

Data and Replication Guide

May 29, 2014

1 Data

All data sets used in this article are available online and were imported for processing in their original state.

- PRIO-GRID (prio.no/Data/PRIO-GRID/): Contains grid-cell shapefiles and attribute data on population and distance to the state capital for each cell. Population data in PRIO-GRID is taken from the CIESIN Gridded Population of the World data set for the years 2000 and 2005. Our analysis relies on the 2000 statistics.
- UCDP-PRIO Armed Conflict Data Set v4 (www.pcr.uu.se/research/ucdp/datasets/ucdp_prio_armed_conflict_dataset/): Contains country-conflict-year data for all African states. The data set was used to identify which country-years to observe – countries that did not have an ongoing conflict in a given year were not included for analysis.
- UCDP Georeferenced Event Dataset (GED) v1.5 (www.pcr.uu.se/research/ucdp/datasets/ucdp_ged/): Contains data on conflict events for Africa between the years 1989 and 2011. This data set was used as one of the outcome variables in our analysis.
- Armed Conflict and Location Event Dataset (ACLED) v3 (www.acleddata.com): Contains data on conflict events for Africa and a set of Asian nations between the years 1997 and 2013. This data set was used as one of the outcome variables in our analysis.
- Global Database of Events, Language, and Tone (GDELT) (www.gdeltproject.org): Contains data on a wide range of events for the entire world between the years 1979 and 2013. This data set was used as one of the independent variables in our analysis. Due to the size of the GDELT, we did not download yearly event data files directly: instead, we used an R package (GDELTtools) to directly subset the data prior to loading it into SQL for analysis. The commands used to download the GDELT subset are available with the rest of our replication code, but the results will be identical if the user chooses to download and process the full yearly GDELT files.

These data sets were loaded into a spatial PostgreSQL 9.3 database and processed using PostGIS 2.1. The processed data was then loaded into R 3.0.2 for analysis. The processed data is available online and uses two R scripts to create all tables and figures found in the main paper and appendices.

1. To replicate the tables and figures included in the paper, download the main data set ("analysisdata.csv") run the R script "gdeltanalysis.R". This will generate a set of figures as .pdf files, as well as a set of tables in LaTeX format for export. This analysis relies on a large number of external packages, so be sure these are installed properly before running.
2. To replicate the materials included in Appendix D, download the data sets "analysisdata.csv" and "analysisdata_AppendixD.csv" and run the R script "gdeltappendix.R". This will generate a set of figures as .pdf files, as well as a set of tables in LaTeX format for export.