", "Essential Shopping", "Exercise", "Walk Dog", "Other")
freq_reasons <- c(sum(tidy_data$leavehomereason___1, na.rm = TRUE), sum(tidy_data$leavehomereason___2, na.rm = TRUE), sum(tidy_data$leavehomereason___3, na.rm = TRUE), sum(tidy_data$leavehomereason___4, na.rm = TRUE), sum(tidy_data$leavehomereason___5, na.rm = TRUE), sum(tidy_data$leavehomereason___6, na.rm = TRUE), sum(tidy_data$leavehomereason___7, na.rm = TRUE))
reasons_for_leaving = data.frame(reasons, freq_reasons)
```

```
print(reasons_for_leaving)
```

```
##           reasons freq_reasons
## 1           Work           543
## 2 Provide Care for Others       215
## 3   Grocery Shopping      1643
## 4   Essential Shopping       811
## 5           Exercise      1286
## 6           Walk Dog        839
## 7              Other        257
```

```
pie_chart <- tidy_data %>%
  group_by(race) %>%
  count() %>%
  ungroup() %>%
  mutate(perc = n / sum(n))
```

```
tidy_data$race[tidy_data$race == 0] <- "American Indian"
tidy_data$race[tidy_data$race == 1] <- "Asian"
tidy_data$race[tidy_data$race == 2] <- "Native Hawaiian"
tidy_data$race[tidy_data$race == 3] <- "African American"
tidy_data$race[tidy_data$race == 4] <- "White"
tidy_data$race[tidy_data$race == 5] <- "Mixed"
```

```

tidy_data$race[tidy_data$race == 6] <- "Unknown"

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

```

Warning: Removed 1 rows containing missing values (position_stack).

