R Clinic

2013/05/23

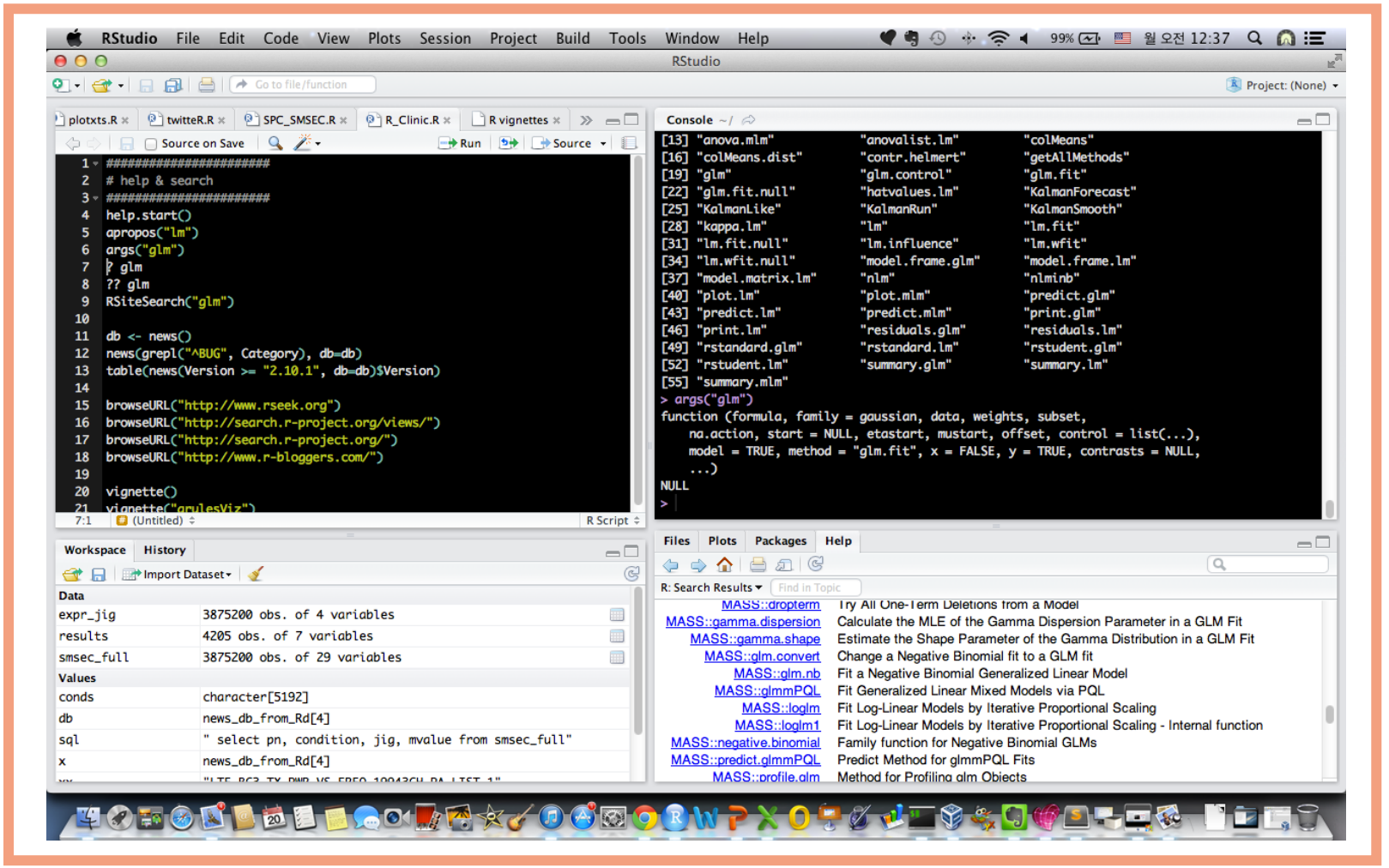
유충현 - KRUG 대표

Agenda

1. 유용한 도구
2. R Basic
3. performance tuning
4. tips

1. 유용한 도구

RStudio를 이용한 분석 환경



Help & Search

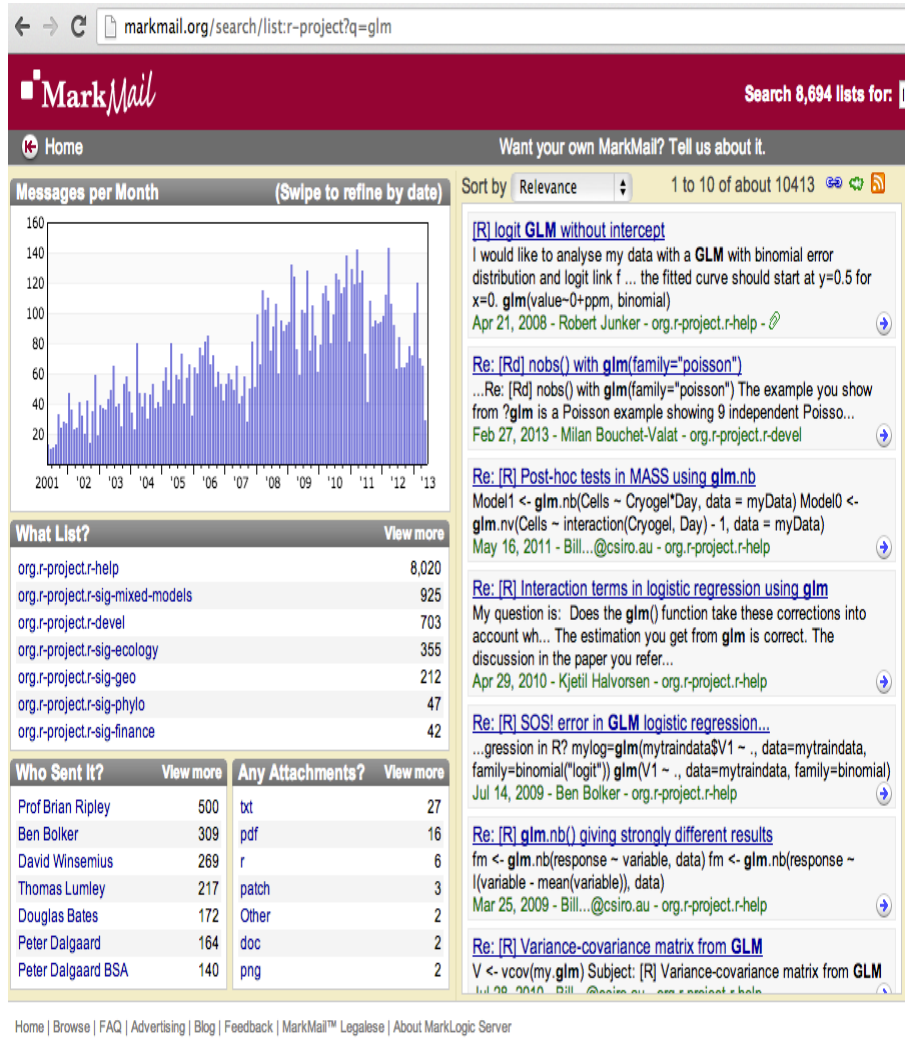
R 학습의 첫 걸음은 Help와 Search로부터

```
#########################
# help & search
#########################
help.start()
apropos("lm")
args("glm")
? glm
?? glm
RSiteSearch("glm")

db <- news()
news(grepl("^BUG", Category), db=db)
table(news(Version >= "2.10.1", db=db)$Version)

browseURL("http://www.rseek.org")
browseURL("http://search.r-project.org/views/")
browseURL("http://search.r-project.org/")
browseURL("http://www.r-bloggers.com/")

vignette()
vignette("arulesViz")
```



for Data Manipulation

Perform SQL Selects on R Data Frames

Documentation for package 'sqldf' version 0.4-6.4

- DESCRIPTION file.
- Code demos. Use demo() to run them.
- Package NEWS.

Help Pages

sqldf-package	sqldf package overview
read.csv.sql	Read File Filtered by SQL
read.csv2.sql	Read File Filtered by SQL
sqldf	SQL select on data frames

data.frame

CSV file

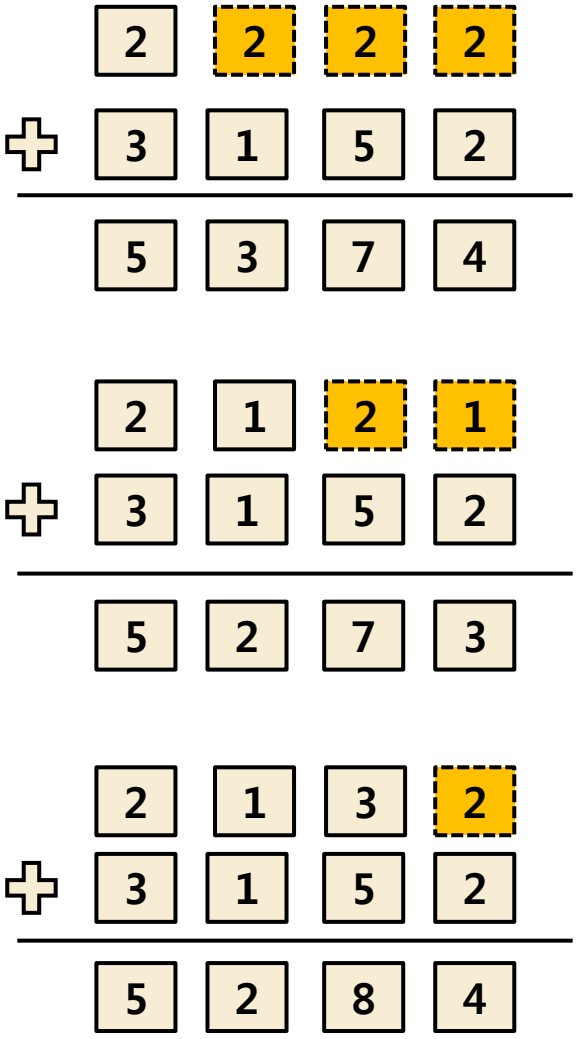
Manipulation

Easy Fast

2. R Basic

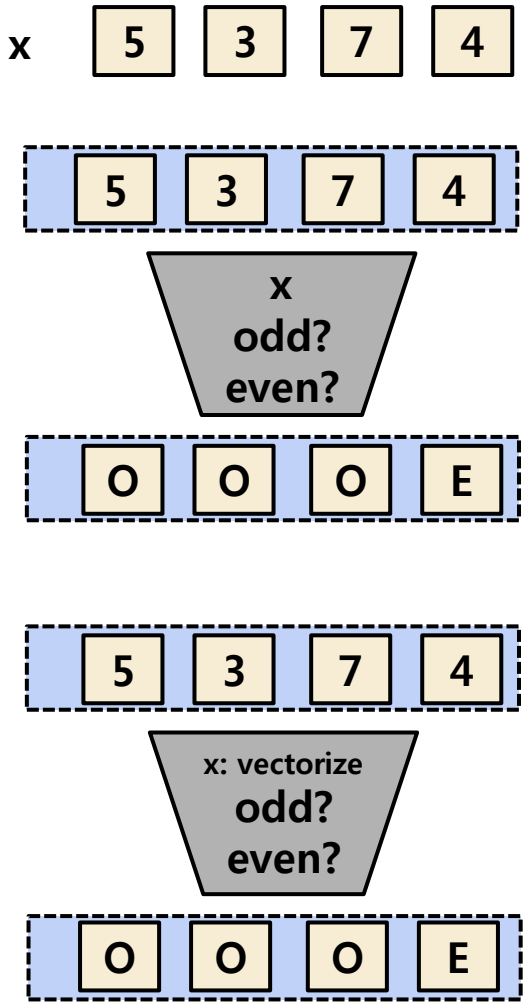
for same length

```
60 ##########################
61 # recycleing rule
62 ##########################
63 (x <- 1)
64 (y <- 2:5)
65 x + y
66
67 (x <- 1:2)
68 (y <- 2:5)
69 x + y
70
71 (x <- 1:2)
72 (y <- 2:4)
73 x + y
74
75 (x <- rnorm(5))
76 plot(x, col="red", pch=16)
77 plot(x, col=1:5, pch=15:19)
78 plot(x, col=1:2, pch=15:16)
```



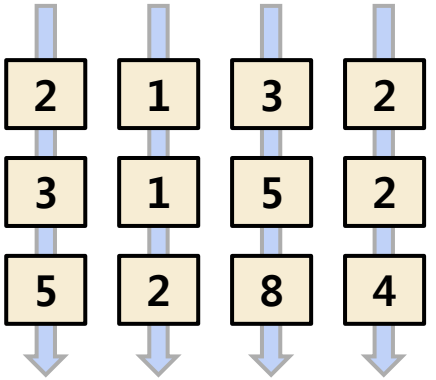
not looping

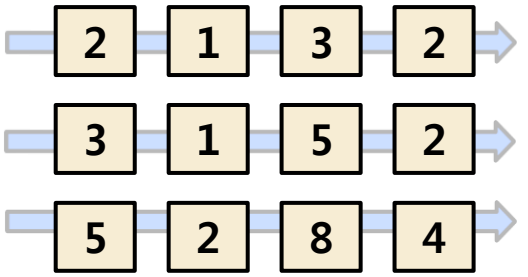
```
86  ########################
87  # vectorization
88  ########################
89  1:5 + 6:10
90
91  x <- 1:10
92  even <- logical(length(x))
93
94  for (i in x) {
95    if (i %% 2 == 0) {
96      even[i] <- TRUE
97    } else {
98      even[i] <- FALSE
99    }
100 }
101 even
102
103 (even_vectorization <- x %% 2 == 0)
104
105 if (x %% 2 == 0) "EVEN" else "ODD"
106
107 ifelse(x %% 2 == 0, "EVEN", "ODD")
```



apply 함수군의 이용

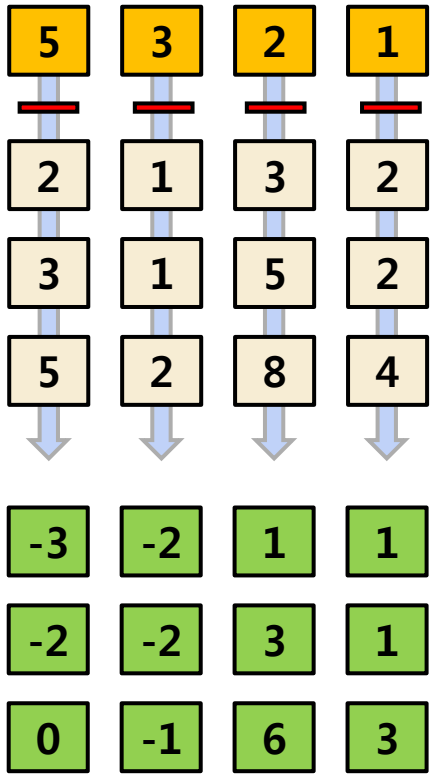
```
110 #-------------------------------
111 # apply function
112 #-------------------------------
113 (mat <- matrix(1:12, ncol=4, byrow=T))
114
115 apply(mat, 1, sum)
116 apply(mat, 2, mean)
117 (tmp <- apply(mat, 2, FUN=function(x, value) grep(value, x), 7))
118 do.call("rbind", tmp)
119
120 colMeans(mat)
121 rowSums(mat)
122
123 (x <- list(a = 1:10, beta = exp(-3:3), logic = c(TRUE,FALSE,FALSE,TRUE)))
124 (tmp <- lapply(x, mean))
125 do.call("cbind", tmp)
126
127 #pc <- by(pn_condition[, "item"], pn_condition[, "user"],
128 #         function(x) as.character(x))
129 #pc <- lapply(pc, c)
130 #pc <- as(pc, "transactions")
131
132 # sapply
133 # tapply
134 # mapply
```





sweep fuction

```
137 #------------------------------
138 # sweep function
139 #------------------------------
140 (med.att <- apply(attitude, 2, median))
141 sweep(data.matrix(attitude), 2, med.att)
142
143 set.seed(1)
144 (mat <- matrix(sample(12), nrow=3))
145
146 # propotion
147 idx <- 2
148 total <- apply(mat, idx, sum)
149 sweep(mat, idx, total, FUN="/")
150
151 prop.table(mat, idx)
152
153 idx <- 1
154 total <- apply(mat, idx, sum)
155 sweep(mat, idx, total, FUN="/")
156
157 prop.table(mat, idx)
158
```



3. performance tuning

performance tuning tips

```
235 ##########################################
236 # performance tuning
237 ##########################################
238
239 #------------------------------------------
240 # looping  사용 않기 (for, while, repeat)
241 # ---> vectorize, apply
242 #------------------------------------------
243
244 vec <- sample(10000000)
245 over.thresh <- function(x, threshold)
246 {
247   for (i in 1:length(x))
248     if (x[i] < threshold)
249       x[i] <- 0
250   x
251 }
252
253 system.time(tmp <- over.thresh(vec, 100))
254
255 over.thresh2 <- function(vec, threshold)
256 {
257   ifelse(x < threshold, 0, x)
258 }
259
260 system.time(tmp <- over.thresh2(vec, 100))
```

- looping 사용하지 않기
- dataset을 키우지 않기
- 계산 결과의 재사용
- 재귀호출 사용하지 않기

performance tuning tips

```
350 ##########################################
351 # performance tuning
352 ##########################################
353
354 #------------------------------------------
355 # multicore 사용하기
356 #------------------------------------------
357 library(foreach)
358 library(doMC)
359 registerDoMC(cores=2)
360
361 set.seed(1)
362 m <- matrix(rnorm(9000000), 3000, 3000)
363
364 system.time(result <- foreach(i=1:nrow(m), .combine=rbind) %dopar%
365   (m[i,] / mean(m[i,])))
366
367
368 result <- matrix(0, ncol=3000, nrow=3000)
369 system.time(for(i in 1:nrow(m))
370   result[i, ] <- (m[i,] / mean(m[i,])))
371
372
373 library(plyr)
374 (dfx <- data.frame(
375   group = c(rep('A', 8), rep('B', 15), rep('C', 6)),
376   sex = sample(c("M", "F"), size = 29, replace = TRUE),
377   age = runif(n = 29, min = 18, max = 54)
```

- multi-core 사용하기
  - foreach
  - doMC
  - plyr

memory 절약

```
#-----------------------------------------
# Memory 절약
#-----------------------------------------

gets <- function(n=500000) {
  tmp <- runif(n)
  tmp1 <- 2 * tmp
  tmp2 <- trunc(tmp1)
  mean(tmp2 > 0.05)
}

gets2 <- function(n=500000) {
  tmp <- runif(n)
  tmp <- 2 * tmp
  tmp <- trunc(tmp)
  mean(tmp > 0.05)
}

gets(10000000)
gets2(10000000)
```

- Same name
- Same Size

4. tips

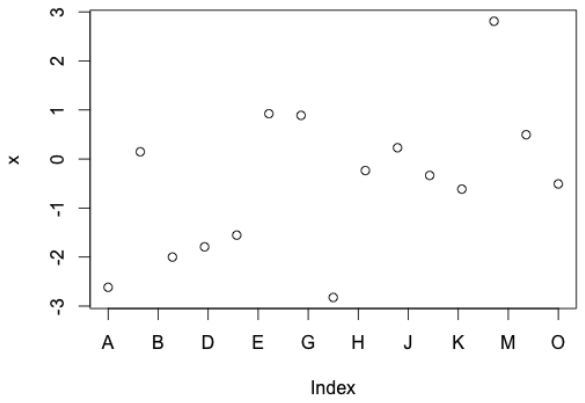
기타 유용한 팁

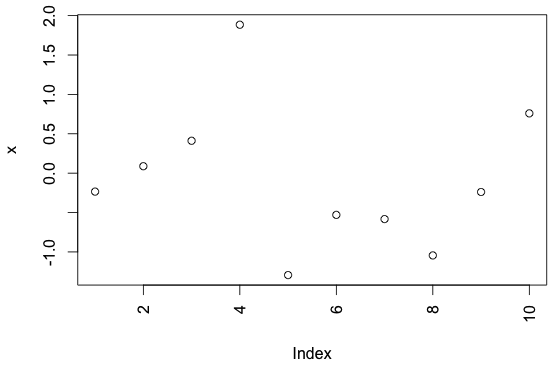
```
161 ################################
162 # tips
163 ################################
164
165 #--------------------------------------
166 # interactive object create & reference
167 #--------------------------------------
168
169 (idx <- 1:5)
170 (obj.names <- paste("var", idx, sep="."))
171
172 for (i in idx) assign(obj.names[i], 1:i)
173
174 ls(pat="^var\\.")
175
176 var.1
177 var.2
178 var.3
179 var.4
180 var.5
181
182 for (i in idx) print(get(obj.names[i]))
```

- 동적 이름 할당한 객체 생성
- if~else 주의 사항
- 논리연산자
- on.exit

유용한 graphics 팁

```
165 #---------------------------------------
166 # interactive object create & reference
167 #---------------------------------------
168
169 (idx <- 1:5)
170 (obj.names <- paste("var", idx, sep="."))
171
172 for (i in idx) assign(obj.names[i], 1:i)
173
174 ls(pat="^var\\.")
175
176 var.1
177 var.2
178 var.3
179 var.4
180 var.5
181
182 for (i in idx) print(get(obj.names[i]))
183
```





감사합니다.

bdboy@r-project.kr
http://r-project.kr