GEOU9SP GIS Workbook

Thiago Silva

2025-12-07

Table of contents

Welcome!

This e-book was created as a lab companion for the *GEOU9SP - Geographic Information Systems* module at University of Stirling. It contains all the practical exercises that will be covered throughout the module.

This book is **not a replacement to Canvas**, so make sure you are engaging with the material on Canvas on a weekly basis!

The content of this book is organised into sections, which will correspond to the teaching weeks and practical lab sessions. You will always see the main content in the center window, with a detailed menu covering all entries in the book on the left, and a chapter-specific table of contents on the right. The Canvas pages for each lab session will always link to the corresponding chapter in this book.

If you have any questions follow the instructions on Canvas on how to contact the module coordinator.

You can download PDf versions of each week's material here:

Week 1 Week 2 Week 3

Part I

Week 1 - Introduction

In our first week, we will get acquainted to with basic GIS data manipulation using the QGIS software.

ILOs covered

1. Understand the structure of spatial data and choose appropriate data types and models for storing and representing it;
2. Obtain and assess the quality of spatial data from online and offline sources and produce new spatial data using computer and field methods;
3. Create map visualisations that adhere to cartographic principles and can be easily and unambiguously interpreted by the non-specialist public;

What will you learn

For every week, we will list the main theoretical and practical learning goals. Use these as a 'checklist' to gauge your learning for each week. If you don't feel confident you have learned any specific topic, then revisit the week's material!

Theoretical knowledge for Week 1:

- What are Geomatics, GIS, Remote Sensing and associated terms and fields?
- What is spatial data?
- What is a datum?
- What is a map projection?
- What is a Coordinate Reference System?

Practical knowledge:

Chapter 1

- How to launch QGIS
- How to load a spatial data file
- How to navigate the main QGIS interface
- How to identify features
- How to do basic measurements
- How to do basic styling of spatial data
- How to enter the layout editor
- How to make and export a simple map layout

Chapter 2

- How to check the Coordinate Reference System of a spatial file
- How to set the CRS of a QGIS project
- How to set the CRS of a spatial file when it is already known.
- How to reproject(convert) a spatial file from one CRS to another
- What are the main CRSs you should be familiar with.

1 Lab 1: QGIS overview and file management

The purpose of this lab is to give you a first overview of what a proper GIS workflow looks like, from start to end. As you progress on your exercises the projects will become more complex but the general workflow will not change.

Developing proper project and file management habits from the start is the *best* thing you can do to succeed in GIS. Speaking as someone who has been teaching and working with Geomatics for more than a decade, poor file/project management is the underlying cause of at least 50% of the GIS problems you may encounter.

1.1 Before you start!

1. Go through the Week 1 preparatory session on Canvas, and watch the seminar recording if you have missed it.
2. Read this document to understand how the British National Grid indexing system works.

1.2 Guided Exercise 1 - Managing and loading files

In this exercise, you will download some data, prepare a *folder structure* to organise the files in your GIS projects, then create a QGIS project file and add some GIS data to it.

1.2.1 Downloading the required data

For this practical you will use the "OS open road" and "OS Terrain 50 Digital Terrain Model" spatial datasets, specifically for the **NS89** tile of the Ordnance Survey British National Grid. (Read about what these datasets are here: OS Open Roads , OS Terrain 50).

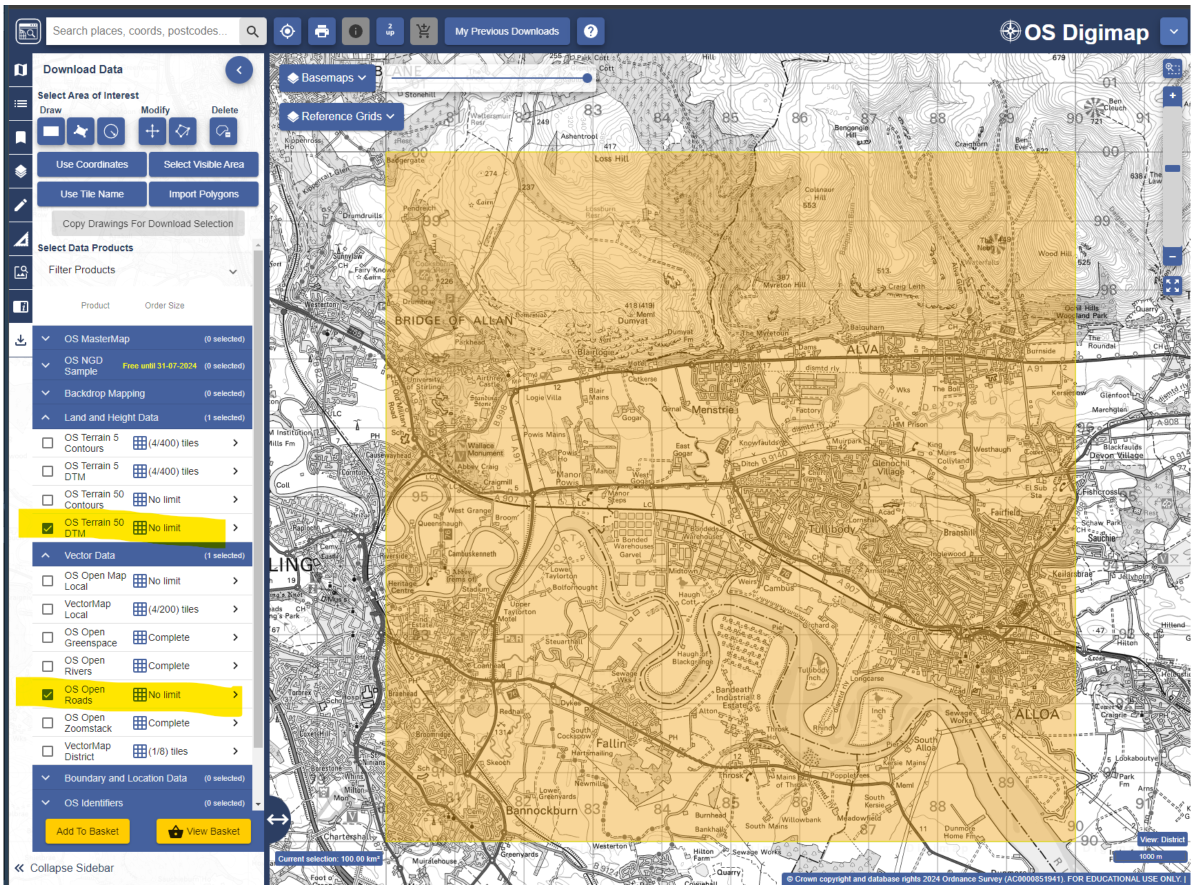
(1) Head to the Digimap website, and if you haven't already, make sure you accept the user licenses for all datasets, as show on the instructions video avaiable on Canvas.

(2) Go the `Ordnance Survey` section of the site, and then pick the `View maps and download data` option on the right.

(3) Following the steps shown on the instructions video, download the following datasets *for the NS89 BNG tile only*:

- OS Open Roads (under *Vector Data*) - choose the `SHP` format
- "OS Terrain 50 Digital Terrain Model" (*NOT Terrain 5 and NOT Contours!*, under Land and Height Data) - choose the `ASC` format

If you have correctly searched for tile NS89 and checked the required datatsets, you will see a screen like this:

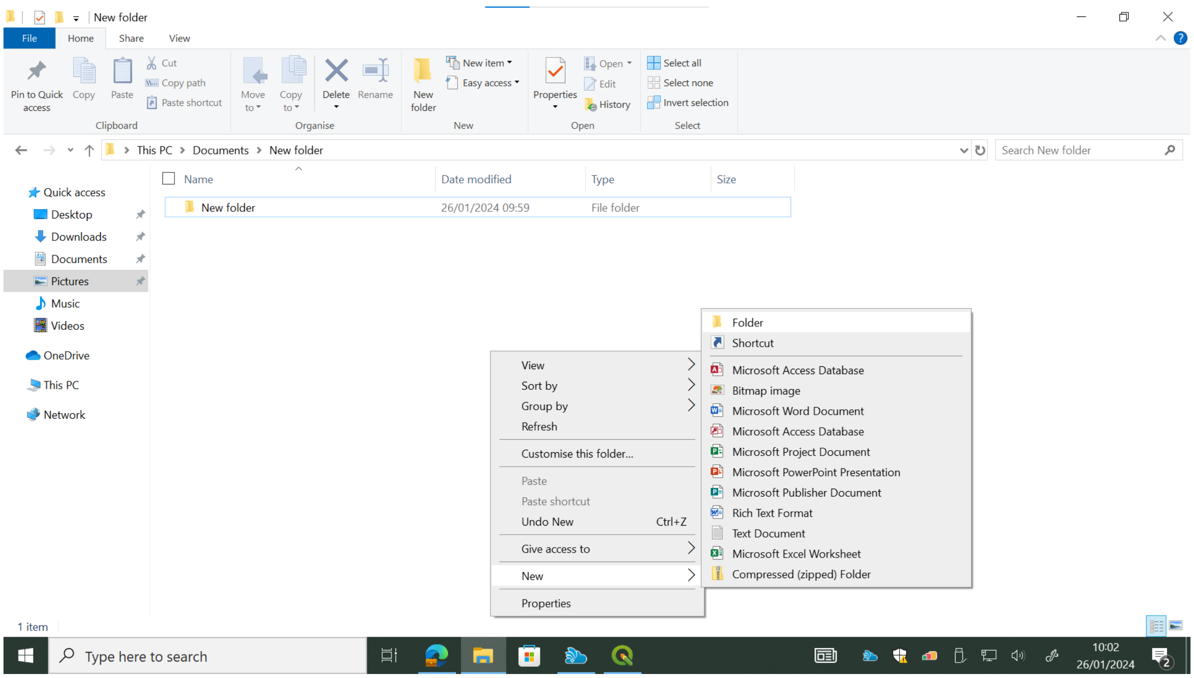


1.2.2 Creating a project structure

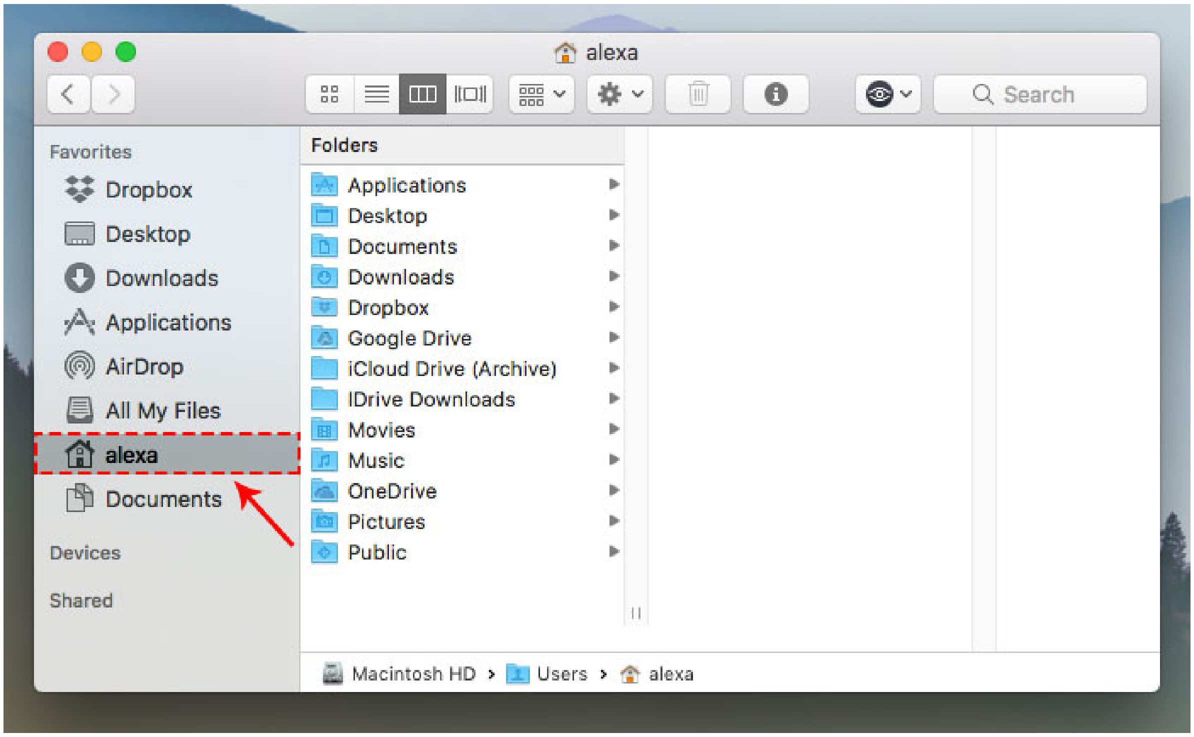
GIS projects generate a lot of different files, and quickly, so project organization is **essential**. The folder structure below is my personal suggestion for organizing GIS projects files and associated data. As you get comfortable managing your own projects, feel free to change the organization structure to something that best suits your own workflow and the specific project you are working on!

(4) Create a folder on your computer with the module name (`GEOU9SP`), off of your main work folder. For **Windows** this will usually be `Documents`. For **Linux** and **Mac**, create it on you `home` folder. (If you prefer and are comfortable with saving at another location on your computer, go for it!)

Windows:



mac OS:



Warning

- Having complex file paths can sometimes create problems. That is why we try to keep our main projects on a base folder. If you would like to keep a copy of your data on your OneDrive folder you are encouraged to do so, but I don't recommend working directly from it. That is because the full path to a OneDrive folder will to be very long and likely contain spaces. Mine, for example is `C:\Users\Thiago\OneDrive - University of Stirling\Documents`. Those spaces in the name may create problems later. So I recommend instead to *copy* the completed lab data to your OneDrive folder at the end of each lab.

- For the same reason, **never** use spaces or special symbols on your folder and file names. Limit yourself to using letters A-Z and a-z, numbers 1-9 and just the *underline* (_) and *dash* (-) symbols.

- Windows **does not** differentiate letter case, so if you create a file named `filename1.txt` and then a second file named `Filename1.txt` in the same folder, the second will overwrite the first. Mac OS does differentiate upper and lower case letters, so `Filename1.txt` and `filename1.txt` are considered different names and can co-exist in the same folder.

(5) Inside the new `GEOU9SP` folder, create a subfolder called `lab_1` (notice I am avoiding

spaces by using the underline)

(6) Inside `lab_1`, create the following folder structure. On the diagram below, the folder `raster` is a folder inside the folder `01_raw_data`, which is inside `lab_1` and so on.

```
.
 GEOU9SP/
     lab_1/
         00_qgis
         01_raw_data/
             raster
             vector
         02_processing
         03_final_products
         04_notes
```

I like to start folder names start with double digits so they are kept in order. These folders will be used as follows:

- `00_qgis`: we will use this folder to save our QGIS project files.
- `01_raw_data`: this folder will keep all the original data files you download. This way you can always go back to the start if something goes wrong. You can optionally use the subfolders `vector` and `raster` to easily know which kind of data you are working with, but this may not be necessary for small projects. As you obtain data you will create additional subfolders for each individual dataset to keep data organized.
- `02_processing`: here we will keep all the intermediate files you generate as part of your work. Make ample use of subfolders to identify each step of the workflow.
- `03_final_products`: here we will keep the final products of our intended analysis. This makes it easy for us to find the latest version of our final results, without risking confusing it with intermediate files. This can also include maps, reports and any other 'deliverable'resulting from your analysis.
- `04_notes`: here we will keep all our *non-GIS* files. For example, it may be a good idea to keep a text file documenting the project steps as you work on it. You could also save here screen grabs of specific steps. Again, use subfolders as needed to keep your data tidy.

(7) Now move the data you've downloaded into your organized project folders and extract (unzip) them if necessary. The terrain data folder (`terrain-50-dtm_xxxxxxx`) should go into the `raster` folder within `01_raw_data` and the roads folder (`open-roads_xxxxxxx`) into the `vector` folder. The `citations_orders.xxxx.txt` and `contents_order_xxx` text files should go into your `04_notes` folder. Your folder structure should look like this:

```
.
  GEOU9SP/
      lab_1/
          00_qgis
          01_raw_data/
              raster/
                  OS_terrain_50/
                      ...
              vector/
                  OS_roads/
                      ...
          02_processing
          03_final_products
          04_notes
```

! Stop and Think

What are the `citations_orders.xxxx.txt` and `contents_order_xxx` files? What information do they contain?

Click for answer

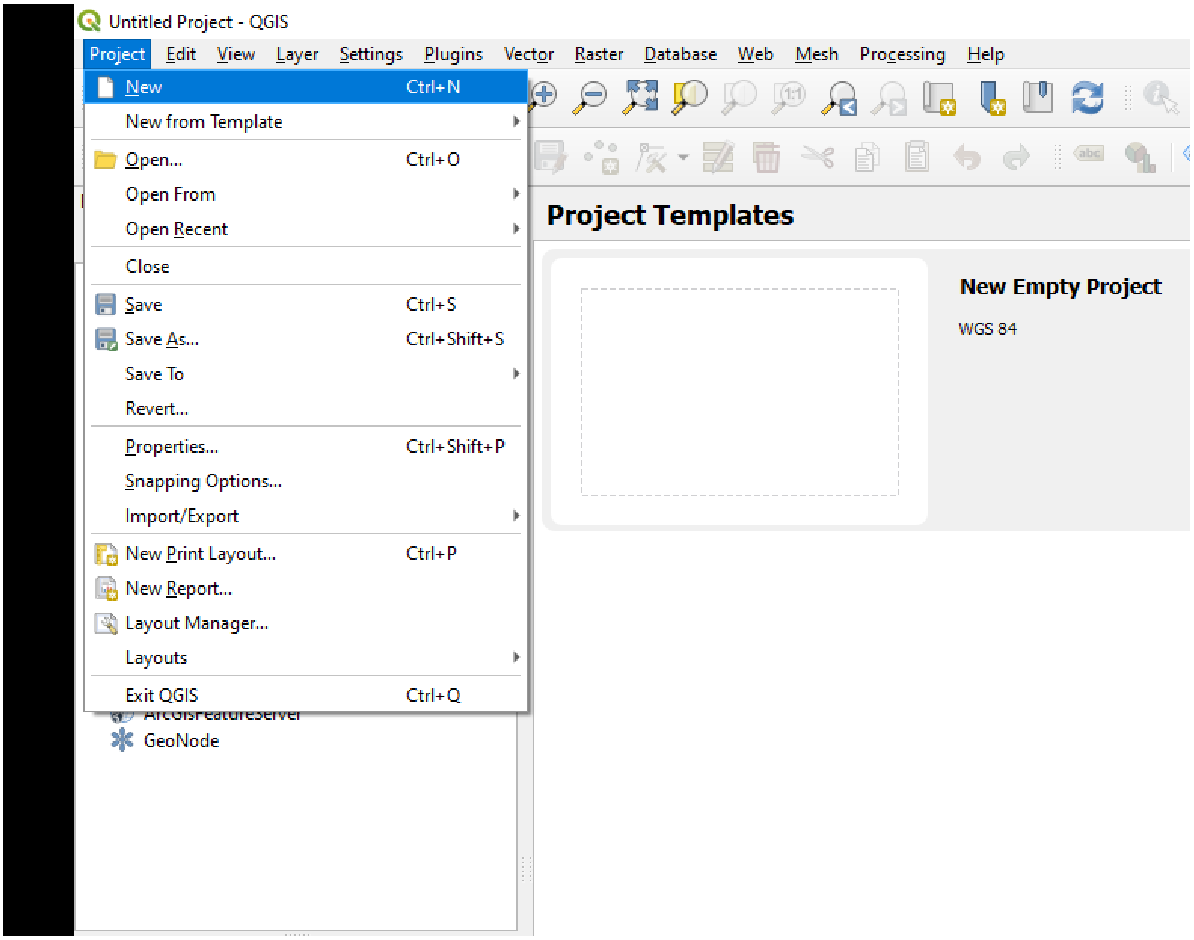
They are both plain text files. `citations_orders.xxxx.txt` is very handy as it gives us the proper way to cite the data sources on reports. `contents_order_xxx` is just a summary of your Digimap order.

(8) Start a new text file (a Word *.doc* file or a Notepad *.txt* file) in your `04_notes` folder, and write a few lines documenting the steps you took until now. If you prefer handwritten notes on a paper notebook, feel free to use it instead! The important thing is to keep track of the steps you are taking. As projects get more complicated, it is easy to forget which steps we took, and in what order!

1.2.3 Creating a QGIS project and adding data to it

A QGIS project is a file, which will remember all the data layers you have loaded, their stacking order, the styling of each layer, and some other information such as the coordinate reference system. It will also keep any map layouts that you create. **But he project file will NOT store the data files themselves!!**. It will only link to the data. That is why we need good project folder organisation. If a data file is moved or renamed, the QGIS project file will lose track of it.

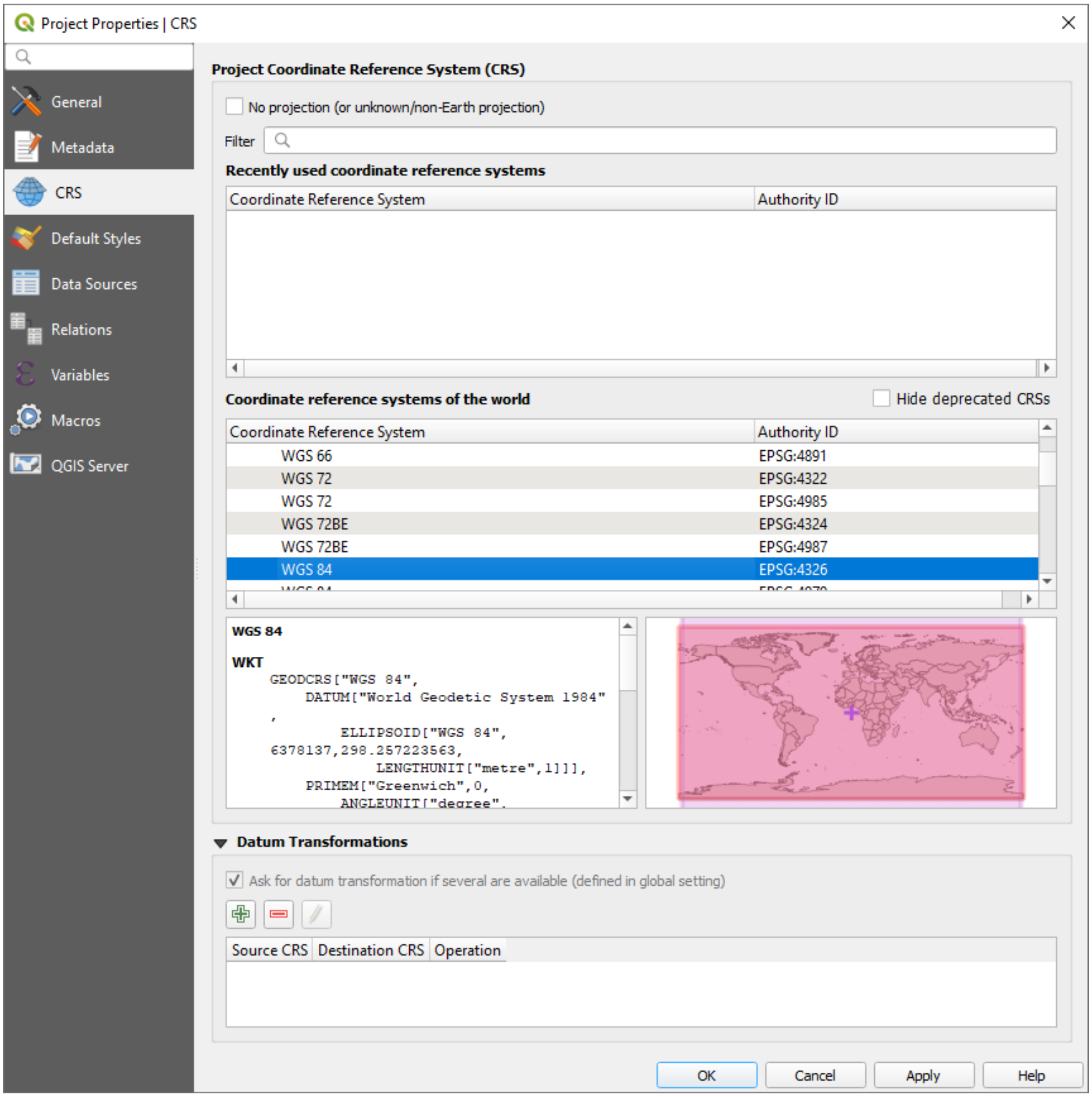
6. Create a new project in QGIS by clicking on `Project > New...` or pressing *Ctrl-N*:



(9) Add to your project the layers `NS_RoadLink.shp` from the OS Roads dataset and the `NS89.asc` terrain data from your organized project folders. Don't worry if the extents don't match - it's just that the OS DEM and the OS Roads are distributed in different sized 'chunks'.

(10) Save your project inside the `00_qgis` folder. You can save your project by clicking on the *Save* button, going to `Project > Save` menu, or by holding the keys `Ctrl+S (Command+S on Mac)`.

(11) Open the project settings by clicking on the menu `Project > Properties...`. You will then see this window:



(12) The main things to set on your new project are the project home folder, the coordinate reference system (CRS) and the measurement units.

- As the *project home* on the `General` tab, select the `lab_1` folder you created. This helps quick navigation when opening and saving data.
- For *measurement units*, make sure distance units are set in meters, and area units in squared meters, also on the `General` tab. Any QGIS operations that calculate distances or areas will use the units set by the project.

- For the CRS, go to the `CRS` tab (to the left) and search for **27700** in the `Filter` box. That is the *EPSG code* that identifies the *OSGB 1936 British National Grid* coordinate reference system (CRS). Select it by clicking on it - you may need to expand your search results by clicking on the small arrows besides `Projected` and `Transverse Mercator`.

> **Note**
>
> EPSG stands for "European Petroleum Survey Group", and designates a database with standard codes for hundreds of coordinate reference systems. Over time, you will probably memorize the EPSG codes for the CRSs you use more often.

(13) Save your project again.

1.3 Guided Exercise 2 - Organizing and styling your layers

One of the most powerful aspects of GIS software is the ability to *style* spatial data in very specific ways, by specifying colors, line widths, line types, simbols, etc. As we progress in the module you will learn more and more ways to style your data.

(14) On the `Layers panel` to the left of the screen, select the `NS89` layer, and drag it to the bottom of the layers list if not there already.

(15) Turn the roads layer off for now, by unchecking the box besides its name.

(16) Right-click on the `NS89` layer name and choose `Rename Layer`. Rename it to `Digital Elevation Model (50m)`.

(17) Open the file explorer in your system, and find the `NS89.asc` file that holds the actual elevation data, which you downloaded from Digimap.

> **Stop and Think**
>
> Will changing a layer name in the layer panel also change the name of the source data file for that layer?

> **Click for answer**
>
> No, layer names within the project are independent of file names - but as default QGIS will use the file name as layer name when you add new data. But you should always change them into nice, readable and properly spelled names within your project, where you are free to use spaces. The actual file name linked by the project will stay the same and can always be seen by right-clicking on the layer name and selecting `Properties...> Information`.

(18) Right click on the newly renamed terrain layer and choose `Zoom to Layer`. This is a very handy tool to "find yourself" if you end up zooming or panning the map too far.

(19) Right click on the terrain layer name again and select `Properties...`, then go to the `Symbology` tab.

(20) On the `Min` and `Max` boxes, type `0` and `500` respectively. This determines the range of layer values to be visualised.

(21) Select `Rendering type` to be `Single Band Pseudocolor`, and change the `Color ramp` option by clicling on the down-arrow button to the right and picking the `spectral` colour ramp. If you click on the colour ramp itself it will open a new window to custmize it. This is now what we need for now, so just close it and use the down-arrow.

(22) Click on `Classify` (under the dropdown menu that says * Mode: continuous*). The colours will be matched to the range of elevation values in the layer.

(23) Click again on the little down-arrow button beside the colour palette button and select `Invert Color Ramp` at the top of the options, so that the lowest heights are coloured blue. Then click on `OK`.

(24) Note how the legend for you terrain layer has changed on the Layers panel. it now shows the minimum and maximum elevations and the colour ramp. Save your project.

! Stop and Think

Why do we bother inverting the colour ramp for this dataset?

Click for answer

We should always try to use colours that reinforce map interpretation. The color blue is usually associated with water, and water accumulates on the lowest elevations, so setting the lowest elevations to blue helps map users read and interpret the map.

(25) Rename the `NS_RoadLink` layer to `Road Network`.

(26) Go to its `Symbology` properties, like you did for the terrain layer, Notice that the symbology options are data-specific (vector vs. raster). Seclet `Simple Line` in the symbology window (under the main `Line` option). Then under the window, go to the `Color` option an click on the down-arrow menu to change the line color to a dark grey. If you click on the actual colour, a more complex colour-picking window will appear. You can use either the quicker colour dropdown or the main colour window, whatever you prefer.

(27) Change the `Stroke width` (line width) to 0.3. Click `OK`. Reactivate the layer if needed to visualise it.

1.4 Guided Exercise 3: Processing data using GIS operations

The core of GIS work is to use the many built-in operations (also known as functions or tools) of GIS software to *process* the data in some way, and thus create additional information. For this exercise, we will use an operation that creates a new layer representing the boundaries of the terrain data layer, and then use a second operation to cut the roads layer to the same shape and extent as this new layer.

(28) Go to the `Vector` menu and select `Research Tools > Extract Layer Extent...`. Select the terrain layer as your `Input layer`, and click `Run` to generate a *temporary layer*. Temporary layers are not kept once you close QGIS, unless you save it manually later. The `Extract Layer Extent` window will not close automatically after you run the operation, so click on `Close` when you are done.

(29) Go to `Vector > Geoprocessing Tools > Clip...`. Select the roads layer as the `Input Layer` and the new temporary layer as the `Overlay Layer`.

(30) This time, we will save the output. Click on the `...` button to the right of the `Clipped` text box, and then choose `Save to file`. Save your new layer on the folder `GEOU9SP/lab_1/02_processing/`, naming it `clipped_roads.shp`. Make sure the `SHP file` format is selected below the file name. Then click on `Run` to execute the operation. Close the window.

(31) Turn the original roads layer on and off to see the result of your operation. Then right-click on the original roads layer, and select `Styles > Copy Style > All Style Categories`. Then right click on the new (`Clipped`) roads layer and select `Styles > Paste Style > All Style Categories`. This is a great way top style several layers ion the same way wirthout effort.

(32) Remove the original roads layer from your project by right clicking on it and selecting `Remove Layer...`, and then rename you clipped layer to “Roads”. Save your project.

(33) Close QGIS. **It will give you a warning** - read it carefully and then confirm it.

(34) Reopen QGIS, and load back your project. Notice it remembers exactly where you last saved it, including zoom level, layer names and layer styles. The “Extent” layer will still be on your list, but will appear empty - it has been deleted once you closed QGIS. Remove it from the project and save again.

! Stop and Think

a) What are the names of the two GIS functions you just used in this exercise?

b) Why did you have to create a new layer representing the extent of the terrain data before *clipping* the roads layer?

c) What does the warning given by QGIS when you closed your project means?

Click for answer

a) The functions are called `Clip` and `Extract Layer Extent`.

b) The `Clip` function requires two *vector* files as input, but the terrain data is a *raster* file (i.e. an image). The `Extract Layer Extent` creates a vector file representing the extent of any given layer. Each specific tool will require different kinds of data to work properly. We will learn more about vectors and rasters on weeks 2 and 3.

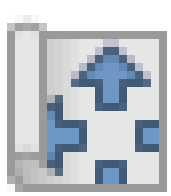
c) When we created the extent layer, we produced a *temporary layer*, which is discarded by QGIS when the program is closed. QGIS was letting you know that will happen, to give you the chance to go back and save it. Temporary layers are also lost when QGIS crashes (yes when, not if - it will happen). So never use them for important stuff - only for quick tests if you are not sure what the output of a function will be.

1.5 Guided Exercise 4: Creating a map layout

The main QGIS interface (or any other GIS software) is developed and optimised for interactive work. But very often as part of our GIS analysis we will want to generate nice maps and figures following proper design rules (and not just grabbing a screen capture of the QGIS window and dumping it on a page - hint for your assignment reports). For that, we use the QGIS **Map Layout Editor**.

(35) Click on `Project > New Print Layout` and name it `Lab 1 Layout`. A new window will open showing the QGIS Layout Editor. Notice that the main QGIS window remains open as well. These windows are 'linked', so that the map layout reflects any styling changes you make on the main window.

(36) Add a new map to the layout by clicking on the  icon. Drag it through the page so it covers about 2/3 of it horizontally, and the full height of the page (minus some borders).

(37) Use the `Interactive Extent` tool  to pan (click and grag) and zoom (mouse wheel) until your data covers the entire map box. But make sure you don't hide

the edges of the data by zooming in too much!

(38) Now fine tune the map scale by changing the `Scale` value on the bottom right panel. Remember that this value means “1:value”, i.e. one unit on the page is equal to that many units (value) in the real world. This means larger numbers will “zoom out”, and smaller numbers will “zoom in”. Try to make the map fill as much as possible of the map box, without clipping the edges.

(39) Go to the menu `Add Item > Add Legend...` in the Map Layout Window (or guess the icon for this option on the side toolbar). Click the area beside the map box and drag to add a legend.

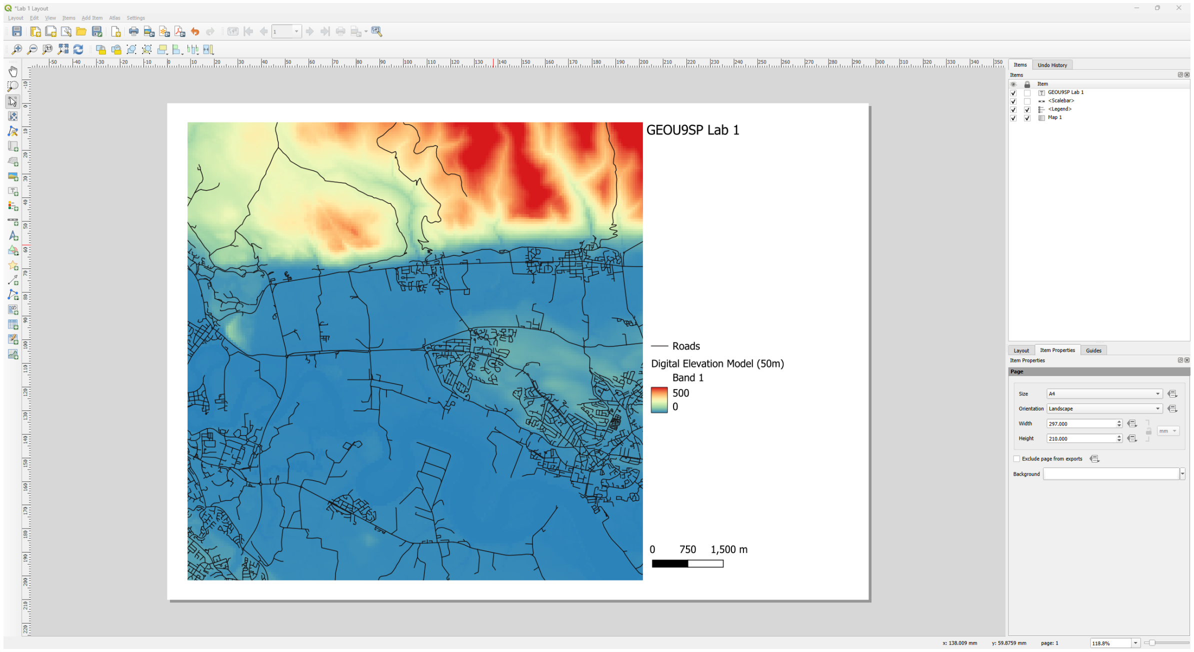
(40) Go back to the main QGIS window, right click on the terrain layer name, then select `Properties...`, and on the `Symbology` tab, change the `Mode` under the class color box to `Equal Interval`. Then select the number of `Classes` to 10 (to the right of the `Equal Interval` box). Go back top the Map Layout editor.

(41) Go to the menu `Add Item > Add Scalebar` in the Map Layout Window (or guess the icon for this option on the side toolbar). Click and drag in the area below the legend to add it to the map layout.

(42) Add a title to your map using the `Add Label...` tool (either on the `Add Item` menu or selecting the tool directly from the left sidebar). You can change the text by replacing the “Lorem ipsum” placeholder text with your own text on the left pane. Try to make it **bold** with a font size of 16 (hint: always use the down-arrow buttons for ‘quick settings’).

(43) Rearrange the items on the page until you are pleased with the results. Then go to `Layout > Export as PDF...`, and export your map, naming it properly and saving it on `GEOU9SP/lab_1/final_products`. It is a good idea to export as PDF if you want your map to be a standalone document. If you are exporting to insert the map into a report, then use `Layout > Export as Image...` instead.

For reference, my map looked like this at this stage:



(44) Then close the Map Layout window, save your project on the main QGIS window, and close QGIS. Reopen QGIS, and go to `Project > Layout Manager`. The layout you created previously should appear on the list. Select it and click on `Show` to reopen the Map Layout window.

Congratulations! You have successfully finished your first GIS project, using proper file management practices! As a final suggestion, create a "workflow_notes" file on `GEOU9SP/lab1/notes` and write up a quick overview of what you did, along with any specific notes you would like to remember later.

1.6 Independent Exercise 1

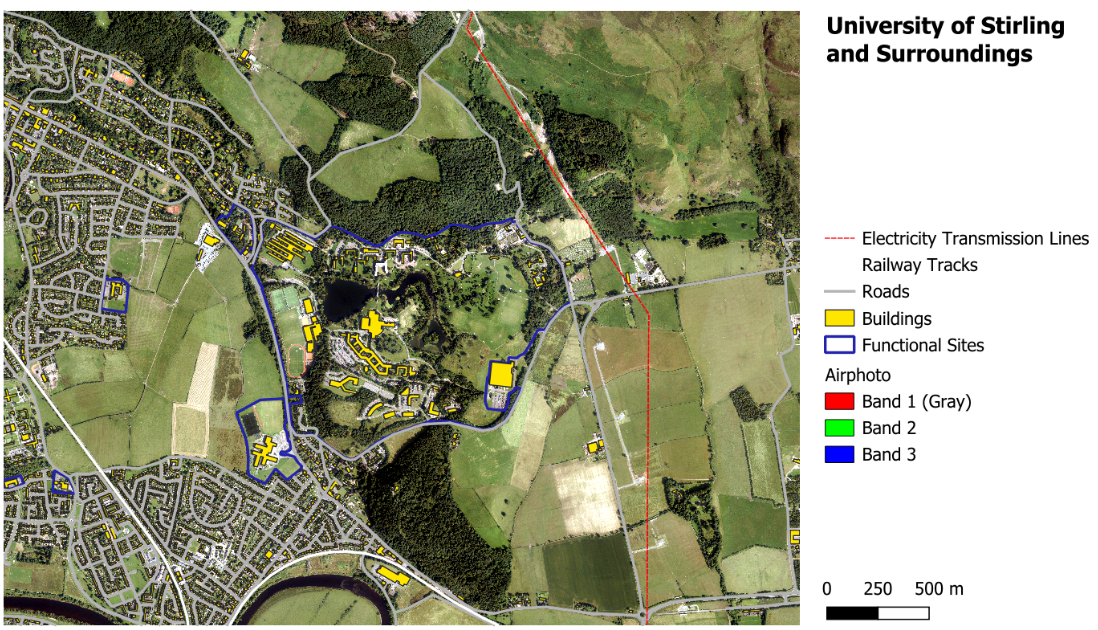
At the end of each lab, you will have the opportunity to reinforce what you learned by going through 'independent' exercises. These will use the same operations you learned with the guided exercises, but it may require you to figure out small bits of new functionality, and will not have step by step instructions.

As the module progresses, the independent exercises will give you less and less directions, to reflect real-world GIs usage. **Make sure you do the independent exercises** - it is easy to fall into a false sense of '*I got this*' when only following step-by-step directions.

1. Create a new project folder structure for this project, under your main GEOU9SP work folder.

2. Download the zipfile containing the Air Photo Mosaic and the Stirling Council Geospatial Data from here.
3. Extract the data from the zipfiles and move them to the proper project folders.
4. Open QGIS and create a new project that uses the OSGB British Grid (EPSG 27700) coordinate reference system.
5. Import the airphoto raster image and the vector files for Buildings, Roads, Railway Track, Electricity Transmission Lines, and Functional Sites into your QGIS project.
6. Organize your layers so that, from top to bottom, you have transmission lines, railways, roads, buildings, functional sites, and then the airphoto image.
7. Style your layers so that transmission lines are styled as thin dashed red lines, railways as thick white lines, roads as thick grey lines, buildings as filled yellow polygons and functional sites with a thick dark blue outline and no fill colour (hint: check the `Fill style` options).
8. `Clip` all your vector layers to the extent of the airphoto image.
9. Create a map layout covering the entire extent of the airphoto. Add a title, legend and scale bar to your layout and export it as a PNG image. Then insert your exported file as a figure into a MS Word document.

Click here to see what the final layout should look like.



10. Now answer the questions below:

- How far is the Pathfoot Building from the Cottrell Bulding? Calculate it both “as the crow flies” (linear distance) and as if you were walking. (Hint: use the Measuring tool).
- What is the total surface area of the water bodies in the University of Stirling Campus? (Hint: also the Measuring tool).
- What is the classification of the UoS campus polygon within the “Functional Site” layer? (Hint: use the Indentify Features tool).

2 Lab 2: Coordinate Reference Systems

The purpose of this lab is to help you understand why we need to pay attention to Coordinate Reference Systems (CRS) when working with spatial data. CRS's are what make data *spatial* - they associate the actual data to locations on the surface of the Earth (or other planets!). But there are dozens of CRS's in existence, each adapted for a specific world region and purpose. So quite often you will obtain spatial data in different coordinate systems, which can cause problems if not normalised before analysis.

2.1 Guided Exercise 1: Understanding Coordinate Reference Systems

In this exercise, you will learn how to use QGIS to identify the coordinate reference system (CRS, sometimes wrongly called as just "projection") of spatial data you acquire, and also how to manage data and project coordinate reference systems.

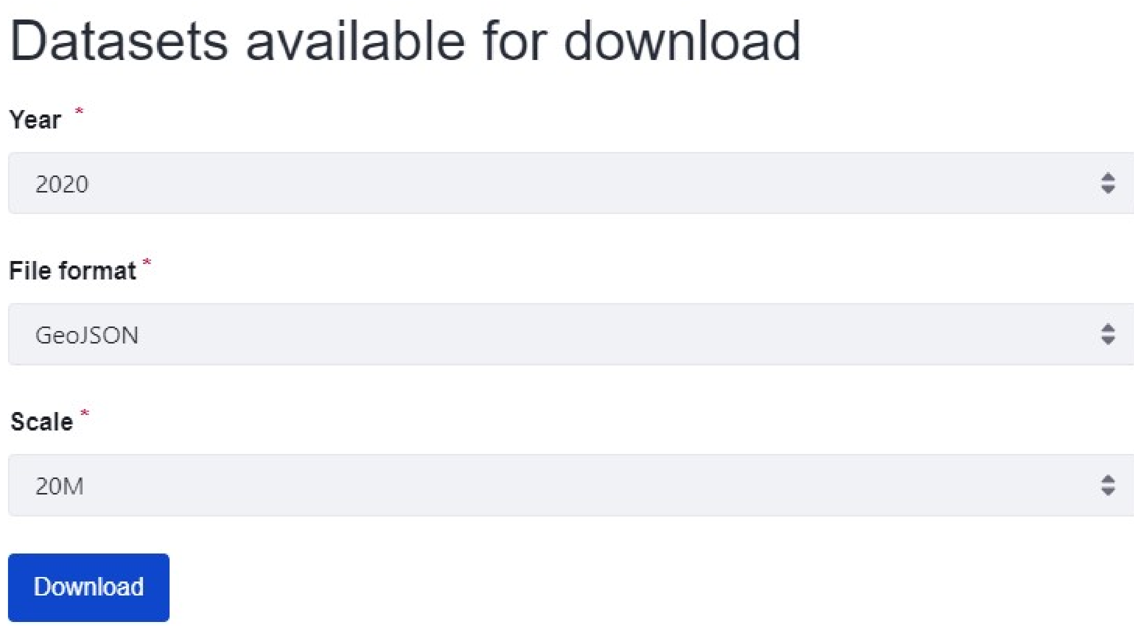
> **!** Stop and Think
>
> Why are coordinate reference systems and 'projections' not the same thing?

> **Click for answer**
>
> Coordinate reference systems combine a **datum**, which defines a geometric representation of the Earth's shape and how it 'intersects' with the real surface of the Earth, a **coordinate system** (for example latitude and longitude or northings and eastings) and a **map projection** which is a set of mathematical rules to project the 3D surface of the datum's *ellpisoid* into a flat plane (such as a screen or a map).

2.1.1 Obtaining the required data

(45) Download the **2020** country boundary data in **GeoJSON** format, at the **1:20** million scale, from the link below. The GeoJSON format is a more recent GIS file format, commonly used for web-based mapping. It is derived from the JavaScript Object Notation (JSON) data file format, widely used to exchange data among websites and web servers. More info here.

https://ec.europa.eu/eurostat/web/gisco/geodata/administrative-units/countries



! Stop and Think

a) What is the source of the data you are downloading? Does it seem reliable?

b) What are the conditions (provisions) of use for the data?

Click for answer

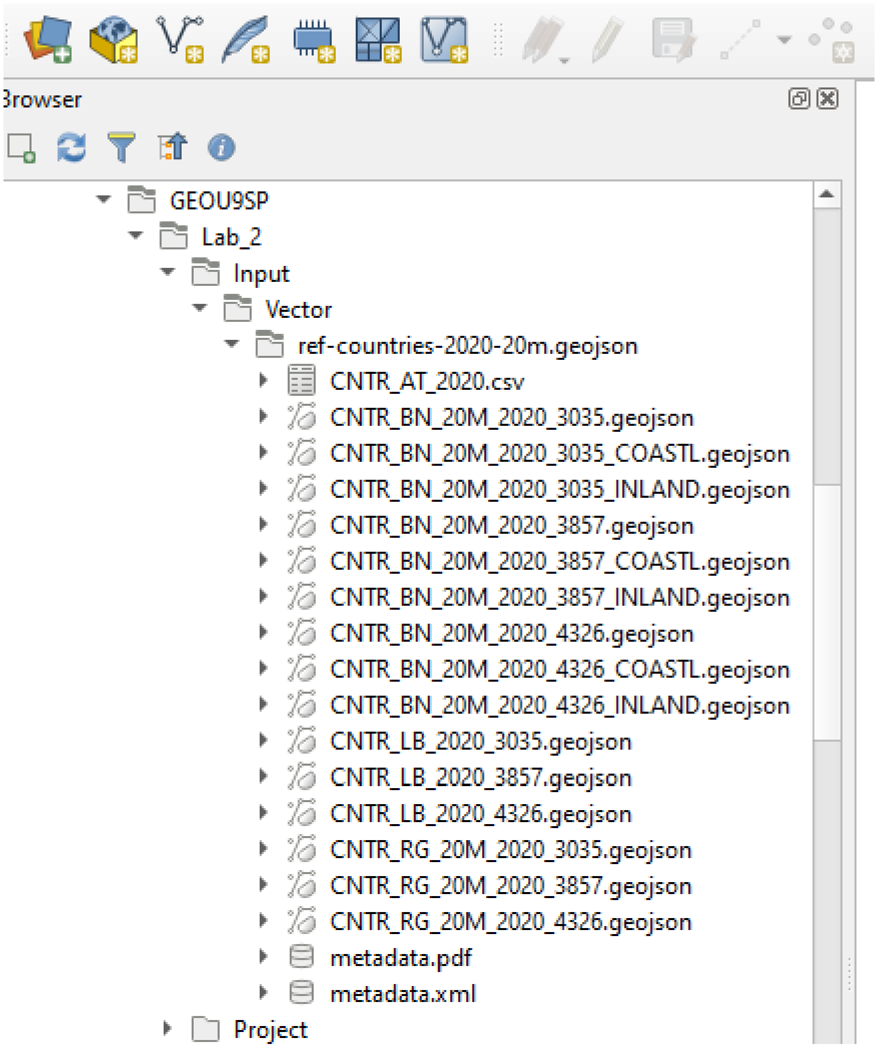
a) The page providing the data is managed by Eurostat, the statistical office of the European Union. Therefore you would be inclined to trust in the quality and correctness of the data provided.

b) The webpage presents a link to "dowload rules' which describe the authorised uses of the data. This informartion can also be found in the data's *metadata* file.

(46) Create a lab_2 folder in your GEOU9SP main folder. Then create a simple folder structure to organise the data, and extract the zipped downloaded data in its proper location. If you are not sure how to do this, please revisit Chapter 1.

(47) Open QGIS and start a new project. Save it as `lab_2` in its proper folder. Then look at the contents of the folder holding the country data, using the QGIS browser panel. If the panel is not available, you can enable it by going to the `View > Panels` menu and

checking the box for `browser`. You should see several layers on tghe folder:



(48) Load the `CNTR_BN_20M_2020_4326` file into your project. Pay attention to the file name as there are many files with similar names.

! Stop and Think

Good file names are always informative of their content. Can you gess the contents of the different GeoJSON files you have downloaded based on their names?

Click for answer

All file names start with `CNTR` for 'countries', followed by a two letter code. `BN` seems to stand for ‘boundary’(vector lines), `RG` for ‘region’(vector polygons), and `LB` for ‘labels’ (vector points). Then `2020` specified the reference year, and `20M` indicates the 1:20 million mapping scale. the final four-letter number indicates the EPSG code for the data CRS: 4326 (‘unprojected’ WGS84), 3857 (WGS 84 with Pseudo-Mercator projection) or 3035 (ETRS89-extended / Lambert Azimuthal Equal Area for Europe).

2.1.2 Visualising data with different CRS

(49) Set the symbology for the outline and fill as you prefer. Try to manipulate more visual variables than just colour.

! Stop and Think

Why can’t you set a fill color for the countries?

Click for answer

Because the `BN` files are vector lines, not polygons. Lines only have one dimension, and thus the inner parts of the countries in this dataset are actually empty. If you load the `RG` dataset instead, you can set the fill, as vector polygons represent 2D areas.

(50) Right-click on this layer’s name and go to `Properties > Information`.

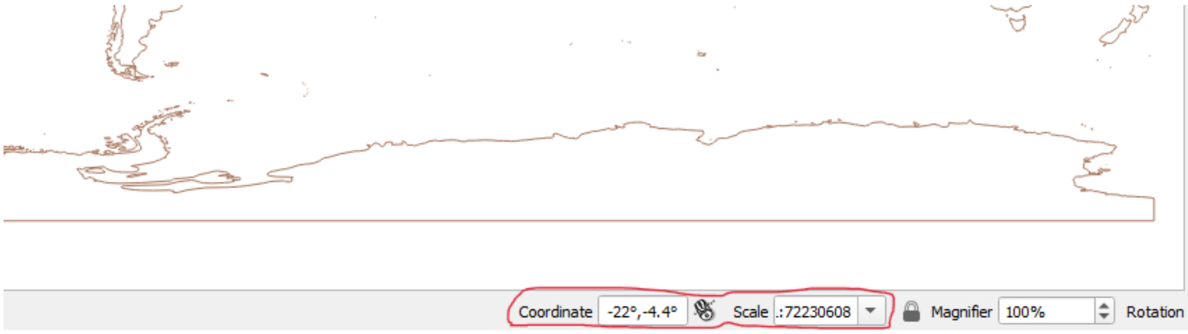
! Stop and Think

What is the Coordinate Reference System (CRS) for this dataset?

Click for answer

The information tab will have a section called Coordinate Reference System, as below: Name: EPSG:4326 - WGS 84 Units: Geographic (uses latitude and longitude for coordinates) Type: Geographic (2D) Method: Lat/long (Geodetic alias) Celestial Body: Earth Accuracy: Based on World Geodetic System 1984 ensemble (EPSG:6326), which has a limited accuracy of at best 2 meters. Reference: Dynamic (relies on a datum which is not plate-fixed)

(51) Note, on the bottom QGIS status bar, that as you move your mouse pointer around, the coordinates for the mouse position are updated in real time. Also note what the map scale is, how it changes as zoom in and out. You can also type the second part of a scale nuber to zoom at the desired map scale (for example 50000 if you want to see the map at a 1:50000 scale)



! Stop and Think

a) Why doesn't the scale shown on the bar match the "advertised" scale for the dataset (1:10 million)?

b) The box on the very bottom right of the QGIS status bar tells you what is the current project CRS. How is it different from a layer CRS?

Click for answer

a) The 20M scale refers to the scale used when digitising the coastline, i.e., what is the 'closest' you can view the dataset without loss of detail.

b) The project CRS defines the 'viewing' CRS for the map canvas. Any data that uses a different CRS than the project will be re-projected 'on the fly' to match the CRS of the project - but continue with the exercise to learn why that can be a problem.

(52) Zoom to the UK in the shown layer. Note how the scale at the bottom status bar changes with your zoom.

! Stop and Think

Does the shape of the UK look "right" to you? If not, what is the issue and what is the cause?

Click for answer

The UK looks 'squished' vertically. That is because the data is being viewed in the EPSG 4326 ('unprjected' WGS84) CRS. EPSG 4236 uses what is effectively the *Plate Carrée* or *Equidistant cylindrical* projection, the simplest possible map projection - lat and long degrees are just linearly converted to x,y coordinates. This projection does not preserve area nor shape (conformal) and increasingly distorts features as you approach the poles.

(53) Click on the project projection box at the bottom right of the status bar (or go to `Project > Properties... > CRS tab`). On the `Filter` text box, search for EPSG:3035. Select this projection for the project and click `OK`. A warning box will appear, make sure you read it through before selecting `OK` again.

! Stop and Think

a) what is the name of the Coordinate Reference System specified by EPSG 3035?

b) What did the warning window warned you about?

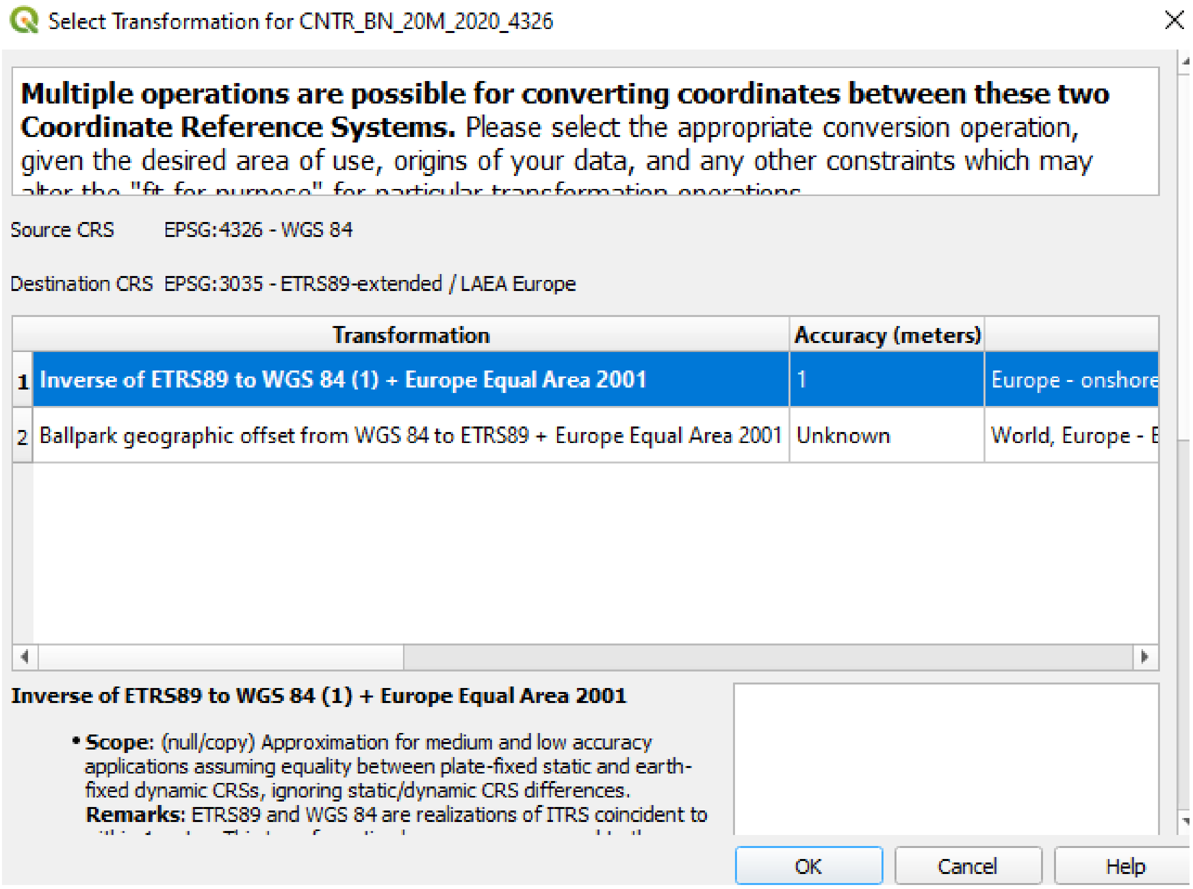
Click for answer

a) EPSG 3035 is called ETRS89-extended / LAEA Europe and is the official projection for cartographic data from the European Union. It uses the European Terrestrial Reference System 1989 datum and the Lambert Azimuthal Equal Area projection, which preserves areas and can be considered conformal for Europe.

b) It warned you that there is more than one option for the 'on the fly' reprojection of your layer from ESPG 4326 to EPSG 3035. It showed you the options with the most accurate (1m) selected by default. But this option is only valid for Europe.

It is important to not "freak out" when an unexpected warning or error appear. Take a breath, and read through the window or error message, most often the explanation is right there. You just have to dare to look.

If you did click through it without looking, here is a screen capture of it:



(54) Look at the shape of the UK again after changing the project CRS. Then right click on the layer name and select "Zoom to Layer(s)."

! Stop and Think

a) How does the rest of the world look now? Why?

b) When you move your mouse, what unit are the coordinates in?

Click for answer

a) As you move further away from the centre of the projection, shape gets progressively more distorted. This is because the LAEA projection is only conformal at its centre. But areas are all correct.

b) In metres. You can check that on the layer's `Properties > Information` tab.

(55) Now add to the project the file called CNTR_BN_10M_2020_3035.geojson. Notice the different last four numbers on the file name.

! Stop and Think

What is the CRS for this new layer, and how well dos it visually align with the p[revious layer.

Click for answer

This second layers uses the EPSG 3035 CRS, while the previous layer used EPSG 4326. Notice how ech layer maintains its original CRS when added to the project, but if necessary they are reprojected 'on the fly' to match up visually.

(56) Change the project CRS back to EPSG:4326.

(57) Now go back to the project CRS properties and check the box that says `No CRS` at the top of the window. This disables the on-the-fly projection. Then click OK and go back to your map.

(58) Right click on the 4326 layer and select `Zoom to Layer`. Then select the zoom out tool (the loupe with a minus sign) at the top button row, and start clicking at the centre of the map. Keep clicking as it gets really smalll - you should click about 17 times until the second dataset is fully visible. Check the properties of each layer to make sure they still have the same CRS of when you loaded them.

! Stop and Think

What has happened? Why are the two datasets suddenly very different in size?

Click for answer

Since you turned off on-the-fly reprojection, each dataset is now drawn at their original coordinates - but one is in meters and the other in degrees, so their x,y positions and scale become very different.

2.1.3 Potential issues with using mismatched data

(59) Download this vector shapefile (link), unzip it and add it to your QGIS project. Check what is the CRS of this layer.

(60) Set the Project CRS back to EPSG:3035. Then go to the top menu bar and select `Vector > Geoprocessing > Clip`. Select the layer that has the EPSG 3035 projection as your

`Input` Layer, and the new "clip_bounds" layer as your `Overlay` Layer. You can just leave the output as a temporary file. Click Run.

(61) Turn off the visibility of all layers except the new "Clipped" layer to see the result of the Clip operation.

(62) Rename the "Clipped" layer to "Clipped_3035" by right clicking on it and selecting `Rename layer`. Then repeat the `Clip` operation, this time selecting the 4326 world layer as `Input`, and "clip_bounds" as `Overlay` again. Rename the result to "Clipped_4326."

Using what you learned on the previous lab activities, pick two contrasting colours for each "Clipped_..." layer, and make the lines thicker. Zoom in at the lines of each clipped layer and check if they overlap perfectly.

> **!** Stop and Think
>
> Why are the clipping results different even though the initial 3035 and 4236 layers looked perfectly aligned?

> **Click for answer**
>
> Because although they have been reprojected visually to match on screen, the files still have different CRSs, and this affects the lining up of the `Input` and overlay layers. To get proper results, you need permanently translate (reproject) both datasets to the same CRS.

(63) Now go to `Vector > Data Management Tools > Reproject Data`. Select the 4326 world layer as your `Input Layer`, and EPSG:3035 as your `Target CRS`. Let the result be a temporary file and click `OK`. The new layer will be automatically named as "Reprojected". What is the CRS of this new layer (check on the layer properties window)?

(64) Now repeat the use of the `Clip` tool using "Reprojected" as the `Input Layer` and "clip_bounds" as the `Overlay Layer`. Rename the resulting layer to 'Clipped_Reprojected'. Which of the two originally clipped layers ("Clipped_3035" or "Clipped_4236?) better matches the"Clipped_Reprojected" layer?

> **!** Stop and Think
>
> What does the `Reproject` operation do?

> **Click for answer**
>
> It actually applies a permament mathematical transformation to the coordinates of the input data, translating the data from one CRS to another. For any GIS project that involves multiple data layers with different CRSs, you should pick the CRS that makes

> more sense for your project as the 'project CRS', and then reproject all layers to the same CRS before anything else.

This is the end of Lab 2! You shoudl now understand why different datasets may have different Coordinate Reference Systems, what are the problems with working with data that has mismatched CRSs, and how to reproject data to match a given CRS. As this is your first week, there are no additional independent exercises - lets take it easy!

If you still want to practice more, check the exercises from the QGIS Training Manual!

Part II

Week 2 - Vector data

In our second week, we will take a deeper dive into understanding GIS spatial vector data. This is the most common data model you will encounter while working with GIS, so make sure you take your time to understand it fully.

ILOs covered

1. Understand the structure of spatial data and choose appropriate data types and models for storing and representing it;
2. Obtain and assess the quality of spatial data from online and offline sources and produce new spatial data using computer and field methods;
3. Create map visualisations that adhere to cartographic principles and can be easily and unambiguously interpreted by the non-specialist public;
4. Plan and execute GIS analytical steps to solve spatial problems successfully;

What will you learn

Use this list as a 'check-list' to gauge your learning for each week. If you don't feel confident you have learned any specific topic, then revisit the week's material!

Theoretical knowledge for Week 2:

- What are data models vs data formats (file formats) vs data types (attribute types)?
- What is the vector data model?
  - How are vectors represented?
  - What are the components of a vector file (geometry + attribute)?
- What are the main vector file formats?
- What are the main data types we use to represent attributes?
- What kinds of data can we represent using vector data?
- What kinds of questions can we answer using attributes?
- What kinds of questions can we answer using geometries?
- What kinds of questions can we answer using both (i.e. geometry-based attribute calculations).

Practical knowledge:

Chapter 3

- How to identify a vector file format
- How to load a vector file
- How to use the identify tool to get information on the fly
- The concept of 'selecting' a feature.
- How to select features using the click tools in QGIS
- How to access the attribute table
- How to use the statistical summary tool
  - Global summaries vs summaries on selected data
- How to select using simple queries (filter) on attributes
  - Boolean operators
  - String based operators
- How to use attributes to set symbology
  - Single symbol
  - Categorised
  - Graduated

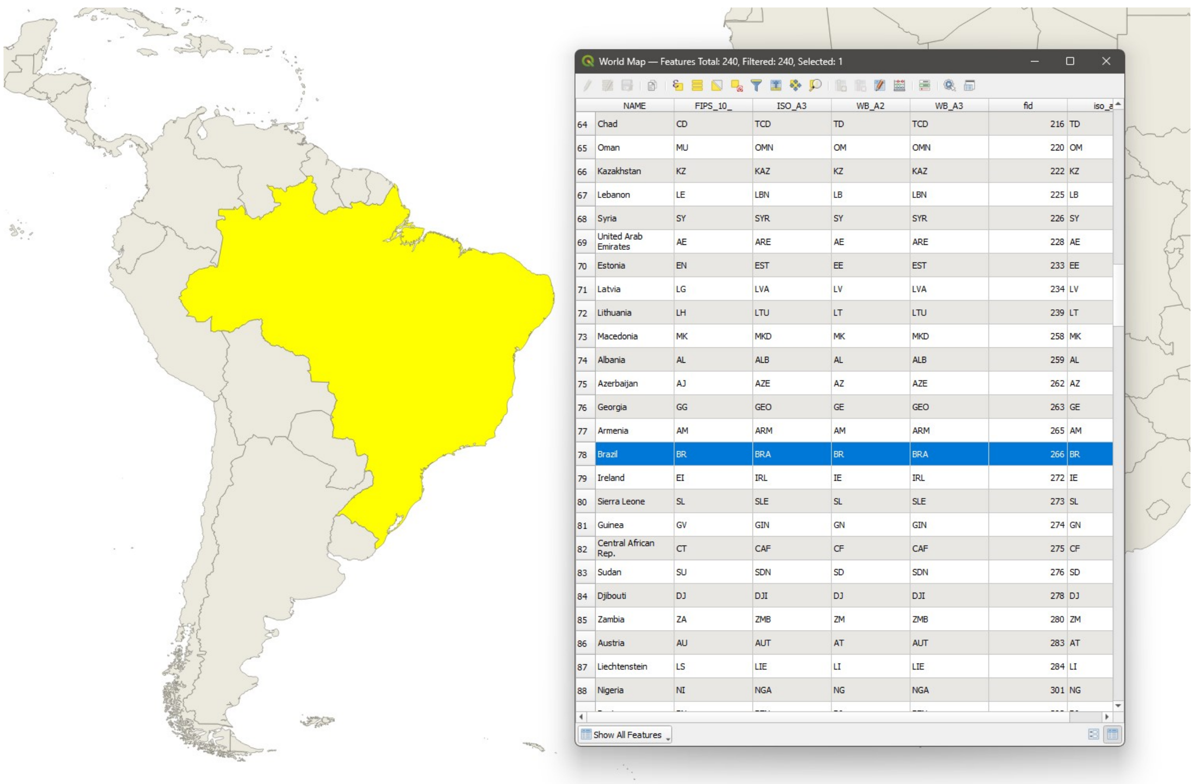
Chapter 4

- How to do combined attribute queries (AND / OR)
- How to calculate new attributes
  - Numeric attributes
  - String attributes
  - Geometry attributes
- How to convert data types if needed
- Attribute selection vs attribute filtering
- How to select based on location
  - The different selection options
  - Importance of CRS matching
  - Select by location vs clipping -How to combine spatial and attribute selection
  - Selecting within selections or saving intermediate results

3 Lab 3: Working with vector attributes

Vector data consists of discrete observations called features. For example, on a vector layer representing all protected areas in Scotland, each individual protected area would comprise a feature. With regular data, observations are just rows on a table, but with vector spatial data, features are composed of two elements: the geometry (the visual component + coordinates) and attributes (columns on a table holding data about each feature, also known as *fields*).

On the figure below, each country is a feature - the shape of the country is represented by the geometry (left) and the corresponding *attributes* are shown in the right. Notice that both the geometry and the corresponding attributes of one specific feature (Brazil) are *selected*. Selections are a very important component of GIS analysis, as they can narrow down the targets for your calculations.



For this lab, we will focus mainly on the following GIS operations: filtering attributes, summarising attributes, and creating and modifying attribute data. Remember to apply, from here on, all the steps you have already learned in previous labs: create a project folder, organize your data, save a named project with proper CRS info, etc. Each week's labs will build upon the previous activities, so I will not be repeating instructions for things covered in previous sessions.

3.1 Before you start!

1. Go through the Week 2 preparatory session on Canvas, and watch the seminar recording if you have missed it.

3.2 Guided Exercise 1 - Basic work with Vector Data

In this exercise, you will learn how to open and read vector data.

(65) Download the data for this exercise from here, then extract the zip files (make sure all the data are unzipped! Sometimes you have zipfiles inside zipfiles...) and organise them to your preference (use a similar folder structure as from week 1).

(66) Load all three datasets in QGIS and inspect them. The file names are: `global_earthquakes_2011.gpkg`, `MajorRivers.shp`, and `ne_50m_admin_0_countries.shp`.

! Stop and Think

a) What are the datasets you have?

b) What are the *data models* used by these datasets?

c) What are the *file formats* you have to work with?

d) Which of these files contain *metadata* about your datasets?

e) What is the *CRS* of each data layer you have?

Click for answer

a) You can use the file names, the metadata, and the visual appearance of the layers to answer data. The `global_earthquakes_2011.gpkg` file seems to hold point data on recorded Earthquake locations in the year 2011. The `MajorRivers.shp` layer seems to hold line data on the world's largest rivers. The `ne_50m_admin_0_countries.shp` seems to have boundaries for the world countries.

b) All three files use the *vector* data model, being a point vector, line vector and polygon vector respectively.

c) The earthquakes layer is given in the geopackage file format, while the other two layers are given in the shapefile file format. Notice that data model and file format are two different things, and they don't necessarily imply each other. Geopackages, for example, can hold both *vector* and *raster* data.

d) The rivers layer has a plain text metadata file with a link that points to the source of the data, where more information can be found. The countries layer has an HTML file that holds metadata about the file. The earthquakes data has no metadata.

e) All three layers are using EPSG 4236 (WGS84 'unprojected') as CRS. You check it by right clicking on the layer name and selecting `Layer CRS`, or by selecting `Properties > Information`.

(67) Check your Project properties to make sure they are correct (does the project CRS match the layers? What are the measurement units set for this project? Have you set the base folder?). Then save your project file within your folder structure as in previous labs.

(68) Now inspect the *attribute table* for each layer, by right-clicking on each layer name and then on `Open Attribute Table`. Then answer the following questions:

! Stop and Think

a) How many *features* does each layer have?

b) How many *attributes* does each feature have?

Tip: if you go to `Properties > Fields`, you get a list of all attributes ordered by ID, which is a sequential number. That makes it easier to count attributes when there are many.

Click for answer

a) Earthquakes: 15272; Rivers: 98; Countries: 241.

b) Earthquakes: 5 (fid, Event, latitude, longitude, Magnitude, Date); Rivers: 4 (NAME, SYSTEM, MILES, KILOMETERS); Countries: 94 (featurecla, scalerank, etc...).

If you looked at `Properties > Fields`, the last ID is 93, but the first one is 0, so 94 in total.

(69) Rename your layers on the layer list to human-readable, informative names, then save your project. Remember, these new names will appear within this project only, the name of the source files in your folders will not change.

(70) Organize the layer order and play with different layer superpositions to make the most readable visualisation for all datasets involved. Then experiment with the symbology of each layer to improve your visualization.

3.3 Guided Exercise 2: Visualising layer attributes

One of the main applications of vector data is the ability to *select* and then *summarise* the different attributes of each layer to extract relevant information.

(71) Turn off the "Rivers" and "Earthquakes" layers. (Tip: to turn multiple layers off or on at once, highlight all the layers you need to make hidden (or visible) by holding down the `Control` key of your keyboard while you click, then hit the spacebar on your keyboard).

(72) Go to `Layer Properties > Symbology` for the World Countries layer, and change the top option from `Single Symbol` to `Categorized`. This lets you assign different colours based on attribute values. We will colour the countries based on the main region where they are. So for `Value`, choose the `REGION_UN` attribute. Leave the `Symbol` option as is, and for `Color Ramp`, select `Random Colors` if not already. Then click on the `Classify` button on the bottom left of the large white space in the middle of the window. Then click on `OK`.

(73) Look at your layer and notice how it has been styled. Try to manually change the colours of each region to your liking.

Now lets use visualisation to understand the distribution of world population.

(74) Return to the `Symbology` window and select `Graduated` instead of `Categorized`. Change your `Value` to the `POP_EST` attribute. Choose `Magma` as your colour ramp (click on the little arrow to the right), and select the `Invert Color Ramp` option.

(75) Change the classification `Mode` (above the classify button) to `Equal interval`, leave the number of `Classes` as 5 (to the left of `Mode`), and then click on `Classify`, then `OK`.

! Stop and Think

a) What is the difference between the `Categorized` and `Graduated` options?

b) Does the `Equal Interval` classification give a good visualization of the distribution of world population?

Click for answer

a) `Categorized` is for *categorical*, non-numeric variables (i.e. names, classes, etc.). `Graduated` is for *continuous* variables (i.e. quantities, measurements).

b) No, because China, India and to a lesser extent the US have much numbers than the rest, which biases the breakpoints. We will fix it in the next step.

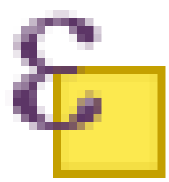
(76) Return to the `Symbology` window, and change the `Mode` from `Equal Interval` to `Natural Breaks (Jenks)`, and increase the number of `Classes` to 10. Click on `OK`.

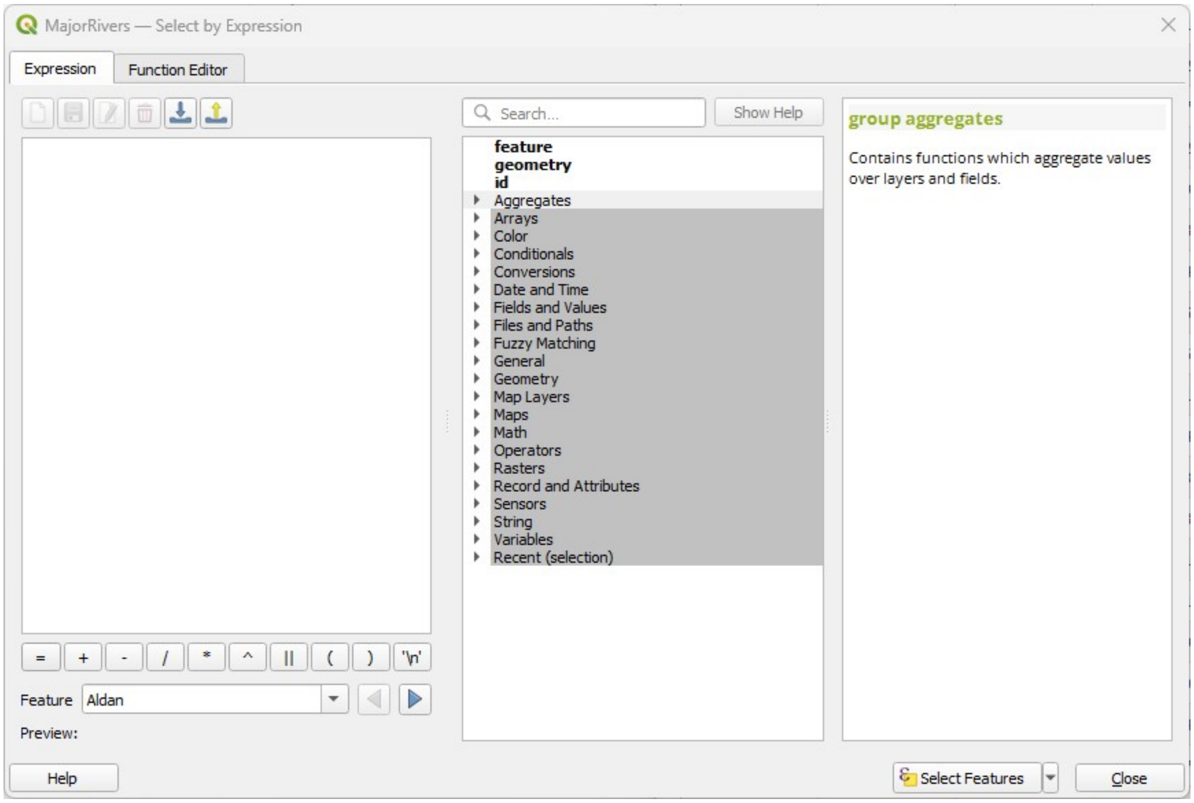
When you are mapping your own data, make sure you explore the different methods for calculating breakpoints, and also play with the number of classes. You can see the ful explanation of each `Mode` on the QGIS Documentation.

3.4 Guided Exercise 3: Selections based on layer attributes

Now we will look into using *expressions* to search and select specific features according to their attributes. As GIS vector data emerged from the database world, these searches as sometimes refered to as *queries*.

(77) Change the symbology of the Countries layer to `Single Symbol`, and pick a dark gray. Then turn on the Rivers layer, change its symbology to a light blue, and make sure it is on top of the Countries layer.

(78) Open the Rivers layer attribute table (right click on its name then on `Open attribute table`). Then click on the `Select features using expression` button (). You will see a new window with three panels, like the one below. If you only see two panels, then click on `Show help`:



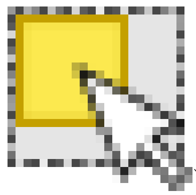
You will encounter this expression window in other parts of QGIS as well. The way it works is that you type your expression on the left window, using the middle and right panels to browse and select operators to add to your expression. It will make more sense with an example:

(79) In the middle panel, expand the `Fields and Values` item and then double click on `KILOMETERS`. It will add this attribute name to the left panel. Notice that it is enclosed by double quotes. In the expression window, **any word between double quotes means it is an attribute's name**. You can also type names directly in the window if you prefer, and QGIS will offer autocomplete suggestions based on the existing attributes.

(80) Now complete the expression by typing the remainder, so that the final expression is `"KILOMETERS" > 5000`. This expression means "Select all River features that have a length - represented by the KILOMETERS attribute - larger than 5000". Then click on the `Select features` button.

(81) Check the results of your selection on the Attribute Table. The number of selected features should be shown on the top of the window (11), and selected features will be highlighted in blue. On the bottom left of the window, you can change from `Show all features` to `Show selected features` if you only want to see the selected features.

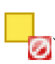
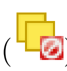
(82) Also check the results of your selection on the Map canvas. All selected features will be highlighted in yellow. Tip: if you ever set the symbology of a feature yellow, remember to not confuse it with selected objects. When in doubt, check the attribute table.

> **!** Stop and Think
>
> It seems the selection is missing a few of the longest rivers in the world, such as the Amazon River. Why would that be?
>
> Hint: try to use the manual selection tool () to click on the Amazon River and investigate.

> **Click for answer**
>
> The different segments of the Amazon River officially receive different names: Amazonas (lower Amazon), Solimões (central Amazon) and Ucayali (upper Amazon). In this dataset, it is broken down into two features: Amazon (with 3042 km) and Ucayali (with 2088 km). Since these are two separate features, neither is selected by our expression.

(83) Return to the attribute table and make sure that the rivers larger than 5000km are still selected. Then create a new expression: `SYSTEM" = 'Amazon'`. Notice that we enclose the word Amazon with single quotes. This identifies this as a *string*, i.e. a character value, and differentiates it from an attribute name. Then instead of clicking on `Select Features`, click on the small arrow to the right and click in `Add to Current Selection`. Now your selection should include all river features that are longer than 5000 km *or* belong to the Amazon system (26 features in total).

Tip: If you can't remember all the possible values of an attribute, select the attribute under `Fields and Values` and then on the right panel, double click on `All Unique`. QGIS will list all possible value options for that particular attribute.

(84) Now deselect all features by clicking on the `Deselect` button in the Attribute Table () or the `Deselect from all layers` button in the main QGIS toolbar (). It is always good to clear selections when you are done with a certain analysis, to avoid unexpected consequences.

We can use several operators to create expressions. For numeric values, we can use all logical operators: 'greater than' (`>`), 'lesser than' (`<`), their 'or equal' variants (`<=`, `>=`) as well as 'equal' (`=`) or 'not equal' (`<>`). For strings (text), `=` and `<>` also work, but you can use the operators `IS` and `IS NOT` (all upper case) instead. In the example above, we could have used `"SYSTEM" IS 'Amazon'` to get the same result.

Another class of useful operators are called *Boolean* operators: `AND`, `OR` and `NOT`. They allow us to create compound expressions with multiple criteria:

(85) Return to the attribute table and this time use the following expression: `"KILOMETERS" > 5000 OR "SYSTEM" = 'Amazon'`. You should get the same results as when you used two separate selections with `Add to Current Selection`. But boolean operators can be more powerful.

(86) Clear your selection and create a new one with the expression `"KILOMETERS" > 1000 AND "SYSTEM" = 'Amazon'`. When you use the `AND` operator, each feature must fulfil *both* criteria (like an intersection in set theory, if you remember your maths). When you use the `OR` keyword, then each feature can fulfil *either* criteria (a mathematical *union*).

(87) Now change the expression to `"KILOMETERS" > 1000 AND "SYSTEM" IS NOT 'Amazon'` and see what you get. Do you understand the effect of using the `NOT`operator?

Finally, we have two useful operators for *partial matching* on strings. They are useful when you need to select based on a subset of string (word) values of an attribute:

(88) Clear your selection and create a new one with the expression `"NAME" LIKE 'Am%'`. This should select all three rivers whose name starts with 'Am' (Amazon, Amu Darya and Amur). The '%' symbol in this case is what we call a 'wildcard', and it means 'anything else'.

(89) Now use the expression `"NAME" LIKE 'C____'` (four underscores, '_'). This should select all rivers whose name starts with 'C' followed by any four characters (So it picks Congo and Chire).

! Stop and Think

How many rivers would you get if you changed the above expression to `"NAME" LIKE 'C%'`?

Click for answer

Four (Columbia, Colorado, Congo, Chire). When you use the '%' wildcard it means 'any amount of any character'.

(90) Finally, create a selection with the expression `"NAME" LIKE 'AM%'` (notice the uppercase). You won't get any results. Then change the expression to `"NAME" ILIKE 'AM%'`. You should get the same three rivers starting with 'Am' again.

! Stop and Think

What is the difference between `LIKE` and `ILIKE`?

Click for answer

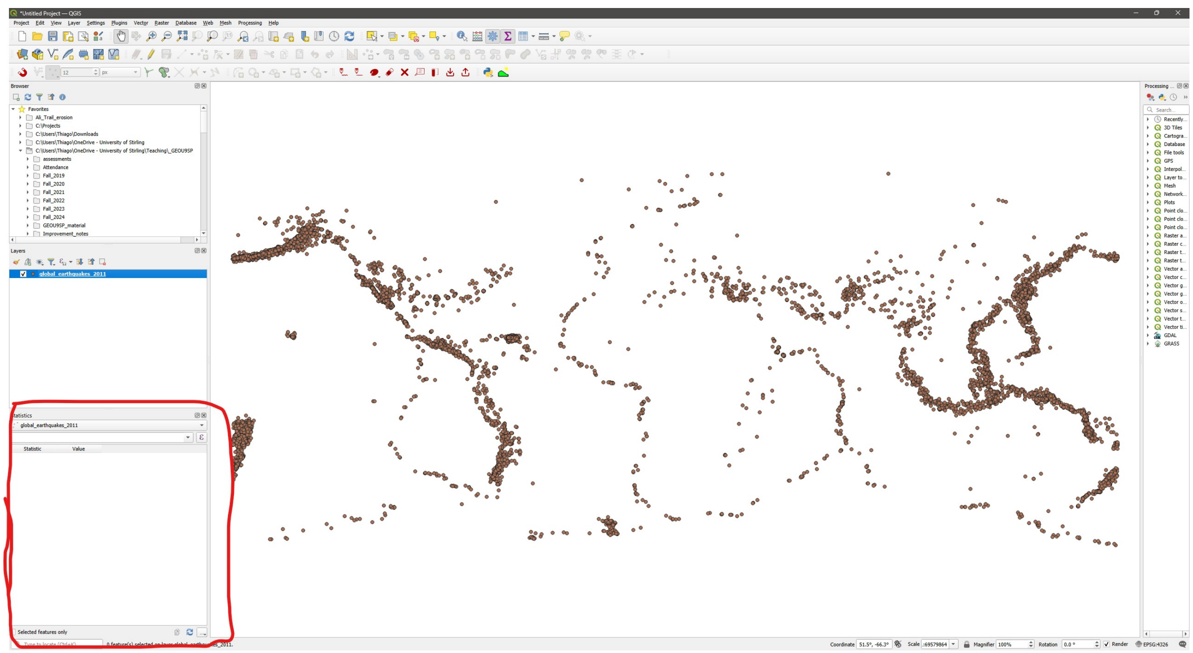
The `LIKE` operator is case sensitive, while `ILIKE` is case insensitive.

(91) Before you move on to the next exercise, make sure you clear your selection.

3.5 Guided Exercise 3: Summarising layer attributes

The `Statisitcal Summary Tool` (Σ) is a quick way to summarise attribute values, and can be quite powerful when combined with attribute selections.

(92) On the main QGIS window, click on the `Statistical Summary` tool button. A new panel will open on the bottom left corner of the QGIS window.



(93) Select the "Rivers" layer as input, then select the `KILOMETERS` attribute on the drop-down menu. You will get a table with several summary statistics calculated for all features in the layer.

! Stop and Think

What are the longest, shortest, and mean kilometre lengths in the dataset?

Click for answer

Max: 9207.1 km; Min: 194.9 km; Mean: 2663.6 km

(94) Now repeat the attribute selection you did before using the expression `"KILOMETERS" > 5000`, and then return to the `Statistical Summary` panel. Then check the `Selected Features Only` box at the bottom of the window. Now the stats will be re-calculated for the selected features only.

(95) The `Statistical Summary` panel is smart enough to know how to summarise different *data types*. For example, select the `SYSTEM` attribute, and see how the stats change - now it gives you how many features in total (`Count`), how many unique values (`Count(Distinct)`), how many missing values (`Count(Missing)`), the `Minimum` and `Maximum` string values (they don't have a clear meaning here), the least (`Minority`) and most (`Majority`) common unique values, and the `Minimum length` and `Maximum length` in number of characters.

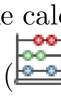
! Stop and Think

Why do you get a blank value for `Majority` and a 0 for `Minimum Lenght`?

Click for answer

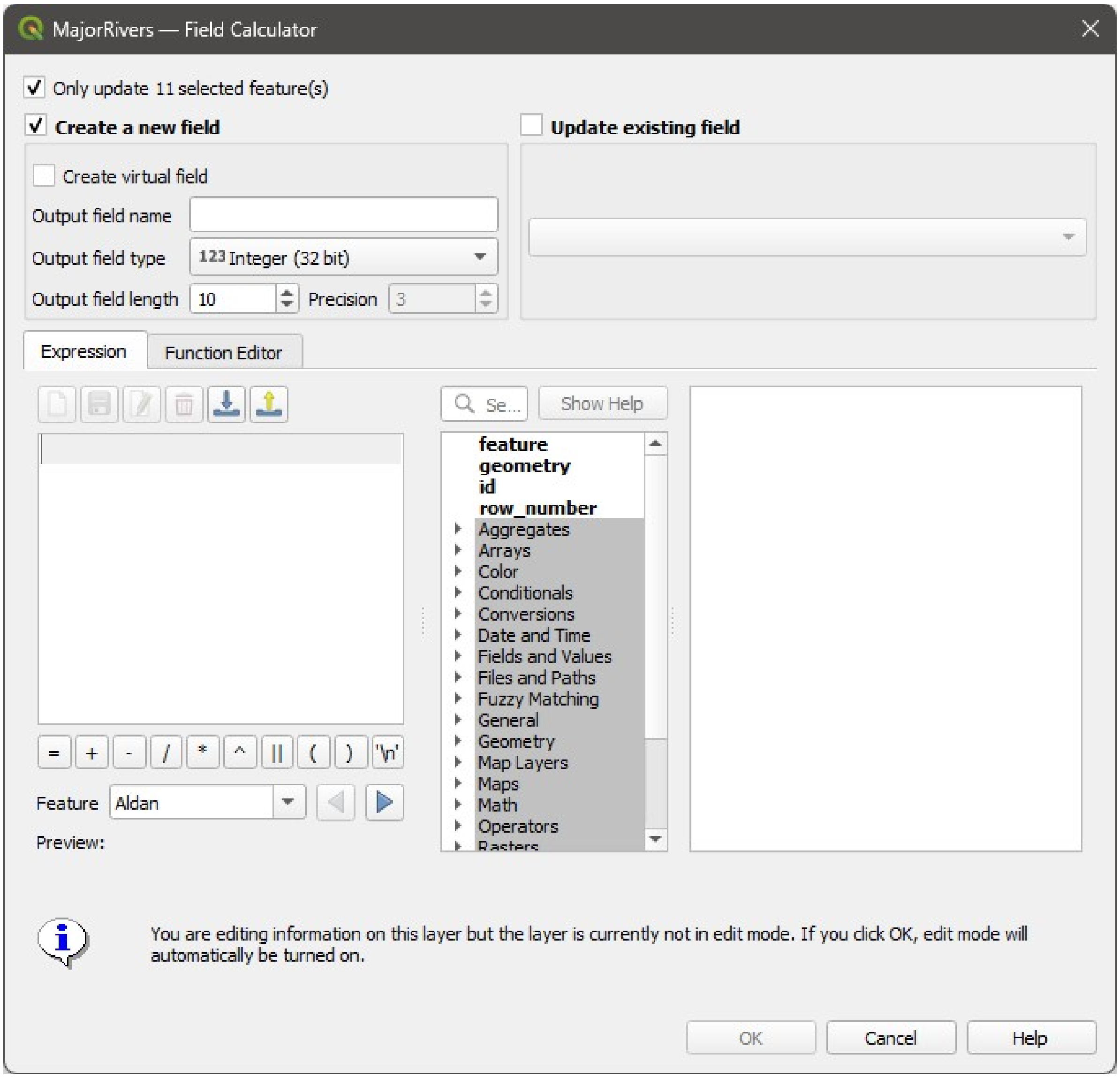
For the rivers dataset, `Majority` is blank and `Minimum length` is 0 because we have several `NULL` (missing) values. They have zero length, and they are in fact the most common unique value.

3.6 Guided Exercise 4: Calculating new attributes

Another powerful way to derive information from vector attributes is to make calculations involving multiple attributes. For that, we can use the `Field Calculator` tool (), accessible either from the `Attribute table` or the main QGIS toolbar.

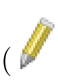
(96) Before your start, make sure you clear your selections.

(97) Go back to the Attribute Table window for the Rivers layer and click on the Field Calculator button to open it. It should look like this:



(98) On the new window, make sure the `Only update selected features` option is **not** checked. Then name the new field `Mi_to_Km`, and change the output *data type* to `Decimal Number`. `Output field length` tells you the maximum number of digits that can be stored per attribute value, and the `Precision` field tells you how many of these digits should be decimal places.

(99) On the expression window, write `"MILES" / "KILOMETERS"`. Note the warning at the bottom. Now click on `OK`.

(100) Now **turn off editing mode** by clicking on the Toggle Editing () button in the Attribute Table. When asked, confirm you want to save the layer changes.

Warning

Turning on editing mode is one of the **most dangerous** options in QGIS, as it lets you freely change both the geometry and the attribute values of a layer. Using the `Field calculator` automatically puts you on editing mode, so always make sure you turn it off immediately after you have finished a calculation. Once you make any changes and then save the changes, *there is no turning back*. I'll often first export a copy of the layer if I need to do any edits, so I aways have the original as a backup if something goes wrong. We'll revisit editing mode on week 4, when we learn how to digitise and edit geometries by hand.

Stop and Think

Does the Miles to Km proportion you calculated seem right?

Click for answer

Yes, 1 Mile = 1.609 kilometres.

(101) This new attribute you calculated is not very useful. Let us get rid of it. Put the attribute table back into editing mode, then click on the `Delete Field` button (), and select your `Mi_to_Km` field. Click `OK`. It's gone! Remember to save the changes and exit edit mode.

Tip: if you want to rename an attribute, create a new one with the new name and just use the name of the old one as the expression on the `Field Calculator`. It will copy all values to the new attribute. Then remove the old one.

(102) We will continue working with this data in the next lab. If you want to keeo your project, this is the best way to do it: on your computer's file explorer, find the root folder for the project (`lab_3`, or whatever you have named it as). Then right click on it and select (on Windows) `Compress to Zip file`. That will create a new zipfile of the folder contents, with the same name as the folder. Then you can copy this zip file to your OneDrive folder or to an external drive.

Good job, we have now finished our first guided tour of vector attributes. We will revisit it in the next lab, when we also learn about geometry-based selections.

Make sure you have a go at the independent exercise below, to make sure you feel comfortable with attribute selections, calculations and summaries. You will keep using these skills for the rest of the module (and your GIS life).

3.7 Independent Exercise

Using the earthquakes layer you downloaded (`global_earthquakes_2011.gpkg`), do the following:

1) Find out how many earthquakes of magnitude equal or larger than 7 have occurred in the Northern Hemisphere in 2011.
2) What was the average magnitude of all earthquakes that occurred in Japan in 2011? (Tip: make sure you enlarge the `event` column of the attribute table to the whole values).
3) Create a new `Text (string)` attribute that indicates if an earthquake is located on the western (`W`) or eastern (`E`) hemisphere.

4 Lab 4: Working with vector geometries

The purpose of this lab is to continue developing your knowledge about vector spatial data and vector-based GIS operations. Today you will learn how to make selections based on the overlap of two vector layers - called spatial selection, location selection or spatial query. This is a powerful way to combine data from different sources to generate new information.

4.1 Before you start!

1. Go through the Week 2 preparatory session on Canvas, and watch the seminar recording if you have missed it. Also make sure you have completed Lab 3.

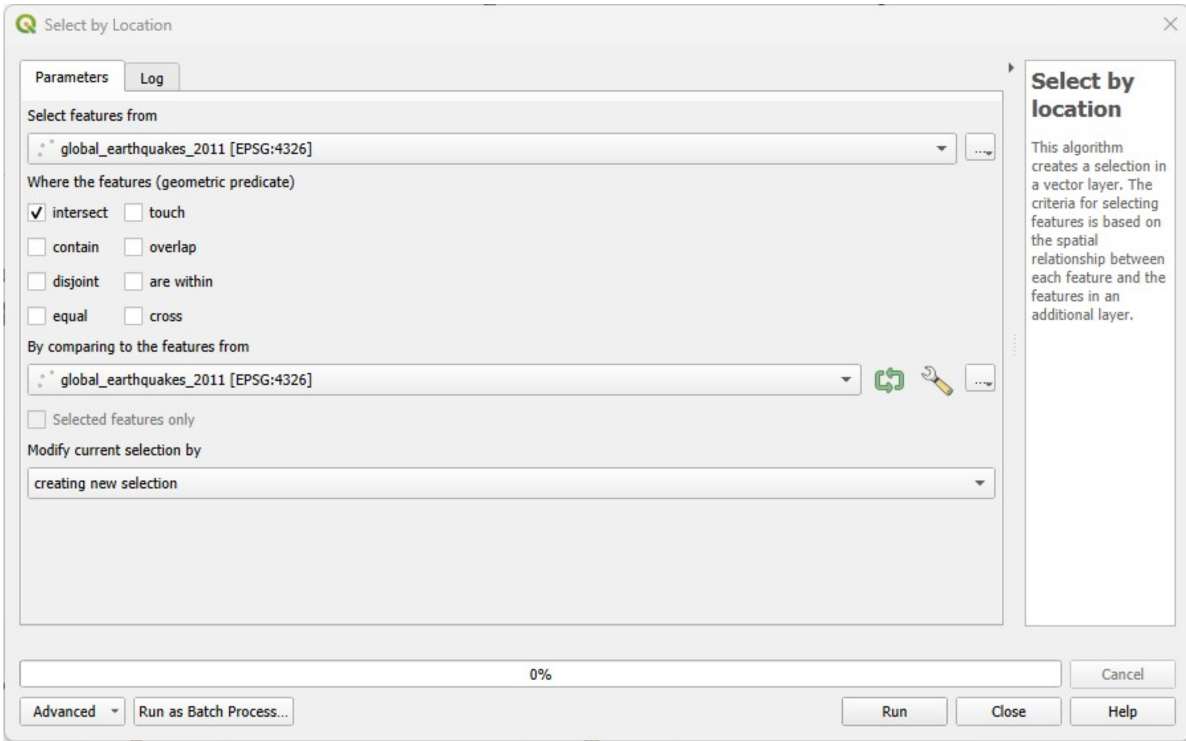
4.2 Guided Exercise 1 - Spatial Queries

We will continue with the Earthquakes, Rivers and Countries dataset from the previous lab.

(103) If you have zipped your project from the last session, just unzip it and get to work. If not, then download the data again, and redo your project organization. As a reminder, the files to be loaded are: `global_earthquakes_2011.gpkg`, `MajorRivers.shp`, and `ne_50m_admin_0_countries.shp`.

For our next step in the analysis, we would like to focus only on earthquakes on *land* (i.e. not ocean). This information is not contained in the Earthquakes layer, but we do have a Countries layer that separates land from ocean. Could we use that to make a selection?

(104) Go to the menu `Vector > Research Tools > Select by Location...`, or click on the  button on the main QGIS toolbar. You will get this window:



(105) Select the Earthquakes layer as `Select features from`. Then check *only* the `Are within` query option. Then pick your World Countries layer for `By comparing to the fetaures from`. Click `Run`, and after it is finished, click `Close`.

(106) Visually inspect the results of your selection, and then open the Attribute Table for the Earthquakes layer.

! Stop and Think

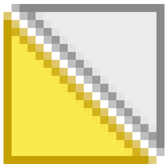
How many Earthquakes were originated on land in 2011?

Click for answer

3742 earthquakes.

Now that we identified the earthquakes on the continents, we may want to add this information to the Earthquakes layer as an attribute, in case we need the information later.

(107) Still on the Attribute Table of the Earthquakes layer, open the `Field Calculator`, and create a new `Text(string)` field called `Origin`. Make sure the option `Only update selected features` is enabled this time. Then on the Expression window, just write `'Land'`. That means the word `Land` will be added as the attribute value for every selected feature.

(108) Now back on the Attribute Table window, click on the `Invert Selection` button ().

(109) Return to the `Field Calculator` and this time use the `Update Existing Field` option to the right, instead of `Create a new field`. Make sure the option `Only update selected features` is still enabled. Pick `Origin` as the field to be updated, and now just write `'Ocean'` on the Expression window. Click `OK`.

(110) Now use the `Origin` attribute to style the colour or shape of your Earthquake points differently for land and ocean, using the `Symbology` options.

! Stop and Think

Do you understand why you did the steps above?

Click for answer

We wanted to create a new string attribute designating an Earthquake as either `Ocean` or `Land`. But the information on where the Earthquake is depends on whether it overlaps with the Countries layer. So we need to combine `select by location` with the `Field Calculator`. First we create a selection of land quakes based on location, and then we add a new field to the Earthquakes layer, to be filled with the word `Land`. But since we checked the `Only update selected features` option, only the selected features will be filled, and the rest will remain blank (`Null`). We then invert our selection so that only ocean quakes are selected, and update the existing field by filling in the word `Ocean` just for these selected features.
Creating this new attribute adds the possibility of styling the points by `Origin`, which would not be possible just based on the selection.

Now, let us find details about earthquake originating on a specific continent.

(111) On the World layer, use Attribute selection (as in Exercise 3) to select all countries from South America. Then go to `Vector > Research Tools > Select by Location...` and select all Earthquakes that `are within` the Countries layer (same process as above), but this time *make sure you turn on the option* `Selected features only`.

(112) Another way to make a selection 'permanent' is to export it as a new layer. While you still have the above selection on, right-click on the Earthquakes layer name and choose `Export > Save selected features as...`. Name your exported file as `Land_earthquakes_2011_south_america`. Choose an appropriate folder to save it

within your project structure by clicking on the ... button to the left of the file name box, and select your format of choice (geopackage or shapefile). Click on OK. Save your project.

4.3 Guided Exercise 2 - Geometry-based attribute calculations

Another way to relate geometries to attributes is when we want to store some property of the geometry as an attribute itself, such as area, length or perimeter. QGIS also has tools for that.

As an example, let us calculate the areas and border length (perimeter) of all the World's countries.

(113) Before using any of the geometry-based operators, we need to set up our desired units for the project. Go to the menu `File > Properties...` to open your project properties. Then on the `General` tab, set the `Units for distance measurement` to `Kilometers` and the `Units for area measurement` to `Square Kilometers` (yes, we will use SI units because we scientists!).

It is always important to think about your problem before choosing the units for length and area calculations. In this case, using the default of meters and squared meters would not make sense for things as large as countries.

(114) Now open the Attribute Table for the Countries layer and launch the `Field Calculator`. Create a new attribute called `Area_km2` (it is always a good idea to add the units to the name to avoid any confusion) that is a decimal number with two decimal places. Then on the centre panel, find and expand the `Geometry` heading, and double click on the `$area` option to add it to the expression window. Then click on `OK` to calculate the field.

(115) Remember to deactivate editing mode!

! Stop and Think

a) What is the difference between `$area` and `area`? Hint: look at the help text on the right panel.

b) Why should we use `$area` instead of `area` here?

Click for answer

a) The `$area` operator calculates `geodesic` area, meaning it takes into consideration the curved surface of the ellipsoid (defined by the projects CRS). It will also use the units specified in the project properties. The `area` operator will calculate *planimetric* area (i.e. a flat surface) and will derive the units from the CRS of the layer.

b) Since we are calculating very large areas, geodesic area will be more correct than planimetric area (unless the dataset uses an *equal-area* map projection). Moreover, the data is in EPSG24326 (WGS 84), which has degrees as units. If we used `area` we would get areas in squared degrees, which don't make a lot of sense.

(116) Now repeat the process above, and calculate `Per_km` as a new attribute, using the `$perimeter` operator (we use it instead of `perimeter` for the same reason above).

! Stop and Think

The perimeter/area (PA) ratio is often used as a measurement of polygon complexity. Can you answer which country has the most complex border in the world?

Click for answer

Use the `Field Calculator` to calculate areas and perimeters (you just did), then use it to create a new `PA_ratio` attribute (decimal number) using the expression `"Per_km" / "Area_km2"` (assuming these are the attribute names you used). Then use the `Summary Statistics tool` to find the maximum `PA_ratio` value in the dataset (4.592). The use `Select by expression` to select which feature has `"PA_ratio" = 4.592. That would be the Vatican!  A quicker way to answer this (that only works for min/max values) is to click on thePA`_ratio' attribute name on the Attribute Table twice, to sort it in ascending and then descending order, and then checking which country is the top row after sorting.

Good job! You now know all you need to query, create, delete, update and summarise attributes. You will use these tools a lot whenever you are doing any GIS work, so it is important to know them well.

To further solidify your knowledge, do the Independent Exercise below, which reviews all your learning from the past week and this week.

4.4 Independent Exercise - Supporting Wildcat conservation in Scotland.

You want to investigate how priority areas for wildcat conservation (WPAs) overlap with protected areas, and the risk of wildcat roadkill.

For that, obtain the following layers from the NatureScot Open Data Hub:

- Wildcat Priority Areas (WPA)
- Sites of Special Scientific Interest (SSSI)

Then obtain the Open Roads layer for all of the UK from the Ordnance Survey OS Open Data hub (it is a large file and it may take a while to download).

Finally, obtain the UK country boundaries from the Global Administrative Boundaries (GADM) hub. Get the *GeoJSON* file at *Level 1* (Level 0 is the UK border without countries, level 1 is countries, level 2 is counties/council areas, etc.)

Organize the data and create a new project where you load all four datasets. Order and style them as you prefer. Then:

1) Answer the questions:

    a) What is the data model of each layer?
    b) What is the file format of each layer?
    c) What is the CRS of each layer?
    d) How many features does the WPA dataset have?
    e) How many attributes does the WPA dataset have?

2) The UK boundaries file has a different CRS from the rest, and it also contains all countries. Reproject this layer to the same CRS as the other layers, and then create a new layer containing the Scotland boundary only. Renove the UK layer from the porject after that.

3) The roads layer covers all of the UK, making it very heavy and slow to use. But our analysis concerns Scotland only, so create a new layer containing only the road links (lines) inside the Scottish boundaries. Remove the full UK layer from the project after that.

4) The attribute `Shape_Area` of the WPA dataset does not indicate any units. It is always useful to have field names that hint at the unit for the values, making the data more self-explanatory. Recalculate the area of each WPA *in square km*, naming the new field 'Area_km2'. Then delete the 'Shape_Area' attribute from the dataset to avoid confusion for future users.

5) The Sites of Specific Scientific Interest (SSSIs) are a category of protected area in Scotland. These may offer additional protection to wildcats. Find all the SSSIs that overlap with WPAs, and create a new layer containing only these sites. Then calculate the total area of these SSSIs.

6) A major threat to wildlife is roadkill. Using the OS Open Data Open Roads dataset, answer the questions below:

    a) How many road segments overlap with WPAs? (Create a new layer containing only, the selected data, to use below).

b) What is the total length of roads within WPAs?

c) Not all roads present the same risk. In the UK, roads are classified as:

    - Motorways: high-speed expressways typically reserved for longer journeys between major cities;
    - A roads: major roads intended to provide large-scale transport links;
    - B roads: roads intended to connect different areas, and to feed traffic between A roads and smaller roads on the network;
    - Classified Unnumbered – smaller roads intended to connect together unclassified roads with A and B roads, and often linking a housing estate or a village to the rest of the network;
    - Unclassified – local roads intended for local traffic. The vast majority (60%) of roads in the UK fall within this category.

What is the total length per road class for all roads intersecting WPAs?

You will notice a 'problem' with the terms used in the road classification attribute - its unique values don't match the terms above. You will need to create a new field called `class_fix` that keeps the classification of `Motorway A`, `B`, `Classified Unnumbered` and `Unclassified` unchanged, but changes the records labelled as both `Not Classified` and `Unknown` to `Unclassified`.

d) Along with road class, the length of the road segment is also important in determining overall traffic speed. To take that into consideration, select all roads within WPAs that are of either Motorway, A or B class and have 100m or more in length. Create a new layer from this selection.

e) GIS analyses will almost always have a visual component, as maps are a natural way to tell a 'spatial story'. To finalise your exercise, organise and style your layers to show, in the most readable way possible (i.e. think about orders, line colours, widths and styles, fill colours, etc):

- The Scottish boundary;
- The WPAs;
- The boundaries of the SSSIs, distinguishing between those that do and do not overlap with WPAs;
- All Scottish roads differentiated by class using line width;
- All the road segments that are Motorways, A or B roads within WPAs that are longer than 100m. These should use the same line widths as above to differentiate road class, but have a different colour from other roads.

Part III

Week 3 - Raster data

For our third week, we will learn about the second most common data model for spatial data: raster files. While vectors are good at representing separated and/or well-defined individual regions, rasters excel at representing gradual, continuous data such as surface temperatures or elevations.

ILOs covered

1. Understand the structure of spatial data and choose appropriate data types and models for storing and representing it;
2. Obtain and assess the quality of spatial data from online and offline sources and produce new spatial data using computer and field methods;
3. Create map visualisations that adhere to cartographic principles and can be easily and unambiguously interpreted by the non-specialist public;
4. Plan and execute GIS analytical steps to solve spatial problems successfully;

What will you learn

For every week, we will list the main theoretical and practical learning goals. Use these as a 'checklist' to gauge your learning for each week. If you don't feel confident you have learned any specific topic, then revisit the week's material!

Theoretical knowledge for Week 3:

- What is the raster data model?
  - How are rasters represented?
- What are the components of a raster file (rows, columns, bands/channels, resolution)?
- What are the main raster file formats?
- What data types can rasters represent?
- What kinds of data can we represent using raster data?
- Raster / vector component equivalence
- Raster / vector tool equivalence
  - No concept of ' selection'
- What kinds of questions can we answer using rasters?
- How can we transform rasters to vectors and vice-versa
  - The concept of raster resampling
  - The concept of raster contrast manipulation

Practical knowledge:

Chapter 5

- How to identify raster files
- How to load raster files
- How to get raster properties
- How to reproject raster data
- How to mask rasters
- How to set raster symbology
  - Continuous variables
  - Categorical variables
- How to use the identify tool to get raster information on the fly
- Working with Digital Elevation Models
  - Calculating slope, aspect and viewshed

Chapter 6

- How to calculate raster statistics
- How to 'select' using the raster calculator
  - Boolean operators on the same band
  - Boolean operators between bands
  - Band arithmetic
  - Making flowcharts
- Mosaicking rasters
- How to adjust the stretch of a raster
- Styling raster Images
- Resampling methods

5 Lab 5: Working with raster data

So far, we have been working within the realm of vector data: beautiful topological combinations of vertices and lines to represent the complexity of the Earth's surface. But there are plenty of situations where vectors are not the best choice; any information that varies gradually and continuously over the surface can be better represented by a raster.

You know rasters well - any digital image is a raster. While vectors connect dots with lines, rasters are grid (i.e. a matrix) of numbers, called *pixels*. The numeric value of each pixel can be used the indicate a colour (like on digital photos), or can actually represent a physical quantity, such as temperature or elevation.

Raster files can contain multiple 'images' within them, which call *channels* or *bands*. Digital photos for example, actually contain three bands: one specifying the amount of red colour per pixel, one for the green colour, and one for the blue colour. When you look at a photo you are seeing a *color composition* mixing the amounts of the three primary colours according to the information help by each pixel.

Figure 5.1: Source: https://doi.org/10.5772/63028

For GIS data, multiple raster bands can be thought as vector attributes - each band will hold the values of a specific variable (i.e. one band for temperature, and one for humidity). In fact, it may help to frame rasters in terms of their vector equivalents:

	Vector	Raster
Data variables	Vector Attributes	Raster Bands
Data values	Attribute values	Pixel values
Select ranges of data values	`Select by expression`	`Raster calculator` and/or `Mask`
Calculate new variables	`Field Calculator`	`Raster Calculator`
Calculate feature areas	`Field Calculator` (`$area`)	`Raster layer unique values`
Select by spatial location	`Select By Location` (Vec > Vec)	Extract Raster Values (Points > Raster) Zonal Statistics (Polygons > Raster) Zonal Statistics (Raster > Raster)

We will cover most of the above this week, and then cover how to extract values from raster based on other layers on Week 6. For today, we will learn how to inspect and style raster data, and also how to do some mathematical operations among raster layers using the raster equivalent of the `Field Calculator` - the `Raster calculator`.

5.1 Before you start!

1. Go through the Week 3 preparatory session on Canvas, and watch the seminar recording if you have missed it.

5.2 Guided Exercise 1 - Opening and inspecting raster data

(117) For this exercise we will look at three different raster layers. The data has been pre-packaged and can be downloaded here. It includes the following datasets:

- UK SRTM: This is a single-band raster containing data from the [Shuttle Radar Topography Mission (SRTM)] (https://en.wikipedia.org/wiki/Shuttle_Radar_Topography_Mission), a global dataset of surface elevation. The data can be obtained from a variety of sources online.

- UK Bioclim: The BIOCLIM suite of climate data has been developed to support biodiversity studies and species distribution modelling. It is a single multiband raster file. Read this page for a description of which variable each band represents.
- CORINE land cover: CORINE is a EU-wide land cover map that is produced by the Copernicus Space Program. The pixel values are numerical codes that specify the different land cover classes. A description of each land cover number is given here.
- UK administrative boundaries from the GADM website, in geopackage vector file format.

(118) Download the required data and organise it as you have learned. Create a new QGIS project and add the `UK_SRTM.tif`, `UK_Bioclim.tif` and `CLC2018_CLC2018_V2018_20_UKclip.tif` raster files to your project. All files are in *GeoTIFF*' format, which is the most common and standard spatial raster file format.

(119) For each of the raster layers, right-click on the layer name and go to `Properties > Information`, then:

! Stop and Think

a) What is the *data model* of these files?

b) What is the *file format* of these files?

c) What are the *data types* of each raster file? Why is that important?

d) What are the CRSs of each dataset you have?

e) What are the dimensions (rows, columns) of each raster dataset?

f) How many *bands* does each raster dataset have?

g) What are the *pixel sizes* (spatial resolution) of each raster dataset?

Click for answer

a) raster (vector for the boundaries)

b) GeoTIFF for the rasters, geopackage for the vector

c) The data types can be found when looking at `Properties > Information` for a raster layer. The SRTM and CORINE datasets are sixteen bit signed integer (INT16), the BIOCLIM dataset is a 32-but floating point raster. Raster data types are important as they determine the range and precision of the data that can be stored in the raster, and also impact the size of the raster file. For the same number of rows, columns and bandas, a 32-bit raster is twice as large as a 16-bit raster.

d) GADM, SRTM, BIOCLIM: EPSG 4236 (WGS-84); CORINE: EPSG3035 (ETRS89 / Europe LAEA)

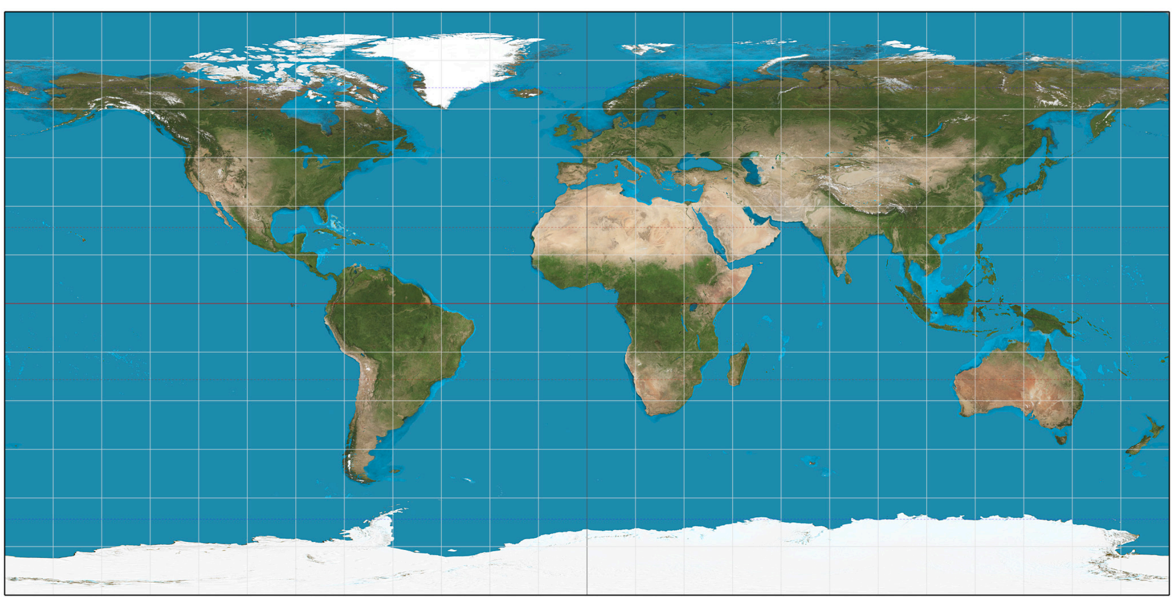
e) These are also found in `Properties > Information`, as `Width` (columns) and `Height`(rows) or in `Dimensions`. SRTM: 16355 cols x 12036 rows; CORINE: 10473 cols x 12261 rows; BIOCLIM: 1471 cols x 1084 rows.

f) These are also found in `Properties > Information`, as a list of bands, or in `Dimensions`. SRTM: 1 band, CORINE 1 band; BIOCLIM: 19 bands.

g) These are also found in `Properties > Information` in `Pixel Size`. SRTM: 0.0008084837557075693617 degrees; CORINE: 100m; BIOCLIM: 0.008983152841195215371 degrees.

5.3 Guided Exercise 2 - Reprojecting raster data

As we learned in week 1, the UK looks "squished" because the data is projected using only the WGS-84 datum and geographic coordinates (EPSG:4236). Although data in geographic coordinates is often referred to as "unprojected," this is not actually true (you are looking at it on a flat screen, right?). For these "unprojected" datasets, what most GIS software do is to simply use a linear function to convert latitudes and longitudes in degrees to x and y values on your screen. This projection can be referred to as Plate Carrée or Equirectangular projection, which has a heavy amount of distortion towards the poles:



(120) To reproject raster data, go to `Raster > Projections > Warp (Reproject...)`. Select your SRTM layer as the Input Layer, and the EPSG: `27700 - OSGB 1936 / British National Grid` as the `Target CRS`. (If you don't see it as an option, click on the small button to the right to bring up the CRS selection window). Set the `Output file resolution...` to `90` (this will be in meters, as meters are the units of the BNG projection). Leave everything else as default. Then click on the `...` button to pick an appropriate folder location. Name uour file `UK_SRTM_BG.tif`. `Run` the algorithm and then `Close` when finished.

(121) Repeat the process for the Bioclim layer (use a cell size of 1000m) and the CORINE land cover map (cell size of 100m). These cell sizes are the meter equivalent of the degree pixel sizes the data currently has. Make sure you add the `_BG` suffix to the new file names to keep track of what has changed from one file to the other.

(122) If it seems nothing has really changed, remember that QGIS does reprojections "on the fly" to make sure data on the screen are all aligned to the project CRS. So as you learned in Week 1, change the project CRS to EPSG:27700 as well. Now everything should look good.

(123) Remove the original layers from the project and keep only the reprojected ones, then save your project.

5.4 Guided Exercise 3 - Masking rasters using vectors

We often want to remove portions (specific cells) of a raster, a process called *masking*. For example, some of our datasets include the Republic of Ireland and bits of mainland Europe, and some datasets include the Shetland Islands, while others don't. Let use the UK boundaries from the GADM dataset to mask our data to the Island of Great Britain (Scotland, England and Wales) only.

(124) First add the GADm data (`gadm36_GBR.gpkg`) to your project. This geopackage holds multiple layers for the different levels of admin boundaries. You can just add the level 0 layer, which gets the UK as a whole.

> ❗ Stop and Think
>
> Why did you get a warning window when you added the layer to the project?

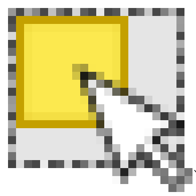
> **Click for answer**
>
> Because this layer has a different CRS than the project, and QGIS it is asking you how to deal with it.

As we learned on Week 1, trying to do clipping (and masking) operations between layers with mismatched CRSs gives us wrong results. So before you do anything, reproject the UK boundaries layer to EPSG 27700. You also need to extract the main British Island from the rest of the dataset. Turns out there is a quick way to do both at once, but first we need to deal with a little issue with the vector layer:

(125) Open the Attribute Table of the UK boundaries layer. How many features does it have?

This is what is called a `multipart` polygon - a set of disjoint polygons all treated as a single feature. Before we can use the layer, we need to split the individual polygons apart to be able to select the main island only:

(126) Go to `Vector > Geometry Tools > Multipart to Singleparts`, and select the GADM layer. You can leave it as a temporary layer since we will only export one polygon for it in the next step. A new temp layer called `Single parts` will be created. Inspect its Attribute Table.

(127) Now use the feature selection tool () to select Great Britain only, and then right-click on the Single Part layer name and choose `Export > Save selected features as...`. Pick a folder location and name the file `GB_Island.shp`. Before you click on OK, however, change the `CRS` drop-down menu to EPSG:27000. That is it, QGIS will now reproject the layer before saving it! Click `OK`to save, and then remove the original GADM and the temp layer from your project to keep things tidy. Save your project.

(128) Double check what is the CRS of the new `GB_island` layer you have created.

(129) You are now ready to mask the raster data. Go to `Raster > Extraction > Clip Raster by Mask Layer...`. As `Input layer`, select the SRTM layer, and as `Mask layer`, select the GB island layer. Then on the option `Assign a specific nodata value to output bands`, enter the number `-32768`. *Make sure you include the minus sign*! Leave the other options as default, but do check both the `Match the extent of the clipped raster to the extent of the mask layer` and `Keep resolution of input raster` options. Pick a folder and name your file `UK_SRTM_BG_GBmask.tif`. Now we can know at a glance that this is the SRTM data, reprojected to British Grid and masked to Great Britain.

! Stop and Think

a) Why did we pick a value of -999 for 'nodata'?

b) Why should we check the `Match the extent...` and `Keep Resolution..` boxes?

Click for answer

a) Since raster layers are grids (matrices) they must always be rectangular in shape. So when masking a raster using an irregular shape, the pixels outside the polygon will receive a value indicating *nodata*, but they will still exist. The default nodata value for QGIS is zero, but since we are dealing with elevations, zero is still a valid elevation value, and we could even have areas inland that are at sea level or lower, and thus have an elevation of zero or below. We also know our data is a signed 16 bit integer, that can take values from -32,768 to +32,767. So we pick the most negative elevation number within this range, which also would never occur on land (-32768 meters) to use as nodata.

b) By selecting the `Match the extent of the clipped raster to the extent of the mask layer` option when masking, the extent of the raster rectangle will be just enough to cover the GB polygon. If we left that unchecked, QGIS would set the non GB pixels to 'nodata', but leave the raster with the original dimensions. That would be a waste of disk space. We also check `Keep resolution...` to make sure QGIS doesn;t change the pixel sizes to better 'fit' the rectangle to the extent. The spatial resolution of a raster file often has a reason to be, and we don't want to arbitrarily change that.

(130) Use the `Identify Features` tool () to click on blank area of the masked SRTM layer (near the shore). Make sure you have the SRTM layer selected on the Layers panel before you click the tool. What pixel values do you get? -32768 or 'nodata'?

(131) Now go to the `Properties > Transparency` tab of the SRTM raster layer, and uncheck the `No data value -999` box. Look at your raster again, and probe the same areas with the `Identify Features` tool. Can you see now how the actual raster is still a rectangle, with the pixels outside the mask set to -32768? Before you proceed, go back and check the `No data -999` box again.

(132) Now repeat the masking for the CORINE and BIOCLIM layers. You can use the same nodata number for the CORINE and BIOCLIM layers. the CORINE layer actually already has nodata defined as `-32768` so we just keep it. And for Bioclim, it is an unlikely value for any of the climatic variables, and because the climate data is a floating point number (decimal), it is almost impossible that a perfect value of -32768.000000 would exist in the data naturally.

(133) Remove the non-masked raster datasets from your project and save it.

(134) Now zoom in into any place on the coastline, and answer:

> **!** Stop and Think
>
> Why don't the edges of each raster dataset line up perfectly?

> **Click for answer**
>
> a) Again, this is a side effect from raster layers being grids. As each one has a different pixel size, the `decision` of which pixel is inside our outside the GB vector polygon will be different for each layer. Also, the coastline will always appear 'ragged' because we can't represent any detail smaller than a single pixel on a raster file - and the pixel grids of the three layers do not align because of the pixel sizes.

5.5 Guided Exercise 4 - Styling raster data

Just as with vector data, QGIS offers many options for the symbology of raster data. You will see some similarities between the symbology options for vectors and rasters, but also some differences because of the nature of each data model.

(135) Turn off all layers except the CORINE land cover, and then go to its `Properties > Symbology`. Note that the default `Render Type` for single-band rasters is `Singleband gray`. That means pixels are coloured by a shade of grey that is proportional to its numeric value, with higher values being closer to white. Change it to `Paletted / Unique Values`. Maintain the `Band 1` selection (there is only one band anyway) and `Random Colors` options, and then click on `Classify`. Apply the result to your raster and visualise it.

> **!** Stop and Think
>
> a) How many categories are there on this raster?
>
> b) What are the pixel values representing?
>
> c) What would be the vector symbology equivalent of `Paletted / Unique Values`?

> **Click for answer**
>
> a) 36 categories (unique pixel values)

b) Each pixel value is a numerical code that indicates a land cover type.

c) the `Categorized` option.

WHoa, there are a lot of classes, and the random colours selection picked colours that are very similar to one another for some of the classes. Fortunately for us, the data producers of CORINE include a colour map file as part of the metadata. Let us use it!

(136) Back on the `Properties > Symbology` window, click on the `...` button to the right of `Delete All` and choose `Load Color Map from File...`. Within the CORINE files you unzipped, find and select the metadata file named `CLC2018_CLC2018_V2018_20_QGIS.txt`. `Apply`and then click `OK`. Not only we get the colours selected by the CORIEN mapping team, but also all the proper class names! That was nice!

(137) Now let us have a look at the BIOCLIM dataset. First, read the metadata included with the data files (`Bioclim_metadata.txt`) to understand what each band of this dataset represents. Then turn on the BIOLCIM layer and turn the remaining layers off.

! Stop and Think

Why does the BIOCLIM layer appears to be coloured?

Click for answer

When QGIS opens a multiband, single file raster, it always thinks that the file is an image, and thus it automatically assigns the first three bands of the file to the Red, Green, and Blue colour channels. We will look at raster images in the next session, but this color composition clearly doesn't make sense for our climate data, so let us change it.

(138) Go to the `Properties > Symbology` window for the BIOCLIM dataset. Notice the default symbology choice of `Multiband color`. Since each band of our data represents a different climatic variable, with continuous numeric values (not categories or classes), we need to pick the `Single band - pseudocolor` option. Choose the `bio1` band, and for the colour ramp, click on the down-arrow button to the right of the colour palette button, and select the `magma` colour ramp. Classify the existing values and then `Apply` and click `OK`. How does the layer look like now?

! Stop and Think

a) What would be the vector symbology equivalent of `Single band - pseudocolor`?

b) What climatic variable is "bio01" representing?

c) What units are the minimum and maximum values shown on the colour scale?

d) Why such a strange choice for the units?

Click for answer

a) The `Graduated` option.

b) The metadata file tells us it is Annual Mean Temperature.

c) The metadata file also tells us it is °C * 10 (deegrees celsius times ten)

d) This is an old `trick` to reduce file sizes while keeping some precision. Let's say the data creators wanted to record temperatures with one decimal place of precision. If they stored the data as a floating point, we would need 32-bit files. But if we consider that mean annual temperatures on Earth will easily be contained in the range of -100 to +100 degrees Celsius, we could instead multiply all numbers by ten, and store them as 16-bit integers instead, cutting file sizes in half. For example, a temperature of 32.7°C becomes a pixel value of 327.

(139) Explore some of the other climatic variables contained in the BIOCLIM dataset, using different colour ramps.

Finally, let's work on the symbology for the SRTM data. For that, we will take advantage of some nice scientific colour palettes that are built-in on QGIS. We will talk aboout what makes these palettes special on Week 4, but let us just use them now.

(140) On the `Properties > Symbology` window of the SRTM layer, choose the `Singleband Pseudocolor` option, and then on th downarrow button besides the `Color Ramp` option, select Create New Color Ramp. On the small options window that comes up, select Catalog:cpt-city.

(141) Once the catalogue window opens, go to the `Topography` list and select the `cd-a` palette. Then manually enter `Min` and `Max` values, using `0` and `900` respectively. `Apply` and see how it changes. But before clicking on `OK`, go to the `Transparency` tab and drag the slider at the top to around `60%`. That will let other layers under it to show through and blend with the colours.

(142) Right-click on the SRTM layer name and select `Duplicate layer`. This option creates a 'virtual' copy of the layer - it will still point to the same file on disk, but you are able to select a different symbology for it.

(@)Now, on the `Properties > Symbology` of the copied layer, change the render type to `Hillshade`. Leave everything as default, then `Apply` and close the window. Make sure both the coloured and the hillshade SRTM layers are turned on on the Layer Panel, and that the hillshade layer is immediately under the coloured SRTM layer. Zoom in and turn each layer on and off in turn to understand how this neat '3-D' visual effect works.

! Stop and Think

What does the `hillshade` render style does?

Click for answer

It uses the elevation information to simulate how different areas of the terrain would be illuminated or shaded by sunlight coming from a certain sun elevation and azimuth (direction).

5.6 Guided Exercise 5 - Terrain Calculations

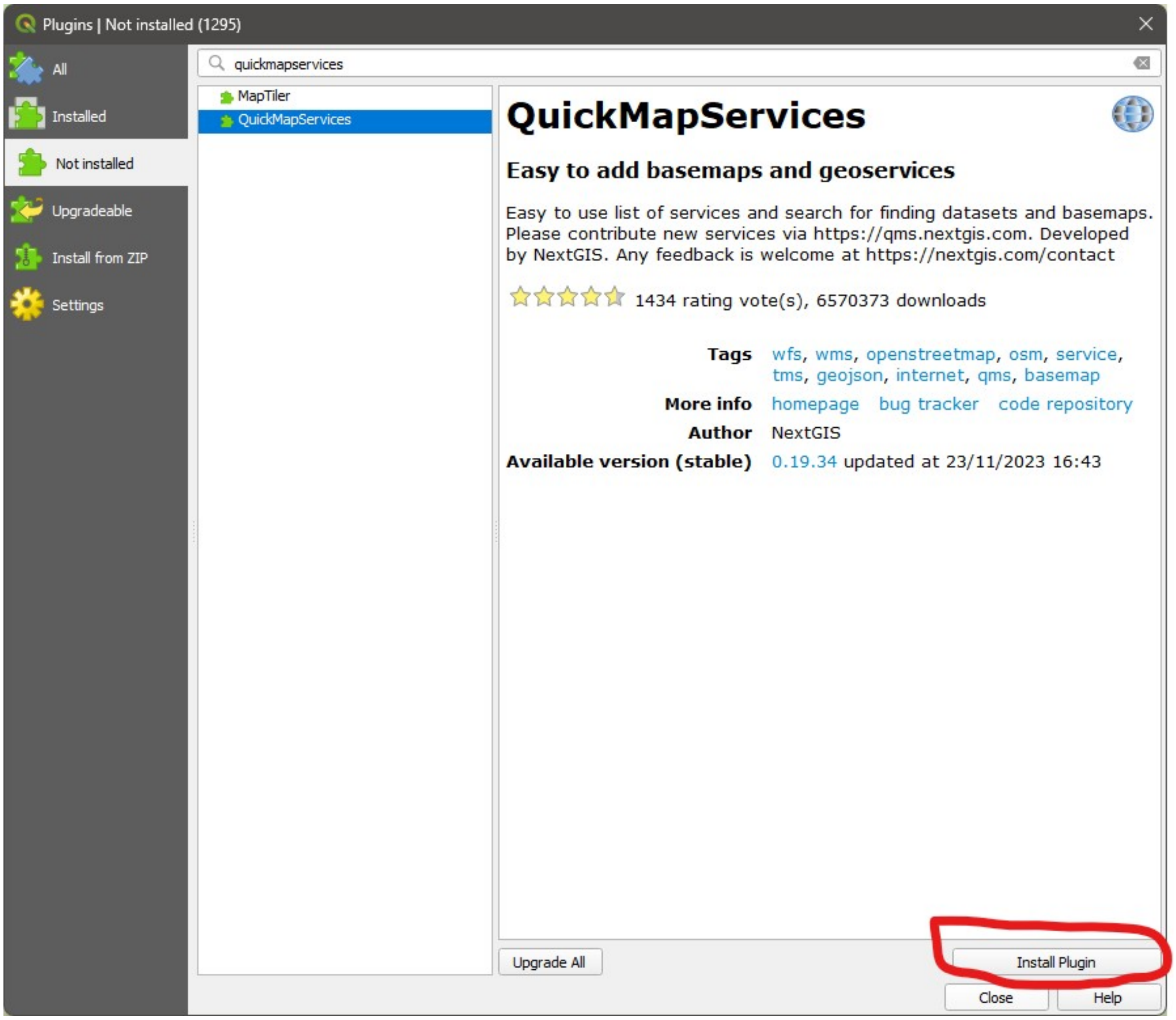
One of the most common raster datasets used in GIS are Digital Elevation Models or DEMS. The structure of raster data is particularly well suited to represent terrain, in its continuous and highly variable nature. Moreover, terrain and elevation data is often a key variable in GIS analysis, as it directly influences most biological, geological and anthropic processes.

For this reason, a few specific terrain analysis tools that are always included in GIS software, and used often. These are the methods to calculate slopes(a.k.a grades, gradients), aspects and viewsheds, using DEMs as the base data.

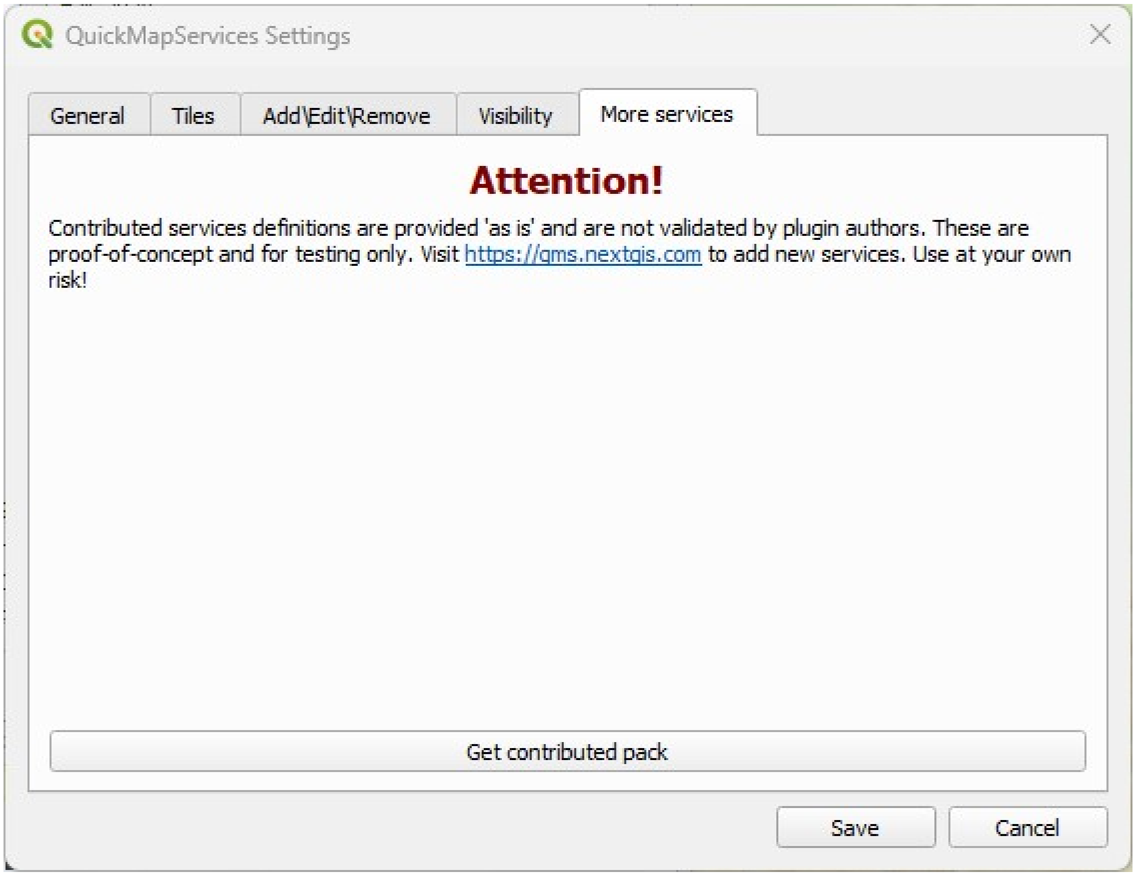
5.6.1 Plugins

In this exercise, you will also learn about one of the most powerful aspects of QGIS: *plugins*. As you may know, QGIS is a free and open-source software, meaning anyone can contribute with the code. QGIS also makes it easy for it's functionality to be extended via plugins, separate little 'apps' that add new tools and capabilities to the main QGIS app. For this exercise, we will install and use two official QGIS plugins, the `QuickMapServices` and `Visibility Analysis`.

(143) Head to the menu `Plugins > Manage and Install Plugins`. You will se a window like the one below. Then select the `Not Installed` tab, and then on the search bar search for `QuickMapServices`. Select it on the left panel, and then click on `Install Plugin`. Then repeat this process to find and install the `Visibility Analysis` plugin.



(144) The `QuickMapServices` plugin will add a new menu entry: `Web > QuickMapServices`. Once yo8u click on it, you will see a list of web map services - but there is more. Go to `Web > QuickMapServices > Settings`, and then click on the `More services` tab. Then read the warning and click on `Get contributed pack`, and once you get a confirmation message, click on `Save` and exit the window.



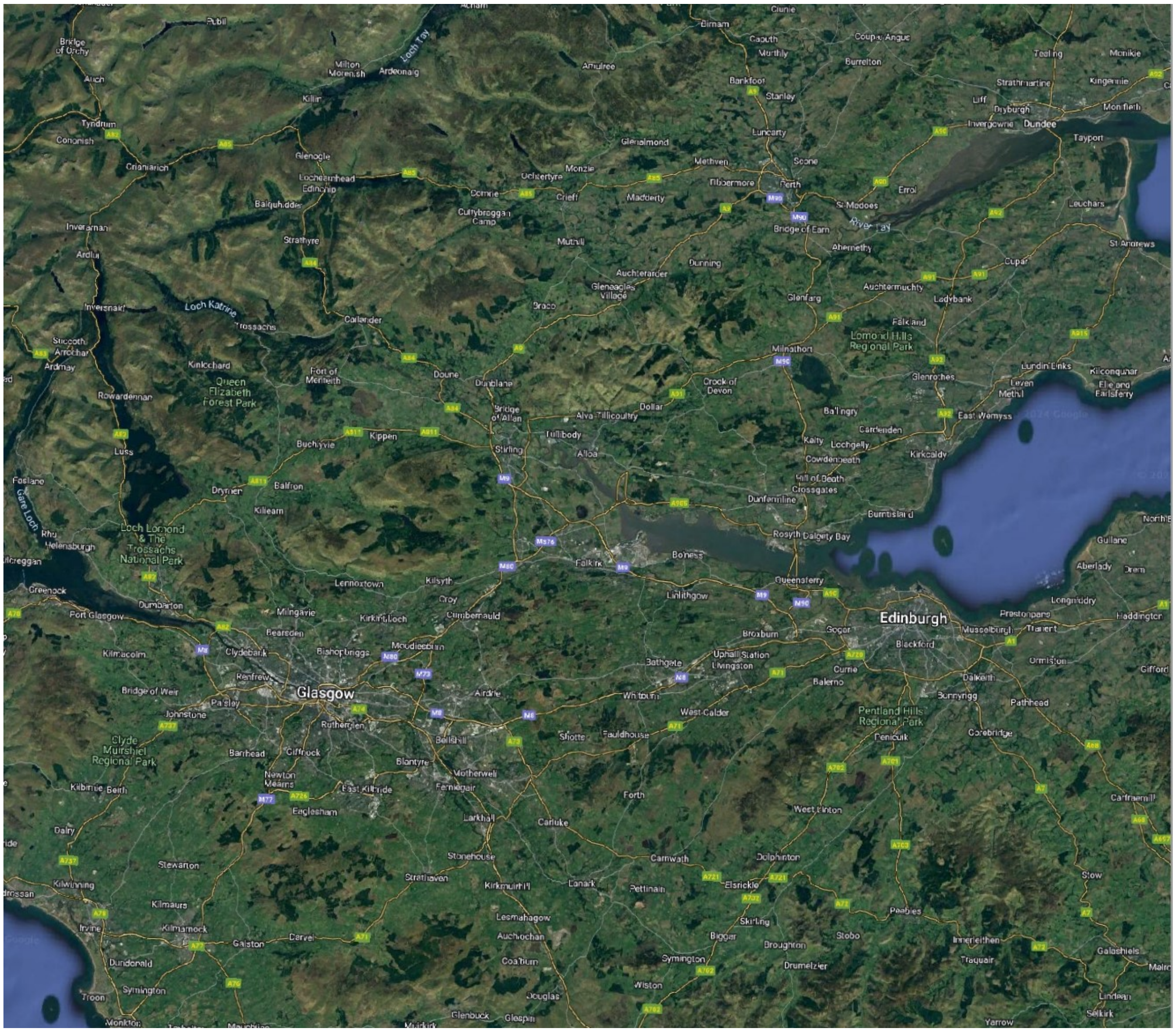
(145) Now go back to `Web > QuickMapServices > Settings`, and you will see a list of providers. For example, go to the `Web > QuickMapServices > Google` option and add the `Google Hybrid` dataset directly as layer in your project! These layers will require you to be connected to Internet to work, and they won;t give you many symbology options, but they are very handy to help in navigation when working on a project.

(146) Take some time to explore the layers available in this plugin.

5.6.2 Masking by area

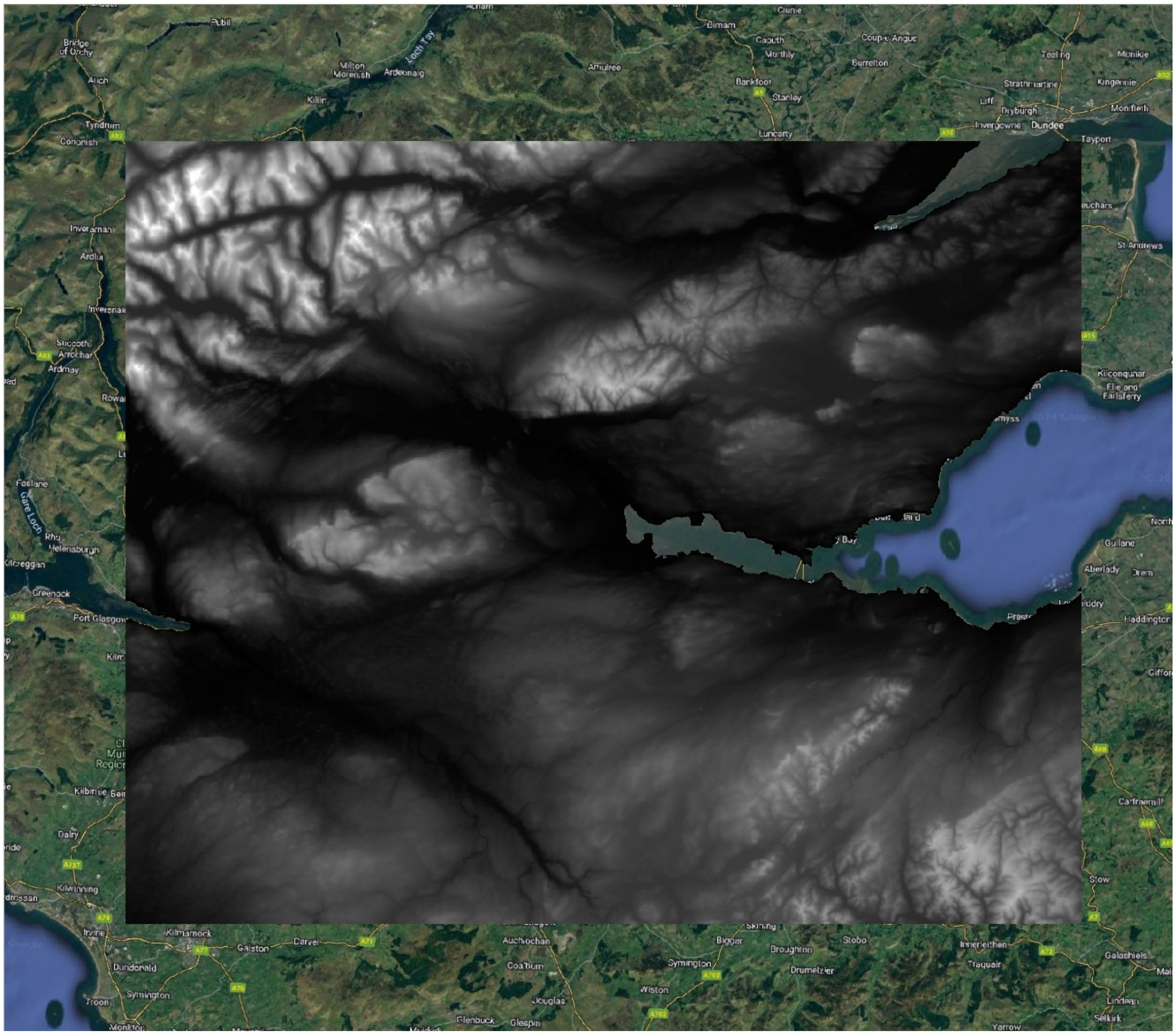
Earlier today you have learned how to use a vector polygon to mask a raster layer. But sometimes we just want to 'freeform' cut a piece of a raster, without the need to be too precise. In that case, we can use the `Clip Raster by Extent` function.

(147) Make sure you `Google Hybrid` web layer is visible, and navigate until you frame the 'Stirling-Glasgow-Edinburgh triangle' in your canvas, like the image below:



(@) Then go to `Raster > Extract > Clip Raster by Extent...`. Pick the SRTM layer as `Input layer`, and then for `Clipping Extent`, click on the small button to the right with a small arrow figure. That will set the `cut area` to be exactly what you are viewing on the canvas. But there are other options. If you click on the small down-arrow button to the right, you can use the extent (bounding box) of another layer, as well some more advanced options. You can also click on `Draw on map canvas` to be allowed to drag a rectangle over your map canvas that sets the extent of the cut.

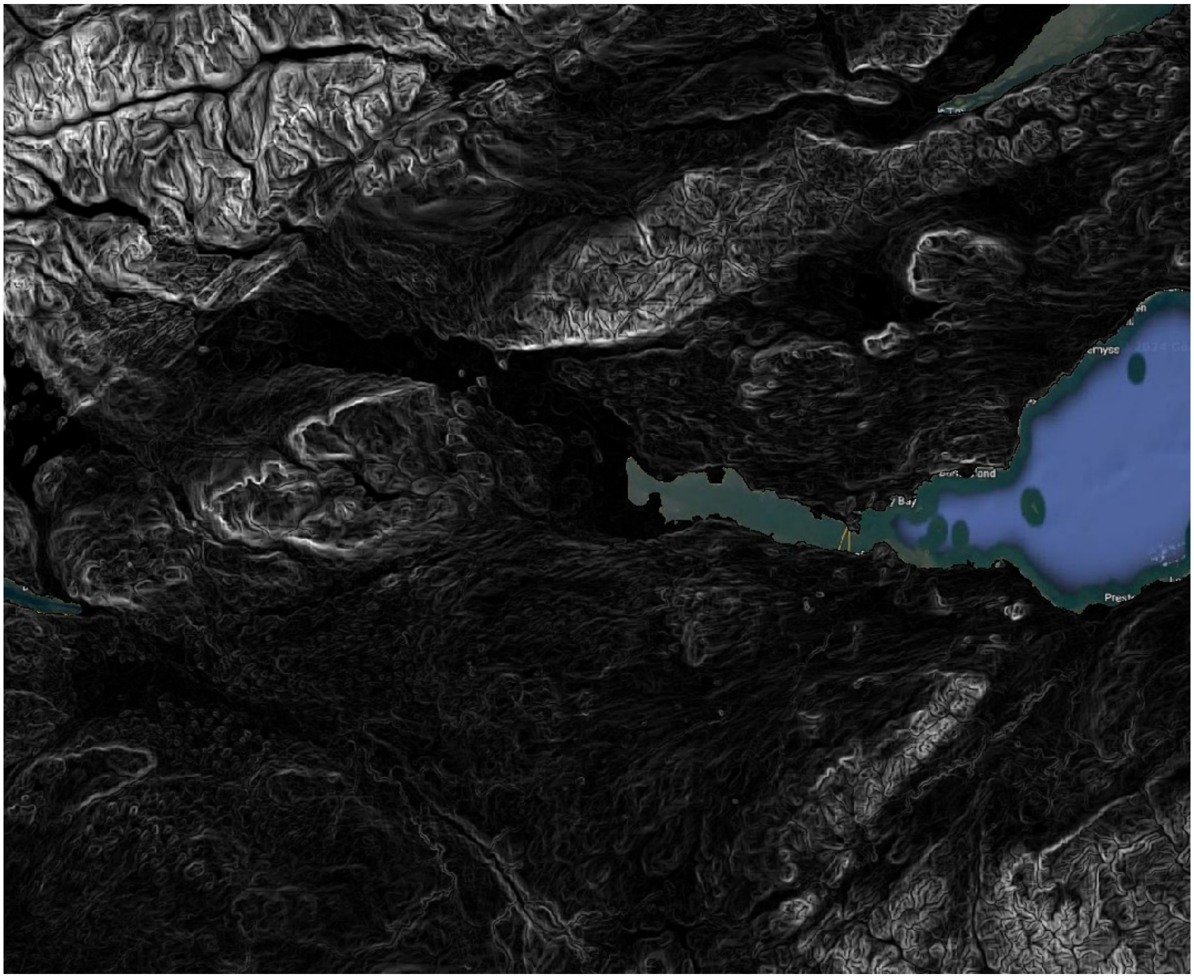
(148) Mask the SRTM layer to the region including Stirling, Glasgow and Edinburgh, either by setting you map canvas zoom and using it as extent, or by clicking and dragging to set the extent. Then pick a proper folder and save it as `UK_SRTM_BG_centralscotland.tif`. You should end up with something like this (but your extent will likely vary):



5.6.3 Calculating slope and aspect

We can now use our Central Scotland subset to demonstrate how to calculate slope and aspect.

(149) Go to `Raster > Analysis > Slope...`, and pick the Central Scotland DEM you created as `Input Layer`. Leave everything else as default and then pick a proper folder to save the new file as `UK_SRTM_BG_centralscot_slope.tif`. Then `Run` and `Close`. You will get a new raster layer, where the pixel values indicate the steepness of each pixel, in degrees. Steep slopes will appear as light grey/white, and flat areas will appear dark:

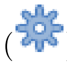


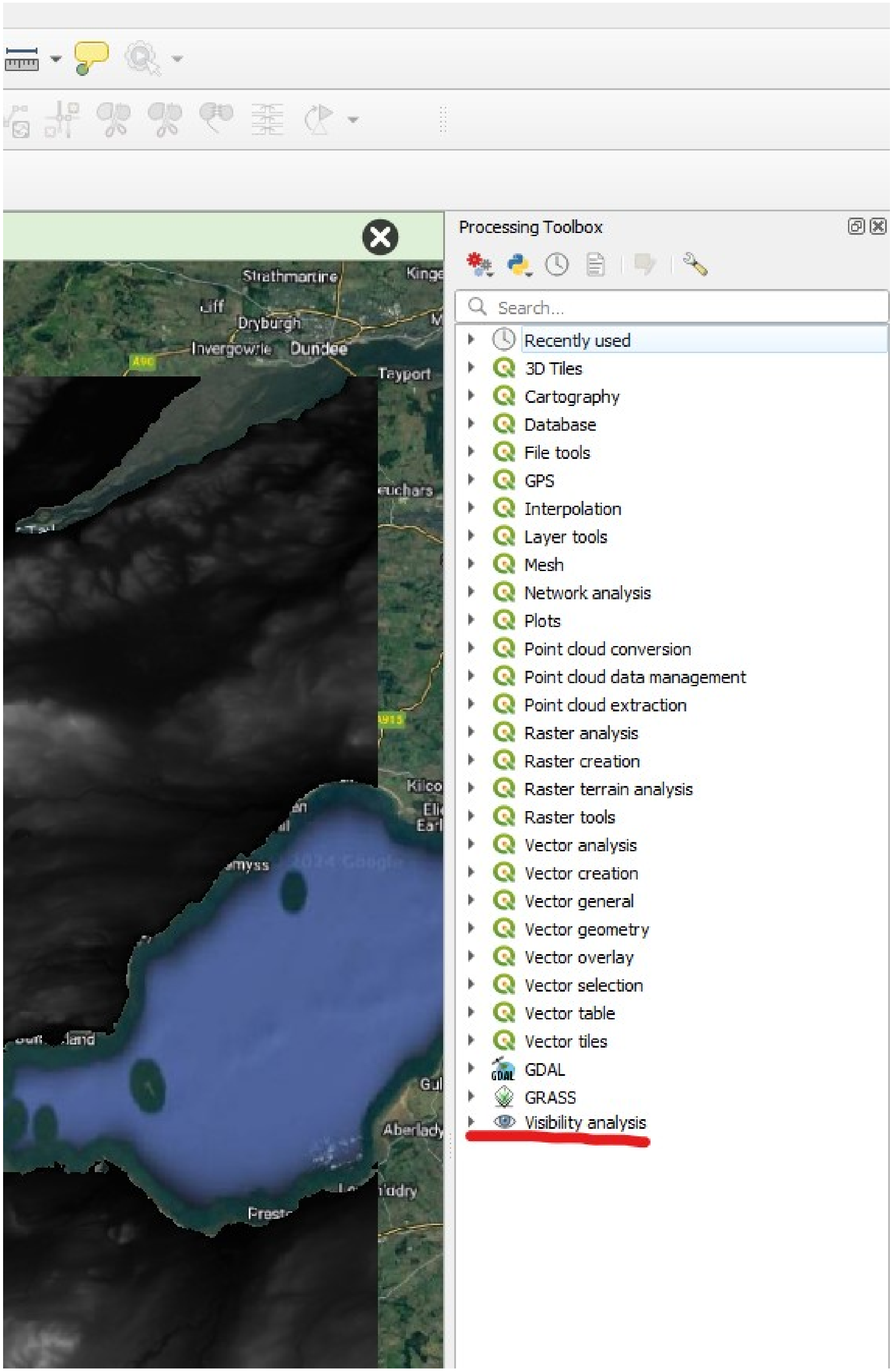
(150) Now go to `Raster > Analysis > Aspect...`, and again pick the Central Scotland DEM you created as `Input Layer`. Leave everything else as default and pick a folder to save the new file as `UK_SRTM_BG_centralscot_aspect.tif`. `Run` and `Close`. This layer will now tell you the cardinal direction, in degrees (North = 0/360),, that each slope is facing.

(151) Use the `Indentify Features...` tool to explore some of the aspect and slope values you have calculated.

5.6.4 Viewshed analysis

Finally, let us calculate a 'viewshed' - raster indicating which points in the Earth surface are visible from a given point, considering the topography. Viewshed analysis is often used for landscape planning - for example, determining from where a wind power turbine may be visible or not.

The `Visibility Analysis` plugin adds some option to the `Processing panel` of QGIS. To see it, click on the `Processing Toolbox` button on the main QGIS toolbar (), and the panel will open to the right. This panel houses many, many more GIS functions beyond those available on the `Vector` and `Raster` menus, and we will use it often during the last weeks of the module.


Processing Toolbox
Search...
Recently used
3D Tiles
Cartography
Database
File tools
GPS
Interpolation
Layer tools
Mesh
Network analysis
Plots
Point cloud conversion
Point cloud data management
Point cloud extraction
Raster analysis
Raster creation
Raster terrain analysis
Raster tools
Vector analysis
Vector creation
Vector general
Vector geometry
Vector overlay
Vector selection
Vector table
Vector tiles
GDAL
GRASS
Visibility analysis
Strathmartine
Dryburgh
Invergowrie
Dundee
Tayport

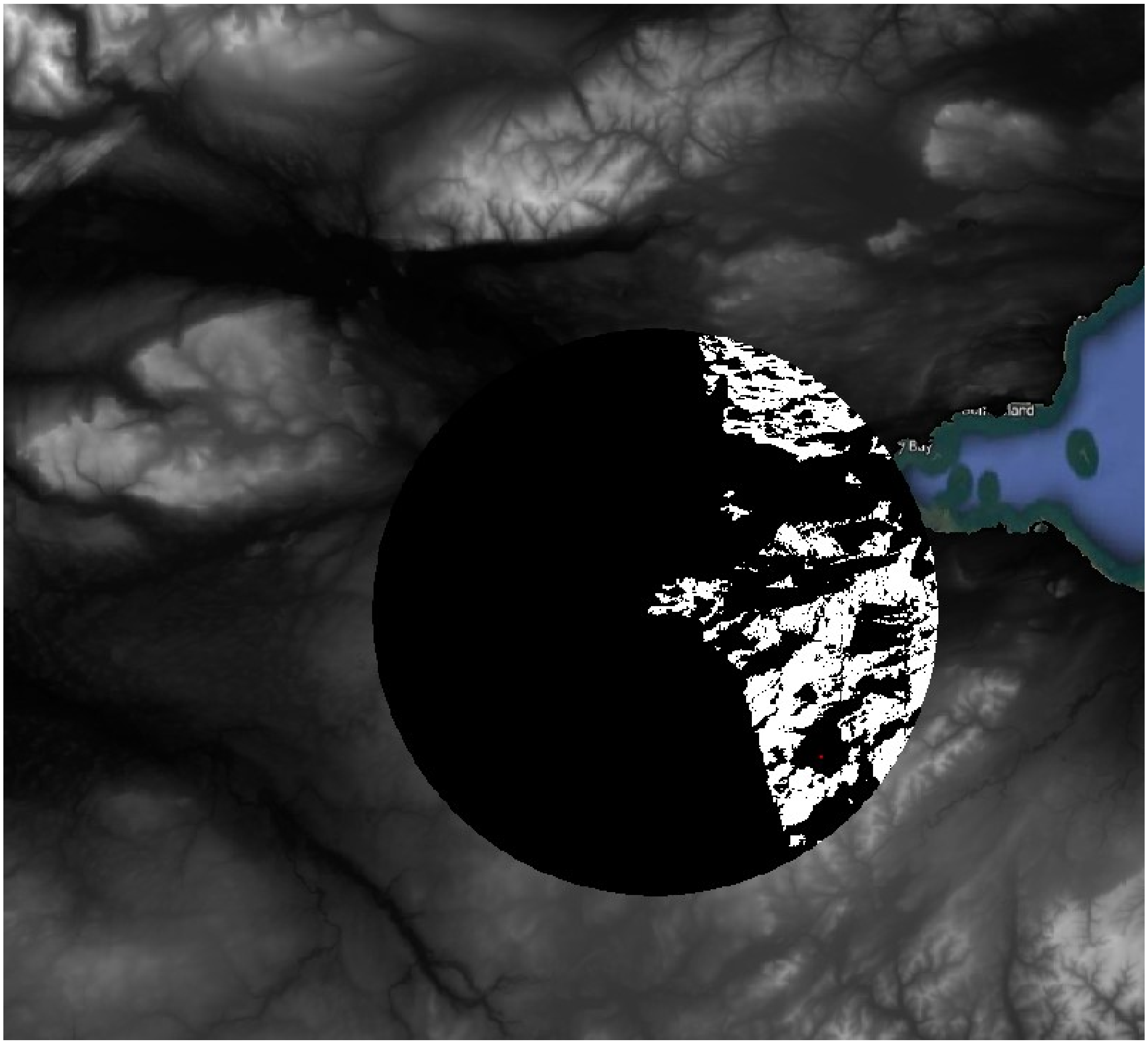
(152) Load the `observation_point,shp` file you have downloaded as part of the lab data.

(153) Open the `Visibility Analysis` heading on the `Processing Toolbox`, and then double click on `Create Viewpoints`. A new window will open. On this window, you need to pick the `Observer Location`, which will be the Observation point layer, and then the DEM, which will be the Central Scotland SRTM (make sure you *don't* pick the slope or aspect layers by mistake!).

(154) There are many options that you don't need to worry about for now, but three warrant some explanation: `Radius of Analysis` specifies how far you want to calculate the viewshed - it is an expensive computation so it may make sense to limit it. Let's use 20,000 meters for our analysis. Then pat attention to `Observer Height` and `Target Height`. They determine what is the height of the observation being made (above the elevation of the observation point), and also what would be the height of the target. So again, if you are wondering if a wind power turbine would be visible from that location, you should add the height of the turbine to the `Target Height` field. You can leave these options at 1.6m and zero, respectively. Then pick a folder and save the viewpoint as `viewpoint.shp`, and `Run` the tool.

(155) Now go back to the `Processing Toolbox` and double click on `Viewshed`. Pick the `Viewpoint` layer as `Observer Location`, and the Central Scotland DEM as `Digital Elevation Model`. Leave everything else as default, and pick a folder to save your viewshed analysis as `viewshed.tif` (it wil be a raster file).

You should get an output similar to the figure below, where white pixels (value `1`) indicate 'visible' and black pixels (value `0`) indicate "not visible'.



Guided Exercise 6 - Cleaning up and saving your project for the next lab

We will use the data you produced on this lab to get started on the next, so let us 'package'it properly.

(156) Remove all layers from the project except for the SRTM, CORINE, BIOCLIM and GADM layers that you have reprojected and masked for the GB island. Save your project

(157) Delete from the folder all files except for the files for the four data layers above. If you saved the GB island vector as a shapefile, remember you need to keep all files with the same name but different extensions (`.shp`, `.shx`, `.dbf`, `.prj`)

(158) On your systems file explorer, find the base folder for today's lab according to your organization, and right click on it. On Windows, choose `Compress to...` to create a zipfile containing your entire project. Name it as `lab_5_final_results.zip` or something similar.

(159) Now store this zipfile on your university's OneDrive or on some external drive, so that you can re-download it when you need it for the next session.

Congratulations, you reached the end of Lab 5. You should now understand the raster data model and how to mask it and style it. You have also learned some common terrain analysis tools. In the next lab, we will learn about the `Raster Calculator`, which fulfils the role of both the `Field Calculator` and `Select by Attribute` for rasters.

6 Lab 6: The raster calculator and other rastery bits

In this lab, we will continue working with rasters, and will learn how to do several new operations using them, including using the all-powerful *Raster Calculator*. We will also see the effects of *raster resampling* in practice.

6.1 Before you start!

1. Go through the Week 3 preparatory session on Canvas, and watch the seminar recording if you have missed it. Also make sure you have completed all labs prior to this one.

6.2 Guided Exercise 1 - Global raster statistics

In this exercise, we will use the layers from the previous lab to learn about raster statistics and the raster calculator.

6.2.1 Recovering your data

At the end of last week's lab you saved the final results of your project as a zipfile. Retrieve this file, extract the contents, and re-open your project. It should have these four layers in it:

- Polygon vector layer of the Island of Great Britain
- Raster layer of the SRTM digital elevation model (reprojected to EPSG:27700 and masked to the GB island)
- Raster file of the CORINE land cover dataset (reprojected to EPSG:27700 and masked to the GB island)
- Raster file of the BIOCLIM climatic dataset (reprojected to EPSG:27700 and masked to the GB island)

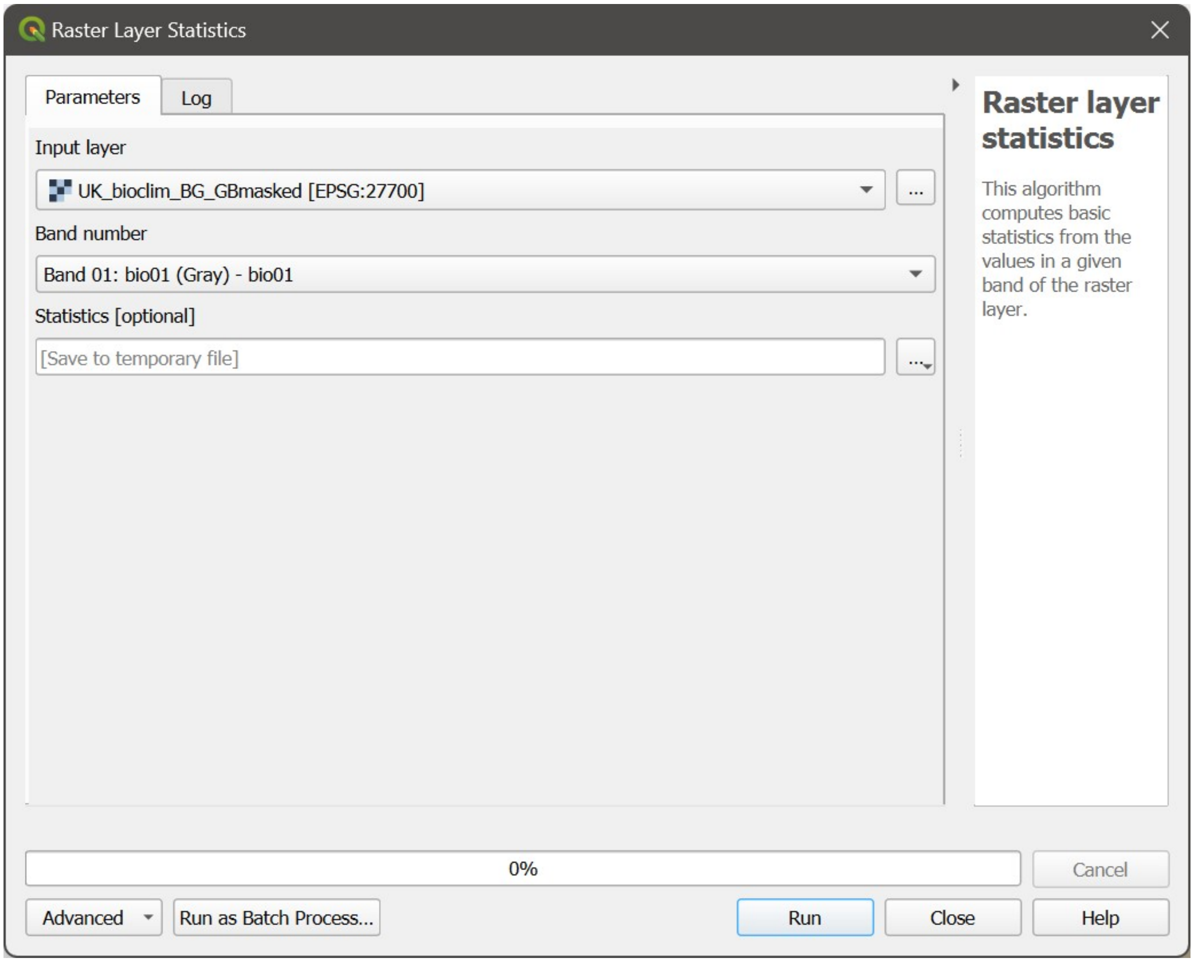
6.2.2 Global raster statistics

Winter is coming, and you want to escape the worst of the cold and rain. You therefore decide to use your GIS skills to find the ideal place to spend your winter in the UK - an urban area with lower precipitation and higher minimum temperatures. For convenience, we will restrict oour search to the Island of Great Britain, as we already have data for it ready to go.

To get started, you would like to know what is the overall range of precipitation and temperature values for the entire island, using the BIOCLIM data. If this was a vector dataset, you would use the `Statistical Summary Tool` - but it is a raster. What would be the equivalent operation?

(160) The tool we need is called `Raster Layer Statistics`, but it is *not* located in the `Raster` menu. Instead, we need to launch the `Processing` panel (. This panel contains a lot of additional GIS functions, and we will use it often now.

(161) Once you have the `Processing` panel open, you can either search for `Raster Layer Statistics` or find it under the `Raster analysis` heading. Double click on it to launch the tool window. It should look like this:



Looking at the BIOCLIM data description, the two useful pieces of climatic data you need to get statistics for are *BIO6 - Minimum Temperature of the Coldest Month, and BIO12 - Annual Precipitation*. The BIO variables are ordered in the file, so *band 6* is BIO6, and *Band 12* is BIO12.

(162) In the `Raster Layer Statistics` window, select the BIOCLIM layer as `Input Layer` and `band 06` as band number. You can keep the results as a temporaty file. `Run` it and `Close` the window. A new panel will have appeared under the `Processing` panel, called `Results Viewer`. It will have an entry called `Statistics`. Double click on it and take note of the minimum and maximum BIO6 values (-75 and 33).

! Stop and Think

Do these values seem too extreme for the UK? What is going on here?

Click for answer

The metadata for the BIO layers states that temperature values are scaled by a factor of 10, so you need to divide these numbers by 10 to get proper Celsius units.

(163) Now repeat the process to get the Min and Max values for BIO12 (532 and 2311)

! Stop and Think

What are the units for precipitation?

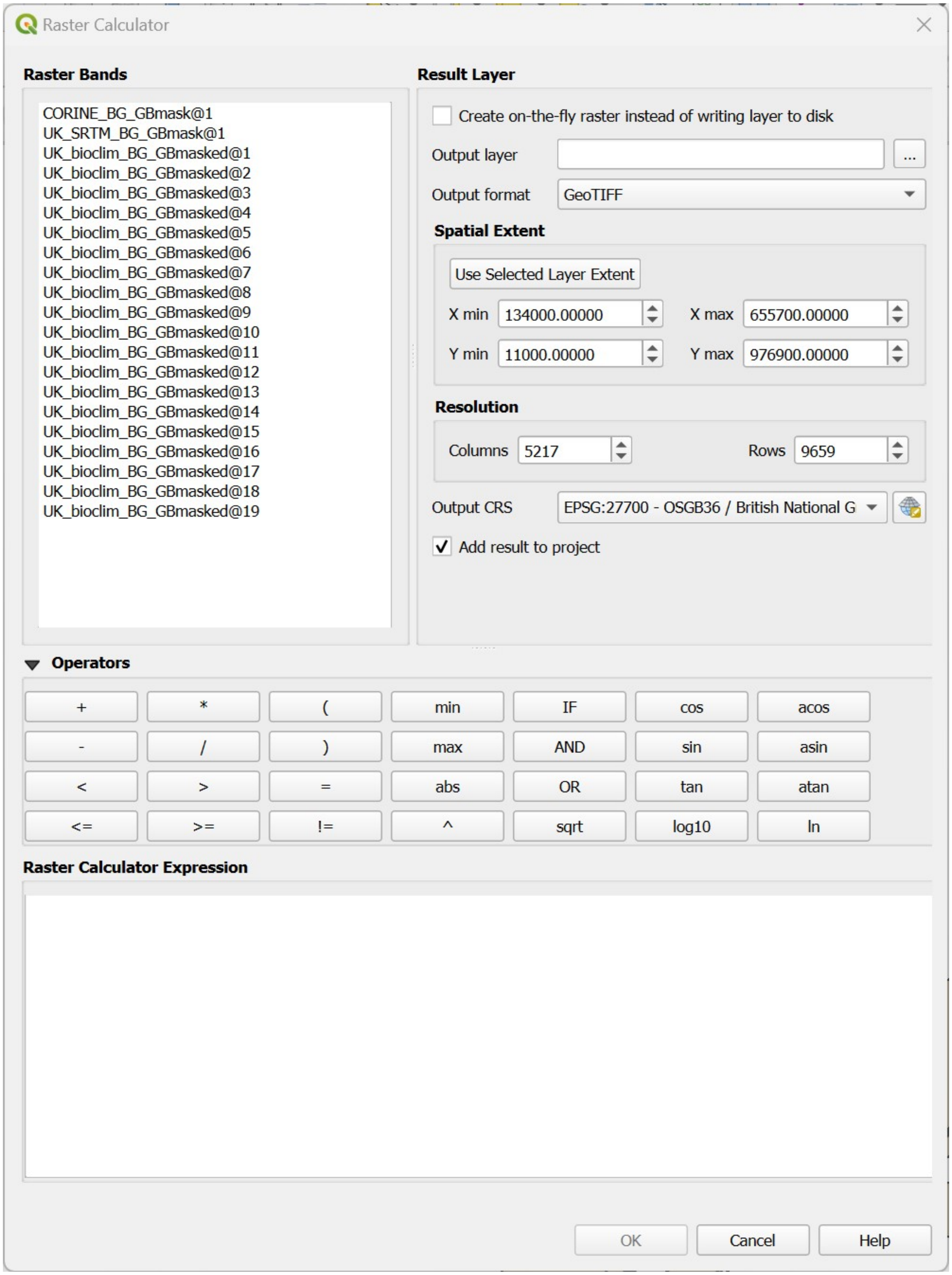
Click for answer

The metadata for the BIO layers states that annual precipitation is given in mm. One milimitre of rain is equal to one liter per squared metre.

6.3 Guided Exercise 2 - Per-pixel calculations with the raster calculator

Although it is easy enough to covert the temperature pixel values to Celsius in our heads, it would be useful to have the data in the proper units, especially if we want to map it later. Let us use the `Raster Calculator` to apply the conversion factor to all pixels.

(164) Go to the menu `Raster > Raster Calculator` to launch the calculator. It will look like this:


Raster Calculator
Raster Bands
CORINE_BG_GBmask@1
UK_SRTM_BG_GBmask@1
UK_bioclim_BG_GBmasked@1
UK_bioclim_BG_GBmasked@2
UK_bioclim_BG_GBmasked@3
UK_bioclim_BG_GBmasked@4
UK_bioclim_BG_GBmasked@5
UK_bioclim_BG_GBmasked@6
UK_bioclim_BG_GBmasked@7
UK_bioclim_BG_GBmasked@8
UK_bioclim_BG_GBmasked@9
UK_bioclim_BG_GBmasked@10
UK_bioclim_BG_GBmasked@11
UK_bioclim_BG_GBmasked@12
UK_bioclim_BG_GBmasked@13
UK_bioclim_BG_GBmasked@14
UK_bioclim_BG_GBmasked@15
UK_bioclim_BG_GBmasked@16
UK_bioclim_BG_GBmasked@17
UK_bioclim_BG_GBmasked@18
UK_bioclim_BG_GBmasked@19
Result Layer
Create on-the-fly raster instead of writing layer to disk
Output layer
...
Output format
GeoTIFF
Spatial Extent
Use Selected Layer Extent
X min
134000.00000
X max
655700.00000
Y min
11000.00000
Y max
976900.00000
Resolution
Columns
5217
Rows
9659
Output CRS
EPSG:27700 - OSGB36 / British National G
Add result to project
Operators
+
*
(
min
IF
cos
acos
-
/
)
max
AND
sin
asin
<
>
=
abs
OR
tan
atan
<=
>=
!=
^
sqrt
log10
ln
Raster Calculator Expression
OK
Cancel
Help

The general idea is similar to the `Field Calculator`, with a slightly different layout. The top right panel lists all raster *bands* in the project (so `UK_bioclim_BG_GBmasked@19` means band 19 of the BIOCLIM raster). The top right panel lets us pick some options for our raster creation, and the bottom panel lets us type expressions.

(165) We need to divide all BIO6 temperature values by 10. Double click on the BIOCLIM band 6 in the bands list to add it to the expression panel, and then add the division by `10`. On my project, the final expression panel reads as `"UK_bioclim_BG_GBmasked@6" / 10`, but your BIOCLIM layer may not be named the same as mine. Note that the bottom of the expression panel should say `Expression Valid`. If it says `Expression Invalid`, check your typing (did you use the double quotes?)

(166) For `Output Layer`, click on the `...` button and pick a folder to save the new file. Name it `UK_bioclim_BG_GBmasked_BIO6_celsius.tif`. Leave the rest as default and click in `OK`.

! Stop and Think

How many bands has your newly created file?

Click for answer

Only one band, holding the results of the calculation.

Every time you enter an expression involving a raster band and a single number in the `Raster Calculator` you are telling QGIS to apply the same expression to each pixel of that band. In our case, we divided all pixel values by 10.

6.3.1 Guided Exercise 3 - Boolean (logical) operations with the raster calculator

The `Raster Calculator` also fulfills the role of the vector `Select by expression` tool for rasters. The result of any raster logical operation is either `1` (True) or `0` (False), i.e. a *binary raster*, sometimes called a *raster mask*.

Let us find the warmer places in Great Britain. Our highest minimum temperature of the coldest month (BIO6) value was 3.3 degrees, not too far from zero. So let us limit our search to any places that don't go below 0 Celsius.

(167) Open the `Raster Calculator` again, and this time use the expression `"UK_bioclim_BG_GBmask_BIO6_cels` `> 0`. Notice we used here the layer that we just converted to degrees Celsius. Save the results in a proper folder with the name `GB_mintemp_gt_0.tif` (gt for 'greater than') and run the calculation. You should get something like this:

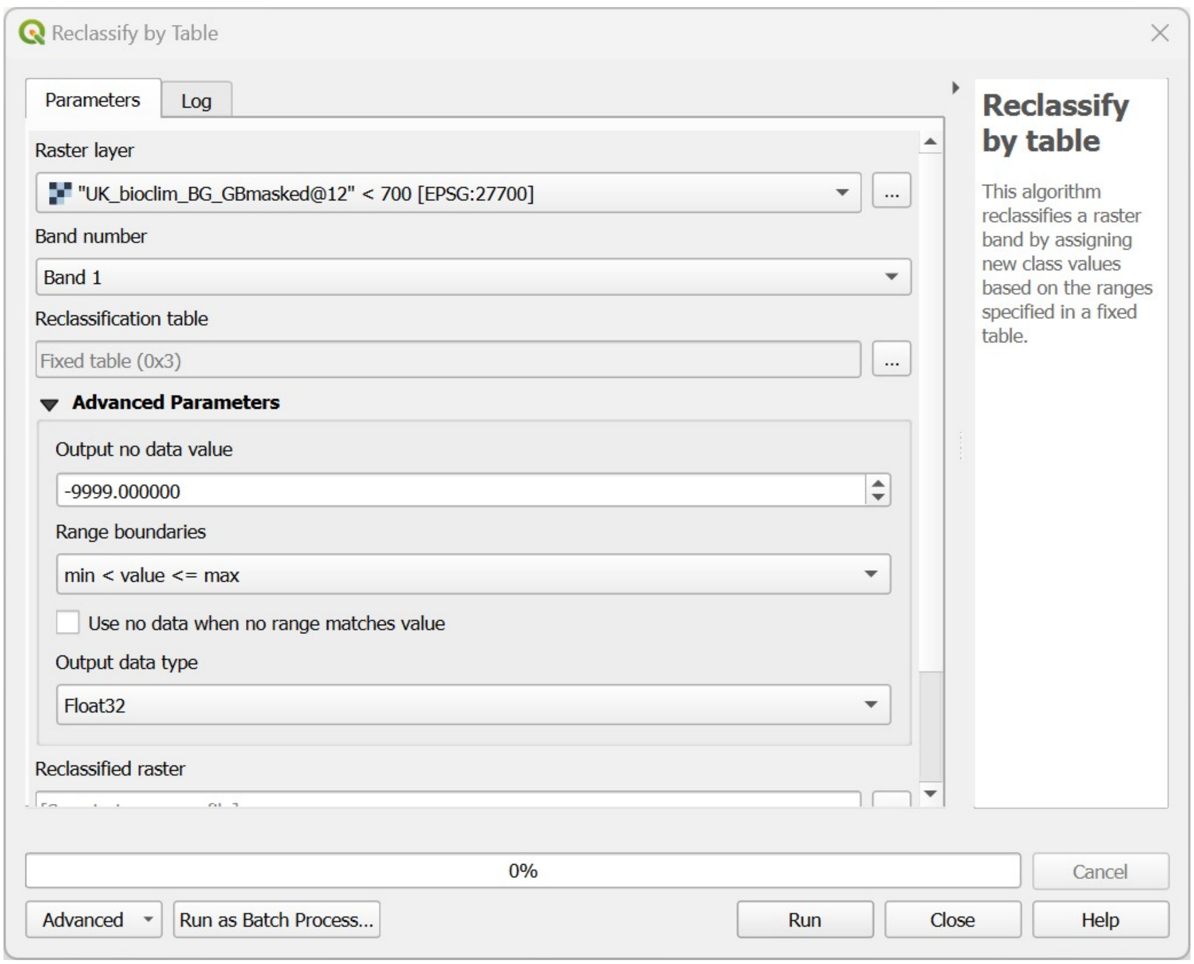
By default,

QGIS will paint pixels with a value of `1` white, and `2` black. So the white areas now show us all regions of Great Britain that *on average* don't go below zero in the worst of winter (these are long term climatic averages). Use the `Identify Features` tool to check if the pixel values are really `0` and `1`.

(168) Now repeat the process for the precipitation layer. Our minimum annual precipitation value was just above 500, so let us limit us to areas under 700mm of precipitation. Save the result as `GB_anprec_lt_700.tif`.

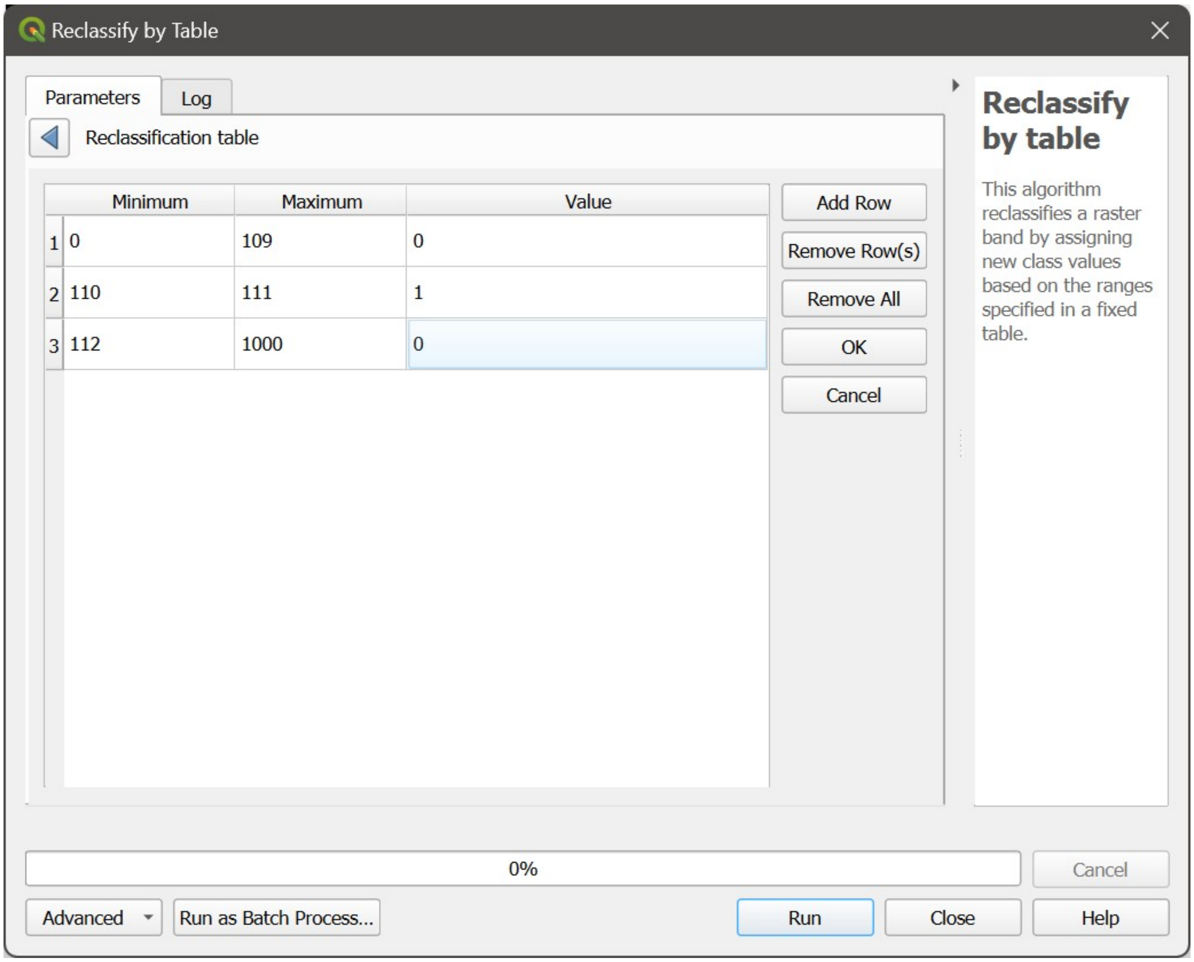
6.3.2 Guided Exercise 4 - Raster reclassification

An alternative to the `Raster Calculator` that can be handier when you have multiple ranges to recode is the `Reclassify by table` tool in the `Processing` panel, also under the `Raster Analysis` heading. Find the tool and open it, and you will see this:



Let us use this tool to extract the urban areas from the CORINE dataset. looking at the metadata, the two classes od interest are `Continuous Urban Fabric` (pixel value 111) and `Discontinuous Urban Fabric` (112) classes, which are the classes were we would expect existing housing to be located.

(169) On the `Reclassify by table` window, pick the CORINE layer as your `Raster layer`, and `band 1` as your `Band Number`. Then click on the `...` button besides the box to open the `Reclassification table`, and fill the table like the figure below. Click to add rows as necessary.



(170) Click on te blue arrow to go back to main tool window, then set the `nodata value` back to `-32768`, the `Range Boundaries` to `min <= value <= max` and the `Output Data Type` to `Byte`. Save it as `GB_urban_areas.tif`.

(171) Repeat the process above with the expression .

! Stop and Think

How could you have done the analysis above using the `Raster Calculator` instead?

Click for answer

You could do `"CORINE_BG_GBmask@1" = 111 OR "CORINE_BG_GBmask@1" = 112`. Notice how you can use `AND` and `OR` to chain logical expressions as you can for vectors - but because computers are dumb, you need to repeat the name of the layer after the `OR` (and for each additional expression you add).
You could also do ranges, using an expression such as `"CORINE_BG_GBmask@1" > 110 AND "CORINE_BG_GBmask@1" < 113` (notice the use of `AND` instead of `OR` for ranges) since they are consecutive values, or use `"CORINE_BG_GBmask@1" >=111 AND "CORINE_BG_GBmask@1" <= 112`). Using ranges is more quicker than using `=` and `OR` when you have many consecutive values to search for.

6.3.3 Guided Exercise 5 - raster vs. raster calculations

We now have three *raster masks* showing us 1) all the GB areas that don't go below zero Celsius, 2) the areas that have less than 700mm of rain, and 3) the areas that are urban. How can we combine them into a single layer?

The answer to the above question is still the `Raster Calculator`. We know the pixels we 'want' are numbered `1` on each layer, otherwise they are `0`. So if we could multiply one layer by the other, i.e. multiply the value of each set of overlapping pixels, we would get a new raster whith values of $1 \times 1 \times 1 = 1$ when both conditions are met, and a result of 0 if any of the the conditions is not met. This is why binary rasters are also called raster *masks* - when you multiply any other raster by it, it *masks* (i.e. zeroes out) anything that is not `True` in the mask.

Doing it using the `Raster Calculator` is as simple as it sounds:

(172) Open the `Raster Calculator` and enter the expression `"GB_anprec_lt_700@1" * "GB_mintemp_gt_0@1" * GB_urban_areas@1`. Then save you layer with name `UKs_winter_havens.tif`.

6.3.4 Guided Exercise 6 - Vectorizing rasters

The final result of your analysis should consist of only a handful of pixels - such is lifev in the UK. And since these are discrete locations, it may occur to you they would be better represented and visualised as vectors. Turns out we can convert raster to vectors and vice versa:

(173) Go to `Raster > Conversion > Polygonize (Raster to Vector)...`. In the new window, pick your results layer as the `Input Layer`, `Band 1` as the `Band number`, and leave the rest as default (`DN` here means *digital number*, another name for pixel values. Then pick a folder ans save the results as `UK_winter_havens_poly.shp` (or geopackage if you prefer).

(174) You will notice that a very large polygon was also created for the 0 areas of the raster. You can then put the new vector layer in editing mode, use the `Selection Tool` to select this polygon, and then hit the `Delete` key on your keyboard to get rid of it. Then save the changes and exit edit mode.

> ⚠ Warning
>
> Polygonising rasters can be a very computationally-heavy operation if you have large rasters and/or many scattered pixel values (for example, vectorising the original CORINE layer with all classes). And it is virtually useless for continuous rasters such as temperature or elevation (you would end up with one polygon per pixel as there would be no adjacent pixels with the same value).
> So if you start an analysis with raster data, try to do as much as you can in the raster domain, before vectorising anything. But if your raster results represent isolated small regions within a large matrix of 'no data' or zero values, then vectorising is a good idea.

You now have polygons representing all adjacent areas that were values as '1' in your raster. But many of them are still only one pixel. What if you wanted to have points instead of polygons?

(175) In the `Processing` panel, you will find a tool called `Raster Pixels to Points`, which you can use to generate the points. I'll leave the detauils to you, but here is a tip: before you use the tool, go the `Properties > Transparency` of your raster results layer, and set 0 as an `Additional Nodata Value`. Otherwise `Raster Pixels to Points` will create one point for every pixel, including the pixels with a value of 0. That's a lot of points.

6.3.5 Guided Exercise 7 - Mosaicking and Stacking rasters

Raster files are often very large and 'heavy', so it is common to distribute them as *tiles* or *scenes*, i.e. smaller adjacent pieces that can be merged back together to cover a certain area of interest. For satellite images, it is also common for each image band to be stored as a separate file (unlike regular photos that only have Blue, Red and Geeen bands, satellite images can have several more bands, covering areas of the spectrum we can't normally see. The Sentinel-2 satellite, for example, has a total of 13 bands:

Band	Wavelength	Description
B1	443 nm	Ultra Blue (Coastal and Aerosol)
B2	490 nm	Blue
B3	560 nm	Green
B4	665 nm	Red
B5	705 nm	Visible and Near Infrared (VNIR)
B6	740 nm	Visible and Near Infrared (VNIR)
B7	783 nm	Visible and Near Infrared (VNIR)
B8	842 nm	Visible and Near Infrared (VNIR)
B8a	865 nm	Visible and Near Infrared (VNIR)
B9	940 nm	Short Wave Infrared (SWIR)
B10	1375 nm	Short Wave Infrared (SWIR)
B11	1610 nm	Short Wave Infrared (SWIR)
B12	2190 nm	Short Wave Infrared (SWIR)

In this exercise, we will learn how to a) *mosaic* rasters - 'glue' them together side by side to cover a larger area and b) *stack* rasters - merge the different band files into a single file.

(176) Download the data for this specific exercise from this link. It contains four aerial photographs covering the University of Stirling campus and the Wallace Monument area. Extract the data and organise it as usual.

> Tip
>
> These images were originally downloaded from the 'Aerial' section of Digimap - if you ever need very detailed imagery of the UK land, check it out!

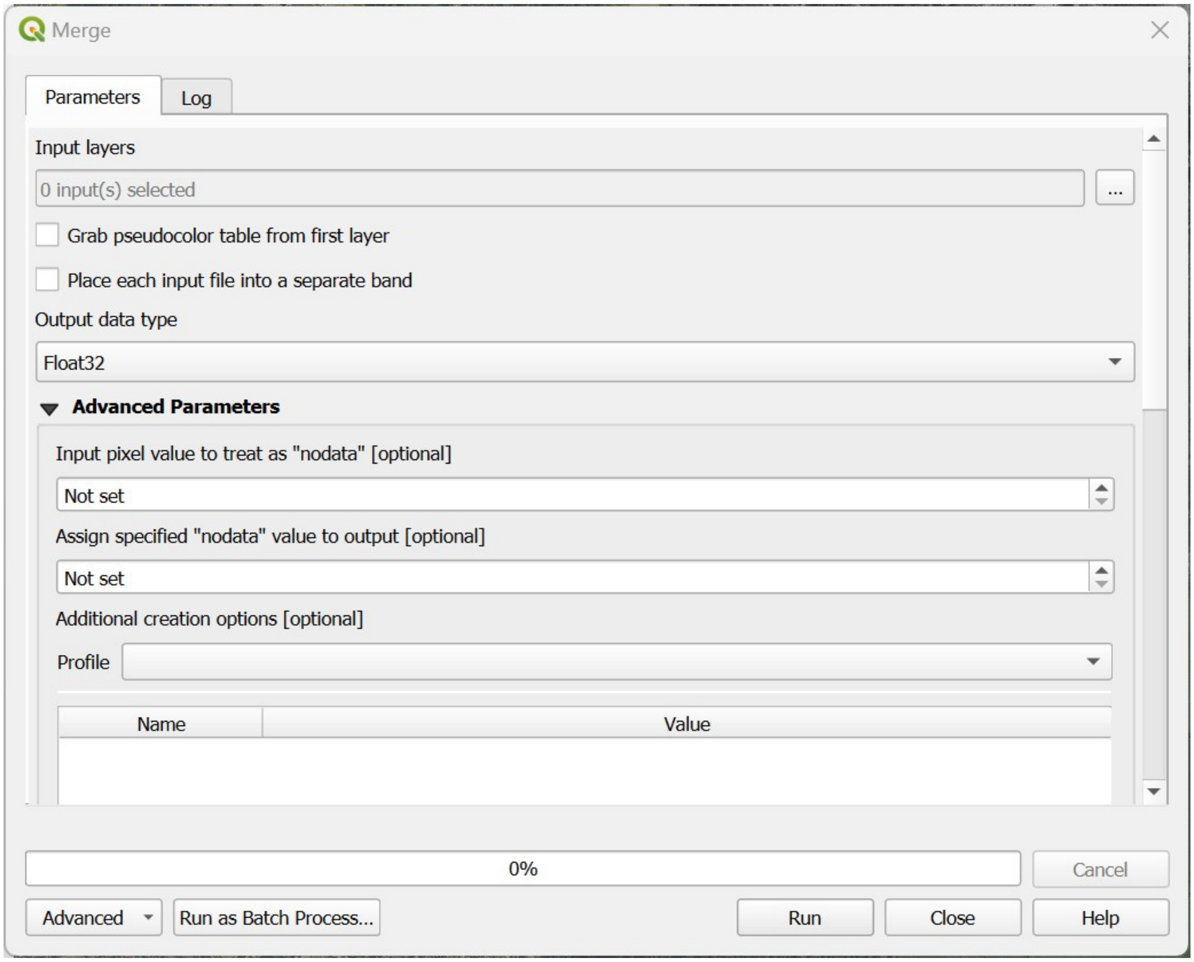
(177) Each image is stored within its own folder, named `nsXXXX`, where `XXXX` is a four digit number. *Start a new project* and load into QGIS all the `.tif` files in all four folders. There should be six files in total.

> Stop and Think
>
> What do you think is the meaning of the `nsXXXX` folder names?

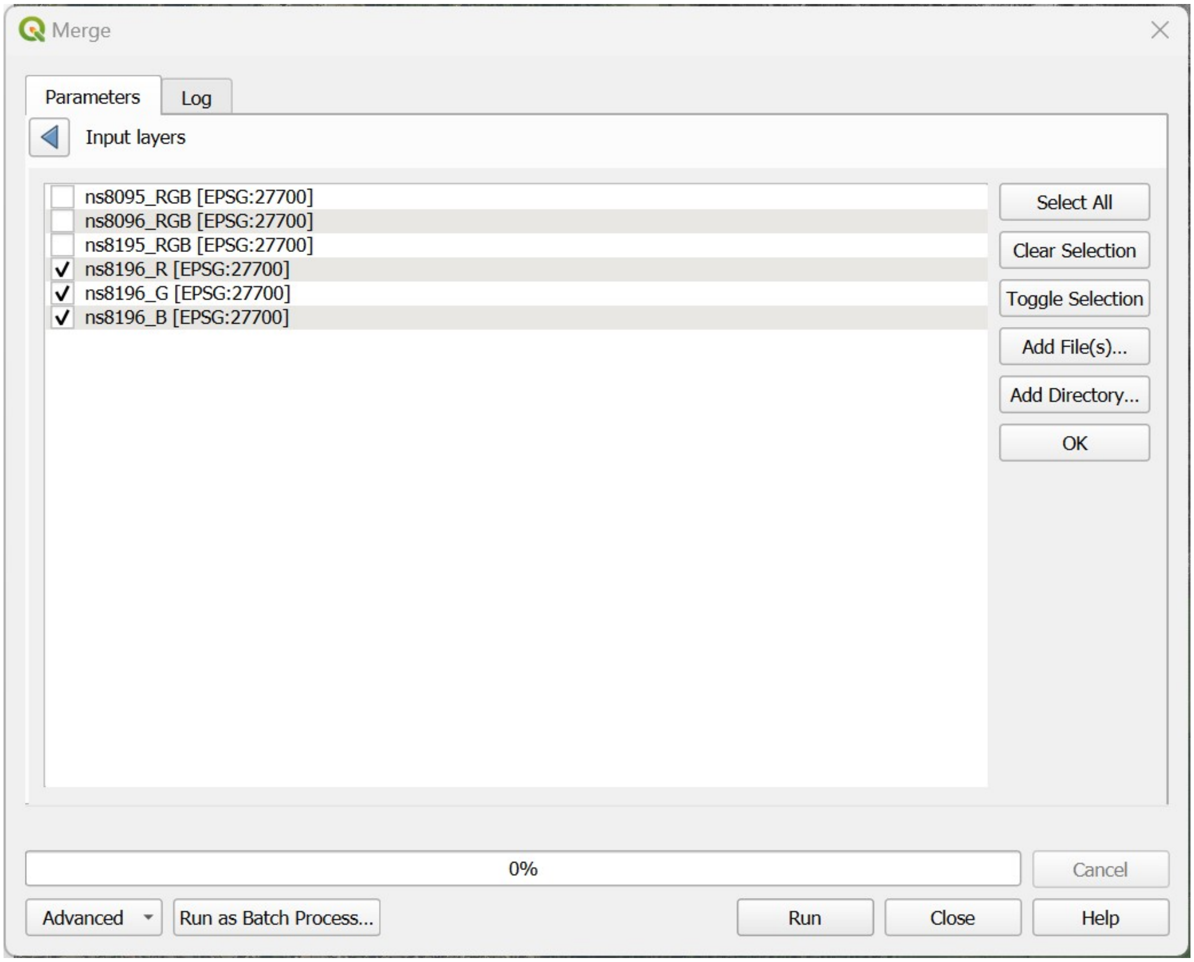
> **Click for answer**
>
> They indicate tiles in the OS British Grid system.

Notice that the airphoto for tile ns8196 has been provided with its three colour bands (RGB) as separate files - simulating what you would expect for satellite files. To be able to see this aerial photo in colour, we must first *stack* these bands into a single file. The operation is called *stack* because you could visualise it as physically stacking them one on top of the other.

(178) Go to `Raster > Miscellaneous > Merge...`. You will get a window like this:



(@) For `Input Layers`, click on the `...` button, and a list of all available raster layers in your project will appear. You can drag the layer names to reorder them, and your final selection should the `ns8196_R`, `ns8196_G`, and `ns8196_B` layers, in the order below from top to bottom. This sets the order of the bands inside the file.



(@) Once you have selected the layers, click on the blue arrow to go back to the main window, then set the following parameters: *check the box* for `Place each input file into a separate band` (that tells QGIS you want a *stack*), and the `Output Data Type` to `Byte`. Then save the resulting file as `ns8196_RGB.tif`, in the same folder where you had the three separate band files.

! Stop and Think

Why do we change `Output Data Type` to `Byte`?

Click for answer

The RGB colour system uses one byte (8-bits, 0-255 integer numbers) per colour channel to designate any colour. For example, the University of Stirling official green colour () is R=0 G=105 B=56. The `Byte` option ensures we only use 8-bits per band, and keep the file size smaller. If we left it as `Float32`, the same 0-225 integer numbers would be stored as 32-bit decimal numbers, quadrupling the size of the file for no reason.

(179) You should now have a properly colored image for tile ns8196. Remove the three original single band layers from the project and then save it.

Everything looks good colour-wise now, bur our airphoto is still actually four separate adjacent photos. That means any processing we may want to do involving the entire area would have to be repeated four times. We can however merge these four tiles in a single airphoto *mosaic*.

(180) Go back to `Raster > Miscellaneous > Merge...`. Yes, it is a bit confusing, QGIS uses the same tool for both stacking and mosaicking.

(181) This time, check the four tiled RGB images as your `Input Layers` - order doesn't matter. Then make sure to **not check** the `Place each input file into a separate band` option. That will tell QGIS you want a mosaic and not a stack. Set the `Output Data Type` to `Byte` again, and save it as `UoS_RGB_mosaic.tif`. You should now have a single airphoto layer covering the entire area. Remove the previous individual layers from the project and save it again.

6.3.6 Guided Exercise 8 - Raster image stretching

When working with multiband color images in a GIS environment, we can often manipulate the contrast of an image by applying different *contrast stretches* to each band. Especially when working with satellite images, which tend to look a bit 'faded' because of the atmospheric effect, stretching can make our images more vivid and easier to interpret.

(182) Right-click on the airphoto mosaic layer you created and go to `Properties > Symbology`. You will se that the symbology type is `Multiband color`, which is the correct choice for representing images. Note that each band is associated to a colour channel (R, G, B), but this association can be changed at will. Change the `Red band` to band 2, and the `Green band` to band 1, then `Apply`. Now the vegetation appears red! Change it back to the proper colour order and `Apply` again.

Tip

This is a useful feature when working with satellite images, which often have bands representing spectral regions that we cannot see, such as Near Infrared or Microwave. We can then freely associate these bands to colour channels and create `false colour compositions` that highlight different surface features. The example below shows the UoS campus as seen from the Sentinel-2 satellite. Here we place Band 13 - Short Wave Infared (SWIR) in the blue channel, Band 8 - Near-Infrared band (NIR) in the green channel, and Band 11 - SWIR in the red channel. Since water does not reflect IR, the loch appears black. Plants reflect strongly in the NIR region so they appear bright green, and bare ground / concrete reflects mostly in the SWIR, thus appearing purplish (red+blue).

(183) Notice how the `Min` and `Max` values for each band are `0` and `255`, meaning we are mapping our screen colours to the entire range of possible values (0-255). But quite often the colours in an image do not cover the entire 8-bit range, so we are 'wasting' colour discrimination.

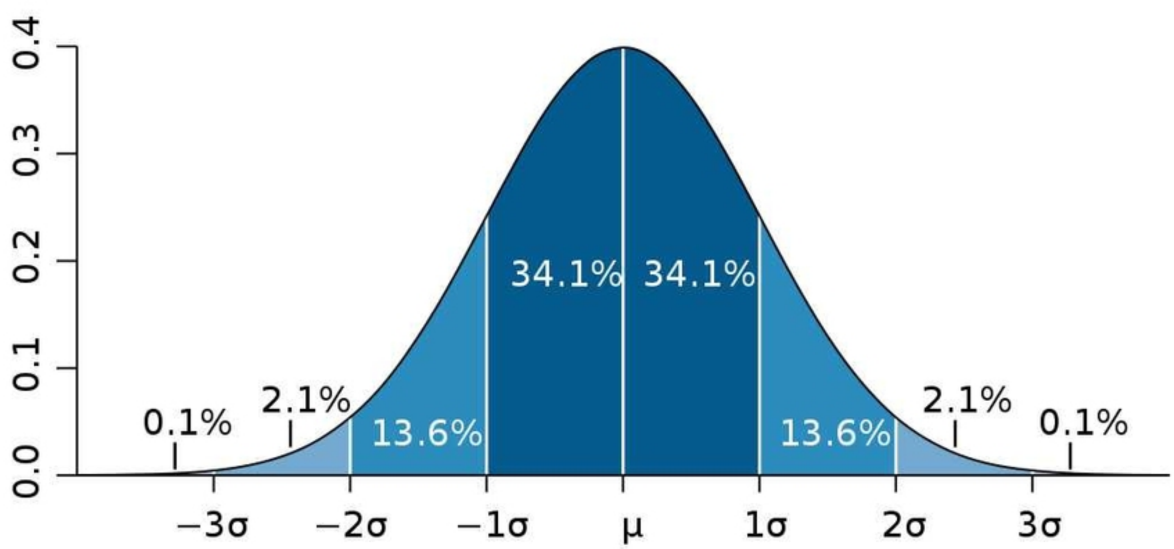
(184) Change the `No Enhancement` option to `Stretch to Min/Max`, and then expand the `Min/Max Value Settings` heading. Pick the `Min/Max` option, and select `Whole Raster` for `Statistics Extent` and `Actual(slower)` for `Accuracy`, then click on `Apply`. You will see that while band 1 and band 3 do use the entire 0-255 range, band 2 uses the 2-255 range only.

To further enhance the contrast of images, we can tell QGIS to 'cut off' the extreme values of the range, using different methods.

(185) Change the `Min/Max Value Settings` option to `Cumulative Count Cut`, and then`Apply`. Now the `Min` and `Max` colour values for each band are (13-189, 29-187,and 31-164 respectively), and the image should look more vivid. What QGIS did was remove the values in the lowest 2% (0.2 percentile) and highest 2% (0.98 percentile) of the existing range before calculating Min/Max values, thus *stretching* this smaller number range to the full colour range of the screen.

(186) Progressively increase the lower percentile (i.e. from 2% to 3%) and decrease the upper percentile (i.e. from 98% to 97%) and notice how the contrast gets progressively stronger. If you go too far (i.e. too narrow a range) you will start to see *pixel saturation* - many of the pixels will be outside the range and thus mapped to fully black or fully white.

A second way to calculate the stretch is to use standard deviations. You may remember from statistics that when you assume a Normal distribution, a distance of $\pm 1\sigma$ from the mean will capture about 68% of the data, $\pm 2\sigma$ will capture 98%, $\pm 3\sigma$ will capture X % and so on:



(187) Change the `Min/Max Value Settings` option to `Mean +/- Standard Deviation x`, and leave the multiplier at `2`. Aplly and check the contrast. Then change the multiplier to `1` and `Apply`, and compare the contrast of using $\pm 1\sigma$ vs. $\pm 2\sigma$.

(188) Set the final image contrast to your linking, and exit the `Properties` window. Save your project.

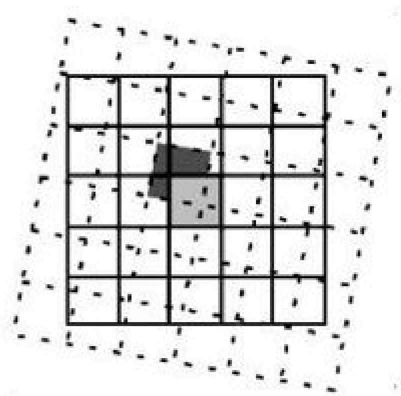
6.3.7 Independent Exercise - Raster resampling

As you have learned, raster data consists of a *grid* of pixel values (i.e. a matrix). So quite often, when working with rasters, you may need to shift the pixels from one grid to another. One main example is raster reprojection - if you reproject your raster to a different CRS, the new raster grid will be aligned with this new CRS, and won't match the original perfectly, as in the figure below.

FIGURE mismatched grids.

That is why the raster reprojection tool is called `Warp(Reproject)` - it is as if you are *warping* the image during the transformation. Other times you may need to make two different raster datasets match the same extent and pixel size, which would also require warping one of the rasters to match the other.

A key step when doing these transformations is choosing a *resampling* method, i.e. in which way will the original pixel values be transferred to the new grid. The most common and fast method is called *Nearest Neighbour* (NN) - if you imagine the old and new grids partially overlapping, NN resampling just looks for the closest old pixel to the new pixel, and transfers the value:



NN is fast because it requires no additional computations, and also has the benefit of not changing the original data values. But if your data has a lot of fine and/or linear elements, it may result in a 'jagged' appearance. For that reason, other methods exist that try to *interpolate* the new pixel values based on a combination of the original ones. Two common methods are the *bilinear* and *bicubic* interpolations, which will average the 4 and 16 closest pixels, respectively. That has the benefit of creating smoother resulting rasters:

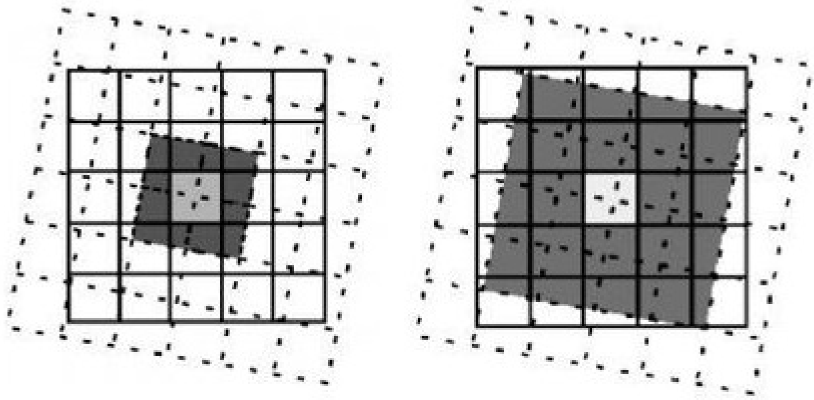
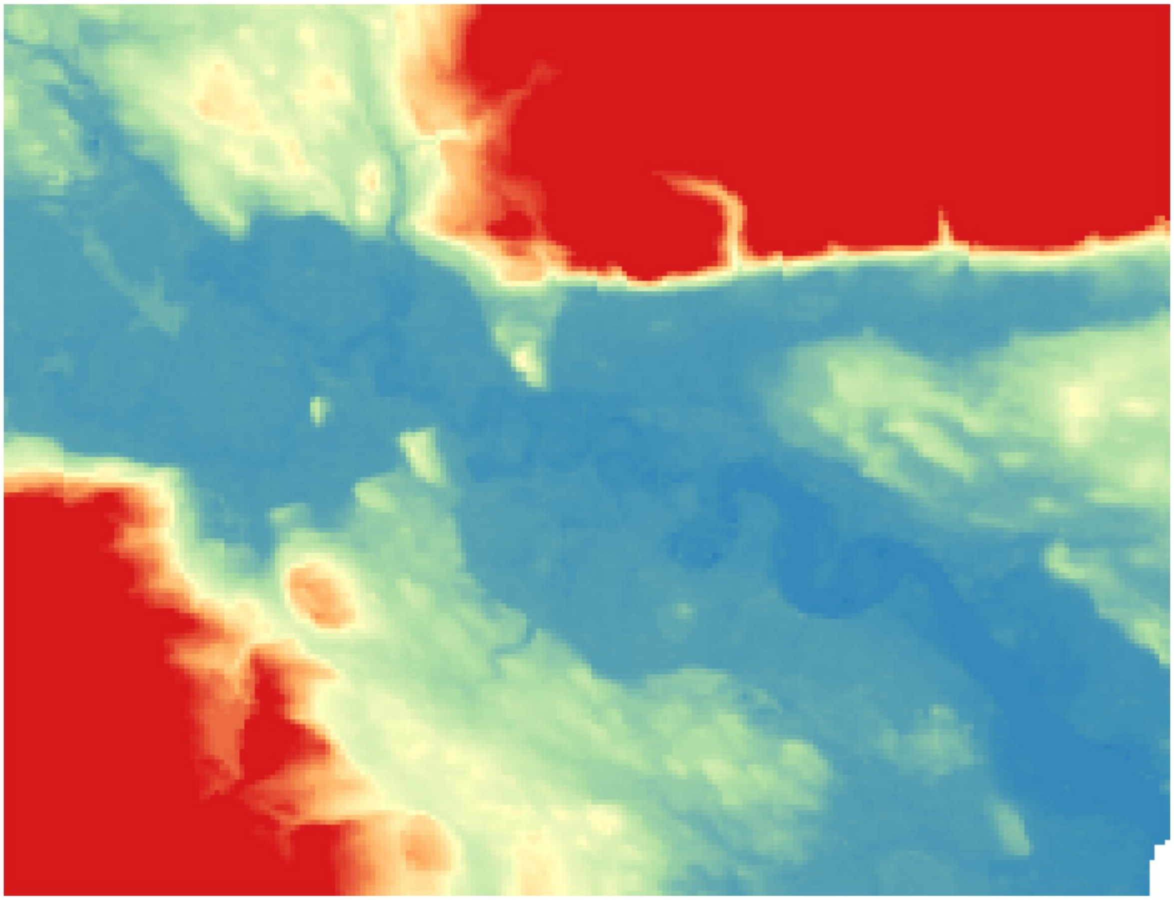


Figure 6.1: Source:10.13140/RG.2.2.16038.27207

Let us now practice our raster skills in an independent exercise, while also demonstrating the effect of resampling:

1. On your UK SRTM layer, find the location of Stirling and the university campus, and then use the `Raster > Extraction > Clip Raster by Extent` function to clip it to an area similar to the image below. For reference, I am using a `Singleband Pseudolcolor` symbology with the inverted `Spectral` colour palette, *stretched* to the values between 0 and 200m. If you find it hard to find Stirling, use the OpenStreetMap online layer as a reference. Don't worry about matching my area perfectly, the clip is just to speed up our computations.



2. Now use `Warp(Reproject)` to reproject this dataset back to EPSG:4326, but twice. Do it the first time using `Nearest Neighbour` as the `Resampling method to use`, and then a seocnd time using the `Cubic(4x4 kernel)` resampling method. Remember to pick the correct layer to reproject (it should be your Stirling-clipped SRTM layer). **Hint:** If you save it as a temporaty file, it may not recognise the CRS after the warp - you will see a question mark to the right of the layer name. If that happens, just click on the question mark and pick ESPG:4326 as the CRS.

3. Apply a symbology to the original Stirling SRTM that makes it easier for you to see all terrain features. Then use `Copy Style` and `Paste Style` to make sure both of the new warped versions have the same symbology.

4. Flick back and forth between the original, NN resampled and cubic resampled rasters - can you see how NN results in more 'jagged' edges where you have sharp terrain transitions?

5. Calculate the global raster statistics for the three layers and assess how much impact NN vs. cubic resampling has had on the new layers, compared to the original.

One instance where we should *always* use NN resampling is when we have categorical rasters such as the CORINE dataset. In these cases, any sort of resampling that averages multiple pixel values will create 'fake' new numbers that don't represent any class. Let us test it:

6. Using `Raster > Extract > Clip Raster by Extent`, clip a portion of the CORINE dataset that matches the extent of the clipped Stirling SRTM layer. Hint: ask QGIS to use the extents of the SRTM clip for the CORINE clip.
7. Again, reproject this clipped CORINE dataset using both `Nearest Neighbor` and `Cubic(4x4 kernel)`, as you did for the terrain.

8.Now use the `Raster Layer Unique Values Report` tool from the `Processing` panel to count the unique pixel values in the original CORINE clip and the two reprojections. What happened when you used the bicubic resampling?

Return to the Symbology window, and expand the option Min / Max Value Settings. What is the default option selected? Experiment with the other options (with different percent clips and standard deviations) and see how the Symbology changes. Remember to Apply the changes every time you change any options. Still, on the Symbology window, change the Classification options (drop-down menu to the left, under the classes) from Continuous to Equal Interval. To the right of it, select the number of classes as 3. Then Apply and evaluate. Without changing your number of classes, change the Interpolation option above the classes from Linear to Discrete. Apply and evaluate. Take some time to play with these visualisation options to analyse the distribution of different Bioclimatic variables in the UK, and think about what they mean for plants, (non-human) animals and in terms of social and economic factors. Finally, let's work on the Symbology for the SRTM data. On the Symbology window, choose the Singleband Pseudocolor option, and then on the Color Ramp option, select Create New Color Ramp. On the small options window that comes up, select Catalog:cpt-city. Once the catalogue window opens, go to the Topography list and select the cd-a palette. Then classify your elevation values using the User Defined Min /Max option, and type 0 as Min and 900 as Maximum. Apply and visualise. Then, before closing the Properties window, go to the Transparency table and drag the slider at the top to around 60%.

Right-click on the SRTM layer name and select Duplicate layer (this step is useful to create a copy of a layer and retain symbology. Remember, this is a temporary layer). Now, on the Symbology of the copied layer, change the render type to Hillshade. Apply and close the window. Make sure the copied layer is immediately under the original layer, and alternate each layer between on and off.

Stop and think:

What does the hillshade render style does? You may zoom in on an area of interest and examine how the topographical data appear now. Guided Exercise 3 - Mathematical and boolean operations using the Raster CalculatorLinks to an external site. Go to the menu Raster > Raster Calculator…. This tool allows you to apply several mathematical functions to raster layer values and even to do calculations among values.

First, find the Bio1 layer of the Bioclim raster. From the metadata, we know it is the first band, so double-click UK_bioclim@1 to add it to the expression area. In this context @1, @2, etc. indicate the respective raster band number.

As we saw above, the Bio1 layer corresponds to Annual Average Temperature, with a unit of degrees Celsius multiplied by ten. Let's convert it to regular degree Celsius units. On the expression area, after UK_bioclim@1, use the operator / then write 10. Then, in the top right corner, select a folder to save your file and name it appropriately (e.g. UK_bioclim_C.tif). Then click OK.

Style your new layer in the same style used for the original Bio1 layer, using the full Min / Max range, and observe the laughable maximum average temperature values for the UK (sorry, couldn't resist). Can you spot the London heat island effect (urban heat island - UHILinks to an external site.)? Can you recognise other heat islands?

Let us say we would like to relocate to the hottest (in average) locations in the UK. Go back to the raster calculator, and enter the expression UK_bioclim@1 > 100 (or use your new layer and select areas > 10). Save the result as UKs_last_hope.tif. (or more likely future_tropical_UK.tif, considering the heat waves during the last summer 2022).

Go to the Symbology layer of your newly created layer and select Paletted / Unique Values, then Classify. Then, remove the 0 values from the legend using the - (minus) button. Click on the colour box for the 1 values and select a strong red colour. Finally, go to the Transparency tab and drag the slider to around 50%. Then Apply and close the window. Position your temperature range layer just above the terrain + hillshade layers, and make sure only these three layers are visible. Pretty, isn't it?

Optional: add the UK counties layer from hereLinks to an external site.. Symbolise using "outline", no fill. Make sure the counties layer appears on the top. Can you identify the counties with the highest average temperature?

For example, let's say we want to get the mean average annual temperature and visualize the temperature ranges for our "last hope" region only. Let us go to the Raster Calculator, and use the expression (UK_Biolcim@1 / 10) * UKs_last_hope@1. Name the resulting layer and save it, then style the result using a Pseudocolor colour ramp. However, this time manually specify the Min value as 10. It may take a few minutes, so be patient.

Stop and think:

What happened to the temperature values when we applied the expression above? Use the information tool to probe a few values of the new raster to help you think about it. Return to the Properties of the new masked layer, and select the Transparency tab. On the table named Transparent pixel list, click on the + (plus) button to add a new line. Fill the line with the values from = 0, to = 0 and Percent Transparent = 100. Apply and evaluate the result. Stop and think: did you actually change the values of the raster layer when you set the transparent pixels?

Vectors can also be used as masks. Go to Raster > Extraction > Clip Raster by Mask Layer…. Select your Celsius converted average temperature raster as input layer, the gadm36_GBR_0 as your mask layer, and check the keep the resolution of the input raster box. Run the algorithm (it will take a while) then Close the window when finished. Set the transparency to 60%. Zoom to the original UK_SRTM layer extent and compare the results before and after the masking.

Now click on the menu Processing > Toolbox…. Welcome to the QGIS toolbox! You will find many additional functions here to process vectors and rasters, as well as functions from external software that can be accessed through QGIS. On the search bar at the top, search for Raster layer zonal statistics. Double-click on the tool with the name that comes up from the search. As Input layer, select the degree Celsius temperature layer you created, and as Zones layer, select the "last hope" layer (the one with the 0/1 values, not the masked temperature layer). Run (it will take a while) and then Close.

Open the attribute table of the new "statistics" layer and see what the number represents. Independent Exercise Is there any difference in mean values of average annual temperature between urban areas and forests in the UK? Make a map (including legend and scale) showing your results visually.

Hint 1: To ensure raster operations work properly, all rasters must be in the same projection.

Hint 2: You can also use AND and OR to create compound expressions (such as raster@1 > 20 OR raster@2 < 50) on the raster calculator (just make sure you type them in all caps).

Optional: Open a new project. Add the online Mapzen Global Elevation layer by going to the QGIS Browser and choosing XYZ Tiles > Mapzen Global Elevation, and then download the plate tectonic data from HERE Download HERE.

Image showing where to load the Mapzen layer on the QGIS Browser

Open the properties of the DEM and select the Hillshade symbology, with a Z factor of 10. Add the earthquake and country data from Week 2. Now see if there are any clear relationships between high elevations, plate boundaries and earthquake locations.

Part IV

Week 4 - Making maps

This week is about visualising spatial data. For as long as Geography existed, maps have been a key way of communicating spatial information to other people. Most rules of visual design also apply to map making, but there are some additional constraints: maps need to have a proper scale, a coordinate grid (or graticule), and often a north arrow. Maps, like statistical plots, must also have a comprehensive legend, associating all visual variables present on the map to an interpretation key. Maps also often make use of textual labels to help give spatial context to the displayed information.

ILOs covered

2. Obtain and assess the quality of spatial data from online and offline sources and produce new spatial data using computer and field methods;
3. Create map visualisations that adhere to cartographic principles and can be easily and unambiguously interpreted by the non-specialist public;

What will you learn

For every week, we will list the main theoretical and practical learning goals. Use these as a 'checklist' to gauge your learning for each week. If you don't feel confident you have learned any specific topic, then revisit the week's material!

Theoretical knowledge for Week 4:

Practical knowledge:

?@sec-carto1

?@sec-carto2

Part V

Week 5 - Spatial analysis I

This week, we will

ILOs covered

2. Obtain and assess the quality of spatial data from online and offline sources and produce new spatial data using computer and field methods;
3. Plan and execute GIS analytical steps to solve spatial problems successfully;

What will you learn

For every week, we will list the main theoretical and practical learning goals. Use these as a 'checklist' to gauge your learning for each week. If you don't feel confident you have learned any specific topic, then revisit the week's material!

Theoretical knowledge for Week 1:

Practical knowledge:

?@sec-spatan1

?@sec-spatan2

7 Lab 9: Geoprocessing tools

7.1 Guided Exercise: assessing children access to schools

You are a GIS analyst at the Stirling council, and you have been asked to perform an analysis regarding children access to schools. Your manager has indicated which data sources you should use, and they want you to answer **how many children within the relevant age range live within less than 1km, between 1km and 2km, and further than 2km of *any* primary (4 to 10 years old) and secondary (11 to 17 years old) school. You have been instructed to use the census data at the most detailed level (Output Area, OA), and to limit your analysis to the OAs and schools that are within Stirling Council Area limits**. You should report the results as a table.

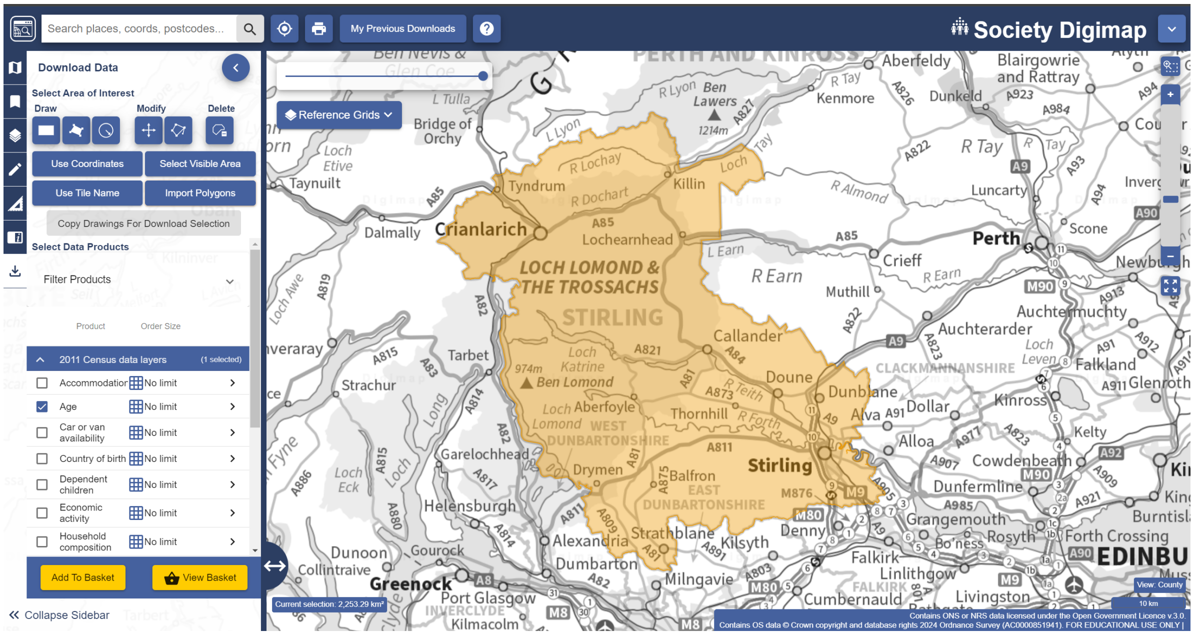
7.1.1 Obtaining the data

(189) The first piece of data needed for this exercise is of course the limits of the Stirling Council Area. The official administrative boundary data for the UK is contained within the Ordnance Survey Boundary Line dataset. It can be downloaded from the Ordnance Survey Open Data Hub or from Digimap, but the files cover all of the UK, and thus are a very large download. To save precious lab time, we have already isolated the Stirling Council boundary as a separate file that you can download here. You are welcome!

You can then download the two other datasets needed:

(190) **Scottish School Roll and Locations**: this dataset is provided by the Scottish Government, and contains information on all Scottish schools. Download the '*Scottish School Roll and Locations ZIP*' (NOT the one with 'Table' on the name) from here.

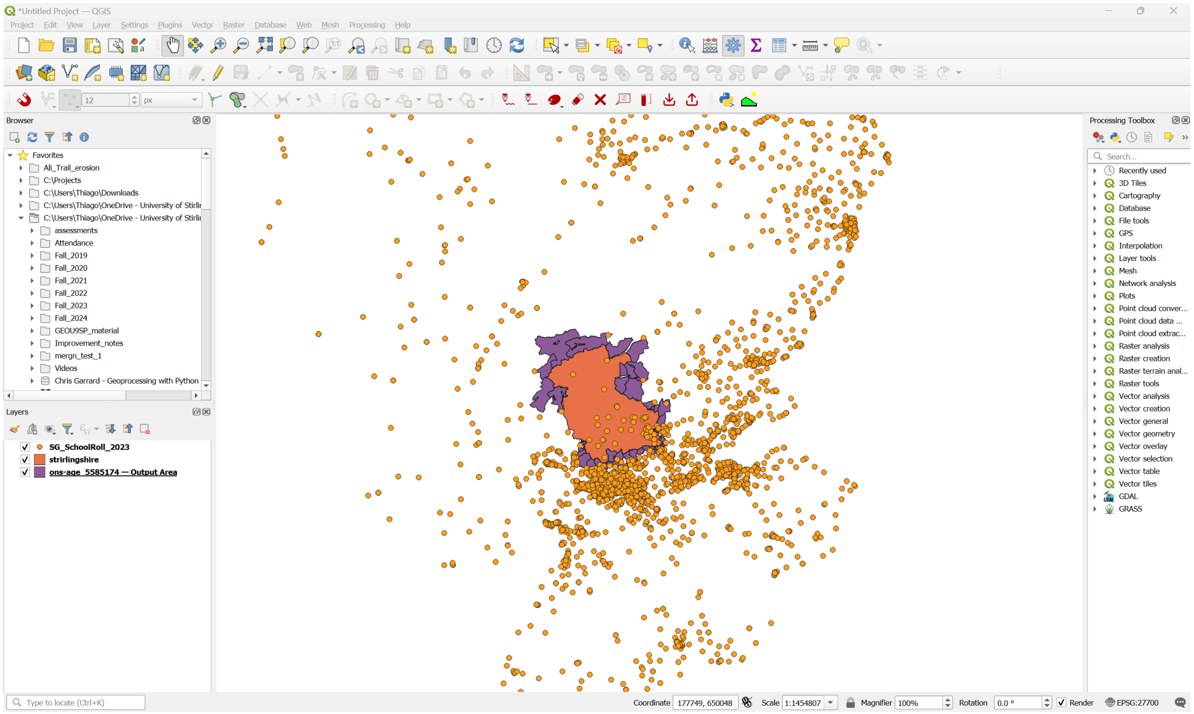
(191) **Census data on population age**: the Digimap platform offers access to the 2011 Census variables in spatial format. To obtain it, go to the *Society* tab, and then select the 'Download' option on the leftmost toolbar (a down arrow going into a 'box'). That will bring up the familiar Digimap download interface. This time we will use our Stirling Council shapefile to set the search area. For that, click on the `Import Polygons` button, and then on `Choose File`. Then choose the *zipped* Stirling dataset you have downloaded and click on import. Finally, under the `2011 Census data layers` tab, select the `Age` dataset. Your window should look like this:



> **Warning**
>
> The data for the 2022 census have been officially published by National Records of Scotland, but are not available on Digimap yet. Since we are only practising our GIS knowledge, it is OK to use mismatched data (schools from 2024 vs. age data from 2011). In a real application scenario, it would be very important to make sure the temporal coverage of the datasets matches as well as possible.

(192) Add your data to the basket and finish the order, then download the data as you have done in previous labs.

(193) Now create a folder structure for your project and extract and organise the contents of the three datasets. Then create a new QGIS project and import the three datasets. The census geopackage will contain layers for different levels of aggregation, remember you want the Output Area polygons. If everything worked correctly, you should have something like this:



! Stop and Think

1. What are the data models and file formats of each dataset?
2. What are the attributes contained in each dataset?
3. What is the CRS of each dataset? Are they all the same?

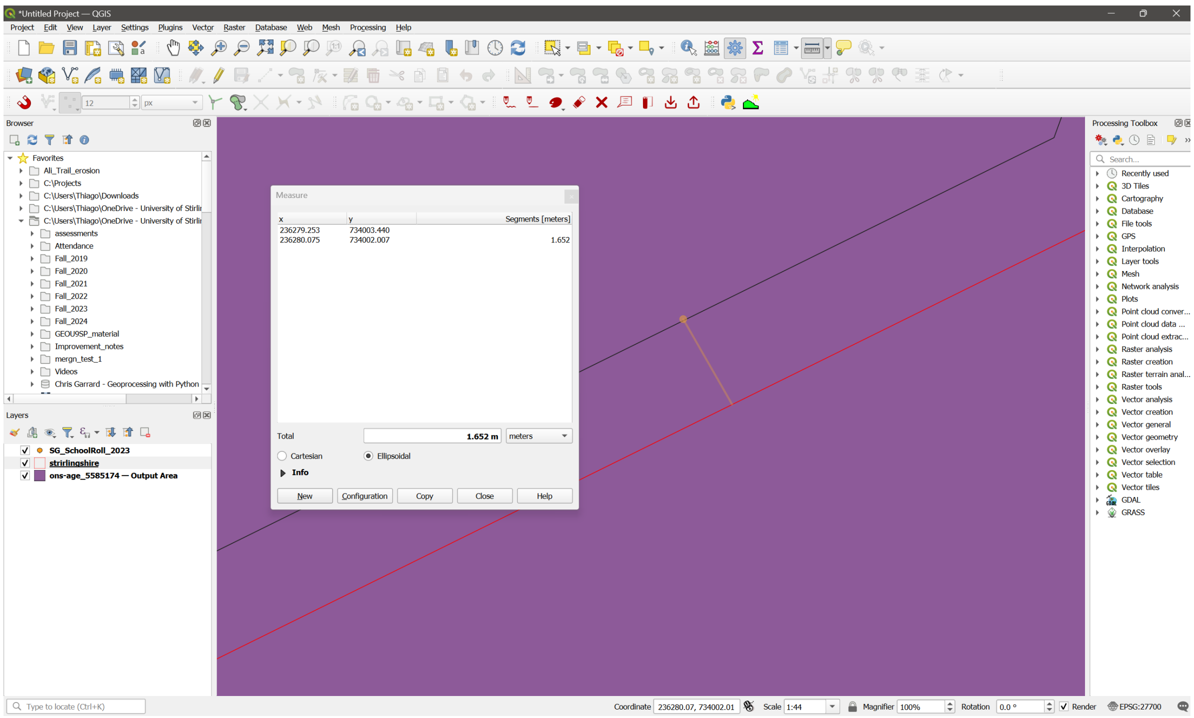
Click for answer

1. They all use the vector data model. The schools are vector points, the other are vector polygons.
2. The attributes of each dataset can be seen by either inspecting the *attribute table* of each dataset or by looking at the layer `Properties > Fields` tab.
3. All three datasets are in OS British Standard Grid (EPSG 27700). You can find this out by either going to the `Properties > Information` tab or by right-clicking on the layer name then on `Layer CRS`.

7.1.2 Preparing the data

You may have noticed that your Digimap census data download includes many OA polygons that are outside of Stirling Council. This is because of a (sadly) very common problem with spatial data from different sources - the census output areas (OAs) do not perfectly line up with the administrative boundaries of the council. So when the Digimap interface tries to determine which OAs overlap with Stirling, we get more data than we need.

(194) Set the symbology of the Stirling boundaries to a red outline without fill. Then make sure this layer is on top of the census layer. Then zoom in very close to the Stirling boundary. You will see they are offset by about 1.5m:



We thus need to further subset our data to match the actual extent of the Stirling coucil.

! Stop and Think

What are the options of GIS operations that you have to subset the datasets?

Click for answer

You could use `Clip` or `Intersection` if you wanted to change ('cut') the geometries of the target layer, or use `Select by Location` if you wanted to fidn the overlapping polygins without changing geometries.

After consulting with your manager, you both decide that clipping or intersecting the data would not be a good option as it would change the geometry of the OAs, and there are official limits. So you decide to use `Select by Location`:

(195) Go to `Vector > Research Tools > Select by Location...` or click on the button.

(196) Select all polygons from the census layer that `are within` the Stirling bounds layer. Check the selection results.

(197) Change your criteria to include both `intersect` and `are within` the Stirling bounds layer. Check the selection results.

! Stop and Think

Why both selection attempts give you wrong results? How would you fix it?

Click for answer

Because of the mismatch between the boundaries of the OAs and the Stirlingshire boundaries. So inner OAs that are the edges of the council area are not considered fully `within` (they extend a sliver outwards from the actual bounds), and outer OAs also at the edges will be considered to `intersect` Stirling since they extend a sliver within the bounds. Think a bit on how to fix it, then continue below for a solution.

You could manually deselect the polygons that are not part of Stirling council, but then you have an idea: what if you buffer the Stirlingshire boundaries to expand them a tiny bit, and then try selecting the polygons `within` the buffer. That way OAs inside the council area that slightly extend outwards will fall within the buffer, while OAs outside the council area that slightly extend inward will still not be fully within it.

(198) Go to `Vector > Geoprocessing > Buffer` and create a 20m buffer around the Stirling layer. You can just leave it as a temporary layer, as we will only use it for our spatial selection.

(199) Now repeat the `Select by location` operation for OA polygons `within` the buffered area. Bingo!

(200) Save the selection as a new layer by right clicking on the census layer name and selecting `Export > Save Selected Features As...` Once this new census layer is added to your project, remove the original one to avoid confusion.

7.1.3 Answering the question

Now we can answer the actual study question. As a reminder, it is **how many children in the relevant age range live within less than 1km, between 1km and 2km, and further than 2km of *any* primary (4 to 10 years old) and secondary (11 to 17 years old) school?**. You will need to use some of the Geoprocessing tools you just learned to answer that, as well as tools from previous weeks.

> **!** Stop and Think
>
> Before you continue, try to think about the solution yourself. Look at the attributes of each layer, and then try to sketch a step by step plan to achieve the results you need. Then continue.

(201) You will first need to add up the number of children at each 1yr age bracket to compute the totals for the primary and secondary school age ranges (4-10 yo and 11-17 yo). For that, you can use the `Field Calculator`, as in previous labs. Just add up the numbers for columns `age_4` to `age_10` into a new `age_4-10` column, and then add columns `age_11` to `age_17` into a new `age_11-17` column.

(202) Then you will need to select only the schools within the Stirling council area. You can use `Select by location` as above, using the Stirling bounds, or by using the `LAname` field in the attribute table of the School Roll layer. Both methods should give you a selection of 48 schools. Create a new layer containing just these schools, then remove the original school layer from the project.

(203) The next step is to calculate the area covered by each distance bracket (1km, 2km) from the schools. For that, you should create a `Buffer` for the 1km distance, then a second `Buffer` for the 2km distance. Make sure you select the `Dissolve result` option when creating the buffers, since many locations will be within less than the specified distance from multiple schools, and we don't want to repeat them.

(204) Then compute the `Difference` between the 2km and 1km buffers to create "rings" covering the range of 1km to 2km. This is important so we don't "double count" the 1km bracket - see below.

Now you have to deal with the biggest limitation of your analysis - the OAs are still very big, and many will span areas both within and outside the 1k and 2km distances. You discuss the problem with your manager again, and you decide that the best approach is to *weight* the children count of each OA by the respective *area* of the OA that is *within* each distance bracket. So if an OA has 15 children aged 11-17, and 40% of the OA area is within 1km of a school, then you should multiply 15 by 0.4 to obtain the estimated number of children within 1km.

(205) The first step is to calculate the total area of each OA. You can do that by using the `Field Calculator` tool and creating a new field called `oa_area` that takes the `$area` parameter under the `Geometry` calculator operations. Remember to make this new field a **Decimal number** with two decimal places.

! Stop and Think

What units were used for the area calculation? Does it matter?

Click for answer

Unless you have set it differently in your *Project properties*, areas are calculated in squared metres by default. This is not an ideal unit for such large areas (km^2 would have been better), but since we will only need to calculate the percentage of this total area within each distance bracket, it shouldn't matter much.

Now that you have the total areas for each OA, it is time to find out the portions within each distance bracket. For that, we can either `Clip` or get the `Intersection` of the buffer layers with the census layer.

! Stop and Think

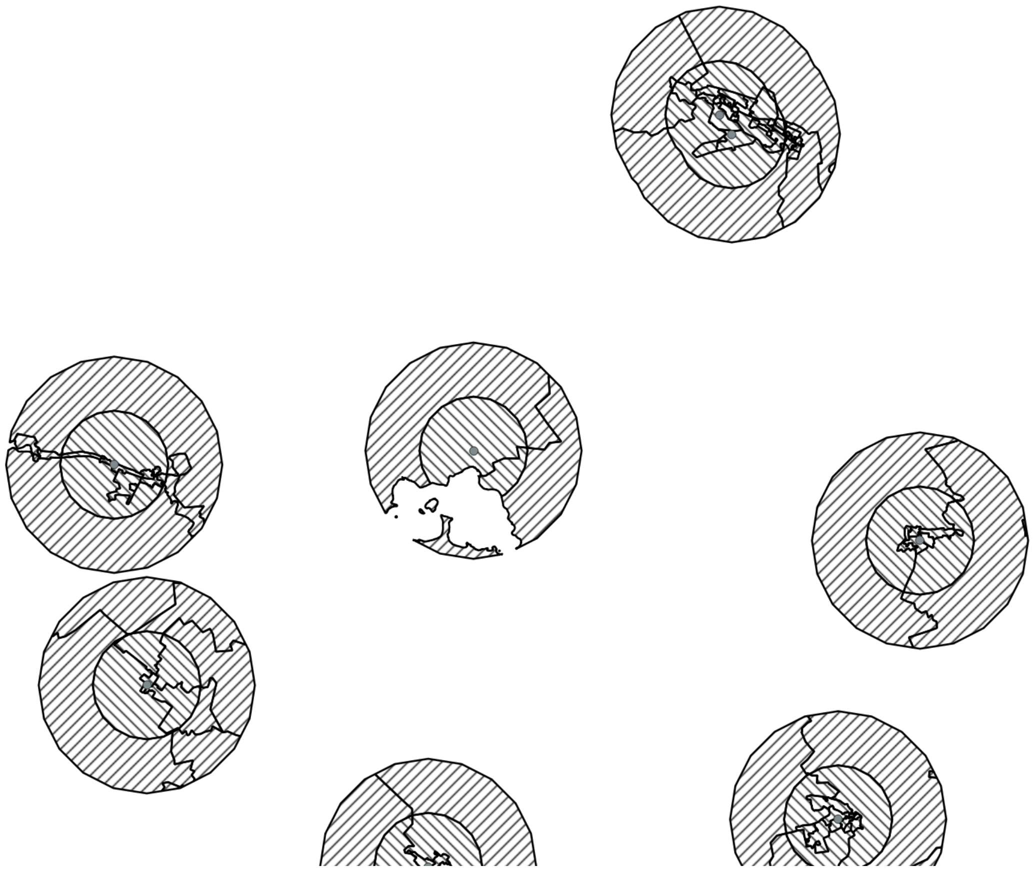
What will change in your results if you use `Intersection` instead of `Clip`?

Click for answer

`Intersection` will combine the attributes of both layers, while `CLip` will only preserve the attributes of the target layer.

(206) Since the buffers have been *dissolved*, they can't be linked to individual schools anymore - whenever the buffers of two schools overlap their attribute tables will just keep the name of one of the schools when they are dissolved. So no need to preserve the buffer attributes, and we can use `Clip`. Create two new layers, the first being the clipping of the census layer by the 1km buffer layer, and the second being the clipping of the census layer by the 1km-2km 'donuts' layer. We use it instead of the 2km buffer otherwise when we calculate the percentage of area within 2km we will be double counting the area within 1km - which is already covered by the 1km buffer.

You should end up with something like the figure below. To make sure the layers don't overlap, style them using hatching in different directions. If you see a cross-hatch, then they are overlapping, and it means you used the full 2km buffer instead of the 1km-2km 'donut':



(207) Now we can calculate the area percentages. First, re-calculate the polygon areas for the clipped layers, using the `Field Calculator`. Create a new attribute called `clip_area`, and assign the `$area` operator to it again. Remember to set it as a decimal number.

(208) Still using the `Field Calculator`, you can now calculate the percent area by creating a new attribute called `perc_area`, which will be `clip_area / oa_area`. It should also be a decimal number.

(209) Finally, calculate the weighted number of children by creating new fields. First calculate `perc_4-10` as `age_4-10 * perc_area`, and then `perc_11-17` as `age_11-17 * perc_area`. These fields should be integers, unless you are sending partial children to school.

Great, you are almost there! The last step is to aggregate all the data at the council level, and then also determine the number of children beyond 2km:

(210) Use the `Statistical Summary` tool to calculate the respective sums of `perc_4-10` and `perc_11-17`. These are the final answers for **how many children in the relevant age range live within less than 1km, between 1km and 2km?**.

(211) To get the number of children beyond 2km, use the `Statistical Summary` tool to calculate the total number of children within each age bracket (`age_4-10` and `age_11-17`) in the council, and then subtract the number of children within 1km and between 1km and 2km from this total.

Your final results should be:

	1km	2km	>2km
Primary	6451	6656	89
Secondary	7572	7880	91

Part VI

Week 6 - Spatial analysis II

This week, we will

ILOs covered

2. Obtain and assess the quality of spatial data from online and offline sources and produce new spatial data using computer and field methods;
3. Plan and execute GIS analytical steps to solve spatial problems successfully;

What will you learn

For every week, we will list the main theoretical and practical learning goals. Use these as a 'checklist' to gauge your learning for each week. If you don't feel confident you have learned any specific topic, then revisit the week's material!

Theoretical knowledge for Week 1:

Practical knowledge:

?@sec-spatan1

?@sec-spatan2

Part VII

Week 8 - Spatial analysis III

This week, we will

ILOs covered

2. Obtain and assess the quality of spatial data from online and offline sources and produce new spatial data using computer and field methods;
3. Plan and execute GIS analytical steps to solve spatial problems successfully;

What will you learn

For every week, we will list the main theoretical and practical learning goals. Use these as a 'checklist' to gauge your learning for each week. If you don't feel confident you have learned any specific topic, then revisit the week's material!

Theoretical knowledge for Week 1:

Practical knowledge:

?@sec-spatan1

?@sec-spatan2

Part VIII

Week 10 - Obtaining spatial data in the field

This week, we will

ILOs covered

2. Obtain and assess the quality of spatial data from online and offline sources and produce new spatial data using computer and field methods;
3. Plan and execute GIS analytical steps to solve spatial problems successfully;

What will you learn

For every week, we will list the main theoretical and practical learning goals. Use these as a 'checklist' to gauge your learning for each week. If you don't feel confident you have learned any specific topic, then revisit the week's material!

Theoretical knowledge for Week 1:

Practical knowledge:

?@sec-spatan1

?@sec-spatan2