

Open for Innovation

KNIME

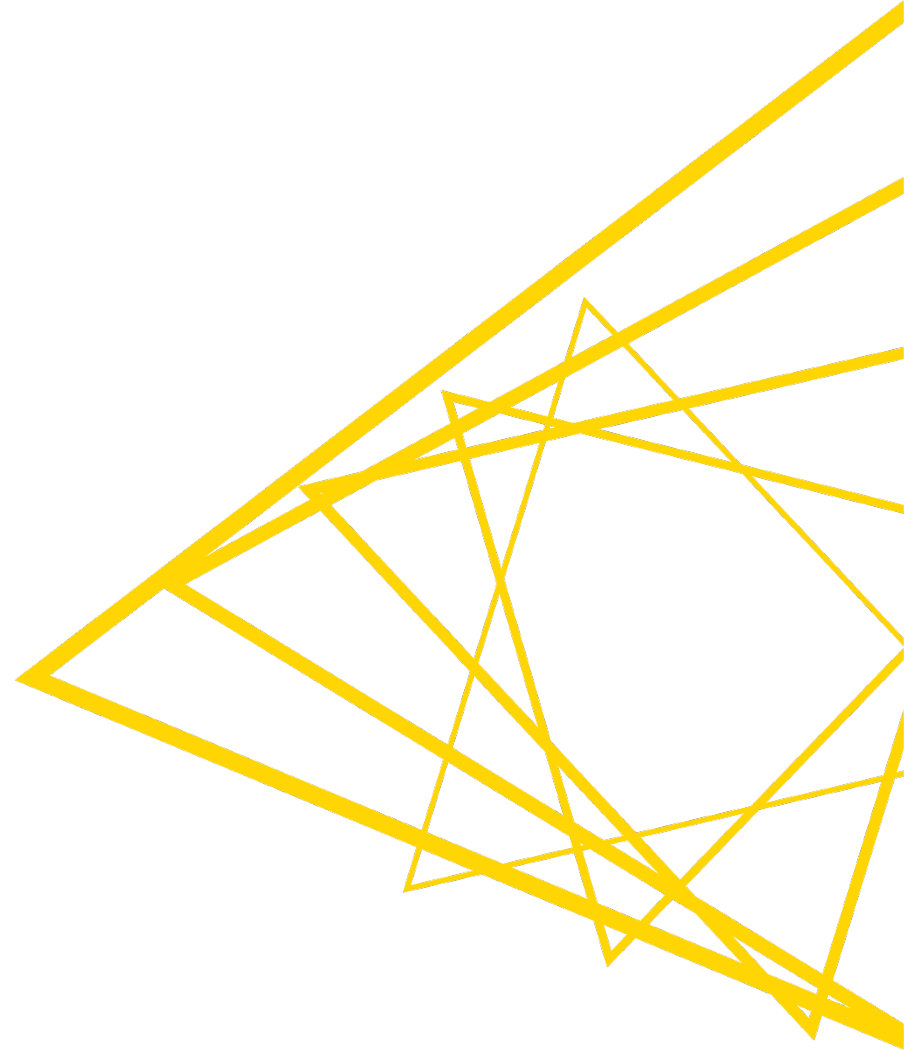
[L1-LS] KNIME Analytics Platform for Data Scientists (Life Science): Basics

KNIME AG



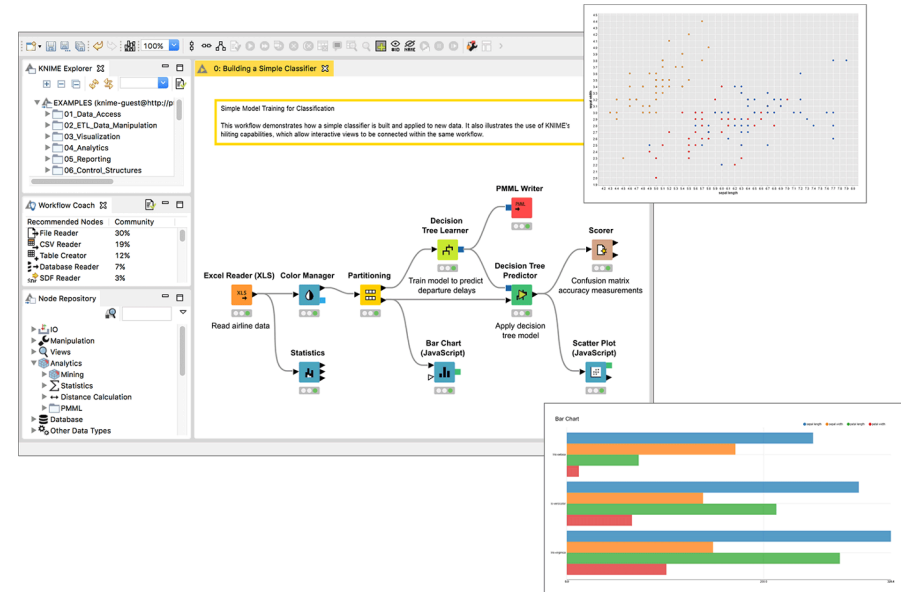
Overview

KNIME Analytics Platform



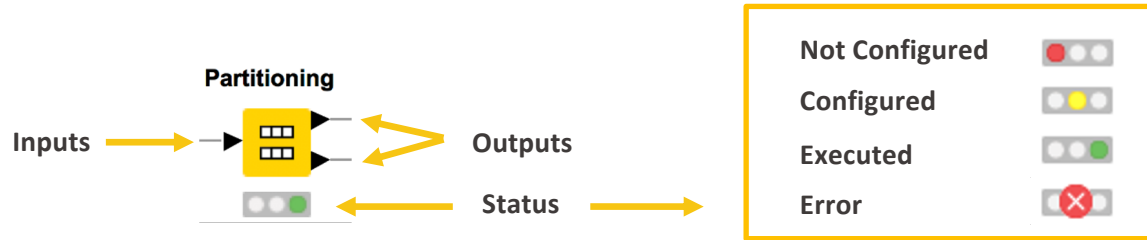
What is KNIME Analytics Platform?

- 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
 - Many integrations, such as Java, R, Python, Weka, Keras, Plotly, H2O, etc.

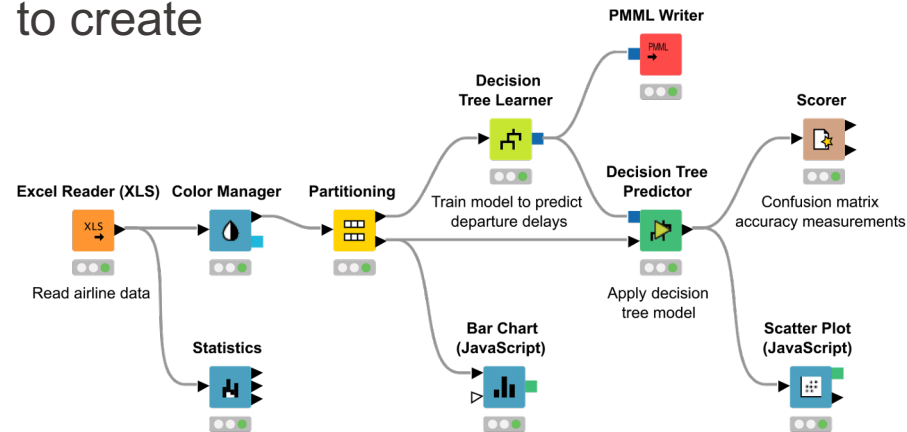


Visual KNIME Workflows

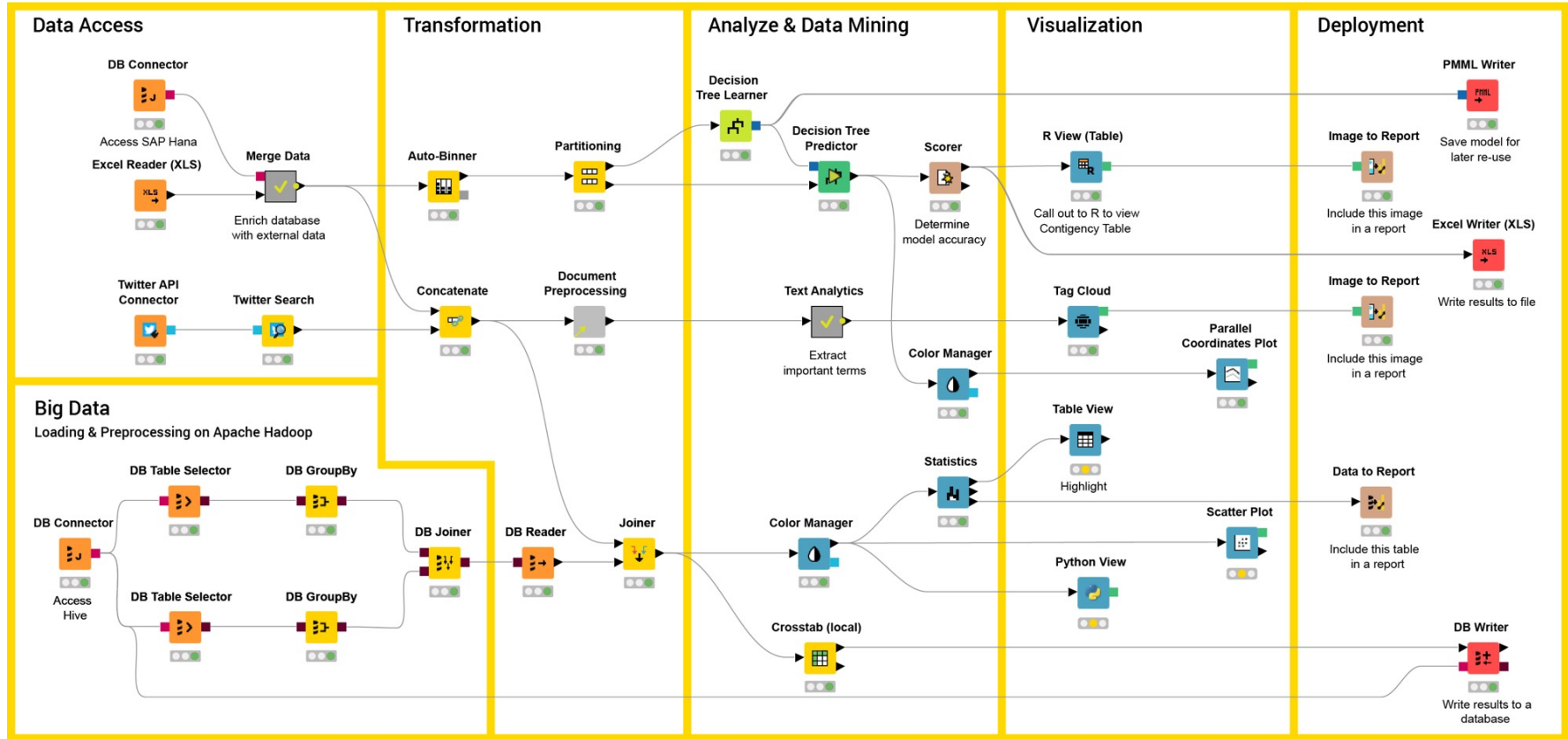
NODES perform tasks on data



Nodes are combined to create
WORKFLOWS



4000+ Nodes for all Steps of End-To-End Data Science



Overview

- Installing KNIME Analytics Platform
- The KNIME Workspace
- The KNIME File Extensions
- The KNIME Workbench
 - Workflow editor
 - Explorer
 - Node Repository
 - Node Description
- Installing new extensions

Install KNIME Analytics Platform

- Select the KNIME version for your computer:
 - Mac
 - Windows – 32 or 64 bit
 - Linux
- Download archive and extract the file, or download installer package and run it

Windows		
KNIME Analytics Platform for Windows (installer)	64 Bit	(441.03 MB)
<i>The installer adds an icon to the desktop and suggests suitable memory settings</i>	32 Bit	(437.42 MB)
KNIME Analytics Platform for Windows (self-extracting archive)	64 Bit	(444.58 MB)
<i>The self-extracting archive only creates a folder holding the KNIME installation</i>	32 Bit	(441.15 MB)
KNIME Analytics Platform for Windows (zip archive)	64 Bit	(529.54 MB)
	32 Bit	(525.59 MB)

Linux		
KNIME Analytics Platform for Linux	64 Bit	(554.2 MB)

Mac		
KNIME Analytics Platform for Mac OSX (10.11 and above)	64 Bit	(522.98 MB)

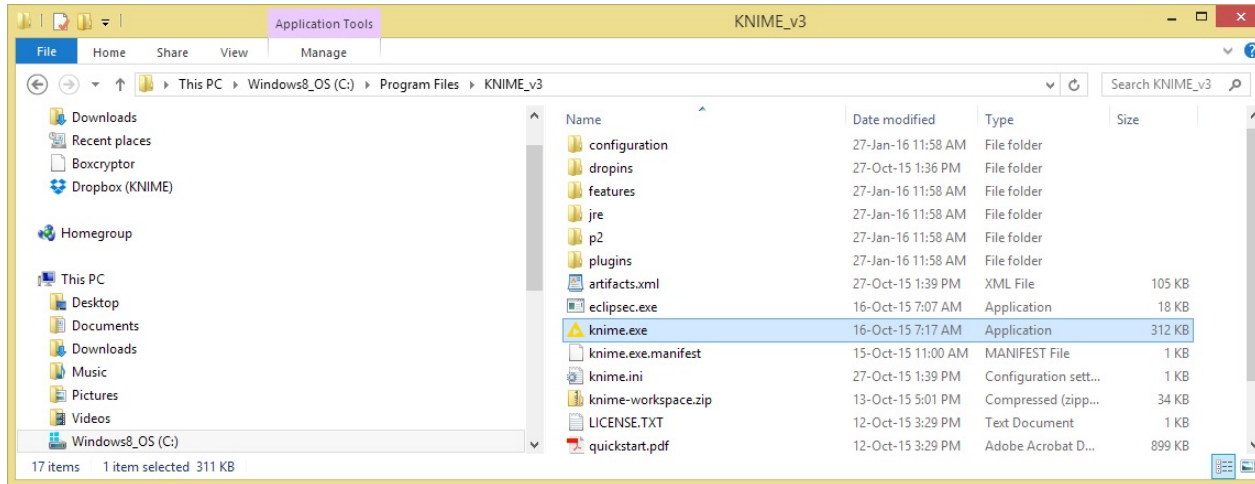
Download from here: <https://www.knime.com/downloads>

Start KNIME Analytics Platform

- Use the shortcut created by the installer

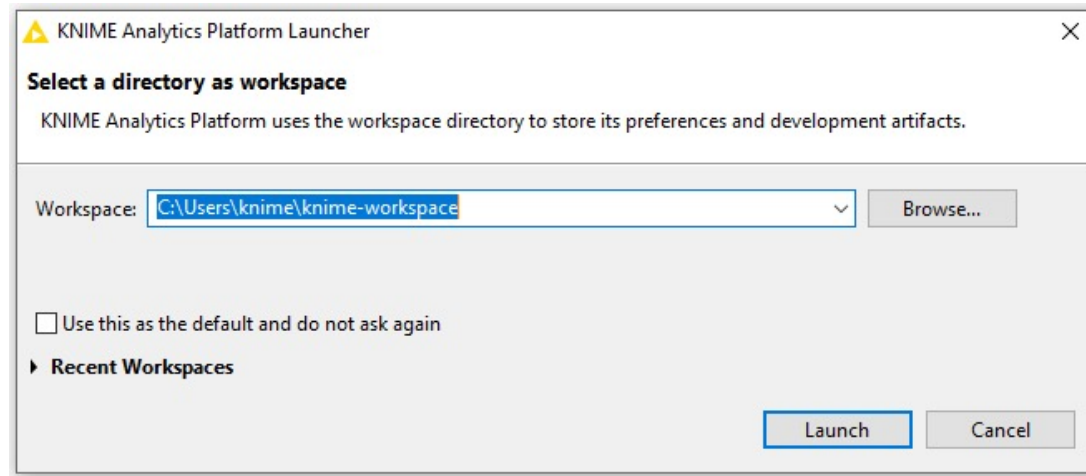


- Or go to the installation directory and launch KNIME via the knime.exe



The KNIME Workspace

- The workspace is the **folder/directory** in which workflows (and potentially data files) are stored for the current KNIME session.
- Workspaces are portable (just like KNIME)



The KNIME Analytics Platform Workbench

The screenshot displays the KNIME Analytics Platform Workbench interface. The main workspace is titled "My first Workflow" and contains a workflow diagram with four nodes: "File Reader" (read adult.csv), "Row Filter" (keep only records born in the US), "Column Filter" (remove gender), and "Table Writer" (write table). The interface is divided into several panels:

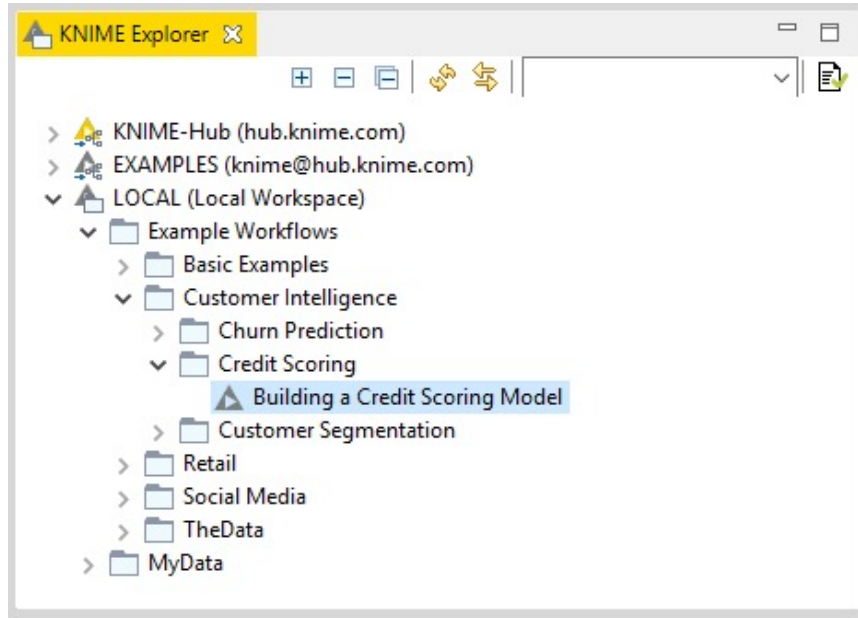
- KNIME Explorer:** Located on the top left, it shows a tree view of the project structure, including "My-KNIME-Hub", "EXAMPLES", and "LOCAL (Local Workspace)".
- Workflow Coach:** Located on the middle left, it displays a list of recommended nodes and their usage percentages, such as "GroupBy" (12%), "Joiner" (9%), and "Column Filter" (8%).
- Node Repository:** Located on the bottom left, it provides a comprehensive list of nodes categorized by function, such as "IO", "Manipulation", "Views", "Analytics", and "DB".
- Node Description:** Located on the top right, it provides detailed information about the selected "Row Filter" node, including its purpose, configuration options, and dialog options.
- Outline:** Located at the bottom left, it shows a small thumbnail of the current workflow.
- Console & Node Monitor:** Located at the bottom right, it displays the execution status of the selected node ("Row Filter (3:6)") and its output data. The output is shown in a table with columns: ID, age, workclass, fnlwgt, education, education-num, marital-status, occupation, relationship, race, sex, capital-gain, capital-loss, and hours.



The workflow diagram in the center shows the following steps:

```
graph LR; A[File Reader: read adult.csv] --> B[Row Filter: keep only records born in the US]; B --> C[Column Filter: remove gender]; C --> D[Table Writer: write table];
```

ID	age	workclass	fnlwgt	education	education-num	marital-status	occupation	relationship	race	sex	capital-gain	capital-loss	hours
Row0	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White	Male	2174	0	40
Row1	50	Self-emp-not-inc	83311	Bachelors	13	Married-civ-spouse	Exec-managerial	Husband	White	Male	0	0	13
Row2	38	Private	215646	HS-grad	9	Divorced	Handlers-cleaners	Not-in-family	White	Male	0	0	40

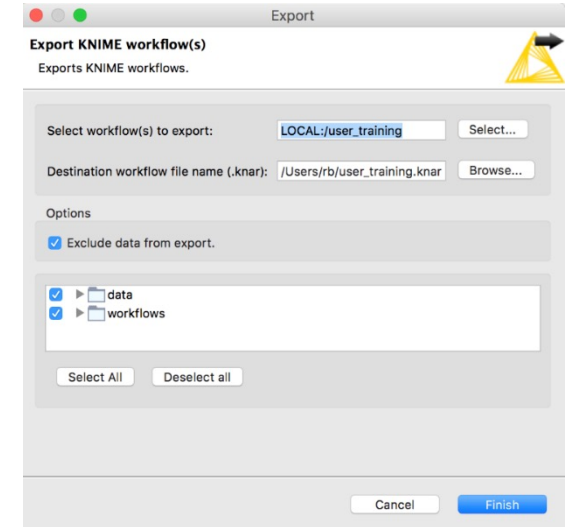
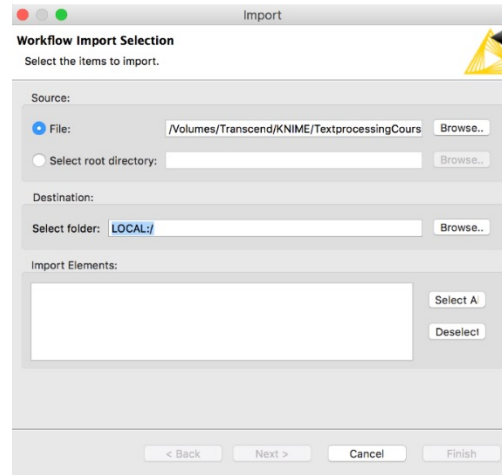
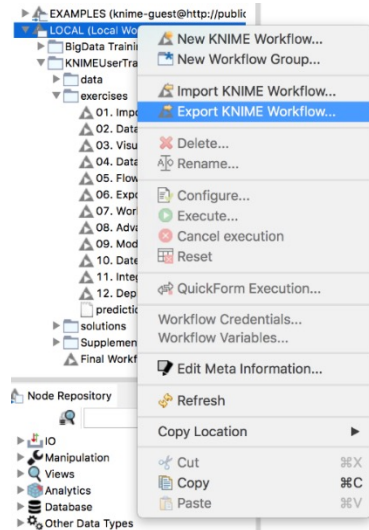
KNIME Explorer



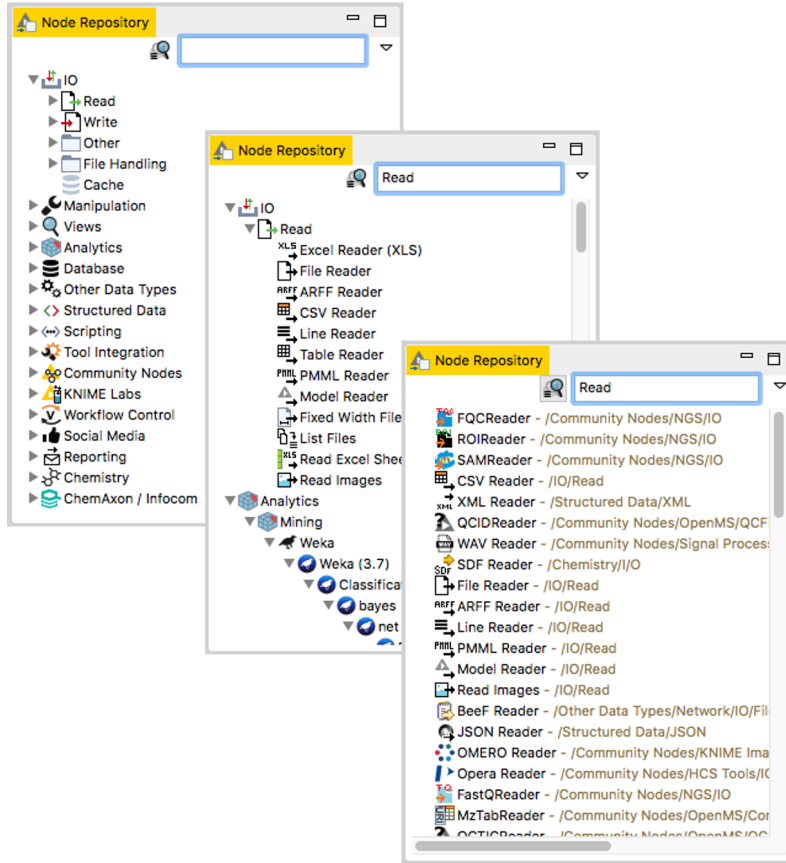
- In LOCAL you can access your own workflow projects.
- Other mountpoints allow you to connect to
 - EXAMPLE Server
 - KNIME Hub
 - KNIME Server
- The Explorer toolbar on the top has a search box and buttons to
 -  select the workflow displayed in the active editor
 -  refresh the view
- The KNIME Explorer can contain 4 types of content:
 - Workflows
 - Workflow groups
 - Data files
 - Shared Components



Creating New Workflows, Importing and Exporting

- Right-click inside the KNIME Explorer to create a new workflow or a workflow group, or to import a workflow
- Right-click the workflow or workflow group to export

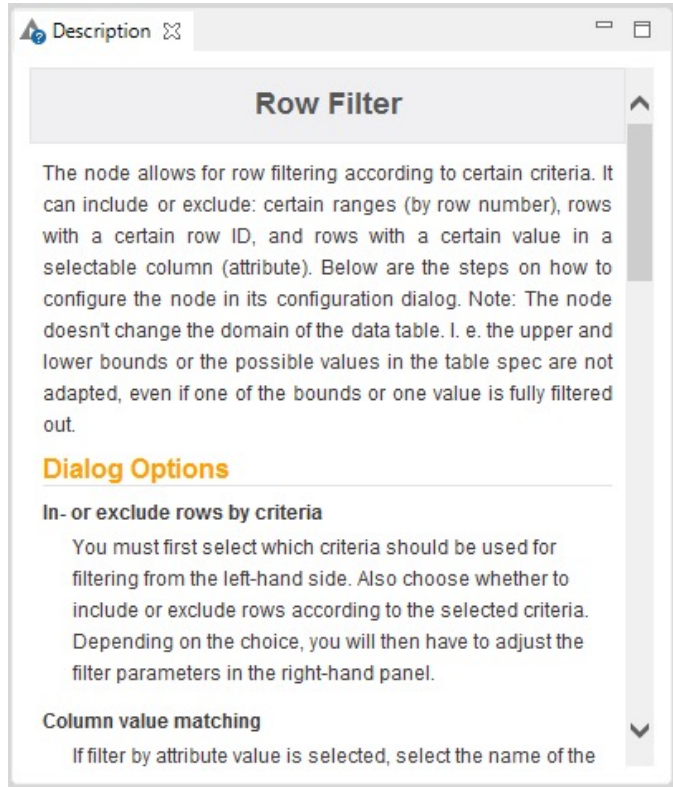


Node Repository



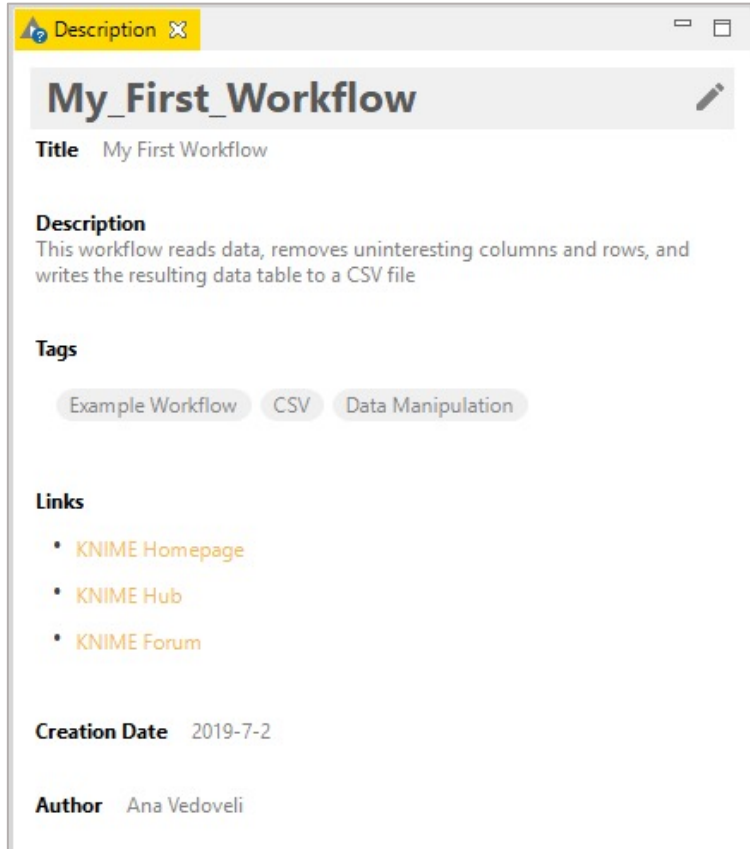
- The Node Repository lists all KNIME nodes
- The search box has 2 modes
 -  Standard Search – exact match of node name
 -  Fuzzy Search – finds the most similar node name
- Nodes can be added by drag and drop from the Node Repository to the Workflow Editor.

Description



- The Description window gives information about:
 - Node Functionality
 - Input & Output
 - Node Settings
 - Ports
 - References to literature

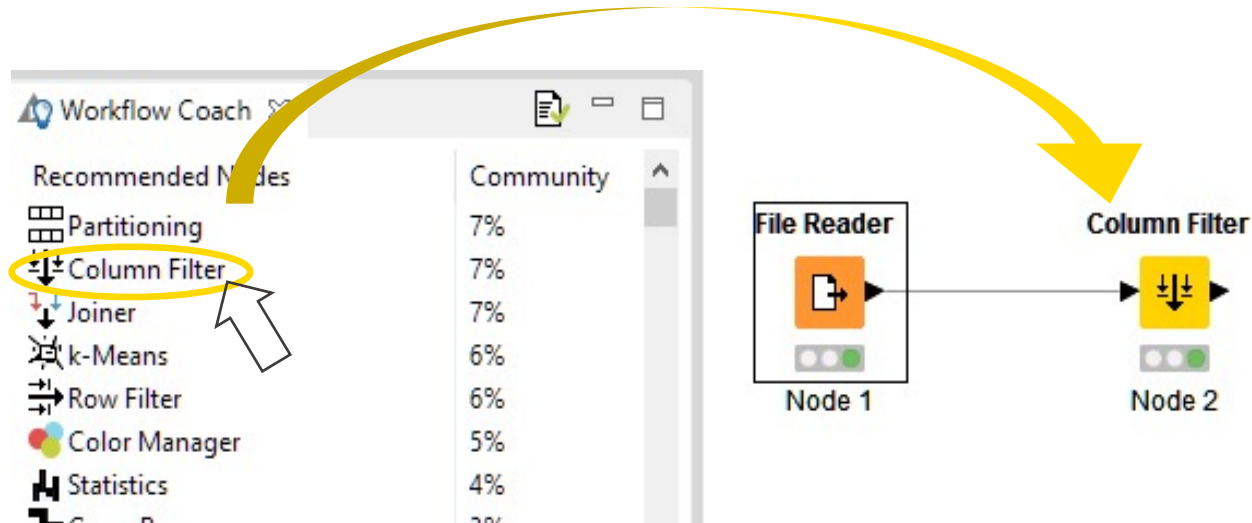
Workflow Description



- When selecting the workflow, the Description window gives information about the workflow's:
 - Title
 - Description
 - Associated Tags and Links
 - Creation Date
 - Author

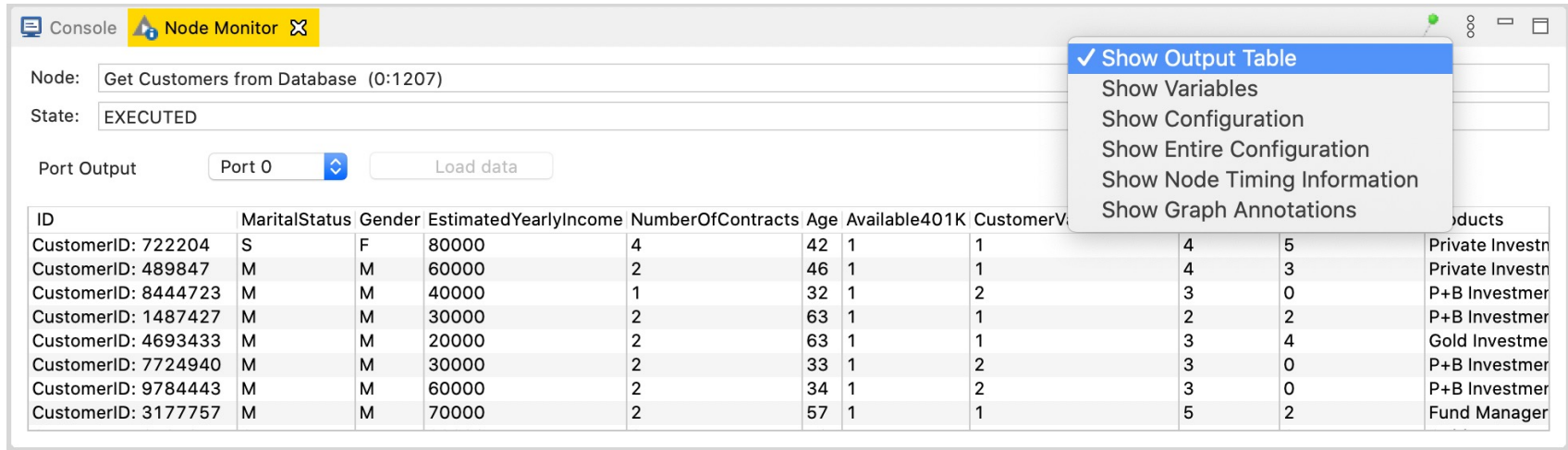
Workflow Coach

- Node recommendation engine
 - Gives hints about which node use next in the workflow
 - Based on KNIME communities' usage statistics
 - Based on own KNIME workflows



Node Monitor

- By default the Node Monitor shows you the output table of the node selected in the workflow editor
- Click on the three dots on the upper right to show the flow variables, configuration, etc.



Node: Get Customers from Database (0:1207)

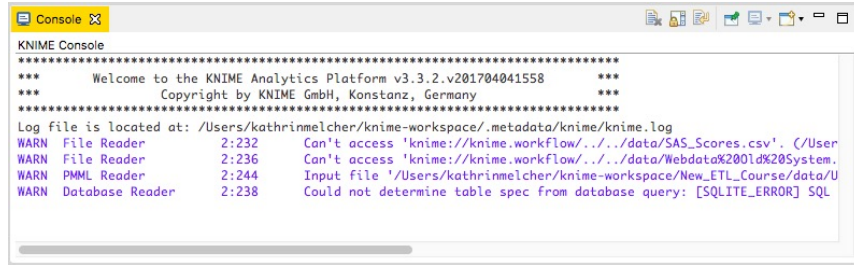
State: EXECUTED

Port Output: Port 0 [Load data]

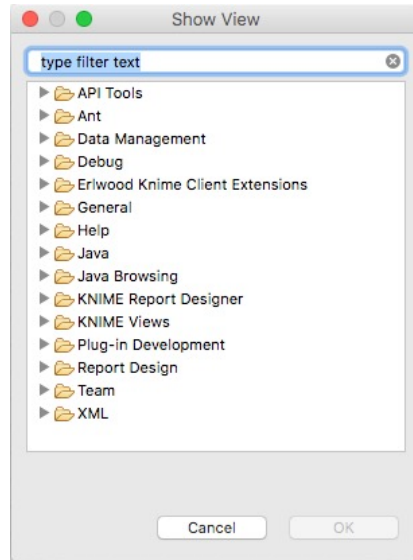
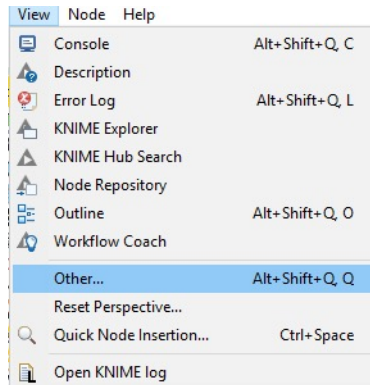
ID	MaritalStatus	Gender	EstimatedYearlyIncome	NumberOfContracts	Age	Available401K	CustomerV				Products
CustomerID: 722204	S	F	80000	4	42	1	1		4	5	Private Investn
CustomerID: 489847	M	M	60000	2	46	1	1		4	3	Private Investn
CustomerID: 8444723	M	M	40000	1	32	1	2		3	0	P+B Investmer
CustomerID: 1487427	M	M	30000	2	63	1	1		2	2	P+B Investmer
CustomerID: 4693433	M	M	20000	2	63	1	1		3	4	Gold Investme
CustomerID: 7724940	M	M	30000	2	33	1	2		3	0	P+B Investmer
CustomerID: 9784443	M	M	60000	2	34	1	2		3	0	P+B Investmer
CustomerID: 3177757	M	M	70000	2	57	1	1		5	2	Fund Manager

- ✓ Show Output Table
- Show Variables
- Show Configuration
- Show Entire Configuration
- Show Node Timing Information
- Show Graph Annotations

Console and Other Views



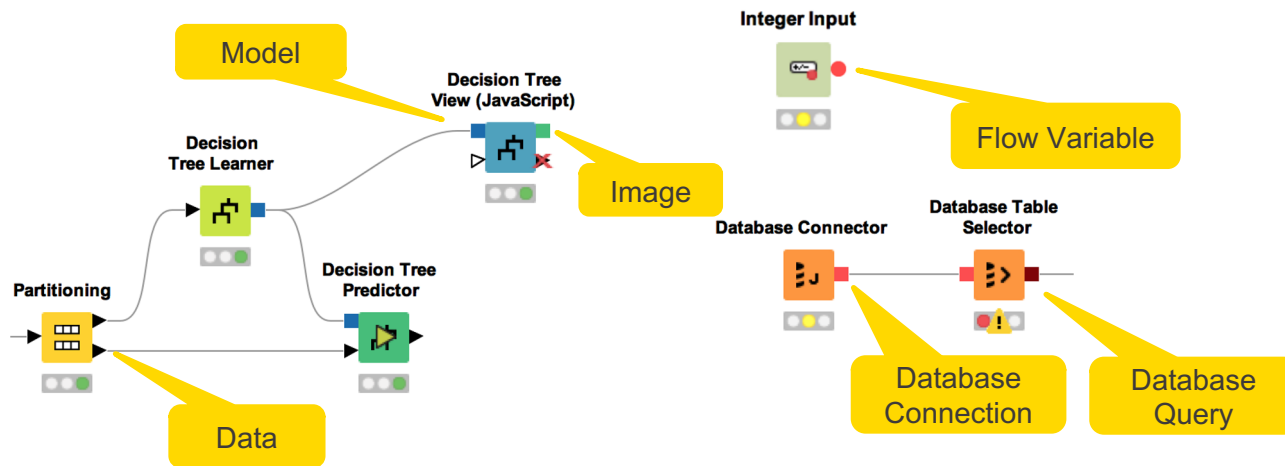
```
KNIME Console
*****
*** Welcome to the KNIME Analytics Platform v3.3.2.v201704041558 ***
*** Copyright by KNIME GmbH, Konstanz, Germany ***
*****
Log file is located at: /Users/kathrinmelcher/knime-workspace/.metadata/knime/knime.log
WARN File Reader      2:232 Can't access 'knime://knime.workflow/../../data/SAS_Scores.csv'. (User
WARN File Reader      2:236 Can't access 'knime://knime.workflow/../../data/Webdata%2001d%20System.
WARN PMML Reader       2:244 Input file '/Users/kathrinmelcher/knime-workspace/New_ETL_Course/data/U
WARN Database Reader   2:238 Could not determine table spec from database query: [SQLITE_ERROR] SQL
```



- Console view prints out error and warning messages about what is going on under the hood
- Click on View and select Other... to add different views
 - Node Monitor, Licenses, etc.

Inserting and Connecting Nodes

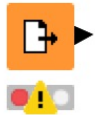
- Insert nodes into workspace by dragging them from Node Repository or by double-clicking in Node Repository
- Connect nodes by left-clicking output port of Node A and dragging the cursor to (matching) input port of Node B
- Common port types:



More on Nodes...

- A node can have 4 states:

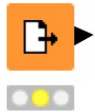
File Reader



Not Configured:

The node is waiting for configuration or incoming data.

File Reader



Configured:

The node has been configured correctly, and can be executed.

File Reader



Executed:

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

File Reader

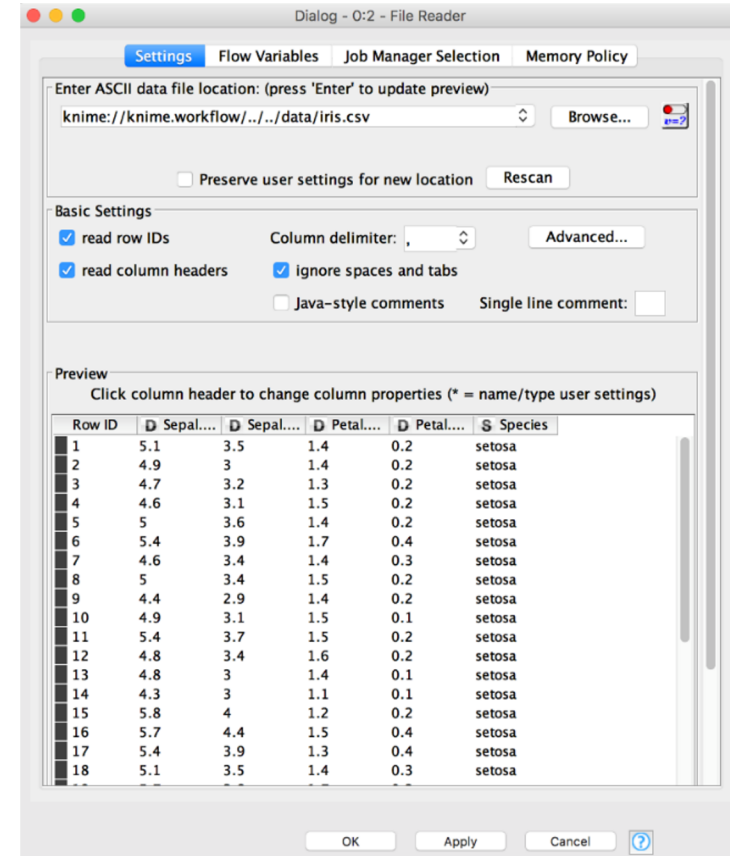


Error:

The node has encountered an error during execution.

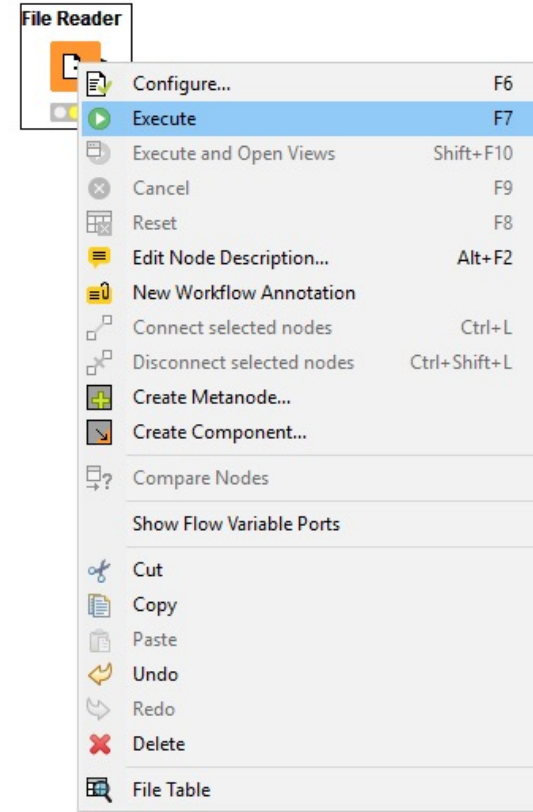
Node Configuration

- Most nodes require configuration
- To access a node configuration window:
 - Double-click the node
 - 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



Tool Bar

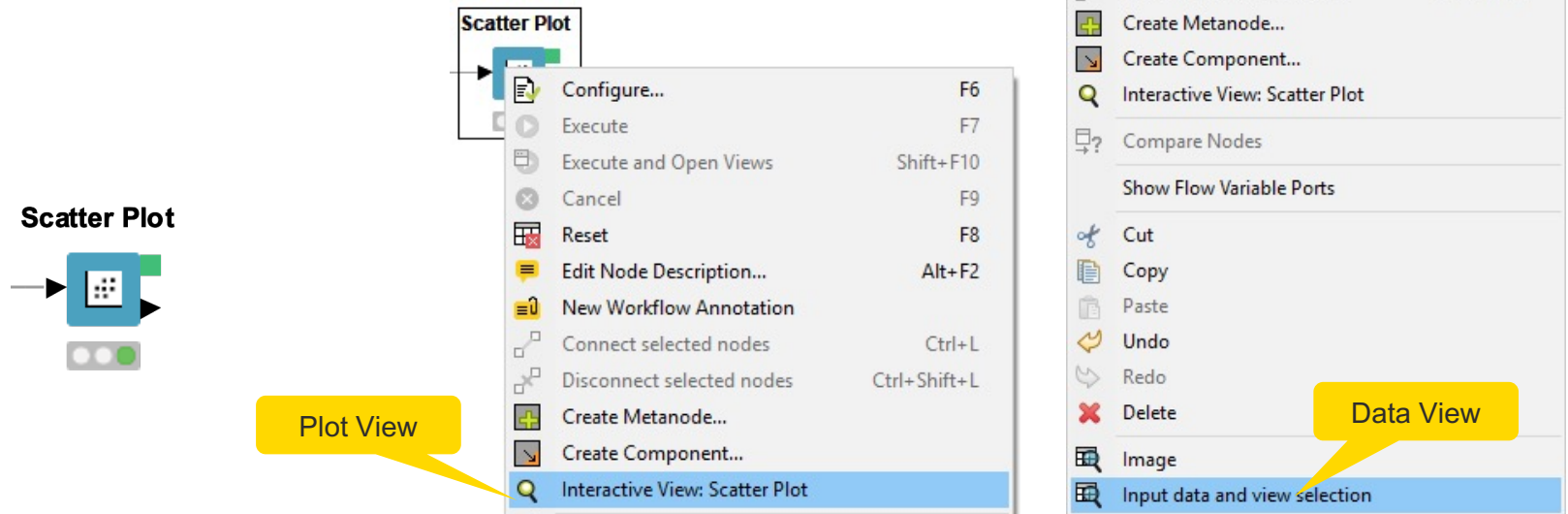


The buttons in the toolbar can be used for the active workflow. The most important buttons:

-  Execute selected and executable nodes (F7)
-  Execute all executable nodes
-  Execute selected nodes and open first view
-  Cancel all selected, running nodes (F9)
-  Cancel all running nodes

Node Views

- Right-click node
- Select Views in context menu
- Select output port to inspect execution results



KNIME File Extensions

- Dedicated file extensions for Workflows and Workflow groups associated with KNIME Analytics Platform

- ***.knwf** for KNIME Workflow Files



- ***.knar** for KNIME Archive Files



Hot Keys (for Future Reference)

Task	Hot key	Description
Node Configuration	F6	opens the configuration window of the selected node
Node Execution	F7	executes selected configured nodes
	Shift + F7	executes all configured nodes
	Shift + F10	executes all configured nodes and opens all views
	F9	cancels selected running nodes
	Shift + F9	cancels all running nodes
Node Connections	Ctrl + L	connects selected nodes
	Ctrl + Shift + L	disconnects selected nodes
Move Nodes and Annotations	Ctrl + Shift + Arrow	moves the selected node in the arrow direction
	Ctrl + Shift + PgUp/PgDown	moves the selected annotation in the front or in the back of all overlapping annotations
Workflow Operations	F8	resets selected nodes
	Ctrl + S	saves the workflow
	Ctrl + Shift + S	saves all open workflows
	Ctrl + Shift + W	closes all open workflows
	Ctrl + F	search workflow for nodes
Metanode	Shift + F12	opens metanode wizard

Getting Started: KNIME Example Server

- Connect via KNIME Explorer to a public repository with large selection of example workflows for many, many applications
- Workflows also available on KNIME Hub

The screenshot displays the KNIME Analytics Platform interface. On the left, the 'KNIME Explorer' sidebar shows a tree view of example workflows. The selected workflow is '01_Example_for_Standard_Preprocessing' under '02_ETL_Data_Manipulation'. The main workspace shows a workflow diagram titled 'Simple Preprocessing Example' with the description 'Shows the use of different filter nodes.' The workflow starts with a 'File Reader' node, followed by a 'Row Filter' node (labeled 'keep rows where "native-country" is missing'). The workflow then splits into two parallel paths. The top path consists of a 'Column Filter' node, followed by a 'Reference Column Filter' node, a 'Numeric Binner' node (labeled 'Replace numeric column age with an attribute column'), and a 'Nominal Value Row Filter' node (labeled 'Filter by this new attribute'). The bottom path consists of a 'Reference Row Filter' node (labeled 'Exclude those rows from the original table'), followed by a 'Row Filter' node (labeled 'filter example with a regular expression'), and a 'Concatenate' node. Both paths converge at the 'Concatenate' node. A yellow tooltip is visible over the 'Simple Preprocessing Example' title, stating: 'This is a temporary copy of "knime://EXAMPLES/Users/knime/Examples/02_ETL_Data_Manipulation/00_Basic_Examples/01_Example_for_Standard_Preprocessing". Use "Save As..." to save a permanent copy of the workflow to your local workspace, or a mounted KNIME Server.'

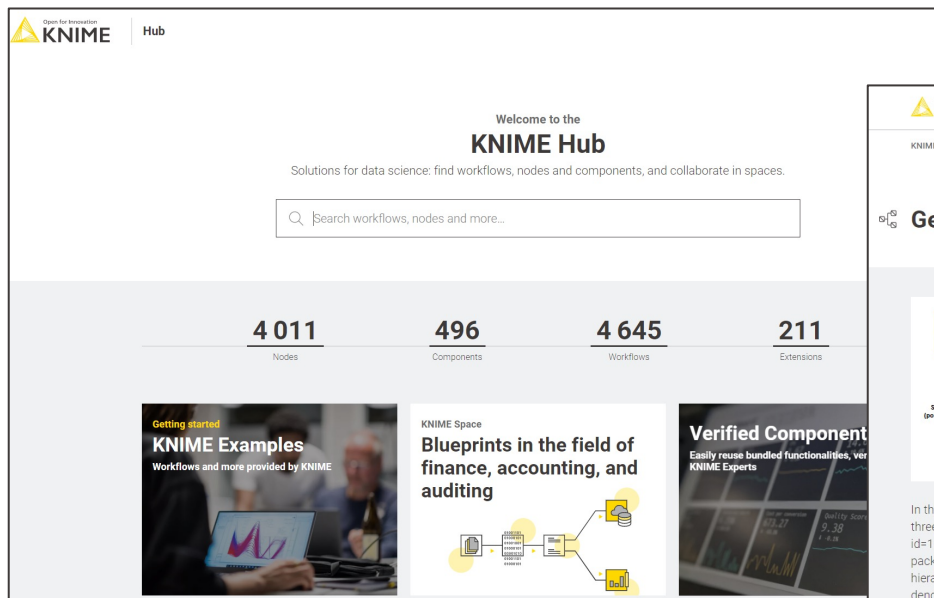
KNIME Explorer sidebar contents:

- My-KNIME-Hub (hub.knime.com)
- EXAMPLES (knime@hub.knime.com)
 - 00_Components
 - 01_Data_Access
 - 02_ETL_Data_Manipulation
 - 00_Basic_Examples
 - 00_Visual_Analysis_of_Sales_Data
 - 01_Example_for_Standard_Preprocessing
 - 02_ETL_Basics
 - 01_Filtering
 - 02_Aggregations
 - 03_Joining_and_Concatenating
 - 04_Transformation
 - 05_Indexing_Searching
 - 06_Date_and_Time_Manipulation
 - Random Forest Model Workflow 15 Plates
 - 03_Visualization
 - 04_Analytics
 - 05_Reporting
 - 06_Control_Structures
 - 07_Scripting
 - 08_Other_Analytics_Types
 - 09_Enterprise
 - 10_Big_Data
 - 20_Strange_but_Educational
 - 40_Partners
 - 50_Applications
 - 99_Community
- LOCAL (Local Workspace)

KNIME Explorer main panel contents:

- My-KNIME-Hub (hub.knime.com)
- EXAMPLES (knime@hub.knime.com)
 - Double-click to see the examples
- LOCAL (Local Workspace)
 - Example Workflows
 - MyData

KNIME Hub



Open for innovation
KNIME Hub

Welcome to the
KNIME Hub

Solutions for data science: find workflows, nodes and components, and collaborate in spaces.

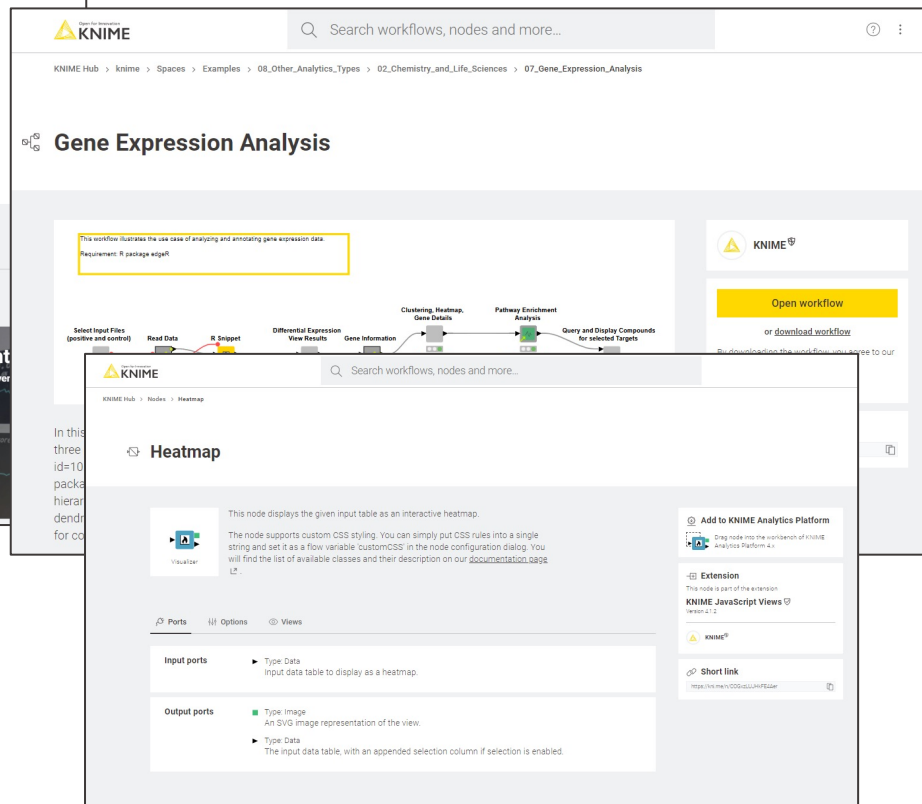
Search workflows, nodes and more...

4 011 Nodes 496 Components 4 645 Workflows 211 Extensions

Getting started
KNIME Examples
Workflows and more provided by KNIME

KNIME Space
Blueprints in the field of finance, accounting, and auditing

Verified Component
Easily reuse bundled functionalities, verified by KNIME Experts



Open for innovation
KNIME

Search workflows, nodes and more...

KNIME Hub > knime > Spaces > Examples > 08_Other_Analytics_Types > 02_Chemistry_and_Life_Sciences > 07_Gene_Expression_Analysis

Gene Expression Analysis

This workflow illustrates the use case of analyzing and annotating gene expression data.
Requirement: R package edgeR

Open workflow
or download workflow
By downloading the workflow, you agree to our

KNIME Hub > Nodes > Heatmap

Heatmap

This node displays the given input table as an interactive heatmap.

The node supports custom CSS styling. You can simply put CSS rules into a single string and set it as a flow variable customCSS in the node configuration dialog. You will find the list of available classes and their description on our [documentation page](#).

Ports Options Views

Input ports

- Type: Data
Input data table to display as a heatmap.

Output ports

- Type: Image
An SVG image representation of the view.
- Type: Data
The input data table, with an appended selection column if selection is enabled.

Add to KNIME Analytics Platform
Drag node into the workspace of KNIME Analytics Platform 4.x

Extension
This node is part of the extension
KNIME JavaScript Views
Version 1.2

Short link
<https://www.knime.com/KNIMEHub/Heatmap>

A place to share knowledge
about Workflows and Nodes
<https://hub.knime.com>

The KNIME Hub

The screenshot displays the KNIME Hub homepage. At the top left is the KNIME logo with the tagline 'Open for Innovation'. At the top right are help and menu icons. The main heading reads 'Welcome to the KNIME Hub', followed by a subtitle: 'The place to find and collaborate on KNIME workflows and nodes. Here you can find solutions for your data science questions.' Below this is a search bar containing the text 'heatmap', which is circled in yellow. Under the search bar, four statistics are presented: 3 680 Nodes, 196 Components, 2 300 Workflows, and 196 Extensions. The bottom section features three promotional cards: 'Extended KNIME Summit' with the text 'Online courses, workshops, and webinars during April and May'; 'How to Getting started' with the text 'From downloading through to building your first workflow' and a diagram of a workflow; and 'Forum' with the text 'Get help from our community and help others' and an image of people at a laptop.

Open for Innovation
KNIME

Welcome to the
KNIME Hub

The place to find and collaborate on KNIME workflows and nodes. Here you can find solutions for your data science questions.

🔍 heatmap

3 680
Nodes

196
Components

2 300
Workflows

196
Extensions

Extended KNIME Summit
Online courses, workshops, and webinars during April and May

How to Getting started
From downloading through to building your first workflow


Forum
Get help from our community and help others

Searching Nodes and Workflows

The screenshot shows the KNIME Hub search interface. At the top, the KNIME logo is on the left, a search bar with the text 'heatmap' is in the center, and a close button (X) and a help icon (?) are on the right. Below the search bar, the text 'KNIME Hub > Search' is displayed. The main section shows '42 results'. Below this, there are tabs for 'All', 'Nodes', 'Components', 'Workflows' (which is selected), and 'Extensions'. The search results are listed in a table-like format. The second result, 'Gene Expression Analysis', is highlighted with a dashed yellow border. Each result includes a title, a description, a path, and a user profile picture.

Icon	Title	Description	Path	User
	Visualization of screening data with HCS-Tools High-content screening	The workflow shows how to use the 'Plate Heatmap Viewer' node from the HCS-Tools plugin for visualization	knime > Examples > 99_Community > 02_HCS_Tools > 01_Visualization_of_screening_data	
	Gene Expression Analysis Life Sciences Bioinformatics R Shared components Interactive views	In this workflow, we analyze RNA-Seq data from tumors and matched normal tissue from three patients with oral squamous cell carcinomas (https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0137110)	knime > Examples > 08_Other_Analytics_Types > 02_Chemistry_and_Life_Sciences > 07_Gene_Expression_Analysis	
	Create Heatmaps in KNIME and use R to add Dendrograms knime R heatmap dendrogram	It is possible to create heatmaps with dendrograms with the help to R. You would have to check the code if the distance and cluster settings fit your needs. You can also export the numbers behind the...	mllauber71 > Public > kn_example_r_heatmap_cars	
	heatmap_test		jtyler > Public > heatmap_test	


Opening a Workflow from the Hub

Open for Innovation

Hub

Search workflows, nodes and more...

KNIME Hub > knime > Spaces > Examples > 08_Other_Analytics_Types > 02_Chemistry_and_Life_Sciences > 07_Gene_Expression_Analysis

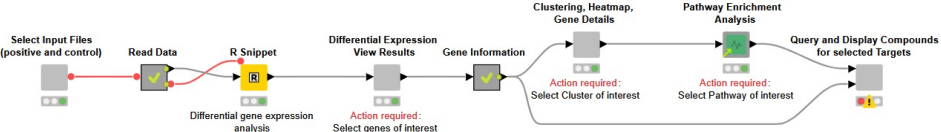
 Workflow

Gene Expression Analysis

Life Sciences | Bioinformatics | R | Shared components | Interactive views

Last edited: 24 Apr 2020

This workflow illustrates the use case of analyzing and annotating gene expression data.
Requirement: R package edgeR



```
graph LR; A[Select Input Files  
(positive and control)] --> B[Read Data]; B --> C[R Snippet  
Differential gene expression analysis]; C --> D[Differential Expression View Results  
Action required: Select genes of interest]; D --> E[Gene Information]; E --> F[Clustering, Heatmap, Gene Details  
Action required: Select Cluster of interest]; E --> G[Pathway Enrichment Analysis  
Action required: Select Pathway of interest]; F --> H[Query and Display Compounds for selected Targets]; G --> H;
```

In this workflow, we analyze RNA-Seq data from tumors and matched normal tissue from three patients with oral squamous cell carcinomas (<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0009317>). Differentially expressed genes are discovered using the R package edgeR and are then displayed in an interactive view. Subsequently, genes are hierarchically clustered based on their expression pattern, and the results are shown via a dendrogram alongside a heatmap. We then perform a pathway enrichment analysis and look for compounds targeting the gene product of interest.


Short link
<https://kni.me/w/vUgr-iyGudXur-1B>

Used extensions & nodes
Legal
Discussion

Edit the Workflow

Row Filter

The node allows for row filtering a criteria. It can include or exclude: row number), rows with a certain a certain value in a selectable col are the steps on how to configure configuration dialog. Note: the domain of the data table. I. e. the upper and lower possible values in the table spec are not adapted, ever or one value is fully filtered out.



Manipulator

KNIME Analytics Platform

File Edit View Node Help

100%

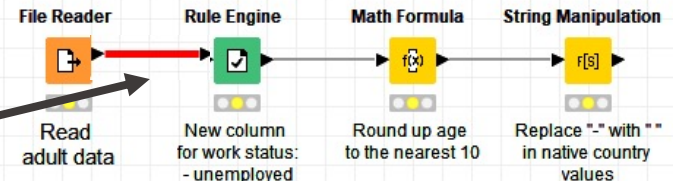
Quick Access

*0: 02_StringManipulation_MathFormula_RuleEngine

String Manipulation, Math Formula and Rule Engine Example

This workflow shows three different data manipulation operations, namely:

- creating three categories of people based on their weekly work hours with the Rule Engine node
- rounding up people's age to the nearest 10 with the Math Formula node
- replacing hyphens with " " characters in the native country column



File Reader
Read adult data

Rule Engine
New column for work status:
- unemployed
- employed part-time
- employed full-time

Math Formula
Round up age to the nearest 10

String Manipulation
Replace "-" with " " in native country values

Drag & Drop

Sharing the Workflow on the Hub

The screenshot displays the KNIME Analytics Platform interface. The top menu bar includes File, Edit, View, Node, and Help. The left sidebar contains the KNIME Explorer and Node Repository. The main workspace shows a workflow titled 'String Manipulation, Math Formula and Rule Engine Example'.

1. Save your Edits

2. Connect to KNIME Hub

String Manipulation, Math Formula and Rule Engine Example

This workflow shows three different data manipulation operations, namely:

- creating three categories of people based on their weekly work hours with the Rule Engine node
- rounding up people's age to the nearest 10 with the Math Formula node
- replacing hyphens with " " characters in the native country column

File Reader
Read adult data

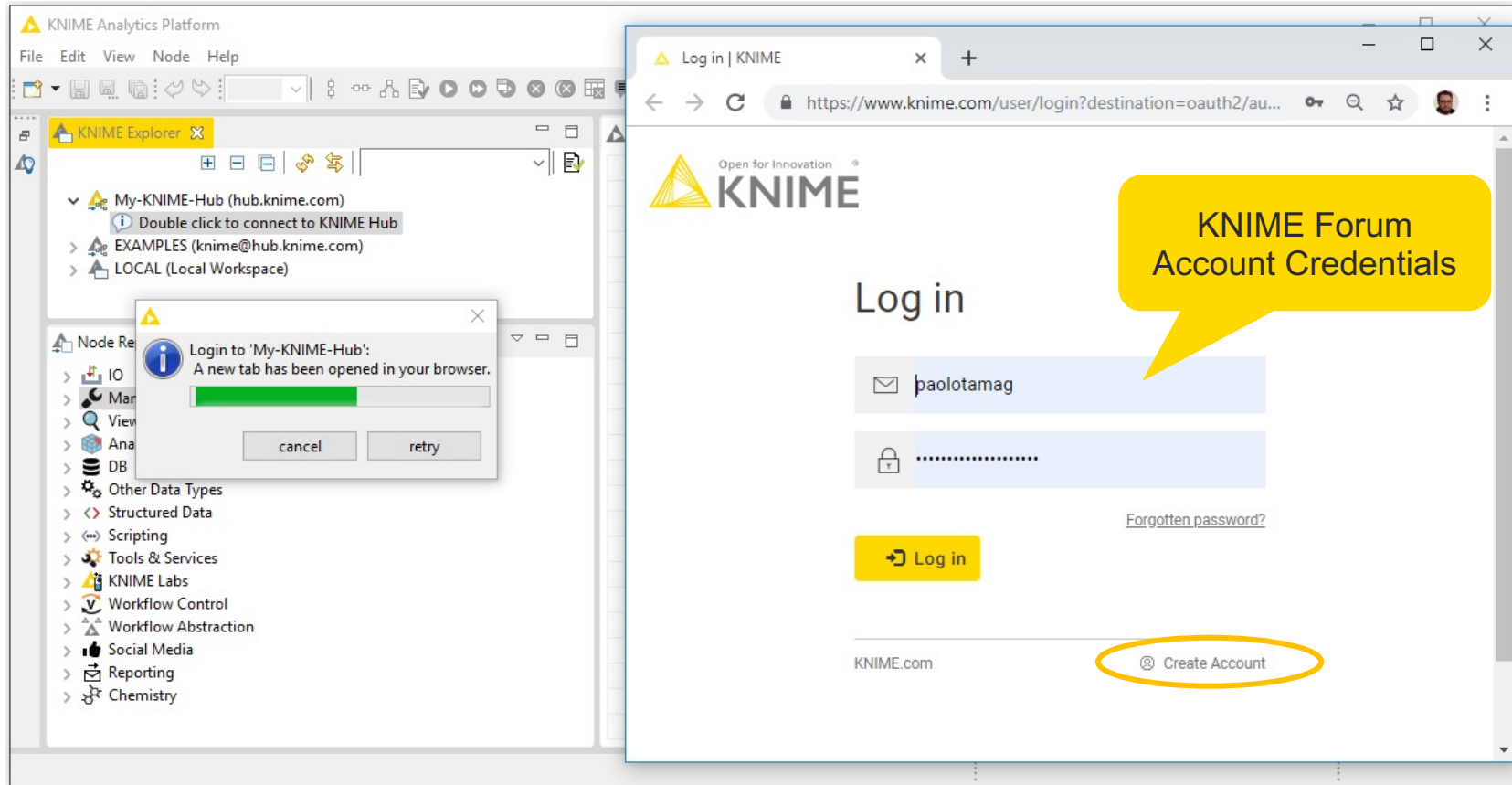
Row Filter
filter data

Rule Engine
New column for work status:
- unemployed
- employed part-time
- employed full-time

Math Formula
Round up age to the nearest 10

String Manipulation
Replace "-" with " " in native country values

Log in the Hub



Stay connected with KNIME



Blog: knime.com/blog



Forum: forum.knime.com



KNIME Hub: hub.knime.com



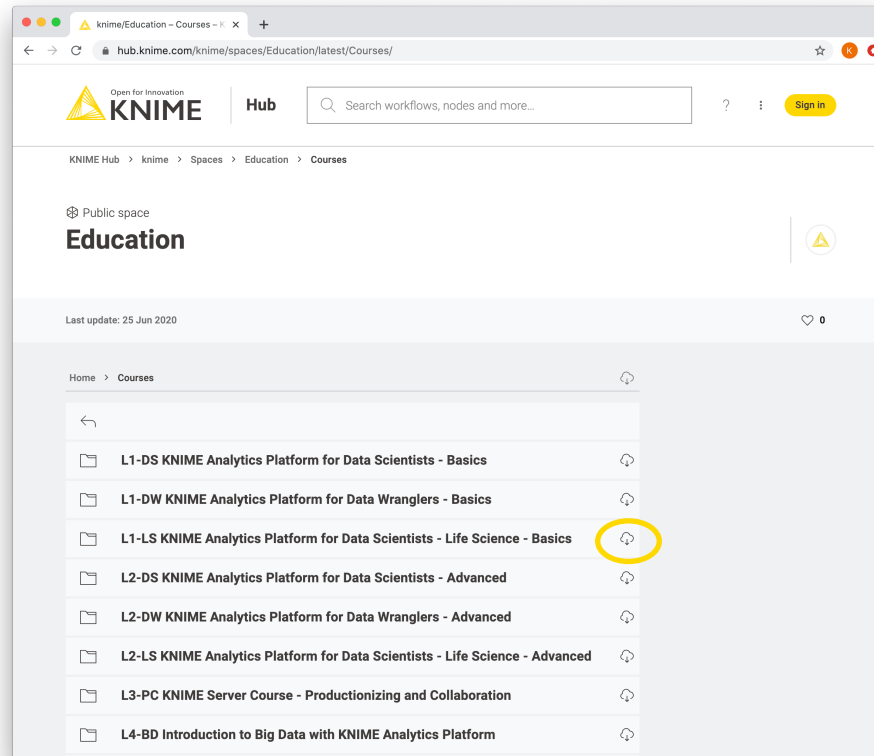
KNIME E-Learning Course:
www.knime.com/e-learning-course

Follow us on social media:



Exercise L1-LS:

- Download the course material from the KNIME Hub
<https://hub.knime.com/knime/spaces/Education/latest/Courses/>



KNIME Extensions

- ▶  KNIME & Extensions
- ▶  KNIME Big Data Extensions
- ▶  KNIME Community Extensions - Bioinformatics & NGS
- ▶  KNIME Community Extensions - Cheminformatics
- ▶  KNIME Community Extensions - Image Processing and Analysis
- ▶  KNIME Community Extensions - Other
- ▶  KNIME Community Extensions Sources
- ▶  KNIME Labs Extensions
- ▶  KNIME Node Development Tools
- ▶  KNIME Partner Extensions
- ▶  KNIME Server Extensions
- ▶  Marvin Chemistry Extensions (donated by Infocom & Chemaxon)

Selected Extensions from the Life Science Community

Cheminformatics

- ▼ RDKit
 - > Converters
 - > Modifiers
 - > Calculators
 - > Geometry
 - > Fingerprints
 - > Fragments
 - > Searching
 - > Reactions
 - > Viewing
 - > Experimental
 - > Testing
- ▼ Vernalis
 - > Chemistry
 - > Collections
 - > Databases
 - > European PubMed Central
 - > Fingerprints
 - > Flow Control
 - > IO
 - > Matched Molecular Pairs (MMPs)
 - > Plotting (Views)
 - > RCSB PDB Tools
 - > Local PDB Tools
 - > Sequence Tools
 - > Speedy SMILES
 - > Testing
 - > Miscellaneous
- ▼ CDK
 - > 3D
 - > AMBIT
 - > I/O
 - > 2D Coordinates
 - > Sgn Atom Signatures
 - > ChemSpider
 - > Connectivity
 - > Depiction
 - > Element Filter
 - > Fingerprint Similarity
 - > Fingerprints
 - > Hydrogen Manipulator
 - > Lipinski's Rule-of-Five
 - > Mass Calculator
 - > Molecular Properties
 - > OPSIN
 - > SMARTS Query
 - > Structure Sketcher
 - > Substructure Search
 - > Sugar Remover
 - > CH₃ Sum Formula
 - > Symmetry
 - > XLogP
- ▼ Lhasa Limited
 - > Generic
 - > Metabolism
 - > Reaction
- ▼ Chemistry
 - > I/O
 - > Mining
 - > Misc
 - > Translators
- ▼ Erwood Nodes
 - > IO
 - > Structure Data Format Converters
 - > Structure Similarity
 - > Structure Properties
 - > Virtual Screening
 - > Evaluation and Ranking
 - > SAR Analysis
 - > Viewers
 - > Testing

Bioinformatics

- ▼ SeqAn
 - > BS-Seq Analysis
 - > Benchmarking
 - > Data Mining
 - > Databases
 - > Error Correction
 - > Local Alignment
 - > Metagenomics
 - > NGS Quality Control
 - > NGS ROI Analysis
 - > Phylogeny
 - > Read Mapping
 - > Sequence Alignment
 - > Sequence Comparison
 - > Simulators
 - > Utilities
 - > Variant Detection
 - > Gustaf
 - > GustafMateJoining
- ▼ NGS
 - > IO
 - > ROI
 - > tools
- ▼ NgsToolbox
 - > Read Mapping
 - > SAM and BAM Manipulation
 - > SNP Calling
 - > VCF and BCF

Mass Spectrometry

- ▼ OpenMS
 - > Conversion
 - > File Handling
 - > ID Processing
 - > Identification
 - > Map Alignment
 - > Peptide property prediction
 - > QCFileHandling
 - > Quantitation
 - > Signal processing and preprocessing
 - > Targeted Experiments
 - > Utilities

High Content Screening

- ▼ HCS Tools
 - > Data Manipulation
 - > Data Views
 - > IO
 - > Normalization
 - > Population analysis
 - > Pre-Processing
 - > Quality Control
 - > Screen Mining
 - > Utilities

Selected Commercial Life Science Extensions



- ▼ BioSolveIT Nodes
 - ▶ CoLibri (Chemistry Spaces)
 - ▶ IO
 - Assess Affinity with Hyde in SeeSAR
 - Compute FTrees Similarity
 - Compute FlexS Alignments
 - Compute LeadIT Docking
 - Convert Molecules with Naomi
 - FTrees Query Generator
 - Filter Molecules with Naomi
 - FlexX Docking
 - Generate 3D Coordinates
 - Generate Protomers / Tautomers with Naomi
 - Interactive BioSolveIT Table
 - Interactive SeeSAR Viewer
 - Prepare Receptor with LeadIT
 - Run ReCore Interactively
 - Search FTrees Fragment Space
 - SeeSAR Project Generator



- ▼ ChemAxon / Infocom
 - ▼ JChem
 - ▶ IO
 - ▶ Converter
 - ▶ Marvin
 - ▶ Calculator Plugins
 - ▶ JChem Base
 - ▶ JChem Cartridge
 - ▶ Standardizer
 - ▶ Structure Checker
 - ▶ Name to Structure
 - ▶ Screen
 - ▶ JKlustor
 - ▶ Reactor
 - ▶ Markush Viewer
 - ▶ Metabolizer
 - ▶ Fragmenter
 - ▶ Marvin



- ▼ Cresset
 - ▼ Forge
 - ▶ Models
 - ▶ Project
 - ▶ Forge Align
 - ▶ Activity Miner
 - ▶ FieldTemplater
 - ▼ Spark
 - ▶ Spark Fragment Selector
 - ▶ Generate Spark Database
 - ▶ Spark Database Search
 - ▼ XedTools
 - ▶ XedMin
 - ▶ XedeX
 - ▶ Torch/Forge Molecule Viewer



- ▼ MOE
 - ▶ Input
 - ▶ Output
 - ▶ Convert
 - ▶ Transform
 - ▶ Process
 - ▶ Calculate
 - ▶ QuaSAR
 - ▶ Fingerprints
 - ▶ Simulations
 - ▶ Bioinformatics
 - ▶ Fragment Based Design
 - ▶ CombiChem
 - ▶ Miscellaneous
 - ▶ Pharmacophore
 - ▶ Materials



- ▶ Schrödinger
 - ▶ Readers/Writers
 - ▶ Converters
 - ▶ Ligand Preparation
 - ▶ Property Generation
 - ▶ Cheminformatics
 - ▶ Pharmacophore Modeling
 - ▶ Protein Structure Prediction
 - ▶ Docking and Scoring
 - ▶ Molecular Mechanics
 - ▶ Molecular Dynamics
 - ▶ Quantum Mechanics
 - ▶ Workflows
 - ▶ Filtering
 - ▶ Reporting
 - ▶ Scripting
 - ▶ Tools

inte:ligand

Your partner for in-silico drug discovery.

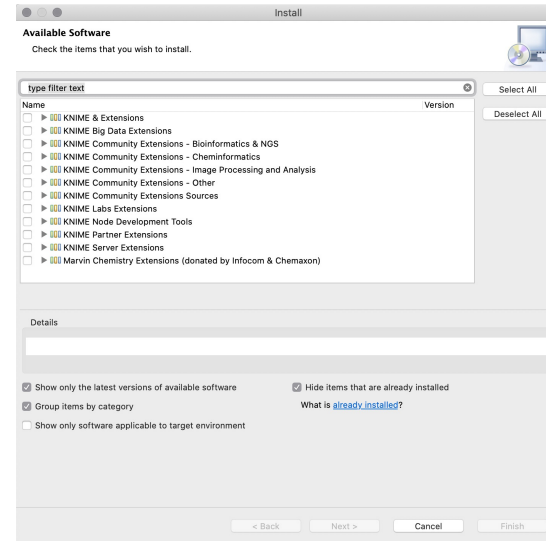
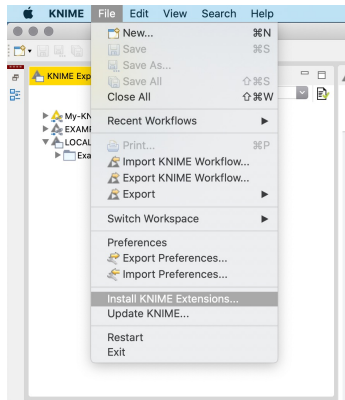


Further extensions including detailed descriptions can be found at <https://hub.knime.com>

Installing KNIME Extensions

There are different ways to install KNIME extensions:

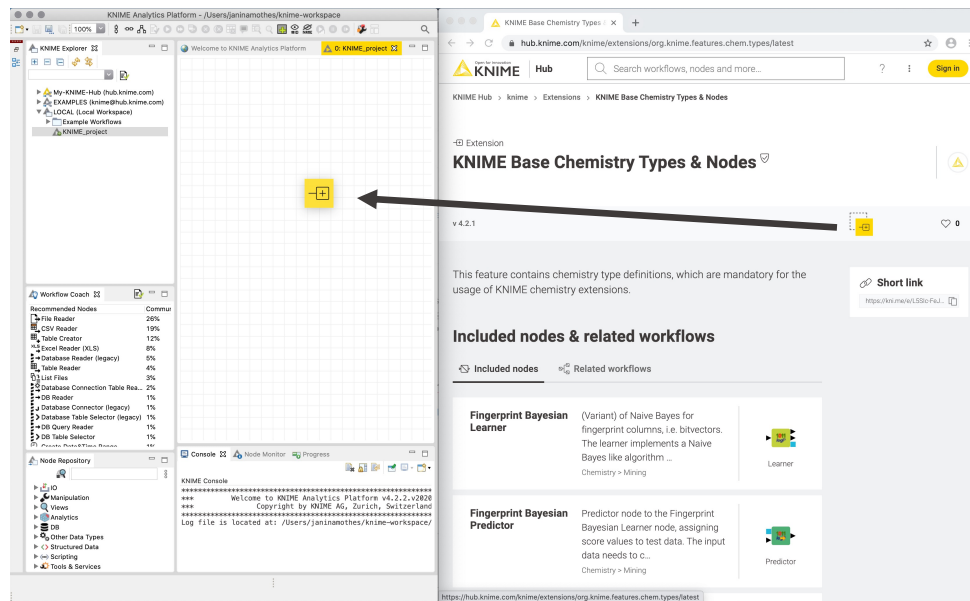
Option 1: Via the menu bar in the KNIME Analytics Platform and select your extension from the list (feasible if you know the name of the extension)



Installing KNIME Extensions

There are different ways to install KNIME extensions:

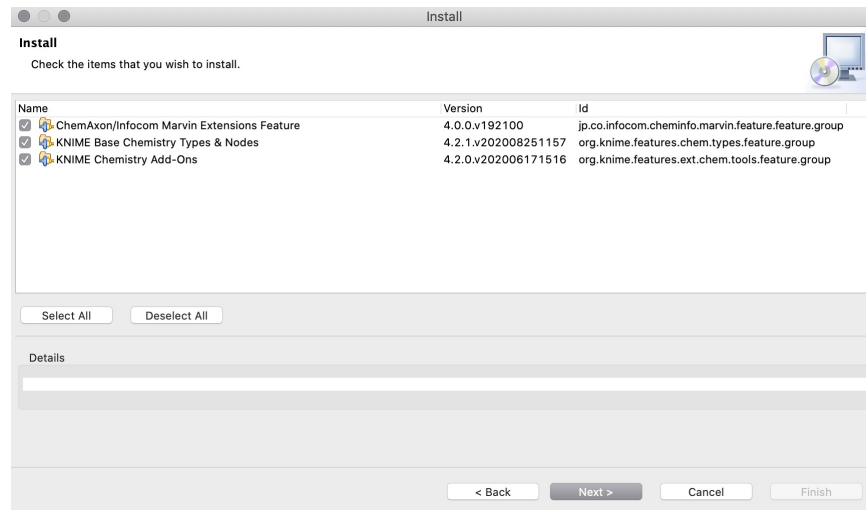
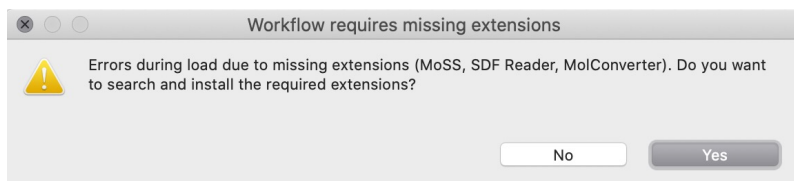
Option 2: Via the KNIME Hub (hub.knime.com): Drag-and-drop the required extension (or node) to your workflow



Installing KNIME Extensions

There are different ways to install KNIME extensions:

Option 3: Opening a workflow that contains a node from an extension (KNIME automatically detects the necessary extensions)



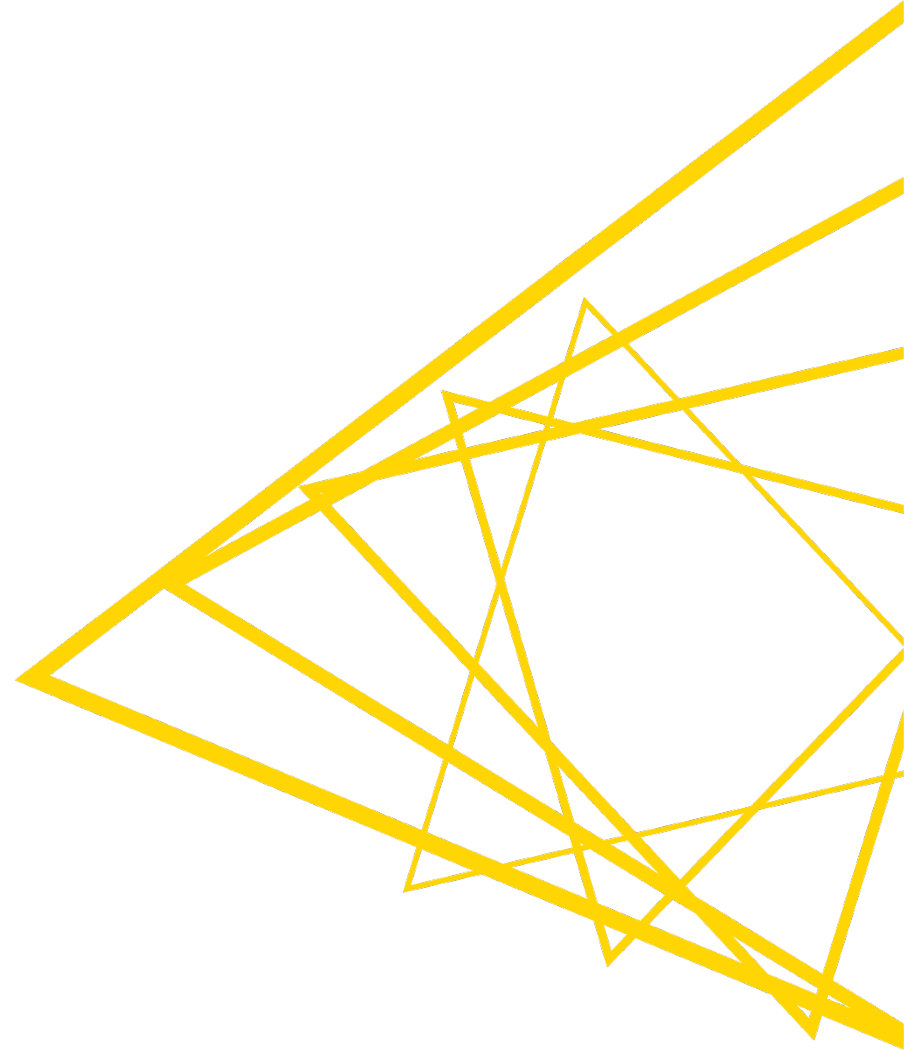
Installing Extensions Exercise

Open Exercise: *00. Installing Extensions*

Activity 1: Install the following Extensions

- KNIME Base Chemistry Types & Nodes
- KNIME JavaScript Views (Labs)
- ...

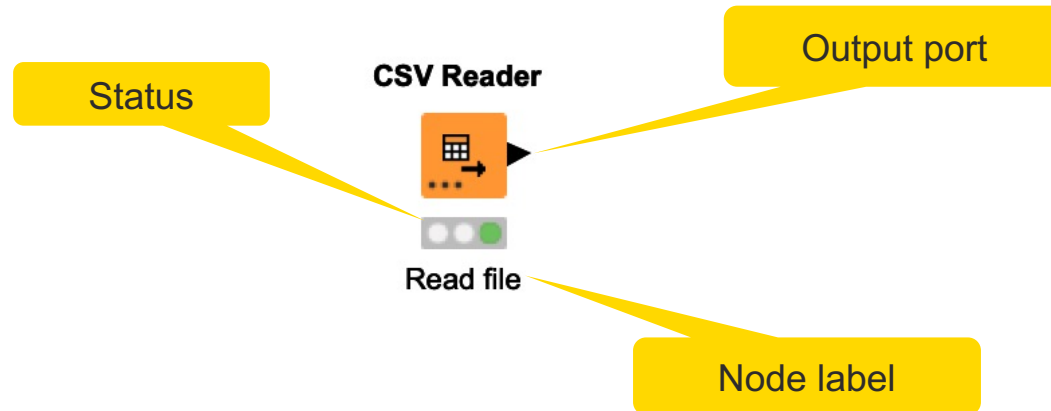
Importing Data



Data Source Nodes

Typically characterized by:

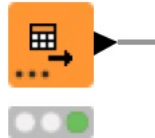
- Orange color
- By default no input ports, 1-2 output ports
- New file handling with KNIME 4.3.
 - Consistent user experience across all nodes and file systems
 - Managing of various file systems within the same workflow
 - Performance improvements



CSV Reader

- Reads either one or multiple .csv and .txt files
- Further tabs to
 - limit the rows
 - select encoding

CSV Reader



File system

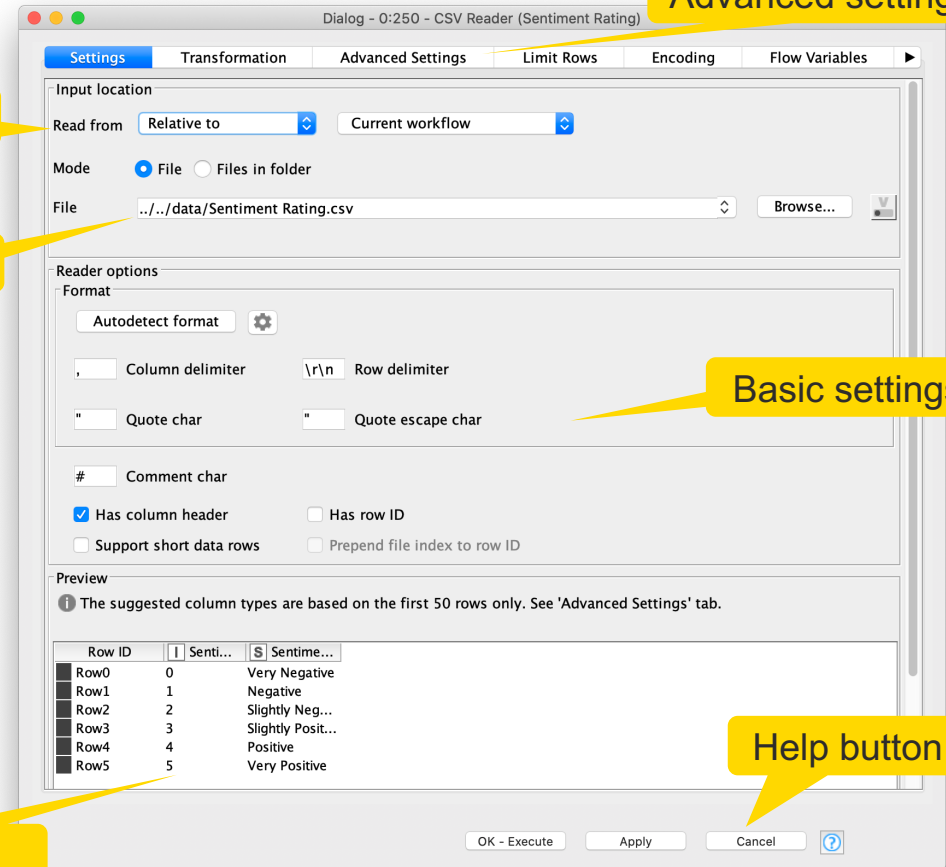
File path

Advanced settings

Basic settings

Help button

Preview



Dialog - 0:250 - CSV Reader (Sentiment Rating)

Settings Transformation Advanced Settings Limit Rows Encoding Flow Variables

Input location

Read from Relative to Current workflow

Mode ☒ File ☐ Files in folder

File ../../data/Sentiment Rating.csv Browse...

Reader options

Format

Autodetect format

Column delimiter , Row delimiter \r\n

Quote char " Quote escape char "

Comment char

☒ Has column header ☐ Has row ID

☐ Support short data rows ☐ Prepend file index to row ID

Preview

The suggested column types are based on the first 50 rows only. See 'Advanced Settings' tab.

Row ID	I Senti...	S Senti...
Row0	0	Very Negative
Row1	1	Negative
Row2	2	Slightly Neg...
Row3	3	Slightly Posit...
Row4	4	Positive
Row5	5	Very Positive

OK - Execute Apply Cancel ?


Common Settings: Four Default File Systems

■ Local File System


Input location


Read from:

Mode: ☒ File ☐ Files in folder

File: 



■ Relative to ...


Read from: 

File: 


Current mountpoint
Current workflow data area
Current workflow


■ Mountpoint

Read from:  

File: 

■ Custom URL

Read from: 

URL: 

File Path Options Old File Handling

- Local path

Enter ASCII data file location: (press 'Enter' to update preview)

- Absolute URL

Enter ASCII data file location: (press 'Enter' to update preview)

- Mountpoint-relative URL

Enter ASCII data file location: (press 'Enter' to update preview)

New file handling

Input location

Read from

Mode ☒ File ☐ Files in folder

File

Input location

Read from

Mode ☒ File ☐ Files in folder

File

Input location

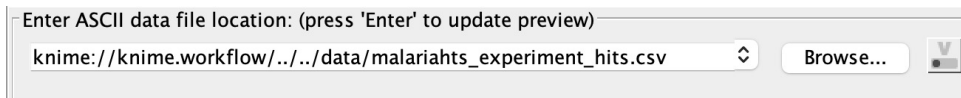
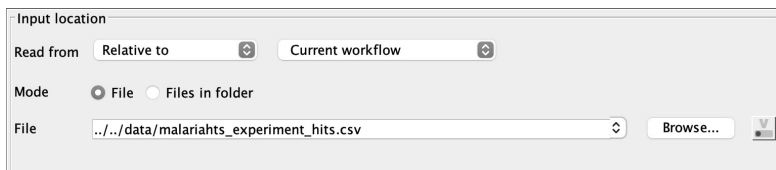
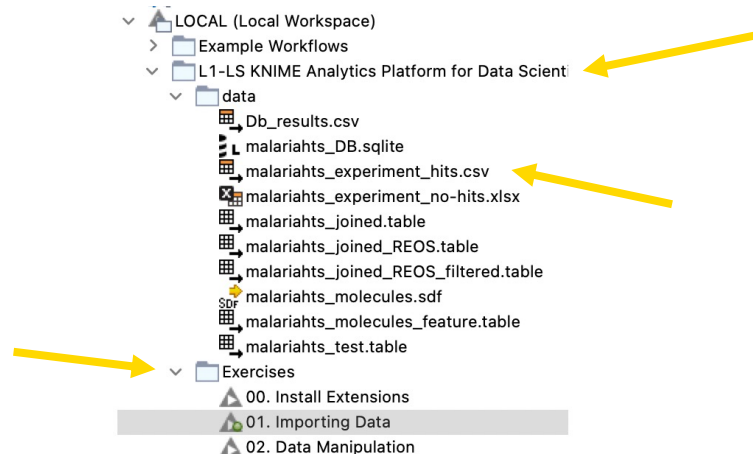
Read from

Mode ☒ File ☐ Files in folder

File

Workflow-Relative File Paths

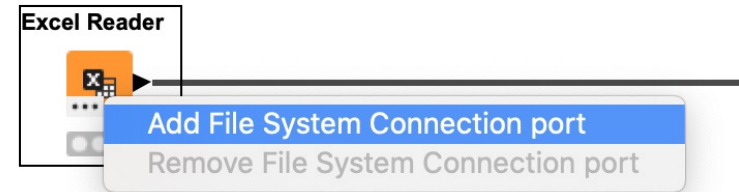
- Best choice if workflows are to be shared
- Requires matching folder structure within workflow group
 - Independent of environment outside of workflow group
- Example: Path to „Sentiment Analysis.table“
 - Local path:
 - C:\Users\rb\knime-workspace\KNIMEUserTraining\data\malariahts_experiments_hits.csv
 - Workflow relative:



YouTube KNIME TV Channel:
<https://youtu.be/U9sP4g4yGwY>

Common Settings: Connecting to other File Systems

- Add file system connection port to connect to another file system
 - Click on the three dots on the lower left to add or remove a dynamic port.
- Supported file systems
 - Microsoft Azure
 - Google
 - Amazon
 - Databricks
 - BigData file systems (hdfs, httpFS, ...)
 - On-premise (e.g. ssh, ftp, ...)



The image shows the configuration window for the 'Excel Reader' widget, specifically the 'Input location' section. It includes a 'Read from' dropdown menu currently set to 'Amazon S3'. Below this, the 'Mode' is set to 'File' (selected with a radio button) instead of 'Files in folder'. At the bottom, there is a text field labeled 'File' which is currently empty.

Common Settings: Read Single or Multiple Files

■ Single file

Input location

Read from: Local File System

Mode: ☒ File ☐ Files in folder

File: /Users/kathrinmelcher/Desktop/course_data.csv

Browse...

■ Files in a folder

Input location

Read from: Relative to Current workflow

Mode: ☐ File ☒ Files in folder

Filter options ☐ Include subfolders

Folder: ../../data/

Browse...

Selected 22 of 22 files

- Option to include subfolder
- Option to define filter criterions

Filter options

File filter options

☒ File extension(s) .csv

☐ Case sensitive

☐ File name *

☐ Case sensitive ☒ Wildcard ☐ Regular expression

☐ Include hidden files

Folder filter options

☒ Folder name month

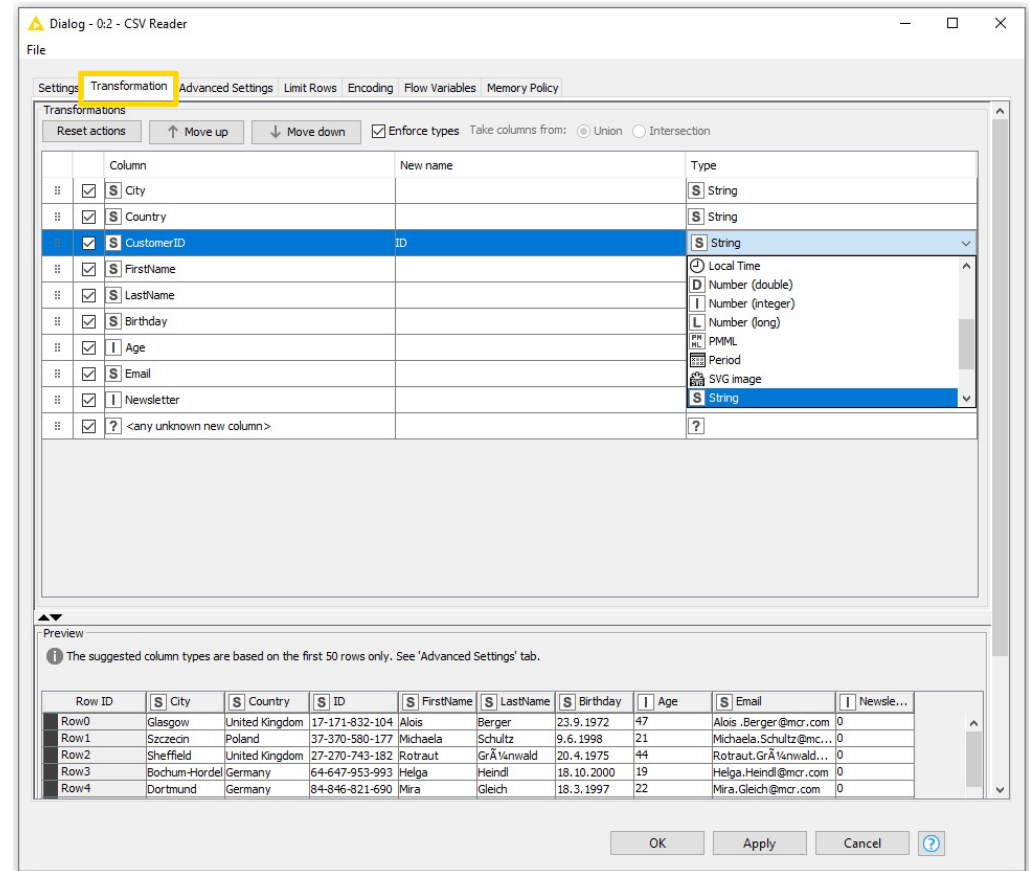
☐ Case sensitive ☒ Wildcard ☐ Regular expression

☐ Include hidden folders

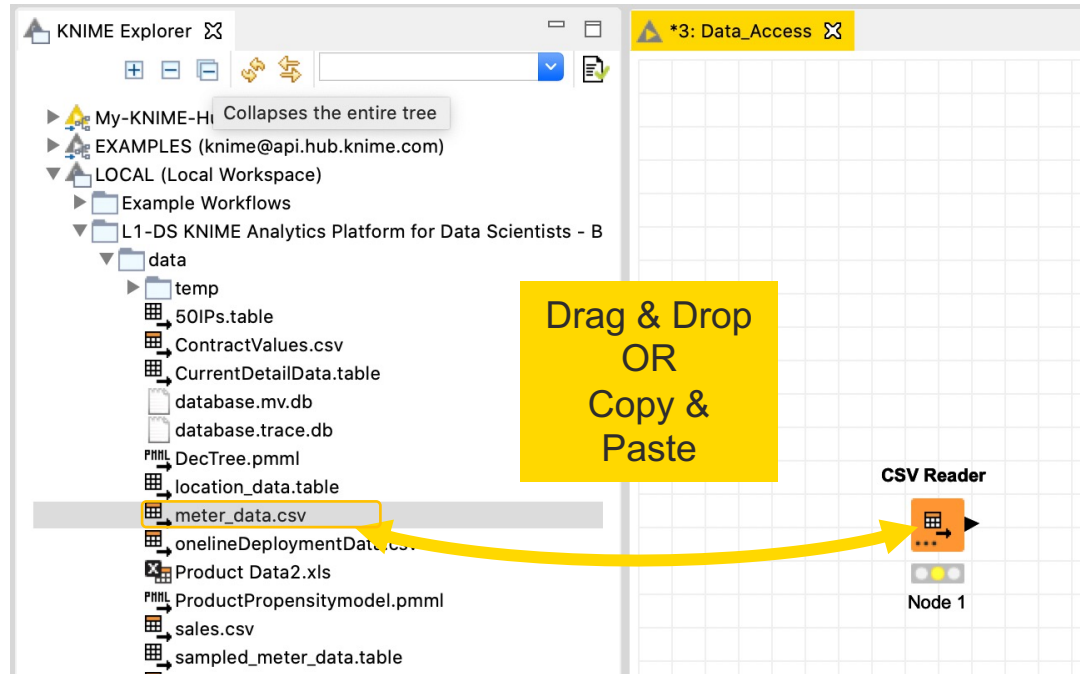
OK Cancel

Common Settings: Transformation Tab

- Supported operations
 - Column filtering
 - Column sorting
 - Column renaming
 - Column type mapping
 - Select between union or intersection of columns (in case of reading many files)



Alternative Faster Way ...

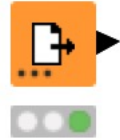


File Reader/ File Reader (Complex Format)

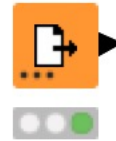
Good option, if the CSV Reader node can't read your file

- Reads all text-based files (e.g. csv, txt, etc.)
- Many advanced features allow it to read most 'weird' files
 - Short lines, inline comments, headers and special encoding

File Reader



**File Reader
(Complex Format)**



YouTube KNIME TV Channel video:

<https://youtu.be/flaHQw-Qhlq>

File Reader

Dialog - 0:8 - File Reader

Settings Transformation Advanced Settings Limit Rows Encoding Flow Variables

Input location

Read from Relative to Current workflow

Mode ☒ File ☐ Files in folder

File ../../data/malariahts_experiment_hits.csv Browse...

Reader options

Format

Autodetect format

Column delimiter Row delimiter ☒ Line break ☐ Custom \n

Quote char Quote escape char

Comment char

☒ Has column header ☐ Has row ID

☐ Support short data rows ☐ Prepend file index to row ID

Preview

The suggested column types are based on the first 10000 rows only. See 'Advanced Settings' tab.

Row ID	S Sample	D Pf3D7...	D Pf3D7...	S Pf3D7...	D Pf3D7...
Row0	Sj000259230-1	102.523	80.893	true	?
Row1	Sj000282539-1	113.112	88.908	true	6.227
Row2	Sj000033142-1	122.304	104.185	true	6.069
Row3	Sj000079671-1	105.88	97.726	true	4.824
Row4	Sj000179372-1	98.684	84.861	true	6.142
Row5	Sj000276817-1	98.44	86.477	true	5.14
Row6	Sj000273047-1	100.665	80.332	true	?
Row7	Sj000260256-1	102.336	96.501	true	4.824
Row8	Sj000123502-1	113.261	101.839	true	5.433
Row9	Sj000170548-1	107.117	111.154	true	5.416
Row10	Sj000092590-1	105.653	98.151	true	6.121
Row11	Sj000033131-1	116.077	96.145	true	7.284
Row12	Sj000257328-1	111.211	122.335	true	?
Row13	si000117911-1	114.424	84.23	true	6.239

OK Apply Cancel ?

File system

File path

Basic settings

Autodetect
Format

Preview

Help Button

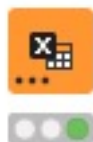
Excel Reader (XLS)

- Reads .xls and .xlsx file from Microsoft Excel
- Supports reading from multiple sheets



Excel Reader

Excel Reader



File system

File path

Sheet specific settings

Preview

Dialog - 0:1 - Excel Reader

File

Settings Transformation Advanced Settings Flow Variables Memory Policy

Input location

Read from: Relative to Current workflow

Mode: ☒ File ☐ Files in folder

File: ../../data/Product Data2.xls Browse...

Sheet selection

☒ Select first sheet with data (Product Data.xls_defa...)

☐ Select sheet with name Product Data.xls_defa...

☐ Select sheet at index 0 (Sheet indexes start with 0.)

Column header

☒ Table contains column names in row number 1 (Row numbers start with 1. See "File Content" tab to identify row numbers.)

Row ID

☒ Generate row IDs ☐ Table contains row IDs in column A

Sheet area

☒ Read entire data of the sheet ☐ Read only data in columns from A to and rows from 1 to . (See "File Content" tab to identify columns and rows.)

Preview File Content

Preview with current settings

The suggested column types are based on the first 50 rows only. See "Advanced Settings" tab.

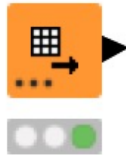
Row ID	I Custom...	S Products
Row0	11000	Private Investment
Row1	11001	Private Investment
Row2	11002	Private Investment
Row3	11003	Private Investment
Row4	11004	Private Investment

OK Apply Cancel ?

Table Reader

- Reads tables from the native KNIME Format
- Maximum performance, minimum configuration

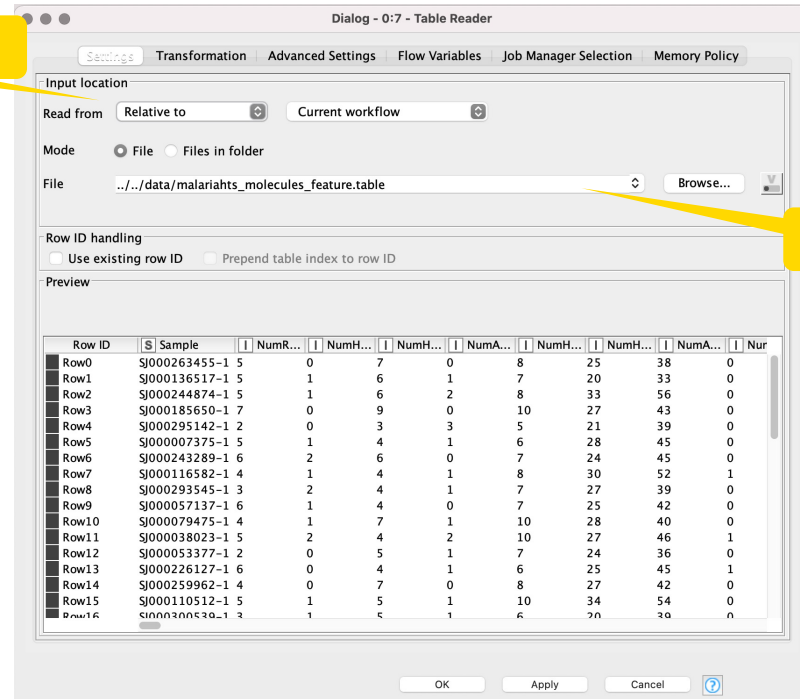
Table Reader



File system

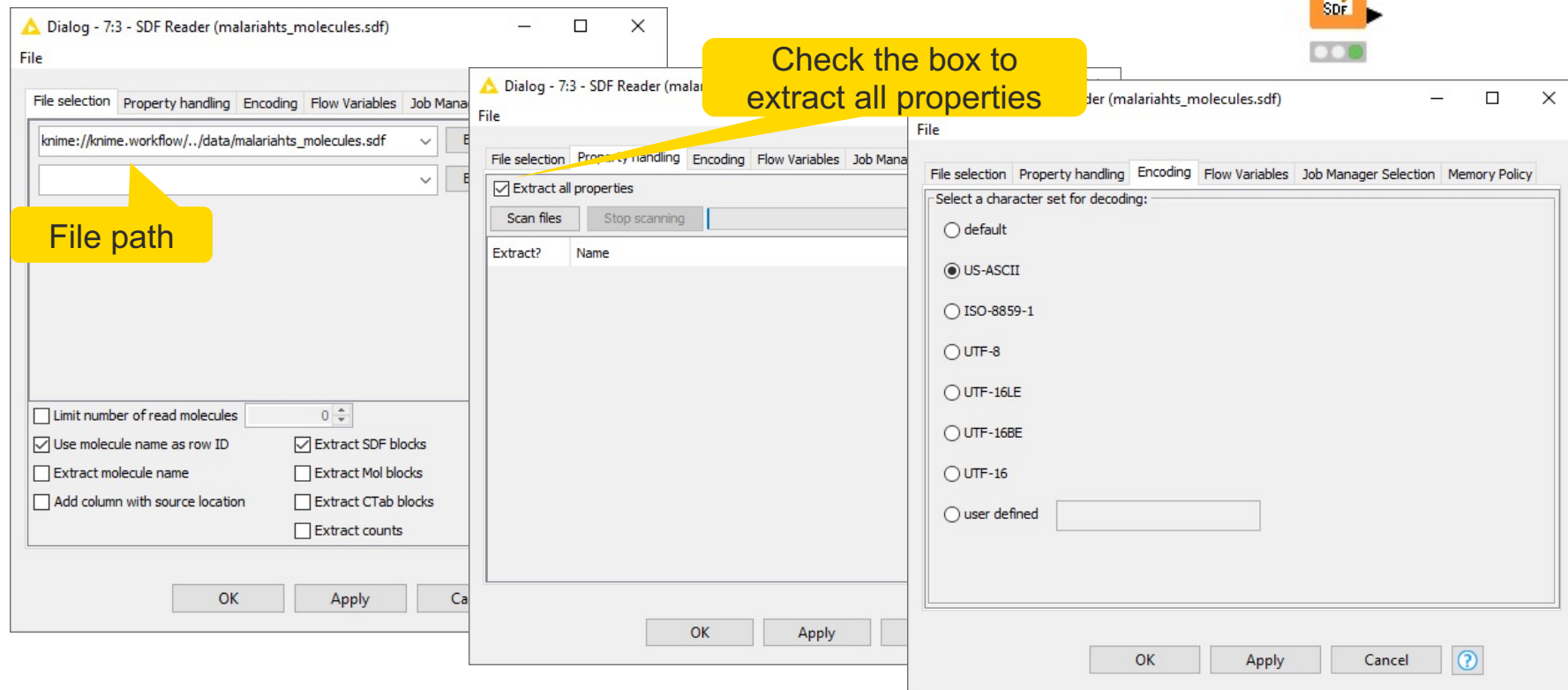
File path

YouTube KNIME TV channel video:
<https://youtu.be/tid1qi2HAOo>



SDF Reader


- Reads chemical data from the .sdf file



required Extension: KNIME Base Chemistry Types & Nodes

Table Creator

Table Creator



Dialog - 0:1 - Table Creator

File

Table Creator Settings | Flow Variables | Job Manager Selection | Memory Policy

Input line: O=C(O)c1ccccc1O

	S ChemblID	SMI SMILES	S assay_type
Row0	CHEMBL10	<chem>O=C(O)c1ccccc1O</chem>	Binding
Row1	CHEMBL1006	<chem>CC(=O)Oc1ccccc1C(=O)O</chem>	ADME
Row2	CHEMBL23455	<chem>O=C(O)c1cc(N=N/c2ccc(S(=O)(=O)cc2)cc1</chem>	Functional
Row3	CHEMBL10982	<chem>O=C(O)c1cc(-c2ccc(F)cc2F)ccc1O</chem>	Binding
Row4			
Row5			
Row6			
Row7			
Row8			
Row9			
Row10			
Row11			
Row12			

OK Apply Cancel ?

New settings for column 'S...' X

Column Properties

☐ DON'T include column in output table

Name: SMILES

Type: Smiles

miss. value pattern:

Format:





Domain...

OK Cancel

Manually created table - 0:1 - Table Creator

File Hilite Navigation View

Table "default" - Rows: 4 Spec - Columns: 3 Properties Flow Variables

Row ID	S ChemblID	SMI SMILES	S assay_...
Row0	CHEMBL10		Binding
Row1	CHEMBL1006		ADME
Row2	CHEMBL23455		Functional
Row3	CHEMBL10982		Binding

Other Useful Data Sources

- KNIME Analytics Platform provides many more options to access data:
 - PMML Reader – reads standard predictive models
 - XML Reader with XPATH support
 - Python/R Source nodes
 - Tika Parser – extracts textual data from 200+ file types
 - REST Web Services, and many more

XML Reader



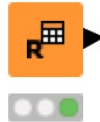
Tika Parser



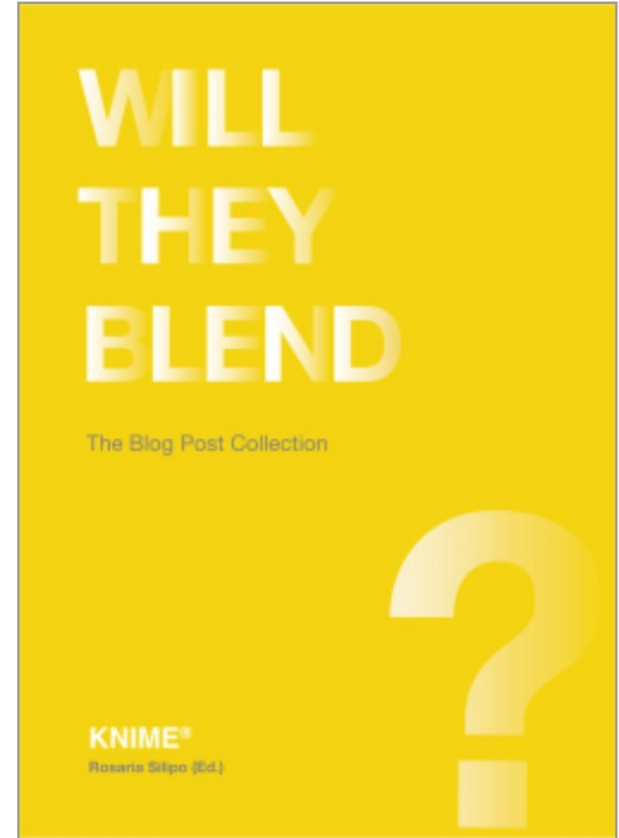
PMML Reader



R Source (Table)



- Find out more in by downloading the free book “Will they blend”
<https://www.knime.com/knimepress/download-will-they-blend>



Our example data set



- Malaria High-throughput Screening data
- HTS hit list and compounds with confirmed IC50 data in malaria *Plasmodium falciparum* whole cell assay
- Content of the original dataset:

Column Name	Description
SAMPLE	sample identifier
Pf3D7_ps_green	primary screen, measuring green fluorescence intensity
Pf3D7_ps_red	primary screen, measuring red fluorescence intensity
Pf3D7_ps_hit	standardized call on hits: 'true' if activity in red AND green 80% $\leq x < 250\%$; 'false' if activity in red AND green $< 20\%$; 'ambiguous' for all other compounds (20 – 80%, or $> 250\%$)
Pf3D7_pEC50	Reported pEC50 value (NA for compounds not submitted for dose-response confirmation)
Canonical_Smiles	standardized structure information

<http://www.tdtproject.org/challenge-1---malaria-hts.html#>

Importing Data Exercise

Open exercise 01. Importing Data

Activity I: Importing Data

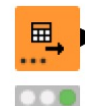
- Read the following files:
 - *malariahts_experiment_hits.csv*
 - *malariahts_experiment_no-hits.xlsx*
 - *malariahts_molecules.sdf*
 - *malariahts_molecules_feature.table*

Hint: drag and drop the files from the KNIME Explorer panel to get started

You can download the training workflows from the KNIME Hub:

<https://hub.knime.com/knime/spaces/Education/latest/Courses/>

CSV Reader



SDF Reader



Excel Reader

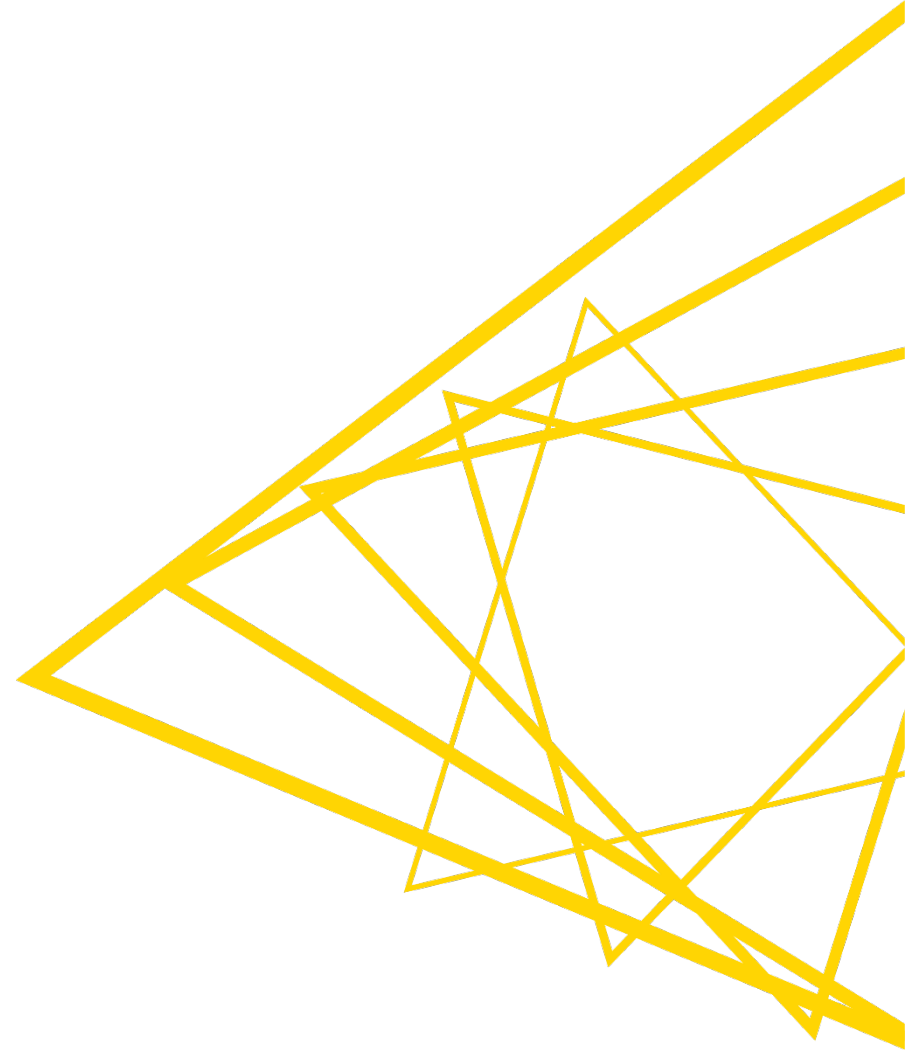


Table Reader



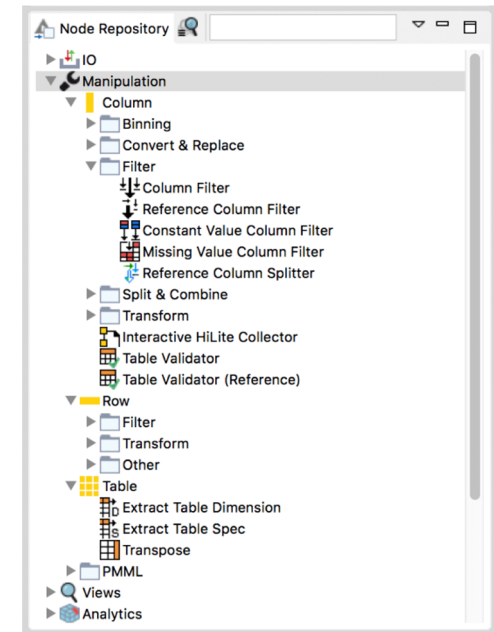
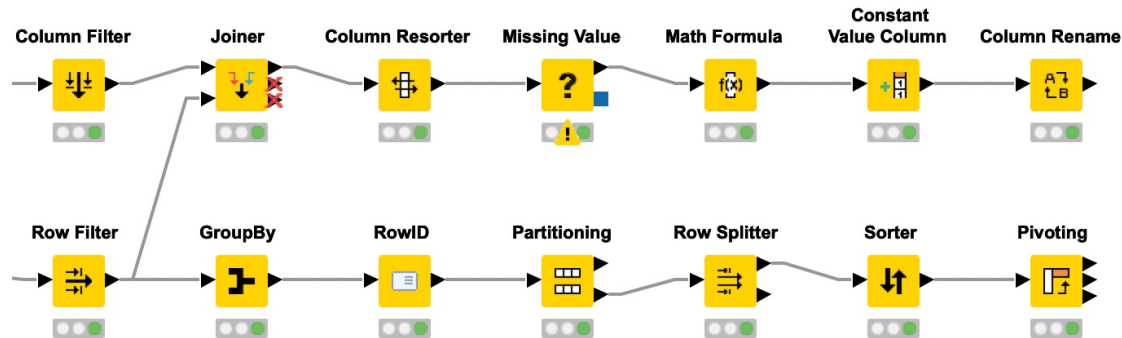
Data Manipulation

Clean, Join, Aggregate



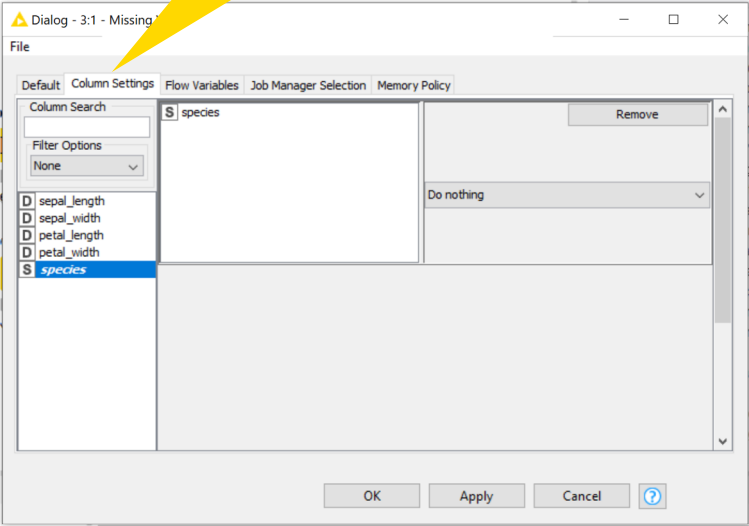
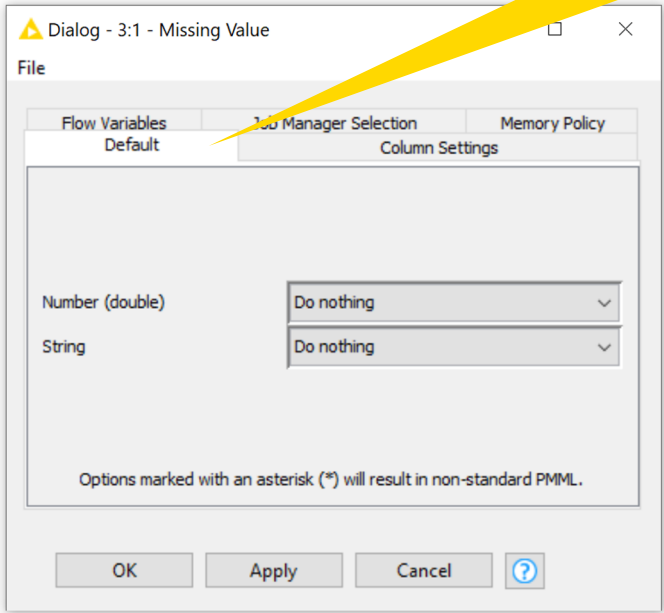
Data Manipulation Nodes

- Yellow color with a variety of input and output ports
- Apply a transformation to input data
- Many, many nodes!



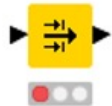
Missing Value

helps to handle missing values



Row Filter

Row Filter



Dialog - 7:28 - Row Filter (Pf3D7_ps_red < 150)

File

Filter Criteria | Flow Variables | Job Manager Selection | Memory Policy

Column value matching

Column to test:

☐ filter based on collection elements

Matching criteria

☐ use pattern matching

☐ case sensitive match ☐ contains wild cards

☐ regular expression

☒ use range checking

lower bound:

upper bound:

☐ only missing values match

☒ Include rows by attribute value

☐ Exclude rows by attribute value

☐ Include rows by number

☐ Exclude rows by number

☐ Include rows by row ID

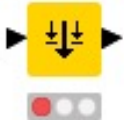
☐ Exclude rows by row ID

OK Apply Cancel ?

- Row
 - Filter
 - Duplicate Row Filter
 - Filter Apply
 - Filter Apply Row Splitter
 - Filter Definition Merger
 - HiLite Row Splitter
 - Nominal Value Row Filter
 - Nominal Value Row Splitter
 - Numeric Row Splitter
 - Reference Row Filter
 - Reference Row Splitter
 - Row Filter
 - Row Splitter
 - Rule-based Row Filter
 - Rule-based Row Filter (Dictionary)
 - Rule-based Row Splitter
 - Rule-based Row Splitter (Dictionary)

Column Filter

Column Filter



- Manipulation
 - Column
 - Filter
 - Column Filter
 - Reference Column Filter
 - Constant Value Column Filter
 - Missing Value Column Filter
 - Reference Column Splitter

Dialog - 7:25 - Column Filter (Remove)

File

Column Filter | Flow Variables | Job Manager Selection | Memory Policy

☒ Manual Selection ☐ Wildcard/Regex Selection ☐ Type Selection

Exclude

Filter

No columns in this list

☒ Enforce exclusion

Include

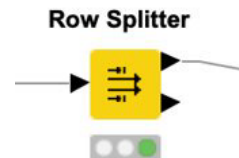
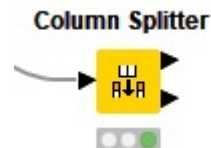
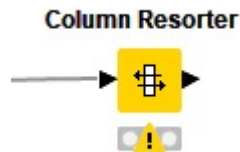
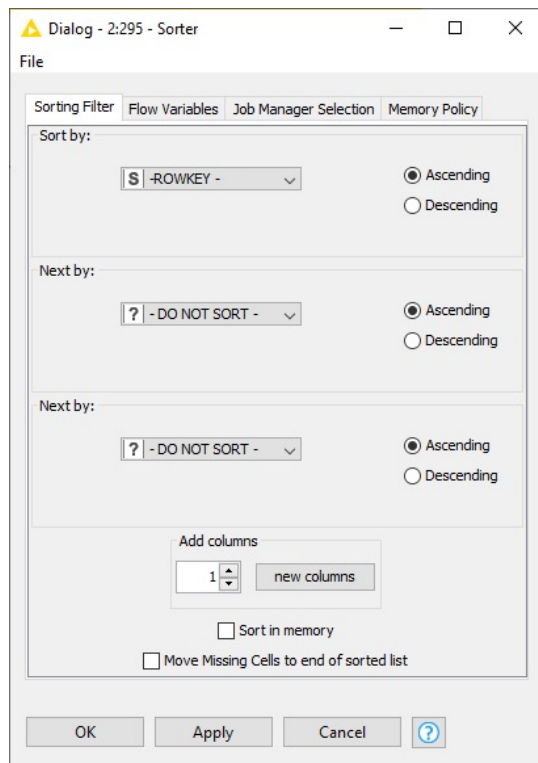
Filter

S	Sample
D	Pf3D7_ps_green
D	Pf3D7_ps_red
S	Pf3D7_ps_hit
D	Pf3D7_pEC50

☐ Enforce inclusion

OK Apply Cancel ?

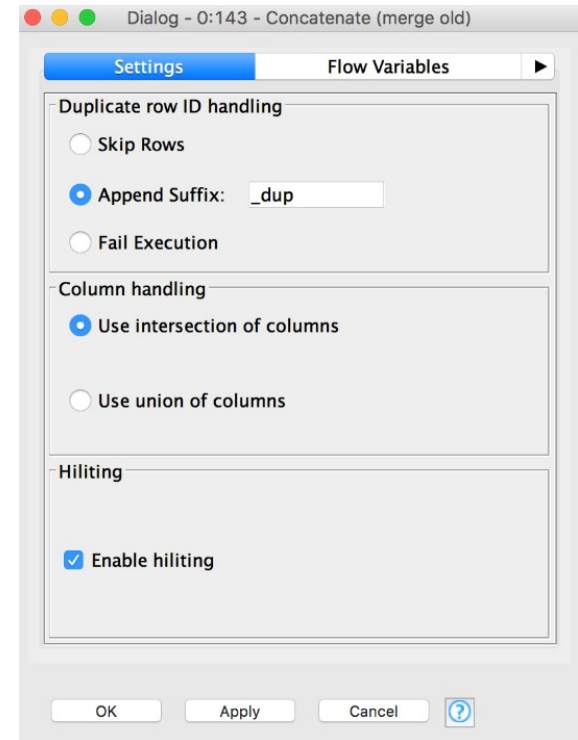
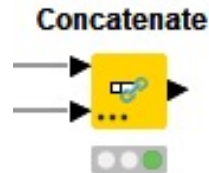
Other Useful Nodes for Row and Column Handling



Concatenate

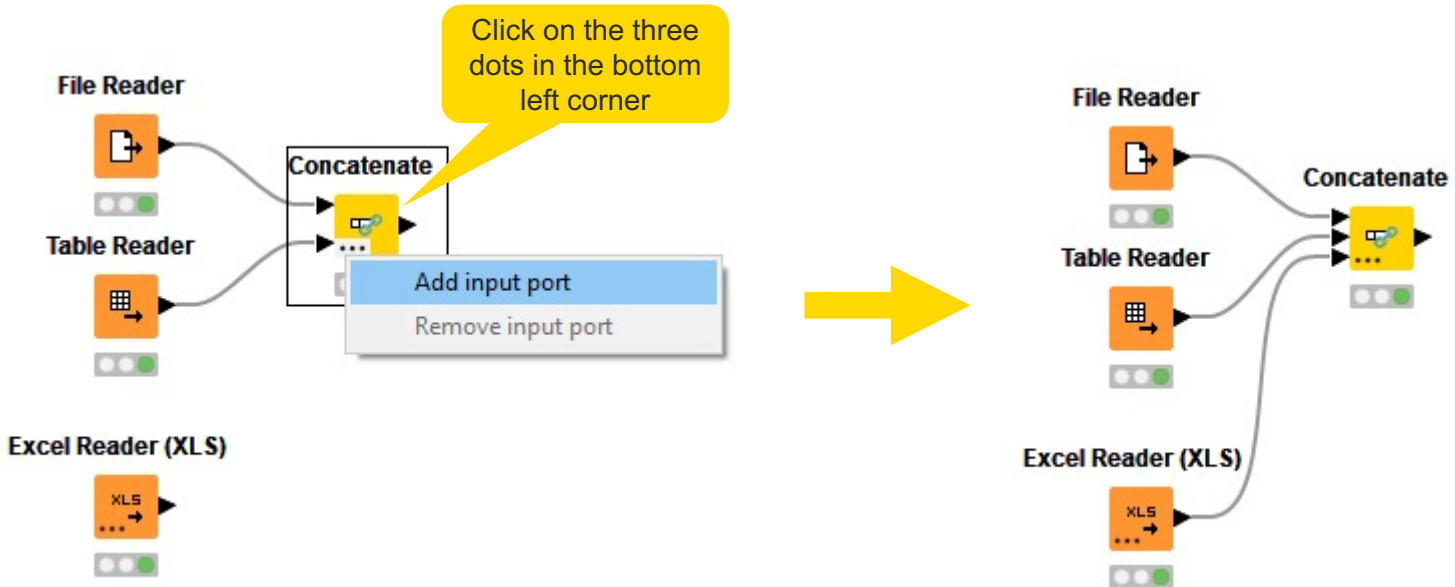
Combine rows from 2 or more tables with shared columns

- Handles duplicate row keys gracefully
- Take the union or intersection of columns



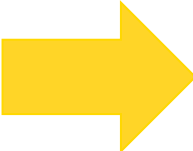
Dynamic Ports

Add and remove node ports based on your needs, e.g. in order to concatenate three or more tables

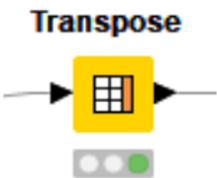


Transpose

Row ID	S column1	S column2	S column3
Row0	1		3
Row1	1		3
Row2	1		3
Row3	1		3



Row ID	S Row0	S Row1	S Row2	S Row3
column1	1	1	1	1
column2	2	2	2	2
column3	3	3	3	3



Dialog - 3:4 - Transpose

File

Options

Flow Variables

Job Manager Selection

Memory Policy

Chunk size (columns): 10

OK

Apply

Cancel

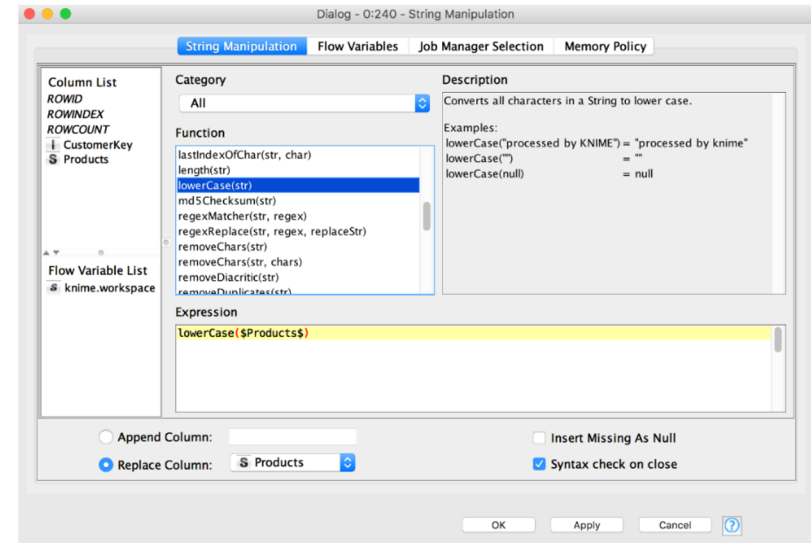
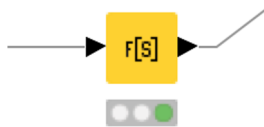
?

String Manipulation

Create and edit values in String columns

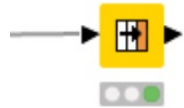
- Clean up capitalization (e.g. Lowercase)
- Replace strings
- Modify existing strings or create new columns

String Manipulation



More Nodes

String Replacer



Dialog - 2:297 - String Replacer

File

Standard settings | Flow Variables | Job Manager Selection | Memory Policy

Target column:

Pattern type:
☒ Wildcard pattern
☐ Regular expression

Pattern:

Replacement text:

Replace ...:
☒ ... whole string
☐ ... all occurrences

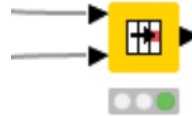
Case sensitive search: ☒

Use backslash as escape character: ☐

Append new column: ☐

OK Apply Cancel ?

Cell Replacer



Dialog - 2:298 - Cell Replacer

File

Options | Flow Variables | Job Manager Selection | Memory Policy

Input table:
Target column:

Dictionary table:
Input (Lookup):
Output (Replacement):

Append/Replace Result Column:
☒ Append new column

If no element matches use:
☐ Input ☒ Missing

Metadata in Output:
☒ Copy metadata from replacement column

OK Apply Cancel ?

Joining Columns of Data

Left Table

Mol Reg No	Chembl ID	SMILES
22	CHEMBL1794855	CCCN(CCC)
24	CHEMBL278751	CCN(C)
15	CHEMBL103772	CCCN1CC
10	CHEMBL328107	C1CN(CCN1)

Join by Mol Reg No

Inner Join

Right Table

Mol Reg No	Ki value	Ki relation	Ki unit
17	76.0	=	nM
65	6.56	=	nM
35	100	>	nM
15	8	=	nM
10	95.8	=	nM

Left Outer Join

Mol Reg No	Chembl ID	SMILES	Ki value	Ki relation	Ki unit
15	CHEMBL103772	CCCN1CC	8	=	nM
10	CHEMBL328107	C1CN(CCN1)	95.8	=	nM

Right Outer Join

Mol Reg No	Chembl ID	SMILES	Ki value	Ki relation	Ki unit
22	CHEMBL1794855	CCCN(CCC)	?	?	?
24	CHEMBL278751	CCN(C)	?	?	?
15	CHEMBL103772	CCCN1CC	8	=	nM
10	CHEMBL328107	C1CN(CCN1)	95.8	=	nM

Mol Reg No	Chembl ID	SMILES	Ki value	Ki relation	Ki unit
17	?	?	76.0	=	nM
65	?	?	6.56	=	nM
35	?	?	100	>	nM
15	CHEMBL103772	CCCN1CC	8	=	nM
10	CHEMBL328107	C1CN(CCN1)	95.8	=	nM

Joining Columns of Data

Left Table

Mol Reg No	Chembl ID	SMILES
22	CHEMBL1794855	CCCN(CCC)
24	CHEMBL278751	CCN(C)
15	CHEMBL103772	CCCN1CC
10	CHEMBL328107	C1CN(CCN1)

Join by Mol Reg No

Full Outer Join

Right Table

Mol Reg No	Ki value	Ki relation	Ki unit
17	76.0	=	nM
65	6.56	=	nM
35	100	>	nM
15	8	=	nM
10	95.8	=	nM

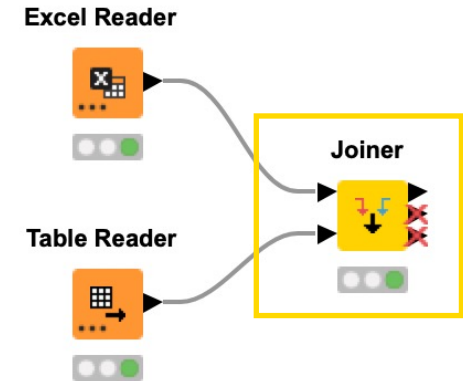
Missing values in the left table

Mol Reg No	Chembl ID	SMILES	Ki value	Ki relation	Ki unit
17	?	?	76.0	=	nM
65	?	?	6.56	=	nM
35	?	?	100	>	nM
15	CHEMBL1794855	CCCN(CCC)	8	=	nM
10	CHEMBL278751	CCN(C)	95.8	=	nM
22	CHEMBL103772	CCCN1CC	?	?	?
24	CHEMBL328107	C1CN(CCN1)	?	?	?

Missing values in the right table

Joiner

- Combines columns from two different tables
 - Top input port: “Left” data table
 - Bottom input port: “Right” data table
- Outputs:
 - Top port: Resulting joined table
 - Middle port: Unmatched rows from the left input table (top input port)
 - Bottom port: Unmatched rows from the right input table (bottom input port)
- By default the two bottom output ports are deactivated



Joiner Configuration – Linking Rows

Values to join on.
Multiple joining columns
are allowed

Select the rows which
should be included in the
joined table

Activate this checkbox to
activate the bottom
output ports

The screenshot shows the 'Joiner Settings' dialog box in KNIME. The 'Join columns' section is highlighted with a yellow box and contains the following elements:

- Match:** Radio buttons for 'all of the following' (selected) and 'any of the following'.
- Top Input ('left' table):** A dropdown menu showing 'StoreID'.
- Bottom Input ('right' table):** A dropdown menu showing 'StoreID'.
- Buttons:** '+' and '-' buttons to add or remove columns.

Below the 'Join columns' section, the 'Compare values in join columns by' section has radio buttons for 'value and type' (selected), 'string representation', and 'making integer types compatible'.

The 'Include in output' section has three checked checkboxes: 'Matching rows', 'Left unmatched rows', and 'Right unmatched rows'. To the right of this section is a Venn diagram labeled 'Full outer join'.

The 'Output options' section has a checked checkbox for 'Route unmatched rows to separate ports' and two unchecked checkboxes: 'Merge join columns' and 'Hilting enabled'.

The 'Row Keys' section has two radio buttons: 'Concatenate original row keys with separator' (selected) and 'Assign new row keys sequentially'.

At the bottom of the dialog are buttons for 'OK', 'Apply', 'Cancel', and a help icon.

Joiner Configuration – Column Selection

Dialog - 0:303 - Joiner

Joiner Settings **Column Selection** Performance Flow Variables Memory Policy

Top Input (left table)

☒ Manual Selection ☐ Wildcard/Regex Selection ☐ Type Selection

Exclude

Filter

No columns in this list

☒ Enforce exclusion

Include

Filter

☒ Enforce inclusion

Bottom Input (right table)

☒ Manual Selection ☐ Wildcard/Regex Selection ☐ Type Selection

Exclude

Filter

☒ Enforce exclusion

Include

Filter

☐ Enforce inclusion

Duplicate column names

☐ Do not execute

☒ Append custom suffix (right)

OK Apply Cancel ?

Columns from top table for joined table

Columns from lower table for joined table

Data Aggregation (GroupBy)

Type	Name	Weight
NSAID	paracetamol	151.17
NSAID	aspirin	180.16
NSAID	ibuprofen	206.29
NSAID	diclofenac	296.15
PPI	omeprazole	345.42
PPI	pantoprazole	383.38
SSRI	fluoxetine	309.33
SSRI	paroxetine	329.37
SSRI	citalopram	324.40
SSRI	sertraline	342.70



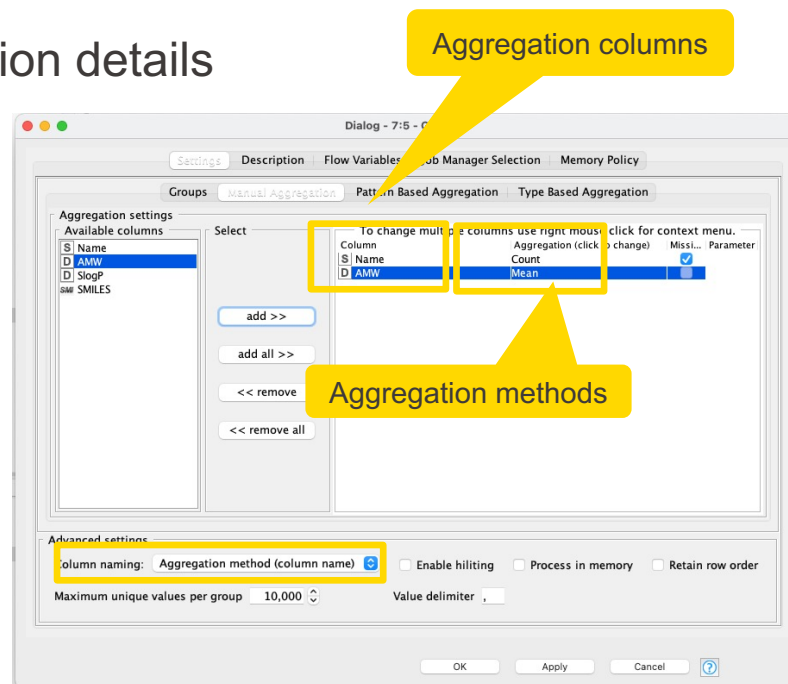
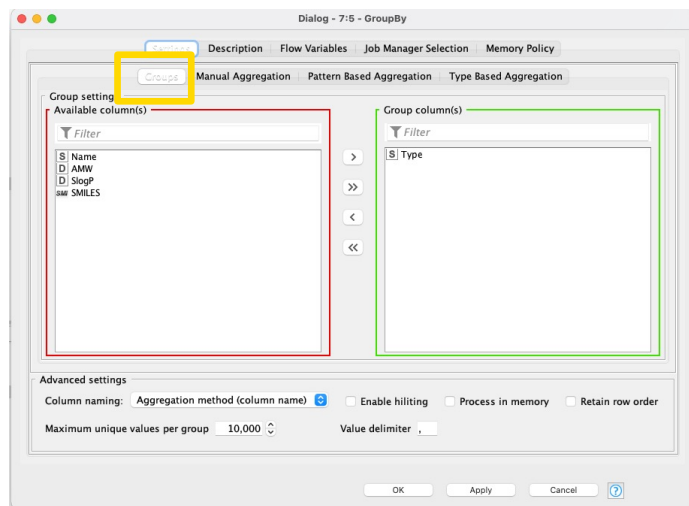
Type	Count(Name)	Mean(Weight)
NSAID	4	208.44
PPI	2	364.40
SSRI	4	326.45

Aggregated on Type (**group**) by
Count (**aggregation method**) and
Mean (**aggregation method**)

GroupBy

Aggregate to summarize data

- First tab provides grouping options
- Second tab provides control over aggregation details



YouTube KNIME TV video: <https://youtu.be/bDwF-TOMtWw>

Data Aggregation (Pivoting)

Type	Name	Safety
NSAIDs	paracetamol	irritant
NSAIDs	aspirin	irritant
NSAIDs	ibuprofen	health hazard
NSAID	diclofenac	acute toxic
PPIs	omeprazole	irritant
PPIs	pantoprazole	irritant
SSRIs	fluoxetine	acute toxic
SSRIs	paroxetine	health hazard
SSRIs	citalopram	health hazard
SSRIs	sertraline	health hazard



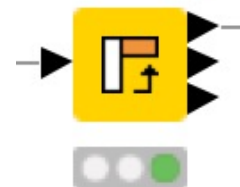
Type	Acute toxic	Health hazard	Irritant
NSAIDs	1	1	2
PPIs	?	?	2
SSRIs	1	3	?

Pivoting Node: **Group** - **Pivot** - **Aggregate**

Pivoting with Two Aggregation Methods

Row ID	S Type	S Safety	S Name
Row0	NSAID	irritant	paracetamol
Row1	NSAID	irritant	aspirin
Row2	NSAID	health hazard	ibuprofen
Row3	NSAID	acute toxic	diclofenac
Row4	PPIs	irritant	omeprazole
Row5	PPIs	irritant	pantoprazole
Row7	SSRIs	acute toxic	fluoxetine
Row8	SSRIs	health hazard	paroxetine
Row9	SSRIs	health hazard	citalopram
Row10	SSRIs	health hazard	sertraline

Pivoting



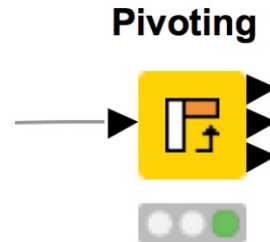
Row ID	S Type	I acute toxic+Count(Name)	I health hazard+Count(Name)	I irritant+Count(Name)
Row0	NSAID	1	1	2
Row1	PPIs	?	?	2
Row2	SSRIs	1	3	?

Pivoting Node: **Group** - **Pivot** - **Aggregate**

Pivoting

Performs pivoting on selected columns for grouping and pivoting

- Values of group columns become unique rows
- Values of the pivot columns become unique columns for each set of column combination together with each aggregation
- Many aggregation methods are provided (similar to GroupBy)



Pivoting

Groups ~ Rows

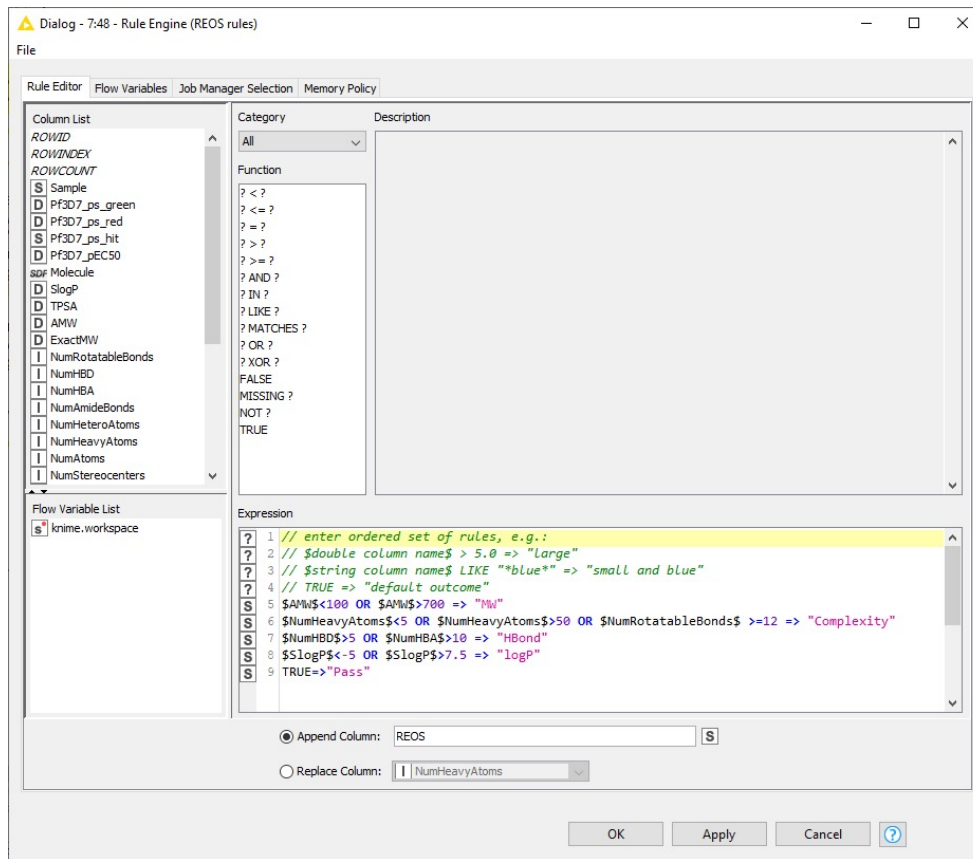
Pivots ~ Columns

Aggregation

Pivot table - 0:35 - Pivoting

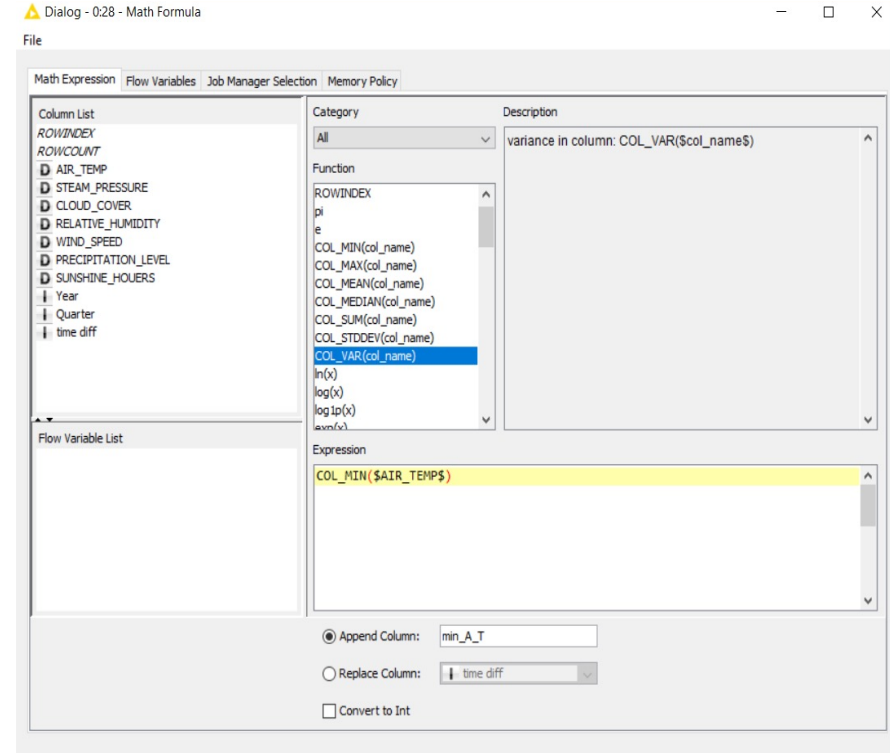
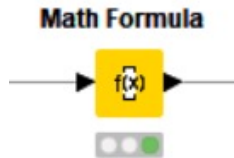
Row ID	Category	Online+Sum(OrderedItems)	Onsite+Sum(OrderedItems)
Row0	Clothing	11823	7604
Row1	Electronics	10754	6624
Row2	Home	7180	5109

Rule Engine



Math Formula

- Row-wise calculations
- Some column-wise statistics
- Lots of mathematical functions
- Double click on function, then select column



Column Expression

- Append or modify an arbitrary number of columns using expressions
- Many different functions are available
- No restriction on number of lines per expression allow to write complex expressions
- Part of the KNIME Labs extension

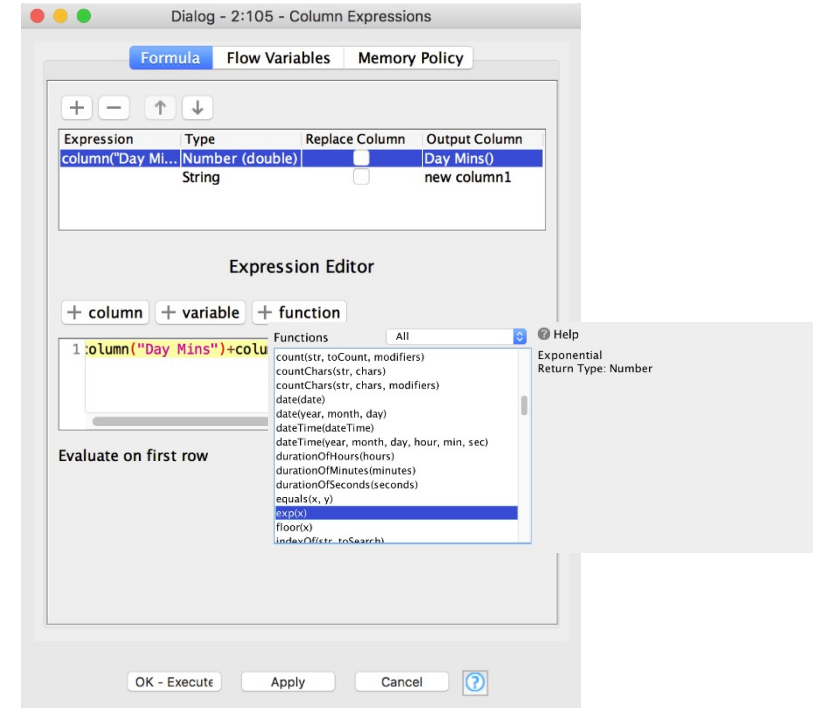
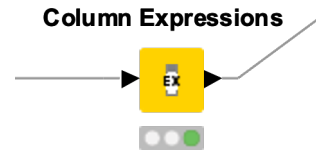
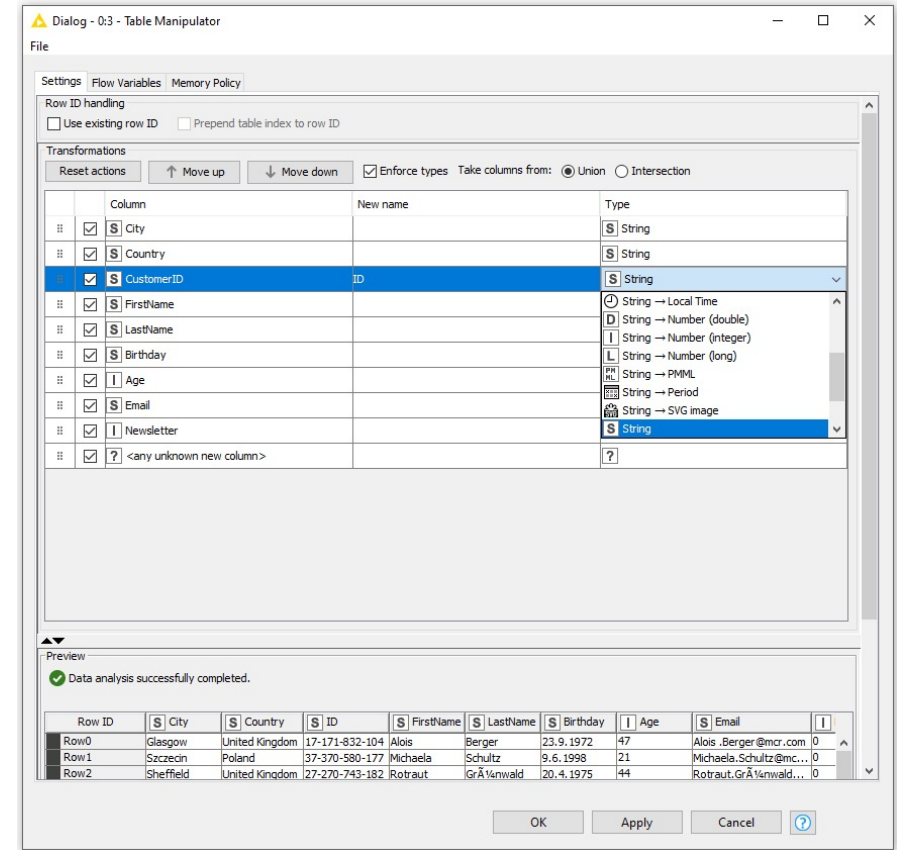
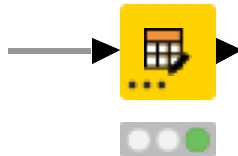


Table Manipulator

Allows for

- Concatenation of multiple files/tables
- Column filtering
- Column sorting
- Column renaming
- Column type mapping

Table Manipulator

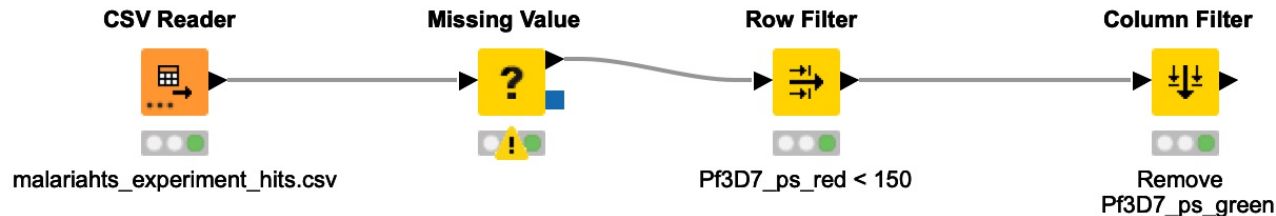


Data Manipulation Exercise

Open Exercise 02. Data Manipulation

Activity I: Filtering

- Remove rows where column **Pf3D7_pEC50** contains missing values
- Filter rows with higher values than 150 in column **Pf3D7_ps_red**
- Remove column **Pf3D7_ps_green** from the result



Data Manipulation Exercise

Still in Exercise: *02. Data Manipulation*

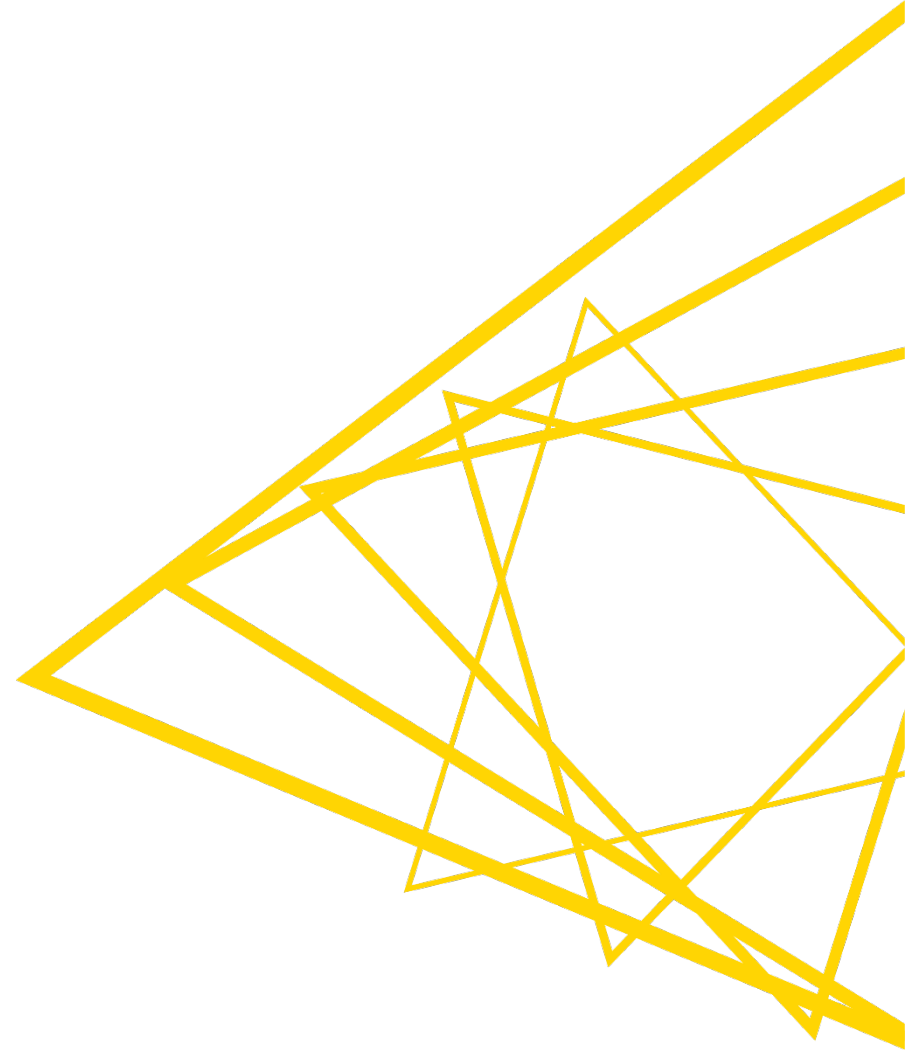
Activity II: Data Manipulation & Aggregation

- Concatenate, Join and Manipulate the data according to the instructions in the annotations

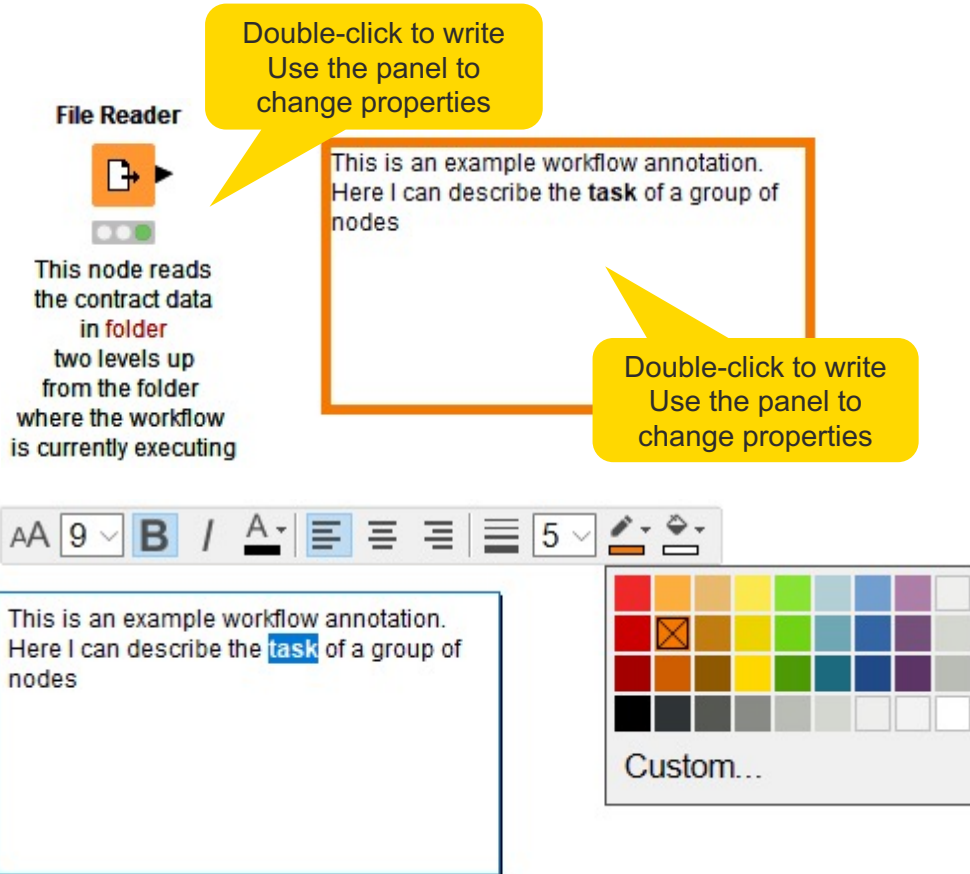
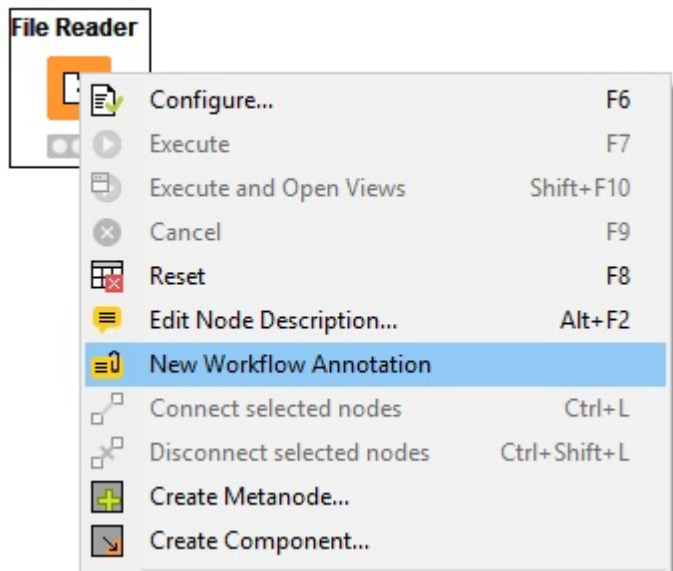
Activity III: Data Manipulation (Optional)

- Use the Rule Engine node to add the following tags in a new column named REOS (according to rules in the exercise)
- Filter columns

Workflow Organization and Documentation



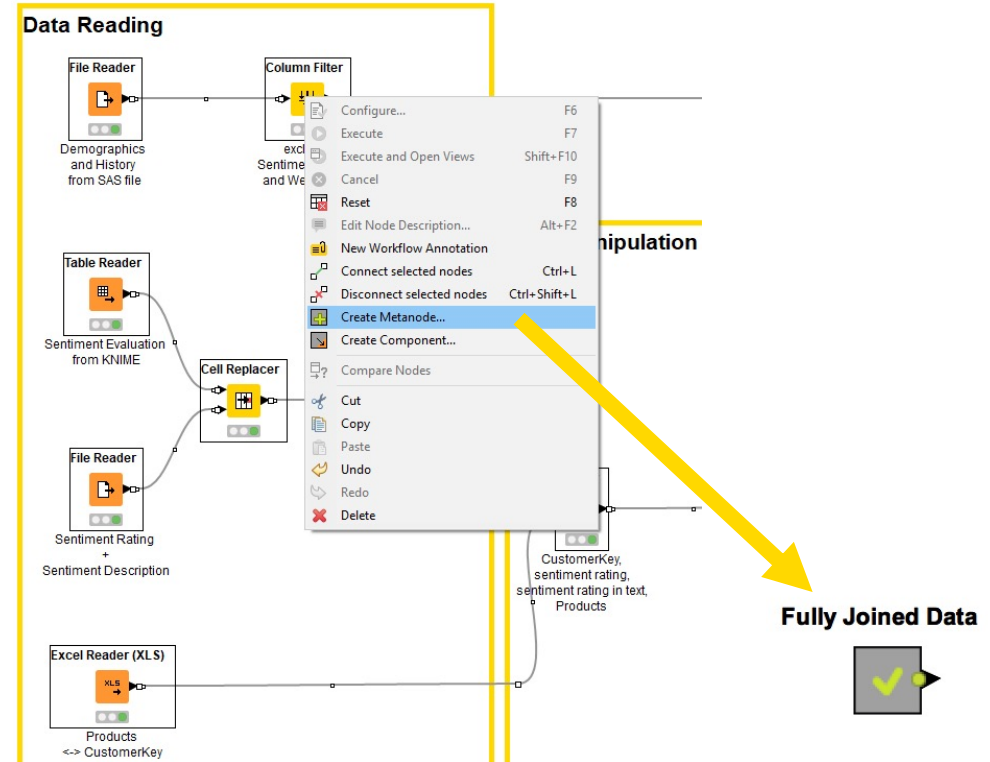
Comments & Annotations



YouTube KNIME TV Channel:
https://youtu.be/AHURYB_O8sA

Workflow Organisation – Good Practices

- Workflow annotations
- Node labels
- Metanodes
 - Right click -> Create Metanode...
 - Organize workflow by task
 - Hide complexity & improve readability



Workflow Organisation – Components

- Component encapsulates a reusable functionality as a KNIME workflow
- Components can be configured as any KNIME nodes
- Access and share components on the KNIME Hub

The screenshot shows the KNIME Hub search results for 'column filter'. The search bar at the top contains 'column filter' and shows 36 results. The 'Components' tab is selected and highlighted with a yellow circle. Below the search bar, two components are listed:

- Interactive Column Filter**: This Component creates an interactive view to filter and select columns for your model. (paolotamag > Public > Interactive Column Filter)
- Column Filter (by Index)**: This component allows to filter columns by their index (i.e. position). An example can be: select the second and the last columns (e.g. with '2,-1'). Configuration takes a string of column positions. ... (knime > Examples > 00_Components > Data Manipulation > Column Filter (by Index))

A yellow callout bubble points to the 'Column Filter (by Index)' component with the text: "Drag and drop from the KNIME Hub to your workflow".

Below the callout, a workflow diagram for the 'Column Filter (by Index)' component is shown. The workflow starts with a 'Component Input' node, followed by a 'Select column subset' node, then a 'Reference Column Filter' node, and finally a 'Component Output' node. The 'Select column subset' node is configured with a 'String Configuration' node, a 'Variable to Table Row' node, and a 'Cell Splitter' node. The 'Reference Column Filter' node is configured with an 'Error treatment' node. The workflow is titled 'Get a table with column indices' and 'Handle situations, when something went wrong'.

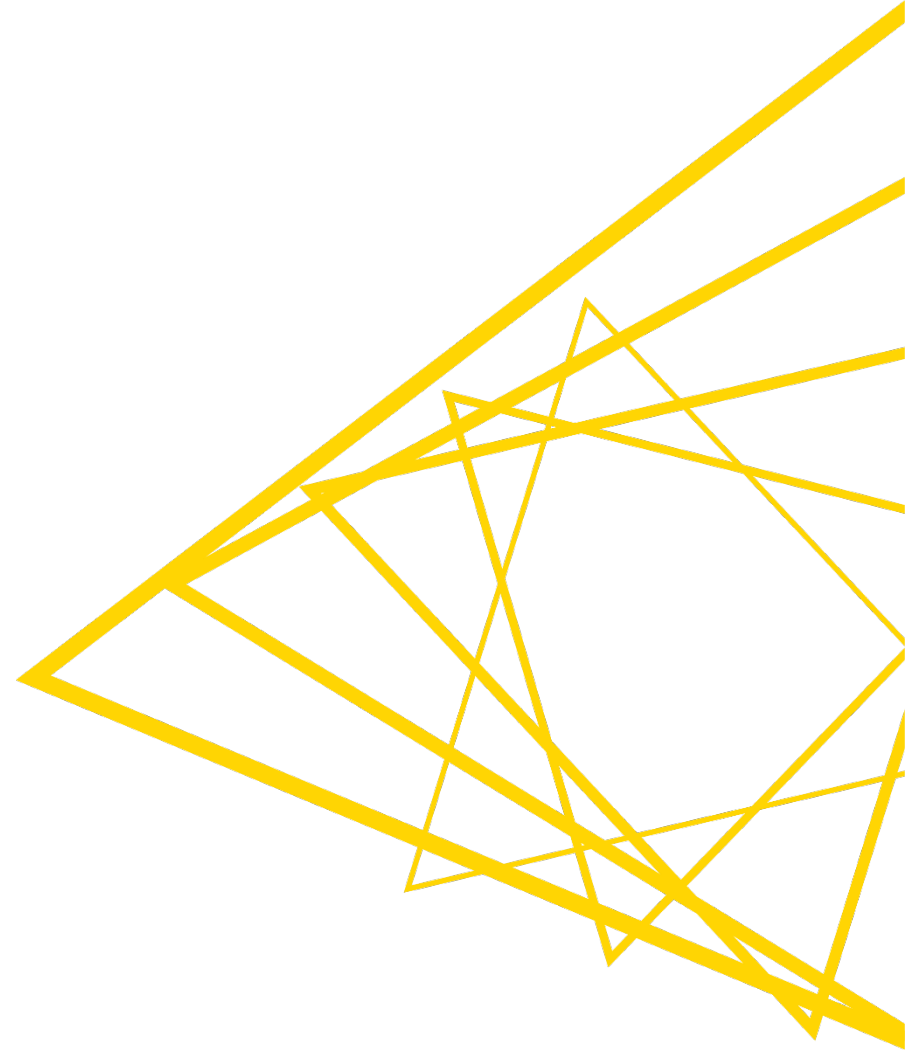
Input is a string of comma-separated column indices.
Negative numbers mean counting from the back, e.g. -1 is the last column.
Zeros is not a valid input!

Examples of the configuration:
1,3,5: 1st, 3rd and 5th columns
1,-1: 1st and last columns

The component requires the following extensions:
- KNIME Expressions (<https://hub.knime.com/knime/extensions/org.knime.features.expressions/latest>)

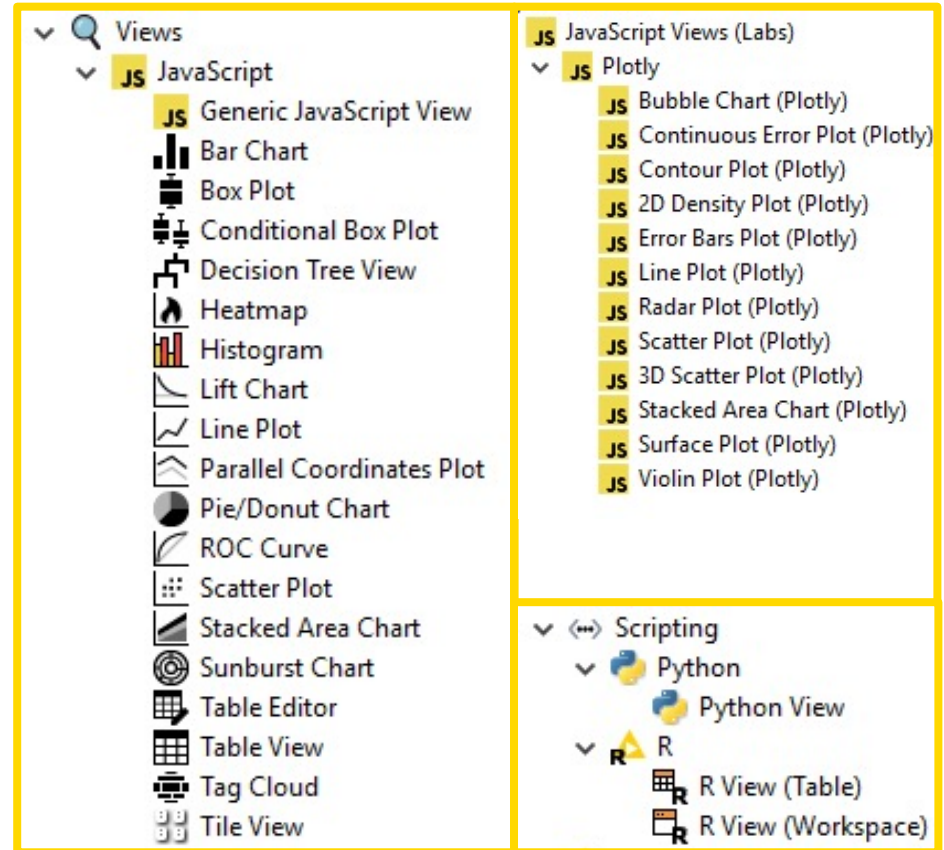
Data Visualization

Charts and Tables



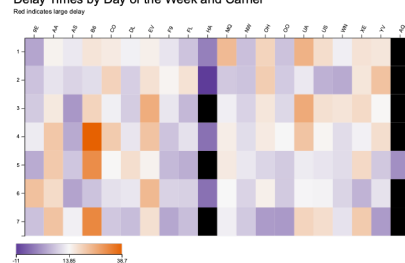
Data Visualization

- Large selection of easy to use visualization nodes
 - Web-based and interactive
 - Dedicated nodes, no scripting required
- Plotly nodes
 - Similar but integrated from an external library
- R and Python View nodes for highly customizable graphics
 - Require scripting

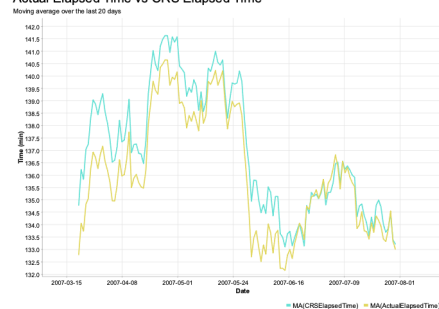


Visualization

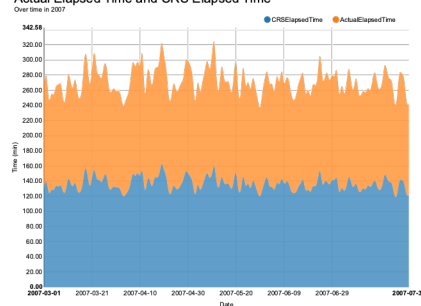
Delay Times by Day of the Week and Carrier



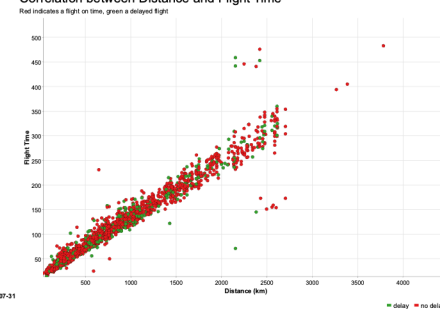
Actual Elapsed Time vs CRS Elapsed Time



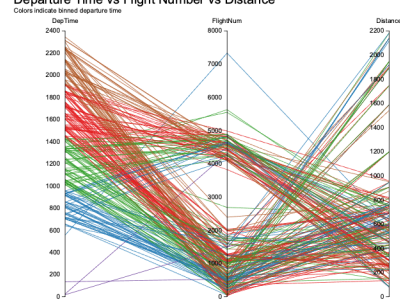
Actual Elapsed Time and CRS Elapsed Time



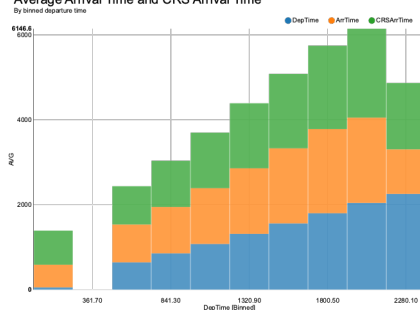
Correlation between Distance and Flight Time



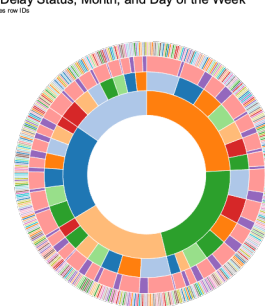
Departure Time vs Flight Number vs Distance



Average Arrival Time and CRS Arrival Time



Flights by Delay Status, Month, and Day of the Week



Distribution of Elapsed Time

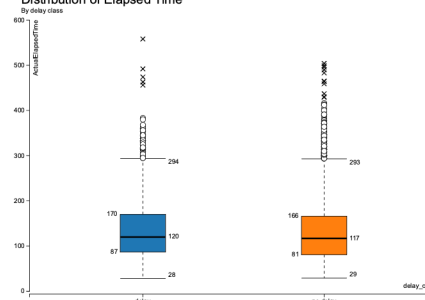
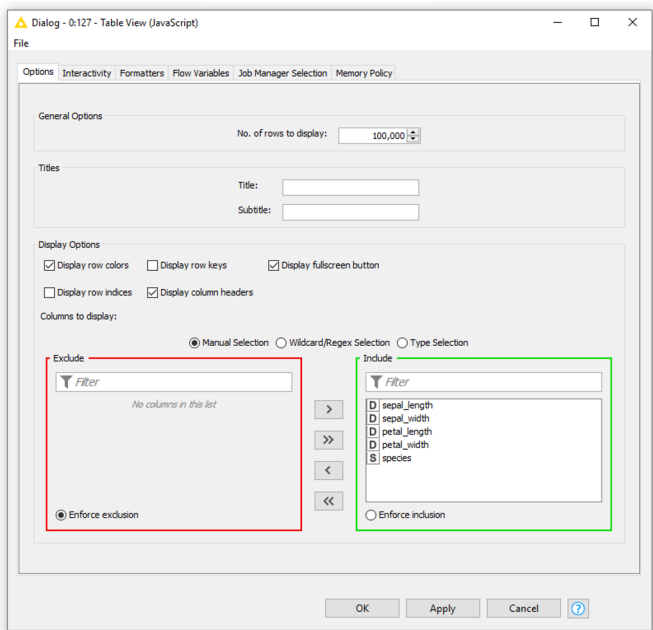
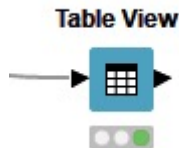


Table View



JavaScript Table View

Show 10 entries Search:

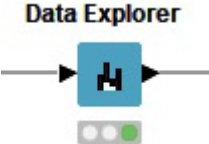
		sepal_length	sepal_width	petal_length	petal_width	species
		5.1	3.5	1.4	0.2	setosa
		4.9	3	1.4	0.2	setosa
		4.7	3.2	1.3	0.2	setosa
		4.6	3.1	1.5	0.2	setosa
		5	3.6	1.4	0.2	setosa
		5.4	3.9	1.7	0.4	setosa
		4.6	3.4	1.4	0.3	setosa
		5	3.4	1.5	0.2	setosa
		4.4	2.9	1.4	0.2	setosa
		4.9	3.1	1.5	0.1	setosa

Showing 1 to 10 of 150 entries

Previous 1 2 3 4 5 ... 15 Next

Reset Apply Close

Data Explorer



Data Explorer View

Numeric

Nominal

Data Preview

Search:

Column	Exclude Column	Minimum	Maximum	Mean	Standard Deviation	Variance	Skewness	Kurtosis	Overall Sum	No. zeros	No. missings	No. NaN	No. +∞	No. -∞	Histogram
AMW	<input type="checkbox"/>	163.184	700.812	390.278	69.540	4835.749	-0.215	0.364	596344.599	0	0	0	0	0	
NumRotatableBonds	<input type="checkbox"/>	0	16	5.228	2.242	5.027	0.558	0.926	7988	6	0	0	0	0	
NumHBD	<input type="checkbox"/>	0	4	1.302	0.895	0.800	0.325	-0.386	1990	285	0	0	0	0	
NumHBA	<input type="checkbox"/>	0	16	4.694	1.721	2.963	0.198	0.880	7172	3	0	0	0	0	
NumHeavyAtoms	<input type="checkbox"/>	12	52	27.425	4.977	24.772	0.021	0.549	41905	0	0	0	0	0	
FractionCSP3	<input type="checkbox"/>	0	1	0.261	0.148	0.022	0.837	1.786	398.142	45	0	0	0	0	

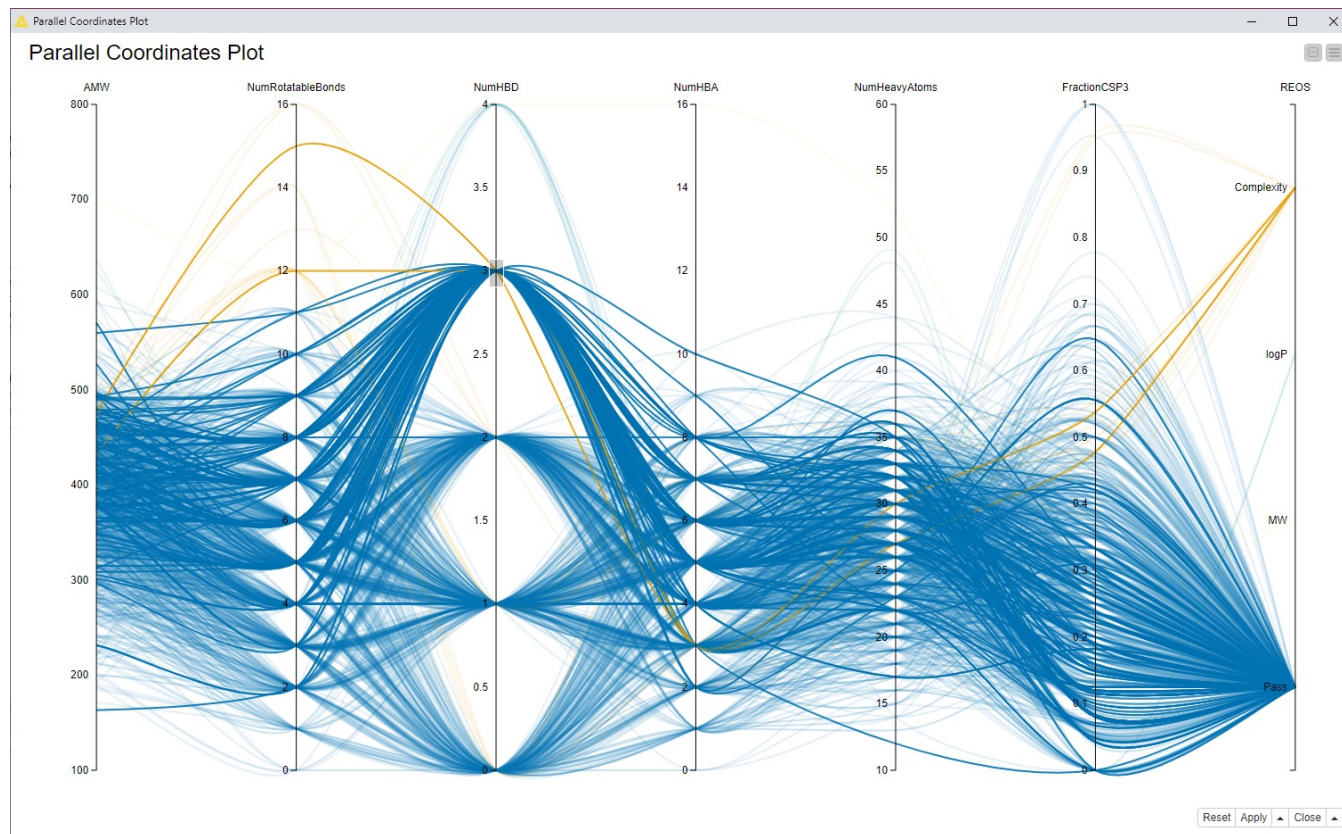
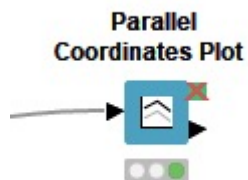
Showing 1 to 6 of 6 entries

Reset

Apply

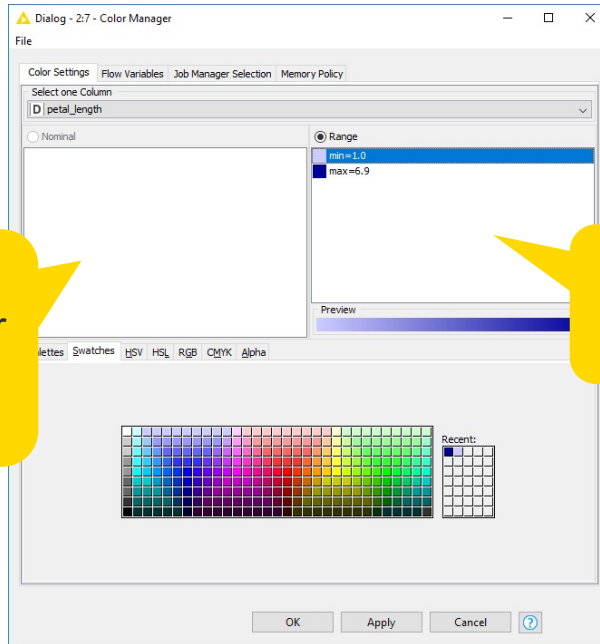
Close

Parallel Coordinates Plot



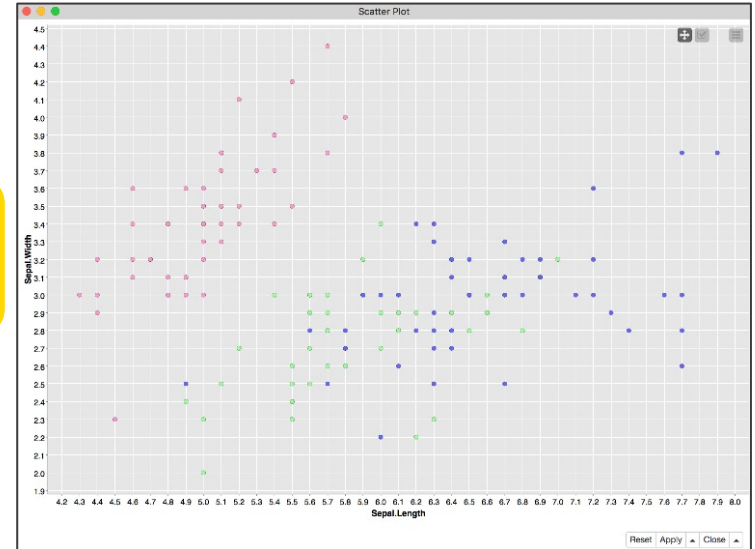
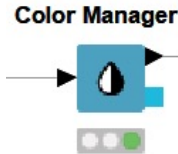
Color Manager

- Color by nominal or continuous values
- Sync colors between views using the color model port and Color Appender node



Discrete colors for nominal values

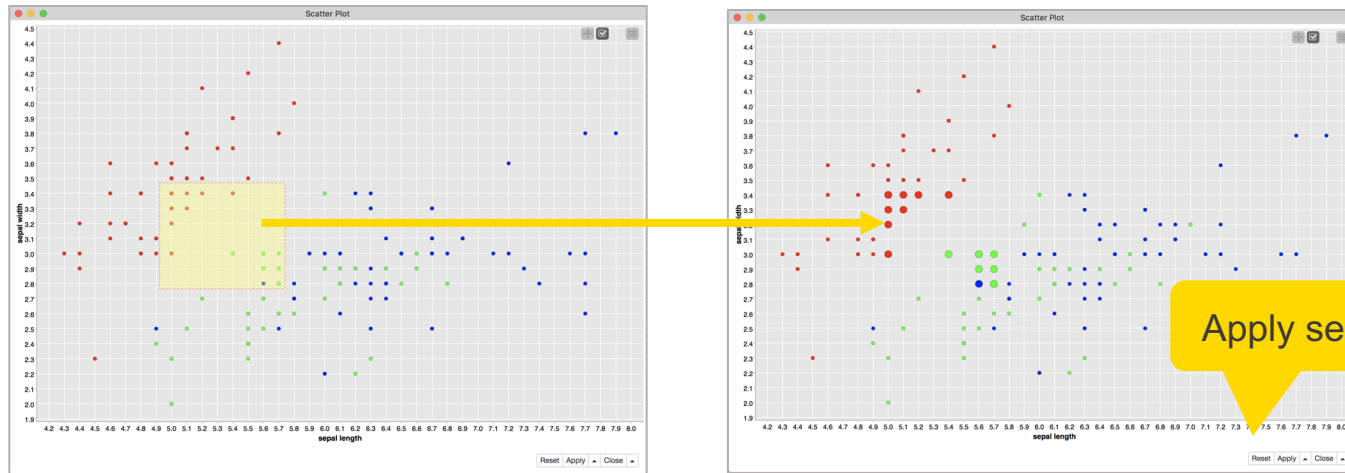
Color range for numerical values



Selection & Filtering in JavaScript Views

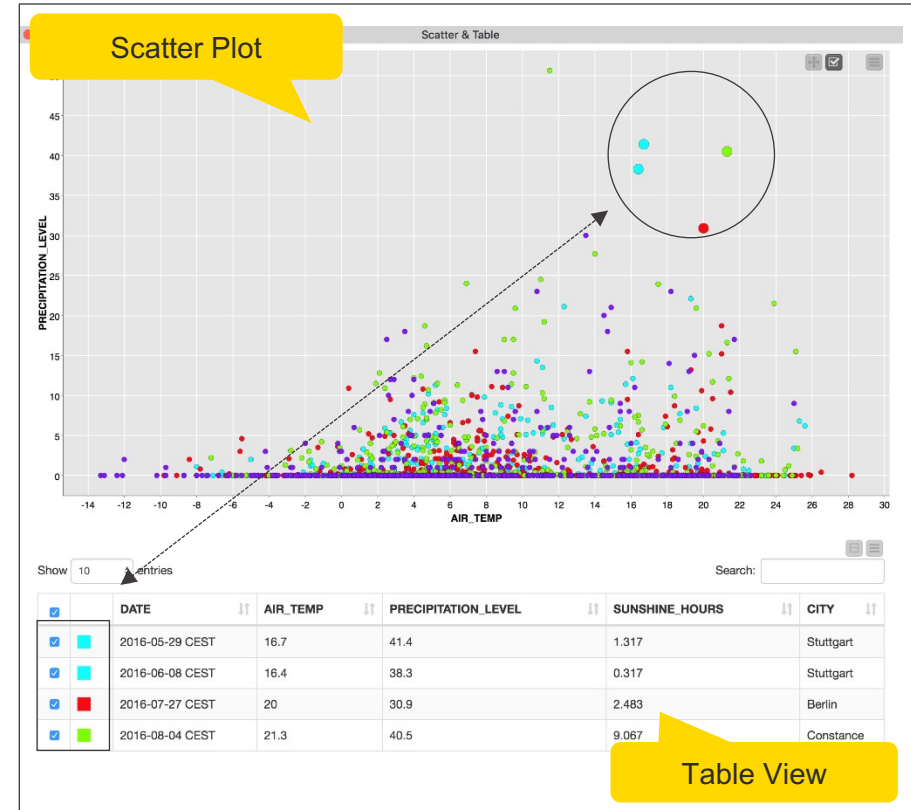
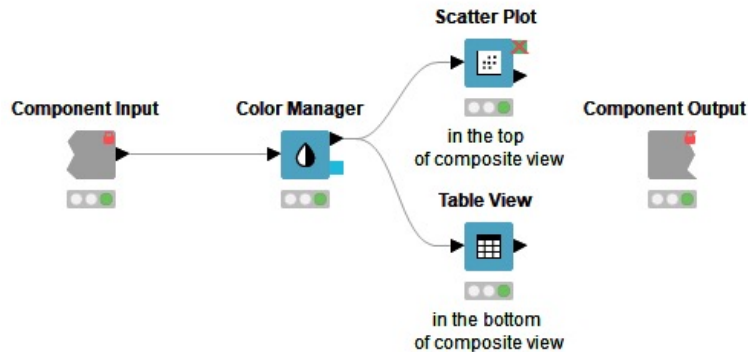
Interactivity allows you to select data points in views

- Selection is propagated to other views.
- Highlight selected rows or filter them
- Click “Apply” to add column to data that indicates selection (true/false) for use in downstream nodes

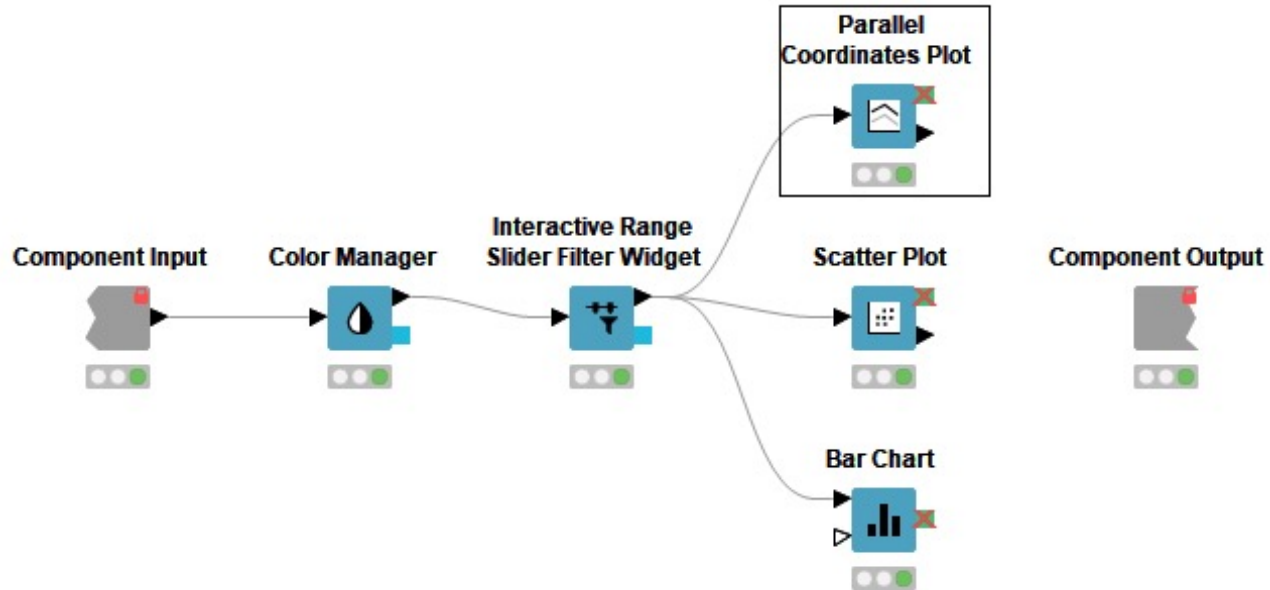


Components – Combined Views

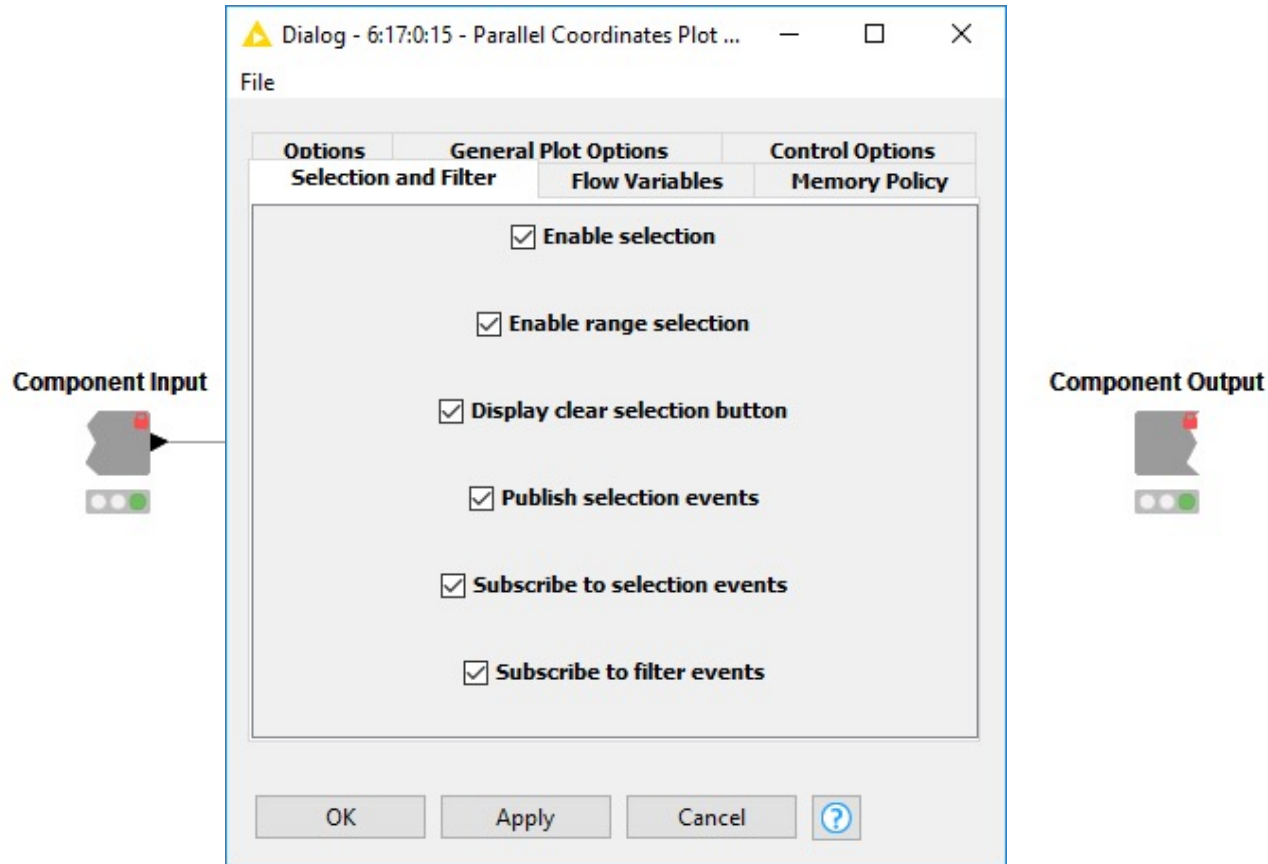
- Multiple JavaScript View nodes can be combined in Components
- Selections are transmitted to all other views
- Also for use on the KNIME WebPortal



Interactivity across Charts: Selection and Filter Events



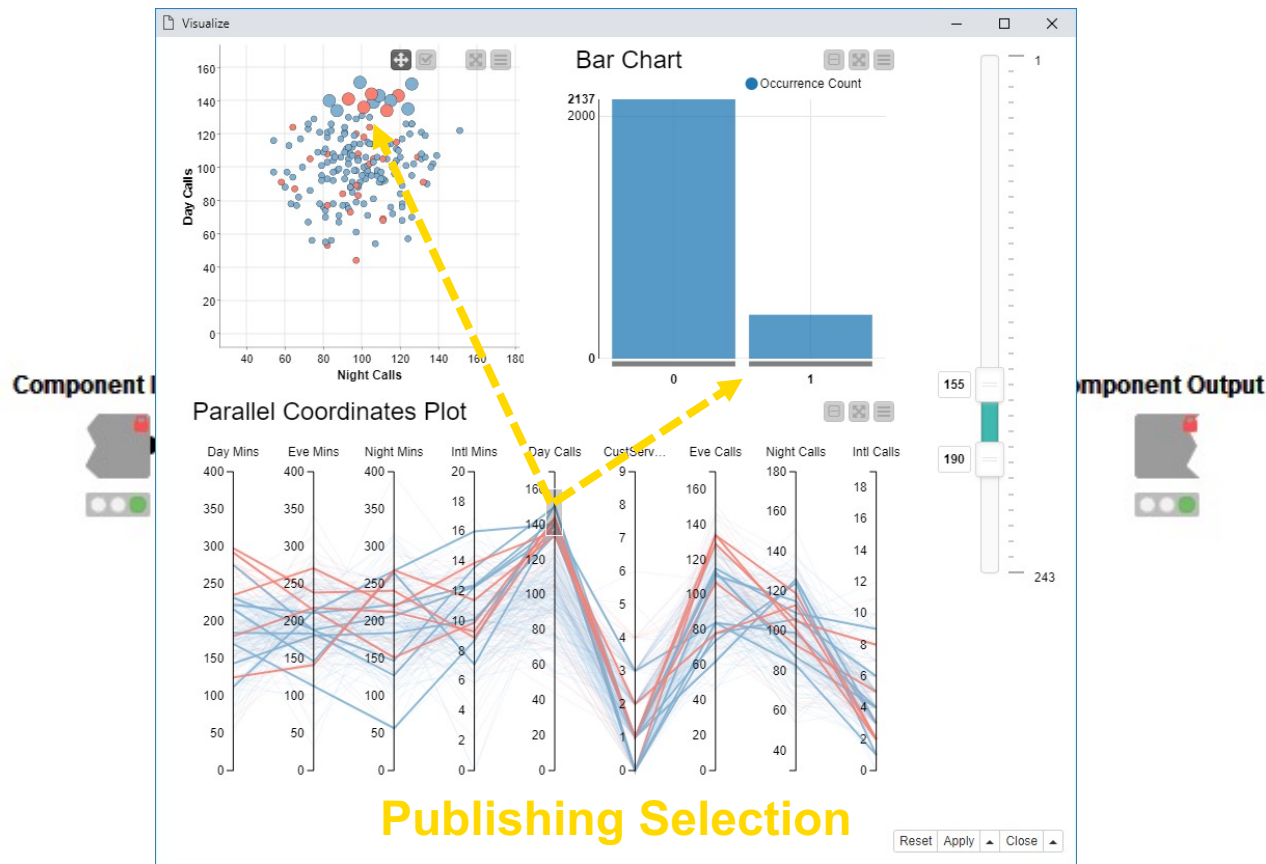
Interactivity across Charts: Selection and Filter Events



Interactivity across Charts: Selection and Filter Events



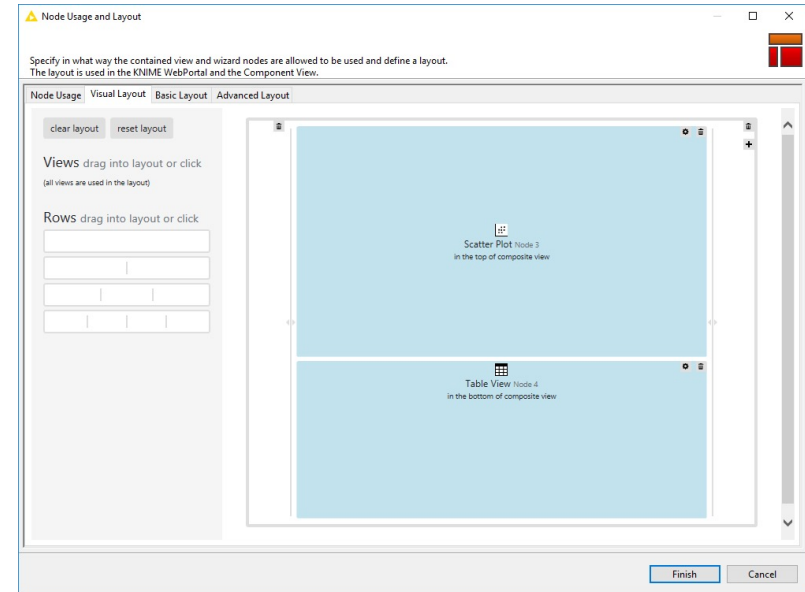
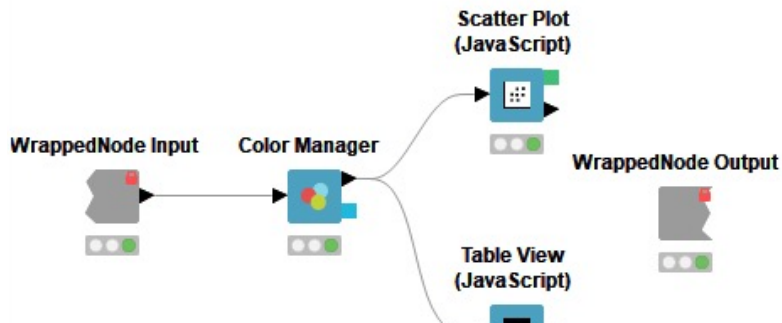
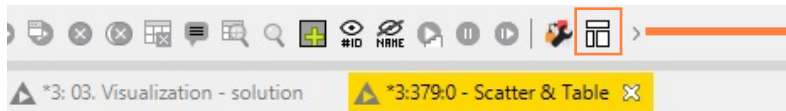
Interactivity across Charts: Selection and Filter Events



Publishing Selection

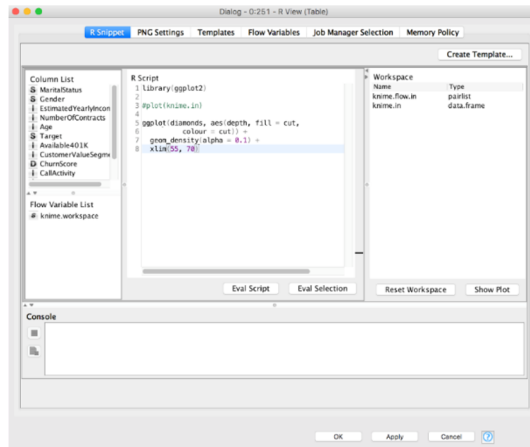
Configure Content and Views Layout

- Click layout button when inside Component to assign views to rows and columns
 - Add views and rows via *drag&drop*
 - Add columns using + buttons

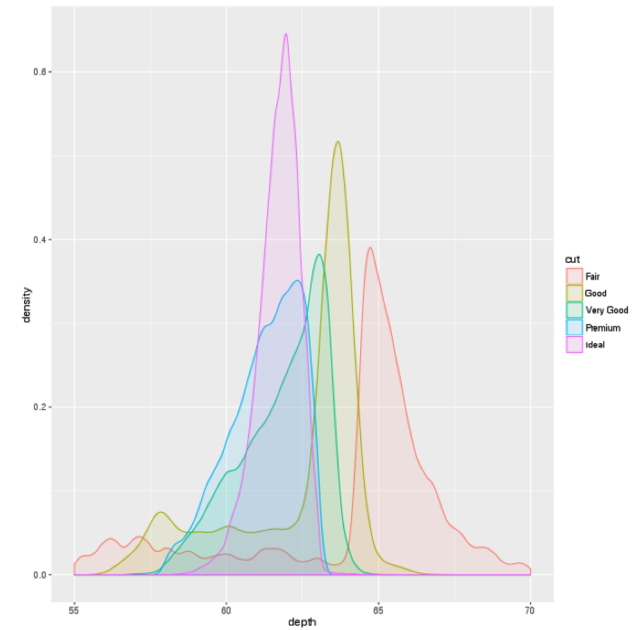


Script-based View Nodes

- R View nodes for greater customizability
 - Use your favorite libraries, e.g. ggplot2
- If you prefer Python: Python View node
- For JS developers: Generic JavaScript View

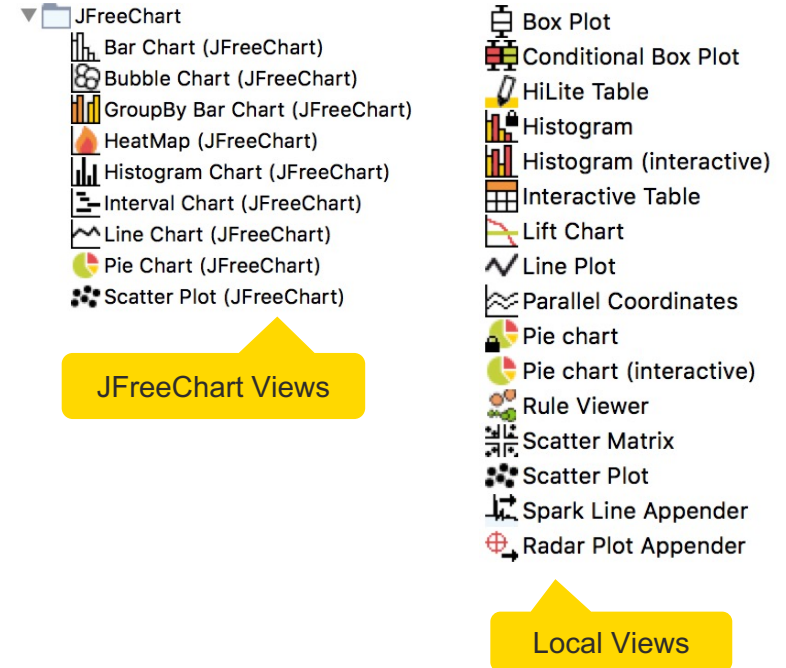


R View (Table)



Legacy View Nodes: JFreeChart & KNIME Views

- KNIME provides three types of visualizations
 - **JavaScript Views**
 - JFreeChart Views
 - Local Views
- Active development only for JavaScript Views -> use those!
- JFreeChart and Local Views still useful when visualizing locally



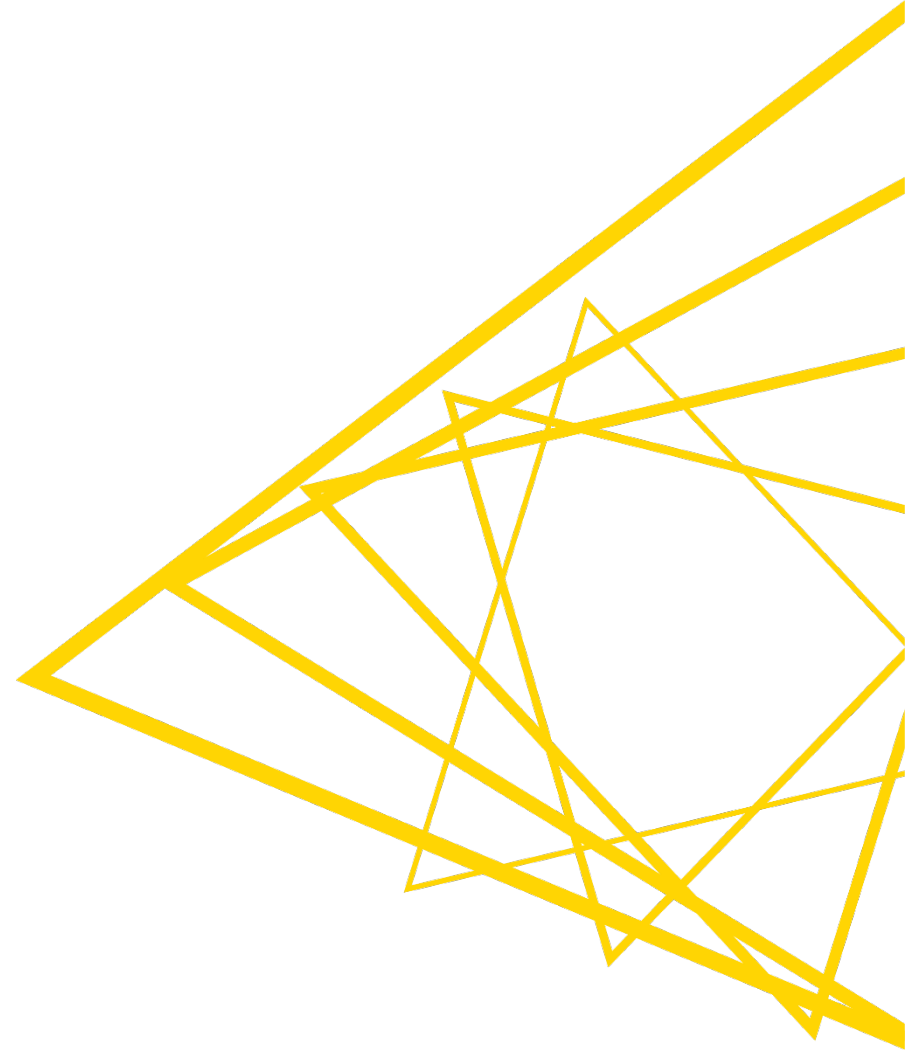
Visualization Exercise

Open Exercise: 03. Visualization

Activity: Visualization

- Use the interactive view of the Data Explorer node to inspect the data.
 - Connect a Parallel Coordinates Plot node and adjust the configuration to display the REOS category and other two numeric columns of interest.
 - Display the other columns of interest in a Table View node and add a title to the table
 - Count the number of molecules in each REOS category using a GroupBy node and display the result in a second Table View node.
 - Create a component containing the newly added nodes (select desired nodes-> right click-> create component)
 - Look into the component (Ctrl + doubleclick) and adjust the layout. Make sure that selection between the views works (hint: enable hiliting in the GroupBy node)
-
- *Optional:*
 - Add a Color Manager before the Table View node and Color the rows using the REOS category.
 - Add a Range Slider Filter Definition node before the Table View node and configure it to use FractionCSP3 column.

Image Processing



KNIME Image Processing

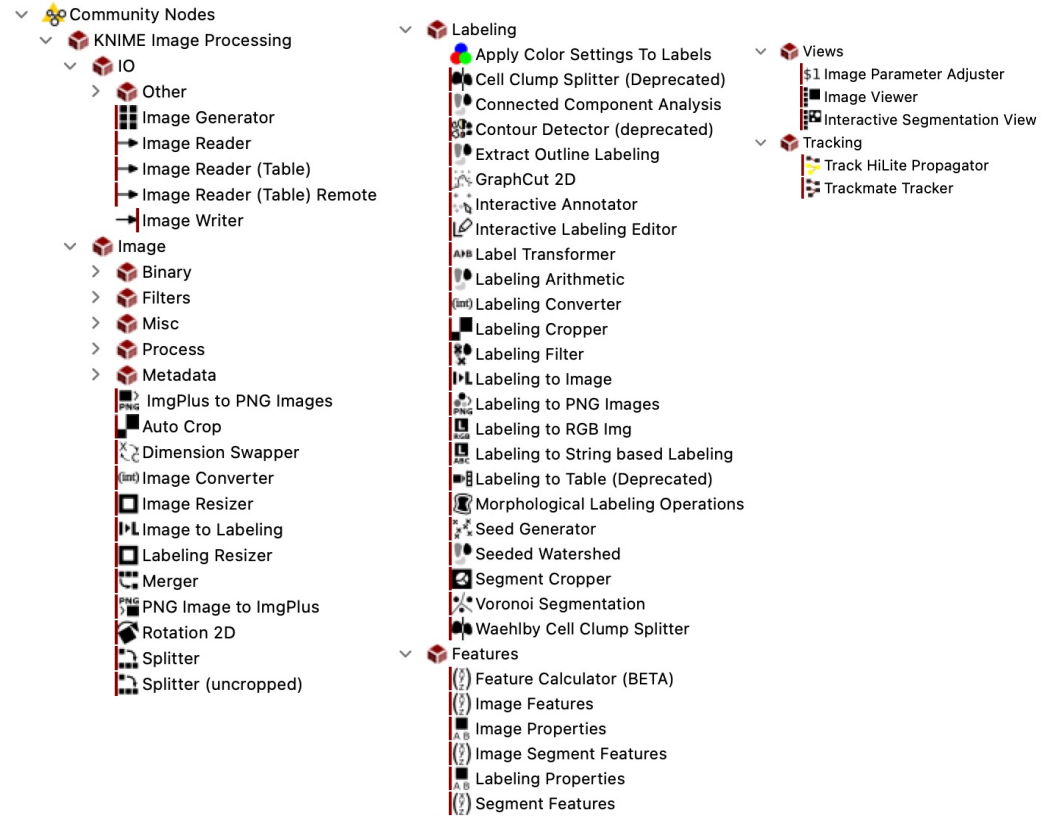
■ KNIME Image Processing Extension

- these nodes operate on multi-dimensional image data (e.g. videos, 3D images, multi-channel images or even a combination of them)

More infos:

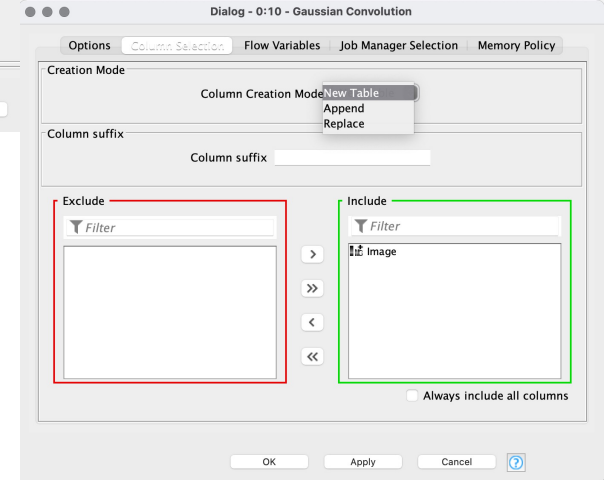
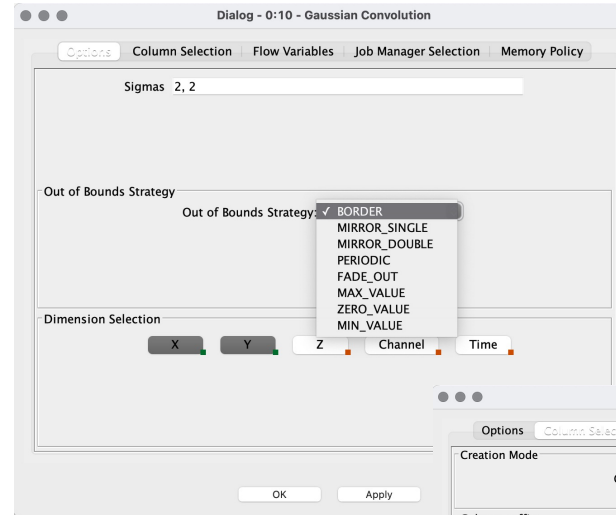
<https://www.knime.com/community/image-processing>

Video: [Introduction to KNIME for Image Processing](#)



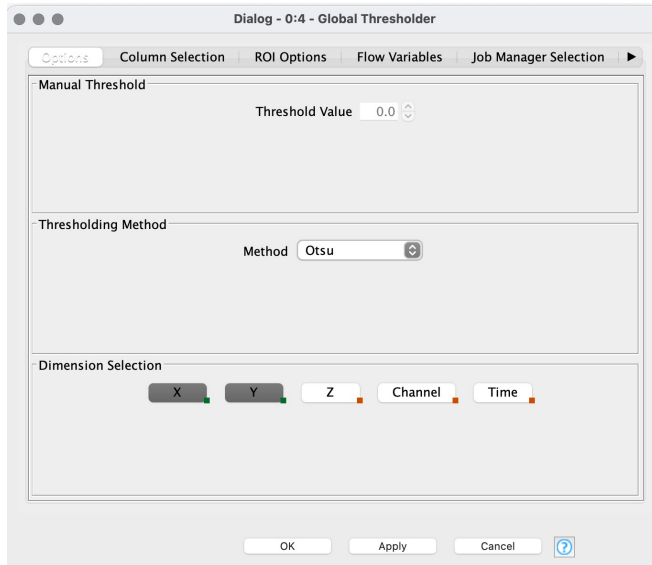
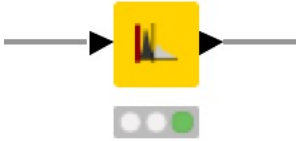
Gaussian Convolution

- Applies a Gaussian blur to reduce image noise and reduce detail
 - Sigma: standard deviation used for each dimension
 - Out of Bounds Strategy: defines how to handle pixels which lie outside of an image (for convolutions)
 - Dimension Selection



Global Thresholder

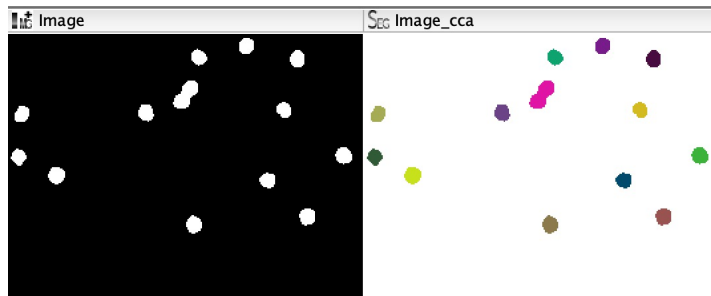
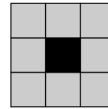
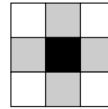
Global Thresholder



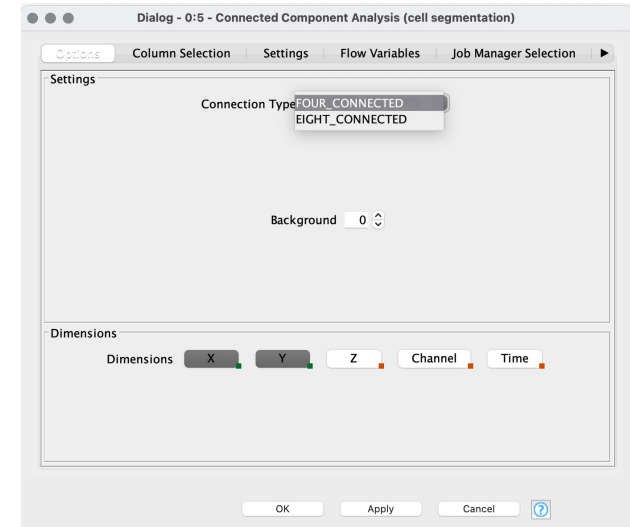
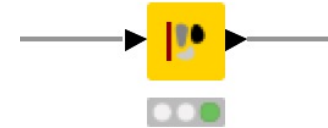
- The selected threshold algorithm determines a threshold
 - any pixel above or below the threshold is marked as “1” or “0”, respectively
- Different Threshold methods are available, e.g.
 - Otsu: threshold that minimizes the intra-class variance, defined as a weighted sum of variances of the two classes

Connected Component Analysis

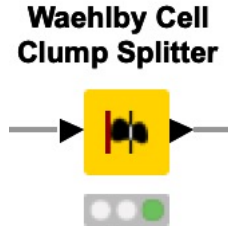
- Identifies connected components in an image and assigns a label
 - FOUR_CONNECTED:** These pixels are neighbours to every pixel that touches one of their edges. These pixels are connected horizontally and vertically.
 - EIGHT_CONNECTED:** These pixels are neighbours to every pixel that touches one of their edges or corners.



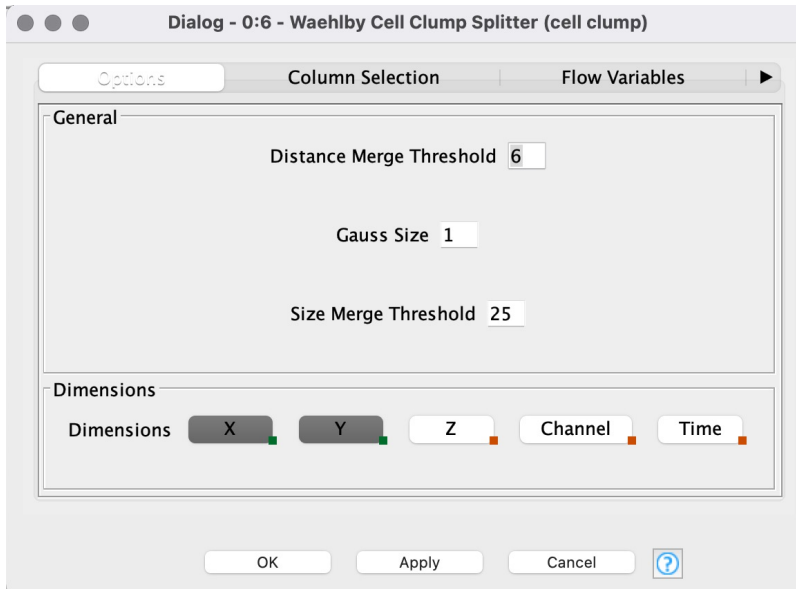
Connected Component Analysis



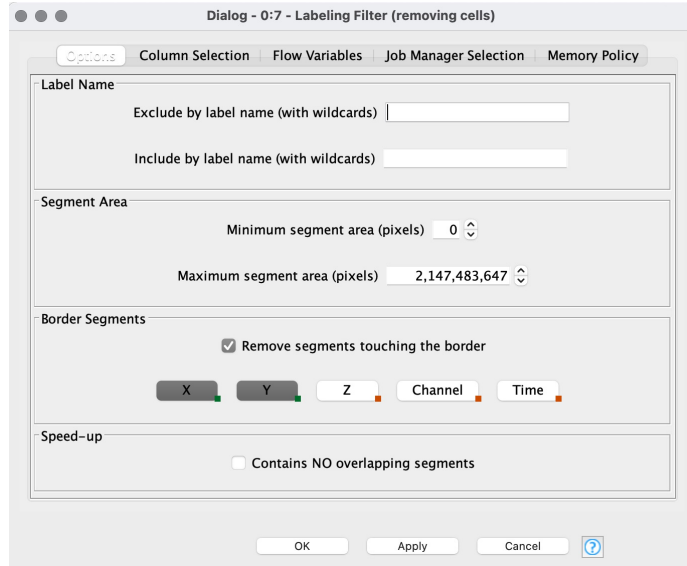
Waehlby Cell Clump Splitter



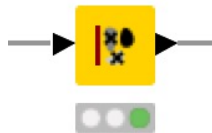
- Waehlby Cell Clump Splitter performs object split and merge by shape on a labeling.
- Distance Merge Threshold (Objects closer than this distance threshold will be considered candidates for merging)
- Gauss Size (size of gaussian convolution for pre-processing step of split and merge method)
- Size Merge Threshold (Objects which contain less pixels than threshold will be merged regardless of their shape)



Labeling Filter



Labeling Filter



- Filters the segments of a labeling according to some filter criteria like
 - segment label,
 - segment area, or
 - whether the segment is touching the labeling border

Image Segment Features

- Calculate features of an image under consideration of a given labeling.
 - Tamura features: Granularity, Maximal directionality, Skewness,...
 - Segment Geometry: size, perimeter, centroid, convexity, ...
 - First order statistics: min, max, mean, sum, ...
 - Haralick: Statistical features based on gray-level co-occurrence matrix

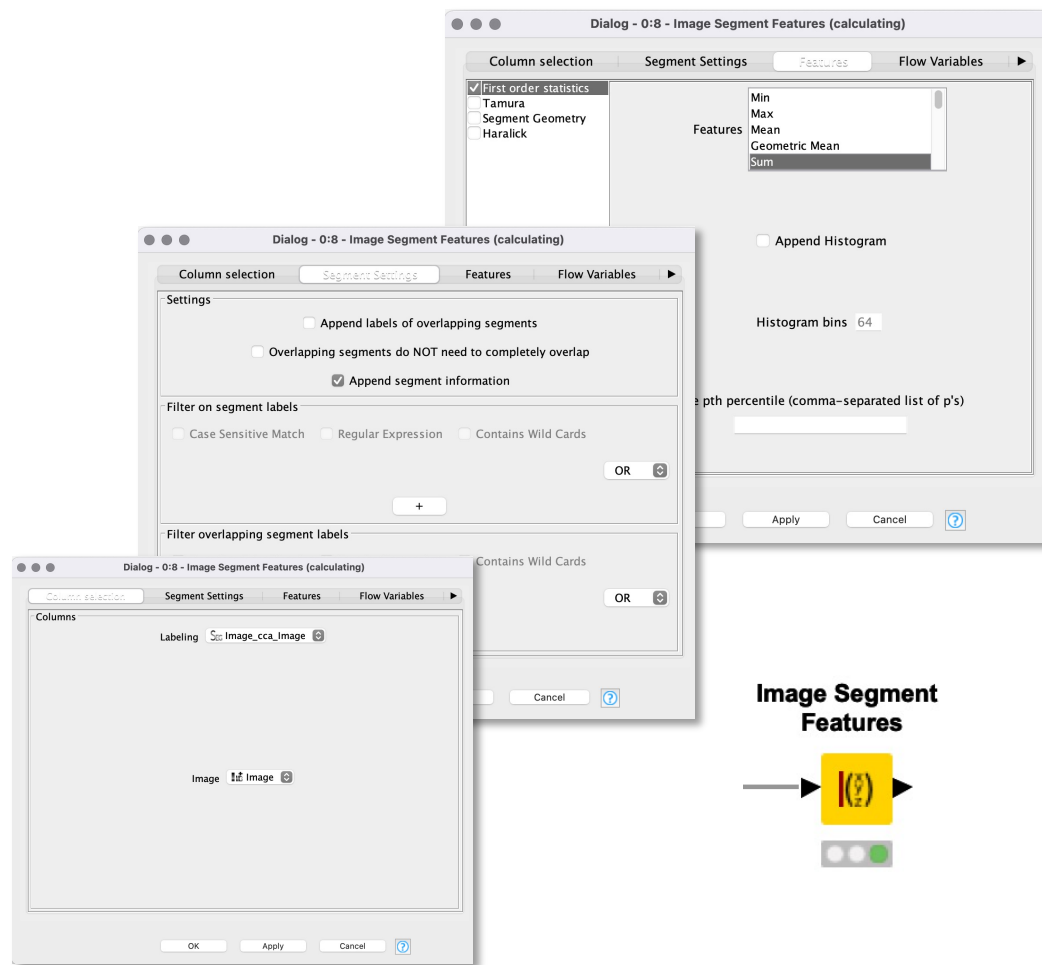
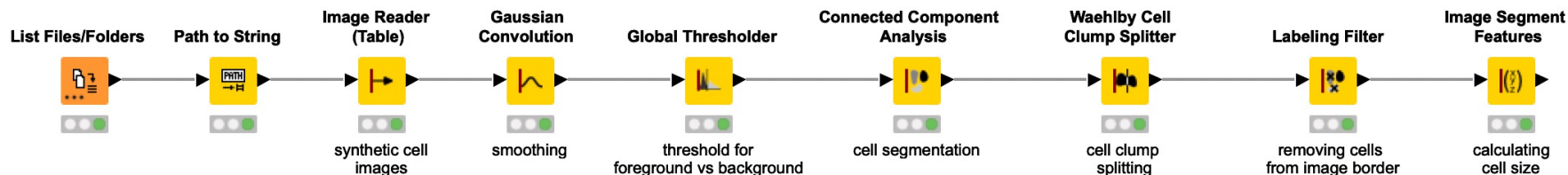


Image Processing Exercise

Open Exercise: 04. Image Processing

Activity: Image Processing of synthetic cells

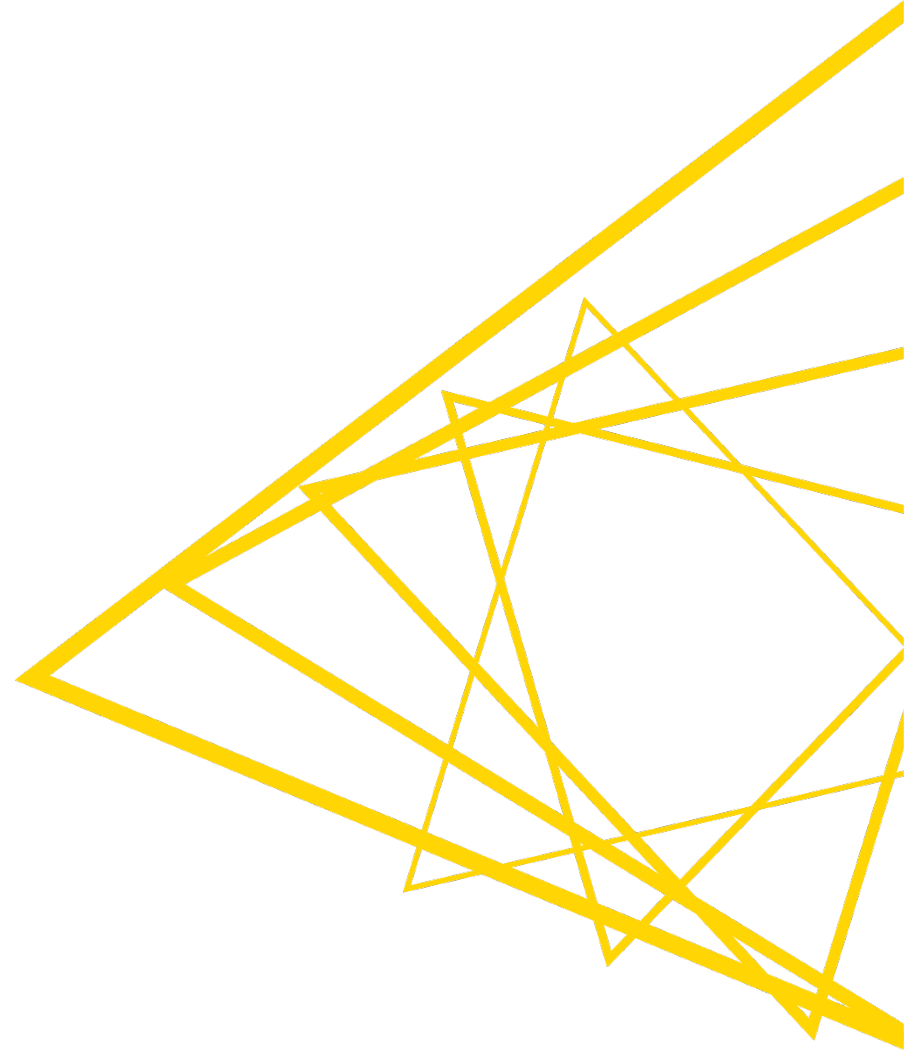
- Calculate the size of each cell using
 - Gaussian Convolution node to smooth your image and reduce noise
 - Global Thresholder node to determine foreground from background
 - Connected Component Analysis node to segment cells and apply labels
 - Waehlby Cell Clump Splitter node to segment cell clumps
 - Labeling Filter node to remove cells on image border
 - Image Segmentation Feature node to calculate the size of each cell



Broad Bioimage Benchmark Collection data set: <https://bbbc.broadinstitute.org/bbbc/BBBC005>

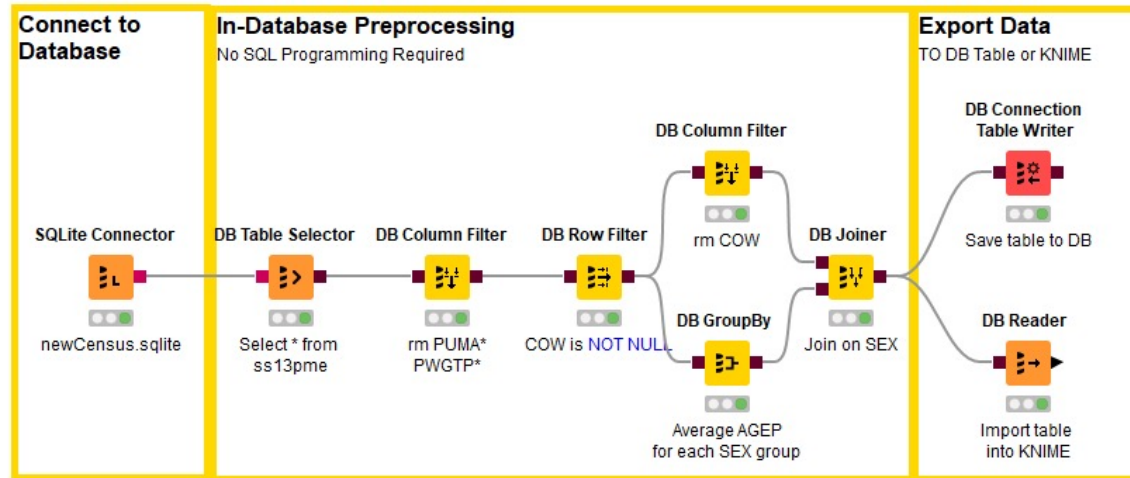
Database

Access, Read, Manipulate, Write

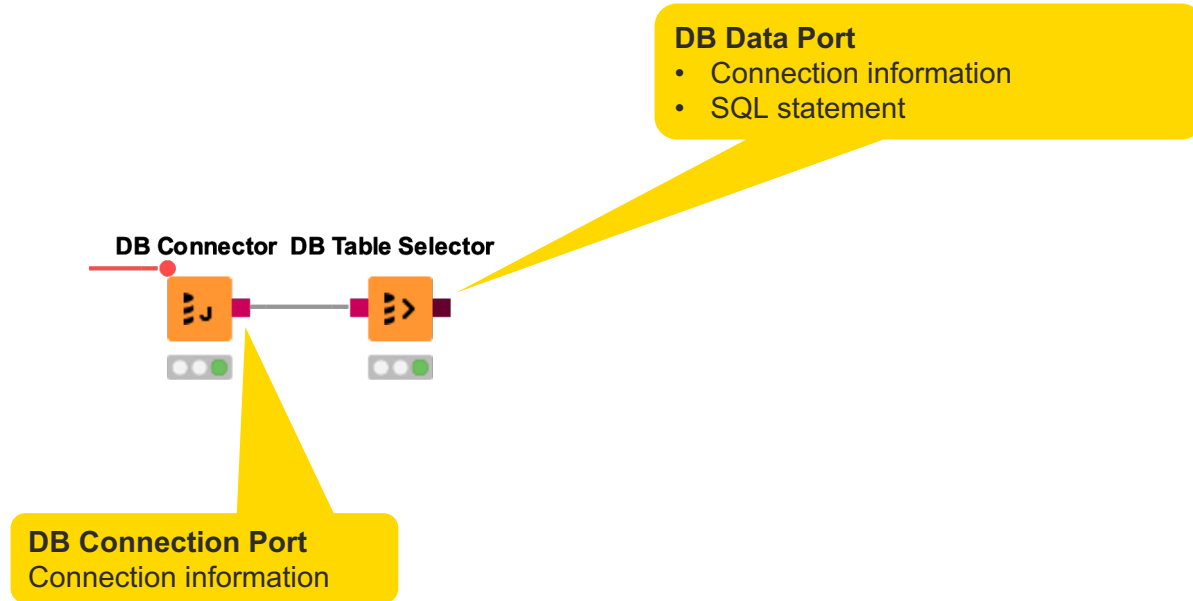


Database Extension

- Visually assemble complex SQL statements (no SQL coding needed)
- Connect to all JDBC-compliant databases
- Harness the power of your database within KNIME
- Complete rewrite in KNIME Analytics Platform 4.0

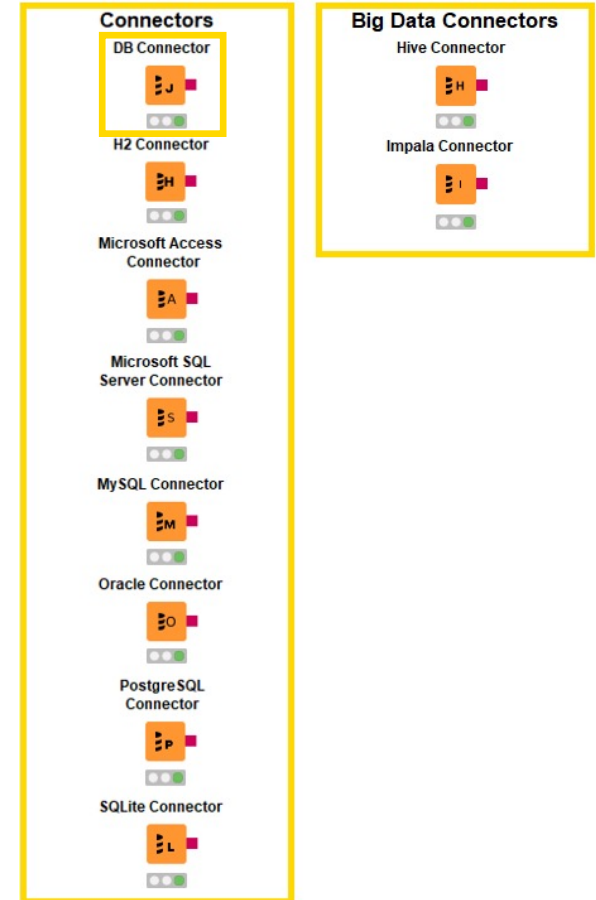


Database Port Types



Database Connectors

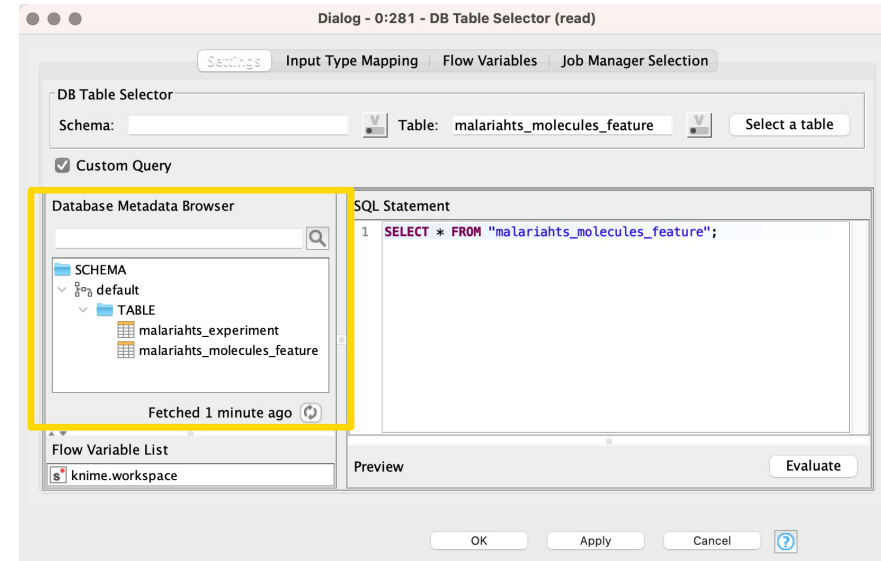
- Dedicated nodes to connect to specific Databases
 - Necessary JDBC driver included
 - Easy to use
 - Import DB specific behavior/capability
- Hive, Impala connectors part of the KNIME Big Data Connectors extension
- General DB Connector
 - Can connect to any JDBC source
 - Register new JDBC driver via
File -> Preferences -> KNIME -> Databases



DB Table Selector

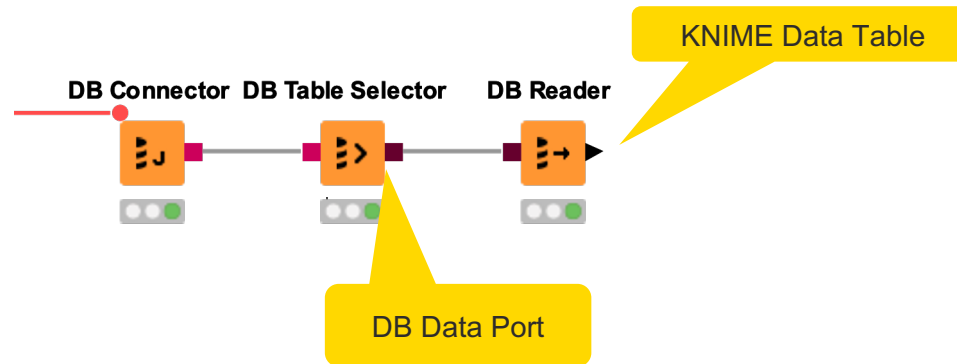
- Takes connection information and constructs a query
- Explore DB metadata
- Outputs a SQL query

DB Connector DB Table Selector



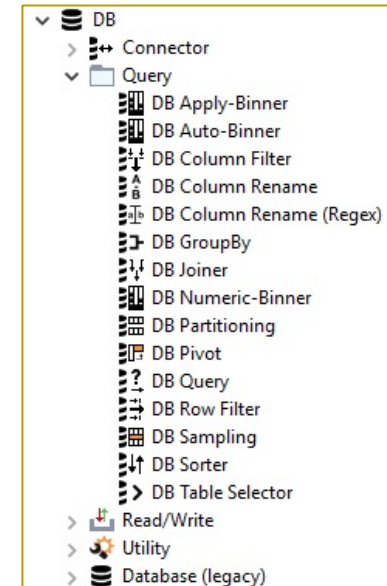
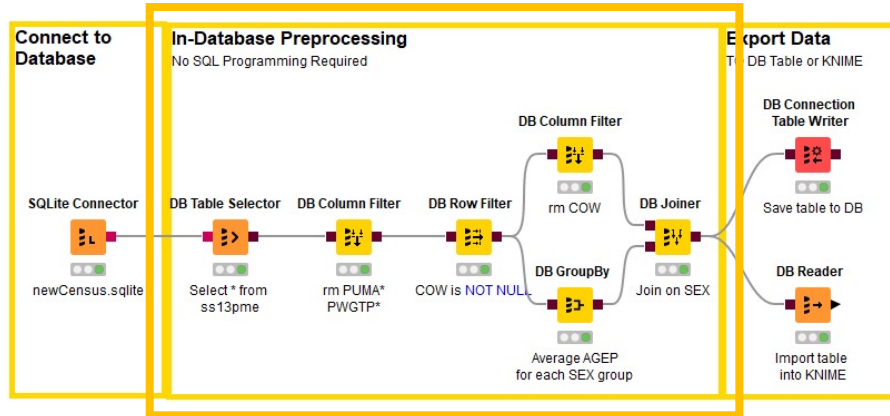
DB Reader

- Executes incoming SQL Query on Database
- Reads results into a KNIME data table



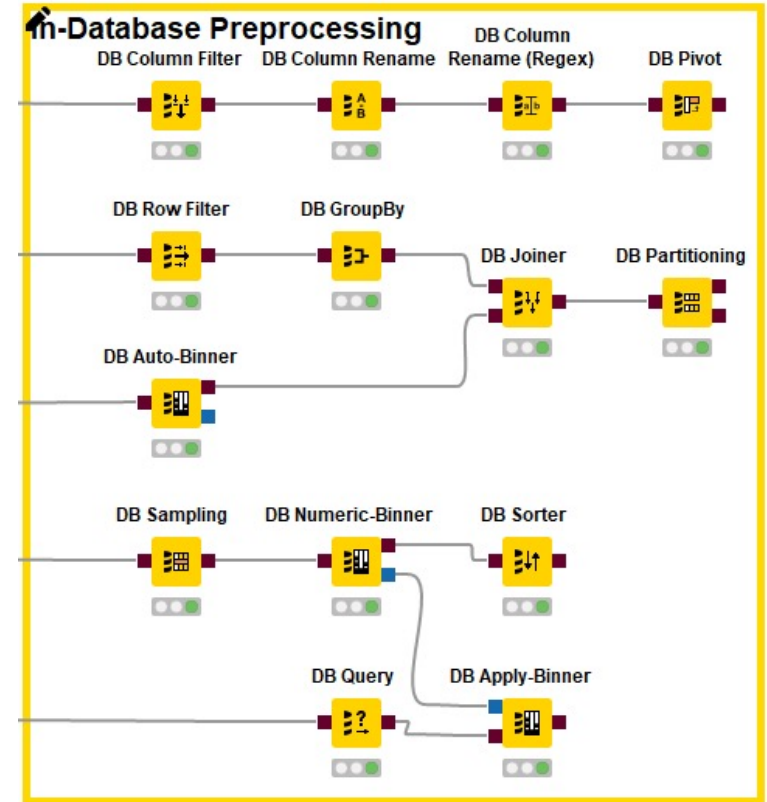
In-Database Processing

- Database Manipulation node generates a SQL query on top of the input SQL query (brown square port)



Query Nodes

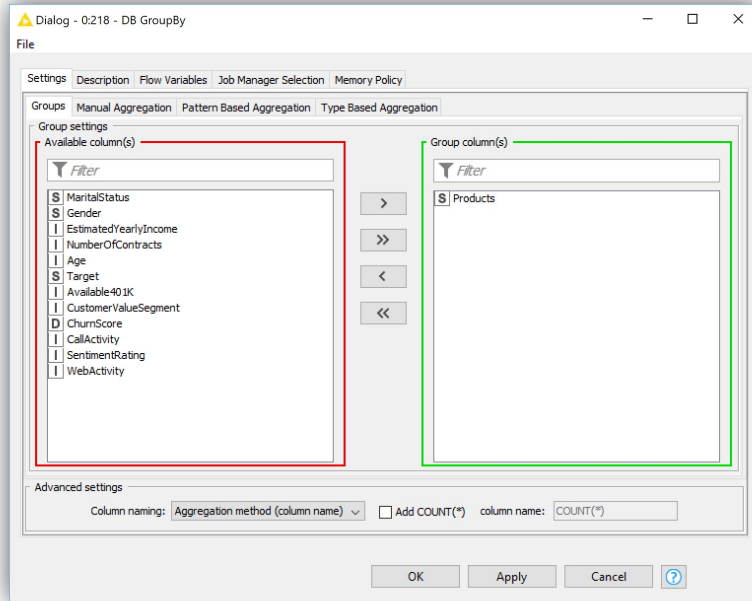
- Filter rows and columns
- Join tables/queries
- Extract samples
- Bin numeric columns
- Sort your data
- Write your own query
- Aggregate your data
- Partition your data



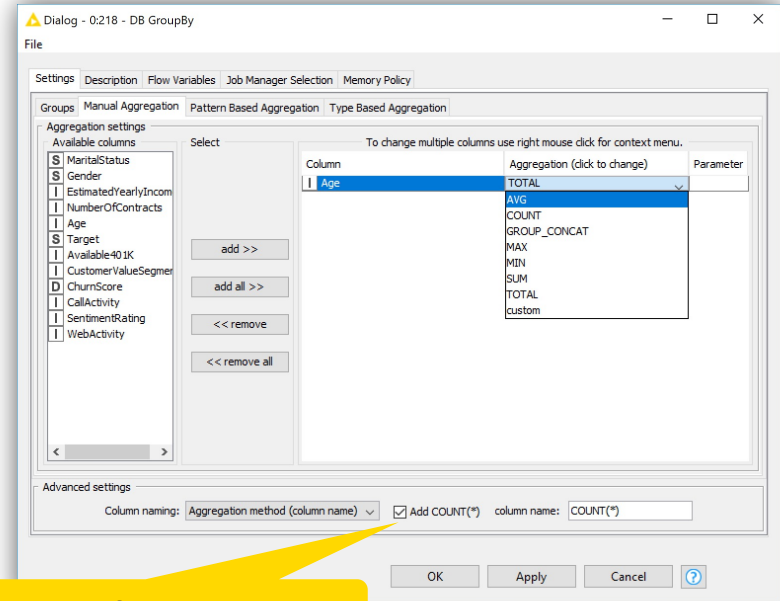
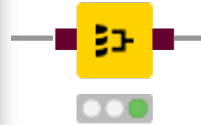
DB GroupBy

Aggregate rows to summarize data

- First tab provides grouping options
- Second tab provides control over aggregation details

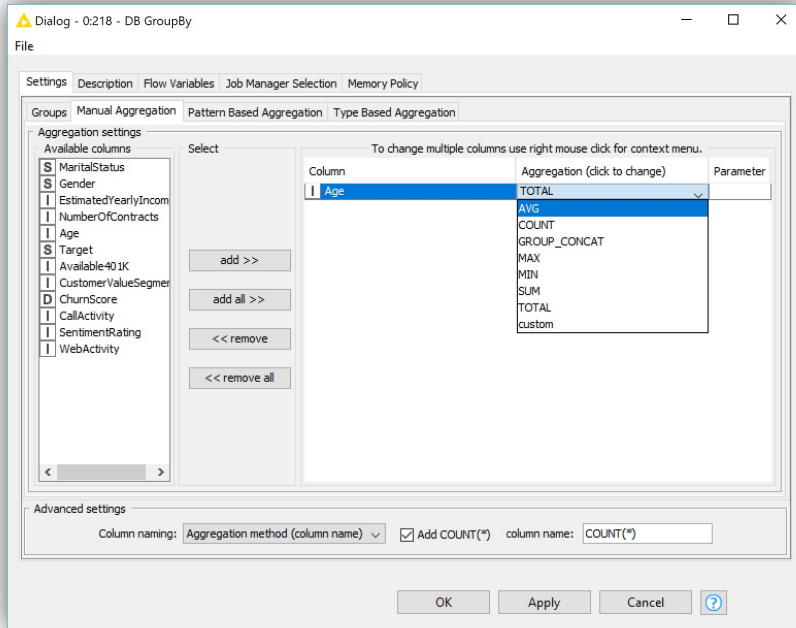


DB GroupBy

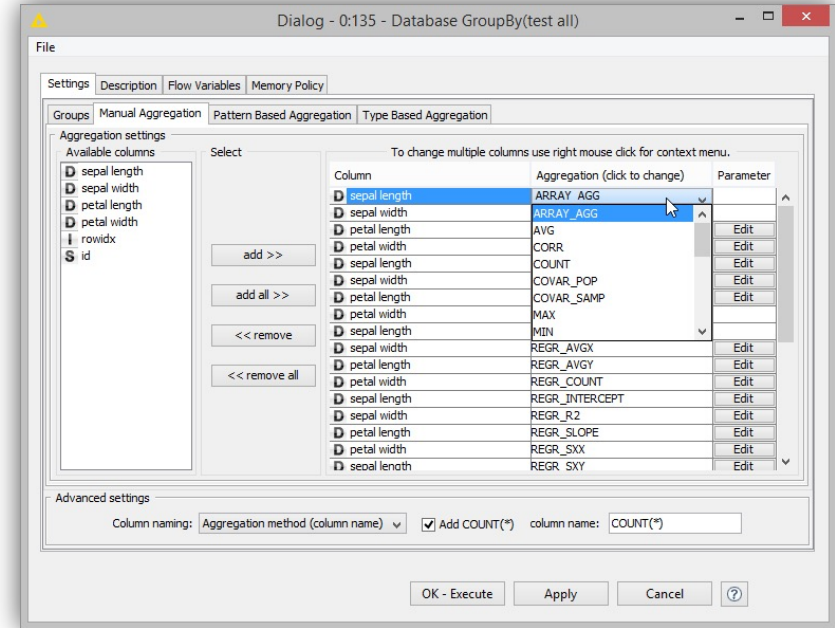


Returns number of rows per group

DB GroupBy – DB Specific Aggregation Methods



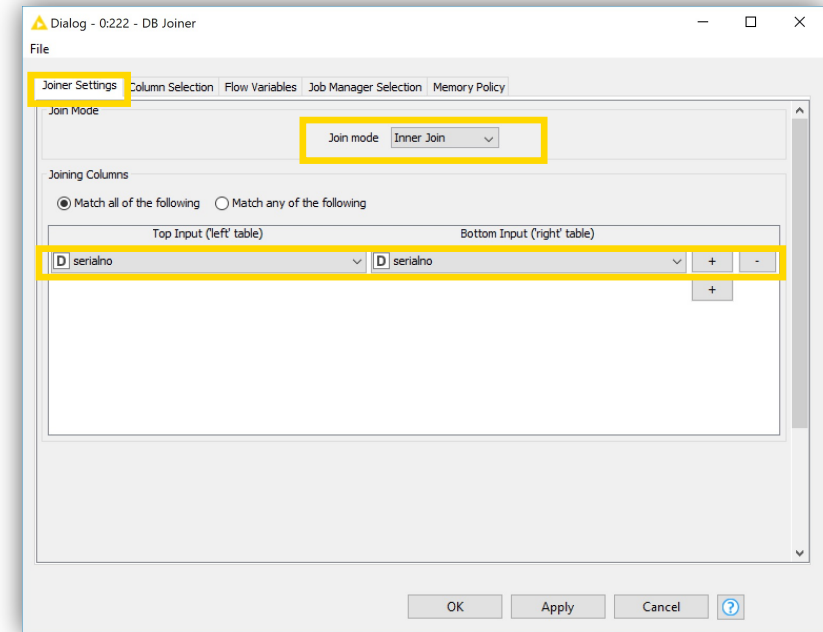
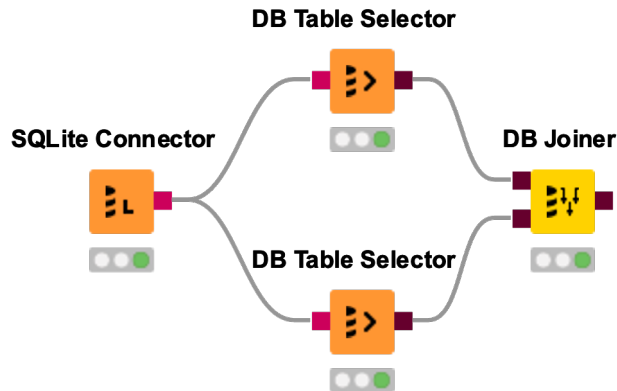
SQLite: 7 aggregation functions



PostgreSQL: 25 aggregation functions

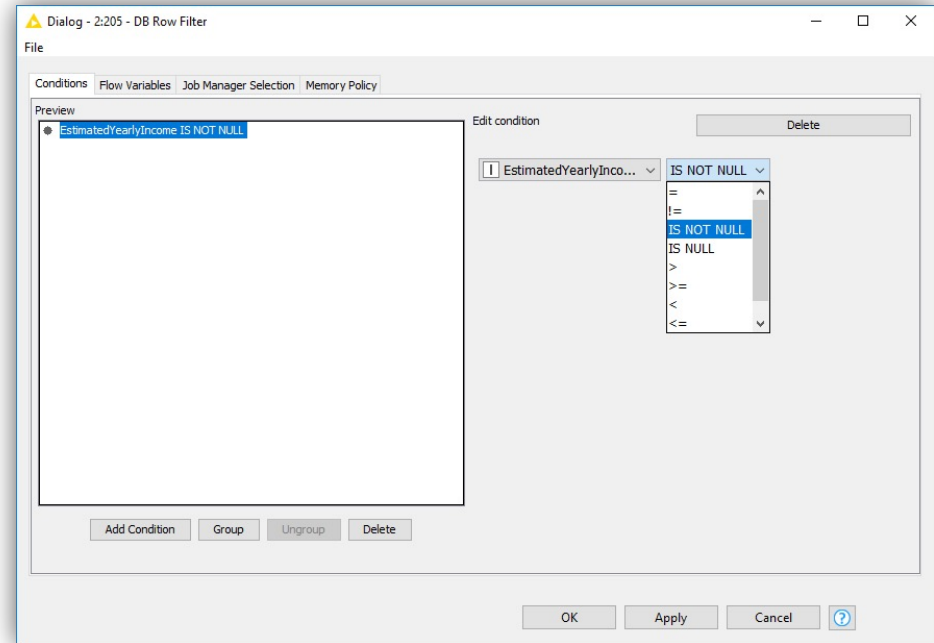
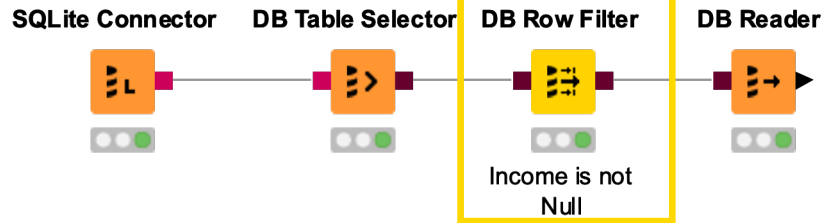
DB Joiner

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



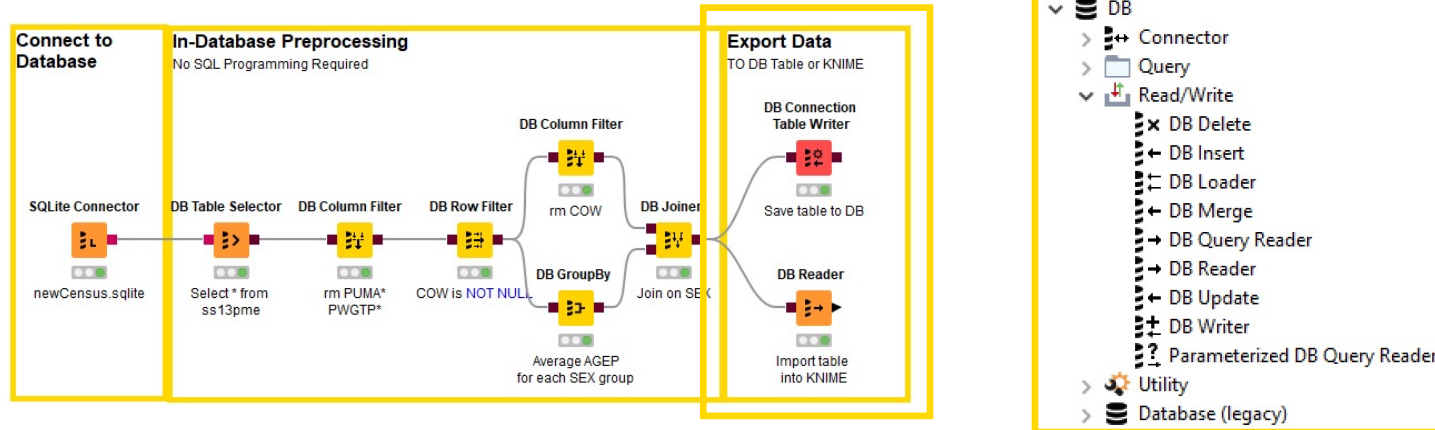
DB Row Filter

- Filters rows that do not match the filter criteria
- Use the *IS NULL* or *IS NOT NULL* operator to filter missing values



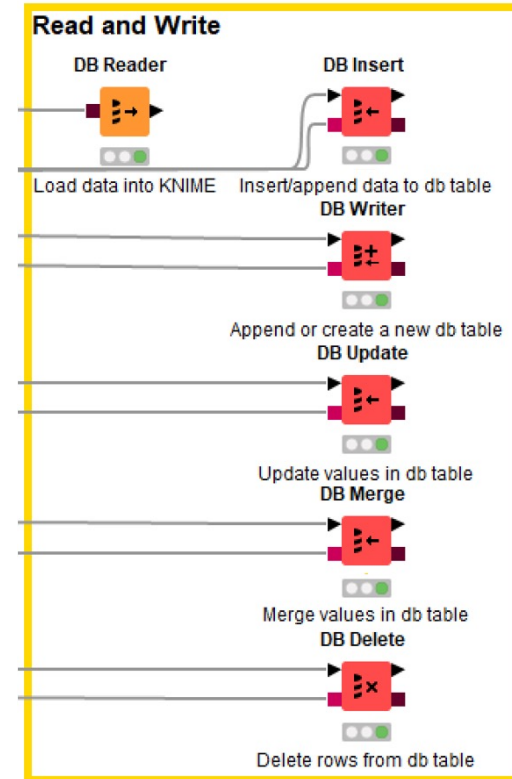
Export Data

- Writing data back into database
- Exporting data into KNIME
- SQL operations are **executed on the database!**



Database Writing Nodes

- Create table as select
- Insert/append/merge data
- Update values in table
- Delete rows from table



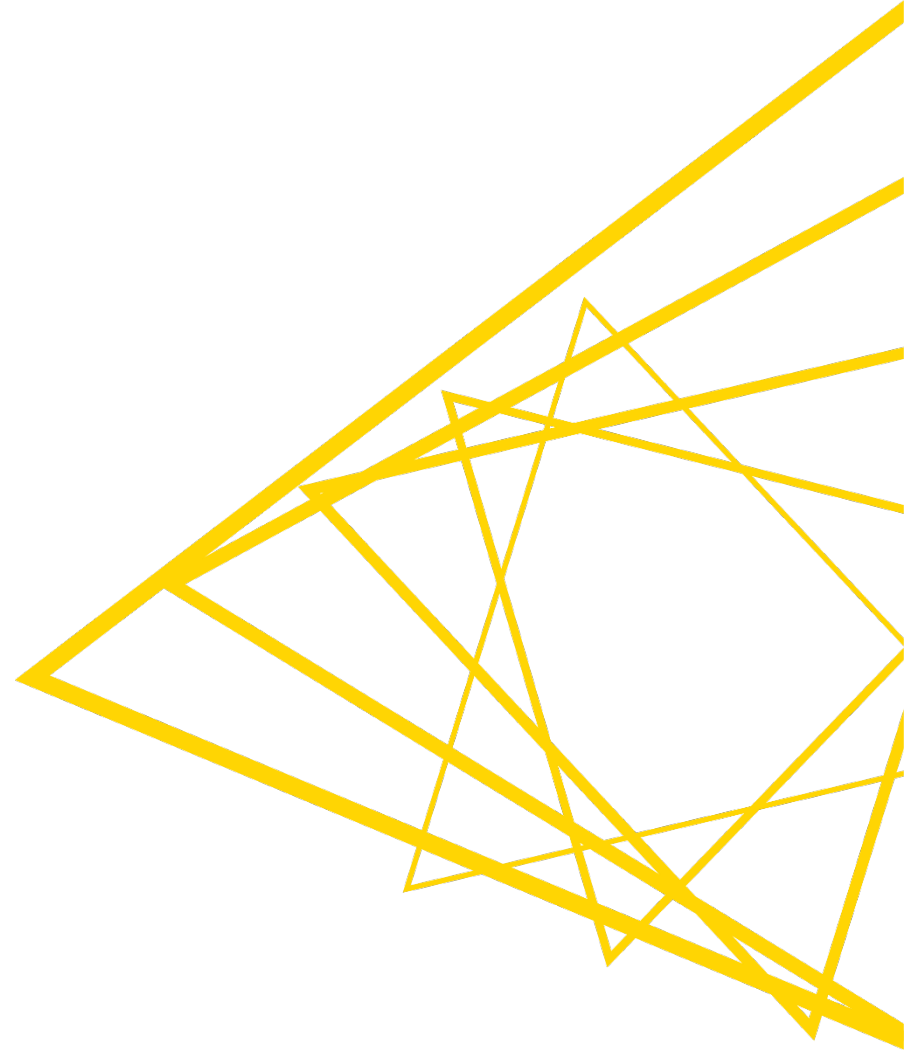
Databases Exercise

Open Exercise 05. Database

Activity I: Databases

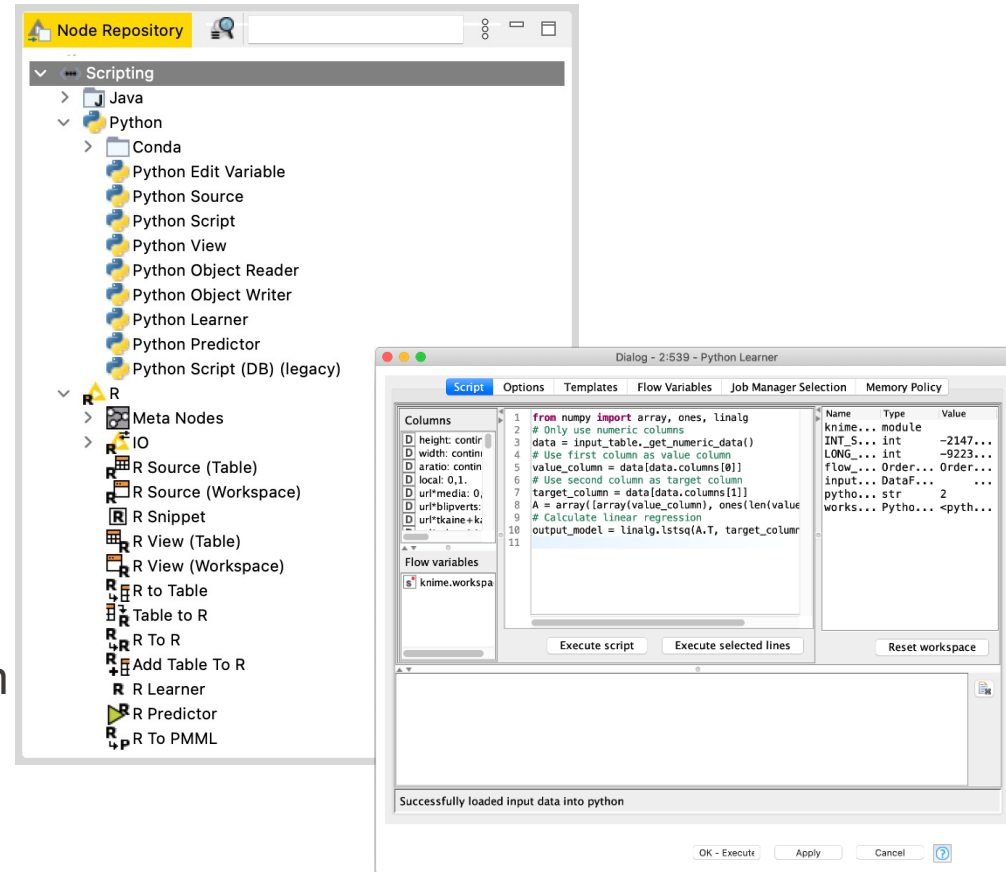
- Drag & drop the malariahts_DB.sqlite file into your workspace
- Connect malariahts_DB.sqlite to the two **DB Table Selector** nodes to read the "malariahts_experiment" and "malariahts_molecules_feature" table, respectively
- Filter the Column called "Pf3D7_pEC50" in the "malariahts_experiment" table with the **Database Column Filter** node
- Join the two tables together with the **DB Joiner** node using "Sample" column
- Read the data into KNIME using the **DB Reader** node.
- Write the data into a csv file

Integrating External Tools



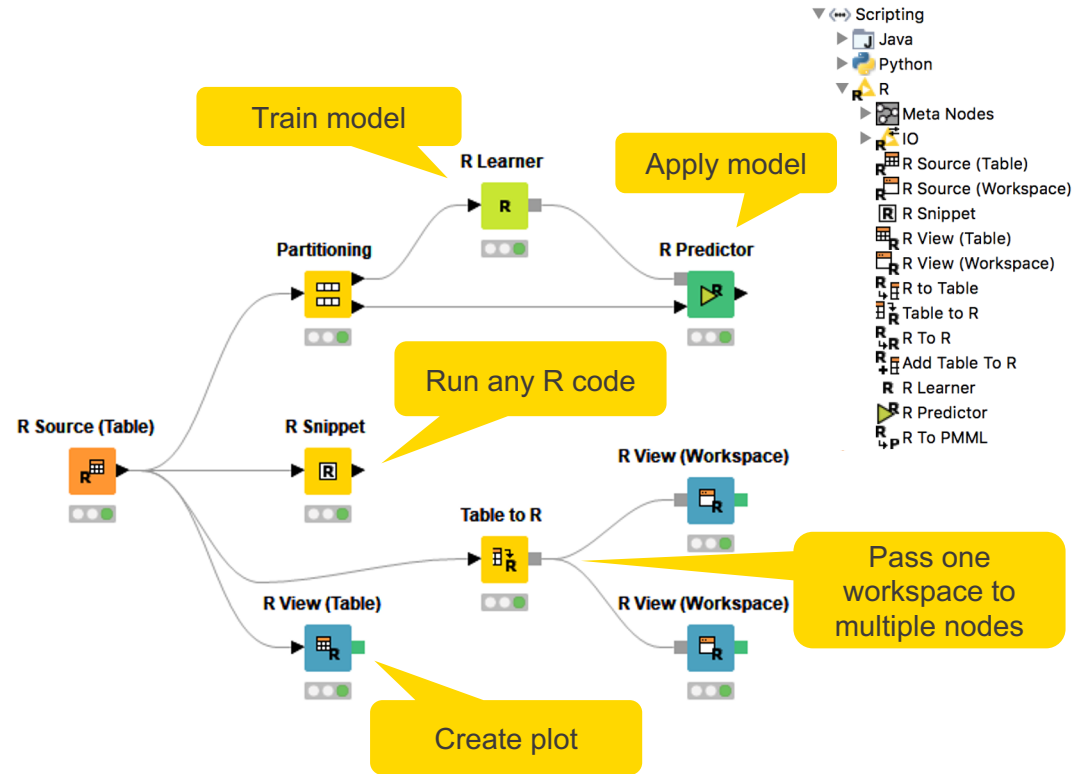
Scripting Integrations: R and Python

- Run R or Python code in KNIME Analytics Platform
- Works with existing Python and R installations
- Syntax highlighting support
- Different nodes for many tasks e.g training a model using an algorithm available in Python



R Integration

- Run any R code from KNIME
- Works with existing R installations
- Nodes for many tasks
- First run:
`install.packages('Rserve')`
and
`install.packages('Cairo')*`
*mac only



R Integration

The screenshot shows the KNIME RView (Table) window, which is used for running R scripts within the KNIME environment. The window is divided into several panes:

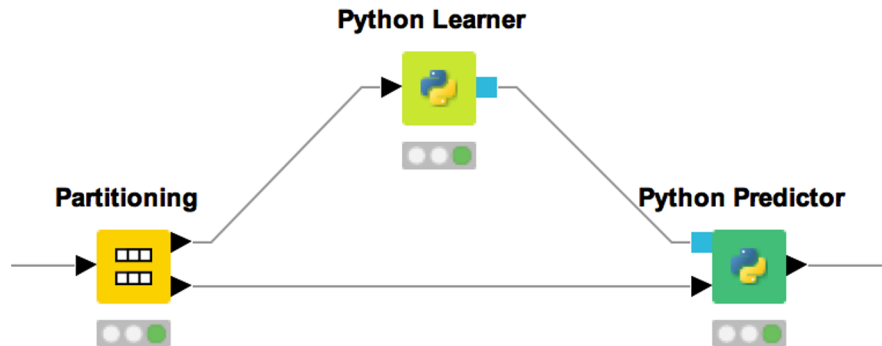
- Column List:** A list of columns available for the script, including Universe_0_0, Universe_0_1, Universe_1_0, Universe_1_1, and Cluster Membership.
- R Script:** A text editor where R code is written. The code includes comments, library calls for ggplot2 and grid, and a ggplot2 script that uses column references and a flow variable (kname.workspace) to create a plot. Syntax highlighting is applied to the code.
- Workspace:** A table showing the current R workspace variables. It has two columns: Name and Type. The variables listed are kname.flow.in (pairlist) and kname.in (data.frame).
- Flow Variable List:** A list of flow variables, including kname.workspace.
- Buttons:** Buttons for 'Eval Script', 'Eval Selection', 'Reset Workspace', and 'Show Plot' are located at the bottom of the R Script pane.
- Console:** A pane at the bottom for displaying the output of the R script.

Yellow callout boxes with arrows point to specific features:

- Syntax highlighting:** Points to the R script code.
- Create and store templates:** Points to the 'Create Template...' button.
- R workspace:** Points to the Workspace table.
- Show results:** Points to the 'Show Plot' button.
- Evaluate script:** Points to the 'Eval Script' button.
- R console output:** Points to the Console pane.

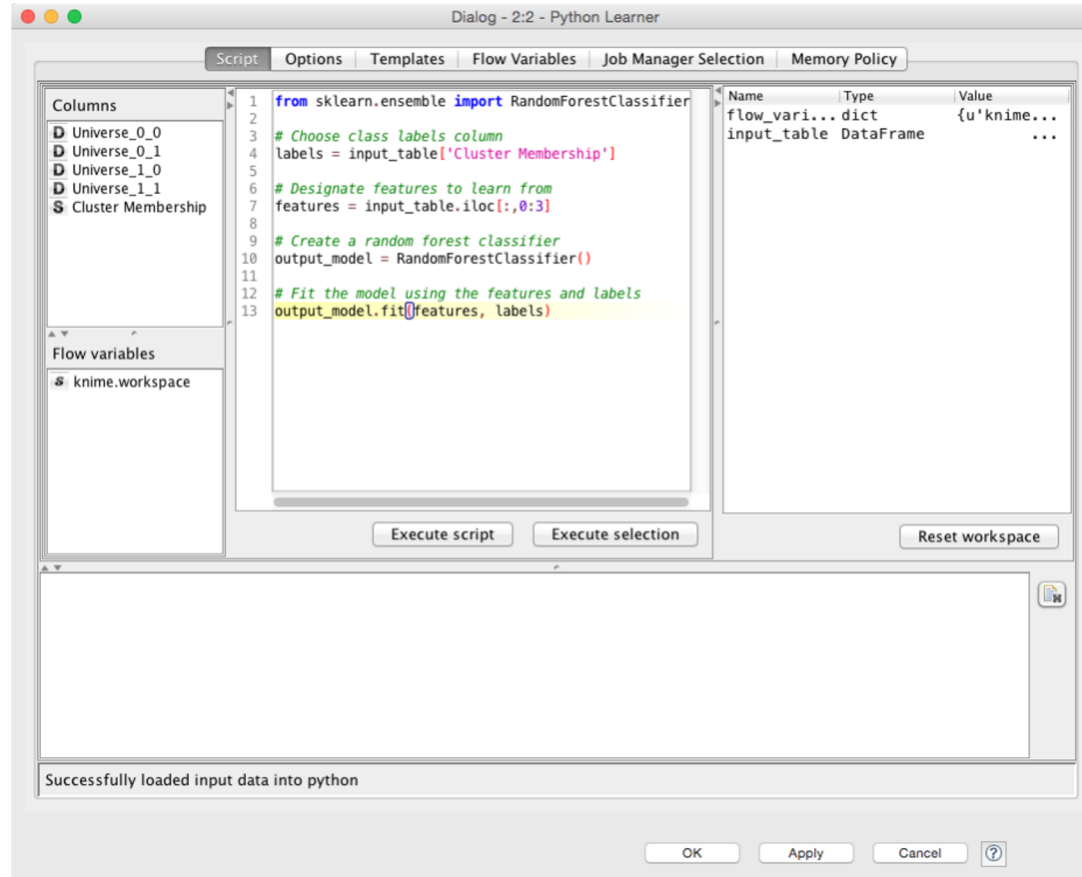
Python Integration

- Run Python inside KNIME
- Works with existing installations
- UI modeled after R integration



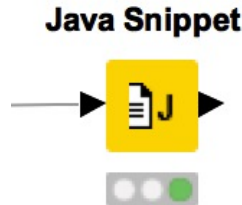
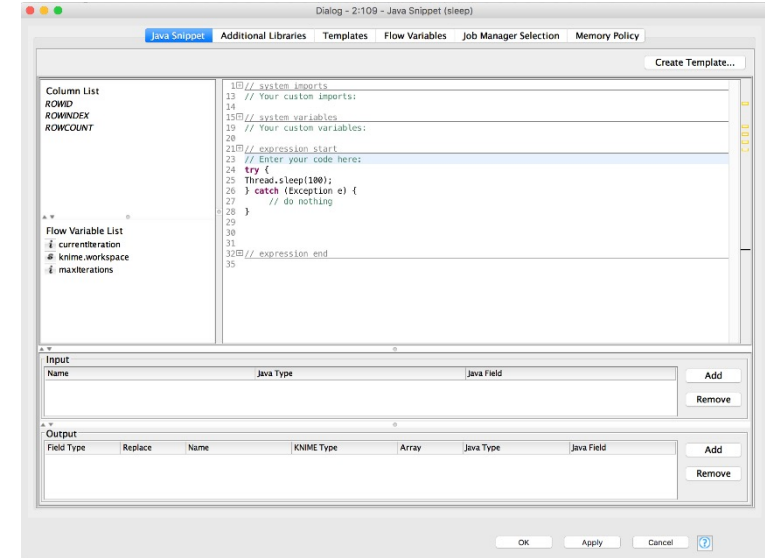
- ✓ Scripting
 - > Java
 - ✓ Python
 - > Conda
 - Python Edit Variable
 - Python Source
 - Python Script
 - Python View
 - Python Object Reader
 - Python Object Writer
 - Python Learner
 - Python Predictor
 - Python Script (DB) (legacy)

Python Scripting UI



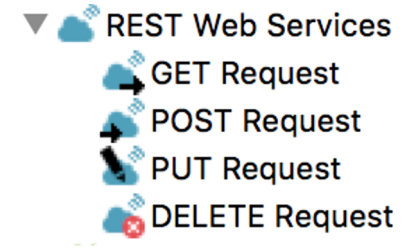
Java Snippet

- Fastest running scripting node in KNIME
- Syntax highlighting, auto completion, error checking
- Templates allow you to save scripts for later re-use
- Import custom libraries

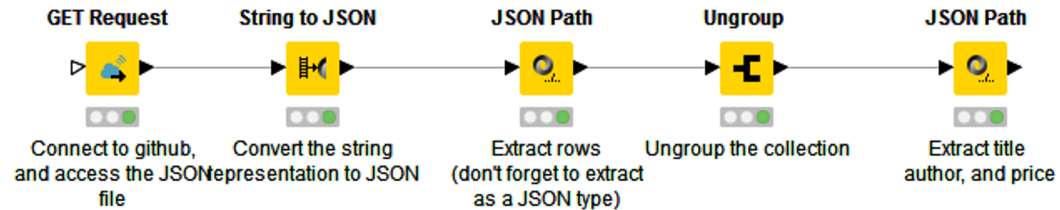


RESTful Web Services

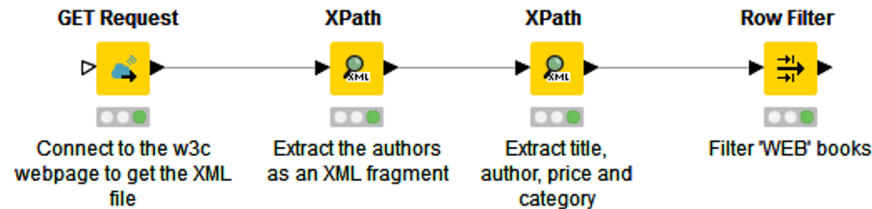
- Use KNIME nodes to interact with RESTful web services
- Send requests using standard HTTP methods



JSON Response:



XML Response:



RESTful Web Services

GET Request



Enter URL, or
use from column

Add delay between
individual requests

Provide authentication
if necessary

Dialog - 2:26 - GET Request (Connect to the w3c)

File

Connection Settings Authentication Request Headers Response Headers Flow Variables Job Manager Selection Memory Policy

☒ URL:

☐ URL column:

☐ Delay (ms):

Concurrency:

SSL

☐ Ignore hostname mismatches

☐ Trust all certificates

☐ Fail on connection problems (e.g. timeout, certificate errors, ...)

☐ Fail on http errors (e.g. page not found)

☒ Follow redirects

Timeout (s):

Body column:

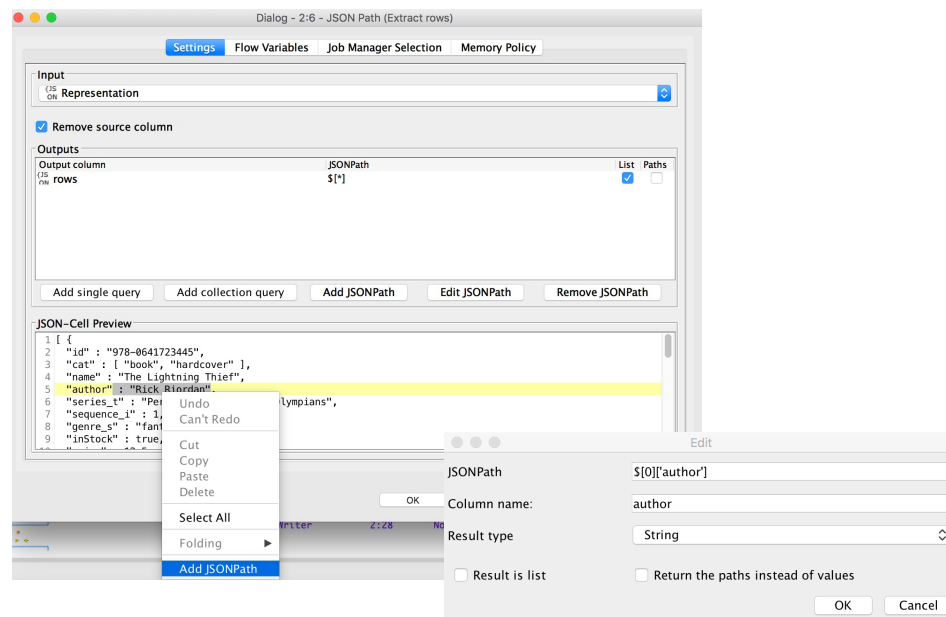
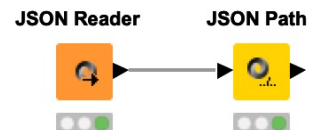
OK Apply Cancel ?

<https://www.knime.com/blog/a-restful-way-to-find-and-retrieve-data>

<https://www.knime.com/blog/OSM-meets-CSV-file-and-Google-API>

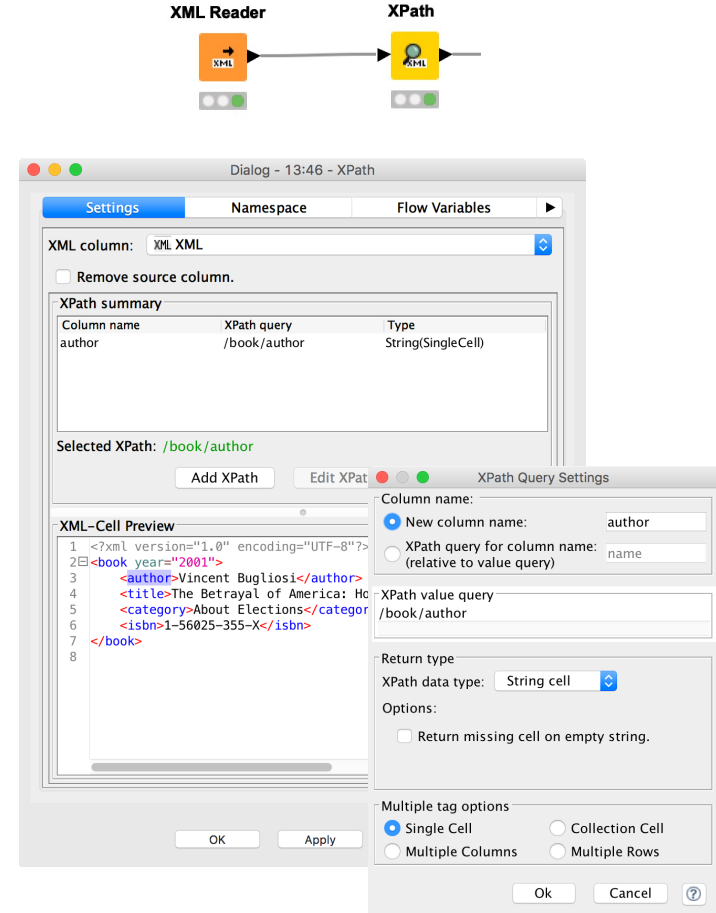
JSON Reader and JSON Path nodes

- Use the JSON Reader (or GET Request) node to get a JSON cell
- Use the JSON Path node to query the JSON file and extract parameters
- Editor window simplifies construction of JSON queries by auto-generating them (click on properties)



XML Reader and XPath nodes

- Use the XML Reader (or GET Resource) nodes to get an XML cell
- Use XPath nodes to query the XML and extract certain parameters
- Editor window simplifies construction of XPath queries by auto-generating them (click on XML elements)



External Tools Exercise

Open *Exercise 06. Integrating External Tools*

Activity: Use REST node to call an external Web Service

- Configure the **String Manipulation** node so that it creates a new column called `resource_url` with the following format.
<https://www.ebi.ac.uk/chembl/api/data/molecule/<ChEMBL ID>.xml>
- Access the ChEMBL Rest API with the **GET Request** node
- Use the **XPath** node to extract `standard_inchi` and `standard_inchi_key`
- Use **RDKit From InChI** node to convert the `standard_inchi` column to RDKit molecules type.
- Use **Renderer to Image** node to convert the RDKit molecules into images.
- Display the result in a table view

Optional:

- Use the R Snippet to filter the `standard_inchi` and `standard_inchi_key`. Hint: You can use the Regex column filter template in the snippet with the expression: `standard.*`

Exporting Data & Deployment



Exporting Data

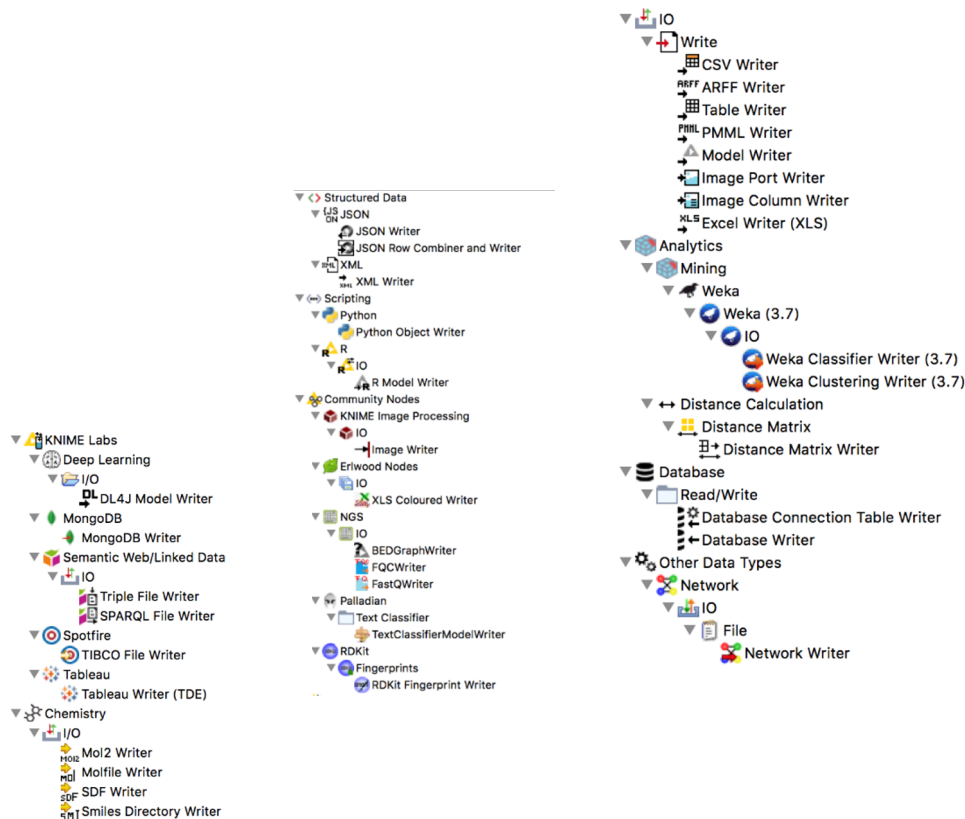
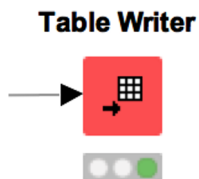
After an analysis is completed, what next?

- Write results to a file
- Create/update a database
- Upload results to a Cloud Storage
- Generate a report
- Send your data to Tableau, Spotfire, PowerBI to create a report
- Deploy via KNIME WebPortal (KNIME Server required)

Data Export Nodes

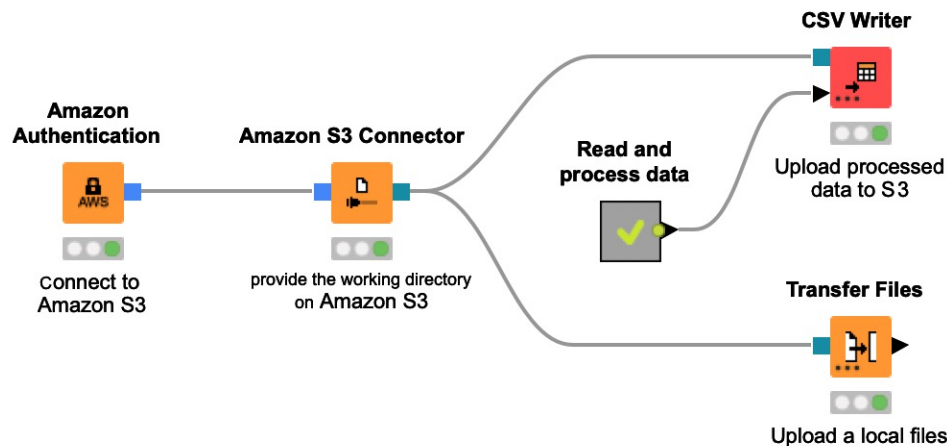
Typically characterized by:

- Magenta color
- 1 input port, no output ports
- Create file on file system or write to database

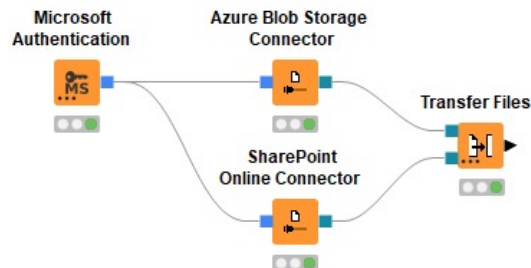


Write Files to a Remote File System

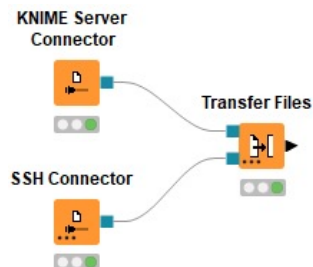
- The new file handling framework makes it easy to upload data to remote file systems
 - Write processed data directly with a writer node
 - Upload local files with the Transfer Files node
- Supported file systems
 - Microsoft Azure
 - Google
 - Amazon
 - Databricks
 - BigData file systems (hdfs, httpFS, ...)
 - On-premise (e.g. ssh, ftp, ...)



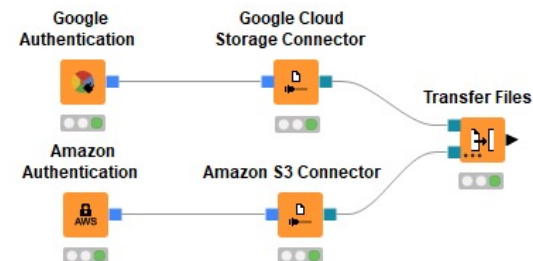
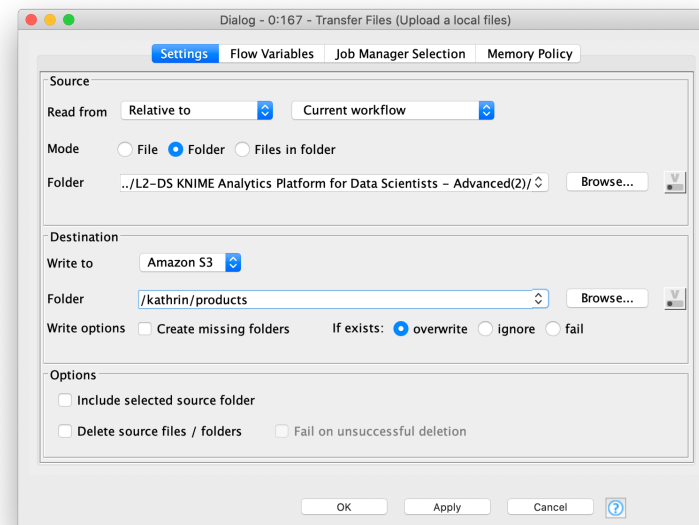
Full Flexibility with the Transfer Files node



Same cloud environment



On-premise

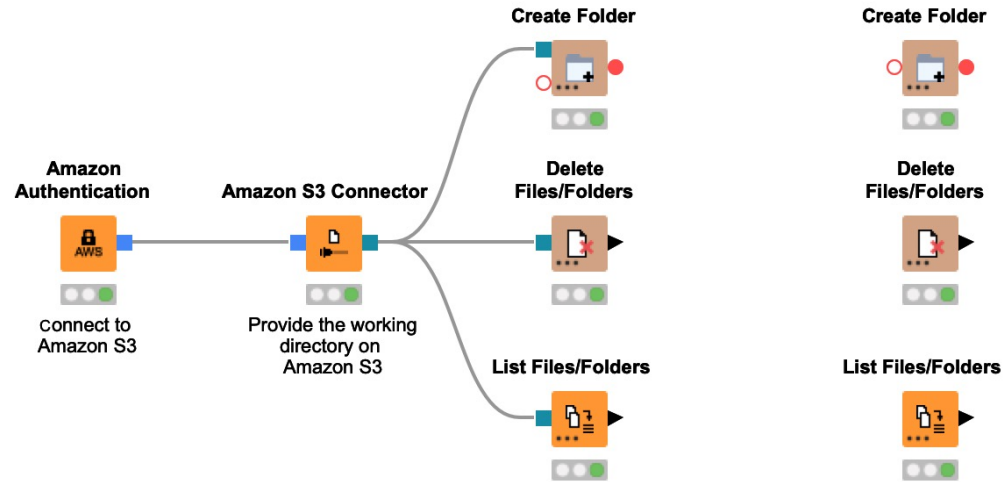


Cross cloud environments

Other Utility Nodes

Can be used local and with remote file systems

- Create a folder
- Delete files or folders
- List all files in a folder



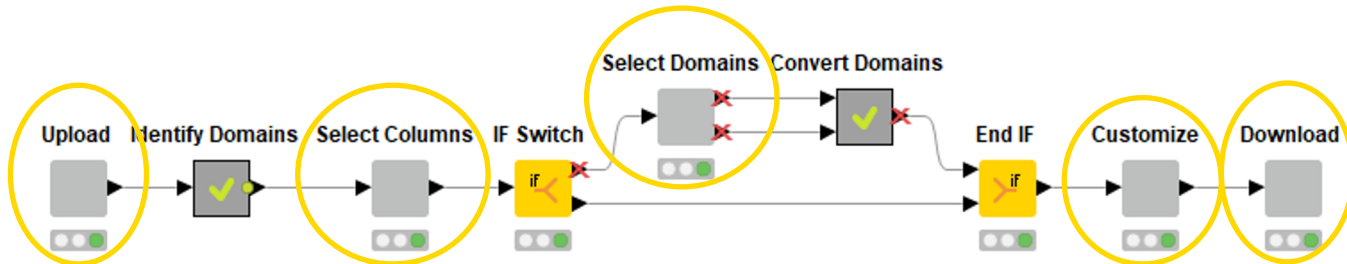
- Further information about file handling

https://docs.knime.com/latest/analytics_platform_file_handling_guide/index.html

To Dashboard on WebPortal

The Process Step by Step

1. Upload your data / Select one of the available datasets
2. Select the columns to visualize (maximum 3)
3. Convert the domain of the columns (OPTIONAL)
4. Customize the visualizations interactively
5. Download the images of the customized charts



Step 1
Upload File

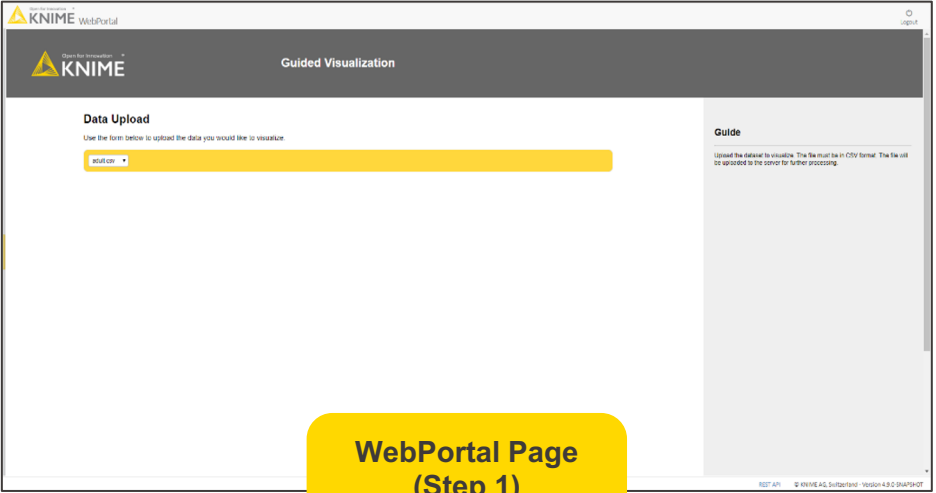
Step 2
Select Columns

Step 3
Customize
Column Domains

Step 4
Interactive View

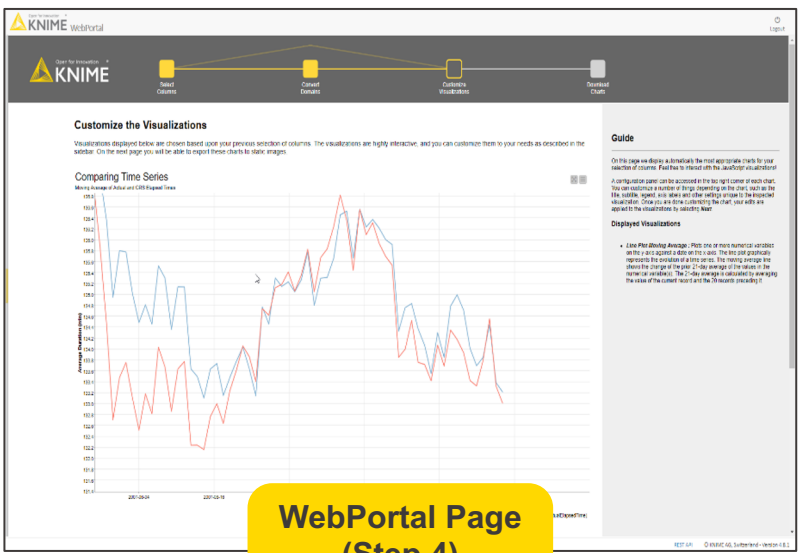
Step 5
Download Image

Workflow on KNIME WebPortal



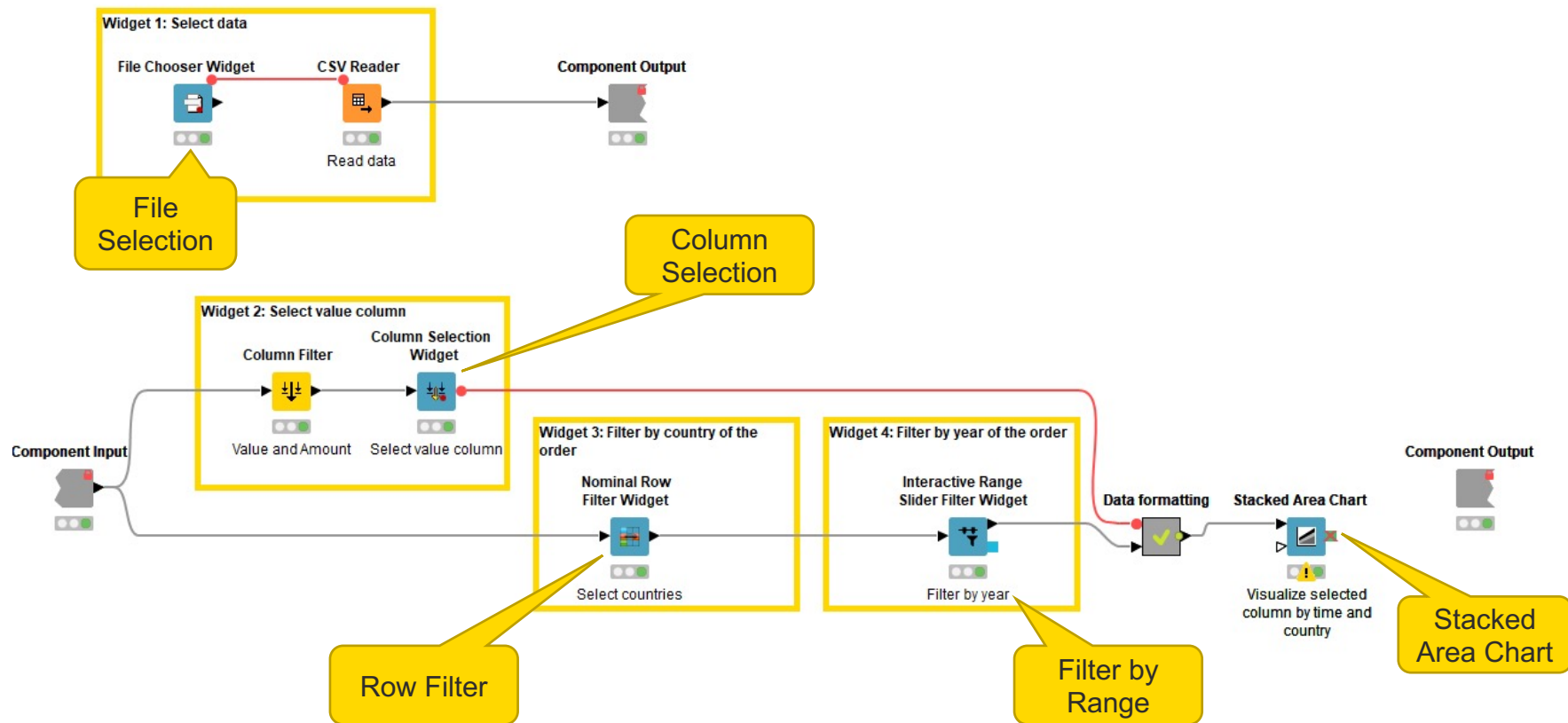
**WebPortal Page
(Step 1)
Upload File**

**Available in
KNIME Server**



**WebPortal Page
(Step 4)
Interactive View**

Components to Produce Dashboard on Web Page



Demo: Deploy a Web App (KNIME Server required)

- Deploy a workflow on KNIME Server and execute it on the KNIME Web Portal

Stay connected with KNIME



Blog: knime.com/blog



Forum: forum.knime.com



KNIME Hub: hub.knime.com



KNIME E-Learning Course:
www.knime.com/e-learning-course

Follow us on social media:



Thank You!
KNIME AG

