



Introduction to GIS

A Visual Companion

for Waldo Learning in ArcGIS Pro

The ideas behind every module of the course, in pictures: what a GIS is, how the Earth becomes a flat map, how data is stored, drawn, queried and analysed.



Waldo GIS

Introduction to GIS — A Visual Companion

CC BY-NC-ND 4.0 © the author. You may copy and share this book for free, in any medium, as long as you credit it, do not sell it and do not change it. Licence text: creativecommons.org/licenses/by-nc-nd/4.0/.

The licence covers the book's own text and the figures made for it. Figures and data from other sources keep their own licences; every one is listed with its credit line under **Figure credits** at the back of the book, and the course's data layers under **Course data sources**.

Waldo for ArcGIS Pro is made by Waldo GIS LLC, an independent company. ArcGIS and ArcGIS Pro are trademarks of Esri; this book is not an Esri publication and is not endorsed by Esri.

Waldo is named for the geographer Waldo Tobler, whose First Law of Geography reads: "Everything is related to everything else, but near things are more related than distant things."

The same content is published as this PDF and as a web edition; both carry the edition line below. Where the two differ, the later edition is the correct one.

How to read this book

This book is the concept companion to *Introduction to GIS*, the course that Waldo teaches inside ArcGIS Pro. You do the practical work in the lessons in the Waldo pane, such as opening data, drawing maps and running tools. This book explains the ideas behind those lessons and includes the figures that do not fit in the pane beside Pro.

One chapter for each module. Chapter 3 goes with Module 3, Chapter 8 with Module 8. Chapter 0 is the exception. It covers the setup from a new computer to the first lesson, and it is also published separately as the getting-started guide. You can read a chapter before its module, after it, or alongside it, because the chapters do not assume that you have done the lessons.

The book does not repeat the steps of a lesson. It explains what a coordinate system is, but it does not say which button sets one. Lesson boxes indicate where the module puts the concept to work.

The boxes are the same in every chapter:

WHY THIS MATTERS

Appears at the start of each chapter and gives the reason for studying the chapter's ideas.

YOU SHOULD BE ABLE TO EXPLAIN

Appears after it on the opening page and lists three ideas that you should be able to explain to someone in a sentence.

RULE OF THUMB

A working rule that the course follows, such as mapping rates rather than counts. The course's own maps and Waldo's advice follow these rules.

IN THE PANE

Lists the lessons in the matching module that use the idea and what they do with it. It names lessons but does not give the clicks.

TRY IT

A question to ask Waldo in the pane that demonstrates the idea on your own screen.

READING THE RESULT

Appears at the end of a worked example and states the question, the representation, what the result shows and what it cannot show.

NOTE

A short aside, such as a detail, an exception or a reference to another chapter.

Figures are numbered by chapter; their sources and licences are under Figure credits at the back. **Key terms** close each chapter and are collected in the Glossary.

The two editions have the same text. The PDF prints on US Letter and A4 without scaling. The web edition has one page per chapter and uses the same section names, so *Chapter 3, Scale* is 3-spatial-

reference#scale on the web and a bookmark in the PDF.

Contents

How to read this book	3
0 Start here	8
What you need	9
Installing ArcGIS Pro	10
Starting Pro the first time	10
Installing Waldo	11
Finding Waldo	11
Settings	12
The first thing to click	13
The course map	14
A project is a folder	15
1 What a GIS is, and the Pro workspace	18
The layer model	19
The ArcGIS Pro window	20
Maps, layouts and scenes	22
Moving around: extent and bookmarks	23
A feature has a row	24
Where data comes from: a first look	25
2 Data models and where data lives	28
The project and its data	29
Paths, and how links break	29
The vector model	30
The raster model	31
Vector or raster?	33
Containers: the geodatabase and the shapefile	34
Layers, datasets and other names	36
Where data comes from	37
Metadata	39
3 Spatial reference	42
The shape of the Earth	43
Datums: fitting the ellipsoid	44
Latitude and longitude	45
Map scale	47
Projection: from round to flat	48
What a projection keeps	49
Cylinder, cone and plane	51
The famous projections	52
Choosing a coordinate system	53
Web Mercator: fine for looking, wrong for measuring	54
Map units, on-the-fly projection and transformations	56
Why it matters	57
4 Cartographic design	61
Every map is generalized	62

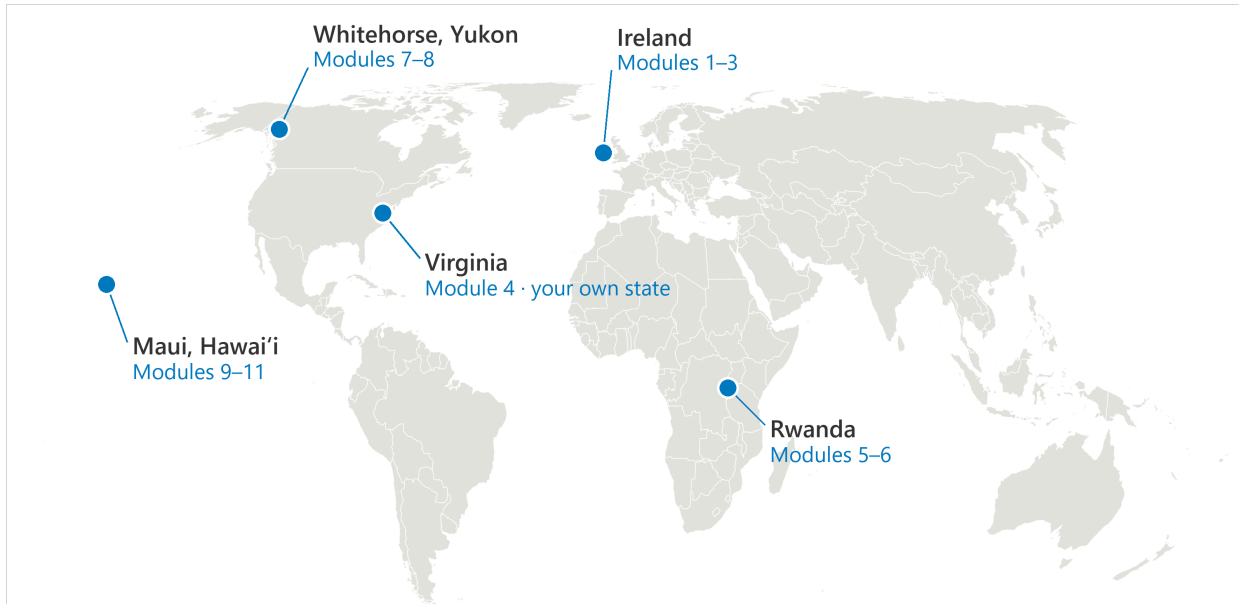
Reference and thematic maps	62
Visual variables	63
Colour	64
Thematic map types	65
Counts and rates	66
Classification	67
Typography and labels	69
Map elements and the layout	70
Design principles	71
5 Attribute data	75
Tables: records and fields	76
Three kinds of table	77
What a field can hold	77
Making and filling fields	79
Looking at a column	80
Joins: destination, source and key	81
Cardinality	82
Tables from outside the GIS	84
A rate from joined data	84
6 Exploring spatial data	88
Asking questions of data	89
John Snow's map, 1854	89
Attribute queries and SQL	90
AND, OR, NOT and the selection types	91
Keeping, viewing and clearing a selection	92
Spatial relationships	93
Select By Location in Rwanda	94
Attribute and spatial together	96
Summary statistics	97
Charts: which one answers what	97
7 Vector analysis	101
Analysis makes new data	102
Buffers	103
Clip and erase: the cookie cutters	104
Union and intersect: overlays with attributes	106
Dissolve, merge and append	107
A site-selection chain	108
Spatial joins: attributes by location	110
8 Raster analysis	116
The raster data model	117
Cell size is resolution	118
Bands and images	119
Drawing a raster	120
Making your own copy	121
Terrain: hillshade, slope and aspect	122
Map algebra	124
Boolean rasters	125
Reclassify	126

	Suitability: the habitat chain	127
9	Spatial patterns	131
	Tobler's first law	132
	Spatial autocorrelation	132
	Rain on Maui	133
	From samples to a surface	135
	Thiessen polygons	137
	Inverse distance weighting	138
	Checking a surface	139
	Kernel density	141
10	Creating and editing data	144
	Where new data comes from	145
	Vertices, segments and nodes	146
	Design before drawing	147
	Templates, tools and attributes	149
	Snapping	150
	Changing what you drew	151
	Saving edits	152
	Editing hygiene	152
11	3D scenes and your final map	155
	Two kinds of scene	156
	The ground is a DEM	157
	Moving in a scene	158
	Draping and extruding	159
	The summit's buildings	160
	One number or a field	161
	Your final map	162
	Where to make it	163
	Glossary	166
	Figure credits	173
	Course data sources	180
	Further reading	188

CHAPTER 0

Start here

Accompanies Module 0 · Setup and Foundations



Countries: World Countries (Generalized), Esri, ArcGIS Living Atlas of the World

WHY THIS MATTERS

This course is taught inside ArcGIS Pro by Waldo, a chat pane that can read your project. Before the first lesson you need three things on your computer: ArcGIS Pro, the Waldo add-in, and a model provider that supplies the language model Waldo uses. This chapter covers the setup up to the start of Module 0's first lesson, and then explains what Module 0 teaches: what a project is and how to keep it safe.

YOU SHOULD BE ABLE TO EXPLAIN

- why a project is a folder, and why deleting the .aprx keeps your data while moving the data breaks a map
- how Waldo checks your progress by reading your saved project, so you can stop at any step and come back
- what Ask mode lets Waldo do and not do, and why every lesson runs in it

IN THIS CHAPTER

What you need
Installing ArcGIS Pro
Starting Pro the first time
Installing Waldo
Finding Waldo

Settings
The first thing to click
The course map
A project is a folder

What you need

A Windows PC. ArcGIS Pro runs on 64-bit Windows 11 (and Windows Server). Esri lists 8 GB of memory as the minimum and 32 GB as recommended, with at least 32 GB of free disk space, ideally on a solid-state drive. On a Mac, Pro runs only inside a Windows virtual machine such as Parallels Desktop. Waldo needs **ArcGIS Pro 3.7 or later**; earlier releases cannot load it. Check Esri's system requirements page before you install (see Further reading).

An ArcGIS Pro licence. Pro will not start without one. There are four common ways to get one:

Route	Who it suits	Where to look
Your school or employer	Anyone whose university or organisation already licenses ArcGIS	Ask the GIS or IT office for an ArcGIS account; you sign in with it when Pro starts
ArcGIS Pro free trial	A work or government email address	Esri's trial page (21 days at the time of writing)
ArcGIS for Student Use	Students building skills while in school	Esri's Student Use page
ArcGIS for Personal Use	Personal, non-commercial learning	Esri's Personal Use page

The trial requires a business or government email address, and Esri directs students and hobbyists without one to the two non-commercial licences. Esri sets the prices and terms and changes them, so check them on Esri's pages rather than in this book.

A model provider. Waldo does not include its own artificial intelligence. You connect it to a model provider, and the provider's model reads your project and writes the explanations. Most providers identify you with an **API key**, a long secret string that you create on their website and paste into Waldo once. The free options are listed first in the table below.

Provider (as named in Waldo's Settings)	Cost	What you need
Google Gemini (free or paid tier)	Free tier available	A Google account; create a key in Google AI Studio
OpenAI-compatible → OpenRouter	Free models (ids ending in :free) or pay as you go	An OpenRouter account and key
GitHub Copilot (installed CLI)	Free for verified students (Copilot Student); paid plans	A GitHub account and the Copilot command-line tool
OpenAI-compatible → Ollama or LM Studio	Free, runs on your own computer	A powerful computer; slower and less capable than hosted models
Anthropic (Claude, API key)	Pay as you go	An Anthropic console account and key
OpenAI-compatible → OpenAI, DeepSeek, Groq, Mistral, xAI	Pay as you go	An account and key with that company
ChatGPT (sign in)	Your ChatGPT plan	A paid ChatGPT plan; no key, you sign in instead

The free tiers have usage limits, and Google's allows only a few requests a minute. Because a single Waldo reply can use several requests, expect pauses. Google also states that content sent on the free tier may be used to improve its products. OpenRouter's free models are capped at a small number of requests per day unless you have bought credits. Any provider is enough for the lessons, but a more capable model gives better explanations.

NOTE

Treat an API key like a password. Paste it into Waldo's Settings and nowhere else; Waldo stores it encrypted on your computer. If a key leaks, delete it on the provider's website and make a new one.

Installing ArcGIS Pro

Esri documents the installation of Pro, and you should follow its instructions (see Further reading). In outline, you sign in to your ArcGIS account, download the installer from the **Licenses** page, run the setup wizard, and then start Pro and sign in with the same account. On a school or company licence, an administrator often controls the download. If you see no **Download ArcGIS Pro** link, ask the person who gave you the account.

Starting Pro the first time

When Pro starts, it opens the **start page**, which lists your recent projects and the default project templates (Map, Catalog, Global Scene, Local Scene), with **Start without a template** beside them. Click **Start without a template**. Pro opens an empty, **Untitled** project with no map in it yet. The course begins from this empty project. Module 0 gives the project a name, and Module 1 adds the first map.

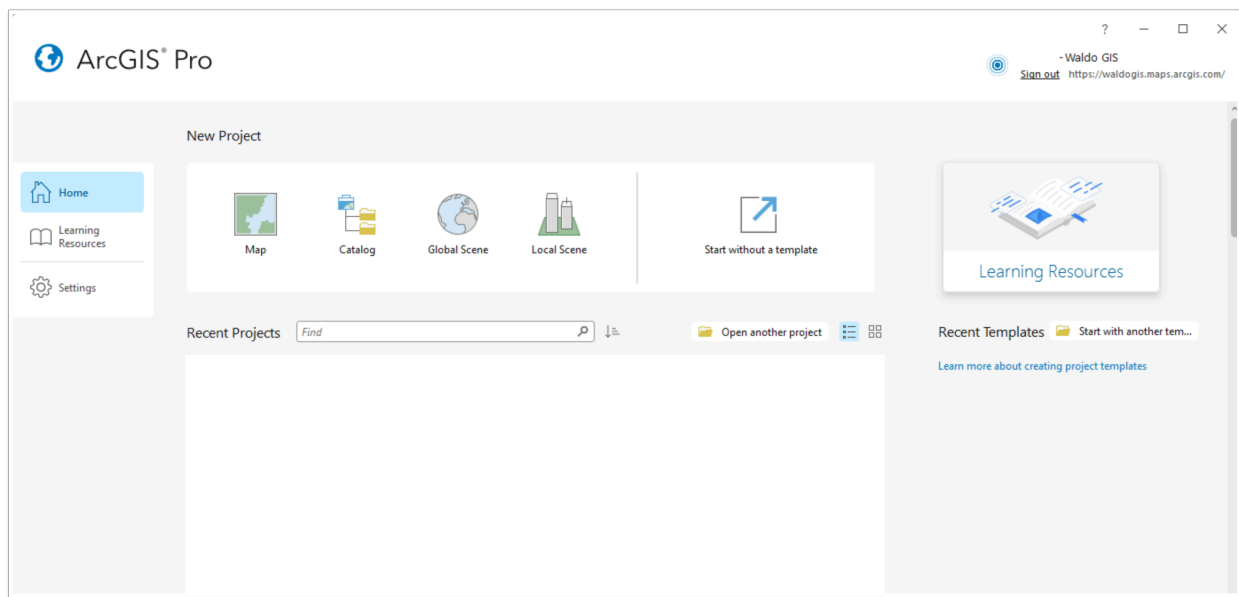


Figure 0.1 The ArcGIS Pro start page. Start without a template is beside the default project templates.

An Untitled project is temporary. Esri's documentation notes that it is deleted when you end the session unless you save it, so the first lesson of Module 0 is a Save As.

Installing Waldo

Waldo is an ArcGIS Pro **add-in**, distributed as a single file ending in `.esriaddinX` that you download from waldogis.com. The course lessons are included in that file, so there is no separate course download.

1. Close ArcGIS Pro. The add-in installs while Pro is closed.
2. Double-click the `.esriaddinX` file.
3. In the prompt that appears, click **Install Add-In**. No administrator rights are needed. The prompt also shows a notice in red that the file is not digitally signed. This notice is expected for this release, because Waldo is distributed directly from its own site rather than through a store.
4. Start Pro again.

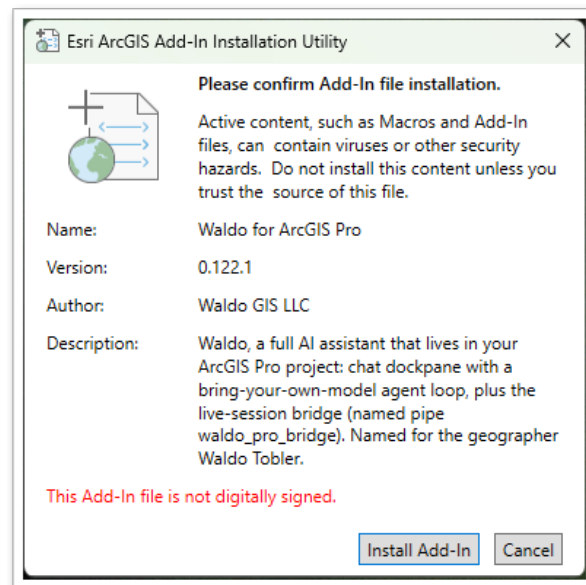


Figure 0.2 The ArcGIS Pro add-in installer, with the prompt to install Waldo.

To update Waldo later, repeat these steps with the newer file. The new version replaces the old one, and your settings and key are kept.

Finding Waldo

With a project open, look at the ribbon, the row of tabs across the top of Pro. The Waldo button is on the **Add-In** tab, in a group called **Waldo**. Clicking the **Waldo** button opens the pane, docked beside the map.

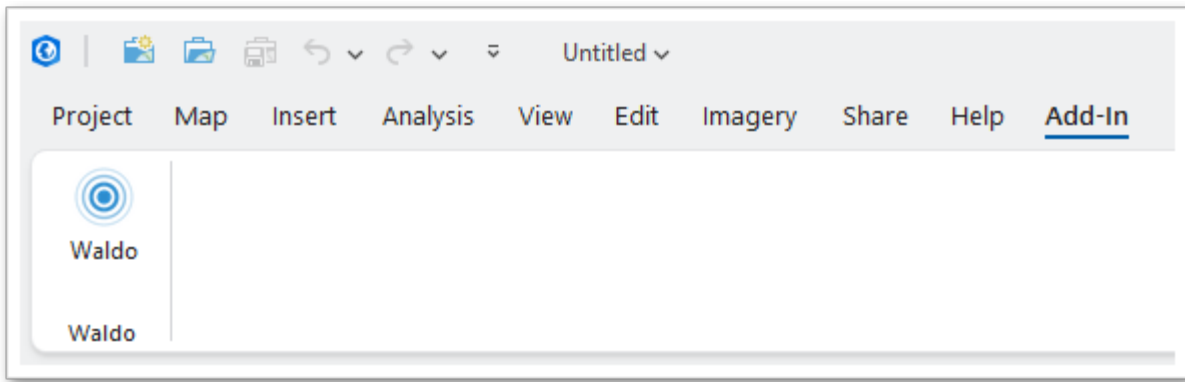


Figure 0.3 The Add-In tab of the ribbon, with the Waldo button.

If you close the pane, it stays closed until you reopen it, and Pro gives no indication of where it went. You always reopen it the same way, with the **Waldo** button on the **Add-In** tab. On a narrow window the ribbon may hide some tabs behind a » at its right end.

Settings

The first time the pane opens, click  **Settings** at the top of it. Under **Model provider**, choose your provider from the list, paste your API key into the key box (or, for the ChatGPT and GitHub Copilot entries, sign in as the panel explains), pick a model, and click **Save**. A link under the key box opens the provider's page for creating keys.

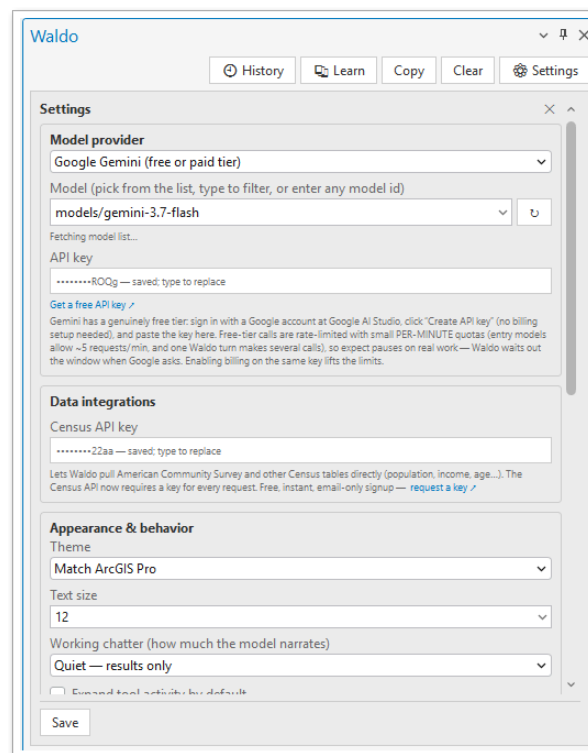


Figure 0.4 Waldo's Settings panel: the model provider, the model and the API key.

A new installation of Waldo starts a free seven-day trial in which Waldo can both explain and act. After the trial it continues in **Ask mode**, in which it can read your project and teach but cannot change anything. Every lesson in the course runs in Ask mode, so you can complete the course with or without a Waldo licence (licences are sold at waldogis.com).

The first thing to click

At the top of the Waldo pane, in the row of buttons beside **History** and **Settings**, is the **Learn** button (its icon is two small books). Click it, click the row **Module 0: Setup and Foundations**, and click **Start** beside its first lesson.

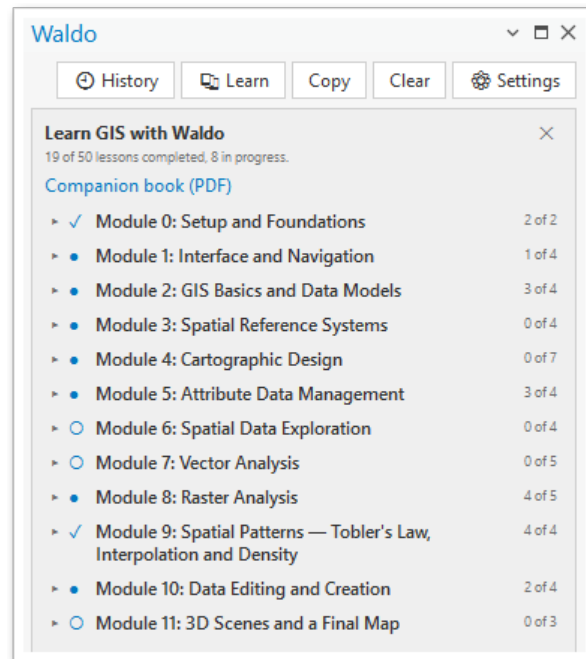


Figure 0.5 The Learn panel: every module of the course, each marked ○ not started, ● in progress or ✓ completed, with how many of its lessons are done.

The Learn panel lists the whole course. Each module row has a mark (○ not started, ● in progress, ✓ completed) and a count of the lessons you have done, and clicking a row shows its lessons. Each lesson has a **Start** button, which becomes **Resume** once you have begun. A started lesson also has a **Check my work** button. Clicking it runs the lesson's checks against your project and reports which of them pass. To come back another day, click **Learn**, then the module, then **Resume**. The **Companion book** link under the progress line opens this book at waldogis.com/learn, where it is free as a web edition and a PDF.

Module 0 has two short lessons. In the first you name and save your project, and the second explains how the lessons work. A lesson is a short series of steps that you carry out in Pro while Waldo explains them. Each of Waldo's replies ends with a few buttons that you can click instead of typing. You can also do everything the panel does by typing into the box at the bottom of the pane:

Type	What happens
start module 0	Begins a module at its first lesson
start lesson m01-103	Begins one particular lesson
list lessons	Shows every module and lesson
resume module 1	Picks up where you left off, even weeks later
explain more	Goes deeper on the current step
skip ahead	Moves past the current step
(close the pane or Pro)	Stops; your progress is kept, because Waldo reads it from the project (save the project first)

A **checkpoint** is the point at the end of a step where Waldo checks your work. Waldo does not rely on what you tell it. It reads your actual project, including the maps, layers and files in it, and reports what it found. Nothing is graded. A step is either done or not yet done, and because your progress is recorded in the state of your project, you can stop at any point and come back.

In a lesson Waldo acts as a teaching assistant and does not do the work for you. It explains, answers questions and checks your work, and you carry out the steps in Pro. When you carry out each operation yourself, you learn what it is for, how to predict its result and what the result does and does not show. A checkpoint checks your project, and the questions in this book test your reasoning.

TRY IT

Open the **Learn** panel after Module 0 and compare its list with the course map below. Then ask Waldo about any button you do not recognise. You can ask questions in the middle of any step.

The course map

The course has twelve modules and fifty lessons, which take roughly fourteen hours in all. Each module uses data from one place, and the course moves to a new place as the topics change. It uses Irish river basins for the basics, a US state for cartography, Rwanda's census for tables and queries, Whitehorse in the Yukon for vector and raster analysis, and Maui for spatial patterns, editing and 3D. The map on this chapter's opening page shows them. Module 4 uses your own US state, with Virginia as the worked example, and Module 11's final map can be of any place in the course.

Module	Title	Lessons	Minutes	Place
0	Setup and Foundations	2	16	Your own project
1	Interface and Navigation	4	52	Ireland
2	GIS Basics and Data Models	4	38	Ireland
3	Spatial Reference Systems	4	51	Ireland and the world
4	Cartographic Design	7	96	A US state (Virginia)

Module	Title	Lessons	Minutes	Place
5	Attribute Data Management	4	63	Rwanda
6	Spatial Data Exploration	4	69	Rwanda
7	Vector Analysis	5	122	Whitehorse, Yukon
8	Raster Analysis	5	112	Whitehorse, Yukon
9	Spatial Patterns	4	70	Maui, Hawai'i
10	Data Editing and Creation	4	57	Haleakalā, Maui
11	3D Scenes and a Final Map	3	75	Haleakalā, then your choice

The minutes are estimates for a first pass. Most lessons add their own data, so you can start almost anywhere. Later modules build on earlier ones, however (Module 8 continues Module 7's project, and Module 11 continues Module 10's), so it is easiest to take the modules in order.

A project is a folder

This idea will be important for the rest of the modules. An ArcGIS Pro **project** is a folder on your computer, and the file you open, the `.aprx`, is only one thing in it.

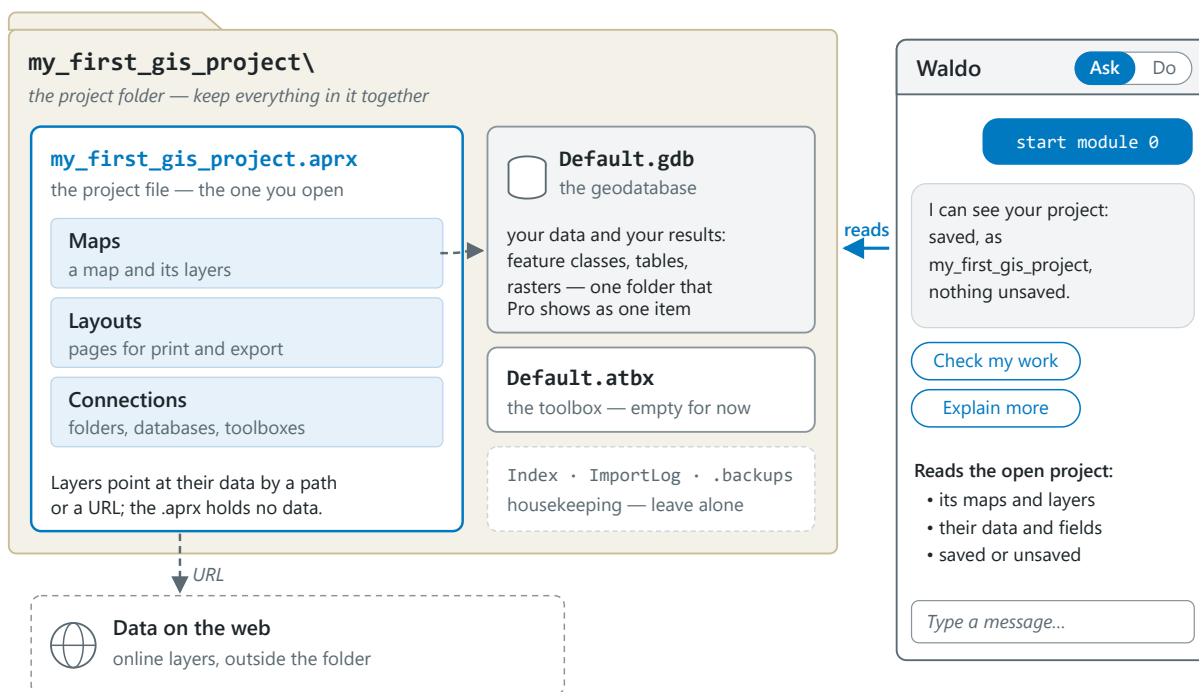


Figure 0.6 An ArcGIS Pro project is a folder. The `.aprx` holds maps, layouts and connections that point to data, the geodatabase holds the data, and the Waldo pane reads the open project.

- The `.aprx` file is the project itself: its maps, layouts and connections. A map's layers do not contain data; each one holds a path to data on your computer or a web address for data online.
- The **geodatabase** (`Default.gdb`) is where your own data and results are stored. In Windows Explorer it appears as a folder of files with unfamiliar names, but Pro shows it as a single item.

Never edit its contents from Explorer.

- The **toolbox** (Default.atbx) holds any tools you build. It stays empty in this course.
- Folders such as Index, ImportLog and .backups hold Pro's own housekeeping files.

This arrangement has three consequences. First, deleting the .aprx does not delete your data. Second, moving or renaming data breaks the path that the project stores, and the layer shows as broken until it is repaired. Third, your own work belongs in the geodatabase inside the project folder, so that the whole folder can be moved or copied as one unit.

RULE OF THUMB

Name projects, folders and data in lowercase, with words joined by underscores and no spaces: my_first_gis_project. Spaces and capital letters can cause problems in paths, scripts and some tools, and names written this way avoid those problems.

Save and Save As. **Save As** (on the **Project** tab) gives the project a name and a folder, and you use it once, at the start. **Save** writes your changes to that same project. You use it often, from **Project > Save** or from the small disk button on the Quick Access Toolbar at the top-left of the window. The button is greyed out when there is nothing new to save and becomes active when you change something. Data that you draw or edit is saved separately, as Chapter 10 explains.

RULE OF THUMB

Save often, and always save before closing Pro. A saved project reopens as you left it, although layers from the web still depend on their services. An Untitled project that was not saved is deleted.

Ask and Do. At the top of the Waldo pane is a two-way switch. In **Ask** mode Waldo can read your project and explain anything, but it cannot change anything. In **Do** mode it can make changes, such as running tools, styling layers and building layouts. The lessons run in Ask mode because you learn by doing the steps yourself. Outside a lesson, you can also learn by switching to Do, watching Waldo perform a task once and then doing it yourself.

IN THE PANE

In Module 0 lesson 1 you save the project with Save As, open the project folder in Explorer to see the .aprx, the geodatabase and the toolbox, and find both Save buttons. Waldo then checks that the project has a real name and nothing unsaved. Lesson 2 shows the course map, the Ask/Do switch (on some installs it is fixed on Ask, and the lessons never need Do), the Learn panel and its Companion book link, the words that move you between lessons and the Add-In tab, and Waldo demonstrates a checkpoint by reading your project live.

Key terms

project	an ArcGIS Pro workspace: a folder holding the .aprx file, a default geodatabase, a toolbox and housekeeping folders.
.aprx	the project file you open; it stores maps, layouts and connections, and points at data rather than containing it.
geodatabase	a container for GIS data (feature classes, tables, rasters); a file geodatabase is a .gdb folder that Pro shows as one item.
add-in	a package that extends ArcGIS Pro with new buttons and panes, installed from an .esriAddinX file.
pane	a dockable window inside ArcGIS Pro, such as Contents, Catalog or Waldo.
provider	the company or program that runs the language model Waldo uses.
API key	a secret string that identifies you to a provider; treat it like a password.
Ask mode	Waldo reads and explains your project but cannot change it; every lesson runs in Ask.
Do mode	Waldo can act on your project: run tools, change symbology, build layouts.
checkpoint	the check at the end of a lesson step, made by reading your actual project.
Learn panel	the list of every module and lesson in the Waldo pane, with your progress marked and Start, Resume and Check my work buttons.
lesson	a short sequence of steps, done in Pro, that teaches one idea.
module	a group of lessons on one topic; the course has twelve, numbered 0 to 11.

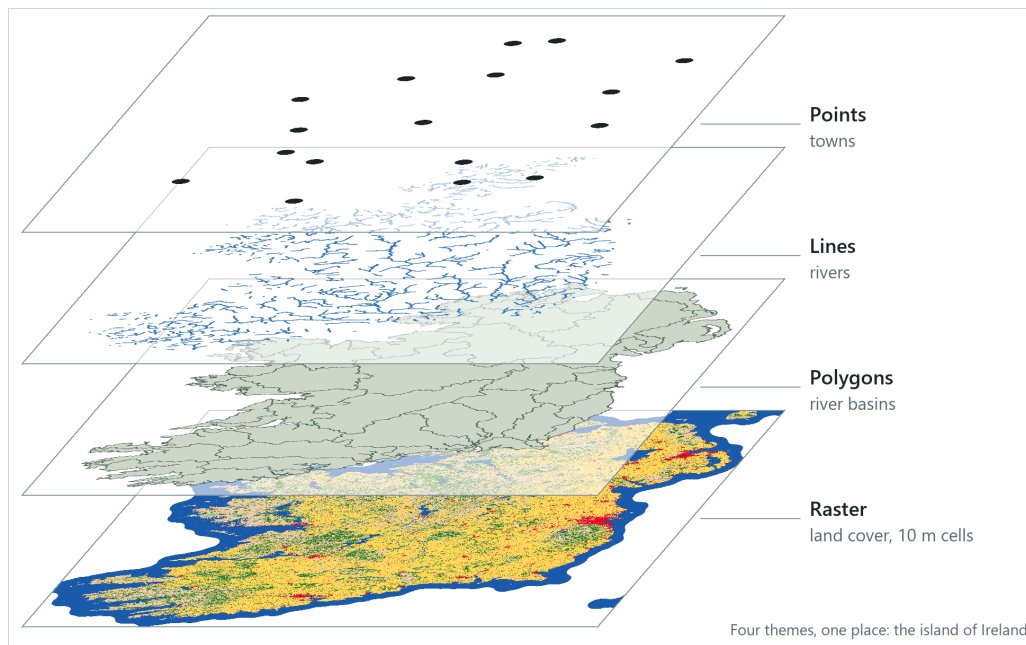
Further reading

- Esri. *ArcGIS Pro 3.7 system requirements*. ArcGIS Pro documentation. <https://doc.esri.com/en/arcgis-pro/latest/get-started/arcgis-pro-system-requirements.html>
- Esri. *Download ArcGIS Pro and Install ArcGIS Pro*. ArcGIS Pro documentation. <https://doc.esri.com/en/arcgis-pro/latest/get-started/download-arcgis-pro.html> and <https://doc.esri.com/en/arcgis-pro/latest/get-started/install-and-sign-in-to-arcgis-pro.html>
- Esri. *Create a project*. ArcGIS Pro documentation (the start page, templates and Start without a template). <https://doc.esri.com/en/arcgis-pro/latest/help/projects/create-a-new-project.html>
- Esri. *Run ArcGIS Pro on a Mac*. ArcGIS Pro documentation. <https://doc.esri.com/en/arcgis-pro/latest/get-started/run-pro-on-a-mac.html>
- Esri. ArcGIS Pro free trial, ArcGIS for Student Use and ArcGIS for Personal Use. <https://www.esri.com/en-us/arcgis/products/arcgis-pro/trial>, <https://www.esri.com/en-us/arcgis/products/arcgis-for-student-use/overview> and <https://www.esri.com/en-us/arcgis/products/arcgis-for-personal-use/overview>
- The course's data sources: every layer the lessons use, who publishes it and its licence.

CHAPTER 1

What a GIS is, and the Pro workspace

Accompanies Module 1 · Interface and Navigation



Basins: Hydrometric Areas Ireland & Northern Ireland; rivers: Rivers Network of Ireland (ArcGIS Living Atlas of the World). Source: Environmental Protection Agency (Ireland), CC BY 4.0. Land cover: Sentinel-2 10m Land Use/Land Cover 2024, Impact Observatory, Microsoft and Esri, ArcGIS Living Atlas of the World (source data CC BY 4.0)

WHY THIS MATTERS

A GIS stores information about a place in layers, one theme each, and draws them together because they share the same locations. The figure above shows Ireland as four stacked layers: towns, rivers, river basins and land cover. This chapter explains the layer model, the parts of the ArcGIS Pro window and how a map view moves, and introduces an idea the rest of the course depends on: every feature on the map has a row of information in a table.

YOU SHOULD BE ABLE TO EXPLAIN

- why a layer is a set of drawing rules pointing at a dataset, and why switching it off deletes nothing
- why the order of the Contents list controls what you see, and how one filled layer can hide the rest
- why every feature on the map is also a row in a table, and what the table shows that the map does not

IN THIS CHAPTER

The layer model
The ArcGIS Pro window
Maps, layouts and scenes

Moving around: extent and bookmarks
A feature has a row
Where data comes from: a first look

The layer model

A GIS links the locations of things to information about them, so that you can ask and answer questions about places. In GIS work, an analyst starts with a question, chooses data for it, carries out an analysis and interprets what the result shows and what it cannot show. The people doing the work decide on the question and judge the answer. An example of the kind of question this course asks is which river basins have few survey observations, and whether that means the rivers there are healthy or that nobody has looked. This chapter starts with how a GIS organises its information.

A GIS does not store a map as one picture. It divides the information about a place into **layers**, each holding a single theme for the same area: one layer of towns, one of rivers, one of river basins, one of land cover. Because every layer records where its contents are located, the layers line up and can be drawn, compared and analysed together. The figure on the opening page shows the idea as a cutaway, with each theme drawn on its own sheet.

Layers come in two basic kinds. Most hold **features**: points (towns), lines (rivers) or polygons (river basins), each a separate object with its own shape. Others hold a grid of **cells**, a **raster**, in which every cell holds a value. The land-cover sheet is a raster of cells 10 m across, each one coded as water, trees, crops, built area and so on. Chapter 2 explains the difference. For now, it is enough to know that both kinds are stacked in the same way.

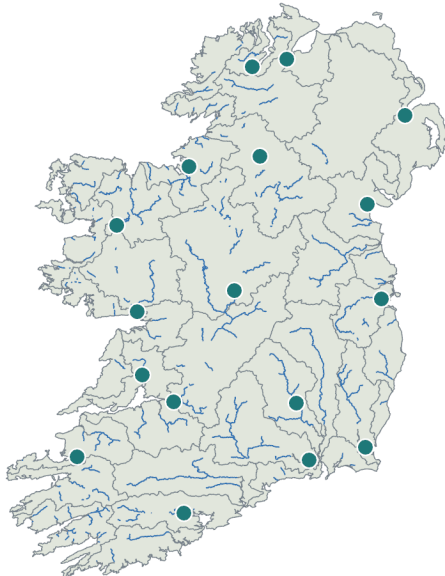
A layer displays a dataset without containing it. The layer holds a name, a set of drawing rules (its **symbolology**) and a reference to the dataset, which stays where it is stored. As a result, the same dataset can appear in many maps, in different colours and under different names, and turning a layer off with the checkbox beside it only stops it from being drawn. Nothing is deleted.

The layers of a map are listed in the **Contents** pane, and the order of the list is the order in which they are drawn. The layer at the top of the list is drawn on top of the others, and the layer at the bottom is drawn underneath everything. The **basemap**, the background reference map, is therefore placed at the bottom. The order affects the map wherever layers overlap, and only filled shapes hide the layers beneath them. A polygon layer drawn as outlines only can be at the top of the list without covering anything.

Towns and rivers above the basins

Drawing order (top of the list draws on top):

1. towns
2. rivers
3. river basins



River basins dragged to the top

Drawing order (top of the list draws on top):

1. river basins
2. towns
3. rivers

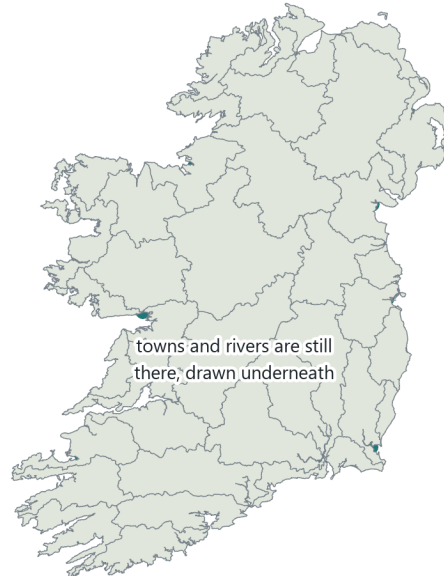


Figure 1.1 The same map in two drawing orders. With the towns and rivers above the basins, all three themes are visible. When the filled basins are dragged to the top of the list, they cover the other layers. The towns and rivers are still in the map, drawn underneath.

Basins: Hydrometric Areas Ireland & Northern Ireland; rivers: Rivers Network of Ireland (ArcGIS Living Atlas of the World). Source: Environmental Protection Agency (Ireland), CC BY 4.0

RULE OF THUMB

Keep points above lines, lines above polygons, and the basemap at the bottom. If a layer seems to have disappeared, check the drawing order and the checkbox first.

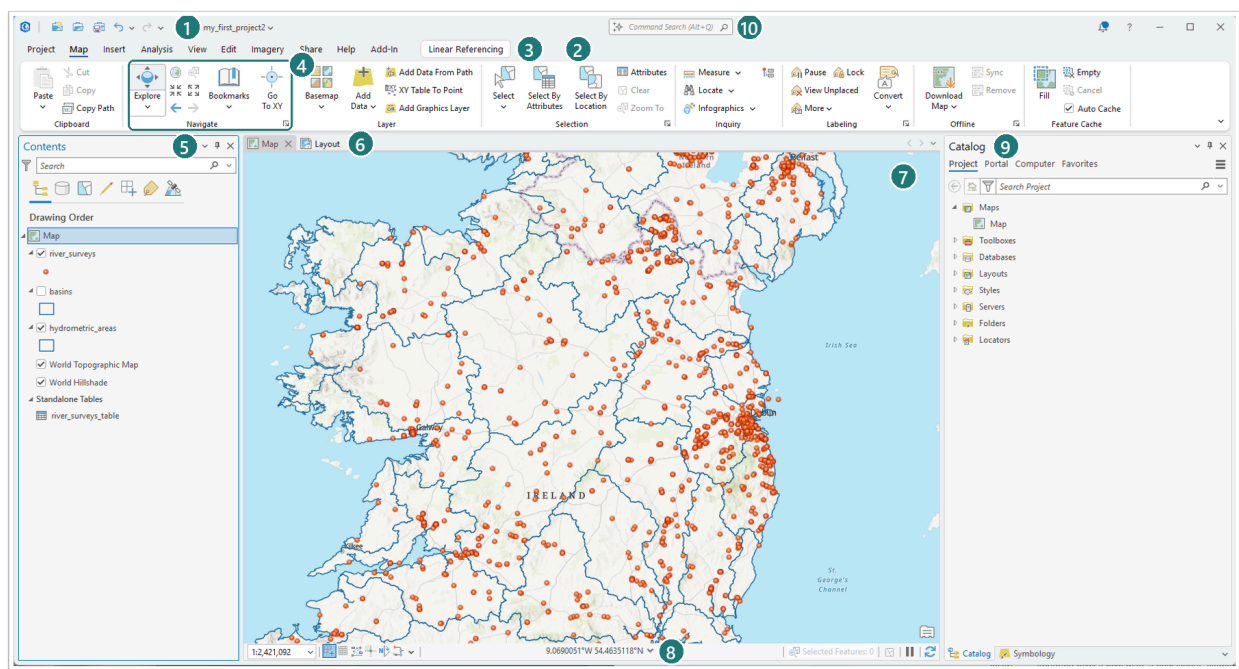
IN THE PANE

Module 1 lesson 1 is a tour of the window, which includes the Contents pane and explains what a layer is and how the list sets the drawing order. In lesson 4 you drag the river basins above the river survey points and back again. The basins added from the Living Atlas are drawn as outlines only, so they hide nothing when they are on top. The outlines are drawn over the points, and every point is still visible. Waldo reads the basins' symbology and the new order from your map.

The ArcGIS Pro window

The Pro window has three kinds of component. Across the top is the **ribbon**. In the middle are **views**, where maps and layouts are drawn. At the sides are **panes**, dockable windows that list and manage the contents of the project. If you know which kind of component something is, you know where to look for it.

1 What a GIS is, and the Pro workspace



- 1 **Quick Access Toolbar**
save, undo, redo: always in reach
- 2 **Ribbon tabs**
the row Project, Map, Insert, Analysis, View ... each a set of commands
- 3 **Contextual tab**
shown only while something needs it (here Linear Referencing)
- 4 **Group**
related commands on a tab; Navigate holds Explore and Bookmarks
- 5 **Contents pane**
the layers of the active map, in drawing order
- 6 **View tabs**
each open map, layout or table is a view; one is active
- 7 **Map view**
the map itself, drawn from the Contents list
- 8 **Scale and coordinates**
the view's current scale and the pointer's position
- 9 **Catalog pane**
everything the project holds or can reach
- 10 **Command Search**
find any command or tool by name

Figure 1.2 The ArcGIS Pro window with the Module 1 map open. The ribbon's tabs and groups run across the top, the Contents pane is on the left, the Catalog pane on the right and the map view in the middle, with a tab for each open map or layout. The numbers are explained in the key below the picture.

On the ribbon, commands are organised on **tabs**, and each tab arranges related commands in **groups**. The course uses four tabs most of the time:

Tab	What it contains
Map	Moving around the map (the Explore tool), bookmarks, the basemap gallery, adding data, selecting features
Insert	Anything new: a new map, a new layout, elements to place on a layout
Analysis	The geoprocessing tools, more than a thousand of them, and the pane that runs them
View	Which panes are open, and the Pane Sets menu that restores a standard arrangement

The ribbon also has the Project tab (save, options), Edit, Imagery, Share, Help and, once Waldo is installed, the Add-In tab that holds its button (Chapter o). A **contextual tab** appears only while something that it applies to is selected. When you select a layer in Contents, a Feature Layer tab is added to the row, and when you select something else, the tab disappears.

In a new map, the two open panes are Contents and Catalog. **Contents** lists the layers in the active map, in drawing order. **Catalog** lists the items in the project. Its Project tab lists the maps,

databases, folders and toolboxes in the project, and its Portal tab gives access to data online. A pane that has been closed is reopened from its own button on the View tab. The status row under the map shows the view's current scale and the position of the pointer, and Chapter 3 explains both of these numbers.

RULE OF THUMB

When Pro seems not to be working correctly, the cause is usually a pane that has been closed or moved. Choosing Mapping from the Pane Sets menu on the View tab returns every standard pane to its usual position.

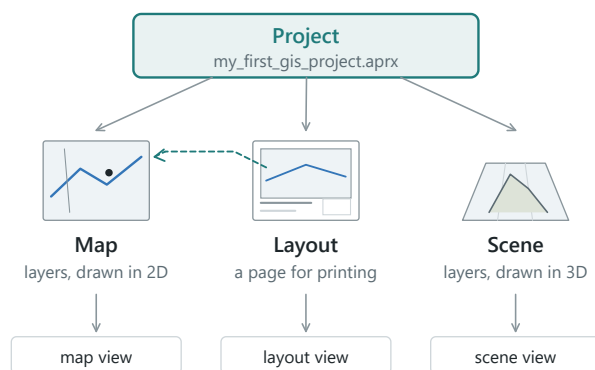
IN THE PANE

Module 1 lesson 1 goes through the ribbon one tab at a time (Map, Insert, Analysis, View) and shows the Pane Sets menu without choosing anything from it. It then moves to the Contents pane, the map view and the Catalog pane. The lesson asks you to look rather than to answer questions, and each step ends with a confirmation such as "I see it" or "My screen looks different".

Maps, layouts and scenes

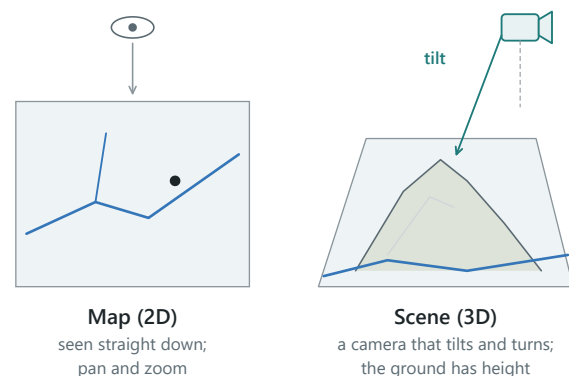
A project can hold many items, and the three that you will draw with are maps, layouts and scenes. A **map** is a set of layers drawn flat, in two dimensions, in one coordinate system. A **layout** is a page that arranges one or more maps in **map frames**, with a title, legend, scale bar and other elements, for printing or export. A **scene** draws layers in three dimensions over a ground surface that has elevation.

Each item opens in a view of its own: a map view, a layout view, a scene view, and a table view for an attribute table. Several can be open at once as tabs across the middle of the window, but only one is the **active view**, and the Contents pane always shows the layers of the active view. The course works in maps for most of its length. Layouts are covered in Chapter 4, and scenes in Chapter 11.



A layout's map frame shows one of the project's maps.
Tables, toolboxes and connections live in the project too.

Figure 1.3 A project holds maps, layouts and scenes, and each opens in its own view. A layout's map frame shows one of the project's maps.



The same layers can be drawn either way.

Figure 1.4 A map is viewed from directly above and is moved by panning and zooming. A scene has a camera that can also tilt and turn.

Moving around: extent and bookmarks

A map view shows one part of the map at a time. The part that it shows at any moment is its **extent**, the rectangle of the map in view, at a particular scale. **Panning** moves the view across the map without changing the scale. **Zooming** changes the scale, and therefore how much of the map fits in the view. In Pro you do both with the Explore tool, the hand icon at the left end of the Map tab. You drag to pan and turn the mouse wheel to zoom. In a scene the camera can also tilt and turn.

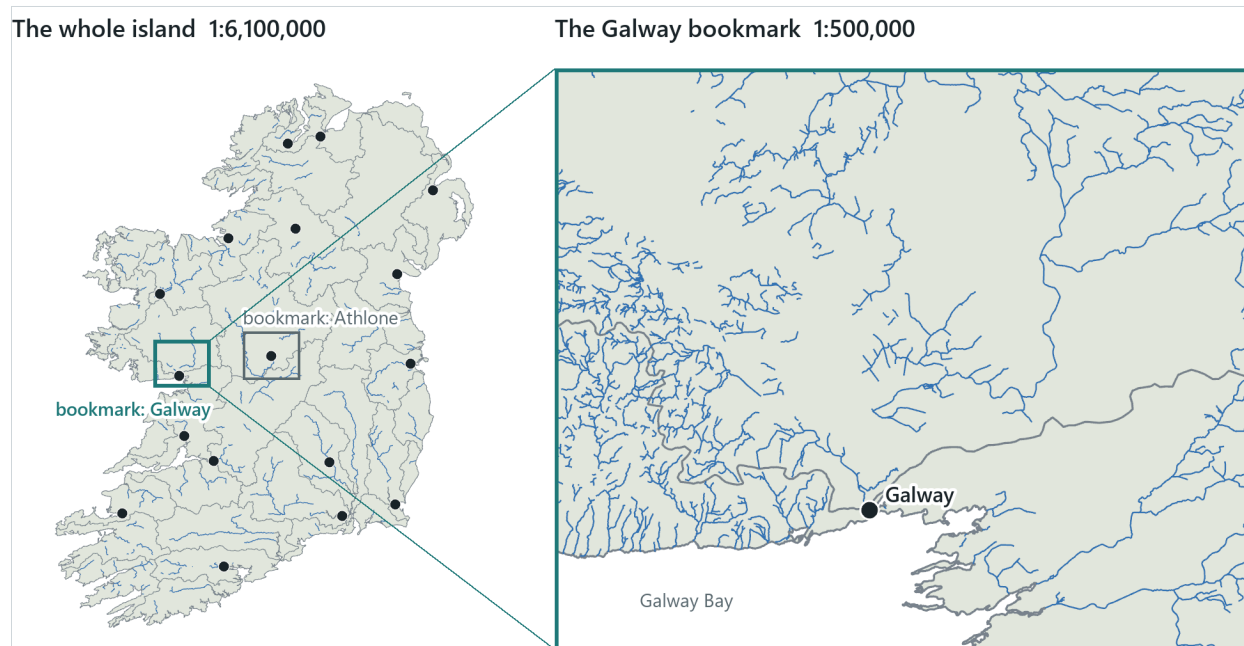


Figure 1.5 Two extents of the same map. On the left the whole island fits at about 1:6,100,000, with two saved extents drawn on it. On the right is the view that the Galway bookmark returns to, about 46 km across at 1:500,000. Both scales are true at the printed size.

Basins: Hydrometric Areas Ireland & Northern Ireland; rivers: Rivers Network of Ireland (ArcGIS Living Atlas of the World). Source: Environmental Protection Agency (Ireland), CC BY 4.0

A **bookmark** saves an extent under a name so that the view can return to it at once. It belongs to the map in which it was made and is saved with the project. In a scene, a bookmark saves the camera position, including its tilt. The Galway bookmark in the figure stores the extent of the rectangle drawn on the left. Chapter 3 explains scale itself, and what "large" and "small" scale mean.

RULE OF THUMB

Name a bookmark for its place ("Galway", "Home"), never "Bookmark 1". Descriptive names make the list of bookmarks easy to use.

IN THE PANE

In Module 1 lesson 2 you practise moving around the map. You pan and zoom to a place you know and then save a bookmark with a real name. Waldo compares the view before and after to confirm that it has changed, and checks that your bookmark is saved in the project.

A feature has a row

GIS data links location to information. Every feature on the map has a row of **attributes** that describe it, and there are two ways to read them. A **pop-up** shows one feature's attributes when you click its shape. The **attribute table** shows every feature at once, with one row per feature and one column per field.

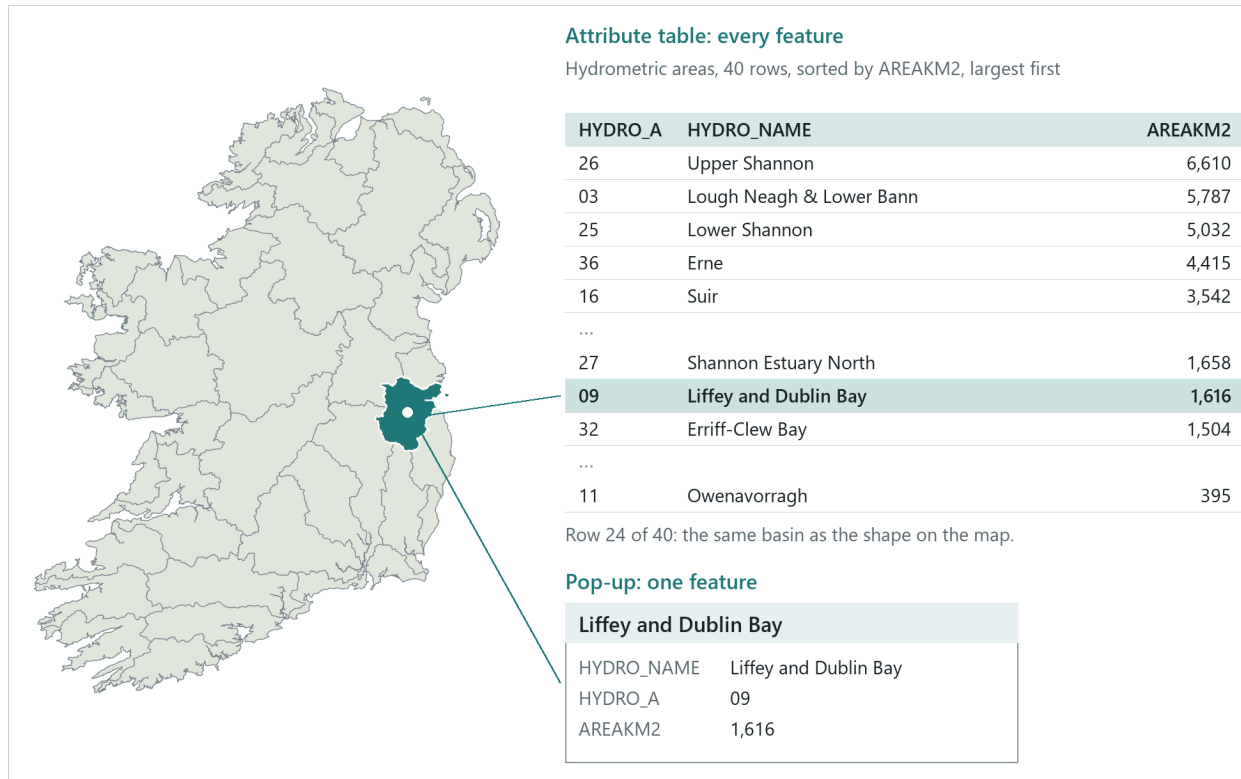


Figure 1.6 The same basin in a pop-up and in the attribute table. When you click the Liffey and Dublin Bay hydrometric area, its pop-up shows its name, code and area. In the attribute table, sorted by area, the same basin is row 24 of 40.

Basins: Hydrometric Areas Ireland & Northern Ireland. Source: Environmental Protection Agency (Ireland), CC BY 4.0

The first table in the course is Ireland's hydrometric areas. These are the river basins that the water agencies of Ireland and Northern Ireland use to organise the island's hydrology, in one layer that covers both jurisdictions. It has exactly 40 rows, one for each of the island's 40 hydrometric areas. A hydrometric area is a reporting unit and does not always correspond to the basin of one river. One area may be a single large basin and another may group several small coastal catchments, and the Shannon's basin is split between two areas. The course calls them the river basins for short. The area that covers Dublin, named Liffey and Dublin Bay, has the code 09 and an area of 1,616 km². The map and the table show the same data. When you select a row in the table, its shape is highlighted on the map.

Some questions can be answered from the table but not from the map. Counting the rows gives the number of hydrometric areas. Sorting the area column from largest to smallest puts the Upper Shannon at the top with 6,610 km², followed by Lough Neagh & Lower Bann at 5,787 km², and the Owenavorrhagh at the bottom with 395 km². Chapter 5 covers tables in more depth, including fields and their types, calculations and joins.

IN THE PANE

In Module 1 lesson 3 you add the hydrometric areas to your map from the Living Atlas, click a basin you know and read its pop-up, and then open the attribute table to count the rows and sort by area. Waldo checks each answer against the layer itself before saying whether it is right.

Where data comes from: a first look

Data can be added to a map in four main ways. You search a **portal** such as ArcGIS Online, and the **Living Atlas**, Esri's curated collection of ready-to-use layers within it, from the Portal tab of the Catalog pane. A **service URL**, the web address of one published layer, is added directly. The project's own **geodatabase** holds your data and every result you make, and is empty in a new project. A **folder connection** points Pro at any folder of shapefiles, rasters or spreadsheets on a disk.

Four doors in, one place they all lead: a layer in the map

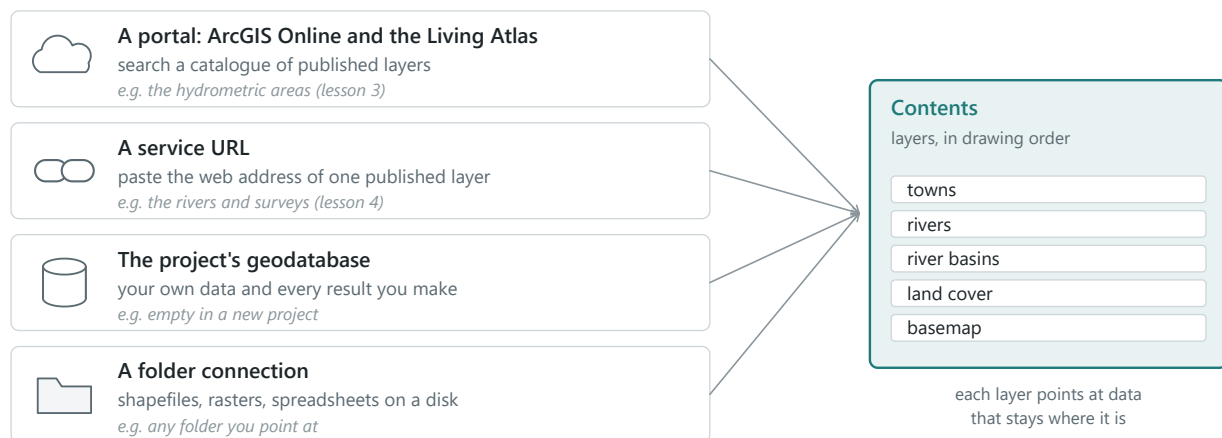


Figure 1.7 The four ways of adding data, drawn as doors that lead to the same place. Whichever way you add data, it appears as a layer in the Contents list, and the layer points to data that stays where it is.

Module 1 uses two of these ways. You add the river basins from the Living Atlas by searching for them by name on the Portal tab. You add the rivers and the river surveys by URL. They are two small copies that the course keeps on ArcGIS Online so that they draw quickly and do not change from week to week. The rivers are the larger rivers of the Environmental Protection Agency's Rivers Network of Ireland, 22,856 of its 102,108 segments, with their lines simplified. The full network is in the Living Atlas, but it is slow to draw over the web at the scale of the whole island. The surveys are about 1,800 points from The Big River Watch, a layer of citizen river surveys from Ireland and Britain, limited to those inside the island's river basins. The full layer holds about 20,000 points and grows every week. Chapter 2 explains what these datasets are and where each one is stored.

IN THE PANE

In Module 1 lesson 4 you add the rivers and the river surveys by their links (Add Data, Data From Path) and drag the basins over the survey points to see how drawing order affects an outline. You also look at the project's geodatabase in the Catalog pane, and the lesson names folder connections. Waldo confirms that each new layer is in your map rather than part of the basemap.

TRY IT

With the Module 1 map open, ask Waldo in Ask mode: list the layers in my map from top to bottom, and tell me which one would hide the others if it were on top. Waldo answers by reading the layers in your Contents pane.

Key terms

layer	one dataset drawn in a map with its own name and symbology; the data stays where it is.
feature	one point, line or polygon in a layer, together with its row of attributes.
raster	a layer made of a regular grid of cells, each holding one value.
attribute	a value describing a feature, stored in one field of its row.
attribute table	one row per feature, one column per field.
pop-up	the window that shows one feature's attributes when the feature is clicked.
drawing order	the order of layers in the Contents pane; a layer higher in the list draws on top.
basemap	the reference layers at the bottom of a map that give it context.
symbology	the drawing rules of a layer: its colours, symbols and sizes.
ribbon	the band of tabs across the top of Pro, each holding groups of commands.
contextual tab	a ribbon tab that appears only while something that needs it is selected.
view	a work area showing a map, layout, scene or table; one view is active.
map	a project item holding layers drawn in two dimensions.
layout	a page in a project that arranges maps and map elements for printing or export.
scene	a project item that draws layers in 3D, seen through a camera that tilts and turns.
extent	the rectangle of the map shown in a view at a given moment.
bookmark	a saved extent (in a scene, a saved camera) that a view can return to by name.
Living Atlas	Esri's curated collection of ready-to-use layers, reached from the Portal tab.
service URL	the web address of one published layer, which can be added to a map directly.

Further reading

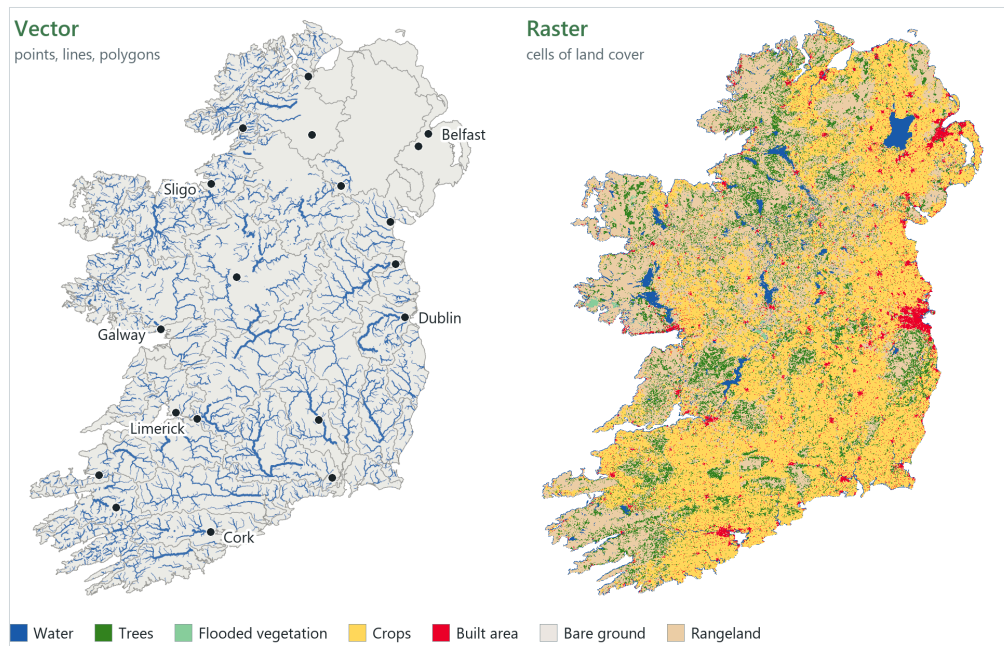
- Olaya, V. *Introduction to GIS*, the chapter on data (the layer concept). <https://volaya.github.io/gis-book/en/>
- Esri. *ArcGIS Pro user interface*. ArcGIS Pro documentation. <https://doc.esri.com/en/arcgis-pro/latest/get-started/user-interface.html>
- Esri. *Bookmarks*. ArcGIS Pro documentation. <https://doc.esri.com/en/arcgis-pro/latest/help/mapping/navigation/bookmarks.html>

- The course's data sources: every layer used, with its publisher and licence.

CHAPTER 2

Data models and where data lives

Accompanies Module 2 · GIS Basics and Data Models



Basins: Hydrometric Areas Ireland & Northern Ireland (ArcGIS Online); rivers: Rivers Network of Ireland (ArcGIS Living Atlas of the World). Source: Environmental Protection Agency (Ireland), CC BY 4.0. Land cover: Sentinel-2 10m Land Use/Land Cover Time Series, most recent year, Impact Observatory, Microsoft and Esri (ArcGIS Living Atlas of the World; source data CC BY 4.0). Towns: Natural Earth (public domain)

WHY THIS MATTERS

A map looks like one picture, but it is assembled from separate datasets. Each dataset is stored in a particular place and describes the world in one of two ways. The figure on the opening page shows Ireland twice, once as points, lines and polygons and once as a grid of land-cover cells. This chapter explains the two models, the formats in which data is stored, how a project locates its data and where to find more data.

YOU SHOULD BE ABLE TO EXPLAIN

- why a project stores pointers rather than data, and why a relative path lets the whole folder be moved
- when a theme suits the vector model and when the raster model, and what sets the detail in each
- why more detail is not more accuracy, and how metadata shows whether a dataset fits your purpose

IN THIS CHAPTER

The project and its data

Paths, and how links break

The vector model

The raster model

Vector or raster?

Containers: the geodatabase and the shapefile

Layers, datasets and other names

Where data comes from

Metadata

The project and its data

A Pro **project** is a folder (Chapter 0). The file you open, the **.aprx**, holds maps, layouts and connections. Because every layer draws when you open it, the file appears to hold the data as well, but it does not. For each layer the project stores only a **pointer**, the path or web address where the data is actually stored, together with instructions for drawing it. The data itself stays outside the file.

A pointer can refer to three kinds of location. It can be a path to a geodatabase or a shapefile on your own computer, a path to a folder on a shared network drive, or the web address (URL) of a service on a server. A layer that points to the web draws only while the server responds, and it goes blank if the internet connection is lost. A copy on your own disk still draws when you are offline.

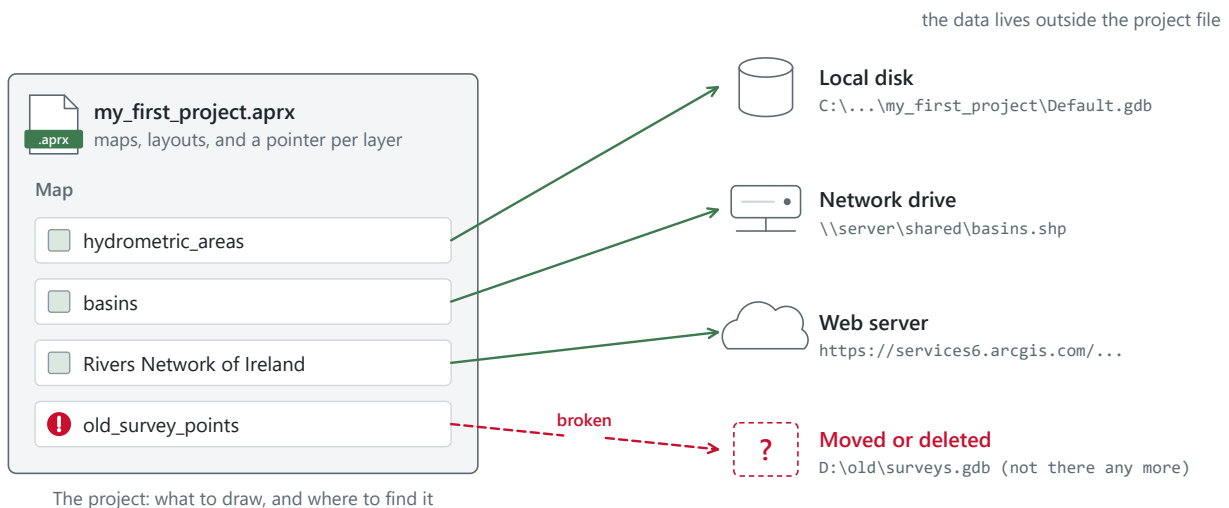


Figure 2.1 A project stores a set of pointers. Each layer in the map points to its data: a geodatabase on the local disk, a shapefile on a network drive, a service on a web server. When the data that a pointer refers to is moved or deleted, the link breaks.

The layer's **Source** tab shows its pointer. For a web layer it is a URL beginning `https://`. For a local copy it is a **Database** path ending in a geodatabase name, with the dataset's name beside it. Because the project contains only the pointers, an **.aprx** sent on its own to someone else contains none of your local data. The layers that point to web services still draw, but the layers that point to files on your disk do not.

IN THE PANE

In Module 2 lesson 1 you read the pointer of the web basins layer on its Source tab. You then copy the basins and the river surveys into the project's own geodatabase with the Export Features tool and read the new, local pointer. Waldo checks that the copies are in the geodatabase and that the path you read ends in the right place.

Paths, and how links break

A path can be written in two ways. An **absolute path** starts at the top of the file tree, the drive letter, and names every folder on the way down to the data. A **relative path** starts from the folder that holds the project file and gives only the route from there to the data, for example to a

geodatabase in the same folder as the project file. Pro stores paths relative to the project whenever the data is on the same drive as the project.

As a result, a project folder can be copied to another disk, zipped and emailed, or opened on a different computer and still work, as long as the whole folder is moved together. Links break when the parts are separated, for example when the .aprx is sent without its folder, a geodatabase is renamed or data is moved to another drive. The layer keeps its name in the Contents pane but shows a red exclamation mark, which means that Pro still has the layer but cannot find its data. You repair the link on the same Source tab, where **Set Data Source** points the layer to the data's new location.



Figure 2.2 Absolute and relative paths to the same feature class. Because the project stores the relative path, the whole project folder can be moved to another drive or computer and every layer can still find its data. Moving the geodatabase out of the folder on its own breaks the link.

RULE OF THUMB

Keep the project folder together, with its data inside it, and share the whole folder rather than the .aprx file alone.

The vector model

A GIS represents the world in two ways. The **vector** model represents discrete objects, called **features**, as shapes built from coordinates, and stores one row of attributes for each feature. There are three shapes. A **point** is a single coordinate pair, such as a town, a gauging station or a survey site. A **line** is a list of coordinate pairs, the **vertices**, joined by straight segments, such as a river or a road. A **polygon** is a line that closes on itself and encloses an area, such as a river basin, a county or a lake.

In this data, as in most vector data, a line that looks curved is made of straight pieces. The Environmental Protection Agency's river network of Ireland is made of 102,108 line segments built from 4,730,464 vertices. Each segment is straight between its vertices, and the rivers look smooth only because the vertices are close together. If you zoom in to about 1:5,000, the corners become visible. The course's own copy of the network, made so that it draws quickly, keeps the 22,856 segments of the larger rivers and simplifies them to a tolerance of 10 m, which reduces them from

1,117,805 vertices to 257,044. On that copy the corners become visible at a smaller scale. The 40 river basins, which look like simple shapes on a national map, have 433,845 vertices in total. A geodatabase can also store true curves (circular arcs and similar shapes), which are sometimes used in surveyed and engineered data. The course's data has none.

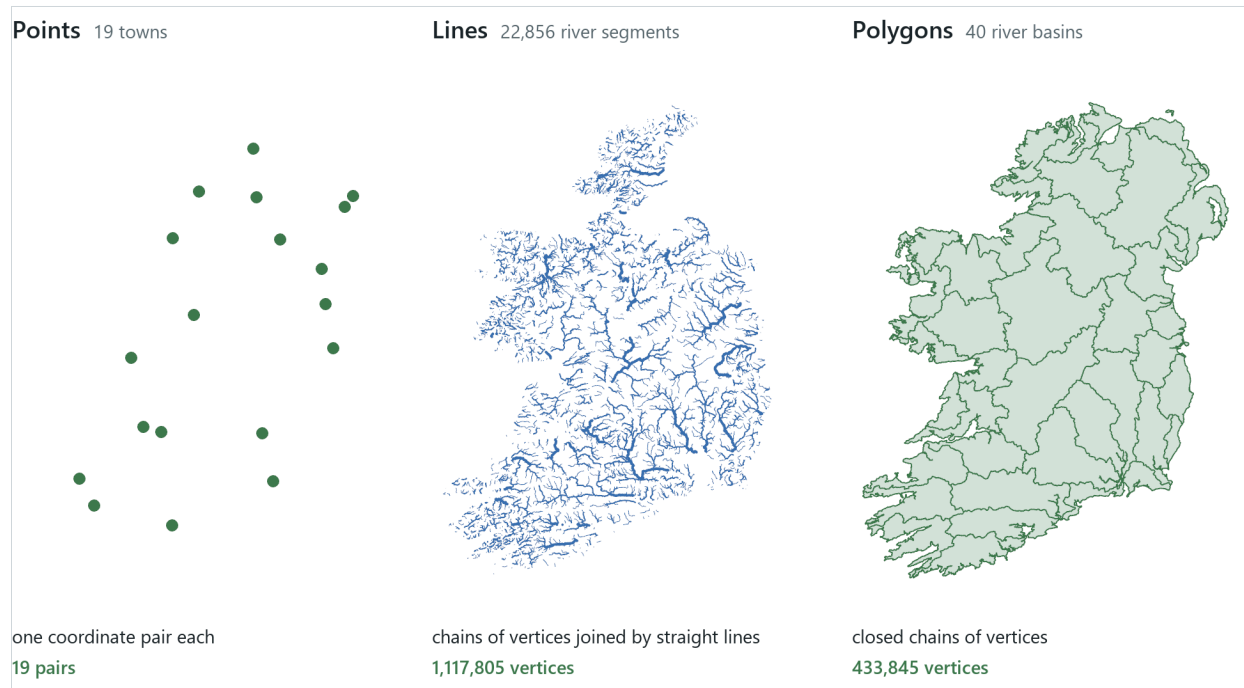


Figure 2.3 The three vector geometries, each shown with one Irish layer. Points: towns, each one coordinate pair. Lines: the river network, drawn here for the larger rivers (Strahler order 3 and up, the same segments as the course's copy, at the original's full detail), with each segment a chain of vertices. Polygons: the 40 hydrometric areas, each a closed chain of vertices. The counts are those of the data drawn.

Towns: Natural Earth (public domain). Rivers: Rivers Network of Ireland (ArcGIS Living Atlas of the World); basins: Hydrometric Areas Ireland & Northern Ireland (ArcGIS Online). Source: Environmental Protection Agency (Ireland), CC BY 4.0

The raster model

The **raster** model covers the ground with a grid of equal squares, the **cells** (or pixels), arranged in rows and columns. Each cell holds one value, such as a class code, an elevation or a temperature. A raster has no features or shapes, only a value at every position. A raster can hold several layers of values over the same grid, called **bands**. A colour photograph has red, green and blue bands, and a land-cover map has one.

The raster in this module is a global land-cover map made from Sentinel-2 satellite images. It has 10 m cells, one band, one map for each year from 2017 to 2025, and nine classes such as Water, Trees, Crops, Built area and Rangeland. Ireland's grass pasture is split between two of these classes. In the 2025 map most of the improved pasture is classed as Crops, and the rougher grazing of the west and the uplands as Rangeland. This is why the map on the opening page is mostly yellow, with tan along the Atlantic coast. It also shows that a class name is chosen by the mapmaker and may not describe the ground accurately. At a small scale the individual cells are too small to see, and the map looks continuous. When you zoom in, the separate square cells become visible.

The **cell size** determines how much detail a raster can show. A raster with 10 m cells cannot show a path 2 m wide or the exact curve of a riverbank, and a single square kilometre contains 10,000 cells of that size. Halving the cell size quadruples the number of cells, so choosing a cell size means balancing detail against the amount of data. Chapter 8 returns to this subject.

A finer grid or denser vertices can hold more detail without being more accurate, because how well a dataset matches the world depends on how it was captured. Resampling a raster to smaller cells does not create new observations, and each 10 m land-cover cell has a single class even where the ground has several types of cover.

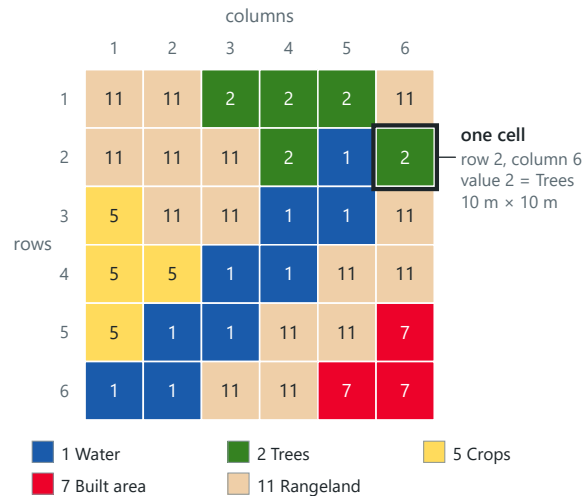


Figure 2.4 A raster grid. Each cell is at a row and a column and holds one number. Here the numbers are land-cover classes, so the value 2 means Trees. The cell size, 10 m by 10 m, is the smallest area the grid can show.

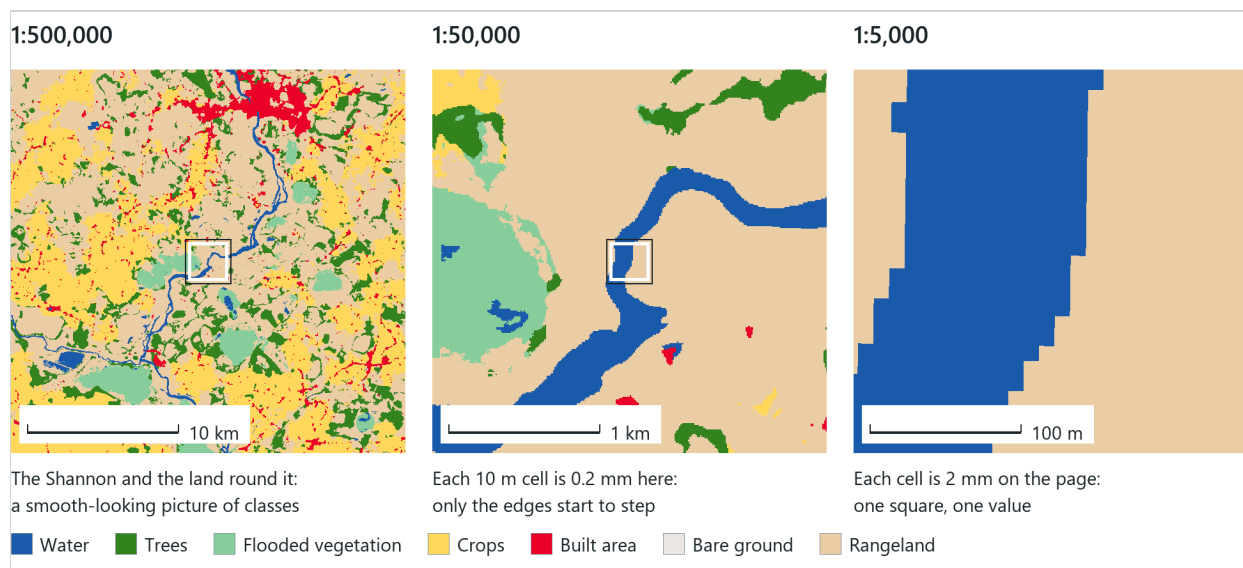


Figure 2.5 The same land cover along the River Shannon at three scales. At 1:500,000 it looks smooth. At 1:50,000 each 10 m cell is 0.2 mm across, and the edges begin to look stepped. At 1:5,000 each cell is 2 mm across and the squares are clearly visible. The boxes show the area of the next frame.

Land cover: Sentinel-2 10m Land Use/Land Cover Time Series, most recent year, Impact Observatory, Microsoft and Esri (ArcGIS Living Atlas of the World; source data CC BY 4.0)

IN THE PANE

In Module 2 lesson 2 you identify the point, line and polygon layers by their symbols in the Contents pane and zoom to a river bend at about 1:5,000 to see its vertices. You then add the Sentinel-2 land cover by its link (Add Data, Data From Path) and zoom in again until the cells are visible. Waldo checks the line layer's geometry and checks that the land cover is a layer you added rather than part of the basemap.

Vector or raster?

The two models are two ways of representing the world, and most themes can be represented in either. The vector model suits objects with boundaries and individual identities, such as a survey site, a river, a basin or a building, each of which you can select, name and count. The raster model suits quantities that vary continuously, such as elevation and rainfall, and categories mapped on a grid, such as the module's land cover. In a raster there is no single feature to select, only a value at every position. A river can be represented as a line for questions about the network or as a band of water cells for questions about area, so the choice of model depends on the question as well as on the theme. Most maps contain both, as the figure on the opening page shows.

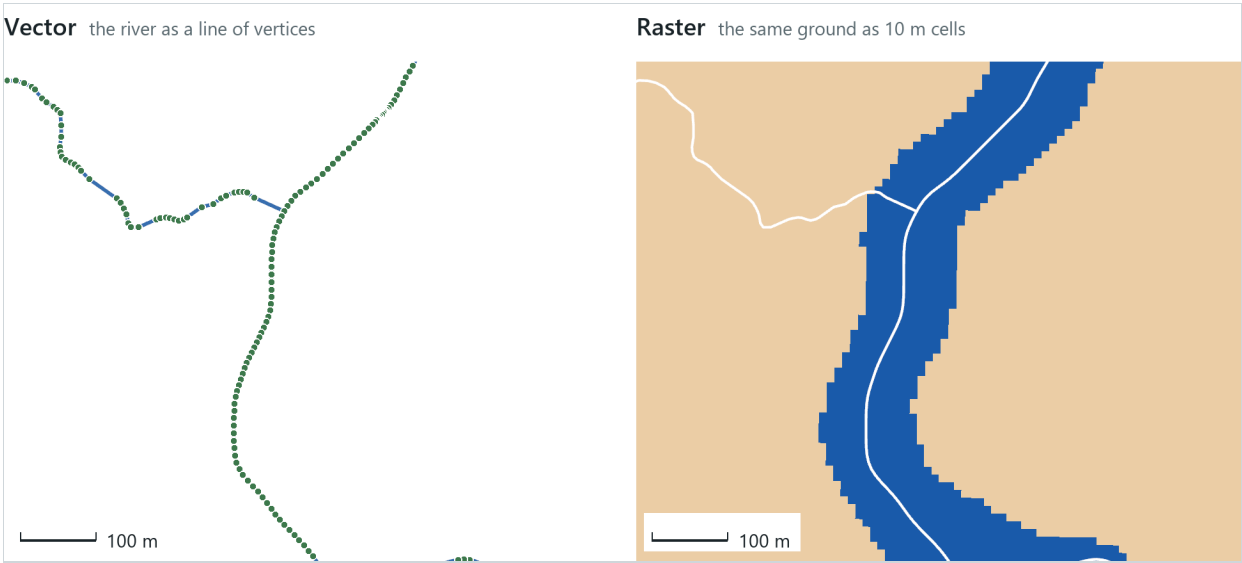


Figure 2.6 One bend of the Shannon at 1:10,000 in both models. The vector river is a line through its vertices, with a tributary joining from the west. The raster river is a band of 10 m Water cells whose banks form steps from square to square. The vector line runs down the middle of it.

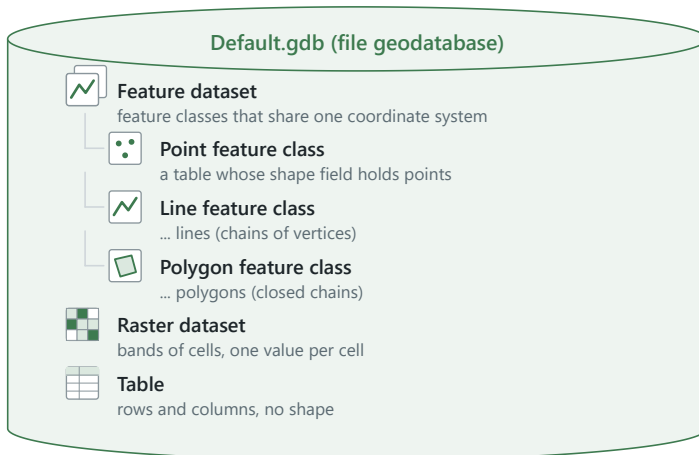
River: Rivers Network of Ireland (ArcGIS Living Atlas of the World), Environmental Protection Agency (Ireland), CC BY 4.0. Land cover: Sentinel-2 10m Land Use/Land Cover Time Series, Impact Observatory, Microsoft and Esri (ArcGIS Living Atlas of the World; source data CC BY 4.0)

	Vector	Raster
Represents	features built from coordinates: discrete things with boundaries and identities	a value at every grid position: categories (land cover) or continuous quantities (elevation)
Building block	points, lines, polygons from vertices	cells in rows and columns
Attributes	a row per feature, many fields	one value per cell in each band
Detail set by	the spacing of the vertices	the cell size
Curves and edges	followed closely	become steps
Examples in the course	basins, rivers, survey points	land cover, elevation

Containers: the geodatabase and the shapefile

Data is stored in files, and two formats are used most often. The **file geodatabase** is Pro's own format and the format the course uses to store its data. It holds several kinds of item: **feature classes**, each a table whose shape field stores point, line or polygon geometry; **feature datasets**, groups of feature classes that share one coordinate system; **raster datasets**; and plain **tables**, rows and columns with no shape at all.

In Catalog: one container, many kinds of item



In Windows Explorer: a folder

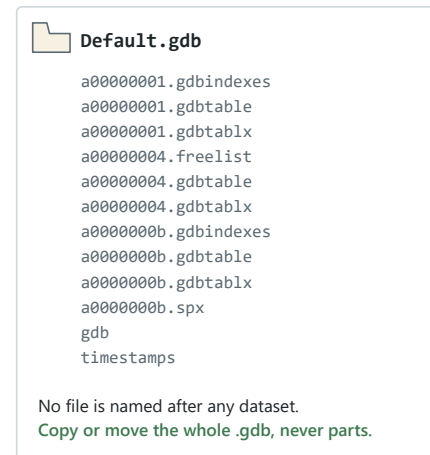


Figure 2.7 A file geodatabase in the Catalog pane and in Windows Explorer. In the Catalog pane it is one container holding feature datasets, feature classes, raster datasets and tables, each with its own icon. In Windows Explorer it is a folder whose name ends in .gdb and which contains internal files, none of them named after a dataset.

The internal files of a geodatabase are managed by the software and are not designed to be read directly. Inside Pro the geodatabase always appears as one item, and you should rename, copy or delete its contents there. From Windows Explorer the only safe operation is to move or copy the whole .gdb folder.

The **shapefile** is an older vector format from the 1990s. Nearly every GIS can read it, so data is still often exchanged in this format. A shapefile is made up of several files with the same name and different extensions. Three are required: the .shp holds the shapes, the .shx an index of them and the .dbf the attribute table. Other files are common: the .prj records the coordinate system, the .cpg the text encoding, the .sbn and .sbx hold a spatial index and the .shp.xml the metadata. A shapefile exported from Pro in the course has eight files.

In a folder: eight files

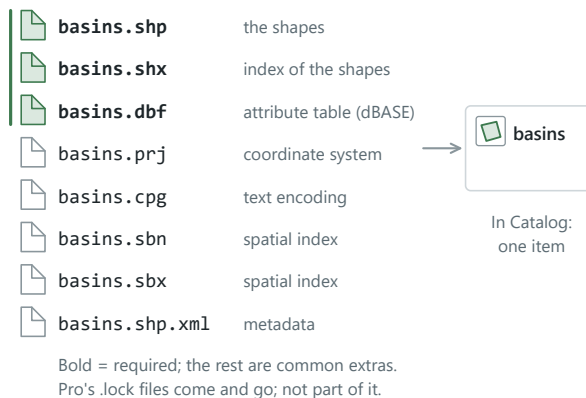


Figure 2.8 A shapefile made of eight files. The .shp, .shx and .dbf are required, and the others are common additional files. In the Catalog pane the set of files appears as a single item.

The files of a shapefile must be kept together. If you send only the .shp, the attributes are lost, and without the .prj the coordinate system of the shapes is unknown. Shapefiles also shorten field names to ten characters, which can rename an attribute without warning when data is saved to a shapefile. The geodatabase was designed to avoid both problems.

Two other exchange formats are also common. **KML**, first made for Google Earth, is an XML format for geographic display that is now a standard of the Open Geospatial Consortium. **GeoJSON** is a plain-text format for features built on JSON, a standard of the Internet Engineering Task Force since 2016, with coordinates always in WGS 84 longitude and latitude. Pro reads both. Chapter 5 covers tables in CSV files and spreadsheets.

RULE OF THUMB

Keep your own data in a file geodatabase. When a shapefile arrives, move or copy every file that shares its name, never the .shp alone.

Layers, datasets and other names

The terms for the parts of a GIS project are easily confused, and the most important distinction is between a layer and its data. A **dataset** (a feature class, a shapefile, a raster, a table) is the stored data. A **layer** is a map's reference to a dataset together with its name in the Contents pane, its colours and its other drawing settings. Two maps can contain layers of the same dataset with different names and colours. Renaming a layer or changing its symbology does not change the data.

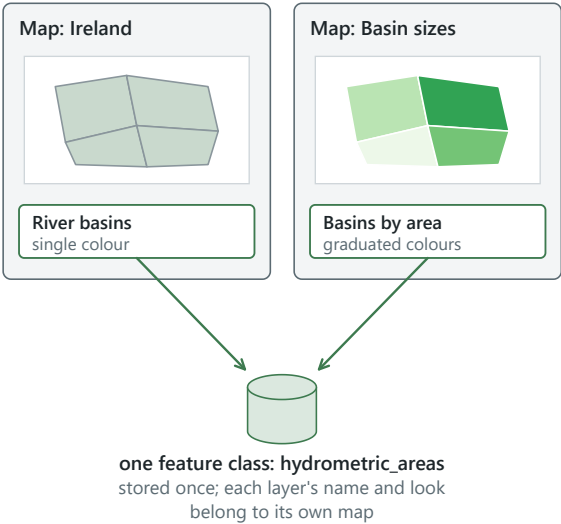


Figure 2.9 One dataset shown as two layers. The same feature class appears in two maps under different names and with different symbology. The data is stored once, and each map has its own name and symbology for the layer.

Name	What it is	Where it is stored
Project	the .aprx and its folder	a folder on disk
Map	layers drawn and symbolised together	inside the project

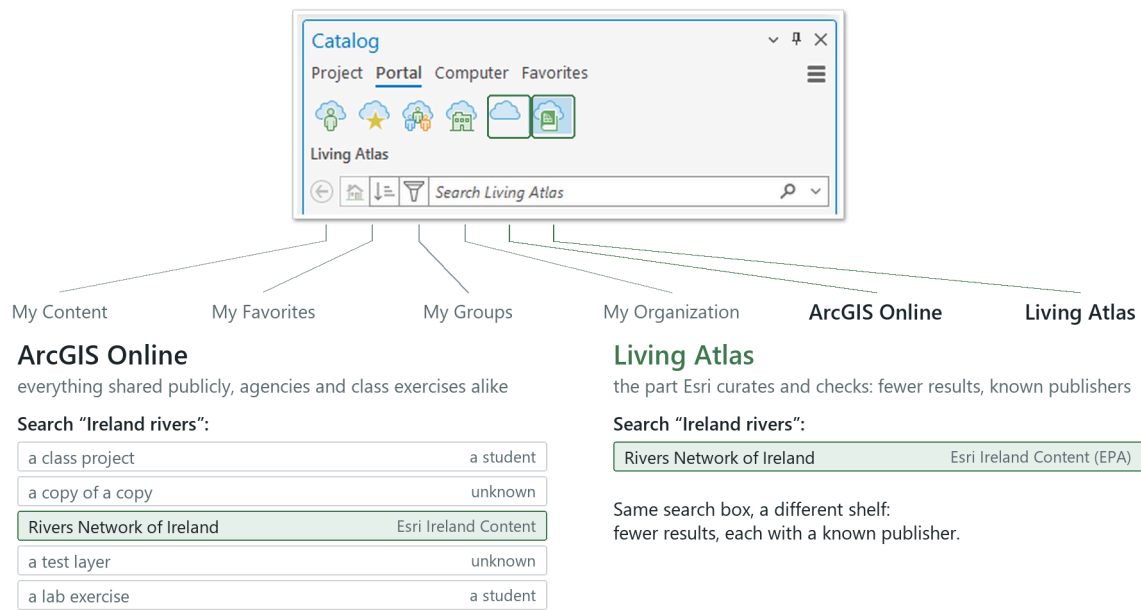
Name	What it is	Where it is stored
Layout	a page for a finished map, with title, legend and scale bar	inside the project
Layer	a map's pointer to a dataset, plus how to draw it	inside a map
Feature class, shapefile, raster dataset, table	the data	a geodatabase, a folder, or a server
Web layer	a layer whose data is a service on a server	on the server

IN THE PANE

In Module 2 lesson 3 you open the project geodatabase in Catalog and read its icons, export a standalone table and a shapefile, and look at the geodatabase and the shapefile in Windows Explorer. Waldo checks that the table and the shapefile exist, and then checks your count of the shapefile's files against the folder, leaving out the temporary lock files that Pro adds while the layer is open.

Where data comes from

The person making a map rarely collects the data on it, and most GIS work uses data made by someone else. The skill required is finding that data and judging it. In Pro you start the search on the **Portal** tab of the Catalog pane, which offers two similar-looking places to search. **ArcGIS Online** is Esri's cloud platform, where organisations and individuals share maps, apps and data. A search of ArcGIS Online covers everything shared publicly, from the data of national agencies to class exercises. The **Living Atlas** is the part of it that Esri curates and checks. Its content comes in three tiers: public, open to anyone without signing in; subscriber, which needs an organisational account; and premium, which also uses credits.



A third route: download portals

data.gov.ie, EPA GeoPortal, Tailte Éireann → ZIP file → unzip to a folder → folder connection → add to the map

Figure 2.10 The Portal tab of the Catalog pane, with its six sources labelled; the two outlined are the doors to shared data. Below it, what one search for Ireland rivers returns through each: in ArcGIS Online a mix of publishers, some unknown; in the Living Atlas fewer items, each with a known publisher. Much authoritative data is published instead on download portals, as ZIP files that reach the map through a folder connection.

What you can do with an online item depends on its type.

Layer type	What it is	What you can do in Pro
Feature layer	features and attributes served live	display, select, query; copy to your own disk when the publisher allows it
Tile layer	a pre-drawn picture of a map	display as a backdrop (basemaps are tile layers)
Map image layer	a picture drawn by the server on request	display; its sublayers can sometimes be queried like features
Imagery layer	a raster served live	display and query cell values
Web map	a whole map someone built	opens as a new map, not as a layer
App	a web application	cannot be added to Pro

The publisher decides whether a layer can be copied. The course's basins service allows queries but not extraction, so Pro's right-click export option is unavailable. The Export Features tool needs only queries, however, so it can still copy the layer.

A great deal of authoritative data is not published on ArcGIS Online. Government portals provide it as downloads, usually a ZIP file containing a shapefile or a geodatabase. For Ireland these include data.gov.ie, Tailte Éireann's open data, the EPA's GeoPortal (the source of the course's basins and rivers) and OpenDataNI. Geofabrik offers OpenStreetMap extracts, and Natural Earth offers free world-scale data. In the United States, the main sources include data.gov, the Census Bureau's

TIGER files, the USGS National Map and the state clearinghouses. To add a download to a map, you unzip it into a folder that Pro can access through a **folder connection**.

 IN THE PANE

In Module 2 lesson 4 you run the same search in ArcGIS Online and in the Living Atlas, and read the owner of the Rivers Network of Ireland item from the card that appears over a search result. The lesson then explains the kinds of web layer and the Irish download portals, and you compare the land cover's full Metadata tab with that of the course's own copy of the river surveys. Waldo checks the owner you read against the item itself.

Metadata

Before trusting a dataset, ask who made it, what it represents, when it was made or last updated, where it covers, how it was made and why. The answers are **metadata**, data about the data, and they help you decide whether a dataset is fit for your purpose. They are recorded on an item's page online and on the **Metadata** tab of a layer's properties in Pro.

Metadata Facts	
Rivers Network of Ireland	
ArcGIS Online item · feature layer	
Who	Environmental Protection Agency (source); shared by Esri Ireland Content
What	the river and stream network of Ireland: 102,108 line segments, Strahler order
When	first shared 1 May 2024; last updated 11 April 2025
Where	the Republic of Ireland: 10.47° to 6.01° W, 51.45° to 55.38° N
How	digitised from the Ordnance Survey 1:50,000 maps, updated from local surveys; copied from a shapefile to a geodatabase and re-projected to Web Mercator
Why	the EPA's river network, used to define the Water Framework Directive's river water bodies
Licence	Creative Commons CC BY 4.0
From the item page and the service, read September 2026. Where a field is empty, the question is yours.	

Figure 2.11 The metadata of Rivers Network of Ireland, laid out like a nutrition label: the answers to who, what, when, where, how and why, taken from its own item page and service.

Facts: the Rivers Network of Ireland item on ArcGIS Online (Esri Ireland Content; source Environmental Protection Agency, CC BY 4.0)

Some publishers document their data fully. The land-cover layer's metadata names its makers, classes and method, and the river network's item page, shown above, records where its lines were first digitised. Other publishers provide no documentation, and the Metadata tab is blank. A blank

tab does not mean that the data is poor, but you have to answer the questions yourself from the source. Read the dates carefully. The date that the data describes, when the ground was observed, surveyed or digitised, is different from the date on which an item was published or last updated. The river network's item gives its sharing and update dates and states that its lines were digitised from source maps, but it does not say when those maps were surveyed. That date is therefore unknown and should not be assumed to be recent, because an item updated last month can describe old conditions. The metadata of a copy may only refer to its source. The course's copy of the river surveys describes the copy but says little about the surveys themselves, and its description refers you to The Rivers Trust's live dataset for the rest. Metadata is also carried over to copies, so data exported from a service keeps the service's title and summary.

RULE OF THUMB

Ask who, what, when, where, how and why of every dataset before you use it, and prefer data whose publisher answers them.

TRY IT

With a course project open, ask Waldo in Ask mode: for each layer on my map, where does its data live, and is it a feature layer, an imagery layer or a local copy? Waldo reads every layer's data source and type. Compare its answer with the Source tab of one layer.

Key terms

project (.aprx)	the Pro file that holds maps, layouts and connections and a pointer to each layer's data, but not the data itself.
data source	the path or web address a layer uses to find its data, shown on the layer's Source tab.
absolute path	a path that starts at the drive letter and names every folder down to the data.
relative path	a path written from the folder that holds the project file, so the project folder can move as a whole.
broken link	a layer whose data can no longer be found at its stored path, marked with a red exclamation mark and repaired from the Source tab.
vector	a data model that represents features as points, lines and polygons built from coordinates, with one row of attributes per feature.
vertex	one stored coordinate pair along a line or a polygon's boundary; straight segments join the vertices.
raster	a data model that covers the ground with a grid of equal cells in rows and columns, each holding one value.
cell (pixel)	one square of a raster grid; its size sets the smallest thing the raster can show.
band	one layer of values in a raster; a land-cover map has one, a colour image has several.
layer	a map's reference to a dataset together with its name, symbology and other drawing settings.
file geodatabase	Pro's container for data, stored as a folder ending in .gdb and managed through Pro.

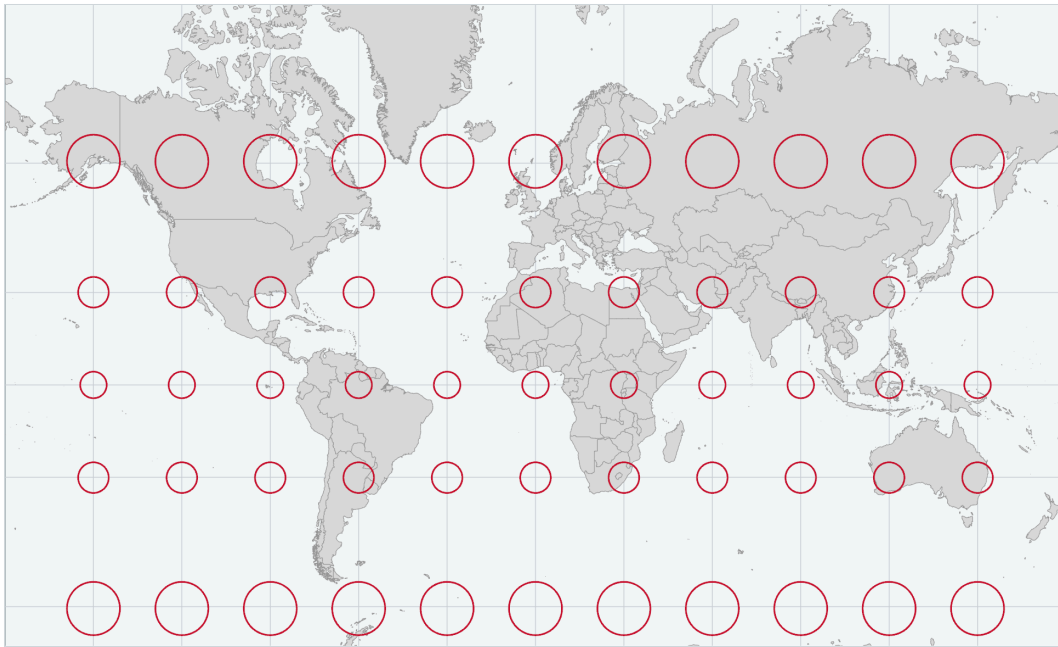
feature class	a table in a geodatabase whose shape field holds point, line or polygon geometry.
feature dataset	a group of feature classes in a geodatabase that share one coordinate system.
table (standalone table)	rows and columns with no geometry.
shapefile	an older vector format stored as a family of same-named files (.shp, .shx, .dbf and optional others) that must be kept together.
KML	an XML format for geographic display that began with Google Earth and is now an OGC standard.
GeoJSON	a plain-text format for features built on JSON, an IETF standard since 2016.
ArcGIS Online	Esri's cloud platform where maps, apps and data are shared, reached in Pro through the Portal tab.
Living Atlas	Esri's curated collection of maps and data, in public, subscriber and premium tiers.
feature layer	a web layer that serves features and attributes that can be queried and, when the publisher allows it, copied.
tile layer / map image layer	a web layer served as a picture of a map, pre-drawn in tiles or drawn by the server on request.
imagery layer	a raster served live from a server.
folder connection	a shortcut in the Catalog pane to a folder on disk, the way downloaded data reaches a map.
metadata	data about data: who made it, what it is, when, where, how and why.

Further reading

- Olaya, V. *Introduction to GIS*, the chapters on data and on data sources. <https://volaya.github.io/gis-book/en/>
- Butler, H., et al. (2016). *The GeoJSON Format*. RFC 7946, Internet Engineering Task Force. <https://www.rfc-editor.org/rfc/rfc7946>
- Open Geospatial Consortium. *KML standard*. <https://www.ogc.org/standards/kml/>
- Harder, C., & Brown, C. (eds.) (2017). *The ArcGIS Book* (2nd ed.), on the Living Atlas and web layers. Esri Press. <https://learn.arcgis.com/en/arcgis-book/>
- Esri. *ArcGIS Pro documentation*, on geodatabases, shapefiles and data sources. <https://pro.arcgis.com/>
- The course's data sources: every layer the lessons and this chapter's maps use, with its publisher and licence.

Spatial reference

Accompanies Module 3 · Spatial Reference Systems



Countries: World Countries (Generalized), ArcGIS Living Atlas of the World. Sources: Esri; Garmin International, Inc.; U.S. Central Intelligence Agency (The World Factbook); National Geographic Society

WHY THIS MATTERS

Every layer in a GIS stores locations as numbers, and those numbers identify a place only through the model of the Earth they refer to. If the model changes, the same numbers refer to a different place. Any flat map of the round Earth stretches some part of it: on the Mercator map above, every red circle is the same size on the ground. This chapter explains the models, which properties each one preserves, and how to choose a suitable one before you measure anything.

YOU SHOULD BE ABLE TO EXPLAIN

- why the same latitude and longitude name different places on different datums
- why no flat map keeps shape, area, distance and direction at once, and which a thematic map should keep
- why Dublin to Galway measures 311 km planar in Web Mercator but 187 km in Irish Transverse Mercator

IN THIS CHAPTER

The shape of the Earth

Datums: fitting the ellipsoid

Latitude and longitude

Map scale

Projection: from round to flat

What a projection keeps

Cylinder, cone and plane

The famous projections

Choosing a coordinate system

Web Mercator: fine for looking, wrong for measuring

Map units, on-the-fly projection and transformations

Why it matters

The shape of the Earth

Describing where something is requires a model of the Earth's shape. For a desk globe, a sphere with a radius of about 6,371 km is the right model. The Earth's rotation, however, makes it bulge at the equator and flatten at the poles, so its shape is an **ellipsoid** (also called a spheroid), the shape produced by rotating an ellipse about its shorter axis.

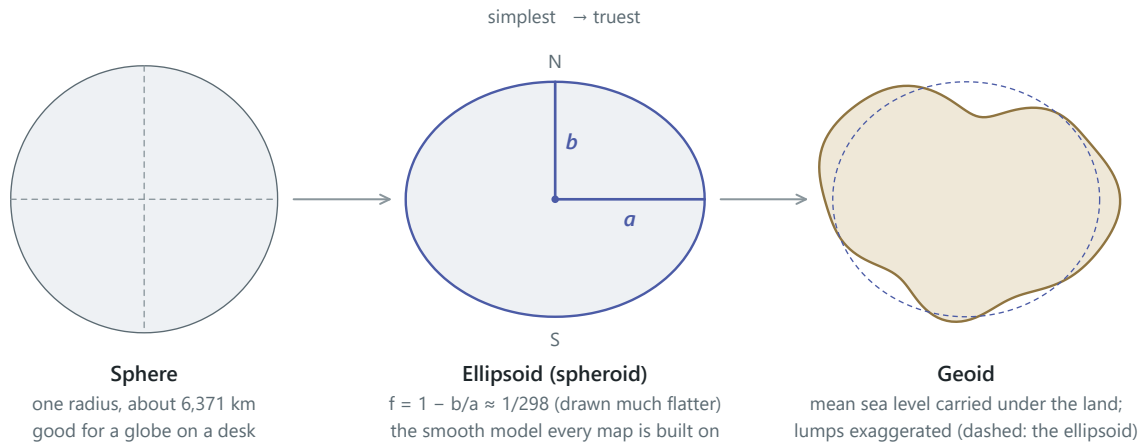


Figure 3.1 Three models of the Earth's shape: a sphere with one radius, an ellipsoid with two axes *a* and *b*, and the irregular geoid. The flattening and the irregularities are drawn far larger than they are.

Two numbers define an ellipsoid: the equatorial radius **a** and the polar radius **b**. The degree to which it is squashed is the **flattening**, $f = 1 - b/a$. For the Earth *f* is about one part in three hundred, so GIS dialogs show its inverse, $1/f$, which is easier to read. The WGS 84 ellipsoid, on which GPS reports positions, has $a = 6,378,137$ m exactly and $1/f = 298.257223563$, which puts the poles about 21.4 km closer to the centre than the equator. On a globe 20 cm across that difference would be two-thirds of a millimetre, so globes are made as spheres; on maps the difference becomes significant at scales of about 1:100,000 and larger.

Surveyors have fitted many ellipsoids to the Earth over the past two centuries. The following are among those you will see on Pro's Source tab:

Ellipsoid	<i>a</i> (m)	1/ <i>f</i>	Used by
WGS 84	6,378,137	298.257223563	GPS, the WGS 1984 datum
GRS 80	6,378,137	298.257222101	NAD 1983, ETRS89, Ireland's IRENET95
Clarke 1866	6,378,206.4	294.98	NAD 1927 (North America)
Airy Modified 1849	6,377,340.189	299.3249646	Ireland 1965 (the old Irish Grid)

WGS 84 and GRS 80 differ by about a tenth of a millimetre, so for mapping they can be treated as the same shape.

The ellipsoid is a smooth surface, but the real Earth is irregular. Mass is spread unevenly through the planet, so gravity pulls slightly harder in some places than in others. The **geoid** is the surface the oceans would take under the influence of gravity alone, extended under the continents; it corresponds to mean sea level everywhere. It rises and falls relative to a well-fitting ellipsoid by no

more than about a hundred metres, but it is too irregular to be used for computing map coordinates. Each of the three shapes is therefore used for a different purpose: the sphere for globes and small world maps, the ellipsoid for latitude, longitude and every projection, and the geoid for heights, which are measured from sea level.

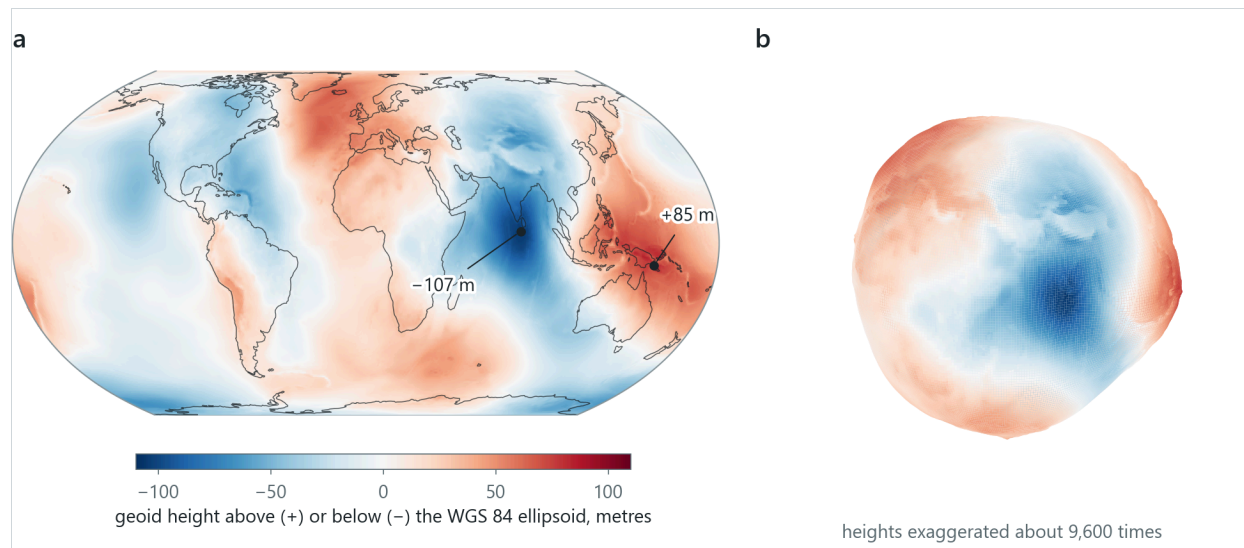


Figure 3.2 The geoid against the WGS 84 ellipsoid (EGM96). (a) Its height on an Equal Earth map, from about -107 m in the Indian Ocean south of India to about $+85$ m near New Guinea. (b) The same heights exaggerated about 9,600 times on a globe, a figure often called the "potato" Earth.

Geoid heights: EGM96, U.S. National Geospatial-Intelligence Agency (public domain), 15' grid via PROJ-data. Coastlines: Natural Earth (public domain). Colours: ColorBrewer RdBu, Cynthia Brewer (Apache 2.0)

Datums: fitting the ellipsoid

An ellipsoid defines a shape only. To use it as a reference for coordinates, you must also specify its position: where its centre is and how it is oriented. That specification is a **datum**, and a latitude and longitude identify one definite place only when the datum is known.

Before satellites, a country fixed its datum by surveying: the ellipsoid was fixed to one carefully measured point on the ground and every other survey point was computed from it. The North American Datum of 1927 (NAD 1927) was fixed at Meades Ranch, Kansas, on the Clarke 1866 ellipsoid; Ireland 1965 used Airy Modified 1849. Such **local datums** fit the geoid closely across their own region and depart from it elsewhere.

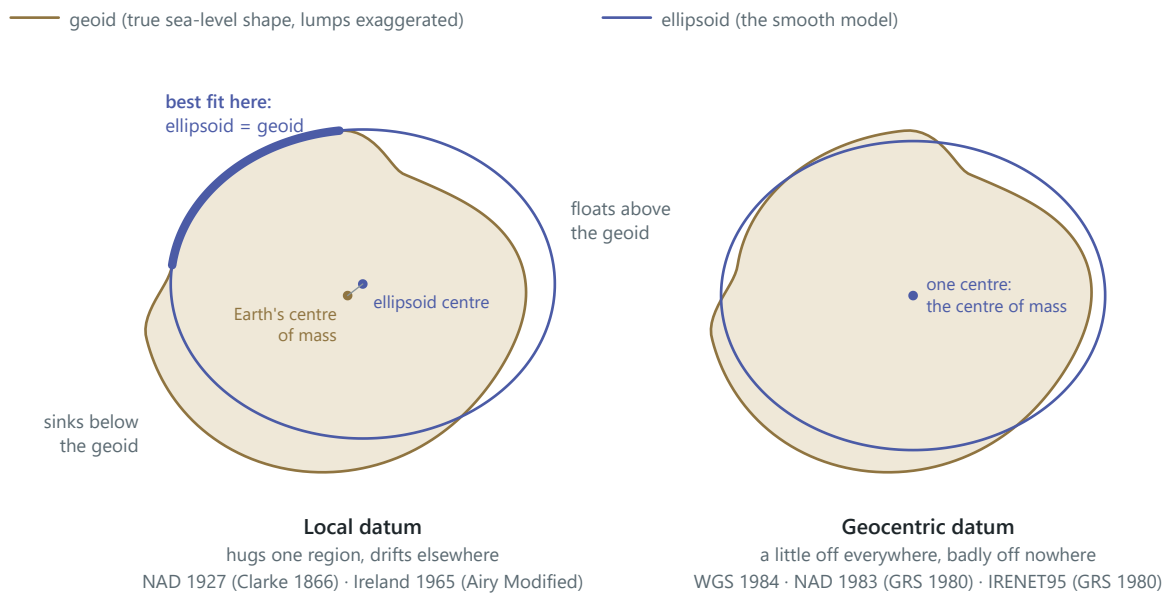


Figure 3.3 Local versus geocentric datums, fitted to the same lumpy geoid. A local datum matches the geoid along one region and lies above or below it elsewhere; a geocentric datum is centred on the Earth's centre of mass and departs slightly from the geoid everywhere, but by a large amount nowhere.

Satellite geodesy made **geocentric datums** possible. In a geocentric datum the ellipsoid's centre is placed at the Earth's centre of mass, and the ellipsoid fits the whole planet well without being the best fit for any one region. WGS 1984, NAD 1983 (on GRS 80) and IREN95, Ireland's GPS-observed realisation of the European system ETRS89, are all geocentric. Datums are also revised from time to time as measurements improve.

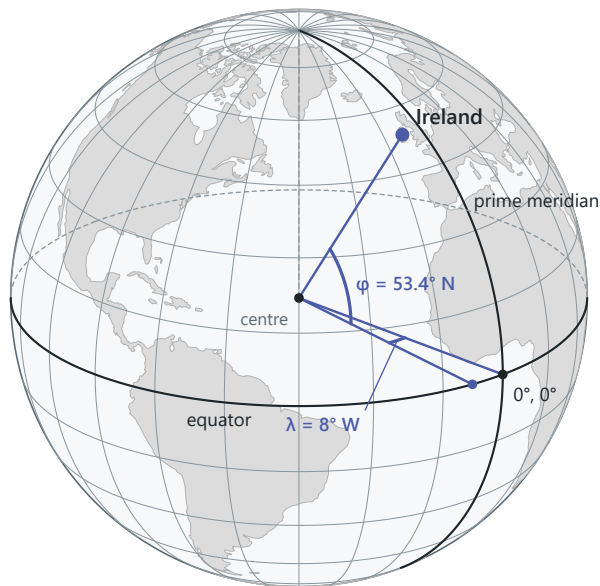
Because the fit differs, the same physical point has different latitude and longitude on different datums. Moving from NAD 1927 to NAD 1983 shifts coordinates by amounts that reach 300 m; in Ireland the same spot on Ireland 1965 and on WGS 1984 differs by about 50 to 70 m. A coordinate recorded without its datum is therefore incomplete.

IN THE PANE

Module 3 lesson 1 opens with a quiz on the Earth's shape, then covers the sphere, ellipsoid and geoid and the difference between local and geocentric datums. You read these properties on a real layer: its **Source** tab lists the Datum, the Spheroid, the two semi-axes and the Inverse Flattening, 298.257223563 for WGS 1984, and Waldo checks what you read against the layer.

Latitude and longitude

On the ellipsoid, a position is two angles. **Latitude** is the angle north or south of the equator, 0° to 90° at each pole. **Longitude** is the angle east or west of a prime meridian, 0° to 180° each way; longitude has no natural zero, and in 1884 an international conference chose the meridian through Greenwich. Lines of equal latitude are **parallels**, lines of equal longitude **meridians**, and the network of both is the **graticule**.



Angles, not distances

Latitude (φ)
the angle up or down from the equator's plane,
0° at the equator to 90° at each pole.

Longitude (λ)
the angle east or west of the prime meridian
(Greenwich), 0° to 180°.

Graticule
every parallel and meridian drawn on the surface;
here every 15°.

One degree of ...

latitude, anywhere	≈ 111 km
longitude at the equator	≈ 111 km
longitude at 53.4° N (Ireland)	≈ 67 km
longitude at the poles	0 km

Figure 3.4 Latitude and longitude as angles at the centre of the Earth, with Ireland at about 53.4° N, 8° W. A degree of latitude is about 111 km everywhere; a degree of longitude shrinks from about 111 km at the equator to about 67 km at Ireland's latitude and to nothing at the poles. Drawn on a sphere; on an ellipsoid, geodetic latitude is measured from the line perpendicular to the surface, as the text below explains.
Land: Natural Earth (public domain)

Dublin is at 53.3498° N, 6.2603° W in decimal degrees, or 53°20′59″ N, 6°15′37″ W in degrees, minutes and seconds (60 minutes to a degree, 60 seconds to a minute). Without letters, south and west are negative (−6.2603). Pro's coordinate readout prints longitude first, each value with its hemisphere letter.

Degrees measure angles rather than lengths. Because the meridians converge towards the poles, a degree of longitude shrinks with the cosine of the latitude: 111 km at the equator, 66.7 km at Dublin and 55.5 km at 60°, half the equatorial value. You can locate anything in latitude and longitude, but you cannot measure distance or area on them with flat geometry. (Strictly, on an ellipsoid the stored latitude is measured from the line perpendicular to the surface, not from the centre as on a sphere; the two differ by at most about a fifth of a degree.)

A **geographic coordinate system (GCS)** bundles the pieces: a datum (and so an ellipsoid), a prime meridian and an angular unit, degrees. The term "datum" is often used for the whole GCS, but the datum is only one of its parts.

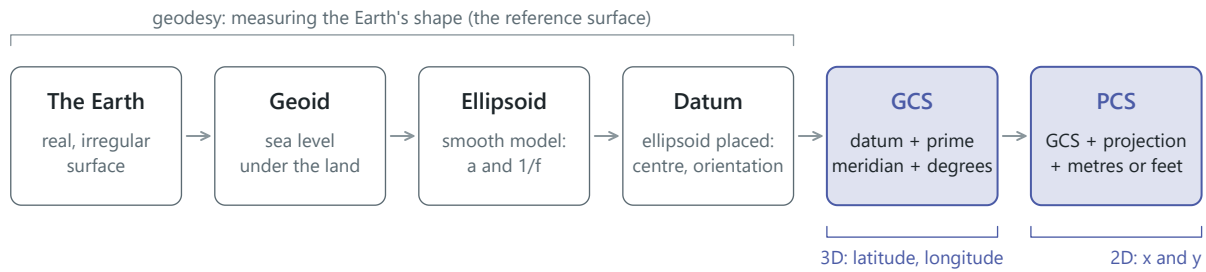


Figure 3.5 From the real Earth to map coordinates. Geodesy models the surface (geoid, ellipsoid, datum); a geographic coordinate system gives positions in degrees on it; a projected coordinate system adds a projection and metres or feet.

IN THE PANE

Module 3 lesson 2 has you watch the coordinate readout, switch it from decimal degrees to degrees-minutes-seconds, and add the **Firefly Grid Lines** graticule from the Living Atlas. You read Dublin's position, then open **Map Properties > Coordinate Systems** and read the **Horizontal** line under *Current Map Coordinate Systems*. This is the map's own coordinate system, which can differ from that of every layer.

Map scale

Every map shows the ground at a reduced size. **Map scale** is the ratio between a distance on the map and the same distance on the ground, and it can be written in three ways. A **representative fraction** (RF) such as 1:50,000 means that one unit on the map represents 50,000 of the same unit on the ground, whatever the unit. A **verbal scale** states the ratio in words ("one centimetre to five hundred metres"). A **bar scale** shows it as a graphic, and it is the only form that remains correct when a map is enlarged or reduced.

One scale, three ways to write it

Representative fraction

1:24,000

1 unit on the map = 24,000 of the same unit on the ground

Verbal scale

"1 cm to 240 m"

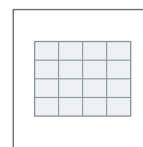
the same ratio in words, often in mixed units

Bar (graphic) scale

0 500 1,000 m

true at the printed size of this page (1 km at 1:24,000 = 41.7 mm);
the only one of the three that stays true when a map is enlarged or shrunk

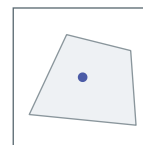
Large scale or small scale?



Large scale

a large fraction (1/10,000)
zoomed in: small area,
much detail

1:10,000



Small scale

a small fraction (1/1,000,000)
zoomed out: large area,
little detail

1:1,000,000



A globe 1 m around models an Earth about 40,000 km around:
1 m : 40,000,000 m, a scale of 1:40,000,000 — every map is a reduction,
and its scale says by how much.

Figure 3.6 One scale written in three ways, and the meaning of "large" and "small" scale. A globe 1 m around is a model of the Earth at about 1:40,000,000.

The terms can seem reversed until you recall that a scale is a fraction: 1:10,000 is a larger fraction than 1:1,000,000, so a **large-scale** map shows a small area in much detail and a **small-scale** map shows a large area with little. The equator is 40,075 km round, so a globe one metre round is a map at about 1:40,000,000.

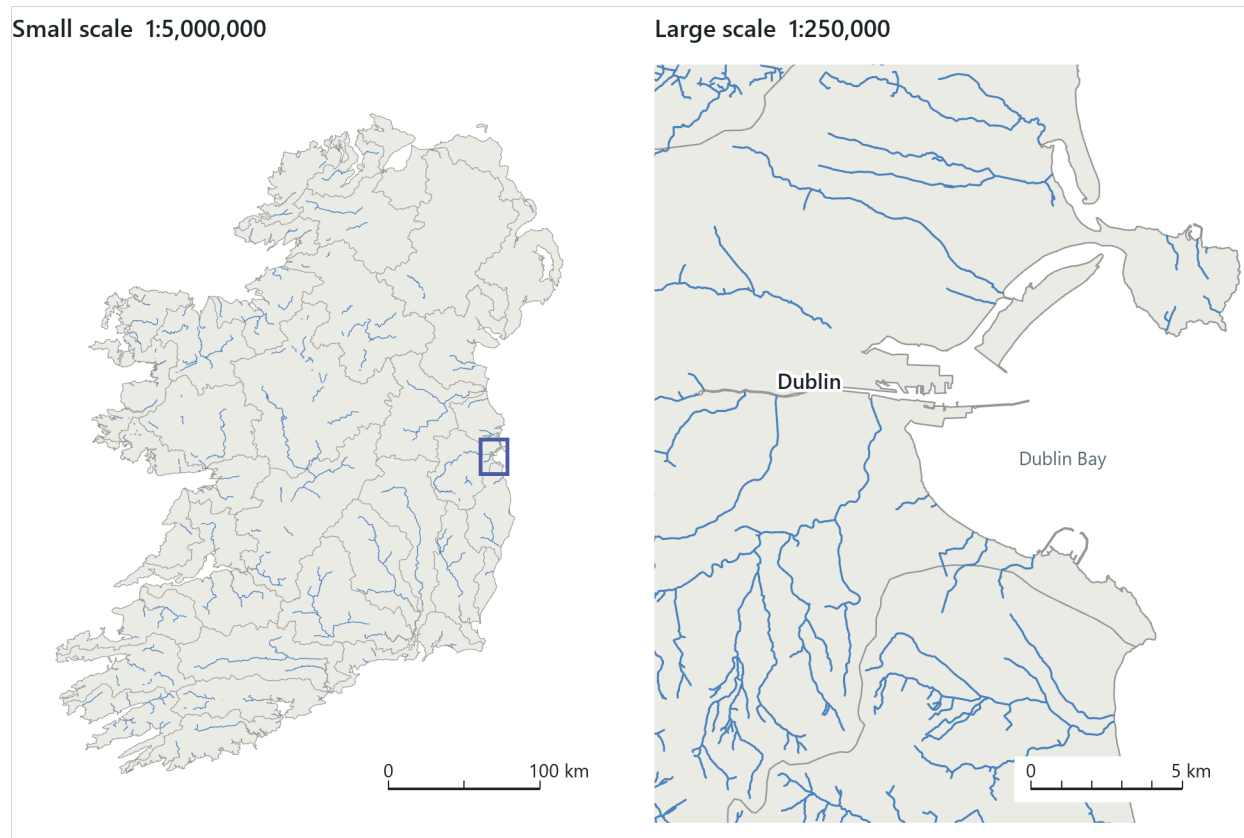


Figure 3.7 The same Irish river basins and rivers at two scales. At 1:5,000,000 all of Ireland fits and only the larger rivers can be shown; at 1:250,000 the box around Dublin fills the frame and every stream appears. Scale bars are true at the printed size.

Basins: Hydrometric Areas Ireland & Northern Ireland; rivers: Rivers Network of Ireland (ArcGIS Living Atlas of the World). Source: Environmental Protection Agency (Ireland)

Scale limits the detail a map can show. The eye can separate marks about 0.2 mm apart, so at 1:50,000 a mark 0.2 mm wide represents 10 m of ground. Anything narrower, such as a path or a stream, can be shown only with an enlarged or simplified symbol, and the drawn width of a river is not its measured width. Zooming in a GIS never adds detail beyond the detail with which the data were captured. Pro shows the current scale as an RF at the bottom of the map view; typing a value sets it.

On a flat map the stated scale is exactly true only along certain lines or at certain points. The ratio of the actual scale at a place to the stated scale is the **scale factor**. The rest of this chapter is about where, and by how much, the scale factor departs from 1.

Projection: from round to flat

A **map projection** is a systematic, mathematical way of representing the curved Earth on a plane: a formula that takes each latitude and longitude to an x and a y. The result is a **projected**

coordinate system (PCS), which combines a GCS, a projection with its parameters and a linear unit, usually metres or feet.

A curved surface cannot be flattened without stretching; pressing a piece of orange peel flat shows the same effect. Every projection distorts at least one of four properties: **shape** (the angles that make small shapes look right), **area**, **distance** and **direction**. No flat map keeps all four. Each projection is designed to keep some of them in certain places, at the cost of greater distortion elsewhere. Over a town the distortion can be too small to measure, but over a continent or the world it is clearly visible.

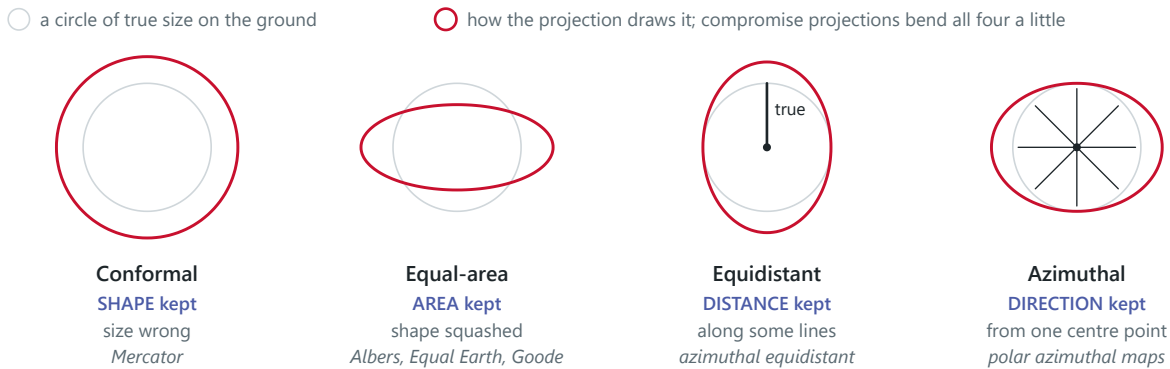


Figure 3.8 The four properties a projection can keep. A conformal map keeps shape but not size, an equal-area map keeps area but not shape, an equidistant map keeps distance along some lines, and an azimuthal map keeps direction from one centre point; compromise projections distort all four slightly.

The standard way to show distortion is **Tissot's indicatrix**: an infinitely small circle on the Earth projects as a small ellipse whose shape and size show how the map stretches at that location. The course applies the same idea with circles large enough to see: 55 circles, each 500 km in radius on the ground, placed every 30° of latitude and longitude. Because every circle has the same true size, any change in their size or shape on a map is caused by the projection. On a conformal map they stay round but change size; on an equal-area map they keep their area but are squashed into ovals.

What a projection keeps

Projections are grouped by the property they preserve.

- **Conformal** projections keep local angles, so small shapes are right; areas are enlarged or reduced. Nearly all large-scale topographic mapping is conformal.
- **Equal-area** projections keep relative sizes: a coin placed anywhere covers the same true area. Shapes are distorted over most of the map.
- **Equidistant** projections keep true scale from one or two points, or along the meridians; no map keeps every distance true.
- **Azimuthal** projections keep true directions from one centre point. Azimuthal combines with the others: an azimuthal map can also be conformal, equal-area or equidistant.
- **Compromise** projections keep nothing exactly but balance the errors so none is large.
- **Interrupted** projections cut the globe, usually through the oceans, into lobes with their own central meridians, so that the shapes of the land areas are less distorted.

Only one pair of properties is mutually exclusive: no map can be both conformal and equal-area. The course's azimuthal example is also equidistant, and its interrupted one is also equal-area.

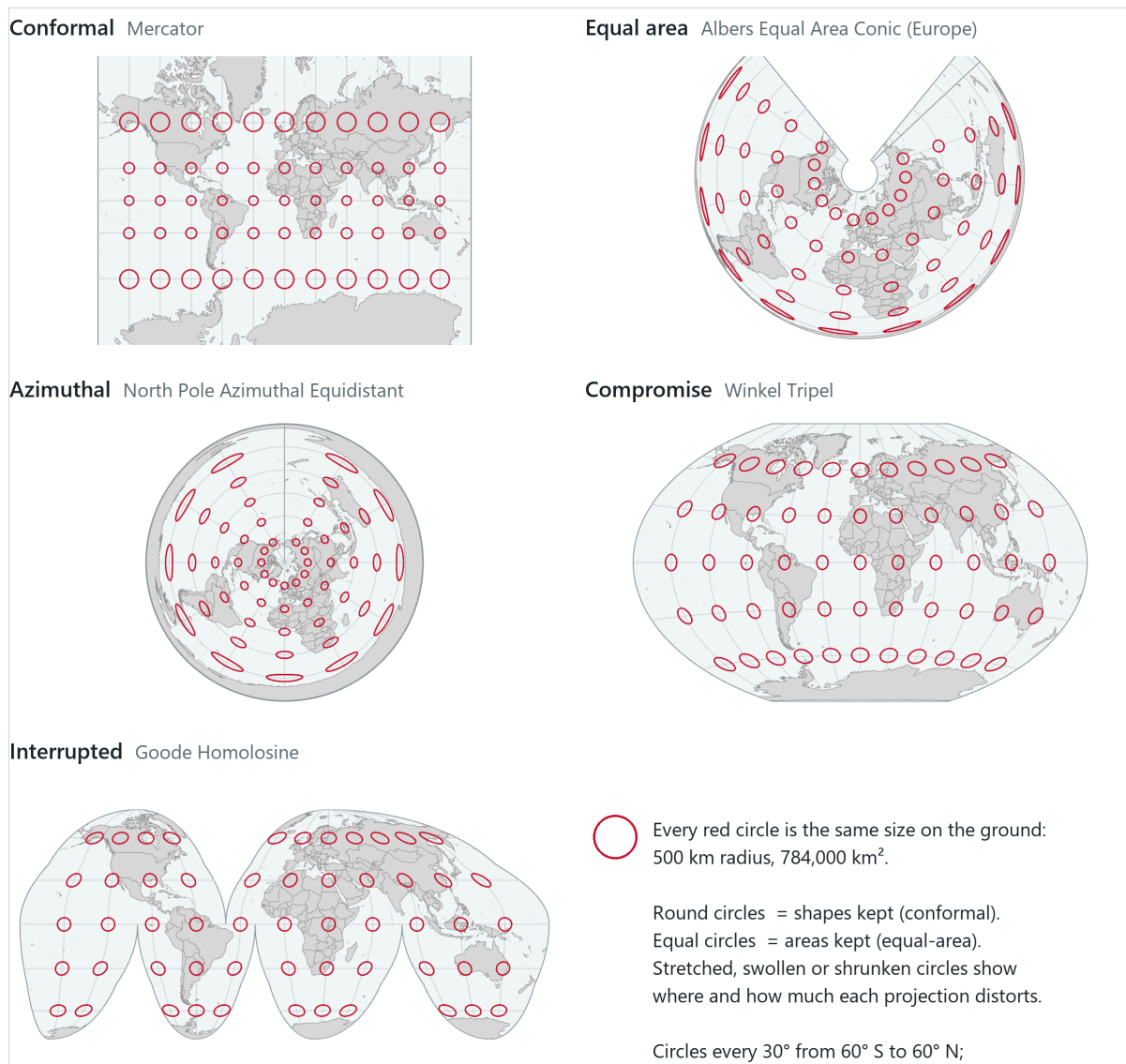


Figure 3.9 The course's gallery, one projection per class, each with the same 55 circles of true size: Mercator (conformal), Albers Equal Area Conic centred on Europe (equal-area), North Pole Azimuthal Equidistant (azimuthal), Winkel Tripel (compromise) and Goode Homolosine (interrupted and equal-area).

Countries: World Countries (Generalized), ArcGIS Living Atlas of the World. Sources: Esri; Garmin International, Inc.; U.S. Central Intelligence Agency (The World Factbook); National Geographic Society

The circles show how each projection in the gallery distorts. On Mercator they stay round and grow larger towards the poles. On the Albers they all cover the same area, but only those near Europe are round. On the azimuthal map they stretch towards the edge, where Antarctica is stretched around the whole circumference. On the Winkel Tripel they are slightly oval everywhere; on the Goode they are equal and fairly round across the land.

RULE OF THUMB

Choose the projection to suit the purpose of the map. Anything about area or density (a choropleth, a density surface, a world thematic map) should be mapped on an equal-area projection; conformal projections suit reference and navigation maps. Never use Web Mercator for a world thematic map.

Cylinder, cone and plane

Most projections belong to one of three families, named after a surface that can be unrolled flat without stretching, a **developable surface**. A **cylinder** wrapped round the globe unrolls into a rectangle, with straight, evenly spaced meridians. A **cone** set over the globe unrolls into a fan, meridians radiating from its apex and parallels as arcs. A **plane** touching the globe at one point gives a disc, meridians radiating from the centre at their true angles. The surface gives the family its name and determines the general appearance of the graticule, but the formulas themselves are derived mathematically.

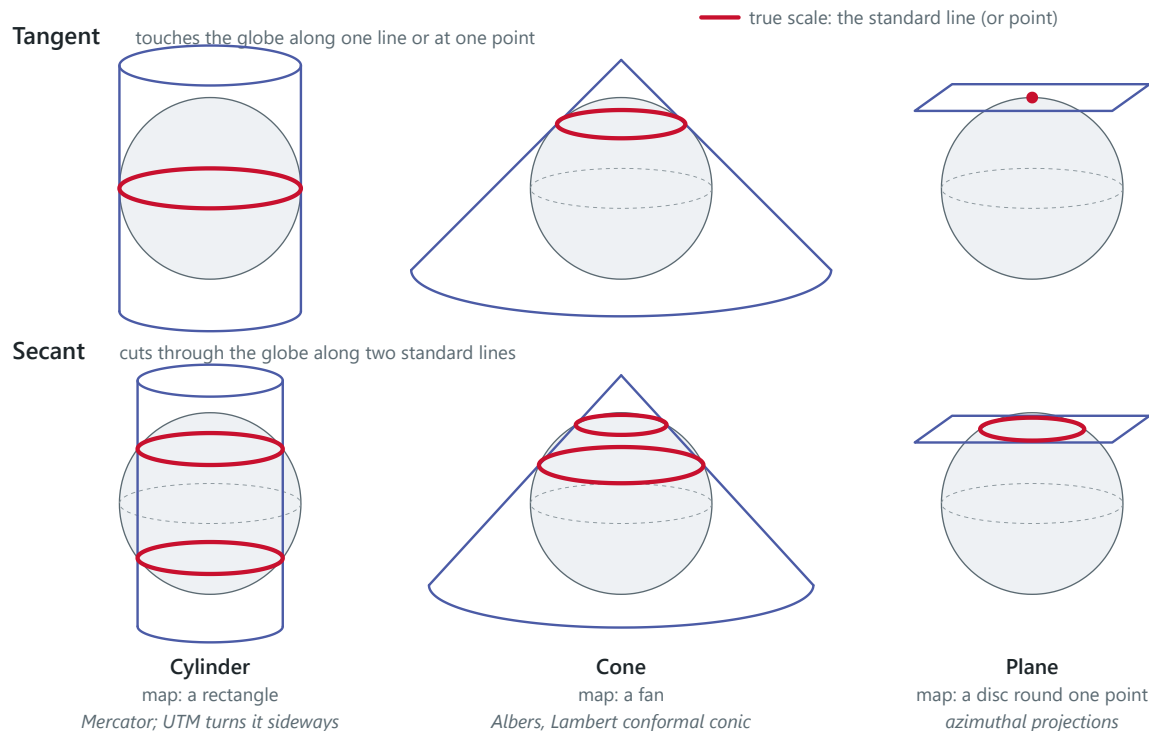


Figure 3.10 Cylinder, cone and plane, each tangent to the globe (touching along one line or at one point) and secant (cutting through it). Scale is true along the standard lines in red, where the surface meets the globe.

Where the surface meets the globe, scale is true: the scale factor is 1. A **tangent** surface touches along one line (a plane, at one point). A **secant** surface cuts through the globe along two lines, the **standard lines** (for a cone or cylinder, two standard parallels; a secant plane cuts along one circle). Between them the scale is slightly too small, outside them slightly too large, and the low-distortion band is wider than a tangent surface gives. True scale on a standard line does not mean that the map is free of distortion there. The gallery's Albers is a secant cone with standard parallels at 43° N and 62° N, so its circles are round across Europe.

A cylinder can also be turned on its side to touch the globe along a meridian, a **transverse** cylinder, so that the band of true scale runs north–south. UTM and the national grids described below are based on this arrangement.

The famous projections

A small number of projections are used for most of the world's mapping, and each one makes a different trade-off between the four properties.

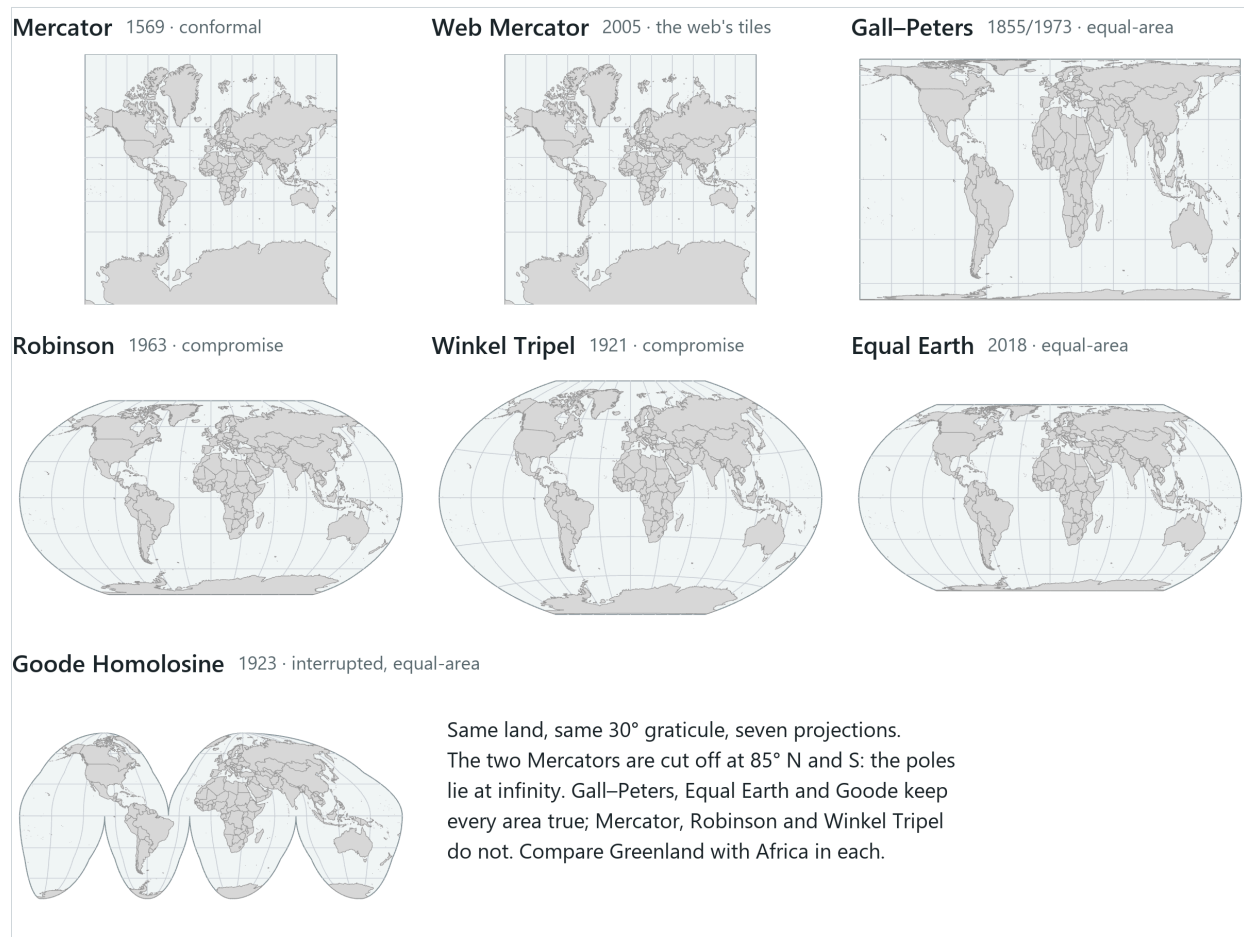


Figure 3.11 Seven world projections from the same land and the same 30° graticule. The two Mercators are cut at 85° N and S because the poles lie at infinity; Gall-Peters, Equal Earth and Goode keep every area true, Mercator, Robinson and Winkel Tripel do not.

Countries: World Countries (Generalized), ArcGIS Living Atlas of the World. Sources: Esri; Garmin International, Inc.; U.S. Central Intelligence Agency (The World Factbook); National Geographic Society

- **Mercator** (1569) is a conformal cylinder whose parallels are spaced ever wider towards the poles, so that a straight line is a line of constant compass bearing, which made it the standard map for navigation. Scale grows as $1/\cos(\text{latitude})$ and area as its square; Greenland, one-eighth the area of South America, looks almost as large.
- **Web Mercator** applies the Mercator formula for a sphere to WGS 84 latitudes and longitudes. It is a projected coordinate system (EPSG 3857), the one most web basemaps are drawn in, and not exactly conformal.

- **Gall–Peters** is a cylindrical equal-area projection with standard parallels at 45° N and S (James Gall, 1855; promoted by Arno Peters from the late 1960s). Areas are true; shapes near the equator stretch north–south.
- **Robinson** (1963) is a compromise built from tables to make the world "look right"; National Geographic used it from 1988 to 1998.
- **Winkel Tripel** (1921) is a compromise with small errors everywhere; National Geographic's world maps have used it since 1998.
- **Equal Earth** (2018) is an equal-area world projection designed to look like the Robinson while keeping every area true.
- **Goode Homolosine** (1923) is equal-area and almost always shown interrupted, in the form familiar from world atlases.

IN THE PANE

Module 3 lesson 3 covers the four properties, the three surfaces and tangent versus secant, then shows five gallery pictures with the same circles as above. You zoom your own map out to the whole world, switch it to **Equal Earth** in Map Properties > Coordinate Systems and watch it redraw, compare Greenland with South America, and switch it back to Web Mercator. Last, you drag Ireland north and south on thetruesize.com, a Mercator web map.

Choosing a coordinate system

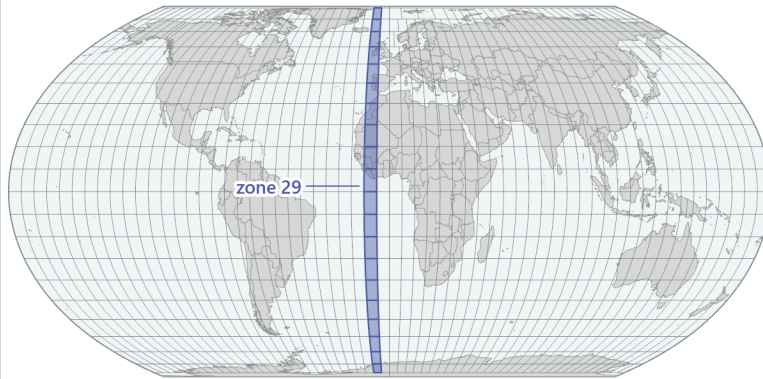
A world projection has some distortion everywhere. For a city, county or country the usual choice is a zoned system, in which the world is divided into zones small enough that the distortion inside each stays below a set limit.

Universal Transverse Mercator (UTM) cuts the Earth into 60 zones, each 6° of longitude wide, from 84° N to 80° S. Each zone is a transverse Mercator with the scale on its central meridian reduced to 0.9996, which makes it secant: scale is true along two lines about 180 km either side, slightly too small between, slightly too large beyond. Coordinates are metres, with a false easting of 500,000 m (and a false northing of 10,000,000 m in the southern hemisphere) so that none is negative. UTM is a grid scheme rather than a datum, so there are WGS 1984, NAD 1983 and ETRS89 versions of each zone. Galway and Dublin both fall in zone 29, whose central meridian, 9° W, runs along Ireland's west coast.

In the United States the **State Plane Coordinate System** divides each state into zones along state and county lines, each designed to keep scale error within 1 part in 10,000.

Many countries have a **national grid**. Ireland's is **Irish Transverse Mercator** (ITM, EPSG 2157), introduced in 2001 as the GPS-compatible successor to the Irish Grid: a transverse Mercator on IRENET95 with its central meridian at 8° W down the middle of the island and a scale factor of 0.99982, which puts the two lines of true scale about 121 km either side. Its origin, 53.5° N 8° W, is given the false coordinates 600,000 m E, 750,000 m N, chosen so that old Irish Grid coordinates interpreted as ITM fall outside the island, which makes a mix-up immediately visible.

60 UTM zones, 6° wide



Irish Transverse Mercator (ITM)



Figure 3.12 Zoned systems. Left: the 60 UTM zones, each 6° wide, with zone 29, which holds Ireland. Right: Ireland in ITM, its central meridian at 8° W and its false origin, 53.5° N 8° W, at 600,000 m E, 750,000 m N.

UTM zones: World UTM Grid, Esri (ArcGIS Online). Countries: World Countries (Generalized), ArcGIS Living Atlas of the World (Sources: Esri; Garmin International, Inc.; U.S. Central Intelligence Agency (The World Factbook); National Geographic Society). Basins: Hydrometric Areas Ireland & Northern Ireland, Environmental Protection Agency (Ireland)

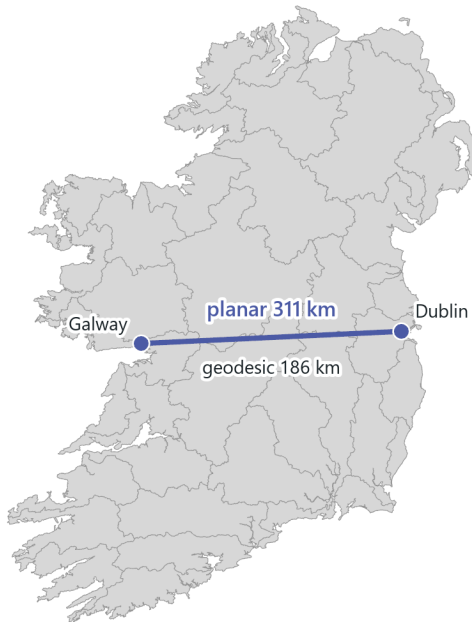
RULE OF THUMB

For a local study area, use the national grid where one exists (ITM in Ireland), otherwise the UTM zone that holds the area. Keep world projections for world maps. The software can draw any data in any coordinate system, so choosing a suitable system is the analyst's responsibility.

Web Mercator: fine for looking, wrong for measuring

Web basemaps are cut into square tiles in Web Mercator, and Pro's maps often start in it because the basemap is the first layer added. It is well suited to displaying the world on a screen but not to measurement. Like Mercator, it stretches lengths by $1/\cos(\text{latitude})$: at Ireland's 53.3° N that is 1.67, so every length is two-thirds too long and every area about 2.8 times too large.

Web Mercator (EPSG 3857)



Irish Transverse Mercator (EPSG 2157)



Both frames drawn at the same map scale (1:9,000,000 in map units).

Figure 3.13 Dublin to Galway in two projected coordinate systems, drawn at the same map scale. In Web Mercator the straight line measures 311 km; in Irish Transverse Mercator it measures 186 km, the true (geodesic) distance.

Basins: Hydrometric Areas Ireland & Northern Ireland, Environmental Protection Agency (Ireland)

A planar measurement from Dublin to Galway in Web Mercator gives 311.65 km. The **geodesic** distance, the shortest path along the curved surface of the ellipsoid, is 186.6 km. With the map in ITM, a planar measurement gives 186.57 km. **Planar** measurement applies flat geometry to the coordinates of the system it is given, so it is only as accurate as that projection is in the area where you work. Many tools can measure either way: Buffer, the Measure tool and the distance tests of Chapter 6 offer a geodesic option, while others (the area and length fields, densities, most overlays) are planar in the data's own system. Check which method a tool uses before relying on its result. A 5 km buffer drawn planar in Web Mercator near Dublin reaches only about 3 km on the ground, and the map gives no visible sign of the error.

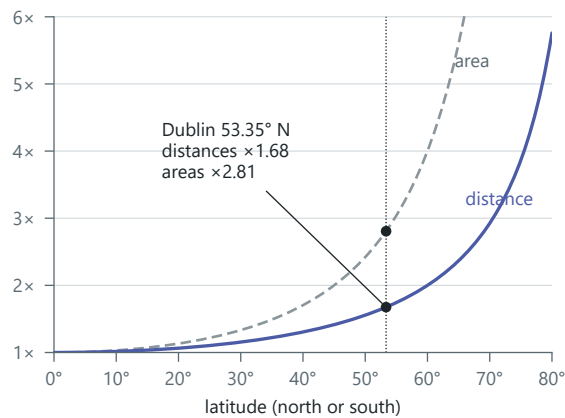


Figure 3.14 Web Mercator's scale factor against latitude: distances are stretched by $1/\cos(\text{latitude})$, areas by its square.

RULE OF THUMB

Never measure planar in Web Mercator. Project the data to an analysis coordinate system before measuring, or use a geodesic method; when a single distance or area is asked for, report both the geodesic and the planar number and why they differ.

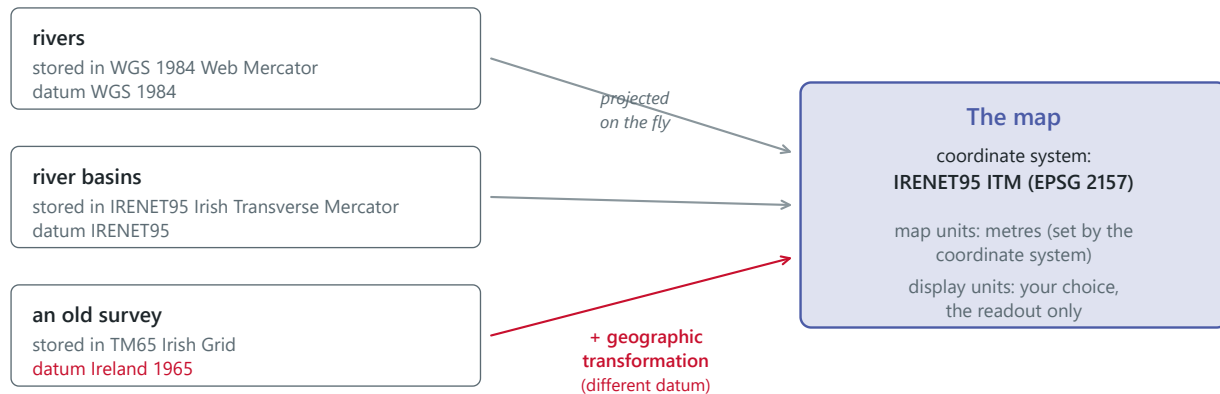
IN THE PANE

Module 3 lesson 4 starts with large versus small scale and adds the World UTM Grid to find Galway's zone. You measure Dublin to Galway with the **Measure** tool, planar then geodesic, in Web Mercator; then you switch the map to ITM by searching **2157** (Esri's list spells it "IRENET95 Irish Tranverse Mercator") and measure again. Waldo checks your readings and runs its projection check, which recommends ITM for Irish data.

Map units, on-the-fly projection and transformations

A layer keeps the coordinate system it was stored in. The map has one of its own, usually set by the first layer added, and each time it draws, the software reprojects every layer into it. This **on-the-fly projection** allows layers from different sources to line up, and it changes nothing stored on disk.

Stored three ways, drawn one way



Nothing on disk changes: each layer keeps its own coordinate system; only the drawing is projected. A layer on another datum also needs a transformation between datums, or it lands tens to hundreds of metres off.

Figure 3.15 Three layers stored in three coordinate systems are projected on the fly into the map's system for drawing; the layer on another datum also needs a geographic transformation.

The map's coordinate system fixes its **map units**: metres for Web Mercator or ITM, degrees for a GCS. **Display units** only change how the readout and measurements are shown (metres, feet, decimal degrees, degrees-minutes-seconds); they change no coordinates.

When a layer's datum differs from the map's, reprojecting is not enough, because latitude and longitude themselves differ between datums. A **geographic transformation** converts coordinates from one datum to another, by a published set of shifts, rotations and a scale change or by a grid of corrections, each good to a stated accuracy. NAD 1927 to NAD 1983 needs one; so does Ireland 1965 to WGS 1984. Pro picks a default, or you choose one, on the **Transformation** tab of the map's properties; choosing the wrong one leaves a layer misplaced by tens to hundreds of metres.

Three operations are easy to confuse. **Defining** a coordinate system, or repairing a missing or wrong one, tells the software what the stored numbers mean and moves nothing; the right answer is the system the data were made in, never the one that makes the layer look plausible. **Projecting** computes new coordinate numbers for the same places and writes a new dataset; this is done with the Project tool, covered in a later module. Changing the **map's** coordinate system changes only how the layers are drawn. The next section shows the effect of a wrong definition: the numbers stay the same, but the layer is drawn in a different place.

Why it matters

The choice of coordinate system affects analysis as well as display. Overlay needs every layer in a common coordinate system on compatible datums. Measurements are only as accurate as the projection in which they are made. A datum mismatch without a transformation shifts one layer against another by tens of metres, enough to put a point on the wrong side of a river or a boundary. A layer whose coordinate system is labelled wrongly keeps its numbers, but they are interpreted differently, and the software draws the data at the location the label implies.

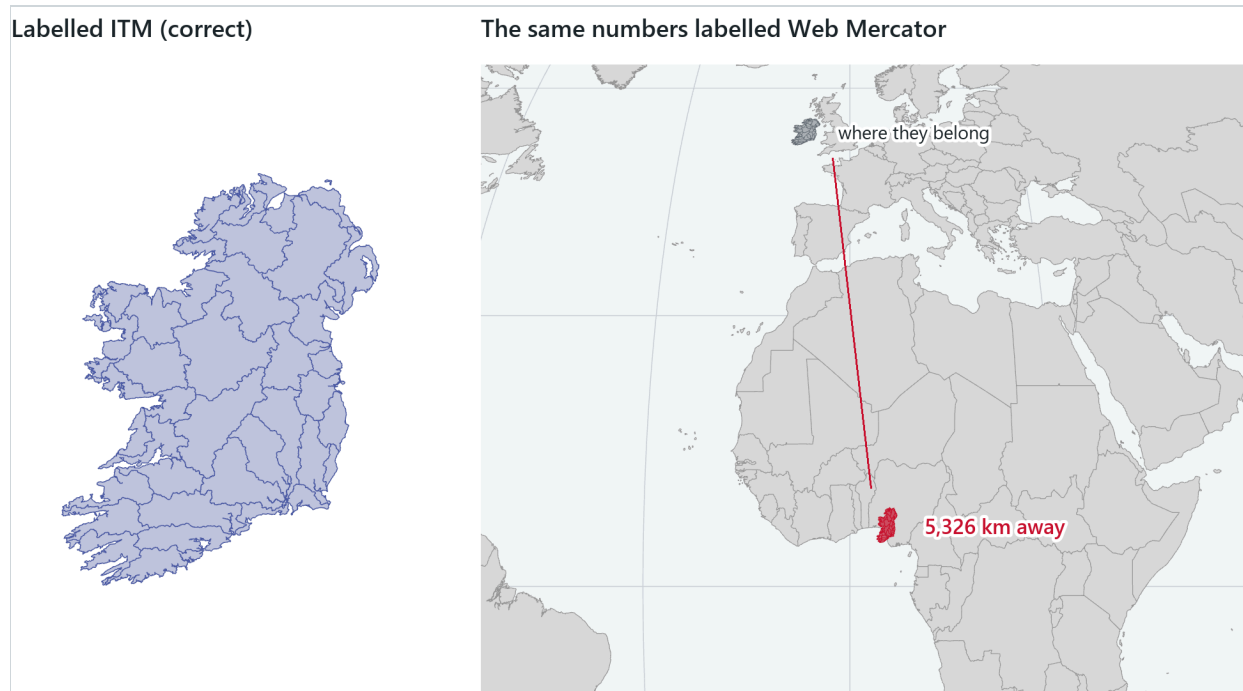


Figure 3.16 The same coordinate numbers read in two coordinate systems. The Irish river basins' ITM coordinates, correctly labelled, draw over Ireland; labelled as Web Mercator metres, the same numbers put the basins 5,326 km away, off West Africa.

Basins: Hydrometric Areas Ireland & Northern Ireland, Environmental Protection Agency (Ireland). Countries: World Countries (Generalized), ArcGIS Living Atlas of the World (Sources: Esri; Garmin International, Inc.; U.S. Central Intelligence Agency (The World Factbook); National Geographic Society)

The software will compute in the wrong system and draw a map that looks normal. Checking the coordinate system, the datum and the measurement method (planar or geodesic) is the responsibility of the person doing the analysis, in every analysis.

RULE OF THUMB

Check the projection before any overlay or measurement: read every layer's datum and coordinate system, choose one analysis system suited to the place and the purpose, and project to it before you measure.

TRY IT

With a course project open, ask Waldo in Ask mode: which coordinate system is my map in, and which would you use to measure distances here? Waldo reads the map, runs its projection check on a layer and names a system. Compare its answer with the rules of thumb in this chapter.

Key terms

ellipsoid (spheroid)

the smooth shape made by rotating an ellipse about its shorter axis; the mathematical model of the Earth, defined by its two semi-axes.

flattening

$f = 1 - b/a$, how much shorter the polar radius is than the equatorial; about 1/298 for the Earth, shown in dialogs as its inverse, 1/f.

geoid

the irregular surface of equal gravity potential that mean sea level follows; the reference for heights, within about ± 100 m of the ellipsoid.

datum	the ellipsoid and its position relative to the Earth, which give latitude and longitude a definite meaning; local datums fit one region, geocentric ones the whole Earth.
graticule	the network of meridians and parallels drawn on a globe or a map.
geographic coordinate system (GCS)	latitude and longitude in degrees, defined by a datum, a prime meridian and an angular unit.
projected coordinate system (PCS)	a GCS plus a map projection and a linear unit, giving flat x and y coordinates in metres or feet.
map projection	a mathematical transformation of the curved Earth onto a plane; it always distorts shape, area, distance or direction somewhere.
representative fraction (RF)	a map scale written as 1:n, one unit on the map for n of the same unit on the ground.
large scale / small scale	a large-scale map (1:10,000) shows a small area in detail; a small-scale map (1:1,000,000) a large area with little detail.
scale factor	the actual scale at a point of a projected map divided by the stated scale; 1 on the standard lines.
conformal	preserving local angles, so small shapes are true; areas are distorted.
equal-area	preserving the relative sizes of areas everywhere; shapes are distorted.
equidistant	preserving true scale along certain lines, such as from one or two points to all others.
azimuthal	preserving true directions from a centre point; can also be conformal, equal-area or equidistant.
compromise projection	a projection that keeps no property exactly but every distortion moderate, such as Robinson or Winkel Tripel.
developable surface	a cylinder, cone or plane: a surface that unrolls flat without stretching, used to build and classify projections.
tangent / secant	a tangent surface touches the globe along one line or at one point; a secant surface cuts it along two standard lines (a plane along one circle).
UTM (Universal Transverse Mercator)	60 transverse Mercator zones, each 6° wide, with a central-meridian scale factor of 0.9996 and coordinates in metres.
Irish Transverse Mercator (ITM)	Ireland's national grid (EPSG 2157): transverse Mercator on IRENET95, central meridian 8° W, scale factor 0.99982, false origin 600,000 m E, 750,000 m N.
Web Mercator	the spherical Mercator formula applied to WGS 84 coordinates (EPSG 3857); the system of most web basemaps, stretching lengths by $1/\cos(\text{latitude})$.
geodesic	the shortest path between two points along the curved surface of the ellipsoid; geodesic measurement ignores the map's projection.
planar	measured with flat geometry in a projected coordinate system, so only as accurate as that projection is where you measure.
on-the-fly projection	the software's redrawing of every layer in the map's coordinate system at display time, without changing the stored data.
geographic transformation	a calculation that moves coordinates from one datum to another, such as NAD 1927 to NAD 1983, so that data line up.

Further reading

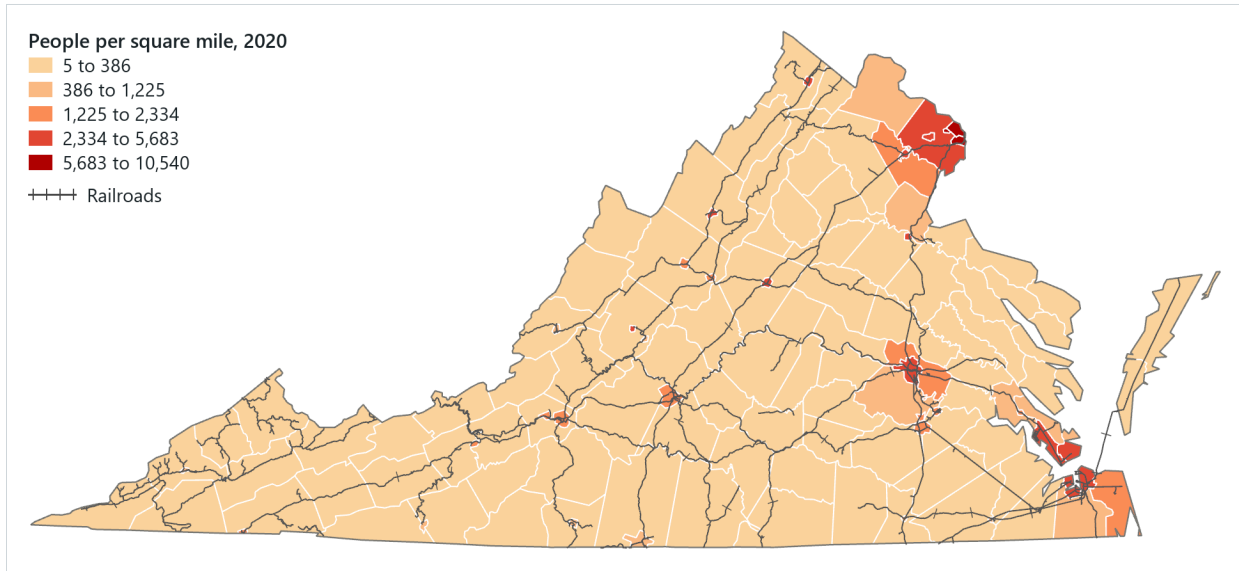
- Snyder, J. P. (1987). *Map Projections: A Working Manual*. U.S. Geological Survey Professional Paper

1395. <https://doi.org/10.3133/pp1395>
- Snyder, J. P., & Voxland, P. M. (1989). *An Album of Map Projections*. U.S. Geological Survey Professional Paper 1453. <https://pubs.usgs.gov/publication/pp1453>
 - U.S. Geological Survey. *Map Projections* (poster). <https://pubs.usgs.gov/publication/70047422>
 - NOAA National Ocean Service. *Global Positioning Tutorial: Geodesy* (the geoid and datums). https://oceanservice.noaa.gov/education/tutorial_geodesy/
 - Olaya, V. *Introduction to GIS*, the chapter on cartography and geodesy. <https://volaya.github.io/gis-book/en/>
 - Cory, M. J., Morgan, A., Bray, C., & Greenway, I. (2001). "A New Coordinate System for Ireland." FIG Working Week, Seoul.
 - Šavrič, B., Patterson, T., & Jenny, B. (2019). "The Equal Earth map projection." *International Journal of Geographical Information Science* 33(3). <https://doi.org/10.1080/13658816.2018.1504949>
 - Esri ArcGIS Blog. "Geographic vs. projected coordinate systems." <https://www.esri.com/arcgis-blog/products/arcgis-pro/mapping/coordinate-systems-difference>
 - Esri ArcGIS Blog. "Projection on the fly and geographic transformations." <https://www.esri.com/arcgis-blog/products/arcgis-pro/mapping/projection-on-the-fly-and-geographic-transformations>
 - The True Size Of..., an interactive Mercator map for dragging countries north and south. <https://thetruesize.com/>
 - The course's data sources: every layer the lessons and this chapter's maps use, with its publisher and licence.

CHAPTER 4

Cartographic design

Accompanies Module 4 · Cartographic Design



Counties: USA Counties Generalized Boundaries, ArcGIS Living Atlas of the World (Sources: Esri; U.S. Census Bureau). Railroads: North American Rail Network Lines, Bureau of Transportation Statistics (public domain). Colours: ColorBrewer OrRd, Cynthia Brewer (Apache 2.0)

WHY THIS MATTERS

A map can be correct and still mislead its readers or fail to be read. Every map leaves out most of the world, and the choices of symbol, colour, classes, lettering and layout determine what a reader learns from it. The map above shows Virginia's 133 counties and independent cities by people per square mile, a rate, in five classes of one hue. This chapter explains each of those choices and the reasoning behind it.

YOU SHOULD BE ABLE TO EXPLAIN

- why a choropleth of counts mostly maps the size of the areas, and what a rate shows instead
- why the same data can tell a different story under a different classification method
- why size, weight and value suit amounts while hue and shape suit kinds of thing

IN THIS CHAPTER

Every map is generalized

Reference and thematic maps

Visual variables

Colour

Thematic map types

Counts and rates

Classification

Typography and labels

Map elements and the layout

Design principles

Every map is generalized

A map is a representation of geographic data that communicates a point of view. **Cartography** is the art and science of designing maps, and the cartographer's decisions determine what a reader can learn from a map. The real world is far too detailed to show in full, so a map leaves some things out and simplifies and groups others, so that the important features can be seen. The geographer Mark Monmonier called the result a set of "white lies" told in the reader's interest.

This process is called **generalization**: selecting, organizing and transforming geographic information into a form that can be mapped at a given scale and for a given purpose. Its operations are **selection** (what to include), **simplification** (simpler shapes for the scale), **classification** (grouping features by a measured attribute) and **symbolization** (turning information into graphic marks). Classification and symbolization take up most of this chapter.

The U.S. Census Bureau publishes the same state boundaries generalized for three display scales, as cartographic boundary files at 1:500,000, 1:5,000,000 and 1:20,000,000. Virginia's outline has 326 points in the smallest-scale file and 5,933 in the largest.



Figure 4.1 The mouth of the Chesapeake Bay drawn from the Census Bureau's three cartographic boundary files, all at the same size (about 1:2,900,000). The file made for 1:20,000,000 omits the rivers and the detail of the Eastern Shore's shape; the one made for 1:500,000 includes every inlet.

State outlines: U.S. Census Bureau cartographic boundary files, 2023, at 1:20,000,000, 1:5,000,000 and 1:500,000 (public domain)

At a small scale, detail that cannot be seen only adds clutter, and a boundary simplified for a national map looks angular when viewed close up. The county layer used in Module 4 is itself generalized (Accomack County, on the Eastern Shore, is drawn with 36 points), so it draws quickly for the whole country.

Reference and thematic maps

A **reference map** shows where things are, natural and human-made, with many features at roughly equal weight: a street map, a topographic map, a basemap. It serves many purposes. A **thematic map** has one subject, such as population density or disease rates, and one purpose; the other features on it provide context. Every map made in Module 4 is thematic.

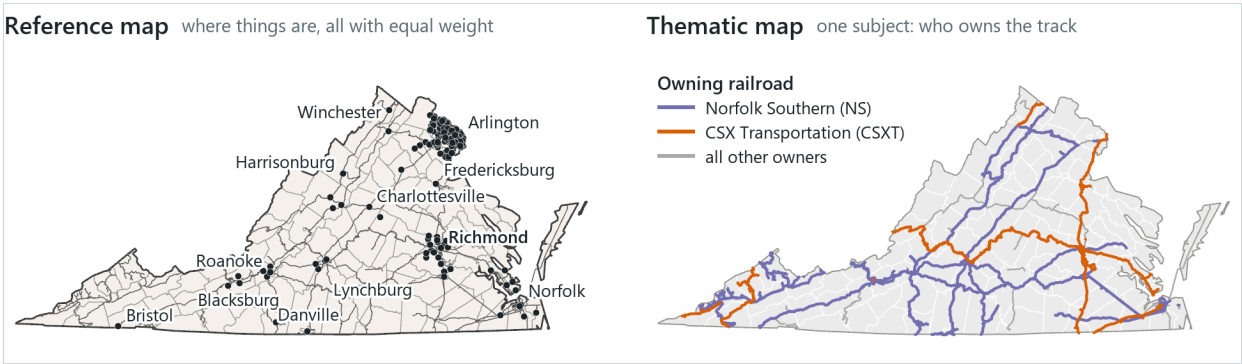


Figure 4.2 Virginia as a reference map and as a thematic map. The reference map gives counties, railroads and places equal weight and names the main cities. The thematic map has one subject, which company owns each railroad, and shows everything else in pale grey.

Counties: USA Counties Generalized Boundaries; places: USA Major Cities (ArcGIS Living Atlas of the World; Sources: Esri; U.S. Census Bureau). Railroads: North American Rail Network Lines, Bureau of Transportation Statistics (public domain). Colours: ColorBrewer Dark2, Cynthia Brewer (Apache 2.0)

The second decision is what kind of attribute is being shown. A **qualitative** attribute distinguishes kinds of things; a **quantitative** attribute measures an amount.

Layer (Virginia)	Qualitative attributes	Quantitative attributes
Counties (133)	NAME, FIPS code	POPULATION, SQMI, POP_SQMI
Places (129)	NAME, CLASS (city, town, Census Designated Place)	POPULATION
Railroad segments (5,176)	owning company (RROWNER1)	MILES

The categories depend on the state. The USA Major Cities layer holds places of 10,000 people or more plus the state capitals, so Virginia's places fall into three classes while a state such as Idaho has only one, city. A map with one class is still a correct map of that attribute, but it shows no differences between the features.

A code made of digits, such as a FIPS code, is still qualitative, because it serves as a label rather than measuring an amount. Of Virginia's 5,176 rail segments (about 4,306 miles), Norfolk Southern owns 2,699 and CSX Transportation 1,658; the thematic map above shows this ownership.

Visual variables

Map symbols are points, lines and areas, and each is designed by adjusting a small number of graphic properties. These properties are the **visual variables**, the characteristics of a mark that the eye perceives as a difference; the idea is usually traced to the French cartographer Jacques Bertin in 1967.

	Point	Line	Area	Reads as
Size bigger = more				more / less
Weight line or outline thickness				more / less
Value light to dark				more / less
Saturation greyed to pure				more / less
Hue which colour				a kind
Shape (character) circle, square, dashed, ties				a kind
Orientation the angle of a mark				a kind
Texture / pattern fill patterns				a kind

Figure 4.3 Eight visual variables applied to point, line and area symbols. Size, weight, value and saturation read as "more" or "less"; hue, shape, orientation and texture read as "a different kind".

Hues: ColorBrewer Dark2, Cynthia Brewer (Apache 2.0)

The variables fall into two groups according to what the eye perceives from them. **Size**, **weight** (the thickness of a line or outline) and **value** (light versus dark) read as order, so they suit amounts.

Hue and **shape** (the character of a symbol: circle versus square, solid versus dashed, a plain line versus one with ties) read as differences in kind, so they suit categories. A real symbol usually combines several variables: the gallery's Railroad symbol in ArcGIS Pro is two layers, a grey stroke and a marker (the ties) repeated along the line.

RULE OF THUMB

Match the variable to the attribute. Use size, weight or value for amounts and hue or shape for kinds. A rainbow of hues for a count, or different sizes for categories, makes the symbols harder for the reader to interpret.

IN THE PANE

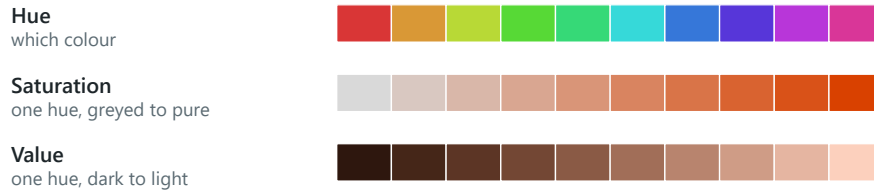
Module 4 lesson 1 builds a new project with three Living Atlas layers (counties, major cities and the rail network), each filtered to one state by a definition query, and compares reference and thematic maps. Lesson 2 reads qualitative and quantitative columns from the counties table, shows the visual variables and gives each layer a single symbol, the railroads the gallery Railroad symbol.

Colour

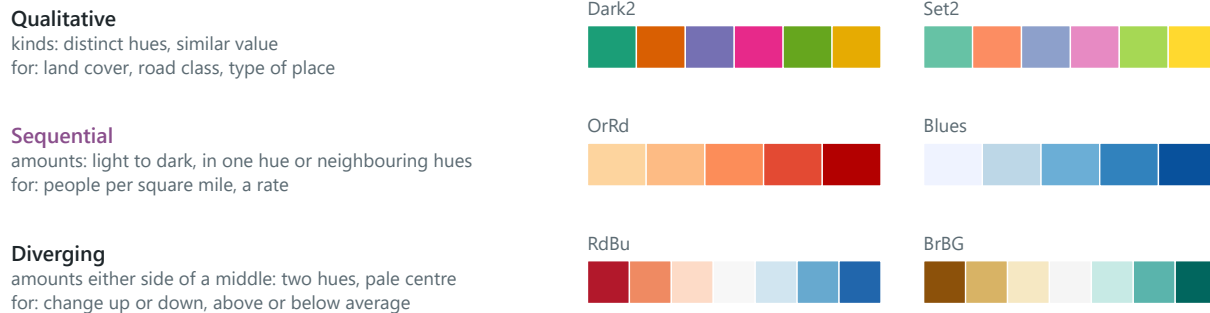
Colour has three dimensions. **Hue** is which colour it is (red, blue). **Saturation** is how pure or greyed it is. **Value** is how light or dark it is. Screens mix red, green and blue light, and Pro's colour picker shows hue, saturation and value beside the RGB numbers.

Schemes for data come in three families. A **qualitative** scheme uses distinct hues of similar value, for categories. A **sequential** scheme runs from light to dark, in one hue or a few neighbouring ones, for amounts: the order of lightness represents the order of the values, and darker colours are read as larger amounts. A **diverging** scheme joins two sequential ramps at a pale middle, for amounts with a meaningful midpoint such as zero change or an average.

Three dimensions of colour



Three families of scheme



Sequential and diverging schemes carry order, so they suit amounts; a rainbow of hues carries none.

Figure 4.4 The three dimensions of colour, and the three families of scheme with two ColorBrewer examples each. OrRd, the sequential scheme on this chapter's maps, starts at a pale tint rather than white.

Colour schemes: ColorBrewer, Cynthia Brewer, Mark Harrower and The Pennsylvania State University (Apache 2.0). This product includes color specifications and designs developed by Cynthia Brewer (<http://colorbrewer.org/>).

ColorBrewer, by Cynthia Brewer and Mark Harrower at Penn State, is a widely used source of tested schemes, and Pro's colour ramps include many of them. A choropleth usually needs three to five classes in one ordered scheme; one hue from light to dark is a reliable first choice, and beyond about seven classes the eye cannot separate them. If a scheme's lightest class is white, the lowest class is hard to see between white borders, so choose one that starts at a tint. Also check whether the scheme suits readers with colour-vision deficiency; ColorBrewer marks this for each scheme, because not every scheme stays distinct at every class count. Earth tones usually look better than strident colours, and emphasis comes from colour, size and line width.

Thematic map types

Each symbology type in Pro represents one attribute with one visual variable. The Primary symbology menu lists them in order: Single Symbol, Unique Values, then the quantity types Graduated Colors, Bivariate Colors, Unclassed Colors, Graduated Symbols and Proportional Symbols, and further down Dot Density, Charts and Heat Map.

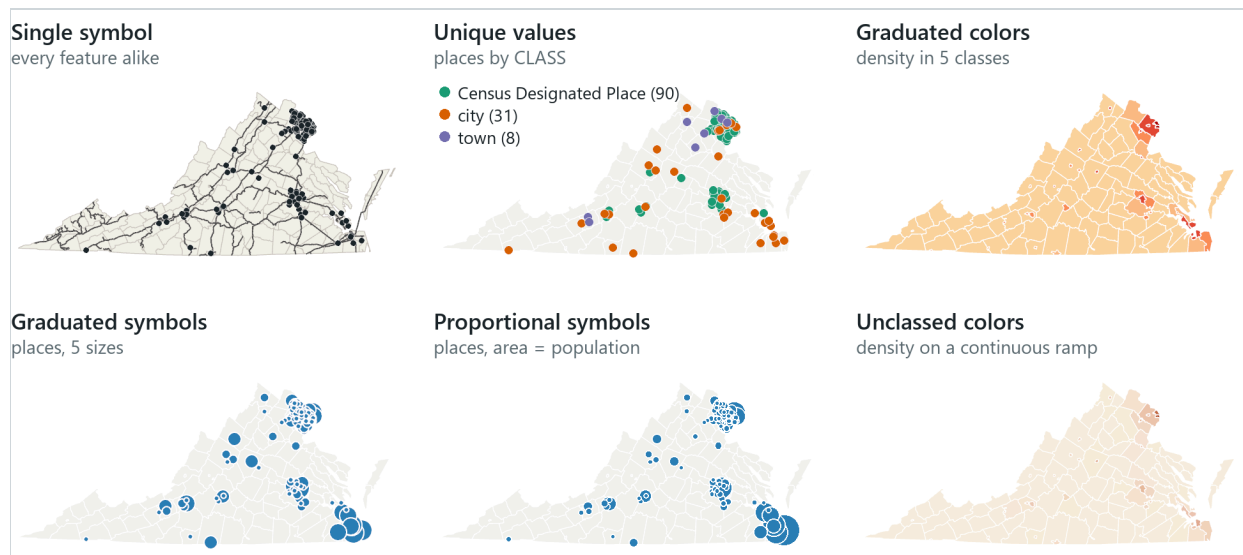


Figure 4.5 Six symbology types on the same Virginia layers. Single symbol draws every feature alike; unique values gives each category its own hue; graduated colors shades counties by class; graduated symbols sizes places in five steps; proportional symbols make each circle's area proportional to population; unclassified colors shade on a continuous ramp.

Counties: USA Counties Generalized Boundaries; places: USA Major Cities (ArcGIS Living Atlas of the World; Sources: Esri; U.S. Census Bureau). Railroads: North American Rail Network Lines, Bureau of Transportation Statistics (public domain). Colours: ColorBrewer OrRd, Dark2 and Blues, Cynthia Brewer (Apache 2.0)

- **Single symbol** draws every feature alike; labels distinguish them.
- **Unique values** gives each category its own symbol, usually its own hue.
- **Classed** types sort values into a few classes first: graduated colors (the **choropleth** map of shaded areas), graduated symbols and bivariate colors.
- **Unclassed** types do not use classes: unclassified colors, proportional symbols (each symbol sized in proportion to its value) and dot density.

Contour lines are used for surfaces and are covered with rasters in Module 8. Virginia's largest place in the USA Major Cities layer is Virginia Beach, with 459,470 people; on the graduated-symbol map it falls in the top class, which starts above 186,247.

Counts and rates

Colouring areas by a count, such as total population, mostly shows how large each area is, because a large county tends to contain more of most things and the eye responds to the area rather than to the number. A **rate** divides the count by the area or by another count, so it describes the area regardless of its size. Pro's **normalization** option divides one field by another on the fly.

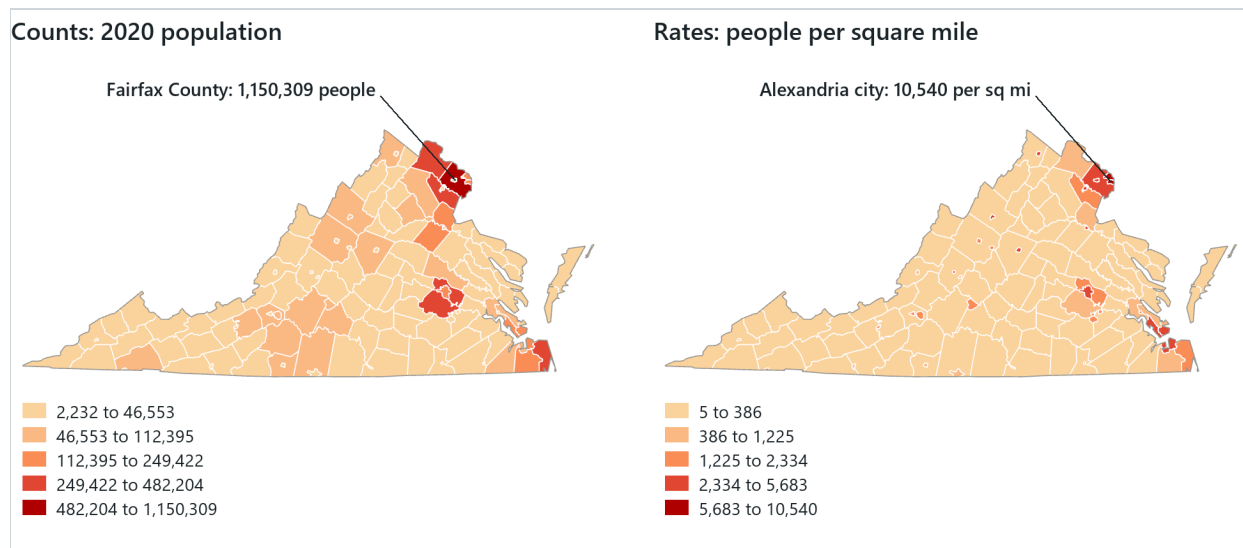


Figure 4.6 The same counties by a count and by a rate, same scheme and same method. Fairfax County has the most people, 1,150,309; Alexandria city, 15.1 square miles, is the most crowded at 10,540 people per square mile.

Counties: USA Counties Generalized Boundaries, ArcGIS Living Atlas of the World (Sources: Esri; U.S. Census Bureau). Colours: ColorBrewer OrRd, Cynthia Brewer (Apache 2.0)

In Virginia, Fairfax County (397 square miles) has the highest population and Alexandria city the highest people per square mile. The two maps show different things about the same state. The rate map compares how crowded the counties are. The count is also a legitimate question (how many people live here?), and proportional symbols show it without letting the size of each area stand in for the amount. The denominator must match the question: people per square mile, cases per person and the share of households measure three different things.

The shade of an area represents an average over the whole area; it does not mean that people are spread evenly inside it. If the boundaries are redrawn, for example counties into districts or tracts into counties, the pattern can change; geographers call this the **modifiable areal unit problem**. What is true of an area on average need not be true of the people in it, and assuming otherwise is the **ecological fallacy**: a county with a high rate contains many people to whom the rate does not apply.

RULE OF THUMB

On a map of areas, map rates rather than counts. Areas shaded by a count mostly show the size of each area; if the count itself matters, use proportional symbols instead.

Classification

Classification groups features into a small number of classes by their attribute values and a stated method. The method affects both the appearance of the map and the conclusions it appears to support, so examine the numbers before relying on the map.

Classification starts with a **histogram**, a chart of how many features have each value; Pro's Symbology pane has one on its Histogram tab. Data can be normal (a symmetric bell), uniform (flat), **skewed** (piled at one end with a long tail) or multi-modal. Virginia's 133 people-per-square-mile values run from 5.4 (Highland County) to 10,539.8 (Alexandria city). The median is 102.6, the

mean 894.9, and 84 of the 133 are below 250, so the distribution is strongly right-skewed. The **standard deviation**, 1,681.2 here, measures how spread out the values are around the mean; in a normal distribution about 68 % of values lie within one standard deviation of the mean.

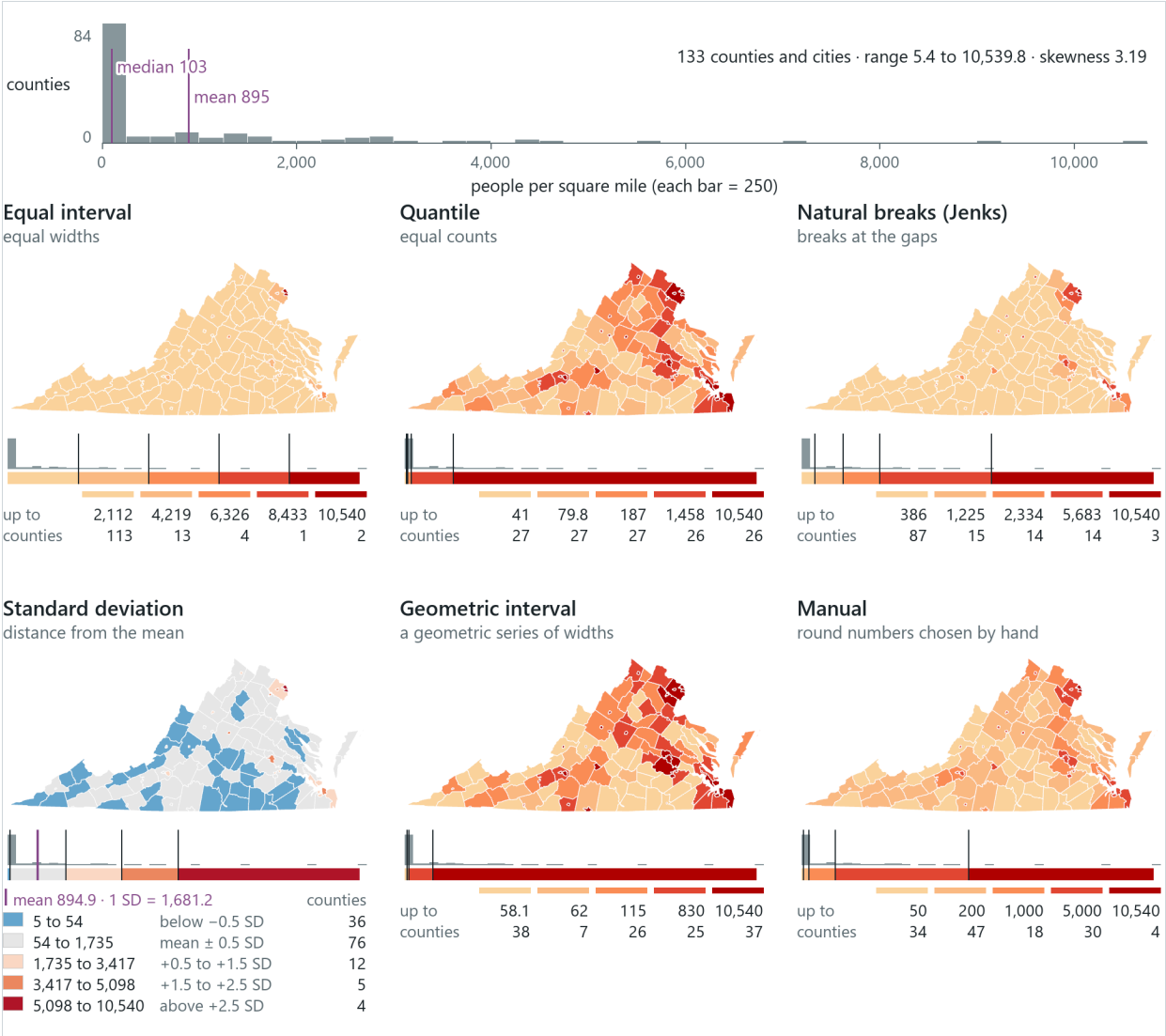


Figure 4.7 One variable classified by six methods. Across the top, the histogram of people per square mile for Virginia's 133 counties and cities; below, a map per method with its class breaks on a strip under the histogram, then each class's upper limit and its number of counties. The standard-deviation map has its own legend: each class in people per square mile, its distance from the mean in standard deviations, and the mean marked.

Counties: USA Counties Generalized Boundaries, ArcGIS Living Atlas of the World (Sources: Esri; U.S. Census Bureau). Colours: ColorBrewer OrRd and RdBu, Cynthia Brewer (Apache 2.0)

Method	How it cuts	Suits	Watch for
Equal interval	equal-width classes across the range	familiar ranges (percentages, temperature); general readers	clustered data lands in one class
Quantile	the same number of features per class	ranks ("the top fifth"); evenly spread data	alike values split, very different values share a class

Method	How it cuts	Suits	Watch for
Natural breaks (Jenks)	breaks where the data thin out	most data; Pro's default for graduated renderers	breaks fit one dataset, so two maps are hard to compare
Standard deviation	distance from the mean	bell-shaped data	the legend shows deviations, not values
Geometric interval	class widths follow a geometric series, which Pro may reverse once to fit the data	a compromise for continuous data	hard to explain
Manual	breaks chosen by hand	an outside standard, round numbers	check the histogram so no class is empty

The equal-interval map places 113 of the 133 in its lightest class, so it shows almost no variation. Quantile puts the same number of features in every class but shows most of the state in dark colours. Pro's natural breaks for these values are 386, 1,225, 2,334 and 5,683 people per square mile. The geometric-interval map's second class is only 4 people per square mile wide (58 to 62). This narrow class results from the series fitting a long tail; it is not an error, but it is hard to explain to a reader. The manual map uses breaks of 50, 200, 1,000 and 5,000.

RULE OF THUMB

Use natural breaks by default; use quantile when the reader thinks in ranks, and geometric interval only on request. Keep to three to five classes, read the histogram first, and state which method the map uses.

IN THE PANE

Module 4 lesson 3 puts unique values on the cities' class (as many classes as your state has), covers the three scheme families and the symbology menu, shows that Fairfax has the highest count and Alexandria the highest rate, and shades the counties by people per square mile. Lesson 4 opens the Histogram tab, applies equal interval, quantile and natural breaks in turn, reads their class breaks, and asks you to choose a method and give a reason.

Typography and labels

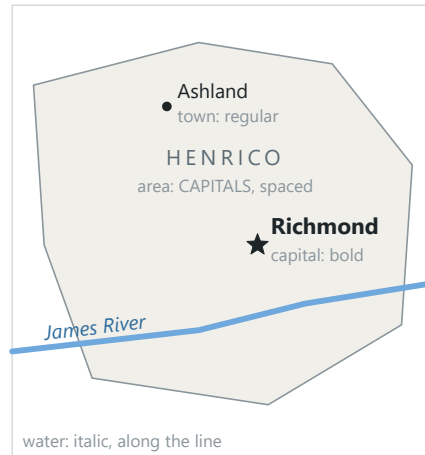
Text on a map is also a symbol. Type varies in **weight** (light to heavy), posture (roman or italic), width and case (capitals or mixed case). **Legibility** is whether the text can be read; **harmony** means using one or two typefaces and varying them rather than mixing many. Choose the typeface deliberately, because the default text font can differ from one computer to another, and a page built on defaults can end up with three typefaces that nobody selected.

Type as a symbol

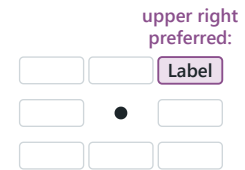
Weight	Light	Regular	Semibold	Bold
Posture	Roman	<i>Italic</i>		
Case	Mixed case	ALL CAPITALS		

One or two typefaces per map (harmony);
vary weight, posture and case within them.

Conventions



Point labels



Beside the point, upper right first; lines along the line; areas inside, spread across.

Labels that do not fit are left off, not forced.

Figure 4.8 Type varied as a symbol; the usual conventions on a small drawn map (an area name in spaced capitals, water in italic along the river, the capital in bold, a town in regular type); and the preferred position for a point label, upper right.

Map readers are familiar with a set of conventions: water features in italic, land features in roman; area features in capitals (FRANCE), point and line features in mixed case; bold for importance, such as a capital city. Point labels are placed beside the point, preferably at the upper right; line labels run along the line; area labels are placed inside the area and spread across it.

A map can carry two kinds of text. **Graphic text** (the title, a note) is not tied to features. **Labels** are tied to features: **dynamic labels** are placed automatically from a field for every feature and replaced at each zoom, while **annotation** is text placed and stored piece by piece. Pro omits duplicate and overlapping labels; forcing every label onto the map makes it unreadable. A **halo**, a thin light outline around letters, separates text from a busy background.

RULE OF THUMB

Give a halo to one label layer per map at most. Capitalize area names through the label expression (such as Upper(\$feature.NAME) in Arcade), never by editing the data.

IN THE PANE

Module 4 lesson 5 separates graphic text from labels and dynamic labels from annotation, then labels the cities and the counties in one typeface, Tahoma, the counties in capitals through an expression. The railroads are deliberately left unlabelled, and only one layer gets a halo.

Map elements and the layout

A finished map includes more than the map body. A **layout** is a fixed-size page; a map frame places the map on it, and the other elements give the reader the information needed to interpret the map:

- **Title:** what the map shows, usually without the word "map".
- **Legend:** how to read the symbols; its symbols must match the map exactly.
- **Scale:** a bar, verbal scale or representative fraction; required on large-scale maps and any map meant for measuring.

- **North arrow:** required when north is not up; placed together with the scale bar.
- **Source and date** of the data, the cartographer or agency, the date made, and the projection when it matters.
- Optional: a **neatline** enclosing the set, a locator or inset map, explanatory text.

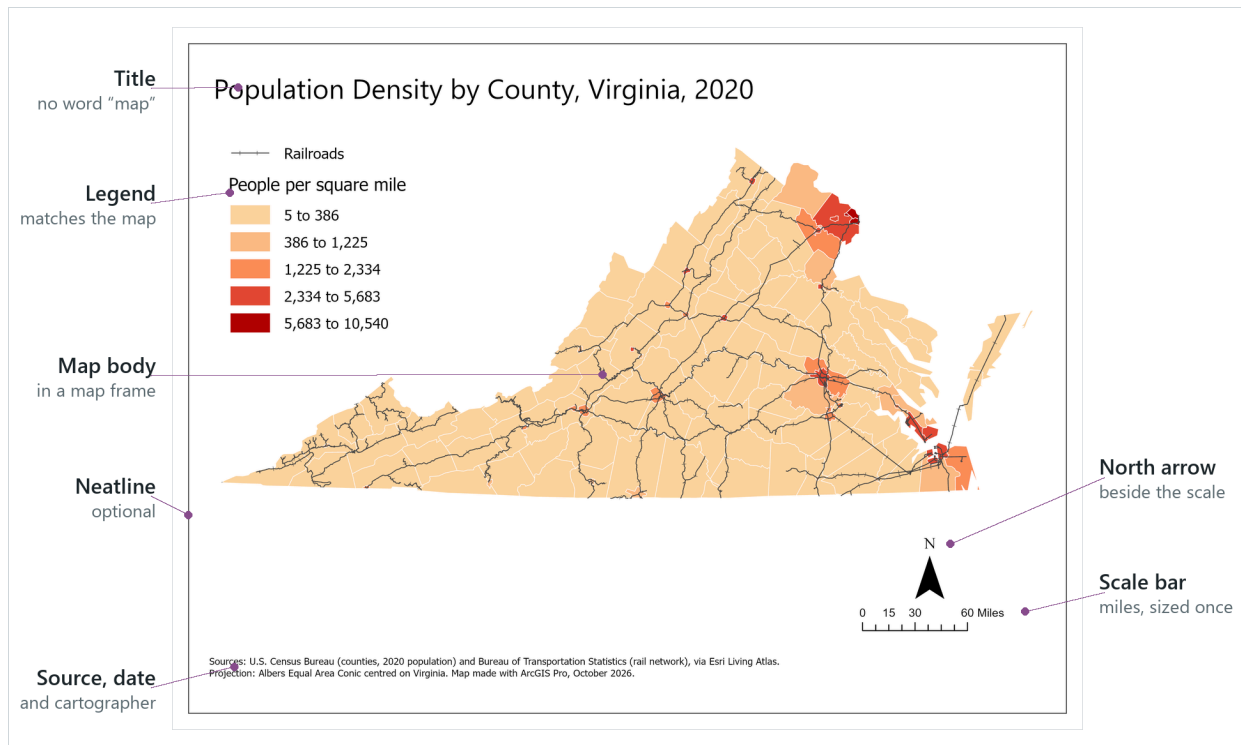


Figure 4.9 A letter-size landscape layout of the density map with every element named around it. In Pro the legend is dynamic: if you rename a layer or change a class, the legend updates.

Counties: USA Counties Generalized Boundaries, ArcGIS Living Atlas of the World (Sources: Esri; U.S. Census Bureau). Railroads: North American Rail Network Lines, Bureau of Transportation Statistics (public domain). Colours: ColorBrewer OrRd, Cynthia Brewer (Apache 2.0)

Legend entries follow the order of the Contents pane, so points, lines and polygons listed in that order give a legend in that order. The scale bar is sized once, in units the reader uses (miles, in Virginia).

IN THE PANE

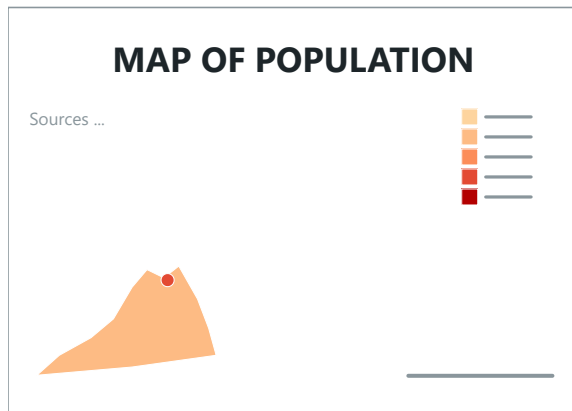
Module 4 lesson 6 builds a layout one element at a time, on a portrait or landscape page chosen from the shape of your state: the map frame, a bold title in the same typeface as the labels, a legend, a scale bar in miles with the north arrow beside it, and a source line, then exports it as an image.

Design principles

Layout design follows a few general practices. Choose the page orientation to fit the area: landscape for a state wider than it is tall, such as Virginia or Tennessee, portrait for one taller than wide, such as Idaho or California. Make the map body as large as the page allows. Put the focal element near the **optical centre**, a little above the geometric centre, where the eye tends to look first. **Balance** the elements so that no area is empty or crowded, and **align** their edges. Build a **visual hierarchy**

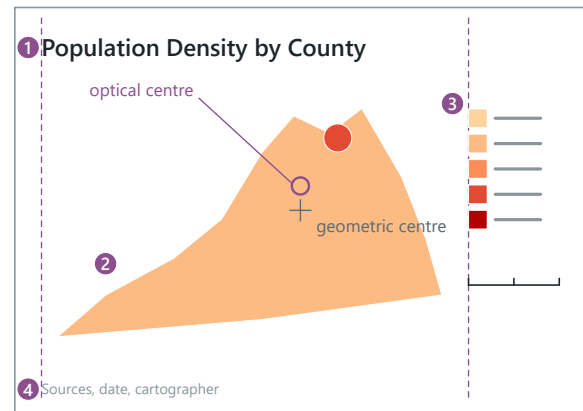
so that the most important elements are the most prominent: the title, the map and the legend, then the fine print.

Unbalanced



- small map in a corner
- empty middle, crowded edges
- title too big, says "map"
- nothing lines up

Designed



- 1 title 2 map body 3 legend 4 fine print:
the visual hierarchy; dashed lines = shared edges

Figure 4.10 One page laid out two ways. The unbalanced page leaves its middle empty, uses an oversized title and aligns none of its elements; the designed page gives the map most of the space, places the focus near the optical centre and aligns the elements along shared edges, numbered in order of importance.

Colours: ColorBrewer OrRd, Cynthia Brewer (Apache 2.0)

Figure-ground is the contrast that lets the reader separate the subject (the figure) from its background (the ground). It comes from colour (hue, value, saturation, warm against cool), size, texture, and effects such as a glow. In practice this usually means a muted basemap, such as a light grey canvas, so that the coloured counties stand out.

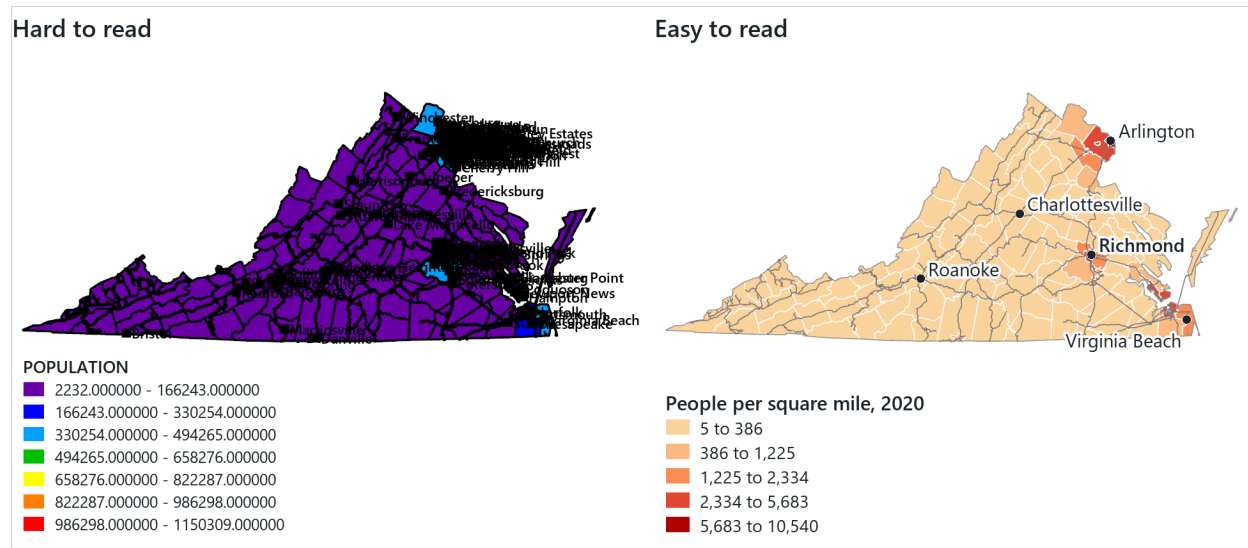


Figure 4.11 The same state mapped badly and well. Left: a count, equal interval, seven rainbow classes, heavy black borders, every place named and a legend of raw six-decimal numbers. Right: a rate, natural breaks, one hue from light to dark, white borders, five names and a legend in words.

Counties: USA Counties Generalized Boundaries; places: USA Major Cities (ArcGIS Living Atlas of the World; Sources: Esri; U.S. Census Bureau). Railroads: North American Rail Network Lines, Bureau of Transportation Statistics (public domain). Colours: ColorBrewer OrRd, Cynthia Brewer (Apache 2.0)

An ugly map with crude colours and careless lettering may be as accurate as a beautiful one, but readers are less likely to trust it; the geographer John K. Wright made the point in 1944. Design therefore affects whether a correct map is also convincing. Some aspects of design can be checked by software (text size, contrast between classes, elements that overlap or run off the page), but a check that finds no problems does not show that a map is well designed; the rest has to be checked by looking at the map.

IN THE PANE

Module 4 lesson 7 covers the design habits and figure-ground, switches to a quiet Light Gray Canvas basemap, and runs Waldo's design check on your layout, fixing what it finds until no high-priority finding remains, before exporting a second version.

TRY IT

With your Module 4 layout open, ask Waldo in Ask mode: which classification method is my county layer using, and how many counties are in each class? Compare the answer with the histogram, and decide whether the method suits your data.

Key terms

cartography	the art and science of designing maps.
generalization	selecting, simplifying, classifying and symbolizing geographic information so it can be shown at a given scale and for a given purpose.
reference map	a map of where things are, with many features at roughly equal weight, for many purposes.
thematic map	a map of one subject for one purpose, with everything else as context.
qualitative attribute	an attribute that tells kinds of things apart: a category, a name, a code.
quantitative attribute	an attribute that measures an amount: a count, an area, a rate, a percentage.
visual variable	a graphic property of a symbol, such as size, weight, shape, orientation, texture, hue, value or saturation, that the eye reads as a difference.
hue / saturation / value	which colour; how pure versus greyed; how light or dark.
qualitative / sequential / diverging scheme	distinct hues for categories; an ordered light-to-dark run for amounts; two hues meeting at a pale middle for amounts with a meaningful midpoint.
choropleth map	a map of areas shaded by class of a rate or proportion (Graduated Colors in Pro).
graduated / proportional symbols	point symbols sized by class; point symbols sized in exact proportion to each value.
normalization	dividing a count by another field, such as area or population, to make a rate.
modifiable areal unit problem	the way a mapped pattern can change when the same data are summarized over differently drawn areas.
ecological fallacy	the mistake of assuming that what is true of an area on average is true of the individuals in it.
classification	grouping features into a few classes by attribute value, with a stated method.
histogram	a chart of how many features have each value.
skewed distribution	values piled at one end with a long tail at the other.

natural breaks (Jenks)	a classification method that places breaks where the data thin out, keeping each class's values alike.
quantile	a classification method that puts the same number of features in every class.
equal interval	a classification method that divides the range into classes of equal width.
legibility / harmony	whether text can be read; one or two typefaces used consistently.
dynamic label / annotation	text placed automatically from a field; text placed and stored piece by piece.
halo	a light outline around lettering that separates it from a busy background.
layout	a fixed-size page holding map frames and map elements.
neatline	an optional line enclosing the map or the whole layout.
optical centre	the point a little above the geometric centre of a page where the eye first lands.
visual hierarchy	design that makes the most important elements the most prominent.
figure-ground	the contrast that separates a map's subject from its background.

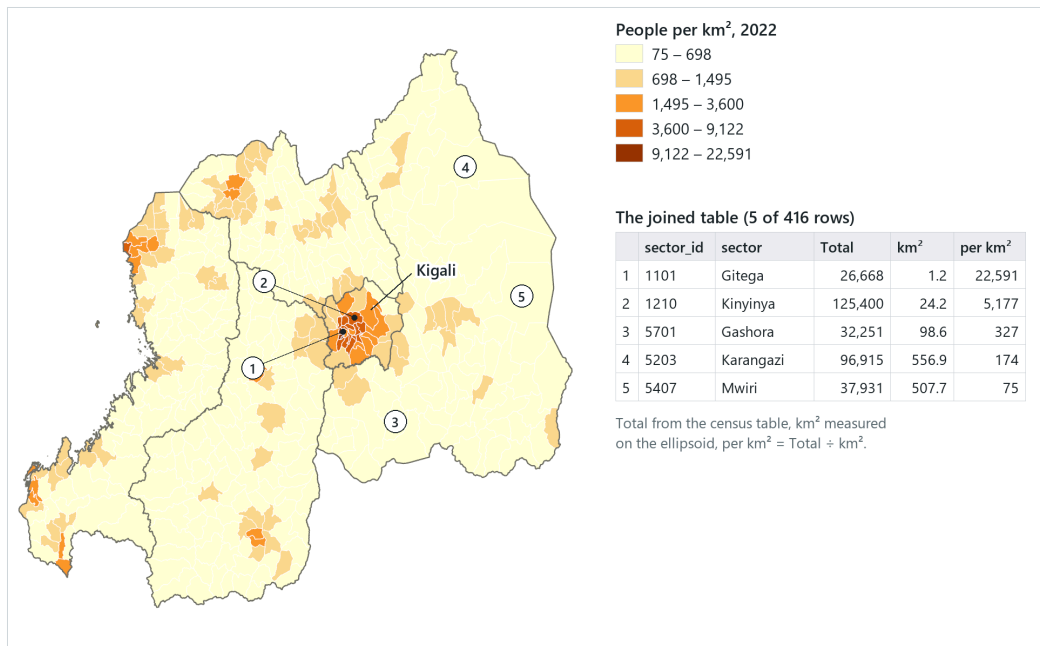
Further reading

- Brewer, C. A., Harrower, M., & The Pennsylvania State University. *ColorBrewer 2.0*: colour advice for maps. <https://colorbrewer2.org/>
- Ingram, U., & Rhodes, J. *Introduction to Cartography (KSU)*. GALILEO Open Learning Materials. <https://oer.galileo.usg.edu/geo-textbooks/3>
- Schlosser, S. (2024). *Making Effective Maps: Cartographic Visualization for GIS*. University of Colorado Boulder. <https://colorado.pressbooks.pub/makingmaps/>
- Axis Maps. *Cartography Guide*. <https://www.axismaps.com/guide>
- DiBiase, D., et al. *The Nature of Geographic Information*. Penn State. <https://www.e-education.psu.edu/natureofgeoinfo/>
- Monmonier, M. (1996). *How to Lie with Maps* (2nd ed.). University of Chicago Press; 3rd ed. 2018.
- Slocum, T. A., McMaster, R. B., Kessler, F. C., & Howard, H. H. *Thematic Cartography and Geovisualization*. Pearson / CRC Press.
- U.S. Census Bureau. *Cartographic Boundary Files*: the generalized boundaries behind this chapter's generalization figure. <https://www.census.gov/geographies/mapping-files/time-series/geo/cartographic-boundary.html>
- Esri. *ArcGIS Pro documentation*: the data classification methods page. <https://pro.arcgis.com/>
- The course's data sources: every layer the lessons and this chapter's maps use, with its publisher and licence.

CHAPTER 5

Attribute data

Accompanies Module 5 · Attribute Data Management



Sectors: Sector Boundary 2022; population: population_count (2022 census). Source: National Institute of Statistics of Rwanda (NISR), ArcGIS Online. Colours: ColorBrewer YlOrBr, Cynthia Brewer (Apache 2.0)

WHY THIS MATTERS

Every feature on a map is also a row in a table, and most of the information a GIS holds about a place is stored in that row. The map above could not be drawn from the boundaries alone: the population figures came from a census table with no geometry, joined to the sectors through a shared code and divided by an area measured on the Earth. This chapter covers the table behind every layer: the values a field can hold, how to fill a field, and how to join two tables safely.

YOU SHOULD BE ABLE TO EXPLAIN

- why a code that looks like a number is stored as text, and why a population needs a Long
- why a one-to-many join repeats or drops rows, and what a good key must satisfy
- why a null is not a zero, and why an area in square degrees is no use as a measurement

IN THIS CHAPTER

Tables: records and fields
Three kinds of table
What a field can hold
Making and filling fields
Looking at a column

Joins: destination, source and key
Cardinality
Tables from outside the GIS
A rate from joined data

Tables: records and fields

A table consists of rows and columns. In a GIS attribute table each row is a **record**, and each record is one **feature**: a single point, line or polygon. Each column is a **field**, and each field holds one **attribute**, one thing known about every feature: a name, a code, an area, a population. The terms come in two sets: row, record and feature refer to one direction of the table, and column, field and attribute to the other. All six are in common use.

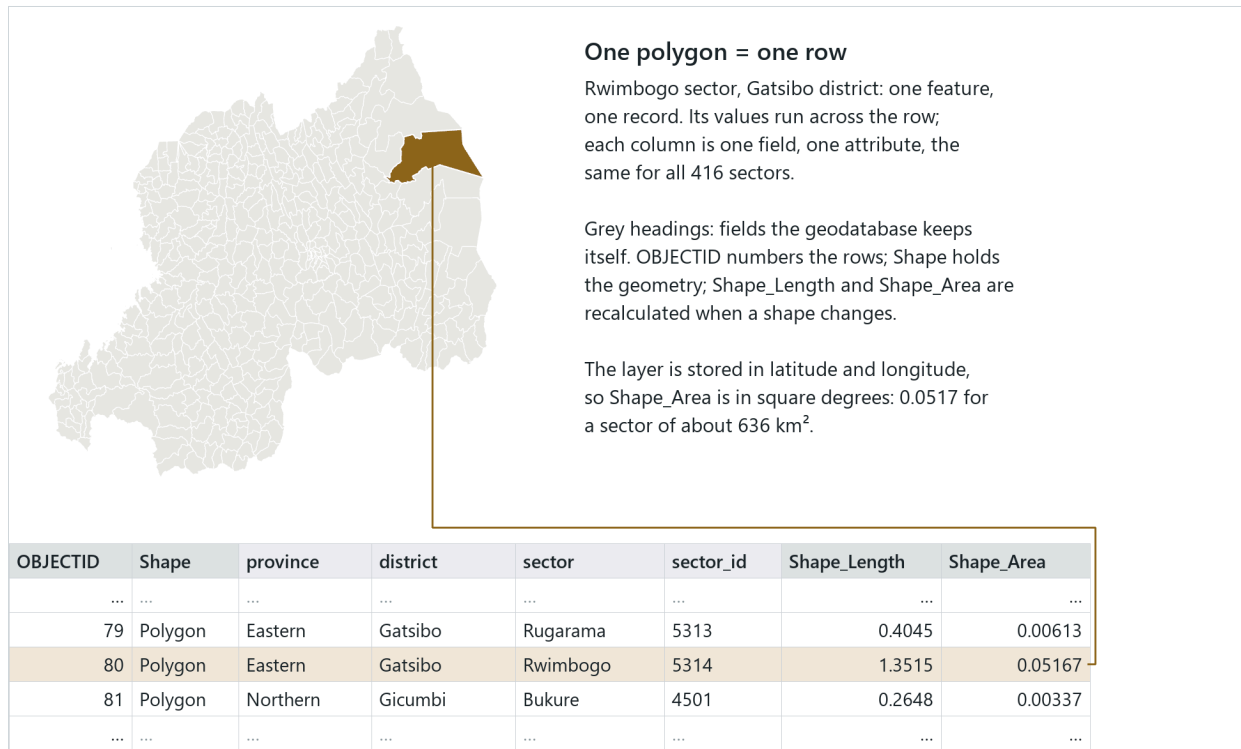


Figure 5.1 One polygon corresponds to one row. Rwimbogo sector in eastern Rwanda is one feature of the 416 in the national sector layer and one record in its attribute table. The grey headings are fields that the geodatabase maintains automatically.

Sectors: Sector Boundary 2022. Source: National Institute of Statistics of Rwanda (NISR), ArcGIS Online

Rwanda's statistics office publishes its **sectors**, the country's third administrative level, as one layer: 416 polygons under 30 districts and 5 provinces. Its table has 416 records. The record count at the bottom of any table view is the first thing to check, to confirm that you are looking at the data you expect.

Some fields are filled automatically and never typed. A feature class in a geodatabase keeps an **OBJECTID** that numbers every row, a **Shape** field that holds the geometry, and **Shape_Length** and **Shape_Area**, recalculated whenever a shape changes. A shapefile keeps only a row number (FID). Those two measurements are in the units of the layer's coordinate system. The sector layer is stored in latitude and longitude (WGS 1984), so its Shape_Area is in **square degrees**: 0.0517 for Rwimbogo, a sector of about 636 km². A square degree is not a practical unit of area, and its size changes with latitude; this is the problem described in Chapter 3, appearing here in a table.

Three kinds of table

Not every table belongs to a layer. An **attribute table** is tied to a layer's features and opens from it. A **standalone table** has no geometry at all: a census table, a spreadsheet, a list of readings from a monitoring station. It is listed in the project beside the layers but is not drawn on the map. Standalone tables come from many sources: a geodatabase table, an Excel workbook, a comma-separated text file, a dBASE file.

Attribute table

one row per feature, tied to the layer

OBJECTID	Shape	sector	sector_id
...	...	...	...
79	Polygon	Rugarama	5313
80	Polygon	Rwimbogo	5314
81	Polygon	Bukure	4501
...	...	...	...

sectors: 416 rows, 416 polygons.
Shape holds each row's geometry.

Standalone table

rows with no geometry: not drawn

sector_id	period	Total
5701	2012	22,920
5701	2022	32,251
5701	2002	15,248
5702	2012	24,695
...	...	...

population_count: 1,248 rows,
three censuses for each sector.

Raster attribute table

one row per cell VALUE, with a COUNT

1	1	2	2	OID	VALUE	COUNT
0	1	2	4			
1	3	2	5			
3	3	3	5			
1	3	4	2			

16 cells, 4 values, 4 rows.
No Shape column: the cells are the geometry.
VALUE is always a number.

kept by the software: you never type these (a geodatabase also keeps Shape_Length and Shape_Area)

Figure 5.2 Three kinds of table. An attribute table has one row per feature and a Shape column; a standalone table has rows but no geometry; a raster attribute table has one row per cell value, with a count of the cells that hold it.

Values: National Institute of Statistics of Rwanda (NISR), Sector Boundary 2022 and population_count

The census table used in Module 5 is a standalone table published by the same office: 1,248 rows, the 416 sectors counted three times, in 2002, 2012 and 2022.

A raster can also have a table, but it is built differently. It does not have one row per cell, which for a satellite scene would mean millions of rows. A **raster attribute table** has one row for each distinct cell **VALUE** and a **COUNT** of the cells that hold it. There is no Shape field, because the cells are the geometry. Only a raster of whole numbers, such as land-cover class codes, has such a table. In a raster of decimals, such as elevation, hardly any two cells have the same value, so it has statistics instead. The table may carry extra columns that describe each value, such as a class name, but the value is the data and the description is attached to it. Chapter 8 works with real raster attribute tables.

What a field can hold

Before a field is created, two questions need to be answered about it. The first is what kind of value it holds, its **level of measurement**:

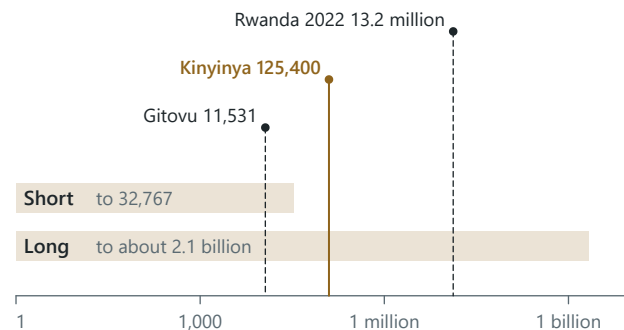
Level	What the values are	Example in the sector data	Store as
Nominal	names or categories, with no order	sector name; the code 5314	Text
Ordinal	categories with a meaningful order	a density class, 1 (lowest) to 5	Text or a small whole number (Short)
Interval	quantities whose zero is arbitrary	a date; a year	Date, or a whole number
Ratio	quantities with a true zero: counts, measures, rates	population; area in km ² ; people per km ²	Long for counts; Double for measures and rates

Interval and ratio values are both quantities, and many texts treat them together; the difference is only whether twice the number means twice as much. Arithmetic applies only to quantities. A code that looks like a number is still a name: sector 5314 is never added to sector 5313. Stored as text, the code also keeps any leading zero that a number field would drop.

The second question is how the value is stored, the field's **storage type**. The types used in this course are:

- **Short:** whole numbers from $-32,768$ to $32,767$.
- **Long:** whole numbers up to about 2.1 billion.
- **Float:** decimals, with six or seven significant digits.
- **Double:** decimals, with about fifteen significant digits.
- **Text:** characters, up to the field's **length** (4 for the sector code, 50 for the names).
- **Date:** dates, times, or both, so that "before 2020" works as a query.
- **Blob:** binary data such as an image or a document, never typed by hand.

Whole numbers: how big?



Kinyinya's population overflows a Short field; so do 145 of Rwanda's 416 sectors. A population count is a Long.

Decimals: how many digits?

Rwimbogo's Shape_Area, 0.0516705435380795... square degrees

Float 0.0 5 1 6 7 0 5 4
32 bits: six or seven significant digits

Double 0.0 5 1 6 7 0 5 4 3 5 3 8 0 7 9 5
64 bits: about fifteen

Both hold decimals. Double is the safe default for anything a calculation produces: areas, rates, densities.

Figure 5.3 Storage types compared by size and by number of digits. Short and Long differ in how large a whole number they can hold: Kinyinya sector's 125,400 people are too many for a Short. Float and Double differ in how many digits of a decimal they keep.

Values: National Institute of Statistics of Rwanda (NISR), Sector Boundary 2022 and population_count

The number types come in pairs. Short and Long both hold whole numbers and differ in size: a Short holds values only up to 32,767, and in the 2022 census 145 of Rwanda's 416 sectors were larger than that, so a population count is a Long. Float and Double both hold decimals, and differ in precision. A Double is the safe choice for anything a calculation will produce, an area or a rate, because you cannot know in advance how many digits it will need.

field	example value	level of measurement	storage type	why
sector	Rwimbogo	nominal	Text (50)	a name: sorted and matched, never added up
sector_id	5314	nominal (a code)	Text (4)	a code: text keeps it exact, and keeps any leading zero
Total	49,975	ratio (a count)	Long	whole numbers, and more than a Short's 32,767
area_km2	635.79	ratio (a measure)	Double	decimals; a calculation decides how many digits
pct_female	51.6	ratio (a percentage)	Float	decimals, but few digits: the smallest that fits
density_class	5 (of 5)	ordinal (a rank)	Short	a small whole number whose order matters
created_date	2024-09-03	interval (a date)	Date	sorts and queries as a date, not as text

Names and codes go in Text; counts in Long (Short if small); anything measured or calculated in Double (Float if it stays short); dates in Date.

Figure 5.4 Seven fields from the sector and census tables, each with an example value, its level of measurement and the storage type it belongs in.

Values: National Institute of Statistics of Rwanda (NISR), Sector Boundary 2022 and population_count

RULE OF THUMB

Choose the smallest storage type that fits, and choose it once: the type of a field cannot be changed after it holds data. Names and codes go in Text, counts in Long, anything measured or calculated in Double.

IN THE PANE

Module 5 lesson 1 adds the statistics office's 416 sectors by their web address, counts the rows, and exports a copy into the project geodatabase so the automatic fields appear, square degrees included; the web layer is then removed from the map, leaving the copy. In the Fields view, which lists the table's fields and their properties, you read the types of the code and the area fields, then answer two ungraded questions on which type a population and a percentage need. A final picture compares the vector table with a raster's.

Making and filling fields

A field has a definition as well as values: its name, its type, its length, whether it may be empty, a default. A **field name** follows rules the software enforces: letters, numbers and underscores only; a letter first; no spaces or punctuation; at most 64 characters in a geodatabase and 10 in a shapefile. Shorter is better, because the name is what you type in every expression and query. The readable version, with spaces and units, goes in the field's **alias**: area_km2, shown as "Area (km²)".

A new field starts empty. There are two ways to fill it. **Calculate Field** fills it from an expression over each row's other fields: a label built from two names, a rate from two numbers, a category from a threshold. **Calculate Geometry** fills it from each feature's shape: area, length, perimeter or the coordinates of a centroid.

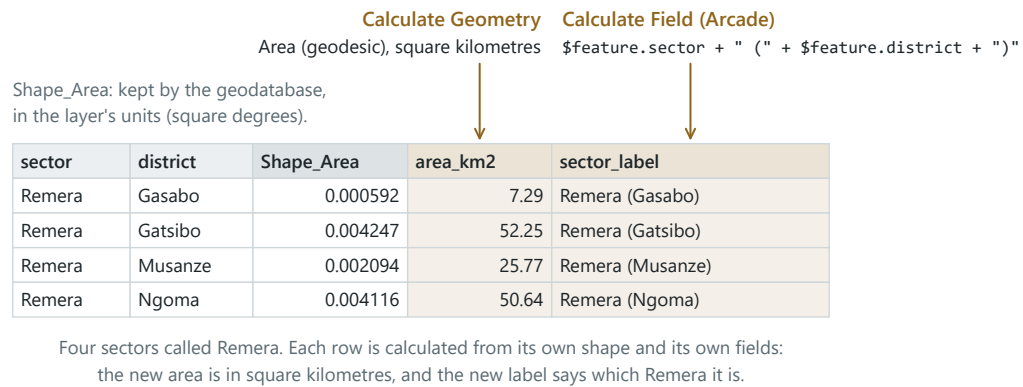


Figure 5.5 The four sectors named Remera, before and after two calculations. Calculate Geometry fills area_km2 with each sector's area measured on the ellipsoid; Calculate Field fills sector_label from each row's own sector and district.

Values: National Institute of Statistics of Rwanda (NISR), Sector Boundary 2022 and population_count

Module 5 uses both, for reasons that come from the data. Thirty sector names repeat: there are four Remeras, and three each of Kageyo, Murambi, Muganza, Karama and Ngoma. A label that combines the sector with its district, "Remera (Gatsibo)", identifies which one is meant. Because the layer is stored in latitude and longitude, an ordinary area calculation would again give square degrees. A **geodesic** area is measured on the ellipsoid (Chapter 3) and is accurate anywhere without first projecting the data. Added up, the 416 geodesic areas come to 24,306 km²; Rwanda's official area, 26,338 km², includes the open water of Lake Kivu, which lies in no sector.

Two cautions apply to every calculation. First, it takes effect immediately and cannot be undone unless an edit session is open. Second, when rows are selected, only the selected rows are calculated. Clicking a row in a table to read it selects it, so a quick look can leave a selection behind; clear it before the next calculation.

RULE OF THUMB

Calculate into a new, empty field, with nothing selected, and check the result against something you know: a sum of areas against the country's area, a density against a place you know is crowded. Check the expression before you run it.

Looking at a column

Look at a column before you map it, join it or rely on it. The count of values, the number that are empty, the smallest and largest, the mean, the median and the spread describe a field in one line, and a histogram shows its shape. Pro's **Data Engineering view** shows all of these for a field at once and writes nothing to the data; it is used only for inspection.

An empty cell, called a **null**, is different from a zero. A sector with no recorded cases, one whose cases were never reported and one whose census row failed to join look alike on a map if they fall in one class, although they mean three different things. A rate with a null or zero denominator has no value, and summaries such as a mean skip nulls, so check the number of empty values first.

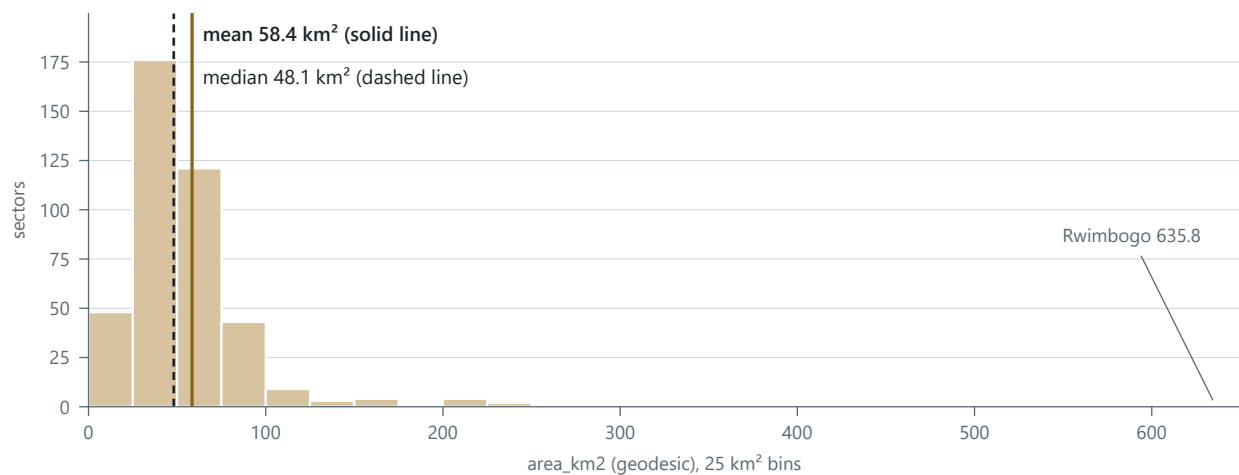


Figure 5.6 The 416 sectors' areas in 25 km² bins. Most sectors are under 100 km²; a few large eastern ones, the biggest Rwimbogo at 635.8 km², pull the mean above the median.

Values: National Institute of Statistics of Rwanda (NISR), Sector Boundary 2022 and population_count

The sector areas illustrate why both averages are useful. The mean is 58.4 km² and the median 48.1 km²: half the sectors are smaller than 48.1, but a small number of large sectors on the eastern savanna raise the mean. The smallest, Rwezamenyo in Kigali, covers 1.02 km². A distribution this skewed indicates that a map of it will need a careful choice of classes (Chapter 4).

IN THE PANE

Module 5 lesson 2 gives the naming rules, adds a Double field and fills it with geodesic area in square kilometres, checked by the national total, then finds the largest sector by sorting. A text field is filled by an Arcade expression that distinguishes the four Remeras. The lesson ends in the Data Engineering view, reading the mean area of the 416 sectors.

Joins: destination, source and key

A **join** attaches the columns of one table to the rows of another. The table you add to is the **destination**, usually a layer's attribute table; the table whose columns are added is the **source**, which may be another attribute table, a standalone table, or a spreadsheet. Rows are matched through a **key field**: a field present in both tables, holding the same kind of value, in which each value identifies one row.

Destination			Source			Joined				
sectors: the features			population_by_sector, period = 2022			same rows, the source's columns added				
sector_id	sector	district	sector_id	period	Total	sector_id	sector	district	period	Total
5701	Gashora	Bugesera	5314	2022	49,975	5701	Gashora	Bugesera	2022	32,251
1210	Kinyinya	Gasabo	1101	2022	26,668	1210	Kinyinya	Gasabo	2022	125,400
5314	Rwimbogo	Gatsibo	5701	2022	32,251	5314	Rwimbogo	Gatsibo	2022	49,975
1101	Gitega	Nyarugenge	5407	2022	37,931	1101	Gitega	Nyarugenge	2022	26,668
5407	Mwiri	Kayonza	1210	2022	125,400	5407	Mwiri	Kayonza	2022	37,931

The key, `sector_id`, holds the same codes on both sides and each code once, so every sector finds exactly one census row, in whatever order the source stores them. Five rows in, five rows out.

Figure 5.7 A join on `sector_id`. Each sector is matched to the one census row with the same code, whatever that row's position in the source; the destination keeps its five rows and gains the source's columns.

Values: National Institute of Statistics of Rwanda (NISR), Sector Boundary 2022 and `population_count`

A join also has a direction. It is made from the destination, the layer you are adding to; if it is started from the source table instead, it attaches the layer's columns to the table, and the layer receives no new columns.

Codes and identifiers make good keys, but names rarely do. The census table's province column spells the east "Estern Province" in all 285 of its eastern rows, so a join on the province name would fail to match the whole east without any warning. Joined columns carry the source table's name as a prefix, such as `population_by_sector.Total`, so that you can tell where each came from.

A join is a property of the layer in the current project. It is lost if the layer is removed or the source table moves, and it changes nothing on disk. To keep the combined data, export the joined layer as a new feature class: the export writes every column, joined ones included, into a dataset of its own. (Like every export, it copies only the selected records when there is a selection.)

Cardinality

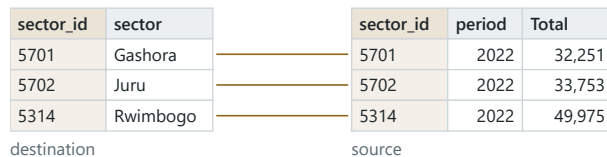
The central question for every join is how many rows one key value matches on each side; this is the join's **cardinality**.

- **One-to-one (1:1)**: one sector, one census row. After the census table is filtered to 2022, every one of the 416 codes appears once on each side.
- **Many-to-one (m:1)**: many sectors share one row of the source, as the sectors of a district would share a table of district facts. Each sector is matched to one row, and the source row is copied as often as needed. This is also a valid join.
- **One-to-many (1:m)**: one sector, several census rows, one per census. A table cell cannot hold three rows, so the software has to handle the extra rows in one of two ways, and neither gives a map of 2022. One-to-many is a legitimate structure for data (one sector, three census years, when the rows wanted are sector-years), but it is the wrong structure for a join that should give each sector one row. A *one-to-first* join keeps whichever matching row it finds first. Joined that way, the unfiltered census gives 151 sectors their 2012 figure, 150 their 2022 figure and 115 their 2002 figure, which produces a map that mixes three censuses but looks like a map of one. A *one-to-many* join instead repeats each sector once per match, so 416 sectors become 1,248 rows; ArcGIS Pro makes this kind of join when both tables are in the same geodatabase. There are

three ways to resolve the problem: **filter** the source to one row per key (a definition query of period = 2022 on the table), **aggregate** it first (sum or average the rows per key, then join), or **relate** the tables, which links them without merging, although a related table cannot be used for symbology, labels or queries.

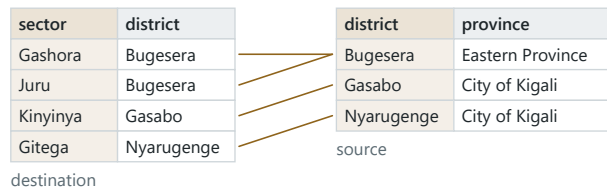
- **Many-to-many (m:n)**: several rows on both sides share a value, as the four Remeras in the sectors match the twelve Remera rows of the census. For a join meant to give each sector one row of facts, this key is unsuitable because it does not identify a single sector.

One-to-one (1:1)



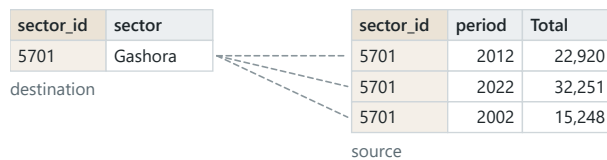
a join: one row each side

Many-to-one (m:1)

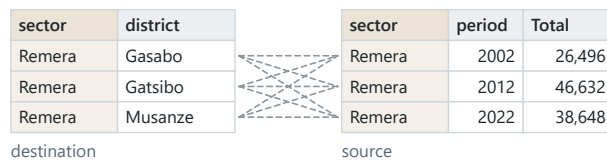


a join: the district row is copied to each sector

One-to-many (1:m)



Many-to-many (m:n)



filter to one row per key, aggregate, or relate

Unfiltered, a join keeps the first row it meets (2012 here), or repeats Gashora once per row: 416 sectors become 1,248.

the wrong key: use sector_id, not the name

Remera is four sectors in four districts; by name, every Remera matches every Remera row. No join is defined.

Count the rows one key value picks out on each side. A key field holds the same kind of value on both sides and is unique on the source side. The 1:m panel is Module 5's census table as it arrives: a definition query, period = 2022, turns it into the 1:1 panel.

Figure 5.8 Cardinality in the Module 5 tables. One-to-one and many-to-one are the cardinalities a join to a layer is designed for; the census table as published is one-to-many, three rows per sector; joining by name is many-to-many, because the same name belongs to several sectors.

Values: National Institute of Statistics of Rwanda (NISR), Sector Boundary 2022 and population_count

A good key therefore holds the same kind of value on both sides and is unique on the source side. The sector code passes both tests once the census is filtered; the sector name fails the second.

RULE OF THUMB

For a join that gives each feature one row of facts: one-to-one and many-to-one are joins; one-to-many needs a filter, an aggregate or a relate; many-to-many means the key does not identify your features. Test the key before you join: count the matches, the unmatched rows and the duplicates on each side. Export the joined layer to make the join permanent.

IN THE PANE

Module 5 lesson 3 adds the census as a table by its web address and copies it into the project. Waldo tests two candidate keys against the data: the sector code (one-to-many, three censuses per sector) and the sector name (many-to-many, the Remeras). A definition query on the table leaves only 2022, the same test then reports a clean match, 416 for 416, and the join is made from the sectors layer, read on the Joins tab, and exported as a new feature class under a name of its own, which replaces the joined layer on the map.

Tables from outside the GIS

Many tables are first created in a spreadsheet, typed for people to read rather than for software to parse. A GIS reads the first row as field names and each column as one field, so a table needs a little preparation first: the naming rules apply to the header row, and every column must hold one kind of value.

As typed

	A	B	C	D	E
1	Census 2022 by sector				
2	Sector ID	Sector name	2022 Population	% Female	People/km ²
3	5701	Gashora	32,251	50.3	327.2
4	5702	Juru	33 753	51.3	#DIV/0!
5	5314	Rwimbogo	49,975	51.6	78.6

Ready for a GIS

	A	B	C	D	E
1	sector_id	sector	pop2022	pct_female	density
2	5701	Gashora	32251	50.3	327.2
3	5702	Juru	33753	51.3	
4	5314	Rwimbogo	49975	51.6	78.6

- ❶ The first row holds the field names: no title row or blank row above them.
- ❷ Names start with a letter and use only letters, numbers and underscores: no spaces or symbols such as % / () - or a full stop, and at most 64 characters (10 if the table becomes a shapefile).
- ❸ One kind of value per column, formatted the same way all the way down: numbers as numbers.
- ❹ No error values such as #DIV/0!: leave the cell empty.
- ❺ Codes stay text, so a code keeps its exact form and any leading zero.

Figure 5.9 A spreadsheet as typed and the same table ready for a GIS. The numbered markers match the checklist beneath.
Values: National Institute of Statistics of Rwanda (NISR), Sector Boundary 2022 and population_count

The figure above shows the full checklist: the field names in the first row, with nothing above them; names that follow the field-naming rules; one kind of value per column, formatted the same way all the way down; no error values such as #DIV/0!; and codes kept as text. A comma-separated (.csv) or other delimited text file follows the same rules: its first line is the header.

A rate from joined data

A join is most useful when it lets you compute something that neither table contained on its own. The sector layer has areas but no population figures, and the census has population figures but no areas. Once they are joined, they give **population density**, the 2022 total divided by the geodesic area, a new Double field filled by Calculate Field over a joined column.

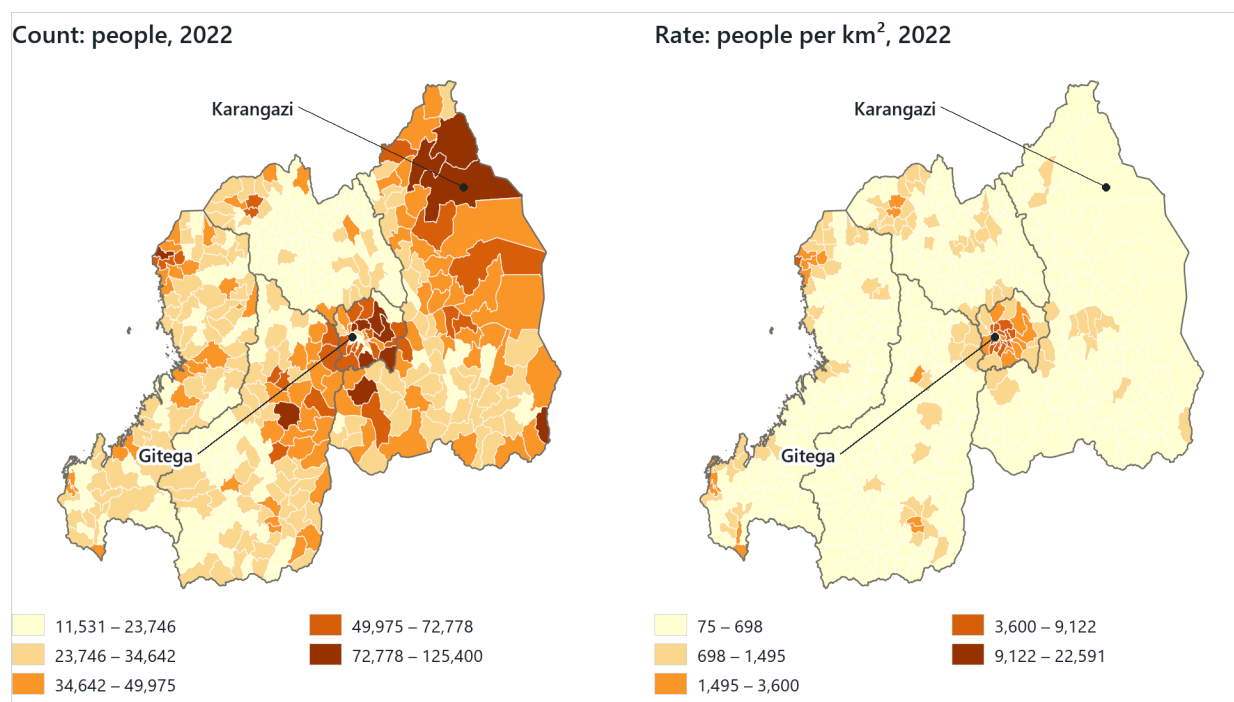


Figure 5.10 The same joined data mapped as a count and as a rate. By count, big rural Karangazi is among the darkest sectors; by rate it is among the palest, and the few square kilometres of central Kigali stand out.

Sectors: Sector Boundary 2022; population: population_count (2022 census). Source: National Institute of Statistics of Rwanda (NISR), ArcGIS Online. Colours: ColorBrewer YlOrBr, Cynthia Brewer (Apache 2.0)

The densities run from 75 people per km² in Mwiri, on the edge of Akagera National Park (37,931 people on 507.7 km²), to 22,591 in Gitega in central Kigali (26,668 people on 1.18 km²): a three-hundred-fold range in one small country that has 545 people per km² across all its sectors together. The count map gives a different and misleading impression. Karangazi, in the north-east, is the third most populous sector in the country with 96,915 people, and one of the emptiest, 174 per km² over 556.9 km². Mapped by count, its size alone makes it look important. The map on the opening page shows the rate, with five of the sectors named in the table beside it.

READING THE RESULT

- The question: where in Rwanda do people live most densely, sector by sector?
- The representation: the 2022 census total joined on sector_id, divided by geodesic area.
- What it shows: 75 people per km² in Mwiri to 22,591 in Gitega; by rate, Karangazi is among the palest sectors.
- What it cannot establish: how people are spread inside a sector; each rate assumes an even spread.

RULE OF THUMB

Map rates rather than counts on areas of different sizes (Chapter 4). A count on a choropleth mostly maps the size of the areas.

IN THE PANE

Module 5 lesson 4 adds a density field to the exported layer, fills it by dividing the joined census total by the measured area, and checks both ends of the range, Gitega and Mwiri. The field is then mapped with graduated colours in natural-breaks classes, and briefly mapped as a count, which shows Karangazi as one of the darkest sectors.

TRY IT

With the Module 5 project open, ask Waldo in Ask mode: could I join `population_by_sector` to `sectors` on the `sector name`? Waldo tests the key against both tables and reports how many values match, repeat or go unmatched. Compare the answer with the cardinality section of this chapter.

Key terms

record	one row of a table; in an attribute table, one feature.
field	one column of a table: one attribute of every record, with a name, a type and a definition.
attribute table	the table tied to a layer, with one row per feature.
standalone table	a table with no geometry, such as a census table or an imported spreadsheet.
raster attribute table	an integer raster's table: one row per distinct cell VALUE, with a COUNT of the cells that hold it.
OBJECTID	the unique number a geodatabase keeps for every row of a table.
level of measurement	whether values are nominal, ordinal, interval or ratio, which determines the arithmetic that makes sense.
nominal / ordinal	values that are names or categories / categories with a meaningful order.
interval / ratio	quantities whose zero is arbitrary / quantities with a true zero, where twice the number means twice as much.
storage type	how a field stores its values: Short, Long, Float, Double, Text, Date or Blob.
Short / Long	whole-number storage types, up to 32,767 and up to about 2.1 billion.
Float / Double	decimal storage types keeping six or seven and about fifteen significant digits.
field length	the number of characters a Text field can hold.
alias	a readable label shown in place of a field's name.
Calculate Field	a tool that fills a field from an expression over each row's other fields.
Calculate Geometry	a tool that fills a field with each feature's area, length, perimeter or coordinates.
null	an empty field value, which records that nothing is known and is not the same as zero.
join	attaching the columns of a source table to the rows of a destination table through a key field.
destination / source table	in a join, the table that is added to / the table whose columns are added.
key field	a field in both tables whose values match a row on one side to a row on the other.
cardinality	how many rows one key value matches on each side of a join: one-to-one, many-to-one, one-to-many or many-to-many.
relate	a link between two tables that keeps them separate instead of merging their columns.
population density	people per unit of area; a rate, suited to a choropleth where a count is not.

Further reading

- Olaya, V. *Introduction to GIS*, the chapters on data and on databases (records, keys and relations). <https://volaya.github.io/gis-book/en/>
- Esri. "ArcGIS field data types." ArcGIS Pro documentation. <https://pro.arcgis.com/en/pro-app/latest/help/data/geodatabases/overview/arcgis-field-data-types.htm>
- Esri. "Joins and relates" and the Add Join tool reference. ArcGIS Pro documentation. <https://pro.arcgis.com/en/pro-app/latest/help/data/tables/joins-and-relates.htm>
- National Institute of Statistics of Rwanda, geodata hub: the sector boundaries and the census tables used in this chapter. <https://geodata-nisr-rwanda.opendata.arcgis.com/>

CHAPTER 6

Exploring spatial data

Accompanies Module 6 · Spatial Data Exploration



Map: John Snow, 1855 (lithograph by C. F. Cheffins), public domain; georeferenced scan "Snow cholera map image", Esri_Tutorials, ArcGIS Online. Deaths and pumps: Cholera Cases and Water Pumps, digitised from Snow's map by Esri_Tutorials (ArcGIS Online)

WHY THIS MATTERS

A GIS is most useful when it answers questions: which places pass a test on their values, which lie near, inside or beside something else, how many or how much there is, and whether two things vary together. In 1854 John Snow asked questions of these kinds about a few streets in Soho with a pen, and the map above is his answer redrawn in a GIS. This chapter introduces a tool for each kind of question: attribute queries, spatial queries, summaries and charts.

YOU SHOULD BE ABLE TO EXPLAIN

- why a selection highlights rows while a definition query hides them, and what each does to a tool
- how the selection type makes two queries combine as AND, OR or NOT
- why a passed 5 km test or a high R^2 is a reason to look further and does not establish access or cause

IN THIS CHAPTER

Asking questions of data

John Snow's map, 1854

Attribute queries and SQL

AND, OR, NOT and the selection types

Keeping, viewing and clearing a selection

Spatial relationships

Select By Location in Rwanda

Attribute and spatial together

Summary statistics

Charts: which one answers what

Asking questions of data

Exploring data means asking it a question and looking at the answer, usually in more than one view. The map shows where the answer is, the table shows the rows and their values, and a chart shows how the values are spread or whether two of them go together. The views are linked. When you select features on the map, the same rows are highlighted in the table and the same dots in any chart made from the layer, and a selection made in a chart is shown on the map.

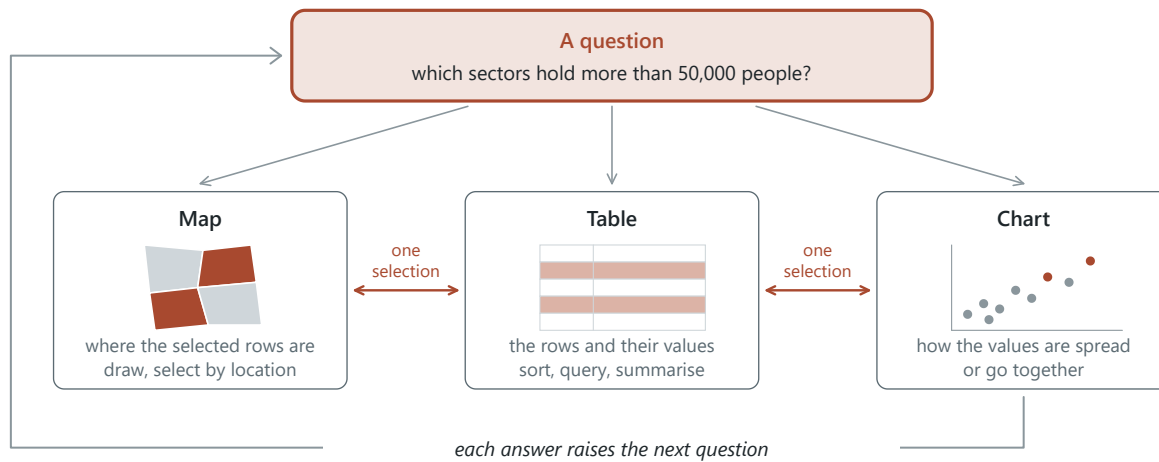


Figure 6.1 Exploration drawn as a loop. A question is examined in three linked views, a selection made in one view appears in all of them, and each answer usually raises another question.

Exploration is iterative, because the first answer is rarely the last. A map of the most populous places raises the question of whether they are also the largest, a chart can answer that question and reveal an outlier, and you then return to the map to look at the outlier. This way of working is called spatial data exploration, or exploratory spatial data analysis (ESDA). It comes before any formal analysis, and the main finding of a study is often made at this stage.

John Snow's map, 1854

On the night of 31 August 1854 cholera broke out around Broad Street, near Golden Square in Soho, London. Snow counted more than five hundred fatal attacks within 250 yards of one spot in ten days, and in under a week more than three-quarters of the people in the worst streets had fled. The accepted explanation was miasma, bad air rising from filth. Snow, a physician, believed that cholera travelled in water.

He listed the registered deaths, went to the houses, and drew each death on a street map as a short black bar stacked at the address, with the public water pumps marked. The bars were concentrated around the pump in Broad Street. The exceptions supported the same conclusion. The workhouse in Poland Street, almost surrounded by deaths, lost 5 of its 535 inmates, and it had its own well; the Broad Street brewery's seventy-odd workmen escaped, drinking the beer they were allowed and water from the brewery's own deep well. Snow put his case to the parish board on 7 September, and the pump's handle was removed the next day. By then the outbreak was already fading (his own table shows new attacks falling from 143 on 1 September to 12 on the 8th), and it was decades before the water theory was generally accepted. The map was published in 1855.

Esri's digitisation of that map holds 322 addresses and 578 deaths; Snow's table counts 616, because his map could only show deaths whose house was known. The worst address, 45 Broadwick Street (Broad Street was renamed in 1936), had 18.

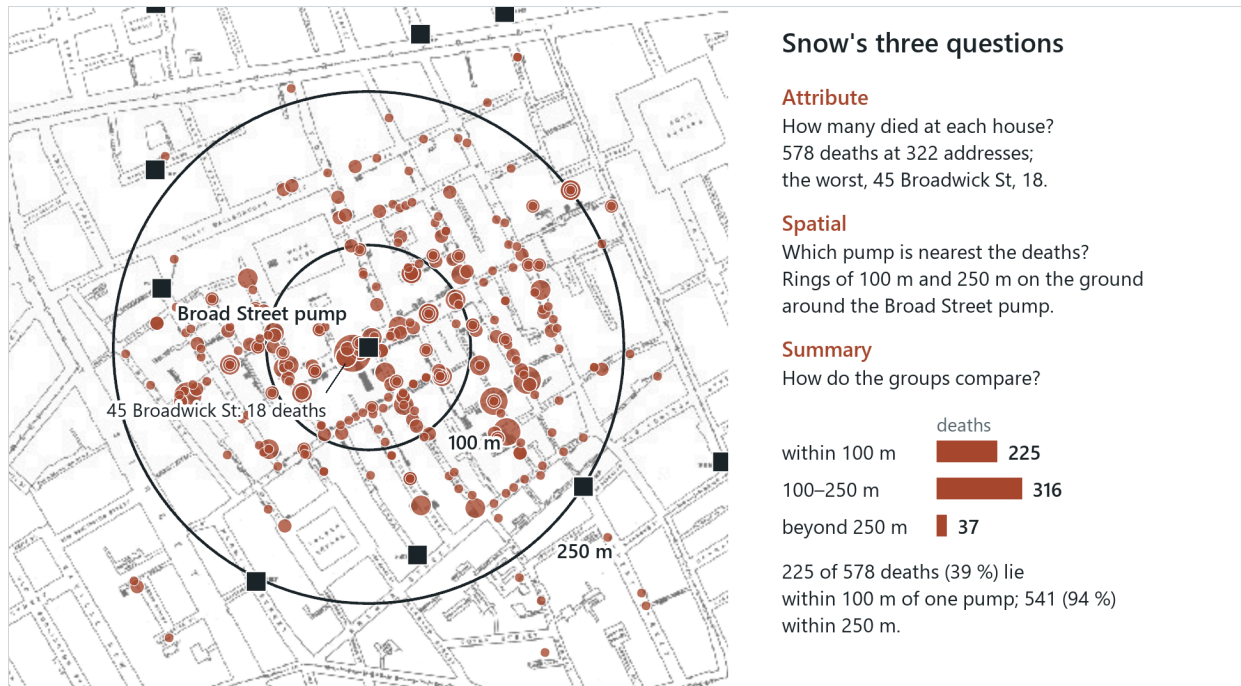


Figure 6.2 Snow's three questions applied to his own data and drawn on his map. Of the 578 deaths, 225 lie within 100 m of the Broad Street pump and 541 within 250 m.

Map: John Snow, 1855 (lithograph by C. F. Cheffins), public domain; georeferenced scan "Snow cholera map image", Esri_Tutorials, ArcGIS Online. Deaths and pumps: Cholera Cases and Water Pumps, digitised from Snow's map by Esri_Tutorials (ArcGIS Online)

The steps Snow took follow the same outline as this chapter. He asked an **attribute** question of every house: how many died here? He asked a **spatial** question: which pump is nearest these deaths? He also **summarised**, comparing the count for the houses around one pump with the count for the rest. When the digitised deaths are sized by the number of deaths, as on the opening page, the cluster around one pump is clearly visible.

IN THE PANE

Module 6 lesson 1 adds Snow's deaths, pumps and the scanned 1855 map to your own map by URL, and switches the map to Web Mercator so that the scan is drawn in the correct position. You read the sum of the Deaths column (578), sort it to find the worst address, select that row to see the pump beside it, and size the deaths with Graduated Symbols so that the cluster is visible. The London layers are removed at the end.

Attribute queries and SQL

An **attribute query** selects the records whose values pass a test: which of Rwanda's 416 sectors have more than 50,000 people? The language behind it is **SQL** (Structured Query Language), the query language of relational databases. Its syntax is strict, and every query tool writes an SQL statement, however simple its dialog looks.

The full statement is `SELECT * FROM sectors_census WHERE Total > 50000`: select every field (the `*`) from this table where the test is true. A query builder fills in the first two parts itself and writes only the **WHERE clause**. The clause has three pieces: a field name (the name, not its alias), a comparison (`=`, `<>`, `>`, `<`, `>=`, `<=`) and a value. Numbers are written without quotes, and text goes in single quotes, as in `province = 'City of Kigali'`. On the 2022 census, `Total > 50000` selects 42 sectors.

<code>SELECT *</code> <small>every field</small>			<code>FROM sectors_census</code> <small>this table</small>	<code>WHERE Total > 50000</code> <small>this test</small>	<small>the builder writes only the WHERE clause</small>
In the builder			The SQL it writes		sectors
Total	is greater than	50000	→	<code>Total > 50000</code> <small>numbers stand bare</small>	42
district	includes the value(s)	☒ Gasabo ☒ Kicukiro ☒ Nyarugenge	→	<code>district IN ('Gasabo', 'Kicukiro', 'Nyarugenge')</code> <small>a list; text in single quotes</small>	35
sector	begins with	Nya	→	<code>sector LIKE 'Nya%'</code> <small>a pattern: % means any characters</small>	31

Figure 6.3 The SQL statement and its parts, and three queries as the builder shows them beside the SQL it writes, with the number of sectors each selects.

Two operators are worth learning by their SQL names, because query builders show them under other labels. **IN** tests a field against a list: `district IN ('Gasabo', 'Kicukiro', 'Nyarugenge')` selects the 35 sectors of the City of Kigali's three districts. A builder shows it as a set of tick boxes ("includes the value(s)"); in SQL it is one line, shorthand for three equal tests joined by **OR**. **LIKE** matches text against a pattern in which `%` stands for any run of characters: `sector LIKE 'Nya%'` finds the 31 sectors whose names begin with Nya, `'%ru%'` would find names containing "ru", and `'%mata'` names ending in "mata". In a file geodatabase **LIKE** is case-sensitive, so `'nya%'` finds nothing.

RULE OF THUMB

Build the query in the dialog, then read the SQL it wrote. The builder only helps you type the query. The SQL is the query itself, and it is the text you can check, copy and reuse.

AND, OR, NOT and the selection types

Tests are combined with three operators. **AND** keeps rows that pass both tests, **OR** rows that pass either or both, and **NOT** reverses a test. They can join clauses inside one query: `Total > 50000 OR district = 'Nyagatare'` selects 52 sectors, the 42 big ones plus the ten of Nyagatare's fourteen that were not already among them.

The same logic works between queries. ArcGIS Pro's **selection type** determines how a new query combines with what is already selected: a *new selection* replaces it, *add to the current selection* is OR, *remove from the current selection* is NOT, and *select subset from the current selection* is AND. Two more entries switch the selection (everything selected becomes unselected, and the reverse) and clear it.

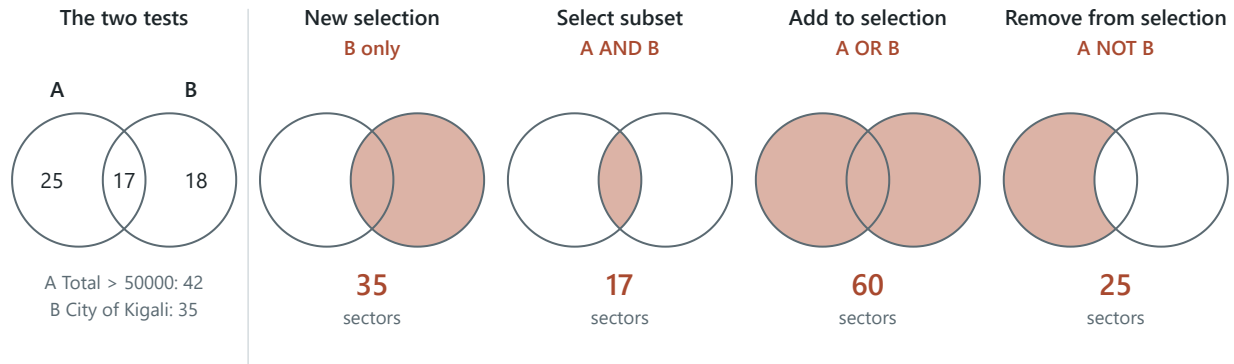


Figure 6.4 The selection types as Venn diagrams, with A the 42 sectors of more than 50,000 people and B the 35 sectors of the City of Kigali. The shaded region is what each selection type leaves selected.

With the 42 big sectors selected, a second query for the City of Kigali gives four different answers depending on the selection type: 35 as a new selection, 17 as a subset (big and in Kigali), 60 added (big or in Kigali) and 25 removed (big but not in Kigali).

Keeping, viewing and clearing a selection

A selection is temporary, and the next new query replaces it. The number of selected features is shown in the Contents pane's List By Selection view and at the foot of the map. A query that reports zero has selected nothing, whatever the dialog showed. This is a real, empty selection, which is different from having no selection, and a tool run on the layer now processes zero rows. In the List By Selection view the box beside each layer controls whether the layer can be selected; it does not control whether the layer is drawn. To switch a layer off, go back to the List By Drawing Order view first.

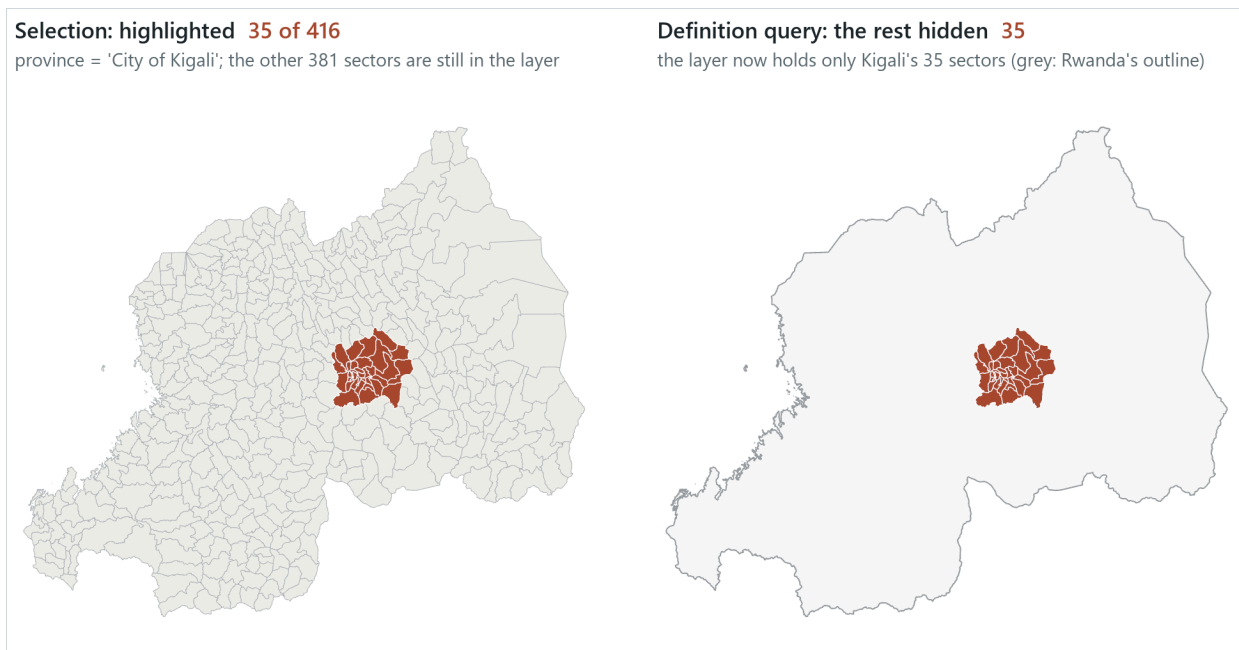


Figure 6.5 The 35 sectors of the City of Kigali as a selection, highlighted among all 416, and as a definition query, which leaves the layer holding only those 35.

Sectors: Sector Boundary 2022; population: population_count, 2022 census. Source: National Institute of Statistics of Rwanda (NISR), ArcGIS Online

There are three ways to keep a selection. **Make a layer from the selection:** a new layer that holds only those rows, in this map only, with nothing written to disk. **Export it:** exporting a layer while a selection is active copies only the selected rows into a new, permanent feature class. **Turn it into a definition query:** a filter on the layer that hides every row that fails the test.

A selection **highlights** rows and does not remove the others. The other rows stay in the layer, but while the selection is active, tools that honour selections (most of them) use only the selected rows. Clear the selection to use all rows. A layer is therefore in one of three states: no selection, some rows selected, or zero rows selected after a test nothing passed. A definition query **filters:** the layer behaves in the map, in its table and in every tool as though the other rows did not exist, until the query is removed. Neither changes the data on disk.

IN THE PANE

Module 6 lesson 2 runs every query on sectors_census with Select By Attributes and has you open the SQL Editor view after each one: Total > 50000 (42), a subset with the City of Kigali (17), one OR clause (52), an IN list (35) and a LIKE pattern (31). Waldo checks each count. You then find the selection in List By Selection, switch back to List By Drawing Order, and keep the 31 as a map-only layer.

Spatial relationships

A **spatial query** selects features by where they are relative to other features, not by their values: sectors that contain a hospital, hospitals inside Kigali, sectors within 5 km of a hospital. The simplest is the cursor: drawing a box, a lasso or a polygon on the map selects whatever it touches. A tick box for each layer controls which layers the cursor can select, so a box drawn around a city can pick sectors without also picking the hospitals inside them.

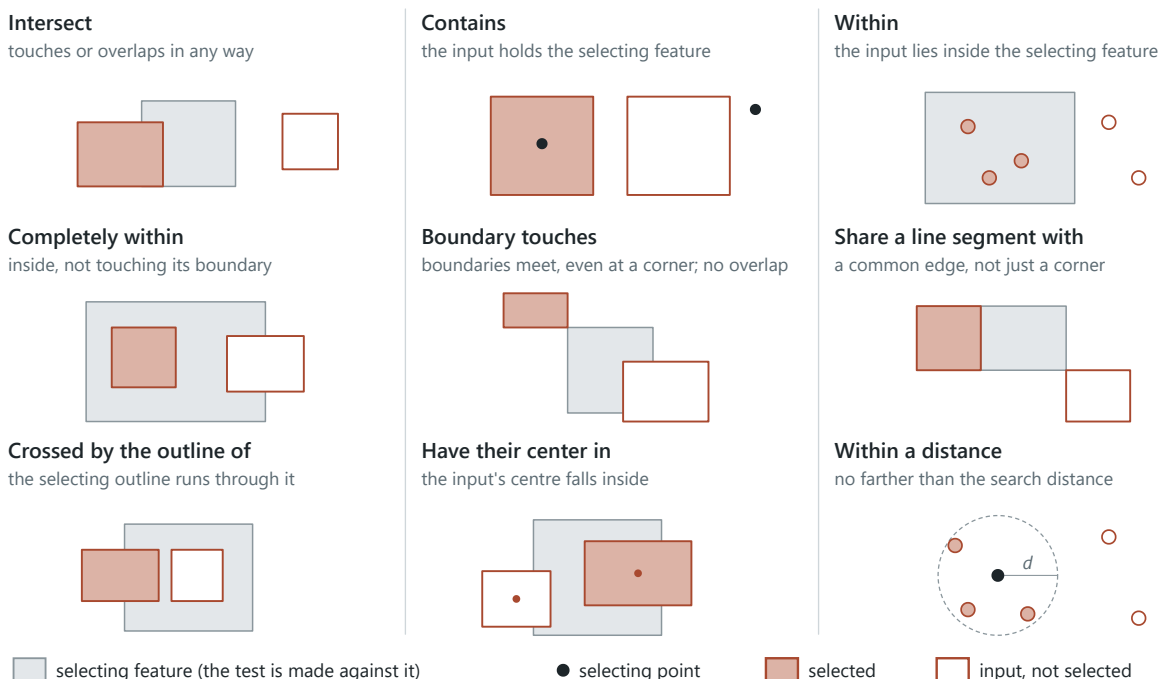


Figure 6.6 Nine spatial relationships, each shown with one small pair of features. The grey feature is the selecting feature; the filled input feature passes the test and the outlined one does not.

Select By Location applies a spatial test between two layers. The **input features** are the layer whose rows are selected, the **selecting features** are the layer the test is made against, and the **relationship** is the test itself. The tool also takes a search distance and the same selection types as attribute queries, and an invert option selects the features that fail the test instead.

Intersect is the broadest relationship: any contact or overlap. **Containment** has two directions, an input that *contains* the selecting feature or one that lies *within* it, and a stricter *completely within* that may not touch the boundary. **Proximity** covers distance (*within a distance*) and adjacency, where features meet without overlapping (*boundary touches*, even at a corner, and *share a line segment*, which needs a common edge). Two more are useful with polygons: *crossed by the outline of* selects what a boundary runs through, and *have their center in* selects features whose centre falls inside.

Select By Location in Rwanda

Asked of Rwanda's 416 sectors and the 55 hospitals mapped in the 2022 census, these tests give the following counts: 50 sectors intersect a hospital, meaning that a hospital lies inside them. Within 5 km of a hospital lie 248 sectors; inverting the test leaves the 168 that have no hospital within 5 km.

These counts need careful interpretation. A sector passes when any part of it lies within 5 km of a hospital point, so a long sector passes if its far corner is within 5 km, even when most of its people live well beyond that distance. The test measures how close polygons lie to the mapped hospitals; whether people can reach care depends on where in the sector they live, on roads and travel time, on the hospital's capacity and on clinics the layer does not hold.

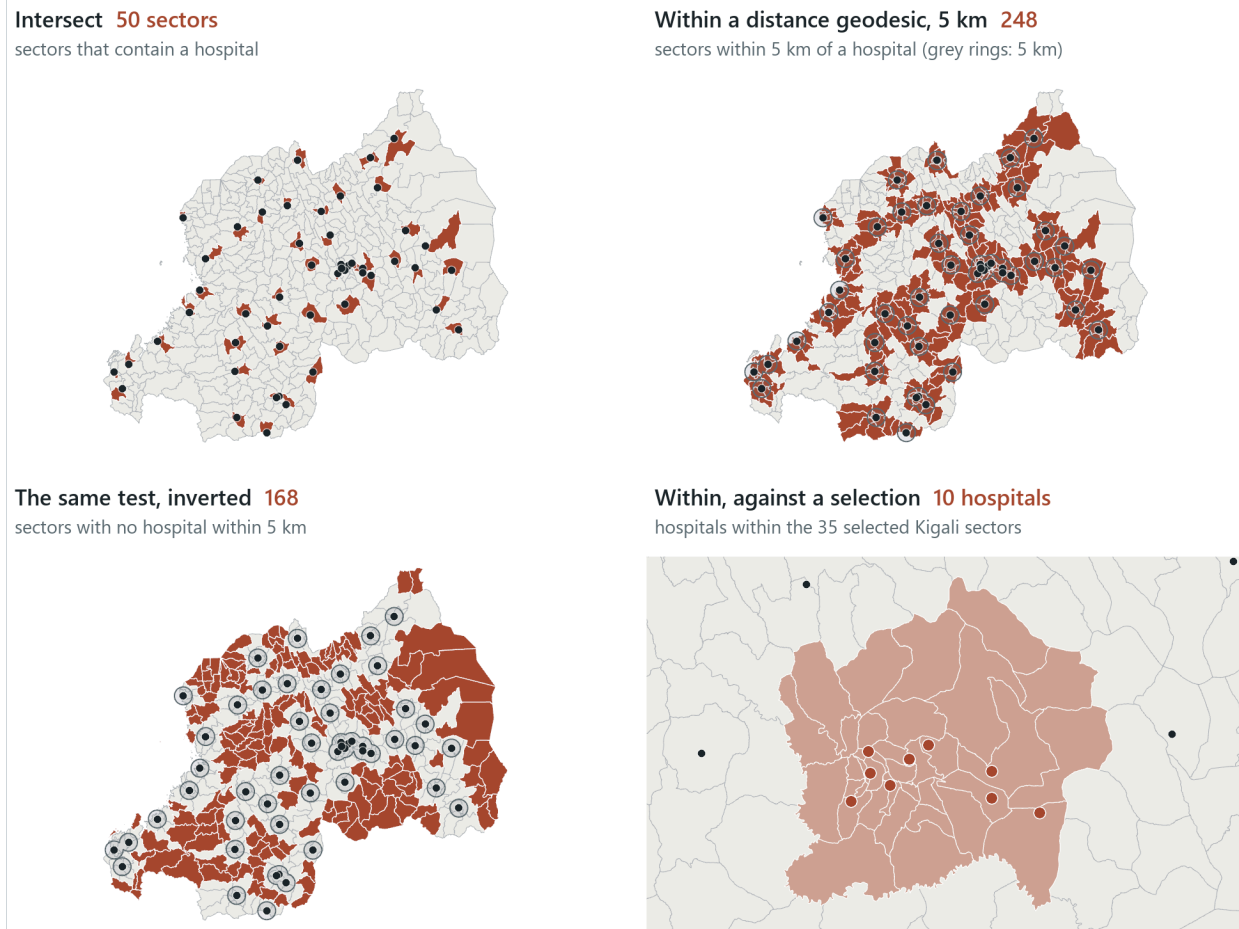


Figure 6.7 Select By Location on the Rwanda sectors and hospitals: intersect, within a distance, the same test inverted, and hospitals within a selected set of sectors.

Sectors: Sector Boundary 2022; population: population_count, 2022 census; hospitals: Health_Establishments_hospital. Source: National Institute of Statistics of Rwanda (NISR), ArcGIS Online

These layers are stored in latitude and longitude, so the search distance box starts in decimal degrees, and a value of "5" left in degrees is about 550 km, a distance that every sector passes. Set the unit to kilometres, and on data stored in degrees use the geodesic form of the test, which measures on the ellipsoid (Chapter 3); here the plain planar test gives 247 instead of 248.

The selecting layer can itself have a selection. With the 35 Kigali sectors selected, a query for hospitals *within* the sectors tests the hospitals against those 35 only and finds Kigali's 10 hospitals. This is useful, but it can also cause mistakes, because a selection left on the selecting layer narrows every later spatial test without any warning.

RULE OF THUMB

Before a spatial query, check the units of the search distance and clear any selection you do not mean to use. After it, read the count before you rely on what the map shows.

Attribute and spatial together

Many real questions have an attribute condition and a spatial one. The classic example is funding for earthquake preparedness: cities within 50 miles of a past earthquake and with more than 250,000 people. The two queries run one after the other, and the order does not matter, as long as the second combines with the first instead of replacing it.

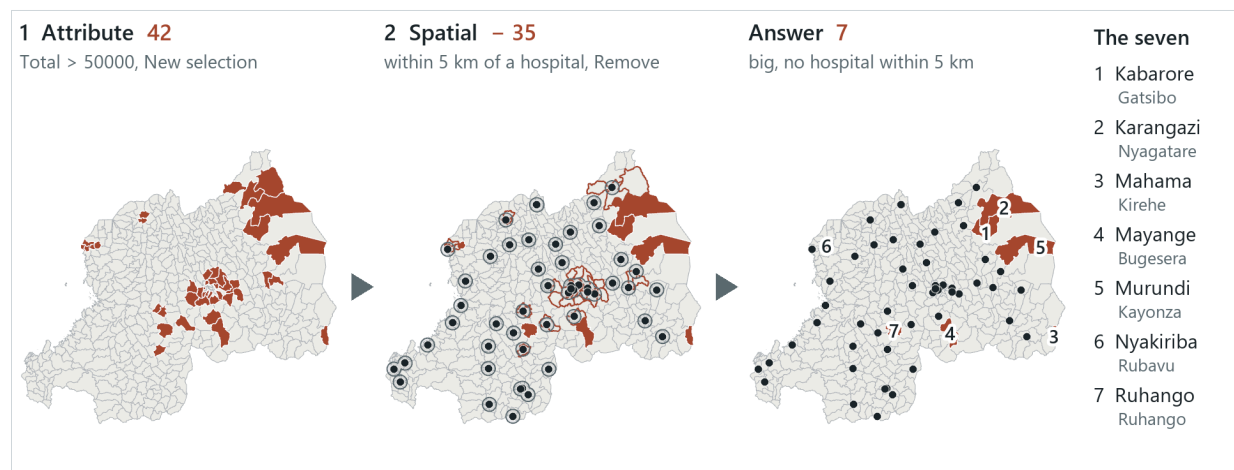


Figure 6.8 Two queries combined to answer one question. The 42 sectors of more than 50,000 people, minus the 35 of them within 5 km of a hospital, leave 7 big sectors with no hospital within 5 km.

Sectors: Sector Boundary 2022; population: population_count, 2022 census; hospitals: Health_Establishments_hospital. Source: National Institute of Statistics of Rwanda (NISR), ArcGIS Online

In Rwanda the health-planning version is: which big sectors have no hospital within reach? Select the 42 sectors of more than 50,000 people, then run *Select By Location* for sectors within 5 km of a hospital with the selection type *remove from the current selection*. Seven sectors remain: Kabarore, Karangazi, Mahama, Mayange, Murundi, Nyakiriba and Ruhango. They are a shortlist produced by a proximity test, and the test alone does not show that their access to care is poor. Using *select subset* instead would keep the 35 big sectors that do have a hospital nearby, which answers the opposite question, and a new selection would discard the first result. Running the spatial test first (168) and the attribute test second, as a subset, gives the same seven. Exporting the selection saves the result as a feature class.

READING THE RESULT

- The question: which big sectors have no hospital within 5 km?
- The representation: hospital points and a 5 km geodesic distance from any part of a sector.
- What it shows: 248 sectors pass, 168 fail, and seven sectors of over 50,000 people fail.
- What it cannot establish: access to care, which depends on homes, roads, capacity and clinics.

IN THE PANE

Module 6 lesson 3 adds the hospitals by URL, selects with the cursor, then runs *Select By Location* four ways: intersect (50), within a distance geodesic of 5 km (248), inverted (168) and hospitals within the selected Kigali sectors (10). With *Invert* unticked again, it ends with the combined question and exports the seven sectors as `big_sectors_no_hospital`, and you read the output path before clicking OK, because Pro reuses the last Output Location it used, which can be another project's geodatabase.

Summary statistics

A summary reduces many rows to a few numbers. **Summary statistics** computes statistics of a field (sum, mean, minimum, maximum and others) for groups of records defined by a **case field**, one output row per value of that field. If a selection is active, only the selected rows are used. The result is always a table, because a summary has no geometry of its own. Besides the statistics it carries a **FREQUENCY** column, the number of rows in each group, and it names each statistic after its field, such as **SUM_Total** and **MEAN_Total**.

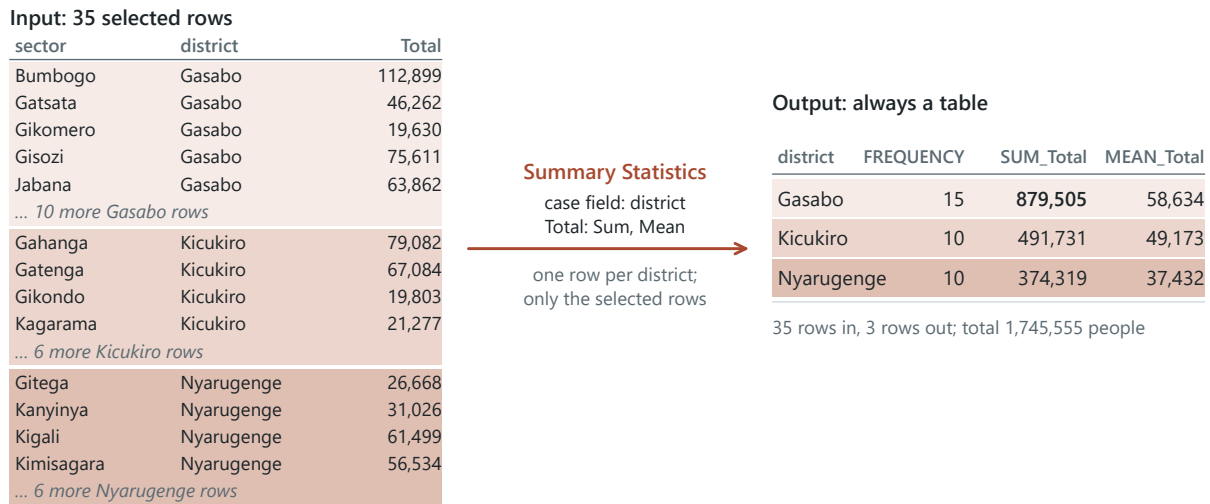


Figure 6.9 Summary statistics on the 35 selected Kigali sectors with district as the case field: 35 rows in, three rows out, one per district.

Summarised by district, the 35 Kigali sectors become three rows: Gasabo's 15 sectors hold 879,505 people, about half the city's 1,745,555. Run again on all 416 sectors with province as the case field and no selection, the same tool gives five rows, and the Eastern Province, with 3,563,145 people, is the most populous, more than twice the City of Kigali, whose sectors are the densest. A summary table can be joined back to its polygons by the case field to be mapped, as in Chapter 5.

Charts: which one answers what

A chart is a drawing of the values in a table, and each kind of chart suits a different question. A **bar chart** shows an amount per category. A **line chart** shows change over time. A **histogram** shows how the values of one numeric field are spread, as counts in bins. A **box plot** shows the same spread as a median, quartiles and range, and is suited to comparing groups side by side. A bar chart and a histogram look alike, but a bar chart compares categories and a histogram describes a distribution.

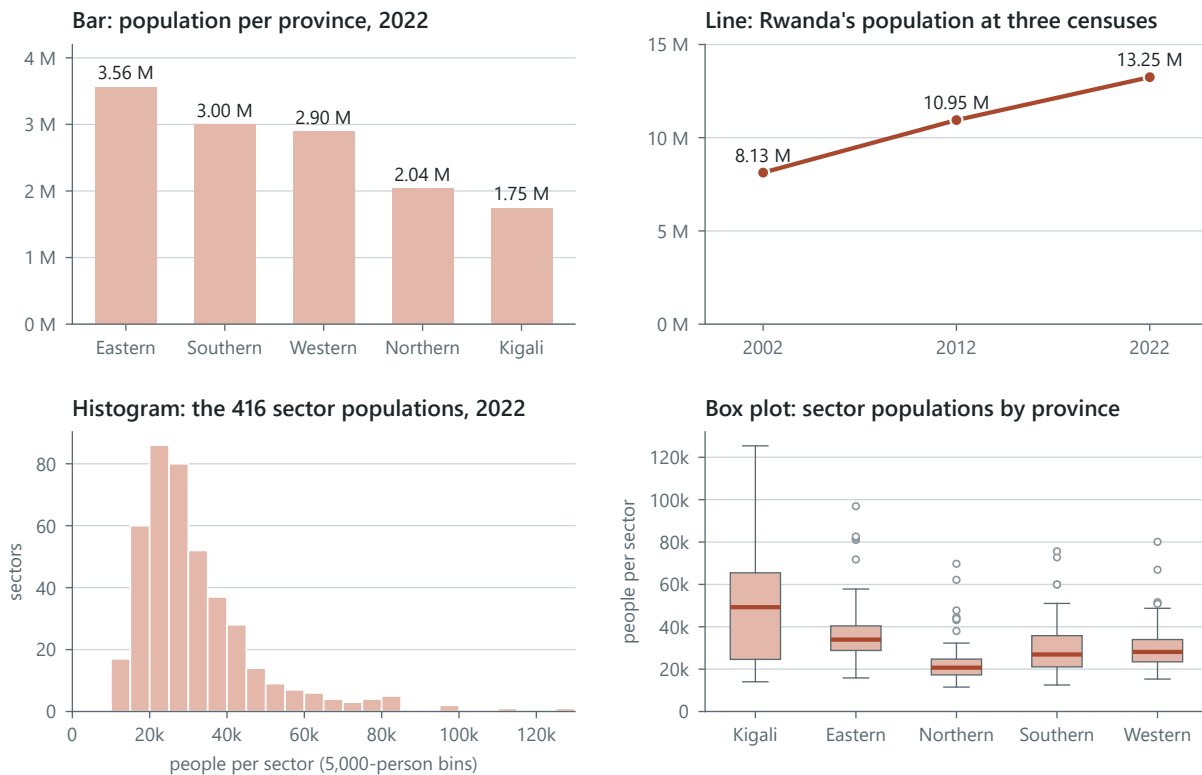


Figure 6.10 The same Rwanda data four ways: population per province (bar), change across three censuses (line), the spread of sector populations (histogram) and that spread compared by province (box plot).

Data: Sector Boundary 2022 and population_count (2002, 2012, 2022 censuses), National Institute of Statistics of Rwanda (NISR), ArcGIS Online

A **scatter plot** draws one dot per feature by two numeric fields, to show whether and how they go together. A fitted trend line helps to show this, and its R^2 gives the share of the variation that the line accounts for: near 1 the dots lie close to the line, near 0 the line explains little (the two may still be related in a way no straight line fits). Sector area against population is weak, with R^2 about 0.07: Rwimbogo, the largest sector at 636 km², holds under 50,000 people, while Kinyinya, the most populous with 125,400, covers about 24 km². Men against women in the same sectors is strong, with R^2 about 0.96. A strong relationship is not necessarily a meaningful one. Both counts rise with a sector's population, so they move together, and a relationship between areas tells you nothing about cause or about individuals (Chapter 4). The outliers are often the most interesting dots. In ArcGIS Pro a chart and its map share one selection, so clicking an outlier in the chart selects that feature on the map at once.

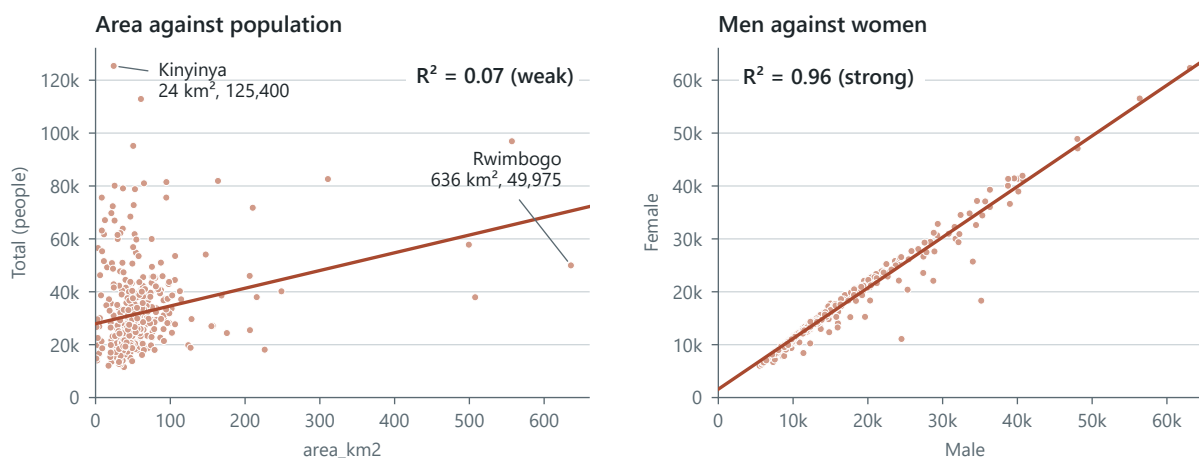


Figure 6.11 A weak and a strong relationship among Rwanda's 416 sectors: area against population, and men against women, each with its linear trend line and R^2 .

Data: Sector Boundary 2022 and population_count (2022 census), National Institute of Statistics of Rwanda (NISR), ArcGIS Online

READING THE RESULT

- The question: does population go with area, and men with women, across the sectors?
- The representation: one dot per sector for two fields, with a straight trend line and its R^2 .
- What it shows: weak for area (R^2 about 0.07), strong for men and women (R^2 about 0.96).
- What it cannot establish: cause or anything about individuals; both counts track population.

RULE OF THUMB

Choose the chart by the question: a bar for amounts per category, a line for change over time, a histogram or box plot for a distribution, a scatter plot for whether two fields go together. Then look for the dots that lie far from the general pattern.

IN THE PANE

Module 6 lesson 4 runs Summary Statistics from a column heading twice, on the Kigali selection by district (Gasabo 879,505) and on everything by province (the Eastern Province), each time reading the Output Table path before OK, then makes a scatter plot of area against population with a trend line, names the far-right dot and selects in the chart to show the selection on the map. A second scatter plot, men against women, shows a strong relationship for comparison.

TRY IT

With the Module 6 project open, ask Waldo in Ask mode: which three districts have the most people, and how many sectors does each have? Waldo summarises the layer for you. Then ask it to write the SQL for the sectors in those districts, and read the clause before you run it yourself.

Key terms

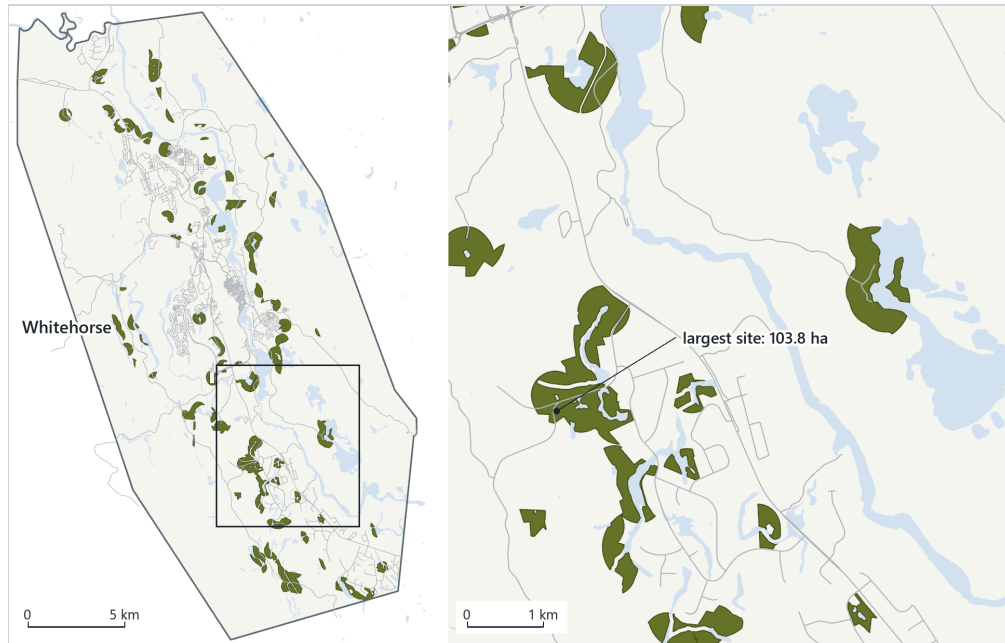
spatial data exploration (ESDA)	interactive, iterative questioning of data through linked maps, tables and charts.
attribute query	a query that selects records by their attribute values, written as an SQL WHERE clause.
SQL	Structured Query Language, the language of relational databases: SELECT fields FROM a table WHERE a test is true.
WHERE clause	the test part of an SQL query: a field, a comparison and a value, joined to others by AND, OR and NOT.
IN	an SQL operator that tests a field against a list of values.
LIKE	an SQL operator that matches text against a pattern, with % standing for any run of characters.
selection	the features currently highlighted, shown on the map, in the table and in the count; replaced by the next new query.
selection type	how a new query combines with the current selection: new, add (OR), remove (NOT), subset (AND), switch or clear.
definition query	a layer filter that hides the features failing a test, without changing the stored data.
spatial query	a query that selects features by their spatial relationship to other features.
spatial relationship	the geometric test in a spatial query, such as intersect, within, contains, touches or within a distance.
adjacency	features meeting along a boundary without overlapping.
summary statistics	statistics of a field computed per group of records; the output is always a table.
case field	the field whose values define the groups, one output row per value, in a summary.
histogram	a chart of one numeric field's distribution, as counts in value bins.
box plot	a chart of a distribution's median, quartiles and range, suited to comparing groups.
scatter plot	a chart of two numeric fields, one point per feature, for reading their relationship.
R^2	the share of the variation in the vertical-axis variable that the fitted line accounts for, from 0 (none) to 1 (all).

Further reading

- Snow, J. (1855). *On the Mode of Communication of Cholera* (2nd ed.). London: John Churchill. Public domain; Wellcome Collection <https://wellcomecollection.org/works/uqa27qrt> and Project Gutenberg <https://www.gutenberg.org/ebooks/72894>
- Olaya, V. *Introduction to GIS*, the chapters on databases (queries and logical operators) and analysis (spatial queries). <https://volaya.github.io/gis-book/en/>
- Esri. ArcGIS Pro documentation: *Select features by attributes* and *Select features by location*, with the full list of spatial relationships. <https://pro.arcgis.com/en/pro-app/latest/help/mapping/navigation/select-features-by-location.htm>
- The course's data sources: the Rwanda sectors, census table and hospitals, and John Snow's data, with their publishers and licences.

Vector analysis

Accompanies Module 7 · Vector Analysis



Government of Yukon, GeoYukon (Municipal Boundaries; NG911 Roads, Hydrology – Line and – Polygon – Whitehorse; Land Parcels – Surveyed). Contains information licensed under the Open Government Licence – Yukon.

WHY THIS MATTERS

Queries find features that are already in your data, while analysis creates new data, such as the land within 30 metres of a stream, the geology inside a city or the wells on each rock unit. Each analysis tool reads layers, applies a spatial rule and writes a new feature class. When a few of these tools are run in sequence, they can answer practical questions. The map on this page shows 97 possible campground sites that a sequence of eight tools selected from a 420 km² city.

YOU SHOULD BE ABLE TO EXPLAIN

- why undissolved buffers count the ground where they overlap more than once
- why a clipped piece keeps its parent's count while its measured area is recomputed
- why a site-selection result depends on one set of chosen thresholds and gives a shortlist for further judgement rather than a decision

IN THIS CHAPTER

Analysis makes new data

Buffers

Clip and erase: the cookie cutters

Union and intersect: overlays with attributes

Dissolve, merge and append

A site-selection chain

Spatial joins: attributes by location

Analysis makes new data

The queries and selections of the last two chapters picked out rows and changed nothing on disk. The tools in this chapter are different: each one reads one or two layers, applies a rule such as "everything within 30 metres" or "everything inside this boundary", and writes a **new feature class**. The inputs are left as they were; of the tools in this chapter only Append, which adds features into an existing dataset, changes something that already exists. Vector analysis is therefore often described as a set of **transformations**, in which simple spatial rules turn existing layers into more useful products.

Because every result is a feature class, every result has a name, a place on disk and an attribute table of its own. If a run is wrong, you can delete its output and run the tool again. The software also keeps a **history** of every tool run with its inputs, outputs and settings, so the whole analysis can be reviewed and repeated.

Every tool: one or two layers in → a spatial rule → a NEW feature class out. The inputs are never changed.

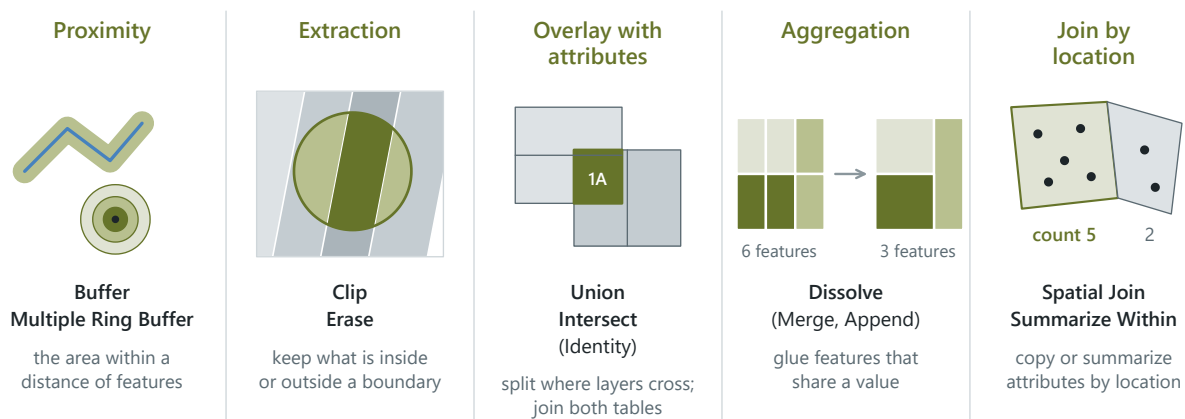


Figure 7.1 The families of vector operations used in this chapter. Proximity tools create shapes from distances, extraction tools cut one layer with another, overlay tools split shapes and combine their attribute tables, aggregation combines features that share a value, and a join by location copies or summarizes attributes from one layer to another.

The families differ in what they do to shapes and to attributes. A buffer creates new shapes from one layer. Clip and erase cut one layer with another and leave the attributes unchanged. Union and intersect split shapes where they cross and combine both attribute tables. Dissolve combines features and keeps only the attributes you ask for. A spatial join leaves the shapes unchanged and transfers attributes between layers by location.

Two practical points apply to all of them. Distances and areas are measured in the units of the data's coordinate system, so analysis should be done in a projected system suited to measurement (Chapter 3). The Yukon layers used here are stored in NAD 1983 Yukon Albers, in metres. The location of the output is as important as its content. The same tool can write into the project geodatabase or into a folder as a shapefile, and a shapefile shortens field names to ten characters, so a field such as WELL_DEPTH_FTBGS becomes WELL_DEPTH.

RULE OF THUMB

Read the output path and name before running any tool. Save results in the project geodatabase, under names you will recognise later.

IN THE PANE

Module 7 lesson 1 starts a new project for Whitehorse, Yukon, and cuts the city out of the territory's municipal boundaries as a study area. It opens the Geoprocessing pane, which gives access to about 1,600 tools, and shows the History pane that records every run.

Buffers

A **buffer** is a polygon covering every place within a given distance of a feature. Points, lines and polygons can all be buffered. Typical uses are setbacks in zoning, protective zones along streams and around other sensitive features, and capturing what lies around a feature, such as the customers near a store.

Whitehorse keeps new buildings 30 metres back from watercourses. The city's stream layer holds 1,368 stream lines, and a 30 m buffer around them marks the setback. The tool offers a choice in how the result is stored. Without dissolving, the tool writes one polygon per stream, and neighbouring polygons overlap wherever streams meet or run close. The areas of those 1,368 polygons add up to 85.0 km², about twice the true figure, because the overlaps are counted more than once. Dissolved into one feature, the setback covers 40.4 km², roughly a tenth of the city.

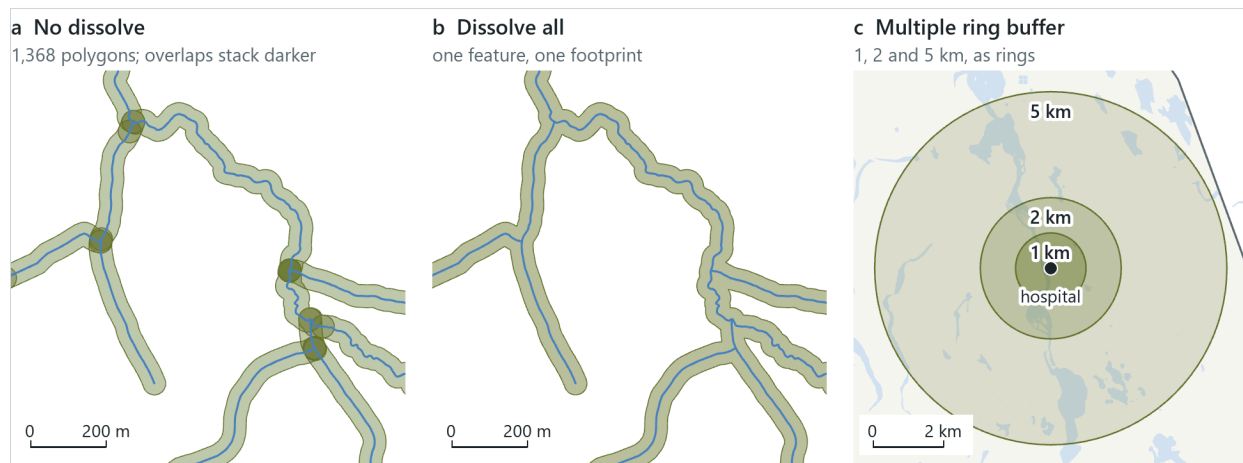


Figure 7.2 Buffers on Whitehorse data. (a) A 30 m buffer around each stream line: each stream has its own polygon, and the polygons overlap at junctions. (b) The same buffer dissolved into a single feature. (c) A multiple ring buffer around the hospital at 1, 2 and 5 km, stored as rings.

Government of Yukon, GeoYukon (NG911 Hydrology – Line and – Polygon – Whitehorse; Health care facilities; Municipal Boundaries). Contains information licensed under the Open Government Licence – Yukon.

READING THE RESULT

- The question: how much of Whitehorse lies inside the 30 m stream setback?
- The representation: 1,368 streams buffered by 30 m, one polygon each or dissolved into one.
- What it shows: 85.0 km² when overlaps are counted again, 40.4 km² once dissolved.
- What it cannot establish: whether a site can be built on; the setback is only one rule.

A **multiple ring buffer** creates buffers at several distances in one run. Around the city's hospital, rings at 1, 2 and 5 km cover 3.14, 9.42 and 65.93 km². Stored as **rings**, each feature is the band between two distances, the areas do not overlap, and the output's distance field can be used to map distance classes. Stored as overlapping **disks**, each feature is everything within its distance (3.14, 12.57 and 78.54 km²). A **negative buffer**, a negative distance applied to a polygon, shrinks it inward instead.

Tools honour a selection, so with only the hospital selected among the territory's 17 health facilities, the rings are drawn around the hospital alone.

 **IN THE PANE**

Module 7 lesson 1 buffers the streams by 30 m with the dissolve switched on, explains why the undissolved version double-counts, and reads the setback's area from its table. It then selects the hospital and runs a multiple ring buffer at 1, 2 and 5 km as rings, and names the negative buffer without running it.

Clip and erase: the cookie cutters

An **overlay** combines two feature classes into a new one. The simplest overlays are extraction tools, which are often compared to cookie cutters. **Clip** keeps the parts of the input features that fall inside a boundary; **Erase** keeps the parts that fall outside it. In both, the input's attributes are kept unchanged, and the cutting layer provides only its shape and none of its columns.

Unchanged attributes are correct for descriptions, because a rock type is still the same rock type on a smaller piece. They are wrong for counts: a district clipped in half still has its whole population in the table. A measured field such as Shape_Area is recomputed for the new shape; a stored count is not. Estimating the people in a fragment from its share of the area assumes that they were spread evenly, and the overlay has no information about where they live.

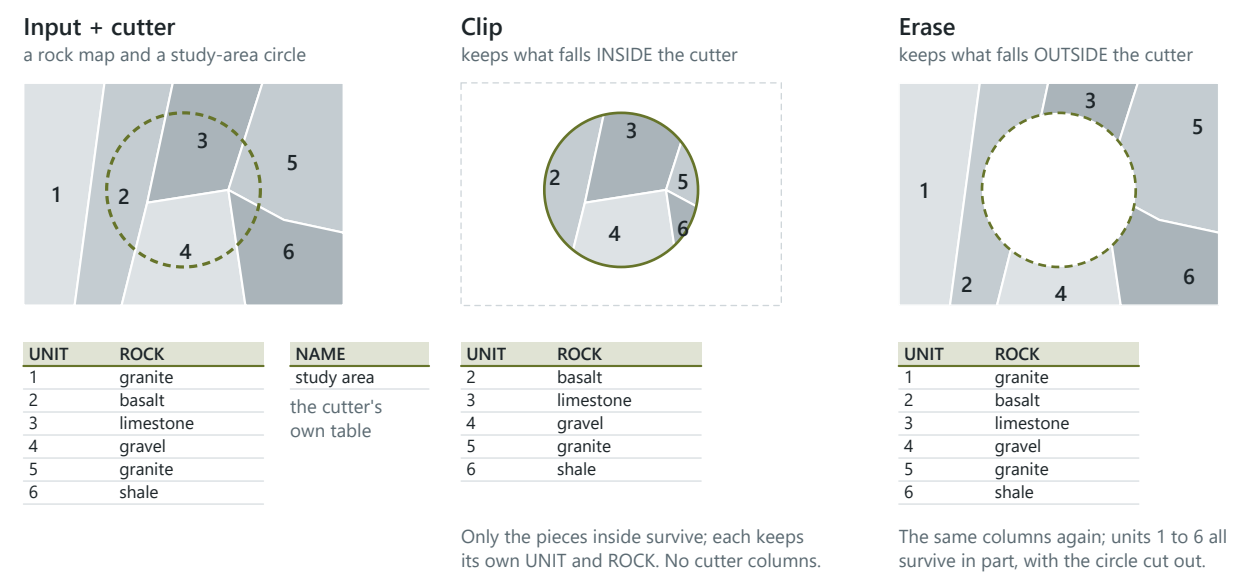


Figure 7.3 One cutting shape used by two tools. Clip keeps the pieces of the rock units inside the circle; Erase keeps the pieces outside it. Both outputs carry the input's own columns (UNIT, ROCK) and nothing from the cutter.

Clip is commonly used to prepare data. The Yukon's bedrock geology map has 28,565 polygons for the whole territory; clipped to Whitehorse it has 64, each still carrying its unit code, geological period and rock type. Points can be clipped too. A point is either inside or outside the boundary, so each point is either kept whole or dropped. Of the territory's 3,244 registered water wells, 1,102 lie in the city, and their table still records the purpose of each well; the most common purpose is private domestic supply, 532 wells. Selecting the wells by location would find the same ones, but a clip writes a layer that you keep, whereas a selection is temporary and remains part of the whole territory's layer.

Erase is commonly used in suitability analysis to remove land that cannot be used. Erasing the city's lakes, wetlands and river from the city boundary leaves 395.45 km² of dry land; 24.3 km² of the city is water.

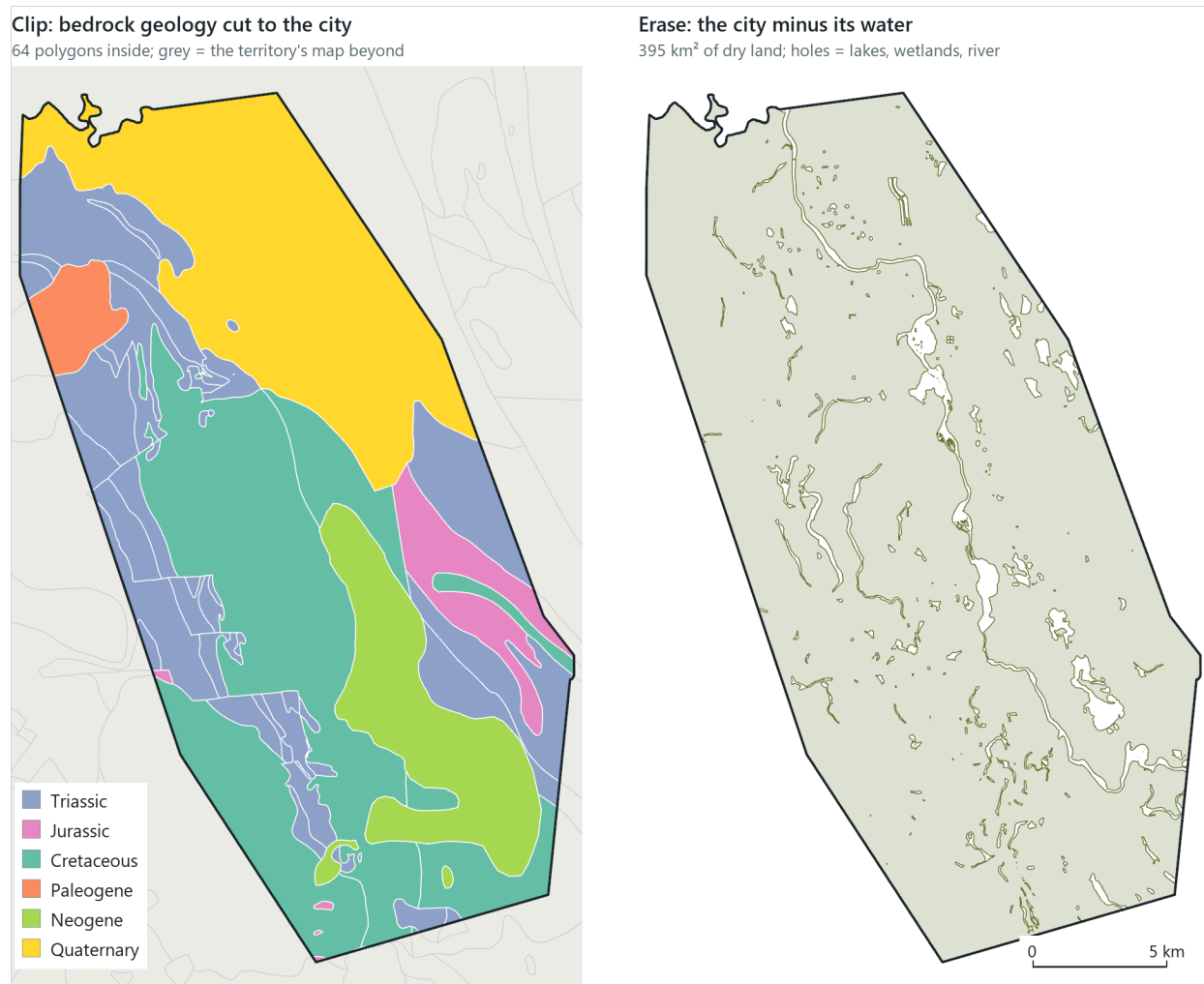


Figure 7.4 Clip and erase on Whitehorse. Left: the territory's bedrock geology clipped to the city boundary, the pieces coloured by geological period. Right: the city with its water erased, the holes where lakes, wetlands and the river were.

Government of Yukon, GeoYukon (Bedrock Geology, Yukon Geological Survey; NG911 Hydrology – Polygon – Whitehorse; Municipal Boundaries). Contains information licensed under the Open Government Licence – Yukon. Colours: ColorBrewer Set2, Cynthia Brewer (Apache 2.0)

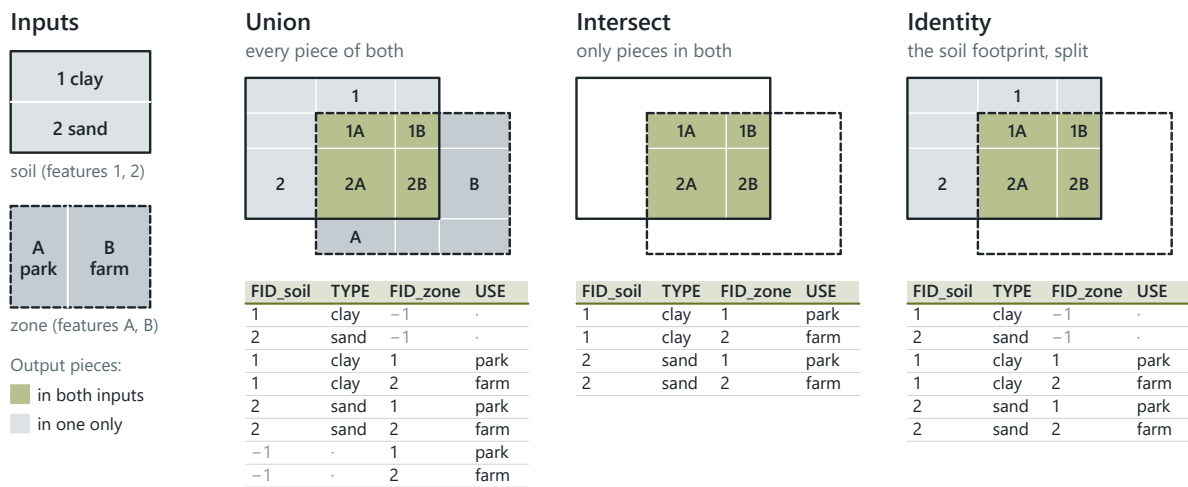
A large territory-wide layer served over the web can take a long time to clip. A geoprocessing **environment**, a setting applied around a tool, can limit the processing extent to the study area so that only the nearby features are read.

IN THE PANE

Module 7 lesson 2 clips the bedrock geology and the wells to the city, sets the Extent environment for the big geology layer, and reads the most common well purpose to show that the attributes are kept. It then erases the water from the city and compares the dry-land area with the city's.

Union and intersect: overlays with attributes

The second kind of overlay joins attributes as well as shapes. Where features of the two inputs overlap, the tool splits them at the shared edges, and every resulting piece carries the attributes of both inputs. **Union** keeps every piece of both inputs; **Intersect** keeps only the pieces that lie in both. **Identity** keeps the first input's footprint and splits it wherever the second covers it. The order in which the inputs are listed makes no difference to a union or an intersect.



FID_ columns record which input feature each piece came from; -1 means none, and that input's columns stay empty (.).

Figure 7.5 Overlay with attributes. Two inputs, soil (features 1 and 2) and zone (A and B), overlap. Union keeps all eight pieces, Intersect the four pieces in both, Identity the six pieces of the soil footprint. Each output table carries both inputs' columns; -1 in an FID_ column means the piece is not in that input.

The output table records where each piece came from. The overlay adds an **FID_** column for each input, named after the input dataset, holding the ID of the feature the piece came from. A value of -1 means "not in this input", and that input's other columns are empty for the piece. The attributes from the two tables appear side by side in each row.

In Whitehorse, the union of the 30 m stream setback with the city's water produces a layer of everything a building site should avoid: 1,668 pieces covering 103.66 km², more than the setback and the city's 24.3 km² of water together, because the water layer runs beyond the city limits and a union keeps everything in both inputs; clipping the result to the city would restrict it to the city. Of these, 1,055 are water that lies away from any stream (-1 in the setback's FID_ column), 612 are places where the setback crosses water, and one is the setback on dry land (one piece, because the dissolved setback was a single feature). Intersecting the city's 36 subdivisions with the clipped geology gives more than 80 pieces, each one subdivision on one rock unit, with the subdivision's name beside the rock's period and type.

IN THE PANE

Module 7 lesson 3 unions the stream setback with the water into a no-build layer and reads the -1 in its FID_ columns. It intersects the subdivisions with the clipped geology, so that every piece carries both its subdivision and its rock unit, and names Identity without running it.

Dissolve, merge and append

Dissolve does the opposite of the overlays. Instead of splitting features, it combines the features of one layer that share a value in a chosen field and removes the boundaries between them. The output keeps the dissolve field and any summary statistics you ask for, such as the sum of the areas, and drops every other attribute.

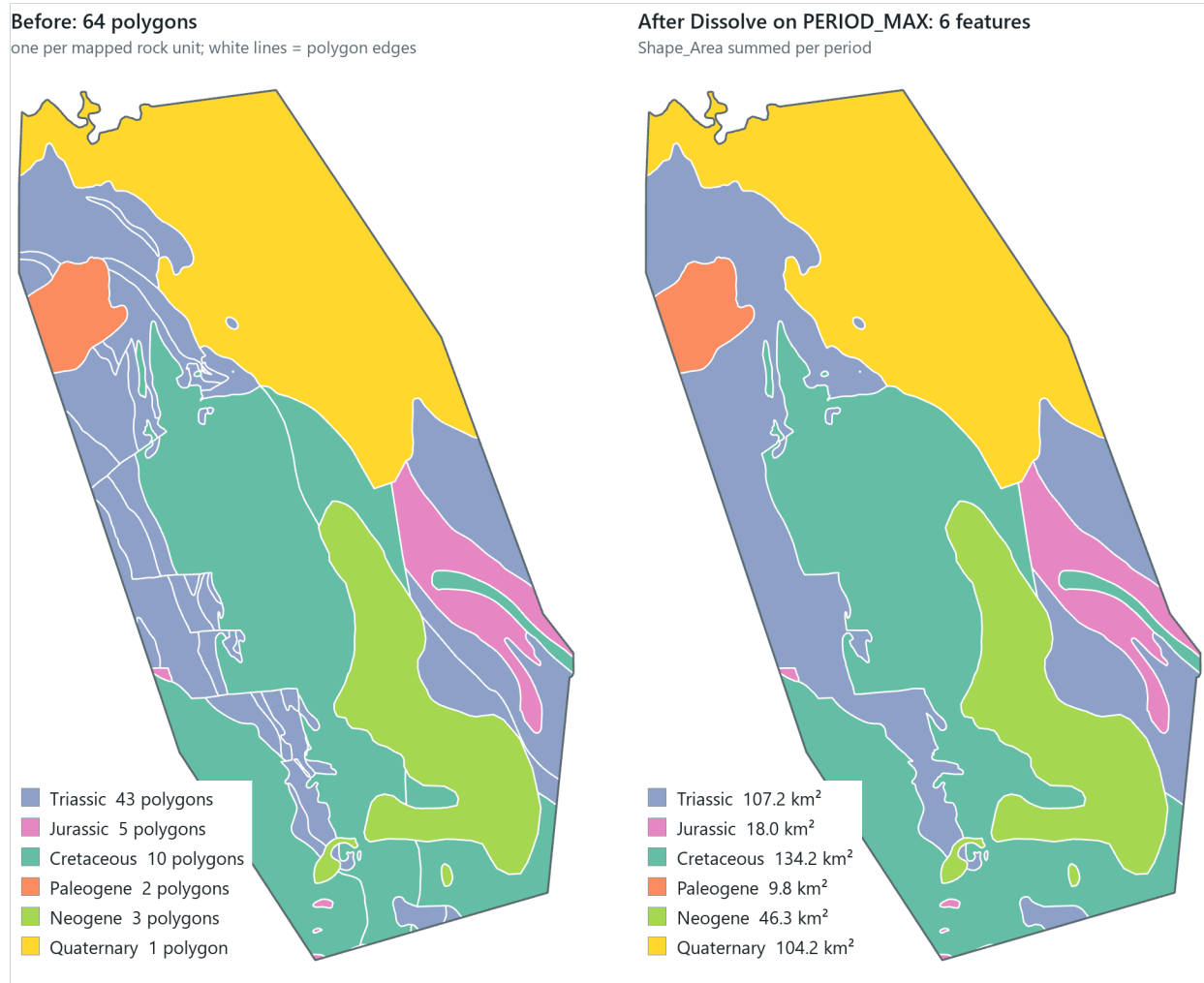


Figure 7.6 Dissolve on the clipped geology. Left: 64 polygons, one per mapped rock unit. Right: dissolved on the geological period with the areas summed, six features remain. Triassic has the most polygons; Cretaceous covers the most area.

Government of Yukon, GeoYukon (Bedrock Geology, Yukon Geological Survey; Municipal Boundaries). Contains information licensed under the Open Government Licence – Yukon. Colours: ColorBrewer Set2, Cynthia Brewer (Apache 2.0)

The 64 geology polygons dissolved on their period become six features. The counts and the areas give different results. Triassic rocks form 43 of the 64 polygons but cover 107.2 km², while the 10 Cretaceous polygons cover 134.2 km², the most of any period. Aggregating in this way replaces a

count of pieces with a total for each period. If the output is allowed to hold only single-part features, each separate patch of a period remains a separate feature, and the output does not have one feature per period.

Merge and **Append** also combine features, but whole layers of the same geometry type: merge writes a new feature class from several inputs, append adds features into an existing one. Neither splits anything, so where the inputs overlap the result holds two stacked features and the overlapping area is counted twice. They are suited to joining adjacent map sheets or several years of the same data. They are not suited to combining layers that overlap, which is done with a union.

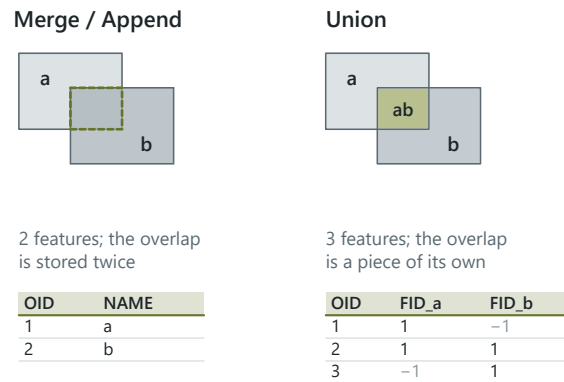


Figure 7.7 Merge or append against union for two overlapping polygons a and b. Merge keeps two whole features, so the overlap is stored twice; union splits them into three pieces, one of which is the overlap.

IN THE PANE

Module 7 lesson 3 dissolves the clipped geology on its period with the area summed and asks which period covers the most of the city. Merge and append are discussed but not run.

A site-selection chain

These tools are most useful when they are combined. A **site-selection model** is used to answer the question "where could we put it?". You write down what a good site must have and must avoid, turn each rule into a spatial operation, and combine the operations until only the places that pass every rule remain.

The question here is where Whitehorse could put a new lakeside campground inside the city. The criteria are six: within 500 m of a road, within 300 m of a lake, outside the 30 m stream setback, outside the wetlands and the water, not on a surveyed parcel (a simplification: a surveyed parcel is land with registered interests, but land outside them is not necessarily free, and First Nation settlement lands, for example, are mapped separately), and at least 2 ha. The kind of rule determines which tool is used. **Requirements**, which a site must meet together, are buffered and intersected. **Constraints**, which a site must avoid, are combined and erased. A **threshold** on the result is a selection.

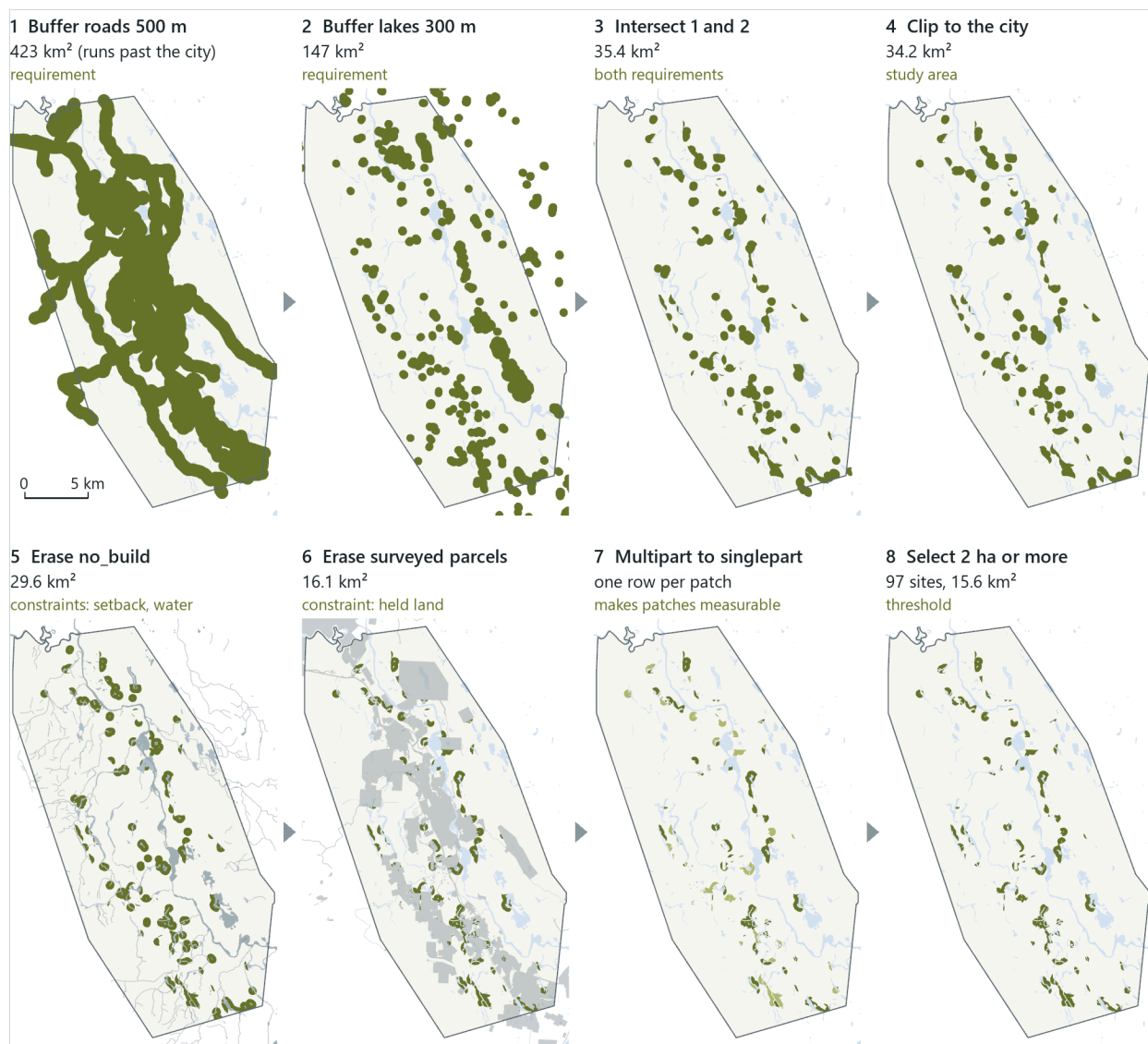


Figure 7.8 The campground chain as eight small maps, each the output of one tool. The requirements (1 and 2) are intersected, the result is clipped to the city, the constraints are erased (the no-build layer, then the surveyed parcels), the multipart result is split into patches, and the patches of 2 ha or more are selected: 97 sites.

Government of Yukon, GeoYukon (Municipal Boundaries; NG911 Roads, Hydrology – Line and – Polygon – Whitehorse; Land Parcels – Surveyed). Contains information licensed under the Open Government Licence – Yukon.

The area left after each step shows how the chain narrows the result. Everything within 500 m of a road covers 423 km², reaching well beyond the city limits; one stray road segment far to the north stretches the layer's extent, and dissolving thousands of overlapping buffers across that extent is slow, so limiting the processing extent to the city halves the area, takes seconds instead of a minute, and here changes nothing inside the city, because every road that matters lies within that rectangle. The extent is a rectangle that selects which inputs are read and does not clip them. In general it must reach a buffer distance beyond the study area, because a road just outside an extent that stops at the boundary loses its 500 m of influence inside it. This edge effect is not visible on the map. Everything within 300 m of a lake covers 147 km². Their intersection, land near both, is 35.4 km², and 34.2 km² of it lies inside the city. Erasing the no-build layer of the setback and water leaves 29.6 km², and erasing the surveyed parcels leaves 16.1 km². Because every buffer upstream was dissolved into one feature, that result is a single **multipart feature**, one row made of many separate shapes;

splitting it into single parts gives one row per patch, which can be measured. Selecting the patches of 2 ha or more leaves **97 sites** covering 15.6 km², about 4 % of the city, 44 of them 10 ha or more; the largest is 103.8 ha.

The result is a shortlist of possible sites and is not a decision. It is also the result of one particular model. Each threshold (500 m, 300 m, 2 ha) is a choice someone made, so if you change the road distance to 400 m or the minimum to 3 ha, the count changes, and a patch just inside a line is not better than one just outside it. State the criteria with the result, and say which are rules (the 30 m setback is the city's) and which are judgements. The analysis narrows 420 km² to 97 places that pass these mapped criteria and are worth a site visit, and people must then decide which of them, if any, is suitable. Every intermediate layer is a named feature class and every step is in the history, so the reasoning can be checked, changed and rerun. The finished map shows the result and its context and leaves out the intermediate layers.

READING THE RESULT

- The question: where inside Whitehorse could a new lakeside campground go?
- The representation: six criteria made into buffers, overlays, erases and a 2 ha threshold.
- What it shows: 97 patches pass every criterion; the largest is 103.8 ha.
- What it cannot establish: which site is right; the thresholds are choices that someone made.

RULE OF THUMB

Write the criteria down before running any tool, and name the kind of each one: requirements are intersected, constraints are erased, thresholds are selected. After each step, look at the output before running the next.

IN THE PANE

Module 7 lesson 4 builds this model in its own map, one criterion at a time: two buffers (the road buffer with the processing extent set to the city, the lake buffer on a selection of lakes only), an intersect, a clip, two erases, a multipart-to-singlepart split and a selection of 2 ha or more. You read the largest site's area, read the finished map against the criteria, and export a picture of it.

Spatial joins: attributes by location

A table join (Chapter 5) needs a key: a value in one table must equal a value in the other. Without a key, two tables cannot be joined in this way, but a GIS can still join two layers by the locations of their features. A **spatial join** attaches attributes from one layer to another based on their spatial relationship: one feature inside another, one closest to another, or one intersecting another. It always works on two layers and always writes a new, permanent layer, and the output has the geometry type of the **target** layer, the one that keeps its shapes and gains columns, and, when each target is matched once and unmatched targets are kept (the case in these examples), its row count too. A one-to-many join repeats a target for each match, and a join that keeps only matched targets drops the rest, so predict the row count from the cardinality (Chapter 5) before reading it. The other layer, the **join features**, supplies the attributes.

There are three kinds. A **simple join** gives each target feature the attributes of the one join feature it matches: every well takes the attributes of the rock unit it stands on. A **summarized join** gives each target feature a count or statistic of the many join features it contains: every subdivision gets the number of wells inside it and their mean depth. A **distance join** matches each target feature to the closest join feature and records how far away it is. Joins can link any combination of points, lines and polygons; the examples here stay with points and polygons.

The simple join of the city's wells (the target) to the clipped geology (the join features) returns 1,102 rows, exactly the number of wells, and each row now carries the geology's columns beside the well's own. A join count field records how many join features each target matched: here 1 in every row, because every well sits on exactly one rock unit. Most of the wells, 579, stand on Cretaceous rock, followed by 272 on Neogene, 224 on Quaternary and 27 on Triassic. The join makes it possible to answer a question that neither layer could answer alone: the Cretaceous wells average 99 feet deep and the Triassic ones 180. Swapping the target and the join features gives a different result: 64 geology polygons, each with a count of wells.

The summarized join reverses the direction, with the subdivisions as the target. Wolf Creek, a subdivision where houses have their own wells, holds the most, 126 wells, followed by Mary Lake with 87 and the industrial area of Marwell with 79. Only 588 of the 1,102 wells fall inside a named subdivision, and subdivisions with no well depths recorded have no mean. The count gives the number of registered wells of every kind in a subdivision: Marwell's 79 are mostly shallow monitoring wells, Wolf Creek's are domestic supply. The mean describes the wells on record, from 26 feet at Marwell to 273 at Cowley Creek, and differences between subdivisions mix geology with the purpose of the wells, how many were recorded and how many had a depth at all; the depth a new domestic well would need can only be judged from the domestic wells, and then by a driller. Comparing the choropleth with the points is a good way to check any summarized join.

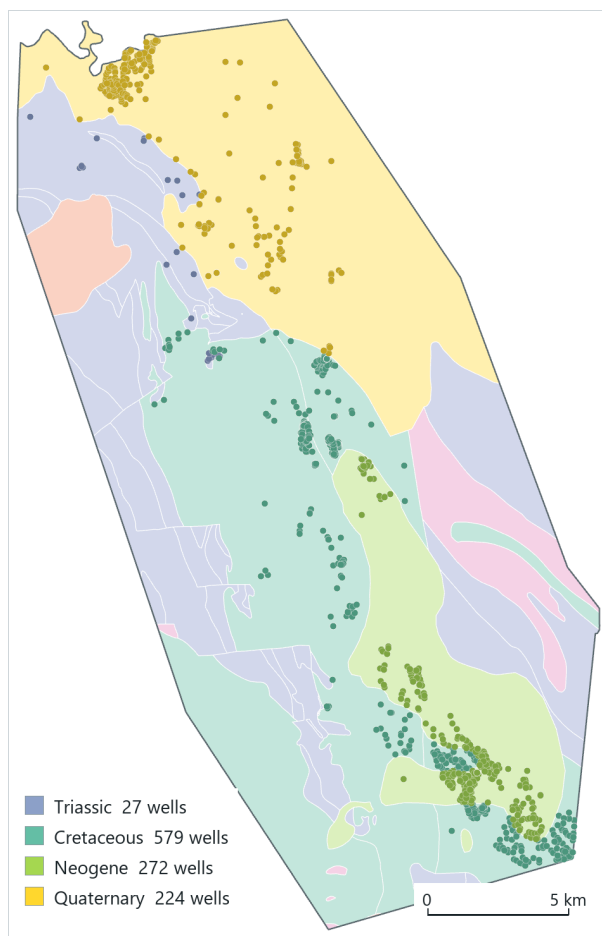


Figure 7.9 A simple join: each of the 1,102 wells coloured by the geological period it took from the rock beneath it.

Government of Yukon, GeoYukon (Water Wells; Bedrock Geology, Yukon Geological Survey; Municipal Boundaries). Contains information licensed under the Open Government Licence – Yukon. Colours: ColorBrewer Set2, Cynthia Brewer (Apache 2.0)

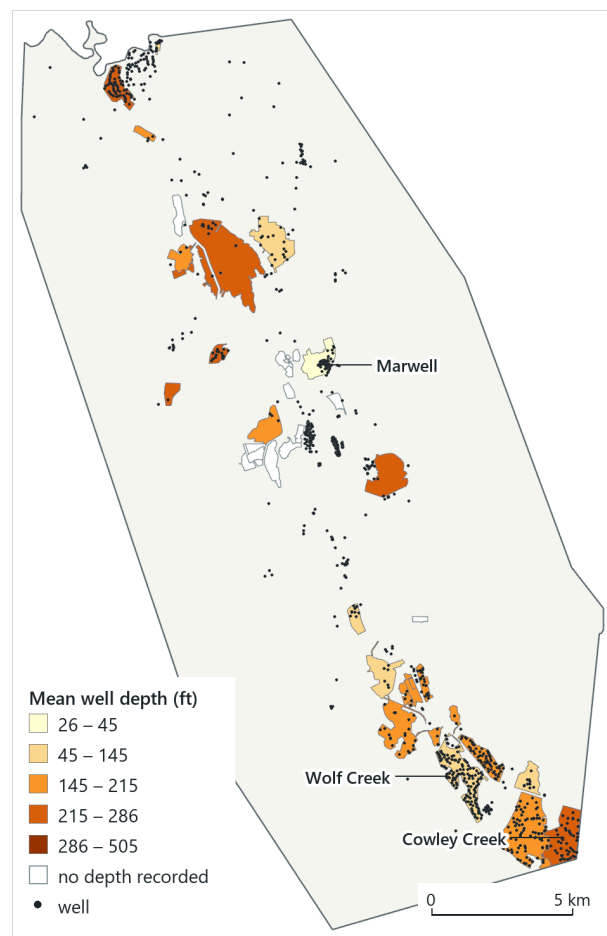


Figure 7.10 A summarized join: the city's subdivisions shaded by the mean depth of the wells inside them (natural breaks), with the wells on top.

Government of Yukon, GeoYukon (Water Wells; NG911 Subdivisions – Whitehorse; Municipal Boundaries). Contains information licensed under the Open Government Licence – Yukon. Colours: ColorBrewer YlOrBr, Cynthia Brewer (Apache 2.0)

READING THE RESULT

- The question: how many registered wells lie in each subdivision, and how deep are they?
- The representation: wells counted per subdivision, with the mean of the depths on record.
- What it shows: Wolf Creek has the most, 126; mean depth runs from 26 to 273 feet.
- What it cannot establish: a new well's depth; the means mix geology, purpose and missing depths.

RULE OF THUMB

In a spatial join, choose the target first, because it determines the geometry type of the result and, with one match per target, the number of rows. Map a summary that describes the place (a rate or a mean), format the class labels as readable numbers, and then check how the rounding changed them.

IN THE PANE

Module 7 lesson 5 joins the wells to the geology and asks which period most wells stand on, then counts the wells in each subdivision with the Summarize Within tool and draws the mean well depth as graduated colours with formatted class labels: one significant digit first, which collapses two of these classes to a single number, then whole feet. The distance join is named, not run.

TRY IT

With the Whitehorse project open, ask Waldo in Ask mode: which tool would you use to find the wells within 1 km of the hospital, and what would the output contain? Compare its answer with the families at the start of this chapter.

Key terms

transformation	an analysis operation that reads existing layers and writes a new feature class, leaving the inputs as they were (Append, which adds to an existing dataset, is the exception).
buffer	a polygon covering every place within a given distance of a feature.
dissolve (buffer option)	merging overlapping buffer polygons into one feature so that no area is counted twice.
multiple ring buffer	buffers at several distances around the same features, stored as non-overlapping rings or overlapping disks.
overlay	combining two feature classes by location into a new feature class.
clip	keeps the parts of the input features inside a boundary; the input's attributes are unchanged.
erase	keeps the parts of the input features outside a boundary; the input's attributes are unchanged.
union	an overlay that keeps every piece of both inputs, each piece carrying both inputs' attributes.
intersect	an overlay that keeps only the pieces lying in both inputs, each carrying both inputs' attributes.
identity	an overlay that keeps the first input's footprint, split wherever the second input covers it.
FID_ column	a column an overlay adds to record which feature of an input each piece came from; -1 means none.
dissolve	merging the features of one layer that share a field value, keeping that field and any requested statistics.
merge / append	combining feature classes of the same type into a new one (merge) or into an existing one (append), without splitting; overlaps become duplicate features.
multipart feature	one feature, one row, made of several separate shapes.
site-selection model	a chain of spatial operations that keeps only the places meeting every written criterion.
spatial join	attaching one layer's attributes to another's features by their spatial relationship.
target / join features	in a spatial join, the layer that keeps its shapes and rows / the layer that supplies the attributes.
summarized join	a spatial join that gives each target feature a count or statistic of the join features it matches.
geoprocessing environment	a setting applied around a tool, such as the processing extent.

Further reading

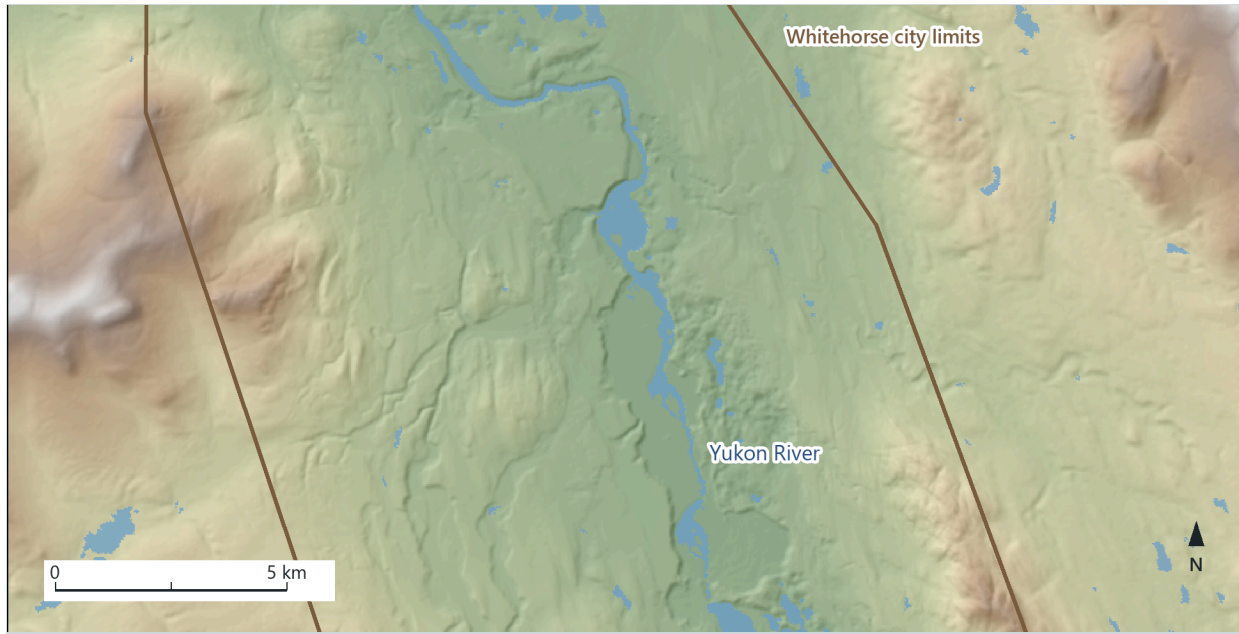
- Olaya, V. *Introduction to GIS*, the chapter on spatial analysis. <https://volaya.github.io/gis-book/en/>
- Fang, Y., Shandas, V., & Arriaga Cordero, E. (2014). *Spatial Thinking in Planning Practice: An Introduction to GIS*. Portland State University Library. <https://pdxscholar.library.pdx.edu/pdxopen/4/>
- Gimond, M. *Intro to GIS and Spatial Analysis*. <https://mgimond.github.io/Spatial/>

- The course's data sources: the nine GeoYukon layers the lessons and this chapter's maps use, published by the Government of Yukon under the Open Government Licence – Yukon.

CHAPTER 8

Raster analysis

Accompanies Module 8 · Raster Analysis



Elevation: Medium Resolution Digital Elevation Model (MRDEM), Natural Resources Canada; contains Copernicus WorldDEM-30 © DLR e.V. 2010-2014 and © Airbus Defence and Space GmbH 2014-2018, provided under COPERNICUS by the European Union and ESA. Water: North America Land Cover 2020 30m, Commission for Environmental Cooperation, NALCMS (CC BY 4.0). City limits: Government of Yukon, GeoYukon. Contains information licensed under the Open Government Licence – Canada and the Open Government Licence – Yukon.

WHY THIS MATTERS

Vector data describes discrete features: wells, roads, parcels. Raster data describes the ground as a surface, one number for every square of it: a height, a land-cover class, a satellite's brightness. The map above is only numbers on a 30 m grid over Whitehorse, drawn in colour and as shading. This chapter explains the grid, how to draw it to suit its values, and how arithmetic on whole grids answers questions that cannot practically be measured in the field.

YOU SHOULD BE ABLE TO EXPLAIN

- why an integer raster has a table of counts while an elevation surface has only statistics
- why halving the cell size quadruples the data, and why resampling to smaller cells adds no detail
- why multiplying Boolean rasters finds the cells that pass every test, and what a NoData cell means

IN THIS CHAPTER

The raster data model

Cell size is resolution

Bands and images

Drawing a raster

Making your own copy

Terrain: hillshade, slope and aspect

Map algebra

Boolean rasters

Reclassify

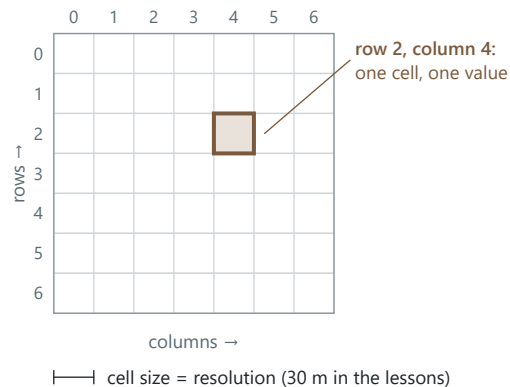
Suitability: the habitat chain

The raster data model

A **raster** covers an area with a grid of equal squares, the **cells** (in an image, pixels), arranged in rows and columns. Every cell holds exactly one value, and the value describes the ground at that cell's location: an elevation, a temperature, a rainfall total, a land-cover class, the reflected light a sensor recorded. There are no features and no rows of attributes as in a vector layer; the position of a value in the grid is its location. Rasters are therefore well suited to things that vary continuously over a surface, where every place has a value, and to all images.

Rows, columns, cells

every cell the same size, one value each



Integer: a table of counts

land-cover codes

1	1	1	8	16
1	1	8	8	16
18	1	1	8	8
18	18	17	1	1
18	17	17	1	1

Value	Count	Class
1	11	Needleleaf forest
8	5	Shrubland
16	2	Barren land
17	3	Urban
18	4	Water

one row per value, not per cell: 5 rows for 25 cells

Floating point: statistics instead

702.4	715.9	741.2	788.6	846.1
695.8	709.3	730.7	769.0	820.5
688.1	699.6	717.4	748.3	791.8

elevation in metres

minimum 688.1

maximum 846.1

mean 744.3

15 different values in 15 cells:

nothing to count, no table

Figure 8.1 The raster model. Left: rows, columns and cells, each cell holding one value. Top right: an integer raster of land-cover codes and its attribute table, one row per value with a count of cells. Bottom right: a floating-point raster of elevations, where almost every cell differs and statistics take the place of a table.

Cell values come in two kinds, and the kind determines what you can do with a raster. An **integer raster** holds whole numbers, usually codes: 1 for needleleaf forest, 18 for water. A **floating-point raster** holds decimals, usually measurements: 694.7 m in one cell, 695.2 m in the next. The land-cover layer the course uses stores each cell as an 8-bit whole number, 0 to 255, which is room for 256 codes; its elevation layers store 32-bit decimals.

Only an integer raster can have a **raster attribute table**. The table has one row per *value*, not one row per cell, and two fields that are always there: **Value** and **Count**, the number of cells holding it. Any other column, such as a class name, is an addition someone made. A floating-point raster has nothing to count by value, because hardly any two cells share a value, so it has no table; you read it through its statistics instead, such as the minimum, maximum and mean. The same distinction applies to selection: you can select `value = 18` on an integer raster, but there is no row to select on an elevation surface.

You query a raster by pointing at it. When you click a place, the software reports the value of the cell under the cursor for every raster drawn there. On the valley floor at Whitehorse the elevation cells hold values of about 640 m, and on the benches above the river about 700 m.

IN THE PANE

Module 8 lesson 1 is a looking lesson. You add the Commission for Environmental Cooperation's land cover of North America and zoom in until the individual square cells are visible, read its cell size, pixel type and band count off the layer's **Source** tab, compare it with the floating-point elevation layer, see the table of counts Waldo fetches from the service, and click the ground with the Explore tool to read cell values.

Cell size is resolution

Every cell covers the same square of ground, and the length of its side is the **cell size**, the raster's **spatial resolution**. A 30 m cell represents $30 \times 30 = 900 \text{ m}^2$ of ground, and the raster cannot represent anything smaller than a cell. A pond 20 m across either determines the value of the cell it falls in or is lost among the surrounding cells. A land-cover grid behaves in this way, but in an image a small bright object can make a pixel brighter without being recorded as a separate feature.

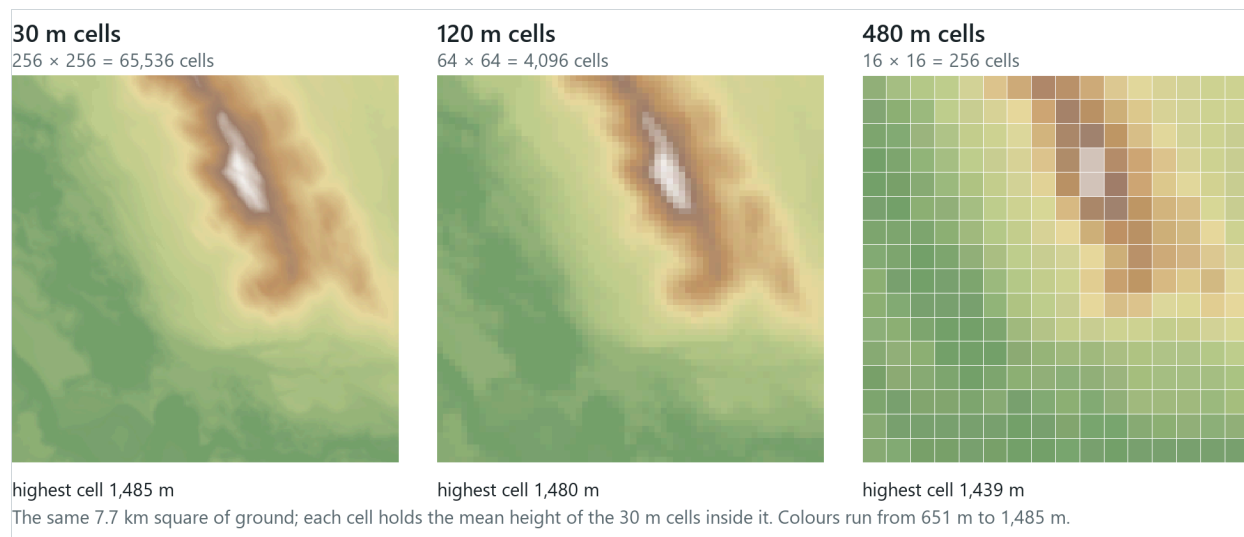


Figure 8.2 One hill east of the Yukon River at three cell sizes. The same 7.7 km square holds 65,536 cells at 30 m, 4,096 at 120 m and 256 at 480 m; the coarse cells average the fine ones, so the summit has values of 1,485 m, 1,480 m and 1,439 m.

Elevation: MRDEM, Natural Resources Canada (contains Copernicus WorldDEM-30 © DLR e.V. 2010-2014 and © Airbus Defence and Space GmbH 2014-2018, provided under COPERNICUS by the European Union and ESA). Contains information licensed under the Open Government Licence – Canada.

Halving the cell size quadruples the number of cells, and with it the storage and the processing time, so a finer resolution has a cost. The course's elevation model of Whitehorse, 689 columns by 1,080 rows of 30 m cells in 4-byte decimals, is about 3 MB, and at 15 m it would be about 12 MB. Coarser cells are cheaper but smooth the surface. In the hill above, the summit is almost 50 m lower at 480 m cells because the coarse cells in the figure were made by averaging the fine ones (how a real DEM's cell value relates to the ground inside it depends on how it was made), and small features disappear completely. Choose a cell size that can represent the smallest feature the analysis needs. Cell size, the detail the source supports and the accuracy of the values are separate properties. Resampling a 30 m grid to 15 m cells makes four cells where there was one and adds no detail.

Bands and images

A raster can carry one value per cell or several. The elevation model and the land cover have one **band**. A satellite image has many: most bands record the light reflected in one slice of the spectrum, over the same cells, and some record the heat the ground gives off. The Landsat 8 and 9 satellites record eleven bands: nine of reflected light (coastal aerosol, blue, green, red, near infrared, two shortwave infrared and a cirrus band at 30 m, and a panchromatic band at 15 m) and two thermal bands captured at 100 m and delivered resampled to 30 m. The course's Multispectral Landsat service also holds eleven numbers per cell, but not the same eleven: it leaves out the panchromatic band and adds a quality band, a code describing clouds and other problems with the observation. The eleven bands of a multiband raster are therefore not necessarily measurements of one kind at one resolution.

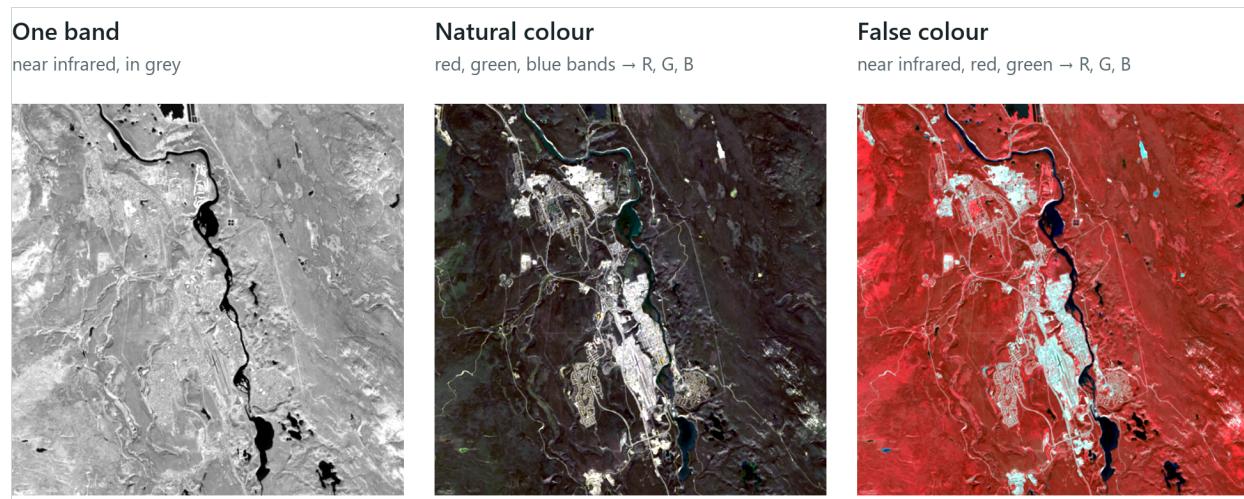


Figure 8.3 The same 16.5 km square of Whitehorse from one Landsat 9 scene of 5 August 2024. Left: the near-infrared band alone, drawn in grey. Middle: the red, green and blue bands sent to the screen's red, green and blue, a natural-colour composite. Right: near infrared, red and green sent to red, green and blue, a false-colour composite in which healthy vegetation appears red.

Landsat 9 image (5 August 2024, scene LC09_L2SP_059017_20240805) courtesy of the U.S. Geological Survey, via Microsoft Planetary Computer.

A screen shows colour by mixing red, green and blue light, so a colour image of a raster uses three of its bands, one in each of the screen's channels, and you choose which three. Red, green and blue bands sent to red, green and blue give **natural colour**, the picture the eye would see. Shifting each assignment one band along the spectrum, with near infrared in red, red in green and green in blue, gives the standard **false-colour composite**. Plants reflect near-infrared light strongly, so healthy forests and lawns appear red, while the town, the gravel and the water appear pale blue and black. Arithmetic on two bands also produces new rasters: the vegetation index **NDVI** is $(\text{NIR} - \text{Red}) / (\text{NIR} + \text{Red})$, computed cell by cell, which is map algebra (below).

Land cover is a raster made from such images. The North America Land Cover 2020 layer the course uses was classified from Landsat 8 imagery by Canadian, US and Mexican agencies into 19 classes at 30 m. Each cell holds one class code, which makes it an integer raster with a table; inside Whitehorse's city limits eight classes occur, and needleleaf forest covers 70 % of the 420 km².

Drawing a raster

A raster is drawn by a **colorizer**, and the right one depends on what the numbers mean. **Unique values** gives each code its own colour. It suits categories, where a value "between shrubland (8) and grassland (10)" has no meaning. **Stretch** spreads a colour ramp from the lowest value to the highest, so that every cell gets its own shade, with no classes. It suits a continuous surface that is read by eye. **Classify** groups the values into a few classes (natural breaks by default) with a readable legend. It suits a map from which a reader must be able to say "this area is above 1,264 m". **RGB** sends three bands to the screen's three colours, as in the images above.

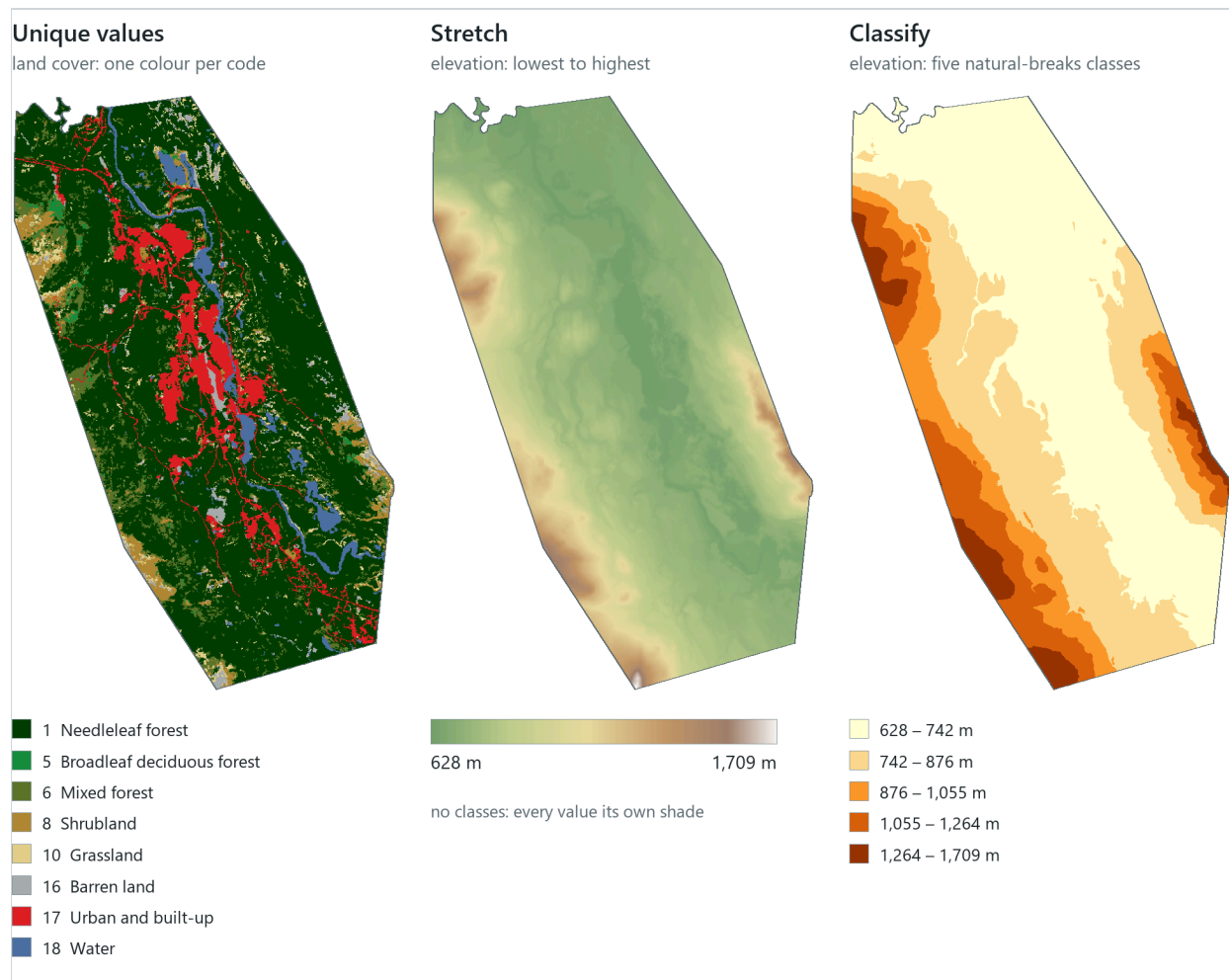


Figure 8.4 Three colorizers on Whitehorse, clipped to the city limits. Left: land cover by unique values, one colour per class code. Middle: elevation stretched from 628 m to 1,709 m. Right: the same elevation in five natural-breaks classes.

Land cover: North America Land Cover 2020 30m, Commission for Environmental Cooperation, NALCMS (CC BY 4.0), in the CEC's own class colours. Elevation: MRDEM, Natural Resources Canada (contains Copernicus WorldDEM-30 © DLR e.V. 2010-2014 and © Airbus Defence and Space GmbH 2014-2018, provided under COPERNICUS by the European Union and ESA). City limits: Government of Yukon, GeoYukon. Contains information licensed under the Open Government Licence – Canada and the Open Government Licence – Yukon. Colours: ColorBrewer YlOrBr, Cynthia Brewer (Apache 2.0)

The software offers every colorizer for every raster, but only some of them represent the data correctly. A stretch on land-cover codes would draw water (18) as "more" than forest (1); unique values on an elevation surface would need thousands of colours.

RULE OF THUMB

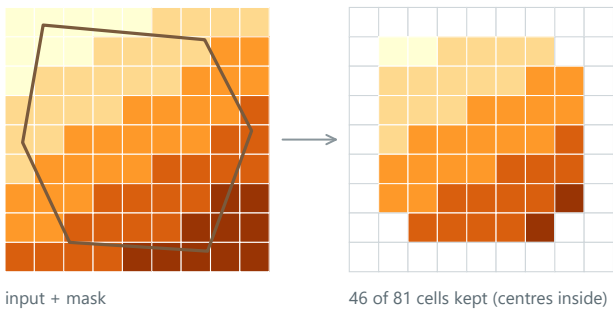
Unique values for integer categories, stretch or classify for floating-point quantities, RGB for multiband images. When a map has to be read in numbers, classify it so that the legend shows the class values.

Making your own copy

Rasters from the internet often cover a continent and are published by someone else. For analysis you need a copy of your own, cut to the study area, in the project's coordinate system, at a cell size you chose, and saved in your geodatabase. **Extract by Mask** is the raster equivalent of Clip. Given a raster and a polygon (or another raster), it keeps the cells inside, and the cells outside become **NoData**, cells that hold no value at all. Exporting a raster with a clipping geometry does the same job without the Spatial Analyst extension.

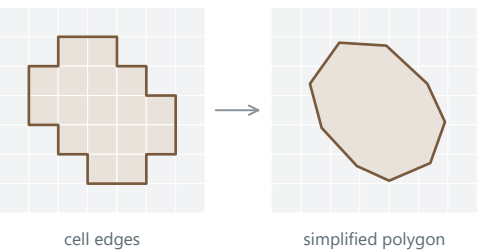
Extract by Mask

a polygon cuts a raster: cells outside become NoData



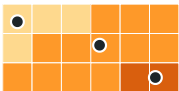
Raster to Polygon

integer cells → shapes; Simplify smooths the steps



Extract Values to Points

each point takes the value of the cell under it and writes it into its own table as a new field



well	dem_m
W1	699.0
W2	761.9
W3	790.3

Figure 8.5 Cutting and converting. Top left: Extract by Mask keeps the cells whose centres fall inside the polygon and sets the rest to NoData. Top right: Raster to Polygon turns groups of equal integer cells into shapes; the simplify option smooths the stepped cell edges. Bottom: Extract Values to Points gives each point the value of the cell under it, in a new field of its own table.

Colours: ColorBrewer YlOrBr, Cynthia Brewer (Apache 2.0)

Three settings, the raster **environments**, determine the properties of the copy. The first is the **output coordinate system**. Web Mercator stretches lengths by $1/\cos(\text{latitude})$ (Chapter 3), so at Whitehorse's 60° N a "30 m" Web Mercator cell is only 15 m of real ground. The course projects to NAD 1983 Yukon Albers instead. It is an equal-area projection designed for the territory, so a 30 m cell covers 900 m² of ground everywhere in it and counting cells gives correct areas, although its distances are true only near its standard parallels, as in any projection (Chapter 3). The second is the **cell size**. An elevation service that combines the finest data it has may report cells a fraction of a metre wide, and an extraction without a cell size of your own would try to copy all of them. The third is the **snap raster**. Two rasters with the same cell size can still be offset by part of a cell, and snapping aligns the new grid's cells exactly with an existing raster's cells, so that when one raster is later multiplied by another, every cell has a matching cell.

When a grid is projected or re-aligned, every new cell needs a value taken from the old ones, and the method must suit the data. A category keeps the code of the nearest old cell (nearest neighbour), because the average of codes 1 and 18 is not a land-cover class. A measured surface such as elevation may be interpolated between neighbouring cells (bilinear), which smooths it a little. In either case, resampling changes the cell values.

RULE OF THUMB

Before extracting, set the output coordinate system and the cell size, and snap to the raster you will combine it with. Then read the result back: cell size, coordinate system, rows and columns.

IN THE PANE

Module 8 lesson 2 makes the learner's own dem with Export Raster (Yukon Albers, 30 m cells, clipped to the city) and landcover with Extract by Mask, snapped to the DEM. It then adds Multispectral Landsat, turns off its processing template to reach all eleven bands, switches the band pickers to false colour, and compares stretch and classify on the DEM with the land cover's unique values.

Terrain: hillshade, slope and aspect

A **digital elevation model** (DEM) is a raster of heights. From it a family of new surfaces can be derived, each by a tool that reads the DEM and writes a new raster, cell by cell.

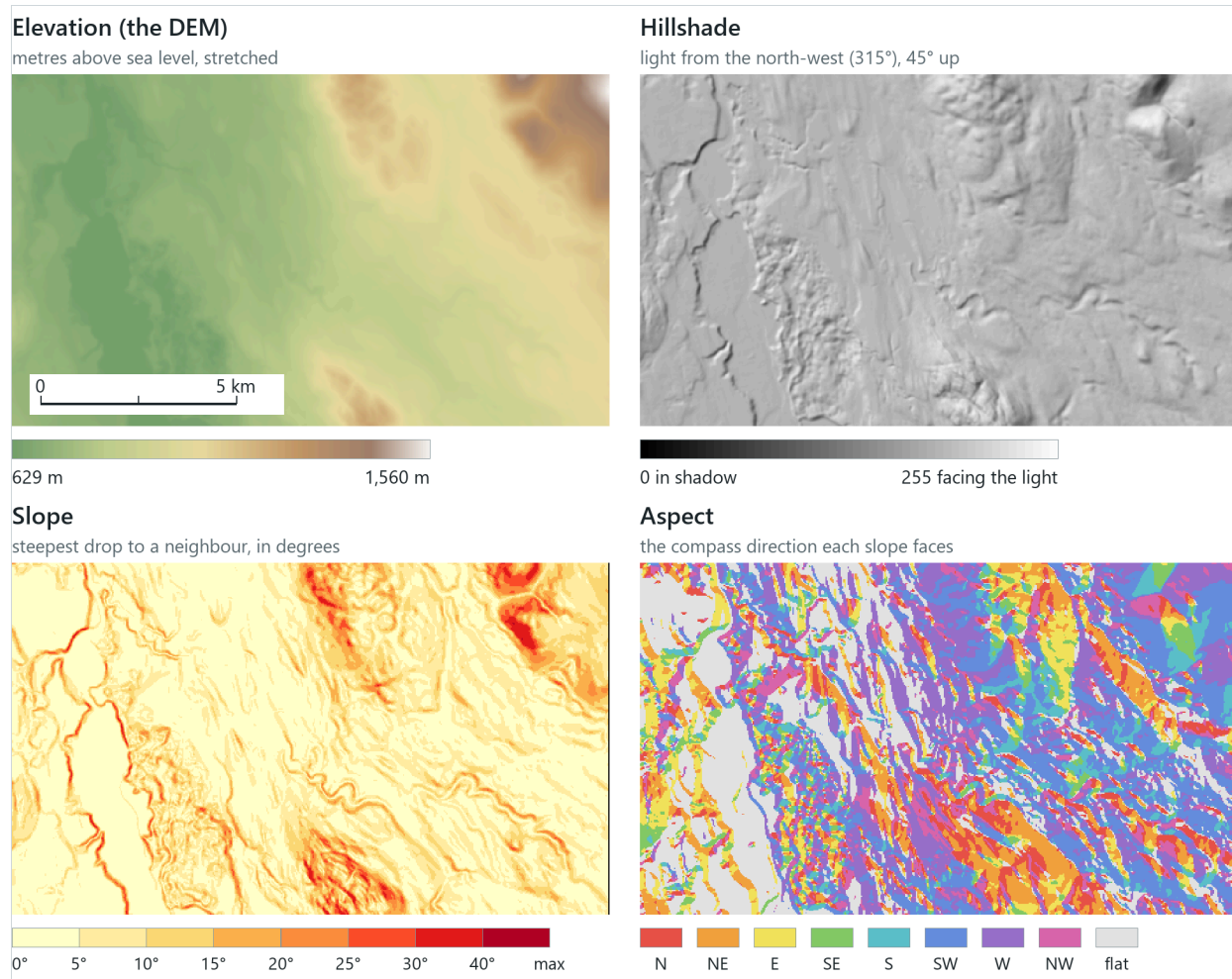


Figure 8.6 Four surfaces over the Yukon River and the hills east of it. The DEM (elevation, 629 to 1,564 m in this window); its hillshade, lit from the north-west at 45°; slope, the steepest drop from each cell to a neighbour, in degrees; aspect, the compass direction each slope faces, with near-flat ground in grey.

Elevation: MRDEM, Natural Resources Canada (contains Copernicus WorldDEM-30 © DLR e.V. 2010-2014 and © Airbus Defence and Space GmbH 2014-2018, provided under COPERNICUS by the European Union and ESA). Contains information licensed under the Open Government Licence – Canada. Colours: ColorBrewer YlOrRd, Cynthia Brewer (Apache 2.0)

A **hillshade** shows the pattern of light and shadow on the terrain. The tool places a light source at a given direction and height, by convention an azimuth of 315° (the north-west) and an altitude of 45°, and records how brightly each cell's slope would be lit, from 0 in full shadow to 255 facing the light. The light comes from the north-west because a light from the south makes most readers see valleys as ridges. A hillshade contains no heights, only brightness values.

The **relief map** is the combination nearly every terrain map uses: the DEM, in an elevation colour scheme, drawn about 40 % transparent over its own hillshade, so that the colour shows the height and the shading shows the shape of the land. The map on the opening page of this chapter is made in this way.

A hillshade can look like a map of steepness, but it is not one, because a steep face is bright or dark depending on the direction it faces. **Slope** measures steepness. For each cell the tool compares the cell with its eight neighbours, finds the steepest change in height, and writes it as an angle: a rise of 25 m over a run of 60 m is a slope of about 23°, or 42 % (the same hill written as rise over run times 100). Slope in Whitehorse runs from 0 on the valley floor to 51.5° on the mountain faces, with a

mean of only 7°, which reflects a flat-floored valley with mountains at its edges. **Aspect** gives the compass direction each slope faces, 0° to 360° clockwise from north, and flat cells, which have no direction, are marked -1. In the Yukon, the snow melts first on south-facing slopes.

IN THE PANE

Module 8 lesson 3 runs Hillshade (315° and 45°) on the learner's DEM and builds the relief map by setting the DEM to 40 % transparency over it. It explains slope as rise over run, runs Slope in degrees and asks for the steepest cell, names Aspect as an optional run, and reclassifies the slope into three classes.

NOTE

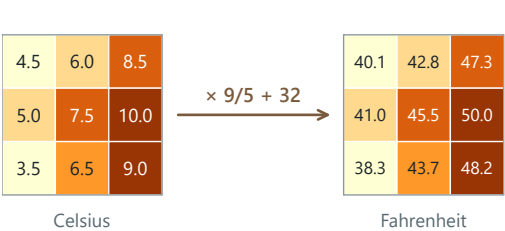
The book's rasters are made with the lessons' own tools and settings but from public sources: Natural Resources Canada's 30 m elevation model instead of the lessons' elevation service, and the same USGS Landsat imagery from a public copy. The two elevation products agree on the valley floor but not to the metre on the summits, so the book's numbers differ slightly from the ones the lessons report.

Map algebra

A raster is a grid of numbers, and two rasters with the same cells align value over value. This means you can do arithmetic on them, cell by cell: multiply a whole grid by a constant, add one raster to another, subtract, or compare every cell with a threshold. This is called **map algebra**, and each operation writes a new raster.

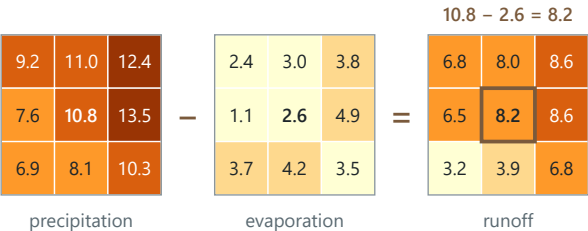
One raster in

every cell through the same formula



Two rasters in

matching cells lined up, cell by cell



The Raster Calculator's grammar

"dem" * 3.28084	a raster's name in double quotes; + − * / for arithmetic
"dem" > 1100	comparisons > < >= <= and == (two equals signs) give 1 or 0
("dem" > 1100) & ("slope" < 15)	& is AND, is OR, ~ is NOT; each side in parentheses

Figure 8.7 Map algebra. Left: a single input raster, with every cell passed through the same formula (Celsius to Fahrenheit). Right: two input rasters, combined cell by cell (precipitation minus evaporation, a simplified water balance; the centre cell is worked out above the grids). Bottom: the Raster Calculator's grammar.
Colours: ColorBrewer YlOrBr, Cynthia Brewer (Apache 2.0)

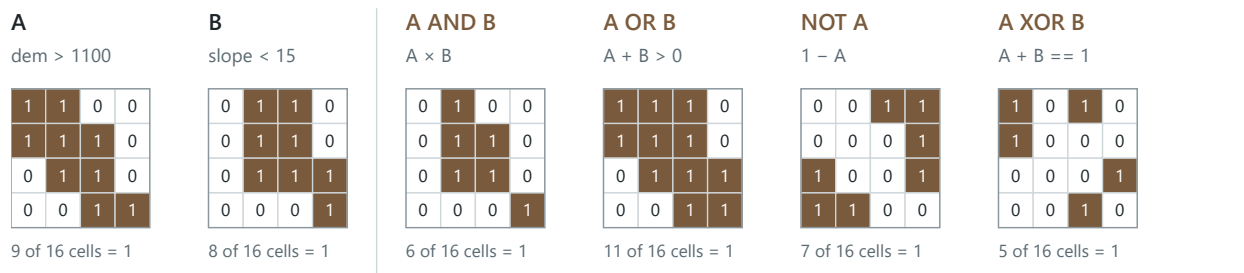
With one input raster, the operation is a conversion: degrees Celsius $\times 9/5 + 32$ gives degrees Fahrenheit in every cell, and the course's "dem" * 3.28084 turns an elevation model in metres into one in feet (Whitehorse's 628 to 1,709 m become 2,062 to 5,608 ft). With two or more input rasters, the operation is an overlay: the figure subtracts evaporation from precipitation cell by cell, a

simplified water balance. Real runoff also depends on what the soil stores, what plants take up and what flows in from upslope, so the subtraction is an example of two-raster arithmetic and is not a hydrological model.

ArcGIS Pro's calculator for this is the **Raster Calculator**, and its grammar is strict. A raster's name goes in double quotes ("dem"); arithmetic is + - * /; comparisons are > < >= <= and ==, two equals signs, since one equals sign is an error; AND is &, OR is |, NOT is ~, and each side of an & or | is placed in its own parentheses. The most common mistakes are a name without quotes, = in place of ==, and the word AND in place of &.

Boolean rasters

A comparison applied to every cell gives a **Boolean raster**: 1 where the condition is true, 0 where it is false. "dem" > 1100 marks every cell above 1,100 m, roughly the tree line on the mountains around Whitehorse; "slope" < 15 marks every cell gentler than 15°. A Boolean raster is an integer raster with a two-row table, and the Count beside the value 1 gives the size of the result: 55,538 cells, or 50.0 km², lie above 1,100 m inside the city.



Cell by cell

A	B	AND A × B	OR A + B > 0	XOR A + B == 1	NOT A 1 - A
0	0	0	0	0	1
0	1	0	1	1	1
1	0	0	1	1	0
1	1	1	1	0	0

Figure 8.8 Two Boolean rasters, A and B, and four ways to combine them, cell by cell, with the arithmetic that produces each. Filled cells are 1 (true).

Because the values are only 0 and 1, the logical operators are ordinary arithmetic. **AND** is multiplication: $A \times B$ is 1 only where both are 1. **OR** is addition tested against zero: $A + B > 0$ wherever either is 1. **NOT** is subtraction from one: $1 - A$. **XOR**, "one or the other but not both", is $A + B == 1$. As a result, $(\text{"dem"} > 1100) \& (\text{"slope"} < 15)$ and $\text{"high"} * \text{"gentle"}$ give the same raster.

A cell that is NoData in either input stays NoData in the result, because arithmetic with an unknown value gives an unknown value rather than 0. A blank cell in a suitability map therefore means that there is no evidence for that cell, which is different from "unsuitable", and the area of NoData is worth reporting beside the area of 1s.

IN THE PANE

Module 8 lesson 4 teaches the Raster Calculator's grammar in detail, runs one conversion ("dem" * 3.28084 into dem_ft, compared legend against legend), and then two Boolean rasters, high and gentle, whose two-row tables Waldo reads back with the count of 1s and its area.

Reclassify

Reclassify replaces one set of values with another. Given ranges, it groups a continuous surface into classes: slope from 0 to 5° becomes 1, 5 to 15° becomes 2, 15 to 90° becomes 3, which a planner can read as flat, moderate and steep. Given values, it recodes categories: every land-cover code becomes 1 or 0. The output is always an integer raster, whatever the input, so Reclassify is the usual way to turn a floating-point surface into a raster with a table that you can count, select or convert to polygons.

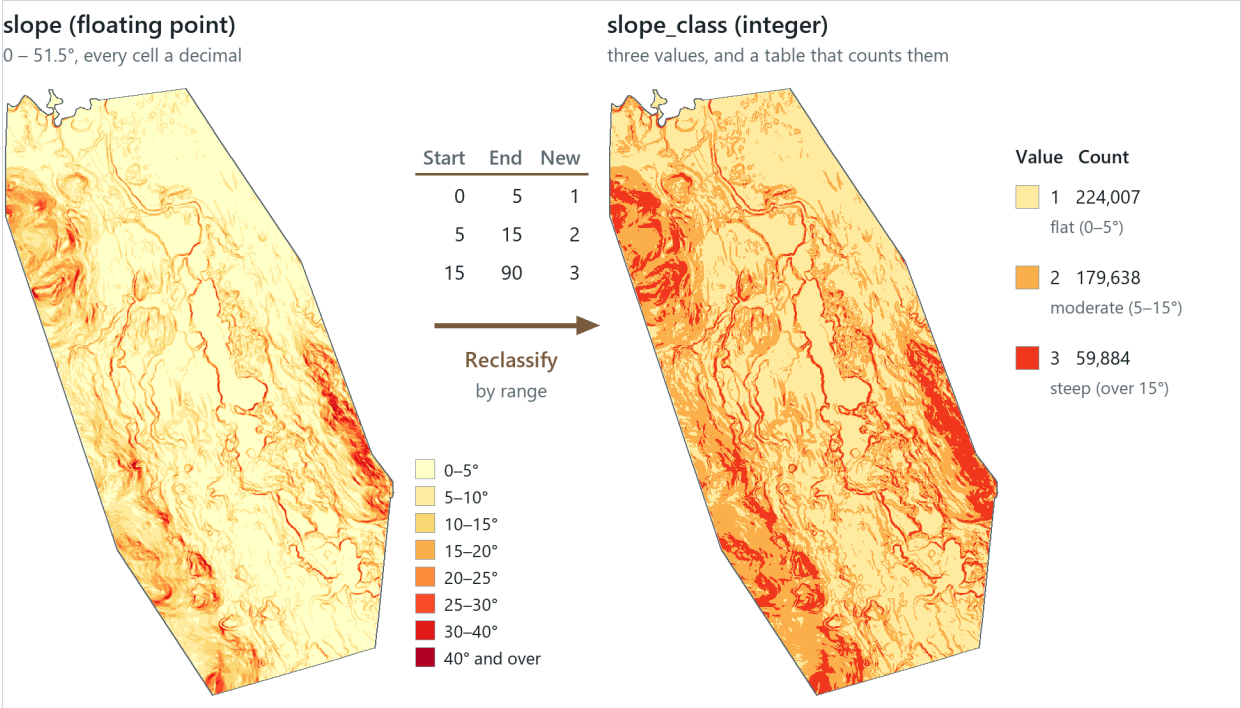


Figure 8.9 Reclassify by range. The floating-point slope of Whitehorse becomes an integer raster of three classes, and the new raster has a table: 224,007 flat cells, 179,638 moderate and 59,884 steep.

Slope derived from MRDEM, Natural Resources Canada (contains Copernicus WorldDEM-30 © DLR e.V. 2010-2014 and © Airbus Defence and Space GmbH 2014-2018, provided under COPERNICUS by the European Union and ESA). City limits: Government of Yukon, GeoYukon. Contains information licensed under the Open Government Licence – Canada and the Open Government Licence – Yukon. Colours: ColorBrewer YlOrRd, Cynthia Brewer (Apache 2.0)

On a floating-point raster the ranges must meet: each class ends where the next begins (0–5, 5–15, 15–90). A table written as 0–5, 6–20, 21–90 leaves a slope of 5.5° or 20.4° in no class at all, and what happens to those cells depends on a setting. By default they keep their original value, which then appears among the new codes without explanation, and they are removed only if the tool's change-missing-values-to-NoData option is ticked. In either case the output's table contains values you did not ask for, so reading the table reveals the problem. A shared break belongs to one class: 5

goes to the first range, 5.001 to the second. In Whitehorse the three classes cover 201.6, 161.7 and 53.9 km².

READING THE RESULT

- The question: how much of Whitehorse is flat, moderate or steep?
- The representation: slope in degrees on 30 m cells, split at 5° and 15° into three classes.
- What it shows: 201.6 km² flat, 161.7 km² moderate and 53.9 km² steep.
- What it cannot establish: slope within a 30 m cell, or buildability; the breaks are choices.

RULE OF THUMB

Reclassify ranges on a decimal surface share their break values; make sure every class in the input is given a new value. Then read the output's table, which should contain the classes you asked for and nothing else.

Suitability: the habitat chain

Boolean rasters and multiplication can be combined into a method. You write a question as criteria that each cell either passes or fails, make one Boolean raster per criterion, and multiply them. A cell has the value 1 in the result only if it passes every criterion. Site selection, habitat mapping and risk mapping are done this way with rasters, the raster counterpart of the vector overlay in Chapter 7.

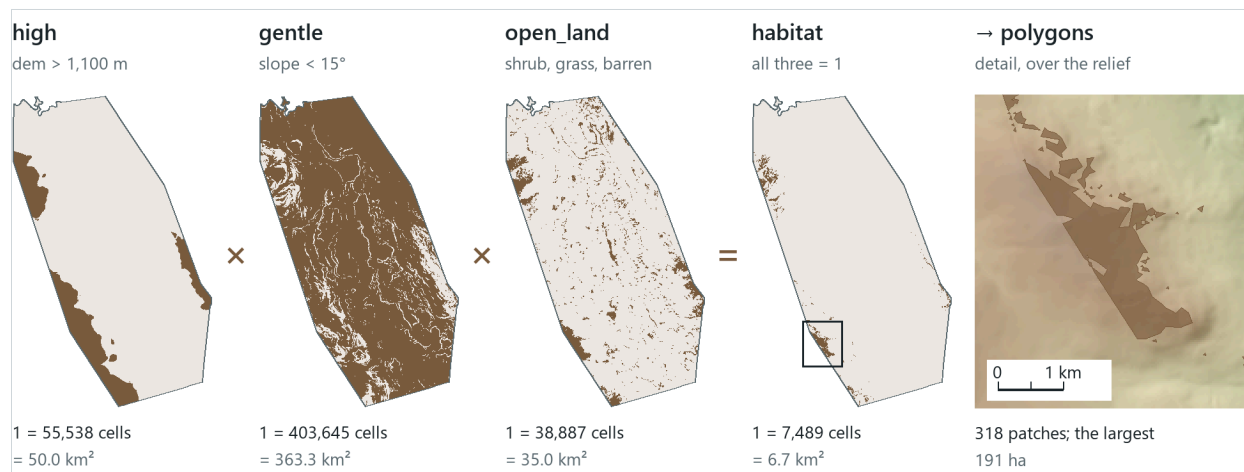


Figure 8.10 The course's habitat question as a chain of rasters over Whitehorse: above 1,100 m, times gentler than 15°, times open ground (shrubland, grassland or barren land) equals habitat, 6.7 km² of the city. Right: the habitat cells converted to polygons around the largest patch, over the relief.

Elevation and slope: MRDEM, Natural Resources Canada (contains Copernicus WorldDEM-30 © DLR e.V. 2010-2014 and © Airbus Defence and Space GmbH 2014-2018, provided under COPERNICUS by the European Union and ESA). Land cover: North America Land Cover 2020 30m, Commission for Environmental Cooperation, NALCMS (CC BY 4.0). City limits: Government of Yukon, GeoYukon. Contains information licensed under the Open Government Licence – Canada and the Open Government Licence – Yukon.

The course asks where Whitehorse's alpine habitat lies, the open ground above the trees where wild sheep graze and ptarmigan nest. There are three criteria: high enough to be alpine (the DEM above 1,100 m), gentle enough to use (slope under 15°), and open rather than forest, town or water (the land cover reclassified by value, shrubland, grassland and barren land to 1, everything else to 0). When the three rasters are multiplied, 7,489 cells, 6.7 km² of the city's 420 km², pass all three. The thresholds were chosen by the course as an illustration and were not set by a biologist. Changing one

threshold changes the map: 1,000 m instead of 1,100 m brings in lower slopes, and 10° instead of 15° removes patches. The sharp edge of the habitat area is set by the thresholds and does not mark where any animal actually lives. Hard exclusions (water, town) and graded preferences (gentler is better) are different kinds of criteria, and the Boolean method treats both as yes or no. If the largest value in the result is above 1, the rasters were added instead of multiplied.

A raster result consists of cells. To buffer it, join it or give it to someone as shapes, you convert it.

Raster to Polygon turns each group of touching cells with the same integer value into a polygon, one row per polygon, and needs an integer raster, since a float would first be reclassified. Converting only the cells with value 1 gives 318 habitat patches in this book's data, the largest 191 ha. The opposite direction exists too: converting features to a raster writes one of their fields into the cells at a chosen cell size, with NoData where no feature lies. Points can also take values from a raster directly. **Extract Values to Points** writes the value of the cell under each point into the point's table. A well and an elevation cell are related only by *where* they are, so only a spatial operation can transfer the value from one to the other.

READING THE RESULT

- The question: where in Whitehorse is open ground that is both high and gentle?
- The representation: three Booleans (above 1,100 m, under 15°, open land cover) multiplied.
- What it shows: 7,489 cells, 6.7 km² of the city's 420 km², pass all three criteria.
- What it cannot establish: where animals live; the edges come from the thresholds.

IN THE PANE

Module 8 lesson 5 explains the habitat question before any tool is run, reclassifies the learner's land cover by class name into open_land, multiplies the three Booleans into habitat in the Raster Calculator, converts the value-1 cells to polygons and asks for the largest patch. Last, Extract Multi Values to Points adds each well's ground elevation from the DEM to Module 7's city_wells as a new field (a learner who started Module 8 in a new project exports the wells into it first, because the tool writes into the layer it is given).

TRY IT

With the Module 8 project open, ask Waldo in Ask mode: what are the cell size, pixel type and value range of my dem, and does it have an attribute table? Waldo reads the raster and answers; check its answer against the integer and floating-point rules in this chapter.

Key terms

raster	a grid of equal cells in rows and columns covering an area, each cell holding one value.
cell (pixel)	one square of a raster, standing for one square of ground and holding one value.
cell size (spatial resolution)	the ground length of one cell's side; nothing smaller than a cell exists in the raster.
integer raster	a raster of whole numbers, usually codes such as land-cover classes; it can have an attribute table.
floating-point raster	a raster of decimal values, usually measurements such as elevation; it has statistics instead of a table.
raster attribute table	one row per cell value of an integer raster, with the fields Value and Count and any added columns.
NoData	a cell that holds no value, such as a cell outside the mask of an extraction.
band	one layer of values in a raster; a multiband image such as Landsat holds several per cell.
false-colour composite	an image with bands other than red, green and blue sent to the screen's colours, such as near infrared into red.
colorizer	the rule that draws a raster: unique values, stretch, classify or RGB.
digital elevation model (DEM)	a raster of ground heights.
hillshade	a raster of the brightness each cell would have under a lamp at a set direction and height.
relief map	a DEM in an elevation colour scheme drawn partly transparent over its own hillshade.
slope	the steepest rate of change of height from a cell to its neighbours, in degrees or percent.
aspect	the compass direction a slope faces, 0° to 360° from north, -1 for flat cells.
map algebra	arithmetic and comparisons applied to rasters cell by cell, each writing a new raster.
Raster Calculator	ArcGIS Pro's tool for map-algebra expressions, with quoted raster names and strict operators.
Boolean raster	a raster of 1 (true) and 0 (false) cells, the answer to a condition asked of every cell.
Reclassify	a tool that replaces values or ranges of values with new values; its output is always integer.
Extract by Mask	a tool that keeps a raster's cells inside a polygon or raster mask and sets the rest to NoData.
snap raster	an environment that aligns a new raster's cells exactly with an existing raster's cells.
Raster to Polygon	a tool that turns groups of equal integer cells into polygons, one row per polygon.
Extract Values to Points	a tool that writes the value of the cell under each point into the point's table.
suitability analysis	answering "where?" from criteria every cell can pass or fail; the course uses the simplest form, one Boolean raster per criterion, multiplied.

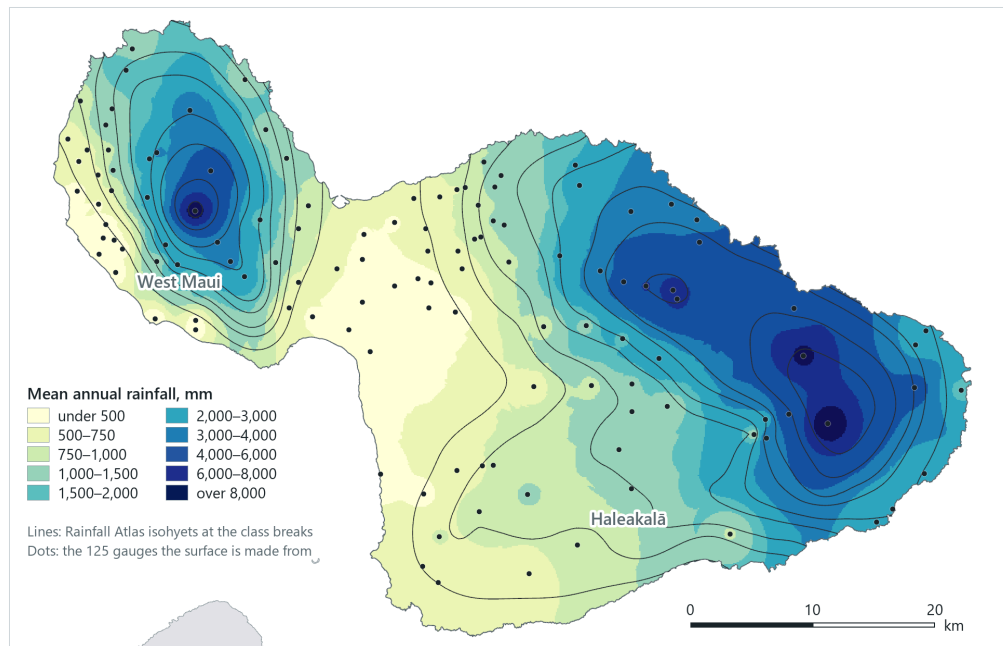
Further reading

- Olaya, V. *Introduction to GIS*, the chapters on the raster model and map algebra.
<https://volaya.github.io/gis-book/en/>
- U.S. Geological Survey. Landsat 8–9 Collection 2 Level-2 surface reflectance.
<https://doi.org/10.5066/P9OGBGM6>
- Natural Resources Canada. *Medium Resolution Digital Elevation Model (MRDEM) – CanElevation Series*.
<https://open.canada.ca/data/en/dataset/18752265-bda3-498c-a4ba-9dfe68cb98da>
- Commission for Environmental Cooperation. *North American Environmental Atlas – Land Cover 2020 30m*. <https://www.arcgis.com/home/item.html?id=eoff7c071cb740409c7fe01f63f4fef8>
- The course's data sources: every layer the lessons and this chapter's maps use, with its publisher and licence.

CHAPTER 9

Spatial patterns

Accompanies Module 9 · Spatial Patterns — Tobler's Law, Interpolation and Density



Coastline: USGS; Hawaii Statewide GIS Program. Rain gauges: Rainfall Atlas of Hawai'i, University of Hawai'i at Mānoa, Department of Geography. Isohyets: Giambelluca et al. (2013), Online Rainfall Atlas of Hawai'i, via the Hawaii Statewide GIS Program (public domain). Colours: ColorBrewer YlGnBu, Cynthia Brewer (Apache 2.0)

WHY THIS MATTERS

Rain is measured at only a few places. On Maui, 125 gauges record it at 125 points, yet the map above gives a value for every hectare of the island. Moving from samples to a surface depends on one idea: values at nearby places are more alike than values at distant places. This chapter applies the idea in three tools (Thiessen polygons, inverse distance weighting and kernel density) and identifies the setting in each that you should report to the readers of your map.

YOU SHOULD BE ABLE TO EXPLAIN

- why near samples are more alike than far ones, and why that allows a whole island to be mapped from 125 gauges
- why an interpolated surface looks equally reliable everywhere, and how you would test it
- why the search radius shapes a density map, and why a density is not a rate

IN THIS CHAPTER

Tobler's first law

Spatial autocorrelation

Rain on Maui

From samples to a surface

Thiessen polygons

Inverse distance weighting

Checking a surface

Kernel density

Tobler's first law

In 1970 the geographer Waldo Tobler, writing about a computer simulation of the growth of Detroit, stated a central principle of geography in one sentence:

"Everything is related to everything else, but near things are more related than distant things." (Waldo Tobler, 1970)

Geographers call it the **first law of geography**, and the course treats it as the most important idea in GIS, because every map made from samples depends on it. (Waldo, the assistant in the pane, is named after him.)

The law is easiest to understand with an example from the weather. If it is raining at your house, you can be almost certain it is raining across the street, and fairly, but less, confident that it is raining across town. Elevation behaves the same way: knowing the height of the ground where you stand, you can predict the height 100 m away well and 500 m away less well. The relationship is real, and it weakens with distance. This weakening is called **distance decay**, and every tool in this chapter measures or uses it.

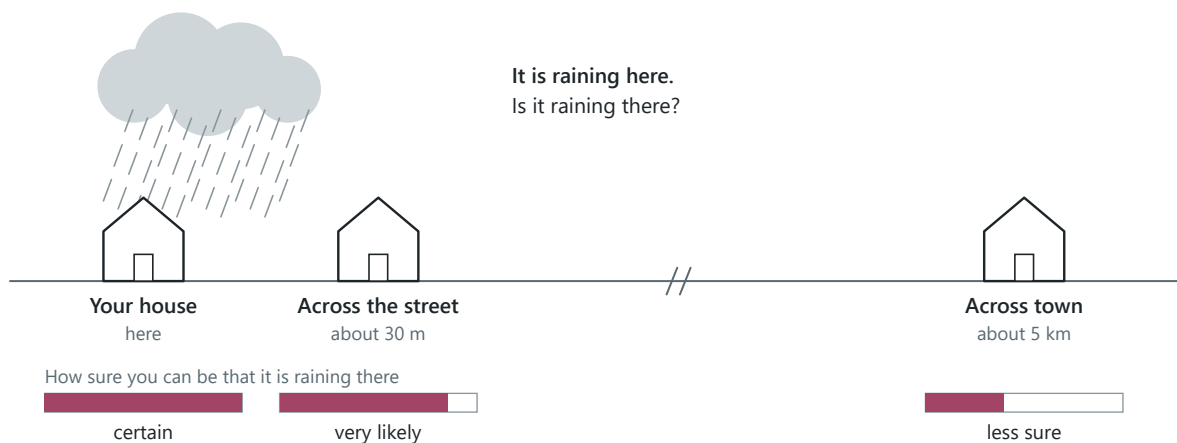


Figure 9.1 Tobler's first law illustrated with rainfall. Rain at your house is a good guide to the rain at the house across the street and a weaker guide to the rain at a house across town, because the similarity weakens with distance.

Spatial autocorrelation

When Tobler's law appears in data, it is called **spatial autocorrelation**: the tendency of values at nearby places to be related. The prefix "auto" indicates that the variable is correlated with itself at neighbouring places, rather than with a second variable. Spatial autocorrelation comes in three kinds.

- **Positive** autocorrelation: similar values lie next to each other, so the map shows clusters and smooth gradients. This is the most common case. Examples include rainfall, elevation, house prices, the votes for one party, the stands of a plant species, and a disease spreading outwards from its first cases, as John Snow's cholera deaths spread around the Broad Street pump (Chapter 6).

- **Random:** values are scattered with no relation to their neighbours. This is the arrangement that chance alone would produce, and the other two kinds are judged against it.
- **Negative** autocorrelation: neighbours tend to be unlike each other, with high values beside low ones. The extreme case is a checkerboard. Negative autocorrelation is rare in nature and is usually the result of human design.

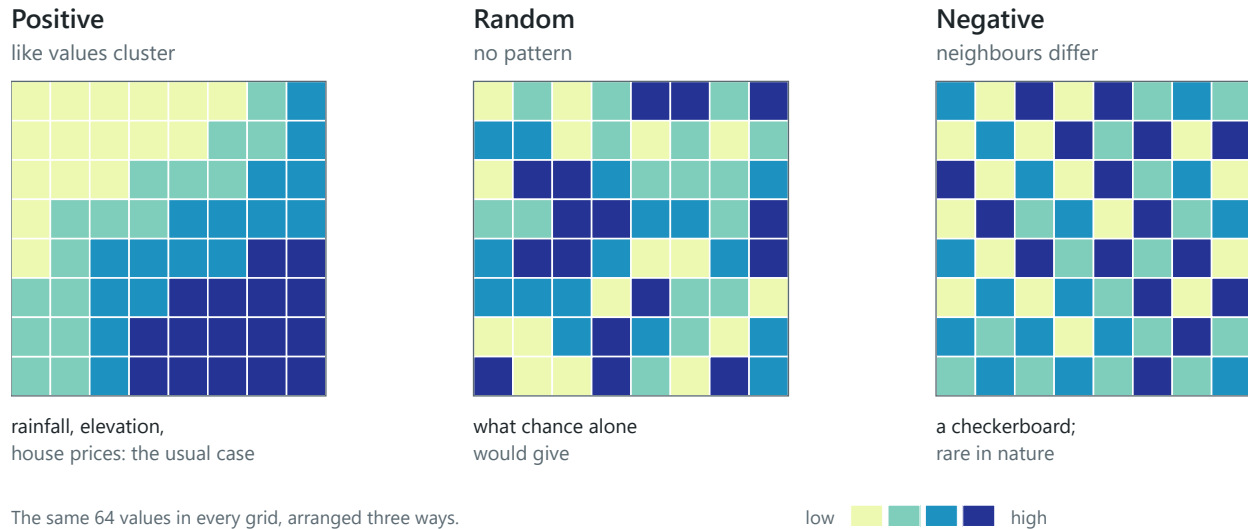


Figure 9.2 Positive, random and negative spatial autocorrelation. Each grid holds the same 64 values; only the arrangement of the values differs, and the arrangement determines the kind of autocorrelation.

Colours: ColorBrewer YlGnBu, Cynthia Brewer (Apache 2.0)

Positive autocorrelation makes it possible to predict values in the gaps between samples. If rainfall were random across an island, a gauge reading would give no information about the ground a kilometre away, and no map could be made from gauges. Because rainfall is positively autocorrelated, a value in between is a reasonable first estimate for a place half-way between a wet gauge and a dry one. It is only an estimate, however: a ridge, the wind or the sea between two gauges can break the pattern, as the example of Maui's rain shadow below shows.

Rain on Maui

Maui, the second-largest of the Hawaiian Islands, is two volcanoes joined by a low valley: the West Maui mountains to the west and the broad dome of Haleakalā to the east. The trade winds blow from the north-east. They push moist air up the windward slopes, where it cools and drops its rain; the leeward coasts to the south and west lie in the rain shadow. The contrast in rainfall is extreme: Big Bog, on Haleakalā's windward slope at 1,645 m, averages 10,199 mm of rain a year, one of the wettest places on Earth. Kahului Airport, 39 km away on the valley floor (computed), gets 487 mm, Kīhei on the south coast 328, and Olowalu on the leeward coast of West Maui 309.

The numbers come from the rain gauges of the 2011 Rainfall Atlas of Hawai'i, served by the state's GIS program. Each gauge's value is its mean annual rainfall over the Atlas's common period, 1978 to 2007, with records adjusted and gap-filled to that period by the Atlas's authors. Each value is therefore a climate average rather than the rainfall of any one year. The layer holds 1,107 gauges

across the islands, each with its mean monthly and annual rainfall. Of Maui's 224, 125 have a usable annual mean; the rest carry a code meaning "no value", which is explained in the next section.

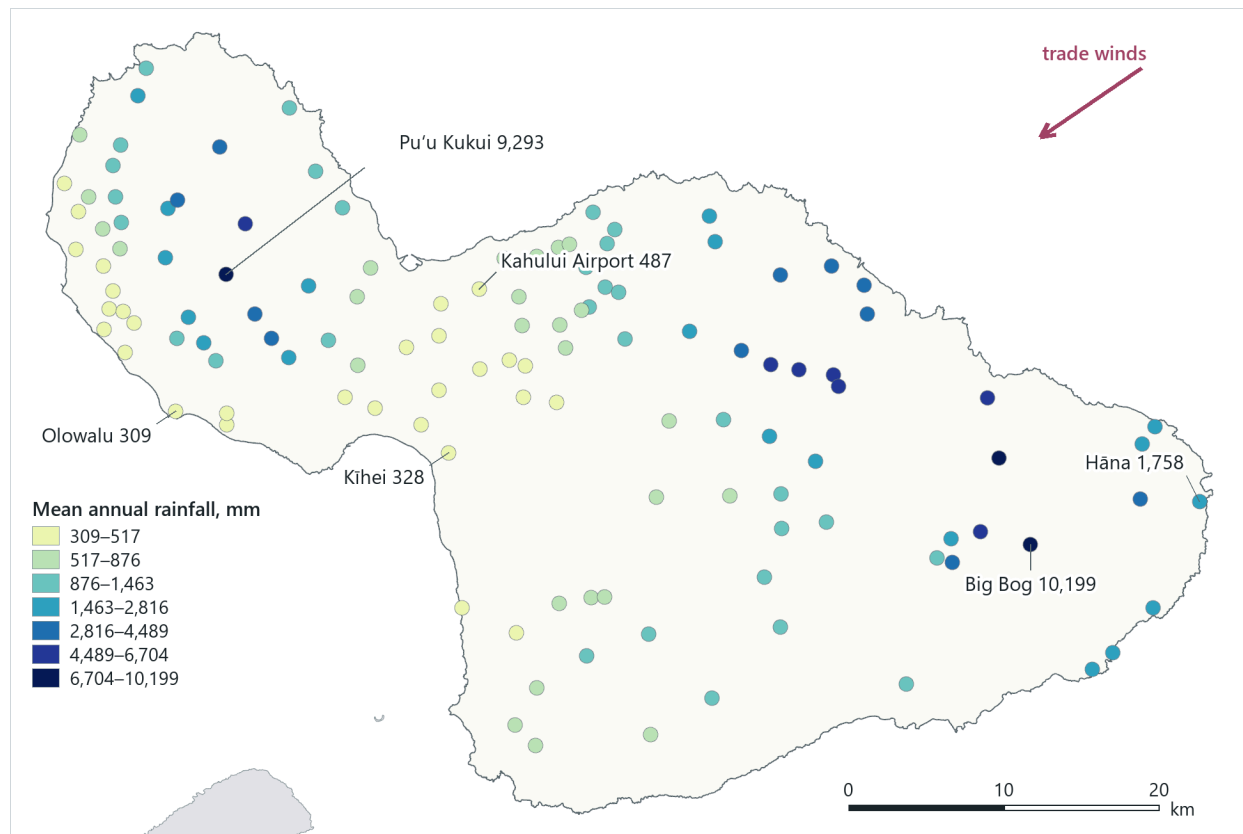


Figure 9.3 Maui's 125 rain gauges, coloured by mean annual rainfall (1978–2007) in seven natural-breaks classes. Wet gauges lie beside other wet gauges on the windward slopes of both mountains, and dry gauges beside other dry gauges on the leeward coasts. No gauge recording 10,000 mm is next to one recording 300 mm.

Coastline: USGS; Hawaii Statewide GIS Program. Rain gauges: Rainfall Atlas of Hawai'i, University of Hawai'i at Mānoa, Department of Geography, via the Hawaii Statewide GIS Program (public domain). Colours: ColorBrewer YIGnBu, Cynthia Brewer (Apache 2.0)

The map shows positive autocorrelation visually, and the pattern can also be measured. Take each gauge and compare its rainfall with that of its nearest neighbour: half of the 125 pairs differ by less than 159 mm. Now compare any two gauges on the island, near or far: the typical difference is 889 mm (computed over all 7,750 pairs). Near gauges are more alike than distant ones, by a factor of more than five. This is a direct measurement of Tobler's law for Maui's rainfall.

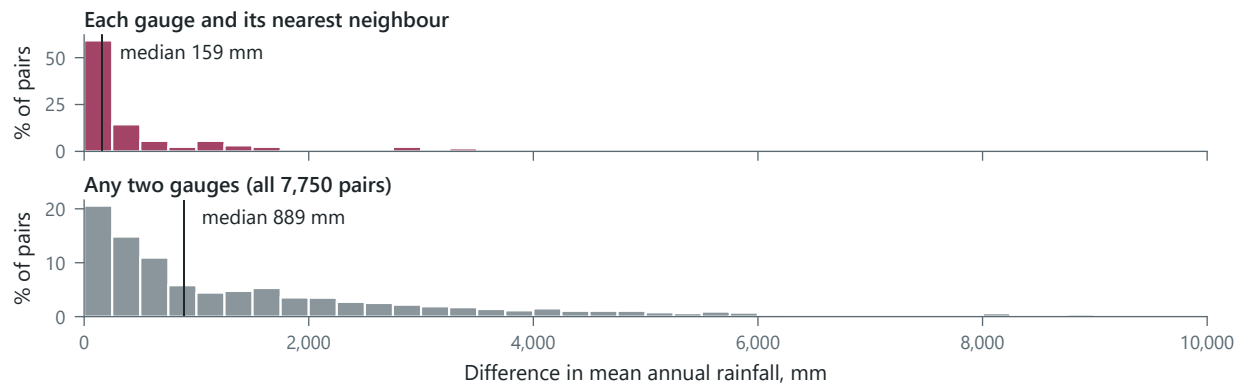


Figure 9.4 Tobler's law measured on the 125 gauges. Top: how much each gauge differs from its nearest neighbour. Bottom: how much any two gauges differ. The nearest-neighbour differences are concentrated close to zero.

Rain gauges: Rainfall Atlas of Hawai'i, University of Hawai'i at Mānoa, Department of Geography (public domain)

IN THE PANE

Module 9 lesson 1 starts a new project, `maui_rainfall`, adds the state's coastline and filters it to Maui, then adds the Rainfall Atlas gauges, finds the missing-value code, keeps the 125 valid gauges on the island and colours them by rainfall. You look for the windward and leeward pattern yourself, and Waldo then gives the measured result, 159 mm between nearest neighbours. The lesson names the three kinds of autocorrelation but does not compute a statistic.

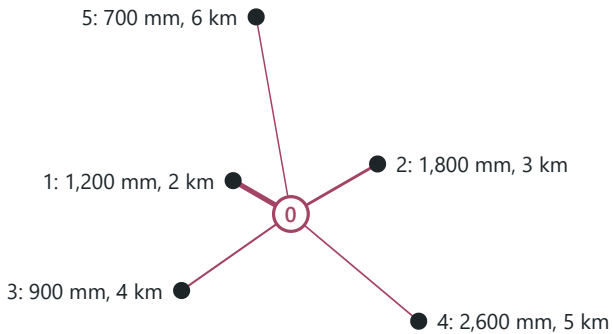
From samples to a surface

Interpolation is estimating a value at a place where nothing was measured from **sample points** where something was: rain gauges, wells, soil cores, air-quality monitors. In a GIS the result is almost always a raster, one estimate per cell, called a **continuous surface**: rainfall, elevation, a chemical concentration or a noise level, known everywhere in the study area instead of at a scatter of points. Every interpolation method depends on Tobler's law. The distance-based methods in this chapter differ in how much weight they give to nearness, and other methods add terrain, wind or other predictors to distance.

The basic problem is the same for every cell. A cell, which we can call point *o*, has no value, but its neighbours do, and the method has to decide how to combine their values.

What is the rainfall at point 0?

Line width = the weight each neighbour gets (power 2)



point	rainfall	distance	1 / d ²	weight
1	1,200	2 km	0.250	50.9 %
2	1,800	3 km	0.111	22.6 %
3	900	4 km	0.062	12.7 %
4	2,600	5 km	0.040	8.1 %
5	700	6 km	0.028	5.7 %

Estimate = the weighted average	1,383 mm
Nearest neighbour only (Thiessen)	1,200 mm
Power 1 (far points count more)	1,422 mm
Power 3 (near points count more)	1,339 mm

Figure 9.5 Estimating point 0 from five neighbouring gauges. Nearer neighbours receive more weight: with inverse distance weighting at power 2 the nearest gauge carries half the weight and the estimate is 1,383 mm. Using only the nearest gauge gives 1,200 mm, and a lower or higher power gives other results (computed, illustrative values).

Missing values are data until you remove them

Before you run any method, the samples must be cleaned. Of the Atlas's 1,107 gauges, 583 have no annual mean. Instead of leaving that field empty, the layer stores **−9999**, a **sentinel value** meaning "no data here". A person recognises the code, but an averaging tool treats it as an ordinary reading, as if the gauge had recorded minus 9,999 mm of rain. On Maui, 99 gauges carry the code. If they are left in, the interpolated surface ranges from −9,999 mm to 10,199 mm, with a mean of −2,693 mm of rain a year, and the software gives no warning.

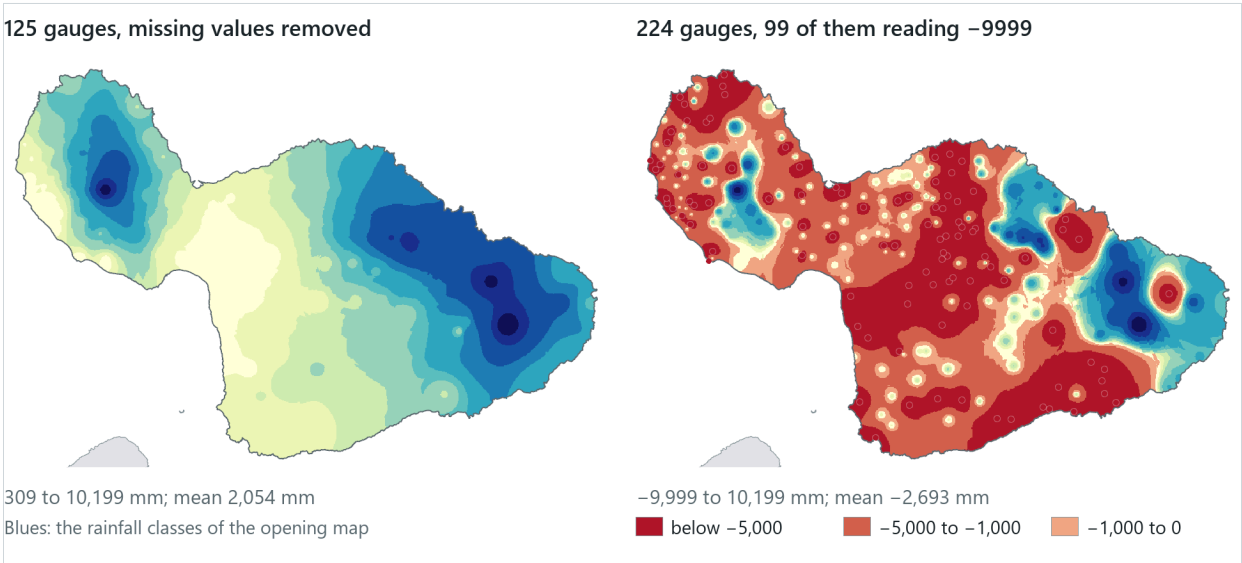


Figure 9.6 The same interpolation with and without the missing-value code. Left, the 125 valid gauges: 309 to 10,199 mm. Right, all 224 gauges on Maui, 99 of them reading −9999: most of the island has values below zero.

Coastline: USGS; Hawaii Statewide GIS Program. Rain gauges: Rainfall Atlas of Hawai'i, University of Hawai'i at Mānoa, Department of Geography, via the Hawaii Statewide GIS Program (public domain). Colours: ColorBrewer YIGnBu and RdBu, Cynthia Brewer (Apache 2.0)

RULE OF THUMB

Before you interpolate, summarise, or average any sampled data, sort the value field and look at both ends. Codes such as -9999, -999 or 0 for "not measured" must be filtered out first, usually with a definition query that keeps only real values.

Thiessen polygons

The simplest way to estimate point o is to use only the nearest sample. Drawing that rule on a map produces **Thiessen polygons**, named after the meteorologist A. H. Thiessen, who used them in 1911 to average rainfall over an area from scattered gauges. They are also called **Voronoi polygons** or proximal zones.

Around each gauge, the polygon's edges run half-way to each neighbouring gauge, so every location inside it is closer to that gauge than to any other. Together the polygons tile the whole study area with no gaps and no overlaps. When each polygon is given its gauge's rainfall, every place on the island has a value. The result is a set of vector polygons rather than a raster. It is also the special case of the next method in which only the single nearest point is used.

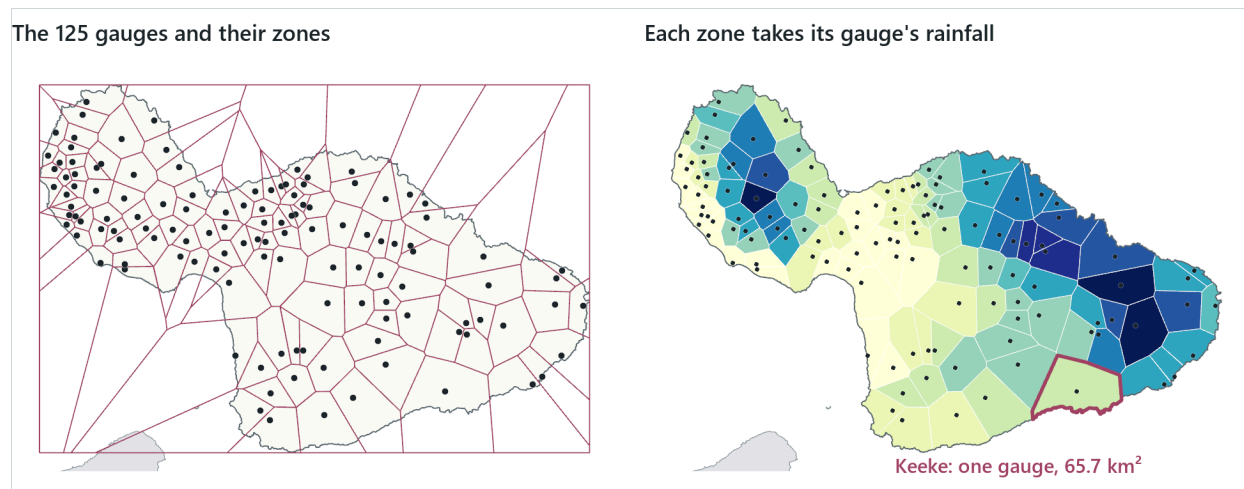


Figure 9.7 From 125 gauges to 125 zones. Left, the Thiessen edges, drawn out to the rectangle around the island. Right, the zones clipped to the coast and coloured by their gauge's rainfall, in the same classes as the opening map. Keeke's zone, on the dry south slope, covers 65.7 km^2 and is represented by a single gauge.

Coastline: USGS; Hawaii Statewide GIS Program. Rain gauges: Rainfall Atlas of Hawai'i, University of Hawai'i at Mānoa, Department of Geography, via the Hawaii Statewide GIS Program (public domain). Colours: ColorBrewer YIGnBu, Cynthia Brewer (Apache 2.0)

The map is crude but covers the whole island, and it shows two things. First, its level of detail depends on the density of the samples. On the windward coast east of Kahului, where the sugar and pineapple plantations once kept a gauge on nearly every field, the zones are tiny. On Haleakalā's dry southern slope one discontinued gauge, Keeke, represents 65.7 km^2 , the largest zone on the island; the smallest is 2.66 km^2 . Second, the sharp steps at the zone edges are produced by the method: rainfall does not jump from 6,000 mm to 2,000 mm at a line half-way between two gauges.

Thiessen polygons can also be used for any "which is the nearest one?" question, such as finding the nearest hospital or school to every address.

READING THE RESULT

- The question: what rainfall does each place get from its nearest gauge?
- The representation: 125 gauges, each given the land closer to it than to any other gauge.
- What it shows: zone size follows gauge density; Keeke's one gauge represents 65.7 km².
- What it cannot establish: rainfall between gauges; the steps at the edges are produced by the method.

IN THE PANE

Module 9 lesson 2 runs Create Thiessen Polygons on the 125 gauges, with the study area set as the processing extent so the zones reach the coast instead of stopping at the outermost gauges, then clips them to the island and colours them by rainfall. You then compare the zones: they are tiny where gauges are dense, and a single zone covers kilometres where gauges are sparse.

Inverse distance weighting

Rainfall changes gradually from place to place rather than in steps. **Inverse distance weighting (IDW)** uses several gauges around each cell, rather than only the nearest, and averages their values with weights that decrease as distance increases. The weight is one over the distance raised to a **power**, p :

$$\text{weight} = 1 / \text{distance}^p$$

With the default power, 2, a gauge twice as far from the cell counts one quarter as much; with power 1 it counts half as much; with power 4, one sixteenth. The power states, as a number, how quickly the relationship in Tobler's law weakens with distance. A higher power gives the nearest gauges more influence, and in the limit IDW produces the Thiessen map again. A lower power gives distant gauges more influence and produces a smoother surface.

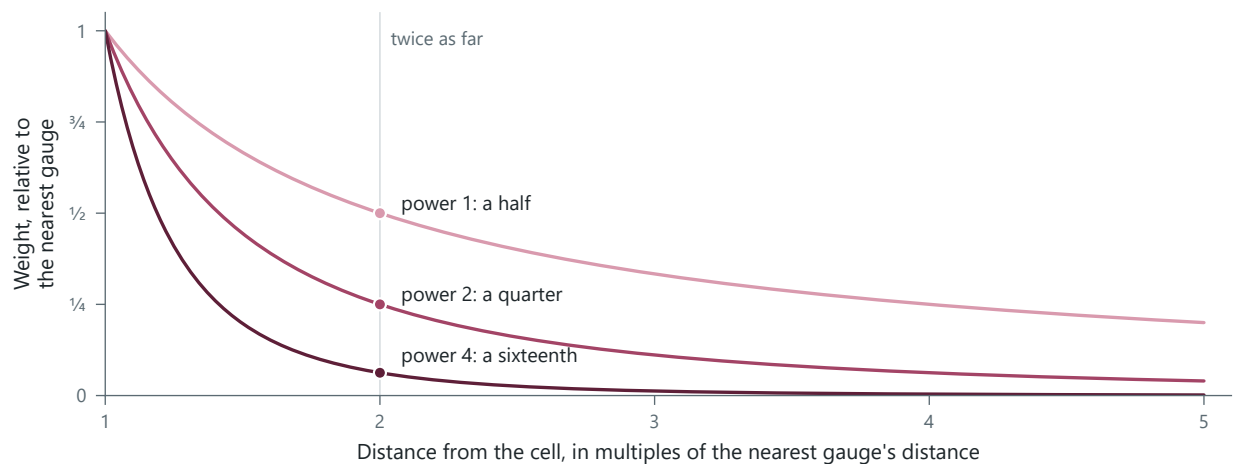


Figure 9.8 How a gauge's weight falls with distance for powers 1, 2 and 4. At twice the distance of the nearest gauge its weight is a half, a quarter or a sixteenth. The curves are inverse powers of distance and are steepest at short range.

A second setting, the **search neighbourhood**, determines which gauges are used at all. A variable neighbourhood uses a fixed number of the nearest gauges, 12 by default, however far away they are; a fixed neighbourhood uses every gauge within a set radius, however many that is.

The map on the opening page is IDW with the default power, 2, and the default 12 nearest gauges, on 100 m cells, limited to the island by a **mask**. Its 502 rows by 751 columns hold values from 309 to 10,199 mm, within the range of the gauges, because a weighted average can never go above its highest input or below its lowest; it reaches both ends here because IDW returns a gauge's own value at the gauge, and a cell sits on each of the two extreme gauges. Raising or lowering the power changes the surface, but only slightly here: its mean moves from 2,051 mm at power 1 to 2,054 mm at power 2 and 2,063 mm at power 3. The power has a larger effect on the peaks. At power 1 the surface does not reach Big Bog's own 10,199 mm; its maximum is 10,144.

An interpolated value can differ from the gauge value nearby. At the town of Kīhei the surface gives 367 mm, while the Kīhei gauge, a short distance away, records 328, so its Thiessen zone also gives 328. IDW returns a gauge's own value only at the gauge. Every other cell holds the estimate at that cell's centre, and at the town's centre the estimate includes the wetter gauges up the slope. Neither number is wrong; the two methods estimate the rainfall between gauges in different ways.

IDW is simple, easy to explain and fast; the whole island is computed in about five seconds. Its weaknesses are a consequence of that simplicity. It is sensitive to outliers: a single unusual gauge appears as a ring of its own colour, as the wettest gauges do on the opening map. It also gives no measure of how uncertain an estimate is.

IN THE PANE

Module 9 lesson 3 runs IDW, a Spatial Analyst tool, on the 125 gauges at power 2 with the 12 nearest neighbours and 100 m cells, masked to Maui. Pro first draws the surface in classes; you view it as a continuous stretch and then return to classes. Pro's History records the run with its settings, so you can correct a wrong setting by reopening the run from History. You then read the value at Kīhei and lay the Rainfall Atlas's own rainfall lines over your surface. The lesson explains the power and the neighbourhood but does not rerun the tool with other values.

Checking a surface

An interpolated surface looks equally reliable everywhere, so it needs to be checked against another source. For Maui there is a good reference, although it has a limitation that is described below. The Rainfall Atlas published its own map of mean annual rainfall as **isohyets**, lines joining places of equal rainfall, the rain version of elevation contours. The lines on the opening page are the Atlas's isohyets at the same values as the colour breaks, so where the two agree, the lines run along the colour edges.

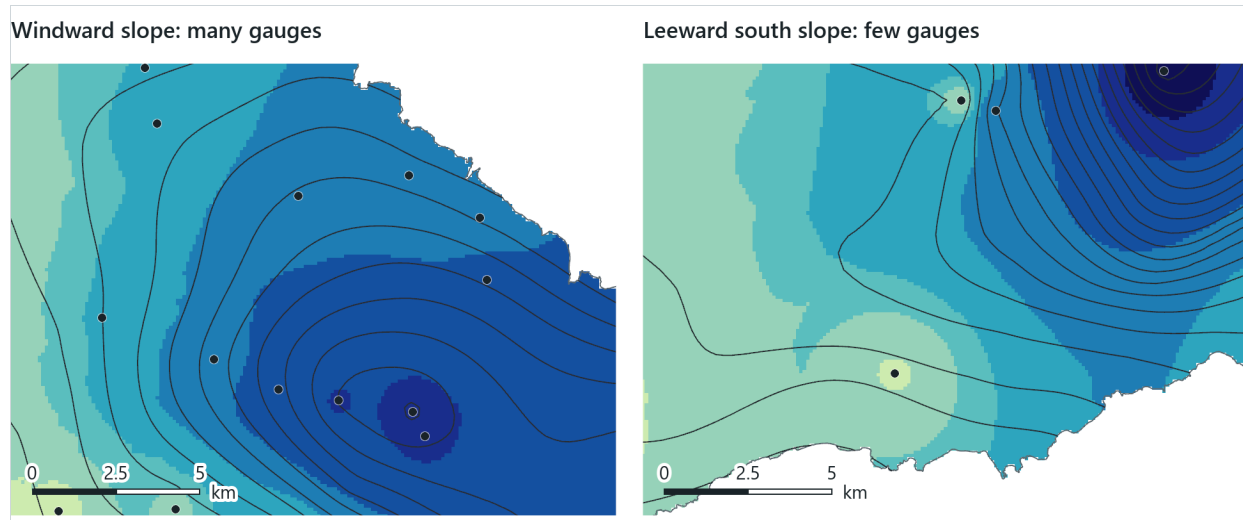


Figure 9.9 Two enlarged areas of the same surface with every Atlas isohyet drawn. On the windward slope (14 gauges in the area) the lines follow the colour bands. On the leeward south slope (4 gauges) they cross the bands, because the Atlas shaped its lines to follow the mountain where there are no gauges.

Coastline: USGS; Hawaii Statewide GIS Program. Rain gauges: Rainfall Atlas of Hawai'i, University of Hawai'i at Mānoa, Department of Geography. Isohyets: Giambelluca et al. (2013), Online Rainfall Atlas of Hawai'i, via the Hawaii Statewide GIS Program (public domain). Colours: ColorBrewer YlGnBu, Cynthia Brewer (Apache 2.0)

Where gauges are dense, the surface computed in five seconds and the University's multi-year atlas agree. Where gauges are sparse, the Atlas used more information (radar, weather models and the shape of the terrain), and the two maps differ. This is a basic limitation of interpolation: a surface can only be as accurate as its samples, and IDW does not show where its estimates are poorly supported. The Atlas is a reference model rather than an independent measurement. It was built from the same gauges and others, with radar, weather models and the terrain, and its isohyets are estimates with their own uncertainty, so agreement between the two maps shows that they are consistent but does not prove that either is correct. The most direct test of an interpolation does not need a second map: you leave one gauge out, predict its value from the others and compare the prediction with the measurement. Repeating this for every gauge in turn is called **cross-validation**. A family of statistical methods called **kriging** addresses IDW's lack of an uncertainty measure. Kriging fits the weights to the way similarity actually decreases with distance in the data and produces an estimate of its own uncertainty beside each estimate (an estimate of uncertainty, not the true error at places nobody measured); the course names kriging but does not run it.

READING THE RESULT

- The question: does the IDW rainfall surface agree with the Atlas's own map?
- The representation: IDW at power 2 from the 12 nearest gauges, with the Atlas isohyets over it.
- What it shows: isohyets follow the colour bands where gauges are dense but not where they are sparse.
- What it cannot establish: true rainfall; the Atlas shares the gauges, so agreement is not proof.

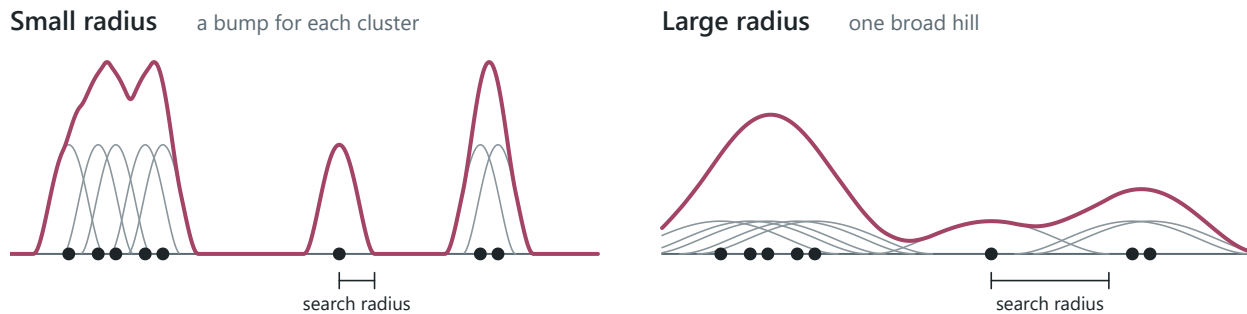
RULE OF THUMB

Report an interpolated map with its method and settings (for IDW, the power and the neighbourhood), show the sample points on it, and do not read detail into the places between sparse samples.

Kernel density

Some surfaces are not made from measurements. A **kernel density** surface starts from events or objects, each recorded as one point, and is used to answer a different question: "how many points are near here?" rather than "what is the value between the points?"

The method replaces each point with a small hill-shaped surface, the **kernel**, which is highest over the point and falls smoothly to zero at a chosen distance, the **search radius**. It then adds the kernels together. A cell near many points lies under many overlapping kernels and receives a high value; a cell with no point within the search radius receives zero. The result is measured in points per unit of area. A **population field** can make one point count as several, for example a building with four homes. If the field is left empty, every point counts once.



Grey: each point's own hill, zero at the search radius. Rose: the hills added, the density.

Figure 9.10 Kernel density in one dimension. Each point is replaced by a hill-shaped kernel that reaches zero at the search radius, and the kernels add up to the density. A small radius gives a separate peak for each cluster; a large radius merges them into one broad hill with a lower peak.

Module 9 applies kernel density to cesspools. Hawai‘i has more cesspools than any other state, and a state law requires every one to be converted by 2050. The Hawaii Cesspool Prioritization Tool maps 82,141 of them, 11,105 on Maui.

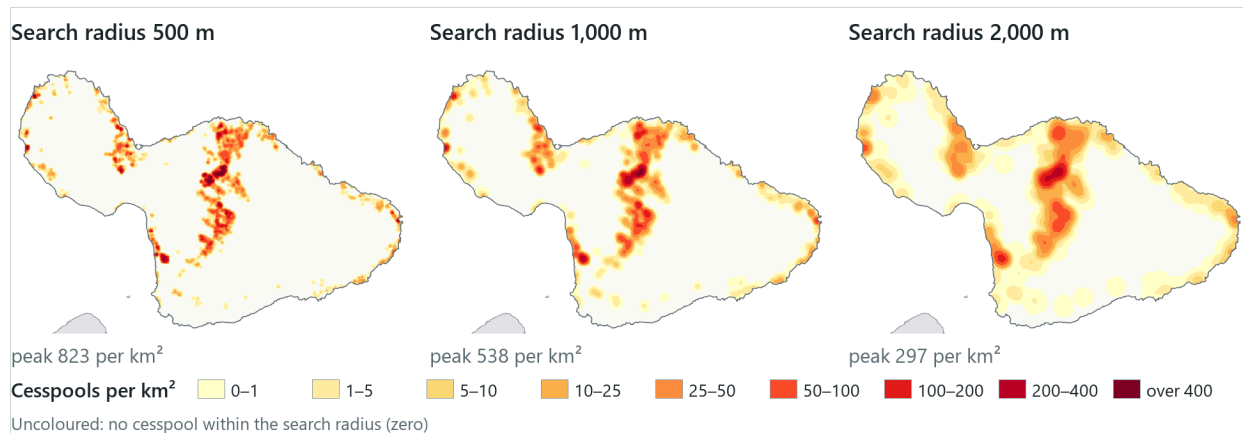


Figure 9.11 The same 11,105 Maui cesspools at three search radii, on one colour scale. At 500 m the neighbourhoods separate and the peak reaches 823 cesspools per km²; at 1,000 m it is 538; at 2,000 m the clusters merge into broad areas and the peak falls to 297. Uncoloured ground has no cesspool within the radius.

Coastline: USGS; Hawaii Statewide GIS Program. Cesspools: Hawaii Cesspool Prioritization Tool (2022), UH Water Research Center, via the Hawaii Statewide GIS Program (public domain). Colours: ColorBrewer YlOrRd, Cynthia Brewer (Apache 2.0)

The search radius has more effect on the map than any other setting. When the same points are spread over wider kernels, the kernels are lower, so doubling the radius lowers the peak and halving it raises the peak. None of the three maps is the correct one; the radius is chosen to match the scale of the question. A 500 m radius suits the question "which street?", 1,000 m "which neighbourhood?" and 2,000 m "which part of the island?". The number of zero cells also changes: at 1,000 m about 70 per cent of Maui's cells are zero (computed), and with a wider radius less of the island is empty. At every radius the main pattern is the same. The cesspools are concentrated Upcountry, on Haleakalā's lower northern slope around Makawao, Pukalani and Ha'ikū, where 5,253 of the 11,105 are found; the towns with sewers have very low values.

Density is not a rate

Two cautions apply to every density map. First, on a stretched colour ramp any colour means only "more than zero": a coastal town can look tinted and still hold a fraction of a cesspool per square kilometre, so read the legend or the cell's value before interpreting the colour. Second, a density divides a count by **ground area**, not by people or homes. A neighbourhood of large lots can look pale on this map and still be entirely on cesspools. "What share of homes are on cesspools?" is a question about a rate, which requires the number of homes and a division rather than a kernel. Kernel density is well suited to the question "how many, near here?" and should not be used to answer other questions.

READING THE RESULT

- The question: where on Maui are cesspools concentrated?
- The representation: 11,105 points as kernels with search radii of 500, 1,000 and 2,000 m.
- What it shows: at every radius the peak is Upcountry (823, 538 and 297 per km²).
- What it cannot establish: what share of homes use cesspools; density is per km², not per home.

RULE OF THUMB

State the search radius on every density map, and the power and neighbourhood on every IDW map. These settings have the largest effect on the map, and a reader cannot see them on the map itself.

IN THE PANE

Module 9 lesson 4 adds the state's cesspool layer, filters it to Maui and runs Kernel Density with no population field, 100 m cells and a 1,000 m search radius, masked to the island. You redraw it as a stretch with the zeros transparent, run it once more at 2,000 m from History and compare the two, then read a coastal patch and the Upcountry peak side by side to see what the colours indicate and what they do not.

TRY IT

With the Module 9 project open, ask Waldo in Ask mode: how much does my IDW surface differ from the Kahului Airport gauge, and why? Then ask which search radius your density surface used, because your map should state that setting.

Key terms

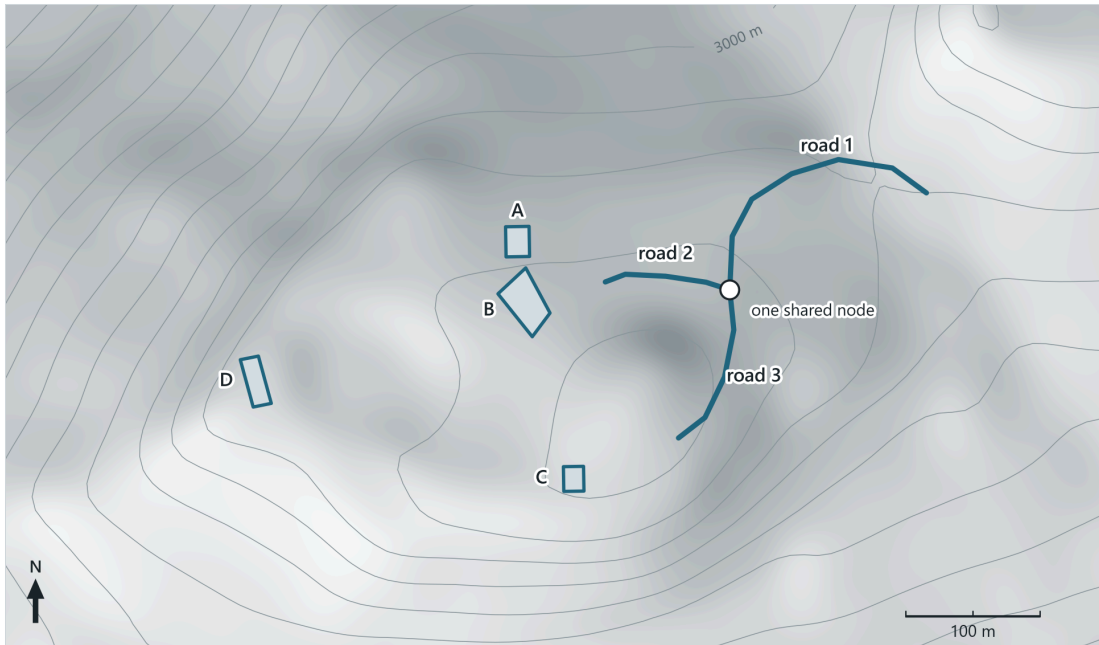
Tobler's first law of geography	"everything is related to everything else, but near things are more related than distant things" (Waldo Tobler, 1970).
distance decay	the weakening of a relationship or an influence as distance grows.
spatial autocorrelation	the tendency of values at nearby places to be related: alike (positive) or unlike (negative), rather than random.
interpolation	estimating values at unsampled places from nearby sample points, usually into a raster.
sentinel value	a code such as -9999 stored in place of a missing measurement; it must be removed before any calculation.
Thiessen (Voronoi) polygons	polygons that assign every location to its nearest sample point; also called proximal zones.
inverse distance weighting (IDW)	interpolation by a weighted average of nearby samples, each weight one over the distance to a power.
power (IDW)	the exponent that sets how fast a sample's weight falls with distance; 2 by default.
search neighbourhood	the samples an estimate may use: a set number of the nearest, or all within a set radius.
mask	a setting that limits a raster tool's output to the cells inside a study-area polygon.
isohyet	a line joining places of equal rainfall; the rain version of an elevation contour.
cross-validation	testing an interpolation by leaving out each sample in turn, predicting its value from the others and comparing the prediction with the measurement.
kriging	a statistical interpolation method that fits its weights to how similarity falls off with distance and also estimates its own uncertainty (named only).
kernel density	a surface made by spreading each point into a hill of a chosen radius and adding the hills, giving points per unit of area.
search radius	in kernel density, the distance at which each point's hill falls to zero; the setting that most changes the map.

Further reading

- Tobler, W. R. (1970). "A computer movie simulating urban growth in the Detroit region." *Economic Geography* 46, 234–240. <https://doi.org/10.2307/143141>
- Giambelluca, T. W., et al. (2013). "Online Rainfall Atlas of Hawai'i." *Bulletin of the American Meteorological Society* 94, 313–316; the atlas itself at <https://rainfall.geography.hawaii.edu/>
- Olaya, V. *Introduction to GIS*, the chapters on creating raster layers and interpolation. <https://volaya.github.io/gis-book/en/>
- Esri. "How IDW works" and "How Kernel Density works." ArcGIS Pro documentation. <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/how-idw-works.htm>
- The course's data sources: the Hawaii Statewide GIS Program layers this chapter's maps use, with their publishers and terms.

Creating and editing data

Accompanies Module 10 · Data Editing and Creation



Relief: USGS 10 m DEM, served by the Hawaii Statewide GIS Program (public domain); hillshade and 10 m contours computed for this book

WHY THIS MATTERS

Every layer used in this course so far was made by someone else. At some point the data you need will not exist, and you will have to make it: you decide what is there, click its shape and type its attributes. The map on the opening page shows this work at the Haleakalā summit on Maui: three road lines that meet at one node and four roofs, traced by hand. This chapter covers what a digitized feature is made of, how to design the feature class before the first click, and the habits that keep hand-made data reliable.

YOU SHOULD BE ABLE TO EXPLAIN

- why lines that should meet must share one node, and how snapping prevents undershoots and overshoots
- why the geometry type, fields and coordinate system are decided before the first vertex
- why saving the project does not save your edits, and what Undo and Discard can still take back

IN THIS CHAPTER

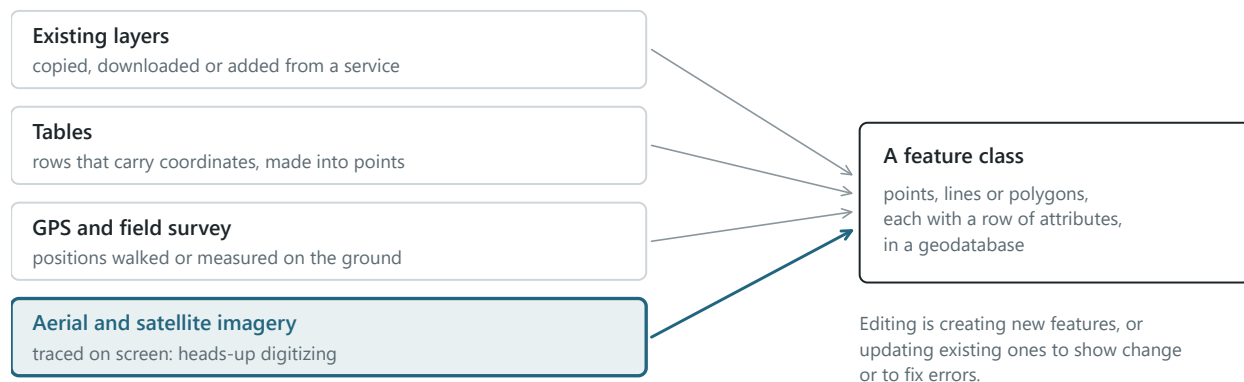
Where new data comes from
Vertices, segments and nodes
Design before drawing
Templates, tools and attributes

Snapping
Changing what you drew
Saving edits
Editing hygiene

Where new data comes from

Most GIS work starts from layers that already exist: a service added from ArcGIS Online, a feature class copied from a colleague, a table whose rows carry coordinates and can be turned into points. When nothing suitable exists, the data has to be created, by walking the ground with a GPS receiver, by survey, or by tracing a picture. Editing includes both creating new features and updating existing ones to show what has changed since they were made or to fix their errors.

Where new vector data comes from



Module 10 uses the last route: two empty feature classes, filled by tracing an aerial picture of the Haleakalā summit.

Figure 10.1 Four ways of getting data into a feature class. Module 10 uses the last one: heads-up digitizing, tracing an aerial picture on screen.

Tracing is one of the oldest ways of turning maps and pictures into GIS data, and manual tracing is still the most common way vector layers are made from them. On a digitizing tablet the operator traces the paper map itself; in **heads-up digitizing** the picture is on screen, the operator's eyes stay up on it, and each mouse click adds a vertex. The geometry and the attributes are separate tasks: the operator first draws the shape and then records what it represents.

The quality of the result is limited by the picture. On an aerial mosaic made to look straight down, each roof appears close to its footprint, although only a true orthophoto removes the lean of tall buildings entirely, and a roof outline differs from the ground footprint where the roof overhangs the walls. On an oblique image, tall buildings lean, their walls are visible, and the roof appears a few metres away from the ground it stands on. White roofs on pale concrete can also be hard to distinguish, so the display's brightness and contrast are adjusted until the edges are visible. These adjustments change how the image is drawn, not the data. The features traced are roofs, and the layer's description should state this.

IN THE PANE

Module 10 lesson 1 starts a new project (or a new map in the project already open), adds the State of Hawaii's 2022 aerial mosaic by URL, goes to the summit observatories with Go To XY and darkens the picture with the layer's Enhancement settings until the roof edges are visible. The lesson ends by showing the features to trace: three road lines and four roofs, A to D. Heads-up digitizing is the only task in the course that Waldo cannot do for you; instead, Waldo reads what you have drawn.

Vertices, segments and nodes

A vector feature is a list of coordinates, its **vertices**. A **point** is one coordinate pair. A **line** is two or more vertices joined by straight **segments**; it has a direction, running from a **start node** to an **end node**, with any number of vertices between. A **polygon** is a line that closes on itself: its last vertex is its first, so it encloses an area and needs at least three distinct vertices.

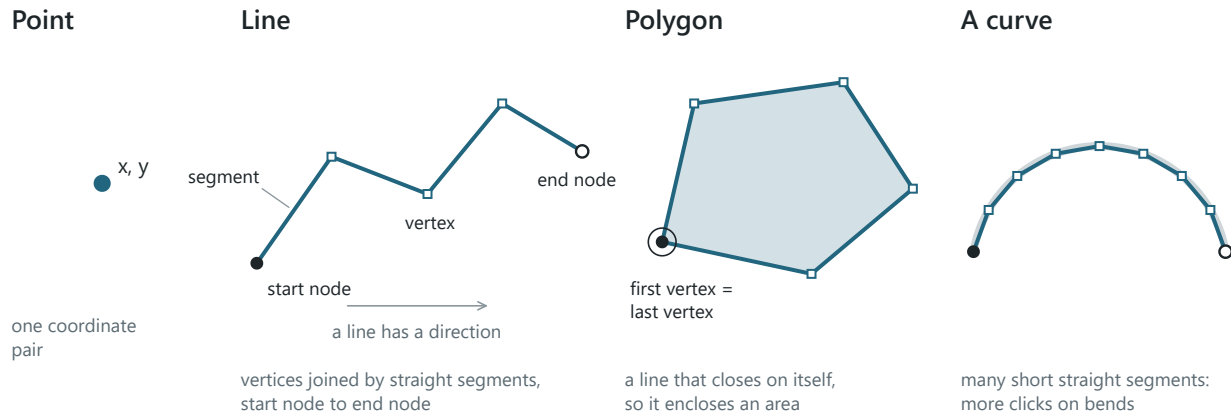


Figure 10.2 Point, line and polygon as vertices. A line runs from a start node to an end node; a polygon's first vertex is also its last; a curve approximated by short straight segments.

Digitizing consists of choosing those vertices. When you trace a road, each click adds one vertex. A bend needs many short segments and a straight section needs few, because this feature class stores no true curves, only more vertices (a geodatabase can store arcs, as the Arc construction tool below shows, but a traced road does not need them). When you trace a roof, the corners you click form the polygon. The module's four roofs are stored as five points each: four corners and a closing vertex that repeats the first. Its longest road line, 215 m of the access road, is seven points: a start node, five vertices on the bends, and an end node.

Where lines meet, they should share one node. On the summit, the access road is drawn as three lines: road 1 comes in from the north-east and ends at a junction, and roads 2 and 3 both start at that same point. Because the three roads share one node, the data itself records that they connect.

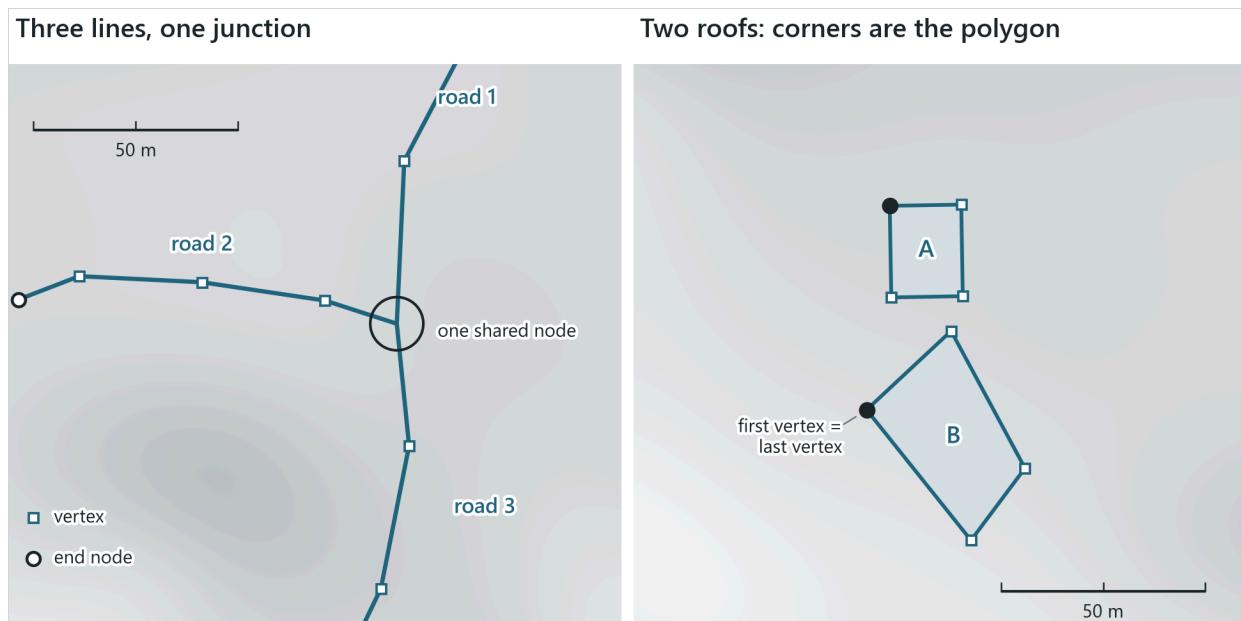


Figure 10.3 The summit's features close up, with every stored point shown. Left: road 1 ends where roads 2 and 3 begin, on one shared node. Right: roofs A and B, each with four corners and a first vertex that is also the last. Relief: USGS 10 m DEM, served by the Hawaii Statewide GIS Program (public domain); hillshade computed for this book

Design before drawing

New features have to be drawn into a **feature class**, and a feature class is a set of decisions made before the first vertex: what kind of shapes it holds, what it records about each one, and where on Earth its coordinates are measured. In Pro those decisions are the pages of the Create Feature Class wizard, and the empty class is usually stored in a new file geodatabase set as the project's default, so that new data created in the project is saved there.

Designing a feature class before the first vertex

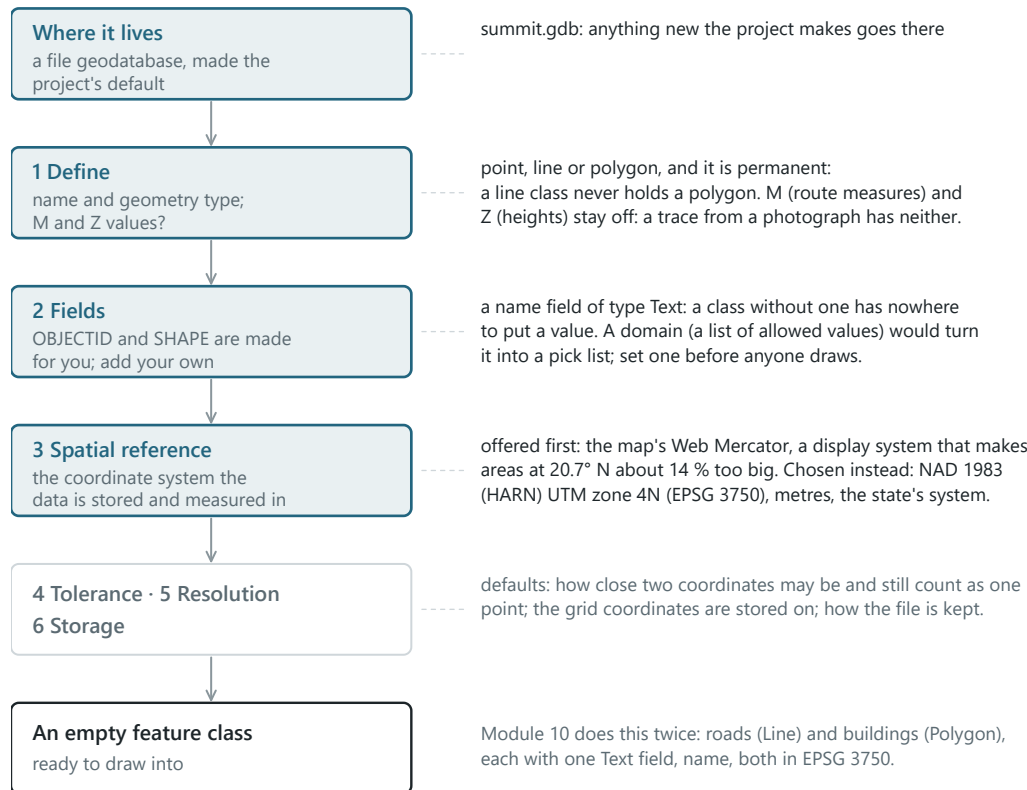


Figure 10.4 The Create Feature Class wizard as a decision flow. The first three pages hold the main decisions; the last three are left at their defaults.

The **geometry type** (point, line or polygon) is permanent: a line class never holds a polygon, so the road and the roofs need a class each. Two other options are set with it. **M values** store a measure with each vertex, used for routes; **Z values** store a height. A trace from a photograph has neither, so both are left off.

The fields also have to be defined before any features are drawn. Pro creates OBJECTID and the shape field itself; any other field, such as a Text field for a name, must exist before you can type a value into it. A **domain**, a list of allowed values, would turn that field into a pick list and prevent typing errors such as "Obseatory"; in real editing projects, domains are set up before anyone draws.

The coordinate system should be chosen deliberately rather than accepted as offered. The wizard offers the map's own system first, and a map built on web imagery is in Web Mercator, a display system that at the summit's latitude, 20.7° N, makes every area about 14 % too big (the stretch is $1/\cos^2$ of the latitude, 1.143 there). The module stores both classes in NAD 1983 (HARN) UTM zone 4N (EPSG 3750), which is Maui's UTM zone, measured in metres, and the system in which the State of Hawaii publishes its data. The map itself can stay in Web Mercator to match the imagery; Pro projects the new layers on the fly for drawing, as Chapter 3 describes, while storing and measuring them in UTM.

The last three pages, **tolerance**, **resolution** and storage, are kept at their defaults. The tolerance is how close two coordinates may be before they count as one point, the resolution is the fine grid on which coordinates are stored, and the storage settings control how the file is kept.

RULE OF THUMB

Design before you draw. Decide the geometry type, the fields and the coordinate system the data will be measured in before the first vertex; the map's display system is rarely the right one to store in.

IN THE PANE

Module 10 lesson 2 makes a file geodatabase, `summit.gdb`, sets it as the project default and runs the wizard twice: roads (lines) and buildings (polygons), each with one Text field, name, and both in EPSG 3750. The lesson mentions domains but does not create one. Waldo checks the geometry type, the field and the coordinate system of each class.

Templates, tools and attributes

Pro gathers everything for editing on the Edit tab: saving and discarding edits, snapping, creating features, selecting them, and a gallery of tools for changing features that already exist. New features start from a **feature template**, one for each layer you can draw into. A template stores the settings for a new feature: the layer the feature goes to, the symbol it is drawn with, the construction tools available and any attribute values to fill in automatically. Templates are made from a layer's symbology or set up by hand, and they differ for points, lines and polygons.

A feature template, and the tools it offers

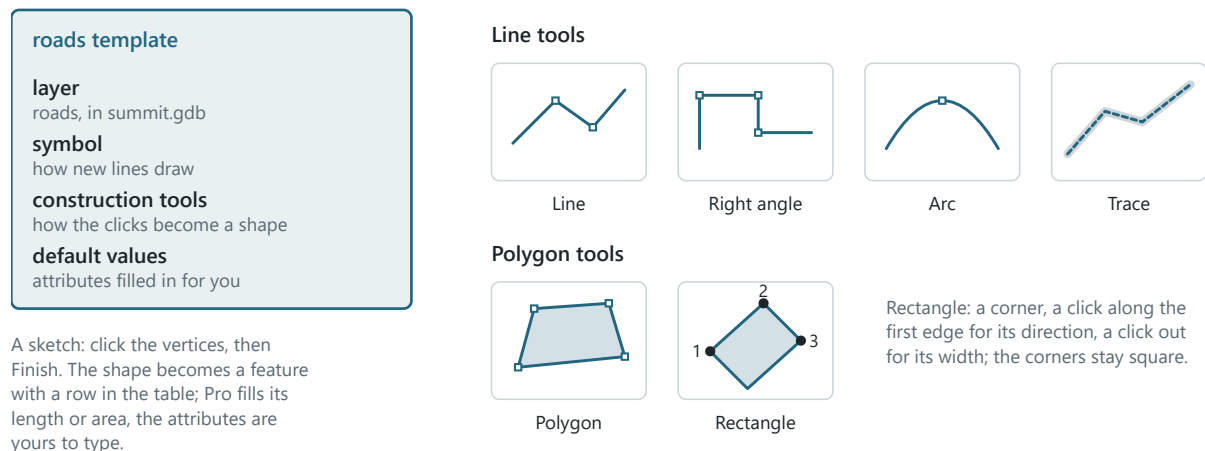


Figure 10.5 A feature template and the tools it offers. The construction tool determines how clicks are turned into a shape: straight segments, right angles, arcs, traced edges, free polygons or square-cornered rectangles.

The shape being drawn is a **sketch**: click the vertices, then finish it, and the sketch becomes a feature. A **construction tool** determines how the clicks are turned into geometry. For lines there are straight segments, right angles, arcs and tracing along an existing edge; for polygons a free polygon, where each click is a corner, and a rectangle, where three clicks (a corner, a direction along the first edge, a width) give four square corners in any orientation. The Rectangle tool is suited to the summit's rectangular roofs, and the free polygon to other shapes.

A feature consists of its shape and its attributes. Every new feature gets a row in the table, and Pro fills in the values it can compute, such as the length of a line or the area of a polygon. The editor types the other values, such as the name, usually in the Attributes pane with the new features

selected. Immediately after digitizing, a table of three new roads has three lengths and three empty names.

IN THE PANE

Module 10 lesson 3 tours the Edit tab, picks the roads template and draws the three road lines; its last step opens the table, where the lengths are filled and the names are not. Lesson 4 uses the buildings template, with the Polygon and Rectangle tools, for the four roofs, and names them A to D in the Attributes pane. Waldo reads the counts, lengths, areas and blank names with its `check_edits` tool.

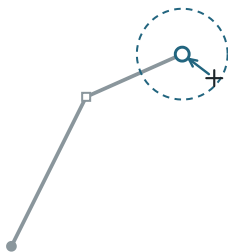
Snapping

Features that are meant to touch must meet exactly. A sewer connection has to fall on the house's edge, neighbouring parcels have to share their corners, a stream gauge has to lie on its stream, and two stream segments have to connect. Clicking a mouse by hand is not that precise, and near-misses are among the most common errors in digitized data.

Snapping places vertices precisely. Within a small distance of a target, the **snap tolerance** (a few pixels on screen, or a distance in map units), the cursor jumps exactly onto it. What counts as a target is set by snap type: **end** snapping onto the end of a line, **vertex** snapping onto any vertex, **edge** snapping onto any point along a segment. Pro starts with point, end, vertex and edge snapping on; the lesson turns point and edge off so that a new road can snap only to another road's end or to a vertex.

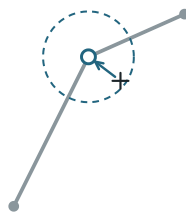
Three snap types, inside a snap tolerance

Inside the dashed circle (the tolerance: a few pixels), the cursor jumps onto the target.



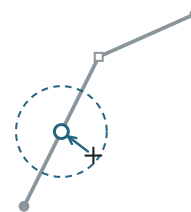
End snapping

onto the end of a line



Vertex snapping

onto a vertex



Edge snapping

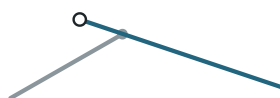
onto any point of a segment

Where two roads meet



Undershoot

the new road stops short:
a gap, two nodes



Overshoot

the new road crosses past:
a dangling stub



Snapped

the new road starts on the
old one's end: one node

Figure 10.6 Snap types, each with its tolerance, and what they are for. Without snapping, a road drawn to a junction stops short (an undershoot) or crosses past it (an overshoot); snapped, it starts on the other road's end, and the two share one node.

Without snapping, the second road at the summit junction would miss the first by a very small distance. The result would be an **undershoot**, which stops short and leaves a gap too small to see, or an **overshoot**, which crosses past and leaves a stub. Both errors look correct on screen but are wrong in the data, because the lines do not share a node. With snapping on, roads 2 and 3 each start exactly on road 1's end.

RULE OF THUMB

Turn snapping on before drawing features that must meet, and limit it to the types the job needs (ends and vertices for a road network), so the cursor never jumps somewhere unintended.

Changing what you drew

Digitizing also includes correcting features. The tools gallery on the Edit tab changes features that already exist, and each change is an edit like any other.

Changing features that already exist

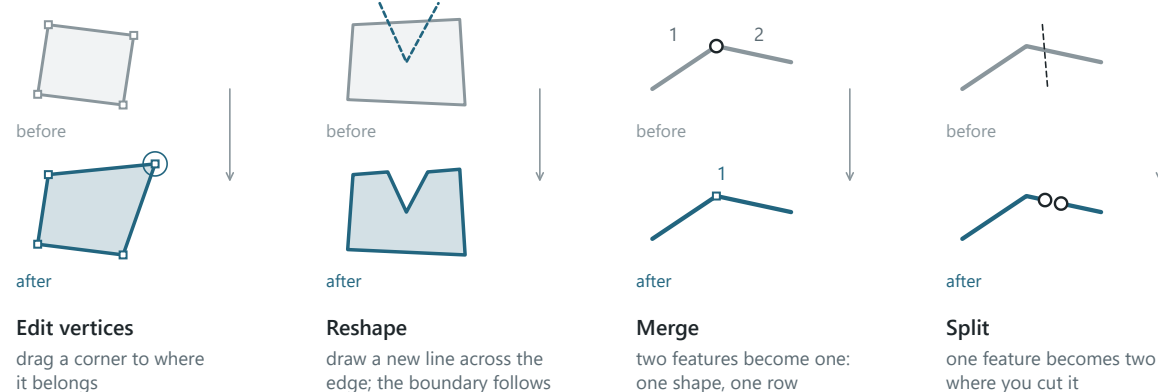


Figure 10.7 Four tools for features that already exist, before and after: move a vertex, reshape an edge, merge two features into one, split one into two.

Edit Vertices shows a feature's vertices as small squares; dragging one moves that corner to where it belongs. **Reshape** redraws part of an edge: draw a new line across the feature and its boundary follows it. **Merge** makes one feature of several: two road lines selected together become one line with one row in the table. **Split** does the opposite and cuts one feature into two. Moving a vertex changes the geometry's measurements, so a roof whose corner has been corrected has a new area. A check by numbers can therefore detect that something changed without looking at the map. It cannot confirm that the right corner moved to the right place; only a comparison with the picture can do that.

IN THE PANE

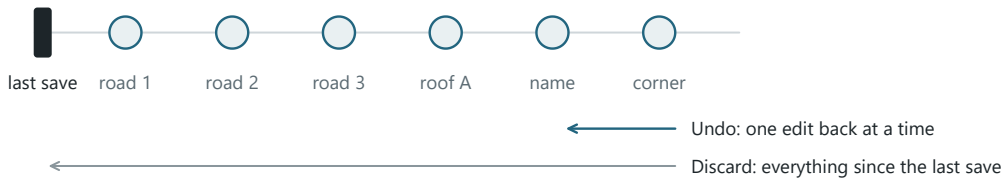
Module 10 lesson 4 moves one roof corner with Edit Vertices and names Reshape, Merge and Split without using them. Waldo compares the roofs' total area before and after to check whether the corner moved.

Saving edits

Edits are saved separately from the project. Features are stored in the geodatabase; the map, its layers, symbology and layouts are stored in the project file (the .aprx). The **Save** in the Edit tab's Manage Edits group writes the edits into the geodatabase. Saving the project writes the project file and, unless Pro has been set to save edits along with the project, leaves the edits unsaved. Pro warns about unsaved edits when a project closes, but the safe habit is to save edits as you go.

Two saves, two different files

Edits since the last save



Save, on the Edit tab

writes the features into the geodatabase:
summit.gdb > roads, buildings
the save that keeps what you drew

Save, on the Project tab

writes the project file (.aprx): maps,
layers, symbology, layouts; not the edits,
unless Pro is set to save them with it

Once the edits are saved they are permanent: Undo and Discard stop at the save.

Figure 10.8 The project and the edits are saved separately, to two different files. Until the edits are saved, Undo takes them back one at a time and Discard removes them all; the Edit tab's Save writes them into the geodatabase.

Until you save, every edit can be reversed. **Undo** steps back one edit at a time, as far as the last save. **Discard** removes every edit since the last save in one step. After the save the edits are permanent, which is the purpose of saving and the reason a backup is important.

RULE OF THUMB

Saving the project does not save your edits. After drawing or changing features, save the edits from the Edit tab, and confirm that nothing is left unsaved before closing.

IN THE PANE

Module 10 lesson 3 stops after the three roads are drawn and has Waldo count how many of them are on the map but not yet in the geodatabase. The save step is then always taught, even when a read suggests it is already done, and Waldo reads again to confirm that no lines are left unsaved and that Pro's unsaved-edits flag is off. Lesson 4 saves once the roofs are drawn and named, and again after the moved corner. Because a read showing no unsaved edits does not prove that you saved, Waldo asks whether you did.

Editing hygiene

A saved edit changes the data permanently, so editing requires careful habits. The list below gathers common editing tips and the practice of Module 10.

Before, during and after an editing session

- | | |
|--|--|
| <p>☒ Back up the data first
a copy of the geodatabase you are about to change</p> <p>☒ Edit in a map of its own
only the layers the job needs</p> <p>☒ Lock what you are not editing
make other layers not editable (List By Editing)</p> <p>☒ Design before you draw
geometry type, fields, coordinate system: set up front</p> | <p>☒ Know your coordinate systems
store edited layers alike; know when the map projects them</p> <p>☒ Snap only to what you need
e.g. End and Vertex; fewer targets, fewer surprises</p> <p>☒ Save edits often
nothing saves them for you by default</p> <p>☒ Check what you made
counts, lengths, areas, blank attributes</p> |
|--|--|

Figure 10.9 An editing checklist: protect the data, set up to be precise, save deliberately, and check the result by numbers.

Protect the data first: keep a backup copy of the geodatabase you are about to change, edit in a map of its own with only the layers the job needs, and make every other layer not editable (the List By Editing view of the Contents pane shows which are). Set up to be precise: design the feature class before drawing, and keep the layers being edited in one stored coordinate system. The module draws UTM classes into a Web Mercator map to match the imagery, so in practice the rule is to know when the map is projecting a layer on the fly. Turn snapping on, limited to what the job needs.

Save edits deliberately, because Pro does not save them for you by default. Then check the result the way Waldo does in this module, by numbers as well as by eye: how many features, how long, how large, which attributes are still blank, and then, against the picture, whether the shapes sit where they should and the lines that should meet do meet. On the summit, the three traced roads come to about 435 m and the four roofs to about 2,120 m², the range a careful trace of these features gives. A total within this range is a rough check and does not prove that the trace is good.

TRY IT

With the Module 10 project or map open, ask Waldo in Ask mode: how many roads and buildings have I drawn, how long and how big are they, and are my edits saved? Waldo runs its edit check and answers with counts, geodesic lengths and areas, blank names, and what is still unsaved (Pro's unsaved-edits flag, and how many features are on the map but not yet on disk), and never claims to have looked at the map.

Key terms

heads-up digitizing	tracing features on screen by clicking their vertices over an image.
vertex	one x, y coordinate of a line or polygon; the shape is the list of its vertices.
segment	the straight piece of a line or polygon boundary between two consecutive vertices.
node	the first (start) or last (end) vertex of a line; lines that meet should share one.
feature class	a set of features of one geometry type, with the same fields and coordinate system, stored in a geodatabase.
geometry type	point, line or polygon; fixed when the feature class is created.
M value / Z value	an optional measure (for routes) or height stored with each vertex.

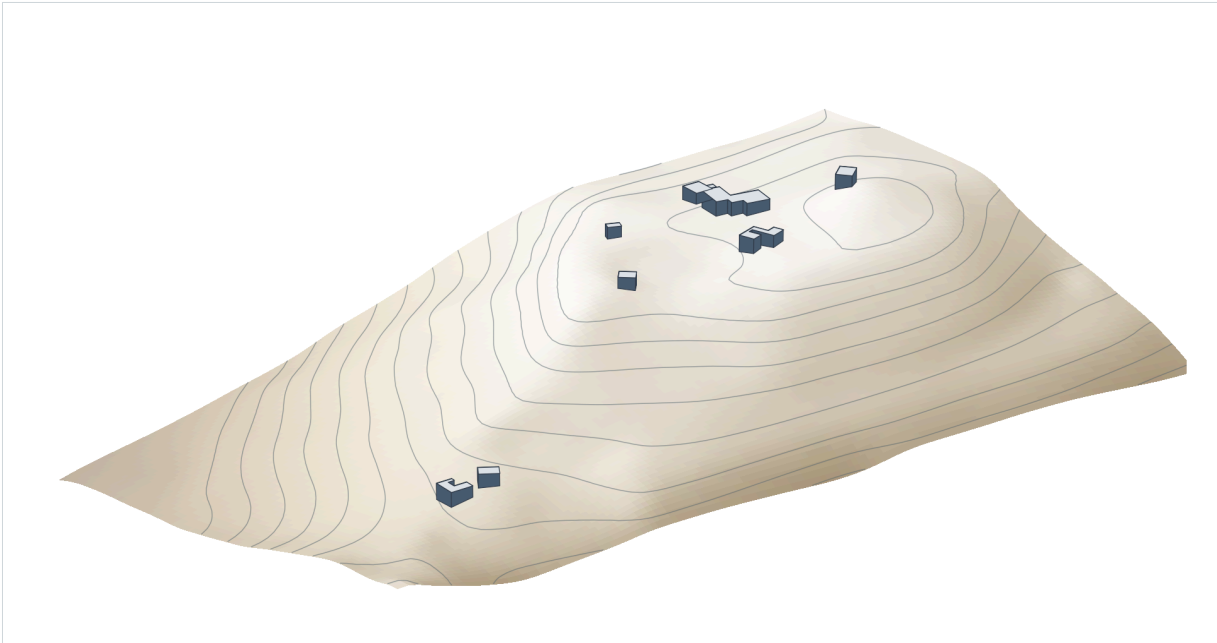
domain	a list or range of allowed values for a field, offered as a pick list when editing.
XY tolerance	how close two coordinates may be before they are treated as the same point.
XY resolution	the fine grid on which a feature class stores its coordinates.
feature template	a recipe for new features: the target layer, its symbol, its construction tools and default attribute values.
construction tool	the tool that turns clicks into a shape, such as Line, Arc, Polygon or Rectangle.
sketch	the shape being drawn, before finishing turns it into a feature.
snapping	moving a vertex exactly onto an end, vertex or edge of another feature when the pointer comes within a tolerance.
snap tolerance	the distance, in pixels or map units, within which snapping pulls the pointer onto a target.
undershoot / overshoot	a line that stops short of the line it should meet, or crosses past it.
save edits	writing the edits made since the last save into the geodatabase; separate from saving the project.
undo / discard	taking back the last edit, or every edit since the last save.
reshape / merge / split	redrawing part of a feature's edge; making one feature of several; cutting one feature into two.

Further reading

- Olaya, V. *Introduction to GIS*, the chapter on data sources (digitization and its errors).
https://volaya.github.io/gis-book/en/Data_sources.html
- Esri. ArcGIS Pro documentation: editing, feature templates, snapping and editing options.
<https://pro.arcgis.com/en/pro-app/latest/help/editing/>
- U.S. Geological Survey. The 10 m digital elevation models behind this chapter's relief (public domain), served for Hawaii by the Hawaii Statewide GIS Program.
- The course's data sources: every layer the lessons use, with its publisher and licence, including the Module 10 imagery and its terms.

3D scenes and your final map

Accompanies Module 11 · 3D Scenes and a Final Map



Elevation: USGS 10 m DEM, State of Hawaii copy (Hawaii Statewide GIS Program; public domain). Building footprints: USA Structures, Oak Ridge National Laboratory (ORNL) and FEMA Geospatial Response Office (CC BY 4.0), extruded 10 m for this figure

WHY THIS MATTERS

Every map in this book so far has been flat. A scene draws the same layers in three dimensions: a ground surface built from an elevation model, with flat layers laid over it and features raised into blocks. The picture above shows the Haleakalā summit on Maui, drawn from the state's 10 m elevation model, with its eight federal building footprints raised 10 m each. The chapter ends, as the last lesson of the course does, with a checklist for a finished map of your own.

YOU SHOULD BE ABLE TO EXPLAIN

- why a scene's ground is a DEM, and why a 10 m cell can give a sharp summit a lower height than it has
- why Max Height suits a building on a slope, and why one height typed for every building is a guess
- why the map is finished before the layout, and why a layout is judged by its exported image

IN THIS CHAPTER

Two kinds of scene

The ground is a DEM

Moving in a scene

Draping and extruding

The summit's buildings

One number or a field

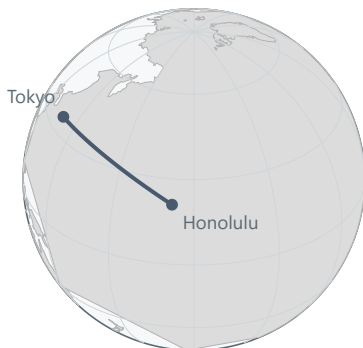
Your final map

Where to make it

Two kinds of scene

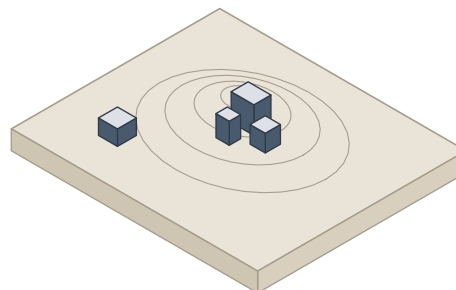
A **scene** is a map drawn in three dimensions. ArcGIS Pro has two kinds. A **global scene** shows the whole round Earth. It suits large areas where the curvature of the planet matters, such as a flight route across an ocean, the track of a hurricane or a continent seen from space. A **local scene** shows a flat piece of the world in a projected coordinate system. It suits a site, a town or a mountain, where the curvature does not matter.

Global scene



The whole round Earth. For large areas, where the curve matters: a flight route, a hurricane's track.

Local scene



A flat piece of the world in a projected coordinate system. For a site, a town, a mountain.

Figure 11.1 A global scene draws the round Earth, so a route between Honolulu and Tokyo follows the curve of the planet; a local scene is a flat, projected piece of ground with the features standing on it.

Land: Natural Earth (public domain)

The difference is the same as the one Chapter 3 described between a globe and a projected map. A global scene keeps latitude and longitude on a sphere and is never flattened. A local scene is a projected map with heights added, so everything Chapter 3 says about choosing a projected coordinate system for a place applies to it.

A scene is usually made from an existing map. Converting a map to a local scene makes a **new item** beside it (named after the map with "_3D" on the end); the flat map is not changed. The scene's Contents then sorts its layers under three headings: **3D Layers** for things that stand up off the ground, **2D Layers** for flat things drawn on the ground, and **Elevation Surfaces**, where a surface called **Ground** contains the elevation of the ground. A newly converted scene has everything under 2D Layers and one source under Ground.

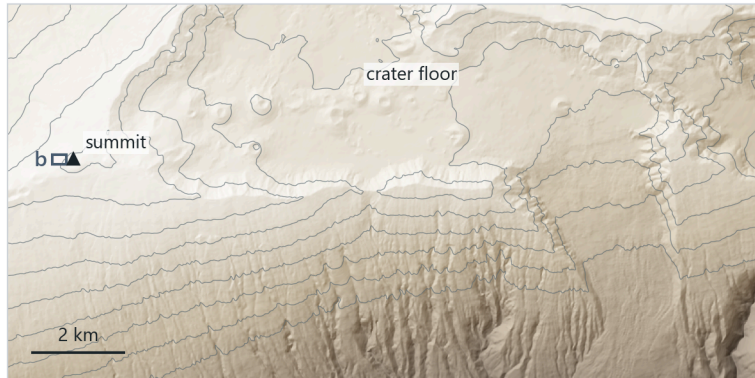
IN THE PANE

Module 11 lesson 1 builds a fresh map of the Haleakalā summit with two layers added by URL, the state's imagery and the federal building footprints, then converts it to a local scene. The lesson explains global scenes but does not use one. You look at which heading each layer is listed under in Contents, and Waldo reads the scene to confirm that its ground is Esri's world DEM.

The ground is a DEM

Every scene is drawn on an elevation surface, the **ground**, and the ground is a DEM: the grid of elevations from Chapter 8, one height for every cell. A new local scene gets Esri's elevation model of the whole world as its ground unless another is chosen. The flat layers are then laid over that surface, so imagery that was flat on the map has slopes in the scene.

a The DEM: one height for every 10 m cell (contours every 200 m)



b Twenty cells on the slope (m)

2,954	2,957	2,960	2,963	2,967
2,956	2,959	2,962	2,965	2,968
2,959	2,962	2,964	2,967	2,970
2,961	2,964	2,967	2,969	2,972

Each square is one 10 × 10 m cell; the ground rises towards the summit.

c The same grid as a surface, seen from the south-west, heights true to scale: the ground

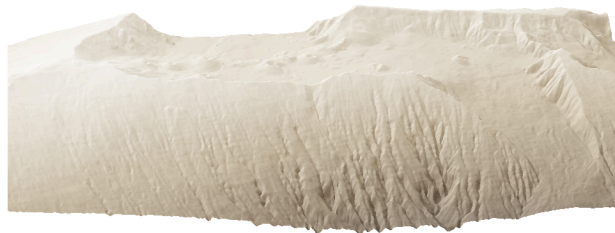


Figure 11.2 The ground of a scene is a DEM. (a) The State of Hawaii's copy of the USGS 10 m DEM for the Haleakalā summit and crater, shaded, with contours every 200 m. (b) Twenty of its cells, 10 m on a side, beside the summit's buildings. (c) The same grid drawn as a surface in perspective, heights true to scale: this is the surface a scene is drawn on.

Elevation: USGS 10 m DEM, State of Hawaii copy (Hawaii Statewide GIS Program; public domain)

A scene shows the height of the ground under the pointer, so you can read the elevation wherever you point. At the observatory buildings the state's DEM gives about 3,033 m. The highest cell of that grid is 3,051 m, a little below the 3,055 m of Pu'u 'Ula'ula, the summit itself. A 10 m cell holds one height for a hundred square metres, so a sharp peak can appear lower than it is, and a height also depends on the vertical datum it is measured from (the surface that the height is measured above). Agreement between the scene's own ground and the state's DEM is a consistency check rather than independent proof, because Esri's world elevation model uses USGS data for Hawaii, as the state's layer does, so the two are related.

East of the summit the ground falls away into Haleakalā crater, a valley about 11 km long. Along the row of cells through the buildings its floor lies at about 2,200 m, some 800 m below the rim, dotted with cinder cones.

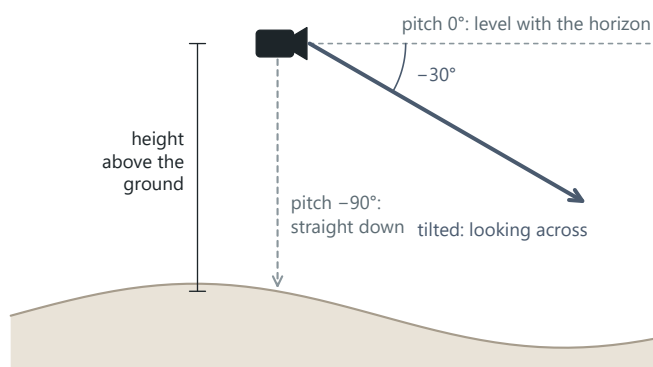
A ground surface can be stretched upwards with **vertical exaggeration**: a factor of 2 doubles every height. In gently rolling terrain, a small exaggeration makes the hills easier to see. A 3,000 m volcano does not need it, and an exaggerated view would mislead anyone judging its slopes, so the summit is shown at a factor of 1.

Moving in a scene

On a flat map you pan and zoom. In a scene you also move a **camera**: a position in three dimensions, with a height above the ground, a **pitch** (how far it looks down from the horizon) and a **heading** (the compass direction it faces).

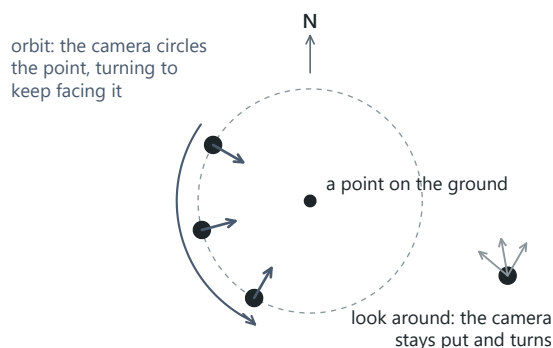
A heading is an angle counted clockwise from north: 0° is north, 90° east, 180° south and 270° west, so a heading of 78° faces east-north-east, a little north of east. ArcGIS Pro shows the numbers in a small **Camera** panel (View tab, Navigation group, beside the Navigator button): the camera's position, its heading, pitch and roll. Without the panel, you can use the on-screen navigator: its N always points north, so when you face east it points to your left.

Pitch and height above the ground



A new scene often opens looking straight down; a camera only a few metres above the ground shows nothing but ground.

Orbit and look around, from above



Heading is the compass direction the camera faces. Pan moves it sideways; zoom moves it along its view.

Figure 11.3 The scene camera. Pitch runs from -90° , straight down, to 0° , level with the horizon; tilting means raising the pitch so the view looks across the ground. Orbiting circles a point while facing it; looking around turns the camera in place.

Pan and zoom still work as on a map. **Tilt** changes the pitch, **rotate** changes the heading, **orbit** circles the camera round a point so a building can be seen from every side, and **look around** keeps the camera where it is and turns it, like standing on the summit and turning your head. The mouse, the keyboard and a small on-screen navigator in the corner of the view all do the same things; which one you use depends on your preference and your hardware.

Two situations commonly confuse new users. A converted scene often opens looking straight down, so it looks exactly like the flat map until you tilt it. And if the camera ends up only a few metres above the ground, the view is a patch of brown or grey with nothing recognisable in it: zoom out until the features appear.

IN THE PANE

Module 11 lesson 1 reads the summit's elevation off the scene, and Waldo checks your number against the state's USGS 10 m DEM (about 3,033 m). Then you tilt, turn and zoom out towards the crater; Waldo reads the camera's pitch to confirm the view is tilted. To face the crater you turn east, using the Camera panel's heading, and if you are unsure which way you are looking, Waldo reads the camera's heading and says it in words ("facing east-north-east, 78°"). The lesson points out vertical exaggeration but does not use it.

Draping and extruding

A scene holds two kinds of layer. A flat layer is **draped**: drawn on the ground surface so that it follows every rise and dip. The imagery and the outlines of the buildings are draped when a map is converted, and a draped layer stays under 2D Layers.

A layer can also be **extruded**: each feature is pushed up by a height into a solid, a polygon into a block, a line into a wall, a point into a column. Extrusion exists only in scenes and changes only how the layer is drawn; no value in the data changes. Pro moves an extruded layer from 2D Layers to 3D Layers automatically.

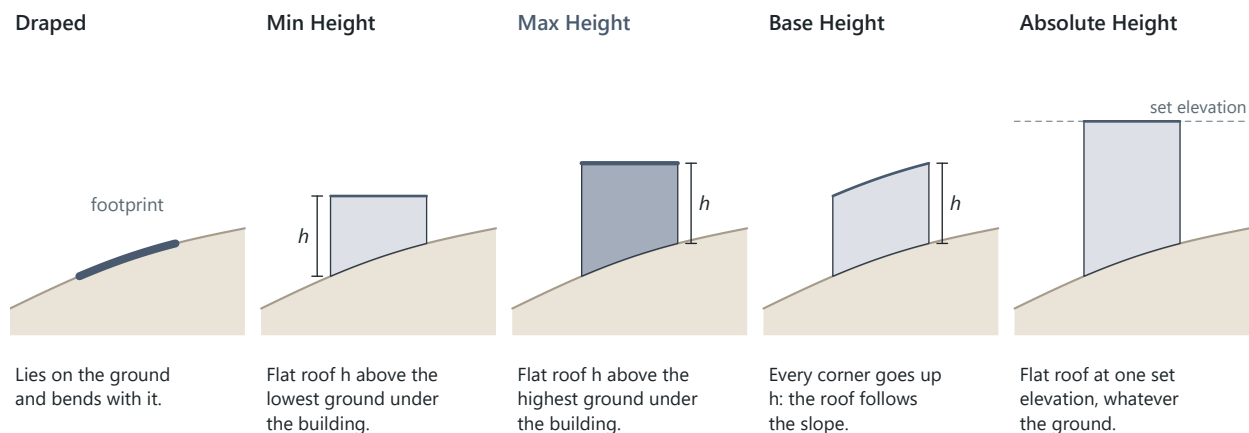


Figure 11.4 Draping and extruding a building on a slope, in section. Pro's four extrusion types differ in where they measure the height h from: the lowest ground under the building, the highest, every corner, or a set elevation.

On flat ground the three relative types give the same block. Absolute Height differs everywhere, because it sets an elevation rather than a height above the ground: a top at 10 m would be 3,000 m below the summit. On a slope the three relative types differ, and choosing one is a decision about how the building meets the ground. **Min Height** puts a flat roof h above the lowest ground under the feature, so the uphill wall is shorter than h . **Max Height** puts it h above the highest ground, so no wall is shorter than h ; it is the sensible choice for buildings on a slope, and the one the lesson uses. **Base Height** raises every corner by h , so the top follows the slope. **Absolute Height** puts the top at one set elevation, whatever the ground.

The summit's buildings

The buildings in the opening picture come from **USA Structures**, a national set of building footprints traced by Oak Ridge National Laboratory for FEMA. At the summit it holds eight footprints, from 83 m² to 1,121 m², all traced automatically from a 2017 satellite image. The layer has a field for each building's height, and at the summit it is empty on all eight.

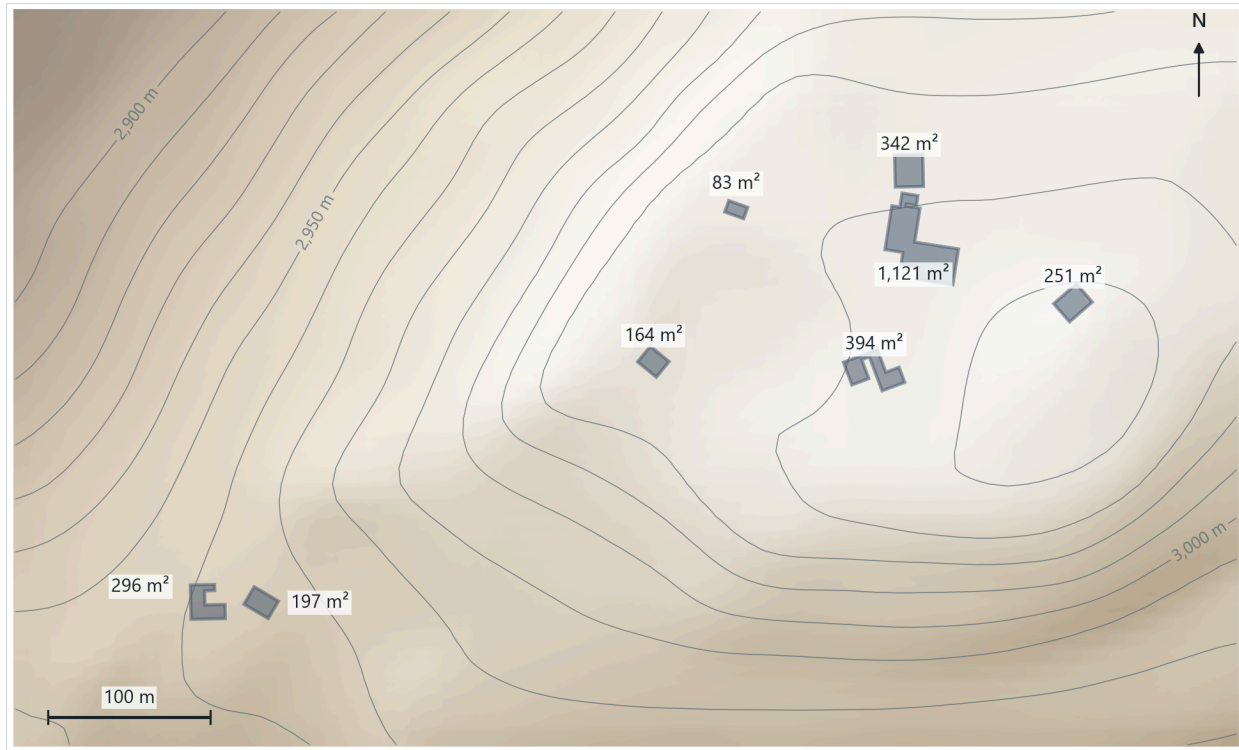


Figure 11.5 The eight USA Structures footprints at the Haleakalā summit, with their areas, over the state's 10 m DEM (contours every 10 m). The main cluster stands on the summit plateau at about 3,025 to 3,050 m; two buildings sit some 60 m lower on the western slope.

Elevation: USGS 10 m DEM, State of Hawaii copy (Hawaii Statewide GIS Program; public domain). Building footprints: USA Structures, Oak Ridge National Laboratory (ORNL) and FEMA Geospatial Response Office (CC BY 4.0)

The traced footprints are accurate enough to extrude, but they are not survey data. When they are drawn over different imagery, the outlines are a few metres off the roofs, and some of the summit's smaller structures, dishes and sheds, have no footprint at all. These are the limits of an automated national dataset traced from one image, and on a summit this steep every image is shifted a little.

The ground under a footprint is rarely level. Under the eight summit buildings it falls between 0.9 m and 7.1 m from one side to the other. The largest footprint stands on ground from 3,028.7 m to 3,035.8 m, so with Max Height and 10 m its roof is at 3,045.8 m: 10 m above its uphill corner and about 17 m above its downhill one. With Min Height the same roof would be 10 m above the downhill corner and barely 3 m above the uphill one.

READING THE RESULT

- The question: how do the summit's buildings stand on the mountain?
- The representation: eight traced footprints extruded with Max Height by a single estimated height of 10 m.
- What it shows: where the buildings stand and how the ground falls under them.
- What it cannot establish: any building's real height; the height field is empty at the summit.

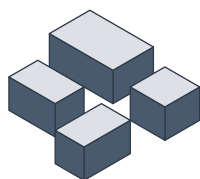
IN THE PANE

Module 11 lesson 2 looks at the draped layers first, then opens a footprint's pop-up to show that the height field is empty. You extrude all eight footprints with Max Height and a single value of 10 m, watch the layer move into 3D Layers, orbit and look around, and export a picture of the view. Waldo reads the scene to confirm the footprints are extruded, and the project folder to confirm the picture is there.

One number or a field

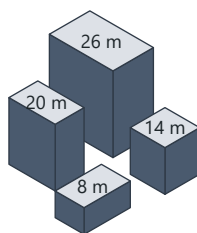
A height can come from a single value typed for every feature or from a field. When the data records each building's height, in a height field or as a count of storeys that can be multiplied by a storey height, extruding by that field gives every building its own height. When it does not, as at the summit, one value for all is the usual choice, and that value is a guess (leaving the footprints draped is the other honest choice): the large telescope domes are almost certainly taller than 10 m. A survey, a lidar surface or a storeys field would show by how much.

One number



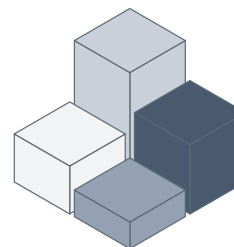
Every building 10 m: the data has no heights, so one value is typed.

A height field



Each building its own height, read from a field (heights here invented).

An amount and a rate



height = an amount (people)
colour = a rate (share over 65)



low high
Height reads like size: extrude an amount, show a rate with colour.

Figure 11.6 Three ways to give features a height. One number: every building the same height, because the data has none. A height field: each building its own height, read from the data. An amount and a rate: height shows a count, colour shows a rate (values illustrative).

Extrusion can also show a thematic value, turning a choropleth map into a set of prisms. Height then works as a visual variable in the same way as size: readers see a taller prism as more of something. The lecture therefore advises extruding by a count, an amount of something, and showing a rate or a ratio with colour, as Chapter 4 does on flat maps. One caution applies: readers judge a prism's bulk as well as its height, so a large area with a moderate value can look like more

than a small area with a high one. Module 11 does not build such maps; the idea is included so that you can interpret one.

RULE OF THUMB

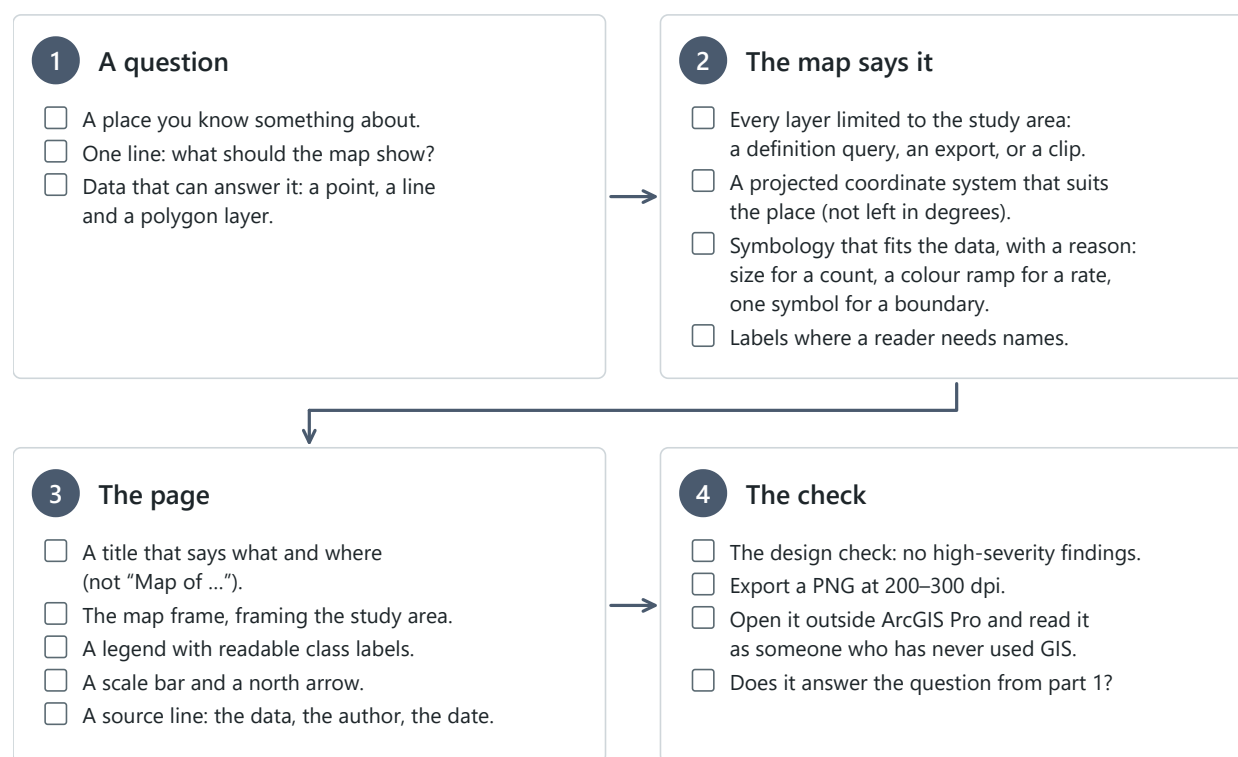
Extrude by a field when the data holds heights. When it does not, one value for everything is fine, but say that it is a guess, on the picture itself and not only in the text.

Your final map

The last lesson of the course works differently from the others. You choose a place and a question, and make a map of it on your own, with anything the course has taught; the lesson gives the brief and a checklist rather than step-by-step instructions. The figure below shows that brief on one page.

Your final map

The brief in four parts, in order: the map before the page. The map is finished when every box is ticked.



Part 4 sends you back: fix what the check or your own eyes find, then export and look again. It is the same loop as Chapter 6's: ask, look, refine.

Figure 11.7 The final map's brief as a checklist in four parts: a question, a map that answers it, a page, and a check. After the check you return to the map as often as needed.

The order of the work is important. Finish **the map** before you start the page. National or continental layers added straight from a service cover far more than the study area; limit each layer to the place with a definition query, an exported selection or a clip (Chapters 4, 6 and 7). A map made from ArcGIS Pro's Map template starts in Web Mercator, a projected system that is fine for

looking at a state or a county; if the map measures anything, or asks the reader to compare areas by eye, give it a projection chosen for the place (Chapter 3). A map that started empty takes its coordinate system from the first layer added, which is often latitude and longitude, and everything on it looks stretched; give that map a single projected coordinate system that suits the place, chosen for a reason (Chapter 3). The map's coordinate system reprojects the layers on the fly, so nothing needs to be rewritten on disk. Then choose symbology that fits the data, for a reason you could explain (a size for a count, a colour ramp for a rate, one symbol for a boundary), and add labels only where a reader needs names.

Only then make **the page**: a title that states what and where (the words "Map of" at the start of a title add no information), the map frame, a legend, a scale bar and a north arrow, and a small source line with the data, the author and the date (Chapter 4). A layout started from a template includes its own map frame, linked to the template's own map; set the frame to show your map before you do anything else.

Then run **the check**. Waldo's design check measures what can be measured on the layout: whether the frame fits the study area, whether text is large enough to read, contrast, overlaps and margins. Resolving any high-severity finding is part of the brief; the other findings are suggestions. Export the layout as a PNG at 200 to 300 dpi, open the file outside ArcGIS Pro, and look at it as someone who has never used GIS would. The last box on the checklist is the question you started with: does the picture answer it? If it does not, return to the map. Add one line about what no check can measure: why these data and this method fit the question, and what the map does not show. This is the cycle of asking, looking and refining from Chapter 6, applied one last time to your own work.

RULE OF THUMB

Make the map answer the question before you make the page, and judge the page by the exported image rather than by the layout view.

Where to make it

You can use any place, and the lesson offers the ones the course has already visited, each with a point, a line and a polygon layer registered and ready to add. Working in a place you already know saves searching for data and makes the question the new part of the work.

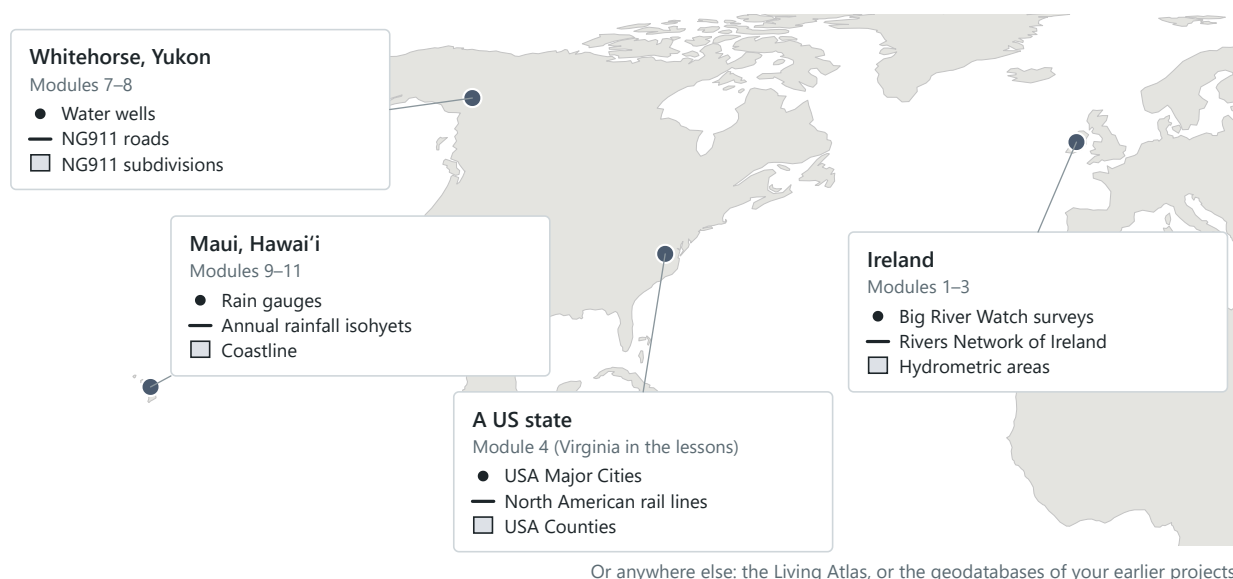


Figure 11.8 The final map's places menu: four places the course has worked in, each with a point, a line and a polygon layer ready to add. Any other place can also be used.

Land: Natural Earth (public domain)

A good question is one line that a map can answer: where the rain falls on Maui, which Virginia counties the railways miss, where Whitehorse's deep wells are. The question determines which layers are needed, which one is shown by colour and what the title says. The same data can support many maps, and the question determines which one you make.

IN THE PANE

Module 11 lesson 3 opens with the places menu and asks for your place and question in one line. After that you make the decisions: Waldo reminds you how anything was done and names the lesson that taught it. Each time you say "check my work", it checks the whole brief in one pass (your layers, coordinate system and layout, its design check, the PNG) and names anything still missing with where to find it; it does not move you on to the export while a part of the page is missing or a high-severity finding is open. It never builds the map for you. If the model in use can see pictures, it can also review the exported PNG itself, as suggestions, never a grade.

TRY IT

In any course project with a map open, ask Waldo in Ask mode: if I made this map into a local scene, which of my layers could I extrude, and by what? Waldo reads the layers and their fields and says which have a height, a count or neither. Compare its answer with the rules of thumb in this chapter.

Key terms

scene

a map drawn in three dimensions, with a ground surface, flat layers draped on it and features that can stand up from it.

global scene

a scene of the whole round Earth, for large areas where the Earth's curvature matters.

local scene	a scene of a flat piece of the world in a projected coordinate system, for a site, a town or a mountain.
ground	the elevation surface a scene stands on, built from one or more DEMs.
vertical exaggeration	a factor that multiplies every height of the ground; 2 doubles the relief, 1 shows it true.
camera pitch	how far a scene's camera looks down from the horizon: -90° straight down, 0° level.
heading	the compass direction a scene's camera faces, counted clockwise from north: 0° north, 90° east, 180° south, 270° west.
orbit	moving a scene's camera in a circle round a point while it keeps facing the point.
draping	drawing a flat layer on the ground surface of a scene, so that it follows the terrain.
extrusion	drawing each feature pushed up by a height into a solid (point to column, line to wall, polygon to block); a display setting, not a change to the data.
Max Height	the extrusion type that puts a flat top at the given height above the highest ground under each feature.
2D Layers / 3D Layers	the headings a scene's Contents sorts layers under: flat layers drawn on the ground, and layers that stand up from it.
study area	the place a map is about; the data on the map is limited to it.
source line	the small text on a finished layout naming the data, the author and the date.

Further reading

- Esri. "Extrude features to 3D symbology." ArcGIS Pro documentation. <https://doc.esri.com/en/arcgis-pro/latest/help/mapping/layer-properties/extrude-features-to-3d-symbology.html>
- Esri. "Author and share a local scene," the Get Started tutorial the lecture's scene exercise follows. ArcGIS Pro documentation. <https://pro.arcgis.com/en/pro-app/latest/get-started/author-and-share-a-local-scene.htm>
- Harder, C., & Brown, C. (eds.) (2017). *The ArcGIS Book*, the chapter on 3D. Esri Press. <https://learn.arcgis.com/en/arcgis-book/>
- The course's data sources: the State of Hawaii's copy of the USGS 10 m DEM, USA Structures and every layer on the places menu, each with its publisher and licence.

Glossary

2D Layers / 3D Layers — the headings a scene's Contents sorts layers under: flat layers drawn on the ground, and layers that stand up from it. 11

absolute path — a path that starts at the drive letter and names every folder down to the data. 2

add-in — a package that extends ArcGIS Pro with new buttons and panes, installed from an .esriAddinX file. 0

adjacency — features meeting along a boundary without overlapping. 6

alias — a readable label shown in place of a field's name. 5

API key — a secret string that identifies you to a provider; treat it like a password. 0

.aprx — the project file you open; it stores maps, layouts and connections, and points at data rather than containing it. 0

ArcGIS Online — Esri's cloud platform where maps, apps and data are shared, reached in Pro through the Portal tab. 2

Ask mode — Waldo reads and explains your project but cannot change it; every lesson runs in Ask. 0

aspect — the compass direction a slope faces, 0° to 360° from north, -1 for flat cells. 8

attribute — a value describing a feature, stored in one field of its row. 1

attribute query — a query that selects records by their attribute values, written as an SQL WHERE clause. 6

attribute table

- one row per feature, one column per field. 1
- the table tied to a layer, with one row per feature. 5

azimuthal — preserving true directions from a centre point; can also be conformal, equal-area or equidistant. 3

band

- one layer of values in a raster; a land-cover map has one, a colour image has several. 2
- one layer of values in a raster; a multiband image such as Landsat holds several per cell. 8

basemap — the reference layers at the bottom of a map that give it context. 1

bookmark — a saved extent (in a scene, a saved camera) that a view can return to by name. 1

Boolean raster — a raster of 1 (true) and 0 (false) cells, the answer to a condition asked of every cell. 8

box plot — a chart of a distribution's median, quartiles and range, suited to comparing groups. 6

broken link — a layer whose data can no longer be found at its stored path, marked with a red exclamation mark and repaired from the Source tab. 2

buffer — a polygon covering every place within a given distance of a feature. 7

Calculate Field — a tool that fills a field from an expression over each row's other fields. 5

Calculate Geometry — a tool that fills a field with each feature's area, length, perimeter or coordinates. 5

camera pitch — how far a scene's camera looks down from the horizon: -90° straight down, 0° level. 11

cardinality — how many rows one key value matches on each side of a join: one-to-one, many-to-one, one-to-many or many-to-many. 5

cartography — the art and science of designing maps. 4

case field — the field whose values define the groups, one output row per value, in a summary. 6

cell (pixel)

- one square of a raster grid; its size sets the smallest thing the raster can show. 2
- one square of a raster, standing for one square of ground and holding one value. 8

cell size (spatial resolution) — the ground length of one cell's side; nothing smaller than a cell exists in the raster. 8

checkpoint — the check at the end of a lesson step, made by reading your actual project. 0

choropleth map — a map of areas shaded by class of a rate or proportion (Graduated Colors in Pro). 4

classification — grouping features into a few classes by attribute value, with a stated method. 4

clip — keeps the parts of the input features inside a boundary; the input's attributes are unchanged. 7

colorizer — the rule that draws a raster: unique values, stretch, classify or RGB. [8](#)

compromise projection — a projection that keeps no property exactly but every distortion moderate, such as Robinson or Winkel Tripel. [3](#)

conformal — preserving local angles, so small shapes are true; areas are distorted. [3](#)

construction tool — the tool that turns clicks into a shape, such as Line, Arc, Polygon or Rectangle. [10](#)

contextual tab — a ribbon tab that appears only while something that needs it is selected. [1](#)

cross-validation — testing an interpolation by leaving out each sample in turn, predicting its value from the others and comparing the prediction with the measurement. [9](#)

data source — the path or web address a layer uses to find its data, shown on the layer's Source tab. [2](#)

datum — the ellipsoid and its position relative to the Earth, which give latitude and longitude a definite meaning; local datums fit one region, geocentric ones the whole Earth. [3](#)

definition query — a layer filter that hides the features failing a test, without changing the stored data. [6](#)

destination / source table — in a join, the table that is added to / the table whose columns are added. [5](#)

developable surface — a cylinder, cone or plane: a surface that unrolls flat without stretching, used to build and classify projections. [3](#)

digital elevation model (DEM) — a raster of ground heights. [8](#)

dissolve — merging the features of one layer that share a field value, keeping that field and any requested statistics. [7](#)

dissolve (buffer option) — merging overlapping buffer polygons into one feature so that no area is counted twice. [7](#)

distance decay — the weakening of a relationship or an influence as distance grows. [9](#)

Do mode — Waldo can act on your project: run tools, change symbology, build layouts. [0](#)

domain — a list or range of allowed values for a field, offered as a pick list when editing. [10](#)

draping — drawing a flat layer on the ground surface of a scene, so that it follows the terrain. [11](#)

drawing order — the order of layers in the Contents pane; a layer higher in the list draws on top. [1](#)

dynamic label / annotation — text placed automatically from a field; text placed and stored piece by piece. [4](#)

ecological fallacy — the mistake of assuming that what is true of an area on average is true of the individuals in it. [4](#)

ellipsoid (spheroid) — the smooth shape made by rotating an ellipse about its shorter axis; the mathematical model of the Earth, defined by its two semi-axes. [3](#)

equal interval — a classification method that divides the range into classes of equal width. [4](#)

equal-area — preserving the relative sizes of areas everywhere; shapes are distorted. [3](#)

equidistant — preserving true scale along certain lines, such as from one or two points to all others. [3](#)

erase — keeps the parts of the input features outside a boundary; the input's attributes are unchanged. [7](#)

extent — the rectangle of the map shown in a view at a given moment. [1](#)

Extract by Mask — a tool that keeps a raster's cells inside a polygon or raster mask and sets the rest to NoData. [8](#)

Extract Values to Points — a tool that writes the value of the cell under each point into the point's table. [8](#)

extrusion — drawing each feature pushed up by a height into a solid (point to column, line to wall, polygon to block); a display setting, not a change to the data. [11](#)

false-colour composite — an image with bands other than red, green and blue sent to the screen's colours, such as near infrared into red. [8](#)

feature — one point, line or polygon in a layer, together with its row of attributes. [1](#)

feature class

— a table in a geodatabase whose shape field holds point, line or polygon geometry. [2](#)

— a set of features of one geometry type, with the same fields and coordinate system, stored in a geodatabase. [10](#)

feature dataset — a group of feature classes in a geodatabase that share one coordinate system. [2](#)

feature layer — a web layer that serves features and attributes that can be queried and, when the publisher allows it, copied. 2

feature template — a recipe for new features: the target layer, its symbol, its construction tools and default attribute values. 10

FID_ column — a column an overlay adds to record which feature of an input each piece came from; -1 means none. 7

field — one column of a table: one attribute of every record, with a name, a type and a definition. 5

field length — the number of characters a Text field can hold. 5

figure-ground — the contrast that separates a map's subject from its background. 4

file geodatabase — Pro's container for data, stored as a folder ending in .gdb and managed through Pro. 2

flattening — $f = 1 - b/a$, how much shorter the polar radius is than the equatorial; about 1/298 for the Earth, shown in dialogs as its inverse, 1/f. 3

Float / Double — decimal storage types keeping six or seven and about fifteen significant digits. 5

floating-point raster — a raster of decimal values, usually measurements such as elevation; it has statistics instead of a table. 8

folder connection — a shortcut in the Catalog pane to a folder on disk, the way downloaded data reaches a map. 2

generalization — selecting, simplifying, classifying and symbolizing geographic information so it can be shown at a given scale and for a given purpose. 4

geodatabase — a container for GIS data (feature classes, tables, rasters); a file geodatabase is a .gdb folder that Pro shows as one item. 0

geodesic — the shortest path between two points along the curved surface of the ellipsoid; geodesic measurement ignores the map's projection. 3

geographic coordinate system (GCS) — latitude and longitude in degrees, defined by a datum, a prime meridian and an angular unit. 3

geographic transformation — a calculation that moves coordinates from one datum to another, such as NAD 1927 to NAD 1983, so that data line up. 3

geoid — the irregular surface of equal gravity potential that mean sea level follows; the reference for heights, within about ± 100 m of the ellipsoid. 3

GeoJSON — a plain-text format for features built on JSON, an IETF standard since 2016. 2

geometry type — point, line or polygon; fixed when the feature class is created. 10

geoprocessing environment — a setting applied around a tool, such as the processing extent. 7

global scene — a scene of the whole round Earth, for large areas where the Earth's curvature matters. 11

graduated / proportional symbols — point symbols sized by class; point symbols sized in exact proportion to each value. 4

graticule — the network of meridians and parallels drawn on a globe or a map. 3

ground — the elevation surface a scene stands on, built from one or more DEMs. 11

halo — a light outline around lettering that separates it from a busy background. 4

heading — the compass direction a scene's camera faces, counted clockwise from north: 0° north, 90° east, 180° south, 270° west. 11

heads-up digitizing — tracing features on screen by clicking their vertices over an image. 10

hillshade — a raster of the brightness each cell would have under a lamp at a set direction and height. 8

histogram

— a chart of how many features have each value. 4

— a chart of one numeric field's distribution, as counts in value bins. 6

hue / saturation / value — which colour; how pure versus greyed; how light or dark. 4

identity — an overlay that keeps the first input's footprint, split wherever the second input covers it. 7

imagery layer — a raster served live from a server. 2

IN — an SQL operator that tests a field against a list of values. 6

integer raster — a raster of whole numbers, usually codes such as land-cover classes; it can have an attribute table. 8

interpolation — estimating values at unsampled places from nearby sample points, usually into a raster. 9

intersect — an overlay that keeps only the pieces lying in both inputs, each carrying both inputs' attributes. 7

interval / ratio — quantities whose zero is arbitrary / quantities with a true zero, where twice the number means twice as much. 5

inverse distance weighting (IDW) — interpolation by a weighted average of nearby samples, each weight one over the distance to a power. [9](#)

Irish Transverse Mercator (ITM) — Ireland's national grid (EPSG 2157): transverse Mercator on IRENET95, central meridian 8° W, scale factor 0.99982, false origin 600,000 m E, 750,000 m N. [3](#)

isohyet — a line joining places of equal rainfall; the rain version of an elevation contour. [9](#)

join — attaching the columns of a source table to the rows of a destination table through a key field. [5](#)

kernel density — a surface made by spreading each point into a hill of a chosen radius and adding the hills, giving points per unit of area. [9](#)

key field — a field in both tables whose values match a row on one side to a row on the other. [5](#)

KML — an XML format for geographic display that began with Google Earth and is now an OGC standard. [2](#)

kriging — a statistical interpolation method that fits its weights to how similarity falls off with distance and also estimates its own uncertainty (named only). [9](#)

large scale / small scale — a large-scale map (1:10,000) shows a small area in detail; a small-scale map (1:1,000,000) a large area with little detail. [3](#)

layer

— one dataset drawn in a map with its own name and symbology; the data stays where it is. [1](#)

— a map's reference to a dataset together with its name, symbology and other drawing settings. [2](#)

layout

— a page in a project that arranges maps and map elements for printing or export. [1](#)

— a fixed-size page holding map frames and map elements. [4](#)

Learn panel — the list of every module and lesson in the Waldo pane, with your progress marked and Start, Resume and Check my work buttons. [0](#)

legibility / harmony — whether text can be read; one or two typefaces used consistently. [4](#)

lesson — a short sequence of steps, done in Pro, that teaches one idea. [0](#)

level of measurement — whether values are nominal, ordinal, interval or ratio, which determines the arithmetic that makes sense. [5](#)

LIKE — an SQL operator that matches text against a pattern, with % standing for any run of characters. [6](#)

Living Atlas

— Esri's curated collection of ready-to-use layers, reached from the Portal tab. [1](#)

— Esri's curated collection of maps and data, in public, subscriber and premium tiers. [2](#)

local scene — a scene of a flat piece of the world in a projected coordinate system, for a site, a town or a mountain. [11](#)

M value / Z value — an optional measure (for routes) or height stored with each vertex. [10](#)

map — a project item holding layers drawn in two dimensions. [1](#)

map algebra — arithmetic and comparisons applied to rasters cell by cell, each writing a new raster. [8](#)

map projection — a mathematical transformation of the curved Earth onto a plane; it always distorts shape, area, distance or direction somewhere. [3](#)

mask — a setting that limits a raster tool's output to the cells inside a study-area polygon. [9](#)

Max Height — the extrusion type that puts a flat top at the given height above the highest ground under each feature. [11](#)

merge / append — combining feature classes of the same type into a new one (merge) or into an existing one (append), without splitting; overlaps become duplicate features. [7](#)

metadata — data about data: who made it, what it is, when, where, how and why. [2](#)

modifiable areal unit problem — the way a mapped pattern can change when the same data are summarized over differently drawn areas. [4](#)

module — a group of lessons on one topic; the course has twelve, numbered 0 to 11. [0](#)

multipart feature — one feature, one row, made of several separate shapes. [7](#)

multiple ring buffer — buffers at several distances around the same features, stored as non-overlapping rings or overlapping disks. [7](#)

natural breaks (Jenks) — a classification method that places breaks where the data thin out, keeping each class's values alike. [4](#)

neatline — an optional line enclosing the map or the whole layout. [4](#)

NoData — a cell that holds no value, such as a cell outside the mask of an extraction. 8

node — the first (start) or last (end) vertex of a line; lines that meet should share one. 10

nominal / ordinal — values that are names or categories / categories with a meaningful order. 5

normalization — dividing a count by another field, such as area or population, to make a rate. 4

null — an empty field value, which records that nothing is known and is not the same as zero. 5

OBJECTID — the unique number a geodatabase keeps for every row of a table. 5

on-the-fly projection — the software's redrawing of every layer in the map's coordinate system at display time, without changing the stored data. 3

optical centre — the point a little above the geometric centre of a page where the eye first lands. 4

orbit — moving a scene's camera in a circle round a point while it keeps facing the point. 11

overlay — combining two feature classes by location into a new feature class. 7

pane — a dockable window inside ArcGIS Pro, such as Contents, Catalog or Waldo. 0

planar — measured with flat geometry in a projected coordinate system, so only as accurate as that projection is where you measure. 3

population density — people per unit of area; a rate, suited to a choropleth where a count is not. 5

pop-up — the window that shows one feature's attributes when the feature is clicked. 1

power (IDW) — the exponent that sets how fast a sample's weight falls with distance; 2 by default. 9

project — an ArcGIS Pro workspace: a folder holding the .aprx file, a default geodatabase, a toolbox and housekeeping folders. 0

project (.aprx) — the Pro file that holds maps, layouts and connections and a pointer to each layer's data, but not the data itself. 2

projected coordinate system (PCS) — a GCS plus a map projection and a linear unit, giving flat x and y coordinates in metres or feet. 3

provider — the company or program that runs the language model Waldo uses. 0

qualitative / sequential / diverging scheme — distinct hues for categories; an ordered light-to-dark run for amounts; two hues meeting at a pale middle for amounts with a meaningful midpoint. 4

qualitative attribute — an attribute that tells kinds of things apart: a category, a name, a code. 4

quantile — a classification method that puts the same number of features in every class. 4

quantitative attribute — an attribute that measures an amount: a count, an area, a rate, a percentage. 4

R² — the share of the variation in the vertical-axis variable that the fitted line accounts for, from 0 (none) to 1 (all). 6

raster

— a layer made of a regular grid of cells, each holding one value. 1

— a data model that covers the ground with a grid of equal cells in rows and columns, each holding one value. 2

— a grid of equal cells in rows and columns covering an area, each cell holding one value. 8

raster attribute table

— an integer raster's table: one row per distinct cell VALUE, with a COUNT of the cells that hold it. 5

— one row per cell value of an integer raster, with the fields Value and Count and any added columns. 8

Raster Calculator — ArcGIS Pro's tool for map-algebra expressions, with quoted raster names and strict operators. 8

Raster to Polygon — a tool that turns groups of equal integer cells into polygons, one row per polygon. 8

Reclassify — a tool that replaces values or ranges of values with new values; its output is always integer. 8

record — one row of a table; in an attribute table, one feature. 5

reference map — a map of where things are, with many features at roughly equal weight, for many purposes. 4

relate — a link between two tables that keeps them separate instead of merging their columns. 5

relative path — a path written from the folder that holds the project file, so the project folder can move as a whole. 2

relief map — a DEM in an elevation colour scheme drawn partly transparent over its own hillshade. 8

representative fraction (RF) — a map scale written as 1:n, one unit on the map for n of the same unit on the ground. [3](#)

reshape / merge / split — redrawing part of a feature's edge; making one feature of several; cutting one feature into two. [10](#)

ribbon — the band of tabs across the top of Pro, each holding groups of commands. [1](#)

save edits — writing the edits made since the last save into the geodatabase; separate from saving the project. [10](#)

scale factor — the actual scale at a point of a projected map divided by the stated scale; 1 on the standard lines. [3](#)

scatter plot — a chart of two numeric fields, one point per feature, for reading their relationship. [6](#)

scene

— a project item that draws layers in 3D, seen through a camera that tilts and turns. [1](#)

— a map drawn in three dimensions, with a ground surface, flat layers draped on it and features that can stand up from it. [11](#)

search neighbourhood — the samples an estimate may use: a set number of the nearest, or all within a set radius. [9](#)

search radius — in kernel density, the distance at which each point's hill falls to zero; the setting that most changes the map. [9](#)

segment — the straight piece of a line or polygon boundary between two consecutive vertices. [10](#)

selection — the features currently highlighted, shown on the map, in the table and in the count; replaced by the next new query. [6](#)

selection type — how a new query combines with the current selection: new, add (OR), remove (NOT), subset (AND), switch or clear. [6](#)

sentinel value — a code such as -9999 stored in place of a missing measurement; it must be removed before any calculation. [9](#)

service URL — the web address of one published layer, which can be added to a map directly. [1](#)

shapefile — an older vector format stored as a family of same-named files (.shp, .shx, .dbf and optional others) that must be kept together. [2](#)

Short / Long — whole-number storage types, up to 32,767 and up to about 2.1 billion. [5](#)

site-selection model — a chain of spatial operations that keeps only the places meeting every written criterion. [7](#)

sketch — the shape being drawn, before finishing turns it into a feature. [10](#)

skewed distribution — values piled at one end with a long tail at the other. [4](#)

slope — the steepest rate of change of height from a cell to its neighbours, in degrees or percent. [8](#)

snap raster — an environment that aligns a new raster's cells exactly with an existing raster's cells. [8](#)

snap tolerance — the distance, in pixels or map units, within which snapping pulls the pointer onto a target. [10](#)

snapping — moving a vertex exactly onto an end, vertex or edge of another feature when the pointer comes within a tolerance. [10](#)

source line — the small text on a finished layout naming the data, the author and the date. [11](#)

spatial autocorrelation — the tendency of values at nearby places to be related: alike (positive) or unlike (negative), rather than random. [9](#)

spatial data exploration (ESDA) — interactive, iterative questioning of data through linked maps, tables and charts. [6](#)

spatial join — attaching one layer's attributes to another's features by their spatial relationship. [7](#)

spatial query — a query that selects features by their spatial relationship to other features. [6](#)

spatial relationship — the geometric test in a spatial query, such as intersect, within, contains, touches or within a distance. [6](#)

SQL — Structured Query Language, the language of relational databases: SELECT fields FROM a table WHERE a test is true. [6](#)

standalone table — a table with no geometry, such as a census table or an imported spreadsheet. [5](#)

storage type — how a field stores its values: Short, Long, Float, Double, Text, Date or Blob. [5](#)

study area — the place a map is about; the data on the map is limited to it. [11](#)

suitability analysis — answering "where?" from criteria every cell can pass or fail; the course uses the simplest form, one Boolean raster per criterion, multiplied. [8](#)

summarized join — a spatial join that gives each target feature a count or statistic of the join features it matches. [7](#)

summary statistics — statistics of a field computed per group of records; the output is always a table. [6](#)

symbolology — the drawing rules of a layer: its colours, symbols and sizes. [1](#)

table (standalone table) — rows and columns with no geometry. [2](#)

tangent / secant — a tangent surface touches the globe along one line or at one point; a secant surface cuts it along two standard lines (a plane along one circle). [3](#)

target / join features — in a spatial join, the layer that keeps its shapes and rows / the layer that supplies the attributes. [7](#)

thematic map — a map of one subject for one purpose, with everything else as context. [4](#)

Thiessen (Voronoi) polygons — polygons that assign every location to its nearest sample point; also called proximal zones. [9](#)

tile layer / map image layer — a web layer served as a picture of a map, pre-drawn in tiles or drawn by the server on request. [2](#)

Tobler's first law of geography — "everything is related to everything else, but near things are more related than distant things" (Waldo Tobler, 1970). [9](#)

transformation — an analysis operation that reads existing layers and writes a new feature class, leaving the inputs as they were (Append, which adds to an existing dataset, is the exception). [7](#)

undershoot / overshoot — a line that stops short of the line it should meet, or crosses past it. [10](#)

undo / discard — taking back the last edit, or every edit since the last save. [10](#)

union — an overlay that keeps every piece of both inputs, each piece carrying both inputs' attributes. [7](#)

UTM (Universal Transverse Mercator) — 60 transverse Mercator zones, each 6° wide, with a central-meridian scale factor of 0.9996 and coordinates in metres. [3](#)

vector — a data model that represents features as points, lines and polygons built from coordinates, with one row of attributes per feature. [2](#)

vertex

— one stored coordinate pair along a line or a polygon's boundary; straight segments join the vertices. [2](#)

— one x, y coordinate of a line or polygon; the shape is the list of its vertices. [10](#)

vertical exaggeration — a factor that multiplies every height of the ground; 2 doubles the relief, 1 shows it true. [11](#)

view — a work area showing a map, layout, scene or table; one view is active. [1](#)

visual hierarchy — design that makes the most important elements the most prominent. [4](#)

visual variable — a graphic property of a symbol, such as size, weight, shape, orientation, texture, hue, value or saturation, that the eye reads as a difference. [4](#)

Web Mercator — the spherical Mercator formula applied to WGS 84 coordinates (EPSG 3857); the system of most web basemaps, stretching lengths by $1/\cos(\text{latitude})$. [3](#)

WHERE clause — the test part of an SQL query: a field, a comparison and a value, joined to others by AND, OR and NOT. [6](#)

XY resolution — the fine grid on which a feature class stores its coordinates. [10](#)

XY tolerance — how close two coordinates may be before they are treated as the same point. [10](#)

Figure credits

Every figure not listed here (53 in this edition) was made for this book from the course's registered data and public-domain sources and is covered by the book's licence, CC BY-NC-ND 4.0 © the author. Of the figures listed below, those marked “book figure, drawn from the credited data” were also drawn for this book, from the data named in the credit, whose publishers keep their own terms; the rest reuse someone else's picture under the licence shown.

Chapter 0 · [Start here](#)

Figure	Credit	Licence
Chapter 0 opener	Countries: World Countries (Generalized), Esri, ArcGIS Living Atlas of the World	book figure, drawn from the credited data

Chapter 1 · What a GIS is, and the Pro workspace

Figure	Credit	Licence
Chapter 1 opener	Basins: Hydrometric Areas Ireland & Northern Ireland; rivers: Rivers Network of Ireland (ArcGIS Living Atlas of the World). Source: Environmental Protection Agency (Ireland), CC BY 4.0. Land cover: Sentinel-2 10m Land Use/Land Cover 2024, Impact Observatory, Microsoft and Esri, ArcGIS Living Atlas of the World (source data CC BY 4.0)	book figure, drawn from the credited data
Figure 1.1	Basins: Hydrometric Areas Ireland & Northern Ireland; rivers: Rivers Network of Ireland (ArcGIS Living Atlas of the World). Source: Environmental Protection Agency (Ireland), CC BY 4.0	book figure, drawn from the credited data
Figure 1.5	Basins: Hydrometric Areas Ireland & Northern Ireland; rivers: Rivers Network of Ireland (ArcGIS Living Atlas of the World). Source: Environmental Protection Agency (Ireland), CC BY 4.0	book figure, drawn from the credited data
Figure 1.6	Basins: Hydrometric Areas Ireland & Northern Ireland. Source: Environmental Protection Agency (Ireland), CC BY 4.0	book figure, drawn from the credited data

Chapter 2 · Data models and where data lives

Figure	Credit	Licence
Chapter 2 opener	Basins: Hydrometric Areas Ireland & Northern Ireland (ArcGIS Online); rivers: Rivers Network of Ireland (ArcGIS Living Atlas of the World). Source: Environmental Protection Agency (Ireland), CC BY 4.0. Land cover: Sentinel-2 10m Land Use/Land Cover Time Series, most recent year, Impact Observatory, Microsoft and Esri (ArcGIS Living Atlas of the World; source data CC BY 4.0). Towns: Natural Earth (public domain)	book figure, drawn from the credited data
Figure 2.3	Towns: Natural Earth (public domain). Rivers: Rivers Network of Ireland (ArcGIS Living Atlas of the World); basins: Hydrometric Areas Ireland & Northern Ireland (ArcGIS Online). Source: Environmental Protection Agency (Ireland), CC BY 4.0	book figure, drawn from the credited data
Figure 2.5	Land cover: Sentinel-2 10m Land Use/Land Cover Time Series, most recent year, Impact Observatory, Microsoft and Esri (ArcGIS Living Atlas of the World; source data CC BY 4.0)	book figure, drawn from the credited data

Figure	Credit	Licence
Figure 2.6	River: Rivers Network of Ireland (ArcGIS Living Atlas of the World), Environmental Protection Agency (Ireland), CC BY 4.0. Land cover: Sentinel-2 10m Land Use/Land Cover Time Series, Impact Observatory, Microsoft and Esri (ArcGIS Living Atlas of the World; source data CC BY 4.0)	book figure, drawn from the credited data
Figure 2.11	Facts: the Rivers Network of Ireland item on ArcGIS Online (Esri Ireland Content; source Environmental Protection Agency, CC BY 4.0)	book figure, drawn from the credited data

Chapter 3 · Spatial reference

Figure	Credit	Licence
Chapter 3 opener	Countries: World Countries (Generalized), ArcGIS Living Atlas of the World. Sources: Esri; Garmin International, Inc.; U.S. Central Intelligence Agency (The World Factbook); National Geographic Society	book figure, drawn from the credited data
Figure 3.2	Geoid heights: EGM96, U.S. National Geospatial-Intelligence Agency (public domain), 15' grid via PROJ-data. Coastlines: Natural Earth (public domain). Colours: ColorBrewer RdBu, Cynthia Brewer (Apache 2.0)	book figure, drawn from the credited data
Figure 3.4	Land: Natural Earth (public domain)	book figure, drawn from the credited data
Figure 3.7	Basins: Hydrometric Areas Ireland & Northern Ireland; rivers: Rivers Network of Ireland (ArcGIS Living Atlas of the World). Source: Environmental Protection Agency (Ireland)	book figure, drawn from the credited data
Figure 3.9	Countries: World Countries (Generalized), ArcGIS Living Atlas of the World. Sources: Esri; Garmin International, Inc.; U.S. Central Intelligence Agency (The World Factbook); National Geographic Society	book figure, drawn from the credited data
Figure 3.11	Countries: World Countries (Generalized), ArcGIS Living Atlas of the World. Sources: Esri; Garmin International, Inc.; U.S. Central Intelligence Agency (The World Factbook); National Geographic Society	book figure, drawn from the credited data
Figure 3.12	UTM zones: World UTM Grid, Esri (ArcGIS Online). Countries: World Countries (Generalized), ArcGIS Living Atlas of the World (Sources: Esri; Garmin International, Inc.; U.S. Central Intelligence Agency (The World Factbook); National Geographic Society). Basins: Hydrometric Areas Ireland & Northern Ireland, Environmental Protection Agency (Ireland)	book figure, drawn from the credited data
Figure 3.13	Basins: Hydrometric Areas Ireland & Northern Ireland, Environmental Protection Agency (Ireland)	book figure, drawn from the credited data
Figure 3.16	Basins: Hydrometric Areas Ireland & Northern Ireland, Environmental Protection Agency (Ireland). Countries: World Countries (Generalized), ArcGIS Living Atlas of the World (Sources: Esri; Garmin International, Inc.; U.S. Central Intelligence Agency (The World Factbook); National Geographic Society)	book figure, drawn from the credited data

Chapter 4 · Cartographic design

Figure	Credit	Licence
Chapter 4 opener	Counties: USA Counties Generalized Boundaries, ArcGIS Living Atlas of the World (Sources: Esri; U.S. Census Bureau). Railroads: North American Rail Network Lines, Bureau of Transportation Statistics (public domain). Colours: ColorBrewer OrRd, Cynthia Brewer (Apache 2.0)	book figure, drawn from the credited data

Figure	Credit	Licence
Figure 4.1	State outlines: U.S. Census Bureau cartographic boundary files, 2023, at 1:20,000,000, 1:5,000,000 and 1:500,000 (public domain)	book figure, drawn from the credited data
Figure 4.2	Counties: USA Counties Generalized Boundaries; places: USA Major Cities (ArcGIS Living Atlas of the World; Sources: Esri; U.S. Census Bureau). Railroads: North American Rail Network Lines, Bureau of Transportation Statistics (public domain). Colours: ColorBrewer Dark2, Cynthia Brewer (Apache 2.0)	book figure, drawn from the credited data
Figure 4.3	Hues: ColorBrewer Dark2, Cynthia Brewer (Apache 2.0)	book figure, drawn from the credited data
Figure 4.4	Colour schemes: ColorBrewer, Cynthia Brewer, Mark Harrower and The Pennsylvania State University (Apache 2.0). This product includes color specifications and designs developed by Cynthia Brewer (http://colorbrewer.org/).	book figure, drawn from the credited data
Figure 4.5	Counties: USA Counties Generalized Boundaries; places: USA Major Cities (ArcGIS Living Atlas of the World; Sources: Esri; U.S. Census Bureau). Railroads: North American Rail Network Lines, Bureau of Transportation Statistics (public domain). Colours: ColorBrewer OrRd, Dark2 and Blues, Cynthia Brewer (Apache 2.0)	book figure, drawn from the credited data
Figure 4.6	Counties: USA Counties Generalized Boundaries, ArcGIS Living Atlas of the World (Sources: Esri; U.S. Census Bureau). Colours: ColorBrewer OrRd, Cynthia Brewer (Apache 2.0)	book figure, drawn from the credited data
Figure 4.7	Counties: USA Counties Generalized Boundaries, ArcGIS Living Atlas of the World (Sources: Esri; U.S. Census Bureau). Colours: ColorBrewer OrRd and RdBu, Cynthia Brewer (Apache 2.0)	book figure, drawn from the credited data
Figure 4.9	Counties: USA Counties Generalized Boundaries, ArcGIS Living Atlas of the World (Sources: Esri; U.S. Census Bureau). Railroads: North American Rail Network Lines, Bureau of Transportation Statistics (public domain). Colours: ColorBrewer OrRd, Cynthia Brewer (Apache 2.0)	book figure, drawn from the credited data
Figure 4.10	Colours: ColorBrewer OrRd, Cynthia Brewer (Apache 2.0)	book figure, drawn from the credited data
Figure 4.11	Counties: USA Counties Generalized Boundaries; places: USA Major Cities (ArcGIS Living Atlas of the World; Sources: Esri; U.S. Census Bureau). Railroads: North American Rail Network Lines, Bureau of Transportation Statistics (public domain). Colours: ColorBrewer OrRd, Cynthia Brewer (Apache 2.0)	book figure, drawn from the credited data

Chapter 5 · Attribute data

Figure	Credit	Licence
Chapter 5 opener	Sectors: Sector Boundary 2022; population: population_count (2022 census). Source: National Institute of Statistics of Rwanda (NISR), ArcGIS Online. Colours: ColorBrewer YlOrBr, Cynthia Brewer (Apache 2.0)	book figure, drawn from the credited data
Figure 5.1	Sectors: Sector Boundary 2022. Source: National Institute of Statistics of Rwanda (NISR), ArcGIS Online	book figure, drawn from the credited data
Figure 5.2	Values: National Institute of Statistics of Rwanda (NISR), Sector Boundary 2022 and population_count	book figure, drawn from the credited data
Figure 5.3	Values: National Institute of Statistics of Rwanda (NISR), Sector Boundary 2022 and population_count	book figure, drawn from the credited data

Figure	Credit	Licence
Figure 5.4	Values: National Institute of Statistics of Rwanda (NISR), Sector Boundary 2022 and population_count	book figure, drawn from the credited data
Figure 5.5	Values: National Institute of Statistics of Rwanda (NISR), Sector Boundary 2022 and population_count	book figure, drawn from the credited data
Figure 5.6	Values: National Institute of Statistics of Rwanda (NISR), Sector Boundary 2022 and population_count	book figure, drawn from the credited data
Figure 5.7	Values: National Institute of Statistics of Rwanda (NISR), Sector Boundary 2022 and population_count	book figure, drawn from the credited data
Figure 5.8	Values: National Institute of Statistics of Rwanda (NISR), Sector Boundary 2022 and population_count	book figure, drawn from the credited data
Figure 5.9	Values: National Institute of Statistics of Rwanda (NISR), Sector Boundary 2022 and population_count	book figure, drawn from the credited data
Figure 5.10	Sectors: Sector Boundary 2022; population: population_count (2022 census). Source: National Institute of Statistics of Rwanda (NISR), ArcGIS Online. Colours: ColorBrewer YlOrBr, Cynthia Brewer (Apache 2.0)	book figure, drawn from the credited data

Chapter 6 · Exploring spatial data

Figure	Credit	Licence
Chapter 6 opener	Map: John Snow, 1855 (lithograph by C. F. Cheffins), public domain; georeferenced scan "Snow cholera map image", Esri_Tutorials, ArcGIS Online. Deaths and pumps: Cholera Cases and Water Pumps, digitised from Snow's map by Esri_Tutorials (ArcGIS Online)	book figure, drawn from the credited data
Figure 6.2	Map: John Snow, 1855 (lithograph by C. F. Cheffins), public domain; georeferenced scan "Snow cholera map image", Esri_Tutorials, ArcGIS Online. Deaths and pumps: Cholera Cases and Water Pumps, digitised from Snow's map by Esri_Tutorials (ArcGIS Online)	book figure, drawn from the credited data
Figure 6.5	Sectors: Sector Boundary 2022; population: population_count, 2022 census. Source: National Institute of Statistics of Rwanda (NISR), ArcGIS Online	book figure, drawn from the credited data
Figure 6.7	Sectors: Sector Boundary 2022; population: population_count, 2022 census; hospitals: Health_Establishments_hospital. Source: National Institute of Statistics of Rwanda (NISR), ArcGIS Online	book figure, drawn from the credited data
Figure 6.8	Sectors: Sector Boundary 2022; population: population_count, 2022 census; hospitals: Health_Establishments_hospital. Source: National Institute of Statistics of Rwanda (NISR), ArcGIS Online	book figure, drawn from the credited data
Figure 6.10	Data: Sector Boundary 2022 and population_count (2002, 2012, 2022 censuses), National Institute of Statistics of Rwanda (NISR), ArcGIS Online	book figure, drawn from the credited data
Figure 6.11	Data: Sector Boundary 2022 and population_count (2022 census), National Institute of Statistics of Rwanda (NISR), ArcGIS Online	book figure, drawn from the credited data

Chapter 7 · Vector analysis

Figure	Credit	Licence
Chapter 7 opener	Government of Yukon, GeoYukon (Municipal Boundaries; NG911 Roads, Hydrology – Line and – Polygon – Whitehorse; Land Parcels – Surveyed). Contains information licensed under the Open Government Licence – Yukon.	book figure, drawn from the credited data
Figure 7.2	Government of Yukon, GeoYukon (NG911 Hydrology – Line and – Polygon – Whitehorse; Health care facilities; Municipal Boundaries). Contains information licensed under the Open Government Licence – Yukon.	book figure, drawn from the credited data
Figure 7.4	Government of Yukon, GeoYukon (Bedrock Geology, Yukon Geological Survey; NG911 Hydrology – Polygon – Whitehorse; Municipal Boundaries). Contains information licensed under the Open Government Licence – Yukon. Colours: ColorBrewer Set2, Cynthia Brewer (Apache 2.0)	book figure, drawn from the credited data
Figure 7.6	Government of Yukon, GeoYukon (Bedrock Geology, Yukon Geological Survey; Municipal Boundaries). Contains information licensed under the Open Government Licence – Yukon. Colours: ColorBrewer Set2, Cynthia Brewer (Apache 2.0)	book figure, drawn from the credited data
Figure 7.8	Government of Yukon, GeoYukon (Municipal Boundaries; NG911 Roads, Hydrology – Line and – Polygon – Whitehorse; Land Parcels – Surveyed). Contains information licensed under the Open Government Licence – Yukon.	book figure, drawn from the credited data
Figure 7.9	Government of Yukon, GeoYukon (Water Wells; Bedrock Geology, Yukon Geological Survey; Municipal Boundaries). Contains information licensed under the Open Government Licence – Yukon. Colours: ColorBrewer Set2, Cynthia Brewer (Apache 2.0)	book figure, drawn from the credited data
Figure 7.10	Government of Yukon, GeoYukon (Water Wells; NG911 Subdivisions – Whitehorse; Municipal Boundaries). Contains information licensed under the Open Government Licence – Yukon. Colours: ColorBrewer YlOrBr, Cynthia Brewer (Apache 2.0)	book figure, drawn from the credited data

Chapter 8 · Raster analysis

Figure	Credit	Licence
Chapter 8 opener	Elevation: Medium Resolution Digital Elevation Model (MRDEM), Natural Resources Canada; contains Copernicus WorldDEM-30 © DLR e.V. 2010-2014 and © Airbus Defence and Space GmbH 2014-2018, provided under COPERNICUS by the European Union and ESA. Water: North America Land Cover 2020 30m, Commission for Environmental Cooperation, NALCMS (CC BY 4.0). City limits: Government of Yukon, GeoYukon. Contains information licensed under the Open Government Licence – Canada and the Open Government Licence – Yukon.	book figure, drawn from the credited data
Figure 8.2	Elevation: MRDEM, Natural Resources Canada (contains Copernicus WorldDEM-30 © DLR e.V. 2010-2014 and © Airbus Defence and Space GmbH 2014-2018, provided under COPERNICUS by the European Union and ESA). Contains information licensed under the Open Government Licence – Canada.	book figure, drawn from the credited data
Figure 8.3	Landsat 9 image (5 August 2024, scene LC09_L2SP_059017_20240805) courtesy of the U.S. Geological Survey, via Microsoft Planetary Computer.	book figure, drawn from the credited data
Figure 8.4	Land cover: North America Land Cover 2020 30m, Commission for Environmental Cooperation, NALCMS (CC BY 4.0), in the CEC's own class colours. Elevation: MRDEM, Natural Resources Canada (contains Copernicus WorldDEM-30 © DLR e.V. 2010-2014 and © Airbus Defence and Space GmbH 2014-2018, provided under COPERNICUS by the European Union and ESA). City limits: Government of Yukon, GeoYukon. Contains information licensed under the Open Government Licence – Canada and the Open Government Licence – Yukon. Colours: ColorBrewer YlOrBr, Cynthia Brewer (Apache 2.0)	book figure, drawn from the credited data

Figure	Credit	Licence
Figure 8.5	Colours: ColorBrewer YlOrBr, Cynthia Brewer (Apache 2.0)	book figure, drawn from the credited data
Figure 8.6	Elevation: MRDEM, Natural Resources Canada (contains Copernicus WorldDEM-30 © DLR e.V. 2010-2014 and © Airbus Defence and Space GmbH 2014-2018, provided under COPERNICUS by the European Union and ESA). Contains information licensed under the Open Government Licence – Canada. Colours: ColorBrewer YlOrRd, Cynthia Brewer (Apache 2.0)	book figure, drawn from the credited data
Figure 8.7	Colours: ColorBrewer YlOrBr, Cynthia Brewer (Apache 2.0)	book figure, drawn from the credited data
Figure 8.9	Slope derived from MRDEM, Natural Resources Canada (contains Copernicus WorldDEM-30 © DLR e.V. 2010-2014 and © Airbus Defence and Space GmbH 2014-2018, provided under COPERNICUS by the European Union and ESA). City limits: Government of Yukon, GeoYukon. Contains information licensed under the Open Government Licence – Canada and the Open Government Licence – Yukon. Colours: ColorBrewer YlOrRd, Cynthia Brewer (Apache 2.0)	book figure, drawn from the credited data
Figure 8.10	Elevation and slope: MRDEM, Natural Resources Canada (contains Copernicus WorldDEM-30 © DLR e.V. 2010-2014 and © Airbus Defence and Space GmbH 2014-2018, provided under COPERNICUS by the European Union and ESA). Land cover: North America Land Cover 2020 30m, Commission for Environmental Cooperation, NALCMS (CC BY 4.0). City limits: Government of Yukon, GeoYukon. Contains information licensed under the Open Government Licence – Canada and the Open Government Licence – Yukon.	book figure, drawn from the credited data

Chapter 9 · Spatial patterns

Figure	Credit	Licence
Chapter 9 opener	Coastline: USGS; Hawaii Statewide GIS Program. Rain gauges: Rainfall Atlas of Hawai'i, University of Hawai'i at Mānoa, Department of Geography. Isohyets: Giambelluca et al. (2013), Online Rainfall Atlas of Hawai'i, via the Hawaii Statewide GIS Program (public domain). Colours: ColorBrewer YlGnBu, Cynthia Brewer (Apache 2.0)	book figure, drawn from the credited data
Figure 9.2	Colours: ColorBrewer YlGnBu, Cynthia Brewer (Apache 2.0)	book figure, drawn from the credited data
Figure 9.3	Coastline: USGS; Hawaii Statewide GIS Program. Rain gauges: Rainfall Atlas of Hawai'i, University of Hawai'i at Mānoa, Department of Geography, via the Hawaii Statewide GIS Program (public domain). Colours: ColorBrewer YlGnBu, Cynthia Brewer (Apache 2.0)	book figure, drawn from the credited data
Figure 9.4	Rain gauges: Rainfall Atlas of Hawai'i, University of Hawai'i at Mānoa, Department of Geography (public domain)	book figure, drawn from the credited data
Figure 9.6	Coastline: USGS; Hawaii Statewide GIS Program. Rain gauges: Rainfall Atlas of Hawai'i, University of Hawai'i at Mānoa, Department of Geography, via the Hawaii Statewide GIS Program (public domain). Colours: ColorBrewer YlGnBu and RdBu, Cynthia Brewer (Apache 2.0)	book figure, drawn from the credited data
Figure 9.7	Coastline: USGS; Hawaii Statewide GIS Program. Rain gauges: Rainfall Atlas of Hawai'i, University of Hawai'i at Mānoa, Department of Geography, via the Hawaii Statewide GIS Program (public domain). Colours: ColorBrewer YlGnBu, Cynthia Brewer (Apache 2.0)	book figure, drawn from the credited data

Figure	Credit	Licence
Figure 9.9	Coastline: USGS; Hawaii Statewide GIS Program. Rain gauges: Rainfall Atlas of Hawai'i, University of Hawai'i at Mānoa, Department of Geography. Isohyets: Giambelluca et al. (2013), Online Rainfall Atlas of Hawai'i, via the Hawaii Statewide GIS Program (public domain). Colours: ColorBrewer YlGnBu, Cynthia Brewer (Apache 2.0)	book figure, drawn from the credited data
Figure 9.11	Coastline: USGS; Hawaii Statewide GIS Program. Cesspools: Hawaii Cesspool Prioritization Tool (2022), UH Water Research Center, via the Hawaii Statewide GIS Program (public domain). Colours: ColorBrewer YlOrRd, Cynthia Brewer (Apache 2.0)	book figure, drawn from the credited data

Chapter 10 · Creating and editing data

Figure	Credit	Licence
Chapter 10 opener	Relief: USGS 10 m DEM, served by the Hawaii Statewide GIS Program (public domain); hillshade and 10 m contours computed for this book	book figure, drawn from the credited data
Figure 10.3	Relief: USGS 10 m DEM, served by the Hawaii Statewide GIS Program (public domain); hillshade computed for this book	book figure, drawn from the credited data

Chapter 11 · 3D scenes and your final map

Figure	Credit	Licence
Chapter 11 opener	Elevation: USGS 10 m DEM, State of Hawaii copy (Hawaii Statewide GIS Program; public domain). Building footprints: USA Structures, Oak Ridge National Laboratory (ORNL) and FEMA Geospatial Response Office (CC BY 4.0), extruded 10 m for this figure	book figure, drawn from the credited data
Figure 11.1	Land: Natural Earth (public domain)	book figure, drawn from the credited data
Figure 11.2	Elevation: USGS 10 m DEM, State of Hawaii copy (Hawaii Statewide GIS Program; public domain)	book figure, drawn from the credited data
Figure 11.5	Elevation: USGS 10 m DEM, State of Hawaii copy (Hawaii Statewide GIS Program; public domain). Building footprints: USA Structures, Oak Ridge National Laboratory (ORNL) and FEMA Geospatial Response Office (CC BY 4.0)	book figure, drawn from the credited data
Figure 11.8	Land: Natural Earth (public domain)	book figure, drawn from the credited data

Course data sources

The course uses data from several parts of the world: Ireland, a state of the United States, Rwanda, the Yukon and Maui. The data for each place are published by the agency or organisation that made them. This page lists every layer the lessons ask you to add, and the base data behind the book's own maps: what each layer is, who publishes it, what the course does with it and the licence as the publisher states it. The licences were read on the publishers' pages in September 2026; publishers can change them, so the licence on a layer's own item page takes precedence. The *Where to find it* column gives each layer's ArcGIS Online item id, or its service or download address, with the data version or period where the course's registry records one; the ids and addresses were copied on 16 October 2026 from the course's data registry (and, for the rows the book alone uses, from the book's source ledger), and where the book's own figures use a substitute dataset, the Figure credits page names it.

A few lesson layers have licences that prevent their use in the book's pictures. The Big River Watch surveys in Modules 1 and 2 are shared under a ShareAlike licence, and the Vivid 2022 imagery in Modules 10 and 11 may not be used for profit or fee. The lessons use both layers in your own project, but the book does not print them. Where a lesson layer could not be used in a picture, the book's maps use public or openly licensed data instead, credited under each figure and on the Figure credits page.

Layers from Esri's ArcGIS Living Atlas of the World and Esri's other reference layers are used under Esri's terms of use (the Esri Master License Agreement), the same terms under which ArcGIS Pro accesses them. Two Module 8 services need you to be signed in to ArcGIS; every other layer here is public.

Ireland — Modules 1–3

The first three modules use Irish data: river basins, the rivers inside them and the citizen scientists who survey them, then land cover mapped from satellite imagery. Module 3 adds world reference layers for its maps of the whole Earth. The rivers and the surveys are added to your map from two small copies that the course hosts publicly on ArcGIS Online, so that they draw quickly and do not change from week to week; each is listed below beside the original it was made from, and keeps the original's licence.

Layer	Publisher	Where to find it	What the course uses it for	Licence
Hydrometric Areas Ireland & Northern Ireland (the river basins)	Environmental Protection Agency (Ireland), with the Northern Ireland hydrological agencies; published on ArcGIS Online by Esri Ireland	item f82f9b52b88b4b4c9dea75c86d732eb3	Module 1, lesson 3: the first layer added from the Living Atlas, and the first attribute table. The basins in the Irish maps of Chapters 1–3	Creative Commons CC BY 4.0 ("for public use")

Layer	Publisher	Where to find it	What the course uses it for	Licence
Rivers Network of Ireland	Environmental Protection Agency (Ireland); in the ArcGIS Living Atlas through Esri Ireland	item aaa9a7087b6845e9abe54d266117d013	The original of the course's rivers copy. Module 2, lesson 4: the Living Atlas item whose owner and metadata you read. The rivers in Chapters 1–3	Creative Commons CC BY 4.0 ("for public use")
Rivers of Ireland (Waldo Learning)	The course, derived from the Rivers Network of Ireland: the 22,856 segments of stream order 3 and up (of 102,108), simplified to 10 m. Credit: Environmental Protection Agency, via Esri Ireland	Course-hosted copy of the Rivers Network of Ireland, made 2026-10-15; item 827dd08cdd544695a1efc5e741c85672	Module 1, lesson 4: rivers added by URL; Module 2: the line layer whose vertices you zoom in to see	Creative Commons CC BY 4.0, as the original
The Big River Watch (river surveys)	The Rivers Trust, from surveys by citizen scientists	item cb49c4d15ecc40a5b290c3d40dae5970; a live layer that keeps growing	The original of the course's survey copy: a live layer of about 20,000 surveys from Ireland and Britain that grows every week. Not in the book's figures	Creative Commons Attribution-ShareAlike 4.0 (CC BY-SA 4.0)
Big River Watch surveys – Ireland (Waldo Learning)	The course, derived from The Big River Watch: the 1,817 surveys inside the island's river basins, Ireland and Northern Ireland, without photo attachments. Credit: The Rivers Trust	Course-hosted copy of The Big River Watch, a snapshot of 2026-10-15; item aeefb39126c64fd982f19a5f44255b85	Module 1, lesson 4, and Module 2: points to add, export and question about who made them. Not in the book's figures	Creative Commons Attribution-ShareAlike 4.0 (CC BY-SA 4.0), as the original
Sentinel-2 10m Land Use/Land Cover Time Series	Esri and Impact Observatory, with Microsoft; ArcGIS Living Atlas	item cfcb7609de5f478eb7666240902d4d3d; a time series, 2017–2025	Module 2, lesson 2: the first raster, zoomed in until its cells show; lesson 4: a full Metadata tab. The land cover in Chapters 1 and 2	Esri Master License Agreement; the source land-cover data CC BY 4.0
World Countries (Generalized)	Esri (sources: Garmin International; U.S. Central Intelligence Agency, The World Factbook; National Geographic Society); ArcGIS Living Atlas	item 2b93b06dc0dc4e809d3c8db5cb96ba69	Module 3, lesson 3: the countries in the projection-gallery pictures. The world maps in Chapters 0 and 3	Esri Master License Agreement
Firefly Grid Lines v2 (a graticule)	Esri (John Nelson); ArcGIS Living Atlas	item 7b19a92d0c984a4da2e06889c2b4ea89	Module 3, lesson 2: lines of latitude and longitude drawn on the map	Esri Master License Agreement

Layer	Publisher	Where to find it	What the course uses it for	Licence
World UTM Grid	Esri (sources: U.S. Geological Survey; U.S. National Geospatial-Intelligence Agency); ArcGIS Online	item b294795270aa4fb3bd25286bf09edc51	Module 3, lesson 4: finding the UTM zone Galway falls in. The zone map in Chapter 3	Esri Master License Agreement
USA State Plane Zones NAD 1983	Esri (source: NOAA); ArcGIS Living Atlas	item 23178a639bdcd4d658816b3ea8ee6c3ae	Module 3, lesson 4: an optional look at the zones for learners in the United States	Esri Master License Agreement

A US state, Virginia as the worked example — Module 4

Module 4 makes a finished map of one state, the learner's choice; the lessons and the book use Virginia as the worked example. Each layer is national and is filtered to the state with a definition query.

Layer	Publisher	Where to find it	What the course uses it for	Licence
USA Counties Generalized Boundaries	Esri, from U.S. Census Bureau data; ArcGIS Living Atlas	item 3c164274a80748dda926a046525da610; 2020 census population	Module 4: the county choropleth, its classification and labels. The Virginia maps in Chapter 4	Esri Master License Agreement
USA Major Cities	Esri, from U.S. Census Bureau data; ArcGIS Living Atlas	item 9df5e769bfe8412b8de36a2e618c7672	Module 4: point symbols and city labels. The reference and design maps in Chapter 4	Esri Master License Agreement
North American Rail Network Lines	Bureau of Transportation Statistics and Federal Railroad Administration; ArcGIS Living Atlas	item feb7060cb7ba49ba94e53651fec963b9; National Transportation Atlas Database (NTAD)	Module 4: a line layer with a distinctive railroad symbol. The railroads in Chapter 4	A U.S. government work, "available for unrestricted public use" (public domain)
Cartographic boundary files, 2023 (state outlines at three scales)	U.S. Census Bureau	2023 files cb_2023_us_state_500k, _5m and _20m, from www2.census.gov , GENZ 2023	The book only: the generalisation figure in Chapter 4	Public domain

Rwanda — Modules 5–6

Modules 5 and 6 work with Rwanda's 416 sectors and its census, both published by the country's statistics office. Module 6 also uses data from London in 1854, where John Snow mapped a cholera outbreak.

Layer	Publisher	Where to find it	What the course uses it for	Licence
Sector Boundary 2022	National Institute of Statistics of Rwanda (NISR)	item ded3c363069f4dd2b11275e2510f4ddd; 2022 boundaries	Modules 5 and 6: the sectors whose table you explore, calculate and join. The maps and tables in Chapters 5 and 6	"Public access and no constraints"
population_count (census table)	National Institute of Statistics of Rwanda (NISR)	item 445bf6f37cb648318a52088dc620ba91; the 2002, 2012 and 2022 censuses	Module 5: a standalone table of the 2002, 2012 and 2022 censuses, joined to the sectors. Chapters 5 and 6	No licence text on the item itself; published on NISR's open-data hub beside the sector boundaries
Health_Establishments_hospital	National Institute of Statistics of Rwanda (NISR)	item 4840f443228b4052bca41513911f7f3f; from the 2022 census mapping	Module 6, lesson 3: selecting sectors by their distance from a hospital. Chapter 6	"Public access and no constraints"
Cholera Cases (London, 1854)	Esri Tutorials, digitised from John Snow's map	item f39a169103da4149ae577ccceec9d125; the deaths of 1854	Module 6, lesson 1: the deaths at each address, looked at and summed. Chapter 6	"Digitized from a map that is in the public domain"
Water Pumps (London, 1854)	Esri Tutorials, digitised from John Snow's map	item edd427fecaa242a689b74d7e470fa821	Module 6, lesson 1: the pumps, the Broad Street pump among them. Chapter 6	"Digitized from a map that is in the public domain"
Snow cholera map image	Esri Tutorials: a georeferenced scan of John Snow's 1855 map (lithograph by C. F. Cheffins)	item 0f74f92ada58401bba2c593bd4ca593f; the map of 1855	Module 6, lesson 1: the original map under the two point layers. The opening map of Chapter 6	"This image is in the public domain"

Whitehorse, Yukon — Modules 7–8

Module 7's vector analysis uses the Government of Yukon's GeoYukon services, all in one projected coordinate system measured in metres. In the table, a GeoYukon layer is given by its map service and layer number under <https://mapservices.gov.yk.ca/arcgis/rest/services/GeoYukon/> (Municipal Boundaries, for example, is GY_LandPlanning/MapServer/6). Module 8 uses the same place and adds three public rasters. In the book's elevation and imagery figures, openly licensed national data are used in place of the two Module 8 services that need a sign-in.

Layer	Publisher	Where to find it	What the course uses it for	Licence
Municipal Boundaries	Government of Yukon (GeoYukon)	GeoYukon GY_LandPlanning map service, layer 6	Module 7, lesson 1: the Whitehorse city limits, the study area of both modules. Chapters 7 and 8	Open Government Licence – Yukon
NG911 Hydrology – Line – Whitehorse (streams)	Government of Yukon (GeoYukon)	GeoYukon GY_EmergencyManagement map service, layer 36	Module 7, lesson 1: stream setbacks by buffer. Chapter 7	Open Government Licence – Yukon
Health care facilities	Government of Yukon (GeoYukon)	GeoYukon GY_PeopleStructures map service, layer 0	Module 7, lesson 1: rings around the hospital. Chapter 7	Open Government Licence – Yukon
Bedrock Geology	Government of Yukon, Yukon Geological Survey (GeoYukon)	GeoYukon GY_Geological map service, layer 113	Module 7, lessons 2, 3 and 5: clip, overlay, dissolve and spatial join. Chapter 7	Open Government Licence – Yukon
Water Wells	Government of Yukon (GeoYukon)	GeoYukon GY_EnvironmentalMonitoring map service, layer 17	Module 7, lessons 2 and 5: clipped to the city, joined to geology and counted by subdivision; Module 8 reads elevations at the wells. Chapter 7	Open Government Licence – Yukon
NG911 Hydrology – Polygon – Whitehorse (lakes, wetlands, rivers)	Government of Yukon (GeoYukon)	GeoYukon GY_EmergencyManagement map service, layer 35	Module 7, lessons 2–4: erase, union and the campground search. Chapter 7	Open Government Licence – Yukon
NG911 Subdivisions – Whitehorse	Government of Yukon (GeoYukon)	GeoYukon GY_EmergencyManagement map service, layer 27	Module 7, lessons 3 and 5: intersect, and counting wells per subdivision. Chapter 7	Open Government Licence – Yukon
NG911 Roads – Whitehorse	Government of Yukon (GeoYukon)	GeoYukon GY_EmergencyManagement map service, layer 29	Module 7, lesson 4: the campground search. Chapter 7	Open Government Licence – Yukon
Land Parcels – Surveyed	Government of Yukon (GeoYukon)	GeoYukon GY_LandTenure map service, layer 27	Module 7, lesson 4: land already taken, erased from the campground search. Chapter 7	Open Government Licence – Yukon

Layer	Publisher	Where to find it	What the course uses it for	Licence
Terrain (elevation)	Esri; ArcGIS Living Atlas	item 58a541efc59545e6b7137f961d7de883	Module 8: the elevation raster the lessons make for Whitehorse. Needs an ArcGIS sign-in; not in the book's figures	Esri Master License Agreement
North America Land Cover 2020 30m	Commission for Environmental Cooperation, North American Land Change Monitoring System (CCRS, USGS, CONABIO, CONAFOR, INEGI)	item e0ff7c071cb740409c7fe01f63f4fef8; land cover of 2020, edition 2.0	Module 8: land-cover classes, reclassification and the habitat model. Chapter 8	Creative Commons CC BY 4.0
Multispectral Landsat	Esri, from U.S. Geological Survey Landsat imagery; ArcGIS Living Atlas	item d9b466d6a9e647ce8d1dd5fe12eb434b	Module 8: band combinations. Needs an ArcGIS sign-in; not in the book's figures	Esri Master License Agreement
Medium Resolution Digital Elevation Model (MRDEM)	Natural Resources Canada	MRDEM-30; open.canada.ca data set 18752265-bda3-498c-a4ba-9dfe68cb98da	The book only: every elevation figure in Chapter 8, in place of Terrain	Open Government Licence – Canada; contains Copernicus WorldDEM-30 (© DLR e.V. and © Airbus Defence and Space GmbH, provided under COPERNICUS by the European Union and ESA)
Landsat 9 scene, 5 August 2024 (Collection 2, Level-2)	U.S. Geological Survey, read from Microsoft Planetary Computer's copy	Scene LC09_L2SP_059017_20240805, Planetary Computer collection landsat-c2-l2; product doi.org/10.5066/P9OGBGM6	The book only: the band-combination figure in Chapter 8, in place of Multispectral Landsat	Public domain

Maui, Hawai'i — Modules 9–11

The last three modules use data from Maui. Module 9's layers come from the Hawaii Statewide GIS Program, whose State of Hawaii Terms of Use place the contents in the public domain, with no warranty. Where the table gives a Hawaii service rather than an item, the path is under <https://geodata.hawaii.gov/arcgis/rest/services/> (the coastline, for example, is `Terrestrial1/MapServer/3`). Module 10 traces the Haleakalā summit on aerial imagery, and Module 11 shows the same summit as a 3D scene. Module 11's final map uses layers from any of the places above, or your own.

Layer	Publisher	Where to find it	What the course uses it for	Licence
Coastline	Hawaii Statewide GIS Program, from U.S. Geological Survey data	Hawaii Terrestrial map service, layer 3	Module 9, lesson 1: Maui as the study area. The maps of Chapter 9	State of Hawaii Terms of Use: public domain
Rain Gauges	Hawaii Statewide GIS Program; data from the Rainfall Atlas of Hawai'i, University of Hawai'i at Mānoa, Department of Geography	item ed0eca4ac4554febb86e0df2d30c59df; the 2011 Rainfall Atlas of Hawai'i	Module 9, lessons 1–3: Tobler's law, Thiessen polygons and IDW interpolation. Chapter 9	State of Hawaii Terms of Use: public domain
Annual Rainfall (millimeters) (isohyets)	Hawaii Statewide GIS Program; from the Online Rainfall Atlas of Hawai'i (Giambelluca et al., 2013)	Hawaii Climate map service, layer 26	Module 9, lesson 3: a reference map to compare the interpolated surface with. Chapter 9	State of Hawaii Terms of Use: public domain
HCPT Cesspools	Hawaii Statewide GIS Program; Hawaii Cesspool Prioritization Tool (2022), UH Water Research Center	item 7a57005f120b4027977d9cdbc9451ea7; the tool's 2022 data	Module 9, lesson 4: kernel density. Chapter 9	State of Hawaii Terms of Use: public domain
Vivid 2022 (aerial imagery)	Hawaii Statewide GIS Program; USDA Farm Production and Conservation Business Center; imagery by Maxar, licensed to the USDA	Hawaii image service SoH_Imagery/Vivid_2022; 2022 imagery, 0.6 m cells	Module 10: tracing roads and roofs at the Haleakalā summit; Module 11, lessons 1 and 2: the picture draped on the scene. Viewed live only; not in the book's figures	The metadata prohibits use of the product for commercial purposes, "for profit or fee"
USGS 10 m DEM (State of Hawaii copy)	U.S. Geological Survey, served by the Hawaii Statewide GIS Program	Hawaii image service SoH_Imagery/USGS_10m_DEM_Elevation	Module 11, lesson 1: the elevation check at the summit. The relief in Chapter 10 and the ground in Chapter 11's figures	Public domain
USA Structures (building footprints)	Oak Ridge National Laboratory and FEMA Geospatial Response Office; on ArcGIS Online through Esri US Federal Data	item 0ec8512ad21e4bb987d7e848d14e7e24; the summit footprints traced from 2017 imagery	Module 11, lessons 1 and 2: the summit buildings, extruded in the scene. Chapter 11	Creative Commons CC BY 4.0

World base data for the book's figures

Two world datasets are used only in the book's own figures.

Layer	Publisher	Where to find it	What the book uses it for	Licence
Natural Earth (land, coastlines, towns)	Natural Earth	naturalearthdata.com, Downloads page	The globes, world maps and town points in Chapters 2, 3 and 11	Public domain

Layer	Publisher	Where to find it	What the book uses it for	Licence
EGM96 geoid heights	U.S. National Geospatial-Intelligence Agency	The 15' grid <code>us_nga_egm96_15.tif</code> from PR OJ-data	The geoid map in Chapter 3	Public domain

Further reading

This page gathers the reading lists from the end of each chapter in one place, in chapter order. Within each group the openly licensed and free-to-read sources come first; the few commercial books come last.

Start here (Chapter 0)

For installing and licensing ArcGIS Pro and starting a first project.

- Esri. *ArcGIS Pro 3.7 system requirements*. ArcGIS Pro documentation. <https://doc.esri.com/en/arcgis-pro/latest/get-started/arcgis-pro-system-requirements.html>
- Esri. *Download ArcGIS Pro* and *Install ArcGIS Pro*. ArcGIS Pro documentation. <https://doc.esri.com/en/arcgis-pro/latest/get-started/download-arcgis-pro.html> and <https://doc.esri.com/en/arcgis-pro/latest/get-started/install-and-sign-in-to-arcgis-pro.html>
- Esri. *Create a project*. ArcGIS Pro documentation (the start page, templates and Start without a template). <https://doc.esri.com/en/arcgis-pro/latest/help/projects/create-a-new-project.html>
- Esri. *Run ArcGIS Pro on a Mac*. ArcGIS Pro documentation. <https://doc.esri.com/en/arcgis-pro/latest/get-started/run-pro-on-a-mac.html>
- Esri. ArcGIS Pro free trial, ArcGIS for Student Use and ArcGIS for Personal Use. <https://www.esri.com/en-us/arcgis/products/arcgis-pro/trial>, <https://www.esri.com/en-us/arcgis/products/arcgis-for-student-use/overview> and <https://www.esri.com/en-us/arcgis/products/arcgis-for-personal-use/overview>

The Pro workspace and where data lives (Chapters 1–2)

For layers, the vector and raster models, file formats and the places data comes from.

- Olaya, V. *Introduction to GIS*, the chapters on data and on data sources. <https://volaya.github.io/gis-book/en/>
- Butler, H., et al. (2016). *The GeoJSON Format*. RFC 7946, Internet Engineering Task Force. <https://www.rfc-editor.org/rfc/rfc7946>
- Open Geospatial Consortium. *KML standard*. <https://www.ogc.org/standards/kml/>
- Esri. *ArcGIS Pro user interface*. ArcGIS Pro documentation. <https://doc.esri.com/en/arcgis-pro/latest/get-started/user-interface.html>
- Esri. *Bookmarks*. ArcGIS Pro documentation. <https://doc.esri.com/en/arcgis-pro/latest/help/mapping/navigation/bookmarks.html>
- Esri. *ArcGIS Pro documentation*, on geodatabases, shapefiles and data sources. <https://pro.arcgis.com/>
- Harder, C., & Brown, C. (eds.) (2017). *The ArcGIS Book* (2nd ed.), on the Living Atlas and web layers. Esri Press. <https://learn.arcgis.com/en/arcgis-book/>

Spatial reference (Chapter 3)

For the shape of the Earth, datums, coordinate systems and map projections.

- Snyder, J. P. (1987). *Map Projections: A Working Manual*. U.S. Geological Survey Professional Paper 1395. <https://doi.org/10.3133/pp1395>
- Snyder, J. P., & Voxland, P. M. (1989). *An Album of Map Projections*. U.S. Geological Survey Professional Paper 1453. <https://pubs.usgs.gov/publication/pp1453>
- U.S. Geological Survey. *Map Projections* (poster). <https://pubs.usgs.gov/publication/70047422>
- NOAA National Ocean Service. *Global Positioning Tutorial: Geodesy* (the geoid and datums). https://oceanservice.noaa.gov/education/tutorial_geodesy/
- Olaya, V. *Introduction to GIS*, the chapter on cartography and geodesy. <https://volaya.github.io/gis-book/en/>
- Cory, M. J., Morgan, A., Bray, C., & Greenway, I. (2001). "A New Coordinate System for Ireland." FIG Working Week, Seoul.
- Esri ArcGIS Blog. "Geographic vs. projected coordinate systems." <https://www.esri.com/arcgis-blog/products/arcgis-pro/mapping/coordinate-systems-difference>
- Esri ArcGIS Blog. "Projection on the fly and geographic transformations." <https://www.esri.com/arcgis-blog/products/arcgis-pro/mapping/projection-on-the-fly-and-geographic-transformations>
- The True Size Of..., an interactive Mercator map for dragging countries north and south. <https://thetruesize.com/>
- Šavrič, B., Patterson, T., & Jenny, B. (2019). "The Equal Earth map projection." *International Journal of Geographical Information Science* 33(3). <https://doi.org/10.1080/13658816.2018.1504949>

Cartographic design (Chapter 4)

For symbols, colour, classification, labels, layouts and the principles of map design.

- Brewer, C. A., Harrower, M., & The Pennsylvania State University. *ColorBrewer 2.0*: colour advice for maps. <https://colorbrewer2.org/>
- Ingram, U., & Rhodes, J. *Introduction to Cartography (KSU)*. GALILEO Open Learning Materials. <https://oer.galileo.usg.edu/geo-textbooks/3>
- Schlosser, S. (2024). *Making Effective Maps: Cartographic Visualization for GIS*. University of Colorado Boulder. <https://colorado.pressbooks.pub/makingmaps/>
- Axis Maps. *Cartography Guide*. <https://www.axismaps.com/guide>
- DiBiase, D., et al. *The Nature of Geographic Information*. Penn State. <https://www.e-education.psu.edu/natureofgeoinfo/>
- Esri. *ArcGIS Pro documentation*: the data classification methods page. <https://pro.arcgis.com/>
- Slocum, T. A., McMaster, R. B., Kessler, F. C., & Howard, H. H. *Thematic Cartography and Geovisualization*. Pearson / CRC Press.

Attribute data (Chapter 5)

For tables, field types, keys and joins.

- Olaya, V. *Introduction to GIS*, the chapters on data and on databases (records, keys and relations). <https://volaya.github.io/gis-book/en/>
- National Institute of Statistics of Rwanda, geodata hub: the sector boundaries and the census tables used in Chapter 5. <https://geodata-nisr-rwanda.opendata.arcgis.com/>
- Esri. "ArcGIS field data types." ArcGIS Pro documentation. <https://pro.arcgis.com/en/pro-app/latest/help/data/geodatabases/overview/arcgis-field-data-types.htm>
- Esri. "Joins and relates" and the Add Join tool reference. ArcGIS Pro documentation. <https://pro.arcgis.com/en/pro-app/latest/help/data/tables/joins-and-relates.htm>

Exploring spatial data (Chapter 6)

For asking questions of data: attribute and spatial queries, summaries and charts, and the classic cholera map.

- Snow, J. (1855). *On the Mode of Communication of Cholera* (2nd ed.). London: John Churchill. Public domain; Wellcome Collection <https://wellcomecollection.org/works/uqa27qrt> and Project Gutenberg <https://www.gutenberg.org/ebooks/72894>
- Olaya, V. *Introduction to GIS*, the chapters on databases (queries and logical operators) and analysis (spatial queries). <https://volaya.github.io/gis-book/en/>
- Esri. ArcGIS Pro documentation: *Select features by attributes* and *Select features by location*, with the full list of spatial relationships. <https://pro.arcgis.com/en/pro-app/latest/help/mapping/navigation/select-features-by-location.htm>

Vector analysis (Chapter 7)

For buffers, overlay, dissolve, spatial joins and chaining tools into a site search.

- Olaya, V. *Introduction to GIS*, the chapter on spatial analysis. <https://volaya.github.io/gis-book/en/>
- Fang, Y., Shandas, V., & Arriaga Cordero, E. (2014). *Spatial Thinking in Planning Practice: An Introduction to GIS*. Portland State University Library. <https://pdxscholar.library.pdx.edu/pdxopen/4/>
- Gimond, M. *Intro to GIS and Spatial Analysis*. <https://mgimond.github.io/Spatial/>

Raster analysis (Chapter 8)

For the raster model, terrain, map algebra and the public rasters behind the chapter.

- Olaya, V. *Introduction to GIS*, the chapters on the raster model and map algebra. <https://volaya.github.io/gis-book/en/>
- U.S. Geological Survey. Landsat 8–9 Collection 2 Level-2 surface reflectance. <https://doi.org/10.5066/P9OGBGM6>

- Commission for Environmental Cooperation. *North American Environmental Atlas – Land Cover 2020 30m*. <https://www.arcgis.com/home/item.html?id=eoff7c071cb740409c7fe01f63f4fef8>
- Natural Resources Canada. *Medium Resolution Digital Elevation Model (MRDEM) – CanElevation Series*. <https://open.canada.ca/data/en/dataset/18752265-bda3-498c-a4ba-9dfe68cb98da>

Spatial patterns (Chapter 9)

For Tobler's first law, Thiessen polygons, interpolation and kernel density.

- Olaya, V. *Introduction to GIS*, the chapters on creating raster layers and interpolation. <https://volaya.github.io/gis-book/en/>
- Giambelluca, T. W., et al. (2013). "Online Rainfall Atlas of Hawai'i." *Bulletin of the American Meteorological Society* 94, 313–316; the atlas itself at <https://rainfall.geography.hawaii.edu/>
- Esri. "How IDW works" and "How Kernel Density works." ArcGIS Pro documentation. <https://pro.arcgis.com/en/pro-app/latest/tool-reference/spatial-analyst/how-idw-works.htm>
- Tobler, W. R. (1970). "A computer movie simulating urban growth in the Detroit region." *Economic Geography* 46, 234–240. <https://doi.org/10.2307/143141>

Editing data and 3D scenes (Chapters 10–11)

For digitising and editing features, and for turning a map into a 3D scene.

- Olaya, V. *Introduction to GIS*, the chapter on data sources (digitization and its errors). https://volaya.github.io/gis-book/en/Data_sources.html
- U.S. Geological Survey. The 10 m digital elevation models behind Chapter 10's relief (public domain), served for Hawaii by the Hawaii Statewide GIS Program.
- Esri. ArcGIS Pro documentation: editing, feature templates, snapping and editing options. <https://pro.arcgis.com/en/pro-app/latest/help/editing/>
- Esri. "Extrude features to 3D symbology." ArcGIS Pro documentation. <https://doc.esri.com/en/arcgis-pro/latest/help/mapping/layer-properties/extrude-features-to-3d-symbology.html>
- Esri. "Author and share a local scene," a Get Started tutorial. ArcGIS Pro documentation. <https://pro.arcgis.com/en/pro-app/latest/get-started/author-and-share-a-local-scene.htm>
- Harder, C., & Brown, C. (eds.) (2017). *The ArcGIS Book*, the chapter on 3D. Esri Press. <https://learn.arcgis.com/en/arcgis-book/>

The course's data

For every layer the lessons and the book's maps use; the Course data sources page lists them by place.

- The course's data sources: every layer the lessons use, who publishes it and its licence.