Workshop proposal: Data management, visualization and analysis for epidemiological 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 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. That's all.

METHODS: The course will be active and participatory. It will be challenging, but fun. Using real demographic, meteorological, geospatial, and epidemiological data, each participant will progressively build a "capstone" project: a formal data-driven report with maps, charts, interactive visualizations, and bibliography. The course will be one week in duration.

COST AND OPTIONS: Cost is a function of the number of teachers (1 vs 2). We recommend two teachers. However, if funds are very limited, costs can be reduced by having only one teacher. But our experience suggests that students do best when two teachers are available (one for instruction and one for simultaneous hands-on help).

Course Outline

Day 1
- Introduction to R and RStudio
- Basic calculations and object oriented programming
- Best practices for workspace and file management
- Creating and managing a cloud-based survey for data collection

Day 2
- Reading in data (flat files) and working with Excel.
- Introduction to libraries
- R syntax, data structures, and variable types
- Exploratory analysis through visualization

Day 3
- Interactive visualization
- Maps and GIS
- Introduction to "reproducible research" and Rmarkdown
- Dashboard basics

Day 4
- Deep-dive into dashboards and applications in surveillance
- Widgets, pop-ups, and user experience
- Automating analysis "pipelines"
- Principles of documentation

Day 5
- Basic hypothesis testing through statistics
- Bibliographic management in the R environment
- LaTeX, Rmarkdown, html generation, and other formats for data communication
- How to help yourself - resources for further learning

INSTRUCTORS:

Joe Brew | Data scientist | joe@databrew.cc | databrew.cc/joe.pdf

Ben Brew | Statistician, Economist | ben@databrew.cc | databrew.cc/ben.pdf

ESTIMATED COSTS:

Item	Cost for 1 instructor option	Cost for 2 instructor option
Travel to/from Maputo at $1000	$1000	$2000
5 days lodging at $50 per day	$250	$500
Course fees	$2500	$3750
Total	**$3750**	**$6250**

Comment on pricing and timing:

DataBrew generally charges market price for data science consulting services. However, inspired by our roots in the Univ. of Chicago's "Data Science for Social Good", we strive to discount our fees for organizations which are involved in delivering a "social good". In other words, some underfunded "social good" collaborations are effectively subsidized by fully funded (or non social good) projects.

The discount is a function of the client's ability to pay, the type of work being undertaken, the landscape of alternative partners available to the client, and the capacity of DataBrew at the time of the project. This discounting is made possible by organizations paying market price when able, and by both DataBrew and clients being transparent and clear about capacity and expectations when full funding is not available.

Ministries of Health and health-related projects meet the definition of "social good." If budget-constrained, we should discuss a discount.