READ ME FILE to replicate the results contained in "Pre-Colonial Warfare and Long-Run Development in India" by Mark Dincecco, James Fenske, Anil Menon, and Shivaji Mukherjee.

October 11, 2021.

Data availability and provenance statements

Below, we describe the data that is provided in the replication folder. Details on each variable are provided in the paper. If not specified, the file extension is .dta.

Figures made in Stata

/*Table 1: Main Results*/
conflictexposure
- Conflict exposure computed from our geocoding of Jaques (2007).
- Luminosity from National Geophysical Data Center (2010).

/*Table 2: IV*/
conflictexposure
- Ruggedness from Nunn and Puga (2012).

/*Table 3: Pre-Colonial-Era State-Making*/
conflictexposure
- Important Mughal Sites from Schwartzberg (1978).
- Territories of emperors from Jha (2013).

/*Table 4: Colonial Fiscal Development*/
conflictexposure_native
leetaxdata
- Fiscal data for 1881 from Baness (1881).
- Fiscal data for 1931 from Lee (2019)/

/*Table 5: Colonial and Post-Colonial Conflict*/
conflictexposure
- Conflict exposure in all periods computed from our geocoding of Jaques (2007).

/*Table 6: Conflict and Post-Colonial Political Violence*/
conflictexposure
mukherjeedata
- Political violence from Raleigh et al (2010).
- Maoist violence from Mukherjee (2018).
- Fractionalization from Census of India (2001).

/*Table 7: Irrigation Infrastructure*/
vdsadataclean
exposure_bi
conflictexposure
- Irrigation in 1931 from Bharadwaj and Mirza (2019).
- Later irrigation from Village Dynamics in South Asia (2012).

- Yield from Banerjee and Iyer (2005).
- % Non Agriculture from Census of India (2011).

/*Table 8: Literacy*/
literacy1881
literacy1921
exposure_bi_literacy
devcensuswgis
- Literacy in 1881, 1921, and 2011 from Census of India (1881, 1921, 2011).
- Literacy 1961-91 from Banerjee and Iyer (2005).

/*Table 9: Education*/
bi_educ_exposure
bi_infmort_exposure
- Education and mortality from Banerjee and Iyer (2005).

/*Figure 1: Pre-Colonial Conflict Exposure and Luminosity: India*/
IND_adm2.shp (and .cpg, .dbf, .prj, .sbn, .sbx, and .shx)
conflictexposure
- As in Table 1.

/*Figure 2: Pre-Colonial Conflict Exposure and Luminosity: Uttar Pradesh*/
IND_adm2.shp (and .cpg, .dbf, .prj, .sbn, .sbx, and .shx)
conflictexposure
- As in Table 1.

/*TABLE A.1: Summary Statistics: Main Analysis*/
conflictexposure
conflictexposure_native
leetaxdata
mukherjeedata
vdsadataclean
exposure_bi
conflictexposure
literacy1881
literacy1921
exposure_bi_literacy
devcensuswgis
bi_educ_exposure
bi_infmort_exposure
- Sources of all variables are outlined in the tables where they are used for analysis.

/*TABLE A.2: Summary Statistics: Appendix Analysis*/
cells
tehsildata
conflictexposure
districtgdp
bgdpak
biclean
- Sources of all variables are outlined in the tables where they are used for analysis.

/*TABLE A.3: Unit of Analysis: Grid Cell or Sub-District (i.e. Tehsil)*/
cells
tehsildata
- As in Table 1.

/*TABLE A.4: Alternative Specifications of Dependent Variable*/
conflictexposure
- As in Table 1.

/*TABLE A.5: Population Density Polynomials*/
conflictexposure
- Population density taken from Center for International Earth Science Information Network (2018).

/*TABLE A.6: No Population Density Control*/
conflictexposure
- As in Table 1.

/*TABLE A.7: Historical Population Density Control*/
conflictexposure
- Historical population density from Klein Goldewijk et al. (2010).

/*TABLE A.8: Grid Cell Fixed Effects*/
conflictexposure
- As in Table 1.

/*TABLE A.9: Conley Spatial Standard Errors*/
conflictexposure
- As in Table 1.

/*TABLE A.10: Alternative Standard Errors*/
x_ols.ado
conflictexposure
- As in Table 1.

/*TABLE A.11: Latitude and Longitude Polynomials*/
conflictexposure
- As in Table 1.

/*TABLE A.12: Conflict Types*/
conflictexposure
- As in Table 1.

/*TABLE A.13: 1500-1757 Conflict Sample*/
conflictexposure
- As in Table 1.

/*TABLE A.14: Alternative Conflict Data*/

conflictexposure
- Additional conflicts from Clodfelter (2002) And Naravane (1997).

/*TABLE A.15: 5,000 km Radius*/
conflictexposure
- As in Table 1.

/*TABLE A.16: Running End-Date Cutoff*/
conflictexposure
- Date of conquest from Banerjee and Iyer (2005).

/*TABLE A.17: VIIRS Data*/
conflictexposure
- Data on luminosity from National Aeronautics and Space Administration (2020).

/*TABLE A.18: GDP Outcome*/
districtgdp
- India district GDP from Planning Commission (2014).

/*TABLE A.19: Initial State Capacity Controls*/
conflictexposure
- Settlements and sites from Important Mughal Sites from Schwartzberg (1978).
- Urban population from Chandler (1987).
- States from Nag (2007).

/*TABLE A.20: Additional Geographic Features*/
conflictexposure
- Coast and River from Patterson and Kelso (2019).
- Irrigation potential from Bentzen et al. (2017).
- Rainfall variation from Matsuura and Willmott (2009).
- Resource deposits from Tollefsen et al (2012).
- Forest from India Institute of Forest Management, Center for Geoinformatics (2015).

/*TABLE A.21: Colonial Controls*/
conflictexposure
biclean
- Direct rule from Iyer (2010).
- Non-landlord from Banerjee and Iyer (2005).

/*TABLE A.22: Control for Colonial Railroad Network*/
conflictexposure
- Railroad from Fenske et al. (2021).

/*TABLE A.23: Exclude Canal Colonies*/
conflictexposure
- As in Table 1.

/*TABLE A.24: Fractionalization Controls*/
conflictexposure

- Ports and Muslim rule from Jha (2013).
- Data on religion, caste, and tribe from Census of India (2001).
- Data on language from Omid (2015).
- Ganges from Patterson and Kelso (2019).

/*TABLE A.25: Post-1757 Conflict Controls*/
conflictexposure
- As in Table 1.

/*TABLE A.26: Controls for Urban Distances*/
conflictexposure
- As in Table 1.

/*TABLE A.27: Asian Highway Control*/
conflictexposure
- Asian Highway from Environmental Systems Research Institute (1992).

/*TABLE A.28: Pairwise Correlations: Alternative Exposure Measures*/
conflictexposure
- As in Table 1.

/*TABLE A.29: Alternative Conflict Exposure Measures*/
conflictexposure
- As in Table 1.

/*TABLE A.30: IV: Reduced Form*/
conflictexposure
- Ruggedness from Nunn and Puga (2012).

/*TABLE A.31: IV: Balance Tests: Historical Trade*/
conflictexposure
- Historical Trade Routes from Raychaudhuri (1982).
- Silk Road Sites from United Nations Educational, Scientific, and Cultural Organization (2010).
- Medieval Ports from Jha (2013).

/*TABLE A.32: IV: Exclude Districts with Historical Trade*/
conflictexposure
- As in Table A31

/*TABLE A.33: IV: Historical Trade Controls*/
conflictexposure
- As in Table A31

/*TABLE A.34: IV: Historical Trade Controls (Cost Distance)*/
conflictexposure
- As in Table A31

/*TABLE A.35: IV: Placebo Entry Points*/

conflictexposure
- As in Table A31

/*TABLE A.36: IV: Alternative Cost-Distance*/
conflictexposure
- HMI from Ozak (2018).

/*TABLE A.37: OLS: Include Pakistan and Bangladesh*/
conflictexposure
bgdpak
- As in Table 1.

/*TABLE A.38: IV: Include Pakistan and Bangladesh*/
conflictexposure
bgdpak
- As in Table 2.

/*TABLE A.39: IV: Control for Colonial Railroad Network*/
conflictexposure
- As in Table A22.

/*TABLE A.40: Pre-Colonial-Era State-Making: Control for Sieges*/
conflictexposure
- As in Table 1.

/*TABLE A.41: Colonial Fiscal Development: Control for Pre-Colonial State-Making*/
conflictexposure
- As in Table 3.

/*TABLE A.42: Colonial and Post-Colonial Conflict: Sub-District as Unit of Analysis*/
tehsildata
- As in Table 5.

/*TABLE A.43: Colonial and Post-Colonial Conflict: Control for Pre-Colonial State-Making*/
tehsildata
- As in Table 5.

/*TABLE A.44: Do Historical Investments Persist?*/
exposure_bi
exposure_bi_literacy
devcensuswgis
- As in Tables 7, 8 and 9.

/*Figure A.3: Artificial Spatially-Correlated Noise Placebo Variables*/
conflictexposure
- As in Table 1.

/*Figure A.4: Exclude States One by One*/
conflictexposure

- As in Table 1.

/*Figure A.5: Exclude Colonial Provinces*/
conflictexposure
- As in Table 1.

/*Figure A.6: Yearly Luminosity Data*/
conflictexposure
- As in Table 1.

/*Figure A.8: Cost Distance from Khyber Pass*/
conflictexposure
- As in Table 2.

/*Figure A.9: Khyber Proximity IV*/
conflictexposure
- As in Table 2.

/*Section 5.1: Compute Standardized Beta from Banerjee-Iyer 2005, Table 1, Column 1*/
yld_sett_aug03
- Data on all variables from Banerjee and Iyer (2005).

/*Footnote: Exclude Districts for which Conflict Exposure=0*/
conflictexposure
- As in Table 1.

/*Footnote: Quadratics for Geographic Controls*/
conflictexposure
- As in Table 1.

/*Footnote: 1857 as Cutoff for Colonial Sub-Periods*/
conflictexposure
- As in Table 1.

/*Footnote: Distance to Presidency City*/
conflictexposure
- As in Table 1.

Figures made in ArcGIS

/*Figure A.1: Conflict Locations, 1000-2010*/
formapping.csv
- Conflicts taken from Jaques (2007).

/*Figure A.2: Conflict Locations by Sub-Period*/
formapping.csv
- As in Figure A1.

/*Figure A.7: Convex Hull Example: Pre-Colonial Land Battles Involving Shah Jahan*/

shahjahan.shp (and .cpg, .dbf, .prj, .sbn, .sbx, and .shx)
- As in Figure A1

/*Figure A.10: Locations of Conflicts Fought By Invaders from Central Asia, 1000-1757*/
centasian.csv
- As in Figure A1

Dataset list

Data file	Source	Notes	Provided
conflictexposure.dta	As listed	The main dataset of the paper, containing data on conflict exposure, luminosity, and controls.	Yes
bgdpak.dta	As listed	Conflict exposure, luminosity, and control variables for Bangladesh and Pakistan.	Yes
biclean.dta	As listed	Conflict exposure, luminosity, and control variables for the districts in Banerjee and Iyer 2005, as well as several variables from their work.	Yes
bi_educ_exposure.dta	As listed	Conflict exposure, luminosity, and control variables for the districts in Banerjee and Iyer 2005, as well as education variables from their work.	Yes
bi_infmort_exposure.dta	As listed	Conflict exposure, luminosity, and control variables for the districts in Banerjee and Iyer 2005, as well as mortality variables from their work.	Yes
cells.dta	As listed	Conflict exposure, luminosity, and control variables at the grid cell level.	Yes
centasian.csv	As listed	Conflict events with Central Asian actors for mapping in ArcGIS.	Yes
conflictexposure_native.dta	As listed	Conflict exposure, luminosity, and control variables for princely states, as well as fiscal data from Baness (1881).	Yes
devcensuswgis.dta	As listed	Conflict exposure, luminosity, and control variables for census districts, as well as development data from DevInfo and CensusInfo.	Yes
districtgdp.dta	As listed	Conflict exposure, district GDP, and control variables.	Yes
exposure_bi.dta	As listed	Conflict exposure, luminosity, and control variables for the districts in Banerjee and Iyer 2005, as well as several variables from their work.	Yes
exposure_bi_literacy.dta	As listed	Conflict exposure, luminosity, and control variables for the districts in	Yes

		Banerjee and Iyer 2005, as well as literacy variables from their work.	
formapping.csv	As listed	Conflict events and start dates for mapping in ArcGIS.	Yes
IND_adm2.cpg	As listed	District-level boundaries of India from gadm.org.	Yes
IND_adm2.dbf	As listed	District-level boundaries of India from gadm.org.	Yes
IND_adm2.prj	As listed	District-level boundaries of India from gadm.org.	Yes
IND_adm2.sbn	As listed	District-level boundaries of India from gadm.org.	Yes
IND_adm2.sbx	As listed	District-level boundaries of India from gadm.org.	Yes
IND_adm2.shp	As listed	District-level boundaries of India from gadm.org.	Yes
IND_adm2.shx	As listed	District-level boundaries of India from gadm.org.	Yes
leetaxdata.dta	As listed	Conflict exposure, luminosity, and control variables, as well as fiscal data from Lee (2019).	Yes
literacy1881.dta	As listed	Conflict exposure, luminosity, and control variables for 1881 districts, with literacy from the colonial census.	Yes
literacy1921.dta	As listed	Conflict exposure, luminosity, and control variables for 1921 districts, with literacy from the colonial census.	Yes
mukherjeedata.dta	As listed	Conflict exposure, luminosity, and control variables, as well as data on left-wing extremism from Mukherjee (2018).	Yes
shahjahan.cpg	As listed	Conflict events with Shah Jahan as an actor for mapping in ArcGIS.	Yes
shahjahan.dbf	As listed	Conflict events with Shah Jahan as an actor for mapping in ArcGIS.	Yes
shahjahan.prj	As listed	Conflict events with Shah Jahan as an actor for mapping in ArcGIS.	Yes
shahjahan.sbn	As listed	Conflict events with Shah Jahan as an actor for mapping in ArcGIS.	Yes
shahjahan.sbx	As listed	Conflict events with Shah Jahan as an actor for mapping in ArcGIS.	Yes
shahjahan.shp	As listed	Conflict events with Shah Jahan as an actor for mapping in ArcGIS.	Yes
shahjahan.shx	As listed	Conflict events with Shah Jahan as an actor for mapping in ArcGIS.	Yes
tehsildata.dta	As listed	Conflict exposure, luminosity, and control variables at the sub-district level.	Yes

		Conflict exposure, luminosity, and control variables for the districts in the Village Dynamics Studies in South Asia, as well as agricultural variables	
vdsadataclean.dta	As listed	from this source.	Yes
x_ols.ado	As listed	An ado file written by Jean-Pierre Dube that helps implement spatial standard errors.	Yes
yld_sett_aug03.dta	As listed	Agricultural data from Banerjee and Iyer, used to compute the standardized beta from Banerjee-Iyer 2005, Table 1, Column 1.	Yes

Description of programs and instructions to replicators

The folder "Replication/Data/" contains all files that are needed to replicate the paper. It includes Stata .dta and .ado files, Excel .csv files, and spatial data that we have used in arcmap (.cpg, .dbf, .sbn, .shp, .shx, , and .prj). The folder "Replication" contains the do file wii_replication.do, which replicates almost all results reported in the paper and appendix using the datasets listed above. The exceptions, indicated in the do file, are maps created by opening the relevant file listed above and displaying it in ArcGIS. To run the file, change the path on line 17 of the do file. All tables and figures will be saved to "Replication/Output/".

Software used, packages and libraries, and running time

The Operating System is Windows 10 Pro. The software used is Stata 16.1 and ArcGIS Desktop 10.5.1. The running time is approximately half an hour.

References

Baness, J. Frederick. 1881. *Geographicus Indicus*. Edward Stanford.

Banerjee, Abhijit and Lakshmi Iyer. 2005. "History, Institutions, and Economic Performance: The Legacy of Colonial Land Tenure Systems in India." *American Economic Review* 95(4):1190–1213.

Bentzen, Jeanet, Nicolai Kaarsen, and Asger Wingender. 2017. "Irrigation and Autocracy." *Journal of the European Economic Association* 15(1):1–53.

Bharadwaj, Prashant and Rinchan Ali Mirza. 2019. "Displacement and Development: Long-Term Impacts of Population Transfer in India." *Explorations in Economic History* 73:101273.

Census of India. 1881. *Office of the Registrar General & Census Commissioner, India*.

Census of India. 1921. *Office of the Registrar General & Census Commissioner, India*.

Census of India. 2001. *Office of the Registrar General & Census Commissioner, India*.

Census of India. 2011. *Office of the Registrar General & Census Commissioner, India*.

Center for International Earth Science Information Network (CIESIN), Columbia University. 2018. "Documentation for the Gridded Population of the World, Version 4 (GPWv4), Revision 11 Data Sets." NASA Socioeconomic Data and Applications Center (SEDAC).

Chandler, Tertius. 1987. *Four Thousand Years of Urban Growth: A Historical Census*. St. David's University Press.

Clodfelter, Michael. 2002. *Warfare and Armed Conflicts: A Statistical Reference to Casualty and Other Figures, 1500-2000*. McFarland.

Environmental Systems Research Institute. 1992. "Digital Chart of the World."

Fenske, James, Namrata Kala and Jinlin Wei. 2021. "Railways and Cities in India." Centre for Competitive Advantage in the Global Economy Working Paper 559.

India Institute of Forest Management, Center for Geoinformatics. 2015. "India State of Forest Report."

Iyer, Lakshmi. 2010. "Direct versus Indirect Colonial Rule in India: Long-Term Consequences." *Review of Economics and Statistics* 92(4):693–713.

Jaques, Tony. 2007. *Dictionary of Battles and Sieges: A Guide to 8,500 Battles from Antiquity through the Twenty-First Century*. Greenwood Press.

Jha, Saumitra. 2013. "Trade, Institutions, and Ethnic Tolerance: Evidence from South Asia." *American Political Science Review* 107(4):806–32.

Klein Goldewijk, Kees, Arthur Beusen, and Peter Janssen. 2010. "Long-Term Dynamic Modeling of Global Population and Built-Up Area in a Spatially Explicit Way: HYDE 3.1." *The Holocene* 20(4):565–73.

Lee, Alexander. 2019. "Land, State Capacity, and Colonialism: Evidence from India." *Comparative Political Studies* 52(3):412–444.

Matsuura, Kenji and Cort Willmott. 2009. "Terrestrial Air Temperature: 1900-2008 Gridded Monthly Time Series." Center for Climatic Research, University of Delaware.

Mukherjee, Shivaji. 2018. "Colonial Origins of Maoist Insurgency in India: Historical Institutions and Civil War." *Journal of Conflict Resolution* 62(10):2232–2274.

Nag, Prithvish. 2007. *Historical Atlas of India*. NATMO.

National Aeronautics and Space Administration. 2020. "Visible Infrared Imaging Radiometer Suite."

Naravane, M.S. 1997. *Battles of Medieval India: AD 1295-1850*. APH Publishing.

National Geophysical Data Center. 2010. "Version 4 DMSP-OLS Nighttime Lights Time Series." National Oceanic and Atmospheric Administration.

Nunn, Nathan and Diego Puga. 2012. "Ruggedness: The Blessing of Bad Geography in Africa." *The Review of Economics and Statistics* 94(1):20–36.

Omid. 2015. "Peoples of South Asia Database."

Ozak, Omer. 2018. "Distance to the Pre-Industrial Technological Frontier and Economic Development." *Journal of Economic Growth* 23(2):175–221.

Patterson, Tom and Nathaniel Kelso. 2019. "Natural Earth."

Planning Commission. 2014. "Districtwise GDP and Growth Rate based at Current Price from 1999-00 to 2007-08."

Raleigh, Clionadh, Andrew Linke, Havard Hegre, and Joakim Karlsen. 2010. "Introducing ACLED-Armed Conflict Location and Event Data." *Journal of Peace Research* 47(5) 651-60.

Raychaudhuri, Tapan. 1982. "Inland Trade." In *Cambridge Economic History of India*, *Volume 1: c.1200-c.1750*, ed. Tapan Raychaudhuri and Irfan Habib. Cambridge University Press, chapter 11, pp. 325–359.

Schwartzberg, Joseph. 1978. *A Historical Atlas of South Asia*. University of Chicago Press.

Tollefsen, Andreas, Havard Strand and Halvard Buhaug. 2012. "PRIO-GRID: A Unified Spatial Data Structure." *Journal of Peace Research* 49(2):363–74.

United Nations Educational, Scientific, and Cultural Organization. 2010. "Silk Road Sites in India."

Village Dynamics in South Asia District Level Database Documentation. 2012. "19 States of India." Updated Database (Apportioned). Objective 2 of VDSA Project. Documentation of Files: 1966-67 to 2007-08. (1966 district boundaries). February 2012. Village Dynamics in South Asia (VDSA). ICRISAT-ICAR-IRRI Collaborative Research Project. Module I (ICRISAT Component).