

Gentle GIS Introduction

QGIS Project

Nov 24, 2026

CONTENTS

1 Preamble **3**
- 1.1 A word from the editor . . . 4

2 Introducing GIS **5**
- 2.1 Overview . . . 5
- 2.2 More about GIS . . . 6
- 2.3 What is GIS Software / a GIS Application? . . . 6
- 2.4 Getting a GIS Application for your own computer(s) . . . 10
- 2.5 GIS Data . . . 10
- 2.6 What have we learned? . . . 12
- 2.7 Now you try! . . . 12
- 2.8 Something to think about . . . 13
- 2.9 Further reading . . . 13
- 2.10 What's next? . . . 13

3 Vector Data **15**
- 3.1 Overview . . . 15
- 3.2 Point features in detail . . . 18
- 3.3 Polyline features in detail . . . 19
- 3.4 Polygon features in detail . . . 19
- 3.5 Vector data in layers . . . 20
- 3.6 Editing vector data . . . 20
- 3.7 Scale and vector data . . . 20
- 3.8 Symbology . . . 21
- 3.9 What can we do with vector data in a GIS? . . . 23
- 3.10 Common problems with vector data . . . 23
- 3.11 What have we learned? . . . 24
- 3.12 Now you try! . . . 25
- 3.13 Something to think about . . . 26
- 3.14 Further reading . . . 26
- 3.15 What's next? . . . 26

4 Vector Attribute Data **27**
- 4.1 Overview . . . 27
- 4.2 Attributes in detail . . . 30
- 4.3 Single Symbols . . . 31
- 4.4 Graduated Symbols . . . 32
- 4.5 Continuous Colour Symbols . . . 33
- 4.6 Unique Value Symbols . . . 35
- 4.7 Things to be aware of . . . 36
- 4.8 What have we learned? . . . 37
- 4.9 Now you try! . . . 37
- 4.10 Something to think about . . . 38
- 4.11 Further reading . . . 38

4.12 What's next? . . . 38

5 Data Capture **39**
5.1 Overview . . . 39
5.2 How does GIS digital data get stored? . . . 39
5.3 Planning before you begin . . . 40
5.3.1 Example 1: Creating a tourism map . . . 40
5.3.2 Example 2: Creating a map of pollution levels along a river . . . 41
5.4 Creating an empty shapefile . . . 41
5.5 Adding data to your shapefile . . . 43
5.6 Heads-up digitising . . . 45
5.7 Digitising using a digitising table . . . 45
5.8 After your features are digitised… . . . 46
5.9 Common problems / things to be aware of . . . 46
5.10 What have we learned? . . . 47
5.11 Now you try! . . . 47
5.12 Something to think about . . . 47
5.13 Further reading . . . 47
5.14 What's next? . . . 48

6 Raster Data **49**
6.1 Overview . . . 49
6.2 Raster data in detail . . . 50
6.3 Georeferencing . . . 51
6.4 Sources of raster data . . . 51
6.5 Spatial Resolution . . . 52
6.6 Spectral resolution . . . 53
6.7 Raster to vector conversion . . . 53
6.8 Vector to raster conversion . . . 53
6.9 Raster analysis . . . 54
6.10 Common problems / things to be aware of . . . 54
6.11 What have we learned? . . . 54
6.12 Now you try! . . . 54
6.13 Something to think about . . . 55
6.14 Further reading . . . 55
6.15 What's next? . . . 55

7 Topology **57**
7.1 Overview . . . 57
7.2 Topology errors . . . 57
7.3 Topology rules . . . 58
7.4 Topological tools . . . 58
7.5 Snapping distance . . . 59
7.6 Search Radius . . . 59
7.7 Common problems / things to be aware of . . . 59
7.8 What have we learned? . . . 60
7.9 Now you try! . . . 60
7.10 Something to think about . . . 60
7.11 Further reading . . . 60
7.12 What's next? . . . 60

8 Coordinate Reference Systems **61**
8.1 Overview . . . 61
8.2 Map Projection in detail . . . 61
8.3 The three families of map projections . . . 62
8.4 Accuracy of map projections . . . 62
8.4.1 Map projections with angular conformity . . . 63
8.4.2 Map projections with equal distance . . . 64
8.4.3 Projections with equal areas . . . 65

8.5 Coordinate Reference System (CRS) in detail . . . 66
8.5.1 Geographic Coordinate Systems . . . 66
8.5.2 Projected coordinate reference systems . . . 67
8.6 Universal Transverse Mercator (UTM) CRS in detail . . . 67
8.6.1 The northing (Y) value . . . 68
8.6.2 The easting (X) value . . . 69
8.7 On-The-Fly Projection . . . 69
8.8 Common problems / things to be aware of . . . 69
8.9 What have we learned? . . . 69
8.10 Now you try! . . . 70
8.11 Something to think about . . . 70
8.12 Further reading . . . 70
8.13 What's next? . . . 70

9 Map Production **71**
9.1 Overview . . . 71
9.2 Title in detail . . . 72
9.3 Map Border in detail . . . 72
9.4 Map Legend in detail . . . 72
9.5 North arrow in detail . . . 72
9.6 Scale in detail . . . 72
9.7 Acknowledgment in detail . . . 74
9.8 Graticule in detail . . . 74
9.9 Name of the map projection in detail . . . 74
9.10 Common problems / things to be aware of . . . 75
9.11 What have we learned? . . . 75
9.12 Now you try! . . . 75
9.13 Something to think about . . . 76
9.14 Further reading . . . 76
9.15 What's next? . . . 76

10 Vector Spatial Analysis (Buffers) **77**
10.1 Overview . . . 77
10.2 Buffering in detail . . . 77
10.3 Variations in buffering . . . 80
10.3.1 Multiple buffer zones . . . 80
10.3.2 Buffering with intact or dissolved boundaries . . . 81
10.3.3 Buffering outward and inward . . . 81
10.4 Common problems / things to be aware of . . . 82
10.5 More spatial analysis tools . . . 82
10.6 What have we learned? . . . 82
10.7 Now you try! . . . 83
10.8 Something to think about . . . 83
10.9 Further reading . . . 83
10.10 What's next? . . . 84

11 Spatial Analysis (Interpolation) **85**
11.1 Overview . . . 85
11.2 Spatial interpolation in detail . . . 85
11.3 Inverse Distance Weighted (IDW) . . . 86
11.4 Triangulated Irregular Network (TIN) . . . 87
11.5 Common problems / things to be aware of . . . 88
11.6 Other interpolation methods . . . 88
11.7 What have we learned? . . . 88
11.8 Now you try! . . . 89
11.9 Something to think about . . . 89
11.10 Further reading . . . 89
11.11 What's next? . . . 89

12 About the authors & contributors 91

13 GNU Free Documentation License 93

CHAPTER

ONE

PREAMBLE

A Gentle Introduction to GIS

Brought to you with QGIS, a Free and Open Source Software GIS Application for everyone, hosted by OSGeo.



T. Sutton, O. Dassau, M. Sutton

Sponsored by: Chief Directorate: Spatial Planning & Information, Department of Land Affairs, Eastern Cape, South Africa.



In partnership with: Spatial Information Management Unit, Office of the Premier, Eastern Cape, South Africa.



Copyright (c) 2009 Chief Directorate: Spatial Planning & Information, Department of Land Affairs, Eastern Cape.

Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.2 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts.

A copy of the license is included in the section *GNU Free Documentation License*.

The above copyright notice excludes the QGIS User Manual which may be appended to this document. Consult the QGIS Manuals for further copyright and licensing information.

1.1 A word from the editor

This project was sponsored by the Chief Directorate: Spatial Planning & Information, Department of Land Affairs (DLA), Eastern Cape, in conjunction with the Spatial Information Management Unit, Office of the Premier, Eastern Cape, South Africa.

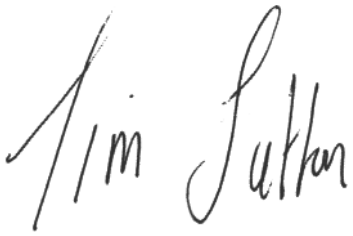
GIS is becoming an increasingly important tool in environmental management, retail, military, police, tourism and many other spheres of our daily lives. If you use a computer or a cell phone, you have probably already used a GIS in some form without even realising it. Maybe it was a map on a web site, Google Earth, an information booth or your cell phone telling you where you are. Proprietary GIS software (software that cannot be freely shared or modified) is available that will let you do everything we describe in these worksheets and a lot more. However this software is usually very expensive or otherwise limits your freedom to copy, share and modify the software. GIS vendors sometimes make an exception for educational activities, providing cheaper or free copies of their software. They do this knowing that if teachers and learners get to know their software, they will be reluctant to learn other packages. When learners leave school they will go into the workplace and buy the commercial software, never knowing that there are free alternatives that they could be using.

With QGIS, we offer an alternative — software that is free of cost and free in a social sense. You can make as many copies as you like. When learners leave school one day they can use this software to build their skills, solve problems at work and make the world a better place.

When you buy commercial software, you limit your options for the future. By learning, using and sharing Free and Open Source Software, you are building your own skills, freeing money to be spent on important things like food and shelter and boosting our own economy.

By sponsoring the creation of this resource, the DLA has created a foundation to which young minds can be exposed. Exciting possibilities lie ahead when principles of free sharing of knowledge and data are embraced. For this we give our heartfelt thanks to the DLA!

We hope you enjoy using and learning QGIS in the spirit of Ubuntu!



Tim Sutton, April 2009

CHAPTER

TWO

INTRODUCING GIS



Objectives:	Understanding what GIS is and what it can be used for.
Keywords:	GIS, Computer, Maps, Data, Information System, Spatial, Analysis

2.1 Overview

Just as we use a word processor to write documents and deal with words on a computer, we can use a **GIS application** to deal with **spatial information** on a computer. GIS stands for **'Geographical Information System'**.

A GIS consists of:

- **Digital Data** — the geographical information that you will view and analyse using computer hardware and software.
- **Computer Hardware** — computers used for storing data, displaying graphics and processing data.
- **Computer Software** — computer programs that run on the computer hardware and allow you to work with digital data. A software program that forms part of the GIS is called a GIS Application.

With a GIS application you can open digital maps on your computer, create new spatial information to add to a map, create printed maps customised to your needs and perform spatial analysis.

Let's look at a little example of how GIS can be useful. Imagine you are a health worker and you make a note of the date and place of residence of every patient you treat.

Longitude	Latitude	Disease	Date
26.870436	-31.909519	Mumps	13/12/2008
26.868682	-31.909259	Mumps	24/12/2008
26.867707	-31.910494	Mumps	22/01/2009
26.854908	-31.920759	Measles	11/01/2009
26.855817	-31.921929	Measles	26/01/2009
26.852764	-31.921929	Measles	10/02/2009
26.854778	-31.925112	Measles	22/02/2009
26.869072	-31.911988	Mumps	02/02/2009
26.863354	-31.916406	Chicken Pox	26/02/2009

If you look at the table above you will quickly see that there were a lot of measles cases in January and February. Our health worker recorded the location of each patient's house by noting its latitude and longitude in the table. Using this data in a GIS Application, we can quickly understand a lot more about the patterns of illness:

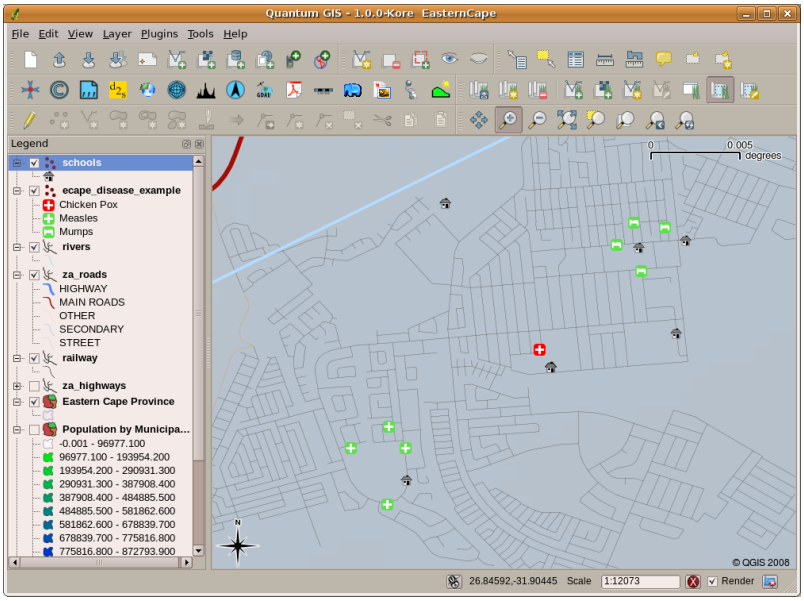


Fig. 2.1: Example showing disease records in a GIS application. It is easy to see that the mumps patients all live close to each other.

2.2 More about GIS

GIS is a relatively new field — it started in the 1970's. It used to be that computerised GIS was only available to companies and universities that had expensive computer equipment. These days, anyone with a personal computer or laptop can use GIS software. Over time GIS Applications have also become easier to use — it used to require a lot of training to use a GIS Application, but now it is much easier to get started in GIS even for amateurs and casual users. As we described above, GIS is more than just software, it refers to all aspects of managing and using digital geographical data. In the tutorials that follow we will be focusing on GIS Software.

2.3 What is GIS Software / a GIS Application?

You can see an example of what a **GIS Application** looks like in Fig. 2.1. GIS Applications are normally programs with a graphical user interface that can be manipulated using the mouse and keyboard. The application provides **menus** near to the top of the window (*Project*, *Edit*, etc.) which, when clicked using the mouse, show a panel of **actions**. These actions provide a way for you to tell the GIS Application what you want to do. For example you may use the menus to tell the GIS Application to add a new layer to the display output, as in Fig. 2.2.

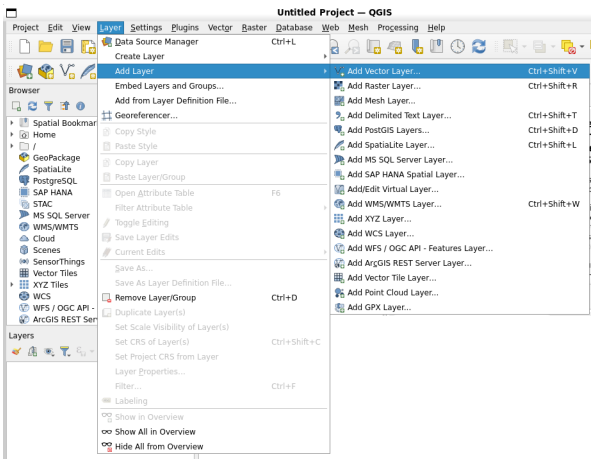


Fig. 2.2: Application menus, when clicked with the mouse, expand to show a list of actions that can be carried out.

As seen in Fig. 2.3, **toolbars** (rows of small pictures that can be clicked with the mouse) normally sit just below the menus and provide a quicker way to use frequently needed actions.

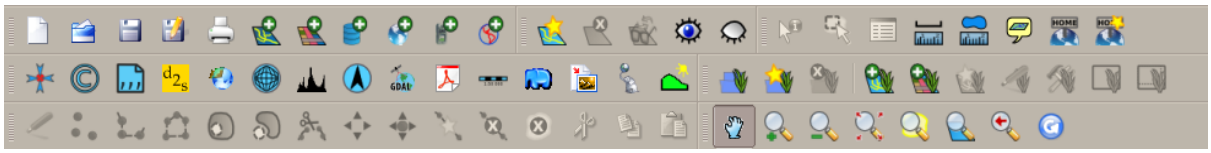


Fig. 2.3: Toolbars provide quick access to commonly used functions. Holding your mouse over a picture will usually tell you what will happen when you click on it.

A common function of GIS Applications is to display **map layers**. Map layers are stored as files on a disk or as records in a database. Normally each map layer will represent something in the real world — a roads layer for example will have data about the street network.

When you open a layer in the GIS Application it will appear in the **map view**. The map view shows a graphic representing your layer. When you add more than one layer to a map view, the layers are overlaid on top of each other. Look at figures Fig. 2.4, Fig. 2.5, Fig. 2.6 and Fig. 2.7 to see a map view that has several layers being added to it.

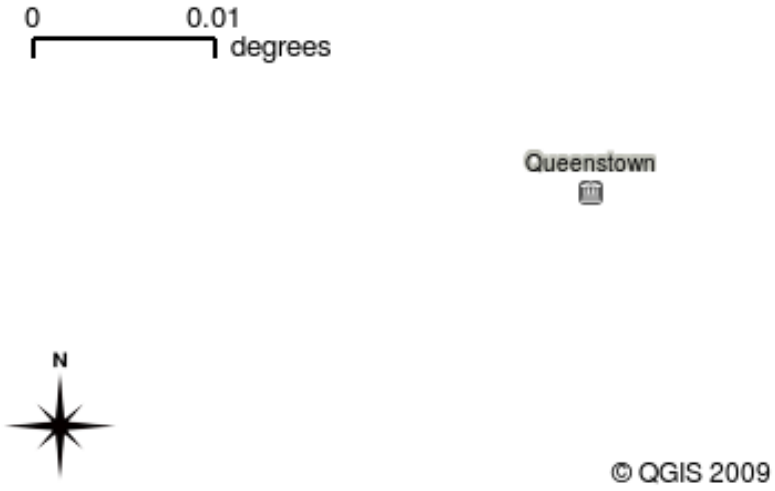


Fig. 2.4: A towns layer added to the map view

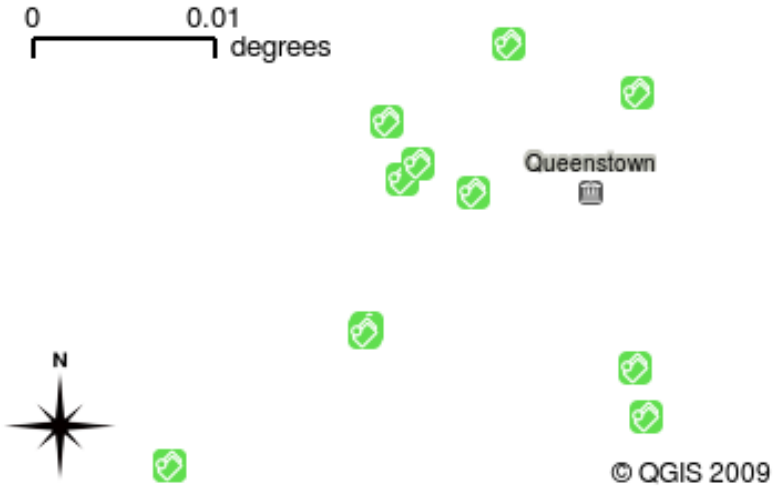


Fig. 2.5: A schools layer added to the map view

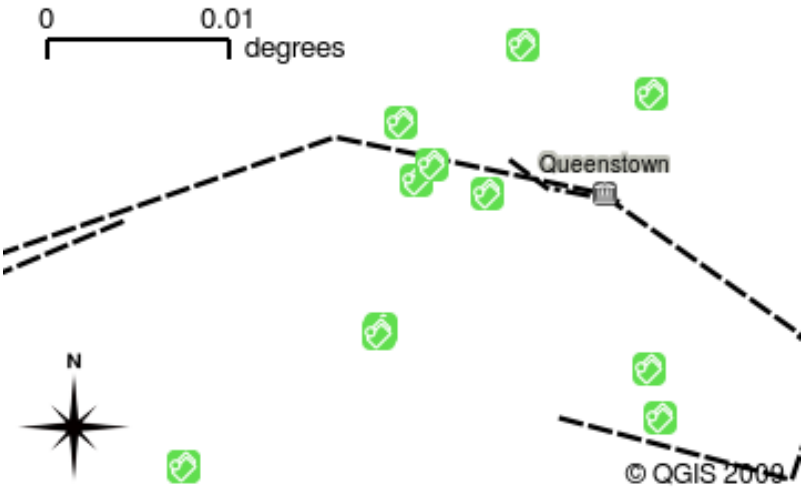


Fig. 2.6: A railways layer added to the map view

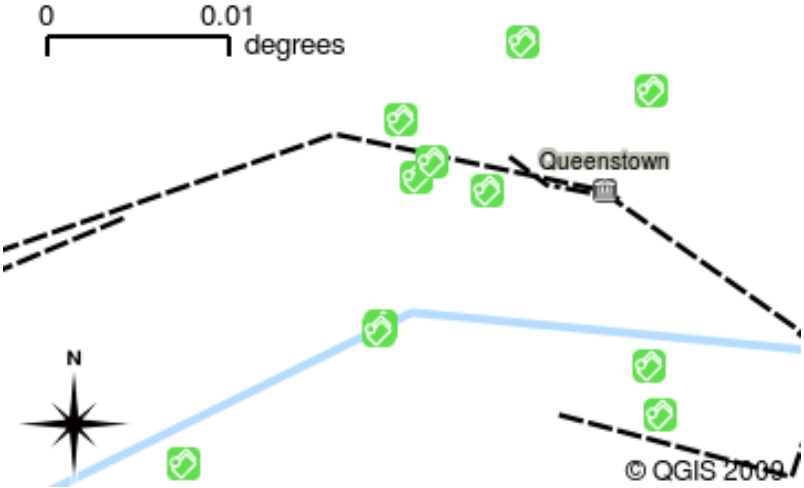


Fig. 2.7: A rivers layer added to the map view

An important function of the map view is to allow you to zoom in to magnify, zoom out to see a greater area and move around (panning) in the map. GIS software also lets you easily change symbology — the way information is displayed. Fig. 2.8 shows the map view after panning to the side and changing the symbology of the layers.

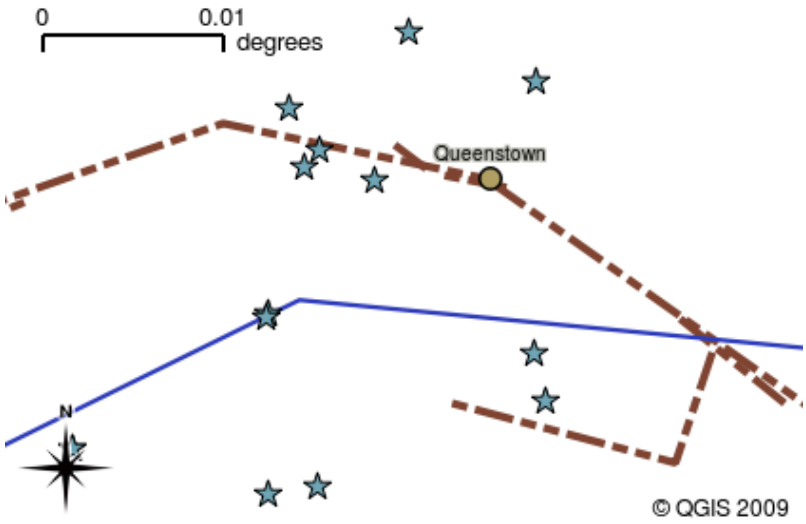


Fig. 2.8: A map view after moving around (panning) and changing the symbology

Another common feature of GIS Applications is the **map legend**. The map legend provides a list of layers that have been loaded in the GIS Application. Unlike a paper map legend, the map legend or 'layers list' in the GIS Application provides a way to re-order, hide, show and group layers. Changing the layer order is done by clicking on a layer in the legend, holding the mouse button down and then dragging the layer to a new position. In Fig. 2.9 and Fig. 2.10, the map legend is shown as the area to the left of the GIS Application window. By changing the layer order, the way that layers are drawn can be adjusted — in this case so that rivers are drawn below the roads instead of over them.

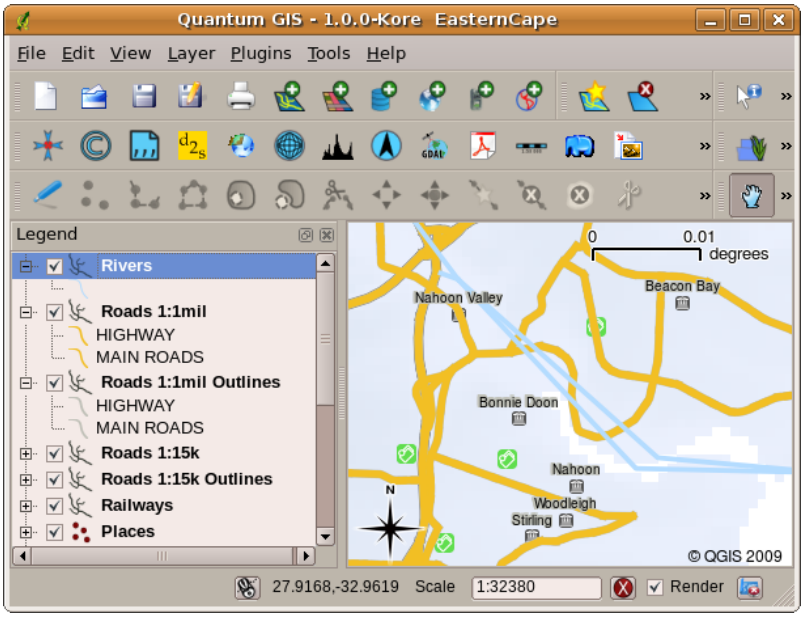


Fig. 2.9: Changing the layer order allows to adjust the way that layers are drawn. Before changing the layer order, rivers are drawn on top of roads

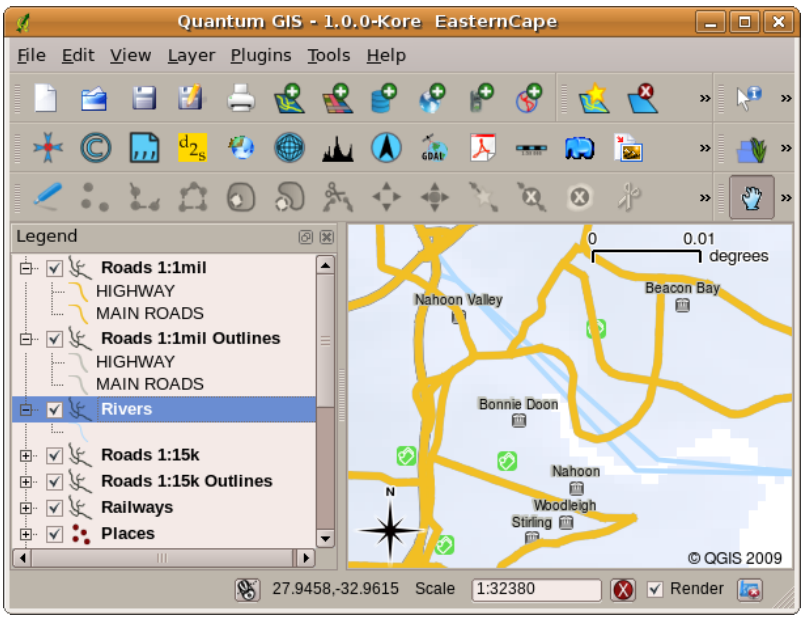


Fig. 2.10: Changing the layer order allows to adjust the way that layers are drawn. After changing the layer order, rivers are drawn underneath roads

2.4 Getting a GIS Application for your own computer(s)

There are many different GIS Applications available. Some have many sophisticated features and cost tens of thousands of Rands for each copy. In other cases, you can obtain a GIS Application for free. Deciding which GIS Application to use is a question of how much money you can afford and personal preference. For these tutorials, we will be using the QGIS Application. QGIS is completely free and you can copy it and share it with your friends as much as you like. If you received this tutorial in printed form, you should have received a copy of QGIS with it. If not, you can always visit https://www.qgis.org/ to download your free copy if you have access to the internet.

2.5 GIS Data

Now that we know what a GIS is and what a GIS Application can do, let's talk about **GIS data**. Data is another word for **information**. The information we use in a GIS normally has a geographical aspect to it. Think of our example above, about the health care worker. She created a table to record diseases that looked like this:

Longitude	Latitude	Disease	Date
26.870436	-31.909519	Mumps	13/12/2008

The longitude and latitude columns hold **geographical data**. The disease and date columns hold **non-geographical data**.

A common feature of GIS is that they allow you to associate information (non-geographical data) with places (geographical data). In fact, the GIS Application can store many pieces of information which are associated with each place — something that paper maps are not very good at. For example, our health care worker could store the person's age and gender on her table. When the GIS Application draws the layer, you can tell it to draw the layer based on gender, or based on disease type, and so on. So, with a GIS Application we have a way to easily change the appearance of the maps we created based on the non-geographical data associated with places.

GIS Systems work with many different types of data. **Vector data** are stored as a series of `X, Y` coordinate pairs inside the computer's memory. Vector data are used to represent points, lines and areas. Fig. 2.11 shows different types of vector data being viewed in a GIS application. In the tutorials that follow we will be exploring vector data in more detail.



Fig. 2.11: Vector data are used to represent points (e.g., towns), lines (e.g., rivers) and polygons (e.g., municipal boundaries).

Raster data are stored as a grid of values. There are many satellites circling the earth and the photographs they take are a kind of raster data that can be viewed in a GIS. One important difference between raster and vector data is that if you zoom in too much on a raster image, it will start to appear 'blocky' (see Fig. 2.12 and Fig. 2.13). In fact these blocks are the individual cells of the data grid that makes up the raster image. We will be looking at raster data in greater detail in later tutorials.



Fig. 2.12: Raster data are often images taken by satellites. Here we can see mountains in the Eastern Cape.



Fig. 2.13: The same raster data, but this time zoomed in. The grid nature of the data can be seen.

2.6 What have we learned?

Let's wrap up what we covered in this worksheet:

- A **GIS** is a system of computer hardware, computer software and geographical data.
- A **GIS Application** allows you to view geographical data and is an important part of the GIS.
- A GIS Application normally consists of a **menu bar**, **toolbars**, a **map view** and a **legend**.
- **Vector** and **raster** data are geographical data used in a GIS application.
- **Geographical** data can have associated **non-geographical** data.

2.7 Now you try!

Here are some ideas for you to try with your learners:

- **Geography:** Describe the concept of GIS to your learners as outlined in this tutorial. Ask them to try to think of 3 reasons why it might be handy to use a GIS instead of paper maps. Here are some that we could think of:
  - GIS Applications allow you to create many different types of maps from the same data.
  - GIS is a great visualisation tool that can show you things about your data and how they are related in space (e.g. those disease outbreaks we saw earlier).
  - Paper maps need to be filed and are time consuming to view. The GIS can hold a very large amount of map data and make it quick and easy to find a place you are interested in.
- **Geography:** Can you and your learners think of how raster data from satellites could be useful? Here are some ideas we had:
  - During natural disasters, raster data can be useful to show where the impacted areas are. For example a recent satellite image taken during a flood can help to show where people may need rescuing.
  - Sometimes people do bad things to the environment, like dumping dangerous chemicals that kill plants and animals. Using raster data from satellites can help us to monitor for these types of problems.
  - Town planners can use raster data from satellites to see where informal settlements are and to help in planning infrastructure.

2.8 Something to think about

If you don't have a computer available, many of the topics we cover in this tutorial can be reproduced using an overhead and transparency as it uses the same technique of layering information. However, to properly understand GIS it is always better to learn it using a computer.

2.9 Further reading

Book: Desktop GIS: Mapping the Planet with Open Source Tools. **Author:** Gary Sherman. **ISBN:** 9781934356067

The QGIS User Guide also has more detailed information on working with QGIS.

2.10 What's next?

In the sections that follow we are going to go into more detail, showing you how to use a GIS Application. All of the tutorials will be done using QGIS. Next up, let's look at vectors!

CHAPTER

THREE

VECTOR DATA



Objectives:	Understanding of vector data models as used in GIS.
Keywords:	Vector, Point, Polyline, Polygon, Vertex, Geometry, Scale, Data Quality, Symbology, Data Sources

3.1 Overview

Vector data provide a way to represent real world **features** within the GIS environment. A feature is anything you can see on the landscape. Imagine you are standing on the top of a hill. Looking down you can see houses, roads, trees, rivers, and so on (see Fig. 3.1). Each one of these things would be a **feature** when we represent them in a GIS Application. Vector features have **attributes**, which consist of text or numerical information that **describe** the features.



Fig. 3.1: Looking over a landscape you can see the main features, such as roads, houses and trees.

A vector feature has its shape represented using **geometry**. The geometry is made up of one or more interconnected **vertices**. A vertex describes a position in space using an **X**, **Y** and optionally **Z** axis. Geometries which have vertices with a `Z` axis are often referred to as **2.5D** since they describe height or depth at each vertex, but not both.

When a feature's geometry consists of only a single vertex, it is referred to as a **point** feature (see illustration Fig. 3.2). Where the geometry consists of two or more vertices and the first and last vertex are not equal, a **polyline** feature is formed (see illustration Fig. 3.3). Where four or more vertices are present, and the last vertex is equal to the first, an enclosed **polygon** feature is formed (see illustration Fig. 3.4).

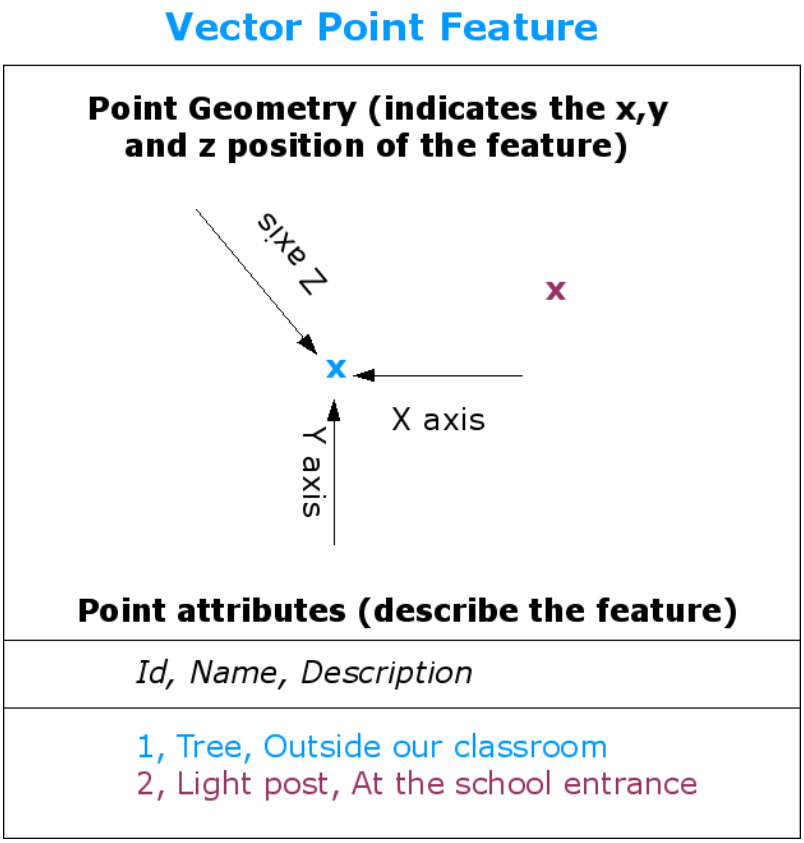


Fig. 3.2: A point feature is described by its X, Y and optionally Z coordinate. The point attributes describe the point e.g. if it is a tree or a lamp post.

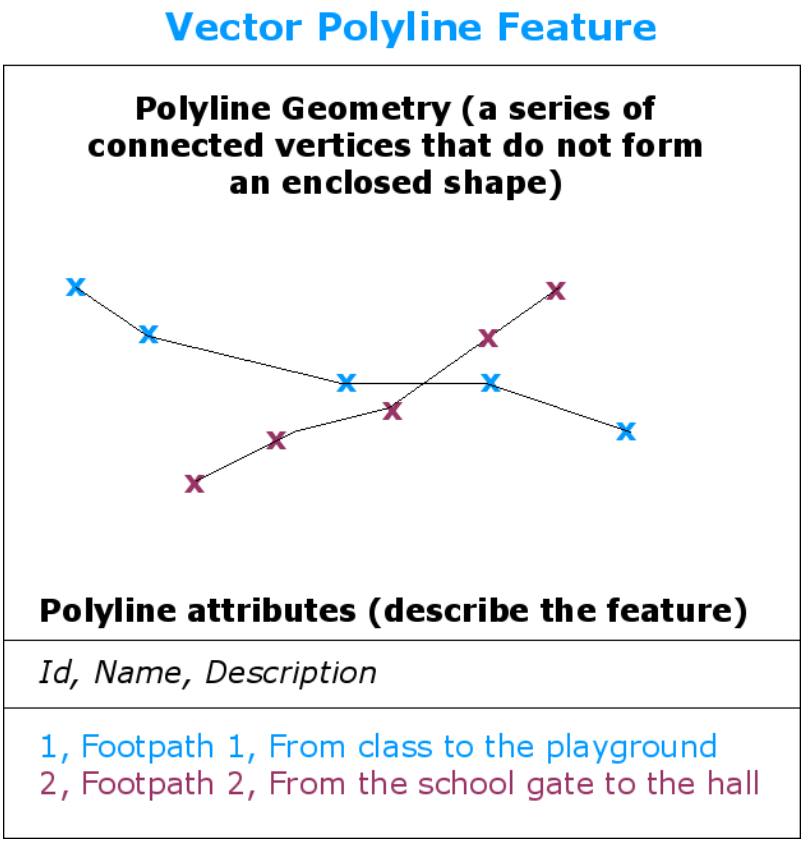


Fig. 3.3: A polyline is a sequence of joined vertices. Each vertex has an X, Y (and optionally Z) coordinate. Attributes describe the polyline.

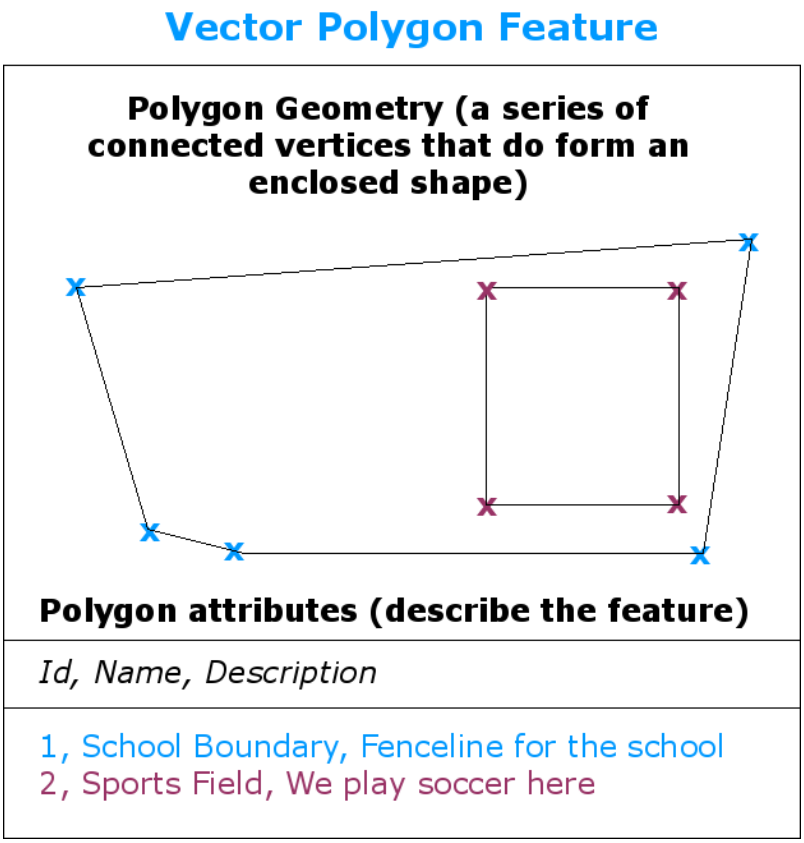


Fig. 3.4: A polygon, like a polyline, is a sequence of vertices. However in a polygon, the first and last vertices are always at the same position.

Looking back at the picture of a landscape we showed you further up, you should be able to see the different types of features in the way that a GIS represents them now (see illustration Fig. 3.5).

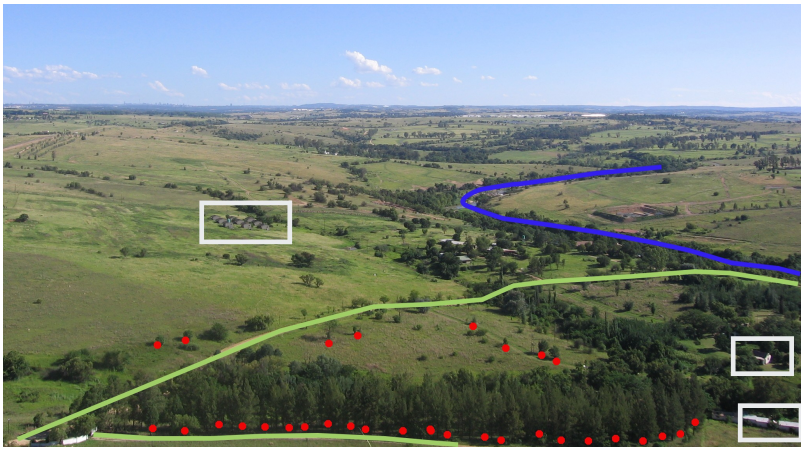


Fig. 3.5: Landscape features as we would present them in a GIS. Rivers (blue) and roads (green) can be represented as lines, trees as points (red) and houses as polygons (white).

3.2 Point features in detail

The first thing we need to realise when talking about point features is that what we describe as a point in GIS is a matter of opinion, and often dependent on scale. Let's look at cities for example. If you have a small scale map (which covers a large area), it may make sense to represent a city using a point feature. However as you zoom in to

the map, moving towards a larger scale, it makes more sense to show the city limits as a polygon.

When you choose to use points to represent a feature is mostly a matter of scale (how far away are you from the feature), convenience (it takes less time and effort to create point features than polygon features), and the type of feature (some things like telephone poles just don't make sense to be stored as polygons).

As we show in illustration Fig. 3.2, a point feature has an X, Y and optionally, Z value. The X and Y values will depend on the **Coordinate Reference System** (CRS) being used. We are going to go into more detail about Coordinate Reference Systems in a later tutorial. For now let's simply say that a CRS is a way to accurately describe where a particular place is on the earth's surface. One of the most common reference systems is **Longitude and Latitude**. Lines of Longitude run from the North Pole to the South Pole. Lines of Latitude run from the East to West. You can describe precisely where you are at any place on the earth by giving someone your Longitude (X) and Latitude (Y). If you make a similar measurement for a tree or a telephone pole and mark it on a map, you will have created a point feature.

Since we know the earth is not flat, it is often useful to add a Z value to a point feature. This describes how high above sea level you are.

3.3 Polyline features in detail

Where a point feature is a single vertex, **a polyline has two or more vertices**. The polyline is a continuous path drawn through each vertex, as shown in Fig. 3.3. When two vertices are joined, a line is created. When more than two are joined, they form a 'line of lines', or **polyline**.

A polyline is used to show the geometry of **linear features** such as roads, rivers, contours, footpaths, flight paths and so on. Sometimes we have special rules for polylines in addition to their basic geometry. For example contour lines may touch (e.g. at a cliff face) but should never cross over each other. Similarly, polylines used to store a road network should be connected at intersections. In some GIS applications you can set these special rules for a feature type (e.g. roads) and the GIS will ensure that these polylines always comply with these rules.

If a curved polyline has very large distances between vertices, it may appear **angular** or jagged, depending on the scale at which it is viewed (see Fig. 3.6). Because of this it is important that polylines are digitised (captured into the computer) with distances between vertices that are small enough for the scale at which you want to use the data.

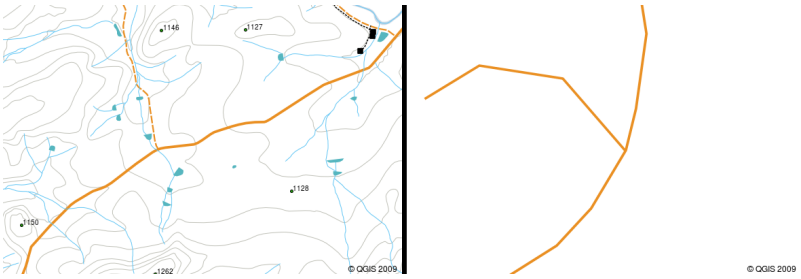


Fig. 3.6: Polylines viewed at a smaller scale (1:20 000 to the left) may appear smooth and curved. When zoomed in to a larger scale (1:500 to the right) polylines may look very angular.

The **attributes** of a polyline describe its properties or characteristics. For example a road polyline may have attributes that describe whether it is surfaced with gravel or tar, how many lanes it has, whether it is a one way street, and so on. The GIS can use these attributes to symbolise the polyline feature with a suitable colour or line style.

3.4 Polygon features in detail

Polygon features are **enclosed areas** like dams, islands, country boundaries and so on. Like polyline features, polygons are created from a series of vertices that are connected with a continuous line. However because a polygon always describes an enclosed area, the first and last vertices should always be at the same place! Polygons often have **shared geometry** — boundaries that are in common with a neighbouring polygon. Many GIS applications have the capability to ensure that the boundaries of neighbouring polygons exactly coincide. We will explore this in the *Topology* topic later in this tutorial.

As with points and polylines, polygons have **attributes**. The attributes describe each polygon. For example a dam may have attributes for depth and water quality.

3.5 Vector data in layers

Now that we have described what vector data is, let's look at how vector data is managed and used in a GIS environment. Most GIS applications group vector features into **layers**. Features in a layer have the same geometry type (e.g. they will all be points) and the same kinds of attributes (e.g. information about what species a tree is for a trees layer). For example if you have recorded the positions of all the footpaths in your school, they will usually be stored together on the computer hard disk and shown in the GIS as a single layer. This is convenient because it allows you to hide or show all of the features for that layer in your GIS application with a single mouse click.

3.6 Editing vector data

The GIS application will allow you to create and modify the geometry data in a layer — a process called **digitising** — which we will look at more closely in a later tutorial. If a layer contains polygons (e.g. farm dams), the GIS application will only allow you to create new polygons in that layer. Similarly if you want to change the shape of a feature, the application will only allow you to do it if the changed shape is correct. For example it won't allow you to edit a line in such a way that it has only one vertex — remember in our discussion of lines above that all lines must have at least two vertices.

Creating and editing vector data is an important function of a GIS since it is one of the main ways in which you can create personal data for things you are interested in. Say for example you are monitoring pollution in a river. You could use the GIS to digitise all outfalls for storm water drains (as point features). You could also digitise the river itself (as a polyline feature). Finally you could take readings of pH levels along the course of the river and digitise the places where you made these readings (as a point layer).

As well as creating your own data, there is a lot of free vector data that you can obtain and use. For example, you can obtain vector data that appears on the 1:50 000 map sheets from the Chief Directorate: Surveys and Mapping.

3.7 Scale and vector data

Map **scale** is an important issue to consider when working with vector data in a GIS. When data is captured, it is usually digitised from existing maps, or by taking information from surveyor records and global positioning system devices. Maps have different scales, so if you import vector data from a map into a GIS environment (for example by digitising paper maps), the digital vector data will have the same scale issues as the original map. This effect can be seen in illustrations Fig. 3.7 and Fig. 3.8. Many issues can arise from making a poor choice of map scale. For example using the vector data in illustration Fig. 3.7 to plan a wetland conservation area could result in important parts of the wetland being left out of the reserve! On the other hand if you are trying to create a regional map, using data captured at 1:1000 000 might be just fine and will save you a lot of time and effort capturing the data.

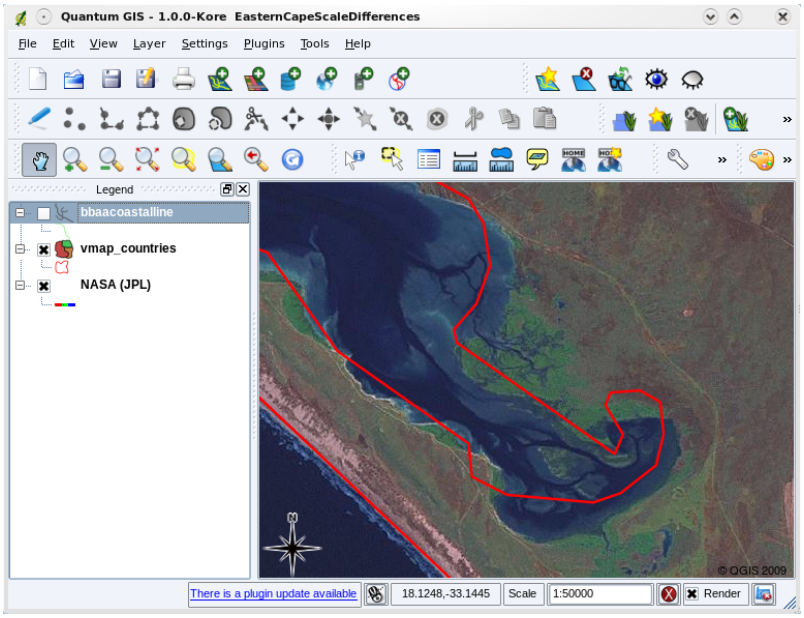


Fig. 3.7: Vector data (red lines) that was digitised from a small scale (1:1000 000) map.

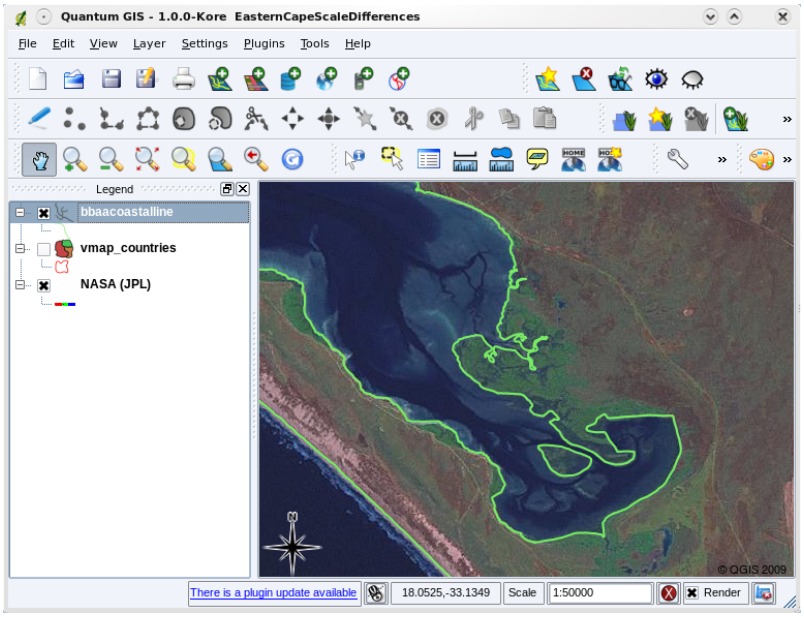


Fig. 3.8: Vector data (green lines) that was digitised from a large scale (1:50 000) map.

3.8 Symbology

When you add vector layers to the map view in a GIS application, they will be drawn with random colours and basic symbols. One of the great advantages of using a GIS is that you can create personalised maps very easily. The GIS program will let you choose colours to suit the feature type (e.g. you can tell it to draw a water bodies vector layer in blue). The GIS will also let you adjust the symbol used. So if you have a trees point layer, you can show each tree position with a small picture of a tree, rather than the basic circle marker that the GIS uses when you first load the layer (see illustrations Fig. 3.9, Fig. 3.10 and Fig. 3.11).

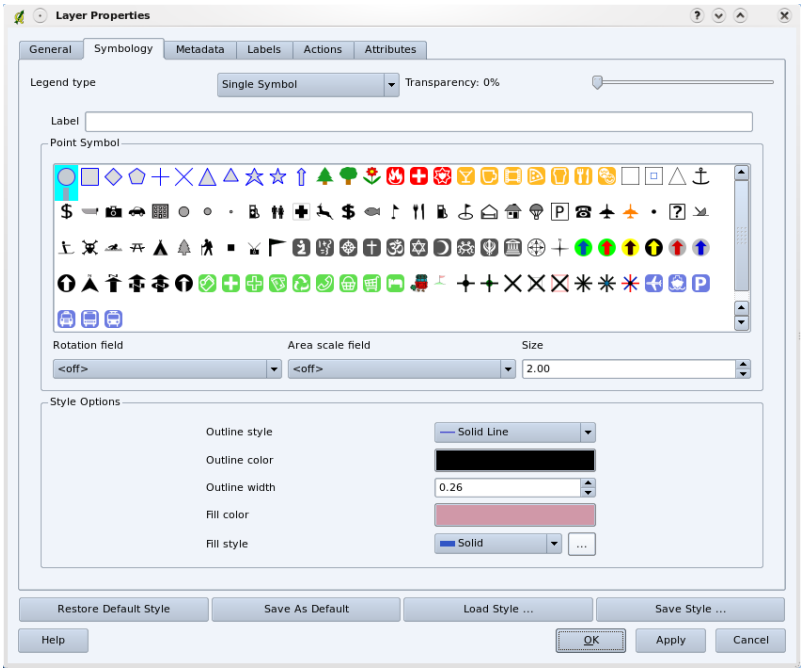


Fig. 3.9: In the GIS, you can use a panel (like the one above) to adjust how features in your layer should be drawn.

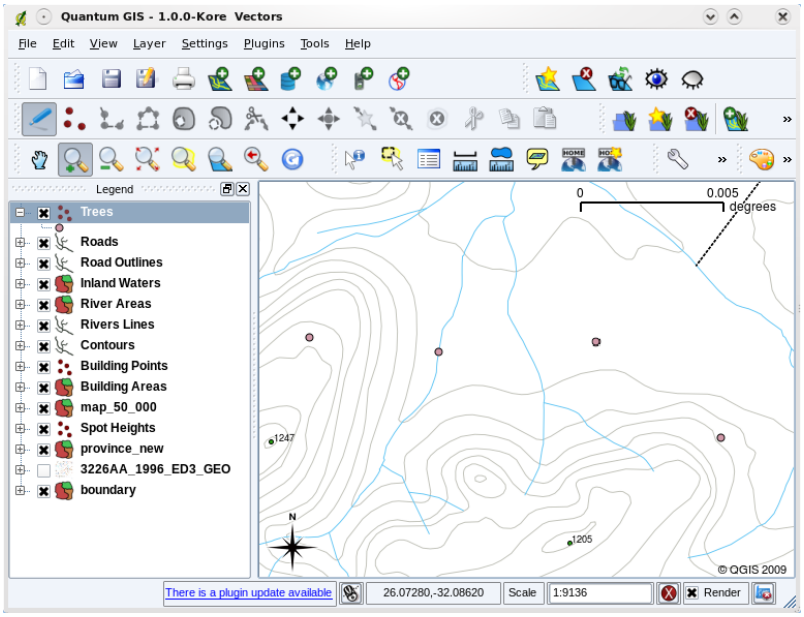


Fig. 3.10: When a layer (for example the trees layer above) is first loaded, a GIS application will give it a generic symbol.

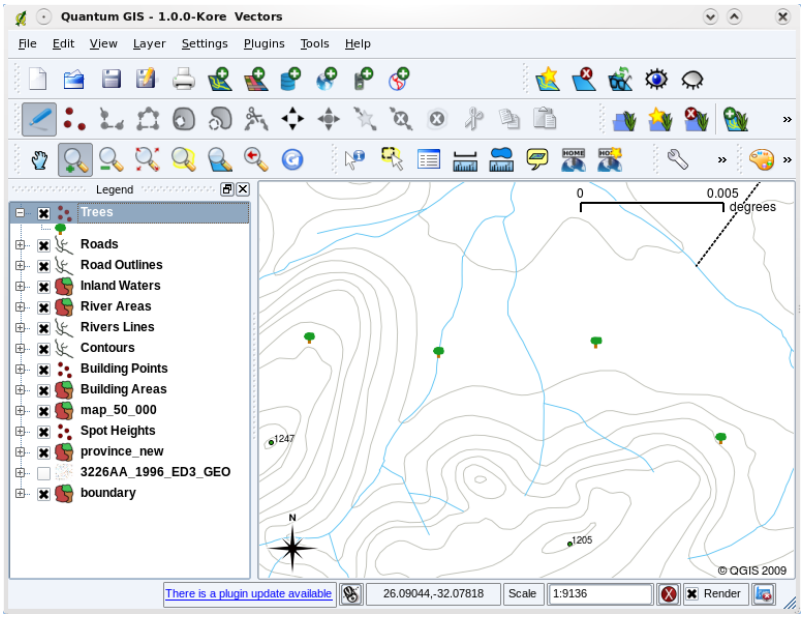


Fig. 3.11: After making our adjustments it is much easier to see that our points represent trees.

Symbology is a powerful feature, making maps come to life and the data in your GIS easier to understand. In the topic that follows (*Vector Attribute Data*) we will explore more deeply how symbology can help the user to understand vector data.

3.9 What can we do with vector data in a GIS?

At the simplest level we can use vector data in a GIS Application in much the same way you would use a normal topographic map. The real power of GIS starts to show itself when you start to ask questions like 'which houses are within the 100 year flood level of a river?'; 'where is the best place to put a hospital so that it is easily accessible to as many people as possible?'; 'which learners live in a particular suburb?'. A GIS is a great tool for answering these types of questions with the help of vector data. Generally we refer to the process of answering these types of questions as **spatial analysis**. In later topics of this tutorial we will look at spatial analysis in more detail.

3.10 Common problems with vector data

Working with vector data does have some problems. We already mentioned the issues that can arise with vectors captured at different scales. Vector data also needs a lot of work and maintenance to ensure that it is accurate and reliable. Inaccurate vector data can occur when the instruments used to capture the data are not properly set up, when the people capturing the data aren't being careful, when time or money don't allow for enough detail in the collection process, and so on.

If you have poor quality vector data, you can often detect this when viewing the data in a GIS. For example **slivers** can occur when the edges of two polygon areas don't meet properly (see Fig. 3.12).

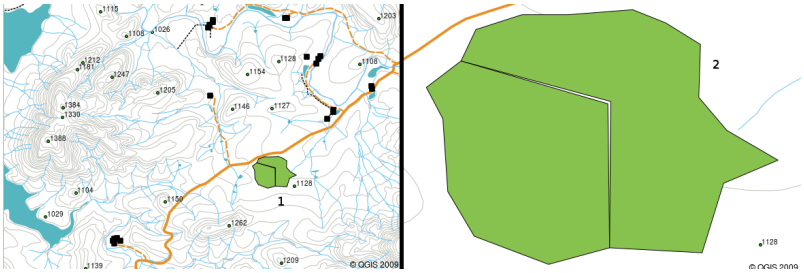


Fig. 3.12: Slivers occur when the vertices of two polygons do not match up on their borders. At a small scale (e.g. 1 on left) you may not be able to see these errors. At a large scale they are visible as thin strips between two polygons (2 on right).

Overshoots can occur when a line feature such as a road does not meet another road exactly at an intersection. **Undershoots** can occur when a line feature (e.g. a river) does not exactly meet another feature to which it should be connected. Figure Fig. 3.13 demonstrates what undershoots and overshoots look like.



Fig. 3.13: Undershoots (1) occur when digitised vector lines that should connect to each other don't quite touch. Overshoots (2) happen if a line ends beyond the line it should connect to.

Because of these types of errors, it is very important to digitise data carefully and accurately. In the upcoming topic on **topology**, we will examine some of these types of errors in more detail.

3.11 What have we learned?

Let's wrap up what we covered in this worksheet:

- **Vector data** is used to represent real world **features** in a GIS.
- A vector feature can have a **geometry** type of **point**, **line** or a **polygon**.
- Each vector feature has **attribute data** that describes it.
- Feature geometry is described in terms of **vertices.**
- Point geometries are made up of a **single vertex** (X,Y and optionally Z).
- Polyline geometries are made up of **two or more** vertices forming a connected line.
- Polygon geometries are made up of **at least four vertices** forming an enclosed area. The first and last vertices are always in the same place.

- Choosing which geometry type to use depends on scale, convenience and what you want to do with the data in the GIS.
- Most GIS applications do not allow you to mix more than one geometry type in a single layer.
- Digitising is the process of creating digital vector data by drawing it in a GIS application.
- Vector data can have quality issues such as **undershoots**, **overshoots** and **slivers** which you need to be aware of.
- Vector data can be used for **spatial analysis** in a GIS application, for example to find the nearest hospital to a school.

We have summarised the GIS Vector Data concept in Figure Fig. 3.14.

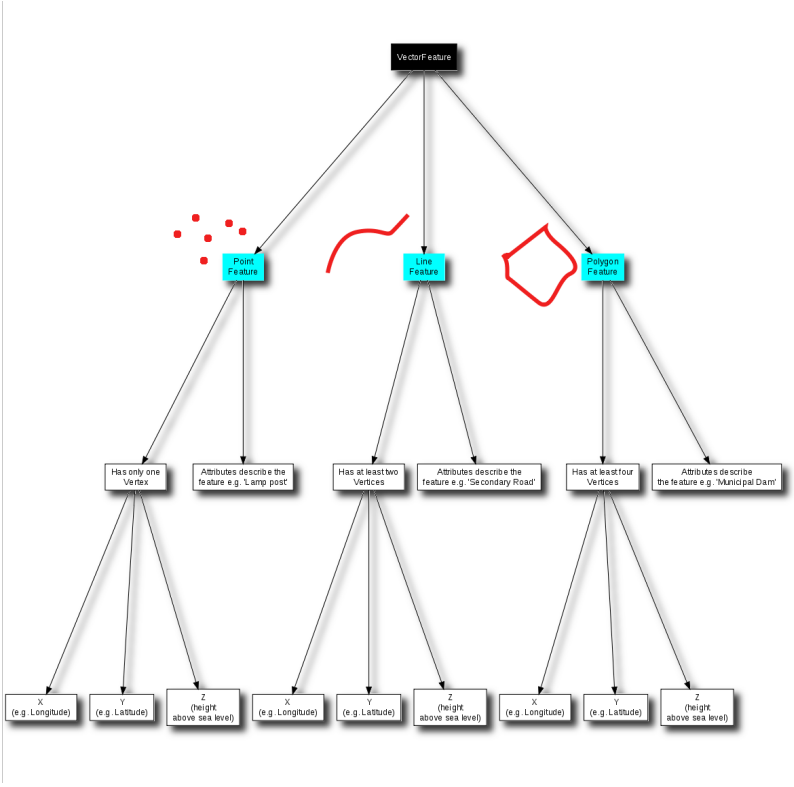


Fig. 3.14: This diagram shows how GIS applications deal with vector data.

3.12 Now you try!

Here are some ideas for you to try with your learners:

- Using a copy of a toposheet map for your local area (like the one shown in Fig. 3.15), see if your learners can identify examples of the different types of vector data by highlighting them on the map.
- Think of how you would create vector features in a GIS to represent real world features on your school grounds. Create a table of different features in and around your school and then task your learners to decide whether they would be best represented in the GIS as a point, line or polygon. See Table 3.1 for an example.

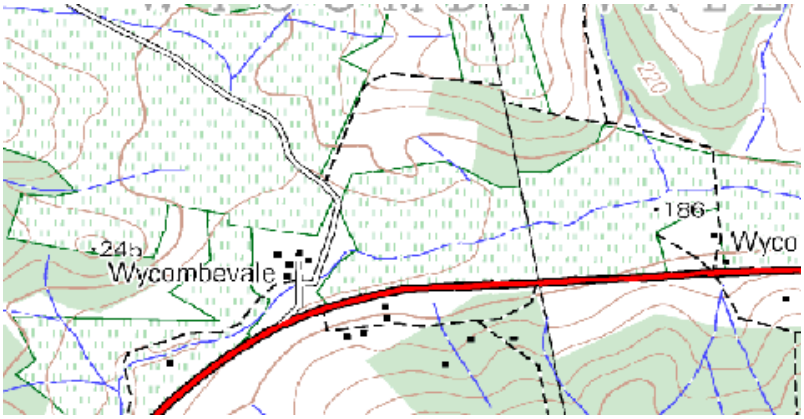


Fig. 3.15: Can you identify two point features and one polygon feature on this map?

Table 3.1: Vector data and geometry types

Real world feature	Suitable Geometry Type
The school flagpole	
The soccer field	
The footpaths in and around the school	
Places where taps are located	
Etc.	

Create a table like Table 3.1 (leaving the "geometry type" column empty) and ask your learners to decide on suitable geometry types.

3.13 Something to think about

If you don't have a computer available, you can use a toposheet and transparency sheets to show your learners about vector data.

3.14 Further reading

The QGIS User Guide also has more detailed information on working with vector data in QGIS.

3.15 What's next?

In the section that follows we will take a closer look at **attribute data** to see how it can be used to describe vector features.

CHAPTER

FOUR

VECTOR ATTRIBUTE DATA

Objectives:	In this topic we describe how attribute data are associated with vector features and can be used to symbolise data.
Keywords:	Attribute, database, fields, data, vector, symbology

4.1 Overview

If every line on a map was the same colour, width, thickness, and had the same label, it would be very hard to make out what was going on. The map would also give us very little information. Take a look at Fig. 4.1 for example.

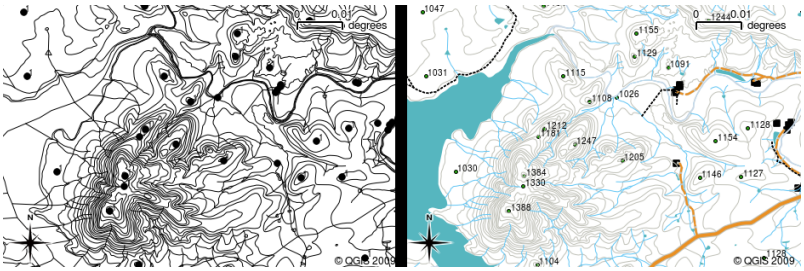


Fig. 4.1: Maps come to life when colour and different symbols are used to help you to tell one type of feature from the next. Can you tell the difference between rivers, roads and contours using the map on the left? Using the map on the right it is much easier to see the different features.

In this topic we will look at how attribute data can help us to make interesting and informative maps. In the previous topic on vector data, we briefly explained that **attribute data** are used to **describe vector features**. Take a look at the house pictures in Fig. 4.2.



Fig. 4.2: Every feature has characteristics that we can describe. These can be visible things, or things we know about the feature (e.g. year built).

The geometry of these house features is a polygon (based on the floor plan of the house), the attributes we have recorded are roof colour, whether there is a balcony, and the year the house was built. Note that attributes don't have to be visible things — they can describe things we know about the feature such as the year it was built. In a GIS Application, we can represent this feature type in a houses polygon layer, and the attributes in an attribute table (see Fig. 4.3).

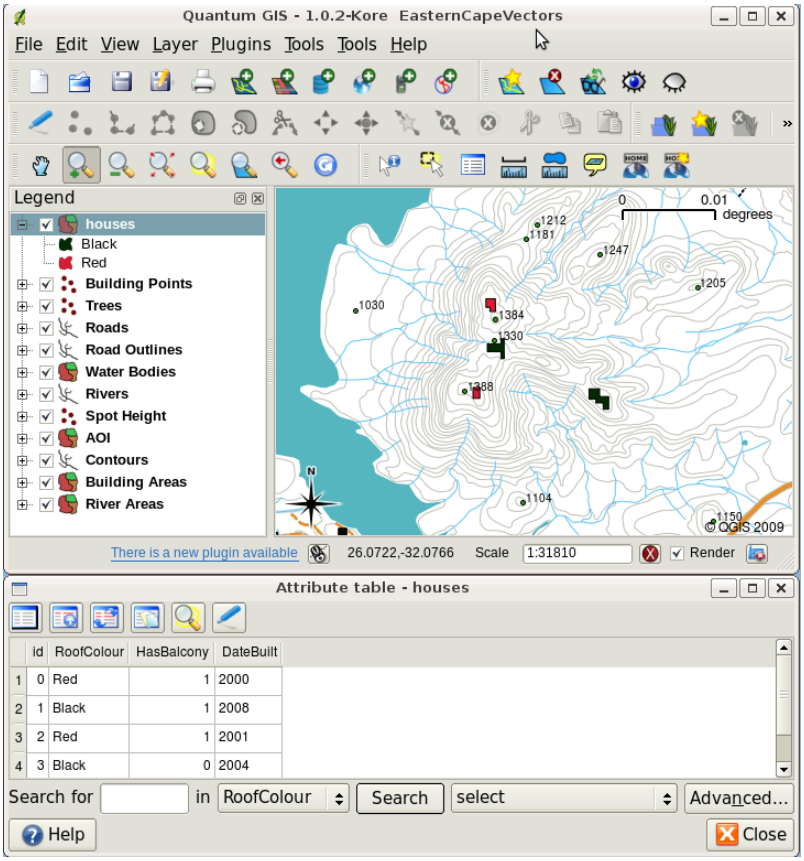


Fig. 4.3: A houses layer. House features have attributes that describe the houses' roof colour and other properties. The attribute table (lower image) lists the attributes for the house areas shown on the map. When a feature is highlighted in the table, it will appear as a yellow polygon on the map.

The fact that features have attributes as well geometry in a GIS Application opens up many possibilities. For example we can use the attribute values to tell the GIS what colours and style to use when drawing features (see Fig. 4.4). The process of setting colours and drawing styles is often referred to as setting feature **symbology**.

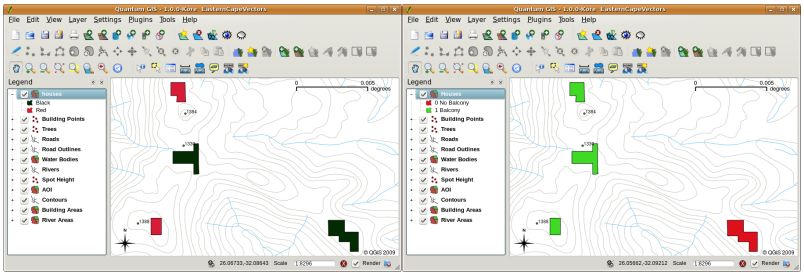


Fig. 4.4: In a GIS Application, we can draw features differently depending on their attributes. On the left we have drawn house polygons with the same colour as the roof attribute. On the right we colour coded houses according to whether they have a balcony or not.

Attribute data can also be useful when creating **map labels**. Most GIS Applications will have a facility to select an attribute that should be used to label each feature.

If you have ever **searched a map** for a place name or a specific feature, you will know how time consuming it can be. Having attribute data can make searching for a specific feature quick and easy. In Fig. 4.5 you can see an example of an attribute search in a GIS.

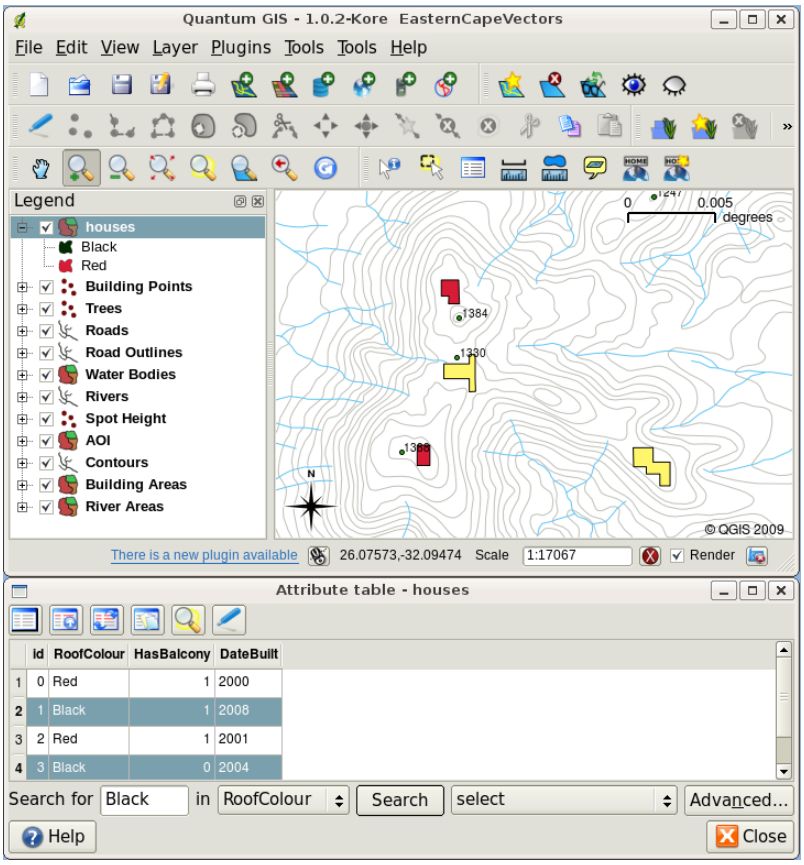


Fig. 4.5: In a GIS Application, we can also search for features based on their attributes. Here we see a search for houses with black roofs. Results are shown in yellow in the map, turquoise on the table.

Finally, attribute data can be very useful in carrying out **spatial analysis**. Spatial analysis combines the spatial information stored in the geometry of features with their attribute information. This allows us to study features and how they relate to each other. There are many types of spatial analysis that can be carried out, for example, you could use GIS to find out how many red roofed houses occur in a particular area. If you have tree features, you could use GIS to try to find out which species might be affected if a piece of land is developed. We can use the attributes stored for water samples along a river course to understand where pollution is entering into the stream. The possibilities are endless! In a later topic we will be exploring spatial analysis in more detail.

Before we move on to attribute data in more detail, let's take a quick recap.

Features are real world things such as roads, property boundaries, electrical substation sites and so on. A **feature** has a **geometry** (which determines if it is a **point**, **polyline** or **polygon**) and **attributes** (which describe the feature). This is shown in Fig. 4.6.

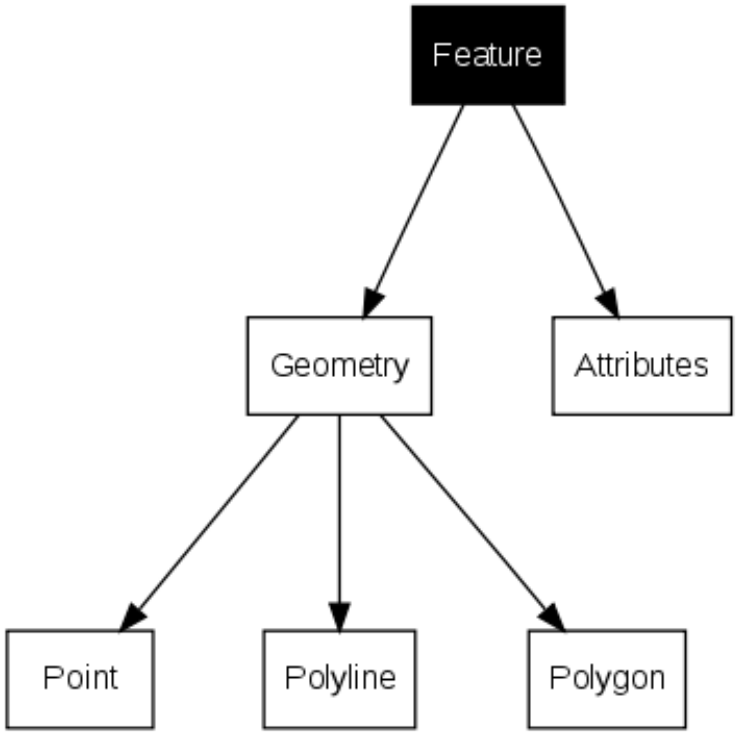


Fig. 4.6: Vector features at a glance.

4.2 Attributes in detail

Attributes for a vector feature are stored in a **table**. A table is like a spreadsheet. Each column in the table is called a **field**. Each row in the table is a **record**. Table 4.1 shows a simple example of how an attribute table looks in a GIS. The records in the attribute table in a GIS each correspond to one feature. Usually the information in the attribute table is stored in some kind of database. The GIS application links the attribute records with the feature geometry so that you can find records in the table by selecting features on the map, and find features on the map by selecting features in the table.

Table 4.1: House Attributes: An attribute table has fields (columns) and records (in rows).

Attribute Table	Field 1 : YearBuilt	Field 2: RoofColour	Field 3: Balcony
Record 1	1998	Red	Yes
Record 2	2000	Black	No
Record 3	2001	Silver	Yes

Each field in the attribute table contains a specific type of data — text, numeric or date. Deciding what attributes to use for a feature requires some thought and planning. In our house example earlier on in this topic, we chose roof colour, presence of a balcony and year of construction as attributes of interest. We could just as easily have chosen other aspects of a house such as:

- number of levels
- number of rooms
- number of occupants
- type of dwelling (RDP House, block of flats, shack, brick house, etc)
- month the house was built

- area of floor space in the house
- and so on….

With so many options, how do we make a good choice as to what attributes are needed for a feature? It usually boils down to what you plan to do with the data. If you want to produce a colour coded map showing houses by age, it will make sense to have a 'Year Built' attribute for your feature. If you know for sure you will never use this type of map, it is better to not store the information. Collecting and storing unneeded information is a bad idea because of the cost and time required to research and capture the information. Very often we obtain vector data from companies, friends or the government. In these cases it is usually not possible to request specific attributes and we have to make do with what we get.

4.3 Single Symbols

If a feature is symbolised without using any attribute table data, it can only be drawn in a simple way. For example with point features you can set the colour and **marker** (circle, square, star etc.) but that is all. You cannot tell the GIS to draw the features based on one of its properties in the attribute table. In order to do that, you need to use either a **graduated**, **continuous** or **unique value** symbol. These are described in detail in the sections that follow.

A GIS application will normally allow you to set the symbology of a layer using a **dialog box** such as the one shown in in Fig. 4.7. In this dialog box you can choose colours and symbol styles. Depending on the geometry type of a layer, different options may be shown. For example with point layers you can choose a **marker style**. With line and polygon layers there is no marker style option, but instead you can select a **line style** and **colour** such as dashed orange for gravel roads, solid orange for minor roads, and so on (as shown in Fig. 4.8). With polygon layers you also have the option of setting a **fill style** and color.

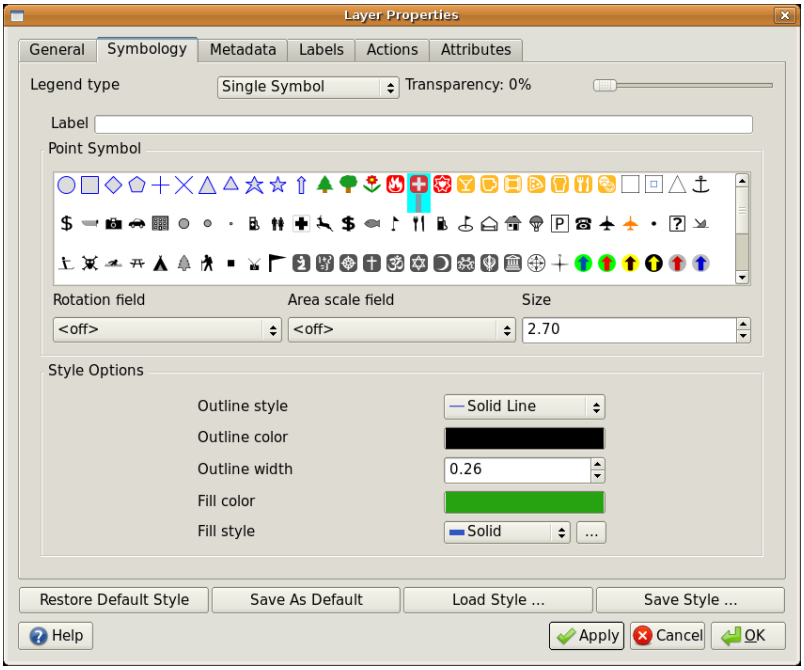


Fig. 4.7: When using simple symbols, the feature is drawn without using an attribute to control how it looks. This is the dialog for point features.

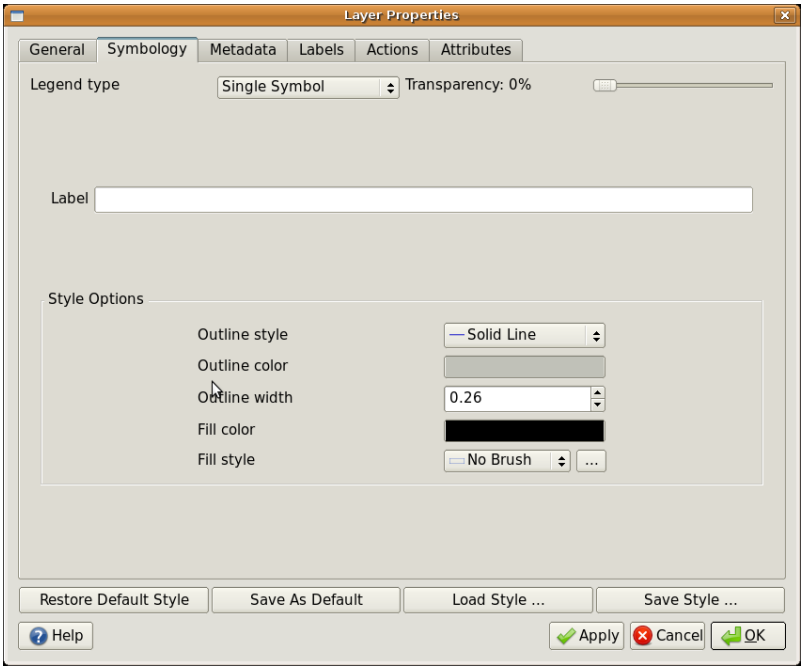


Fig. 4.8: There are different options when defining simple symbols for polyline and polygon features.

4.4 Graduated Symbols

Sometimes vector features represent things with a changing numerical value. Contour lines are a good example of this. Each contour usually has an attribute value called 'height' that contains information about what height that contour represents. In earlier in this topic we showed contours all drawn with the same colour. Adding colour to the contours can help us to interpret the meanings of contours. For example we can draw low lying areas with one colour, mid-altitude areas with another and high-altitude areas with a third.

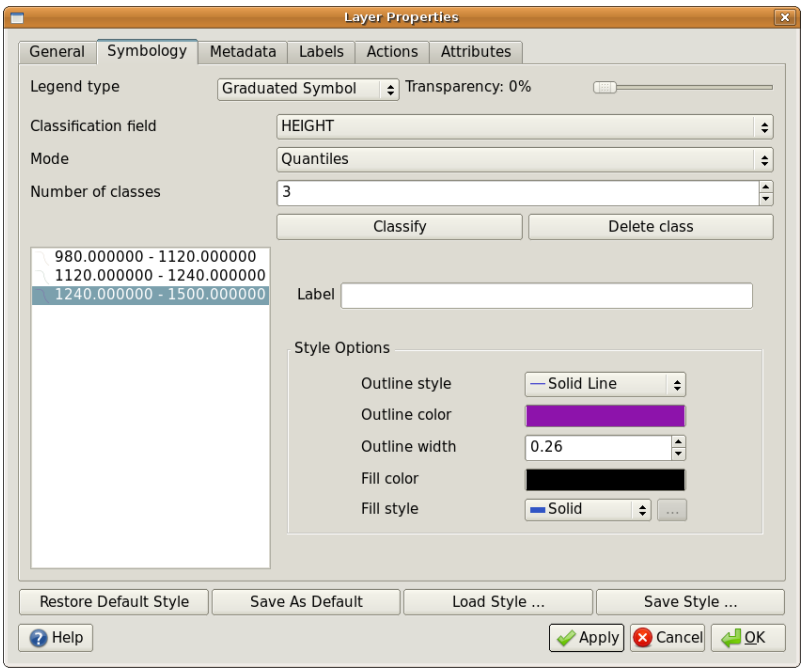


Fig. 4.9: The height attribute of contours can be used to separate the contours into 3 classes. Contours between 980 m and 1120 m will be drawn in brown, those between 1120 m and 1240 m in green and those between 1240 m and 1500 m in purple.

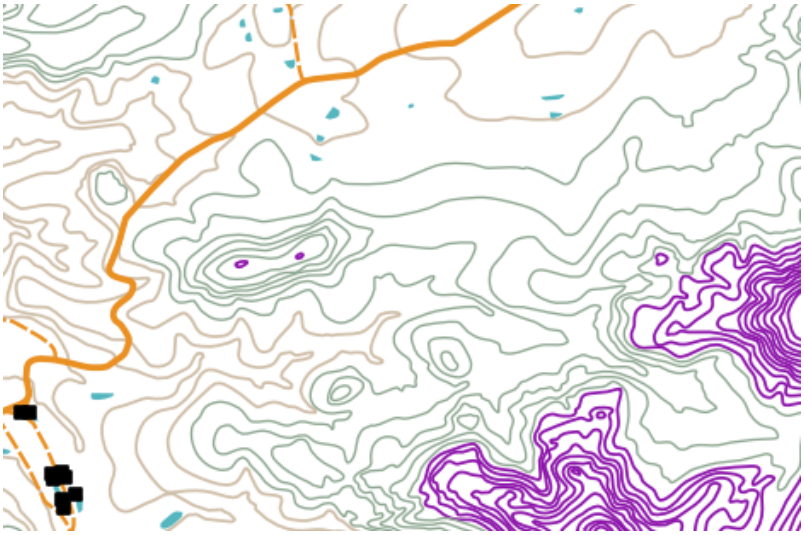


Fig. 4.10: Our map after setting graduated colours for our contours.

Setting colours based on discrete groups of attribute values is called Graduated Symbology in QGIS. The process is shown in Illustrations Fig. 4.9 and Fig. 4.10. **Graduated symbols** are most useful when you want to show clear differences between features with attribute values in different value ranges. The GIS Application will analyse the attribute data (e.g. height) and, based on the number of classes you request, create groupings for you. This process is illustrated in Table 4.2: graduated colour breaks up the attribute value ranges into the number of classes you select; each class is represented by a different colour.

Table 4.2: Graduated Symbols

Attribute Value	Class and Colour
1	Class 1
2	Class 1
3	Class 1
4	Class 2
5	Class 2
6	Class 2
7	Class 3
8	Class 3
9	Class 3

4.5 Continuous Colour Symbols

In the previous section on Graduated Colour symbols we saw that we can draw features in discrete groups or classes. Sometimes it is useful to draw features in a **colour range** from one colour to another. The GIS Application will use a numerical attribute value from a feature (e.g. contour heights or pollution levels in a stream) to decide which colour to use. Table 4.3 shows how the attribute value is used to define a continuous range of colours. Continuous colour symbology uses a start colour (e.g. light orange shown here) and an end colour (e.g. dark brown shown here) and creates a series of shades between those colours.

Table 4.3: Continuous colour symbology

Attribute Value	Colour (no classes or grouping)
1	
2	
3	
4	
5	
6	
7	
8	
9	

Using the same contours example we used in the previous section, let's see how a map with continuous colour symbology is defined and looks. The process starts by setting the layers properties to continuous colour using a dialog like the one shown in Fig. 4.11.

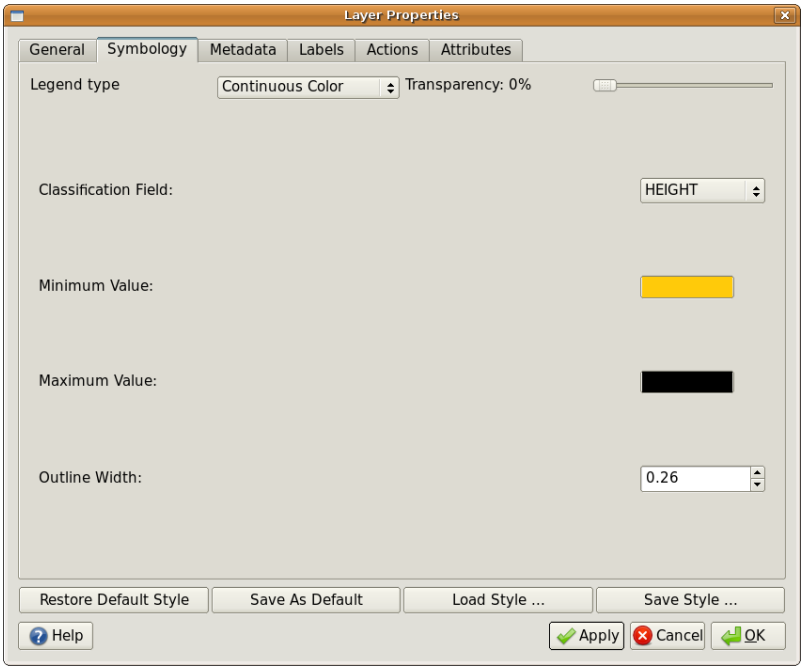


Fig. 4.11: Setting up continuous colour symbology. The contour height attribute is used to determine colour values. Colours are defined for the minimum and maximum values. The GIS Application will then create a gradient of colours for drawing the features based on their heights.

After defining the minimum and maximum colours in the colour range, the colour features are drawn in will depend on where the attribute lies in the range between minimum and maximum. For example if you have contour features with values starting at 1000 m and ending at 1400 m, the value range is 1000 to 1400. If the colour set for the minimum value is set to orange and the colour for the maximum value is black, contours with a value of close to 1400 m will be drawn close to black. On the other hand contours with a value near to 1000 m will be drawn close to orange (see Fig. 4.12).

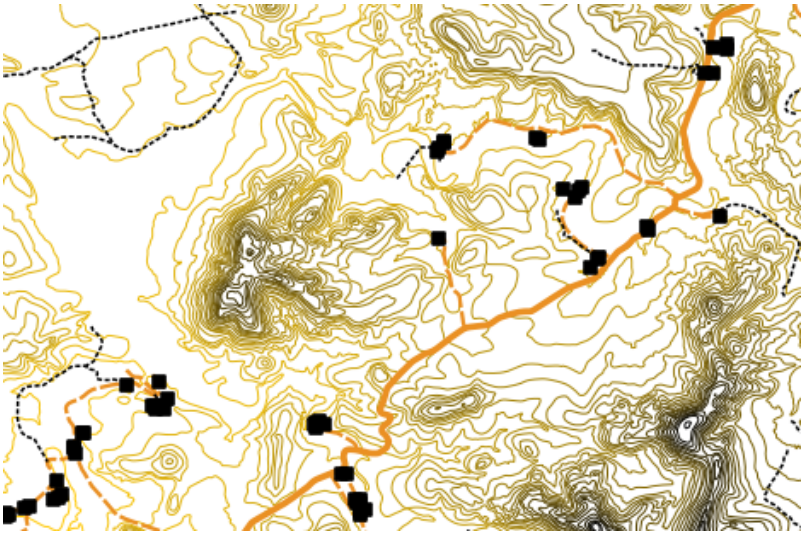


Fig. 4.12: A contour map drawn using continuous colour symbology

4.6 Unique Value Symbols

Sometimes the attributes of features are not numeric, but instead **strings** are used. 'String' is a computer term meaning a group of letters, numbers and other writing symbols. Strings attributes are often used to classify things by name. We can tell the GIS Application to give each unique string or number its own colour and symbol. Road features may have different classes (e.g. 'street', 'secondary road', 'main road' etc.), each drawn in the map view of the GIS with different colours or symbols. This is illustrated in Table 4.4: unique attribute values for a feature type (e.g. roads) can each have their own symbol.

Table 4.4: Unique values and symbols

Attribute Value	Colour class and symbol
Arterial route	
Main road	
Secondary road	
Street	

Within the GIS Application we can open/choose to use Unique Value symbology for a layer. The GIS will scan through all the different string values in the attribute field and build a list of unique strings or numbers. Each unique value can then be assigned a colour and style. This is shown in Fig. 4.13.

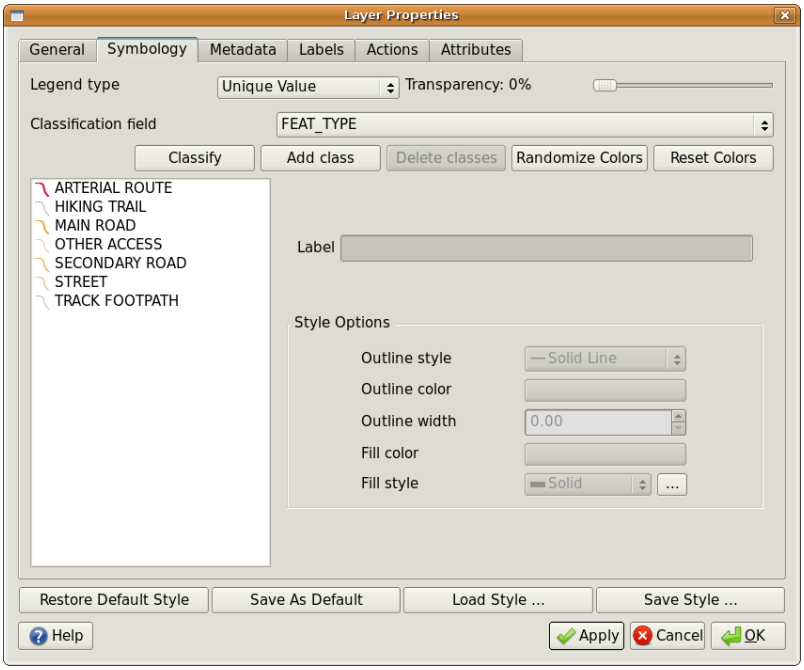


Fig. 4.13: Defining unique value symbology for roads based on the road type.

When the GIS draws the layer, it will look at the attributes of each feature before drawing it to the screen. Based on the value in the chosen field in the attribute table, the road line will be drawn with suitable colour and line style (and fill style if its a polygon feature). This is shown in Fig. 4.14.



Fig. 4.14: A roads vector layer symbolised using a unique value per road type.

4.7 Things to be aware of

Deciding which attributes and symbology to use requires some planning. Before you start collecting any **GeoSpatial** data, you should ensure you know what attributes are needed and how it will be symbolised. It is very difficult to go back and re-collect data if you plan poorly the first time around. Remember also that the goal of collecting attribute data is to allow you to analyse and interpret spatial information. How you do this depends on the questions you are trying to answer. Symbology is a visual language that allows people to see and understand your attribute data based on the colours and symbols you use. Because of this you should put a lot of thought into how you symbolise your maps in order to make them easy to understand.

4.8 What have we learned?

Let's wrap up what we covered in this worksheet:

- Vector features have **attributes**
- Attributes **describe** the **properties** of the feature
- The attributes are stored in a **table**
- Rows in the table are called **records**
- There is **one record per feature** in the vector layer
- Columns in the table are called **fields**
- Fields represent **properties** of the feature e.g. height, roof colour etc.
- Fields can contain **numerical**, **string** (any text) and **date** information
- The attribute data for a feature can be used to determine how it is **symbolised**
- **Graduated colour** symbology groups the data into discrete classes
- **Continuous colour** symbology assigns colours from a colour range to the features based on their attributes
- **Unique value** symbology associates each different value in the chosen attribute column with a different symbol (colour and style)
- If the attribute of a vector layer is not used to determine its symbology, it is drawn using a **single symbol** only

4.9 Now you try!

Here are some ideas for you to try with your learners:

- Using the table that you created in the last topic, add a new column for the symbology type you would use for each feature type and have the learners identify which symbology type they would use (see Table 4.5 for an example).
- Try to identify which symbology types you would use for the following types of vector features:
  - points showing pH level of soil samples taken around your school
  - lines showing a road network in a city
  - polygons for houses with an attribute that shows whether it is made of brick, wood or 'other' material.

Table 4.5: Example Feature Types and Symbology

Real world feature	Geometry Type	Symbology Type
The school flagpole	Point	Single Symbol
The soccer field	Polygon	Single Symbol
The footpaths in and around the school	Polyline	Have your learners count the number of learners using each footpath in the hour before school and then use **graduated symbols** to show the popularity of each footpath
Places where taps are located	Point	Single symbol
Classrooms	Polygon	**Unique value** based on the grade of the learners in the classroom
Fence	Polyline	Have your learners rate the condition of the fence around your school by separating it into sections and grading each section on a scale of 1-9 based on its condition. Use **graduated symbols** to classify the condition attribute.
Classrooms	Polygon	Count the number of learners in each classroom and use a **continuous colour symbol** to define a range of colours from red to blue.

4.10 Something to think about

If you don't have a computer available, you can use transparency sheets and a 1:50 000 map sheet to experiment with different symbology types. For example place a transparency sheet over the map and using different coloured koki pens, draw in red all contour lines below 900 m (or similar) and in green all lines above or equal to 900 m. Can you think of how to reproduce other symbology types using the same technique?

4.11 Further reading

Website: https://en.wikipedia.org/wiki/Cartography

The QGIS User Guide also has more detailed information on working with attribute data and symbology in QGIS.

4.12 What's next?

In the section that follows we will take a closer look at **data capture.** We will put the things we have learned about vector data and attributes into practice by creating new data.

CHAPTER

FIVE

DATA CAPTURE



Objectives:	Learn how to create and edit vector and attribute data.
Keywords:	Editing, data capture, heads-up, table, database.

5.1 Overview

In the previous two topics we looked at vector data. We saw that there are two key concepts to vector data, namely: **geometry** and **attributes**. The geometry of a vector feature describes its **shape** and **position**, while the **attributes** of a vector feature describe its **properties** (colour, size, age etc.).

In this section we will look more closely at the process of creating and editing vector data — both the geometry and attributes of vector features.

5.2 How does GIS digital data get stored?

Word processors, spreadsheets and graphics packages are all programs that let you create and edit digital data. Each type of application saves its data into a particular file format. For example, a graphics program will let you save your drawing as a `.jpg` JPEG image, word processors let you save your document as an `.odt` OpenDocument or `.doc` Word Document, and so on.

Just like these other applications, GIS Applications can store their data in files on the computer hard disk. There are a number of different file formats for GIS data, but the most common one is probably the 'shape file'. The name is a little odd in that although we call it a shape file (singular), it actually consists of at least three different files that work together to store your digital vector data, as shown in Table 5.1.

Table 5.1: The Basic Files that make up a 'Shapefile'

Extension	Description
`.shp`	The geometry of vector features are stored in this file
`.dbf`	The attributes of vector features are stored in this file
`.shx`	This file is an index that helps the GIS Application to find features more quickly.

When you look at the files that make up a shapefile on the computer hard disk, you will see something like Fig. 5.1. If you want to share vector data stored in shapefiles with another person, it is important to give them all of the files for that layer. So in the case of the trees layer shown in Fig. 5.1, you would need to give the person `trees.shp`, `trees.shx`, `trees.dbf`, `trees.prj` and `trees.qml`.

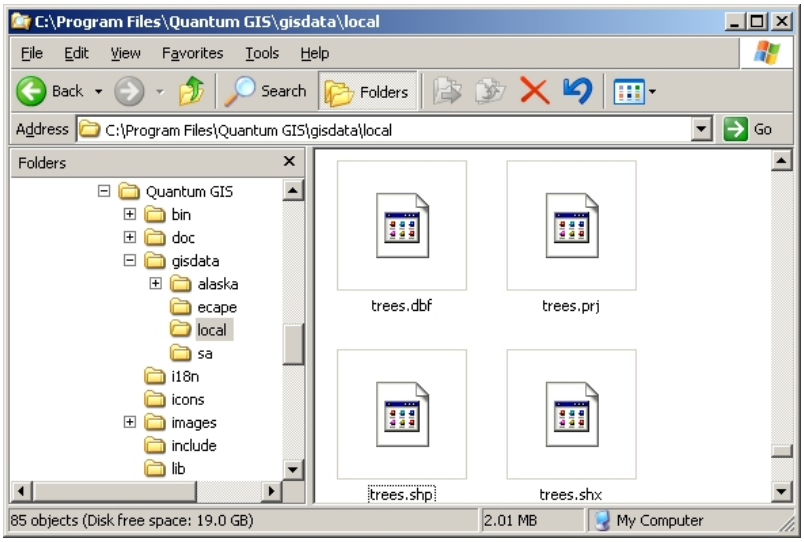


Fig. 5.1: The files that make up a 'trees' shapefile as seen in the computer's file manager.

Many GIS Applications are also able to store digital data inside a **database**. In general storing GIS data in a database is a good solution because the database can store **large amounts** of data **efficiently** and can provide data to the GIS Application quickly. Using a database also allows many people to work with the same vector data layers at the same time. Setting up a database to store GIS data is more complicated than using shapefiles, so for this topic we will focus on creating and editing shapefiles.

5.3 Planning before you begin

Before you can create a new vector layer (which will be stored in a shapefile), you need know what the geometry of that layer will be (point, polyline or polygon), and you need to know what the attributes of that layer will be. Let's look at a few examples and it will become clearer how to go about doing this.

5.3.1 Example 1: Creating a tourism map

Imagine that you want to create a nice tourism map for your local area. Your vision of the final map is a 1:50 000 toposheet with markers overlaid for sites of interest to tourists. First, let's think about the geometry. We know that we can represent a vector layer using point, polyline or polygon features. Which one makes the most sense for our tourism map? We could use points if we wanted to mark specific locations such as look out points, memorials, battle sites and so on. If we wanted to take tourists along a route, such as a scenic route through a mountain pass, it might make sense to use polylines. If we have whole areas that are of tourism interest, such as a nature reserve or a cultural village, polygons might make a good choice.

As you can see it's often not easy to know what type of geometry you will need. One common approach to this problem is to make one layer for each geometry type you need. So, for example, if you look at digital data provided by the Chief Directorate: Surveys and Mapping, South Africa, they provide a river areas (polygons) layer and a rivers polyline layer. They use the river areas (polygons) to represent river stretches that are wide, and they use river polylines to represent narrow stretches of river. In Fig. 5.2 we can see how our tourism layers might look on a map if we used all three geometry types.

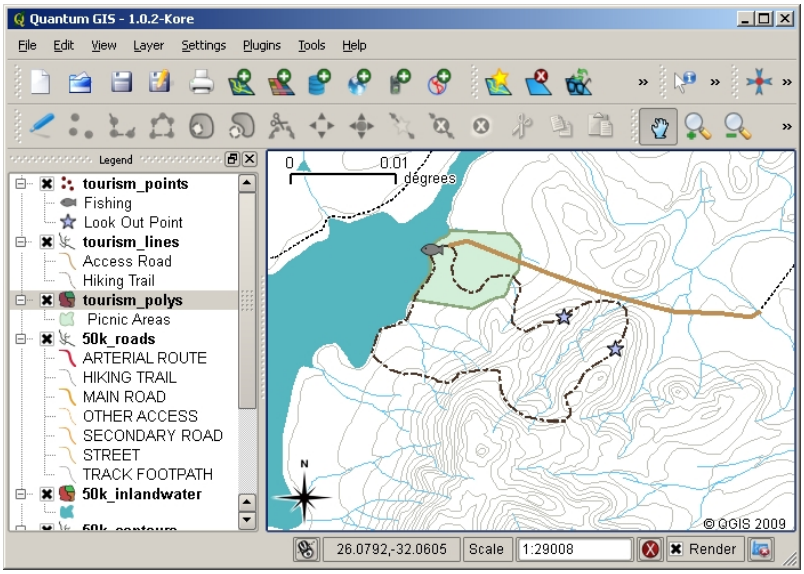


Fig. 5.2: A map with tourism layers. We have used three different geometry types for tourism data so that we can properly represent the different kinds of features needed for our visitors, giving them all the information they need.

5.3.2 Example 2: Creating a map of pollution levels along a river

If you wanted to measure pollution levels along the course of a river you would typically travel along the river in a boat or walk along its banks. At regular intervals you would stop and take various measurements such as Dissolved Oxygen (DO) levels, Coliform Bacteria (CB) counts, Turbidity levels and pH. You would also need to make a map reading of your position or obtain your position using a GPS receiver.

To store the data collected from an exercise like this in a GIS Application, you would probably create a GIS layer with a point geometry. Using point geometry makes sense here because each sample taken represents the conditions at a very specific place.

For the attributes we would want a **field** for each thing that describes the sample site. So we may end up with an attribute table that looks something like Table 5.2.

Table 5.2: River Attributes

SampleNo	pH	DO	CB	Turbidity	Collector	Date
1	7	6	N	Low	Patience	12/01/2009
2	6.8	5	Y	Medium	Thabo	12/01/2009
3	6.9	6	Y	High	Victor	12/01/2009

Drawing a table like Table 5.2 before you create your vector layer will let you decide what attribute fields (columns) you will need. Note that the geometry (positions where samples were taken) is not shown in the attribute table — the GIS Application stores it separately!

5.4 Creating an empty shapefile

Once you have planned what features you want to capture into the GIS, and the geometry type and attributes that each feature should have, you can move on to the next step of creating an empty shapefile.

The process usually starts with choosing the 'new vector layer' option in your GIS Application and then selecting a geometry type (see Fig. 5.3). As we covered in an earlier topic, this means choosing either point, polyline or polygon for the geometry.

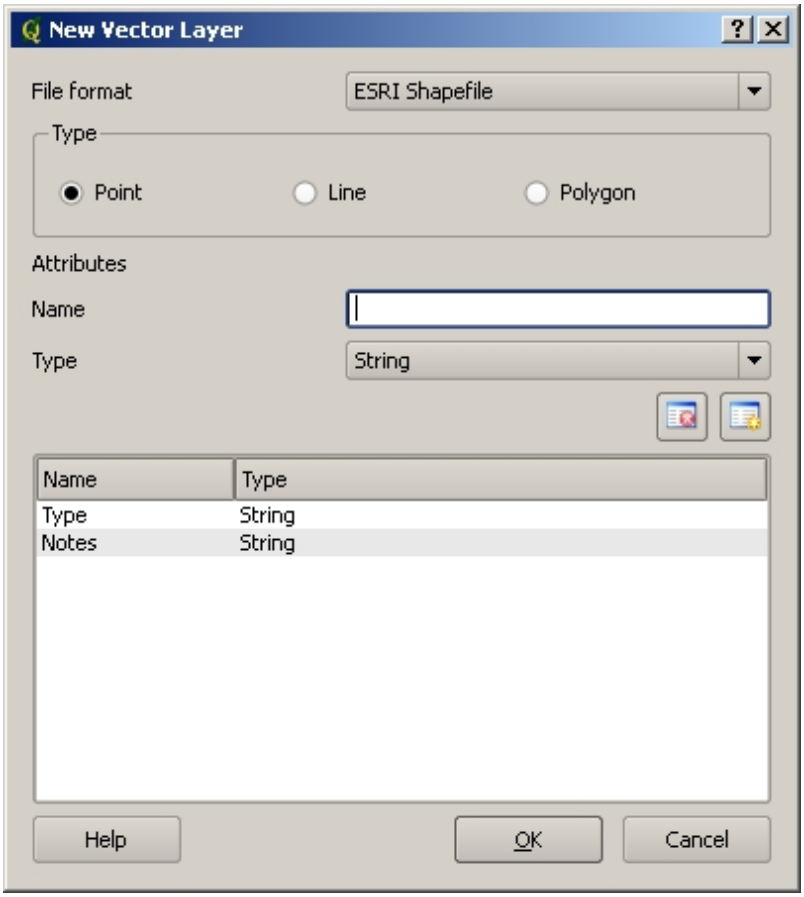


Fig. 5.3: Creating a new vector layer is as simple as filling in a few details in a form. First you choose the geometry type, and then you add the attribute fields.

Next you will add fields to the attribute table. Normally we give field names that are short, have no spaces and indicate what type of information is being stored in that field. Example field names may be 'pH', 'RoofColour', 'RoadType' and so on. As well as choosing a name for each field, you need to indicate how the information should be stored in that field — i.e. is it a number, a word or a sentence, or a date?

Computer programs usually call information that is made up of words or sentences '**strings**', so if you need to store something like a street name or the name of a river, you should use 'String' for the field type.

The shapefile format allows you to store the numeric field information as either a whole number (**integer**) or a decimal number (**floating point**) — so you need to think before hand whether the numeric data you are going to capture will have decimal places or not.

The final step (as shown in Fig. 5.4) for creating a shapefile is to give it a name and a place on the computer hard disk where it should be created. Once again it is a good idea to give the shapefile a short and meaningful name. Good examples are 'rivers', 'watersamples' and so on.

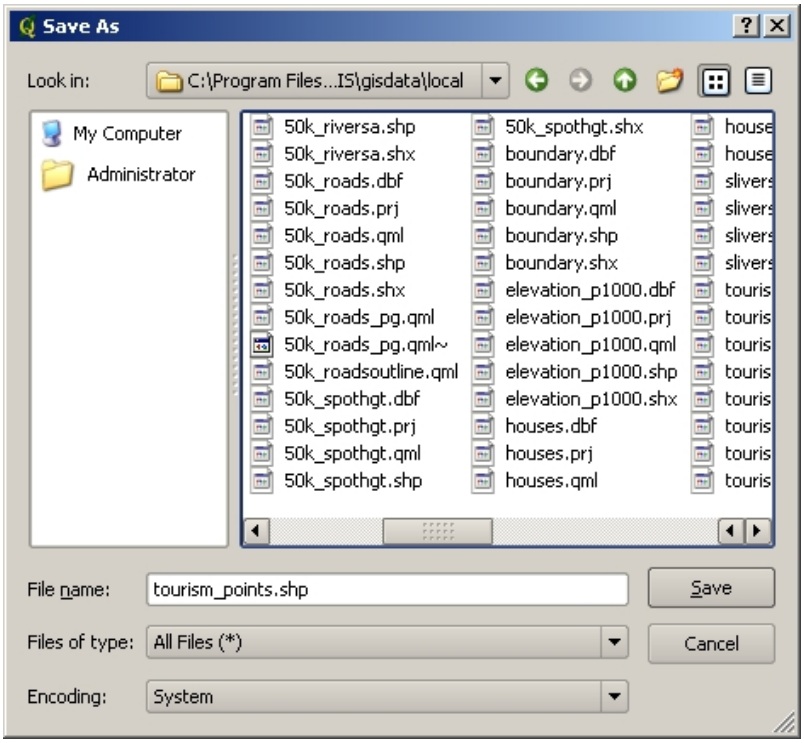


Fig. 5.4: After defining our new layer's geometry and attributes, we need to save it to disk. It is important to give a short but meaningful name to your shapefile.

Let's recap the process again quickly. To create a shapefile you first say what kind of geometry it will hold, then you create one or more fields for the attribute table, and then you save the shapefile to the hard disk using an easy to recognise name. Easy as 1-2-3!

5.5 Adding data to your shapefile

So far we have only created an empty shapefile. Now we need to enable editing in the shapefile using the 'enable editing' menu option or tool bar icon in the GIS Application. Shapefiles are not enabled for editing by default to prevent accidentally changing or deleting the data they contain. Next we need to start adding data. There are two steps we need to complete for each record we add to the shapefile:

1. Capturing geometry
2. Entering attributes

The process of capturing geometry is different for points, polylines and polygons.

To **capture a point**, you first use the map pan and zoom tools to get to the correct geographical area that you are going to be recording data for. Next you will need to enable the point capture tool. Having done that, the next place you click with the **left mouse button** in the map view, is where you want your new point **geometry** to appear. After you click on the map, a window will appear and you can enter all of the **attribute data** for that point (see Fig. 5.5). If you are unsure of the data for a given field you can usually leave it blank, but be aware that if you leave a lot of fields blank it will be hard to make a useful map from your data!

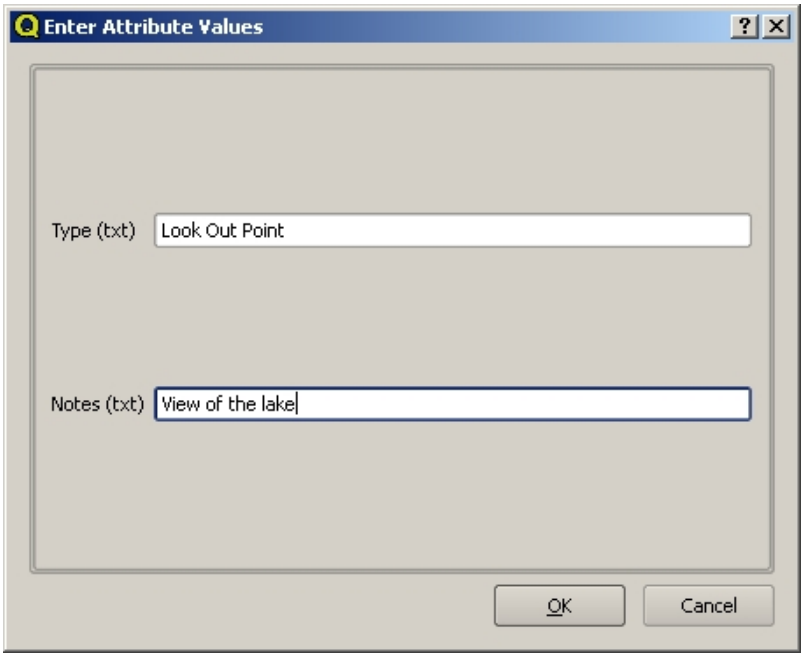


Fig. 5.5: After you have captured the point geometry, you will be asked to describe its attributes. The attribute form is based on the fields you specified when you created the vector layer.

To **capture a polyline** the process is similar to that of a point, in that you need to first use the pan and zoom tools to move the map in the map view to the correct geographical area. You should be zoomed in enough so that your new vector polyline feature will have an appropriate scale (see *Vector Data* for more details on scale issues). When you are ready, you can click the polyline capture icon in the tool bar and then start drawing your line by clicking on the map. After you make your first click, you will notice that the line stretches like an elastic band to follow the mouse cursor around as you move it. Each time you click with the **left mouse button**, a new vertex will be added to the map. This process is shown in Fig. 5.6.

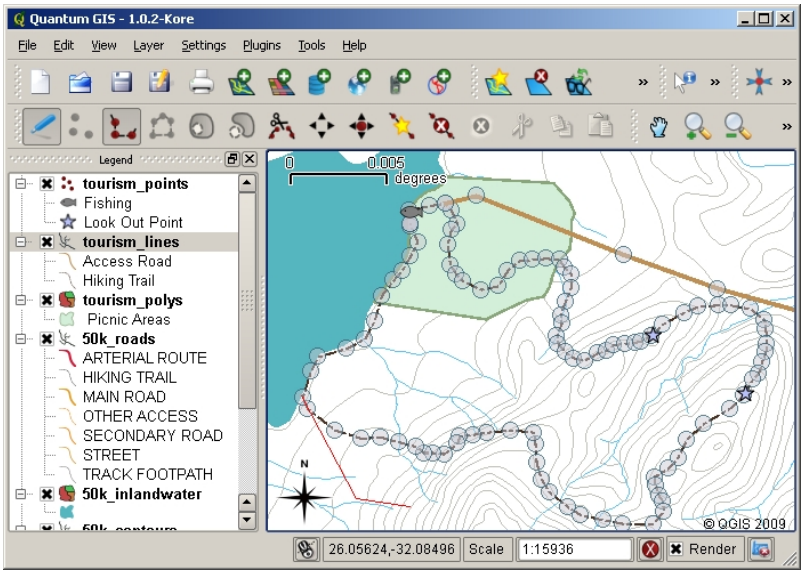


Fig. 5.6: Capturing lines for a tourism map. When editing a line layer, the vertices are shown with circular markers which you can move about with the mouse to adjust the line's geometry. When adding a new line (shown in red), each click of the mouse will add a new vertex.

When you have finished defining your line, use the **right mouse button** to tell the GIS Application that you have completed your edits. As with the procedure for capturing a point feature, you will then be asked to enter in the

attribute data for your new polyline feature.

The process for **capturing a polygon** is almost the same as capturing a polyline except that you need to use the polygon capture tool in the toolbar. Also, you will notice that when you draw your geometry on the screen, the GIS Application always creates an enclosed area.

To add a new feature after you have created your first one, you can simply click again on the map with the point, polyline or polygon capture tool active and start to draw your next feature.

When you have no more features to add, always be sure to click the 'allow editing' icon to toggle it off. The GIS Application will then save your newly created layer to the hard disk.

5.6 Heads-up digitising

As you have probably discovered by now if you followed the steps above, it is pretty hard to draw the features so that they are **spatially correct** if you do not have other features that you can use as a point of reference. One common solution to this problem is to use a raster layer (such as an aerial photograph or a satellite image) as a backdrop layer. You can then use this layer as a reference map, or even trace the features off the raster layer into your vector layer if they are visible. This process is known as 'heads-up digitising' and is shown in Fig. 5.7.

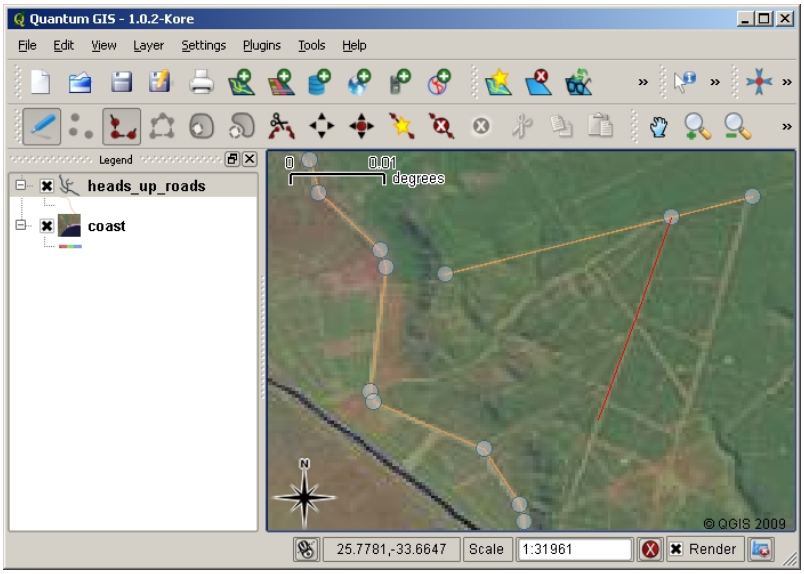


Fig. 5.7: Heads-up digitising using a satellite image as a backdrop. The image is used as a reference for capturing polyline features by tracing over them.

5.7 Digitising using a digitising table

Another method of capturing vector data is to use a digitising table. This approach is less commonly used except by GIS professionals, and it requires expensive equipment. The process of using a digitising table, is to place a paper map on the table. The paper map is held securely in place using clips. Then a special device called a 'puck' is used to trace features from the map. Tiny cross-hairs in the puck are used to ensure that lines and points are drawn accurately. The puck is connected to a computer and each feature that is captured using the puck gets stored in the computer's memory. You can see what a digitising puck looks like in Fig. 5.8.



Fig. 5.8: A digitising table and puck are used by GIS professionals when they want to digitise features from existing maps.

5.8 After your features are digitised...

Once your features are digitised, you can use the techniques you learned in the previous topic to set the symbology for your layer. Choosing an appropriate symbology will allow you to better understand the data you have captured when you look at the map.

5.9 Common problems / things to be aware of

If you are digitising using a backdrop raster layer such as an aerial photograph or satellite image, it is very important that the raster layer is properly georeferenced. A layer that is georeferenced properly displays in the correct position in the map view based on the GIS Application's internal model of the Earth. We can see the effect of a poorly georeferenced image in Fig. 5.9.

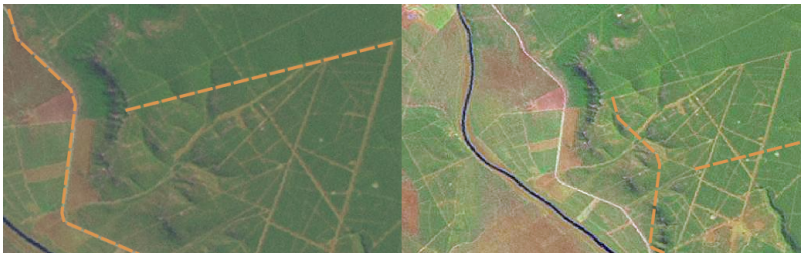


Fig. 5.9: The importance of using properly georeferenced raster images for heads-up digitising. On the left we can see the image is properly georegistered and the road features (in orange) overlap perfectly. If the image is poorly georeferenced (as shown on the right) the features will not be well aligned. Worse still, if the image on the right is used as a reference when capturing new features, the newly captured data will be inaccurate!

Also remember that it is important that you are zoomed in to an appropriate scale so that the vector features you create are useful. As we saw in the previous topic on vector geometry, it is a bad idea to digitise your data when you are zoomed out to a scale of 1:1000 000 if you intend to use the data you capture at a scale of 1:50 000 later.

5.10 What have we learned?

Let's wrap up what we covered in this worksheet:

- **Digitising** is the process of capturing knowledge of a feature's **geometry** and **attributes** into a **digital format** stored on the computer's disk.
- GIS Data can be stored in a **database** or as **files**.
- One commonly used file format is the **shapefile** which is actually a group of three or more files (`.shp`, `.dbf` and `.shx`).
- Before you create a new vector layer you need to plan both what **geometry** type and **attribute** fields it will contain.
- Geometry can be point, polyline or polygon.
- Attributes can be **integers** (whole numbers), **floating points** (decimal numbers), **strings** (words) or **dates**.
- The digitising process consists of **drawing** the geometry in the map view and then entering its attributes. This is repeated for each feature.
- **Heads-up digitising** is often used to provide orientation during digitising by using a raster image in the background.
- Professional GIS users sometimes use a **digitising table** to capture information from paper maps.

5.11 Now you try!

Here are some ideas for you to try with your learners:

- Draw up a list of features in and around your school that you think would be interesting to capture. For example: the school boundary, the position of fire assembly points, the layout of each class room, and so on. Try to use a mix of different geometry types. Now split your learners into groups and assign each group a few features to capture. Have them symbolise their layers so that they are more meaningful to look at. Combine the layers from all the groups to create a nice map of your school and its surroundings!
- Find a local river and take water samples along its length. Make a careful note of the position of each sample using a GPS or by marking it on a toposheet. For each sample take measurements such as pH, dissolved oxygen etc. Capture the data using the GIS application and make maps that show the samples with a suitable symbology. Could you identify any areas of concern? Was the GIS Application able to help you to identify these areas?

5.12 Something to think about

If you don't have a computer available, you can follow the same process by using transparency sheets and a notebook. Use an aerial photo, orthosheet or satellite image printout as your background layer. Draw columns down the page in your notebook and write in the column headings for each attribute field you want to store information about. Now trace the geometry of features onto the transparency sheet, writing a number next to each feature so that it can be identified. Now write the same number in the first column in your table in your notebook, and then fill in all the additional information you want to record.

5.13 Further reading

The QGIS User Guide has more detailed information on digitising vector data in QGIS.

5.14 What's next?

In the section that follows we will take a closer look at **raster data** to learn all about how image data can be used in a GIS.

CHAPTER

SIX

RASTER DATA



Objectives:	Understand what raster data is and how it can be used in a GIS.
Keywords:	Raster, Pixel, Remote Sensing, Satellite, Image, Georeference

6.1 Overview

In the previous topics we have taken a closer look at vector data. While vector features use geometry (points, polylines and polygons) to represent the real world, raster data takes a different approach. Rasters are made up of a matrix of pixels (also called cells), each containing a value that represents the conditions for the area covered by that cell (see Fig. 6.1). In this topic we are going to take a closer look at raster data, when it is useful and when it makes more sense to use vector data.

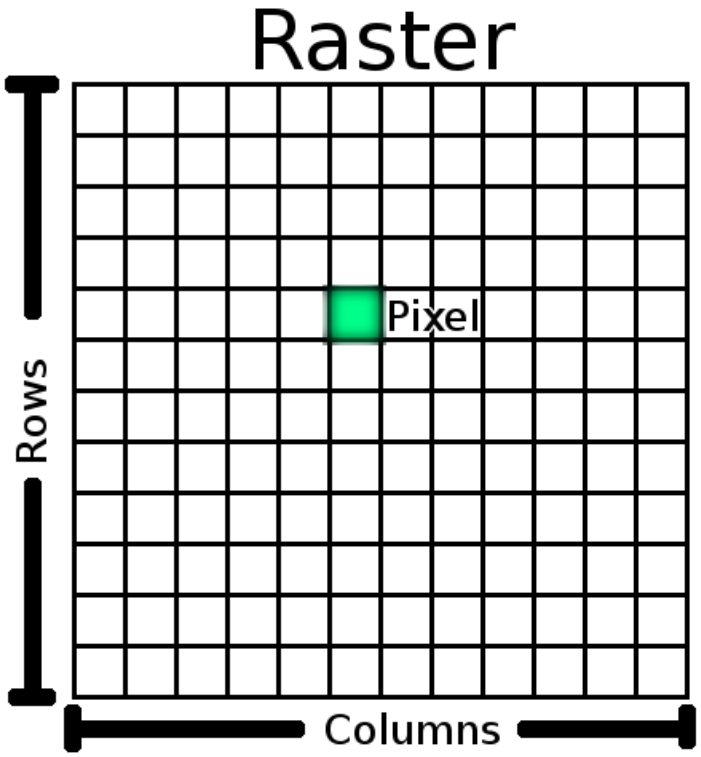


Fig. 6.1: A raster dataset is composed of rows (running across) and columns (running down) of pixels (also know as cells). Each pixel represents a geographical region, and the value in that pixel represents some characteristic of that region.

6.2 Raster data in detail

Raster data is used in a GIS application when we want to display information that is continuous across an area and cannot easily be divided into vector features. When we introduced you to vector data we showed you the image in Fig. 6.2. Point, polyline and polygon features work well for representing some features on this landscape, such as trees, roads and building footprints. Other features on a landscape can be more difficult to represent using vector features. For example the grasslands shown have many variations in colour and density of cover. It would be easy enough to make a single polygon around each grassland area, but a lot of the information about the grassland would be lost in the process of simplifying the features to a single polygon. This is because when you give a vector feature attribute values, they apply to the whole feature, so vectors aren't very good at representing features that are not homogeneous (entirely the same) all over. Another approach you could take is to digitise every small variation of grass colour and cover as a separate polygon. The problem with that approach is that it will take a huge amount of work in order to create a good vector dataset.



Fig. 6.2: Some features on a landscape are easy to represent as points, polylines and polygons (e.g. trees, roads, houses). In other cases it can be difficult. For example how would you represent the grasslands? As polygons? What about the variations in colour you can see in the grass? When you are trying to represent large areas with continuously changing values, raster data can be a better choice.

Using raster data is a solution to these problems. Many people use raster data as a **backdrop** to be used behind vector layers in order to provide more meaning to the vector information. The human eye is very good at interpreting images and so using an image behind vector layers, results in maps with a lot more meaning. Raster data is not only good for images that depict the real world surface (e.g. satellite images and aerial photographs), they are also good for representing more abstract ideas. For example, rasters can be used to show rainfall trends over an area, or to depict the fire risk on a landscape. In these kinds of applications, each cell in the raster represents a different value e.g. risk of fire on a scale of one to ten.

An example that shows the difference between an image obtained from a satellite and one that shows calculated values can be seen in Fig. 6.3.

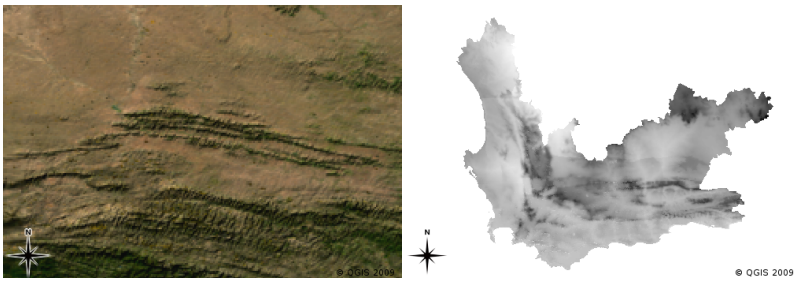


Fig. 6.3: True colour raster images (left) are useful as they provide a lot of detail that is hard to capture as vector features but easy to see when looking at the raster image. Raster data can also be non-photographic data such as the raster layer shown on the right which shows the calculated average minimum temperature in the Western Cape for the month of March.

6.3 Georeferencing

Georeferencing is the process of defining exactly where on the earth's surface an image or raster dataset was created. This positional information is stored with the digital version of the aerial photo. When the GIS application opens the photo, it uses the positional information to ensure that the photo appears in the correct place on the map. Normally this positional information consists of a coordinate for the top left pixel in the image, the size of each pixel in the X direction, the size of each pixel in the Y direction, and the amount (if any) by which the image is rotated. With these few pieces of information, the GIS application can ensure that raster data are displayed in the correct place. The georeferencing information for a raster is often provided in a small text file accompanying the raster.

6.4 Sources of raster data

Raster data can be obtained in a number of ways. Two of the most common ways are aerial photography and satellite imagery. In aerial photography, an aeroplane flies over an area with a camera mounted underneath it. The photographs are then imported into a computer and georeferenced. Satellite imagery is created when satellites orbiting the earth point special digital cameras towards the earth and then take an image of the area on earth they are passing over. Once the image has been taken it is sent back to earth using radio signals to special receiving stations such as the one shown in Fig. 6.4. The process of capturing raster data from an aeroplane or satellite is called **remote sensing**.



Fig. 6.4: The CSIR Satellite Applications Center at Hartebeeshoek near Johannesburg. Special antennae track satellites as they pass overhead and download images using radio waves.

In other cases, raster data can be computed. For example an insurance company may take police crime incident reports and create a country wide raster map showing how high the incidence of crime is likely to be in each area. Meteorologists (people who study weather patterns) might generate a province level raster showing average temperature, rainfall and wind direction using data collected from weather stations (see Fig. 6.4). In these cases, they will often use raster analysis techniques such as interpolation (which we describe in Topic *Spatial Analysis (Interpolation)*).

Sometimes raster data are created from vector data because the data owners want to share the data in an easy to use format. For example, a company with road, rail, cadastral and other vector datasets may choose to generate a raster version of these datasets so that employees can view these datasets in a web browser. This is normally only useful if the attributes, that users need to be aware of, can be represented on the map with labels or symbology. If the user

needs to look at the attribute table for the data, providing it in raster format could be a bad choice because raster layers do not usually have any attribute data associated with them.

6.5 Spatial Resolution

Every raster layer in a GIS has pixels (cells) of a fixed size that determine its spatial resolution. This becomes apparent when you look at an image at a small scale (see Fig. 6.5) and then zoom in to a large scale (see Fig. 6.6).



Fig. 6.5: This satellite image looks good when using a small scale…



Fig. 6.6: …but when viewed at a large scale you can see the individual pixels that the image is composed of.

Several factors determine the spatial resolution of an image. For remote sensing data, spatial resolution is usually determined by the capabilities of the sensor used to take an image. For example SPOT5 satellites can take images where each pixel is 10 m x 10 m. Other satellites, for example MODIS take images only at 500 m x 500 m per pixel. In aerial photography, pixel sizes of 50 cm x 50 cm are not uncommon. Images with a pixel size covering a small area are called '**high resolution**' images because it is possible to make out a high degree of detail in the image. Images with a pixel size covering a large area are called '**low resolution**' images because the amount of detail the images show is low.

In raster data that is computed by spatial analysis (such as the rainfall map we mentioned earlier), the spatial density of information used to create the raster will usually determine the spatial resolution. For example if you want to create a high resolution average rainfall map, you would ideally need many weather stations in close proximity to each other.

One of the main things to be aware of with rasters captured at a high spatial resolution is storage requirements. Think of a raster that is 3 x 3 pixels, each of which contains a number representing average rainfall. To store all the information contained in the raster, you will need to store 9 numbers in the computer's memory. Now imagine you want to have a raster layer for the whole of South Africa with pixels of 1 km x 1 km. South Africa is around 1,219,090 km ^{2}. Which means your computer would need to store over a million numbers on its hard disk in order to hold all of the information. Making the pixel size smaller would greatly increase the amount of storage needed.

Sometimes using a low spatial resolution is useful when you want to work with a large area and are not interested in looking at any one area in a lot of detail. The cloud maps you see on the weather report, are an example of this — it's useful to see the clouds across the whole country. Zooming in to one particular cloud in high resolution will not tell you very much about the upcoming weather!

On the other hand, using low resolution raster data can be problematic if you are interested in a small region because you probably won't be able to make out any individual features from the image.

6.6 Spectral resolution

If you take a colour photograph with a digital camera or camera on a cellphone, the camera uses electronic sensors to detect red, green and blue light. When the picture is displayed on a screen or printed out, the red, green and blue (RGB) information is combined to show you an image that your eyes can interpret. While the information is still in digital format though, this RGB information is stored in separate colour **bands**.

Whilst our eyes can only see RGB wavelengths, the electronic sensors in cameras are able to detect wavelengths that our eyes cannot. Of course in a hand held camera it probably doesn't make sense to record information from the **non-visible** parts of the spectrum since most people just want to look at pictures of their dog or what have you. Raster images that include data for non-visible parts of the light spectrum are often referred to as multi-spectral images. In GIS recording the non-visible parts of the spectrum can be very useful. For example, measuring infra-red light can be useful in identifying water bodies.

Because having images containing multiple bands of light is so useful in GIS, raster data are often provided as multi-band images. Each band in the image is like a separate layer. The GIS will combine three of the bands and show them as red, green and blue so that the human eye can see them. The number of bands in a raster image is referred to as its **spectral resolution**.

If an image consists of only one band, it is often called a **grayscale** image. With grayscale images, you can apply false colouring to make the differences in values in the pixels more obvious. Images with false colouring applied are often referred to as **pseudocolour images**.

6.7 Raster to vector conversion

In our discussion of vector data, we explained that often raster data are used as a backdrop layer, which is then used as a base from which vector features can be digitised.

Another approach is to use advanced computer programs to automatically extract vector features from images. Some features such as roads show in an image as a sudden change of colour from neighbouring pixels. The computer program looks for such colour changes and creates vector features as a result. This kind of functionality is normally only available in very specialised (and often expensive) GIS software.

6.8 Vector to raster conversion

Sometimes it is useful to convert vector data into raster data. One side effect of this is that attribute data (that is attributes associated with the original vector data) will be lost when the conversion takes place. Having vectors converted to raster format can be useful though when you want to give GIS data to non GIS users. With the simpler raster formats, the person you give the raster image to can simply view it as an image on their computer without needing any special GIS software.

6.9 Raster analysis

There are a great many analytical tools that can be run on raster data which cannot be used with vector data. For example, rasters can be used to model water flow over the land surface. This information can be used to calculate where watersheds and stream networks exist, based on the terrain.

Raster data are also often used in agriculture and forestry to manage crop production. For example with a satellite image of a farmer's lands, you can identify areas where the plants are growing poorly and then use that information to apply more fertilizer on the affected areas only. Foresters use raster data to estimate how much timber can be harvested from an area.

Raster data is also very important for disaster management. Analysis of Digital Elevation Models (a kind of raster where each pixel contains the height above sea level) can then be used to identify areas that are likely to be flooded. This can then be used to target rescue and relief efforts to areas where it is needed the most.

6.10 Common problems / things to be aware of

As we have already mentioned, high resolution raster data can require large amounts of computer storage.

6.11 What have we learned?

Let's wrap up what we covered in this worksheet:

- Raster data are a grid of regularly sized **pixels.**
- Raster data are good for showing **continually varying information.**
- The size of pixels in a raster determines its **spatial resolution.**
- Raster images can contain one or more **bands**, each covering the same spatial area, but containing different information.
- When raster data contains bands from different parts of the electromagnetic spectrum, they are called **multi-spectral images**.
- Three of the bands of a multi-spectral image can be shown in the colours Red, Green and Blue so that we can see them.
- Images with a single band are called grayscale images.
- Single band, grayscale images can be shown in pseudocolour by the GIS.
- Raster images can consume a large amount of storage space.

6.12 Now you try!

Here are some ideas for you to try with your learners:

- Discuss with your learners in which situations you would use raster data and in which you would use vector data.
- Get your learners to create a raster map of your school by using A4 transparency sheets with grid lines drawn on them. Overlay the transparencies onto a toposheet or aerial photograph of your school. Now let each learner or group of learners colour in cells that represent a certain type of feature e.g. building, playground, sports field, trees, footpaths etc. When they are all finished, overlay all the sheets together and see if it makes a good raster map representation of your school. Which types of features worked well when represented as rasters? How did your choice in cell size affect your ability to represent different feature types?

6.13 Something to think about

If you don't have a computer available, you can understand raster data using pen and paper. Draw a grid of squares onto a sheet of paper to represent your soccer field. Fill the grid in with numbers representing values for grass cover on your soccer field. If a patch is bare give the cell a value of 0. If the patch is mixed bare and covered, give it a value of 1. If an area is completely covered with grass, give it a value of 2. Now use pencil crayons to colour the cells based on their values. Colour cells with value 2 dark green. Value 1 should get coloured light green, and value 0 coloured in brown. When you finish, you should have a raster map of your soccer field!

6.14 Further reading

Book:

- Chang, Kang-Tsung (2006). Introduction to Geographic Information Systems. 3rd Edition. McGraw Hill. ISBN: 0070658986
- DeMers, Michael N. (2005). Fundamentals of Geographic Information Systems. 3rd Edition. Wiley. ISBN: 9814126195

Website: https://en.wikipedia.org/wiki/GIS_file_formats#Raster_formats

The QGIS User Guide also has more detailed information on working with raster data in QGIS.

6.15 What's next?

In the section that follows we will take a closer look at **topology** to see how the relationship between vector features can be used to ensure the best data quality.

CHAPTER

SEVEN

TOPOLOGY

	Objectives:	Understanding topology in vector data
	Keywords:	Vector, topology, topology rules, topology errors, search radius, snapping distance, simple feature

7.1 Overview

Topology expresses the spatial relationships between connecting or adjacent vector features (points, polylines and polygons) in a GIS. Topological or topology-based data are useful for detecting and correcting digitising errors (e.g. two lines in a roads vector layer that do not meet perfectly at an intersection). Topology is necessary for carrying out some types of spatial analysis, such as network analysis.

Imagine you travel to London. On a sightseeing tour you plan to visit St. Paul's Cathedral first and in the afternoon Covent Garden Market for some souvenirs. Looking at the Underground map of London (see Fig. 7.1) you have to find connecting trains to get from Covent Garden to St. Paul's. This requires topological information (data) about where it is possible to change trains. Looking at a map of the underground, the topological relationships are illustrated by circles that show connectivity.

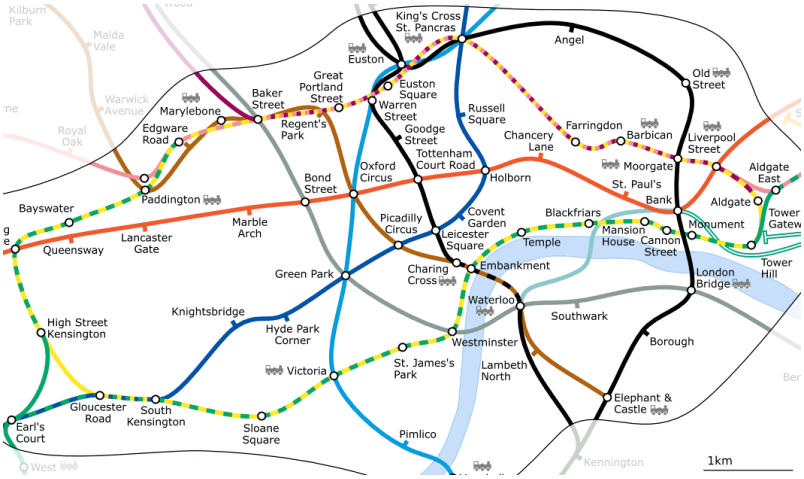


Fig. 7.1: Topology of London Underground Network.

7.2 Topology errors

There are different types of topological errors and they can be grouped according to whether the vector feature types are polygons or polylines. Topological errors with **polygon** features can include unclosed polygons, gaps between polygon borders or overlapping polygon borders. A common topological error with **polyline** features is that they do

not meet perfectly at a point (node). This type of error is called an **undershoot** if a gap exists between the lines, and an **overshoot** if a line ends beyond the line it should connect to (see Fig. 7.2).

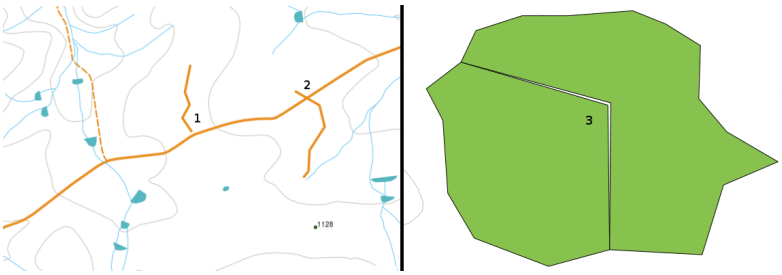


Fig. 7.2: Undershoots (1) occur when digitised vector lines that should connect to each other don't quite touch. Overshoots (2) happen if a line ends beyond the line it should connect to. Slivers (3) occur when the vertices of two polygons do not match up on their borders.

The result of overshoot and undershoot errors are so-called 'dangling nodes' at the end of the lines. Dangling nodes are acceptable in special cases, for example if they are attached to dead-end streets.

Topological errors break the relationship between features. These errors need to be fixed in order to be able to analyse vector data with procedures like network analysis (e.g. finding the best route across a road network) or measurement (e.g. finding out the length of a river). In addition to topology being useful for network analysis and measurement, there are other reasons why it is important and useful to create or have vector data with correct topology. Just imagine you digitise a municipal boundaries map for your province and the polygons overlap or show slivers. If such errors were present, you would be able to use the measurement tools, but the results you get will be incorrect. You will not know the correct area for any municipality and you will not be able to define exactly, where the borders between the municipalities are.

It is not only important for your own analysis to create and have topologically correct data, but also for people who you pass data on to. They will be expecting your data and analysis results to be correct!

7.3 Topology rules

Fortunately, many common errors that can occur when digitising vector features can be prevented by topology rules that are implemented in many GIS applications.

Except for some special GIS data formats, topology is usually not enforced by default. Many common GIS, like QGIS, define topology as relationship rules and let the user choose the rules, if any, to be implemented in a vector layer.

The following list shows some examples of where topology rules can be defined for real world features in a vector map:

- Area edges of a municipality map must not overlap.
- Area edges of a municipality map must not have gaps (slivers).
- Polygons showing property boundaries must be closed. Undershoots or overshoots of the border lines are not allowed.
- Contour lines in a vector line layer must not intersect (cross each other).

7.4 Topological tools

Many GIS applications provide tools for topological editing. For example in QGIS you can **enable topological editing** to improve editing and maintaining common boundaries in polygon layers. A GIS such as QGIS 'detects' a shared boundary in a polygon map so you only have to move the edge vertex of one polygon boundary and QGIS will ensure the updating of the other polygon boundaries as shown in Fig. 7.3 (1).

Another topological option allows you to prevent **polygon overlaps** during digitising (see Fig. 7.3 (2)). If you already have one polygon, it is possible with this option to digitise a second adjacent polygon so that both polygons overlap and QGIS then clips the second polygon to the common boundary.

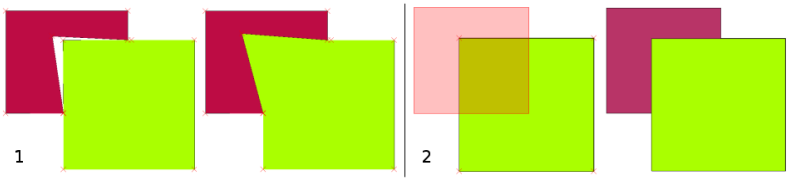


Fig. 7.3: (1) Topological editing to detect shared boundaries, when moving vertices. When moving a vertex, all features that share that vertex are updated. (2) To avoid polygon overlaps, when a new polygon is digitised (shown in red) it is clipped to avoid overlapping neighbouring areas.

7.5 Snapping distance

Snapping distance is the distance a GIS uses to search for the closest vertex and / or segment you are trying to connect when you digitise. A **segment** is a straight line formed between two vertices in a polygon or polyline geometry. If you aren't within the snapping distance, a GIS such as QGIS will leave the vertex where you release the mouse button, instead of snapping it to an existing vertex and / or segment (see Fig. 7.4).

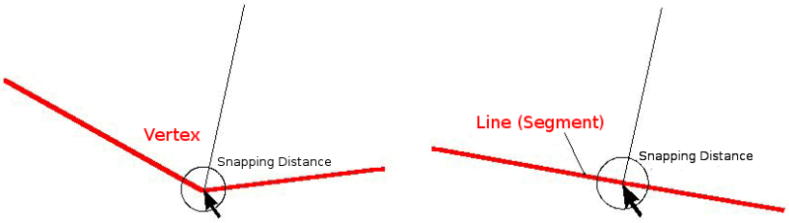


Fig. 7.4: The snapping distance (black circle) is defined in map units (e.g. decimal degrees) for snapping to either vertices or segments.

7.6 Search Radius

Search radius is the distance a GIS uses to search for the closest vertex you are trying to move when you click on the map. If you aren't within the search radius, the GIS won't find and select any vertex of a feature for editing. In principle, it is quite similar to the snapping distance functionality.

Snapping distance and search radius are both set in map units so you may need to experiment to get the distance value set right. If you specify a value that is too big, the GIS may snap to a wrong vertex, especially if you are dealing with a large number of vertices close together. If you specify the search radius too small the GIS application won't find any feature or vertex to move or edit.

7.7 Common problems / things to be aware of

Mainly designed for simplicity and for fast rendering but not for data analysis that requires topology (such as finding routes across a network), many GIS applications are able to show topological and simple feature data together and some can also create, edit and analyse both.

7.8 What have we learned?

Let's wrap up what we covered in this worksheet:

- **Topology** shows the spatial relation of neighbouring vector features.
- Topology in GIS is provided by **topological tools**.
- Topology can be used to **detect and correct digitizing errors**.
- For some tools, such as **network analysis**, topological data is essential.
- **Snapping distance** and **search radius** help us to digitise topologically correct vector data.
- **Simple feature** data is not a true topological data format but it is commonly used by GIS applications.

7.9 Now you try!

Here are some ideas for you to try with your learners:

- Mark your local bus stops on a toposheet map and then task your learners to find the shortest route between two stops.
- Think of how you would create vector features in a GIS to represent a topological road network of your town. What topological rules are important and what tools can your learners use in QGIS to make sure that the new road layer is topologically correct?

7.10 Something to think about

If you don't have a computer available, you can use a map of a bus or railway network and discuss the spatial relationships and topology with your learners.

7.11 Further reading

Books:

- Chang, Kang-Tsung (2006). Introduction to Geographic Information Systems. 3rd Edition. McGraw Hill. ISBN: 0070658986
- DeMers, Michael N. (2005). Fundamentals of Geographic Information Systems. 3rd Edition. Wiley. ISBN: 9814126195

Websites:

- http://www.innovativegis.com/basis/
- https://en.wikipedia.org/wiki/Geospatial_topology

The QGIS User Guide also has more detailed information on topological editing provided in QGIS.

7.12 What's next?

In the section that follows we will take a closer look at **Coordinate Reference Systems** to understand how we relate data from our spherical earth onto flat maps!

CHAPTER

EIGHT

COORDINATE REFERENCE SYSTEMS



Objectives:	Understanding of Coordinate Reference Systems.
Keywords:	Coordinate Reference System (CRS), Map Projection, On the Fly Projection, Latitude, Longitude, Northing, Easting

8.1 Overview

Map projections try to portray the surface of the earth, or a portion of the earth, on a flat piece of paper or computer screen. In layman's term, map projections try to transform the earth from its spherical shape (3D) to a planar shape (2D).

A **coordinate reference system** (CRS) then defines how the two-dimensional, projected map in your GIS relates to real places on the earth. The decision of which map projection and CRS to use depends on the regional extent of the area you want to work in, on the analysis you want to do, and often on the availability of data.

8.2 Map Projection in detail

A traditional method of representing the earth's shape is the use of globes. There is, however, a problem with this approach. Although globes preserve the majority of the earth's shape and illustrate the spatial configuration of continent-sized features, they are very difficult to carry in one's pocket. They are also only convenient to use at extremely small scales (e.g. 1:100 million).

Most of the thematic map data commonly used in GIS applications are of considerably larger scale. Typical GIS datasets have scales of 1:250 000 or greater, depending on the level of detail. A globe of this size would be difficult and expensive to produce and even more difficult to carry around. As a result, cartographers have developed a set of techniques called **map projections** designed to show, with reasonable accuracy, the spherical earth in two-dimensions.

When viewed at close range the earth appears to be relatively flat. However when viewed from space, we can see that the earth is relatively spherical. Maps, as we will see in the upcoming map production topic, are representations of reality. They are designed to not only represent features, but also their shape and spatial arrangement. Each map projection has **advantages** and **disadvantages**. The best projection for a map depends on the **scale** of the map, and on the purposes for which it will be used. For example, a projection may have unacceptable distortions if used to map the entire African continent, but may be an excellent choice for a **large-scale (detailed) map** of your country. The properties of a map projection may also influence some of the design features of the map. Some projections are good for small areas, some are good for mapping areas with a large East-West extent, and some are better for mapping areas with a large North-South extent.

8.3 The three families of map projections

The process of creating map projections is best illustrated by positioning a light source inside a transparent globe on which opaque earth features are placed. Then project the feature outlines onto a two-dimensional flat piece of paper. Different ways of projecting can be produced by surrounding the globe in a **cylindrical** fashion, as a **cone**, or even as a **flat surface**. Each of these methods produces what is called a **map projection family**. Therefore, there is a family of **planar projections**, a family of **cylindrical projections**, and another called **conical projections** (see Fig. 8.1)

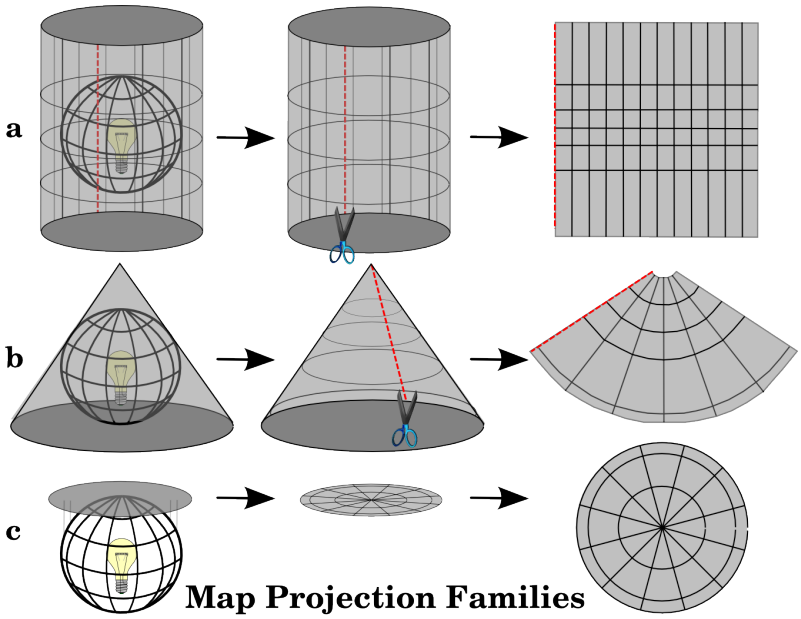


Fig. 8.1: The three families of map projections. They can be represented by a) cylindrical projections, b) conical projections or c) planar projections.

Today, of course, the process of projecting the spherical earth onto a flat piece of paper is done using the mathematical principles of geometry and trigonometry. This recreates the physical projection of light through the globe.

8.4 Accuracy of map projections

Map projections are never absolutely accurate representations of the spherical earth. As a result of the map projection process, every map shows **distortions of angular conformity, distance and area**. A map projection may combine several of these characteristics, or may be a compromise that distorts all the properties of area, distance and angular conformity, within some acceptable limit. Examples of compromise projections are the **Winkel Tripel projection** and the **Robinson projection** (see Fig. 8.2), which are often used for producing and visualizing world maps.

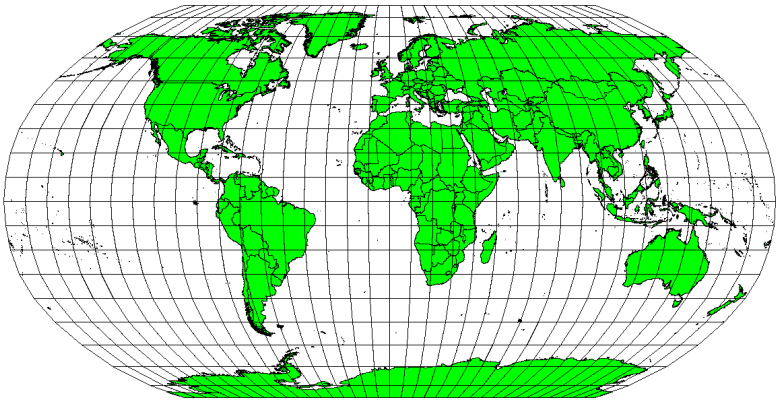


Fig. 8.2: The Robinson projection is a compromise where distortions of area, angular conformity and distance are acceptable.

It is usually impossible to preserve all characteristics at the same time in a map projection. This means that when you want to carry out accurate analytical operations, you need to use a map projection that provides the best characteristics for your analyses. For example, if you need to measure distances on your map, you should try to use a map projection for your data that provides high accuracy for distances.

8.4.1 Map projections with angular conformity

When working with a globe, the main directions of the compass rose (North, East, South and West) will always occur at 90 degrees to one another. In other words, East will always occur at a 90 degree angle to North. Maintaining correct **angular properties** can be preserved on a map projection as well. A map projection that retains this property of angular conformity is called a **conformal** or **orthomorphic projection**.

These projections are used when the **preservation of angular relationships** is important. They are commonly used for navigational or meteorological tasks. It is important to remember that maintaining true angles on a map is difficult for large areas and should be attempted only for small portions of the earth. The conformal type of projection results in distortions of areas, meaning that if area measurements are made on the map, they will be incorrect. The larger the area the less accurate the area measurements will be. Examples are the **Mercator projection** (as shown in Fig. 8.3) and the **Lambert Conformal Conic projection**. The U.S. Geological Survey uses a conformal projection for many of its topographic maps.

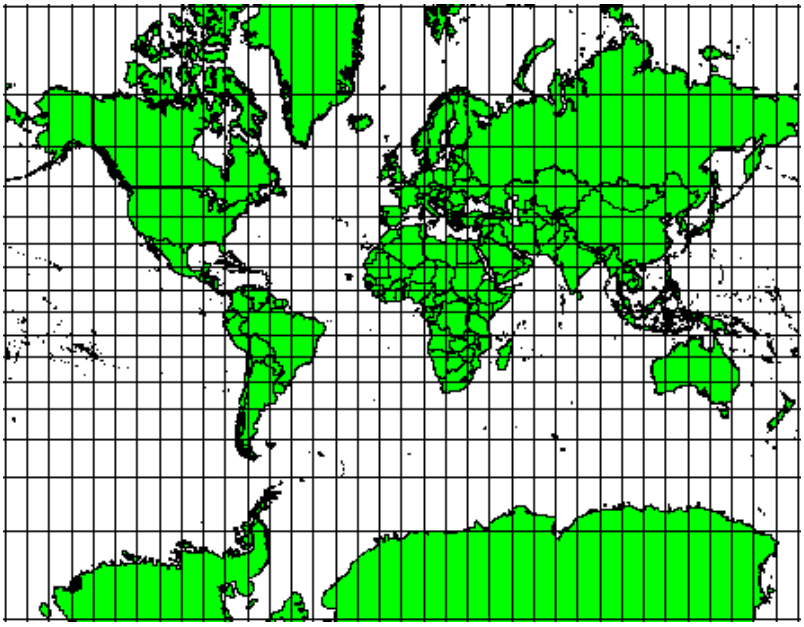


Fig. 8.3: The Mercator projection, for example, is used where angular relationships are important, but the relationship of areas are distorted.

8.4.2 Map projections with equal distance

If your goal in projecting a map is to accurately measure distances, you should select a projection that is designed to preserve distances well. Such projections, called **equidistant projections**, require that the **scale** of the map is **kept constant**. A map is equidistant when it correctly represents distances from the centre of the projection to any other place on the map. **Equidistant projections** maintain accurate distances from the centre of the projection or along given lines. These projections are used for radio and seismic mapping, and for navigation. The **Plate Carree Equidistant Cylindrical** (see Fig. 8.4) and the **Equirectangular projection** are two good examples of equidistant projections. The **Azimuthal Equidistant projection** is the projection used for the emblem of the United Nations (see Fig. 8.5).

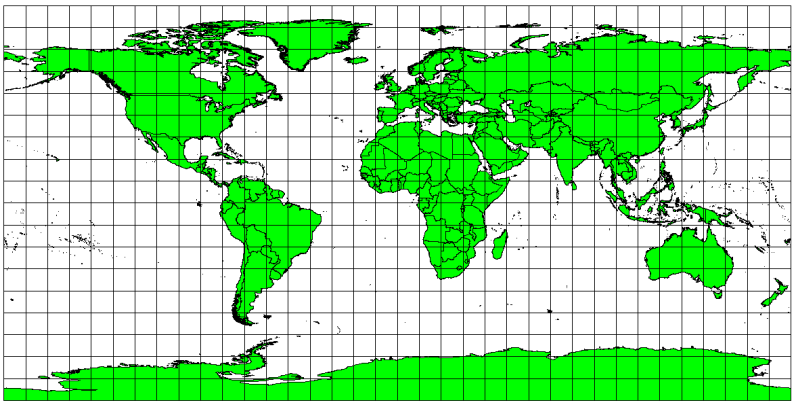


Fig. 8.4: The Plate Carree Equidistant Cylindrical projection, for example, is used when accurate distance measurement is important.



Fig. 8.5: The United Nations Logo uses the Azimuthal Equidistant projection.

8.4.3 Projections with equal areas

When a map portrays areas over the entire map, so that all mapped areas have the same proportional relationship to the areas on the Earth that they represent, the map is an **equal area map**. In practice, general reference and educational maps most often require the use of **equal area projections**. As the name implies, these maps are best used when calculations of area are the dominant calculations you will perform. If, for example, you are trying to analyse a particular area in your town to find out whether it is large enough for a new shopping mall, equal area projections are the best choice. On the one hand, the larger the area you are analysing, the more precise your area measures will be, if you use an equal area projection rather than another type. On the other hand, an equal area projection results in **distortions of angular conformity** when dealing with large areas. Small areas will be far less prone to having their angles distorted when you use an equal area projection. **Alber's equal area**, **Lambert's equal area** and **Mollweide Equal Area Cylindrical projections** (shown in Fig. 8.6) are types of equal area projections that are often encountered in GIS work.

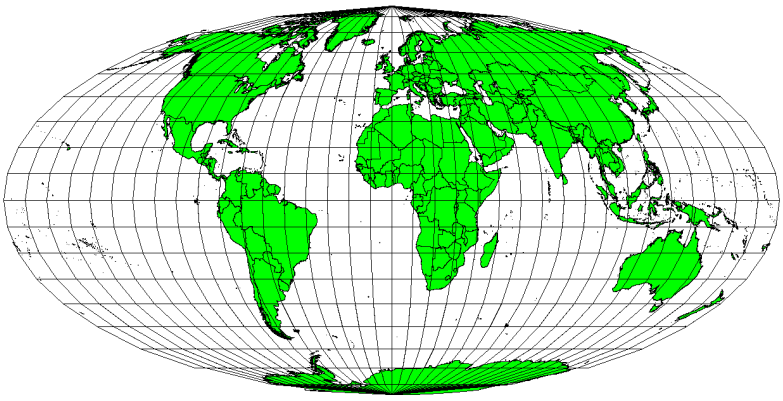


Fig. 8.6: The Mollweide Equal Area Cylindrical projection, for example, ensures that all mapped areas have the same proportional relationship to the areas on the Earth.

Keep in mind that map projection is a very complex topic. There are hundreds of different projections available world wide each trying to portray a certain portion of the earth's surface as faithfully as possible on a flat piece of

paper. In reality, the choice of which projection to use, will often be made for you. Most countries have commonly used projections and when data is exchanged people will follow the **national trend**.

8.5 Coordinate Reference System (CRS) in detail

With the help of coordinate reference systems (CRS) every place on the earth can be specified by a set of three numbers, called coordinates. In general CRS can be divided into **projected coordinate reference systems** (also called Cartesian or rectangular coordinate reference systems) and **geographic coordinate reference systems**.

8.5.1 Geographic Coordinate Systems

The use of Geographic Coordinate Reference Systems is very common. They use degrees of latitude and longitude and sometimes also a height value to describe a location on the earth's surface. The most popular is called **WGS 84**.

Lines of latitude run parallel to the equator and divide the earth into 180 equally spaced sections from North to South (or South to North). The reference line for latitude is the equator and each **hemisphere** is divided into ninety sections, each representing one degree of latitude. In the northern hemisphere, degrees of latitude are measured from zero at the equator to ninety at the north pole. In the southern hemisphere, degrees of latitude are measured from zero at the equator to ninety degrees at the south pole. To simplify the digitisation of maps, degrees of latitude in the southern hemisphere are often assigned negative values (0 to -90°). Wherever you are on the earth's surface, the distance between the lines of latitude is the same (60 nautical miles). See Fig. 8.7 for a pictorial view.

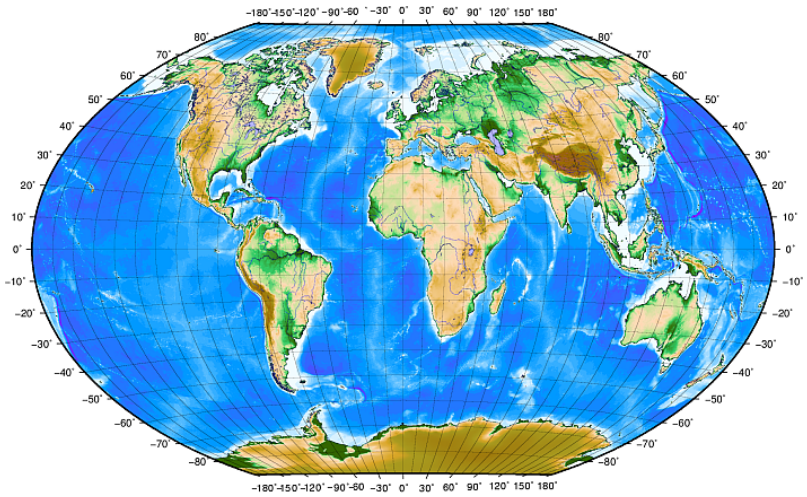


Fig. 8.7: Geographic coordinate system with lines of latitude parallel to the equator and lines of longitude with the prime meridian through Greenwich.

Lines of longitude, on the other hand, do not stand up so well to the standard of uniformity. Lines of longitude run perpendicular to the equator and converge at the poles. The reference line for longitude (the prime meridian) runs from the North pole to the South pole through Greenwich, England. Subsequent lines of longitude are measured from zero to 180 degrees East or West of the prime meridian. Note that values West of the prime meridian are assigned negative values for use in digital mapping applications. See Fig. 8.7 for a pictorial view.

At the equator, and only at the equator, the distance represented by one line of longitude is equal to the distance represented by one degree of latitude. As you move towards the poles, the distance between lines of longitude becomes progressively less, until, at the exact location of the pole, all 360° of longitude are represented by a single point that you could put your finger on (you probably would want to wear gloves though). Using the geographic coordinate system, we have a grid of lines dividing the earth into squares that cover approximately 12363.365 square kilometres at the equator — a good start, but not very useful for determining the location of anything within that square.

To be truly useful, a map grid must be divided into small enough sections so that they can be used to describe (with an acceptable level of accuracy) the location of a point on the map. To accomplish this, degrees are divided into

minutes (') and **seconds** (") . There are sixty minutes in a degree, and sixty seconds in a minute (3600 seconds in a degree). So, at the equator, one second of latitude or longitude = 30.87624 meters.

8.5.2 Projected coordinate reference systems

A two-dimensional coordinate reference system is commonly defined by two axes. At right angles to each other, they form a so called **XY-plane** (see Fig. 8.8 on the left side). The horizontal axis is normally labelled **X**, and the vertical axis is normally labelled **Y**. In a three-dimensional coordinate reference system, another axis, normally labelled **Z**, is added. It is also at right angles to the **X** and **Y** axes. The **Z** axis provides the third dimension of space (see Fig. 8.8 on the right side). Every point that is expressed in spherical coordinates can be expressed as an **X Y Z** coordinate.

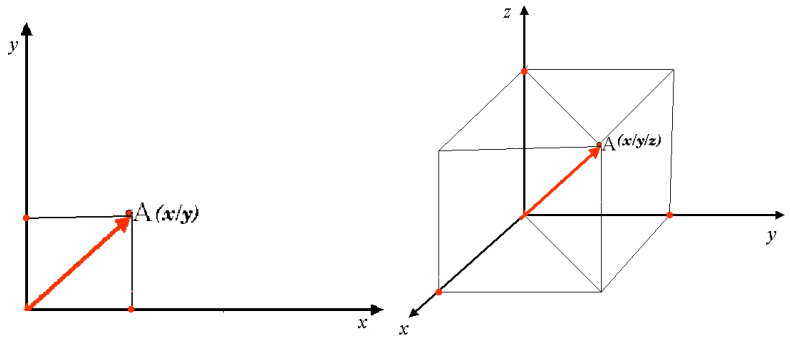


Fig. 8.8: Two and three dimensional coordinate reference systems.

A projected coordinate reference system in the southern hemisphere (south of the equator) normally has its origin on the equator at a specific **Longitude**. This means that the Y-values increase southwards and the X-values increase to the West. In the northern hemisphere (north of the equator) the origin is also the equator at a specific **Longitude**. However, now the Y-values increase northwards and the X-values increase to the East. In the following section, we describe a projected coordinate reference system, called Universal Transverse Mercator (UTM) often used for South Africa.

8.6 Universal Transverse Mercator (UTM) CRS in detail

The Universal Transverse Mercator (UTM) coordinate reference system has its origin on the **equator** at a specific **Longitude**. Now the Y-values increase southwards and the X-values increase to the West. The UTM CRS is a global map projection. This means, it is generally used all over the world. But as already described in the section 'accuracy of map projections' above, the larger the area (for example South Africa) the more distortion of angular conformity, distance and area occur. To avoid too much distortion, the world is divided into **60 equal zones** that are all **6 degrees** wide in longitude from East to West. The **UTM zones** are numbered **1 to 60**, starting at the **antimeridian** (**zone 1** at 180 degrees West longitude) and progressing East back to the **antemeridian** (**zone 60** at 180 degrees East longitude) as shown in Fig. 8.9.

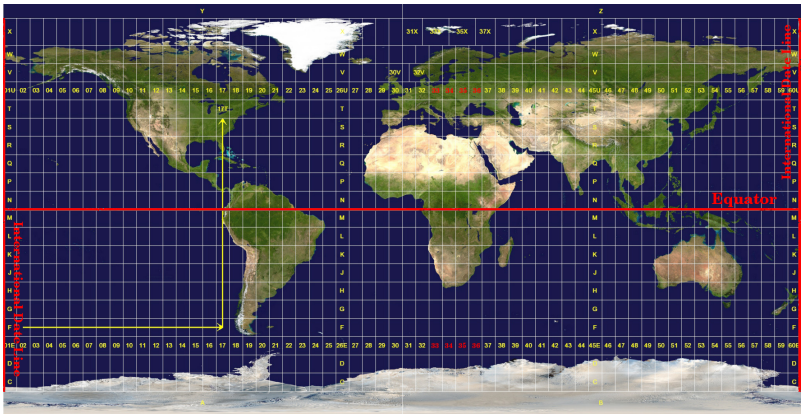


Fig. 8.9: The Universal Transverse Mercator zones. For South Africa UTM zones 33S, 34S, 35S, and 36S are used.

As you can see in Fig. 8.9 and Fig. 8.10, South Africa is covered by four **UTM zones** to minimize distortion. The **zones** are called **UTM 33S**, **UTM 34S**, **UTM 35S** and **UTM 36S**. The **S** after the zone means that the UTM zones are located **south of the equator**.

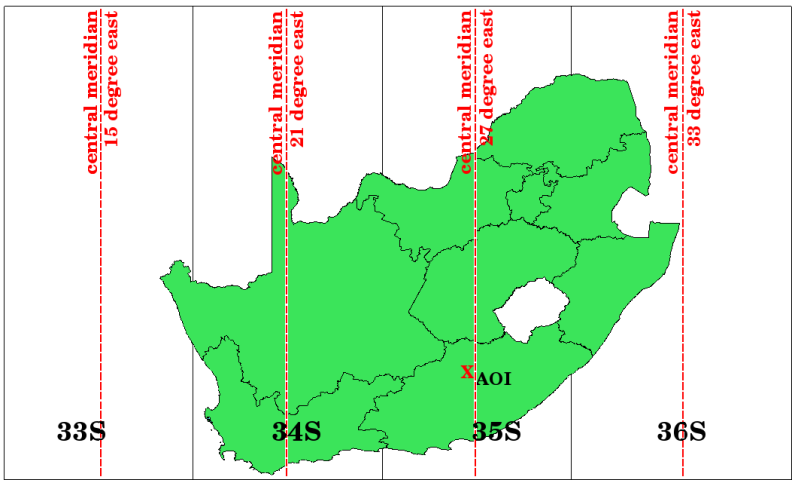


Fig. 8.10: UTM zones 33S, 34S, 35S, and 36S with their central longitudes (meridians) used to project South Africa with high accuracy. The red cross shows an Area of Interest (AOI).

Say, for example, that we want to define a two-dimensional coordinate within the **Area of Interest (AOI)** marked with a red cross in Fig. 8.10. You can see, that the area is located within the **UTM zone 35S**. This means, to minimize distortion and to get accurate analysis results, we should use **UTM zone 35S** as the coordinate reference system.

The position of a coordinate in UTM south of the equator must be indicated with the **zone number** (35) and with its **northing (Y) value** and **easting (X) value** in meters. The **northing value** is the distance of the position from the **equator** in meters. The **easting value** is the distance from the **central meridian** (longitude) of the used UTM zone. For UTM zone 35S it is **27 degrees East** as shown in Fig. 8.10. Furthermore, because we are south of the equator and negative values are not allowed in the UTM coordinate reference system, we have to add a so called **false northing value** of 10,000,000 m to the northing (Y) value and a false easting value of 500,000 m to the easting (X) value. This sounds difficult, so, we will do an example that shows you how to find the correct **UTM 35S** coordinate for the **Area of Interest**.

8.6.1 The northing (Y) value

The place we are looking for is 3,550,000 meters south of the equator, so the northing (Y) value gets a **negative sign** and is -3,550,000 m. According to the UTM definitions we have to add a **false northing value** of 10,000,000 m. This means the northing (Y) value of our coordinate is 6,450,000 m (-3,550,000 m + 10,000,000 m).

8.6.2 The easting (X) value

First we have to find the **central meridian** (longitude) for the **UTM zone 35S**. As we can see in Fig. 8.10 it is **27 degrees East**. The place we are looking for is **85,000 meters West** from the central meridian. Just like the northing value, the easting (X) value gets a negative sign, giving a result of **-85,000 m**. According to the UTM definitions we have to add a **false easting value** of 500,000 m. This means the easting (X) value of our coordinate is 415,000 m (-85,000 m + 500,000 m). Finally, we have to add the **zone number** to the easting value to get the correct value.

As a result, the coordinate for our **Point of Interest**, projected in **UTM zone 35S** would be written as: **35 415,000 m E / 6,450,000 m N**. In some GIS, when the correct UTM zone 35S is defined and the units are set to meters within the system, the coordinate could also simply appear as **415,000 6,450,000**.

8.7 On-The-Fly Projection

As you can probably imagine, there might be a situation where the data you want to use in a GIS are projected in different coordinate reference systems. For example, you might get a vector layer showing the boundaries of South Africa projected in UTM 35S and another vector layer with point information about rainfall provided in the geographic coordinate system WGS 84. In GIS these two vector layers are placed in totally different areas of the map window, because they have different projections.

To solve this problem, many GIS include a functionality called **on-the-fly** projection. It means, that you can **define** a certain projection when you start the GIS and all layers that you then load, no matter what coordinate reference system they have, will be automatically displayed in the projection you defined. This functionality allows you to overlay layers within the map window of your GIS, even though they may be in **different** reference systems. In QGIS, this functionality is applied by default.

8.8 Common problems / things to be aware of

The topic **map projection** is very complex and even professionals who have studied geography, geodetics or any other GIS related science, often have problems with the correct definition of map projections and coordinate reference systems. Usually when you work with GIS, you already have projected data to start with. In most cases these data will be projected in a certain CRS, so you don't have to create a new CRS or even re project the data from one CRS to another. That said, it is always useful to have an idea about what map projection and CRS means.

8.9 What have we learned?

Let's wrap up what we covered in this worksheet:

- **Map projections** portray the surface of the earth on a two-dimensional, flat piece of paper or computer screen.
- There are global map projections, but most map projections are created and **optimized to project smaller areas** of the earth's surface.
- Map projections are never absolutely accurate representations of the spherical earth. They show **distortions of angular conformity, distance and area.** It is impossible to preserve all these characteristics at the same time in a map projection.
- A **Coordinate reference system** (CRS) defines, with the help of coordinates, how the two-dimensional, projected map is related to real locations on the earth.
- There are two different types of coordinate reference systems: **Geographic Coordinate Systems** and **Projected Coordinate Systems**.
- **On the Fly** projection is a functionality in GIS that allows us to overlay layers, even if they are projected in different coordinate reference systems.

8.10 Now you try!

Here are some ideas for you to try with your learners:

1. Start QGIS
2. In *Project* ► *Properties…* ► *CRS* check *No projection (or unknown/non-Earth projection)*
3. Load two layers of the same area but with different projections
4. Let your pupils find the coordinates of several places on the two layers. You can show them that it is not possible to overlay the two layers.
5. Then define the coordinate reference system as Geographic/WGS 84 inside the *Project Properties* dialog
6. Load the two layers of the same area again and let your pupils see how setting a CRS for the project (hence, enabling "on-the-fly" projection) works.
7. You can open the *Project Properties* dialog in QGIS and show your pupils the many different Coordinate Reference Systems so they get an idea of the complexity of this topic. You can select different CRSs to display the same layer in different projections.

8.11 Something to think about

If you don't have a computer available, you can show your pupils the principles of the three map projection families. Get a globe and paper and demonstrate how cylindrical, conical and planar projections work in general. With the help of a transparency sheet you can draw a two-dimensional coordinate reference system showing X axes and Y axes. Then, let your pupils define coordinates (X and Y values) for different places.

8.12 Further reading

Books:

- Chang, Kang-Tsung (2006). Introduction to Geographic Information Systems. 3rd Edition. McGraw Hill. ISBN: 0070658986
- DeMers, Michael N. (2005). Fundamentals of Geographic Information Systems. 3rd Edition. Wiley. ISBN: 9814126195
- Galati, Stephen R. (2006): Geographic Information Systems Demystified. Artech House Inc. ISBN: 158053533X

Websites:

- https://foote.geography.uconn.edu/gcraft/notes/mapproj/mapproj_f.html
- http://geology.isu.edu/wapi/geostac/Field_Exercise/topomaps/index.htm

The QGIS User Guide also has more detailed information on working with map projections in QGIS.

8.13 What's next?

In the section that follows we will take a closer look at **Map Production**.

CHAPTER

NINE

MAP PRODUCTION



Objectives:	Understanding of map production for spatial data
Keywords:	Map production, map layout, scale bar, north arrow, legend, map body, map unit

9.1 Overview

Map production is the process of arranging map elements on a sheet of paper in a way that, even without many words, the average person can understand what it is all about. Maps are usually produced for presentations and reports where the audience or reader is a politician, citizen or a learner with no professional background in GIS. Because of this, a map has to be effective in communicating spatial information. Common elements of a map are the title, map body, legend, north arrow, scale bar, acknowledgement, and map border (see Fig. 9.1).

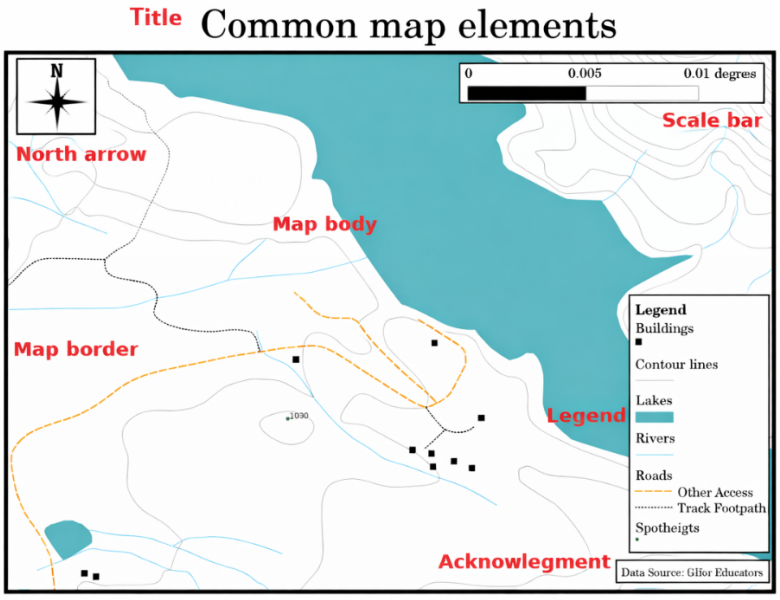


Fig. 9.1: Common map elements (labelled in red) are the title, map body, legend, north arrow, scale bar, acknowledgement and map border.

Other elements that might be added are e.g. a **graticule**, or **name of the map projection** (CRS). Together, these elements help the map reader to interpret the information shown on the map. The map body is, of course, the most important part of the map because it contains the map information. The other elements support the communication process and help the map reader to orientate himself and understand the map topic. For example, the title describes the subject matter and the legend relates map symbols to the mapped data.

9.2 Title in detail

The map title is very important because it is usually the first thing a reader will look at on a map. It can be compared with a title in a newspaper. It should be short but give the reader a first idea of what the map is about.

9.3 Map Border in detail

The map border is a line that defines exactly the edges of the area shown on the map. When printing a map with a graticule (which we describe further down), you often find the coordinate information of the graticule lines along the border lines, as you can see in Fig. 9.5.

9.4 Map Legend in detail

A map is a simplified representation of the real world and **map symbols** are used to represent real objects. Without symbols, we wouldn't understand maps. To ensure that a person can correctly read a map, a map legend is used to provide a key to all the symbols used on the map. It is like a dictionary that allows you to understand the meaning of what the map shows. A map legend is usually shown as a little box in a corner of the map. It contains icons, each of which will represent a type of feature. For example, a *house* icon will show you how to identify houses on the map (see Fig. 9.2).

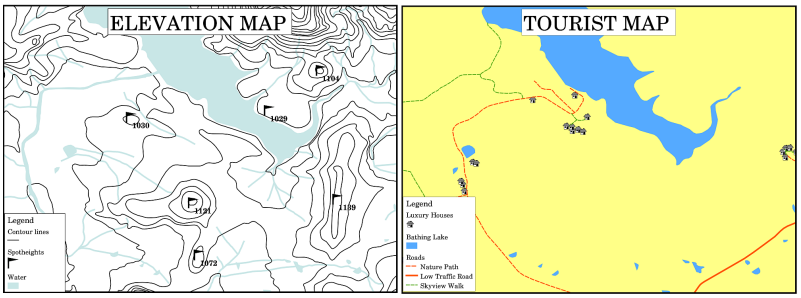


Fig. 9.2: Two maps from the same area, both with a water body in the background but with different themes, map symbols and colours in the legend.

You can also use different symbols and icons in your legend to show different themes. In Fig. 9.2 you can see a map with a lake in light blue overlaid with contour lines and spot heights to show information about the terrain in that area. On the right side you see the same area with the lake in the background but this map is designed to show tourists the location of houses they can rent for their holidays. It uses brighter colours, a house icon and more descriptive and inviting words in the legend.

9.5 North arrow in detail

A north arrow (sometimes also called a compass rose) is a figure displaying the main directions, **North**, **South**, **East** and **West**. On a map it is used to indicate the direction of North.

For example, in GIS this means that a house that is located north from a lake can be found on top of the lake on a map. The road in the east will then be to the right of the water body on the map, a river in the south will be below the water body and if you are searching for a train station to the west of the lake you will find it on the left side on the map.

9.6 Scale in detail

The scale of a map, is the value of a single unit of distance on the map, representing distance in the real world. The values are shown in map units (meters, feet or degrees). The scale can be expressed in several ways, for example, in words, as a ratio or as a graphical scale bar (see Fig. 9.3).

Expressing a scale in words is a commonly used method and has the advantage of being easily understood by most map users. You can see an example of a word based scale in a Fig. 9.3 (a). Another option is the **representative fraction (RF)** method, where both the map distance and the ground distance in the real world are given in the same map units, as a ratio. For example, a RF value 1:25,000 means that any distance on the map is 1/25,000 *th* of the real distance on the ground (see Fig. 9.3 (b)). The value 25,000 in the ratio is called the **scale denominator**. More experienced users often prefer the representative fraction method, because it reduces confusion.

When a representative fraction expresses a very small ratio, for example 1:1000 000, it is called a **small scale map**. On the other hand if the ratio is very large, for example a 1:50 000 map, it is called a **large scale map**. It is handy to remember that a small scale map covers a **large area**, and a large scale map covers a **small area**!

A **scale expression as a graphic or bar scale** is another basic method of expressing a scale. A bar scale shows measured distances on the map. The equivalent distance in the real world is placed above as you can see in Fig. 9.3 (c).

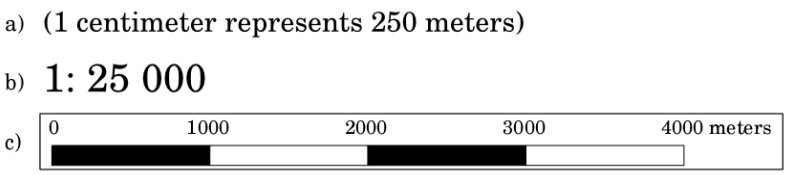


Fig. 9.3: A map scale can be expressed in words (a), as a ratio (b), or as graphic or bar scale (c)

Maps are usually produced at standard scales of, for example, 1:10 000, 1:25 000, 1:50 000, 1:100 000, 1:250 000, 1:500 000. What does this mean to the map reader? It means that if you *multiply* the distance measured on the **map** by the **scale denominator**, you will know the distance in the **real world**.

For example, if we want to measure a distance of 100mm on a map with a scale of 1:25,000 we calculate the real world distance like this:

```
100 mm x 25,000 = 2,500,000 mm
```

This means that 100 mm on the map is equivalent to 2,500,000 mm (2500 m) in the real world.

Another interesting aspect of a map scale, is that the lower the map scale, the more detailed the feature information in the map will be. In Fig. 9.4, you can see an example of this. Both maps are the same size but have a different scale. The image on the left side shows more details, for example the houses south-west of the water body can be clearly identified as separate squares. In the image on the right you can only see a black clump of rectangles and you are not able to see each house clearly.

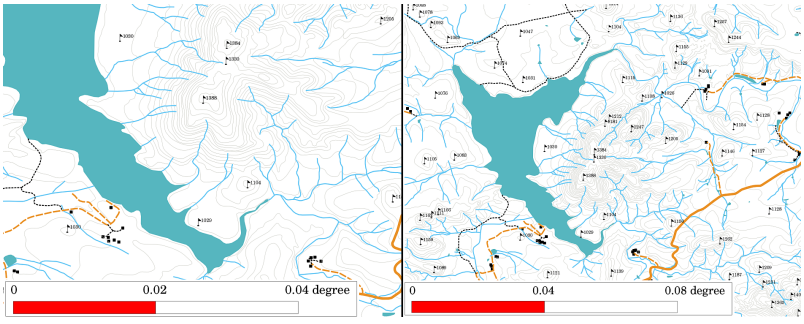


Fig. 9.4: Maps showing an area in two different scales. The map scale on the left is 1:25,000. The map scale on the right is 1:50,000.

9.7 Acknowledgment in detail

In the acknowledgment area of a map it is possible to add text with important information. For example information about the quality of the used data can be useful to give the reader an idea about details such as how, by whom and when a map was created. If you look at a topographical map of your town, it would be useful to know when the map was created and who did it. If the map is already 50 years old, you will probably find a lot of houses and roads that no longer exist or maybe never even existed. If you know that the map was created by an official institution, you could contact them and ask if they have a more current version of that map with updated information.

9.8 Graticule in detail

A graticule is a network of lines overlain on a map to make spatial orientation easier for the reader. The lines can be used as a reference. As an example, the lines of a graticule can represent the earth's parallels of latitude and meridians of longitude. When you want to refer to a special area on a map during your presentation or in a report you could say: 'the houses close to latitude 26.04 / longitude -32.11 are often exposed to flooding during January and February' (see Fig. 9.5).

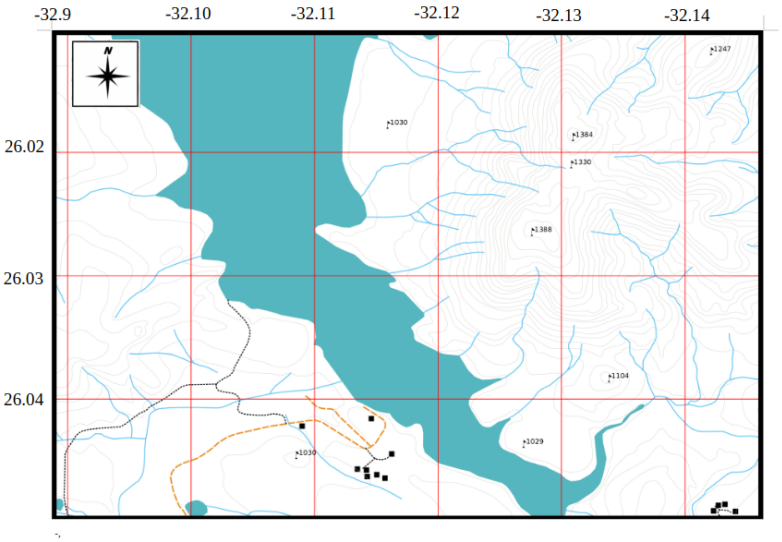


Fig. 9.5: Graticules (red lines) representing the Earth's parallels of latitude and meridians of longitude. The latitude and longitude values on the map border can be used for better orientation on the map.

9.9 Name of the map projection in detail

A map projection tries to represent the 3-dimensional Earth with all its features like houses, roads or lakes on a flat sheet of paper. This is very difficult as you can imagine, and even after hundreds of years there is no single projection that is able to represent the Earth perfectly for any area in the world. Every projection has advantages and disadvantages.

To be able to create maps as precisely as possible, people have studied, modified, and produced many different kinds of projections. In the end almost every country has developed its own map projection with the goal of improving the map accuracy for their territorial area (see Fig. 9.6).

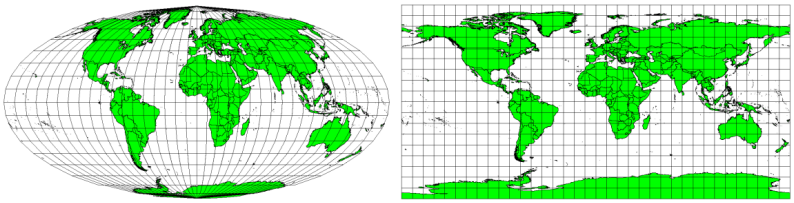


Fig. 9.6: The world in different projections. A Mollweide Equal Area projection left, a Plate Carree Equidistant Cylindrical projection on the right.

With this in mind, we can now understand why it makes sense to add the name of the projection on a map. It allows the reader to see quickly, if one map can be compared with another. For example, features on a map in a so-called Equal Area projection appear very different to features projected in a Cylindrical Equidistant projection (see Fig. 9.6).

Map projection is a very complex topic and we cannot cover it completely here. You may want to take a look at our previous topic: Coordinate Reference Systems if you want to know more about it.

9.10 Common problems / things to be aware of

It is sometimes difficult to create a map that is easy to understand and well laid out whilst still showing and explaining all the information that the reader needs to know. To achieve this, you need to create an ideal arrangement and composition of all the map elements. You should concentrate on what story you want to tell with your map and how the elements, such as the legend, scale bar and acknowledgements should be ordered. By doing this, you will have a well designed and educational map, that people would like to look at and be able to understand.

9.11 What have we learned?

Let's wrap up what we covered in this worksheet:

- **Map production** means arranging **map elements** on a sheet of paper.
- **Map elements** are the title, map body, map border, legend, scale, north arrow and the acknowledgement.
- **Scale** represents the ratio of a distance on the map to the actual distance in the real world.
- Scale is displayed in **map units** (meters, feet or degrees)
- A **legend** explains all the symbols on a map.
- A map should **explain complex information as simply as possible**.
- Maps are usually always displayed '**North up**'.

9.12 Now you try!

Here are some ideas for you to try with your learners:

- Load some vector layers in your GIS for your local area. See if your learners can identify examples of different types of legend elements such as road types or buildings. Create a list of legend elements and define what the icons should look like, so a reader can most easily figure out their meaning in the map.
- Create a map layout with your learners on a sheet of paper. Decide on the title of the map, what GIS layers you want to show and what colors and icons to have on the map. Use the techniques you learned in Topics *Vector Data* and *Vector Attribute Data* to adjust the symbology accordingly. When you have a template, open the QGIS print layout and try to arrange a map layout as planned.

9.13 Something to think about

If you don't have a computer available, you can use any topographical map and discuss the map design with your learners. Figure out if they understand what the map wants to tell. What can be improved? How accurately does the map represent the history of the area? How would a map from 100 years ago differ from the same map today?

9.14 Further reading

Books:

- Chang, Kang-Tsung (2006). Introduction to Geographic Information Systems. 3rd Edition. McGraw Hill. ISBN: 0070658986
- DeMers, Michael N. (2005). Fundamentals of Geographic Information Systems. 3rd Edition. Wiley. ISBN: 9814126195

Website: Scale (map)

The QGIS User Guide also has more detailed information on map production provided in QGIS.

9.15 What's next?

In the section that follows we will take a closer look at **vector analysis** to see how we can use a GIS for more than just making good looking maps!

CHAPTER

TEN

VECTOR SPATIAL ANALYSIS (BUFFERS)



Objectives:	Understanding the use of buffering in vector spatial analysis.
Keywords:	Vector, buffer zone, spatial analysis, buffer distance, dissolve boundary, outward and inward buffer, multiple buffer

10.1 Overview

Spatial analysis uses spatial information to extract new and additional meaning from GIS data. Usually spatial analysis is carried out using a GIS Application. GIS Applications normally have spatial analysis tools for feature statistics (e.g. how many vertices make up this polyline?) or geoprocessing such as feature buffering. The types of spatial analysis that are used vary according to subject areas. People working in water management and research (hydrology) will most likely be interested in analysing terrain and modelling water as it moves across it. In wildlife management users are interested in analytical functions that deal with wildlife point locations and their relationship to the environment. In this topic we will discuss buffering as an example of a useful spatial analysis that can be carried out with vector data.

10.2 Buffering in detail

Buffering usually creates two areas: one area that is **within** a specified distance to selected real world features and the other area that is **beyond**. The area that is within the specified distance is called the **buffer zone**.

A **buffer zone** is any area that serves the purpose of keeping real world features distant from one another. Buffer zones are often set up to protect the environment, protect residential and commercial zones from industrial accidents or natural disasters, or to prevent violence. Common types of buffer zones may be greenbelts between residential and commercial areas, border zones between countries (see Fig. 10.1), noise protection zones around airports, or pollution protection zones along rivers.



Fig. 10.1: The border between the United States of America and Mexico is separated by a buffer zone. (Photo taken by SGT Jim Greenhill 2006).

In a GIS Application, **buffer zones are** always represented as **vector polygons** enclosing other polygon, line or point features (see Fig. 10.2, Fig. 10.3, Fig. 10.4).

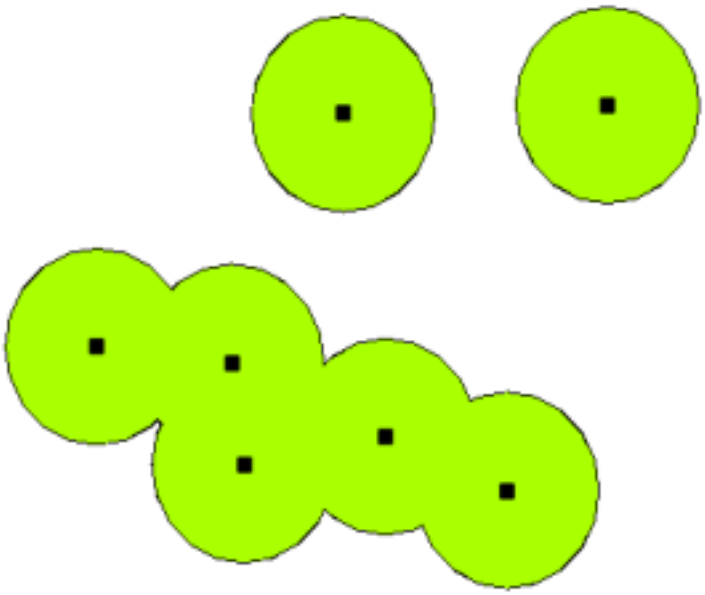


Fig. 10.2: A buffer zone around vector points.



Fig. 10.3: A buffer zone around vector polylines.

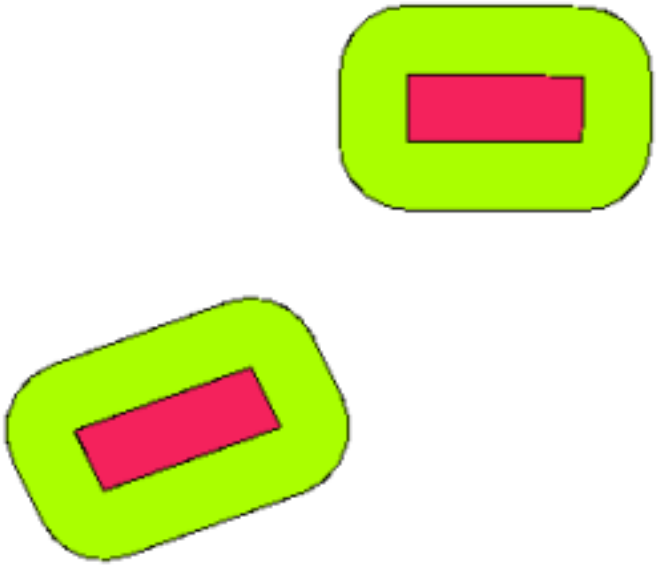


Fig. 10.4: A buffer zone around vector polygons.

10.3 Variations in buffering

There are several variations in buffering. The **buffer distance** or buffer size **can vary** according to numerical values provided in the vector layer attribute table for each feature. The numerical values have to be defined in map units according to the Coordinate Reference System (CRS) used with the data. For example, the width of a buffer zone along the banks of a river can vary depending on the intensity of the adjacent land use. For intensive cultivation the buffer distance may be bigger than for organic farming (see Figure Fig. 10.5 and Table 10.1).

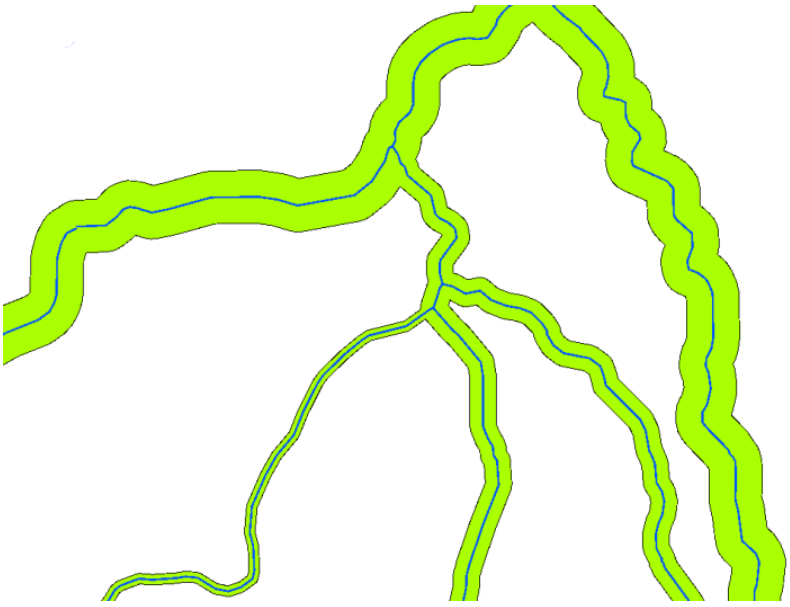


Fig. 10.5: Buffering rivers with different buffer distances.

Table 10.1: Different buffer distances to rivers based on adjacent land use.

River	Adjacent land use	Buffer distance (meters)
Breede River	Intensive vegetable cultivation	100
Komati	Intensive cotton cultivation	150
Oranje	Organic farming	50
Telle river	Organic farming	50

Buffers around polyline features, such as rivers or roads, do not have to be on both sides of the lines. They can be on either the left side or the right side of the line feature. In these cases the left or right side is determined by the direction from the starting point to the end point of the line during digitising.

10.3.1 Multiple buffer zones

A feature can also have more than one buffer zone. A nuclear power plant may be buffered with distances of 10, 15, 25 and 30 km, thus forming multiple rings around the plant as part of an evacuation plan (see Fig. 10.6).

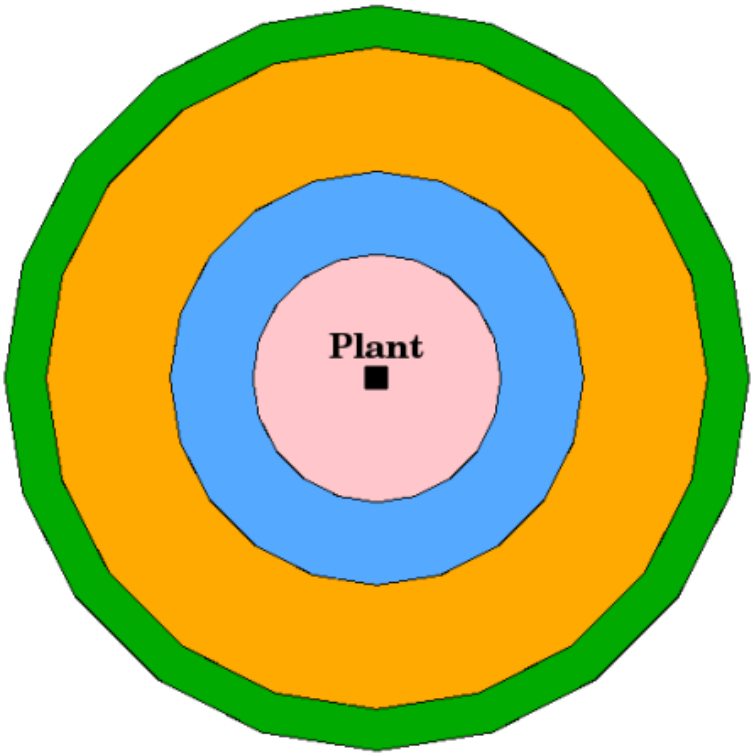


Fig. 10.6: Buffering a point feature with distances of 10, 15, 25 and 30 km.

10.3.2 Buffering with intact or dissolved boundaries

Buffer zones often have dissolved boundaries so that there are no overlapping areas between the buffer zones. In some cases though, it may also be useful for boundaries of buffer zones to remain intact, so that each buffer zone is a separate polygon and you can identify the overlapping areas (see Figure Fig. 10.7).

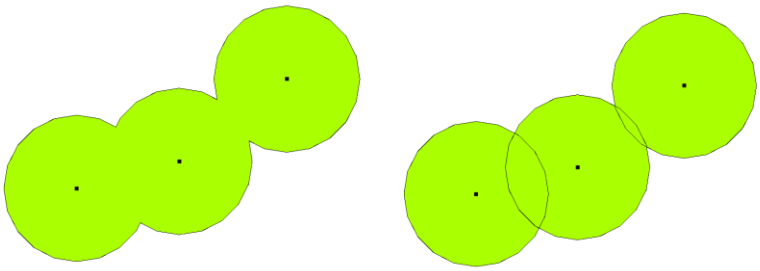


Fig. 10.7: Buffer zones with dissolved (left) and with intact boundaries (right) showing overlapping areas.

10.3.3 Buffering outward and inward

Buffer zones around polygon features are usually extended outward from a polygon boundary but it is also possible to create a buffer zone inward from a polygon boundary. Say, for example, the Department of Tourism wants to plan a new road around Robben Island and environmental laws require that the road is at least 200 meters inward from the coastline. They could use an inward buffer to find the 200 m line inland and then plan their road not to go beyond that line.

10.4 Common problems / things to be aware of

Most GIS Applications offer buffer creation as an analysis tool, but the options for creating buffers can vary. For example, not all GIS Applications allow you to buffer on either the left side or the right side of a line feature, to dissolve the boundaries of buffer zones or to buffer inward from a polygon boundary.

A buffer distance always has to be defined as a whole number (integer) or a decimal number (floating point value). This value is defined in map units (meters, feet, decimal degrees) according to the Coordinate Reference System (CRS) of the vector layer.

10.5 More spatial analysis tools

Buffering is an important and often used spatial analysis tool but there are many others that can be used in a GIS and explored by the user.

Spatial overlay is a process that allows you to identify the relationships between two polygon features that share all or part of the same area. The output vector layer is a combination of the input features information (see Fig. 10.8).

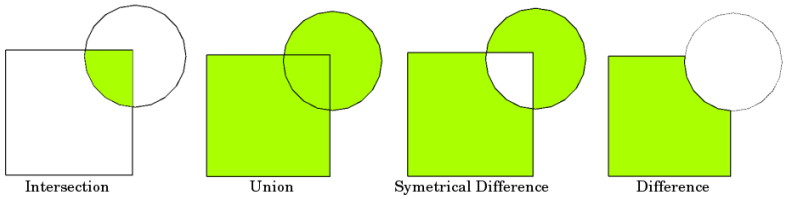


Fig. 10.8: Spatial overlay with two input vector layers (a_input = rectangle, b_input = circle). The resulting vector layer is displayed in green.

Typical spatial overlay examples are:

- **Intersection**: The output layer contains all areas where both layers overlap (intersect).
- **Union**: the output layer contains all areas of the two input layers combined.
- **Symmetrical difference**: The output layer contains all areas of the input layers except those areas where the two layers overlap (intersect).
- **Difference**: The output layer contains all areas of the first input layer that do not overlap (intersect) with the second input layer.

10.6 What have we learned?

Let's wrap up what we covered in this worksheet:

- **Buffer zones** describe areas around real world features.
- Buffer zones are always **vector polygons**.
- A feature can have **multiple** buffer zones.
- The size of a buffer zone is defined by a **buffer distance**.
- A buffer distance has to be an **integer** or **floating point** value.
- A buffer distance can be different for each feature within a vector layer.
- Polygons can be buffered **inward** or **outward** from the polygon boundary.
- Buffer zones can be created with **intact** or **dissolved** boundaries.
- Besides buffering, a GIS usually provides a variety of vector analysis tools to solve spatial tasks.

10.7 Now you try!

Here are some ideas for you to try with your learners:

- Because of dramatic traffic increase, the town planners want to widen the main road and add a second lane. Create a buffer around the road to find properties that fall within the buffer zone (see Fig. 10.9).
- For controlling protesting groups, the police want to establish a neutral zone to keep protesters at least 100 meters from a building. Create a buffer around a building and colour it so that event planners can see where the buffer area is.
- A truck factory plans to expand. The siting criteria stipulate that a potential site must be within 1 km of a heavy-duty road. Create a buffer along a main road so that you can see where potential sites are.
- Imagine that the city wants to introduce a law stipulating that no bottle stores may be within a 1000 meter buffer zone of a school or a church. Create a 1 km buffer around your school and then go and see if there would be any bottle stores too close to your school.

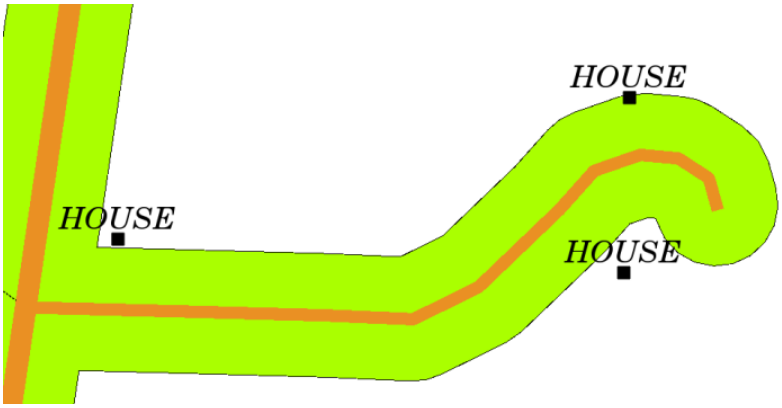


Fig. 10.9: Buffer zone (green) around a roads map (brown). You can see which houses fall within the buffer zone, so now you could contact the owner and talk to him about the situation.

10.8 Something to think about

If you don't have a computer available, you can use a toposheet and a compass to create buffer zones around buildings. Make small pencil marks at an equal distance all along your feature using the compass, then connect the marks using a ruler!

10.9 Further reading

Books:

- Galati, Stephen R. (2006). Geographic Information Systems Demystified. Artech House Inc. ISBN: 158053533X
- Chang, Kang-Tsung (2006). Introduction to Geographic Information Systems. 3rd Edition. McGraw Hill. ISBN: 0070658986
- DeMers, Michael N. (2005). Fundamentals of Geographic Information Systems. 3rd Edition. Wiley. ISBN: 9814126195

The QGIS User Guide also has more detailed information on analysing vector data in QGIS.

10.10 What's next?

In the section that follows we will take a closer look at **interpolation** as an example of spatial analysis you can do with raster data.

CHAPTER

ELEVEN

SPATIAL ANALYSIS (INTERPOLATION)



Objectives:	Understanding of interpolation as part of spatial analysis
Keywords:	Point data, interpolation method, Inverse Distance Weighted, Triangulated Irregular Network

11.1 Overview

Spatial analysis is the process of manipulating spatial information to extract new information and meaning from the original data. Usually spatial analysis is carried out with a Geographic Information System (GIS). A GIS usually provides spatial analysis tools for calculating feature statistics and carrying out geoprocessing activities as data interpolation. In hydrology, users will likely emphasize the importance of terrain analysis and hydrological modelling (modelling the movement of water over and in the earth). In wildlife management, users are interested in analytical functions dealing with wildlife point locations and their relationship to the environment. Each user will have different things they are interested in depending on the kind of work they do.

11.2 Spatial interpolation in detail

Spatial interpolation is the process of using points with known values to estimate values at other unknown points. For example, to make a precipitation (rainfall) map for your country, you will not find enough evenly spread weather stations to cover the entire region. Spatial interpolation can estimate the temperatures at locations without recorded data by using known temperature readings at nearby weather stations (see Fig. 11.1). This type of interpolated surface is often called a **statistical surface**. Elevation data, precipitation, snow accumulation, water table and population density are other types of data that can be computed using interpolation.

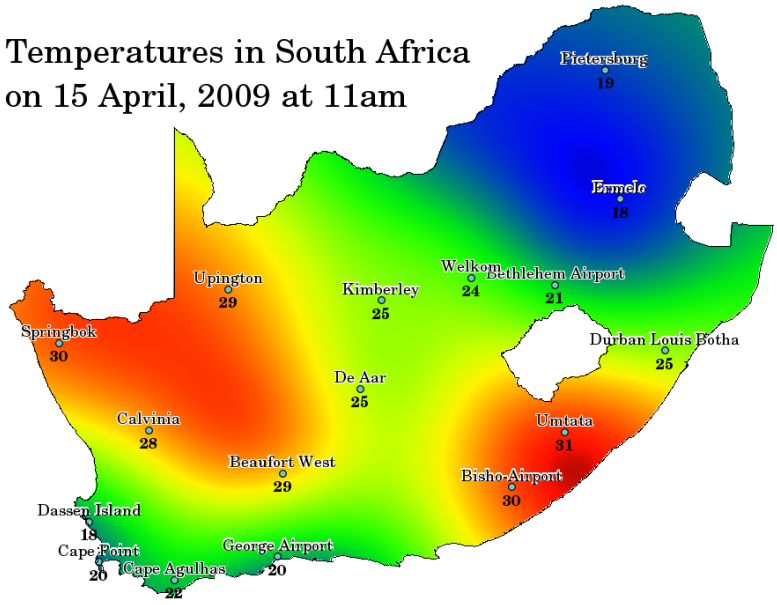


Fig. 11.1: Temperature map interpolated from South African Weather Stations.

Because of high cost and limited resources, data collection is usually conducted only in a limited number of selected point locations. In GIS, spatial interpolation of these points can be applied to create a raster surface with estimates made for all raster cells.

In order to generate a continuous map, for example, a digital elevation map from elevation points measured with a GPS device, a suitable interpolation method has to be used to optimally estimate the values at those locations where no samples or measurements were taken. The results of the interpolation analysis can then be used for analyses that cover the whole area and for modelling.

There are many interpolation methods. In this introduction we will present two widely used interpolation methods called **Inverse Distance Weighting** (IDW) and **Triangulated Irregular Networks** (TIN). If you are looking for additional interpolation methods, please refer to the 'Further Reading' section at the end of this topic.

11.3 Inverse Distance Weighted (IDW)

In the IDW interpolation method, the sample points are weighted during interpolation such that the influence of one point relative to another declines with distance from the unknown point you want to create (see Fig. 11.2).

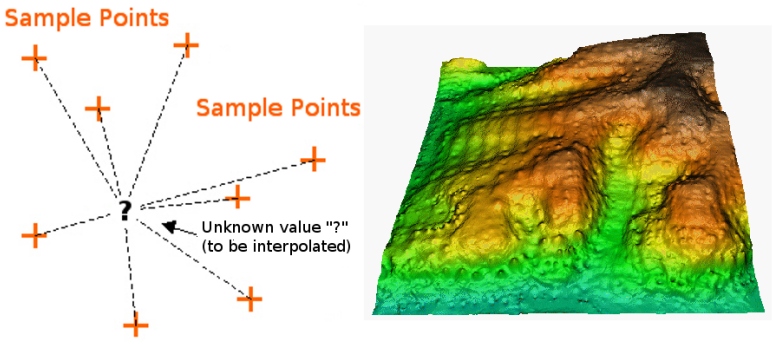


Fig. 11.2: Inverse Distance Weighted interpolation based on weighted sample point distance (left). Interpolated IDW surface from elevation vector points (right). Image Source: Mitas, L., Mitasova, H. (1999).

Weighting is assigned to sample points through the use of a weighting coefficient that controls how the weighting influence will drop off as the distance from new point increases. The greater the weighting coefficient, the less the

effect points will have if they are far from the unknown point during the interpolation process. As the coefficient increases, the value of the unknown point approaches the value of the nearest observational point.

It is important to notice that the IDW interpolation method also has some disadvantages: the quality of the interpolation result can decrease, if the distribution of sample data points is uneven. Furthermore, maximum and minimum values in the interpolated surface can only occur at sample data points. This often results in small peaks and pits around the sample data points as shown in Fig. 11.2.

In GIS, interpolation results are usually shown as a 2 dimensional raster layer. In Fig. 11.3, you can see a typical IDW interpolation result, based on elevation sample points collected in the field with a GPS device.

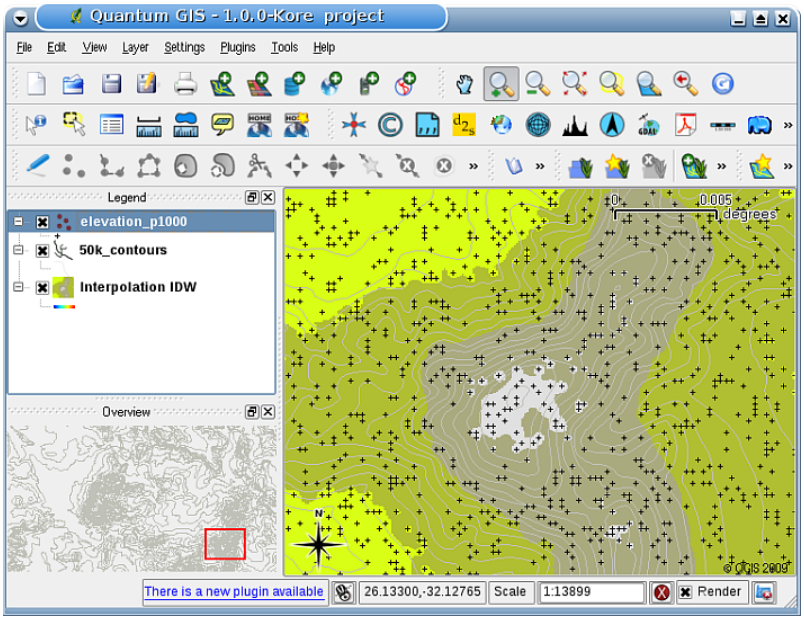


Fig. 11.3: IDW interpolation result from irregularly collected elevation sample points (shown as black crosses).

11.4 Triangulated Irregular Network (TIN)

TIN interpolation is another popular tool in GIS. A common TIN algorithm is called **Delaunay triangulation**. It tries to create a surface formed by triangles of nearest neighbour points. To do this, circumcircles around selected sample points are created and their intersections are connected to a network of non overlapping and as compact as possible triangles (see Fig. 11.4).

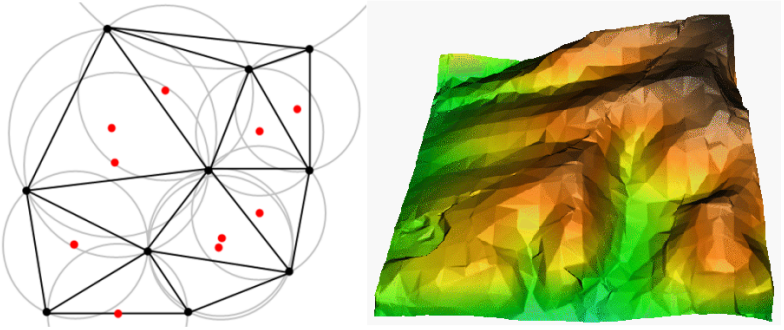


Fig. 11.4: Delaunay triangulation with circumcircles around the red sample data. The resulting interpolated TIN surface created from elevation vector points is shown on the right. Image Source: Mitas, L., Mitasova, H. (1999).

The main disadvantage of the TIN interpolation is that the surfaces are not smooth and may give a jagged appearance. This is caused by discontinuous slopes at the triangle edges and sample data points. In addition, triangulation is generally not suitable for extrapolation beyond the area with collected sample data points (see Fig. 11.5).

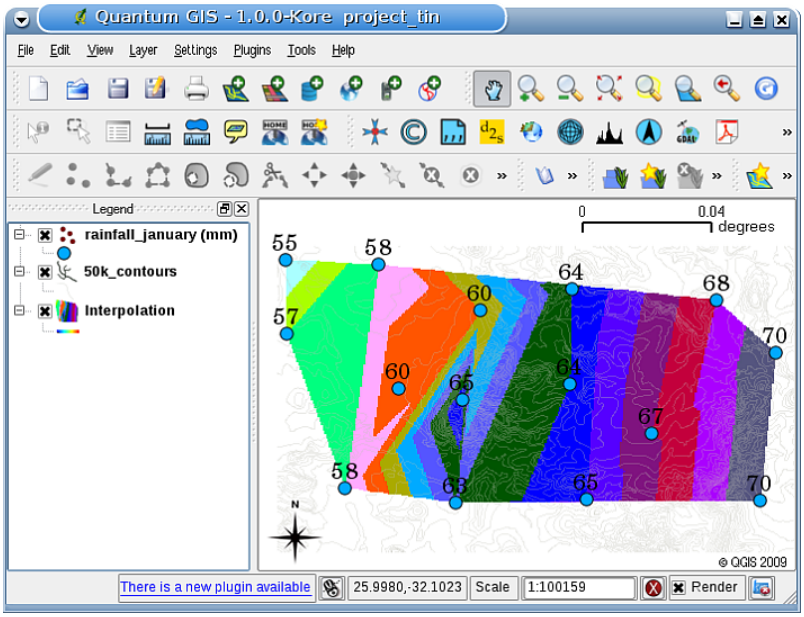


Fig. 11.5: Delaunay TIN interpolation result from irregularly collected rainfall sample points (blue circles)

11.5 Common problems / things to be aware of

It is important to remember that there is no single interpolation method that can be applied to all situations. Some are more exact and useful than others but take longer to calculate. They all have advantages and disadvantages. In practice, selection of a particular interpolation method should depend upon the sample data, the type of surfaces to be generated and tolerance of estimation errors. Generally, a three step procedure is recommended:

1. Evaluate the sample data. Do this to get an idea on how data are distributed in the area, as this may provide hints on which interpolation method to use.
2. Apply an interpolation method which is most suitable to both the sample data and the study objectives. When you are in doubt, try several methods, if available.
3. Compare the results and find the best result and the most suitable method. This may look like a time consuming process at the beginning. However, as you gain experience and knowledge of different interpolation methods, the time required for generating the most suitable surface will be greatly reduced.

11.6 Other interpolation methods

Although we concentrated on IDW and TIN interpolation methods in this worksheet, there are more spatial interpolation methods provided in GIS, such as Regularized Splines with Tension (RST), Kriging or Trend Surface interpolation. See the additional reading section below for a web link.

11.7 What have we learned?

Let's wrap up what we covered in this worksheet:

- **Interpolation** uses vector points with known values to estimate values at unknown locations to create a raster surface covering an entire area.
- The interpolation result is typically a **raster** layer.
- It is important to **find a suitable interpolation method** to optimally estimate values for unknown locations.
- **IDW interpolation** gives weights to sample points, such that the influence of one point on another declines with distance from the new point being estimated.

- **TIN interpolation** uses sample points to create a surface formed by triangles based on nearest neighbour point information.

11.8 Now you try!

Here are some ideas for you to try with your learners:

- The Department of Agriculture plans to cultivate new land in your area but apart from the character of the soils, they want to know if the rainfall is sufficient for a good harvest. All the information they have available comes from a few weather stations around the area. Create an interpolated surface with your learners that shows which areas are likely to receive the highest rainfall.
- The tourist office wants to publish information about the weather conditions in January and February. They have temperature, rainfall and wind strength data and ask you to interpolate their data to estimate places where tourists will probably have optimal weather conditions with mild temperatures, no rainfall and little wind strength. Can you identify the areas in your region that meet these criteria?

11.9 Something to think about

If you don't have a computer available, you can use a toposheet and a ruler to estimate elevation values between contour lines or rainfall values between fictional weather stations. For example, if rainfall at weather station A is 50 mm per month and at weather station B it is 90 mm, you can estimate, that the rainfall at half the distance between weather station A and B is 70 mm.

11.10 Further reading

Books:

- Chang, Kang-Tsung (2006). Introduction to Geographic Information Systems. 3rd Edition. McGraw Hill. ISBN: 0070658986
- DeMers, Michael N. (2005): Fundamentals of Geographic Information Systems. 3rd Edition. Wiley. ISBN: 9814126195
- Mitas, L., Mitasova, H. (1999). Spatial Interpolation. In: P.Longley, M.F. Goodchild, D.J. Maguire, D.W.Rhind (Eds.), Geographical Information Systems: Principles, Techniques, Management and Applications, Wiley.

Websites:

- https://en.wikipedia.org/wiki/Interpolation
- https://en.wikipedia.org/wiki/Delaunay_triangulation

The QGIS User Guide also has more detailed information on interpolation tools provided in QGIS.

11.11 What's next?

This is the final worksheet in this series. We encourage you to explore QGIS and use the accompanying QGIS manual to discover all the other things you can do with GIS software!

CHAPTER

TWELVE

ABOUT THE AUTHORS & CONTRIBUTORS



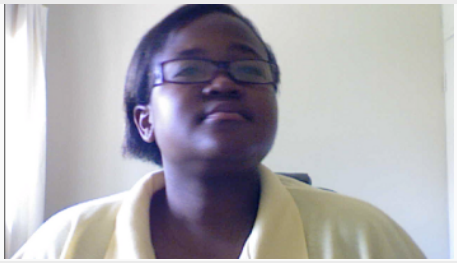
Tim Sutton — Editor & Lead Author. Tim is a developer and project steering committee member of the QGIS project. He is passionate about seeing GIS being Freely available to everyone. Tim is also a founding member of Linfiniti Consulting CC. — a small business set up with the goal of helping people to learn and use opensource GIS software. **Web**: https://kartoza.com **Email**: tim@kartoza.com



Otto Dassau — Assistant Author. Otto Dassau is the documentation maintainer and project steering committee member of the QGIS project. Otto has considerable experience in using and training people to use Free and Open Source GIS software. **Web**: http://www.nature-consult.de **Email**: otto.dassau@gmx.de



Marcelle Sutton — Project Manager. Marcelle Sutton studied English and drama and is a qualified teacher. Marcelle is also a founding member of Linfiniti Consulting CC. — a small business set up with the goal of helping people to learn and use opensource GIS software. **Web**: https://kartoza.com **Email**: marcelle@kartoza.com



Lerato Nsibande — Video Presenter. Lerato is a grade 12 scholar living in Pretoria. Lerato learns Geography at school and has enjoyed learning GIS with us!

continues on next page

Table 12.1 – continued from previous page

	Sibongile Mthombeni — Video Presenter. Sibongile lives near Johannesburg with her young daughter. Her goal is to continue her studies and become a nurse. Working on this project was the first time Sibongile used a computer.

CHAPTER

THIRTEEN

GNU FREE DOCUMENTATION LICENSE

Version 1.3, 3 November 2008

Copyright 2000, 2001, 2002, 2007, 2008 Free Software Foundation, Inc https://www.fsf.org/

Everyone is permitted to copy and distribute verbatim copies of this license document, but changing it is not allowed.

Preamble

The purpose of this License is to make a manual, textbook, or other functional and useful document "free" in the sense of freedom: to assure everyone the effective freedom to copy and redistribute it, with or without modifying it, either commercially or noncommercially. Secondarily, this License preserves for the author and publisher a way to get credit for their work, while not being considered responsible for modifications made by others.

This License is a kind of "copyleft", which means that derivative works of the document must themselves be free in the same sense. It complements the GNU General Public License, which is a copyleft license designed for free software.

We have designed this License in order to use it for manuals for free software, because free software needs free documentation: a free program should come with manuals providing the same freedoms that the software does. But this License is not limited to software manuals; it can be used for any textual work, regardless of subject matter or whether it is published as a printed book. We recommend this License principally for works whose purpose is instruction or reference.

1. APPLICABILITY AND DEFINITIONS

This License applies to any manual or other work, in any medium, that contains a notice placed by the copyright holder saying it can be distributed under the terms of this License. Such a notice grants a world-wide, royalty-free license, unlimited in duration, to use that work under the conditions stated herein. The **Document**, below, refers to any such manual or work. Any member of the public is a licensee, and is addressed as "**you**". You accept the license if you copy, modify or distribute the work in a way requiring permission under copyright law.

A "**Modified Version**" of the Document means any work containing the Document or a portion of it, either copied verbatim, or with modifications and/or translated into another language.

A "**Secondary Section**" is a named appendix or a front-matter section of the Document that deals exclusively with the relationship of the publishers or authors of the Document to the Document's overall subject (or to related matters) and contains nothing that could fall directly within that overall subject. (Thus, if the Document is in part a textbook of mathematics, a Secondary Section may not explain any mathematics.) The relationship could be a matter of historical connection with the subject or with related matters, or of legal, commercial, philosophical, ethical or political position regarding them.

The "**Invariant Sections**" are certain Secondary Sections whose titles are designated, as being those of Invariant Sections, in the notice that says that the Document is released under this License. If a section does not fit the above definition of Secondary then it is not allowed to be designated as Invariant. The Document may contain zero Invariant Sections. If the Document does not identify any Invariant Sections then there are none.

The "**Cover Texts**" are certain short passages of text that are listed, as Front-Cover Texts or Back-Cover Texts, in the notice that says that the Document is released under this License. A Front-Cover Text may be at most 5 words, and a Back-Cover Text may be at most 25 words.

A "**Transparent**" copy of the Document means a machine-readable copy, represented in a format whose specification is available to the general public, that is suitable for revising the document straightforwardly with generic text editors

or (for images composed of pixels) generic paint programs or (for drawings) some widely available drawing editor, and that is suitable for input to text formatters or for automatic translation to a variety of formats suitable for input to text formatters. A copy made in an otherwise Transparent file format whose markup, or absence of markup, has been arranged to thwart or discourage subsequent modification by readers is not Transparent. An image format is not Transparent if used for any substantial amount of text. A copy that is not "Transparent" is called **Opaque**.

Examples of suitable formats for Transparent copies include plain ASCII without markup, Texinfo input format, LaTeX input format, SGML or XML using a publicly available DTD, and standard-conforming simple HTML, PostScript or PDF designed for human modification. Examples of transparent image formats include PNG, XCF and JPG. Opaque formats include proprietary formats that can be read and edited only by proprietary word processors, SGML or XML for which the DTD and/or processing tools are not generally available, and the machine-generated HTML, PostScript or PDF produced by some word processors for output purposes only.

The "**Title Page**" means, for a printed book, the title page itself, plus such following pages as are needed to hold, legibly, the material this License requires to appear in the title page. For works in formats which do not have any title page as such, "Title Page" means the text near the most prominent appearance of the work's title, preceding the beginning of the body of the text.

The "**publisher**" means any person or entity that distributes copies of the Document to the public.

A section "**Entitled XYZ**" means a named subunit of the Document whose title either is precisely XYZ or contains XYZ in parentheses following text that translates XYZ in another language. (Here XYZ stands for a specific section name mentioned below, such as "**Acknowledgements**", "**Dedications**", "**Endorsements**", or "**History**".) To "**Preserve the Title**" of such a section when you modify the Document means that it remains a section "Entitled XYZ" according to this definition.

The Document may include Warranty Disclaimers next to the notice which states that this License applies to the Document. These Warranty Disclaimers are considered to be included by reference in this License, but only as regards disclaiming warranties: any other implication that these Warranty Disclaimers may have is void and has no effect on the meaning of this License.

2. VERBATIM COPYING

You may copy and distribute the Document in any medium, either commercially or noncommercially, provided that this License, the copyright notices, and the license notice saying this License applies to the Document are reproduced in all copies, and that you add no other conditions whatsoever to those of this License. You may not use technical measures to obstruct or control the reading or further copying of the copies you make or distribute. However, you may accept compensation in exchange for copies. If you distribute a large enough number of copies you must also follow the conditions in section 3.

You may also lend copies, under the same conditions stated above, and you may publicly display copies.

3. COPYING IN QUANTITY

If you publish printed copies (or copies in media that commonly have printed covers) of the Document, numbering more than 100, and the Document's license notice requires Cover Texts, you must enclose the copies in covers that carry, clearly and legibly, all these Cover Texts: Front-Cover Texts on the front cover, and Back-Cover Texts on the back cover. Both covers must also clearly and legibly identify you as the publisher of these copies. The front cover must present the full title with all words of the title equally prominent and visible. You may add other material on the covers in addition. Copying with changes limited to the covers, as long as they preserve the title of the Document and satisfy these conditions, can be treated as verbatim copying in other respects.

If the required texts for either cover are too voluminous to fit legibly, you should put the first ones listed (as many as fit reasonably) on the actual cover, and continue the rest onto adjacent pages.

If you publish or distribute Opaque copies of the Document numbering more than 100, you must either include a machine-readable Transparent copy along with each Opaque copy, or state in or with each Opaque copy a computer-network location from which the general network-using public has access to download using public-standard network protocols a complete Transparent copy of the Document, free of added material. If you use the latter option, you must take reasonably prudent steps, when you begin distribution of Opaque copies in quantity, to ensure that this Transparent copy will remain thus accessible at the stated location until at least one year after the last time you distribute an Opaque copy (directly or through your agents or retailers) of that edition to the public.

It is requested, but not required, that you contact the authors of the Document well before redistributing any large number of copies, to give them a chance to provide you with an updated version of the Document.

4. MODIFICATIONS

You may copy and distribute a Modified Version of the Document under the conditions of sections 2 and 3 above, provided that you release the Modified Version under precisely this License, with the Modified Version filling the role of the Document, thus licensing distribution and modification of the Modified Version to whoever possesses a copy of it. In addition, you must do these things in the Modified Version:

A. Use in the Title Page (and on the covers, if any) a title distinct from that of the Document, and from those of previous versions (which should, if there were any, be listed in the History section of the Document). You may use the same title as a previous version if the original publisher of that version gives permission.

B. List on the Title Page, as authors, one or more persons or entities responsible for authorship of the modifications in the Modified Version, together with at least five of the principal authors of the Document (all of its principal authors, if it has fewer than five), unless they release you from this requirement.

C. State on the Title page the name of the publisher of the Modified Version, as the publisher.

D. Preserve all the copyright notices of the Document.

E. Add an appropriate copyright notice for your modifications adjacent to the other copyright notices.

F. Include, immediately after the copyright notices, a license notice giving the public permission to use the Modified Version under the terms of this License, in the form shown in the Addendum below.

G. Preserve in that license notice the full lists of Invariant Sections and required Cover Texts given in the Document's license notice.

H. Include an unaltered copy of this License.

I. Preserve the section Entitled "History", Preserve its Title, and add to it an item stating at least the title, year, new authors, and publisher of the Modified Version as given on the Title Page. If there is no section Entitled "History" in the Document, create one stating the title, year, authors, and publisher of the Document as given on its Title Page, then add an item describing the Modified Version as stated in the previous sentence.

J. Preserve the network location, if any, given in the Document for public access to a Transparent copy of the Document, and likewise the network locations given in the Document for previous versions it was based on. These may be placed in the "History" section. You may omit a network location for a work that was published at least four years before the Document itself, or if the original publisher of the version it refers to gives permission.

K. For any section Entitled "Acknowledgements" or "Dedications", Preserve the Title of the section, and preserve in the section all the substance and tone of each of the contributor acknowledgements and/or dedications given therein.

L. Preserve all the Invariant Sections of the Document, unaltered in their text and in their titles. Section numbers or the equivalent are not considered part of the section titles.

M. Delete any section Entitled "Endorsements". Such a section may not be included in the Modified Version.

N. Do not retitle any existing section to be Entitled "Endorsements" or to conflict in title with any Invariant Section.

O. Preserve any Warranty Disclaimers.

If the Modified Version includes new front-matter sections or appendices that qualify as Secondary Sections and contain no material copied from the Document, you may at your option designate some or all of these sections as invariant. To do this, add their titles to the list of Invariant Sections in the Modified Version's license notice. These titles must be distinct from any other section titles.

You may add a section Entitled "Endorsements", provided it contains nothing but endorsements of your Modified Version by various parties—for example, statements of peer review or that the text has been approved by an organization as the authoritative definition of a standard.

You may add a passage of up to five words as a Front-Cover Text, and a passage of up to 25 words as a Back-Cover Text, to the end of the list of Cover Texts in the Modified Version. Only one passage of Front-Cover Text and one of Back-Cover Text may be added by (or through arrangements made by) any one entity. If the Document already includes a cover text for the same cover, previously added by you or by arrangement made by the same entity you are acting on behalf of, you may not add another; but you may replace the old one, on explicit permission from the previous publisher that added the old one.

The author(s) and publisher(s) of the Document do not by this License give permission to use their names for publicity for or to assert or imply endorsement of any Modified Version.

5. COMBINING DOCUMENTS

You may combine the Document with other documents released under this License, under the terms defined in section 4 above for modified versions, provided that you include in the combination all of the Invariant Sections of all of the original documents, unmodified, and list them all as Invariant Sections of your combined work in its license notice, and that you preserve all their Warranty Disclaimers.

The combined work need only contain one copy of this License, and multiple identical Invariant Sections may be replaced with a single copy. If there are multiple Invariant Sections with the same name but different contents, make the title of each such section unique by adding at the end of it, in parentheses, the name of the original author or publisher of that section if known, or else a unique number. Make the same adjustment to the section titles in the list of Invariant Sections in the license notice of the combined work.

In the combination, you must combine any sections Entitled "History" in the various original documents, forming one section Entitled "History"; likewise combine any sections Entitled "Acknowledgements", and any sections Entitled "Dedications". You must delete all sections Entitled "Endorsements".

6. COLLECTIONS OF DOCUMENTS

You may make a collection consisting of the Document and other documents released under this License, and replace the individual copies of this License in the various documents with a single copy that is included in the collection, provided that you follow the rules of this License for verbatim copying of each of the documents in all other respects.

You may extract a single document from such a collection, and distribute it individually under this License, provided you insert a copy of this License into the extracted document, and follow this License in all other respects regarding verbatim copying of that document.

7. AGGREGATION WITH INDEPENDENT WORKS

A compilation of the Document or its derivatives with other separate and independent documents or works, in or on a volume of a storage or distribution medium, is called an "aggregate" if the copyright resulting from the compilation is not used to limit the legal rights of the compilation's users beyond what the individual works permit. When the Document is included in an aggregate, this License does not apply to the other works in the aggregate which are not themselves derivative works of the Document.

If the Cover Text requirement of section 3 is applicable to these copies of the Document, then if the Document is less than one half of the entire aggregate, the Document's Cover Texts may be placed on covers that bracket the Document within the aggregate, or the electronic equivalent of covers if the Document is in electronic form. Otherwise they must appear on printed covers that bracket the whole aggregate.

8. TRANSLATION

Translation is considered a kind of modification, so you may distribute translations of the Document under the terms of section 4. Replacing Invariant Sections with translations requires special permission from their copyright holders, but you may include translations of some or all Invariant Sections in addition to the original versions of these Invariant Sections. You may include a translation of this License, and all the license notices in the Document, and any Warranty Disclaimers, provided that you also include the original English version of this License and the original versions of those notices and disclaimers. In case of a disagreement between the translation and the original version of this License or a notice or disclaimer, the original version will prevail.

If a section in the Document is Entitled "Acknowledgements", "Dedications", or "History", the requirement (section 4) to Preserve its Title (section 1) will typically require changing the actual title.

9. TERMINATION

You may not copy, modify, sublicense, or distribute the Document except as expressly provided under this License. Any attempt otherwise to copy, modify, sublicense, or distribute it is void, and will automatically terminate your rights under this License.

However, if you cease all violation of this License, then your license from a particular copyright holder is reinstated (a) provisionally, unless and until the copyright holder explicitly and finally terminates your license, and (b) permanently, if the copyright holder fails to notify you of the violation by some reasonable means prior to 60 days after the cessation.

Moreover, your license from a particular copyright holder is reinstated permanently if the copyright holder notifies you of the violation by some reasonable means, this is the first time you have received notice of violation of this License (for any work) from that copyright holder, and you cure the violation prior to 30 days after your receipt of the notice.

Termination of your rights under this section does not terminate the licenses of parties who have received copies or rights from you under this License. If your rights have been terminated and not permanently reinstated, receipt of a copy of some or all of the same material does not give you any rights to use it.

10. FUTURE REVISIONS OF THIS LICENSE

The Free Software Foundation may publish new, revised versions of the GNU Free Documentation License from time to time. Such new versions will be similar in spirit to the present version, but may differ in detail to address new problems or concerns. See https://www.gnu.org/copyleft/.

Each version of the License is given a distinguishing version number. If the Document specifies that a particular numbered version of this License "or any later version" applies to it, you have the option of following the terms and conditions either of that specified version or of any later version that has been published (not as a draft) by the Free Software Foundation. If the Document does not specify a version number of this License, you may choose any version ever published (not as a draft) by the Free Software Foundation. If the Document specifies that a proxy can decide which future versions of this License can be used, that proxy's public statement of acceptance of a version permanently authorizes you to choose that version for the Document.

11. RELICENSING

"Massive Multiauthor Collaboration Site" (or "MMC Site") means any World Wide Web server that publishes copyrightable works and also provides prominent facilities for anybody to edit those works. A public wiki that anybody can edit is an example of such a server. A "Massive Multiauthor Collaboration" (or "MMC") contained in the site means any set of copyrightable works thus published on the MMC site.

"CC-BY-SA" means the Creative Commons Attribution-Share Alike 3.0 license published by Creative Commons Corporation, a not-for-profit corporation with a principal place of business in San Francisco, California, as well as future copyleft versions of that license published by that same organization.

"Incorporate" means to publish or republish a Document, in whole or in part, as part of another Document.

An MMC is "eligible for relicensing" if it is licensed under this License, and if all works that were first published under this License somewhere other than this MMC, and subsequently incorporated in whole or in part into the MMC, (1) had no cover texts or invariant sections, and (2) were thus incorporated prior to November 1, 2008.

The operator of an MMC Site may republish an MMC contained in the site under CC-BY-SA on the same site at any time before August 1, 2009, provided the MMC is eligible for relicensing.

ADDENDUM: How to use this License for your documents

To use this License in a document you have written, include a copy of the License in the document and put the following copyright and license notices just after the title page:

> Copyright © YEAR YOUR NAME. Permission is granted to copy, distribute and/or modify this document under the terms of the GNU Free Documentation License, Version 1.3 or any later version published by the Free Software Foundation; with no Invariant Sections, no Front-Cover Texts, and no Back-Cover Texts. A copy of the license is included in the section entitled "GNU Free Documentation License".

If you have Invariant Sections, Front-Cover Texts and Back-Cover Texts, replace the "with … Texts." line with this:

> with the Invariant Sections being LIST THEIR TITLES, with the Front-Cover Texts being LIST, and with the Back-Cover Texts being LIST.

If you have Invariant Sections without Cover Texts, or some other combination of the three, merge those two alternatives to suit the situation.

If your document contains nontrivial examples of program code, we recommend releasing these examples in parallel under your choice of free software license, such as the GNU General Public License, to permit their use in free software.