Introduction to plyr

Hadley Wickham

May 3, 2008

plyr is a set of tools that solves a common set of problems: you need to break a big problem down into manageable pieces, operate on each pieces and then put all the pieces back together. This paper describes the components that make up plyr.

1 Introduction

This avoids any ambiguity about what you'll get back from one of these functions.
divide and conquer, Divide and marriage before conquest, divide et impera

2 Input

2.1 Data frames (d)

2.2 Arrays (a)

2.3 Lists (l)

mply

3 Process

error checking explode/splat

4 Output

4.1 Data frames (d)

4.2 Arrays (a)

4.3 Lists (l)

4.4 Ignored (_)

5 Labels

6 Equivalence to existing R functions

```
* aggregate(mtcars, list(mtcars$cyl), median)
  daply(mtcars, .(cyl), colwise(median))
  daply(mtcars, .(cyl), colwise(median, .if = is.numeric))

* p <- function(df) coef(lm(mpg ~ wt, data = df))
  do.call("rbind", lapply(split(mtcars, mtcars$cyl), p))
  ddply(mtcars, .(cyl), p)
```

The cast function in the reshape package is essentially a wrapper around aaply.

7 Function tools

each colwise failwith

8 Strategy

Take a small dataset, that you can easily solve.

8.1 Case study: baseball data