

# Package ‘toponym’

May 8, 2026

**Title** Analyze and Visualize Toponyms

**Version** 2.0.1

**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 information on the study of toponyms, see Wichmann & Chevallier (2025) <[doi:10.5195/names.2025.2616](https://doi.org/10.5195/names.2025.2616)>.

**Depends** R (>= 4.1)

**License** GPL (>= 3)

**Encoding** UTF-8

**RoxygenNote** 7.3.3

**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>

**NeedsCompilation** no

**Author** Lennart Chevallier [aut, cre] (ORCID:  
<<https://orcid.org/0009-0009-6800-1206>>),  
Søren Wichmann [aut] (ORCID: <<https://orcid.org/0000-0002-3257-3087>>)

**Maintainer** Lennart Chevallier <[mail@lchevallier.de](mailto:mail@lchevallier.de)>

**Repository** CRAN

**Date/Publication** 2026-04-13 14:40:02 UTC

## Contents

country	2
createPolygon	3
getData	5
mapper	6
ortho	8
top	9
topComp	10
topFreq	12
toponymOptions	13
topZtest	15

<b>Index</b>	<b>17</b>
--------------	-----------

---

country	<i>Country designations</i>
---------	-----------------------------

---

### 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: <ul style="list-style-type: none"> <li>regions numeric. If 1, outputs the region designations of the respective countries. By default, it is 0.</li> <li>toponym_path character string. Path name for downloaded data.</li> </ul>

### Details

If you enter an individual country designation, you receive the three different designations (ISO2, 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. If regions is set to value higher than 0, map data by the geodata is required to be downloaded. With toponymOptions(), users can specify the path for toponym and map data downloaded by this package across sessions. See help(toponymOptions).

**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**

```
country(query = "IS03")
## returns a vector of all IS03 codes

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

## We recommend setting a persistent path for downloaded data by using toponymOptions()
## Users can always set the path manually when a function is used
## For illustration purposes,
## 1. the path is manually set each time
## 2. and wrapped in donttest because data will be downloaded in the following example:

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

---

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: <ul style="list-style-type: none"> <li>regions numeric. Specifies the level of administrative borders. By default 0 for displaying only country borders.</li> <li>region_ID character string vector with region IDs.</li> <li>region_name character string vector with region names.</li> <li>retrieve logical. If TRUE, the coordinates of the region or country are returned. No map will be drawn.</li> <li>toponym_path character string. Path name for downloaded data.</li> </ul>

## 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 map data is incomplete in this regard. For mapping purposes, geodata is used throughout this package.

Parameter `toponym_path` accepts `"pkgdir"` for the package directory or a full, alternative path. With `toponymOptions()`, users can specify the path for toponym and map data downloaded by this package across sessions. See `help(toponymOptions)`.

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

```
## We recommend setting a persistent path for downloaded data by using toponymOptions()
## Users can always set the path manually when a function is used
## For illustration purposes in the following examples,
## 1. the path is manually set each time
## 2. and wrapped in donttest because data will be downloaded
## 3. or if(interactive) because it is interactive:
if(interactive()){
  createPolygon("NA", region_ID = "NAM.7_1", toponym_path = tempdir())
  # a plot of the region Ohangwena in Namibia appears.
  # by point-and-click a polygon can be created
  # upon completion, a data frame with the coordinates of the polygon returns
}
```

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

---

getData	<i>Downloads GeoNames data</i>
---------	--------------------------------

---

### Description

This function downloads toponym data for the package.

### Usage

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

### 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.
toponym_path	character string. Path name for downloaded data. If not specified, this function will call <code>toponymOptions()</code> and try use the persistent path.

### 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")`. If "all", data from all countries stored in the package folder is selected. Parameter `toponym_path` accepts "pkgdir" for the package directory or a full, alternative path. With `toponymOptions()`, users can specify the path for toponym data downloaded by `getData()` across sessions. See `help(toponymOptions)`.

### Value

No return value.

### See Also

[GeoNames download page](#)

### Examples

```
## We recommend setting a persistent path for downloaded data by using toponymOptions()
## Users can always set the path manually when a function is used
## For illustration purposes,
## 1. the path is manually set each time
## 2. and wrapped in donttest because data will be downloaded in the following examples:

getData(countries = "NL", toponym_path = tempdir())
## downloads and extracts data for NL to the temporary directory
```

```
getData(countries = c("DK", "DE"), toponym_path = tempdir())
## downloads and extracts data for DK and DE to the temporary directory
```

```
getData(countries = c("DK", "DE"), overwrite = TRUE, toponym_path = tempdir())
## downloads, extracts, and overwrites data for DK and DE in the temporary directory
```

---

 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: <ul style="list-style-type: none"> <li>• <code>color</code> character string vector indicating, which color is assigned to each string. It is prioritized over colors based on the column <code>color</code>.</li> <li>• <code>regions</code> numeric. Specifies the level of administrative borders. By default 0 for displaying only country borders.</li> <li>• <code>title</code> character string. Text for the title of the plot.</li> <li>• <code>legend_title</code> character string. Text for the title of the legend. It is prioritized over titles based on column group.</li> <li>• <code>show_legend</code> logical. If TRUE, a legend with all unique strings in the column group will be displayed, provided there is a column group. If FALSE, no legend will be displayed. By default, TRUE.</li> <li>• <code>frame</code> data frame. Sets the frame of the map.</li> <li>• <code>plot_size</code> numeric. Specifies the value by which the size of the map is scaled.</li> <li>• <code>toponym_path</code> character string. Path name for downloaded data.</li> </ul>

## Details

This function's purpose is to allow users to provide data frames by the function `top()`, edited ones as well as own data frames.

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

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. However, if specified, the additional parameter `color` will be used instead of the column `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 greater than 0, the data frame must have a column `country code`.

Parameter `frame` accepts data frames containing coordinates which define the frame of the plot. The data frame must have a column called `latitude` & a column called `longitude`. The latitudinal and longitudinal ranges define the frame of the plot.

Parameter `plot_size` accepts numeric values of greater than -1. The plot's size is scaled by the given value. Thus, a value of 0 extends the size by 0%. A value of .1 extends the size by 10%. A value of -.1 reduces the size by 10% and so on.

Parameter `toponym_path` accepts "pkgdir" for the package directory or a full, alternative path. With `toponymOptions()`, users can specify the path for toponym and map data downloaded by this package across sessions. See `help(toponymOptions)`.

## Value

A plot.

## Examples

```
## We recommend setting a persistent path for downloaded data by using toponymOptions()
## Users can always set the path manually when a function is used
## For illustration purposes,
## 1. the path is manually set each time
## 2. and wrapped in donttest because data will be downloaded in the following examples:
```

```
mapper(
  top("itz$", "DE", toponym_path = tempdir()),
  toponym_path = tempdir())
# returns a plot with all populated places
# in Germany ending in "itz"
```

```
UG_data <- top(c("et$", "wa$"), "UG", toponym_path = tempdir())
UG_data$color <- "blue"
UG_data[UG_data$group == "wa", "color"] <- "grey"
mapper(UG_data,
  legend_title = "two strings",
  title = "Some locations in grey and blue",
  toponym_path = tempdir())
# returns a plot with all populated places
# in Uganda ending in "wa" (grey) and "et" (blue)
# the plot is titled "Some locations in grey and blue"
```

```
# the legend title is "two strings"
```

---

```
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.
- `toponym_path` character string. Path name for downloaded data.

## 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 `"asciname"`. It outputs an ordered frequency table of all symbols used in a given column of the GeoNames data for one or more countries specified.

Parameter `toponym_path` accepts `"pkgdir"` for the package directory or a full, alternative path. With `toponymOptions()`, users can specify the path for toponym and map data downloaded by this package across sessions. See `help(toponymOptions)`. The data used is downloaded by `getData()` and is accessible on the [GeoNames download server](#).

## Value

A table with frequencies of all symbols.

## Examples

```
## We recommend setting a persistent path for downloaded data by using toponymOptions()
## Users can always set the path manually when a function is used
## For illustration purposes,
## 1. the path is manually set each time
## 2. and wrapped in donttest because data will be downloaded in the following example:

ortho(countries = "MC", toponym_path = tempdir())
# returns a table with frequencies of all symbols
# in the "alternatenames" column for the Monaco data set
```

---

top	<i>Selection of Toponyms</i>
-----	------------------------------

---

**Description**

This function returns coordinates of selected toponyms (strings).

**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: <ul style="list-style-type: none"><li>• feat.class character string vector. Selects data only of those feature classes (check <a href="http://download.geonames.org/export/dump/readme.txt">http://download.geonames.org/export/dump/readme.txt</a> for the list of all feature classes). By default, it is P.</li><li>• polygon data frame. Selects toponyms only inside the polygon.</li><li>• column character string vector. Selects the column(s) for query.</li><li>• toponym_path character string. Path name for downloaded data.</li></ul>

**Details**

This function selects locations which match 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 toponym\_path accepts "pkgdir" for the package directory or a full, alternative path. With toponymOptions(), users can specify the path for toponym and map data downloaded by this package across sessions. See help(toponymOptions). The data used is downloaded by getData() and is accessible on the [GeoNames download server](#).

**Value**

A data frame of selected toponym(s).

**Examples**

```
## We recommend setting a persistent path for downloaded data by using toponymOptions()
## Users can always set the path manually when a function is used
## For illustration purposes,
## 1. the path is manually set each time
## 2. and wrapped in donttest because data will be downloaded in the following examples:

itz_data <- top("itz$", "DE", toponym_path = tempdir())
# returns a data frame with all populated places
```

```

# in Germany ending in "itz"

vlad_data <- top("^Vlad", "RU", toponym_path = tempdir())
# returns a data frame with all populated places
# in Russia starting with "Vlad" (case sensitive)

itz_ice_data <- top(c("itz$", "ice$"), c("DE", "PL"), toponym_path = tempdir())
# returns a data frame with all populated places
# in Germany and Poland ending in either "itz" or "ice"

maw_data <- top("Maw$", "MM", column = "alternatenames", toponym_path = tempdir())
# returns a data frame with all populated places
# in Myanmar listed in the "alternatenames" column
# and ending in "Maw" (case sensitive)

```

---

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: <ul style="list-style-type: none"> <li>• type character string. Either by default "\$" (ending) or "^" (beginning).</li> <li>• feat.class character string vector. Selects data only of those feature classes (check <a href="http://download.geonames.org/export/dump/readme.txt">http://download.geonames.org/export/dump/readme.txt</a> for the list of all feature classes). By default, it is P.</li> </ul>

- `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.
- `toponym_path` character string. Path name for downloaded data.

## 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. Parameter `toponym_path` accepts "pkgdir" for the package directory or a full, alternative path. With `toponymOptions()`, users can specify the path for toponym and map data downloaded by this package across sessions. See `help(toponymOptions)`. The data used is downloaded by `getData()` and is accessible on the [GeoNames download server](#).

## 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

```
## We recommend setting a persistent path for downloaded data by using toponymOptions()
## Users can always set the path manually when a function is used
## For illustration purposes,
## 1. the path is manually set each time
## 2. and wrapped in donttest because data will be downloaded in the following examples:

topComp("GB",
  limit = 100,
  len = 4,
  rat = .7,
  polygon = toponym::danelaw_polygon,
  toponym_path = tempdir()
)
## returns 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",
```

```

    toponym_path = tempdir()
  )
  ## returns 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,
  toponym_path = tempdir()
)
## returns 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.

```

---

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: <ul style="list-style-type: none"> <li>• type character string. Either by default "\$" (ending) or "^" (beginning).</li> <li>• feat.class character string vector. Selects data only of those feature classes (check <a href="http://download.geonames.org/export/dump/readme.txt">http://download.geonames.org/export/dump/readme.txt</a> for the list of all feature classes). By default, it is P.</li> <li>• polygon data frame. Selects toponyms only inside the polygon.</li> <li>• toponym_path character string. Path name for downloaded data.</li> </ul>

**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. Parameter `toponym_path` accepts `"pkgdir"` for the package directory or a full, alternative path. With `toponymOptions()`, users can specify the path for toponym and map data downloaded by this package across sessions. See `help(toponymOptions)`. The data used is downloaded by `getData()` and is accessible on the [GeoNames download server](#).

**Value**

A table with toponym substrings and their frequency.

**Examples**

```
## We recommend setting a persistent path for downloaded data by using toponymOptions()
## Users can always set the path manually when a function is used
## For illustration purposes,
## 1. the path is manually set each time
## 2. and wrapped in donttest because data will be downloaded in the following examples:
```

```
topFreq(
  countries = "Ecuador",
  len = 3,
  limit = 10,
  toponym_path = tempdir())
## returns the top 10 most frequent toponym endings
## of three-character length in Ecuador
```

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

---

toponymOptions

*Manage Options of toponym*


---

**Description**

This function allows users to set the download path for the toponym package. Downloaded data includes toponym data and map data.

## Usage

```
toponymOptions(toponym_path = NULL)
```

## Arguments

`toponym_path` character string. Path name for downloaded data. This setting is saved across sessions.

## Details

Most functions require external data which will be downloaded and stored for later use. This is described in the respective functions. For this reason, after installation, users will be asked to specify the path for downloaded data. Parameter `toponym_path` accepts either the character string `"pkgdir"` or a full, alternative path. `"pkgdir"` is interpreted as the `extdata` folder in the toponym package directory, i.e.: `system.file("extdata", package = "toponym")`

Thus, users can set the path to the package directory with this command:

```
toponymOptions(toponym_path = "pkgdir")
```

If a path is provided, users are prompted to confirm their choice. This function will write the new path into `toponym_options.rds` in the package directory; the path is saved across sessions.

To locate `toponym_options.rds`, enter:

```
system.file("extdata", package = "toponym")
```

To check the path that is currently set, enter:

```
toponymOptions()
```

## Value

A character string indicating the current path for downloaded data.

## Examples

```
if(interactive()){  
  # Set the path to the temporary directory  
  # Users are prompted to confirm their choice.  
  # Upon confirmation, toponym_options.rds will be edited in the package directory  
  toponymOptions(toponym_path = tempdir())  
  # Show the current path  
  toponymOptions()  
}
```

---

topZtest	<i>Applies Z-test</i>
----------	-----------------------

---

### 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: <ul style="list-style-type: none"> <li>• <code>feat.class</code> character string vector. Selects data only of those feature classes (check <a href="http://download.geonames.org/export/dump/readme.txt">http://download.geonames.org/export/dump/readme.txt</a> for the list of all feature classes). By default, it is P.</li> <li>• <code>toponym_path</code> character string. Path name for downloaded data.</li> </ul>

### 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. Parameter `toponym_path` accepts `"pkgdir"` for the package directory or a full, alternative path. With `toponymOptions()`, users can specify the path for toponym and map data downloaded by this package across sessions. See `help(toponymOptions)`. The data used is downloaded by `getData()` and is accessible on the [GeoNames download server](#).

### Value

An object of class `htest` containing the results.

### Examples

```
## We recommend setting a persistent path for downloaded data by using toponymOptions()
## Users can always set the path manually when a function is used
## For illustration purposes,
## 1. the path is manually set each time
## 2. and wrapped in donttest because data will be downloaded in the following example:

topZtest("thorpe$",
```

```
    "GB",  
    toponym::danelaw_polygon,  
    toponym_path = tempdir()  
## returns an object of class htest containing the results.
```

# Index

country, [2](#)  
createPolygon, [3](#)  
  
getData, [5](#)  
  
mapper, [6](#)  
  
ortho, [8](#)  
  
top, [9](#)  
topComp, [10](#)  
topFreq, [12](#)  
toponymOptions, [13](#)  
topZtest, [15](#)