Visual programming for R

Anup Parikh (anup@red-r.org)
Kyle Covington (kyle@red-r.org)
Red-R Motivation

• Hide the code complexity and improve readability

• Create a more interactive platform for data exploration

• Improve data and analysis sharing between users

• Provide a community repository of analysis pipelines
Architecture

Red-R

- Visual Programming
- Analysis interpretability and sharing
- Interactive Graphics

Qt UI Framework

Python

RPy

R Graphics And User interfaces
Red-R Motivation

• Hide the code complexity and improve readability

• Create a more interactive platform for data exploration

• Improve data and analysis sharing between users

• Provide a community repository of analysis pipelines
Visual Programming

• Visual programming interface
  – Analysis is performed by linking a series of widgets together

• Widgets correspond to R function
  – Read, manipulate or visualize data
R vs. Red-R

define `diff_data` as the data frame read from a CSV file.
define `filtered` as the subset of `diff_data` where the `limma.pval.KAWvsBSW` is less than 0.05, the absolute value of `FC.KAWvsBSW` is greater than or equal to 2, and the `samplesWithExpression` count is greater than 0.
write the `filtered` data frame to a new CSV file.
Red-R Overview

Canvas

Widget

All Widget

Widget Suggestions
Widget
Widget
Red-R Motivation

• Hide the code complexity and improve readability

• Create a more interactive platform for data exploration

• Improve data and analysis sharing between users

• Provide a community repository of analysis pipelines
Creating a Workflow
Interactive Widgets
Red-R Motivation

- Hide the code complexity and improve readability
- Create a more interactive platform for data exploration
- Improve data and analysis sharing between users
- Provide a community repository of analysis pipelines
Data Sharing
Data Sharing

One Shareable File
Workflow Parameters Outputs Notes

R
Import Existing R Sessions
Red-R Motivation

• Hide the code complexity and improve readability

• Create a more interactive platform for data exploration

• Improve data and analysis sharing between users

• Provide a community repository of analysis pipelines
Community Repository: Packages

### Available Packages

<table>
<thead>
<tr>
<th>Package</th>
<th>Stability</th>
<th>Summary</th>
<th>Version</th>
</tr>
</thead>
<tbody>
<tr>
<td>stats</td>
<td>Good</td>
<td>Base statistical widgets and signals.</td>
<td>1</td>
</tr>
<tr>
<td>pls</td>
<td>Prototype</td>
<td>Red-R package for the R pls package.</td>
<td>1</td>
</tr>
<tr>
<td>rsqlitedataframe</td>
<td>Good but for advanced users</td>
<td>Interface for sqlite in Red-R.</td>
<td>1</td>
</tr>
<tr>
<td>affy</td>
<td>Good, some experimental</td>
<td>Affymetrix microarray analysis.</td>
<td>1.1</td>
</tr>
<tr>
<td>RedRLME4</td>
<td>Good, intermediate</td>
<td>Widget implementations of the lme4 package.</td>
<td>1</td>
</tr>
<tr>
<td>survival</td>
<td>Good</td>
<td>Widgets for making and analyzing survival data.</td>
<td>1</td>
</tr>
<tr>
<td>base</td>
<td>Good</td>
<td>Base widgets, signals, and templates.</td>
<td>1</td>
</tr>
<tr>
<td>blank</td>
<td>Good</td>
<td>A blank package for development.</td>
<td>1</td>
</tr>
</tbody>
</table>

Interface for sqlite in Red-R. This is used to interact with sqlite databases and to create them from R objects or text files. Use this package to work with large datasets and condense them into things that are manageable with R. Many of these widgets are for advanced users as they require some knowledge of sql syntax. This package is distributed with Core but may be removed with no loss to functionality. Updates may become available to allow novice users to have better access to these tools.
Community Repository: Templates

**TEMPLATES**

Templates hold a set of widgets that work together to perform a specific task.

**Neural Net Plot and Summary**
This template makes a neural net from data read into the schema using read files. The data is used to form a neural net using the Make Neural Net widget.

**Boxplot**
This template reads in data using read files and plots the data using the RedR Mstplotlib boxplot widget. Boxplots are printable and interactive. D ... (continued)

**RDataframe to Database**
Reads in a file using read files and saves it to a database. This database will be in the Red-R session by default but can be saved to other location ...

**Connect and Query**
This template connects to an SQLite database and then allows querying of that database using the Run SQLite Query widget. The output can be used as ...

**Principal Components Plotting**
This template generates Principal Components fits to data. Data should be numeric (no text). If any ...

**Spline Fit**
This template generates a Spline fit to X and Y data. X and Y data are set using the List Selector
Community Repository: Templates

TEMPLATES
Templates hold a set of widgets that can be used to form a Neural Net using the Neural Net ...
Current Functionality

**Base R functionality**
- Read/View Data
- Subsetting
  - Merge/Intersect/Filter
- Manipulations
  - Math/Apply
- Plotting
  - Interactive Scatter Plot
  - Most R plots
- Stats
  - Parametric
  - Non-Parametric

**Additional R packages**
- Bioconductor microarray analysis
- Survival analysis
- Spatial Stats
- SQLite
- ROCR – ROC Curves
- Neural Nets
- LME4
Expanding Functionality

• How do you make it easier to transition from R to Red-R?
Expanding Functionality

- How do you make it easier to transition from R to Red-R?
Expanding Functionality

How do you make it easier to transition from R to Red-R?
Highlights

• Reduced learning curve for access to R functionality

• Analysis methods easier to read and understand and share
  – Hopefully leads to analysis reproducibility

• Increase productivity with interactivity