



GeoEPR

Polygons describing ethnic groups

Geo-referencing Ethnic Power Relations (GeoEPR)

Codebook

Version 2021

Guy Schvitz

GeoEPR 2021 codes the settlement patterns of politically relevant ethnic groups in independent states from 1946 to 2021 based on the group list in the Ethnic Power Relations dataset version 2021 (Vogt et al., 2014; see <http://www.icr.ethz.ch/data>).

1 Definitions and general coding procedure

GeoEPR starts with a list of politically relevant groups from the EPR Core 2021 dataset and codes their settlement type and settlement areas over time within each country. Ethnic groups are assigned one of six different settlement types:

1. **Regionally based:** group members reside in regions that are easily distinguishable on a map. In line with the Minorities at Risk dataset, we use the following definition to define a group's regional base: A spatially contiguous region larger than an urban area that is part of the country, in which the predominant share of group members reside. Note that the settlement areas of different groups can overlap.
2. **Urban:** at least 60% of the group is concentrated in cities and not in particular regions.
3. **Regional and urban:** Group members are both concentrated in cities and in a particular region / in particular regions.
4. **Migrant:** groups with a frequent location change, such as nomadic groups or some Roma groups.

5. **Dispersed:** groups without a clear territorial basis, i.e. they are not concentrated in particular regions or cities but instead are scattered across the country.
6. **Statewide:** some groups have a strong presence in virtually every part of the country. They do not need to constitute a majority of the population everywhere, and small pockets of territory without a significant presence (and uninhabited areas) can be ignored. This settlement pattern is typical for titular nations in Europe such as the Italians in Italy, the Bulgarians in Bulgaria, or the Hungarians in Hungary.
7. **Aggregate:** a particular group which during a given period is aggregated from several smaller ones (provided that their ethnic geography remains constant). This corresponds to the EPR dataset's coding of "umbrella" groups, which can become the main politically relevant identity for multiple subgroups in some periods (e.g. Blacks in South-Africa during Apartheid).

Only when groups have a distinct settlement pattern, that is, when they are either coded as (1) regionally based, (3) regional and urban, (6) statewide or (7) aggregate, we assign them a polygon that represents their settlement area.

Statewide groups get their country's polygon from the CShapes 2.0 dataset. For all other groups, we rely on the most adequate map material available. In a many cases, the GREG/Atlas Narodov Mira Dataset (Weidmann et al, 2009) was the appropriate source, either because EPR groups corresponded to GREG groups (and their settlements had remained constant), or because various GREG groups could be treated as components of an overarching EPR group. In many other cases, it was necessary to consult external sources, such as Gordon (2005) and Levinson (1998).

In contrast to previous datasets on ethnic geography, GeoEPR 2018 is dynamic, in that it captures major changes in ethnic settlement patterns over time. This includes changes in international borders, changes due to mass migration and/or ethnic cleansing, as well as changes in the composition of particular ethnic groups.

The split of Yugoslavia into Slovenia, Macedonia, Croatia, Bosnia, Montenegro, and Kosovo is an example for changing state borders. We follow the CShapes 2.0 data set (Weidmann, Kuse & Gleditsch, 2010) for changes in the international system.

The expulsion of Georgians from South Ossetia following the Russian invasion in 2008 is an example for changing settlement patterns of an ethnic group within a given country.

Finally, politically relevant ethnic groups can split into sub-groups and merge into an umbrella group. Blacks in South Africa that split into the constituent language groups after the end of Apartheid in 1994 are an example for a break-up. The Lari-Bakongo in Congo-Brazzaville that merged from two sub-groups - Lari and Bakongo - in 1969 are an example for aggregation.

2 Changelog

2.1 Changes from GeoEPR-2019 to GeoEPR-2021

GeoEPR 2021 builds on the previous versions by Schvitz and Müller-Crepon (2019), Vogt et al (2014) and Wucherpfennig et al (2011). The following changes were made in this update:

- Added geometries for 16 new EPR groups
- Coded post-2018 settlement area changes for 2 EPR groups: Muslim Arakanese in Myanmar (2018) and Armenians in Azerbaijan (2021)
- Updated settlement type coding for 9 EPR groups
- Fixed error concerning umbrella groups (some were incorrectly assigned a statewide polygon). All umbrella group polygons now consist only of sub-group polygons
- Fixed minor issues in previous version for 36 EPR groups: For some groups, we removed areas that did not belong to them and removed tiny “sliver” polygons resulting from changes in state borders. In a few cases, we updated the previous coding with data from more adequate maps

2.2 Changes from GeoEPR 2014 to GeoEPR 2019

GeoEPR 2019 builds directly on the previous versions by Vogt et al (2014) and Wucherpfennig et al (2011). The following changes were made in this version:

- As in the new EPR Core dataset, GeoEPR 2018 drops the minimum population threshold of 500'000 and instead extends the coding to all independent states since 1946 according to Gleditsch and Ward (1999). As a result, GeoEPR 2018 includes new ethnic settlement data for 8 countries: Bahamas, Belize, Brunei, Comoros, Equatorial Guinea, Surinam, Tibet, Zanzibar.

- All coding errors that were reported by users of GeoEPR 2.0 were fixed in the new version. This includes an erroneous polygon for the group ‘Mizrahim (Jewish)’ in Israel, an incomplete polygon for the ‘Mayas’ in Mexico, and several other coding errors.
- Ethnic settlement polygons are now adjusted to country borders from the updated Cshapes 2.0 dataset, which differs from the original Cshapes data in a few instances (e.g. the border between Chad and Libya since 1946 or between Russia and Ukraine after 2014).
- GeoEPR 2018 uses a slightly different definition of the settlement type ‘Regional and Urban’ than previous versions.