Table of Contents

Preface

Part 1: The Essentials of Geospatial Data Science

Introducing Geographic Information Systems and Geospatial Data Science

4

5

What is GIS?	
What is data science?	
Mathematics	
Computer science	

Industry and domain knowledge	8
Soft skills	9
What is geospatial data science?	9
Summary	10

ArcGIS geodatabase

What Is Geospatial Data and Where Can I Find It?

24

25

xiii

3

Static and dynamic geospatial data	12
Geospatial file formats	12
Vector data	12
Raster data	20
Introducing geospatial databases and	
storage	24
PostgreSQL and PostGIS	24

Exploring open geospatial data assets	
Human geography	25
Physical geography	28
Country- and area-specific data	30
Summary	30

3

Working with Geographic and Projected Coordinate Systems 33

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

4

5

Exploring Geospatial Data Science Packages 55 **Technical requirements** 55 Packages for producing productionquality spatial visualizations 81 Packages for working with geospatial data ipyLeaflet 81 56 84 Folium GeoPandas 56 geoplot 86 GDAL 67 GeoViews 88 67 Shapely Datashader 90 Fiona 72 Rasterio 72 Reviewing foundational data science packages 90 Packages enabling spatial analysis and modeling 90 78 pandas scikit-learn 96 PySAL 70

VSAL	/9	

Summary

96

Part 2: Exploratory Spatial Data Analysis

Exploratory Data Visualiz	ation	et	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	Ran	domness	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

Spatial Feature Engineering

Technical requirements	
Defining spatial feature engineering	144
Performing a bit of geospatial magic	145
Engineering summary spatial features	146
Summary spatial features using one dataset	146

Summary spatial features using two datasets	150
Engineering proximity spatial feature	s152
Proximity spatial features - NYC attractions	154
Summary	161

Part 3: Geospatial Modeling Case Studies

166

170

171

Spatial Clustering and Regionalization

Technical requirements166Collecting geodemographic data for
modeling166

Extracting data using the Census API Cleaning the extracted data Conducting EDA and ESDA

Developing geodemographic clusters 177 K-means geodemographic clustering 178

Agglomerative hierarchical geodemographic	c
clustering	184
Spatially constrained agglomerative	
hierarchical geodemographic clustering	187
Measuring model performance	189
Summary	192

165

143

9

Developing Spatial Regression Models		193	
Technical requirements	194	Introduction to GWR models	215
A refresher on regression models Constructing an initial regression model	194 194	Fitting a GWR model to predict nightly Airbnb prices	215
Exploring unmodeled spatial relationships	201	Introduction to Multiscale Geographically Weighted Regression	217
Incorporating spatial fixed effects within the	208	Fitting an MGWR model to predict nightly Airbnb prices	217
model	212	How do I choose between these models?	218
		Summary	219

10

1

Developing Solutions for Spatial Optimization Problems			221
Technical requirements	222	Setting up the Google Maps API	234
Exploring the Location Set Covering		Solving the TSP	234
Problem (LSCP)	222	Exploring a single-vehicle Vehicle Routing	
Understanding the math behind the LSCP Solving LSCPs	223 224	Problem (VRP) Exploring a Capacitated Vehicle Routing	241
		Problem (CVRP)	245
Exploring route-based combinatorial optimization problems	231	Summary	247
TT I I II II II II II MOD	001		

251

Understanding the math behind the TSP 231

Advanced Topics in Spatial Data Science

Technical requirements Efficient operations with spatial indexing

Implementing R-tree indexing in GeoPandas Introducing the H3 spatial index

249 Estimating unknowns with spatial interpolation 258 Applying Inverse Distance Weighted (IDW) 249 interpolation 258 Introduction to Kriging-based interpolation 264 254

249

284

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

Other Books You May Enjoy

extens the globe in near real time, and this map was likely the first time must the second were expression to the part of special montrains special door visualization, and spatial data science. How every spatial advantances a map of choices deaths in London within ethewei him to show that chok as was press because grants in water wells and cont through minutes in the site of a moury thright dinter that the

the start of the second second of the start of the start of the start of the start of the second sec

the new altor leving through a time in which more and more people are willing to stand up for equity call out inequalies when they see them. It die United States and serves the world, teams of are researching high-proble inequifies in arms of the global food aupply, healthcare areas, people, enough a proble inequifies in arms of the global food aupply, healthcare areas, are researching high proble inequifies in arms of the global food aupply, healthcare areas, are researching to the order of this kind are people are the future generations

The base the data model and technology to begin to do with thing about each of these problems, and a surprise and the actence have the potential to provide moreous value in helping as and interpretermine realismus planeting, and better educate conservers and there around us However, the relation perform realismus planeting, and better educate conservers and there around us However, the relation perform realismus planeting, and better educate conservers and there around us However, the relation perform realismus planeting, and producing compelling visualizations to new easier then even it is not which in 1995. By nature, maps and producing compelling visualizations to new easier then even it is not a strain realization in producing compelling visualizations to new easier then even it is not a manipulation happened in an event that has ance been dailed. Sharple gain, is which which is not a manipulation happened in an event that has ance been dailed. Sharple gain, is which is realized to the date of the manipulation event that has an event with a filterial of spatial and the realized to the date of the manipulation of the truth can be date of the realized to the date of the manipulation of the truth are an event to spatial and the state community. While this manipulation are real worked, there are many manipulate to spatial and the spatial comparison which the manipulation is be even of as a burgeneous provid data sciences.