

Readme – Replication package for “Shift, Not Stasis: The Geography of Post-Civil Rights Racial Inequality”

Overview

This replication package contains code and intermediate results for the paper “Shift, Not Stasis: The Geography of Post-Civil Rights Racial Inequality” (Manduca 2026). The paper is based entirely on publicly available data, primarily US Decennial Census and American Community Survey data downloaded from IPUMS USA (Ruggles et al. 2025) and the National Historical Geographic Information System (Schroeder et al. 2025). The bulk of the analysis is conducted in R, with one short program in Stata.

The entry point for running the pipeline is `code/source_calls.R`, which sources each build and analysis program in order.

Software requirements

- **R** (tested with version 4.x). Required packages, all installed from CRAN:
 - Build phase: `tidyverse`, `segregation`, `Hmisc`, `weights`, `readstata13`, `haven`, `ipumsr`
 - Analysis phase: `tigris`, `sf`, `ggrepel`, `scales`, `lm.beta`, `estimatr`, `car`, `gridExtra`
- **Stata** (any modern version). One do-file (`prep_12_racial_oppression_factor.do`) is run in Stata to compute the HRR factor with `pca`.

Setup

Before running this replication package:

1. **Create a working directory** and copy all folders in the replication package to that directory.
2. **Edit** `code/source_calls.R`:
 - Set the working directory path on line 4 (`setwd('PATH/TO/replication_package/')`).
 - Set the seven extract-name placeholders on lines 6–12 to match the file names of the data extracts you downloaded (see the data section below). These are used by the build scripts to construct the actual file paths.
3. **Edit** `code/build_code/prep_12_racial_oppression_factor.do`: Set the path local on line 4 to the same replication-package path you used in `source_calls.R`.
4. **Download the publicly available datasets** described below into the corresponding folders inside `raw_data/`.

5. **Run** `code/source_calls.R` from start to finish. When the script pauses (after `prep_11_racial_oppression.R`), open Stata, run `code/build_code/prep_12_racial_oppression_factor.do`, return to R, and press Enter to continue.

Reproducing the analysis without rebuilding from raw data

This package includes the analysis-ready intermediate data (`intermediate_data/sumstats/` and `intermediate_data/regression_data/`), which allows users to reproduce every figure and table **without** downloading the raw microdata or running the build. To do this, set the working directory on line 4 of `code/source_calls.R`, run the library-loading lines, and then run only the analysis programs — the `source()` calls for `analysis_1_regressions.R` through `graphs_6_statistics_for_paper.R` near the end of `source_calls.R` — skipping the build programs (`prep_1–prep_14`). To rebuild everything from the original sources instead, follow the full setup below.

Required data downloads

Each row below lists a dataset, where to save it, and which placeholder in `source_calls.R` (if any) needs to be set to the extract name.

IPUMS USA microdata The following IPUMS USA variables are required:

- Sample/weights: YEAR, MULTYEAR, SAMPLE, SERIAL, CBSERIAL, HHWT, PERWT, STATEFIP, CNTYGP97, CNTYGP98, PUMA, GQ, GQTYPE, FAMUNIT, PERNUM
- Demographic: AGE, SEX, RACE, RACED, HISPAN, BPL, CITIZEN
- Education / labor force: EDUC, EMPSTAT
- Income: INCTOT, FTOTINC, INCSS, INCWELFR, INCSUPP

Download a single extract covering the years 1970, 1980, 1990, 2000, 2005–2009 ACS, 2010–2014 ACS, and 2015–2019 ACS. Save the resulting .Rdata file in `raw_data/ipums/`, and set `ipumsname` on line 6 of `source_calls.R` to the file name without the .Rdata extension (e.g. 'usa_00037').

IPUMS USA 1970 county-group composition table A separate supplementary table from IPUMS USA listing the counties that make up each 1970 county group is required to build the 1970 county-group / county crosswalk by race. Download the county-group composition spreadsheet (an .xls file) from <https://usa.ipums.org/usa/volii/t1970maps.shtml#composition>, convert it to CSV, and save it as `ipums_usa_1970_county_group_composition.csv` in `raw_data/1970_crosswalk/`. The file should have columns labeled County Group (CNTYGP70), State FIPS Code (STATEFIP), and County FIPS Code* (COUNTYFIP).

IPUMS CPS data (unionization) Download an IPUMS CPS Annual Social and Economic Supplement (ASEC) extract covering 1989–2019 with the following variables:

- YEAR, ASECWT, STATEFIP, METAREA, METFIPS
- UNION

Save both the .xml (DDI) and .dat files in `raw_data/union_cps/`, and set `cpsname` on line 7 of `source_calls.R` to the file name without the extension (e.g. 'cps_00030').

In addition, place `cps_metro_xw_1.csv` (a crosswalk from old METAREA codes to the current `met_standard` definitions) in `raw_data/union_cps/`. This file is included with the replication package.

ACS Supplemental Poverty Measure (SPM) microdata Download annual SPM files `spm_2010_pu.dta` through `spm_2019_pu.dta` (US Census Bureau 2025) and save them in `raw_data/spm_acs/`.

NHGIS aggregate Census data (five separate extracts) The build code expects five NHGIS extracts. For each one, save the extract folder/files to the indicated location, and set the corresponding placeholder in `source_calls.R` to the extract name (e.g. 'nhgis0076'). The placeholder variables (`nhgis_1970pop_name`, `nhgis_ind_name`, `nhgis_seg_name`, `nhgis_slavery_name`, `nhgis_sharecrop_name`) are commented in `source_calls.R` with the corresponding extract description.

Extract	Contents	Save to
1970 county race population	County-level total / white / Black / other population from the 1970 Census (ds94)	<code>raw_data/1970_crosswalk/</code>
Industry	County-level employment by industry for 1970, 1980, 1990, 2000, 2005–2009 ACS, 2010–2014 ACS, 2015–2019 ACS	<code>raw_data/nhgis_industries/</code>
Segregation	Tract-level population by race for 1970, 1980, 1990, 2000, 2005–2009 ACS, 2010–2014 ACS, 2015–2019 ACS	<code>raw_data/nhgis_segregation/</code>
Slavery (1860)	County-level slave and total population from the 1860 Census	<code>raw_data/nhgis_racial_oppression/</code>
Sharecropping (1930)	County-level cropper-farm counts from the 1930 Census of Agriculture	<code>raw_data/nhgis_racial_oppression/</code>

Additional datasets to download The following datasets are not included in the replication package and must be downloaded by the user into the indicated folders inside `raw_data/`.

- **Historical county crosswalks (Eckert et al. 2020).** Download `eglp_county_crosswalk_endyr_2010.csv` from the authors' replication materials (linked from the NBER working paper page, <https://doi.org/10.3386/w26770>) and save it in `raw_data/county_hist/`.
- **Longitudinal Tract Database (Logan, Xu, and Stults 2014).** Download the full-count tract files for 1970, 1980, 1990, 2000, 2010, and 2020 from the LTDB website at <https://s4.ad.brown.edu/projects/diversity/Researcher/LTDB.htm>. Save them in `raw_data/ltdb_std_all_fullcount/` with the file names expected by the code: `LTDB_Std_1970_fullcount.csv`, `LTDB_Std_1980_fullcount.csv`, `LTDB_Std_1990_fullcount.csv`, `LTDB_Std_2000_fullcount.csv`, `LTDB_Std_2010_fullcount.csv`, and `ltdb_std_2020_fullcount.csv`.
- **State-level union membership rates.** Download `State_Union_Membership_Density_1964-2018.xlsx` from the Union Membership and Coverage Database at <https://unionstats.com/> (Hirsch, Macpherson, and Even 2026; early years from Hirsch, Macpherson, and Vroman 2001), convert it to CSV (`State_Union_Membership_Density_1964-2018.csv`), and save it in `raw_data/unionstats/`. `prep_7_union_cps.R` reshapes this wide spreadsheet into the long-form table consumed by `prep_14`.
- **State-level incarceration rates (Vera Institute of Justice 2025).** Download the Incarceration Trends dataset from <https://trends.vera.org/> (or the GitHub repository at <https://github.com/vera-institute/incarceration-trends>) and save the state-level file as `incarceration_trends_state.csv` in `raw_data/incarceration/`.

Other inputs included with the replication package The following files are included with the replication package and do not need to be re-downloaded; they are listed here for reference.

- **Geocorr PUMA-to-county crosswalks** (1980, 1990, 2000, 2010) in `raw_data/geocorr/`.
- **CBSA delineations (Office of Management and Budget 2020):** `list1_2020.csv` and `msa_names_2020.csv` in `raw_data/msas/`.
- **MSA-to-state crosswalk:** `msa_to_states_xwalk.csv` in `raw_data/msa_to_state_xwalk/`.
- **Census region codes:** `state-geocodes-v2016.csv` in `raw_data/census_regions/`.
- **Baker (2022) historical-racial-regime measures:** `hrr_baker.csv` in `raw_data/baker_hrr/`.

Build code

All build programs are in `code/build_code/` and are run in order by `source_calls.R`.

`prep_1_xwalk_1970_byrace.R`

Builds a crosswalk from 1970 county groups (which sometimes cross state and county boundaries) to 1970 counties, by race. For each county group × county pair, it computes race-specific and all-race population weights from the 1970 Census full-count tabulations, so that downstream code can allocate county-group-level statistics to counties (and ultimately MSAs) in a race-aware way.

- **Inputs:**

- `raw_data/1970_crosswalk/ipums_usa_1970_county_group_composition.csv`
- `raw_data/1970_crosswalk/<nhgis_1970pop_name>.csv/<nhgis_1970pop_name>_ds94_1970_county.csv`

- **Outputs:** `intermediate_data/crosswalks/xwalk_1970_cg_cnty_byrace.csv`.

`prep_2_pumas_to_msas.R`

Creates crosswalks that allocate 1970 county groups, 1980 county groups, and 1990/2000/2010 PUMAs to current (2020) CBSAs. Within each state, all counties not in a metropolitan CBSA are pooled into a single “non-metropolitan” pseudo-MSA coded `nm_<state FIPS>`.

- **Inputs:**

- `raw_data/msas/list1_2020.csv`
- `intermediate_data/crosswalks/xwalk_1970_cg_cnty_byrace.csv` (from `prep_1`)
- `raw_data/geocorr/cg98stat.csv`
- `raw_data/geocorr/geocorr1990_puma_to_county.csv`
- `raw_data/geocorr/geocorr2000_puma_to_county.csv`
- `raw_data/geocorr/geocorr2014_puma_to_county.csv`

- **Outputs:** `output/crosswalks/puma_msa_<year>_byrace.csv` for each year 1970–2010, plus a combined `puma_msa_allyrs_byrace.csv`.

`prep_3_spm.R`

Reads the annual ACS SPM files and combines them into two 5-year pooled datasets, adding `dispinc` (disposable income, adding back child-care/work/medical expenses that SPM subtracts) and `posttax` (income after taxes) variables.

- **Inputs:** raw_data/spm_acs/spm_<year>_pu.dta for <year> in 2010–2019.
- **Outputs:** intermediate_data/spm_acs/spm_2010_2014.Rdata, intermediate_data/spm_acs/spm_2015_2019.Rdata.

prep_4_ipums.R

Cleans the raw IPUMS USA microdata one year (or 5-year ACS) at a time: constructs income, family, household, race, education, employment, group-quarters, migration, and citizenship variables; merges on PUMA-to-MSA crosswalks (producing population-weighted observations across MSAs); and, for the two most recent ACS samples, merges on the SPM data prepared in prep_3_spm.R.

- **Inputs:**
 - raw_data/ipums/<ipumsname>.Rdata
 - output/crosswalks/puma_msa_allyrs_byrace.csv (from prep_2)
 - intermediate_data/spm_acs/spm_2010_2014.Rdata and spm_2015_2019.Rdata (from prep_3)
- **Outputs:** intermediate_data/ipums/ipums_clean_<year>.Rdata for <year> in {1970, 1980, 1990, 2000, 2009, 2014, 2019}.

prep_5_sumstats.R

Computes summary statistics (mean, median, 5th/10th/25th/75th/90th/95th percentiles, population, weighted observations, and a rank-gap measure) at the national, regional (South vs. rest of country), and MSA levels, separately for all residents and for “expansive” and “narrow” race classifications, and for a variety of population subsets (full sample, household population, group-quarters subsets, education subsets, employment subsets, migration subsets, citizenship subsets).

- **Inputs:**
 - intermediate_data/ipums/ipums_clean_<year>.Rdata (from prep_4)
 - raw_data/census_regions/state-geocodes-v2016.csv
- **Outputs (in intermediate_data/sumstats/):**
 - sumstats_nat_all.csv, sumstats_nat_raceex.csv, sumstats_nat_racenr.csv
 - sumstats_region_all.csv, sumstats_region_raceex.csv, sumstats_region_racenr.csv
 - sumstats_msa_all.csv, sumstats_msa_raceex.csv, sumstats_msa_racenr.csv
 - sumstats_msa_raceex_matchg.csv (statistics within MSA × age-bucket × sex × race cells)

`prep_6_matched_sumstats_weighted.R`

Takes the age/sex-stratified MSA summary statistics from `prep_5` and computes population-weighted averages (weighted by Black population, and separately by white population) to produce one observation per MSA $\times$ year that uses the Black (or white) age/sex distribution as the standard.

- **Inputs:** `intermediate_data/sumstats/sumstats_msa_raceex_matchg.csv` (from `prep_5`).
- **Outputs:** `intermediate_data/sumstats/msa_matched_bpop.csv`,
`intermediate_data/sumstats/msa_matched_wpop.csv`.

`prep_7_union_cps.R`

Uses IPUMS CPS data to compute unionization rates by MSA $\times$ year for 1989, 1999, 2009, 2014, and 2017 (the years covered in the analysis). Where an MSA has fewer than 50 observations in a given year, the corresponding state-level rate is used instead. Also reshapes the Hirsch–Macpherson state-level union density spreadsheet from wide to long form for use as the pre-1989 state-level union rate.

- **Inputs:**
 - `raw_data/union_cps/<cpsname>.xml` and `<cpsname>.dat`
 - `raw_data/union_cps/cps_metro_xw_1.csv`
 - `raw_data/unionstats/State_Union_Membership_Density_1964-2018.csv`
- **Outputs:**
 - `intermediate_data/predictors/cps_metro_xw.csv`
 - `intermediate_data/predictors/union_50oneyr_cbsa.csv`
 - `intermediate_data/predictors/union_50oneyr_st.csv`
 - `intermediate_data/predictors/union_statewide.csv` (long-form fips $\times$ year $\times$ unionrate)

`prep_8_cen_industry.R`

Constructs MSA-level employment shares in manufacturing, public administration, professional services, and FIRE (finance/insurance/real estate) for 1969, 1979, 1989, 1999, 2007, 2012, and 2017, using NHGIS county-level industry tables and the Eckert et al. (2020) county crosswalks to convert pre-2010 counties to 2010 county boundaries.

- **Inputs:**
 - `raw_data/msas/list1_2020.csv`
 - `raw_data/county_hist/eglp_county_crosswalk_endyr_2010.csv`
 - `raw_data/nhgis_industries/<nhgis_ind_name>_ds98_1970_county.csv` (and the analogous files for 1980, 1990, 2000, 2009 ACS, 2014 ACS, 2019 ACS)

- **Outputs:** `intermediate_data/predictors/msa_census_emp_all.csv`.

`prep_9_segregation.R`

Computes Black-white dissimilarity, isolation, Theil (mutual information), and exposure indices for each MSA $\times$ year using NHGIS tract-level race counts at the original tract definitions of each Census year.

- **Inputs:**
 - `raw_data/msas/list1_2020.csv`
 - `raw_data/nhgis_segregation/<nhgis_seg_name>_ds98_1970_tract.csv` (and the analogous files for the 1980–2000 time-series and the 2009/2014/2019 ACS).
- **Outputs:** `intermediate_data/predictors/msa_segregation_nhgis.csv`.

`prep_10_segregation_ltdb.R`

Computes the same four segregation indices, but using the Longitudinal Tract Database (Logan et al. 2014), which projects all years onto 2010 tract boundaries for a more comparable time series.

- **Inputs:**
 - `raw_data/msas/list1_2020.csv`
 - `raw_data/ltdb_std_all_fullcount/LTDB_Std_<year>_fullcount.csv` for 1970–2010, plus `ltdb_std_2020_fullcount.csv`.
- **Outputs:** `intermediate_data/predictors/msa_segregation_ltdb.csv`.

`prep_11_racial_oppression.R`

Builds MSA- and state-level versions of the historical-racial-regime (HRR) inputs: 1860 share of population enslaved, 1930 share of sharecropping farms operated by Black households, the number of disfranchisement devices, and support for the 1956 Southern Manifesto (from Baker 2022).

- **Inputs:**
 - `raw_data/msas/list1_2020.csv`
 - `raw_data/county_hist/eglp_county_crosswalk_endyr_2010.csv`
 - `raw_data/nhgis_racial_oppression/<nhgis_slavery_name>_csv/<...>_ds14_1860_county.csv`
 - `raw_data/nhgis_racial_oppression/<nhgis_sharecrop_name>_csv/<...>_ds212_1930_county.csv`
 - `raw_data/msa_to_state_xwalk/msa_to_states_xwalk.csv`
 - `raw_data/baker_hrr/hrr_baker.csv`
- **Outputs:**
 - `intermediate_data/predictors/racialoppr_msa.csv`

- intermediate_data/predictors/racialoppr_st.csv

prep_12_racial_oppression_factor.do (Stata)

Runs principal components analysis on the slavery and sharecropping measures (and on the 4-variable version that also includes disfranchisement and Southern Manifesto) to produce the HRR factor used in the regressions. **This is the one place the pipeline switches to Stata.** Edit the path local in the do-file to point to your replication directory, then run.

- **Inputs:**

- intermediate_data/predictors/racialoppr_msa.csv
- intermediate_data/predictors/racialoppr_st.csv

- **Outputs:**

- intermediate_data/predictors/racialoppr_msa_wfactor.csv
- intermediate_data/predictors/racialoppr_st_wfactor.csv

After this program runs, return to R and press Enter to resume source_calls.R.

prep_13_regressions_outcomes.R

Pivots the summary statistics into a wide MSA × year panel of outcome variables: Black/white levels, ratios, log differences, rank gap, and matched-sample equivalents.

- **Inputs:**

- intermediate_data/sumstats/sumstats_msa_all.csv, sumstats_msa_raceex.csv, sumstats_msa_racenr.csv
- intermediate_data/sumstats/msa_matched_bpop.csv, msa_matched_wpop.csv
- raw_data/msa_to_state_xwalk/msa_to_states_xwalk.csv
- raw_data/msas/msa_names_2020.csv

- **Outputs:**

- intermediate_data/regression_data/msa_pred_rawdata.csv (population by race × subpop for use as predictors)
- intermediate_data/regression_data/msa_reg_outcomes.csv

prep_14_regressions_predictors.R

Assembles the MSA × year panel of predictor variables: population shares (Black/other/college-educated/foreign-born/migrant), industry shares, segregation indices, unionization rate, incarceration rates (state Census-based for 1969/1979, Vera-based for later years), and historical racial regime measures.

- **Inputs:**

- intermediate_data/regression_data/msa_pred_rawdata.csv (from prep_13)

- intermediate_data/predictors/msa_census_emp_all.csv (from prep_8)
- intermediate_data/predictors/msa_segregation_nhgis.csv (from prep_9)
- intermediate_data/predictors/msa_segregation_ltdb.csv (from prep_10)
- intermediate_data/predictors/union_50oneyr_cbsa.csv, union_50oneyr_st.csv, union_statewide.csv (from prep_7)
- intermediate_data/predictors/racialoppr_msa_wfactor.csv, racialoppr_st_wfactor.csv (from prep_12)
- raw_data/incarceration/incarceration_trends_state.csv
- **Outputs:** intermediate_data/regression_data/msa_reg_predictors.csv.

Analysis code

All analysis programs are in `code/analysis_code/` and produce output (figures, tables, intermediate CSVs) under `output/`.

`analysis_1_regressions.R`

Estimates the baseline regression and all robustness specifications described in the paper. The baseline model regresses the log Black/white median family-income ratio (and separately, log Black and log white medians, and the income rank gap) on MSA population, demographic composition, industry mix, unionization, incarceration, segregation, and the HRR factor. Twenty-plus alternative specifications cover different income definitions, samples, segregation measures, HRR specifications, and weighting schemes. Models are estimated separately for each year (1969, 1979, 1989, 1999, 2007, 2012, 2017) using `lm_robust` (heteroskedasticity-robust SEs) and, in parallel, `lm.beta` for standardized coefficients. Computes variance inflation factors for the baseline specification (Appendix Table A4.4).

- **Inputs:**
 - intermediate_data/regression_data/msa_reg_outcomes.csv
 - intermediate_data/regression_data/msa_reg_predictors.csv
- **Outputs (in `output/regressions/`):**
 - `all_coefs.csv` (one row per coefficient × model × year × outcome)
 - `all_r2s.csv` (one row per model × year × outcome)
 - `vifs_baseline.csv`
 - Per-sample coefficient files in `output/regressions/fulloutput/<incsamp>/`

`analysis_2_reg_sigchange.R`

For every model and every coefficient, computes pairwise z-tests of the change in coefficient between years, and creates a heatmap of statistically significant changes (Appendix Figure A4.1).

- **Inputs:** `output/regressions/all_coefs.csv` (from `analysis_1`).
- **Outputs (in `output/regressions/sig_ch/`):** `sig_test_<incsamp_mod>.csv` and `coef_ch_<incsamp_mod>.pdf` for each model, plus a combined `output/regressions/coef_ch_three_panels.pdf` (median ratio, Black median, white median).

`analysis_3_decomp_rr.R`

Performs the regression-based decomposition reported in the paper: holding the consistent-MSA sample fixed, decomposes the change in the South-vs.-non-South gap in log income ratios into components attributable to changing covariate distributions, changing coefficients, and changing residuals. Uses the standardized version of the baseline model from `analysis_1`. Produces Figure 7 in the main text.

- **Inputs:**
 - `intermediate_data/regression_data/msa_reg_outcomes.csv`
 - `intermediate_data/regression_data/msa_reg_predictors.csv`
- **Outputs (in `output/simulations/`):**
 - `slong_std.csv` (full decomposition table)
 - `dta_explode_17_std.csv`, `dta_explode_69_std.csv` (waterfall data)
 - `chart_simulations_explode_17dist_std.pdf`, `chart_simulations_explode_69dist_std.pdf` (waterfall charts)

`analysis_4_cormatrix.R`

Computes population-weighted correlation matrices among the regression covariates (Appendix Table A4.5)

- **Inputs:** same as `analysis_3`.
- **Outputs (in `output/regressions/`):** `cormat_pop_b_<year>.csv` for each year, plus `cormat_allyears.csv`.

`graphs_1_trends.R`

Creates Appendix Figure A1.1, graphing average Black-white median family income ratios over time by region of the country.

- **Inputs:** `intermediate_data/sumstats/sumstats_nat_raceex.csv`, `sumstats_region_raceex.csv` (from `prep_5`).
- **Outputs:** `output/trends/trends_south.pdf`.

`graphs_2_maps_outcomes.R`

Maps the Black-white median income ratio (and the income rank gap) for each MSA, separately for each year and for the change between 1969 and 2017.

Creates Figures 1 and 2 in the main text and Appendix Figures A2.1-A2.5.

- **Inputs:**
 - intermediate_data/regression_data/msa_reg_outcomes.csv
 - raw_data/msas/list1_2020.csv
 - County shapefiles downloaded via the tigris package
- **Outputs (in output/maps/outcomes/ and output/maps/):** PDF maps for each <incsamp>/<racedef>/<year> combination, plus the paper-facing map_paper_panel.pdf and map_fig2_ch_<...>.pdf.

graphs_3_maps_medinc.R

Maps each MSA's Black and white median family income as a fraction of the national Black and white medians, for each year and as a change between 1969 and 2017 (Appendix Figures A3.1 and A3.2). Also produces the Figure 3 scatterplot of metro-level Black-income growth vs. white-income growth.

- **Inputs:**
 - intermediate_data/regression_data/msa_reg_outcomes.csv
 - intermediate_data/sumstats/sumstats_nat_raceex.csv
 - raw_data/msas/list1_2020.csv, raw_data/msas/msa_names_2020.csv
 - County shapefiles via tigris
- **Outputs:** PDF maps in output/maps/medinc_vs_nat/, the scatter output/maps/fig3_sc_grow.pdf, and a CSV output/trends/mchange.csv (which is consumed by graphs_6).

graphs_4_reg_coefs.R

Creates the regression coefficient plots for the main text (Figures 4–6) and Appendix (Appendix Figures A5.1-A6.1).

- **Inputs:** output/regressions/all_coefs.csv (from analysis_1).
- **Outputs:** PDF coefficient plots in output/regressions/coef_plots_paper/.

graphs_5_reg_table.R

Outputs regression coefficient tables. Creates Appendix Tables A4.1-A4.3 (baseline specifications) and A5.1-A6.1 (alternative specifications).

- **Inputs:** output/regressions/all_coefs.csv, output/regressions/all_r2s.csv (from analysis_1).
- **Outputs:** CSV tables in output/regressions/tables/.

`graphs_6_statistics_for_paper.R`

Recomputes every numerical statistic quoted in the body of the paper (example MSAs, ratios, decomposition shares, illustrative coefficient effects, R^2 values, etc.) and writes them to a single CSV. Each row records the page of the manuscript where the statistic appears, a short description, and the computed value, so a reviewer can trace every quoted number back to the underlying data.

- **Inputs:**

- `intermediate_data/regression_data/msa_reg_outcomes.csv`
- `intermediate_data/regression_data/msa_reg_predictors.csv`
- `intermediate_data/sumstats/sumstats_nat_raceex.csv`,
`sumstats_region_raceex.csv`
- `output/trends/mchange.csv` (from `graphs_3`)
- `output/regressions/all_r2s.csv` (from `analysis_1`)
- `output/regressions/fulloutput/all_faminc_sqrt/coefs_r_all_faminc_sqrt_med_ldiff_expansive_Baseline.csv` (from `analysis_1`)
- `output/simulations/dta_explode_17_std.csv`, `dta_explode_69_std.csv` (from `analysis_3`)
- `raw_data/msas/msa_names_2020.csv`

- **Outputs:** `output/paper_statistics/paper_statistics.csv`.

Output

All output goes under `output/`:

- `output/crosswalks/` – PUMA-to-MSA crosswalks (from `prep_2`).
- `output/regressions/` – Regression coefficient files, R^2 , VIFs, correlation matrices, significance-change tests, coefficient plots, and tables.
- `output/maps/` – Choropleth maps of outcomes and change.
- `output/trends/` – Trend chart and the `mchange.csv` change dataset.
- `output/simulations/` – Decomposition output (data and waterfall charts).
- `output/paper_statistics/` – CSV listing every numerical statistic quoted in the paper, with page references.

References

- Baker, Regina S. 2022. “The Historical Racial Regime and Racial Inequality in Poverty in the American South.” *American Journal of Sociology* 127 (6): 1721–81. <https://doi.org/10.1086/719653>.
- Eckert, Fabian, Andrés Gvirtz, Jack Liang, and Michael Peters. 2020. *A Method to Construct Geographical Crosswalks with an Application to US Counties since 1790*. NBER Working Paper No. 26770. <https://doi.org/10.3386/w26770>.

- Hirsch, Barry T., David A. Macpherson, and William E. Even. 2026. "Union Membership, Coverage, and Earnings from the CPS." <https://unionstats.com/>.
- Hirsch, Barry T., David A. Macpherson, and Wayne G. Vroman. 2001. "Estimates of Union Density by State." *Monthly Labor Review* 124 (7): 51–55.
- Logan, John R., Zengwang Xu, and Brian Stults. 2014. "Interpolating U.S. Decennial Census Tract Data from as Early as 1970 to 2010: A Longitudinal Tract Database." *The Professional Geographer* 66 (3): 412–20. <https://doi.org/10.1080/00330124.2014.905156>.
- Manduca, Robert. 2026. "Shift, Not Stasis: The Geography of Post–Civil Rights Racial Inequality." *American Journal of Sociology*, ahead of print. <https://doi.org/10.1086/741817>.
- Office of Management and Budget. 2020. *Revised Delineations of Metropolitan Statistical Areas, Micropolitan Statistical Areas, and Combined Statistical Areas, and Guidance on Uses of the Delineations of These Areas*. OMB Bulletin No. 20–01. <https://www.whitehouse.gov/wp-content/uploads/2020/03/Bulletin-20-01.pdf>.
- Ruggles, Steven, Sarah Flood, Matthew Sobek, et al. 2025. "IPUMS USA: Version 16.0." Minneapolis, MN: IPUMS. <https://doi.org/10.18128/D010.V16.0>.
- Schroeder, Jonathan, David Van Riper, Steven Manson, et al. 2025. "National Historical Geographic Information System: Version 20.0." Minneapolis, MN: IPUMS. <https://doi.org/10.18128/D050.V20.0>.
- US Census Bureau. 2025. *ACS Supplemental Poverty Measures (SPM) Research Files: 2009 to 2019, 2021–2023*. <https://www.census.gov/data/datasets/time-series/demo/supplemental-poverty-measure/acs-research-files.html>.
- Vera Institute of Justice. 2025. "Incarceration Trends." <https://trends.vera.org/>.