
Contents

List of Figures	xv
Preface	xxiii
Acknowledgments	xxv
About the Author	xxvii
1 Introduction	1
1.1 Overview of Volume 1	2
1.2 A Quick Tour of GeoDa	4
1.2.1 Data entry	4
1.2.2 Data manipulation	5
1.2.3 GIS operations	5
1.2.4 Weights manager	5
1.2.5 Mapping and geovisualization	5
1.2.6 Exploratory data analysis	6
1.2.7 Space-time analysis	6
1.2.8 Spatial autocorrelation analysis	6
1.2.9 Cluster analysis	7
1.3 Sample Data Sets	7
I Spatial Data Wrangling	9
2 Basic Data Operations	11
2.1 Topics Covered	11
2.2 Spatial Data	12
2.2.1 GIS files	13
2.2.2 Tabular files	17
2.2.3 Other spatial data input	18
2.3 Creating Spatial Layers	19
2.3.1 Point layers from coordinates	19
2.3.2 Grid	22
2.4 Table Manipulations	22
2.4.1 Variable properties	23
2.4.2 Calculator	26
2.4.3 Merging tables	29
2.5 Queries	30
2.5.1 Selection Tool	31
2.5.2 Indicator variable	33
2.5.3 Save selected observations	33
2.5.4 Spatial selection	34

3	GIS Operations	37
3.1	Topics Covered	37
3.2	Projections	38
3.2.1	Coordinate reference system	39
3.2.2	Selecting a projection	39
3.2.3	Reprojection	41
3.3	Converting Between Points and Polygons	41
3.3.1	Mean centers and centroids	42
3.3.2	Tessellations	44
3.4	Minimum Spanning Tree	45
3.4.1	Concept	45
3.4.2	Minimum Spanning Tree options	49
3.5	Aggregation	49
3.5.1	Dissolve	50
3.5.2	Aggregation in table	51
3.6	Multi-Layer Support	52
3.6.1	Loading multiple layers	52
3.6.2	Automatic reprojection	54
3.6.3	Selection in multiple layers	54
3.7	Spatial Join	55
3.7.1	Spatial assign	56
3.7.2	Spatial count	57
3.8	Linked Multi-Layers	58
3.8.1	Specifying an inter-layer linkage	59
3.8.2	Visualizing linked selections	59
II	EDA and ESDA	61
4	Geovisualization	63
4.1	Topics Covered	64
4.2	From EDA to ESDA	64
4.2.1	Exploratory data analysis	65
4.2.2	Mapping as exploration	66
4.2.3	Exploratory spatial data analysis	66
4.2.4	Linking and brushing	66
4.3	Thematic Maps – Overview	67
4.3.1	Choropleth map	67
4.3.2	Map classification	67
4.3.3	Legend and color	68
4.3.4	Implementation	69
4.4	Common Map Classifications	69
4.4.1	Quantile map	69
4.4.2	Equal intervals map	71
4.4.3	Natural breaks map	72
4.5	Map options	73
4.5.1	The map legend	74
4.5.2	The map toolbar	75
4.5.3	Map base layer (base map)	75
4.5.4	Saving the classification as a categorical variable	78
4.5.5	Saving the map as an image	79

4.5.6	Other map options	80
4.6	Custom Classifications	81
4.6.1	Category editor	81
4.6.2	Applications of custom categories	83
4.6.3	Saving the custom categories – the Project File	83
5	Statistical Maps	87
5.1	Topics Covered	87
5.2	Extreme Value Maps	88
5.2.1	Percentile map	88
5.2.2	Box map	89
5.2.3	Standard deviation map	91
5.3	Mapping Categorical Variables	92
5.3.1	Unique values map	93
5.3.2	Co-location map	94
5.4	Cartogram	96
5.4.1	Principle	96
5.4.2	Creating a cartogram	97
5.5	Map Animation	98
6	Maps for Rates	101
6.1	Topics Covered	101
6.2	Choropleth Maps for Rates	102
6.2.1	Spatially extensive and spatially intensive variables	102
6.2.2	Raw rate map	104
6.3	Excess Risk – SMR – LQ	106
6.3.1	Relative risk	106
6.3.2	Excess risk map	106
6.4	Rate Smoothing	108
6.4.1	Variance instability	109
6.4.2	Borrowing strength	109
6.4.3	Empirical Bayes smoothed rate map	112
7	Univariate and Bivariate Data Exploration	115
7.1	Topics Covered	115
7.2	Analyzing the Distribution of a Single Variable	116
7.2.1	Histogram	116
7.2.2	Box plot	121
7.3	Bivariate Analysis – The Scatter Plot	123
7.3.1	Scatter plot basics	123
7.3.2	Smoother option – local regression	128
7.4	Scatter Plot Matrix	130
7.4.1	Implementation	130
7.5	Spatial Heterogeneity	134
7.5.1	Averages chart	134
7.5.2	Brushing the scatter plot	138
8	Multivariate Data Exploration	141
8.1	Topics Covered	141
8.2	The Curse of Dimensionality	142
8.2.1	The empty space phenomenon	143

8.3	Three Variables: Bubble Chart and 3-D Scatter Plot	144
8.3.1	Bubble chart	144
8.3.2	3-D scatter plot	148
8.4	Conditional Plots	151
8.4.1	Implementation	152
8.4.2	Conditional statistical graphs	153
8.4.3	Conditional maps	155
8.5	Parallel Coordinate Plot	156
8.5.1	Implementation	156
8.5.2	Clusters and outliers in PCP	159
9	Space-Time Exploration	161
9.1	Topics Covered	161
9.2	Time Editor	162
9.2.1	Creating grouped variables	162
9.3	Time Player – Comparative Statics	166
9.3.1	Box plot over time	166
9.3.2	Scatter plot with a time-lagged variable	167
9.3.3	Thematic map over time	169
9.4	Treatment Effect Analysis – Averages Chart	171
9.4.1	Difference in means	172
9.4.2	Difference in difference	172
III	Spatial Weights	177
10	Contiguity-Based Spatial Weights	179
10.1	Topics Covered	179
10.2	The Concept of Spatial Weights	180
10.2.1	Spatial weights matrix	181
10.2.2	Types of contiguity	183
10.2.3	Higher-order contiguity	185
10.2.4	Practical considerations	187
10.3	Creating Contiguity Weights	188
10.3.1	Weights manager	188
10.3.2	Rook contiguity	189
10.3.3	Queen contiguity	191
10.3.4	Higher-order contiguity	192
10.3.5	Project file	192
10.3.6	Using existing weights	192
10.3.7	Space-time weights	193
10.4	Weights Characteristics	196
10.4.1	Summary characteristics	196
10.4.2	Connectivity histogram	196
10.4.3	Connectivity map	198
10.4.4	Connectivity graph	199
10.4.5	Connectivity option	200
11	Distance-Based Spatial Weights	203
11.1	Topics Covered	203
11.2	Distance Metrics	204

11.2.1	Distance in a Cartesian coordinate system	204
11.2.2	Great circle distance	205
11.2.3	General distance	205
11.3	Distance-Based Weights	206
11.3.1	Distance-band weights	206
11.3.2	K-nearest neighbor weights	207
11.3.3	General distance weights	208
11.4	Implementation	209
11.4.1	Distance-band weights	211
11.4.2	Isolates	213
11.4.3	K-nearest neighbor weights	215
11.4.4	General distance matrix	216
11.5	Broadening the Concept of Contiguity	217
11.5.1	Contiguity-based weights for points	218
11.5.2	Distance-based weights for polygons	220
11.5.3	Block weights	220
11.5.4	Space-time weights	222
11.6	Manipulating Weights	223
11.6.1	Set operations	223
11.6.2	Make symmetric	224
12	Special Weights Operations	225
12.1	Topics Covered	225
12.2	Spatial Weights as Distance Functions	226
12.2.1	Inverse distance weights	226
12.2.2	Kernel weights	229
12.3	Spatial Transformations	233
12.3.1	Spatially lagged variables	233
12.3.2	Using inverse distance weights – <i>potential</i> measure	236
12.3.3	Using kernel weights	238
12.4	Spatial Weights in Rate Smoothing	239
12.4.1	Spatial rate smoothing	240
12.4.2	Spatial Empirical Bayes smoothing	242
IV	Global Spatial Autocorrelation	245
13	Spatial Autocorrelation	247
13.1	Topics Covered	247
13.2	Spatial Randomness	248
13.2.1	Parametric approach	248
13.2.2	Randomization	249
13.2.3	Simulating spatial randomness	250
13.3	Positive and negative spatial autocorrelation	250
13.3.1	Positive spatial autocorrelation	250
13.3.2	Negative Spatial Autocorrelation	251
13.4	Spatial Autocorrelation Statistic	252
13.4.1	Attribute similarity	252
13.4.2	Locational similarity	253
13.4.3	Examples of spatial autocorrelation statistics	253
13.5	Moran's I	254

13.5.1	Statistic	254
13.5.2	Inference	255
13.5.3	Interpretation	257
13.5.4	Moran scatter plot	259
13.6	Visualizing spatial autocorrelation with the Moran scatter plot	261
13.6.1	Creating a Moran scatter plot	261
13.6.2	Moran scatter plot options	262
13.6.3	Brushing the Moran scatter plot	264
14	Advanced Global Spatial Autocorrelation	267
14.1	Topics Covered	267
14.2	Specialized Moran Scatter Plots	267
14.2.1	Differential Moran scatter plot	268
14.2.2	Moran scatter plot for EB rates	271
14.3	Bivariate Spatial Correlation	273
14.3.1	Bivariate Moran scatter plot	273
14.4	The Many Faces of Space-Time Correlation	276
14.4.1	Serial (temporal) correlation	276
14.4.2	Serial (temporal) correlation between spatial lags	276
14.4.3	Space-time regression	277
14.4.4	Serial and space-time regression	279
15	Nonparametric Spatial Autocorrelation	281
15.1	Topics Covered	281
15.2	Non-Parametric Approaches	282
15.3	Spatial Correlogram	282
15.3.1	Creating a spatial correlogram	283
15.3.2	Spatial correlogram options	285
15.4	Smoothed Distance Scatter Plot	288
15.4.1	Multivariate extension	290
15.4.2	Creating a smoothed distance scatter plot	290
15.4.3	Smoothed distance scatter plot options	291
V	Local Spatial Autocorrelation	295
16	LISA and Local Moran	297
16.1	Topics Covered	297
16.2	LISA Principle	298
16.3	Local Moran	298
16.3.1	Formulation	299
16.3.2	Implementation	300
16.4	Clusters and Spatial Outliers	302
16.4.1	Clusters	302
16.4.2	Spatial outliers	304
16.5	Significance and Interpretation	305
16.5.1	Multiple comparisons	306
16.5.2	Interpretation of significance	309
16.5.3	Interpretation of clusters	310
16.6	Conditional Local Cluster Maps	311

17 Other Local Spatial Autocorrelation Statistics	313
17.1 Topics Covered	313
17.2 Extensions of the Local Moran	314
17.2.1 Median Local Moran	314
17.2.2 Differential Local Moran	318
17.2.3 Local Moran with EB Rate	320
17.3 Local Geary	322
17.3.1 Implementation	323
17.3.2 Clusters and spatial outliers	325
17.3.3 Comparing Local Geary and Local Moran	326
17.4 Getis-Ord Statistics	329
17.4.1 Implementation	329
17.4.2 Clusters and outliers	330
17.4.3 Comparing G statistics and Local Moran	331
17.5 Which Local Statistic to Use?	333
18 Multivariate Local Spatial Autocorrelation	335
18.1 Topics Covered	335
18.2 The Multivariate Spatial Autocorrelation Problem	335
18.3 Bivariate Local Moran	336
18.3.1 Implementation	337
18.3.2 Interpretation	340
18.4 Multivariate Local Geary	340
18.4.1 Implementation	342
18.4.2 Interpretation	344
18.5 Local Neighbor Match Test	345
18.5.1 Implementation	346
18.5.2 Local neighbor match and Multivariate Local Geary	349
19 LISA for Discrete Variables	351
19.1 Topics Covered	351
19.2 Univariate Local Join Count Statistic	351
19.2.1 Implementation	353
19.3 Bivariate Local Join Count Statistic	355
19.3.1 Implementation	356
19.4 Co-Location Local Join Count Statistic	357
19.4.1 Implementation	357
19.5 Quantile LISA	358
19.5.1 Implementation	359
19.5.2 Univariate Quantile LISA	360
19.5.3 Bivariate and Multivariate Quantile LISA	360
20 Density-Based Clustering Methods	363
20.1 Topics Covered	364
20.2 Heat Map	364
20.2.1 Implementation	364
20.3 DBSCAN	366
20.3.1 Important concepts	366
20.3.2 DBSCAN algorithm	368
20.3.3 Implementation	370
20.4 DBSCAN*	373

20.4.1	Mutual reachability distance	373
20.4.2	Implementation	375
20.5	HDBSCAN	376
20.5.1	Important concepts	377
20.5.2	HDBSCAN algorithm	379
20.5.3	Cluster membership	382
20.5.4	Outliers	382
20.5.5	Implementation	383
VI Epilogue		389
21 Postscript – The Limits of Exploration		391
A Appendix A – GeoDa Preference Settings		395
B Appendix B – Menu Structure		399
C Appendix C – Scripting with GeoDa via the geodalib Library		403
Bibliography		405
Index		417