

Base R

备忘单

获取帮助

访问帮助文件

`?mean`

获得特定的帮助。

`help.search('weighted mean')`

在帮助文件中搜索单词或短语。

`help(package = 'dplyr')`

查找包的帮助。

有关对象的更多信息

`str(iris)`

有关对象的详细信息。

`class(iris)`

查找对象所属的类。

使用包

`install.packages('dplyr')`

从CRAN下载并安装软件包。

`library(dplyr)`

将包加载到会话中，使其所有函数都可用。

`dplyr::select`

使用包中的特定函数。

`data(iris)`

将内置数据集加载到环境中。

工作目录

`getwd()`

找到当前工作目录(查找输入和发送输出的位置)。

`setwd('C://file/path')`

更改当前工作目录。

使用RStudio中的项目将工作目录设置为正在工作的文件夹。

向量

创建向量

<code>c(2, 4, 6)</code>	2 4 6	将元素连接成一个向量
<code>2:6</code>	2 3 4 5 6	整数序列
<code>seq(2, 3, by=0.5)</code>	2.0 2.5 3.0	一个复杂的序列
<code>rep(1:2, times=3)</code>	1 2 1 2 1 2	对向量进行重复
<code>rep(1:2, each=3)</code>	1 1 1 2 2 2	对向量中的元素进行重复

关于向量的函数

`sort(x)`

返回已排序过的x。

`table(x)`

查看值的计数。

`rev(x)`

返回反转后的x。

`unique(x)`

查看所有唯一的值。

选择向量中的元素

按位置

<code>x[4]</code>	第4个元素
<code>x[-4]</code>	除了第4个外的所有元素
<code>x[2:4]</code>	第2个到第4个元素
<code>x[-(2:4)]</code>	除了第2个到第4个之外的所有元素
<code>x[c(1, 5)]</code>	第1个和第5个元素

按值

<code>x[x == 10]</code>	等于10的元素
<code>x[x < 0]</code>	所有小于零的元素
<code>x[x %in% c(1, 2, 5)]</code>	包含于数据集1, 2, 5中的元素

按名字

`x['apple']` 名称为 'Apple的' 元素

编程

For循环

```
for (variable in sequence){
  Do something
}
```

示例

```
for (i in 1:4){
  j <- i + 10
  print(j)
}
```

If语句

```
if (condition){
  Do something
} else {
  Do something different
}
```

示例

```
if (i > 3){
  print('Yes')
} else {
  print('No')
}
```

While循环

```
while (condition){
  Do something
}
```

示例

```
while (i < 5){
  print(i)
  i <- i + 1
}
```

函数

```
function_name <- function(var){
  Do something
  return(new_variable)
}
```

示例

```
square <- function(x){
  squared <- x*x
  return(squared)
}
```

数据的读取与写入

另请参阅 `readr` 包

输入	输出	描述
<code>df <- read.table('file.txt')</code>	<code>write.table(df, 'file.txt')</code>	读写带分隔符的文本文件。
<code>df <- read.csv('file.csv')</code>	<code>write.csv(df, 'file.csv')</code>	读写逗号分隔值文件。这是read.table/write.table的特例。
<code>load('file.RData')</code>	<code>save(df, file = 'file.Rdata')</code>	读写R数据文件，这是R专用的文件类型。

条件判断

<code>a == b</code>	相等的	<code>a > b</code>	大于	<code>a >= b</code>	大于等于	<code>is.na(a)</code>	是缺失?
<code>a != b</code>	不等的	<code>a < b</code>	小于	<code>a <= b</code>	小于等于	<code>is.null(a)</code>	是未初始化?

类型

R 中常见数据类型之间的转换。总是可以从表中的较高值转换为较低值。

as.logical	TRUE, FALSE, TRUE	布尔值 (TRUE 或 FALSE)。
as.numeric	1, 0, 1	整数或浮点数。
as.character	'1', '0', '1'	字符串。一般优先于因素。
as.factor	'1', '0', '1', levels: '1', '0'	具有预设级别的字符串。某些统计模型需要。

数学函数

log(x)	取自然对数	sum(x)	求和
exp(x)	取自然指数	mean(x)	求平均
max(x)	取最大的元素	median(x)	取中位数
min(x)	取最小的元素	quantile(x)	取百分位数
round(x, n)	取n位小数	rank(x)	计算秩
signif(x, n)	取n位有效数字	var(x)	计算方差
cor(x, y)	计算相关性系数	sd(x)	计算标准差

变量赋值

```
> a <- 'apple'
> a
[1] 'apple'
```




环境

ls()	列出环境中的所有变量。
rm(x)	从环境中移除 x。
rm(list = ls())	从环境中移除所有变量。

您可以使用 RStudio 中的环境面板来浏览环境中的变量。

矩阵

```
m <- matrix(x, nrow = 3, ncol = 3)
通过x创建一个矩阵。
```

 <code>m[2,]</code> - 选择1横行	<code>t(m)</code> 转置
 <code>m[, 1]</code> - 选择1竖列	<code>m %% n</code> 矩阵乘法
 <code>m[2, 3]</code> - 选择1个元素	<code>solve(m, n)</code> 计算x使得: $m * x = n$

列表

```
l <- list(x = 1:5, y = c('a', 'b'))
列表是可以是不同类型的元素的集合。
```

<code>l[[2]]</code> 返回l的第二个元素 (类型向下退化)	<code>l[1]</code> 返回一个新得列表, 其仅包含第一个元素	<code>l\$x</code> 获取名字为x的元素	<code>l['y']</code> 返回一个新得列表, 其仅包含名字为y的元素
---	--	--------------------------------	--


另请参阅 **dplyr** 包

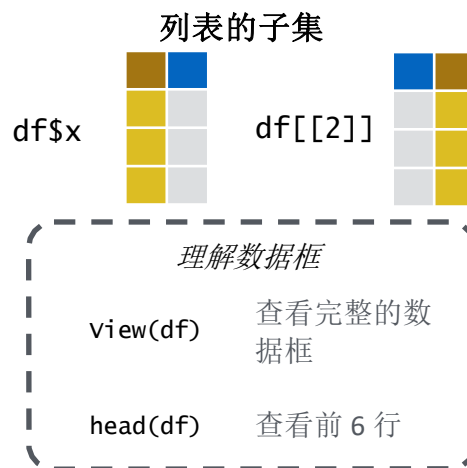
数据框

```
df <- data.frame(x = 1:3, y = c('a', 'b', 'c'))
列表的一种特殊情况, 其中所有元素的长度都相同。
```

x	y
1	a
2	b
3	c

矩阵的子集

<code>df[, 2]</code>	
<code>df[2,]</code>	
<code>df[2, 2]</code>	



`nrow(df)`
行的数量

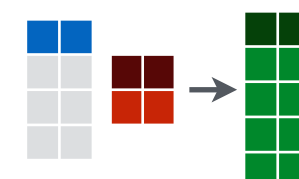
`ncol(df)`
列的数量

`dim(df)`
行与列的数量

`cbind` - 列追加



`rbind` - 行追加



字符串

另请参阅 **stringr** 包

<code>paste(x, y, sep = ' ')</code>	将多个向量连接在一起
<code>paste(x, collapse = ' ')</code>	将向量的元素连接在一起
<code>grep(pattern, x)</code>	在x中查找正则表达式匹配
<code>gsub(pattern, replace, x)</code>	用字符串替换x中的匹配项
<code>toupper(x)</code>	转换为大写
<code>tolower(x)</code>	转换为小写
<code>nchar(x)</code>	字符串中的字符数

因子

<code>factor(x)</code>	将向量转换为因子。可以设置因子的级别和顺序。
<code>cut(x, breaks = 4)</code>	通过“切割”成部分, 将数字向量变成一个因子。

统计

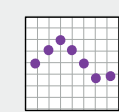
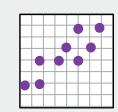
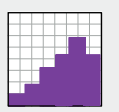
<code>lm(y ~ x, data=df)</code> 线性模型。	<code>t.test(x, y)</code> 对均值之间的差异执行 t 检验。	<code>prop.test</code> 检验几个组中的比值是否相同或是否等于某值。
<code>glm(y ~ x, data=df)</code> 广义线性模型。	<code>pairwise.t.test</code> 对配对数据执行 t 检验。	<code>aov</code> 方差分析。
<code>summary</code> 从模型中获取更多详细信息。		

分布

	随机变量	概率密度函数	累计分布	分位数
Normal (正态分布)	<code>rnorm</code>	<code>dnorm</code>	<code>pnorm</code>	<code>qnorm</code>
Poisson (泊松分布)	<code>rpois</code>	<code>dpois</code>	<code>ppois</code>	<code>qpois</code>
Binomial (二项分布)	<code>rbinom</code>	<code>dbinom</code>	<code>pbinom</code>	<code>qbinom</code>
Uniform (均匀分布)	<code>runif</code>	<code>dunif</code>	<code>punif</code>	<code>qunif</code>

绘图

另请参见 **ggplot2** 包

 <code>plot(x)</code> x的值按顺序排列	 <code>plot(x, y)</code> x对y的散点图	 <code>hist(x)</code> x的直方图 (分布图)
--	--	---

日期

另请参见 **lubridate** 包