R and spreadsheets – combining different programming paradigms

Erich Neuwirth

erich@statconn.com
Topics covered

Spreadsheet examples
Conceptual issues
User profiles
Why spreadsheets as client?

Well known user interface
Easy data manipulation
Integration into daily working environment
Widely accessible environment for numerical computations
Simple (simplistic?) toolkit for static and animated graphics
User profiles

Methods developer

Sophisticated methods user
and small scale application developers

Naïve user knowledgeable about the spreadsheet model
Design issues for user interface

Closed application
Some “end user programming”
Open development environment
Enhancement
  of core spreadsheet functionality
Language paradigms

Main difficulty
Coping with different paradigms in one application

In R, the user has total control over “calculation flow”
In Excel, the application triggers calculation
Excel decides about the order of calculation!
Language paradigms

Common data types
statconnDCOM server (by Thomas Baier) supports arrays which may even contain different scalar data types
These arrays can be accessed both from R and from VBA.
Excel “by nature” only supports 2-dimensional arrays (ranges)
Integration

Excel keeps state in the worksheet (cells and/or ranges)

R keeps state in variables

State in both applications may become “out of sync”

Excel does computations “on the safe side”, better too often than not often enough
Integration

So far Excel was client and R was server

rcom also allows R as client and Excel as server
This way, R can control Excel transfer data in both directions, create spreadsheet formulas ...

Excel also could become the data editor for R
Tools for subtasks

**Excel (more general: spreadsheets)**
- Data preparation and manipulation
- Exploratory methods (possibly)
- Presentation

**R**
- Powerful analytical methods
- Exploratory methods (numerically intensive)
- Preparing data for visual presentation
- Advanced statistical graphics (additional libraries)

**VBA**
- Writing user interfaces on top of Excel
- Transfer large amount of data outside of recalculation cycle
More information

R:
http://www.R-project.org
   CRAN repositories

Spreadsheets:
http://sunsite.univie.ac.at/Spreadsite

RExcel and R(D)COM
http://rcom.univie.ac.at
http://www.statconn.com