

Installation

1. Install KNIME Analytics Platform (from thumb drive)
2. Help > Install New Software > Add (> Archive):
 - 00_InstallationFiles/CommunityContributions_trunk.zip
 - <https://update.knime.org/community-contributions/trunk>
3. Available Software Sites:
 - Enable KNIME AP 3.5 Update Site
 - Enable KNIME Community Contributions (trunk)
4. Select *KNIME Community Contributions (trunk)*
5. Install *KNIME Image Processing* **and** *KNIME Image Processing - OMERO Integration*



OMERO and KNIME Workshop

Stefan Helfrich, Christian Dietz

KNIME

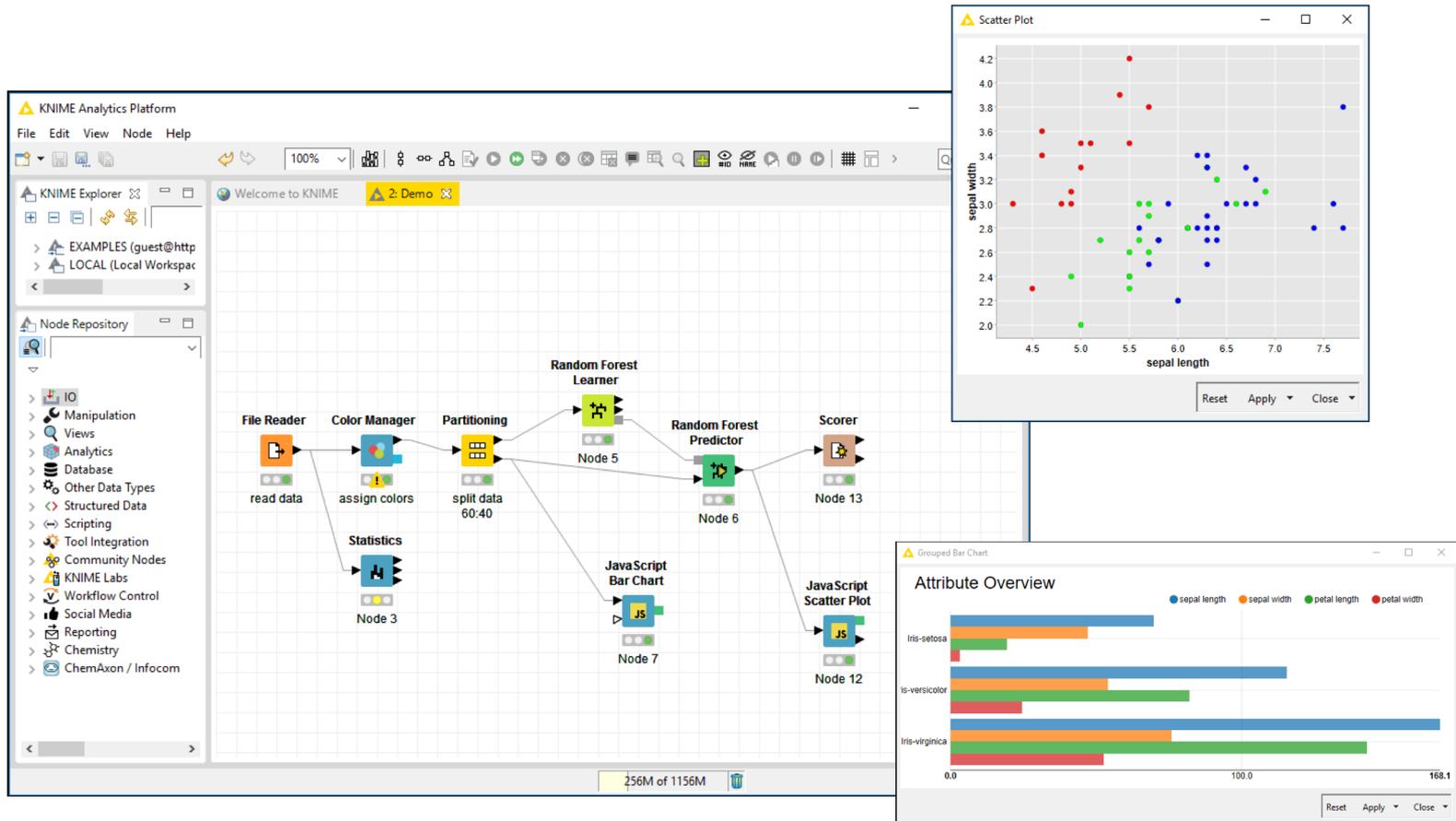
Learning goals

- How to get around KNIME Analytics Platform (AP)
- How to get your images from OMERO into KNIME AP
- How to extract quantitative data from your images
- How to use analytics and visualization functionality in KNIME AP

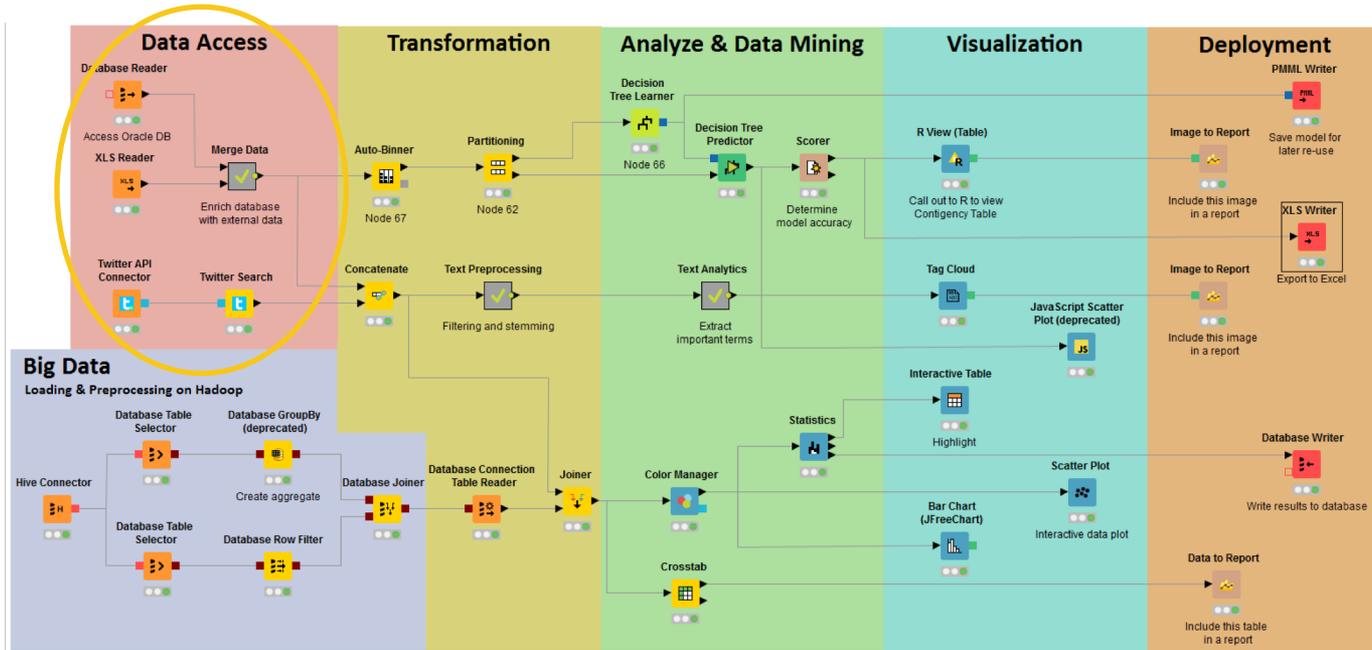
What is KNIME Analytics Platform?

- A tool for data analysis, manipulation, visualization, and reporting
- Based on the graphical programming paradigm
- Provides a diverse array of extensions:
 - Text Mining
 - Network Mining
 - Cheminformatics
 - Image Informatics
 - Many integrations, such as Java, R, Python, Weka, etc.

The KNIME[®] Analytics Platform



Over 2000 native and embedded nodes included:



Data Access

MySQL, Oracle, ...
 SAS, SPSS, ...
 Excel, Flat, ...
 Hive, Impala, ...
 XML, JSON, PMML
 Text, Doc, Image, ...
 Web Crawlers
 Industry Specific
 Community / 3rd

Transformation

Row,
 Column
 Matrix
 Text, Image
 Time Series
 Java
 Python
 Community / 3rd

Analysis & Mining

Statistics
 Data Mining
 Machine Learning
 Web Analytics
 Text Mining
 Network Analysis
 Social Media Analysis
 R, Weka, Python
 Community / 3rd

Visualization

R
 JFreeChart
 JavaScript
 Community / 3rd

Deployment

via BIRT
 PMML
 XML, JSON
 Databases
 Excel, Flat, etc.
 Text, Doc, Image
 Industry Specific
 Community / 3rd

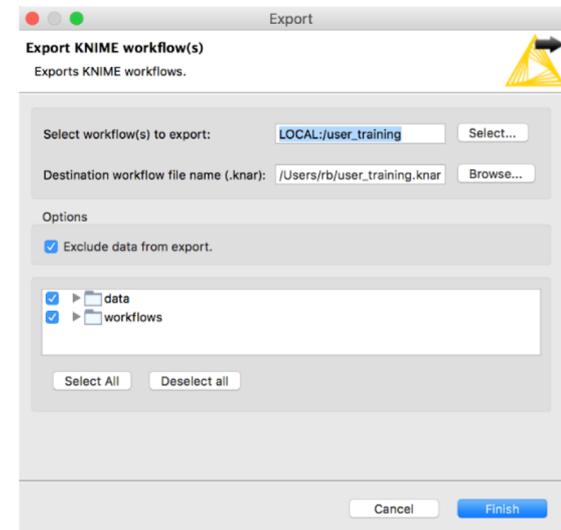
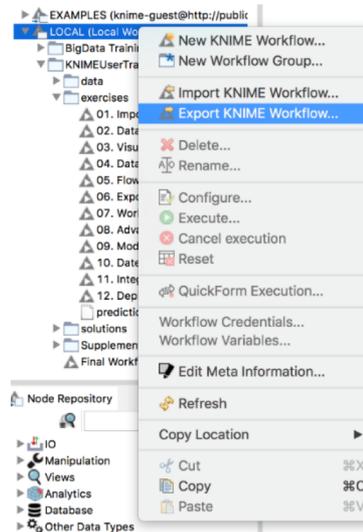
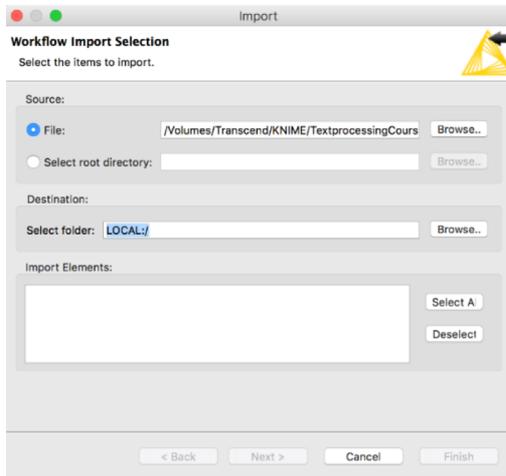
The KNIME Workbench

The screenshot displays the KNIME Analytics Platform interface with several key components highlighted by yellow callout boxes:

- Servers and Workflows:** Points to the KNIME Explorer on the left, which shows a project tree with folders like 'LOCAL (Local Workspace)', 'BigData Training', and 'KNIMEUserTraining', and a list of solutions including '01. Importing Data', '02. Data Manipulation', '03. Visualization', '04. Data Mining', '05. Flow Variables', and '06. Exporting Data'.
- Workflow Editor:** Points to the central workspace where a workflow is being built. The workflow starts with 'Fully Joined Data', followed by 'Partitioning', 'Decision Tree Learner', 'Decision Tree Predictor', 'Scorer', 'JavaScript ROC Curve', 'Node 278', and 'ROC Curve'.
- Node Recommendations:** Points to the 'Recommended Nodes' panel on the left, which lists nodes such as 'Decision Tree Predictor', 'Decision Tree Learner', 'Naive Bayes Predictor', 'PIRI Naive Bayes Learner', and 'Regression Predictor'.
- Node Repository:** Points to the 'Node Repository' panel on the left, which provides a hierarchical view of nodes categorized by function like 'IO', 'Manipulation', 'Views', 'Analytics', 'Database', 'Other Data Types', 'Structured Data', 'Scripting', 'Tool Integration', 'Community Nodes', and 'KNIME Labs'.
- Node Description:** Points to the 'Node Description' panel on the right, which shows the configuration options for the 'Partitioning' node. It includes sections for 'Partitioning', 'Dialog Options', 'Absolute' (specifying row counts), 'Relative' (specifying percentages), 'From top', and 'Linear sampling'.
- Console:** Points to the 'Console' panel at the bottom right, which displays the KNIME console output, including a welcome message and license information.
- Outline:** Points to the 'Outline' panel at the bottom left, which provides a visual overview of the workflow structure.

Creating New Workflows, Importing and Exporting

- Right-click anywhere in KNIME Explorer to create a new workflow or workflow group or to import a workflow
- Right-click on workflow or workflow group to export the selected workflow

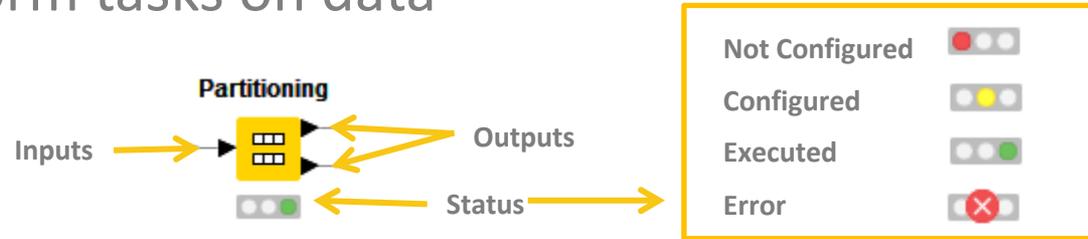


Exercise 00 – Importing the Exercises

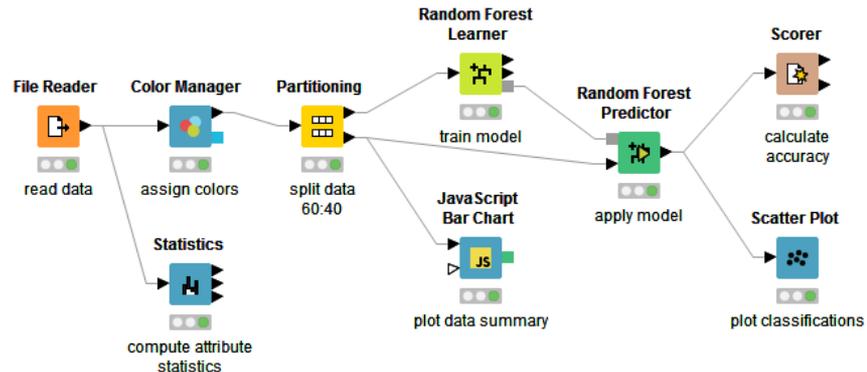
- File > Import KNIME Workflow > Browse
 - 01_KNIME-OMERO/OME-UGM-2018.knar

Visual KNIME Workflows

NODES perform tasks on data



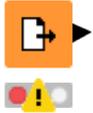
Nodes are combined to create **WORKFLOWS**



More on Nodes...

A node can have 3 states:

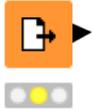
File Reader



Not Configured:

The node is not yet configured and cannot be executed with its current settings.

File Reader



Configured:

The node has been set up correctly, and may be executed at any time

File Reader

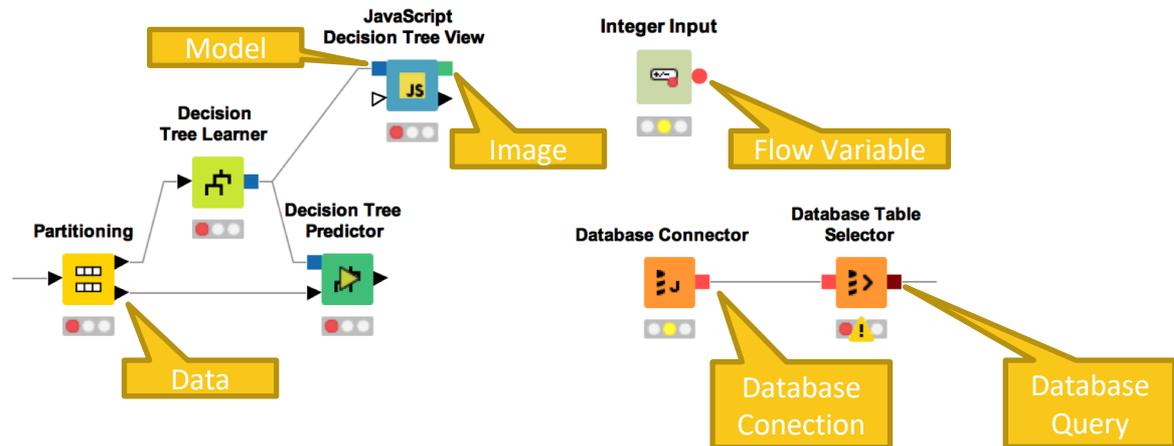


Executed:

The node has been successfully executed. Results may be viewed and used in downstream nodes.

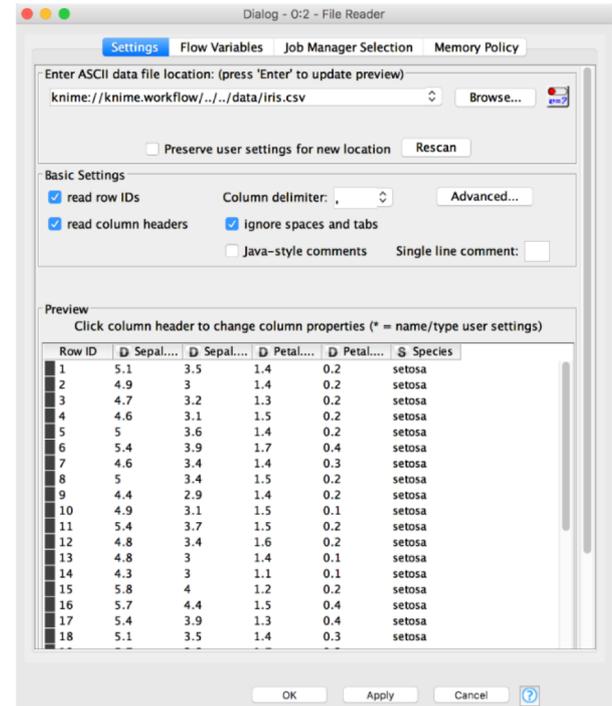
Inserting and Connecting Nodes

- Insert nodes into workspace by dragging them from the Node Repository or by double-clicking in the Node Repository
- Connect nodes by left-clicking output port of Node A and dragging the cursor to (matching) input port of Node B
- Common port types:



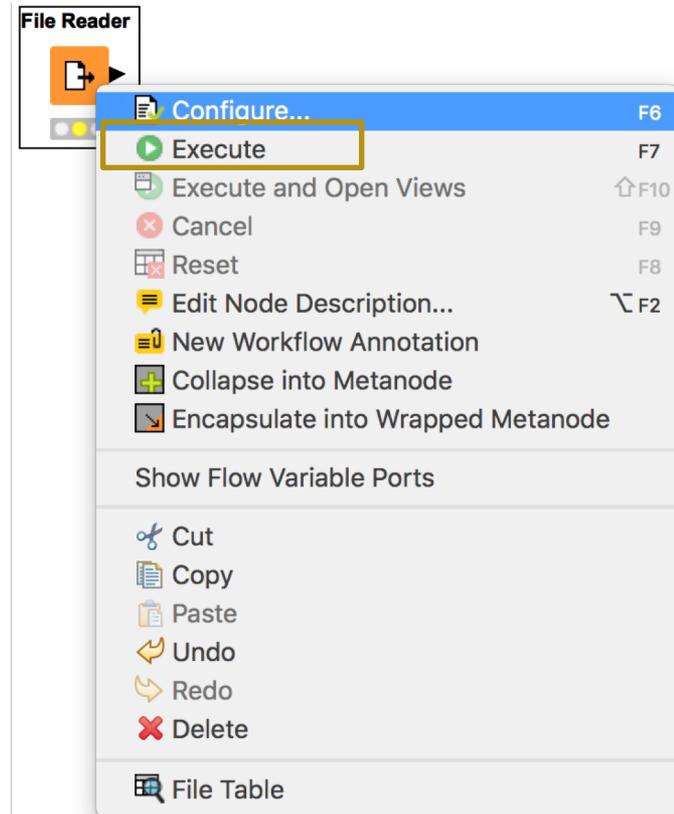
Node Configuration

- Most nodes require configuration
 - To access a node configuration window:
 - Double-click the node
- OR
- Right-click > Configure



Node Execution

- Right-click node
- Select Execute in context menu
- If execution is successful, status shows green light
- If execution encounters errors, status shows red light

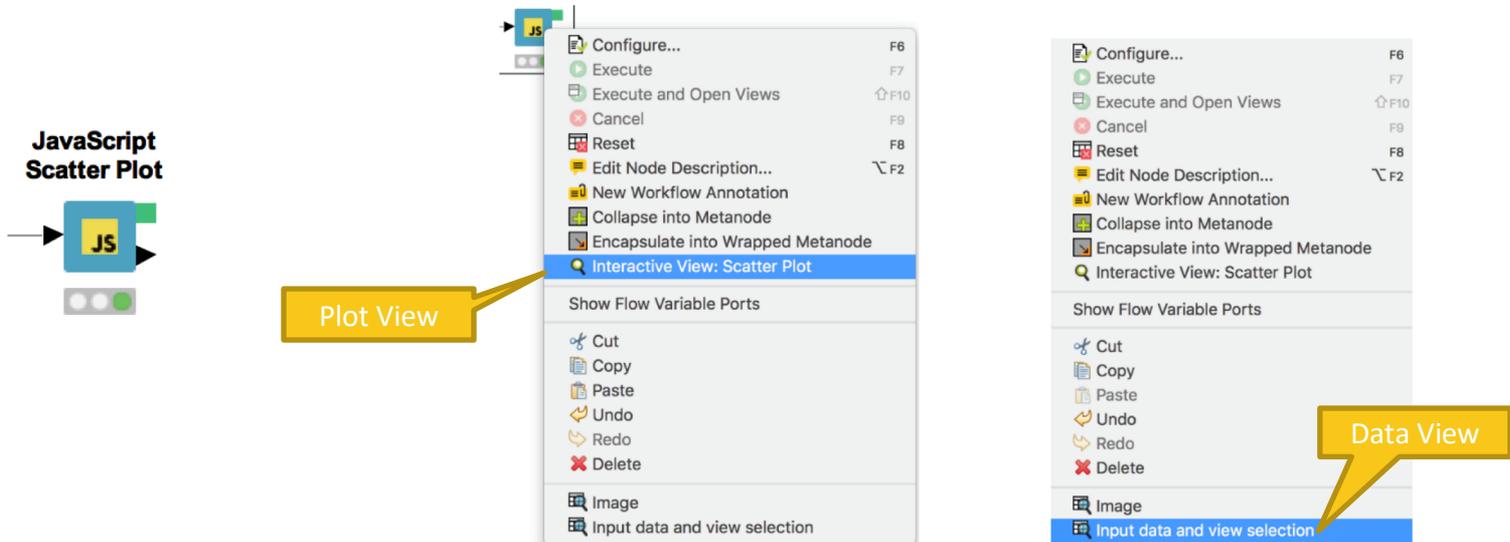


Node Outputs and Views

- Right-click executed node
- Select View option in context menu

OR

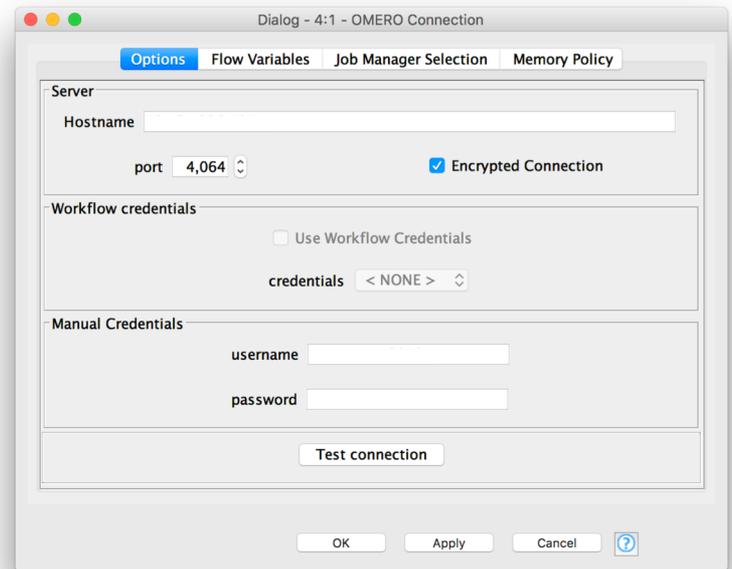
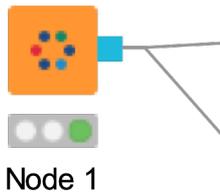
- Select output port (last item) to inspect execution results



New node: OMERO Connection

- Uses server settings (IP address, port) and credentials to establish a connection
- Doesn't download any images

OMERO Connection



Exercise 01 – Connecting to OMERO

- Create a connection to an OMERO instance:
 - IP: ...
 - User: ...
 - PW: ...
- Which port(s) do you see on the *OMERO Connection* node?

Exercise 01 - Solution

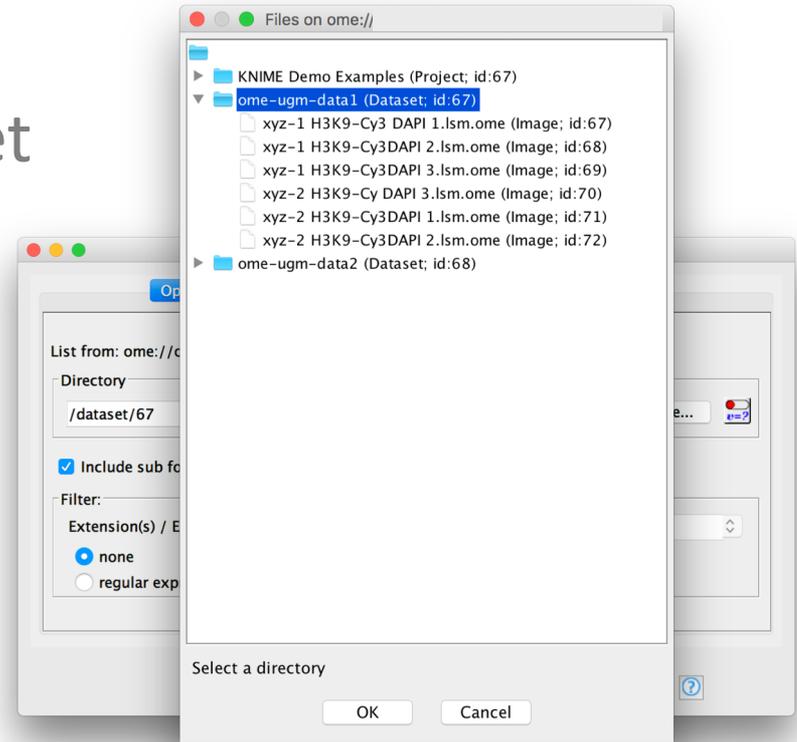
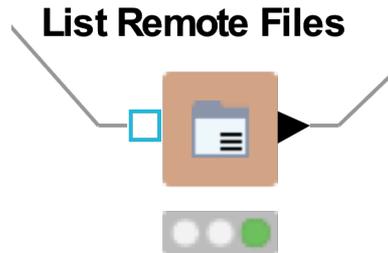
OMERO Connection



Node 1

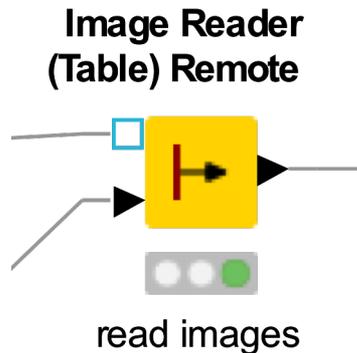
New node: List Remote Files

- Creates file locations for a set of files (usually all files in a folder → *List Files*)
- Doesn't load any images yet



New node: Image Reader (Table) Remote

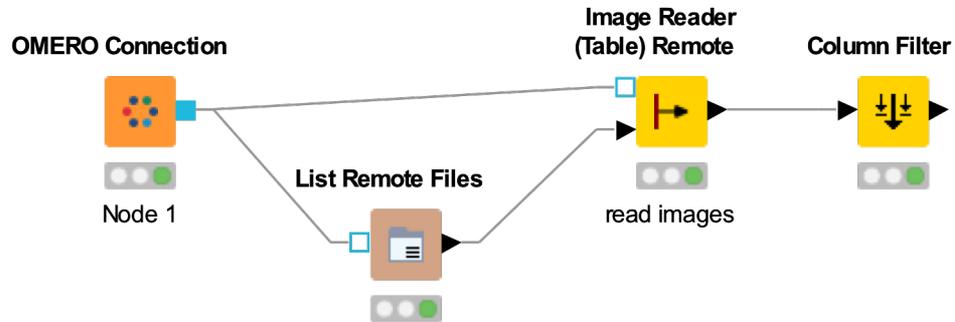
- Input table has to contain file locations
- Uses Bio-Formats to open images and store them into an `Img` column



Exercise 02 – Downloading Images

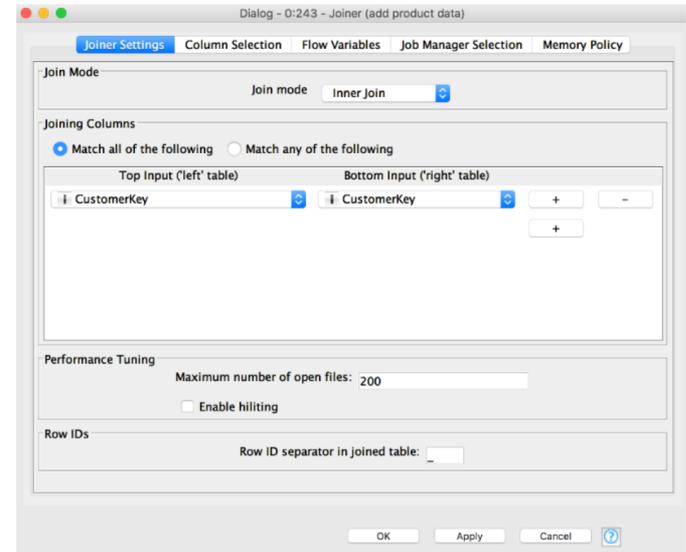
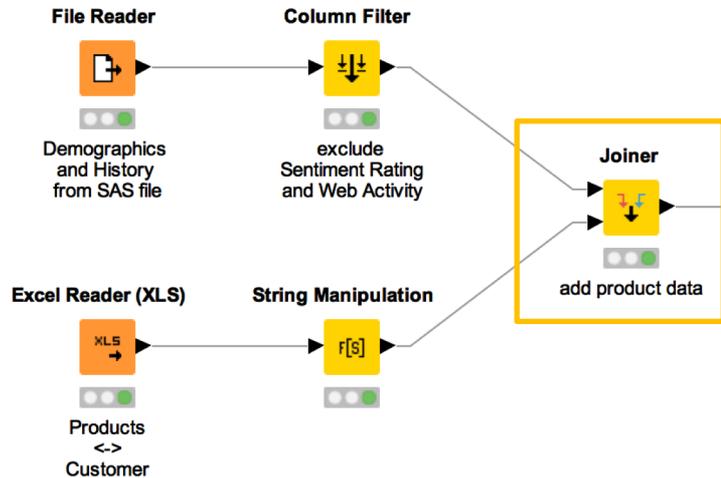
- Use the existing connection to select a dataset from our OMERO instance:
 - *ome-ugm-data1* or *ome-ugm-data2*
- What does the output table look like?
- Download images from the selected dataset and inspect the images

Exercise 02 - Solution



New Node: Joiner

- Combines columns from 2 different tables
- Top port contains “Left” data table
- Bottom port contains the “Right” data table



Joining Columns of Data

Left Table

Row ID	ID	age	Income	class
Row1	1	23	<=50K	F1
Row2	3	25	<=50K	F3
Row3	6	22	>50K	A4
Row4	8	21	<=50K	C3

Right Table

Row ID	ID	age	Income	sex
Row1	1	23	<=50K	M
Row2	2	25	<=50K	F
Row3	4	23	>50K	M
Row4	5	21	<=50K	F
Row5	6	25	>50K	M
Row6	7	24	<=50K	M

Join by ID

Inner Join

Row ID	ID	age	Income	class	age(*)	Income(*)	sex
Row1_Row1	1	23	<=50K	F1	23	<=50K	M
Row3_Row5	6	22	>50K	A4	25	>50K	M

Left Outer Join

Row ID	ID	age	Income	class	age(*)	Income(*)	sex
Row1_Row1	1	23	<=50K	F1	23	<=50K	M
Row3_Row5	6	22	>50K	A4	25	>50K	M
Row2_?	3	25	<=50K	F3	?	?	?
Row4_?	8	21	<=50K	C3	?	?	?

Missing values in the right table.

Right Outer Join

Row ID	age	Income	class	ID	age(*)	Income(*)	sex
Row1	23	<=50K	F1	1	23	<=50K	M
Row3_Row5	22	>50K	A4	6	25	>50K	M
?_Row2	?	?	?	2	25	<=50K	F
?_Row3	?	?	?	4	23	>50K	M
?_Row4	?	?	?	5	21	<=50K	F
?_Row6	?	?	?	7	24	<=50K	M

Missing values in the left table.

Joining Columns of Data

Left Table

Row ID	ID	age	Income	class
Row1	1	23	<=50K	F1
Row2	3	25	<=50K	F3
Row3	6	22	>50K	A4
Row4	8	21	<=50K	C3

Join by ID

Right Table

Row ID	ID	age	Income	sex
Row1	1	23	<=50K	M
Row2	2	25	<=50K	F
Row3	4	23	>50K	M
Row4	5	21	<=50K	F
Row5	6	25	>50K	M
Row6	7	24	<=50K	M

Full Outer Join

Row ID	ID	age	Income	class	ID(*)	age(*)	Income(*)	sex
Row1_Row1	1	23	<=50K	F1	1	23	<=50K	M
Row3_Row5	6	22	>50K	A4	6	25	>50K	M
?_Row2	3	25	<=50K	F3	?	?	?	?
?_Row4	8	21	<=50K	C3	?	?	?	?
?_Row2	?	?	?	?	2	25	<=50K	F
?_Row3	?	?	?	?	4	23	>50K	M
?_Row4	?	?	?	?	5	21	<=50K	F
?_Row6	?	?	?	?	7	24	<=50K	M

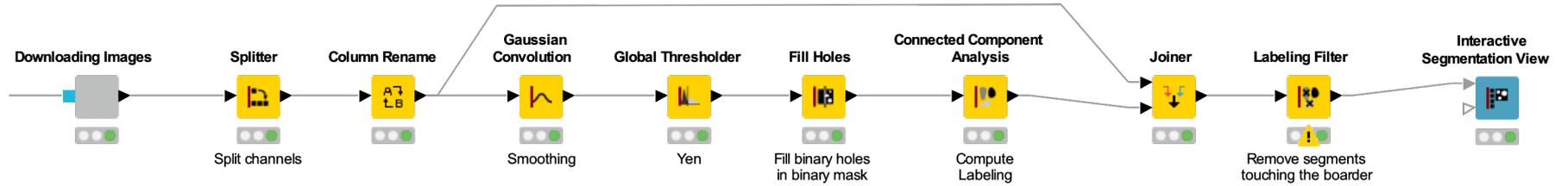
Missing values in the left table.

Missing values in the right table.

Exercise 03 – Processing Images

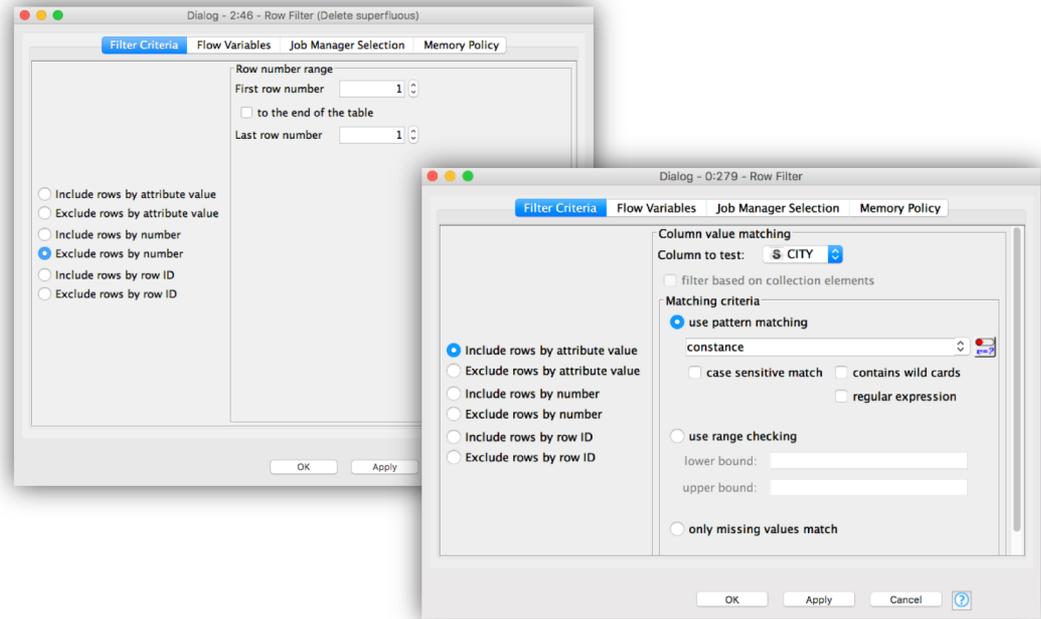
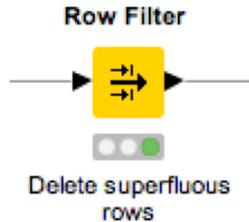
- Segment the first channel of each image:
 - Smoothing (*Gaussian Convolution*), *Global Thresholding* (Yen), *Fill Holes*, and *Connected Component Analysis*
 - Replace the result from the previous step!
- *Join* the resulting Labeling with the input data to recover a raw version of the first channel
- Remove cells that are touching the border image
 - Use *Labeling Filter*

Exercise 03 - Solution



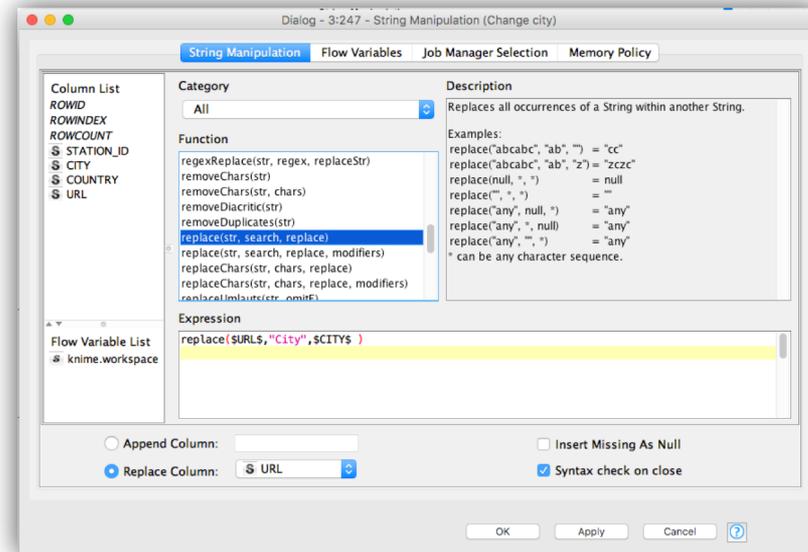
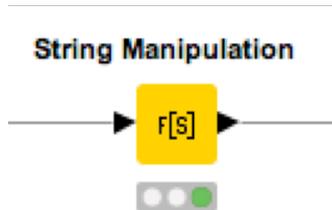
New Node: Row Filter

- Row filtering with in- and exclude option according to certain criteria
 - Certain value or pattern in a selectable column
 - Row number
 - Row ID



New Node: String Manipulation

- Create and edit values in string columns
 - Clean up capitalization (e.g. Lowercase)
 - Search and Replace, join
- Modify existing strings or create new columns



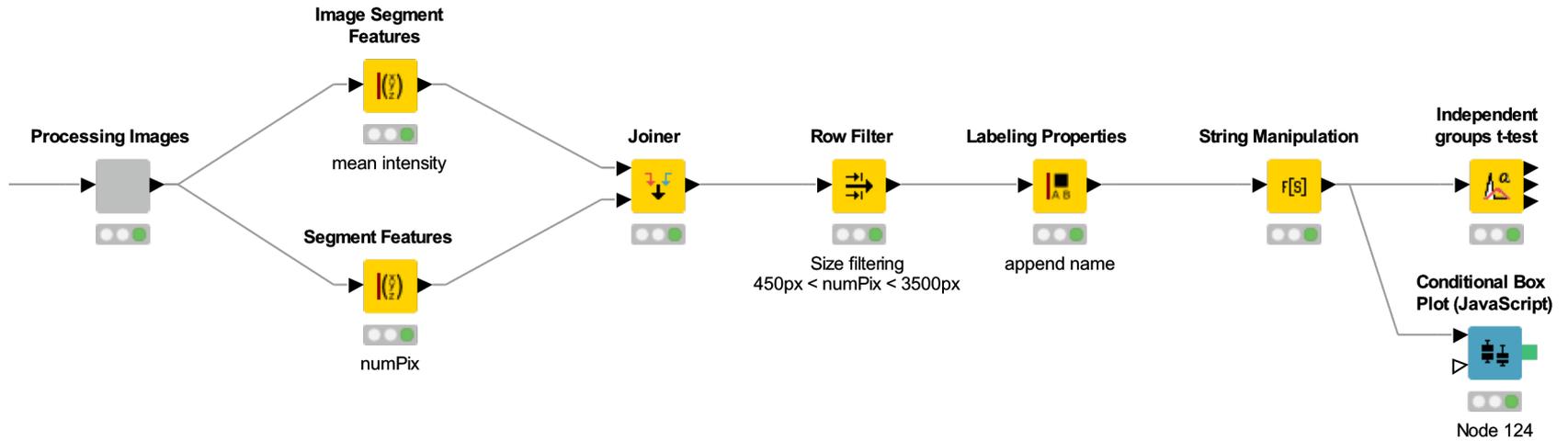
Exercise 04 – Visualizing Results

- Extract mean intensity from 2nd channel for each cell in an image (*Image Segment Features*)
- Extract the cell area (NumPix) for each cell in an image (*Segment Features*)
- *Join* the information for each cell
- Filter cells according to their size (*Row Filter*)
- Extract original file name for each cell (*Labeling Properties*)

Exercise 04 – Visualizing Results (continued)

- Extract class (*xyz-1* or *xyz-2*) from Name column
 - *String Manipulation* to extract first 5 characters
- Compute a p-value for the (independent) mean intensities distributions (*Independent groups t-test*)
- Generate a box plot of the NumPix properties (per group):
 - Use the *Conditional Box Plot (JavaScript)*

Exercise 04 - Solution



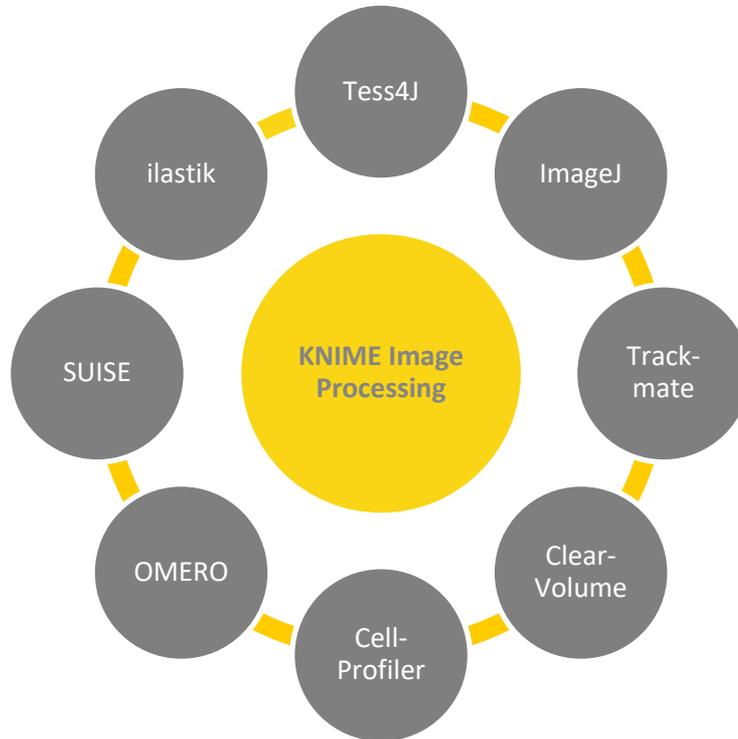
Additional Resources

- Learning/**LEARNING HUB:**
<https://www.knime.com/learning-hub>
- Learning/**NODE GUIDE:**
<https://www.knime.com/nodeguide>
- Community/**Forum:**
<https://forum.knime.com/>

KNIME TV on YouTube

<https://www.youtube.com/user/KNIMETV>

KNIME Image Processing and the ImageJ Ecosystem



The KNIME® trademark and logo and OPEN FOR INNOVATION® trademark are used by KNIME AG under license from KNIME GmbH, and are registered in the United States. KNIME® is also registered in Germany.