# Workshop proposal: Data management, visualization and analysis for entomological surveillance, research, and action

**COURSE OBJECTIVE:** Empower participants to fully engage with their data by learning the steps of the data lifecycle: data collection, storage, processing, visualization, analysis, and communication. By the end of the course, participants will feel comfortable using the R statistical environment to carry out basic research and epidemiological/entomological surveillance related tasks such as data cleaning, merging, mapping, and basic statistical hypothesis testing.

**PREREQUISITES:** Students need a laptop computer, basic proficiency in English, and a desire to learn.

**METHODS:** The course will be participatory, and require student activity and engagement. Using demographic, meteorological, spatial, and epidemiological data, each participant will progressively build a "capstone" project: either a semi-formal report/write-up, or a dashboard.

**NUMBER OF INSTRUCTORS AND COURSE LENGTH:**

- 1 or 2 two teachers
- 1 or 2 weeks

Whether the course has 1 or 2 teachers and lasts 1 or 2 weeks is a function of (a) funding, (b) number of participants. What follows is a "basic package" (ie, a 1 week course, or the first week of the 2 week course), and an "advanced package" (ie, the second week of the 2 week course).  The course content is independent of the number of instructors: if the number of participants is large, however, a 2nd instructor is recommended so as to provide additional hands-on help to students who fall behind, etc.

---

## Course Outline

### Week 1: basic package

The one week "basic package" can be a stand-alone course, or the first half of the "advanced package" (2 week course).

- Day 1
    - Introduction to R and RStudio
    - Basic calculations and object oriented programming
    - Best practices for workspace and file management
    - Introduction to libraries/packages.
    - Creating and managing a basic survey for data collection using a manually created spreadsheet
- Day 2
    - Reading in data (flat files), online files, and Excel files
    - Connecting to a database and reading/writing data

- - R syntax, data structures, and variable types
    - Exploratory analysis through visualization: introduction to ggplot
- Day 3
    - Interactive visualization: ggplotly, googleVis, and rcharts
    - Maps and GIS: sp, rgdal, leaflet, and maptools
    - Introduction to "reproducible research": rmarkdown and knitr
    - Dashboard basics in rmarkdown
- Day 4
    - Dashboards and applications in surveillance
    - Widgets, pop-ups, and user experience
    - Automating analysis "pipelines"
    - Principles of documentation
- Day 5
    - Basic hypothesis testing through statistics: chisq.test, ks.test, t.test, lm, and glm
    - Bibliographic management in rmarkdown using bibtex
    - LaTeX, Rmarkdown, html, word, and other formats for data communication
    - How to help yourself - resources for further learning
    - Capstone projects

## Week 2: advanced package only

The two week "advanced" package starts with the one week "basic package" (above) before digging deeper into matters of interactivity, collaboration, versioning, advanced data "munging" (cleaning, re-shaping), and deploying knowledge products (dashboards and reports) to the web.

- Day 1
    - Deep-dive into dplyr: the verbs of data processing
    - Simulating excel operations in R: pivot tables, transformations, summations
- Day 2
    - Advanced data visualization: color palettes, variable-aspect mapping, and facetting
    - Deeper-dive into interactive visualizations: leaflet, rCharts, highcharter, networkD3
- Day 3
    - The shiny web framework: bringing data visualization to life
- Day 4
    - "Literate programming" and "reproducible research": applying the principles to academia and public health
    - Collaboration using Git and version control
- Day 5
    - Writing functions and documentation.
    - Building an R package
    - Deploying projects, reports, and apps to the web