Package ‘dataSDA’

February 11, 2026

Type Package

Title Datasets and Basic Statistics for Symbolic Data Analysis

Version 0.1.8

Date 2026-02-11

Author Po-Wei Chen [aut],
Chun-houh Chen [aut],
Han-Ming Wu [cre]

Maintainer Han-Ming Wu <wuhm@g.nccu.edu.tw>

Description Collects a diverse range of symbolic data and offers a comprehensive set of functions that facilitate the conversion of traditional data into the symbolic data format.

License GPL (>= 2)

Encoding UTF-8

LazyData true

RoxygenNote 7.3.2

Depends R (>= 4.0.0)

Suggests testthat (>= 2.1.0), knitr, rmarkdown

VignetteBuilder knitr

Imports magrittr, tidyr, dplyr, RSDA, HistDAWass

NeedsCompilation no

Repository CRAN

Date/Publication 2026-02-11 15:00:02 UTC

Contents

`abalone.iGAP` *Abalone Dataset (iGAP Format)*

Description

Interval-valued dataset of 24 units from the UCI Abalone dataset, aggregated by sex and age. iGAP format for matrix visualization.

Usage

```
data(abalone.iGAP)
```

Format

An object of class `data.frame` with 24 rows and 7 columns.

Source

UCI Machine Learning Repository.

References

Kao, C.-H. et al. (2014). Exploratory data analysis of interval-valued symbolic data with matrix visualization. *CSDA*, 79, 14-29.

Examples

```
data(abalone.iGAP)
```

abalone.int *Abalone Interval Dataset*

Description

Interval-valued dataset of 24 units from the UCI Abalone dataset, aggregated by sex and age. Standard data frame format.

Usage

```
data(abalone.int)
```

Format

An object of class `data.frame` with 24 rows and 14 columns.

Source

UCI Machine Learning Repository.

Examples

```
data(abalone.int)
```

acid_rain.int *Acid Rain Pollution Indices Interval Dataset*

Description

Interval-valued acid rain pollution indices for sulphates and nitrates (kg/hectares) by US state.

Usage

```
data(acid_rain.int)
```

Format

A data frame with 2 observations and 2 interval-valued variables:

- `sulphate`: Sulphate pollution index range (kg/hectares).
- `nitrate`: Nitrate pollution index range (kg/hectares).

References

Billard, L. and Diday, E. (2006). *Symbolic Data Analysis*. Wiley. Table 2.21.

Examples

```
data(acid_rain.int)
```

age_cholesterol_weight.int *Age-Cholesterol-Weight Interval Dataset*

Description

Interval-valued dataset relating age, cholesterol, and weight measurements.

Usage

```
data(age_cholesterol_weight.int)
```

Format

An object of class `symbolic_tbl` (inherits from `tbl_df`, `tbl`, `data.frame`) with 7 rows and 4 columns.

References

Billard, L. and Diday, E. (2006). *Symbolic Data Analysis*. Wiley.

Examples

```
data(age_cholesterol_weight.int)
```

airline_flights.hist *JFK Airport Airline Flights Histogram-Valued Dataset*

Description

Histogram-valued dataset of 16 airlines flying into JFK Airport. Six variables (Flight Time, Taxi In, Arrival Delay, Taxi Out, Departure Delay, Weather Delay) recorded as frequency distributions.

Usage

```
data(airline_flights.hist)
```

Format

An object of class `data.frame` with 16 rows and 17 columns.

References

Billard, L. and Diday, E. (2006). *Symbolic Data Analysis*. Wiley. Table 2.7.

Examples

```
data(airline_flights.hist)
```

airline_flights2 *JFK Airport Airline Flights Modal-Valued Dataset*

Description

Modal-valued version of the airline flights dataset. See airline_flights.hist.

Usage

```
data(airline_flights2)
```

Format

An object of class symbolic_tbl (inherits from tbl_df, tbl, data.frame) with 16 rows and 6 columns.

References

Billard, L. and Diday, E. (2006). *Symbolic Data Analysis*. Wiley. Table 2.7.

Examples

```
data(airline_flights2)
```

bank_rates *Bank Interest Rates AR Model Symbolic Dataset*

Description

Symbolic dataset of autoregressive time series models for 4 banks. Each bank is described by AR model order, parameters, and noise variance.

Usage

```
data(bank_rates)
```

Format

An object of class data.frame with 4 rows and 6 columns.

References

Billard, L. and Diday, E. (2006). *Symbolic Data Analysis*. Wiley. Table 2.9.

Examples

```
data(bank_rates)
```

`baseball.int` *Baseball Teams Interval Dataset*

Description

Interval-valued data for baseball teams with player statistics.

Usage

```
data(baseball.int)
```

Format

An object of class `symbolic_tbl` (inherits from `tbl_df`, `tbl`, `data.frame`) with 19 rows and 3 columns.

References

Billard, L. and Diday, E. (2006). *Symbolic Data Analysis*. Wiley.

Examples

```
data(baseball.int)
```

`bats.int` *Bat Species Interval Dataset*

Description

Interval-valued data for 21 bat species described by 4 morphological measurements. Benchmark dataset for matrix visualization.

Usage

```
data(bats.int)
```

Format

A data frame with 21 observations and 4 interval-valued variables:

- `head`: Head length range (mm).
- `tail`: Tail length range (mm).
- `height`: Ear height range (cm).
- `forearm`: Forearm length range (mm).

Details

Used to demonstrate color coding schemes, the HCT-R2E seriation algorithm, and distance measure comparisons (Gowda-Diday, Hausdorff, City-Block, L1, L2, etc.) for interval data.

References

Kao, C.-H. et al. (2014). Exploratory data analysis of interval-valued symbolic data with matrix visualization. *CSDA*, 79, 14-29.

Examples

```
data(bats.int)
```

`bird.mix` *Bird Species Mixed Symbolic Dataset*

Description

Mixed symbolic data for bird species with interval-valued morphological measurements and categorical symbolic variables (habitat, color).

Usage

```
data(bird.mix)
```

Format

An object of class `symbolic_tbl` (inherits from `tbl_df`, `tbl`, `data.frame`) with 20 rows and 2 columns.

References

Billard, L. and Diday, E. (2006). *Symbolic Data Analysis*. Wiley. Table 2.5.

Examples

```
data(bird.mix)
```

`bird_species.mix` *Bird Species Mixed Symbolic Dataset*

Description

Symbolic data for 3 bird species (Swallow, Ostrich, Penguin) with interval-valued size, categorical flying, and categorical migration. Foundational SDA example from 600 individual bird observations.

Usage

```
data(bird_species.mix)
```

Format

A data frame with 3 observations and 3 symbolic variables:

- `flying`: Flying ability (Yes/No), categorical.
- `size`: Size range as interval (cm).
- `migration`: Migratory behavior, categorical.

References

Diday, E. and Noirhomme-Fraiture, M. (Eds.) (2008). *Symbolic Data Analysis and the SODAS Software*. Wiley. Table 1.2, p.6.

Examples

```
data(bird_species.mix)
```

`bird_species_extended.mix` *Bird Species Extended Mixed Symbolic Dataset*

Description

Three bird species (Geese, Ostrich, Penguin) with interval-valued height, histogram-valued color distribution, and categorical flying/migratory variables.

Usage

```
data(bird_species_extended.mix)
```

Format

A data frame with 3 observations and 4 symbolic variables.

References

Billard, L. and Diday, E. (2006). *Symbolic Data Analysis*. Wiley. Table 2.19.

Examples

```
data(bird_species_extended.mix)
```

blood_pressure.int *Blood Pressure Interval Dataset*

Description

Interval-valued blood pressure measurements by patient groups.

Usage

```
data(blood_pressure.int)
```

Format

An object of class `symbolic_tbl` (inherits from `tbl_df`, `tbl`, `data.frame`) with 15 rows and 3 columns.

References

Billard, L. and Diday, E. (2006). *Symbolic Data Analysis*. Wiley.

Examples

```
data(blood_pressure.int)
```

car.int *Car Models Interval Dataset*

Description

Interval-valued data for car models with price, engine, speed, acceleration.

Usage

```
data(car.int)
```

Format

An object of class `symbolic_tbl` (inherits from `tbl_df`, `tbl`, `data.frame`) with 8 rows and 5 columns.

References

Billard, L. and Diday, E. (2006). *Symbolic Data Analysis*. Wiley.

Examples

```
data(car.int)
```

`cars.int` *Cars Interval Dataset*

Description

Interval-valued data for 27 car models classified into four classes (Utilitarian, Berlina, Sportive, Luxury), described by Price, EngineCapacity, TopSpeed and Acceleration intervals.

Usage

```
data(cars.int)
```

Format

A data frame with 27 observations and 5 variables.

Source

https://CRAN.R-project.org/package=MAINT.Data

References

Duarte Silva, A.P., Brito, P., Filzmoser, P. and Dias, J.G. (2021). MAINT.Data: Modelling and Analysing Interval Data in R. *R Journal*, 13(2).

Examples

```
data(cars.int)
```

china_temp.int *China Meteorological Stations Quarterly Temperature Interval Dataset*

Description

Interval-valued temperature data (Celsius) for 60 Chinese meteorological stations observed over the four quarters of years 1974 to 1988. One outlier observation (YinChuan_1982) has been discarded.

Usage

```
data(china_temp.int)
```

Format

A data frame with 899 observations and 5 variables.

Details

Originates from the Long-Term Instrumental Climatic Database of the People's Republic of China. Widely used in the SDA literature for demonstrating standardization, clustering, self-organizing maps, MLE and MANOVA.

Source

https://CRAN.R-project.org/package=MAINT.Data

References

Brito, P. and Duarte Silva, A.P. (2012). Modelling interval data with Normal and Skew-Normal distributions. *J. Appl. Stat.*, 39(1), 3-20.

Kao, C.-H. et al. (2014). Exploratory data analysis of interval-valued symbolic data with matrix visualization. *CSDA*, 79, 14-29.

Examples

```
data(china_temp.int)
```

clean_colnames *clean_colnames*

Description

This function is used to clean up variable names to conform to the RSDA format.

Usage

```
clean_colnames(data)
```

Arguments

data	The conventional data.

Value

Data after cleaning variable names.

Examples

```
data(mushroom)
mushroom.clean <- clean_colnames(data = mushroom)
```

credit_card.int *Credit Card Expenses Interval Dataset*

Description

Interval-valued credit card spending aggregated by person-month. Three individuals' (Jon, Tom, Leigh) monthly expenditures across five categories.

Usage

```
data(credit_card.int)
```

Format

A data frame with person-month rows and 5 interval-valued columns:

- `food`: Food expenditure range (USD).
- `social`: Social expenditure range (USD).
- `travel`: Travel expenditure range (USD).
- `gas`: Gas expenditure range (USD).
- `clothes`: Clothes expenditure range (USD).

Details

The original classical dataset (Table 2.3) records individual transactions. The symbolic version (Table 2.4) aggregates into interval-valued observations for each person-month combination.

References

Billard, L. and Diday, E. (2006). *Symbolic Data Analysis*. Wiley. Tables 2.3-2.4.

Examples

```
data(credit_card.int)
```

`crime` *Crime Demographics Dataset*

Description

Crime-related demographic variables with symbolic data types.

Usage

```
data(crime)
```

Format

An object of class `data.frame` with 15 rows and 7 columns.

References

Billard, L. and Diday, E. (2006). *Symbolic Data Analysis*. Wiley.

Examples

```
data(crime)
```

`crime2` *Crime Demographics Modal-Valued Dataset*

Description

Modal-valued version of the crime demographics dataset.

Usage

```
data(crime2)
```

Format

An object of class `symbolic_tbl` (inherits from `tbl_df`, `tbl`, `data.frame`) with 15 rows and 3 columns.

References

Billard, L. and Diday, E. (2006). *Symbolic Data Analysis*. Wiley.

Examples

```
data(crime2)
```

`employment.int` *European Employment by Gender and Age Interval Dataset*

Description

Interval-valued proportions for 12 sex-age population groups across employment variables (employment type, education, industry sector, occupation, marital status). Used for factorial discriminant analysis.

Usage

```
data(employment.int)
```

Format

A data frame with 12 sex-age group observations and interval-valued proportion variables.

References

Diday, E. and Noirhomme-Fraiture, M. (Eds.) (2008). *Symbolic Data Analysis and the SODAS Software*. Wiley. Table 18.1.

Examples

```
data(employment.int)
```

`energy_consumption.distr`

US Energy Consumption Distribution-Valued Dataset

Description

Distribution-valued dataset of energy consumption across US states. Each energy type described by Normal distribution parameters (mean, SD).

Usage

```
data(energy_consumption.distr)
```

Format

A data frame with 5 observations and 3 variables:

- `type`: Energy type.
- `mean`: Mean consumption across 50 states.
- `sd`: Standard deviation.

Details

Five types: Petroleum, Natural Gas, Coal, Hydroelectric, Nuclear Power. Values are rescaled consumption from the US Census Bureau (2004).

References

Billard, L. and Diday, E. (2006). *Symbolic Data Analysis*. Wiley. Table 2.8.

Examples

```
data(energy_consumption.distr)
```

face.iGAP *Face Dataset (iGAP Format)*

Description

Symbolic data matrix with all interval-type variables for facial measurements, in iGAP format.

Usage

```
data(face.iGAP)
```

Format

An object of class `data.frame` with 27 rows and 6 columns.

References

Kao, C.-H. et al. (2014). Exploratory data analysis of interval-valued symbolic data with matrix visualization. *CSDA*, 79, 14-29.

Examples

```
data(face.iGAP)
```

finance.int *Finance Sector Interval Dataset*

Description

Interval-valued data for 14 business sectors described by job-related financial variables (job cost codes, activity codes, budgets). Used for PCA demonstrations.

Usage

```
data(finance.int)
```

Format

An object of class `symbolic_tbl` (inherits from `tbl_df`, `tbl`, `data.frame`) with 14 rows and 7 columns.

References

Billard, L. and Diday, E. (2006). *Symbolic Data Analysis*. Wiley. Table 5.2.

Examples

```
data(finance.int)
```

fuel_consumption *Fuel Consumption by Region Dataset*

Description

Modal-valued dataset describing fuel consumption patterns across 10 regions by proportions of heating fuel types (gas, oil, electricity, coal, none) and central heating presence.

Usage

```
data(fuel_consumption)
```

Format

An object of class `symbolic_tbl` (inherits from `tbl_df`, `tbl`, `data.frame`) with 10 rows and 3 columns.

References

Billard, L. and Diday, E. (2006). *Symbolic Data Analysis*. Wiley. Table 3.7.

Examples

```
data(fuel_consumption)
```

health_insurance.mix *Health Insurance Mixed Symbolic Dataset*

Description

Health insurance data grouped by disease type and gender with classical and symbolic variables of mixed types.

Usage

```
data(health_insurance.mix)
```

Format

An object of class `data.frame` with 51 rows and 30 columns.

References

Billard, L. and Diday, E. (2006). *Symbolic Data Analysis*. Wiley. Tables 2.1-2.2.

Examples

```
data(health_insurance.mix)
```

health_insurance2 *Health Insurance Modal-Valued Dataset*

Description

Modal-valued version of the health insurance dataset.

Usage

```
data(health_insurance2)
```

Format

An object of class `symbolic_tbl` (inherits from `tbl_df`, `tbl`, `data.frame`) with 6 rows and 6 columns.

References

Billard, L. and Diday, E. (2006). *Symbolic Data Analysis*. Wiley. Table 2.2b.

Examples

```
data(health_insurance2)
```

hierarchy *Hierarchy Dataset*

Description

Classical dataset illustrating hierarchical data structures.

Usage

```
data(hierarchy)
```

Format

An object of class `data.frame` with 20 rows and 6 columns.

References

Billard, L. and Diday, E. (2006). *Symbolic Data Analysis*. Wiley. Table 2.15.

Examples

```
data(hierarchy)
```

hierarchy.int *Hierarchy Interval Dataset*

Description

Interval-valued version of the hierarchy dataset.

Usage

```
data(hierarchy.int)
```

Format

An object of class `symbolic_tbl` (inherits from `tbl_df`, `tbl`, `data.frame`) with 20 rows and 6 columns.

References

Billard, L. and Diday, E. (2006). *Symbolic Data Analysis*. Wiley. Table 2.15.

Examples

```
data(hierarchy.int)
```

histogram_stats *Statistics for Histogram Data*

Description

Functions to compute the mean, variance, covariance, and correlation of histogram-valued data.

Usage

```
hist_mean(x, var_name, method = "BG", ...)

hist_var(x, var_name, method = "BG", ...)

hist_cov(x, var_name1, var_name2, method = "BG")

hist_cor(x, var_name1, var_name2, method = "BG")
```

Arguments

`x`	histogram-valued data object.
`var_name`	the variable name or the column location.
`method`	methods to calculate statistics: mean and var: BG (default), L2W; cov and cor: BG (default), BD, B, L2W.
`...`	additional parameters.
`var_name1`	the variable name or the column location.
`var_name2`	the variable name or the column location.

Details

...

Value

A numeric value: the mean, variance, covariance, or correlation.

Author(s)

Po-Wei Chen, Han-Ming Wu

See Also

int_mean int_var int_cov int_cor

Examples

```
library(HistDAWass)
```

`horses.int` *Horse Breeds Interval Dataset*

Description

Interval-valued data for 8 horse breeds (CES, CMA, PEN, TES, CEN, LES, PES, PAM) described by 6 variables: minimum/maximum weight, minimum/maximum height, cost of mares, cost of fillies.

Usage

```
data(horses.int)
```

Format

A data frame with 8 observations and 6 interval-valued variables.

Details

Extensively used in SDA for demonstrating divisive clustering, distance computation, hierarchy/pyramid construction, and complete objects.

References

Billard, L. and Diday, E. (2006). *Symbolic Data Analysis*. Wiley. Table 7.14.

Examples

```
data(horses.int)
```

---

iGAP_to_MM *iGAP to MM*

Description

To convert iGAP format to MM format.

Usage

```
iGAP_to_MM(data, location)
```

Arguments

`data`	The dataframe with the iGAP format.
`location`	The location of the symbolic variable in the data.

Value

Return a dataframe with the MM format.

Examples

```
data(abalone.iGAP)
abalone <- iGAP_to_MM(abalone.iGAP, 1:7)
```

iGAP_to_RSDA *iGAP to RSDA*

Description

To convert iGAP format interval dataframe to RSDA format (symbolic_tbl).

Usage

```
iGAP_to_RSDA(data, location)
```

Arguments

`data`	The dataframe with the iGAP format.
`location`	The location of the symbolic variable in the data.

Value

Return a symbolic_tbl dataframe with complex-encoded interval columns.

Examples

```
data(abalone.iGAP)
rsda <- iGAP_to_RSDA(abalone.iGAP, 1:7)
```

interval_distance *Distance Measures for Interval Data*

Description

Functions to compute various distance measures between interval-valued observations.

`int_dist_all` computes all available distance measures at once.

Usage

```
int_dist(x, method = "euclidean", gamma = 0.5, q = 1, p = 2, ...)

int_dist_matrix(x, method = "euclidean", gamma = 0.5, q = 1, p = 2, ...)

int_pairwise_dist(x, var_name1, var_name2, method = "euclidean", ...)

int_dist_all(x, gamma = 0.5, q = 1)
```

Arguments

`x` interval-valued data with symbolic_tbl class, or an array of dimension [n, p, 2]

`method` distance method: "GD", "IY", "L1", "L2", "CB", "HD", "EHD", "nEHD", "snEHD", "TD", "WD", "euclidean", "hausdorff", "manhattan", "city_block", "minkowski", "wasserstein", "ichino", "de_carvalho"

`gamma` parameter for the Ichino-Yaguchi distance, 0 <= gamma <= 0.5 (default: 0.5)

`q` parameter for the Ichino-Yaguchi distance (Minkowski exponent) (default: 1)

`p` power parameter for Minkowski distance (default: 2)

`...` additional parameters

`var_name1` first variable name or column location

`var_name2` second variable name or column location

Details

Available distance methods:

- `GD`: Gowda-Diday distance (Gowda & Diday, 1991)
- `IY`: Ichino-Yaguchi distance (Ichino, 1988)
- `L1`: L1 (midpoint Manhattan) distance
- `L2`: L2 (Euclidean midpoint) distance
- `CB`: City-Block distance (Souza & de Carvalho, 2004)
- `HD`: Hausdorff distance (Chavent & Lechevallier, 2002)
- `EHD`: Euclidean Hausdorff distance
- `nEHD`: Normalized Euclidean Hausdorff distance
- `snEHD`: Span Normalized Euclidean Hausdorff distance
- `TD`: Tran-Duckstein distance (Tran & Duckstein, 2002)
- `WD`: L2-Wasserstein distance (Verde & Irpino, 2008)
- `euclidean`: Euclidean distance on interval centers (same as L2)
- `hausdorff`: Hausdorff distance (same as HD)
- `manhattan`: Manhattan distance (same as L1)
- `city_block`: City-block distance (same as CB)
- `minkowski`: Minkowski distance with parameter p
- `wasserstein`: Wasserstein distance (same as WD)
- `ichino`: Ichino-Yaguchi distance (simplified version)
- `de_carvalho`: De Carvalho distance

Value

A distance matrix (class 'dist') or numeric vector

Author(s)

Han-Ming Wu

References

Gowda, K. C., & Diday, E. (1991). Symbolic clustering using a new dissimilarity measure. *Pattern Recognition*, 24(6), 567-578.

Ichino, M. (1988). General metrics for mixed features. *Systems and Computers in Japan*, 19(2), 37-50.

Chavent, M., & Lechevallier, Y. (2002). Dynamical clustering of interval data. In *Classification, Clustering and Data Analysis* (pp. 53-60). Springer.

Tran, L., & Duckstein, L. (2002). Comparison of fuzzy numbers using a fuzzy distance measure. *Fuzzy Sets and Systems*, 130, 331-341.

Verde, R., & Irpino, A. (2008). A new interval data distance based on the Wasserstein metric.

Kao, C.-H. et al. (2014). Exploratory data analysis of interval-valued symbolic data with matrix visualization. *CSDA*, 79, 14-29.

See Also

int_dist_matrix int_dist_all int_pairwise_dist

Examples

```
# Using symbolic_tbl format
data(mushroom.int)
d1 <- int_dist(mushroom.int[, 3:4], method = "euclidean")
d2 <- int_dist(mushroom.int[, 3:4], method = "hausdorff")
d3 <- int_dist(mushroom.int[, 3:4], method = "GD")

# Using array format: 4 concepts, 3 variables
x <- array(NA, dim = c(4, 3, 2))
x[,,1] <- matrix(c(1,2,3,4, 5,6,7,8, 9,10,11,12), nrow=4)
x[,,2] <- matrix(c(3,5,6,7, 8,9,10,12, 13,15,16,18), nrow=4)
d4 <- int_dist(x, method = "snEHD")
d5 <- int_dist(x, method = "IY", gamma = 0.3)
```

---

interval_geometry *Geometric Properties of Interval Data*

Description

Functions to compute geometric characteristics of interval-valued data.

Usage

```
int_width(x, var_name, ...)

int_radius(x, var_name, ...)

int_center(x, var_name, ...)

int_overlap(x, var_name1, var_name2, ...)

int_containment(x, var_name1, var_name2, ...)

int_midrange(x, var_name, ...)
```

Arguments

`x`	interval-valued data with symbolic_tbl class.
`var_name`	the variable name or the column location (multiple variables are allowed).
`...`	additional parameters
`var_name1`	the first variable name or column location.
`var_name2`	the second variable name or column location.

Details

These functions compute basic geometric properties:

- `int_width`: Width of each interval (upper - lower)
- `int_radius`: Radius of each interval (width / 2)
- `int_center`: Center point of each interval ((lower + upper) / 2)
- `int_overlap`: Overlap measure between two interval variables
- `int_containment`: Check if one interval contains another

Value

A numeric matrix or value

Author(s)

Han-Ming Wu

See Also

int_width int_radius int_center int_overlap

Examples

```
data(mushroom.int)

# Calculate interval widths
int_width(mushroom.int, var_name = "Pileus.Cap.Width")
int_width(mushroom.int, var_name = 2:3)

# Calculate interval radius
int_radius(mushroom.int, var_name = c("Stipe.Length", "Stipe.Thickness"))

# Get interval centers
int_center(mushroom.int, var_name = 2:4)

# Measure overlap between two variables
int_overlap(mushroom.int, "Pileus.Cap.Width", "Stipe.Length")
```

`interval_position` *Position and Scale Measures for Interval Data*

Description

Functions to compute position and scale statistics for interval-valued data.

Usage

```
int_median(x, var_name, method = "CM", ...)

int_quantile(x, var_name, probs = c(0.25, 0.5, 0.75), method = "CM", ...)

int_range(x, var_name, method = "CM", ...)

int_iqr(x, var_name, method = "CM", ...)

int_mad(x, var_name, method = "CM", ...)

int_mode(x, var_name, method = "CM", breaks = 30, ...)
```

Arguments

`x`	interval-valued data with symbolic_tbl class.
`var_name`	the variable name or the column location (multiple variables are allowed).
`method`	methods to calculate statistics: CM (default), VM, QM, SE, FV, EJD, GQ, SPT.
`...`	additional parameters
`probs`	numeric vector of probabilities with values in [0,1].
`breaks`	number of histogram breaks for mode estimation (default: 30).

Details

These functions provide position and scale measures:

- `int_median`: Median of interval data
- `int_quantile`: Quantiles of interval data
- `int_range`: Range (max - min) of interval data
- `int_iqr`: Interquartile range (Q3 - Q1)
- `int_mad`: Median absolute deviation

Value

A numeric matrix or value

Author(s)

Han-Ming Wu

See Also

int_mean int_var int_median int_quantile

Examples

```
data(mushroom.int)

# Calculate median
int_median(mushroom.int, var_name = "Pileus.Cap.Width")
int_median(mushroom.int, var_name = 2:3, method = c("CM", "EJD"))

# Calculate quantiles
int_quantile(mushroom.int, var_name = 2, probs = c(0.25, 0.5, 0.75))

# Calculate interquartile range
int_iqr(mushroom.int, var_name = c("Stipe.Length", "Stipe.Thickness"))

# Calculate MAD
int_mad(mushroom.int, var_name = 2:3, method = "CM")
```

`interval_robust` *Robust Statistics for Interval Data*

Description

Functions to compute robust statistics for interval-valued data.

Usage

```
int_trimmed_mean(x, var_name, trim = 0.1, method = "CM", ...)

int_winsorized_mean(x, var_name, trim = 0.1, method = "CM", ...)

int_trimmed_var(x, var_name, trim = 0.1, method = "CM", ...)

int_winsorized_var(x, var_name, trim = 0.1, method = "CM", ...)
```

Arguments

`x`	interval-valued data with symbolic_tbl class.
`var_name`	the variable name or the column location (multiple variables are allowed).
`trim`	the fraction (0 to 0.5) of observations to be trimmed from each end.
`method`	methods to calculate statistics: CM (default), VM, QM, SE, FV, EJD, GQ, SPT.
`...`	additional parameters

Details

These functions provide robust alternatives to standard statistics:

- `int_trimmed_mean`: Mean after trimming extreme values
- `int_winsorized_mean`: Mean after winsorizing extreme values
- `int_trimmed_var`: Variance after trimming extreme values
- `int_winsorized_var`: Variance after winsorizing extreme values

Trimming vs Winsorizing:

- Trimming: Remove extreme values
- Winsorizing: Replace extreme values with less extreme values

Value

A numeric matrix

Author(s)

Han-Ming Wu

See Also

int_mean int_var int_trimmed_mean

Examples

```
data(mushroom.int)

# Trimmed mean (10% from each end)
int_trimmed_mean(mushroom.int, var_name = "Pileus.Cap.Width", trim = 0.1)

# Winsorized mean
int_winsorized_mean(mushroom.int, var_name = 2:3, trim = 0.05, method = "CM")

# Trimmed variance
int_trimmed_var(mushroom.int, var_name = c("Stipe.Length"), trim = 0.1)
```

`interval_shape` *Distribution Shape Measures for Interval Data*

Description

Functions to compute shape statistics (skewness, kurtosis) for interval-valued data.

Usage

```
int_skewness(x, var_name, method = "CM", ...)

int_kurtosis(x, var_name, method = "CM", ...)

int_symmetry(x, var_name, method = "CM", ...)

int_tailedness(x, var_name, method = "CM", ...)
```

Arguments

`x`	interval-valued data with symbolic_tbl class.
`var_name`	the variable name or the column location (multiple variables are allowed).
`method`	methods to calculate statistics: CM (default), VM, QM, SE, FV, EJD, GQ, SPT.
`...`	additional parameters

Details

These functions measure distribution shape:

- `int_skewness`: Measure of asymmetry (skewness)
- `int_kurtosis`: Measure of tail heaviness (kurtosis)
- `int_symmetry`: Symmetry coefficient

Skewness interpretation:

- = 0: Symmetric distribution

- > 0: Right-skewed (positive skew)
- < 0: Left-skewed (negative skew)

Kurtosis interpretation (excess kurtosis):

- = 0: Normal distribution (mesokurtic)
- > 0: Heavy tails (leptokurtic)
- < 0: Light tails (platykurtic)

Value

A numeric matrix

Author(s)

Han-Ming Wu

See Also

int_mean int_var int_skewness int_kurtosis

Examples

```
data(mushroom.int)

# Calculate skewness
int_skewness(mushroom.int, var_name = "Pileus.Cap.Width")
int_skewness(mushroom.int, var_name = 2:3, method = c("CM", "EJD"))

# Calculate kurtosis
int_kurtosis(mushroom.int, var_name = c("Stipe.Length", "Stipe.Thickness"))

# Check symmetry
int_symmetry(mushroom.int, var_name = 2:4, method = "CM")
```

`interval_similarity` *Similarity Measures for Interval Data*

Description

Functions to compute similarity measures between interval-valued observations.

Usage

```
int_jaccard(x, var_name1, var_name2, ...)

int_dice(x, var_name1, var_name2, ...)

int_cosine(x, var_name1, var_name2, ...)

int_overlap_coefficient(x, var_name1, var_name2, ...)

int_tanimoto(x, var_name1, var_name2, ...)

int_similarity_matrix(x, method = ”jaccard”, ...)
```

Arguments

`x`	interval-valued data with symbolic_tbl class.
`var_name1`	the first variable name or column location.
`var_name2`	the second variable name or column location.
`...`	additional parameters
`method`	similarity method for int_similarity_matrix: "jaccard", "dice", or "overlap".

Details

These functions compute various similarity measures:

- `int_jaccard`: Jaccard similarity coefficient
- `int_dice`: Dice similarity coefficient
- `int_cosine`: Cosine similarity
- `int_overlap_coefficient`: Overlap coefficient
- `int_tanimoto`: Tanimoto coefficient (generalized Jaccard)

All similarity measures range from 0 (no similarity) to 1 (perfect similarity).

Value

A numeric matrix or value

Author(s)

Han-Ming Wu

See Also

int_dist int_cor int_jaccard

Examples

```
data(mushroom.int)

# Jaccard similarity
int_jaccard(mushroom.int, "Pileus.Cap.Width", "Stipe.Length")

# Dice coefficient
int_dice(mushroom.int, 2, 3)

# Cosine similarity
int_cosine(mushroom.int,
           var_name1 = c("Pileus.Cap.Width"),
           var_name2 = c("Stipe.Length", "Stipe.Thickness"))

# Overlap coefficient
int_overlap_coefficient(mushroom.int, 2, 3:4)
```

`interval_stats` *Statistics for Interval Data*

Description

Functions to compute the mean, variance, covariance, and correlation of interval-valued data.

Usage

```
int_mean(x, var_name, method = "CM", ...)

int_var(x, var_name, method = "CM", ...)

int_cov(x, var_name1, var_name2, method = "CM", ...)

int_cor(x, var_name1, var_name2, method = "CM", ...)
```

Arguments

`x`	interval-valued data with symbolic_tbl class.
`var_name`	the variable name or the column location (multiple variables are allowed).
`method`	methods to calculate statistics: CM (default), VM, QM, SE, FV, EJD, GQ, SPT.
`...`	additional parameters
`var_name1`	the variable name or the column location (multiple variables are allowed).
`var_name2`	the variable name or the column location (multiple variables are allowed).

Details

...

Value

A numeric value: the mean, variance, covariance, or correlation.

Author(s)

Han-Ming Wu

See Also

int_mean int_var int_cov int_cor

Examples

```
data(mushroom.int)
int_mean(mushroom.int, var_name = "Pileus.Cap.Width")
int_mean(mushroom.int, var_name = 2:3)

var_name <- c("Stipe.Length", "Stipe.Thickness")
method <- c("CM", "FV", "EJD")
int_mean(mushroom.int, var_name, method)
int_var(mushroom.int, var_name, method)

var_name1 <- "Pileus.Cap.Width"
var_name2 <- c("Stipe.Length", "Stipe.Thickness")
method <- c("CM", "VM", "EJD", "GQ", "SPT")
int_cov(mushroom.int, var_name1, var_name2, method)
int_cor(mushroom.int, var_name1, var_name2, method)
```

interval_uncertainty *Uncertainty and Variability Measures for Interval Data*

Description

Functions to compute uncertainty and variability measures for interval-valued data.

Usage

```
int_entropy(x, var_name, method = "CM", base = 2, ...)

int_cv(x, var_name, method = "CM", ...)

int_dispersion(x, var_name, method = "CM", ...)

int_imprecision(x, var_name, ...)

int_granularity(x, var_name, ...)

int_uniformity(x, var_name, ...)

int_information_content(x, var_name, method = "CM", ...)
```

Arguments

`x` interval-valued data with symbolic_tbl class.

`var_name` the variable name or the column location (multiple variables are allowed).

`method` methods to calculate statistics: CM (default), VM, QM, SE, FV, EJD, GQ, SPT.

`base` logarithm base for entropy calculation (default: 2)

`...` additional parameters

Details

These functions measure uncertainty and variability:

- `int_entropy`: Shannon entropy (information content)
- `int_cv`: Coefficient of variation (CV = SD / Mean)
- `int_dispersion`: General dispersion index
- `int_imprecision`: Imprecision based on interval width
- `int_granularity`: Variability in interval sizes

Value

A numeric matrix or value

Author(s)

Han-Ming Wu

See Also

int_var int_entropy int_cv

Examples

```
data(mushroom.int)

# Calculate entropy
int_entropy(mushroom.int, var_name = "Pileus.Cap.Width")

# Coefficient of variation
int_cv(mushroom.int, var_name = c("Stipe.Length", "Stipe.Thickness"), method = c("CM", "EJD"))

# Measure imprecision
int_imprecision(mushroom.int, var_name = c("Stipe.Length", "Stipe.Thickness"))

# Check data granularity
int_granularity(mushroom.int, var_name = 2:4)
```

int_convert_format *Convert Interval Data Format*

Description

Automatically detect the format of interval data and convert it to the target format.

Usage

```
int_convert_format(x, to = "MM", from = NULL, ...)
```

Arguments

`x` interval data in one of the supported formats

`to` target format: "MM", "iGAP", "RSDA", "SODAS" (default: "MM")

`from` source format (optional): "MM", "iGAP", "RSDA", "SODAS". If NULL, will auto-detect.

`...` additional parameters passed to specific conversion functions

Details

This function provides a unified interface for all interval format conversions. It automatically detects the source format (unless specified) and applies the appropriate conversion function.

Supported conversions:

- RSDA → MM (via `RSDA_to_MM`)
- RSDA → iGAP (via `RSDA_to_iGAP`)
- iGAP → MM (via `iGAP_to_MM`)
- SODAS → MM (via `SODAS_to_MM`)
- SODAS → iGAP (via `SODAS_to_iGAP`)
- MM → iGAP (via `MM_to_iGAP`)
- MM → RSDA (via `MM_to_RSDA`)
- iGAP → RSDA (via `iGAP_to_RSDA`)

Value

Interval data in the target format

Author(s)

Han-Ming Wu

See Also

int_detect_format int_list_conversions RSDA_to_MM iGAP_to_MM MM_to_iGAP MM_to_RSDA iGAP_to_RSDA

Examples

```
# Auto-detect and convert to MM
data(mushroom.int)
data_mm <- int_convert_format(mushroom.int, to = "MM")

# Explicitly specify source format
data(abalone.iGAP)
data_mm <- int_convert_format(abalone.iGAP, from = "iGAP", to = "MM")

# Convert MM to iGAP
data_igap <- int_convert_format(data_mm, to = "iGAP")

 # Convert multiple datasets to MM
datasets <- list(mushroom.int, abalone.int, car.int)
mm_datasets <- lapply(datasets, int_convert_format, to = "MM")

# Check what conversions are available
int_list_conversions()
```

---

int_detect_format *Detect Interval Data Format*

Description

Automatically detect the format of interval data.

Usage

```
int_detect_format(x)
```

Arguments

x	interval data in unknown format

Details

Detection rules:

- RSDA: has class "symbolic_tbl" and contains complex columns
- MM: data.frame with paired "_min" and "_max" columns
- iGAP: data.frame with columns containing comma-separated values (e.g., "1.2,3.4")
- SODAS: character string ending with ".xml" (file path)
- SDS: alias for SODAS

Value

A character string indicating the detected format: "RSDA", "MM", "iGAP", "SODAS", or "unknown"

Examples

```
data(mushroom.int)
int_detect_format(mushroom.int)  # Should return "RSDA"

data(abalone.iGAP)
int_detect_format(abalone.iGAP)  # Should return "iGAP"
```

`int_list_conversions` *List Available Format Conversions*

Description

List all available format conversion functions.

Usage

```
int_list_conversions(from = NULL, to = NULL)
```

Arguments

`from`	source format (optional): "RSDA", "MM", "iGAP", "SODAS"
`to`	target format (optional): "RSDA", "MM", "iGAP", "SODAS"

Value

A data.frame showing available conversions

Examples

```
# List all conversions
int_list_conversions()

# List conversions from RSDA
int_list_conversions(from = "RSDA")

# List conversions to MM
int_list_conversions(to = "MM")
```

lackinfo.int *Lack of Information Questionnaire Interval Dataset*

Description

Interval-valued dataset from a lack-of-information questionnaire. Contains biographical data and responses to 5 items measuring perception of lack of information, collected via an interval-valued Likert scale.

Usage

```
data(lackinfo.int)
```

Format

A data frame with 50 observations and 8 variables:

- `id`: Identification number.
- `sex`: Sex of the respondent (`male` or `female`).
- `age`: Respondent's age (in years).
- `item1`: Interval-valued answer to item 1.
- `item2`: Interval-valued answer to item 2.
- `item3`: Interval-valued answer to item 3.
- `item4`: Interval-valued answer to item 4.
- `item5`: Interval-valued answer to item 5.

Details

An educational innovation project was carried out for improving teaching-learning processes at the University of Oviedo (Spain) for the 2020/2021 academic year. A total of 50 students answered an online questionnaire about biographical data (sex and age) and their perception of lack of information by selecting the interval that best represents their level of agreement on a scale bounded between 1 (strongly disagree) and 7 (strongly agree).

The 5 items measuring perception of lack of information are:

- I1: I receive too little information from my classmates.
- I2: It is difficult to receive relevant information from my classmates.
- I3: It is difficult to receive relevant information from the teacher.
- I4: The amount of information I receive from my classmates is very low.
- I5: The amount of information I receive from the teacher is very low.

Source

https://CRAN.R-project.org/package=IntervalQuestionStat

Examples

```
data(lackinfo.int)
```

loans_by_purpose.int *Loans by Purpose Interval Dataset*

Description

Interval-valued data for loan characteristics aggregated by their purpose. Original microdata contains 887,383 loan records from Kaggle.

Usage

```
data(loans_by_purpose.int)
```

Format

A data frame with 14 observations and 4 interval-valued variables:

- `ln_inc`: Natural logarithm of self-reported annual income.
- `ln_revolbal`: Natural logarithm of total credit revolving balance.
- `open_acc`: Number of open credit lines.
- `total_acc`: Total number of credit lines.

Source

https://CRAN.R-project.org/package=MAINT.Data

Examples

```
data(loans_by_purpose.int)
```

lung_cancer.hist *Lung Cancer Treatments by State Histogram-Valued Dataset*

Description

Histogram-valued distribution of lung cancer treatment counts by US state.

Usage

```
data(lung_cancer.hist)
```

Format

An object of class `data.frame` with 2 rows and 2 columns.

References

Billard, L. and Diday, E. (2006). *Symbolic Data Analysis*. Wiley. Table 2.20.

Examples

```
data(lung_cancer.hist)
```

MM_to_iGAP *MM to iGAP*

Description

To convert MM format to iGAP format.

Usage

```
MM_to_iGAP(data)
```

Arguments

`data`	The dataframe with the MM format.

Value

Return a dataframe with the iGAP format.

Examples

```
data(face.iGAP)
face <- iGAP_to_MM(face.iGAP, 1:6)
MM_to_iGAP(face)
```

MM_to_RSDA *MM to RSDA*

Description

To convert MM format interval dataframe to RSDA format (symbolic_tbl).

Usage

```
MM_to_RSDA(data)
```

Arguments

`data`	The dataframe with the MM format (paired _min/_max columns).

Value

Return a symbolic_tbl dataframe with complex-encoded interval columns.

Examples

```
data(mushroom.int)
mm <- RSDA_to_MM(mushroom.int, RSDA = FALSE)
rsda <- MM_to_RSDA(mm)
```

---

mushroom *Mushroom Species Dataset (Original Format)*

Description

Interval-valued data for 23 mushroom species of the genus Agaricus with 3 morphological measurements from the Fungi of California Species.

Usage

```
data(mushroom)
```

Format

A data frame with 23 observations and 5 variables:

- `Species`: Mushroom species name.
- `Pileus.Cap.Width`: Pileus cap width range (cm).
- `Stipe.Length`: Stipe length range (cm).
- `Stipe.Thickness`: Stipe thickness range (cm).
- `Edibility`: Edibility code (U/Y/N/T).

Details

Classic SDA dataset used for descriptive statistics, histogram construction, and clustering of interval-valued data.

Source

Billard, L. and Diday, E. (2006), Table 3.2.

References

Billard, L. and Diday, E. (2006). *Symbolic Data Analysis: Conceptual Statistics and Data Mining*. Wiley, Chichester. Table 3.2.

Examples

```
data(mushroom)
```

`mushroom.int` *Mushroom Species Interval Dataset*

Description

Interval-valued version of the mushroom dataset. See `mushroom`.

Usage

```
data(mushroom.int)
```

Format

A data frame with 23 observations and interval-valued variables.

References

Billard, L. and Diday, E. (2006). *Symbolic Data Analysis*. Wiley. Table 3.2.

Examples

```
data(mushroom.int)
```

`mushroom_fuzzy` *Mushroom Species Fuzzy/Symbolic Dataset*

Description

Extended mushroom data with fuzzy stipe thickness (Small/Average/Large), numerical stipe length, interval cap size, and categorical cap colour for two Amanita species.

Usage

```
data(mushroom_fuzzy)
```

Format

An object of class `data.frame` with 4 rows and 9 columns.

References

Diday, E. and Noirhomme-Fraiture, M. (Eds.) (2008). *Symbolic Data Analysis and the SODAS Software*. Wiley. Tables 1.14-1.16.

Examples

```
data(mushroom_fuzzy)
```

`nycflights.int` *New York City Flights Interval Dataset*

Description

Interval-valued dataset with 142 units and four interval-valued variables from the nycflights13 package, aggregated by month and carrier.

Usage

```
data(nycflights.int)
```

Format

A list containing FlightsDF, FlightsUnits, and FlightsIdt.

Source

`https://CRAN.R-project.org/package=MAINT.Data`

References

Duarte Silva, A.P., Brito, P., Filzmoser, P. and Dias, J.G. (2021). MAINT.Data: Modelling and Analysing Interval Data in R. *R Journal*, 13(2).

Examples

```
data(nycflights.int)
```

`occupations` *Occupation Salaries Dataset*

Description

Salary ranges for different occupations.

Usage

```
data(occupations)
```

Format

An object of class `data.frame` with 9 rows and 11 columns.

References

Billard, L. and Diday, E. (2006). *Symbolic Data Analysis*. Wiley.

Examples

```
data(occupations)
```

`occupations2` *Occupation Salaries Modal-Valued Dataset*

Description

Modal-valued version of the occupation salaries dataset.

Usage

```
data(occupations2)
```

Format

An object of class `symbolic_tbl` (inherits from `tbl_df`, `tbl`, `data.frame`) with 9 rows and 4 columns.

References

Billard, L. and Diday, E. (2006). *Symbolic Data Analysis*. Wiley.

Examples

```
data(occupations2)
```

`ohtemp.int` *Ohio River Basin 30-Year Trimmed Mean Daily Temperatures Interval Dataset*

Description

Interval-valued dataset of 30-year trimmed mean daily temperatures for the Ohio river basin. Intervals are defined by the mean daily maximum and minimum temperatures from January 1, 1988 to December 31, 2018.

Usage

```
data(ohtemp.int)
```

Format

A data frame with 161 rows and 7 variables:

- `ID`: Global Historical Climatological Network (GHCN) station identifier.
- `NAME`: GHCN station name.
- `STATE`: Two-digit state designation.
- `LATITUDE`: Latitude coordinate position.
- `LONGITUDE`: Longitude coordinate position.
- `ELEVATION`: Elevation of the measurement location (meters).
- `TEMPERATURE`: 30-year mean daily temperature (tenths of degrees Celsius).

Source

https://CRAN.R-project.org/package=intkrige

Examples

```
data(ohtemp.int)
```

---

oils.int *Oils and Fats Interval Dataset*

Description

Classic benchmark interval-valued data for 8 oils and fats described by 4 physico-chemical properties. Originally from Ichino (1988).

Usage

```
data(oils.int)
```

Format

A data frame with 8 observations and 4 interval-valued variables:

- `specific_gravity`: Specific gravity of the oil/fat.
- `freezing_point`: Freezing point (degrees Celsius).
- `iodine_value`: Iodine value.
- `saponification_value`: Saponification value.

Details

The 8 samples are: Linseed oil, Perilla oil, Cottonseed oil, Sesame oil, Camellia oil, Olive oil, Beef tallow, Hog fat. The expected 3-cluster structure is: {Beef tallow, Hog fat}, {Cottonseed, Sesame, Camellia, Olive}, and {Linseed, Perilla}. Widely used for comparing clustering methods and distance measures in symbolic data analysis.

References

Ichino, M. (1988). General metrics for mixed features. *Proc. IEEE Conf. Systems, Man, and Cybernetics*, pp. 494-497.

Diday, E. and Noirhomme-Fraiture, M. (Eds.) (2008). *Symbolic Data Analysis and the SODAS Software*. Wiley. Table 13.7, p.253.

Examples

```
data(oils.int)
```

profession.int *Profession Work Salary Time Interval Dataset*

Description

Interval-valued data for professional categories by salary and working time.

Usage

```
data(profession.int)
```

Format

An object of class `symbolic_tbl` (inherits from `tbl_df`, `tbl`, `data.frame`) with 15 rows and 4 columns.

References

Billard, L. and Diday, E. (2006). *Symbolic Data Analysis*. Wiley.

Examples

```
data(profession.int)
```

RSDA_format *RSDA Format*

Description

This function changes the format of the data to conform to RSDA format.

Usage

```
RSDA_format(data, sym_type1 = NULL, location = NULL, sym_type2 = NULL, var = NULL)
```

Arguments

data	A conventional data.
sym_type1	The labels I means an interval variable and $S means set variable.
location	The location of the sym_type in the data.
sym_type2	The labels I means an interval variable and $S means set variable.
var	The name of the symbolic variable in the data.

Value

Return a dataframe with a label added to the previous column of symbolic variable.

Examples

```
data("mushroom")
mushroom.set <- set_variable_format(data = mushroom, location = 8, var = "Species")
mushroom.tmp <- RSDA_format(data = mushroom.set, sym_type1 = c("I", "S"),
                              location = c(25, 31), sym_type2 = c("S", "I", "I"),
                            var = c("Species", "Stipe.Length_min", "Stipe.Thickness_min"))
```

RSDA_to_iGAP *RSDA to iGAP*

Description

To convert RSDA format interval dataframe to iGAP format.

Usage

```
RSDA_to_iGAP(data)
```

Arguments

data	The RSDA format with interval dataframe.

Value

Return a dataframe with the iGAP format.

Examples

```
data(mushroom.int)
RSDA_to_iGAP(mushroom.int)
```

RSDA_to_MM *RSDA to MM*

Description

To convert RSDA format interval dataframe to MM format.

Usage

```
RSDA_to_MM(data, RSDA)
```

Arguments

data	The RSDA format with interval dataframe.
RSDA	Whether to load the RSDA package.

Value

Return a dataframe with the MM format.

Examples

```
data(mushroom.int)
RSDA_to_MM(mushroom.int, RSDA = FALSE)
```

---

set_variable_format *Set Variable Format*

Description

This function changes the format of the set variables in the data to conform to the RSDA format.

Usage

```
set_variable_format(data, location, var)
```

Arguments

data	A conventional data.
location	The location of the set variable in the data.
var	The name of the set variable in the data.

Value

Return a dataframe in which a set variable is converted to one-hot encoding.

Examples

```
data("mushroom")
mushroom.set <- set_variable_format(data = mushroom, location = 8, var = "Species")
```

soccer_bivar.int *French Soccer Championship Bivariate Interval Dataset*

Description

Interval-valued data for 20 teams from the French premier soccer championship. Contains ranges of Weight (response), Height and Age (explanatory variables).

Usage

```
data(soccer_bivar.int)
```

Format

A data frame with 20 rows and 3 interval-valued variables:

- y: Weight (response variable, kg).
- t1: Height (explanatory variable, cm).
- t2: Age (explanatory variable, years).

Source

https://CRAN.R-project.org/package=iRegression

References

Lima Neto, E. A., Cordeiro, G. and De Carvalho, F.A.T. (2011). Bivariate symbolic regression models for interval-valued variables. *Journal of Statistical Computation and Simulation*, 81, 1727-1744.

Examples

```
data(soccer_bivar.int)
```

SODAS_to_iGAP *SODAS to iGAP*

Description

To convert SODAS format interval dataframe to the iGAP format.

Usage

```
SODAS_to_iGAP(XMLPath)
```

Arguments

XMLPath	Disk path where the SODAS *.XML file is.

Value

Return a dataframe with the iGAP format.

Examples

```
## Not run:
data(abalone.int)
```

SODAS_to_MM *SODAS to MM*

Description

To convert SODAS format interval dataframe to the MM format.

Usage

```
SODAS_to_MM(XMLPath)
```

Arguments

XMLPath	Disk path where the SODAS *.XML file is.

Value

Return a dataframe with the MM format.

Examples

```
## Not run:
data(abalone.int)
```

teams.int *Pickup League Teams Interval Dataset*

Description

Interval-valued data for 5 teams in a local pickup league, classified by season performance. Each team is described by ranges of player age, weight, and speed.

Usage

```
data(teams.int)
```

Format

A data frame with 5 observations and 4 variables:

- `team_type`: Performance category (Very Good, Good, Average, Fair, Poor).
- `age`: Player age range (years).
- `weight`: Player weight range (pounds).
- `speed`: Speed range – time to run 100 yards (seconds).

Details

The symbolic results are more informative than classical midpoint analyses: the Very Good team has homogeneous players, whereas the Poor team has players varying widely in age, weight, and speed. Used for symbolic principal component analysis.

References

Billard, L. and Diday, E. (2006). *Symbolic Data Analysis*. Wiley. Table 2.24, p.63.

Examples

```
data(teams.int)
```

temperature_city.int *World Cities Monthly Temperature Interval Dataset*

Description

Interval-valued monthly temperatures for major cities worldwide. Benchmark dataset for comparing distance measures (Hausdorff, L2, Wasserstein) in dynamic clustering algorithms.

Usage

```
data(temperature_city.int)
```

Format

A data frame with city rows and 12 interval-valued monthly temperature variables (Jan-Dec), plus an expert class assignment.

Details

Expert partition into 4 classes: Class 1 (tropical/warm), Class 2 (temperate European and Asian), Class 3 (Mauritius), Class 4 (Tehran).

References

Verde, R. and Irpino, A. (2008). A new interval data distance based on the Wasserstein metric. *Proc. COMPSTAT 2008*, pp. 705-712.

Examples

```
data(temperature_city.int)
```

tennis.int *Tennis Court Types Interval Dataset*

Description

Interval-valued data for tennis players aggregated by court type (Hard, Grass, Indoor, Clay) with weight, height, and racket tension.

Usage

```
data(tennis.int)
```

Format

A data frame with 4 observations and 4 variables:

- `court_type`: Type of court (Hard, Grass, Indoor, Clay).
- `player_weight`: Player weight range (kg).
- `player_height`: Player height range (m).
- `racket_tension`: Racket tension range.

Details

Clustering on weight and height separates grass courts from the rest (decision rule: Weight <= 74.75 kg). When all three variables are used, clustering separates by racket tension instead.

References

Billard, L. and Diday, E. (2006). *Symbolic Data Analysis*. Wiley. Table 2.25, p.64.

Examples

```
data(tennis.int)
```

town_services.mix *Town Services Concatenated Mixed Symbolic Dataset*

Description

Symbolic data for 3 towns (Paris, Lyon, Toulouse) combining school and hospital databases. Contains interval-valued, multi-valued, and modal-valued variables.

Usage

```
data(town_services.mix)
```

Format

A data frame with 3 observations and 5 symbolic variables:

- `no_pupils`: Number of pupils range (interval).
- `type`: School type (modal).
- `level`: Coded level (multi-valued).
- `no_beds`: Number of beds range (interval).
- `specialty`: Specialty code (multi-valued).

References

Diday, E. and Noirhomme-Fraiture, M. (Eds.) (2008). *Symbolic Data Analysis and the SODAS Software*. Wiley. Table 1.21, p.19.

Examples

```
data(town_services.mix)
```

trivial_intervals.int *Trivial and Non-Trivial Intervals Example Dataset*

Description

Simple 5x3 example illustrating different interval types: full intervals (hyperrectangles), degenerate intervals (lines), and trivial intervals (points). Used for vertices PCA demonstration.

Usage

```
data(trivial_intervals.int)
```

Format

A data frame with 5 observations and 3 interval-valued variables.

References

Billard, L. and Diday, E. (2006). *Symbolic Data Analysis*. Wiley. Table 5.1, p.146.

Examples

```
data(trivial_intervals.int)
```

veterinary.int *Veterinary Interval Dataset*

Description

Interval-valued veterinary dataset with animal measurements.

Usage

```
data(veterinary.int)
```

Format

An object of class `symbolic_tbl` (inherits from `tbl_df`, `tbl`, `data.frame`) with 10 rows and 3 columns.

References

Billard, L. and Diday, E. (2006). *Symbolic Data Analysis*. Wiley.

Examples

```
data(veterinary.int)
```

world_cup.int *World Cup Soccer Teams Interval Dataset*

Description

Interval-valued data for soccer teams grouped by World Cup qualification status. Includes age, weight, height ranges and covariance.

Usage

```
data(world_cup.int)
```

Format

A data frame with 2 observations and 5 variables.

References

Diday, E. and Noirhomme-Fraiture, M. (Eds.) (2008). *Symbolic Data Analysis and the SODAS Software*. Wiley. Table 1.9, p.13.

Examples

```
data(world_cup.int)
```

write_csv_table *Write Symbolic Data Table*

Description

This function write (save) a symbolic data table from a CSV data file.

Usage

```
write_csv_table(data, file, output)
```

Arguments

`data`	The conventional data.
`file`	The name of the CSV file.
`output`	This is an experimental argument, with default TRUE, and can be ignored by most users.

Value

Write in CSV file the symbolic data table.

Examples

```
data(mushroom)
mushroom.set <- set_variable_format(data = mushroom, location = 8, var = "Species")
mushroom.tmp <- RSDA_format(data = mushroom.set, sym_type1 = c("I", "S"),
                            location = c(25, 31), sym_type2 = c("S", "I", "I"),
                          var = c("Species", "Stipe.Length_min", "Stipe.Thickness_min"))
mushroom.clean <- clean_colnames(data = mushroom.tmp)
# We can save the file in CSV to RSDA format as follows:
write_csv_table(data = mushroom.clean, file = "mushroom_interval.csv", output = FALSE)
```

Index