



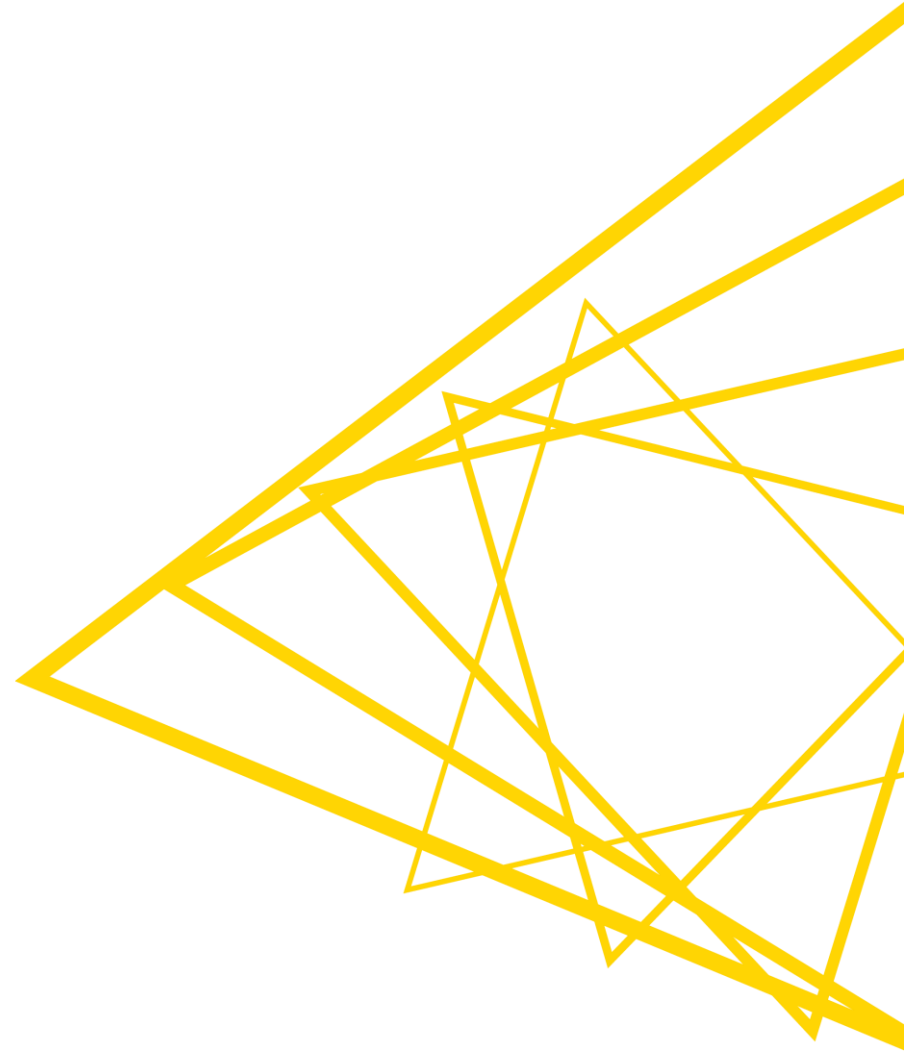
Open for Innovation

KNIME

KNIME Analytics Platform

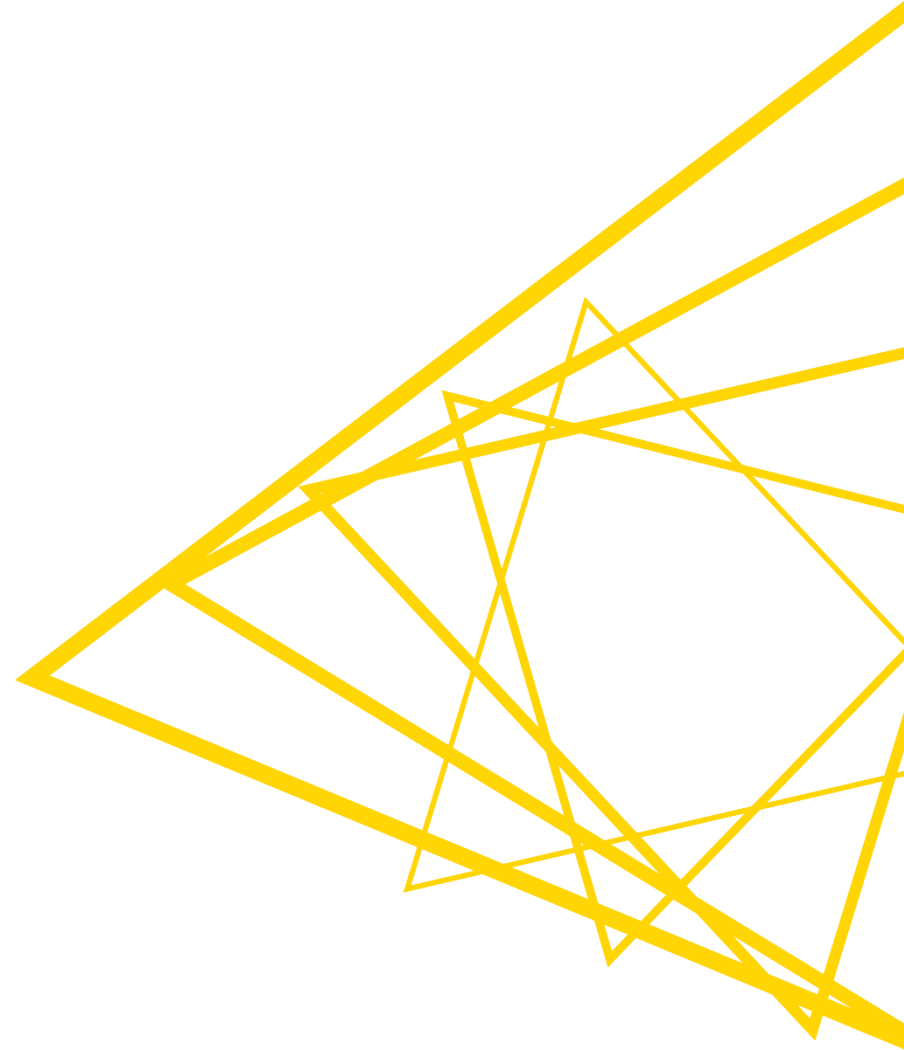
Dr. Alice Krebs
September 12, 2023

@ Syngenta



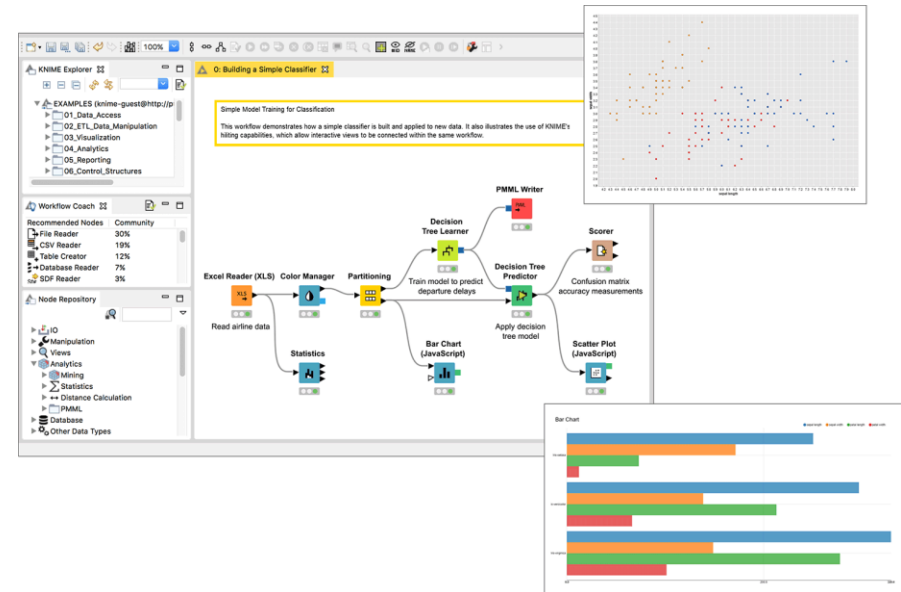
Overview

KNIME Analytics Platform



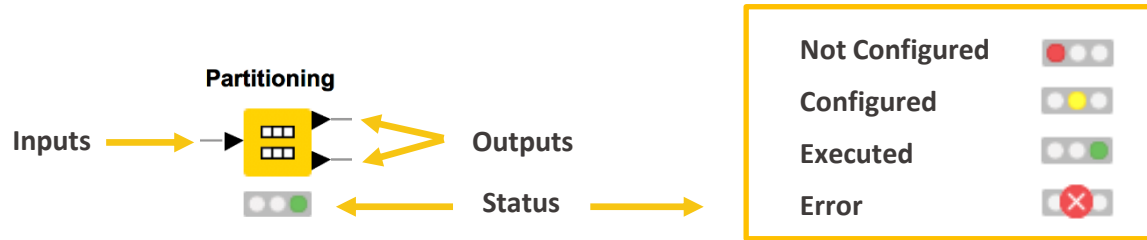
What is KNIME Analytics Platform?

- 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.

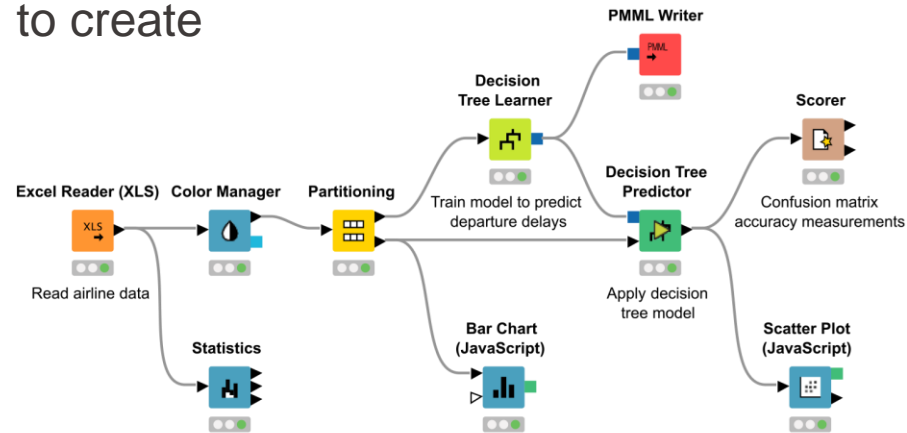


Visual KNIME Workflows

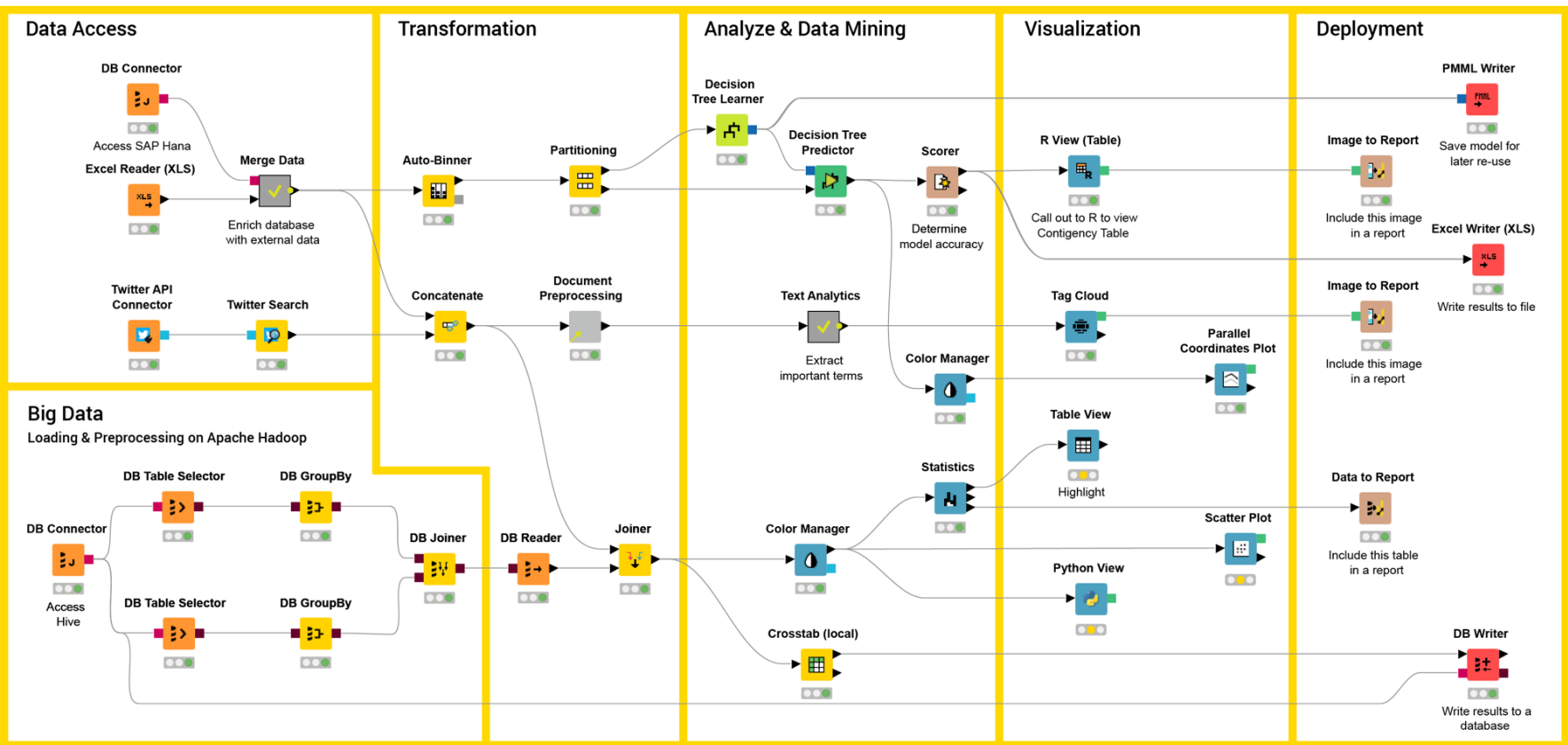
NODES perform tasks on data



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)

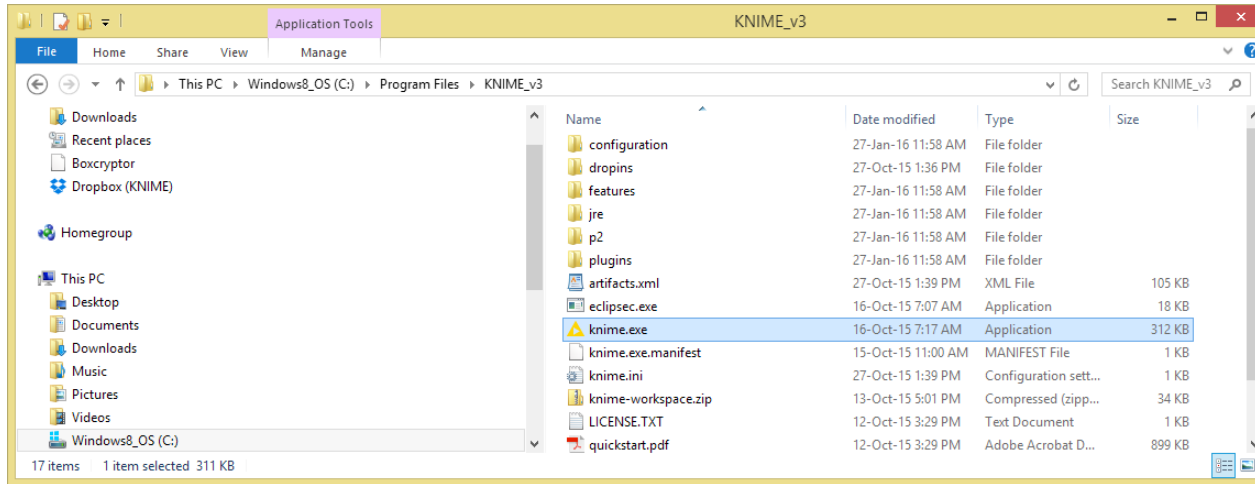
Download from here: <https://www.knime.com/downloads>

Start KNIME Analytics Platform

- Use the shortcut created by the installer

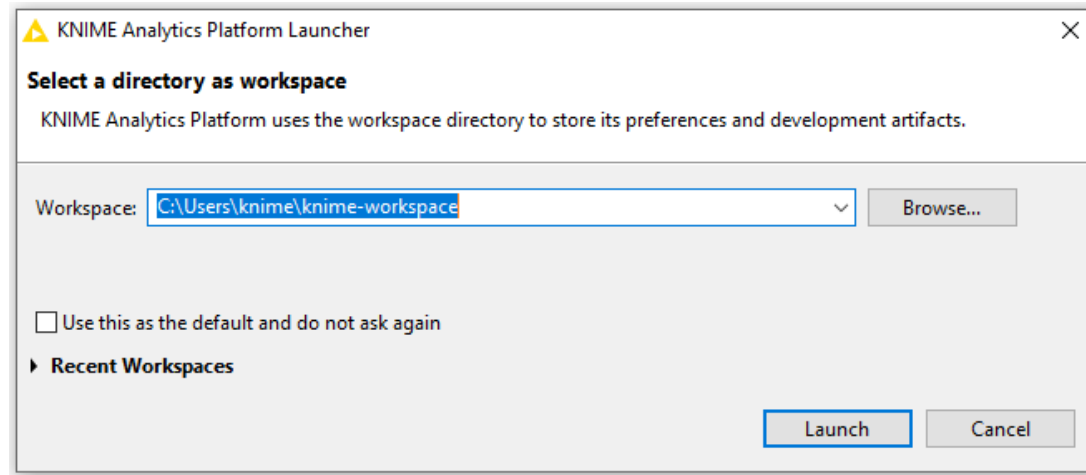


- Or go to the installation directory and launch KNIME via the knime.exe



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

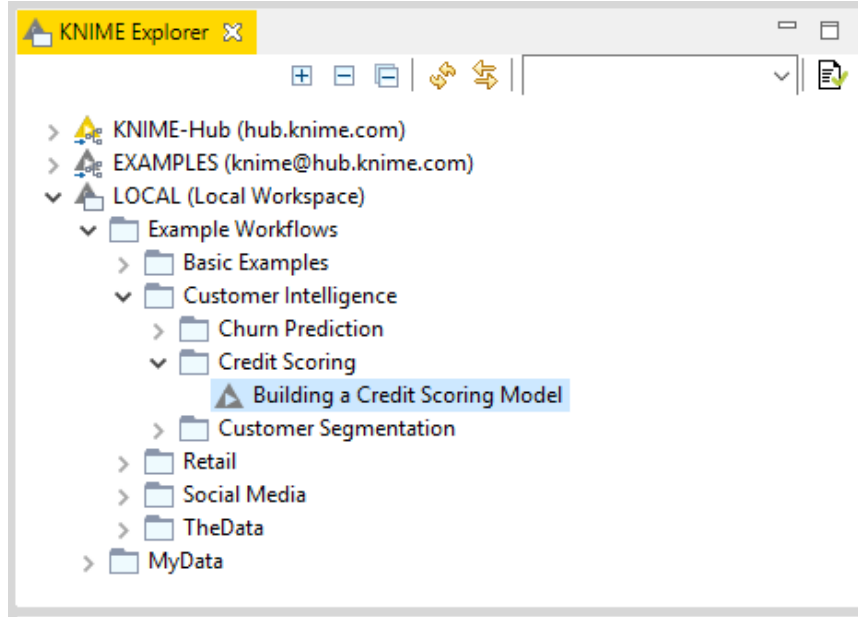
The screenshot displays the KNIME Analytics Platform Workbench interface. The main workspace shows a workflow titled "My first Workflow" with the following steps: File Reader (read adult.csv) → Row Filter (keep only records born in the US) → Column Filter (remove gender) → Table Writer (write table).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



- KNIME Explorer:** Located on the top left, it shows a file tree with folders like "My-KNIME-Hub", "EXAMPLES", and "LOCAL (Local Workspace)".
- Workflow Coach:** Located on the middle left, it displays a list of recommended nodes such as "GroupBy", "Joiner", "Column Filter", and "Concatenate".
- Node Repository:** Located on the bottom left, it provides a hierarchical view of all available nodes, categorized by function like "IO", "Manipulation", and "Analytics".
- Node Description:** Located on the top right, it provides detailed information about the selected "Row Filter" node, including its purpose and configuration options.
- KNIME Hub:** Located on the middle right, it features a search bar and a "Sign in" button for accessing the KNIME Hub.
- Outline:** Located at the bottom left of the main workspace, it provides a small overview of the current workflow.
- Console & Node Monitor:** Located at the bottom right, it displays the execution status of the selected node and its output data.

The "Console & Node Monitor" panel shows the output of the "Row Filter" node (3:6) in a table format:

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

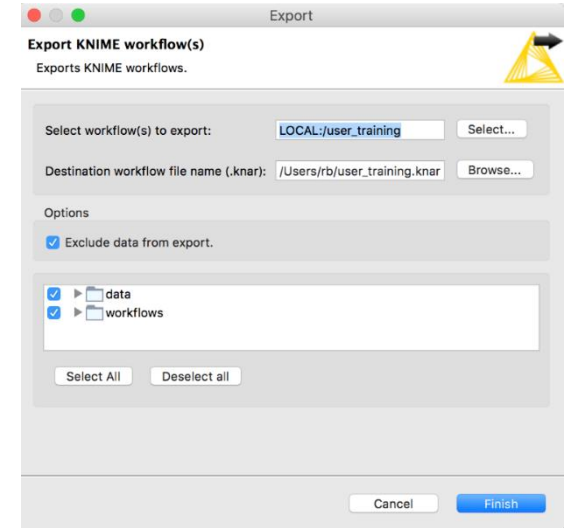
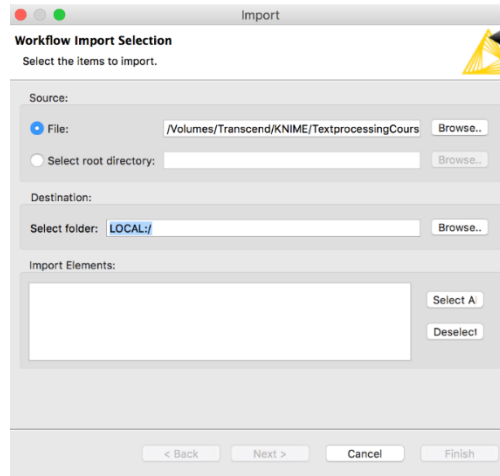
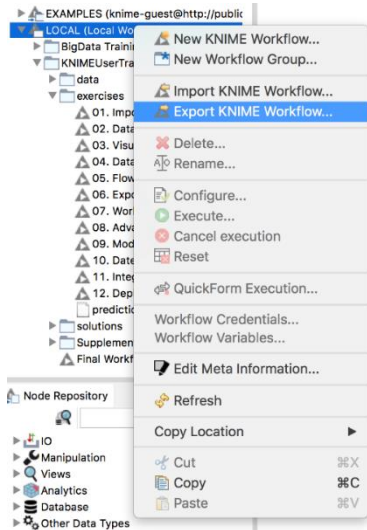
KNIME Explorer



- 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



KNIME File Extensions

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

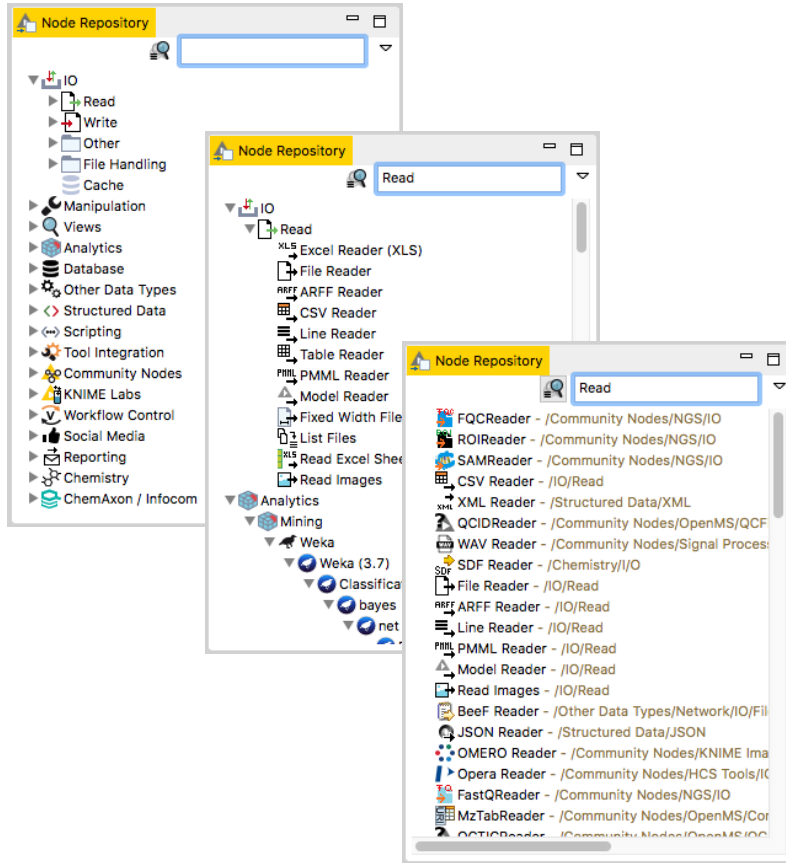
- ***.knwf** for KNIME Workflow Files





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

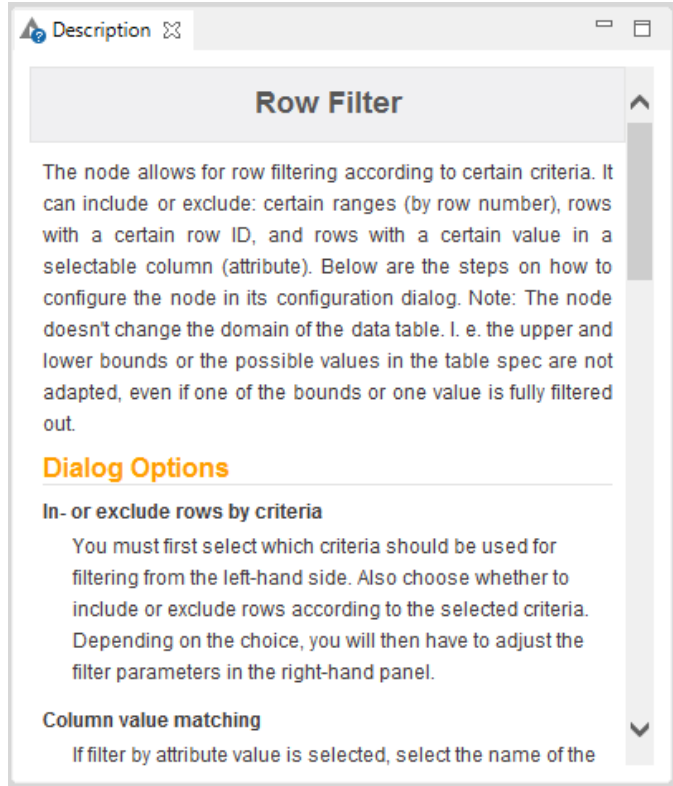


Node Repository



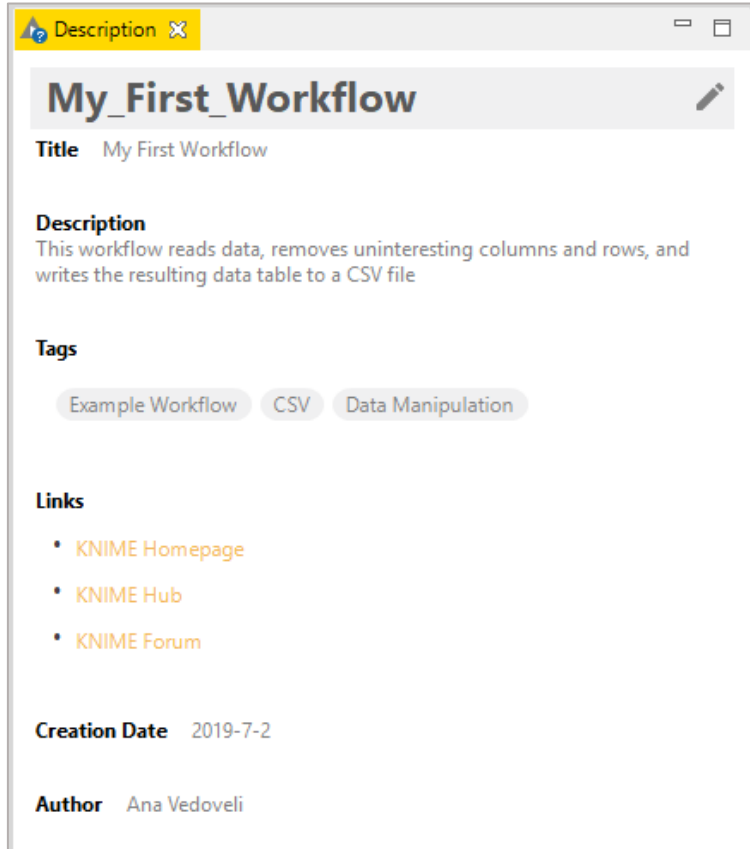
- The Node Repository lists all KNIME nodes
- The search box has 2 modes
 -  Standard Search – exact match of node name
 -  Fuzzy Search – finds the most similar node name
- Nodes can be added by drag and drop from the Node Repository to the Workflow Editor.

Description



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

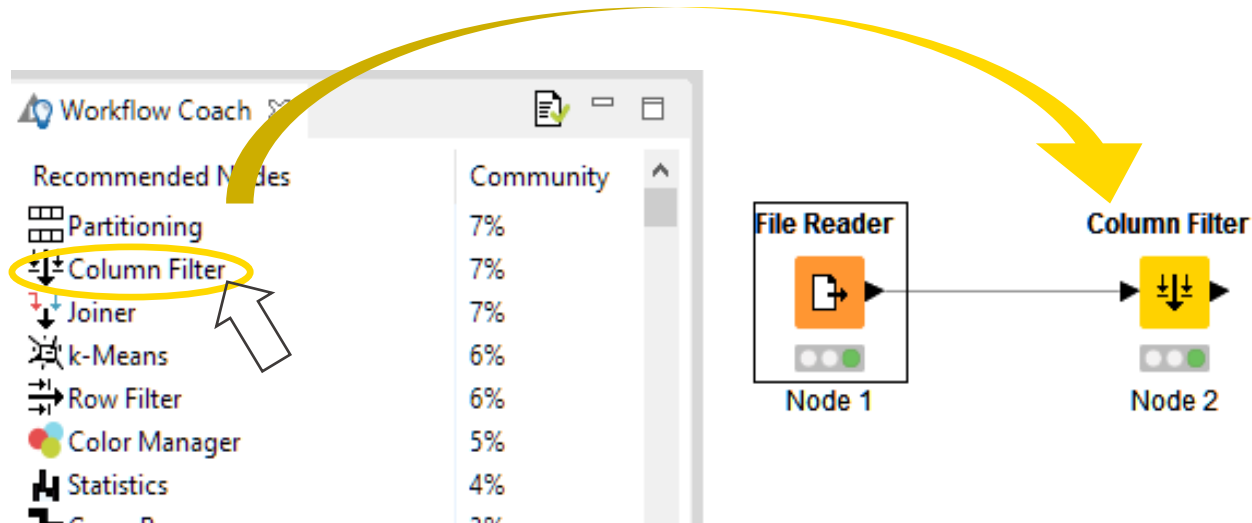
Workflow Description



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

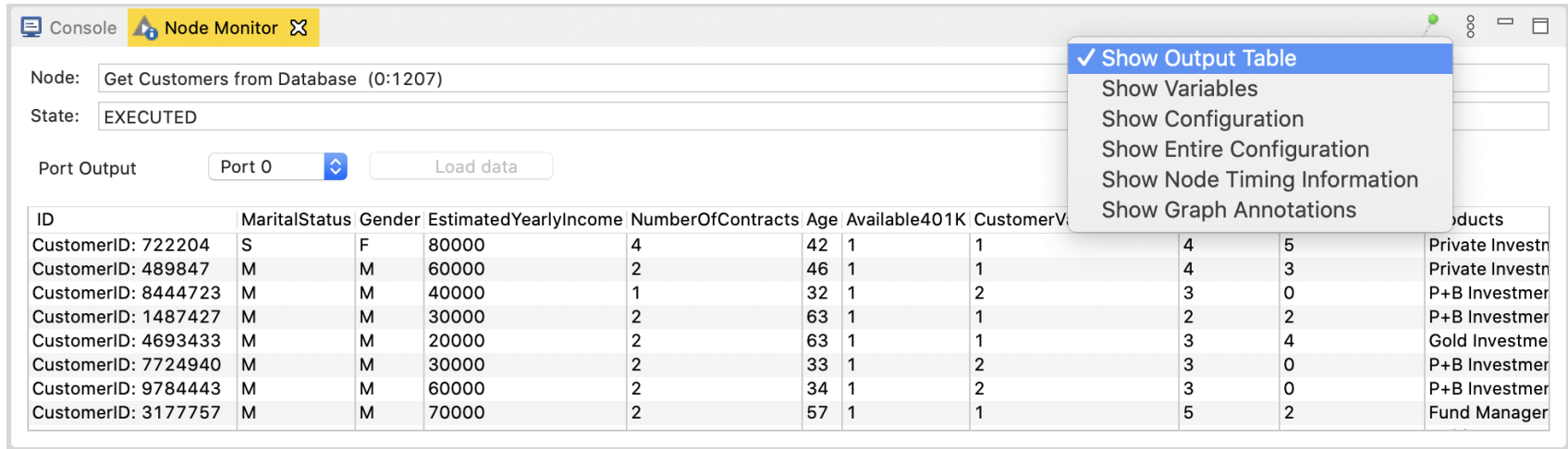
Workflow Coach

- 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

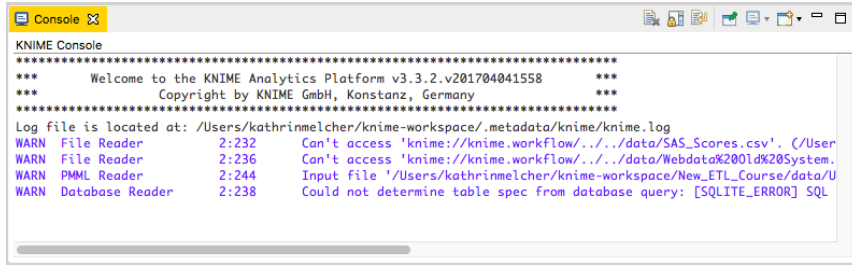
- By default the Node Monitor shows you the output table of the node selected in the workflow editor
- Click on the three dots on the upper right to show the flow variables, configuration, etc.



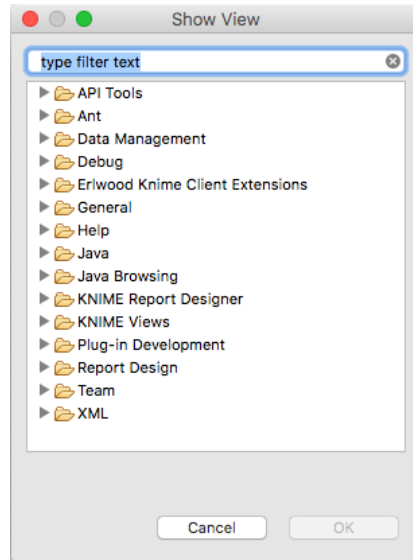
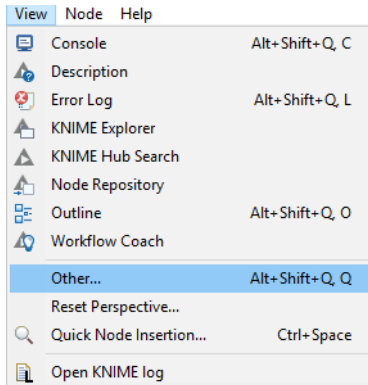
The screenshot displays the KNIME Node Monitor window. The selected node is 'Get Customers from Database (0:1207)' with a state of 'EXECUTED'. The 'Port Output' is set to 'Port 0'. A context menu is open, showing options: 'Show Output Table' (checked), 'Show Variables', 'Show Configuration', 'Show Entire Configuration', 'Show Node Timing Information', and 'Show Graph Annotations'.

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

Console and Other Views



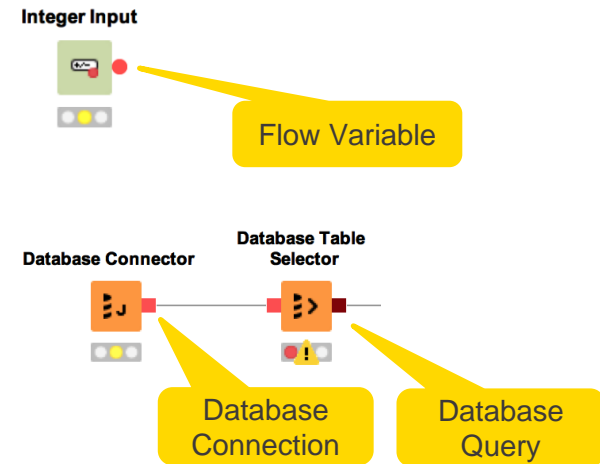
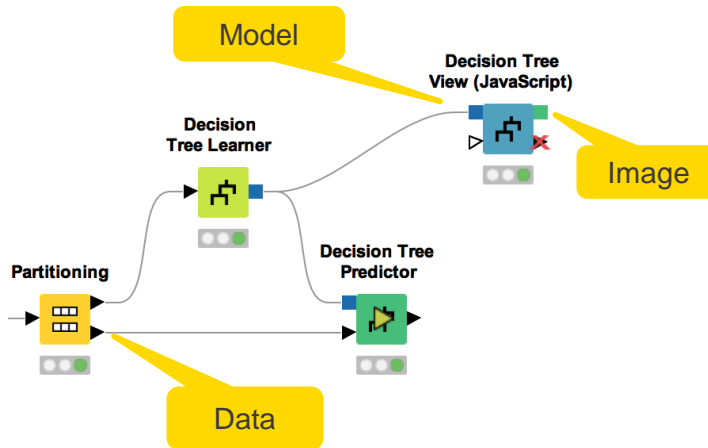
```
KNIME Console
*****
*** Welcome to the KNIME Analytics Platform v3.3.2.v201704041558 ***
*** Copyright by KNIME GmbH, Konstanz, Germany ***
*****
Log file is located at: /Users/kathrinmelcher/knime-workspace/.metadata/knime/knime.log
WARN File Reader 2:232 Can't access 'knime://knime.workflow/../../data/SAS_Scores.csv'. (/User
WARN File Reader 2:236 Can't access 'knime://knime.workflow/../../data/Webdata%20old%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

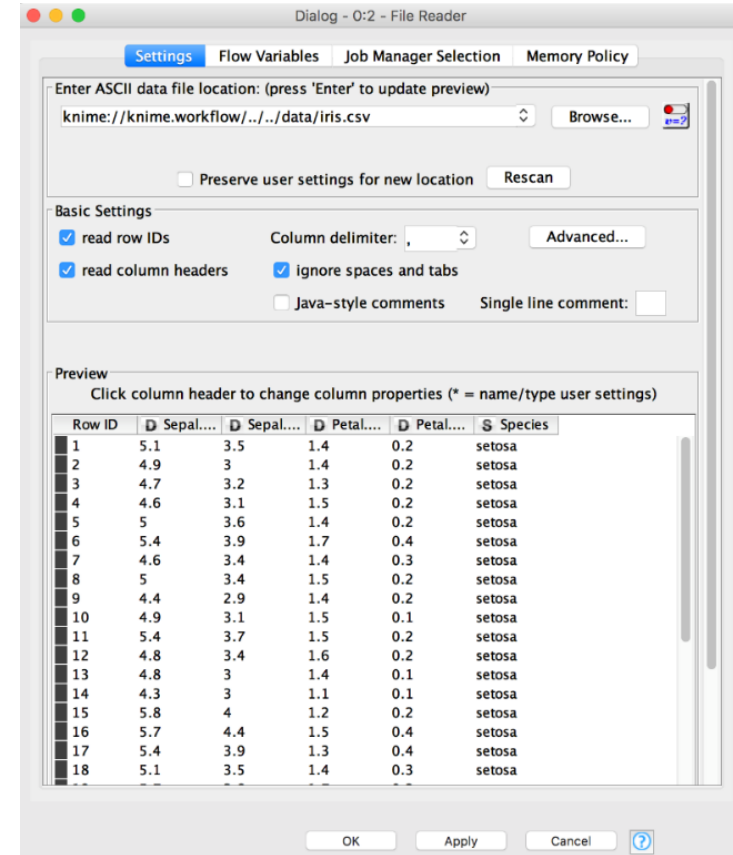


Error:

The node has encountered an error during execution.

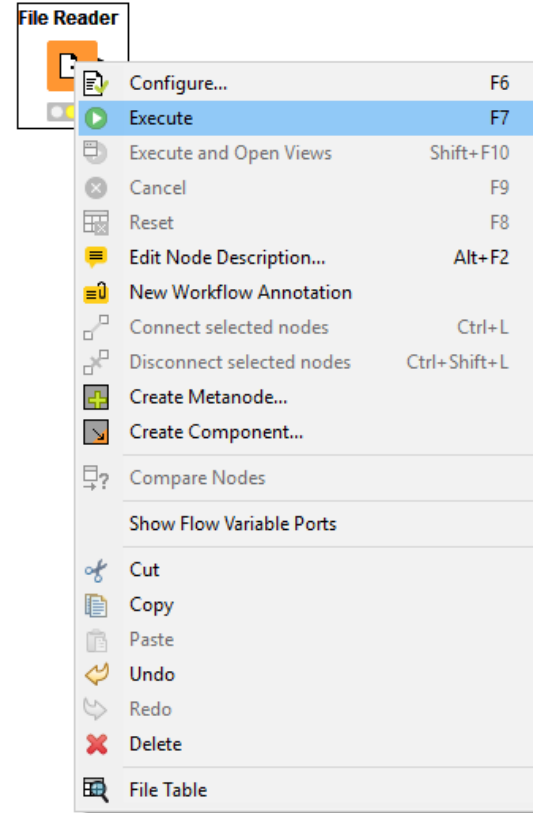
Node Configuration

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

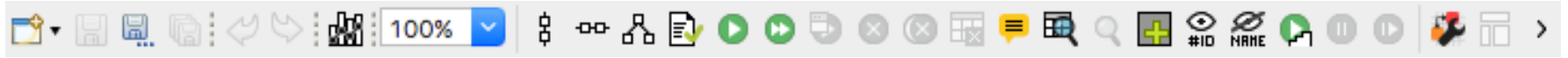


Node Execution






- 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

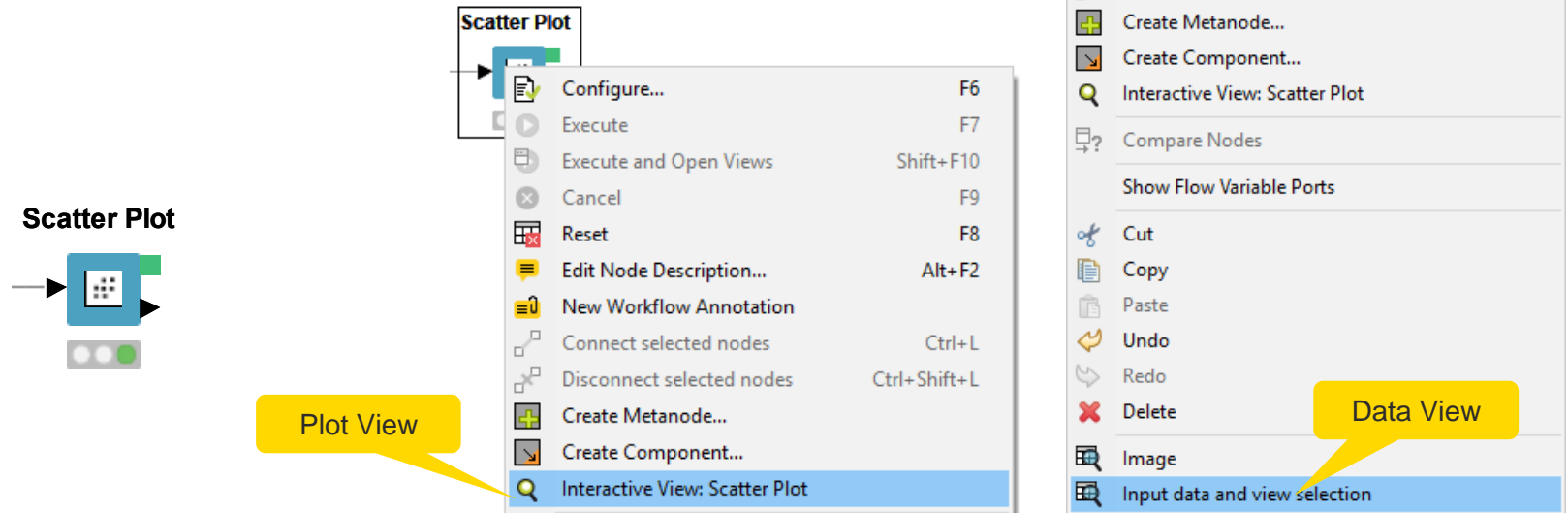


The buttons in the toolbar can be used for the active workflow. The most important buttons:

-  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



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

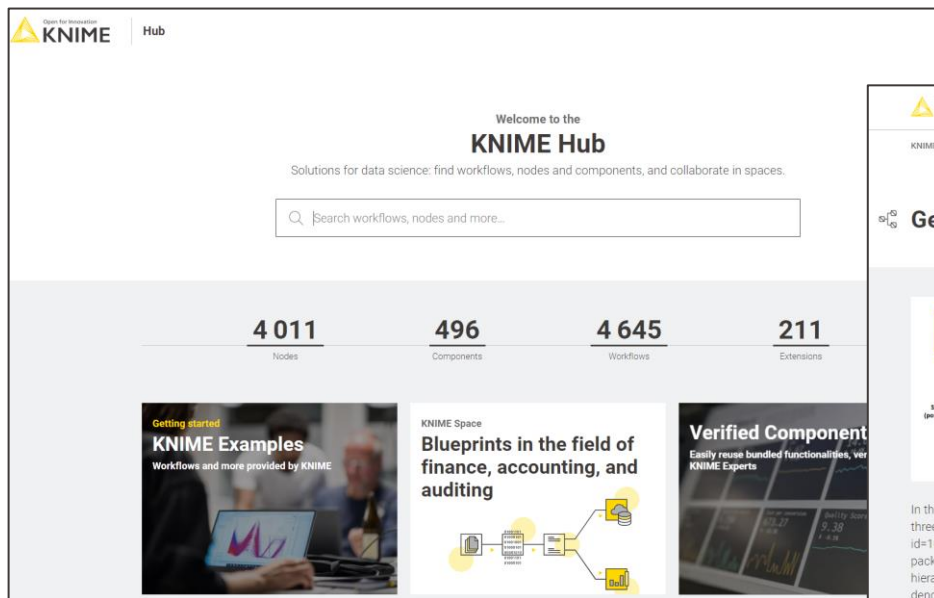
- Connect via KNIME Explorer to a public repository with large selection of example workflows for many, many applications
- Workflows also available on KNIME Hub

The screenshot displays the KNIME Analytics Platform interface. On the left, the 'KNIME Explorer' sidebar shows a tree view of example workflows. The 'EXAMPLES (knime@hub.knime.com)' folder is expanded, showing a hierarchy of folders like '00_Components', '01_Data_Access', '02_ETL_Data_Manipulation', and '02_ETL_Basics'. The '01_Example_for_Standard_Preprocessing' workflow is selected. The main workspace shows a 'Simple Preprocessing Example' workflow diagram. The workflow starts with a 'File Reader' node, followed by a 'Row Filter' node (labeled 'keep rows where "native-country" is missing'). The output of the 'Row Filter' is split into two paths. The top path goes through a 'Column Filter' node, then 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 goes through a 'Reference Row Filter' node (labeled 'Exclude those rows from the original table'), a 'Row Filter' node (labeled 'filter example with a regular expression'), and finally a 'Concatenate' node. A yellow box highlights the 'Simple Preprocessing Example' title and description. A yellow banner at the top of the workspace states: '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

- 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



KNIME Hub

Welcome to the
KNIME Hub

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

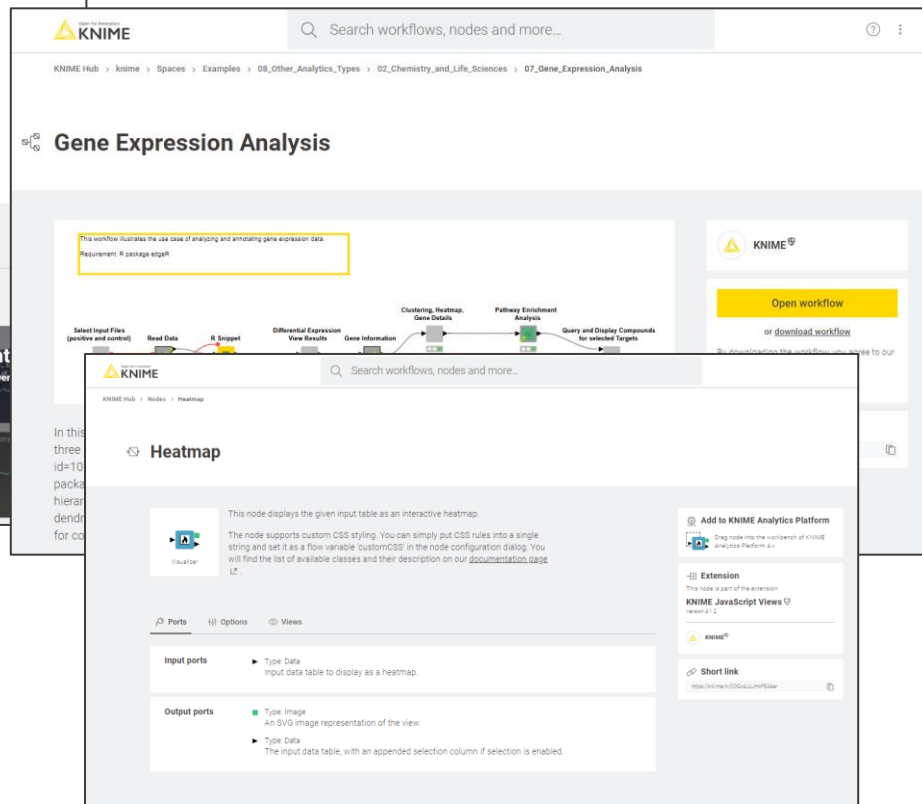
Search workflows, nodes and more...

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



KNIME Hub

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

Search workflows, nodes and more...

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 4.0

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

Short link
[https://knime.com/extension/Heatmap](#)

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

The KNIME Hub

Open for Innovation

?

⋮

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

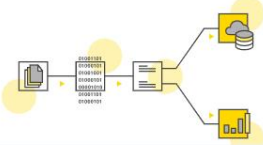
Online courses, workshops, and webinars during April and May



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 and 'Open for Innovation' tagline are on the left, and a search bar with the text 'heatmap' is on the right. Below the search bar, the text 'KNIME Hub > Search' is displayed. The search results section shows '42 results'. Below this, there are tabs for 'All', 'Nodes', 'Components', 'Workflows', and 'Extensions'. The 'Workflows' tab is selected. The search results are listed in a table-like format. The first result is 'Visualization of screening data with HCS-Tools' with a tag 'High-content screening'. The second result, 'Gene Expression Analysis', is highlighted with a dashed yellow box and has tags 'Life Sciences', 'Bioinformatics', 'R', and 'Shared components'. The third result is 'Create Heatmaps in KNIME and use R to add Dendrograms' with tags 'knime', 'heatmap', and 'dendrogram'. The fourth result is 'heatmap_test'. Each result includes a brief description, a path to the workflow, and a small circular icon.


KNIME Hub > Search

42 results

All Nodes Components **Workflows** Extensions

 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.0137111) 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 heatmap dendrogram	It is possible to create heatmaps with dendrograms with the help of 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... mlauber71 > Public > kn_example_r_heatmap_cars	
 heatmap_test	jtyler > Public > heatmap_test	

Opening a Workflow from the Hub



Hub

Search workflows, nodes and more...

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

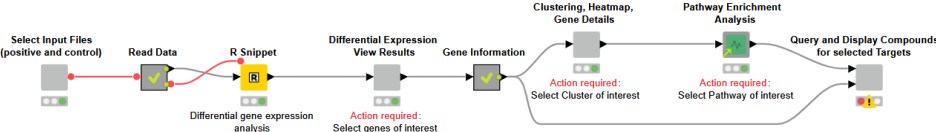
Workflow

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



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 – KNIME Hub

https://hub.knime.com/knime/extension...

Search work

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 → **Rule Engine** → **Math Formula** → **String Manipulation**

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. On the left, the 'KNIME Explorer' pane shows a tree view with 'My-KNIME-Hub (hub.knime.com)' selected. A yellow callout bubble points to this entry with the text '1. Save your Edits'. Below it, a second yellow callout bubble points to the same entry with the text '2. Connect to KNIME Hub'. The 'Node Repository' pane on the left lists various node categories like IO, Manipulation, Views, Analytics, DB, etc. The main workspace shows a workflow titled 'String Manipulation, Math Formula and Rule Engine Example'. A yellow box highlights a text description of the workflow: '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'. Below this text, a diagram illustrates the workflow steps: 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).

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

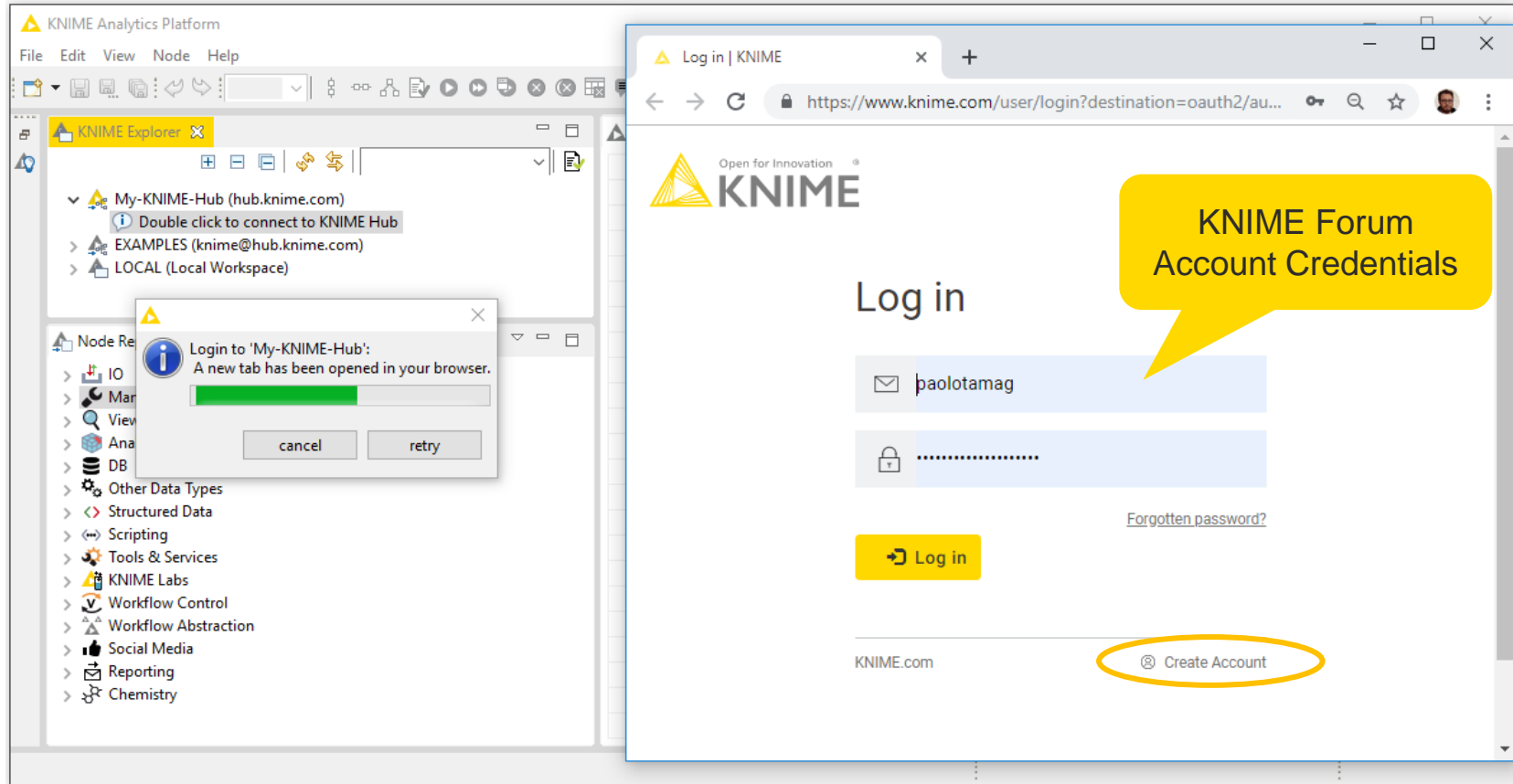
Row Filter
filter data

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

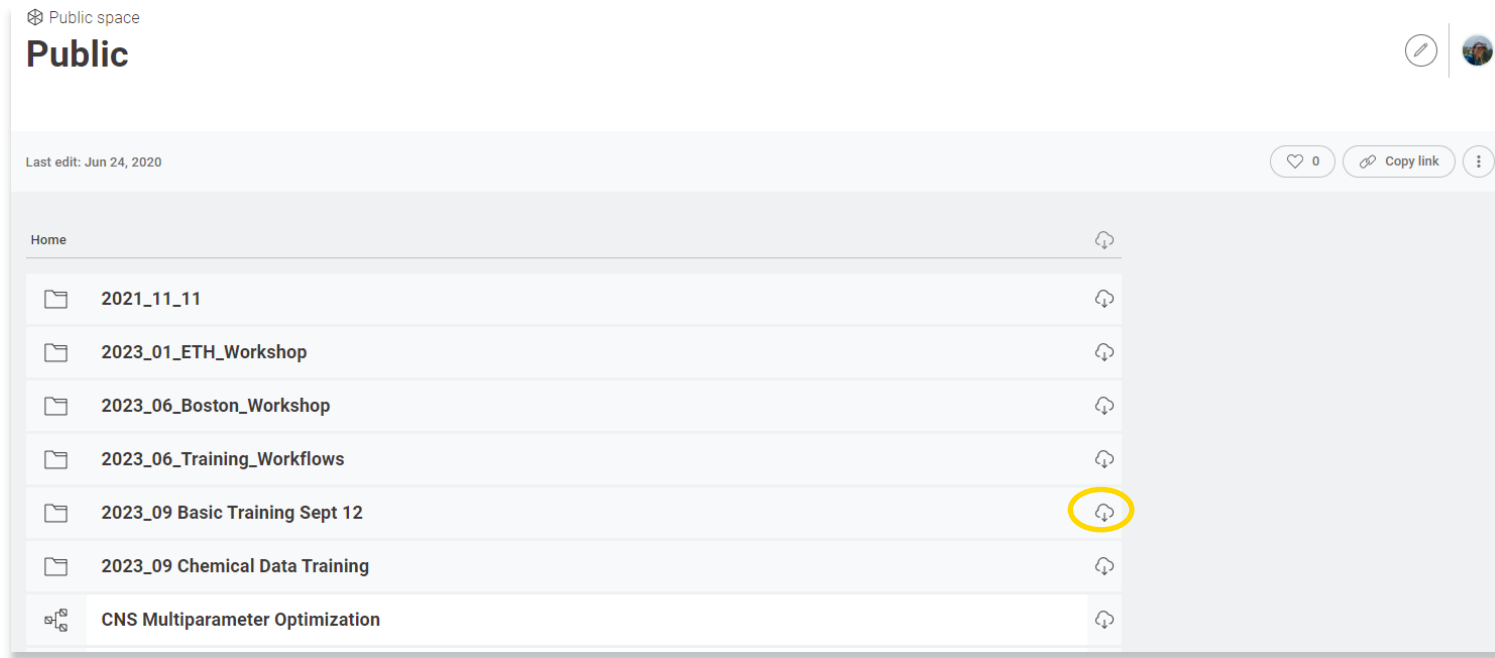
Log in the Hub



Exercises for the training:

- Download the course material from the KNIME Hub

<https://hub.knime.com/-/spaces/-/latest/~0Zau7lq7Cl4a5jXg/>



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 Labs Extensions
- >  KNIME Node Development Tools
- >  KNIME Partner Extensions
- >  KNIME Server Extensions

Selected Extensions from the Life Science Community

Cheminformatics

- ▼ RDKit
 - > Converters
 - > Modifiers
 - > Calculators
 - > Geometry
 - > Fingerprints
 - > Fragments
 - > Searching
 - > Reactions
 - > Viewing
 - > Experimental
 - > Testing
- ▼ 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
- ▼ 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
- ▼ 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.

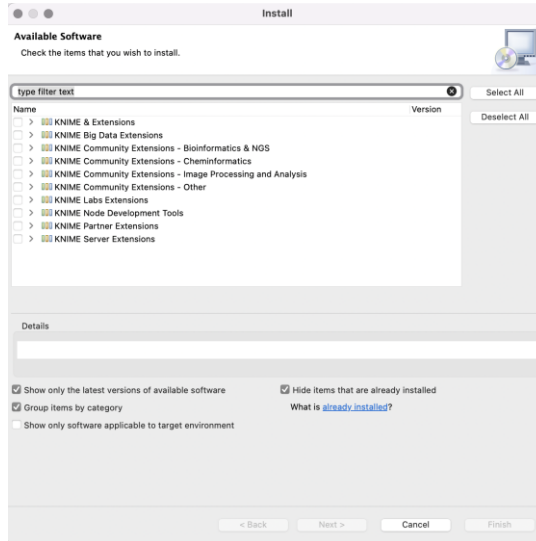
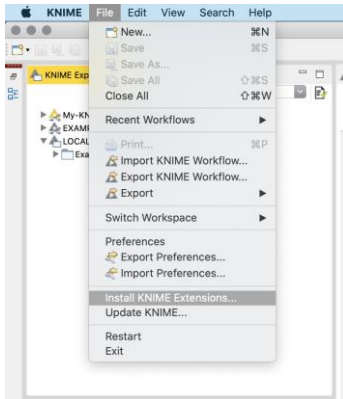


Further extensions including detailed descriptions can be found at <https://hub.knime.com>

Installing KNIME Extensions

There are different ways to install KNIME extensions:

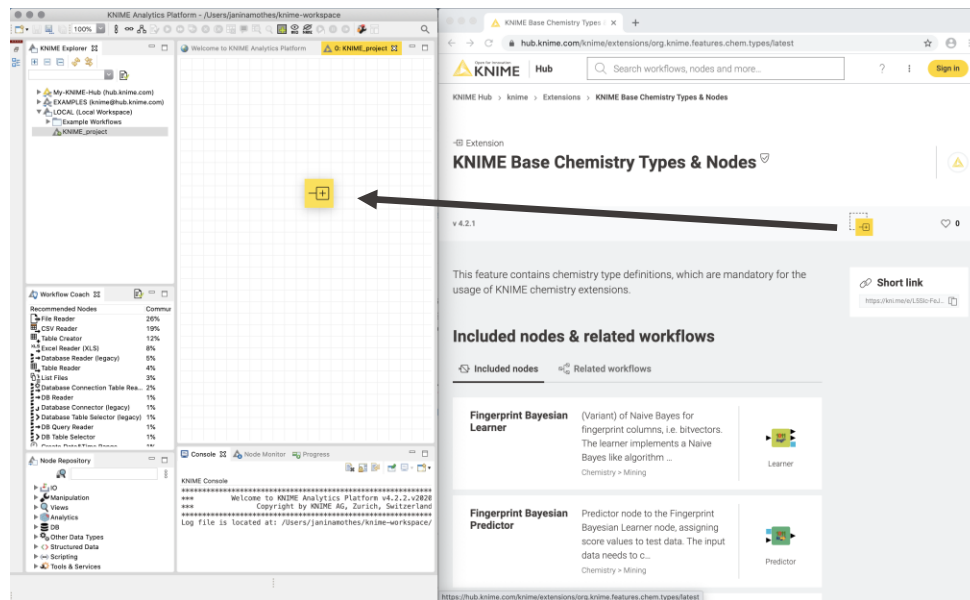
Option 1: Via the menu bar in the KNIME Analytics Platform and select your extension from the list (feasible if you know the name of the extension)



Installing KNIME Extensions

There are different ways to install KNIME extensions:

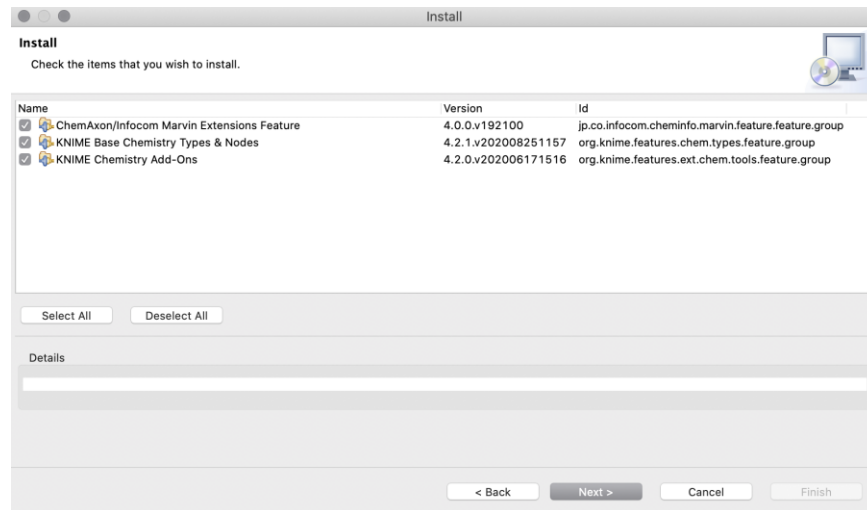
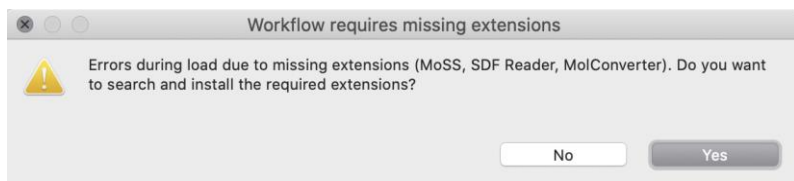
Option 2: Via the KNIME Hub (hub.knime.com): Drag-and-drop the required extension (or node) to your workflow



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)



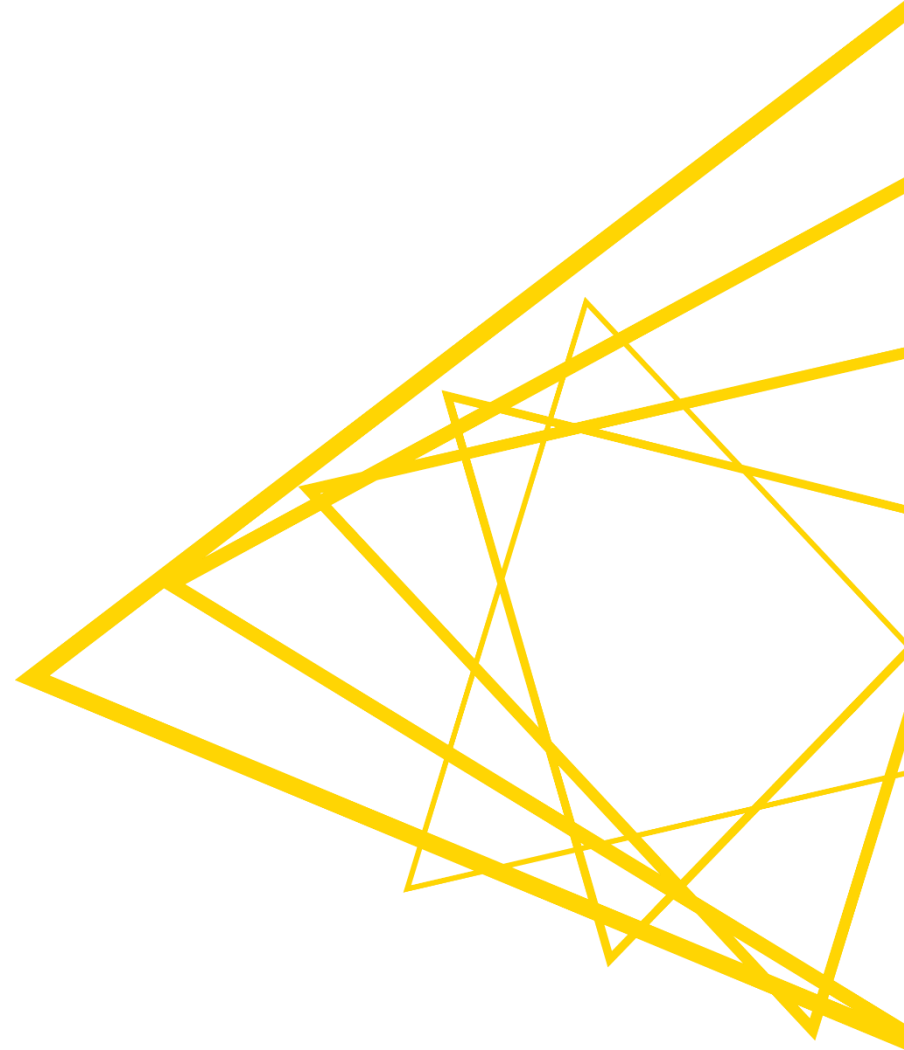
Installing Extensions Exercise

Open Exercise: *00. Installing Extensions*

Activity 1: Install the following Extensions

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

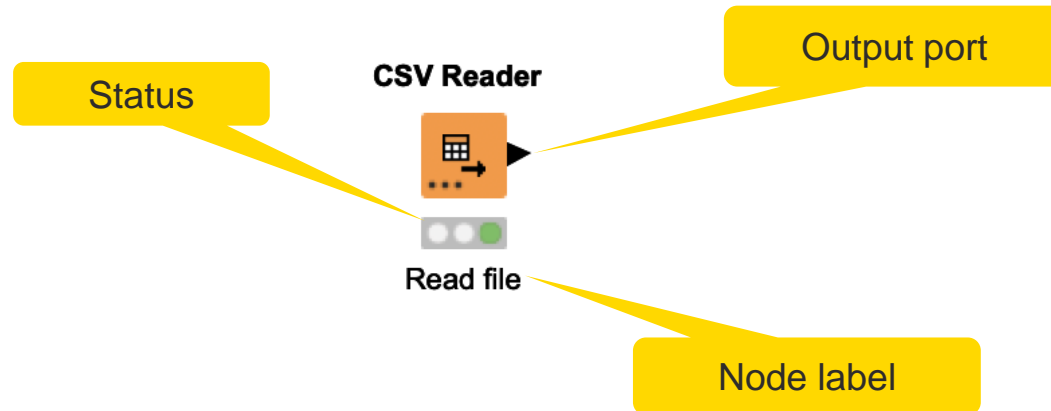
Importing Data



Data Source Nodes

Typically characterized by:

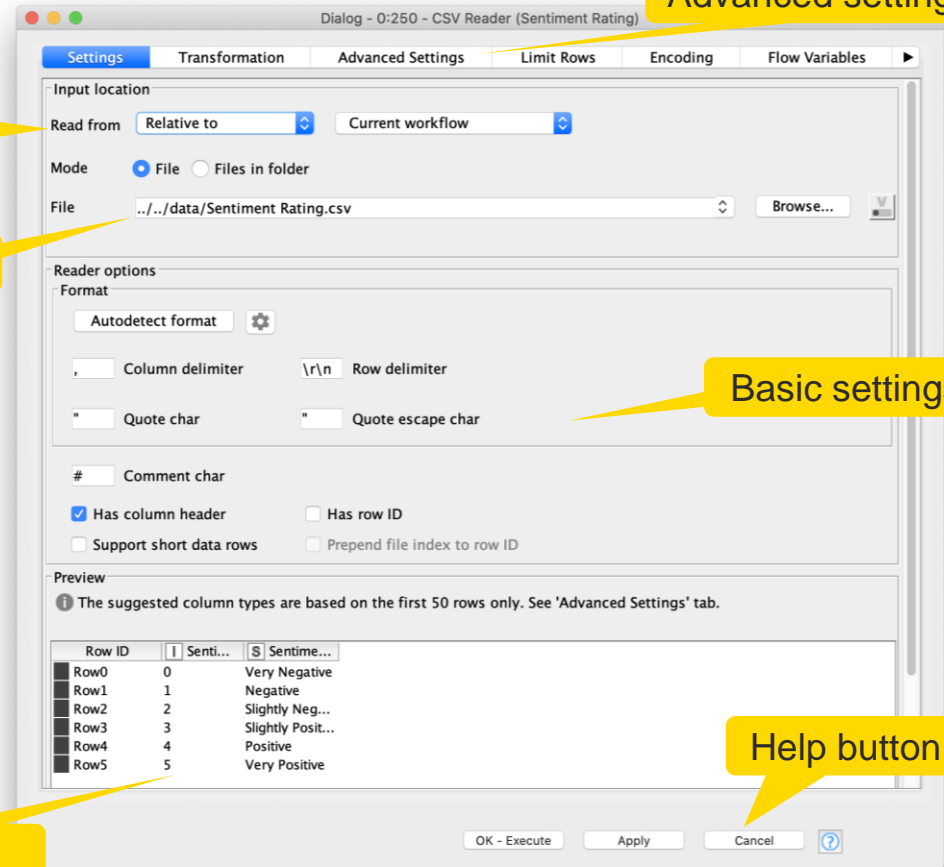
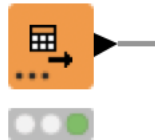
- Orange color
- By default no input ports, 1-2 output ports
- New file handling with KNIME 4.3.
 - Consistent user experience across all nodes and file systems
 - Managing of various file systems within the same workflow
 - Performance improvements



CSV Reader

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

CSV Reader




Common Settings: File Path Options

■ Local File System

Input location

Read from:

Mode: ☒ File ☐ Files in folder

File: 


■ Relative to ...

Read from:

Current mountpoint


Current workflow data area

Current workflow

File: 


■ Mountpoint

Read from:

File: 

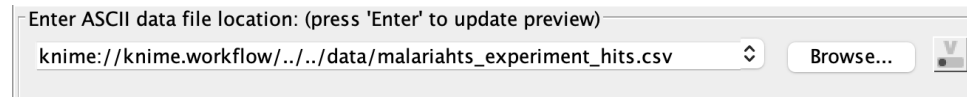
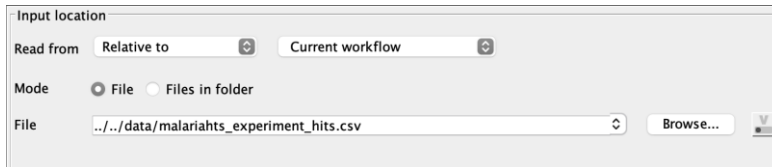
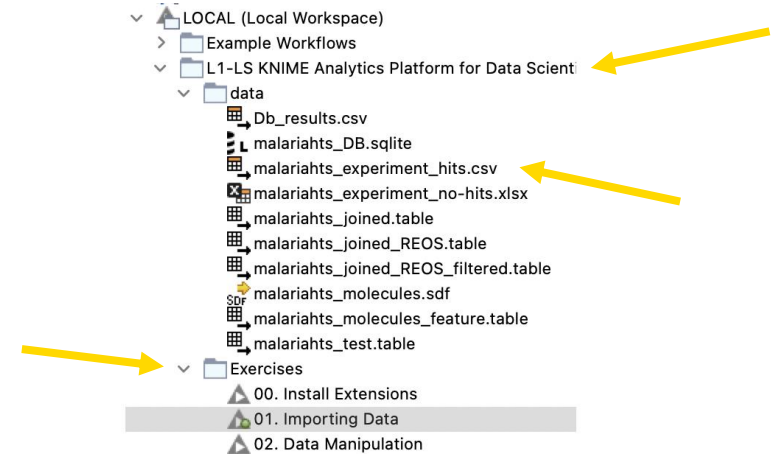
■ Custom URL

Read from:

URL: 

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:



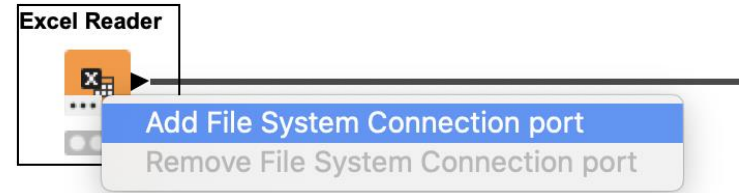
YouTube KNIME TV Channel:

<https://youtu.be/liZsOnhZgzk>

Common Settings: Connecting to other File Systems

- Add file system connection port to connect to another file system

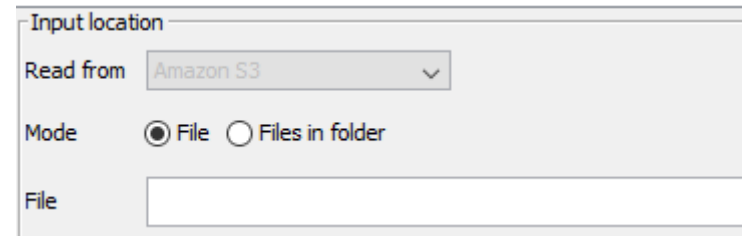
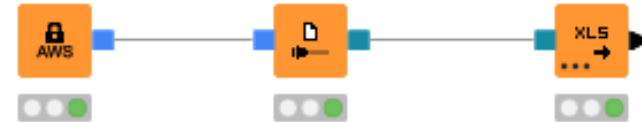
- Click on the three dots on the lower left to add or remove a dynamic port.



- Supported file systems

- Microsoft Azure
- Google
- Amazon
- Databricks
- BigData file systems (hdfs, httpFS, ...)
- On-premise (e.g. ssh, ftp, ...)

Amazon
Authentication Amazon S3 Connector Excel Reader (XLS)



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 criteria

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

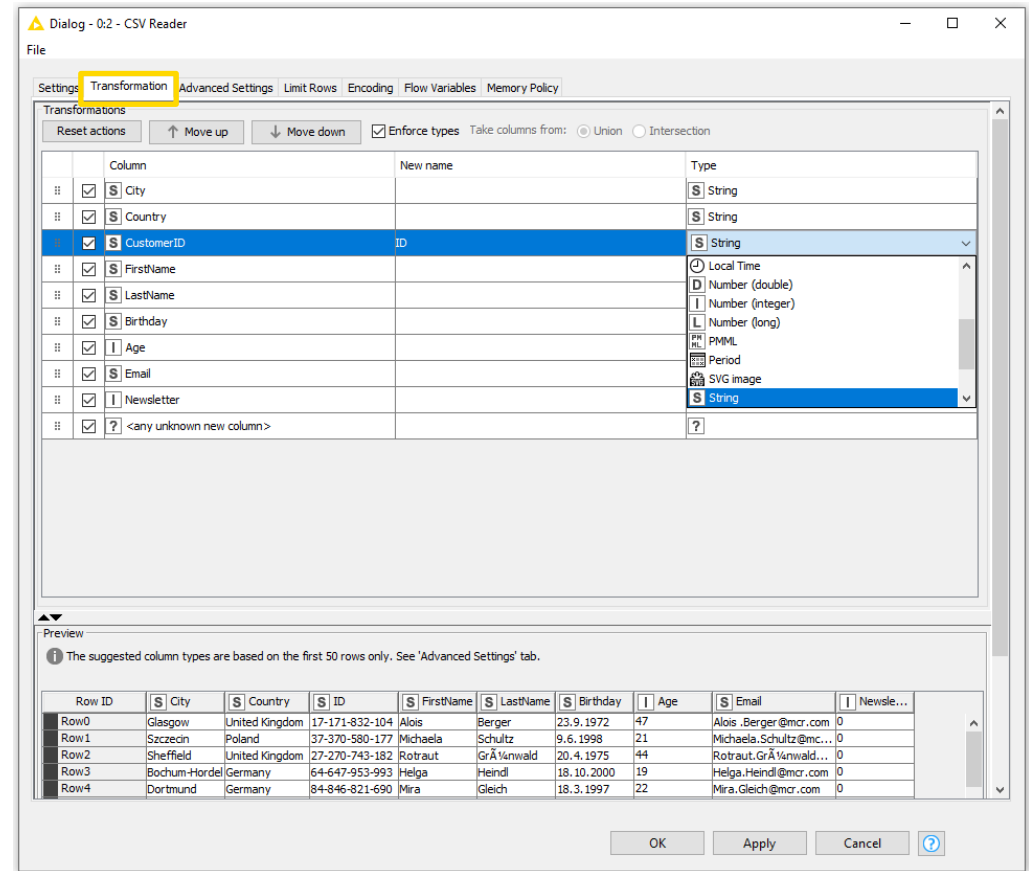
☐ Case sensitive ☒ Wildcard ☐ Regular expression

☐ Include hidden folders

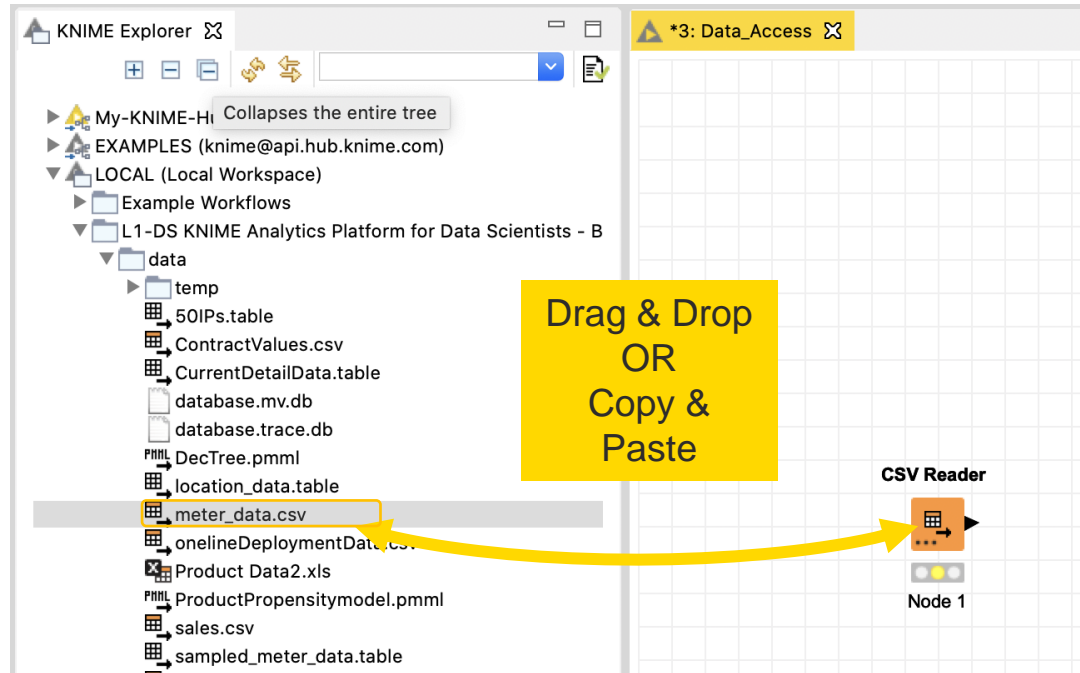
OK Cancel

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)**



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



Read Excel
Sheet Names



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	Custom...	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

Dialog - 0:7 - Table Reader

Settings Transformation Advanced Settings Flow Variables Job Manager Selection Memory Policy

Input location

Read from Relative to Current workflow

Mode ☒ File ☐ Files in folder

File ../../data/malariahts_molecules_feature.table Browse...

Row ID handling

☐ Use existing row ID ☐ Prepend table index to row ID

Preview

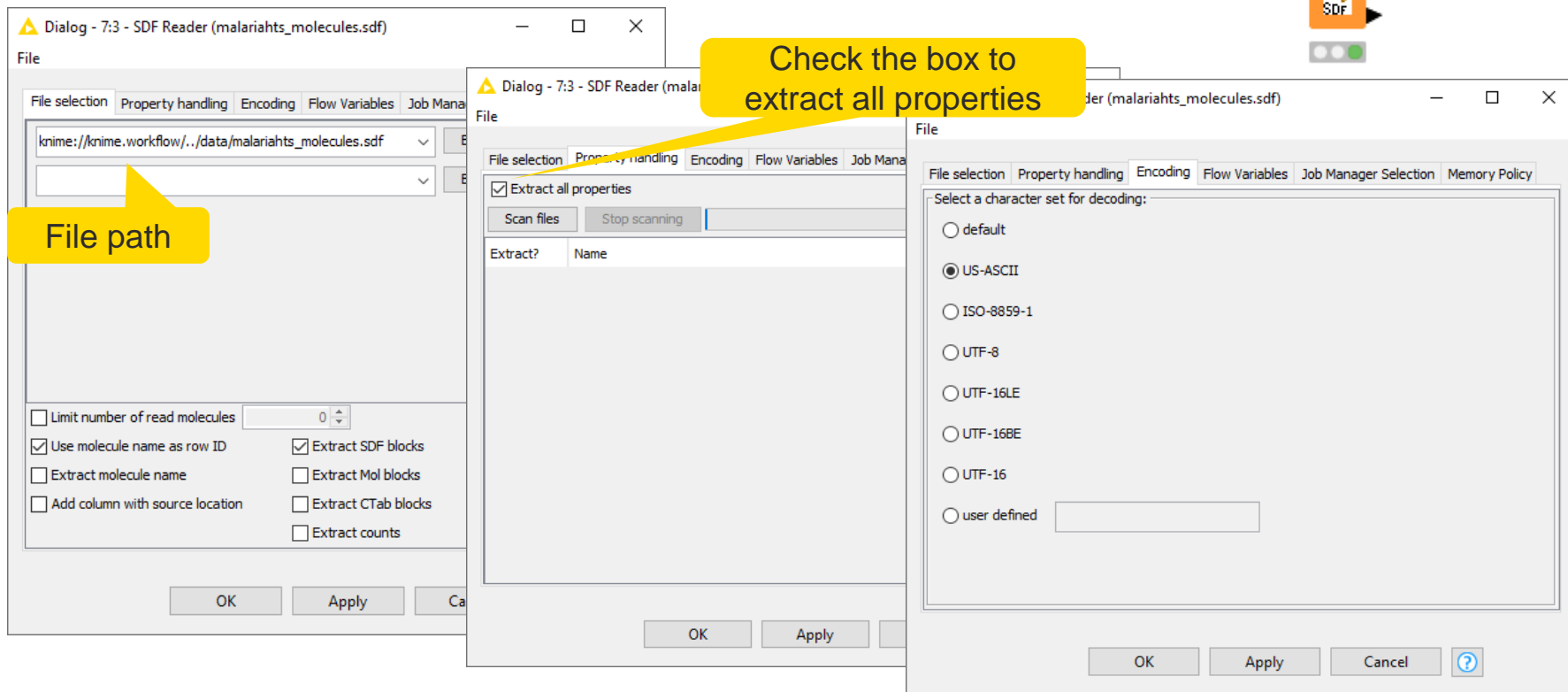
Row ID	Sample	NumR...	NumH...	NumH...	NumA...	NumH...	NumH...	NumA...	Nur
Row0	Sj000263455-1 5	0	7	0	8	25	38	0	
Row1	Sj000136517-1 5	1	6	1	7	20	33	0	
Row2	Sj000244874-1 5	1	6	2	8	33	56	0	
Row3	Sj000185650-1 7	0	9	0	10	27	43	0	
Row4	Sj000295142-1 2	0	3	3	5	21	39	0	
Row5	Sj000007375-1 5	1	4	1	6	28	45	0	
Row6	Sj000243289-1 6	2	6	0	7	24	45	0	
Row7	Sj000116582-1 4	1	4	1	8	30	52	1	
Row8	Sj000293545-1 3	2	4	1	7	27	39	0	
Row9	Sj000057137-1 6	1	4	0	7	25	42	0	
Row10	Sj000079475-1 4	1	7	1	10	28	40	0	
Row11	Sj000038023-1 5	2	4	2	10	27	46	1	
Row12	Sj000053377-1 2	0	5	1	7	24	36	0	
Row13	Sj000226127-1 6	0	4	1	6	25	45	1	
Row14	Sj000259962-1 4	0	7	0	8	27	42	0	
Row15	Sj000110512-1 5	1	5	1	10	34	54	0	
Row16	Sj000300530-1 3	1	5	1	6	20	30	0	

OK Apply Cancel

File path

SDF Reader


- Reads chemical data from the .sdf file



required Extension: KNIME Base Chemistry Types & Nodes

Table Creator

Table Creator



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

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 ?

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

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



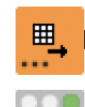
SDF Reader



Excel Reader

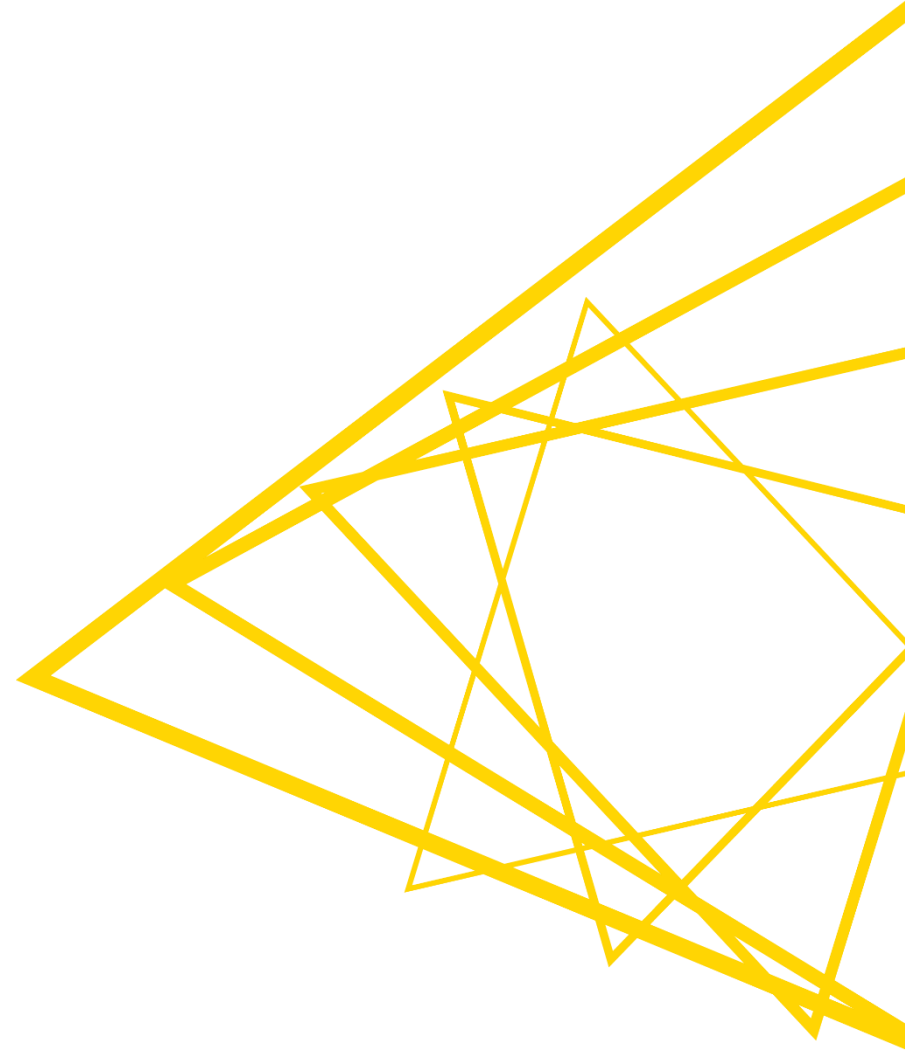


Table Reader



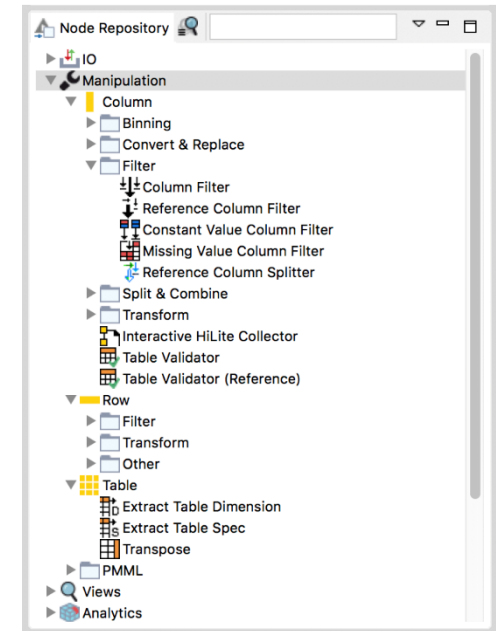
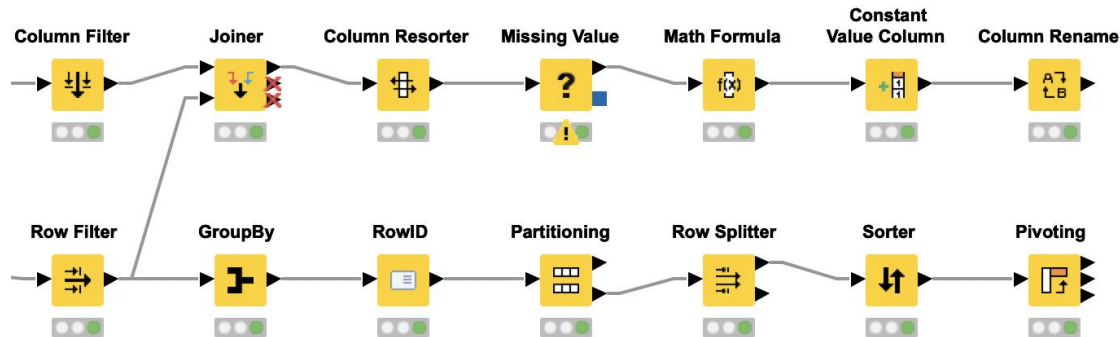
Data Manipulation

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!



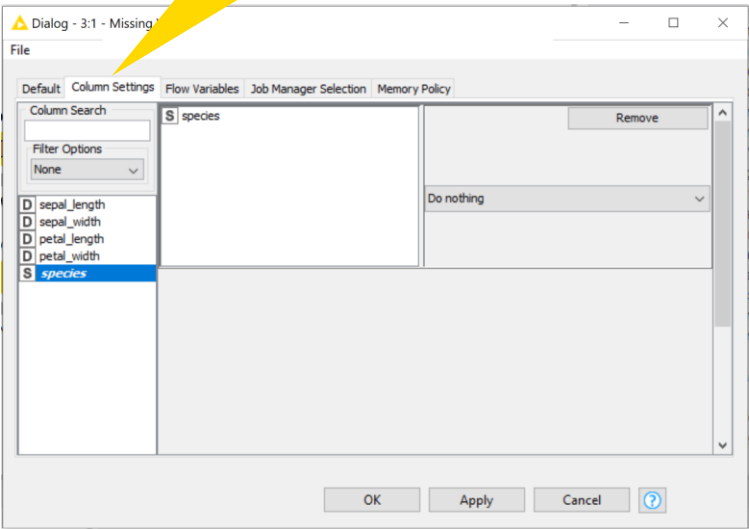
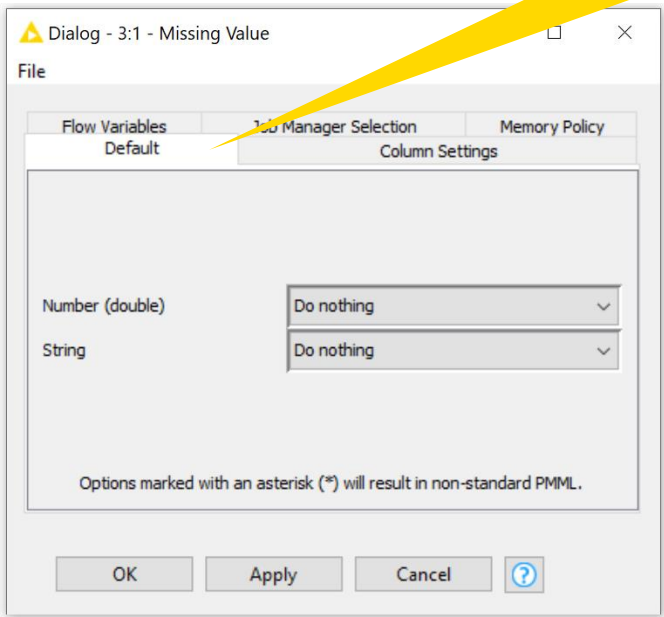
Missing Value

helps to handle missing values

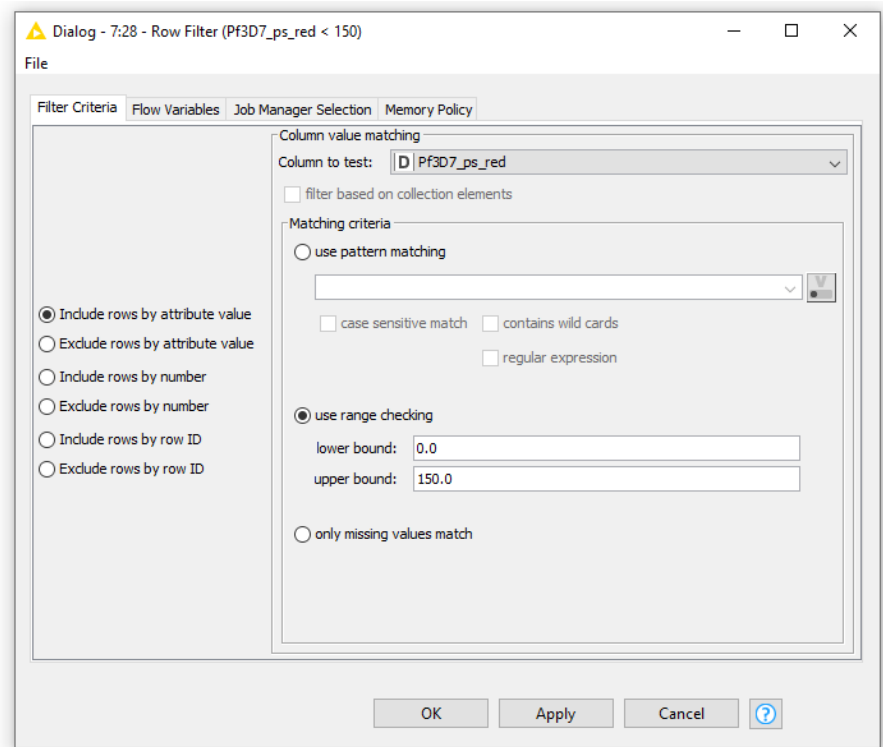
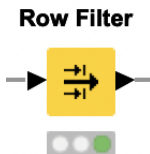


Define default values for all columns of specific type

Define values for every single column



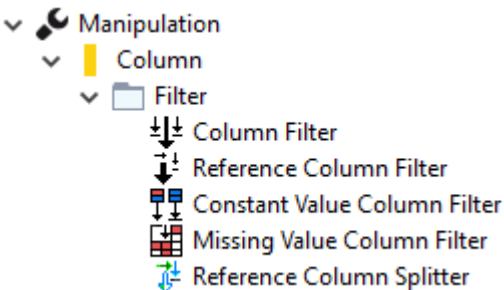
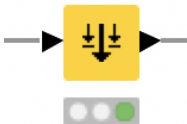
Row Filter



- ▼ 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



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



Dialog - 2:295 - Sorter

File

Sorting Filter Flow Variables Job Manager Selection Memory Policy

Sort by:

S - ROWKEY - Ascending
Descending

Next by:

? - DO NOT SORT - Ascending
Descending

Next by:

? - DO NOT SORT - Ascending
Descending

Add columns

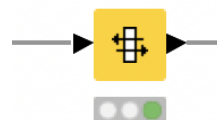
1 new columns

☐ Sort in memory

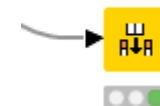
☐ Move Missing Cells to end of sorted list

OK Apply Cancel ?

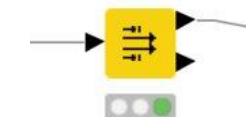
Column Resorter



Column Splitter



Row Splitter



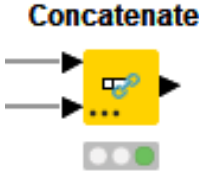
Concatenate Tables

Table A

RowID	Mol Reg No	Chembl ID	Ki value
0	35	CHEMBL15435	100.0
1	15	CHEMBL1794855	8.0

Table B

RowID	Mol Reg No	Chembl ID
0	15	CHEMBL1794855
1	10	CHEMBL278751
2	22	CHEMBL103772



union of columns

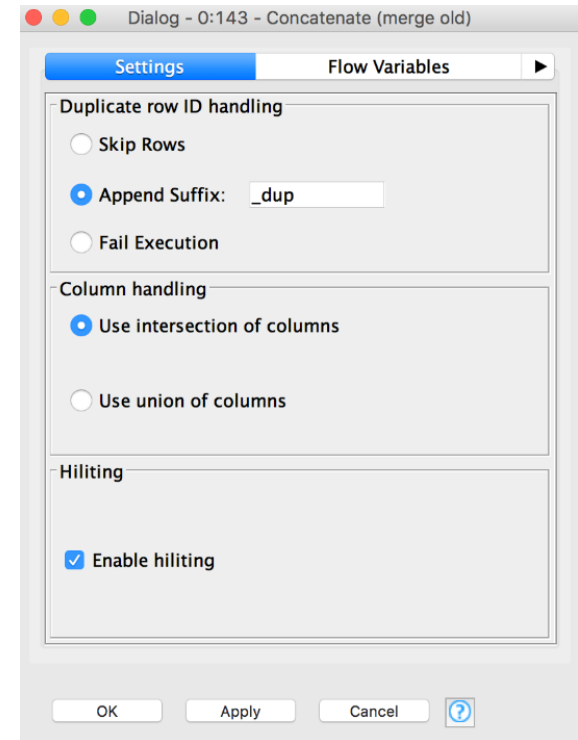
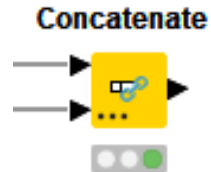
RowID	Mol Reg No	Chembl ID	Ki value
0	35	CHEMBL15435	100.0
1	15	CHEMBL1794855	8.0
0_dup	15	CHEMBL1794855	
1_dup	10	CHEMBL278751	
2	22	CHEMBL103772	

intersection of columns

Concatenate

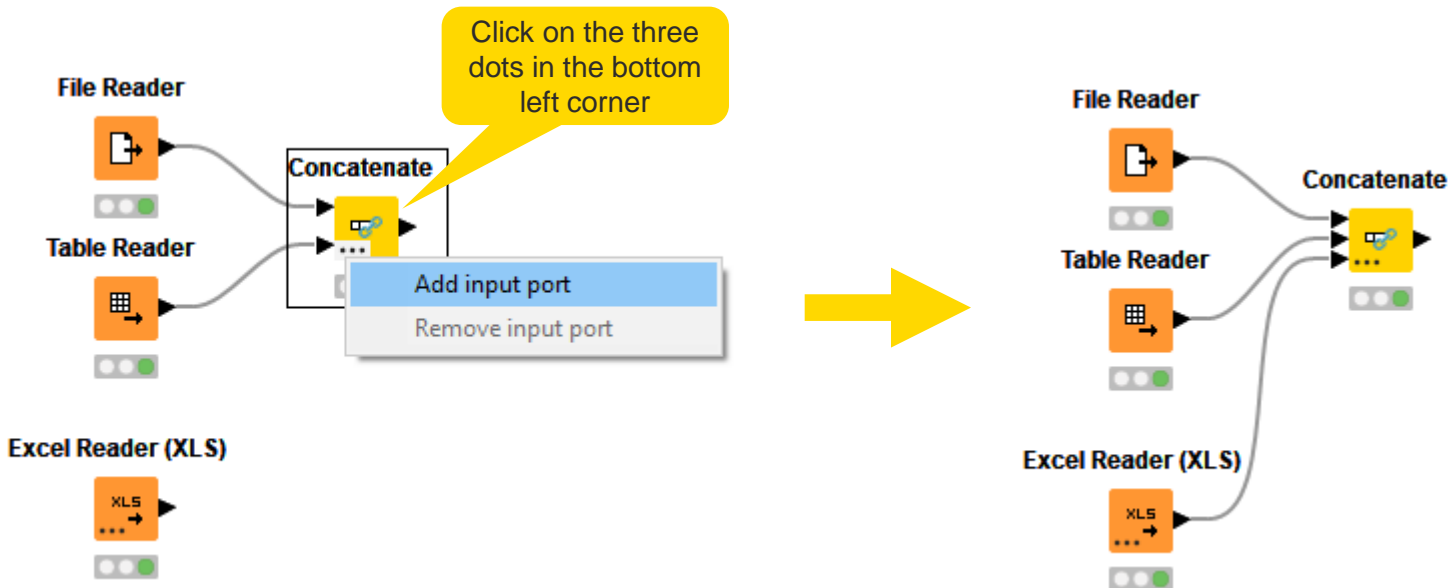
Combine rows from 2 or more tables with shared columns

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



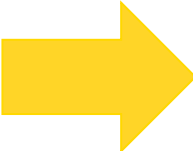
Dynamic Ports

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

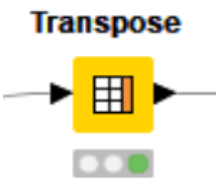


Transpose

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

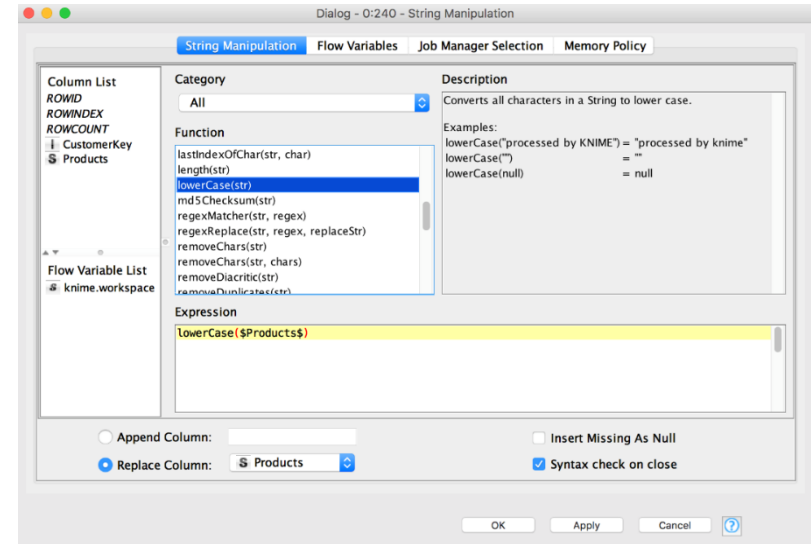
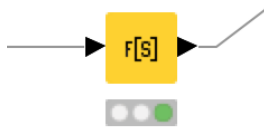
?

String Manipulation

Create and edit values in String columns

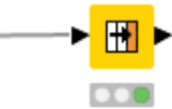
- Clean up capitalization (e.g. Lowercase)
- Replace strings
- Modify existing strings or create new columns

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

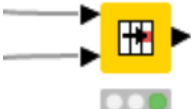
Case sensitive search: ☒

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

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

Left 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

Right Outer Join

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

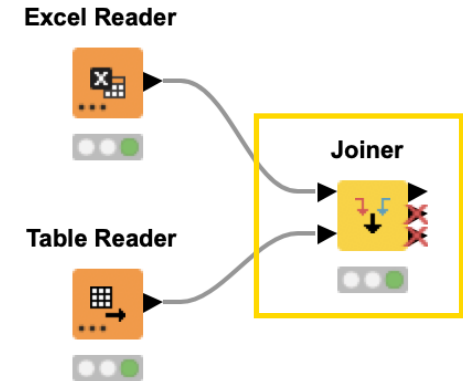
Mol Reg No	Chembl ID	SMILES	Ki value	Ki relation	Ki unit
17	?	?	76.0	=	nM
65	?	?	6.56	=	nM
35	?	?	100	>	nM
15	CHEMBL1794855	CCCN(CCC)	8	=	nM
10	CHEMBL278751	CCN(C)	95.8	=	nM
22	CHEMBL103772	CCCN1CC	?	?	?
24	CHEMBL328107	C1CN(CCN1)	?	?	?

Missing values in the left table

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 'Dialog - 3:8 - Joiner' window. The 'Joiner Settings' tab is active. The 'Join columns' section is highlighted with a yellow box. It contains two dropdown menus, 'Top Input (left' table)' and 'Bottom Input (right' table)', both set to 'Sample'. Below these are buttons for adding (+) and removing (-) columns. The 'Compare values in join columns by' section has three radio buttons: 'value and type' (selected), 'string representation', and 'making integer types compatible'. The 'Include in output' section has three checkboxes: 'Matching rows' (checked), 'Left unmatched rows', and 'Right unmatched rows'. To the right of these is a Venn diagram labeled 'Inner join' showing the intersection of two sets. The 'Output options' section has three checkboxes: 'Split join result into multiple tables (top = matching rows, middle = left unmatched rows, bottom = right unmatched rows)' (checked), 'Merge join columns', and 'Hilting enabled'. At the bottom are 'OK', 'Apply', 'Cancel', and a help icon.

Dialog - 3:8 - Joiner

File

Joiner Settings | Column Selection | Performance | Flow Variables | Job Manager Selection | Memory Policy

Join columns

Match ☒ all of the following ☐ any of the following

Top Input (left' table) Bottom Input (right' table)

Sample Sample + -

+

Compare values in join columns by ☒ value and type ☐ string representation ☐ making integer types compatible

Include in output

☒ Matching rows

☐ Left unmatched rows

☐ Right unmatched rows

Inner join

Output options

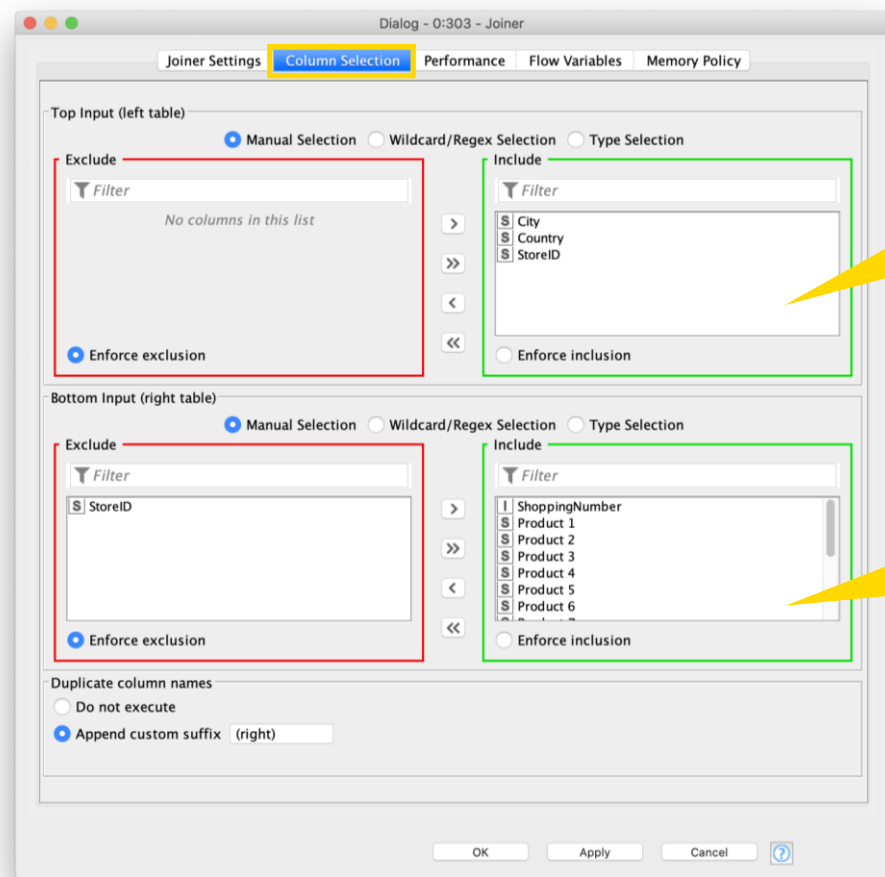
☒ Split join result into multiple tables (top = matching rows, middle = left unmatched rows, bottom = right unmatched rows)

☐ Merge join columns

☐ Hilting enabled

OK Apply Cancel ?

Joiner Configuration – Column Selection



Columns from top table for joined table

Columns from lower table for joined table

Data Aggregation (GroupBy)

Type	Name	Weigt
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



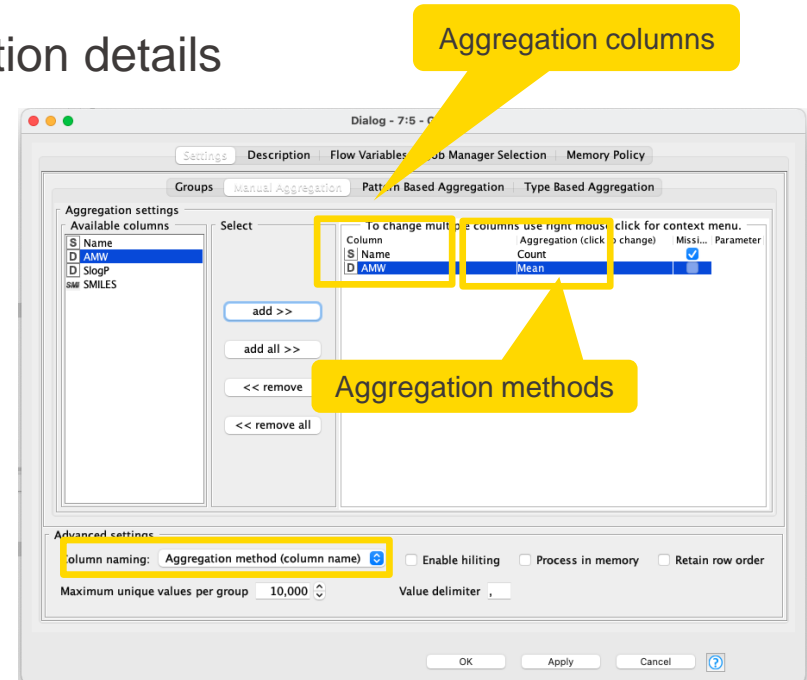
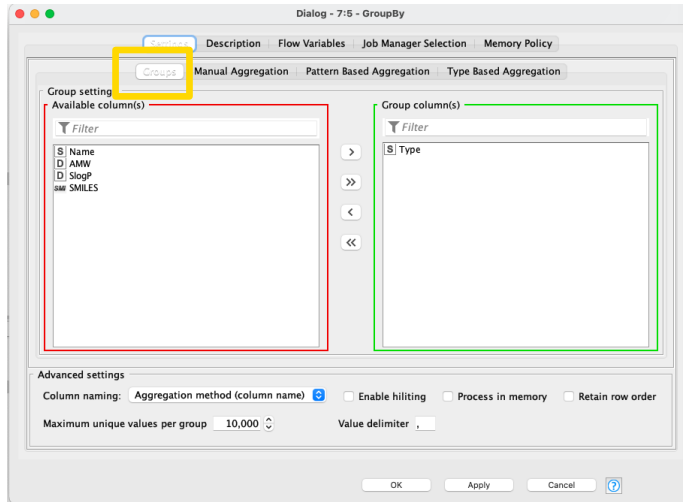
Type	Count(Name)	Mean(Weight)
NSAID	4	208.44
PPI	2	364.40
SSRI	4	326.45

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



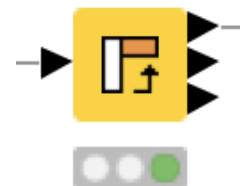
Type	Acute toxic	Health hazard	Irritant
NSAIDs	1	1	2
PPIs	?	?	2
SSRIs	1	3	?

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



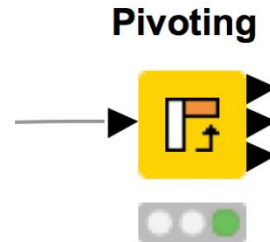
Row ID	S Type	I acute toxic+Count(Name)	I health hazard+Count(Name)	I irritant+Count(Name)
Row0	NSAID	1	1	2
Row1	PPIs	?	?	2
Row2	SSRIs	1	3	?

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

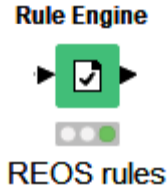
Pivots ~ Columns

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



Dialog - 7:48 - Rule Engine (REOS rules)

File

Rule Editor | Flow Variables | Job Manager Selection | Memory Policy

Column List

- ROWID
- ROWINDEX
- ROWCOUNT
- [S] Sample
- [D] Pf307_ps_green
- [D] Pf307_ps_red
- [S] Pf307_ps_hit
- [D] Pf307_pEC50
- ssr Molecule
- [D] SlogP
- [D] TPSA
- [D] AMW
- [D] ExactMW
- [I] NumRotatableBonds
- [I] NumHBD
- [I] NumHBA
- [I] NumAmideBonds
- [I] NumHeteroAtoms
- [I] NumHeavyAtoms
- [I] NumAtoms
- [I] NumStereocenters

Category

All

Description

Function

- ? < ?
- ? <= ?
- ? = ?
- ? > ?
- ? >= ?
- ? AND ?
- ? IN ?
- ? LIKE ?
- ? MATCHES ?
- ? OR ?
- ? XOR ?
- FALSE
- MISSING ?
- NOT ?
- TRUE

Flow Variable List

- [S] knime.workspace

Expression

```
1 // enter ordered set of rules, e.g.:
2 // $double column name$ > 5.0 => "Large"
3 // $string column name$ LIKE "*blue*" => "small and blue"
4 // TRUE => "default outcome"
5 $AMW$<100 OR $AMW$>700 => "MW"
6 $NumHeavyAtoms$<5 OR $NumHeavyAtoms$>50 OR $NumRotatableBonds$ >=12 => "Complexity"
7 $NumHBD$>5 OR $NumHBA$>10 => "HBond"
8 $SlogP$<-5 OR $SlogP$>7.5 => "logP"
9 TRUE=>"Pass"
```

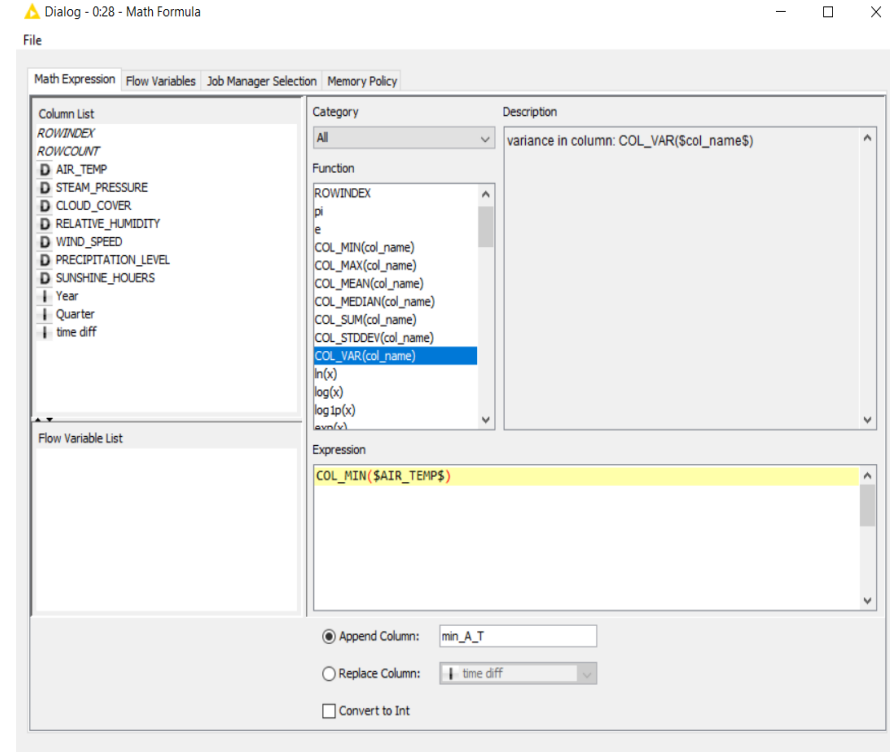
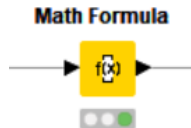
☒ Append Column: REOS [S]

☐ Replace Column: NumHeavyAtoms

OK Apply Cancel ?

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

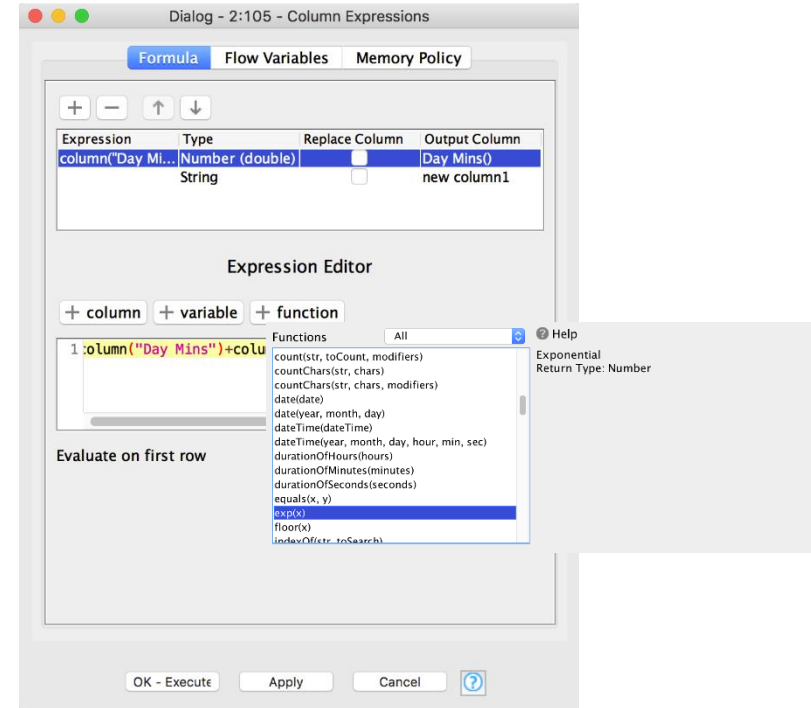
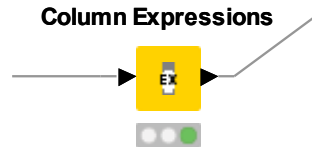
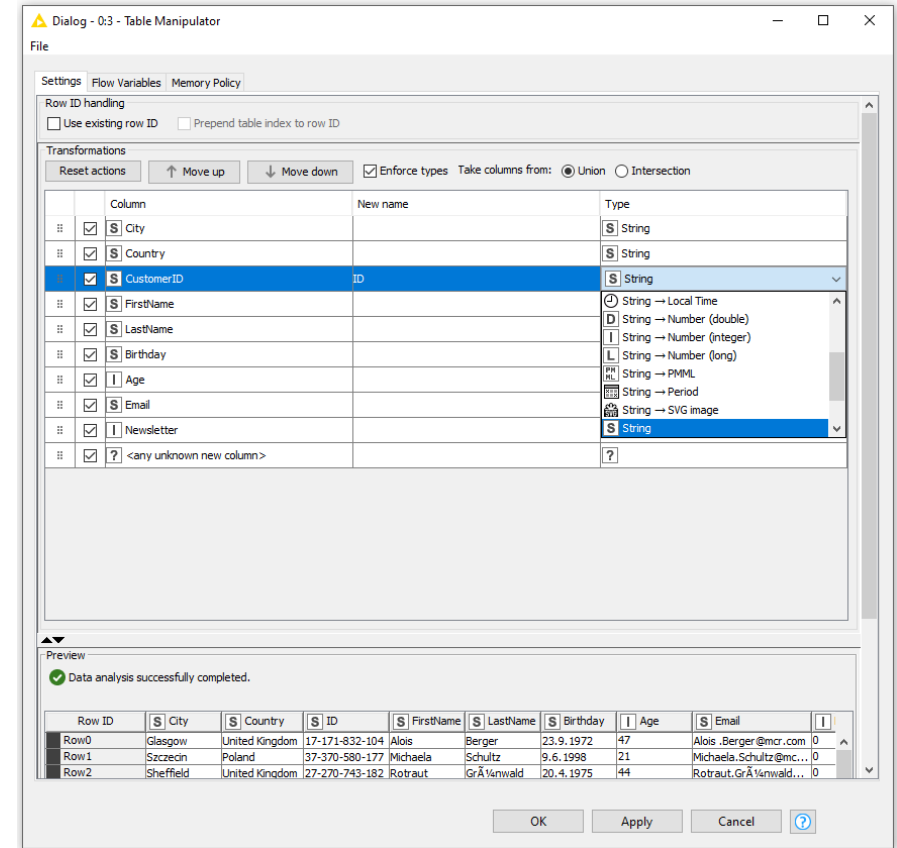
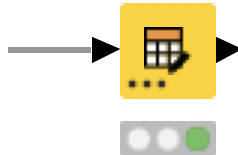


Table Manipulator

Allows for

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

Table Manipulator

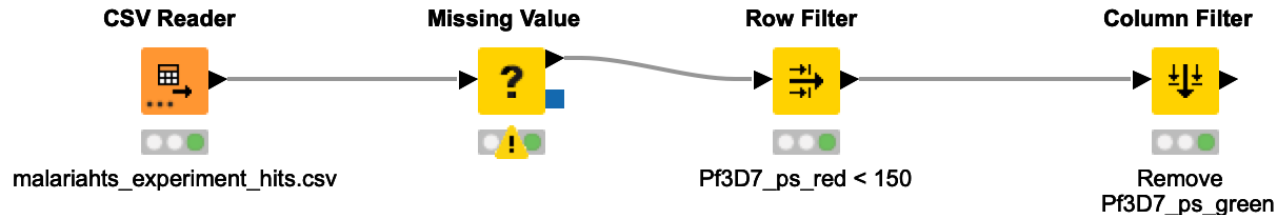


Data Manipulation Exercise

Open Exercise 02. Data Manipulation

Activity 1: 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

Data Manipulation Exercises

Open Exercise 02. *Data Manipulation*

Activity I: Filtering

- Remove rows missing values, filter rows and columns

Activity II: Data Manipulation & Aggregation

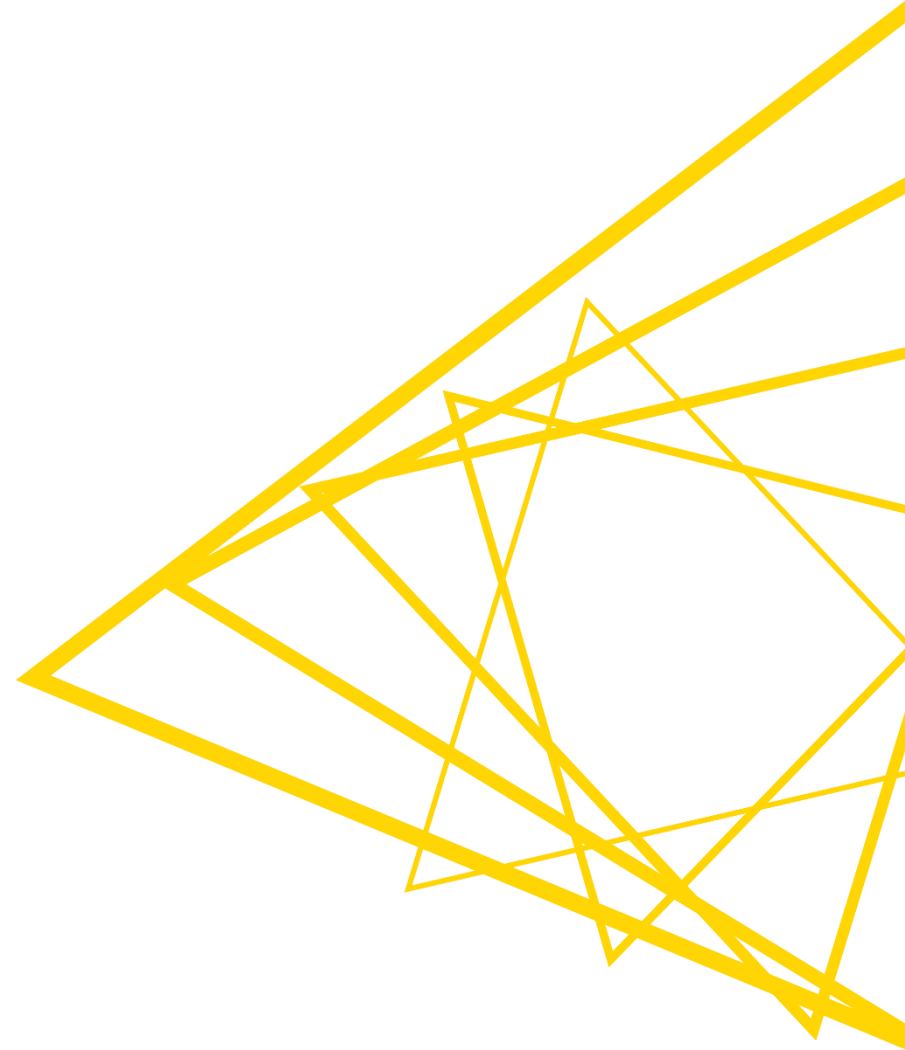
- Concatenate, Join and Manipulate the data according to the instructions in the annotations

Optional: Activity III: Data Manipulation

- Use the Rule Engine node and Table Manipulator node

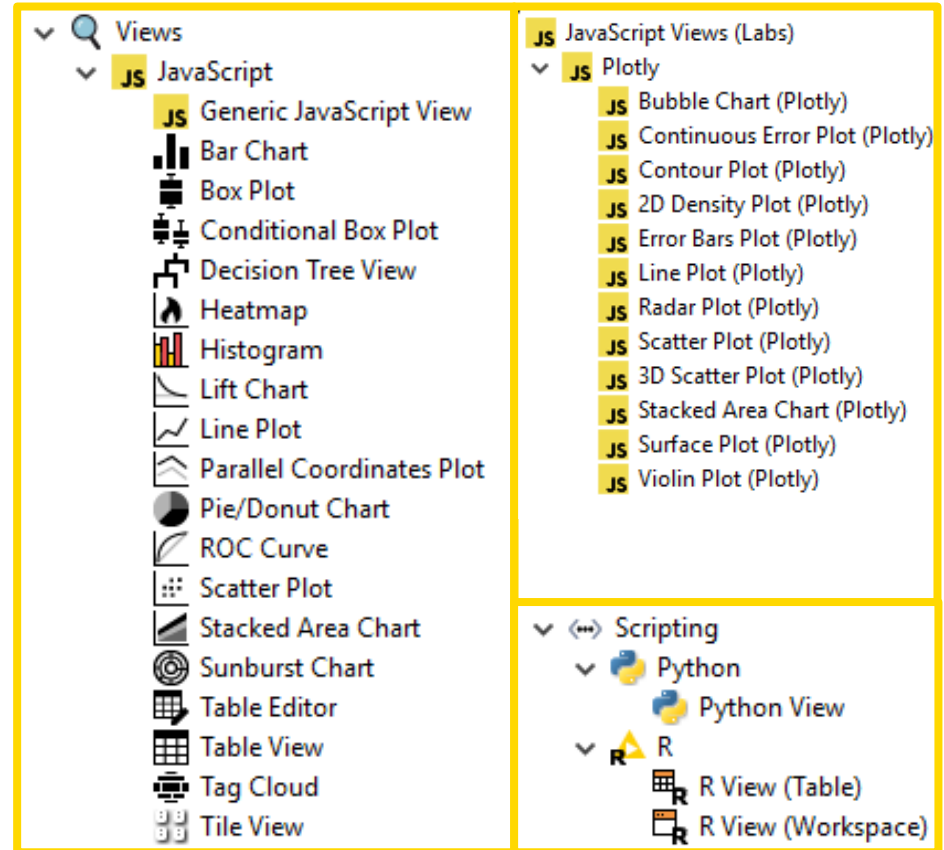
Data Visualization

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



© 2023 KNIME AG. All rights reserved.

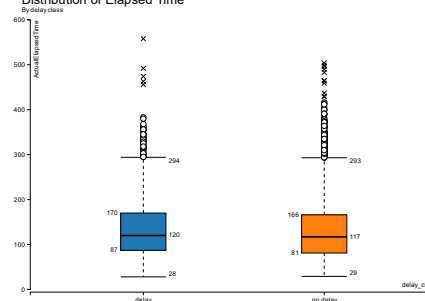
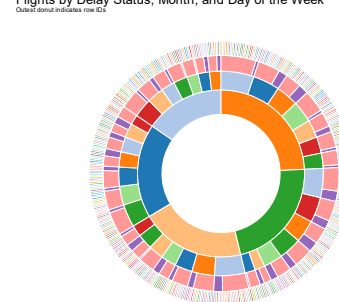
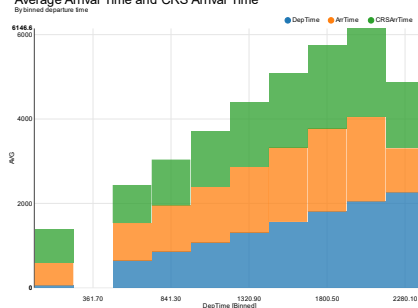
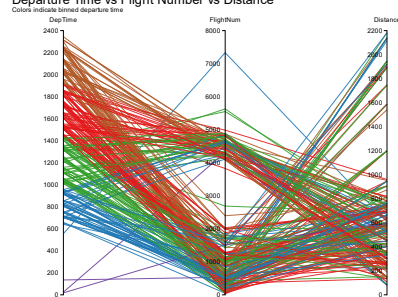
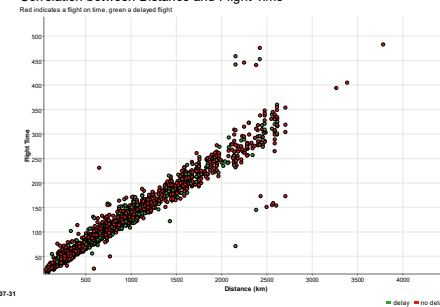
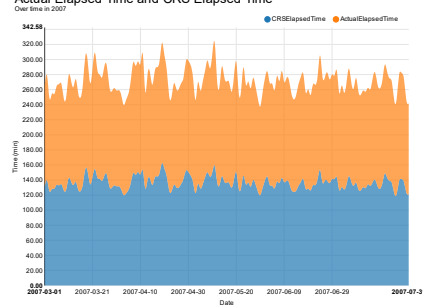
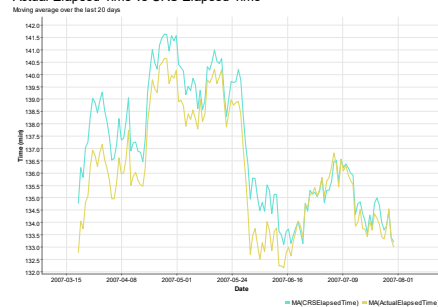
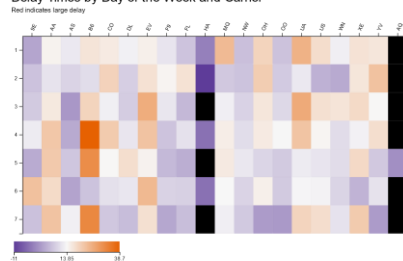
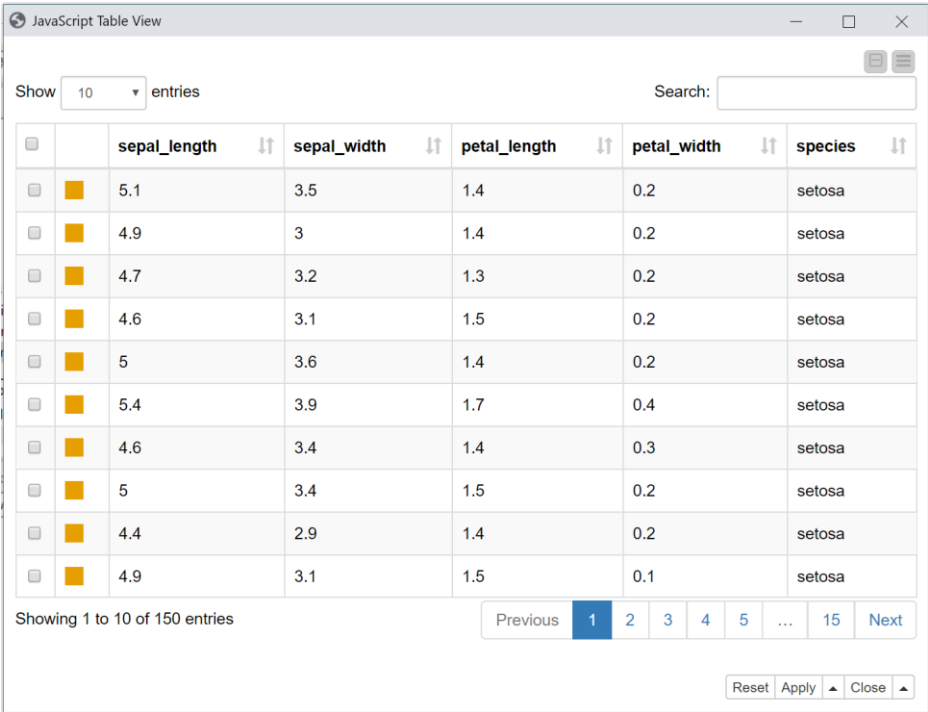
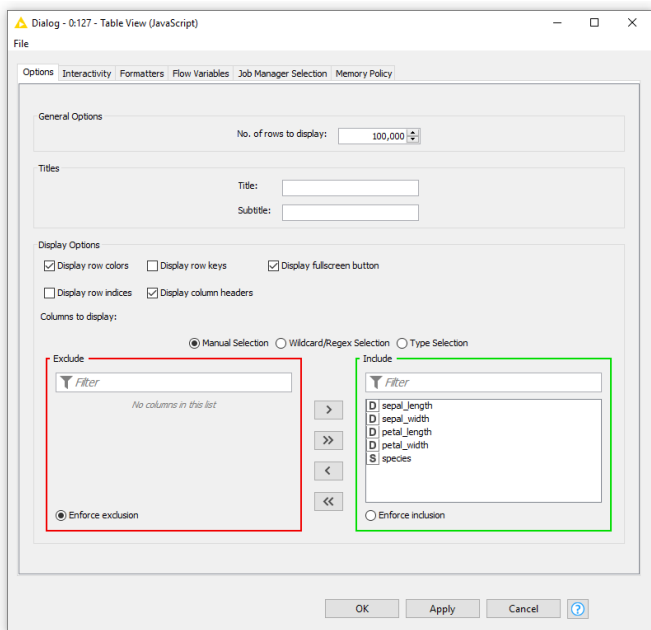
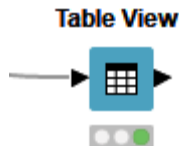
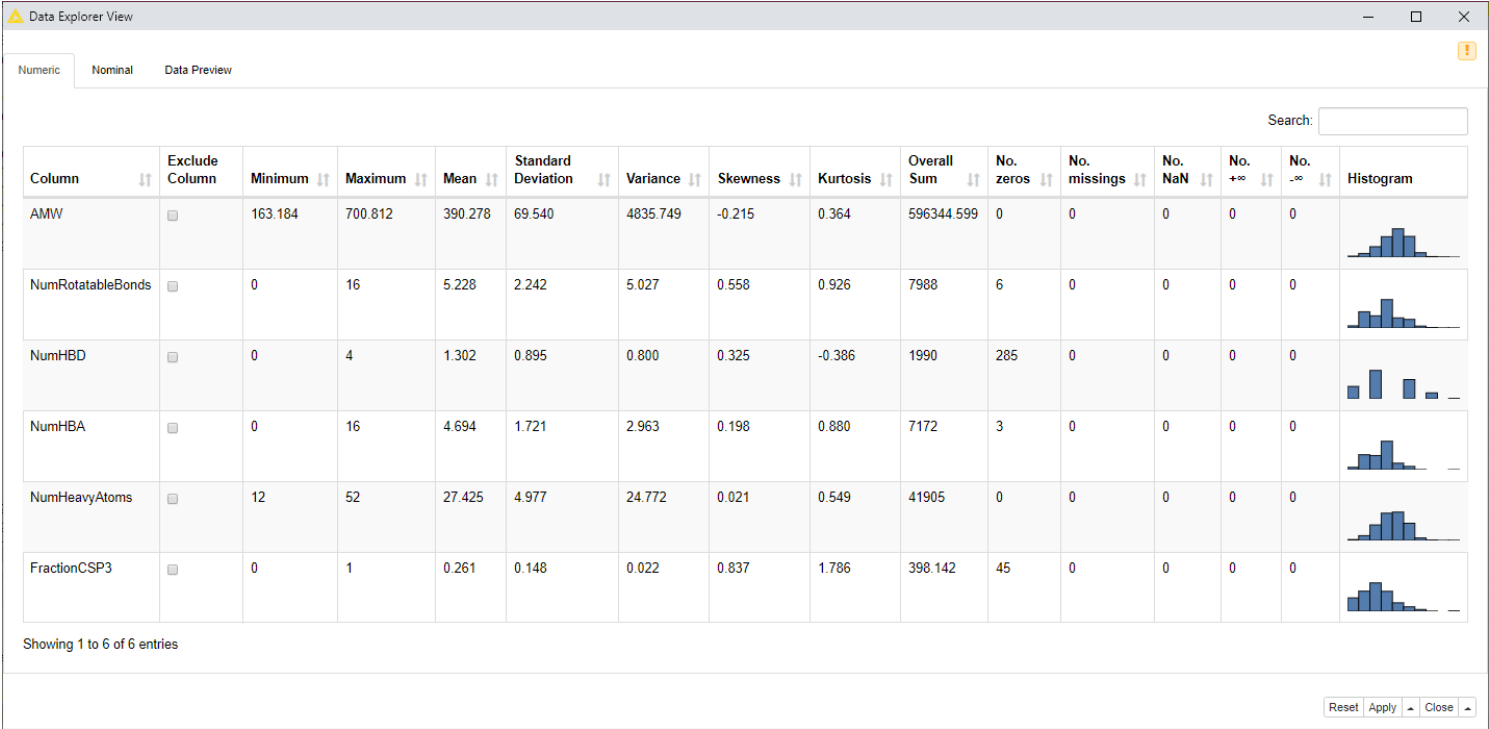
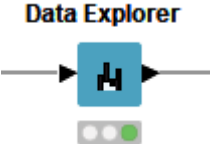


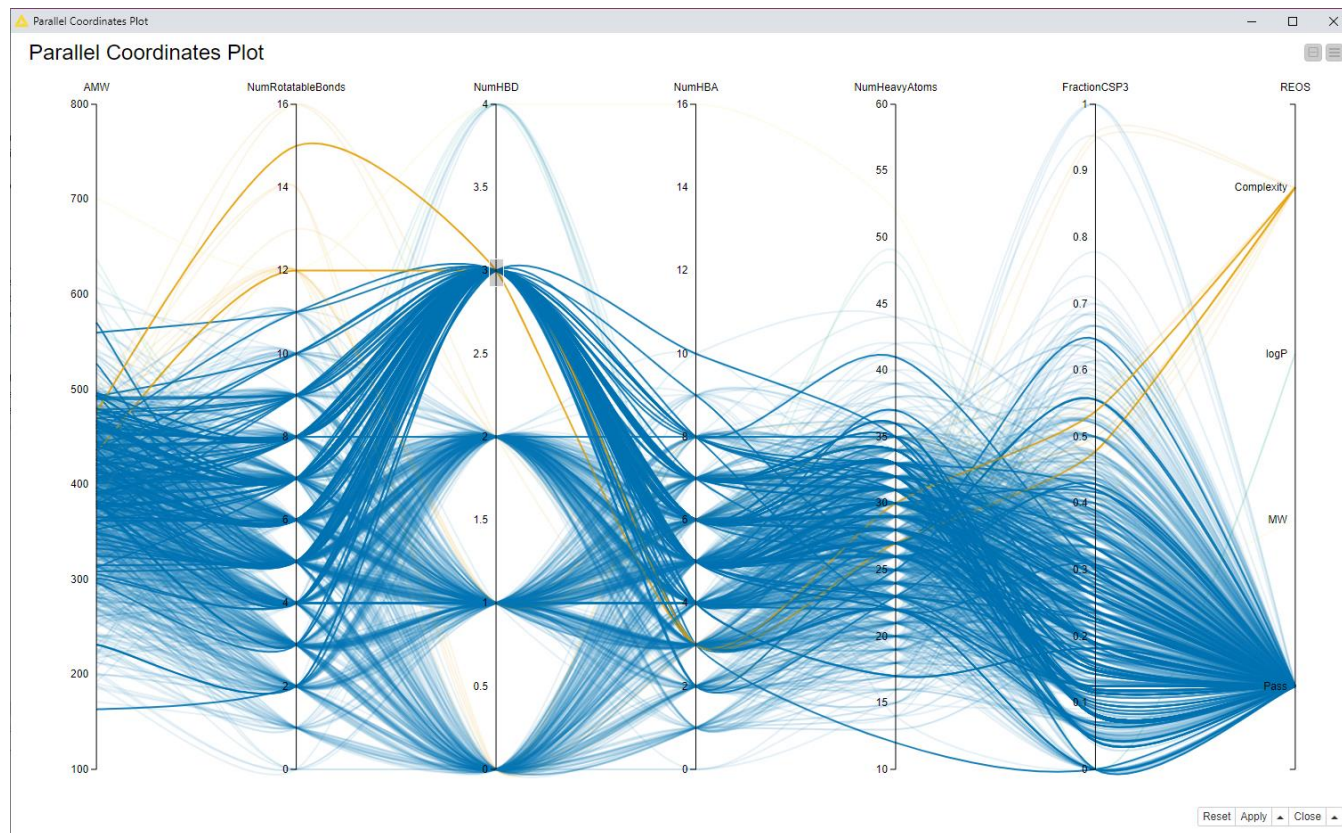
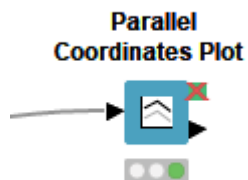
Table View



Data Explorer

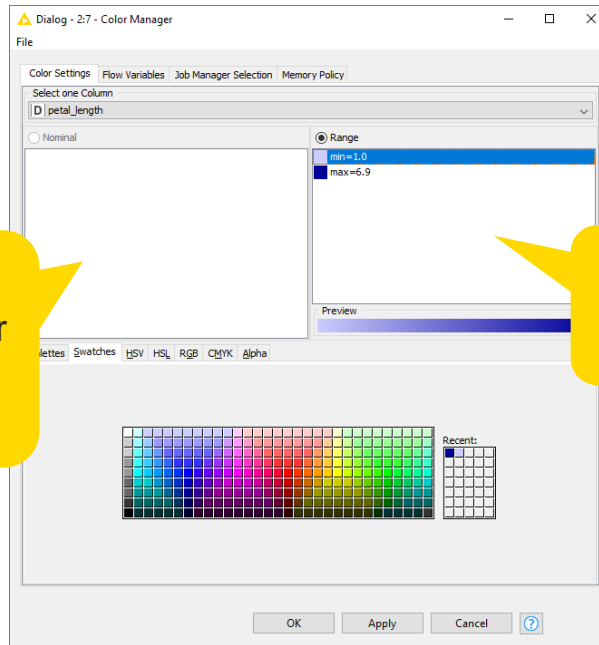


Parallel Coordinates Plot



Color Manager

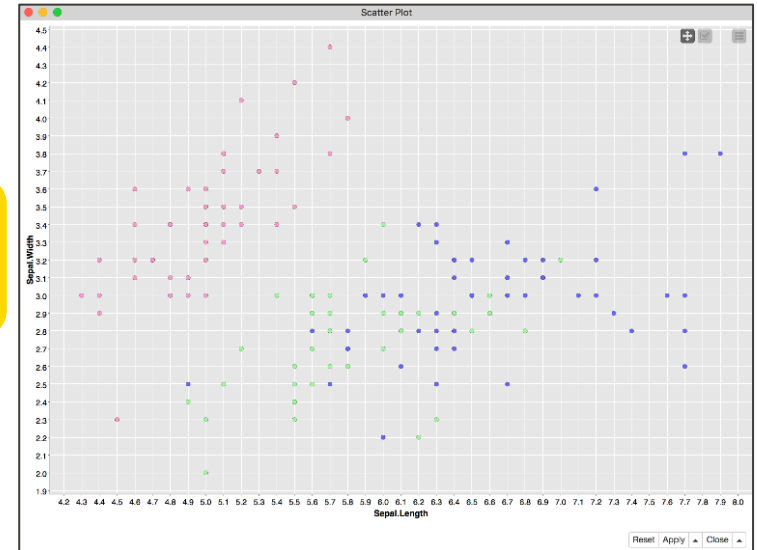
- Color by nominal or continuous values
- Sync colors between views using the color model port and Color Appender node



Discrete
colors for
nominal
values

Color Manager

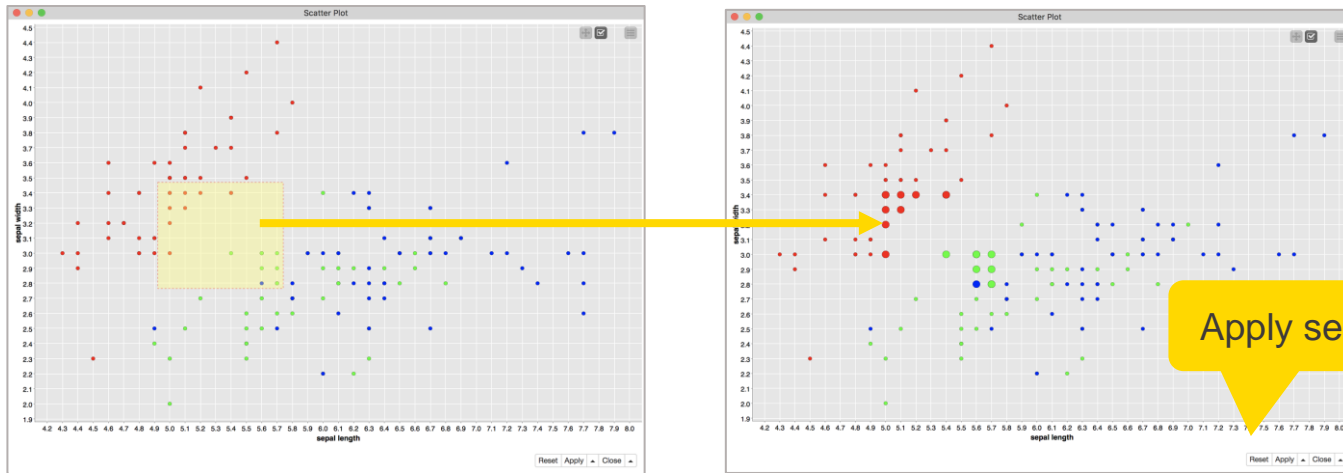
Color range
for numerical
values



Selection & Filtering in JavaScript Views

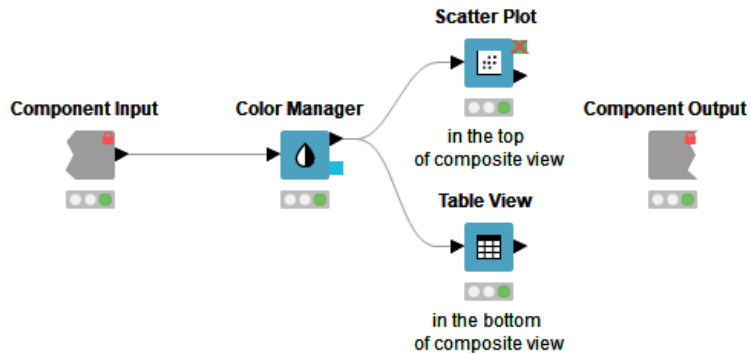
Interactivity allows you to select data points in views

- 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



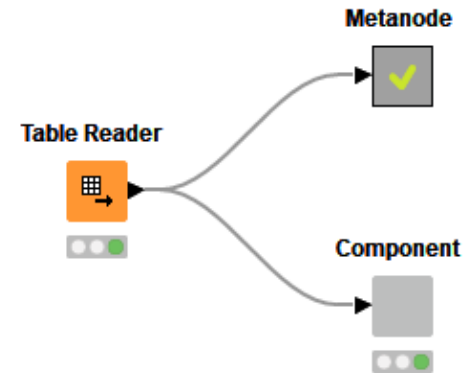
Components – Combined Views

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



Components

- Components encapsulate functionality for reuse and sharing
- Components main features:
 - Local Flow Variable scope
 - Configurable via Configuration nodes
- Components are the key to advanced functionality in KNIME products:
 - Components corresponds to a KNIME WebPortal page
 - Configurations on a WebPortal page are defined using Widget nodes
 - Possibility to be shared via KNIME Hub



Component Description

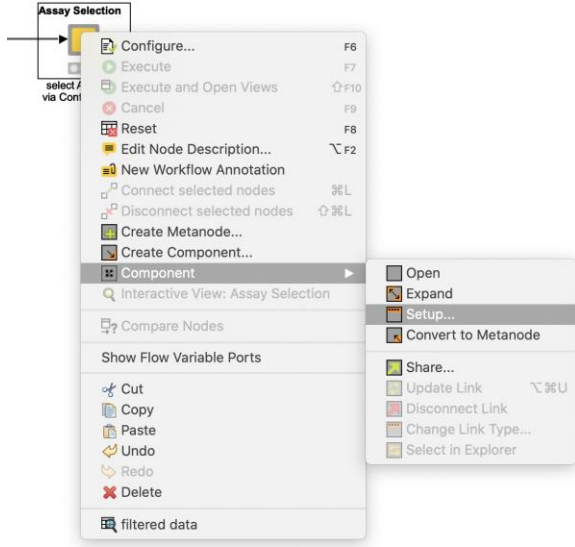
- Make your component look like a KNIME node

The image displays two screenshots of the KNIME software interface, illustrating the process of creating a custom component.

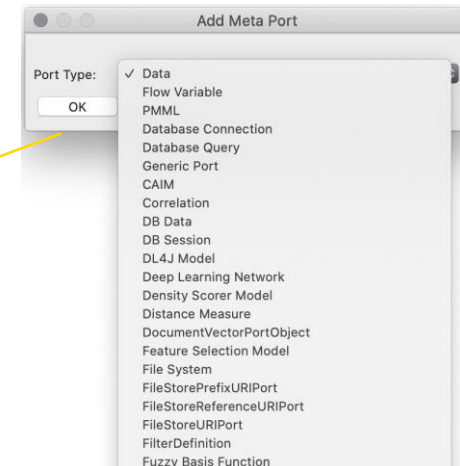
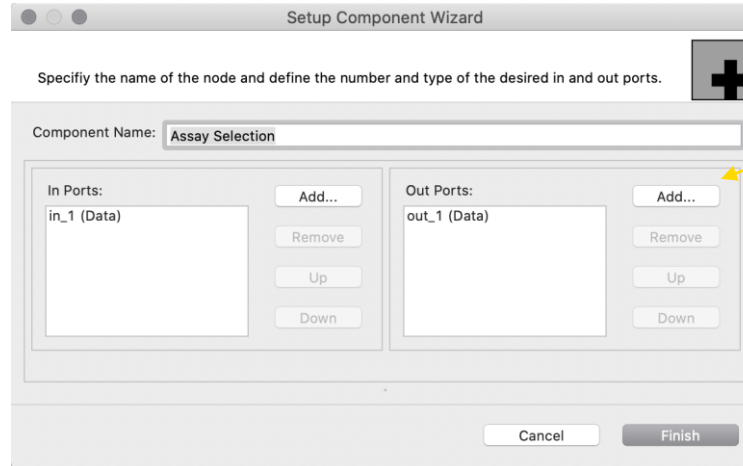
Top Screenshot: Shows the 'Assay Selection' component configuration window. The 'Description' field contains the text: 'This component filters the data by the selected AssayID'. The 'Component Icon' section shows a dropdown menu with options: Learner, Manipulator, Predictor, Sink, Source, and Visualizer. The 'Port #1' field is set to 'all data' with the description 'Data containing records for all AssayIDs'. The 'Output Port #1' field is set to 'filtered data' with the description 'Data containing records for the selected AssayID'. A yellow callout points to the 'Description' field with the text: 'Add description of the component'. Another yellow callout points to the 'Port #1' field with the text: 'Add description of the input and output ports'. A third yellow callout points to the 'Component Icon' dropdown with the text: 'Add background color or icon'.

Bottom Screenshot: Shows the 'Assay Selection' component in a workflow. The component is connected to a 'Table Reader' (input) and an 'Assay Selection' (output). The 'Table Reader' is labeled 'CHEMBLID228_SERV_ligands.table'. The 'Assay Selection' is labeled 'select AssayID via Configuration'. The 'Assay Selection' component's description window is open, showing the same configuration as the top screenshot. A yellow callout points to the 'Assay Selection' component with the text: 'Add description of the component'.

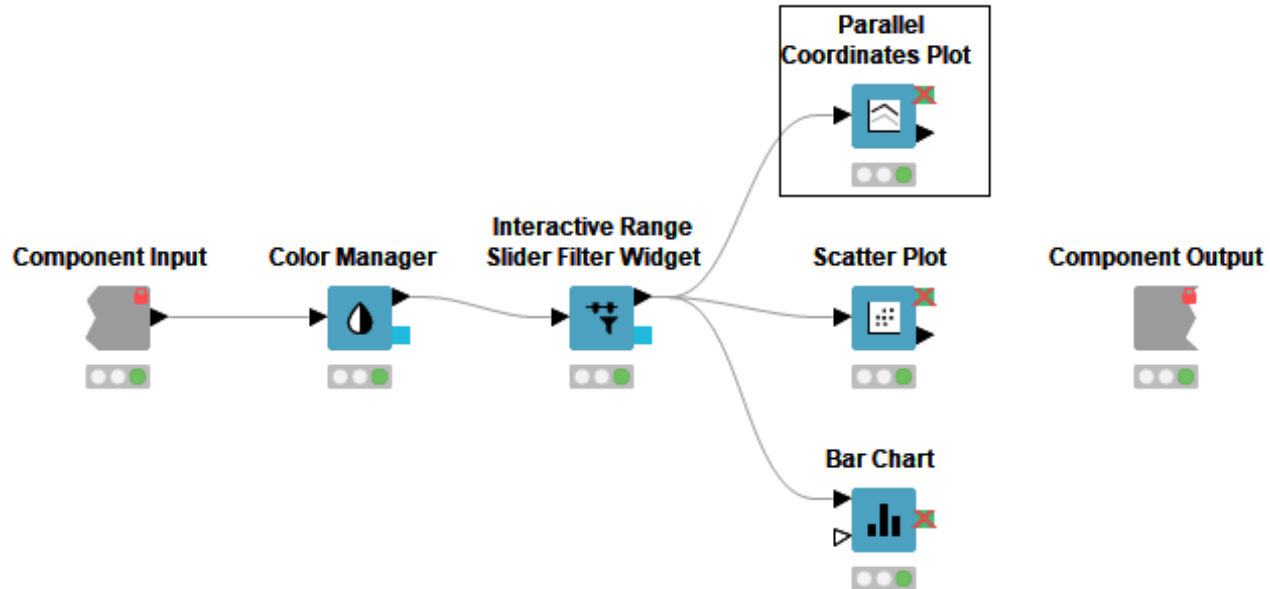
Configure Component Ports



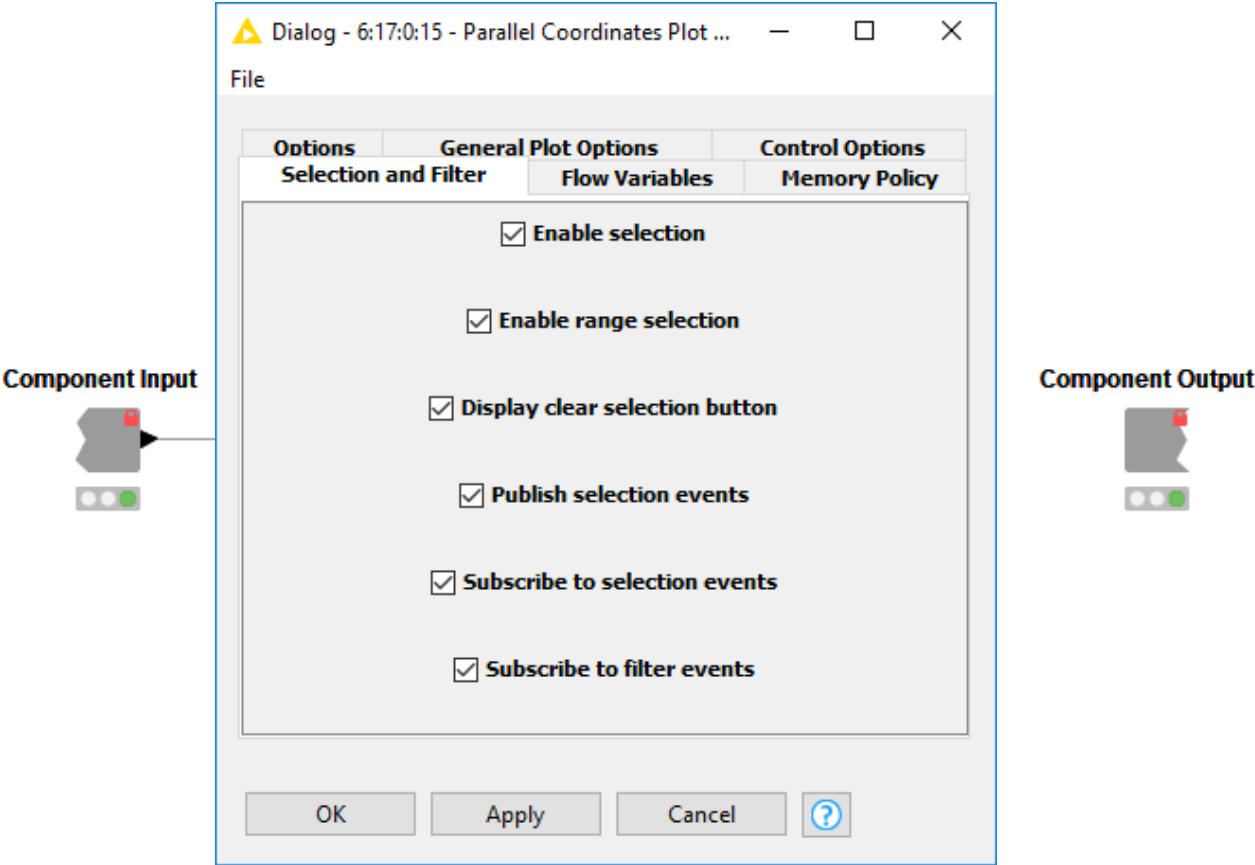
- Add input and output points to Metanodes/Components
- Remove ports to adapt to changes after creation of Metanodes/Components



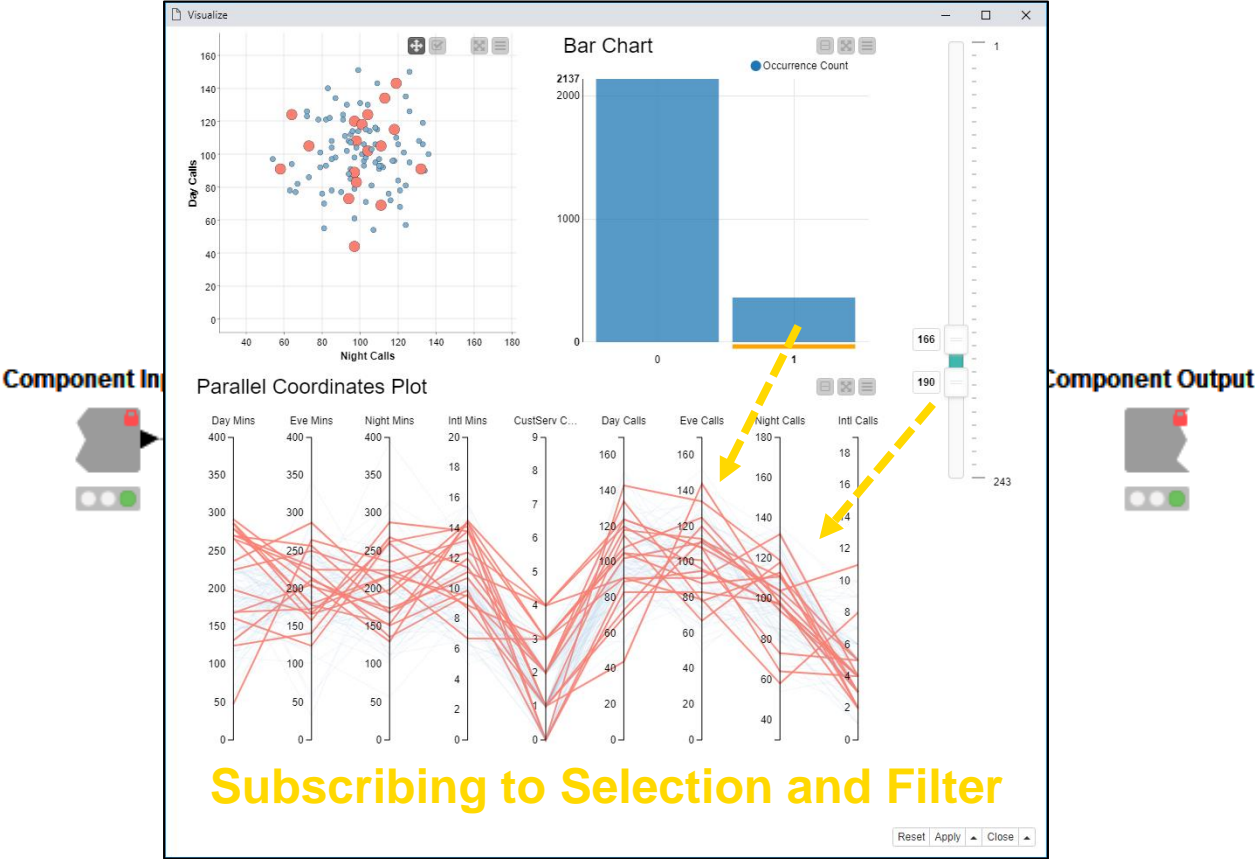
Interactivity across Charts: Selection and Filter Events



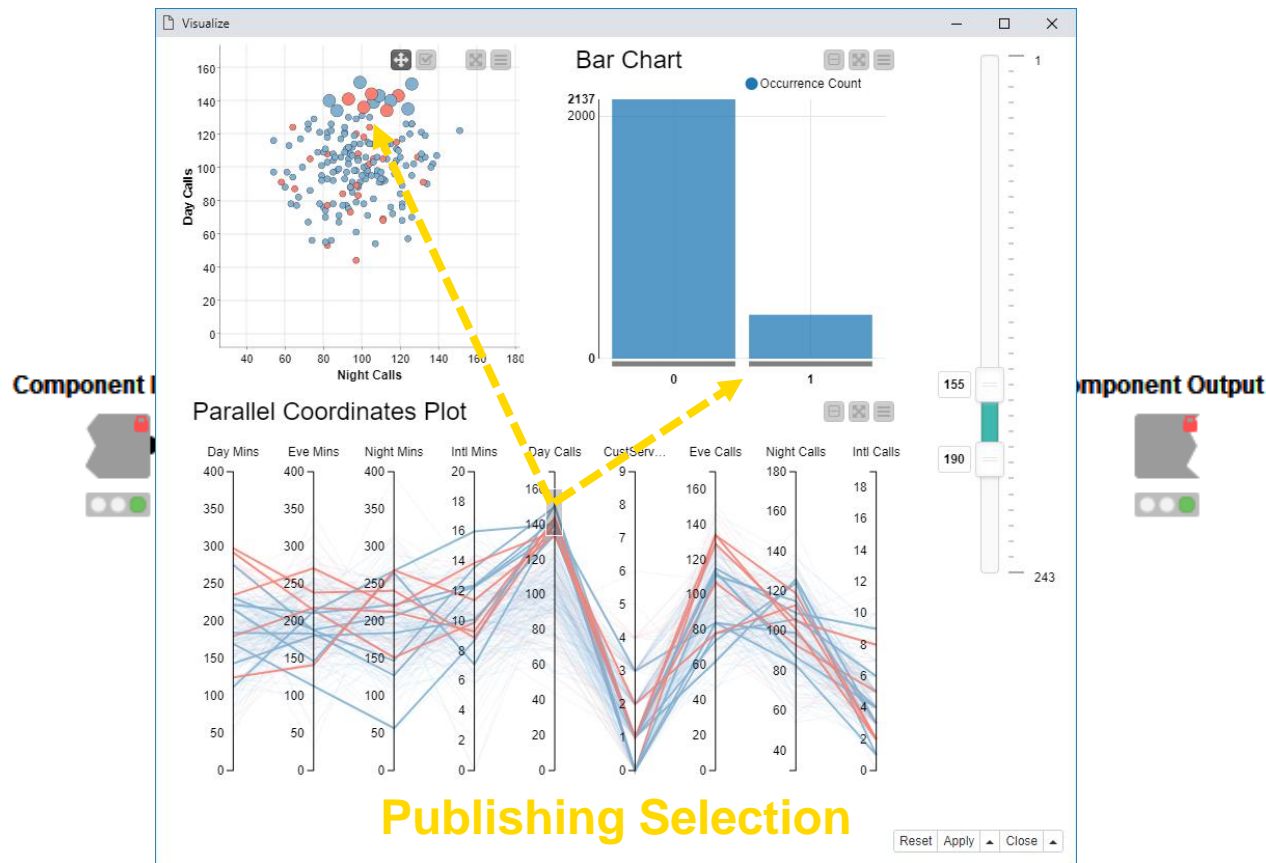
Interactivity across Charts: Selection and Filter Events



Interactivity across Charts: Selection and Filter Events

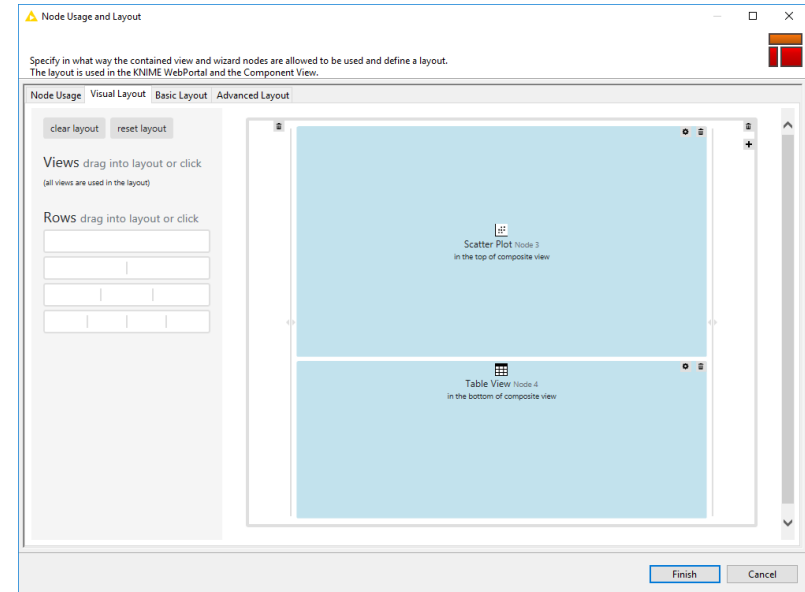
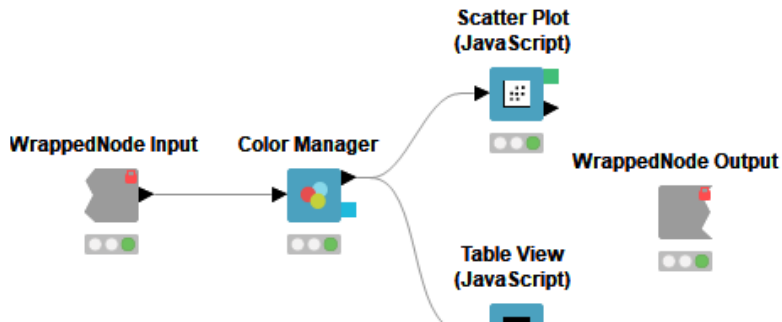
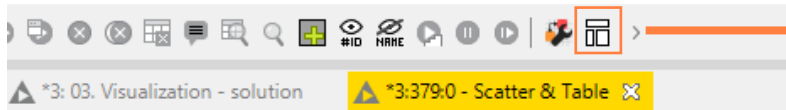


Interactivity across Charts: Selection and Filter Events



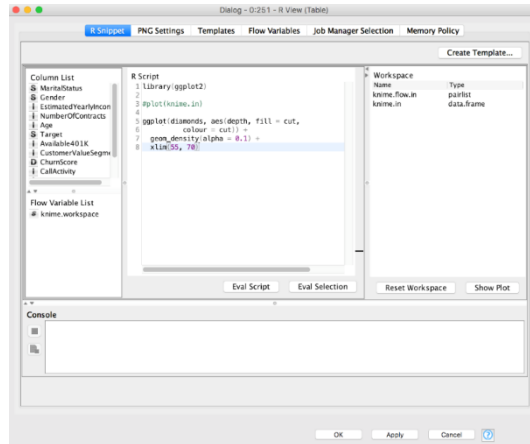
Configure Content and Views Layout

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

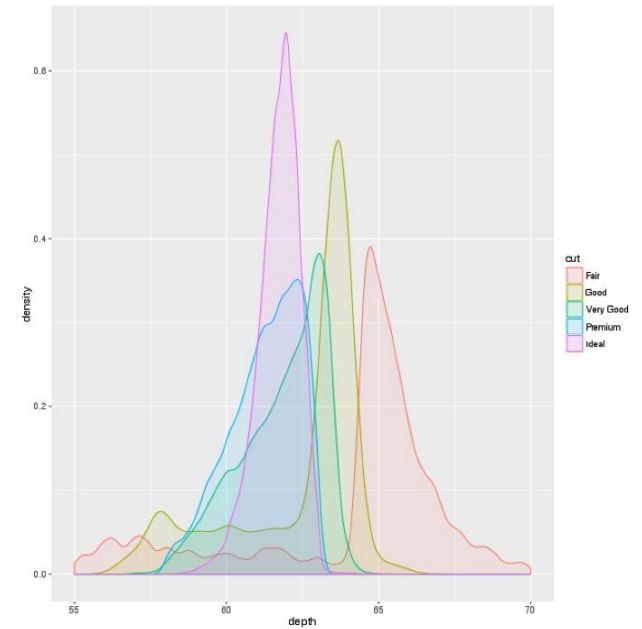


Script-based View Nodes

- 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

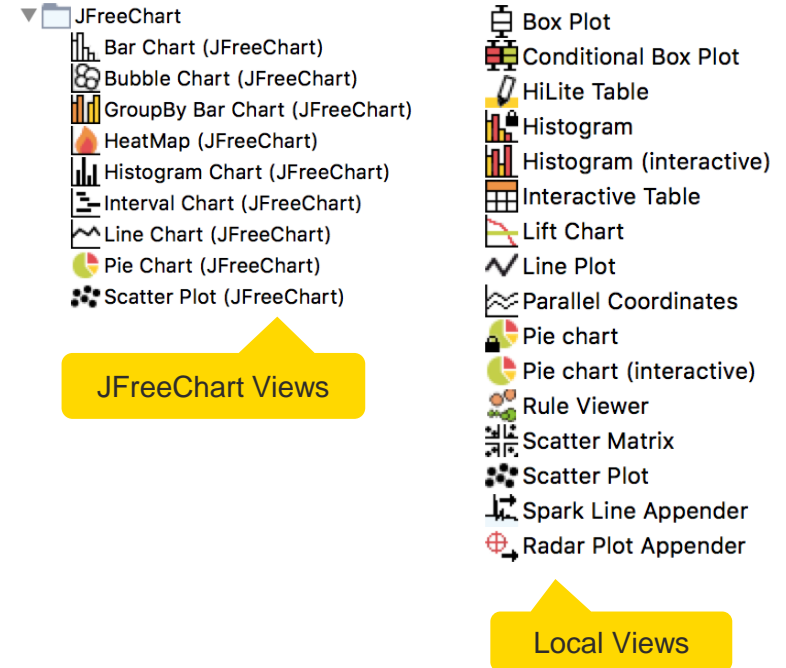


R View (Table)



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

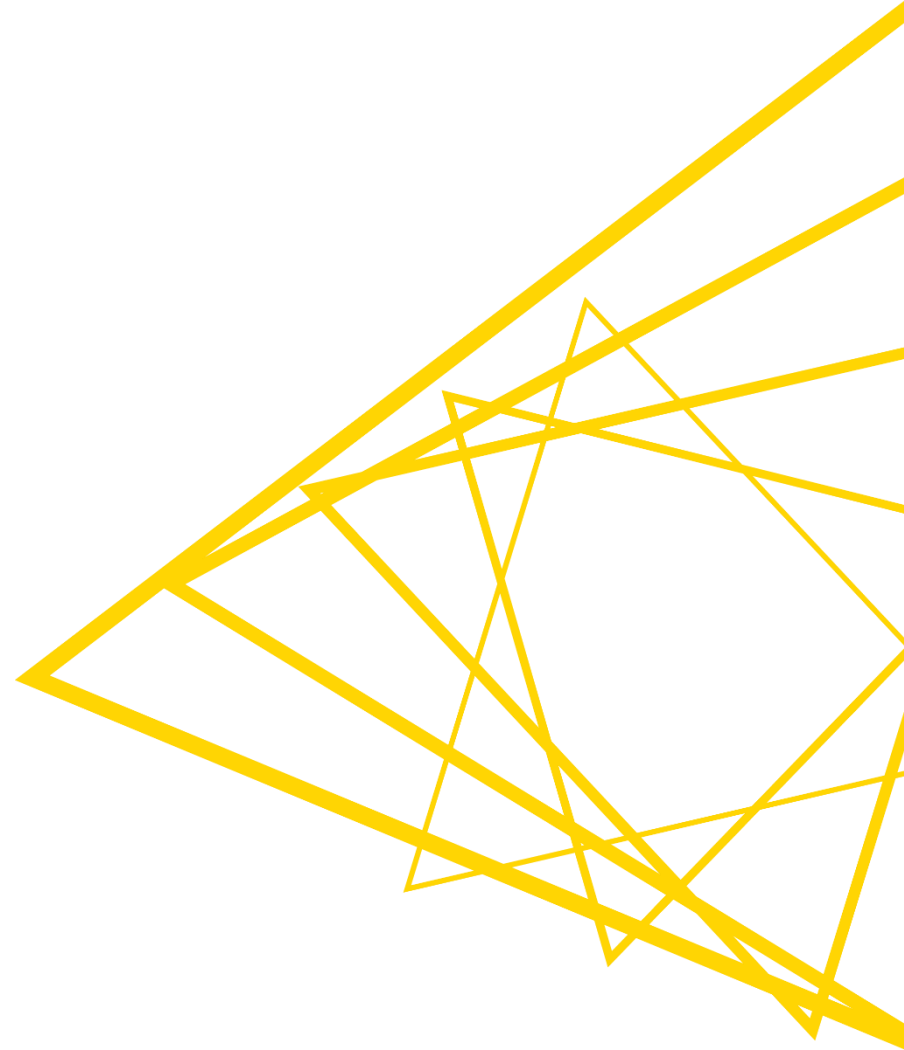
Open Exercise: 03. Visualization

Activity: Visualization

- Use the Parallel Coordinates Plot node and the Table View node to visualize the data table
 - Use the Groupby node to count the number of entries in each REOS category and display the result in a second Table View node.
 - Create a component containing the JavaScript Views 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.

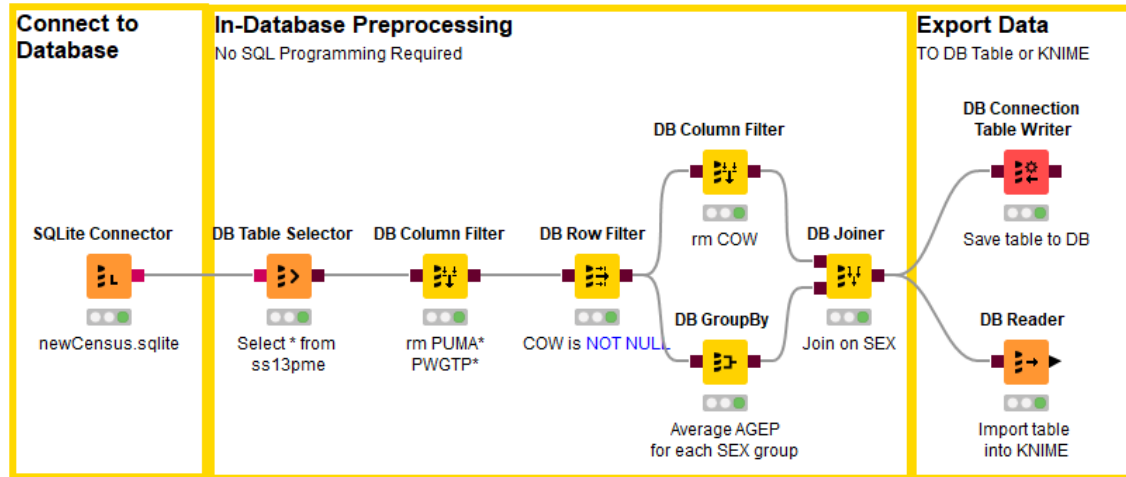
Database

Access, Read, Manipulate, Write

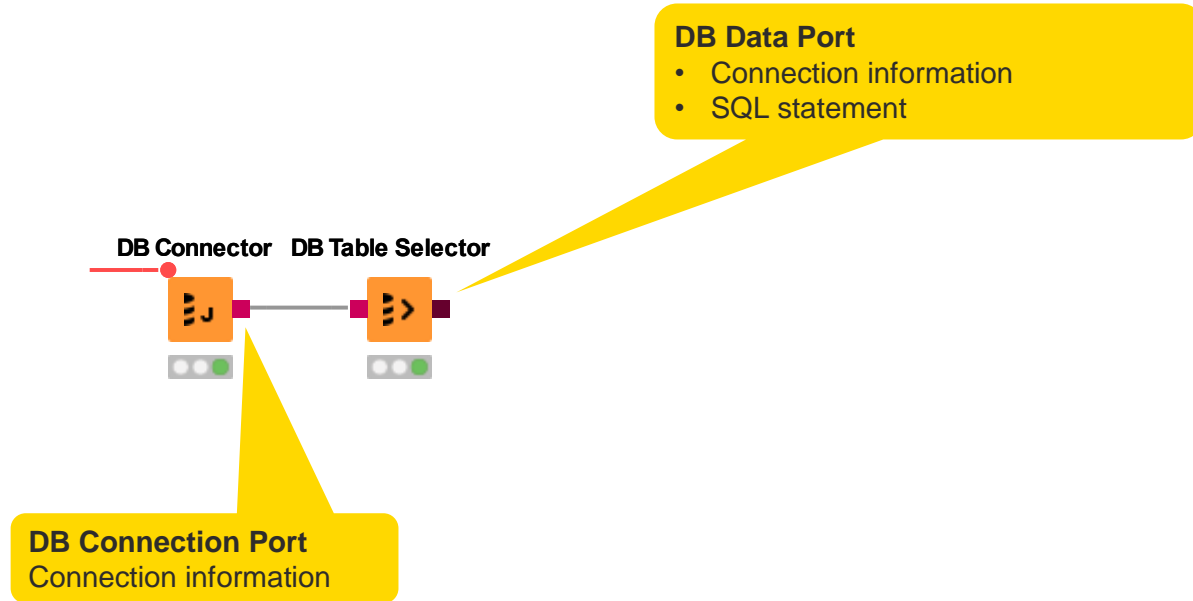


Database Extension

- 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

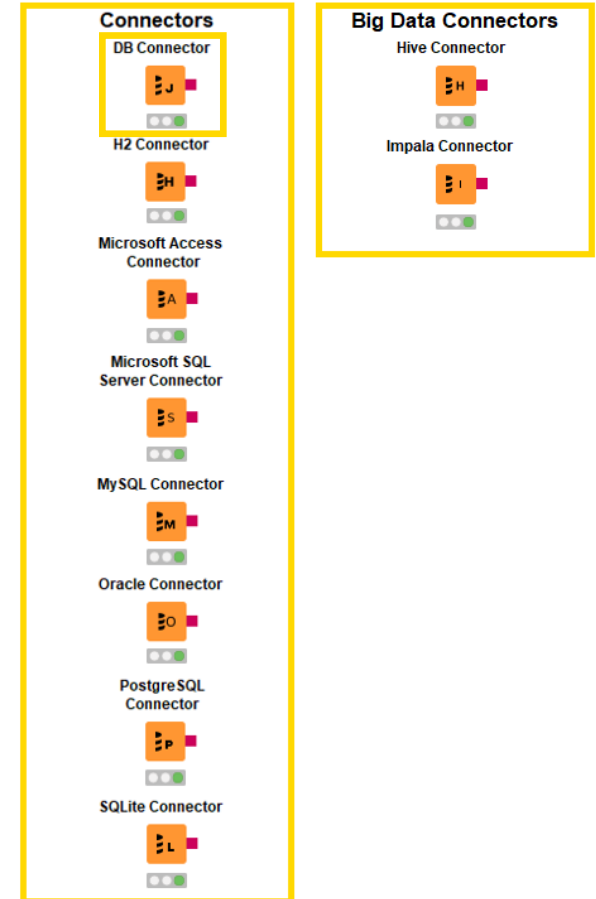


Database Port Types



Database Connectors

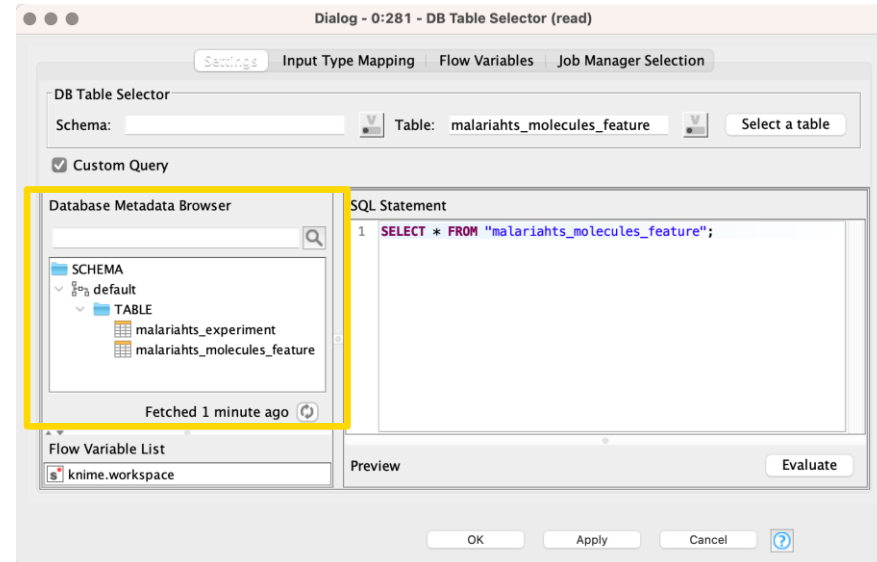
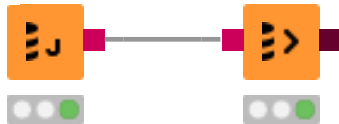
- Dedicated nodes to connect to specific Databases
 - Necessary JDBC driver included
 - Easy to use
 - Import DB specific behavior/capability
- Hive, Impala connectors part of the KNIME Big Data Connectors extension
- General DB Connector
 - Can connect to any JDBC source
 - Register new JDBC driver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



DB Table Selector

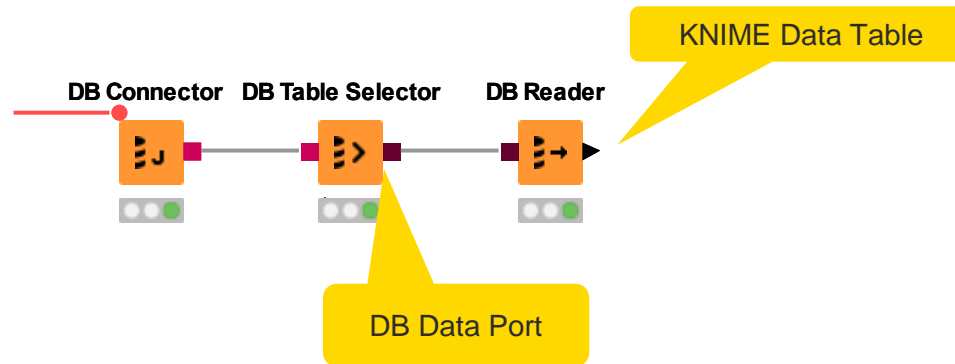
- Takes connection information and constructs a query
- Explore DB metadata
- Outputs a SQL query

DB Connector DB Table Selector



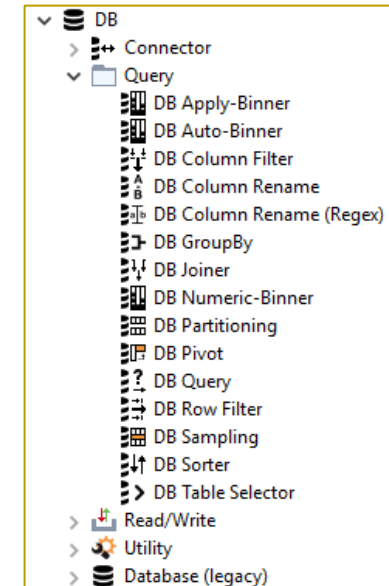
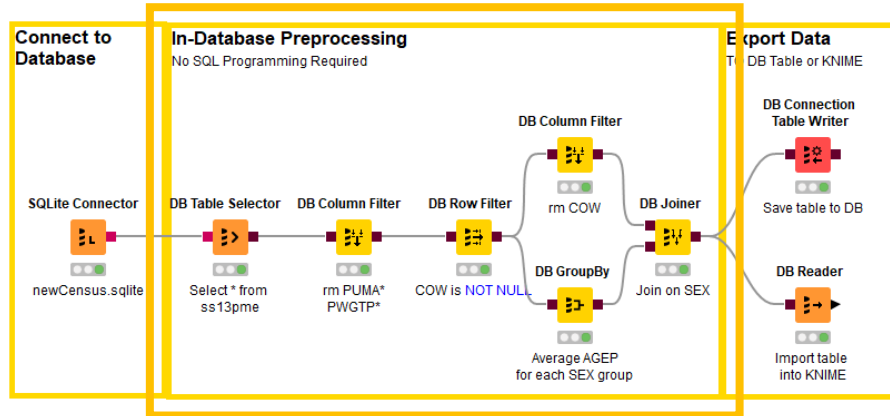
DB Reader

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



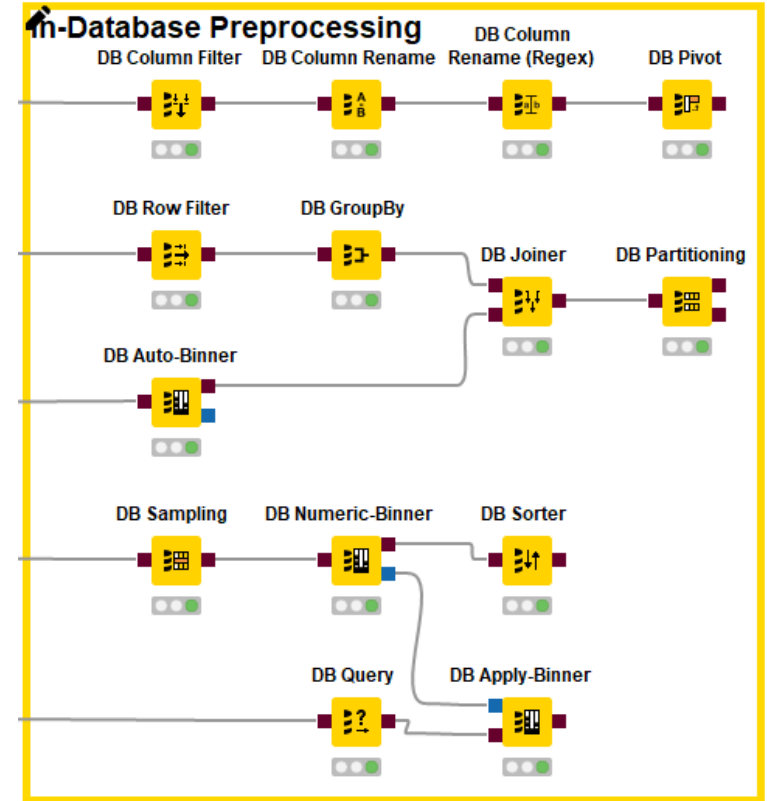
In-Database Processing

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



Query Nodes

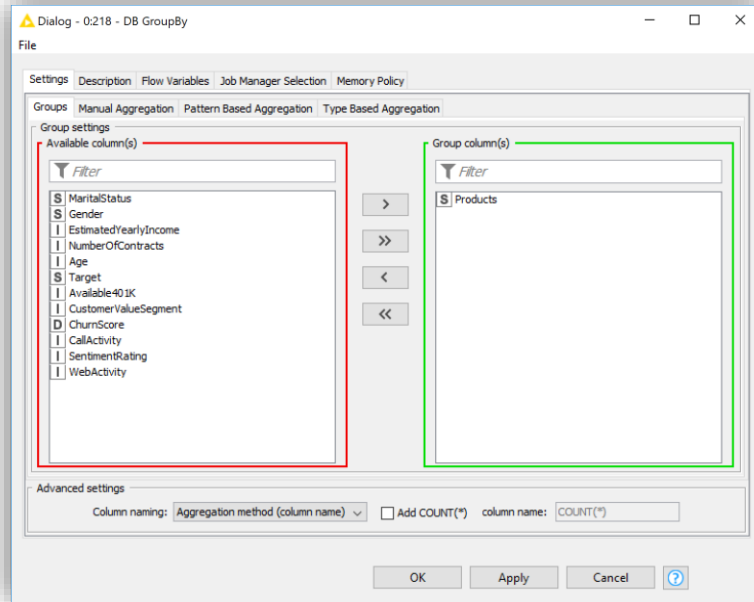
- Filter rows and columns
- Join tables/queries
- Extract samples
- Bin numeric columns
- Sort your data
- Write your own query
- Aggregate your data
- Partition your data



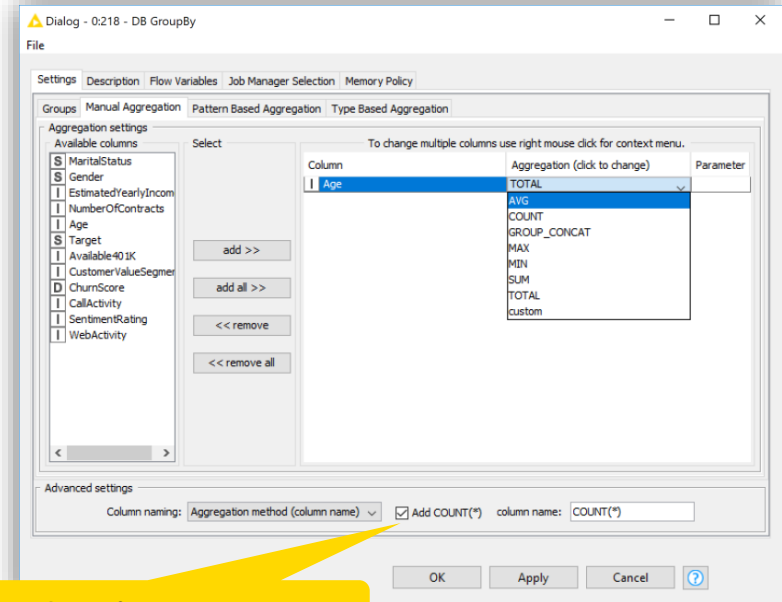
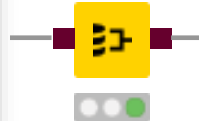
DB GroupBy

Aggregate rows to summarize data

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

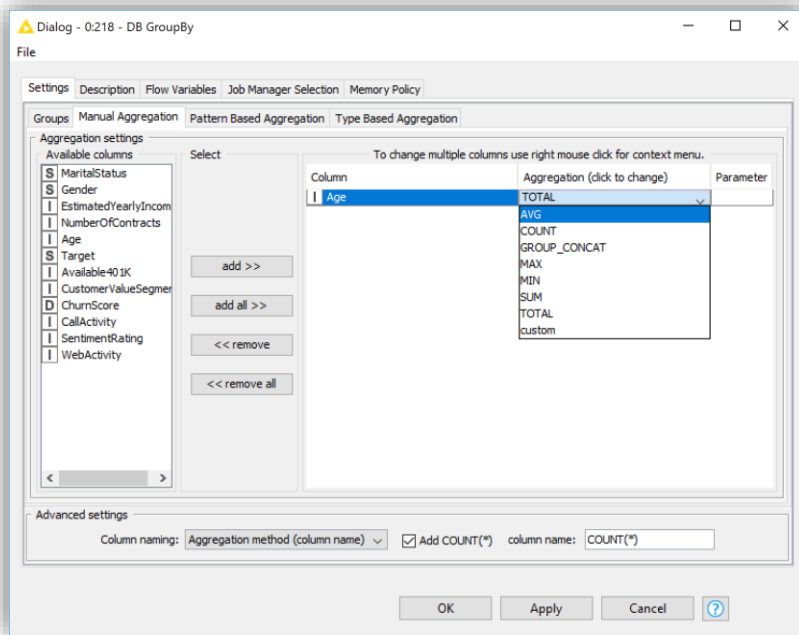


DB GroupBy

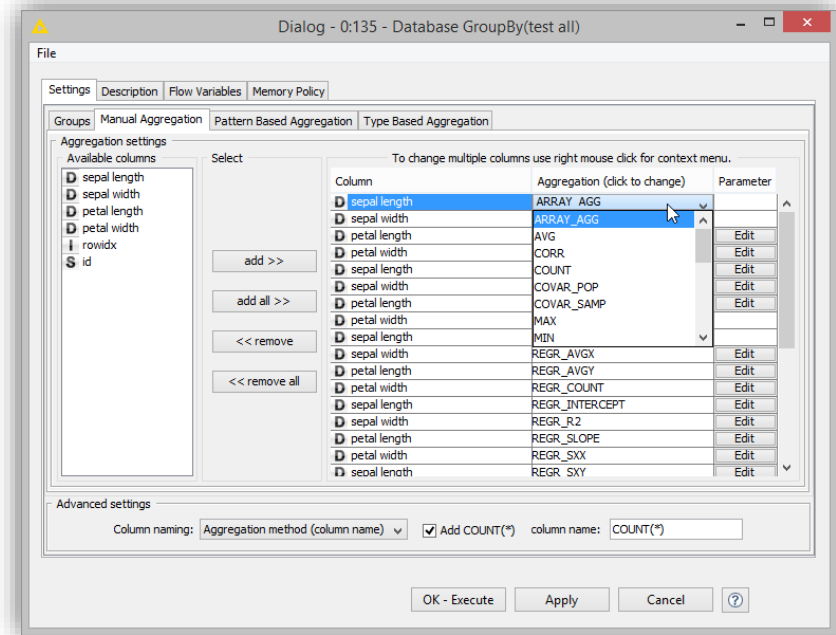


Returns number of rows per group

DB GroupBy – DB Specific Aggregation Methods



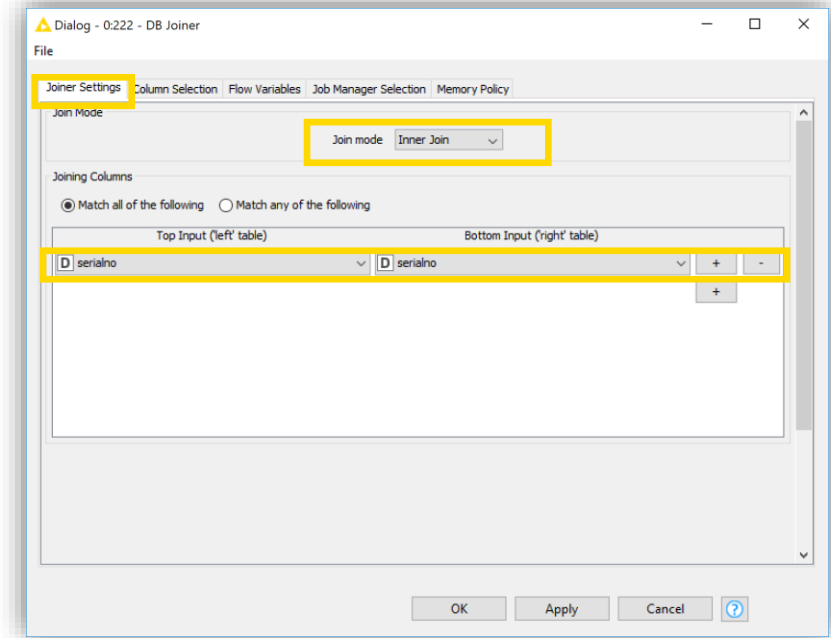
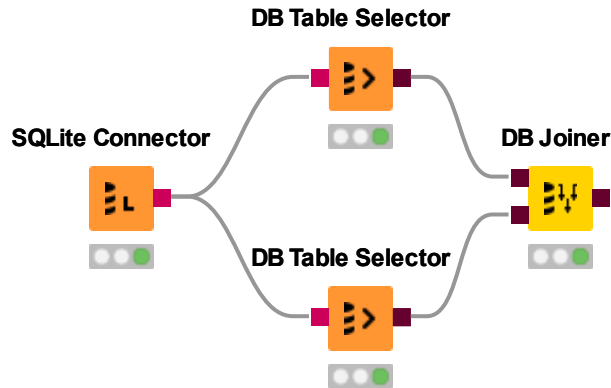
SQLite: 7 aggregation functions



PostgreSQL: 25 aggregation functions

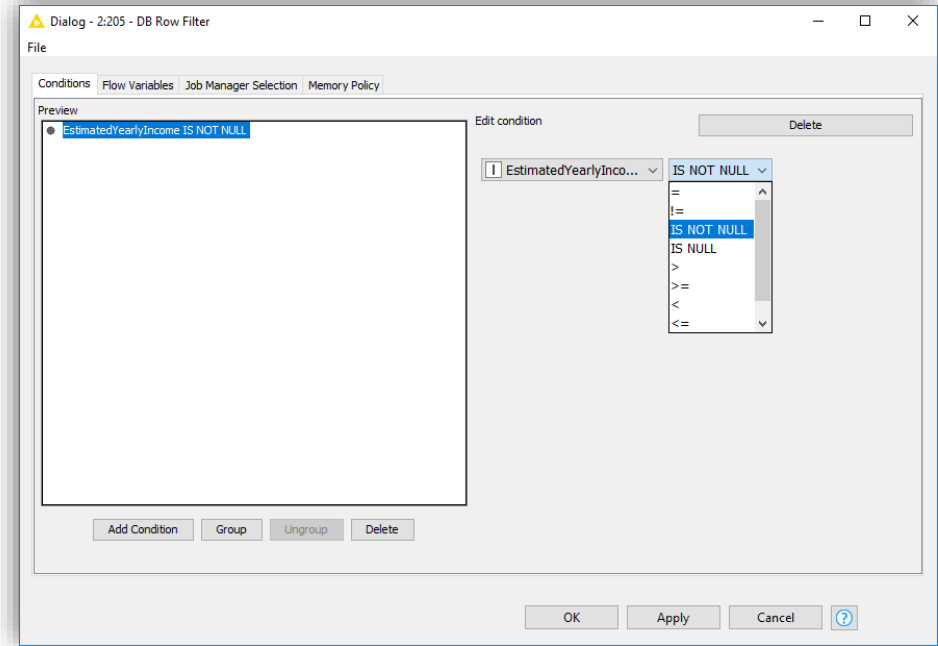
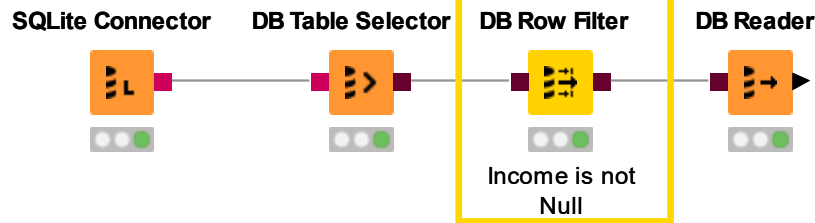
DB Joiner

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



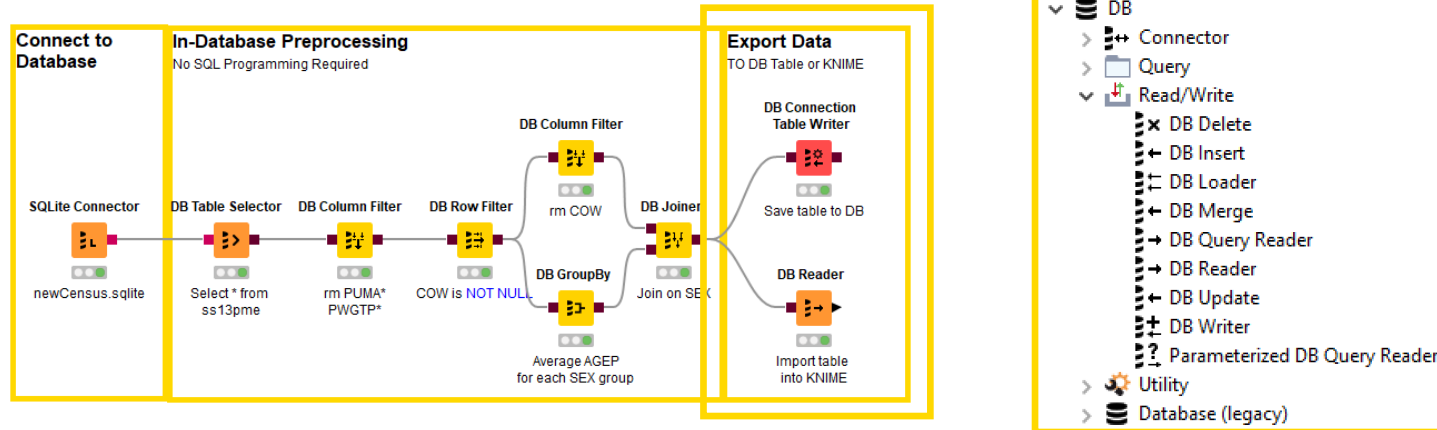
DB Row Filter

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



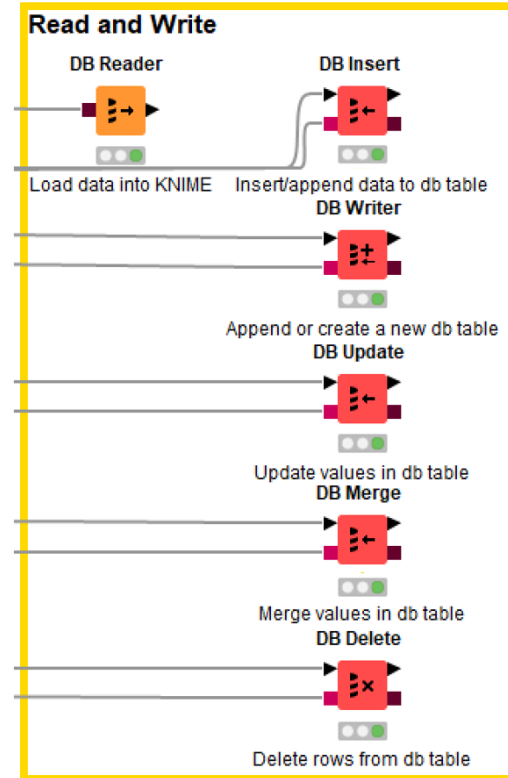
Export Data

- 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



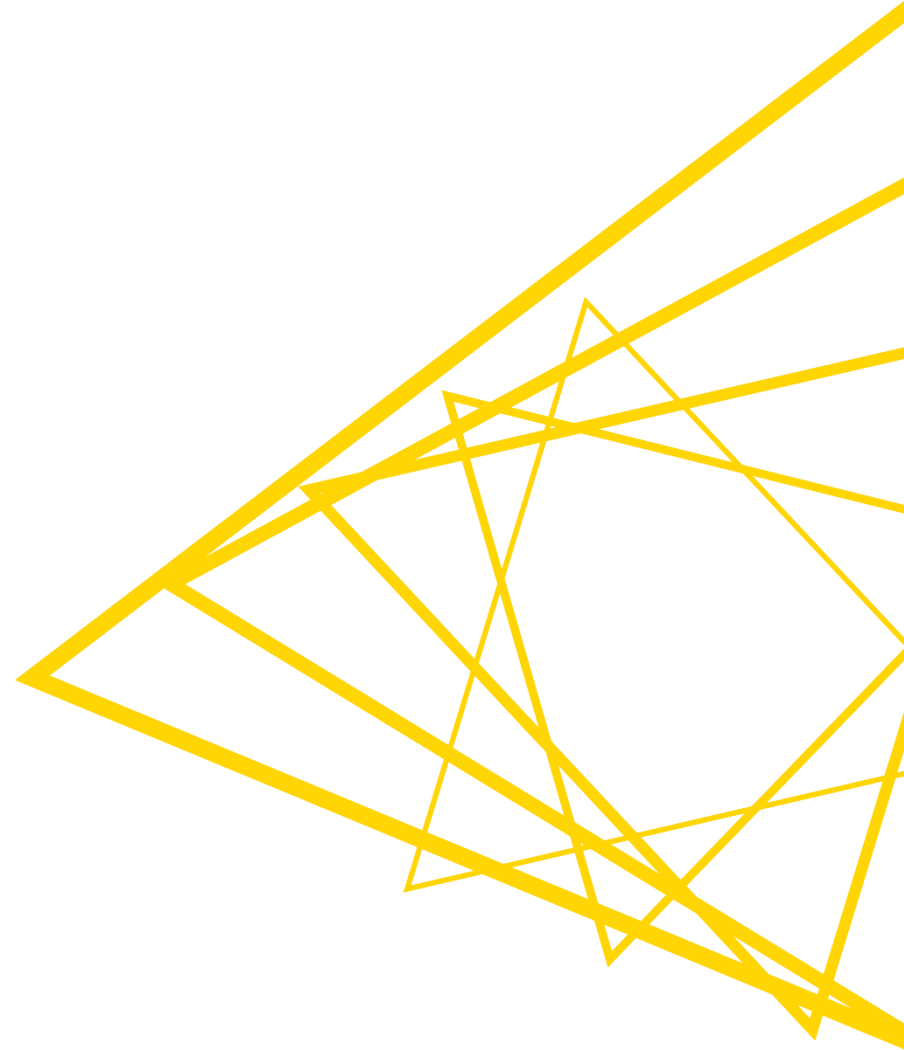
Databases Exercise

Open Exercise 04. 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

Flow Variables



Goal of this Session

- What is a Flow Variable?
- How to create a Flow Variable?
- How to use a Flow Variable to overwrite node settings?
- How to use a Configuration node to parameterize a Component?
- How to use a Widget node to parameterize a Component via an interactive view?

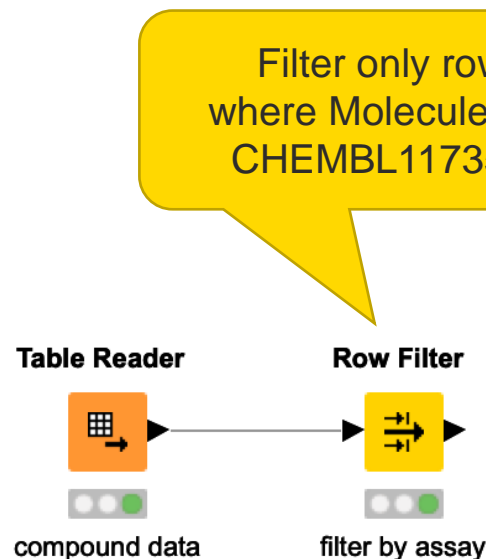
Goal of this Session

- What is a Flow Variable?
- How to create a Flow Variable?
- How to use a Flow Variable to overwrite node settings?
- How to use a Configuration node to parameterize a Component?
- How to use a Widget node to parameterize a Component via an interactive view?

Flow Variables: Usage Examples

- I want to filter for a different Molecule ID in each of my projects based on certain criteria
- I want to filter the molecule that was tested the most

Assay ID	Molecule ID	Ki
CHEMBL853187	CHEMBL208069	8
CHEMBL853187	CHEMBL382554	4.6
...
CHEMBL1176921	CHEMBL1173532	19.95
CHEMBL1176921	CHEMBL1173530	7.943
...
CHEMBL1176750	CHEMBL1173532	1.047



Flow Variables: Usage Examples

- I want to filter for a different Molecule ID in each of my projects based on certain criteria
- I want to filter the molecule that was tested the most

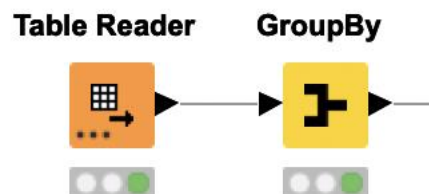
Assay ID	Molecule ID	Ki
CHEMBL853187	CHEMBL208069	8
CHEMBL853187	CHEMBL382554	4.6
...
CHEMBL1176921	CHEMBL1173532	19.95
CHEMBL1176921	CHEMBL1173530	7.943
...
CHEMBL1176750	CHEMBL1173532	1.047

Solution

- I need to aggregate/group my data based on the Molecule ID
- Count the number of entries for each Molecule ID

Flow Variables: Usage Examples

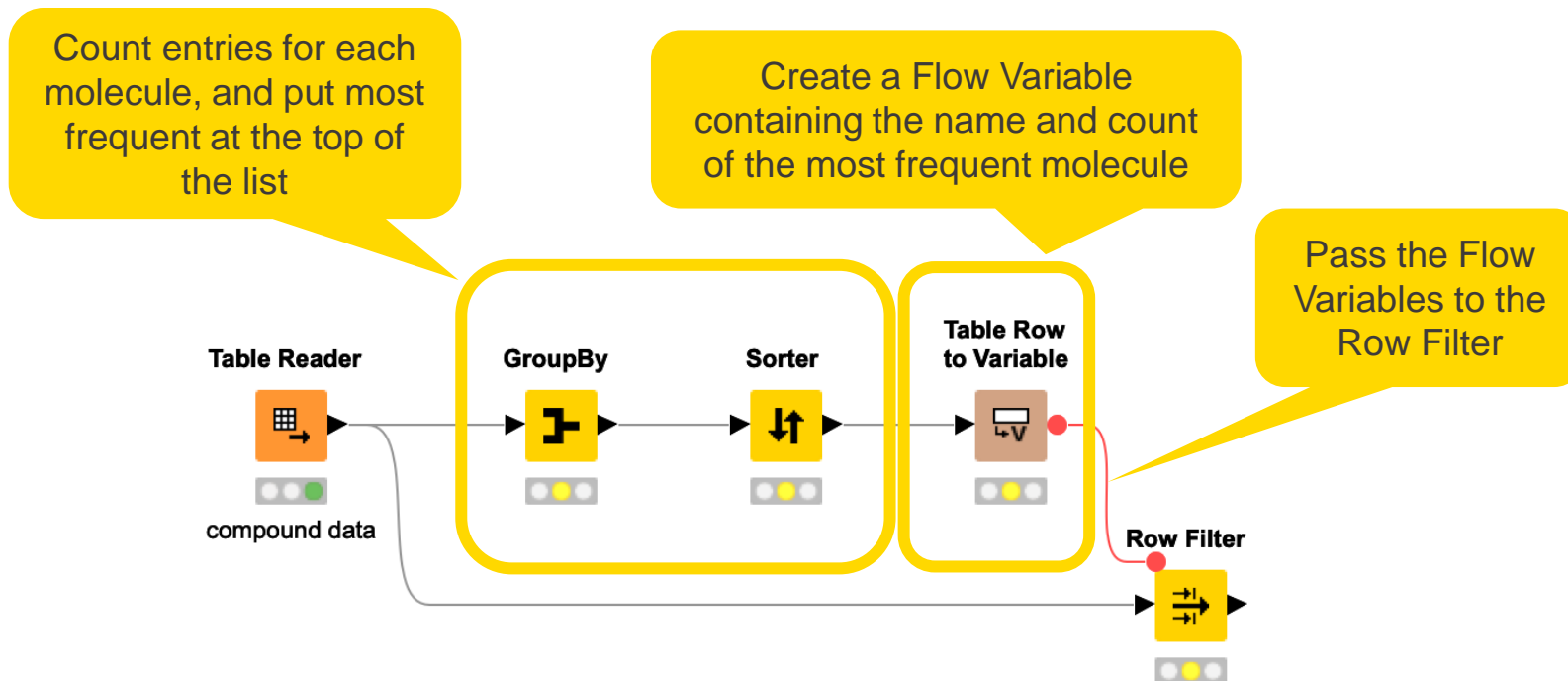
- Each time I need to launch the Analytics Platform....
- Use a GroupBy node and find the most tested molecule to update the Row Filter...



Molecule ID	count
CHEMBL208069	1
CHEMBL382554	1
CHEMBL1173532	2
CHEMBL1173530	1
...	...

- Or do I? Perhaps Flow Variables can help ...

Automatically filter by most frequent tested molecule



Goal of this Session

- What is a Flow Variable?
- How to create a Flow Variable?
- How to use a Flow Variable to overwrite node settings?
- How to use a Configuration node to parameterize a Component?
- How to use a Widget node to parameterize a Component via an interactive view?

Create Flow Variables using Table Row to Variable node

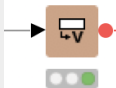
- Takes a table as input and converts the first row to Flow Variables
 - Column names -> Flow Variable names
 - Column values -> Flow Variable values
- Only the first row is transformed, additional rows are discarded

Sorted Table - 8:469 - Sorter

Table "default" - Rows: 327 Spec - Columns: 2 Properties Flow Variables

Row ID	[s] molecule_chembl_id	[i] Count(assay_chembl_id)
Row27	CHEMBL1173532	4
Row2	CHEMBL11	3
Row9	CHEMBL1172744	3
Row11	CHEMBL1172850	3
Row17	CHEMBL1173341	3
Row26	CHEMBL1173530	3
Row28	CHEMBL1173533	3
Row30	CHEMBL1173595	3
Row43	CHEMBL1237064	3
Row164	CHEMBL1370805	3

Table Row to Variable



Variables Output - 3:493 - Table Row to Variable

Flow Variables

Index	Owner ID	Name	Value
0	3:493	[s] molecule_chembl_id	CHEMBL1173532
0	3:493	[i] Count(assay_chembl_id)	4
0	3:493	[s] RowID	Row27
0		[s] knime.workspace	/Users/janinamothos/knime-training

Dialog - 3:493 - Table Row to Variable

Settings Flow Variables Job Manager Selection

Missing values Handling

- ☐ Fail
- ☒ Use defaults if available
- ☐ Omit

Defaults

String missing

Boolean false

Integer 0

Long 0

Double 0.0

Column selection

☒ Manual Selection ☐ Wildcard/Regex Selection ☐ Type Selection

Exclude

Filter

No columns in this list

☒ Enforce exclusion

Include

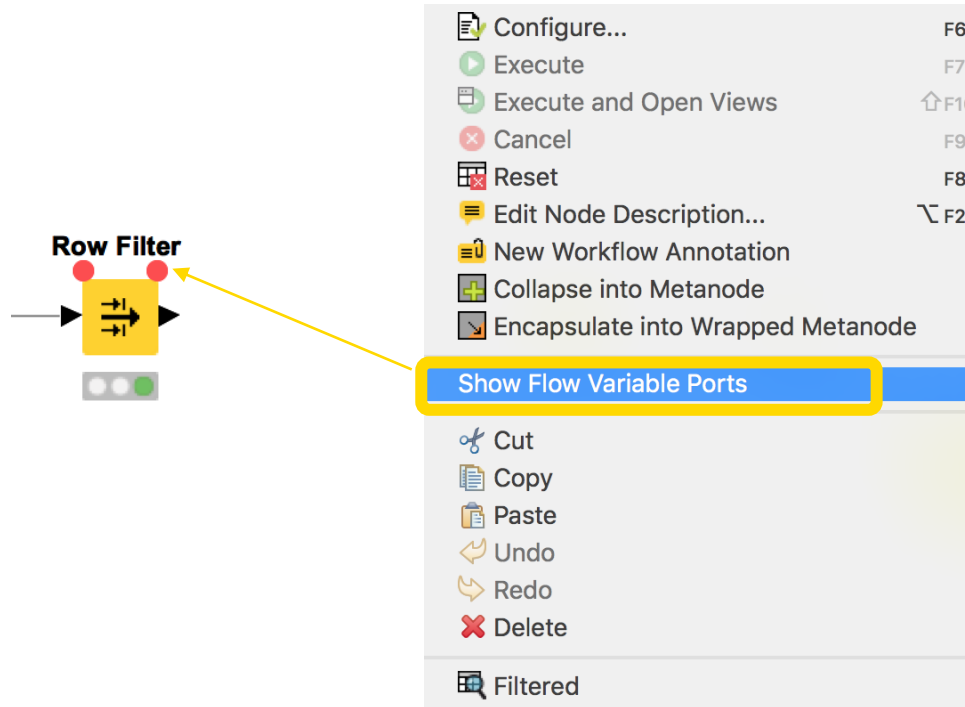
Filter

- [s] molecule_chembl_id
- [i] Count(assay_chembl_id)

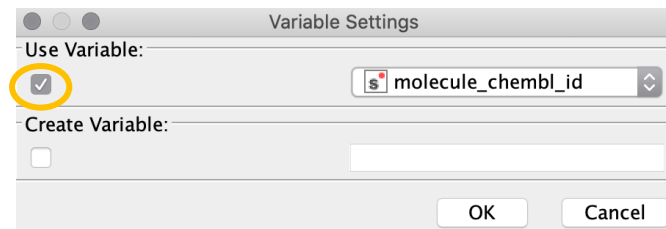
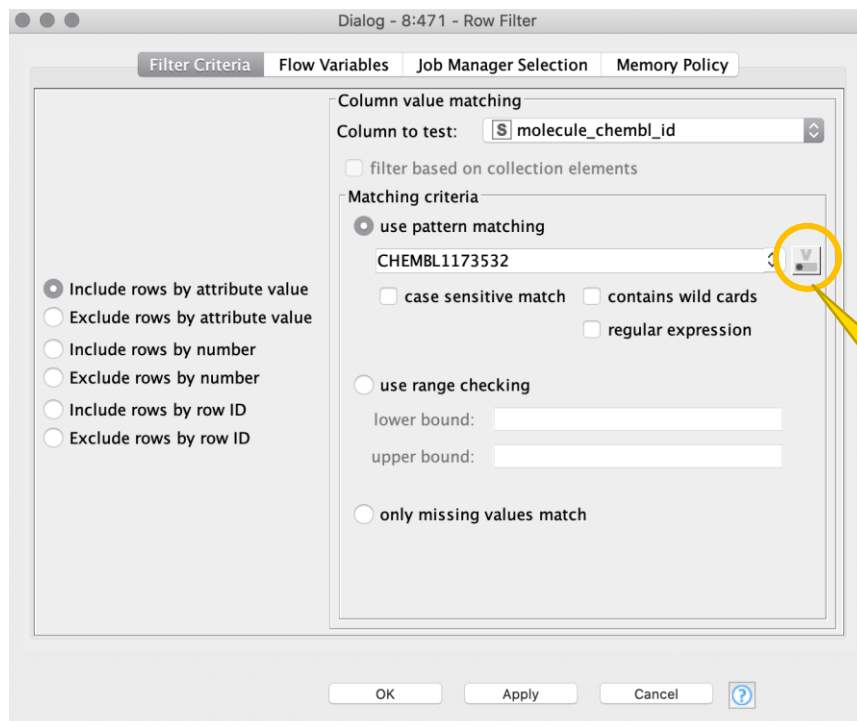
☐ Enforce inclusion

OK Apply Cancel

Flow Variable Ports



Apply a Flow Variable (Button)



The Flow Variable button

Apply a Flow Variable (Advanced)

The screenshot shows a software dialog box titled "Dialog - 8:471 - Row Filter". It has four tabs: "Filter Criteria", "Flow Variables", "Job Manager Selection", and "Memory Policy". The "Flow Variables" tab is selected and highlighted with a yellow circle. Below the tabs, there is a section labeled "rowFilter" with a dropdown arrow. Under this section, there are several rows of settings, each with a small icon (s for string, b for boolean) and a text field. The "ColumnName" row is highlighted with a yellow circle, and its dropdown menu is open, showing a list of available flow variables. The variable "molecule_chembl_id" is selected in this list. Two yellow callout boxes with arrows point to the "Flow Variables" tab and the list of available flow variables.

Dialog - 8:471 - Row Filter

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

▼ ? rowFilter

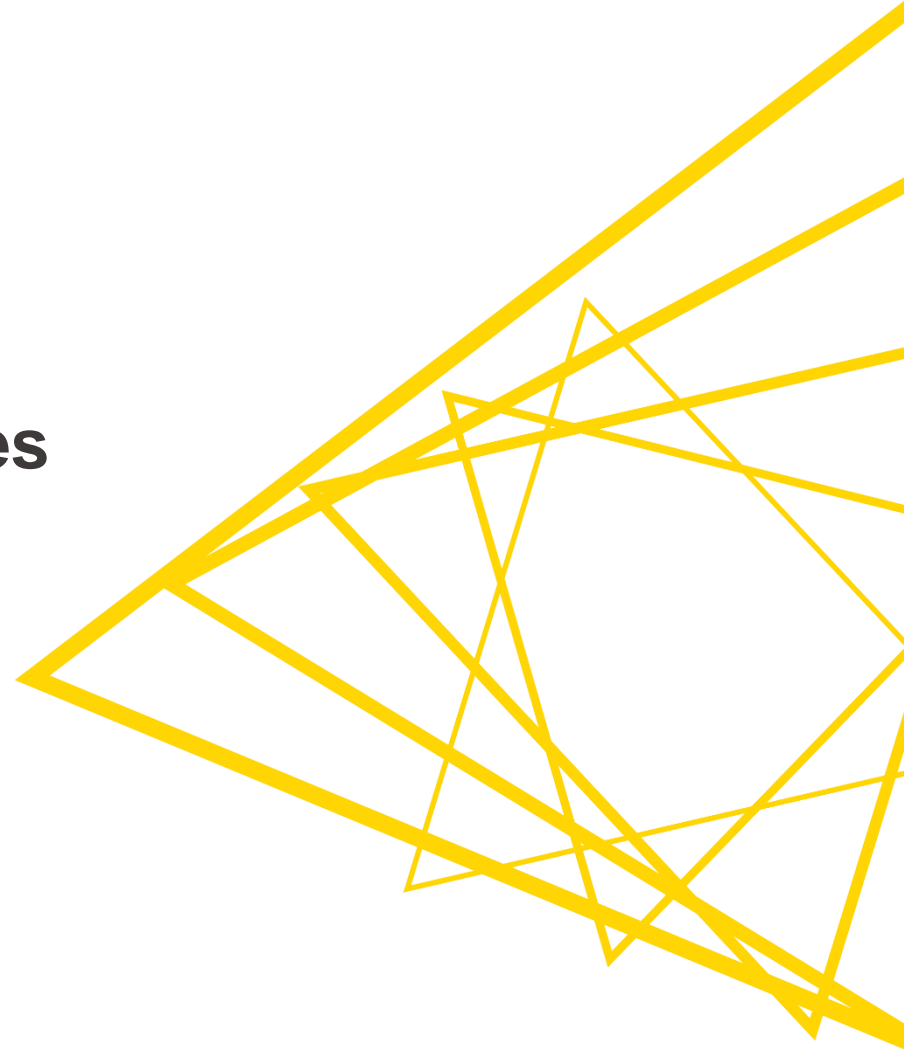
- ☒ RowFilter_TypeID
- ☒ ColumnName ☒ molecule_chembl_id
- ☒ include
- ☒ deepFiltering
- ☒ CaseSensitive
- ☒ Pattern
- ☒ hasWildCards
- ☒ isRegExpr

OK Apply Cancel ?

The Flow Variables tab

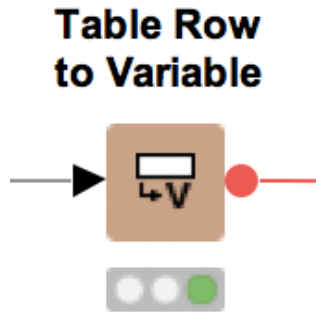
List of available Flow Variables

Ways to create Flow Variables

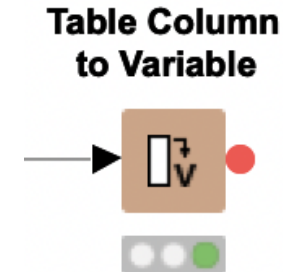


Nodes: From Table to Flow Variables

Takes a table as input and converts the first row to Flow Variables



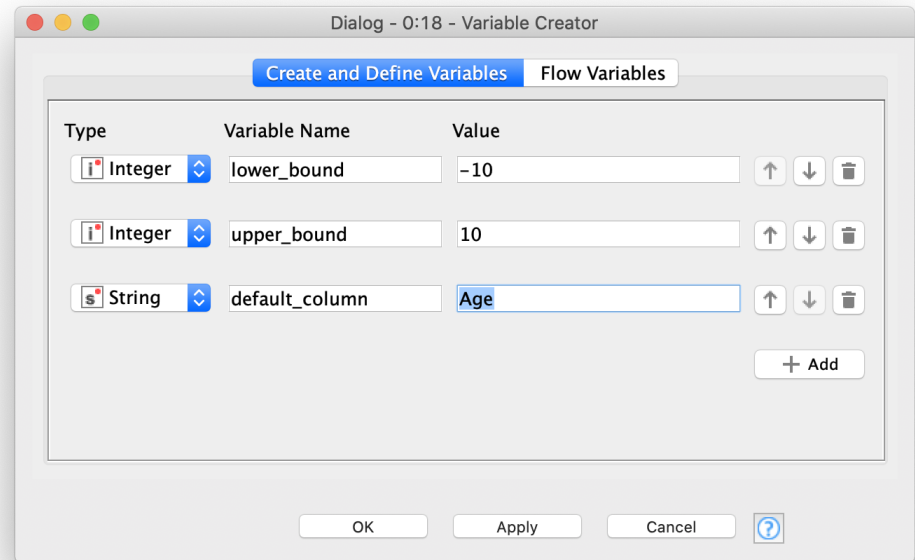
Takes a table as input and converts selected column to Flow Variables



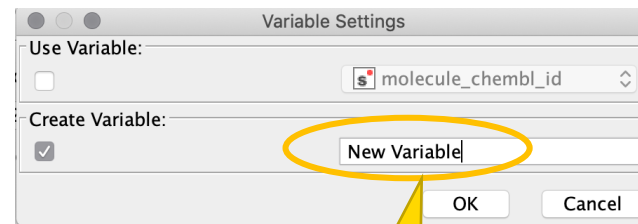
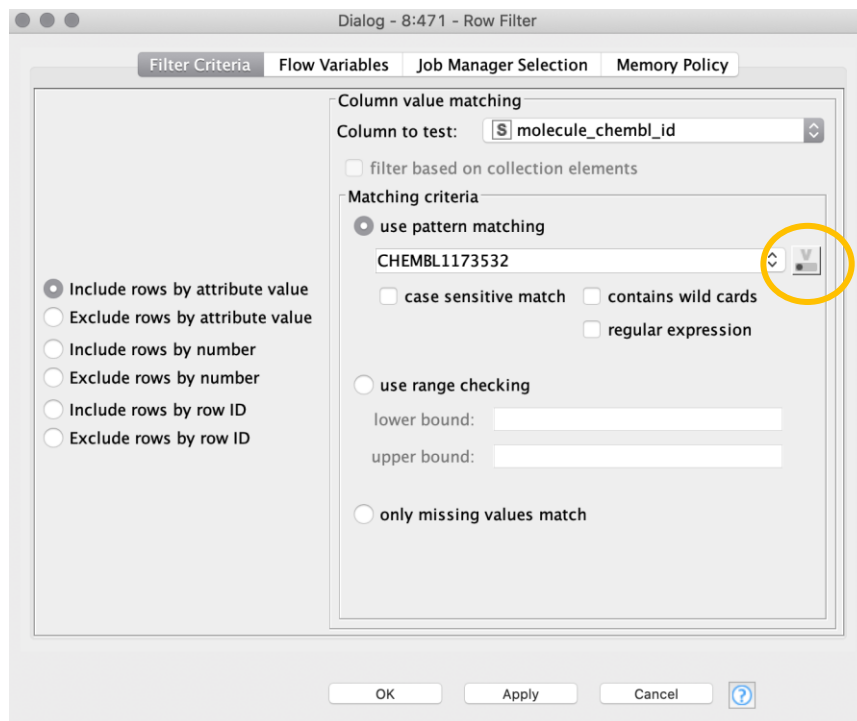
Create Flow Variables using the Variable Creator

- Allows to create flow variables of different types
- Click on “+ Add” to add a new variable and define a custom
 - Variable Name
 - Variable Value

Variable Creator



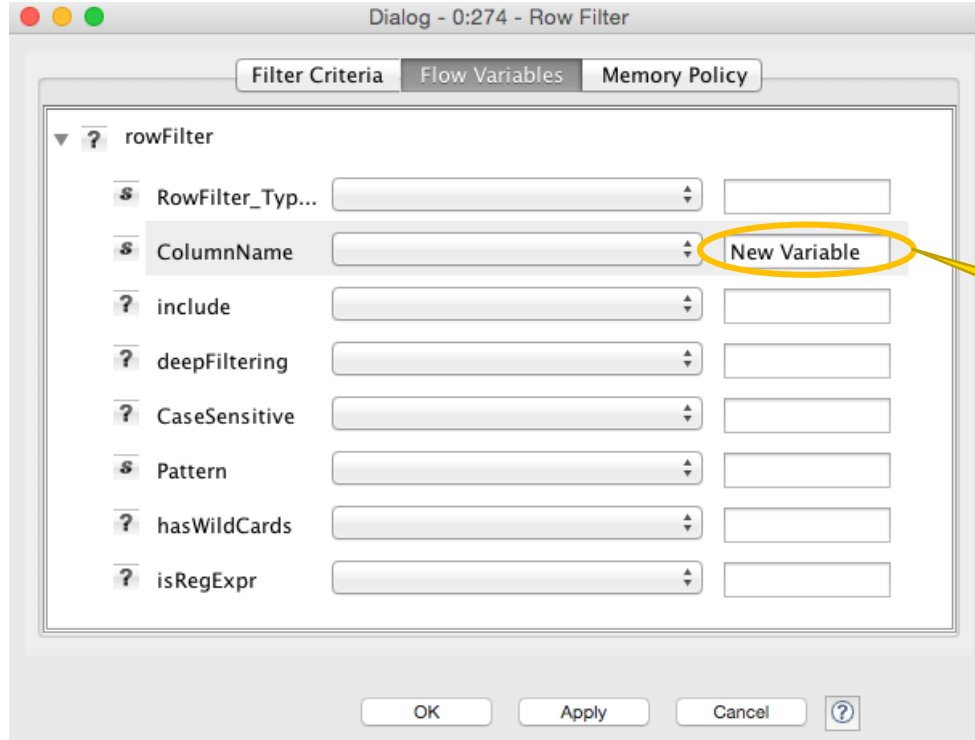
Create a Flow Variable (Button)



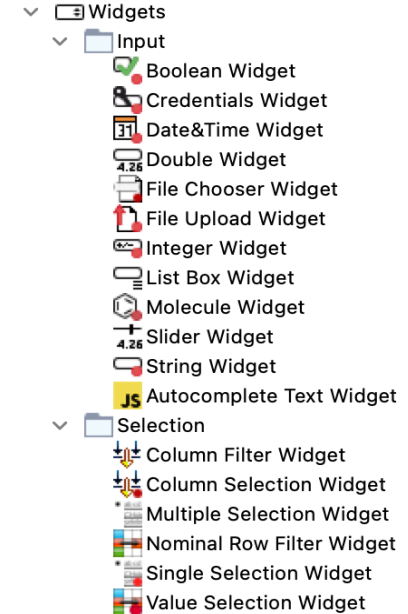
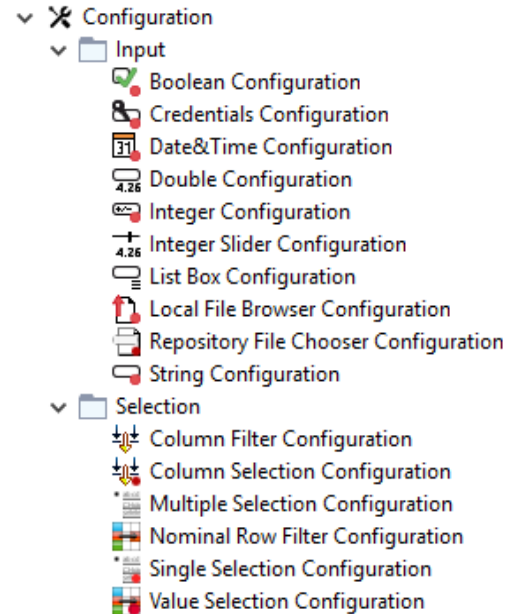
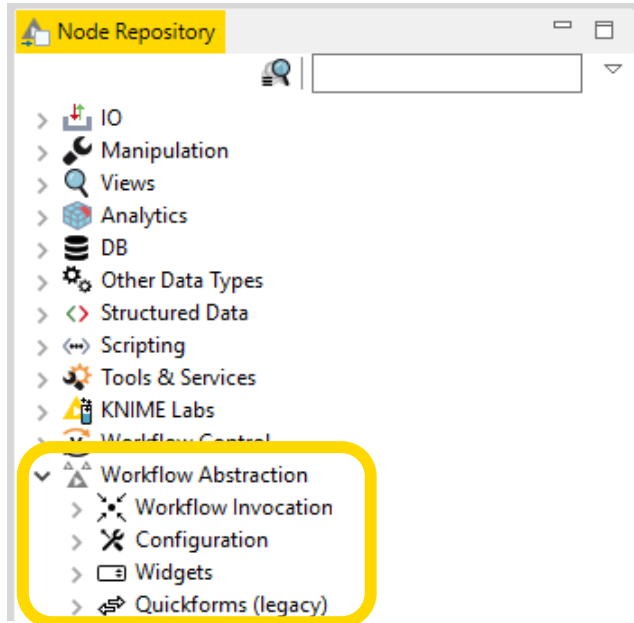
Name of the new
Flow Variable

Create a Flow Variable in any node

- Converting a setting value into a Flow Variable



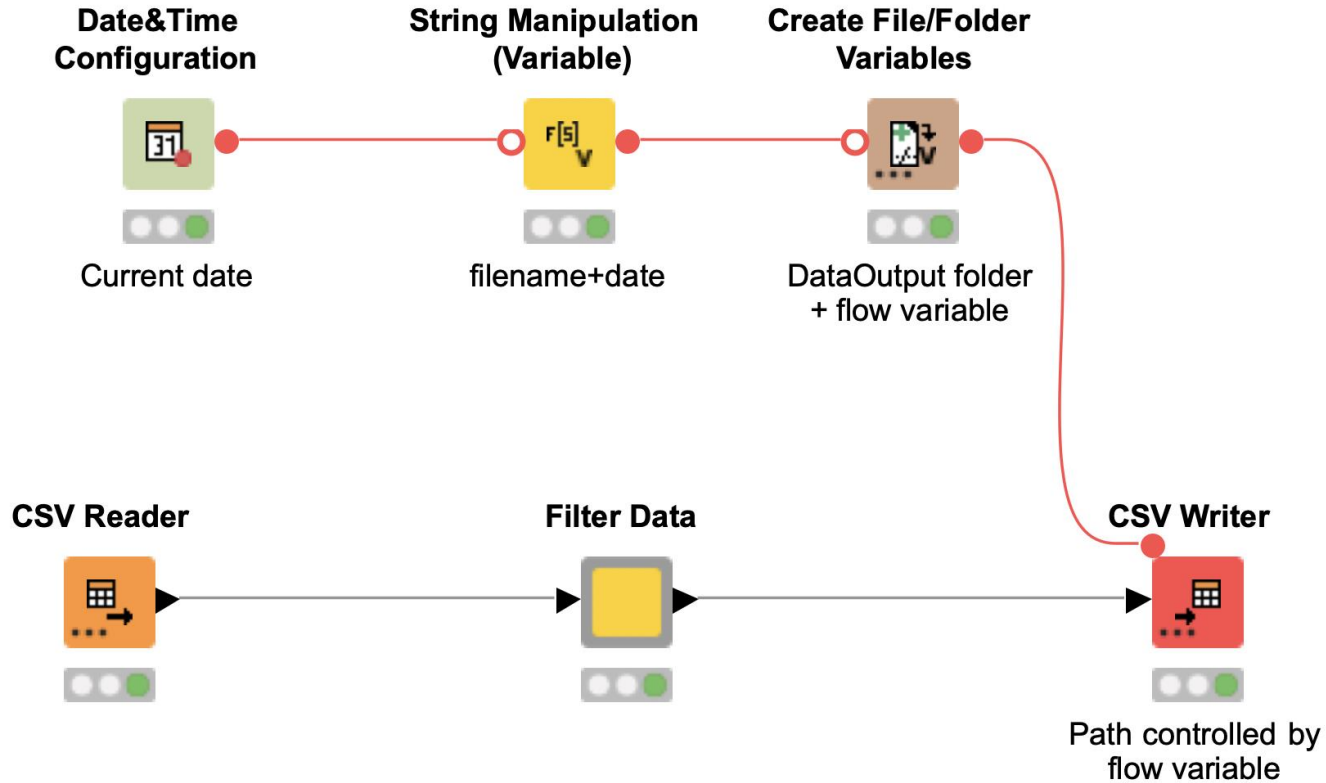
Create Flow Variables using Configuration/Widget Nodes



Usage Example: Configuration Node creates Flow Variable

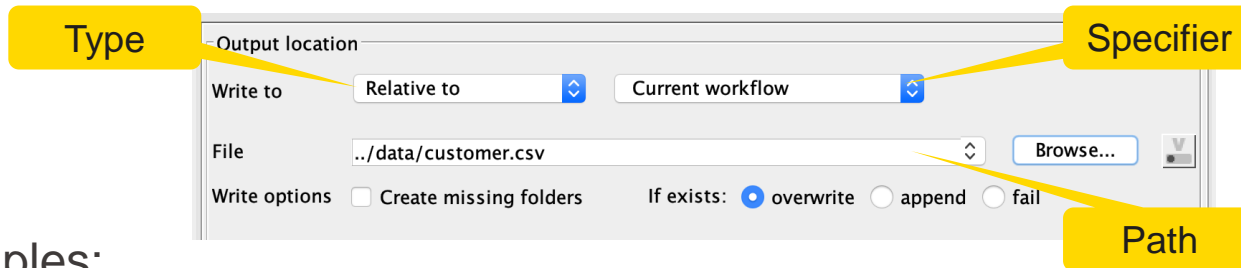


Example: Add Execution Date to File Name



Path Variables

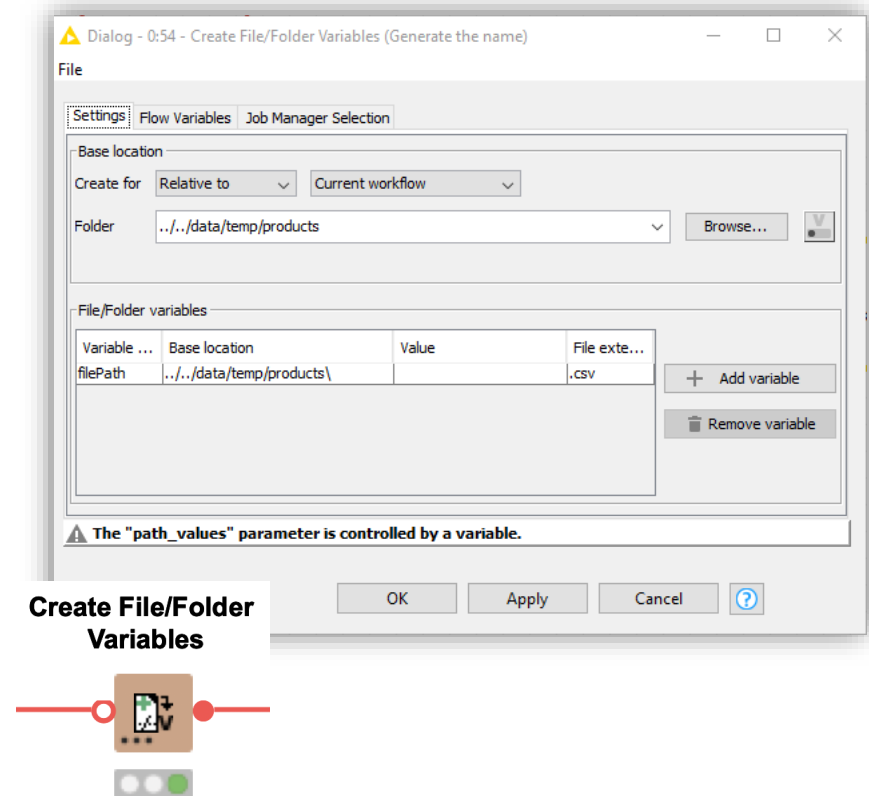
- Special flow variable type to point to a file or folder
 - E.g. to control output location of a file
- A path type consists of three parts:
 - **Type**: Specifies the file system type e.g. local, relative, mountpoint, custome_url or connected.
 - **Specifier**: Optional string with additional file system specific information e.g. relative to which location (knime.workflow)
 - **Path**: Specifies the location within the file system



- Examples:
 - (LOCAL, , C:\Users\username\Desktop)
 - (RELATIVE, knime.workflow, file1.csv)
 - (MOUNTPOINT, MOUNTPOINT_NAME, /path/to/file1.csv)
 - (CONNECTED, amazon-s3:eu-west-1, /mybucket/file1.csv)

Create File/Folder Variables

- Creates one or multiple path flow variable(s) pointing to files / folders
- Inputs:
 - Base location
 - Flow variable name(s)
 - Value (file name or path relative to base location)
 - File extension (optional)
- Output variables can be used to control the output location in writer nodes.



Goal of this Session

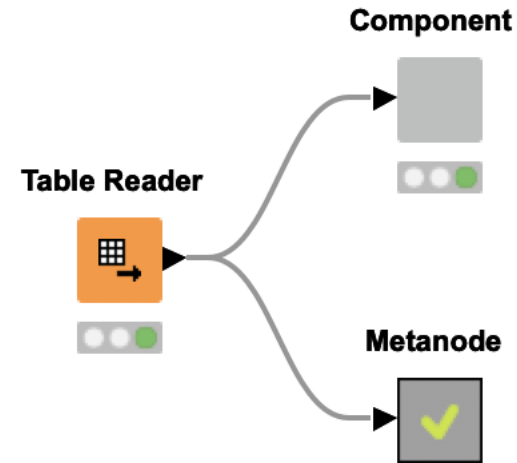
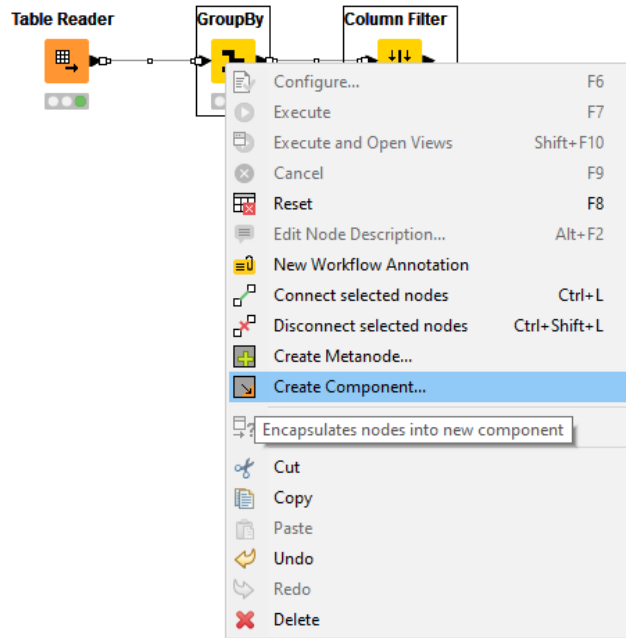
- What is a Flow Variable?
- How to create a Flow Variable?
- How to use a Flow Variable to overwrite node settings?
- How to use a Configuration node to parameterize a Component?
- How to use a Widget node to parameterize a Component via an interactive view?

Recap: Component vs. Metanode

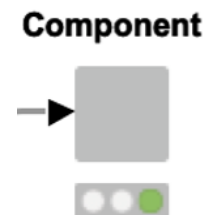
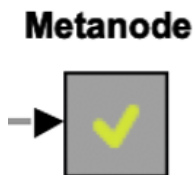


What are components?

- Components encapsulate functionalities that can be reused as your personal customized KNIME nodes, to perform tasks that you often repeat.
- They can also be shared with others via KNIME Hub and KNIME Server.

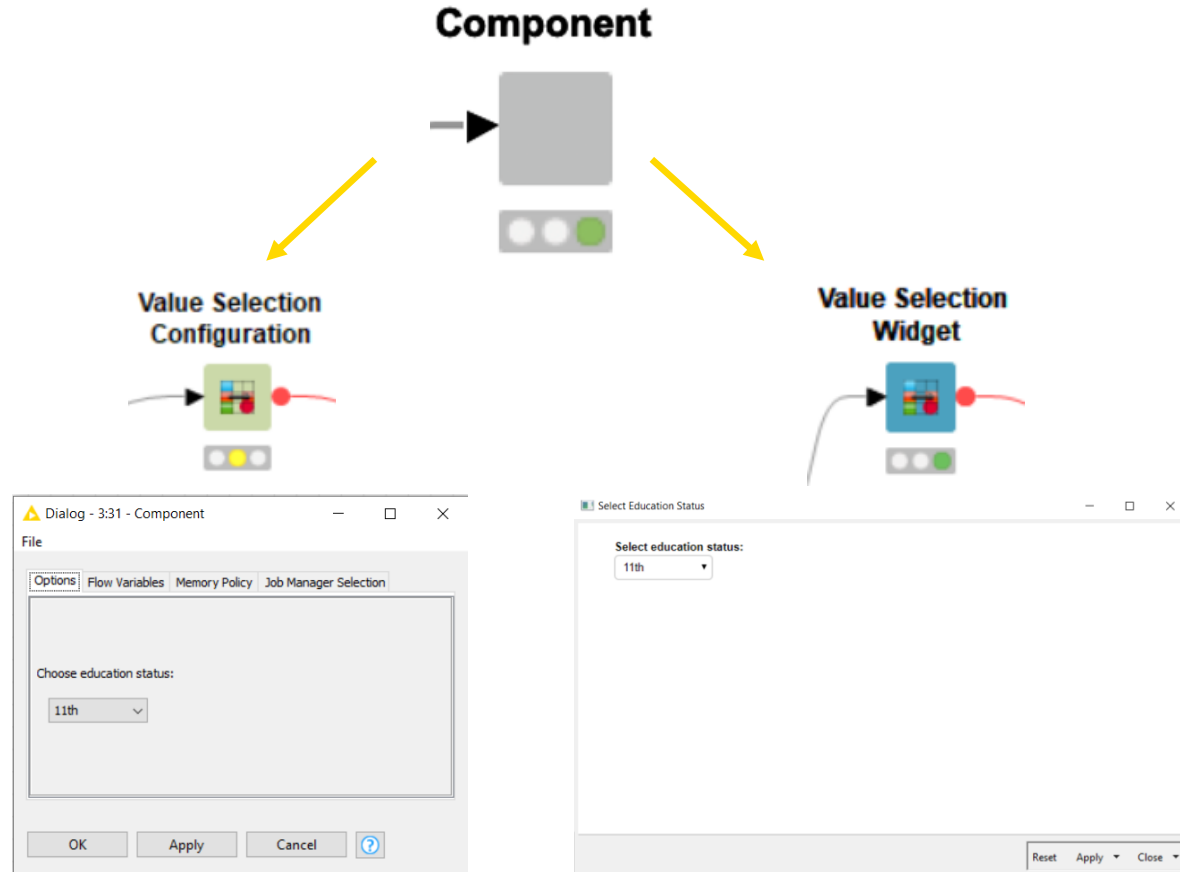


Metanodes vs. Components



	Metanodes	Components
Configuration	Not configurable	Via Configuration nodes (local workflow) and Widget nodes (KNIME WebPortal)
Variable scope	Global	Configurable: Local or global
Interactivity	Executed in the background	JavaScript Views and Widgets inside the component are shown in the Interactive View or a WebPortal page
Execution mode	Normal execution	Allows Simple Streaming execution
Recommended uses	Workflow cleaning	Enabling custom interactions, producing interactive views, sharing functionalities

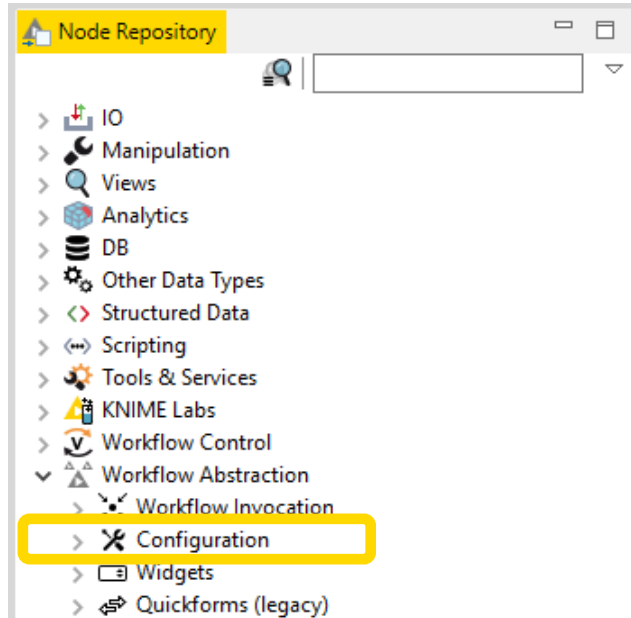
Recap: Configuration vs. Widget



Goal of this Session

- What is a Flow Variable?
- How to create a Flow Variable?
- How to use a Flow Variable to overwrite node settings?
- How to use a Configuration node to parameterize a Component?
- How to use a Widget node to parameterize a Component via an interactive view?

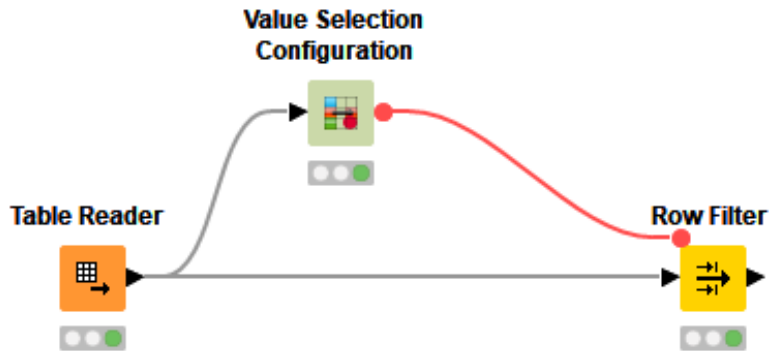
Configuration Nodes for Variable Creation and Output



- Configuration
 - Input
 - Boolean Configuration
 - Credentials Configuration
 - Date&Time Configuration
 - Double Configuration
 - Integer Configuration
 - Integer Slider Configuration
 - List Box Configuration
 - Local File Browser Configuration
 - Repository File Chooser Configuration
 - String Configuration
 - Selection
 - Column Filter Configuration
 - Column Selection Configuration
 - Multiple Selection Configuration
 - Nominal Row Filter Configuration
 - Single Selection Configuration
 - Value Selection Configuration

Configuration Nodes

- Use Configurations to create Flow Variables



Dialog - 8:488 - Value Selection Configuration

Control | **Flow Variables** | Job Manager Selection

Label:

Description:

Parameter/Variable Name:

Selection Type:

Lock Column: ☒

Default Column:

Default Value:

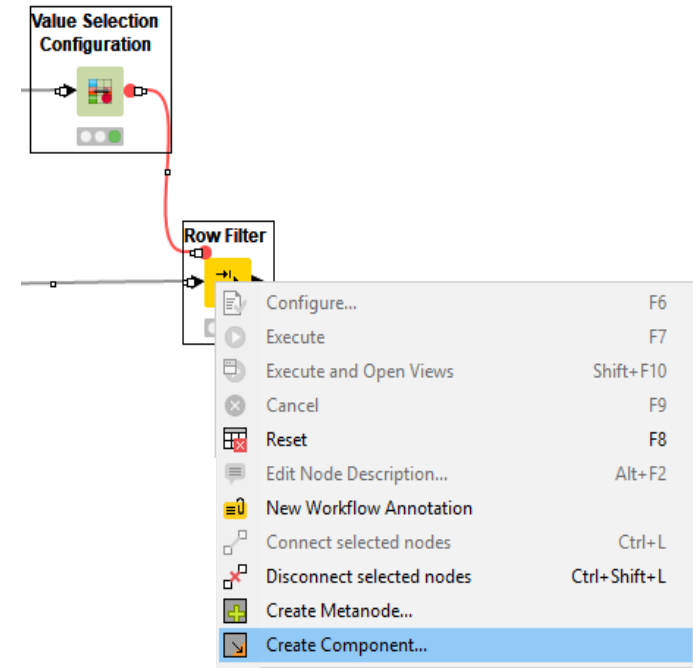
Limit number of visible options: ☐

Number of visible options:

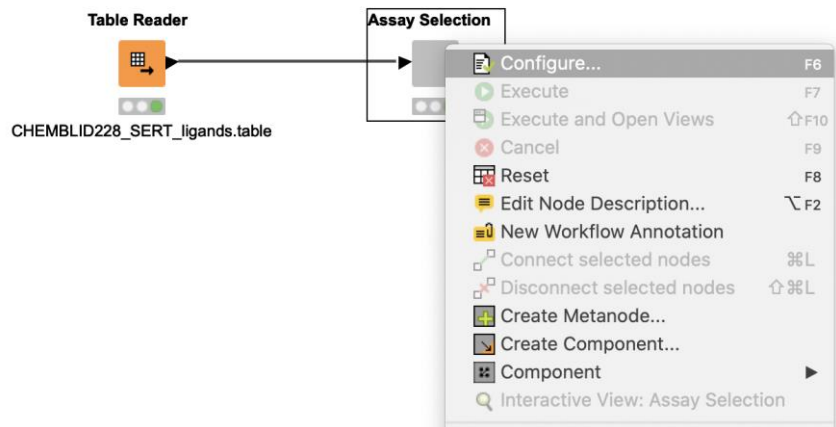
OK Apply Cancel ?

Create a Component

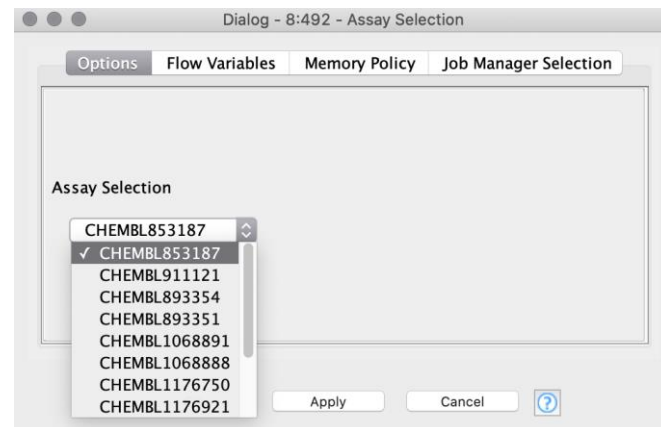
- Select nodes to encapsulate into a Component
- Right click a node
- Select “Create Component...”



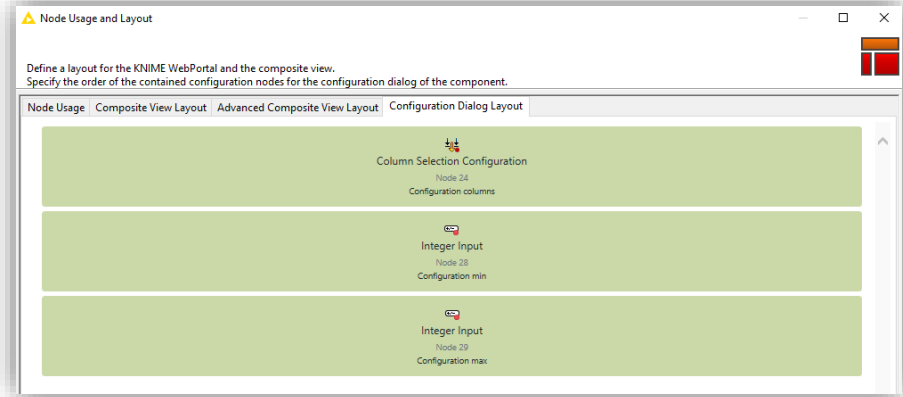
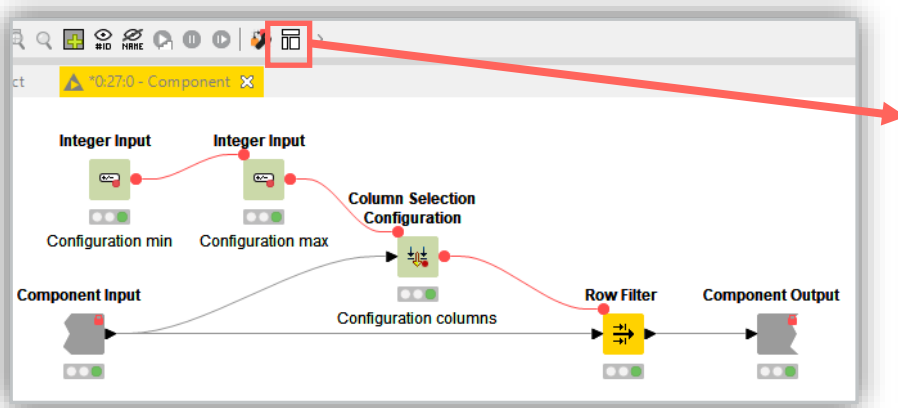
Simple Configuration of Component



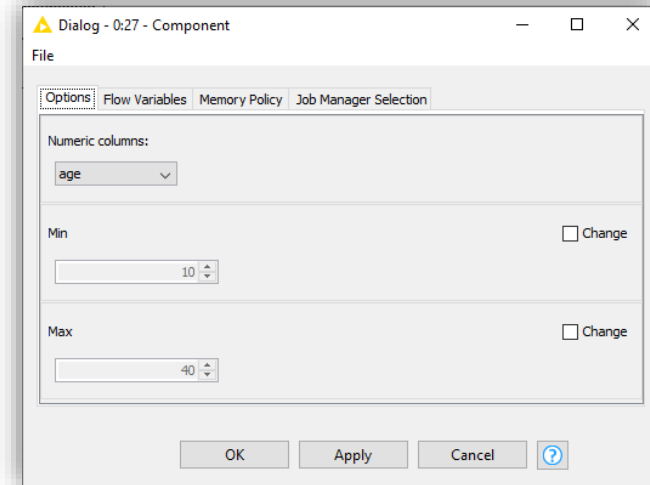
- Double click on Component to configure
- For use in Webportal replace Configuration nodes by Widget nodes



Configuration Dialog Layout



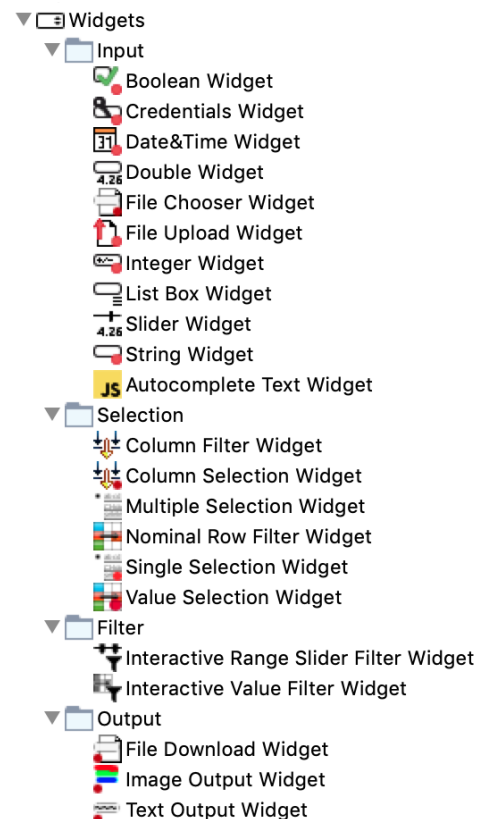
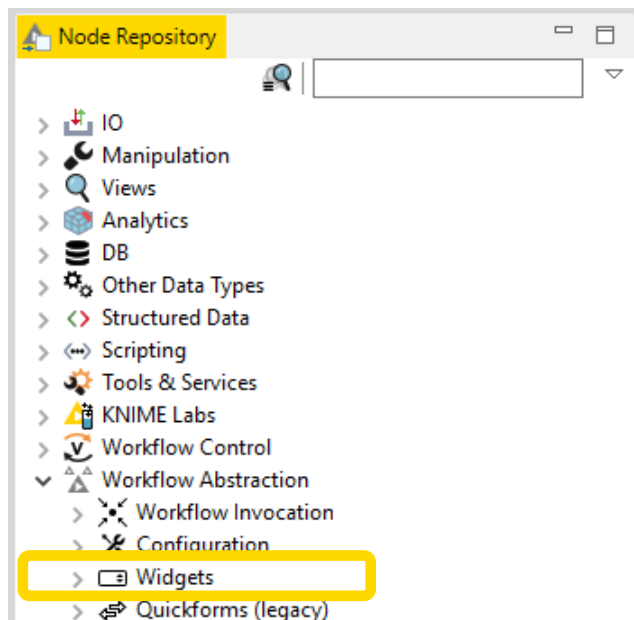
- Click layout button when inside component to modify the order of the setting options in configuration window of the component



Goal of this Session

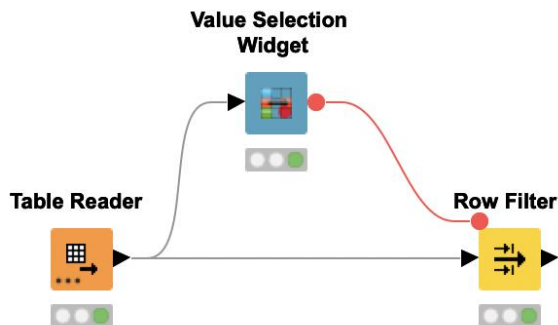
- What is a Flow Variable?
- How to create a Flow Variable?
- How to use a Flow Variable to overwrite node settings?
- How to use a Configuration node to parameterize a Component?
- How to use a Widget node to parameterize a Component via an interactive view?

Widget Nodes for Variable Creation and Output



Widget Nodes

- Use Widget to create Flow Variables



Dialog - 3:504 - Value Selection Widget

Control Re-execution Flow Variables

Label:

Description:

Variable Name:

Selection Type:

Lock Column: ☒

Default Column:

Default Value:

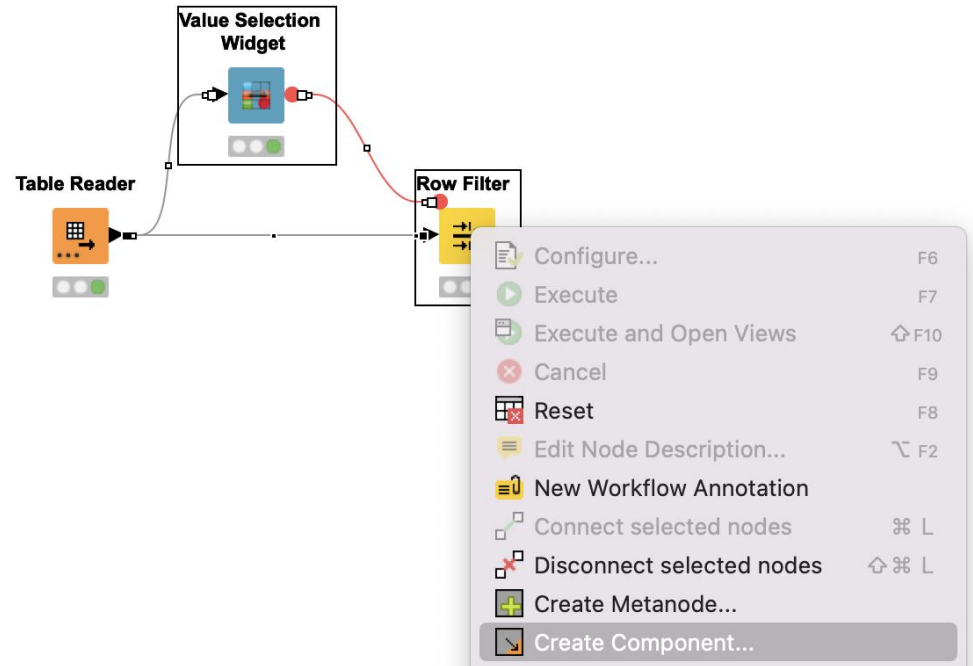
Limit number of visible options: ☐

Number of visible options:

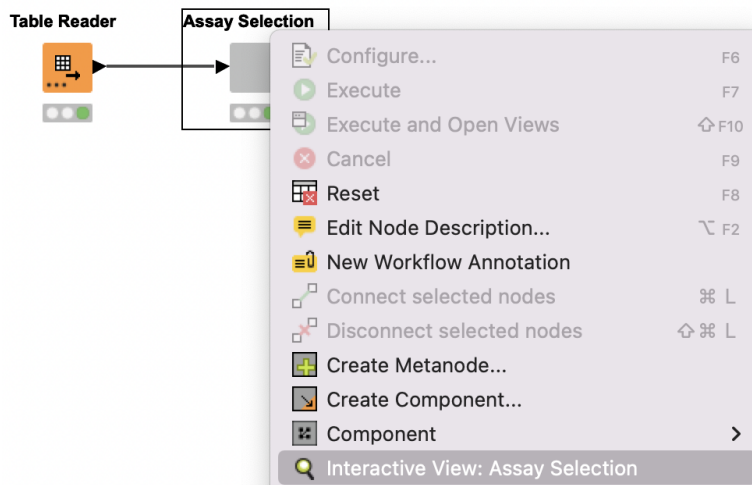
OK - Execute Apply Cancel ?

Create a Component

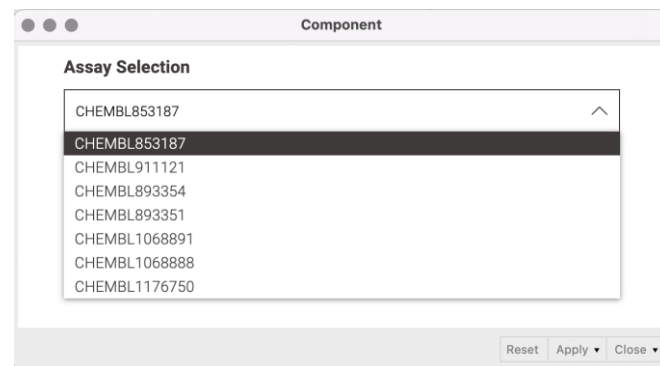
- Select nodes to encapsulate into a Component
- Right click a node
- Select “Create Component...”



Interactive View of Component



- Right click on Component to show Interactive View
- Select options in view or provide input

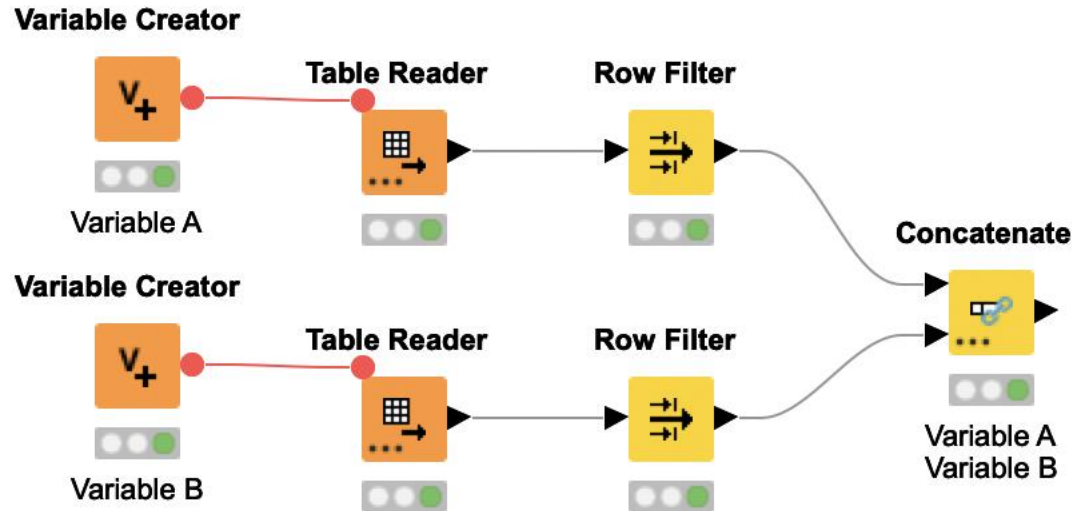


Key Features: Flow Variables

- Flow Variables are workflow parameters used to overwrite existing node settings
- Flow Variables can be of type String, Integer, Double, Boolean, Long, Array and Path
- Flow Variables can be created
 1. Using the "Table Row/Column to Variable"
 2. In the "Flow Variable" tab of any node
 3. Using the "Variable Creator" node
 4. Using Configuration and Widget nodes
- A Flow Variable is carried along workflow branches (parallel branches don't share local Flow Variables)

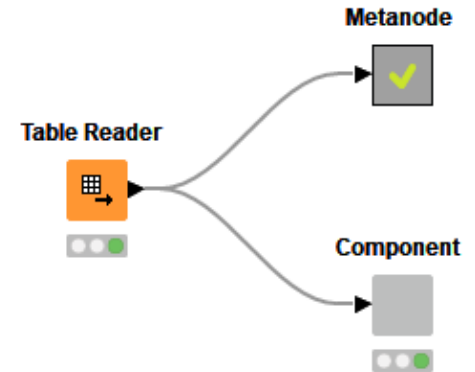
Flow Variables carried along on Branches

- Flow Variable A is carried along the top branch
- Flow Variable B is carried along the bottom branch
- In the Concatenate Node both Variables are available



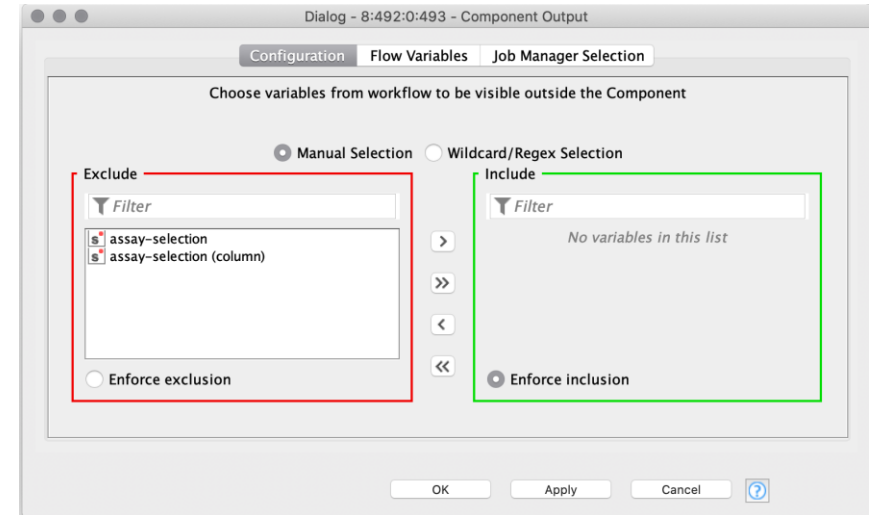
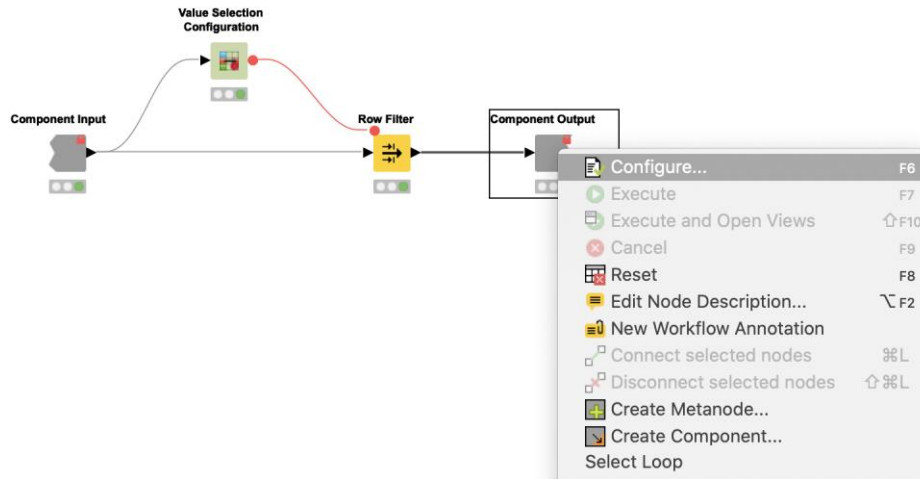
Components/Metanodes: local/global Flow Variables

- Flow Variables that are created inside of a Metanode are globally available along the branch
- Flow Variables that are created inside of a Component are only available locally
- UNLESS: you actively make them available globally along the branch

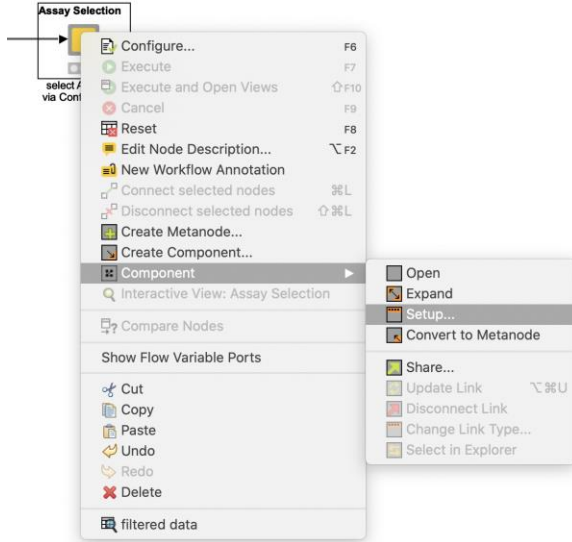


Passing Variable from Components

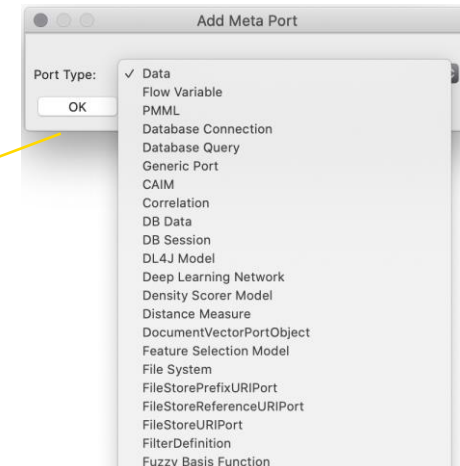
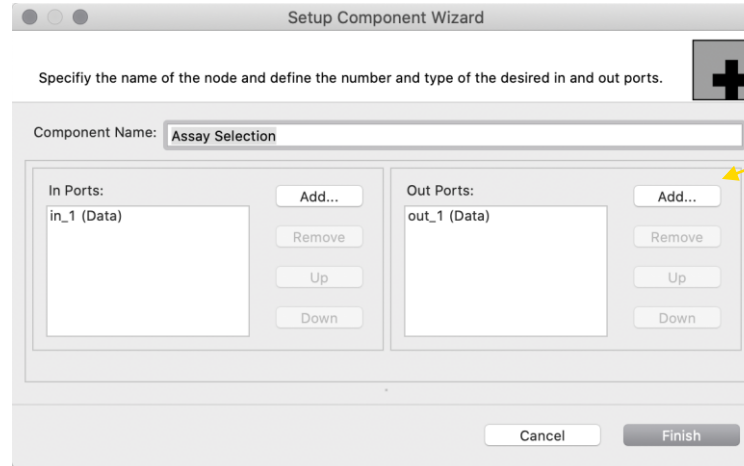
- Flow Variables by default available locally inside Component
- Configure Component Input/Output to pass Flow Variables from/to outside Component



Configure Component Ports



- Add input and output points to Metanodes/Components
- Remove ports to adapt to changes after creation of Metanodes/Components



Component Description

- Make your component look like a KNIME node

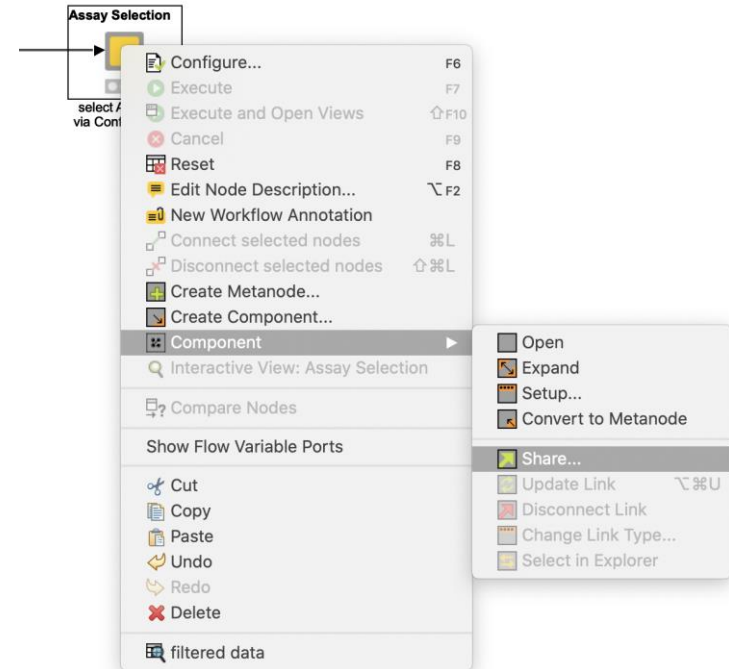
The image displays two screenshots of the KNIME software interface, illustrating how to configure and describe a custom component.

Top Screenshot: Shows the 'Assay Selection' component configuration window. The 'Description' field contains the text: 'This component filters the data by the selected AssayID'. The 'Component Icon' section shows a dropdown menu with options: Learner, Manipulator (selected), Predictor, Sink, Source, and Visualizer. Below this, the 'Port #1' configuration shows 'Name: all data' and 'Description: Data containing records for all AssayIDs'. The 'Output Port #1' configuration shows 'Name: filtered data' and 'Description: Data containing records for the selected AssayID'. A yellow callout points to the 'Description' field with the text: 'Add description of the component'. Another yellow callout points to the 'Port #1' configuration with the text: 'Add description of the input and output ports'. A third yellow callout points to the 'Component Icon' dropdown with the text: 'Add background color or icon'.

Bottom Screenshot: Shows the 'Assay Selection' component description window. The 'Description' field contains the text: 'This component filters the data by the selected AssayID'. Below this, the 'Dialog Options' section shows 'Assay Selection' and 'Select AssayID to filter data'. The 'Ports' section shows 'Input Ports' and 'Output Ports'. The 'Input Ports' section shows '0 Data containing records for all AssayIDs'. The 'Output Ports' section shows '0 Data containing records for the selected AssayID'. A yellow callout points to the 'Description' field with the text: 'Add description of the component'. Another yellow callout points to the 'Port #1' configuration with the text: 'Add description of the input and output ports'. A third yellow callout points to the 'Component Icon' dropdown with the text: 'Add background color or icon'.

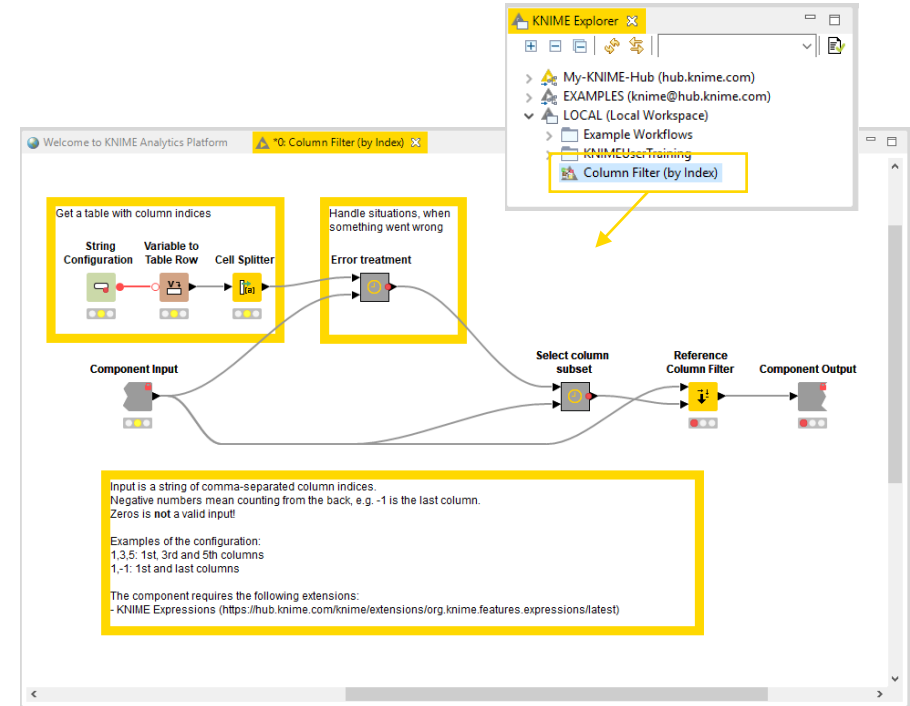
What is a shared component?

- Components can be saved in your KNIME workspace, KNIME Server or the KNIME Hub for later reuse
- To do this, simply right-click any Component and select “Share...”
- Shared Components are read-only instances of a Component
- Public Shared Components are available on EXAMPLES Server and on the KNIME Hub



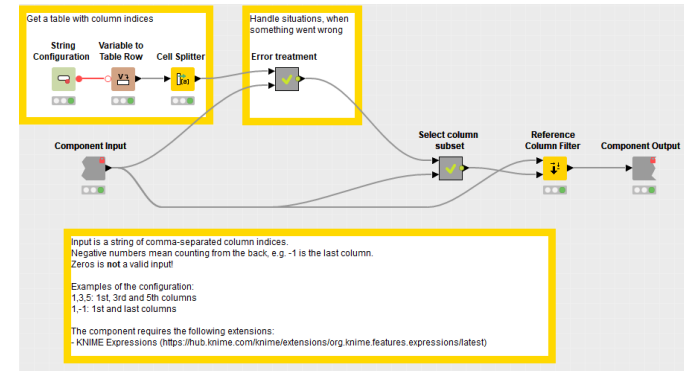
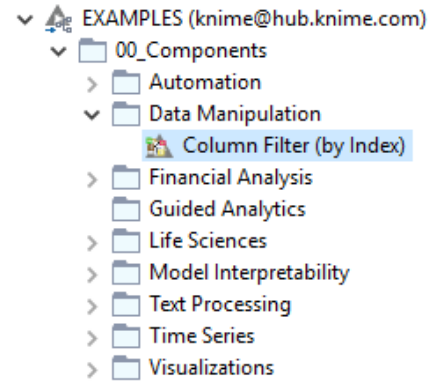
How can you edit a shared component?

- Components can be edited using the Component Editor similar to workflows
- To edit a Component using the Component Editor, double-click the Component in its location in the KNIME Explorer
- To ensure Components are executable when opened in the Component Editor, chose the option “Include input data with component” when sharing it



How can you use a shared component?

- To use a Shared Component, drag and drop it to the workflow editor
- Instances of Shared Components can be updated either manually or when workflow is opened
- Shared Components can also be unlinked from its original location, which makes it editable in the workflow directly
- Update Shared Components by overwriting them



Flow Variable Exercise

Open Exercise 05. Flow Variable

Activity I: Flow Variables

- *Filter the compound data to*
 1. *Contain the compounds that were tested in the highest number of assays*
 2. *Contain only compounds tested for AssayID "ChEMBL853197"*

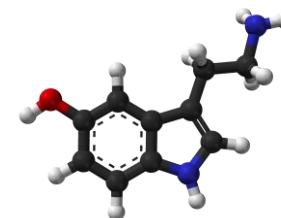
Activity II: Using Flow Variables for Substructure Search

- Pass a drawn substructure from the Molecular Sketcher component as a Flow Variable to the RDKit Substructure Filter node. Write the found molecules into an Excel file, which contains the execution time in the file name.

Exercise Data

- Serotonin transporter (SERT) ligand data from ChEMBL
 - one target ID
 - one target organism: homo sapiens
 - 7 different documents IDs
 - 12 different assays IDs
 - 327 different molecules IDs
 - K_i values (inhibitor constant describe the binding affinity of molecule to target)
 - Range: 0.043 nM - 9120000 nM

Serotonin



Exporting Data & Deployment



Exporting Data

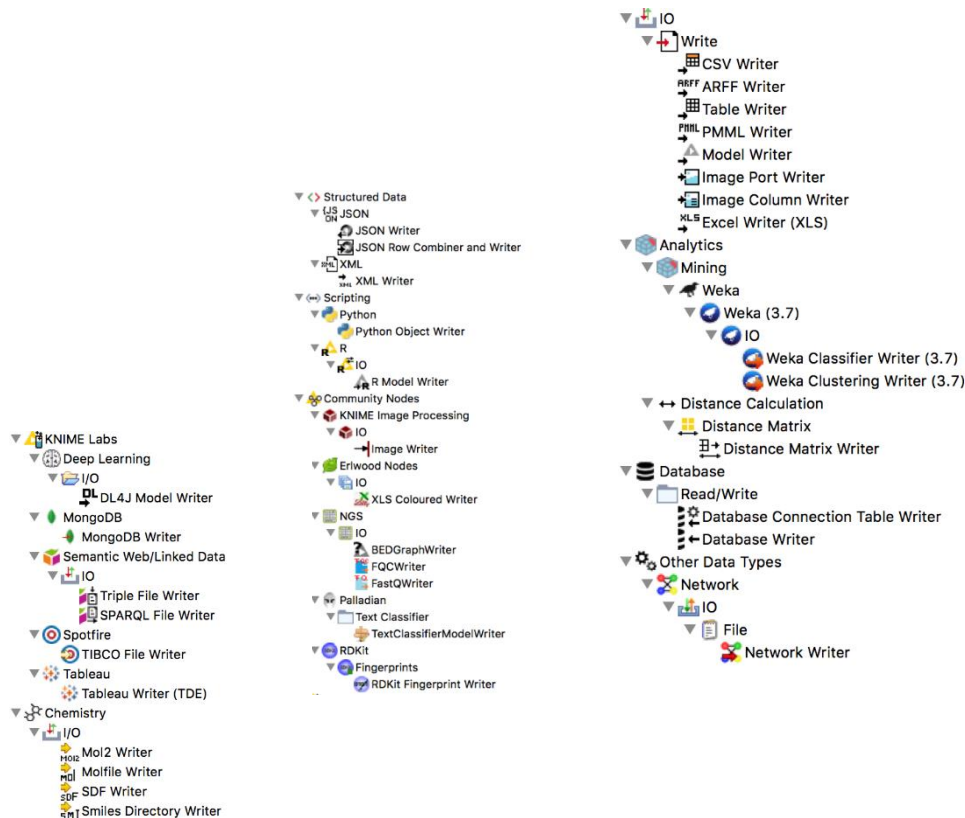
After an analysis is completed, what next?

- 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

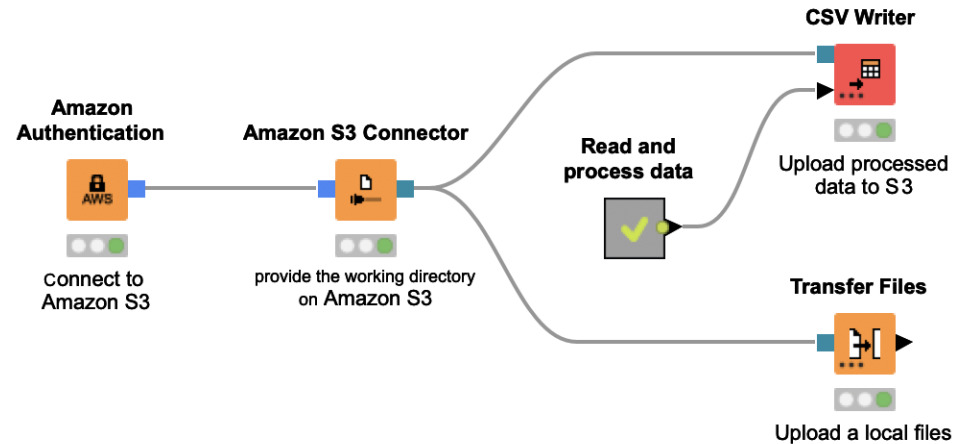
Typically characterized by:

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

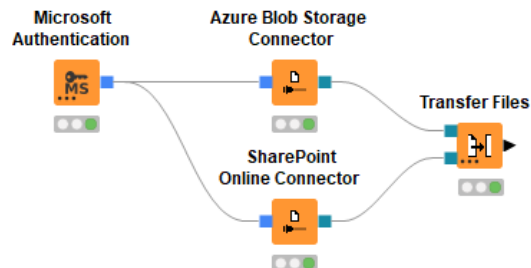


Write Files to a Remote File System

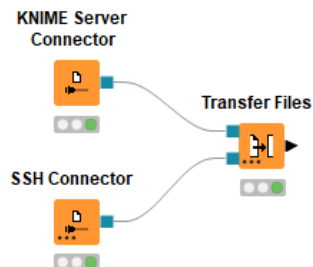
- The new file handling framework makes it easy to upload data to remote file systems
 - Write processed data directly with a writer node
 - Upload local files with the Transfer Files node
- Supported file systems
 - Microsoft Azure
 - Google
 - Amazon
 - Databricks
 - BigData file systems (hdfs, httpFS,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



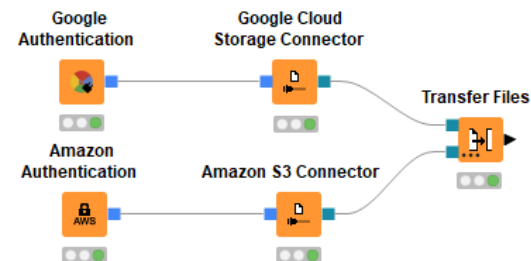
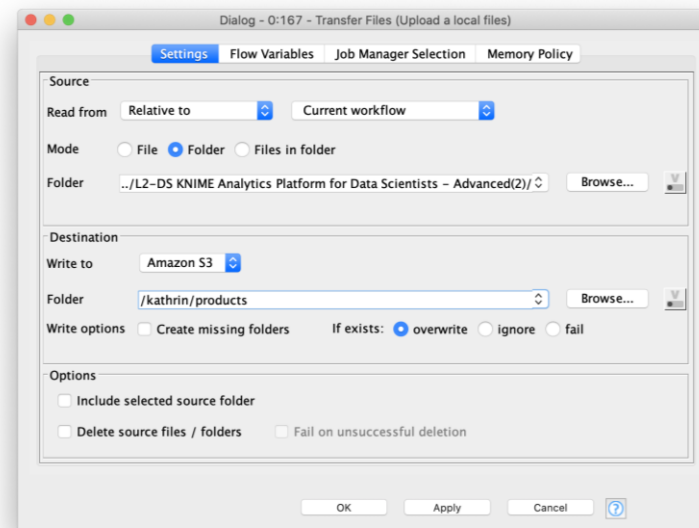
Full Flexibility with the Transfer Files node



Same cloud environment



On-premise

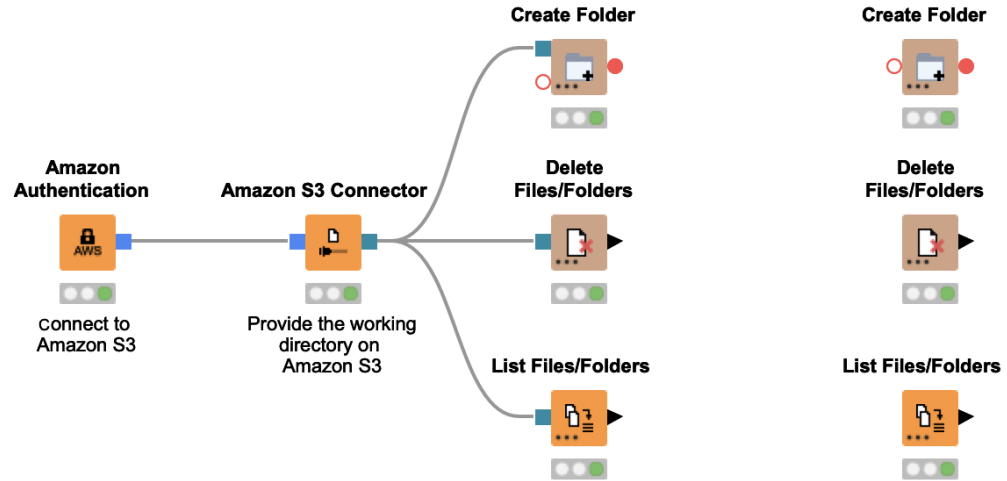


Cross cloud environments

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

Thank You!
KNIME AG

