

# 第 1 章

## 初识 R 语言

### 学习目标

- 1 了解 R 语言简史。
- 2 掌握 R 语言在大数据中的应用。
- 3 了解 R 语言在线资源。
- 4 了解 R 语言的开发。
- 5 掌握集成开发环境的安装。

### 知识导图



 本章导读

R 是一套完整的数据处理、计算和制图软件系统，R 语言是目前世界上使用最广泛的统计编程语言，它是数据科学家的第一选择。由于 R 是一种专业性很强的统计语言，所以学习者必须有统计学的知识基础，不然很多东西会掌握得比较慢。

本章将对 S 语言的诞生、R 语言在大数据中的应用，以及 R 语言手册、在线资料、R 语言联盟和集成开发环境等进行介绍，使读者对 R 语言基础有初步的了解，为后面的学习打下基础。

## 1.1 R 语言概述

### 1.1.1 S 语言的诞生

S 语言是一种用来进行数据探索、统计分析、制图的解释型语言，它于 1975—1976 年由贝尔实验室的 Rick Becker、John Chambers 和 Allan Wilks 开发。

第一个可运行的 S 语言版本在 1976 年发表，在 GCOS 操作系统上运行，当时它还没有正式名称，曾经被人称为互动式 SCS (interactive SCS, ISCS)、统计运算系统 (statistical computing system)、统计分析系统 (statistical analysis system, SAS)。直到 1979 年，它才被正式定名为 S 语言。

S 语言丰富的数据类型 (如向量、数组、列表、对象等) 特别有利于实现新的统计算法，其交互式运行方式、强大的图形交互功能使得我们可以方便地探索数据。

S 语言的实现版本主要是 S-PLUS。它基于 S 语言，并由 MathSoft 公司的统计科学部进一步完善。作为统计学家及一般研究人员的通用方法工具箱，S-PLUS 强调演示图形、探索性数据分析、统计方法、开发新统计工具的计算方法以及可扩展性。

S-PLUS 可以直接进行标准的统计分析，并得到所需结果，它的主要特点是可以交互地从各个方面发现数据中的信息，并可以很容易地实现一个新的统计方法。

### 1.1.2 R 语言简史

在 S 语言的基础上，新西兰奥克兰大学的 Robert Gentleman 和 Ross Ihaka 及其他志愿人员在 S 语言的基础上开始构思一种新的用于统计学分析的开源语言，因为他们名字的第一个字母都是 R，所以这门语言就叫作 R 语言。R 是 S 语言的一个分支，一般人认为 R 就是 S 语言的一种实现。S 语言的实现版本主要是 S-PLUS，这是一个由 MathSoft 公司开发的一种基于 S 语言的统计学软件，而 R 也是 S-PLUS 的基础，所以它们在程序语法上几乎一样，可能只是在函数方面有细微差别，所以两门语言只要稍加修改，就可以非常容易地移植到对方程序中。

### 1.1.3 关于大数据

一般认为，大数据 (big data) 是指无法在一定时间范围内用常规软件工具进行捕捉、管理和处理的数据集合。这里也有许多不同的定义，比如研究机构 Gartner 认为，大数据是需要新处理模式才能具有更强的决策力、洞察发现力和流程优化能力的海量、高增长



率和多样化的信息资产。而麦肯锡全球研究所给出的定义是“一种规模大到在获取、存储、管理、分析方面大大超出传统数据库软件工具能力范围的数据集合，具有海量的数据规模、快速的数据流转、多样的数据类型和价值密度低四大特征。”总之，大数据的核心不在于掌握庞大的数据信息，而在于对这些含有意义的数据进行专业化处理。换言之，如果把大数据比作一种产业，那么这种产业实现盈利的关键在于提高对数据的“加工能力”，通过“加工”实现数据的“增值”。

如果把数据比喻为蕴藏能量的煤矿，煤炭按照不同的性质可分为焦煤、无烟煤、肥煤、贫煤等类，而露天煤矿、深山煤矿的挖掘成本又不一样。与此类似，大数据并不在“大”，而在于“有用”，在于让数据产生更高的价值。所以，数据价值含量、挖掘成本比数据的量更重要。对于很多行业而言，如何利用这些大规模数据是赢得竞争的关键。

大数据的价值主要体现在以下几个方面：

(1) 为企业提供基础的数据统计报表分析服务。分析师能够轻易获取数据，生成分析报告，指导产品和运营。产品经理能够通过统计数据完善产品功能，改善用户体验。运营人员可以通过数据发现运营问题，并确定运营的策略和方向。管理层可以通过数据掌握公司业务运营状况，从而进行一些战略决策。

(2) 人工智能离不开数据。数据作为人工智能发展的重要基础，在未来的智能化时代也将扮演着重要的角色，所以数据的价值也必然会随着人工智能技术的发展而得到提升。由此可以预测，一家公司如果拥有海量的用户数据，那无异于拥有一座取之不尽的矿山。

(3) 未来的大数据能更好地解决社会、商业、科学各类问题。大部分数据都与人类有关，要通过大数据解决人的问题。比如，建立个人的数据中心，每个人的日常生活习惯，身体状况，社交网络，知识能力，爱好性情，情绪波动记录等数据可以充分利用，医疗机构将实时地监测用户的身体健康状况；教育机构针对用户制订培训计划；服务行业提供符合用户习惯的服务；社交网络为志同道合的人群相识相知而提供便利；政府能在用户心理健康出现问题时及时干预；金融机构能为用户的资金提供更有效的使用建议和规划；道路交通平台可以为用户提供合适的出行线路。

### 1.1.4 R 语言在大数据中的应用

R 语言（简称 R）是专门为数据科学而生的软件，能帮用户完成该领域内统计分析和开发与维护等工作。R 语言拥有 15 000 多个成熟稳定的 R 包，几乎所有类型的数据分析任务都可以在 R 中完成，更强大的是，它还提供了各种数学计算、统计计算的函数，从而使使用者能灵活机动地进行数据分析，甚至创造出符合需要的新的统计计算方法。R 语言是一套完整的数据处理、计算和制图软件系统，其功能叙述如下。

(1) 完整连贯的统计分析工具：R 内建多种统计学及数字分析功能，是统计分析、绘图功能的自由开源软件，拥有完整体系的数据分析和挖掘工具，能够有效地存储和处理数据。

(2) 向量、矩阵运算等强大的数组运算工具：虽然 R 主要用于统计分析或者开发统计相关的软件，但也有人将其用作矩阵计算。其分析速度可媲美 GNU Octave，甚至商业软件 MATLAB。

(3) 优秀的统计制图功能：R 具有丰富的数据挖掘工具包（packages），拥有完整体系的数据统计和分析工具，为数据分析和显示提供了强大的图形功能。

(4) 简便而强大的编程语言：可操纵数据的输入和输出，可实现分支、循环，用户可





自定义功能。

最后，在 R 语言里的软件包生态系统，特别在机器学习方面，R 语言体现出强大的联动效应，即在大数据领域的任何新型研究成果可能都会马上以 R 软件包的形式体现出来。因此，从这个角度看，R 语言始终站在技术发展的尖端位置。

## 1.2 R 语言社区

### 1.2.1 R 语言手册

关于 R 语言手册，初学者或者英文水平不高的读者可以借助各个搜索引擎，搜索中文版的语言手册，这些资料很容易获得。但如果想更深入地学习，还是建议看官方的文档，因为翻译过来的内容可能有偏差。R 语言的官方文档地址是 <https://www.r-project.org/>，官方文档中提供了 R 相关的学习手册，单击“Manuals”按钮即可浏览，建议所有读者都看一看。另外，如果有不会的函数，在寻找搜索引擎搜索之后，也可以看 R 函数在线帮助 (<https://www.rdocumentation.org/>)，进而获得对某个函数或者包的帮助信息。

### 1.2.2 在线资源

R 语言的在线资源有很多，相信读者在学习的过程中也能搜集到很多资源。常见的如菜鸟教程、W3Cschool 以及慕课等学习网站上都有很多免费的学习资源。在此为大家推荐如下资源网站，供学习使用。

丰富的 R 语言博客资源：<http://www.r-bloggers.com>

R 语言资源汇总：<https://github.com/qinwf/awesome-R>

R 语言搜索引擎：<http://www.rseek.org>

R 语言函数在线帮助：<http://www.rdocumentation.org>

关于 R 语言的问答：<http://stackoverflow.com/questions/tagged/r>

一个入门级的 R 语言在线教程：<http://tryr.codeschool.com>

一个交互式的 R 语言在线教程：<https://www.datacamp.com>

统计之都：<http://cos.name>

如果大家在学习或工作中遇到 R 语言的一些问题，除了国内的搜索引擎外，这里推荐 Stack Overflow 网站 (<https://stackoverflow.com>)。Stack Overflow 是一个与程序相关的 IT 技术问答网站，用户可以在该网站免费提交问题，浏览问题，索引相关内容。

### 1.2.3 R 语言联盟

2015 年 7 月，Linux 基金会宣布成立了 R 语言联盟 (R Consortium)，用以加强技术和用户交流的合作计划，其宗旨是为 R 社区、R 基金会、团体以及个人使用、维护和分发 R 软件提供帮助。R 语言联盟将与 R 基金会和其他组织共同工作，并向其提供开发、维护和分销 R 软件服务，以及为 R 用户社区提供统一框架。R 语言联盟的创始公司和组织包括 R 基金会、微软、RStudio、谷歌和惠普等。

R 语言被统计学家、分析师以及数据科学家用来发掘数据价值。这是一种免费开源的



编程语言，用于统计计算和为数据分析、建模以及可视化提供互动环境。R 语言联盟将会补充 R 基金会工作的不足，帮助这家位于奥地利的非营利性机构维护这种语言。R 语言联盟将会着力用户外延以及其他项目设计，帮助 R 语言用户。

R 语言联盟将打造一个新的代码托管平台 R-Hub，用于 R 语言的开发和分发软件包，这项计划得到 Linux 基金会的资助。R-Hub 平台将向 R 软件包提供开发、构建、测试和验证服务，目标是将其打造成“满足 R 社区所有需求的构建平台”。R-Hub 将基于并兼容现有的 CRAN 和 R-Forge 网站。

## 1.2.4 用户活动

R 语言作为统计和数据挖掘界广泛应用的工具，每年 R 的官方机构都会举办多场 R 语言的学术会议，各个国家及地区也定期有 R 用户的交流活动。在中国的北京、上海、杭州、广州等地，每年都会有多场 R 语言会议，会议内容覆盖数据科学在各行各业的应用，包括天文、地理、医疗、生物、金融、能源、互联网等领域，在高校和业界均有深远影响，促进了 R 语言乃至数据科学在中国的推广和发展。如今 R 语言会议成为 R 语言社区在国内影响力较大的交流盛会，聚学术专家、业界精英、技术大咖于一堂，让更多的数据人参与其中，促进社区内部的交流和进步。

比如，在活动家网站 (<https://www.huodongjia.com>) 上，可以搜索到近期将要举办的线上或线下的 R 语言学术活动，有兴趣的读者可以报名参加。当然，大多数学术会议都是收费的，也有部分是免费的。

## 1.3 集成开发环境

### 1.3.1 R GUI

本节以 Windows x64 系统为例安装 R GUI，安装包 R 的官方网站是 <https://cran.r-project.org/>，界面如图 1-1 所示。从这里可以看到当前最新的 R 版本号，如果了解最新版本的详细信息，可以单击“[What's new?](#)”链接查看。

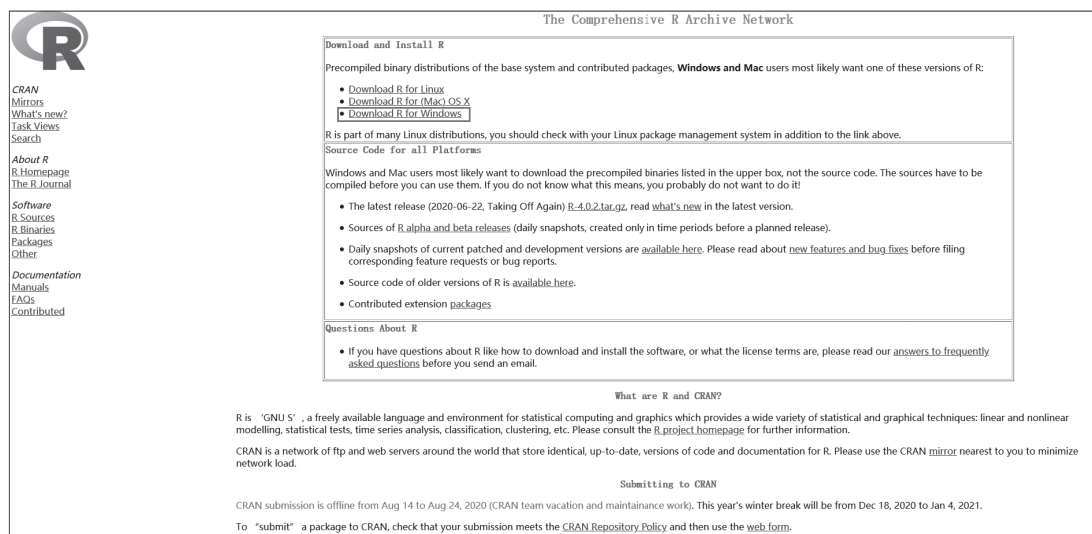


图 1-1 R 的官方网站



笔记



此时选择自己对应的计算机合适的版本。本节以 Windows 系统为例，所以单击“Download R for Windows”链接。之后进入下一界面，如图 1-2 所示。

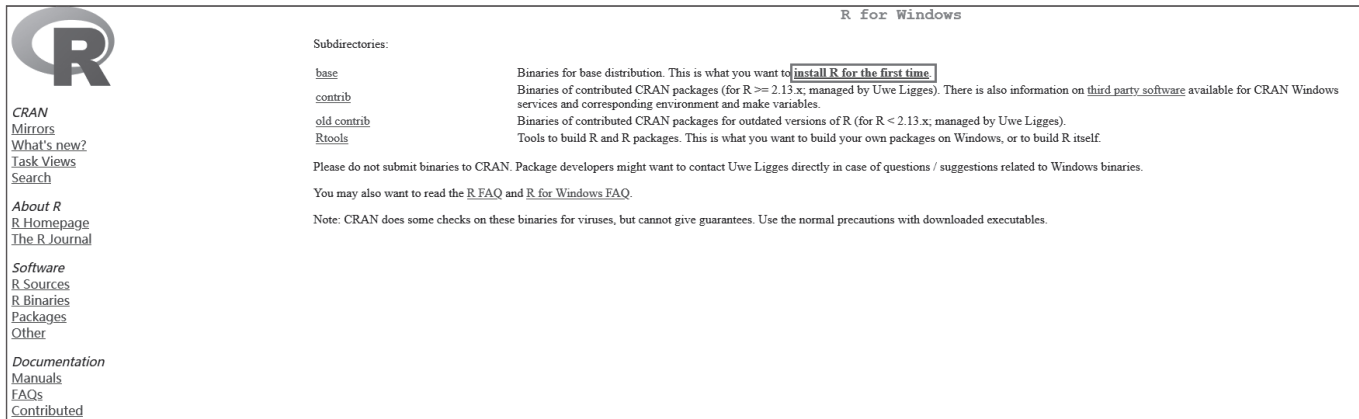


图 1-2 Download R for Windows

单击“install R for the first time”链接，进入新的界面，如图 1-3 所示。

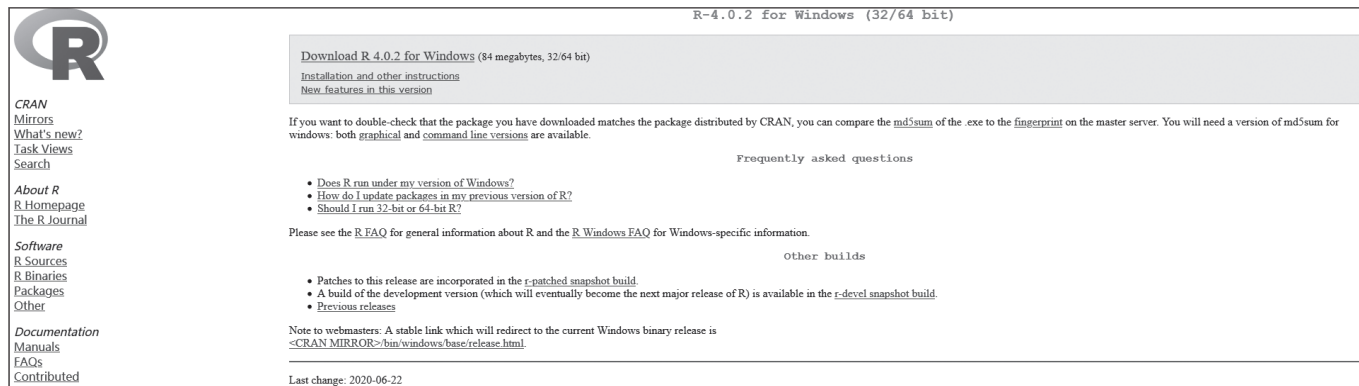


图 1-3 install R the first time

单击“Download R 4.0.2 for Windows ( 84 megabytes, 32/64 bit)”链接，在弹出的对话框中单击“保存文件”按钮，然后等几分钟，下载完成。

打开下载的 R-4.0.2-win.exe。若打开 R-4.0.2-win.exe 的时候出现问题，可以尝试单击鼠标右键，以管理员身份运行。在“选择语言”对话框中，建议使用“中文（简体）”，如图 1-4 所示。

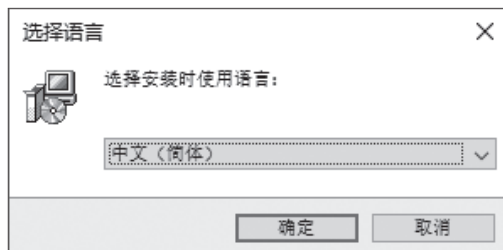


图 1-4 选择语言

选择语言后，单击“确定”按钮，在后边的安装向导对话框中单击“下一步”按钮，当需要选择安装的路径时，建议放在 D 盘，此处将其装至“D:\Program Files (x86)\R-4.0.2”，接着单击右下角的“下一步”按钮，进入“选择组件”对话框，如图 1-5 所示。

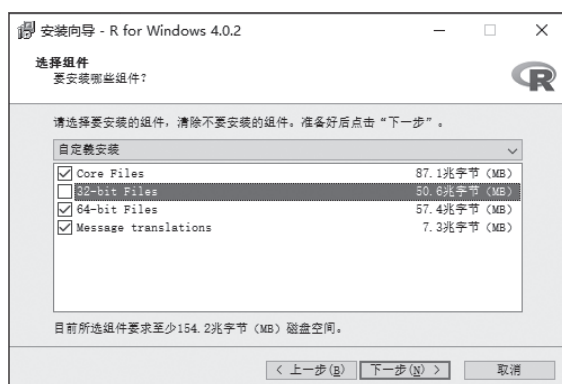


图 1-5 选择组件

在图 1-5 中选择安装的组件，即以 Windows x64 为例，需要取消 32-bit Files 前的“√”，接着一直单击右下角的“下一步”按钮，直到进入“选择附加任务”对话框，如图 1-6 所示。

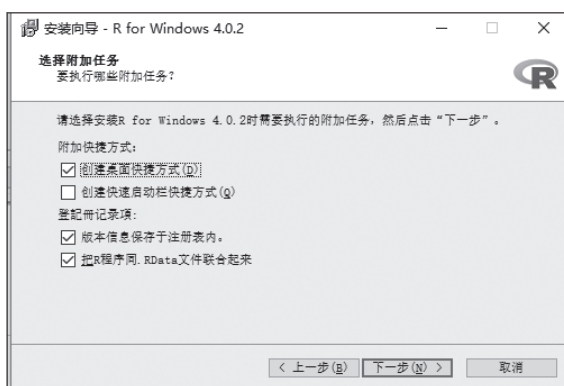


图 1-6 选择附加任务

在选择附加任务中勾选“创建桌面快捷方式”，安装完成后，直接单击 R 桌面的快捷方式就可以打开，这样会更方便一点。当然，如果此处没有选择创建快捷方式，以后可通过程序启动文件，重新创建快捷方式。单击右下角的“下一步”按钮，进入安装等待，直到成功安装完成。

成功安装后，打开桌面的快捷方式，就可以看到如图 1-7 所示的界面。

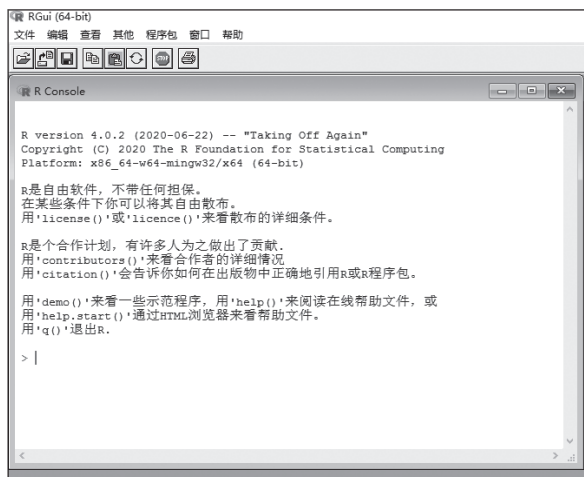


图 1-7 RGui 界面





## 1.3.2 RStudio IDE

在 1.3.1 节中，我们成功地安装了 R。使用 R 自带的环境可以进行操作，但 R 自带的环境操作起来可能不是很方便，而 RStudio 很好地解决了这个问题，而且它还具有调试、可视化等功能，支持纯 R 脚本、Rmarkdown（脚本文档混排）、Bookdown（脚本文档混排成书）、Shiny（交互式网络应用）等。

RStudio 是 R 语言的集成开发环境，是一款 R 语言的 IDE，分为面向桌面用户 IDE 和 Linux R 服务器版编辑器两种编辑器，采用 AGPL v3 与 RStudio License Agreement 双协议授权。R 是 RStudio 的基础，必须先安装 R，再安装 RStudio。即使只使用 RStudio，也需要事先为计算机安装好 R。RStudio 只是辅助你使用 R 进行编辑的工具，因为它自身并不附带 R 程序。

下面讲一下如何安装 RStudio。首先，打开 RStudio 官方网站 <https://www.rstudio.com/products/rstudio/download/>，如图 1-8 所示。

RStudio Desktop	RStudio Desktop	RStudio Server	RStudio Server Pro
Open Source License	Commercial License	Open Source License	Commercial License
<b>Free</b>	<b>\$995</b>	<b>Free</b>	<b>\$4,975</b>
	/year		/year
			(5 Named Users)
<b>DOWNLOAD</b>	<b>BUY</b>	<b>DOWNLOAD</b>	<b>BUY</b>
Learn more	Learn more	Learn more	Evaluation   Learn more

图 1-8 RStudio 官方网站

选择 Free（免费）软件，然后单击“DOWNLOAD”下载，进入如图 1-9 所示的界面。本书成稿时，RStudio 最新版本是 RStudio Desktop 1.3.1073，支持的最低 R 版本是 3.0.1，在 1.3.1 节我们已安装过。此处直接选择 All Installers 中的 RStudio-1.3.1073.exe-Windows 10/8/7，单击开始下载。

OS	Download	Size	SHA-256
Windows 10/8/7	<input type="checkbox"/> RStudio-1.3.1073.exe	171.62 MB	2fea472a
macOS 10.13+	<input type="checkbox"/> RStudio-1.3.1073.dmg	148.66 MB	087bb305
Ubuntu 16	<input type="checkbox"/> rstudio-1.3.1073-amd64.deb	124.07 MB	6d71c5ff
Ubuntu 18/Debian 10	<input type="checkbox"/> rstudio-1.3.1073-amd64.deb	126.78 MB	86be9352
Fedora 19/Red Hat 7	<input type="checkbox"/> rstudio-1.3.1073-x86_64.rpm	146.95 MB	01abb3d8
Fedora 28/Red Hat 8	<input type="checkbox"/> rstudio-1.3.1073-x86_64.rpm	151.04 MB	4b4e4878
Debian 9	<input type="checkbox"/> rstudio-1.3.1073-amd64.deb	126.98 MB	0226b0c2
SLES/OpenSUSE 12	<input type="checkbox"/> rstudio-1.3.1073-x86_64.rpm	119.43 MB	7c3aef2c
OpenSUSE 15	<input type="checkbox"/> rstudio-1.3.1073-x86_64.rpm	128.39 MB	29078f11

图 1-9 RStudio 下载

打开刚下载的安装包，如图 1-10 所示。在安装向导中，根据提示单击“下一步”按钮，再选择安装位置，这里仍然建议在 D 盘！再次单击“下一步”按钮，然后直接单击“安装”就可以了。

安装成功后，在自己安装的路径下选择 bin 文件→rstudio.exe，右击鼠标，从弹出的快捷菜单中执行“发送到”→“桌面快捷方式”命令，如图 1-11 所示。

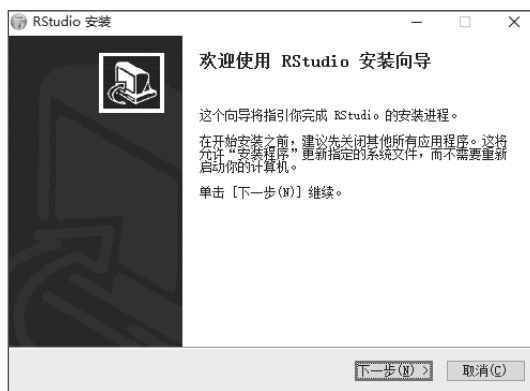


图 1-10 RStudio 安装

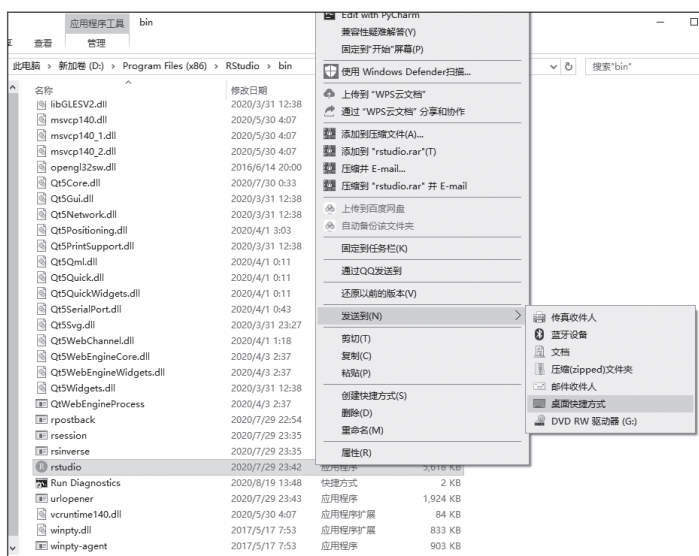


图 1-11 RStudio 快捷方式

然后进入桌面，双击 RStudio 快捷键即可打开如图 1-12 所示的界面。

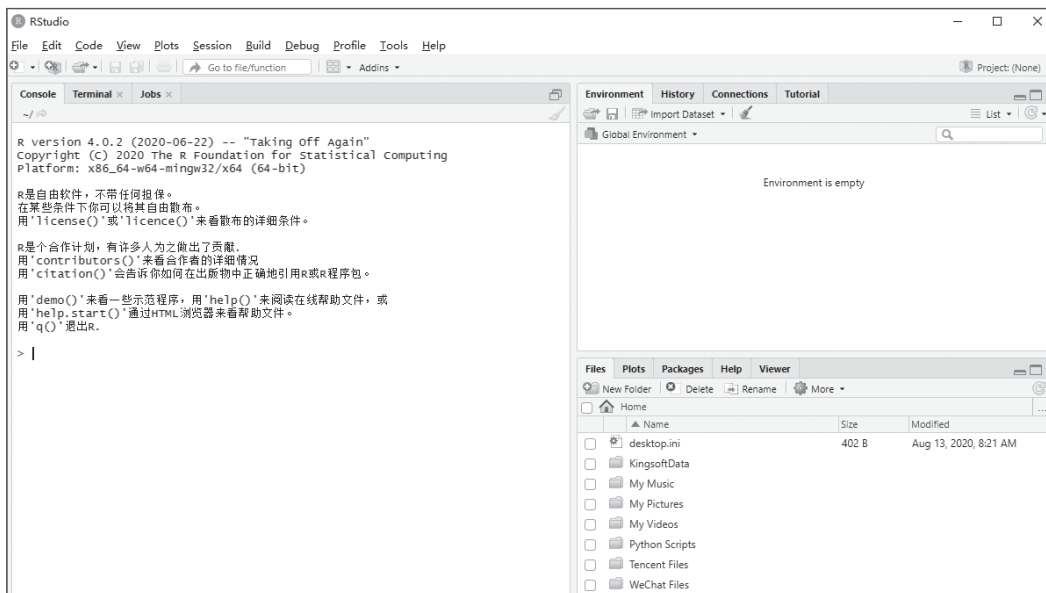


图 1-12 RStudio 界面

至此，RStudio 安装成功。



笔记

### 1.3.3 Jupyter Notebook

学过 Python 的读者，应该接触或使用过 Jupyter Notebook。Notebook 是一个交互式笔记本，支持运行 40 多种编程语言。按照 Jupyter 创始人的说法，起初他是想做一个综合 Julia)、Py (Python) 和 R 三种科学运行语言的计算工具平台，所以将其命名为 Jupyter。

它可以内置很多编程语言内核 (kernel)，默认的内核是 Python，所以，当安装 Notebook 后，就可以直接运行 Python 代码。

Jupyter 发展到现在，已经成为一个几乎支持所有语言，能够把软件代码、计算输出、解释文档、多媒体资源整合在一起的多功能科学运行平台。Jupyter Notebook 的工作界面如图 1-13 所示。可以看出，当前还不支持创建 R 语言项目。

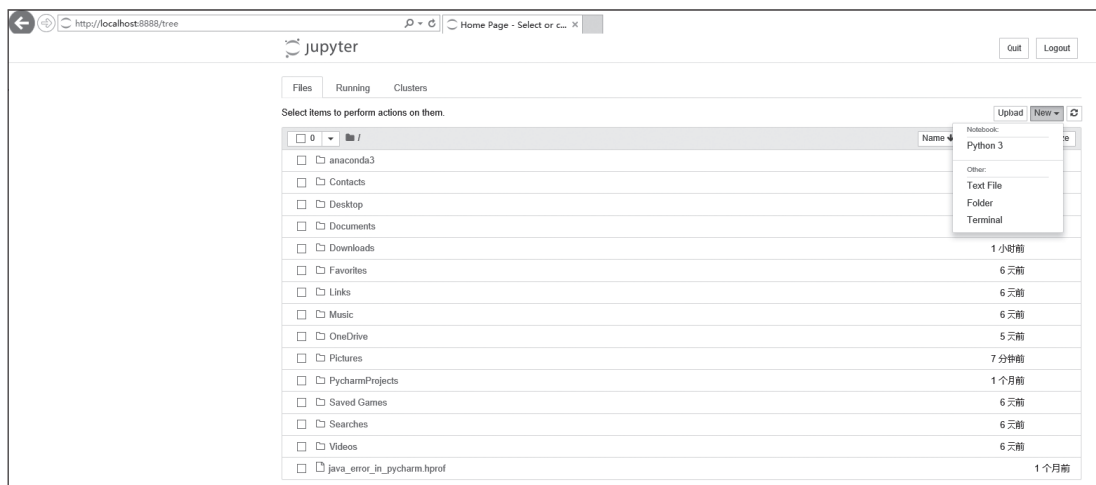


图 1-13 Jupyter Notebook 的工作界面

下面总结了几个 Jupyter Notebook 的优势。

#### 1. 整合了所有资源

在软件开发过程中频繁地进行上下文切换，会影响生产效率。举一个例子，假设需要切换窗口看一些文档，再切换窗口用另一个工具画图，不断地切换窗口就会成为影响效率的因素。而 Jupyter Notebook 则不同，它会将所有用软件编写的资源全部放在一个地方，无须切换窗口就可以轻松找到。

#### 2. 交互性编程体验

在机器学习和数据统计领域，Python 编程的实验性特别强，比如为了测试 100 种不同的方法，有时就需要将一小块代码重写 100 遍。在这种情况下，如果使用传统的 Python 开发流程，每一次测试都要将所有代码重新运行一遍，这样会花费开发者很多时间。

Jupyter Notebook 引进了 Cell 的概念。每次测试可以只运行一小部分代码，并且在代码下方立刻就能看到运行结果。如此强的交互性，满足了 Python 程序员可以专注问题本身，不会被频繁的工具链拖累，也不用在命令行之间来回切换，所有工作都能在 Jupyter Notebook 上完成。

#### 3. 轻松运行他人编写的代码

同样是在机器学习和数学统计领域，我们可能会借鉴他人分享的代码，但当复制过来想运行时，却需要使用 pip 安装一大堆依赖的库，这足以让人抓狂。而 Jupyter Notebook



就可以解决这个问题。

例如，Jupyter 官方的 Binder 平台以及 Google 提供的 Google Colab 环境，它们可以让 Jupyter Notebook 变得和 Google Doc 在线文档一样。比如用 Binder 打开一份 GitHub 上的 Jupyter Notebook 时，就不需要安装任何 Python 库，直接打开代码就能运行。

通过以上介绍，我们对 Jupyter Notebook 有了初步的了解，下面介绍 Jupyter Notebook 的安装。安装 Jupyter Notebook，推荐使用 Anaconda。Anaconda 是一个基于 Python 的数据处理和科学计算平台，它已经内置了许多非常有用的第三方库，装上 Anaconda，就相当于把 Python 和一些如 Numpy、Pandas、Scrip、Matplotlib 等常用的库自动安装好了，使得安装比常规 Python 安装要容易。当然，Python 不是本书讲述的内容，但因为它是内置于 Jupyter Notebook 的，所以安装 Jupyter Notebook 后默认就安装了 Python。

首先，进入 Anaconda 的官网 <https://www.anaconda.com/download/#windows>，如图 1-14 所示，下载对应的版本，这里选择的是 64-Bit Graphical Installer (466 MB)。

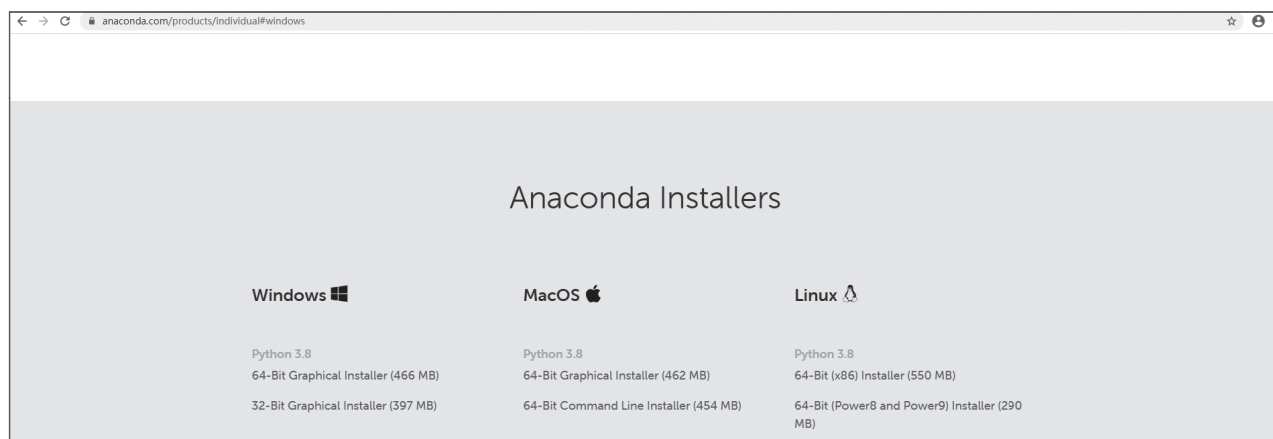


图 1-14 Anaconda 官网

下载完成后，双击下载好的 .exe 文件进行安装，单击“Next”按钮，单击“I agree”，选择“Just Me”，之后单击“Next”按钮。注意，在选择安装目录对话框中，仍然建议安装在 D 盘，单击 Next 按钮进入如图 1-15 所示的界面，建议勾选红色框中的“Add Anaconda to my PATH environment variable”，自动配置环境变量，单击“Install”按钮等待完成。

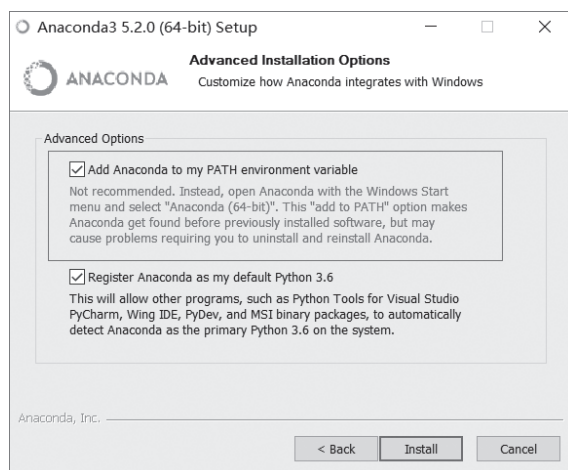


图 1-15 Anaconda 环境变量





在接下来的对话框中单击“skip”，进入图 1-16，取消两个勾选，最后单击“Finish”按钮。

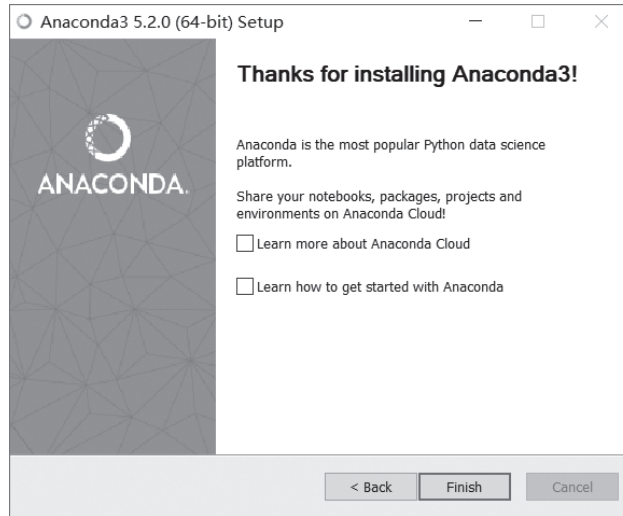


图 1-16 安装完成

安装成功后，在桌面左下方的搜索框中输入“jupyter notebook”，可看到图 1-17，双击打开 Jupyter Notebook 即可看到如图 1-13 所示的界面。



图 1-17 打开 Jupyter Notebook

下面介绍如何在 Jupyter Notebook 中安装 R 语言内核。要想在 Jupyter Notebook 中运行 R 语言其实非常简单，打开 RStudio，依次输入以下代码，安装对应的扩展包即可。

```
install.packages('repr','IRdisplay','evaluate','crayon','pbdZMQ','devtools','uuid','digest')
library(devtools)
install_github("IRkernel/IRkernel")
IRkernel::installspec()
```

执行完毕，打开 Jupyter Notebook，就可以新建 R 项目了，如图 1-18 所示。

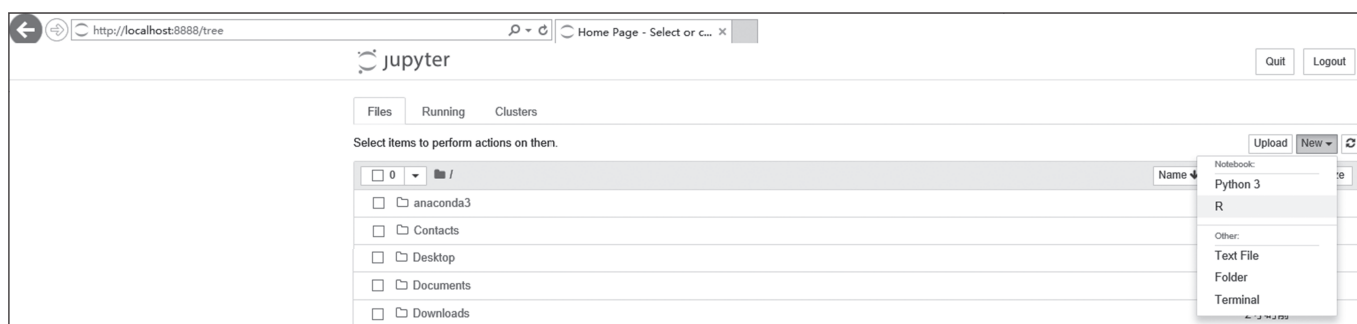


图 1-18 Notebook 中显示 R

## 1.4 R 语言的开发

在本书的编排上，首先介绍给读者的是 R 语言的初级内容，主要包括了解 R 语言的对象，掌握程序的基本编写逻辑，怎样进行数据导入、整理、转换，怎样进行假设检验，然后由浅入深进一步介绍如何分析，以及如何使用各种包。

R 语言的学习同样遵循二八定律，即百分之八十的精力需要花费在软件之外的统计学理论背景、业务知识学习上，而需要使用 R 软件实现的部分，不要单纯地学 R 语言的基础语法，尽管这些很重要，需要明确的是，统计学的理论背景更加重要，如果理论理解透了，很多事情就水到渠成，迎刃而解了。这一点特别体现在对统计与数据分析的学习上。

对于数据可视化，读者需要在牢固掌握基础语法运用、数据清洗技能的基础上，能够熟练运用一套图形语法。掌握这些之后，就不要将过多精力放在工具和代码本身，而是要更多地关注可视化素养，提升设计审美水平。数据可视化除了依赖实现的工具和平台语法之外，更多的是对于数据源的理解、对于可视化的理解、对于设计理念的融会贯通（怎么配色、怎么排版、怎么搭配字体等）。当然，对于设计、审美、创意这些柔性的东西，很难通过一两本书或者一两套课程完全解决，这些是内化于生活，积累于日常的点点滴滴。当然，如果有意识地通过一些课程、书籍慢慢培养，日积月累也会见效的。

R 语言是统计学家开发的，问世之初就决定了它的使命是统计计算和数据可视化，这算是 R 语言核心功能的两个大方向。

R 包是 R 的精髓，说是 R 的全部，一点也不为过。2006 年 3 月 15 日，第一个 R 包（coxrobust）正式上线 CRAN。近年来，R 包出现了井喷的趋势，目前已经有超过 20000 个 R 包，如果再算上 GitHub 上托管的个人开发的小众包，可能有好几万了，而且还在不断增长，每天都在增长，每一个 R 包都有各自的功能，R 的强大，正是因为这些程序包。对于这两个方向而言，统计计算的学习，基础都在课堂理论与专业背景上，说实话，R 语言只是提供了一个实现的平台而已，它并不改变或者创造新的理论、模型。

虽然这些包里的函数很多，使初学者眼花缭乱，但实际上，随着学习的深入，这些统计计算使用的公式、用到的模型算法，大部分都被封装到一个个扩展包里，导入包之后，仅调用对应函数、设置对应参数即可。这些函数与 Excel 里面的函数没有区别，不必恐惧。

至于参数的调优、模型的检验与优化，这些操作依赖的知识背景也基本来源于课堂学





习和专业背景，与 R 软件的关系并不大，对于需要自己写算法的情况，也仅是在函数的基础上按照成熟的理论算法进行调优和计算，这与软件无关，而与软件之外的专业背景和行业经验有关。

说到底，对于统计学习这一块，重要的是理论背景、业务经验，真正需要 R 实现的仅是内置的扩展包函数、基础语法。

类比一下 SPSS 的学习，一个不懂统计学的人很难学好 SPSS，尽管他知道各种功能模块和菜单。同样，一个不懂统计学和数学的人也很难学好 R 语言，尽管他很熟悉 R 语言的基础语法和很多扩展包所能实现的功能。