Replication of a Research Claim from Kavanagh et al. (2020),
from medRxiv

Replication Team: Andrew Tyner and Nicholas Huntington-Klein

Research Scientist: Nick Fox

Action Editor: Kevin Esterling

Independent Reviewers
(add name below when you initiate review, comment “DONE” on your name when you finish):

Reviewer #1: Nathaniel Porter

Reviewer #2: Jan Philipp Röer

Reviewer #3: Landon Schnabel]

Review Period: October 13 - October 19

View-only links to: Original Paper, Original Materials, Replication Data, Replication Analysis

Privacy Statement: Other teams are making predictions about the outcomes of many different studies, not knowing which studies have been selected for replication. As a consequence, the success of this project requires full confidentiality of this peer review process. This includes privacy about which studies have been selected for replication and all aspects of the discussion about these replication designs.

Instructions for Data Analysts

The preregistration for this replication study was started by a separate team of researchers who were responsible for identifying data sources and constructing them into a replication dataset(s) for your use in the analysis. They have completed sections 1-13 of the preregistration below, and included additional materials in the OSF project that document how the dataset was constructed.

You'll be responsible for filling out sections 16-25 of the preregistration below. Before you do so, **please review the original study, sections 1-15 of the preregistration, and the materials provided on the OSF**, so that you are familiar with all of the decisions that have been made to date. In many cases, the 'data preparer' will have left you instructions and suggestions on how the provided data can be used in the analysis, as well as idiosyncrasies and discrepancies in the data that you should be aware of. The data preparers have tried to be thorough in including all variables that you might need, but please keep in mind the following:

- Some of the variables included in the constructed dataset(s) may not be needed in the final analysis, so please do not feel the need to necessarily use all of the provided variables.
- Some of the variables needed might have mistakenly been excluded from the constructed datasets. If you find that this is the case, please let Andrew know, and he will work with you to supplement the datasets as needed.

For these secondary data replications, we would like the analysis plan to be completed before the preregistration goes through review, so that after review, the only remaining steps are registration and running the analysis code on the full datasets. To facilitate that, we are asking that you include in section 19 a link to the code you will use that takes the constructed dataset(s) provided to you and produces the focal analyses (including all of the cleaning, merging, and transforming required). When developing your analysis plan and code, please randomly sample 5% of the data for use in your work and demonstrate that the focal analyses produce sensible results using just that random sample by providing a screenshot of the output (see section 19 for details). **Do not use the rest of the data until after your study is registered and it is time to run the final analyses**. In section 19, you will find a statement that we are asking you to bold that confirms you've only used 5% of the data when developing and testing your code. If this approach will not work for any reason, please let Andrew know and disclose deviations from this plan somewhere in the preregistration.

- In cases where we are providing you a complete dataset, you can just sample out 5% of the observations and hold the rest out until you are ready to perform the final analysis.
- In cases where we are providing you multiple datasets that need to be combined prior to analysis, please sample out 5% of the observations in whatever way is most sensible.
    - For example, in cases where each dataset contains complete observations on its own (a typical 'row bind' situation), it makes the most sense to sample out 5% of each dataset separately and then combine them together to develop and test your code.
    - In cases where datasets need to be merged in order to create complete observations (a typical 'column bind' situation), it makes the most sense to merge the separate datasets into a full dataset first, and then sample out the 5% before proceeding with the rest of the analysis code.
- We leave the decision on how to sample out the random subset of data to you, so long as (a) you are not performing any analyses on the complete dataset until after your study is registered and (b) whatever decision you make is documented in the preregistration.

Finally, in cases where the replication data combines observations from the original study with observations that were not used in the original study (what we are calling 'hybrid replications'), please perform up to three analyses (details immediately below). This will likely require you to subset your data, based on the description of the original analysis provided in the study.

- When the 'new' data alone can clear the minimum power threshold, please perform one analysis that relies only on the 'new data' (the focal replication analysis), one analysis that relies on all available data, and a third analysis that relies only on the original data (the focal reproduction analysis). Please make sure all three analyses are documented (with code) in section 19 below.
- When the 'new' data alone *cannot* clear the minimum power threshold, please perform one analysis that combines all available data (the focal replication analysis), and a second that only uses the old data (the focal reproduction analysis). Please make sure both analyses are documented (with code) in section 19 below.

Please contact Andrew if you have any questions. After you've completed the remaining sections of the preregistration and uploaded all the necessary materials to the OSF, please contact the SCORE coordinators regarding next steps.

Preregistration of Kavanagh_covid_BNrQ

Existing Data Replication

Study Information

1. Title (provided by SCORE)

RR TEAM INSTRUCTIONS: *This has been determined by SCORE.*

Replication of a research claim from Kavanagh et al. (2020).

2. Authors and affiliations

RR TEAM INSTRUCTIONS: *Fill in the names and affiliations of your team below.*

Andrew Tyner [Data identification & preparation][1]
Nick Huntington-Klein[2]

1 Center for Open Science, Charlottesville, VA
2 Seattle University, Seattle, Wa

3. Description of study (provided by SCORE)

RR TEAM INSTRUCTIONS: *This description has been provided by SCORE. Please review and make a SCORE project coordinator aware of any edits, additions, and corrections you would suggest to the paragraph. You are free to add additional descriptions of your project in a separate paragraph.*

Greater Republican political orientation is associated with reduced social distancing among U.S. counties. This reflects the following statement from the paper's abstract: "Using 15–17 million anonymized cell phone records, we find that lower per capita income and greater Republican orientation were associated with significantly reduced social distancing among U.S. counties." This claim was tested as follows: "Multivariable ordinary least squares regression for percentage-point change in average county mobility, given an interquartile change in given characteristics." An interquartile increase in support for Trump (I.Q.R. = 20.3%) resulted in a 4.1 percentage point decrease in social distancing (95% C.I. = 3.0–5.2) [From Table 1: $p < 0.001$].

4. Hypotheses (provided by SCORE with possible Data Analyst additions)

RR TEAM INSTRUCTIONS: *The focal test for SCORE is indicated as H*. If you will test additional hypotheses (or use alternate analyses) that help you to evaluate the claim your*

replication/reproduction is testing, number them H1, H2, H3 etc. (You can place H in the list wherever makes sense). Please make sure that any additional hypotheses are logical deductions/operationalizations of the selected SCORE claim or are necessary to properly interpret the focal H* hypothesis. Research that is outside this scope should be described in a separate preregistration.*

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Are the listed hypotheses specific, concise, clearly testable, and specified at the level of operationalized variables?*
- *Are hypotheses identified as directional or non-directional, and, if applicable, have the direction of hypotheses been stated? (Example: "Customers' mean choice satisfaction will be higher in the CvSS architecture condition than in the standard attribute-by-attribute architecture condition.")*
- *Does the list of hypotheses/tests indicate whether additional hypotheses are taken from the original study or modified/added by the team?*

H*: At the level of U.S. counties, support for Donald Trump in the 2016 presidential election will be negatively associated with social distancing behavior.

Design Plan

5. Study type

NOTE: *The study type selected should be based on the data collected for the replication, and not necessarily the data used in the original study.*

- Experiment - A researcher randomly assigns treatments to study subjects, this includes field or lab experiments. This is also known as an intervention experiment and includes randomized controlled trials.
- **Observational Study - Data is collected from study subjects that are not randomly assigned to a treatment. This includes surveys, natural experiments, and regression discontinuity designs.**
- Meta-Analysis - A systematic review of published studies.
- Other

6. Blinding

RR TEAM INSTRUCTIONS: *Select any/all of the below that apply for your study by bolding them. You will give a longer description in the next question.*

- **No blinding is involved in this study.**
- For studies that involve human subjects, they will not know the treatment group to which they have been assigned.
- Personnel who interact directly with the study subjects (either human or non-human subjects) will not be aware of the assigned treatments. (Commonly known as “double blind”)
- Personnel who analyze the data collected from the study are not aware of the treatment applied to any given group.

[QUESTION 6 - BOLD YOUR RESPONSE ABOVE]

7. Blinding

RR TEAM INSTRUCTIONS: *Since all existing data replications are based on data that has already been collected, in most cases it will not be necessary to comment on participant blinding. In the rare instance when an existing experiment is being re-analyzed for an existing data replication and blinding is a relevant consideration, please provide below any details regarding blinding that are important for a reviewer to be aware of.*

Blinding is not relevant to this replication study.

8. Study Design

RR TEAM INSTRUCTIONS: *Please describe how data was collected in the original study and how it compares to the data that was selected for the replication attempt. Explain why the data selected for the replication study is suitable for a replication and if any substantial deviations exist between the two.*

If the data used in the replication combines observations from the original study with new observations (e.g. if the data selected for the replication attempt comes from the same longitudinal survey as the original study), describe how 'original' and 'new' observations relate to each other and an estimate for what proportion of the final dataset's observations will be comprised of original vs. new observations.

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration specify the unit of analysis?*
- *Does the preregistration provide sufficient detail about how the data selected for the replication attempt deviates from or is congruent with the data employed in the original study?*
- *Does the preregistration describe whether and how 'original' and 'new observations' are combined together for the replication dataset?*

The original study (https://osf.io/n3csj/?view_only=a682bd32535b4bbf8ba6c816b0120d36) was a preprint downloaded by the SCORE team on 4/13. The authors found that 2016 county-level vote share for Donald Trump is negatively associated with social distancing behavior, which was "conceptualized as changes in average distance traveled for March 19–28, 2020 (most recent data available), relative to matched days of a pre- COVID-19 reference week," based on data from Unacast (p. 3). The authors test for this association while controlling for a set of additional county-level variables. Please see Table 1 on page 7 for details.

The replication study will similarly test for an association between social distancing behavior and support for Donald Trump in 2016, but using a more recent set of dates and a different data source for social distancing: specifically, the Bureau of Transportation Statistics' 'Daily Travel' dataset, which is described here and downloadable here. Unacast was contacted about providing an updated dataset, but was unable to provide a custom county-level dataset due to demand. Even so, the BTS' data has the advantage of being open access, more thoroughly documented, and more comprehensive, since it extends backward in time through January 1, 2019.

The BTS provides a number of different social distancing measures that the data analyst can select from (e.g. 'Population Not Staying at Home,' 'Number of Trips between 1-3 miles,' 'Number of trips between 100-250 miles'). These are all measured as counts by day, so the analyst should probably standardize them using a measure of county population (see below).

Alternatively, the raw counts can be used, and the dependent variable can be computed as a percentage change from a baseline within each county.

At the outset, the following note from the BTS data site should be surfaced for the review team to consider: “These data are experimental and may not meet all of our quality standards. Experimental data products are created using new data sources or methodologies that benefit data users in the absence of other relevant products. We are seeking feedback from data users and stakeholders on the quality and usefulness of these new products. Experimental data products that meet our quality standards and demonstrate sufficient user demand may enter regular production if resources permit.” via: https://www.bts.gov/browse-statistical-products-and-data/trips-distance/explore-us-mobility-during-covid-19-pandemic

9. Randomization (free response)

RR TEAM INSTRUCTIONS: *If the variables used for this replication attempt were randomized, state how they were randomized, and at what level.*

Randomization is not relevant to this replication study.

Sampling Plan

This section describes how the data sources for the replication were selected, how they were prepared into a replication dataset, and the number of observations that will be analyzed from these data. Please keep in mind that the data described in this section are the actual data used for analysis, so if you are using a subset of a larger dataset, please describe the subset that will actually be used in your study.

10. Existing data

1.1.1. Registration prior to creation of data
1.1.2. Registration prior to any human observation of the data
1.1.3. Registration prior to accessing the data
1.1.4. Registration prior to analysis of the data
1.1.5. Registration following analysis of the data

11. Explanation of existing data

NOTE: *For replications that rely on existing data sources, this question refers to the data that will be used for the replication analysis (i.e. the final replication dataset), and not (a) the data from the original study or (b) the data sources accessed to construct the replication dataset.*

Since no new data will be created for 'existing data replications,' 1.1.1 should never be selected. Since all analyses will occur after registration, 1.1.5 should also never be selected.

County-level data from a wide variety of datasets have been accessed, cleaned, and merged by the data finder prior to being sent to the data analyst. All steps have been documented in section 12 below. All measures were selected based on their (a) expected relevance to the replication analysis and (b) the record of the original variables used, as determined from the authors' script: https://osf.io/5c438/?view_only=086301fb801f4ecda7a7e796d4a51e10. None of the variables were selected because of their likelihood (or not) of leading to a confirmatory result.

12. Data collection procedures

RR TEAM INSTRUCTIONS: *Please describe the process for constructing the replication dataset in as much detail as you can. The sections below should be used to provide the following information:*

- *Which variables are needed from the original study to perform a good-faith, high-quality replication.*
- *Which data sources were used, why they were selected, any deviations between the original study design and the replication study design that these selections present, and the procedures used to access the data.*
- *Which of the variables from the original study are available in the replication data sources, including relevant details about each measure.*
- *The procedure for creating the replication dataset, in both narrative and script form.*
- *A data dictionary that documents each variable included in the replication dataset.*

In the sections below, please provide links to the original materials whenever possible -- including descriptions of the original datasets and corresponding codebooks. If materials can be shared on the OSF, please do so, and provide view-only links to those materials.

Specific points to keep in mind for reviewers:

- *Does the preregistration describe which data sources were selected for the replication study and why each is suitable?*
- *Does the preregistration make clear how the data sources were used to construct the replication dataset?*

(a) Data Needed

RR TEAM INSTRUCTIONS: *List below the datasets and variables the original author used to analyze the focal claim. Include details regarding the sample size, waves or years used, and other details pertinent to finding an existing dataset for replication. Please include page numbers when excerpting from the original article. If possible, categorize the list of variables as one of the following: dependent variable, focal independent variable, control variable, or sample*

parameters/clustering variable. Finally, include the sample size of the original study's focal analysis, if it is available.

All references to the authors' R script refer to the following file, which helps identify how certain variables were collected:

- https://osf.io/5c438/?view_only=086301fb801f4ecda7a7e796d4a51e10
- The script confirms that all American Community Survey variables except the rurality variables are 5-year averages, 2014-2018. Rurality variables are drawn from the 2010 decennial file for table H2.

Dependent Variable(s)

Social distancing

- Data from Unacast, "who measure county-level averages of distance traveled per person" (p. 3)
- Specifically, "social distancing was conceptualized as changes in average distance traveled for March 19–28, 2020 (most recent data available), relative to matched days of a pre-COVID-19 reference week" (p. 3)
- The manuscript does not state which reference week was used.

Focal Independent Variable(s)

Political preferences

- Political preferences were "operationalized by the 2016 county-level vote share for President Trump" using data from the MIT Election Data and Science Lab (p. 3).

Control Variable(s)

Socioeconomic status

- Socioeconomic status was "operationalized by income per capita" using 5-year averages data (2014-2018) from the American Community Survey (ACS).
- This is the other focal independent variable of the study, though not a focal variable for the purposes of this replication study.
- The variable used is S1902_C03_019E from ACS table S1902.

Percent male; percent Black; percent Hispanic; share of adults with college degrees

- Obtained from the ACS, 5-year averages from 2014-2018.
- Percent male is DP05_0002PE from table DP05
- Percent Black is DP05_0065PE from table DP05.
- Share of adults with college degrees is S1501_C02_015E from table S1501.

Age distribution

- Obtained from the ACS, 5-year averages from 2014-2018.

- Defined as “percentage for each decade of life” (p. 7) in Table 1 and further defined in the authors’ R script as 0-14, 15-24, 25-34, 35-44, 45-54, 55-64, 65-74, and 75+.
- Variables used are DP05_0005PE (5 and under) through DP05_0017PE (85 and over) from table DP05.

Share of the workforce in retail; share of the workforce in transportation; share of the workforce in health, educational, or social services

- Obtained from the ACS, 5-year averages from 2014-2018. All variables are from table DP03.
    - Percent in retail is DP03_0037PE.
    - Percent in transportation is DP03_0038PE.
    - Percent in health, education, and social services is DP03_0042PE.

Rurality

- Defined specifically in Table 1 (p. 7) as ‘percentage rural.’
- Uses H002005 [total rural] and H002001 [Total] from table H2 [2010 decennial file].
- As noted below, the H2 data measures housing units rather than population. The replication dataset makes these same items available, as well as items from P2, which instead measures total population.

State fixed effects

Sample size of analysis has 3,037 observations (see caption of Table 1).

(b) Data Access

RR TEAM INSTRUCTIONS: *Describe below the data sources that will provide the replication variables. Include information such as the name of the data source (e.g., Indonesian Family Life Survey), the description and link of the data source, and the waves needed to create a final replication dataset.*

Also describe the process for accessing the data sources that will be used to create the final replication dataset; specify how long long it took for the registration to be approved and what information was required (e.g., writeup of the purpose of the project, email address from an IPCSR institution, etc.); and verify that the data can be opened as expected. If applicable, provide a link to the page where you registered to access the data.

Describe in detail any restrictions on data access and data-sharing, as well as any additional terms of data use that will be relevant for the replication study and final report (e.g. citations that will need to be made). If you were able to access the data because of special permissions that you have, but that you expect other researchers might not have, please document those as well.

Bureau of Transportation Statistics Data

Full ‘Trips by Distance’ dataset was downloaded (csv format) on 9/21, via https://data.bts.gov/Research-and-Statistics/Trips-by-Distance/w96p-f2qv

- As of download date and time, the data was last updated on 9/21. The latest date of county-level data it contains is 9/12/20.
- Note: “Data analysis is conducted at the aggregate national, state, and county levels. To assure confidentiality and support data quality, no data are reported for a county if it has fewer than 50 devices in the sample on any given day.”
- Additionally: “Trips are defined as movements that include a stay of longer than 10 minutes at an anonymized location away from home. Home locations are imputed on a weekly basis. A movement with multiple stays of longer than 10 minutes before returning home is counted as multiple trips. Trips capture travel by all modes of transportation. including driving, rail, transit, and air.”
    - This applies to all definitions of ‘trips’ captured in the variables below.
- The license links to the following page suggesting that the data can be freely shared and distributed as part of a U.S. government creative work: https://www.usa.gov/government-works

Trump Support Data

- Accessed via: https://electionlab.mit.edu/data
- Data was downloaded (9/3) as countypres_2000-2016.csv via https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/VOQCHQ
- Description provided: “This dataset contains county-level returns for presidential elections from 2000 to 2016. (2018-10-11)”
- Harvard dataverse lists the citation as follows: MIT Election Data and Science Lab, 2018, "County Presidential Election Returns 2000-2016", https://doi.org/10.7910/DVN/VOQCHQ, Harvard Dataverse, V6, UNF:6:ZZe1xuZ5H2l4NUiSRcRf8Q== [fileUNF]
- This is the only ‘raw’ data file not appearing on the OSF. Please download from the Harvard dataverse if you would like a version of the raw data.

American Community Survey Data

- Please see this site for Census Bureau citation preferences: https://ask.census.gov/prweb/PRServletCustom/YACFBFye-rFIz_FoGtyvDRUGg1Uzu5Mn*/!STANDARD?pyActivity=pyMobileSnapStart&ArticleID=KCP-3282
- The following ACS tables have been downloaded for use in the variables below:
    - S1902 [downloaded 9/3]:
    - S1501 [downloaded 9/8]
    - DP05 [downloaded 10/19]
    - DP03 [downloaded 10/19]
    - P2 [downloaded 10/19]
    - H2 [downloaded 10/20]
- For each table above, the same (general) approach was used to access the data:

    - From https://data.census.gov/cedsci/, search for the table name.
    - Click 'View all Tables'
    - Click 'Customize tables'
    - Under Geographies, select 'County,' then 'All counties in the United States,' then close
    - Click 'Download Table,' then change the selection so that only 2018/5-Year is selected.
        - Exceptions are P2 and H2, where 2010 is the selected period.
    - Download table

2019 County Population

Note: This file does not have a use in the replication dataset, but it was used to validate that, in the transportation data for 2019, the sum of the 'population staying home' and 'population not staying home' in each county in each year provides a good approximation of county population.

Accessed via County Population by Characteristics: 2010-2019

- Click datasets at bottom of page.
- Click '2010-2019/' folder.
- Click 'counties/' folder.
- Click 'totals' folder.
- Download co-est2019-alldata.csv [file downloaded 9/10/20]
- Use POPESTIMATE2019 as 2019 population estimate

(c) Variable Availability

RR TEAM INSTRUCTIONS: *For each variable required for the replication analysis (listed above), describe the variables from the replication data that can be used to measure it (including which data files or sources each measure is found in),* ***any notes a data analyst should consider when using the measure in a replication analysis****, and any important differences between the original variable and the proposed replication variable.*

If there are multiple variables in the replication data that correspond to a required variable (e.g. two different measures of education in the replication data), include all of those options below. If a variable from the original study ***cannot*** *be measured using the replication data, please make that clear as well.* ***Finally, include a description of the identifiers used to merge multiple datasets, if applicable.***

Mobility

- As discussed above, the data source for the mobility data has changed between the original study (Unacast) and this replication study (Bureau of Transportation Statistics). Accordingly, it won't be possible to re-create the exact measures used in the original

study. Instead, the data finder (A. Tyner) has made available all of the measures from the Bureau of Transportation Statistics, so that the data analyst can decide which to use.

- Observations in the transportation data are county-days, extending from 2019-01-01 to 2020-09-12. The dates from before the pandemic are being retained in the replication dataset, in case the analyst wants to use any of them as a reference period.
- The measures available are (variable names as they are in replication dataset):
    - pop_home: Number of residents staying at home, i.e., persons who make no trips with a trip end more than one mile away from home
    - pop_not_home: Number of residents not staying at home
    - num_trips: Number of trips made by residents, i.e., movements that include a stay of longer than 10 minutes at an anonymized location away from home
    - trips_under_1: Number of trips by residents shorter than one mile
    - trips_1_3: Number of trips by residents greater than one mile and shorter than 3 miles (1 ≤ trip distance < 3 miles)
    - trips_3_5: Number of trips by residents greater than 3 miles and shorter than 5 miles (3 ≤ trip distance < 5 miles)
    - trips_5_10: Number of trips by residents greater than 5 miles and shorter than 10 miles (5 ≤ trip distance < 10 miles)
    - trips_10_25: Number of trips by residents greater than 10 miles and shorter than 25 miles (10 ≤ trip distance < 25 miles)
    - trips_25_50: Number of trips by residents greater than 25 miles and shorter than 50 miles (25 ≤ trip distance < 50 miles)
    - trips_50_100: Number of trips by residents greater than 50 miles and shorter than 100 miles (50 ≤ trip distance < 100 miles)
    - trips_100_250: Number of trips by residents greater than 100 miles and shorter than 250 miles (100 ≤ trip distance < 250 miles)
    - trips_250_500: Number of trips by residents greater than 250 miles and shorter than 500 miles (250 ≤ trip distance < 500 miles)
    - trips_over_500: Number of trips by residents greater than 500 miles (trip distance ≥ 500 miles)
- **Data finders' recommendation:** All of the distance measures seem to suffer from the same limitation -- namely, that they're affected by the density of the county, since trips will be longer on average if the destination is farther away. Since density is likely correlated with the focal IV (Trump support), it doesn't seem like the best choice for a DV. The cleanest measure seems to be the proportion of the population staying home: pop_home / (pop_home + pop_not_home).
- Note as well that 'pop_home + pop_not_home' provides a reasonable approximation of county population. The 'data prep' script linked below demonstrates that: (a) among non-missing observations for pop_home + pop_not_home, the sum is stable within each county-year; and (b) when compared to the Census' 2019 estimates for county population (also provided in the replication dataset), the two measures are quite similar. The mean difference expressed as a fraction of the 2019 Census estimates is 0.97% [min = 0.00%; third quartile = 1.26% max = 56.4%]. The max (Concho County, Texas) is

an outlier; the next two biggest differences (Gulf County, Florida and Loving County, Texas) are off by 18.2% and 10.1%, and the rest are all below 10%.

Share of Trump votes

- Uses MIT election data; filtered to only include observations from 2016 and about Trump, in geographic units with non-missing FIPS values.
    - The following geographic units with missing FIPS values have thus been excluded: Federal Precinct, Maine UOCAVA, Statewide writein
- Within each county, Trump's share of the vote (trump_share) is calculated as candidatevotes / totalvotes.

Per capita income

- Uses 5-year, 2018 ACS data from table S1902.
- Selection is S1902_C03_019E
    - PER CAPITA INCOME BY RACE AND HISPANIC OR LATINO ORIGIN -- Total population
    - Converted to numeric in the replication dataset
- Notes: FIPS # 35039 is listed as 'null,' and filtered out for easier merging

Percent male

- Uses 5-year, 2018 ACS data from table DP05
- Selection is DP05_0002PE
    - Percent Estimate!!SEX AND AGE!!Total population!!Male
    - Converted to numeric in the replication dataset.

Percent Black

- Uses 5-year, 2018 ACS data from table DP05
- Selection is DP05_0065PE
    - Percent Estimate!!Race alone or in combination with one or more other races!!Total population!!Black or African American
    - Converted to numeric in the replication dataset.

Percent Hispanic

- Uses 5-year, 2018 ACS data from table DP05
- Selection is DP05_0071PE
    - Percent Estimate!!HISPANIC OR LATINO AND RACE!!Total population!!Hispanic or Latino (of any race)
    - Converted to numeric in the replication dataset.

Percent with college degree

- Uses 5-year, 2018 ACS data from table S1501.
- Selection is S1501_C02_015E
    - Estimate!!Percent!!Population 25 years and over!!Bachelor's degree or higher

    - Converted to numeric in the replication dataset.

Employment in industries

- Uses 5-year, 2018 ACS data from table DP03
- For all three measures below, FIPS # 35039 is listed as '(X),' and will be NA in the replication dataset.
- Percent in retail: uses DP03_0037PE
    - Percent Estimate!!INDUSTRY!!Civilian employed population 16 years and over!!Retail trade
    - Converted to numeric in the replication dataset.
- Percent in transportation: uses DP03_0038PE
    - Percent Estimate!!INDUSTRY!!Civilian employed population 16 years and over!!Transportation and warehousing, and utilities
    - Converted to numeric in the replication dataset.
- Percent in health care, education, and social services: uses DP03_0042PE
    - Percent Estimate!!INDUSTRY!!Civilian employed population 16 years and over!!Educational services, and health care and social assistance
    - Converted to numeric in the replication dataset.

Percent rural [option 1]

- Uses the P2 decennial file from the ACS (2010).
    - This is a departure from the original study, which uses the H2 decennial file. H2 measures housing units rather than population. P2 measures population.
    - To confirm that P2 is the right file, 10 random counties were selected to compare the value of P002001 ['Total'] to the value for 'Population, Census, April 1, 2010' on each county's respective 'Quick Facts' site. The results are contained in the following link. In each case, the two values match exactly: https://osf.io/x5jng/?view_only=72073cd358194ca099dd27905aca209c
- Percent rural is defined as P002005 / P002001
    - Total!!Rural / Total
- Between 2010 and the present, Shannon County, SD was changed to Oglala Lakota County with a new FIPS (46102). The FIPS value has been changed in the rural dataset but the county name ('county_rural') has not, to show the record of how this value was changed.
- Between 2010 and the present, Wade Hampton Census Area, AK was changed to Kusilvak Census Area with a new FIPS of 02158. The FIPS value has been changed in the rural dataset but the county name ('county_rural') has not, to show the record of how this value was changed.
- Additionally, Bedford city, VA (FIPS 51515) has since been consolidated with Bedford County, Virginia (FIPS 51019). It was left unaltered in the rural dataset, since the merging procedure will drop it from the final set of county-level variables anyways.

    - Note that the data for Bedford County, Virginia currently reflect its population and percent rural with its 2010 definition. The data finder can recalculate with the inclusion of Bedford city, VA data on request.
- **Note:** The original P2 contains 54 counties with a string appended to the P002001 variable. As documented immediately below, the data cleaning procedure documented in the original authors' script would appear to create values of NA for those counties if the strings were not removed in pre-processing (noting, though, that the original authors used items from H2 rather than P2). This would make the 'percent rural' variable NA as well for those observations.
    - Note that it's only 53 counties in the replication data affected by this issue, since FIPS 72063 (Gurabo Municipio, Puerto Rico) was not measured in the transportation data and thus was not joined to the replication dataset.
    - The data cleaning procedure for the replication dataset **does** remove those strings prior to creating the percent rural variable. To record which variables were affected, a dummy variable has been added (pop_with_char) that has the value of T for observations where a string was removed from P002001 to create the denominator variable (total_pop).

Percent rural (option 2)

- Uses the H2 decennial file from the ACS (2010) to match the original study. Note that H2 items are measuring housing units, rather than total population.
- Percent rural is defined as H002005 / H002001
    - Total!!Rural / Total
- **Note:** The original H2 data file contains 40 counties with a string appended to the H002001 variable. The data cleaning procedure documented in the original authors' script would appear to create values of NA for those counties if the strings were not removed in pre-processing. This would make the 'percent rural' variable NA as well for those observations.
    - As above, note that it's only 39 counties in the replication data, since FIPS 72063 (Gurabo Municipio, Puerto Rico) was not measured in the transportation data and thus was not joined to the replication dataset.
    - The data cleaning procedure for the replication dataset **does** remove those strings prior to creating the percent rural variable. To record which variables were affected, a dummy variable has been added (housing_with_char) that has the value of T for observations where a string was removed from H002001 to create the denominator variable (total_housing).

Age brackets

- Uses 5-year, 2018 ACS data from table DP05
- Replication dataset contains the following discrete age bins, all converted to numeric:
    - DP05_0005PE: Percent Estimate!!SEX AND AGE!!Total population!!Under 5 years
    - DP05_0006PE: 5 to 9

    - DP05_0007PE: 10 to 14
    - DP05_0008PE: 15 to 19
    - DP05_0009PE: 20 to 24
    - DP05_0010PE: 25 to 34
    - DP05_0011PE: 35 to 44
    - DP05_0012PE: 45 to 54
    - DP05_0013PE: 55 to 59
    - DP05_0014PE: 60 to 64
    - DP05_0015PE: 65 to 74
    - DP05_0016PE: 75 to 84
    - DP05_0017PE: 85 and older
- Additionally, the authors' script includes the following combined age brackets, which have been recreated in the replication dataset:
    - 14 and under: under 5 + 5-9 + 10-14
    - 15 to 24: 15-19 + 20-24
    - 55 to 64: 55-59 + 60-64
    - 75+: 75-84 + 85+

State fixed effects

- State data is provided as a postal code from the transportation data ('state' in the replication data) as well as a separate postal code variable and state name variable from the MIT data ('state_po' and 'state_votes,' respectively).
    - Only one should be needed, but three are provided for data quality/verification purposes.

Additional variables contained in the dataset

There are additional supplementary variables contained in the replication dataset for the analyst to use, in case they find them useful. The data analyst should review the data dictionary (linked below) for a full account of each variable. Some notable variables include:

- County name variables for each of the separate datasets that are merged together below. These are included as a data quality and verification check, to ensure that merging by FIPS values always worked as intended. Note that the county names corresponding to the rural data ('county_rural' and 'county_housing_rural') reflect the county names as they existed in 2010. This means the entries for county_rural and county_housing_rural for FIPS 46102 is Shannon County, South Dakota rather than Oglala Lakota County, South Dakota, and the entries for FIPS 2158 are Wade Hampton Census Area, Alaska rather than Kusilvak Census Area, Alaska.
- Two different measures of county population:
    - total_pop: The 2010 county population estimate, used in the calculation of the percent rural. This is drawn from the ACS' P2 decennial file.
    - population, which is a 5-year average (2014-2018) drawn from the DP05 table in the ACS.

- As noted above, within each county for each year, the sum of pop_home and pop_not_home provides a reasonable approximation of county population. If the analyst wanted a county population estimate specifically for 2020, taking the sum of those variables for 2020 is an option.

(d) Data Creation

RR TEAM INSTRUCTIONS: *Create a dataset using the data sources and variables listed above. Provide a detailed narrative describing how the various datasets were cleaned and merged into a final replication dataset. Provide a view-only link to a clearly commented script on the OSF that produces the replication data as described in the narrative. Our preference is that this be either an R script or a script from another language that similarly allows for open and reproducible analyses. Please let the SCORE team know if this is not possible.*

- *If the data can be freely shared and posted to OSF, please post it in your OSF project and provide a link to the completed dataset below.*
- *If any part of the dataset cannot be shared between researchers or posted to the OSF, please leave the final dataset off the OSF. Instead, include either below or in your script (commented out at the bottom) two pieces of information that will help an independent team verify they have created the dataset according to your instructions:*
  - *The dimensions of the final dataset(s) you've created (# of rows, # of columns)*
  - *A summary of 8-10 variables in the replication dataset. For numeric variables, the summary should include the mean, standard deviation, and count of NAs. For categorical variables, the summary should include each level present in the data and its count, as well as a count of NAs. If multiple datasets are submitted as part of your work, at least one variable should be included from each dataset.*

The data from the replication sources should be preserved in as 'raw' a form as possible, in order to give the data analyst the most latitude to clean the variables as they see fit. Variables from the original source should be preserved in their original form (e.g. do not recode values of 99 to NA). New variables should only be created when they're needed to complete the merge or combine the datasets; in those cases, please preserve a version of the original, unaltered variable in the new dataset.

Please also use this section to describe:

- *Any deviations between the original study design and the replication design that would result from using this replication dataset.*
- *Any notes about using these variables that you would like to pass along to the data analyst.*

Note: The replication_data.csv that's constructed by the procedure below is a country-day level dataset that spans January 1, 2019 - September 12, 2020 (621 days, representing all of the dates available in the Bureau of Transportation Statistics dataset) for each of the available 3142 counties. Accordingly, the combined dataset has 1951182 rows, with 63 columns. The vast

majority of these columns are stable county-level covariates. The full set of dates was preserved in order to give the data analyst the most options for how to construct a 'pre-COVID' comparison (should they wish to), but nonetheless most observations won't be necessary for the focal analysis.

- The size of the combined dataset (1.36 GB) was too large for the data finder to upload to the OSF. Instead, the data finder has uploaded two files: one with the stable, county-level variables (3142 rows, county_variables.csv); and the other the county-day level mobility measures (1951182 rows, transportation.csv).
- The OSF contains a short script that takes transportation.csv and county_variables.csv and produces the replication dataset (replication_data).
- For easier handling, the data finder recommends that after creating 'replication_data,' the data analyst should select the dates they want to work with for the replication analysis and then filter accordingly. The focal analysis will be performed at the county level, so the county-day level observations will eventually need to reduce to a single mobility or mobility change estimate for each county.

The procedure to construct the replication dataset input files is as follows:

Load in each of the raw data files documented in section 12b and the README file (https://osf.io/xftdv/?view_only=5d2e876e6b7245e090c3f6a2eb0b5d88).

For each dataframe, clean as needed to prepare for an eventual merge:

- In the transportation data from the BTS, convert all FIPS values with a leading 0 to 4-digit FIPS for easier merging; additional light processing to prepare dataframe for merge.
- In the voting data from MIT, select observations from 2016, about Donald Trump, and for geographic units with a non-NA FIPS value. Compute trump_share as Trump's votes divided by total votes, and perform additional light processing.
- For each file from the ACS, convert the GEO_ID to a FIPS value and then compute and select the focal variable(s) as needed. See section 12c for details on each variable.
    - Note: The first 54 rows of the P2 decennial file contain a string at the end of the P002001 (total population) variable. It's unclear what the string represents or why it's only included for that subset of rows, but it doesn't affect the population estimates. The population totals are just as accurate for this subset as for the rest of the rows. The version of total population in the replication dataset has removed the string, but there's a T/F dummy that identifies those cases (pop_with_char). As noted previously, it's only 53 counties affected by this issue in the replication dataset, for the reasons documented above.
    - Additionally: 40 rows of the H2 decennial file contain a string at the end of the H002001 (total housing) variable. It's unclear what the string represents or why it's only included for that subset of rows. The version of total housing in the replication dataset has removed the string, but there's a T/F dummy that identifies those cases (housing_with_char). As noted previously, it's only 39

counties affected by this issue in the replication dataset, for the reasons documented above.

- Create a dataset of county-level variables by merging each of the county-level datasets, with counties corresponding to the unique fips in the cleaned transportation dataset (county_variables.csv): https://osf.io/83m7y/
- Create a county-day level dataset from the cleaned BTS data (transportation.csv): https://osf.io/2pqn7/
- The script also contains some data validation code; specifically, validating that (a) in the transportation data, the sum of the population staying at home and the population not staying at home is stable within each county-year, suggesting it's a good estimate of county population; and (b) comparing the 2019 sum of home + not at home to an alternative measure of county population of 2019.

The data analyst will need to merge transportation.csv and county_variables.csv using the code found here: https://osf.io/d4whs/. The resulting dataset (replication_data) has variables in the same order as the data dictionary referenced below. To ensure the merged dataset matches what was intended, a variable summary file is also provided: https://osf.io/zjfg4/

Raw file links

- All raw data files except for the MIT voting data have been uploaded to the OSF. The MIT voting dataset is available in its original form through the Harvard Dataverse site linked above.
- Transportation data from the Bureau of Transportation Statistics: https://osf.io/y4pj3/?view_only=5d2e876e6b7245e090c3f6a2eb0b5d88
- Income per capita data from ACS: https://osf.io/j6d3w/?view_only=5d2e876e6b7245e090c3f6a2eb0b5d88
- College degree data from ACS: https://osf.io/ng7vt/?view_only=5d2e876e6b7245e090c3f6a2eb0b5d88
- Demographic data from ACS: https://osf.io/e9r2t/?view_only=72073cd358194ca099dd27905aca209c
- Employment industry data from ACS: https://osf.io/bdfhx/?view_only=72073cd358194ca099dd27905aca209c
- Population data (including rural population) from ACS: https://osf.io/vky8u/?view_only=72073cd358194ca099dd27905aca209c
- 2019 Census population estimates: https://osf.io/gz4t2/?view_only=5d2e876e6b7245e090c3f6a2eb0b5d88

Non-Matching Counties

The vast majority of counties could be matched across the various datasets documented above using the provided or derived FIPS value. There were exceptions, though, and these were omitted from the analysis data, which are documented here. The discrepancies between the DV dataset (from the BTS) and the IV dataset (from MIT) were investigated most deeply.

- Among the counties present in the transportation/DV dataset, 28 could not be matched to the voting dataset from MIT:
    - 26 of these non-matching cases were census areas, boroughs, and municipalities from Alaska (e.g. Juneau City and Borough, Fairbanks North Star Borough, Bethel Census Area).
        - Note: One of these counties is Kusilvak Census Area, Alaska (FIPS 2158), which appears in the rural population and housing data (tables P2 and H2) with an older name (Wade Hampton Census Area, Alaska) and older FIPS (2270). Before being merged with the replication dataset, the FIPS was changed to 2158 but the names in the 'county_rural' and 'county_rural_housing' variables were retained as Wade Hampton Census Area, Alaska. However, this county does not appear in the MIT voting data with either FIPS, so it'll likely drop out of the replication analysis anyways.
    - One is Oglala Lakota County, SD (FIPS 46102). This is because the voting data from MIT was using the FIPS associated with the previous name of the county (Shannon County): 46113. The FIPS code was changed to 46102 in the MIT voting data before the merge to accommodate this deviation.
        - Oglala Lakota County is fully available in the replication data. It already had the new name in the MIT voting data but with the old FIPS (46113), and the FIPS was manually updated to 46102 before the voting data was joined to the transportation data. It had both the old name and the old FIPS in the 2010 rural population and housing data (tables P2 and H2, respectively). Before being merged with the replication dataset, the FIPS was changed to 46102 but the names in the 'county_rural' and 'county_rural_housing' variables were retained as Shannon County, South Dakota. So, Oglala Lakota is a complete observation in the replication data.
    - The final non-matching case is Kalawao County, Hawaii (FIPS 15005), which does not appear in the voting data from MIT. The data from MIT includes four Hawaiin counties: Hawaii, Honolulu, Kauai, and Maui, but not Kalawao.
- Among the counties present in the voting/IV dataset, 41 could not be matched to the transportation dataset from BTS:
    - 38 of these were districts in Alaska named by district number (e.g. District 1, District 2). These include districts from 1 to 40 and District 99, except for Districts 13, 16, and 20 which do have matching FIPS values in the transportation data (see below). The number 40 could correspond to the number of districts in the Alaska House of Representatives, but it's not clear from the data.
    - As mentioned above, the FIPS value for Oglala Lakota, SD reflects the former name (Shannon County), and was changed to 46102 to facilitate the merge.
    - Bedford, VA (FIPS 51515) also does not appear in the transportation data, likely reflecting a change in Bedford's county status. However, Bedford did not record

any votes cast for presidential candidates in 2016, so nothing is lost from its exclusion from the dataset.
    - The final non-matching case is FIPS 36000, listed as 'Kansas City, MO.' This is a puzzling case, because: (a) Kansas City is not itself a county and (b) 36000 does not appear consistently on other lists of U.S. FIPS codes (for example, it's not in this list at all, and it refers to New York state in this list).
- The transportation FIPS values have matches for the FIPS values in all of the other datasets referenced above, save for the 'per capita income' and industry datasets, where Rio Arriba County, NM was excluded because of missingness (see above). Those variables will continue to be missing after the merge anyways, so it does not affect the replication dataset either way.
- In sum, the only substantial discrepancy is the inability to match most of Alaska to the voting data. Only three FIPS values from the transportation data have a match in the voting data: 2013, 2016, and 2020. These are called Aleutians East Borough, Aleutians West Census Area, and Anchorage Municipality in the transportation data; and they're called District 13, District 16, and District 20 in the voting data.

(e) Data Dictionary

RR TEAM INSTRUCTIONS: *Create a data dictionary following this template. Provide below a view-only link to the completed data dictionary included in the OSF project. If the Data Analyst will need to create new variables using the variables in the final replication dataset (e.g. recoding the provided education variable to be in a better format for analysis), please document below your recommendation on how the analyst should do so. Please also document any additional notes regarding the variables in the dataset that do not fit within the provided data dictionary template or the other sections above.*

The data dictionary for replication_data.csv is found here:
https://osf.io/ceb36/?view_only=5d2e876e6b7245e090c3f6a2eb0b5d88

13. Sample size

RR TEAM INSTRUCTIONS: *Please report below the analytic sample size(s) in the replication dataset, with reference to however many units or levels are in the data. Please report as much information here as will be helpful for the review committee to be aware of, including differences in sample size resulting from various analytic decisions (e.g. listwise deletion vs multiple imputation).* ***Finally, when the replication combines observations from the original study with new observations, please estimate what proportion of the analytic sample's observations will be comprised of original vs. new observations.***

As mentioned above, the replication dataset described above is a 1951182 x 63 dataframe with observations measured at the county-day level. After the data analyst computes a single mobility measure for each county, the data will reduce to the county-level prior to analysis.

There are 3142 unique counties in the full dataset. 3115 counties have a non-missing observation for the focal independent variable (trump_share). 3114 have a non-missing observation for each independent variable, including trump_share. Rio Arriba County, NM accounts for the difference, since it is missing a number of variables derived from the ACS (see section 12c for details).

The original study consisted of 3,037 counties. These almost certainly largely overlap with the counties in the replication dataset, since there are only ~3140 counties in the United States, but the reporting in the original study does not allow for a comparison of the non-overlapping counties. Regardless, this replication analysis should **not** be considered a reproduction, regardless of which dates of mobility data are selected by the data analyst, since the dependent variable is not the same.

Required sample size [to be filled out by the SCORE team]: The primary unit of analysis is the county. An estimate of the minimum viable sample size for the data analytic replication is: 863. For comparison, the stage1 required sample size would be: 2990 and the stage2 sample size would be: 4454.

14. Sample size rationale (provided by SCORE)

For data analytic replications in SCORE, three sample sizes are calculated:

- *A minimum threshold sample size, defined as the sample size required for 50% power of 100% of the original effect*
- *A stage 1 sample size, defined as the sample size needed to have 90% power to detect 75% of the original effect*
- *A stage 2 sample size, defined as the sample size needed to have 90% power to detect 50% of the original effect*

Details about how those sample sizes were calculated for this project are found here: https://osf.io/zuh7a/?view_only=60ef0141f40a4e4e8164bced2be540cd

15. Stopping rule (provided by SCORE)

RR TEAM INSTRUCTIONS: *For replications and reproductions involving existing data, this section describes which analyses the SCORE team is recommending be performed. Most often, this corresponds to analyses involving new data, original data, or a combination of new and old data.*

Since the measurement of the dependent variable is different from the original study, no reproduction analysis is possible. Therefore, the SCORE team recommends that a single, focal replication analysis be performed for this study.

Variables

RR TEAM INSTRUCTIONS: *The preregistration form divides variables across three questions: manipulated variables, measured variables, and indices (i.e. analytic variables derived from raw variables). For existing data replications, only fill out the "Measured variables' and 'Indices' sections. Please do not fill out anything in the 'Manipulated variables' section.*

The raw data of any transformed variable (e.g. reaction time → log reaction time) or any created index should be defined in the 'Measured variables' section. Details regarding the variable transformation should be specified in the 'Transformations' section. Details regarding the creation of an index should be specified in the 'Indices' section.

Across these questions, you should define all variables that will later be used during your analysis (including data preparation/processing). You can describe all variables in the preregistration and/or summarize and link to a data dictionary (codebook) in your repository to answer these questions.

If you will share data from your replication, this is also the place to state whether any variables will be removed prior to sharing the dataset (e.g. to reduce risk of participant identification or comply with copyright restrictions on scale items.)

16. Manipulated variables

RR TEAM INSTRUCTIONS: *Manipulated variables in this preregistration refer specifically to variables that have been randomly assigned in an experiment. The use of data from an experiment should be rare in existing data replications. If your existing data replication relies on experimental data, please document each manipulated variable as a measured variable, and use the codebook to indicate what each level of the variable corresponds to (e.g. participants assigned to the treatment condition = 1; participants assigned to the control condition = 0). The default language in bold below has been copied into all existing data replication preregistrations.*

N/A -- not documented for existing data replications.

17. Measured variables

RR TEAM INSTRUCTIONS: *Please use this section to document each variable that was used in the original study's analysis and the role it served (e.g. dependent variable, control variable,*

sample parameter, etc). For each variable, provide the description of the variable offered in the paper and/or codebook of the original study, the variable in the replication dataset that it corresponds to, and explain any deviations between the two. In cases where an equivalent replication variable was not found, explain how, if at all, you expect it will affect the replication attempt. In cases where you are adding a variable that was not present in the original study, please explicitly state that you are doing so, and explain how, if at all, you expect it will affect the replication attempt.

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration surface all of the variables needed to replicate the focal analysis?*
- *Are deviations between the original variables and replication variables documented when needed?*

Trump Vote Share

- Focal regression predictor
- Original: Share of votes in the county that were cast for Donald Trump in 2016, from the MIT Election Data and Science Lab
- Original: Share of votes in the county that were cast for Donald Trump in 2016, from the MIT Election Data and Science Lab
- No deviation between original and replication study

Income per Capita

- Regression predictor
- Original: 2014-2018 ACS county-level estimate of income per capita
- Replication: 2014-2018 ACS county-level estimate of income per capita
- No deviation between original and replication study

Percent Rural

- Regression predictor
- Original: 2010 Census county-level estimate of the proportion of the county's housing that is rural
- Replication: 2010 Census county-level estimate of the proportion of the county's housing that is rural. Since it appears that the proportion of the *population* that is rural was intended, all analyses will additionally be run using rural population proportion.
- No deviation between original and replication study in the focal analysis. Some observations were dropped in the original analysis due to string mishandling - these are dropped in replication as well.

Percent Male / Black / Hispanic / Has a College Degree / Works in Retail / Works in transportation / Works in Health, Education, or Social Services / age 0-4 / age 5-9 / age 10-14 /

age 15-19 / age 20-24 / age 25-34 / age 35-44 / age 45-54 / age 55-59 / age 60-64 / age 65-74 / age 75-84 / age 85+

- Regression predictor
- Original: ACS county-level estimate of percent of a county that fits the listed description, presumed from 2014-2018
- Replication: 2014-2018 ACS county-level estimate of percent of a county that fits the listed description
- No intended deviation between original and replication, but description of original is imprecise

State

- Regression predictor
- Original: State that each county is in, from ACS
- Replication: State that each county is in, from ACS
- No deviation between original and replication study

Number Staying At Home / Not Staying at Home

- For Calculation of Dependent Variable
- Original: Unacast data was used to calculate the distance traveled from home each day by county
- Replication: BTS data from mobile phones tracks the number of people in the population recorded as staying home or not staying home each day
- The data source is completely different, although in both cases mobile phone data is used to track mobility.

18. Indices

RR TEAM INSTRUCTIONS: *If any of the measured variables described in Section 17 will be combined into a composite measure (including simply a mean), describe in detail what measures you will use and how they will be combined. Please be sure this preregistration includes a link to a clearly commented script that constructs the index according to the narrative.*

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration specify each of the composite measures (e.g. mean scores, factor scores) that are needed for the focal analysis, and which of the measured variables in Section 17 are used in each one (e.g. the happiness, joy, and satisfaction items will be used to create the 'positive feelings' measure)?*
- *Does the preregistration link to a clearly commented script that constructs the indices according to the narrative description?*

Growth in Stay-at-home rates

- Dependent Variable

- The BTS data is at the day-county level, and analysis is at the county level.
- Data is first limited to three time periods: a set of “reference weeks” from February 16-29, 2020, a set of post-COVID weeks from March 19-April 1, 2020, extending the range of the original study slightly so as to cover two full weeks, and (for a supplementary analysis) a “most-recent” set of post-COVID weeks from August 16-29, selected as these are the most recent set of full weeks in the data before missing data becomes more prevalent in the very latest data.
- Then, the “proportion staying home” is calculated in each county-day as (number staying home)/(number staying home + number not staying home)
- Then, proportion staying home is averaged within each county-weekset (referring to the three different two-week sets mentioned above)
- Then, the pre-COVID to post-COVID growth rate (on a -100 to 100 scale) for proportion staying home is calculated for the two post-COVID weeksets, compared to the pre-COVID weekset
- Then the data is reshaped so that there is one observation per county, and two growth rates: one for March and one for August.
- This process can be seen in the analysis script: https://osf.io/eygkv

Analysis Plan

19. Statistical models

RR TEAM INSTRUCTIONS: *This section should describe in detail the analysis that will be performed to replicate the focal result. This analysis must align as closely as possible with the original study’s analysis, even if you have identified limitations in the original study. The level of detail should allow anyone to reproduce your analyses from your description below. Examples of what should be specified: the model; each variable; adjustments made to the standard errors and to case weighting; additional analyses that are required to set up the focal analysis; and the software used.*

Beyond the replication of the focal analysis from the original study, it is at your discretion to test the claim using other analytic approaches as a check of the robustness of the claim. The original test should be listed first and be clearly distinguished from any other tests. If you are testing additional confirmatory hypotheses, describe them in the same order as you numbered them in the “Hypotheses” section above and make clear reference to the specific hypothesis being tested for each.

Please provide a link to a clearly commented script that performs the analysis described in the narrative provided below. Our preference is that this be either an R script or a script from another language that similarly allows for open and reproducible analyses. Please let the SCORE team know if this is not possible.

For each analysis specified in section 15 (and particularly the analyses labeled as 'focal'), please test that the code runs without error on a random subset of 5% of the relevant data. When more than one analysis is listed in section 15, this could require separate 5% samples (e.g. a replication sample and a reproduction sample). Please provide verification that the code has produced sensible results by providing a screenshot(s) of the output (please upload the screenshot(s) to the OSF as well). Finally, please confirm that you have only developed and tested your analysis plan and code using 5% of the dataset (noting that that could be 5% of the replication observations; 5% of the reproduction observations; and/or 5% of the combined observations, as relevant).

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration specify which statistical model will be used to provide the 'focal evidence' for the SCORE test (e.g. a regression coefficient in a larger multiple regression model), and does it correspond closely to the model and evidence from the original study?*
- *Does the preregistration describe each variable that will be included in the focal analysis, and what role each variable has (e.g. dependent variable, independent variable)?*
- *Does the preregistration include a detailed specification of the focal analysis, including interactions, lagged terms, controls, etc., in both narrative form and in a clearly commented script?*
- *Does the preregistration verify that the code runs without error on a random subset of the replication dataset? Is there a separate verification for each analysis specified in section 15?*

This statement confirms that only 5% of the data have been randomly sampled in developing the analysis plan and code contained in this preregistration.

All of the following analyses can be seen in the analysis script: https://osf.io/eygkv

The focal statistical models will be a regression of (dependent variable) growth in proportion home on (independent variables) share of Trump voters, income per capita, and percent Male / Black / Hispanic / Has a College Degree / Works in Retail / Works in transportation / Works in Health, Education, or Social Services / Rural (by housing) / age 0-4 / age 5-9 / age 10-14 / age 15-19 / age 20-24 / age 25-34 / age 35-44 / age 45-54 / age 55-59 / age 60-64 / age 65-74 / age 75-84 / age 85+, with state fixed effects.

This matches the statistical model used in the original analysis. However, notably, in the original study, a higher value of the dependent variable indicated less stay-at-home behavior. In the replication, a higher value indicates more stay-at-home behavior. So the anticipated coefficients reverse sign.

I will run this model using the growth by March. Then, I will calculate the effect of a one-IQR change in Trump vote share on growth in mobility.

Screenshots of these analyses performed using a 5% sample (the regression table is a partial screen shot because it is long). The August analysis pictured will be described as a supplemental analysis:

	March 19-April 1	August 16-29
Income per Capita (Thousands)	0.632 *	0.027
	(0.294)	(0.363)
Share of Trump Voters	0.003	0.056
	(0.091)	(0.112)
Percent Male	-1.384 *	-1.080
	(0.568)	(0.700)
Percent Black	-0.012	0.143
	(0.096)	(0.118)
Percent Hispanic	-0.142	-0.032
	(0.090)	(0.111)
Percent with College Degree	0.124	0.478
	(0.203)	(0.250)
Percent in Retail	0.240	0.368
	(0.288)	(0.356)

```
> summary(glht(m1, paste0(trumpIQR,'*trump_share = 0')))

         Simultaneous Tests for General Linear Hypotheses

Fit: felm(formula = formula_maker("prop_home_change_March",
    flat_data), data = flat_data)

Linear Hypotheses:
                                    Estimate Std. Error z value Pr(>|z|)
0.209767609392536 * trump_share == 0  -0.1643     2.0191  -0.081    0.935
(Adjusted p values reported -- single-step method)

> summary(glht(m2, paste0(trumpIQR,'*trump_share = 0')))

         Simultaneous Tests for General Linear Hypotheses

Fit: felm(formula = formula_maker("prop_home_change_September",
    flat_data), data = flat_data)

Linear Hypotheses:
                                    Estimate Std. Error z value Pr(>|z|)
0.209767609392536 * trump_share == 0   0.7084     2.3372   0.303    0.762
(Adjusted p values reported -- single-step method)
```

20. Transformations

RR TEAM INSTRUCTIONS: *This section should describe how any of the measured variables or composite measures mentioned above will be transformed prior to the analyses listed in Section 19. These are adjustments made to variables **after** measurement or measure creation, and might include centering, logging, lagging, rescaling etc. Please provide enough detail such that anyone else could reproduce the transformations based on the description below. Please be sure this preregistration includes a link to a clearly commented script that performs the transformations described in the narrative provided below.*

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration specify which of the measured variables or composite measures will need to be transformed prior to the focal analysis?*
- *For each variable needing transformation, does the preregistration adequately describe the transformations, including any centering, logging, lagging, recoding, or implementation of a coding scheme for categorical variables?*
- *Does the preregistration link to a clearly commented script that performs each transformation?*

Income per capita is scaled to Income per capita (thousands) by dividing by 1000. The original study does not describe this transformation, but it seems highly likely given the coefficient in the original study.

21. Inference criteria

RR TEAM INSTRUCTIONS: *This section describes the precise criteria that will be used to assess whether the hypotheses listed above were confirmed by the analyses in Section 19. The default language below only applies to the test of the SCORE claim, **H***. *It is at your discretion to describe the inferential criteria you will use for any additional analyses. They need not rely on p-values and/or the same alpha level we have specified for **H***. *Following section 15, if you are performing multiple analyses corresponding to different subsets of the data, please specify whether the same criteria will be used for each analysis (e.g. the same coefficient is expected to be positive and significant in each subset). If the inference criteria differ across analyses, please make that clear below.*

If the additional analyses will use multiple comparisons, the inference criteria is a question with few "wrong" answers. In other words, transparency is more important than any specific method of controlling the false discovery rate or false error rate. One may state an intention to report all tests conducted or one may conduct a specific correction procedure; either strategy is acceptable.

Criteria for a successful replication attempt for the SCORE project is a statistically significant effect (alpha = .05, two tailed) in the same pattern as the original study on the focal hypothesis test (**H***). For this study, this criteria is met by a negative and significant (at the 95% level) coefficient on Trump Vote Share in the March regression.

22. Data exclusion

RR TEAM INSTRUCTIONS: *The section below should describe the rules you will follow to exclude collected cases from the analyses described in Section 19. Note that this refers to exclusions **after** the creation of the replication dataset; exclusion criteria that prevent a case from entering the replication dataset in the first place should be detailed in the 'Data Collection Procedure' section above. Please be as detailed as possible in describing the rules you will follow (e.g. What is the specific definition of outliers you will use? Exactly how many attention checks does a participant need to fail before their removal from the analytic sample?).*

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration comment on whether any cases included in the replication dataset will be excluded prior to data analysis?*
- *If yes, does the preregistration provided detailed instructions on how the exclusions will be performed (e.g. Is the definition of outlier provided? Is the number of attention checks failed before a participant is excluded specified?)*

In the original study, analysis of the original code shows that two counties were probably dropped due to character-processing issues. These observations will also be dropped here, for consistency with the original study.

23. Missing data

RR TEAM INSTRUCTIONS: *The section below should describe how missing or incomplete data will be handled. Please be as detailed as possible in describing the exact procedures you will follow (e.g. last value carried forward; mean imputation) and any software required (e.g. We will use Amelia II in R to perform the imputation).*

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration comment on how missing or incomplete data will be addressed (e.g. casewise removal, missing data imputation)?*
- *If applicable, does the preregistration specify how many missing variables will lead to a case's removal (e.g. If a subject does not complete any of the three indices of tastiness, that subject will not be included in the analysis.)?*
- *If applicable, does the preregistration describe how missing data imputation will be performed, including relevant software?*

Missing data during data construction and analysis is handled by casewise deletion. This will occur if county data cannot be matched to stay-at-home data.

24. Exploratory analysis (Optional)

RR TEAM INSTRUCTIONS: *If you plan to explore your data set to look for unexpected differences or relationships, you may describe those tests here. An exploratory test is any test where a prediction is not made up front, or there are multiple possible tests that you are going to use. A statistically significant finding in an exploratory test is a great way to form a new confirmatory hypothesis, which could be registered at a later time. If any exploratory analyses involve additions to the data collection procedure beyond what was performed in the original study (e.g. additional items on the survey; running another condition in the experiment), please describe them below.*

To see whether the result persists over time, I will repeat the analysis using growth to August 16-29 rather than to March 19-April 1 (screenshot above). Then, to test the robustness of the results, I will add a control for spatial autocorrelation. Both stay-at-home behavior and Trump support are strongly geographically clustered, in a way that is unlikely to be handled by fixed effects for state. I calculate county neighbors using 5-nearest-neighbors clustering on county centroid latitude and longitude. Then, I estimate a standard spatial autocorrelation model, repeating the original analysis but with a spatial autocorrelation term. Screenshot:

	March 19-April 1	August 16-29
Income per Capita (Thousands)	0.779 ***	0.225
	(0.232)	(0.287)
Share of Trump Voters	0.008	0.039
	(0.069)	(0.083)
Percent Male	-1.516 ***	-1.205 *
	(0.420)	(0.516)
Percent Black	0.027	0.149
	(0.074)	(0.089)
Percent Hispanic	-0.154 *	-0.047
	(0.073)	(0.089)
Percent with College Degree	0.046	0.344
	(0.160)	(0.198)
Percent in Retail	0.206	0.351
	(0.216)	(0.266)

Additionally, because the usage of “percent of housing that is rural” appears to have been an error in the original study, **all** analyses will be rerun using “percent of population that is rural”. Screenshots:

	March 19-April 1	August 16-29
Income per Capita (Thousands)	0.632 *	0.027
	(0.294)	(0.363)
Share of Trump Voters	0.003	0.056
	(0.091)	(0.112)
Percent Male	-1.384 *	-1.080
	(0.568)	(0.700)
Percent Black	-0.012	0.143
	(0.096)	(0.118)
Percent Hispanic	-0.142	-0.032
	(0.090)	(0.111)
Percent with College Degree	0.124	0.478
	(0.203)	(0.250)
Percent in Retail	0.240	0.368
	(0.288)	(0.356)

	March 19-April 1	August 16-29
Income per Capita (Thousands)	0.765 **	0.205
	(0.233)	(0.288)
Share of Trump Voters	0.007	0.038
	(0.069)	(0.083)
Percent Male	-1.496 ***	-1.165 *
	(0.426)	(0.523)
Percent Black	0.025	0.147
	(0.074)	(0.089)
Percent Hispanic	-0.151 *	-0.045
	(0.073)	(0.089)
Percent with College Degree	0.052	0.352
	(0.160)	(0.198)
Percent in Retail	0.210	0.353
	(0.216)	(0.266)

25. Other

RR TEAM INSTRUCTIONS: *This section serves two purposes. First, please use this section to discuss any features of your replication plan that are not discussed elsewhere. Literature cited, disclosures of any related work such as replications or work that uses the same data, plans to make your data and materials public, or other context that will be helpful for future readers would be appropriate here. Second, please also re-surface any major deviations from earlier in*

the preregistration that you expect a reasonable reviewer could flag for concern. Give a summary of these deviations, focusing on larger changes and any possible challenges for comparing the results of the original and replication study.

Specific points to keep in mind (please also consult the Reviewer Criteria):

- *Does the preregistration reference other sections of the preregistration where substantial deviations from the original study have been described (including deviations due to differences in location or time compared to the original study)?*
- *Does the preregistration comment on plans to make the data and materials from the replication study public?*

Final review checklist

REVIEWER INSTRUCTIONS: *For the following questions, reviewers please indicate whether you can 'sign off' on the following items by adding a comment. You can update this response as the lab moves through revisions during the review period!*

- Included in this pre-registration are specific materials needed to create a replication dataset:
    - Is the final replication dataset that the research team constructed suitable for performing a high-quality, good-faith replication of the focal claim selected from the original study?
    - Is the procedure for constructing the final replication dataset sufficiently documented that an independent researcher could construct the same dataset following the procedures and code they lay out?
- Included with this pre-registration is a narrative description of how the replication dataset will be used to perform the focal replication analysis, as well as the specific analytic scripts/code/syntax that will be used:
    - Is the analysis plan (including code) that's documented in the preregistration consistent with a high-quality, good-faith replication of the focal claim selected from the original study?
    - Has the data analyst demonstrated that the analysis code works as expected on a random 5% of the final replication dataset?
- I have reviewed all sections of this pre-registration, and I believe it represents a good-faith replication attempt of the original focal claim.

Additionally, please consider the following if the preregistration includes a reproduction analysis:

- The observations used for the reproduction analysis were collected and measured in the same way as the original study.
- The observations used for the reproduction analysis were analyzed in the same way as the original study.
- The data analyst has demonstrated that their analysis code works as expected on a random 5% of the reproduction data.
- I believe this preregistration represents a good-faith reproduction attempt of the original focal claim.