

课程复习



课堂测试时间

- 1、写一个函数，输入有三个参数 x 、 y 、 z ，该函数计算 x 、 y 与 z 的和与乘，返回值是包括和与乘的一个向量。
- 2、写出完成如下功能的语句：从文件名为“testing01.txt”的文件读取一个数据框TT1（分隔符为*，包括header），然后将该TT1以追加模式写到“testing02.csv”中。
- 3、分别用if-else、ifelse、switch写出完成如下功能的语句： x 大于0小于等于10时， y 等于 x 的 $1/2$ ； x 大于10小于等于20时， y 等于 x 的平方根； x 大于20小于30时， y 等于1到30的和； x 等于30时， y 等于20到30的和减去10到20的和。
- 4、分别用while、repeat、for写出完成如下功能的语句：随便赋给 i 一个正整数，如果 i 是偶数，将 i 除以2，并赋给 i ，否则将 $3i + 1$ 赋给 i ，不断循环，直到 $i = 1$ ，停止。

- 流程控制: *if-else; ifelse;*
- 循环控制: *repeat; for; while;*
- 数据输入输出函数: *read.table(); write.table(); read.csv(); write.csv();*
- 函数: *function();*
- *apply*族函数: *lapply(); sapply(); vapply(); tapply();*

课堂测试复习

安装*Rmarkdown*和*Knitr*

- 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能区？.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、按照要求写出相应的命令：(1). 创建向量x，其元素为1、2、3，创建向量y,其元素为"a""b""c",定义一个由x, y组成的数据框df; (2). 将数据框df与一个有且仅有一个数值列z，数值列元素为3、2、1的数据框进行列连接; (3). 按如下方式：data.frame(x = 10, y = "d",c=10))定义一个数据框df2,然后将其与df进行连接。
- 2、按照题目要求写出命令：(1). 创建矩阵a_mat,矩阵元素为1到12之间的整数，创建矩阵b_mat,矩阵元素为1到25之间的间隔为2的整数; (2).计算a_mat和b_mat的内积和外积; (3).计算a_mat的转置、逆矩阵和特征值、特征向量。
- 3、按要求写出命令：(1). 创建字符串str,其值为I'm Shouting，将str所有的字母大写; (2). 截取str的第3个到最后一个字符; (3). 以空格为分割符，分割str为字符串向量strs 4、使用paste函数将strs与向量c(1,2)连接，指定参数分割符为"-".
- 4、求出当前的日期与1970-1-1之间相差的秒数、分钟数、小时数、天数和周数。
- 5、写出产生50个学号 (DA-2017-01到DA-2017-50) 的语句。
- 6、test <- c(20, 9, 6, NA, 99, 88, NA, 46)，写出完成如下功能的语句，将test按照行优先行程一个4列的矩阵Mt，检查哪些位置是NA，求每一列和每一行的总和，并将行列的总和增加到Mt中。

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, 完成三个课堂练习 (估计30-60分钟)

`install_course_github("pkusdatanalysis","C1_R_Basis")`

- 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



Learn R, in R.

swirl teaches you R programming and data science
interactively, at your own pace, and right in the R console!

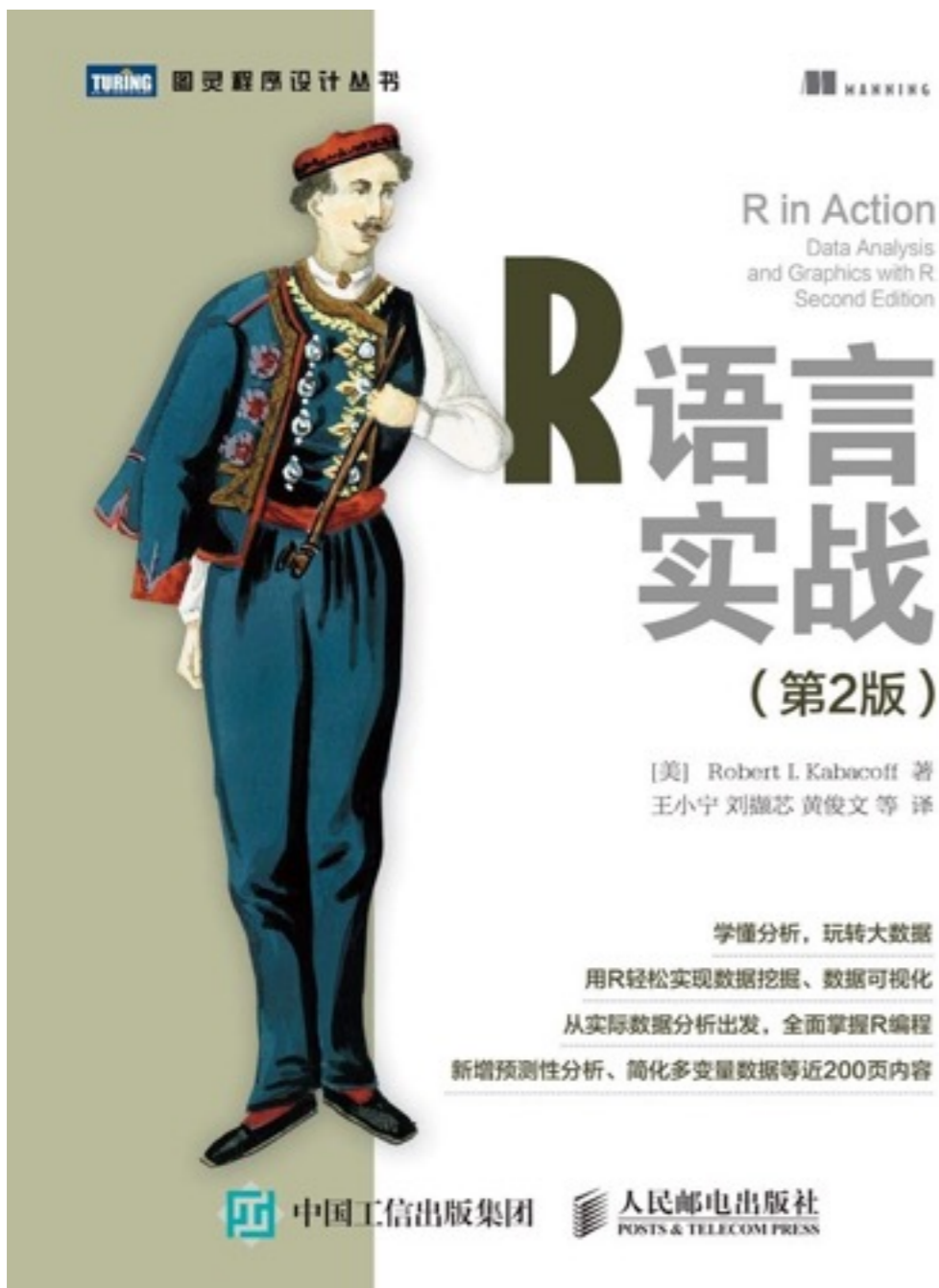
```
install_course_github("pkussdatanalysis","C2_Data_Structure_01")
```

The screenshot shows the DataCamp website interface. At the top, there is a navigation bar with the DataCamp logo and the tagline 'We're hiring!'. The navigation menu includes 'Dashboard', 'Courses', 'Pricing', 'Business', 'Community', a notification bell, and a user profile 'sunhp@ss.pk...'. Below the navigation bar is a search bar with a magnifying glass icon and a right arrow. To the right of the search bar are two dropdown menus: 'All Technologies' and 'All Topics'. The main content area displays a grid of course cards. The first card, 'Introduction to R', is highlighted with a red border. It features a blue header, a description: 'Master the basics of data analysis by manipulating common data structures such as vectors, matrices and data frames.', a circular profile picture of Jonathan Cornelissen, and his name and title: 'JONATHAN CORNELISSEN, Co-founder and CEO of DataCamp'. Below the name is a small circular icon. The other cards in the grid include 'Intro to Python for Data Science' by Filip Schouwenaars, 'Intermediate R' by Filip Schouwenaars, 'Intermediate Python for Data Science', 'Data Visualization with ggplot2 (Part 1)', and 'Introduction to Machine Learning'.

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

The screenshot shows two chapters from the 'Introduction to R' course. The first chapter, 'Intro to basics', is marked as 100% complete and includes a description: '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!'. Below the description are links for 'Hide Chapter Details' and a 'Continue Chapter' button. The second chapter, 'Vectors', is also marked as 100% complete and includes a description: '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.'. Below the description are links for 'View Chapter Details' and a 'Continue Chapter' button.

The screenshot shows a user profile page for 'sunhp'. At the top is a circular profile picture placeholder. Below it, the username 'sunhp' is displayed. The profile statistics are: 3100 XP Earned, 0 Courses Completed, and 32 Exercises Aced. Under the heading 'sunhp's Skills', there are two skill progress bars: 'Programming' with 3000/48450 XP and 'Importing & Cleaning Data' with 100/34750 XP.



第二章



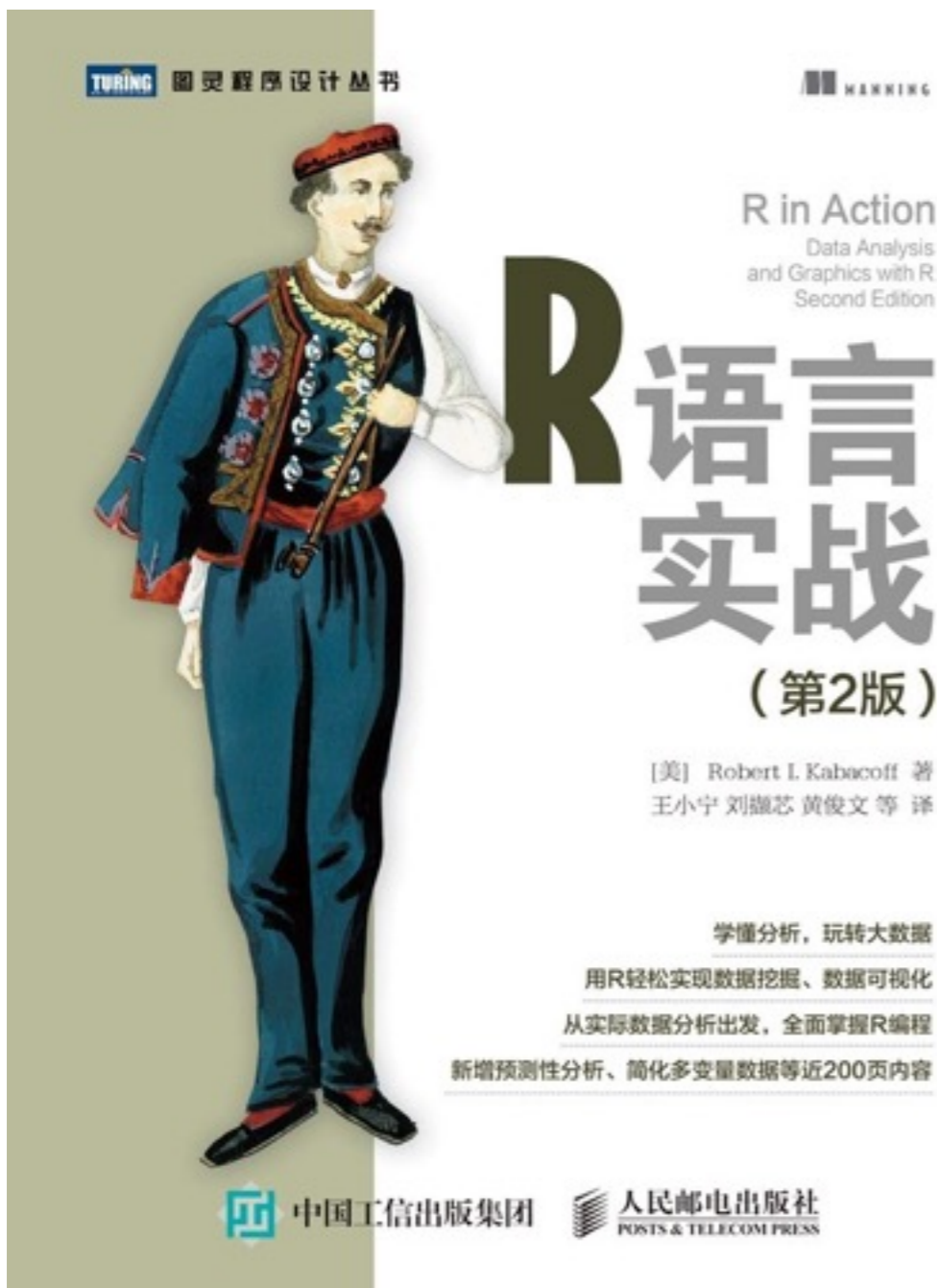
第一章到第四章



Learn R, in R.

swirl teaches you R programming and data science
interactively, at your own pace, and right in the R console!

```
install_course_github("pkusdatanalysis","C3_Data_Structure_02")
```



第四章和第五章



第五章、第六章、第七章

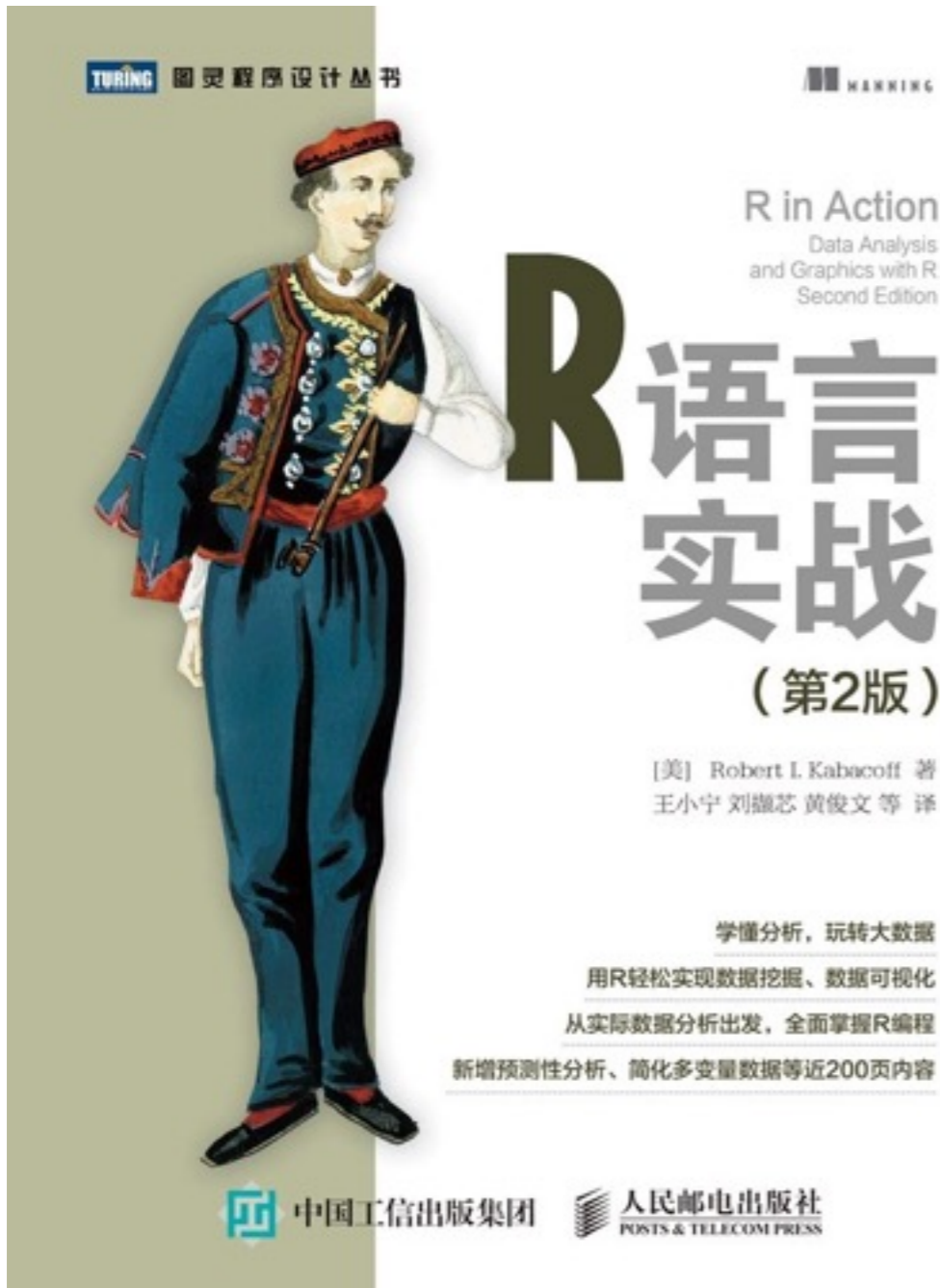
- 完成课后大作业0001
- 按照要求提交



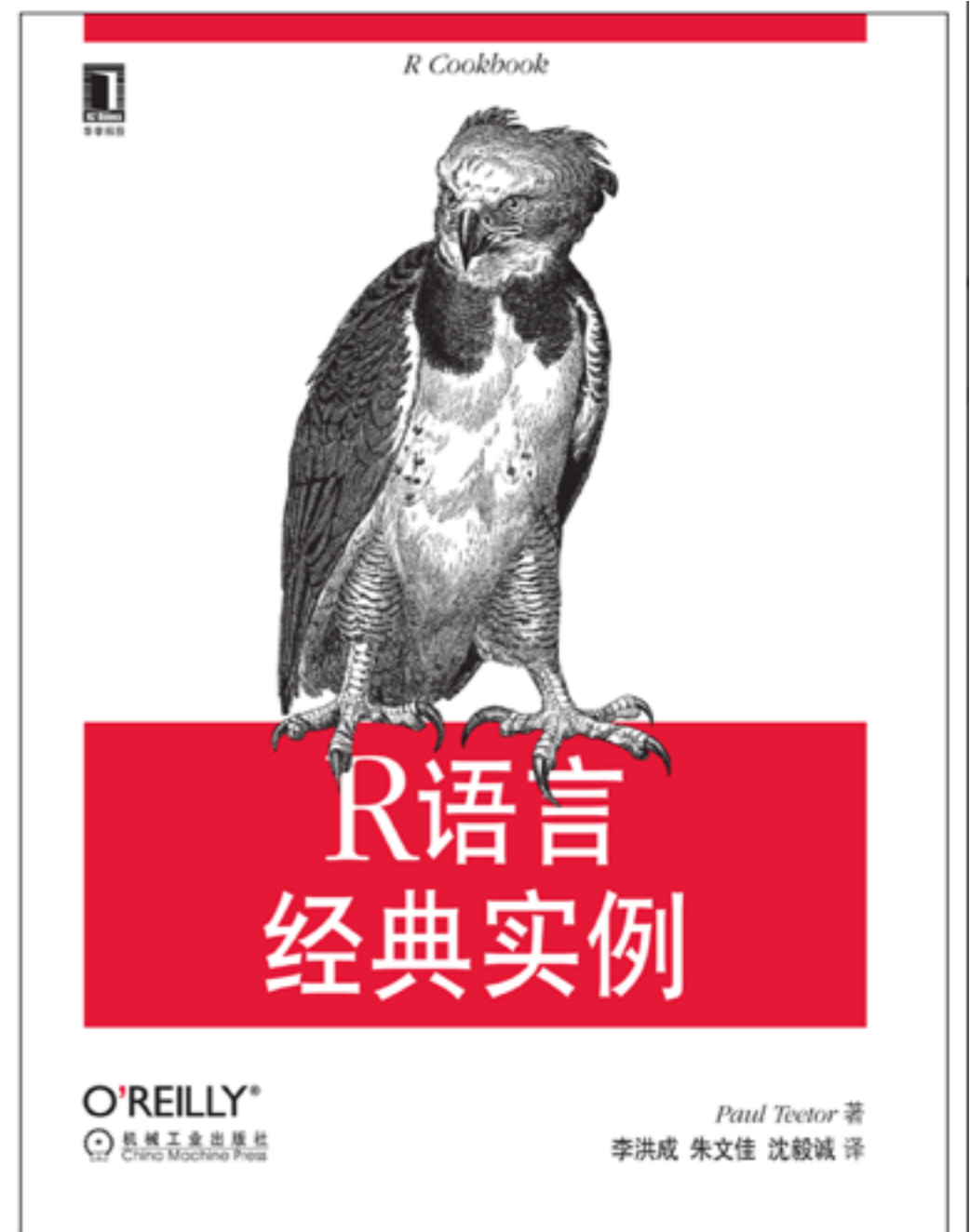
Learn R, in R.

swirl teaches you R programming and data science
interactively, at your own pace, and right in the R console!

```
install_course_github("pkusdatanalysis","C4_Grammar_01")  
install_course_github("pkusdatanalysis","C5_Grammar_02")
```

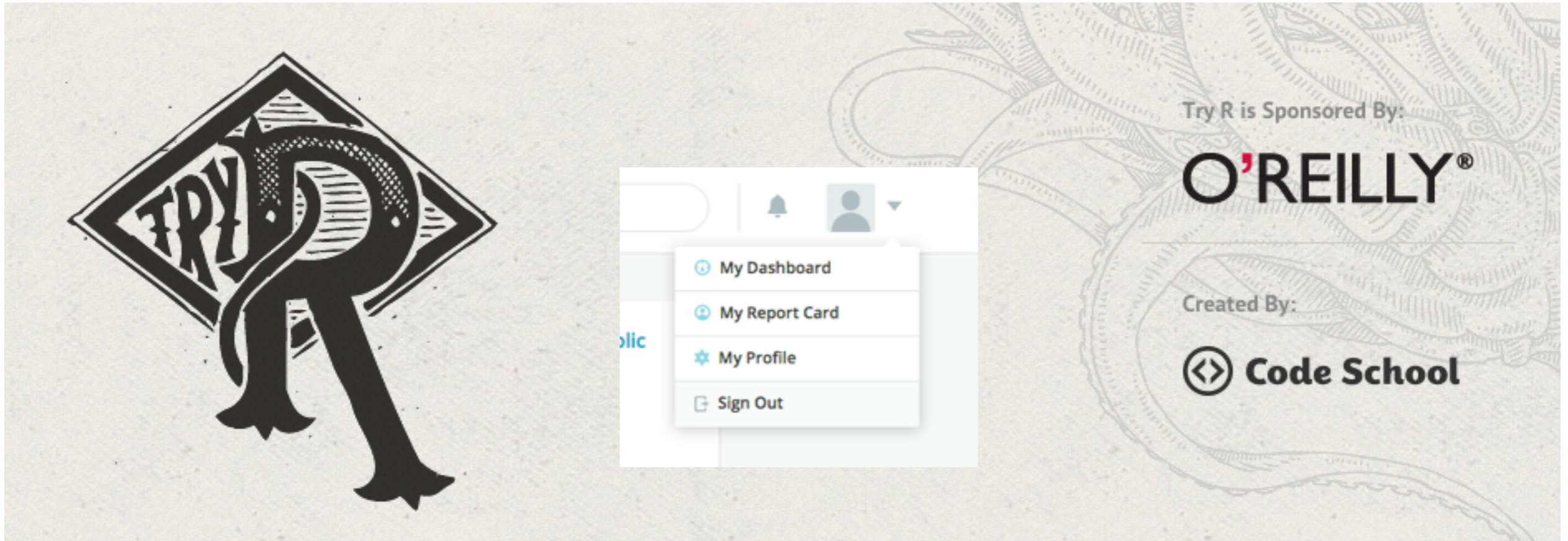


2.3、5.4、5.5
例子5-6、5-8



第四章

<http://tryr.codeschool.com/>



code_cademy

 Code School

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.

● 0015-1

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

● 0015-2

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

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.

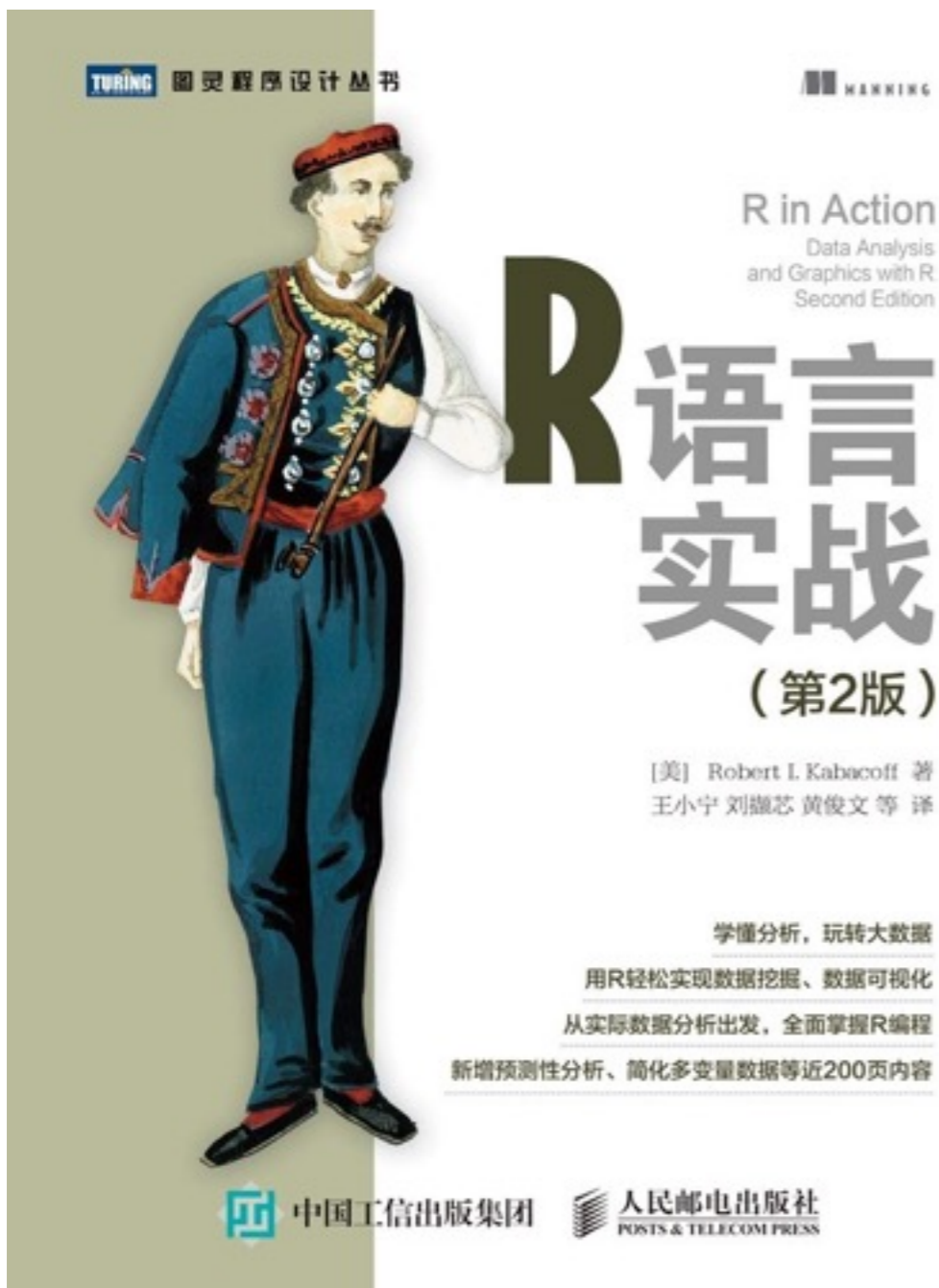
- 有48位应聘者应聘公司某职位，公司为这些应聘者的15个指标打分，分数从0到10，0最低，10最高，具体分数见0016_applicant.csv，公司要录用其中优秀的8名，写一个程序来选择：
 - 如果各个指标权重是同样的，怎么选择
 - 随机产生15个指标的一个权重表，按照权重进行选择
 - 随机制定3个指标，这三个指标不能低于8，如何选择

- *0017_test.txt*文件中包含两个班学生的6门课程成绩，要求：
 - 在最后增加一列班级 (*class*)，前35个为1班，后边35个为3班，中间的为2班，班级形式为“01, 02, 03”
 - 在最前面增加一列学生序号 (*number*)，序号形式为“*student_class_number*”
 - 统计成绩缺失的信息，包括哪些学生成绩出现缺失，每个课程有多少学生成绩有缺失，每个班级有多少学生成绩有缺失，使用本班级该课程的平均分作为该缺失的值
 - 统计每个课程的平均分、每个班级的平均分，每个人的总分
 - 求1班每科成绩均大于80的学生集合，2班每科成绩均小于90的学生集合
 - 给每个学生评定优、良、中、差四个等级，按照每个班级划分，按照所有班级一起划分，找出两种划分中等级不同的学生

- 从 `0018_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 `0014grade.txt` 中

看教材例子 5-6,
然后完成练习 `0017` 和 `0018`
和大作业 `0002` 一起提交

练习



- 完成课后大作业0002
- 按照要求提交