

06. 데이터 처리 성능 향상 # src: 7_R/07_데이터처리성능향상 참고**1) plyr 패키지;** 데이터의 분할, 함수 적용, 재조합 등에 사용하는 함수 포함하는 패키지

apply 계열 함수 대체 가능, 함수명은 데이터 구조의 종류에 따라 달라짐

#plyr패키지에 대한 설명: <https://cran.r-project.org/web/packages/plyr/plyr.pdf>**xyply(.data, ...)**

x: 입력데이터 타입 지정(a: array, l: list, d: data.frame, m: multiple inputs, r: repeat multiple times)

y: 출력데이터 타입 지정(a, l, d, m, r, _: nothing)

(1) ddply: `ddply(.data, .variables=(), .fun= , ...)`#ex `ddply(iris, .(Species), function(group){data.frame(mean(group$Sepal.Length))})`(2) adply: `adply(.data, .margins, .fun= , ...)`#ex `adply(iris[1:4], 2, function(col){sum(col)})`**2) reshape2 패키지;** 데이터 구조를 변경하기 위한 함수 제공, reshape보다 속도 빠름

(1) melt; 열 이름과 값을 variable, value 열에 저장된 형태로 변환하는 함수

`melt(data, ..., [na.rm=FALSE, value.name="value"])`#ex `melt(airquality, id=c('Month', 'Day'), na.rm=T)` #na.rm=T: 결측치(NA) 제외(F가 기본값)(2) cast; melt된 데이터를 되돌릴 때 사용: `dcast()`, `acast()`

① dcast의 결과 타입: 데이터 프레임(data.frame)

② acast의 결과 타입: 벡터(vector), 행렬(matrix), 배열(array)

3) 데이터 테이블(data.table) ; 짧고 유연한 구문 사용을 위해 데이터 프레임에서 상속 받음

데이터 프레임을 대신하여 사용할 수 있는 더 빠르고 편리한 데이터 타입

(1) 생성: `fread()`, `data.table()` 등