R-Shared: A Collaborative Platform for R Code Development and Sharing

8th March 2011

Richard Pugh, Commercial Director, Mango Solutions UK, rpugh@mango-solutions.com
Agenda

- Overview
- Projects that Inspired ...
  - ModSpace
  - Commercial R Support
- Introducing LondonR Ltd!
- The R-Shared Platform
- Summary
- Questions
Overview
Overview

• More teams starting to use R in commercial situations
  • How do these people communicate?
  • Central knowledge sharing
• IT teams needing to work with R:
  • Improve performing with use of hardware
  • Control R versions and packages
Overview

- R used for more and more vital decisions
  - More focus on validation/verification
  - Code control and versioning
- Support framework needed as user base increases
  - R-Help fantastic but not commercial focus
  - Commercial training & support becoming more important
Overview

• Today I will present Mango’s “R-Shared” platform
• Builds on other Mango projects
• A “concept” at this stage
Projects that Inspired ...
ModSpace

- Built for Novartis, but owned by Mango
- Knowledge Management system that:
  - Understands the difference between “code” and other files (e.g. documents)
  - Backed by version control system
  - Uses the high performance “Lucene” search
- Used heavily at Mango
New entry: Automated R installation script for Linux
Installs R 2.10 and pre-requisites for rjava, and all required packages for PME 2.4.

step-07-04.sh installs R
step-07-06.sh installs all R libraries
R installation scripts used in PME, installs R and R packages.

Mike Creed, 08-Mar-11 12:08PM

New entry: Logging Module for Bash Installation
A modular script that provides logging functionality, which can be included in bash scripts to provide the following:
- api for creating logs: DEBUG, INFO, ERROR, FATAL
- Termination of a process...
Useful logging script for bash installation script - writes to console and log file, displays errors in colour.

Mike Creed, 08-Mar-11 12:03PM

New entry: NONMEM 6 and 7 installation scripts
Installation scripts used by PME to automate the installation of NONMEM 6 and 7. These are modular scripts that can be called with parameters to determine default and custom sizes.

The PME install...
<table>
<thead>
<tr>
<th>Name</th>
<th>Last Modified</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Forest Research R Training 2010</td>
<td>03 Jun 2010</td>
<td>3%</td>
</tr>
<tr>
<td>Chapter 04 - Producing Graphics.doc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Richard Pugh</td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACOP Graphic Scripts</td>
<td>16 Feb 2011</td>
<td>2%</td>
</tr>
<tr>
<td>newBoxplotFunction.R</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Richard Pugh</td>
<td></td>
<td></td>
</tr>
<tr>
<td>R for Actuaries</td>
<td>20 Feb 2011</td>
<td>2%</td>
</tr>
<tr>
<td>6.TraditionalGraphics.doc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Richard Pugh</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pfizer 5 Day S-PLUS Training 2004</td>
<td>29 Jun 2010</td>
<td>2%</td>
</tr>
<tr>
<td>Chapter8.doc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Richard Pugh</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pfizer 5 Day S-PLUS Training 2004</td>
<td>29 Jun 2010</td>
<td>2%</td>
</tr>
<tr>
<td>Tip_Sheet.doc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Richard Pugh</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pfizer 5 Day S-PLUS Training 2004</td>
<td>29 Jun 2010</td>
<td>1%</td>
</tr>
<tr>
<td>Tip_Sheet.doc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Richard Pugh</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abbott Intermediate S-PLUS Training 2007</td>
<td>29 Jun 2010</td>
<td>1%</td>
</tr>
<tr>
<td>Tip_Sheet.doc</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Richard Pugh</td>
<td></td>
<td></td>
</tr>
<tr>
<td>GSK Webinars for S and R (2008 - 2010)</td>
<td>23 Jan 2010</td>
<td>1%</td>
</tr>
<tr>
<td>Training Slides - 2010 - 2012</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Richard Pugh</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Useful R functions

**Description**

R memory management seems to be quite good. However, code should be created to make best use of RAM. This function shows the time cost of stretching (cbinding to) a vector rather than populating a planned matrix.

This entry now also includes a Mango branded lattice theme and relative speeds of apply, ifelse, factors and subsetting

**Assumptions**

R

**Projects**

Not applicable

**Keywords**

R, memory, lattice

**Language**

S, R
mango.theme <- function(safe = FALSE) {
mango.palette <- matrix(c(  
  c("#FF80BF", "#E8112D", "#00A3DD",  
    "#0D88C8", "#4BC977", "#00C9C9"),  
  c("#FF80BF", "#CC0000", "#0000FF",  
    "#0D88C8", "#4BC977", "#00C9C9")),  
  ncol = 2, dnames = list("Solutions", "CRM", "Finance",  
    "Pharma", "Sensory", "Energy"), c("True", "Safe")))
  # define Mango colours
  column <- 1
  if(safe) {column <- 2}
  # toggle true or browser supported colours
  mango.hsv <- rgb2hsv(col2rgb(mango.palette[1:6, column]))
  mango.hsv[2, ] <- 0.3
  mango.light <- hsv(h = mango.hsv[1, ], s = mango.hsv[2, ],
    v = mango.hsv[3, ])
  # create vector of lightened colours
  # saturated colours influence perception of data
  par <- col.whitebg()
  # show settings() row 1
  par$superpose.symbolCol <- mango.palette[c(6:1, 6), column]
  par$superpose.linecol <- rep(1, 7)
  par$superpose.linecol <- mango.palette[c(6:1, 6), column]
  par$strip.borderCol <- "black"
Daiichi Sankyo

- Ongoing Commercial R Support
- Provision of “validated” R build
- Optimisation of R performance
- Performance of regression testing to provide evidence of verification
Daiichi Sankyo

Daiichi Sankyo

Test Reports

DSI CRAN

Mango CRAN

Mango

CRAN

www.mango-solutions.com
Introducing LondonR Ltd
R User Community at LondonR Ltd
Where do they get R?
Most CRAN
All using different versions, packages, IDEs, etc
Where did they get help?
R-Help sometimes
Otherwise they ask each other

R User Community at LondonR Ltd
How about that package that Bob wrote!
Sent around via email
Code maintained on Bob’s machine or R-Forge
Everyone has different versions

R User Community at LondonR Ltd
Where do they get training?
Group 1 had a private training course from Mango
Someone from Group 2 went on a public course
Others are self taught
Where is their code?
On their own machines mostly
Emailing of scripts
Server to put some code but somewhat disorganised
How about some IT support?
IT team is excellent, but don’t know R and are not analysts
R is open source, which adds concern
Worried about installing on central servers
They google for commercial support and don’t find a lot …

R User Community at LondonR Ltd
How do the teams communicate?
They don’t really (no-one even knows about Group 3)
Sometimes see questions on R-Help
Other potentially see them at LondonR or other events
The R-shared Platform
The R-shared Platform

- Integration of the ModSpace software with DSI structures
- On development path at Mango for May 2011
- Looking for interested parties to get involved
**My Favourites**

- My Entries
- My Groups

**Questions**

- How do I calculate the risk of an asset within R?
  - By rpuh on 1st Feb 2011
  - See 8 responses

- Does anyone have a script to create a boxplot using the mean as the mid-point of the box instead of the median?
  - By mcreed on 10th Jan 2011
  - See 4 responses

**News**

- **New R Package: ggplot2 v1.0.1 now available for use**
  - Hadley’s Wickham’s ggplot2 library has now been tested and is available for download
  - Richard Pugh, 1st Feb 2011 15:30pm

- **Feedback: Interactive Graphic Windows**
  - “Really good video Mike - got any more you can post?”
  - Jonathan Chard, 30th Jan 2011 10:01am

- **New R Training Entry: Interactive Graphic Windows**
  - Video showing how to set your graph windows to be interactive, allowing you to scroll through plots within the R environment
  - Michael Creed, 29th Jan 2011 16:05pm

- **New Entry: Scripts for Risk Assessment**
  - Set of basic scripts that can be used to perform a variety of risk assessment tasks
  - Michael Creed, 29th Jan 2011 15:34pm
Where do they get R?

- CRAN
- Mango CRAN
- LondonR CRAN

News:
New R Package: ggplot2 v1.0.1 now available for use

Hadley's Wickham's ggplot2 library has now been tested and is available for download

Richard Push, 1st Feb 2011 15:30pm
Where do they get Help?

Search for an answer

Ask Mango a question

Internal Q&A

IT Team

Hardware

Group 1

Group 2

Group 3

www.mango-solutions.com

R User Community at LondonR Ltd
How about that package that Bob wrote!

It lives in ModSpace

Can be downloaded

Can be further developed by everyone with access

Updates are announced in R-Shared

IT Team

Hardware

Group 1

Group 2

Group 3

R User Community at LondonR Ltd
Where do they get training?

- Training announcements part of news feed
- Training materials stored in R-shared

R User Community at LondonR Ltd
Where is their code?
In R-shared, so stored in version control system
How about some IT support?
IT users have admin and management interfaces
How do the teams communicate?

Using “groups” within R-shared and view news feed
Summary
Summary

• R-shared Platform in development (although there isn’t a lot to do!)
• Should be available May/June 2011
• Looking for companies who want to get involved at this stage
Questions