

Package ‘tangles’

October 11, 2019

Type Package

Title Anonymization of Spatial Point Patterns and Raster Objects

Version 0.8.1

Date 2019-10-10

Description Spatial data anonymization preserves confidentiality. Using methods described in Zandbergen (2014) <doi:10.1155/2014/567049>, spatial data anonymization is achieved by dithering original spatial coordinates with combinations of randomized vertical, horizontal and rotational shifts. This can apply to non-grid spatial point patterns and raster objects, and the methods preserve the same spatial characteristics and relationships of the original data. Unique hash keying enables data subjected to anonymization sequences to be re-identified where required.

Depends R (>= 3.0.1)

Imports raster, digest, sp

Suggests rasterVis, rgdal, knitr, rmarkdown

VignetteBuilder knitr

License GPL-2

LazyData no

Encoding UTF-8

NeedsCompilation no

Author Brendan Malone [cre, aut]

Maintainer Brendan Malone <brendan.malone@csiro.au>

Repository CRAN

Date/Publication 2019-10-11 11:30:02 UTC

R topics documented:

detangler_data	2
detangles	2
Hunter Valley covariates (subset area)	4
Hunter Valley subsoil pH points	6
tangler	8
tangles	9

Index**12**

detangler_data	<i>Example object that is output from tangles function.</i>
----------------	---

Description

This output that would be derived from the `tangles` function. It contains the instructions for anonymization of associated point pattern or raster data, or the re-identify tangled data.

Usage

```
data(detangler_data)
```

Format

`detangler_data` is a list with 3 elements. The first element is a unique hash key for tracking data processing provenance. The second element contains the integer sequence of steps that were used to anonymize a given data sets. There are 3 possible anonymizing procedures: shift vertical (1), shift (horizontal), and rotate (3). The last element is a data frame that encodes the values to perform either anonymization or re-identification, given the particular anonymization step used.

Note

This data is the sort of output that one would expect when the `tangles` function is used.

References

- CM O’Keefe, S Otorepec, M Elliot, E Mackey, and K O’Hara (2017) [The De-Identification Decision Making Framework](#). CSIRO Reports EP173122 and EP175702.

Examples

```
data(detangler_data)
```

detangles	<i>Revert anonymized point patterns or raster objects</i>
-----------	---

Description

Anonymized spatial data can be reverted back to their original coordinates with a given `detangler` object. For correct re-identification, there is an assumed match in the hash key between anonymized data and `detangler` object.

Usage

```
detangles(data=NULL, tanglerInfo=NULL, raster_object = FALSE,  
stub = NULL, hash_key=NULL,  
saveTangles = FALSE, path = NULL)
```

Arguments

data	Can be either a 2 column matrix of spatial coordinates or a raster object.
tanglerInfo	output from tangles which encodes the steps and values of the re-identification process.
raster_object	logical; If TRUE the data to be re-identified is a raster object. This means the data associated with the raster are carried through to the outputs.
stub	Character; A string of user-defined setting for establishing naming conventions.
hash_key	Character; A string of the hash key that was used for the original anonymization.
saveTangles	logical; Save function output to file. Default is FALSE, but it is recommended to always save outputs.
path	character; Path to where outputs are saved.

Value

The transformed coordinates or raster object (dependent on input data).

Note

Outputs can be saved to any specified directory with the hash key associated with the detangler object and user-defined stub in the file name. This will be saved as an .rds file.

Author(s)

Brendan Malone

References

- CM O'Keefe, S Otorepec, M Elliot, E Mackey, and K O'Hara (2017) [The De-Identification Decision Making Framework](#). CSIRO Reports EP173122 and EP175702.

Examples

```
## POINT DATA tangles  
library(digest)  
data("HV_subsoilpH")  
str(HV_subsoilpH)  
dat.xy<- HV_subsoilpH[,1:2]  
xyData<- as.matrix(dat.xy)  
  
# anonymize with 5 levels of abstraction  
tangles.out<- tangles(data = xyData,
```

```
depth = 5,
rasterdata = TRUE,
raster_object = FALSE,
saveTangles = FALSE)

# detangle
xyData<- as.matrix(tangles.out[[1]])

point_detang<- detangles(data=xyData,
tanglerInfo=tangles.out[[2]],
raster_object = FALSE,
stub = "hv_fix",
hash_key = tangles.out[[2]][1],
saveTangles = FALSE)
```

Hunter Valley covariates (subset area)

*Suite of selected environmental covariates for a subset the Lower
Hunter Valley, NSW*

Description

A RasterStack of selected environmental covariates from a subset area of the Lower Hunter Valley, NSW.

Usage

```
data(hunterCovariates_sub)
```

Format

hunterCovariates_sub is a RasterStack of selected environmental covariate rasters. The rasters have a pixel resolution of 25m x 25m. The CRS of each raster is WGS 84 UTM Zone 56. The RasterStack contains the following rasters:

- **Terrain_Ruggedness_Index**: A quantitative measure of topographic heterogeneity. High values indicate terrain is rugged or heterogeneous.
- **AACN**: Difference between elevation and an interpolation of a channel network base level elevation. Knowledge of the spatial distribution of channel networks (lines) is therefore necessary for this parameter.
- **Landsat_Band1**: Earth surface reflectance information derived from Landsat 7 ETM+ satellite. Band 1 spectral range is .45 to .52 microns (visible blue).
- **Elevation**: Meters above sea level; derived from a digital elevation model.
- **Hillshading**: Analytic hill shading derived from digital elevation model and a fixed sun degree angle.

- **Light_insolation**: Measure of potential incoming solar radiation, and used as a parameter for evaluating the positional aspect effect. Derived from digital elevation model, this parameter was evaluated over the duration of a single calendar year with a 5 day time step.
- **Mid_Slope_Positon**: A relative slope position parameter which gives a classification of the slope position in both valley and crest positions.
- **MRVBF**: Multi-resolution valley bottom flatness is derived using slope and elevation to classify valley bottoms as flat, low areas (Gallant and Dowling 2003). This is accomplished through a series of neighborhood operations at progressively coarser resolutions with the goal of identifying both small and large valleys. MRVBF has been used extensively for the delineation and grading of valley floor units corresponding to areas of alluvial and colluvial deposits. High values of MRVBF indicate relatively low, flat areas of the landscape.
- **NDVI**: Normalized difference vegetation index. Derived from Landsat 7 data $(B4-B3)/(B4+B3)$. High values indicate actively growing vegetation.
- **TWI**: A secondary land form parameter which estimates for each pixel, its tendency to accumulate water.
- **Slope**: Measured in degrees, is the first derivative of elevation in the direction of greatest slope.

Details

The area in question is the Hunter Wine Country Private Irrigation District (HWCPID), situated in the Lower Hunter Valley, NSW (32.83S 151.35E), and covers an area of approximately 220 km². The HWCPID is approximately 140 km north of Sydney, NSW, Australia. Climatically, the HWCPID is situated in a temperate climatic zone, and experiences warm humid summers, and relatively cooler yet also humid winters. Rainfall is mostly uniformly distributed throughout the year. On average the HWCPID receives just over 750 mm of rainfall annually. In terms of land use, an expansive viticultural industry is situated in the area and is possibly most widespread of rural industries, followed by dry land agricultural grazing systems.

Note

This RasterStack is used in the Use R for Digital Soil Mapping manual in the section about quantification of uncertainties.

References

- Gallant, J. C., Dowling, T. I., (2003) **A multiresolution index of valley bottom flatness for mapping depositional areas**. *Water Resources Research* 39(12), 1347.
- Malone, B.P., Hughes, P., McBratney, A.B., Minasny, B. (2014) **A model for the identification of terrons in the Lower Hunter Valley, Australia**. *Geoderma Regional* 1, 31-47.
- This data is copyright of the Soil Security Lab, The University of Sydney. If any part of this data is to be used in any publication or report, please provide a citation: Soil Security Laboratory, 2015. *Use R for Digital Soil Mapping Manual*. The University of Sydney, Sydney, Australia.

Examples

```
library(raster)
data(hunterCovariates_sub)

#plot 1 of the rasters in the \code{RasterStack}
plot(hunterCovariates_sub[["Elevation"]])
```

Hunter Valley subsoil pH points

Hunter Valley subsoil pH data with environmental covariates

Description

A dataframe carrying point coordinates for 506 observations of soil pH from the Lower Hunter Valley, NSW, Australia. The depth interval of the observation is 60-100cm. Together with the soil pH information is environmental covariate data that have been intersected with the point data. The environmental covariates have been sourced from a digital elevation model and Landsat 7 spectral band reflectance.

Usage

```
data(HV_subsoilpH)
```

Format

HV_subsoilpH is 506 row dataframe with the first 2 columns indicating the projected coordinate locations. The coordinate reference system is WGS 84 Zone 56. Soil pH for the 60-100cm depth interval is recorded in the following column. A suite of intersected environmental covariate data fills the remaining columns. This environmental covariate information refers to:

- **Terrain_Ruggedness_Index:** A quantitative measure of topographic heterogeneity. High values indicate terrain is rugged or heterogeneous.
- **AACN:** Difference between elevation and an interpolation of a channel network base level elevation. Knowledge of the spatial distribution of channel networks (lines) is therefore necessary for this parameter.
- **Landsat_Band1:** Earth surface reflectance information derived from Landsat 7 ETM+ satellite. Band 1 spectral range is .45 to .52 microns (visible blue).
- **Elevation:** Meters above sea level; derived from a digital elevation model.
- **Hillshading:** Analytic hill shading derived from digital elevation model and a fixed sun degree angle.
- **Light_insolation:** Measure of potential incoming solar radiation, and used as a parameter for evaluating the positional aspect effect. Derived from digital elevation model, this parameter was evaluated over the duration of a single calendar year with a 5 day time step.

- Mid_Slope_Positon: A relative slope position parameter which gives a classification of the slope position in both valley and crest positions.
- MRVBF: Multi-resolution valley bottom flatness is derived using slope and elevation to classify valley bottoms as flat, low areas (Gallant and Dowling 2003). This is accomplished through a series of neighborhood operations at progressively coarser resolutions with the goal of identifying both small and large valleys. MRVBF has been used extensively for the delineation and grading of valley floor units corresponding to areas of alluvial and colluvial deposits. High values of MRVBF indicate relatively low, flat areas of the landscape.
- NDVI: Normalized difference vegetation index. Derived from Landsat 7 data $(B4-B3)/(B4+B3)$. High values indicate actively growing vegetation.
- TWI: A secondary land form parameter which estimates for each pixel, its tendency to accumulate water.
- Slope: Measured in degrees, is the first derivative of elevation in the direction of greatest slope.

Details

The area in question is the Hunter Wine Country Private Irrigation District (HWCPID), situated in the Lower Hunter Valley, NSW (32.83S 151.35E), and covers an area of approximately 220 km². The HWCPID is approximately 140 km north of Sydney, NSW, Australia. Climatically, the HWCPID is situated in a temperate climatic zone, and experiences warm humid summers, and relatively cooler yet also humid winters. Rainfall is mostly uniformly distributed throughout the year. On average the HWCPID receives just over 750 mm of rainfall annually. In terms of land use, an expansive viticultural industry is situated in the area and is possibly most widespread of rural industries, followed by dry land agricultural grazing systems.

Note

This data set is used in the Use R for Digital Soil Mapping manual in the section about quantification of uncertainties.

References

- Gallant, J. C., Dowling, T. I., (2003) [A multiresolution index of valley bottom flatness for mapping depositional areas](#). *Water Resources Research* 39(12), 1347.
- Malone, B.P., Hughes, P., McBratney, A.B., Minasny, B. (2009) [A model for the identification of terrons in the Lower Hunter Valley, Australia](#). *Geoderma Regional* 1, 31-47.
- This data is copyright of the Soil Security Lab, The University of Sydney. If any part of this data is to be used in any publication or report, please provide a citation: Soil Security Laboratory, 2015. Use R for Digital Soil Mapping Manual. The University of Sydney, Sydney, Australia.

Examples

```
data(HV_subsoilpH)
summary(HV_subsoilpH)
```

tangler	<i>Spatial point patterns and raster objects anonymized with a given detangler object.</i>
---------	--

Description

For a given detangler object, spatial coordinates are anonymized accordingly.

Usage

```
tangler(data=NULL, tanglerInfo = NULL, raster_object = FALSE, stub = NULL,
saveTangles = FALSE, path = NULL)
```

Arguments

data	Can be either a 2 column matrix of spatial coordinates or a raster object.
tanglerInfo	output from tangles which encodes the steps and values of the anonymization process.
raster_object	logical; If TRUE the data to be anonymized is a raster object. This means the data associated with the raster are carried through to the outputs.
stub	Character; A string of user-defined setting for establishing naming conventions.
saveTangles	logical; Save function output to file. Default is FALSE, but it is recommended to always save outputs.
path	character; Path to where outputs are saved.

Value

The transformed coordinates or raster object (dependent on input data).

Note

Outputs can also saved to any specified directory with the hash key associated with the anonymization and user-defined stub in the file name. This will be saved as an .rds file that will also have tanglerXY in the file name.

tangles is coordinate reference system agnostic. Anonymization will proceed regardless of whether data is in geographic or projected space. Anonymized coordinates will not exist in reality, but the spatial properties of the original data will be maintained.

Author(s)

Brendan Malone

References

- CM O'Keefe, S Otorepec, M Elliot, E Mackey, and K O'Hara (2017) [The De-Identification Decision Making Framework](#). CSIRO Reports EP173122 and EP175702.

Examples

```
## POINT DATA tangles
library(digest)
data("HV_subsoilpH")
str(HV_subsoilpH)
dat.xy<- HV_subsoilpH[,1:2]
xyData<- as.matrix(dat.xy)

# anonymize with 5 levels of abstraction
tangles.out<- tangles(data = xyData,
depth = 5,
rasterdata = TRUE,
raster_object = FALSE,
saveTangles = FALSE)

# tangle associated RASTER DATA
library(raster)
data("hunterCovariates_sub")

tangled.origi<- tangler(data = hunterCovariates_sub,
tanglerInfo = tangles.out[[2]],
raster_object = TRUE,
stub = "hv1",
saveTangles = FALSE)
```

tangles

Anonymize spatial point patterns and raster objects

Description

Spatial coordinates are anonymized by using three possible techniques of altering locations without loss of the spatial relationships between data. The three techniques are vertical, horizontal and rotational shifts. User defines how many steps to use in the anonymization. The actual steps taken are randomized and saved for later use.

Usage

```
tangles(data=NULL, depth=3, rasterdata = FALSE, raster_object = FALSE,
saveTangles = FALSE, path = NULL)
```

Arguments

data	Can be either a 2 column matrix of spatial coordinates or a raster object.
depth	numeric; the number of anonymization steps to take (levels of abstraction). Default is 3.

rasterdata	logical; If TRUE the rotational degrees will be taken as either 90, 180 or 270 degrees rather than randomly selected. This is to preserve the properties of a raster if raster data is to be anonymized.
raster_object	logical; If TRUE the data to be anonymized is a raster object. This means the data associated with the raster are carried through to the outputs.
saveTangles	logical; Save function output to file. Default is FALSE, but it is recommended to always save outputs.
path	character; Path to where outputs are saved.

Value

A list object that contains: 1. The transformed coordinates or transformed raster object (dependent on the inputs). 2. A separate list object to be used for untangling the transformed coordinates. The outputs can be written to file to a specified directory with file stub names of tangledXY and detangler respectively. These files have a common hash key as part of their file name. The hash key is generated from the detangler object using the sha256 hash algorithm

Note

Raster data can in practice be rotated for any given angle, yet the linear horizontal and vertical properties of the raster will be lost, meaning that no software will be able to recognize it as a raster. This why there is the rasterdata parameter exists, and can be used for both raster and non-raster data. It just constrains the possible degree values that can be used in the anonymization. If raster data is being anonymized both rasterdata and raster_object will be set to TRUE in most cases.

The most important output to come from this function is the detangler object (saved to file with unique hash key), as this encodes the anonymization steps that can be used for re-identification, or anonymization of associated data. This enables parties to share pertinent information without the requirement to share actual spatial coordinates.

tangles is coordinate reference system agnostic. Anonymization will proceed regardless of whether data is in geographic or projected space. Anonymized coordinates will not exist in reality, but the spatial properties of the original data will be maintained.

Author(s)

Brendan Malone

References

- CM O'Keefe, S Otorespec, M Elliot, E Mackey, and K O'Hara (2017) [The De-Identification Decision Making Framework](#). CSIRO Reports EP173122 and EP175702.

Examples

```
## POINT DATA
library(digest)
data("HV_subsoilpH")
str(HV_subsoilpH)
dat.xy<- HV_subsoilpH[,1:2]
xyData<- as.matrix(dat.xy)
```

```
# anonymize with 5 levels of abstraction
tangles.out<- tangles(data = xyData,
depth = 5,
rasterdata = FALSE,
raster_object = FALSE,
saveTangles = FALSE)

str(tangles.out)
head(tangles.out[[1]])

## RASTER DATA
library(raster)
data("hunterCovariates_sub")
str(hunterCovariates_sub)
raster_object<- hunterCovariates_sub

tangles.out<- tangles(data = hunterCovariates_sub,
depth = 5,
rasterdata = TRUE,
raster_object = TRUE,
saveTangles = FALSE)

str(tangles.out)
```

Index

*Topic **datasets**

detangler_data, [2](#)

Hunter Valley covariates (subset area), [4](#)

Hunter Valley subsoil pH points, [6](#)

*Topic **methods**

detangles, [2](#)

tangler, [8](#)

tangles, [9](#)

detangler_data, [2](#)

detangles, [2](#)

Hunter Valley covariates (subset area), [4](#)

Hunter Valley subsoil pH points, [6](#)

hunterCovariates_sub(Hunter Valley covariates (subset area)), [4](#)

HV_subsoilpH(Hunter Valley subsoil pH points), [6](#)

tangler, [8](#)

tangles, [9](#)