

Data processing and handling techniques for the purposes of efficient biodiversity data management and reporting

Exercise 4.2 - Cleaning data

Objective - Learn how to import data to QGIS and clean them for common mistakes

Document version: 2017-09-19
Document author: Marin Grgurev, PhD

Task 1: Import csv (comma separated values) data to QGIS.

Objective: Learn how to import csv files into QGIS

Steps to do:

1. Open QGIS.
 2. Click on Layer > Add Layer > Add Delimited Text Layer
 3. In the Create a Layer from a Delimited Text File dialog, click on Browse and specify the path to the text file located in data folder named `import_file.csv`. In the File format section, select Custom delimiters and check Tab. The Geometry definition section will be auto-populated if it finds a suitable X and Y coordinate fields. In our case they are *decimallongitude* and *decimallatitude*. Click OK.
 4. Next, a Coordinate Reference System Selector will ask you to select a coordinate reference system. Since this species dataset coordinates are in latitudes and longitudes (it was downloaded from GBIF), you should select WGS 84.
 5. You will now see that the data will be imported and displayed in the QGIS canvas.
 6. Now let’s select another file to import.
 7. Click again on Layer > Add Layer > Add Delimited Text Layer.
 8. In the Create a Layer from a Delimited Text File dialog, click on Browse and specify the path to the text file located in data folder named `import_file_dirty.csv`. In the File format section, select Custom delimiters and check Tab. The Geometry definition section will not be auto-populated as before so choose from the dropdown menu *decimallongitude* for *X field* and *decimallatitude* for *Y field*. Click OK.
 9. Do you see anything unusual? Check the entry 19, 28, 46 for example. Whats wrong with them?
 10. In next task you’ll learn to deal with such problematic entries and how to clean them.
-

Task 2: .

Objective: Learn how to clean the data before importing them to QGIS

Steps to do:

1. In this assignment you’re going to work without instructions. Your assignment is to correctly import the species data file in QGIS without any errors.
2. The file that needs to be cleansed is `data/cleaning_data_dirty.csv`
3. You can import the file in QGIS several times to check if you clean all the errors and that everything is OK.
4. Ask instructors to help you with functions you don’t know well in both Calc or QGIS.
5. The final result is to have 61 entries of 27 unique species correctly imported in QGIS without errors.