

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...
tidy_data$race[is.na(tidy_data$race) == TRUE] <- "6"
number_of_hours <- tidy_data %>%
  group_by(race) %>%
  # summarize(localsiphours) %>%
  summarise_at(vars(localsiphours), list(hours = mean), na.rm = TRUE) %>% #to summarize count
print()
```

```
## # A tibble: 7 x 2
##   race  hours
##   <chr> <dbl>
## 1 0      21.8
## 2 1      21.9
## 3 2      21.3
## 4 3      21.7
## 5 4      21.2
## 6 5      21.2
## 7 6      20.3
```

Introduction and Data, including Research Questions

(The introduction should introduce your general research question and your data (where it came from, how it was collected, what are the cases, what are the variables, etc.). Your research questions should be clearly specified. The motivation for your research question should be clear, with citations to relevant literature as appropriate.)

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                <chr> "5", "4", "4", "5", "1", "4", "5", "4", "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      <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,~
## $ 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,~
```

```
summary(tidy_data)
```

```
## Participant_ID      age          usres          state          race
## Min.   : 1      Min.   :18.00      Min.   :1      Min.   : 1.00      Length:2441
## 1st Qu.: 673    1st Qu.:32.00    1st Qu.:1      1st Qu.:37.00      Class :character
## Median :1371    Median :40.00    Median :1      Median :44.00      Mode  :character
## Mean   :1377    Mean   :41.91    Mean   :1      Mean   :36.97
## 3rd Qu.:2082    3rd Qu.:51.00    3rd Qu.:1      3rd Qu.:44.00
## Max.   :2789    Max.   :86.00    Max.   :1      Max.   :51.00
##
##                NA's   :266
## sex          localsip    localsip2    localsip3
## Min.   :1.000      Min.   :1.000      Min.   :1.000      Min.   :1.000
## 1st Qu.:1.000      1st Qu.:1.000      1st Qu.:1.000      1st Qu.:1.000
## Median :2.000      Median :1.000      Median :1.000      Median :1.000
## Mean   :1.674      Mean   :1.117      Mean   :1.063      Mean   :1.042
## 3rd Qu.:2.000      3rd Qu.:1.000      3rd Qu.:1.000      3rd Qu.:1.000
## Max.   :2.000      Max.   :7.000      Max.   :2.000      Max.   :2.000
```

```

## NA's :248 NA's :478 NA's :2362 NA's :2417
## leavehomeact___1 leavehomeact___2 leavehomeact___3 leavehomeact___4
## Min. :0.0000 Min. :0.0000 Min. :0.000 Min. :0.0000
## 1st Qu.:1.0000 1st Qu.:0.0000 1st Qu.:0.000 1st Qu.:0.0000
## Median :1.0000 Median :1.0000 Median :0.000 Median :1.0000
## Mean :0.7841 Mean :0.5449 Mean :0.186 Mean :0.6436
## 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:0.000 3rd Qu.:1.0000
## Max. :1.0000 Max. :1.0000 Max. :1.000 Max. :1.0000
##
## leavehomeact___5 leavehomeact___6 leavehomeact___7 leavehomereason___1
## Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. :0.0000
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000
## Median :0.0000 Median :1.0000 Median :0.0000 Median :0.0000
## Mean :0.4441 Mean :0.7395 Mean :0.1266 Mean :0.2962
## 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:0.0000 3rd Qu.:1.0000
## Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :1.0000
## NA's :608
## leavehomereason___2 leavehomereason___3 leavehomereason___4
## Min. :0.00000 Min. :0.0000 Min. :0.0000
## 1st Qu.:0.00000 1st Qu.:0.0000 1st Qu.:0.0000
## Median :0.00000 Median :1.0000 Median :0.0000
## Mean :0.08808 Mean :0.6731 Mean :0.3322
## 3rd Qu.:0.00000 3rd Qu.:1.0000 3rd Qu.:1.0000
## Max. :1.00000 Max. :1.0000 Max. :1.0000
##
## leavehomereason___5 leavehomereason___6 leavehomereason___7 localsiphours
## Min. :0.0000 Min. :0.0000 Min. :0.0000 Min. : 0.00
## 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.:0.0000 1st Qu.: 21.00
## Median :1.0000 Median :0.0000 Median :0.0000 Median : 23.00
## Mean :0.5268 Mean :0.3437 Mean :0.1053 Mean : 21.19
## 3rd Qu.:1.0000 3rd Qu.:1.0000 3rd Qu.:0.0000 3rd Qu.: 23.00
## Max. :1.0000 Max. :1.0000 Max. :1.0000 Max. :528.00
## NA's :563
## covidssick hhcovidssick ffcovidssick Classification
## Min. :1.000 Min. :1.000 Min. :1.000 Rural :539
## 1st Qu.:2.000 1st Qu.:2.000 1st Qu.:2.000 Suburban:694
## Median :2.000 Median :2.000 Median :2.000 Urban :895
## Mean :2.113 Mean :2.065 Mean :2.221 NA's :313
## 3rd Qu.:2.000 3rd Qu.:2.000 3rd Qu.:2.000
## Max. :3.000 Max. :3.000 Max. :4.000
## NA's :244 NA's :247 NA's :247
## covidtest educ hhincome
## Min. :1.000 Min. :3.000 Min. : 1.000
## 1st Qu.:2.000 1st Qu.:6.000 1st Qu.: 8.000
## Median :2.000 Median :6.000 Median :11.000
## Mean :1.982 Mean :5.812 Mean : 9.848
## 3rd Qu.:2.000 3rd Qu.:6.000 3rd Qu.:12.000
## Max. :2.000 Max. :7.000 Max. :12.000
## NA's :250 NA's :246 NA's :281

```

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$race[number_of_hours$race == 0] <- "American Indian"
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",
  )
```

