

dplyr : Grammar of data manipulation

What is dplyr?

- ▶ dplyr is mainly authored by Hadley Wickham and Romain Francois. It is designed to be intuitive and easy to learn, thereby making “doing things” in R more user-friendly.
- ▶ dplyr is a new package which provides a set of tools for efficiently manipulating datasets in R.
- ▶ dplyr is the next iteration of plyr, focussing on only data frames.
- ▶ dplyr is faster, has a more consistent API and should be easier to use.

dplyr; Single Table Verbs

dplyr aims to provide a function for each basic verb of data manipulating:

- ▶ `filter()` (and `slice()`)
- ▶ `arrange()`
- ▶ `select()` (and `rename()`)
- ▶ `distinct()`
- ▶ `mutate()` (and `transmute()`)
- ▶ `summarise()`
- ▶ `sample_n()` and `sample_frac()`