

# Final Report

due November 16, 2021 by 11:59 PM

Grace Lee and Jiyun Hyo

November 16, 2021

## 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 (73%) 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 asked if the average number of hours spent at home differed between different populations in order to understand which demographic characteristics were associated with greater hours spent at home. Specifically, we asked if people of different races, income levels, and education levels, among other variables, differed significantly in their mean number of hours spent at home. Upon finding significant results, we did preliminary exploration into why results were significant gusing the responses in our dataset.

To answer these questions, 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 (Burford, 2020). The data was collected between April 15-May 5, 2020 through an anonymous 15-minute self-completed online questionnaire of U.S. adults (N = 2,441). Participants were approached and recruited through social media platforms such as Twitter, Instagram, and Facebook, aged over 18, currently residing in the U.S., and did not include essential service workers, who were excluded due to their need to leave the home for employment (Burford, 2020).

To focus our analysis on hours spent at home and the associated demographic characteristics and reasons for staying home, we chose to focus only on a subset of variables. The dataset had 66 variables corresponding to the questionnaire questions. We chose to focus on the survey responses pertaining to (1) age, (2) country & (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 or (8) even if they didn't know if an order existed, (9) how the participant protected themselves in public, (10) reasons for leaving home during the order, (11) average hours spent at home per day 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.

## Data Analysis Plan

We excluded people who did not respond to race (response = NA) and number of hours spent at home (response = NA) from the dataset. We excluded Hawaiians and those with only grades 9-12 education in our visualizations and analysis (though not from the dataset) due to low sample size, which skewed results. We created new variables for each demographic examined in our analysis. Specifically, we created variables for each race as binaries where the value of the variable was “1” if the participant was the race denoted by the variable name (e.g. “Asian”) and “0” otherwise: for example, Asian people were denoted “1” in the variable, “Asian,” and non-Asian people were denoted as “0.” We created similar variables for white, African American, Native American, and Mixed. We did the same for each level of education and level of income.

In order to explore the relationship between certain demographic characteristics (predictor variables) and hours stayed at home during the pandemic (response variable), we conducted multiple two-sample t-tests comparing the mean number of hours spent at home during the pandemic between different races (e.g. Asian vs. non-Asian), levels of education, income levels, and sex. We used two-sample t-tests as the population parameters are unknown, and the groups being compared are sufficiently different and independent from each other that they could not be paired. We constructed 95% confidence intervals regarding the number of hours spent at home for different races, levels of education, income levels, and sex.

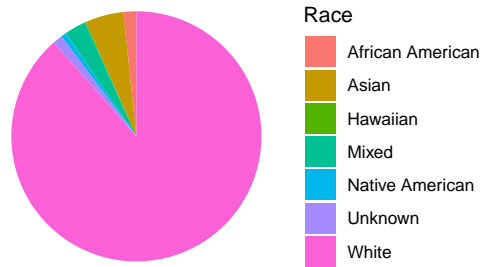
We conducted a linear regression to see if age had any significant association with the number of hours spent at home. Upon finding significant differences between sexes in the number of hours spent at home, we conducted a logistic regression to determine the differences between sexes in odds ratios of reasons for leaving the home, including grocery shopping, exercise, and work in order to examine which factors led to a significant difference between sexes in number of hours spent at home. We conducted these tests while setting  $\alpha = 0.05$ .

## Visualizations

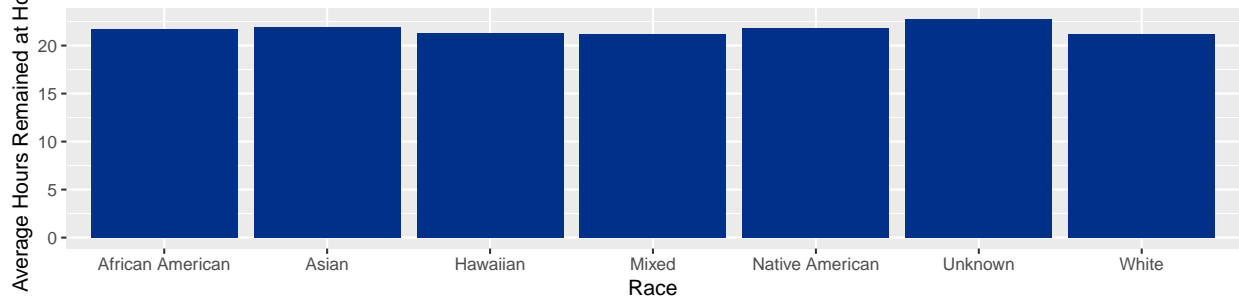
To understand the population, we visualized the racial composition of the sample and the associated mean number of hours spent at home with each race. We found that 88.5% of the sample was composed of Hispanic non-white individuals, with the next most represented racial groups being Asians (4.93%) followed by African Americans (1.66%).

We also visualized the most cited reasons for leaving home during a stay-in-place order. We found that the most commonly cited reason for leaving home was grocery shopping followed by exercise, walking the dog, and other essential shopping.

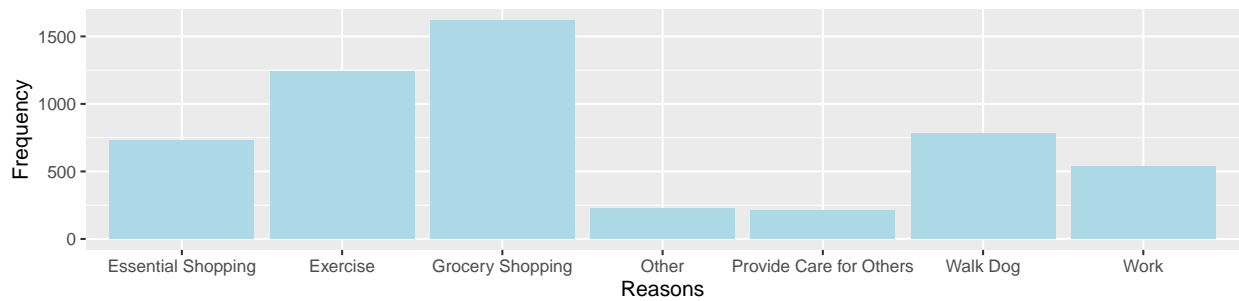
Sample Composition by Race



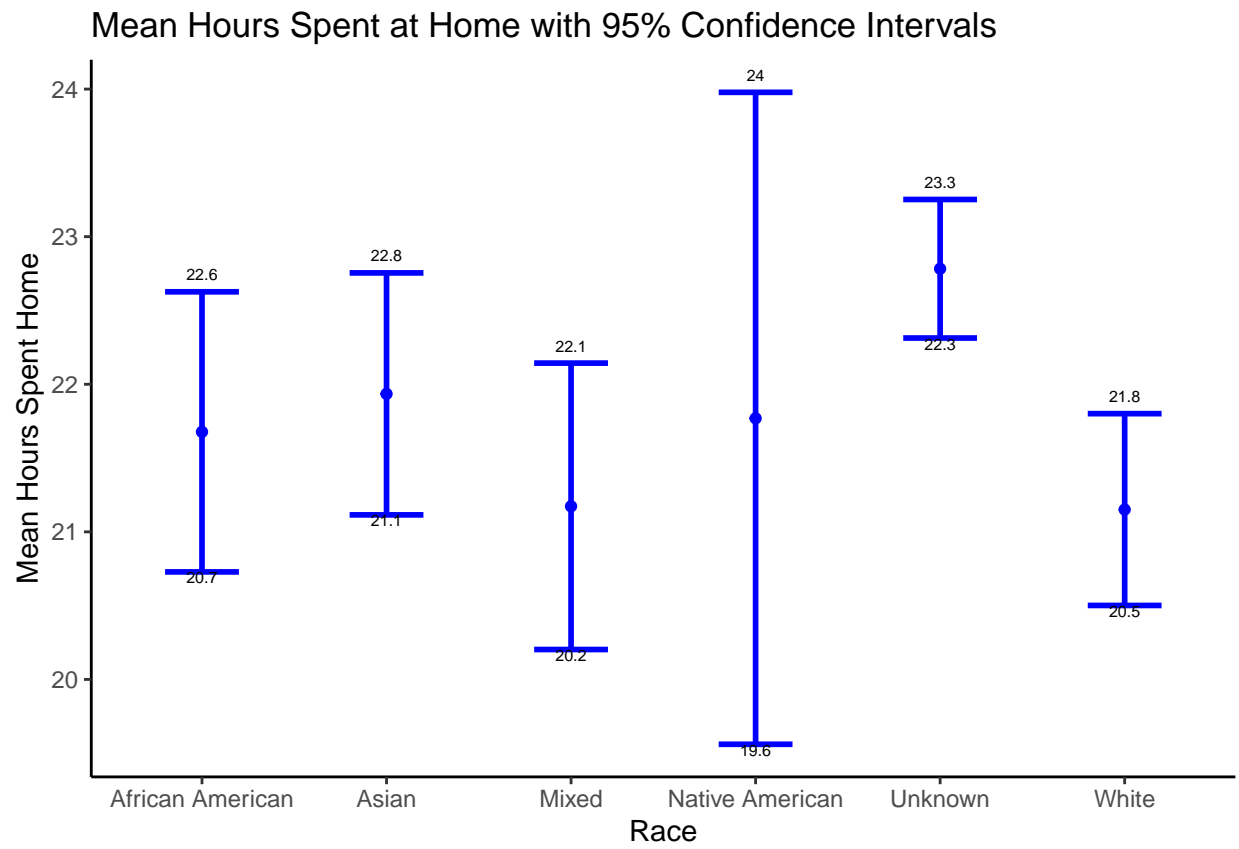
Number of Hours Remained at Home by Race

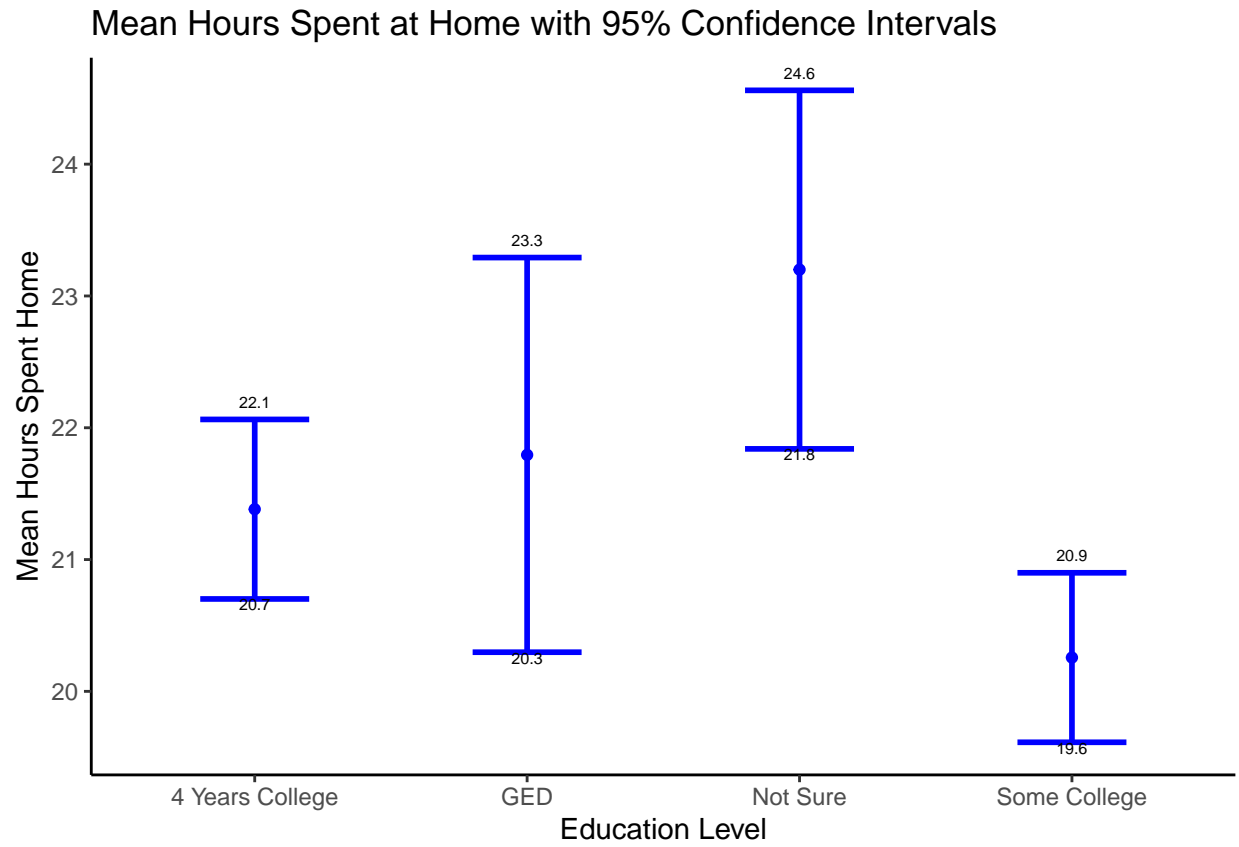


Reasons for Leaving the House

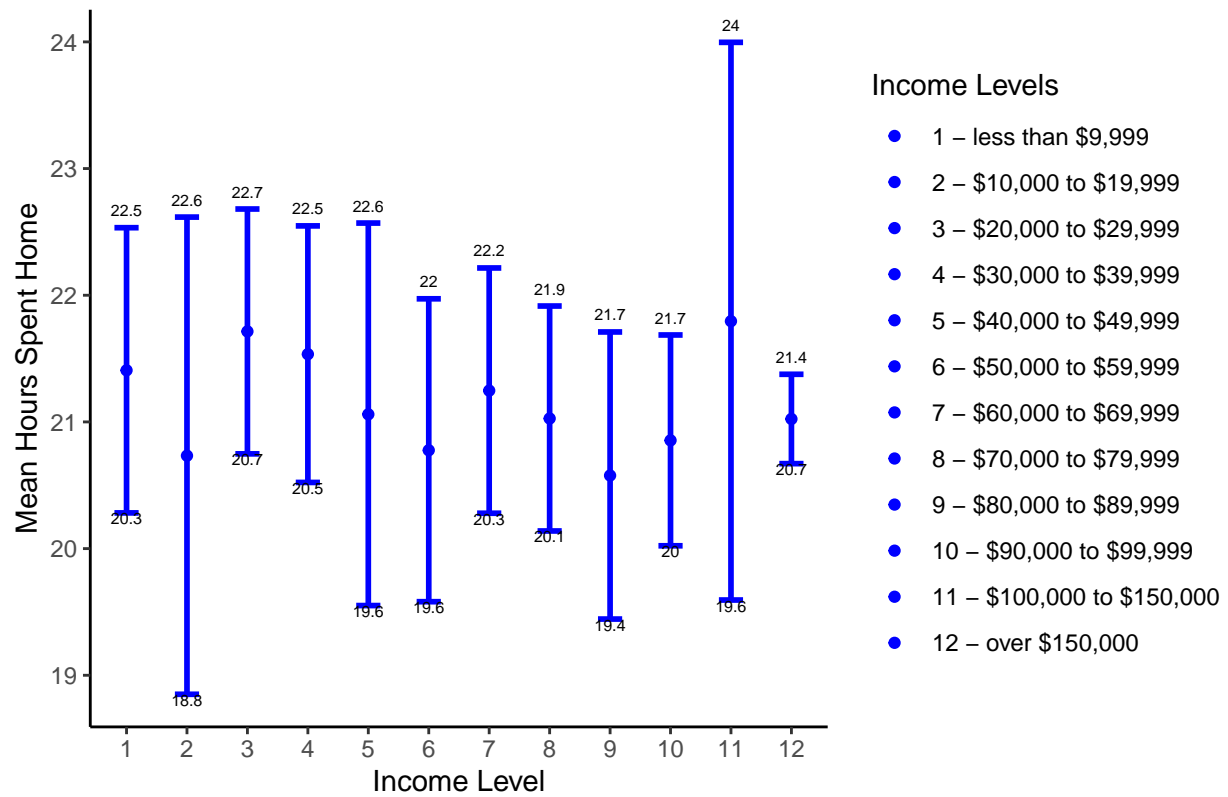


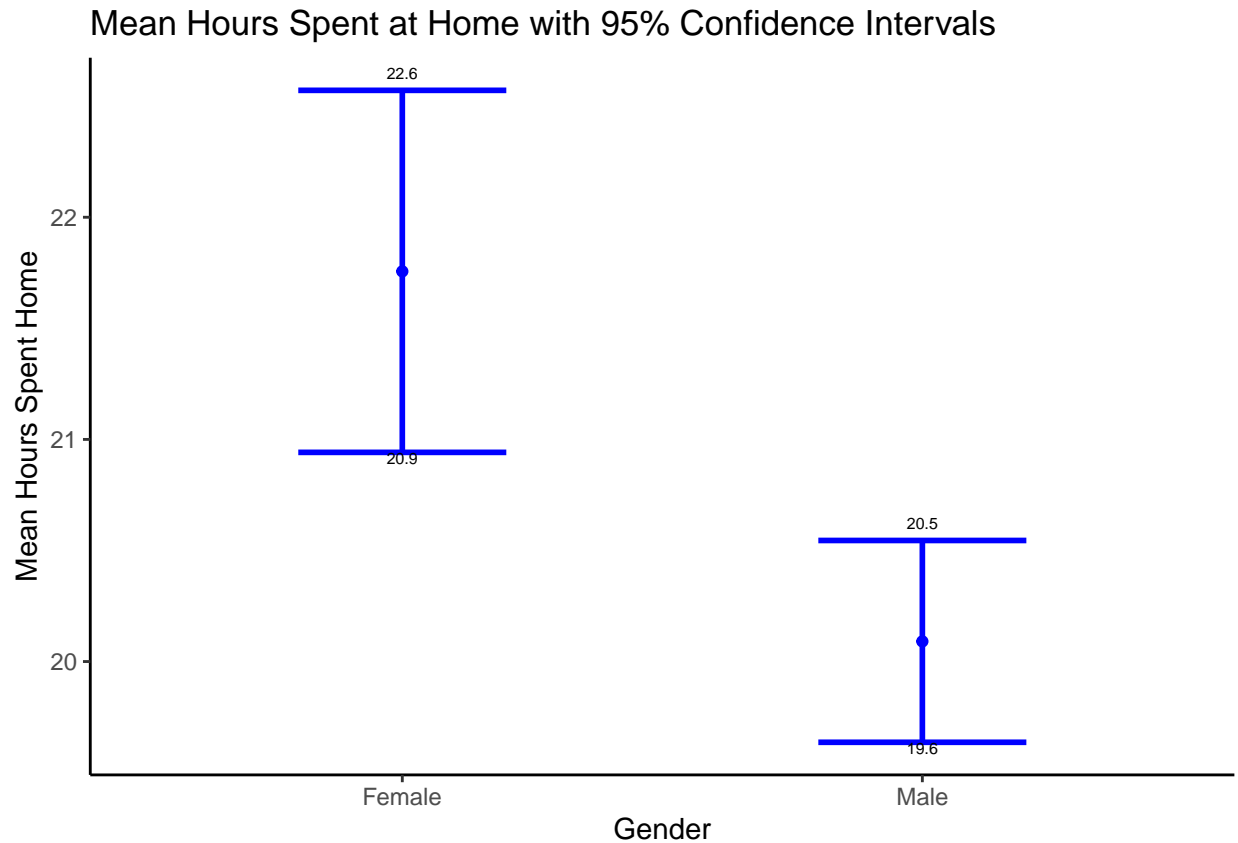
We visualized 95% confidence intervals of the mean number of hours spent at home to see if there were any immediately noticeable differences between demographic groups. Specifically, we visualized confidence intervals for different levels of education and income.





Mean Hours Spent at Home with 95% Confidence Intervals





## Results

We found no significant differences between races in terms of the mean hours spent at home. To test this, we conducted multiple two-sample t-tests comparing each race against the rest of the population, for example, White compared to non-White. No one race had a significantly different mean number of hours spent home compared to the rest of the population.

In our linear regression, we did not find that age significantly impacted the mean number of hours spent at home.

We also found significant differences in the mean number of hours spent at home between people who had attended a 4-year college and completed their degree and those who had not. WRITE THE 95% CONFIDENCE INTERVAL FOR IT HERE. ALSO CONSTRUCT ONE AS A VISUALIZATION?

We found significant differences in the mean number of hours spent at home between the sexes, finding that on average, males spent less time at home than females did with a 95% confidence interval of  $[-2.598, -0.733]$  and rejecting the null hypothesis that the mean number of hours spent is the same between the two populations. To examine for which reasons males spent less time at home than females, we conducted a logistic regression for each reason examined in the study. We found that, compared to males, females had a odds ratio of exercising, grocery shopping, and working (95% CI: ).

## Conclusion, Limitations, and Future Directions

In examining the relationship between mean hours spent at home and certain demographic characteristics, we found that there were no significant differences in the number of hours spent at home between different races. We found our results surprising given that the COVID cases and hospitalization rates differ between races. We had hypothesized that Asian people would differ in their mean hours spent at home compared to the rest of the population, as they had 0.7 times the case rate of white people, who comprised nearly 90% of the sample size (CDC, 2020).

We found no significant differences in mean hours spent at home comparing through two-sample t-tests comparing those at the highest level of income compared to the rest of the population, which we found surprising given that we had hypothesized that people with incomes over \$150,000 would have different/greater average hours spent at home due to having access to grocery and meal delivery services such as InstaCart, which would decrease the number of hours needed to be spent outside. This is further supported by the data, where grocery shopping was the highest cited reason for leaving the home and by reports that meal delivery services had increased by approximately 70% in March 2020 (Hobbs, 2020).

We also found significant differences in the mean number of hours spent at home between those who had completed a four-year college degree and those who had not. Level of education may therefore have an impact on the number of hours

We found significant differences in the mean number of hours spent at home during the pandemic between males and females. We delved deeper into the reasons for why the significant number of hours spent at home differed between males and females. Further analysis showed that males were WRITE ODDS RATIOS This is unsurprising, as men were reported to participate in more hours of exercise and in a greater variety of exercise activities than women, since exercise was one of the most highly cited reasons for leaving the home (Hickey & Mason, 2017).

The sample over-represented Hispanic non-white individuals (88.5% in sample vs. 60.1% in U.S. population) while under-representing other races such as Black and Asian people according to the U.S. Census data. Thus, our results may be skewed due to having small sample sizes for certain demographics.

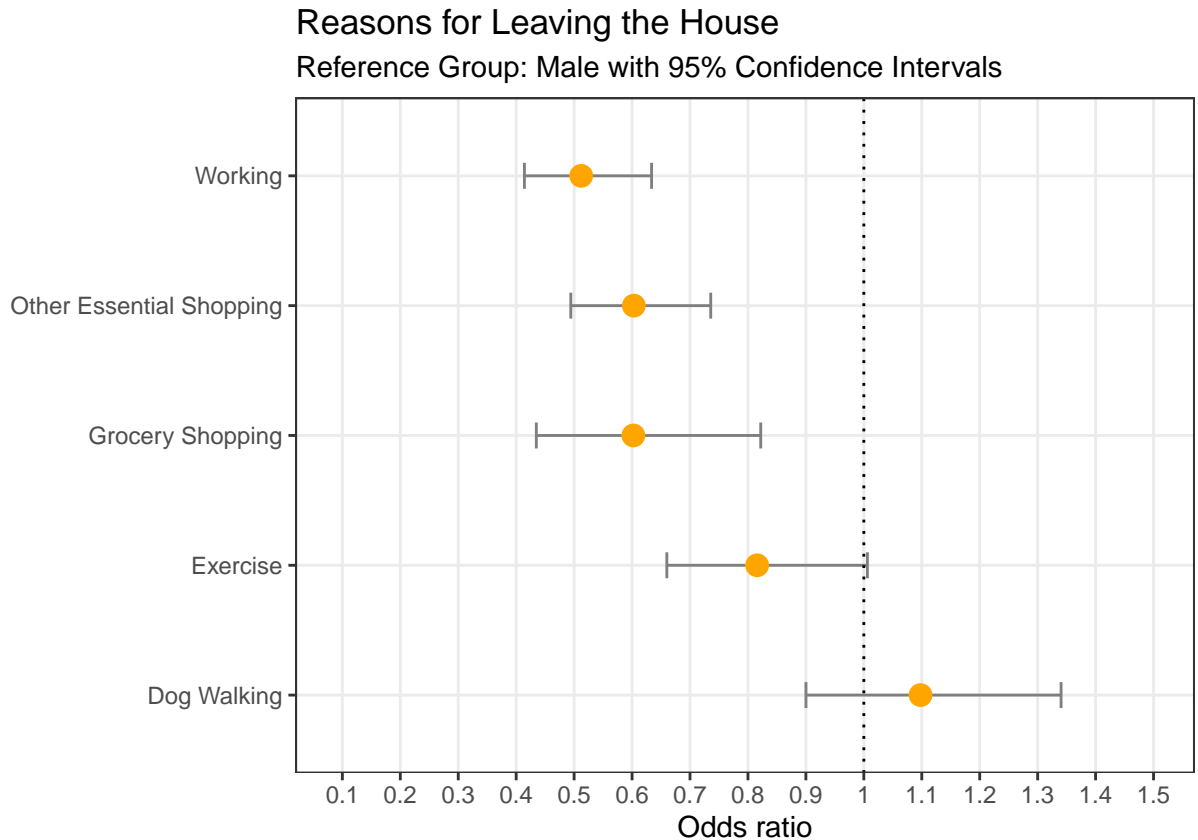
In future analysis, examining data from later periods of quarantine may also yield different results as restlessness and “cabin fever” during quarantine could lead to people leaving home for other, non-essential reasons more often, and it could be possible that different demographics may be disproportionately prone to this behavior due to different living conditions.

```
# grocery shopping
grocery_by_gender <- tidy_data %>%
  filter(is.na(leavehomereason__3) == FALSE)
grocery_by_gender$sex = relevel(grocery_by_gender$sex, ref = "Male")
grocery_by_gender_fit <- logistic_reg() %>%
  set_engine("glm") %>%
  fit(leavehomereason__3 ~ sex, data = grocery_by_gender, family = "binomial")
tidy(grocery_by_gender_fit, conf.int=TRUE, exponentiate = TRUE)
```

```
## # A tibble: 2 x 7
##   term                estimate std.error statistic  p.value conf.low conf.high
##   <chr>                <dbl>    <dbl>    <dbl>    <dbl>    <dbl>    <dbl>
```



## 1 (Intercept)	9.62	0.142	16.0	1.76e-57	7.36	12.8
## 2 sexFemale	0.602	0.162	-3.13	1.77e- 3	0.435	0.822



## References

- Burford, K. G. (2020). Replication data for: Associations of urbanicity and sociodemographic characteristics with protective health behaviors and reasons for leaving the home during COVID-19. Harvard Dataverse. <https://doi.org/10.7910/DVN/7FA07D>
- Bourassa, K. J. (2020). State-level stay-at-home orders and objectively measured movement in the United States during the COVID-19 pandemic. *Psychosomatic Medicine*, 83(4), 358–362. <https://doi.org/10.1097/psy.0000000000000905>
- Centers for Disease Control and Prevention. (n.d.). Risk for COVID-19 infection, hospitalization, and death by Race/Ethnicity. Centers for Disease Control and Prevention. <https://www.cdc.gov/coronavirus/2019-ncov/covid-data/investigations-discovery/hospitalization-death-by-race-ethnicity.html>.
- Hickey, M. E. & Mason, S. E. (2017). Age and gender differences in participation rates, motivators for, and barriers to exercise. *Modern Psychological Studies*, 22(2), 10-19. <https://scholar.utc.edu/mps/vol22/iss2/3>.
- Hobbs, J. E. (2020). Food supply chains during the COVID-19 pandemic. *Canadian Journal of Agricultural Economics/Revue Canadienne D'agroeconomie*, 68(2), 171–176. <https://doi.org/10.1111/1/cjag.12237>

Thunström, L., Newbold, S. C., Finnoff, D., Ashworth, M., & Shogren, J. F. (2020). The benefits and costs of using social distancing to flatten the curve for COVID-19. *Journal of Benefit-Cost Analysis*, 11(2), 179–195. <https://doi.org/10.1017/bca.2020.12>

U.S. Census Bureau quickfacts: United States. (n.d.). <https://www.census.gov/quickfacts/fact/table/US/PST045219>.