

Table of Contents

Preface

xiii

Part 1: The Essentials of Geospatial Data Science

1

Introducing Geographic Information Systems and Geospatial Data Science

3

What is GIS?	4	Industry and domain knowledge	8
What is data science?	5	Soft skills	9
Mathematics	7	What is geospatial data science?	9
Computer science	7	Summary	10

2

What Is Geospatial Data and Where Can I Find It?

11

Static and dynamic geospatial data	12	Exploring open geospatial data assets	25
Geospatial file formats	12	Human geography	25
Vector data	12	Physical geography	28
Raster data	20	Country- and area-specific data	30
Introducing geospatial databases and storage	24	Summary	30
PostgreSQL and PostGIS	24		
ArcGIS geodatabase	25		

3

Working with Geographic and Projected Coordinate Systems 33

Technical requirements	34	Common types of projected coordinate systems	40
Exploring geographic coordinate systems	35	Working with GCS and PCS in Python	46
Understanding GCS versions	37	PyProj	46
		GeoPandas	46
Understanding projected coordinate systems	39	Summary	53

4

Exploring Geospatial Data Science Packages 55

Technical requirements	55	Packages for producing production-quality spatial visualizations	81
Packages for working with geospatial data	56	ipyLeaflet	81
GeoPandas	56	Folium	84
GDAL	67	geoplot	86
Shapely	67	GeoViews	88
Fiona	72	Datashader	90
Rasterio	72	Reviewing foundational data science packages	90
Packages enabling spatial analysis and modeling	78	pandas	90
PySAL	79	scikit-learn	96
		Summary	96

Part 2: Exploratory Spatial Data Analysis

5

Exploratory Data Visualization 99

Technical requirements	100	The fundamentals of ESDA	100
		Example – New York City Airbnb listings	101

Conducting EDA	102	Summary	114
ESDA	106		

6

Hypothesis Testing and Spatial Randomness 115

Technical requirements	116	Local spatial autocorrelation	127
Constructing a spatial hypothesis test	116	Point pattern analysis	134
Understanding spatial weights and spatial lags	117	Ripley's alphabet functions	135
Global spatial autocorrelation	120	Summary	142

7

Spatial Feature Engineering 143

Technical requirements	144	Summary spatial features using two datasets	150
Defining spatial feature engineering	144	Engineering proximity spatial features	152
Performing a bit of geospatial magic	145	Proximity spatial features – NYC attractions	154
Engineering summary spatial features	146	Summary	161
Summary spatial features using one dataset	146		

Part 3: Geospatial Modeling Case Studies

8

Spatial Clustering and Regionalization 165

Technical requirements	166	Agglomerative hierarchical geodemographic clustering	184
Collecting geodemographic data for modeling	166	Spatially constrained agglomerative hierarchical geodemographic clustering	187
Extracting data using the Census API	166	Measuring model performance	189
Cleaning the extracted data	170	Summary	192
Conducting EDA and ESDA	171		
Developing geodemographic clusters	177		
K-means geodemographic clustering	178		

9

Developing Spatial Regression Models 193

Technical requirements	194	Introduction to GWR models	215
A refresher on regression models	194	Fitting a GWR model to predict nightly Airbnb prices	215
Constructing an initial regression model	194	Introduction to Multiscale Geographically Weighted Regression	217
Exploring unmodeled spatial relationships	201	Fitting an MGWR model to predict nightly Airbnb prices	217
Teaching the model to think spatially	208	How do I choose between these models?	218
Incorporating spatial fixed effects within the model	212	Summary	219

10

Developing Solutions for Spatial Optimization Problems 221

Technical requirements	222	Setting up the Google Maps API	234
Exploring the Location Set Covering Problem (LSCP)	222	Solving the TSP	234
Understanding the math behind the LSCP	223	Exploring a single-vehicle Vehicle Routing Problem (VRP)	241
Solving LSCPs	224	Exploring a Capacitated Vehicle Routing Problem (CVRP)	245
Exploring route-based combinatorial optimization problems	231	Summary	247
Understanding the math behind the TSP	231		

11

Advanced Topics in Spatial Data Science 249

Technical requirements	249	Estimating unknowns with spatial interpolation	258
Efficient operations with spatial indexing	249	Applying Inverse Distance Weighted (IDW) interpolation	258
Implementing R-tree indexing in GeoPandas	251	Introduction to Kriging-based interpolation	264
Introducing the H3 spatial index	254		

Ethical spatial data science	268	Example 3 – COVID-19 contact tracing	269
Example 1 – Sharpiegate	268	Example 4 – United States Census Bureau disclosure avoidance system	269
Example 2 – Human mobility: The New York Times investigative report	269	Summary	270
Index			273
Other Books You May Enjoy			284