

The RAI_ROKKAN dataset on regional distinctiveness

Version November 22, 2020
(contact: hooghe@unc.edu)

This document introduces a new dataset on regional distinctiveness, which we call in memory of Stein Rokkan, the RAI_ROKKAN dataset. The data are available on our websites: <https://garymarks.web.unc.edu> and <https://hooghe.web.unc.edu>.

Contents

Introduction: what it covers, how to reference, where to find it?	2
1. Difference	3
Language	3
Religion	4
2. Dependence	6
Former state core	6
Early overarching governance	6
3. Distance	7
Travel distance	7
Travel time	8
Island-region	8
Table A.1: Correlation matrix of Rokkan characteristics of regions	9
Table A.2: Regions and their distinctiveness by country	10
4. Codebook for excel file	13
References	14

Introduction: what it covers, how to reference, where to find it?

The **ROKKAN DATASET** estimates normative, political, and geographical distinctiveness of 1766 regions in 95 countries that have a regional government as of 2018.¹ The unit of analysis is the territorial unit with the most authoritative general purpose government between the local and the national level in a country. Each region is assessed on how distinctive it is from the national state's core with respect to language, religion, political history, and geography (Hooghe and Marks 2016).

The dataset operationalizes Stein Rokkan's concepts of **difference**, **dependence**, and **distance** as indications of a territory's peripherality (Rokkan and Urwin 1983) on the hypothesis that these structural characteristics that might facilitate a demand for regional authority.

FOR USERS

How to reference the data?

Liesbet Hooghe and Gary Marks. 2016. *Community, Scale, and Regional Governance*. Oxford University Press.

&

Sarah Shair-Rosenfield, Arjan H. Schakel, Sara Niedzwiecki, Sandra Chapman-Osterkat, Liesbet Hooghe, and Gary Marks. 2021. "Language difference and regional authority." *Regional and Federal Studies*.

Where to find the data?

The data and supporting material are available online on <https://garymarks.web.unc.edu/data/regional-authority>

¹ For regions with stable jurisdictional boundaries these estimates could be extended backwards in time.

1. Difference

Difference refers to the normative distinctiveness of a community. Our coding operationalizes two sources of distinctiveness: language and religion.² In Brubaker’s words (2013: 3), “both language and religion sort people into distinct, bounded and largely self-reproducing ‘communities . . . Language, religion or both are . . . constitutive of most ethnic and national identifications, and they frequently serve as the key diacritical markers, emblems or symbols of such identifications.”

Language

A region is a **language region** if a majority of its population speaks one or more mother tongues that are different from the mother-tongue of the majority in the country. In the face of considerable debate about what constitutes a separate language, we code language difference for standard languages according to the ISO-639 classification—level 2, which encodes languages for library cataloguing and bibliographic purposes and is curated by the US Library of Congress. This level seems by and large appropriate to sift standardized languages from dialectic variants. Dialects of a standard language (e.g. mutually intelligible variants of Malay) or diglossia involving a codified and vernacular version of a common language (e.g. Mandarin Chinese and its local variants) do not meet this criterion. Some gray cases remain, such as Mandarin Chinese and Cantonese Chinese, which ISO-639/2 considers cases of diglossia because of a shared written form but many language specialists consider separate languages because they are mutually unintelligible. We follow dominant practice

² Our coding decisions are documented in detailed profiles for 95 countries for language (291pp) and religion (225pp) respectively. See <https://garymarks.web.unc.edu/data/regional-authority/>

with linguistic and area specialists. We codify a person’s so-called “mother tongue,” i.e. the first language that a person grows up with in childhood. Where data is available, reported language use is averaged over the course of the past three decades, and triangulated with information from census data, Ethnologue (Eberhard et al. 2020), Wikipedia, and scholarly sources. This is operationalized in two ways:

- A **language region L1**: a majority speaks a mother-tongue that differs from a majority in the country.
- A **language region L2**: a majority speaks a single mother-tongue that differs from the majority in the country.

Religion

Rokkan (in Flora et al. 1999: 171) regarded language as decisive because it is “a focal point of identity . . . a collective act in which everyone in a territory must share,” but other scholars emphasize the return of religion as a source of difference and conflict about authority (Brubaker 2013, 2015; Gorski 2019; Safran 2008).

A **religion region** has a majority of its population declaring to adhere to one or more religions that differ from the majority in the country. We collect separate information on whether a country has a state religion, which is defined as a religion that is formally endorsed by the state (“established church”). Any country that had a state religion in the 1990s (or introduced one after 1990) is conceived to have a state religion for our 2018 coding purposes. State atheism is treated as equivalent to a state religion. The constitutional status of a religion is usually fairly easy to assess. We base our assessment on a detailed list provided on

Wikipedia,³ which documents its categorization for most cases. For ambiguous or poorly documented cases, we complement with research that draws on primary and secondary sources (for a historical review, see Barro and McCleary 2005).

We follow established practice to distinguish between Sunni Islam and Shia Islam, and Catholicism, Protestantism, and Eastern Christianity (orthodox churches), but we do not differentiate among different forms of Sunni, Shia, Catholicism, Protestantism, or Eastern Christianity. In calculating the proportion of a population attesting a religion we include those who attest no religion.

This produces three operationalizations:

- A **religiously distinctive region R1**: the region resides in a country that has a state religion, and the population in the region attesting the state religion is a minority.
- A **religiously distinctive region R2**: the region resides in a country that has a state religion, and a majority of the population in the region attests a single religion that is not the state religion.
- A **religiously distinctive region R3**: the region resides in a country that has no state religion, and a majority of the population attests one or more religions that are different from the majority religion in the country.

R1, R2, and R3 can be combined to identify a **religiously distinctive region R5**, in which a majority of a region's population adheres to one or more religions that differ from the majority religion in the country.

³ https://simple.wikipedia.org/wiki/State_religion

2. Dependence

Dependence refers to the political reliance of the periphery on the center and the periphery's corresponding vulnerability to homogenization (Flora et al. 1999: 115). To what extent is a region able to sustain its distinctiveness against a centralizing core—a state, an empire, a colonizer? We code two plausible scenarios.

Former state core

A region is coded as a **former state core** if it meets the following criteria: (a) it was part of a prior independent state or empire for a continuous period of thirty years or more since AD1200; (b) it encompasses the core or capital of the prior state; and (c) at least half of its territory was part of the prior state. We apply a thin definition of statehood that focuses on the extent to which there is jurisdictional coherence—a single regime—controlling the legitimate means of coercion (Marks 1997a; 1997b: 24; Weber 1948). Statehood emerged in Europe from multicephalous quasi-feudal orders between the 12th-14th centuries (Strayer 1970), and given the durable effect of statehood, we code former state core from AD1200.

Early overarching governance

A region is coded as having **early overarching governance** if the region was encompassed in a durable overarching polity by the time the region was colonized by a European power (with Laos most recently in 1892), or absent colonization, by 1600AD. The criterion of durable overarching governance is met if it existed uninterruptedly for at least thirty years. An overarching polity can take on a variety of forms including that of a state or empire. Our measure has some affinity with Borcan et al.'s (2018) State History Index, which measures the

presence and duration of supra-tribal government since 3000BC. However, while Borcan et al. estimate governance for a country as a whole in half-century increments, our measure estimates the structure of governance region by region prior to European colonization (see also Wimmer and Min 2006). The converse of early overarching governance is **tribal governance** at the regional level. The coding for former state core and early overarching governance relies on *Wikipedia*, historical atlases, and *Encyclopedia Britannica*, crosschecked with scholarly sources.

3. Distance

Distance refers to the spatial transaction costs that impede interaction and which may sustain distinctiveness even in the face of a sustained national strategy of assimilation. We estimate three variables that assess these costs.

Travel distance

Travel distance is the distance in kilometers as-the-crow-flies between a region's capital and the capital of the country in which the region is located. We use a publicly available online search tool to calculate distance.⁴ Two regions in Poland and one in Spain have two capitals; distance is the average between these and the national capital. One region in Panama, and all but one region in North Macedonia do not have a formal capital, and we calculate the distance from each region's largest town or city to the national capital. We code the distance of the median region in their country for indigenous communities in Bolivia, Canada, Colombia, Costa Rica, and the United States, which are composed of multiple tribes in non-

contiguous territories across the country.

Travel time

Travel time is the travel time in minutes by road or water, calculated by using <https://maps.google.com> (taking the shortest travel time), and when google does not provide estimates, <https://www.rome2rio.com/> was used. Forty-eight regions can only be reached by air (commercial, private, or military air transport), and since calculated travel time there is incommensurate with travel by road or water, we allocate a missing value.

Island-region

Island-region takes a value of 1 if the region is 30km or more removed from any other region of its state. We measure the shortest distance between the coastal shore of the region and that of the nearest region of the state. Two neighboring regions of the same state on the same island do not meet the 30km distance criterion.

A region can have none, one, or several of the characteristics described above. Table A.3 summarizes the correlations among these properties for the regions in our sample.

⁴ <https://www.freemaptools.com/how-far-is-it-between.htm>

Table A.1: Correlation matrix of Rokkan characteristics of regions

	Language region L1	Language region L2	Religion region R1	Religion region R2	Religion region R3	Religion region R5	Former state core	Early overarching governance	Travel distance (kms)	Travel time (minutes)
Language L1	1.00									
Language L2	0.79	1.00								
Religion R1	0.26	0.23	1.00							
Religion R2	0.26	0.24	0.89	1.00						
Religion R3	0.28	0.26	0.64	0.73	1.00					
Religion R5	0.32	0.28	0.79	0.77	0.91	1.00				
Former state core	0.22	0.22	0.10	0.14	0.20	0.21	1.00			
Early overarching governance	-0.15	-0.04	0.05	0.06	0.05	0.03	0.05	1.00		
Travel distance	0.13	0.10	0.05	0.06	0.10	0.09	0.07	-0.18	1.00	
Travel time	0.23	0.14	0.04	0.06	0.12	0.13	0.07	-0.17	0.85	
Island-region	0.17	0.13	0.00	0.01	0.09	0.11	0.06	-0.05	0.20	0.28

Note: n=1749 (n=1706 for travel time).

Table A.2: Regions and their distinctiveness by country

Country	Language region L1	Language region L2	Religion region R1	Religion region R2	Religion region R3	Religion region R5	Former statehood	Early overarching governance	Median travel distance (km)	Median travel time (minutes)	Island-region
Albania	3	0	0	0	0	0	0	12	70	105	0
Argentina	0	0	0	0	0	0	13	6	925.5	720	1
Australia	0	0	0	0	0	0	6	0	908.5	1005	1
Austria	0	0	0	0	0	0	4	9	155	150	0
Bangladesh	4	0	4	2	2	4	1	62	147	350	0
Belgium	4	4	0	0	0	0	0	5	1	1	0
Bolivia	4	3	0	0	1	1	1	5	290	535	0
Bosnia & Herzegovina	1	1	0	0	1	1	1	2	62	100.5	0
Brazil	0	0	0	0	0	0	0	0	1315	1240	0
Bulgaria	2	1	0	0	1	1	0	28	178	175	0
Cambodia	0	0	0	0	0	0	0	23	148	210	0
Canada	4	4	0	0	3	3	9	0	1674	1710	0
Chile	0	0	0	0	0	0	0	8	645.5	475	0
China	14	11	2	2	0	2	6	31	1126	1063	1
Colombia	5	1	0	0	0	0	1	0	406	615	1
Costa Rica	1	1	1	1	0	1	0	0	116	91	0
Croatia	0	0	0	0	0	0	0	21	130	110	0
Cuba	0	0	0	0	0	0	0	0	290.5	240	1
Czech Republic	0	0	0	0	0	0	0	14	107	80	0
Denmark	2	2	0	0	0	0	1	7	215	180	2
Dominican Republic	0	0	0	0	0	0	0	32	103	125	0
Ecuador	1	0	0	0	0	0	0	11	172	335	1
El Salvador	0	0	0	0	0	0	0	9	54	70	0
Finland	1	1	0	0	0	0	0	19	234	180	1
France	2	2	0	0	0	0	2	13	343	250	1
Germany	1	1	0	0	0	0	6	16	275.5	210	0

Country	Language region L1	Language region L2	Religion region R1	Religion region R2	Religion region R3	Religion region R5	Former statehood	Early overarching governance	Median travel distance (km)	Median travel time (minutes)	Island-region
Greece	0	0	0	0	0	0	2	14	267	285	2
Guatemala	9	6	0	0	0	0	0	0	90.5	175	0
Haiti	0	0	0	0	0	0	0	10	120.5	210	0
Honduras	1	1	0	0	0	0	0	0	140.5	240	1
Hungary	0	0	0	0	0	0	0	20	119.5	97.5	0
India	27	20	0	0	6	6	19	30	1212	1530	2
Indonesia	25	8	0	0	4	4	10	22	950	2370	5
Ireland	0	0	0	0	0	0	0	31	110	95	0
Israel	1	1	0	0	0	0	0	6	62.5	65	0
Italy	6	6	0	0	0	0	8	21	394	310	1
Japan	0	0	0	0	0	0	0	47	390	360	0
Laos	9	0	1	1	1	1	3	18	319	525	0
Latvia	2	0	0	0	0	0	1	5	41	45	0
Lithuania	0	0	0	0	0	0	0	10	127.5	105	0
Malaysia	4	3	2	0	0	2	10	15	287.5	225	3
Mexico	4	2	0	0	0	0	3	17	481	445	0
Mongolia	0	0	0	0	0	0	0	1	1	1	0
Myanmar	12	11	3	3	3	3	4	14	352	410	0
Nepal	2	0	0	0	0	0	1	4	185	400	0
Netherlands	1	1	0	0	0	0	1	12	81	80	0
New Zealand	0	0	0	0	0	0	0	0	313.5	390	0
Nicaragua	2	0	0	0	0	0	2	0	111	140	0
North-Macedonia	2	1	0	0	1	1	0	8	81	77.5	0
Norway	0	0	0	0	0	0	0	20	277.5	220	1
Pakistan	8	7	1	1	1	1	5	6	327	335	0
Panama	5	4	4	4	0	4	0	0	150	210	0
Papua New Guinea	20	0	0	0	0	0	0	0	524	1	4

Country	Language region L1	Language region L2	Religion region R1	Religion region R2	Religion region R3	Religion region R5	Former statehood	Early overarching governance	Median travel distance (km)	Median travel time (minutes)	Island-region
Paraguay	0	0	0	0	0	0	0	0	189.5	305	0
Peru	5	4	0	0	0	0	1	21	533.5	860	0
Philippines	63	47	0	5	6	6	7	17	466	935	4
Poland	0	0	0	0	0	0	0	16	252.5	200.5	0
Portugal	0	0	0	0	0	0	0	25	219	150	2
Romania	2	2	0	0	0	0	0	42	222	240	0
Russia	8	6	9	8	8	9	13	81	1165	1010	1
Serbia	0	0	0	0	0	0	1	4	83	80	0
Slovakia	0	0	0	0	0	0	0	8	136	97.5	0
South Korea	0	0	0	0	1	1	0	17	187	110	1
Spain	7	7	0	0	0	0	5	19	339	250	4
Sri Lanka	2	2	2	1	1	2	3	5	101	200	0
Sweden	0	0	0	0	0	0	0	21	271	260	1
Switzerland	7	7	0	0	7	7	23	26	88	90	0
Taiwan	1	1	0	0	0	0	0	22	155	150	3
Thailand	25	25	4	4	4	4	2	78	394.5	360	0
Trinidad and Tobago	1	1	0	0	0	0	0	0	104	230	1
Turkey	16	14	26	26	26	26	3	81	415	390	0
Ukraine	7	7	0	0	2	2	0	25	367	340	0
United Kingdom	0	0	1	0	0	1	2	4	364.5	290	1
United States	2	2	0	0	2	2	15	0	1231	760	3
Uruguay	0	0	0	0	0	0	0	0	173	170	0
Venezuela	1	0	0	0	0	0	0	0	301	330	0
Vietnam	9	3	1	0	0	1	3	63	605	840	0
Total (*median)	345	231	61	58	81	96	198	1239	275*	300*	50

4. Codebook for excel file

worldregion	1 Asia
	2 Latin America
	3 New World
	4 Europe

countryname
countryabbr
countryid
regionname
regionid
year

ROKKAN	DIFFERENCE, DEPENDENCE, DISTANCE VARIABLES
language1 L1	a majority speaks a mother-tongue that differs from a majority in the country
language2 L2	a majority speaks a single mother-tongue that differs from the majority in the country
statereligion	country has a state religion, defined as a religion that is formally endorsed by the state
religion1 R1	the region resides in a country that has a state religion, and the population in the region attesting the state religion is a minority
religion2 R2	the region resides in a country that has a state religion, and a majority of the population in the region attests a single religion that is not the state religion
religion3 R3	the region resides in a country that has no state religion, and a majority of the population attests one or more religions that are different from the majority religion in the country
religion5 R5	a majority of a region's population adheres to one or more religions that differ from the majority religion in the country
statehood	(a) it was part of a prior independent state or empire for a continuous period of thirty years or more since AD1200; (b) it encompasses the core or capital of the prior state; and (c) at least half of its territory was part of the prior state
unilevel	tribal = the region was <u>not</u> encompassed in a durable overarching polity by the time the region was colonized by a European power (with Laos most recently in 1892), or absent colonization, by 1600AD
multilevel	the region was encompassed in a durable overarching polity by the time the region was colonized by a European power (with Laos most recently in 1892), or absent colonization, by 1600AD
travel1	the distance in kilometers as-the-crow-flies between a region's capital and the capital of the country in which the region is located
travel1log	log10 of travel1
travel2	the travel time in minutes by road or water in 2020 (google maps)
islandregion	the region is non-contiguous and 30km or more removed from any other region of its state
DEMOGRAPHICS	ANCILLARY INFORMATION
popregion	population of the region around 2018
arearegion	area of the region around 2018
popregionlog	log10 of popregion
arearegionlog	log10 of arearegion
creationunit	year in which region as a unit was created (1949 if prior to 1950)

References

- Barro, Robert J. and Rachel M. McCleary. 2005. "Which countries have state religions?" *Quarterly Journal of Economics*, 20 (4):1331-1370.
- Brubaker, Rogers. 2013. "Language, Religion and the Politics of Difference." *Nations and Nationalism* 19(1): 1-20.
- Brubaker, Rogers. 2015. *Grounds for Difference*. Cambridge, MA: Harvard University Press.
- Elazar, Daniel J. 1987. *Exploring Federalism*. Tuscaloosa: The University of Alabama Press.
- Flora, Stein Kuhnle, and Derek Urwin (eds). 1999. *State Formation, Nation-building, and Mass Politics in Europe*. Oxford: Oxford University Press.
- Gorski, Philip S. 2019. "The Politics of Religion and Language: Similar or Different?" *Social Science History* 43(2).
- Hooghe, Liesbet, and Gary Marks. 2003. "Unraveling the Central State, But How? Types of Multi-Level Governance." *American Political Science Review* 97(2): 233-43.
- Hooghe, Liesbet, and Gary Marks. 2010. "Types of Multilevel Governance." In Hendrik Enderlein, Sonja Wälti, and Michael Zürn (eds.), *Handbook on Multilevel Governance*, 17-31. Cheltenham: Edward Elgar.
- Hooghe, Liesbet, Gary Marks, and Arjan H. Schakel. 2008. "Regional Authority in 42 Democracies, 1950–2006." *Regional and Federal Studies* 18 (2-3): 111-302.
- Hooghe, Liesbet, Gary Marks, Arjan H. Schakel. 2010. *The Rise of Regional Authority: A Comparative Study of 42 Democracies (1950-2006)*. London: Routledge.
- Keating, Michael. 1998. *The New Regionalism in Western Europe. Territorial Restructuring and Political Change*. Aldershot: Edward Elgar.
- Keating, Michael. 2001. *Plurinational Democracy: Stateless Nations in a Post-Sovereignty Era*. Oxford: Oxford University Press.
- Lane, Jan-Erik, and Svante Errson. 1999. *Politics and Society in Western Europe*. London: Sage.
- Liesbet Hooghe, Gary Marks, Arjan H. Schakel, Sara Niedzwiecki, Sandra Chapman Osterkatz, and Sarah Shair-Rosenfield. 2016. *Measuring Regional Authority*. Oxford University Press.
- Marks, Gary. 1997a. "A Third Lens: Comparing European Integration and State Building." In Jytte Klausen and Louise A. Tilly, eds., *European Integration in Social and Historical Perspective: 1850 to the Present* (New York: Rowman & Littlefield), 23-50.
- Marks, Gary. 1997b. *An Initial Attempt to Explain Reallocation of Authority Among Regimes, with Reference to Diffusion of Authority in European Integration and the Disintegration of the Carolingian Empire*, APSA meeting, Washington DC, August 1997.
- Niedzwiecki, Sara, Sandra Chapman-Osterkatz, Gary Marks, and Liesbet Hooghe. 2018. "The RAI Travels to Latin America: Measuring Regional Authority Under Regime Change." *Regional & Federal Studies*, DOI: 10.1080/13597566.2018.1489248, online 6 July 2018.
- Riker, William H. 1964. *Federalism: origin, operation, significance*. Boston: Little, Brown.

- Rokkan, Stein, and Derek Urwin. 1983. *Economy, Territory, Identity: Politics of West European Peripheries*. London: Sage.
- Safran, William. 2008. "Language, Ethnicity and Religion: A Complex and Persistent Linkage." *Nations and Nationalism* 14(1): 171-190.
- Schakel, Arjan H. 2020.
- Shair-Rosenfield, Sarah, Gary Marks, and Liesbet Hooghe. 2014. "A Comparative Measure of Decentralization for Southeast Asia." *Journal of East Asian Studies* 14(1): 85-107.
- Strayer, Joseph. 1970. *On the Medieval Origins of the Modern State*. Princeton: Princeton University Press.
- Weber, Max. 1968. *Economy and Society*. Edited by Guenther Roth and Claus Wittich. Bedminster: New York.
- Wimmer, Andreas, and Brian Min. 2006. "From Empire to Nation-State. Explaining Wars in the Modern World, 1816-2001." *American Sociological Review* 71 (6): 867-897.