

unit 1 detailed for data structure

2023-08-26

R data structure

```
#specific for data structure

#基本类型: 继承c语言类型, 数值 (numeric)、字符串 (character)、逻辑型 (logical)
##numeric: integer double (浮点) 双精度
a<-1
a_integer<-1
is.integer(a_integer) #jugde the data type
```

[1] FALSE

```
as.integer(a) #transfer to integer
```

[1] 1

```
is.double(a)
```

[1] TRUE

```
#character:标识 " "
is.character(a)
```

[1] FALSE

```
is.numeric(a)
```

[1] TRUE

```
as.numeric(a) #长得像数据类型的字符串转换为数值型
```

[1] 1

```
paste("hello_world","123",sep="_") #指定间隔符
```

[1] "hello_world_123"

```
#result:"hello_world_123"
substr("hello_world",start=2,stop=5) #字符串取值
```

[1] "ello"

```
#result: "ello"

#logical Boolean 逻辑型 内置变量
vari<-TRUE
is.logical(vari)
```

[1] TRUE

```
### multi-dimentional data structure
#vector, matrix, array, dataframe(行列有含义, 行为样本, 列为属性), list (mixed types)

#vector
# c() 向量
x<-c(2,3,7,1,8,9,20)

#vector 取值
length(x)
```

[1] 7

```
x[2]
```

[1] 3

```
#vector替换 取多少给多少 善用length
x[3]<-87
x[c(1,4)]<-c(0,9)

#use boolean
x[c(TRUE,FALSE,TRUE,FALSE)]<-c(5,8)

#多个向量组成一个矩阵 matrix
rbind(x)
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7]
## x      5      3      8      9      5      9      8
```

```
cbind(x) #从行或者列维度合并向量
```

```
##      x
## [1,] 5
## [2,] 3
## [3,] 8
## [4,] 9
## [5,] 5
## [6,] 9
## [7,] 8
```

```
#matrix operation 需要利用一维数据生成, 替换: 拿多少还多少
matrix1<-matrix(c(1:50),nrow=5)
matrix2<-matrix(c(1:50),nrow=5,byrow=TRUE) # vector排列方式变化

#计算
matrix2*2
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
## [1,]      2      4      6      8     10     12     14     16     18     20
## [2,]     22     24     26     28     30     32     34     36     38     40
## [3,]     42     44     46     48     50     52     54     56     58     60
## [4,]     62     64     66     68     70     72     74     76     78     80
## [5,]     82     84     86     88     90     92     94     96     98    100
```

```
matrix1*matrix2
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
## [1,]      1      12     33     64    105    156    217    288    369    460
## [2,]     22     84    156    238    330    432    544    666    798    940
## [3,]     63    176    299    432    575    728    891   1064   1247   1440
## [4,]    124    288    462    646    840   1044   1258   1482   1716   1960
## [5,]    205    420    645    880   1125   1380   1645   1920   2205   2500
```

```
sum(matrix1)
```

[1] 1275

```
apply(matrix2,1,sum) #apply() 对行列进行操作
```

[1] 55 155 255 355 455

```
#dataframe 构造函数 data.frame 接受等长的向量, 参数的名字将会转换为列名
name<-c("bob","alice","sam","fred")
score<-c(34,72,67,98)
class<-c(1,2,2,1)
student_score<-data.frame(name=name,
                           score=score,
                           class=class)

#与matrix的区别: m需要数据类型全都一致, d需要每一列的数据类型一致
#data.frame取值 利用行列名
rownames(student_score)<-student_score[,1]
student_score_ba<-student_score[c("bob","alice"),]
student_name<-student_score$name #取列

#which: bealoon to numeric
which(c(TRUE,FALSE,TRUE,FALSE)) #直接取值或镶嵌在向量及dataframe中的数据
```

[1] 1 3

```
test_frame<-student_score[c(1,3,4),]
#等价于
test_frame2<-student_score[which(c(TRUE,FALSE,TRUE,TRUE)),]

#in% 判断character

#行列名的替换 rownames/colnames 实质是向量的取值和替换问题
rownames(student_score)<-c('a','b','c','d')

#list as special vector
list1<-list(1,2,3,4,5)
list1[1]
```

[[1]]
[1] 1

```
list1[c(2,4)]
```

[[1]]
[1] 2
##
[[2]]
[1] 4

```
list2<-list(c(1:3),
            c(2,2),
            "hello",
            matrix(c(1:100),nrow=10),
            data.frame(name=c(1:100),height=c(1:100)),
            list(1,23,6,8,90))

length(list2)
```

[1] 6

```
t.test_result=t.test(c(1:20),c(10:15))

is.list(t.test_result)
```

[1] TRUE

```
names(list2)<-c("vec1","vec2","vec3","vec4","vec5","vec6")
t.test_result$value
```

[1] 0.2029565

reference from bilibiliupshengxinbanmayu

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.