

Introduction to R - Data Management prt 2

Christopher Maronga

March 01, 2019

Type Conversion

In some cases, you will need to change the type of a variable to suit the purpose for which you want to utilize it for. And for this, R provides functions to enable you achieve just that. These group of functions obey certain rules in the sense that at some point they do not give you desired results as we are going to see in examples.

For instance, you might have a character type of a variable that you want to convert into a factor variable, in this scenario you will simply call the function `as.factor(variable_name)`.

These family of `as.***` are essential and widely used for all type conversions.

NOTE: You can check the type of a variable by simply calling the function `class(variable_name)`

Introduction to dplyr for data transformation/Wrangling

We are going to focus on the **dplyr** package which is a core member of the group of packages called tidyverse.

dplyr package makes data transformation and manipulation easy by providing simple verbs that correspond to common data manipulation tasks which helps you translate your **thoughts** into code.

The main verbs for data manipulation provided by dplyr include:-

- `select()` and `rename()` to select variables based on their names/rename variables.
- `mutate()` and `transmute()` to add/create new variables.
- `filter()` to select cases/rows based on their values.
- `arrange()` to reorder the cases/rows.
- `summarise()` to condense multiple values to a single value. Can be used together with `group_by()`
- `sample_n()` and `sample_frac()` to take random samples.

Data Wrangling – Selecting columns (1/4)

For the task of selecting column names, we are going to cover the below approaches on how to achieve that and clearly explain the differences in each one of them:

- basic selecting/deselecting (*named approach* and *chunk approach*)
- selecting based on column partial name
- selecting columns based on regular expression
- re-ordering columns using `select()`
- renaming column names using `select()`