Documentation for Do-files and Datasets

Replication files for "Neighborhoods Matter: Assessing the Evidence for Place Effects"

Eric Chyn[1], and Lawrence F. Katz[2]

Files last updated: August 11, 2021

Project ID: openicpsr-142621

Overview

The code in this replication package runs all the analyses in Chyn and Katz (2021) using Stata, R, and the data provided. For Stata, a master file runs all code to create Figure 1 (which has 4 panels) and Table 1 in the paper. The replicator should expect this code to run for about 8 seconds. Figure 2 is a series of maps that were created in R by the Opportunity Insights team using the code and data included in this package.

Data Availability

All data used in the paper is provided in the replication package. The data was assembled by the authors from publicly available sources (documented below). The archive for this paper contains all data necessary for our analysis. The data included for the Stata analysis has detailed labels for all variables.

Data Sources

1. Health Inequality Project data: The Health Inequality Project provides public-data with information on life-expectancy at the commuting zone (CZ) level.

   Currently, the data can be downloaded from: https://www.healthinequality.org/data/

2. Opportunity Atlas data: The Opportunity Atlas provides public-data with information on upward mobility at the census tract level. Note that the Opportunity

---

[1] Dartmouth College and NBER
[2] Harvard University and NBER

Atlas files also include U.S. Census Bureau statistics from the 2000 decennial census as part of the files available for download.

Currently, the data can be downloaded from: https://www.opportunityatlas.org/

3. Stanford Education Data Archive (SEDA) data: SEDA provides public-data with information on standardized test scores at the school district level. Note that the SEDA files also include Census statistics from the 2007-2011 and 2012-2016 American Community Surveys as part of the SEDA files available for download. The data are listed as being licensed under the Creative Commons Attribution-Non-commercial-Share Alike 3.0 Unported License.

Currently, the SEDA data can be downloaded from: https://purl.stanford.edu/db586ns4974

Description of the Data

The cleaned Stata data files that we use for our analysis are contained in the folder "Replication_Data." The data were created by processing data files downloaded from the websites listed above. The datasets are unique at different geographic levels (i.e., Commuting Zones, School Districts, or Census Tracts). We provide a full list of all the datasets we use at the end of this document. The variables in all Stata files are labelled.

Note that there are two data files necessary to create the maps for Figure 2. These files are in CSV format and are included in the "Maps" directory. The sources of these data are the same as those listed above.

Reuse and Citation

Please cite the original sources if you are re-using the data provided in this replication package. Below we list the several suggested citations for the data and relevant publications associated with the data:

SEDA data citation:

Fahle, E. M., Shear, B. R., Kalogrides, D., Reardon, S. F., Chavez, B., & Ho, A. D. (2018). "Stanford Education Data Archive: Dataset (Version 3.0)". Retrieved

from http://purl.stanford.edu/db586ns4974 (accessed December 12, 2019).

Health Inequality paper and data citation:

Chetty, Raj, Michael Stepner, Sarah Abraham, Shelby Lin, Benjamin Scuderi, Nicholas Turner, Augustin Bergeron, and David Cutler. 2016. "The Association Between Income and Life Expectancy in the United States, 2001-2014." JAMA 315 (16): 1750.

Chetty, Raj, Michael Stepner, Sarah Abraham, Shelby Lin, Benjamin Scuderi, Nicholas Turner, Augustin Bergeron, and David Cutler. 2016. "The Health Inequality Project: Dataset" Retrieved from https://www.healthinequality.org/data/ (Accessed January 17, 2020).

Opportunity Atlas paper and data citation:

Chetty, R., Friedman, J., Hendren, N., Jones, M., and Porter, S. (2020). "The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility." Opportunity Insights, Harvard University, January.

Chetty, R., Friedman, J., Hendren, N., Jones, M., and Porter, S. (2020). "The Opportunity Atlas: Mapping the Childhood Roots of Social Mobility: Dataset." Retrieved from https://www.opportunityatlas.org/ (Accessed November 11, 2019).

General Description of the Code

The "Code" folder consists of two do-files. The file "master.do" sets directories for accessing the data, storing figures and tables, and calls the program "analysis.do" which creates the Figure 1 and Table 1 in the paper. Note that the R code necessary to create the maps for Figure 2 is included in a separate folder called "Maps."

General Organization of Top Folder

1. Code
2. Figures
3. Replication_Data

4. Tables
5. Maps

Note: The code will automatically create the "Figures" and "Tables" folders if they are not present.

Replication Instructions:

To replicate the results in "Neighborhoods Matter: Assessing the Evidence for Place Effects", you must install Stata (for Figure 1 and Table 1) and Base R (for Figure 2). Stata must be purchased and is available at https://download.stata.com/download/. Base R is available for free at https://cloud.r-project.org/.

Stata Instructions

Once Stata is installed, run the "master.do" file after modifying the directories in the file. All the data provided for this replication package must be stored in the appropriate directories. The "master.do" file will automatically install the "binscatter.ado" and "outreg2.ado" packages if these are not already installed. It will also call the do-file "analysis.do", which will create Table 1 and Figure 1 in the paper.

The last replication was run in Stata 16.1 on a Windows 10 Enterprise version 10.0.19042 Build 19042 computer.

R Instructions

Figure 2 in the paper is a series of maps created by the team at Opportunity Insights using R. For replication purposes, the code and data used to create the maps is included in a folder called "Maps."

Once R is installed, open the program and load the R-program "atlas_maps" included in the Maps directory. Run this program after modifying the directories at the top to specify the location of the Maps directory and entering your Census API key. A Census API key can be obtained from https://api.census.gov/data/key_signup.html.

The "atlas_maps" program requires the following libraries: Rcpp; sf; maggritr; leaflet; mapview; webshot; dplyr; tidyverse; tidycensus; conflicted. The top of the program will install these packages.

Note that the "atlas_maps" program must be run from the same location as the program "simple_shot" which is included in the replication code. In addition, the "atlas_maps" program uses two CSV files to generate maps: "r_chicago_2000census.csv" and "r_chicago_atlas.csv." The sources of these data are detailed above, and detailed descriptions of these datasets are provided at the end of this document.

Once the "atlas_maps" program has been run completely, it will generate maps that are included as Figure 2 in the paper. Labels and a legend must be manually added to these maps.

Contact:

For any further questions, contact Eric Chyn (eric.t.chyn@dartmouth.edu), or Lawrence F. Katz (lkatz@harvard.edu.).

Full List of Computational Requirements

- Stata (code was last run with version 16.1)
    - binscatter (as of 2020-06-09)
    - The analysis.do file will install the program above.
- R x64 4.0.3
    - Rcpp_1.0.7 (as of 2021-08-10)
    - sf_1.0-2 (as of 2021-08-10)
    - maggritr_2.0.01 (as of 2021-08-10)
    - leaflet_2.0.4.1 (as of 2021-08-10)
    - mapview_2.10.0 (as of 2021-08-10)
    - webshot_0.5.2 (as of 2021-08-10)
    - dplyr_1.0.7 (as of 2021-08-10)
    - tidyverse_1.3.1 (as of 2021-08-10)
    - tidycensus_1.0 (as of 2021-08-10)
    - conflicted_1.0.4 (as of 2021-08-10)

Full List of Datasets with Brief Descriptions

1. cz_adult_emp_pov.dta: Data at the CZ-level with adult employment and poverty rates (both from the 2000 decennial Census). Used for Table 1 and Figure 1.
2. cz_life_expect.dta: Data at the CZ-level with life expectancy (from the Health Inequality Project) and poverty rates (from the 2000 decennial Census). Used for

Table 1 and Figure 1.

3. cz_upward_mobility_pov: Data at the CZ-level with upward mobility (from the Opportunity Atlas) and poverty rates (from the 2000 decennial Census). Used for Table 1 and Figure 1.

4. district_test_scores_pov: Data at the school district level with test scores (from SEDA) and poverty rates (from the 2007-2011 and 2012-2016 American Community Surveys). Used for Table 1 and Figure 1.

5. tract_adult_emp_pov: Data at the tract-level with adult employment and poverty rates (both from the 2000 decennial Census). Used for Table 1 and Figure 1.

6. tract_upward_mobility_pov: Data at the tract-level with upward mobility (from the Opportunity Atlas) and poverty rates (from the 2000 decennial Census). Used for Table 1 and Figure 1.

7. r_chicago_2000census.csv: Data at the tract-level with median household income and the adult employment rate (both from the 2000 decennial Census). Used for Figure 2.

8. r_chicago_atlas.csv: Data at the tract-level with upward mobility (from the Opportunity Atlas). Used for Figure 2.