Package: gdverse (via r-universe)

July 7, 2024

Title Geographical Detector Models

Version 0.3.0

Description Family of geographical detector models in R.

License GPL-3

Encoding UTF-8

Roxygen list(markdown = TRUE)

RoxygenNote 7.3.2

Imports classInt, dplyr, forcats, geosphere, ggplot2, kableExtra, magrittr, parallel, patchwork, purrr, reticulate, rpart, scatterpie, stats, tibble, tidyr, utils

URL https://spatlyu.github.io/gdverse/, https://github.com/SpatLyu/gdverse

Suggests cowplot, knitr, PieGlyph, Rdimtools, rmarkdown, sf, terra, tidyterra, tidyverse

VignetteBuilder knitr

Depends R (>= 4.1.0)

LazyData true

Repository https://spatlyu.r-universe.dev

RemoteUrl https://github.com/SpatLyu/gdverse

RemoteRef HEAD

RemoteSha 6df0a635ba6ea1870788d524414e79e5e3d59b46

Contents

`cpsd_disc` *optimal spatial data discretization based on SPADE q-statistics*

Description

Function for determining the optimal spatial data discretization based on SPADE q-statistics.

Usage

```
cpsd_disc(
  formula,
  data,
  wt,
  discnum = NULL,
  discmethod = NULL,
  strategy = 2L,
  increase_rate = 0.05,
  cores = 1,
  return_disc = TRUE,
  seed = 123456789,
  ...
)
```

Arguments

`formula`	A formula of optimal spatial data discretization.
`data`	A data.frame or tibble of observation data.
`wt`	The spatial weight matrix.
`discnum`	(optional) A vector of number of classes for discretization. Default is `3:22`.
`discmethod`	(optional) A vector of methods for discretization,default is all using `quantile`in `gdverse`. see `method` parameter in `st_unidisc()` for more details.
`strategy`	(optional) Discretization strategy. When `strategy` is `1L`, choose the highest SPADE model q-statistics to determinate optimal spatial data discretization parameters. When `strategy` is `2L`, The optimal discrete parameters of spatial data are selected by combining LOESS model.
`increase_rate`	(optional) The critical increase rate of the number of discretization. Default is `5%`.

cores	(optional) A positive integer(default is 1). If cores > 1, a 'parallel' package cluster with that many cores is created and used. You can also supply a cluster object.
return_disc	(optional) Whether or not return discretized result used the optimal parameter. Default is TRUE.
seed	(optional) Random seed number, default is 123456789.Setting random seed is useful when the sample size is greater than 3000(the default value for largeN) and the data is discretized by sampling 10%(the default value for samp_prop in st_unidisc()).
...	(optional) Other arguments passed to st_unidisc().

Value

A list with the optimal parameter in the provided parameter combination with k, method and disc(when return_disc is TRUE).

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

References

Yongze Song & Peng Wu (2021) An interactive detector for spatial associations, International Journal of Geographical Information Science, 35:8, 1676-1701, DOI:10.1080/13658816.2021.1882680

Examples

```
## Not run:
library(sf)
usfi = read_sf(system.file('extdata/USFI_Xian.gpkg',package = 'gdverse')) |>
  dplyr::select(dplyr::all_of(c("NDVI","BH","SUHI")))
coord = usfi |>
  st_centroid() |>
  st_coordinates()
wt = inverse_distance_weight(coord[,1],coord[,2])
usfi = st_drop_geometry(usfi)
cpsd_disc(SUHI ~ NDVI + BH, data = usfi,wt = wt,cores = 6)

## End(Not run)
```

---

cpsd_spade *compensated power of spatial determinant(CPSD)*

Description

Function for calculate compensated power of spatial determinant Q_s.

Usage

```
cpsd_spade(yobs, xobs, xdisc, wt)
```

Arguments

yobs	Variable Y
xobs	The original un-discretized covariable X.
xdisc	The discretized covariable X.
wt	The spatial weight matrix.

Details

The power of compensated spatial determinant formula is $Q_s = \frac{q_s}{q_{s_{inforkep}}} = \frac{1-\frac{\sum_{h=1}^{L} N_h \Gamma_{kdep}}{N\Gamma_{totaldep}}}{1-\frac{\sum_{h=1}^{L} N_h \Gamma_{hind}}{N\Gamma_{totalind}}}$

Value

A value of compensated power of spatial determinant Q_s.

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

References

Xuezhi Cang & Wei Luo (2018) Spatial association detector (SPADE),International Journal of Geographical Information Science, 32:10, 2055-2075, DOI: 10.1080/13658816.2018.1476693

Examples

```
## Not run:
library(sf)
usfi = read_sf(system.file('extdata/USFI_Xian.gpkg',package = 'gdverse')) |>
  dplyr::select(dplyr::all_of(c("NDVI","BH","SUHI")))
coord = usfi |>
  st_centroid() |>
  st_coordinates()
wt = inverse_distance_weight(coord[,1],coord[,2])
BH = usfi$BH
BH_disc = st_unidisc(usfi$BH,12)
SUHI = usfi$SUHI
cpsd_spade(SUHI,BH,BH_disc,wt)

## End(Not run)
```

ecological_detector *ecological detector*

Description

Compare the effects of two factors X_1 and X_2 on the spatial distribution of the attribute Y.

Usage

```
ecological_detector(y, x1, x2, alpha = 0.95)
```

Arguments

y	Dependent variable, continuous numeric vector.
x1	Covariable X_1, factor, character or discrete numeric.
x2	Covariable X_2, factor, character or discrete numeric.
alpha	(optional) Confidence level of the interval,default is 0.95.

Value

A list contains F statistics, P-values, and is there a significant difference between the two factors X_1 and X_2 on the spatial distribution of the attribute Y.

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

Examples

```
ecological_detector(y = 1:7,
                    x1 = c('x',rep('y',3),rep('z',3)),
                    x2 = c(rep('a',2),rep('b',2),rep('c',3)))
```

factor_detector *factor detector*

Description

The factor detector q-statistic measures the spatial stratified heterogeneity of a variable Y, or the determinant power of a covariate X of Y.

Usage

```
factor_detector(y, x)
```

Arguments

y	Variable Y, continuous numeric vector.
x	Covariable X, factor, character or discrete numeric.

Value

A list contains the Q-statistic and the P-value.

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

Examples

```
factor_detector(y = 1:7,x = c('x',rep('y',3),rep('z',3)))
```

F_informationloss *measure information loss by information entropy*

Description

Function for measure information loss by information entropy.

Usage

```
F_informationloss(xvar, xdisc)
```

Arguments

xvar	The original un-discretized vector.
xdisc	The discretized vector.

Details

The information loss measured by information entropy formula is $F = -\sum_{i=1}^{N} p_{(i)} \log_2 p_{(i)} - \left(-\sum_{h=1}^{L} p_{(h)} \log_2 p_{(h)}\right)$

Value

A numeric value of information loss measured by information entropy.

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

Examples

```
F_informationloss(1:7,c('x',rep('y',3),rep('z',3)))
```

---

gd *original geographical detector model*

Description

Function for original geographical detector model.

Usage

```
gd(formula, data, type = "factor", ...)
```

Arguments

formula	A formula of geographical detector model.
data	A data.frame or tibble of observation data.
type	(optional) The type of geographical detector,which must be one of factor(default), interaction, risk, ecological.
...	(optional) Specifies the size of the alpha (confidence level).Default is 0.95.

Value

A tibble of the corresponding result is stored under the corresponding detector type.

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

References

Jin-Feng Wang, Xin-Hu Li, George Christakos, Yi-Lan Liao, Tin Zhang, XueGu & Xiao-Ying Zheng (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, DOI: 10.1080/13658810802443457

Examples

```
gd(y ~ x1 + x2,
   tibble::tibble(y = 1:7,
                  x1 = c('x',rep('y',3),rep('z',3)),
                  x2 = c(rep('a',2),rep('b',2),rep('c',3))))

gd(y ~ x1 + x2,
   tibble::tibble(y = 1:7,
```

```
                x1 = c('x',rep('y',3),rep('z',3)),
                x2 = c(rep('a',2),rep('b',2),rep('c',3))),
   type = 'interaction')

gd(y ~ x1 + x2,
   tibble::tibble(y = 1:7,
                x1 = c('x',rep('y',3),rep('z',3)),
                x2 = c(rep('a',2),rep('b',2),rep('c',3))),
   type = 'risk',alpha = 0.95)

gd(y ~ x1 + x2,
   tibble::tibble(y = 1:7,
                x1 = c('x',rep('y',3),rep('z',3)),
                x2 = c(rep('a',2),rep('b',2),rep('c',3))),
   type = 'ecological',alpha = 0.95)
```

`gd_bestunidisc` *best univariate discretization based on geodetector q-statistic*

Description

Function for determining the best univariate discretization based on geodetector q-statistic.

Usage

```
gd_bestunidisc(
  formula,
  data,
  discnum = NULL,
  discmethod = NULL,
  cores = 1,
  return_disc = TRUE,
  seed = 123456789,
  ...
)
```

Arguments

`formula`	A formula of best univariate discretization.
`data`	A data.frame or tibble of observation data.
`discnum`	(optional) A vector of number of classes for discretization. Default is `3:22`.
`discmethod`	(optional) A vector of methods for discretization,default is using `c("sd","equal","pretty","quantile` `gdverse`.
`cores`	(optional) A positive integer(default is 1). If cores > 1, a 'parallel' package cluster with that many cores is created and used. You can also supply a cluster object.

return_disc	(optional) Whether or not return discretized result used the optimal parameter. Default is TRUE.
seed	(optional) Random seed number, default is 123456789.Setting random seed is useful when the sample size is greater than 3000(the default value for largeN) and the data is discretized by sampling 10%(the default value for samp_prop in st_unidisc()).
...	(optional) Other arguments passed to st_unidisc().

Value

A list with the optimal parameter in the provided parameter combination with k, method and disc(when return_disc is TRUE).

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

Examples

```
## Not run:
library(terra)
library(tidyverse)
fvcpath = "https://github.com/SpatLyu/rdevdata/raw/main/FVC.tif"
fvc = terra::rast(paste0("/vsicurl/",fvcpath))
fvc = terra::aggregate(fvc,fact = 5)
fvc = as_tibble(terra::as.data.frame(fvc,na.rm = T))
g = gd_bestunidisc(fvc ~ .,data = select(fvc,-lulc),discnum = 3:15,cores = 6)
g

## End(Not run)
```

---

generate_subsets *generate subsets of a set*

Description

generate subsets of a set

Usage

```
generate_subsets(set, empty = TRUE, self = TRUE)
```

Arguments

set	A vector including the empty set and the set itself. Default is TRUE.
empty	(optional) When empty is TRUE,the generated subset includes the empty set, otherwise the empty set is removed. Default is TRUE.
self	(optional) When self is TRUE,the resulting subset includes the set itself, otherwise the set itself is removed. Default is TRUE.

Value

A list with the subsets

Examples

```
generate_subsets(letters[1:3])
generate_subsets(letters[1:3],empty = FALSE)
generate_subsets(letters[1:3],self = FALSE)
generate_subsets(letters[1:3],empty = FALSE,self = FALSE)
```

---

gozh *geographically optimal zones-based heterogeneity(GOZH) model*

Description

Function for geographically optimal zones-based heterogeneity(GOZH) model

Usage

```
gozh(formula, data, cores = 1, type = "factor", alpha = 0.95, ...)
```

Arguments

`formula`	A formula of GOZH model.
`data`	A data.frame or tibble of observation data.
`cores`	(optional) A positive integer(default is 1). If cores > 1, a 'parallel' package cluster with that many cores is created and used. You can also supply a cluster object.
`type`	(optional) The type of geographical detector,which must be `factor`(default), `interaction`, `risk`, `ecological`.You can run one or more types at one time.
`alpha`	(optional) Specifies the size of confidence level.Default is `0.95`.
`...`	(optional) Other arguments passed to `rpart_disc()`.

Value

A list of GOZH model result.

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

References

Luo, P., Song, Y., Huang, X., Ma, H., Liu, J., Yao, Y., & Meng, L. (2022). Identifying determinants of spatio-temporal disparities in soil moisture of the Northern Hemisphere using a geographically optimal zones-based heterogeneity model. ISPRS Journal of Photogrammetry and Remote Sensing: Official Publication of the International Society for Photogrammetry and Remote Sensing (ISPRS), 185, 111–128. https://doi.org/10.1016/j.isprsjprs.2022.01.009

Examples

```
data('ndvi')
g = gozh(NDVIchange ~ ., data = ndvi)
g
```

`gozh_detector` *geographically optimal zones-based heterogeneity detector*

Description

Function for geographically optimal zones-based heterogeneity detector.

Usage

```
gozh_detector(formula, data, cores = 1, type = "factor", alpha = 0.95, ...)
```

Arguments

`formula`	A formula of GOZH detector.
`data`	A data.frame or tibble of observation data.
`cores`	(optional) A positive integer(default is 1). If cores > 1, a 'parallel' package cluster with that many cores is created and used. You can also supply a cluster object.
`type`	(optional) The type of geographical detector,which must be one of `factor`(default), `interaction`, `risk`, `ecological`.
`alpha`	(optional) Confidence level of the interval,default is `0.95`.
`...`	(optional) Other arguments passed to `rpart_disc()`.

Value

A list of the corresponding result is stored under the corresponding detector type.

Note

Only one type of detector is supported in a `gozh_detector()` run at a time.

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

References

Luo, P., Song, Y., Huang, X., Ma, H., Liu, J., Yao, Y., & Meng, L. (2022). Identifying determinants of spatio-temporal disparities in soil moisture of the Northern Hemisphere using a geographically optimal zones-based heterogeneity model. ISPRS Journal of Photogrammetry and Remote Sensing: Official Publication of the International Society for Photogrammetry and Remote Sensing (ISPRS), 185, 111–128. https://doi.org/10.1016/j.isprsjprs.2022.01.009

Examples

```
data('ndvi')
g = gozh_detector(NDVIchange ~ ., data = ndvi)
g
```

---

idsa *interactive detector for spatial associations(IDSA)*

Description

Function for interactive detector for spatial associations model.

Usage

```
idsa(
  formula,
  data,
  wt = NULL,
  overlaymethod = "and",
  locations = NULL,
  discnum = NULL,
  discmethod = NULL,
  strategy = 2L,
  increase_rate = 0.05,
  cores = 6,
  seed = 123456789,
  alpha = 0.95,
  ...
)
```

Arguments

formula	A formula of IDSA model.
data	A data.frame or tibble of observation data.
wt	(optional) The spatial weight matrix.When wt is not provided, must provide locations. And gdverse will use locations columns to construct spatial weight use inverse_distance_weight().
overlaymethod	(optional) Spatial overlay method. One of and, or, intersection. Default is and.
locations	(optional) The geospatial locations coordinate columns name which in data. Useful and must provided when wt is not provided. When wt is provided, locations is not need.
discnum	(optional) Number of multilevel discretization.Default will use 3:22.
discmethod	(optional) The discretization methods. Default all use quantile. More details to see st_unidisc().
strategy	(optional) Discretization strategy. When strategy is 1L, choose the highest SPADE model q-statistics to determinate optimal spatial data discretization parameters. When strategy is 2L, The optimal discrete parameters of spatial data are selected by combining LOESS model.
increase_rate	(optional) The critical increase rate of the number of discretization. Default is 5%.
cores	(optional) A positive integer(default is 1). If cores > 1, a 'parallel' package cluster with that many cores is created and used. You can also supply a cluster object.
seed	(optional) Random number seed, default is 123456789.
alpha	(optional) Specifies the size of confidence level.Default is 0.95.
...	(optional) Other arguments passed to cpsd_disc().

Value

A list with PID values tibble under different spatial overlays and performance evaluation indicators.

Note

The IDSA model requires at least $2^n - 1$ calculations when has n explanatory variables. When there are more than 10 explanatory variables, carefully consider the computational burden of this model. When there are a large number of explanatory variables, the data dimensionality reduction method can be used to ensure the trade-off between analysis results and calculation speed.

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

References

Yongze Song & Peng Wu (2021) An interactive detector for spatial associations, International Journal of Geographical Information Science, 35:8, 1676-1701, DOI:10.1080/13658816.2021.1882680

Examples

```
## Not run:
data('sim')
g = idsa(y ~ ., data = sim, locations = c('lo','la'),
         discvar = c("xa","xb","xc"), cores = 6)
g

## End(Not run)
```

`interaction_detector` *interaction detector*

Description

Identify the interaction between different risk factors, that is, assess whether factors X1 and X2 together increase or decrease the explanatory power of the dependent variable Y, or whether the effects of these factors on Y are independent of each other.

Usage

```
interaction_detector(y, x1, x2)
```

Arguments

`y`	Dependent variable, continuous numeric vector.
`x1`	Covariable X_1, `factor`, `character` or `discrete numeric`.
`x2`	Covariable X_2, `factor`, `character` or `discrete numeric`.

Value

A list contains the Q statistic when the factors X_1 and X_1 act on Y alone and the Q statistic when the two interact on Y together with the result type of the interaction detector.

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

Examples

```
interaction_detector(y = 1:7,
                     x1 = c('x',rep('y',3),rep('z',3)),
                     x2 = c(rep('a',2),rep('b',2),rep('c',3)))
```

`inverse_distance_weight` *calculate inverse distance weight*

Description

Function for calculate inverse distance weight.

Usage

```
inverse_distance_weight(locx, locy, power = 1, is_arc = FALSE)
```

Arguments

`locx`	The x axis location.
`locy`	The y axis location.
`power`	(optional) Default is 1. Set to 2 for gravity weights.
`is_arc`	(optional) FALSE (default) or TRUE, whether to compute arc distance.

Details

The inverse distance weight formula is $w_{ij} = 1/d_{ij}^{\alpha}$

Value

A inverse distance weight matrices with class of `matrix`.

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

Examples

```
x = 1:10
y = 1:10
inverse_distance_weight(x,y)
inverse_distance_weight(x,y,is_arc = TRUE)
```

`lesh` *locally explained heterogeneity(LESH) model*

Description

Function for locally explained heterogeneity model.

Usage

```
lesh(formula, data, cores = 1, ...)
```

Arguments

`formula`	A formula of LESH model.
`data`	A data.frame or tibble of observation data.
`cores`	(optional) A positive integer(default is 1). If cores > 1, a 'parallel' package cluster with that many cores is created and used. You can also supply a cluster object.
`...`	(optional) Other arguments passed to `rpart_disc()`.

Value

A list of LESH model result.

Note

The LESH model requires at least $2^n - 1$ calculations when has n explanatory variables. When there are more than 10 explanatory variables, carefully consider the computational burden of this model. When there are a large number of explanatory variables, the data dimensionality reduction method can be used to ensure the trade-off between analysis results and calculation speed.

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

References

Li, Y., Luo, P., Song, Y., Zhang, L., Qu, Y., & Hou, Z. (2023). A locally explained heterogeneity model for examining wetland disparity. International Journal of Digital Earth, 16(2), 4533–4552. https://doi.org/10.1080/17538947.2023.2271883

Examples

```
## Not run:
data('ndvi')
g = lesh(NDVIchange ~ ., data = ndvi, cores = 6)
g

## End(Not run)
```

`loess_optdiscnum` *determine optimal spatial data discretization for individual variables*

Description

Function for determining optimal spatial data discretization for individual variables based on locally estimated scatterplot smoothing (LOESS) model.

Usage

```
loess_optdiscnum(qvec, discnumvec, increase_rate = 0.05)
```

Arguments

`qvec`	A numeric vector of q statistics.
`discnumvec`	A numeric vector of break numbers corresponding to `qvec`.
`increase_rate`	(optional) The critical increase rate of the number of discretization. Default is 5%.

Value

A optimal number of spatial data discretization.

Note

When `increase_rate` is not satisfied by the calculation, `increase_rate*0.1` is used first. At this time, if `increase_rate*0.1` is not satisfied again, the discrete number corresponding to the highest Q-statistic is selected as a return.

Note that `gdverse` sorts `discnumvec` from smallest to largest and keeps `qvec` in one-to-one correspondence with `discnumvec`.

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

References

Yongze Song & Peng Wu (2021) An interactive detector for spatial associations, International Journal of Geographical Information Science, 35:8, 1676-1701, DOI:10.1080/13658816.2021.1882680

Examples

```
library(sf)
usfi = read_sf(system.file('extdata/USFI_Xian.gpkg',package = 'gdverse')) |>
  dplyr::select(dplyr::all_of(c("NDVI","BH","SUHI")))
3:10 %>%
purrr::map_dbl(\(.k) st_unidisc(usfi$NDVI,.k) %>%
               factor_detector(usfi$SUHI,.) %>%
               {.[[1]]}) %>%
 loess_optdiscnum(3:10)
```

---

loess_optscale *determine optimal spatial data analysis scale*

Description

Function for determining optimal spatial data analysis scale based on locally estimated scatter plot smoothing (LOESS) model.

Usage

```
loess_optscale(qvec, spscalevec, increase_rate = 0.05)
```

Arguments

qvec	A numeric vector of q statistics.
spscalevec	A numeric vector of spatial scales corresponding to qvec.
increase_rate	(optional) The critical increase rate of the number of discretization. Default is 5%.

Value

A optimal number of spatial scale

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

Examples

```
## Not run:
library(tidyverse)
fvcpath = "https://github.com/SpatLyu/rdevdata/raw/main/FVC.tif"
fvc = terra::rast(paste0("/vsicurl/",fvcpath))
fvc1000 = fvc %>%
  terra::as.data.frame(na.rm = T) %>%
  as_tibble()
fvc5000 = fvc %>%
```

```
  terra::aggregate(fact = 5) %>%
  terra::as.data.frame(na.rm = T) %>%
  as_tibble()
qv1000 = factor_detector(fvc1000$fvc,
                         st_unidisc(fvc1000$premax,10,'quantile'))[[1]]
qv5000 = factor_detector(fvc5000$fvc,
                         st_unidisc(fvc5000$premax,10,'quantile'))[[1]]
loess_optscale(c(qv1000,qv5000),c(1000,5000))

## End(Not run)
```

---

ndvi *dataset of NDVI changes and its influencing factors*

Description

dataset of NDVI changes and its influencing factors, modified from GD package.

Usage

```
ndvi
```

Format

ndvi: A tibble with 713 rows and 7 variables

NTDs *NTDs data*

Description

The data were obtained by preprocessing use sf and tidyverse.

Usage

```
NTDs
```

Format

NTDs: A tibble with 185 rows and 4 variable columns and 2 location columns, modified from geodetector package.

opgd *optimal parameters-based geographical detector(OPGD) model*

Description

Function for optimal parameters-based geographical detector(OPGD) model.

Usage

```
opgd(
  formula,
  data,
  discvar,
  discnum = NULL,
  discmethod = NULL,
  cores = 1,
  type = "factor",
  alpha = 0.95,
  ...
)
```

Arguments

formula	A formula of OPGD model.
data	A data.frame or tibble of observation data.
discvar	Name of continuous variable columns that need to be discretized.Noted that when formula has discvar, data must have these columns.
discnum	(optional) A vector of number of classes for discretization. Default is 3:22.
discmethod	(optional) A vector of methods for discretization,default is using c("sd","equal","pretty","quantile gdverse.
cores	(optional) A positive integer(default is 1). If cores > 1, a 'parallel' package cluster with that many cores is created and used. You can also supply a cluster object.
type	(optional) The type of geographical detector,which must be factor(default), interaction, risk, ecological.You can run one or more types at one time.
alpha	(optional) Specifies the size of confidence level.Default is 0.95.
...	(optional) Other arguments passed to gd_bestunidisc().A useful parameter is seed, which is used to set the random number seed.

Value

A list of the OPGD model result.

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

References

Song, Y., Wang, J., Ge, Y. & Xu, C. (2020) An optimal parameters-based geographical detector model enhances geographic characteristics of explanatory variables for spatial heterogeneity analysis: Cases with different types of spatial data, GIScience & Remote Sensing, 57(5), 593-610. doi: 10.1080/15481603.2020.1760434.

Examples

```
## Not run:
library(terra)
library(tidyverse)
fvcpath = "https://github.com/SpatLyu/rdevdata/raw/main/FVC.tif"
fvc = terra::rast(paste0("/vsicurl/",fvcpath))
fvc = terra::aggregate(fvc,fact = 5)
fvc = as_tibble(terra::as.data.frame(fvc,na.rm = T))
opgd(fvc ~ ., data = fvc,
     discvar = names(select(fvc,-c(fvc,lulc))),
     cores = 6, type =c('factor','interaction'))

## End(Not run)
```

---

pid_idsa *IDSA Q-saistics* PID

Description

IDSA Q-saistics PID

Usage

```
pid_idsa(formula, rawdata, discdata, wt, overlaymethod = "and")
```

Arguments

formula	A formula for IDSA Q-saistics
rawdata	Raw observation data
discdata	Discrete explanatory variables data
wt	Spatial weight matrix
overlaymethod	(optional) Spatial overlay method. One of and, or, intersection. Default is and.

Details

$Q_{IDSA} = \frac{\theta_r}{\phi}$

Value

The value of IDSA Q-saistics PID.

Examples

```
## Not run:
library(sf)
usfi = read_sf(system.file('extdata/USFI_Xian.gpkg',package = 'gdverse')) |>
  dplyr::select(dplyr::all_of(c("NDVI","BH","WAR","SUHI")))
coord = usfi |>
  st_centroid() |>
  st_coordinates()
wt = inverse_distance_weight(coord[,1],coord[,2])
usf = usfi |>
  st_drop_geometry() |>
  dplyr::mutate(dplyr::across(1:3,\(.x) st_unidisc(.x,12)))
pid_idsa('NDVI~.',rawdata = usfi,discdata = usf,wt = wt)

## End(Not run)
```

---

plot.ecological_detector

plot ecological detector

Description

S3 method to plot output for ecological detector in `gd()`.

Usage

```
## S3 method for class 'ecological_detector'
plot(x, ...)
```

Arguments

`x`	Return by `gd()`.
`...`	(optional) Other arguments passed to `ggplot2::theme()`.

Value

A ggplot2 layer

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

plot.factor_detector *plot factor detector result*

Description

S3 method to plot output for factor detector in gd().

Usage

```
## S3 method for class 'factor_detector'
plot(x, slicenum = 2, alpha = 0.95, ...)
```

Arguments

x	Return by gd().
slicenum	(optional) The number of labels facing inward. Default is 2.
alpha	(optional) Confidence level.Default is 0.95.
...	(optional) Other arguments passed to ggplot2::theme().

Value

A ggplot2 layer.

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

plot.idsa_result *plot IDSA risk result*

Description

S3 method to plot output for IDSA risk result in idsa().

Usage

```
## S3 method for class 'idsa_result'
plot(x, ...)
```

Arguments

x	Return by idsa().
...	(optional) Other arguments passed to ggplot2::theme().

Value

A ggplot2 layer

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

`plot.interaction_detector` *plot interaction detector result*

Description

S3 method to plot output for interaction detector in `gd()`.

Usage

```
## S3 method for class 'interaction_detector'
plot(x, alpha = 1, ...)
```

Arguments

`x`	Return by `gd()`.
`alpha`	(optional) Picture transparency. Default is `1`.
`...`	(optional) Other arguments passed to `ggplot2::theme()`.

Value

A ggplot2 layer

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

`plot.lesh_result` *plot LESH model result*

Description

S3 method to plot output for LESH model interaction result in `lesh()`.

Usage

```
## S3 method for class 'lesh_result'
plot(x, pie = TRUE, scatter = FALSE, ..., scatter_alpha = 1)
```

Arguments

`x`	x Return by `lesh()`.
`pie`	(optional) Whether to draw the interaction contributions. Default is `TRUE`.
`scatter`	(optional) Whether to draw the interaction direction diagram. Default is `FALSE`.
`...`	(optional) Other arguments passed to `ggplot2::theme()`.
`scatter_alpha`	(optional) Picture transparency. Default is `1`.

Details

When both `scatter` and `pie` are set to `TRUE` in RStudio, enlarge the drawing frame for normal display.

Value

A ggplot2 layer.

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

Examples

```
## Not run:
data('ndvi')
g = lesh(NDVIchange ~ ., data = ndvi, cores = 6)
plot(g)

## End(Not run)
```

plot.risk_detector *plot risk detector*

Description

S3 method to plot output for risk detector in `gd()`.

Usage

```
## S3 method for class 'risk_detector'
plot(x, ...)
```

Arguments

`x`	Return by `gd()`.
`...`	(optional) Other arguments passed to `ggplot2::theme()`.

Value

A ggplot2 layer

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

plot.sesu_gozh *plot gozh sesu*

Description

S3 method to plot output for gozh sesu in `sesu_gozh()`.

Usage

```
## S3 method for class 'sesu_gozh'
plot(x, ...)
```

Arguments

`x`	Return by `sesu_gozh()`.
`...`	(optional) Other arguments passed to `ggplot2::theme()`.

Value

A ggplot2 layer.

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

---

plot.sesu_opgd *plot opgd sesu*

Description

S3 method to plot output for opgd sesu in `sesu_opgd()`.

Usage

```
## S3 method for class 'sesu_opgd'
plot(x, ...)
```

Arguments

`x`	Return by `sesu_opgd()`.
`...`	(optional) Other arguments passed to `ggplot2::theme()`.

Value

A ggplot2 layer.

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

---

plot.spade_result *plot SPADE power of spatial and multilevel discretization determinant*

Description

S3 method to plot output for SPADE power of spatial and multilevel discretization determinant from `spade()`.

Usage

```
## S3 method for class 'spade_result'
plot(x, slicenum = 2, alpha = 0.95, ...)
```

Arguments

x	Return by spade().The number of labels facing inward.
slicenum	(optional) The number of labels facing inward. Default is 2.
alpha	(optional) Confidence level.Default is 0.95.
...	(optional) Other arguments passed to ggplot2::theme().

Value

A ggplot2 layer.

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

---

print.ecological_detector *print ecological detector*

Description

S3 method to format output for ecological detector in gd().

Usage

```
## S3 method for class 'ecological_detector'
print(x, ...)
```

Arguments

x	Return by gd().
...	(optional) Other arguments passed to knitr::kable().

Value

Formatted string output

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

print.factor_detector *print factor detector*

Description

S3 method to format output for factor detector in `gd()`.

Usage

```
## S3 method for class 'factor_detector'
print(x, ...)
```

Arguments

`x`	Return by `gd()`.
`...`	(optional) Other arguments passed to `knitr::kable()`.

Value

Formatted string output

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

print.idsa_result *print IDSA result*

Description

S3 method to format output for IDSA model from `idsa()`.

Usage

```
## S3 method for class 'idsa_result'
print(x, ...)
```

Arguments

`x`	Return by `idsa()`.
`...`	(optional) Other arguments passed to `knitr::kable()`.

Value

Formatted string output

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

---

print.interaction_detector *print interaction detector*

Description

S3 method to format output for interaction detector in gd().

Usage

```
## S3 method for class 'interaction_detector'
print(x, ...)
```

Arguments

x	Return by gd().
...	(optional) Other arguments passed to knitr::kable().

Value

Formatted string output

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

---

print.lesh_result *print LESH model interaction result*

Description

S3 method to format output for LESH model interaction result in lesh().

Usage

```
## S3 method for class 'lesh_result'
print(x, ...)
```

Arguments

x	Return by lesh().
...	(optional) Other arguments passed to knitr::kable().

Value

Formatted string output

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

---

print.rid_result *print RID result*

Description

S3 method to format output for RID model from `rid()`.

Usage

```
## S3 method for class 'rid_result'
print(x, ...)
```

Arguments

`x`	Return by `rid()`.
`...`	(optional) Other arguments passed to `knitr::kable()`.

Value

Formatted string output

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

---

print.risk_detector *print risk detector*

Description

S3 method to format output for risk detector in `gd()`.

Usage

```
## S3 method for class 'risk_detector'
print(x, ...)
```

Arguments

x	Return by gd().
...	(optional) Other arguments passed to knitr::kable().

Value

Formatted string output

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

---

print.sesu_gozh *print gozh sesu*

Description

S3 method to format output for gozh sesu from sesu_gozh().

Usage

```
## S3 method for class 'sesu_gozh'
print(x, ...)
```

Arguments

x	Return by sesu_gozh().
...	(optional) Other arguments passed to knitr::kable().

Value

Formatted string output

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

print.sesu_opgd *print opgd sesu*

Description

S3 method to format output for opgd sesu from `sesu_opgd()`.

Usage

```
## S3 method for class 'sesu_opgd'
print(x, ...)
```

Arguments

`x`	Return by `sesu_opgd()`.
`...`	(optional) Other arguments passed to `knitr::kable()`.

Value

Formatted string output

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

print.spade_result *print SPADE power of spatial and multilevel discretization determinant*

Description

S3 method to format output for SPADE power of spatial and multilevel discretization determinant from `spade()`.

Usage

```
## S3 method for class 'spade_result'
print(x, ...)
```

Arguments

`x`	Return by `spade()`.
`...`	Other arguments.

Value

Formatted string output

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

---

psd_iev *PSD of an interaction of explanatory variables (PSD-IEV)*

Description

PSD of an interaction of explanatory variables (PSD-IEV)

Usage

```
psd_iev(discdata, spzone, wt)
```

Arguments

discdata	Discreted explanatory variables data. A tibble or dataframe .
spzone	Fuzzy overlay spatial zones. Returned from st_fuzzyoverlay().
wt	Spatial weight matrix

Details

$\phi = 1 - \frac{\sum_{i=1}^{m}\sum_{k=1}^{n_i} N_{i,k}\tau_{i,k}}{\sum_{i=1}^{m} N_i\tau_i}$

Value

The Value of PSD-IEV

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

References

Yongze Song & Peng Wu (2021) An interactive detector for spatial associations, International Journal of Geographical Information Science, 35:8, 1676-1701, DOI:10.1080/13658816.2021.1882680

Examples

```
data('NTDs')
wt = inverse_distance_weight(NTDs$X,NTDs$Y,power = 2)
sz = st_fuzzyoverlay(incidence ~ watershed + elevation + soiltype,
                     data = NTDs)
psd_iev(dplyr::select(NTDs,-c(X,Y,incidence)),sz,wt)
```

psd_pseudop *calculate power of spatial determinant(PSD) and the corresponding pseudo-p value*

Description

Function for calculate power of spatial determinant q_s.

Usage

```
psd_pseudop(y, x, wt, cores = 6, seed = 123456789, permutations = 0)
```

Arguments

y	Variable Y, continuous numeric vector.
x	Covariable X, `factor`, `character` or `discrete numeric`.
wt	The spatial weight matrix.
cores	(optional) A positive integer(default is 6). If cores > 1, use parallel computation.
seed	(optional) Random seed number, default is `123456789`.
permutations	(optional) The number of permutations for the PSD computation. Default is `0`, which means no pseudo-p values are calculated.

Details

The power of spatial determinant formula is $q_s = 1 - \frac{\sum_{h=1}^{L} N_h \Gamma_h}{N\Gamma}$

Value

A tibble of power of spatial determinant and the corresponding pseudo-p value.

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

References

Xuezhi Cang & Wei Luo (2018) Spatial association detector (SPADE),International Journal of Geographical Information Science, 32:10, 2055-2075, DOI: 10.1080/13658816.2018.1476693

Examples

```
## Not run:
data('NTDs')
wt = inverse_distance_weight(NTDs$X,NTDs$Y,power = 2)
tictoc::tic()
pp = psd_pseudop(NTDs$incidence,NTDs$soiltype,wt)
tictoc::toc()
```

```
pp

## End(Not run)
```

---

psd_spade *power of spatial determinant(PSD)*

Description

Function for calculate power of spatial determinant q_s.

Usage

```
psd_spade(y, x, wt)
```

Arguments

y	Variable Y, continuous numeric vector.
x	Covariable X, `factor`, `character` or `discrete numeric`.
wt	The spatial weight matrix.

Details

The power of spatial determinant formula is $q_s = 1 - \frac{\sum_{h=1}^{L} N_h \Gamma_h}{N\Gamma}$

Value

A value of power of spatial determinant q_s.

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

References

Xuezhi Cang & Wei Luo (2018) Spatial association detector (SPADE),International Journal of Geographical Information Science, 32:10, 2055-2075, DOI: 10.1080/13658816.2018.1476693

Examples

```
data('NTDs')
wt = inverse_distance_weight(NTDs$X,NTDs$Y,power = 2)
psd_spade(NTDs$incidence,NTDs$soiltype,wt)
```

psmd_pseudop *power of spatial and multilevel discretization determinant(PSMD) and the corresponding pseudo-p value*

Description

Function for calculate power of spatial and multilevel discretization determinant and the corresponding pseudo-p value.

Usage

```
psmd_pseudop(
  formula,
  data,
  wt = NULL,
  locations = NULL,
  discnum = NULL,
  discmethod = NULL,
  cores = 6,
  seed = 123456789,
  permutations = 0,
  ...
)
```

Arguments

formula	A formula of calculate power of spatial and multilevel discretization determinant $PSMDQ_s$.
data	A data.frame or tibble of observation data.
wt	(optional) The spatial weight matrix.When wt is not provided, must provide locations. And gdverse will use locations columns to construct spatial weight use inverse_distance_weight().
locations	(optional) The geospatial locations coordinate columns name which in data. Useful and must provided when wt is not provided.
discnum	(optional) Number of multilevel discretization.Default will use 3:22.
discmethod	(optional) The discretization methods. Default will use quantile. When discmethod is robust use robust_disc(), others use st_unidisc().Now only support one discmethod at one time.
cores	(optional) A positive integer(default is 6). If cores > 1, use parallel computation.
seed	(optional) Random seed number, default is 123456789.
permutations	(optional) The number of permutations for the PSD computation. Default is 0, which means no pseudo-p values are calculated.
...	(optional) Other arguments passed to st_unidisc() or robust_disc().

Details

The power of spatial and multilevel discretization determinant formula is $PSMDQ_s = MEAN(Q_s)$

Value

A tibble of power of spatial and multilevel discretization determinant and the corresponding pseudo-p value.

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

References

Xuezhi Cang & Wei Luo (2018) Spatial association detector (SPADE),International Journal of Geographical Information Science, 32:10, 2055-2075, DOI: 10.1080/13658816.2018.1476693

Examples

```
## Not run:
library(sf)
usfi = read_sf(system.file('extdata/USFI_Xian.gpkg',package = 'gdverse')) |>
  dplyr::select(dplyr::all_of(c("NDVI","BH","SUHI")))
coord = usfi |>
  st_centroid() |>
  st_coordinates()
usfi = usfi |>
  dplyr::bind_cols(coord) |>
  st_drop_geometry()
tictoc::tic()
pp = psmd_pseudop('SUHI ~ BH',data = dplyr::select(usfi,SUHI,BH,X,Y),
                  locations = c('X','Y'),cores = 6)
tictoc::toc()
pp

## End(Not run)
```

---

psmd_spade *power of spatial and multilevel discretization determinant(PSMD)*

Description

Function for calculate power of spatial and multilevel discretization determinant $PSMDQ_s$.

Usage

```
psmd_spade(
  formula,
  data,
  wt = NULL,
  locations = NULL,
  discnum = NULL,
  discmethod = NULL,
  cores = 1,
  seed = 123456789,
  ...
)
```

Arguments

formula	A formula of calculate power of spatial and multilevel discretization determinant $PSMDQ_s$.
data	A data.frame or tibble of observation data.
wt	(optional) The spatial weight matrix.When wt is not provided, must provide locations. And gdverse will use locations columns to construct spatial weight use inverse_distance_weight().
locations	(optional) The geospatial locations coordinate columns name which in data. Useful and must provided when wt is not provided.
discnum	(optional) Number of multilevel discretization.Default will use 3:22.
discmethod	(optional) The discretization methods. Default will use quantile. When discmethod is robust use robust_disc(), others use st_unidisc().Now only support one discmethod at one time.
cores	(optional) A positive integer(default is 1). If cores > 1, use parallel computation.
seed	(optional) Random seed number, default is 123456789.
...	(optional) Other arguments passed to st_unidisc() or robust_disc().

Details

The power of spatial and multilevel discretization determinant formula is $PSMDQ_s = MEAN(Q_s)$

Value

A value of power of spatial and multilevel discretization determinant $PSMDQ_s$.

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

References

Xuezhi Cang & Wei Luo (2018) Spatial association detector (SPADE),International Journal of Geographical Information Science, 32:10, 2055-2075, DOI: 10.1080/13658816.2018.1476693

Examples

```
## Not run:
library(sf)
usfi = read_sf(system.file('extdata/USFI_Xian.gpkg',package = 'gdverse')) |>
  dplyr::select(dplyr::all_of(c("NDVI","BH","SUHI")))
coord = usfi |>
  st_centroid() |>
  st_coordinates()
usfi = usfi |>
  dplyr::bind_cols(coord) |>
  st_drop_geometry()
psmd_spade('SUHI ~ BH',data = dplyr::select(usfi,SUHI,BH,X,Y),
           locations = c('X','Y'),cores = 6)

## End(Not run)
```

---

rescale_vector *rescale continuous vector to specified minimum and maximum*

Description

rescale continuous vector to specified minimum and maximum

Usage

```
rescale_vector(x, to_left = 0, to_right = 1)
```

Arguments

x	A continuous numeric vector.
to_left	(optional) Specified minimum. Default is 0.
to_right	(optional) Specified maximum. Default is 1.

Value

A continuous vector which has rescaled.

Examples

```
rescale_vector(c(-5,1,5),0.01,0.99)
```

rgd *robust geographical detector(RGD) model*

Description

Function for robust geographical detector(RGD) model.

Usage

```
rgd(
  formula,
  data,
  discvar,
  discnum = NULL,
  minsize = NULL,
  cores = 1,
  type = "factor",
  alpha = 0.95
)
```

Arguments

formula	A formula of RGD model.
data	A data.frame or tibble of observation data.
discvar	Name of continuous variable columns that need to be discretized.Noted that when formula has discvar, data must have these columns.
discnum	A numeric vector of discretized classes of columns that need to be discretized. Default all discvar use 10.
minsize	(optional) The min size of each discretization group.Default all use 1.
cores	(optional) Positive integer(default is 1). If cores > 1, use python's joblib package to parallel computation.
type	(optional) The type of geographical detector,which must be factor(default), interaction, risk, ecological.You can run one or more types at one time.
alpha	(optional) Specifies the size of confidence level.Default is 0.95.

Value

A list of the RGD model result.

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

References

Zhang, Z., Song, Y.*, & Wu, P., 2022. Robust geographical detector. International Journal of Applied Earth Observation and Geoinformation. 109, 102782. DOI: 10.1016/j.jag.2022.102782.

Examples

```
## Not run:
data('ndvi')
reticulate::use_condaenv('geocompy')
g = rgd(NDVIchange ~ ., data = ndvi, dicsvar = names(ndvi)[-1],
        cores = 6, type =c('factor','interaction'))

## End(Not run)
```

`rid` *robust interaction detector(RID) model*

Description

Function for robust interaction detector(RID) model.

Usage

```
rid(
  formula,
  data,
  overlaymethod = "and",
  discvar,
  discnum = NULL,
  minsize = NULL,
  cores = 1
)
```

Arguments

`formula`	A formula of RID model.
`data`	A data.frame or tibble of observation data.
`overlaymethod`	(optional) Spatial overlay method. One of `and`, `or`, `intersection`. Default is `and`.
`discvar`	Name of continuous variable columns that need to be discretized.Noted that when `formula` has `discvar`, `data` must have these columns.
`discnum`	A numeric vector of discretized classes of columns that need to be discretized. Default all `discvar` use `10`.
`minsize`	(optional) The min size of each discretization group.Default all use `1`.
`cores`	(optional) Positive integer(default is 1). If cores > 1, use parallel computation.

Value

A list of the RID model result.

Note

For bivariate spatial interactions, use the RGD function and specify the type parameter as interaction.

The RID model requires at least $2^n - 1$ calculations when has n explanatory variables. When there are more than 10 explanatory variables, carefully consider the computational burden of this model. When there are a large number of explanatory variables, the data dimensionality reduction method can be used to ensure the trade-off between analysis results and calculation speed.

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

References

Zhang, Z., Song, Y., Karunaratne, L., & Wu, P. (2024). Robust interaction detector: A case of road life expectancy analysis. Spatial Statistics, 59(100814), 100814. https://doi.org/10.1016/j.spasta.2024.100814

Examples

```
## Not run:
data('sim')
reticulate::use_condaenv('geocompy')
g = rid(y ~ ., data = sim %>% dplyr::select(-dplyr::any_of(c('lo','la'))),
        discvar = c("xa","xb","xc"), discnum = 4, cores = 6)
g

## End(Not run)
```

---

risk_detector *risk detector*

Description

Determine whether there is a significant difference between the attribute means of two subregions.

Usage

```
risk_detector(y, x, alpha = 0.95)
```

Arguments

y	Variable Y, continuous numeric vector.
x	Covariable X, factor, character or discrete numeric.
alpha	(optional) Confidence level of the interval,default is 0.95.

Value

A tibble contains different combinations of covariate X level and student t-test statistics, degrees of freedom, p-values, and whether has risk (Yes or No).

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

Examples

```
risk_detector(y = 1:7,
              x = c('x',rep('y',3),rep('z',3)))
```

---

robust_disc *univariate discretization based on offline change point detection*

Description

Determines discretization interval breaks using an optimization algorithm for variance-based change point detection.

Usage

```
robust_disc(formula, data, discnum, minsize = NULL, cores = 1)
```

Arguments

formula	A formula of univariate discretization.
data	A data.frame or tibble of observation data.
discnum	A numeric vector of discretized classes of columns that need to be discretized.
minsize	(optional) The min size of each discretization group.Default all use 1.
cores	(optional) A positive integer(default is 1). If cores > 1, use python's joblib package to parallel computation.

Value

A tibble of discretized classes of columns which need to be discretized.

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

Examples

```
## Not run:
library(terra)
library(tidyverse)
fvcpath = "https://github.com/SpatLyu/rdevdata/raw/main/FVC.tif"
fvc = terra::rast(paste0("/vsicurl/",fvcpath))
fvc = terra::aggregate(fvc,fact = 5)
fvc = as_tibble(terra::as.data.frame(fvc,na.rm = T))
```

```
new.fvc = robust_disc(fvc ~ .,data = select(fvc,-lulc),discnum = 10,cores = 6)
new.fvc

## End(Not run)
```

`rpart_disc` *discretization of variables based on recursive partitioning*

Description

discretization of variables based on recursive partitioning

Usage

```
rpart_disc(formula, data, ...)
```

Arguments

`formula`	A formula.
`data`	A data.frame or tibble of observation data.
`...`	(optional) Other arguments passed to `rpart::rpart()`.

Value

A vector that being discretized.

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

References

Luo, P., Song, Y., Huang, X., Ma, H., Liu, J., Yao, Y., & Meng, L. (2022). Identifying determinants of spatio-temporal disparities in soil moisture of the Northern Hemisphere using a geographically optimal zones-based heterogeneity model. ISPRS Journal of Photogrammetry and Remote Sensing: Official Publication of the International Society for Photogrammetry and Remote Sensing (ISPRS), 185, 111–128. https://doi.org/10.1016/j.isprsjprs.2022.01.009

Examples

```
## Not run:
data('ndvi')
rpart_disc(NDVIchange ~ ., data = ndvi)

## End(Not run)
```

sesu_gozh *comparison of size effects of spatial units based on GOZH*

Description

Function for comparison of size effects of spatial units in spatial heterogeneity analysis based on geographically optimal zones-based heterogeneity(GOZH) model.

Usage

```
sesu_gozh(
  formula,
  datalist,
  su,
  cores = 1,
  strategy = 2L,
  increase_rate = 0.05,
  alpha = 0.95,
  ...
)
```

Arguments

formula	A formula of comparison of size effects of spatial units.
datalist	A list of data.frame or tibble.
su	A vector of sizes of spatial units.
cores	(optional) A positive integer(default is 1). If cores > 1, a 'parallel' package cluster with that many cores is created and used. You can also supply a cluster object.
strategy	(optional) Calculation strategies of Q statistics at different scales. Default is 2L, see details for more contents.
increase_rate	(optional) The critical increase rate of the number of discretization. Default is 5%.
alpha	(optional) Specifies the size of confidence level. Default is 0.95.
...	(optional) Other arguments passed to rpart_disc().

Details

When strategy is 1, use the same process as sesu_opgd().If not, all explanatory variables are used to generate a unique Q statistic corresponding to the data in the datalist based on rpart_disc() and gd(), and then loess_optscale()is used to determine the optimal analysis scale.

Value

A list with sesu(size effects of spatial units),optsu(optimal spatial unit) and strategy(A number that represents the optimal analytical scale selection strategy).

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

References

Song, Y., Wang, J., Ge, Y. & Xu, C. (2020) An optimal parameters-based geographical detector model enhances geographic characteristics of explanatory variables for spatial heterogeneity analysis: Cases with different types of spatial data, GIScience & Remote Sensing, 57(5), 593-610. doi: 10.1080/15481603.2020.1760434.

Luo, P., Song, Y., Huang, X., Ma, H., Liu, J., Yao, Y., & Meng, L. (2022). Identifying determinants of spatio-temporal disparities in soil moisture of the Northern Hemisphere using a geographically optimal zones-based heterogeneity model. ISPRS Journal of Photogrammetry and Remote Sensing: Official Publication of the International Society for Photogrammetry and Remote Sensing (ISPRS), 185, 111–128. https://doi.org/10.1016/j.isprsjprs.2022.01.009

Examples

```
## Not run:
library(tidyverse)
fvcpath = "https://github.com/SpatLyu/rdevdata/raw/main/FVC.tif"
fvc = terra::rast(paste0("/vsicurl/",fvcpath))
fvc1000 = fvc %>%
  terra::as.data.frame(na.rm = T) %>%
  as_tibble()
fvc5000 = fvc %>%
  terra::aggregate(fact = 5) %>%
  terra::as.data.frame(na.rm = T) %>%
  as_tibble()
sesu_gozh(fvc ~ .,
          datalist = list(fvc1000,fvc5000),
          su = c(1000,5000),
          cores = 6)

## End(Not run)
```

---

sesu_opgd *comparison of size effects of spatial units based on OPGD*

Description

Function for comparison of size effects of spatial units in spatial heterogeneity analysis based on optimal parameters geographical detector(OPGD) model.

Usage

```
sesu_opgd(
  formula,
  datalist,
```

```
  su,
  discvar,
  discnum = NULL,
  discmethod = NULL,
  cores = 1,
  increase_rate = 0.05,
  alpha = 0.95,
  ...
)
```

Arguments

formula	A formula of comparison of size effects of spatial units.
datalist	A list of data.frame or tibble.
su	A vector of sizes of spatial units.
discvar	Name of continuous variable columns that need to be discretized.Noted that when formula has discvar, data must have these columns.
discnum	(optional) A vector of number of classes for discretization. Default is 3:22.
discmethod	(optional) A vector of methods for discretization,default is used c("sd","equal","pretty","quantile" gdverse.
cores	(optional) A positive integer(default is 1). If cores > 1, a 'parallel' package cluster with that many cores is created and used. You can also supply a cluster object.
increase_rate	(optional) The critical increase rate of the number of discretization. Default is 5%.
alpha	(optional) Specifies the size of confidence level. Default is 0.95.
...	(optional) Other arguments passed to gd_bestunidisc().

Details

Firstly, the OPGD model is executed for each data in the datalist (all significant Q statistic of each data are averaged to represent the spatial connection strength under this spatial unit), and then the loess_optscale function is used to select the optimal spatial analysis scale.

Value

A list with sesu(size effects of spatial units) and optsu(optimal spatial unit).

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

References

Song, Y., Wang, J., Ge, Y. & Xu, C. (2020) An optimal parameters-based geographical detector model enhances geographic characteristics of explanatory variables for spatial heterogeneity analysis: Cases with different types of spatial data, GIScience & Remote Sensing, 57(5), 593-610. doi: 10.1080/15481603.2020.1760434.

Examples

```
## Not run:
library(tidyverse)
fvcpath = "https://github.com/SpatLyu/rdevdata/raw/main/FVC.tif"
fvc = terra::rast(paste0("/vsicurl/",fvcpath))
fvc1000 = fvc %>%
  terra::as.data.frame(na.rm = T) %>%
  as_tibble()
fvc5000 = fvc %>%
  terra::aggregate(fact = 5) %>%
  terra::as.data.frame(na.rm = T) %>%
  as_tibble()
sesu_opgd(fvc ~ .,
          datalist = list(fvc1000,fvc5000),
          su = c(1000,5000),
          discvar = names(select(fvc5000,-c(fvc,lulc))),
          cores = 6)

## End(Not run)
```

---

shuffle_vector *randomly shuffling vector*

Description

randomly shuffling vector

Usage

```
shuffle_vector(x, shuffle_rate, seed = 123456789)
```

Arguments

x	A vector.
shuffle_rate	The shuffling rate.
seed	(optional) Random seed number. Default is 123456789.

Value

A shuffled vector.

Examples

```
shuffle_vector(1:100,0.15)
```

sim *Simulation data.*

Description

Simulation data.

Usage

```
sim
```

Format

`sim`: A tibble with 80 rows and 6 variables, modified from `IDSA` package.

spade *spatial association detector (SPADE) model*

Description

Function for spatial association detector (SPADE) model.

Usage

```
spade(
  formula,
  data,
  wt = NULL,
  locations = NULL,
  discnum = NULL,
  discmethod = NULL,
  cores = 6,
  seed = 123456789,
  permutations = 0,
  ...
)
```

Arguments

`formula`	A formula of spatial association detector (SPADE) model.
`data`	A data.frame or tibble of observation data.
`wt`	(optional) The spatial weight matrix.When `wt` is not provided, must provide `locations`. And `gdverse` will use `locations` columns to construct spatial weight use `inverse_distance_weight()`.

locations	(optional) The geospatial locations coordinate columns name which in data. Useful and must provided when wt is not provided. When wt is provided, locations is not need.
discnum	(optional) Number of multilevel discretization.Default will use 3:22.
discmethod	(optional) The discretization methods. Default all use quantile. When discmethod is robust use robust_disc(), others use st_unidisc()
cores	(optional) A positive integer(default is 6). If cores > 1, use parallel computation.
seed	(optional) Random number seed, default is 123456789.
permutations	(optional) The number of permutations for the PSD computation. Default is 0, which means no pseudo-p values are calculated.
...	(optional) Other arguments passed to st_unidisc() or robust_disc().

Value

A list of the SPADE model result.

Note

The columns in the locations part of data are only used to construct spatial weight matrix when wt is NULL, and are not considered as explanatory variables. If you need to include spatial locations as explanatory variables, build a spatial weight matrix ahead of time, leaving the locations parameter to NULL.The most recommended method is to explicitly specify all variables in formula instead of using .!

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

References

Xuezhi Cang & Wei Luo (2018) Spatial association detector (SPADE),International Journal of Geographical Information Science, 32:10, 2055-2075, DOI: 10.1080/13658816.2018.1476693

Examples

```
## Not run:
library(sf)
usfi = read_sf(system.file('extdata/USFI_Xian.gpkg',package = 'gdverse')) |>
  dplyr::select(dplyr::all_of(c("NDVI","BH","SUHI")))
coord = usfi |>
  st_centroid() |>
  st_coordinates()
wt = inverse_distance_weight(coord[,1],coord[,2])
usfi = usfi |>
  dplyr::bind_cols(coord) |>
  st_drop_geometry()
spade('SUHI~.', data = usfi,locations = c('X','Y'),cores = 6)
spade('SUHI~.', data = usfi, wt = wt,locations = c('X','Y'),
      discmethod = c('sd','equal'),cores = 6)
```

```
## End(Not run)
```

---

spd_lesh *SHAP power of determinants (SPD)*

Description

Function for calculate SHAP power of determinants SPD.

Usage

```
spd_lesh(formula, data, cores = 1, ...)
```

Arguments

formula	A formula of calculate SHAP power of determinants SPD.
data	A data.frame or tibble of observation data.
cores	(optional) A positive integer(default is 1). If cores > 1, a 'parallel' package cluster with that many cores is created and used. You can also supply a cluster object.
...	(optional) Other arguments passed to rpart_disc().

Details

The power of SHAP power of determinants formula is

$\theta_{x_j}(S) = \sum_{s \in M \setminus \{x_j\}} \frac{|S|!(|M|-|S|-1)!}{|M|!} \left(v\left(S \cup \{x_j\}\right) - v\left(S\right)\right).$

SHAP power of determinants (SPD) is the contribution of variable x_j to the power of determinants.

Value

A tibble with variable and its corresponding SPD value.

Note

The SHAP power of determinants (SPD) requires at least $2^n - 1$ calculations when has n explanatory variables. When there are more than 10 explanatory variables, carefully consider the computational burden of this model. When there are a large number of explanatory variables, the data dimensionality reduction method can be used to ensure the trade-off between analysis results and calculation speed.

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

References

Li, Y., Luo, P., Song, Y., Zhang, L., Qu, Y., & Hou, Z. (2023). A locally explained heterogeneity model for examining wetland disparity. International Journal of Digital Earth, 16(2), 4533–4552. https://doi.org/10.1080/17538947.2023.2271883

Examples

```
## Not run:
data('ndvi')
g = spd_lesh(NDVIchange ~ ., data = ndvi, cores = 6)
g

## End(Not run)
```

`spvar` *spatial variance*

Description

Function for calculate inverse distance weight.

Usage

```
spvar(yn, wtn)
```

Arguments

`yn`	The numerical vector of a response variable.
`wtn`	The spatial weight matrix.

Details

The spatial variance formula is $\Gamma = \frac{\sum_i \sum_{j \neq i} \omega_{ij} \frac{(y_i - y_j)^2}{2}}{\sum_i \sum_{j \neq i} \omega_{ij}}$

Value

The spatial variance

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

References

Xuezhi Cang & Wei Luo (2018) Spatial association detector (SPADE),International Journal of Geographical Information Science, 32:10, 2055-2075, DOI: 10.1080/13658816.2018.1476693

Examples

```
y = c(42,56,73)
wt1 = inverse_distance_weight(1:length(y),1:length(y))
wt2 = matrix(1,ncol = length(y),nrow = length(y))
diag(wt2) = 0
spvar(y,wt1)
spvar(y,wt2)
var(y)
```

---

st_fuzzyoverlay *spatial fuzzy overlay*

Description

Function for spatial fuzzy overlay.

Usage

```
st_fuzzyoverlay(formula, data, method = "and")
```

Arguments

formula	A formula of spatial fuzzy overlay.
data	A data.frame or tibble of discretized data.
method	(optional) Overlay methods. When method is and, use min to do fuzzy overlay;and when method is or,use max to do fuzzy overlay. Default is and.

Value

A spatial fuzzy overlay vector.

Note

Independent variables in the data provided to st_fuzzyoverlay() must be discretized variables, and dependent variables are continuous variables.

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

References

Yongze Song & Peng Wu (2021) An interactive detector for spatial associations, International Journal of Geographical Information Science, 35:8, 1676-1701, DOI:10.1080/13658816.2021.1882680

Examples

```
data('sim')
sim = sim %>%
  dplyr::mutate(dplyr::across(4:6,\(.x) st_unidisc(.x,4,"quantile")))
fo1 = st_fuzzyoverlay(y~xa+xb+xc,data = sim, method = 'and')
fo2 = st_fuzzyoverlay(y~xa+xb+xc,data = sim, method = 'or')
fo1
fo2
```

`st_unidisc` *univariate discretization*

Description

Function to classify univariate vector to interval,a wrapper of `classInt::classify_intervals()`.

Usage

```
st_unidisc(x, k, method = "quantile", factor = FALSE, seed = 123456789, ...)
```

Arguments

`x`	A continuous numerical variable.
`k`	(optional) Number of classes required, if missing, `grDevices::nclass.Sturges()` is used; see also the "dpih" and "headtails" styles for automatic choice of the number of classes. `k` must greater than `3` !
`method`	Chosen classify style: one of "fixed", "sd", "equal", "pretty", "quantile", "kmeans", "hclust", "bclust", "fisher", "jenks", "dpih", "headtails", "maximum", or "box". Default is `quantile`.
`factor`	(optional) Default is `FALSE`, if `TRUE` returns cols as a factor with intervals as labels rather than integers.
`seed`	(optional) Random seed number, default is `123456789`.Setting random seed is useful when the sample size is greater than `3000`(the default value for `largeN`) and the data is discretized by sampling `10%`(the default value for `samp_prop`).
`...`	(optional) Other arguments passed to `classInt::classify_intervals()`, see `?classInt::classify_intervals()`.

Value

A discrete vectors after being discretized.

Author(s)

Wenbo Lv <lyu.geosocial@gmail.com>

Examples

```
xvar = c(22361, 9573, 4836, 5309, 10384, 4359, 11016, 4414, 3327, 3408,
         17816, 6909, 6936, 7990, 3758, 3569, 21965, 3605, 2181, 1892,
         2459, 2934, 6399, 8578, 8537, 4840, 12132, 3734, 4372, 9073,
         7508, 5203)
st_unidisc(xvar,k = 6,method = 'sd')
```

`weight_assign` *assign values by weight*

Description

assign values by weight

Usage

```
weight_assign(x, w, list = FALSE)
```

Arguments

`x`	A numeric value
`w`	A weight vector
`list`	(optional) Return list or not. if `list` is `TRUE`, return a list, otherwise return a vector. Default is `FALSE`.

Value

A numeric Vector.

Examples

```
weight_assign(0.875,1:3)
```

Index