

data.table

data.table이란 무엇인가?

- data.table패키지는 2008년 [Matt Dowle](#)에 만들어짐.
- data.table을 data.frame의 고급 버전이라고 할 수 있음.
- data.frame을 상속하며, data.frame의 구문을 그대로 이용 가능함.
- data.frame패키지는 data.frame을 허용하는 다른 패키지와 호환됨.

- DT[i, j, by]

- DT는 data table의 레퍼런스

- i ⇔ where

- j ⇔ select

- by ⇔ group_by

data.table이란 무엇인가?

- `DT <- data.table(ID = 1:50,`
- `Capacity = sample(100:1000, size = 50, replace = F),`
- `Code = sample(LETTERS[1:4], 50, replace = T),`
- `State = rep(c("Alabama","Indiana","Texas","Nevada"), 50))`

- `DT[Code == "C", mean(Capacity), State]`

- `DT[Code == "D"]`
- `DT[, mean(Capacity), by = State]`
- `DT[Code == "A", mean(Capacity)]`

왜 data.frame 대신 data.table을 사용해야 하는가?

- 데이터 로딩할 때 빠른 속도를 제공함
 - `system.time(dt <- read.csv("data.csv"))`
 - `system.time(dt <- fread("data.csv"))`
- 데이터 조작에 사용되는 dplyr, plyr 패키지보다 훨씬 빠름
 - `system.time(dt %>% group_by(Store_ID) %>% filter(Gender == "F") %>%`
 - `system.time(dt[Gender == "F", .(sum(Transaction_Amount), mean(Var2)), by = Store_ID])`
- 파일을 읽는 것뿐만 아니라 파일을 저장할 때도 빠름
- 자동 인덱싱, 롤링 조인, 오버래핑 레인지 조인을 빠르게 처리함