README FILE

To replicate the results contained in

"Soil Heterogeneity, Social Learning, and the Formation of Close-knit Communities"

Itzchak Tzachi Raz, The Hebrew University of Jerusalem, iraz@mail.huji.ac.il

(November 2024)

This README file describes the contents of the replication folder. The folder contains all code needed to replicate the results of the paper and online appendix. Two separate *.R files that reproduce the tables and figures are located in the ./code/ folder. Multiple *.csv files contain the data needed to replicate the results are located in the ./data/ folder.

I am unable to publicly post some of the data used in the analysis. First, I obtained access to individual-level, complete-count 1850-1940 Population Census data including string variables through an agreement with the Minnesota Population Center. Other researchers can apply to use the restricted version of the data, as described here [link accessed November 20, 2024]. Second, I obtained access to individual-level data from the "Moral Foundations Questionnaire" (MFQ) surveyed on www.yourmorals.org. To request access to the data, researchers can contact YourMorals research team [link accessed November 20, 2024].

CODES

- ./code/Tables.R – a code to produce all of the tables in the paper and the online appendix.
- ./code/Figures.R – a code to produce all of the figures in the paper and the online appendix.

There is no need to change paths to run the codes.

OUTPUT

All replicated tables and figures are saved in the relevant subdirectories of the ./output/ folder. Tables are saved as *.tex files and figures as *.jpg files.

DATA

The following data files are needed to produce the figures and tables in the paper and online appendix. I flag restricted-access data with red color text and an *.

- CountyLevelData.csv – a county-level dataset 1850-1940 used in Table 1, Figure 2, Appendix Tables C.1, C.2, C.8, C.9, C.10, C.11, C.12, C.13, D.1, D.2, D.3, D.4, D.5, and Appendix Figures B.22, C.1, C.2, C.3, C.9, and D.1.
- SelectionFarmingData.csv* – a subset of restricted-access, individual-level, complete-count linked Census data 1850-1880 used in Table 2 and Appendix Table D.6.
- MigrantsNative_2.csv* – a subset of restricted-access, individual-level, complete-count Census data 1850-1880 used in Tables 2 and 3, Figure 3, Appendix Tables C.4, C.5, D.6, and D.7, and Appendix Figures C.4, C.5, C.6, C.7, C.8, C.9, D.2, and D.3.
- MigrantsChildrenData.csv* – a subset of restricted-access, individual-level, complete-count linked Census data 1850-1910 used in Table 4, Appendix Tables D.9, D.10, D.11, and D.12, and Appendix Figure C.9.
- FertilizersData.csv – a county-level dataset 1910-1930 used in Table 5, Appendix Tables D.13, D.14, D.16, D.17, and D.18 and Appendix Figure D.4.
- WheatShareData.csv – a county-level dataset 1880-1935 used in Table 5, Appendix Tables D.15, D.16, D.17, and D.18 and Appendix Figure D.4.
- OutMigrationData.csv* - a subset of restricted-access, individual-level, complete-count linked Census data 1850-1880 used in Table 6, and Appendix Tables C.6, C.7, D.19, D.20, and D.21.
- AgriculturalDiversityData.csv – a county-level dataset 1880-1935 used in Appendix Table A.2.
- LongRunData.csv – a county-level dataset used in Figure 1, Appendix Tables A.1, A.2, and C.3, and Appendix Figure A.1.

- LNIexamples.csv – a subset of state X name-level dataset 1940 used in Appendix Table B.1.
- MFQData.csv* – an individual-level dataset of the "Moral Foundations Questionnaire" (MFQ) surveyed on www.yourmorals.org, used in Appendix Table C.3.
- ValidationData.csv – a county-level dataset used in Appendix Figures B.1-B.21.
- ValidationDataStates.csv – a state-level dataset used in Appendix Figures B.15-B.21.
- SlaveShareData.csv – a county-level dataset 1850-1940, harmonized to 1940 county borders, used in Appendix Table C.11.
- CountyLevelDataFrontierStatus.csv – a county-level dataset 1850-1880, harmonized to 2000 county borders, used in Appendix Table C.14.
- CountyLevelData1940TFE.csv – a county-level dataset 1940, harmonized to 2000 county borders, used in Appendix Table C.14.
- Migrants_2.csv* – a subset of restricted-access, individual-level, complete-count Census data 1850-1880 used in Appendix Table D.8.
- MigrantsNativeAllRace_2.csv* – a subset of restricted-access, individual-level, complete-count Census data 1850-1880 used in Appendix Table D.8.
- MigrantsNative.csv* – a subset of restricted-access, individual-level, complete-count Census data 1850-1880 used in Appendix Table D.8.
- WheatProductionData.csv – a county-level dataset 1840-1935 used in Appendix Table D.13.

Details on the original raw data used and the construction of each variable are provided in Appendix E.