

Contents

List of figures	xi
List of tables	xvii
List of contributors	xix
Introduction	xxi
Statement by the American Statistical Association on statistical significance and <i>p</i> -value – use in the book	xxiii
Acknowledgements	xxv
1 Basic operations in the R software	1
MATEUSZ KOPYT	
1.1 About the R software	1
1.2 The R software interface	1
1.2.1 R Commander	2
1.2.2 RStudio	3
1.3 Using help	4
1.4 Additional packages	7
1.5 R language – basic features	9
1.6 Defining and loading data	9
1.7 Basic operations on objects	11
1.8 Basic statistics of the dataset	18
1.9 Basic visualisations	24
1.9.1 Scatterplot and line chart	24
1.9.2 Column chart	27
1.9.3 Pie chart	29
1.9.4 Boxplot	29
1.10 Regression in examples	31
2 Data, spatial classes and basic graphics	37
KATARZYNA KOPCZEWSKA	
2.1 Loading and basic operations on spatial vector data	37
2.2 Creating, checking and converting spatial classes	48
2.3 Selected colour palettes	53
2.4 Basic contour maps with a colour layer	57
Scheme 1 – with colorRampPalette() from the grDevices:: package	57
Scheme 2 – with choropleth() from the GISTools:: package	58
Scheme 3 – with findInterval() from the base:: package	59
Scheme 4 – with findColours() from the classInt:: package	60
Scheme 5 – with spplot() from the sp:: package	61
2.5 Basic operations and graphs for point data	62
Scheme 1 – with points() from the graphics:: package – locations only	62

	Scheme 2 – with <code>spplot()</code> from the <code>sp::</code> package – locations and values	63
	Scheme 3 – with <code>findInterval()</code> from the <code>base::</code> package – locations, values, different size of symbols	64
2.6	Basic operations on rasters	67
2.7	Basic operations on grids	73
2.8	Spatial geometries	80
3	Spatial data with Web APIs	87
	MATEUSZ KOPYT AND KATARZYNA KOPCZEWSKA	
3.1	What is an application programming interface (API)?	87
3.2	Creating background maps with use of an application programming interface	88
3.3	Ways to visualise spatial data – maps for point and regional data	102
	Scheme 1 – with <code>bubbleMap()</code> from the <code>RgoogleMaps::</code> package	102
	Scheme 2 – with <code>ggmap()</code> from the <code>ggmap::</code> package	104
	Scheme 3 – with <code>PlotOnStaticMap()</code> from the <code>RgoogleMaps::</code> package	109
	Scheme 4 – with <code>RGoogleMaps::GetMap()</code> and conversion of <code>staticMap</code> into a raster	109
3.4	Spatial data in vector format – example of the OSM database	110
3.5	Access to non-spatial internet databases and resources via application programming interface – examples	117
3.6	Geocoding of data	133
4	Spatial weights matrix, distance measurement, tessellation, spatial statistics	151
	KATARZYNA KOPCZEWSKA AND MARIA KUBARA	
4.1	Introduction to spatial data analysis	151
4.2	Spatial weights matrix	153
	4.2.1 General framework for creating spatial weights matrices	153
	4.2.2 Selection of a neighbourhood matrix	155
	4.2.3 Neighbourhood matrices according to the contiguity criterion	156
	4.2.4 Matrix of k nearest neighbours (<code>knn</code>)	159
	4.2.5 Matrix based on distance criterion (neighbours in a radius of d km)	161
	4.2.6 Inverse distance matrix	163
	4.2.7 Summarising and editing spatial weights matrix	164
	4.2.8 Spatial lags and higher-order neighbourhoods	169
	4.2.9 Creating weights matrix based on group membership	170
	### Example ###	170
	### Example ###	173
4.3	Distance measurement and spatial aggregation	174
	### Example ###	177
4.4	Tessellation	182
4.5	Spatial statistics	185
	4.5.1 Global statistics	188
	4.5.1.1 Global Moran's I statistics	188
	4.5.1.2 Global Geary's C statistics	194
	4.5.1.3 Join-count statistics	195
	4.5.2 Local spatial autocorrelation statistics	199
	4.5.2.2 Local Moran's I statistics (local indicator of spatial association)	199
	4.5.2.3 Local Geary's C statistics	201

4.5.2.4	Local Getis-Ord G_i^* statistics	202
4.5.2.5	Local spatial heteroscedasticity	203
4.6	Spatial cross-correlations for two variables	206
4.7	Correlogram	208
5	Applied spatial econometrics	213
KATARZYNA KOPCZEWSKA		
5.1	Added value from spatial modelling and classes of models	213
5.2	Basic cross-sectional models	216
5.2.1	Estimation	216
	### Example ###	219
5.2.2	Quality assessment of spatial models	230
5.2.2.1	Information criteria and pseudo- R^2 in assessing model fit	230
5.2.2.2	Test for heteroscedasticity of model residuals	232
5.2.2.3	Residual autocorrelation tests	234
5.2.2.4	Lagrange multiplier tests for model type selection	236
5.2.2.5	Likelihood ratio and Wald tests for model restrictions	238
5.2.3	Selection of spatial weights matrix and modelling of diffusion strength	240
5.2.4	Forecasts in spatial models	243
5.2.5	Causality	245
5.3	Selected specifications of cross-sectional spatial models	246
5.3.1	Unidirectional spatial interaction models	246
5.3.2	Cumulative models	255
5.3.3	Bootstrapped models for big data	261
	### Example ###	261
5.3.4	Models for grid data	269
	### Example ###	269
5.4	Spatial panel models	274
	### Example ###	278
6	Geographically weighted regression – modelling spatial heterogeneity	289
PIOTR ĆWIAKOWSKI		
6.1	Geographically weighted regression	289
6.2	Basic estimation of geographically weighted regression model	291
6.2.1	Estimation of the reference ordinary least squares model	291
6.2.2	Choosing the optimal bandwidth for a dataset	292
6.2.3	Local geographically weighted statistics	295
6.2.4	Geographically weighted regression estimation	297
6.2.5	Basic diagnostic tests of the geographically weighted regression model	298
6.2.6	Testing the significance of parameters in geographically weighted regression	304
6.2.7	Selection of the optimal functional form of the model	305
6.2.8	Geographically weighted regression with heteroscedastic random error	307
6.3	The problem of collinearity in geographically weighted regression models	308
6.3.1	Diagnosing collinearity in geographically weighted regression	308
6.4	Mixed geographically weighted regression	316
6.5	Robust regression in the geographically weighted regression model	318
6.6	Geographically and temporally weighted regression	319

7	Spatial unsupervised learning	323
	KATARZYNA KOPCZEWSKA	
7.1	Clustering of spatial points with k-means, PAM (partitioning around medoids) and CLARA (clustering large applications) algorithms	323
	### Example ###	326
	### Example ###	333
7.2	Clustering with the density-based spatial clustering of applications with noise algorithm	336
	### Example ###	337
7.3	Spatial principal component analysis	345
	### Example ###	346
7.4	Spatial drift	349
	### Example ###	349
7.5	Spatial hierarchical clustering	356
	### Example ###	358
	### Example ###	362
7.6	Spatial oblique decision tree	364
	### Example ###	364
8	Spatial point pattern analysis and spatial interpolation	371
	KATERYNA ZABARINA	
8.1	Introduction and main definitions	373
8.1.1	Dataset	373
8.1.2	Creation of window and point pattern	374
8.1.3	Marks	375
8.1.4	Covariates	381
	### Example ###	381
8.1.5	Duplicated points	382
8.1.6	Projection and rescaling	383
8.2	Intensity-based analysis of unmarked point pattern	386
8.2.1	Quadrat test	387
8.2.2	Tests with spatial covariates	388
8.3	Distance-based analysis of the unmarked point pattern	391
8.3.1	Distance-based measures	392
8.3.1.1	Ripley's K function	392
8.3.1.2	F function	393
8.3.1.3	G function	393
8.3.1.4	J function	393
8.3.1.5	Distance-based complete spatial randomness tests	393
8.3.2	Monte Carlo tests	396
8.3.3	Envelopes	396
8.3.4	Non-graphical tests	397
8.4	Selection and estimation of a proper model for unmarked point pattern	398
8.4.1	Theoretical note	399
8.4.2	Choice of parameters	400
8.4.3	Estimation and results	401
8.4.4	Conclusions	404
8.5	Intensity-based analysis of marked point pattern	404
8.5.1	Segregation test	404
8.6	Correlation and spacing analysis of the marked point pattern	405

8.6.1	Analysis under assumption of stationarity	405
8.6.1.1	K function variations for multitype pattern	405
8.6.1.2	Mark connection function	407
8.6.1.3	Analysis of within- and between-type dependence	409
8.6.1.4	Randomisation test of components' independence	409
8.6.2	Analysis under assumption of non-stationarity	410
8.6.2.1	Inhomogeneous K function variations for multitype pattern	410
8.7	Selection and estimation of a proper model for unmarked point pattern	410
8.7.1	Theoretical note	412
8.7.2	Choice of optimal radius	412
8.7.3	Within-industry interaction radius	412
8.7.4	Between-industry interaction radius	414
8.7.5	Estimation and results	415
8.7.6	Model with no between-industry interaction	415
8.7.7	Model with all possible interactions	418
8.8	Spatial interpolation methods – kriging	421
8.8.1	Basic definitions	421
8.8.2	Description of chosen kriging methods	424
8.8.3	Data preparation for the study	424
8.8.4	Estimation and discussion	425
9	Spatial sampling and bootstrapping	433
	KATARZYNA KOPCZEWSKA AND PIOTR ĆWIAKOWSKI	
9.1	Spatial point data – object classes and spatial aggregation	434
9.2	Spatial sampling – randomisation/generation of new points on the surface	437
9.3	Spatial sampling – sampling of sub-samples from existing points	440
9.3.1	Simple sampling	441
9.3.2	The options of the <code>sperrorest::</code> package	443
9.3.3	Sampling points from areas determined by the k -means algorithm – block bootstrap	448
9.3.4	Sampling points from moving blocks (moving block bootstrap)	456
9.4	Use of spatial sampling and bootstrapping in cross-validation of models	462
	### Example ###	462
10	Spatial big data	477
	PIOTR WÓJCIK	
10.1	Examples of big data applications	478
10.2	Spatial big data	478
10.2.1	Spatial data types	479
10.2.2	Challenges related to the use of spatial big data	479
10.2.2.1	Processing of large datasets	479
10.2.2.2	Mapping and reduction	480
10.2.2.3	Spatial data indexing	480
10.3	The <code>sd::</code> package – simple features	481
10.3.1	<code>sf</code> class – a special data frame	481
10.3.2	Data with POLYGON geometry	482
10.3.3	Data with POINT geometry	488
10.3.4	Visualisation using the <code>ggplot2::</code> package	489
10.3.5	Selected functions for spatial analysis	490

10.4	Use the dplyr:: package functions	494
10.5	Sample analysis of large raster data	505
10.5.1	Measurement of economic inequalities from space	505
10.5.2	Analysis using the raster:: package functions	507
10.5.3	Other functions of the raster:: package	514
10.5.4	Potential alternative – stars:: package	515
11	Spatial unsupervised learning – applications of market basket analysis in geomarketing	517
	ALESSANDRO FESTI	
11.1	Introduction to market basket analysis	517
11.2	Data needed in spatial market basket analysis	518
11.3	Simulation of data	520
11.4	The market basket analysis technique applied to geolocation data	526
11.5	Spatial association rules	530
11.6	Applications to geomarketing	534
11.6.1	Finding the best location for a business	534
11.6.2	Targeting	536
11.6.3	Discovery of competitors	538
11.7	Conclusions and further approaches	538
	Appendix A: Datasets used in examples	541
A1.	Dataset no. 1 / dataset1/ – poviats panel data with many variables	541
A2.	Dataset no. 2 / dataset2/ – geolocated point data	544
A3.	Dataset no. 3 / dataset3/ – monthly unemployment rate in poviats (NTS4)	548
A4.	Dataset no. 4 / dataset4/ – grid data for population	549
A5.	Shapefiles of contour maps – for poviats (NTS4), regions (NTS2), country (NTS0) and registration areas	551
A6.	Raster data on night light intensity on Earth in 2013	552
A7.	Population in cities in Poland	553
	Appendix B: Links between packages	555
	References	561
	Index	577