

# Package ‘dataSDA’

July 22, 2025

**Type** Package

**Title** Datasets and Basic Statistics for Symbolic Data Analysis

**Version** 0.1.2

**Date** 2025-06-07

**Author** Po-Wei 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** 2025-06-07 23:30:09 UTC

## Contents

Abalone . . . . .	2
Abalone.iGAP . . . . .	3
age_cholesterol_weight.int . . . . .	4
airline_flights . . . . .	4
airline_flights2 . . . . .	5
baseball.int . . . . .	5
bird.int . . . . .	6

blood_pressure.int . . . . .	6
car.int . . . . .	7
Cars.int . . . . .	8
ChinaTemp.int . . . . .	8
clean_colnames . . . . .	9
crime . . . . .	9
crime2 . . . . .	10
Face.iGAP . . . . .	10
finance.int . . . . .	11
fuel_consumption . . . . .	11
health_insurance . . . . .	12
health_insurance2 . . . . .	13
hierarchy . . . . .	13
hierarchy.int . . . . .	14
histogram_stats . . . . .	14
horses.int . . . . .	15
iGAP_to_MM . . . . .	16
interval_stats . . . . .	16
lackinfo.int . . . . .	18
LoansbyPurpose.int . . . . .	19
MM_to_iGAP . . . . .	20
mushroom . . . . .	20
mushroom.int . . . . .	21
nycflights.int . . . . .	22
occupations . . . . .	23
occupations2 . . . . .	23
ohtemp.int . . . . .	24
profession.int . . . . .	25
RSDA_format . . . . .	25
RSDA_to_iGAP . . . . .	26
RSDA_to_MM . . . . .	26
set_variable_format . . . . .	27
soccer.bivar.int . . . . .	28
SODAS_to_iGAP . . . . .	29
SODAS_to_MM . . . . .	29
veterinary.int . . . . .	30
write_csv_table . . . . .	30
<b>Index</b>	<b>32</b>

---

Abalone

*Abalone Dataset*


---

### Description

A interval-valued data set containing 24 units, created from from the Abalone dataset (UCI Machine Learning Repository), after aggregating by sex and age.

**Usage**

```
data(Abalone)
```

**Format**

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

**References**

Billard L. and Diday E. (2006). Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.

**Examples**

```
data(Abalone)
```

---

Abalone.iGAP

*Abalone iGAP format Dataset*

---

**Description**

A interval-valued data set containing 24 units, created from the Abalone dataset (UCI Machine Learning Repository), after aggregating by sex and age.

**Usage**

```
data(Abalone.iGAP)
```

**Format**

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

**References**

Billard L. and Diday E. (2006). Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.

**Examples**

```
data(Abalone.iGAP)
```

---

age\_cholesterol\_weight.int  
*Age-cholesterol-weight Interval-Valued Dataset*

---

**Description**

Age-cholesterol-weight Interval-Valued Dataset.

**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: Conceptual statistics and data mining. Wiley, Chichester.

**Examples**

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

---

airline\_flights      *Airline Flights Dataset*

---

**Description**

Airline Flights Dataset.

**Usage**

```
data(airline_flights)
```

**Format**

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

**References**

Billard L. and Diday E. (2006). Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.

**Examples**

```
data(airline_flights)
```

---

airline_flights2	<i>Airline Flights Modal-Valued Dataset</i>
------------------	---

---

**Description**

Airline Flights Modal-Valued Dataset.

**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: Conceptual statistics and data mining. Wiley, Chichester.

**Examples**

```
data(airline_flights2)
```

---

baseball.int	<i>Baseball Interval-Valued Dataset</i>
--------------	---

---

**Description**

Baseball Interval-Valued Dataset.

**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: Conceptual statistics and data mining. Wiley, Chichester.

**Examples**

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

---

bird.int	<i>Bird Interval-Valued Dataset</i>
----------	-------------------------------------

---

**Description**

Bird Interval-Valued Dataset.

**Usage**

```
data(bird.int)
```

**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: Conceptual statistics and data mining. Wiley, Chichester.

**Examples**

```
data(bird.int)
```

---

blood_pressure.int	<i>Blood Pressure Interval-Valued Dataset</i>
--------------------	---

---

**Description**

blood pressure Interval-Valued Dataset.

**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: Conceptual statistics and data mining. Wiley, Chichester.

**Examples**

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

---

car.int

*Car Interval-Valued Dataset*

---

**Description**

Car Interval-Valued Dataset.

**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: Conceptual statistics and data mining. Wiley, Chichester.

**Examples**

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

---

`Cars.int`*Cars Interval Dataset*

---

**Description**

Cars Interval Dataset generated from Cars dataset. This data set consist of the intervals for four characteristics (Price, EngineCapacity, TopSpeed and Acceleration) of 27 cars models partitioned into four different classes (Utilitarian, Berlina, Sportive and Luxury).

**Usage**

```
data(Cars.int)
```

**Format**

A data frame containing 27 observations on 5 variables, the first five with the interval characteristics for 27 car models, the last one a factor indicating the model class.

**Source**

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

**Examples**

```
data(Cars.int)
```

---

`ChinaTemp.int`*China Temperatures Interval Dataset*

---

**Description**

China Temperatures Interval Dataset generated from ChinaTemp dataset. This data set consist of the intervals of observed temperatures (Celsius scale) in each of the four quarters, Q\_1 to Q\_4, of the years 1974 to 1988 in 60 chinese meteorologic stations; one outlier observation (YinChuan\_1982) has been discarded. The 60 stations belong to different regions in China, which therefore define a partition of the 899 stations-year combinations.

**Usage**

```
data(ChinaTemp.int)
```

**Format**

A data frame containing 899 observations on 5 variables, the first four with the temperatures by quarter in the 899 stations-year combinations, the last one a factor indicating the geographic region of each station.



**Source**

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

**Examples**

```
data(ChinaTemp.int)
```

---

clean_colnames	<i>clean_colnames</i>
----------------	-----------------------

---

**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)
```

---

crime	<i>Crime demographics Dataset</i>
-------	-----------------------------------

---

**Description**

Crime demographics Dataset.

**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: Conceptual statistics and data mining. Wiley, Chichester.

**Examples**

```
data(crime)
```

---

crime2	<i>Crime demographics Modal-Valued Dataset</i>
--------	--

---

**Description**

Crime demographics Modal-Valued 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: Conceptual statistics and data mining. Wiley, Chichester.

**Examples**

```
data(crime2)
```

---

Face.iGAP	<i>Face iGAP format Dataset</i>
-----------	---------------------------------

---

**Description**

Symbolic data matrix with all the variables of interval type.

**Usage**

```
data(Face.iGAP)
```

**Format**

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

**References**

Billard L. and Diday E. (2006).Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.

**Examples**

```
data(Face.iGAP)
```

---

finance.int	<i>Finance Interval-Valued Dataset</i>
-------------	--

---

**Description**

Finance Interval-Valued Dataset.

**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: Conceptual statistics and data mining. Wiley, Chichester.

**Examples**

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

---

fuel_consumption	<i>Fuel Consumption Dataset</i>
------------------	---------------------------------

---

**Description**

Fuel Consumption Dataset.

**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: Conceptual statistics and data mining. Wiley, Chichester.

**Examples**

```
data(fuel_consumption)
```

---

health_insurance	<i>Health Insurance Dataset</i>
------------------	---------------------------------

---

**Description**

Health Insurance Dataset.

**Usage**

```
data(health_insurance)
```

**Format**

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

**References**

Billard L. and Diday E. (2006). Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.

**Examples**

```
data(health_insurance)
```

---

health_insurance2	<i>Health Insurance Modal-Valued Dataset</i>
-------------------	--

---

**Description**

Health Insurance Modal-Valued 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: Conceptual statistics and data mining. Wiley, Chichester.

**Examples**

```
data(health_insurance2)
```

---

hierarchy	<i>Hierarchy Dataset</i>
-----------	--------------------------

---

**Description**

Hierarchy Dataset.

**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: Conceptual statistics and data mining. Wiley, Chichester.

**Examples**

```
data(hierarchy)
```

---

hierarchy.int	<i>Hierarchy Interval-Valued Dataset</i>
---------------	--

---

**Description**

Hierarchy Interval-Valued 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: Conceptual statistics and data mining. Wiley, Chichester.

**Examples**

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

---

histogram_stats	<i>Statistics for Histogram Data</i>
-----------------	--------------------------------------

---

**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

*Horses Interval-Valued Dataset*

---

**Description**

Horses Interval-Valued Dataset.

**Usage**

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

**Format**

An object of class `symbolic_tbl` (inherits from `tbl_df`, `tbl`, `data.frame`) with 8 rows and 7 columns.

**References**

Billard L. and Diday E. (2006). Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.

**Examples**

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

---

iGAP_to_MM	<i>iGAP to MM</i>
------------	-------------------

---

**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)
```

---

interval_stats	<i>Statistics for Interval Data</i>
----------------	-------------------------------------

---

**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 <code>symbolic_tbl</code> 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)
```

---

`lackinfo.int`*Lack of information questionnaire interval dataset.*

---

### Description

Lack of information questionnaire interval dataset generated from lackinfo dataset. A dataset containing some biographical data and the responses to 5 items measuring the perception of lack of information in a questionnaire.

### Usage

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

### Format

A data frame with 50 observations of the following 8 variables:

- `id`: identification number.
- `sex`: sex of the respondent (male or female).
- `age`: respondent's age (in years).
- `item1`: respondent's interval-valued answer to item 1.
- `item2`: respondent's interval-valued answer to item 2.
- `item3`: respondent's interval-valued answer to item 3.
- `item4`: respondent's interval-valued answer to item 4.
- `item5`: respondent's 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 have been requested to answer an online questionnaire about some biographical data (sex and age) and their perception of lack of information by selecting the interval that best represents their level of agreement to the statements proposed in a interval-valued scale bounded between 1 and 7, where 1 represents the option 'strongly disagree' and 7 represents the option 'strongly agree'.

These are the 5 items used to measure the perception of lack of information:

- 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)
```

---

LoansbyPurpose.int      *Loans by purpose: Interval Dataset*

---

**Description**

Loans by purpose interval dataset generated from LoansbyPurpose dataset. This data set consist of the lower and upper bounds of the intervals for four interval characteristics of the loans aggregated by their purpose. The original microdata is available at the Kaggle Data Science platform and consists of 887 383 loan records characterized by 75 descriptors. Among the large set of variables available, we focus on borrowers' income and account and loan information aggregated by the 14 loan purposes, wich are considered as the units of interest.

**Usage**

```
data(LoansbyPurpose.int)
```

**Format**

A data frame containing 14 observations on the following 4 variables:

- `ln-inc`: The current loan purpose of natural logarithm of the self-reported annual income provided by the borrower during registration
- `ln-revolbal`: The current loan purpose of natural logarithm of the total credit revolving balance
- `open-acc`: The current loan purpose of the number of open credit lines in the borrower's credit file
- `total-acc`: The current loan purpose, of the total number of credit lines currently in the borrower's credit file

**Source**

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

**Examples**

```
data(LoansbyPurpose.int)
```

---

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)
```

---

mushroom

*Mushroom Data Set*


---

**Description**

The mushroom data set consists of a set of 23 species described by 3 interval variables. These mushroom species are members of the genus *Agaricis*. The specific variables and their values are extracted from the *Fungi of California Species*.

**Usage**

```
data(mushroom)
```

**Format**

A data frame with 23 observations and 5 variables named Species, Pileus Cap Width, Stipe Length, Stipe Thickness, and Edibility.

- Species: The class of mushroom.
- Pileus Cap Width: The pileus cap width of the mushroom.
- Stipe Length: The stipe length of the mushroom.
- Stipe Thickness: The stipe thickness of the mushroom.
- Edibility: The edibility of mushroom (U: unknown, Y: Yes, N: No, T: Toxic).

**Source**

Billard, L. and Diday, E. (2006) Symbolic Data Analysis: Conceptual Statistics and Data Mining John Wiley & Sons, Ltd.

**References**

Billard L. and Diday E. (2006).Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.

**Examples**

```
data(mushroom)
```

---

mushroom.int

*Mushroom Interval Dataset*

---

**Description**

Mushroom interval dataset generated from mushroom dataset. The mushroom data set consists of a set of 23 species described by 3 interval variables. These mushroom species are members of the genus Agaricies. The specific variables and their values are extracted from the Fungi of California Species.

**Usage**

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

**Format**

A data frame with 23 observations and 5 variables named Species, Pileus Cap Width, Stipe Length, Stipe Thickness, and Edibility.

- Species: The class of mushroom.
- Pileus Cap Width: The pileus cap width of the mushroom.
- Stipe Length: The stipe length of the mushroom.
- Stipe Thickness: The stipe thickness of the mushroom.
- Edibility: The edibility of mushroom (U: unknown, Y: Yes, N: No, T: Toxic).

**Source**

Billard, L. and Diday, E. (2006) Symbolic Data Analysis: Conceptual Statistics and Data Mining John Wiley & Sons, Ltd.

**References**

Billard L. and Diday E. (2006).Symbolic data analysis: Conceptual statistics and data mining. Wiley, Chichester.

## Examples

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

---

nycflights.int

*New York City flights Interval Dataset*

---

## Description

New York City flights interval dataset generated from nycflights dataset. A interval-valued data set containing 142 units and four interval-valued variables (dep\_delay, arr\_delay, air\_time and distance), created from from the flights data set in the R package nycflights13 (on-time data for all flights that departed the JFK, LGA or EWR airports in 2013), after removing all rows with missing observations, and aggregating by month and carrier.

## Usage

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

## Format

**FlightsDF** A data frame containing the original 327346 valid (i.e. with non missing values) flights from the nycflights13 package, described by the 4 variables: dep\_delay, arr\_delay, air\_time and distance.

**FlightsUnits** A factor with 327346 observations and 142 levels, indicating the month by carrier combination to which each original flight belongs to.

**FlightsIdt** An IData object with 142 observations and 4 interval-valued variables, describing the intervals formed by agregating the FlightsDF microdata by the 0.05 and 0.95 quantiles of the subsamples formed by FlightsUnits factor.

## Source

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

## References

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

## Examples

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

---

occupations	<i>Occupation Salaries Dataset</i>
-------------	------------------------------------

---

**Description**

Occupation Salaries Dataset.

**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: Conceptual statistics and data mining. Wiley, Chichester.

**Examples**

```
data(occupations)
```

---

occupations2	<i>Occupation Salaries Modal-Valued Dataset</i>
--------------	---

---

**Description**

Occupation Salaries Modal-Valued Dataset.

**Usage**

```
data(occupations)
```

**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: Conceptual statistics and data mining. Wiley, Chichester.

**Examples**

```
data(occupations2)
```

---

ohtemp.int	<i>30 year trimmed mean daily temperatures interval dataset for the Ohio river basin.</i>
------------	---

---

### Description

30 year trimmed mean daily temperatures interval dataset for the Ohio river basin generated from ohtemp dataset. Intervals are defined by the mean daily maximum and minimum temperatures for the Ohio river basin from January 1, 1988 - December 31, 2018. The 116 observations in this dataset all had at least 300 daily observations of temperature in at least 30 of the 31 considered years. The mean was calculated after trimming 10 influence of potential outliers.

### Usage

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

### Format

A data frame with 161 rows and 7 variables:

- ID: The global historical climatological network (GHCN) station identifier
- NAME: The GHCN station name
- STATE: The two-digit designation for the state in which each station resides
- LATITUDE: Latitude coordinate position
- LONGITUDE: Longitude coordinate position
- ELEVATION: Elevation of the measurement location (meters)
- TEMPERATURE: The 30 year mean daily temperature (tenths of degrees Celsius)

### Source

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

### Examples

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



---

profession.int	<i>Profession Work Salary Time Interval-Valued Dataset</i>
----------------	--

---

**Description**

Profession Work Salary Time Interval-Valued Dataset.

**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: Conceptual statistics and data mining. Wiley, Chichester.

**Examples**

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

---

RSDA_format	<i>RSDA Format</i>
-------------	--------------------

---

**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**

<code>data</code>	A conventional data.
<code>sym_type1</code>	The labels I means an interval variable and \$\$ means set variable.
<code>location</code>	The location of the <code>sym_type</code> in the data.
<code>sym_type2</code>	The labels I means an interval variable and \$\$ means set variable.
<code>var</code>	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	<i>Soccer bivar Interval Data Set</i>
------------------	---------------------------------------

---

## Description

Soccer bivar interval dataset generated from soccer.bivar dataset. A real interval-valued data set.

## Usage

```
soccer.bivar.int
```

## Format

A data frame with 20 rows and 3 variables:

- y: The response variable Y (weight)
- t1: The explanatory variable T1 (height)
- t2: The explanatory variable T2 (age)

## Details

This data set concerns the record of the Weight (Y), Height (T1) and Age (T2) from 20 soccer teams of the premiere French championship.

## 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 (Print)*, 81, 1727–1744.

## Examples

```
data(soccer.bivar.int)
```

---

SODAS_to_iGAP	<i>SODAS to iGAP</i>
---------------	----------------------

---

**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)
```

---

SODAS_to_MM	<i>SODAS to MM</i>
-------------	--------------------

---

**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)
```

---

veterinary.int	<i>Veterinary Interval-Valued Dataset</i>
----------------	---

---

**Description**

Veterinary Interval-Valued Dataset.

**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: Conceptual statistics and data mining. Wiley, Chichester.

**Examples**

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

---

write_csv_table	<i>Write Symbolic Data Table</i>
-----------------	----------------------------------

---

**Description**

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

**Usage**

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

**Arguments**

<code>data</code>	The conventional data.
<code>file</code>	The name of the CSV file.
<code>output</code>	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

## \* datasets

Abalone, 2  
Abalone.iGAP, 3  
age\_cholesterol\_weight.int, 4  
airline\_flights, 4  
airline\_flights2, 5  
baseball.int, 5  
bird.int, 6  
blood\_pressure.int, 6  
car.int, 7  
Cars.int, 8  
ChinaTemp.int, 8  
crime, 9  
crime2, 10  
Face.iGAP, 10  
finance.int, 11  
fuel\_consumption, 11  
health\_insurance, 12  
health\_insurance2, 13  
hierarchy, 13  
hierarchy.int, 14  
horses.int, 15  
lackinfo.int, 18  
LoansbyPurpose.int, 19  
mushroom, 20  
mushroom.int, 21  
nycflights.int, 22  
occupations, 23  
occupations2, 23  
ohtemp.int, 24  
profession.int, 25  
soccer.bivar.int, 28  
veterinary.int, 30

Abalone, 2  
Abalone.iGAP, 3  
age\_cholesterol\_weight.int, 4  
airline\_flights, 4  
airline\_flights2, 5

baseball.int, 5  
bird.int, 6  
blood\_pressure.int, 6

car.int, 7  
Cars.int, 8  
ChinaTemp.int, 8  
clean\_colnames, 9  
crime, 9  
crime2, 10

Face.iGAP, 10  
finance.int, 11  
fuel\_consumption, 11

health\_insurance, 12  
health\_insurance2, 13  
hierarchy, 13  
hierarchy.int, 14  
hist\_cor(histogram\_stats), 14  
hist\_cov(histogram\_stats), 14  
hist\_mean(histogram\_stats), 14  
hist\_var(histogram\_stats), 14  
histogram\_stats, 14  
horses.int, 15

iGAP\_to\_MM, 16  
int\_cor(interval\_stats), 16  
int\_cov(interval\_stats), 16  
int\_mean(interval\_stats), 16  
int\_var(interval\_stats), 16  
interval\_stats, 16

lackinfo.int, 18  
LoansbyPurpose.int, 19

MM\_to\_iGAP, 20  
mushroom, 20  
mushroom.int, 21  
nycflights.int, 22



occupations, [23](#)  
occupations2, [23](#)  
ohtemp.int, [24](#)  
  
profession.int, [25](#)  
  
RSDA\_format, [25](#)  
RSDA\_to\_iGAP, [26](#)  
RSDA\_to\_MM, [26](#)  
  
set\_variable\_format, [27](#)  
soccer.bivar.int, [28](#)  
SODAS\_to\_iGAP, [29](#)  
SODAS\_to\_MM, [29](#)  
  
veterinary.int, [30](#)  
  
write\_csv\_table, [30](#)