

# Package ‘cshapes’

February 14, 2012

**Version** 0.3-1

**Date** 2011-06-26

**Title** CShapes Dataset and Utilities

**Author** Nils B. Weidmann <nils.weidmann@gmail.com>, Doreen Kuse  
<doreen.kuse@gmail.com>, Kristian Skrede Gleditsch <ksg@essex.ac.uk>

**Maintainer** Nils B. Weidmann <nils.weidmann@gmail.com>, Kristian Skrede  
Gleditsch <ksg@essex.ac.uk>

**Depends** R (>= 2.8.0), sp, maptools, rJava

**SystemRequirements** Java (>= 1.5.0)

**LazyLoad** yes

**Description** Package for CShapes, a GIS dataset of country boundaries  
(1946-2008). Includes functions for data extraction and the  
computation of distances matrices and -lists.

**License** GPL (>= 2)

**Repository** CRAN

**Date/Publication** 2011-06-27 07:00:26

## R topics documented:

cshapes-package	2
cshapes2yearly	3
cshp	4
distlist	5
distmatrix	6

<b>Index</b>	<b>8</b>
--------------	----------

## Description

R Package for CShapes, a GIS dataset of country boundaries (1946-2008). Includes functions for data extraction and the computation of weights matrices.

## Details

The cshapes package facilitates the use of CShapes from R. CShapes is a GIS dataset of historical country boundaries (1946-2008) and compatible with two country lists (Gleditsch and Ward 1999 and Correlates of War, see references below). In particular, the package enables access to the dataset directly, as well as distance computations on country polygons for specific points in time. Access to the dataset from within R is done using the `cshp` function. Two functions exist to compute minimum-, capital- and centroid distances between countries: the `distmatrix` function returns these as a matrix (convenient for many spatial statistical applications), and the `distlist` function returns a list of dyadic distances (distances between country pairs). See the examples given in the documentation of the functions in this package. The main cshapes website is located at <http://nils.weidmann.ws/projects/cshapes> and contains additional documentation and examples for the dataset and the R package.

Implementation: The distance matrix computation is implemented in Java using the JAMA, JTS and JUMP libraries, using the rJava package to invoke the computation from R. Warning: If the Java process receives an interrupt from outside R, this will cause R to abort.

## Author(s)

Nils B. Weidmann <nils.weidmann@gmail.com>, Doreen Kuse <doreen.kuse@gmail.com>, Kristian S. Gleditsch <ksg@essex.ac.uk>

Maintainer: Nils B. Weidmann <nils.weidmann@gmail.com>, Kristian S. Gleditsch <ksg@essex.ac.uk>

## References

Correlates of War Project. 2008. "State System Membership List, v2008.1." Online, <http://correlatesofwar.org>.

Gleditsch, Kristian S. & Michael D. Ward. 1999. "Interstate System Membership: A Revised List of the Independent States since 1816." *International Interactions* 25: 393-413. Online, <http://privatewww.essex.ac.uk/~ksg/statelist.html>.

Java Matrix Package JAMA, v. 1.0.2 (<http://math.nist.gov/javanumerics/jama/>)

Java Topology Suite JTS, v. 1.7.2 (<http://www.vividsolutions.com/jts/jtshome.htm>). License: GPL

Java Unified Mapping Platform JUMP, v. 1.2 (<http://www.vividsolutions.com/jump/>). License: GPL

---

cshapes2yearly      *Convert CShapes data frame to a country-year format*

---

## Description

This function converts the original cshapes data frame to a panel structure with country-year observations. The user has to specify the list of variables from CShapes to be included in the new data frame. These can be variables already included in CShapes, but will (in most cases) be user-computed variables appended to the CShapes data frame.

## Usage

```
cshapes2yearly(cshp, vars, useGW=T)
```

## Arguments

cshp	A SpatialPolygonsDataFrame in CShapes format. In almost all cases, this will be the original cshapes frame with some variables appended. These variables can be computed outside R (e.g.) in a GIS and using this function, can be converted to a convenient format for statistical analysis.
vars	A character vector with the variable names to be included in the new frame (besides country code and year).
useGW	Boolean argument specifying the system membership coding. TRUE (Default): Gleditsch and Ward (1999). FALSE: Correlates of War.

## Value

A data frame, with country-year observations. For each year, the function will use the CShapes polygon that was active on June 30 of the respective year. The columns in the data frame are "ctrcode" for the country code (either COW or Gleditsch&Ward, according to the useGW parameter setting), and "year".

## Author(s)

Nils B. Weidmann

## Examples

```
# Convert the standard cshapes dataset
# to a country-year format
cshp.data <- cshp()
cshp.yearly <- cshapes2yearly(cshp.data, vars=c("AREA", "CAPNAME"))
```

---

`cshp`*Access the CShapes dataset in R*

---

### Description

The `cshp` function makes the `cshapes` dataset available for usage in R. If no date is given, it returns a `SpatialPolygonsDataFrame` with the complete CShapes dataset. If specified, the date is used to create a snapshot of the dataset, containing all `cshapes` polygons active at the given date.

### Usage

```
cshp(date=NA, useGW=TRUE)
```

### Arguments

<code>date</code>	The date for which the <code>cshapes</code> polygons should be extracted. This argument must be of type <code>Date</code> and must be in the range 1/1/1946 - 30/6/2008. If omitted, the complete dataset is returned.
<code>useGW</code>	Boolean argument specifying the system membership coding. <code>TRUE</code> (Default): Gleditsch and Ward (1999). <code>FALSE</code> : Correlates of War.

### Value

A `SpatialPolygonsDataFrame`, containing the complete CShapes dataset, or the CShapes snapshot for the specified date.

### Author(s)

Nils B. Weidmann

### See Also

[SpatialPolygonsDataFrame](#)

### Examples

```
# Retrieve the dataset
cshp.data <- cshp()

# Get summary statistics
summary(cshp.data)

# Extract Switzerland
switzerland <- cshp.data[cshp.data$COWCODE==225,]

# Plot Switzerland
plot(switzerland)
```

```
# Extract a snapshot of cshapes as of June 30, 2002
# using the Gleditsch&Ward coding system
cshp.2002 <- cshp(date=as.Date("2002-6-30"), useGW=TRUE)
```

---

distlist

*Compute distance lists on the CShapes dataset*


---

## Description

This function computes a distance list for the given date. It selects all the active CShapes polygons, determines their distances and outputs a distance list. A distance list is a list of dyads of countries and the distances between them. This list is returned as a data frame with three columns:

1. `ccode1` – country 1's code in the coding system specified by the `useGW` parameter
2. `ccode2` – country 2's code in the coding system specified by the `useGW` parameter
3. `capdist`, `centdist`, `mindist` – distance between country 1 and country 2 in km, where distance can be either capital distance, centroid distance or minimum distance, as specified by the `type` parameter. The latter computation is very expensive if polygons have many nodes. For that reason, the function simplifies the country polygons according to the Douglas-Peucker algorithm ([http://en.wikipedia.org/wiki/Ramer-Douglas-Peucker\\_algorithm](http://en.wikipedia.org/wiki/Ramer-Douglas-Peucker_algorithm)), which eliminates points from the polygons and speeds up computation. The `tolerance` parameter specifies the tolerance for the simplification; a value of 0 disables it.

Note that the function returns directed dyads. For example, if there is a country with code 1 and a country with code 2, the resulting data frame contains the dyads (1,2), (2,1), (1,1) and (2,2). It is easy to extract the directed dyads from this data frame (see example below).

## Usage

```
distlist(date, type="mindist", tolerance=0.1, useGW=T)
```

## Arguments

<code>date</code>	The date for which the distance list should be computed. This argument must be of type <code>Date</code> and must be in the range 1/1/1946 - 30/6/2008.
<code>type</code>	Specifies the type of distance list: <code>capdist</code> for capital distances, <code>centdist</code> for centroid distances, and <code>mindist</code> for minimum distances.
<code>useGW</code>	Boolean argument specifying the system membership coding. <code>TRUE</code> (Default): Gleditsch and Ward (1999). <code>FALSE</code> : Correlates of War.
<code>tolerance</code>	Tolerance for polygon simplification according the the Douglas-Peucker algorithm. Only used for <code>mindist</code> computation ( <code>type="mindist"</code> ).

## Value

A distance list with three columns, the first two of which contain the identifiers for the two countries in the dyad, and the third one containing the distance between the two countries.

**Author(s)**

Nils B. Weidmann

**Examples**

```
# Compute a list of minimum distances
# for the international system on 1/1/2008
# using the Correlates of War list and the default accuracy
dl <- distlist(as.Date("2008-1-1"), type="capdist", useGW=FALSE)

# we eliminate duplicate dyads
dl <- subset(dl, ccode1 < ccode2)
```

---

distmatrix

*Compute distance matrices on the CShapes dataset*

---

**Description**

This function computes a distance matrix for the given date. It selects all the active CShapes polygons, determines their distances and outputs a distance matrix in kilometers. The function can compute different types of distance matrices, specified by the "type" parameter: (i) capital distances, and (ii) centroid distances, and (iii) minimum distances between polygons. The latter computation is very expensive if polygons have many nodes. For that reason, the function simplifies the country polygons according to the Douglas-Peucker algorithm ([http://en.wikipedia.org/wiki/Ramer-Douglas-Peucker\\_algorithm](http://en.wikipedia.org/wiki/Ramer-Douglas-Peucker_algorithm)), which eliminates points from the polygons and speeds up computation. The tolerance parameter specifies the tolerance for the simplification; a value of 0 disables it.

**Usage**

```
distmatrix(date, type="mindist", tolerance=0.1, useGW=T)
```

**Arguments**

date	The date for which the distance matrix should be computed. This argument must be of type Date and must be in the range 1/1/1946 - 30/6/2008.
type	Specifies the type of distance matrix: capdist for capital distances, centdist for centroid distances, and mindist for minimum distances.
useGW	Boolean argument specifying the system membership coding. TRUE (Default): Gleditsch and Ward (1999). FALSE: Correlates of War.
tolerance	Tolerance for polygon simplification according to the Douglas-Peucker algorithm. Only used for mindist computation (type="mindist").

**Value**

A quadratic weights matrix, with the row and column labels containing the country identifiers in the specified coding system (COW or G&W).

**Author(s)**

Nils B. Weidmann

**Examples**

```
# Compute a matrix of minimum distances
# for the international system on 1/1/2008
# using the Correlates of War list and the default accuracy
wmat <- distmatrix(as.Date("2008-1-1"), type="capdist", useGW=FALSE)

# For model estimation, our matrix should contain
# the inverted distances
wmat <- 1/wmat

# Fix the values along the diagonale
diag(wmat) <- 0
```

# Index

## \*Topic **package**

[cshapes-package](#), [2](#)

[cshapes \(cshapes-package\)](#), [2](#)

[cshapes-package](#), [2](#)

[cshapes2yearly](#), [3](#)

[cshp](#), [2](#), [4](#)

[distlist](#), [2](#), [5](#)

[distmatrix](#), [2](#), [6](#)

[SpatialPolygonsDataFrame](#), [4](#)