

课程复习01



课堂测试时间

- 1、数据集alpe_d_huez2描述了环法自行车赛期间Alpe d'Huez赛段的最快时间，以及关于年份和吸毒指控的背景信息。绘制出车手最快时间的分布。使用a) 直方图和b) 箱线图显示它们。
- 2、mtcars是datasets包中的数据集。请使用str()函数了解这个数据集的构成，并输出数据集，然后按要求画图：
 - * a. 我们要设置一个蓝色背景和红色的点或线。我们应该使用什么命令
 - * b. 画出cyl和mpg关系的散点图，并将结果输出为plot.png，要求输出为白底，360px*360px,点的大小为72
- 3、obama_vs_mccain数据集描述了2008年美国总统选举中的各州投票信息，以及关于收入，失业，种族和宗教的背景信息。
 - * a. 画出收入Income和参加选举比例Turnout之间的关系的散点图。提示：Turnout存在Na值。
 - * b. 将上述图形点的形状为黑色实心三角形(17)
 - * c. 数据集中有一个因子类型的列regions,请画出每个地区region下的收入Income和参加选举比例Turnout之间的关系的散点图。要求设置布局为5列，行优先。

课堂练习复习

- 安装R
 - 安装RStudio
-
- 熟悉R和RStudio的界面和菜单功能
-
- 安装swirl, 了解如何使用, 可以做几个练习试试

- 输入并执行例子0001
-

- 执行一下前面讲的几个Demo
-

- 熟悉包的发现和安装
-

- 熟悉工作空间
-

- 熟悉输入输出（例子0001执行结果分别或同时送到文件和屏幕）

- 打开帮助文档首页，并查阅其中的“Introduction to R”的第1章和13章。
-
- 安装vcd包。
 - 列出此包中可用的函数和数据集。
 - 载入这个包并阅读数据集Arthritis的描述。
 - 显示数据集Arthritis的内容 (直接输入一个对象的名称将列出它的内容)。
 - 运行数据集Arthritis自带的示例。它基本上显示了接受治疗的关节炎患者较接受安慰剂的患者在病情上有了更多改善。
 - 退出。

- 看RIA第一章
 - 使用帮助看所有出现的函数说明，了解所有函数功能
 - 把第8页、第11页和第12页的三个例子输入到R文件，看执行效果
-
- 看A Introduction to R的第二章和附录A（可选）
-
- 实现一个简单的hello world程序

An Integrated Development Environment for R

Getting Started with

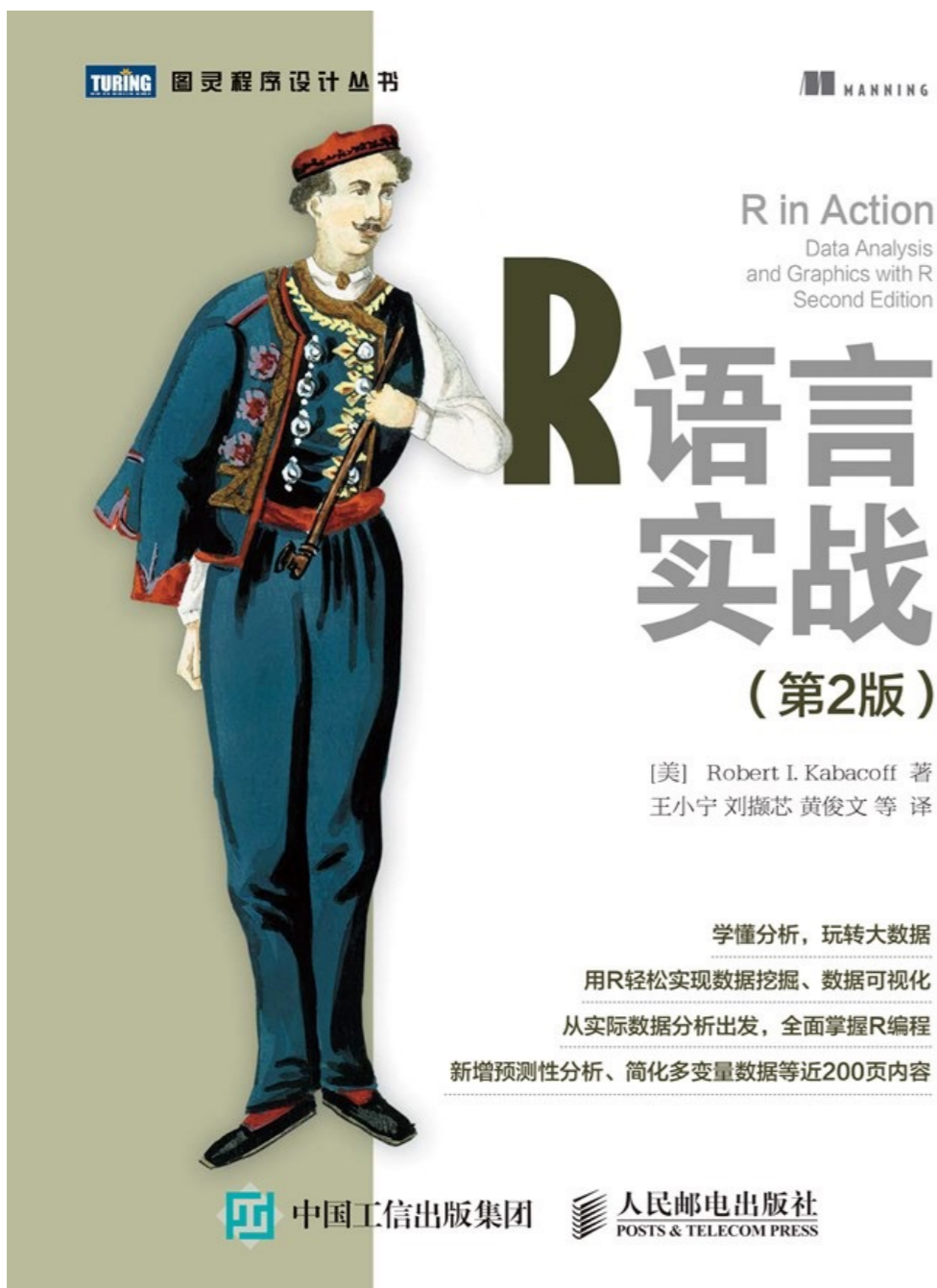
RStudio



O'REILLY®

John Verzani

- 第一章
- 第三章
- 第五章



第二章



第一章到第四章

INTERACTIVE COURSE

Introduction to R

Practice Now

Replay Course

Bookmarked

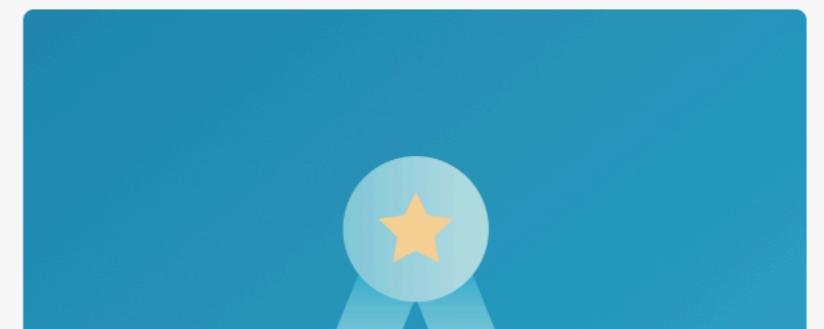


🕒 4 hours | ▶ 0 Videos | </> 62 Exercises | 👥 1,470,780 Participants | 📊 6,200 XP

📱 ALSO AVAILABLE ON MOBILE

Course Description

In Introduction to R, you will master the basics of this widely used open source language, including factors, lists, and data frames. With the knowledge gained in this course, you will be ready to undertake your first very own data analysis. Oracle



<https://www.datacamp.com/courses>

- 注册一个账号
- 完成Introduction to R课程
- 提交一个PDF，包括自己账户页面和课程页面



1 Intro to basics 100%


In this chapter, you will take your first steps with R. You will learn how to use the console as a calculator and how to assign variables. You will also get to know the basic data types in R. Let's get started!

[Hide Chapter Details](#) Continue Chapter

2 Vectors 100%

In this free R course, we'll take you on a trip to Vegas, where you will learn how to analyze your gambling results using vectors in R! After completing this chapter, you will be able to create vectors in R, name them, select elements from them and compare different vectors.

[View Chapter Details](#) Continue Chapter


sunhp

3100 XP Earned **0** Courses Completed **32** Exercises Aced

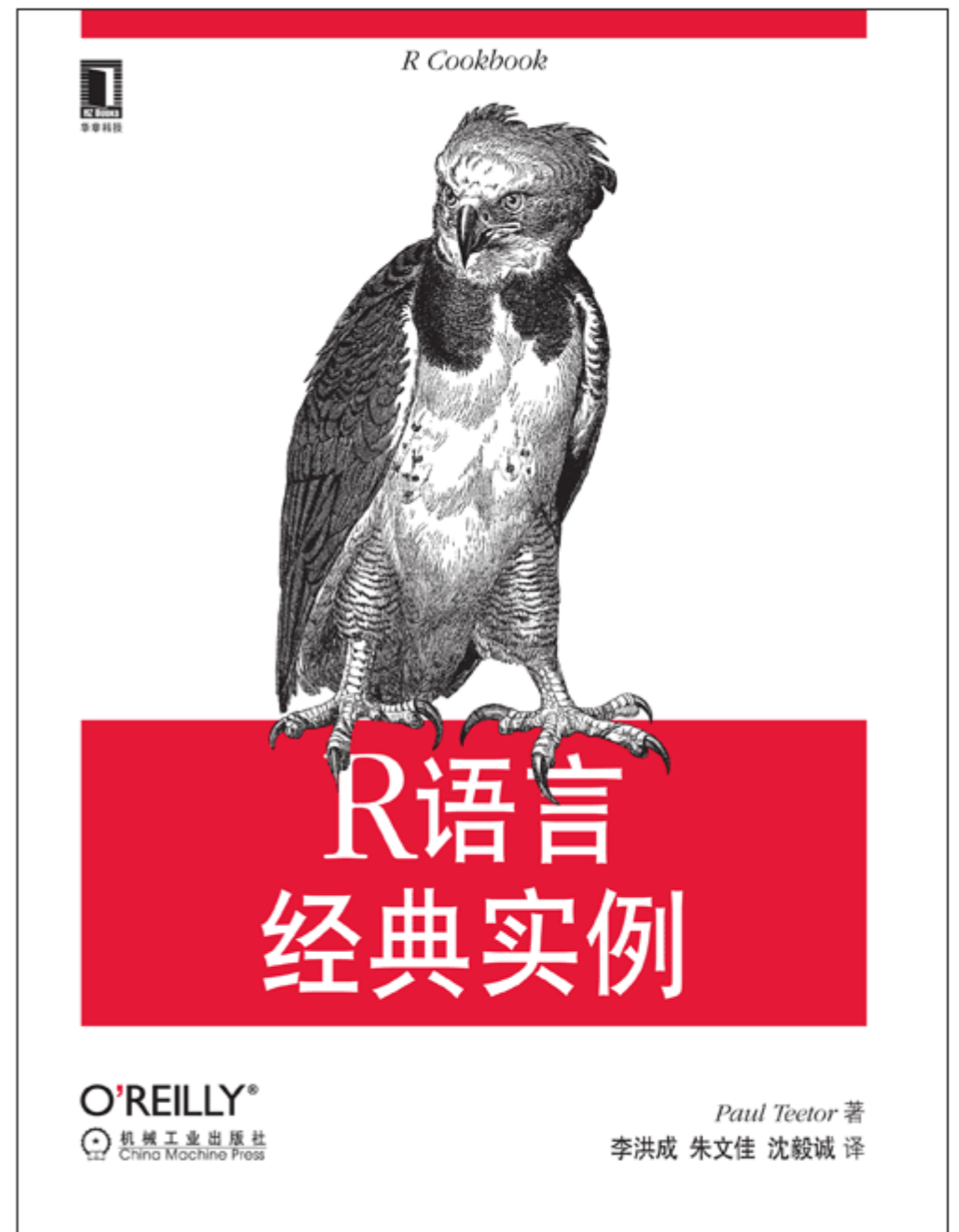
sunhp's Skills

Programming
3000/48450 XP

Importing & Cleaning Data
100/34750 XP



第四章、第五章




第五章-第七章

INTERACTIVE COURSE

Intermediate R

Continue Course Bookmark



6 hours | 14 Videos | 81 Exercises | 354,555 Participants | 6,950 XP


提交方式和上节课一样!

<https://www.datacamp.com/courses>

INTERACTIVE COURSE

Cleaning Data in R

Continue Course Bookmark



4 hours | 15 Videos | 58 Exercises | 108,240 Participants | 4,700 XP

2.5 已知有 5 名学生的数据，如表 2.3 所示。用数据框的形式读入数据。

表 2.3: 学生数据

序号	姓名	性别	年龄	身高 (cm)	体重 (kg)
1	张三	女	14	156	42.0
2	李四	男	15	165	49.0
3	王五	女	16	157	41.5
4	赵六	男	14	162	52.0
5	丁一	女	15	159	45.5

2.6 将例 2.5 中的数据表 2.3 的数据写成一个纯文本文件，用函数 `read.table()` 读该文件，然后再用函数 `write.csv()` 写成一个能用 *Excel* 表能打开的文件，并用 *Excel* 表打开。

2.7 编写一个 R 程序 (函数). 输入一个整数 n , 如果 $n \leq 0$, 则中止运算, 并输出一句话: “要求输入一个正整数”; 否则, 如果 n 是偶数, 则将 n 除 2, 并赋给 n ; 否则, 将 $3n + 1$ 赋给 n . 不断循环, 只到 $n = 1$, 才停止计算, 并输出一句话: “运算成功”. 这个例子是为了检验数论中的一个简单的定理.

● 0011-1

例 2.4 编写一个用二分法求非线性方程根的函数, 并求方程

● 0011-2

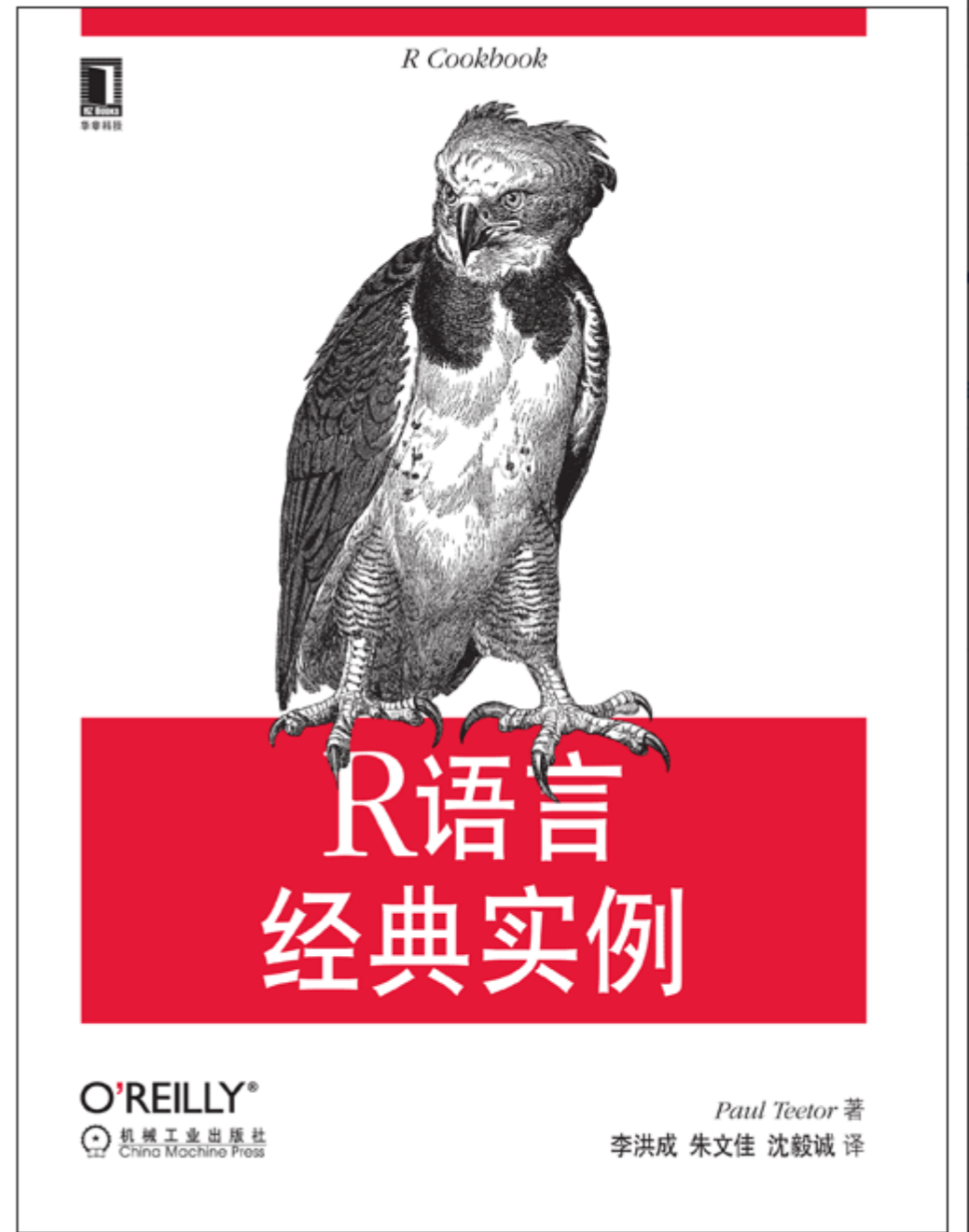
$$x^3 - x - 1 = 0$$

在区间 $[1, 2]$ 内的根, 精度要求 $\varepsilon = 10^{-6}$.

二分法计算过程如下: 取中点 $x = \frac{a+b}{2}$, 若 $f(a)$ 与 $f(x)$ 异号, 则置 $b = x$; 否则 $a = x$. 当区间长度小于指定要求时, 停止计算.



2.3、5.4、5.5
例子5-6、5-8




第四章

INTERACTIVE COURSE

Case Study. Exploring Baseball Pitching Data in R

Start Course For Free | Play Intro Video | Bookmark

4 hours | 14 Videos | 69 Exercises | 7,934 Participants | 5,750 XP



The image shows a course card for 'Exploring Baseball Pitching Data in R'. It features a blue background with a white hexagonal icon on the right containing a folder and a document, with the R logo at the top. The text 'INTERACTIVE COURSE' is in yellow, and the title is in white. Below the title are three buttons: 'Start Course For Free' (yellow), 'Play Intro Video' (blue with a play icon), and 'Bookmark' (blue with a bookmark icon). At the bottom, course statistics are listed in white text.

提交方式和上节课一样!


<https://www.datacamp.com/courses>

INTERACTIVE COURSE

Intermediate R: Practice

Start Course For Free | Bookmark

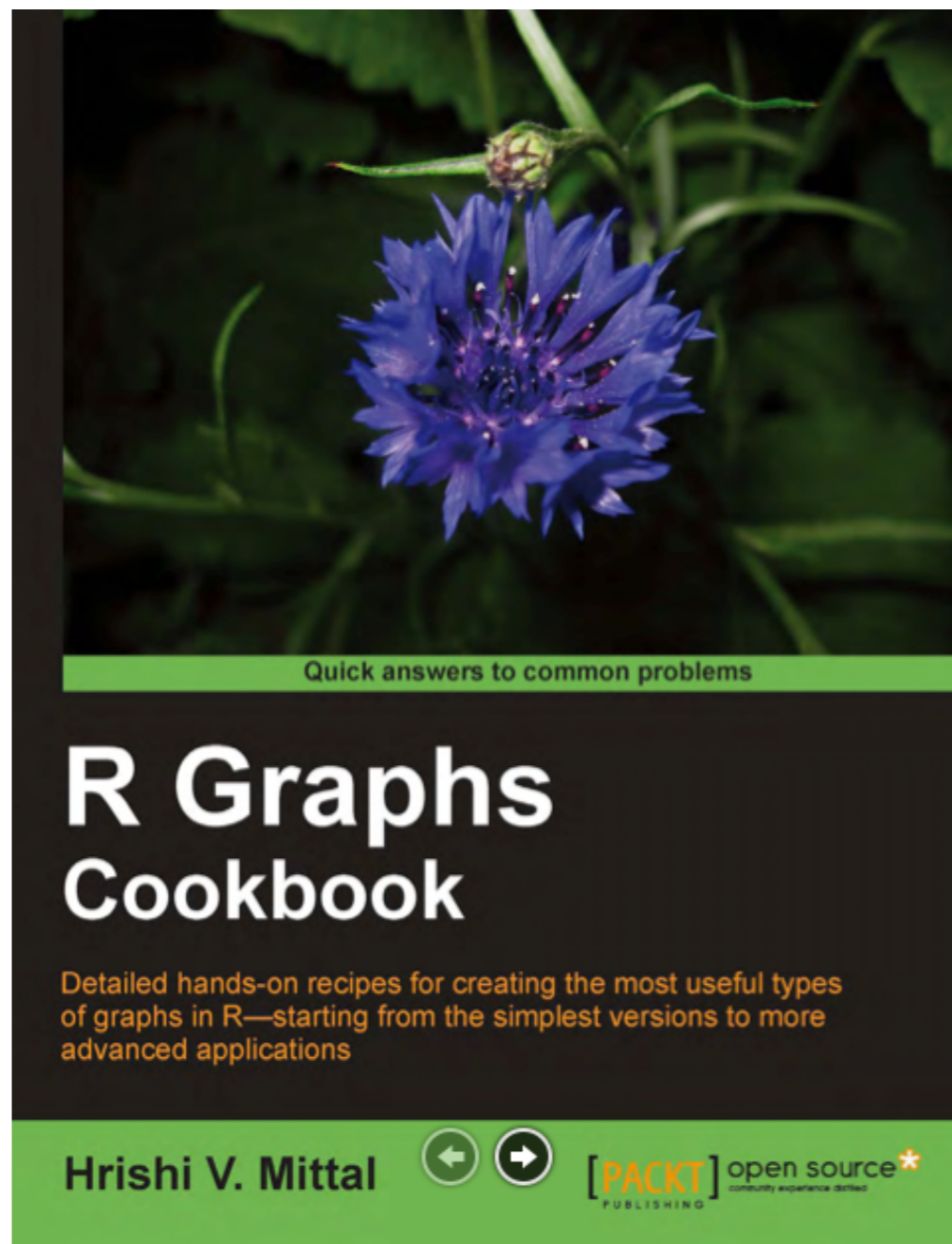
4 hours | 0 Videos | 52 Exercises | 58,845 Participants | 4,800 XP



The image shows a course card for 'Intermediate R: Practice'. It features a blue background with a white hexagonal icon on the right containing a computer monitor and a large R logo, with the R logo at the top. The text 'INTERACTIVE COURSE' is in yellow, and the title is in white. Below the title are two buttons: 'Start Course For Free' (yellow) and 'Bookmark' (blue with a bookmark icon). At the bottom, course statistics are listed in white text.



第三、六章



第一、二章

- 模拟产生100个学号 (1300022001到1300022100)
- 模拟产生三个科目的成绩，要求第一科最大值99，最小值70；第二科平均值81， $sd=7$ ，最大值100；第三科平均值83， $sd=18$ ，最大值100
- 把学号和三科成绩组成一个数据框，显示数据框内容
- 求每个学生的总分、平均分
- 针对三科成绩、总分、平均分，分别做饼图、直方图、条形图，箱线图
- 分别用par和layout把多个图放在一个图中显示：同一个数据的不同类的图形，不同数据的同一类，不同数据的不同图形

- 某校测的19名学生的四项指标：性别、年龄、身高（cm）、体重（磅），具体见0016_student.CSV，要求：
 - * 绘出体重对于身高的散点图
 - * 绘出不同性别情况下，体重与身高的散点图
 - * 绘出不同年龄段的体重与身高的散点图
 - * 绘出不同性别和不同年龄段的体重与身高的散点图
-
- 0016_height01.txt, 画直方图
 - 0016_height02.txt, 画箱式图
 - 0016_marriage.txt, 画散点图
 - 0016_language.txt, 画条形图（母语和日常使用）
 - 0016_language.txt, 画饼图（世界主要语种使用人数比例）

- 从0017_grade.csv中读取两班成绩
- 计算每个班级的均值和标准方差
- 计算每个人的标准化成绩，添加到数据中，写到0017grade.txt中
- 分别画出来两班成绩和标准成绩的箱线图
- 在一张图中画出两班成绩和标准成绩的箱线图

```
plot(rnorm(1000),col="red")
```

- 使用上面的语句，练习颜色的各种表示方法
 - 使用Par和layout函数，分别实现不同颜色的多个图形组合，2*2，3*3，1*1*2*3等
-

课件第12页，citysales.csv

- 输入现有代码，看显示结果
- 用rainbow、top.colors、cm.colors、gray、terrian.colors替换heat.colors，看执行效果
- 练习课件第23页的颜色参数
- 添加图例

cityrain.csv

- 用不同颜色画出不同城市的线图
- 用不同符号画出不同城市的线图
- 用不同颜色画出不同城市的散点图
- 用不同符号画出不同城市的散点图
- 分别加上图例
- 用par和layout把前面四个图放在一张图中，分别为
 $2*2$ ， $1*4$ ， $1+2+1$

INTERACTIVE COURSE

Data Visualization in R

[Start Course For Free](#) [Bookmark](#)

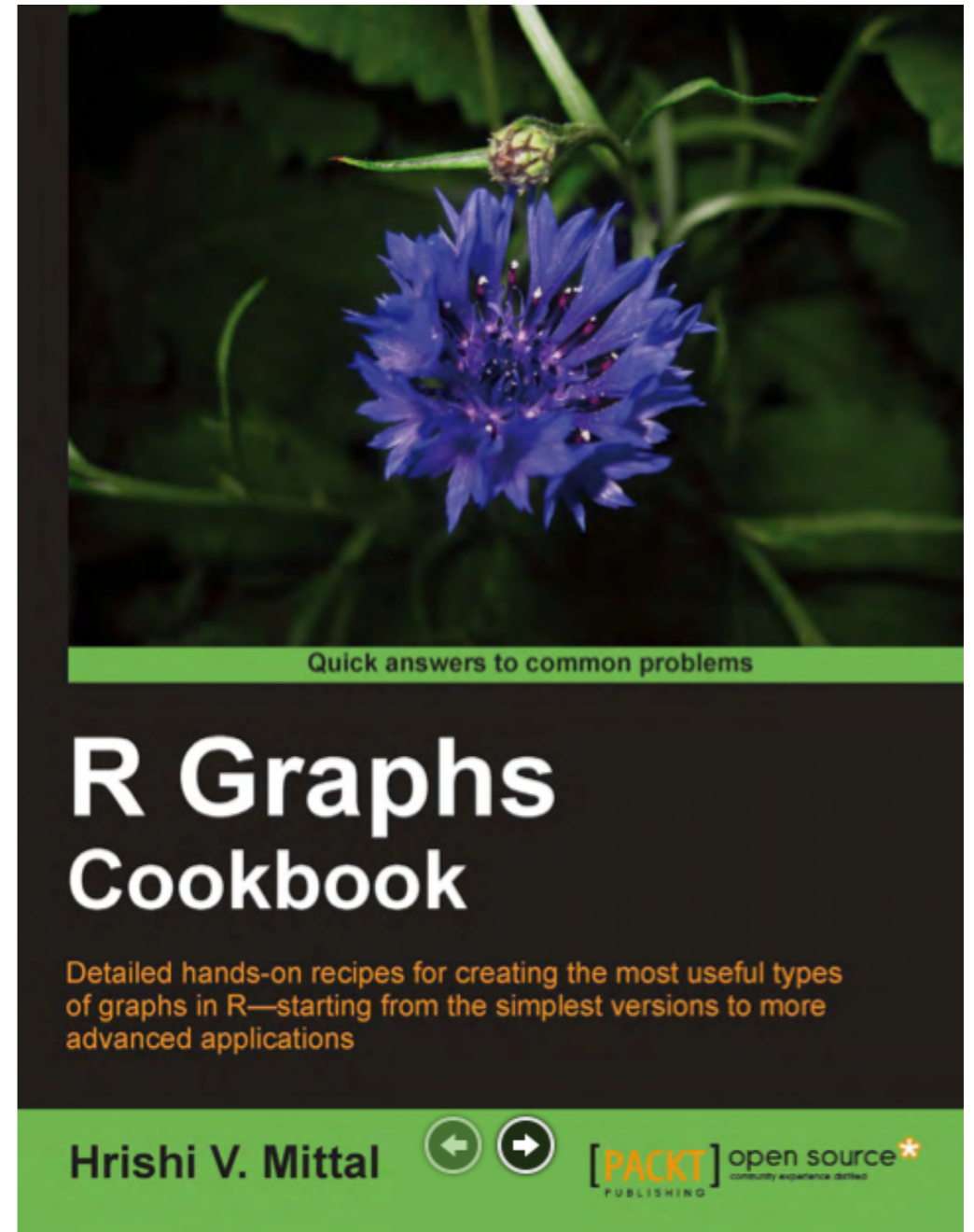
🕒 4 hours ▶ 15 Videos <> 60 Exercises 👤 49,031 Participants 📊 5,250 XP

提交方式和上节课一样！

<https://www.datacamp.com/courses>



第3、6章



第1-6章：看完！！

- `gdp_long.txt`
 - 做折线图（网格、特殊线，图例的不同位置）
 - 条形图（正常、堆积、横向、颜色宽度等、显示数字、误差线）
-
- `cityrain.csv`
 - 做折线图（边界标注，`slide`，`mar`和`bty`的含义）

INTERACTIVE COURSE

Visualizing Time Series Data in R

Start Course For Free

Bookmark

🕒 4 hours ▶ 11 Videos <> 45 Exercises 👤 10,569 Participants 📊 3,550 XP

提交方式和上节课一样！

<https://www.datacamp.com/courses>

课堂测试复习

- 1、数据分析包括几个过程？
- 2、简单描述R语言优点。
- 3、填写下表：

功能	函数	功能	函数
安装vcd包		使用maps包	
显示当前工作目录			demo(image)
帮助		查看函数mean()的示例	
	source()		sink()
	?c		dev.off()

- 4、写出命令，实现如下功能：创建一个从1到9的向量并赋给x，将1付给y，并写出x+y的结果。
- 5、RStudio主界面一般有几个功区？.Rprofile、.Rdata、.Rhistory文件分别有什么作用？

- 1、填写下表：

功能	函数	功能	函数
1:4-1		sep(1.3, 5, 1)	
rep(1:2, 3)		x <- c(1,4,6,7); x[-2]; x*x;	
z <- c(1,5,3,8,1) z[z>3], order(z)		y <- matrix(1:10, nrow=2) y[2, c(4,5)]	
	sort/sort.list()/order		which.min()/which.max()

- 2、请列举R中常见的数据类型数据结构，描述矩阵和数据框的区别？
- 3、使用rep()函数构建一个向量x。向量是由5个1，3个2，4个3和2个4组成的。
- 4、写出命令：1) 创建一个矩阵mat,矩阵的值为1-20之间的整数，四行五列，按行排列
2) 获取矩阵的第三行第四列的元素值 3) 获取矩阵第一行的所有值。
- 5、写出命令：1) 创建字符串向量names,元素值为"zhang san"、"li si"、"wang wu",创建数值型向量scores,元素为70、80、90，创建字符串向量levels,元素值为"C""B""A"; 2)将levels转换成因子类型 3) 创建数据框exam,其列向量为names、scoes、levels; 4) 根据列名称，提出第一列和第三列。

- 1、10位同学的姓名、性别、年龄、身高、体重数据如下：

Name	Sex	Age	Height	Weight
Alice Harden	F	13	56.5	84.0
Sandy Muller	F	11	51.3	50.5
Sharon Wilshere	F	15	62.5	112.5
Tammy Wenger	F	14	62.8	102.5
Alfred Ferguson	M	14	69.0	112.5
Duke klopp	M	14	63.5	102.5
Guido Conte	M	15	67.0	133.0
Robert Mourinho	M	12	64.8	128.0
Thomas Bryant	M	11	57.5	85.0
William Curry	M	15	66.5	112.0

参见 *RIA* 的 93 页
5.3 的例子 5-6

- 1) 根据以上信息构造一个数据框，数据框的名称为 `stuinfos`
- 2) 分别计算全体学生年龄、身高、体重的和、平均值、标准差，并且设置数据的输出格式为小数点后两位
- 3) 依姓氏和名字对数据集进行排序：
- 4) 将上述数据写成(`write.table()`)一个纯文本的文件（文件名为 `class.txt`），并用函数 `read.table()` 读取文件中的身高和体重数据

1. 使用鸢尾花数据iris

- 1) 先用names()观察其结构，然后用花瓣长度和宽度做散点图
- 2) 在plot函数里面添加细节。修改点的形状和颜色由白色空心圆换成红色雪花；修改坐标轴名称并添加标题 "relationship between width and length of Iris petal"。

2. 使用 airquality 数据

- 1) 绘温度 Temp 直方图，加一个横坐标"Temperature",加一个标题"The Distribution of Temperature"
- 2) 频数变频率，并设置颜色为绿色
- 3) 四幅图放在一个面板里，两个一排。并使用MASS包的trueHist函数画出频率直方图：
 - 第一幅图，airquality里温度变量的直方图（频数）
 - 第二幅图，airquality里该变量的直方图（频率）并添加密度曲线，填充红色
 - 第三幅图，airquality里风速变量的直方图（频数）
 - 第四幅图，airquality里该变量的直方图（频率），并添加密度曲线，填充蓝色

3. 使用mtcars里的mpg做箱图

给箱图添加坐标轴：x轴为"Number of Cylinders"，y轴为="Miles Per Gallon"标题"Car Milage Data"。根据不同cyl变量下mpg的箱线图，并添加x轴"Number of Cylinders",y轴"Miles Per Gallon"

4. 按要求作图：

- 1) 创建字符向量colors,元素为"green","orange","brown";创建字符向量months,元素为"一月","二月","三月","四月","五月";创建字符向量regions,元素为"东部地区","西部地区","南部地区";创建矩阵values,元素为值2,9,3,11,9,4,8,7,3,12,5,2,8,10,11，要求3行5列
- 2) 使用矩阵values创建推叠的条形图，添加标题为"总收入"，x轴名称为"月份"，y轴名称为"收入"，条形图的标签为字符向量months(使用names.arg参数)，推叠台型图的颜色设置为创建的字符向量colors
- 3) 添加图例，内容为字符向量regions，分别对应条形图中的三种颜色

练习

INTERACTIVE COURSE

Introduction to Importing Data in R

[Start Course For Free](#) [Bookmark](#)

🕒 3 hours ▶ 11 Videos <> 42 Exercises 👤 138,718 Participants 📖 3,550 XP

提交方式和上节课一样！

<https://www.datacamp.com/courses>

谢谢!

孙惠平

sunhp@ss.pku.edu.cn