Data Management, analysis and mapping

**Module description**

This is a hands-on module which will cover best practices for data entry from field to office, including building a georeferenced database; conducting simple exploratory analysis of the database; and building maps on QGIS. For those participants that are familiar with mapping a georeferenced database, we will replace the last section (building maps) with a more advanced module that will cover the implementation and interpretation of simple spatial analysis in ArcMap.

**Requirements**

Laptop computer with Excel, SOFA statistic and QGIS installed. SOFA statistic and QGIS are open source software and can be downloaded from:

SOFA statistic: https://www.sofastatistics.com/downloads.php

QGIS: https://qgis.org/en/site/forusers/download.html

**Exercises**

**Data management**

Based on the paper by Kache et al (2020), we will use a compiled database of mosquito surveillance data collected at 7 jurisdictions. Each location had a different sampling strategy and recorded the data differently.

*Exercise 1:*

1. Open “Database\_raw.xlsx” to see the different datasets and take note of any problems that you think may appear if we try to merge the data.
2. Observe differences in:
   1. Sampling Effort: How many traps and for how long?
   2. Sampling Methodology: Which traps and baits were used?
   3. Data Collection: How was the data recorded?
3. Complete the table:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Location 1 | Location 2 | Location 3 | Location 4 | Location 5 |
| Coordinate system |  |  |  |  |  |
| Location of the GPS point (town or trap?) |  |  |  |  |  |
| Unique ID |  |  |  |  |  |
| Long vs. wide database |  |  |  |  |  |

*Exercise 2:*

The second step is to merge the mosquito data with the location data database derived from the GPS devices. For that purpose, we first cleaned “Database\_raw.xlsx” to merge in one database called “Database\_clean.xls”. Similarly, we cleaned the GPS data and created a new database called “GPS\_coordinates.xls”. Your task will be to create the final database by merging “Database\_clean.xls” and “GPS\_coordinates.xls” in Excel using Vlookup function.

**Data analysis**

We will use the final database created in Exercise 2, to do data analysis using SOFA or another statistical software of choice. You can use the guide from the presentation to decide which one is the appropriate statistical test.

*Exercise 1:*

1. Compare mosquito abundance between 2 locations of choice
2. Compare mosquito abundance between all locations

*Exercise 2:*

1. Compare mosquito abundance before august and after august.
2. Compare trapping success (yes/no) before and after august.

Hint:

* Consider that you are using the same traps only measuring mosquito abundance at different times.
* For Exercise 2, you will have to create a new variable in the database.