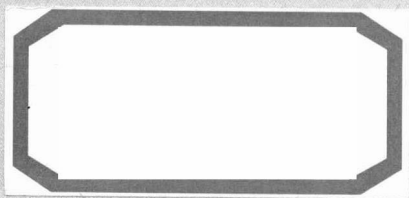


R语言 数据分析与挖掘 实战手册

程 静◎编著

中国铁道出版社有限公司
CHINA RAILWAY PUBLISHING HOUSE CO., LTD.



R语言 数据分析与挖掘 实战手册

程 静◎编著



RFID



中国铁道出版社有限公司
CHINA RAILWAY PUBLISHING HOUSE CO., LTD.

内 容 简 介

本书系统地介绍了利用 R 语言进行数据分析和挖掘的相关技术,采用由浅入深的框架体系:开篇伊始介绍 R 语言的基础操作,进而介绍回归分析、方差分析等数据分析的方法,以更好地探索数据内部结构,获取数据所包含的信息;更重要的是为后续的数据挖掘提供理论依据;最后介绍典型数据挖掘工具和方法,采用理论基础到算法介绍到案例实战的布局,让读者深刻感知数据挖掘的精髓,在了解算法的同时更好地学以致用。

本书提供的案例均给出了较为详尽地分析,以满足不同基础的读者的需要。本书适合的读者包括数据分析师、统计专业的本科生、经管类专业的研究生等。

图书在版编目 (C I P) 数据

R 语言数据分析与挖掘实战手册/程静编著. —北京:
中国铁道出版社有限公司, 2019. 6
ISBN 978-7-113-25745-3

I. ①R… II. ①程… III. ①程序语言-程序设计-手册
②数据处理-手册 IV. ①TP312-62②TP274-62

中国版本图书馆 CIP 数据核字 (2019) 第 081947 号

书 名: R 语言数据分析与挖掘实战手册
作 者: 程 静

责任编辑: 荆 波
责任印制: 赵星辰

读者热线电话: 010-63560056
封面设计: **MX** DESIGN
STUDIO

出版发行: 中国铁道出版社有限公司 (100054, 北京市西城区右安门西街 8 号)
印 刷: 中国铁道出版社印刷厂
版 次: 2019 年 6 月第 1 版 2019 年 6 月第 1 次印刷
开 本: 787mm×1092mm 1/16 印张: 16.5 字数: 347 千
书 号: ISBN 978-7-113-25745-3
定 价: 59.80 元

版权所有 侵权必究

凡购买铁道版图书, 如有印制质量问题, 请与本社读者服务部联系调换。电话: (010) 51873174

打击盗版举报电话: (010) 51873659

随着互联网技术的蓬勃发展，物联网、车联网和云计算等技术的日益成熟，人们的生活环境逐渐由一个数字化的网络体系覆盖。近年来，大数据、机器学习、人工智能等词汇不断出现在大众的视野中，而“数据挖掘”作为实现上述目标的核心利器，不容置疑地成为了数据分析者必须掌握的“关键技术”。

然而，直接接触或学习这项“关键技术”都是较为困难的，因为其本身涵盖了数学、统计学、算法编程等不同专业领域的知识，如何克服这种困难，如何在一本书中既讲清必要的理论知识，又能够使读者能够快速上手操作并在操作中学习更多的知识，成为一名数据分析达人，这是本书要解决的。

本书作者借助于多年的知识积累和实务工作经验，将数据分析和挖掘的各种“干货”浓缩于本书中，其中囊括了大量精美的图表与案例分析，行文深入浅出、图文并茂，将枯燥生硬的理论知识与案例分析相结合，便于读者更快地吸收知识并学以致用。本书抛开深奥的理论化条文，除了必备的基础理论知识介绍外，绝不贪多求全，特别强调实务操作、快速上手，绝不囿于示意与演示，更注重实战展示——从R语言软件的安装、数据的获取、数据的预处理、数据的探索性分析到回归分析等数据分析的方法，再到常规聚类等典型的数据挖掘工具和方法，随着本书内容的一步深入，读者将真正体会到数据挖掘的精髓和乐趣所在。

本书特色

1. 内容体系由浅入深、详略得当，行文安排适用于不同基础的读者

本书内容涵盖了软件安装、数据获取、数据预处理、数据的探索性分析等基础内容以帮助基础较为薄弱的初学者尽快入门，而后介绍了数据分析的一些强有力工具，如回归分析、方差分析、主成分分析、因子分析和判别分析等数据分析的方法，最后将所有知识综合起来形成真正的数据挖掘知识体系。本书对基础部分的细节进行详细介绍，对较为高深的理论知识和算法进行了简单的介绍，这种由浅入深、详略得当的行文安排适用于基础不尽相同的读者。

2. 内容切实可用，辅以大量实例，便于读者更快地掌握核心技术

本书注重实战操作，在基础部分进行详细讲解，如软件安装和数据处理等，均给出了实打实的教程，而在数据分析和挖掘技术的相关章节介绍中均采用了实例分

析，案例中的数据方便易得、真实可靠，在案例分析中按照数据挖掘的基本步骤进行，对分析结果进行详细解读，便于读者更好地理解 and 掌握每一章的核心技术。

3. 将分析结果进行可视化展示，激发读者的阅读兴趣

本书采用较多的可视化展示，从算法介绍到实例分析，尽量采用图片的形式进行解读，以帮助读者从繁琐的文字描述中“解脱”出来，尤其是实例分析部分，将能够进行可视化展示的部分转化成图片形式，并配以简明扼要的文字解说，以便读者更加深刻地理解每一章的内容。需要强调的是，在行业应用中，数据挖掘的结果大多以图片的形式汇总为可视化报告，因此本书的初衷就是建立与行业应用的更多联系。

二维码下载包

为了便于读者习，我们把全书源代码以及书中图片的彩色版放入二维码下载包中，供读者下载使用。

本书读者对象

- 数据分析师
- 统计专业的本科生
- 经管类专业的研究生
- R 语言的编程爱好者

因受作者水平和篇幅所限，本书难免存有疏漏和不当之处，敬请指正。

编者

2019年3月

第 1 章 R 语言简介

1.1 R 语言软件的安装与运行	1
1.1.1 R 语言软件的安装、启动与关闭	1
1.1.2 R 语言编程包的安装和使用	4
1.2 R 语言的数据结构	6
1.2.1 R 语言对象和类型	6
1.2.2 向量	7
1.2.3 数组和矩阵	12
1.2.4 列表	17
1.2.5 数据框	20

第 2 章 数据的读取与保存

2.1 数据的读取	24
2.1.1 读取内置数据集和文本文件	24
2.1.2 读取 Excel 数据和 CSV 格式的数据	30
2.1.3 读取 R 语言格式数据和网页数据	33
2.1.4 读取其他格式的数据	34
2.2 数据保存	36
2.2.1 写出数据	36
2.2.2 使用函数 cat()	37
2.2.3 保存为 R 语言格式文件	38
2.2.4 保存为其他类型文件	39

第 3 章 数据预处理

3.1 缺失值处理	40
3.1.1 缺失值判断	40
3.1.2 缺失模型判断	44
3.1.3 常用处理方法	48
3.2 数据整理	53
3.2.1 数据合并	53
3.2.2 选取子集	56
3.2.3 数据转换	59

第 4 章 数据的探索性分析

4.1 基本绘图函数	66
4.2 探索单个变量	74

4.2.1	单组数据的图形描述	74
4.2.2	单组数据的描述性分析	79
4.3	探索多个变量	81
4.3.1	两组数据的图形描述	81
4.3.2	多组数据的图形描述	85
4.3.3	多组数据的描述性统计	88
4.4	其他图像探索	90

第 5 章 回归分析

5.1	一元线性回归	94
5.1.1	模型简介	94
5.1.2	函数介绍	96
5.1.3	综合案例: iris 数据集的一元回归建模	97
5.2	多元线性回归	99
5.2.1	模型简介	99
5.2.2	综合案例: iris 数据集的多元回归建模	100
5.3	变量的选择	105
5.3.1	逐步回归方法简介及函数介绍	105
5.3.2	综合案例: swiss 数据集的逐步回归建模	106
5.3.3	岭回归的方法简介及函数介绍	109
5.3.4	综合案例: longley 数据集的岭回归探索	110
5.3.5	lasso 回归方法简介及函数介绍	114
5.3.6	综合案例: longley 数据集的 lasso 回归建模	115
5.4	Logistic 回归	117
5.4.1	模型简介	117
5.4.2	函数介绍	119
5.4.3	综合案例: iris 数据集的逻辑回归建模	120

第 6 章 方差分析

6.1	单因素方差分析	124
6.1.1	模型介绍	124
6.1.2	函数介绍	126
6.1.3	综合案例: 不同治疗方法下胆固醇降低效果的差异性分析	127
6.2	双因素方差分析	130
6.2.1	模型介绍	130
6.2.2	综合案例: 不同剂量下老鼠妊娠重量的差异性分析	132
6.3	协方差分析	136
6.3.1	模型简介	136
6.3.2	函数介绍	136
6.3.3	综合案例: hotdog 数据集的协方差分析	137

第 7 章 主成分分析和因子分析

7.1 降维的基本方法: 主成分分析	139
7.1.1 理论基础: 原始变量的线性组合	139
7.1.2 模型介绍	141
7.1.3 函数介绍	143
7.1.4 综合案例: longley 数据集的变量降维及回归	144
7.1.5 综合案例: longley 数据集的变量降维及回归 (主成分回归)	148
7.2 推广发展: 因子分析	150
7.2.1 理论基础: 多个变量综合为少数因子	150
7.2.2 模型介绍	151
7.2.3 函数介绍	153
7.2.4 综合案例: 能力和智商测试的因子分析探索	154

第 8 章 判别分析

8.1 距离判别法	160
8.1.1 理论基础: 离谁近, 就属于谁	160
8.1.2 函数介绍	162
8.1.3 综合案例: 基于距离判别的 iris 数据集分类	164
8.2 Bayes 判别法	168
8.2.1 理论基础: 先验概率与错判损失	168
8.2.2 函数介绍	170
8.2.3 综合案例: 基于 iris 数据集的 Bayes 判别分析	171
8.3 Fisher 判别法	171
8.3.1 理论基础: 投影	171
8.3.2 函数介绍	173
8.3.3 综合案例: 基于 Fisher 判别的 iris 数据集分类	174

第 9 章 常规聚类分析

9.1 深入了解聚类分析	178
9.1.1 差异与分类	178
9.1.2 主流的聚类算法	179
9.2 动态聚类	180
9.2.1 聚类的基本过程	180
9.2.2 函数介绍	183
9.2.3 综合案例: 基于随机生成序列的动态聚类	184
9.3 层次聚类	194
9.3.1 聚类的基本过程	194
9.3.2 函数介绍	197
9.3.3 综合案例: 基于 UScitiesD 数据集的层次聚类	199
9.4 密度聚类	202
9.4.1 聚类的基本过程	202

9.4.2	函数介绍	202
9.4.3	综合案例: 基于随机生成序列的密度聚类	203
9.5	EM 聚类	204
9.5.1	聚类的基本过程	205
9.5.2	函数介绍	205
9.5.3	综合案例: 基于 iris 数据集的 EM 聚类	206

第 10 章 关联规则

10.1	简单关联规则	210
10.1.1	基本概念与表示形式	210
10.1.2	评价简单关联规则的有效性和实用性	211
10.2	序列关联规则	212
10.2.1	差异与基本概念	212
10.2.2	生成序列关联规则	213
10.3	Apriori 算法	214
10.3.1	算法介绍: 挖掘频繁项集	214
10.3.2	函数介绍	215
10.3.3	综合案例: 基于 Titanic 数据集的关联规则挖掘	216
10.4	Eclat 算法	224
10.4.1	算法介绍: 自底向上的搜索	224
10.4.2	函数介绍	224
10.4.3	综合案例: 基于美国人口调查数据的关联规则挖掘	225
10.5	SPADE 算法	230
10.5.1	算法介绍: 基于序列格的搜索和连接	231
10.5.2	函数介绍	232
10.5.3	综合案例: 基于 zaki 数据集的序列关联规则挖掘	233

第 11 章 神经网络

11.1	深入了解人工神经网络	239
11.1.1	生物神经元	240
11.1.2	人工神经元模型	241
11.1.3	人工神经网络种类	244
11.1.4	建立模型的一般步骤	247
11.2	B-P 反向传播网络	248
11.2.1	B-P 反向传播网络模型	248
11.2.2	算法介绍	249
11.2.3	函数介绍	250
11.3	综合案例: 基于 Boston 数据的波士顿郊区房价预测建模	252



第 1 章

R 语言简介

R 语言是用于统计分析和数据挖掘的一门面向对象的编程语言，该语言的前身是由贝尔实验室开发的 S 语言，后由新西兰奥克兰大学的 Ross 和 Robert 共同开发，由于两位开发者的名字都是以 R 开头，故将此语言命名为 R 语言。目前，R 语言已经发展成为开放式的、免费开源的语言和环境。本章将会介绍最基本的 R 语言知识，以及 R 语言中关键术语的简单介绍，以帮助读者建立对 R 语言的整体认识。

1.1 R 语言软件的安装与运行

本节中，读者将会学习到 R 语言软件的最基本操作，即使大家不熟悉计算机编程语言，也同样可以掌握本节知识。

1.1.1 R 语言软件的安装、启动与关闭

首先介绍两个 R 语言的网站：R 语言工程网站和 R 语言的 CRAN 社区。

R 语言工程网站(<http://www.r-project.org/>)目前由 R 开发核心小组(R Development Core Team)维护，他们在该网站上发布 R 语言的相关信息，如 R 语言的简介、R 语言的更新、宏包信息、R 语言的常用手册、已出版的 R 语言的相关书籍、R 语言会议等信息。

R 语言的 CRAN 社区(<https://cran.r-project.org/>)是获取 R 语言软件和源代码，以及其他资源的主要场所，通过它或者其他镜像站点下载最新版本的 R 语言软件和大量的程辑包 (packages)。

由于 R 语言是免费开源的语言和环境，故我们可以从 CRAN 社区免费下载 R 语言并安装到本地计算机。

R 语言可以在大多数桌面环境下使用，目前开发者已将其移植到了各种主流系统，无论用户使用的是 Linux、UNIX、Windows 还是 Mac OS 系统，都可以在 R 语言网站上很方便地下载最新的 R 语言软件。本书主要介绍在 Windows 操作系统上安装并使用 R 语言，其他操作系统上 R 语言的安装和使用方法请参见 R 语言的相关说明。

首先，进入 R 语言的工程网站，界面如图 1.1 所示，可以看到，该页面上有一些蓝色的字体，是导入其他相关网页的超链接，还可以找到“download R”的字样，点击进入 CRAN 镜像网站列表，该列表按照国家名字排序，只需要找到地理位置比较靠近的镜像即可，如 China 下面的几个链接，任意选择一个链接点击就进入 R 语言的下载安装页面，如图 1.2 所示。该页面提供不同操作系统下的 R 语言安装包，选择合适的即可。

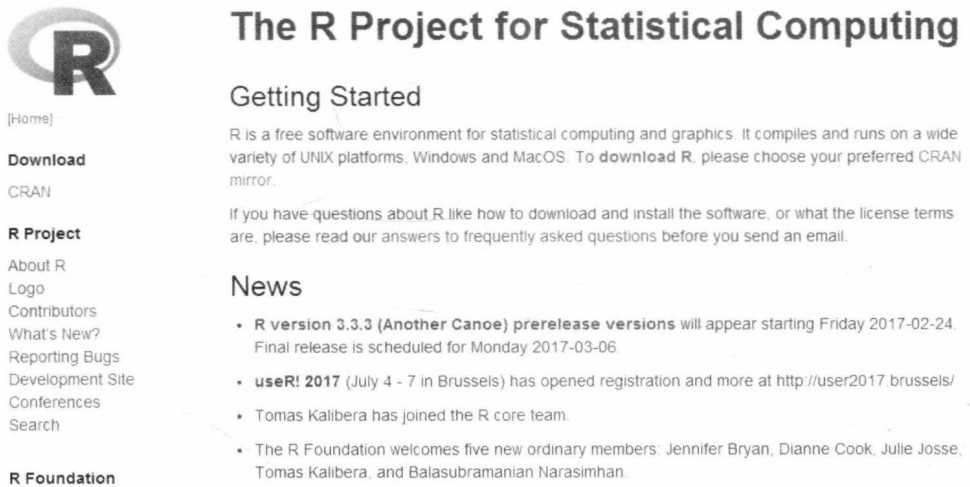


图 1.1 R 语言主页界面展示

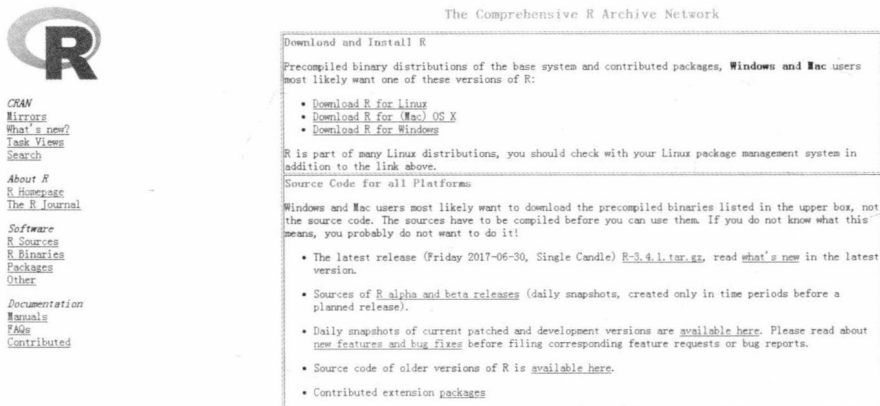


图 1.2 R 语言的安装页面

根据不同的操作系统点击链接，这里选择我们 Windows 系统的安装链接，目前 CARN 上针对 Windows 平台只有 32 位版本的 R 语言程序，但这些版本同样可以在 64 位 Windows 上正常运行，在软件的安装中也会有相应的选项来区分选择。

安装成功后，电脑桌面会出现蓝色的 R 图标，双击该图标就可以启动 R 语言软件。启动后界面如图 1.3 所示。

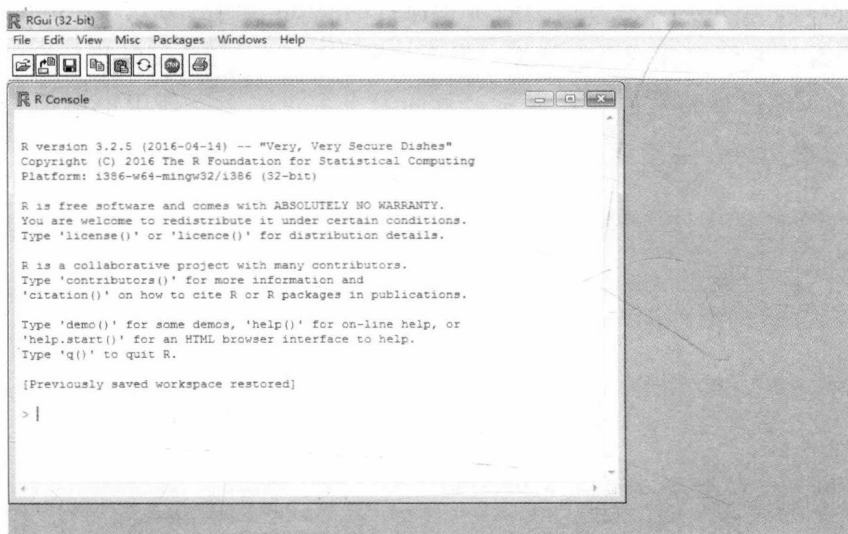


图 1.3 R 语言软件的界面展示

R 语言按照问答方式运行，即在命令提示符“>”后键入命令后回车，R 语言就可以完成相应的操作，例如输入以下命令，就可以得到如图 1.4 所示的结果。

```
> plot(iris)
```

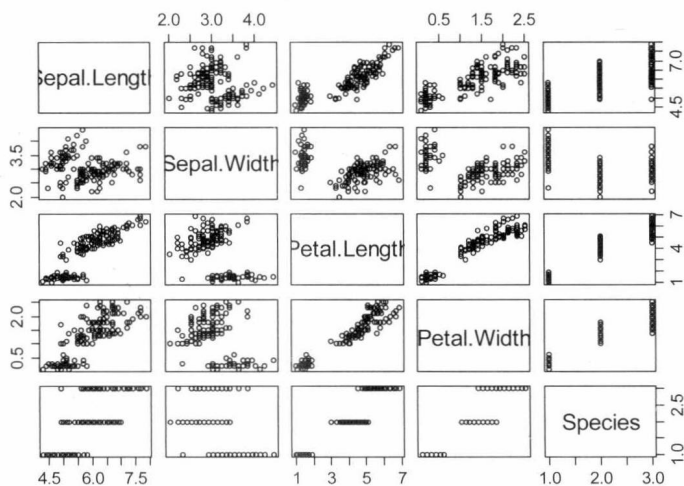


图 1.4 代码运行示例

如果需要关闭该软件，则有两种方式：

在命令提示符“>”后面输入 `q()`，会出现提示：`save workspace image?`即是否保存工作空间的镜像，点击是则保存，则之前在软件上的操作会全部被记录下来，如数据存储、函数编辑等，之后可以直接调用；否则不保存。

点击右上角的关闭图标，同样选择是否保存工作空间镜像后关闭软件。

除去标准的 RGui，现在还有一种非常流行的 R 语言运行方式：RStudio。它是一套免费、开源的 R 语言集成开发环境，界面如图 1.5 所示。RStudio 将所有的窗口都绑定在屏幕上，通过标签页的方式展示，界面看上去更美观，使用也更加方便。关于 RStudio 的安装运行，请参见 RStudio 的主页 (<http://www.rstudio.org/>)。

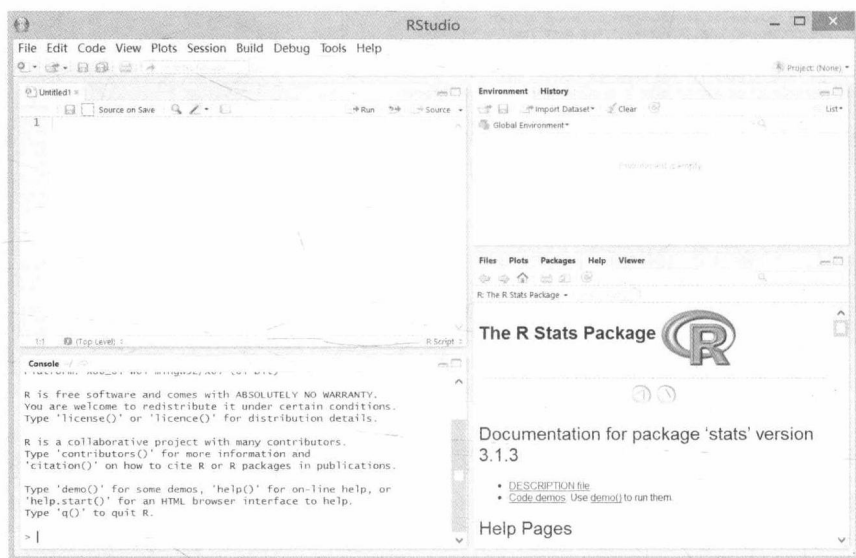


图 1.5 RStudio 的界面展示

1.1.2 R 语言程辑包的安装和使用

已经下载的 R 语言软件附有自带的程辑包，但仅包含最基本的程辑包和函数，在后续使用中会发现远远不够，所以需要读者自行下载一些必要的程辑包。

默认情况下，R 语言不会在启动时自动加载所有安装了的程辑包，因为这样会大大降低软件的启动速度，因此在使用每一个程辑包之前需要先加载该程辑包，否则在调用包里面的函数时 R 语言会报错。

在加载之前，需要确认该程辑包已经安装到电脑上，否则同样会报错。R 语言中可以通过函数 `library()` 来加载所需的程辑包。如：加载 Matrix 包（其中包含一系列关于矩阵的函数）。

```
> library(Matrix)
```

如果加载了没有安装过的程辑包，如在没有安装 kernlab 包（其中包括一系列关于支持向量机的函数）的情况下加载它，R 语言就会报错：

```
> library(kernlab)
Error in library(kernlab) : there is no package called 'kernlab'
```

此时需要先下载并安装 kernlab 程辑包，再加载它，下面将介绍具体步骤。

R 语言程辑包的安装分三种方式：

(1) 菜单方式：在已联网的情况下，按照步骤“packages→install package(s)→选择合适的 CRAN 镜像服务器→选定程辑包”进行下载安装；

(2) 命令方式：在已联网的情况下，在命令提示符“>”后键入以下命令，就可以下载并安装程辑包 kernlab；

```
> install.packages("kernlab")
trying URL 'https://mirrors.tuna.tsinghua.edu.cn/CRAN/bin/windows/contrib/3.2/kernlab_0.9-25.zip'
Content type 'application/zip' length 2135613 bytes (2.0 MB)
downloaded 2.0 MB
```

```
package 'kernlab' successfully unpacked and MD5 sums checked
```

```
The downloaded binary packages are in
  C:\Users\Administrator\AppData\Local\Temp\RtmpCEHoLU\downloaded_packages
```

需要注意的是：此处需要安装的是 kernlab 程辑包，上述输出结果显示了安装过程的基本信息，包括下载路径、宏包大小和存储路径，但有时候会需要安装一些其他相关的程辑包，安装过程会显示还需要先下载哪些包，但是 R 语言会自动下载这些包，不用再次键入命令。同样，在安装完成后会提示所下载的程辑包的存储路径：

(3) 本地安装：在无上网条件的情况下，可以提前从 CRAN 社区下载需要的程辑包及与之相关的程辑包，再按照第一种安装方式通过“packages”菜单中的“install package(s) from local zip files”选定本机上的程辑包（zip 格式）进行安装。

正如前面提到的，在使用 R 语言程辑包之前需要先加载，刚安装的 R 语言软件包含最基本的程辑包，如 base 包，这类程辑包在 R 语言启动时就已经完成自动加载，不用人为加载；其他的新安装的程辑包在使用时必须先加载，下面提供两种加载方式：

(1) 菜单方式：按照步骤“packages→load packages...”再从已有的程辑包中选定需要的进行加载；

(2) 命令方式：在命令提示符后输入 library(),括号内就是需要加载的程辑包名称，需要注意的是，R 语言对字母的大小写敏感，因此在安装和加载程辑包时需要注意名称不能有误，否则 R 语言不能识别，会报错。

此外，还可以对已有的程辑包进行更新，通过步骤“packages→update packages...”

来完成更新，或者直接键入 `old.packages()` 来更新已安装的 R 语言编程包。

1.2 R 语言的数据结构

前文已经介绍了使用 R 语言软件最基本的操作，下面将继续介绍 R 语言的基本数据结构。

1.2.1 R 语言对象和类型

R 语言是通过一些对象来运行的，这些对象可以用其名称和内容来刻画，也可以用对象的数据类型来刻画。所有的对象都有两个内在的属性：类型和长度。在 R 中，数据类型有两种不同的划分方法。

1. 按照存储划分：

(1) 数值型：又分为整型、单精度实型和双精度实型，用于存储身高、体重和年龄等形式的数据。整形是整数的组织形式，根据整数位数的长短确定不同的存储空间，一般需要 2~4 字节的空间；单精度实型通常需要 4 字节的空间存储；双精度实型通常需要 8 字节的空间存储，在 R 语言中数值型数据都默认为双精度实型数据，其一般的表现形式为： $1.234e-2$ ，即 1.234×10^{-2} 。

(2) 字符型：用于存储如姓名、所在单位等字符形式的数据，是夹在单引号 ‘ ’ 或双引号 “ ” 之间的字符串，字符型常量一般形如 “LiLei”、“Beijing” 等。

(3) 逻辑型：用于存储二分类的数据，如是否为男性（女性）、是否成年等，取值只能是 FALSE（或 F）、TRUE（或 T）。

(4) 复数型：一般具有 $a+bi$ 的形式。

2. 按照结构划分

可分为向量、矩阵、数组、列表和数据框，具体内容将在本章后几节分别进行详细介绍。当然，R 语言中还存在其他的类型，如函数或者表达式，但这些类型不能用于表示数据。其中函数是用来进行某些特定操作的 R 语言对象，通常需要进行参数设定，通过执行一系列的操作来产生结果，R 语言中已经包含了大量的函数以供调用，此外我们还可以创建新的函数，即用户自定义函数。

接下来将介绍怎样创建、访问和管理 R 语言数据对象：

(1) 创建数据对象的基本书写格式为：

对象名 \leftarrow R 常量、表达式或函数

其中 \leftarrow 符号称为赋值符号，其功能是将符号右侧的计算结果赋值给左侧的对象名，在 R 语言中，还可以用 “=” 来进行赋值，即对象名=R 常量、表达式或函数。

(2) 访问数据对象的基本书写格式为：

对象名或`print(对象名)`。

指定对象的对象值将会按照顺序输出。

(3) 管理 R 语言的数据对象，即浏览当前的数据对象，并删除不需要的数据对象，浏览数据对象的基本书写格式为：

`ls()`；

(4) 删除指定对象的基本书写格式为：

`rm(对象名或者对象名列表)`、`remove(对象名或者对象名列表)`。

1.2.2 向量

向量是 R 语言中最基本的数据对象，用于存储一组基本类型相同的数据，向量可以是数值型、字符型、逻辑型和复数型，在 R 语言中可以用 `c()` 函数和相应的参数设定来创建一个向量，

`c()` 函数的基本书写格式为：

`c(常量或向量名列表)`

其中，常量和向量名分别用逗号隔开，常量或向量名的数量就是向量的长度，可以用 `length()` 函数来获取向量的长度：

```
> x=c(1,2,3,4) #创建一个向量，并将其赋值给x
> x
[1] 1 2 3 4
> length(x)
[1] 4
```

如果需要查看向量中的元素类型，可以用 `typeof()` 函数：R 语言中的变量类型称为模式，我们知道向量存储的是一组基本类型相同的数据，即向量中所有元素都必须属于相同的模式。否则，R 语言将执行强制类型转换，可以用 `mode()` 函数来查看向量中元素所属的模式。

```
> typeof(x)
[1] "double"
> mode(x)
[1] "numeric"
```

从上述输出可知，向量 `x` 的元素类型是“double”，元素所属的模式为“numeric”；值得注意的是，对于数值型的向量，R 语言一般默认其元素类型为双精度实型(double)。下面我们创建一个向量，使其元素不属于同一个模式。


```
> y=c(2,3,4.5,'z')
> typeof(y)
[1] "character"
> mode(y)
[1] "character"
```

从上述输出可知，向量 y 的元素类型是 "character" 型，因为 R 语言执行了强制类型转化，将所有元素转化成字符型，同样，元素所属的模式也是 "character" 型。

如果想要检查对象 y 是否为向量，可以用 `is.vector()` 函数：

```
> is.vector(y)
[1] TRUE
```

除了上述介绍的创建向量的最基本函数 `c()`，R 语言中还有一些其他的常用函数。如 `seq` 函数、`rep` 函数、`scan` 函数以及利用随机分布函数来创建向量。

(1) `seq` 函数：也称为序列函数，是用来创建含有序列元素的向量，其基本书写格式为：

```
seq(from, to, by, length.out, along.with, ...)
```

参数介绍：

- **from**: 起始值，默认值为 1；
- **to**: 终止值，默认值为 1；
- **by**: 步长，即各元素之间的差值；
- **length**: 向量长度，即元素的个数。

```
> seq(0,3,by=0.5)
[1] 0.0 0.5 1.0 1.5 2.0 2.5 3.0
>seq(0,3,length=4)
[1] 0 1 2 3
```

(2) `rep` 函数：也称为重复函数，是用来创建含有重复元素的向量，其基本书写格式为：

```
rep(x, each = 1, times = 1, length.out = NA)
```

参数介绍：

- **x**: 向量或者类似向量的对象；
- **each**: x 中每个元素重复的次数，默认值为 1；
- **times**: **each** 后的处理，**each** 后生成的序列再重复 **times** 次，默认值为 1；
- **length.out**: **times** 处理后的向量最终输出的长度。

在举例之前，需要先了解一个运算符——冒号运算符。冒号运算符是能够产生序