

R language basics, part 1

HUST Bioinformatics course series

Wei-Hua Chen (CC BY-NC 4.0)

07 September, 2022

section 1: outline

TOC

- ① 数据类型
- ② 基本操作
- ③ 变量
- ④ 函数/子例程 (Functions)
- ⑤ 模型和公式
- ⑥ R 包
- ⑦ 获得帮助

注：不一定按照上述顺序

section 2: R basics - 基本数据类型

基本数据类型

最基本的数据类型包括**数字**和**字符串**，是其它数据类型的基本组成部分。

数字

```
## 整数  
287
```

```
## [1] 287
```

```
## 小数  
99.99
```

```
## [1] 99.99
```

```
## 科学计数法  
1e-3
```

```
## [1] 0.001
```

```
1e2
```

```
## [1] 100
```

逻辑符号

真

```
TRUE  
T
```

假

```
FALSE  
F
```

其本质是数字

```
1 + TRUE
```

```
## [1] 2
```

```
2 * FALSE
```

```
## [1] 0
```

字符串

字符串则是可以是任何字符的组合，由单引号或双引号包括。比如：

```
'a sentence' ## 单括号  
" 一个字符串" ## 双括号  
'1.123'      ## 像是数字的字符串  
'*%*(!)@##&@(9' ## 乱码
```

简单数据类型

简单数据类型包括 `vector` 和矩阵，它们都可以包含某一种基本数据类型的多个数值，比如由多个数字组成的矩阵，多个字符串组成的 `vector` 等。但它们只能包含单一数据类型；这一点稍后会有解释。

```
c(100, 20, 30) ## 整数 vector
c("字符串", "数组", "是我") ## 字符串 vector
c(TRUE, FALSE, TRUE, T, F) ## 一个逻辑 vector
```

如上所示，数组通常用函数 `c()` 来定义。除此之外，还可以用 `:` 操作符号来定义包含连续整数的 `vector`：

```
2:8
```

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


vector 的数据类型转换规则

vector 只能包含一种基本数据类型。因此，在定义数组时，如果输入的数值是混合的，那么某些基本数据类型会自动转换为其它类型，以保证数值类型的一致性；这在英文里称为 `coerce`，有强制转换的意思。这种转换的优先级为：

- 逻辑类型 -> 数字类型
- 逻辑类型 -> 字符串
- 数字类型 -> 字符串

我们可以用 `class()` 或 `str()` 函数来判断 vector 包含的数据类型。以后会介绍两者的不同。

vector 的数据类型转换规则

```
class( c(45, TRUE, 20, FALSE, -100) ); ## 逻辑和数字类型
```

```
## [1] "numeric"
```

```
str( c("string a", FALSE, "string b", TRUE) ); ## 逻辑和字符
```

```
## chr [1:4] "string a" "FALSE" "string b" "TRUE"
```

```
str( c("a string", 1.2, "another string", 1e-3) ); ## 数字和字符
```

```
## chr [1:4] "a string" "1.2" "another string" "0.001"
```

矩阵 (matrix)

矩阵也可看做是一种带有限制的二维数组，其限制是，矩阵内的数值必须是同一种类型。当输入混合有多种基本数据类型时，矩阵会按上面提到的规则进行强制转换。

矩阵由函数 `matrix()` 定义，比如：

```
matrix( c(20, 30.1, 2, 45.8, 23, 14), nrow = 2, byrow = T );
```

```
##      [,1] [,2] [,3]  
## [1,] 20.0 30.1    2  
## [2,] 45.8 23.0   14
```