Package: toponym (via r-universe)

November 4, 2026

Title Toponym Analysis & Visualization Tool

Version 1.0.0

Description A tool to analyze and visualize toponym distributions. This package is intended as an interface to the GeoNames data. A regular expression filters data and in a second step a map is created displaying all locations in the filtered data set. The functions make data and plots available for further analysis—either within R or in a chosen directory. Users can select regions within countries, provide coordinates to define regions, or specify a region within the package to restrict the data selection to that region or compare regions with the remainder of countries. This package relies on the R packages `geodata` for map data and `ggplot2` for plotting purposes. For more details see Wichmann & Chevallier (2025) <doi:10.5195/names.2025.2616>.

Depends R (>= 4.1)

License GPL (>= 3) + file LICENSE

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.3

VignetteBuilder knitr

LazyData true

Imports geodata, ggplot2, grDevices, sf, spatstat.geom, spatstat.utils, stats, terra, utils

URL https://github.com/Lennart05/toponym

BugReports https://github.com/Lennart05/toponym/issues

Suggests knitr (>= 1.50)

Config/pak/sysreqs libabsl-dev cmake libgdal-dev gdal-bin libgeos-dev libssl-dev libproj-dev libsqlite3-dev libudunits2-dev

Repository https://lennart05.r-universe.dev

Date/Publication 2026-03-30 11:10:06 UTC

RemoteUrl https://github.com/lennart05/toponym

RemoteRef HEAD

RemoteSha fb4fc36e26c36c72818c8573630afad9de31e22f

Contents

toponym-package 2
country 3
createPolygon 4
getData 5
mapper 6
ortho 7
top 8
topComp 9
topCompOut 11
topFreq 13
toponymOptions 14
topZtest 15

Index **16**

`toponym-package` *toponym: Toponym Analysis & Visualization Tool*

Description

A package to analyze and visualize toponym distributions.

The main functions are the following:

- `top` returns and plots selected toponyms onto a map.
- `country` helps in navigating designations of countries and regions used by the package.
- `createPolygon` lets users create a polygon by point-and-click or directly retrieve polygon data.
- `mapper` plots a user-specific data frame onto a map.
- `topComp` compares toponym substrings in a polygon and in the remainder of a country (or countries).
- `topCompOut` saves multiple maps and toponym data.
- `topFreq` retrieves most frequent toponym substrings.
- `topZtest` lets users apply a Z-test on toponym distributions
- `toponymOptions` lets users modify settings for managing toponym data

For more detailed descriptions please read the respective documentation.

Author(s)

Maintainer: Lennart Chevallier <mail@lchevallier.de>

Authors:

- Søren Wichmann <wichmannsoeren@gmail.com>

See Also

Useful links:

- https://github.com/Lennart05/toponym
- Report bugs at https://github.com/Lennart05/toponym/issues

---

country *Country designations*

Description

This function returns country and region designations used by the toponym package.

Usage

```
country(query = NULL, ...)
```

Arguments

query	character string vector. Enter queries to access information on countries.
...	Additional parameter: • regions numeric. If 1, outputs the region designations of the respective countries. By default, it is 0.

Details

If you enter an individual country designation, you receive the three different designations (IS02, ISO3, name).

If you enter "ISO2" or "ISO3", you receive a vector of all ISO-codes of the respective length.

If you enter "names", you receive a vector of all country names.

If you enter "country table", you receive a data frame with all three designations for every country.

Region designations are retrieved from the geodata package map data. The list of region designations may be incomplete. For mapping purposes, geodata is used throughout this package.

Value

Returns country designations selected from a data frame. If regions is set to 1, returns region designations in a matrix selected from the geodata map data.

Examples

```
## Not run:
country(query = "ISO3")
## returns a vector of all ISO3 codes

country(query = "Thailand")
## returns a list with a data frame with ISO2 code, ISO3 code and the full name

country(query = "Thailand", regions = 1)
## returns a list with a matrix with all region designations

## End(Not run)
```

`createPolygon` *Creates a polygon*

Description

This function lets users create a polygon by point-and-click or directly retrieve polygon data.

Usage

```
createPolygon(countries, ...)
```

Arguments

`countries`	character string vector with country designations (names or ISO-codes).
`...`	Additional parameters:

- `regions` numeric. Specifies the level of administrative borders. By default `0` for displaying only country borders.
- `region_ID` character string vector with region IDs.
- `region_name` character string vector with region names.
- `retrieve` logical. If `TRUE`, the coordinates of the region or country are returned. No map will be drawn.

Details

Parameter `countries` accepts all designations found in `country(query = "country table")`.

`region_ID` and `region_name` accepts region designations for the selected countries, which can be retrieved by `country()`. The function prioritizes any `region_ID` and ignores `region_name` if users provide both. The matrix from `country()` listing all region designations may be incomplete as the `geodata` mapa data is incomplete in this regard. For mapping purposes, `geodata` is used throughout this package.

In RGui, users exit the point selection by middle-clicking or right-clicking and then pressing stop.

In RStudio, users exit the point selection by pressing ESC or Finish in the top right corner of the plot. Users whose points are shifted away, are advised to set the zoom settings of RStudio and of their device to 100%:

Tools -> Global Options -> Appearance -> Zoom

This function uses the function `spatstatLocator` provided by the `spatstat.utils` package for the point-and-click functionality. For further details on the point-and-click mechanism, please refer to the help page for `spatstatLocator`.

Value

A data frame with the coordinates of the polygon.

Examples

```
## Not run:
createPolygon("NA", region_ID = "NAM.7_1")

# a plot of the region Ohangwena in Namibia
# by point-and-click a polygon can be created
# use country() to find all acceptable region IDs

Ohangwena_polygon <- createPolygon(
"NA", region_ID = "NAM.7_1", retrieve = TRUE
)
# no plot appears
# the coordinates of the region are stored in the object
# and can be used by other functions

## End(Not run)
```

`getData` *Downloads GeoNames data*

Description

This function downloads toponym data for the package.

Usage

```
getData(countries, overwrite = FALSE)
```

Arguments

`countries`	character string vector with country designations (names or ISO-codes).
`overwrite`	logical. If `TRUE`, the data sets (.txt files) in the package folder will be overwritten.

Details

The data is downloaded from the GeoNames download page and thereby made accessible to `readFiles()`. The function allows users to update GeoNames data and to set the date of access to that database to the current date. Parameter `countries` accepts all designations found in `country(query = "country table")`.

See Also

GeoNames download page

Examples

```
## Not run:
getData(countries = c("DK", "DE"), save = FALSE)
## downloads and extracts data for DK and DE to the temporary folder

getData(countries = c("DK", "DE", "PL"), save = TRUE)
## downloads and extracts data for PL but only extracts data for DK and DE
## from the zip files downloaded before to the package folder if used in the same session

## End(Not run)
```

`mapper` *Plots data onto a map*

Description

This function plots a user-specific data frame onto a map.

Usage

```
mapper(mapdata, ...)
```

Arguments

`mapdata`	data frame. A user-specific data frame with coordinates.
`...`	Additional parameters: • `color` character string vector indicating, which color is assigned to each string. • `regions` numeric. Specifies the level of administrative borders. By default `0` for displaying only country borders. • `plot` logical. If `FALSE`, the plot will not be printed but saved as .png in the current working directory. • `title` character string. Text for the title of the plot. • `legend_title` character string. Text for the title of the legend. It is prioritized over titles based on the `color` column or parameter and the `group` column. • `frame` data frame. Sets the frame of the map.

Details

This function's purpose is to allow users to provide own data frames or edited ones exported by this package.

The data frame must have *at least* two columns called `latitude` & `longtitude`.

Data frames output by the function `top()` consist of, among others, a `latitude`, `longitude`, `country code` and `group` column.

If the input data frame has a column `color`, the function will assign every value in that column to the respective coordinates and ignore the additional parameter `color` (see above).

If the input data frame has a column `group`, the function will group data and display a legend.

If the input data frame has a `color` and a `group` column, the assignment must match each other. Every `group` (every unique string in that column) must be assigned a unique color throughout the data frame.

If `regions` is set to a value higher than `0`, the data frame must have a column `country code`.

Parameter `frame` accepts data frames containing coordinates which define the frame. The data frame must to have two columns called `lats` & `lons`. The latitudinal and longitudinal ranges define the frame.

Value

A plot.

---

ortho *Orthographical symbols*

Description

This function retrieves all symbols used in country data.

Usage

```
ortho(countries, ...)
```

Arguments

`countries`	character string vector with country designations (names or ISO-codes).
`...`	Additional parameter: • `column` character string. Selects the column for query.

Details

Parameter `countries` accepts all designations found in `country(query = "country table")`.

The default column is `"alternatenames"`. Other columns of possible interest are `"name"` and `"asciiname"`. It outputs an ordered frequency table of all symbols used in a given column of the GeoNames data for one or more countries specified.

Value

A table with frequencies of all symbols.

Examples

```
## Not run:
ortho(countries = "ID")
# outputs a table with frequencies all symbols
# in the "alternatenames" column for the Indonesia data set

## End(Not run)
```

`top` *Toponym map*

Description

This function returns and plots selected toponyms onto a map.

Usage

```
top(strings, countries, ...)
```

Arguments

`strings`	character string vector with regular expressions to filter data.
`countries`	character string vector with country designations (names or ISO-codes).
`...`	Additional parameters:

- `color` character string vector indicating, which color is assigned to each string.
- `regions` numeric. Specifies the level of administrative borders. By default `0` for displaying only country borders.
- `csv` logical. If `TRUE`, matches will be saved as .csv in the current working directory.
- `tsv` logical. If `TRUE`, matches will be saved as .tsv in the current working directory.
- `plot` logical. If `FALSE`, the plot will not be printed but saved as .png in the current working directory.
- `feat.class` character string vector. Selects data only of those feature classes (check `http://download.geonames.org/export/dump/readme.txt` for the list of all feature classes). By default, it is `P`.
- `polygon` data frame. Selects toponyms only inside the polygon.
- `name` character string. Defines name of output data frame.
- `column` character string vector. Selects the column(s) for query.
- `frame` data frame. Sets the frame of the map.

Details

This function is used to plot all locations matching the regular expression from `strings`. Parameter `countries` accepts all designations found in `country(query = "country table")`. Polygons passed through the `polygon` parameter need to intersect or be within a country specified by the `countries` parameter. Parameter `frame` accepts data frames containing coordinates which define the frame. The data frame must have two columns called `lats` & `lons`. The latitudinal and longitudinal ranges define the frame.

This function calls the internal `simpleMap()` function to generate a map of all locations gotten by `getCoordinates()`. The plot displays additional information if used by `topCompOut()`. The data used is downloaded by `getData()` and is accessible on the GeoNames download server.

Value

A plot of selected toponym(s) with the number of occurrences.

Examples

```
## Not run:
top("itz$", "DE")
# prints a plot with all populated places
# in Germany ending in "itz"
# and saves the locations in a data frame in the global environment.


top("^Vlad", "RU", color = "green", csv = TRUE, plot = FALSE)
# saves a plot with all populated places
# in Russia starting with "Vlad" (case sensitive) colored in green
# and saves it as .png together with the matches as .csv in the working directory.


top(c("itz$", "ice$"), c("DE", "PL"))
# prints a plot with all populated places in Germany and Poland ending in either "itz" or "ice"
# colored in red ("itz") and cyan ("ice")
# and saves matches in the global environment.

## End(Not run)
```

---

topComp *Compares toponyms in a polygon and the remainder of countries*

Description

This function retrieves the most frequent toponym substrings in a given polygon relative to country frequencies.

Usage

```
topComp(countries, len, rat, polygon, ...)
```

Arguments

countries	character string vector with country designations (names or ISO-codes).
len	numeric. The length of the substring within toponyms.
rat	numeric. The cut-off ratio (a number between 0.0 and 1 for freq.type = "abs") of how many occurrences of a toponym string need to be in the polygon relative to the rest of the country (or countries).
polygon	data frame. Defines the polygon for comparison with the remainder of a country (or countries).
...	Additional parameters:

- type character string. Either by default "$" (ending) or "^" (beginning).
- feat.class character string vector. Selects data only of those feature classes (check http://download.geonames.org/export/dump/readme.txt for the list of all feature classes). By default, it is P.
- freq.type character string. If "abs" (the default), ratios of absolute frequencies inside the polygon and in the countries as a whole are computed. If "rel", ratios of relative frequencies inside the polygon and outside the polygon will be computed.
- limit numeric. The number of the most frequent toponym substrings which will be tested.

Details

This function sorts the toponym substrings in the given countries by frequency. It then tests which ones lie in the given polygon and prints out a data frame with those that match the ratio criterion. Parameter countries accepts all designations found in country(query = "country table"). Polygons passed through the polygon parameter need to intersect or be within a country specified by the countries parameter.

Value

A data frame printed out and saved in the global environment. It shows toponym substrings surpassing the ratio, the ratio and the frequency.

Examples

```
## Not run:
topComp("GB",
  limit = 100,
   len = 4,
    rat = .7,
  polygon = toponym::danelaw_polygon
)
## prints and saves a data frame of the top 100 four-character-long endings in the United Kingdom
## if more than 70% of them belong to the polygon
## corresponding to the Danelaw area.


topComp("GB",
```

```
  limit = 100,
  len = 3,
  rat = 1,
  polygon = toponym::danelaw_polygon,
  freq.type = "rel"
)
## prints and saves a data frame of the top 100 three-character-long endings in the United Kingdom
## if they have greater relative frequencies within Danelaw than outside of Danelaw.


topComp(c("BE", "NL"),
  limit = 50,
  len = 3,
  rat = .8,
  polygon = toponym::flanders_polygon
)

## prints and saves a data frame of the top 50 three-character-long endings
## in Belgium and Netherlands viewed as a unit if more than 80% of them belong to the polygon
## corresponding to Flanders.

.

## End(Not run)
```

`topCompOut` *Saves multiple maps and toponym data*

Description

This function retrieves the most frequent toponym substrings in a given polygon relative to country frequencies. It generates maps of them and saves them along with the corresponding data frames.

Usage

```
topCompOut(countries, len, rat, polygon, ...)
```

Arguments

`countries`	character string vector with country designations (names or ISO-codes).
`len`	numeric. The length of the substring within toponyms.
`rat`	numeric. The cut-off ratio (a number between 0.0 and 1 for `freq.type = "abs"`) of how many occurrences of a toponym string need to be in the polygon relative to the rest of the country (or countries).
`polygon`	data frame. Defines the polygon for comparison with the remainder of a country (or countries).
`...`	Additional parameters:

- df logical. If TRUE, the filtered data frames will be saved in the global environment.
- csv logical. If TRUE, the filtered data frames will be saved as .csv in the current working directory.
- tsv logical. If TRUE, the filtered data frames will be saved as .tsv in the current working directory.
- type character string. Either by default "$" (ending) or "^" (beginning).
- feat.class character string vector. Selects data only of those feature classes (check http://download.geonames.org/export/dump/readme.txt for the list of all feature classes). By default, it is P.
- freq.type character string. If "abs" (the default), ratios of absolute frequencies inside the polygon and in the countries as a whole are computed. If "rel", ratios of relative frequencies inside the polygon and outside the polygon will be computed.
- limit numeric. The number of the most frequent toponym substrings which will be tested.

Details

This function applies the list of toponyms returned by topComp() to top(). A series of maps showing the toponym, ratio in percentage and numbers will be generated and locally saved. Parameter countries accepts all designations found in country(query = "country table"). Polygons passed through the polygon parameter need to intersect or be within a country specified by the countries parameter.

Value

Data frames and plots saved in a sub folder (called 'dataframes' and 'plots') in the working directory or global environment.

Examples

```
## Not run:
topCompOut(
  countries = "BE",
   limit = 10,
   len = 3,
   rat = .95,
   df = FALSE,
   polygon = toponym::flanders_polygon
   )

## generates and saves the data frames & maps of the top 10 three-character-long endings
## in Belgium if more than 95% of of them belong to the polygon
## corresponding to Flanders.
## End(Not run)
```

topFreq *Retrieves the most frequent toponyms*

Description

This function returns the most frequent toponym substrings in countries or a polygon.

Usage

```
topFreq(countries, len, limit, ...)
```

Arguments

countries	character string vector with country designations (names or ISO-codes).
len	numeric. The length of the substring within toponyms.
limit	numeric. The number of the most frequent toponym substrings.
...	Additional parameters: • type character string. Either by default "$" (ending) or "^" (beginning). • feat.class character string vector. Selects data only of those feature classes (check http://download.geonames.org/export/dump/readme.txt for the list of all feature classes). By default, it is P. • polygon data frame. Selects toponyms only inside the polygon.

Details

Parameter countries accepts all designations found in country(query = "country table").

Polygons passed through the polygon parameter need to intersect or be within a country specified by the countries parameter.

Value

A table with toponym substrings and their frequency.

Examples

```
## Not run:
topFreq(countries = "Namibia", len = 3, limit = 10)
## returns the top 10 most frequent toponym endings
## of three-character length in Namibia

topFreq(
  countries = "GB", len = 3, limit = 10,
  polygon = toponym::danelaw_polygon
)
## returns the top 10 most frequent toponym endings
## in the polygon which is inside the United Kingdom.

## End(Not run)
```

toponymOptions *Manage Options of* toponym

Description

This function allows users to modify settings for managing toponym data. Users can choose whether to save matches and strings in the global environment or not to save them. Further, users can specify whether toponym data retrieved from GeoNames will be saved in the package folder or in a temporary folder.

Usage

```
toponymOptions(global = NULL, save_data = NULL)
```

Arguments

global	logical. Enter TRUE or FALSE. Allows the user to modify the setting for storing objects in the global environment.
save_data	logical. Enter TRUE or FALSE. Allows the user to modify the setting for saving toponym data sets in the package folder or in a temporary folder.

Details

Parameter global: if the current setting is TRUE, matches from top() and strings from topComp() will be saved in the global environment; if the current setting is FALSE, the results will not be saved.

Parameter save_data: if the current setting is TRUE, toponym data sets will be saved in the package folder; if the current setting is FALSE, toponym data sets will be saved in a temporary folder.

If no parameter is set, i.e. toponymOptions(), the complete data frame with current settings is printed.

Value

A data frame with the value(s) of the respective setting(s).

Examples

```
# Show the current settings
toponymOptions()
```

topZtest *Applies Z-test*

Description

This function applies a Z-test.

Usage

```
topZtest(strings, countries, polygon, ...)
```

Arguments

strings	character string with a regular expression to be tested.
countries	character string vector with country designations (names or ISO-codes).
polygon	data frame. Defines the polygon for comparison with the remainder of a country (or countries).
...	Additional parameter: • feat.class character string vector. Selects data only of those feature classes (check http://download.geonames.org/export/dump/readme.txt for the list of all feature classes). By default, it is P.

Details

This function lets users apply a Z-test (two proportion test), comparing the frequency of a given string in a polygon to the frequency in the rest of the country. Parameter countries accepts all designations found in country(query = "country table"). Polygons passed through the polygon parameter need to intersect or be within a country specified by the countries parameter.

Value

An object of class htest containing the results.

Index

country, *2*, 3
createPolygon, *2*, 4

getData, 5

mapper, *2*, 6

ortho, 7

top, *2*, 8
topComp, *2*, 9
topCompOut, *2*, 11
topFreq, *2*, 13
toponym *(*toponym-package*)*, 2
toponym-package, 2
toponymOptions, *2*, 14
topZtest, *2*, 15