R-Related Features and Integration in STATISTICA

✓ Run native R programs from inside STATISTICA
✓ Enhance STATISTICA with unique R capabilities
✓ Enhance R with unique STATISTICA capabilities
✓ Create and support (FDA) validated installations using R
✓ Use STATISTICA Enterprise Server to create a scalable R server

data analysis  ●  data mining  ●  quality control  ●  web-based analytics
Table of Contents: Comprehensive Native R Support in STATISTICA

- Executive overview
- Running R programs as native STATISTICA macros
- Off-Loading to STATISTICA Enterprise Server
- Capturing detailed results from R into STATISTICA spreadsheets, reports, graphs
- Running R Scripts from STATISTICA using STATISTICA’s flexible UI
- Building new functions for STATISTICA using R libraries
- Integrating R libraries into STATISTICA: Technical details
- Interfacing directly with R through the COM interface
- Creating R-based STATISTICA Data Miner nodes
- Integrating R into STATISTICA Enterprise (using R in validated analytic reporting)
- Using STATISTICA Enterprise Server to create a scalable R server
- Summary: Comprehensive native R support in STATISTICA
- For more information contact

  StatSoft Inc.
  2300 East 14th Street, Tulsa, OK 74104
  Phone: (918) 749-1203
  Fax: (918) 749-2217
  Or visit www.StatSoft.com
Executive Overview

- R is a programming language and environment for statistical computing; R and its source code is freely available under the GNU GPL license (see [http://cran.r-project.org](http://cran.r-project.org))

- With STATISTICA, native R scripts can be run directly within STATISTICA. R output can be retrieved as native STATISTICA spreadsheets and graphs, and managed via highly flexible STATISTICA Workbook containers.

- Thus, enterprises can now use the specialized routines and capabilities of R with STATISTICA, STATISTICA Enterprise, and STATISTICA Enterprise Server:
  - Add new R-based “modules”
  - Leverage STATISTICA’s superior graphics, flexible Spreadsheets, and convenient Workbook containers for various document types to handle output from R
  - Build scalable R servers using STATISTICA Enterprise Server to handle security, load balancing, and to take advantage of multiple-processor servers to run R for demanding and/or validated enterprise applications.
Running R scripts from STATISTICA

R scripts as native STATISTICA macros

- You can now run a complete R script within STATISTICA rather than from the R console:
  - Create new or load existing R scripts
    - Load .R or .S files;
    - STATISTICA will treat them as native macros
  - Simply run the script
  - R console output will be automatically captured into a STATISTICA Report
    - R commands highlighted
    - All graphs are captured
  - You can now
    - Create PDF files
    - Place reports into STATISTICA Document Management System as validated reports
    - ....
Off-Loading to STATISTICA Enterprise Server: Create a Powerful R Server

- STATISTICA Enterprise Server is a powerful web-enabled client-server architecture that based on and code-compatible with all STATISTICA libraries; see also Using STATISTICA Enterprise Server to create a scalable R server

- R scripts (as well as SVB scripts, Data Miner workspaces, etc.) can be off-loaded from STATISTICA desktop to STATISTICA Enterprise Server for execution, taking advantage of powerful multi-processor server hardware

- With STATISTICA and STATISTICA Enterprise Server, R users have available to them a powerful multi-user, multi-processor R server capable of batch processing, scheduled “R-jobs”, …
Running R scripts from STATISTICA
Capturing Detailed Results

- With only small modifications to the R script, you can
  - Pass in a STATISTICA data file
  - Extract results tables into “real” STATISTICA results spreadsheets
  - Extract results graphs into STATISTICA graph objects
  - Put all results into STATISTICA Workbooks, just like native STATISTICA output

- Use language extensions:
  - `ActiveDataSet` and `Spreadsheet(filename)` to transfer spreadsheets to R as “data frames”
  - `RouteOutput (array/matrix-object)` to retrieve vectors, matrices, data frames as STATISTICA tables
  - All R plots are automatically copied to STATISTICA graphs as Metafiles
  - These graphs are scalable vector images which can be annotated with text, arrows, etc. using interactive STATISTICA tools (see next slide)
Running R scripts from STATISTICA
Using STATISTICA’s Flexible UI

- Once R results have been transferred into STATISTICA, the full power of the STATISTICA interactive desktop is available to:
  - Print tables and reports as PDF files
  - Perform follow-up analyses using the comprehensive STATISTICA analytic toolsets
  - Modify, enhance, annotate graphs interactively
  - Manage sets of results in convenient workbooks
  - Archive and version results using STATISTICA Document Management
Building New Functions for STATISTICA

- R code can be executed directly from inside STATISTICA Visual Basic
- Parameters (numbers, strings, arrays, spreadsheets, even additional R code) can be passed to R using the STATISTICA Collection object – they become named R variables
- In this manner, new functions can be built into STATISTICA that are entirely or partially based on R, or that “mix” R and STATISTICA functionality; for example:
  - Create the analysis macro and attach it to the menu so that it becomes a new “STATISTICA module”
  - The macro can show UI so that the user can select variables or set parameters for the R-based analysis
  - Results will be produced inside STATISTICA workbooks
  - Thus, a new Nonlinear Time Series module has been added to STATISTICA
Integrating R libraries into STATISTICA: Technical Details (1)

- To make a module based on R functionality, create these files:
  - An R program that performs the computations (in R), and uses special “extensions” (e.g., RouteOutput, ActiveDataset) so that data (results) and graphs can be exchanged between the STATISTICA and R contexts
    - “Under the hood” this R program will be parsed and executed from within STATISTICA by an SVB support macro, which handles the special keywords to exchange data (results, graphs) with R
  - Note: you need to install the R environment in order to execute R scripts in STATISTICA (see http://cran.r-project.org/ for details)
  - A STATISTICA Visual Basic (SVB) macro that handles UI (accepts parameters, variable selections, ...), and executes the R “macro”
    - The SVB program will load and execute the R program and automatically extract all results into the standard STATISTICA RouteOutput, i.e., into workbooks, stand-alone reports, or individual STATISTICA objects (spreadsheets, graphs)

- The following slides provide details
  - Examples are provided with the STATISTICA installation
Integrating R libraries into STATISTICA: Technical Details (2)

- To make a module based on R functionality, follow these steps:
  - Write the R code as usual with R tools, test and debug your script
    - Or use an existing solution created by the R community
  - Then write the STATISTICA Visual Basic script to create and service the UI, accept variable selections, parameters for the R script, and so on
    - STATISTICA SVB allows you to build complex dialog boxes with all standard Windows controls
  - Functions are available for accepting variable lists, etc.
  - Add a Collection object to the SVB code to pass parameters to the R script (numbers, strings, arrays, spreadsheets)
  - Open the R script in STATISTICA and
    - Check for and retrieve parameters
    - Use ActiveDataSet or Spreadsheet(filename) to transfer data to an R Data Frame
    - Use RouteOutput() to direct output to STATISTICA workbooks
Integrating R libraries into STATISTICA: Technical Details (3)

- The SVB code can then call or “run” the R code; inside the SVB code:
  - Open/create the R script inside the SVB macro; `Macros.Open(filename)`, `Macros.New()`
  - Execute the R script from the SVB macro; e.g.:
    - `Results = Macro.ExecuteNoRouteOutput([Parameters])`
    - `Results` is a `StaDocCollection` object
  - Display the `Results` via the `RouteOutput()` function to send them to workbooks/reports/..., or iterate through the contents to extract specific data

---

```vbnet
' Open/create the R script inside the SVB macro
Macros.Open(filename), Macros.New()

' Execute the R script from the SVB macro; e.g.:
Results = Macro.ExecuteNoRouteOutput([Parameters])

' Results is a StaDocCollection object

' Display the Results via the RouteOutput() function to send them to workbooks/reports/..., or iterate through the contents to extract specific data
```
Interfacing directly with R through the COM Interface

- In general, R programs can (always could) be run from STATISTICA Visual Basic (SVB)
- With STATISTICA, the details of the interface to R are automatically handled when you open an R program file inside STATISTICA (as illustrated on the previous slides)
- However, R can also be accessed directly through COM via the “COMadaptR library”
  - distributed under GNU Lesser Public License
- With COMadaptR library installed (and R), add the COMadaptRLib (1.0) SVB references to the SVB script
- In SVB, instantiate a COMadaptR object:
  ```vba
  Dim s As New COMadaptRLib.COMadaptR
  S.Init("R")
  s.EvaluateNoReturn( " R script command ")
  Dim retval As Variant
  retval = S.Evaluate( " R script command ")
  ```
- When the user opens an R program (.R or .S file name extension), all necessary support to run the script is automatically loaded, and accessible on the second tab of the STATISTICA macro window
  - Users can expand or customize the R (R.SVB) support macro
Creating R-Based Workspace Nodes

- **STATISTICA Workspace** can be expanded with custom-designed STATISTICA Visual Basic (SVB) scripts.
- Simply follow the procedure for writing **STATISTICA Workspace** nodes, and use R functionality in the same way as described on the previous slides.
- Because it is simple to create user interfaces in **STATISTICA Workspace** (nodes), it is easy to create **STATISTICA Workspaces** that incorporate or mix the capabilities of **STATISTICA Workspace** with specialized R functionality.
Integrating R Functionality into STATISTICA Enterprise

- **STATISTICA Enterprise** is an enterprise data analysis platform for role-based secure data analyses and analysis reporting, in use worldwide in often mission critical (FDA) validated and non-validated applications.

- **STATISTICA Enterprise** allows certain administrative users to create data configurations (reusable queries, metadata) and data analysis configurations (reusable analysis templates, analytic reports).

- Using the methods described on earlier slides, **STATISTICA Enterprise** based analyses and reports can now incorporate all R functionality.

- Leverage the specialized power of R in an effective enterprise analysis platform, where end users do not need to know R, or any programming language!
**STATISTICA Enterprise Server creates a scalable R server**

- **STATISTICA Enterprise Server** is a client-server architecture where **STATISTICA** runs as a service
  - Multiple instances of **STATISTICA** can run simultaneously, to handle multiple jobs
  - Individual **STATISTICA** “jobs” can be distributed over multiple processors
  - **STATISTICA Enterprise Server** handles heavy workloads (load-balancing) in a smart way
  - Users (clients) can work with the web server interactively, or by submitting batch jobs which are either automatically scheduled or scheduled by the user

- With **STATISTICA**, you can now run on the server:
  - Native R programs (e.g., submitted from **STATISTICA** desktop)
  - **STATISTICA** Visual Basic (SVB) scripts that call R programs
  - R-based **STATISTICA** Data Miner projects
  - **STATISTICA Enterprise** analysis configurations (templates) based on R functionality

- **STATISTICA Enterprise Server** is in fact a powerful multiprocessor R analysis server
Summary: Comprehensive Native R Support in STATISTICA

- With STATISTICA, users can now take full advantage of the specialized power of R, while using all the powerful STATISTICA and STATISTICA Enterprise Server features (analytics, graphics, flexible handling of results tables, printing/PDF support...)

- With STATISTICA, there are various ways to integrate with R, by:
  - Accessing R COM interfaces for low-level interaction
  - Running R programs directly from STATISTICA, and retrieving results to STATISTICA reports, workbooks and graphs
  - Using STATISTICA datasets in the R environment and retrieving tabular results from R programs into STATISTICA spreadsheets
  - Calling R from STATISTICA Visual Basic (SVB), to create STATISTICA functionality that leverages R libraries
  - Running R from STATISTICA Enterprise (creating reusable R-based analysis configurations/templates, to deliver the power of R to users not familiar with R)
  - Creating and running R-based STATISTICA Workspace nodes, to integrate specialized R routines into STATISTICA Workspace
  - Running R from STATISTICA Enterprise Server (using any of the available methods described above), to create powerful, secure, multi-processor R servers with load balancing, batch-job capabilities (scheduling), etc.
For More Information

- Contact StatSoft Inc.
  2300 East 14th Street, Tulsa, OK 74104
  Phone: (918) 749-1203
  Fax: (918) 749-2217

- Or visit StatSoft (www.StatSoft.com) to contact one of our offices around the world