Replication Package for "Prison Connectivity and Disease Transmission to Neighboring Communities: The Role of Prison Staff"

Overview

This replication package includes the datasets and Stata code necessary to reproduce results from the paper "Prison Connectivity and Disease Transmission to Neighboring Communities: The Role of Prison Staff." The analysis primarily utilizes confidential smartphone geolocation data from Veraset, which we cannot publicly distribute due to privacy restrictions. Researchers interested in the raw ping-level data should contact Veraset directly here. The publicly shareable final analysis datasets, code, and supplementary files needed for replicating the results presented in the paper and appendix are provided in this package.

All analysis was performed using **Stata MP 17**. Ensure the following Stata packages are installed: `gtool`, `geoinpoly`, `csdid`, `drdid`, and `event_plot`.

Directory Structure

1. `raw_data`

This folder contains publicly distributable datasets used in the analysis.

File Name	Description
`bg_boundaries/`	Boundary coordiantes for California Census Block Groups. Includes `tl_2019_06_bg_coor.dta` (Coordinates) and `tl_2019_06_bg_data.dta` (Attributes).
`CDCR_covid_19.csv`	COVID-19 case counts within California Department of Corrections and Rehabilitation (CDCR) facilities.
`LODES_data/ca_od_main_JT01_2020.csv`	LEHD Origin-Destination Employment Statistics (LODES) data from the U.S. Census Bureau for primary job (JT01).
`LODES_data/zcta_tract_rel_10.txt`	Crosswalk file mapping ZCTAs to Census tracts.
`monthly_cdcr_population.csv`	Monthly prison population counts from CDCR prisons.
`nhgis0022_ds254_20215_zcta_codebook.txt`	Codebook for ZCTA-level demographic data from NHGIS.
`nhgis0022_ds254_20215_zcta.csv`	ZCTA-level demographic and socioeconomic data (ACS 2017-2021) obtained from NHGIS.
`prison_boundaries/`	Prison boundaries data. Includes: `prison_boundaries/Prison_Boundaries_data.dta`, `prison_boundaries/Prison_Boundaries_CA_coor.dta`, and `prison_boundaries/prison_boundaries_correct_projection.shp (dbf/cpg/prj/qpj/shx)`.
`Prison_Boundaries.dta`	Prison boundary attribute data, classified to different samples for analysis.
`smartphone_contact_matrix/`	Processed smartphone-derived contact data between prisons and ZCTAs. Includes: `prison_time_from_Jan-OctHome_ZCTA.dta`, `prison_time_from_Jun-OctHome_ZCTA.dta`, `prison_time_from_Jun-OctHome_ZCTA_DropOnePinginJune.dta`
`ZCTA_acs_15_19/`	ACS (2015-2019) socioeconomic variables aggregated to the ZCTA level. Includes: `ZCTA_demographics_acs_15_19.dta`

File Name	Description
`Zipcode_PRA_18185-032923.xlsx`	COVID-19 Case Rate per 100,000 People by Zipcode, Month, and Year in California 2020 from the California Department of Public Health (CDPH).
`ZIPCodetoZCTACrosswalk2021UDS.xlsx`	Crosswalk file mapping ZIP codes to ZCTAs.
`92214_QCEW_BLS.csv`	County-level employment data for Correctional Institutions (NAICS 92214), sourced from the BLS Quarterly Census of Employment and Wages.

2. `intermediate_data`

Contains intermediate datasets generated during the data processing stage (these files are created by provided scripts).

3. `analysis_data`

Contains final datasets ready for statistical analysis (generated by provided scripts).

4. `code`

This directory contains all Stata `.do` files necessary for replication:

- `master.do`: Defines local directories and runs all analysis scripts in sequence.
- `prepare_analysis_data.do`: Processes raw data to generate analysis-ready datasets.
- `figure*.do`: Scripts that generate figures presented in the main paper and appendix.
- `produce_appendix_tables.do`: Generates tables for the appendix.
- `read_prison_data.do`: Imports and processes prison-related data, invoked by `prepare_analysis_data.do`.
- `process_lodes_data.do`: Processes LODES data, invoked by `prepare_analysis_data.do`.

5. `output`

This directory stores the figures and tables produced by the analysis scripts.

Replication Instructions

1. Ensure you have Stata MP 17 and the necessary packages installed.
2. Execute the `master.do` script in Stata, which sequentially runs all required analysis scripts.
3. Outputs (figures and tables) will be saved to the `Output` directory.