Readme File for Replication Package

Changing Ingroup Boundaries: The Effect of Immigration on Race Relations in the U.S.

Vasiliki Fouka & Marco Tabellini

Contents

This replication package is structured as follows:

- Folder *datasets* includes datasets (.dta) used for the analysis in the main paper and appendix. We provide all data with the exception of ANES and GSS data with county identifiers, which are restricted use.

- Folder *figures and tables* includes code (.do) to produce the figures and tables in the main paper and appendix. Output is stored in folder *figures and tables/output* in .xls or .tex format (tables) or as .eps (figures).

- Folder *build* includes the following subfolders:
    - *build/dofiles* includes the code used to generate the final datasets used in the analysis.
    - *build/raw* is the folder storing the raw data used in the construction of the final datasets. We provide raw data for the survey experiment we conducted using Lucid Theorem. Code includes instructions on how to obtain all other raw data files with the appropriate data registration agreements.
    - *Build/intermediate* is a folder that stores intermediate datasets used in the process of constructing final datasets for the paper analysis. It is automatically populated after running the replication master do file (0_0.master.do)**.**

Instructions

The figures and tables of the paper and online appendix can be reproduced by running the do files in *figures and tables/dofiles main* and *figures and tables/dofiles appendix.* File paths are relative. You will need to set the working directory to the folder of the replication package.

To reconstruct the datasets used in the analysis starting from raw data, code is provided in *build/dofiles.* It can be executed in a single piece by running **0_0.master.do** after setting the working directory to the folder of the replication package.

Data sources

The raw data used in this project are obtained from the sources listed below:

1) Integrated Public Use Microdata Series – IPUMS USA (public use data)

The Integrated Public Use Microdata Series consists of a series of compatible-format individual-level representative samples of the United States for the years 1850-1880, 1900-present, and the American Community Surveys of 2000-present. In particular, we make use of data for years 1930 (complete count), 1960, 1970, 1980, 1990, 2000, and 2010.

Data are free and publicly available. They can be downloaded via the IPUMS USA extract system upon registration (see the link: https://usa.ipums.org/usa-action/samples). The extraction system creates a record layout tailored to the needs of each user and the approval is instantaneous.

These data are cited as: *Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover and Erin Meyer. 2019. "Integrated Public Use Microdata Series: Version 9.0 [dataset]" Minneapolis: University of Minnesota*

2) IPUMS National Historical Geographic Information System - NHGIS

NHGIS is one of several IPUMS data integration projects housed with the Minnesota Population Center at the Institute for Social Research & Data Innovation at the University of Minnesota. It provides free online access to summary statistics and GIS files for U.S. censuses. In particular, we make use of data for years 1960, 1970, 1980, 1990, 2000, and 2010.

Data are free and publicly available. They can be downloaded via the IPUMS USA extract system upon registration. The extraction system creates a record layout tailored to the needs of each user and the approval is instantaneous.

These data are cited as: *Manson, Steven, Jonathan Schroeder, David Van Riper and Steven Ruggles. 2019. "IPUMS National Historical Geographic Information System: Version 14.0 [Database]." Minneapolis: University of Minnesota*

3) American Community Surveys – ACS

The American Community Survey (ACS) is an ongoing survey conducted on a yearly basis by the U.S. Census Bureau. To maximize precision in the tract-level analysis, we use 5-year population estimates: 2005 to 2009, 2010 to 2014 and 2015 to 2019.

These data are cited as: *ACS, U.S. Census Bureau. American Community Survey 5-year estimates: 2005-2009; 2010-2014; 2015-2019*

3) American National Election Studies - ANES

The American National Election Studies are a continuous series of survey data investigating electoral behavior and attitudes in the United States. Surveys include a pre-election interview and a post-election interview during years of Presidential elections, and, usually, a post-election interview only during years of midterm elections.

In particular, we make use of ANES "Time Series" from 1972 to 2012 and of ANES Panel Study covering the period 2000-2004

Data are free and publicly available, and they can be downloaded from the website www.electionstudies.org.

County-level ANES post 1978 are restricted use. Applicants must submit a Restricted Data Use Agreement. The Inter-university Consortium for Political and Social Research (ICPSR) handles applications for ANES Restricted-Use Data and provides access to the data through their Virtual Data Enclave (VDE).

These data are cited as:
ANES. "Time Series Cumulative Data File (1948-2012) [database]." Ann Arbor, MI: Inter-university Consortium for Political and Social Research
ANES. "2004 Time Series and Panel Contextual File [database]." Ann Arbor, MI: Inter-university Consortium for Political and Social Research

4) Cooperative Congressional Election Study – CCES

The Cooperative Congressional Election Study is a 50,000+ person national stratified sample survey administered by YouGov. The survey consists of two waves in election years. In the pre-election wave, respondents answer two-thirds of the questionnaire. In non-election years, the survey consists of a single wave conducted in the fall. In particular, we make use of surveys conducted from 2007 to 2018.

Data are free and publicly available, and they can be downloaded from the website https://cces.gov.harvard.edu/

These data are cited as*: Kuriwaki, Shiro. 2018. "Cumulative CCES Common Content (2006-2018)"*

5) General Social Survey - GSS

The General Social Survey (GSS) is a nationally representative survey of adults in the United States conducted since 1972. The GSS collects data on contemporary American society in order to monitor and explain trends in opinions, attitudes and behaviors. It is a project of the independent

research organization NORC at the University of Chicago, with principal funding from the National Science Foundation. In particular, we make use of surveys conducted from 1993 to 2010.

County-level data are restricted use. The GSS geographic identification code files are made available to researchers under special contract with NORC. The process for obtaining the GSS Sensitive Data include the sending of a research proposal, researchers' resumes and a Sensitive Data Protection Plan

These data are cited as*: Smith, Tom W., Davern, Michael, Freese, Jeremy, and Morgan, Stephen. 2018. "General Social Surveys, 1972-2018 [machine-readable data file]" Chicago: NORC.*

6) Hate Crime Data

The Hate Crime data is an FBI data set that is part of the annual Uniform Crime Reporting (UCR) Program data. This data contains information about hate crimes reported in the United State. Since 1930, the Federal Bureau of Investigation (FBI) has compiled the Uniform Crime Reports (UCR) to serve as periodic nationwide assessments of reported crimes not available elsewhere in the criminal justice system. Each year, participating law enforcement agencies contribute reports to the FBI either directly or through their state reporting programs.

Data are free and publicly available, and they can be downloaded from the ICPSR open platform.

These data are cited as: *United States Federal Bureau of Investigation. 2018. "Uniform Crime Reporting Program Data: Hate Crime Data [Record-Type Files]." Ann Arbor: Inter-university Consortium for Political and Social Research*

7) Data on other crimes

The Uniform Crime Reporting Program Data: Offenses Known and Clearances By Arrest data set is a compilation of offenses reported to law enforcement agencies in the United States. Crimes included are criminal homicide, forcible rape, robbery, aggravated assault, burglary, larceny-theft, and motor vehicle theft. In particular, we make use of the data cleaned and concatenated by Jacob Kaplan.

Data are free and publicly available, and they can be downloaded from the ICPSR open platform.

These data are cited as: *Kaplan, Jacob. 2020. "Jacob Kaplan's Concatenated Files: Uniform Crime Reporting Program Data: Offenses Known and Clearances by Arrest, 1960-2019". Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]*

8) Survey experiment

Data comes from Lucid Theorem and is provided in raw form, without the inclusion of geographic identifiers or IP addresses that could compromise the anonymity of respondents.

Statement about Rights

The author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

Software

Results reported in the paper were produced using STATA 15 and R v.4.0.5.

References

ANES. *"Time Series Cumulative Data File (1948-2012) [database]."* Ann Arbor, MI: Inter-university Consortium for Political and Social Research

ANES. *"2004 Time Series and Panel Contextual File [database]."* Ann Arbor, MI: Inter-university Consortium for Political and Social Research

Kaplan, Jacob. 2020. *"Jacob Kaplan's Concatenated Files: Uniform Crime Reporting Program Data: Offenses Known and Clearances by Arrest, 1960-2019"*. Ann Arbor, MI: Inter-university Consortium for Political and Social Research [distributor]

Kuriwaki, Shiro. 2018. "*Cumulative CCES Common Content (2006-2018)*"

Manson, Steven, Jonathan Schroeder, David Van Riper and Steven Ruggles. 2019. *"IPUMS National Historical Geographic Information System: Version 14.0 [Database]."* Minneapolis: University of Minnesota

Ruggles, Steven, Sarah Flood, Ronald Goeken, Josiah Grover and Erin Meyer. 2019. *"Integrated Public Use Microdata Series: Version 9.0 [dataset]."* Minneapolis: University of Minnesota

Smith, Tom W., Davern, Michael, Freese, Jeremy, and Morgan, Stephen. 2018. *"General Social Surveys, 1972-2018 [machine-readable data file]"* Chicago: NORC.

United States Federal Bureau of Investigation. 2018. "*Uniform Crime Reporting Program Data: Hate Crime Data [Record-Type Files]."* Ann Arbor: Inter-university Consortium for Political and Social Research

ACS, U.S. Census Bureau. *"American Community Survey 5-year estimates*," 2005-2009; 2010-2014; 2015-201.