Analyzing Geo-Business Data

A Hands-On Case Study
in Mapping, Spatial Analysis and Data Mining

February, 2003

Joseph K. Berry

W.M. Keck Scholar in Geosciences
University of Denver
Denver, Colorado

Table of Contents

Forward *by Grant Thrall* ……..…… *v*

Preface ……..…….. viii
Book Organization
Companion Software
Case Study Data Set
Acknowledgements

1. Overview of the Case Study …….. 1
1.1 Introduction
1.2 Mapping versus Analysis
1.3 Case Study
1.4 Exercises
Viewing Full Color Figures
Installing ArcExplorer Software
Installing Surfer Software
Installing MapCalc Software
Installing SnagIt Software
Example of Analysis Capabilities

2. Desktop Mapping …….. 11
2.1 Mapping Basics
2.2 Accessing and Displaying Map Data
2.3 Querying a Map
2.4 Vector Data Characteristics
2.5 Analyzing Vector Maps
2.6 Exercises
Interacting with Vector Maps
Vector-based Thematic Mapping
Using Pseudo-Grids

3. Grid-Based Mapping …….. 21
3.1 Raster Data Characteristics
3.2 Visualizing Grid-based Data
3.3 Normalizing and Exchanging Data
3.4 Grid Data Organization
3.5 Exercises
Interacting with Grid Maps
Grid-based Thematic Mapping
Map Summary Statistics and Charts
Map Normalization

4. Spatial Analysis ……..…….. 35
4.1 Classes of Analysis Operations
4.2 Creating Travel-Time Maps
4.3 Transferring Grid and Vector Data
4.4 Creating Visual Exposure Maps
4.5 Exercises
Calculating Simple Proximity
Creating Travel-Time Maps

Transferring Maps
Creating Visual Exposure Maps

5. Analyzing Location 55
5.1 Identifying Competition Zones
5.2 Spatially Characterizing Customer Loyalty
5.3 Connecting with Customers
5.4 Suitability Modeling
5.5 Exercises
Competition Analysis
Connecting With Customers
Suitability Modeling

6. Surface Modeling 71
6.1 Identifying Customer Pockets
6.2 From Point Samples to Map Surfaces
6.3 The Keystone Concept
6.4 IDW and Krig Techniques
6.5 Benchmarking Interpolation Results
6.6 Assessing Interpolation Performance
6.6 Exercises
Deriving Customer Density
Identifying Customer Territories
Creating Interpolated Surfaces
Assessing Interpolation Performance
Using Surfer for Surface Modeling

7. Spatial Data Mining 89
7.1 Characterizing Data Groups
7.2 Calculating Map Similarity
7.3 Identifying Data Zones
7.4 Mapping Data Clusters
7.5 Exporting Mapped Data
7.4 Exercises
Characterizing Map Similarity
Identifying Data Zones
Mapping Data Clusters
Exporting Data

8. Predictive Modeling 103
8.1 Examples of Predictive Modeling
8.2 Spatial Regression
8.3 Assessing Prediction Model Results
8.4 Stratifying Maps for Better Predictions
8.5 Conclusion
8.6 Exercises
Deriving a Dependent Map Variable
Scatter Plots of map Correlation
Spatial Regression
Evaluating Regression Performance

Appendix A. Quick Set of Exercises 113

Appendix B. Creating Your Own Geo-Business Database 125

Preface

"Make strategies fit the situation; don't try to fit the situation to the strategy."
(General George C. Patton)

For years business generally ignored spatial relationships in the data used for making decisions. While common sense considers "location, location, location" a cornerstone of business, traditional analysis procedures forced information, such as customer location, to be aggregated into large generalized reporting units. Pockets of sales on one-side of a county were not differentiated from those on another side.

In a sense, it is like a field general knowing he has 30,000 troops but not sure where they are. Understanding the spatial implications of information is critical for both strategic and tactical levels of planning and operations. The shear magnitude and complexity of spatial information, however, overwhelmed early attempts at its quantitative analysis.

Not until recently did Geographic Information Systems (GIS) technology spur business applications of desktop mapping. The ability to geo-query databases and visualize the results in a variety of thematic map forms is developing an appreciation of "where" as well as "what" in business solutions.

Vector-based desktop mapping is rapidly becoming part of the modern business environment. The close link between these systems and traditional spreadsheet and database management programs has fueled the adoption. In many ways, a "database is simply its picture waiting to happen." The direct link between the attributes described in a database record and its spatial characterization is conceptually easy. Geo-query by clicking on a map to pop-up the attribute record or searching a database then plotting the selected records are extremely useful extensions of traditional database technology. Couple decreasing desktop mapping system costs and complexity with increasing data availability and Internet access makes the adoption of spatial database technology a "no-brainer." Maps in their traditional form of point, lines and polygons identifying discrete spatial objects align with manual mapping concepts and experiences learned as early as girl and boy scouts.

However, the infusion of grid-based map analysis in geo-business has not been as widespread. While vector-based approaches are ideal for spatial database management, grid-based analysis supports mathematical/statistical approaches for quantifying spatial relationships within and among maps that requires a different paradigm of geographic space. Whereas traditional maps emphasize "precise placement of physical features," grid maps seek to "statistically characterize continuous space in both real and cognitive terms." The tools for mapping database attributes are extended to analysis of spatial relationships.

This book focuses on the underlying concepts, considerations and procedures in grid-based data handling and analysis operations as they apply to geo-business applications. *Analyzing Geo-Business Data* provides the blend of theory and practical experience that is needed to "cross the chasm" from simply mapping to map analysis.

Book Organization

Analyzing Geo-Business Data is organized into eight topics and two appendices that lead the reader from an understanding of the fundamental nature of mapped data through a series of basic procedures used in deriving, analyzing and applying spatial information. A case study approach is used with each topic area describing the application of a set of related analysis techniques. The discussion is followed by a series of hands-on exercises providing practical experience in applying the techniques. The exercises include step-by-step instructions that are thoroughly annotated.

Companion Software

SnagIt software by TechSmith is used for screen capture and creation of composite presentation graphics. A 45-day evaluation version of the software is included with this book. Information about the full *SnagIt* package is available at ***www.techsmith.com/***.

ArcExplorer software by ESRI is used for display and query of GIS data and used in vector-based desktop mapping exercises. The free software is included with this book. Information about the full line of GIS software by ESRI is available ***www.esri.com/***.

Surfer software by Golden Software is used for spatial interpolation and advanced graphics. A demonstration version of the software is included with this book. Information about the full *Surfer* package is available at ***www.ssg-surfer.com/***.

MapCalc software by Red Hen Systems is used for grid-based map analysis exercises throughout the book. A 14-day evaluation version of MapCalc is included with this book. The software includes the basic set of data and operations needed to complete the hands-on exercises; **the evaluation version expires after a two-week period**.

The full *MapCalc Learner* software for individual use contains additional materials and provides for extended experience in map analysis procedures and applications. It contains a basic set of functions for import/export of your own data that is constrained to a field-level (100row x 100col) analysis grid configuration. *MapCalc Academic* is a full, multi-seat licensed educational version for classroom/lab use and contains an additional instructor CD with supporting teaching materials including PowerPoint slides, exercises and quizzes for a several workshop and college-level course offerings. *MapCalc Professional* is licensed for commercial use and includes a full range of features including image data routines, coordinate/datum transformation and an extended set of import/export formats for data exchange.

For more information on the Learner, Academic and Professional versions of MapCalc, visit ***www.redhensystems.com/mapcalc/***. Pricing for the MapCalc versions is US$21.95 for Learner, $495 for Academic and $695.00 for Professional plus shipping and handling.
(February, 2003; prices subject to change)

Case Study Data Set

The geographic features of case study area are based on a small Colorado city; however all of the business data is hypothetical. As the result of the mixing of actual and hypothetical information the case study data set is referred to as Smallville, USA. While the analytical procedures described are valid, the outcomes do not describe businesses or interests in the actual community. Any similarities are purely coincidental.

Intended Audience

This book is ideal for professionals and students interested in a basic understanding of the concepts, procedures and considerations in analyzing geo-business data. The material is presented in an informal manner designed so the reader can grasp the broad issues and then delve into hands-on exercises for practical experience in applying the techniques. In a classroom setting, portions of the book can be used for an exposure to grid-based map analysis in an introductory GIS course, or in its entirety for an intermediate course on map analysis with individual/group projects.

Internet Extended Environment

There are several benefits for readers with a self-published book. Traditionally published texts have inordinately slow publishing cycles compared to the speed of software development and application

innovation. This book is printed in small batches and frequently revised to keep the material current. The Internet is a great mechanism for feedback so the book can evolve *with your help*. Please help in identifying sections that need correction or clarification, as well as providing suggestions for extending the current topics or adding entirely new ones—email the author at ***jberry@innovativegis.com***.

In addition, the author's website at ***http://www.innovativegis.com/basis/*** serves as a mechanism for updates, enhancements, extensions and related articles. It is recommended that you periodically check this website for up-to-the-minute information pertaining to the book. Notices and updates for the companion MapCalc software are posted at ***www.redhensystems.com/mapcalc/***.

Acknowledgements

Preparation of a book as complete as this one requires input from several individuals and organizations. Collaboration with *Dr. Kenneth L. Reed* with Accenture Technology Solutions over many years sparked several of the map analysis approaches described in this book. This enduring friendship and interaction dates back to the 1970s when we were junior faculty at Yale University and GIS technology was just a glint in researchers' eyes. Adams Business Media, publishers of *GeoWorld* has supported the "Beyond Mapping" column for over a decade. Several of the topics presented were first introduced through this authoritative resource for spatial information. *Red Hen Systems* provided a critical ingredient—powerful yet inexpensive educational software that is very easy to use. Supporting software by *Golden Software*, *Environmental System Research Institute* and *TechSmith* greatly extend the grid-based exercises and enhance their utility.

Many Thanks!

Joseph K. Berry
University of Denver
Denver, Colorado
May 2003