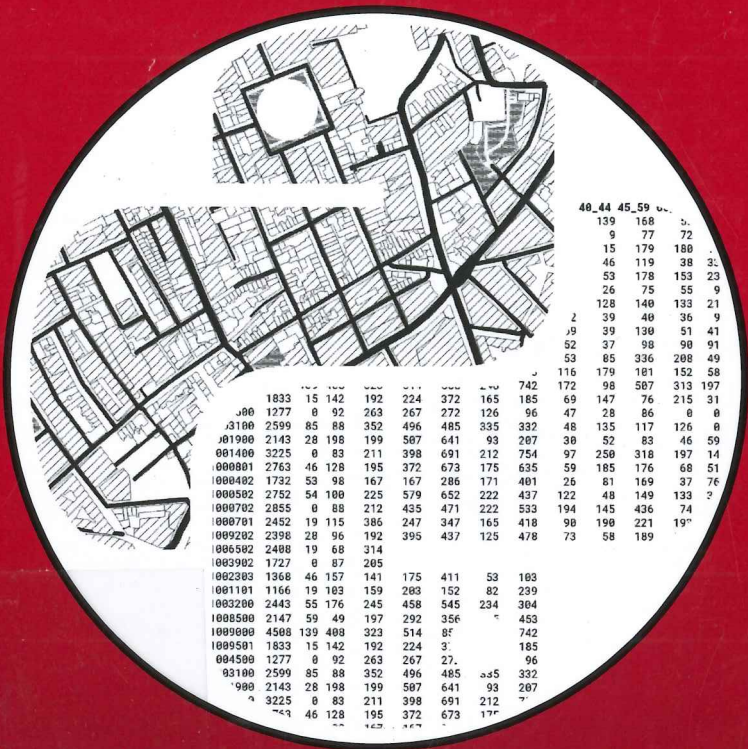


Texts in Statistical Science

Geographic Data Science with Python



สำนักหอสมุด มหาวิทยาลัยเชียงใหม่

b 16704332
o 12576396
i 22684785

Geographic Data Science with Python



By
Sergio Rey, Dani Arribas-Bel and
Levi John Wolf



CRC Press

Taylor & Francis Group

Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business

A CHAPMAN & HALL BOOK

Contents

Preface	xiii
Motivation	xiv
Why this book?	xiv
Who is this book for?	xiv
What this book is not	xv
Content and Purpose of this Book	xv
What we included	xv
What we did not include	xvi
The book in the future	xviii
Acknowledgments	xxi
Author/editor biographies	xxv
List of Figures	xxvii
 I Building Blocks	 1
 1 Geographic Thinking for Data Scientists	 3
1.1 Introduction to geographic thinking	3
1.2 Conceptual representations: models	4
1.3 Computational representations: data structures	5
1.4 Connecting conceptual to computational	10
1.4.1 This categorization is now breaking up (data is data)	10
1.5 Conclusion	11
1.6 Next steps	12
 2 Computational Tools for Geographic Data Science	 13
2.1 Open science	13
2.1.1 Computational notebooks	16
2.1.2 Open source packages	17
2.1.3 Reproducible platforms	17
2.2 The (computational) building blocks of this book	18
2.2.1 Jupyter notebooks and JupyterLab	18
2.2.1.1 Notebooks and cells	19
2.2.1.2 Rich content	20

2.2.1.3	JupyterLab	22
2.2.2	Python and open source packages	24
2.2.2.1	Open source packages	24
2.2.2.2	Contextual help	26
2.2.3	Containerized platform	28
2.2.4	Running the book in a container	29
2.3	Conclusion	32
2.4	Next steps	33
3	Spatial Data	35
3.1	Fundamentals of geographic data structures	36
3.1.1	Geographic tables	36
3.1.2	Surfaces	44
3.1.3	Spatial graphs	48
3.2	Hybrids	51
3.2.1	Surfaces as tables	51
3.2.1.1	One pixel at a time	52
3.2.1.2	Pixels to polygons	55
3.2.2	Tables as surfaces	61
3.2.3	Networks as graphs <i>and</i> tables	62
3.3	Conclusion	64
3.4	Questions	64
4	Spatial Weights	67
4.1	Introduction	67
4.2	Contiguity weights	68
4.2.1	Spatial weights from real-world geographic tables	75
4.2.2	Spatial weights from surfaces	78
4.3	Distance based weights	79
4.3.1	K-nearest neighbor weights	79
4.3.2	Kernel weights	80
4.3.3	Distance bands and hybrid weights	84
4.3.4	Great circle distances	85
4.4	Block weights	86
4.5	Set operations on weights	87
4.5.1	Editing/connecting disconnected observations	88
4.5.2	Using the <code>union</code> of matrices	89
4.6	Visualizing weight set operations	89
4.7	Use case: boundary detection	91
4.8	Conclusion	99
4.9	Questions	99
4.10	Next steps	102

II	Spatial Data Analysis	103
5	Choropleth Mapping	105
5.1	Principles	106
5.2	Quantitative data classification	106
5.2.1	Equal intervals	109
5.2.2	Quantiles	110
5.2.3	Mean-standard deviation	112
5.2.4	Maximum breaks	112
5.2.5	Boxplot	113
5.2.6	Head-tail breaks	114
5.2.7	Jenks-Caspall breaks	115
5.2.8	Fisher-Jenks breaks	115
5.2.9	Max-p	116
5.2.10	Comparing classification schemes	116
5.3	Color	120
5.3.1	Sequential palettes	122
5.3.2	Diverging palettes	122
5.3.3	Qualitative palettes	124
5.4	Advanced topics	126
5.4.1	User-defined choropleths	126
5.4.2	Pooled classifications	128
5.5	Conclusion	130
5.6	Questions	131
5.7	Next steps	131
6	Global Spatial Autocorrelation	133
6.1	Understanding spatial autocorrelation	133
6.2	An empirical illustration: the EU Referendum	135
6.3	Global spatial autocorrelation	138
6.3.1	Spatial lag	140
6.3.2	Binary case: join counts	143
6.3.3	Continuous case: Moran Plot and Moran's I	147
6.3.4	Other global indices	151
6.3.4.1	Geary's C	151
6.3.4.2	Getis and Ord's G	152
6.4	Questions	153
6.5	Next steps	155
7	Local Spatial Autocorrelation	157
7.1	An empirical illustration: the EU Referendum	158
7.2	Motivating local spatial autocorrelation	160
7.3	Local Moran's I_i	163
7.4	Getis and Ord's local statistics	173
7.5	Bonus: local statistics on surfaces	176

7.6	Conclusion	182
7.7	Questions	182
7.8	Next Steps	184
8	Point Pattern Analysis	185
8.1	Introduction	186
8.2	Patterns in Tokyo photographs	186
8.3	Visualizing point patterns	188
8.3.1	Showing patterns as dots on a map	188
8.3.2	Showing density with hexbinning	189
8.3.3	Another kind of density: kernel density estimation	192
8.4	Centrography	193
8.4.1	Tendency	194
8.4.2	Dispersion	196
8.4.3	Extent	198
8.5	Randomness and clustering	203
8.5.1	Quadrat statistics	206
8.5.2	Ripley's alphabet of functions	210
8.6	Identifying clusters	213
8.7	Conclusion	217
8.8	Questions	219
8.9	Next steps	219
III	Advanced Topics	221
9	Spatial Inequality Dynamics	223
9.1	Introduction	223
9.2	Data: U.S. state per capita income 1969-2017	224
9.3	Global inequality	226
9.3.1	20:20 ratio	228
9.3.2	Gini index	230
9.3.3	Theil's index	235
9.4	Personal vs. regional income	236
9.5	Spatial inequality	238
9.5.1	Spatial autocorrelation	238
9.5.2	Regional decomposition of inequality	240
9.5.3	Spatializing classic measures	245
9.6	Conclusion	248
9.7	Questions	250
9.8	Next steps	250
10	Clustering and Regionalization	251
10.1	Introduction	251
10.2	Data	253
10.3	Geodemographic clusters in san diego census tracts	259

10.3.1	K-means	260
10.3.2	Spatial distribution of clusters	261
10.3.3	Statistical analysis of the cluster map	262
10.4	Hierarchical Clustering	268
10.5	Regionalization: spatially constrained hierarchical clustering	272
10.5.1	Contiguity constraint	272
10.5.2	Changing the spatial constraint	274
10.5.3	Geographical coherence	276
10.5.4	Feature coherence (goodness of fit)	278
10.5.5	Solution similarity	279
10.6	Conclusion	280
10.7	Questions	280
10.8	Next steps	281
11	Spatial Regression	283
11.1	What is spatial regression and why should I care?	283
11.2	Data: San Diego Airbnb	285
11.3	Non-spatial regression, a (very) quick refresh	285
11.3.1	Hidden structures	288
11.4	Bringing space into the regression framework	294
11.4.1	Spatial feature engineering: proximity variables	294
11.4.2	Spatial heterogeneity	298
11.4.2.1	Spatial fixed effects	298
11.4.2.2	Spatial regimes	304
11.4.3	Spatial dependence	308
11.4.3.1	Exogenous effects: The SLX model	308
11.4.3.2	Spatial error	315
11.4.3.3	Spatial lag	316
11.4.3.4	Other ways of bringing space into regression	318
11.5	Questions	318
11.5.1	Challenge questions	319
11.5.1.1	The random coast	319
11.5.1.2	The K-neighbor correlogram	321
11.6	Next steps	324
12	Spatial Feature Engineering	325
12.1	What is spatial feature engineering?	326
12.2	Feature engineering using map matching	327
12.2.1	Counting <i>nearby</i> features	328
12.2.2	Assigning point values from surfaces: elevation of Airbnbs	335
12.2.3	Point interpolation using scikit-learn	338
12.2.4	Polygon to point	345
12.2.5	Area to area interpolation	347
12.3	Feature engineering using map synthesis	351
12.3.1	Spatial summary features in map synthesis	352

12.3.1.1	Counting neighbors	352
12.3.1.2	Distance buffers within a single table	353
12.3.1.3	“Ring” buffer features	357
12.3.2	Clustering as feature engineering	359
12.4	Conclusion	363
12.5	Questions	364
12.6	Next steps	364
References		367
Index		375