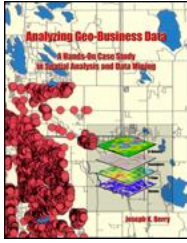

Analyzing Geo-Business Data

**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*



Analyzing Geo-Business Data

Copyright

Copyright © 2003. Joseph K. Berry. All rights reserved.

Trademarks

MapCalc Learner™, MapCalc Academic™ and MapCalc Professional™ are registered trademarks of Red Hen Systems, Incorporated. ArcExplorer™ is a registered trademark of Environmental Systems Research Institute, Incorporated. Surfer™, is a registered trademark of Golden Software, Incorporated. SnagIt™, is a registered trademark of TechSmith, Incorporated.

Publisher

BASIS Press, 2000 South College Avenue, Suite 300
Fort Collins, Colorado, 80525, USA.

For book inquires, please contact the following website (that also contains additional information on spatial data analysis and applications):

<http://www.innovativegis.com/basis/>

Published in the United States.

Edition

1st Edition (February 2003)

ISBN **xxxxxx** (pending)

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

Forward

Business geography aims to improve business decisions through the development and application of geographic reasoning, geographic analysis, and geographic technology. Without improvement of the business judgmental decision, there is no business geography.

Business Geography has not been a crowded playing field, so it is easy to identify relevant intellectual forebears to the discipline. Some might trace business geography back to German geographer, Johann Heinrich von Thünen, as he was the first to perform “what if” scenarios with a map (see Thrall, 2002, ch 3). Thünen recognized that a map was a mathematical abstraction whose parameters and assumptions could be changed, and the map could display the consequences of those changes. The geographic sciences owe an enormous debt to Thünen as contemporary geographic information analysis can be legitimately claimed to be traced directly to his insights at the end of the 18th century. What-if scenarios are today well integrated into the thinking of our society – from spreadsheets to maps. While Thünen’s work did improve the business decision, he and others considered his great contribution to be more along the lines of an exposition of his way of thinking logically and spatially.

We need instead point to the work of William Applebaum in the 1930s, as the principal forbearer of business geography, as his methods were the first to intentionally focus on how geographic reasoning, geographic procedures, and geographic technology, could be used to improve the business decision. Among the methods William Applebaum introduced were the “Customer Spotting Method” and the “Analog Method (see Thrall, 2002, ch 4).” These methods provide the foundations of much of what business geographers do today in retail and real estate market analysis.

Applebaum’s Customer Spotting Method identifies the place of origin of the customer, in contrast to the destination target store. The customer’s origin might be home, work, or a nearby traffic generator such as a regional mall. The technology of the day would be to put pins on a map to represent origins and destinations. Applebaum’s Analog Method introduced a whole new way of utilizing the geographic information. Today, the pins would be points on a map, and if there are many points the analyst might use GIS to create a trend surface of customer density.

Throughout the long history of human geography, each place has been viewed as unique: Cleveland Ohio, and Miami Florida, are distinctively different and very individual places. Armed with their belief in uniqueness of place, geographers have described in great detail the places of their study. The census is one example of such documentation. Satellite imagery is a technology that is appropriately suited for documenting the uniqueness of place.

In the middle of the 20th century, some radical geographers departed from the unique view, and instead believed that of greatest value are those phenomena that we can describe and explain that are in common between places. Without such commonality, there can be no theory. The development of geographic theory, and the verification of theory, requires a belief in, and the existence of, an underlying commonality. Among the geographic theories that have their origins during this time period were Central Place Theory, Gravity and Spatial Interaction Models, and theories of urban land use and land values including my own Consumption and Production Theories of Land Rent whose foundations are based upon the uniquely geographic concept of spatial equilibrium (Thrall, 2002, ch 3).

Applebaum had a third view of a place. He believed that places were repeatable. He was not interested in explaining why places had some areas within them that were similar to areas in other places. Instead, Applebaum was interested in describing the geographic characteristics of the site and nearby geography that were shared among the top performing stores in a multi branch retail chain. Applebaum believed that if descriptions of the site and nearby geography could be found elsewhere, then the store performance of the top performing stores could also be duplicated for those sites and situations. In this way, the poor performing sites and situations could be avoided as well. GIS technology is relied upon

today to calculate the characteristics that make for a successful geographic setting, and then look for those same characteristics, elsewhere.

The geographic technology that business geographers use to improve judgment has come about by either creating a unique new technology to address problems specific to business geography, or relevant technology has been borrowed from the broad GIS industry. Among the technologies developed specifically to satisfy the business geographers' unique needs are lifestyle segmentation profiling (LSP) (Thrall, 2002, ch 4). LSPs are used to categorize consumers into categories, where all within the same category have the same propensity to consume. In other words, propensity to consume is repeatable. Trend surface analysis and grid-based maps, are examples of technology borrowed from the larger GIS industry. Because they are well suited to describe and summarize the geography of a landscape, trend surface analysis and grid-based maps enable the business geographer to better identify those repeatable locations apt to be the most profitable.

Joseph Berry in his book *Analyzing Geo-Business Data: A Hands-on Case Study in Spatial Analysis and Data Mining*, presents geographic technology that is common to the broad GIS industry, while at the same time proven by application to be valuable to improving the business decision. Joseph Berry's focus is upon grid-based map analysis. Examples of grid-based maps include the number of people with a particular LSP within each grid cell and expected revenues to be generated from each grid cell.

The US Census produces a wealth of geographic data, at various levels of geography (see Thrall, 2002, p. 22). Perhaps the most commonly used among business geographers is the US Census Tract. The census tract is an irregular polygon. The best location for a retailer might be where the count of households that exceed \$65,000 per year, and are also under 45 years of age, is greater than 1,000 people per square mile. But census tracts do not come to us as nice square mile geographic chunks. They come in a very unhandy variety of sizes and shapes. Among adjoining census tracts, one might be large, and the adjacent one at the urban periphery might be many times its neighbor's size. Portraying whole number attribute data, such as total population, by such irregular polygons is a too frequently made cartographic error.

Grid-based maps instead portray information in standard sized and shaped grid cells. This standardization is necessary for many kinds of common business decision making. And businesses managers know this. Well before the rise of the GIS industry, businesses purchased custom made data sets and custom maps of that data that were grid based. Indeed, while still a child in the 1960s, I frequently assisted my mother laboring over grid-based maps of population characteristics for General Electric's west coast marketing division.

Grid-based map analysis is not unique to business geography. Applications range from petroleum exploration to identifying the best location of a new retail store. Business geography students and practitioners might know what they want in a grid-based map. But, their complexity of execution has usually been understated in business geographic publications. That leaves the student or practitioner without an understanding of how to actually calculate the grid-based maps. Moreover, the underlying assumptions of the technology used to create the maps might be glossed over, opening the door for improper interpretation and use of the grid-based maps.

If grid-based analysis is a common thread among all consumers of GIS, and spans across all disciplinary barriers, then why would a business geography student and practitioner be better off with a book focusing on grid-based analysis for business geographers, rather than a general purpose GIS book? First, all examples in Joseph Berry's book are relevant to business geography problem solving. Mapping underground petroleum deposits might be intellectually stimulating to some, but business geography students and practitioners might not see the connection and parallel to improving the business decision. Second, the body of general GIS knowledge has expanded during the past two decades where the student and practitioner are best served by those expert in the topic to pre-select what is most important and what will be most relevant.

Business geographers will also appreciate the packaging that Joseph Barry's book brings. Joseph Berry allows the student to become quickly immersed in the spatial reasoning concepts with appropriate technology embedded on the accompanying CD. Having the technology available on your home computer is not only very nice, but it improves productivity, which is what business geography technology should do.

*Grant Ian Thrall, Ph.D.
Gainesville Florida and Vail Colorado
May, 2003*

*References: see Thrall, Grant Ian. 2002, Business Geography and New Real Estate Market Analysis, London: Oxford University Press, www.oup.co.uk, search Business Geography.
Also see <http://web.clas.ufl.edu/users/thrall/oxford2002.html>.*

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 sheer 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