

My Proposal

Using R for Reproducible Analysis

Al Cooper

RAF

RSIG Meeting 6 Oct 2014

Reproducible Analysis

My personal goal: “reproducible analysis.” There is now extensive literature on reproducible research, a lofty goal, but much more difficult to achieve, esp. at the data-collection stage. Instead, I’d like to aim for analysis procedures that can be duplicated by others. The key tools I have chosen are R, RStudio, knitr, \LaTeX (not essential), and git.

My Role:

I’m actively studying this, aiming to use these techniques in RAF analyses and documentation and also for an ongoing research project. As part of my own education, I’m willing to present a series of 30-min sessions, perhaps once a week, that provide an introduction to this structure and involve “homework”, examples, and developing a reference that might be useful in the future.

WHY THESE TOOLS?

A personal choice

- There are other candidates: python (esp. ipython notebooks), Matlab documents, emacs org-mode (a very nice tool)
- I have tried all at various times, most recently the ipython notebook, but RStudio has big advantages for my use.
- Strengths:
 - 1 Free
 - 2 Work on all platforms
 - 3 R: Extensive set of packages available for statistical analysis, graphics, etc.
 - 4 Good help functions and tutorials
 - 5 Expertise elsewhere in RAF and EOL and in the CISL/Image group
 - 6 Much literature available on “reproducible research” using this approach
 - 7 Straightforward to develop specialized packages for our uses
 - 8 I find this approach very convenient to use.

SOME POSSIBLE SESSIONS - 1

Quick Introduction to RStudio

- How to get access to it and how to run it
- Using R as a calculator
- A simple illustrative script
- Some basic concepts:
 - 1 objects - esp. the data.frame
 - 2 packages
 - 3 getting help

SOME POSSIBLE SESSIONS - 2

An Example of How This Fits Together

CONTRASTcalAOA.Rnw

- An ASCII file that mixes R code and document-generating code
- The product is a PDF-format report
- The content of that report includes figures and statistics generated internally via R.
- Can be saved with the data and other key components in a package that is distributable and archived.

SOME POSSIBLE SESSIONS - 3

Some Tools for RAF Analysis Tasks

The RAF package:

- Components
- Illustrations of use
- Opportunities to work together to extend the package

SOME POSSIBLE SESSIONS - 4

Candidate Projects That Could Follow This Analysis Model

- Uncertainty characterization
- Project documentation
- Fitting for sensitivity coefficients
- Analyzing results from maneuvers
- Data quality reviews
show specific examples

SOME POSSIBLE SESSIONS - 5

Working With 'Chunks' in RStudio

- Structuring and sectioning code
- Executing chunks
- Debugging
- Conditional execution and other options
- Generating figures and text for inclusion in reports

WHAT NOW?

Next Steps:

- Worth the time?
- Sufficient interest?
- Gather some books?
- Schedule?

My Opinion:

We need to do this or something like it because these analyses are needed and, without structure, it is difficult to repeat them or transfer them to others.