KNIME SUMMIT 2016

Antje Janosch – Technology Development Studio
MPI-CBG Dresden

Scripting Integration Workshop
History

2009
2010
2013 / 2014
2016

R-Scripting, Python Scripting, Matlab Scripting

R Snippet
Python Script

Antje Janosch, MPI-CBG, Dresden, Germany
General Features

• Server/Client based solution
  • Parallel execution
  • Centralized setup

• Remote template repositories

• Template configuration via “KNIME-like” GUI

• Deployment of KNIME-nodes with scripting behind
Short news

R

- RowID support (both directions)
- Domain values
- Color / Size / Shape model support
- Flow variable support
- Chunked transfer
- ‘Open external’
- R-workspace as generic R port
Client-(server) model

RServe
remote or local

Python Snippet
local
remote or local

Python-Server
local
tremote or local

Matlab-Server
remote or local

R Snippet

Matlab Snippet
Getting started (Client)

• Install Scripting plugins from “Trusted Community Contributions” update site (KNIME >= 3.0 recommended)
Node Types

- Plot
- Snippet
- (OpenIn...) 
- Generic nodes (R)
Plot / Snippet – Template tab

This rank normalization project each group of values per parameter to a normal distribution.
- select type of normalization
- select a group column that should be used for ranking (e.g. barcode, date, project name,...)
- select parameters that should be rank normalized

Author: Martin Stoeter
Provides User Interface: 

Preview (Double-click to enlarge)
Template Repository

This rank normalization project each group of values per parameter to a normal distribution.
- select type of normalization
- select a group column that should be used for ranking (e.g. barcode, date, project name, ...)
- select parameters that should be rank normalized

Preview (Double-click to enlarge)

Author: Martin Stoeter
Provides User-Interface: 

Use this template
Plot / Snippet – Template tab

Template Repository

Description

This rank normalization project each group of values per parameter to a normal distribution.
- select type of normalization
- select a group column that should be used for ranking (e.g. barcode, date, project name,...)
- select parameters that should be rank normalized

Preview (Double-click to enlarge)

Author: Martin Stoeber
Provides User-Interface: ✅

Use this template
Plot / Snippet – Template tab

Template Repository

Template Categories
- pre-processing
- utilities
- Power transformation and Z rank
- Create column name table from
- screenmining
- utils

Description

This rank normalization project each group of values per parameter to a normal distribution.
- select type of normalization
- select a group column that should be used for ranking (e.g. barcode, date, project name,...)
- select parameters that should be rank normalized

Preview (Double-click to enlarge)

Author: Martin Stoeter
Provides User-Interface:

Use this template

OK Apply Cancel

Refresh Gallery
Plot / Snippet – Template tab

Template Repository

Template Description / Source
Script editor tab – template view

Power transformation and Z rank transformation

Description
This rank normalization project each group of values per parameter to a normal distribution.
- select type of normalization
- select a group column that should be used for ranking (e.g., barcode, date, project name, ...)
- select parameters that should be rank normalized

Options

Power transformation
Box-Cox

Group column
barcode

Column selection

Options
Column(s): Search
Select all search hits
PlateRow
plateColumn
Median Nuclei Intensity
Calculation time (seconds)
Number of Cells
Ch2 Median of Maximum
Ch1 Median of Maximum
Median Syto Intensity
Univariate number

Selection
Column(s): Search
Select all search hits
<< remove
<< remove all
Script editor tab – template view

RGG interface of the selected template
Script editor tab – template view

RGG interface of the selected template
Script editor tab – template view

RGG interface of the selected template
Script editor tab – template view

RGG interface of the selected template

#1.0 Parameter selection
# 1.1 select type of transformation
# transformation = "Box-Cox";
# 1.2 select group
group = "barcode";
#group = "barcodePlate"
#subGroup = "library"
# 1.3 select the columns
selectedParams = match(c(), names(kIn));
#selectedParams = names(kIn[,13:18]);
Script editor tab – template view

Power transformation and Z rank transformation

Description
This rank normalization project each group of values per parameter to a normal distribution.
- select type of normalization
- select a group column that should be used for ranking (e.g. barcode, date, project name,...)
- select parameters that should be rank normalized

Options

Power transformation
Box–Cox

Group column
barcode

Column selection

Options
Column(s): Search
Select all search hits
plateRow
plateColumn
Median Nuclei Intensity
Calculation time (seconds)
Number of Cells
Ch2 Median of Maximum
Ch1 Median of Maximum
Median Syto Intensity
Barcode number

Selection
Column(s): Search
Select all search hits

Unlink from Template  Edit Template
Power transformation and Z rank transformation

Description
This rank normalization project each group of values per parameter to a normal distribution.
- select type of normalization
- select a group column that should be used for ranking (e.g. barcode, date, project name, ...)
- select parameters that should be rank normalized

Options
Power transformation

Group column

Column selection

Options
Column(s): Search
Select all search hits

Selection
Column(s): Search
Select all search hits

Plate Row
Plate Column
Median Nuclei Intensity
Calculation time (seconds)
Number of Cells
Ch2 Median of Maximum
Ch1 Median of Maximum
Median Syto Intensity
Recognition number

Unlink from Template Edit Template

modify template (dev)
Script editor tab – template view

modify template (dev)

Antje Janosch, MPI-CBG, Dresden, Germany 23
Tips and Tricks for Editing (R)

- Mouse click = ”Column name”
- Alt + Mouse click = kIn$”Column name”
- Ctrl + Mouse click = Displays possible domain values of the column and offers to insert a selection (comma separated)
- Press Apple/Windows key and select multiple = as soon as you release the key, the selected column names will be inserted “column 1”, ”column 2”, ...

- FLOWVAR(currentColumnName) placeholder to enable the use of flow variables
Plot – Output Options tab

Placeholders for filename

$$\text{DATE}$$ - yymmdd
$$\text{USER}$$ - user name
$$\text{WS}$$ - knime.workspace
FLOWVAR(varName) – flow variable
• Dynamic repainting after mouse click
KNIME tables in R

• Input ports (kIn or kIn1, kIn2, ...)  
  • R data.frame  
  • RowIDs as row names  
  • Domain values as factor levels

• Color model (kIn.color.model)  
  • R data.frame  
  • Colors as hex-values

• Shape model (kIn.shape.model)  
  • R data.frame  
  • Shapes as pch and their KNIME name

<table>
<thead>
<tr>
<th>Petal Width</th>
<th>color</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.1 #000000ff</td>
</tr>
<tr>
<td>2</td>
<td>2.5 #33ff33ff</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Species</th>
<th>shape</th>
<th>pch</th>
</tr>
</thead>
<tbody>
<tr>
<td>virginica</td>
<td>Diamond</td>
<td>18</td>
</tr>
<tr>
<td>setosa</td>
<td>Circle</td>
<td>16</td>
</tr>
<tr>
<td>versicolor</td>
<td>Triangle</td>
<td>17</td>
</tr>
</tbody>
</table>
• Size model (\texttt{knIn.size.model + knIn.size.model.fun})
  • Name of the column having a size model
  • Function to calculate the size based on the KNIME size model

• Flow variables (\texttt{knime.flow.in})
  • R list
KNIME tables in R

- Size model (`knIn.size.model + knIn.size.model.fun`)
  - Name of the column having a size model
  - Function to calculate the size based on the KNIME size model

```r
function (v)
{
  (((v^2 - 4.3^2)/(7.9^2 - 4.3^2)) * (6 - 1)) + 1
}
```

- Flow variables (`knime.flow.in`)
  - R list
KNIME tables in R

- Size model (\texttt{knIn.size.model + knIn.size.model.fun})
  - Name of the column having a size model
  - Function to calculate the size based on the KNIME size model

```r
function (v)
{
  ((v^2 - 4.3^2)/(7.9^2 - 4.3^2))
}
```

- Flow variables (\texttt{knime.flow.in})
  - R list
This is my template description.

R/Python/Matlab code
Optionally embedded in RGG

Another very cool method
Antje Janosch
...
RGG

• R GUI Generator

http://rgg.r-forge.r-project.org/

• Additional element

<panellistbox label="Label"
items="item1,item2,item3"
span="full"/>

<rgg>
some interface elements
definitions
<![CDATA[
R-script
]]>
</rgg>
• Placeholders

$$\text{NUM\_ATTRIBUTES}$$ - for all numeric columns
$$\text{STR\_ATTRIBUTES}$$ - for all string columns
$$\text{ALL\_ATTRIBUTES}$$ - for all columns
Demos (R)

1. I want to use templates
2. I want to write my own R-scripts
3. I want to create knimeR-workflows (generic nodes)
4. I want to create templates
5. I want to create a KNIME node with R behind

Antje Janosch, MPI-CBG, Dresden, Germany
Outlook

• Support of row IDs for Python/Matlab and for both directions
• Console output view for R nodes
• Keep offline copy of template files
• ...
Contact us!

antje.janosch@mpi-cbg.de
stoeter@mpi-cbg.de