

Package ‘sandwichr’

January 27, 2022

Version 1.0.0

Date 2022-01-25

Title Spatial Interpolation Based on Spatial Stratified Heterogeneity

Author Yue Lin [aut, cre],
Chengdong Xu [aut, ctb],
Jinfeng Wang [aut, ctb]

Maintainer Yue Lin <lin.3326@osu.edu>

Depends R (>= 3.5.0)

Imports sf, ggplot2, gridExtra, geodetector, caret, lwgeom, tools, dplyr

Suggests rmarkdown, knitr, MASS, ggpubr

Description Spatial interpolation is a common practice in social and environmental science. This package enables the implementation of SSH-based spatial interpolation in R. It provides functions to output the summarized statistics, the interpolated surface, and the standard error and confidence interval maps.

License GPL (>= 2)

URL https://github.com/linyuehzzz/sandwich_spatial_interpolator/tree/master/r/sandwichr

BugReports https://github.com/linyuehzzz/sandwich_spatial_interpolator/issues

LazyData false

RoxygenNote 7.1.2

VignetteBuilder knitr

Encoding UTF-8

R topics documented:

autoplot	2
bc.data	3
hs.data	3
load.data.shp	4
load.data.txt	4
sandwich.ci	5
sandwich.cv	6
sandwich.model	6
sandwichr	7

ssh.data.shp	8
ssh.data.txt	9
ssh.test	9

Index	11
--------------	-----------

autoplot	<i>Plot interpolation results</i>
----------	-----------------------------------

Description

Plot the estimated mean, standard error, and confidence interval for each reporting unit.

Usage

```
## S3 method for class 'sandwich.ci'
autoplot(
  object,
  titles = c("Confidence Interval (Lower Bound)", "Confidence Interval (Upper Bound)"),
  labels = c("", ""),
  border_color = "darkgray",
  poly_fill_ci = c("white", "red"),
  ...
)

## S3 method for class 'sandwich.model'
autoplot(
  object,
  titles = c("Mean", "Standard Error"),
  labels = c("", ""),
  border_color = "darkgray",
  poly_fill_mean = c("white", "red"),
  poly_fill_sd = c("white", "blue"),
  ...
)
```

Arguments

object	A sandwich.ci object generated by sandwich.ci, or a sandwich.model object generated by sandwich.model.
titles	A list of texts for the titles.
labels	A list of texts for the legend labels.
border_color	Color for outlines of the polygons.
poly_fill_ci	A list of colors for low and high ends of the gradient.
...	Ignored.
poly_fill_mean	A list of colors for low and high ends of the gradient in the mapping of means.
poly_fill_sd	A list of colors for low and high ends of the gradient in the mapping of means.

References

Wang, J. F., Haining, R., Liu, T. J., Li, L. F., & Jiang, C. S. (2013). Sandwich estimation for multi-unit reporting on a stratified heterogeneous surface. *Environment and Planning A*, 45(10), 2515-2534.

Examples

```
library(ggplot2)
data(hs.data)
hs.sw <- sandwich.model(object=hs.data, sampling.attr="Population", type="shp")
ggplot2::autoplot(object=hs.sw)

hs.sw.ci <- sandwich.ci(object=hs.sw, level=.95)
ggplot2::autoplot(object=hs.sw.ci)
```

bc.data

Breast cancer incidence in mainland China

Description

The bc.data dataset consists of two csv data frames:

- bc_sampling_ssh: Breast cancer incidence at 242 sampling units in mainland China, where the SSH stratum ("SSHID") and reporting unit ("GBCODE") that each sample falls into are specified. The sampling attribute is "Incidence".
- bc_reporting_ssh: The county-level administrative divisions in mainland China ("GBCODE"), where the weights of each intersecting stratum ("W1" and "W2") are specified.

hs.data

Heshun human population

Description

The hs.data dataset consists of three data frames:

- hs.sampling: Human population of 167 sampling villages within the Heshun region, China, which serves as the sampling layer.
- hs.ssh: A zonation map with 5 strata that serves as a candidate SSH layer.
- hs.reporting: The administrative division of 10 townships in Heshun county that served as the reporting layer.

References

Li, L., Wang, J., & Wu, J. (2012). A spatial model to predict the incidence of neural tube defects. *BMC Public Health*, 12(1), 951.

load.data.shp	<i>Load shapefiles into sf</i>
---------------	--------------------------------

Description

load.data.shp converts shapefiles into a list of sf objects.

Usage

```
load.data.shp(sampling.file,  
              ssh.file,  
              reporting.file)
```

Arguments

sampling.file Name of the sampling layer. The sampling layer should be a point feature that contains an attribute that is to be interpolated.

ssh.file Name of the SSH layer. The SSH layer should be a polygon feature.

reporting.file Name of the reporting layer. The reporting layer should be a polygon feature.

Examples

```
hs.sampling.name <- system.file("extdata", "hs.sampling.shapefile.shp",  
                               package="sandwichr")  
hs.ssh.name <- system.file("extdata", "hs.ssh.shapefile.shp",  
                           package="sandwichr")  
hs.reporting.name <- system.file("extdata", "hs.reporting.shapefile.shp",  
                                package="sandwichr")  
  
hs.data <- load.data.shp(sampling.file=hs.sampling.name,  
                        ssh.file=hs.ssh.name,  
                        reporting.file=hs.reporting.name)
```

load.data.txt	<i>Load text files into data frames</i>
---------------	---

Description

load.data.txt converts text files into a list of data frames.

Usage

```
load.data.txt(sampling_ssh.file,  
              reporting_ssh.file)
```

Arguments

sampling_ssh.file

Name of the file linking sampling and SSH layers. Each row denotes a sampling unit. At least two columns need to be included in this file: value of the sampling attribute and the stratum each sampling unit falling into.

reporting_ssh.file

Name of the file linking reporting and SSH layers. Each row denotes a reporting unit. The weights of each stratum should be specified in the columns.

Examples

```
bc.sampling_ssh.name <- system.file("extdata", "bc_sampling_ssh.csv",
                                   package="sandwichr")
bc.reporting_ssh.name <- system.file("extdata", "bc_reporting_ssh.csv",
                                   package="sandwichr")

bc.data <- load.data.txt(sampling_ssh.file=bc.sampling_ssh.name,
                        reporting_ssh.file=bc.reporting_ssh.name)
```

sandwich.ci

Calculate confidence interval for the mean

Description

sandwich.ci calculates the confidence interval for the mean value of each reporting unit.

Usage

```
sandwich.ci(object,
            level=.95)
```

Arguments

object An sf object or a data frame generated from the sandwich.model function.

level The confidence level required. By default, level = .95.

Examples

```
data(hs.data)
hs.sw <- sandwich.model(object=hs.data, sampling.attr="Population", type="shp")
hs.sw.ci <- sandwich.ci(object=hs.sw, level=.95)
```

sandwich.cv	<i>Perform k-fold cross validation</i>
-------------	--

Description

sandwich.cv performs k -fold cross validation to evaluate the overall model accuracy and outputs the average root mean square error (RMSE).

Usage

```
sandwich.cv(object,
             sampling.attr,
             k=10,
             type="shp",
             ssh.id.col=NULL,
             reporting.id.col=NULL,
             ssh.weights=NULL)
```

Arguments

object	When type="shp", object is a list of 3 sf objects generated by load.data.shp, including a POINT sf object used as the sampling layer, a POLYGON sf object used as the SSH layer, and a POLYGON sf object used as the SSH layer. When type="txt", object is a list of 2 data frames generated by load.data.txt, including a file linking sampling and SSH layers linking reporting and SSH layers and a file.
sampling.attr	A string denoting the name of the attribute in the sampling layer to be interpolated.
k	The number of folds ($k > 1$). By default, $k = 10$.
type	A string denoting the type of input data. type="shp" denotes shapefiles, and type="txt" denotes text files. By default, type="shp".
ssh.id.col	A string denoting the column that specifies which stratum each sampling unit falls into in the file linking sampling and SSH layers.
reporting.id.col	A string denoting the column that specifies which reporting unit each sampling unit falls into in the file linking sampling and SSH layers.
ssh.weights	A list that specifies the strata in the SSH layer and their corresponding columns of weights in the file linking reporting and SSH layers.

sandwich.model	<i>Perform Sandwich model-based mapping</i>
----------------	---

Description

sandwich.model estimate the mean value of an attribute and its standard error for each reporting unit using the sandwich estimation framework by Wang et al. (2013).

Usage

```
sandwich.model(object,
               sampling.attr,
               type="shp",
               ssh.id.col=NULL,
               ssh.weights=NULL)
```

Arguments

object	When type="shp", object is a list of 3 sf objects generated by load.data.shp, including a POINT sf object used as the sampling layer, a POLYGON sf object used as the SSH layer, and a POLYGON sf object used as the SSH layer. When type="txt", object is a list of 2 data frames generated by load.data.txt, including a file linking sampling and SSH layers linking reporting and SSH layers and a file.
sampling.attr	A string denoting the name of the attribute in the sampling layer to be interpolated.
type	A string denoting the type of input data. type="shp" denotes shapefiles, and type="txt" denotes text files. By default, type="shp".
ssh.id.col	A string denoting the column that specifies which stratum each sampling unit falls into in the file linking sampling and SSH layers.
ssh.weights	A list that specifies the strata in the SSH layer and their corresponding columns of weights in the file linking reporting and SSH layers.

References

Wang, J. F., Haining, R., Liu, T. J., Li, L. F., & Jiang, C. S. (2013). Sandwich estimation for multi-unit reporting on a stratified heterogeneous surface. *Environment and Planning A*, 45(10), 2515-2534.

Examples

```
data(hs.data)
hs.sw <- sandwich.model(object=hs.data, sampling.attr="Population", type="shp")
```

sandwichr	sandwichr: <i>Spatial Interpolation Based on Spatial Stratified Heterogeneity</i>
-----------	---

Description

The package sandwichr provides tools to interpolate a spatially stratified heterogeneous population with high accuracy, even when its spatial autocorrelation is weak. These tools allow the calculation of critical summarized statistics such as geodetector q . They also enable the visualization of the interpolated surface as well as the standard error and confidence interval maps.

Author(s)

Yue Lin, Chengdong Xu, and Jinfeng Wang

References

Wang, J. F., Haining, R., Liu, T. J., Li, L. F., & Jiang, C. S. (2013). Sandwich estimation for multi-unit reporting on a stratified heterogeneous surface. *Environment and Planning A*, 45(10), 2515-2534.

ssh.data.shp	<i>Prepare shapefile data for Geodetector</i>
--------------	---

Description

ssh.data.shp joins a candidate SSH layer to an existing layer.

Usage

```
ssh.data.shp(object,
              ssh.lyr,
              ssh.id,
              ssh.name=NULL)
```

Arguments

object	A POINT sf object to be joint. Its geometry should be consistent with the sampling layer, but it may have additional attributes from the SSH layer(s).
ssh.lyr	A POLYGON sf object used as the candidate SSH layer. Its strata ID will be linked to object.
ssh.id	A string denoting the attribute name of strata ID in the SSH layer.
ssh.name	A string denoting the new attribute name assigned to the strata ID in the output object. If NULL, the attribute name will be the same as that in the old SSH layer.

See Also

[<https://cran.r-project.org/package=geodetector>]geodetector R package

Examples

```
library(sf)
library(tools)
data(hs.data)
hs.ssh2.name <- system.file("extdata", "hs.ssh2.shapefile.shp",
                           package="sandwichr")
hs.ssh2 <- read_sf(dsn=dirname(hs.ssh2.name),
                  layer=file_path_sans_ext(basename(hs.ssh2.name)))
hs.join <- ssh.data.shp(object=hs.data[[1]], ssh.lyr=hs.data[[2]], ssh.id="STR_1")
hs.join <- ssh.data.shp(object=hs.join, ssh.lyr=hs.ssh2, ssh.id="STR_2")
```

ssh.data.txt	<i>Prepare text data for Geodetector</i>
--------------	--

Description

ssh.data.txt prepare the input from text files for the geographical detector model.

Usage

```
ssh.data.txt(object)
```

Arguments

object	A list of 2 data frames generated by load.data.txt, including a file linking sampling and SSH layers linking reporting and SSH layers and a file.
--------	---

See Also

[<https://cran.r-project.org/package=geodetector>]geodetector R package

Examples

```
data(bc.data)
bc.join <- ssh.data.txt(object=bc.data)
```

ssh.test	<i>Calculate Geodetector factor detector q-statistic and interaction detector</i>
----------	---

Description

ssh.test calculates the factor detector q -statistic and the interaction detector in the geographical detector model described by Wang et al. (2010). The q -statistic measures the SSH of the sampling attribute in terms of a given stratification, which can be used for the selection of an SSH layer for Sandwich model-based mapping. The interactive effects indicate whether a combination of two stratifications enhances the SSH of the sampling attribute.

Usage

```
ssh.test(object,
          y,
          x,
          test="factor",
          type="shp")
```

Arguments

object	A POINT sf object linking the attributes of the sampling layer and the strata ID in the SSH layer(s) generated from <code>ssh.data</code> .
y	A string denoting the name of the explained variable (sampling attribute) in object.
x	A string denoting the name(s) of the explanatory variable(s) (stratification(s)) in object.
test	A string denoting the type of test. <code>test="factor"</code> denotes the factor detector, and <code>test="interaction"</code> denotes the interaction detector. By default, <code>test="factor"</code> .
type	A string denoting the type of input data. <code>type="shp"</code> denotes shapefiles, and <code>type="txt"</code> denotes text files. By default, <code>type="shp"</code> .

References

Wang, J. F., Li, X. H., Christakos, G., Liao, Y. L., Zhang, T., Gu, X., & Zheng, X. Y. (2010). Geographical detectors-based health risk assessment and its application in the neural tube defects study of the Heshun Region, China. *International Journal of Geographical Information Science*, 24(1), 107-127.

See Also

[<https://cran.r-project.org/package=geodetector>]geodetector R package

Examples

```
library(sf)
library(tools)
data(hs.data)
hs.ssh2.name <- system.file("extdata", "hs.ssh2.shapefile.shp",
                           package="sandwichr")
hs.ssh2 <- read_sf(dsn=dirname(hs.ssh2.name),
                  layer=file_path_sans_ext(basename(hs.ssh2.name)))
hs.join <- ssh.data.shp(object=hs.data[[1]], ssh.lyr=hs.data[[2]], ssh.id="STR_1")
hs.join <- ssh.data.shp(object=hs.join, ssh.lyr=hs.ssh2, ssh.id="STR_2")
ssh.test(object=hs.join, y="Population", x=c("STR_1", "STR_2"), test="factor")
```

Index

`autoplot`, [2](#)

`bc.data`, [3](#)

`hs.data`, [3](#)

`load.data.shp`, [4](#)

`load.data.txt`, [4](#)

`sandwich.ci`, [5](#)

`sandwich.cv`, [6](#)

`sandwich.model`, [6](#)

`sandwichr`, [7](#)

`ssh.data.shp`, [8](#)

`ssh.data.txt`, [9](#)

`ssh.test`, [9](#)