

R 语言数据分析

Lecture 1 R 基础知识介绍

刘寅

统计与数学学院

大纲

- 1 R软件介绍
- 2 R的安装与启动
- 3 R的基本概念



- 西安交通大学出版社
- 阿兰•F•祖尔(Alain F.Zuur) 主编
- R语言初学者指南
- 参考资料：R绘图系统(第2版) Paul Murrell著，人民邮电出版社



- 1 R软件介绍
- 2 R的安装与启动
- 3 R的基本概念



1. 为什么要使用R软件

1.1 什么是R

R是

- ◆ 一个开放(GPL)的统计编程环境。



1. 为什么要使用R软件

1.1 什么是R

R是

- ◆ 一个开放(GPL)的统计编程环境。
- ◆ 一种语言，是S语言(由AT&T Bell 实验室的Rick Becker, John Chambers, Allan Wilks开发)的一种方言之一，另一个版本是S-plus。



1. 为什么要使用R软件

1.1 什么是R

R是

- ◆ 一个开放(GPL)的**统计编程环境**。
- ◆ 一种**语言**，是S语言(由AT&T Bell 实验室的Rick Becker, John Chambers, Allan Wilks开发)的一种方言之一，另一个版本是S-plus。
- ◆ 一种**软件**，是集统计分析与图形直观显示于一体的统计分析。



1. 为什么要使用R软件

1.1 什么是R

R是

- ◆ 一个开放(GPL)的**统计编程环境**。
- ◆ 一种**语言**，是S语言(由AT&T Bell 实验室的Rick Becker, John Chambers, Allan Wilks开发)的一种方言之一，另一个版本是S-plus。
- ◆ 一种**软件**，是集统计分析与图形直观显示于一体的统计分析。
- ◆ 一个**项目**，最早(1995年)由Auckland大学统计系的Robert Gentleman和Rose Ihaka 开始编制。



Robert Gentleman



Ross Ihaka



1. 为什么要使用R软件

1.1 什么是R

R是

- ◆ 一个开放(GPL)的**统计编程环境**。
- ◆ 一种**语言**，是S语言(由AT&T Bell 实验室的Rick Becker, John Chambers, Allan Wilks开发)的一种方言之一，另一个版本是S-plus。
- ◆ 一种**软件**，是集统计分析与图形直观显示于一体的统计分析。
- ◆ 一个**项目**，最早(1995年)由Auckland大学统计系的Robert Gentleman和Rose Ihaka 开始编制。



Robert Gentleman



Ross Ihaka

大家可以通过网站<http://www.r-project.org> 了解有关R的最新信息和使用说明，下载最新版本的R软件和R package。



1.2 R软件的优势和劣势

R软件的优势

- ◆ 完全免费，资源完全公开。

1.2 R软件的优势和劣势

R软件的优势

- ◆ 完全免费，资源完全公开。
- ◆ 可以在Linux, Windows和Mac环境下运行。

1.2 R软件的优势和劣势

R软件的优势

- ◆ 完全免费，资源完全公开。
- ◆ 可以在Linux，Windows和Mac环境下运行。
- ◆ 具有优秀的在线帮助系统。

1.2 R软件的优势和劣势

R软件的优势

- ◆ 完全免费，资源完全公开。
- ◆ 可以在Linux, Windows和Mac环境下运行。
- ◆ 具有优秀的在线帮助系统。
- ◆ 具有强大的画图功能。

1.2 R软件的优势和劣势

R软件的优势

- ◆ 完全免费，资源完全公开。
- ◆ 可以在Linux，Windows和Mac环境下运行。
- ◆ 具有优秀的在线帮助系统。
- ◆ 具有强大的画图功能。
- ◆ 能够轻松的转至具有商业支持的S-plus 程序。

1.2 R软件的优势和劣势

R软件的优势

- ◆ 完全**免费**，资源完全**公开**。
- ◆ 可以在Linux，Windows和Mac环境下运行。
- ◆ 具有优秀的在线帮助系统。
- ◆ 具有强大的画图功能。
- ◆ 能够轻松的转至具有商业支持的S-plus 程序。
- ◆ 程序语言容易学习，拥有很多内置的统计函数，使用方便。

1.2 R软件的优势和劣势

R软件的优势

- ◆ 完全**免费**，资源完全**公开**。
- ◆ 可以在Linux, Windows和Mac环境下运行。
- ◆ 具有优秀的在线帮助系统。
- ◆ 具有强大的画图功能。
- ◆ 能够轻松的转至具有商业支持的S-plus 程序。
- ◆ 程序语言容易学习，拥有很多内置的统计函数，使用方便。

R软件的劣势

- ◆ 完全免费，没有商业支持，大量数据或运算复杂时处理速度比较慢。

1.2 R软件的优势和劣势

R软件的优势

- ◆ 完全**免费**，资源完全**公开**。
- ◆ 可以在Linux，Windows和Mac环境下运行。
- ◆ 具有优秀的在线帮助系统。
- ◆ 具有强大的画图功能。
- ◆ 能够轻松的转至具有商业支持的S-plus 程序。
- ◆ 程序语言容易学习，拥有很多内置的统计函数，使用方便。

R软件的劣势

- ◆ 完全免费，没有商业支持，大量数据或运算复杂时处理速度比较慢。
- ◆ 需要较多的统计背景知识，需要自己编写代码。

1.3 学习策略

- ◆ 了解R的基本知识，包括下载、安装、编辑器的使用等等



1.3 学习策略

- ◆ 了解R的基本知识，包括下载、安装、编辑器的使用等等
- ◆ 了解R的基本原理，查找、下载，安装包的方法，包的使用方法等



1.3 学习策略

- ◆ 了解R的基本知识，包括下载、安装、编辑器的使用等等
- ◆ 了解R的基本原理，查找、下载，安装包的方法，包的使用方法等
- ◆ R中分析数据的基本流程：数据导入，分析、作图，保存分析结果和图形等



1.3 学习策略

- ◆ 了解R的基本知识，包括下载、安装、编辑器的使用等等
- ◆ 了解R的基本原理，查找、下载，安装包的方法，包的使用方法等
- ◆ R中分析数据的基本流程：数据导入，分析、作图，保存分析结果和图形等
- ◆ 更高级内容(如各种高级统计方法、复杂作图，可复制化研究等等)的学习：根据以后的需要，可能要花费更多的时间精力



2. R的安装与启动

2.1 R的下载和安装

- ◆ 进入网站<http://www.r-project.org>。



2. R的安装与启动

2.1 R的下载和安装

- ◆ 进入网站<http://www.r-project.org>。
- ◆ 点击左边**Download** 下的CRAN，从中选择一个镜像。



[\[Home\]](#)

Download



R Project

[About R](#)

[Logo](#)

[Contributors](#)

[What's New?](#)

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To **download R**, please choose your preferred [CRAN mirror](#).

If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.



◆ 选择适合电脑运行系统的版本: Linux, Windows或者Mac。

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you



◆ 选择适合电脑运行系统的版本：Linux，Windows或者Mac。

The Comprehensive R Archive Network

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Download R for Linux](#)
- [Download R for \(Mac\) OS X](#)
- [Download R for Windows](#)

R is part of many Linux distributions, you should check with your Linux package management system in addition to the link above.

Source Code for all Platforms

Windows and Mac users most likely want to download the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you

◆ 以Windows版本为例，选择base下的install R for the first time。点开，即可看到下载文件。

R for Windows

Subdirectories:

[base](#)

Binaries for base distribution (managed by Duncan Murdoch). This is what you want to [install R for the first time](#).

[contrib](#)

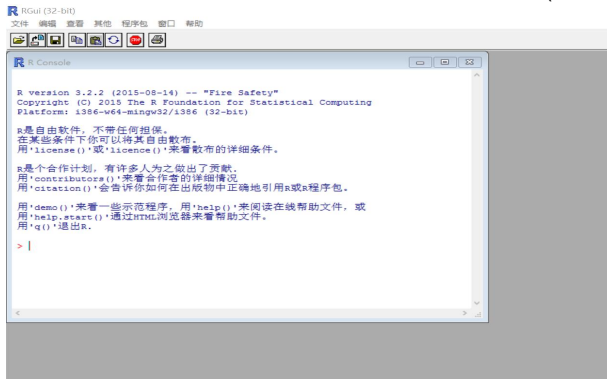
Binaries of contributed CRAN packages (for R \geq 2.11.x; managed by Uwe Ligges). There is also information on [third party software](#) available for CRAN Windows services and corresponding environment and make variables.



2.2 R程序启动

◆ 图形界面方式：运行Rgui.exe。

- 启动R，我们将会看到R Gui (Graphic user's interface) 的主窗口，它由三个部分组成：主菜单、工具条和R console(R的运行窗口)



– R console

- 你的主要工作是在运行窗口通过发布命令来完成，包括数据集的建立、数据分析、作图等。
- 在这里，你可以得到在线帮助：
 - `help.start()` HTML格式的关于R的帮助文件。
 - `help()` 得到相应函数的帮助，例如`help(plot)`。
 - `?XXX` 得到相应函数的帮助，例如`?plot`。

注意：所有的字符和标点符号都是英文状态下的。

- `q()` 退出R。
- 可用上、下方向键重现之前的命令。



2.3 R程序初步印象

- ◆ 开始部分给出了一些基本信息



2.3 R程序初步印象

- ◆ 开始部分给出了一些基本信息
- ◆ 尝试一些例子(demo)

```
> demo()  
> demo(lm.glm)  
> demo(persp)
```



3. R的基本概念

3.1 对象(objects)

- ◆ 对象是一个容器，用来存放R中的各种东西：数据，函数，操作符，公式，分析结果等



3. R的基本概念

3.1 对象(objects)

- ◆ 对象是一个容器，用来存放R中的各种东西：数据，函数，操作符，公式，分析结果等
- ◆ 对象都有一个名称



3. R的基本概念

3.1 对象(objects)

- ◆ 对象是一个容器，用来存放R中的各种东西：数据，函数，操作符，公式，分析结果等
- ◆ 对象都有一个名称
- ◆ 对象名称对字母大小写敏感，如A和a可以用来表示不同的对象



3. R的基本概念

3.1 对象(objects)

- ◆ 对象是一个容器，用来存放R中的各种东西：数据，函数，操作符，公式，分析结果等
- ◆ 对象都有一个名称
- ◆ 对象名称对字母大小写敏感，如A和a可以用来表示不同的对象
- ◆ 可以用对象名称来显示对象的内容



3. R的基本概念

3.1 对象(objects)

- ◆ 对象是一个容器，用来存放R中的各种东西：数据，函数，操作符，公式，分析结果等
- ◆ 对象都有一个名称
- ◆ 对象名称对字母大小写敏感，如A和a可以用来表示不同的对象
- ◆ 可以用对象名称来显示对象的内容
- ◆ 下面我们主要讨论一下数据对象



对象创建、列出和删除

◆ 创建对象：用赋值符(<- 或者=)

```
> a <- 1  
> b = "znufe"  
> b
```

```
[1] "znufe"
```



◆ 列出对象：使用ls函数，ls.str则给出对象的基本信息

```
> ls()
```

```
[1] "a" "b"
```

```
> ls.str()
```

```
a :   num 1
```

```
b :   chr "znufe"
```



◆ 删除对象：使用rm函数

```
> rm(a)
```

```
> ls()
```

```
[1] "b"
```



3.2 函数(Functions)

◆ 函数是一种特殊的对象



3.2 函数(Functions)

- ◆ 函数是一种特殊的对象
- ◆ 函数主要用于操作处理对象



3.2 函数(Functions)

- ◆ 函数是一种特殊的对象
- ◆ 函数主要用于操作处理对象
- ◆ 要带上一对括号，无论这个函数有没有参数，例如`ls()`。没有双括号，将显示该函数的内容(代码)

```
> q
```

```
function (save = "default", status = 0, runLast = TRUE)
.Internal(quit(save, status, runLast))
<bytecode: 0x06f483b8>
<environment: namespace:base>
```



3.2 函数(Functions)

- ◆ 函数是一种特殊的对象
- ◆ 函数主要用于操作处理对象
- ◆ 要带上一对括号，无论这个函数有没有参数，例如`ls()`。没有双括号，将显示该函数的内容(代码)

```
> q
```

```
function (save = "default", status = 0, runLast = TRUE)
.Internal(quit(save, status, runLast))
<bytecode: 0x06f483b8>
<environment: namespace:base>
```

- ◆ 有系统自带函数，也可以自己定义函数



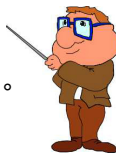
3.2 函数(Functions)

- ◆ 函数是一种特殊的对象
- ◆ 函数主要用于操作处理对象
- ◆ 要带上一对括号，无论这个函数有没有参数，例如ls(). 没有双括号，将显示该函数的内容(代码)

```
> q
```

```
function (save = "default", status = 0, runLast = TRUE)
.Internal(quit(save, status, runLast))
<bytecode: 0x06f483b8>
<environment: namespace:base>
```

- ◆ 有系统自带函数，也可以自己定义函数
- ◆ 系统自带函数都存放在库(library)中，库由包(packages)组成。



3.3 运算符(Operators)

- ◆ 运算符是一种特殊的函数



3.3 运算符(Operators)

- ◆ 运算符是一种特殊的函数
- ◆ 常见运算符：算术运算符，比较运算符，逻辑运算符，其他运算符



3.3 运算符(Operators)

- ◆ 运算符是一种特殊的函数
- ◆ 常见运算符：算术运算符，比较运算符，逻辑运算符，其他运算符
- ◆ 算术运算符：加减乘除幂 $+$ ， $-$ ， $*$ ， $/$ ， $^$



3.3 运算符(Operators)

- ◆ 运算符是一种特殊的函数
- ◆ 常见运算符：算术运算符，比较运算符，逻辑运算符，其他运算符
- ◆ 算术运算符：加减乘除幂 $+$ ， $-$ ， $*$ ， $/$ ， $^$
- ◆ 比较运算符：大于，大于等于，小于，小于等于，等于，不等于 $>$ ， $>=$ ， $<$ ， $<=$ ， $==$ ， $!=$



3.3 运算符(Operators)

- ◆ 运算符是一种特殊的函数
- ◆ 常见运算符：算术运算符，比较运算符，逻辑运算符，其他运算符
- ◆ 算术运算符：加减乘除幂 $+$ ， $-$ ， $*$ ， $/$ ， $^$
- ◆ 比较运算符：大于，大于等于，小于，小于等于，等于，不等于 $>$ ， $>=$ ， $<$ ， $<=$ ， $==$ ， $!=$
- ◆ 逻辑运算符：与或非 $&$ ， $|$ ， $!$



3.3 运算符(Operators)

- ◆ 运算符是一种特殊的函数
- ◆ 常见运算符：算术运算符，比较运算符，逻辑运算符，其他运算符
- ◆ 算术运算符：加减乘除幂 $+$ ， $-$ ， $*$ ， $/$ ， $^$
- ◆ 比较运算符：大于，大于等于，小于，小于等于，等于，不等于 $>$ ， $>=$ ， $<$ ， $<=$ ， $==$ ， $!=$
- ◆ 逻辑运算符：与或非 $&$ ， $|$ ， $!$
- ◆ 其他运算符：赋值符 $<-$ 或 $=$ ，提取符 $$$ ，矩阵乘法 $\%*\%$



3.4 包(Packages)

- ◆ R自动的函数和数据集都存放在包中



3.4 包(Packages)

- ◆ R自动的函数和数据集都存放在包中
- ◆ 包是存放在库中，每个包对应一个文件夹



3.4 包(Packages)

- ◆ R自动的函数和数据集都存放在包中
- ◆ 包是存放在库中，每个包对应一个文件夹
- ◆ 包按照重要程度可以分为：



3.4 包(Packages)

- ◆ R自动的函数和数据集都存放在包中
- ◆ 包是存放在库中，每个包对应一个文件夹
- ◆ 包按照重要程度可以分为：
 - 核心包(Core Packages):共有12个包：base,compiler, datasets, graphics, grDevices, grid, methods,parallel, splines, stats, stats4, tcltk，其中最重要的是base包



3.4 包(Packages)

- ◆ R自动的函数和数据集都存放在包中
- ◆ 包是存放在库中，每个包对应一个文件夹
- ◆ 包按照重要程度可以分为：
 - 核心包(Core Packages):共有12个包：base,compiler, datasets, graphics, grDevices, grid, methods,parallel, splines, stats, stats4, tcltk, 其中最重要的是base包
 - 推荐包(RecommendedPackages):共有15个包：KernSmooth, MASS, Matrix, boot,class, cluster, codetools, foreign, lattice, mgcv, nlme,nnet, rpart, spatial, survival.



3.4 包(Packages)

- ◆ R自动的函数和数据集都存放在包中
- ◆ 包是存放在库中，每个包对应一个文件夹
- ◆ 包按照重要程度可以分为：
 - 核心包(Core Packages):共有12个包：base,compiler, datasets, graphics, grDevices, grid, methods,parallel, splines, stats, stats4, tcltk, 其中最重要的是base包
 - 推荐包(RecommendedPackages):共有15个包：KernSmooth, MASS, Matrix, boot,class, cluster, codetools, foreign, lattice, mgcv, nlme,nnet, rpart, spatial, survival.
 - 贡献包(Contributed Packages):截至2013年5月8日CRAN上共有4499个包



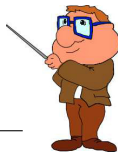
3.4 包(Packages)

- ◆ R自动的函数和数据集都存放在包中
- ◆ 包是存放在库中，每个包对应一个文件夹
- ◆ 包按照重要程度可以分为：
 - 核心包(Core Packages):共有12个包：base,compiler, datasets, graphics, grDevices, grid, methods,parallel, splines, stats, stats4, tcltk, 其中最重要的是base包
 - 推荐包(RecommendedPackages):共有15个包：KernSmooth, MASS, Matrix, boot,class, cluster, codetools, foreign, lattice, mgcv, nlme,nnet, rpart, spatial, survival.
 - 贡献包(Contributed Packages):截至2013年5月8日CRAN上共有4499个包
- ◆ 核心包和推荐包已经安装，贡献包需要自己安装



包按任务类型分类

CRAN Task Views		
Task	Description	中文描述
Bayesian	Bayesian Inference	贝叶斯推断
ChemPhys	Chemometrics and Computational Physics	计量化学和计算物理
ClinicalTrials	Clinical Trial Design, Monitoring, and Analysis	临床诊断
Cluster	Cluster Analysis & Finite Mixture Models	聚类分析
Distributions	Probability Distributions	概率分布
Econometrics	Computational Econometrics	计量经济学
Environmetrics	Analysis of Ecological and Environmental Data	环境计量学
ExperimentalDesign	Design of Experiments (DoE)	试验设计
Finance	Empirical Finance	金融学
Genetics	Statistical Genetics	遗传学
Graphics	Graphic Displays & Visualization	图形展示和可视化
gR	gRaphical Models in R	图论
HighPerformanceComputing	High-Performance and Parallel Computing with R	高性能计算
MachineLearning	Machine Learning & Statistical Learning	机器学习
MedicalImaging	Medical Image Analysis	医学图像分析
Multivariate	Multivariate Statistics	多元统计
NaturalLanguageProcessing	Natural Language Processing	语言处理
OfficialStatistics	Official Statistics & Survey Methodology	官方统计
Optimization	Optimization and Mathematical Programming	优化理论
Pharmacokinetics	Analysis of Pharmacokinetic Data	药理学
Phylogenetics	Phylogenetics, Especially Comparative Methods	种系遗传学
Psychometrics	Psychometric Models and Methods	心理计量学
ReproducibleResearch	Reproducible Research	可再生研究
Robust	Robust Statistical Methods	稳健方法
SocialSciences	Statistics for the Social Sciences	社会统计学
Spatial	Analysis of Spatial Data	空间数据
Survival	Survival Analysis	生存分析
TimeSeries	Time Series Analysis	时间序列



包的查找、安装、更新与删除

- ◆ 查找已经安装的包，使用library函数
 > library()



包的查找、安装、更新与删除

- ◆ 查找已经安装的包，使用library函数
 > library()
- ◆ 如何查找相关包



包的查找、安装、更新与删除

- ◆ 查找已经安装的包，使用library函数
 - > library()
- ◆ 如何查找相关包
 - 文献或其他人推荐



包的查找、安装、更新与删除

- ◆ 查找已经安装的包，使用library函数

```
> library()
```

- ◆ 如何查找相关包

- 文献或其他人推荐
- 按贡献包任务分类查找



包的查找、安装、更新与删除

- ◆ 查找已经安装的包，使用library函数

```
> library()
```

- ◆ 如何查找相关包

- 文献或其他人推荐
- 按贡献包任务分类查找
- 在贡献包列表中按关键字查找



包的查找、安装、更新与删除

- ◆ 查找已经安装的包，使用library函数

```
> library()
```

- ◆ 如何查找相关包

- 文献或其他人推荐
- 按贡献包任务分类查找
- 在贡献包列表中按关键字查找

- ◆ 如何安装包



包的查找、安装、更新与删除

◆ 查找已经安装的包，使用library函数

```
> library()
```

◆ 如何查找相关包

- 文献或其他人推荐
- 按贡献包任务分类查找
- 在贡献包列表中按关键字查找

◆ 如何安装包

- 使用Rgui的菜单安装



包的查找、安装、更新与删除

◆ 查找已经安装的包，使用library函数

```
> library()
```

◆ 如何查找相关包

- 文献或其他人推荐
- 按贡献包任务分类查找
- 在贡献包列表中按关键字查找

◆ 如何安装包

- 使用Rgui的菜单安装
- 使用install.packages函数
> install.packages("sem")



包的查找、安装、更新与删除

◆ 查找已经安装的包，使用library函数

```
> library()
```

◆ 如何查找相关包

- 文献或其他人推荐
- 按贡献包任务分类查找
- 在贡献包列表中按关键字查找

◆ 如何安装包

- 使用Rgui的菜单安装
- 使用install.packages函数
 > install.packages("sem")

◆ 如何更新包



包的查找、安装、更新与删除

◆ 查找已经安装的包，使用library函数

```
> library()
```

◆ 如何查找相关包

- 文献或其他人推荐
- 按贡献包任务分类查找
- 在贡献包列表中按关键字查找

◆ 如何安装包

- 使用Rgui的菜单安装
- 使用install.packages函数
> install.packages("sem")

◆ 如何更新包

- 使用Rgui的菜单更新(推荐)



包的查找、安装、更新与删除

◆ 查找已经安装的包，使用library函数

```
> library()
```

◆ 如何查找相关包

- 文献或其他人推荐
- 按贡献包任务分类查找
- 在贡献包列表中按关键字查找

◆ 如何安装包

- 使用Rgui的菜单安装
- 使用install.packages函数
 > install.packages("sem")

◆ 如何更新包

- 使用Rgui的菜单更新(推荐)
- 使用update.packages函数



包的查找、安装、更新与删除

◆ 查找已经安装的包，使用library函数

```
> library()
```

◆ 如何查找相关包

- 文献或其他人推荐
- 按贡献包任务分类查找
- 在贡献包列表中按关键字查找

◆ 如何安装包

- 使用Rgui的菜单安装
- 使用install.packages函数
> install.packages("sem")

◆ 如何更新包

- 使用Rgui的菜单更新(推荐)
- 使用update.packages函数

◆ 如何删除包：使用函数remove.packages

```
> remove.packages("sem")
```



- ◆ 确保要使用的包已经安装

```
> install.packages("sem")
```



包的使用

- ◆ 确保要使用的包已经安装

```
> install.packages("sem")
```

- ◆ 使用包之前必须加载(load)包，使用library函数

```
> library("sem")
```



包的使用

- ◆ 确保要使用的包已经安装

```
> install.packages("sem")
```

- ◆ 使用包之前必须加载(load)包，使用library函数

```
> library("sem")
```

- ◆ 删除加载的包，使用detach函数. 注：1.detach不会从硬盘删除包，2.base不能删除，3.注意detach的特殊用法

```
> detach("package:stats")
```





第一章结束了!

THANKS