

# Package ‘fastDummies’

August 26, 2017

**Type** Package

**Title** Fast Creation of Dummy (Binary) Columns from Categorical Variables

**Version** 0.1.2

**Description** Creates dummy columns from columns that have categorical variables (character or factor types). You can also specify which columns to make dummies out of, or which columns to ignore. This package provides a significant speed increase from creating dummy variables through `model.matrix()`.

**Depends** R (>= 2.10), data.table

**Imports**

**License** GPL

**Encoding** UTF-8

**LazyData** true

**URL** <https://github.com/jacobkap/fastDummies>

**BugReports** <https://github.com/jacobkap/fastDummies/issues>

**RoxygenNote** 6.0.1

**NeedsCompilation** no

**Author** Jacob Kaplan [aut, cre],  
Benjamin Schlegel [ctb]

**Maintainer** Jacob Kaplan <jacobkap@sas.upenn.edu>

**Repository** CRAN

**Date/Publication** 2017-08-26 11:14:20 UTC

## R topics documented:

dummies_example . . . . .	2
dummy_cols . . . . .	3
dummy_rows . . . . .	4
dummy_rows_example . . . . .	5

<b>Index</b>	<b>6</b>
--------------	----------

---

dummies\_example

*United States Census data for 2015*

---

### **Description**

A dataset containing Census results from the American Community Survey 2015

### **Usage**

dummies\_example

### **Format**

A data frame with 100,000 rows and 17 variables:

**YEAR** Year of the survey

**STATEFIP** FIPS ID for the state

**COUNTYFIPS** FIPS ID for the county

**OWNERSHP** Respondent's ownership status for their home

**PERWT** Survey weight for the respondent

**NCHILD** Number of children in the home

**SEX** Sex of the respondent

**AGE** Age of the respondent

**MARST** Marital status of the respondent

**RACE** Race of the respondent

**HISPAN** If the respondent is hispanic

**BPL** Birthplace of the respondent

**EDUC** Education of the respondent

**EMPSTAT** Employment status of the respondent

**FTOTINC** Family yearly income

**POVERTY** If they are below the poverty line

**MOVEDIN** When the respondent moved into their home

### **Source**

<https://usa.ipums.org/usa-action/variables/group>

---

dummy\_cols                      *Fast creation of dummy variables*

---

## Description

Fast creation of dummy variables

## Usage

```
dummy_cols(dataset, select_columns = NULL, ignore_columns = NULL,
  remove_original = TRUE, dummy_columns_only = FALSE,
  remove_first_dummy = FALSE, conditional_columns = NULL,
  return_type = "data.table")
```

## Arguments

dataset	data.table or data.frame
select_columns	Vector of column names that you want to create dummy variables from. Default uses all character or factor columns.
ignore_columns	Vector of column names to ignore_ Default ignores all numeric columns.
remove_original	Removes the columns used to make dummy variables. Columns that are not used to make dummy variables are not affected.
dummy_columns_only	Removes all columns that didn't create dummy columns (i.e. numeric columns).
remove_first_dummy	Removes the first dummy of every variable that only n-1 Dummies remain
conditional_columns	Select column(s) to multiple other dummy columns created by. Useful to get subcategories of data. e.g. conditional column is gender and other columns are crimes. This will create columns showing the number of each crime for each gender.
return_type	Type of data you want back_ Default is data.table (better for use with large data)_ Other options are data.frame or matrix.

## Value

data.table, data.frame, or matrix depending on input for return\_type. data.table is default.

## Examples

```
data(dummies_example)
example <- dummy_cols(dummies_example)

# Return data.frame
example <- dummy_cols(dummies_example, return_type = "data.frame")
```

```

# Only keep created dummy columns
example <- dummy_cols(dummies_example, dummy_columns_only = TRUE)

# Only keep SEX and RACE columns
example <- dummy_cols(dummies_example, select_columns = c("Sex", "RACE"))

# Keep all except SEX column
example <- dummy_cols(dummies_example, ignore_columns = "SEX")

# Removes the first dummy from every category. Avoids perfect
# multicollinearity issues in models.
example <- dummy_cols(dummies_example, remove_first_dummy = TRUE)

```

---

dummy\_rows

*Fast creation of dummy rows*


---

## Description

Fast creation of dummy rows

## Usage

```

dummy_rows(dataset, select_columns = NULL, add_columns = NULL,
  ignore_columns = NULL, dummy_value = 0, year = FALSE,
  return_type = "data.table", dummy_indicator = FALSE)

```

## Arguments

dataset	data.table or data.frame
select_columns	If NULL, uses character, factor, and Date columns to produce categories to make the dummy rows by. If not NULL, you manually enter a string or vector of columns name(s).
add_columns	String or vector of column name(s) to add to the selected columns. This is only if you want to use all character, factor, and Date columns (selected by default) and are adding additional columns of different data types.
ignore_columns	String or vector of column name(s) to exclude from the selected columns. These excluded columns will get the same dummy value of all non-selected columns
dummy_value	Value of the row for columns that are not selected. Default is a value of 0.
year	TRUE to include a column called year (capitalization is ignored) as one of the selected columns.
return_type	Type of data you want back_ Default is data.table (better for use with large data)_ Other option is data.frame.
dummy_indicator	Adds binary column to say if row is dummy or not (i.e. included in original data or not)

**Value**

data.table or data.frame depending on input for return\_type. data.table is default.

**Examples**

```
data(dummy_rows_example)

# Makes dummy rows using default column selection and year
# to make categories
example <- dummy_rows(dummy_rows_example, year = TRUE)

# Same as above but adds binary column indicating if the row
# is dummy or not
example <- dummy_rows(dummy_rows_example, year = TRUE, dummy_indicator = TRUE)
```

---

dummy\_rows\_example      *National Incident-Based Reporting System crime data*

---

**Description**

A dataset crime information from the 2005-2015 NIBRS

**Usage**

```
dummy_rows_example
```

**Format**

A data frame with 10,000 rows and 9 variables:

**state** State

**year** Year

**simple\_location** Whether the crime happened at home or not

**offender\_used** Drugs or alcohol offender suspected of using

**victim\_injury** Severity of victim's injury

**sexoffender** Sex of the offender

**offender\_age** Age of the offender

**raceoffender** Race of the offender

**spouse\_assault** Number of spousal assault crimes

**Source**

<http://www.icpsr.umich.edu/icpsrweb/NACJD/studies/3449>

# Index

## \*Topic **datasets**

dummies\_example, [2](#)

dummy\_rows\_example, [5](#)

dummies\_example, [2](#)

dummy\_cols, [3](#)

dummy\_rows, [4](#)

dummy\_rows\_example, [5](#)