Geographic Distribution of Linguistic Diversity

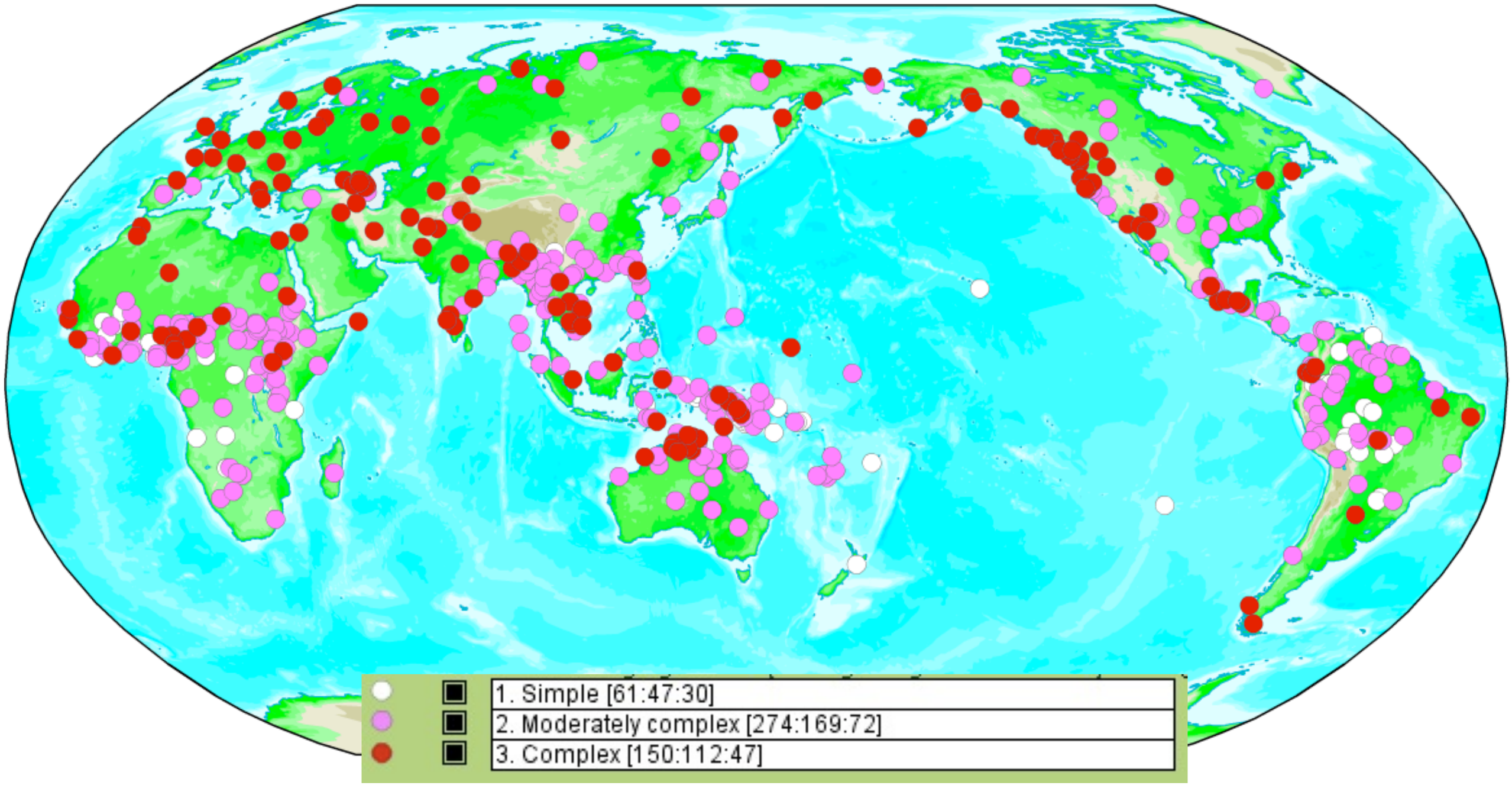
Michael Cysouw
MPI-EVA Leipzig, LMU München

World Atlas of Language Structures (WALS)

- 142 world maps with structural linguistic information about hundreds of languages
- Phonology, Morphology, Syntax, Lexicon
- Somewhat simplistic and at times even redundant classifications

Haspelmath, Dryer, Gil, Comrie (eds.) 2005. The World Atlas of Language Structures. Oxford: OUP.

Syllable Structure



Maddieson, Ian (2005) 'Syllable structure' in: Martin Haspelmath, Matthew S. Dryer, David Gil, & Bernard Comrie (eds.) *World Atlas of Language Structures*. Oxford: Oxford University Press, 54-57.


The World Atlas of Language Structures Online
Home
Features
Languages
References
Authors
Newsblog
Contact
Help
Download WALS data
errata
credits
legal
Welcome to WALS Online
The data and the texts from The World Atlas of Language Structures, published as a book with CD-ROM in 2005 by Oxford University Press, are now freely available online.
WALS Online is a joint project of the Max Planck Institute for Evolutionary Anthropology and the Max Planck Digital Library . It is a separate publication, edited by Martin Haspelmath, Matthew S. Dryer, David Gil and Bernard Comrie (Munich: Max Planck Digital Library, 2008).
What is WALS?
WALS is a large database of structural (phonological, grammatical, lexical) properties of languages gathered from descriptive materials (such as reference grammars) by a team of more than 40 authors (many of them the leading authorities on the subject).
WALS consists of 141 maps with accompanying texts on diverse features (such as vowel inventory size, noun-genitive order, passive constructions, and "hand"/"arm" polysemy), each of which is the responsibility of a single author (or team of authors). Each map shows between 120 and 1370 languages, each language being represented by a symbol, and different symbols showing different values of the feature. Altogether 2,650 languages are shown on the maps, and more than 58,000 datapoints give information on features in particular languages.
WALS thus makes information on the structural diversity of the world's languages available to a large audience, including interested nonlinguists as well as linguists who would not normally read grammars of exotic languages or specialized works by comparative linguists. Although endangered languages are not particularly emphasized, they are automatically foregrounded because of the large sample of languages represented on each map, where each language (independently of its number of speakers) is shown by a single symbol.
Search
powered by Google
WALS News
Scheduled Server Downtime
by robert - May 05, 2009
To fulfill the test regulations of the Land Brandenburg, the data center where the WALS servers are hosted will have to shut ...
WALS Online not reachable
by robert - Apr 14, 2009
From April 13, 16:58 CEST until this morning, April 14, 7:46 CEST WALS Online was not reachable, due to an unplanned downtime ...
Location for Mehri updated
by robert - Mar 04, 2009
The geo-coordinates for Mehri have been corrected - since Mehri is spoken mostly in Yemen, it should be located there. Find the ...
Map for Chapter 141
by robert - Feb 03, 2009
Today, we got a step further in filling the gap between WALS Online and the printed edition of 2005. We incorporated data to be ...
Interactive Reference Tool (WALS program)
The World Atlas of Language Structures was published as a book with a CD-ROM in summer 2005. The CD-ROM contains the "Interactive Reference Tool (WALS program)" as a standalone application for Mac OSX, Mac OS9.2 and Windows 2000, XP written by Hans-Jörg Bibiko. To download the "Interactive Reference Tool (WALS program)" please follow the link http://www.eva.mpg.de/lingua/research/tool.php.

wals.info

Complexity

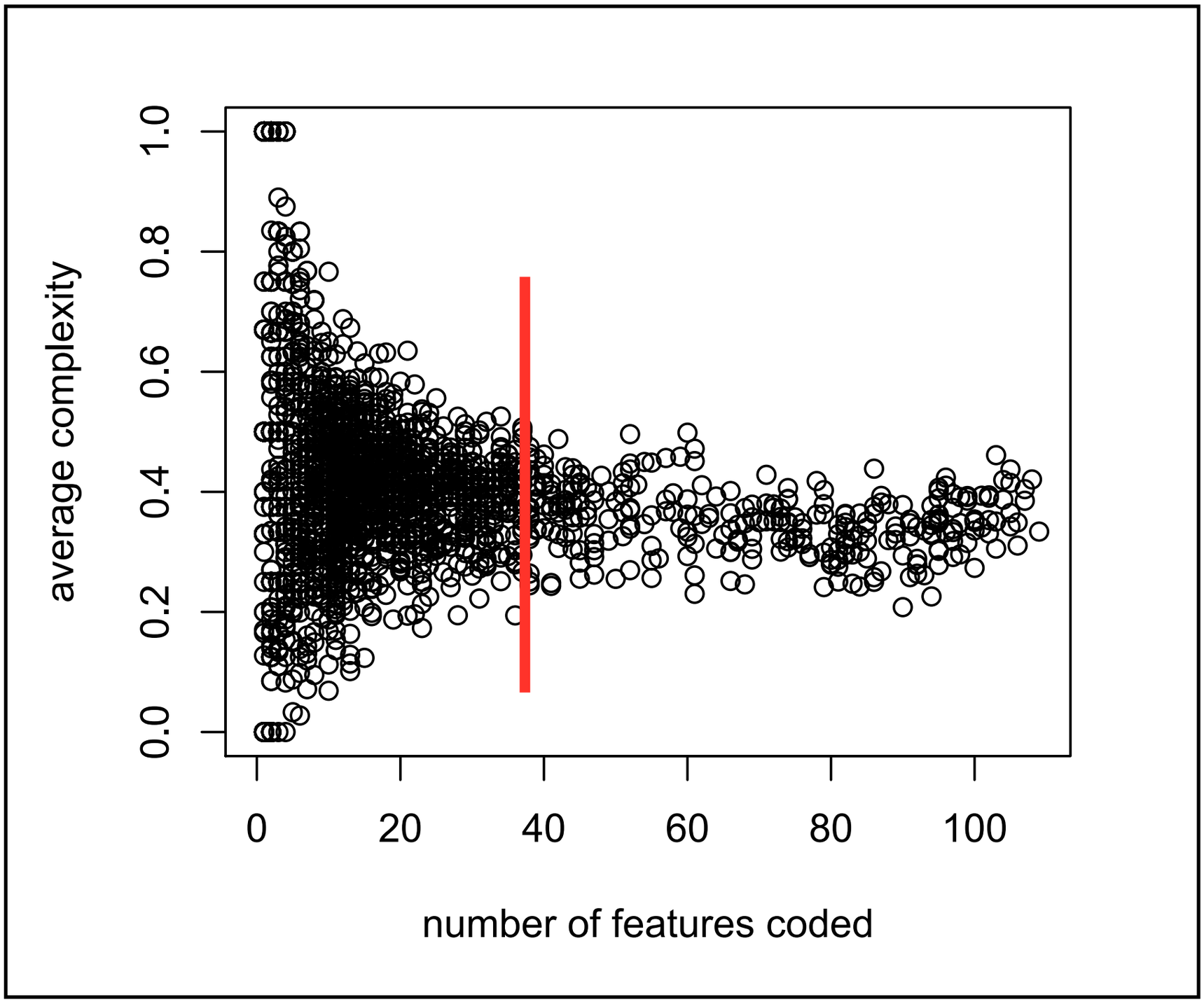
“Complexity”

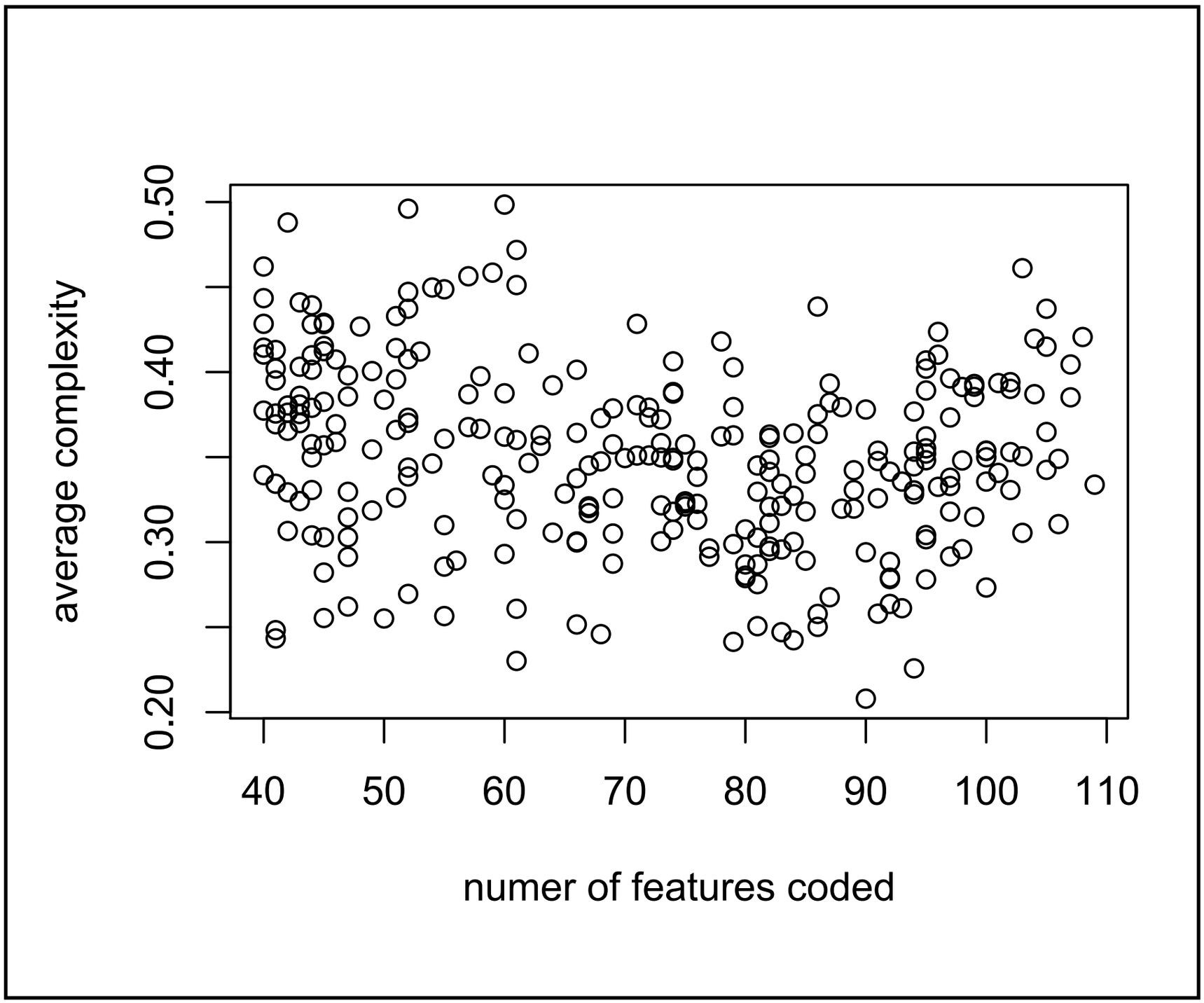
Factor X

- Many different possible definitions for "complexity"
- Here: a simple and easily applicable approach
- **More oppositions** and **strict structure** is interpreted as having more X ("complexity")

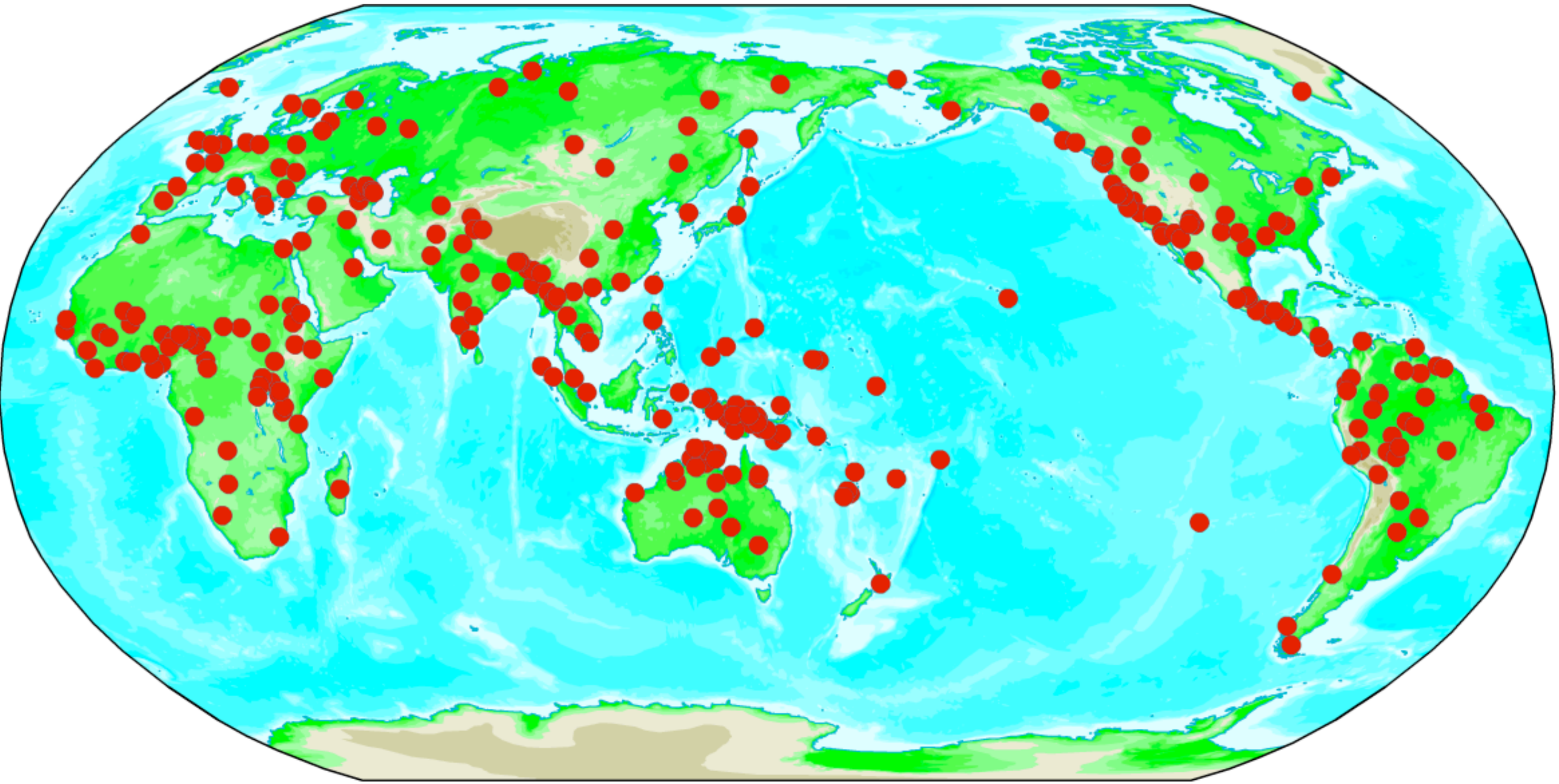
Average Complexity

- Average complexity is easily computed for each language in WALS
- However: WALS is notoriously incomplete
- An ad-hoc selection of languages is necessary

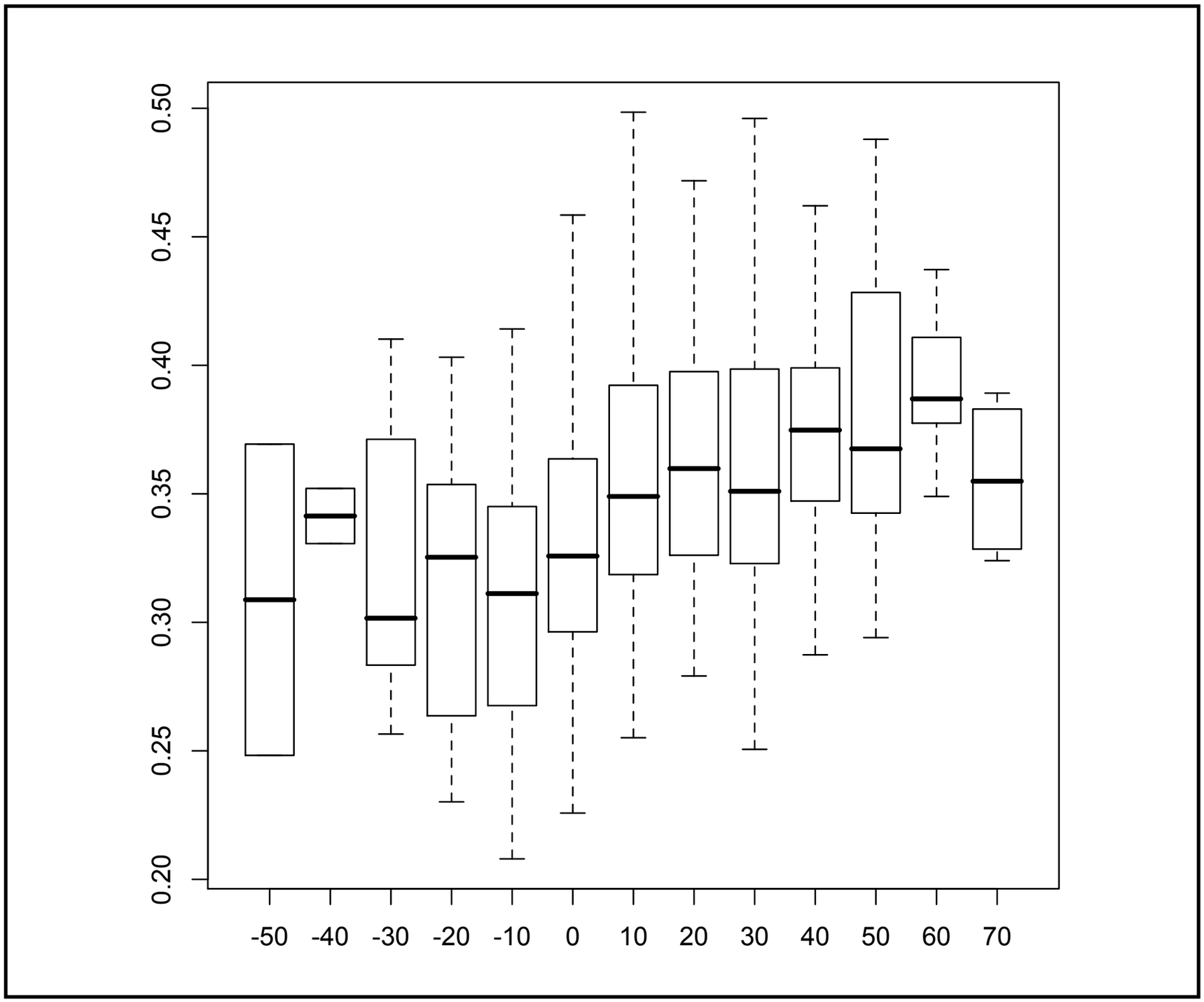

average complexity
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
number of features coded


average complexity
0.20
0.30
0.40
0.50
40
50
60
70
80
90
100
110
numer of features coded

280 language sample

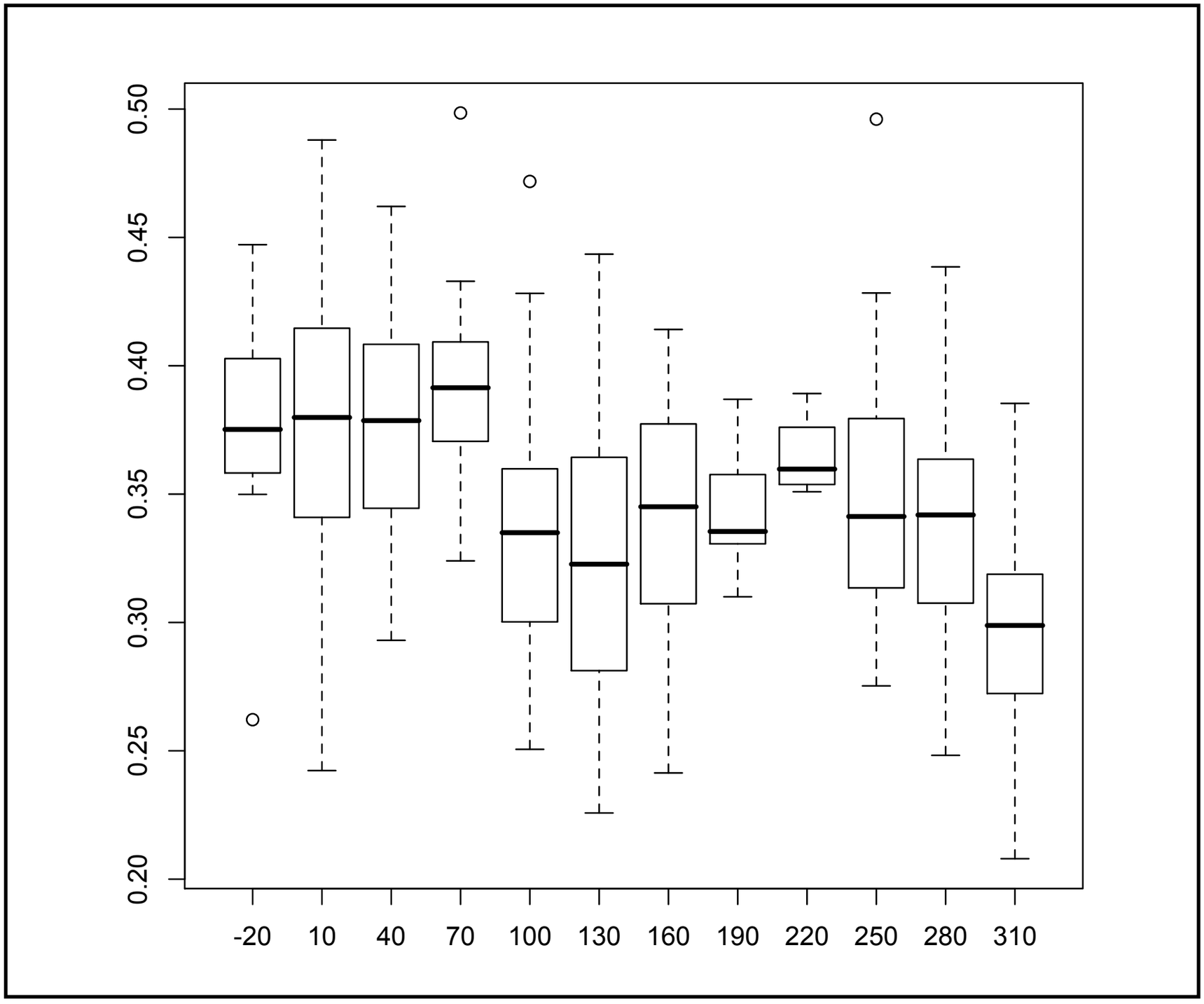


Latitude



South ← → North

Longitude

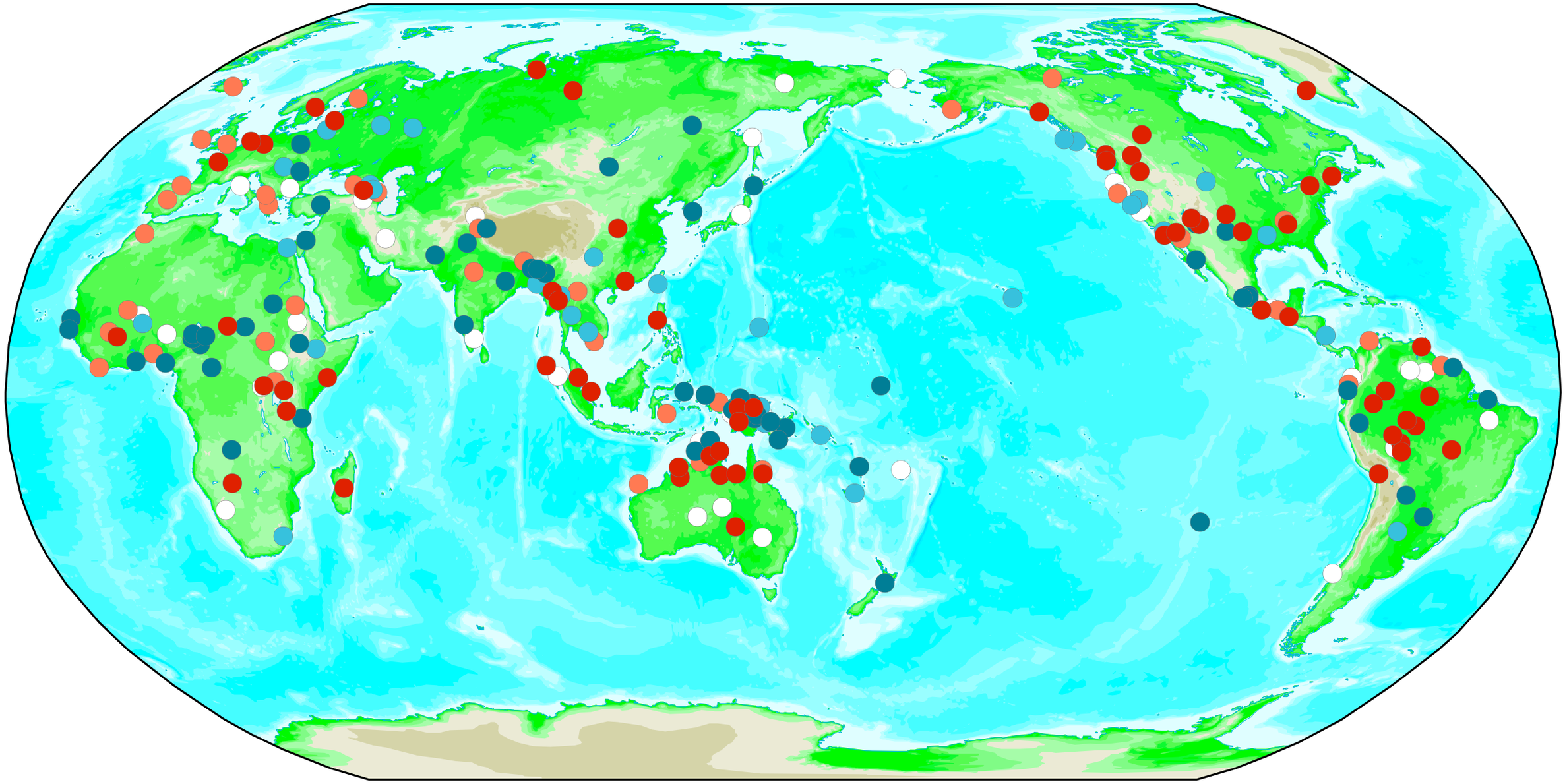


West ← → East

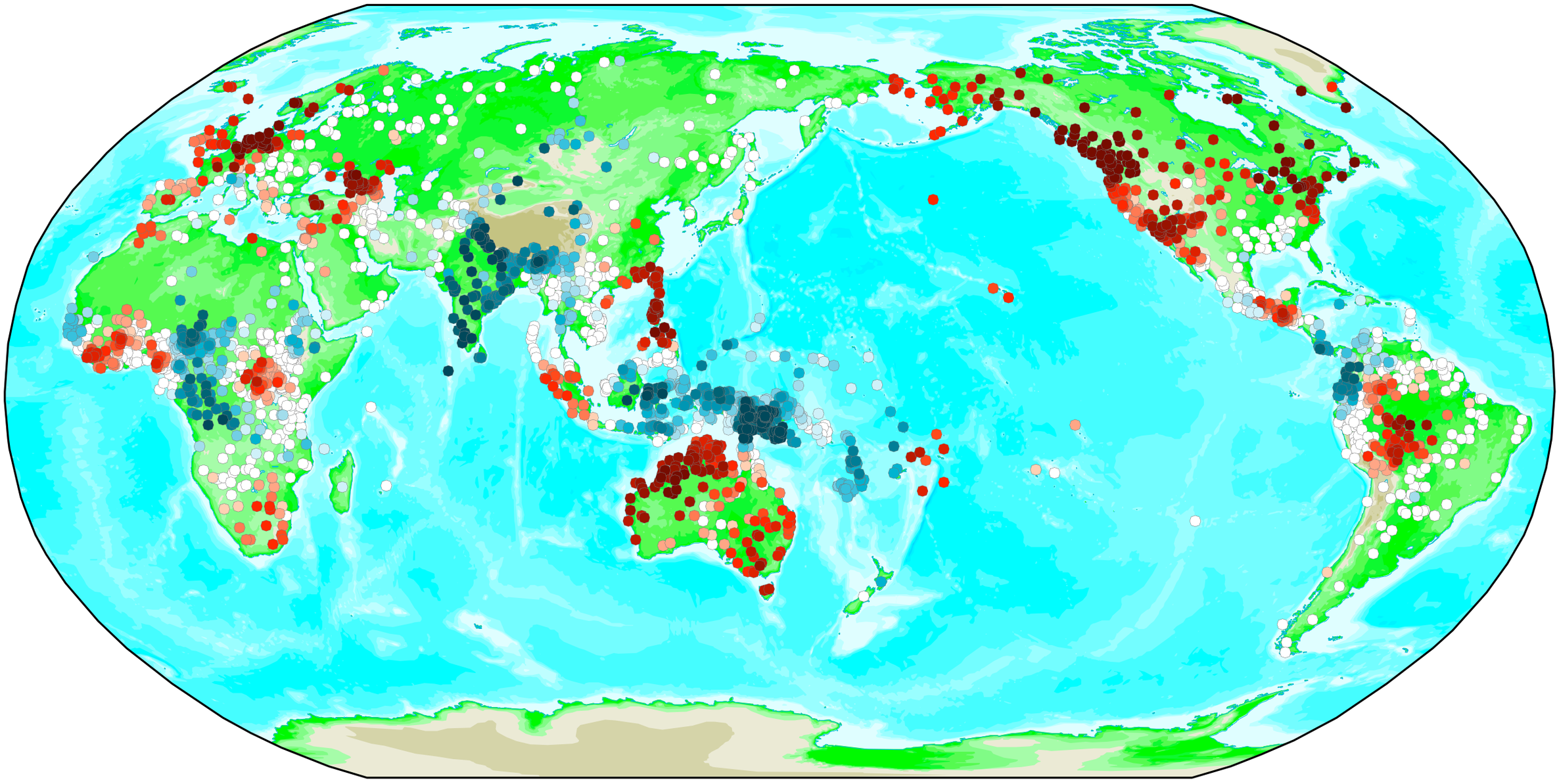
Smoothing Geographical Distributions

- Language variation is geographically rather haphazardly distributed
- To show general trends, take for each language the average of the closest languages
- here: average of the language itself and it's two closest neighbours in the sample

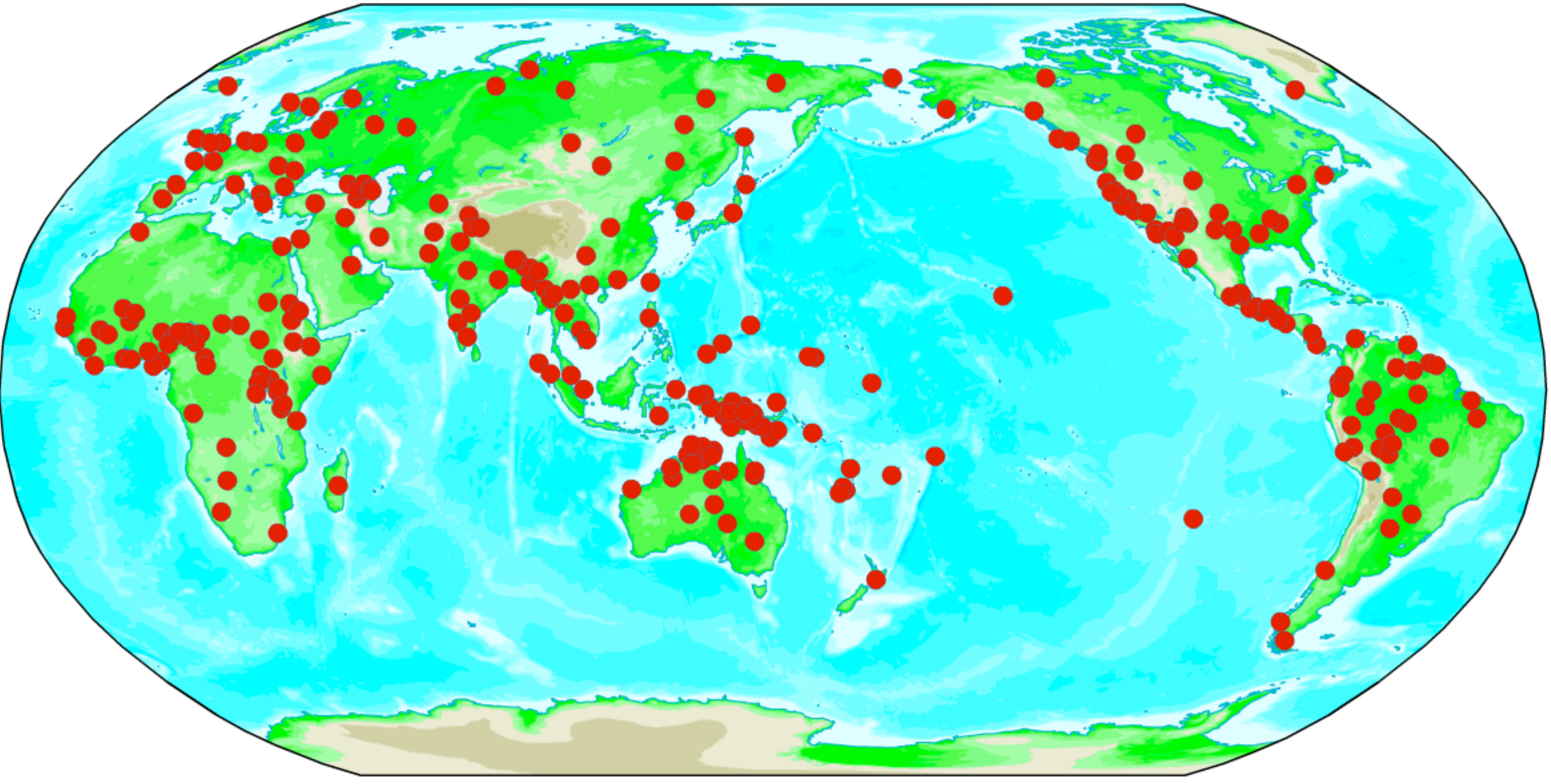
Average Rarity

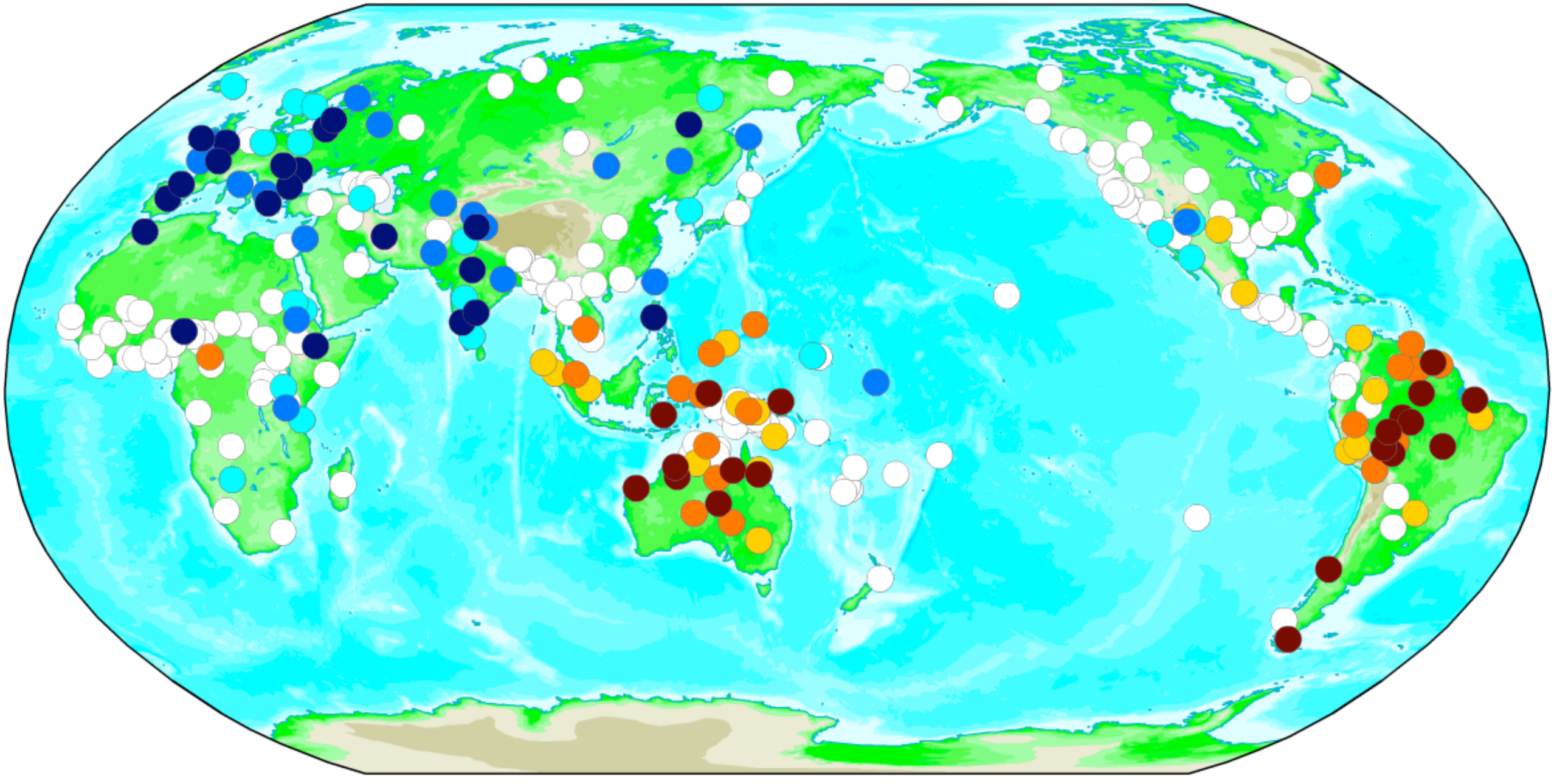


Average Rarity



back to our factor X

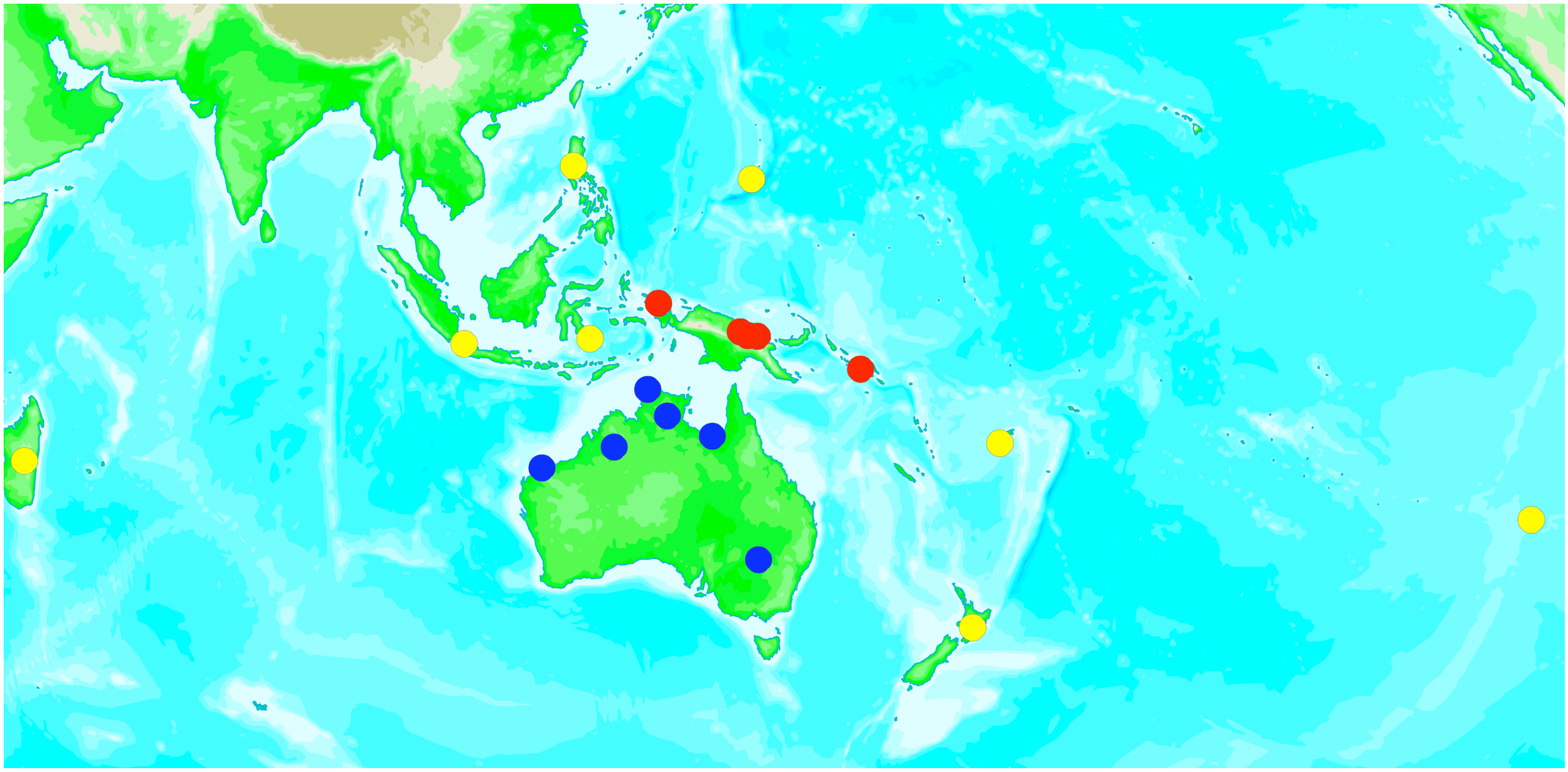


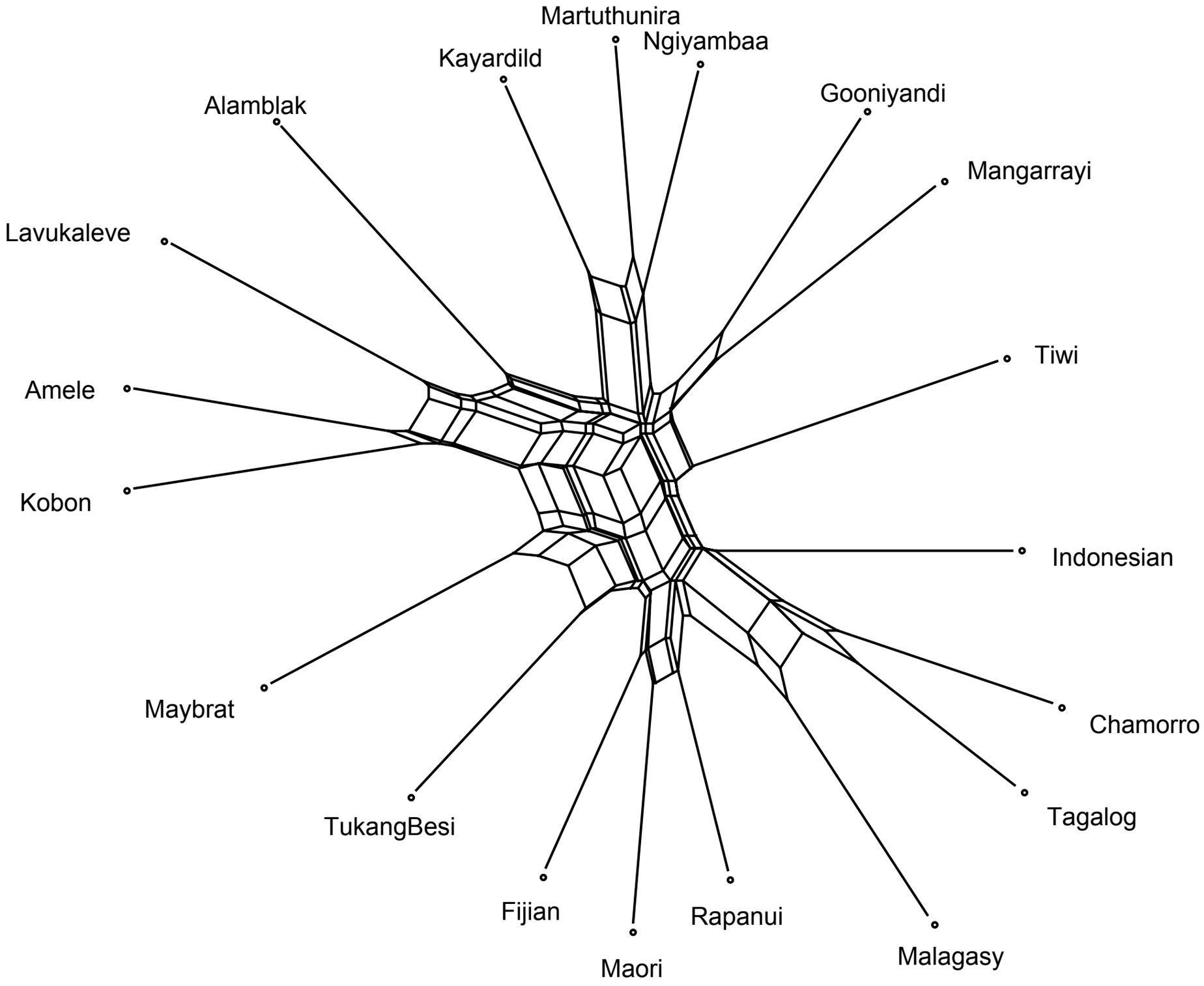


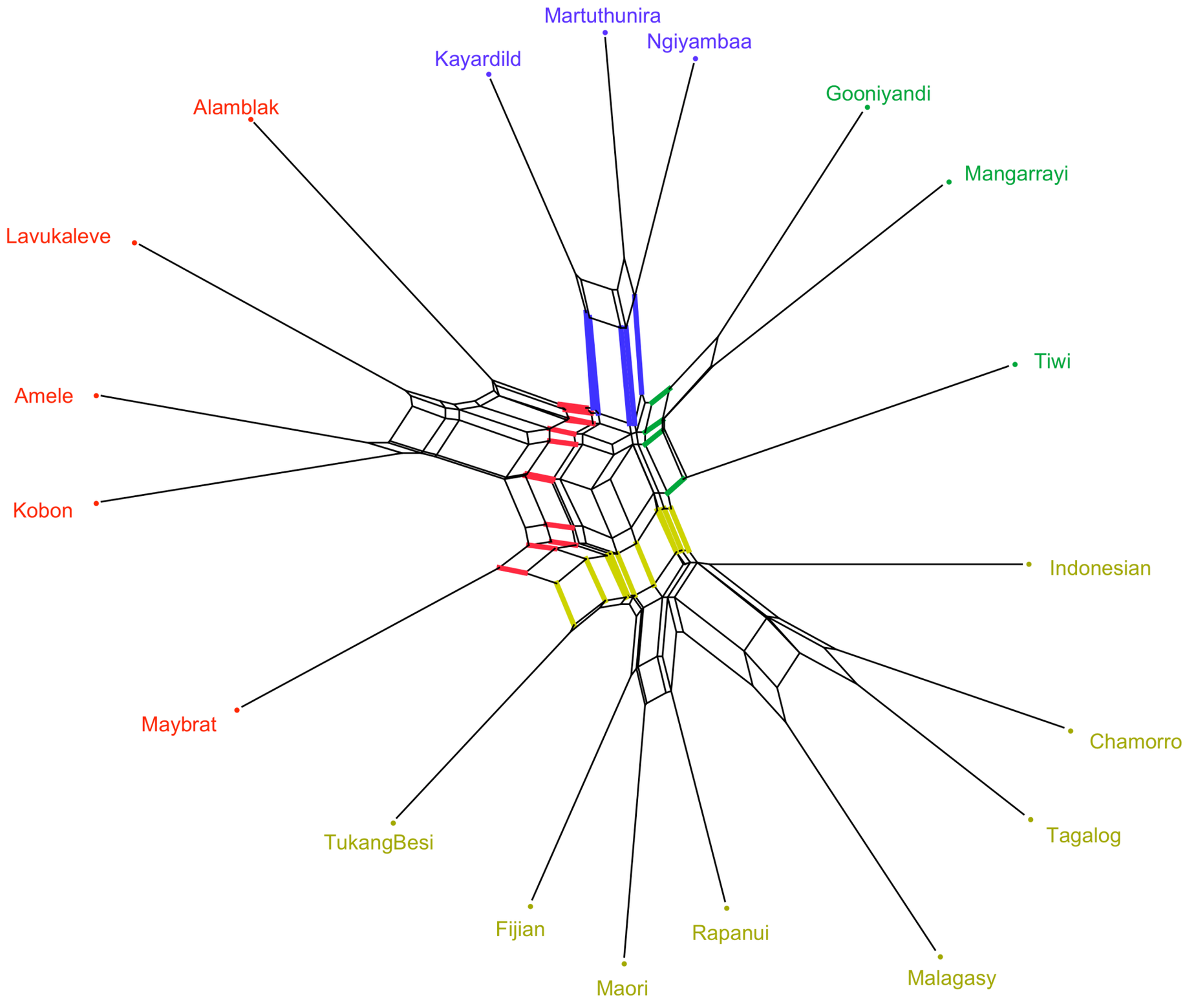
What does this mean?

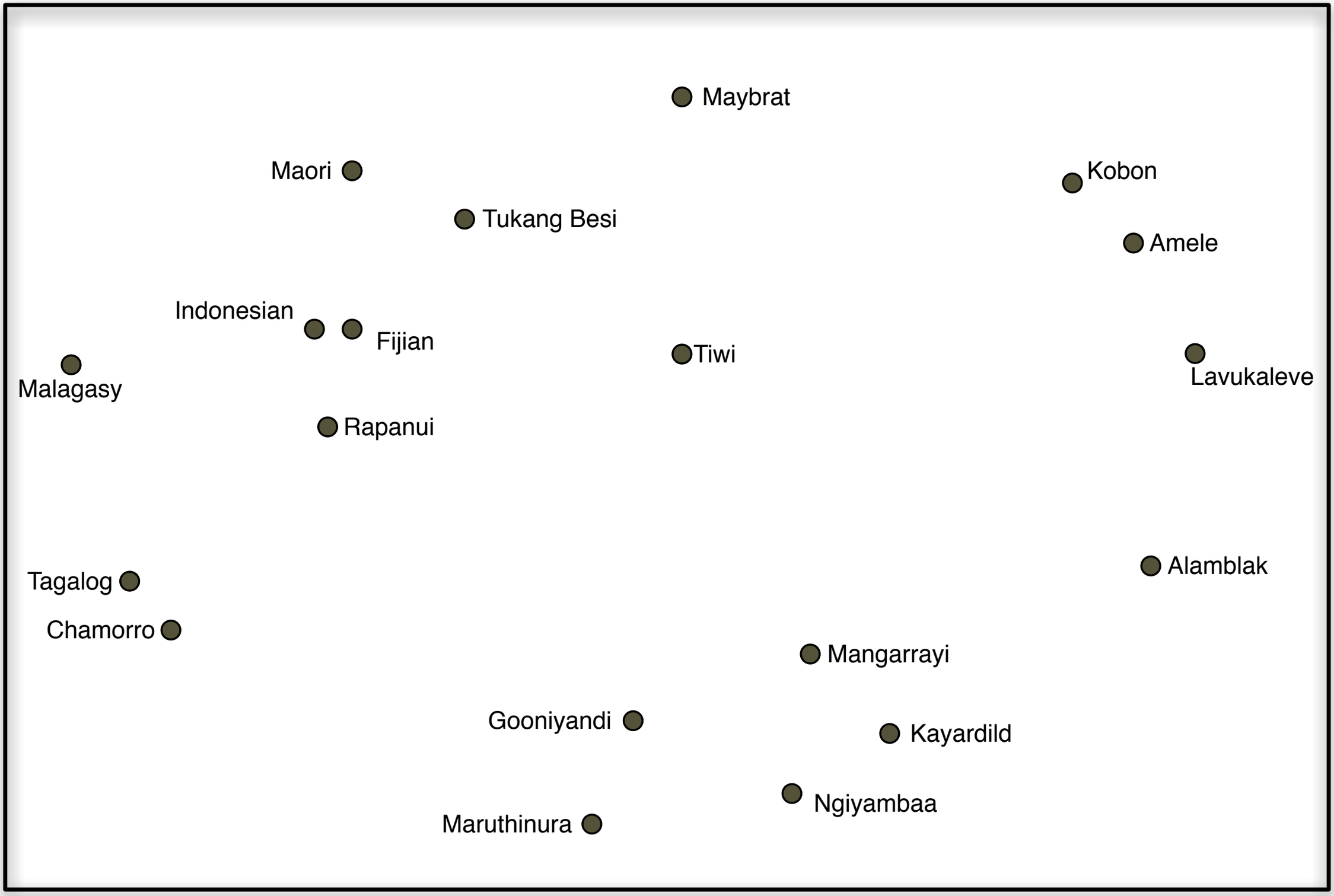
- Language with low complexity are found at the fringes of human settlement on the globe: this might be a sign on old language structure
- Languages with low complexity are the those farthest away from Europe: this indicates the eurocentricity of the feature selection in WALS (and in linguistics in general)

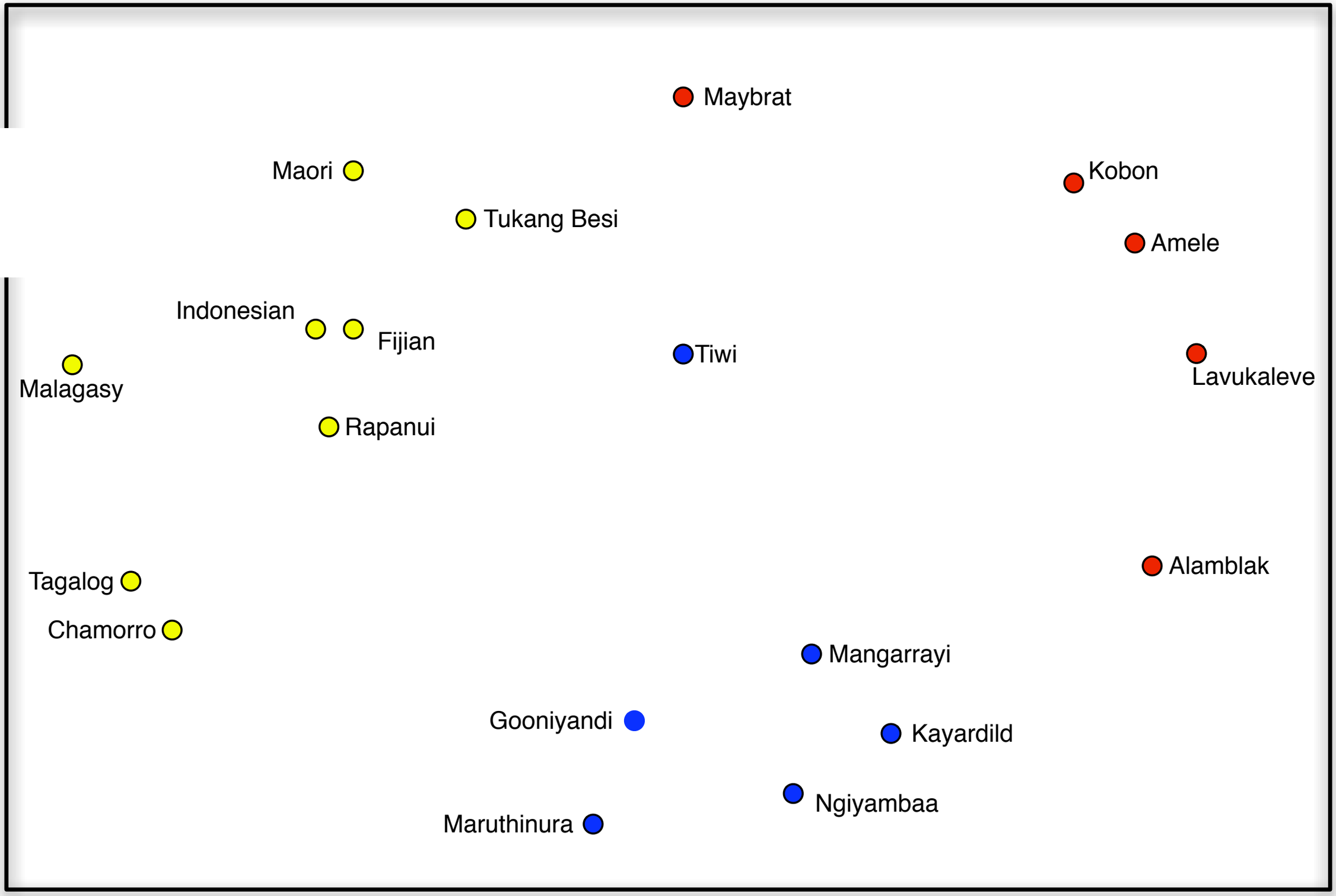
Comparing Typological Profiles

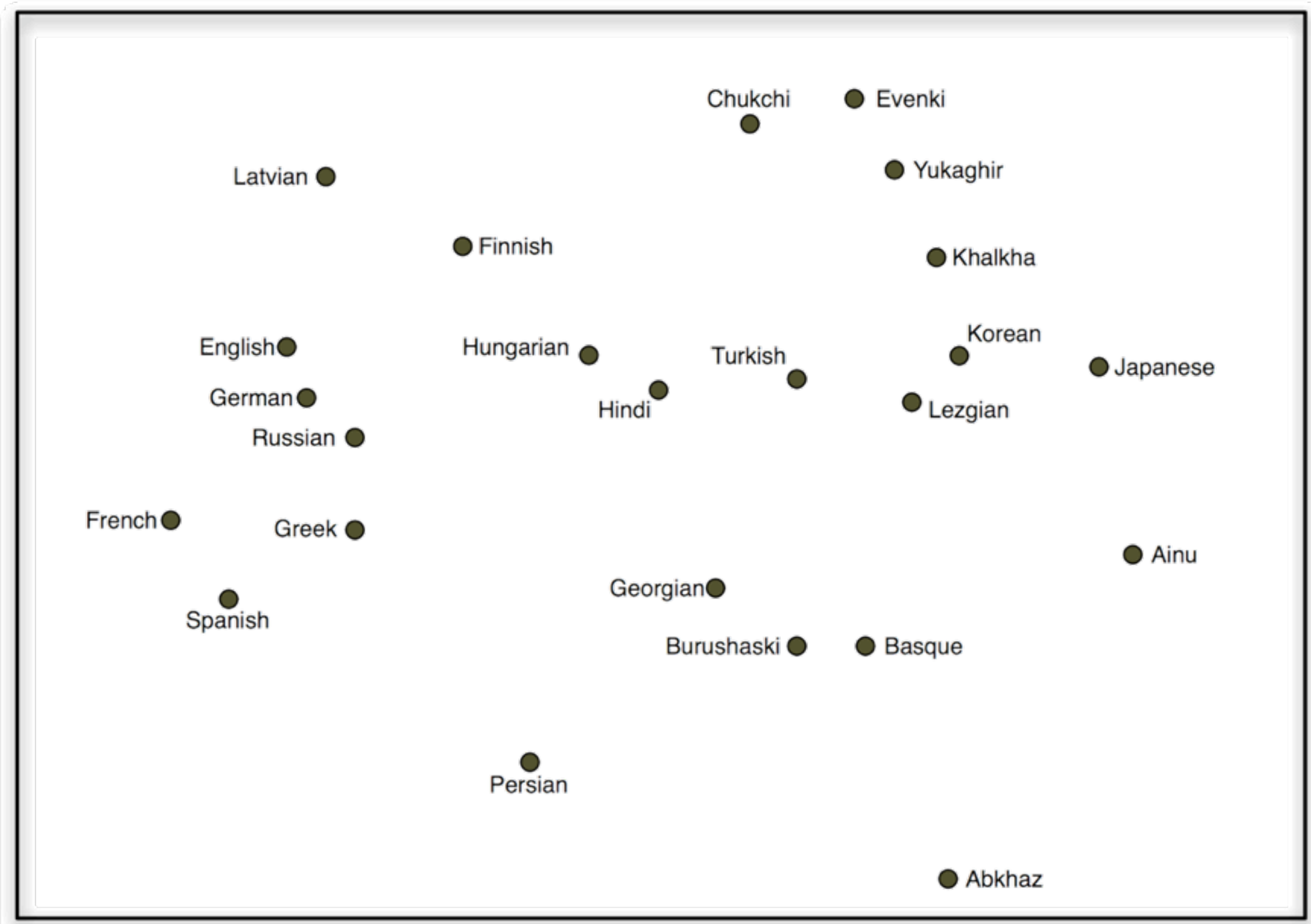


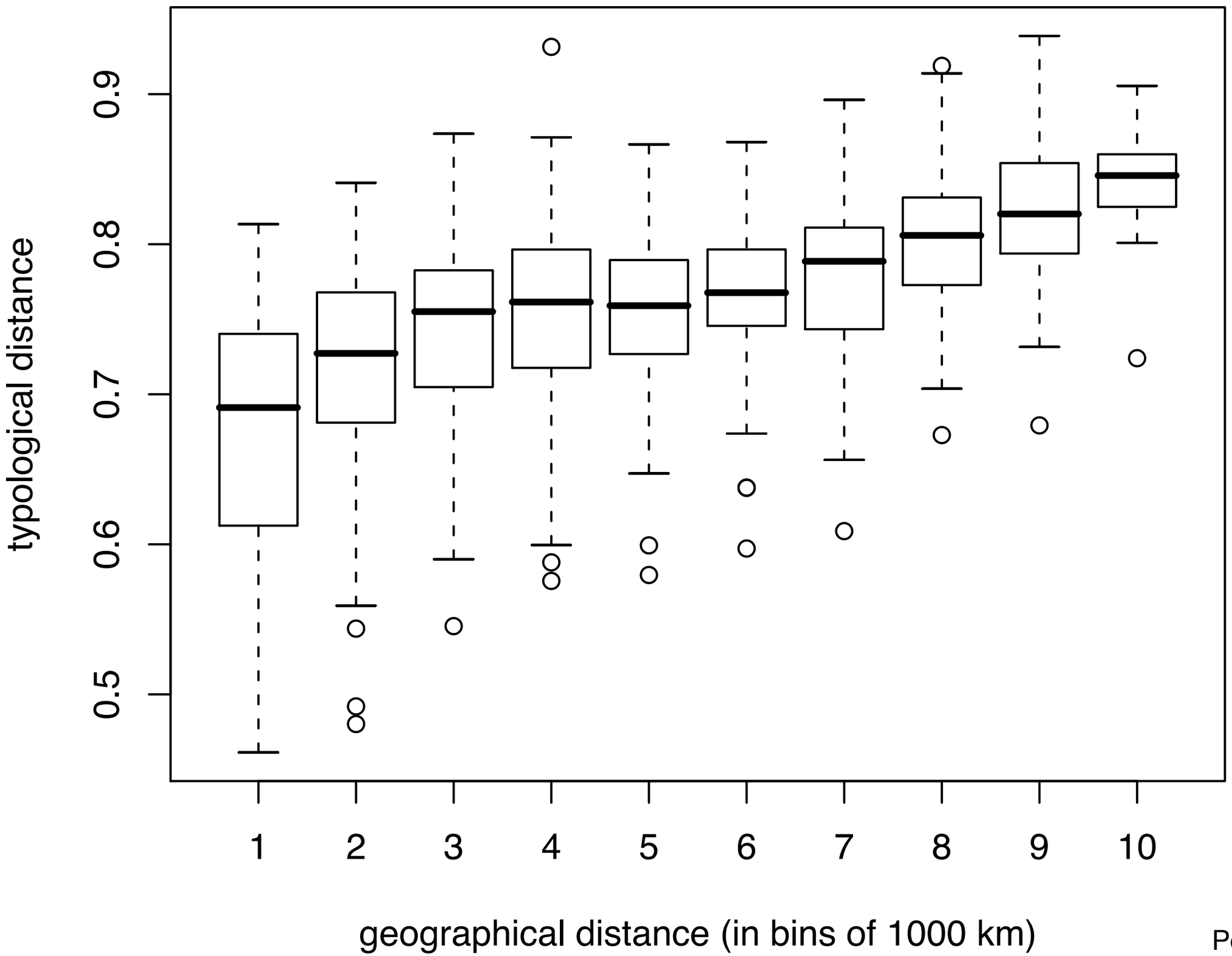

Martuthunira
Ngiyambaa
Kayardild
Alamblak
Gooniyandi
Mangarrayi
Lavukaleve
Tiwi
Amele
Kobon
Indonesian
Maybrat
Chamorro
Tagalog
TukangBesi
Fijian
Rapanui
Maori
Malagasy


Martuthunira
Ngiyambaa
Kayardild
Alamblak
Gooniyandi
Mangarrayi
Lavukaleve
Tiwi
Amele
Kobon
Indonesian
Maybrat
Chamorro
Tagalog
TukangBesi
Fijian
Rapanui
Maori
Malagasy

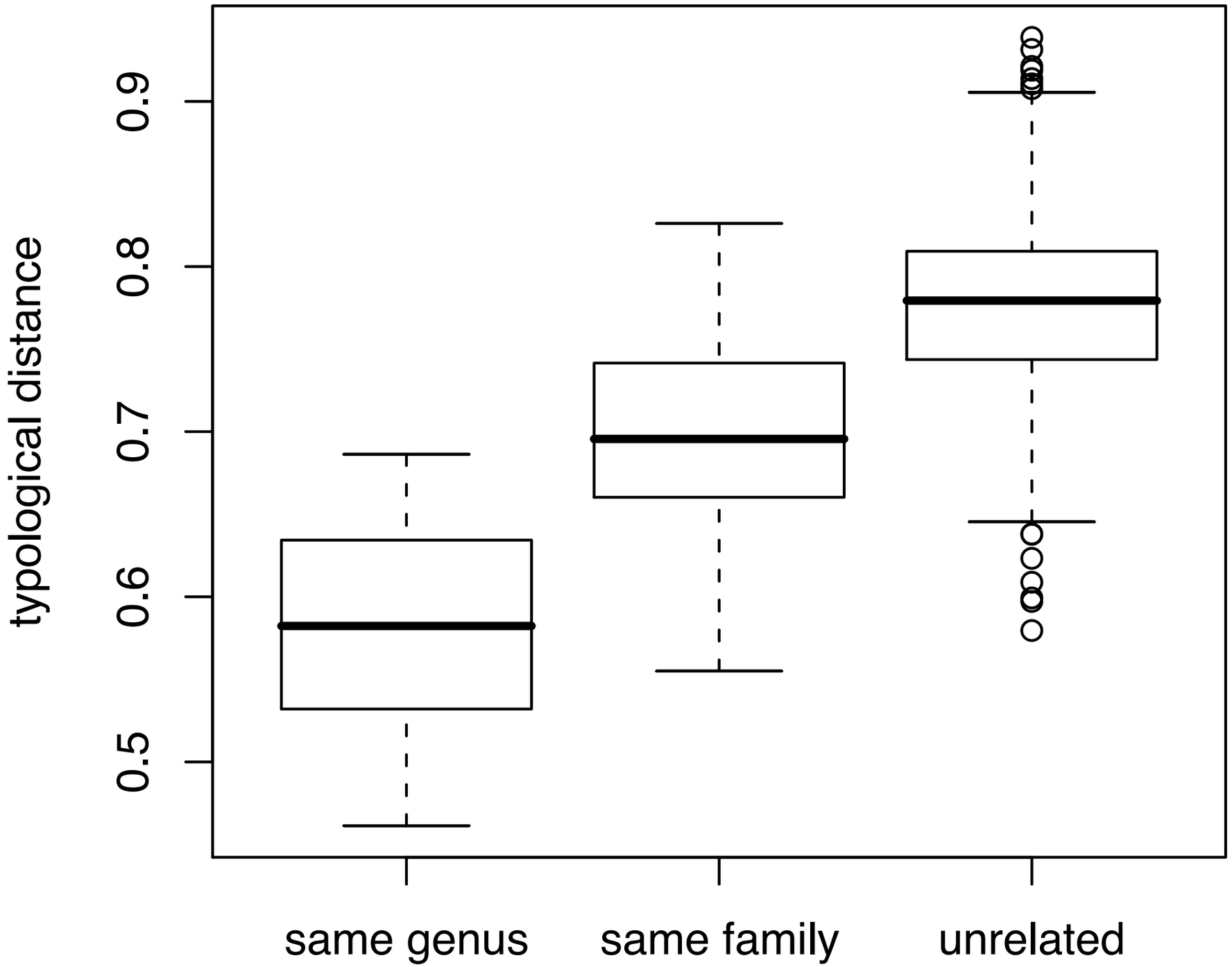

Maybrat
Maori
Kobon
Tukang Besi
Amele
Indonesian
Fijian
Tiwi
Lavukaleve
Malagasy
Rapanui
Alamblak
Tagalog
Chamorro
Mangarrayi
Gooniyandi
Kayardild
Ngiyambaa
Maruthinura


Maybrat
Maori
Kobon
Tukang Besi
Amele
Indonesian
Fijian
Tiwi
Malagasy
Lavukaleve
Rapanui
Alamblak
Tagalog
Chamorro
Mangarrayi
Gooniyandi
Kayardild
Ngiyambaa
Maruthinura


Chukchi
Evenki
Latvian
Yukaghir
Finnish
Khalkha
English
Hungarian
Turkish
Korean
Japanese
German
Hindi
Lezgian
Russian
French
Greek
Ainu
Georgian
Spanish
Burushaski
Basque
Persian
Abkhaz



Pearson r = .52
Mantel p < .001



Pearson r = .61
Mantel p < .001

Mantel Test

Correlation	Pearson *r*	Mantel *p*
Typology ~ Geography	0.52	< .001
Typology ~ Genealogy	0.61	< .001
Geography ~ Genealogy	0.33	< .001

Partial Mantel Test

Correlation	Pearson *r*	Mantel *p*
Typology ~ Geography + Genealogy	0.42	< .001
Typology ~ Genealogy + Geography	0.54	< .001

Multivariate Matrix Regression

	Sums of Sqs	Mean Sqs	F Model	R^2
family : genus	13.65	0.02	-3.29	0.77
latitude : longitude	0.28	0.28	-39.01	0.02
Residuals	3.82	-0.01		0.22

Zapala, M.A. and J. Schork (2006) Multivariate regression analysis of distance matrices for testing associations between gene expression patterns and related variables. PNAS 103(51): 19430–19435

How much geography is there left after factoring out genealogy ?

- Regression Typology ~ Genealogy
- Negative residuals after regression show ‘more similarity than expected by genealogy’
- This surpluss similarity is probably contact

Greek	Bulgarian
German	Dutch
Italian	French
Greek	Albanian
Korean	Japanese
German	French
Russian	Lithuanian
Latvian	Finnish
Swedish	English
French	Dutch
Russian	Finnish
Lezgian	Ingush
Romanian	Albanian

Metrics for geographic language distance ???

- Idea: approximate global probability of contact between languages
  - travel distance – depending on technology
    - walking distance
    - using horses, boat, plane
- However, probability of contact is not necessarily the related to actual contact
  - use language density as proxy to actual contact
  - the more languages in between to languages, the farther apart they are