

《R和Ruby数据分析之旅》

图书基本信息

书名：《R和Ruby数据分析之旅》

13位ISBN编号：9787115303707

10位ISBN编号：7115303703

出版时间：2013-3

出版社：人民邮电出版社

作者：（新加坡）Sau Sheong Chang

页数：224

译者：钱昊,刘熙

版权说明：本站所提供下载的PDF图书仅提供预览和简介以及在线试读，请支持正版图书。

更多资源请访问：www.tushu000.com

《R和Ruby数据分析之旅》

内容概要

如果你对万事万物的运行方式充满好奇，这本有趣的《R和Ruby数据分析之旅》会帮你找到日常生活中某些问题的真正答案。借助基本的数学方法，并使用Ruby和R语言做一些简单的编程工作，你就能学会如何对问题建模，并找出解决方案。

要阅读《R和Ruby数据分析之旅》，对编程有基本了解就足够了。书中会对Ruby和R做一个简明的介绍，之后你就可以随作者一起，对多个领域的问题进行探索，并学习如何收集、处理、模拟和分析可用的数据。你将学会如何通过简单的程序和常识性的逻辑，以不同的视角看待日常的事物。读过本书，你便可以开启自己的探索和发现之。

下面是本书将要探究的一些问题：

- 确定一间拥有70名员工的办公室需要有多少个卫生间隔间才能满足需要；

- 挖掘自己的电子邮件，了解自身特定的邮件习惯；

- 借助简易的音频和视频录制设备来计算自己的心率；

- 创建一个人造的社会，并分析其中的行为模式，找出某些特定因素是如何影响我们的现实社会的。

《R和Ruby数据分析之旅》

作者简介

Sau Sheong Chang，是新加坡惠普实验室的应用研究主管，已经从事软件开发工作17年，主要专注于与云计算和数据相关的系统。他在当地开发者社区颇有名望，并且作为演讲者活跃在众多技术大会上。在本书之前，他出版过《Ruby on Rails Mashup Projects》和《Cloning Internet Applications with Ruby》两部著作。

书籍目录

第1章 握住探险之鞭—认识Ruby

- 1.1 Ruby
 - 1.1.1 为什么用Ruby
 - 1.1.2 安装Ruby
 - 1.1.3 运行Ruby
 - 1.1.4 引用外部库
 - 1.1.5 Ruby基础
 - 1.1.6 一切皆对象
- 1.2 Shoes
 - 1.2.1 何为Shoes
 - 1.2.2 五彩的“鞋子” (Shoes)
 - 1.2.3 安装Shoes
 - 1.2.4 编写Shoes程序
- 1.3 小结

第2章 走进《黑客帝国》—认识R

- 2.1 R的简介
- 2.2 使用R
 - 2.2.1 R控制台
 - 2.2.2 源文件与命令行
 - 2.2.3 程序包
- 2.3 R编程
 - 2.3.1 变量和函数
 - 2.3.2 条件语句与循环
 - 2.3.3 数据结构
 - 2.3.4 导入数据
- 2.4 作图
 - 2.4.1 基本绘图
 - 2.4.2 ggplot2简介
- 2.5 小结

第3章 探讨办公室与卫生间问题

- 3.1 简单的情形
- 3.2 表示卫生间等事物
- 3.3 初次模拟
- 3.4 解释数据
- 3.5 第二次模拟
- 3.6 第三次模拟
- 3.7 最后一个模拟
- 3.8 小结

第4章 当一回经济学家

- 4.1 看不见的手
- 4.2 一个简单的市场经济体系
 - 4.2.1 生产者
 - 4.2.2 消费者
 - 4.2.3 一些使程序简化的方法
 - 4.2.4 模拟
 - 4.2.5 分析模拟结果
- 4.3 价格决定的资源分配

- 4.3.1 生产者
- 4.3.2 消费者
- 4.3.3 市场
- 4.3.4 模拟
- 4.3.5 分析第二个模拟的结果
- 4.3.6 价格控制
- 4.4 小结
- 第5章 对自己的电子邮件进行挖掘
 - 5.1 思想
 - 5.2 抓取和解析
 - 5.3 对电子邮件进行挖掘
 - 5.3.1 每月同一天的邮件数量
 - 5.3.2 邮件挖掘者
 - 5.3.3 每周同一天的邮件数量
 - 5.3.4 每月的邮件数量
 - 5.3.5 一天各小时的邮件数量
 - 5.3.6 交互
 - 5.3.7 交互中的比较
 - 5.4 文本挖掘
 - 5.5 小结
- 第6章 测一测我们的心脏
 - 6.1 跳动的心脏
 - 6.2 听诊
 - 6.2.1 自制数字听诊器
 - 6.2.2 从声音中提取数据
 - 6.2.3 生成心音波形
 - 6.2.4 找出心率
 - 6.3 血氧定量法
 - 6.3.1 自制脉搏血氧计
 - 6.3.2 从视频中提取信息
 - 6.3.3 生成心跳波形并计算心率
 - 6.4 小结
- 第7章 模拟结群飞行的鸟
 - 7.1 Boids的由来
 - 7.2 模拟
 - 7.3 Roid
 - 7.4 boid群的规则
 - 7.5 辅助规则
 - 7.6 规则的变更
 - 7.7 绕圈
 - 7.8 加入障碍物
 - 7.9 小结
- 第8章 是什么让世界运行不止
 - 8.1 生活很美好
 - 8.2 金钱
 - 8.3 性别
 - 8.3.1 生与死
 - 8.3.2 变化
 - 8.4 进化

8.4.1 我们将做哪些改变

8.4.2 实现

8.5 小结

关于作者

封面介绍

章节摘录

版权页：插图：2050项目是由圣菲研究所（Santa Fe Institute）、世界资源研究所（World Resource Institute）和布鲁金斯研究所（Brookings Institution）合资创办的。该研究在爱泼斯坦和阿克斯特尔卓越的著作（*Artificial Societies: Social Science from the Bottom up*）（Brookings Institution出版社 / MIT出版社）已经被详尽地涵盖。我在此完成的是使用Roids模拟程序，对爱泼斯坦和阿克斯特尔的一些思想进行再实现和细化。第7章中创造的roid拥有美好的生活。首先，它们是永生的，可以无忧无虑地在它们的世界中漫步。当然，你会偶尔往它们周围丢一些讨厌的障碍物，但麻烦仅此而已。它们不用吃饭，不用睡觉，而且因为它们不会死，所以它们不需要生儿育女来传宗接代。它们唯一要做的，就是成群地在无垠的空间里晃来晃去。这真是个名副其实的伊甸园。但那并不是一个好的世界，而且作为模拟而言，它很快会给人相当无趣的感觉。我们将要做的是，从这一基本的模拟开始，逐渐地增加roid以及它们所处世界的行为。当我们执行模拟时，我们将得到不同类型的数据，并进行若干分析，随之提出一些问题，并期望能找到正确的解答。让我们从伊甸园中蛇所做的事情“开始：为roids提供食物。我将从第7章的最基本的Roids模拟出发，加入一些有趣的特征，来创建我们的新模拟世界，我将其称为乌托邦（Utopia）。在这个新的模拟中，每一个roid有其能量等级，在被创造时随机赋予。每经过一个时间点（tick），每个roid都会损失一些能量，如果roid失去了它所有的能量，它就会死亡。要想存活，每个roid都要不断地寻找食物，吃下食物以补充能量。这是从绝对不朽层面的一个相当大的坠落，但roid仍是潜在不朽的，只要它们始终能找到足够的事物，它们就会永远存活。自然，我们要在乌托邦中加入食物。为此，我们要创建随机数量的食物，并将其散播到世界当中。经过随机的时间间隔，更多的食物会在别的随机的地点被创造出来。要建立这样的模拟，也是为了让贯穿本章的模拟能简单些，我们需要精简第7章中的基本代码。这实际意味着我们将把roids.rb分拆成一些更小的文件。主体文件将被称为utopia.rb，而剩下的将依据逻辑打散成碎片。

《R和Ruby数据分析之旅》

编辑推荐

如果你对万事万物的运行方式充满好奇，《R和Ruby数据分析之旅》会帮你找到日常生活中某些问题的真正答案。借助基本的数学方法，并使用Ruby和R语言做一些简单的编程工作，你就能学会如何对问题建模，并找出解决方案。

《R和Ruby数据分析之旅》

名人推荐

“这是每一位有好奇心的黑客的梦想！它会引导你把令人生厌的开发项目变得妙趣横生，同时学会R和Ruby语言。这本书比我多年来看过的其他任何一本都更能抓住真正的心。”——Derek Sivers，CD Baby的创始人（sivers.org）

《R和Ruby数据分析之旅》

精彩书评

1、说实的，这本书的组织结构，逻辑上非常之简单。第一章简单介绍ruby，第二章简单介绍R，真得只是简单介绍。从第三章到第八章，都是用实际案例来介绍如何分析案例中的数据，内容格式也很固定，先是用ruby实现对案例的模拟，然后用R分析产生的数据结果，最后是总结。从书的内容和深度来说，这是本适合了解的书，书里面提到的不必一定是R&R，Ruby完全可以换成python，R也完全可以是matlab或者是SPss。但是，这本书最漂亮的是想像，在写了很多年的程序以后，一个程序员还能够保持这种对于世界的单纯探索的好奇心，不知道在我们的国度什么时候能够有这样的环境能够产生这样的程序员。前几个案例都是中规中矩，对于卫生间的随机应用建模，简单的经济生产模型，用安然的案例来分析邮件(虽然话外的意思很多)。真正令我特别感觉震惊的后面三章，let's get a wav, let's get a video，然后就可以测心率了，完全是需要想像的。后面对于boids游戏的改进，让我们看到了对于世界运行问题的好奇心点然后的一种美，其实这个模型还可以演进的更复杂一些，把环境加进去，把前面的经济的案例也加进去，或许还能够跑得更多。对于数据分析，机器学习等方面的各种书都有很多，有薄有厚，用到的语言也各种各样，但是对于我来说，这本书最打动我的是其漂亮的想像力。子曰：“尔爱其羊，我爱其礼。”

章节试读

1、《R和Ruby数据分析之旅》的笔记-第73页

勘误

```
ggplot(data = df) + scale_shape_manual(name="Type", value=c(2,3,4)) +
```

应该为

```
ggplot(data = df) + scale_shape_manual(name="Type", values=c(2,3,4)) +
```

2、《R和Ruby数据分析之旅》的笔记-第69页

我觉得对于图3.2第一次模拟的结果，是不是我这个图更合适点呢？

当超过310人的时候，可以看出队列长度急剧上升。当只有一个卫生间的时候，人数和队列的关系这个好像更真实点：

3、《R和Ruby数据分析之旅》的笔记-第65页

缺少常量DURATION定义；
添加：DURATION=540

4、《R和Ruby数据分析之旅》的笔记-第65页

模拟脚本1的代码片段:
unless queue.empty?
 restroom.enter queue.shift
end
应改为:
until queue.empty?
 restroom.enter queue.shift
end

5、《R和Ruby数据分析之旅》的笔记-第62页

勘误:
def enter(person)
 unoccupied_facility = @facilities.find {|f| not f.occupied?}
 if unoccupied_facility
 unoccupied_facility.occupy(person)
 else
 @queue <<< person
 Person.population.delete person #加入此行
 end
end

6、《R和Ruby数据分析之旅》的笔记-第75页

《R和Ruby数据分析之旅》

我第二次模拟的图3.5，可以看出来，设备超过12时，不用排队

《R和Ruby数据分析之旅》

版权说明

本站所提供下载的PDF图书仅提供预览和简介，请支持正版图书。

更多资源请访问:www.tushu000.com