

CODEBOOK

“Right-Peopling” the State:
Nationalism, Historical Legacies, and Ethnic
Cleansing in Europe, 1886-2020

Carl Müller-Crepon* Guy Schvitz[†] Lars-Erik Cederman[‡]

December 14, 2023

Contents

1	Overview	1
2	Group-segments by year	1
3	HEG Meta Data	2
4	HEG Comparison to Austria-Hungary Census	3
5	Comparison among HEG maps	3
6	Aggregate measures of state to nation alignment	4
7	List of Ethnic Cleansing Cases	5
8	Other data file	5
9	Main data sources	6

*Department of Government, London School of Economics and Political Science. Corresponding author: c.a.muller-crepon@lse.ac.uk

[†]EU Joint Research Center

[‡]Center for Comparative and International Studies, ETH Zürich

1 Overview

The following describes the main data files contained in this replication repository. All data files are stored in the subdirectory `data/` using the filename indicated below.

2 Group-segments by year

File: `country_grp_year_analysis.rds` and `country_grp_year_analysis.csv`

Variables:

- `gwcode`: Gleditsch and Ward code, derived from Cshapes 2.0
- `country`: Country name, from Cshapes 2.0
- `group`: Group name, from HEG data.
- `year`: Year of observation
- `is.capitalrulegrp`: Dummy (0/1) variable for whether group segment is a “dominant” group in the respective country.
- `ethcl_int.onset.do`: Dummy (0/1) variable for an onset of ethnic cleansing.
- `has.tek`: Dummy (0/1) variable for ethnic segments that have TEK.
- `tek.titular`: Dummy (0/1) variable for ethnic segments that have any TEK that dominate their state.
- `tek.nontitular`: Dummy (0/1) variable for ethnic segments that have TEK which exclusively do not dominate their state.
- `tek.status`: Highest “status” of TEK connection of a group segment.
- `prec.homerule.yrsavg`: Average years of past home rule among the population of a segment.
- `terrclaim`: Number of territorial claims towards the host state of a group segment.
- `tek.5col.claim`: Dummy (0/1) variable for whether there is any territorial claim from a state with a TEK connection to an ethnic group segment.
- `WarNum`: Number of interstate wars affecting the host state of a group segment.

- `tek.5col`: Dummy (0/1) variable for whether there is any war with a state with a TEK connection to an ethnic group segment.
- `v2x_polyarchy`: VDEM Polyarchy index for the country-year.
- `capitaldistance`: Average distance of the population of a group segment to its state's capital.
- `grpssize`: Size of the group segment. From HYDE.
- `grpshare`: Population share of the group segment. From HYDE.
- `tek.pop`: Size of the TEK population of the group segment. From HYDE.
- `countrysize`: Population size of the country. From HYDE.
- `altitude`: Average altitude. From FAO GAEZ.
- `slope`: Average terrain slope. From FAO GAEZ.
- `evapotransp`: Average evapotranspiration index. From FAO GAEZ.
- `precipitation`: Average precipitation. From FAO GAEZ.
- `ppetratio`: Average ratio of precipitation over evapotranspiration. From FAO GAEZ.
- `temperature`: Average temperature. From FAO GAEZ.
- `dispersion`: Population dispersion of group segment.
- `borderdistance`: Average distance of segment's population to the closest border of its state.
- `state.frac`: Ethnic fractionalization of the country.
- `tek.frac`: Fractionalization of wider TEK group across state borders.
- `loc.rulegrpavg`: Segregation measure of group with dominant group in country: Share of local population that belongs to the dominant group for the average member of the ethnic segment.
- `loc.rulegrpexp`: Segregation measure of group with dominant group in country: Share of local population that belongs to the ethnic segment for the average member of the dominant group in a country.

3 HEG Meta Data

File: `heg_meta.rds`

Variables:

- `file_id`: ID for the file
- `map_id`: ID for the map
- `file_name`: Name of the file
- `library_id`: ID of the library of origin
- `year_data_end`: Latest date given for the information displayed on a map
- `title`: Title of the map
- `author`: Author of the map
- `scale`: Scale of the map
- `link_map_image`: Link to digital image if available online.

4 HEG Comparison to Austria-Hungary Census

File: `census_AH_comparison.rds`

Variables:

- `Dstrc_C`: District ID
- `Land_Cd`: Land (region) ID
- `Dstrc_N`: District name
- `Land_Nm`: Land (region) name
- `Year`: Census year
- `census_name`: Name of group, concatenated from census
- `census_prop`: Population share of group in census
- `heg_name`: Name of group, concatenated from HEG data
- `heg_prop`: Population share of group in HEG data

5 Comparison among HEG maps

File: `compare_heg_df.rds`

Variables:

- x: Name of comparison map X.
- y: Name of comparison map Y.
- x.year: Year of map X.
- y.year: Year of map Y.
- x.in.y.terr: Territory of group as mapped in map X also mapped as such in map Y. NA = missing data.
- x.in.y.pop: Population of group as mapped in map X also mapped as such in map Y. NA = missing data.
- group: Group for which comparison is carried out.
- x.terr.tot: Total territory of group in map X.
- x.pop.tot: Total population of group in map X.
- y.terr.tot: Total territory of group in map Y.
- y.pop.tot: Total population of group in map Y.
- cov.perc.x: Share of map X covered by map Y.
- cov.perc.y: Share of map Y covered by map X.

6 Aggregate measures of state to nation alignment

Files: `sys_homogeneity.rds` and `sys_homogeneity_tv.rds`, the second containing aggregate measures derived from the temporally smoothed HEG data. The variables in both datasets mirror each other.

Variables:

- name: Name of unit. Can be “all” where measure is system-wide, or a state (for ethnic fractionalization), or an ethnic group (for territorial fractionalization).
- type: Type of measure of state to national alignment: Either normalized mutual information, ethnic fractionalization, or territorial fractionalization.
- value: Value of measure

- `weight`: Population-weight of unit among all units for that year and measure.
- `cs.year`: Year of state borders
- `heg.year`: Year of HEG data

7 List of Ethnic Cleansing Cases

File: `ethnic_cleansing_list.csv`

Variables:

- `group`: Ethnic group affected
- `change_start`: Start of ethnic cleansing campaign
- `change_end`: End of ethnic cleansing campaign (if not same year or unknown)
- `perpetrator_type`: Type of perpetrator: 1 host governments; 2 non-state actor; 3 foreign government
- `perpetrator_id`: Gleditsch and Ward country-codes for perpetrators
- `description`: Short description of event
- `pop_est`: Estimate of victimized population
- `countries`: Countries affected by the ethnic cleansing campaign.
- `countries_event`: Countries in which ethnic cleansing campaign happened.
- `ambiguity`: Flag for ambiguous coding.
- `ambiguity_reason`: Description of ambiguity.

8 Other data file

HEG point data: `heg_map_points.rds` contains geographic points with colors denoting each group, for years 1890 and 2020.

HEG raster data of Poles: `poles_1918.tif` and `poles_1951.tif`

HEG shapefile data of Poles: `poles.shapes.rds`

HEG spatial extent: `heg_extent_main.geojson`

Fearon's List: `Fearon_List.dta`; `fearonlist.csv` contains the match to the language tree, via which Fearon's List is linked to the HEG data.

9 Main data sources

The following describes the sources of the datasets used in the analysis.

HEG data: Data on Historical Ethnic Geography. They are available for download [here](#).

Cshapes 2.0: Data on state borders since 1886 from [Schvitz et al. \(2021\)](#). They are available [here](#).

FAO GAEZ environmental data: Data on environmental variables (for variable details, see the codebook above) used in the article are available in the GAEZ data portal [here](#) (FAO 2015).

HYDE database on historical population data: From [Goldewijk, Beusen and Janssen \(2010\)](#), Version 3.2, available for download [here](#).

References

- FAO. 2015. "Global Agro-Ecological Zones: Crop Suitability Index." *Dataset, available online at: <http://gaez.fao.org>*.
- Goldewijk, Kees Klein, 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* 2010(1):1–9.
- Schvitz, G, S Rüegger, L Girardin, L-E Cederman, N Weidmann and KS Gleditsch. 2021. "Mapping The International System, 1886-2017: The Cshapes 2.0 Dataset." *Journal of Conflict Resolution*.
- URL:** https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3610680