Read me for Flight from urban blight: lead poisoning, crime and suburbanization

Federico Curci and Federico Masera

This document provides replication instructions for the analysis in "Flight from urban blight: lead poisoning, crime and suburbanization" by Federico Curci and Federico Masera in the *Review of Economics and Statistics*.

Software requirements

The software used is Stata 16 and QGIS 3 and the operating system Microsoft Windows 10.

Runtime requirements

The estimated programs runtime to produce all the tables and figure of the paper and appendix can be expected to last around 1 and 45 minutes, respectively.

Instructions to replicators

Be sure to change the working directory.

The replicator needs to install the following additional Stata modules:

1. reghdfe
2. ivreghdfe
3. outreg2
4. grstyle
5. _gwtmean
6. mediation
7. shp2dta

We also include an additional Stata module, called *discretiz*, in the folder *analysis/appendix*

Folder descriptions

1. *analysis*
    a. *paper*: This folder includes all the codes to produce all the tables and figures of the paper.
    b. *appendix*: This folder includes all the codes and instructions to produce all the tables and figures of the appendix.
2. *cleaning*: This folder includes all the codes and instructions to transform the raw data into intermediate data and the intermediate data in final data.
3. *finaldata*: This folder includes all the files needed to produce the tables and figures of the paper and appendix.
4. *intermediatedata*: This folder includes all the intermediate files needed to produce the final datasets.
5. *rawdata*: This folder includes all the raw data.
    a. *AgricultureProductivity*: This folder includes the data from IIASA/FAO (2010). Each subfolder includes potential yield for a specific crop.
        i. *Potential Yield Alfafa*
        ii. *Potential Yield Corn*
        iii. *Potential Yield Cotton*

iv. *Potential Yield Soybean*
v. *Potential Yield Wheat*

b. *AHS*: This folder includes data from the American Housing Survey (AHS) Metropolitan (Metro) and National Surveys. Each subfolder corresponds to a year YYYY of the data.
   i. *AHS_YYYY_Metro*
   ii. *AHS_YYYY_National*
c. *BaumSnow*: This folder includes data from Baum-Snow (2007).
d. *Bridge*: This folder includes data from National Archive of Criminal Justice Data (2005).
e. *CBPCounty*: This folder includes data from the County Business Pattern
   i. *MatchNAICS*: This folder includes data from National Science Board (2014) and the North American Industry Classification System Concordances.
   ii. *NationalArchives: This folder includes data from the County Business Pattern from 1975 to 1985. Each subfolder corresponds to a year YYYY of the data.*
      1. *YYYY*
f. *CDCCounty*: This folder includes data from the Centers for Disease Control and Prevention (CDC) State Childhood Blood Lead Surveillance Data
g. *census_tract70*: This folder includes data about the geography of U.S. Census tracts in 1970 from Manson et al. (2022).
h. *CensusMicroData*: This folder includes Census microdata from Ruggles et al. (2020).
i. *County_and_City_Data_Book*: This folder includes data from the U.S. Census County and City Data Books.
j. *GoldinKatz*: This folder includes data from Goldin and Katz (2008).
k. *inwatera_usa*: This folder includes data from United States Geological Survey (USGS) Inland Water Areas of the United States.
l. *LR_segregation*: This folder includes data about income and racial segregation until 2010 from Manson et al. (2018).
m. *Schooldistrict*: This folder includes data on school district geographies from U.S. Bureau of the Census (2002).
n. *Tetraethyl*: This folder includes data from Bureau of Mines Minerals Yearbooks.
o. *tigercounty*: This folder includes data from United States Census TIGER/Line on the geography of US counties.
p. *tigerdivisions*: This folder includes data from United States Census TIGER/Line on the geography of US Census divisions and regions.
q. *tigerplace*: This folder includes data from United States Census TIGER/Line on the geography of US places. *Each subfolder corresponds to a state ST of the data.*
   i. *tl_2010_ST_place00*
r. *UCR*: This folder includes data from Federal Bureau of Investigation (FBI) Uniform Crime Reporting (UCR) Program Data.
s. *water-polygons-split-4326*: This folder includes data from Water polygons OpenStreetMapData.
t. *wss_gsmsoil_US_[2006-07-06]*: This folder includes data from the U.S. General Soil Map (STATSGO2) - Web Soil Survey.
   i. *Spatial*: This folder includes the shapefile data.
   ii. *Tabular*: This folder includes the additional data.

Programs to produce the tables and figure in the paper

We here detail which dataset and program are needed to produce each table and figure included in the paper.

Section	Illustration	Data used	Program
Paper	Table 1	final_cc_data.dta	table1.do
Paper	Table 2	final_cc_data.dta	table2.do
Paper	Table 3.A	final_cc_data.dta	table3.do
Paper	Table 3.B	final_cc_data.dta	table3.do
		data_cbp.dta	
Paper	Table 4	final_cc_data.dta	table4.do
		data_cbp.dta	
Paper	Figure 1	final_cc_data.dta	figure1.do
		LR_shblack.dta	
		LR_income.dta	
Paper	Figure 2	final_cc_data.dta	figure2.do

Programs to produce the tables and figure in the online appendix

We here detail which dataset and program are needed to produce each table included in the appendix.

Section	Illustration	Data used	Program
Appendix	Table A.1	final_cc_data.dta	Appendix_A.do
Appendix	Table A.2	final_cc_data.dta	Appendix_A.do
		data_cbp_msa.dta	
Appendix	Table A.3	final_cc_data.dta	Appendix_A.do
Appendix	Table A.4	final_cc_data.dta	Appendix_C.do
Appendix	Table A.5	hwy-allyr-state.dta	Appendix_C.do
		final_cc_data.dta	
Appendix	Table A.6	hwy-allyr-state.dta	Appendix_C.do
		final_cc_data.dta	
Appendix	Table A.7	final_cc_data.dta	Appendix_D.do
Appendix	Table A.8	final_cc_data.dta	Appendix_D.do
Appendix	Table A.9	final_cc_data.dta	Appendix_D.do
Appendix	Table A.10	final_cc_data.dta	Appendix_D.do
		lead_blood.dta	
Appendix	Table A.11	final_cc_data.dta	Appendix_E.do
Appendix	Table A.12	final_cc_data.dta	Appendix_E.do
Appendix	Table A.13	final_cc_data.dta	Appendix_E.do
Appendix	Table A.14	final_cc_data.dta	Appendix_E.do
Appendix	Table A.15	final_cc_data.dta	Appendix_F.do
Appendix	Table A.16	final_cc_data.dta	Appendix_F.do
		data_ucr_plc_sd.dta	
Appendix	Table A.17	final_cc_data.dta	Appendix_F.do
		data_ucr_plc_sd.dta	
Appendix	Table A.18	census7080_micro.dta	Appendix_F.do
Appendix	Table A.19	final_cc_data.dta	Appendix_F.do
		educ_micro.dta	
Appendix	Table A.20	STATA Goldin.dta	Appendix_F.do
		final_cc_data.dta	
Appendix	Table A.21	data_ahs_nat.dta	Appendix_F.do
Appendix	Table A.22	final_cc_data.dta	Appendix_F.do
		data_ahs.dta	
Appendix	Table A.23	hwy-allyr-state.dta	Appendix_F.do
		final_cc_data.dta	
Appendix	Table A.24	final_cc_data.dta	Appendix_F.do
Appendix	Table A.25	hwy-allyr-state.dta	Appendix_F.do
		final_cc_data.dta	
Appendix	Table A.26	hwy-allyr-state.dta	Appendix_F.do
		final_cc_data.dta	
Appendix	Table A.27	final_cc_data.dta	Appendix_F.do
Appendix	Table A.28	final_cc_data.dta	Appendix_F.do
Appendix	Table A.29	final_cc_data.dta	Appendix_F.do
Appendix	Table A.30	final_cc_data.dta	Appendix_F.do
Appendix	Table A.31	final_cc_data.dta	Appendix_F.do
Appendix	Table A.32	final_cc_data.dta	Appendix_F.do
Appendix	Table A.33	final_cc_data.dta	Appendix_F.do
Appendix	Table A.34	final_cc_data.dta	Appendix_F.do
Appendix	Table A.35	final_cc_data.dta	Appendix_F.do
Appendix	Table A.36	final_cc_data.dta	Appendix_F.do
Appendix	Table A.37	final_cc_data.dta	Appendix_F.do
Appendix	Table A.38	final_cc_data.dta	Appendix_G.do
Appendix	Table A.39	final_cc_data.dta	Appendix_G.do
Appendix	Table A.40	tetraethyl_lead.xlsx	Appendix_G.do
		final_cc_data.dta	
Appendix	Table A.41	final_cc_data.dta	Appendix_G.do
Appendix	Table A.42	hwy-allyr-state.dta	Appendix_G.do
		final_cc_data.dta	
Appendix	Table A.43	final_cc_data.dta	Appendix_G.do
		ucr_mismeas.dta	
		ucr_nonmiss.dta	
Appendix	Table A.44	hwy-allyr-state.dta	Appendix_G.do
		final_cc_data.dta	
		ucr_mismeas.dta	
		ucr_nonmiss.dta	
Appendix	Table A.45	final_cc_data.dta	Appendix_G.do
Appendix	Table A.46	final_cc_data.dta	Appendix_G.do
Appendix	Table A.47	final_cc_data.dta	Appendix_G.do
Appendix	Table A.48	final_cc_data.dta	Appendix_G.do
Appendix	Table A.49	final_cc_data.dta	Appendix_G.do
Appendix	Table A.50	final_cc_data.dta	Appendix_G.do
Appendix	Table A.51	final_cc_data.dta	Appendix_G.do
Appendix	Table A.52	final_cc_data.dta	Appendix_G.do
Appendix	Table A.53	final_cc_data.dta	Appendix_G.do
Appendix	Table A.54	final_cc_data.dta	Appendix_G.do
Appendix	Table A.55	final_cc_data.dta	Appendix_H.do
		census7080_micro.dta	
Appendix	Table A.56	final_cc_data.dta	Appendix_H.do
		top10_CC.dta	
Appendix	Table A.57	final_cc_data.dta	Appendix_H.do
		top10_CC.dta	
Appendix	Table A.58	final_cc_data.dta	Appendix_H.do
		top10_CC.dta	

We here detail which dataset and program are needed to produce each figure included in the appendix.

Section	Illustration	Data used	Program
Appendix	Figure A.1	final_cc_data.dta	Appendix_A.do
Appendix	Figure A.2	final_cc_data.dta	Appendix_A.do
Appendix	Figure A.3	final_cc_data.dta	Appendix_A.do with further QGIS instructions
		final_cc_data.txt	
		tl_2010_allstates_place00_naec	
		tigerdivisions	
Appendix	Figure A.4	final_cc_data.dta	Appendix_A.do
Appendix	Figure A.5	final_cc_data.dta	Appendix_B.do
Appendix	Figure A.6	final_cc_data.dta	Appendix_B.do
Appendix	Figure A.7	final_cc_data.dta	Appendix_B.do
Appendix	Figure A.8	final_cc_data.dta	Appendix_C.do
Appendix	Figure A.9	final_cc_data.dta	Appendix_C.do
Appendix	Figure A.10	final_cc_data.dta	Appendix_C.do
Appendix	Figure A.11	hwy-allyr-state.dta	Appendix_C.do
		final_cc_data.dta	
Appendix	Figure A.12	hwy-allyr-state.dta	Appendix_C.do
		final_cc_data.dta	
Appendix	Figure A.13	final_cc_data.dta	Appendix_D.do
Appendix	Figure A.14	final_cc_data.dta	Appendix_D.do
Appendix	Figure A.15	final_cc_data.dta	Appendix_D.do
Appendix	Figure A.16	final_cc_data.dta	Appendix_F.do
		data_ucr_plc_sd.dta	
Appendix	Figure A.17	final_cc_data.dta	Appendix_F.do
		educ_micro.dta	
Appendix	Figure A.18	final_cc_data.dta	Appendix_F.do
Appendix	Figure A.19	final_cc_data.dta	Appendix_F.do
Appendix	Figure A.20	final_cc_data.dta	Appendix_G.do
Appendix	Figure A.21	tetraethyl_lead.xlsx	Appendix_G.do
		final_cc_data.dta	
Appendix	Figure A.22	final_cc_data.dta	Appendix_G.do

Data dictionary

The enclosed file data_dictionary.xlsx details all the variables used for producing the tables and figures of the paper. For each variable, we give a description, specify in which dataset is included, the source of the raw data or from which variables it is generated from, and for which illustrations in the paper is used.

Programs to produce the data used in the paper and in the appendix

We here describe the programs to transform the raw data into the final datasets. The enclosed file cleaning_summary.xlsx provide additional details about the source of all the data, the software used, all the necessary file inputs and all the file outputs.

Before running the do files the file named data_ucr_ori.tar needs to be uncompressed.

Here is the order in which the programs need to be run to create the main dataset used in the paper and in the appendix.

1. To construct the data about crimes at ORI level you need to run:
    a. clean_bridgefinal.do

    b. clean_ucr_ori.do
    c. clean_ucr_place.do
2. To construct the data about geology of places you need to run:
    a. Follow the instructions in clean_geologymaps.docx
    b. clean_geology_places.do
    c. clean_geology_places_ncc.do
3. To construct the data about local expenditure you need to run:
    a. clean_countycitydatabook.do
4. To construct the data about undevelopable areas you need to run:
    a. Follow the instructions in clean_developable.txt
    b. clean_developable.do
5. To construct the data about agricultural suitability you need to run:
    a. Follow the instructions in clean_agrimaps.txt
6. To combine the data from 1. to 5. with data about lead poisoning and Baum-Snow (2008) you need to run
    a. mergetofinal.do

These are the programs that create additional datasets used in the paper:

7. To construct data on employment location you need to run:
    a. clean_cbp.do
8. To construct data on long-run income and racial segregation you need to run:
    a. clean_lr_segregation.do

These are the programs that create additional datasets used:

9. To construct data on employment location at MSA level you need to run:
    b. clean_cbp_msa.do
10. To construct data on lead in the blood you need to run:
    c. clean_leadblood.do
11. To construct data on school districts you need to run:
    a. Follow the instructions in clean_schooldistrict.docx
    b. clean_schooldistrict.do
12. To construct data on the profile of people suburbanizing you need to run:
    a. clean_census7080.do
13. To construct data on educational attainments you need to run:
    a. clean_educ_micro.do
14. To construct data from the housing surveys you need to run:
    a. clean_ahs.do
15. To construct data on crime measurement errors you need to run:
    a. clean_ucr_mismeas_nonmiss.do
16. To construct data about crime in neighborhood cities you need to run:
    a. clean_neighbors.txt