

Package ‘geocleaMT’

January 12, 2016

Version 1.0160110

Date 2016-01-10

Title A package to clean/test georeferenced data

Author Viviana R-Alarcon and Daniel R. Miranda-Esquivel

Maintainer Daniel R. Miranda-Esquivel <dmiranda@uis.edu.co>

Depends R (>= 3.2.1)

Import rgbif, RJSONIO, RCurl, maptools, plyr, vegan, modeest

Description GeoCleaMT is a cleaning protocol for distributional data and an automated tool to provide the necessary data mining process before any diversity analysis. The protocol takes into account the most common mistakes found in distributional records and attempts to minimize its possibility, increasing users confidence in their data. All functions use as input a basic table with at least three columns species / decimalLatitude/ decimalLongitude, and save the results in the same format after a specific task is performed. Additionally, all functions return a descriptive report about the process made, that can be also saved as data frame object. Although, the functions could be used in the suggested pipelines, each function is suited to perform an specific task or a set of them, given the package a high level of customizability. See R-Alarcon and Miranda-Esquivel (in prep).

License GPL-2

URL <https://github.com/alarconvv/geocleaMT>

BugReports <https://github.com/alarconvv/geocleaMT/issues>

RoxygenNote 5.0.1

R topics documented:

Africa	2
Altitudes	3
America	4
Antarctica	5
Asia	5

assignElevation	6
ChcBiog	8
checkCoord	8
coordToCountry	10
coordToISO	11
countryNames	12
countryToCentroid	12
countryToISO	13
cutRange	14
delPointsOrSp	15
elevFromGg	17
Europa	18
Example	19
gbifDownSp	20
geocleaMT	21
IBISlist	22
ID_DarwinCore	23
invasiveSp	23
ISO2	25
ISO3	26
ISOTocountry	26
meanPropinquity	27
Oceania	29
pathStructure	29
pointsAtSea	30
readAndWrite	31
readDbBash	33
readDbR	34
splitGeoref	36
spOutPoly	37
stackSp	39
usefulSp	40
Index	42

Africa	<i>Vectorial layer: African continent</i>
--------	-------------------------------------------

Description

Map of the African continent. Shapefile of countries borders. This map was extracted from data(wrld_simpl) from maptools package. The 'Africa' object uses the WGS84 datum.

Usage

data(Africa)

Format

Format class 'SpatialPolygonsDataFrame' [package 'sp'].

Source

<http://r-forge.r-project.org/projects/maptools/>

References

[1] Bivand et al. (2015) maptools: Tools for Reading and Handling Spatial Objects. CRAN. Version 0.8-37. Available online at <http://r-forge.r-project.org/projects/maptools/>

Examples

```
data(Africa)
plot(Africa)
```

Altitudes

Altitudes

Description

Elevation data for the North Andes Block with a resolution of 0.1 decimal degrees. Three columns decimalLongitude, decimalLatitude and elevation. The North Andes block was divided each 0.1 decimal degree, the elevations associated to each 0.1 decimal degree were extracted using the [elevFromGg](#) function. This elevation was downloaded from The Google Maps Elevation API [1] and was used for the example in R-Alrcon and Miranda-Esquivel (2016) [2].

Usage

```
data(Altitudes)
```

Format

A data frame with 28063 observations on the following 3 variables: decimalLatitude, decimalLongitude and elevation.

References

- [1] R-Alarcon V. and Miranda-Esquivel DR.(in prep) geocleaMT: An R package to cleaning geographical data from electronic biodatabases.
- [2] The Google Maps Elevation API. Available online at <https://developers.google.com/maps/documentation/elevation/intro>

Examples

```
data(Altitudes)
str(Altitudes)
data(America)
coordinates(Altitudes) <- Altitudes[, c('decimalLongitude', 'decimalLatitude')]
plot(America)
plot(Altitudes, add=T)
```

America

Vectorial layer: American continent

Description

Map of the American continent. Shapefile of countries borders. This map was extracted from `data(wrld_simpl)` from `maptools` package [1]. The 'America' object uses the WGS84 datum.

Usage

```
data(America)
```

Format

Format class 'SpatialPolygonsDataFrame' [package 'sp'].

Source

<http://r-forge.r-project.org/projects/maptools/>

References

[1] Bivand et al. (2015) `maptools`: Tools for Reading and Handling Spatial Objects. CRAN. Version 0.8-37. Available online at <http://r-forge.r-project.org/projects/maptools/>

Examples

```
data(America)
plot(America)
```

Antarctica

Vectorial layer: Antarctic continent

Description

Map of the Antarctic continent. Shapefile of countries borders. This map was extracted from `data(wrld_simpl)` from `maptools` package [1]. The 'Antarctica' object uses the WGS84 datum.

Usage

```
data(Antarctica)
```

Format

Format class 'SpatialPolygonsDataFrame' [package 'sp'].

Source

<http://r-forge.r-project.org/projects/maptools/>

References

#' [1] Bivand et al. (2015) `maptools`: Tools for Reading and Handling Spatial Objects. CRAN. Version 0.8-37. Available online at <http://r-forge.r-project.org/projects/maptools/>

Examples

```
data(Antarctica)
plot(Antarctica)
```

Asia

Vectorial layer: Asian continent

Description

Map of the Asian continent. Shapefile of countries borders. This map was extracted from `data(wrld_simpl)` from `maptools` package [1]. The 'Asia' object uses the WGS84 datum.

Usage

```
data(Asia)
```

Format

Format class 'SpatialPolygonsDataFrame' [package 'sp'].

Source

<http://r-forge.r-project.org/projects/maptools/>

References

[1] Bivand et al. (2015) maptools: Tools for Reading and Handling Spatial Objects. CRAN. Version 0.8-37. Available online at <http://r-forge.r-project.org/projects/maptools/>

Examples

```
data(Asia)
plot(Asia)
```

assignElevation	<i>Assigning elevation data</i>
-----------------	---------------------------------

Description

Given a geographical coordinate (decimalLongitude/ decimalLatitude) the elevation data is assigned from an elevation database (Headers: decimalLongitude/ decimalLatitude/ elevation).

Usage

```
assignElevation(data = NULL, path.data = NULL, rd.frmt = "readRDS",
  elevations.db = NULL, round.coord = 1, wrt.frmt = "saveRDS",
  save.assigned.in = NULL, save.unassigned.in = NULL)
```

Arguments

data	Vector of characters. Name of the input file.
path.data	Vector of characters. Path to the input file.
rd.frmt	Vector of characters. The file format to read. By default it will be read as a R object using 'readRDS' argument, but it can be read as plain text using 'readTXT' argument. See details.
elevations.db	Data.Frame object. The elevations database, tree columns (decimalLatitude, decimalLongitude, elevation)
round.coord	Numeric vector. Decimal to round the coordinate to assign the elevation. This value must be same of the coordinate resolution of database assigned as argument on elevation.db parameter. Example: the resolution of the coordinates in the databases assigned on elevation.db is one decimal, then round.coord must be equal to 1 (round.coord=1).
wrt.frmt	Vector of characters. Format to save output file. By default it will be written as a R object using 'saveRDS' argument, but it can be saved as plain text using 'saveTXT' argument. See details.

save.assigned.in

Vector of characters. Path to write the output file with only the coordinates whose elevation was assigned. Output file will be called <alt.assigned>.

save.unassigned.in

Vector of characters. Path to write the output file with only the coordinates whose elevation was not assigned. Output file will be called <alt.unassigned>. See details.

Details

For more details about the formats to read and/or write, see the [readAndWrite](#) function.

The coordinates whose elevations were not assigned could be checked with [elevFromGg](#). This function can download data directly from Google Maps Elevation API [1]

The headers of the input file must follow the Darwin Core standard [2]. The user can see the guide using `data('ID_DarwinCore')` command.

Value

A table data.frame class with coordinates and their respective elevation. A table with coordinates whose elevations were not assigned. A table of information.

Note

See: R-Alarcon V. and Miranda-Esquivel DR.(in prep) geocleaMT: An R package to cleaning geographical data from electronic biodatabases.

Author(s)

R-Alarcon Viviana and Miranda-Esquivel Daniel R.

References

[1] The Google Maps Elevation API. Available online at <https://developers.google.com/maps/documentation/elevation/intro>

[2] Wiczorek, J. et al. 2012. Darwin core: An evolving community-developed biodiversity data standard. PloS One 7: e29715.

See Also

[elevFromGg](#)

[readAndWrite](#)

ChcBiog

Biogeographic Choco (Sensu lato: Hernandez 1992) [1]

Description

Polygon of the Biogeographic Choco (Sensu lato: Hernandez 1992)[1] using as example in R-Alrcon and Miranda-Esquivel (In prep)[2]. The 'ChcBiog' object uses the WGS84 datum.

Usage

```
data(ChcBiog)
```

Format

Format class 'SpatialPolygonsDataFrame' [package 'sp'] with 5 slots: data, polygons, plotOrder, bbox, proj4string, keywords.

References

[1] Hernandez-Camacho, J. et al. 1992. Unidades biogeograficas de Colombia, In: Halffter, G. (ed), La Diversidad Biologica de Iberoamerica I. Acta Zoologica Mexicana. Instituto de Ecologia, A.C., pp. 105-151 [2] R-Alarcon V. and Miranda-Esquivel DR.(in prep) geocleaMT: An R package to cleaning geographical data from electronic biodatabases.

Examples

```
data(ChcBiog)
plot(ChcBiog)
```

checkCoord

Check coordinates

Description

Separate coordinates with a format different to decimal degrees and check that the coordinates in this format are into range (latitude:-90, 90 degrees and longitude: -180, 180 degrees)

Usage

```
checkCoord(data = NULL, path = NULL, rd.frmt = "readRDS",
  wrt.frmt = "saveRDS", save.right.in = NULL, save.wrong.in = NULL)
```


Arguments

data	Vector of characters. Name of the input file.
path	Vector of characters. Path to the input file.
rd.frmt	Vector of characters. The file format to read. By default it will be read as a R object using 'readRDS' argument, but it can be read as plain text using 'readTXT' argument. See details.
wrt.frmt	Vector of characters. Format to save output file. By default it will be written as a R object using 'saveRDS' argument, but it can be saved as plain text using 'saveTXT' argument. See details.
save.right.in	Vector of characters. Path to the output file for each species with coordinates in the correct format.
save.wrong.in	Vector of characters. Path to the output file for each species with coordinates in the wrong format.

Details

For more details about the formats to read and/or write, see [readAndWrite](#) function.

The headers of the input file must follow the Darwin Core standard [1]. The user can see the guide using `data('ID_DarwinCore')` command.

Value

For each species, if this is the case, two output files data.frame class: The first one has the correct coordinates for the species and the second one has the wrong coordinates for the same species.

Note

See: R-Alarcon V. and Miranda-Esquivel DR. (in prep.) geocleaMT: An R package to cleaning geographical data from electronic biodatabases.

Author(s)

R-Alarcon Viviana and Miranda-Esquivel Daniel R.

References

[1] Wiczorek, J. et al. 2012. Darwin core: An evolving community-developed biodiversity data standard. PloS One 7: e29715.

See Also

[readAndWrite](#)

coordToCountry	<i>From coordinate in decimal degree format to country name (English)</i>
----------------	---------------------------------------------------------------------------

Description

Give a coordinate in decimal degree format to assign the corresponding country name in English.

Usage

```
coordToCountry(coordinatesTable = NULL)
```

Arguments

`coordinatesTable`
Data.Frame class. A data frame with at least two columns called `decimalLatitude` and `'decimalLongitude'`.

Details

If the coordinate is outside the continents, this can not be assigned and will be assigned as `'NaN'`.

Value

A table with three columns (`decimalLatitude`, `decimalLongitude`, `country`)

Note

See: R-Alarcon V. and Miranda-Esquivel DR.(in prep) geocleaMT: An R package to cleaning geographical data from electronic biodatabases.

Author(s)

R-Alarcon Viviana and Miranda-Esquivel Daniel R.

See Also

[coordToISO](#)

coordToISO	<i>From coordinate in decimal degree format to International Organization for Standardization (ISO 3166-1 alpha-2/alpha-3)</i>
------------	--------------------------------------------------------------------------------------------------------------------------------

Description

Give a coordinate in decimal degree format to assign the corresponding International Organization for Standardization (ISO 3166-1 alpha-2/alpha-3) [1].

Usage

```
coordToISO(coordinates.tab = NULL, iso = 2)
```

Arguments

coordinates.tab	Data.Frame class. A data frame with at least two columns called decimalLatitude and 'decimalLongitude'.
iso	numeric (2 or 3). To define the code whether ISO-2 (iso=2, default value) or ISO-3 (iso=3).

Details

If the coordinate is outside the continent, this can not be assigned and will be assigned as NaN.

Value

A table with three columns (decimalLatitude, decimalLongitude, countryCode)

Note

See: R-Alarcon V. and Miranda-Esquivel DR.(in prep) geocleaMT: An R package to cleaning geographical data from electronic biodatabases.

Author(s)

R-Alarcon Viviana and Miranda-Esquivel Daniel R.

References

[1]ISO 2015. International standard for country codes and codes for their subdivisions. Country codes - iso 3166. International Organization for Standardization. Available online at http://www.iso.org/iso/country_codes

See Also

[coordToCountry](#)

countryNames	<i>Countries Names</i>
--------------	------------------------

Description

List of Countries Names in English.

Usage

```
data(countryNames)
```

Format

A vector class. Factor with 246 levels (Countries Names).

References

R-Alarcon V. and Miranda-Esquivel DR. (in prep) geocleaMT: An R package to cleaning geographical data from electronic biodatabases.

Examples

```
data(countryNames)
```

countryToCentroid	<i>From country name to centroid</i>
-------------------	--------------------------------------

Description

From country name in English to the centroid (Coordinate) of the polygon that represents this country.

Usage

```
countryToCentroid(country = NULL)
```

Arguments

country	data.frame class. a vector with country names in English. See details.
---------	------------------------------------------------------------------------

Details

To see the valid names in English the user can run the command `data(countryNames)`.

Value

A data frame object with the centroid assigned to each country.

Note

See: R-Alarcon V, Miranda-Esquivel DR.(in prep) geocleaMT: An R package to cleaning geographical data from electronic biodatabases.

Author(s)

R-Alarcon Viviana and Miranda-Esquivel Daniel R.

See Also

[countryToISO](#)

countryToISO	<i>From country name to International Organization for Standardization (ISO 3166-1 alpha-2 / alpha-3)</i>
--------------	-----------------------------------------------------------------------------------------------------------

Description

From country name in English to International Organization for Standardization (ISO 3166-1 alpha-2 / alpha-3) [1].

Usage

```
countryToISO(country = NULL, iso = 2)
```

Arguments

country	Data.Frame class. a vector with countries names in English. See details.
iso	numeric (2 or 3). To define the code whether ISO-2 (iso=2, default value) or ISO-3 (iso=3).

Details

To see the valid names in English the user can use the command `data(countryNames)`.

Value

A data frame object with the ISO 3166-1 alpha-2/alpha-3 assigned to each country.

Note

See: R-Alarcon V, Miranda-Esquivel DR.(in prep) geocleaMT: An R package to cleaning geographical data from electronic biodatabases.

Author(s)

R-Alarcon Viviana and Miranda-Esquivel Daniel R.

References

[1]ISO 2015. International standard for country codes and codes for their subdivisions. Country codes - iso 3166. International Organization for Standardization. Available online at http://www.iso.org/iso/country_codes.

See Also

[countryToCentroid](#)

cutRange	<i>Species on a specific elevation range</i>
----------	----------------------------------------------

Description

Get species distributed on a specific range of altitude.

Usage

```
cutRange(data = NULL, path = NULL, rd.frmt = "readRDS", range.from = 0,
  range.to = 1000, wrt.frmt = "saveRDS", save.inside.in = NULL,
  save.outside.in = NULL)
```

Arguments

data	Vector of characters. Name of the input file.
path	Vector of characters. Path to the input file.
rd.frmt	Vector of characters. File format to read. By default it will be read as a R object using 'readRDS' argument, but it can be read as plain text using 'readTXT' argument. See details.
range.from	Numeric vector. Lower bound of the range of the distribution
range.to	Numeric vector. Upper bound of the range of the distribution.
wrt.frmt	Vector of characters. Format to save output file. By default it will be written as a R object using 'saveRDS' argument, but it can be saved as plain text using 'saveTXT' argument. See details.
save.inside.in	Vector of characters. Path to the output file with the species whose occurrences are on the range assigned in range.from to range.to parameters. Output file will be called <inside.range>.
save.outside.in	Vector of characters. Path to the output file with the species whose occurrences are not on the range assigned in range.from and range.to parameters. Output file will be called <outside.range>.

Details

For more details about the formats to read and/or write, see [readAndWrite](#) function.

The headers of the input file must follow the Darwin Core standard [1]. The user can see the guide using `data('ID_DarwinCore')` command.

Value

A table data.frame class with only the species whose occurrences is on range assigned in `range.from` and `range.to` parameters, and a vector with a table with descriptive quantities.

Note

See: R-Alarcon V. and Miranda-Esquivel DR.(in prep) geocleaMT: An R package to cleaning geographical data from electronic biodatabases.

Author(s)

R-Alarcon Viviana and Miranda-Esquivel Daniel R.

References

[1] Wieczorek, J. et al. 2012. Darwin core: An evolving community-developed biodiversity data standard. PloS One 7: e29715.

See Also

[readAndWrite](#)

delPointsOrSp	<i>Delete some/all records of a given species</i>
---------------	---------------------------------------------------

Description

The area of interest is framed as the four coordinates (North, South, West and East), if the species has some points outside of this area the user will be asked to decide whether delete the point or delete the species.

Usage

```
delPointsOrSp(data = NULL, rd.frmt = "readRDS", path = NULL,
  west = -120, east = -65, south = -30, north = 30, plot.distrib = T,
  wrt.frmt = "saveRDS", save.file = NULL)
```

Arguments

<code>data</code>	Vector of characters. Name of the input file.
<code>rd.frmt</code>	Vector of characters. The file format to read. By default it will be read as a R object using 'readRDS' argument, but it can be read as plain text using 'readTXT' argument. See details.
<code>path</code>	Vector of characters. Path to the input file.
<code>west</code>	Numeric vector. Western coordinate in decimal longitude format.
<code>east</code>	Numeric vector. Eastern coordinate in decimal longitude format.
<code>south</code>	Numeric vector. Southern coordinate in decimal latitude format.
<code>north</code>	Numeric vector. Northern coordinate in decimal latitude format.
<code>plot.distrib</code>	Logical vector. If 'TRUE', the distribution of the species will be plotted in red on a world map.
<code>wrt.frmt</code>	Vector of characters. Format to save output file. By default it will be written as a R object using 'saveRDS' argument, but it can be saved as plain text using 'saveTXT' argument. See details.
<code>save.file</code>	Vector of characters. Path to the output file for each species.

Details

The headers of the input file must follow the Darwin Core standard [1]. The user can see the guide using `data('ID_DarwinCore')` command. For more details about the formats to read and/or write, see [readAndWrite](#) function.

Value

A file for each species as data frame class.

Note

See: R-Alarcon V. and Miranda-Esquivel DR.(in prep) `geocleaMT`: An R package to cleaning geographical data from electronic biodatabases.

Author(s)

R-Alarcon Viviana and Miranda-Esquivel Daniel R.

References

[1] Wiczorek, J. et al. 2012. Darwin core: An evolving community-developed biodiversity data standard. *PloS One* 7: e29715.

See Also

[readAndWrite](#)

elevFromGg

*Assign elevation to a coordinates list***Description**

Given a coordinates vector, download the elevation data from The Google Maps Elevation API [1].

Usage

```
elevFromGg(data = NULL, rd.frmt = "readRDS", path = NULL,
  API.Key = NULL, starts.in = 1, round.coord = 4, save.name = "data",
  wrt.frmt = "writeRDS", save.assigned.in = NULL,
  save.unassigned.in = NULL, save.temp.in = NULL)
```

Arguments

<code>data</code>	Vector of characters. Name of the input file.
<code>rd.frmt</code>	Vector of characters. The file format to read. By default it will be read as a R object using 'readRDS' argument, but it can be read as plain text using 'readTXT' argument. See details.
<code>path</code>	Vector of characters. Path to the input file.
<code>API.Key</code>	If the user has a google's API key, it must be used here, otherwise no more than 2000 elevations could be assigned. See details.
<code>starts.in</code>	Numeric vector. Row number where should start the process.
<code>round.coord</code>	Integer. Number of digits to keep in coordinates.
<code>save.name</code>	Vector of characters. Name of the output file.
<code>wrt.frmt</code>	Vector of characters. Format to save the output file. By default it will be written as a R object using 'saveRDS' argument, but it can be saved as plain text using 'saveTXT' argument. See details.
<code>save.assigned.in</code>	Vector of characters. Path to the output file with the coordinates whose elevation was assigned. Output file will be called <alt.assigned>.
<code>save.unassigned.in</code>	Vector of characters. Path to the output file with the coordinates whose elevation was not assigned. Output file will be called <alt.unassigned>. See details.
<code>save.temp.in</code>	Vector of characters. Path to the temporal file as backup of the process. If the process is stopped, the user could reset it assigning the row number in <code>start.in</code> parameter and the process will start again using temporal file.

Details

For more details about the formats to read and/or write, see [readAndWrite](#) function.

The coordinates whose elevations were not assigned could be checked with [assignElevation](#). This function can download directly from Google Maps Elevation API[1]

Additional information at: https://developers.google.com/maps/documentation/elevation/#api_key

The headers of the input file must follow the Darwin Core standard [2]. The user can see the guide using `data('ID_DarwinCore')` command.

Value

A table data.frame class with coordinates and their respective elevation. A table with coordinates whose elevations were not assigned. A table as backup of the process.

Note

See: R-Alarcon V. and Miranda-Esquivel DR.(in prep) geocleaMT: An R package to cleaning geographical data from electronic biodatabases.

Author(s)

R-Alarcon Viviana and Miranda-Esquivel Daniel R.

References

[1] The Google Maps Elevation API. Available online at <https://developers.google.com/maps/documentation/elevation/intro>

[2] Wiczorek, J. et al. 2012. Darwin core: An evolving community-developed biodiversity data standard. PloS One 7: e29715.

See Also

[assignElevation](#)

Europa

Vectorial layer: European continent

Description

Map of the European continent. Shapefile of countries borders. This map was extracted from `data(wrld_simpl)` from `maptools` package [1]. The 'Europa' object uses the WGS84 datum.

Usage

`data(Europa)`

Format

Format class 'SpatialPolygonsDataFrame' [package 'sp'].

Source

<http://r-forge.r-project.org/projects/maptools/>

References

[1] Bivand et al. (2015) maptools: Tools for Reading and Handling Spatial Objects. CRAN. Version 0.8-37. Available online at <http://r-forge.r-project.org/projects/maptools/>

Examples

```
data(Europa)
plot(Europa)
```

Example

Input file to example Rmarkdown

Description

Input file to example Rmarkdown [1].

Usage

```
data(Example)
```

Format

File modified from GBIF, occurrences for 35 species of mammals with distribution on the biogeographic Choco.

References

[1] R-Alarcon V. and Miranda-Esquivel DR. (in prep) geocleaMT: An R package to cleaning geographical data from electronic biodatabases.

Examples

```
data(countryNames)
```

gbifDownSp

*Download occurrences for a list of species***Description**

Download occurrences from the Global Biodiversity Information Facility (GBIF) [1]

Usage

```
gbifDownSp(sp.name = NULL, taxon.key = NULL, genus = NULL,
  epithet = NULL, starts.in = 1, wrt.frmt = "saveRDS",
  save.download.in = NULL, ...)
```

Arguments

<code>sp.name</code>	Vector of characters. A species list to download. See details.
<code>taxon.key</code>	Vector of characters. A codes list, these codes are assigned by GBIF [1] to each species. See details.
<code>genus</code>	Vector of Characters. A genera list to download.
<code>epithet</code>	Vector of characters. A epithet list to download.
<code>starts.in</code>	Numeric vector. Row number to start the process.
<code>wrt.frmt</code>	Vector of characters. Format to save output file. By default it will be written as a R object using 'saveRDS' argument, but it can be saved as plain text using 'saveTXT' argument. See details.
<code>save.download.in</code>	Character. Vector of characters. Path to the output file for each species downloaded.
<code>...</code>	All parameters that can be parsed to occ_search function from <code>rgbif</code> package [2] as <code>limit</code> , <code>basisOfRecord</code> , <code>country</code> , <code>publishingCountry</code> , <code>lastInterpreted</code> , <code>geometry</code> , <code>collectionCode</code> , <code>institutionCode</code> and <code>year</code> .

Details

Additional parameters (...) see [occ_search](#) function from `rgbif` package. The user can only use one parameter at a time: `sp.name`, `taxon.key` or `genus` and/or `epithet`.

For `sp.name` parameter, a list of species names could be assigned. The species names must have genus and epithet, separated by a single space.

For `taxon.key` parameter the list has to be the ID number that assigns GBIF to each taxonomical level.

The headers of the input file must follow the Darwin Core standard [3]. The user can see the guide using `data('ID_DarwinCore')` command.

For more details about the formats to read and/or write, see [readAndWrite](#) function.

Value

Save one file by species. This file is a list of all occurrences recorded from GBIF.

Note

See: R-Alarcon V, Miranda-Esquivel DR.(in prep) geocleaMT: An R package to cleaning geographical data from electronic biodatabases.

Author(s)

R-Alarcon Viviana and Miranda-Esquivel Daniel R.

References

- [1] Global Biodiversity Information Facility. Available online at <http://www.gbif.org/>.
- [2] Chamberlain, S. et al. 2015. rgbif: Interface to the Global Biodiversity Information Facility 'API'. R package version 0.9.0. The Comprehensive R Archive Network (CRAN). Available online at <https://cran.r-project.org/web/packages/rgbif/>.
- [3] Wiecek, J. et al. 2012. Darwin core: An evolving community-developed biodiversity data standard. PloS One 7: e29715.

See Also

[readAndWrite](#)
[occ_search](#)

geocleaMT

A quick summary guide of geocleaMT package

Description

Although, the functions could be used in the suggested pipelines, each function is suited to perform an specific task or a set of them, given the package a high level of customizability [1]

Functions may be used for performing processes as:

Read input files in Darwin Core format using an R based function [[readDbR](#)] or a bash+R function [[readDbBash](#)]

Assign elevations to the data set [[assignElevation](#)]

Eliminate invasive species [[invasiveSp](#)]

Separate occurrences with and without geocode [[splitGeoref](#)]

Check decimal degree coordinates format [[checkCoord](#)]

Delete duplicated points or points with latitude and longitude assigned as 0.0 [[splitGeoref](#)]

Separate land/sea occurrences [[pointsAtSea](#)]

Delete points/species outside of a given distribution [[spOutPoly](#)]

Separate species with a minimum number of occurrences [[usefulSp](#)]

Verify that coordinates are into the range for latitude (-90,90) and longitude (-180,180) [[checkCoord](#)]

Eliminate specific records [[delPointsOrSp](#)]

Calculate for each species some general descriptors (mean propinquity, extreme distances, etc) [[meanPropinquity](#)]

References

[1] R-Alarcon, V. and Miranda-Esquivel, D.R. (in prep.) `geocleaMT` : An R package for cleaning/testing geographical data from electronic biodatabases.

IBISlist

Invasive species list

Description

List of invasive species recorded in Island Biodiversity and Invasive Species (IBIS) [1].

Usage

```
data(IBISlist)
```

Format

A data frame with 65 observations on the following 6 variables.

References

[1] Kells, S. S. and Worswick, C. 1997. An introduction to the IBIS database. Melbourne Institute of Applied Economic and Social Research.

Examples

```
data(IBISlist)
```

ID_DarwinCore	<i>ID Darwin Core standard</i>
---------------	--------------------------------

Description

Headers following the Darwin Core standard [1].

Usage

```
data(ID_DarwinCore)
```

Format

A data frame with 224 observations on the following 2 variables: ID and HeaderName.

References

[1] Wieczorek, J. et al. 2012. Darwin core: An evolving community-developed biodiversity data standard. PloS One 7: e29715.

Examples

```
data(ID_DarwinCore)
```

invasiveSp	<i>Separate invasive species</i>
------------	----------------------------------

Description

Separate species categorized as invasive for The Global Invasive Species Information Network (ISSG) [1] and Island Biodiversity and Invasive Species (IBIS) [2].

Usage

```
invasiveSp(data = NULL, rd.frmt = "readRDS", path = NULL, starts.in = 1,
  save.Sp.list = TRUE, wrt.frmt = "saveRDS", save.foreign.in = NULL,
  save.non.foreign.in = NULL, save.temp.in = NULL)
```

Arguments

<code>data</code>	Vector of characters. Name of the input file.
<code>rd.frmt</code>	Vector of characters. The file format to read. By default it will be read as a R object using 'readRDS' argument, but it can be read as plain text using 'readTXT' argument. See details.
<code>path</code>	Vector of characters. Path to the input file.
<code>starts.in</code>	starts.in Numeric vector. Row number to start the process.
<code>save.Sp.list</code>	Logic. If 'TRUE', two files are saved, one has the list of invasive species and other has the list of non-invasive species, both have two columns genus and specific epithet and it will be used as input for the gbifDownSp function.
<code>wrt.frmt</code>	Vector of characters. Format to save output file. By default it will be written as a R object using 'saveRDS' argument, but it can be saved as plain text using 'saveTXT' argument. See details.
<code>save.foreign.in</code>	Vector of characters. Path to write the output file with species that are categorized as invasive.
<code>save.non.foreign.in</code>	Vector of characters. Path to write the output file with species that are not categorized as invasive.
<code>save.temp.in</code>	Vector of characters. Path to the temporal file as backup of the process. If the process is stopped, the user could reset it assigning the row number in <code>start.in</code> parameter and the process will start again using the temporal file.

Details

For more details about the formats to read and/or write, see [readAndWrite](#) function.

If the argument in the `save.Sp.list` parameter is 'TRUE', the files will be saved in the path assigned in the `path` parameter.

The headers of the input file must follow the Darwin Core standard [3]. The user can see the guide using `data('ID_DarwinCore')` command.

Value

A table `data.frame` class with a list of invasive species. A table `data.frame` class with a list of non-invasive species.

Note

See: R-Alarcon V. and Miranda-Esquivel DR.(in prep) `geocleaMT`: An R package to cleaning geographical data from electronic biodatabases.

Author(s)

R-Alarcon Viviana and Miranda-Esquivel Daniel R.

References

- [1] Lowe, S. et al. 2000. 100 of the world's worst invasive alien species: A selection from the global invasive species database. The Invasive Species Specialist Group (ISSG) a specialist group of the Species Survival Commission (SSC) of the World Conservation Union (IUCN).
- [2] Kells, S. S. and Worswick, C. 1997. An introduction to the IBIS database. Melbourne Institute of Applied Economic and Social Research.
- [3] Wicczorek, J. et al. 2012. Darwin core: An evolving community-developed biodiversity data standard. PloS One 7: e29715.

See Also

[readAndWrite](#)

ISO2

ISO 3166-1 alpha-2

Description

List of International Organization for Standardization (ISO 3166-1 alpha-2) [1].

Usage

```
data(ISO2)
```

Format

A vector class Factor with 246 ISO-2 acronyms.

References

- [1]ISO 2015. International standard for country codes and codes for their subdivisions. Country codes - iso 3166. International Organization for Standardization. Available online at http://www.iso.org/iso/country_codes

Examples

```
data(ISO2)
```

ISO3

*ISO 3166-1 alpha-3***Description**

List of International Organization for Standardization (ISO 3166-1 alpha-3) [1].

Usage

```
data(ISO3)
```

Format

A vector class Factor with 246 ISO-3 acronyms.

References

[1]ISO 2015. International standard for country codes and codes for their subdivisions. Country codes - iso 3166. International Organization for Standardization. Available online at http://www.iso.org/iso/country_codes

Examples

```
data(ISO3)
```

ISOTocountry

Convert from International Organization for Standardization (ISO 3166-1 alpha-2 and alpha-3) to country name (English).

Description

Given the ISO-2 or ISO-3 code [1], it will assign the corresponding country name in English.

Usage

```
ISOTocountry(countryCode = NULL, iso = 2)
```

Arguments

countryCode	Data.Frame class. A vector with the ISO-2or ISO-3 Codes.
iso	numeric (2 or 3). To define the code whether ISO-2 (iso=2, default value) or ISO-3 (iso=3).

Details

The user can see the ISO 3166-1 alpha-2 or alpha-3 valid for each country using the line of code :
data(ISO2) or data(ISO3).

Value

A table with two columns (countryCode, country)

Note

See: R-Alarcon V. and Miranda-Esquivel DR.(in prep) geocleaMT: An R package to cleaning geographical data from electronic biodatabases.

Author(s)

R-Alarcon Viviana and Miranda-Esquivel Daniel R.

References

[1] ISO 2015. International standard for country codes and codes for their subdivisions. Country codes - iso 3166. International Organization for Standardization. Available online at http://www.iso.org/iso/country_codes.

See Also

[countryToCentroid](#)

meanPropinquity	<i>Mean Propinquity</i>
-----------------	-------------------------

Description

Calculate the mean propinquity[1] and some descriptors of the distribution.

Usage

```
meanPropinquity(coord.table = NULL, calculatedBy = "species",  
  wrt.frmt = "saveTXT", save.info.in = NULL, plot.dist = FALSE,  
  plot.onMap = FALSE, save.plot.in = NULL)
```

Arguments

coord.table	data frame class. Table with coordinates (decimalLongitude, decimalLatitude). If the process will be calculated by species, the table must include the species column.
calculatedBy	Vector of characters. If the argument is 'species', the mean propinquity will be calculated for each species. If it is 'whole' the propinquity will be calculated as a whole.

<code>wrt.frmt</code>	Vector of characters. Format to save the output file. By default it will be written as a R object using 'saveRDS' argument, but it can be saved as plain text using 'saveTXT' argument. See details.
<code>save.info.in</code>	Vector of characters. Path to the output file with information about Mean propinquity, median, mode, minimal spanning tree, etc; by species or as a whole.
<code>plot.dist</code>	Logical vector. If 'TRUE' a density plot for the distance between occurrences will be plotted and saved, by species or as a whole. If 'FALSE' there will be no plot.
<code>plot.onMap</code>	Logical vector. If 'TRUE' a map of occurrences will be plotted and saved, by species or as a whole. If 'FALSE' there will be no plot.
<code>save.plot.in</code>	Vector of characters. Path to the output file for each plot by species or as a whole.

Details

If the process uses the parameters `plot.dist` and `plot.onMap` as 'TRUE', the process will be slow. The headers of the input file must follow the Darwin Core standard [2]. The user can see the guide using `data('ID_DarwinCore')` command. For more details about the formats to write, see [readAndWrite](#) function.

Value

The output file will be saved with all species concatenated.

Note

See: R-Alarcon V. and Miranda-Esquivel DR.(in prep) `geocleaMT`: An R package to cleaning geographical data from electronic biodatabases.

Author(s)

R-Alarcon Viviana and Miranda-Esquivel Daniel R.

References

- [1] Rapoport. E.H. 1975 Areografia: Estrategias Geograficas de las Especies. Mexico, Fondo de Cultura Economica
- [2] Wiczorek, J. et al. 2012. Darwin core: An evolving community-developed biodiversity data standard. PloS One 7: e29715.

See Also

[readAndWrite](#)

Oceania	<i>Vectorial layer: Oceanian continent</i>
---------	--------------------------------------------

Description

Map of the Oceanian continent. Shapefile of countries borders. This map was extracted from `data(wrld_simpl)` from `maptools` package [1]. The 'Oceania' object uses the WGS84 datum.

Usage

```
data(Oceania)
```

Format

Format class 'SpatialPolygonsDataFrame' [package 'sp'].

Source

<http://r-forge.r-project.org/projects/maptools/>

References

[1] Bivand et al. (2015) `maptools`: Tools for Reading and Handling Spatial Objects. CRAN. Version 0.8-37. Available online at <http://r-forge.r-project.org/projects/maptools/>

Examples

```
data(Oceania)
plot(Oceania)
```

<code>pathStructure</code>	<i>Folder structure</i>
----------------------------	-------------------------

Description

Create the file structure for the protocol proposed R-Alarcon & Miranda-Esquivel (in prep) [1].

Usage

```
pathStructure(path.dir = NULL, group = NULL)
```

Arguments

<code>path.dir</code>	Vector of characters. Path to the folders' structure for all data/results.
<code>group</code>	Vector of characters. Name or names of group; create a subdirectory for each group into each main directory. See details.

Details

A group is a high taxonomic level such as Amphibia, Reptilia, Mammalia, etc.

Value

Created folders into the assigned path.

Note

See: [1] R-Alarcon V. and Miranda-Esquivel DR.(in prep) geocleaMT: An R package to cleaning geographical data from electronic biodatabases.

Author(s)

R-Alarcon Viviana and Miranda-Esquivel Daniel R.

References

[1] R-Alarcon V. and Miranda-Esquivel DR.(in prep) geocleaMT: An R package to cleaning geographical data from electronic biodatabases.

pointsAtSea	<i>Separate occurrences on Earth and at sea</i>
-------------	-------------------------------------------------

Description

Delete duplicates, separating records on Earth and at sea for each species.

Usage

```
pointsAtSea(data = NULL, path = NULL, rd.frmt = "readRDS",
  wrt.frmt = "saveRDS", save.OnEarth.in = NULL, save.AtSea.in = NULL)
```

Arguments

data	Vector of characters. Name of the input file.
path	Vector of characters. Path to the input file.
rd.frmt	Vector of characters. The file format to read. By default it will be read as a R object using 'readRDS' argument, but it can be read as plain text using 'readTXT' argument. See details.
wrt.frmt	Vector of characters. Format to save output file. By default it will be written as a R object using 'saveRDS' argument, but it can be saved as plain text using 'saveTXT' argument. See details.
save.OnEarth.in	Vector of characters. Path to the output file for each species with the coordinates on earth.
save.AtSea.in	Vector of characters. Path to the output file for each species with the coordinates at sea.

Details

For more details about the formats to read and/or write, see [readAndWrite](#) function.

The headers of the input file must follow the Darwin Core standard [1]. The user can see the guide using `data('ID_DarwinCore')` command.

Value

By each species, if this is the case, two output files data.frame class: The first one has the coordinates at sea for the species and the second one has the coordinates on earth for the same species.

Note

See: R-Alarcon V. and Miranda-Esquivel DR.(in prep) geocleaMT: An R package to cleaning geographical data from electronic biodatabases.

Author(s)

R-Alarcon Viviana and Miranda-Esquivel Daniel R.

References

[1] Wiczorek, J. et al. 2012. Darwin core: An evolving community-developed biodiversity data standard. PloS One 7: e29715.

See Also

[readAndWrite](#)

readAndWrite

Read and write tables

Description

Read or write tables in two specific formats for standardizing the cleaning process. This function reads the input and writes the output.

Usage

```
readAndWrite(action = NULL, frmt = NULL, path = NULL, name = NULL,  
             object = NULL)
```

Arguments

action	Vector of characters. If 'read' string, the function will read the file. If 'write' string, the function will write output file.
frmt	Vector of characters. The format to read or write can be of two types: plain text 'txt' or a R object 'RDS'.
path	Vector of characters. Path to the input file or the destination Path to the output file. See details.
name	Vector of characters. Name of the input file or name of file to write.
object	Vector of characters. Object to write. See details.

Details

It reads and/or writes a file in plain text format (txt) using the functions [read.table](#) and/or [write.table](#). To maintain the standard manipulation the arguments for both functions [read.table](#) and [write.table](#) parameters, are the default, except for the parameters: sep= 'TAB', header= T, quotes= F and row-names = F.

If the argument in the action parameter is 'write', the argument in object and name parameters must be assigned. But, if the argument in the action parameter is 'read', the argument in object parameter must be omitted.

Value

For action = 'read', load the file given for the user. For action = 'write', save an output file. Both depend on the frmt parameter.

Note

See: R-Alarcon V. and Miranda-Esquivel DR.(in prep) geocleaMT: An R package to cleaning geographical data from electronic biodatabases.

Author(s)

R-Alarcon Viviana and Miranda-Esquivel Daniel R.

See Also

[saveRDS](#)

[write.table](#)

readDbBash	<i>Read database table using bash commands.</i>
------------	-------------------------------------------------

Description

Read a database which follows the Darwin Core Standard [1].

Usage

```
readDbBash(data = NULL, path.data = NULL, cut.col = c(1, 78, 79, 200, 218,
  219), del.t.undeterm = TRUE, save.name = NULL, wrt.frmt = "saveRDS",
  save.in = NULL)
```

Arguments

<code>data</code>	Vector of characters. Name of the input file.
<code>path.data</code>	Vector of characters. Path to the input file.
<code>cut.col</code>	Numeric vector. Columns number to read into database. By default, the columns <code>c(1,78,79,200,218,219)</code> are read. These correspond to headers of the Darwin Core standard [1]: <code>gbifID</code> , <code>decimalLongitude</code> , <code>decimalLatitude</code> , <code>elevation</code> , <code>speciesKey</code> and <code>species</code> . See details.
<code>del.t.undeterm</code>	Logical vector. If it is 'TRUE' return a data table with only occurrences that have taxonomic determination until species. Otherwise, it could return all occurrences read into database.
<code>save.name</code>	Vector of characters. Name of the output file.
<code>wrt.frmt</code>	Vector of characters. Format to save output file. By default it will be written as a R object using
<code>save.in</code>	Vector of characters. Path to the output file.

Details

We recommend to use this function when the database has more than one hundred thousand occurrences and / or the computer has low memory. [readDbBash](#) uses the `cut` function from BASH programming language and can be functional on Linux or Mac operative systems. If this is not the case, we recommended to use the [readDbR](#) which runs into the R platform and can be used on any operative system (Linux, Mac, or Windows). However, the [readDbBash](#) function always will be faster than [readDbR](#) (up to four times faster).

Databases downloaded from the Global Biodiversity Information Facility (GBIF) [2] are exported with DarwinCore headers and the separator columns is TAB, and hence all databases read using this functions must be able TAB as separator. See [readAndWrite](#) function.

For `cut.col` parameter, the numbers columns to split must be sorted sequentially. For databases downloaded from GBIF [2], the number for each header can be seen using `data('ID_DarwinCore')` command on console.

For more details about the formats to read and/or write, see [readAndWrite](#) function

Value

writing a data table in Data.frame class and as vector return a table with descriptive quantities.

Note

See: R-Alarcon V. and Miranda-Esquivel DR.(in prep) geocleaMT: An R package to cleaning geographical data from electronic biodatabases.

Author(s)

R-Alarcon Viviana and Miranda-Esquivel Daniel R.

References

- [1] Wieczorek, J. et al. 2012. Darwin core: An evolving community-developed biodiversity data standard. PloS One 7: e29715.
- [2] Global Biodiversity Information Facility. Available online at <http://www.gbif.org/>.

See Also

[readDbR](#)
[readAndWrite](#)

readDbR	<i>Read database table using R</i>
---------	------------------------------------

Description

Read a database which follows the Darwin Core Standard [1].

Usage

```
readDbR(data = NULL, path.data = NULL, cut.col = c(1, 78, 79, 200, 218,
  219), delc.undeterm = TRUE, save.name = NULL, wrt.frmt = "saveRDS",
  save.in = NULL)
```

Arguments

data	Vector of characters. Name of the input file.
path.data	Vector of characters. Path to the input file.
cut.col	Numeric vector. Columns number to read into database. By default, the columns c(1,78,79,200,218,219) are read. These correspond to headers of the Darwin Core standard [1]: gbifID, decimalLongitude, decimalLatitude, elevation, speciesKey and species. See details.

delt.undeterm	Logical vector. If it is 'TRUE' return a data table with only occurrences that have taxonomic determination until species. Otherwise, it could return all occurrences read into database.
save.name	Vector of characters. Name of the output file.
wrt.frmt	Vector of characters. Format to save output file. By default it will be written as a R object using 'saveRDS' argument, but it can be saved as plain text using 'saveTXT' argument. See details.
save.in	Vector of characters. Path to the output file.

Details

We recommend to use this function when the database have fewer than one hundred thousand occurrences. This function works on R platform and can be performed on any operative system (Linux, Mac OS or Windows). If the database to read has more than one hundred thousand occurrences, we recommend to use the [readDbBash](#) function. [readDbBash](#) uses the cut function from BASH programming language and can be functional on Linux or iOS operative systems, but the [readDbBash](#) function always will be faster than [readDbR](#) (until four times faster).

Databases downloaded from Global Biodiversity Information Facility (GBIF)[2] are exported with DarwinCore headers and the column separator is TAB.

See [readAndWrite](#) function.

For cut.col parameter, the numbers columns to split must be sorted sequentially. For database download from GBIF [2], the number for each header can be seen using data('ID_DarwinCore') command on console in the ID column.

For more details about the formats to read and/or write, see [readAndWrite](#) function.

Value

writing a data table as a data.frame class and a vector a table with descriptive statistics.

Note

See: R-Alarcon V. and Miranda-Esquivel DR.(in prep) geocleaMT: An R package to cleaning geographical data from electronic biodatabases.

Author(s)

R-Alarcon Viviana and Miranda-Esquivel Daniel R.

References

- [1] Wieczorek, J. et al. 2012. Darwin core: An evolving community-developed biodiversity data standard. PloS One 7: e29715.
- [2] Global Biodiversity Information Facility. Available online at <http://www.gbif.org/>.

See Also

[readDbR](#)
[readAndWrite](#)

splitGeoref

*Separate geo-referenced records***Description**

Get only georeferenced records from a database downloaded from GBIF[1].

Usage

```
splitGeoref(data = NULL, rd.frmt = "readRDS", path = NULL, min.occ = 3,
  round.coord = 4, wrt.frmt = "saveRDS", save.min.occ.in = NULL,
  save.georef.in = NULL, save.ungeoref.in = NULL)
```

Arguments

data	Vector of characters. Name of the input file.
rd.frmt	Vector of characters. The file format to read. By default it will be read as a R object using 'readRDS' argument, but it can be read as plain text using 'readTXT' argument. See details.
path	Vector of characters. Path to the data file(s).
min.occ	Numeric vector. Minimal number of georeferenced occurrences for that a species can be considered. If the argument is 3, only species with 3 or more occurrences will be considered in the process.
round.coord	Integer. Number of digits to keep in coordinates.
wrt.frmt	Vector of characters. Format to save output file. By default it will be written as a R object using 'saveRDS' argument, but it can be saved as plain text using 'saveTXT' argument. See details.
save.min.occ.in	Character. Path to save the output file of species that do not fulfill the minimum required number of occurrences in the parameter min.occ. Default format is '.RDS'. See details.
save.georef.in	Vector of characters. Path to write the output file for each species with georeferenced records.
save.ungeoref.in	Character. Path to save ungeoreferenced data, a file for each species. Default format is '.RDS'. See details.

Details

For more details about the formats to read and/or write, see [readAndWrite](#) function.

The headers of the input file must follow the Darwin Core standard [2]. The user can see the guide using `data('ID_DarwinCore')` command.

Value

By each species, if this is the case, two output files data.frame class: The first one have the georeferenced and the second one have the ungeoreferenced

Note

See: R-Alarcon V. and Miranda-Esquivel DR.(in prep) geocleaMT: An R package to cleaning geographical data from electronic biodatabases.

Author(s)

R-Alarcon Viviana and Miranda-Esquivel Daniel R.

References

- [1] Global Biodiversity Information Facility. Available online at <http://www.gbif.org/>.
- [2] Wieczorek, J. et al. 2012. Darwin core: An evolving community-developed biodiversity data standard. PloS One 7: e29715.

See Also

[readAndWrite](#)

spOutPoly

Obtain species outside of a polygon

Description

This function separates species with restricted distribution to a polygon or a defined area, following a series of conditions.

Usage

```
spOutPoly(data = NULL, rd.frmt = "readRDS", path = NULL,
  shp.poly = NULL, max.per.out = 5, max.occ.out = 3, execute = T,
  B1 = F, B2 = T, wrt.frmt = "saveRDS", save.inside.in = NULL,
  save.outside.in = NULL)
```

Arguments

data	Vector of characters. Name of the input file.
rd.frmt	Vector of characters. The file format to read. By default it will be read as a R object using 'readRDS' argument, but it can be read as plain text using 'readTXT' argument. See details.
path	Vector of characters. Path to the input file.

shp.poly	SpatialPolygonDataframe Class. Polygon of the referenced areas. The area where the species has to be distributed.
max.per.out	Numeric vector. Maximum percentage of occurrences that can be outside of polygon for the species to be considered as species with restricted distribution. See details.
max.occ.out	Numeric vector. Absolute maximum number of occurrences that can be outside of a polygon for the species to be considered as species with restricted distribution. See details.
execute	Logical vector. If 'TRUE', the process will be executed on species. If 'FALSE' the species will only be classified and an information table will be saved.
B1	Logical vector. If 'TRUE', the condition B1T will be applied. If 'FALSE' the condition B1F will be applied. See details.
B2	Logical vector. If 'TRUE', the condition B2T will be applied. If 'FALSE' the condition B2F will be applied. See details.
wrt.frmt	Vector of characters. Format to save output file. By default it will be written as a R object using
save.inside.in	Vector of characters. Path to the output file for each species whose distribution was considered restricted to polygon.
save.outside.in	Vector of characters. Path to the output file for each species whose distribution was considered widespread.

Details

#' To see the descriptions and details about conditions and parameters the user should check R-Alarcon V. and Miranda-Esquivel DR.(in prep) geocleaMT: An R package to cleaning geographical data from electronic biodatabases.

The condition A, B3 and C for species classification are executed by default, and these can not be modified.

The headers of the input file must follow the Darwin Core standard [1]. The user can see the guide using data(ID_DarwinCore) command. For more details about the formats to read and/or write, see [readAndWrite](#) function.

Value

If execute is 'TRUE' a table data.frame class with information about how was the species classified and an output file by species with occurrences. If execute is FALSE, a table data.frame class will be saved, with information about how was the species classified.

Note

See: R-Alarcon V. and Miranda-Esquivel DR.(in prep) geocleaMT: An R package to cleaning geographical data from electronic biodatabases.

Author(s)

R-Alarcon Viviana and Miranda-Esquivel Daniel R.

References

[1] Wieczorek, J. et al. 2012. Darwin core: An evolving community-developed biodiversity data standard. PloS One 7: e29715.

See Also

[readAndWrite](#)

stackSp	<i>Join species files</i>
---------	---------------------------

Description

Join all independent files into a single multi-species file.

Usage

```
stackSp(data = NULL, rd.frmt = "readRDS", path = NULL, save.name = NULL,  
        save.staking.in = NULL, wrt.frmt = "saveRDS")
```

Arguments

data	Vector of characters. Name of the input file.
rd.frmt	Vector of characters. The file format to read. By default it will be read as a R object using 'readRDS' argument, but it can be read as plain text using 'readTXT' argument. See details.
path	Vector of characters. Path to the input file.
save.name	Vector of characters. Name of the output file.
save.staking.in	Vector of characters. Path to the output file.
wrt.frmt	Vector of characters. Format to save output file. By default it will be written as a R object using 'saveRDS' argument, but it can be saved as plain text using 'saveTXT' argument. See details.

Details

The headers of the input file must follow the Darwin Core standard [1]. The user can see the guide using `data('ID_DarwinCore')` command. For more details about the formats to read and/or write, see [readAndWrite](#) function.

Value

The output file will be saved with all species concatenated.

Note

See: R-Alarcon V. and Miranda-Esquivel DR.(in prep) geocleaMT: An R package to cleaning geographical data from electronic biodatabases.

Author(s)

R-Alarcon Viviana and Miranda-Esquivel Daniel R.

References

[1] Wieczorek, J. et al. 2012. Darwin core: An evolving community-developed biodiversity data standard. PloS One 7: e29715.

See Also

[readAndWrite](#)

usefulSp	<i>Useful species</i>
----------	-----------------------

Description

Get species with a minimum number of occurrences.

Usage

```
usefulSp(data = NULL, path = NULL, cut.off = 3, rd.frmt = "readRDS",
  wrt.frmt = "saveRDS", save.useful.in = NULL, save.useless.in = NULL)
```

Arguments

data	Vector of characters. Name of the input file..
path	The path for the file or files.
cut.off	Numeric. Minimal (absolute) number of occurrences for a given species, the default value is 3, as the minimal number of points to draw a polygon.
rd.frmt	Character. Input file format, by default it uses a .RDS format. See details.
wrt.frmt	Character. Output file format, by default it uses a '.RDS' format, although it could save a txt table. See details.
save.useful.in	String. Output file path for useful species (those with occurrences larger than the cut.off value. See details.
save.useless.in	String. Output file path for useless species (those with occurrences smaller than the cut.off value. See details.

Details

To check input/output formats see [readAndWrite](#). Column names must follow Darwin Core2[1] standard. See `data('ID_DarwinCore')`.

Value

A `data.frame` object and a file for each species.

Note

See: R-Alarcon V. and Miranda-Esquivel DR.(in prep) `geocleaMT`: An R package to cleaning geographical data from electronic biodatabases.

Author(s)

R-Alarcon Viviana and Miranda-Esquivel Daniel R.

References

[1] Wieczorek, J. et al. 2012. Darwin core: An evolving community-developed biodiversity data standard. *PloS One* 7: e29715.

See Also

[readAndWrite](#)

Index

*Topic **datasets**

- Africa, [2](#)
- Altitudes, [3](#)
- America, [4](#)
- Antarctica, [5](#)
- Asia, [5](#)
- ChcBiog, [8](#)
- countryNames, [12](#)
- Europa, [18](#)
- Example, [19](#)
- IBISlist, [22](#)
- ID_DarwinCore, [23](#)
- ISO2, [25](#)
- ISO3, [26](#)
- Oceania, [29](#)

Africa, [2](#)

Altitudes, [3](#)

America, [4](#)

Antarctica, [5](#)

Asia, [5](#)

assignElevation, [6](#), [18](#), [21](#)

ChcBiog, [8](#)

checkCoord, [8](#), [21](#), [22](#)

coordToCountry, [10](#), [11](#)

coordToISO, [10](#), [11](#)

countryNames, [12](#)

countryToCentroid, [12](#), [14](#), [27](#)

countryToISO, [13](#), [13](#)

cutRange, [14](#)

delPointsOrSp, [15](#), [22](#)

elevFromGg, [3](#), [7](#), [17](#)

Europa, [18](#)

Example, [19](#)

gbifDownSp, [20](#), [24](#)

geocleaMT, [21](#)

geocleaMT-package (geocleaMT), [21](#)

IBISlist, [22](#)

ID_DarwinCore, [23](#)

invasiveSp, [21](#), [23](#)

ISO2, [25](#)

ISO3, [26](#)

ISOTocountry, [26](#)

meanPropinquity, [22](#), [27](#)

occ_search, [20](#), [21](#)

Oceania, [29](#)

pathStructure, [29](#)

pointsAtSea, [21](#), [30](#)

read.table, [32](#)

readAndWrite, [7](#), [9](#), [15](#), [16](#), [18](#), [20](#), [21](#), [24](#), [25](#),
[28](#), [31](#), [31](#), [33–41](#)

readDbBash, [21](#), [33](#), [33](#), [35](#)

readDbR, [21](#), [33](#), [34](#), [34](#), [35](#)

saveRDS, [32](#)

splitGeoref, [21](#), [36](#)

spOutPoly, [21](#), [37](#)

stackSp, [39](#)

usefulSp, [22](#), [40](#)

write.table, [32](#)