数据集的行称为“观测”，列称为“变量”。

行/实例标识符。

注释

从#到行尾。

菜单

|  |
| --- |
| 编辑>>清空控制台 |
| 其他>>删除所有对象 |

函数

1、杂项

|  |  |  |
| --- | --- | --- |
| help(**函数名**)  函数名可以加引号。  ?**函数名**  函数名可以加引号。  help(package=**包名**)  包名可以加引号。  查看函数的帮助。 | |  |
| Get Working Directory  getwd()  显示当前的工作目录。 | Set Working Directory  setwd(**字符串**)  修改当前的工作目录。 |  |
| history(**整数**=25)  显示最近用过的命令。 | |  |
| List  ls()  列出当前工作空间中的对象。 | |  |
| Remove  rm(**对象名**,**...**)  对象名可以加引号。  删除当前工作空间中的对象。 | | 1、rm(list=ls())删除当前工作空间中的所有对象。 |
| library()  查看库中的包。 | install.packages("**包名**")  下载和安装包。 |  |
| installed.packages()  查看库中的包的详细信息。 |  |
| search()  查看已经被加载的包。 | library((package=)?**包名**)  包名可以加引号。  载入包。 |  |
| require((package=)?**包名**)  包名可以加引号。  载入包。 |  |
| q()  退出R。 | |  |

2、向量、类别、有序类别

元素下标从1开始。

|  |  |
| --- | --- |
| **向量**[(**正整数向量**|**负整数向量**|**元素名向量**|**布尔向量**)]  例：**向量**[-**正整数向量**] | ！！！向量的元素可以重复。  1、若正整数大于元素个数，产生NA；若布尔向量小于元素个数，循环使用；若布尔向量大于元素个数，产生NA；若没有以元素名为名称的元素，产生NA。 |

|  |  |
| --- | --- |
| length(**向量**) | |
| Replicate  rep(**向量**,**重复次数**) | unique(**向量**)  得到去重后的向量。 |
| Replicate  rep(**因子向量**,**重复次数**) | unique(**因子向量**)  得到去重后的因子向量。 |
| Replicate  rep(**有序因子向量**,**重复次数**) | unique(**有序因子向量**)  得到去重后的有序因子向量。 |

3、向量

！！！没有标量。标量用只含1个元素的向量模拟。

长度不同的2个向量运算，短的向量循环和长的向量运算，得到的向量的长度是长的向量的长度。

|  |  |  |  |
| --- | --- | --- | --- |
| Combine  c((**元素名**=)?(**向量**|**因子向量**|**有序因子向量**)(  ,(**元素名**=)?(**向量**|**因子向量**|**有序因子向量**))\*)  得到所有向量连接起来的向量。  1、若不指定元素名，则没有元素名。  2、若1个元素名对应多个元素，则从第1个开始添加后缀**从1开始的整数**。 | | Concatenate  cat(**字符串向量**(  ,**字符串向量**)?  ,sep=**字符串向量**)  拼接字符串。  1、若sep元素个数小于字符串个数-1，循环使用。 | sum(**向量**(  ,**向量**)?) |
| **实数**:**实数**  步长为。  包括终值。 | Sequence  seq(from,to)  步长默认为。  包括to。  seq(from,to,by)  包括to。  seq(from,to,length.out)  包括to。 | | |

4、类别（名义型变量）

|  |  |
| --- | --- |
| factor(**向量**)  所有可能的类别值由向量确定。  factor((**向量**,)?levels=**类别值向量**)  不在取值范围内的值变成<NA>。 |  |

5、有序类别（有序型变量）

|  |  |
| --- | --- |
| factor(**向量**,ordered=TRUE)  所有可能的类别值由向量按字母顺序确定。  factor((**向量**,)?levels=**类别值向量**,ordered=TRUE)  不在取值范围内的值变成<NA>。 |  |

6、矩阵

|  |  |  |
| --- | --- | --- |
| + | 同型矩阵 |  |
| - | 同型矩阵 |  |
| \* | 同型矩阵 |  |
| / | 同型矩阵 |  |
| %/% | 同型矩阵 | 除以，向下取整。 |
| %% | 同型矩阵 | 求余数，符号和除数相同。 |
| %\*% | 左矩阵列数等于右矩阵行数 |  |

|  |
| --- |
| matrix(**向量**,nrow,ncol,byrow=FALSE,dimnames=list(NULL,NULL))  nrow和ncol还可以1个写1个不写。  写nrow时，nrow的长度整除向量的长度或向量的长度整除nrow的长度；不写nrow时，ncol的长度整除向量的长度或向量的长度整除ncol的长度。  list第1个元素是**行名向量**，第2个元素是**列名向量**。 |
| Transpose  t(**矩阵**)  得到转置后的矩阵。 |
| solve(**矩阵**)  得到逆矩阵。 |
| diag(**矩阵**)  得到对角线上的元素组成的向量。  diag(**向量**,nrow,ncol)  不写nrow、ncol时，得到以向量为对角线的对角矩阵；否则，向量长度小于min{nrol,ncol}时，循环使用向量。 |

6、数组

|  |  |  |
| --- | --- | --- |
| array(**向量**,**各维长度向量**,list=(NULL,...))  list的第i个元素是第i维每个分量（分量个数就是该维的长度）的名字向量。 | 1、填充时，第1维的下标变化最快，第n维的下标变化最慢。例：   |  | | --- | | > array(1:24,c(2,3,4),list(NULL,c("a","b","c")))  , , 1  a b c  [1,] 1 3 5  [2,] 2 4 6  , , 2  a b c  [1,] 7 9 11  [2,] 8 10 12  , , 3  a b c  [1,] 13 15 17  [2,] 14 16 18  , , 4  a b c  [1,] 19 21 23  [2,] 20 22 24 | |

7、数据框（数据帧）

|  |  |
| --- | --- |
| **数据框**[(**正整数向量**|**负整数向量**|**列名向量**|**布尔向量**)?]  **数据框**$**列名**  得到数据框 | ！！！中括号内的向量可以重复。  1、不写会选入所有列。  2、正整数不能大于列数，布尔向量元素个数不能大于列数。  3、将NULL赋给数据框的若干列即可删除这些列。 |
| **数据框**[(**正整数向量**|**负整数向量**|**行名向量**|**布尔向量**)?  ,(**正整数向量**|**负整数向量**|**列名向量**|**布尔向量**)?]  得到数据框或（有1维长度是1时）向量 | ！！！选择行的向量的元素可以重复。  1、不写会选入所有行。  2、若正整数大于元素个数，产生NA；若布尔向量小于元素个数，循环使用；若布尔向量大于元素个数，产生NA；若没有以行名为名称的行，产生NA。  ！！！选择列的向量的元素可以重复。  3、不写会选入所有列。  4、正整数不能大于列数，布尔向量元素个数不能大于列数，列名必须有以列名为名称的列。  5、若布尔向量小于元素个数，循环使用。 |

|  |  |
| --- | --- |
| data.frame((**列名**=)?(**向量**|**因子向量**|**有序因子向量**),...) | 1、若不指定列名，则根据表达式生成列名（例：变量名）。  2、若有重复的列名，则从第2个开始添加后缀.**从1开始的整数**。 |
| str(**数据框**)  得到数据框每列结构。 | 1、例：   |  | | --- | | > a1<-c(1,1,1,1,2,2,2,3,3,4)  > a2<-factor(a1)  > a3<-factor(a1,ordered=T)  > a4<-data.frame(a1,a2,a3)  > str(a4)  'data.frame': 10 obs. of 3 variables:  $ a1: num 1 1 1 1 2 2 2 3 3 4  $ a2: Factor w/ 4 levels "1","2","3","4": 1 1 1 1 2 2 2 3 3 4  $ a3: Ord.factor w/ 4 levels "1"<"2"<"3"<"4": 1 1 1 1 2 2 2 3 3 4  > summary(a4)  a1 a2 a3  Min. :1.00 1:4 1:4  1st Qu.:1.00 2:3 2:3  Median :2.00 3:2 3:2  Mean :2.00 4:1 4:1  3rd Qu.:2.75  Max. :4.00 | |
| summary(**数据框**)  得到数据框每列常用的统计量。 |

8、列表

|  |  |
| --- | --- |
| **列表**[**正整数向量**]  **列表**[**对象名向量**] | ！！！中括号内的向量可以重复。  1、将NULL赋给数据框的若干对象即可删除这些对象。 |
| **列表**[[**正整数**]]  **列表**[[**对象名**]]  **列表**$**对象名** | 1、将NULL赋给数据框的1个对象即可删除这个对象。 |

长度不同的列表和向量进行关系运算（<、>、==、<=、>=、!=），短的循环和长的运算，得到的向量的长度是长的的长度。

|  |  |
| --- | --- |
| list((**对象名**=)?**对象**,...) | 1、若不指定对象名，则没有对象名。 |
| unlist(**列表**)  得到列表中每个对象的每个元素组成的向量。 |  |