

# 工作记录10之R数据子集提取

2016-09-14 17:18:39

用R处理数据时，常常并不需要访问整个数据集，而是选取数据中的的一部分。在处理数据时遇到缺失的值，这些缺失的值需要去掉，这就需要使用一个索引去找出需要的数据或者剔除缺失值，学会以下的方法可以对大部分的数据进行子集处理。

1. 单层方括号索引 单层方括号索引包含数值型索引和逻辑型索引,使用符号[]来对数据进行子集处理。单层方括号索引返回的值的类型和源对象的类型相同。例如：  
`> x <- c(1,2,3,4,5) > x[2] [1] 2 > y <- list(a="aaa", b="foo", c=1 : 4) > y[2] > mode(x[2]) [1]"numeric"`  
`> mode(y[3]) [1]"list"` 这里分别给x,y赋值一个向量和一个列表，返回的对象和源对象的数据类型是相同的，单层方括号[]中的单个数字表示数据的第几个元素。

1.1 数值型索引  
`> m <- x[2 : 4] > n <- x[c(1, 5)] > m [1] 2 3 4 > n [1] 1 5 > y`  
`> mymatrix <- matrix(1 : 6, nrow = 2, ncol = 3) > mymatrix [ , 1][ , 2][ , 3] [1, ] 1 3 5 [2, ] 2 4 6`  
`> fe <- mymatrix[2, ] > fe [1] 2 4 6` 用单层方括号来提取向量，列表，因子和矩阵的数据子集，也可以用-号加括号内的值来剔除你不需要的元素，-2表示不需要第二个 元素，放在矩阵中就是不需要第二行。矩阵的数据按行或者按列可以看成是数字索引，如 `mymatrix[1,]`表示第一行的所有列

1.2 逻辑型索引  
`> data <- c(2, 4, 6, 9, 4, 8, 19) > data[data > 6] [1] 9 8 19 > data > 6 [1] FALSE FALSE FALSE TRUE FALSE TRUE TRUE`

2. 双层方括号索引 双层方括号被用来列表或者数据框中提取元素,但返回的对象不一定是列表或数据框。  
`> num = 1 : 4 > names = c("lly", "lucy", "ziggs", "ben") > sex = c("F", "F", "M", "M")`  
`> score = c(75, 89, 90, 68) > stu = data.frame(number = num, name = names, sex = sex, score = score, ) > stu`  

number	name	sex	score
1	lly	F	75
2	lucy	F	89
3	ziggs	M	90
4	ben	M	68

  
`> stu[["sex"]] [1] F F M M Levels: F M > stu[[1]] [1] 1 2 3 4` 使用双层方括号索引提取一个元素

3. \$符号索引 美元\$索引可以从有命名的列表或者数据框中提取元素，在一定程度上美元符号和双层中括号的用途一样。  
`> stu$name [1] lly lucy ziggs ben Levels: ben lly lucy zigg`  
`> stu$name[score > 80] [1] lucy ziggs Levels: ben lly lucy ziggs > stu[["name"]][score > 80] [1] lucy ziggs`  
Levels: ben lly lucy ziggs 提取了score在80分以上的学生，使用attach函数可以在处理多变量数据框时频繁使用\$索引

4. NA值的移除  
`> x <- c(1, 2, NA, NA, 5) > m <- is.na(x) > m [1] FALSE FALSE TRUE TRUE FALSE > x[!m] [1] 1 2 5` 用于得到x向量中非NA值 函数complete.cases()的用法：  
`> data(airquality) > airquality[1 : 6, ]`  

Ozone	Solar.R	Wind	Temp	Month	Day
41	190	7.4	67	5	1
2	36	118	8.0	72	5
2	3	12	149	12.6	74
5	3	4	18	313	11.5
62	5	4	5	NA	NA
14.3	56	5	5		

  
`> good <- complete.cases(airquality) > airquality[good, ][1 : 6, ]`  

Ozone	Solar.R	Wind	Temp	Month	Day
41	190	7.4	67	5	1
2	36	118	8.0	72	5
2	3	12	149	12.6	74
5	3	4	18	313	11.5
62	5	4	7	23	299
8.6	65	5	7		

  
`> airquality[good, ][1 : 6, ]`  

Ozone	Solar.R	Wind	Temp	Month	Day
41	190	7.4	67	5	1
2	36	118	8.0	72	5
2	3	12	149	12.6	74
5	3	4	18	313	11.5
62	5	4	7	23	299
8.6	65	5	7		