Mastering ArcGIS

MARIBETH H. PRICE

SEVENTH EDITION

Mc Graw Hill Education

Seventh Edition

Maribeth Price
South Dakota School of Mines and Technology

Mastering ArcGIS

Mc Graw Hill Education

MASTERING ArcGIS, SEVENTH EDITION

This book is printed on acid-free paper.

1 2 3 4 5 6 7 8 9 0 RMN/RMN 1 0 9 8 7 6 5

ISBN 978-0-07-809514-6
MHID 0-07-809514-X

Senior Vice President, Products & Markets: *Kurt L. Strand*
Managing Director: *Thomas Timp*
Brand Manager: *Michelle Vogler*
Director of Development: *Rose Koos*
Development Editor: *Tammy Ben*
Marketing Manager: *Matthew Garcia*
Content Production Manager: *Faye Schilling*
Content Project Manager: *Melissa M. Leick, Tammy Juran, and Sandy Schnee*
Buyer: *Susan K. Culbertson*
Content Licensing Specialist (Text): *Deanna Dausener*

Cover Image: © *Image Source, all rights reserved*
Compositor: *MPS Limited*
Typeface: *11/13 Times Roman*
Printer: *R. R. Donnelley*

Library of Congress Cataloging-in-Publication Data

Price, Maribeth Hughett, 1963-
Mastering ArcGIS / Maribeth Price, South Dakota School of Mines and Technology. -- Seventh edition.
pages cm
ISBN 978-0-07-809514-6 (alk. paper)
1. ArcGIS. 2. Geographic information systems. I. Title.
G70.212.P74 2015
910.285'53--dc23
2014041471

www.mhhe.com

Table of Contents

Preface

Welcome to *Mastering ArcGIS*, a detailed primer on learning the ArcGIS™ software by ESRI®, Inc. This book is designed to offer everything you need to master the basic elements of GIS.

> **Notice:** ArcGIS™, ArcMap™, ArcCatalog™, ArcGIS Desktop™, ArcInfo Workstation™, and the other program names used in this text are registered trademarks of ESRI, Inc. The software names and the screen shots used in the text are reproduced by permission. For ease of reading only, the ™ symbol has been omitted from the names; however, no infringement or denial of the rights of ESRI® is thereby intended or condoned by the author.

What's new in the seventh edition?

The primary goals for this edition included strengthening the data management content and placing more emphasis on managing and compiling GIS data, as well as writing metadata. I am finding that students these days know less about basic computer file system concepts than I expect, and they need more background and instruction to work effectively with GIS data. I added a new chapter on GIS data management to cover these concepts, and reorganized the chapters to cover data management first, which appears to be the way most instructors like to structure their courses. Metadata is more strongly emphasized throughout, first as the simple Item Description and later as standards-based metadata in Chapter 14.

The chapter on data presentation is almost entirely rewritten to present basic map design principles and processes from an academic cartographic design perspective. I have also made numerous additions and improvements to other chapters, such as adding weighted overlay to raster analysis and moving erase and clip to the chapter on queries.

With the book length fixed by the publisher, however, I had to eliminate material to make room for these improvements. The chapters on networks and geocoding have been removed from the text, since most people I have talked to don't use them in an introductory class. However, they will remain available on the McGraw-Hill web site in PDF form. The data files needed for these chapters remain in the data provided on the book web site.

I have found welcome stability in the procedures and menus throughout ArcGIS Versions 10.1, 10.2, and 10.2.2, and this edition will work well with any of them and most likely 10.3 as well.

I would like to thank the many people who have used and commented on this book, and I hope that it continues to serve their needs in the rapidly evolving world of GIS.

Looking ahead

ESRI is releasing a completely redesigned ArcGIS Desktop product in late fall 2014, called ArcGIS Professional. It has a 64-bit, multithreaded architecture, uses ribbon-style menus, will integrate 2D and 3D applications, and is closely tied to ArcGIS Online. The eighth edition of *Mastering ArcGIS* will be written for ArcGIS Professional. It seemed premature to use it for the seventh edition because it is still in development, does not yet include all the functionality currently available, and requires a 64-bit computer, which may not be standard equipment in all campus labs yet. ArcMap and ArcCatalog will exist side-by-side with this new program for many years yet, and we anticipate keeping the seventh edition of *Mastering ArcGIS* in print as long as needed for those continuing to use them.

Previous experience

This book assumes that the reader is comfortable using Windows™ to carry out basic tasks such as copying files, moving directories, opening documents, exploring folders, and editing text and word processing documents. Previous experience with maps and map data is also helpful. No previous GIS experience or training is necessary to use this book.

Elements of the package

This learning system includes a textbook and web site, including

- Fourteen chapters on the most important capabilities of ArcGIS
- Comprehensive tutorials in every chapter to learn the skills, with each step demonstrated in a video clip
- A set of exercises, map documents, and data for practicing skills independently
- Reference sections on skills with video clips demonstrating each one

This book assumes that the student has access to ArcGIS Desktop Basic (formerly ArcView). A few optional topics are introduced that require an ArcGIS Desktop Standard (formerly ArcEditor) license. The Spatial Analyst extension is required for Chapter 11.

Philosophy

This text reflects the author's personal philosophies and prejudices developed from 20 years of teaching GIS at an engineering school. The main goal is not to train geographers but to provide students in any field with GIS skills and knowledge. It is assumed that most students using this book already have a background of discipline-specific knowledge and skills upon which to draw and are seeking to apply geospatial techniques within their own knowledge domains.

- **GIS is best learned by doing it, not by studying it.** The laboratory is THE critical component of the book, and theory is introduced sparingly and integrated with experience. Hence, this book is heavy on experience and lighter on theory.
- **Independent work and projects are critical to learning GIS.** This book includes a wealth of exercises in which the student must find solutions independently without a cookbook recipe of steps. A wise instructor will also require students to develop an independent project.

Chapter sequence

The book contains an introduction and 14 chapters. Each chapter includes roughly one week's work for a three-credit semester course. This book intentionally contains more material than the average GIS class can cover during a single semester; instructors may choose what to emphasize.

An introductory chapter describes GIS and gives some examples of how it is used. It also provides an overview of GIS project management and how to develop a project. Chapters 1–11 follow a roughly project-based sequence: data compilation, data exploration and mapping, tables and basic editing, and analysis. These chapters are the core of an introductory GIS class and, by the end of it, students should have little difficulty developing and carrying out an independent GIS project. The final chapters introduce several more advanced topics in data management.

Chapter layout

Each chapter is organized into the following sections:

- **Concepts:** provides basic background material for understanding geographic concepts and how they are specifically implemented within ArcGIS. Most chapters have two sections, one (GIS Concepts) covering general GIS concepts and theory, and another (About ArcGIS) covering the specific implantation of those concepts within ArcGIS Desktop. A set of review questions and important terms follows the concepts section.

- **Tutorial:** contains a step-by-step tutorial demonstrating the concepts and skills. The tutorials begin with detailed instructions, which gradually become more general as mastery is built. Every step in the tutorial is demonstrated by accompanying video clips.

- **Exercises:** presents a series of problems to build skill in identifying the appropriate techniques and applying them without step-by-step help. Through these exercises, the student builds an independent mastery of GIS processes. Brief solution methods are included for all exercises, and maps of the results are shown when applicable. A full answer and methods document is available for instructors at the McGraw-Hill Instructor web site.

The web site also contains all the data needed to follow the tutorials and complete the exercises.

Instructors should use judgment in assigning exercises, as the typical class would be stretched to complete all the exercises in every chapter. Very good students can complete the entire chapter in 3–6 hours, most students would need 6–8 hours, and a few students would require 10 or more hours. Students with more computer experience generally find the material easier than others.

Using this text

In working through this book, the following sequence of steps is suggested:

- READ through the Concepts sections to get familiar with the principles and techniques.
- ANSWER the Chapter Review Questions to test comprehension of the material.
- WORK the Tutorial section for a step-by-step tutorial and explanation of key techniques.
- REREAD the Concepts section to reinforce the ideas.
- PRACTICE by doing the Exercises.

Using the tutorial

The tutorial provides step-by-step practice and introduces details on how to perform specific tasks. Students should be encouraged to think about the steps as they are performed and not just race to get to the end.

It is important to follow the directions carefully. Skipping a step or doing it incorrectly may result in a later step not working properly. Saving often will make it easier to go back and correct a mistake in order to continue. Occasionally, a step will not work due to differences between computer systems or software versions. Having an experienced user nearby to identify the problem can help. If one isn't available, however, just skip the step and move on without it.

Using the videos

The web site contains two types of videos. The Tutorial Videos demonstrate each step of the tutorial. They are numbered in the text for easy reference. The Skills Reference Videos show how to perform generic tasks, such as deleting a file. The videos are intended as an alternate learning strategy. It would be tedious to watch all of them. Instead, use them in the following situations:

- When the student does not understand the written instructions or cannot find the correct menu or button
- When a step cannot be made to work properly
- When a reminder is needed to do a previously learned skill in order to complete a step
- When a student finds that watching the videos enhances learning

As you work through a chapter, keep the web site video listing on the screen, and click the appropriate link to see a video. The tutorial clips are distinguished by numbered steps and the Skills Reference Videos by their headings.

Using video and data components

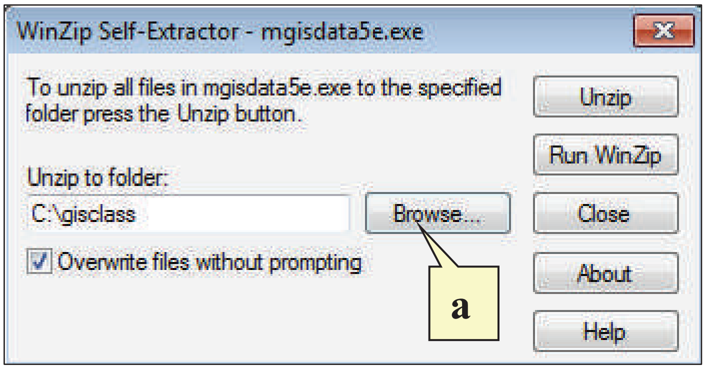
Using the videos:

The videos are available for download from the book web site. Each chapter may be downloaded separately. The videos for Chapter 1 must be downloaded first, for they contain the instructions and folder hierarchy needed for the subsequent chapters. Please review the instructions in the Chapter 1 download before downloading the rest of the videos.

To install the training data:

The mgisdata.exe archive contains a folder with the documents and data needed to do the tutorials and exercises. The student must copy this folder to his own hard drive. If more than one person on the computer is using this book, then each person should make her own copy of the data in a separate folder. The data in the exe archive are a self-extracting zip file that requires approximately 350 MB of disk space. Follow these instructions to install the data:

- Click on the link to Install Data. If a dialog window appears asking whether to Run or Save the data, choose Run.
- When a dialog box appears asking whether to Open or Save the data, choose Open. Don't choose Save because it will only copy the data archive instead of extracting it.
- Click the Browse button to set the folder to which to extract the data (a). The data will be placed in a



Installing the training data

folder called mgisdata in whatever location you choose. In other words, if you select C:\gisclass as the target folder, then the data will be placed in C:\gisclass\mgisdata.

➢ Click Start to begin installing the data. It may take several minutes. Wait until you see the Finished window and then click OK.

System requirements

To use the tutorials and do the exercises in this book, the student must have access to a computer with the following characteristics:

Minimum hardware:

PC-Intel™-platform computer with 2.2-GHz processor or better and 2 GB RAM

Suitable sound/graphics card with 24-bit color depth and 1024 × 768 minimum screen

Software:

Windows 8 (Basic, Professional, Enterprise), Windows 7™ (Ultimate, Enterprise, Professional, Home Premium), Windows Vista™ (Ultimate, Enterprise, Business, Home Premium), Windows 2000™, or Windows XP™ (Home Edition , Professional); requires Microsoft .NET framework to be installed.

A web browser, such as Netscape or Internet Explorer, or Microsoft® Word

A zip utility such as WinZip or 7zip

A media player that is able to display the .mp4 video format (such as Windows Media Player 12 [Windows 7 only] or QuickTime™)

ArcGIS Desktop™ 10.1 or higher (Basic Level); Standard Level and Spatial Analyst extension required for some exercises

For more information, please consult: http://www.esri.com/software/arcgis/arcgis-for-desktop/system-requirements

Internet access is required for ArcGIS installation and for exercises requiring use of ArcGIS Online. Exercises do not require an ArcGIS Online account, although a public or organizational subscription will provide access to more capabilities and content.

For assistance in acquiring or installing these components, contact your system administrator, hardware/software provider, or local computer store.

Acknowledgments

I would like to thank many people who made this book possible. Governor Janklow of South Dakota funded a three-month summer project in 2000 that got the book started, as part of his Teaching with Technology program. Many students in my GIS classes between 2000 and 2014 tested the text and exercises and helped immensely in making sure the tutorials were clear and worked correctly. Reviewers of previous editions, including Richard Aspinall, Joe Grengs, Tom Carlson, Susan K. Langley, Henrietta Loustsen, Xun Shi, Richard Lisichenko, John Harmon, Michael Emch, Jim Sloan, Sharolyn Anderson, Talbot Brooks, Qihao Weng, Jeanne Halls, Mark Leipnik, Michael Harrison, Ralph Hitz, Olga Medvedkov, James W. Merchant, Raymond L. Sanders, Jr., Yifei Sun, Fahui Wang, Michael Haas, Jason Kennedy, Dafna Kohn, Jessica Moy, James C. Pivirotto, Peter Price, Judy Sneller, Dave Verbyla, Birgit Mühlenhaus, Jason Duke, Darla Munroe, Wei-Ning Xiang, L. Joe Morgan, Samantha Arundel, Christopher A. Badurek, Tamara Biegas, John E. Harmon, Michael Hass, Nicholas Kohler, David Long, Jaehyung Yu, Sarah Battersby, Gregory S. Bohr, Kelly R. Dubure, Colleen Garrity, Raymond Greene, Eileen Johnson, James Leonard, and Tao Tang provided detailed and helpful comments, and the book is better than it would have been without their efforts. I also thank the reviewers who provided valuable advice for the seventh edition.

Reviewers for the Seventh Edition:

John Benhart, Jr., *Indiana University of Pennsylvania*
Carsten Braun*, Westfield State University*
Stephanie Deitrick, *Arizona State University,*
Christina Hupy, *University of Wisconsin Eau Claire*
Maction Komwa, *George Mason University*
Timothy LeDoux, *Westfield State University*
Robert Legg, *Northern Michigan University*
Tom Mueller, *California University of Pennsylvania*
Curtis Price, *United States Geological Survey*
Amy Rock, *Ohio University,* and five anonymous reviewers.

Thanks to presenters at the 2014 ESRI Educational and International Users Conference who gave me a crash review on cartography: Allen Carroll, Damien Demaj, Kenneth Field, Makram Murad-Al-Shaikh, and Larry Orman. ESRI, Inc. was prompt and generous in its granting of permission to use the screen shots, data, and other materials throughout the text. They also provided beta and prerelease versions of ArcGIS 10.1 for early development of the text. I extend heartfelt thanks to the City of Austin, Texas, for putting their fine GIS data sets in the public domain. I thank George Sielstad, Eddie Childers, Mark Rumble, Tom Junti, and Patsy Horton for their generous donations of data. I am grateful to Tom Leonard and Steve Bauer for their long-term computer lab administration, without which I could not have taught GIS courses or developed this book. I thank Linda Heindel for organizing student feedback and assisting with the initial round of edits on the first draft. I thank editors Michelle Vogler and Melissa Leick of McGraw-Hill for their unfailing encouragement and enthusiasm about the book as it took shape, as well as for their excellent feedback. I thank the McGraw-Hill team working on the seventh edition, especially Tammy Ben. I am grateful to Daryl Pope, who first started me in GIS, and to John Suppe, who encouraged me to return to graduate school and continue doing GIS on a fascinating study of Venus. I thank my partner, David Stolarz, who provided unfailing encouragement when it seemed as though the editing on this edition would never end. Last, and certainly not least, I thank Curtis Price and my daughters, Virginia and Madeleine, for their understanding and support during the many, many hours I spent working on this book.

Introduction

What Is GIS?

Objectives

- Developing a basic understanding of what GIS is, its operations, and its uses
- Getting familiar with GIS project management
- Learning to plan a GIS project
- Finding resources to learn more about GIS

Concepts

What is GIS?

GIS stands for Geographic Information System. In practical terms, a GIS is a set of computer tools that allows people to work with data that are tied to a particular location on the earth. Although many people think of a GIS as a computer mapping system, its functions are broader and more sophisticated than that. A GIS is a database that is designed to work with map data.

Consider the accounting department of the local telephone company. They maintain a large computer database of their customers, in which they store the name, address, phone number, type of service, and billing information for each customer. This information is only incidentally tied to where customers live; they can carry out most of the important functions (billing, for example) without needing to know where each house is. Of course, they need to have addresses for mailing bills, but it is the post office that worries about where the houses actually ARE. This type of information is called **aspatial data**, meaning that it is not tied, or is only incidentally tied, to a location on the earth's surface.

Employees of the service department, however, need to work with **spatial data** to provide the telephone services. When hundreds of people call in after a power outage, the service department must analyze the distribution of the calls and isolate the location where the outage occurred. When a construction company starts work on a street, workers must be informed of the precise location of buried telephone cables. If a developer builds a new neighborhood, the company must be able to determine the best place to tie into the existing network so that the services are efficiently distributed from the main trunk lines. When technicians prepare lists of house calls for the day, they need to plan the order of visits to minimize the amount of driving time. In these tasks, location is a critical aspect of the job, and the information is spatial.

In this example, two types of software are used. The accounting department uses special software called a *database management system*, or *DBMS*, which is optimized to work with large volumes of aspatial data. The service department needs access to a database that is optimized for working with spatial data, a Geographic Information System. Because these two types of software are related, they often work together, and they may access the same information. However, they do different things with the data.

A GIS is built from a collection of hardware and software components.

- ***A computer hardware platform.*** Due to the intensive nature of spatial data storage and processing, a GIS was once limited to large mainframe computers or expensive workstations. Today, it can run on a typical desktop personal computer.
- ***GIS software.*** The software varies widely in cost, ease of use, and level of functionality but should offer at least some minimal set of functions, as described in the next few paragraphs. In this book, we study one particular package that is powerful and widely used, but others are available and may be just as suitable for certain applications.
- ***Data storage.*** Some projects use only the hard drive of the GIS computer. Other projects may require more sophisticated solutions if large volumes of data are being stored or multiple users need access to the same data sets. Today, many data sets are stored in digital warehouses and accessed by many users over the Internet. Compact disc writers and/or USB portable drives are highly useful for backing up and sharing data.
- ***Data input hardware.*** Many GIS projects require sophisticated data entry tools. Digitizer tablets enable the shapes on a paper map to be entered as features in a GIS data file. Scanners create digital images of paper maps. An Internet connection provides easy access to large volumes of GIS data. High-speed connections are preferred, as GIS data sets may constitute tens or hundreds of megabytes or more.
- ***Information output hardware.*** A quality color printer capable of letter-size prints provides the minimum desirable output capability for a GIS system. Printers that can handle map-size output (36 in. × 48 in.) will be required for many projects.
- ***GIS data.*** Data come from a variety of sources and in a plethora of formats. Gathering data, assessing their accuracy, and maintaining them usually constitutes the longest and most expensive part of a GIS project.
- ***GIS personnel***. A system of computers and hardware is useless without trained and knowledgeable people to run it. The contribution of professional training to successful implementation of a GIS is often overlooked.

GIS software varies widely in functionality, but any system claiming to be a GIS should provide the following functions at a minimum:

- ***Data entry*** from a variety of sources, including digitizing, scanning, text files, and the most common spatial data formats; ways to export information to other programs should also be provided
- ***Data management*** tools, including tools for building data sets, editing spatial features and their attributes, and managing coordinate systems and projections
- ***Thematic mapping*** (displaying data in map form), including symbolizing map features in different ways and combining map layers for display
- ***Data analysis*** functions for exploring spatial relationships in and between map layers
- ***Map layout*** functions for creating soft and hard copy maps with titles, scale bars, north arrows, and other map elements

Geographic Information Systems are put to many uses, but providing the means to collect, manage, and analyze data to produce information for better decision making is the common goal and the strength of each GIS. This book is a practical guide to understanding and using a particular Geographic Information System called ArcGIS. Using this book, you can learn what a GIS is, what it does, and how to apply its capabilities to solve real-world problems.

A history of GIS

Geographic Information Systems have grown from a long history of cartography begun in the lost mists of time by early tribesmen who made sketches on hides or formed crude models of clay as aids to hunting for food or making war. Ptolemy, an astronomer and geographer from the second century B.C., created one of the earliest known atlases, a collection of world, regional, and local maps and advice on how to draw them, which remained essentially unknown to Europeans until the 15th century. Translated into Latin, it became the core of Western geography, influencing cartographic giants such as Gerhard Mercator, who published his famous world map in 1569. The 17th and 18th centuries saw many important developments in cartography, including the measurement of a degree of longitude by Jean Picard in 1669, the discovery that the earth flattens toward the poles, and the adoption of the Prime Meridian that passes through Greenwich, England. In 1859, French photographer and balloonist Gaspard Felix Tournachon founded the art of remote sensing by carrying large-format cameras into the sky. In an oft-cited early example of spatial analysis, Dr. John Snow mapped cholera deaths in central London in September 1854 and was able to locate the source of the outbreak—a contaminated well. However, until the 20th century, cartography remained an art and a science carried out by laborious calculation and hand drawing. In 1950, Jacqueline Tyrwhitt made the first explicit reference to map overlay techniques in an English textbook on town and country planning, and Ian McHarg was one of the early implementers of the technique for highway planning.

As with many other endeavors, the development of computers inspired cartographers to see what these new machines might do. The early systems developed by these groups, crude and slow by today's standards, nevertheless laid the groundwork for modern Geographic Information Systems. Dr. Roger Tomlinson, head of an Ottawa group of consulting cartographers, has been called the "Father of GIS" for his promotion of the idea to use computers for mapping and for his vision and effort in developing the Canada Geographic Information System (CGIS) in the mid-1960s. Another pioneering group, the Harvard Laboratory for Computer Graphics and Spatial Analysis, was founded in the mid-1960s by Howard Fisher. He and his colleagues developed a number of early programs between 1966 and 1975, including SYMAP, CALFORM, SYMVU, GRID, POLYVRT, and ODYSSEY. Other notable developers included professors Nystuen, Tobler, Bunge, and Berry from the Department of Geography at Washington University during 1958–1961. In 1970, the US Bureau of the Census produced its first geocoded census and developed the early DIME data format based on the CGIS and POLYVRT data representations. DIME files were widely distributed and were later refined into the TIGER format. These efforts had a pronounced effect on the development of data models for storing and distributing geographic information.

In 1969, Laura and Jack Dangermond founded the Environmental Systems Research Institute (ESRI), which pioneered the powerful idea of linking spatial representation of features with attributes in a table, a core idea that revolutionized the industry and launched the development of Arc/Info, a program whose descendants have captured about 90% of today's GIS market. Other vendors are still active in developing GIS systems, which include packages MAPINFO, MGA from Intergraph, IDRISI from Clark University, and the open-source program GRASS.

What can a GIS do?

A GIS works with many different applications: land use planning, environmental management, sociological analysis, business marketing, and more. Any endeavor that uses spatial data can benefit from a GIS. For example, researchers at the US Department of Agriculture Rocky Mountain Research Station in Rapid City conducted a study of elk habitat in the Black Hills of

South Dakota and Wyoming by placing radio transmitter collars on about 70 elk bulls and cows (Fig. I.1c). Using the collars and a handheld antenna, they tracked the animals and obtained their locations. Several thousand locations were collected (Fig. I.1a), allowing the scientists to study the characteristics of the habitat where elk spend time.

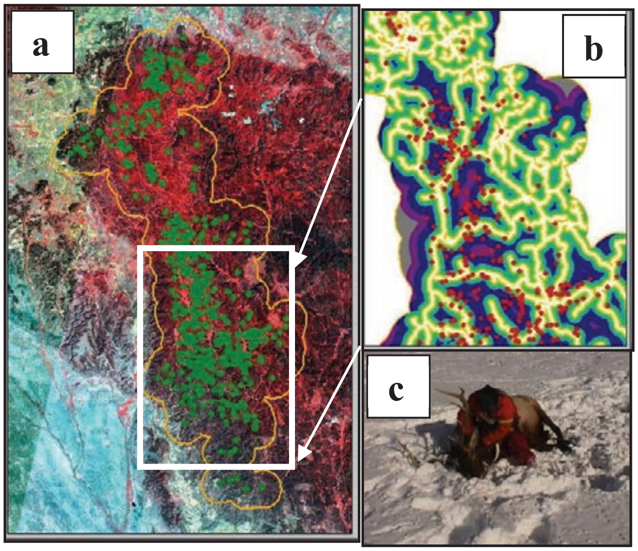


Fig. I.1. Analyzing elk habitat use: (a) elk locations and study area; (b) locations on a map of distance to nearest road; (c) collaring an elk

The elk locations were entered into a GIS system for record keeping and analysis. Each location became a point with attached information, including the animal ID number and the date and time of the sighting. Information about vegetation, slope, aspect, elevation, water availability, and other site factors were derived by overlaying the points on other data layers, allowing the biologists to compare the characteristics of sites utilized by the elk. Figure I.1b shows a map of distances to major roads in the central part of the study area. The elk locations clustered in the darker roadless areas, and statistical analysis demonstrated this observation empirically.

There are many applications of GIS in almost every human endeavor, including business, defense and intelligence, engineering and construction, government, health and human services, conservation, natural resources, public safety, education, transportation, utilities, and communication. In August 2014, the Industries section of the ESRI web site at www.esri.com listed 62 different application areas in these categories, each one with examples, maps, and case studies. Instead of reading through a list here, go to the site and see the latest applications.

New trends and directions in GIS

The GIS industry has grown exponentially since its inception. Starting with the mainframe and then the desktop computer, GIS began as a relatively private endeavor focused on small clusters of specialized workers who spent years developing expertise with the software and data. Since then, the development of the Internet and the rapidly advancing field of computer hardware have been driving some significant changes in the industry.

Proliferation of options for data sharing

Instead of storing data sets locally on individual computers or intra-organization network drives, more people are serving large volumes of data over the Internet to remote locations within an organization, across organizations, and to the general public. In the past, large collections of data were hosted by various organizations through clearinghouses sponsored by the National Spatial Data Infrastructure (NSDI) organization. The data usually existed as GIS data files in various formats and required significant expertise and the right software to download and use. Today, *GIS Servers* are designed to bring GIS data to the general public within a few clicks. Another class of data providers that give wide access to spatial data, although not specifically GIS applications, include web sites such as GoogleEarth, MapQuest, and Microsoft Virtual Earth.

Like many other computer industries, GIS is heading for the clouds. A cloud is a gigantic array of large computers that customers rent pieces of by the hour, instead of purchasing their own physical hardware. ArcGIS Online is a cloud-based platform for users to collaborate and share GIS data with one another, making it suddenly very easy to share data with a colleague, or with the world, even for those with no particular GIS expertise.

Proliferation of options for working with GIS data

In the early days, people who wanted to do GIS had to purchase a large, expensive program and learn to use it. Now, a wide variety of scaled applications permits different levels of use for different levels of need. Not every user must have the full program. There are map servers for people who just need to view and print maps, free download software for viewing interactive map publications, and scaled-down versions of the full program with fewer options.

Many organizations are turning to Server GIS as a less expensive alternative to purchasing large numbers of GIS licenses. Many workers need GIS, but they only use a small subset of GIS functions on a daily basis. A Server GIS can provide both data and a customizable set of viewing and analysis functions to users without a GIS license. All they need is a web browser. Because of more simple and low-cost ways of accessing GIS data and functions, the user base has expanded dramatically.

Expansion of GIS into wireless technology

More people are collecting and sharing data using handheld wireless devices, such as handheld computers, smartphones, and global positioning system (GPS) units. These units can now access Internet data and map servers directly so that users in the field can download background data layers, collect new data, and transmit them back to the large servers. Cell phone technology is advancing rapidly with new geolocation options, web access, and geo-applications arriving daily.

Emphasis on open-source solutions

Instead of relying on proprietary, specialized software, more GIS functions are now implemented within open-source software and hardware. GIS data are now more often stored using engines from commercial database platforms and utilize the same development environments as the rest of the computer industry. This trend makes it easier to have the GIS communicate with other programs and computers and enhances interoperability between systems and parts of systems.

Customization

With emphasis on open-source solutions, new opportunities have developed in creating customized applications based on a fundamental set of GIS tools, such as a hydrology tool or a wildlife management tool. Smartphone and tablet applications programming is burgeoning. These custom applications gather the commonly used functions of GIS into a smaller interface, introduce new knowledge, and formalize best practices into an easy-to-use interface. Customization requires a high level of expertise in object-oriented computer programming.

Enterprise GIS

Enterprise GIS integrates a server with multiple ways to access the same data, including traditional GIS software programs, web browser applications, and wireless mobile devices. The goal is to meet the data needs of many different levels of users and to provide access to nontraditional users of GIS. The Enterprise GIS is the culmination of the other trends and capabilities already mentioned. The costs and challenges in developing and maintaining an Enterprise GIS are significant, but the rewards and cost savings can also be substantial.

What do GIS professionals do?

It's getting so easy to create a map these days that one may wonder why one should bother to learn GIS. However, the easy solutions are based on the work of experts who provide the data and the software systems to handle them. It may not be that hard to learn to use a smartphone, but you still need the engineers and software developers to build the phones and the infrastructure to make them work. GIS is much the same. These days, GIS professionals play a variety of roles. A few broad categories can be defined.

Primary Data Providers create the base data that form the backbone of many GIS installations. Surveyors and land-planning professionals contribute precise measurement of boundaries. Photogrammetrists develop elevation and other data from aerial photography or the newer laser altimetry (LIDAR) systems. Remote sensing professionals extract all kinds of human-made and natural information from a variety of satellite and airborne measurement systems. Experts in global positioning systems (GPS), which provide base data, are also important.

Applications GIS usually involves professionals trained in other fields, such as geography, hydrology, land use planning, business, and utilities, who utilize GIS as part of their work. Specialists in mathematics and statistics develop new ways to analyze and interpret spatial data. For these professionals, GIS is an added skill and a tool to make their work more efficient, productive, and valuable.

Development GIS involves skilled software and hardware engineers who build and maintain the GIS software itself, as well as the hardware components upon which it relies—not only computers and hard drives and plotters but also GPS units, wireless devices, scanners, digitizers, and other systems that GIS could not function without. This group also includes an important class of GIS developers who create customized applications from the basic building blocks of existing GIS software.

Distributed Database GIS involves computer science professionals with a background in networking, Internet protocols, and/or database management systems. These specialists set up and maintain the complex server and network systems that allow data services, Server GIS, and Enterprise GIS to operate.

GIS project management

A GIS project may be a small effort spanning a couple of days by a single person, or it can be an ongoing concern of a large organization with hundreds of people participating. Large or small, however, projects often follow the generalized model shown in Figure I.2, and most new users learn GIS through a project approach. A project usually begins with an assessment of needs. What specific issues must be studied? What kind of information is needed to support decision making? What functions must the GIS perform? How long will the project last? Who will be using the data? What funding is available for start-up and long-term support?

Without a realistic idea of what the system must accomplish, it is impossible to design it efficiently. Users may find that some critical data are absent or that resources have been wasted acquiring data that no one ever uses. In a short-term project the needs are generally clear-cut. A long-term organizational system will find that its needs evolve over time, requiring periodic reassessment. A well-designed system will adapt easily to future modification. The creator of a haphazard system may be constantly redoing previous work when changes arise.

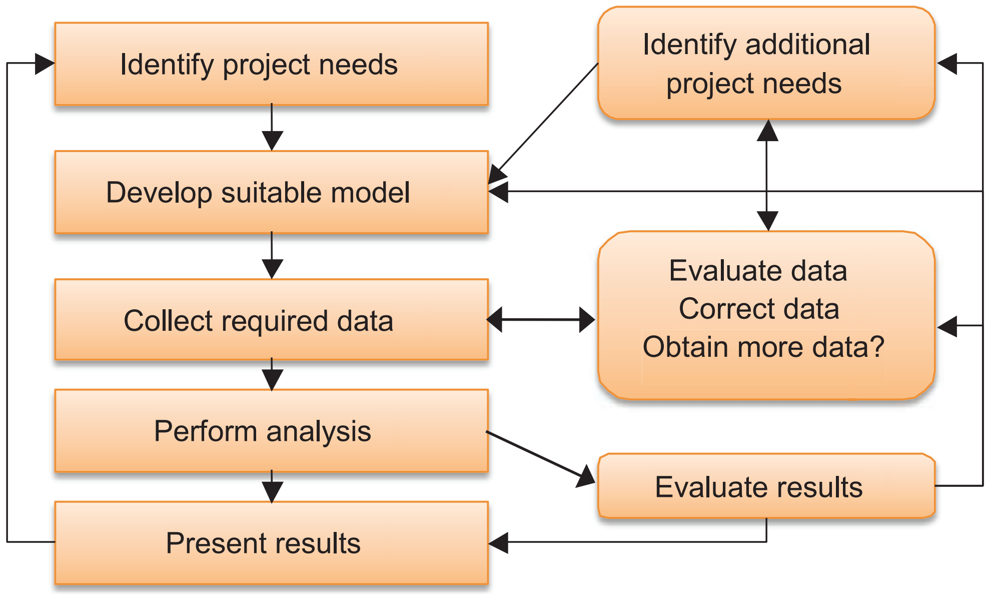


Fig. I.2. Generalized flow diagram representing steps in a GIS project

In studies seeking specific answers to scientific or managerial questions, a methodology or **model** must be chosen. Models convert the raw data of the project into useful information using a well-defined series of steps and assumptions. Creating a landslides hazards map provides a simple example. One might define a model such that, if an area has a steep slope and consists of a shale rock unit, then it should be rated as hazardous. The raw data layers of geology and slope can be used to create the hazard map. A more complex model might also take into account the dip (bedding angle) of the shale units. Models can be very simple, as in this example, or more complex with many different inputs and calculations.

Once needs are known and the appropriate models have been designed, data collection can begin. GIS data are stored as layers, with each layer representing one type of information, such as roads or soil types. The needs dictate which data layers are required and how accurate they must be. A source for each data layer must be found. In some cases data can be obtained free from other organizations. General base layers, such as elevation, roads, streams, political boundaries, and demographics, are freely available from government sources (although the accuracy and level of detail are not always what one might wish). More specific data must often be developed in-house. For example, a utility company would need to develop its own layers showing electric lines and substations—no one else would be likely to have these data.

The spatial detail and accuracy of the data must be evaluated to ensure that they are able to meet the needs of the project. For example, an engineering firm creating the site plan for a shopping mall could download elevation contours of the site for free. However, standard 10-foot contours cannot provide the detailed surface information needed by the engineers. Instead, the firm might contract with a surveying firm to measure contours at half-foot intervals.

After the data are assembled, the analysis can begin. During the analysis phase, it is not unusual to encounter problems that might require making changes to the model and/or data. Thus, the steps of model development, data collection, and analysis often become iterative, as experience gained is used to refine the process. The final result must be checked carefully against reality in order to recognize any shortcomings and to provide guidelines for improving future work.

Finally, no project is complete until the results have been communicated, whether shown informally to a supervisor, published in a scientific journal, or presented during a heated public meeting. Presentations may include maps, reports, slides, or other media.

Learning more about GIS

As you work your way through this text, you will be amazed at all the things that a GIS can do. At times, the abundance of tasks and the flexibility of options may be overwhelming. Even so, this book gives only a bare introduction to what is possible and covers a small portion of what a GIS can do. Moreover, the industry evolves rapidly. Completing the lessons here is only the first step; you will find that you need to seek new information and training constantly as you develop your skills. How can you do this?

- ***Read the Help.*** Maybe you can figure out your cell phone without the manual, but if you ignore the Help files you are turning your back on a wealth of information. They don't just have instructions for doing things but also discussions, diagrams, references, and other ways to help you understand the concepts behind GIS as well as its implementation.
- ***Use the Virtual Campus.*** ESRI has dozens of online courses and seminars to help you learn. Many of the basic ones are free. If your campus has a site license, the GIS administrator can request others for free as well. Talk to your GIS instructor about getting access to these courses.
- ***Build your GIS library.*** GIS integrates many disciplines, including geography, surveying, cartography, statistics, computer science, spatial statistics, and so on. The more you know of these disciplines, the more you will understand, not just the *How* of doing things but also the *Why.*
- ***Join a professional organization.*** Many professional societies cater to GIS and remote sensing. They have newsletters and journals, conferences, and lots of professionals who can help answer your questions and share their experiences.
- ***Join an online forum.*** ESRI has online forums for many aspects of GIS, as well as other useful links, at http://support.esri.com/en/. You can search for answers to questions, and if nothing is there, you can ask a question and generally someone will answer within a few days. It is a great place to go when you get stuck with no one else to ask. You will also learn about bugs and workarounds.
- ***Join ArcGIS Online.*** Not only is it easy and fun to create maps and share them, but you can also search for courses, documents, slideshows, videos, and other GIS-related materials.

Chapter 1. GIS Data

Objectives

- Understanding how real-world features are represented by GIS data
- Knowing the differences between the raster and vector data models
- Getting familiar with the basic elements of data quality and metadata
- Learning the different types of GIS files used by ArcGIS
- Using ArcMap to view GIS data
- Learning about map documents, layers, and data frames

Mastering the Concepts

GIS Concepts

Representing real-world objects on maps

To work with maps on a computer requires developing methods to store different types of map data and the information associated with them. Map data fall into two categories: **discrete** and **continuous. Discrete** data are objects in the real world with specific locations or boundaries, such as cities, roads, or soils units. **Continuous** data represent a quantity that is measured and recorded everywhere over a surface, such as temperature or elevation.

Many different data formats have been invented to encode data for use with GIS programs; however, most follow one of two basic approaches: the **vector** model, which is designed to store discrete data, or the **raster** model, which is designed to store continuous data. In either approach, the critical task includes representing the information at a point, or over a region in space, using *x* and *y* coordinate values (and sometimes *z* for height). The *x* and *y* coordinates are the spatial data. The information being represented, such as a soil type or a chemical analysis of a well, is called the attribute data. Raster and vector data models both store spatial and attribute data, but they do it in different ways.

Both data systems are **georeferenced,** meaning that the information is tied to a specific location on the earth's surface using *x-y* coordinates defined in a standard way: a **coordinate system.** One can choose from a variety of coordinate systems, as we will see in Chapter 3. As long as the coordinate systems match, we can display any two spatial data sets together, and they will appear in the correct spatial relationship to each other.

The vector model

Vector data use a series of *x-y* locations to store information (Fig. 1.1). Three basic vector objects exist: points, lines, and polygons. These objects are called **features. Point** features are used to

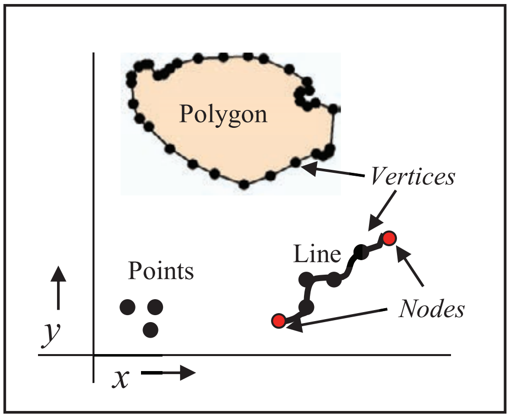


Fig. 1.1. The vector data model uses a series of *x-y* locations to represent points, lines, and polygon areas.

represent objects that have no dimensions, such as a well or a sampling locality. **Line** features represent objects in one dimension, such as a road or a utility line. **Polygons** are used to represent two-dimensional areas, such as a parcel or a state. In all cases, the features are represented using one or more *x-y* coordinate locations (Fig. 1.1). A point consists of a single *x-y* coordinate pair. A line includes two or more pairs of coordinates—the endpoints of the line are termed **nodes**, and each of the intermediate points is called a **vertex**. A polygon is a group of **vertices** that define a closed area.

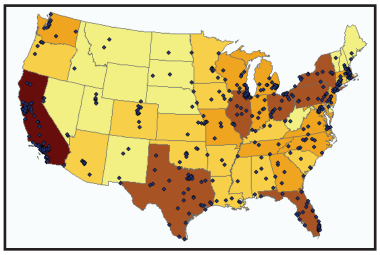


Fig. 1.2. A states feature class and a cities feature class

The type of object used to represent features depends on the scale of the map. A river would be represented as a line on a map of the United States because at that scale it is too small for its width to encompass any significant area on the map. If one is viewing a USGS topographic map, however, the river encompasses an area and would be represented as a polygon.

In GIS, like features are grouped into data sets called **feature classes** (Fig. 1.2). Roads and rivers are different types of features and would be stored in separate feature classes. A feature class can contain only one kind of geometry—it can include point features, line features, or polygon features but never a combination. In addition, objects in a feature class have information stored about them, such as their names or populations. This information is called the **attributes** and is stored in a table (Fig. 1.3). A special field, called the Feature ID (**FID**) or ObjectID (**OID**), links the spatial data with the attributes. Each feature's attributes are stored in one row of the table, and each column has a different type of information, such as population or area. A river and a highway would not be found in the same feature class because their information would be different—flow measurements for one versus pavement type for the other—and would need to be stored in different tables with different columns.

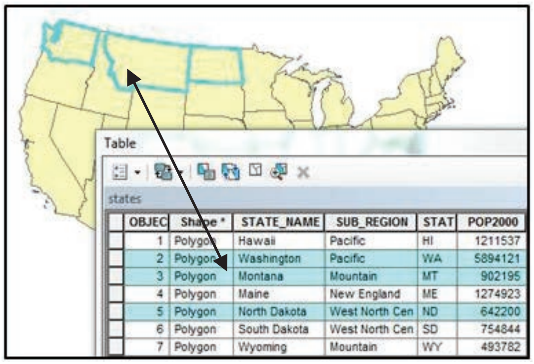


Fig. 1.3. Each state is represented by a spatial feature (polygon), which is linked to the attributes.

When a state is highlighted on the map, its matching attributes are highlighted in the table, and vice versa. It is this live link between the spatial and attribute information that gives the GIS system its power. It enables us, for example, to create a map in which the states are colored based on an attribute field, such as population (Fig. 1.2). This technique is called **thematic mapping** and is one example of how linked attributes can be used to analyze geographic information.

Feature classes can be stored in several different formats. Some data formats only contain one feature class. Others, called **feature datasets**, can contain multiple feature classes that are in some way related to one another. For example, a feature dataset called Transportation might contain the feature classes Roads, Traffic Lights, Railroads, Airports, and Canals.

The benefits of the vector data model are many. First, it can store individual features, such as roads and parcels, with a high degree of precision. Second, the linked attribute table provides great flexibility in the number and type of attributes that can be stored about each feature. Third, the vector model is ideally suited to mapmaking because of the high precision and detail of features that can be obtained. The vector model is also a compact way of storing data, typically requiring a tenth of the space of a raster with similar information. Finally, the vector model is ideally suited to certain types of analysis problems, such as determining perimeters and areas, detecting whether features overlap, and modeling flow through networks.

However, the vector model has some drawbacks. First, it is poorly adapted to storing continuous surfaces, such as elevation or precipitation. Contour lines (as on topographic maps) can be used to represent surfaces, but calculating derived information from contours, such as slope, flow direction, and aspect, is difficult. Finally, some types of analysis are more time-consuming to perform with vectors.

Modeling feature behavior with topology

Two basic vector models exist: **spaghetti models** and **topological models**. A spaghetti model stores features of the file as independent objects, unrelated to each other. Simple and straightforward, this type of model is found in many types of applications that store spatial data. It is also commonly used to transfer vector features from one GIS system to another.

A topological data model stores features, but it also contains information about how the features are spatially related to each other: for example, whether two parcels share a common boundary (**adjacency**), whether two water lines are attached to each other (**connectivity**), whether a company sprayed pesticide over the same area on two different occasions (**overlap**), or whether a highway connects to a crossroad or has an overpass (**intersection**). Although computer algorithms can determine whether these spatial relationships exist between features in a spaghetti model, storing topology can save time if the relationships must be used repeatedly.

Another application of topology involves analyzing the **logical consistency** of features. Logical consistency evaluates whether a data model or data set accurately represents the real-world relationships between features. For example, two adjacent states must share a common boundary that is exactly the same (the real-world situation), even though the states are stored in the data model as two separate features with two boundaries that coincide (Fig. 1.4). Lines representing streets should connect if the roads they represent meet. A line or a polygon boundary should not cross over itself.

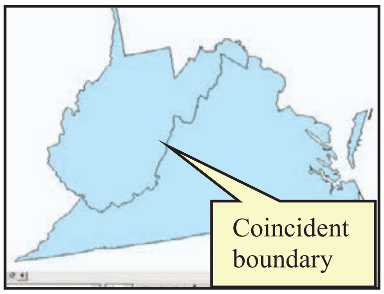


Fig. 1.4. A coincident boundary gets stored twice but is the same for both features.

Finally, topology can be used to model the real-world behavior of features. In a network topology, for example, the connections between features are tracked so that flow through the network can be analyzed. Applications of networks include water in streams, traffic along roads, flights in and out of airline hubs, and utilities through pipes or electrical systems.

The raster model

The raster model has the benefit of simplicity. A set of spatial data is represented as a series of small squares, called **cells** or **pixels** (Fig. 1.5). Each pixel contains a numeric code indicating a single attribute, and the raster is stored as an array of numbers.

Vector features, such as roads or land use polygons, can be converted to raster format by selecting a single attribute to be stored in the cells. In Figure 1.5a, the cells store numeric values representing a land cover type, such as 46 for conifer forest or 23 for hardwood forest. Each value is given a different color for display. The roads shown in Figure 1.5b were originally vector line features with a text attribute indicating a primary, secondary, or primitive road type. When converted to a raster, the number (1, 2, or 3) is used to represent each road that passes through a cell. The cells that don't contain roads are given a null value. Rasters that store vector features in a raster format are sometimes called **discrete** rasters.

However, rasters excel at storing continuous data, which are quantities or variables that change over the earth's surface. A **digital elevation model** (DEM), for example, stores elevation values (Fig. 1.5c). Cells are unlikely to have the same elevation as their neighbors, and the values range smoothly into one another, forming a continuous surface (or continuous field). Therefore, they are commonly called **continuous** rasters.

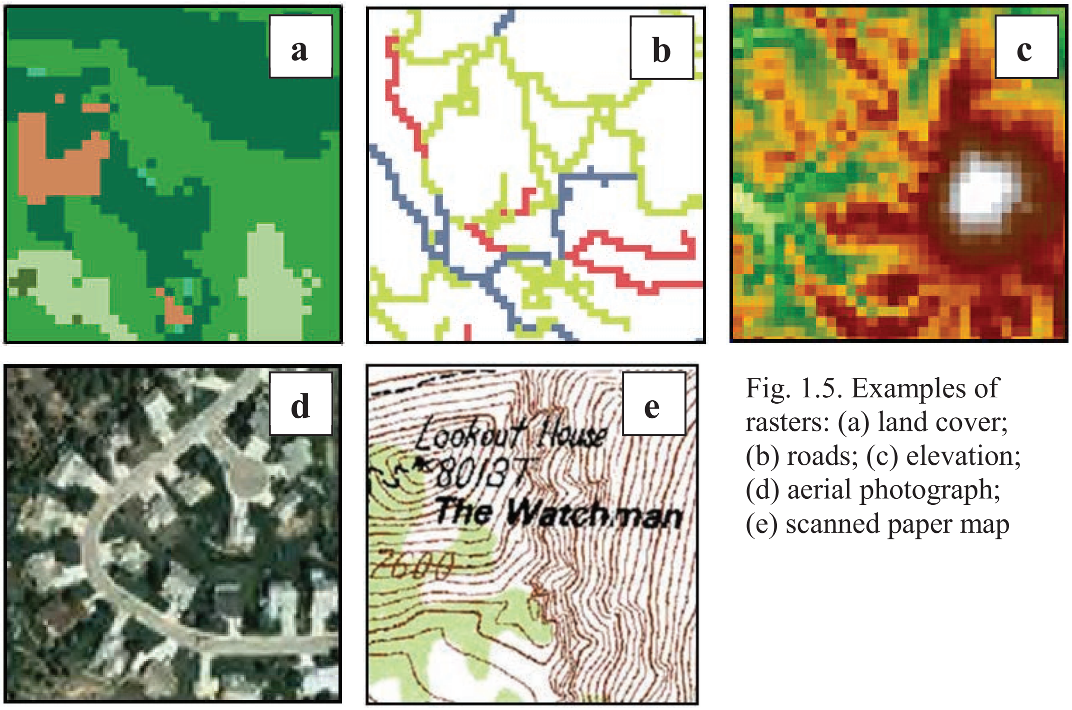


Fig. 1.5. Examples of rasters: (a) land cover; (b) roads; (c) elevation; (d) aerial photograph; (e) scanned paper map

Image rasters (Fig. 1.5d) store brightness values and are commonly used to store aerial photographs or satellite images. Images may contain multiple arrays of values, called **bands**, to represent each pixel. Color images often contain a red band, a green band, and a blue band for each pixel, and the mixture of values in each band defines the color of the pixel. One can also do a digital scan of a paper map and store it as a raster, such as the US Geological Survey topographic map shown in Figure 1.5e, known as a **digital raster graphic** (**DRG**). Each cell stores an index code representing a different color, such as 5 for the brown contours and 1 for the white background.

A raster data set is laid out as a series of rows and columns. Each pixel has an "address" indicated by its position in the array, such as row = 3 and column = 6. Georeferencing a map in an *x-y* coordinate system requires four numbers: an *x-y* location for one pixel in the raster data set and the size of the pixel in the *x* and *y* directions. Usually the upper-left corner is chosen as the known location, and the *x* and *y* pixel dimensions are the same so that the pixels are square. From these four numbers, it is possible to calculate the coordinates of every other pixel based on its row and column position. In this sense, the georeferencing of the pixels in a raster data set is implicit—one need not store the *x-y* location of every pixel.

The *x* and *y* dimensions of each pixel define the **resolution** of the raster data. The higher the resolution, the more precisely the data can be represented. Consider the 90-meter resolution roads raster in Figure 1.6. Since the raster cell dimensions are 90 meters, the roads are represented as much wider than they actually are, and they appear blocky. A 10-meter resolution raster could represent the roads more accurately; however, the file size would increase by 9 × 9, or 81 times.

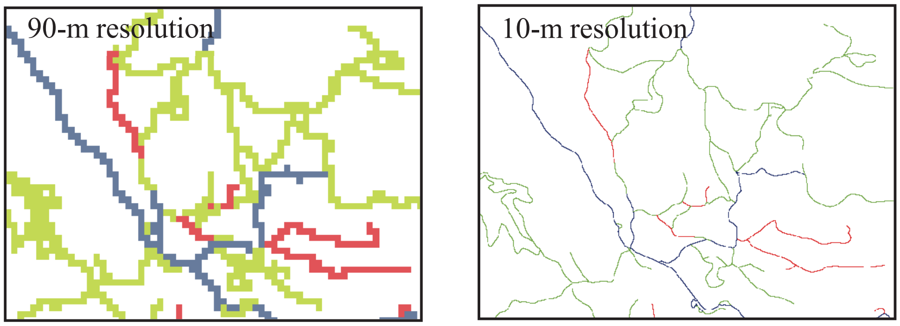


Fig. 1.6. Impact of raster resolution. The 10-meter resolution raster stores the roads more precisely, but it takes 81 times as much space.

The raster model mitigates some of the drawbacks of vectors. It is ideally suited to storing continuous information because each cell can have a value completely different from its neighbors. Many analyses are simple and rapid to perform, and an extensive set of analysis tools for rasters far outstrips those available for vectors.

The drawbacks of rasters lie chiefly in two areas. First, they suffer from trade-offs between precision and storage space to a greater extent than vectors do. The second major drawback of rasters is that they can store only one numeric attribute per raster. Vector files can store hundreds of attribute values for each spatial feature and can handle text data more efficiently.

Coordinate systems

Both raster and vector data rely on *x-y* values to locate data to a particular spot on the earth's surface. The *x-y* values of the coordinate pairs can vary, however. The choice of values and units to store a data set is called its **coordinate system**. Consider a standard topographic map, which actually has three different coordinate systems marked on it. The corners are marked with degrees of latitude and longitude, which is termed a **geographic coordinate system (GCS).** Another set of markings indicates a scale in meters representing the UTM, or Universal Transverse Mercator, coordinate system. A third set of markings shows a scale in feet, corresponding to a State Plane coordinate system. Any location on the map can be represented by three different *x-y* pairs corresponding to one of the three coordinate systems (Fig. 1.7). A global positioning system (GPS) unit also has this flexibility. It can be set to record a location in GCS degrees, UTM meters, State Plane feet, or other coordinate systems.

When creating a vector or raster data set, one must choose a coordinate system and units for storing the *x-y* values. It is also important to label the data so that the user knows which coordinate system has been selected and what the units for the *x-y* values are. Thus, every GIS data set must have a label that records the type of coordinate system and units used to store the *x-y* data inside it.

In older GIS systems, the coordinate systems of different data sets had to match in order for them to be drawn together in the same map. UTM data could only be shown with other UTM data, State Plane data with other State Plane data, and so on. If data were in different coordinate systems, they would need to be converted to the same coordinate system prior to display.

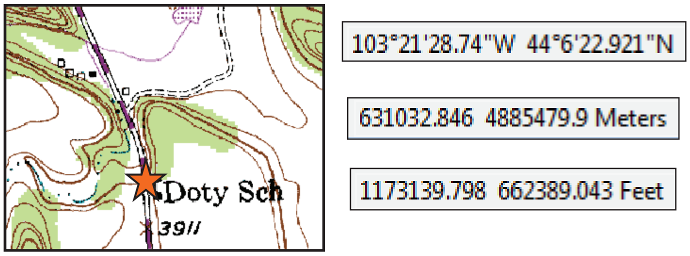


Fig. 1.7. A location can be stored using different coordinate systems and units. The *x-y* location of this school is shown in degrees, UTM meters, and State Plane feet.

Although it is still true that coordinate systems must match for data to be displayed together, many GIS systems can now perform the conversion on the fly. This feature allows data to be stored in different coordinate systems yet to be drawn together. In ArcMap, the user defines a coordinate system for the map, and all of the data are converted to match (Fig. 1.8). The units defined for the map coordinate system, whether they are meters, feet, or degrees, become the **map units** and may differ from the stored units in the files. In Figure 1.8, the UTM data use meters to store the *x-y* coordinates, the GCS data use degrees, and the State Plane data use feet. The Oregon Statewide Lambert coordinate system uses meters, so meters become the map units.

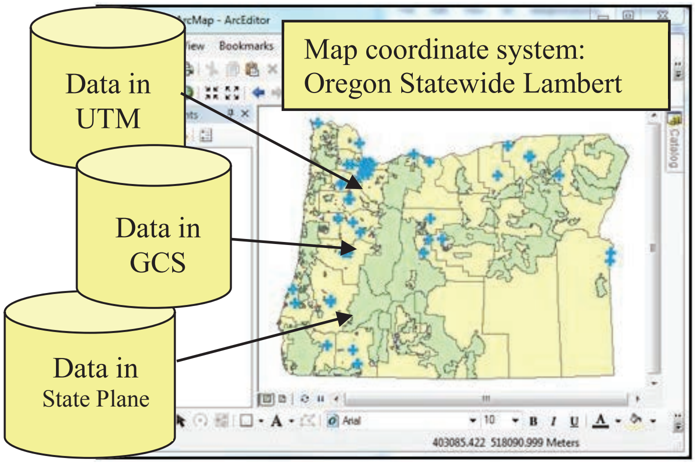


Fig. 1.8. Data in any coordinate system can be displayed together by setting the data frame coordinate system.

Map scale

The act of taking a set of GIS features with *x-y* coordinate values and drawing them on a screen or printing them on paper establishes a map scale. On a paper map, the scale is fixed at the time of printing. Within a computer system that allows interactive display, the scale changes every time the user zooms in or out of the map.

What is map scale?

Map scale is a measure of the size at which features in a map are represented. The scale is expressed as a fraction, or ratio, of the size of objects on the page to the size of the objects on the ground. Because it is expressed as a ratio, it is valid for any unit of measure. So for a common US Geological Survey topographic map, which has a scale of 1:24,000, one inch on the map represents 24,000 inches on the ground. You can use the map scale and a ruler to determine the true distance of any feature on the map, such as the width of a lake (Fig. 1.9). Measure the lake with a ruler and then set up a proportion such that the map scale equals the measured width over the actual width (*x*). Then solve for *x*. Keep in mind that the actual width and the measured width will have the same units. You can convert these units, if necessary.

Often people or publications refer to large-scale maps and small-scale maps. A large-scale map is one in which the *ratio* is large (i.e., the denominator is small). Thus, a 1:24,000 scale map is larger scale than a 1:100,000 scale map. Large-scale maps show a relatively small area, such as a quadrangle, whereas small-scale maps show bigger areas, such as states or countries.

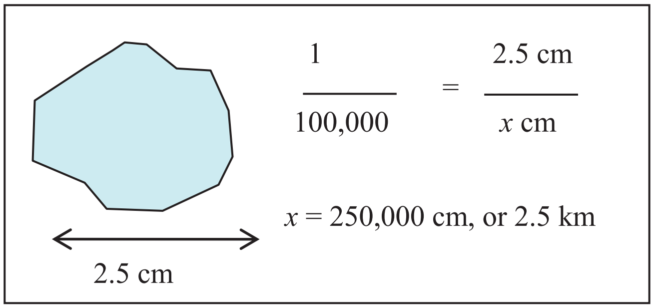


Fig. 1.9. Solving for the size of a lake

Scales for GIS data

When data are stored in a GIS, they technically do not have a scale because only the coordinates are stored. They acquire a scale once they are drawn on the screen or on paper. However, a data set has a **source scale**, or the original scale or resolution at which it was converted to digital form. A 1:1 million scale paper map that is scanned or digitized cannot be used effectively at larger scales. The map in Figure 1.10 shows congressional districts in pink and the state outlines in thick black lines. The state boundaries are more angular and less detailed than the districts because they were digitized at a smaller scale. Thus, although it is possible to take small-scale data and zoom in to large scales, the accuracy of the data will suffer. The source scale of a data set is an important attribute and is included when documenting it.

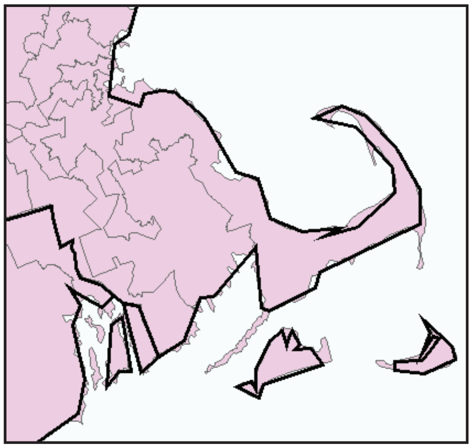


Fig. 1.10. These two layers showing Massachusetts originated from maps with different source scales.

One should exercise caution in using data at scales very different from the original source scale. Zooming in may give a false impression that the data are more precise than they actually are. A pipeline digitized from a 1:100,000 scale map has an uncertainty of about 170 feet in its location due to the thickness of the line on the paper. Displaying the pipeline on a city map at 1:10,000 might look all right, but it is likely to be many feet away from the actual location.

From looking at Figure 1.10, one might conclude that it is desirable to always obtain and use data at the largest possible scale. However, large-scale data require more data points per unit area, increasing data storage space and slowing the drawing of layers. Every application has an optimal scale, and little is gained by using information at a higher scale than needed.

Data quality

Representing real-world objects as points, lines, polygons, or rasters always involves some degree of **generalization**, or simplifying the data for digital storage, such as turning a house into a rectangular polygon, or even a point. No data file can capture all the spatial or attribute qualities of any object. The degree of generalization varies with the map scale. On a standard topographic map, a river has a width and can be modeled as a polygon with two separate banks. A city would be shown as a polygon area. For a national map, however, the river would simply be shown as a line, and a city would be shown as a point.

Even a detailed representation of an object is not always “true.” Rivers and lakes can enlarge in size during a spring flood event or shrink during a drought. The boundary of a city changes over time as the city grows. Users of GIS data must never forget that the data they collect and use will contain flaws, and that the user has an ethical and legal responsibility to ensure that the data used for a particular purpose are appropriate for the task. When evaluating the quality of a data set, geospatial professionals consider the following aspects.

Geometric accuracy describes how close the *x-y* values of a data set correspond to the actual locations on the earth’s surface. Geometric accuracy is a function of the source scale and of how the data were captured. Surveying is one of the most accurate ways to position features. GPS units have an accuracy that ranges from centimeters to tens of meters. Maps derived from aerial photography or satellite imagery can vary widely in geometric quality based on factors such as the scale of the image, the resolution of the image, imperfections and distortions in the imaging system, and the types of corrections applied to the image. In Figure 1.11, notice that the vector road in white is offset in places from the road as it appears in the aerial photo. These differences can arise from digitizing errors, geometric distortions from the camera or satellite, or other factors.

Fig. 1.11. Aerial photo near Woodenshoe Canyon, Utah.
Source: Google Earth and TeleAtlas.

Moreover, not every boundary can be as precisely located as a road. Imagine that you wish to delineate the land-cover types: *forest*, *shrubland*, *grassland*, and *bare rock* in Fig. 1.11. Where would you draw the line between *shrub* and *grassland*? At what point does the *shrubland* become *forest*? Six people given this photo would come up with six slightly different maps. Some boundaries would match closely; others would vary as each person made a subjective decision about where to place each boundary.

Thematic accuracy refers to the attributes. Some types of data are relatively straightforward to record, such as the name of a city or the number of lanes in a road. Even in this situation, the value of a feature might be incorrectly recorded. Other types of information can never be known exactly. Population data, for example, are collected through a process of surveying and self-reporting that takes many months. It is impossible to include every person. Moreover, people are born and die during the survey process, or are moving in and out of towns. Population data can never be more than an estimate. It is important to understand the limitations and potential biases associated with thematic data.

Resolution refers to the sampling interval at which data are acquired. Resolution may be spatial, thematic, or temporal. Spatial resolution indicates at what distance interval measurements are taken or recorded. What is the size of a single pixel of satellite data? If one is collecting GPS points by driving along a road, at what interval is each point collected? Thematic resolution can be impacted by using categories rather than measured quantities: if one is collecting information on the percent crown cover in a forest, is each measurement reported as a continuous value (32%) or as a classified range (10–20%, 20–30%)? Temporal resolution indicates how frequently measurements are taken. Census data are collected every 10 years. Temperature data taken at a climate station might be recorded every 15 minutes, but it might also be reported as a monthly or yearly average.

Precision refers to either the number of significant digits used to record a measurement or the statistical variation of a repeated single measurement. Many people confuse precision with accuracy, but it is important to understand the distinction. Imagine recording your body temperature with an oral digital thermometer that records to a thousandth of a degree and getting the value of 99.894 degrees Fahrenheit. This measurement would be considered precise. However, imagine that you take the reading immediately after drinking a cup of hot coffee. This action throws off the thermometer reading so that it does not record your true body temperature. Thus, the measurement is precise, but it is not accurate.

Evaluating the quality of a data set can be difficult, especially if the data were created by someone else. Professionals who create data usually also provide **metadata**, which stores information about the data set, such as where it came from, how it was developed, who assembled it, how precise it is, and whether it can be given to another person. The user can then decide whether a particular data set is suitable based on the information in the metadata.

Citing GIS data sources

Ethical and professional considerations require that any map, publication, or report should cite the data source(s) used and give proper credit to the originators of the data. The metadata, or the site from which the data were obtained, are good sources of information for citations. The best practice is to record the citation when the data are obtained so the information is available when needed. Generally one cites only data that are publically available (free or purchased). Data created internally within the workplace need not be cited, although often the company name or logo will appear on the map. A data set provided once in response to a personal request should be cited as a personal communication.

Keep in mind that the place you found the data may not always be its source. Your GIS administrator may have placed often-used data sets, such as ESRI Data and Maps, on a workplace fileserver for easy access, but the fact that they were obtained internally does not free you from the need to cite them, and you must cite the original source, not the local server.

General format

The purpose of a citation is to allow other people to find and obtain the data. It is not always possible to find all the information needed for a complete citation, but one should do one's best to make it complete as possible. The following general format for citations may be used:

Data set name (Year published) [source type]. Producer name, producer contact information. *Resource URL: [Date accessed].*

Data set name. The name is assigned by the creator or provider of the data.

Year published. Some data sets are assembled and provided once or at long intervals, and these are considered to have a publication date. For example, the ESRI Data and Maps product is released in revised form with each version of the software and carries a publication year. Aerial photography is flown on a particular date (although mosaics such as Google Earth use multiple sources spread over several years). It may take a little hunting or a few questions to find the publication date. Some data sets are updated at shorter intervals, or are even live. These can be assigned the current year.

Source type. Indicate the format in which the data are available. Types might include physical media (DVD, CD-ROM), a file downloaded from the Internet, or a service that provides live data