

# SPATIAL DATA MANAGEMENT WITH DUCKDB

From SQL Basics to Advanced  
Geospatial Analytics



Qiusheng Wu

# Spatial Data Management with DuckDB

From SQL Basics to Advanced Geospatial  
Analytics

Qiusheng Wu  
2025

# Contents

<b>Preface .....</b>	<b>1</b>
Introduction .....	3
Who This Book Is For .....	3
What This Book Covers .....	5
Getting the Most Out of This Book .....	6
Conventions Used in This Book .....	7
Downloading the Code Examples .....	7
Video Tutorials and Supplementary Resources .....	8
Community and Feedback .....	8
Acknowledgments .....	9
About the Author .....	9
Licensing and Copyright .....	10
<b>I: DuckDB Foundations .....</b>	<b>11</b>
<b>1. Getting Started with DuckDB .....</b>	<b>13</b>
1.1. Introduction .....	13
1.2. Learning Objectives .....	13
1.3. What Makes DuckDB Different from Traditional Databases .....	13
1.4. When (and When Not) to Use DuckDB for Spatial Work .....	15
1.5. Installing DuckDB CLI and Running Your First Query .....	16
1.6. Installing the DuckDB Python Client .....	18
1.7. Installing Visual Studio Code .....	21
1.8. Using the DuckDB UI .....	24
1.9. Installing DBeaver SQL IDE .....	25
1.10. Key Takeaways .....	29
1.11. Exercises .....	30
<b>2. Essential SQL for Spatial Analysis .....</b>	<b>33</b>
2.1. Introduction .....	33
2.2. Learning Objectives .....	33
2.3. Sample Datasets .....	34
2.4. Setting Up the Environment .....	34
2.5. Connecting To DuckDB .....	35
2.6. Install Extensions .....	35
2.7. Reading CSV Files from URLs .....	36
2.8. Creating Tables for Better Performance .....	37
2.9. The SQL SELECT Statement .....	38
2.10. Filtering Data with the WHERE Clause .....	45
2.11. Pattern Matching with LIKE .....	47
2.12. The IN Operator .....	48
2.13. The BETWEEN Operator .....	49
2.14. Combining Data with SQL Joins .....	50
2.15. Query Plans and Performance .....	56
2.16. Aggregating Data for Summary Statistics .....	56
2.17. Conditional Statements .....	60
2.18. Saving Results .....	61
2.19. Advanced SQL Features for Spatial Analysis .....	62
2.20. SQL Comments and Documentation .....	66

2.21. Key Takeaways .....	67
2.22. Exercises .....	68
<b>3. DuckDB Python Integration .....</b>	<b>72</b>
3.1. Introduction .....	72
3.2. Learning Objectives .....	72
3.3. Sample Datasets .....	73
3.4. Installation and Setup .....	74
3.5. Installing and Loading Extensions .....	75
3.6. Reading Data from Multiple Sources .....	76
3.7. Seamless Integration with Pandas DataFrames .....	79
3.8. Polars Interoperability .....	82
3.9. Result Conversion and Output Formats .....	82
3.10. Writing Data To Disk .....	85
3.11. Persistent Storage and Database Files .....	85
3.12. Prepared Statements and Parameters .....	88
3.13. Key Takeaways .....	88
3.14. Exercises .....	90
<b>II: Spatial Data Operations .....</b>	<b>95</b>
<b>4. Loading Spatial Data Formats .....</b>	<b>97</b>
4.1. Introduction .....	97
4.2. Learning Objectives .....	98
4.3. Sample Datasets .....	98
4.4. Installation and Setup .....	99
4.5. Installing and Loading Extensions .....	100
4.6. Downloading Sample Data .....	101
4.7. Loading CSV Files with Coordinates .....	101
4.8. Loading JSON Files .....	104
4.9. Querying Pandas DataFrames Directly .....	106
4.10. Loading Parquet Files for Performance .....	108
4.11. Loading GeoJSON Files with Spatial Geometries .....	110
4.12. Loading Shapefiles into Modern Workflows .....	116
4.13. Loading GeoParquet for Cloud-Native Spatial Analysis .....	118
4.14. Data Loading Performance Strategies .....	120
4.15. Troubleshooting Common Data Loading Issues .....	121
4.16. Key Takeaways .....	121
4.17. Exercises .....	123
<b>5. Exporting and Converting Spatial Data .....</b>	<b>126</b>
5.1. Introduction .....	126
5.2. Learning Objectives .....	127
5.3. Sample Datasets .....	128
5.4. Installation and Setup .....	128
5.5. Installing and Loading Extensions .....	129
5.6. Loading Sample Data .....	130
5.7. Exporting to Pandas DataFrames .....	132
5.8. Exporting to CSV Files .....	136
5.9. Exporting to JSON Files .....	140
5.10. Exporting to Excel Files .....	142
5.11. Exporting to Parquet Files .....	145

5.12. Exporting to GeoJSON Format .....	147
5.13. Exporting to Shapefile Format .....	151
5.14. Exporting to GeoPackage Format .....	152
5.15. Key Takeaways .....	154
5.16. Exercises .....	155
<b>6. <i>Geometry Operations and Functions</i></b> .....	<b>160</b>
6.1. Introduction .....	160
6.2. Learning Objectives .....	160
6.3. Sample Datasets .....	161
6.4. Installation and Setup .....	162
6.5. Connecting to DuckDB and Loading Extensions .....	163
6.6. Understanding Geometry Types .....	165
6.7. Working with Points .....	167
6.8. Working with Linestrings .....	171
6.9. Working with Polygons .....	172
6.10. Working with Collections .....	174
6.11. Visualizing NYC Spatial Data .....	175
6.12. Geometry Processing .....	178
6.13. Geometry Validity and Robustness .....	181
6.14. Key Takeaways .....	183
6.15. Exercises .....	185
<b>7. <i>Spatial Queries and Relationships</i></b> .....	<b>189</b>
7.1. Introduction .....	189
7.2. Learning Objectives .....	190
7.3. Sample Datasets .....	190
7.4. Installation and Setup .....	191
7.5. Connecting to DuckDB and Loading Extensions .....	191
7.6. Understanding Spatial Relationships .....	192
7.7. Testing Geometric Identity .....	193
7.8. Topological Relationships .....	195
7.9. Distance-Based Relationships .....	203
7.10. Filtering by Proximity Threshold .....	205
7.11. Nearest Neighbor Queries .....	206
7.12. Key Takeaways .....	207
7.13. Exercises .....	208
<b>8. <i>Advanced Spatial Joins</i></b> .....	<b>211</b>
8.1. Introduction .....	211
8.2. Learning Objectives .....	211
8.3. Sample Datasets .....	212
8.4. Installation .....	213
8.5. Library Import and Configuration .....	213
8.6. Connecting to DuckDB .....	213
8.7. Intersection Joins .....	214
8.8. Distance Within Joins .....	219
8.9. Advanced Join .....	223
8.10. Coordinate System Transformations .....	225
8.11. Spatial Relationship Functions Reference .....	227
8.12. Key Takeaways .....	228

8.13. Exercises .....	230
<b>9. <i>Interactive Data Visualization</i> .....</b>	<b>234</b>
9.1. Introduction .....	234
9.2. Learning Objectives .....	234
9.3. Sample Datasets .....	235
9.4. Installation and Setup .....	235
9.5. Downloading Sample Data .....	236
9.6. Connecting to DuckDB and Loading Extensions .....	236
9.7. Visualizing Point Data .....	237
9.8. Visualizing Line Data .....	240
9.9. Visualizing Polygon Data .....	243
9.10. Key Takeaways .....	248
9.11. Exercises .....	250
<b>10. <i>Working with Vector Tiles and PMTiles</i> .....</b>	<b>254</b>
10.1. Introduction .....	254
10.2. Learning Objectives .....	255
10.3. Sample Datasets .....	256
10.4. Installation and Setup .....	256
10.5. Visualizing Vector Tiles Directly from Files .....	257
10.6. Converting Vector Data to PMTiles .....	261
10.7. Visualizing PMTiles .....	263
10.8. Key Takeaways .....	272
10.9. Exercises .....	274
<b>III: Real-World Geospatial Analytics .....</b>	<b>279</b>
<b>11. <i>Analyzing US National Wetlands Inventory</i> .....</b>	<b>281</b>
11.1. Introduction .....	281
11.2. Learning Objectives .....	282
11.3. Datasets Used in This Chapter .....	282
11.4. Understanding the Dataset Source .....	285
11.5. Accessing Wetland Data with DuckDB .....	287
11.6. Visualizing Wetland Distributions .....	290
11.7. National-Scale Wetland Analysis .....	294
11.8. Key Takeaways .....	306
11.9. Exercises .....	308
<b>12. <i>Analyzing Global Building Footprints</i> .....</b>	<b>311</b>
12.1. Introduction .....	311
12.2. Learning Objectives .....	312
12.3. About the Dataset .....	313
12.4. Installation and Setup .....	315
12.5. Installing and Loading Extensions .....	316
12.6. Exploring Available Data .....	317
12.7. Single Country Analysis .....	318
12.8. Multi-Country Comparison .....	323
12.9. Confidence Scores and Quality Filters .....	325
12.10. Administrative Subdivision Analysis .....	326
12.11. Urban Density Analysis .....	326
12.12. Continental-Scale Analysis .....	327
12.13. Building Characteristics Analysis .....	328

12.14.	H3-Based Aggregation (Optional) .....	329
12.15.	Joining with Population Data .....	330
12.16.	Performance Considerations .....	331
12.17.	Key Takeaways .....	332
12.18.	Exercises .....	334
12.19.	References .....	338
<b>13.</b>	<b>Analyzing NYC Taxi Data .....</b>	<b>339</b>
13.1.	Introduction .....	339
13.2.	Learning Objectives .....	340
13.3.	About the Dataset .....	340
13.4.	Installation .....	342
13.5.	Library Import .....	343
13.6.	Installing and Loading Extensions .....	343
13.7.	Loading Taxi Data .....	345
13.8.	Temporal Analysis .....	347
13.9.	Spatial Analysis .....	355
13.10.	Trip Flow Analysis .....	360
13.11.	Payment and Economic Analysis .....	363
13.12.	Passenger Behavior Analysis .....	365
13.13.	Multi-Month Analysis .....	366
13.14.	Visualization .....	366
13.15.	Performance Optimization Tips .....	367
13.16.	Key Takeaways .....	368
13.17.	Exercises .....	370
13.18.	References .....	371
<b>14.</b>	<b>Developing Interactive Dashboards with Streamlit and Solara .....</b>	<b>372</b>
14.1.	Introduction .....	372
14.2.	Learning Objectives .....	372
14.3.	Why Dashboards for Geospatial Analysis? .....	372
14.4.	Dashboard Framework Comparison .....	373
14.5.	Building a Streamlit Geospatial Dashboard .....	373
14.6.	Building a Solara Dashboard .....	381
14.7.	Dashboard Best Practices .....	384
14.8.	Deployment Options .....	385
14.9.	Key Takeaways .....	386
14.10.	Exercises .....	386
14.11.	Exercises .....	386
14.12.	References .....	387

# Revolutionize Your Geospatial Data Analysis with DuckDB

Transform how you work with geospatial data using **DuckDB**, the revolutionary analytical database that makes spatial analysis faster, simpler, and more accessible than ever before. This comprehensive guide takes you from SQL fundamentals to advanced geospatial analytics, empowering you to tackle real-world spatial challenges and create scalable analytics pipelines.

## What You Will Learn

- Master DuckDB fundamentals
- Work with diverse data formats
- Perform spatial operations
- Conduct advanced spatial analysis
- Create interactive visualizations
- Handle large-scale datasets
- Integrate with the Python ecosystem
- Apply real-world solutions
- Build production workflows



Scan the QR code to visit the book website and download code examples from <https://duckdb.github.org>