# DanteR reference slides

Tom Taverner, 9-22-10

# Contents


- Motivation
- Installation and running
- Standard data sets
- The DanteR front end
  - DanteR menu items
  - Object Browser
  - History tab
  - Table Display
- Overview of all DanteR functions
- DanteR add-ons




Pacific Northwest NATIONAL LABORATORY

Proudly Operated by Battelle Since 1965

# Motivation: Why should you use DanteR?

- You can perform a wide range of data analyses
- Features are significantly improved over DAnTE in many respects
  - Data flexibility
  - Plotting
  - Statistical functions
  - Multi platform
  - Data exploration
  - Extensibility and scriptability
- You can integrate data from different sources
- You can easily share and reproduce your analysis, algorithms and methods
- Bridges experimentalists and bioinformaticists/ statisticians



Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# What's in these slides

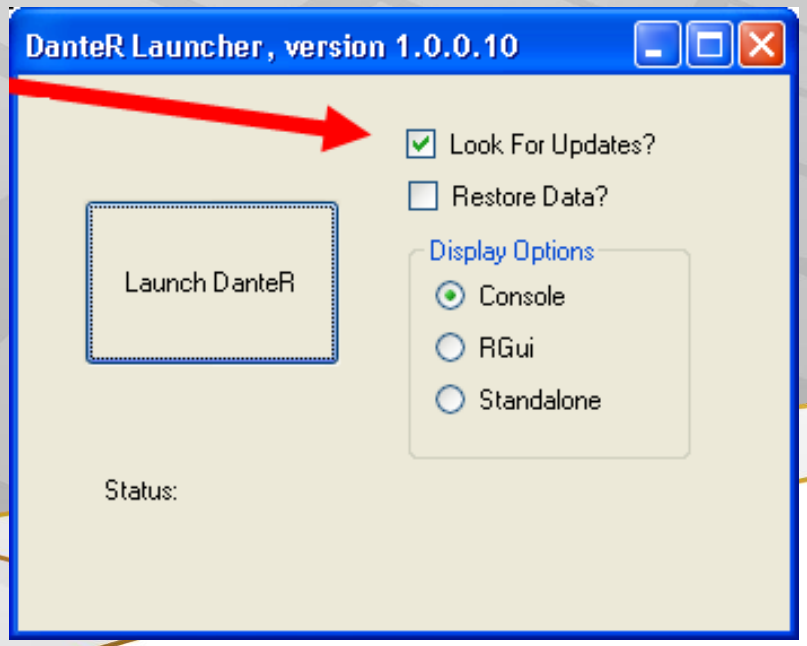
- Getting Started
- Introduction to data vs. metadata
- The DanteR front end
- Loading and setting up data
- Preprocessing and normalizing data
- Generating descriptive plots
- Performing simple statistical analyses
- Data exploration
- Saving output
- Creating your own DanteR add-ons



Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Windows setup

- 1. Install R from http://cran.r-project.org
- 2. Get the DanteR Installer from http://omics.pnl.gov/software
  - Requires .NET 2.0, most Windows XP, Vista and 7 have this already
- 3. Follow the instructions to setup and run DanteR.
- 4. Check “Look For Updates” to update any changes.



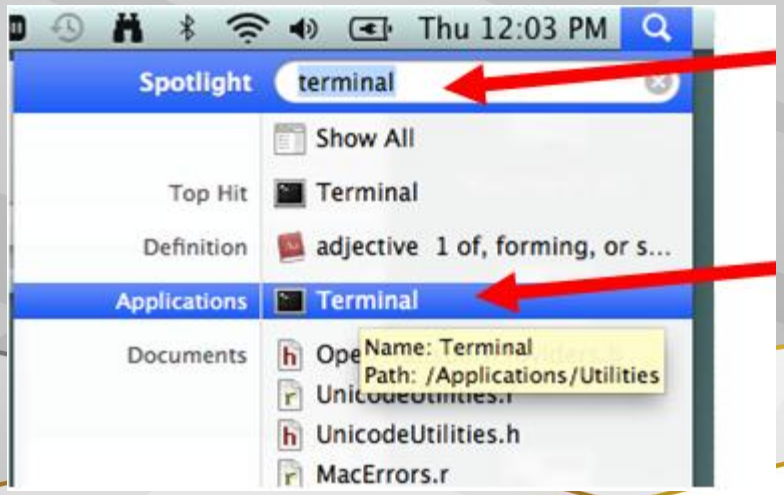


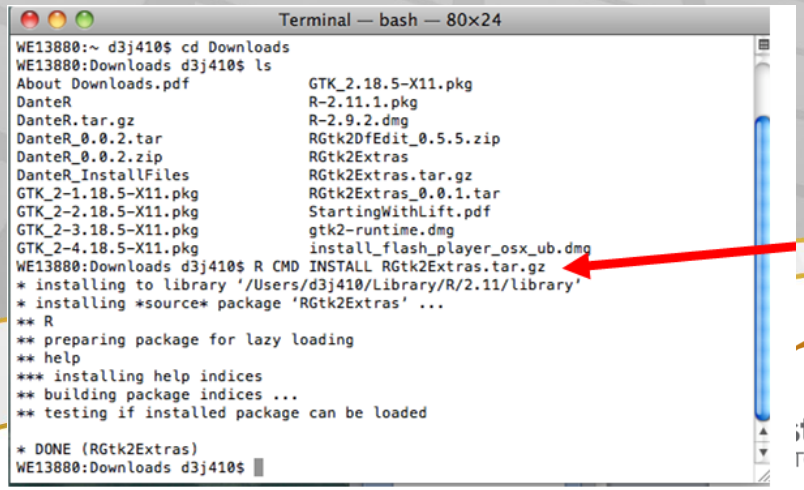

Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Mac and Linux setup

- 1. Install R (www.r-project.org)
  - (Mac Only) Install **GTK+ 2.18.5 framework** for Mac from http://r.research.att.com/
  - (Linux Only) Install `xclip` and `xsel`
- 2. Follow the instructions at http://omics.pnl.gov/software to install DanteR
- 3. To run DanteR, open the terminal (Terminal in Mac), start R and type:

```
library(DanteR); dante()
```

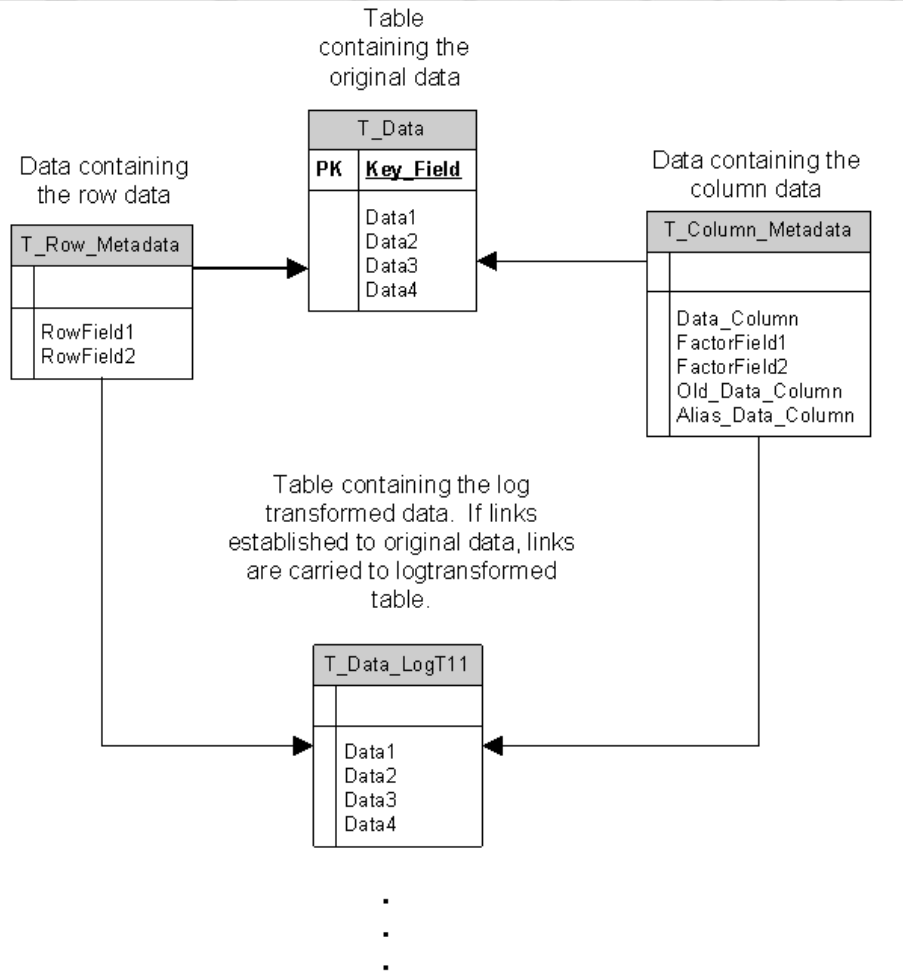





Proudly Operated by Battelle Since 1965

# Data/Metadata Model in DanteR




- Data is kept in crosstabs (rows are peptides, columns are experiments)
- We keep track of data in crosstabs by *linking* data tables to row and column metadata tables via row and column names
- Like Access (but simpler)
- Column metadata (information about datasets) is the same as factors
- Row metadata could contain peptide mass tag ID's, protein ID's, etc



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# DanteR Front End: Menu Options

- File
  - Import Tables
    - Excel/SQLite/etc
    - Nonstandard
  - Close Table
  - Load Workflow/Script
  - Open Session
    - DanteR Session
    - Old Dante Session
  - Save Session
  - Save Session As…
  - Close Session
  - Export Session
  - Quit

- Data
  - Table Display Options
  - Merge
  - Filter
  - Sort
  - Edit Factors
  - Apply Command

- Metadata
  - Define Factors
  - Define Row Metadata
  - Link Metadata
  - Plot Links
  - Create Aliases
  - Apply Aliases

- Pre-process
  - Normalization
    - Eigenvalues
    - Linear Regression
    - LOESS
    - Quantile
  - Central Tendency
  - Impute
    - Impute Missing
    - Model Based ANOVA
- Rollup
  - Rrollup
  - Zrollup
  - Qrollup



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# DanteR Options

- Statistics
  - ANOVA
  - Model Based Imputation
  - Shapiro-Wilks
  - Kruskal-Wallis
  - Fisher Test
  - P-value Adjustment
  - Volcano Plot
  - Split Significant

- Plotting
  - New Window
  - Switch Active
  - Plot Options
  - Matrix/Scatter Plot
  - Plot 3D
  - Histogram
  - QQ Plot
  - Boxplots
  - Scatter/Ellipses
  - Correlation
  - Venn Diagram
  - Row Means

- Explore
  - Clustering
  - Pattern Search
  - PCA Plot
  - Dynamic Row Plot

- Addons
  - Change Addons Location
  - Refresh

Help



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

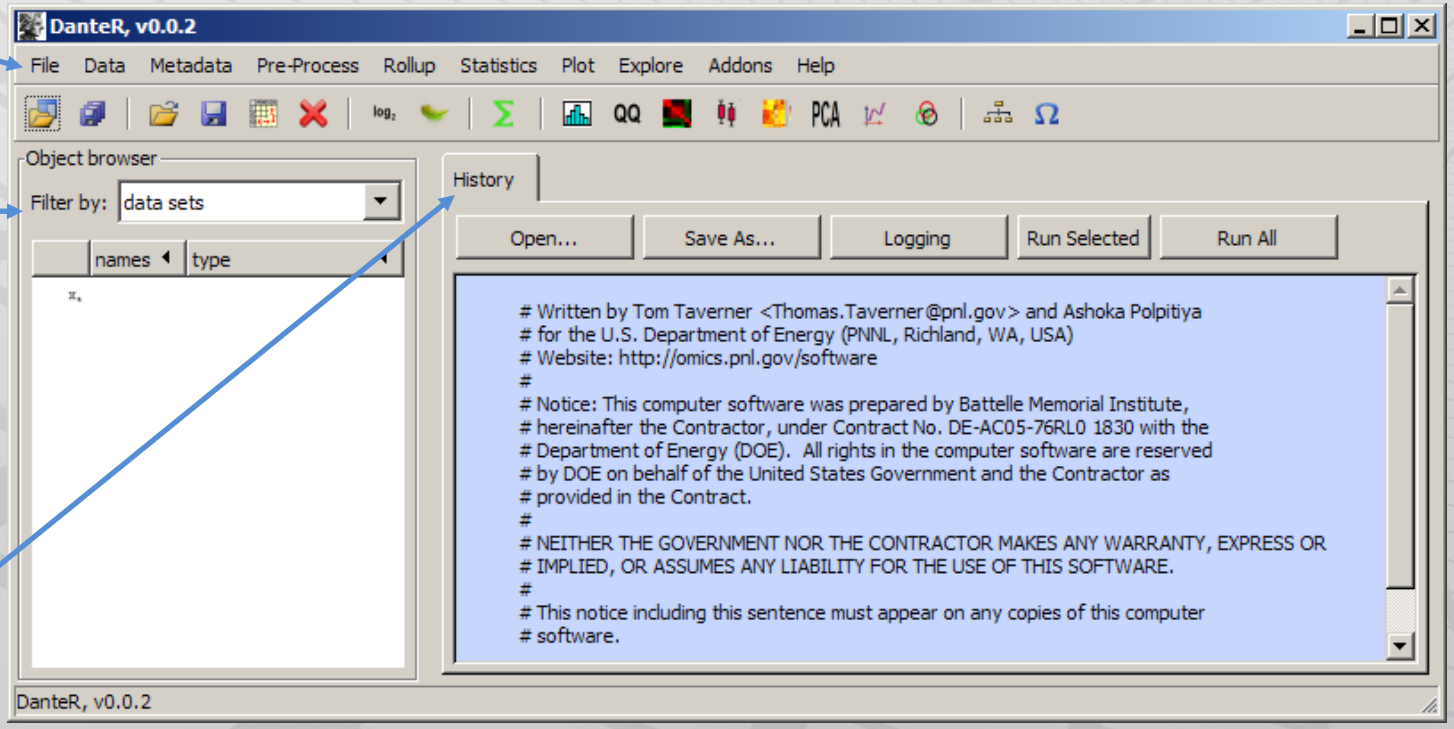
# Anatomy of the DanteR front-end

DanteR Menu Bar

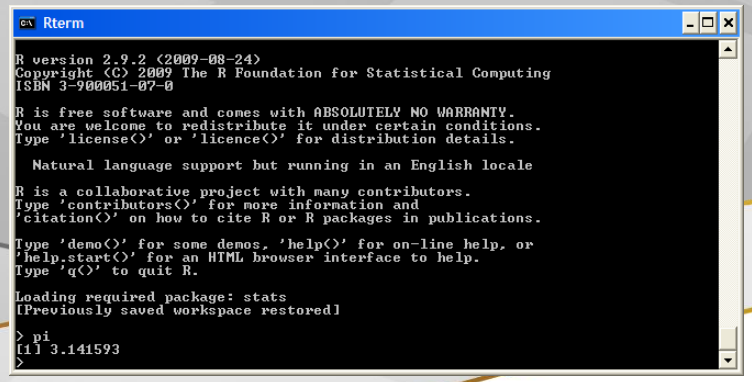
DanteR Icon Bar

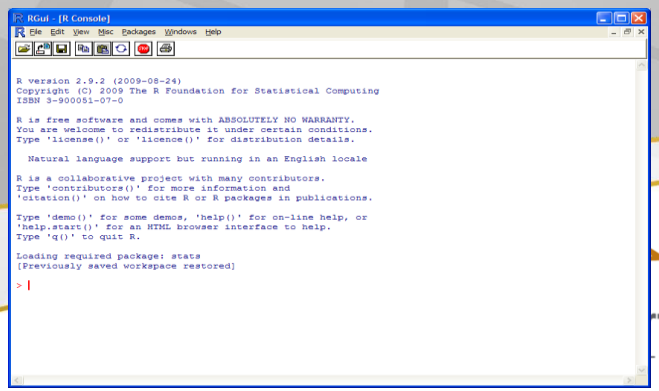
DanteR Object Browser Box

DanteR History Tab




Console Display (optional)

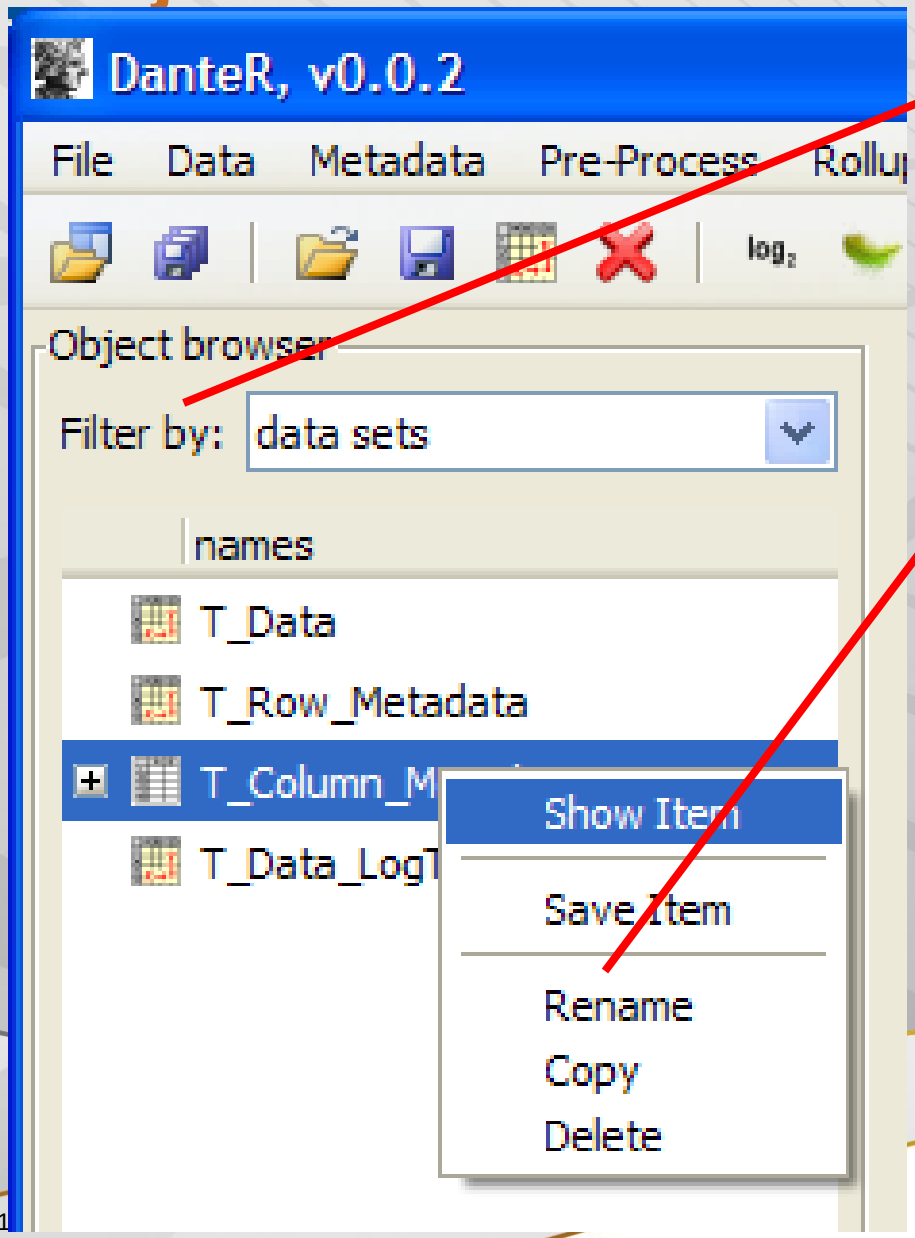





rthwest
LABORATORY
*Proudly Operated by* Battelle *Since 1965*

# Object Browser

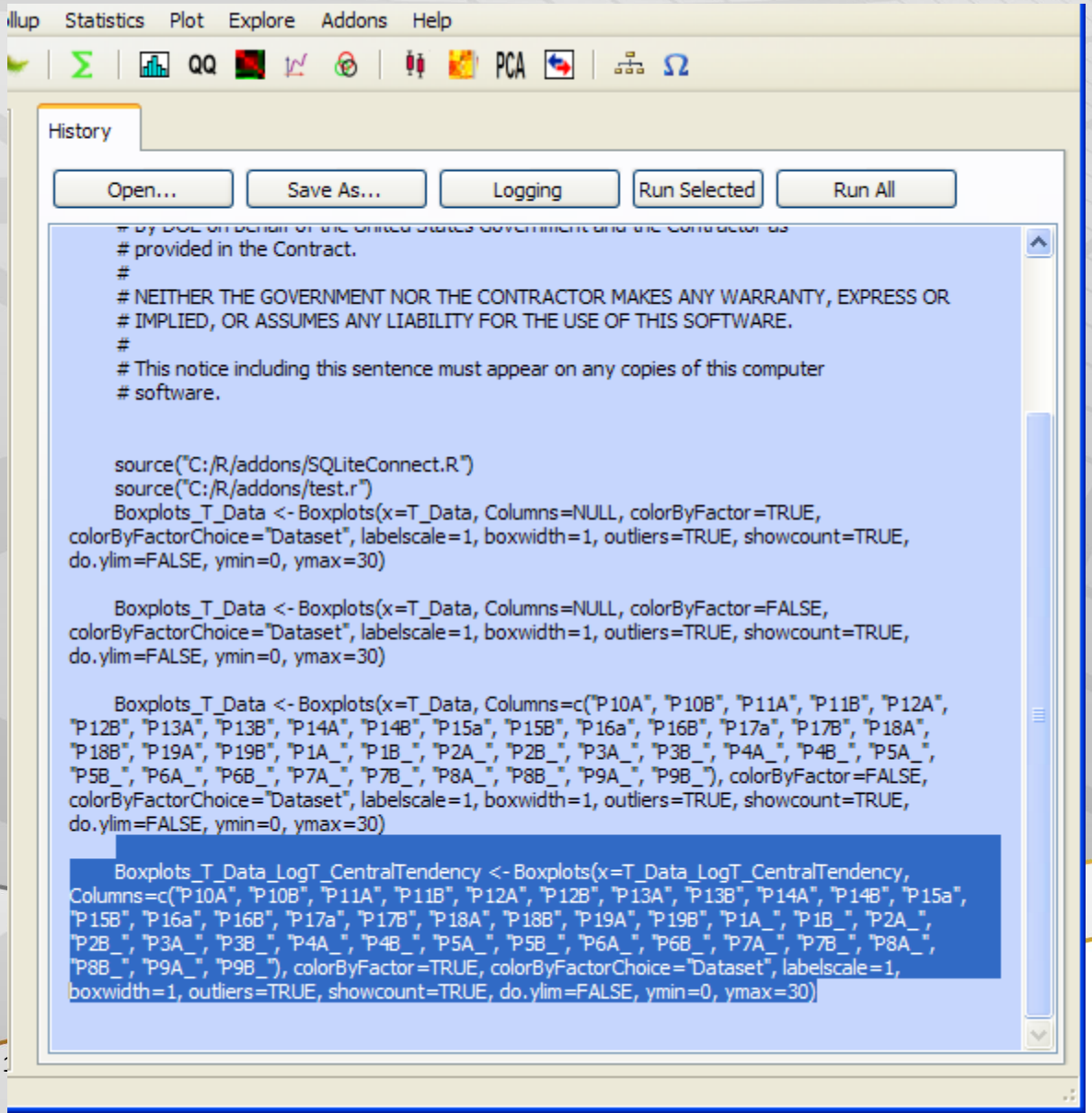



- Filter By lets you choose data sets, functions, or saved plots
- You're interested in data sets most of the time
- Right click for context menu on an item
- Show Item brings it into the table viewer
- Save, Rename, Copy, Delete do pretty much what you'd expect.



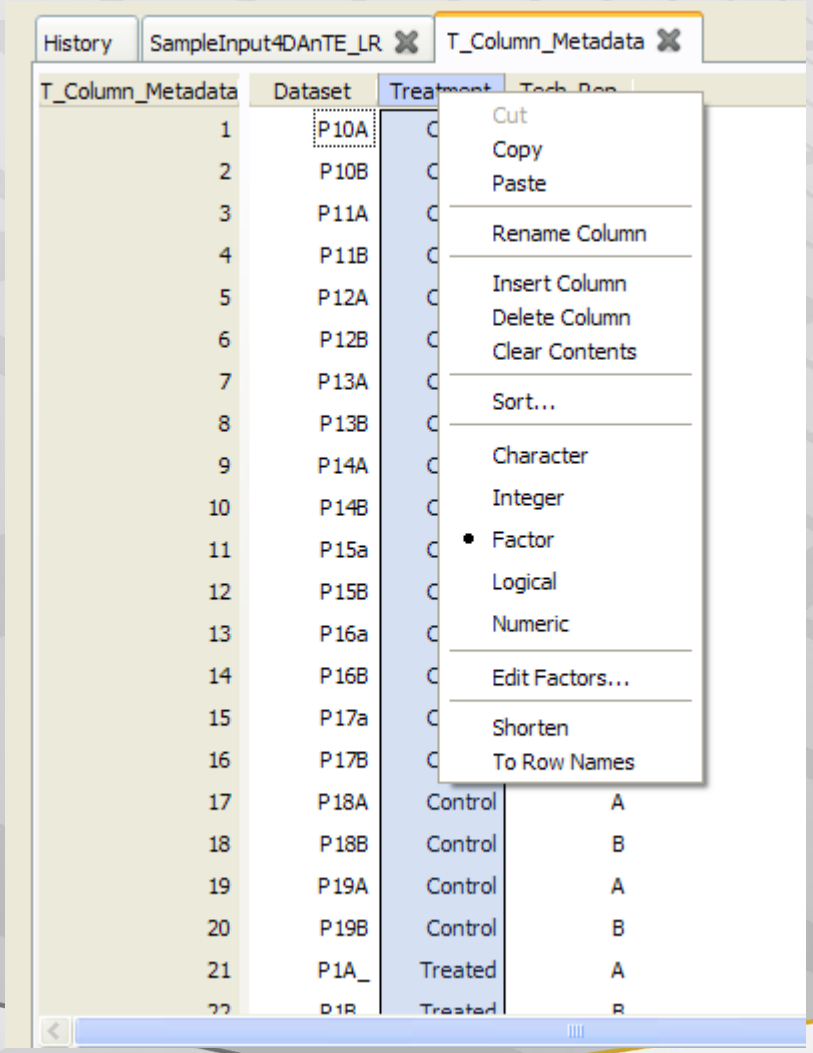
Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# History tab




- Logs all the executed DanteR commands
- Can open, save and close them as *.R files
- Can edit the commands on the window to replay or run loops, etc
- Highlight a command and click “Run Selected” to re-run

# Table display



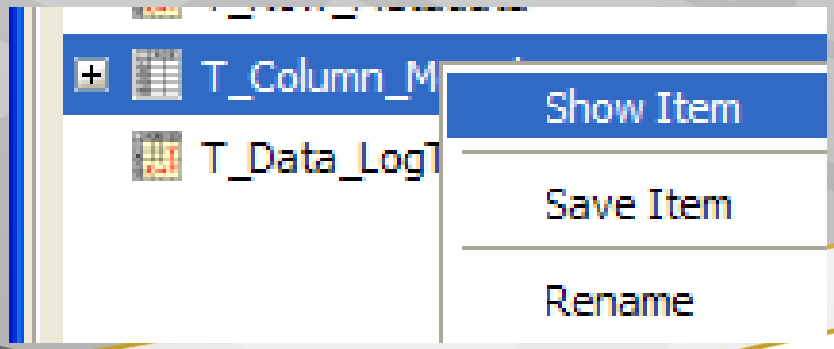

- Similar to other spreadsheets – right-click brings up context menu
- First column are row names, top corner is data set name
- Right click on column headers for
  - Set column type – numeric, factor, etc
  - Insert, Delete, Copy column
  - Rename Column header
- Right click on corner for copy/paste of entire table – useful for Excel



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# DanteR dialogs

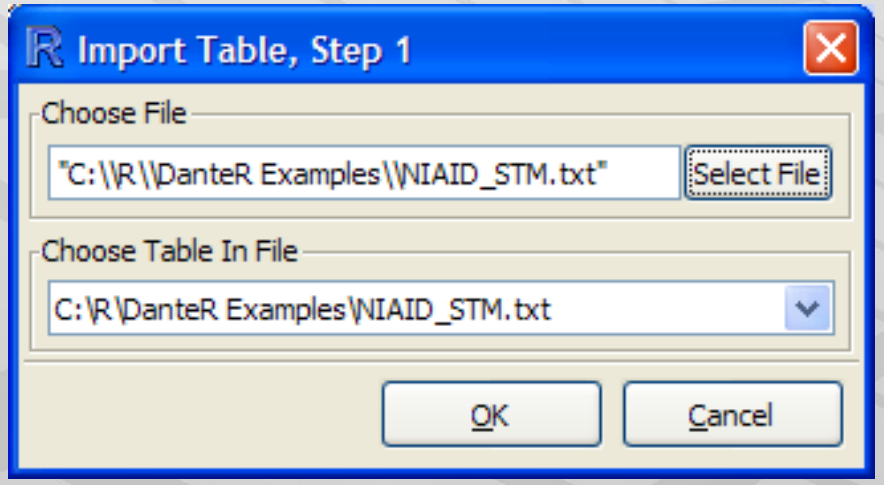
- Functions accessed via dialogs and menus try to give their output sensible names
- For example, ANOVA on MyData produces a table called ANOVA_MyData
- If you want to do ANOVA again in a different way and want to keep the old ANOVA_MyData, right-click-rename the old table to something else.






Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# File/Import Tables/ Excel/CSV/etc - 1

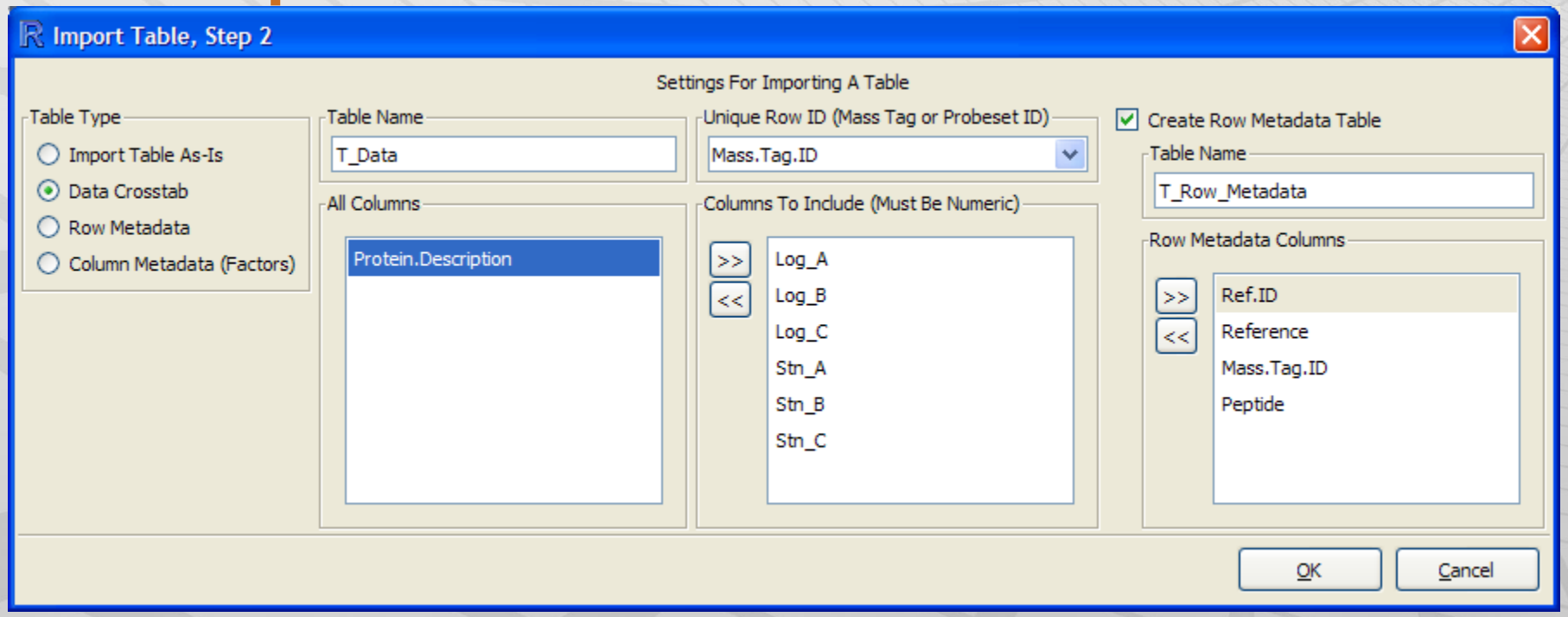



- Import Tables Wizard
- First step is choose file
- File type is auto assigned
- We assume that txt files are tab delimited text with column names
- Choose table for Excel, SQLite files



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# File/Import Tables/Excel/CSV/etc - 2

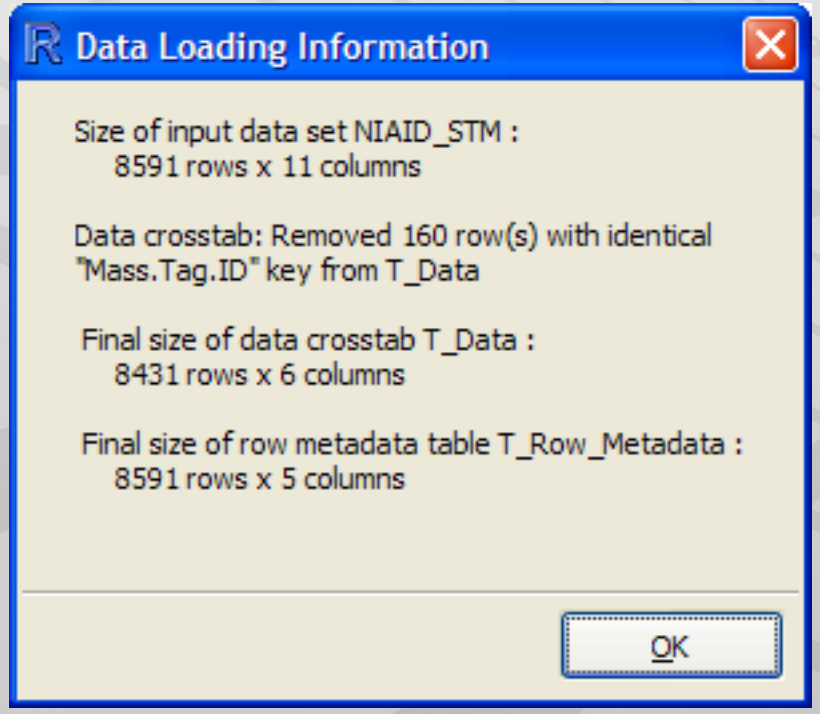



- To get numeric data into DanteR, choose “Data Crosstab”
- Move numeric fields from “All Columns” into “Columns to Include” and call your table name something short.
- Set “Data Row ID” to the unique row index (Mass Tag ID or peptide sequence). Use Row Numbers if you don’t have one.
- Put row information (proteins, pathways etc) to Row Metadata Columns



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# File/Import Tables/Excel/CSV/etc - 3

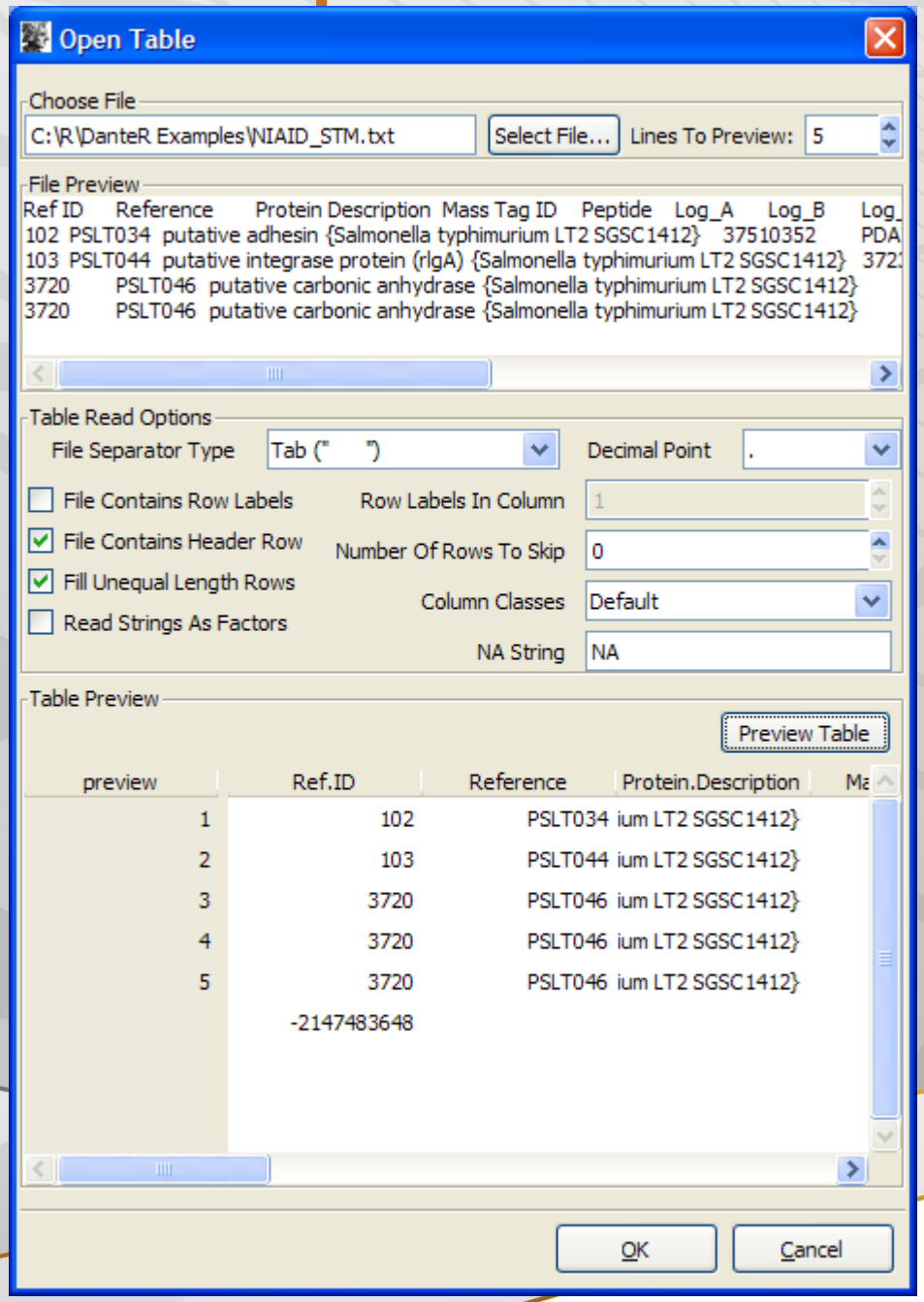



- DanteR will remove redundant information when you load data
- This means data rows with duplicated unique ID's are removed
- A "Data Loading Information" message summarizes the input and output data table sizes
- If non-duplicate rows are removed the program will warn the user



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# File/Import Tables/Nonstandard




- Import Tables makes formatting assumptions which may not be true
- For text-format tables (txt, csv) and crosstabs that are in a weird format (no column names, unusual separator characters, etc) you can try Open Table
- Load them in and hit “Preview Table” to see if R can read them as a table
- Play around with the Table Read Options till Preview works, then hit OK –try previewing more lines if there’s still a problem.

## File/Close Table

- Close the table you have selected

## File/Load Workflow/Script

- You can load a saved R workflow or script and execute it using this command



Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# File/Open Session

- R sessions contain all the objects you are working with
- Save your sessions often
- If you load a session containing MyDataTable and MyDataTable exists in your current R session, it will get written over
- DanteR can load previous old-version DAnTE sessions (.dnt) or R sessions (.RData files)
- If you load an old-version DAnTE session, DanteR will convert table names and formatting so they're compatible
- Save Session saves the current session silently, Save As renames it



Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

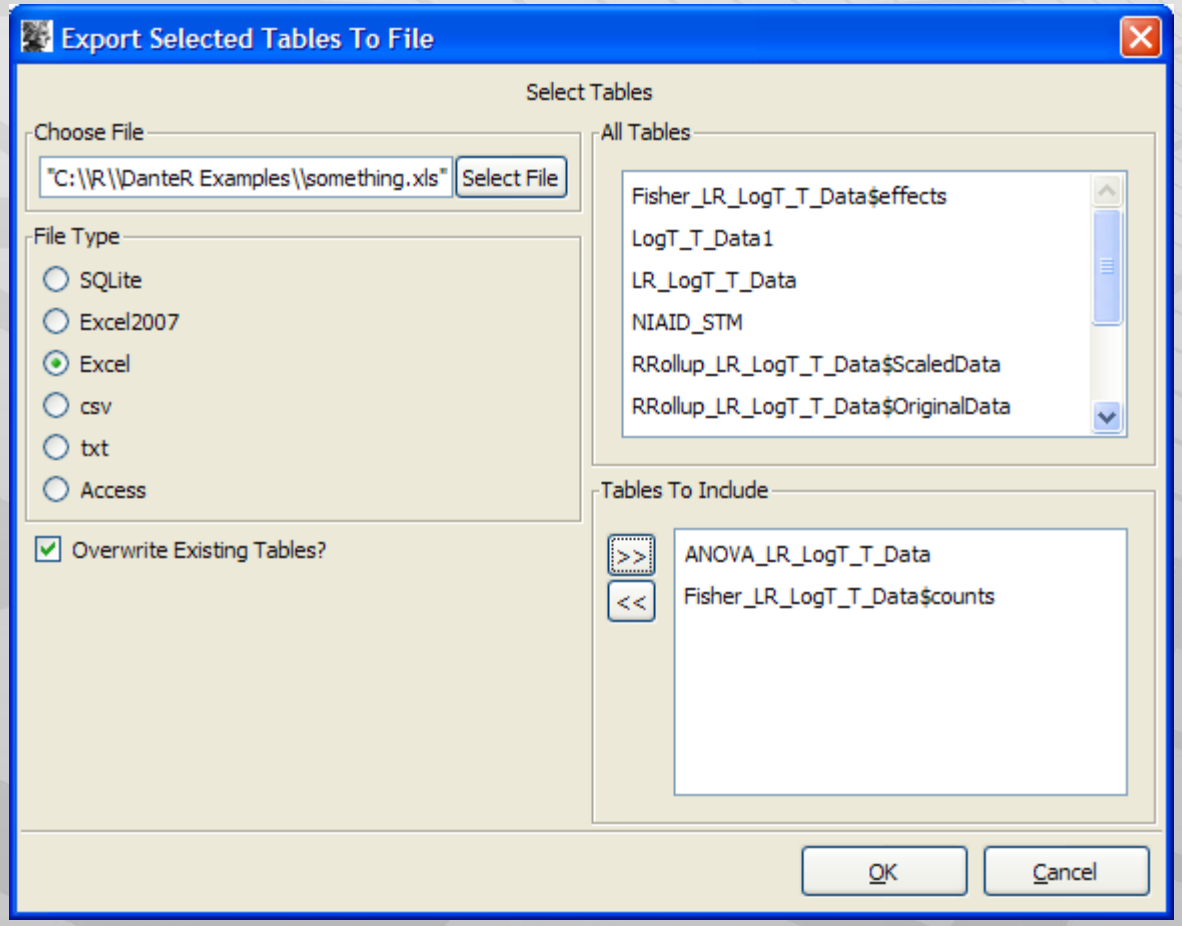
# File/Close Workspace

- Remove everything in the workspace and close any open tables in the viewer
- Won’t remove anything with a name starting with “.” (these are invisible)



Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# File/Export Tables

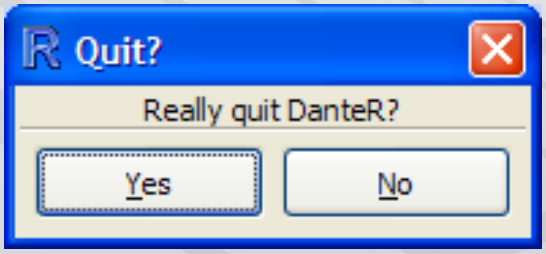


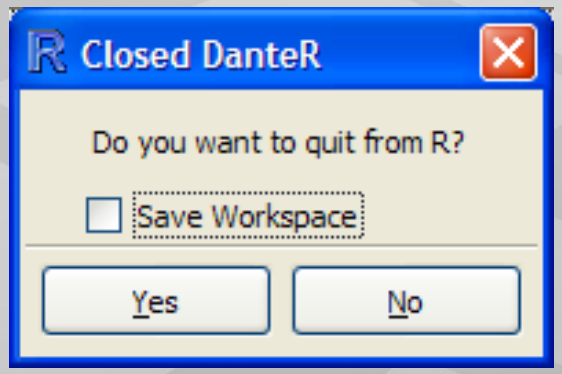

- Export tables in your session to supported formats
- Select tables you want to export
- SQLite, Excel 2007, Excel and Access will save as sheets within the exported file
- CSV and TXT will save as separate files
- For CSV/txt, files are named <filename>_<tablename>.<extension>



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# File/Quit



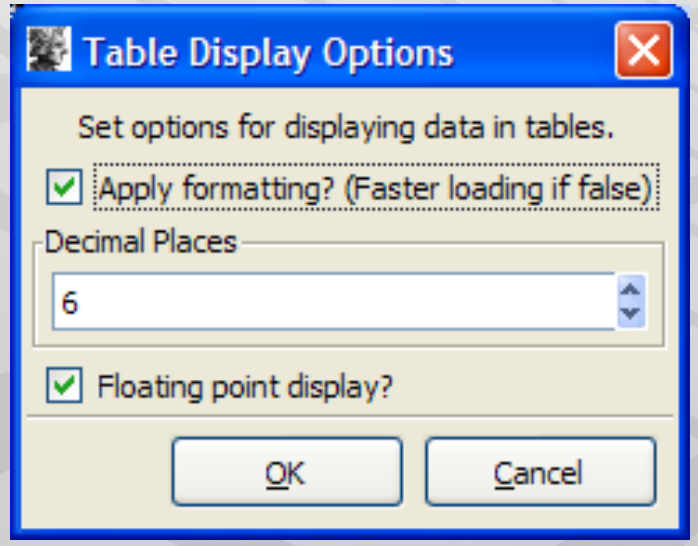




- Quit from R if you don't want R to keep running
- R will quit automatically if it isn't running in a visible window
- Save Workspace if you want to start using the same objects the next time around



Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Data/Table Display Options

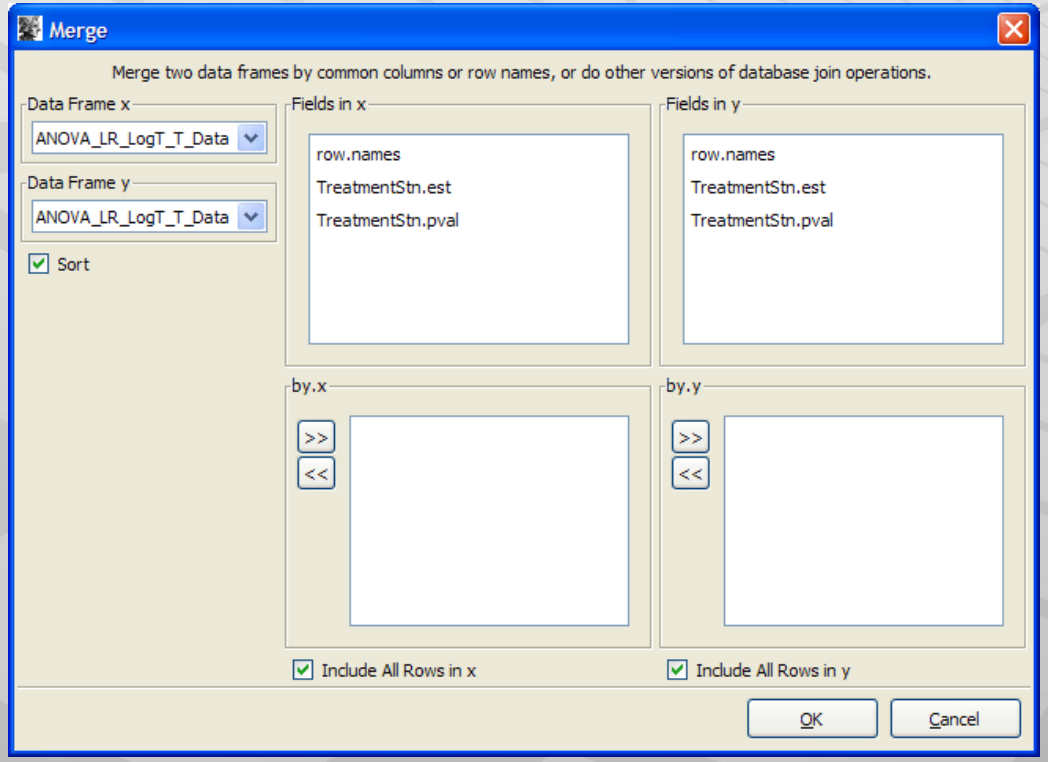



- Display of numbers in tables
- Apply Formatting: prettier table display, but a little slower to load
- Decimal Places
- Floating point display: switches between 314159.26 and 3.1415926E5



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Data/Merge

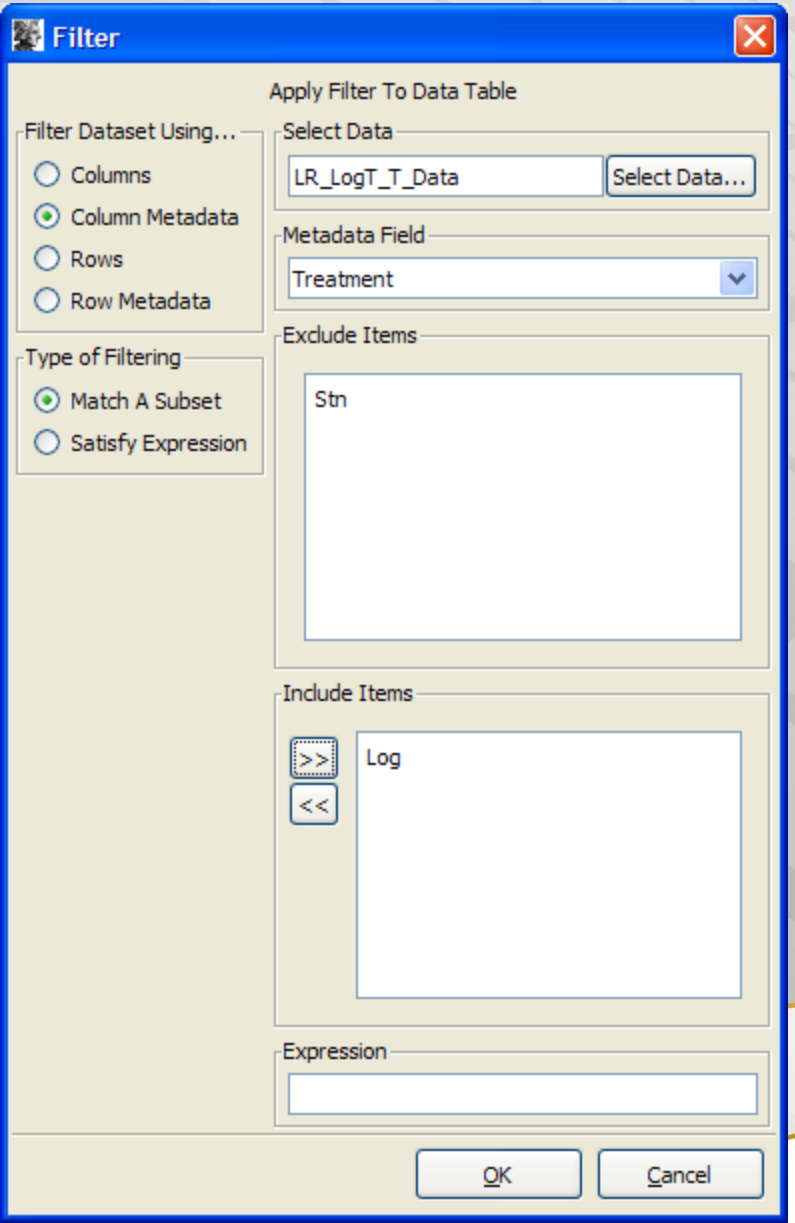



- Perform a left, right or outer join on 2 data tables on selected common columns
- Select row.names for row name merge
- See ?merge in R



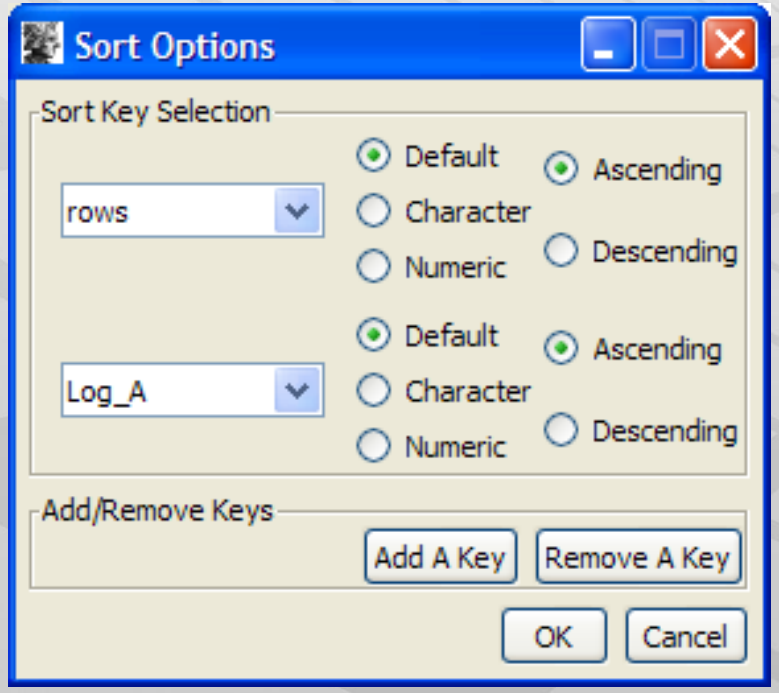
Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Data/Filter




- Filter datasets using either subsets of rows, columns, or linked metadata (see Metadata) – your table needs to be linked for metadata filtering
- Select data frame, filter type, and a subset of items to exclude
- Create a new table containing the filtered items
- For example, filtering T_Data by Column Metadata/Treatment allows only log-phase datasets to be selected.

# Data/Sort




- Must have a table selected in the viewer
- Choose a key to sort by
- Can add additional keys
- Coercion to character, number and ascending/descending sort

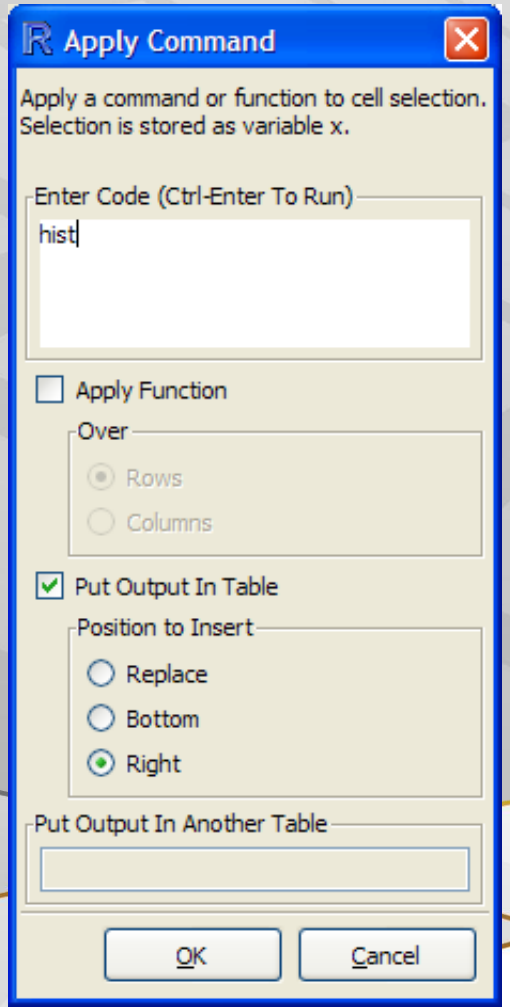


Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Data/Insert or Delete a Column

- Requires an open table

# Data/Apply Command

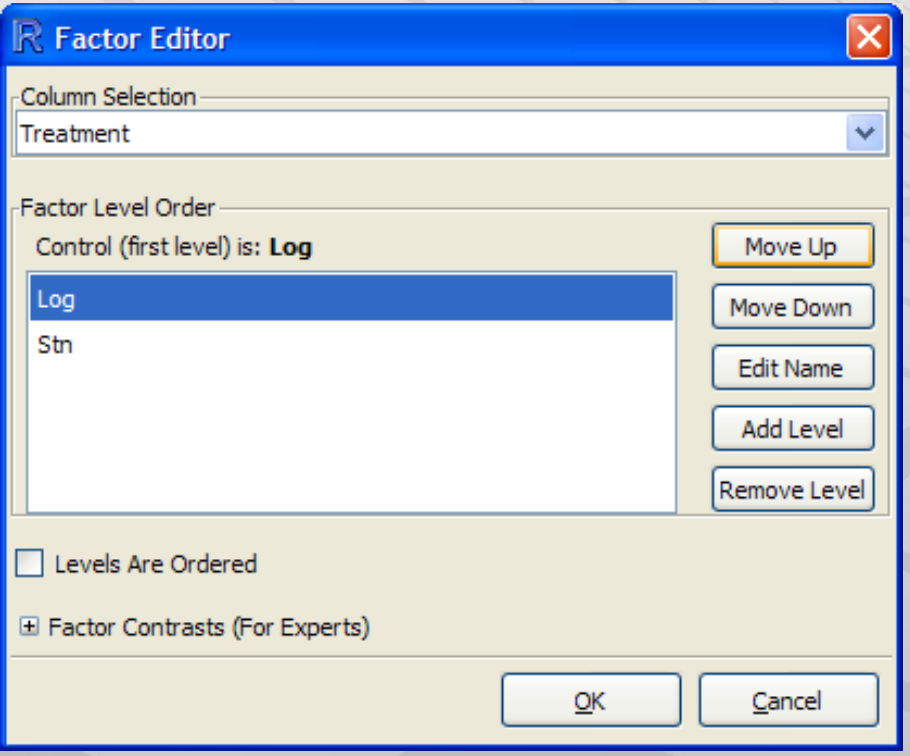



- Apply an R command to any selected data region in an open table
- For example:
  - `hist` displays a histogram
  - `hist(x, 50)` displays histogram with 50 breaks
  - `image` displays a heatmap
- Put Output In Table replaces or inserts the command output
  - `rowSums` inserts row sums to the right.



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Data/Edit Factors

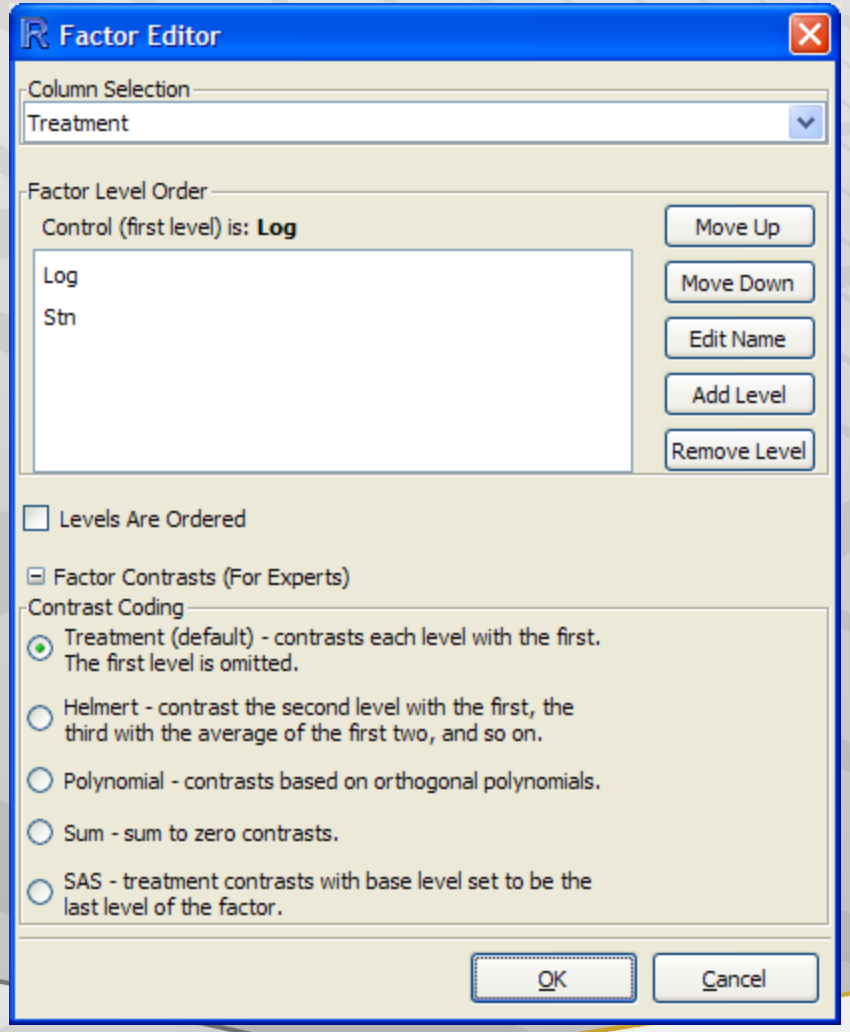



- A *factor* is a set of categories your data sets belong to for the purpose of applying statistical functions
- For example, control versus treated; 0, 30, 60, 120 minutes post treatment; etc
- You can edit factor levels and ordering, and set which factor to treat as the control
- By default, we compare all factor levels to the first.
- See RGtk2Extras help file



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Data/Edit Factors - Contrasts

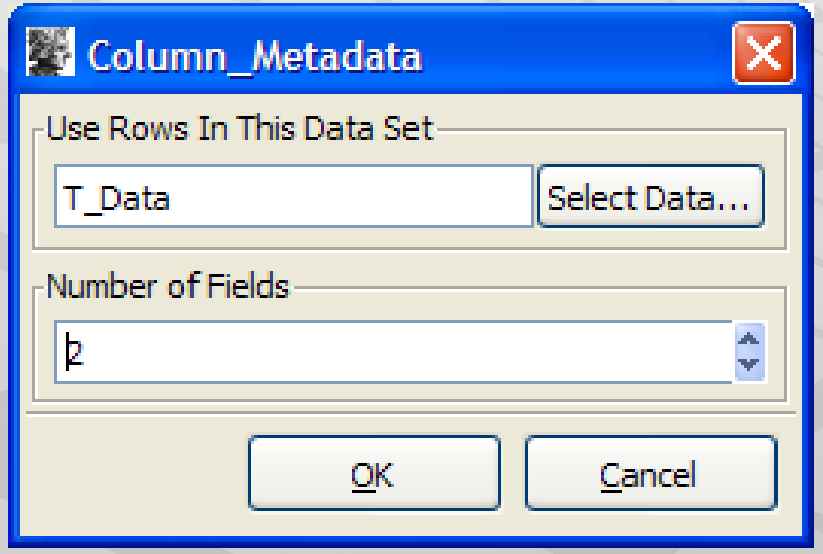



- Advanced concept
- Factor *contrasts* code how your factor is treated by statistical functions
- The default (treatment) treats the first factor level as a baseline
- Sum treats the average of all datasets as a baseline
- See RGtk2Extras help and ?contr.treatment in R



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Metadata/Define Factors

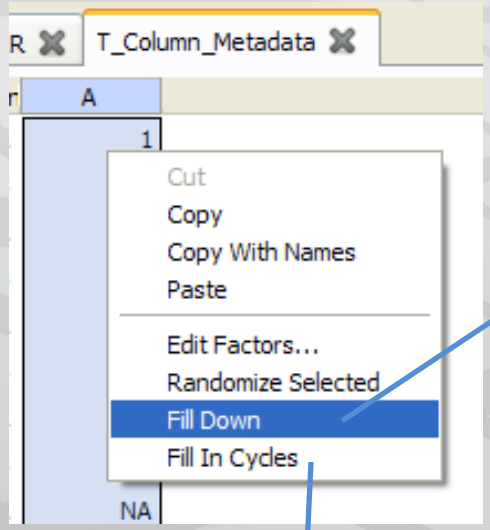


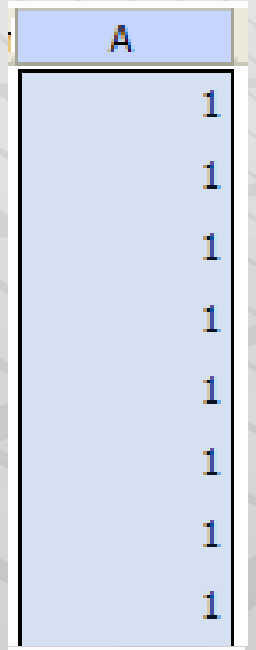

- If you've already defined T_Column_Metadata, this opens the table, otherwise…
- Select the data set you want to use, then put in the number of fields (factors) you want
- For example:
- You have Treatment information and Biological Replicate information = 2 fields = enter in "2" and hit OK

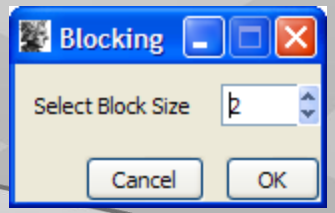


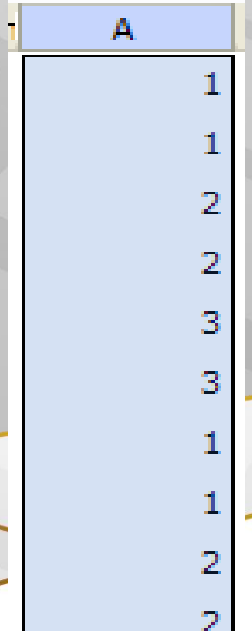
Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Metadata/Define Factors (2)



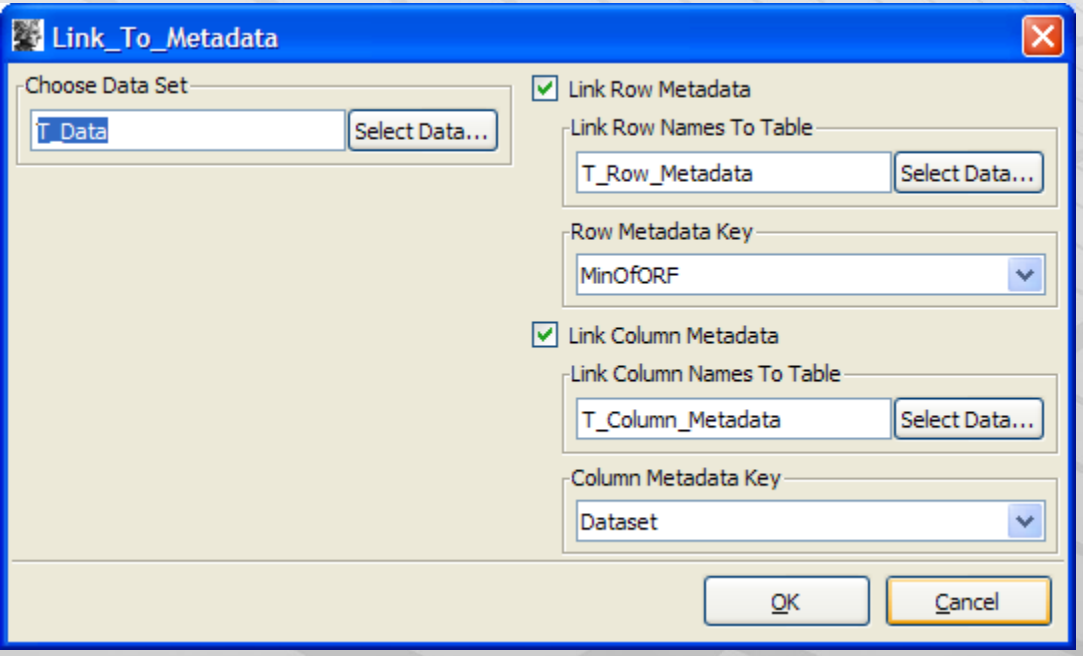




- Rename the new column using right-click on column header
- Ways to put your factor information in:
- 1. Fill some cells with one value
  - Enter a value into a cell
  - Highlight a column region with that cell at the top and right-click
  - Choose “Fill Down”
- 2. Fill cells with a pattern
  - Enter several values into different cells
  - Highlight a column region and right-click
  - Choose “Fill In Cycles” and select the block size
  - The region will be filled with blocks.
- 3. Copy and paste in from Excel or Notepad – use right-click or Ctrl-V



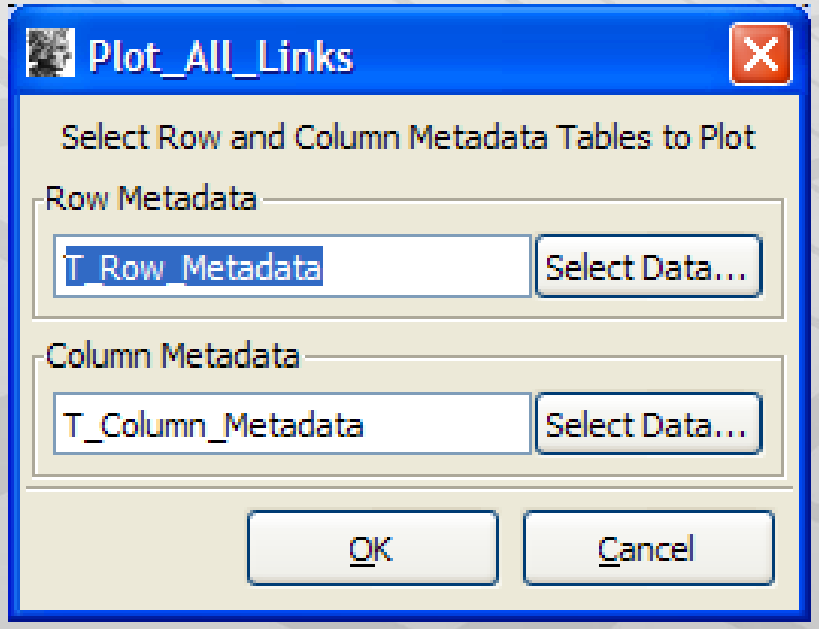
Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Metadata/Link To Metadata




- Select a data set and check it contains linked row and column metadata tables
- The Key choice fields will automatically populate – if they're blank, the links are broken.
- Hit "OK" to make the link.

# Metadata/Plot All Links

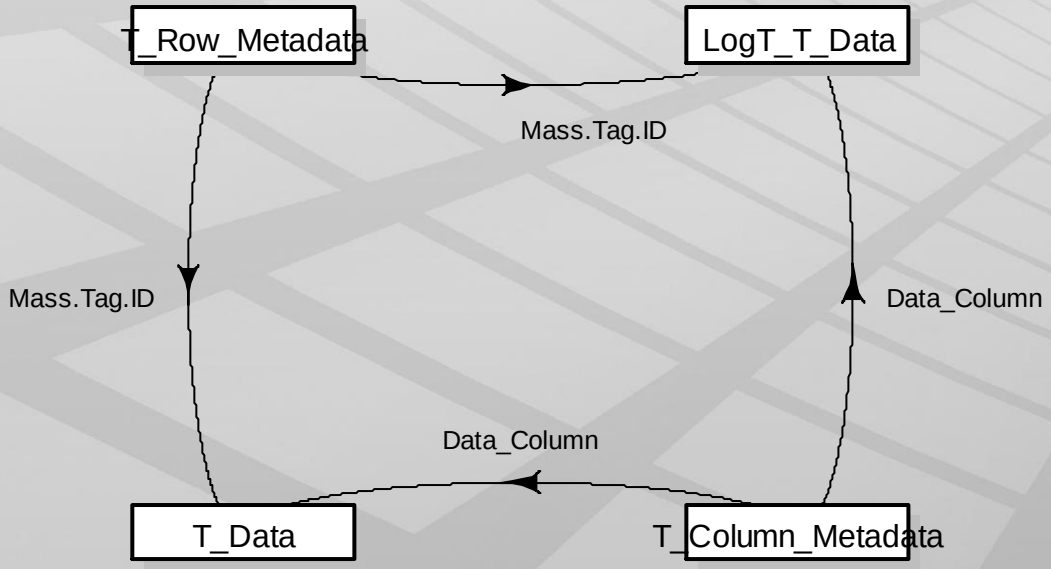



- How can you see all the links at once?
- Select your row and column metadata tables (it should default to their names)
- The function goes through all linked tables and plots how they are linked, highlighting problems



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Metadata/Plot All Links – check data linking




Badly specified links shown in red.

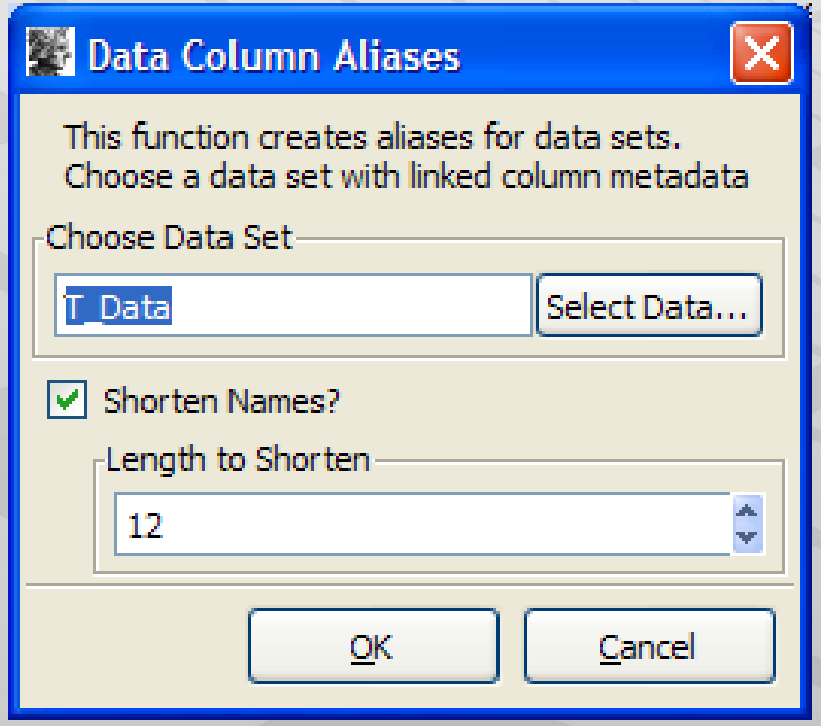
Bad Row/Column Items: means the row/column names of your table aren't contained in the row or column metadata

Bad Row/Column Key: your table is pointing at a nonexistent key.



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Metadata/Create Data Aliases

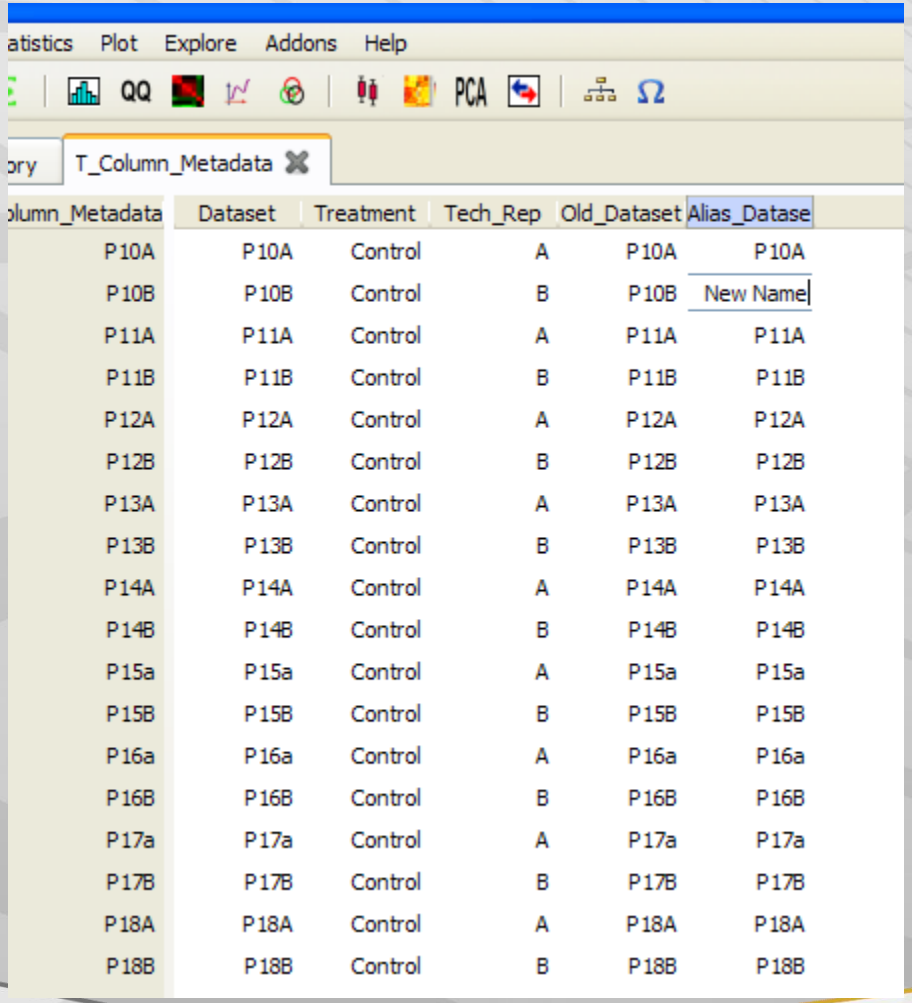



- Long data set names are annoying
- We can change them globally and reversibly by using aliases
- Select a data set and hit OK



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Metadata/Create Data Aliases (2)



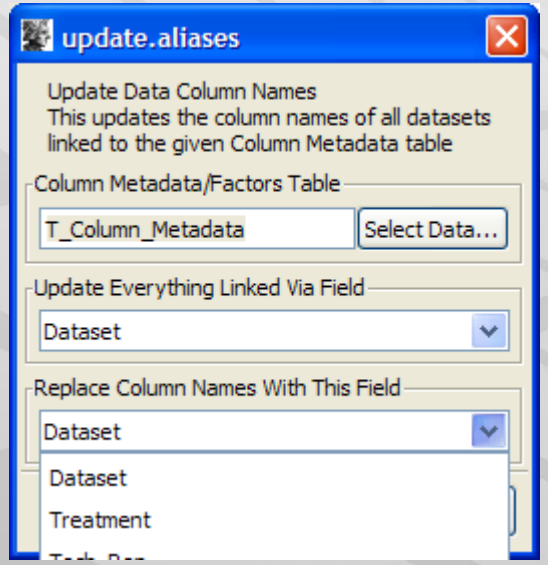

| olumn_Metadata | Dataset | Treatment | Tech_Rep | Old_Dataset | Alias_Datase |
|---|---|---|---|---|---|
| P10A | P10A | Control | A | P10A | P10A |
| P10B | P10B | Control | B | P10B | New Name |
| P11A | P11A | Control | A | P11A | P11A |
| P11B | P11B | Control | B | P11B | P11B |
| P12A | P12A | Control | A | P12A | P12A |
| P12B | P12B | Control | B | P12B | P12B |
| P13A | P13A | Control | A | P13A | P13A |
| P13B | P13B | Control | B | P13B | P13B |
| P14A | P14A | Control | A | P14A | P14A |
| P14B | P14B | Control | B | P14B | P14B |
| P15a | P15a | Control | A | P15a | P15a |
| P15B | P15B | Control | B | P15B | P15B |
| P16a | P16a | Control | A | P16a | P16a |
| P16B | P16B | Control | B | P16B | P16B |
| P17a | P17a | Control | A | P17a | P17a |
| P17B | P17B | Control | B | P17B | P17B |
| P18A | P18A | Control | A | P18A | P18A |
| P18B | P18B | Control | B | P18B | P18B |

- Then open up T_Column_Metadata and notice 2 extra columns:
- Old_Dataset and Alias_Dataset (containing shortened names, if the originals were long)
- Edit the Alias_Dataset column with your new names and go to “Metadata/Apply Aliases”…



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Metadata/Apply Aliases

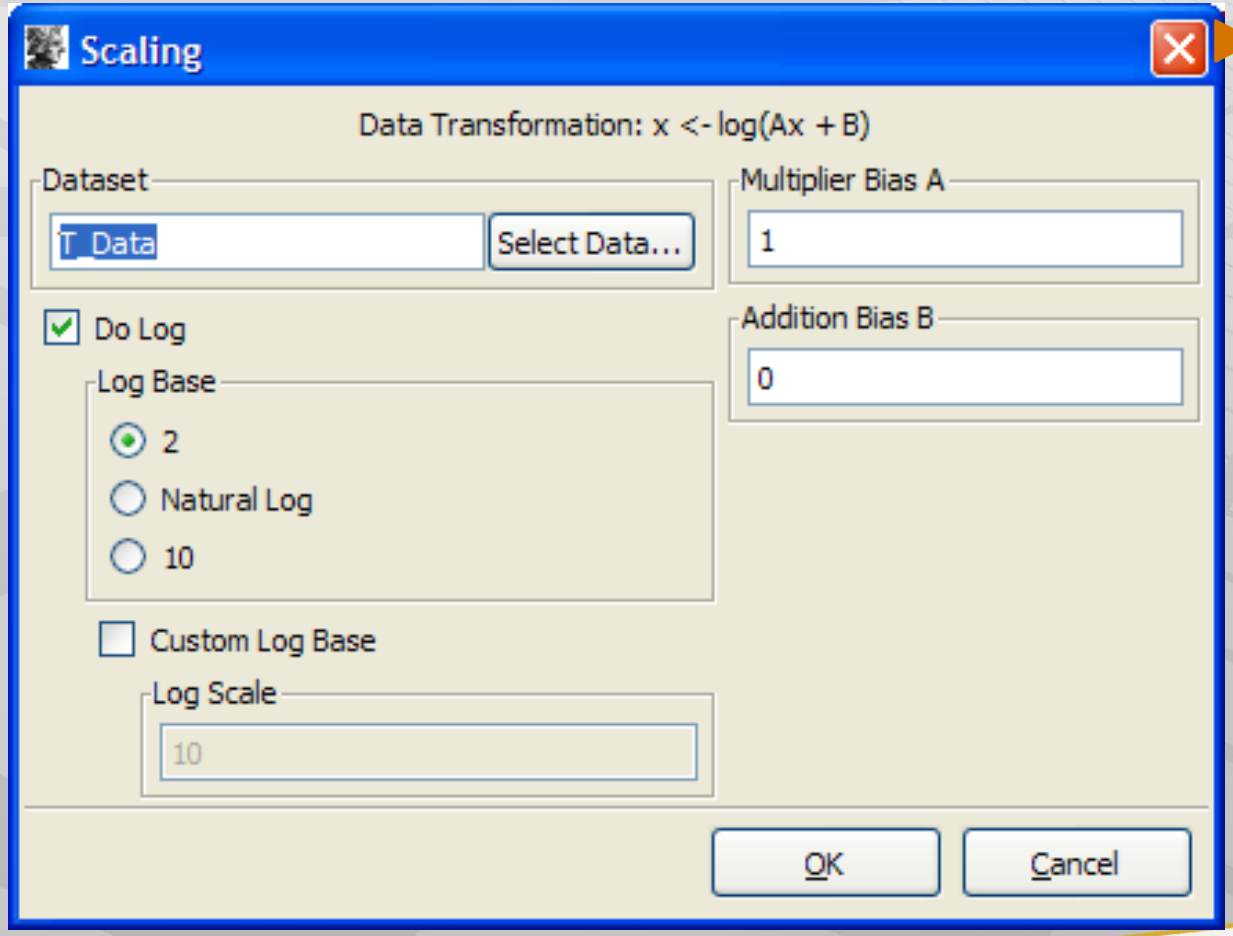



- Select T_Column_Metadata and then update everything linked via the key your dataset was linked with, replacing with "Alias_Dataset"
- This will replace the column names in all linked tables
- To undo this, just replace the same linked field with "Old_Dataset"
- Don't replace column names with non-unique fields.



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Pre-Process/Transform Data

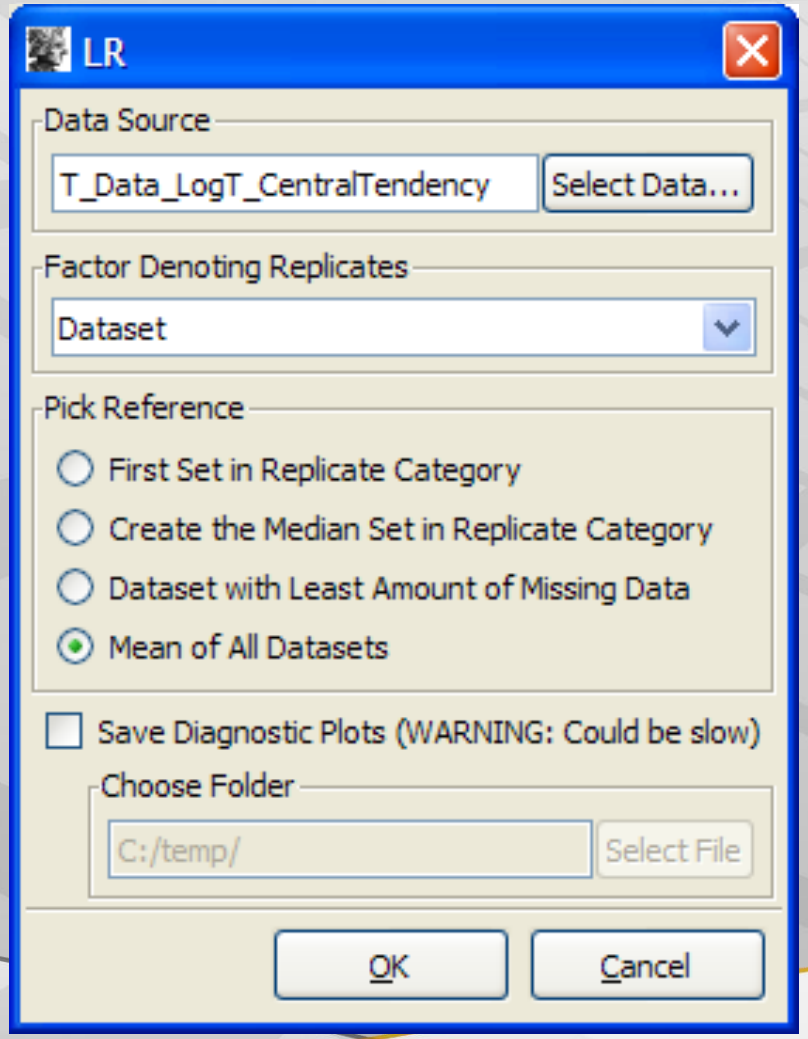



Log Transform and scaling on data set



Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Preprocess/Linear Regression

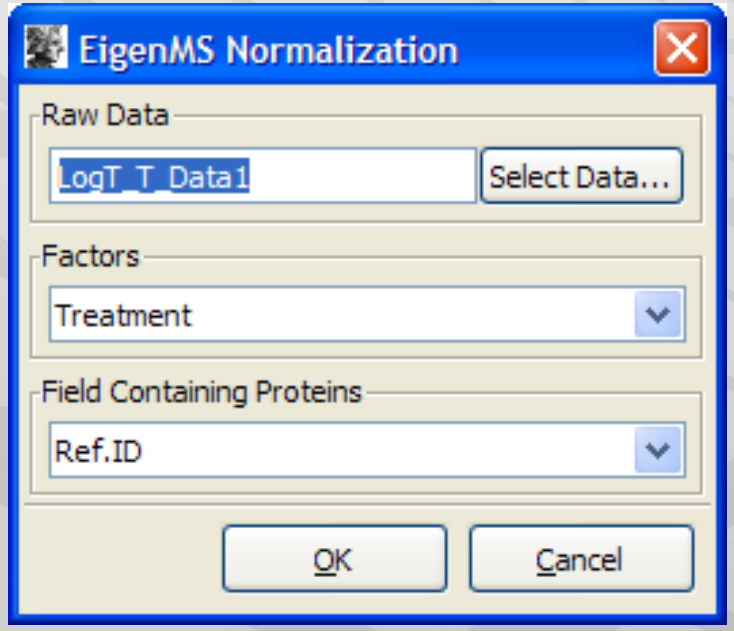



- Correct systematic differences between data sets (columns)
- Linear Regression normalization using various methods (See DAnTE help)
- “Mean of All Datasets” doesn’t require any factor information, the others do
- To do: apply a variance weighting
- Save Plots… doesn’t do anything useful



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Preprocess/Normalization/EigenMS

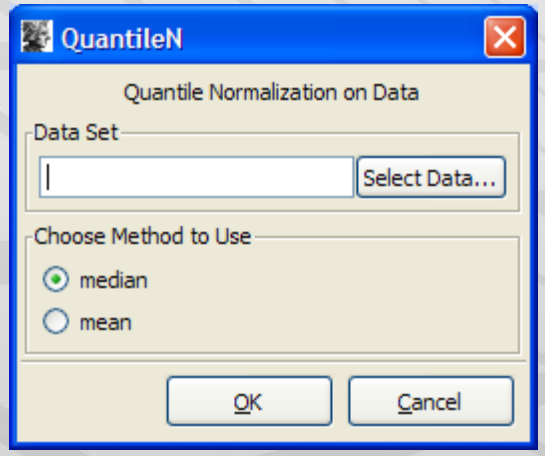



- EigenMS normalization as per Karpievich et al. 2009
- Factor-free normalization using SVD
- Requires entry of field containing proteins (in row metadata table)



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

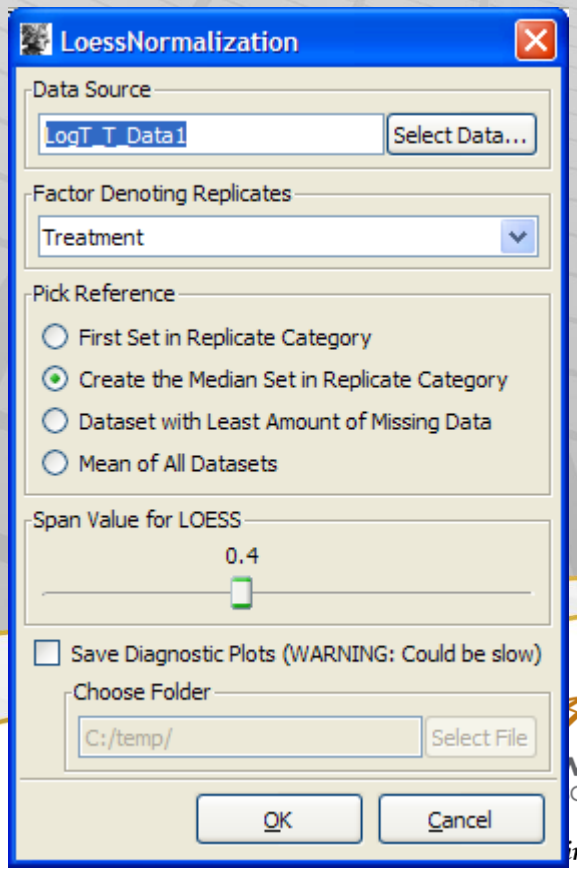
# Preprocess/Normalization/Using Quantiles



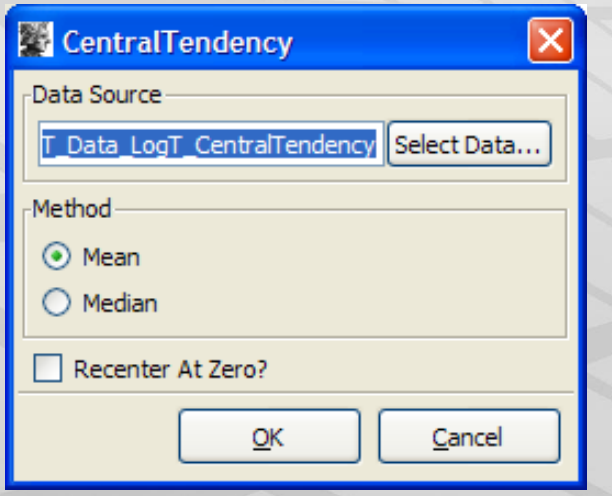

- Quantile normalization
- Nonparametric method from genomics
- Bolstad et al, Bioinformatics (2003)

# Preprocess/Normalization/ Using LOESS

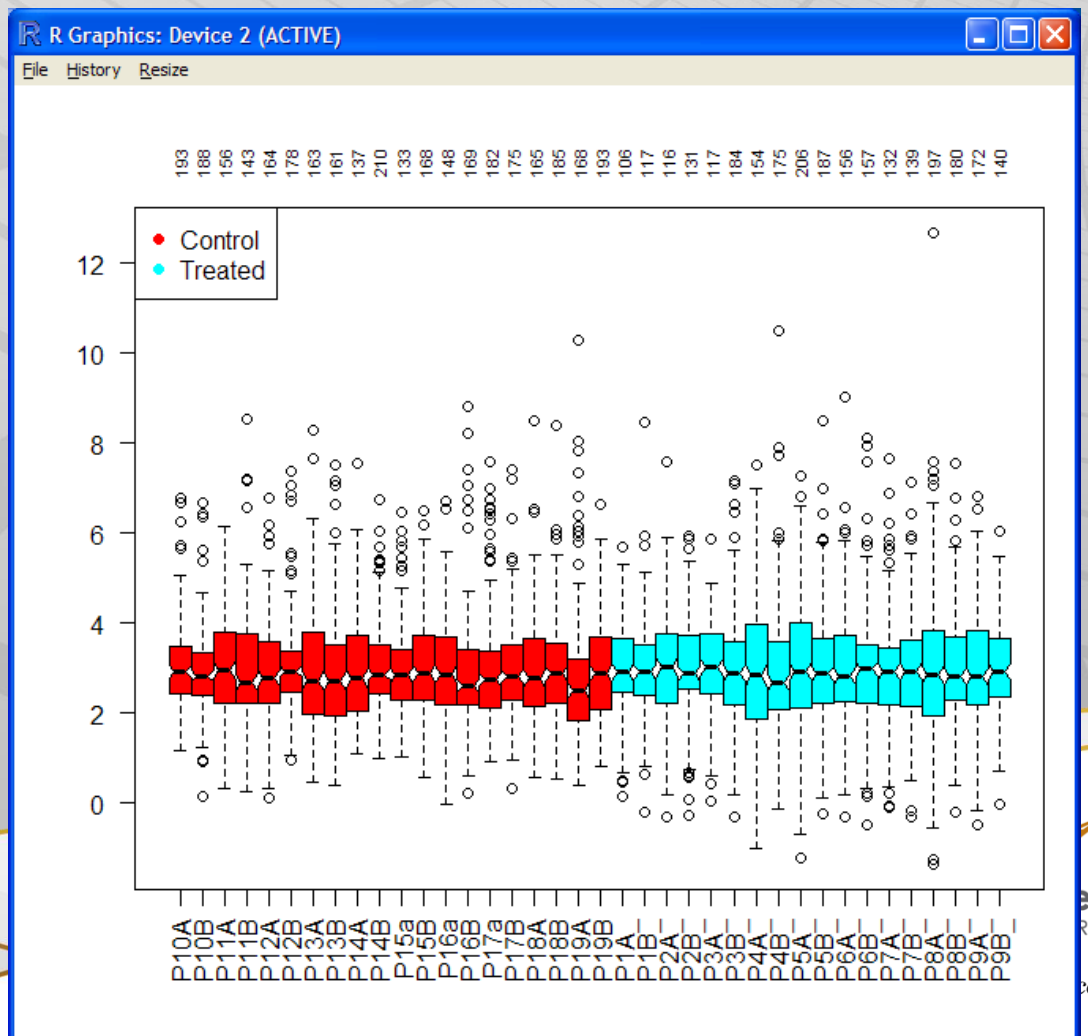
- Locally weighted smoothing
- Correction for nonlinearities in response



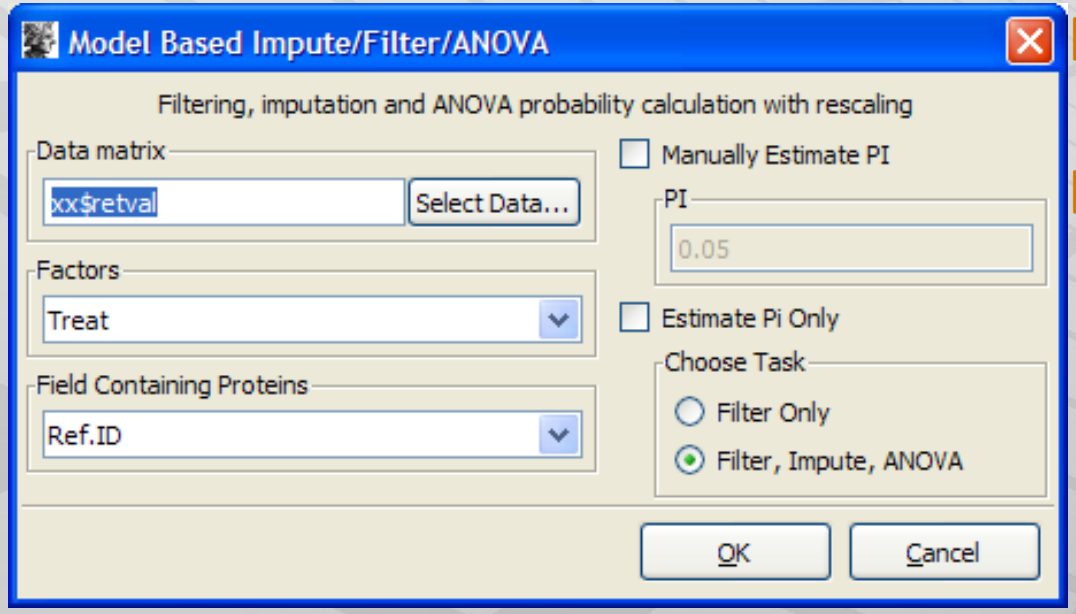
# Preprocess/Central tendency correction




Correct systematic additive differences between data sets (columns)



# Preprocess/Impute/Model Based

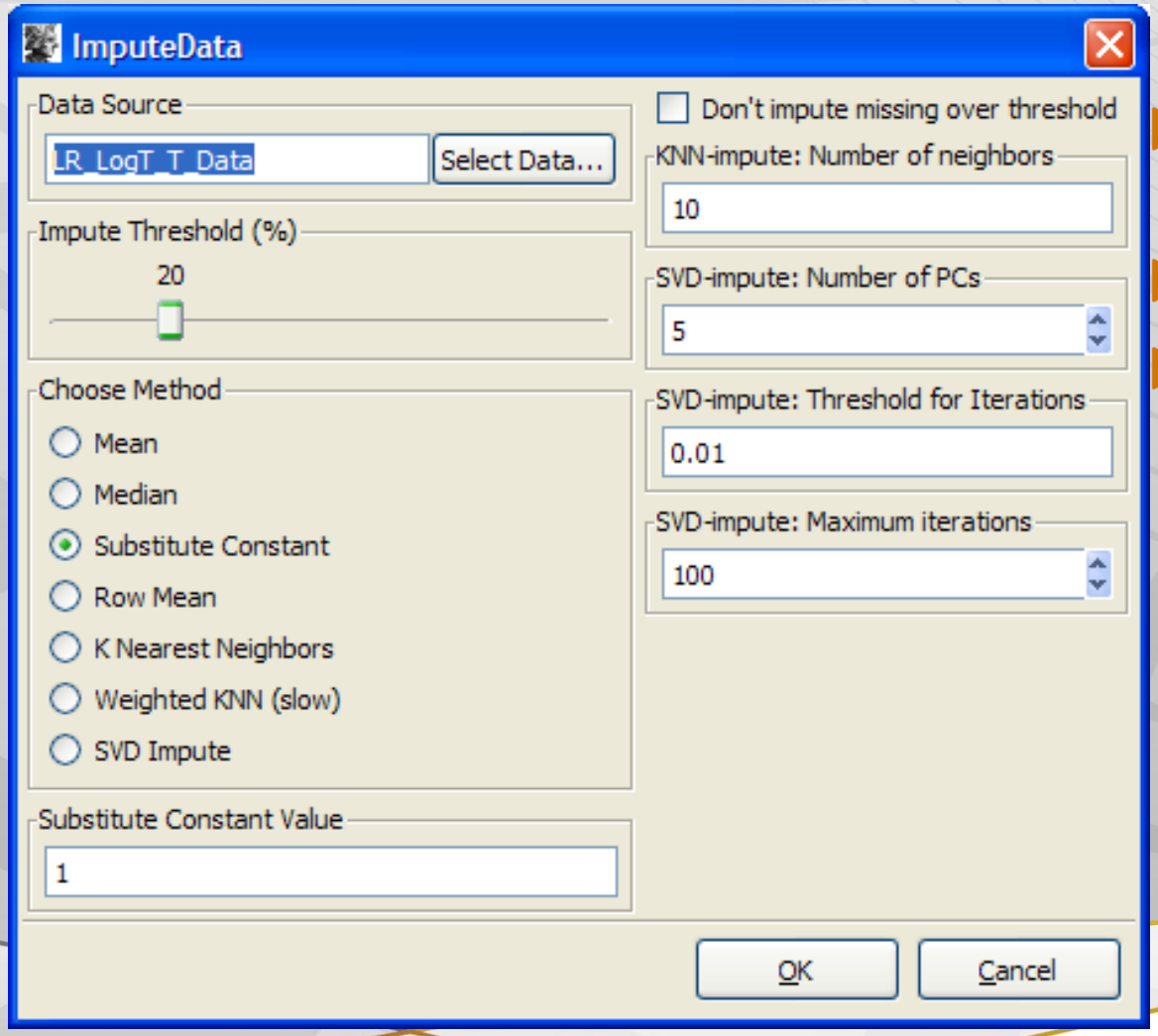



- Model based imputation algorithm
- (Karpievich, 2009)



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Preprocess/Impute Missing Values

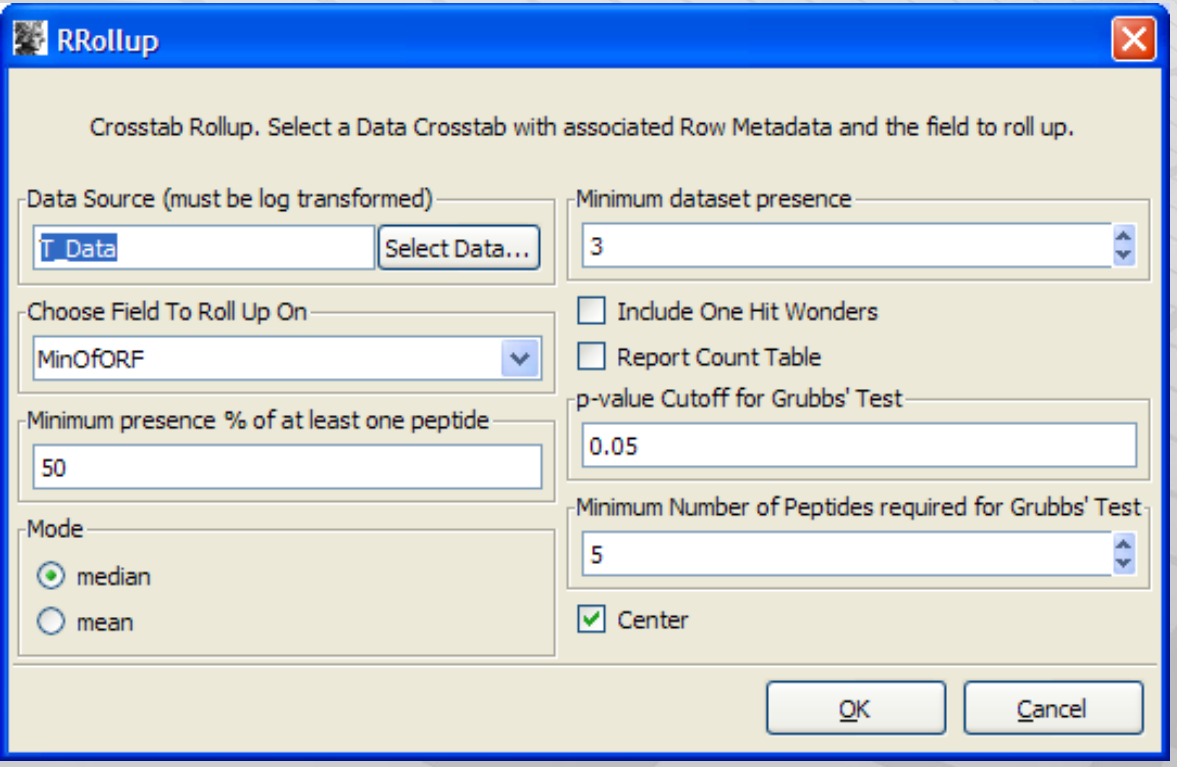



- Methods to replace missing values in data
- Use with caution!
- See original help file for references



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Rollup/Rrollup, Qrollup, ZRollup




- Rollup peptides to proteins
- Take median value over crosstab rows grouped by row metadata
- See DAnTE help for details; similar options



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Statistics

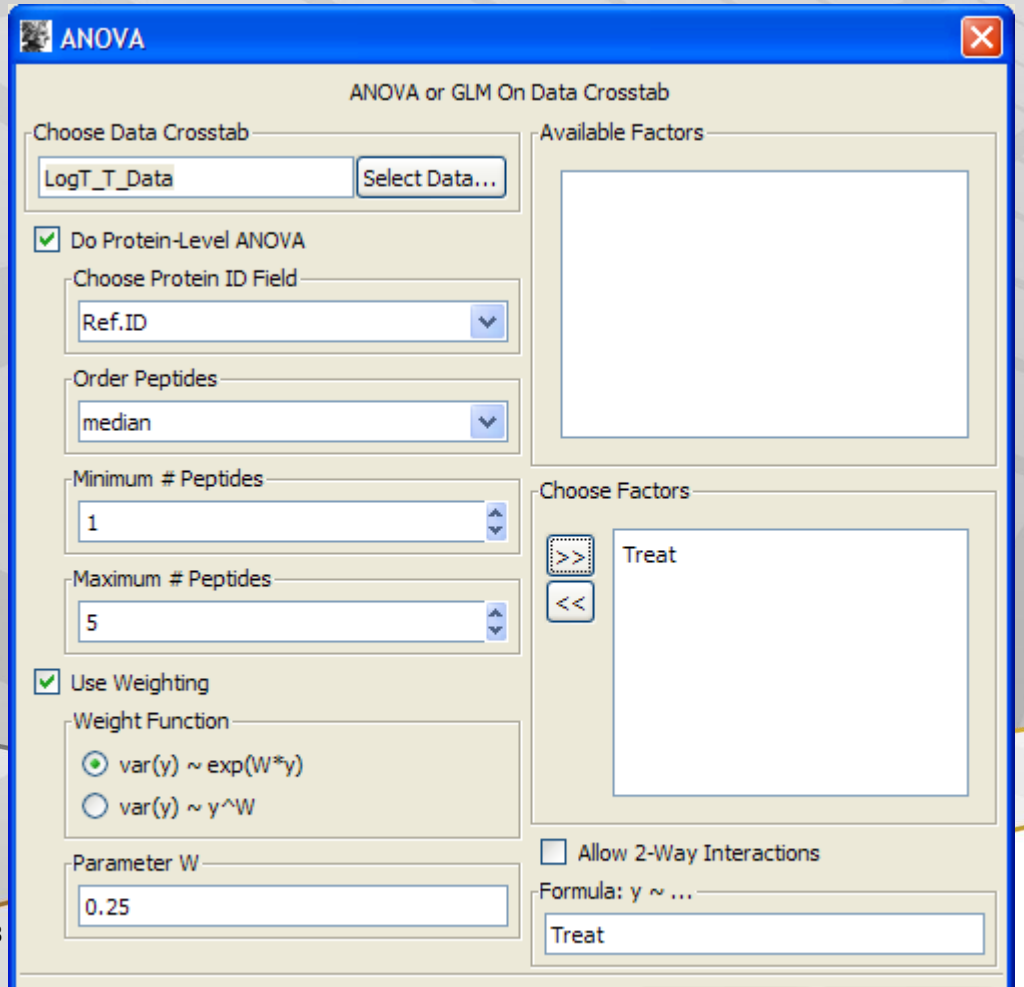


Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Statistics/ANOVA (generalized t-test)

**Two sets of measurements (factors): x is (1.1, 1.3, 1.4) and y is (2.1, 2.3, 2.2)**

* What is the average difference of x and y? (estimate size)
* Could it have occurred by chance? (significance)

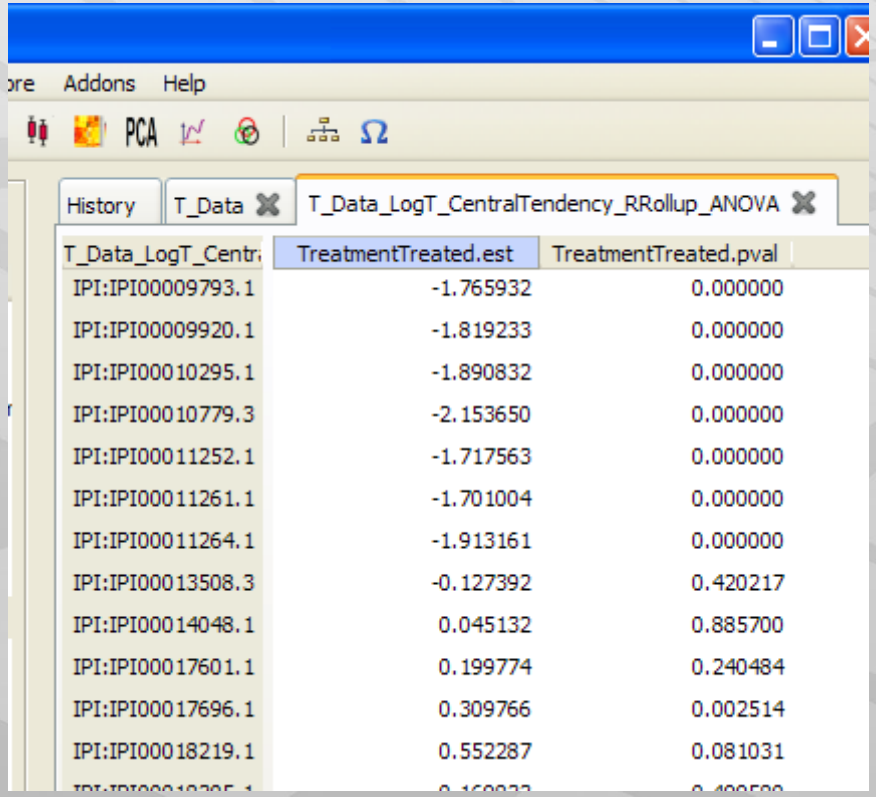



- Available Factors
- Choose Factors
- Display ANOVA Formula



Pacific Northwest
NATIONAL LABORATORY
*Proudly Operated by* Battelle *Since 1965*

# Statistics/ANOVA results



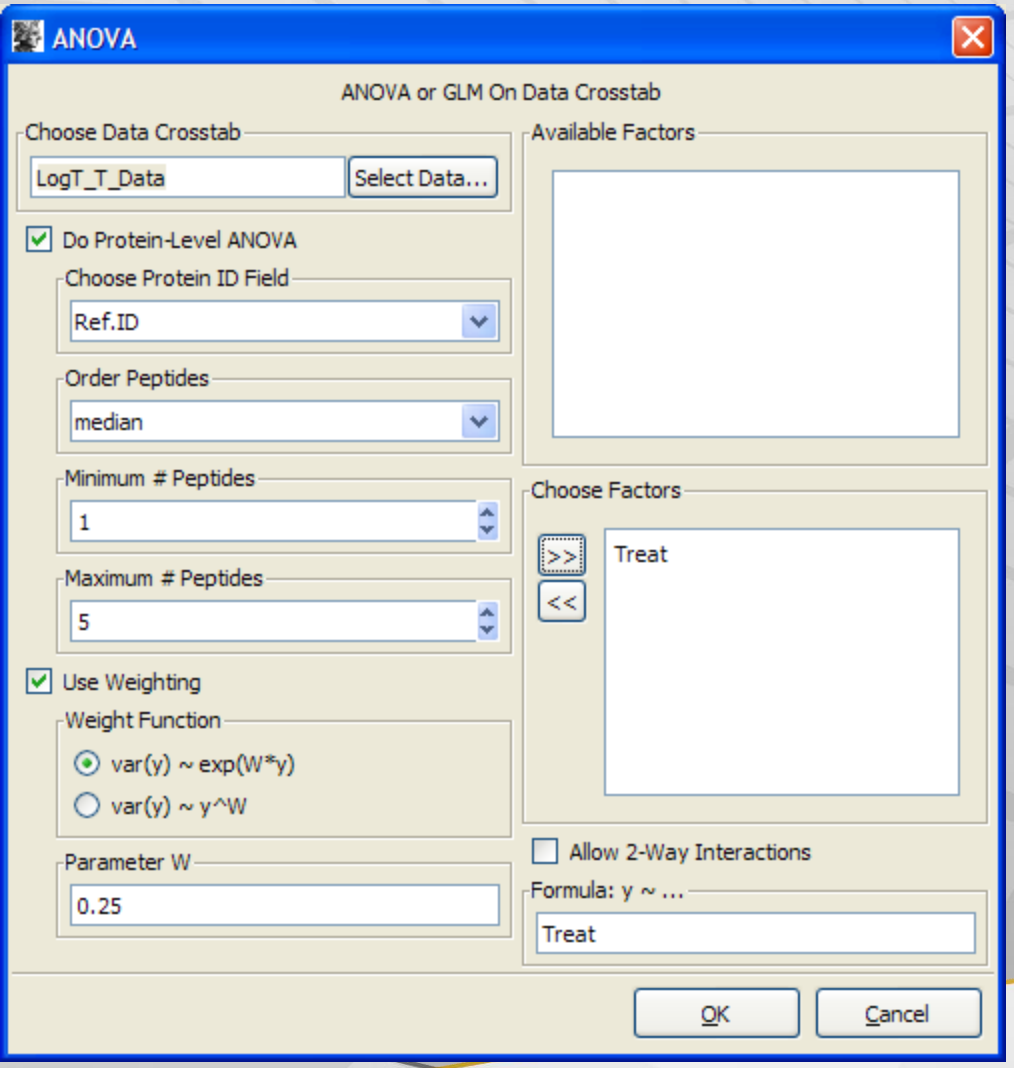

| T_Data_LogT_Centr | TreatmentTreated.est | TreatmentTreated.pval |
|---|---|---|
| IPI:IPI00009793.1 | -1.765932 | 0.000000 |
| IPI:IPI00009920.1 | -1.819233 | 0.000000 |
| IPI:IPI00010295.1 | -1.890832 | 0.000000 |
| IPI:IPI00010779.3 | -2.153650 | 0.000000 |
| IPI:IPI00011252.1 | -1.717563 | 0.000000 |
| IPI:IPI00011261.1 | -1.701004 | 0.000000 |
| IPI:IPI00011264.1 | -1.913161 | 0.000000 |
| IPI:IPI00013508.3 | -0.127392 | 0.420217 |
| IPI:IPI00014048.1 | 0.045132 | 0.885700 |
| IPI:IPI00017601.1 | 0.199774 | 0.240484 |
| IPI:IPI00017696.1 | 0.309766 | 0.002514 |
| IPI:IPI00018219.1 | 0.552287 | 0.081031 |

- 2 columns per combination of factor and factor levels
- “est”: fold change estimate
- “pval”: p-value



Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Statistics/ANOVA: Peptide-level ANOVA

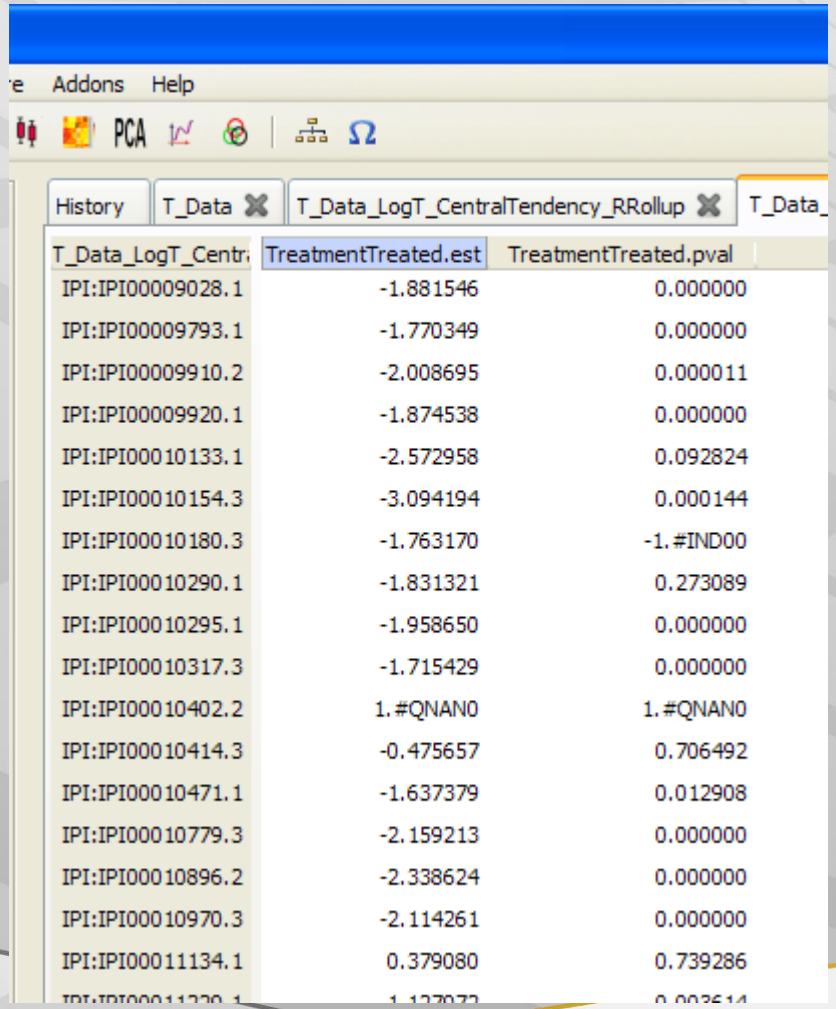



- Check "Do Protein-Level ANOVA" and select protein ID
- Split peptide data into blocks of proteins and apply the ANOVA model
- Gives more sensitivity than rolling up
- C Crescentus: Rrollup "discovered" 17/1097 proteins compared to 181 using peptide-level ANOVA.



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Statistics/ANOVA



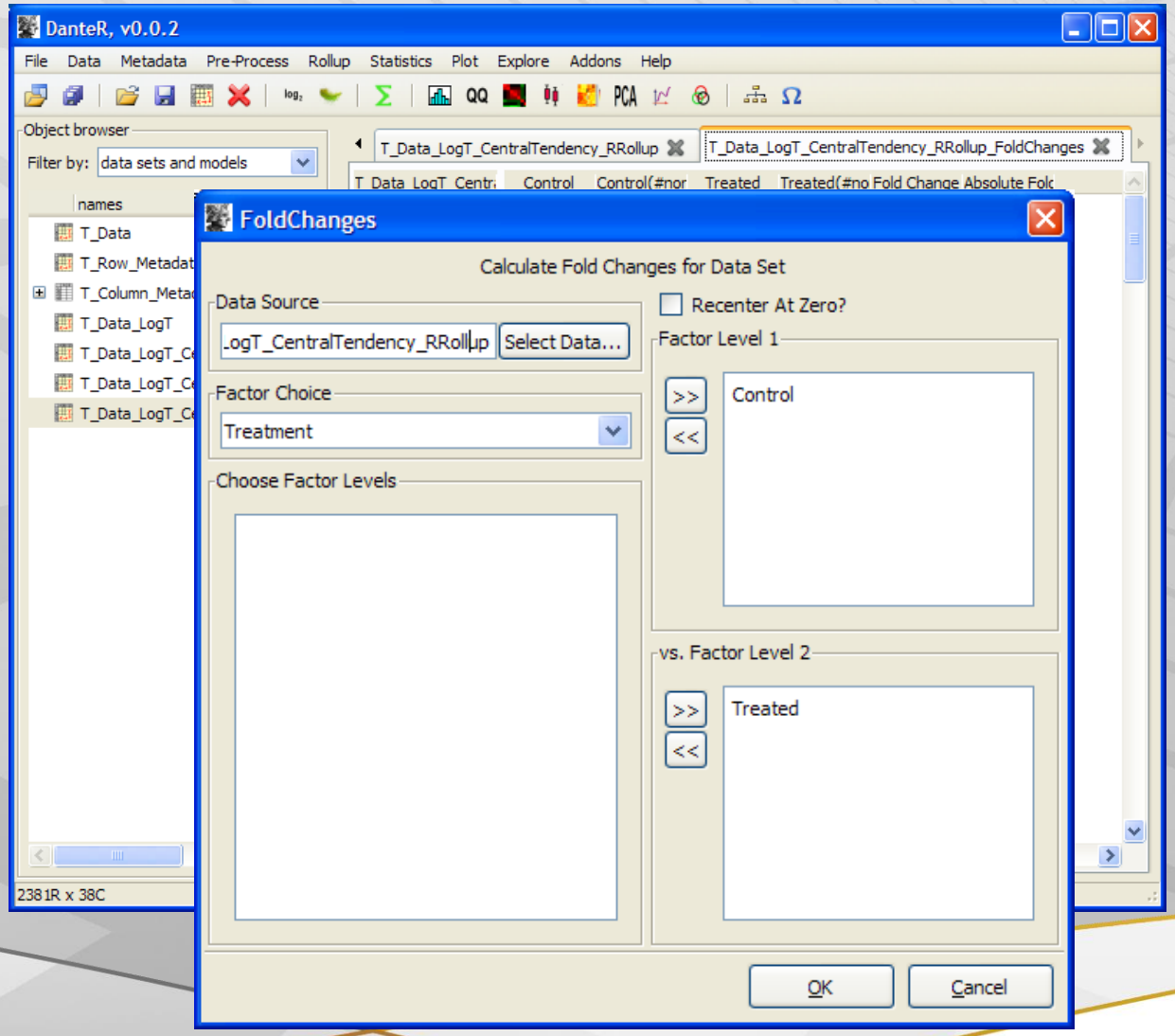

| T_Data_LogT_Centr | TreatmentTreated.est | TreatmentTreated.pval |
|---|---|---|
| IPI:IPI00009028.1 | -1.881546 | 0.000000 |
| IPI:IPI00009793.1 | -1.770349 | 0.000000 |
| IPI:IPI00009910.2 | -2.008695 | 0.000011 |
| IPI:IPI00009920.1 | -1.874538 | 0.000000 |
| IPI:IPI00010133.1 | -2.572958 | 0.092824 |
| IPI:IPI00010154.3 | -3.094194 | 0.000144 |
| IPI:IPI00010180.3 | -1.763170 | -1.#IND00 |
| IPI:IPI00010290.1 | -1.831321 | 0.273089 |
| IPI:IPI00010295.1 | -1.958650 | 0.000000 |
| IPI:IPI00010317.3 | -1.715429 | 0.000000 |
| IPI:IPI00010402.2 | 1.#QNAN0 | 1.#QNAN0 |
| IPI:IPI00010414.3 | -0.475657 | 0.706492 |
| IPI:IPI00010471.1 | -1.637379 | 0.012908 |
| IPI:IPI00010779.3 | -2.159213 | 0.000000 |
| IPI:IPI00010896.2 | -2.338624 | 0.000000 |
| IPI:IPI00010970.3 | -2.114261 | 0.000000 |
| IPI:IPI00011134.1 | 0.379080 | 0.739286 |

- Direct estimate for protein fold changes from peptide data
- More powerful statistics (see my ASMS poster)



Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Statistics/Fold Changes

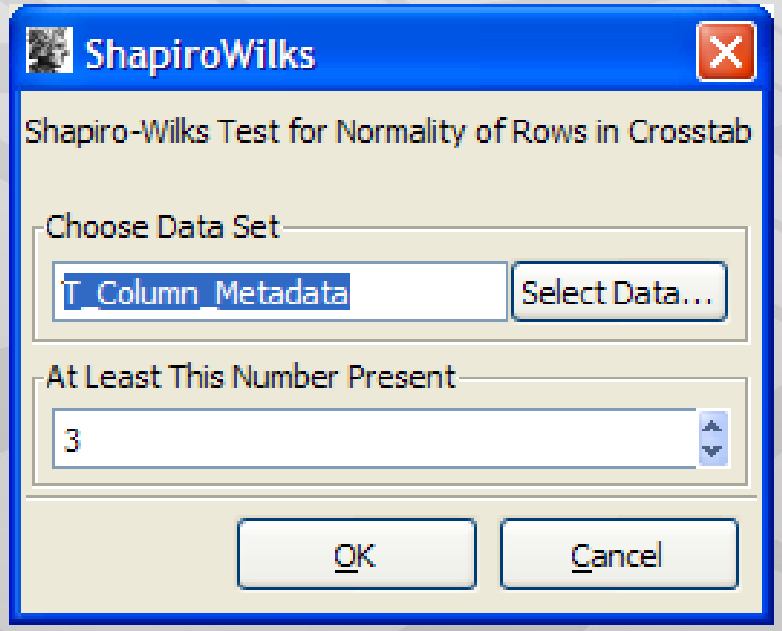



- Compare fold changes between conditions
- Choose a data set and select the factor to choose a fold change over
- Requires linked column metadata
- Select levels from the factor to compare to one another and hit OK



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Statistics/Shapiro-Wilks test

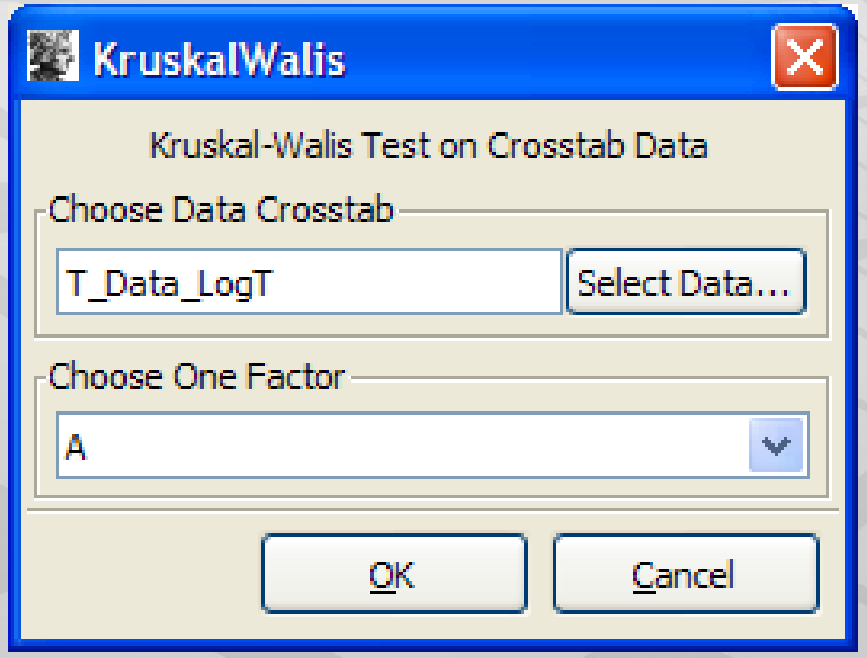



- Performs the Shapiro-Wilk test of normality on each row of a selected data set
- Output is a 2-column table, one row per row of the dataset, containing:
  - Normal: is the data normal? ($p < 0.1$)
  - Pvalue: the p-value of the test



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Statistics/KruskalWalis

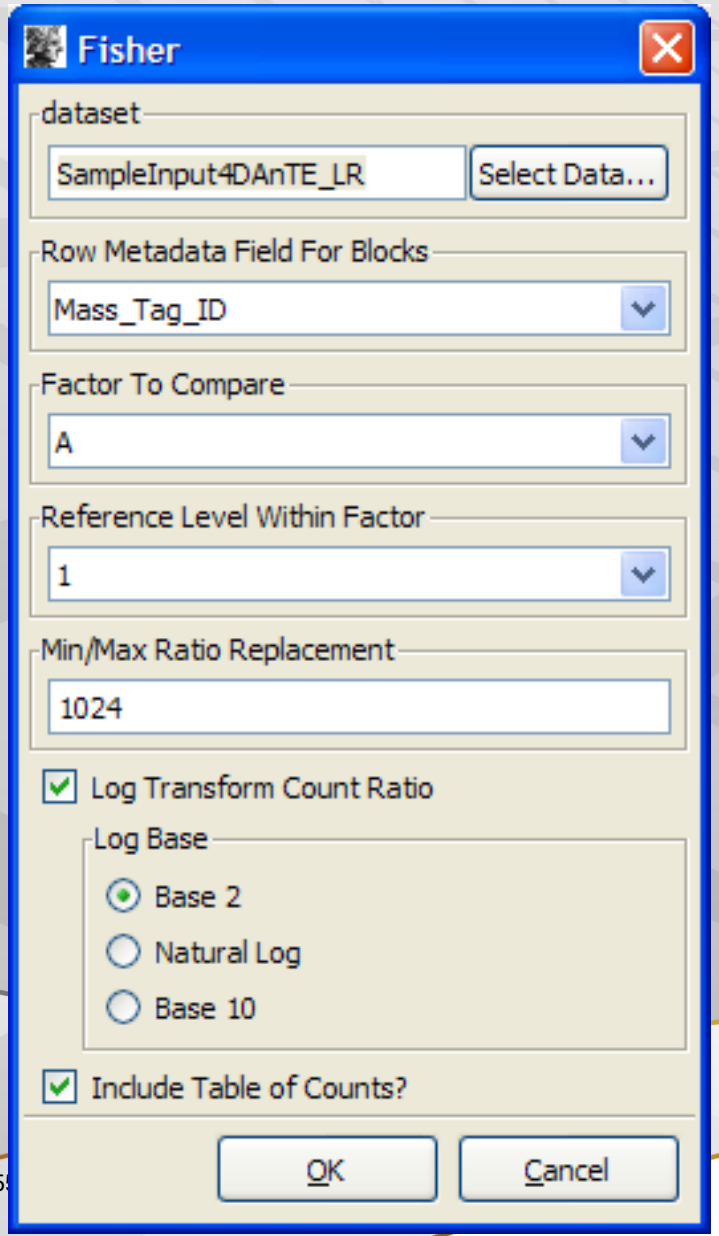



- Nonparametric ANOVA
- Requires a dataset with associated column metadata (factors)
- Performs a Kruskal-Wallis rank sum test on each row
- Choose a dataset and associated factors to perform a comparison over
- Output is a 2-column table, one row per row of the dataset, containing:
  - Normal: is the data normal? ($p < 0.1$)
  - Pvalue: the p-value of the test



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Statistics/Count Fisher Test

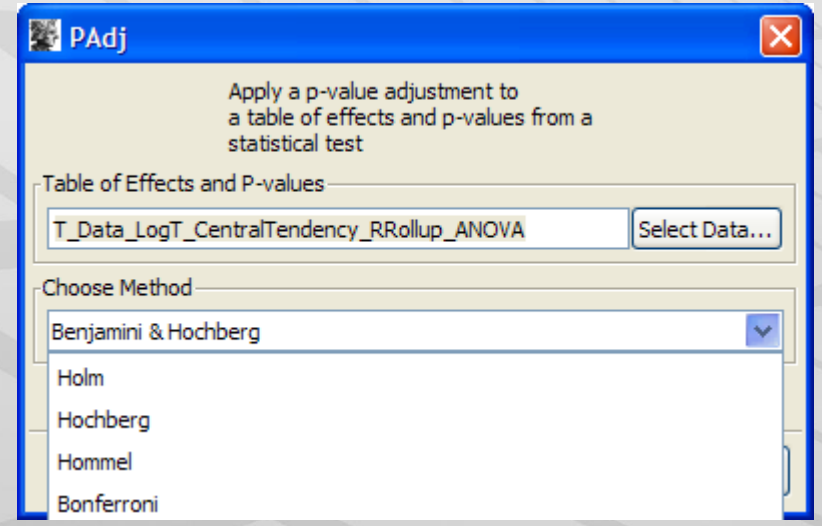



- Statistical test on presence/absence (cf g-test)
- Requires a crosstab with linked row and column metadata
- Cuts dataset into row blocks (“Row Metadata Field for Blocks”) and column blocks (“Factor to Compare”)
- Performs Fisher’s Exact Test to compare absence/presence within sets of row blocks
- Returns table of estimated log2-fold changes and p-values



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Statistics/P-value adjustment

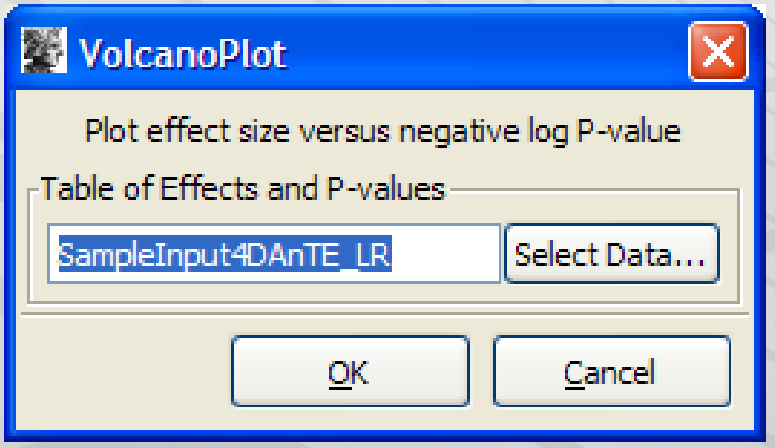


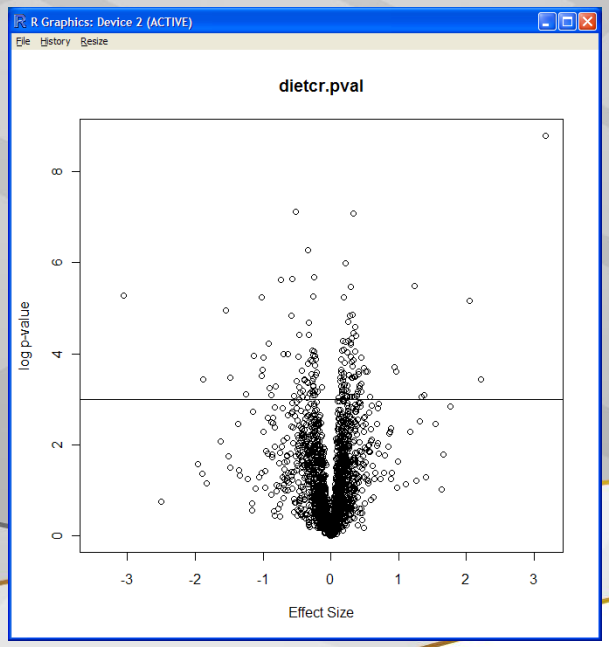

- Adjust for multiple comparisons using standard methods
- Enter the ANOVA or other statistical test table into the data entry
- Choose the P-adjustment method (I recommend Benjamini-Hochberg FDR correction)
- Creates a new table with adjusted P-values



Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Statistics/Volcano Plot



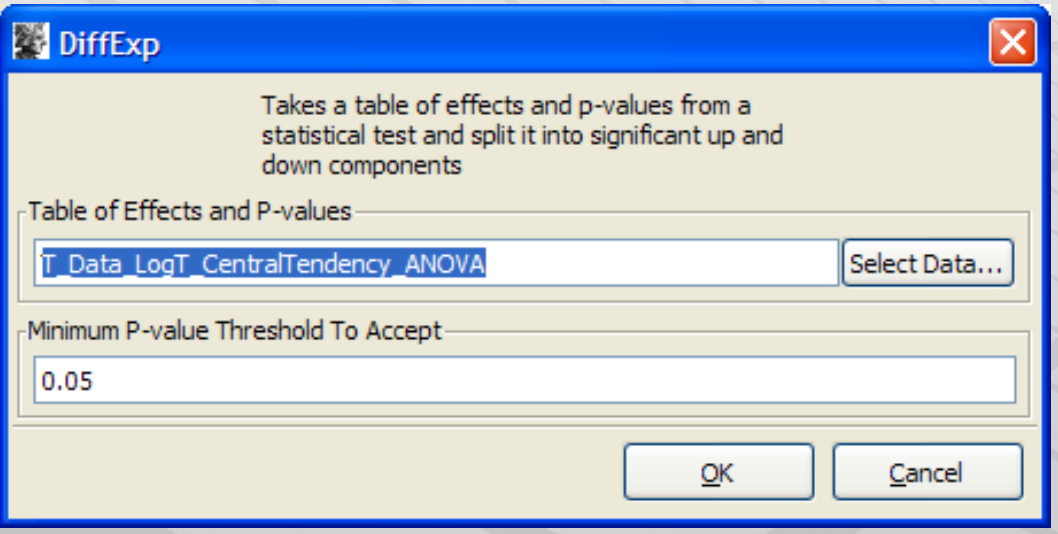




- Useful for assessing statistical output – very asymmetric plots may indicate problems
- Requires output from statistical menu item (ANOVA, Fisher Exact, Shapiro, etc)
- Plots effect size vs. negative-log(P) for each test



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Statistics/Split significant up/down

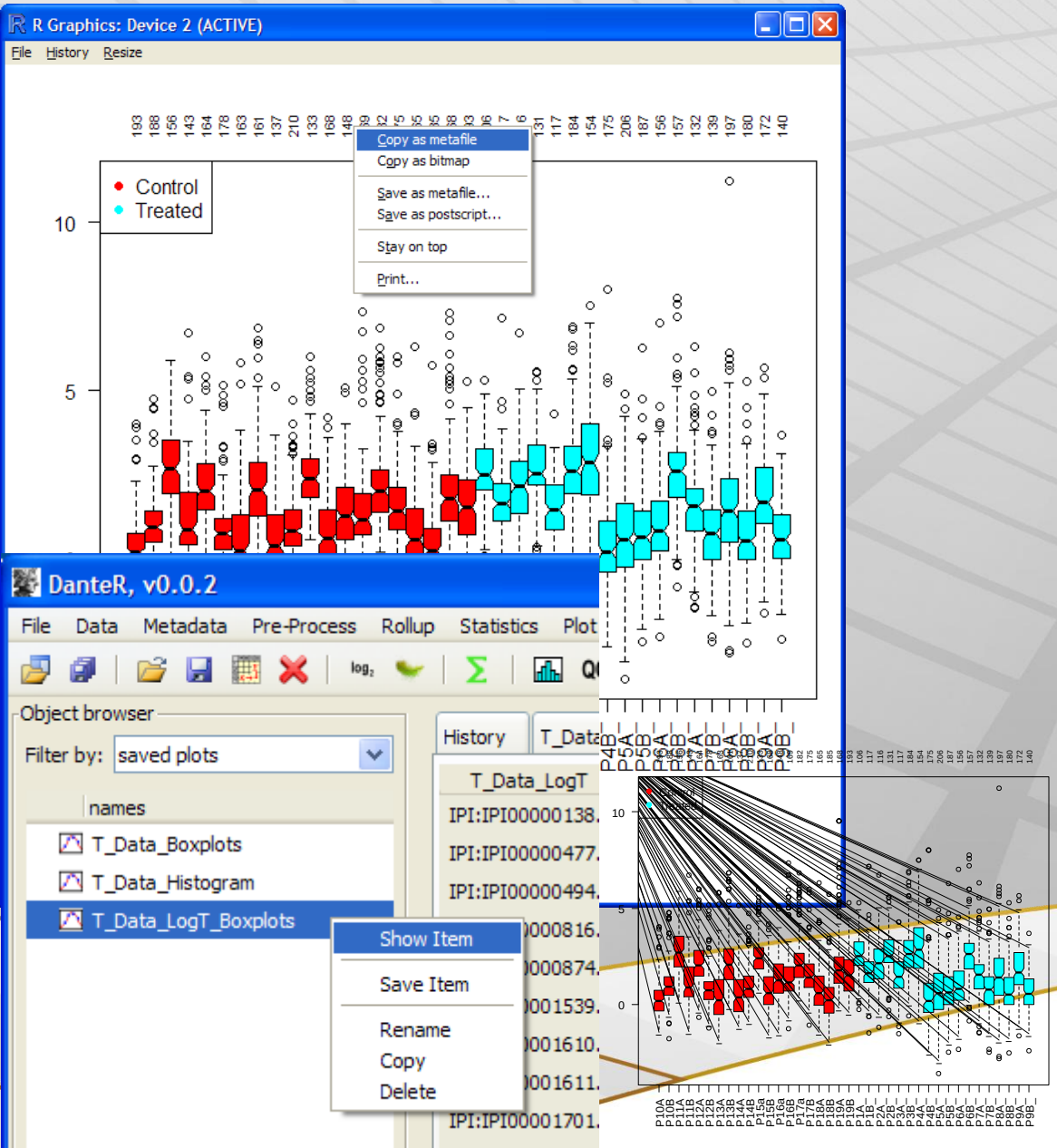



- Create tables of significant up/down
- Requires output from Statistics menu item
- Can plot Venn diagrams from output
- Create tables containing significant up and down expressed items from statistical output



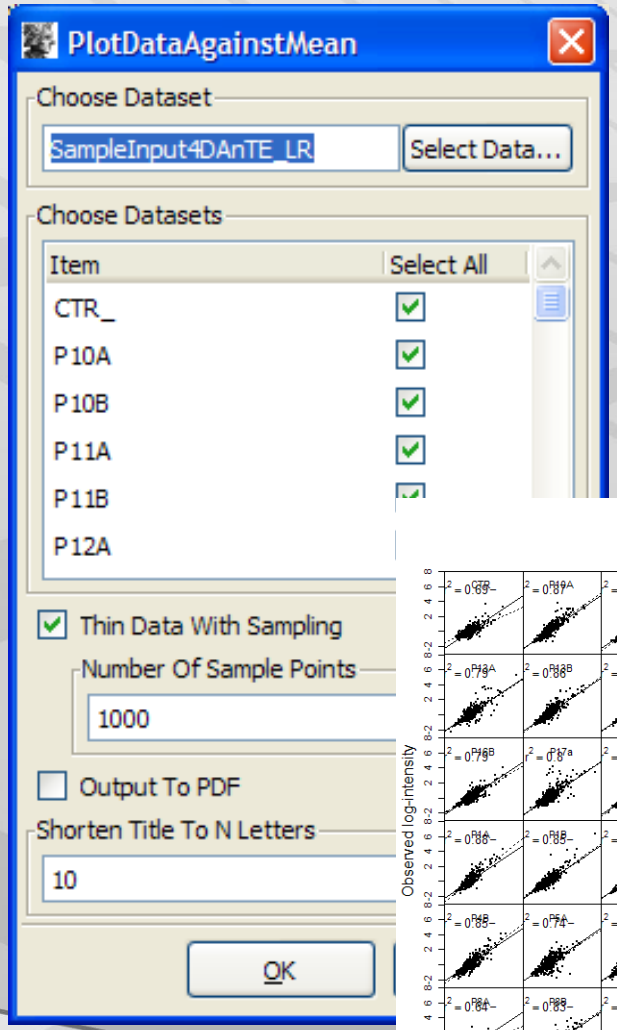
Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

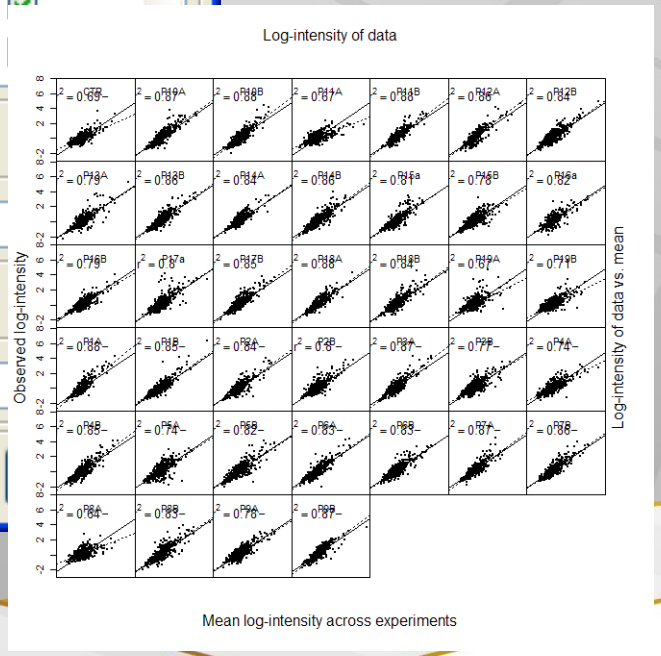
# Plotting menu




- Most plot functions save their output as an R object (accessible via the Object browser/Filter By saved plots)
- Click “Show” or double click to reopen plot
- Rename plots to avoid them being overwritten
- Plots can be resized, exported to powerpoint as metafiles, saved as pdf, etc

# Plot/Plot Against Row Means



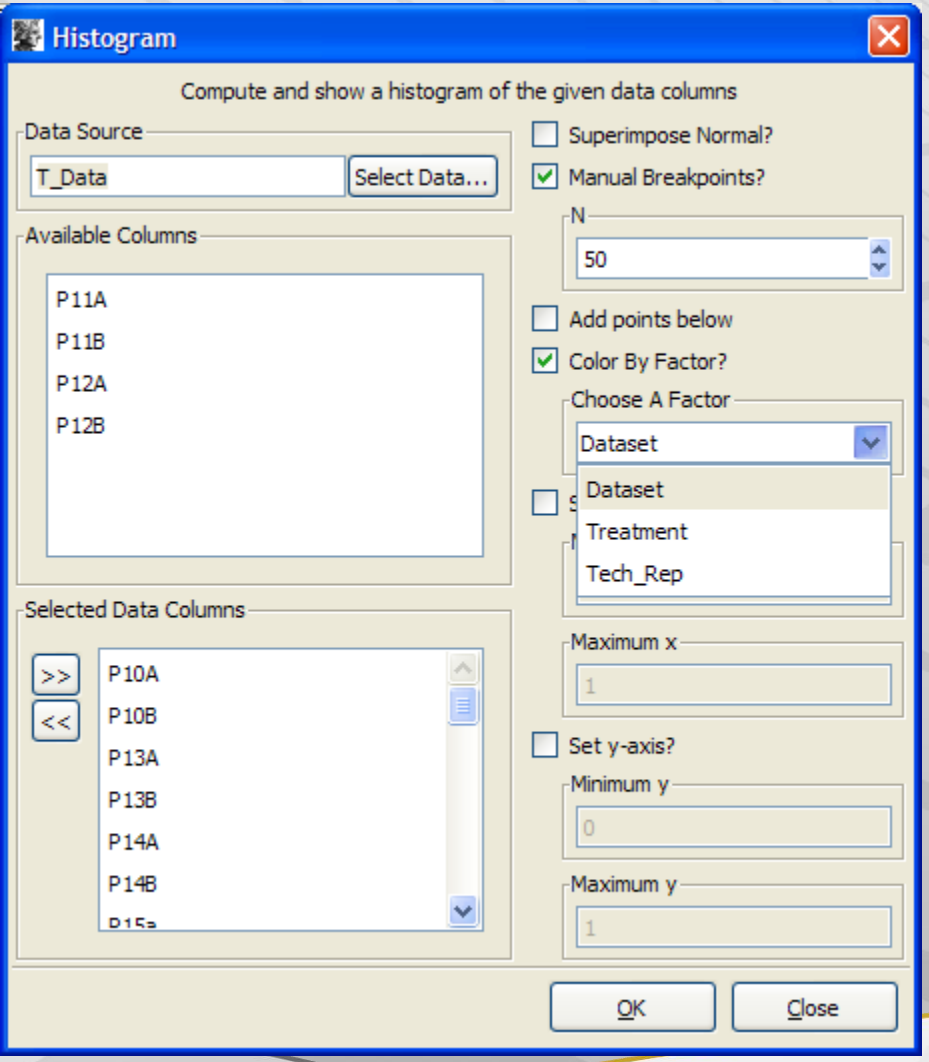




- Useful summary of data quality and seeing poor quality datasets/outliers
- Plot each column of a crosstab against the mean value of each row

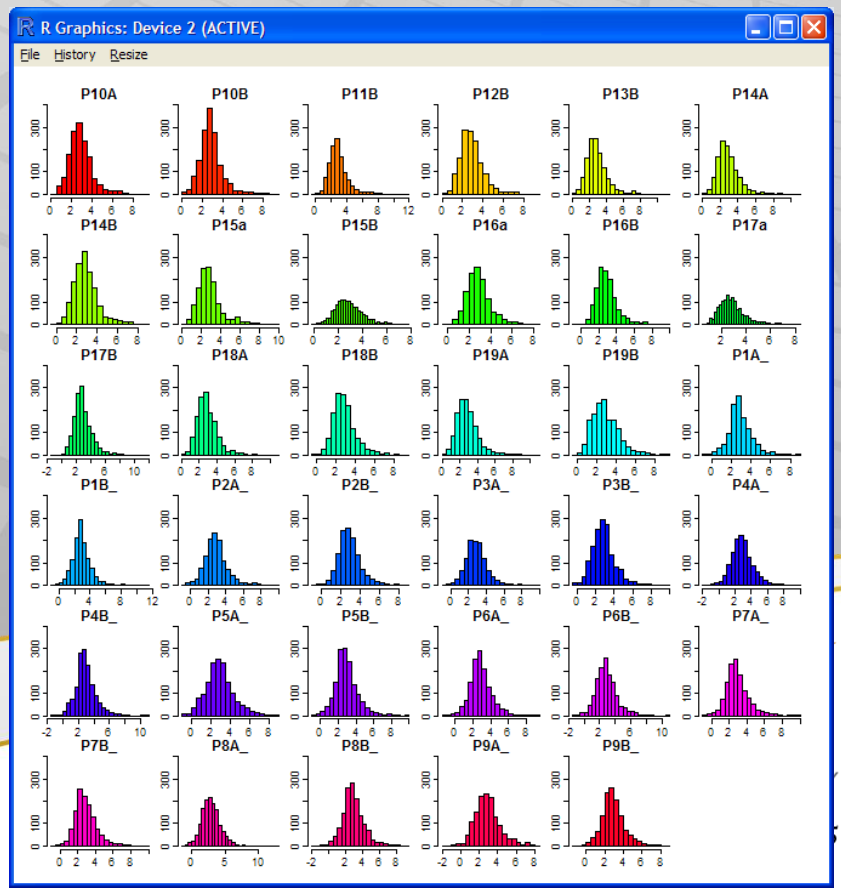


Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

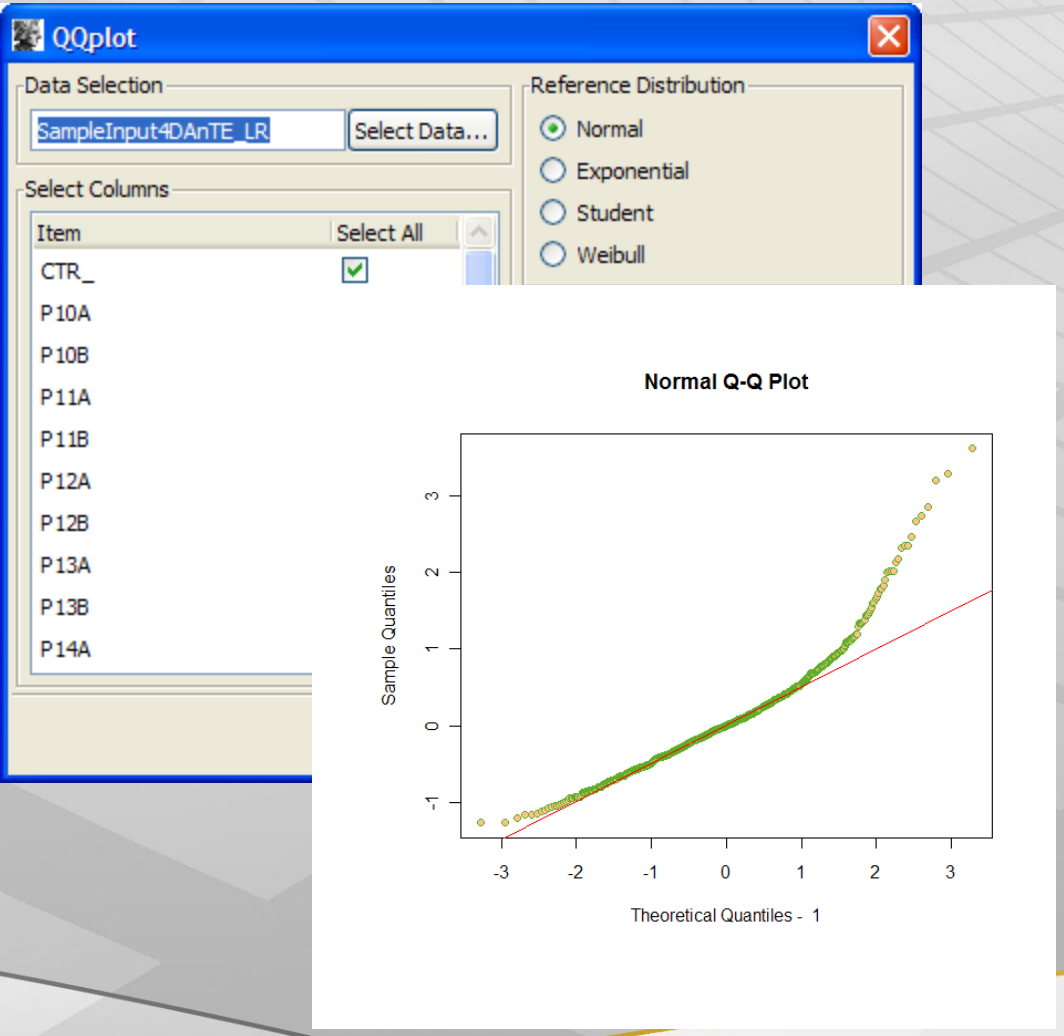
# Plot/Histogram




- Select the columns you're interested in, then hit OK to plot
- Lots of options – see tooltips and ?hist
- If column metadata is defined, you can color by factor



# Plot/Q-Q plot

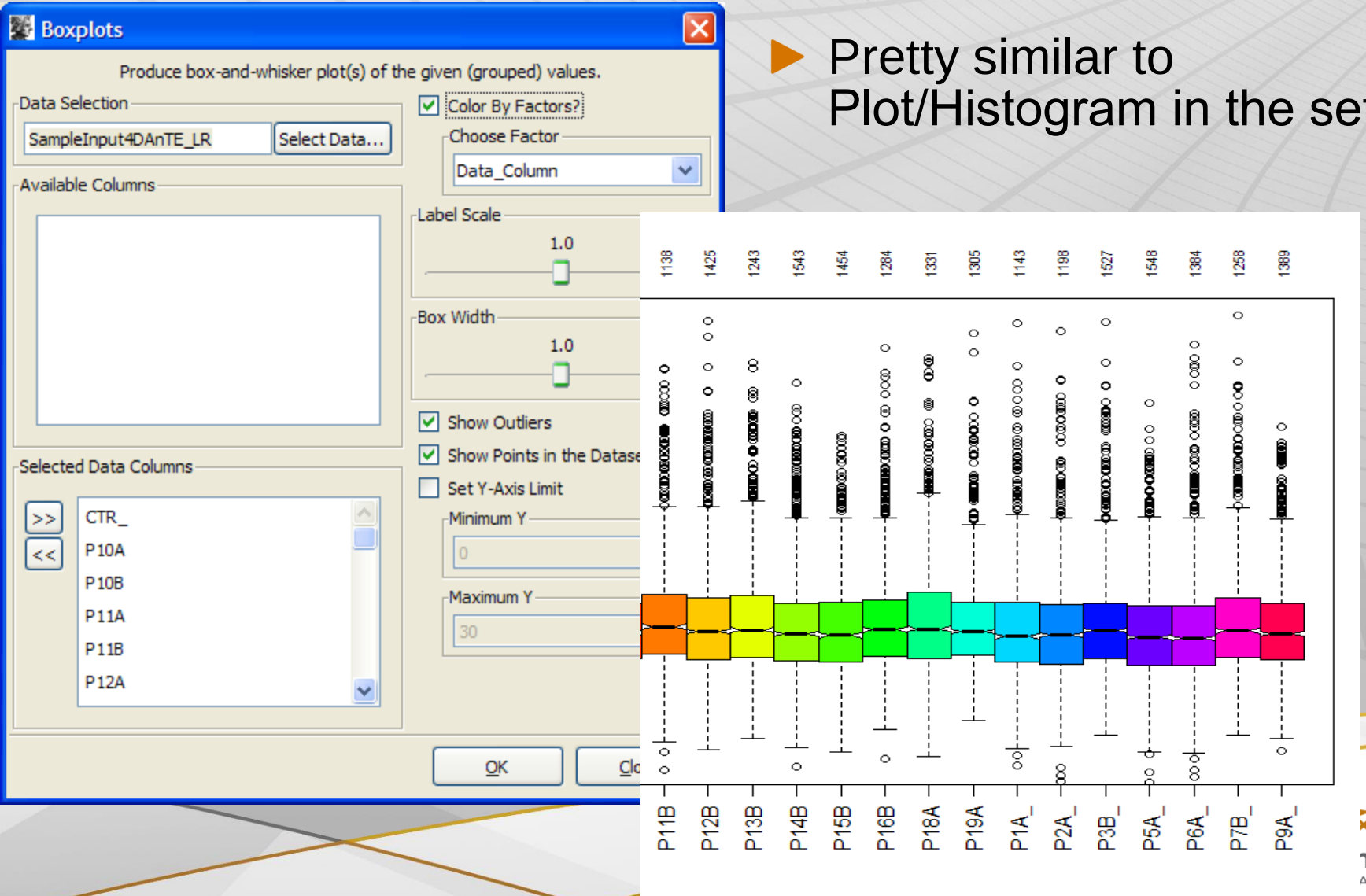



- Assess normality of a data set
- Choose a data crosstab, then select columns to QQ-plot
- Select reference distribution to compare to



Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Plot/Boxplots

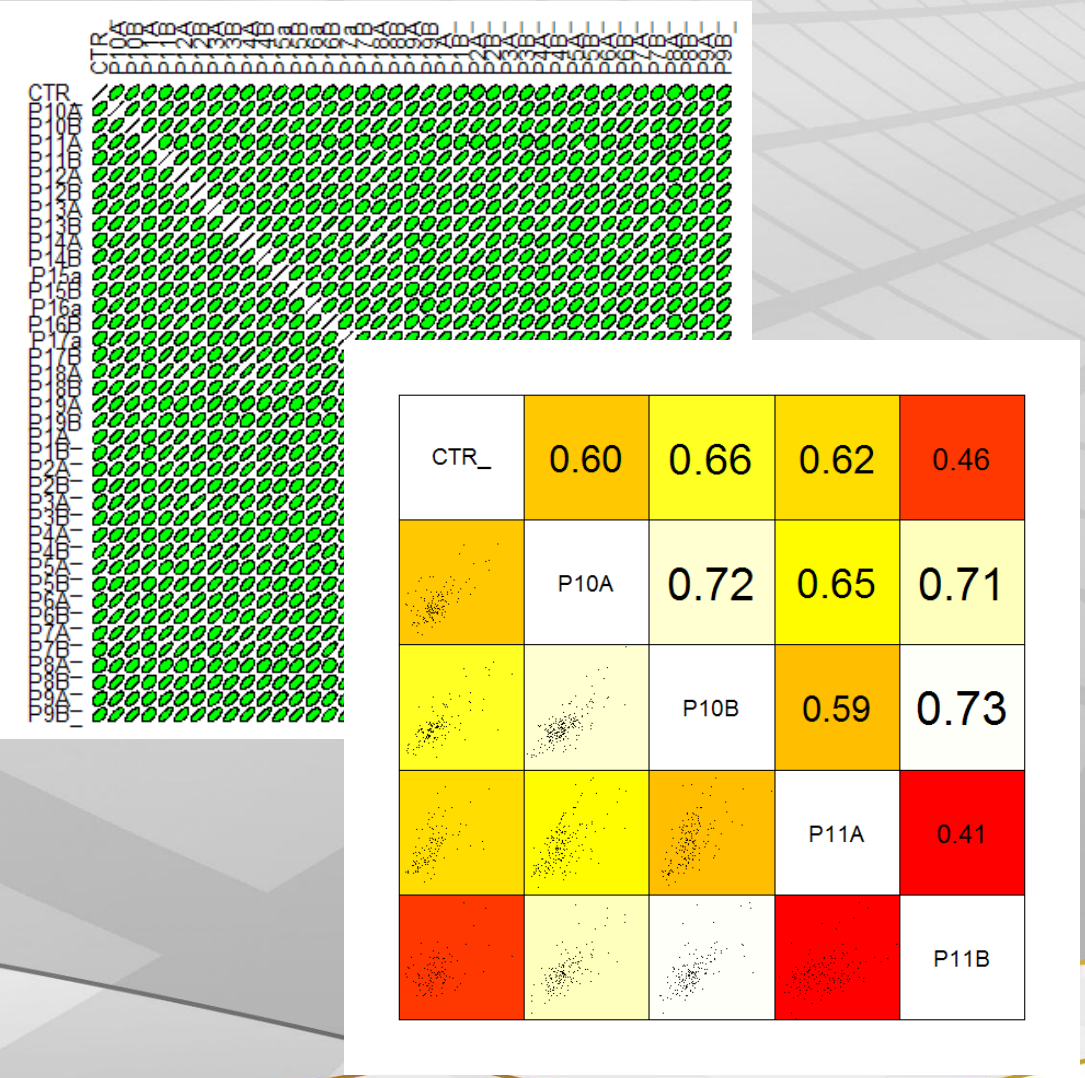



- Pretty similar to Plot/Histogram in the setup…



nwest
ABORATORY
Since 1965

# Plot/Scatter, Ellipses

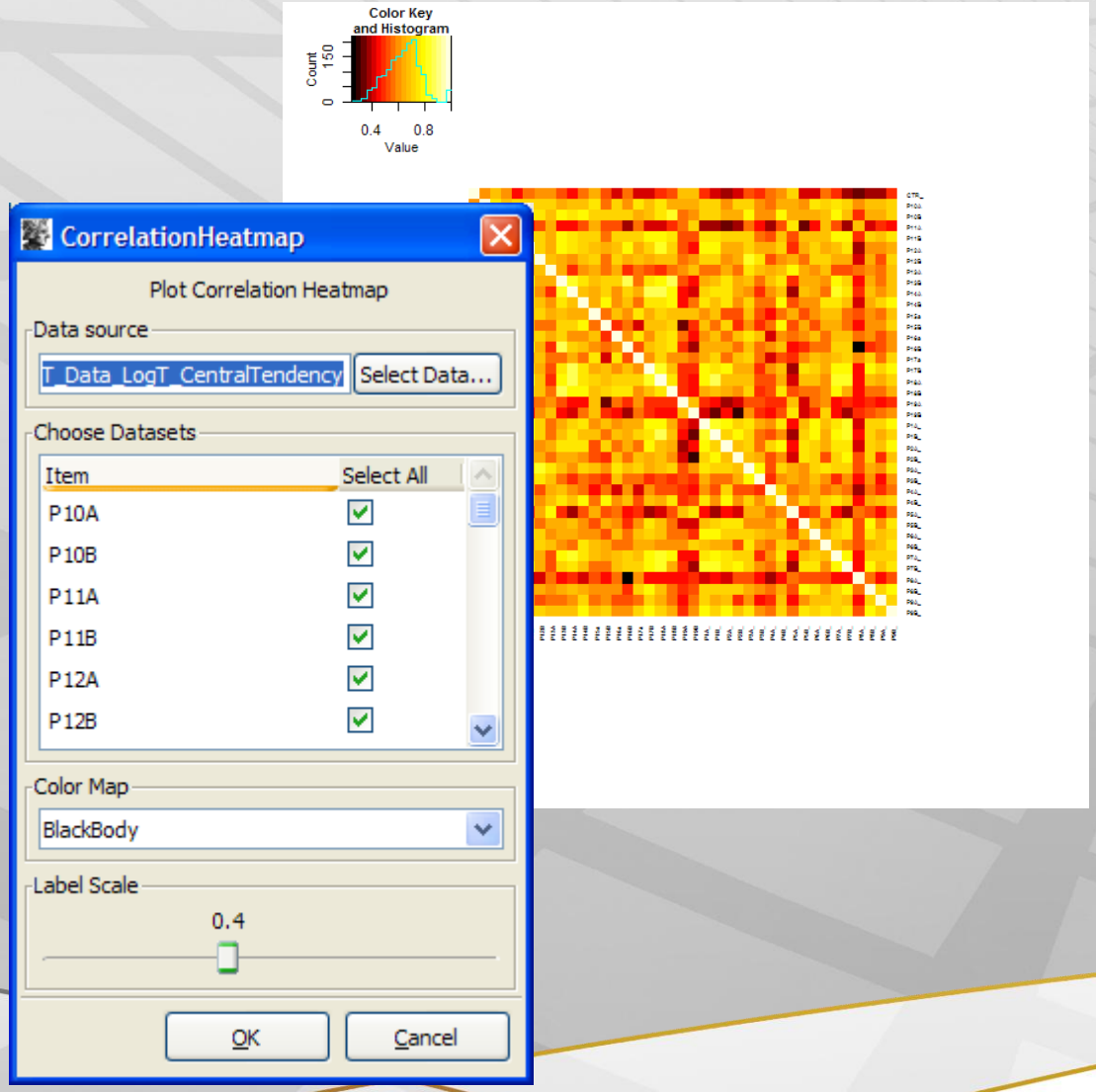



- Additional data set summaries



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Plot/Correlation heatmap

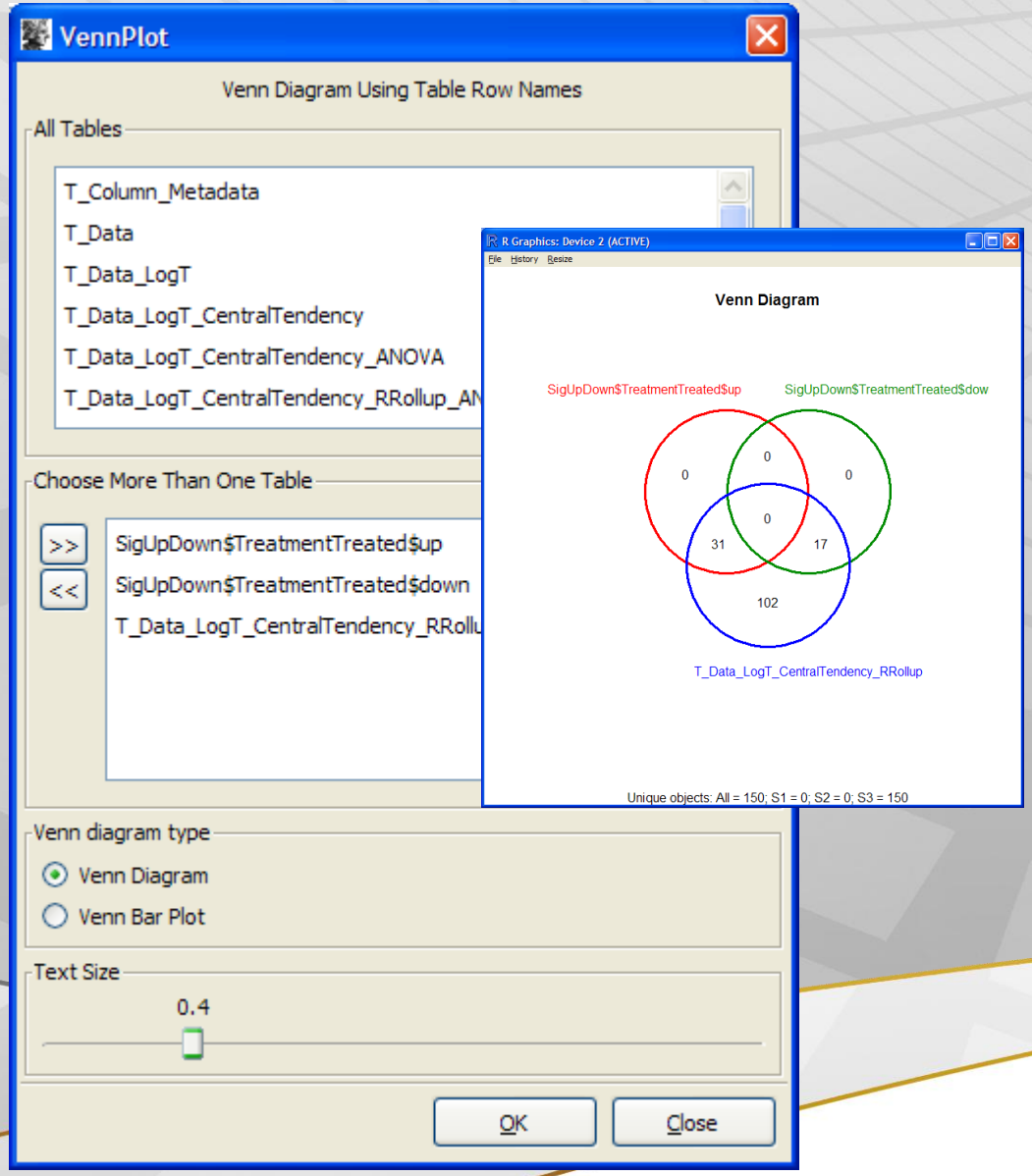



- Assess correlation between data set columns
- Choose a data crosstab, then select columns



Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Plot/Venn diagram plotting

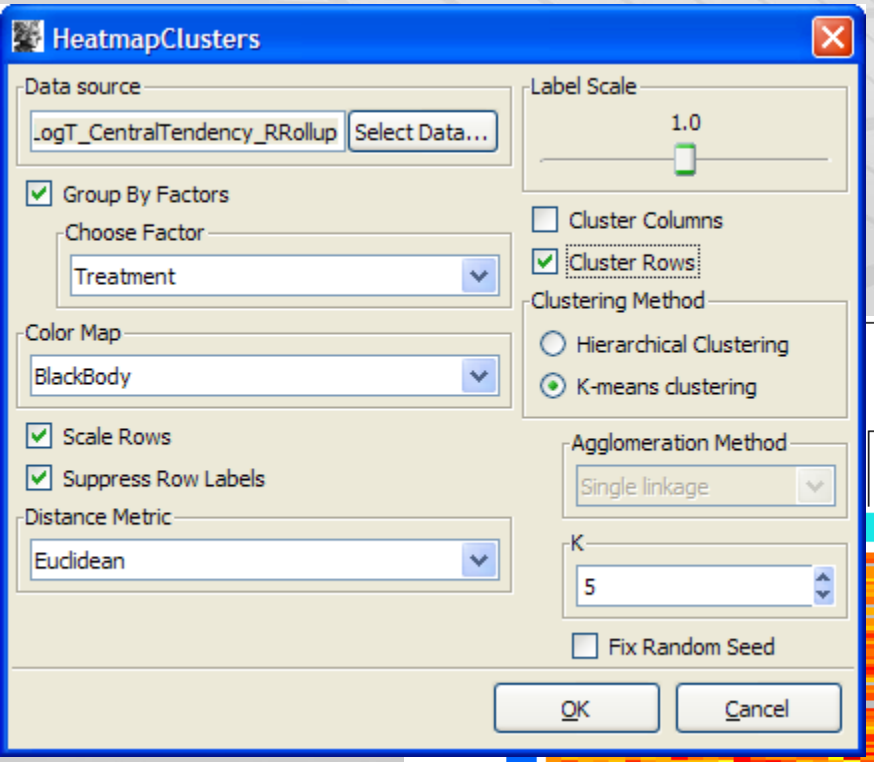


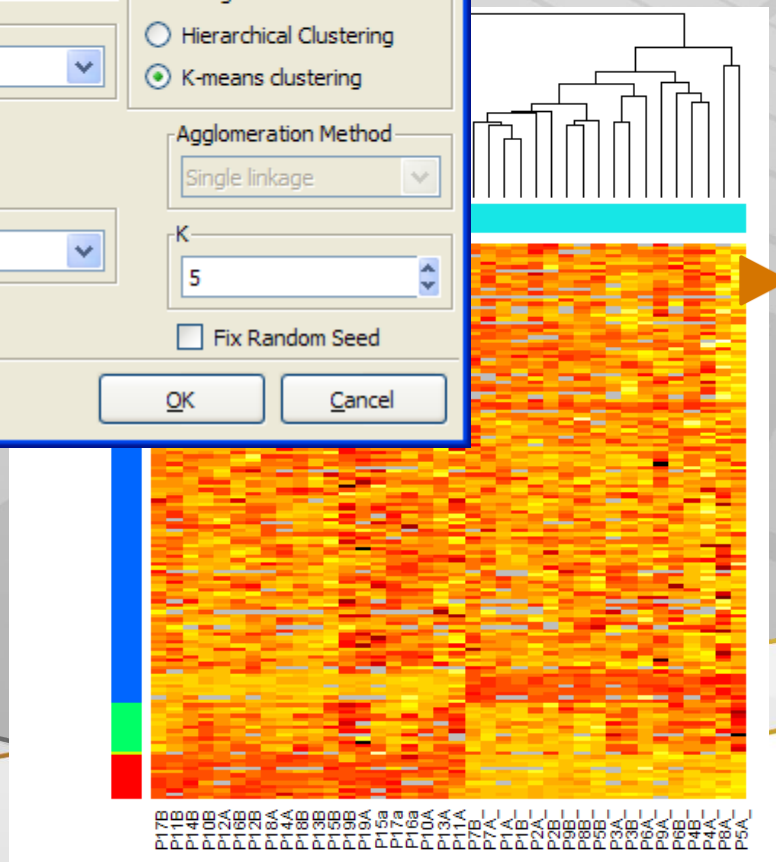

- Summarize data sets
- Choose up to 5 data tables
- Plots area-constant Venn diagrams using row names from each table



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Explore/Clustering



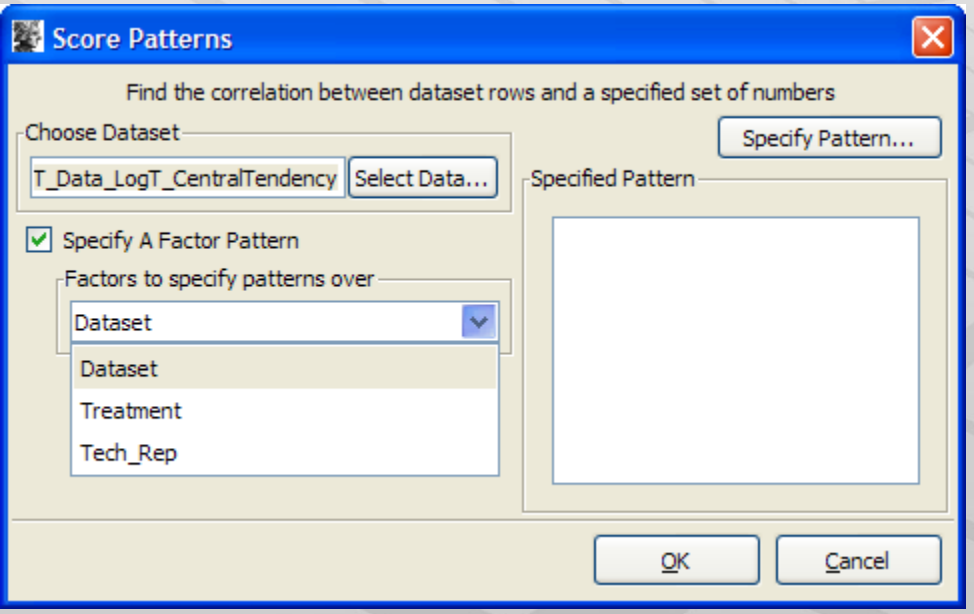



- Generate clusters using hierarchical or K-means clustering
- Requires data crosstab with linked column metadata (factors)
- Plots clusters in data and outputs a table containing row names and clusters.



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Explore/Pattern Scoring

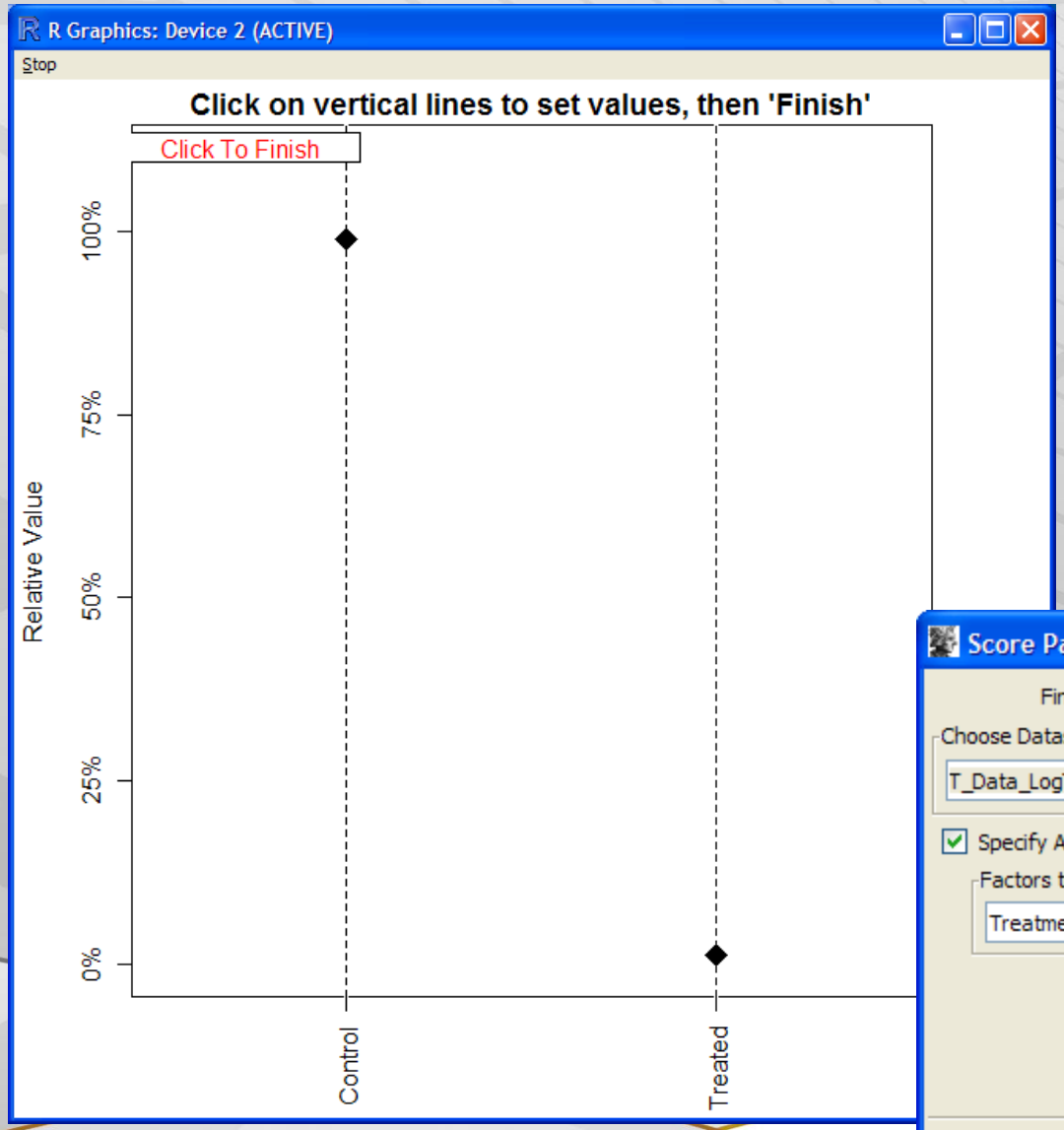



- Search for correlations between data and a specified pattern
- Requires a data crosstab
- If the table has associated factors/column metadata, you can search using “Specify A Factor Pattern”
- Click “Specify Pattern”

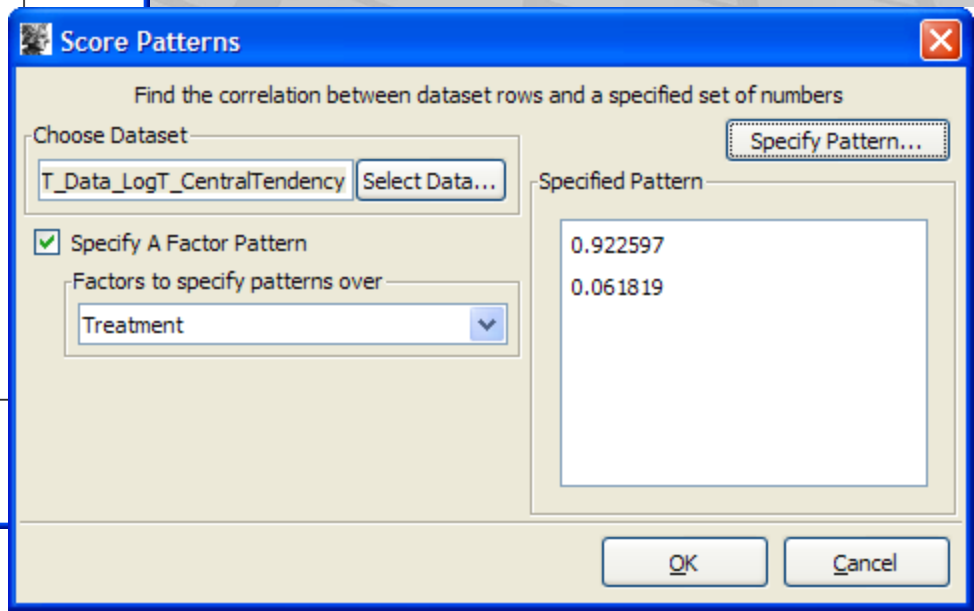


Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

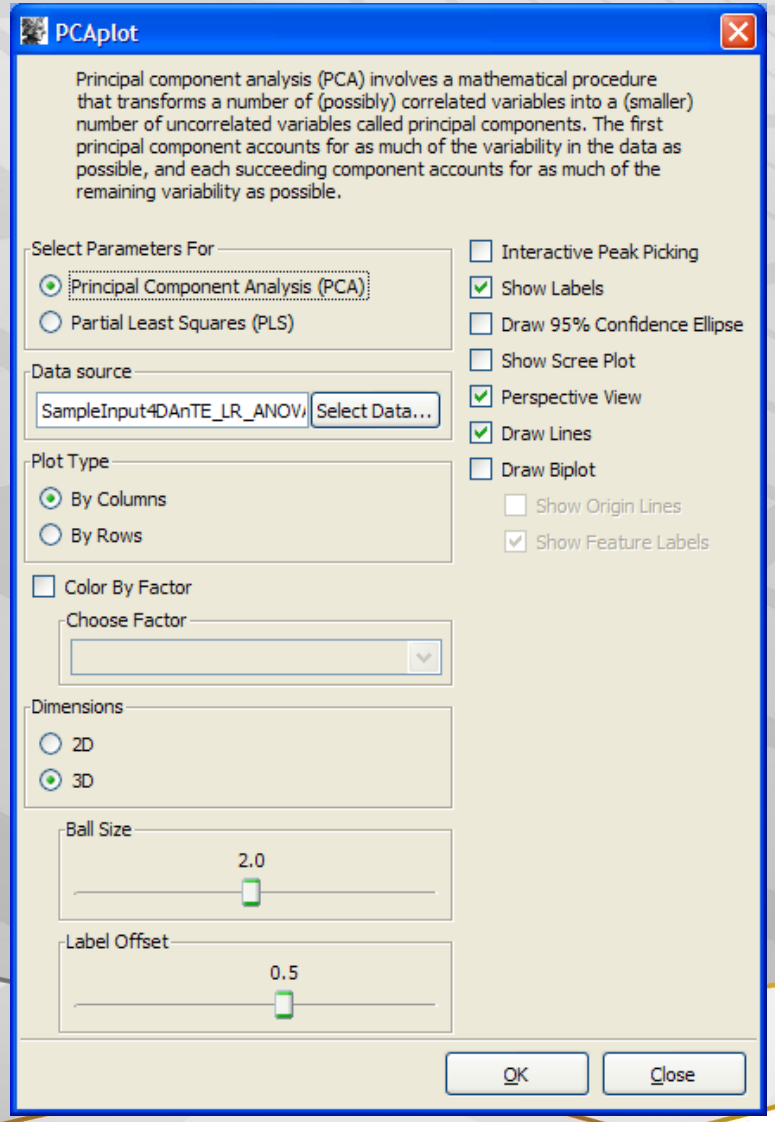
# Explore/Pattern Scoring




- Click on vertical lines to set values for each level of the factor you want to search, then click the “Finish box”, the pattern will appear in the list box, and OK.



# Explore/PCA

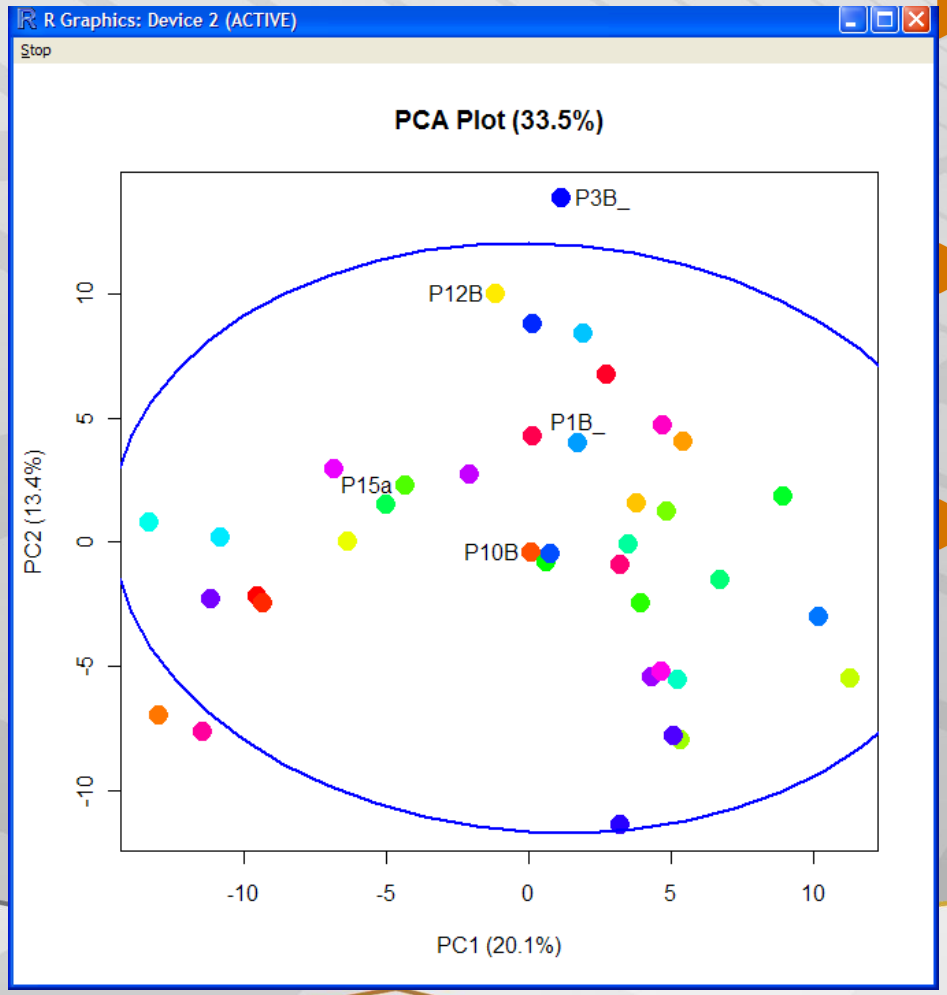



- PCA allows separation of data into “principal components” which account for the most variability.
- Options: columns or rows, color by factor, interactive peak picking, show ellipsoid
- Confidence ellipsoid shows outliers
- Biplots for showing contributions of components



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# 2D PCA

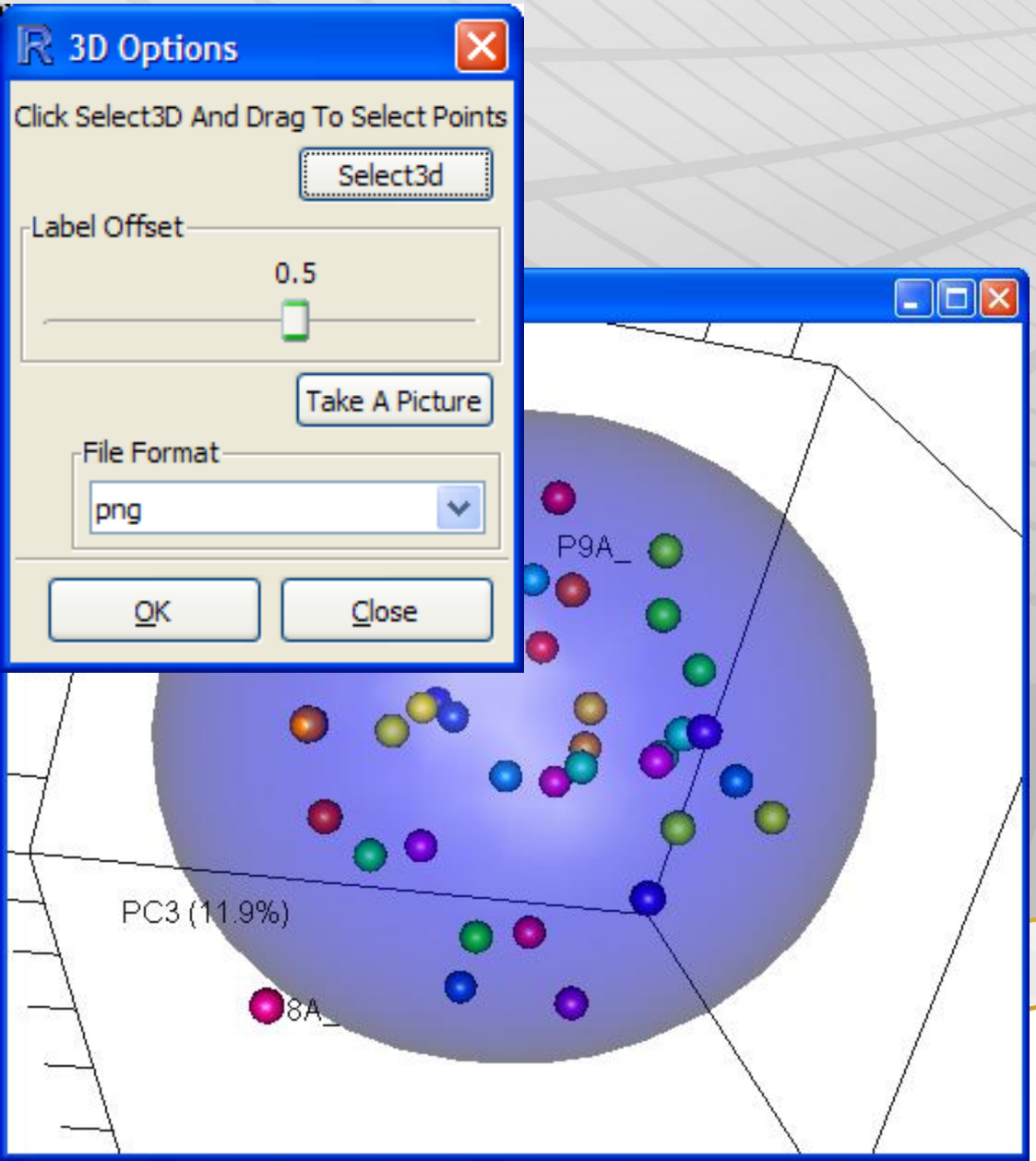



- Separates data rows or columns into principal components
- With interactive peak picking, click on a peak to show the label
- Right-click and choose Stop to or do this from the menu to stop peak picking



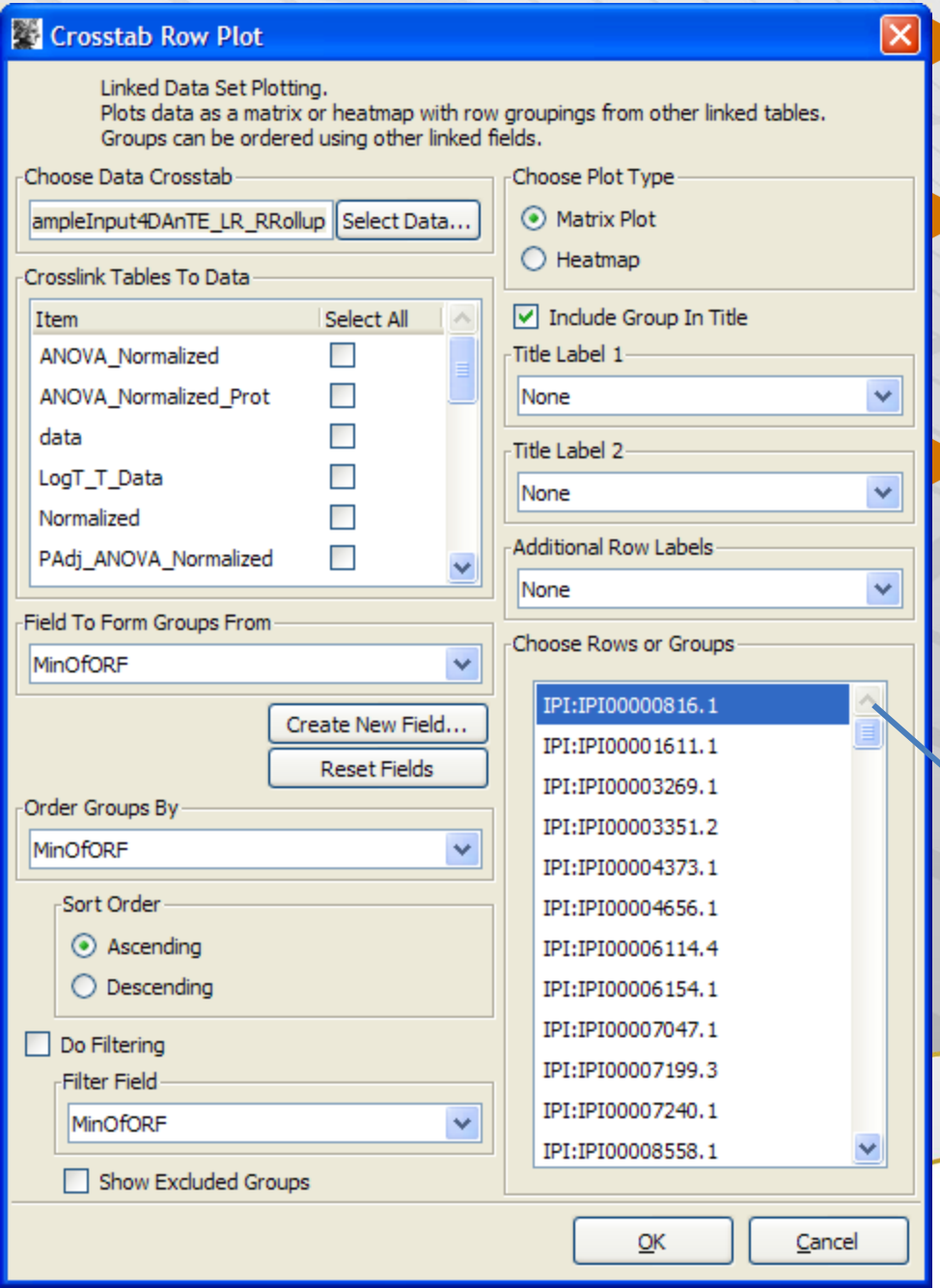
Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# 3D-PCA settings

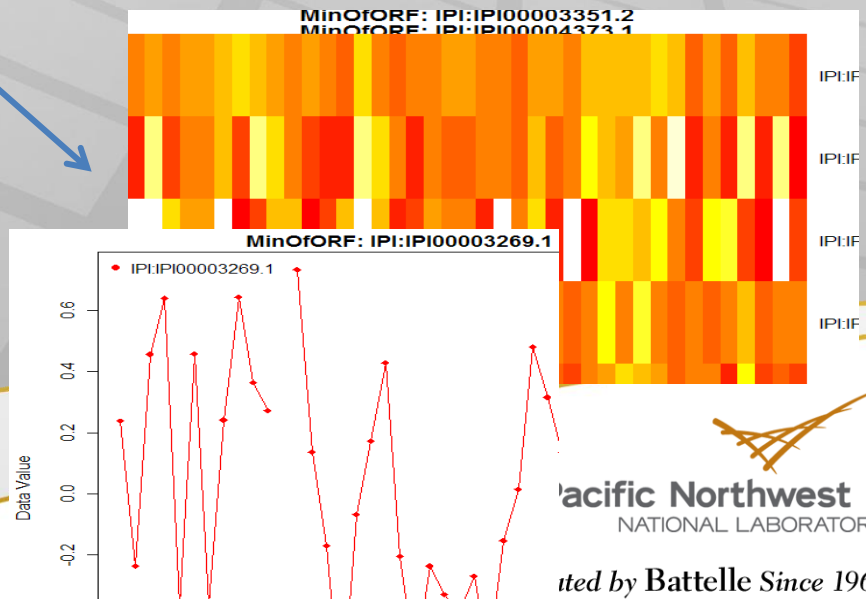



- Turn “3D” on for rotatable plot
- With interactive peak picking turned on, click “Select3D” in window to show labels in complex data
- Shows optional 95% confidence ellipsoid
- Picture save option

# Dynamic Row Plotting

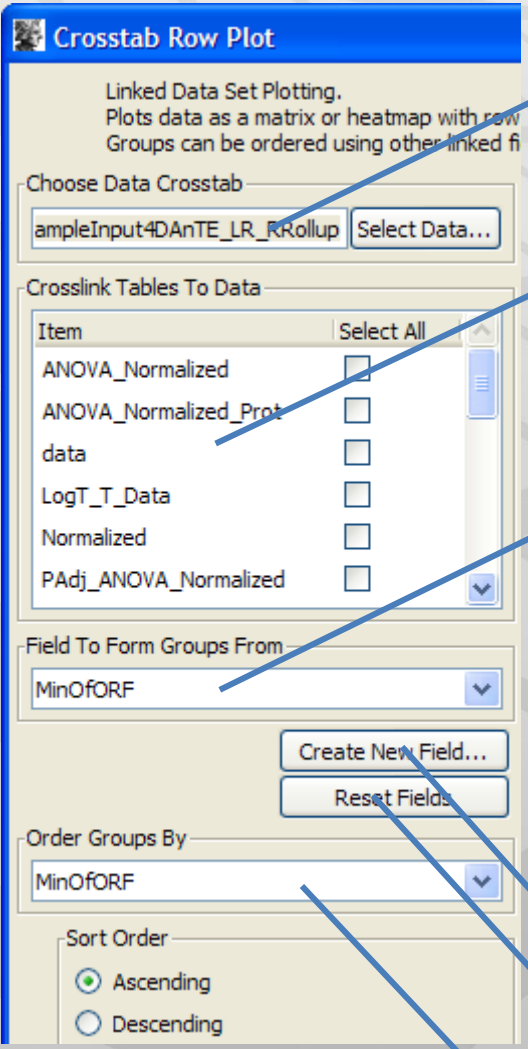



- This is where metadata starts getting really useful
- Uses data crosslinking to allow deep exploration of crosstab row data
- Plotting: Line plots or heatmaps (for now)




Pacific Northwest
NATIONAL LABORATORY
...ted by Battelle Since 1965

# Dynamic Row Plotting (2)

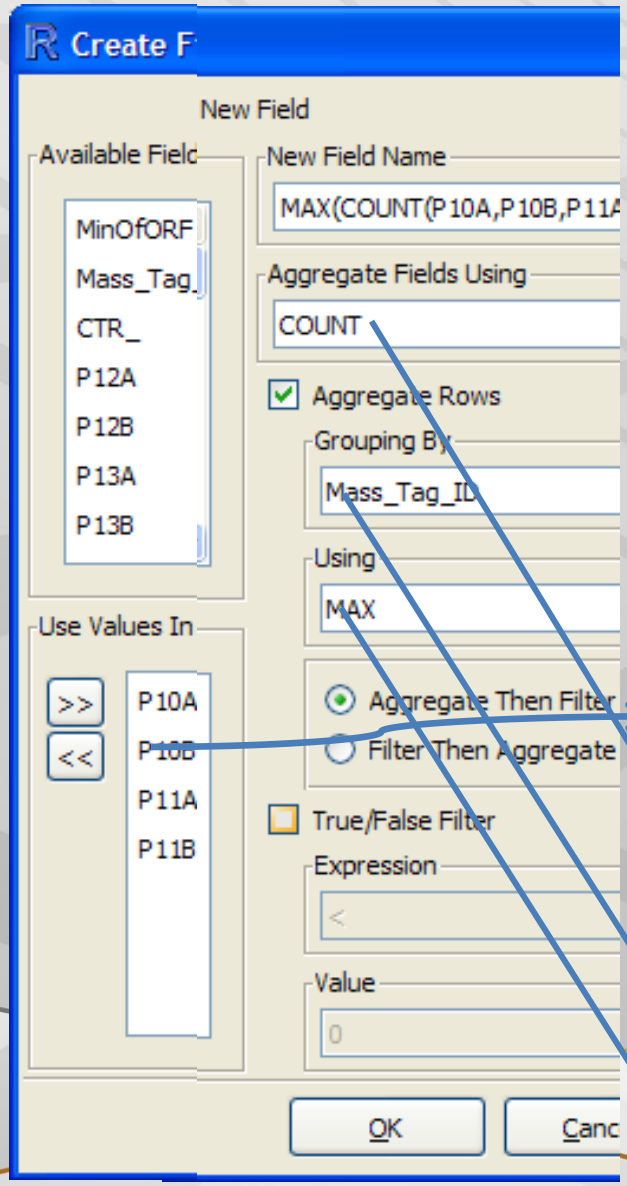



- The data you want to plot – a crosstab
- Additional tables you can crosslink to your data (they share row metadata with your data)
- Group: Collect crosstab rows together by common elements in a field (column). Fields are anything linked to your data – row metadata, something that shares that metadata, the data itself
- Form a new field from existing fields
- Remove any new fields
- Order the groups in the display by a field



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Explore/Dynamic Row Plot/Create Field




| Peptide | Exp A | Exp B | Exp C | Prot | Sum | Avg Sum A,B,C By Prot |
|---|---|---|---|---|---|---|
| 1 | 0.5 | 0.6 | 0.7 | P1 | 1.8 | 2 |
| 2 | 1.1 | 1.2 | 1.3 | P1 | 3.6 | 2 |
| 3 | 0.1 | 0.2 | 0.3 | P1 | 0.6 | 2 |
| 4 | -1 | -1.1 | -0.9 | P2 | -3 | -3 |

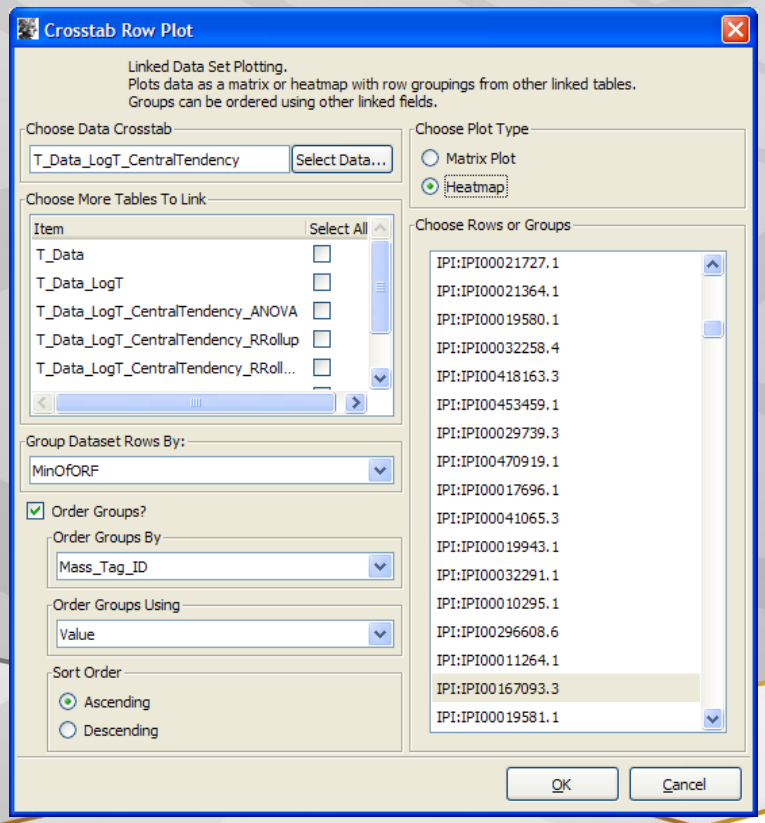
- Example: We want the Average Sum By Prot of fields A, B and C
- Use Values in A, B, C
- Aggregate Fields Using = SUM
- Aggregate Rows = TRUE
- Grouping By = Prot
- Using = AVG

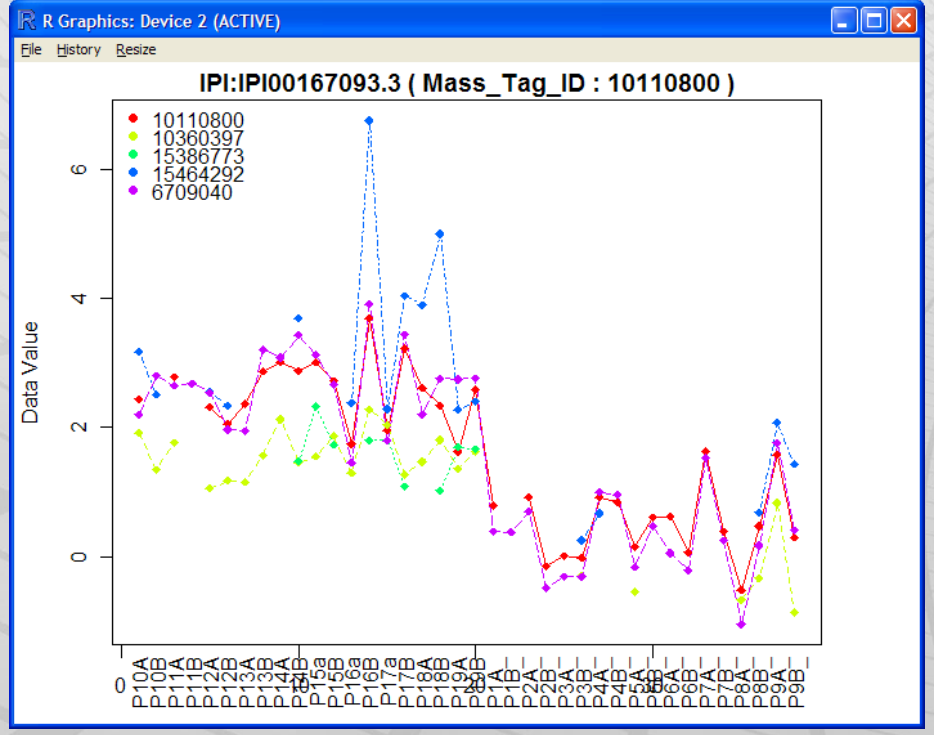


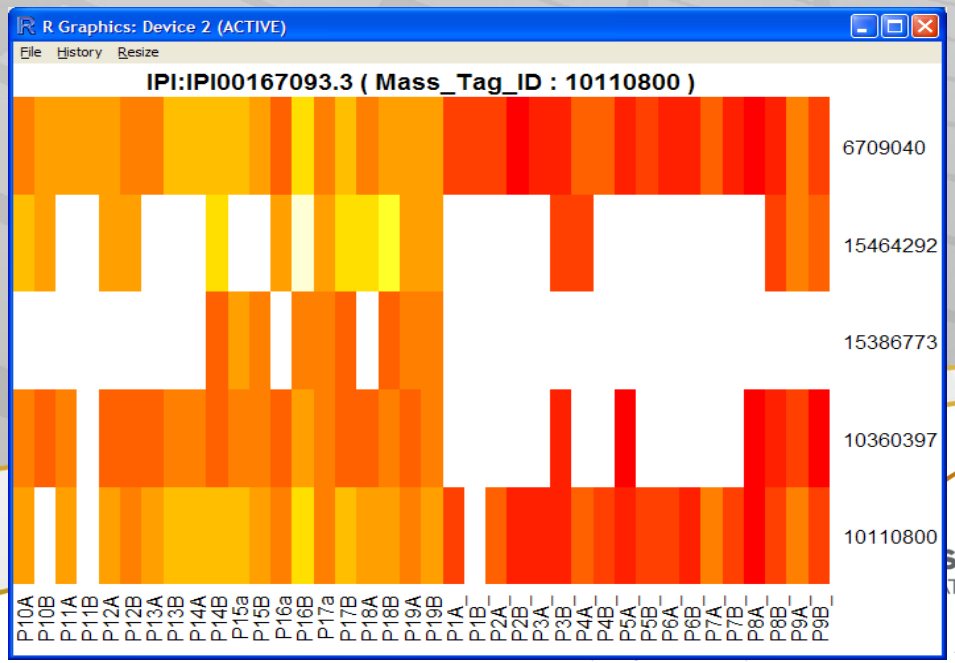
Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Simple row plot

Select Mass Tag crosstab data set and group by protein (MinOfORF field) for plotting protein by protein

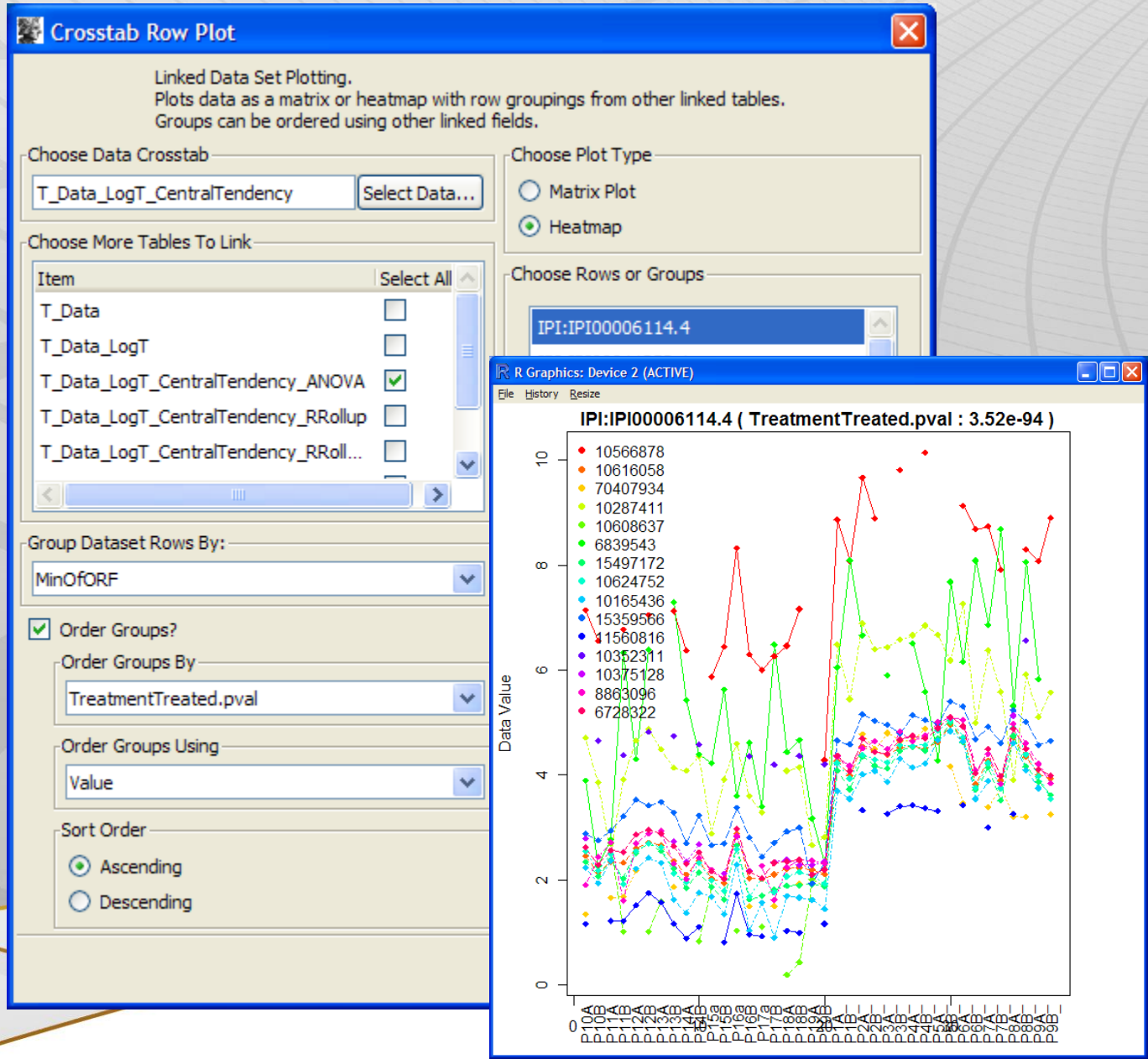


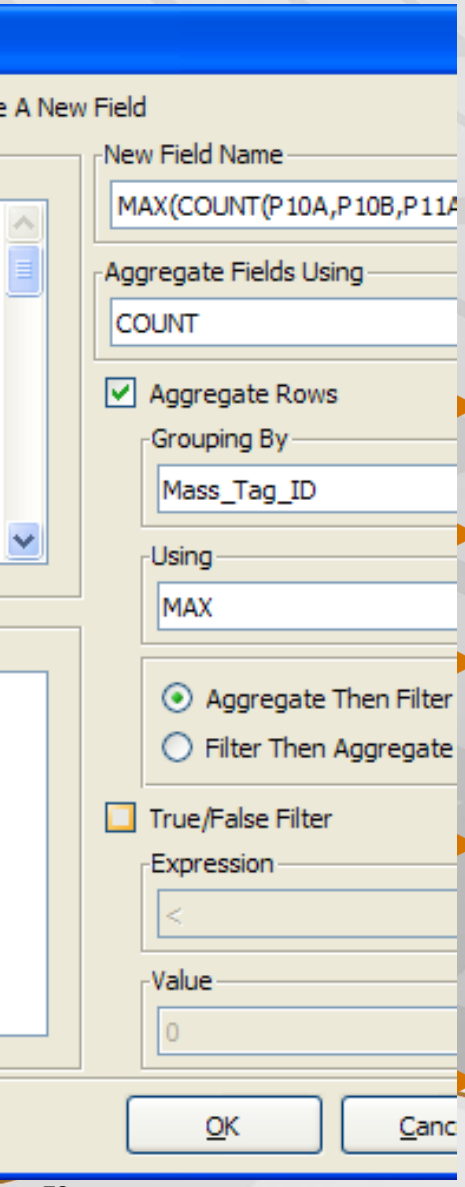



# Crosslinking ANOVA output to your data

- Select Mass Tag crosstab data set
- Crosslink to the statistical ANOVA result table
- Group by protein (MinOfORF field)
- Sort by ANOVA p-value field to highlight proteins with low p-values



# Another Create Field example with ANOVA




| Peptide | Prot | ANOVA P-value | SigP = p < 0.05 | SigPCount = COUNTIFTRUE SigP By Prot |
|---|---|---|---|---|
| 1 | P1 | 0.01 | TRUE | 2 |
| 2 | P1 | 0.05 | TRUE | 2 |
| 3 | P1 | 0.3 | FALSE | 2 |
| 4 | P2 | 0.1 | FALSE | 0 |

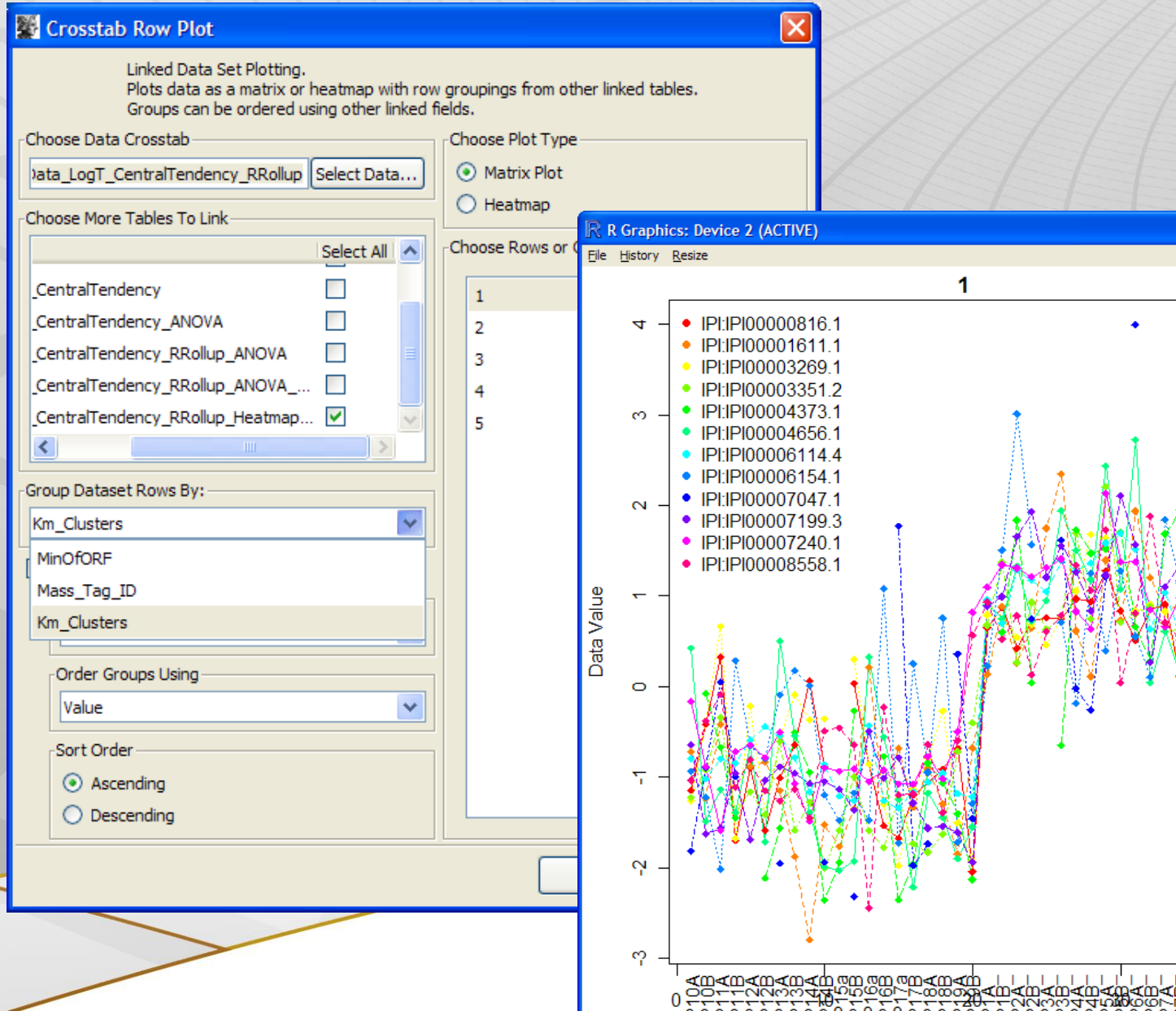
- Say you did ANOVA on all peptides and want to show a count of significant peptides per protein
- In the Row Plot dialog, load your peptide crosstab and select the ANOVA output table to crosslink with
- In the New Field window, create a new field SigP using Values In = ANOVA.pval, Aggregate Fields = Value, True/False Filter = True, Expression = “<“, Value = 0.05
- Reopen the New Field window, create a new field SigPCount using Values In = SigP, Aggregate Fields = Value, Aggregate Rows = True, Grouping By = Prot, Using = COUNTIFTRUE
- In the Row Plot window, group by Prot and select SigPCount as a title field.
- Filter by = SigP to ONLY display significant peptides.



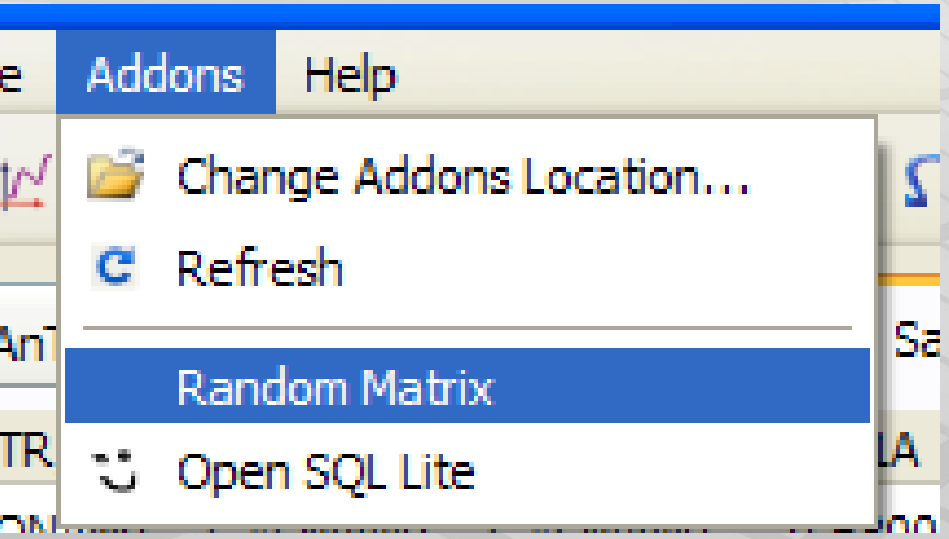
Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Crosslinking cluster output to your data

- Select Mass Tag crosstab data set
- Crosslink to the K-means cluster result table
- Group by K-means cluster field

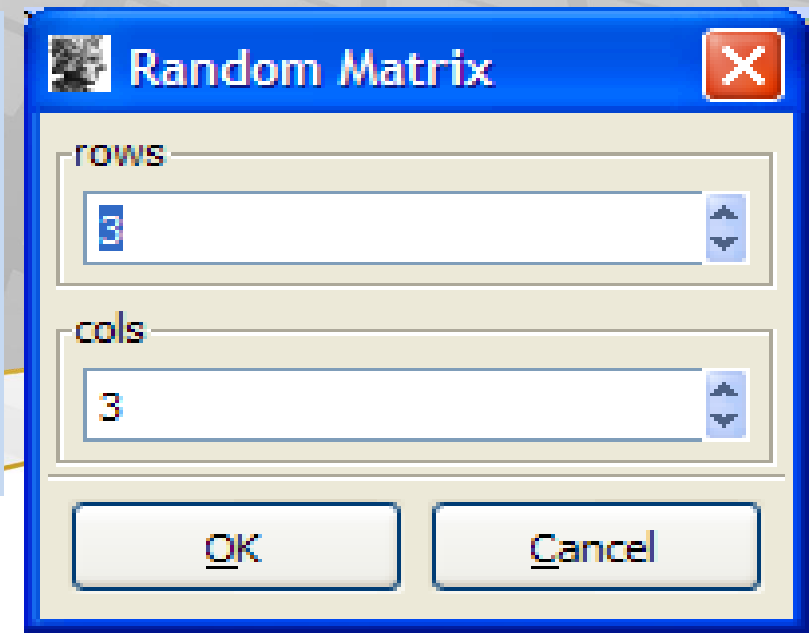


# Addons



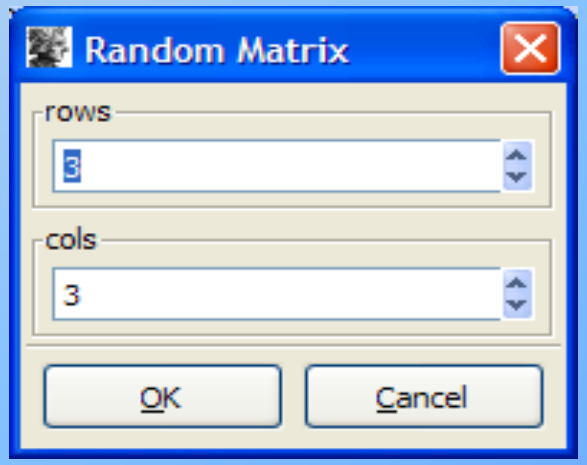

- DanteR will let you put your own R scripts in as menu items
- Create an R script and a dialog, then put it in the Addons folder and hit “Refresh”
- See the Dialog Maker Help

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

random.matrix.dialog = list(
  title = "Random Matrix",
  rows.integerItem = 3,
  cols.integerItem = 3
)
```



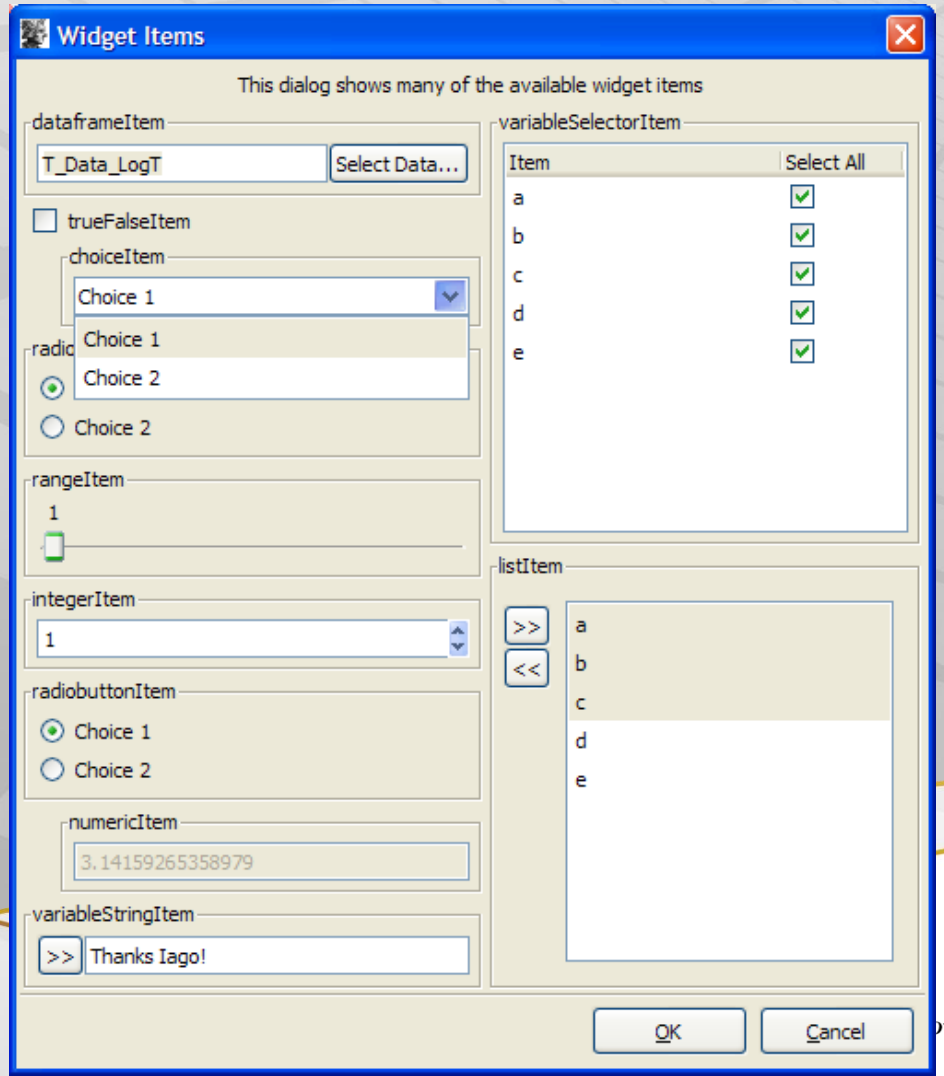
# DialogMaker examples

```
# A function to return a random matrix of some size
random.matrix = function(rows, cols)
  matrix(rnorm(rows*cols), nrow=rows)

# A dialog for random.matrix()
random.matrix.dialog = list(
  title = "Random Matrix",
  rows.integerItem = 3,
  cols.integerItem = 3)
randmat <- run.dialog(random.matrix)$retval
```



# DialogMaker markup

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



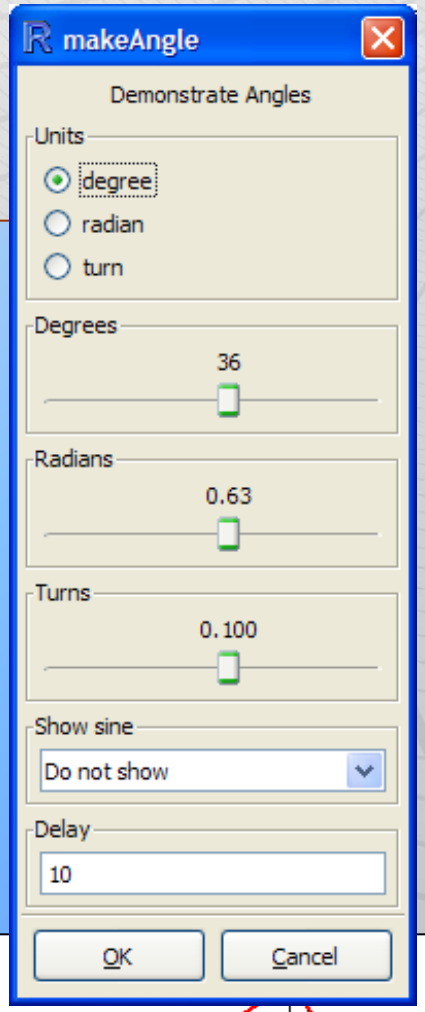


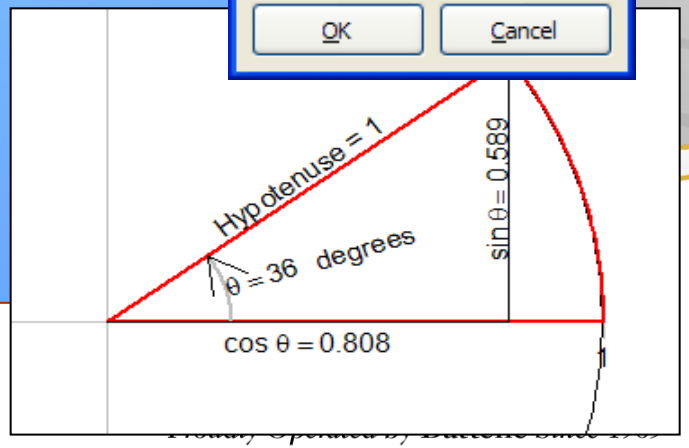

Pacific Northwest
NATIONAL LABORATORY
oudly Operated by Battelle Since 1965

# DialogMaker examples

## 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)
```



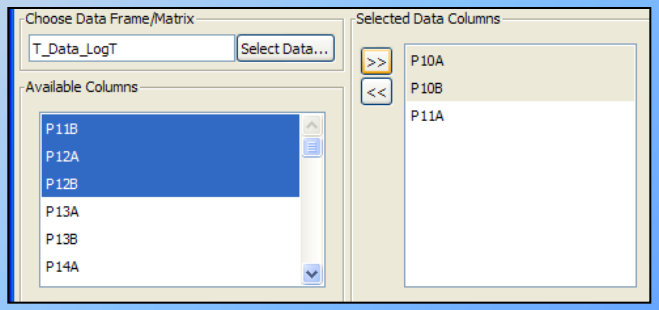



# DialogMaker signaling

## Commonly used callbacks

We wrapped a part of the GSignal API in the markup specification to allow specification of signals in dialog markup.

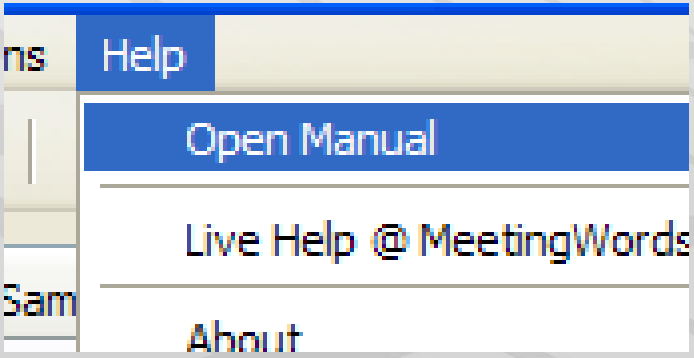
```
myFunction.dialog = list(...
<arg1>.<dialogItem> = value,
  signal = c("signal_type", <function or function name>,
arg2, arg3, ...), ...)

  # Fills item "dummy" with colnames of selected frame
x.dataframeItem = "",
  signal = c("default", "get.colnames", "dummy")
  # Create 2 listItems and allow push/pop of selection
dummy.listItem = NULL, show.arrows = F, suppress = T,
Columns.listItem=NULL, label="Selected Data Columns",
    signal = c("add", "push.selection", "dummy"),
    signal = c("subtract", "pop.selection", "dummy")
```

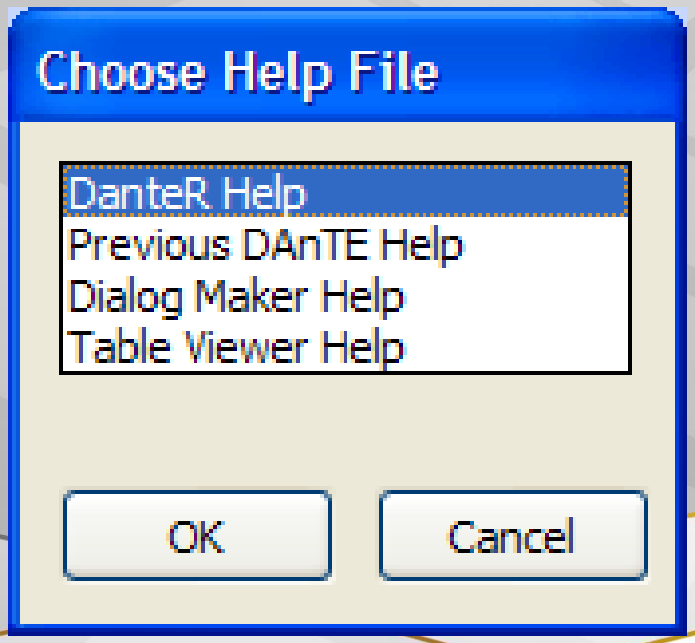





Pacific Northwest
NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Help




- Open Manual lets user choose PDF of DanteR Manual, original DAnTE, reference manual for writing dialogs or reference manual for Table Viewer



Pacific Northwest NATIONAL LABORATORY
Proudly Operated by Battelle Since 1965

# Acknowledgements


- PNNL: Beta testers: John Sandoval, Joe Brown, Tom Angel, Alex Rutledge
- Gordon Anderson, Dick Smith's group
- Others: Alan Dabney and group
- Ashoka Polpitiya
- Iago Conde (Madrid), John Verzani (SUNY), Graham Williams (ATO) and Michael Lawrence (Genentech)
- The R community