

README File for Replication Package

Racial Diversity and Racial Policy Preferences: The Great Migration and Civil Rights

Alvaro Calderon & Vasiliki Fouka & Marco Tabellini

This replication package includes

- Data: datasets (.dta) to replicate all analyses in the main text
- Codes: STATA files (.do) to replicate results reported in all figures and tables of the main text
- Results: to save outputs for all figures (.eps) and tables (.xls) displayed in the main text

Dataset List

In order to replicate all figures and tables of the main text, the following 7 datasets are included:

- a) **main_county_FD**: it includes data needed to produce results at the county level (Figures 2, 7; Tables 1, 2, 4)
- b) **main_CZ**: it includes data needed to produce results at the commuting zone level (Table 2)
- c) **main_CD**: it includes data needed to produce results at the Congressional District level (Figure 4; Tables 1, 3)
- d) **main_ANES**: it includes data needed to produce results relying on ANES survey data at the state level (Figure 1 and Table 5)
- e) **main_newspapers**: it includes data needed to produce results on historical press at the county level (Figure 6; Tables 6, 7)
- f) **dta_forfig_2**: it includes data to be mapped in Figure 2
- g) **dta_forfig_3**: it includes data on discharge petitions to produce Figure 3
- h) **dta_forfig_5**: it includes data on voting behavior by race to produce Figure 5
- i) **uscoord and usdb**: these datasets are instrumental to plot the map in Figure 2

Data Availability and Provenance Statements

The paper uses data extracted from several data sources:

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

This dataset consists of a series of compatible-format individual-level representative samples of the United States for years 1930 and 1940. In particular, we make use of complete count data. All data are coded and anonymous. Data are free and publicly available. They can be downloaded via the IPUMS USA extract system upon registration (at this [link](#)). 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, S., K. Genadek, R. Goeken, J. Grover, and M. Sobek (2020). Integrated Public Use Microdata Series: Version 10.0 [dataset], Minneapolis: University of Minnesota.

2) Migration data

The data contain net migration rates by race, for each county and decade from 1930 to 1970. Data are free and publicly available. They can be downloaded from the [link](#) to the ICPSR 20 and the [link](#) to the ICSPR 8493 studies.

These data are cited as:

Gardner, J. and W. Cohen (1992). Demographic Characteristics of the Population of the United States, 1930-1950: County-Level, Inter-university Consortium for Political and Social Research [distributor];

Bowles, G. K., J. D. Tarver, C. L. Beale, and E. S. Lee (2016). Net Migration of the Population by Age, Sex, and Race, 1950-1970, Inter-university Consortium for Political and Social Research [distributor].

3) Electoral data

The dataset contains number of votes (overall and by political party) in each county and election for several offices and political position (including Presidential and Congressional elections) from 1788 to 1990. Data are free and publicly available from the following [link](#).

These data are cited as:

Clubb, J. M., W. H. Flanigan, and N. H. Zingale (1990). Partisan Realignment: Voters, Parties, and Government in American History. Beverly Hills, CA: SAGE.

4) Legislators' ideology

The dataset includes (race-related) ideology scores for each Congress member, from 1877 to 2011. Legislators are assigned a score that is a function of their past voting behavior on civil rights bills. Data are free and publicly available from the following [link](#).

These data are cited as:

Bateman, D. A., J. D. Clinton, and J. S. Lapinski (2017). "A House Divided? Roll Calls, Polarization, and Policy Differences in the US House, 1877–2011". American Journal of Political Science 61 (3), 698–714.

5) Discharge Petitions data

This dataset contains information on discharge petitions filed in the Congress period 71-94 (1929-1976). In particular, it includes the signatories on each discharge petition, along with the order in which the members signed and information about each member's party, state, ideology, committee assignments, and seniority. Data were kindly shared with the authors by Eric Schickler (at eschickler@berkeley.edu) and Kathryn Pearson (at kpearson@umn.edu).

These data are cited as:

Pearson, K. and E. Schickler (2009). "Discharge Petitions, Agenda Control, and the Congressional Committee System, 1929–76". The Journal of Politics 71 (4), 1238– 1256.

6) CORE demonstrations

This dataset combines a variety of sources collecting information on non-violent demonstrations organized between 1942 and 1970 by 0 by the Congress of Racial Equality (CORE), an inter-racial group of students from the University of Chicago. Data were kindly shared with the authors by James Gregory (at labhist@u.washington.edu).

These data are cited as:

Gregory, J. N. and A. Hermida (2019). "Congress of Racial Equality (CORE) Actions, 1942-1972. Mapping American Social Movement".

7) NAACP chapters

This dataset combines a variety of sources to geolocalize NAACP Chapters. Data are available for years 1940 and 1960 (pooling together 1957 and 1964). Data were kindly shared with the authors by James Gregory (at labhist@u.washington.edu).

These data are cited as:

Gregory, J. N. and J. Estrada (2019). "NAACP History and Geography. Mapping American Social Movement".

8) American National Election Studies – ANES

The American National Election Studies (ANES) is a cross-sectional, nationally representative survey conducted since 1948 by the University of Michigan every two or four years depending on the waves. We make use of ANES "Time Series" from 1952 to 1964. Data are free and publicly available from the following [link](#).

These data are cited as:

ANES Cumulative File (2015). Time Series Cumulative Data File (1948-2012) [database], Inter-university Consortium for Political and Social Research.

9) Newspapers.com

We retrieve words used in historical newspapers from the website <https://www.newspapers.com/>, last accessed in August 2020. This is the largest online newspaper archive consisting of more than 641 million pages of historical newspapers from 20,700 newspapers from around the United States and beyond. The online archive is accessible upon registration.

There is no specific citation for this dataset, except the [website](#) itself.

10) Estimates on voting behavior

We extract statistics on 1940 and 1960 population eligible to vote, by race, from Gardner and Cohen (1992). Estimates on the 1940 Democratic vote share and turnout are taken from Glantz (1960). We use these sources to perform a back-of-the-envelope calculation to estimate the voting behavior of white and African American residents in non-southern counties. The details of this exercise are described in Section 6.2 of the paper. Data are free and publicly available from Glantz (1960)

These data are cited as:

Glantz, O. (1960). "The Negro Voter in Northern Industrial Cities". Western Political Quarterly 13(4), 999–1010.

Data Sources

The datasets in the replication folder are obtained from the sources listed above as follows:

a) main_county_FD

This dataset combines raw data from [1], [2], [3], [6], and [7]. The dataset available in the replication package is restricted to the 1,263 non-southern US counties for which electoral returns in Congressional elections are available for all Census years between 1940 and 1970, and with at least one African American resident in 1940. The dataset is at the county level, covers the period from 1940 to 1970, and includes variables in both levels and changes.

b) main_CZ

This dataset combines raw data from [1], [2], [3], [6], and [7]. The dataset available in the replication package is obtained by collapsing the county level dataset from (a) at the commuting zone level.

c) main_CD

This dataset combines raw data from [1], [2], [3], [4], and [5]. Raw data at the county level are converted at the congressional district level (CD) by applying the crosswalk described in Online Appendix B of the paper. The dataset available in the replication package is restricted to non-southern US CDs and to the Congress period 78-88 (1943-1965). It is at the Congressional District level, and includes variables in both levels and changes.

d) main_ANES

This dataset combines raw data from [1], [2], [3], and [8]. The dataset available in the replication package is restricted to survey years from 1952 to 1964. It is at the individual level and each respondent is matched with corresponding state-level variables.

e) main_newspapers

This dataset combines raw data from [1], [2], [3], and [9]. It is at the county-week-episode level, and is restricted to the 492 counties in our sample for which newspapers data are available. Week 0 refers to the week in which the lynching occurred (and so varies by lynching episode).

f) dta_forfig_2

This dataset is from the main dataset at the county level (a) and restricts attention to the variables mapped in Figure 2.

g) dta_forfig_3

This dataset consists of a summary table of data on discharge petitions [5].

h) dta_forfig_5

This dataset combines raw data from [2], [3], and [10]. It consists of statistics and estimates on the voting behavior of both African Americans and whites, in 1940 and 1960.

Details on each Data Source

All the datasets are in dta (i.e., STATA) format. Data from [1], [2], and [3] were downloaded in May 2020. Newspapers data [9] were last downloaded in August 2020.

Summary of Availability

All data are publicly available, except those shared by individual authors as explained above (in particular, Eric Schickler and Kathryn Pearson for discharge petitions, and James Gregory for NAACP and CORE).

Statement about Rights

The authors of the manuscript have legitimate access to and permission to use the data used in this manuscript.

Description of programs/code and instructions for replication

Results reported in the paper were produced using STATA 17. In order to replicate results, please make reference to the master dofile. First, set the working directory to the replication folder; then, codes should execute. There is one .do file for each table/figure (named accordingly).

In order to run the replication, you may need to install the following packages from the SSC archive: ivreg2, xtivreg2, ftools, reghdfe, ivreghdfe, estout, distinct, coefplot, spmap, shp2dta, mif2dta

List of codes

Figure/Table #	Program	Output File
Figure 1	Figure1.do	Figure1_a.eps, Figure1_b.eps, Figure1_c.eps, Figure1_d.eps
Figure 2	Figure2.do	Figure2.eps
Figure 3	Figure3.do	Figure3.eps
Figure 4	Figure4.do	Figure4
Figure 5	Figure5.do	Figure5.eps
Figure 6	Figure6.do	Figure6.eps
Figure 7	Figure7.do	Figure7.eps
Table 1	Table1.do	Table1_cd.csv, Table1_county.csv
Table 2	Table2.do, Table2_CZ.do	Table2_A.xls, Table2_B.xls, Table2_C.xls, Table2_CZ.xls
Table 3	Table3.do	Table3_A.xls, Table3_B.xls, Table3_C.xls
Table 4	Table4.do	Table4_A.xls, Table4_B.xls
Table 5	Table5.do	Table5_A.xls, Table5_B.xls
Table 6	Table6.do	Table6.xls, Table6_B.xls
Table 7	Table7.do	Table7_A.xls, Table7_B.xls

References

- ANES Cumulative File (2015). *Time Series Cumulative Data File (1948-2012)* [database], Inter-university Consortium for Political and Social Research.
- Bateman, D. A., J. D. Clinton, and J. S. Lapinski (2017). “A House Divided? Roll Calls, Polarization, and Policy Differences in the US House, 1877–2011”. *American Journal of Political Science* 61 (3), 698–714.
- Bowles, G. K., J. D. Tarver, C. L. Beale, and E. S. Lee (2016). *Net Migration of the Population by Age, Sex, and Race, 1950-1970*, Inter-university Consortium for Political and Social Research [distributor].
- Clubb, J. M., W. H. Flanigan, and N. H. Zingale (1990). *Partisan Realignment: Voters, Parties, and Government in American History*. Beverly Hills, CA: SAGE.
- Gardner, J. and W. Cohen (1992). *Demographic Characteristics of the Population of the United States, 1930-1950: County-Level*, Inter-university Consortium for Political and Social Research [distributor]
- Glantz, O. (1960). “The Negro Voter in Northern Industrial Cities”. *Western Political Quarterly* 13(4), 999–1010.
- Gregory, J. N. and J. Estrada (2019). “NAACP History and Geography. Mapping American Social Movement”.
- Gregory, J. N. and A. Hermida (2019). “Congress of Racial Equality (CORE) Actions, 1942-1972. Mapping American Social Movement”.
- Pearson, K. and E. Schickler (2009). “Discharge Petitions, Agenda Control, and the Congressional Committee System, 1929–76”. *The Journal of Politics* 71 (4), 1238– 1256.
- Ruggles, S., K. Genadek, R. Goeken, J. Grover, and M. Sobek (2020). *Integrated Public Use Microdata Series: Version 10.0* [dataset], Minneapolis: University of Minnesota.