" . . . the perfect introduction to the emerging field of spatial data science."

– Luc Anselin, Director, Center for Spatial Data Science; University of Chicago; creator of GeoDa software

" . . . an excellent blend of key theoretical concepts and applications . . . engaging examples, demonstrative code, and laboratory follow-up exercises make this book suitable for both self-learners and traditional academic settings. Highly recommended."

– Giorgos Mountrakis, State University of New York College of Environmental Science and Forestry

"An excellent course text for students of GIS, spatial statistics, quantitative geography, and ecology . . . Essential reading for beginning students as well as those who wish to refresh their knowledge."

– Michael Batty, Centre for Advanced Spatial Analysis, University College London

"Highly valuable and timely book for multidisciplinary professionals and students who aim to work with spatial problems . . . an excellent introduction in the concepts and tools to think and analyze spatially."

– Professor Walter T. de Vries, Technical University of Munich

This is an introductory textbook on spatial analysis and spatial statistics through GIS. Each chapter presents methods and metrics, explains how to interpret results, and provides worked examples. The worked examples link theory to practice through a single real-world case study, with software and illustrated guidance. This is a valuable resource for graduate students and researchers analyzing geospatial data through a spatial analysis lens, including those using GIS in the environmental sciences, geography, and social sciences.

Cover image: Vector globes. Copyright © Josiah Jost, Siah Design Inc.

Cover design by Holly Johnson

CAMBRIDGE
UNIVERSITY PRESS
www.cambridge.org

ISBN 978-1-108-71293-4

Preface *page* xiii

Think Spatially: Basic Concepts of Spatial Analysis and Space Conceptualization 1
Learning Objectives 1
1.1 Introduction: Spatial Analysis 2
1.2 Basic Definitions 6
1.3 Spatial Data: What Makes Them Special? 10
1.4 Conceptualization of Spatial Relationships 15
1.5 Distance Measure 17
1.5.1 Fixed Distance Band (Sphere of Influence) 18
1.5.2 Distance Decay 20
1.6 Contiguity: Adjacency Matrix 23
1.6.1 Polygons Contiguity 23
1.6.2 Adjacency Matrix 25
1.7 Interaction 26
1.8 Neighborhood and Neighbors 27
1.8.1 *k*-Nearest Neighbors (*k*-NN) 27
1.8.2 Space–Time Window 29
1.8.3 Proximity Polygons 30
1.8.4 Delaunay Triangulation and Triangular Irregular Networks (TIN) 31
1.9 Spatial Weights and Row Standardization 31
1.10 Chapter Concluding Remarks 33
Questions and Answers 34
Lab 1 The Project: Spatial Analysis for Real Estate Market Investments 39
Overall Progress 39
Scope of Analysis 39
Dataset Structure 44
Guidelines 45

Section A ArcGIS 45
Exercise 1.1 Getting to Know the Data and Study Region 45

Section B GeoDa 52
Exercise 1.1 Getting to Know the Data and Study Region 52

Exploratory Spatial Data Analysis Tools and Statistics 59
Learning Objectives 59
2.1 Introduction in Exploratory Spatial Data Analysis, Descriptive Statistics, Inferential Statistics and Spatial Statistics 60
2.2 Simple ESDA Tools and Descriptive Statistics for Visualizing Spatial Data (Univariate Data) 63
2.2.1 Choropleth Maps 63
2.2.2 Frequency Distribution and Histograms 65
2.2.3 Measures of Center 69
2.2.4 Measures of Shape 71
2.2.5 Measures of Spread/Variability – Variation 72
2.2.6 Percentiles, Quartiles and Quantiles 75
2.2.7 Outliers 76
2.2.8 Boxplot 78
2.2.9 Normal QQ Plot 81
2.3 ESDA Tools and Descriptive Statistics for Analyzing Two or More Variables (Bivariate Analysis) 82
2.3.1 Scatter Plot 82
2.3.2 Scatter Plot Matrix 84
2.3.3 Covariance and Variance–Covariance Matrix 85
2.3.4 Correlation Coefficient 87
2.3.5 Pairwise Correlation 90
2.3.6 General QQ Plot 91
2.4 Rescaling Data 91
2.5 Inferential Statistics and Their Importance in Spatial Statistics 95
2.5.1 Parametric Methods 96
2.5.2 Nonparametric Methods 100
2.5.3 Confidence Interval 101
2.5.4 Standard Error, Standard Error of the Mean, Standard Error of Proportion and Sampling Distribution 102
2.5.5 Significance Tests, Hypothesis, p-Value and z-Score 104
2.6 Normal Distribution Use in Geographical Analysis 109
2.7 Chapter Concluding Remarks 111
Questions and Answers 112
Lab 2 Exploratory Spatial Data Analysis (ESDA): Analyzing and Mapping Data 117

Overall Progress 117
Scope of the Analysis: Income and Expenses 117

Section A ArcGIS 118
Exercise 2.1 ESDA Tools: Mapping and Analyzing the Distribution of Income 118

Exercise 2.2 Bivariate Analysis: Analyzing Expenditures by Educational Attainment 131

Section B GeoDa 136
Exercise 2.1 ESDA Tools: Mapping and Analyzing the Distribution of Income 136

Exercise 2.2 Bivariate Analysis: Analyzing Expenditures by Educational Attainment 142

Analyzing Geographic Distributions and Point Patterns 147
Learning Objectives 147
3.1 Analyzing Geographic Distributions: Centrography 148
3.1.1 Mean Center 148
3.1.2 Median Center 151
3.1.3 Central Feature 152
3.1.4 Standard Distance 154
3.1.5 Standard Deviational Ellipse 156
3.1.6 Locational Outliers and Spatial Outliers 158
3.2 Analyzing Spatial Patterns: Point Pattern Analysis 163
3.2.1 Definitions: Spatial Process, Complete Spatial Randomness, First- and Second-Order Effects 165
3.2.2 Spatial Process 167
3.3 Point Pattern Analysis Methods 169
3.3.1 Nearest Neighbor Analysis 170
3.3.2 Ripley's *K* Function and the *L* Function Transformation 172
3.3.3 Kernel Density Function 176
3.4 Chapter Concluding Remarks 179
Questions and Answers 180
Lab 3 Spatial Statistics: Measuring Geographic Distributions 183
Overall Progress 183
Scope of the Analysis: Crime Analysis 184
Exercise 3.1 Measuring Geographic Distributions 184

Exercise 3.2 Point Pattern Analysis 191

Exercise 3.3 Kernel Density Estimation 197

Exercise 3.4 Locational Outliers 203

Spatial Autocorrelation 207
Learning Objectives 207
4.1 Spatial Autocorrelation 208
4.2 Global Spatial Autocorrelation 211
4.2.1 Moran's *I* Index and Scatter Plot 211
4.2.2 Geary's C Index 216
4.2.3 General G-Statistic 217
4.3 Incremental Spatial Autocorrelation 219
4.4 Local Spatial Autocorrelation 222
4.4.1 Local Moran's *I* (Cluster and Outlier Analysis) 222
4.4.2 Optimized Outlier Analysis 226
4.4.3 Getis-Ord G_i and G_i^* (Hot Spot Analysis) 227
4.4.4 Optimized Hot Spot Analysis 229
4.5 Space–Time Correlation Analysis 230
4.5.1 Bivariate Moran's *I* for Space–Time Correlation 230
4.5.2 Differential Moran's *I* 232
4.5.3 Emerging Hot Spot Analysis 233
4.6 Multiple Comparisons Problem and Spatial Dependence 234
4.7 Chapter Concluding Remarks 237
Questions and Answers 238
Lab 4 Spatial Autocorrelation 241
Overall Progress 241
Scope of the Analysis 241
Section A ArcGIS 242
Exercise 4.1 Global Spatial Autocorrelation 242
Exercise 4.2 Incremental Spatial Autocorrelation and Spatial Weights Matrix 246
Exercise 4.3 Cluster and Outlier Analysis (Anselin Local Moran's *I*) 252
Exercise 4.4 Hot Spot Analysis (Getis-Ord Gi*I) and Optimized Hot Spot Analysis 255
Exercise 4.5 Optimized Hot Spot Analysis for Crime Events 261
Section B GeoDa 265
Exercises 4.1 and 4.2 Global Spatial Autocorrelation and Spatial Weights Matrix 265
Exercise 4.3 Cluster and Outlier Analysis (Anselin Local Moran's *I*) 270
Exercise 4.4 Hot Spot Analysis (Getis-Ord Gi*I) 273

Multivariate Data in Geography: Data Reduction and Clustering 275
Learning Objectives 275
5.1 Multivariate Data Analysis 276

5.2 Principal Component Analysis (PCA) 279
5.3 Factor Analysis (FA) 289
5.4 Multidimensional Scaling (MDS) 290
5.5 Cluster Analysis 292
5.5.1 Hierarchical Clustering 293
5.5.2 *k*-Means Algorithm (Partitional Clustering) 298
5.6 Regionalization 304
5.6.1 SKATER Method 306
5.6.2 REDCAP Method 309
5.7 Density-Based Clustering: DBSCAN, HDBSCAN, OPTICS 310
5.8 Similarity Analysis: Cosine Similarity 311
5.9 Chapter Concluding Remarks 313
Questions and Answers 314
Lab 5 Multivariate Statistics: Clustering 317
Overall Progress 317
Scope of the Analysis 317
Section A ArcGIS 318
Exercise 5.1 *k*-Means Clustering 318
Exercise 5.2 Spatial Clustering (Regionalization) 329
Exercise 5.3 Similarity Analysis 332
Exercise 5.4 Synthesis 337
Section B GeoDa 345
Exercise 5.1 *k*-Means Clustering 345
Exercise 5.2 Spatial Clustering 348

Modeling Relationships: Regression and Geographically Weighted Regression 351
Learning Objectives 351
6.1 Simple Linear Regression 352
6.1.1 Simple Linear Regression Assumptions 355
6.1.2 Ordinary Least Squares (Intercept and Slope by OLS) 355
6.2 Multiple Linear Regression (MLR) 356
6.2.1 Multiple Regression Basics 356
6.2.2 Model Overfit: Selecting the Number of Variables by Defining a Functional Relationship 359
6.2.3 Missing Values 360
6.2.4 Outliers and Leverage Points 360
6.2.5 Dummy Variables 361
6.2.6 Methods for Entering Variables in MLR: Explanatory Analysis; Identifying Causes and Effects 363

6.3 Evaluating Linear Regression Results: Metrics, Tests and Plots 365
6.3.1 Multiple *r* 365
6.3.2 Variation and Coefficient of Determination *R*-Squared 365
6.3.3 Adjusted *R*-Squared 369
6.3.4 Predicted *R*-Squared 370
6.3.5 Standard Error (Deviation) of Regression (or Standard Error of the Estimate) 371
6.3.6 F-Test of the Overall Significance 373
6.3.7 t-Statistic (Coefficients' Test) 374
6.3.8 Wald Test (Coefficient's Test) 375
6.3.9 Standardized Coefficients (Beta) 376
6.3.10 Residuals, Residual Plots and Standardized Residuals 378
6.3.11 Influential Points: Outliers and High-Leverage Observations 381
6.4 Multiple Linear Regression Assumptions: Diagnose and Fix 383
6.5 Multicollinearity 387
6.6 Worked Example: Simple and Multiple Linear Regression 391
6.7 Exploratory Regression 400
6.8 Geographically Weighted Regression 403
6.8.1 Spatial Kernel Types 405
6.8.2 Bandwidth 406
6.8.3 Interpreting GWR Results and Practical Guidelines 406
6.9 Chapter Concluding Remarks 409
Questions and Answers 411
Lab 6 OLS, Explanatory Regression, GWR 415
Overall Progress 415
Scope of the Analysis 415
Exercise 6.1 Exploratory Regression 416
Exercise 6.2 OLS Regression 428
Exercise 6.3 GWR 437

Spatial Econometrics 451
Learning Objectives 451
7.1 Spatial Econometrics 451
7.2 Spatial Dependence: Spatial Regression Models and Diagnostics 453
7.2.1 Diagnostics for Spatial Dependence 454
7.2.2 Selecting between Spatial Lag or Spatial Error Model 457
7.2.3 Estimation Methods 459

7.3 Spatial Lag Model 459
7.3.1 Spatial Two-Stage Least Squares (S2SLS) 462
7.3.2 Maximum Likelihood 466
7.4 Spatial Error Model 467
7.5 Spatial Filtering 469
7.6 Spatial Heterogeneity: Spatial Regression Models 470
7.7 Spatial Regimes 471
7.8 Chapter Concluding Remarks 474
Questions and Answers 474
Lab 7 Spatial Econometrics 477
Overall Progress 477
Scope of the Analysis 477
Exercise 7.1 OLS 478
Exercise 7.2 Spatial Error Model 487
Exercise 7.3 OLS with Spatial Regimes 491
Exercise 7.4 Spatial Error by Spatial Regimes 501

References 505
Index 513