**Week 2: Databases**

This week, you learn quite a lot of different tools and techniques. You will practice how to insert data into QGIS from Excel. You will learn how to plot data that have coordinates on a map and how to import other datasets. You will learn how to query your data using the information. You will also learn how to join different tables together. Finally, you will practice how to make spatial queries that are specific to GIS.

The data you will work on is blue tit and great tit foraging observations in the Cambridge botanical garden. The birds were captured in their nest boxes and were marked with colour rings. These allowed the researchers to follow the birds and record their foraging behaviours. These observations included the GPS location for each foraging location, the plant species the birds were observed foraging on, the part of the plant (bud, bark etc.) and other information. The dataset also contains information on the individual bird (species, sex, date and size of the clutches) and the plants in the botanical garden.

**A. Download the dataset for this week’s practical**

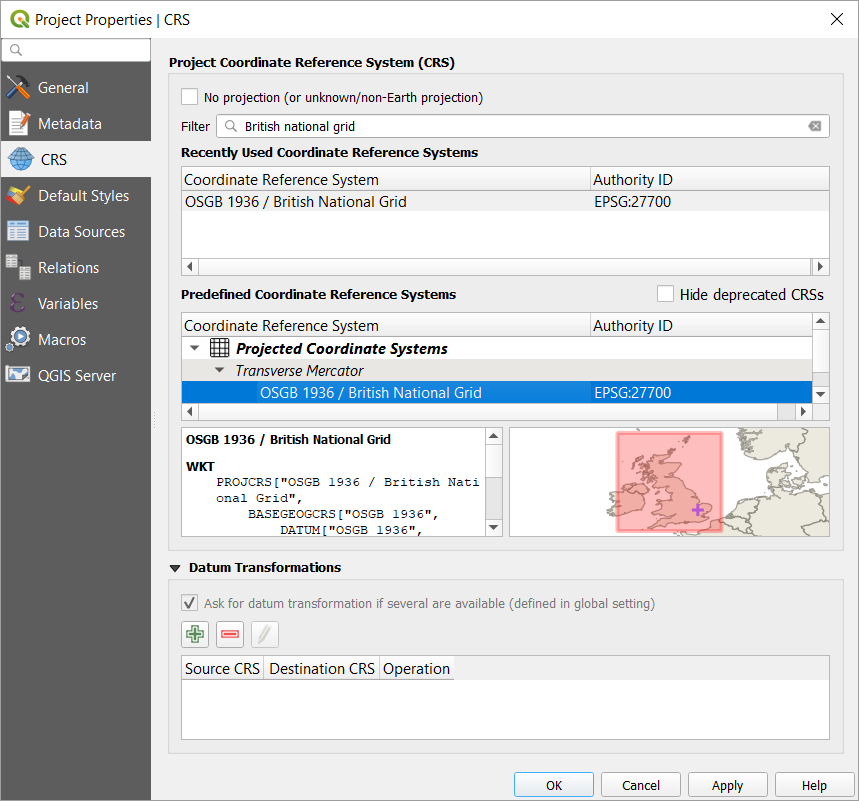
1. On Canvas, from the week 02 Practical page download the archive zip folder (don’t OPEN it, SAVE it!). Save the zip file in an easy location to find. I strongly recommend you to use a memory stick or external hard drive and work from there rather than your personal space on the university network.
2. Extract the archive: right-click on it and select “Extract here”.

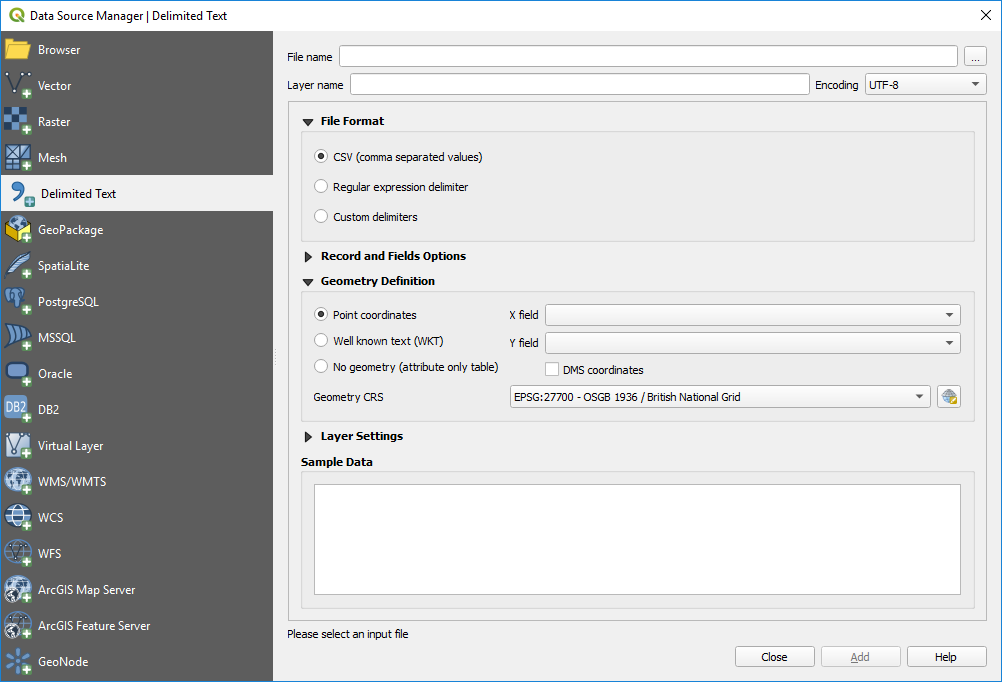
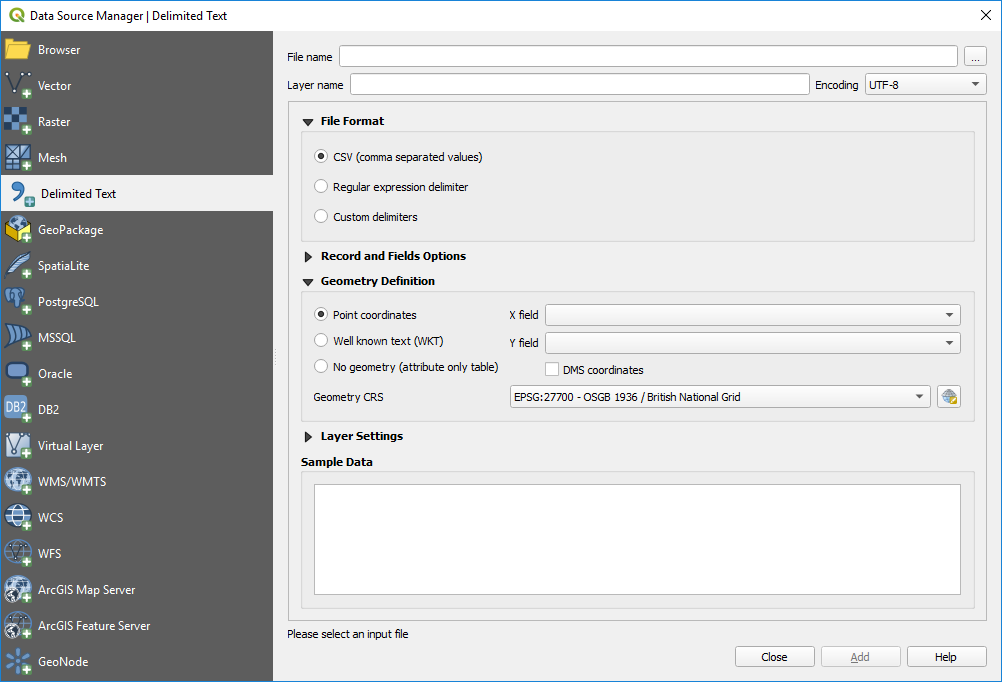
**B. Explore the excel spreadsheet**

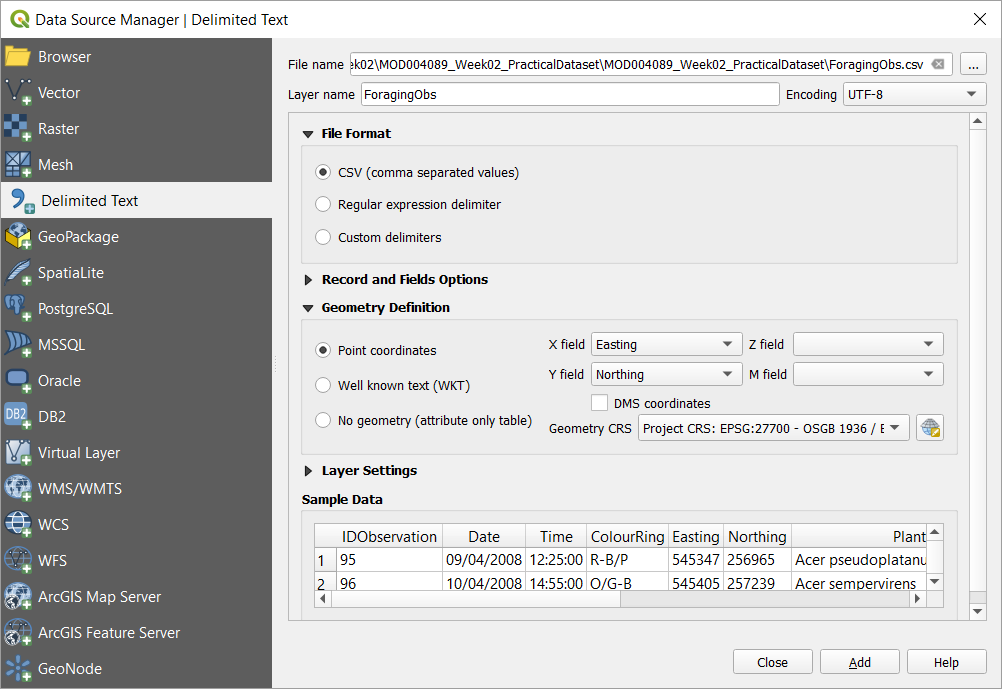
1. Open the BirdsBotanicalGarden2008.xlsx with Excel, have a look at the three worksheets it contains. How do these data relate to each other? Draw on a piece of paper the relational database model for this dataset including the name of the possible joining fields (columns with matching content).
2. Which dataset contains geographical information and can be mapped (coordinates)? QGIS is not very good in handling Excel spreadsheets, so in order map this dataset, it has to be exported as a different file format that is easy to read. A good standard for spreadsheet is the comma delimited file (csv). Select the worksheet that contain the spatial information and save it as a csv file (File > Save as, specify CSV as type of file; use the same name as the worksheet).

**C. Importing the data into QGIS**

1. Open QGIS with an empty project. You are then going to define the projection system for your project (we will talk about that next week, but it is to make sure that our two-dimensional map represents three dimensional reality accurately). Because our data cover an area in the UK, we are going to use a British coordinate reference system (CRS). Open the Project > Properties and go to the CRS tab. In the Filter box, type British National Grid and in the Predefined Coordinate Reference Systems list, select the one that has an EPSG code of 27700. Click on OK. That EPSG code should now appear on the bottom right corner of QGIS.



1. Use one of the 3 methods you tried last week to add the CambridgeBotGarden.jpg layer into QGIS. What type of spatial data is that?
2. If you could answer the B2 question, you know that the dataset containing spatial information is the ForagingObs as it has coordinates columns (Easting and Northing). So you should have saved it as a csv file that you can now add as a layer. Open the Data Source Manager () and click on the Delimited Text tab (). Next to the file name box, click on the  button to browse and select your CSV file. Under the Geometry Definition, The X and Y field where the coordinates are stored should be automatically recognized (if not, X is Easting and Y is Northing). Next to the Geometry CRS, click on the drop-down menu and select Project CRS: EPSG:27700… to indicate that these coordinates are using the British National Grid system. The Sample Data table will give you a preview of the data.

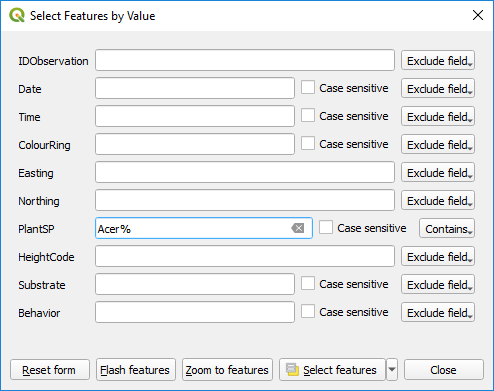


Once you click on Add and then Close, you should get points overlapping your aerial image of the Cambridge botanical garden. You can change the Symbology of the point layer to use a more visible colour if necessary.

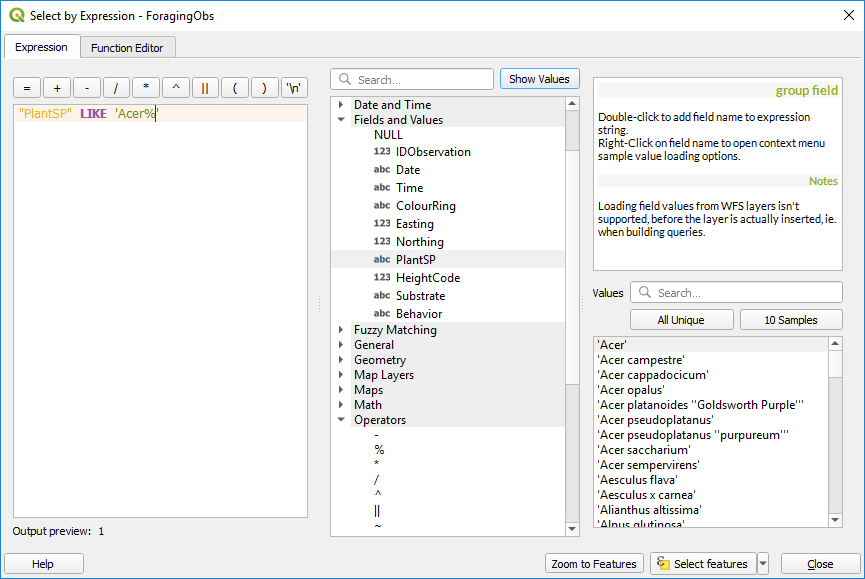
**D. Querying by attribute**

1. We can explore the foraging observation dataset and query the dataset to answer some questions about the attributes of our data. For example: How many foraging observations occurred on species from the genus Acer? To answer that, you will build a query: click on the Select features by Value button in the attribute toolbar ().

The window will list all the different column of your layer; in the PlantSP box, type ‘Acer%’. The “%” symbol acts like a wildcard; it means that all species starting with Acer will be selected. Click on the Select features button and you should see the features matching the query highlighted on the map view. Also on the bottom bar of the software, you should have a message telling you how many features are selected.

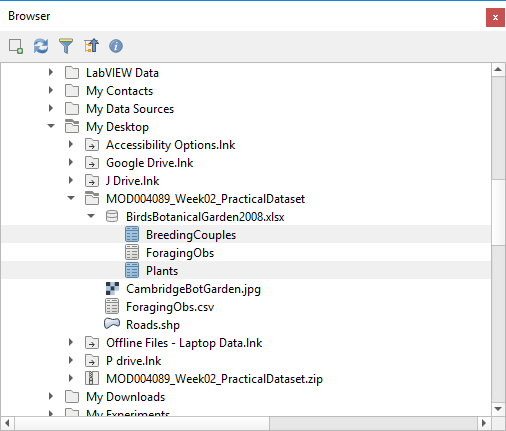
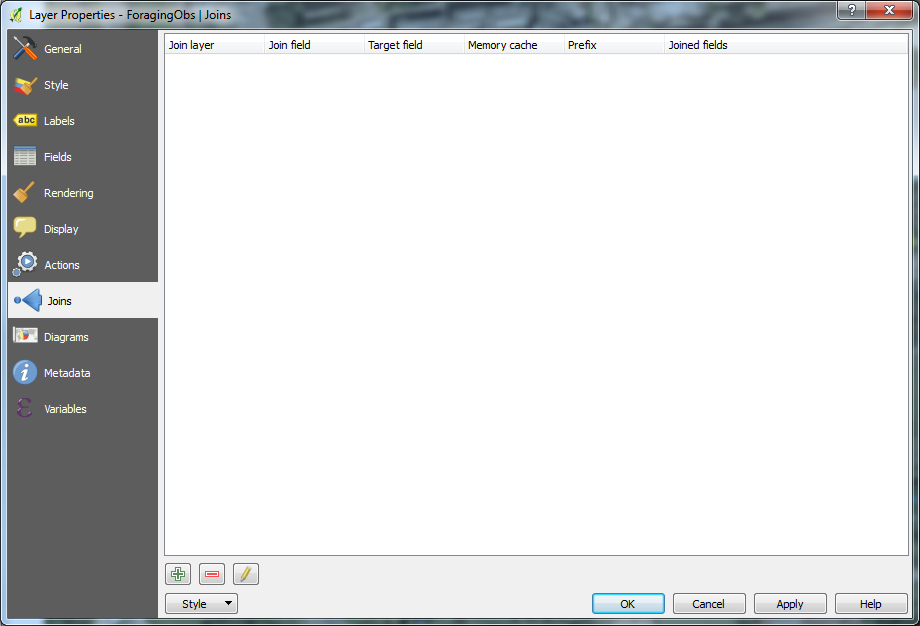


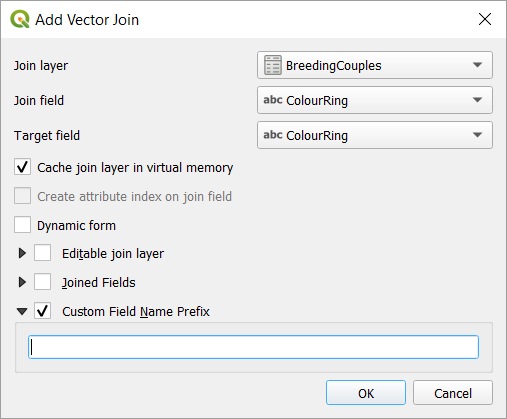
Next to the Select features by Value button, you can use the drop-down menu and use the Select by Expression function (). This tool is a very powerful expression builder. By expanding the Fields and Values group, you can double-click on the PlantSP field; then you can expand the Operators group to double-click on the “LIKE” operator; finally, go back to the Fields and Values group and click on the PlantSP and then on the All Unique button to double-click ‘Acer’. This should give you the following expression: “PlantSP” LIKE ‘Acer’. Add the “%” wild card after “Acer” to get the following expression that should select the same number of features as earlier.



1. Try to answer the following question: “How many foraging observations were done on flowers of species from the *Betula* genus”? You can use any of the previously described selection tools, but if you use the Select by Expression tool, you will have to use the AND operator to combine queries. Clear the query using the Deselect features from all layer button ().

**E. Joining tables**

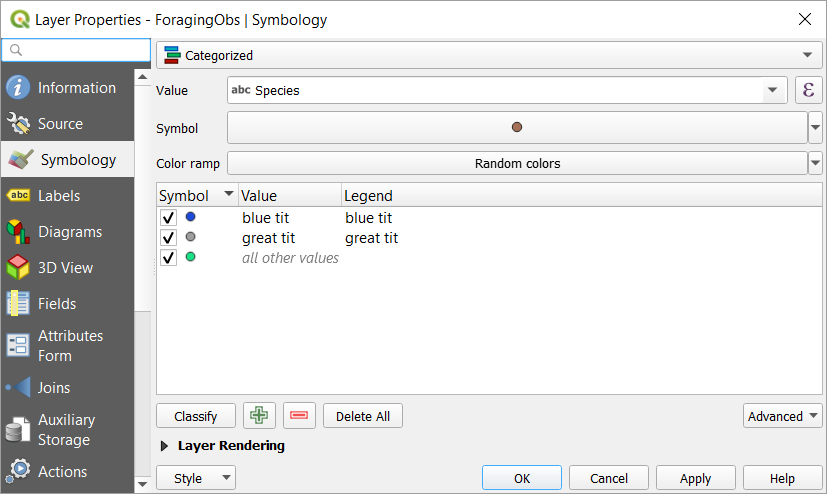
1. On part B.1, you realised that the foraging observations could be linked with the other datasets. Let us do that within QGIS. Spreadsheets can be added as *non-spatial* data (without coordinates); in the Browser panel, find and expand the BirdsBotanicalGarden2008.xlsx spreadsheet to list the three worksheets. Select both Plants and BreedingCouples by using the CTRL+click combination and click on the Add Selected Layers button ().
2. These tables do not contain any spatial information and therefore cannot be mapped. However, we can explore them (Open Attribute table) and link them to the ForagingObs. Open the ForagingObs layer properties window and go to the Joins tab. Click on the  to add a layer, select the BreedingCouples and specify the common fields to be used as Join and Target fields. Tick the Custom Field Name Prefix and empty the text in the box. Click on OK (twice).



Open the attribute table of the ForagingObs layer and you can see that the columns from the BreedingCouples have been added.

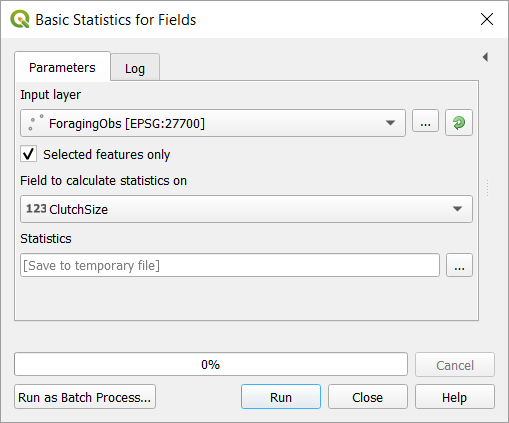
1. Build a query to answer the following question: “Where are located the great tits foraging observations?”

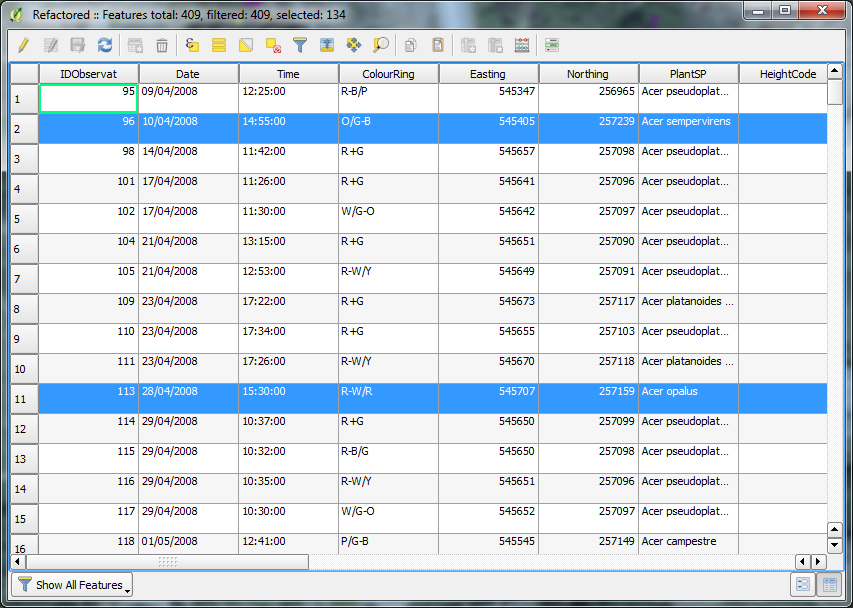
An alternative way to answer that question would be to alter the style of the layer and represent both tit species with a different colour. To do that, open the layer properties and in the Symbology tab, set the style to Categorized, select the column containing the bird species (BirdsBotanicalGarden 2008 BreedingCouples any\_Species) and click on the classify button (you can change the point colours and/or shape by double-clicking on the symbol).

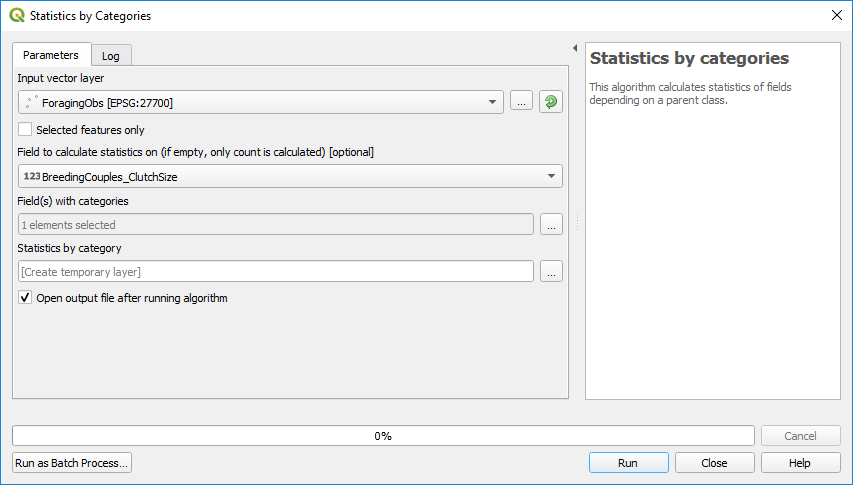


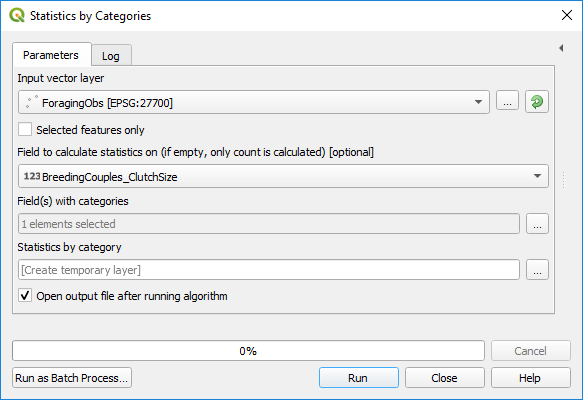
1. Build a query to answer the following question: “How many foraging observations of male blue tit on *Quercus robur*”?
2. To answer the following question “What are the mean clutch sizes of birds foraging on native plant species and birds foraging on non-native plant species?”, you need to first use the same procedure from point 2 to join the Plant dataset (this contains the information whether a given plant is native or not) to the ForagingObs layer. This will add the native information into your joined layer (check the attribute layer).

Then you need to select the foraging observations on native plants. Finally, you can use the Basics Statistics for Fields tool (Vector > Analysis) to calculate some stats on the clutch size field. Make sure you tick the Selected features only box. Write down the mean clutch size.

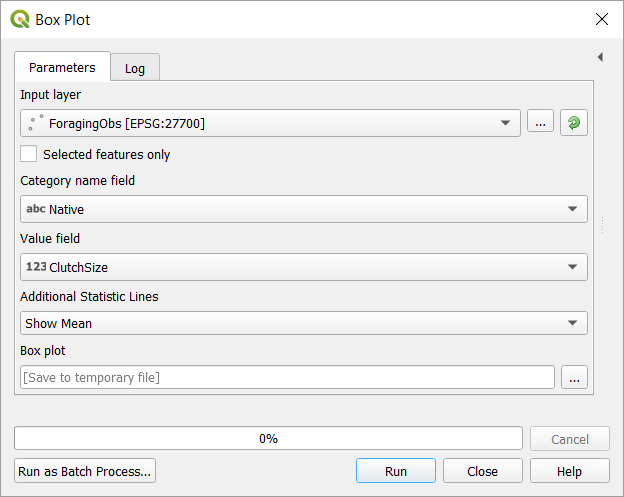


You can then invert the selection by clicking on the Invert Selection button () located in the drop-down menu where the Selection tools are. Once you changed the selection (you can check the attribute table if you are not sure), re-use the Basic statistics for numeric fields tool again to calculate the mean clutch size for birds foraging on non-native plant species.

Alternatively, you can use the Statistics by Categories tool: in the Processing Toolbox (you can enable it in Processing > Toolbox), expand the Vector analysis group and run the Statistics by Categories tool. Select the ForagingObs layer as Input vector layer; the BreedingCouples\_ClutchSize as the field to calculate the statistics on and under Field(s) with categories, click on the button to tick the Plants\_Native field. Click on Run. You will then have a new Statistics by category table in the Layers Panel. Open its attribute table to check the results.

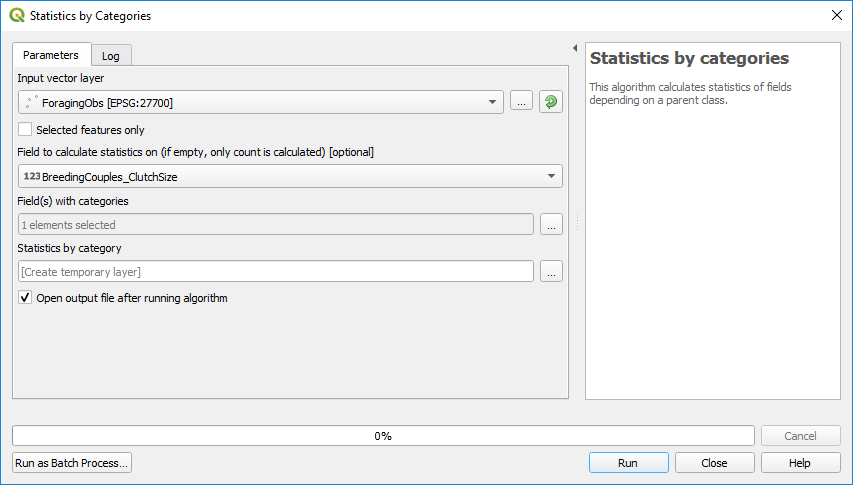


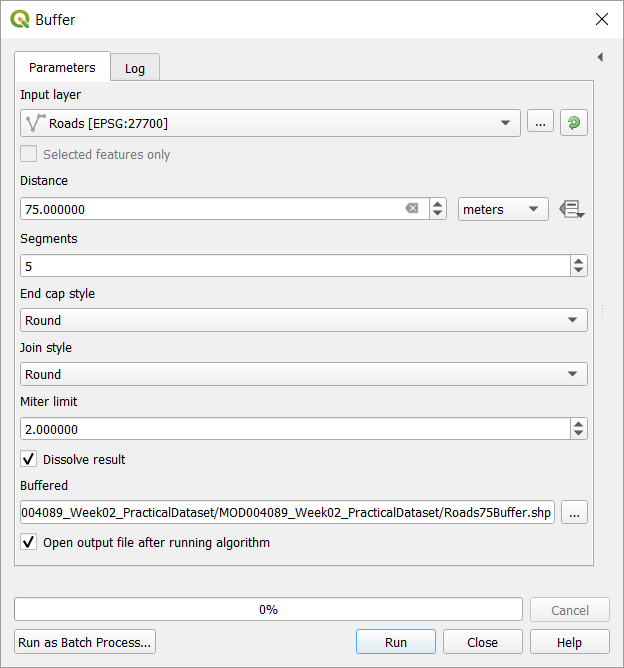
QGIS is also able to plot graphs. Open the Processing Toolbox (Processing > Toolbox), under Plots, run the Box plot function. Select your ForagingObs as Input layer, your Category name field will be Native and the Value field will be ClutchSize. Click on Run and in the Results Viewer, double-click on the Box plot to check it.



**F. Spatial queries**

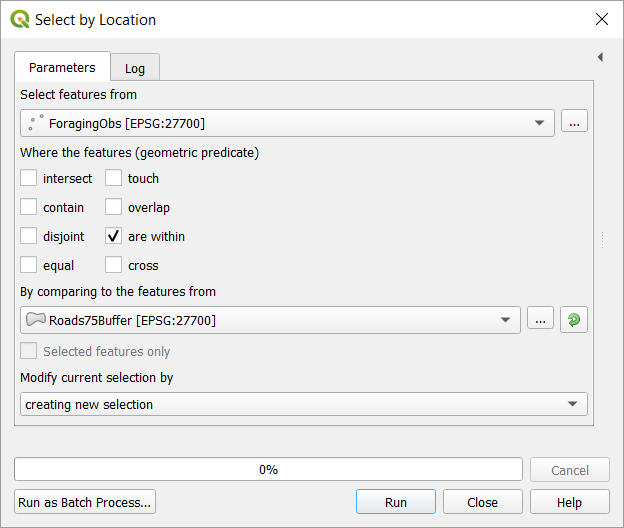
The queries you previously did are not specific to GIS; they were only queries based on the attribute tables and many software can perform them (a GIS software is probably not the best software to perform these and you could do all of them in Excel…). But one of the specificities of GIS is that you can run not only queries based on the information, but also queries based on the spatial distribution of your data.

1. Let us try to see if the proximity to the roads has an influence on the bird fitness. Add the roads vector shapefile.
2. Create a 75 metres buffer around the roads. Run the Buffer tool in the Vector > Geoprocessing Tools menu. Select the Roads layer as Input layer, put 75 meters in the Distance box, tick the Dissolve result box for a better result and under the Buffered box, click on the  button and select Save to file to create a new shapefile called Roads75mBuffer in your week 2 folder. Click on Run.



You created a new vector layer; how different is it from the original roads layer? Rearrange the order of the layers to see the points above the buffers (you can change the Symbology of the Buffer layer to change its Opacity for a nicer map).

1. Use Vector > Research Tools > Select by Location to select which foraging observation points are within the 75 metres buffer around the roads (select features from ForagingObs that are within the Buffered layer).



1. Use the result of this query and the Basic statistics for numeric fields tool used previously to answer the following question: “What is the mean clutch size for birds foraging within 75 metres from the roads?”