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

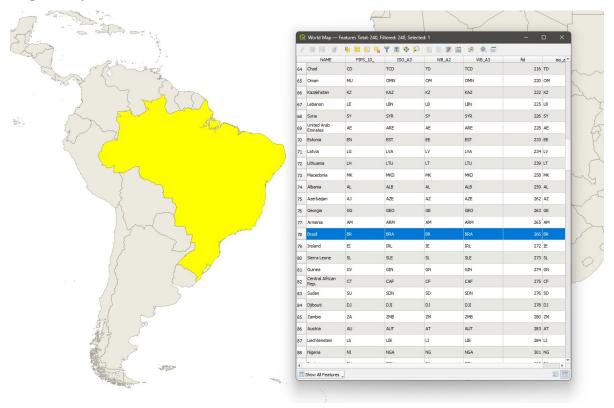
Chapter 4

- 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

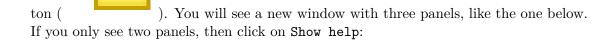
- a) Categorized is for *categorical*, non-numeric variables (i.e. names, classes, etc.). Graduated is for *continuous* variables (i.e. quantities, measurements).
- b) No, because China, India and to a lesser extent the US have much numbers than the rest, which biases the breakpoints. We will fix it 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.

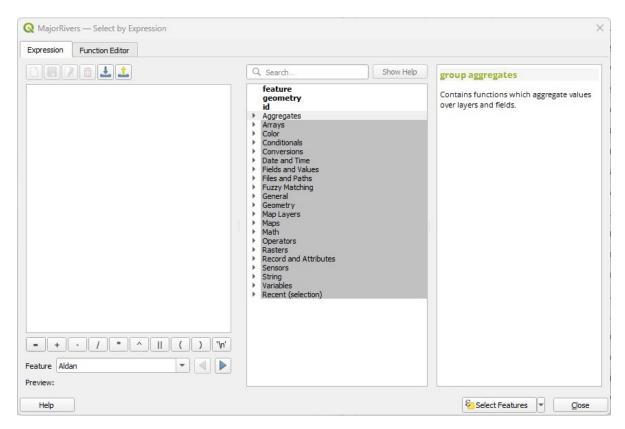
When you are mapping your own data, make sure you explore the different methods for calculating breakpoints, and also play with the number of classes. You can see the ful explanation of each Mode on the QGIS Documentation.

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 referred 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 but-





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?

l () to click on the Amazon River

Hint: try to use the manual selection tool (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 (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 NOToperator?

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

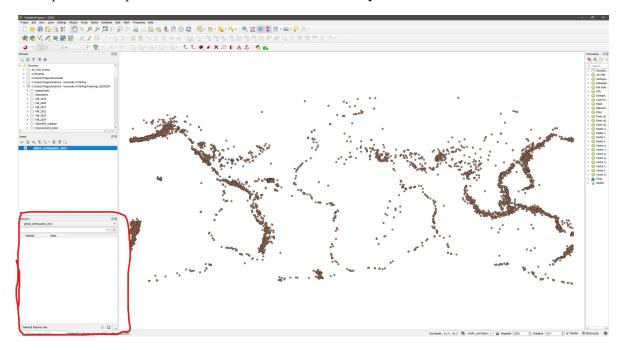
The LIKE operator is case sensitive, while ILIKE is case insensitive.

(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 (2) 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?

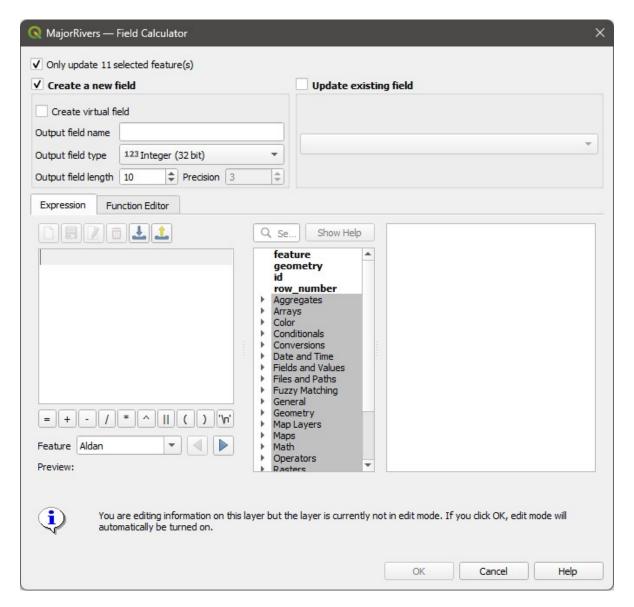
Click for answer

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 (II), 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 keep 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.

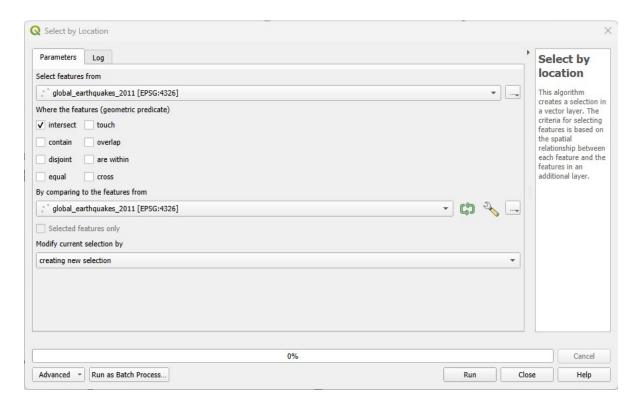
4.2 Guided Exercise 1 - Spatial Queries

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

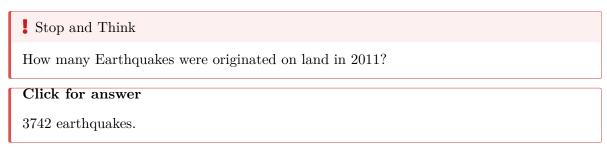
(103) If you have zipped your project from the last session, just unzip it and get to work. If not, then download the data again, and redo your project organization. As a reminder, the files to be loaded are: global_earthquakes_2011.gpkg, MajorRivers.shp, and ne_50m_admin_0_countries.shp.

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.



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 the PA_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.