Small States & Territories, Vol. 5, No. 1, 2022, pp. 159-178

A new classification of small island economies based on geography, demography and sovereignty

Michaël Goujon
Université Clermont Auvergne
CNRS, CERDI
France
michael.goujon@uca.fr

and

Justinien Razafindravaosolonirina
Université de La Réunion
Réunion
France
romain.razafindravaosolonirina@univ-reunion.fr

ABSTRACT: We explore and use correlations (not causations) between geographic and demographic characteristics and current levels of sovereignty in order to propose a new classification of small, island and coastal territories. While previous analyses mostly rely on descriptive statistics between the group of UN-members and subnational jurisdictions, we take advantage of a "formal sovereignty" index developed by Alberti and Goujon (2020) that provides a continuous and multidimensional measure of sovereignty or autonomy for a sample of 100 small island states and coastal/island territories. Huge heterogeneity within such a sample leads us to use a data-driven method of principal component analysis and clustering in order to secure a multidimensional typology of small islands relative to their main geographic and demographic characteristics and their level of sovereignty. The PCA results show that heterogeneity is firstly explained by a combination of geographic and demographic variables, and secondly by sovereignty, associated (positively) with population size and (negatively) with insularity. The clustering analysis leads to divide the 100 territories into four clusters mainly characterized by, respectively: Group 1 (32 territories): high sovereignty associated with a large population; Group 2 (26 territories): high values of latitude and life expectancy (mostly Atlantic and Baltic territories); Group 3 (40 territories): large distance to metropolitan power and high insularity (Pacific Regions); and Group 4: Greenland and Nunavut, two territories with a large land area, high latitude, low populations and large EEZ surface area.

Keywords: clustering, formal sovereignty index, principal component analysis, small island states, small island territories, sovereignty

---

Introduction

Research on the political settings, geographic specificities, and economic performances of small islands economies (SIEs) remains vibrant (e.g. Armstrong and Read, 2021). Particularly, the role of sovereignty or autonomy in public management and performances are subject of much discussion, and so are the determinants of the current levels of sovereignty. However, analytical complexity and huge heterogeneity observed in the group of small islands economies would hardly allow drawing clear causalities between political settings, geographic specificities and economic performances. This is why in this paper we will use correlations (not causations), to get a typology and classification of SIEs based on their main geographic and demographic characteristics augmented by their level of sovereignty.

Complexity can be illustrated when trying to understand the current level of sovereignty or autonomy of SIEs. Analytically, the level of sovereignty could be analyzed as resulting from the interaction between the (colonized territory's) demand for and the (colonizer's) supply of autonomy. However, it is complex to draw a clear separation between demand and supply factors given their expected interaction. Sambanis and Milanovic (2011) list the determinants of the demand for sovereignty, defined as an increasing function of economic, political, cultural, or historical differences between a region and a centre, describing the tradeoff between sovereignty and income that the region faces. Typically, the demand for sovereignty would be higher the larger the size of the region and/or the greater the regional distinctiveness, which can be correlated with distance from the centre (Baldacchino, 2020). Moreover, rich regions, including those less populated, would tend to ask for more local power, while poorer territories would tend to balance a lack of sovereignty by transfers from the centre. In contrast, inter and intra-inequalities, history, or political institutions would have more ambiguous effects (Sambanis and Milanovic, 2011).

Regarding the supply side, Gerring et al. (2018) advocate the role of population size for concentration/dispersion of power in a polity, since more populous territories may have higher incentives to delegate 'down' to local authorities. Delegation would allow adapting the central rule to the local socio-economic environment, in search of greater efficiency in the polity's management and increasing citizens' trust. The economic theory of secessions (e.g. Madies et al., 2018) highlights a key trade-off between (pro-secession) increasing returns to scale in the provision of public goods and (anti-secession) cultural and preference heterogeneity across individuals living in different regions. Then, large (and/or rich) regions would display more secessionist tendencies than small (poor) ones, and so would culturally-distant regions. Rezvani (2014) also demonstrates the role of the relative economic size of the region to the core state in explaining the political dynamics of partially independent territories.

Following this general setting, small islands that are affiliated to a larger country, but are both geographically and culturally distant from the rest of the country, may display significant secessionist tendencies. This should push a central government to grant autonomy without a complete loss of decision-making power. Watts (2009) considers the particular meaning of island sovereignty, because islandness, remoteness, and uniqueness generate pressures for autonomy which are eventually balanced by the advantages of remaining affiliated. Baldacchino (2020), among others, explains that physical distance to the metropolitan power would increase the desire – and the likelihood – by the territory's population for emancipation, the ability to choose its own destiny, and its own development model. The metropolitan power may also desire to grant more or less sovereignty following a cost-benefit calculation. Exclusive Economic Zones (EEZs) then have a specific resonance for small islands (even for uninhabited ones) and their metropolitan power, as they often involve international disputes over sovereignty (e.g. Baldacchino, 2015). The case of France, which possesses the second-largest EEZ in the world thanks to its overseas territories, is a typical illustration (Baldacchino and Hepburn, 2012).

Ferdinand et al. (2020) aim at explaining why a number of small island territories that have been colonized remain non-sovereign (the so-called case of sub-national island jurisdictions: SNIJs[1]) and the local levels of (dis)content, if any, with this political status. They underline the importance of geography (location, climate, size, insularity), demography

[1] "[T]hose non-sovereign territories that qualify as former colonies and whose present post-colonial status implies the possibility of opting out" (Ferdinand et al., 2020, p. 48).

(population size), history (colonialism), economy (standards of living, including ecology) and constitutional and political arrangements (autonomy). They restrict their analysis to 40 SNIJs - thus excluding sovereign island states, and offer a well-documented descriptive analysis; but they do not attempt to draw any statistical correlations.

From these works, we can list a large number of characteristics that can be considered as potential correlates, if not causal factors, of sovereignty for SIEs, in a way that they may affect the demand for/ the supply of sovereignty, directly or indirectly, and for some in an ambiguous way. First, *geography*, such as location (latitude, longitude, ocean or sub-region, remoteness), physical characteristics (land area, perimeter and insularity), climate conditions and resource endowments… Second, *human history*, such as indigenous population, European discovery, colonization (if any) start and mode, slavery and settlement, migrations, which may have been influenced by geography (distance to the colonizers or metropolitan power); and that have resulted in modern demography (size of the population, ethnicity, social capital, local identity)… Third and last, *human development* should have multiple correlations, though with unclear causal directions, with sovereignty, governance, policy choices, structural vulnerabilities and resilience.

In large-scale comparative studies, the measure of the level of sovereignty, or autonomy, or affiliation, is roughly managed as it often reverts to the use of a simplistic binary classification based on international law between United Nations-members (considered as sovereign) versus non-UN-members (considered as affiliated). This is in contrast with an abundant literature based on SIE-case studies or lower scale comparative analyses showing the spectrum of sovereignty or autonomy (e.g. Baldacchino and Hepburn, 2012; Bonilla and Hantel, 2016; Prinsen et al., 2017; Grydehøj, 2016; Pöllath, 2018).

As an alternative to this binary classification, Alberti and Goujon (2020) propose a composite index of what they call "formal sovereignty" for the purpose of a comparative, quantitative analysis on a large sample of small island units. This index uses a simple coding of the main attributes of sovereignty, or of autonomy (diplomacy, mainly based on the UN membership, but also executive, legislative, judiciary, defence and monetary) to generate scores that are aggregated. The index is calculated for a sample of 104 small island and coastal territories[2], and reveals the continuum in the level of formal sovereignty. Moreover, the results from this new composite index show that the binary classification based on UN membership is significantly limited, if not mistaken, by revealing that some territories that are not UN members have a higher formal sovereignty or autonomy score than some UN members.

[2] Including island territories, coastal mainland territories that are Small Island Developing States (SIDS) members or observers, and for the purposes of comparison a number of coastal mainland microstates and small subnational jurisdictions with special territorial status. The 104 small island and coastal territories are: Akrotiri; Aland Islands; American Samoa; Anguilla; Antigua and Barbuda; Aruba; Azores; Bahamas; Bahrain; Barbados; Belize; Bermuda; Bonaire; Bougainville; British Indian Ocean Territory; British Virgin Islands; Cabo Verde; Canary Islands; Cayman Islands; Christmas Island; Cocos (Keeling) Islands; Comoros; Cook Islands; Corsica; Cuba; Curacao; Cyprus; Dhekelia; Dominica; Dominican Republic; Falkland Islands (Islas Malvinas); Faroe Islands; Fiji; French Guiana; French Polynesia; Gibraltar; Greenland; Grenada; Guadeloupe; Guam; Guernsey; Guinea-Bissau; Guyana; Haiti; Hawaii; Hong Kong; Iceland; Isle of Man; Jamaica; Jersey; Kiribati; Macau; Madeira; Maldives; Malta; Marshall Islands; Martinique; Mauritius; Mayotte; Micronesia, Federated States of; Monaco; Montserrat; Nauru; New Caledonia; Niue; Norfolk Island; Northern Mariana Islands; Nunavut; Palau; Papua New Guinea; Pitcairn Islands; Puerto Rico; Réunion; Saba; Saint Barthelemy; Saint Helena, Ascension, and Tristan da Cunha; Saint Kitts and Nevis; Saint Lucia; Saint Martin; Saint Pierre and Miquelon; Saint Vincent and the Grenadines; Samoa; Sao Tome and Principe; Sardinia; Seychelles; Sicily; Singapore; Sint Eustatius; Sint Maarten; Solomon Islands; Sri Lanka; Suriname; Svalbard; Taiwan; Timor-Leste; Tokelau; Tonga; Trinidad and Tobago; Turks and Caicos Islands; Tuvalu; United States Virgin Islands; Vanuatu; Wallis and Futuna; Zanzibar.

In this paper, we take advantage of this new index of "formal sovereignty" to explore the correlations between sovereignty and its potential correlates identified in the SIE literature, being geographic, demographic, or historical, for a large sample of SIEs encompassing both sovereign and non-sovereign territories, limiting selection bias. We gather data for a limited number of characteristics (nine) that allow preserving the initial geographic coverage of Alberti and Goujon (2020), only excluding 4 out of the 104 territories[3]. The 100 territories that we cover in our study significantly encompass the usual samples of large-scaled comparative studies. However, as explained by Alberti and Goujon (2020), there would be room for complementing this sample (easy candidates would include Ceuta, Melilla, etc.).

In such a dataset, composed of 9 variables covering 100 territories that are characterized by huge heterogeneity, and given analytical complexity, correlations can be complex and blurred, and causation would be hard to uncover. We then privilege Principal Component Analysis (PCA) and a Clustering approach in order to elaborate a simple but data-driven typology and a grouping of the territories in our dataset.

The rest of this paper is laid out as follows. Section 2 describes the variables using descriptive statistics and bivariate correlations, showing heterogeneity. Section 3 presents the use of PCA and data grouping, Section 4 proposes the results of the PCA. Section 5 illustrates the clustering approach and the resulting typology and grouping of SIEs. Section 6 concludes.

Descriptive statistics

In the following, out of the 104 territories covered by the index of sovereignty of Alberti and Goujon (2020), four are excluded (the uninhabited Chagos/BIOT, and Akrotiri, Christmas Islands and Norfolk Island for which life expectancy estimates are missing). The following table shows an extreme within-sample heterogeneity in most of the variables.

Table 1: Descriptive statistics (sample = 100 small islands and coastal territories)

	mean	median	min	max	standard error
Latitude (degrees)	13.6	15.1	-51.7 (FLK)	78.2 (SJM)	23.4
Absolute longitude (degrees)	79.1	63.2	2.11 (JEY)	179.2 (TUV)	53.5
Distance to metro power (kms)	7538	7001	273 (GGY)	16744 (NCL)	4654
Land_area (thousands km^2)	59	0.77	0.002 (MCO)	2165 (GRL)	299
Insularity (index)	1.13	0.63	0.015 (SUR)	8.71 (TKL)	1.67
EEZ surface (thousands km^2)	503	160	0.2 (MAC)	4767 (PYF)	793
Population (thousands)	1406	159	0.05 (PCN)	23603 (TWN)	3815
Life_expectancy (years)	77.1	77.9	57.0 (ZAZ)	89.4 (MCO)	5.5
Formal sovereignty (index)	0.57	0.58	0	1	0.34

[3] As specified in Alberti and Goujon (2020), regarding selection criteria, this sample includes island territories, mainland territories that are SIDS members or observers, and a number of mainland small states and small subnational jurisdictions with special territorial status. It excludes island jurisdictions that are states, provinces, or territories of large, mainland federations (e.g., Prince Edward Island in Canada); other island territories having formal (but non-constitutional) or otherwise informal 'special' or asymmetrical arrangements with central powers (such as Jeju Autonomous Province in South Korea); indigenous island-based communities with some sovereignty within another state (such as Torres Strait Islands in Australia); or *de facto* island powers in open confrontation or contestation with a central state (such as Turkish Republic of Northern Cyprus).

Notes: FLK Falklands; SJM Svalbard; JEY Jersey; TUV Tuvalu; GGY Guernsey; NCL New Caledonia; MCO Monaco; GRL Greenland ; SUR Suriname; TKL Tokelau; MAC Macau; PYF French Polynesia ; PCN Pitcairn; TWN Taiwan; ZAZ Zanzibar. *Source*: Authors.

Formal sovereignty

The complexity of the notion of SIEs' sovereignty is highlighted in numerous territory-case studies (Mrgudovic, 2012; Overton et al., 2012; Grydehøj, 2016; Prinsen et al., 2017; Pöllath, 2018; Veenendaal, 2020). In contrast, large-scale studies often fall back to the binary classification of UN-members, considered as sovereign, versus non-UN members, considered as affiliated (Bertram, 2004, 2015; McElroy and Sanborn, 2005; McElroy and Pearce, 2006; Feyrer and Sacerdote, 2009; Dunn, 2011). As an alternative of this binary classification, Alberti and Goujon (2020) propose a composite indicator of "formal sovereignty" based on the coding of formal (de jure) status on six dimensions, or attributes: diplomacy, executive power, judicial power, legislative power, defence and monetary. The six dimensions are simply coded and aggregated to get a composite indicator allowing to highlight the continuous spectrum of multidimensional sovereignty.

The computation is applied to 104 small island and coastal territories. Out of the 43 UN members that are covered, only 24 show full sovereignty; while 19 manifest partial sovereignty. Some of the 61 non-UN members show high level of autonomy. An important result is that some well- or less-known cases of non-UN members (Taiwan, Anguilla, Montserrat, Bermuda and Gibraltar) show higher sovereignty or autonomy scores than some UN members (Marshall Islands, the Federated States of Micronesia, Monaco and Tuvalu).

Geographic data

We collected territories' location attributes such as the longitude, latitude, the ocean and the sub-region. Latitude is a usual variable to approximate geo-climatic conditions and distance to the developed Northern hemisphere[4]. The (absolute) longitude (distance from the Greenwich prime meridian) approximates the distance from the original West European (Dutch, Portuguese, Spanish, French, German or English) colonizers, which should have influenced subsequent colonization history and current demography. Roughly, the SIEs located in the Atlantic Ocean and the Caribbean are less distant than the ones in the Asia-Pacific region.

We compute the distance to the main metropolitan power as the bilateral distance, great-circle or 'as the crow flies' distance between the territory and its main, being former (when fully sovereign) or current, metropolitan power. In the case of the territories with current full sovereignty and multiple historical colonizers, the metropolitan power is defined as the most significant (lengthy and recent) one, that usually corresponds to the ruler during the period of the wave of decolonization / independences that mostly occurred in the 20th century. To build this variable, our main source of information was Wikipedia (accessed in summer 2021) and we may have used other primary sources when cross-checking was necessary (details of the choice are available from the authors[5]).

[4] The distance to the Equator (absolute latitude) is an alternative. Results using this alternative are similar to the ones showed here and are available on request.

[5] Of course, deciding on the main ruler was not so easy for a handful of these 100 SIEs. For instance, for Samoa, one can read that "The country was occupied by the German Empire from 1899 to 1915, and by a joint British and New Zealand colonial administration until 1 January 1962, when it achieved independence", and given that New Zealand gained independence from UK only in 1947, we consider UK as the main ruler / metropolitan power.

The geographic size of the territory (that can be an archipelago) is measured by its perimeter and land area. We complement with a measure of insularity, calculated by dividing the length of the shoreline by the land area[6].

We also gather the geo-political variable of the Exclusive Economic Zones (EEZ) surface. EEZ surface is the marine area on which the territory (or its metropolitan power) exercises its sovereignty and have jurisdiction over resources. Defined as a distance no more than 200 nautical miles from the coastal lines, these areas are not bounded by the principal characteristics of the territories but, since overlaps are not rare, is rather a resultant of political negotiation on maritime boundaries. For all independent territories and for regions far from their metropolis or from other islands of the same country (Martinique, Azores, Jersey, etc.), the EEZ boundaries and surfaces are officially defined and downloadable from Flanders Marine Institute (2019). Ten affiliated SIES out of our sample of 100 does not have officially defined EEZ: Bougainville, Corsica, Hawai´i, Hong Kong, Isle of Man, Macao, Nunavut, Sardinia, Sicily, Zanzibar. We then produced our own estimates. First, when the SIE is close to its mainland and then contribute to the overall EEZ surface of the mainland country, we recalculate the EEZ of the mainland, according to the official definition, disregarding the island, as if it does not exist. The difference between the official EEZ of the mainland and the recalculated EEZ gives the estimate of the island's EEZ. Second, in the cases of affiliated but remote islands, we have estimated their EEZs following the official definition (within the 200 nautical miles limit, the boundary being at equal distance to both coasts when the distance between them is less than 400 nautical miles; the computation is done using ArcGIS software; details from the authors under request).

Demographic data

Various demographic features can be related to the formal sovereignty level. As explained in the introduction, population size is a central variable for the demand for and supply of sovereignty and data are easy to access. Other demographic characteristics related to population composition or cultural distinctiveness would be of interest but are more difficult to define and measure for a large sample of territories.

Life expectancy gives useful information on the level of development as it depends on the health system performance but more generally on living conditions in a territory. It is usually estimated and gathered in international databases[7]. From the Alberti and Goujon's sample of 104 territories, information on life expectancy is missing for only four territories (namely Akrotiri, British Indian Ocean Territory, Christmas Islands and Norfolk Island). Alternative measures of wealth, suggested in Sambanis and Milanovic (2011), Feyrer and Sacerdote (2009) or Armstrong and Read (2021), such as income, are less available (and arguably a more restricted measure of development), particularly for small non-sovereign territories.

Correlates and correlation

Figure 1 displays a basic plot of the correlations between the measure of formal sovereignty (Y-axis) and eight geographic and demographic characteristics (X-axis). Regarding geographic characteristics, the relationship with sovereignty seems (weakly) negative for latitude and insularity, and (weakly) positive for absolute longitude, distance to the metropolitan power and land area, and last, null for EEZ surface.

---

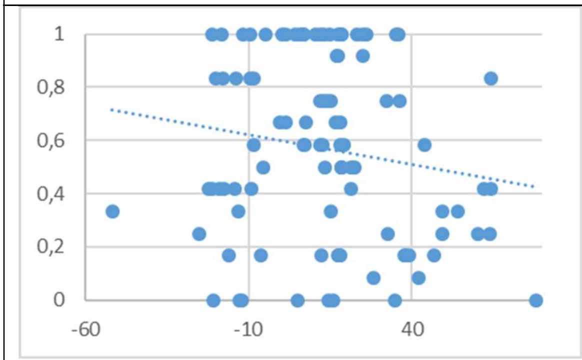
[6] Following the UNEP island indicators on the "coastal index" (http://islands.unep.ch/indicat.htm#Coastal%20Index). Taglioni (2011) discusses a wider notion of insularity by distinguishing three levels of insularity mixing geographic, economic and political contexts.

[7] Demographic data come mainly from Index Mundi or the CIA World Factbook. If missing, we complement information using national or local sources.

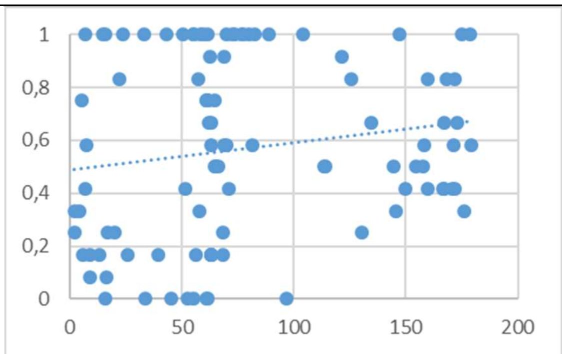
Regarding population, the (log-linearized) relationship is clearly positive: the larger the population, the higher the sovereignty. The relationship between life expectancy and sovereignty is negative, which is also expected, given the negative correlation between development and sovereignty often outlined in the literature. Broadly speaking these results are mostly expected, while the weakness of the relationships is due to a high degree of heterogeneity (R-squared is less than 0.1 for all fitted regression lines except for sovereignty-population that equals 0.24).

<u>Figure 1:</u> Correlations between formal sovereignty index and other characteristics, sample of 100 small islands and coastal territories.

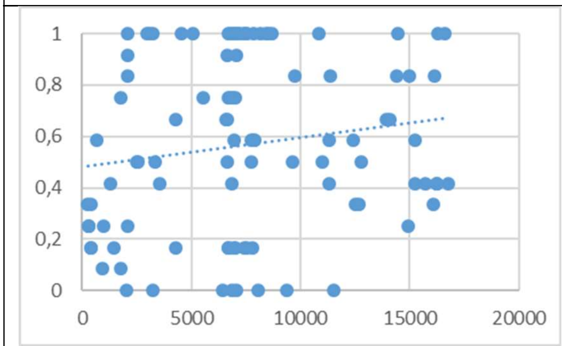
Latitude, degrees (x) vs sovereignty, index (y)



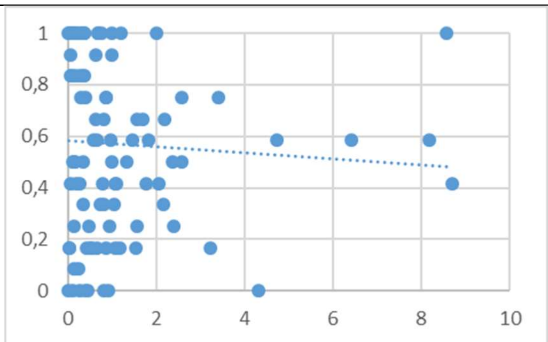
Absolute longitude, degrees (x) vs sovereignty, index (y)



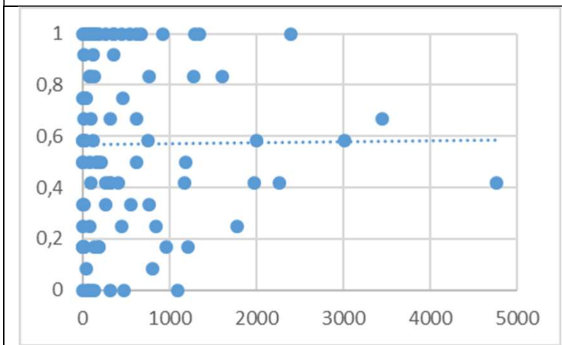
Distance to metropolitan power, kms (x) vs sovereignty, index (y)



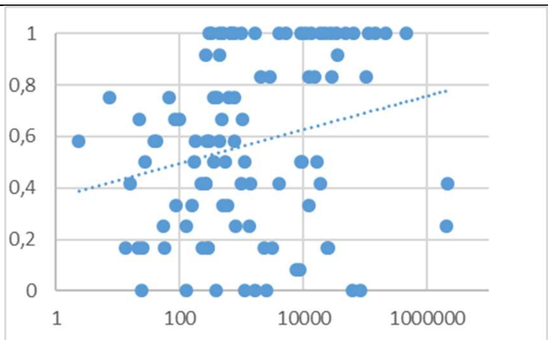
Insularity, index (x) vs sovereignty, index (y)



EEZ surface, thousands km2 (x) vs sovereignty, index (y)



Land area, km2 (x), log scale, vs sovereignty, index (y)



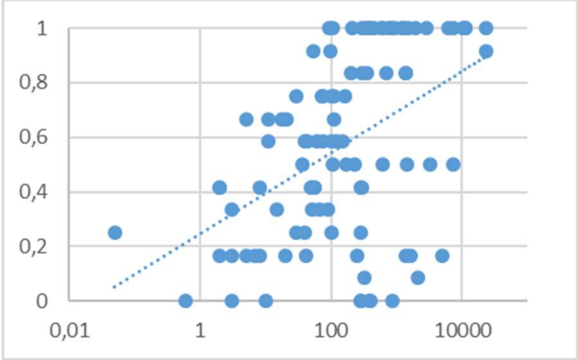
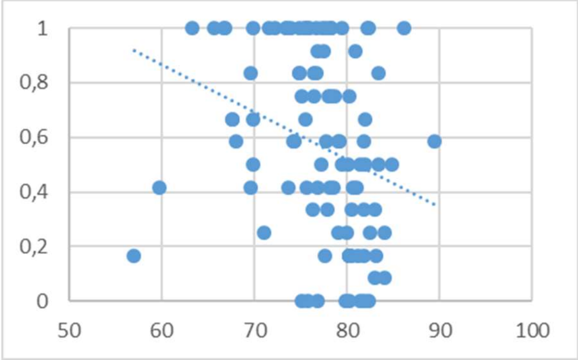
Population, thousands (x), log scale, vs sovereignty, index (y)	Life expectancy, years (x) vs sovereignty, index (y)
	

Table 2 shows that bilateral correlations between the set of sovereignty, geographic and demographic variables are often weak. However, the location variables of latitude, absolute longitude and distance to metropolitan power are highly correlated. Life expectancy is moderately correlated with these three variables of location (the higher the remoteness, the lower the life expectancy). Life expectancy is also moderately negatively correlated with sovereignty. Population size is moderately positively correlated with sovereignty but is not correlated with any other variable. Surprisingly, insularity is not significantly correlated with any other variable.

Table 2: Correlation Matrix: geography, demography and sovereignty, sample of 100 small islands and coastal territories.

	Lat.	Abs. Long	Distance metro	Land area	Insul.	EEZ surf	Pop	Life Exp.	Sover
Latitude	1	**-0.60**	**-0.79**	**0.30**	-0.09	-0.19	0.03	**0.38**	-0.16
Abs. longitude	**-0.60**	1	**0.83**	-0.04	0.21	**0.40**	0.05	**-0.32**	0.16
Distance metro power	**-0.79**	**0.83**	1	-0.13	0.20	**0.38**	-0.06	**-0.33**	0.16
Land area	**0.30**	-0.04	-0.13	1	-0.12	**0.30**	0.00	-0.15	-0.05
Insularity	-0.09	0.21	0.20	-0.12	1	0.15	-0.19	0.00	-0.06
EEZ surf	-0.19	**0.40**	**0.38**	**0.30**	0.15	1	-0.03	-0.20	0.01
Population	0.03	0.05	-0.06	0.00	-0.19	-0.03	1	0.00	**0.30**
Life_expectancy	**0.38**	**-0.32**	**-0.33**	-0.15	0.00	-0.20	0.00	1	**-0.29**
Sovereignty	-0.16	0.16	0.16	-0.05	-0.06	0.01	**0.30**	**-0.29**	1

Principal Component Analysis and data grouping

We now explore the multiple correlations between formal sovereignty, geographic and demographic variables, in order to draw some stylized interpretations. Since the sample is large and heterogeneous, and correlations can be complex and blurred, we privilege a data-driven method to reduce the number of variables or characteristics that mainly and associatively explain the sample variance. This allows us to reveal similarities or dissimilarities regarding these main features and to draw a typology of the selected SIEs. More specifically, it would allow for testing whether sovereignty is one of the main variables that, associated with some other variables, explains the sample variance and is one of the main attributes of the typology.

Principal Component Analysis (PCA) and cluster analysis are well known tools to perform such an analysis, and are used in numerous fields of research (e.g. Husson et al., 2017). Recent applications include: Bakri et al (2018), who build a typology of regional development of Indonesian provinces; Medina et al. (2020) who perform a vulnerability assessment of Sint Maarten; and Augustin and Liaw (2019) who group countries with regard to tourism competitiveness. The aim of PCA is to extract the common factors from a set of multiple variables and data, by finding a few (independent or uncorrelated) linear combinations of a limited number of (significantly) correlated variables that explain most of the sample variance. The choice of the number of linear combinations, or components, is based on the maximization of the share of the sample variance that they are able to explain. Then, based on the identified principal components, a cluster analysis allows for the grouping of territories into homogenous groups according to the values they show on the variables contributing to the principal components.

Finally, we also use categorical variables as "inactive" variables (i.e. variables not playing in the determination of the principal components axis): the location by oceans (Pacific, Atlantic or Indian) or sub-regions (Mediterranean, Baltic, Arabian, Polynesian, etc.), and the identity of the main metropolitan power. These categorical variables may be useful to develop a first insight on the stylized details by regions or by metropolitan powers and / or if the data-driven clusters meet location or metropolitan power identity.

Again, restraining the PCA analysis of the correlates of sovereignty to a few geographical variables (land area, latitude, absolute longitude, EEZ surface, distance to main metropolitan power, insularity) and demographic variables (population size, life expectancy) allows the covering of a large sample of 100 territories.

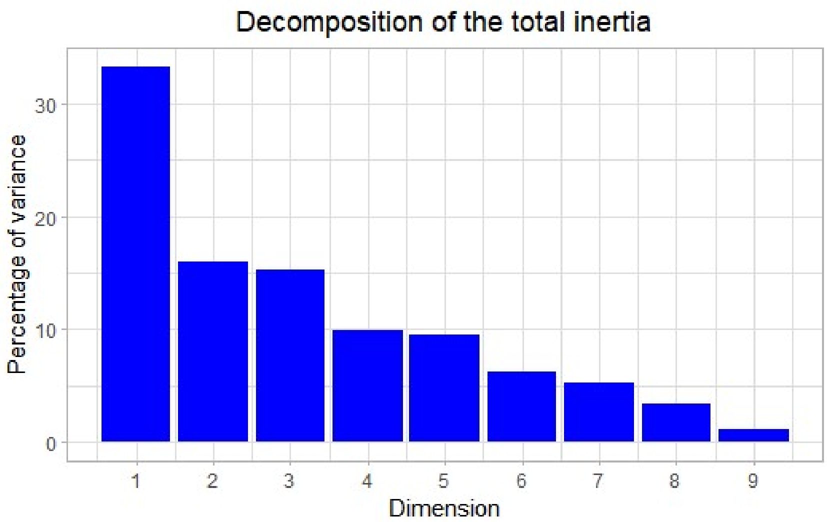
PCA results

The nine variables (Land area, Latitude, Absolute longitude, EEZ surface, Distance to main metropolitan power, Insularity, Population, Life expectancy, Sovereignty) enter into the PCA analysis.

The first important result is that the two first principal components explain approximatively half the total variance within the sample, so that the principal component space can be reasonably represented by a two-dimension space (Figure 2 and Figure 3). The contribution of each variable to the two dimensions is represented graphically by an arrow: the longer the arrow, the greater its contribution to the component (See Appendix 2 for the computed contributions). In a similar way, the closer the variables' direction, the closer the positive association between them; an opposite direction signals a negative, or opposite association.

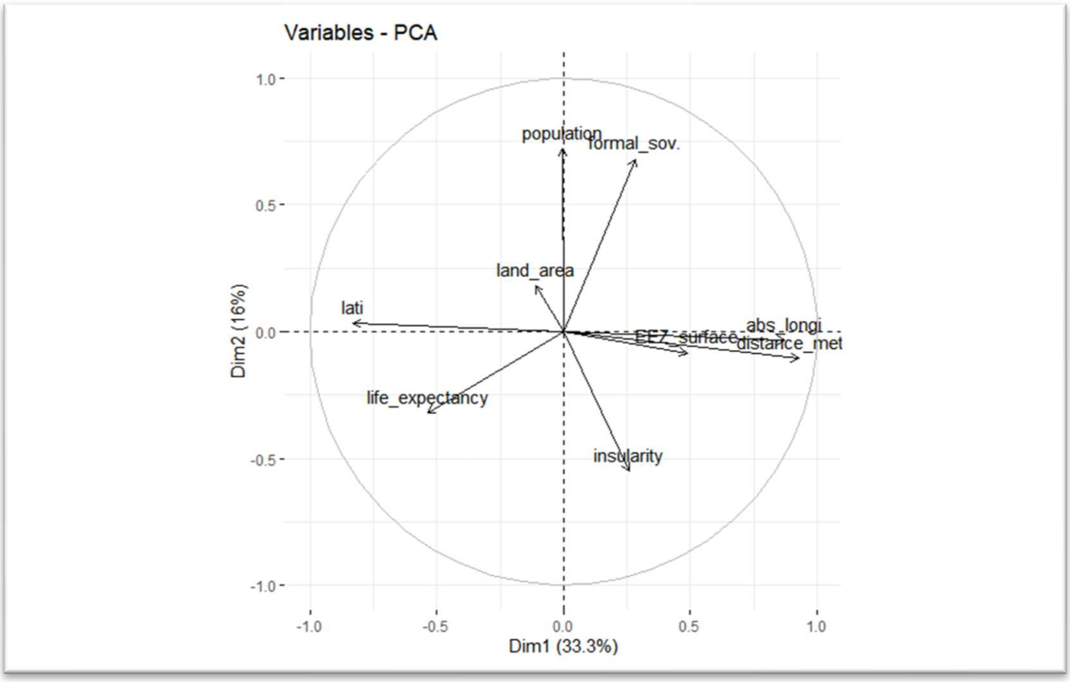
The first principal component (dimension 1) explains approximatively one third ($1/3^{rd}$) of the sample variance. It combines five main variables, the absolute longitude and distance to main metropolitan power that are strongly and negatively correlated with the latitude and life expectancy, and (more moderately) positively correlated with EEZ surface. The second component (dimension 2), that is independent from the dimension 1, explains one sixth ($1/6^{th}$) of the variance. It is significantly driven by the sovereignty index that is strongly and positively associated with the population size, and moderately and negatively with insularity. Besides, the land area is poorly represented in the PCA (the $\cos^2$ falls below 0.12, see Appendix 2, with a minimal contribution to the axes, so that this variable would not be significant in the territories' projection on the plane). A first conclusion is that life expectancy is more correlated with geography, while sovereignty is more correlated with population size. Land area is surprisingly a characteristic that shows a low correlation with the other characteristics and then a low (combined) contribution in explaining the sample variance.

<u>Figure 2:</u> Explained variance by principal components (dimensions).



Note: The first principal component axis explains approximately 33% of the variance within the sample; the second explains approximately 16%.

<u>Figure 3:</u> Projection graph of the two first dimensions.



Note: In the projection graph, the longer the arrow, the greater the contribution to the component and the better the quality of projection; the closer the variables' direction, the closer the positive association between them.

Figure 4 displays the result of the PCA analysis. How territories are located on the graph depends on the construction of both axis. Here, territory's location along the horizontal line – X axis (dimension 1) depends on latitude, absolute longitude, life expectancy, distance to metropolitan power and EEZ. For instance, high values of life expectancy and latitude combined with low values of absolute longitude and distance to metropolitan power, and a small EEZ surface area, locate the territory on the extreme left of the graph.

Location along the vertical line – Y axis (dimension 2) mainly depends on the level of sovereignty, population size and insularity. A high level of sovereignty combined to a large population and a low insularity locate the territory at the top of the graph. For instance, the location of Taiwan, on the upper side of the plan, is mostly explained by a large population combined with a high level of sovereignty, and a high value of latitude and life expectancy. Territories falling around the centre of the plan are characterized by 'moderate' values on the underlying variables.

Figure 4: PCA results: location of the 100 SIEs.

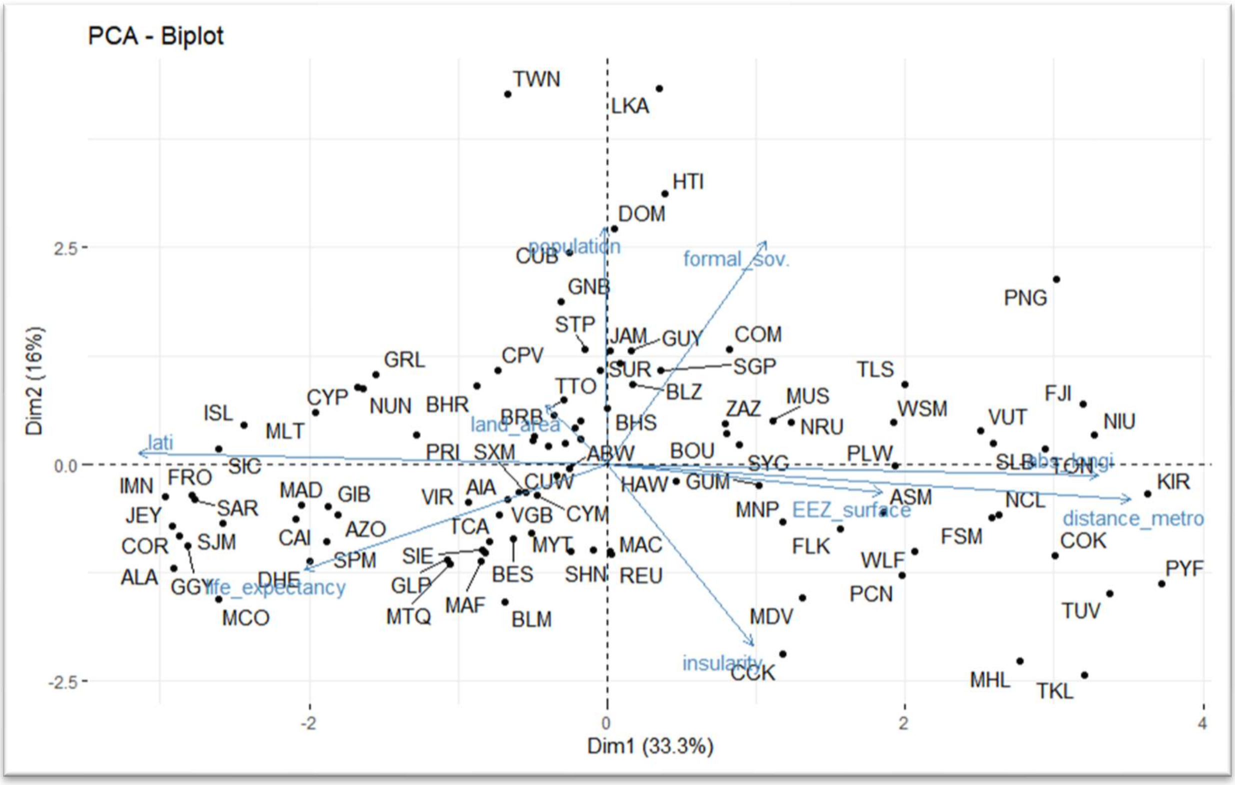
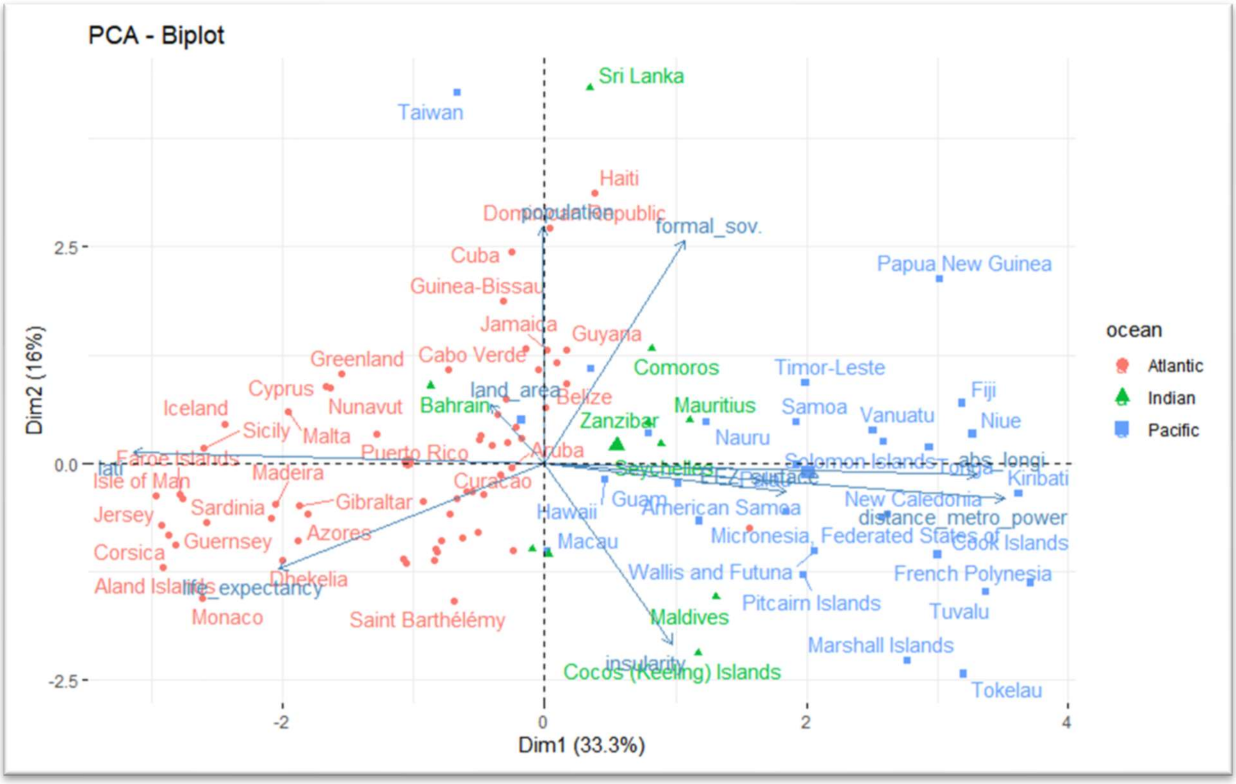


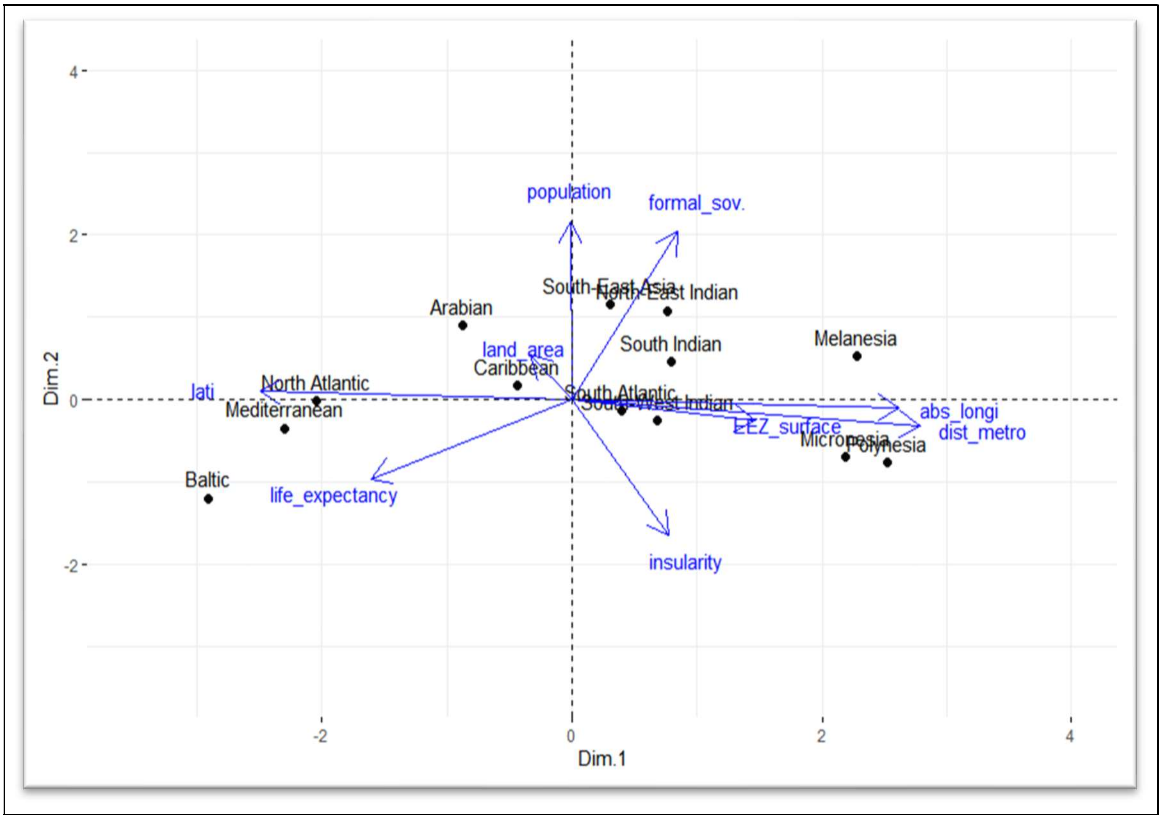
Figure 5 highlights the geographical location of the territories by ocean to reveal regional sub-sample similarities and dissimilarities within the global picture. While heterogeneity is always visible, two distinct groups emerge: the Atlantic territories that are characterized by a larger latitude and longer life expectancy, contrasting with the Pacific territories characterized by a larger insularity, EEZ surface, and distance to metropolitan power. Indian ocean territories show off a more heterogeneous pattern.

Figure 5: PCA results contrasted with location by oceans and sub-regions.



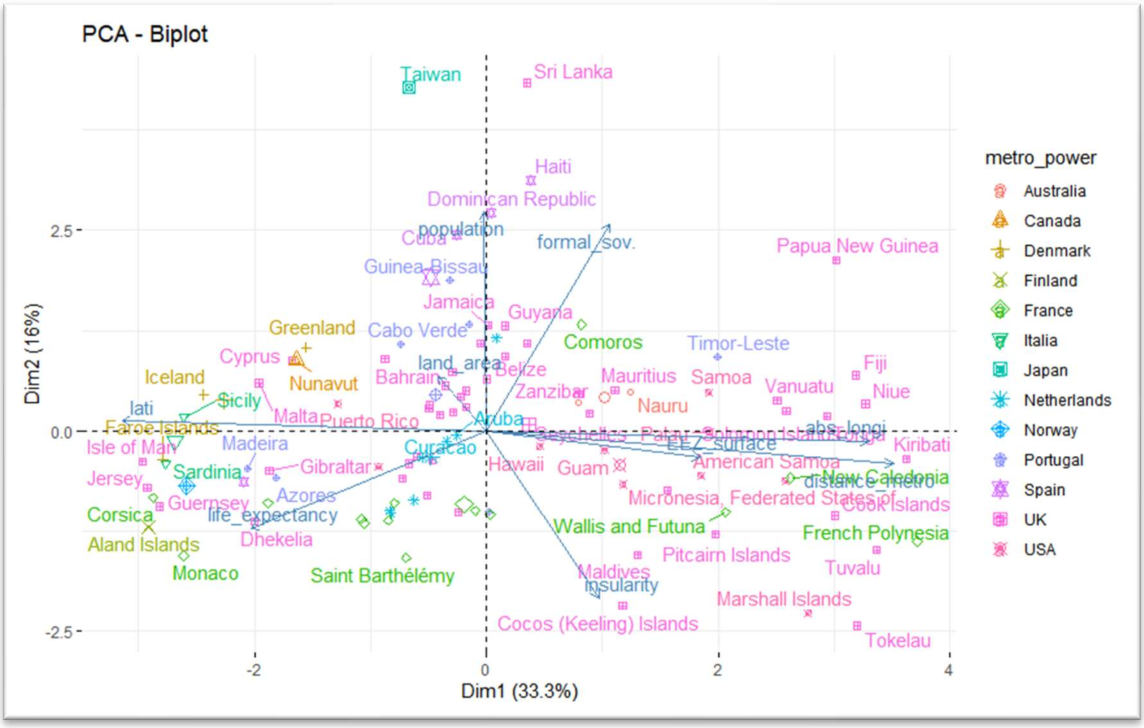
Refining further the geographical location in Figure 6, we report the location of each sub-region at the gravity centre of the locations of all territories that composed it. Thus, the Polynesian, Melanesian and Micronesian territories fall at the extreme boundaries of the spectrum, unveiling the lowest latitude and largest distance to their former colonial power (combined with low life expectancy). At the opposite end, the Baltic, Mediterranean and North Atlantic territories are characterized by the largest latitude and lowest distance to their metropolitan power, combined with higher life expectancy. These two opposed groups mostly display a low to moderate level of formal sovereignty. Conversely, the Caribbean territories mostly located around the centre of the graph show the most moderate level of every variable. Finally, Arabian, South-East Asian and North-East Indian territories stand out with the largest population and formal sovereignty.

<u>Figure 6</u>: Location of each sub-region's gravity centre.



<u>Figure 7</u> highlights the metropolitan power's identity. Contrasting with geography, no clear pattern can be identified. With a few exceptions, like the former Spanish colonies (Cuba, Dominican Republic and Haiti) in the Caribbean region, islands that depend on the same metropolitan power are mostly scattered on the plan.

Figure 7: PCA contrasted with metropolitan powers.



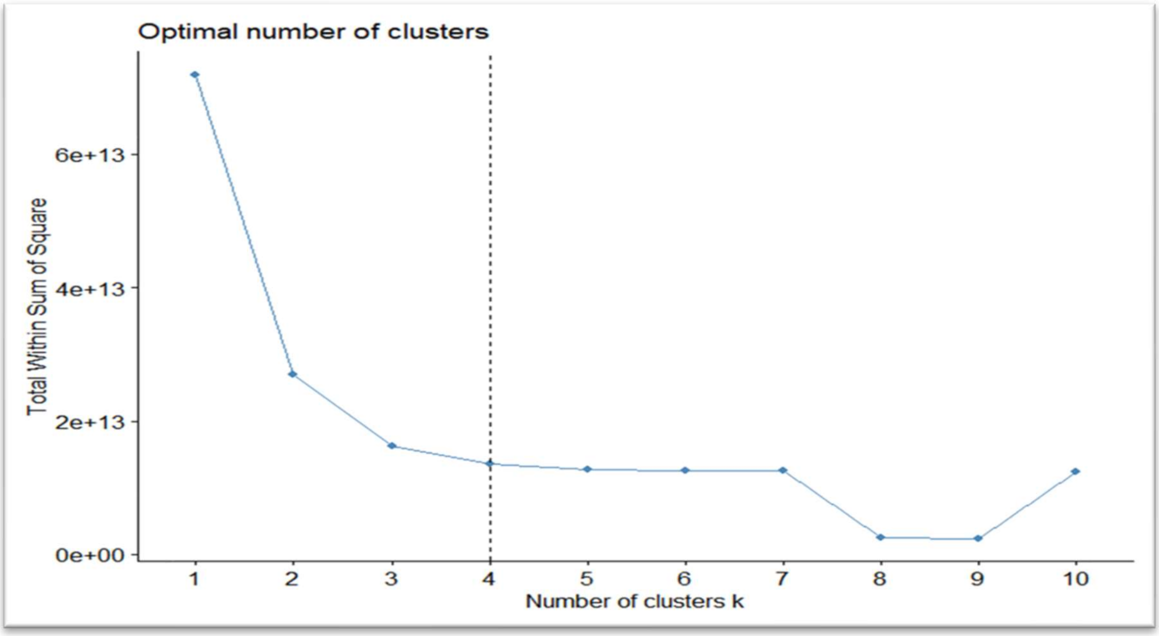
Clustering

We attempt a first data-driven typology of the 100 territories by groups related to similarities and dissimilarities in geographic, demographic and sovereignty terms. Following the results from the PCA, a clustering algorithm is performed to highlight subgroups where members show similarities while showing dissimilarities with other groups, with regard to the variables included in the first two components of the PCA. The k-mean algorithm[8] is used to both minimize the within-group variability and maximize the between-group variability. The clustering algorithm works as follows: (i) individual territories are randomly assigned into a predetermined number of clusters; (ii) centres of gravity are calculated for each cluster; (iii) individual territories are assigned to the nearest cluster; (iv) centres of gravity are recalculated for these new clusters. The second and third process are iteratively implemented until convergence (i.e. no more changes / stability in the individuals' assignment). We follow the elbow method to identify the optimal number of clusters as being four (Figure 8)[9].

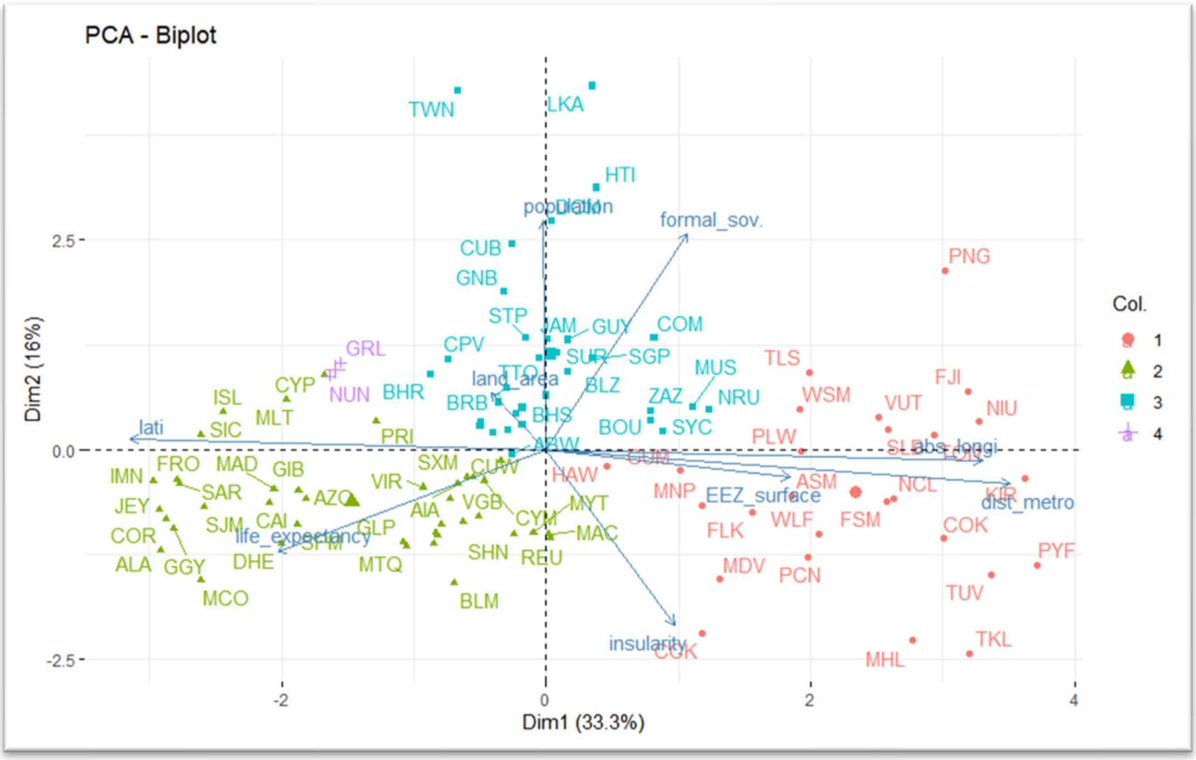
[8] Other algorithms (e.g., dendograms) have been tested and generated similar results.

[9] We tested alternative numbers of clusters, such as 3 and 5, which result in similar findings. Interestingly, the "uniqueness" of Nunavut and Greenland is preserved since they still constitute one cluster when the number of clusters is 3 or 5. However, referring to the gain of inertia, 3 clusters seems too small, and 5 clusters does not result in a significant gain in explaining sample variance.

<u>Figure 8</u>: Optimal number of clusters.



<u>Figure 9</u>: Clustering of the territories.



<u>Figure 9</u> plots territories grouped into the four clusters. The first cluster (blue) gathers 32 territories characterized by relatively high values of formal sovereignty (the group's mean equals 0.92, with a minimum level of 0.58 for Aruba) and population size; and low values of absolute longitude, latitude and insularity. They however show heterogeneous levels of life expectancy.

The second cluster (green) is composed of 26 territories characterized by relatively high values for longitude, distance to metropolitan power, EEZ surface and population, but low values of latitude, and heterogeneous levels of sovereignty.

The third cluster (purple) is opposed to the former cluster, and gathers territories characterized by a large latitude, high levels of life expectancy and short distance to metropolitan power, and a low level of sovereignty (a mean of 0.35). This is the largest group gathering 40 territories (of which a number of Mediterranean, Caribbean and North Atlantic territories).

The remaining cluster gathers two atypical territories, Greenland and Nunavut, of which the uniqueness is characterized by high values of land area, latitude and EEZ surface (and secondarily by a moderate level of sovereignty).

Conclusion

We explored the link between the level of sovereignty and its main geographic and demographic correlates, by performing a Principal Component Analysis followed by a clustering exercise. The resulting typology, though preliminary, is useful to depict the association between territories' geographical and demographic characteristics and their level of sovereignty.

The PCA reveals a large heterogeneity among these SIEs, firstly explained by a first dimension combining geographic and demographic variables such as distance to metropolitan power and EEZ size that are opposed to latitude and life expectancy (in contrast, land area does not contribute). Sovereignty, associated (positively) with population size and (negatively) with insularity, constitutes the second dimension that explains the sample heterogeneity. The PCA and the associated clustering generate a typology and a grouping of these SIEs that rather matches the location by ocean, more than the metropolitan power's identity.

SIEs can be grouped into four clusters mainly characterized by, respectively: Group 1 (32 territories): high sovereignty associated with a large population; Group 2 (26 territories): high values of latitude and life expectancy (mostly Atlantic and Baltic territories); Group 3 (40 territories): large distance to metropolitan power and insularity (Pacific Regions); and Group 4: Greenland and Nunavut, 2 territories with large land area, latitude and EEZ surface area.

This analysis based on a large sample of SIEs reveals interesting associations between geographic, demographic variables and sovereignty, calling for a further in-depth investigation of the determinants of sovereignty. This could be done by including more potential correlates such as colonization history and more economic variables and explore causal relationships based on regressions.

Acknowledgements:

The authors thank Olivier Santoni for his technical support on the calculations of the geographical data, and the participants of the POLIR project (Cemotev-UVSQ-Paris-Saclay) for useful discussions. They also thank two anonymous referees for constructive suggestions. All remaining errors are our own.

Disclaimer

The authors acknowledge the support received from Agence Nationale de la Recherche of the French government through the program “Investissements d’avenir ” (ANR-10-LABX-14-01).

References

Alberti, F. and Goujon, M. (2019). A composite index of formal sovereignty for small islands and coastal territories. *Island Studies Journal*, *15*(1), 3-24.

Armstrong, H. W., and Read, R. (2021). The non-sovereign territories: Economic and environmental challenges of sectoral and geographic over-specialisation in tourism and financial services. *European Urban and Regional Studies*, *28*(3), 213-240.

Augustin, J. L., and Liaw, S. Y. (2019). Grouping the Americas and Asia-Pacific countries based on their ICT readiness, prioritization of travel, tourism and tourist service infrastructures. *International Journal of Marketing Studies*, *11*(2), 38.

Bakri, B., Rustiadi, E., Fauzi, A., and Adiwibowo, S. (2018). Regional sustainable development indicators for developing countries: case study of provinces in Indonesia. *International Journal of Sustainable Development*, *21*(1-4), 102-130.

Baldacchino, G. (2020). How far can one go? How distance matters in island development. *Island Studies Journal*, *15*(1), 25-42.

Baldacchino, G. (Ed.) (2015). *Solution protocols to festering island disputes*. London: Routledge.

Baldacchino, G., and Hepburn, E. (2012). A different appetite for sovereignty? Independence movements in subnational island jurisdictions. *Commonwealth and Comparative Politics*, *50*(4), 555-568.

Bertram, G. (2004). On the convergence of small island economies with their metropolitan patrons. *World Development, 32*(2), 343-364.

Bertram, G. (2015). Is independence good or bad for development in small island economies? A long-run analysis. *Région et Développement, 42*(1), 31-54.

Bonilla, Y., and Hantel, M. (2016). Visualizing sovereignty: Cartographic queries for the Digital Age. *SX: Archipelagos*, *1*, 1-19.

Dunn, L. (2011). The impact of political dependence on small island jurisdictions. *World Development, 39*(12), 2132-2146.

Ferdinand, M., Oostindie, G., and Veenendaal W. (2020). A global comparison of non-sovereign island territories: the search for ‘true equality’. *Island Studies Journal, 15*(1), 43-66.

Feyrer, J., and Sacerdote, B. (2009). Colonialism and Modern Income: Islands as Natural Experiments. *Review of Economics and Statistics*, *91*(2), 245-262.

Flanders Marine Institute (2019). Maritime Boundaries Geodatabase, version 11. Available online at https://www.marineregions.org/

Gerring, J., Maguire, M., and Jaeger, J. (2018). A general theory of power concentration: demographic influences on political organization. *European Political Science Review*, *10(*4), 491-513.

Grydehøj, A. (2016). Navigating the binaries of island independence and dependence in Greenland: Decolonisation, political culture, and strategic services. *Political Geography*, *55*(1), 102-112.

Husson, F., Le, S., and Pagès, J. (2017). *Exploratory Multivariate Analysis by Example Using R*, 2nd edition ed., Boca Raton LA: Chapman and Hall/CRC.

Madiès, T., Rota-Grasiozi, G., Tranchant, J. P., and Trépier, C. (2018). The economics of secession: a review of legal, theoretical, and empirical aspects. *Swiss Journal of Economics and Statistics, 154(*1), 1-18.

McElroy, J. L., and Pearce, K. B., (2006). The advantages of political affiliation: Dependent and independent small-island profiles. *The Round Table, 95*(386), 529-539.

McElroy, J. L., and Sanborn, K. (2005). The propensity for dependence in small Caribbean and Pacific islands. *Bank of Valletta Review (Malta), 31*(1), 1-16.

Medina, N., Abebe, Y. A., Sanchez, A., and Vojinovic, Z. (2020). Assessing socioeconomic vulnerability after a hurricane: A combined use of an index-based approach and principal components analysis. *Sustainability*, *12*(4), 1452.

Mrgudovic, N. (2012). Evolving approaches to sovereignty in the French Pacific. *Commonwealth & Comparative Politics*, *50*(4), 456-473.

Overton, J., Prinsen, G., Murray, W. E., and Wrighton, N. (2012). Reversing the tide of aid: Investigating development policy sovereignty in the Pacific. *Journal de la Société des Océanistes*, 135.

Pöllath, M. (2018). Revisiting island decolonization: The pursuit of self-government in Pacific island polities under US hegemony. *Island Studies Journal*, *13*(1), 235-250.

Prinsen, G., Lafoy Y., and Migozzi J. (2017). Showcasing the sovereignty of non-self-governing islands: New Caledonia. *Asia Pacific Viewpoint*, *58*(3), 331-346.

Rezvani, D. A. (2014). *Surpassing the sovereign state: the wealth, self-rule, and security advantages of partially independent territories*. Oxford: Oxford University Press.

Sambanis, N., and Milanovic, B. (2011). Explaining the demand for sovereignty. *World Bank Policy Research Working Paper*, (5888).

Taglioni, F. (2011). Insularity, political status and small insular spaces. *The International Journal of Research into Island Cultures*, *5*(2), 45-67.

Veenendaal, W. (2020). Islands of democracy. *Area*, *52*(1), 30-37.

Watts, R. L. (2009). Island jurisdictions in comparative constitutional perspective. In G. Baldacchino and D. Milne (Eds.), *The case for non-sovereignty: Lessons from sub-national island jurisdictions* (pp. 21-39). London: Routledge.

<u>Appendix 1</u>: Result of clustering of a large sample of 100 small islands and coastal territories.

Cluster / Characteristics	Territories
Cluster 1 : high scores of formal sovereignty; low values of absolute longitude, latitude and insularity	Antigua and Barbuda, Aruba, Bahamas, Bahrain, Barbados, Belize, Bougainville, Cabo Verde, Comoros, Cuba, Dominica, Dominican Republic, Grenada, Guinea-Bissau, Guyana, Haiti, Hong Kong, Jamaica, Mauritius, Montserrat, Nauru, St Kitts and Nevis, St Lucia, St Vincent and the Grenadines, Sao Tomé and Principe, Seychelles, Singapore, Sri Lanka, Suriname, Taiwan, Trinidad and Tobago, Zanzibar
Cluster 2: high values of land area, latitude and EEZ surface.	Greenland, Nunavut
Cluster 3: high values of longitude, distance to first patron, EEZ surface, insularity and population, and low values of latitude	American Samoa, Samoa, Cocos (Keeling) Islands, Cook Islands, Falkland Islands (Islas Malvinas), Fiji, French Polynesia, Guam, Hawai´i, Kiribati, Maldives, Marshall Islands, Micronesia Fed. St. of, New Caledonia, Niue, Northern Mariana Islands, Palau, Papua New Guinea, Pitcairn Islands, Solomon Islands, Timor-Leste, Tonga, Vanuatu, Wallis and Futuna, Tokelau, Tuvalu
Cluster 4: high values of latitude and life expectancy and low values of distance to metropolitan power, EEZ surface, longitude and population.	Aland Islands, Anguilla, Azores, Bermuda, Bonaire, British Virgin Islands, Canary Islands, Cayman Islands, Corsica, Curacao, Cyprus, Dhekelia, Faroe Islands, French Guiana, Gibraltar, Guadeloupe, Guernsey, Iceland, Isle of Man, Jersey, Macau, Madeira, Malta, Martinique, Mayotte, Monaco, Puerto Rico, Réunion, Saba, St Barthélémy, Saint Martin, Saint Pierre and Miquelon, Sardinia, Sicily, Sint Eustatius, Sint Maarten, Svalbard, Turks and Caicos Islands, United States Virgin Islands, St Helena, Ascension, Tristan da Cunha

Appendix 2: PCA results: Tables of coordinates, contributions and cos2

Variables	Dimension 1			Dimension 2		
	Coord.	Contr.	Cos2	Coord.	Contr.	Cos2
formal_sov.	0.283	2.66	0.080	0.681	**32.13**	**0.463**
insularity	0.258	2.22	0.067	-0.551	**21.08**	**0.304**
Life_expect	-0.536	**9.60**	**0.288**	-0.320	7.09	0.102
Latitude	-0.830	**22.99**	**0.689**	0.033	0.08	0.001
Distance_metro power	0.929	**28.78**	**0.862**	-0.107	0.80	0.011
population	-0.005	0.00	0.000	0.720	**36.00**	**0.519**
Land area	-0.110	0.41	0.012	0.180	2.25	0.032
EEZ surface	0.488	**7.94**	**0.238**	-0.085	0.50	0.007
Abs. longitude	0.872	**25.40**	**0.761**	-0.034	0.08	0.001

Notes: Coordinate ranges between -1 and 1 and gives the position of the variable on the dimension (axis). Contribution is the share of each variable in the construction of the component or dimension (with sum of shares equals 100). Cos2 ranges between 0 and 1: a high cos2 indicates a good representation of the variable on the principal component or dimension.