|  |  |
| --- | --- |
| data.frame {base} | R Documentation |

**Data Frames**

**Description**

The function data.frame() creates data frames, tightly coupled collections of variables which share many of the properties of matrices and of lists, used as the fundamental data structure by most of **R**'s modeling software.

**Usage**

data.frame(..., row.names = NULL, check.rows = FALSE,

check.names = TRUE, fix.empty.names = TRUE,

stringsAsFactors = default.stringsAsFactors())

default.stringsAsFactors()

**Arguments**

|  |  |
| --- | --- |
| ... | these arguments are of either the form value or tag = value. Component names are created based on the tag (if present) or the deparsed argument itself. |
| row.names | NULL or a single integer or character string specifying a column to be used as row names, or a character or integer vector giving the row names for the data frame. |
| check.rows | if TRUE then the rows are checked for consistency of length and names. |
| check.names | logical. If TRUE then the names of the variables in the data frame are checked to ensure that they are syntactically valid variable names and are not duplicated. If necessary they are adjusted (by [make.names](http://127.0.0.1:14695/library/base/help/make.names)) so that they are. |
| fix.empty.names | logical indicating if arguments which are “unnamed” (in the sense of not being formally called as someName = arg) get an automatically constructed name or rather name "". Needs to be set to FALSE even when check.names is false if "" names should be kept. |
| stringsAsFactors | logical: should character vectors be converted to factors? The ‘factory-fresh’ default is TRUE, but this can be changed by setting [options](http://127.0.0.1:14695/library/base/help/options)(stringsAsFactors = FALSE). |

**Details**

A data frame is a list of variables of the same number of rows with unique row names, given class "data.frame". If no variables are included, the row names determine the number of rows.

The column names should be non-empty, and attempts to use empty names will have unsupported results. Duplicate column names are allowed, but you need to use check.names = FALSE for data.frame to generate such a data frame. However, not all operations on data frames will preserve duplicated column names: for example matrix-like subsetting will force column names in the result to be unique.

data.frame converts each of its arguments to a data frame by calling [as.data.frame](http://127.0.0.1:14695/library/base/help/as.data.frame)(optional = TRUE). As that is a generic function, methods can be written to change the behaviour of arguments according to their classes: **R** comes with many such methods. Character variables passed to data.frame are converted to factor columns unless protected by [I](http://127.0.0.1:14695/library/base/help/I) or argument stringsAsFactors is false. If a list or data frame or matrix is passed to data.frame it is as if each component or column had been passed as a separate argument (except for matrices of class "[model.matrix](http://127.0.0.1:14695/library/base/help/model.matrix)" and those protected by [I](http://127.0.0.1:14695/library/base/help/I)).

Objects passed to data.frame should have the same number of rows, but atomic vectors (see [is.vector](http://127.0.0.1:14695/library/base/help/is.vector)), factors and character vectors protected by [I](http://127.0.0.1:14695/library/base/help/I) will be recycled a whole number of times if necessary (including as elements of list arguments).

If row names are not supplied in the call to data.frame, the row names are taken from the first component that has suitable names, for example a named vector or a matrix with rownames or a data frame. (If that component is subsequently recycled, the names are discarded with a warning.) If row.names was supplied as NULL or no suitable component was found the row names are the integer sequence starting at one (and such row names are considered to be ‘automatic’, and not preserved by [as.matrix](http://127.0.0.1:14695/library/base/help/as.matrix)).

If row names are supplied of length one and the data frame has a single row, the row.names is taken to specify the row names and not a column (by name or number).

Names are removed from vector inputs not protected by [I](http://127.0.0.1:14695/library/base/help/I).

default.stringsAsFactors is a utility that takes [getOption](http://127.0.0.1:14695/library/base/help/getOption)("stringsAsFactors") and ensures the result is TRUE or FALSE (or throws an error if the value is not NULL).

**Value**

A data frame, a matrix-like structure whose columns may be of differing types (numeric, logical, factor and character and so on).

How the names of the data frame are created is complex, and the rest of this paragraph is only the basic story. If the arguments are all named and simple objects (not lists, matrices of data frames) then the argument names give the column names. For an unnamed simple argument, a deparsed version of the argument is used as the name (with an enclosing I(...) removed). For a named matrix/list/data frame argument with more than one named column, the names of the columns are the name of the argument followed by a dot and the column name inside the argument: if the argument is unnamed, the argument's column names are used. For a named or unnamed matrix/list/data frame argument that contains a single column, the column name in the result is the column name in the argument. Finally, the names are adjusted to be unique and syntactically valid unless check.names = FALSE.

**Note**

In versions of **R** prior to 2.4.0 row.names had to be character: to ensure compatibility with such versions of **R**, supply a character vector as the row.names argument.

**References**

Chambers, J. M. (1992) *Data for models.* Chapter 3 of *Statistical Models in S* eds J. M. Chambers and T. J. Hastie, Wadsworth & Brooks/Cole.

**See Also**

[I](http://127.0.0.1:14695/library/base/help/I), [plot.data.frame](http://127.0.0.1:14695/library/base/help/plot.data.frame), [print.data.frame](http://127.0.0.1:14695/library/base/help/print.data.frame), [row.names](http://127.0.0.1:14695/library/base/help/row.names), [names](http://127.0.0.1:14695/library/base/help/names) (for the column names), [[.data.frame](http://127.0.0.1:14695/library/base/help/%5B.data.frame) for subsetting methods, [Math.data.frame](http://127.0.0.1:14695/library/base/help/Math.data.frame) etc, about *Group* methods for data.frames; [read.table](http://127.0.0.1:14695/library/base/help/read.table), [make.names](http://127.0.0.1:14695/library/base/help/make.names).

**Examples**

L3 <- LETTERS[1:3]

fac <- sample(L3, 10, replace = TRUE)

(d <- data.frame(x = 1, y = 1:10, fac = fac))

## The "same" with automatic column names:

data.frame(1, 1:10, sample(L3, 10, replace = TRUE))

is.data.frame(d)

## do not convert to factor, using I() :

(dd <- cbind(d, char = I(letters[1:10])))

rbind(class = sapply(dd, class), mode = sapply(dd, mode))

stopifnot(1:10 == row.names(d)) # {coercion}

(d0 <- d[, FALSE]) # data frame with 0 columns and 10 rows

(d.0 <- d[FALSE, ]) # <0 rows> data frame (3 named cols)

(d00 <- d0[FALSE, ]) # data frame with 0 columns and 0 rows