# CODEBOOK
# Replication datasets for
# "Sectoral interests and regional bloc voting in African countries"
# Boone, Crespin-Boucaud, Kim (2024)

General note: Constituencies are our main unit of analysis and almost all variables are at constituency-level.

**Datasets with information on bloc membership and production**

**[country]_blocs_prod_info.dta** ( with country = Kenya, Malawi, or Zambia )

1. Electoral results
   a) morani_bloc: variable that indicates which (if any) bloc the constituency belongs to. The variable takes the value 99 for "non-bloc" constituencies.
   b) Dummies that indicate (Moran's I) bloc membership
2. General constituency information
   a) **global_id**: constituency identifier. It should be used to match the dataset with the shapefile of electoral constituencies.
   b) pop_2015: population of the constituency in 2015
      - See European Commission, Joint Research Centre (JRC) (2015): GHS-POP R2015A - GHS population grid, derived from GPW4, multitemporal (1975, 1990, 2000, 2015) - OBSOLETE RELEASE. European Commission, Joint Research Centre (JRC) [Dataset] PID: http://data.europa.eu/89h/jrc-ghsl-ghs_pop_gpw4_globe_r2015a

   c) area: area of the constituency (square km2, UTM projection)
3. Crop production (from IFPRI – see their codebook for a list of crops abbreviations)
   - See International Food Policy Research Institute, 2020, "Spatially-Disaggregated Crop Production Statistics Data in Africa South of the Sahara for 2017", https://doi.org/10.7910/DVN/FSSKBW, Harvard Dataverse, V2

   a) areal_unit_[crop]: Production per areal unit (constituency-level): This is computed as the ratio of the constituency's production (in % of national production) to the constituency's size (in % of country area). A production per areal unit of 1 indicates that the constituency produces as much of the crop as would be expected under a nation-wide uniform geographic distribution. Values greater than 1 indicate that the constituency produces more than what would be expected given its geographic size.
   b) areal_unit_[crop]_bloc: Production per areal unit (bloc-level): This is computed as the ratio of the bloc's production (in % of national production) to the bloc's size (in % of country area). A production per areal unit of 1 indicates that the bloc produces as much of the crop as would be expected under a nation-wide uniform geographic distribution. Values greater than 1 indicate that the bloc produces more than what would be expected given its geographic size.
   c) pop_[crop]: Production per capita (constituency-level): Total production (total_[crop]) divided by population (pop_2015).
   d) share_[crop]: Share of national production (constituency-level). Share of the country-level production produced in the constituency.
   e) total_[crop]: Total production (constituency-level).
4. Ethnic make-up and employment make-up of constituencies (from DHS)
   - See ICF. 2003-2015. Demographic and Health Surveys (various) [Datasets]. Funded by USAID. Rockville, Maryland: ICF [Distributor].

   a) agric, service, unemployed: share of respondents in the constituency depending on occupation

b) [ethnicgroup]: share of respondents in the constituency who declare belonging to this ethnic group

**Datasets with information on electoral data**

**[country]_electoral_data.dta** ( with country = Kenya, Malawi, or Zambia )

This dataset contains only the vote shares for the main parties in each country. The election results data are sourced from the electoral commissions of Kenya, Malawi, and Zambia.

N.B.:

- The electoral data as well as the electoral shapefiles were "recoded" to create a time-invariant panel of constituencies over the elections considered. See Appendix I for details.

1. **global_id**: constituency identifier. It should be used to match the dataset with the shapefile of electoral constituencies.
2. morani_bloc: variable that indicates which (if any) bloc the constituency belongs to. The variable takes the value 99 for "non-bloc" constituencies.
3. votes_[party/candidate]_[year]: share of votes for [party/candidate] in [year] in the constituency
4. mi_[year]: variable that takes the value 1 if the constituency belongs to the Moran's I cluster in [year] N.B.: A constituency is considered part of a Moran's I bloc if it is part of the Moran's I cluster or a neighbor of the cluster (see Appendix I for definitions) in 75% of the elections considered. Therefore, a constituency can be part of the cluster in a given year, and not be part of the bloc, or the reverse.
5. winner_name_[year]: Party (or candidate) that had the largest share of votes in the constituency in [year].