Week 1: Introduction to Applied Research Methods in Global Health

Prof. Andrew Stokes

January 22, 2019

Course overview

- This course focuses on building skills in data, tools and methods for carrying out global health research.
- The class will introduce you to all stages of the research pipeline, from asking effective questions to conducting a literature review, data collection, analysis, crafting an effective research paper and communicating results.

Course components

- Implement a mixed methods study from start to finish
- Design and implement a questionnaire
- Conduct a literature review and integrate bibliographic software
- Use R for applied data analysis
- Conduct qualitative data interviews and analyses
- Effectively visualize data
- Apply, interpret and communicate results from statistical analyses
- Write a clear and effective research paper
- Clearly communicate results with diverse audiences

Course philosophy

The philosophy of this class is that we learn best by doing. No matter how much theory you take, there some things you can only learn by plunging in and making mistakes.

Course materials

- GH 811 course website: http://sites.bu.edu/gh811/

Contact information & Office Hours for Prof. Stokes

- Email: acstokes@bu.edu
- T: (617) 414-1276
- Office hours: Tuesday 5:00-6:00 PM
- Office: Crosstown 362

TA Contact information

- Katie Berry
- Email: berrykm1@bu.edu
- Weekly R Lab: Thursday 5:00-6:00 PM, CT 460A

Assignments

The key assignments in this class are:

- Course project
- Weekly peer reviews
- Problem sets
- In-class activities
- Preparing for class (e.g. readings, R learning modules)
- Applied research methods journal

Texts

- Peter G. Smith. Field Trials of Health Interventions: A Toolbox
- Other references listed under Logistics on course website.

Software

Becoming proficient in software tools is a major emphasis of the course. These tools will be used for designing and fielding surveys, cleaning and analyzing data, report writing and managing bibliographic information. Some of the tools you will use this semester are:

- R/R Studio
- Kobo Toolbox
- Mendeley
- NVivo

Course communications

- Course listserve

Course Project

- In this class, you will conduct a semester long research project culminating in a final paper.
- Just like in the real world, this will be a highly iterative process, meaning you will be submitting many drafts and receiving continuous feedback throughout the semester, both from the instructor and your peers.
- The paper will be prepared in stages, corresponding to the materials presented in class each week.

Overview of weekly schedule

- Submit project component via email by Sunday 5 pm
- Complete peer review by class time Tuesdays
- Revise based on instructor and peer feedback (you will receive written feedback from me by end of Wednesday each week)
- Re-submit project component for further feedback (optional but encouraged)

Journal

- You will submit three journal entries during the course of the semester (see course website for deadlines)
- Each entry should be no more than 2 pages, double spaced.
- It should include your reflections on your experience in the course (in and out of class). Suggestions for content are provided on the course website, but you are not limited to these topics.

Group Charter and Evaluation

- Each group will develop a charter and evaluation form outside of class.
- At mid-term, groups should schedule a meeting to review the charter, reflect on work dynamics to date, and plan for any needed changes going forward.
- At the end of the semester, each student will complete an evaluation form to assess overall group dynamics and individual contributions.

Grading

Grades will be distributed according to the weights below. 55% of the grade is group-based. The remaining 45% is individual.

- Final written research report (30%)
- Introduction, Methods and Results Sections (5% each x 3)
- Final presentation (10%)
- Problem Sets (5% each x 2)
- Peer Reviews (10%)
- Participation (10%)
- Journal entries (5% each x 3)

Why mixed methods?

What big data won't tell you

- What did this study gain from using a qualitative approach?

What big data won't tell you

- What did this study gain from using a qualitative approach?
- What could a quantitative approach add in this context?

Qualitative research

- Provides depth of understanding
- Asks why and how
- Strives for *emic* perspective (i.e. description of behavior or a belief in terms meaningful to the actor; as opposed to *etic* which provides a description of a belief or behavior by a social analyst or scientific observer in terms that is applicable across cultural contexts)
- Is exploratory
- Enables discovery
- Can allow for insights into behaviors

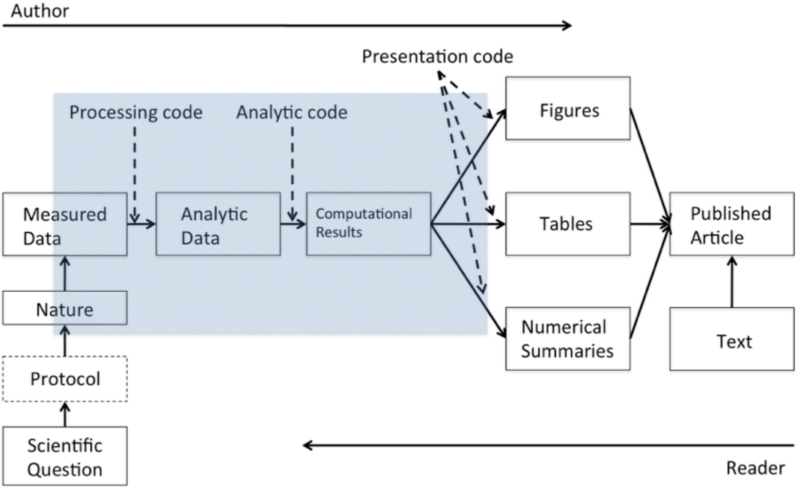
Quantitative research

- Measurement
- Asks how many and how often
- Evaluates strength of relationships between variables
- Causal inference

Value of mixed methods

- Quantitative research can be used to facilitate qualitative research by generalizing findings to a large sample.
- Quantitative research helps to identify groups that warrant in-depth, qualitative study
- But generally not used for finding out why people do or don't do something, so augmenting survey data with interviews or other qualitative methods makes for more complete understanding.

The Research Pipeline[1]



[1]Source: Roger Peng

Question for the class

- How have you performed these steps in your research?
- How are these steps usually performed?

One-off approaches to data processing & analysis

- Research is often performed using a calculator or spreadsheet
- But these types of one-off approaches have serious drawbacks

Example

- Let's say you find out that one of the data collectors made up data. So you go back to the raw data and drop those observations.
- But now you need to re-do all your calculations.
- If you did these by calculator or spreadsheet it may take a long to replicate.
- You may forget what you steps you took
- Mistakes might be made along the way. These can be difficult to detect.

The solution

- Coding means writing scripts in R or another programming language that automate processes of cleaning, processing and analyzing data.
- Automation is key: the code takes the raw data, pre-processes it and then analyzes it, without any intervening steps taken by the analyst.

Why automate?

- Transparency: your code provides a record of the steps you took in processing and analyzing your data that can be shared with others (and with your future self). It thus gets rid of the black box steps that are a feature of most research.
- Accuracy: automation reduces risk of making errors that inevitably arise when procedures are performed manually (e.g. in a spreadsheet)

Why automate (continued)?

- Robustness checks: automation makes it possible to re-run analyses in different ways to examine the robustness of the results to alternative assumptions.
- Replication and reproducibility: automating data processing and analysis not only makes you a more effective and efficient reseearch, it enahnces the reproducibility of your work by independent scientists. Replication scientific enterprise.

Replication and reproducibility

- Ionnadis
- IoM and Science reports
- NYT article on the reproducibility project at the Center for Open Science

The state of science according to John Oliver

Scientific Studies: Last Week Tonight with John Oliver

Case Study: Reinhart-Rogoff

Planet Money Podcast: How much should we trust economics?

Coding in GH 811

- For your course project you will use R and R Studio to pre-process and analyze the data you collect.
- You will write scripts that automate each step you take with the data.
- At the end of the semester, you will append the code you have written to your final paper.

Break for group work on team charters

Ethical considerations in selecting a research topic

IRB guidelines regarding student research projects state that the projects undertaken must be "minimal risk". The regulatory definition of minimal risk is as follows:

- The probability and magnitude of harm or discomfort anticipated in the research are not greater in and of themselves than those ordinarily encountered in daily life or during the performance of routine physical or psychological examinations or tests (45 CFR 46.102(h)(i)).

Ethical considerations in selecting a research topic

Student projects are considered a *classroom experience* if the following criteria hold:

- The purpose of the project is the learning experience
- Not physically or psychologically stressful, invasive
- No more than minimal risk
- Not accessing research or clinical databases/medical records containing identifiable data
- Not designed to contribute to generalizable knowledge (no plans for publication).

If all these criteria hold, IRB review is not required.

What makes a good research question?

- Why is the uptake of improved stoves so low?

What makes a good research question?

- Why is the uptake of improved stoves so low?
- Why is the uptake of improved stoves so low in rural areas?

What makes a good research question?

- Why is the uptake of improved stoves so low?
- Why is the uptake of improved stoves so low in rural areas?
- What are the determinants of uptake of improved stoves in rural Ethiopia?

What makes a good research question?

- Why is the uptake of improved stoves so low?
- Why is the uptake of improved stoves so low in rural areas?
- What are the determinants of uptake of improved stoves in rural Ethiopia?
- Do womens' perceptions of the health hazards of indoor air pollution affect demand for improved stoves in rural Ethiopia?

What makes a good research topic?

A good research question can often be stated as a problem. The IDRC text offers three criteria for problems that make ideal research topics:

- Perceived difference between what exists and the ideal (what is wrong?)
- Reasons for difference unclear (why is it wrong?)
- Often more than one answer to the problem exists (what is the best approach to fix it?)

Example

Let's say the problem you are interested in is low uptake of improved stoves in rural Ethiopia. Does this problem meet the criteria stated above?

Example continued

- The current situation is that few people have adopted improved stoves. Ideally, uptake would be much greater, since the health benefits can be substantial.
- The explanation for why there is low uptake is not obvious.
- The solution could be to make fuel more readily available, design stoves that are more consistent with culture and cooking practices, make the stoves more convenient ot use, or lower the price. In other words, there are many different potential solutions and it's not clear which avenues should be pursued.

Group work: brainstorm for class project

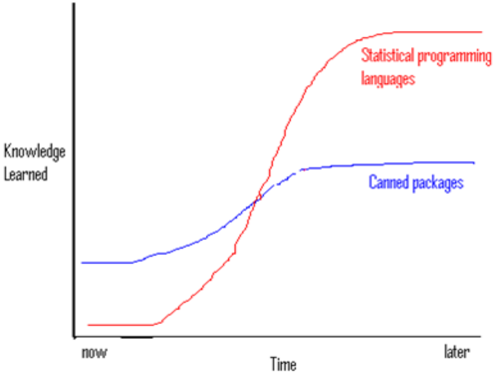
Introduction to R

- R is a programming language for statistical data manipulation and analysis
- R is an alternative to programs like SAS and Stata

Why R is growing in popularity

- Open-source, which means the source code is available for free on the internet
- Free (vs. SAS and Stata, which have expensive licensing fees)
- Large and thriving user community
- Many user-contributed packages that extend functionality and increase relevance to latest methods
- Flexible, useful and compelling graphics

Statistical programming languages vs. canned packages[2]



[2]Source: Gary King

Downloading R and RStudio

- Base R is the default platform for programming in R and can be downloaded from the Comprehensive R Archive Network (CRAN) at https://cran.r-project.org/.
- R Studio (https://www.rstudio.com/) is a platform for programming in R. It manages version control, includes autocomplete features, saves plot history, manipulates graphs interactively and saves search command history.
- To run R Studio, you have to install CRAN R first.

R Studio Interface

The R Studio Interface has four parts:

- Top left: R script window, where you will write your R program.
- Bottom left: this is the console and is where code is implemented and the output appears. Note that you can write code directly in the console, but this should be done sparingly

R Studio Interface

- Top right: this is the Workspace. The Workspace includes every object used in your current R Studio session. Objects include data sets, vectors, regression results, etc.
- Bottom right: This quandrant includes several tabs, which include any plots you generate in R, access to help files and packages you have installed and activated.

Helpful Keyboard Shortcuts in RStudio

Ctrl+Enter	Runs the line of code your cursor is on (does not need to be highlighted). If you want to run multiple lines, you should entirely select all the lines, like you would in SAS.
Tab	Automatically completes code and provides comprehensive lists of functions
Ctrl+1	Moves focus to the R Script
Ctrl+2	Moves focus to the Console
Ctrl+L	Clears the Console
Esc	Interrupts R

Introductory Commands in R

`<-`	Assignment operator which allows you to assign data fields to an object
`#`	Comment (similar to /* */ in SAS) but unlike SAS you have to include a # on every new line of your comments
`####[section name]####`	Enclosing text with four hash tags on either side allows you to segment your code into sections which can be accessed by name from a drop down menu in R studio.
`c()`	Combine allows you to combine elements of the same mode into a vector

Logical Operators in R

==	Equal to
!=	Not Equal to
>	Greater than
>=	Greater than or equal to
<	Less than
<=	Less than or equal to
&	Vectorized AND
|	Vectorized OR
!	Not

The Working Directory

The first thing you should do when you begin an R script is to set the working directory. This is the default location on your hard drive that R will look to in order to read and write files.

```
input <- "C:/Users/acstokes/Desktop/IH 811"
setwd(input)
data<-read.csv("name_of_data_set", header=TRUE)
```

Data Structures in R

R has an object-oriented data type system, with each type of object having its own properties. The major types are:

- scalars
- vectors
- matrices
- lists
- data frames
- factors

Scalars

Scalars are simply an individual number, such as 3.

Vectors

- A vector is an array of elements with the same mode (e.g.numeric, character, or boolean).
- Scalars are stored as a vector of length 1 in R.
- Columns in a matrix or variables in a data frame are also stored as vectors.

Matrix

- A matrix is a rectangular array of same-length vectors, all of which have the same mode.
- R functions that perform statistical models often require the input to be a matrix.

List

- A list is a collection of other objects, none of which has to be of the same mode or structure (e.g. a list could contain both a character vector and a numeric matrix as separate elements).
- Many R functions that perform statistical models return the output as a list containing various components

Data frames

- In a data frame, each column is a vector that can be of different modes
- Data frames are similar to datasets in SAS or Stata

Factors

- Factors are similar to vectors with one difference: factors include a list of its unique values, called levels, sorted alphabetically for character values and in ascending order for numeric values
- This additional information allows R to perform categorical analyses

Examples of R Objects

Object Type	Example
Scalars	> pi [1] 3.141593
Vectors	> x <- c(1,2,3,4) > x [1] 1 2 3 4
Matrices	> X <- matrix(c(1,2,3,4), nrow=2,ncol=2) > X [,1] [,2] [1,] 1 3 [2,] 2 4
Lists	> x [1] "one" "two" "three" > y x1 x2 x3 [1,] 1 2 3 [2,] 2 3 4 [3,] 3 4 5 > L <- list(vector=x, matrix=y) > L $vector [1] "one" "two" "three" $matrix x1 x2 x3 [1,] 1 2 3 [2,] 2 3 4 [3,] 3 4 5
Data Frames	> D <- data.frame(x1, x2) > D x1 x2 1 1 AB 2 2 BC 3 3 CD
Factors	> clothingf <- factor(clothing) > clothingf [1] shirt shirt jeans hat vests jackets vests Levels: hat jackets jeans shirt vests

For next time

- Team charter
- Topic selection sheet