UNIVERSITY OF CALGARY

R Wizardry, Week 4: Data Management (continuation)

February 10th, 2016

Data managment

What makes R powerful?

The ability to apply functions to a variety of data objects (from scalars to data frames!)

The "Apply()" family:

1. apply(x, MARGIN, FUN, …)

"Returns a vector or array or list of values obtained by applying a function to margins of an array or matrix".

https://nsaunders.wordpress.com/2010/08/20/a-brief-introduction-to-apply-in-r/

Data managment

2. by(data, INDICES, FUN, ..., simplify = TRUE)

“Function ‘by’ is an object-oriented wrapper for ‘tapply’ applied to data frames.”

CRAN help files:

“INDICES: a factor or a list of factors, each of length nrow(data).”

Wrapper: The wrapper wraps, i.e. calls, another function that does the real work but provides a different or more convenient (or more convenient for a specific purpose) interface to it or specific syntax. Often the wrapper has the same arguments but different defaults and sometimes that is referred to as a "convenience" wrapper. Sometimes its used in the sense of providing an R interface to a C function with essentially similar arguments.

Data managment

3. eapply(env, FUN, ..., all.names = FALSE, USE.NAMES = TRUE)

“Applies FUN to the named values from an environment and returns the results as a list.”

Trickier, requires knowledge/experience in dealing with environments (self-contained object with its own variables and functions).

E.g.: a) Bioconductor (tools for molecular biology data analyses) uses environments, and b) functions (have their own environments)

Data managment

4. lapply(X, FUN, ...)

Apply a Function over a List or Vector.

“Returns a list of the same length as X, each element of which is the result of applying FUN to the corresponding element of X.”

5. sapply(X, FUN, ...)

“User-friendly version of lapply by default returning a vector or matrix if appropriate.”

Data managment

6. vapply(X, FUN, FUN.VALUE, ..., USE.NAMES = TRUE)

“Similar to sapply, but has a pre-specified type of return value, so it can be safer (and sometimes faster) to use.”

7. replicate(n, expr, simplify = "array")

A wrapper for the common use of sapply for repeated evaluation of an expression (usually to generate random numbers generation).”

Data managment

8. mapply(FUN, ..., MoreArgs = NULL, SIMPLIFY = TRUE,

```
     USE.NAMES = TRUE)
```

“Multivariate version of sapply. mapply applies FUN to the first elements of each (…) argument, the second elements, the third elements, and so on.”

[1,1] + [1,2] +[1,3] +...

[2,1] + [2,2] +[2,3]

9.tapply(X, INDEX, FUN = NULL, ..., simplify = TRUE)

“Apply a function to each cell of a ragged array, that is to each (non-empty) group of values given by a unique combination of the levels of certain factors.”

“Compounds” Dataset

Does water salinity (fresh, brackish or saline) affects the methanogenic microbial degradation of benzene, hexane and toluene?