

Ethnic Power Relations (EPR) Dataset Family



EPR-ED

Ethno-linguistic, -religious, and -racial cleavages

EPR Ethnic Dimensions Data

Codebook

Version 2021

Nils-Christian Bormann

University of Essex, Department of Government, UK. Email: n.bormann@essex.ac.uk

Contents

1 Purpose	2
2 Concepts and Definitions	2
2.1 Language	2
2.2 Religion	3
2.3 Race	3
3 Operationalization	3
3.1 Language	3
3.2 Religion	4
3.3 Race	4
4 Coding Procedure and Reliability	6

1 Purpose

The Ethnic Power Relations - Ethnic Dimensions (EPR-ED) dataset provides information on the linguistic, religious, and racial cleavages that characterize and internally divide the ethnic groups in the Ethnic Power Relations (EPR) Core Dataset 2019. For every EPR group the EPR-ED data codes up to three of the largest sub-segments and their relative sizes on each of the three ethnic dimension. To exemplify the data structure consider the Muslims in India. While EPR identifies them as one politically relevant ethnic group, the EPR-ED dataset identifies three major linguistic sub-segments – namely Urdu, Bengali, and Malayalam – two religious dimensions – Sunni and Shi’a Muslims – as well as a South Asian “racial” identity.¹

2 Concepts and Definitions

The EPR-ED data codes the “contents of ethnicity” for all groups in the EPR Core Dataset. Specifically, it focuses on the three major ethnic dimensions of language, religion, and race. Generally, EPR-ED only considers sub-group heterogeneity on the group-level but not on the individual-level. In other words, the data neither considers multilingualism nor syncretism. This section provides more detail on these concepts.

2.1 Language

The EPR-ED dataset adopts mutual intelligibility as the general criterion for identifying different languages. More specifically, EPR-ED data follow the definitions of the Ethnologue database of languages (version 16) (?). For a more detailed discussion of different practices used to define a language and the ISO 639-3 criterion refer to <http://www.ethnologue.com/about/problem-language-identification> and the paragraphs “Language and dialect” as well as “ISO-639-3 criteria for language identification.”

A number of languages are likely to be mutually intelligible but are separated by distinct alphabets, e.g., Serbian and Croatian, or other characteristics such as pronunciation that goes beyond dialectical differences, e.g., High and Swiss German, so that Ethnologue considers them to be distinct languages. In cases of multilingualism, EPR-ED codes those languages that are most likely to be spoken in private/at home.

¹The next section clarifies our constructivist understanding of race.

2.2 Religion

EPR-ED defines religions as cultural systems which reference transcendental concerns in their attempt to provide a general order of life. As the explicit negation of the transcendental, the data explicitly include the absence of religious belief – atheism – as one sub-segment of the religious dimension. In differentiating distinct religions, EPR-ED considers religious confession to be more important than dogmatic practice because the latter is much shaped by local customs. Catholics in Cameroon, for example, differ significantly in their form of worship from Catholics in Italy.

2.3 Race

Instead of resorting to assumptions about differences in the physical appearance of individual ethnic group members, the EPR-ED dataset defines race as ethnic groups' origins from particular world regions, such as Europe, Sub-Saharan Africa, Oceania, etc. These regional origins – expressed at the individual level through certain phenotypical markers (particularly skin color) – have become relevant as social categories in the context of European colonization of the world and the related process of racial classification (?, p.5-19).

3 Operationalization

3.1 Language

The EPR-ED considers all languages identified by the Ethnologue project (?) and computes the relative language weights from the absolute size estimates provided by Ethnologue.

Name:	<i>language1-3</i>
Type:	Text
Value range:	NA
Description:	Identifies the three largest linguistic segments spoken by group members in descending order. Where groups speak only one or two languages the remaining fields are coded as missing.
Coding details:	Three-letter code that identifies Ethnologue languages as defined by ISO 639-3.

Name: *lang1-3_size*
Type: Real
Value range: (0,1]
Description: Identifies the relative size of each of the three linguistic sub-segments.
Coding details: In linguistically very fragmented groups, the sum of all linguistic segments need not add up to 1.

3.2 Religion

In contrast to the linguistic dimensions, EPR-ED does not rely on an external list of religions. Instead, the data come with their own list of major religions and various sub-categories that was created based on secondary literature from sociology, anthropology, and political science. Table 1 lists all segments and their respective codes.

Information on the relative size codings of religious segments derive from the list of “people’s groups and their religious make-up” assembled by the *Joshua Project* (see ?). The Joshua Project provides information on the religious identities of all Ethnologue language groups and the absolute size of these segments.

Name: *religion1-3*
Type: Text
Value range: NA
Description: Identifies the three largest religious segments spoken by group members in descending order. Where groups adhere to one or two religions the remaining fields are coded as missing.
Coding details: Three- to six-letter code that identifies religions, denominations, and major religious traditions.

Name: *rel1-3_size*
Type: Real
Value range: (0,1]
Description: Identifies the relative size of each of the three religious sub-segments.
Coding details: In religiously very fragmented groups, the sum of all religious segments need not add up to 1.

3.3 Race

Similar to the religious dimension, EPR-ED offers its own classification of “racial” categories that capture major world region. Again, these regions were identified through a close reading of secondary sources from history, sociology, and political science. Table 2 lists the seven racial markers included in EPR-ED and their respective id codes.

Table 1: EPR-ED Classification of Religious Segments

ID	Religion	ID	Religion
ER	EASTERN RELIGIONS	ARC	Christianity
ERH	Hinduism	ARC1	Roman Catholic
ERH1	Vaishnavism	ARC2	Eastern Orthodox
ERH2	Shaivism	ARC2A	Greek Orthodox
ERH3	Shaktism	ARC2B	Romanian Orthodox
ERH4	Smartism	ARC2C	Moldovan Orthodox
ERH5	Vedism	ARC2D	Cypriot Orthodox
ERH5	Vedism	ARC2E	Georgian Orthodox
		ARC2F	Russian Orthodox
ERB	Buddhism	ARC2G	Belarussian Orthodox
ERB1	Theravada	ARC2H	Serbian Orthodox
ERB2	Mahayana	ARC2I	Montenegrine Orthodox
ERB2A	Nichiren Buddhism	ARC2J	Macedonian Orthodox
ERB3	Vajrayana/Tantrayana	ARC2K	Bulgarian Orthodox
ERB4	Korean Zen (Son) Buddhism	ARC2L	Ukrainian Orthodox
ERB5	Zen Buddhism	ARC3	Oriental Orthodox
ERB6	Pure Land Buddhism	ARC3A	Coptic Orthodox
ERB7	Tibetan Buddhism	ARC3B	Ethiopian Orthodox
ERB99	Other Buddhists	ARC3C	Eritrean Orthodox
		ARC3D	Syriac Orthodox
ERS	Sikhism	ARC3E	Malankara Orthodox
		ARC3F	Armenian Apostolic
EAR	EAST ASIAN RELIGIONS	ARC4	Protestantism
EARC	Chinese Religions	ARC4A	Arminian
EARC1	Shenism	ARC4B	Baptism
EARC1A	Taoism	ARC4C	Calvinism
EARC1A1	Southern Taoism	ARC4D	Congregational
EARC1A2	Northern Taoism	ARC4E	Lutheranism
EARC2	Confucianism	ARC4F	Methodism
		ARC4G	Presbyterianism
	OTHER	ARC4H	Reformed
ANI	Animism	ARC4I	Seventh-Day Adventist
ANI1	African Traditional Religions	ARC5	Anglican
ANI1A	Akan	ARC99	Other Christians
ANI1B	Odinani		
ANI1C	Serer	ARI	Islam
ANI199	Other	ARI1	Sunni Islam
ANI2	Shamanism	ARI1A	Hanafi
ZOR	Zoroastrianism	ARI1B	Shafii
ATH	Atheism	ARI1C	Maliki
		ARI1D	Hanbali
AR	ABRAHAMIC RELIGIONS	ARI1D1	Wahhabism
ARJ	Judaism	ARI2	Shia Islam
ARJ1	Orthodox	ARI2A	<i>Twelver Shia (Imami)</i>
ARJ1A	Ultraorthodox	ARI2B	<i>Sevener Shia (Ismaili)</i>
ARJ1B	Modern Orthodox	ARI2B1	Musta'lis
ARJ1C	Open Orthodox	ARI2B2	Nizaris
ARJ1D	Conservadox	ARI2B3	Alawites
ARJ2	Conservative	ARI2B4	Alevis
ARJ3	Reconstructionism	ARI2B5	Druze
ARJ4	Reform	ARI2B6	Yezidis
ARJ5	Renewal	ARI2B7	Bohras
ARJ6	Humanist	ARI2C	<i>Fiver Shia (Zaydi)</i>
ARJ7	Kabbalah	ARI2D	Other Shia
ARJ99	Other Jews	ARI3	Ibadhi Islam
		ARI4	Sufi Islam
		ARI5	Ahmadiyya
		ARI99	Other Islam

Note that, unlike the linguistic and religious dimension, the racial categories identify multiple markers of origin on the individual level. This is necessary because the data’s constructivist understanding of race explicitly applies to such social categories as Mestizos in many Latin American states. Where the data identify more than one racial segment, every user of the data is advised to treat these segments as one combined racial dimension. As a result, the data do not offer information on the size of the various racial segments.

Information mainly come from secondary sources and case studies of individual countries as well as other datasets on ethnicity such as the Minorities at Risk (MAR) project by (?).

Name:	<i>phenotype1-3</i>
Type:	Text
Value range:	NA
Description:	Identifies the socially constructed racial marker of origin.
Coding details:	Three-letter code that identifies various world regions of origins as racial markers.

Table 2: EPR-ED Classification of Racial Segments

ID	Racial Origin
ame	Americas
eas	East Asia
eur	Europe
men	Middle East & Northern Africa
oce	Oceania
sas	South Asia
ssa	Sub-Saharan Africa

4 Coding Procedure and Reliability

In line with previous codings of the EPR data, the Ethnic Dimensions data were collected by a mix of country and region experts, graduate research assistants, and the EPR project coordinators at ETH Zurich. After research assistants provided an original coding of the ethnic dimensions of the EPR groups, the EPR project coordinators double-checked these codings, and relied on the judgment of country and region experts in controversial cases. Moreover, a subset of cases (approximately 10%) was coded twice. Within this set of cases there was a high consistency in coding choices.