RGtk2Extras and DanteR: Rapid GUI development for an "omics" R package

Tom Taverner[1], Ashoka Polpitiya[2], Gordon A Anderson[1], Richard D Smith[1]

[1] Biological Sciences Division, K8-98, Pacific Northwest National Laboratory, P.O. Box 999, Richland, WA 99352 [2]Translational Genomics Research Institute, Phoenix, AZ 85004

Pacific Northwest NATIONAL LABORATORY

Proudly Operated by Battelle *Since 1965*

Abstract

Proteomics research results in large, complex datasets, presenting the researcher with many detected features over hundreds of datasets. The features detected provide evidence of the presence of constituents in the samples, including peptides, proteins, and unidentified features. It can be challenging to normalize and visualize the results of a comprehensive, quantitative "omics" study due to the volume of the data and the need to interactively explore the results to detect changes and patterns. We have developed the DanteR package and some other tools to address these needs.

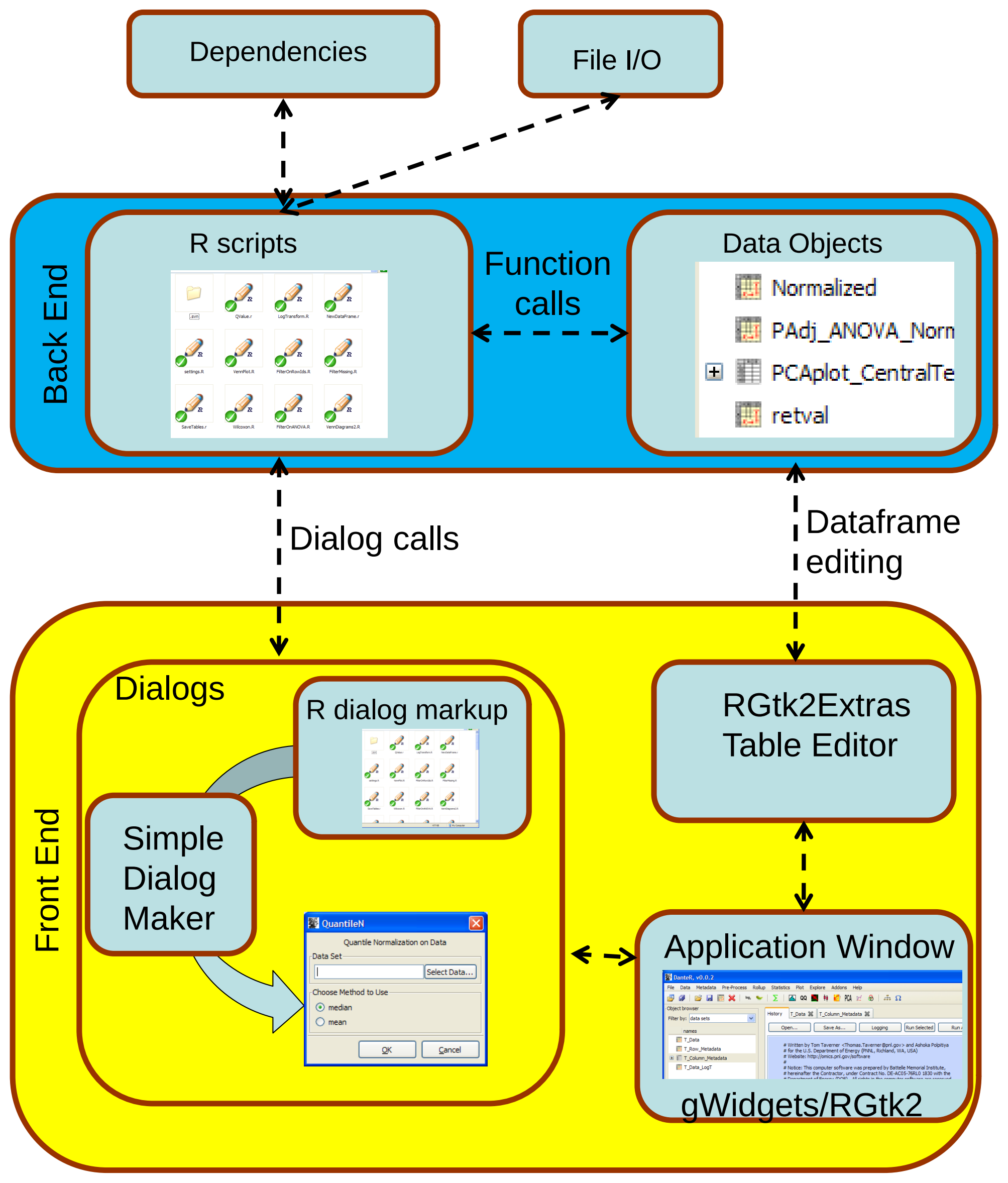
The DanteR Package

DanteR is an R package with a focus on normalization and statistical analysis of quantitative proteomics studies. Its aim is to provide scientists with an interactive, graphical interface that guides users through the preprocessing/analysis process. It is built entirely in the widely used R language and is open-source and platform independent, which are important considerations for a target audience of scientists using a variety of operating systems.

The RGtk2Extras Package

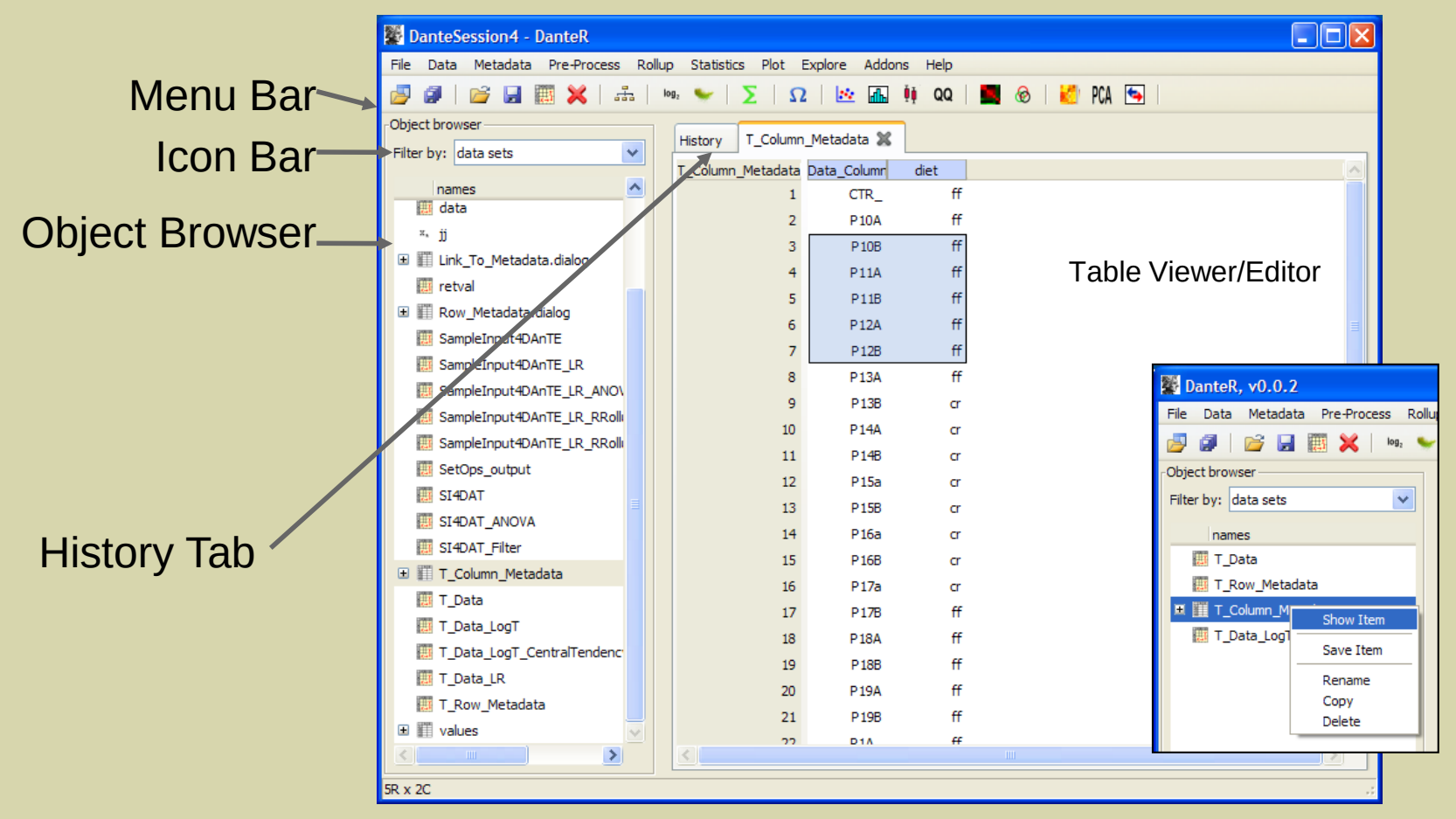
This package allows graphical front-ends for functions and interactive components to be easily specified by users with minimal R experience but no background in GUI development. The package also includes RGtk2DfEdit, an RGtk2-based spreadsheet-like data viewer and editor featuring improved features and capabilities over currently available editor widgets.

DanteR Architecture

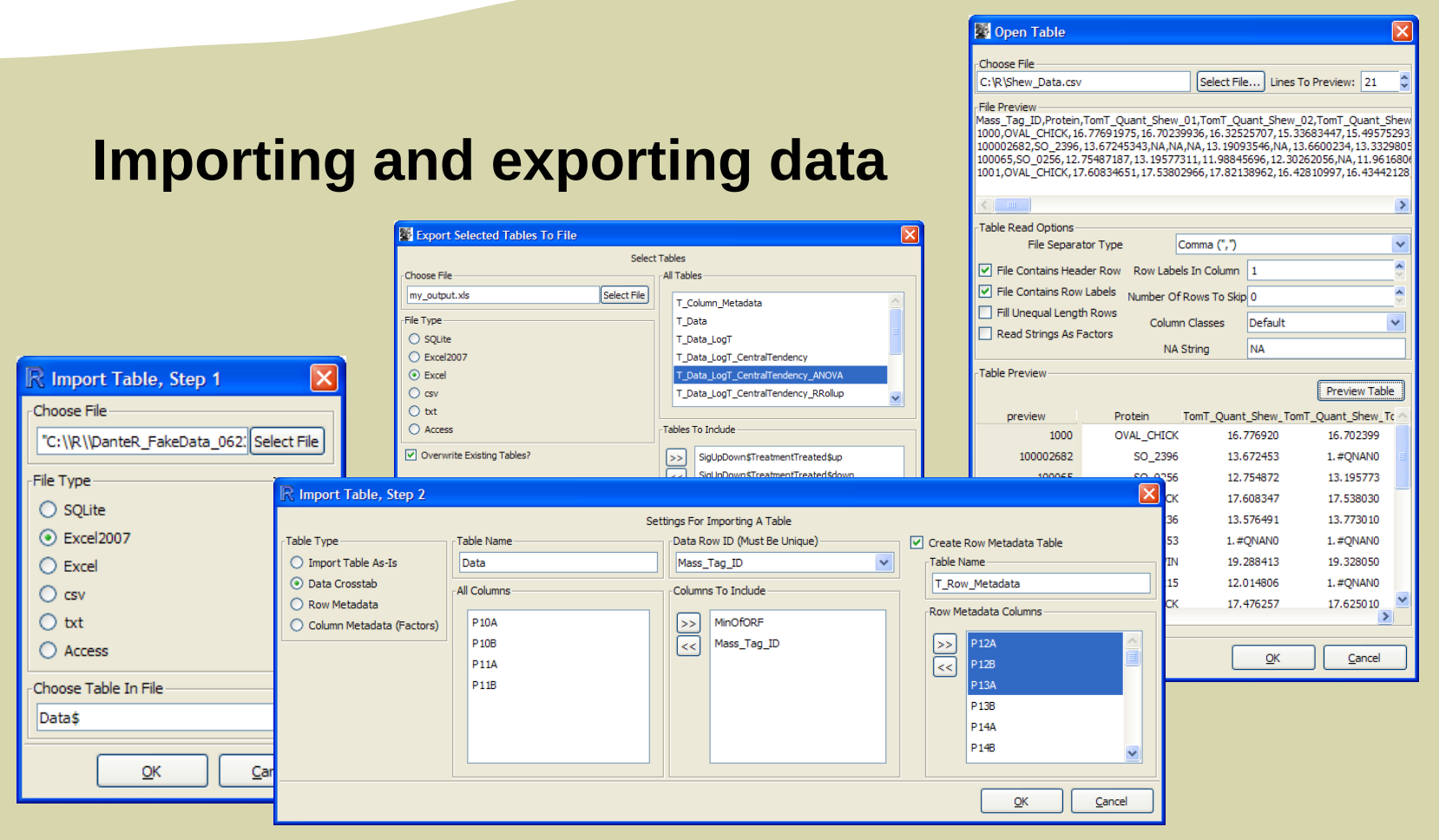


DanteR Package

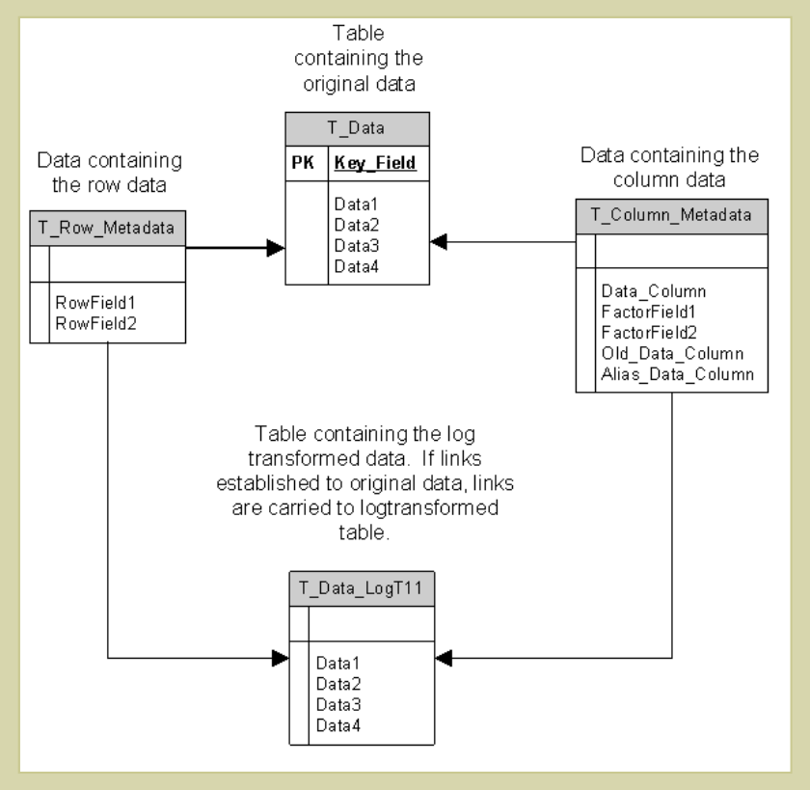
The DanteR front end



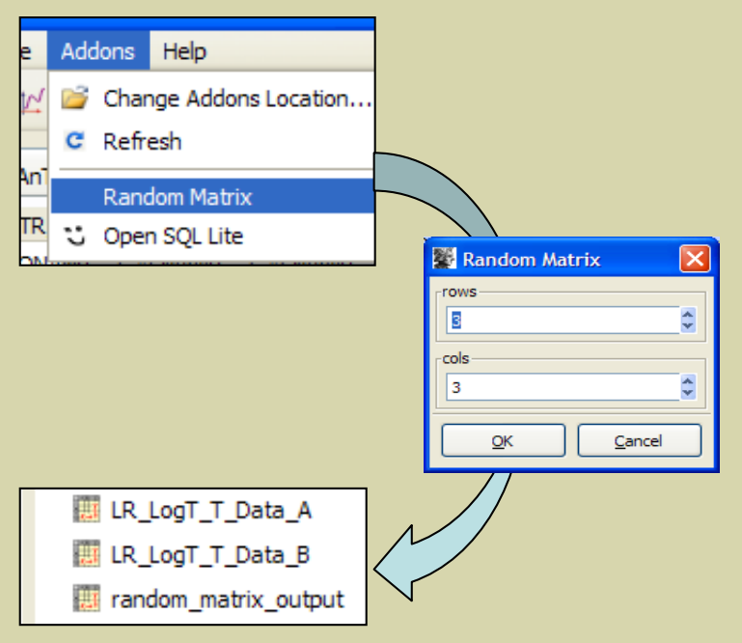
Importing and exporting data



Metadata tracking and link inheritance

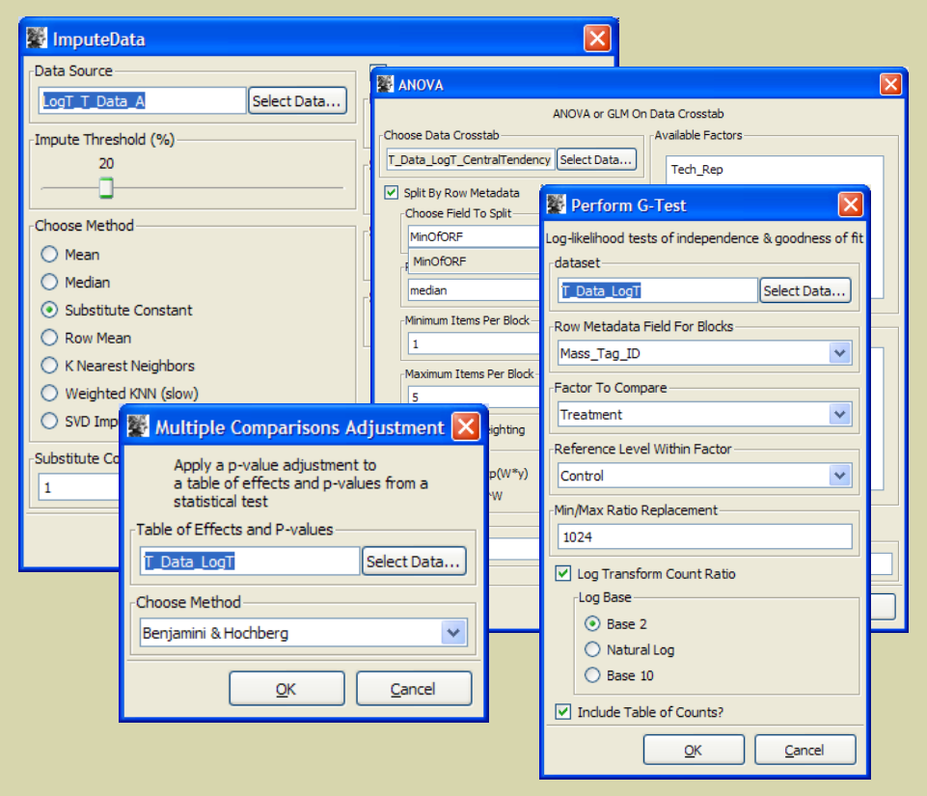


Extensible, user defined add-ons



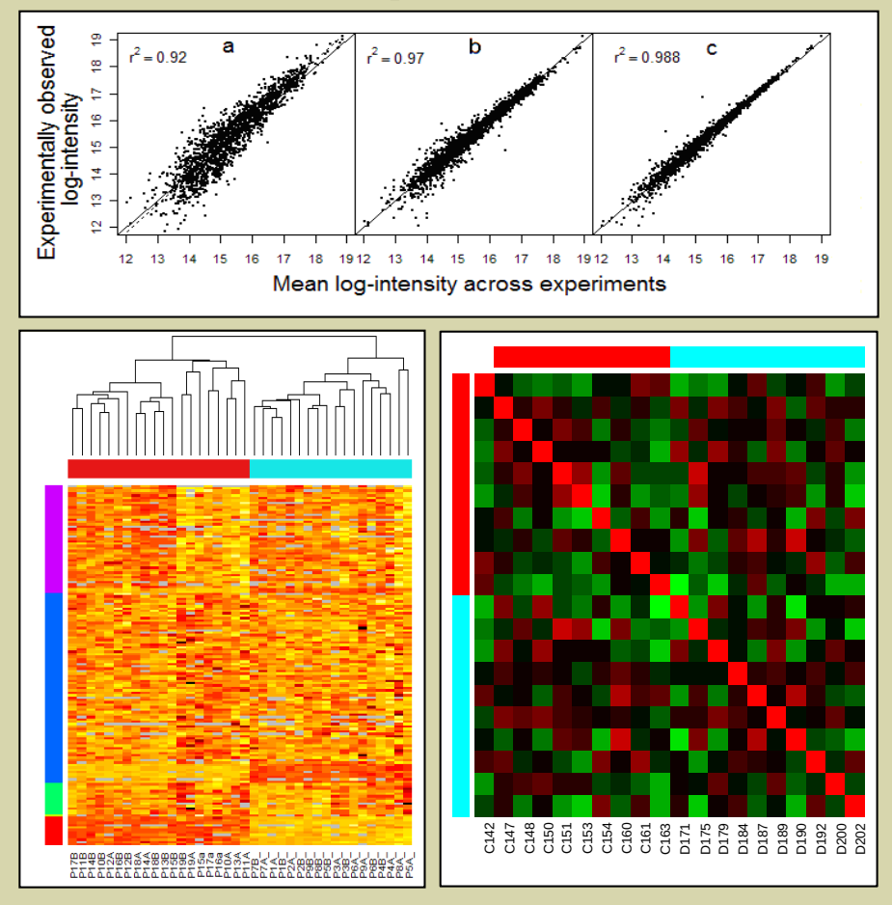
Add-ons can be refreshed to put new functionalities into the menu

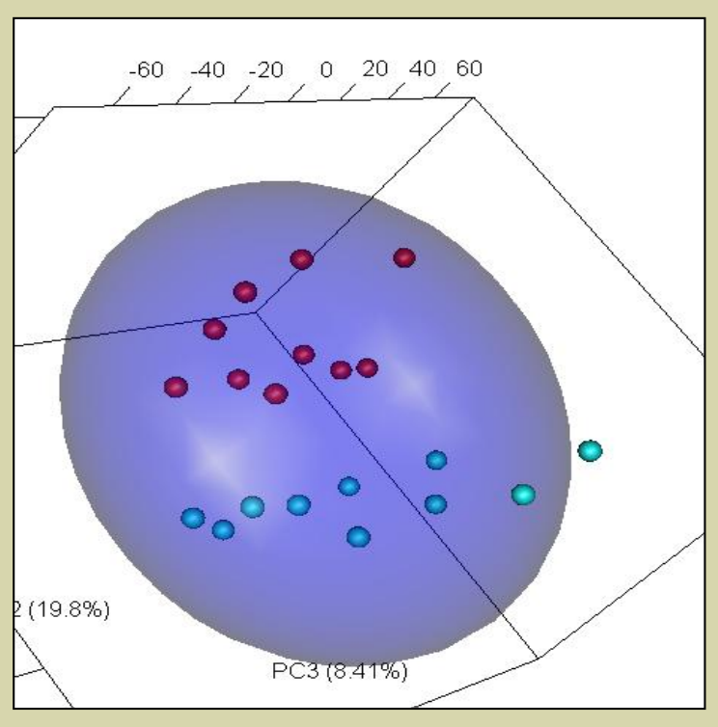
Making the power of R accessible to biologists

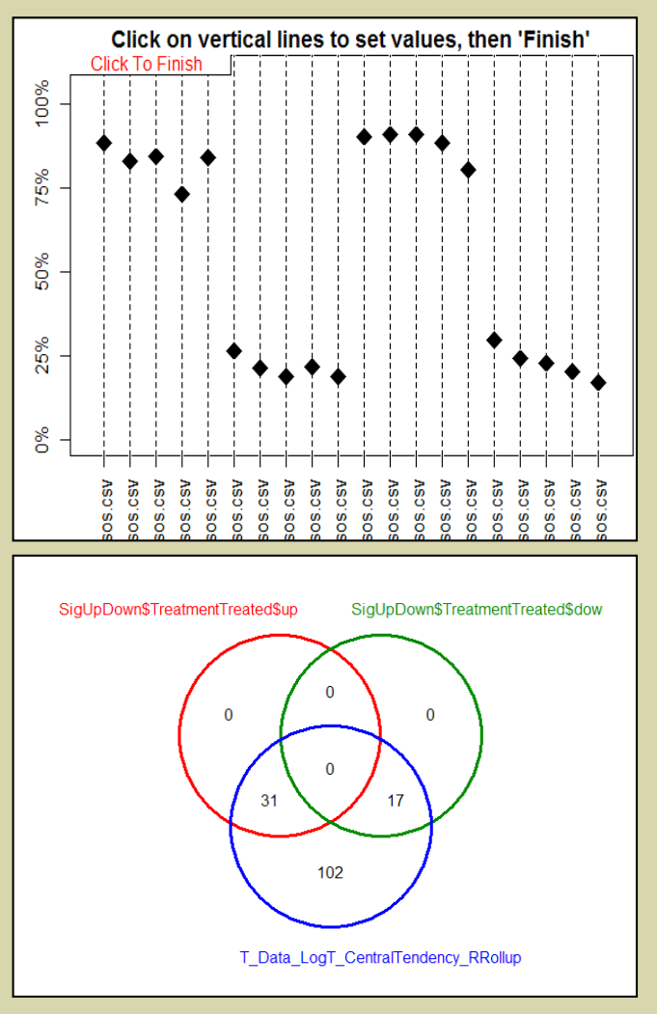


Some available statistical tools

Preprocessing, analysis and data exploration



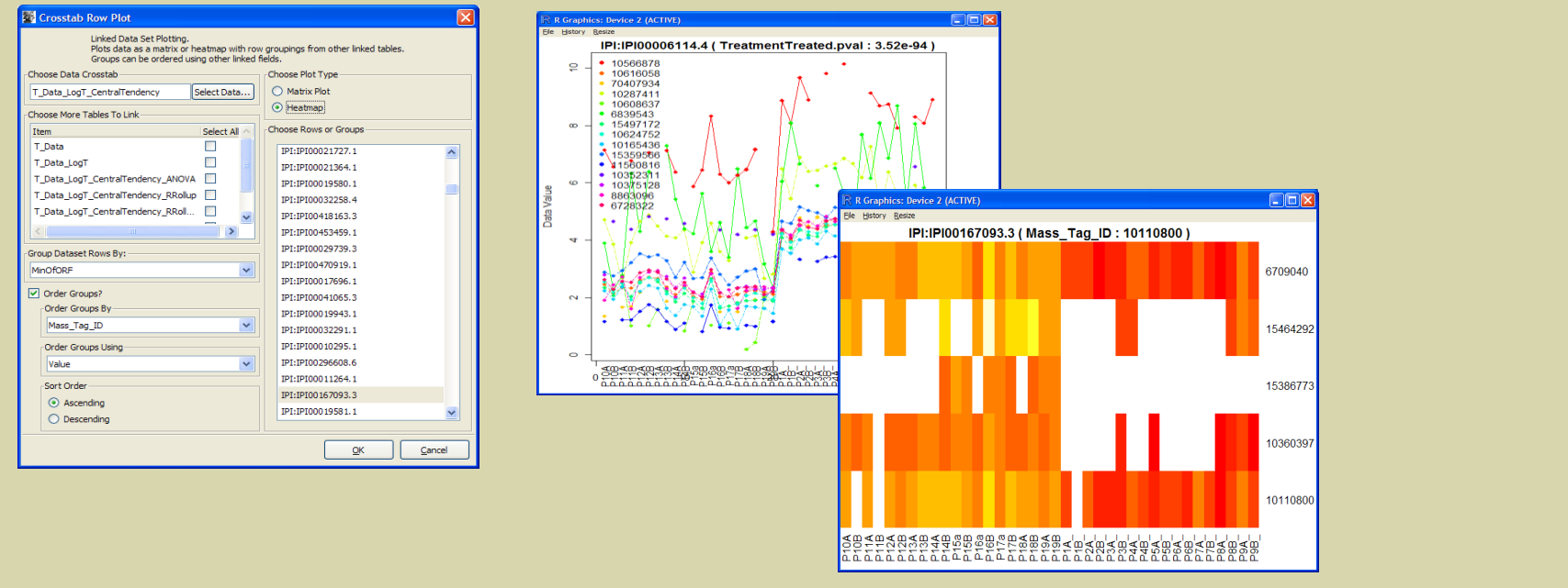




Left to right, top to bottom: Plotting data against mean, 3-D PCA analysis display, pattern search interface, clustering, heatmaps, Venn plotting

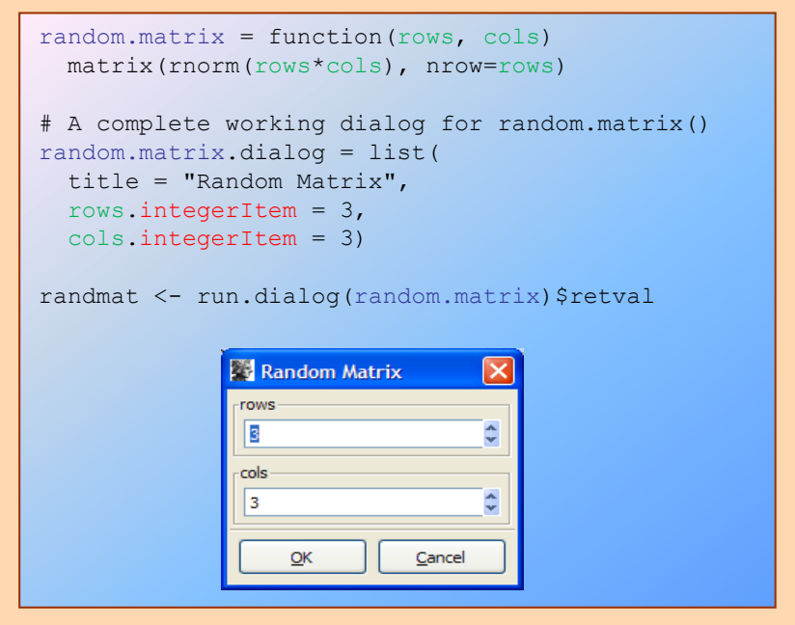
Data exploration using metadata to crosslink tables

Crosstabs with common links to metadata tables can be crosslinked and complex queries performed for data plotting and filtering.



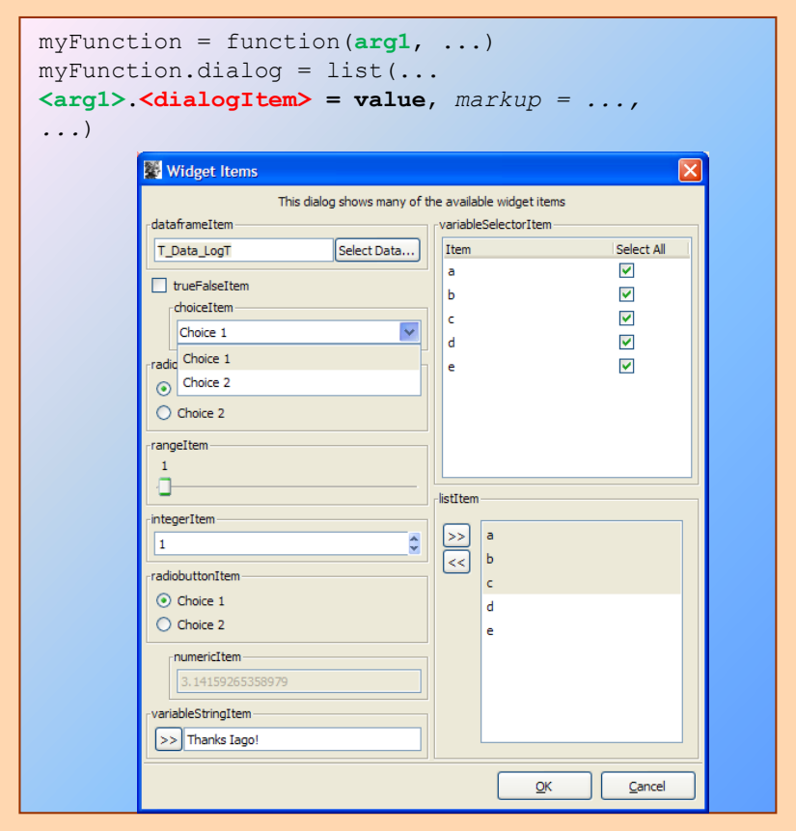
RGtk2Extras Simple Dialog Maker

Fast , functional R dialogs and wizards



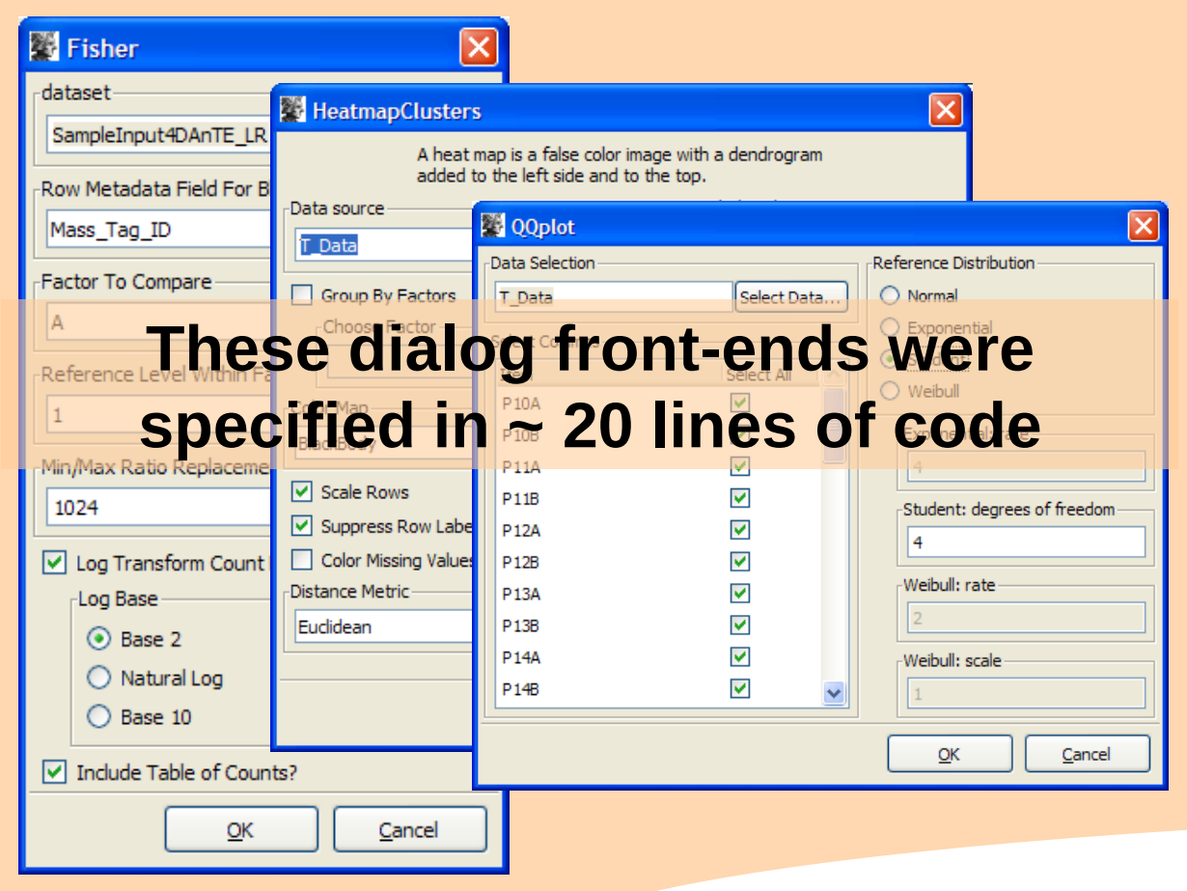
```
random.matrix = function(rows, cols)
  matrix(rnorm(rows*cols), nrow=rows)

# A complete working dialog for random.matrix()
random.matrix.dialog = list(
  title = "Random Matrix",
  rows.integerItem = 3,
  cols.integerItem = 3)

randmat <- run.dialog(random.matrix)$retval
```

A complete example



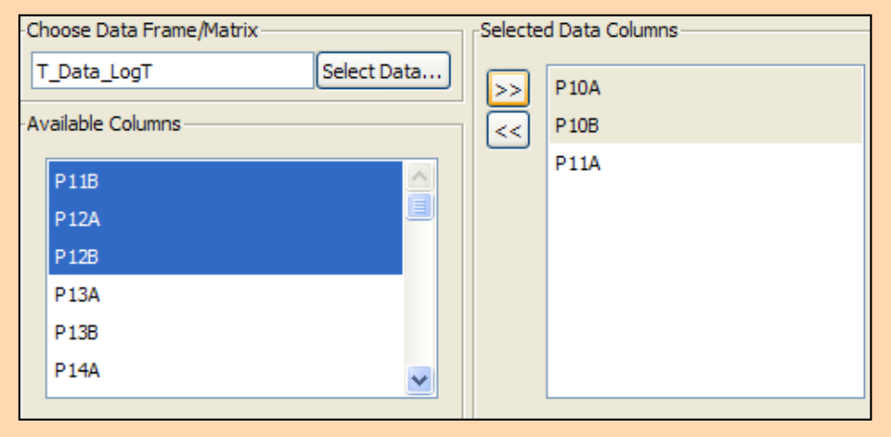
```
myFunction = function(arg1, ...)
myFunction.dialog = list(...
<arg1>.<dialogItem> = value, markup = ...,
...)
```

dialogItem markup (compare to TraitR[3])



These dialog front-ends were specified in ~ 20 lines of code

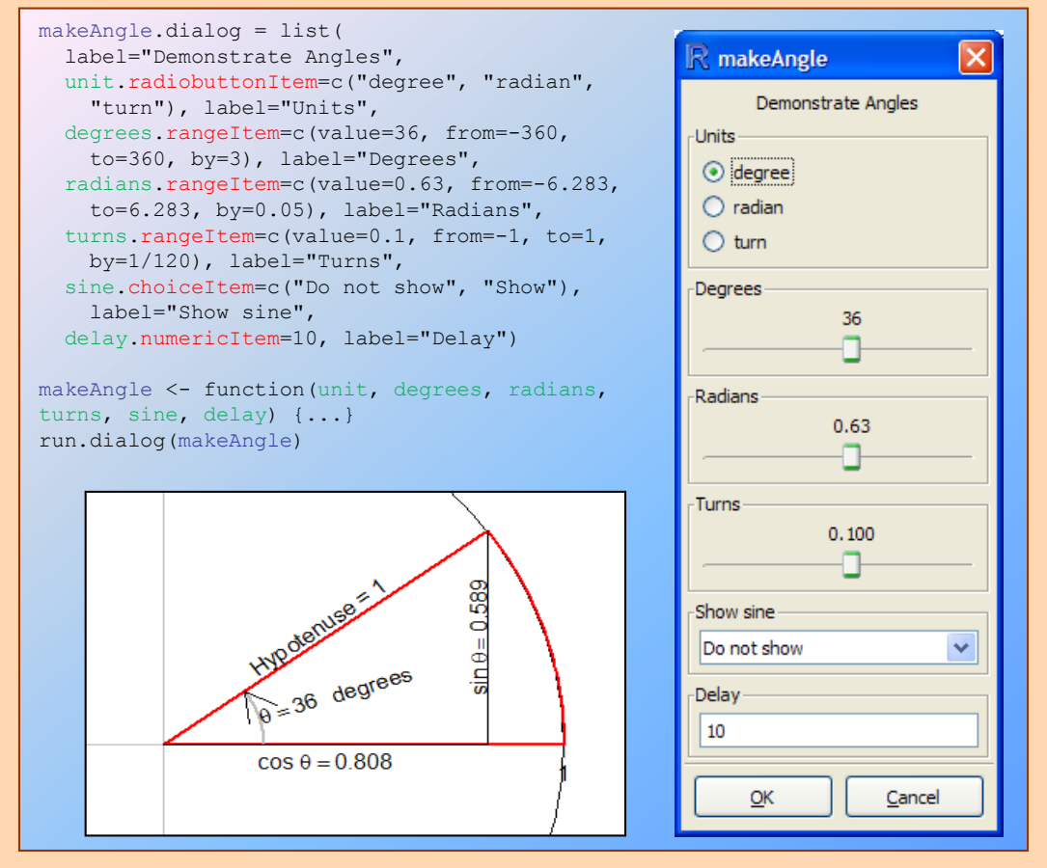
Part of the GSignal API is wrapped to allow callbacks



Additional markup to:

(i) fill a listItem widget with columns from a selected data object
(ii) push/pop selections between listItems

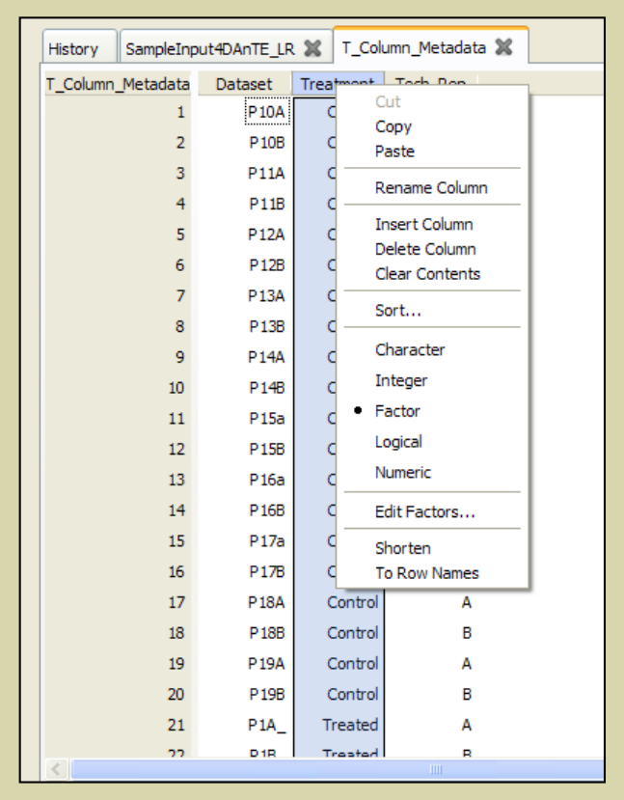
Angle demo (Graham Williams)



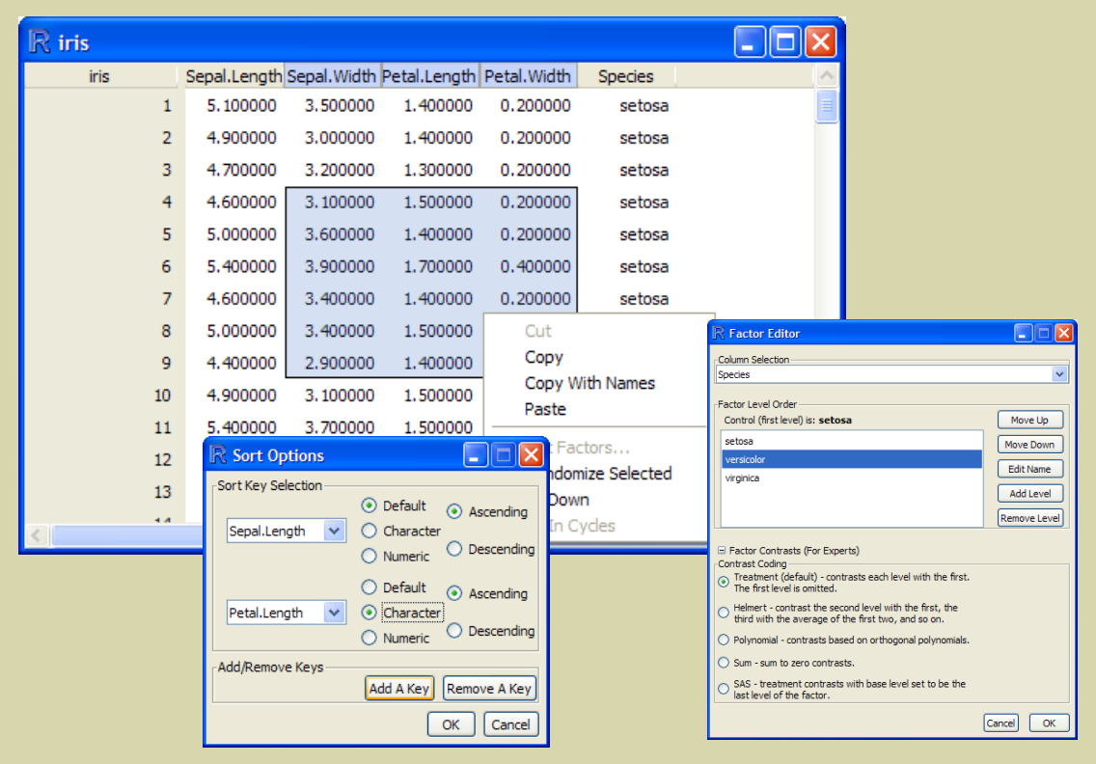
```
makeAngle.dialog = list(
  label="Demonstrate Angles",
  unit.radiobuttonItem=c("degree", "radian",
    "turn"), label="Units",
  degrees.rangeItem=c(value=36, from=-360,
    to=360, by=3), label="Degrees",
  radians.rangeItem=c(value=0.63, from=-6.283,
    to=6.283, by=0.05), label="Radians",
  turns.rangeItem=c(value=0.1, from=-1, to=1,
    by=1/120), label="Turns",
  sine.choiceItem=c("Do not show", "Show"),
    label="Show sine",
  delay.numericItem=10, label="Delay")

makeAngle <- function(unit, degrees, radians,
turns, sine, delay) {...}
run.dialog(makeAngle)
```

RGtk2Extras Table Editor

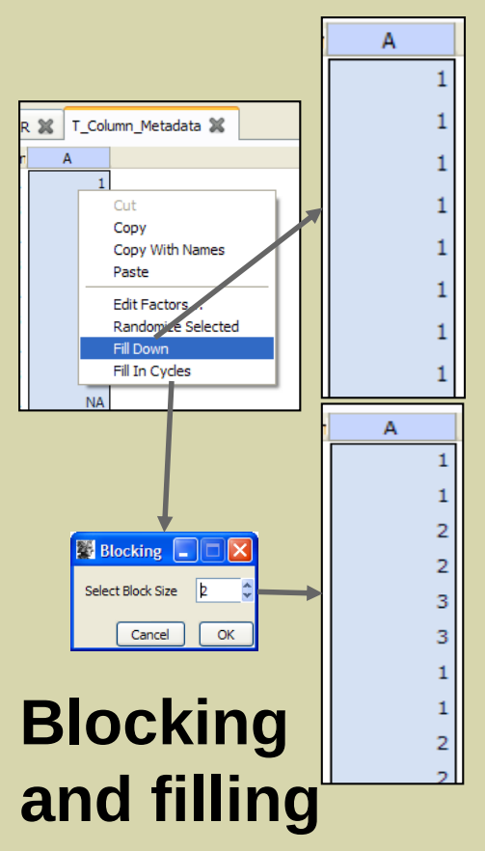
A full-featured spreadsheet



Spreadsheet-style user interaction

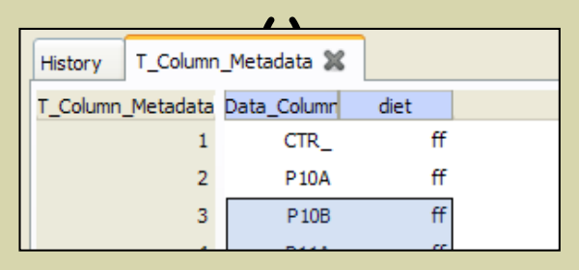


Flexible sorting Factor editing



Blocking and filling

Integration with gWidgets via `as.gWidgetsRGtk2()`



A gtkDfEdit item inside a gnotebook

Features and functionality

- Spreadsheet-style editing, keyboard handling and navigation, sort, undo, copying, pasting and selecting cells, rows and columns
- Named, floating, editable row and column labels
- Backing `RGtk2DataFrame` loads big tables fast, tested on 100,000 x 100, works with 1 million rows
- Data type coercion; support for editing factor levels, order and contrasts, Randomize, Fill Down, Fill Cyclic
- S3 methods include most data-frame like operations
- Function callbacks can be set on user interactions:
  - mouse/keyboard interaction
  - cell content and table size altered
  - selection performed
  - factor attributes changed
  - coercion performed

DanteR Features and Functionality

- History tab, Object Browser, Table Display
- File I/O: Import Tables, Workflow, Open/Save Session, Export Tables
  - Compatible with Excel 2003, Excel 2007, SQLite, Access, CSV, TXT
- Data
  - Sort or filter by data field or linked metadata field, merge data
  - Sort data, edit factors, apply command to selection
- Metadata: Define Column Metadata, Link Metadata, Plot Links
  - Data Aliasing and Reversible Renaming
- Pre-processing: Linear Regression, LOESS, Quantile Regression, Central Tendency, Imputation
  - EigenMS routine[1] for advanced model-based imputation
- Statistics: ANOVA and GLM, Fold Changes, Fisher Test, Nonparametric Tests, G-Test, Post-hoc Adjustment, Filter Statistically Significant
- Plotting: Scatter Plot, 3D Plot, Boxplots, Histogram, QQ Plot, Venn Diagram, many more
- Explore: Clustering, Pattern Search, PCA/PLS, Dynamic Row Plot
  - Dynamic Row Plot allows on-the-fly SQL-like queries on linked data table, for example low p-values
- Dynamically updatable Addons menu using SimpleDialogMaker
- Created in close collaboration with biologists and statisticians
- Extensive help and walkthroughs

References

1. Karpievitch, Y. V.; Taverner, T.; Adkins, J. N.; Callister, S. J.; Anderson, G. A.; Smith, R. D.; Dabney, A. R. (2009) *Bioinformatics 25*, 2573.
2. Michael Lawrence and Duncan Temple Lang (2009). RGtk2: R bindings for Gtk 2.8.0 and above. R package version 2.12.15. http://CRAN.R-project.org/package=RGtk2
3. John Verzani. Based on the iwidgets code of Simon Urbanek, suggestions by Simon Urbanek, Philippe Grosjean and Michael Lawrence (2009). gWidgets: gWidgets API for building toolkit-independent, interactive GUIs. R package version 0.0-39. http://CRAN.R-project.org/package=gWidgets
4. R Development Core Team (2009). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org.

Contact: Thomas.Taverner@pnl.gov

Conclusions and Future Work

We have created R functions for GUI and dialog development which we have found to be useful in the rapid development of a proteomics-related R package and front end, DanteR. This includes a toolkit for simple, rapid dialog development and a data table editor, which are included in the RGtk2Extras package.

Given the power of the R language and its broad use in the statistical community, R provides an excellent foundation to build the data integration framework. For example, one extension to DanteR we are planning is integration with components of the Bioconductor project. Integration with DanteR will allow pathway analysis for proteomic data and will make use of existing work in genomics for integrative projects.

U.S. DEPARTMENT OF ENERGY

Acknowledgements

The authors thank Alan Dabney and group (TAMU); John Verzani, Iago Conde, Graham Williams and Michael Lawrence; Joe Brown and John Sandoval; and the R community, in particular the authors of packages RGtk2, gWidgets, gWidgetsRGtk2, impute, RODBC, e1071, amap, plotrix, reshape, nlme, qvalue, rgl, gplots, car, outliers, RSQLite, scatterplot3d, colorspace, ellipse, MASS, pls, Cairo, Hmisc.

This work was funded by NIH National Center for Research Resources (RR18522) and the U.S. Department of Energy Biological and Environmental Research (DOE/BER). Significant portions of the work were performed in the Environmental Molecular Science Laboratory, a DOE/BER national scientific user facility at Pacific Northwest National Laboratory (PNNL) in Richland, Washington. PNNL is operated for the DOE by Battelle under contract DE- AC05-76RLO-1830.

www.pnl.gov