Overview

The code in this replication package constructs the regression tables and figures using the contained datasets. Five main files run all of the code to generate the data for Table 1 and Figures 2-5 in the main text and Tables D.3-D.5, D.7, and D.10 in the Online Appendix. Each code file should complete in less than 1 minute.

Data Availability and Provenance Statements

The following datasets are used in the analysis:

1. 1992-1994 Multi-City Study of Urban Inequality, 1992-1994, from ICPSR (Bobo et al., 1994). The dataset cannot be publicly distributed under ICPSR's user agreement but are freely available to researchers at member institutions from ICPSR at https://www.icpsr.umich.edu/web/ICPSR/studies/2535. If obtained, the do-file "MCSUI_colourism_do" replicates all analyses using this dataset directly from the raw download.

2. Project on Ethnicity and Race in Latin America (Telles et al., 2015). Accessed from https://www.edwardtelles.com/perla-project-data. The relevant dataset for the analyses is contained in the replication file.

3. 2014 AmericasBarometer survey (all countries merged) from the Latin American Public Opinion Project (LAPOP). A limited dataset containing only the variables necessary to replicate the analyses is included in the replication package. The original source files cannot be transferred under LAPOP's user agreement but are available to researchers on the Center for Global Democracy/LAPOP Lab website at https://www.vanderbilt.edu/lapop/raw-data.php.

4. 1872 Brazilian Census records. Accessed from the Princeton University Library, original scans publicly accessible on the library DataSpace at https://dataspace.princeton.edu/handle/88435/dsp01h989r5980.

5. 1880 Full Count U.S. Census records from IPUMS USA (Ruggles et al., 2024). The dataset cannot be publicly distributed under IPUMS USA's user agreement but are freely available on the IPUMS website at https://usa.ipums.org/usa/. Details on the variables needed for the analyses are listed below.

6. 1800 U.S. National Historical Geographic Information System data from IPUMS NHGIS (Manson et al., 2024). The dataset was obtained from IPUMS NHGIS at https://www.nhgis.org/ and is publicly available.

Statement about Rights

- I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.

- I certify that the author(s) of the manuscript have documented permission to redistribute/publish the data contained within this replication package.

Summary of Availability

- Some data **cannot be made** publicly available.
- Confidential data used in this paper and not provided as part of the public replication package will be preserved for 2 years after publication

Details on each Data Source

Data Name	Data Files	Location	Provided	Citation
"Multi-City Study of Urban Inequality	"mcsui_usa.dta"	ICPSR	FALSE	Bobo et al. (1994)
PERLA Survey	"perla_brazil.dta"	data/	TRUE	Telles et al. (2015)
Americas Barometer Survey	"americas_barometer.dta"	data/	TRUE	
Brazilian Census Records	"brazil_census1872.dta"	data/	TRUE	
1880 Full Count U.S. Census	"us_census1880.dta"	IPUMS USA	FALSE	Ruggles et al. (2024)
1800 U.S. NHGIS County Statistics	"1800_countystats.dta"	data/	TRUE	Manson et al. (2020)

Dataset list

Data file	Source	Notes	Provided
`data/mcsui_usa.dta`	ICPSR	Available to researchers at member institutions	No
`data/perla_brazil.dta`	PERLA Project		Yes
`data/americas_barometer.dta`	LAPOP	Limited dataset, only contains variables used in analysis.	Yes
`data/brazil_census1872.dta`	Princeton Library	Hand coded from original scans on Library DataSpace	Yes
`data/us_census1880.dta`	IPUMS USA	Available from IPUMS USA	No

`data/1800_countystats.dta`	IPUMS NHGIS	Full dataset available from IPUMS NHGIS	Yes

Details on External Downloads

For replication, two datasets must be downloaded which are not included in this replication file: "mcsui_usa.dta" and "us_census1880.dta."

The MCSUI can be downloaded directly from ICPSR as described above and does not require any specific variable request.

The U.S. 1880 full count census data can be downloaded from IPUMS USA and requires the following variables to complete the replication analysis (all variables in 1880 USA Full Count):

- stateicp = State ICPSR code
- countyicp = County ICPSR code
- raced = Detailed race codes (included under variable race) – note: only individuals with raced = 200 or 210 are required for the analysis. The do-file will automatically drop all other records; however, to reduce file size, other records do not need to be downloaded.
- lit = Literacy status
- labforce = Labour force status
- sei = Duncan Socioeconomic Index score
- age = Age
- sex = Sex
- urban = Urban status

Computational requirements

Software Requirements

Stata (code was last run with version MP18.5)

- `binscatter (as of 2025-01-14)`

Memory, Runtime, Storage Requirements

Approximate time needed to reproduce the analyses on a standard 2025 desktop machine is <10 minutes. Approximate storage space needed: < 25 Mbytes. The code was last run on an Intel-based laptop i9-13900H, 2600Mhz, 14 cores with MicroSoft Windows Home with 16.0GB of installed RAM.

Description of programs/code

Do-files in data/ will open and run the associated datasets and construct the tables / figures contained within the paper and online appendix. A list of do-files, their corresponding data files, and the tables / figures they output are as follows:

- mcsui_colourism_do: Calls file "mcsui_usa.dta" and produces Figure 2 (left-hand graph) and results included in Appendix Table D.3 and Appendix Table D.4.
- perla_colourism_do: Calls file "perla_brazil.dta" and produces results included in Appendix Table D.3
- americas_barometer_do: Calls file "americas_barometer.dta" and produces results included in Table 1, Figure 2 (right-hand graph), Figure 3, and results included in Appendix Tables D.3 and D.5.
- brazil_census_do: Calls file "brazil_census1872.dta" and produces Figure 4 and results included in Appendix Table D.7.
- us_census1880_do: Calls files "us_census1880.dta" and "1800_countystats.dta" and produces Figure 5 and results included in Appendix Table D.10.

Instructions to Replicators

- Setup Stata including binscatter package to set up the working environment.
- Download the data files references above. Each should be stored in the working folder using the exact file names as listed above. "mcsui_usa.dta" and "us_census1880.dta" must be downloaded from the listed sources and converted to .dta files before running the do files.
- Each do-file is self-contained and can be run from Stata launch screen to produce the listed Tables / Figures.

List of tables and figures

- The enclosed files reproduce selected tables and figures in the paper, as explained below.

Figure/Table #	Program	Line Number	Note
Figure 1	N/A	N/A	Constructed by hand as described in Appendix C.1
Figure 2	mcsui_colourism_do (left-hand graph) and americas_barometer_do (right-hand graph)	17 / 45	
Figure 3	americas_barometer_do	47 & 48	

Figure 4	brazil_census_do	15	
Figure 5	us_census1880_do	10	Requires 1880 U.S. full-count census dataset.
Table 1	americas_barometer_do	40-43	
Table D.1	N/A	N/A	Summary statistics constructed directly using datasets "mcsui_usa.dta" and "perla_brazil.dta"
Table D.2	N/A	N/A	Summary statistics constructed directly using dataset "americas_barometer.dta"
Table D.3	mcsui_colourism_do (column 1), perla_colourism_do (column 2), and americas_barometer_do (column 3)	19 / 20 / 50	
Table D.4	mcsui_colourism_do	21-25	Summary statistics calculated manually using dataset "mcsui_usa.dta"
Table D.5	americas_barometer_do	52-54	
Table D.6	N/A	N/A	Summary statistics constructed directly using dataset "brazil_census1872.dta"
Table D.7	brazil_census_do	17-18	
Table D.8	N/A	N/A	Summary statistics constructed directly using dataset "us_census1880.dta"
Table D.9	N/A	N/A	Summary statistics constructed directly using dataset "1800_countystats.dta"
Table D.10	us_census1880_do	12-14	Requires 1880 U.S. full-count census dataset.

References

Bobo, Lawrence, James Johnson, Melvin Oliver, Reynolds Farley, Barry Bluestone, Irene Browne, Sheldon Danziger, Gary P. Green, Harry Holzer, Maria Krysan, Michael Massagli, Camille Z. Charles (1994). Multi-city study of urban inequality, 1992-1994. Inter-University Consortium for Political and Social Research.

Manson, Steven, Jonathan Schroeder, David Van Riper, Katherine Knowles, Tracy Kugler, Finn Roberts, and Steven Ruggles. IPUMS National Historical Geographic Information System: Version 19.0 [dataset]. Minneapolis, MN: IPUMS. 2024.

Ruggles, Steven, Sarah Flood, Matthew Sobek, Daniel Backman, Annie Chen, Grace Cooper, Stephanie Richards, Renae Rogers, and Megan Schouweiler. IPUMS USA: Version 15.0 [dataset]. Minneapolis, MN: IPUMS, 2024.

Telles, Edward E., Rene D. Flores, and Fernando Urrea-Giraldo (2015). Pigmentocracies: Educational inequality, skin color and census ethnoracial identification in eight Latin American countries. Research in Social Stratification and Mobility 40, 39-58.