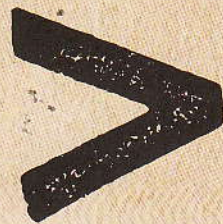



**INNOVATIONS IN**  
**DIGITAL**  
**RESEARCH METHODS**

**PETER HALFPENNY**  
**ROB PROCTER**



companion  
website 





**INNOVATIONS IN**  
**DIGITAL**  
**RESEARCH METHODS**

**PETER HALFPENNY**  
**ROB PROCTER**

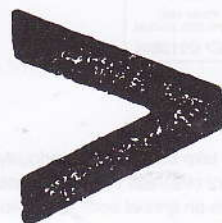
Accession no. **M 0149674**

Date received **29 JUL 2016**

Call no.

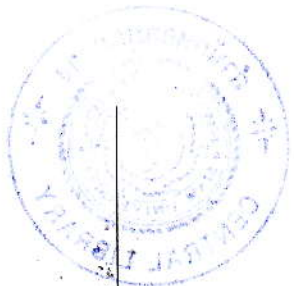


Los Angeles | London | New Delhi  
Singapore | Washington DC





Los Angeles | London | New Delhi  
Singapore | Washington DC



SAGE Publications Ltd  
1 Oliver's Yard  
55 City Road  
London EC1Y 1SP

SAGE Publications Inc.  
2455 Teller Road  
Thousand Oaks, California 91320

SAGE Publications India Pvt Ltd  
B 1/1 1 Mohan Cooperative Industrial Area  
Mathura Road  
New Delhi 110 044

SAGE Publications Asia-Pacific Pte Ltd  
3 Church Street  
#10-04 Samsung Hub  
Singapore 049483

Editor: Jai Seaman  
Assistant editor: Lily Mehrbod  
Production editor: Victoria Nicholas  
Copyeditor: Catjat Pafort  
Proofreader: Rosemary Morlin  
Indexer: David Rudeforth  
Marketing manager: Sally Ransom  
Cover design: Francis Kenney  
Typeset by: C&M Digital (P) Ltd, Chennai, India  
Printed and bound by CPI Group (UK) Ltd,  
Croydon, CR0 4YY

Chapter 1: © Peter Halfpenny and Rob Proctor 2015  
Chapter 2: © Kingsley Purdam and Mark Elliot 2015  
Chapter 3: © Kingsley Purdam and Mark Elliot 2015  
Chapter 4: © Joe Murphy 2015  
Chapter 5: © Paul S. Lambert 2015  
Chapter 6: © Mark Birkin and Nick Malleson 2015  
Chapter 7: © Paul S. Lambert, William J. Browne and Danius T. Michaelides 2015  
Chapter 8: © Lawrence Ampofo, Simon Collister, Ben O'Loughlin and Andrew Chadwick 2015  
Chapter 9: © Andy Crabtree, Paul Tennent, Pat Brundell and Dawn Knight 2015  
Chapter 10: © Rob Ackland and Jonathan Zhu 2015  
Chapter 11: © Michael Batty, Steven Gray, Andrew Hudson-Smith, Richard Milton, Oliver O'Brien and Flora Roumpani 2015  
Chapter 12: © R.J. Anderson and Marina Jirotko 2015  
Chapter 13: © Mike Savage 2015

Apart from any fair dealing for the purposes of research or private study, or criticism or review, as permitted under the Copyright, Designs and Patents Act, 1988, this publication may be reproduced, stored or transmitted in any form, or by any means, only with the prior permission in writing of the publishers, or in the case of reprographic reproduction, in accordance with the terms of licences issued by the Copyright Licensing Agency. Enquiries concerning reproduction outside those terms should be sent to the publishers.

**Library of Congress Control Number: 2014954811**

**British Library Cataloguing In Publication data**

A catalogue record for this book is available from the British Library



ISBN 978-1-4462-0308-8  
ISBN 978-1-4462-0309-5 (pbk)

At SAGE we take sustainability seriously. Most of our products are printed in the UK using FSC papers and boards. When we print overseas we ensure sustainable papers are used as measured by the Egmont grading system. We undertake an annual audit to monitor our sustainability.

# CONTENTS

<b>List of Figures and Tables</b>	<b>vii</b>
<b>List of Contributors</b>	<b>x</b>
<b>Acknowledgements</b>	<b>xvii</b>
<b>Companion Website</b>	<b>xviii</b>
<b>1 Introduction and Overview</b>	<b>1</b>
Peter Halfpenny and Rob Proctor	
<b>2 The Changing Social Science Data Landscape</b>	<b>25</b>
Kingsley Purdam and Mark Elliot	
<b>3 Exploiting New Sources of Data</b>	<b>59</b>
Mark Elliot and Kingsley Purdam	
<b>4 Survey Methods: Challenges and Opportunities</b>	<b>85</b>
Joe Murphy	
<b>5 Advances in Data Management for Social Survey Research</b>	<b>105</b>
Paul S. Lambert	
<b>6 Modelling and Simulation</b>	<b>123</b>
Mark Birkin and Nick Malleson	
<b>7 Contemporary developments in statistical software for social scientists</b>	<b>143</b>
Paul S. Lambert, William J. Browne and Danius T. Michaelides	
<b>8 Text Mining and Social Media: When Quantitative Meets Qualitative and Software Meets People</b>	<b>161</b>
Lawrence Ampofo, Simon Collister, Ben O'Loughlin and Andrew Chadwick	
<b>9 Digital Records and the Digital Replay System</b>	<b>193</b>
Andy Crabtree, Paul Tennent, Pat Brundell and Dawn Knight	



<b>10 Social Network Analysis</b>	<b>221</b>
Robert Ackland and Jonathan J.H. Zhu	
<b>11 Visualizing Spatial and Social Media</b>	<b>245</b>
Michael Batty, Steven Gray, Andrew Hudson-Smith, Richard Milton, Oliver O'Brien and Flora Roumpani	
<b>12 Ethical Praxis in Digital Social Research</b>	<b>271</b>
R.J. Anderson and Marina Jirotko	
<b>13 Sociology and the Digital Challenge</b>	<b>297</b>
Mike Savage	
 <b>Index</b>	 <b>311</b>

# LIST OF FIGURES AND TABLES

## FIGURES

3.1	Example of an electoral candidate's website	62
3.2	Rumour spreading graphic	64
3.3	Geographical distribution of immigrant pupils by local authority district types in England, 2003 to 2007	67
3.4	Observation sheet	68
3.5	Probability of postnatal depression across SES by the number of reactive 5-HTT alleles	73
5.1	Illustration of part of a 'variable by case' matrix	107
5.2	Illustration of online resources associated with agencies involved in the storage and distribution of social survey datasets	111
6.1	The research lifecycle	135
6.2	Components of a research infrastructure for social simulation	135
6.3	Elements of a workflow architecture for social simulation	136
6.4	Indicators/externalities a) baseline, b) projection and c) scenario; pollution in the cities of Leeds, Bristol and Southampton	137
7.1	Illustration of using the SPSS package to perform correspondence analysis	144
7.2	An illustration of using a script ('do file') in Stata	150
7.3	Images of Stat-JR in operation	154
7.4	Using Stat-JR from within a DEEP eBook	156
8.1	Volume of tweets about each leader's response to a question in the third debate	176
8.2	Share of positive sentiment for party leaders	177
8.3	Trend in positive sentiment for Cameron	177
8.4	Trend in positive sentiment for Clegg	177
8.5	Volume of tweets expressing positive sentiment about party leaders in the third debate	179



9.1	Example log file	197
9.2	Re-representing logs: a line chart visualizing heart rate over time	200
9.3	Connected and dynamic interactive charts	200
9.4	Synchronizing heterogeneous data	202
9.5	Annotation schema	203
9.6	Annotation sets and coding tracks	203
9.7	Thick description – creating live texts	204
9.8	Fieldwork tracker	206
9.9	Fieldwork tracker log file (an example)	207
9.10	Representing geo-located data	208
9.11	Creating a frequency table of words	209
9.12	Simultaneous concordance	210
9.13	Visualizing crowdsourced data (an example)	213
9.14	Histogram and event series showing dynamic selection and interrogation	213
9.15	Spatial distribution of events	214
9.16	Medical emergencies	214
9.17	Emergency medical clinics	215
9.18	Creating and exploring new categories of data	216
10.1	A semantic network of top hashtags from Twitter	221
10.2	A directed and outdegree-weighted network	222
10.3	Articles on top 20 online social networks in Web of Science 2004–13	227
11.1	CASA's MapTube Website showing a) population density in 2011 and b) changes in density 2001–2011	251
11.2	2011 Population density at the metropolitan scale in Greater London	252
11.3	Moving to 3D visualization and navigating through the models	255
11.4	Augmenting 3D visualization merging the virtual with the real	256
11.5	a) the Dashboard and b) its display in a visualization wall	258
11.6	Real-time tube train locations	259
11.7	Geometry of the tube network and real-time volumes at stations	260
11.8	Impact of closing a mainline station (Liverpool Street) on flow of travellers passing through related stations	260
11.9	The spatial density of tweets in London	261
11.10	Geo-located tweets captured from Twitter between 15:00 and 22:00 BST on Tuesday 9 August 2011	262

## LIST OF FIGURES AND TABLES

11.11	Spatial crowdsourcing: evolving data in real-time	264
11.12	Visualizing model outputs a) in 2D and 3D with b) data at the metro-region level and c) at the local level	266
11.13	Using procedural modelling in City Engine to visualize radially structured land use activity patterns	268

## TABLES

5.1	Summary of selected recent research projects which embody a digital social research approach to data management challenges linked to social survey research	115
8.1	A summary of some of the more commonly used commercial and free text mining tools	184
10.1	Online networks by direction and manifestation of ties	227
10.2	Tools for collection of online network data	231



# COMPANION WEBSITE

This book is supported by a brand new companion website (<https://study.sagepub.com/halfpennyprocter>). The website offers a wide range of free learning resources, including:

- Chapter summaries
- Links to online sources listed in each chapter
- Links to demos, slides and videos
- Links to current research.



# 1

## INTRODUCTION AND OVERVIEW

PETER HALFPENNY AND ROB PROCTER

### 1.1 INTRODUCTION

The dramatic increase over the last two decades or so in computing power, in wired and wireless connectivity, and in the availability of data has affected all aspects of our lives. Our aim in this book is to provide an accessible introduction to how social science researchers are harnessing innovations in digital technologies to transform their research methods. In this chapter we provide an overview of how and why e-Research methods have emerged, including an account of the drivers that have motivated their development and the barriers to their successful adoption. The chapters that follow examine how innovations in digital technologies are enabling the emergence of more powerful research infrastructure, services and tools, and how social science researchers are exploiting them.

#### 1.1.1 Digital Data

As everyone exposed to the Internet is aware, the amount of digital data available is expanding very rapidly, both through the digitization of past records and by the accretion of 'born digital' materials that are in machine-readable form from the outset. The digital universe – the data we create and copy annually – is estimated to be doubling in size every two years and projected to reach 44 trillion gigabytes by 2020 (where a trillion is a million million, or  $10^{12}$ ) (IDC, 2014). For social scientists, the predictions that more data will be generated in the next five years than in the entire history of human endeavour is both an opportunity and a challenge.

Today, vast amounts of data are generated as people go about their daily activities, both data that is deliberately produced and that which is generated by embedded systems. For example, use of public services is captured in administrative records; in the private sector, patterns of consumption of goods and services are captured in credit and debit card records; patterns of personal communications are captured in telephone

