

格致方法·定量研究系列 吴晓刚 主编



分析制图与地理数据库

[美] G.戴维·加森 (G. David Garson) 著
罗伯特·S.比格斯 (Robert S. Biggs)

曾东林 译 梁海祥 校

- ★ 革新研究理念
- ★ 丰富研究工具
- ★ 最权威、最前沿的定量研究方法指南

格致出版社 上海人民出版社

58

格致方法·定量研究系列 吴晓刚 主编

分析制图与地理数据库

[美] G.戴维·加森(G. David Garson) 著
罗伯特·S.比格斯(Robert S. Biggs) 著
曾东林 译 梁海祥 校

SAGE Publications, Inc.

格致出版社 上海人民出版社

图书在版编目(CIP)数据

分析制图与地理数据库/(美)G.戴维·加森,(美)罗伯特·S.比格斯著;曾东林译.—上海:格致出版社:上海人民出版社,2017.4
(格致方法·定量研究系列)
ISBN 978-7-5432-2725-5

I. ①分… II. ①G… ②罗… ③曾… III. ①数据处理 IV. ①TP274

中国版本图书馆 CIP 数据核字(2017)第 033713 号

责任编辑 张苗凤

格致方法·定量研究系列 分析制图与地理数据库

[美]G.戴维·加森 罗伯特·S.比格斯 著
曾东林 译 梁海祥 校

出版 世纪出版股份有限公司 格致出版社
世纪出版集团 上海人民出版社
(200001 上海福建中路 193 号 www.cwen.co)



编辑部热线 021-63914988
市场部热线 021-63914081
www.hibooks.cn

发行 上海世纪出版股份有限公司发行中心

印刷 上海商务联西印刷有限公司
开本 920×1168 1/32
印张 5.5
字数 92,000
版次 2017年4月第1版
印次 2017年4月第1次印刷

ISBN 978-7-5432-2725-5/C·171

定价:32.00 元

出版说明

由香港科技大学社会科学部吴晓刚教授主编的“格致方法·定量研究系列”丛书,精选了世界著名的 SAGE 出版社定量社会科学研究丛书,翻译成中文,起初集结成八册,于 2011 年出版。这套丛书自出版以来,受到广大读者特别是年轻一代社会科学工作者的热烈欢迎。为了给广大读者提供更多的方便和选择,该丛书经过修订和校正,于 2012 年以单行本的形式再次出版发行,共 37 本。我们衷心感谢广大读者的支持和建议。

随着与 SAGE 出版社合作的进一步深化,我们又从丛书中精选了三十多个品种,译成中文,以飨读者。丛书新增品种涵盖了更多的定量研究方法。我们希望本丛书单行本的继续出版能为推动国内社会科学定量研究的教学和研究作出一点贡献。

总序

2003年,我赴港工作,在香港科技大学社会科学部教授研究生的两门核心定量方法课程。香港科技大学社会科学部自创建以来,非常重视社会科学研究方法论的训练。我开设的第一门课“社会科学里的统计学”(Statistics for Social Science)为所有研究型硕士生和博士生的必修课,而第二门课“社会科学中的定量分析”为博士生的必修课(事实上,大部分硕士生修完第一门课后都会继续选修第二门课)。我在讲授这两门课的时候,根据社会科学研究生的数理基础比较薄弱的特点,尽量避免复杂的数学公式推导,而用具体的例子,结合语言和图形,帮助学生理解统计的基本概念和模型。课程的重点放在如何应用定量分析模型研究社会实际问题上,即社会研究者主要为定量统计方法的“消费者”而非“生产者”。作为“消费者”,学完这些课程后,我们一方面能够读懂、欣赏和评价别人在同行评议的刊物上发表的定量研究的文章;另一方面,也能在自己的研究中运用这些成熟的方法论技术。

上述两门课的内容,尽管在线性回归模型的内容上有少

量重复,但各有侧重。“社会科学里的统计学”从介绍最基本的社会研究方法论和统计学原理开始,到多元线性回归模型结束,内容涵盖了描述性统计的基本方法、统计推论的原理、假设检验、列联表分析、方差和协方差分析、简单线性回归模型、多元线性回归模型,以及线性回归模型的假设和模型诊断。“社会科学中的定量分析”则介绍在经典线性回归模型的假设不成立的情况下的一些模型和方法,将重点放在因变量为定类数据的分析模型上,包括两分类的 logistic 回归模型、多分类 logistic 回归模型、定序 logistic 回归模型、条件 logistic 回归模型、多维列联表的对数线性和对数乘积模型、有关删节数据的模型、纵贯数据的分析模型,包括追踪研究和事件史的分析方法。这些模型在社会科学研究中有着更加广泛的应用。

修读过这些课程的香港科技大学的研究生,一直鼓励和支持我将两门课的讲稿结集出版,并帮助我将原来的英文课程讲稿译成了中文。但是,由于种种原因,这两本书拖了多年还没有完成。世界著名的出版社 SAGE 的“定量社会科学研究”丛书闻名遐迩,每本书都写得通俗易懂,与我的教学理念是相通的。当格致出版社向我提出从这套丛书中精选一批翻译,以飨中文读者时,我非常支持这个想法,因为这从某种程度上弥补了我的教科书未能出版的遗憾。

翻译是一件吃力不讨好的事。不但要有对中英文两种语言的精准把握能力,还要有对实质内容有较深的理解能力,而这套丛书涵盖的又恰恰是社会科学中技术性非常强的内容,只有语言能力是远远不能胜任的。在短短的一年时间里,我们组织了来自中国内地及香港、台湾地区的二十几位

研究生参与了这项工程,他们当时大部分是香港科技大学的硕士和博士研究生,受过严格的社会科学统计方法的训练,也有来自美国等地对定量研究感兴趣的博士研究生。他们是香港科技大学社会科学部博士研究生蒋勤、李骏、盛智明、叶华、张卓妮、郑冰岛,硕士研究生贺光烨、李兰、林毓玲、肖东亮、辛济云、於嘉、余珊珊,应用社会经济研究中心研究员李俊秀;香港大学教育学院博士研究生洪岩璧;北京大学社会学系博士研究生李丁、赵亮员;中国人民大学人口学系讲师巫锡炜;中国台湾“中央”研究院社会学所助理研究员林宗弘;南京师范大学心理学系副教授陈陈;美国北卡罗来纳大学教堂山分校社会学系博士候选人姜念涛;美国加州大学洛杉矶分校社会学系博士研究生宋曦;哈佛大学社会学系博士研究生郭茂灿和周韵。

参与这项工作的许多译者目前都已经毕业,大多成为中国内地以及香港、台湾等地区高校和研究机构定量社会科学方法教学和研究的骨干。不少译者反映,翻译工作本身也是他们学习相关定量方法的有效途径。鉴于此,当格致出版社和 SAGE 出版社决定在“格致方法·定量研究系列”丛中推出另外一批新品种时,香港科技大学社会科学部的研究生仍然是主要力量。特别值得一提的是,香港科技大学应用社会经济研究中心与上海大学社会学院自 2012 年夏季开始,在上海(夏季)和广州南沙(冬季)联合举办“应用社会科学研究方法研修班”,至今已经成功举办三届。研修课程设计体现“化整为零、循序渐进、中文教学、学以致用”的方针,吸引了一大批有志于从事定量社会科学研究博士生和青年学者。他们中的不少人也参与了翻译和校对的工作。他们在

繁忙的学习和研究之余,历经近两年的时间,完成了三十多本新书的翻译任务,使得“格致方法·定量研究系列”丛书更加丰富和完善。他们是:东南大学社会学系副教授洪岩璧,香港科技大学社会科学部博士研究生贺光烨、李忠路、王佳、王彦蓉、许多多,硕士研究生范新光、缪佳、武玲蔚、臧晓露、曾东林,原硕士研究生李兰,密歇根大学社会学系博士研究生王骁,纽约大学社会学系博士研究生温芳琪,牛津大学社会学系研究生周穆之,上海大学社会学院博士研究生陈伟等。

陈伟、范新光、贺光烨、洪岩璧、李忠路、缪佳、王佳、武玲蔚、许多多、曾东林、周穆之,以及香港科技大学社会科学部硕士研究生陈佳莹,上海大学社会学院硕士研究生梁海祥还协助主编做了大量的审校工作。格致出版社编辑高璇不遗余力地推动本丛书的继续出版,并且在这个过程中表现出极大的耐心和高度的专业精神。对他们付出的劳动,我在此致以诚挚的谢意。当然,每本书因本身内容和译者的行文风格有所差异,校对未免挂一漏万,术语的标准译法方面还有很大的改进空间。我们欢迎广大读者提出建设性的批评和建议,以便再版时修订。

我们希望本丛书的持续出版,能为进一步提升国内社会科学定量教学和研究水平作出一点贡献。

吴晓刚

于香港九龙清水湾

序

在社会科学领域，地图是一个被忽视了的分析工具。鉴于地理在人类行为中的影响，这种忽视显得难以解释。究其根源，这看起来很大程度上在于难以有效和方便地使用地图。例如，假设政治学家简·怀特(Jane White)想要研究1988年美国总统大选中关于民主党选票的选举地理学。那么她是应该根据选票百分比在州(区、县?)层级进行晕渲(shading)的等值区域图(choropleth map)吗？如果是的话，那她到哪里去获得地图？应该使用哪一种晕渲方法？以及设置多少精度的晕渲效果呢？当她检查已完成的地图后，会发现在这个国家的地区与投票之间的关系吗？她如何量化这种关系呢？其他非地理因素能否也被添加进来以检验这种关系呢？当根据先前的选举构建类似的地图时，我们能够观察到随着时间发生怎样的变化呢？

要想回答上述问题,怀特教授必须以某种方式收集、存储、展示并且分析这些选举的地理数据。因为在过去,该项工作涉及高强度的手工作业,导致现代的定量政治学家选择回避它。幸运的是,这本由加森(Garson)和比格斯(Biggs)完成的专著展示了这些传统的困难是如何被解决的。当代的分析人员可以获得很多价格合理、基于计算机的地理信息系统(GIS)软件,并将其与许多易于获取的地理数据资源相结合。例如,美国人口普查局(U. S. Bureau of the Census)建立了拓扑集成地理编码与参考(Topologically Integrated Geographic Encoding and Referencing, TIGER)文件,其中包含了第一张数字化的美国街区地图。TIGER可以添加到GIS软件包,如Safari*(来自Geographic Data Technology公司)。一旦研究人员选择了合适的数据集和软件,就可以直接转向分析制图了。

作者回顾了多种类型的地图——仅举几个例子:UBC、密度(dasymeric)、街区(block)、等值(isarithmic)——并解释了大量有趣的、不常见的汇总统计,如地理均值(geographic means)、区位熵(location quotients)和空间相关(areal correspondence)。对统计的理解将引导我们利用地图进行建模,其中将会用到多变量回归技术(multivariate regression),正如GIS软件与标准的统计软件(如SAS)结合一样。

* 由于本书英文版出版时间较早,其中提到的一些GIS软件工具已经更新换代或者不再使用。——译者注

GIS与分析制图的潜在应用超出了常规的社会科学领域。分析制图已经广泛应用于政府部门,规划人员与政策制定者将其用于学校、再分区、税务、自然灾害以及犯罪等问题。例如,应用在华盛顿州塔科马市的犯罪分析制图系统(Crime Analysis Mapping System),使得犯罪数据与人口普查数据能够以精妙的方式展示在地图上。正如加森和比格斯在这本急需的导论中清晰描述的那样,在传统的数据分析向地理驱动分析的转变过程中,我们将受益匪浅。

迈克尔·S.刘易斯—贝克

目录

序	1
第 1 章 导论	1
第 1 节 术语	7
第 2 节 分析制图和地理数据在社会和政策科学领域的案例	20
第 2 章 地理数据	33
第 1 节 地理信息系统	34
第 2 节 不同地理图层的融合	45
第 3 节 普通数据源	49
第 3 章 地图	61
第 1 节 数字化地图	64
第 2 节 图形文件的格式和转换	68

第 4 章	分析制图	71
第 1 节	地图类型	72
第 2 节	缓冲区	91
第 3 节	叠加	92
第 4 节	地图建模	94
第 5 章	展示	123
第 1 节	输出	124
第 2 节	显示	127
第 3 节	图形失真	128
第 6 章	结论	131
附录		135
注释		139
参考文献		141
译名对照表		154

第 **1** 章

导 论

在过去十年中,使用地理数据库的分析制图领域得到了极大的发展。过去十年的显著特征是:(1)出现了大部分社会科学家都负担得起的“微计算机革命”和复杂地理信息系统(GIS)*软件的桌面版本;(2)改善了地理数据收集与传播,其中包括美国人口普查局出版了首幅数字化的全美地图;(3)通过GIS软件将复杂的分析技术应用到地理数据中的方法在不断进步。分析制图如同其他应用到数据分析的可视化方法一样,曾经需要难以置信的繁重工序,但现在可以由计算机来完成。其全面影响在最近才显现出来。例如,据估计在联邦层面广泛使用GIS的政府机构,其数目增加了不止一倍,从1990年的18个增长到1992年的44个(U.S. General Accounting Office, 1991)。分析制图有非常多的应用,不仅仅局限于政府政策层面,也被广泛地应用到社会科学课题中。任何随着空间和时

* GIS与地理信息系统(Geographic Information System)在文中表述等同。——译者注

间的动态扩散与分布的变量,都属于分析制图的范畴。对社会科学家而言,只有统计分析才是最重要的,然而看起来这个领域被社会研究方法的研究生课程刻意回避了,该问题的根源产生于分析制图技术对普通社会科学家而言变得切实可行之前。

遗憾的是,GIS倾向于成为机构专家的特有领域,而非一个通用的工具,正如发生在统计软件上的那样。“通常,具有计算机操作背景的人会听过或者见过地理信息系统或者自动化制图系统,并且通过参加会议或者研讨会,成为该项技术的推动者,”一位分析人员最近写道。“运气好的话,”这位分析人员继续写道,他“将会最终整合足够的资源建立一个系统,并且可以期待成为这个系统的管理者。这个机构(agency)的其他工作人员将会非常高兴地发现他们的部门正在应用这项新技术,但是同样值得高兴的是掌握这个系统的所有责任都落在跟进这个系统的专业人士身上”(Tomlinson Associates, 1989:174)。用一门关于经济学、政治学、心理学或者社会学的学科可替换上述提到的“机构”,并且该声明也同样适用于社会科学。

地理信息系统就是硬件和软件的组合,其为了管理与地理位置有关的数据而整合了计算机图形学和关系数据库(Ripple, 1989)。这些地理数据在本质上既是空间的又是描述性的。描述性的数据和文本被存储在关系数据库中。这个独一无二的GIS组件就是一个系统,能够追踪诸

如一条数据库记录相对于另外一条数据库记录的“距离”(nearness)这样的空间概念,以及其他可能的关系,例如“南边/北边”、“内部/外部”,或者“上方/下方”。在高级应用里,关系可能包括与当时太阳轨迹或风向有关的地理单位的原点,以及噪声分布,或者可能涉及三维地图,正如用于波士顿港清理作业的 GIS 那样(Ardalan, 1988)。

计算机制图在政府中特别重要,因此对于那些研究政府政策的社会科学家尤为如此。据估计在地方政府政策制定者的信息需求中,大约有 80%是与地理空间有关的(Williams, 1987:151)。重新划分区域就是一个常见的例子,如同与公共事务、税务评估、公共事业监管、规划相关的很多活动一样。制图功能的计算机化使得人口普查、税务、犯罪及其他数据可以在地图上表达出来;使得所有地图都能以集中的方式更新,并且不易损坏或丢失;使得地图更容易操作,包括在任意比例尺下的自动输出;也能自动提供与地图关联的数据(如,给位于邻近地块的那些受到再分区听证会影响的所有业主邮寄标签)。

美国人口普查局的拓扑集成地理编码与参考(Topologically Integrated Geographic Encoding and Referencing, TIGER)系统彻底改革了制图,实现了惊人的细节分析。地理信息系统现在是一个几十亿美元的产业,因为政策制定者发现来自 GIS 的地理分析人员比主导社会科学的传统统计分析人员更为实用。即便是在统计分析领域,对数