README

Replication Data for: 'COVID-19 Vaccine Rollouts and the Reproduction of Urban Spatial Inequality: Disparities Within Large U.S. Cities in March and April 2021 by Racial/Ethnic and Socioeconomic Composition'

Nicholas V. DiRago

1 November 2021

This repository contains materials to replicate the spatial interpolation, analyses, and simulations in DiRago et al., "COVID-19 Vaccine Rollouts and the Reproduction of Urban Spatial Inequality: Disparities Within Large U.S. Cities in March and April 2021 by Racial/Ethnic and Socioeconomic Composition."

Full details on the data and methods are available in the online supplement to the study.

File descriptions

R scripts (code)

Scripts assume all files in the repository are downloaded to the same folder. Users should update their working directory to that folder, as indicated in the scripts.

- **base.R** defines functions, file paths, and objects that are called in other scripts.
- **overlay_interpolation.R** reproduces the overlay interpolation (see Section e3.2 of the supplement). This script is memory-intensive. It ran in just under 1.5 hours on a single-core machine with 128 GB of memory.
- **models.R** reproduces the spatial error models (SEMs) and corresponding standard linear models (SLMs), including robust standard errors (see Sections e4.1 and e4.2 of the supplement). It also produces Figure e4.1.
- **simulations.R** reproduces the simulations (see Section e4.3 of the supplement). Designed to be run in parallel, this script is memory-intensive. It ran in just under 2.5 hours on a 16-core machine with 12 GB of memory per core.

Data files

- **analysis_data.csv** contains observations for each unit of analysis on the variables used in the models and simulations.
- **units.geojson** contains geospatial vectors for the areal boundaries of the units of analysis (see Section e3.2 of the supplement).
- **W.txt** is a comma-delimited text file representing the spatial weights matrix (W) used in the main analyses (see Sections e4.1 and e4.2 of the supplement). The row and column names identify units of analysis.
- **W_boot.zip** contains one comma-delimited text file that combines the spatial weights matrices for the 1,000 bootstrapped resamples used in the simulations (see Section e4.3 of the supplement).
- **vax_counts.csv** contains counts of individuals with at least one vaccine dose, aggregated by "ZIP" (ZIP Code or equivalent; see below). These data were collected directly from public sources (see Table e2.1 in the supplement).

Variable descriptions

The suffices "est" and "moe" to the roots listed below denote estimates or margins of error, respectively. The suffices "mar," "apr," and "dif" denote values for March, April, and the difference between March and April, respectively. All percentages are coded as proportions (scaled from zero to one).

Name or root	Description
area_in	The area (in square kilometers) of the intersection of the Census block group (CBG) and "ZIP" (see below)
bg	The 12-digit unique identifier (GEOID) for the Census block group (CBG)
boot_rep	Unique identifier (integer) for bootstrap iterations
city	Name of city in which unit is located
geo	Unique identifier for units of analysis: concatenation of 1) lowercase IATA code for largest airport serving city containing unit and 2) unit-specific three-digit identifier
month	Name of the month (March or April) to which the observation corresponds
mult_bg	The share of the area of the Census block group (CBG) that falls within the "ZIP" (see below)
n_15o	Population age 15+
n_tot	Total population
n_vac_pf	Number of individuals with at least one vaccine dose
pr_15o_vac_pf	Percent individuals age 15+ with at least one vaccine dose
pr_65o	Percent age 65+
pr_a	Percent Asian
pr_b	Percent Black
pr_hig_mtp	Percent with Medicaid or other means-tested public health insurance
pr_hig_ui	Percent without health insurance
pr_hsa	Percent employed in health care or social assistance
pr_l	Percent Hispanic
pr_ni	Percent without internet access
pr_pov	Percent under poverty line
pr_w	Percent White
state	Two-letter abbreviation for state in which unit is located
zip	The five-digit identifier for the ZIP Code Tabulation Area (ZCTA) or New York City Modified ZCTA (MODZCTA)