README for "Opinions as Facts"

This replication package contains the data and code of "Opinions as Facts". To replicate the full analysis creating all figures and table, run the bash files in the main folder in the order indicated by the file names. Note that proprietary data has been replaced with similar-looking datasets with shuffled values and added noise, and that therefore some figures and table will look different.

Data Availability

The data to replicate the figures and tables is provided in this replication package with a few exceptions. All raw data provided in `data/raw` is described in detail. In addition, we provide all GIS datasets used in a separate folder `data/GIS`.

Data collected by the authors   The authors collected data for this project through three surveys. The first survey asks Fox viewers aged 55 and above about their behavior in the early stage of the pandemic. The second survey with 1,000 respondents measures Fox and MSNBC viewers' trust in opinions. The second survey with 13,744 respondents elicits beliefs about election fraud. Additionally, the authors recruited individuals on Amazon MTurk to code TV show transcripts. All data collected by the authors is provided in this replication package.

Publicly available data   Most data used for this project is publicly available. The following list provides an overview of the publicly available data sources:

- Sunset times of counties: sunset times for counties (based on centroids) were downloaded from www.timeanddate.com
- COVID-19 cases and death: daily county-level data on COVID-19 cases and deaths are provided by the Johns Hopkins University and available from https://github.com/CSSEGISandData/COVID-19
- Social distancing: stay-at-home behavior data from the Bureau of Transportation Statistics (BTS) is available from https://www.bts.gov/browse-statistical-products-and-data/trips-distance/daily-travel-during-covid-19-pandemic
- County-level data on US population demographics and attitudes, and county characteristics from various sources:
    - ACS2018: census data on education, age, and racial composition of the population (available from https://www.census.gov/programs-surveys/acs or the US census data explorer)
    - Behavioral Risk Factor Surveillance System (BRFSS): physical health and health insurance (available from https://www.countyhealthrankings.org/explore-health-rankings/rankings-data-documentation/national-data-documentation-2010-2019)
    - Bureau of Labor Statistics (BLS): unemployment rate (available from https://www.bls.gov/lau/#tables)

- Bureau of Transportation Statistics (BTS): airport locations (available from https://data.bts.gov/Research-and-Statistics/County-Transportation-Profiles/qdmf-cxm3)
- CCES 2019: attitudes toward women and minorities (available from https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/WOT7O8)
- CCES preferences: political preferences (available from https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/OSXDQO)
- Dartmouth Atlas of Health Care's Hospital and Physician Capacity: health capacity, e.g. hospital personnel and beds (available from https://data.dartmouthatlas.org/capacity/ and https://data.dartmouthatlas.org/supplemental/)
- MIT Election Lab: voting in the 2016 presidential election (available from https://electionlab.mit.edu/data)
- Small Area Income and Poverty Estimates (SAIPE): census data on income and poverty (available from https://www.census.gov/programs-surveys/saipe.html)
- US Census: population in 2019 (available from https://www.census.gov/data/tables/time-series/demo/popest/2010s-counties-total.html); land area in 2010 (available from the census website); urban-rural classification (available from https://www.census.gov/programs-surveys/geography/guidance/geo-areas/urban-rural.html); city population and location (available from https://www2.census.gov/programs-surveys/popest/datasets/2010-2020/cities/ and https://www2.census.gov/geo/docs/maps-data/data/gazetteer/2020_Gazetteer/)
- US Department of Transportation's Federal Highway Administration: miles of highways (available from https://data-usdot.opendata.arcgis.com/datasets/national-highway-planning-network)
- Yale Climate Opinion Maps: opinions about climate change (https://climatecommunication.yale.edu/visualizations-data/ycom-us/)

Proprietary data Three sets of data are proprietary and can therefore not be shared in this replication package. First, data on TV viewership was bought from Nielsen. Second, data on population characteristics from Gallup. In order to being able to run the code using Nielsen and Gallup data, similar looking files with random values are provided in place of the actual data. Third, the authors downloaded TV show transcripts from LexisNexis. The replication package does not contain the raw transcript data, but it provides the processed datasets: word counts by TV show and date (in `data/raw/transcripts`); similarity of shows (in `data/raw/NLP`). Additionally, we provide the code to derive these processed datasets in `additional-code`.

Dataset list

The following table provides an overview of the data files described above. All files are in the `raw` data folder unless indicated otherwise in the last column.

Data file	Source	Provided
transcripts/transcript-word-counts.csv	constructed from proprietary data	Yes
transcripts/mturk1.csv to mturk6.csv	collected by the authors	Yes
trust-survey/trust_fox_choicetext.csv	collected by the authors	Yes
trust-survey/trust_msnbc_choicetext.csv	collected by the authors	Yes
behavior-survey/survey.csv	collected by the authors	Yes
behavior-survey/coding_qualitative.csv	collected by the authors	Yes
election-survey/survey.dta	collected by the authors	Yes
sunset/counties_sunset_2020jan-mar.csv	ww.timeanddate.com	Yes
cases-deaths/JHU_US_confirmed.dta	John Hopkins University	Yes
cases-deaths/JHU_US_deaths.dta	John Hopkins University	Yes
BTS/Trips_by_Distance.csv	BTS	Yes
nielsen/nielsen-2018.csv	Nielsen	No
nielsen/nielsen-jan-mar-2020.csv	Nielsen	No
gallup/G1K_FINAL_2015_SCI_DATA.dta	Gallup	No
ACS2018/ACSDP5Y2018.DP05	ACS 2018	Yes
ACS2018/ACSST5Y2018.S1501	ACS 2018	Yes
BLS/laucntycur14.xlsx	BLS	Yes
BRFSS/analytic_data2017.csv	BRFSS	Yes
CCES2019/CCES19_Common_OUTPUT.dta	CCES	Yes
CCES-preferences/cumulative_cces_policy_preferences.dta	CCES	Yes
CCES-preferences/cumulative_2006-2020.dta	CCES	Yes
climatechange/YCOM_2020_Data.csv	Yale Climate Opinion Maps	Yes

Data file	Source	Provided
dartmouth/2012_hosp_resource_hrr.xls	Dartmouth Atlas of Health Care	Yes
dartmouth/ZipHsaHrr12.xls	Dartmouth Atlas of Health Care	Yes
population/DEC_10_SF1_GCTPH1.US05PR_with_ann.csv	US Census	Yes
population/co-est2019-alldata.csv	US Census	Yes
rural/County_Rural_Lookup.xlsx	US Census	Yes
SAIPE/est18all.xlsx	SAIPE	Yes
transportation/2020_Gaz_counties_national.txt	US Census	Yes
transportation/bts_county_transportation_profiles.csv	BTS	Yes
transportation/2020_Gaz_place_national.txt	US Census	Yes
transportation/SUB-EST2020_ALL.csv	US Census	Yes
transportation/National_Highway_Planning_Network.csv	US Department of Transportation	Yes
voting/countypres_2000-2016.csv	MIT Election Lab	Yes

Computational requirements

Software Requirements

This section lists all programs and packages necessary to run the code in this replicate package. The packages can be installed by running the file `setup.sh`, which runs the setup files for Stata, R, and Python.

- Stata (code was last run with version 16)
    - `gtools` 1.5.1
    - `keeporder` 1.0
    - `vincenty` 1.0.3
    - `estout` 3.30
    - `ivreg2` 4.1.11
    - `ranktest` 2.0.04

- R 3.4.3
    - `tidyverse` 1.3.1
    - `haven` 2.4.3
    - `lfe` 2.8-7.1
    - `stargazer` `booktabs`
    - `starpolishr` 0.0.0.9004
    - `lubridate` 1.8.0
    - `cowplot` 1.1.1
    - `broom` 0.7.11
    - `extrafont` 0.17
    - `ggpubr` 0.4.0
    - `zoo` 1.8-9
    - `RColorBrewer` 1.1-2
    - `starbility` 0.0.0.9000
    - `fixest` 0.10.4
    - `pscl` 1.5.5
    - `statar` 0.7.3
    - `doParallel` 1.0.16
    - `gridExtra` 2.3
    - `tmap` 3.3-2
    - `sf` 1.0-5
    - `cutr` 0.0.0.900
    - `splines` 4.0.4
    - `devtools` 2.4.3

    These packages require the following dependencies, which will be loaded automatically: `ggplot2` 3.3.5, `tibble` 3.1.6, `tidyr` 1.1.4, `readr` 2.1.1, `purrr` 0.3.4, `dplyr` 1.0.7, `stringr` 1.4.0, `forcats` 0.5.1 `multcomp` 1.4-18, `mvtnorm` 1.1-3, `survival` 3.2-13, `TH.data` 1.1-0, `MASS` 7.3-55, `Matrix` 1.4-0, `foreach` 1.5.1, `iterators` 1.0.13, `parallel` 4.0.4.

- Python 3.9.7
    - `numpy` 1.20.3
    - `pandas` 1.2.4
    - `scipy` 1.7.1
    - `datetime` 4.4

Memory and Runtime Requirements

The code was last run on a **4-core Intel-based laptop with Windows 10 version 21H2** with 16 GB of RAM. Computation took approximately 30 hours. Excluding `code/analysis/cases-deaths/timeseries-randomization.R`, which creates Figures D3, D4, and D5, reduces the time to approximately 1 hour. The additional code in the folder `additional-code`, which can only be run when downloading TV show transcripts from LexisNexis, took approximately 15 hours.

Description of programs

The bash files in the main folder run the whole analysis from raw data to creating all figures and tables:

- `0-setup.sh` creates the folder structure and installs all Stata, R, and Python packages required.
- `1-clean.sh` cleans all raw data files and prepares them for later use.
- `2-construct.sh` combines the prepared data files to the main files used for the analysis.
- `3-analyze.sh` runs the analysis that creates all figures and tables.

Instructions to Replicators

Run the bash files in the main folder in the order indicated by the file names:

- Run `0-setup.sh` to create the folder structure and installs all Stata, R, and Python packages required. This bash file runs all code in `code/setup`. It only has to be run once.
- Run `1-clean.sh` to clean all raw data files and prepare them for later use. The bash files run all code in `code/cleaning` and stores the cleaned data files in `data/working`. This code can be run in any order.
- Run `2-construct.sh` to combine the cleaned data files to the main files used for the analysis. The bash file runs all code in `code/construct` and stores the data files in `data/working`. The order of this code is important.
- Run `3-analyze.sh` to create all figures and tables. This bash file runs all code in code in `code/analysis` and stores the figures and tables in `output/figures` or `output/tables`, respectively.

List of tables and programs

The following table lists all figures and tables with the programs that create them. The programs can be found in `code/analysis` and all figures and tables are stored in `output`.

Figure/Table	Program	Output file
Figure 1	trust/trust-figure-table.R	figures/trust.png
Figure 2, Panel A	transcripts/show-content-cnn-msnbc.R	figures/fox-word-count.png
Figure 2, Panel B	transcripts/show-content-fox.R	figures/fox-seriousness.png
Figure 3, Panel A	behavior/behavior-figures-tables.R	figures/behavior-density.png
Figure 3, Panel B	behavior/behavior-figures-tables.R	figures/behavior-timeseries.png
Figure 4	cases-deaths/timeseries-figures.R	figures/OLS-timeseries.png
Figure 5	iv-intuition/iv-map.R	figures/iv-resid-map.png
Figure 6	cases-deaths/timeseries-figures.R	figures/V1-timeseries.png
Figure 7	cases-deaths/timing-figure.R	figures/V1-timing.png
Figure 8, Panel A	election/election-figures-tables.R	figures/election-allviewers.png
Figure 8, Panel B	election/election-figures-tables.R	figures/election-foxviewers.png
Figure A1, Panel A	nlp/nlp-figures.R	figures/topic-aggregate-2019.png
Figure A1, Panel B	nlp/nlp-figures.R	figures/bert-aggregate-2019.png
Figure A10	cases-deaths/timeseries-figures.R	figures/V1-timeseries-state-clustering.png
Figure A11	cases-deaths/timeseries-figures.R	figures/V1-timeseries-leaveout.png
Figure A12	distancing/social-distancing.R	figures/social-distancing.png
Figure A13	election/election-figures-tables.R	figures/election-stability.png
Figure A2	transcripts/show-content-cnn-msnbc.R	figures/cnn-msnbc-word-count.png
Figure A3	behavior/behavior-categories-figure.R	figures/behavior-categories.png
Figure A4	instrument-validation/exogeneity-figures.R	figures/OLS-exogeneity.png
Figure A5	cases-deaths/timeseries-figures.R	figures/OLS-timeseries-state-clustering.png
Figure A6	cases-deaths/stability-figures.R	figures/OLS-V1-stability.png

Figure/Table	Program	Output file
Figure A7, Panel A	cases-deaths/stability-figures.R	figures/oster-cases.png
Figure A7, Panel B	cases-deaths/stability-figures.R	figures/oster-deaths.png
Figure A8	instrument-validation/exogeneity-figures.R	figures/V1-exogeneity.png
Figure A9	instrument-validation/exogeneity-figures.R	figures/V1-similarity.png
Figure C1	nlp/nlp-figures.R	figures/topic-separate-2019.png
Figure C2	nlp/nlp-figures.R	figures/bert-separate-2019.png
Figure C3	nlp/nlp-figures.R	figures/topic-separate-2020.png
Figure D1, Panel A	cases-deaths/residuals-figures.R	figures/V1-residuals-cases.png
Figure D1, Panel B	cases-deaths/residuals-figures.R	figures/V1-residuals-deaths.png
Figure D2	cases-deaths/timeseries-figures.R	figures/V1-timeseries-unbalanced.png
Figure D3	cases-deaths/timeseries-randomization.R	figures/V1-bootstrap.png
Figure D4	cases-deaths/timeseries-randomization.R	figures/V1-ri.png
Figure D5	cases-deaths/timeseries-randomization.R	figures/V1-permutation.png
Figure D6, Panel A	iv-intuition/iv-intuition-figures.R	figures/iv-intuition-start-country.png
Figure D6, Panel B	iv-intuition/iv-intuition-figures.R	figures/iv-intuition-start-tz.png
Figure D7	instrument-validation/exogeneity-figures.R	figures/V2-exogeneity.png
Figure D8	instrument-validation/exogeneity-figures.R	figures/V3-exogeneity.png
Figure E1, Panel A	epi-model/simulations-figure.R	figures/OLS-simulation-betafigure.png
Figure E1, Panel B	epi-model/simulations-figure.R	figures/OLS-simulation-treatment.png
Figure E2, Panel A	epi-model/simulations-figure.R	figures/V1-simulation-betafigure.png
Figure E2, Panel B	epi-model/simulations-figure.R	figures/V1-simulation-treatment.png
Table A1	behavior/behavior-representativeness.do	tables/behavior-representativeness.tex
Table A2	behavior/behavior-figures-tables.R	tables/behavior-characteristics.tex
Table A3	election/election-figures-tables.R	tables/election-characteristics.tex
Table A4	instrument-validation/first-stage-table.R	tables/V1-first-stage.tex
Table A5	cases-deaths/count-models.R	tables/V1-count-models.tex
Table D1	cases-deaths/iv-variations-table.R	tables/2SLS-robustness.tex

Figure/Table	Program	Output file
Table D2	cases-deaths/misinformation-table.R	tables/V1-misinformation.tex
	instrument-validation/instrument-variations-fstats.do	input for cases-deaths/iv-variations-table.R
	epi-model/run-model.py	input for epi-model/simulations-figure.R
	epi-model/magnitudes.R	input for epi-model/simulations-figure.R