

第七届中国 R 语言会议（上海会场）



自由的统计语言

主办：

华东师范大学金融与统计学院

统计之都

Mango Solutions

赞助：

优酷土豆集团

2014 年 12 月 6 日 - 7 日

R 语言简介

R 是一个有着统计分析功能及强大作图功能的语言环境和软件系统，由新西兰奥克兰大学统计系的 Ross Ihaka 和 Robert Gentleman 共同创立。R 语言可以看作是由 AT&T 贝尔实验室所创的 S 语言发展出的一种方言。

R 是在 GNU 协议 General Public Licence 下免费发行的，它的开发及维护现在则由 R 开发核心小组 *R Development Core Team* 具体负责，这个团队的成员大部分来自大学机构（统计及相关院系），包括牛津大学、华盛顿大学、威斯康星大学、爱荷华大学、奥克兰大学等。除了这些作者之外，R 还拥有一大批贡献者（来自哈佛大学、加州大学洛杉矶分校、麻省理工大学等），他们为 R 编写代码、修正程序缺陷和撰写文档。

R 的功能很大程度上是通过程序包（Package）来实现的，迄今为止，R 语言官网上的程序包数目已经超过 6000 个，广泛地覆盖了数据分析应用到的各类行业和领域。各种统计前沿理论方法的相应计算机程序都会在短时间内以软件包的形式得以实现，这种速度是其它统计软件无法比拟的。

在 KD Nuggets 于 2014 年做的“使用何种编程或统计类语言进行分析、数据挖掘及数据科学的工作”的调查中，R 以 49.0% 的得票率荣登榜首，力压 SAS、Python 和 SQL (www.kdnuggets.com/polls/2014/languages-analytics-data-mining-data-science.html)。连续三年位居榜首。

目前，几乎所有的西方大学与研究机构、以及越来越多的金融机构、制药公司、高科技企业都使用 R。R 的灵活性、开放性以及业界最广泛的支持是其不断完善和发展的根本原因，随着 R 越来越被学术界及业界认可，它也将在数据分析和统计建模中发挥越来越大的作用。

华东师范大学金融与统计学院简介

华东师范大学金融学科、贸易学科、概率论与数理统计学科是国内最早培养国际金融、国际贸易、保险精算、概率论与数理统计高级人才的重要基地之一。在我国金融改革、开放、发展力度不断加大，国家和上海面临新一轮重大发展机遇的大背景下，根据国家和上海中长期发展规划及上海建设国际金融、贸易等四大中心的发展战略，为适应现代市场经济对于高素质复合型金融、贸易、保险精算和统计人才的迫切需求，华东师范大学在原有金融学科、贸易学科、概率论与数理统计学科的基础上，于2007年在国内高校中率先组建金融与统计学院，以培养既具有扎实的经济学与数理基础、又擅长专业应用的复合型人才，并希望通过结构创新和资源整合，来奠定华东师范大学在金融、贸易、概率论与数理统计和保险精算领域走向国内领军地位的基础。

学院下设五个系、一个研究院、一个研究中心和一个研究所：金融系、金融工程系、国际贸易系、风险管理与保险学系、统计与精算学系；华东师范大学应用统计科学研究院；华东师范大学人文社会科学重点研究基地——国际金融与风险管理研究中心和华东师范大学国际金融研究所。学院现设有六个本科专业：统计学（1983）、金融学（1984）、保险（1994）、国际经济与贸易（1999）、金融工程（2009）和经济统计（2012）；五个专业学位硕士点：工商管理硕士（MBA，金融管理方向，2006）、金融硕士（2010）、国际商务硕士（2010）、保险硕士（2010）和应用统计硕士（2010）；两个一级学科博士点：应用经济学（2010）和统计学（2010）（含八个二级学科博士点和八个二级学科硕士点）；两个博士后科研流动站：应用经济学（2012）、统计学（2012）。

目前学院教职员工85人，其中教授30人，博士生导师22人，“千人计划”国家特聘教授3人，长江讲座教授1人，紫江讲座教授5人，副教授30人，有博士学位的教师63人。

学院配备专业资料室、金融实验室、统计与精算实验室。金融实验室和统计与精算实验室拥有先进的金融软件、统计软件和精算软件供教学和科研使用。中国数学类核心期刊、中国数学会概率统计学会《应用概率统计》杂志编辑部设在本学院。

在学校的支持下、在国内外同仁关心和帮助下，在学院全体师生员工的共同参与下，我们将努力把金融与统计学院建设成为文理结合、学科交融、多元发展、国内一流、国际知名、开放与互动的研究型与应用型并重的学院。

统计之都简介*

“统计之都”（Capital of Statistics，简称 COS）网站成立于 2006 年 5 月，其主旨为传播统计学知识并将其应用于实际领域。纵观现今国内统计学理论和应用的发展，一方面我们不难发现统计学在应用领域的巨大潜力——现代管理、咨询、商业、经济、金融、医药、生物等等，无不需数据的力量，而另一方面我们也不得不承认，国内统计学的应用很大程度上受理论的制约——无论是应用界的人们对统计学基础理论知识的欠缺，还是学术界所研究的理论对应用领域问题的轻视。

“统计之都”网站便是基于这样的认识而创建的。我们希望，统计理论研究者能充分关注应用问题，而统计应用者也能正确把握统计学基本知识，将统计学这门应用学科真正的潜力开发出来。

“统计之都”为非赢利性质网站，但大力欢迎所有商界和研究领域的朋友与我们在实际应用问题上合作。我们的口号是：

中国统计学门户网站，免费统计学服务平台

我们怀着“十年磨一剑”的决心，要将“统计之都”创建成中国的统计学门户网站；我们抱着“己欲立而立人、己欲达而达人”的信条，要将“统计之都”以免费统计学服务平台的形式坚持办下去。我们希望“统计之都”在专业知识体系上有真正的王者风范，在面对用户需求时却又以谦恭的态度为大家服务。

想要获取最新的咨询和活动的通知，请关注我们的微信¹：



*统计之都网址：<http://cos.name/>

¹微信号：CapStat

优酷土豆集团简介

优酷土豆集团（NYSE:YOKU），专注于视频领域，是中国第一大视频应用和网络平台，旗下拥有中国排名第一和第二的视频网站优酷（www.youku.com）和土豆（www.tudou.com），以及合一影业。

优酷于2006年12月21日正式上线。一直以来，优酷以“世界都在看”为口号，坚持打造“阳光、真实、主流、有梦想”的品牌形象，现已成为中国互联网领域最具影响力、最受用户喜爱的视频媒体品牌。

土豆于2005年4月15日上线。多年来，土豆通过打造“青春、个性、自主、有趣”的品牌性格，号召“每个人都是生活的导演”，让用户能够充分发挥创造力和互动性，在土豆自由地创作、观看和分享视频节目。

优酷土豆集团拥有庞大的用户群、多元化的内容资源及强大的技术平台优势，为用户群提供最全、最多样的内容，帮助用户多终端、更便捷地观赏高品质视频，充分满足用户日益增长的互动需求及多元化视频体验。

目前，优酷、土豆的应用支持PC、手机、平板电脑、电视等多个终端，兼具UGC、PGC、版权、自制、电影五大内容形态，贯通视频内容制作、播出、发行三大环节，正在打造多屏文化娱乐生态系统，立志于成为全球华人最主要的视频来源，并分享快乐、智慧和感动，成就别人，实现梦想，传递正能量。

第七届中国 R 语言会议（上海会场）

会议指南

1. 日程安排

12 月 6 日	注册和报告	华东师大中山北路校区，科学会堂
12 月 7 日	报告和讨论	华东师大中山北路校区，科学会堂

2. 会议议程

12 月 6 日

时间	内容	嘉宾	主持人
09:00-09:20	会议致辞	汤银才	林祯舜
09:20-09:50	R 语言在推荐系统中的应用	刘思喆	
09:50-10:20	互联网金融大数据与风险	张云松	
10:20-10:45	茶歇		
10:45-11:15	数据科学中的 R 语言	肖凯	李舰
11:15-11:45	新药研发中的建模和模拟（M&S）	杨环	
午餐时间			
13:30-14:00	Lightning Talk	李颖、葛建辉等	张翔
14:00-14:30	一个数据分析师的入门与成长	任华	
14:30-15:00	网站实验设计框架与实战	陈丽云	
15:00-15:20	茶歇		
15:20-15:50	基于 R 的三维人脸绘图系统以及利用 R 实现 三维人脸预测人的性格	呼思乐	陈丽云
15:50-16:20	空气污染与 R	张兵	
16:20-16:50	利用 R 与 node.js 实现可视分析工具	周扬	

12 月 7 日

时间	内容	嘉宾	主持人
09:00-09:20	客户账户分析方法：用 R 实现建模	石宏飞	魏太云
09:20-09:50	基于 R 语言的投资监控平台之构建	邓一硕	
09:50-10:20	Growth Hacker 的 R 工具箱	张翔	
10:20-10:45	茶歇		
10:45-11:15	电商数据化运营中的网站竞争分析	肖嘉敏	刘思喆
11:15-11:45	R-web：大型多人在线数据分析协作平台	林祯舜	

午餐时间			
13:30-14:00	dplyr 包：高效数据清理	郎大为	邓一硕
14:00-14:30	用 R + shiny + rCharts 快速打造数据产品	李殷皓	
14:30-15:00	基于 rpanel 包构建图形用户界面	练勇强	
15:00-15:20	茶歇		
15:20-15:50	绘制地图的另一个选择：baidumap	杜亚磊	蔡锐
15:50-16:20	地理信息与基于位置的服务	李舰	

3. 会议意见和建议



请扫描以上二维码进入本次会议满意度调查页面，提交您宝贵的意见之后可以参与现场的抽奖赠书活动。

4. 会议机构

主办单位：

华东师范大学金融与统计学院
统计之都 (<http://cos.name/>)
Mango Solutions

赞助单位：

优酷土豆集团

会议主席：

汤银才 李舰 陈丽云 魏太云

会议详情请参见：<http://cos.name/2014/11/7th-r-conference-shanghai/>

R语言在推荐系统中的应用

刘思喆^{1,*}

京东商城

摘要

本次分享将从推荐系统实践出发，介绍R语言在推荐系统中的地位以及使用方向。并围绕推荐系统中点击率预测问题，完整介绍如何使用R语言进行数据预处理、模型构建及预测。

*现任京东商城推荐系统算法经理，主要负责Web、App、微信端用户行为分析，跟踪优化各渠道推荐效果，改进落地推荐策略。在加入京东前，供职于亚信联创BOC、神州数码思特奇DSS，主要为电信运营商提供数据挖掘及业务咨询等顾问服务。2014-2016，兼任中国人民大学大数据分析实验班硕士导师。

互联网金融大数据与风险

张云松^{1,*}

¹融360

摘要

互联网金融与传统金融业务的风险控制有相同之处并在原有基础上进行了创新。尤其在大数据的应用上互联网金融风险控制走在了传统金融的前面。 本题目将向听众介绍互联网金融，这一领域的最前沿技术和发展状况。

*融360风控决策总监。

数据科学中的R语言

肖凯^{1,*}

¹1号店

摘要

- 1、什么是数据科学
- 2、数据科学的工作分类
- 3、需要的技能组合
- 4、R语言的角色定位
- 5、数据科学的团队构成
- 6、数据科学家的职业生涯

*1号店商务智能部，数据挖掘工程师，《数据科学中的R语言》作者。

新药研发中的建模和模拟（M&S）

杨环^{1,*}

¹Mango Solutions

摘要

在整个制药行业，每年的销售收入差不多是6万亿人民币。而每年制药行业的研发成本在一万亿左右。每款上市的新药的平均研发时间是12年。在从实验室走出来通过了动物试验的药物中，最终能进入人体试验的概率只有大约千分之一。而通过人体试验最终能够上市的概率只有五分之一。平均每款药物的研发成本在50亿人民币左右。在所有的失败案例中，差不多有一半都是因为在进行人体试验时无法证明其在统计上具有显著的有效性而饮恨的。

通常在决定做试验之前，药厂就开始向FDA等监管机构提交研究性新药申请（Investigational New Drug Application，简称IND），等到做完人体试验再提交新药申请（New Drug Application，简称NDA）。实际上，从试验刚开始，新药研发就开始了。按照FDA的标准流程，需要进行四期的临床试验，

通常第二期试验（PhaseII）是最关键的阶段，此时开始招募对症的病人进行试验，常需要100到300位试验对象。花费2年左右的时间。这个阶段的成本通常占整个流程的40%左右。该阶段是建模和模拟最重要的应用领域，也是R最好的舞台。

在IND申请中，FDA于2003年提出了EOP2A（End of Phase 2A）的申请，鼓励药厂尽早地与FDA交流，并分享临床实验数据的建模和模拟经验。现在越来越多的药厂开始申请和重视EOP2A。对于这个环节的建模和模拟，FDA虽然不会推荐具体的软件，但是FDA发表的很多论文和提供的资料都是使用R实现的。在行业里，药厂负责新药研发的数据科学家们也把R当成了这个阶段标准的分析工具。

本次演讲将针对新药研发中的建模和模拟，对R的应用情况进行介绍。

*杨环，现就职于Mango Solutions (China)，担任咨询顾问。毕业于厦门大学和伦敦政治经济学院。邮箱huan.a.young@gmail.com.

一个数据分析师的入门与成长

任华^{1,*}

¹优酷土豆集团

摘要

一个大型互联网公司数据分析部门的三四个日常实战小故事。

*数据分析总监

网站实验设计框架与实战

陈丽云^{1,*}

¹eBay

摘要

互联网站大量依赖A/B实验来测试新产品性能、评估改进效果等等。随机试验虽然原理简单，但是在具体实践中牵扯到产品及平台设计、实验框架和后期评估以及异常诊断等等。本文以实例说明一般在线实验的设计框架、实施步骤及评估方法。

*巴塞罗那经济研究院经济学硕士、现任eBay资深分析师，负责eBay北美地区卖家业务的客户分析、数量建模和提供数据报告等。常用的工具无非是R和SQL，常玩的数据就是eBay那堆乱糟糟的啥都有的数据，除了本职工作以外还闲来无事写写博客搞搞沙龙之类。

基于R的三维人脸绘图系统以及利用R实现三维人脸 预测人的性格

呼思乐^{1,*}

¹中国科学院上海生命科学研究院

摘要

R内部的绘图系统能够提供功能强大的绘制二维图形的能力，其应用也极为广泛，但是使用R进行三维绘图并将其投入实际应用中的例子则相对较少。为了充分利用R强大的统计与绘图功能，我们基于R的三维绘图包“rgl”开发了一个人脸绘图系统，该系统能够方便地使用R以及“rgl”内部的函数对人脸做各种处理，并绘制出漂亮的三维人脸图形。我们利用这个人脸绘图系统，结合我们采集到的三维人脸数据以及对应个体的性格特征指标，使用一些统计学方法来实现利用三维人脸对性格的预测，并通过我们的人脸绘图系统将提取到的与性格相关的人脸特征绘制出来。

*呼思乐，男，生于1986年，本科毕业于中国科学技术大学数学系，现在是中国科学院上海生命科学研究院计算生物学研究所的博士研究生。研究方向是计算生物学，统计群体遗传学，R爱好者，日常使用R作为重要的研究工具，对R的二维以及三维绘图非常熟悉，参与了“R Graphics – second edition”一书中中文版的翻译（预计于明年由人民邮电出版社出版），并翻译了书中的主要内容。

空气污染与R

张兵^{1,*}

¹浙江省疾病预防控制中心

摘要

研究表明，空气污染物可致多种健康危害，因而研究空气质量的变化以及其产生的健康危害，对于制定空气污染防控措施和保障人群健康显得尤为重要。本次汇报将从以下方面进行叙述：1.空气质量相关概念；2.如何获取空气质量数据；3.空气质量数据简要分析；4.分布非线性滞后模型研究空气质量对人群的健康危害。

*张兵，目前就职于浙江省疾病预防控制中心。毕业于华中科技大学同济医学院，研究生时期研究方向主要为环境流行病学(偏空气污染)、空间统计学和时空数据模型，现今研究方向为传染病动力学模型。

利用R与nodejs实现可视分析工具

周扬^{1,*}

¹JDPOWER

摘要

R作为数据处理、数据分析以及数据建模的利器提供了大量的数据操作（Data Manipulation）和数据分析的函数和方法。同时，R作为数据可视化的重要工具提供了大量高质量的数据展示方法和统计图形，但是R语言默认的绘图系统在实现用户交互方面存在一定不足，因此需要借助于外部工具来实现R与数据可视化工具的连接。本演讲将利用R与基于浏览器作为展示平台的动态可交互图形工具（Javascript包，主要包括D3js, ECharts）进行连接，并且，通过加入Node.js 模块实现R与服务器进行连接通讯构建从图形到图形的可视分析系统。

*JDPOWER 数据分析师。

客户账户分析方法：用R实现建模

石宏飞^{1,*}

¹ 宏源证券

摘要

宏源证券客户账户诊断报告是以第三方视角，基于最原始的客户交易记录，整理统计客户交易及盈亏，借助于R的数学统计算法，对其交易习惯、交易偏好、盈亏特征、交易能力等方面做出完全针对于客户本人的个性化评估，使客户充分、全面地了解其交易中的成败得失，从而有针对性地改进投资技能，提高盈利能力。

*石宏飞，13年金融行业服务经验。精通数据架构及数据平台建设；熟悉数据治理及数据挖掘应用。现任宏源证券数据中心负责人，主要负责宏源证券数据中心的规划和项目管理。

基于R语言的投资监控平台之构建

邓一硕^{1,*}

¹ 统计之都

摘要

R作为一款卓越的数据分析语言，一直为金融分析人士所倚重。R不仅向用户提供了大量用以处理数据和对数据建模的工具，还提供了与MySQL、Redis和MongoDB等常用数据库的接口，加上基础绘图系统、ggplot2、ggvis等可视化包以及shiny，事实上，已经形成了一个可以解决金融投资分析链上的绝大多数问题的生态系统。本演讲将简要介绍如何基于R软件、MySQL、shiny等构建一个实用的金融投资分析界面。

*邓一硕，毕业于中央财经大学统计学专业，统计之都沙龙联合发起人，感兴趣的领域是R语言在金融投资分析中的应用。
微博：dengyishuo。

Growth Hacker的R工具箱

张翔^{1,*}

¹ “在路上”旅行APP

摘要

互联网移动化，把整个数据分析的节奏加快了一倍，消费者决策加快，行销节奏加快，公司运营加快。但是数据却前所未有的分散。web浏览的数据、app行为的数据、微信等社交媒体的数据，已经无法用超链接去串联。本次演讲将会分享Growth Hacker的角色，并介绍如何在当前复杂的情况下创建支撑整个公司运作的数据中心。

* “在路上”旅行APP的Growth Hacker，从事数据工作多年，对互联网、移动互联网各细分领域的的数据有深入的研究和实践。

电商数据化运营中的网站竞争分析

肖嘉敏^{1,*}

¹ 艾瑞咨询

摘要

1. 聚焦电商大促: 流量问题是任何电商都避开不了的现实问题，对于电子商务而言，有流量就意味着一切：交易额、现金流、利润和扩散空间。以致每年总有那么几次大促，细看2014年4.18，6.18，7月年中，8.18，双11的流量，转化及效果。

2. 竞争网站分析: 基于第三方用户行为，问卷调研，行业市场数据，洞察各竞争对手的流量趋势，来源，效果；用户行为，，专注于电商数据运营中的竞争网站分析中目标用户定位，竞争对手营销，运营监控。

*艾瑞咨询全国产品支持中心。

R-web：大型多人在线数据分析协作平台

林祯舜^{1,*}

¹ 先锋信息科技/辰智咨询

摘要

如果你是有经验的数据分析人员(或者数据科学家、数据挖掘工程师)，你会发现要将一个数据进行分析并得到有洞察力的建议，这是一个知识发现的过程，这个过程中有三个重要的组成部分: 数据、工具、人，数据在工具中经由人的探索及验证的过程逐渐积累形成知识，这个过程在未来会经由数据的交叉复用、分析方法(或算法)的镶嵌以及分析人员的协作与交流而加快知识的积累与生产，因此工具(或平台)的协作及弹性就越来越重要，这个报告要介绍第一个由华人统计学家基于R语言研发的数据分析云平台，阐述这个协作平台的设计理念及未来的愿景，希望这个平台在未来能让更多人能学习数据分析，理解分析思维，为培养未来的数据科学人才贡献一份心力。

*林祯舜博士是数据科学及营销科学方面的专家，毕业于人民大学统计学院并获得博士学位，在企业界，目前担任信息技术咨询公司的总经理，在学术界，目前是兰州商学院及吉林大学的兼职教授。林博士学术领域的研究方向包括数据挖掘，机器学习，统计计算，网站效果测量与点击流数据分析。林博士的学术论文是关于互联网点击流的模型应用，这个模型是第一个应用在互联网媒体规划的基础模型，主要的贡献是打破互联网媒体规划和传统电视媒体规划使用相同模型的迷思，对互联网媒体规划的理论与应用找出一个突破口，目前全球互联网监测公司所使用的媒体规划模型，都是在林博士发表的论文基础上加以改进并产品化。这篇论文在2010年五月被美国网站分析协会(Web Analytics association)选为网站分析领域最需要阅读的14篇论文之一。国际顶尖广告研究期刊，广告研究学报(Journal of Advertising Research)在2011创刊50周年的特刊上，针对互动网络的专文中，特别提到林博士论文在互动媒体规划方面的贡献，这是该领域被提到的四篇文章之一。林博士相关的学术研究论文发表在Journal of Advertising(JA)，Journal of the American Society for Information Science and Technology (JASIS&T)，Information Research，营销科学学报等期刊。

dplyr包：高效数据清理

郎大为^{1,*}

¹SupStat

摘要

数据清洗一直是数据挖掘中极为重要，也无法回避的问题，R提供了很多用于数据处理的程序包，dplyr便是其中之一。合理的语言设计，媲美python的处理速度，丰富的数据库接口，正是这些让dplyr包成为数据分析师手中的另一利器。dplyr包是Hadley Wickham于2014年初编写的一款用于清洗数据的程序包，本次演讲将会着重介绍dplyr包的基础函数。

*郎大为，Supstat高级数据分析师，John Hopkins数据科学专项认证，有丰富的R的使用，建模，可视化的经验。擅长的领域为：非参数方法，金融数据挖掘，地理数据的可视化等。邮箱：dawei.lang@supstat.com

用R + shiny + rCharts快速打造数据产品

李殷皓^{1,*}

¹PayPal

摘要

Shiny包是一套基于R语言实现的网页框架，其作用是让不熟悉网页编程语言的R用户也可以快速搭建起基于R的交互式网页应用。rCharts包则提供了对多个javascript数据可视化(highcharts/nvd3/polychart)的R封装。通过Shiny和rCharts，数据分析师和数据科学家们可以很方便的实现一个最小可用的数据产品原型。

本次报告将通过一个简单的例子介绍如何通过shiny和rCharts构造一个基于浏览器的商业智能产品，并分享将shiny用于实际日常工作中的种种经验体会。

*李殷皓，数据分析师，供职于PayPal Seller Risk Business Intelligence Team

基于rpanel包构建图形用户界面

练勇强^{1,*}

¹华东师范大学

摘要

rpanel是建立在tcltk包上用于创建图形用户界面的R包，它可以允许在界面里添加滑动条、复选框和下拉框等元素生动地对R函数进行交互控制。报告首先列出了其他一些类似的R包，然后详细介绍rpanel的一些界面元素，最后通过实例来展现rpanel的应用前景。

*华东师范大学金融与统计学院博士生，研究方向为贝叶斯统计和可靠性统计。

绘制地图的另一个选择：baidumap

杜亚磊^{1,*}

¹SupStat

摘要

ggmap是地图可视化中常用的包，通过Google地图API来绘制地图。由于国内访问Google地图经常出现异常，所以便有了baidumap的诞生。baidumap通过百度地图API获取地图，解决了访问Google不便的问题。另外也提供了坐标转化等其他功能，将常用的非百度坐标（如GPS设备获取的角度坐标，Google地图坐标等）转换成百度地图中使用的坐标。

*杜亚磊，Supstat数据分析师，baidumap包的作者。R user，偶尔用python，awk等其他工具，也折腾过hadoop，spark。擅长数据处理，建模和可视化。邮箱：yalei.du@supstat.com，主页：<http://badbye.github.io/>。

地理信息与基于位置的服务

李舰^{1,*}

¹Mango Solutions

摘要

CRAN上与空间数据（spatial data）相关的R包超过了100个，是R的热门应用领域之一。在实际的应用中，通常把空间数据称为地理信息数据。我们研究的问题是如何将带有经纬度的地理信息数据转化成统计分析中的数据对象，然后使用统计方法和可视化方法对数据进行分析 and 展现。对于这些带有经纬度的位置信息，研究其增值服务也是当今一个热门的应用方向，通常称为基于位置的服务（LBS）。

对于地理信息，我们除了使用传统的工具进行可视化展现与专业的分析之外。搜集和获取地理信息也是一个非常重要的分析需求。尤其是在大数据时代，人们生活在一个数据的世界，将各类数据之间的壁垒打通，可以获取意想不到的好处。地理信息数据可以和各种其他数据相结合，而随着Google和Baidu这样的互联网巨头不断的贡献，普通用户获取地理信息数据也成为了可能。

Google Maps API是业界常用的地理信息接口，但是其免费版提供的查询功能非常有限，如果是查询中国范围内的数据，百度API是更好的选择。演讲者基于该API开发了一个R语言接口：RbaiduLBS包，可以获取更多的地理信息数据。本次演讲将结合该包介绍地理信息和LBS的应用。

*李舰，毕业于中国人民大学统计学院（本科）和北京大学软件与微电子学院（研究生），现就职于Mango Solutions，担任中国区数据总监，负责数据分析相关的咨询项目及公司基于R语言的产品开发。Rweibo、Rwordseg、tmcn等R包的作者。《数据科学中的R语言》的作者，《R语言核心技术手册》的译者。邮箱：lijian.pku@gmail.com，主页：<http://www.jianl.org>。