countrycode: An R package to convert country names and country codes

Vincent Arel-Bundock¹, Nils Enevoldsen², and CJ Yetman³

1 Université de Montréal 2 Massachusetts Institute of Technology 3 Hertie School of Governance

Summary

International organizations, statistical agencies, and research labs use different codes to represent countries. For example, the ISO standard code for Algeria is “DZA”, but Eurostat uses “DZ”, the International Civil Aviation Organization uses “DA”, the Correlates of War project uses “ALG”, and the International Monetary Fund uses “612”. When researchers merge and analyze data from several sources, incompatible country codes can be a major source of frustration.

The countrycode package for R alleviates this problem by making four main contributions. First, it allows bidirectional conversion between more than 30 country code schemes. Second, it includes a set of well-tested regular expressions which can be used to convert long-form English or German country names into country codes. Third, countrycode can convert codes into the names of countries in almost any spoken language. Finally, the package allows users to define custom dictionaries to facilitate the conversion of other identifiers (e.g., provinces or cities).

These functions can support a variety of scientific activities. For instance, countrycode has been used to draw maps (Coene, 2018); to acquire data from sources like the US Census Bureau (Walker, 2018) or the World Bank (Arel-Bundock, 2018); to extract historical weather data from online APIs (Shum, 2018); and to process bird sightings records (Strimas-Mackey, Miller, & Hochachka, 2017). In our own practice as researchers, countrycode has proven to be an invaluable tool to merge datasets produced by organizations which use different country identifiers.

Examples

library(countrycode)

# ISO codes to Correlates of War codes
countrycode(c('DZA', 'USA'), origin = 'iso3c', destination = 'cowc')
[1] "ALG" "USA"

# English names to GENC3 codes
countrycode(c('Antigua and Barbuda', 'Russia'), origin = 'country.name', destination = 'genc3c')
[1] "ATG" "RUS"

# ISO codes to long-form Finnish names
countrycode(c('DOM', 'CAN'),...
Acknowledgements

We acknowledge contributions from Andrew Collier, Christopher Gandrud, @grasshoppermouse, Anh Le, Bastiaan Quast, Étienne Têtreault-Pinard.

References


