

# [L1-DW] KNIME Analytics Platform for Data Wranglers: Basics

KNIME AG

You are free to:

**Share**

copy and redistribute the material  
in any medium or format

**Adapt**

remix, transform, and build upon  
the material for any purpose, even  
commercially.

You must give **appropriate credit**, provide  
a link to the license, and indicate if  
changes were made. You may do so in  
any reasonable manner, but not in any  
way that suggests the licensor endorses  
you or your use.

# Structure of the course

---

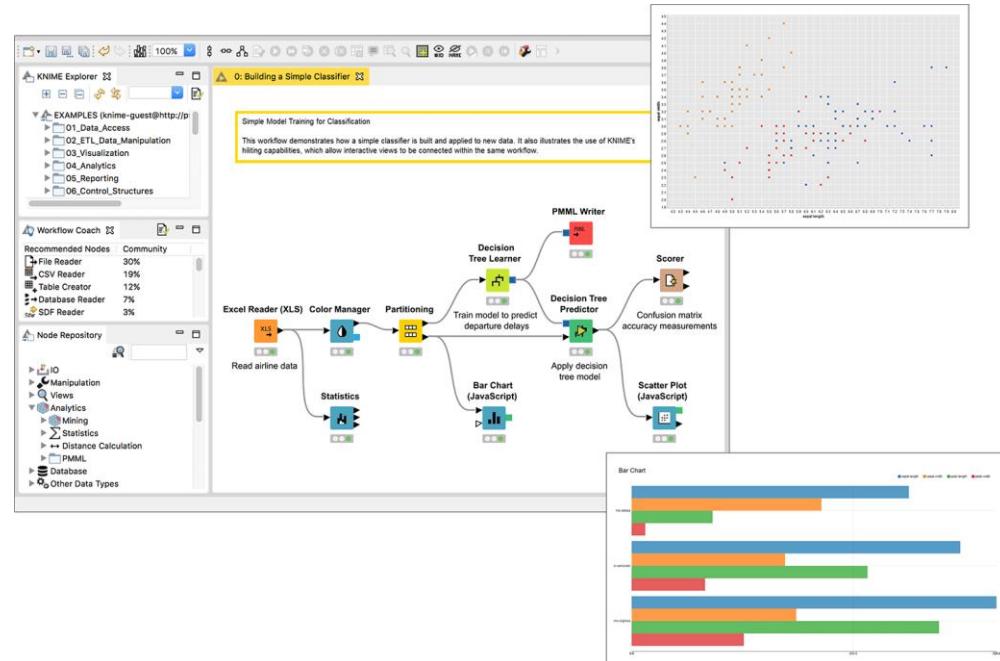
- This course consists of five sessions
  - Introduction to KNIME Analytics Platform and Reading Files
  - Data Merging and Data Cleaning I
  - Data Cleaning II and Data Transformation
  - Data Aggregation and Visualization
  - Q&A and Summary
- Structure of each session
  - Discussion of past exercises
  - Course
  - Introduction of next exercises
- Outside of class time:
  - Exercises

# **Overview**

## **KNIME Analytics Platform**

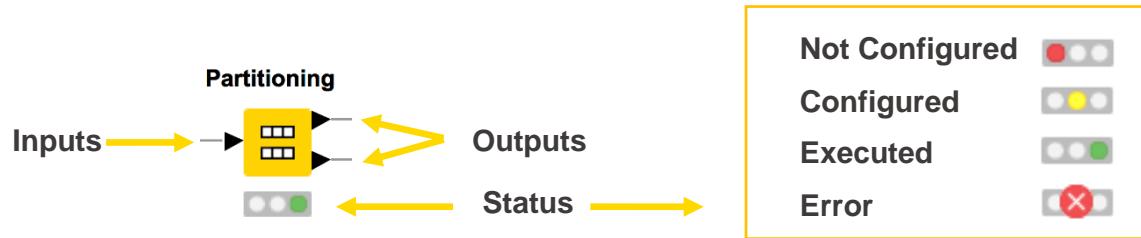
# What is KNIME Analytics Platform?

- A tool for data analysis, manipulation, visualization, and reporting
- Based on a graphical interface
- Provides a diverse array of extensions:
  - Text mining
  - Network mining
  - Cheminformatics
  - Many integrations, such as Java, R, Python, Weka, Keras, Plotly, H2O, etc.

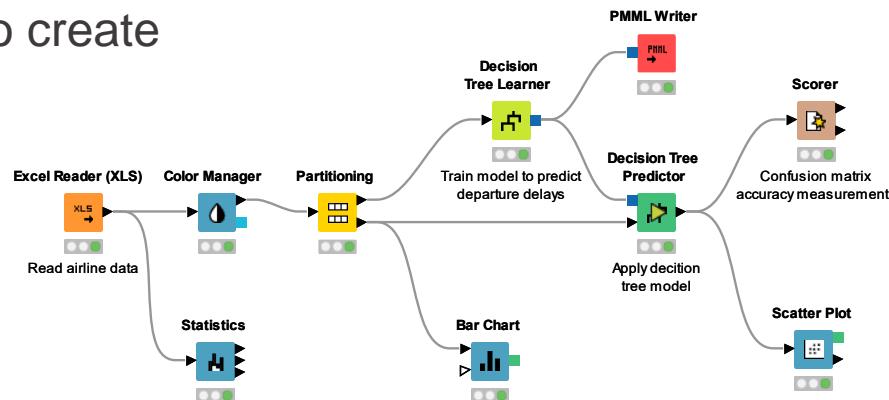


# Visual KNIME Workflows

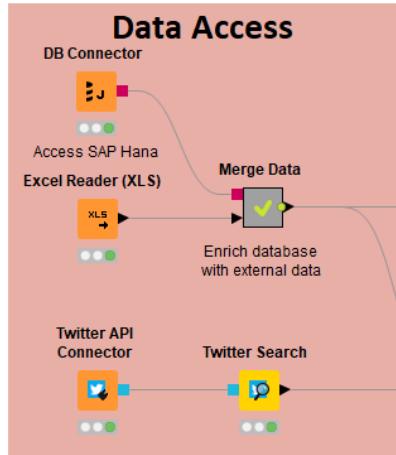
**NODES** perform tasks on data



Nodes are combined to create  
**WORKFLOWS**

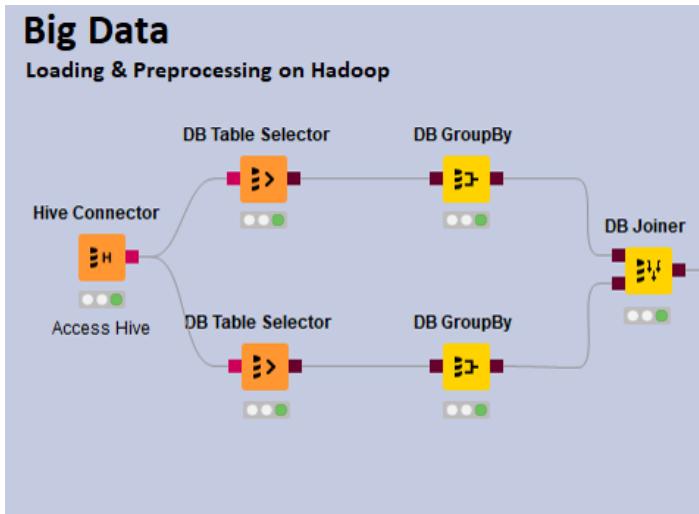


# Data Access



- **Databases**
  - MySQL, PostgreSQL, Oracle
  - Theobald
  - any JDBC (DB2, MS SQL Server)
  - Amazon DynamoDB
- **Files**
  - CSV, txt, Excel, Word, PDF
  - SAS, SPSS
  - XML, JSON, PMML
  - Images, texts, networks
- **Other**
  - Twitter, Google
  - Amazon S3, Azure Blob Store
  - Sharepoint, Salesforce
  - Kafka
  - REST, Web services

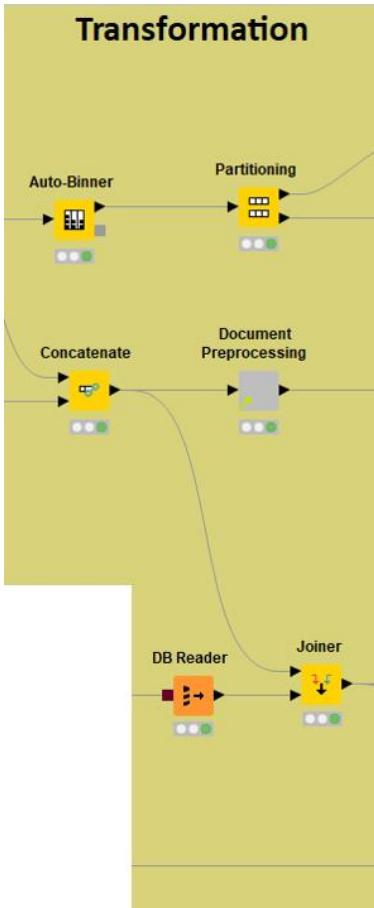
# Big Data



- Spark & Databricks
- HDFS support
- Hive
- Impala
- In-database processing

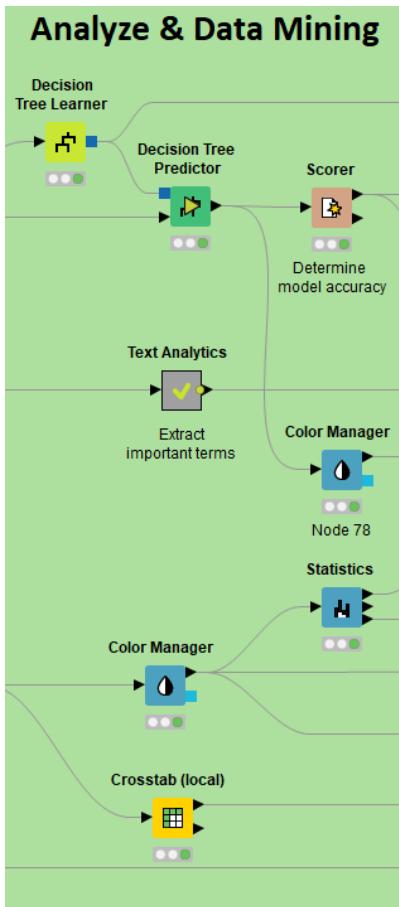


# Transformation



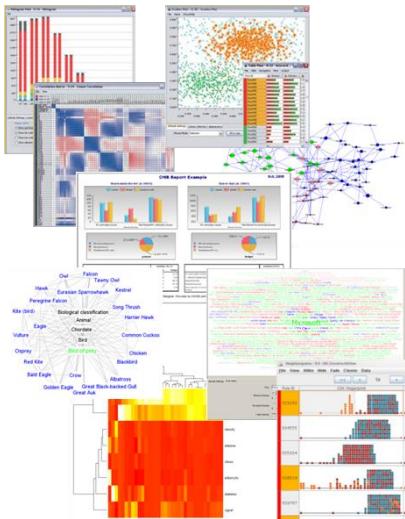
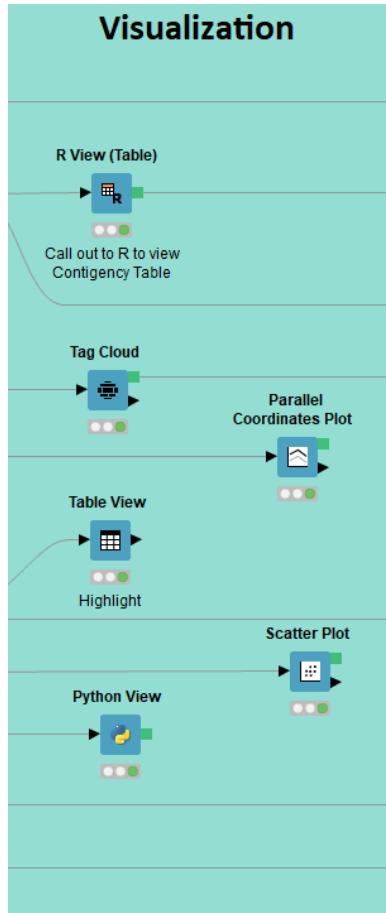
- Preprocessing
  - Row, column, matrix based
- Data blending
  - Join, concatenate, append
- Aggregation
  - Grouping, pivoting, binning
- Feature creation and selection

# Analysis & Data Mining



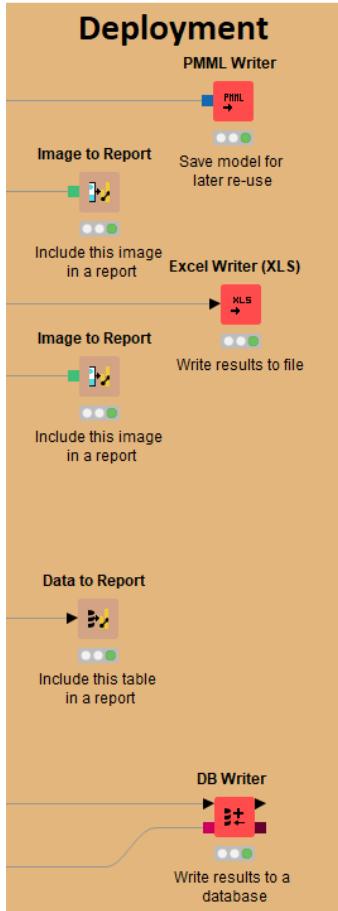
- Regression
  - Linear, regression tree
- Classification
  - Decision tree, ensembles, SVM, MLP, Naïve Bayes, logistic regression
- Clustering
  - k-means, DBSCAN, hierarchical
- Validation
  - Cross-validation, scoring, ROC
- Deep Learning
  - Keras, DL4J
- External
  - R, Python, Weka, H2O, Keras

# Visualization



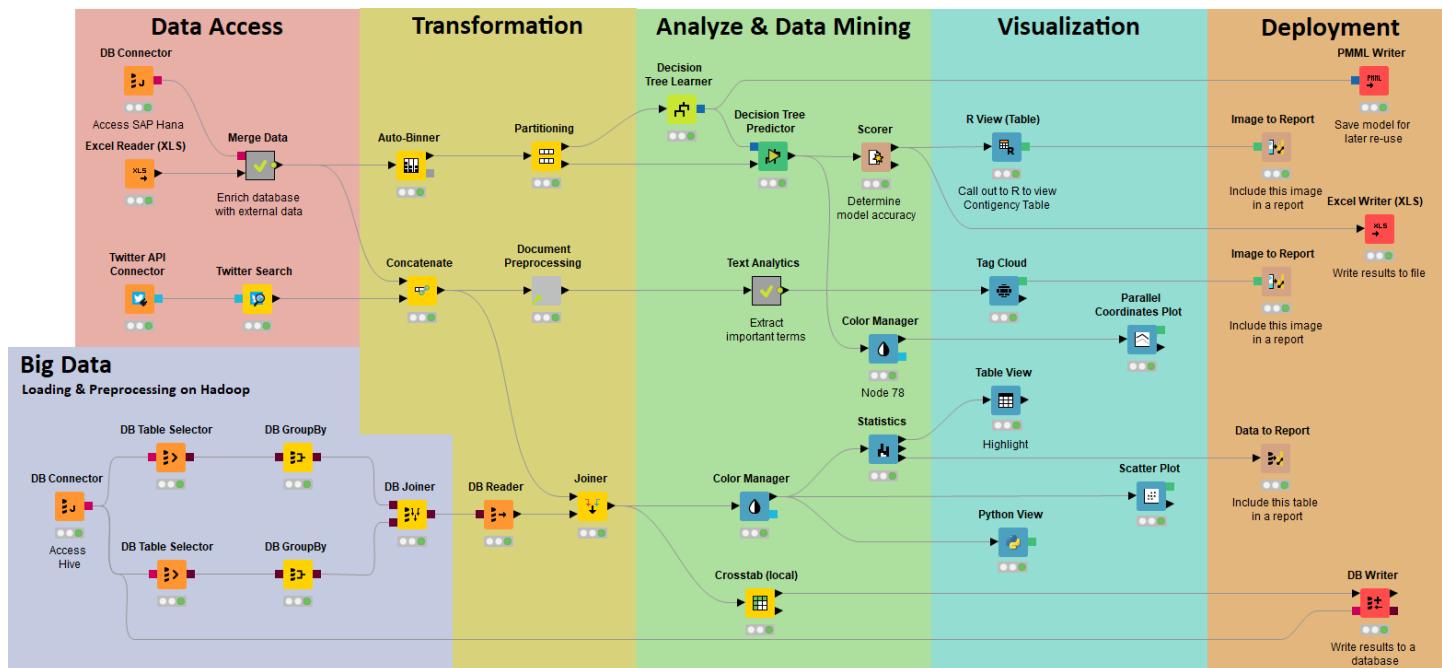
- Interactive visualizations
- JavaScript-based nodes
  - Scatter Plot, Box Plot, Line Plot
  - Networks, ROC Curve, Decision Tree
  - Plotly Integration
  - Adding more with each release!
- Misc
  - Tag cloud, open street map, molecules
- Script-based visualizations
  - R, Python

# Deployment



- Database
- Files
  - Excel, CSV, txt
  - XML
  - PMML
  - to: local, KNIME Business Hub, Amazon S3, Azure Blob Store
- BIRT reporting

# Over 4000 Native and Embedded Nodes Included:



## Data Access

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

## Transformation

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

## Analysis & Mining

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

## Visualization

R  
JFreeChart  
JavaScript  
Plotly  
Community / 3rd

## Deployment

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

# Install KNIME Analytics Platform

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

## Windows

KNIME Analytics Platform for Windows (installer)  
*The installer adds an icon to the desktop and suggests suitable memory settings*

[Download](#) (459 MB)

KNIME Analytics Platform for Windows (self-extracting archive)  
*The self-extracting archive only creates a folder holding the KNIME installation*

[Download](#) (463 MB)

KNIME Analytics Platform for Windows (zip archive)

[Download](#) (547 MB)

## Linux

KNIME Analytics Platform for Linux

[Download](#) (583 MB)

## Mac

KNIME Analytics Platform for macOS (10.13 and above)

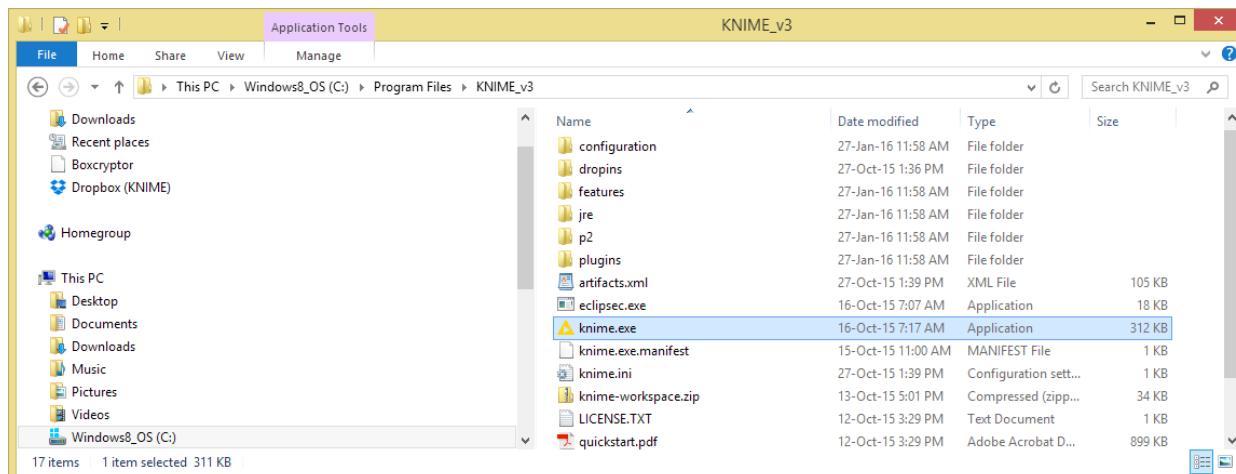
[Download](#) (438 MB)

Find out what's new in the latest KNIME 4.4 release [here](#).

If you are interested in a previous version of KNIME Analytics Platform, please click [here](#).

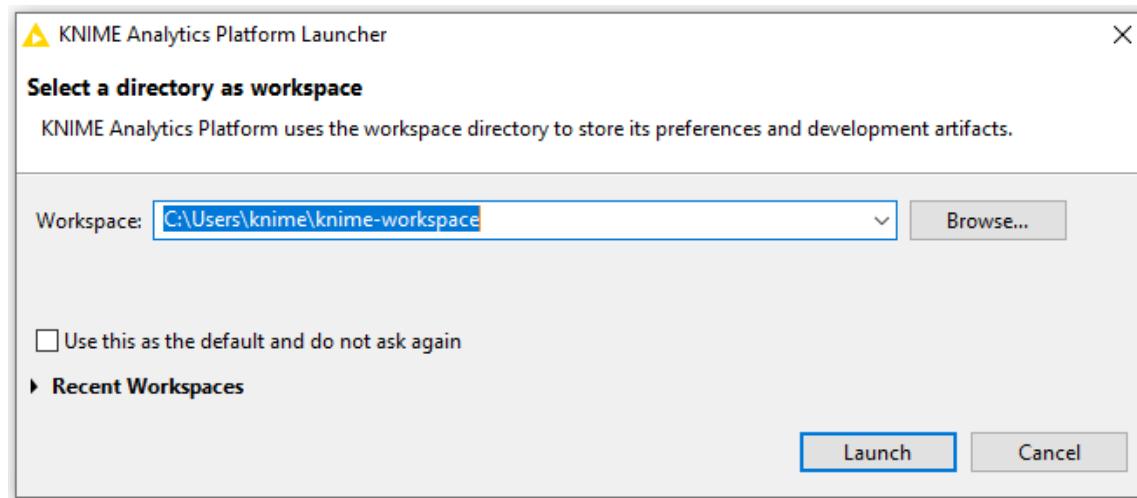
# Start KNIME Analytics Platform

- Use the shortcut created by the installer
- Or go to the installation directory and launch KNIME via the knime.exe

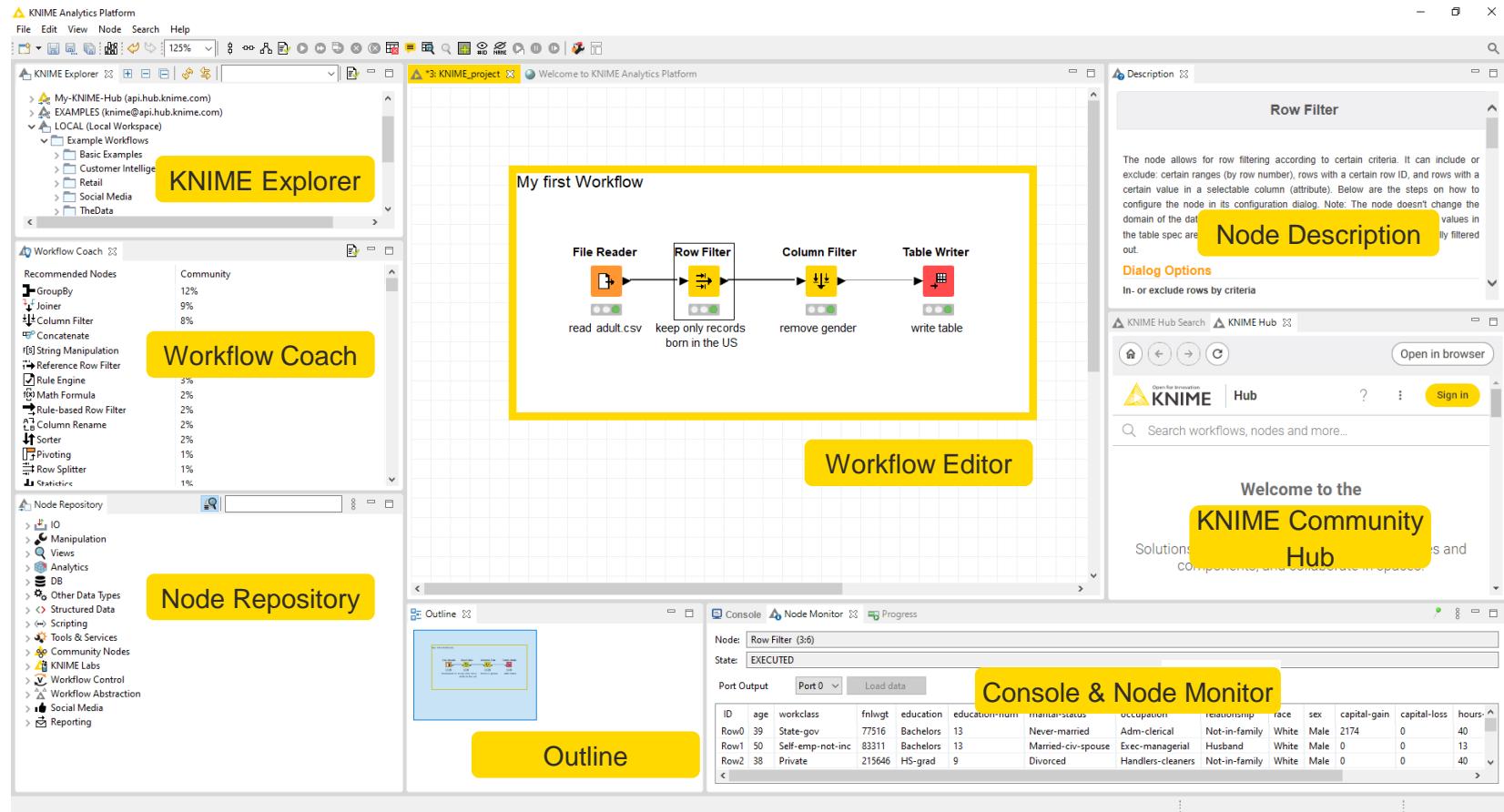


# The KNIME Workspace

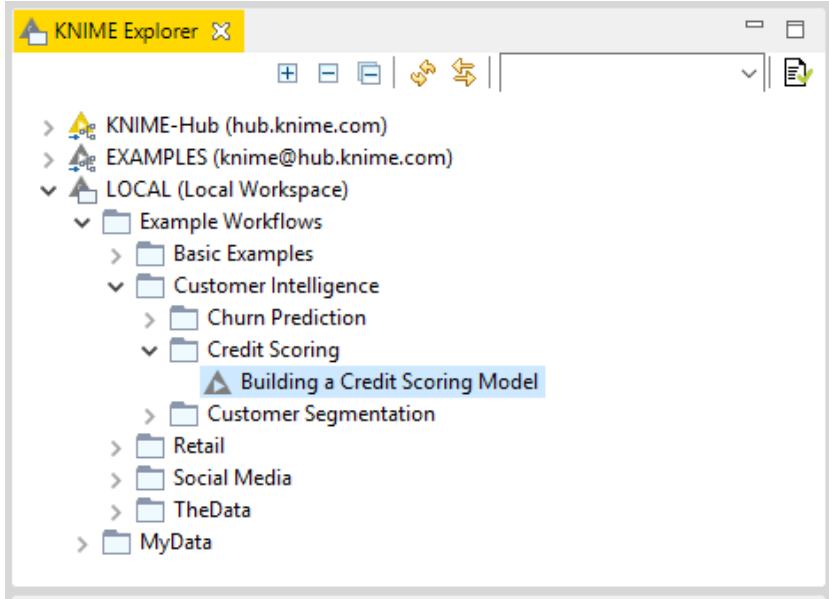
- The workspace is the **folder/directory** in which workflows (and potentially data files) are stored for the current KNIME session
- Workspaces are portable (just like KNIME)



# The KNIME Analytics Platform Workbench



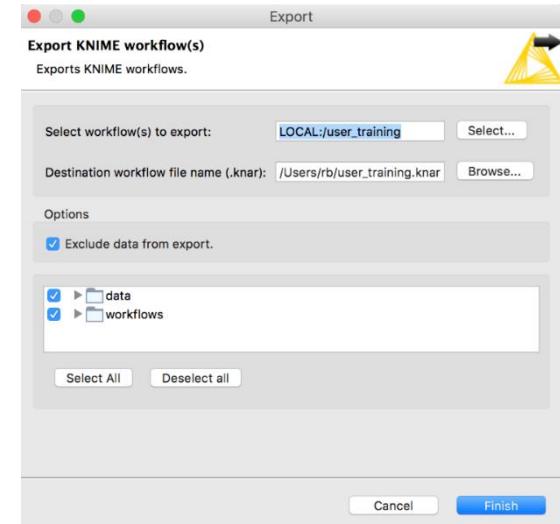
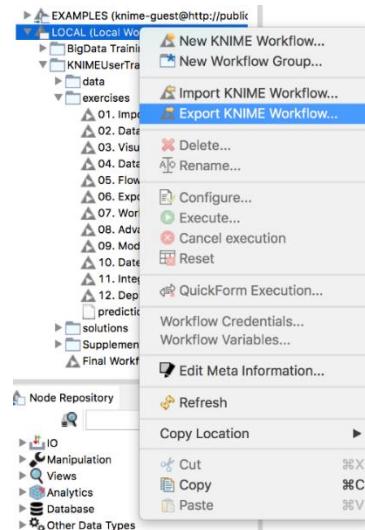
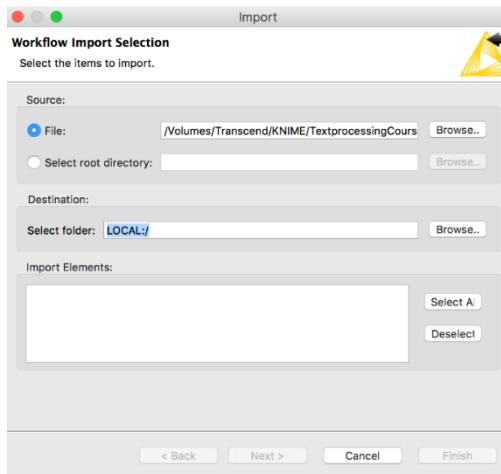
# KNIME Explorer



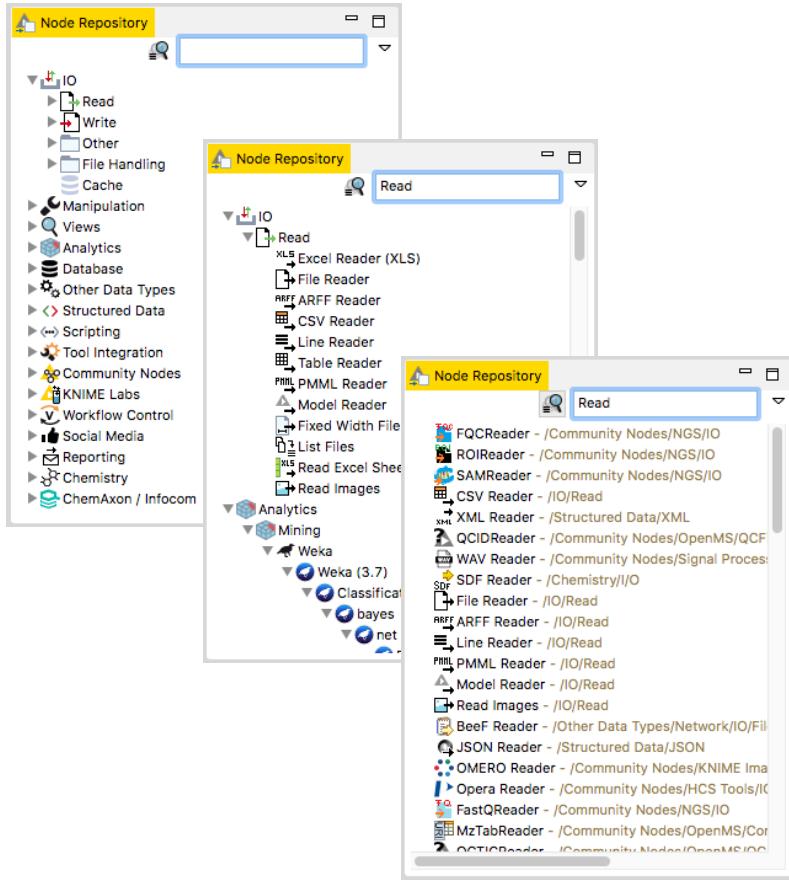
- In LOCAL you can access your own workflow projects.
- Other mountpoints allow you to connect to
  - Examples
  - KNIME Community Hub
  - KNIME Business Hub
- The Explorer toolbar on the top has a search box and buttons to
  - ▶ select the workflow displayed in the active editor
  - ▶ refresh the view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

# Creating New Workflows, Importing, and Exporting

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

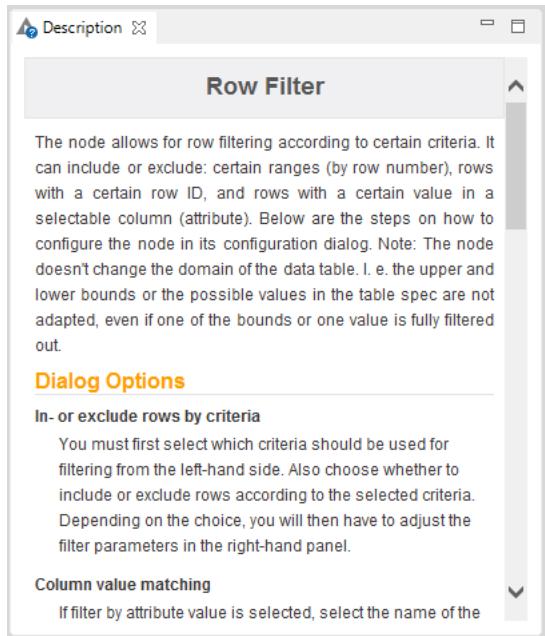


# Node Repository



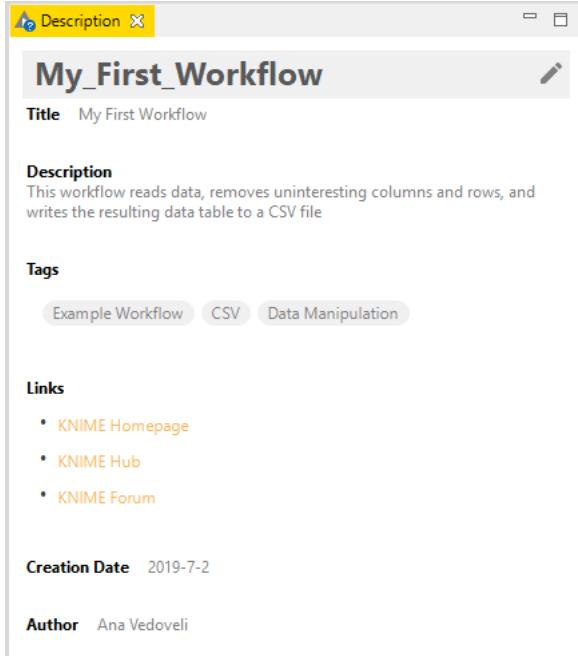
- The Node Repository lists all KNIME nodes
- The search box has 2 modes
  - Standard Search – exact match of node name
  - Fuzzy Search – finds the most similar node name

# Description



- The Description view provides information about:
  - Node functionality
  - Input & output
  - Node settings
  - Ports
  - References to literature

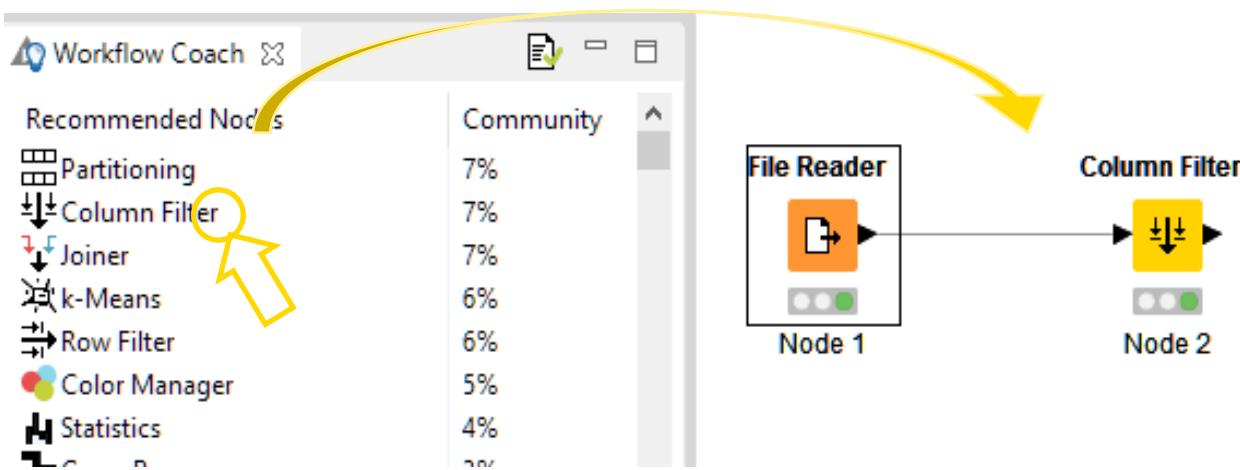
# Workflow Description



- When selecting the workflow, the Description view gives you information about the workflow:
  - Title
  - Description
  - Associated tags and links
  - Creation date
  - Author

# Workflow Coach

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



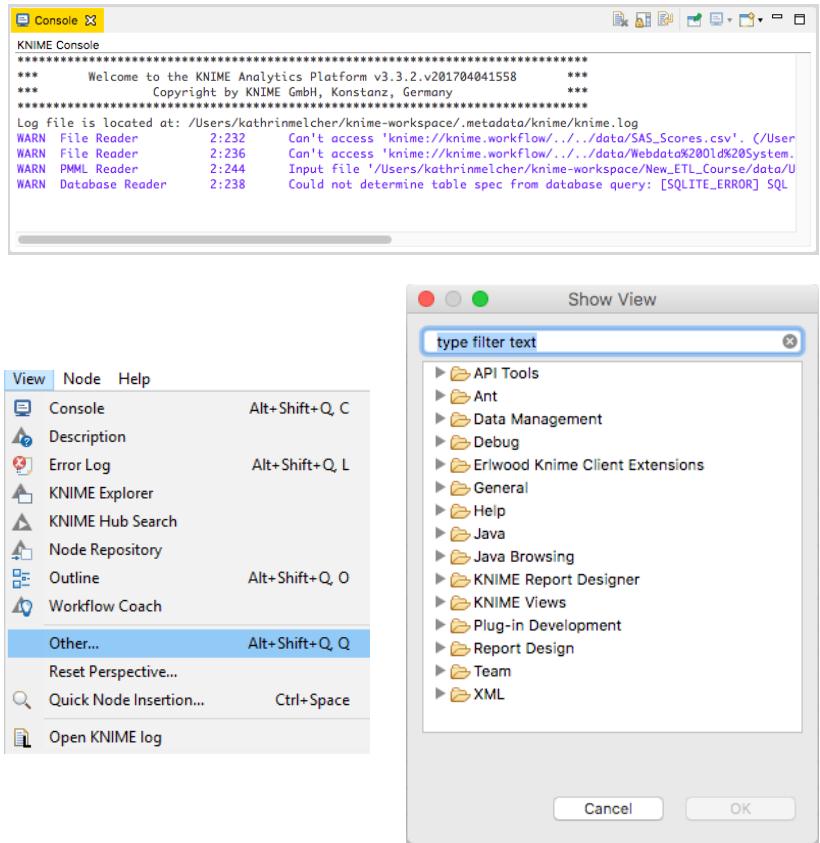
# Node Monitor

- 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 shows the KNIME Node Monitor window. At the top, there are tabs for 'Console' and 'Node Monitor'. Below the tabs, the 'Node' is listed as 'Get Customers from Database (0:1207)' and the 'State' is 'EXECUTED'. A dropdown 'Port Output' is set to 'Port 0' and a 'Load data' button is visible. The main area displays a table of customer data with columns: ID, MaritalStatus, Gender, EstimatedYearlyIncome, NumberOfContracts, Age, Available401K, CustomerV, and Products. The table contains 8 rows of data. To the right of the table, a context menu is open, with the 'Show Output Table' option checked. Other options in the menu include '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

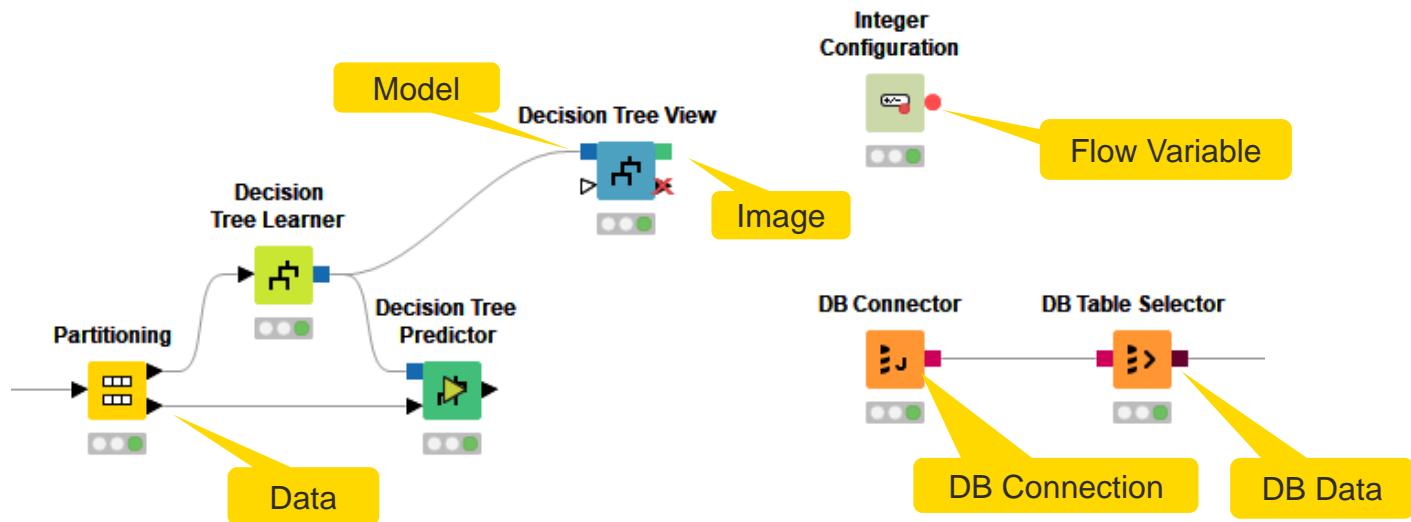
# Console and Other Views



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

# Inserting and Connecting Nodes

- Insert nodes into workspace by dragging them from the Node Repository or by double-clicking in the Node Repository
- Connect nodes by left-clicking the output port of Node A and dragging the cursor to the (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 need to be configured
- To access a node configuration dialog:
  - Double-click the node
  - Right-click -> Configure

The screenshot shows the configuration dialog for a 'File Reader' node. The title bar says 'Dialog - 0:1 - File Reader'. The main tabs are 'Settings', 'Transformation', 'Advanced Settings', 'Limit Rows', 'Encoding', 'Flow Variables', and 'Memory Policy'. The 'Settings' tab is active.

**Input location:**  
Read from: Relative to Current workflow  
Mode:  File  Files in folder  
File: ../../data/CustomerInfoSystem1.csv

**Reader options:**  
Format:  
Autodetect format  Column delimiter: , Row delimiter:  Line break  Custom   
Quote char: "  Quote escape char: "

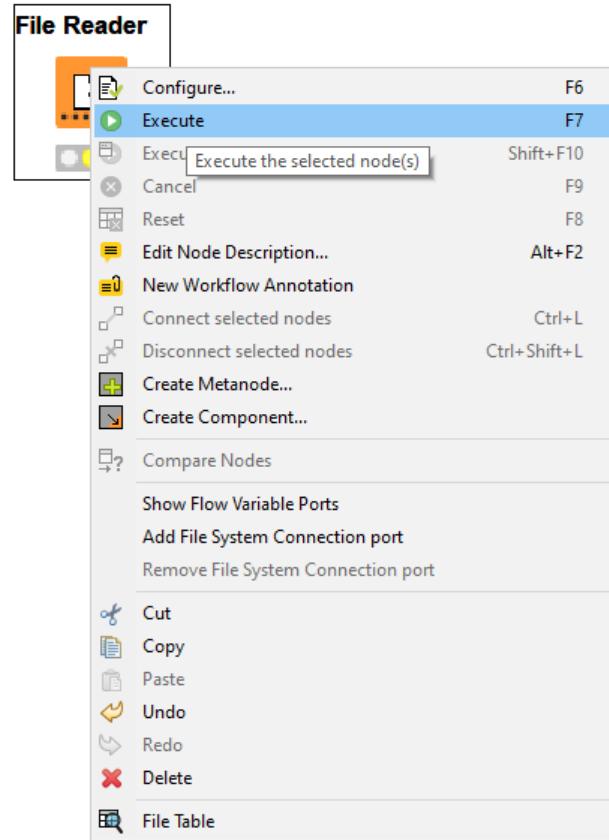
Has column header  Has row ID   
Support short data rows  Prepend file index to row ID

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

Row ID	City	Country	CustomerID	FirstName	LastName	Birthday	Age	Email
Row0	Glasgow	United Kingdom	17-171-832-104	Alois	Berger	23.9.1972	47	Alois.Berger@mcr.com
Row1	Szczecin	Poland	37-370-580-177	Michaela	Schultz	9.6.1998	21	Michaela.Schultz@mcr...
Row2	Sheffield	United Kingdom	27-270-743-182	Rotraut	GrÄ¼nwald	20.4.1975	44	Rotraut.GrÄ¼nwald...
Row3	Bodum-Hordel	Germany	64-647-953-993	Helga	Heindl	18.10.2000	19	Helga.Heindl@mcr.com
Row4	Dortmund	Germany	84-846-821-690	Mira	Gleich	18.3.1997	22	Mira.Gleich@mcr.com
Row5	Valencia	Spain	58-582-352-948	Joanna	Radke	13.12.1995	24	Joanna.Radke@mcr...
Row6	Valenda	Spain	65-655-257-939	Hanspeter	Storch	25.1.1998	21	Hanspeter.Storch@mcr...

# Node Execution

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



# Tool Bar

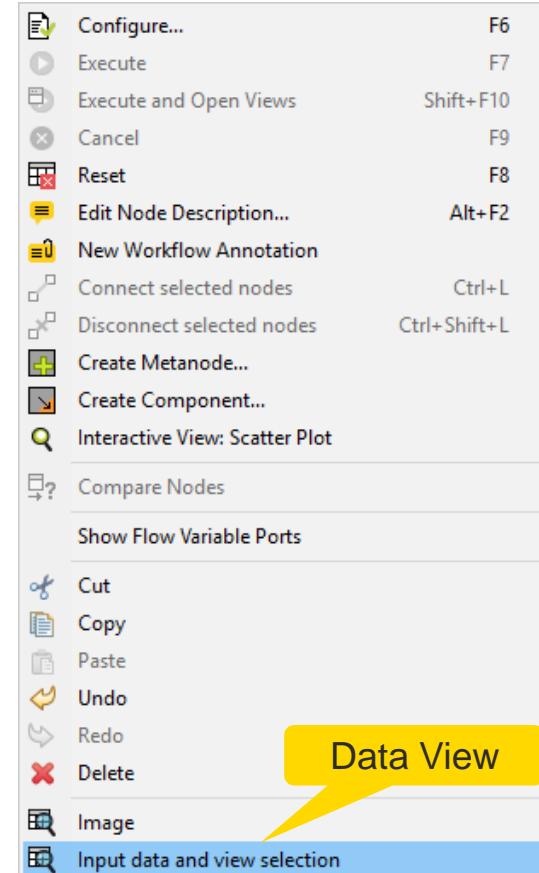
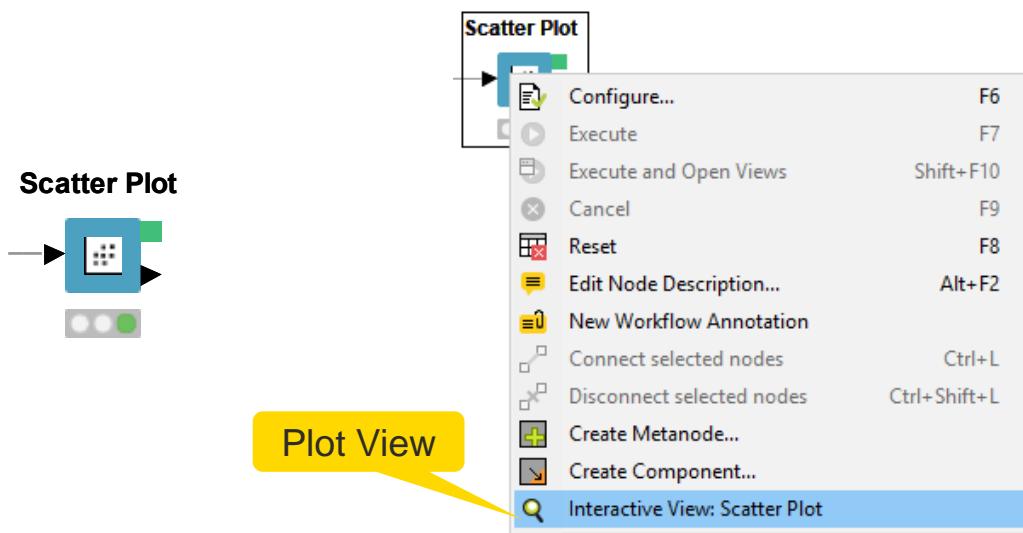


- The buttons in the toolbar can be used for the active workflow. The most important buttons are:

- 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 to inspect the execution results by
  - selecting output ports (last option in the context menu) to inspect tables, images, etc.
  - selecting Interactive View to open visualization results in a browser



# 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



# Getting Started: KNIME Community Hub

- Place to search and share
  - Workflows
  - Nodes
  - Components
  - Extensions

A screenshot of a web browser showing the KNIME Community Hub at [hub.knime.com](https://hub.knime.com). The search bar at the top contains the text "Sentiment Analysis". Below the search bar, the text "350 results" is displayed. Underneath, there are three workflow cards. Each card has a thumbnail, the name "Sentiment Analysis", and a brief description. The first two cards have a yellow "View" button, while the third has a blue "View" button.

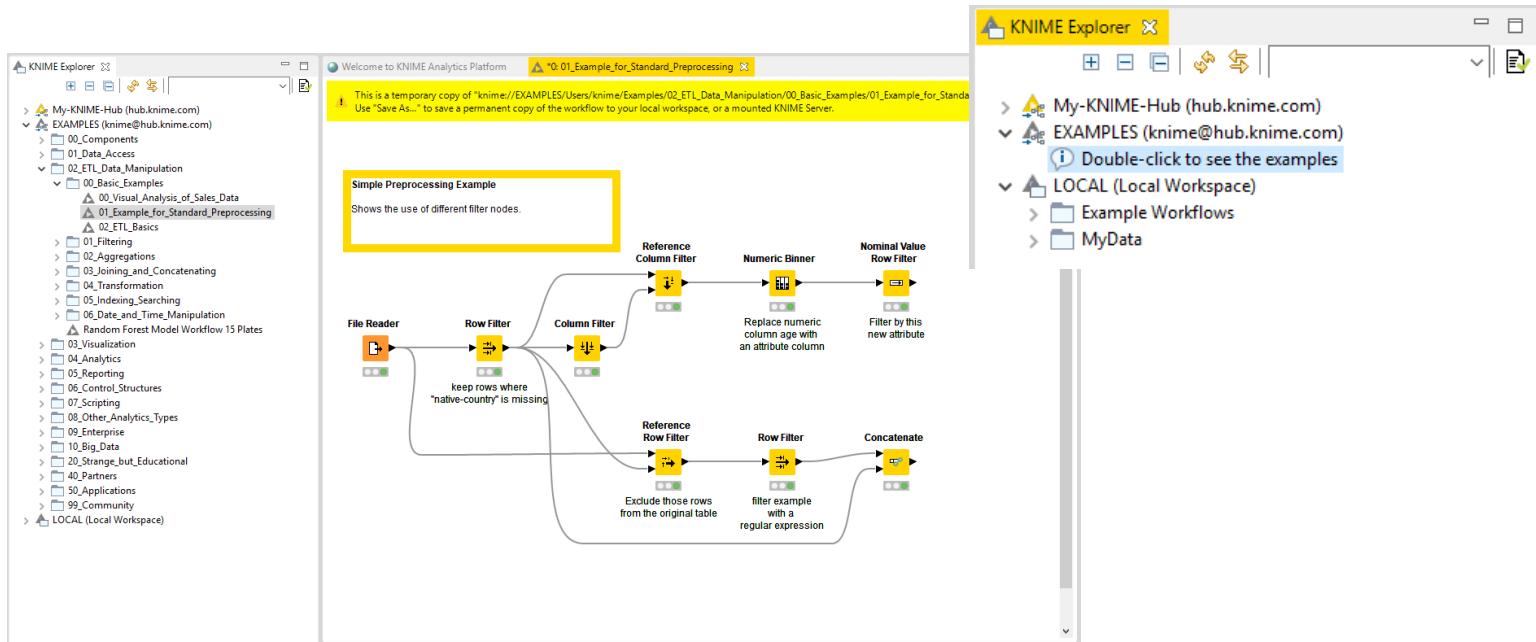
Workflow	Description	Action
Sentiment Analysis	This workflow shows how to train a simple neural network for text classification, in this case sentiment analysis. The used network learns a 128 dimensional word embedding followed by an LSTM. This e...	View
Sentiment Analysis	This workflow shows how to train a simple neural network for text classification, in this case sentiment analysis. The used network learns a 128 dimensional word embedding followed by an LSTM. This e...	View
Sentiment Analysis (Classification) of Documents	This workflow shows how to import text from a csv file, convert it to documents, preprocess the documents and transform them	View

A screenshot of the KNIME Community Hub homepage. At the top, the KNIME logo and the word "Hub" are visible. The main heading is "Welcome to the KNIME Community Hub". Below the heading, it says "Solutions for data science: find workflows, nodes and components, and collaborate in spaces." A search bar with the placeholder "Search workflows, nodes and more..." is present. Below the search bar, four statistics are displayed: 13 056 Workflows, 4 293 Nodes, 1 276 Components, and 222 Extensions.

<https://hub.knime.com>

# Getting Started: KNIME Examples

- Connect via KNIME Explorer to a public repository with large selection of example workflows for many, many applications

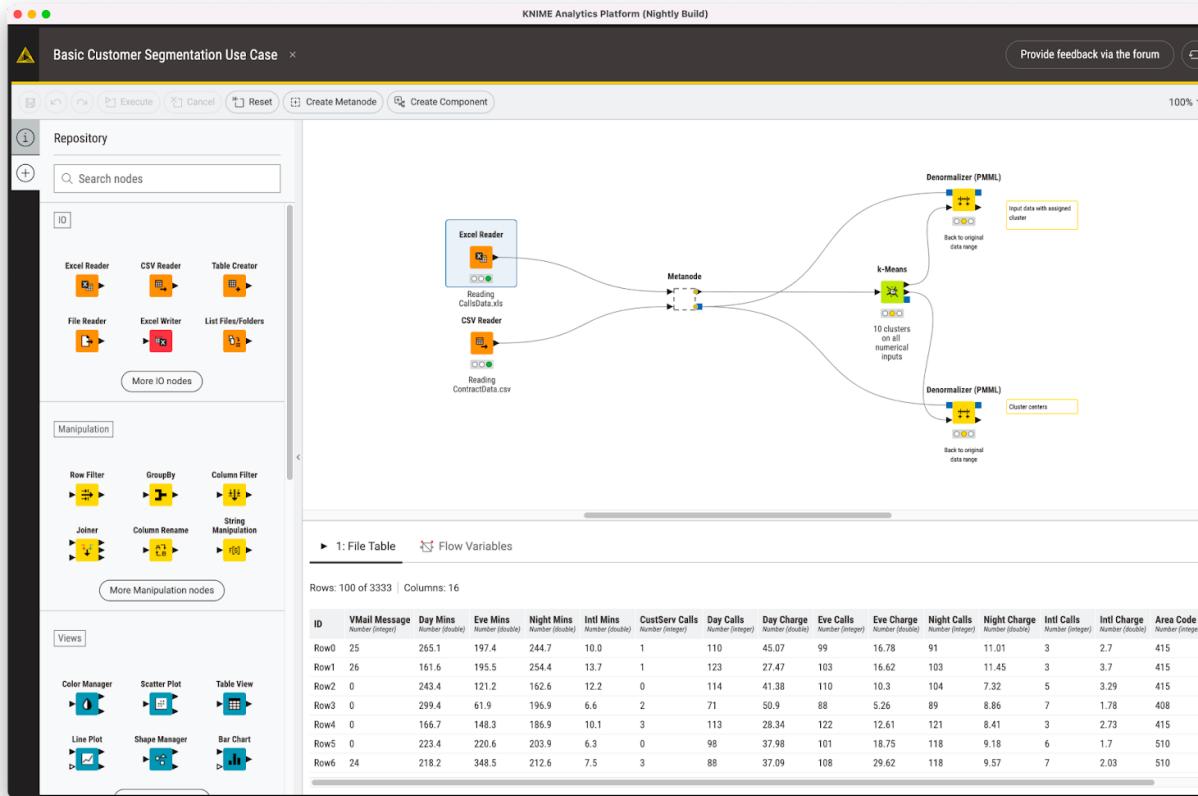


# Hot Keys (for Future Reference)

Task	Hot key	Description
Node Configuration	F6	opens the configuration window of the selected node
	F7	executes selected configured nodes
	Shift + F7	executes all configured nodes
Node Execution	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
Metanode	Shift + F12	opens metanode wizard

# KNIME Modern UI Preview (Labs)

- Preview KNIME Analytics Platform's makeover
  - Install KNIME Modern UI Preview extension and click the “Open KNIME Modern UI Preview”



# Today's Use Case

---

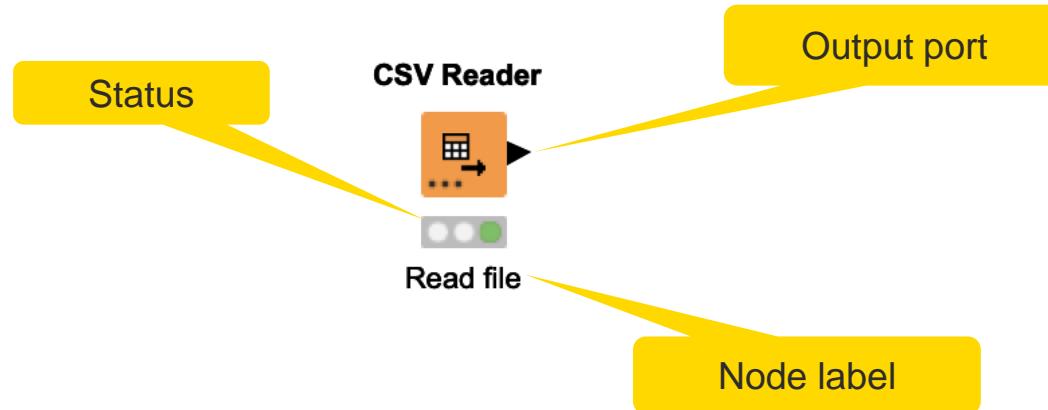
- Analyze data from a retail company, which has an online shop and stores
- Data:
  - Customer information from two different systems (.csv, .table)
  - Purchases from the online store (sqlite database)
  - List of product numbers and prices (sqlite database)
  - Purchases from the stores (.table)
  - Store information (.xls)
- Goal:
  - Single, clean table of our customers
  - Standardized list of all transactions

# Importing Data

# Data Source Nodes

Typically characterized by:

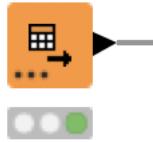
- Orange color
- By default no input ports, 1-2 output ports
- New file handling with KNIME 4.3.
  - Consistent user experience across all nodes and file systems
  - Managing of various file systems within the same workflow
  - Performance improvements



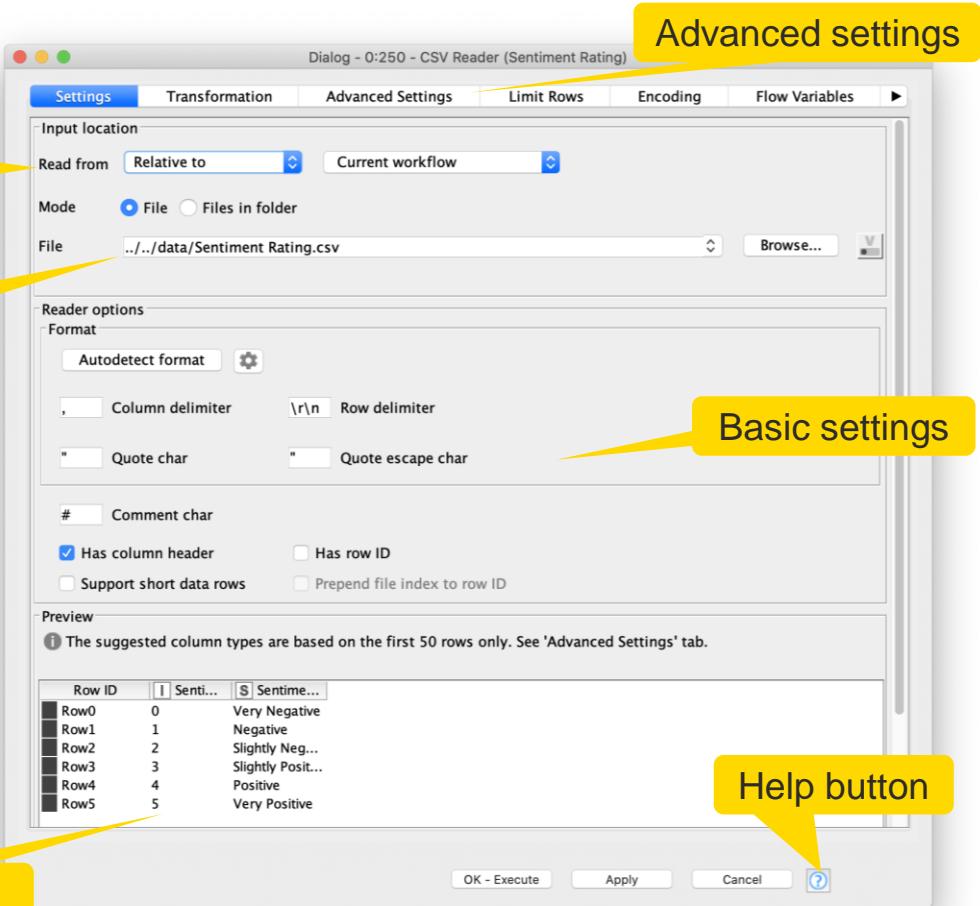
# CSV Reader

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

CSV Reader

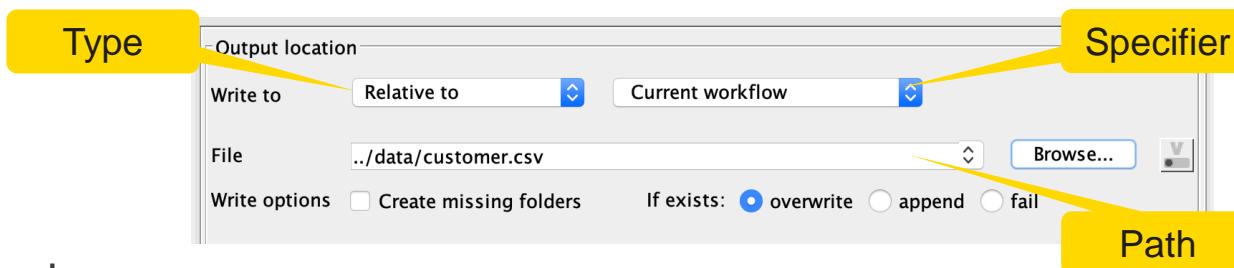


Preview



# Common Settings: File Path

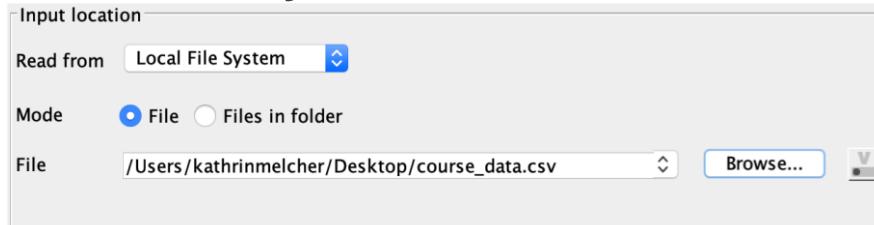
- A path consists of three parts:
  - **Type**: Specifies the file system type e.g. local, relative, mountpoint, custom\_url or connected.
  - **Specifier**: Optional string with additional file system specific information e.g. relative to which location (knime.workflow)
  - **Path**: Specifies the location within the file system



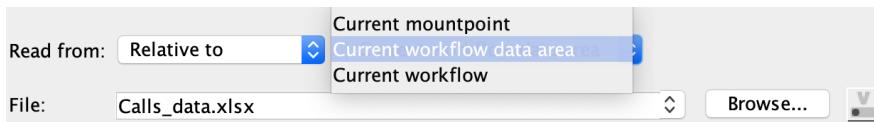
- Examples:
  - (LOCAL, , C:\Users\username\Desktop)
  - (RELATIVE, knime.workflow, file1.csv)
  - (MOUNTPOINT, MOUNTPOINT\_NAME, /path/to/file1.csv)
  - (CONNECTED, amazon-s3:eu-west-1, /mybucket/file1.csv)

# Common Settings: Four Default File Systems

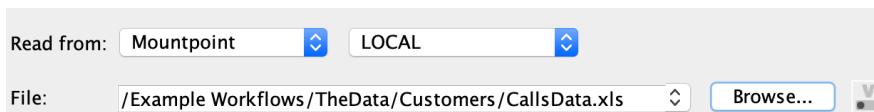
- Local File System



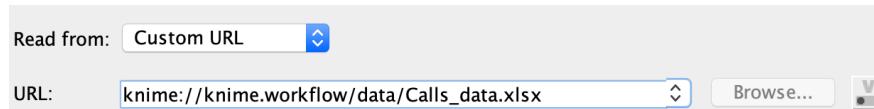
- Relative to ...



- Mountpoint



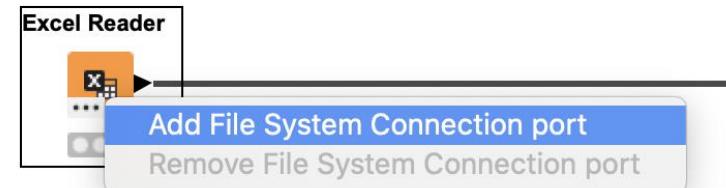
- Custom URL



# 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)



Input location

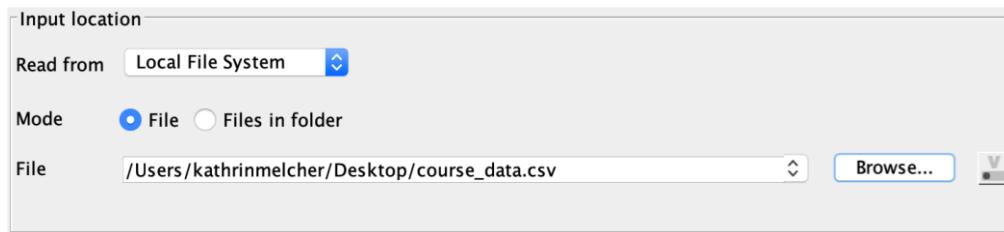
Read from

Mode  File  Files in folder

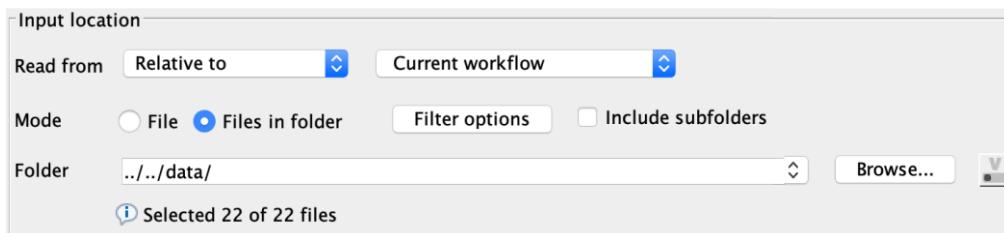
File

# Common Settings: Read Single or Multiple Files

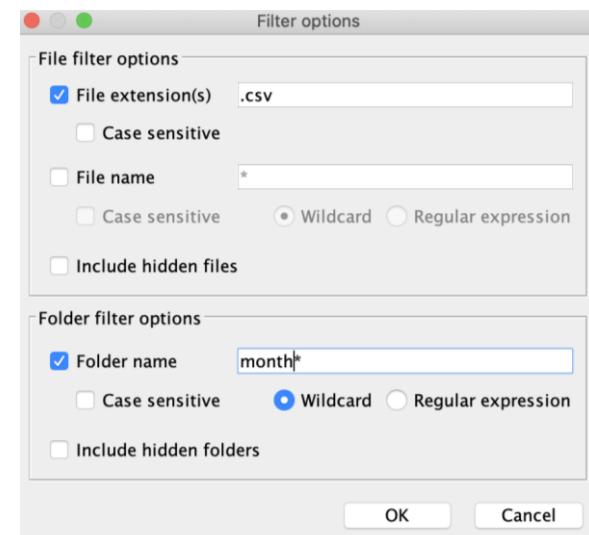
- Single file



- Files in a folder



- Option to include subfolder
- Option to define filter criterions



# Common Settings: Transformation Tab

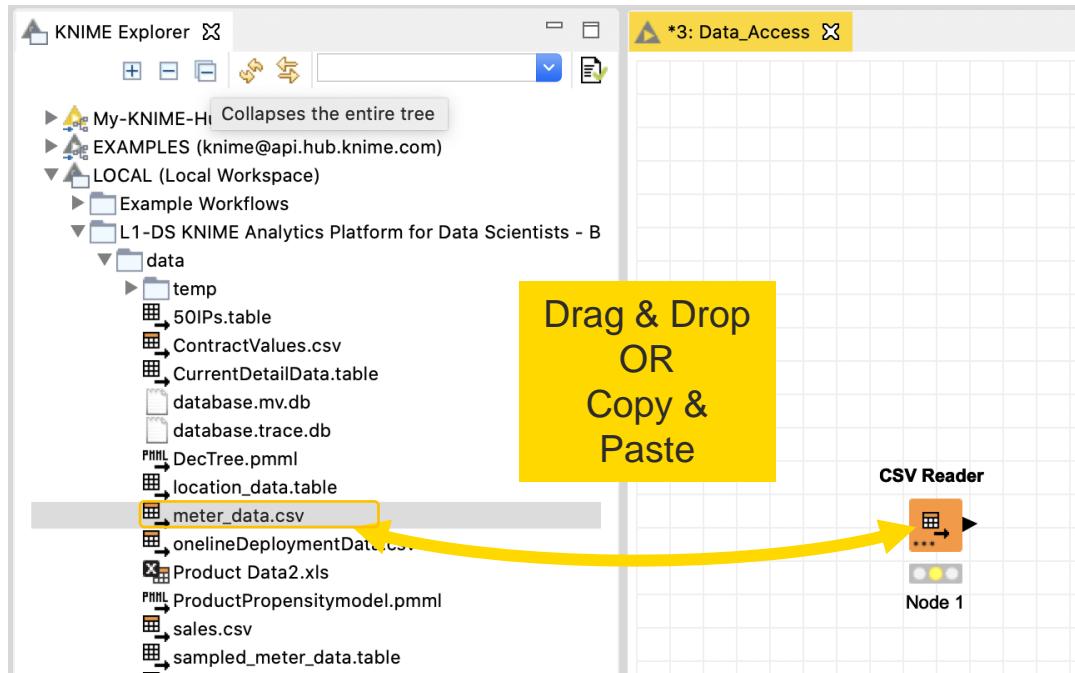
- Supported operations
  - Column filtering
  - Column sorting
  - Column renaming
  - Column type mapping
  - Select between union or intersection of columns (in case of reading many files)

The screenshot shows the 'Dialog - 0:2 - CSV Reader' window with the 'Transformation' tab selected. The 'Transformations' section contains a table with columns 'Column', 'New name', and 'Type'. The table includes rows for City, Country, CustomerID (selected), FirstName, LastName, Birthday, Age, Email, Newsletter, and a placeholder for unknown columns. The 'Type' column shows mappings like String for City and Country, and String for CustomerID. The 'Preview' section at the bottom displays a sample of five rows from the dataset, showing columns for Row ID, City, Country, ID, FirstName, LastName, Birthday, Age, Email, and Newsletter.

Column	New name	Type
\$ City		\$ String
\$ Country		\$ String
<input checked="" type="checkbox"/> \$ CustomerID	ID	\$ String
\$ FirstName		⌚ Local Time
\$ LastName		D Number (double)
\$ Birthday		I Number (integer)
\$ Age		L Number (long)
\$ Email		PMML
\$ Newsletter		Period
<any unknown new column>		SVG image

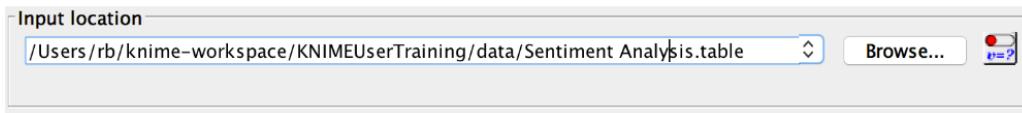
Row ID	City	Country	ID	FirstName	LastName	Birthday	Age	Email	Newsletter
Row0	Glasgow	United Kingdom	17-171-832-104	Alois	Berger	23.9.1972	47	Alois.Berger@mcr.com	0
Row1	Szczecin	Poland	37-370-580-177	Michaela	Schultz	9.6.1998	21	Michaela.Schultz@mcr...	0
Row2	Sheffield	United Kingdom	27-270-743-182	Rotraut	GrÄkmwald	20.4.1975	44	Rotraut.GrÄkmwald...	0
Row3	Bochum-Hordel	Germany	64-647-953-993	Helga	Heindl	18.10.2000	19	Helga.Heindl@mcr.com	0
Row4	Dortmund	Germany	84-846-821-690	Mira	Gleich	18.3.1997	22	Mira.Gleich@mcr.com	0

# Alternative Faster Way ...

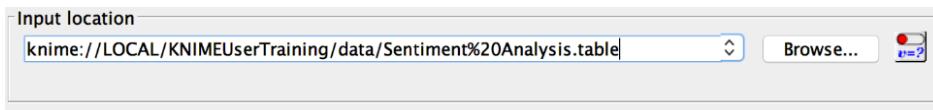


# File Path Options Old File Handling

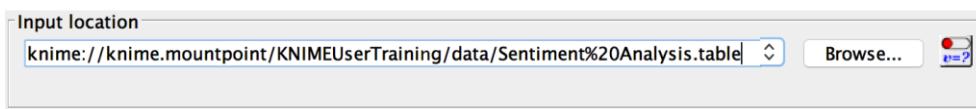
- Local path



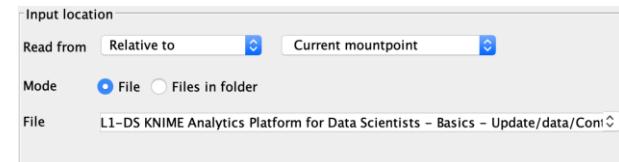
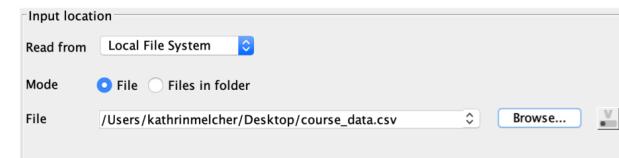
- Absolute URL



- Mountpoint-relative URL

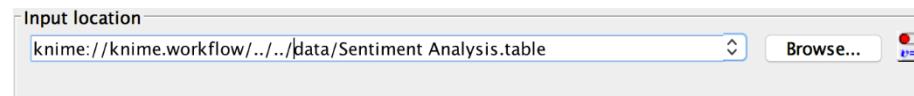
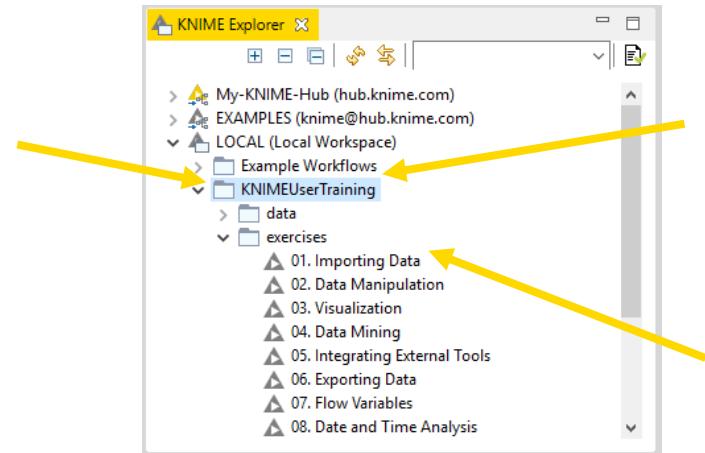


## New file handling



# Workflow-Relative File Paths (Old File Handling)

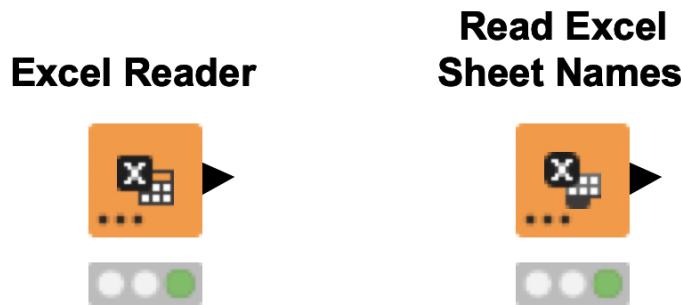
- Best choice if workflows are to be shared
- Requires matching folder structure within workflow group
  - Independent of environment outside of workflow group
- Example: Path to „Sentiment Analysis.table“
  - Local path:
    - C:\Users\rb\knime-workspace\KNIMEUserTraining\data\Sentiment Analysis.table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



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

# Excel Reader (XLS)

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



# Excel Reader

## Excel Reader



File system

Path

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

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 ?

Sheet specific settings

Preview

# Table Reader

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

The image shows the 'Table Reader' dialog box from the KNIME interface. The 'File system' tab is selected. A yellow callout points to the 'File' input field where the path '.../data/CustomerInfoSystem2.table' is entered. Another yellow callout points to the 'Browse...' button next to the file path field. Below the input location section, there's a preview of a table with 11 columns and 11 rows of data. The columns are labeled: Row ID, City, Country, CustomerID, FirstN..., LastN..., Birthday, Age, and Email. The rows contain various customer information such as names, addresses, and birthdates.

File system

Path

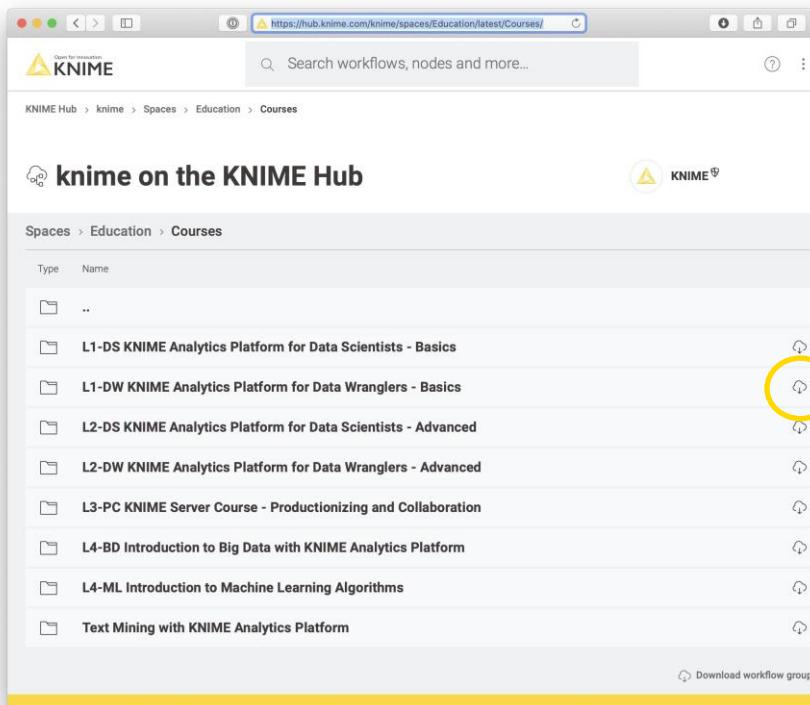
Row ID	City	Country	CustomerID	FirstN...	LastN...	Birthday	Age	Email
Row0	Glasgow	United Kingdom	17-171-832-104	Alois Berger	23.9.1972	47		Alois.Berger
Row1	Szczecin	Poland	37-370-580-177	Michaela Schultz	9.6.1998	21		Michaela.Schultz
Row2	Sheffield	United Kingdom	27-270-743-182	Rotraut Grünwald	20.4.1975	44		Rotraut.Gruenwald
Row3	Bochum...	Germany	64-647-953-993	Helga Heindl	18.10.2000	19		Helga.Heindl
Row4	Dortmund	Germany	84-846-821-690	Mira Gleich	18.3.1997	22		Mira.Gleich
Row5	Valencia	Spain	58-582-352-948	Joanna Radke	13.12.1995	24		Joanna.Radke
Row6	Valencia	Spain	65-655-257-939	Hanspeter Storch	25.1.1998	21		Hanspeter.Storch
Row7	Lodz	Poland	96-969-846-915	Klaus-Peter Heincke	13.4.1963	56		Klaus-Peter.Heincke
Row8	Stuttgart	Germany	48-484-590-744	Guenther Gassner	1.5.1988	31		Guenther.Gassner
Row9	Denver	United States	47-474-558-858	Giuseppina Nitsch	9.12.1975	44		Giuseppina.Nitsch
Row10	Kharkiv	Ukraine	85-859-448-511	Sami Zimmerer	5.10.1982	37		Sami.Zimmerer

Table Reader



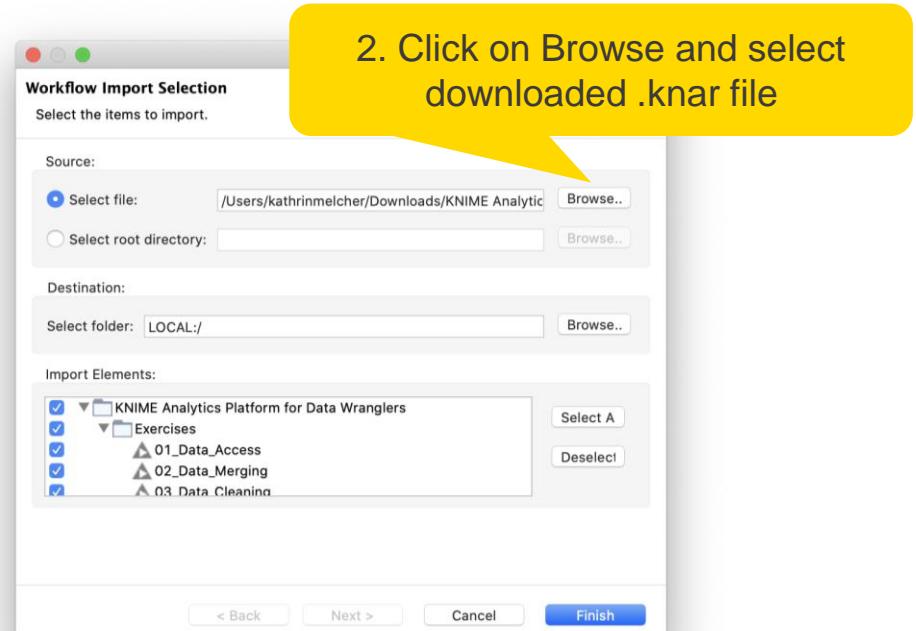
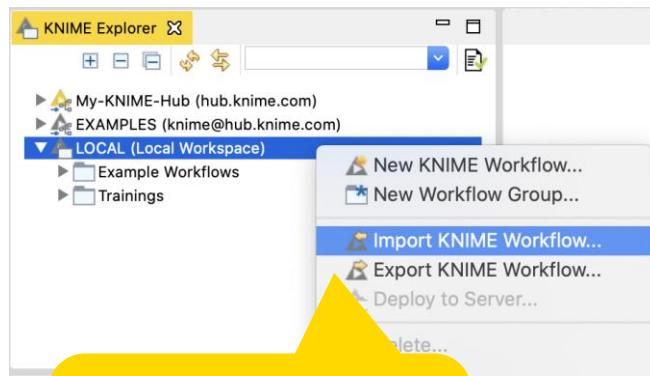
# Downloading Exercises

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



# Importing Exercises

- Import the course material to KNIME Analytics Platform



# Exercise: 01\_Data\_Access

---

Open the workflow 01\_Data\_Access and read the following data files:

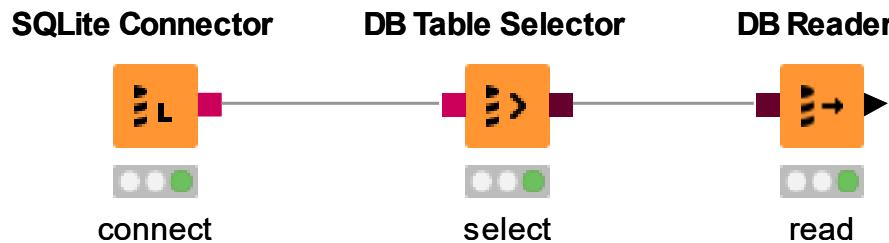
- Customer information
  - CustomerInfoSystem1.csv
  - CustomerInfoSystem2.table
- Online shop transactions, and product number & price information
  - TransactionOnline from Transations.sqlite
  - ProductNrAndPrice from Transations.sqlite
- Store transactions and information
  - Store.xlsx
  - TransactionsStore.table
- Try to use workflow relative-paths

# **Importing Data II**

## **Accessing Databases**

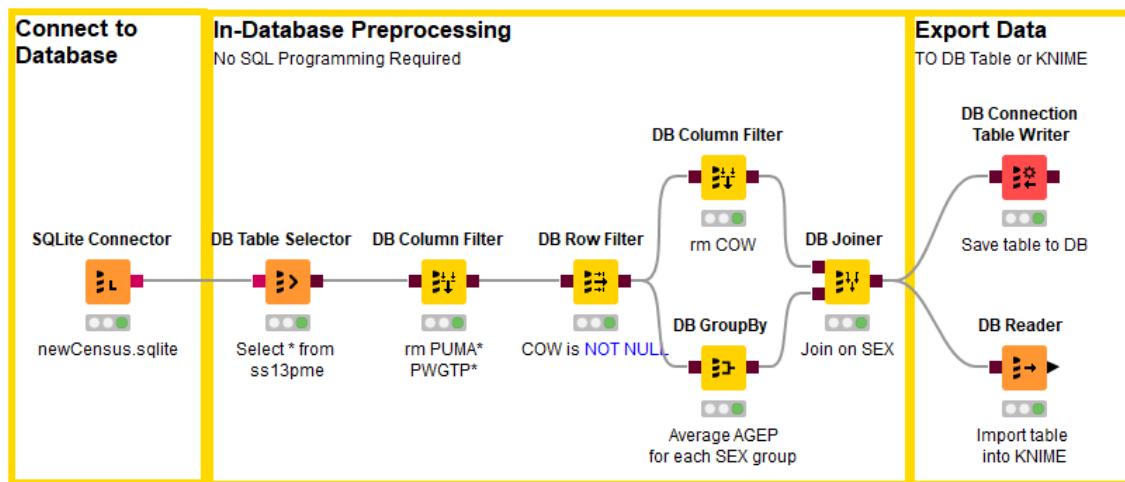
# Database Connectivity

- Read data from any JDBC enabled database
- Write your own SQL or model it using dedicated nodes

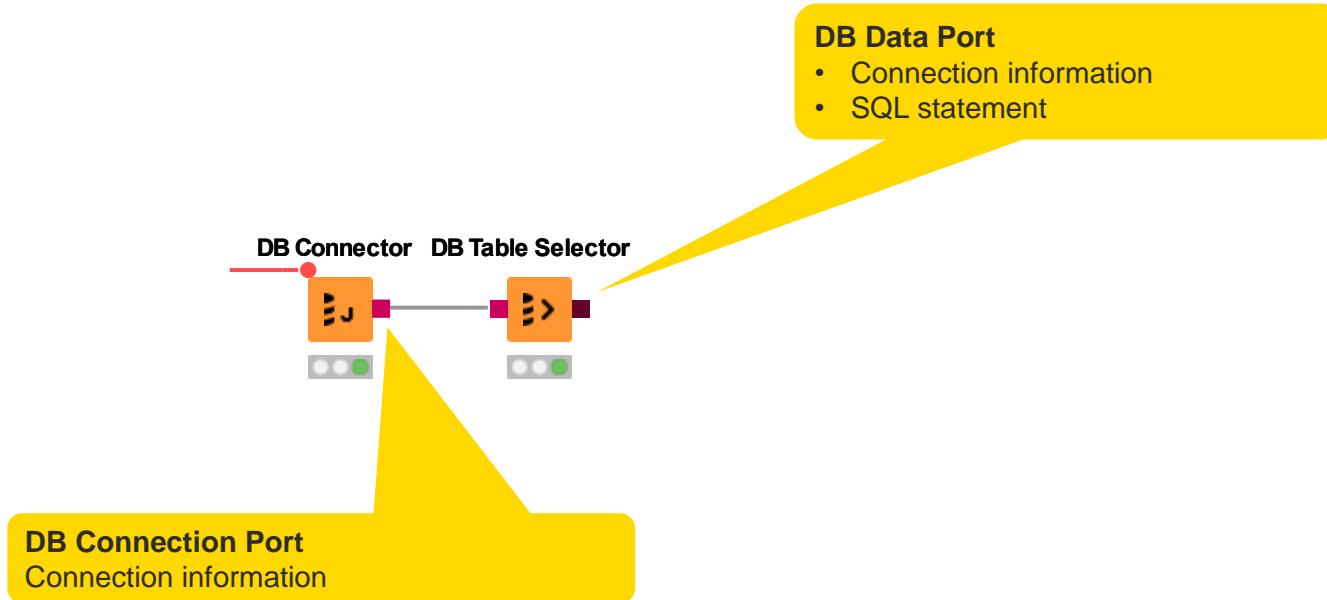


# Database Extension

- 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

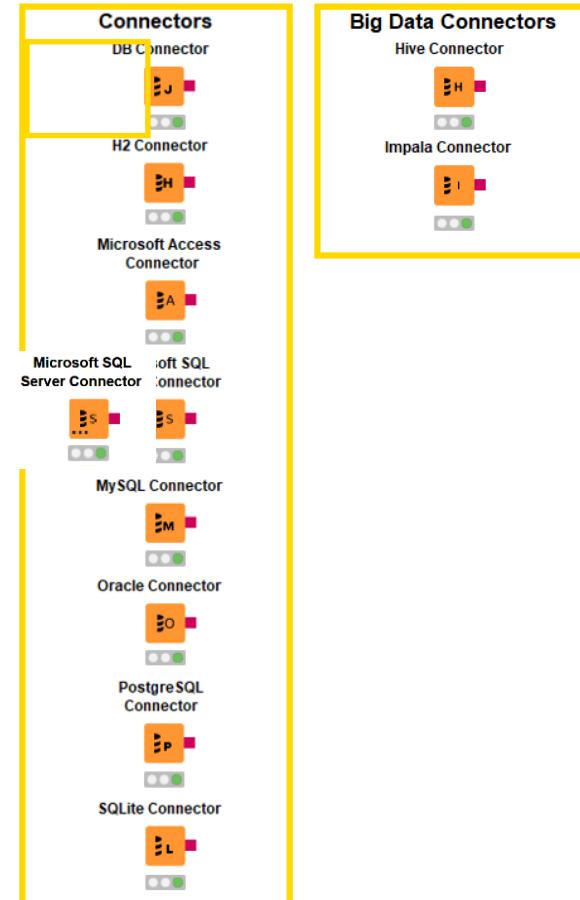


# Database Port Types

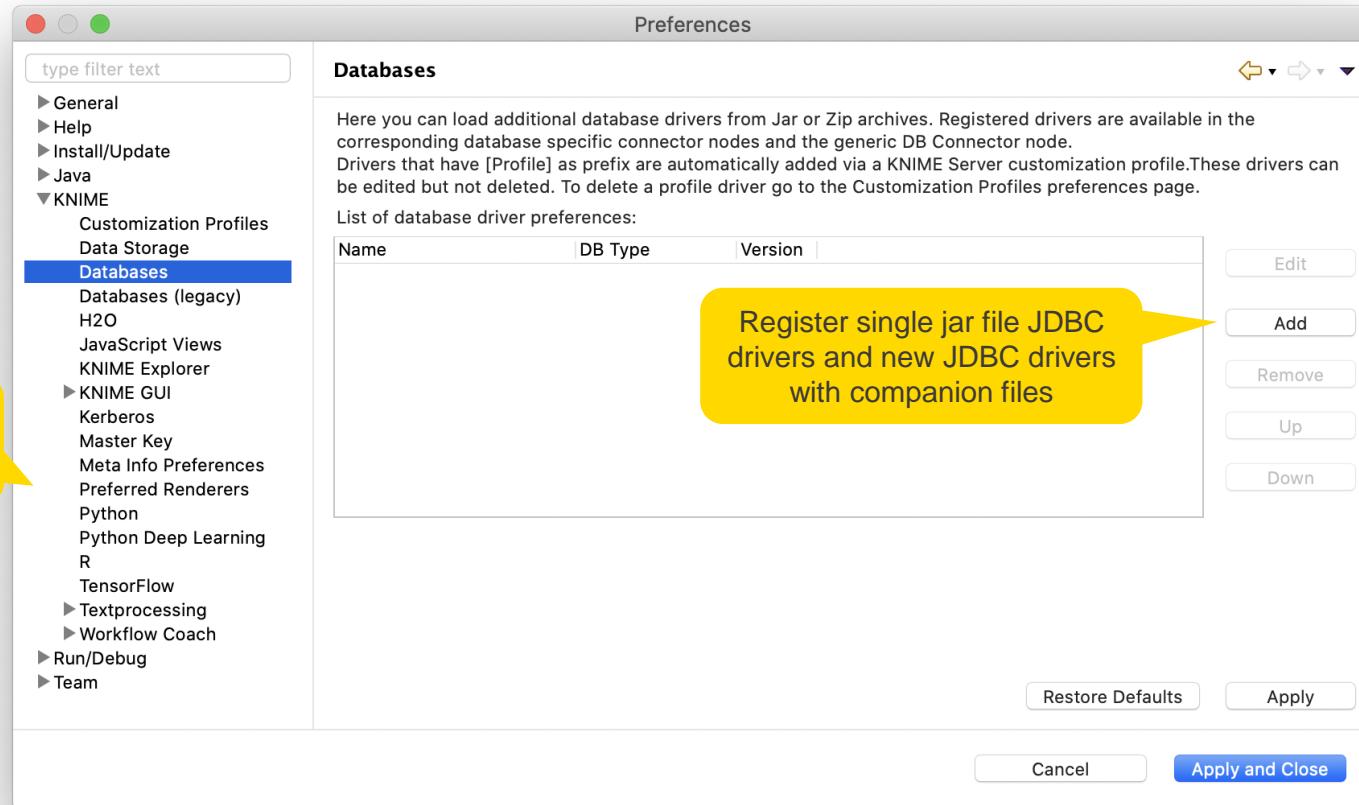


# Database Connectors

- Dedicated nodes to connect to specific databases
  - Necessary JDBC driver included
  - Easy to use
  - Import DB specific behavior/capability
- Hive and Impala connectors are 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

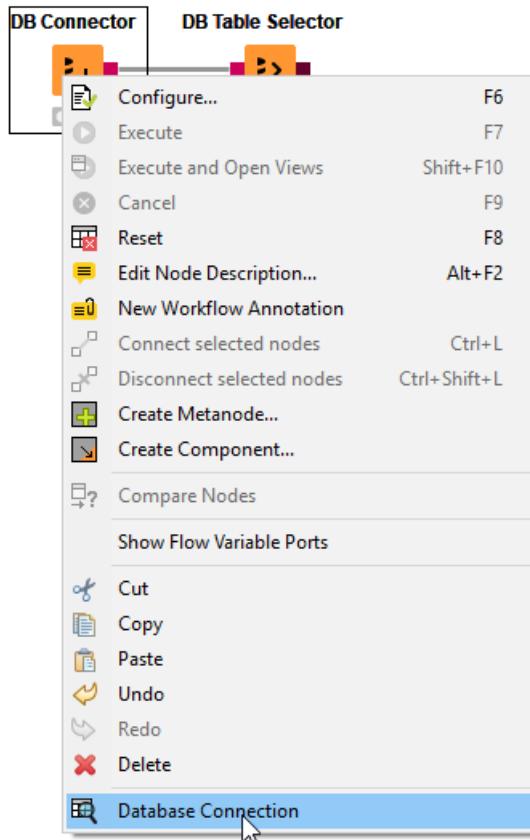


# Register JDBC Driver



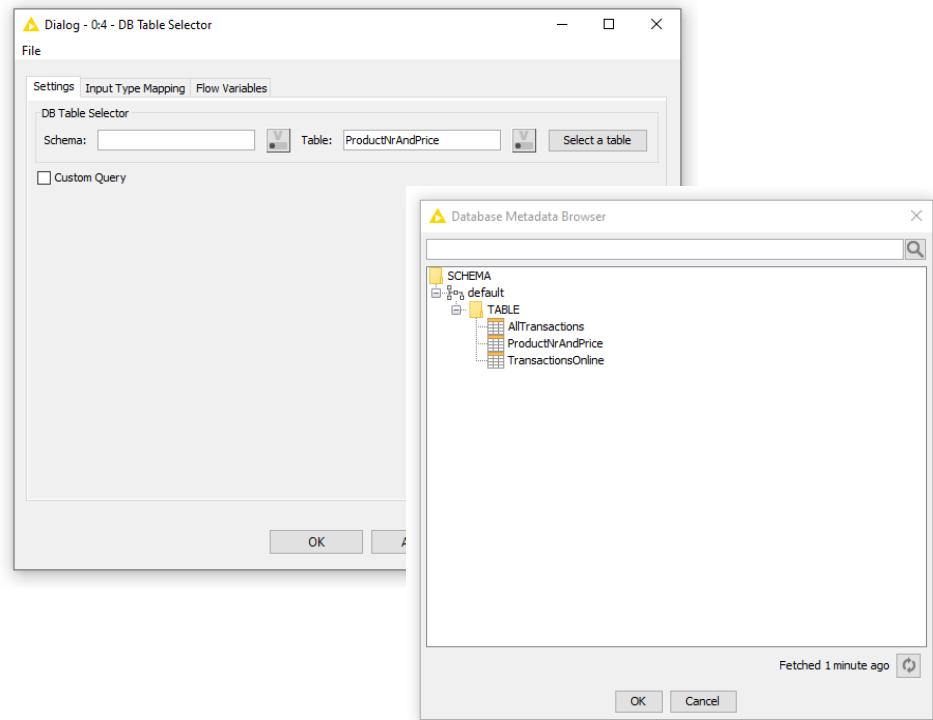
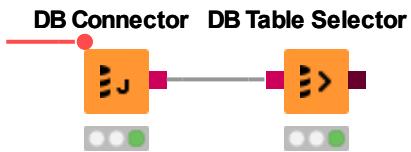
Open KNIME  
and go to File -> Preferences

# Database JDBC Connection Port View

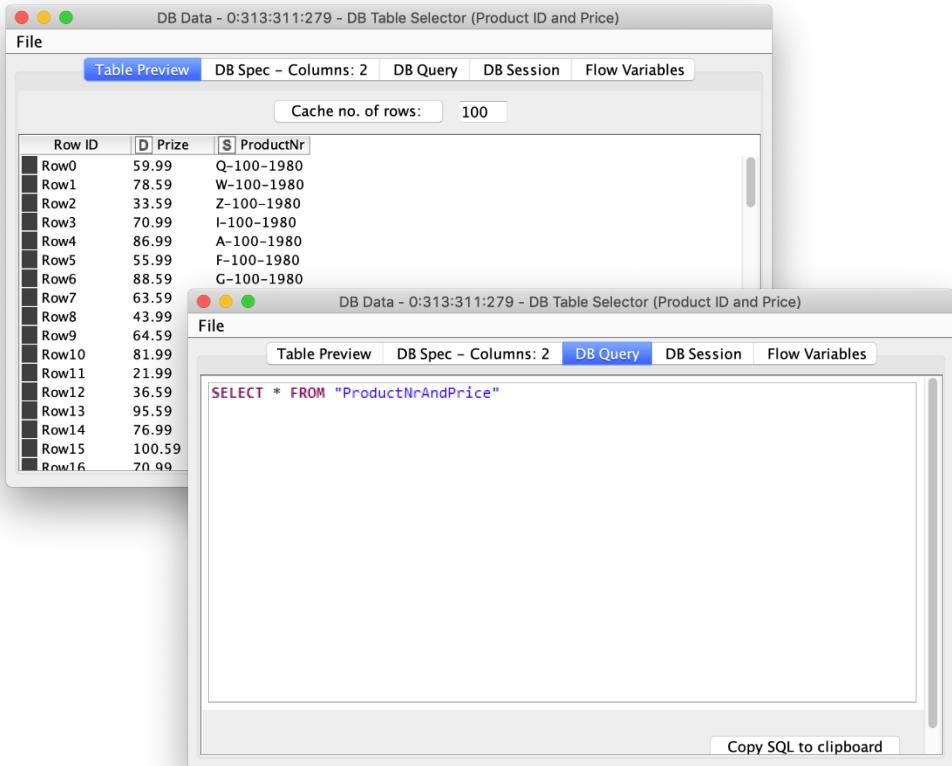
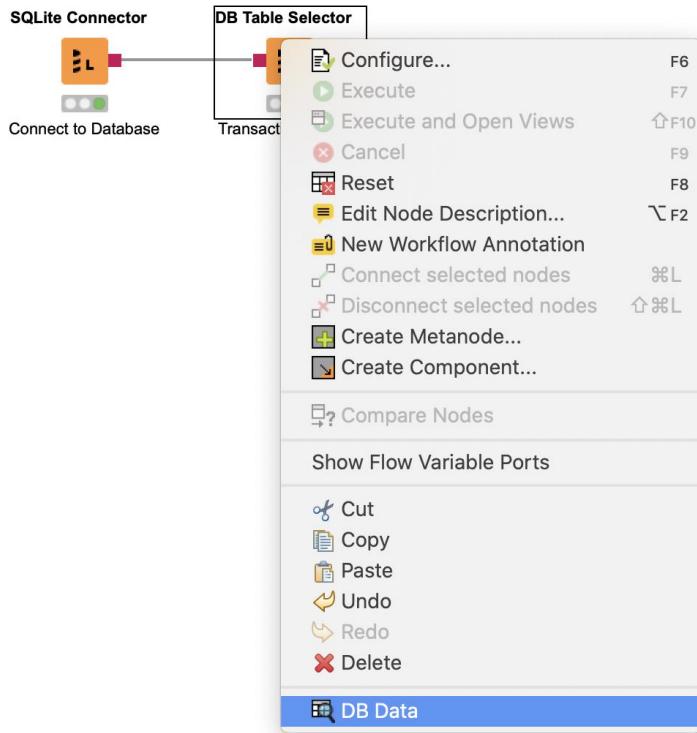


# DB Table Selector

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

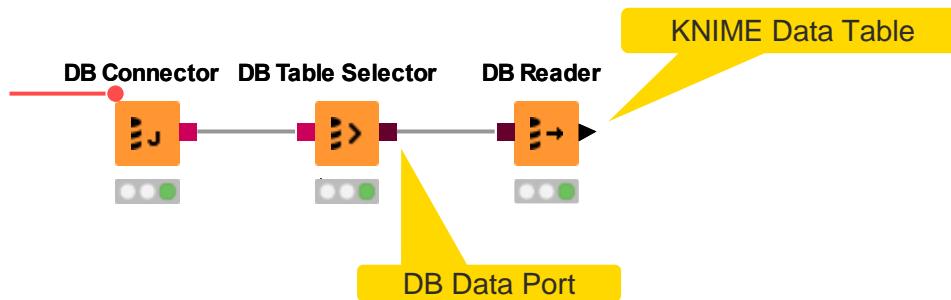


# Database Connection Port View



# DB Reader

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



# Table Creator

- Allows you to create data tables manually
- Data can be entered in a spreadsheet – like the table in the configuration dialog

The image shows the KNIME interface with the 'Table Creator' component on the left and its configuration dialog on the right.

**Table Creator Component:** An orange rectangular component icon with a grid of four squares and a plus sign inside, followed by a black arrow pointing to the right. Below it is a small green circle with three white dots.

**Table Creator Configuration Dialog:** A modal dialog titled "Dialog - 0:3 - Table Creator". It has two tabs at the top: "Table Creator Settings" (selected) and "Flow Variables".

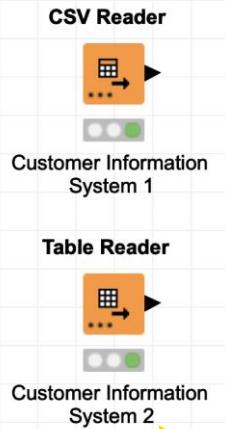
**Table Creator Settings Tab:** Contains an "Input line:" field and a table view.

	Book	Category
Row0	From Alteryx to KNIME	free book
Row1		
Row2		
Row3		

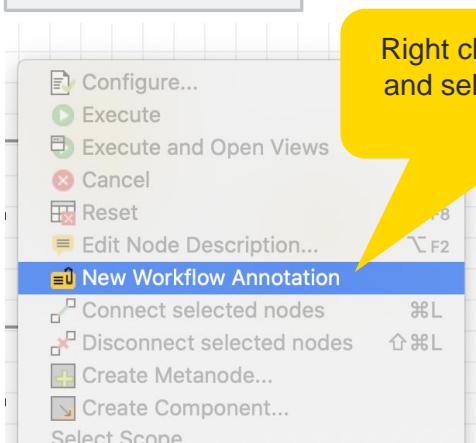
**Buttons at the bottom:** OK - Execute, Apply, Cancel, and a Help button (question mark icon).

# Comments & Annotations

## Comments



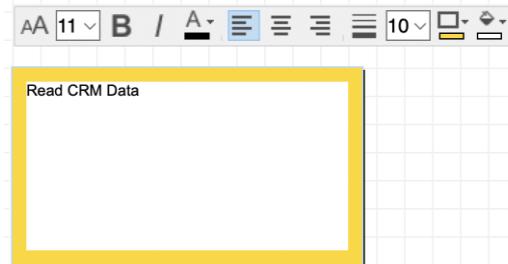
## Annotations



Right click in the workflow  
and select New Workflow  
Annotation



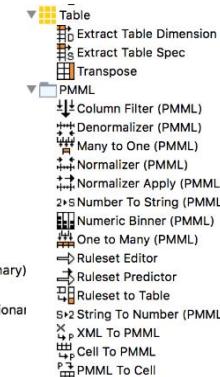
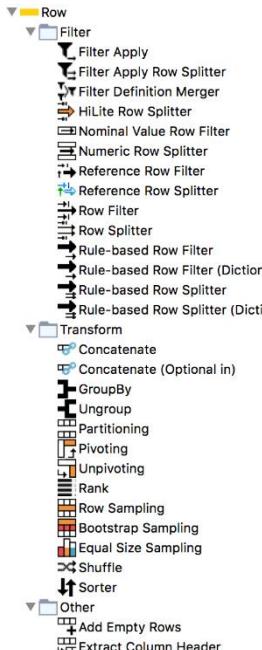
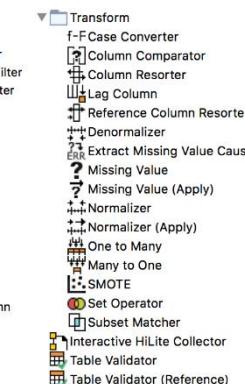
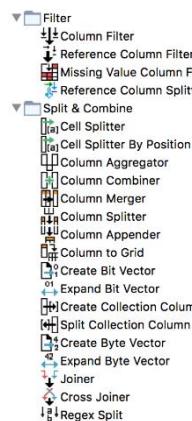
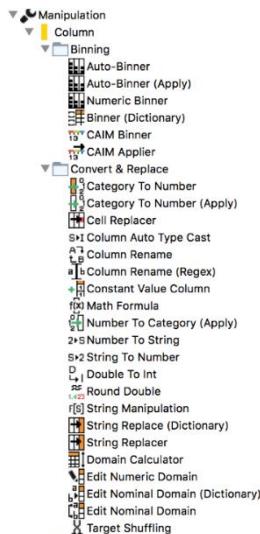
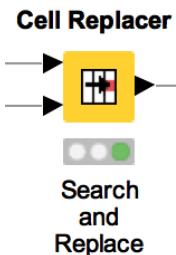
Double-click on the upper  
left corner to open the  
annotation editor



# Data Merging

# Data Manipulation Nodes

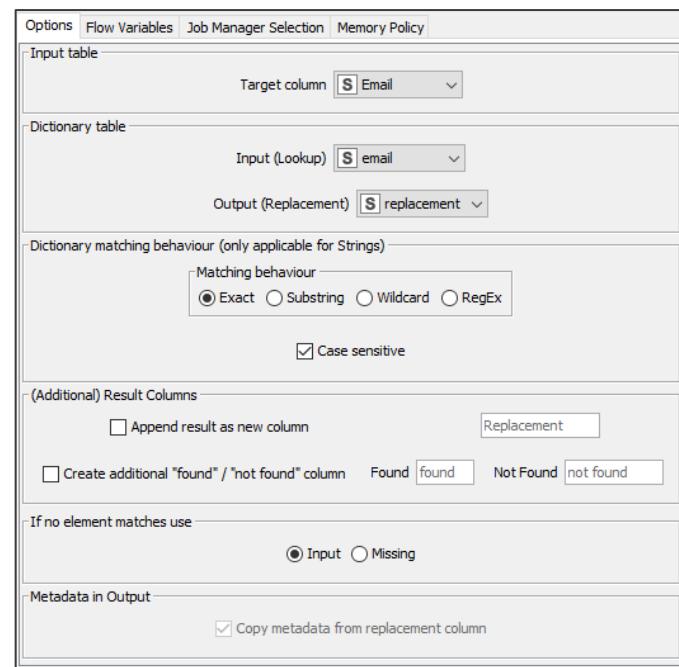
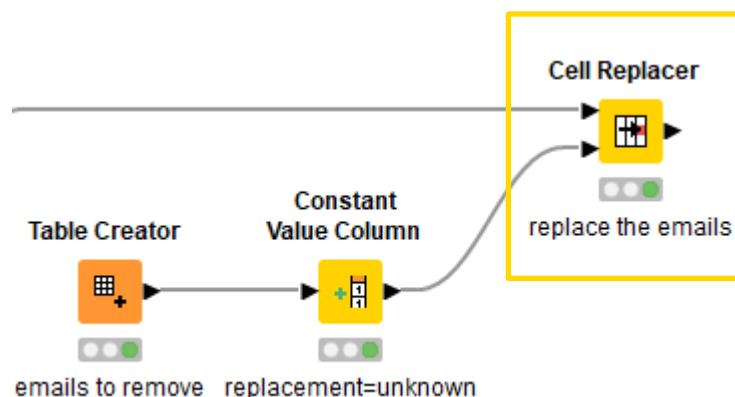
- Yellow color with a variety of input and output ports
- Apply a transformation to input data
- Many, many nodes!



# Cell Replacer

Replaces the content of a column based on a lookup

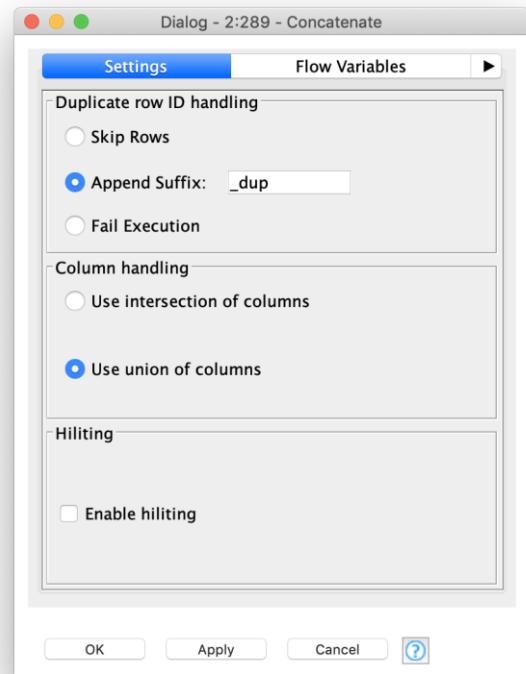
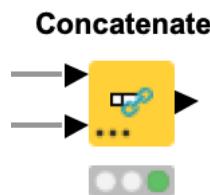
- The top port references the table you want to search
- Bottom port holds the lookup table (search keys and replacement values)



# Concatenate

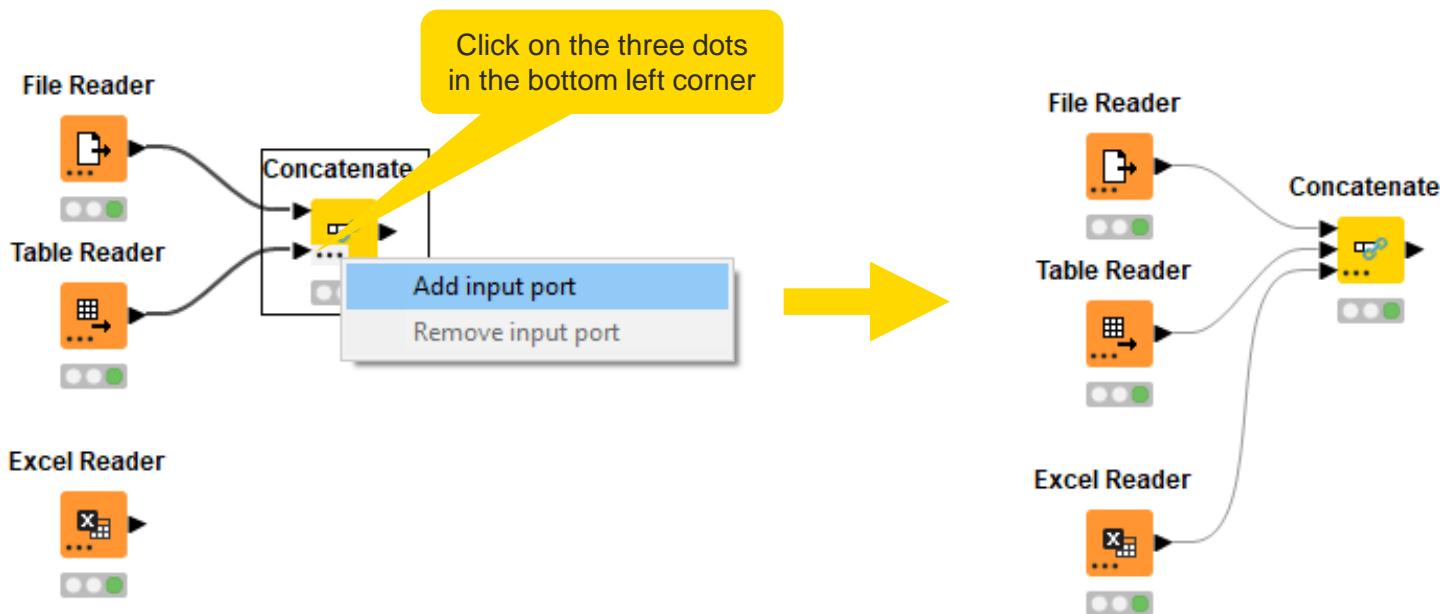
Combine rows from two tables with shared columns

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



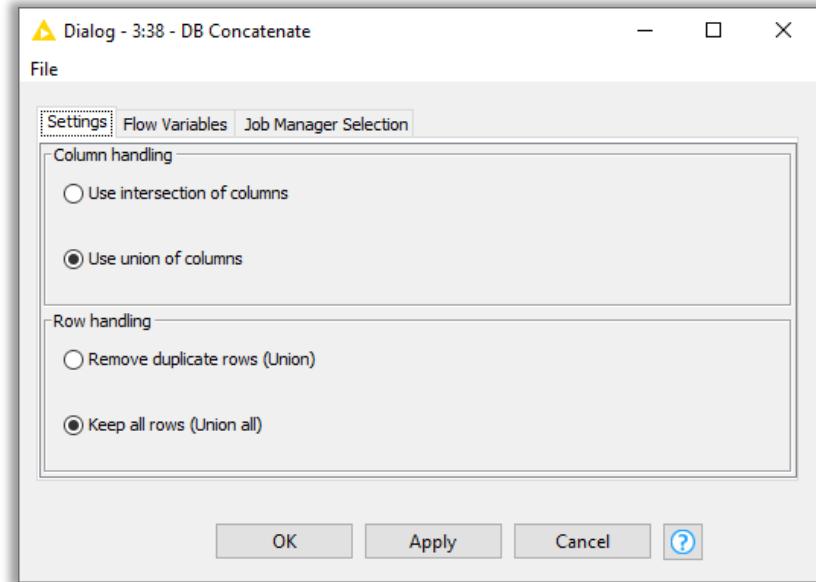
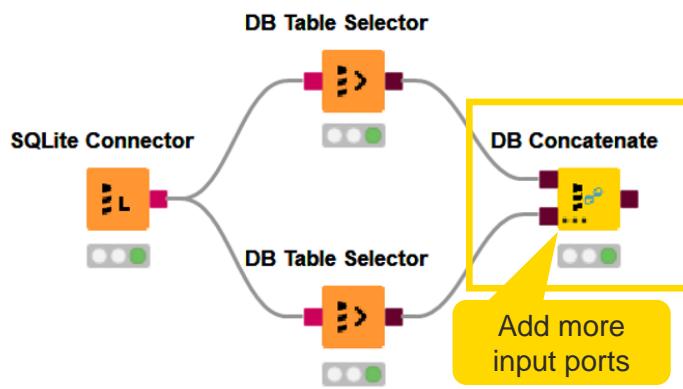
# Dynamic Ports

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

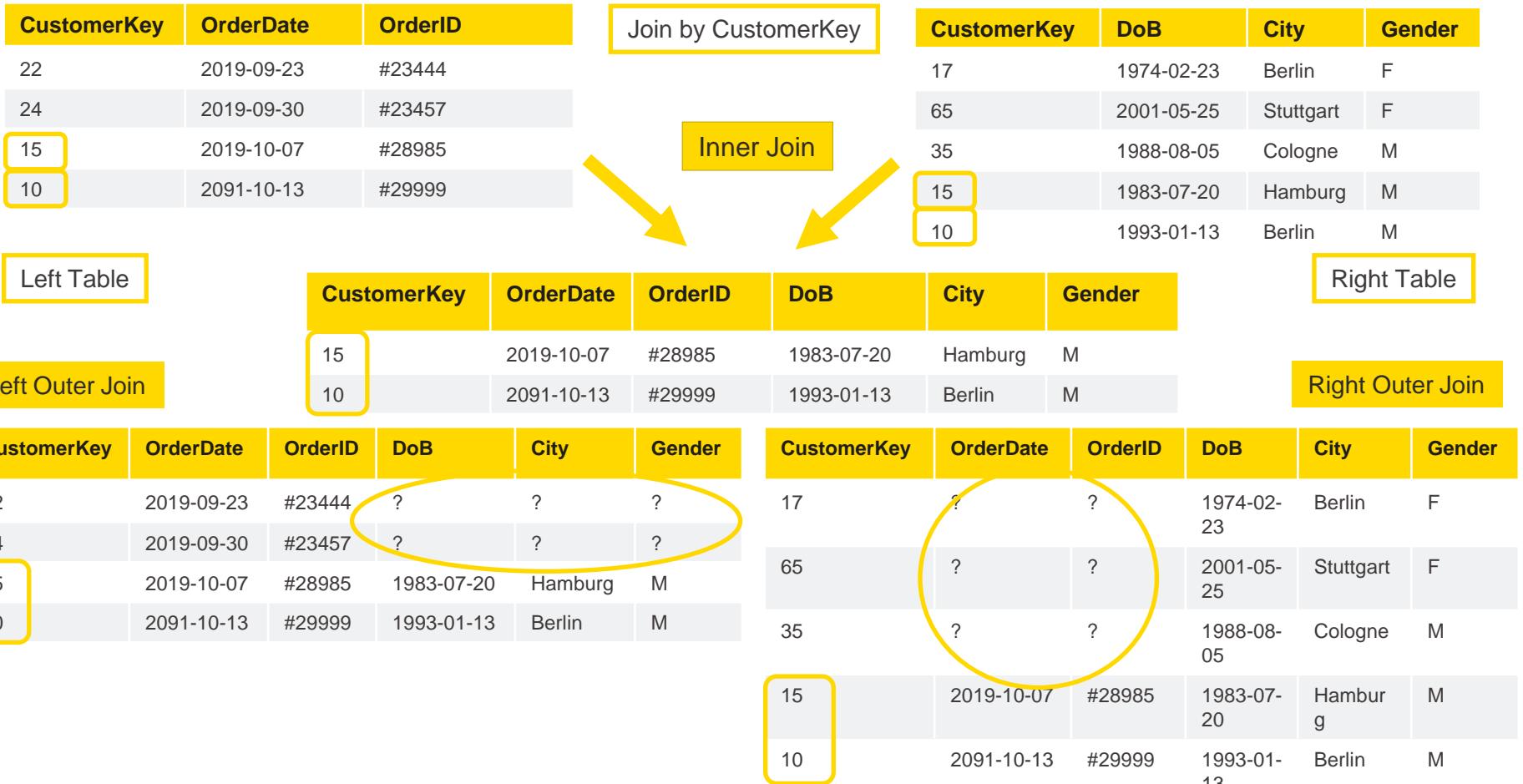


# DB Concatenate

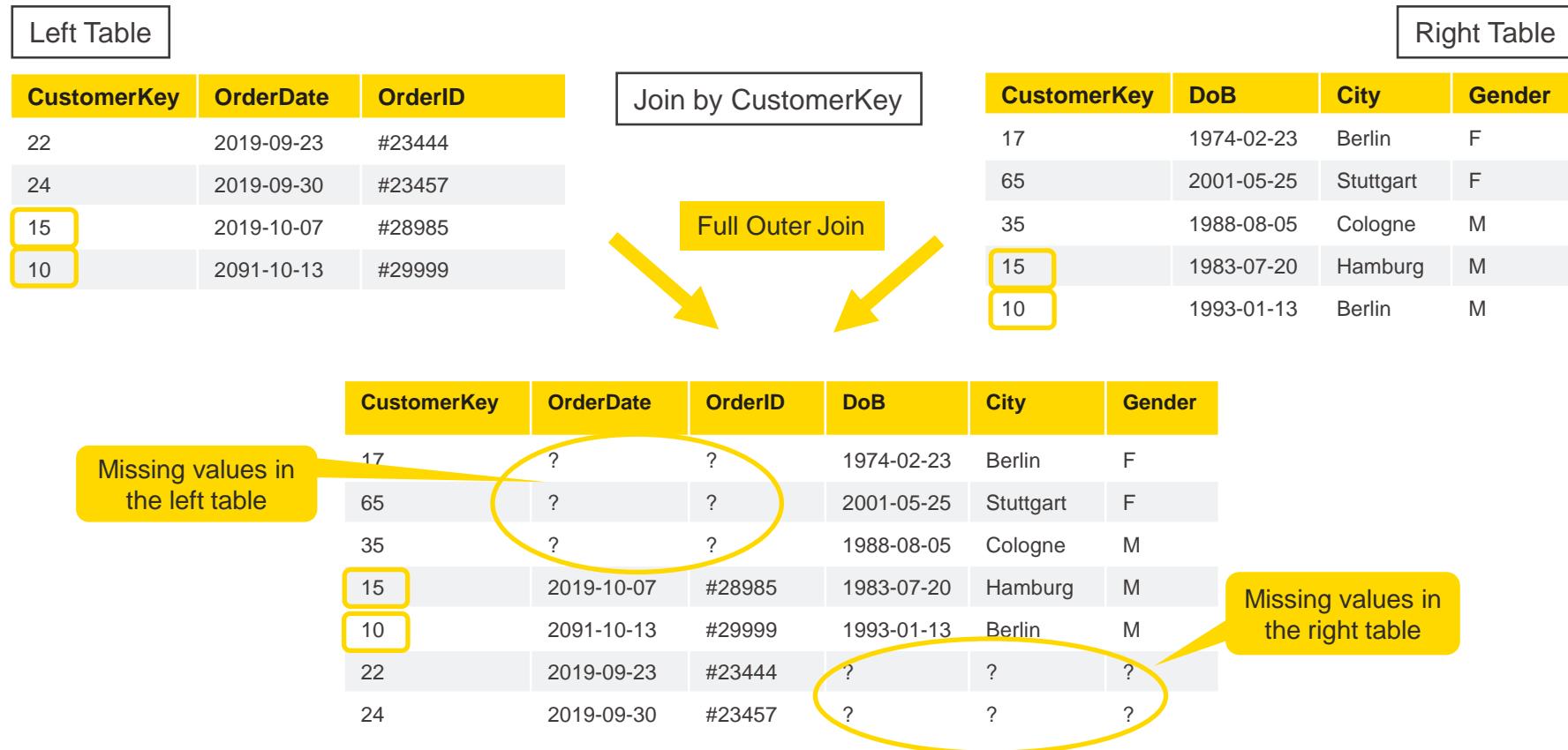
- Combine rows from 2 or more tables with shared columns
- Handles duplicate row keys gracefully
- Take the union or intersection of columns



# Joining Columns of Data

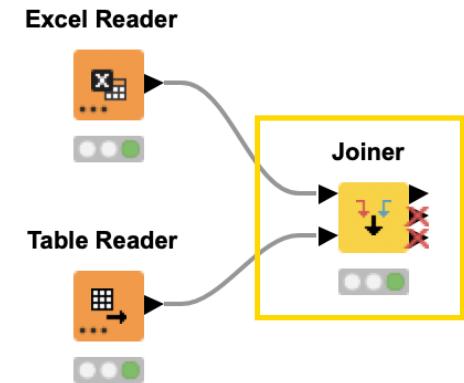


# Joining Columns of Data



# 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

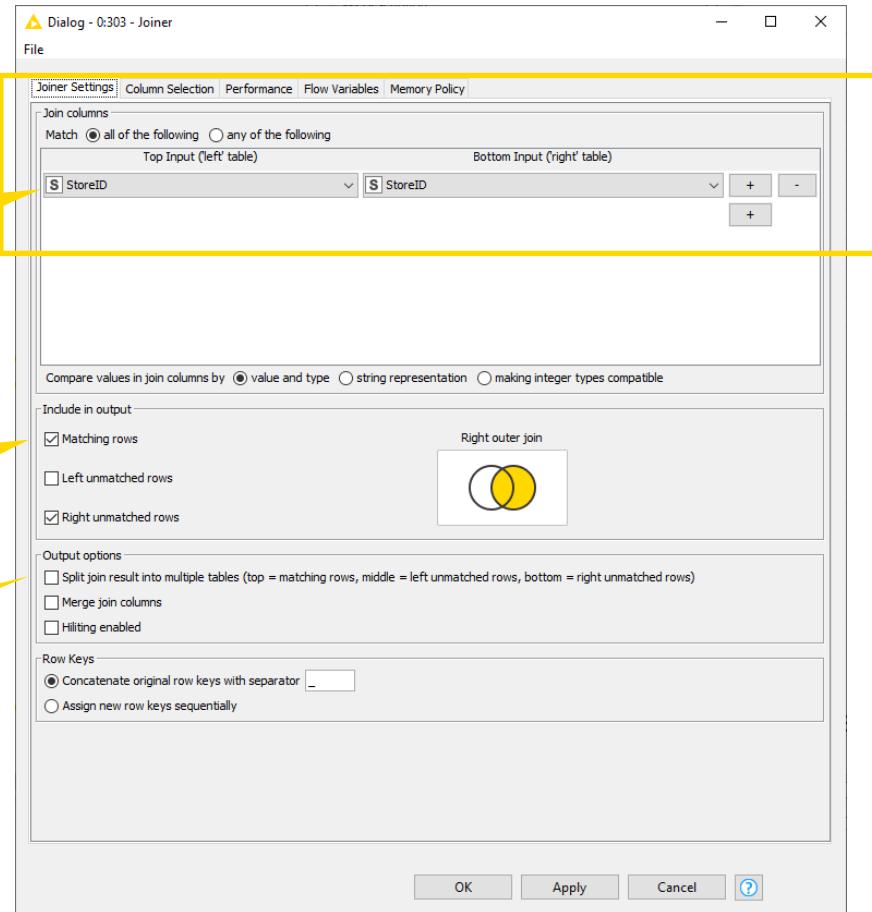


# Joiner Configuration – Linking Rows

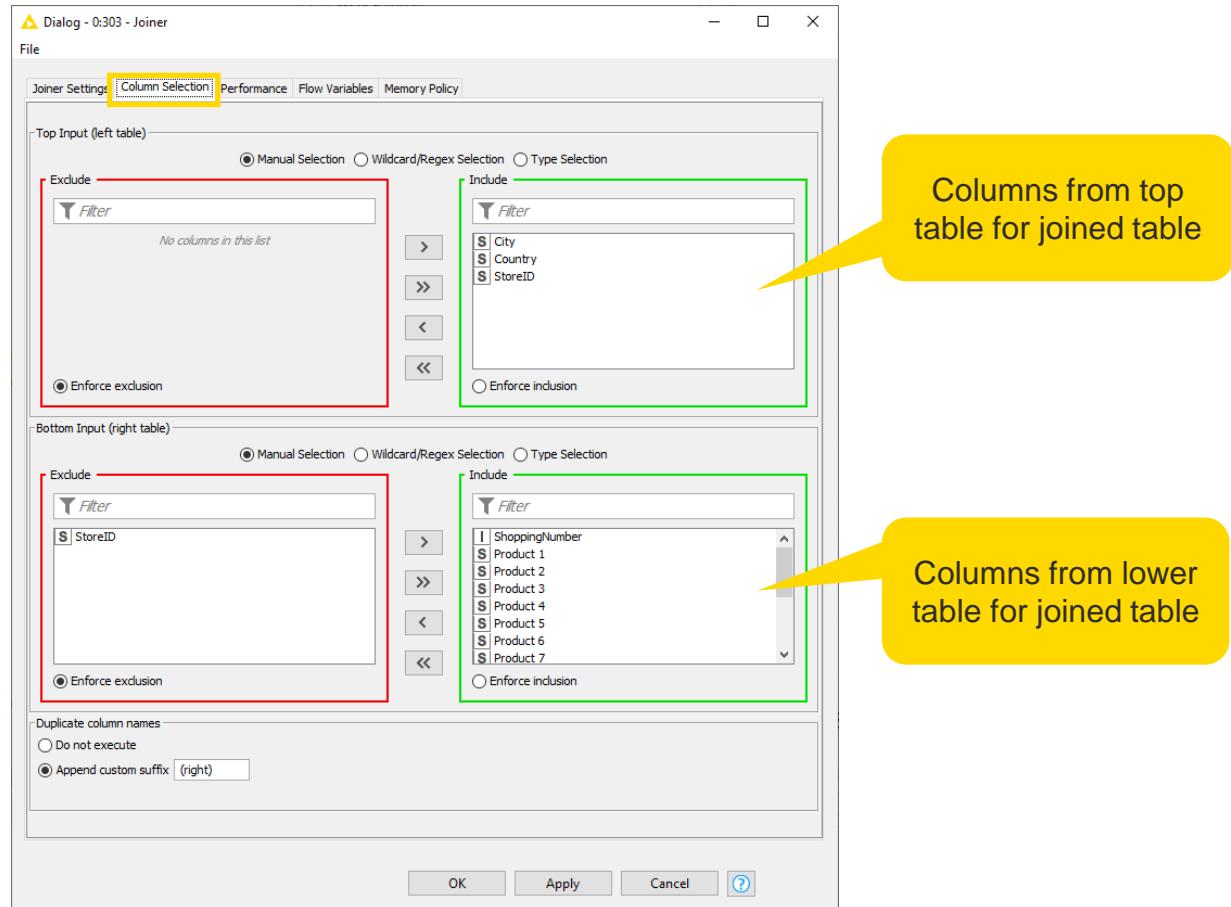
Values to join on.  
Multiple joining columns  
are allowed

Select the rows which  
should be included in the  
joined table

Activate this checkbox to  
activate the bottom  
output ports

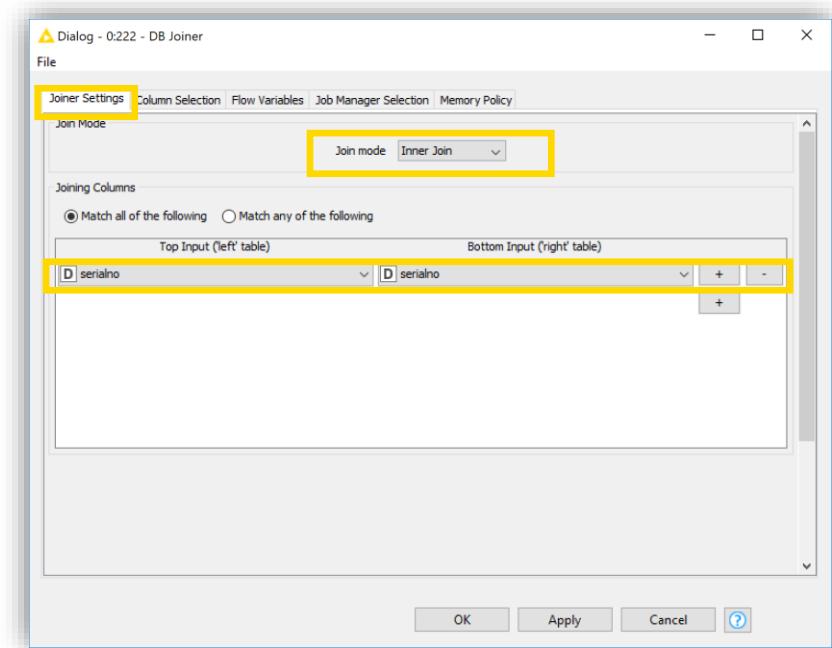
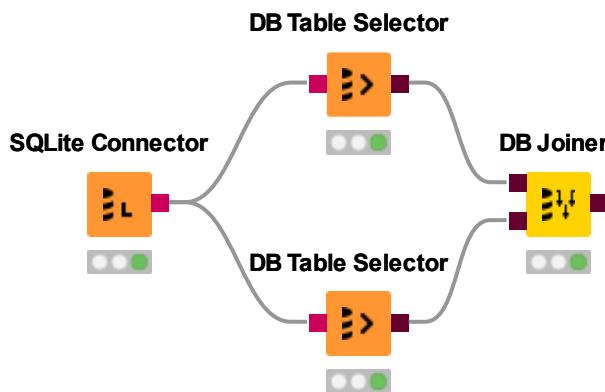


# Joiner Configuration – Column Selection

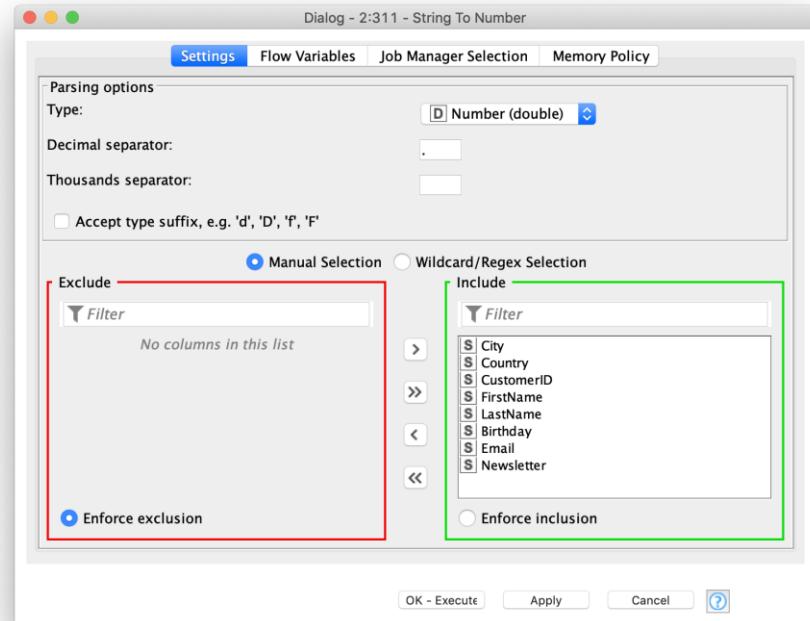
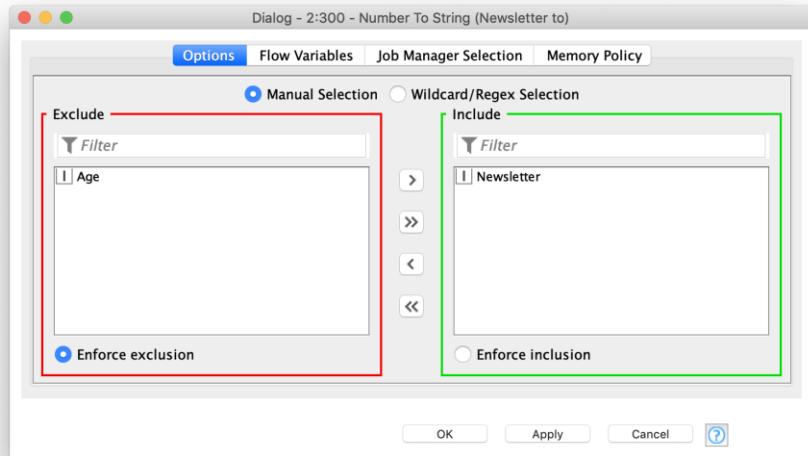


# DB Joiner

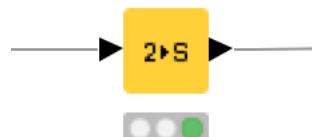
- In-database joiner
- Creates the SQL statement to join two tables stored in the same database
- No coding required



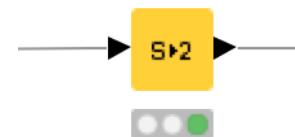
# Type Conversion



**Number To String**

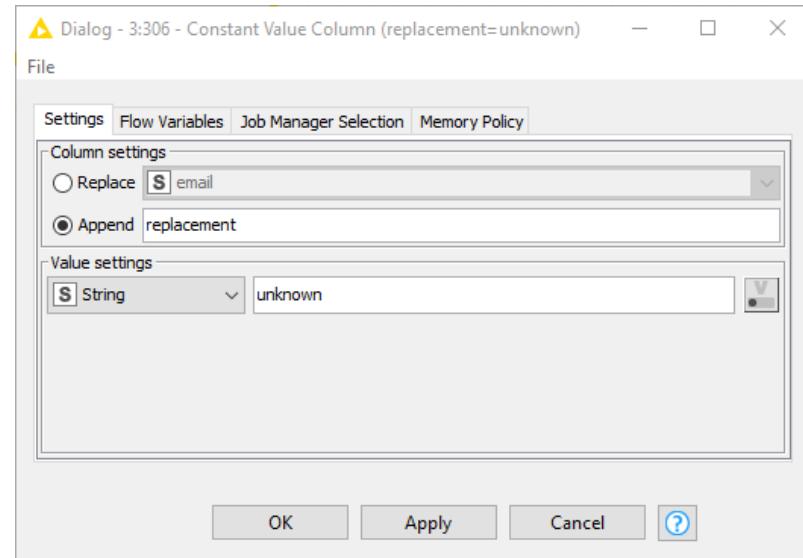
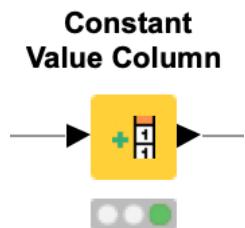


**String To Number**



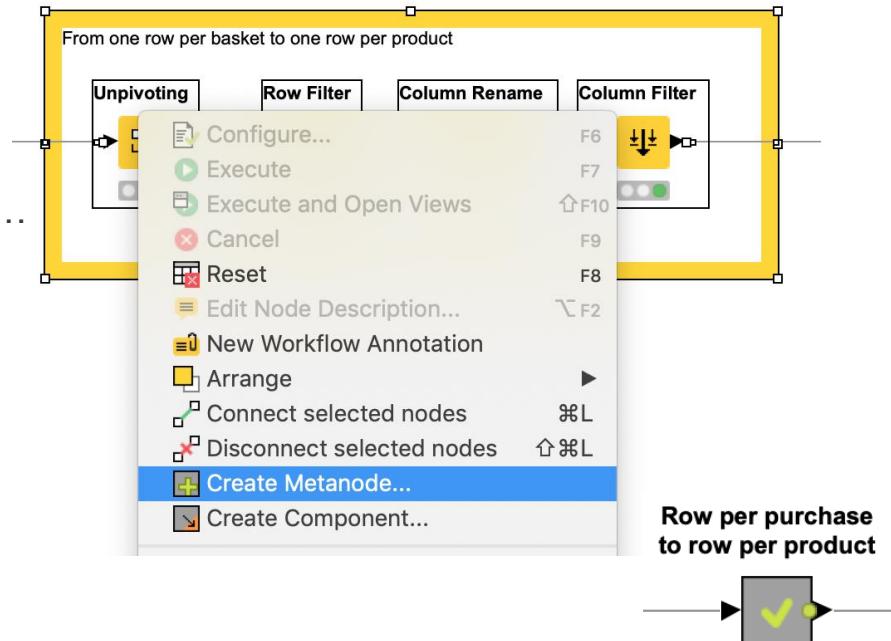
# Constant Value Column

- Adds or replaces a column with a single constant value
- Can be used to add an empty column



# Workflow Organization – Good Practices

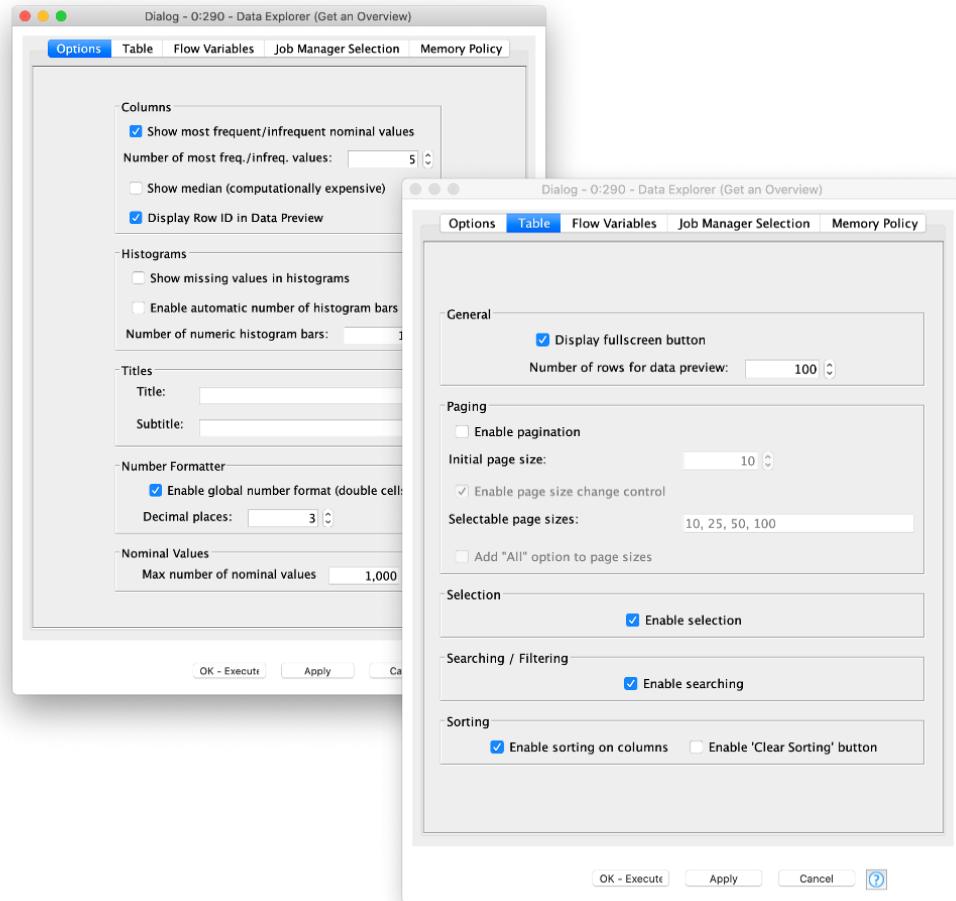
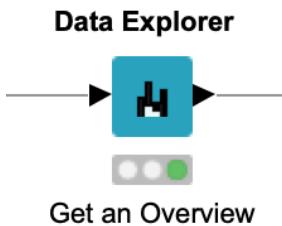
- Workflow annotations
- Node labels
- Metanodes
  - Organize workflow by task
  - Hide complexity & improve readability
  - Select nodes -> Right click -> Create Metanode...



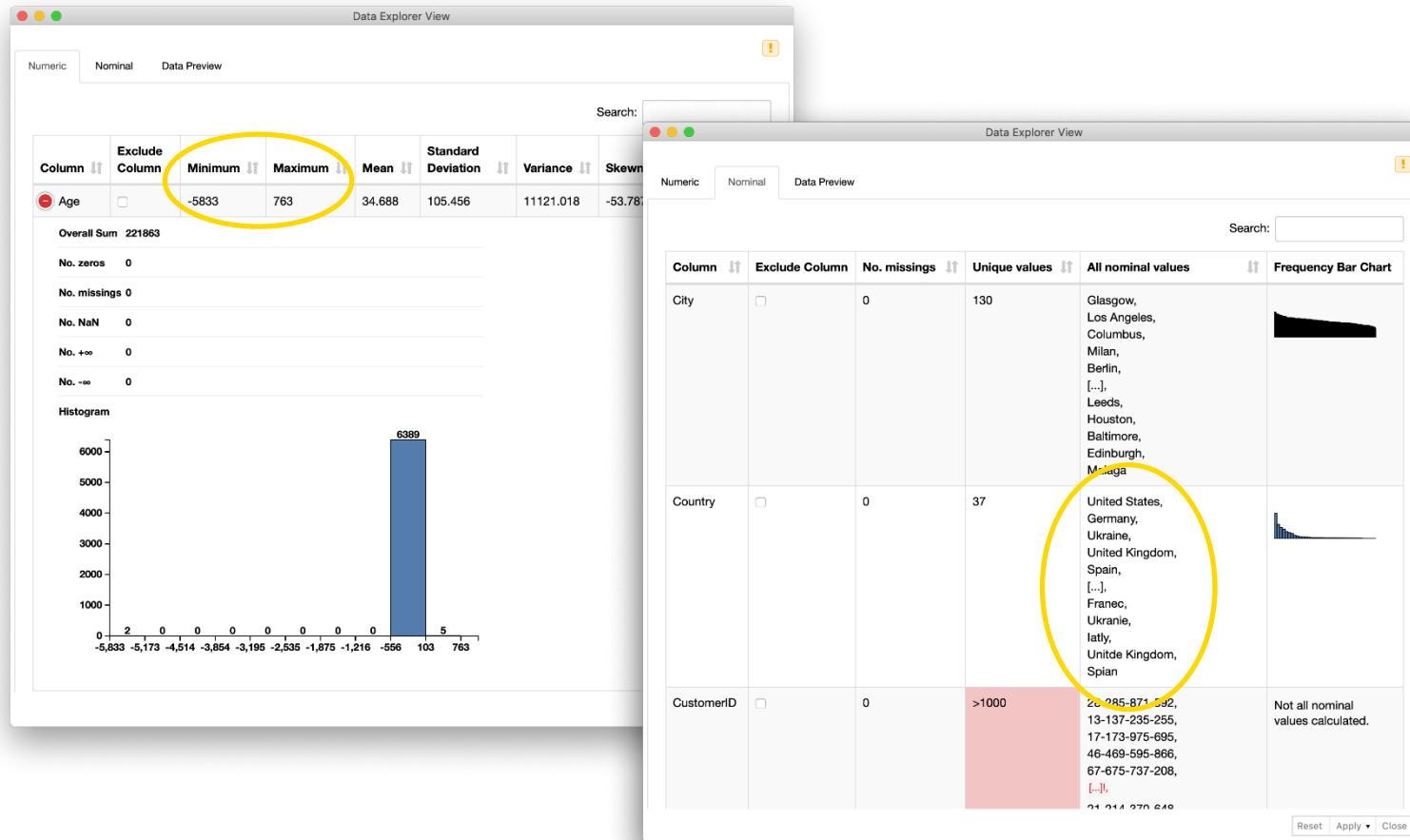
# **Numerical & Nominal Outlier Handling**

# Data Explorer

The Data Explorer node offers a range of options for displaying properties of the input data in an interactive view

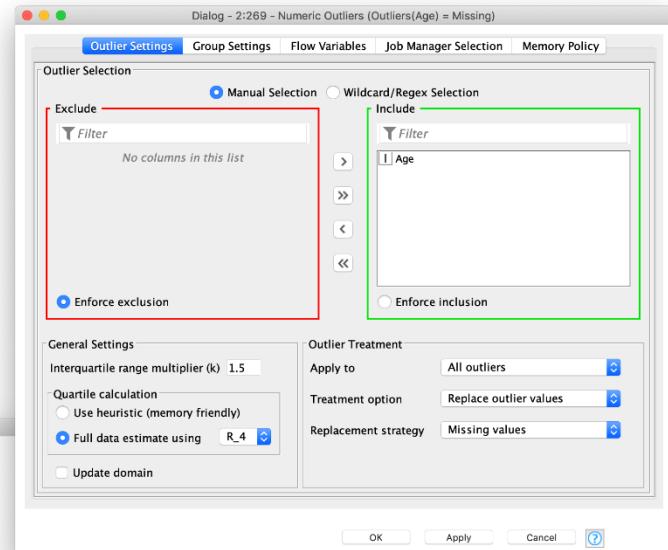
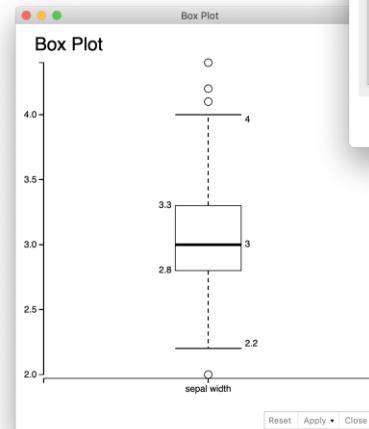


# Customer Data Output

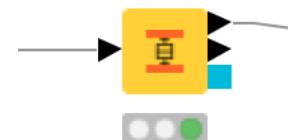


# Numeric Outlier

- Detects and treats outliers
- $x$  is a numeric outlier if
$$x < Q_1 - k * IQR$$
$$x > Q_3 + k * IQR$$
with  $IQR = Q_3 - Q_1$
- For  $k = 1,5$  the boarders correspond to the whiskers of a box plot



## Numeric Outliers



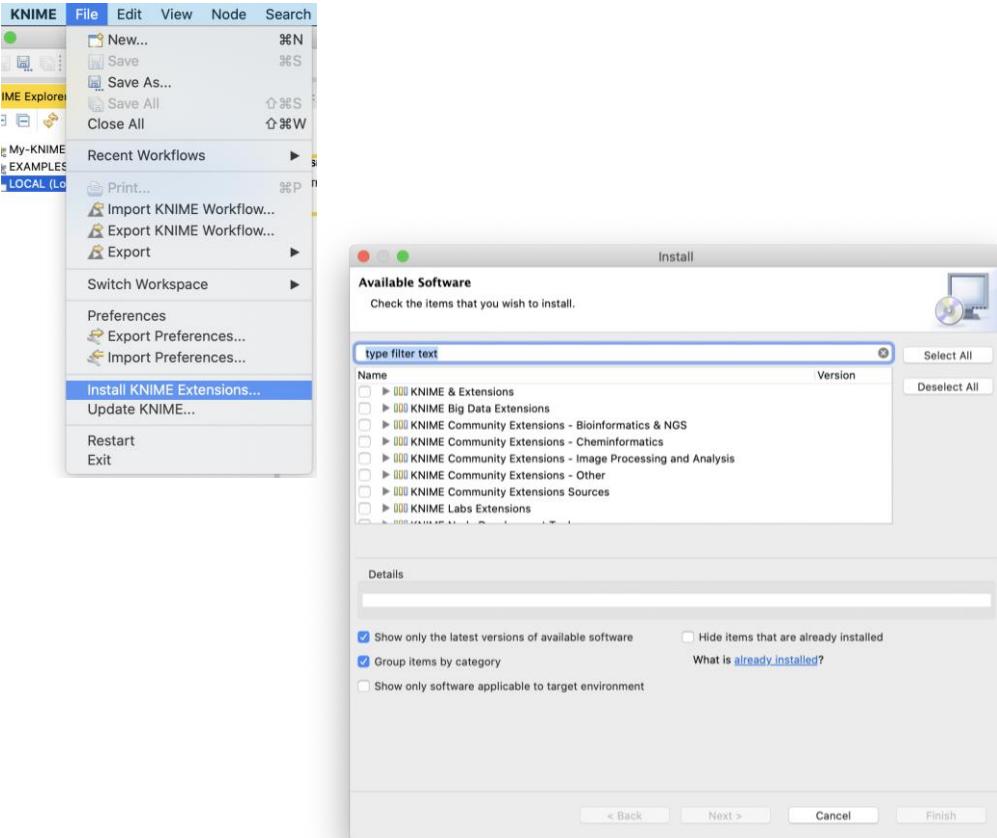
# Calculating the Quantiles - Example

---

- The quantiles  $Q_1$ ,  $Q_2$ , and  $Q_3$  are the three cut points that divide a dataset into four equal-sized groups, after sorting them.
- Example: 3, 6, 7, 8, 8, 10, 13, 15, 16, 20
  - $Q_1$  : Rank =  $10 \times (1/4) = 2.5$ , rounds up to 3 =>  $Q_1 = 7$
  - $Q_2$  : Rank =  $10 \times (2/4) = 5$ , =>  $Q_2 = (8+10)/2=9$
  - $Q_3$  : Rank =  $10 \times (3/4) = 7.5$ , rounds up to 8 =>  $Q_3 = 15$

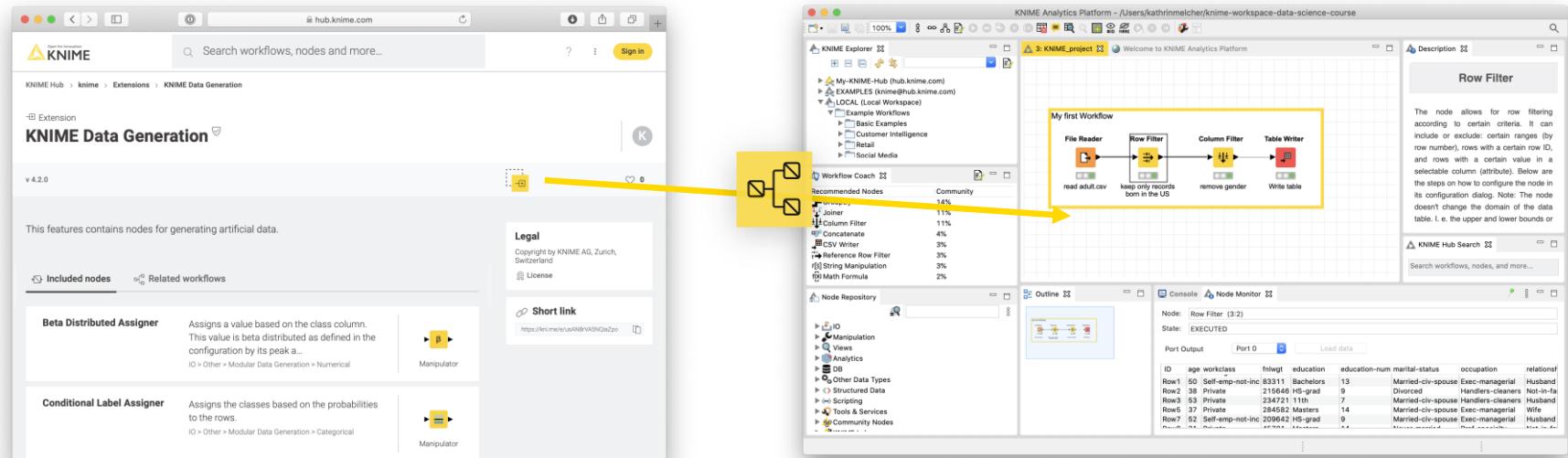
# More Nodes – Installing Extensions

- Go to File -> Install Extensions
- Select the following extensions:
  - KNIME JavaScript (Labs)  
(in the folder KNIME Labs Extensions)
  - KNIME Data Generation  
(in the folder KNIME & Extensions)



# Alternative Way via the KNIME Community Hub

- Drag & Drop extension from the KNIME Community Hub into KNIME Analytics Platform.



# Exercise: 02\_Data Merging

---

- **Concatenate the customer information from the two systems**
  - In addition, create a table with a few emails selected from the data table, to replaced with “unknown”
  - Add a constant column to the newly created table, with the value “unknown”
  - Replace the email addresses in the data with unknown using Cell Replacer
- Add the price information to each online product purchase (DB Joiner) and read the table into KNIME (DB Reader)
- Add the location information to each purchase in a store based on the StoreID (Joiner node)
- Create three metanodes to clean up your workflow
  - Customer data
  - Online transactions & product+price (two output ports)
  - Onsite purchases in stores

# Exercise: 03\_Data Cleaning (Part 1 & 2)

---

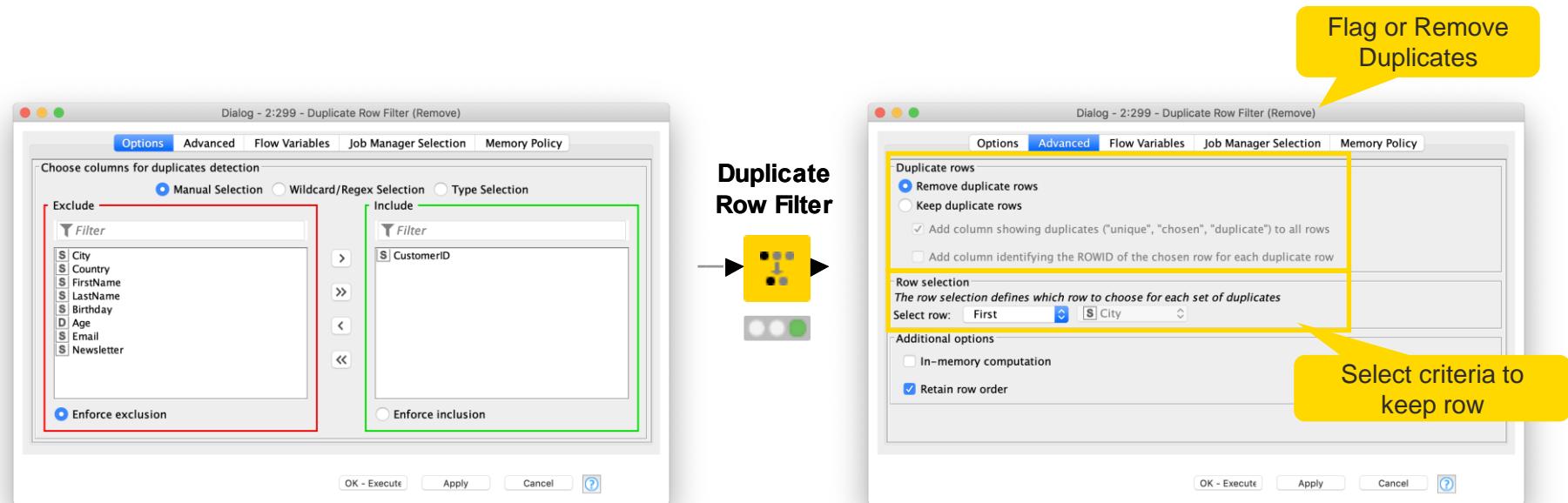
- Explore the data using the Data Explorer node
- Replace numeric outliers in the “Age” column with missing values

# Duplicate and Missing Value Handling

# Duplicate Row Filter

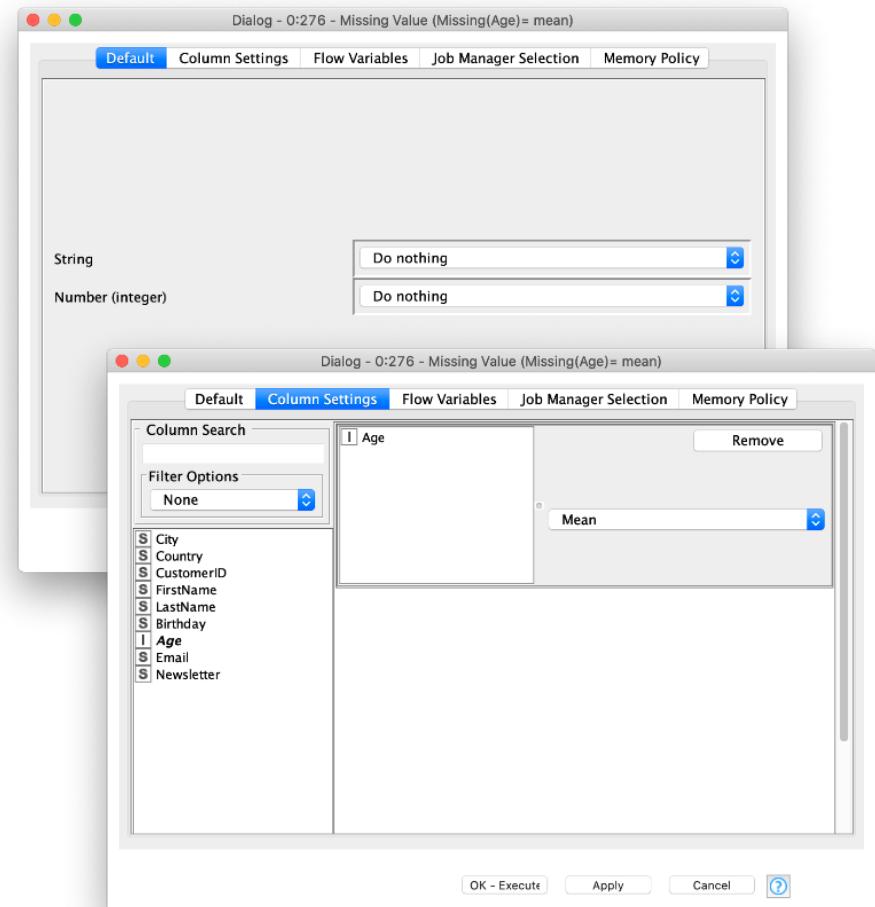
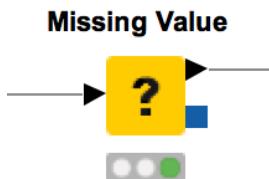
Detects duplicate rows and apply a selected treatment

- First tab provides the option to select columns for duplicate detection
- Second tab provides options for treating duplicated values



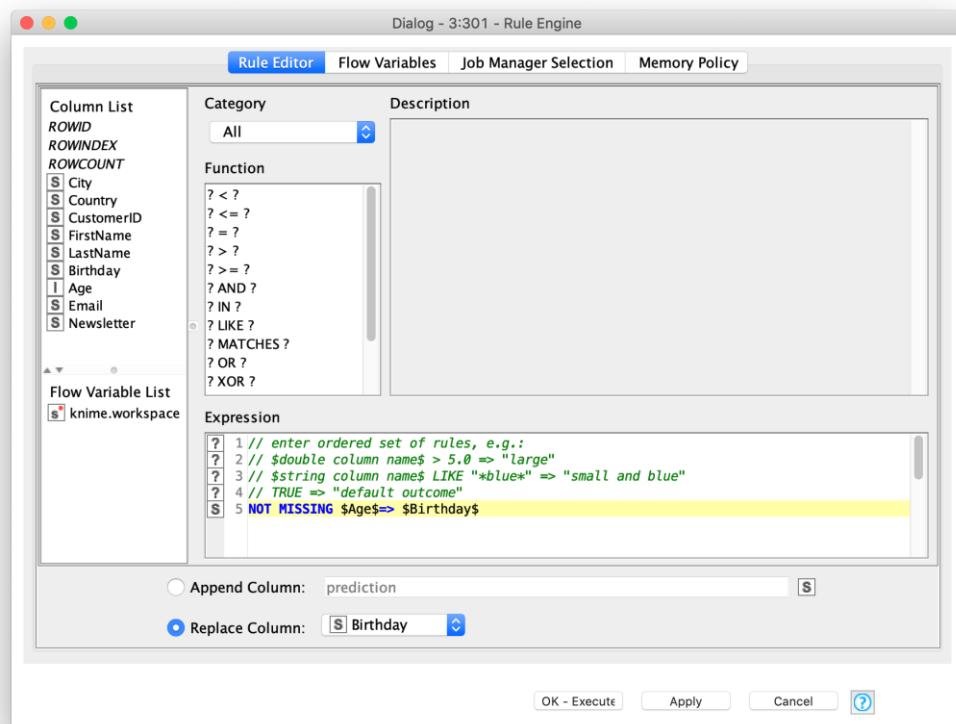
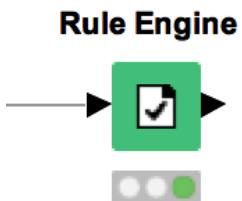
# Missing Value

- Defines how to handle missing values for all columns of a given type
  - Affects all columns that are not explicitly mentioned in the second tab
- Defines how to handle missing values for each available column

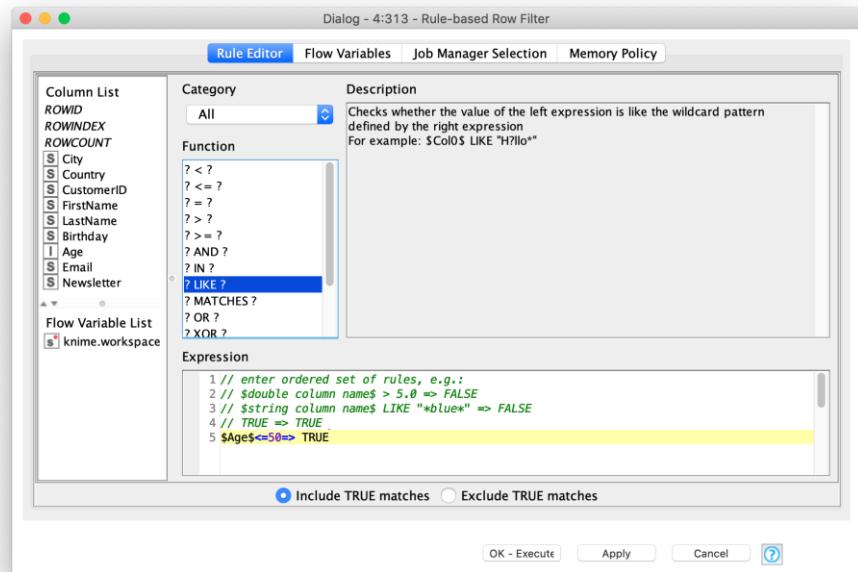
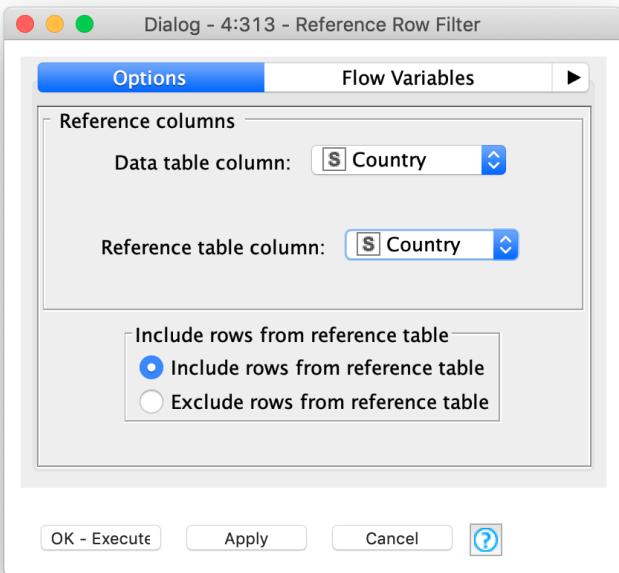


# Rule Engine

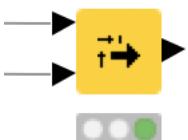
- Defines custom logic to use simple rules
- Rules like: <Antecedent/Condition> => <Consequence>
  - ( $1=1 \Rightarrow$  "true")
- Tries to match rules to each row of the input table



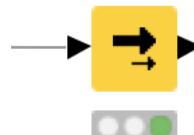
# Other Options to Filter Rows



Reference  
Row Filter



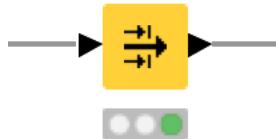
Rule-based  
Row Filter



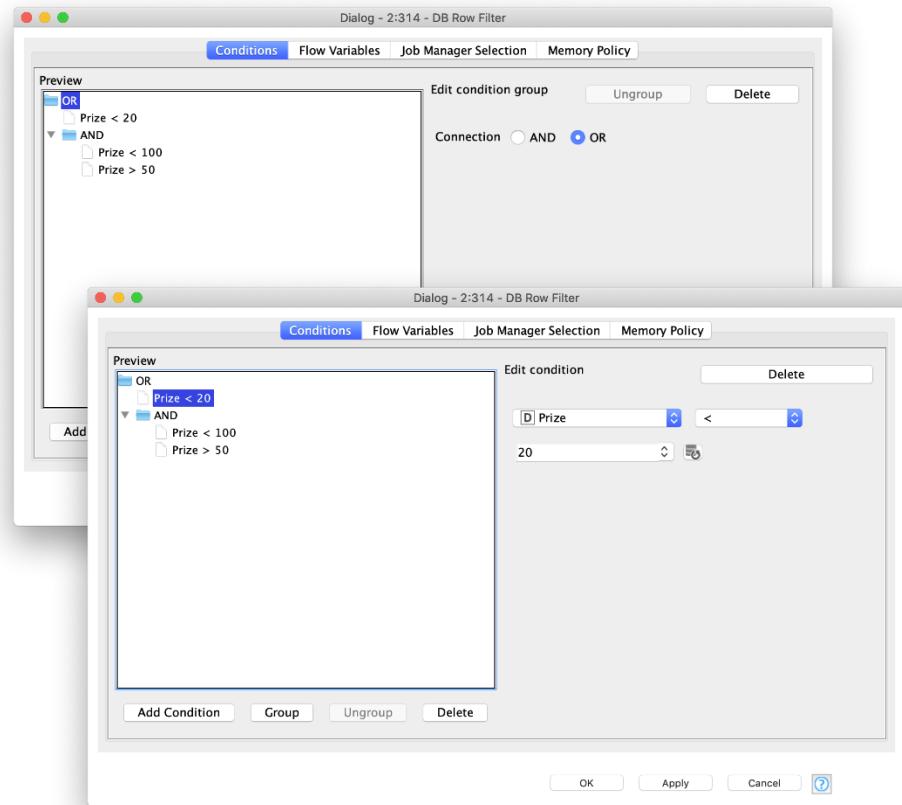
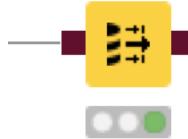
# In-Database Row Filtering – DB Row Filter

- Creates a SQL statement to filter the rows that don't match the conditions
- More than one condition is possible
- Allows you to create logical groups for AND and OR

Row Filter (Labs)

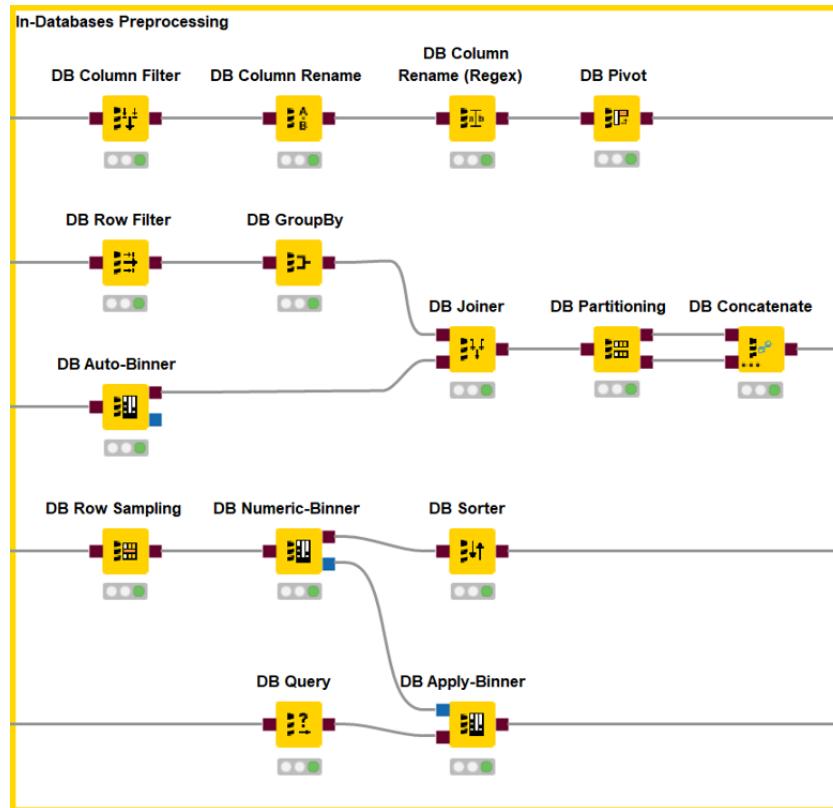


DB Row Filter



# Query Nodes

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



# **Standardization**

# Standardize Table Format

## Onsite transaction data

⚠ Read table - 0:1 - Table Reader

File Edit Hilitc Navigation View

Table "default" - Rows: 20000 Spec - Columns: 16 Properties Flow Variables

Row ID	I Shoppin...	S Product 1	S Product 2	S Product 3	S Product 4	S Product 5	S produ...
Row900	480105758	V-220-2011	A-43-2005	W-52-2003	?	?	?
Row901	480105759	Q-166-2005	I-184-2010	?	?	?	?
Row902	480105760	V-181-1983	W-52-2018	?	?	?	?
Row903	480105761	N-160-1983	I-184-1989	Z-280-2003	C-151-2015	A-91-2002	?
Row904	480105762	H-172-1991	V-265-1991	H-235-1996	H-124-1998	Q-43-2015	?
Row905	480105763	7-160-1994	G-226-2000	N-105-2000	Z-01-1990	U-42-1992	G-40-2014

## Online transaction data

⚠ KNIME data table - 0:6 - DB Reader

File Edit Hilitc Navigation View

Table "database" - Rows: 11679 Spec - Columns: 5 Properties Flow Variables

Row ID	I OrderN...	S Date	S CustomerID	S ProductNr	D Price
Row0	23893756	8-28-2015	69-695-442-229	I-105-2017	12.99
Row1	23893756	8-28-2015	69-695-442-229	B-172-2005	58.99
Row2	23893756	8-28-2015	69-695-442-229	W-181-2003	75.59
Row3	23893756	8-28-2015	69-695-442-229	F-055-2017	111.99
Row4	23893756	8-28-2015	69-695-442-229	F-289-2008	37.59
Row5	23893756	8-28-2015	69-695-442-229	F-289-1990	47.99

## Product Nr & Price

⚠ KNIME data table - 0:4 - DB Reader

File Edit Hilitc Navigation View

Table "database" - Rows: 50700 Spec - Columns: 2 Properties Flow Variables

Row ID	D Price	S ProductNr
Row45899	66.99	C-001-2010
Row45900	19.99	V-061-2010
Row45901	110.59	B-001-2010
Row45902	63.99	N-061-2010
Row45903	54.99	Q-064-2010
Row45904	48.99	W-064-2010

# Data Transformation

ID	City	Product 1	Product 2	ID	City	Product
234	Berlin	Pear	Apple	234	Berlin	Pear
235	London	Nuts	Pear	234	Berlin	Apple
236	Boston	Rice	Grapes	235	London	Nuts
237	Paris	Pasta	Apple	235	London	Pear
				236	Boston	Rice
				236	Boston	Grapes
				237	Paris	Pasta
				237	Paris	Apple

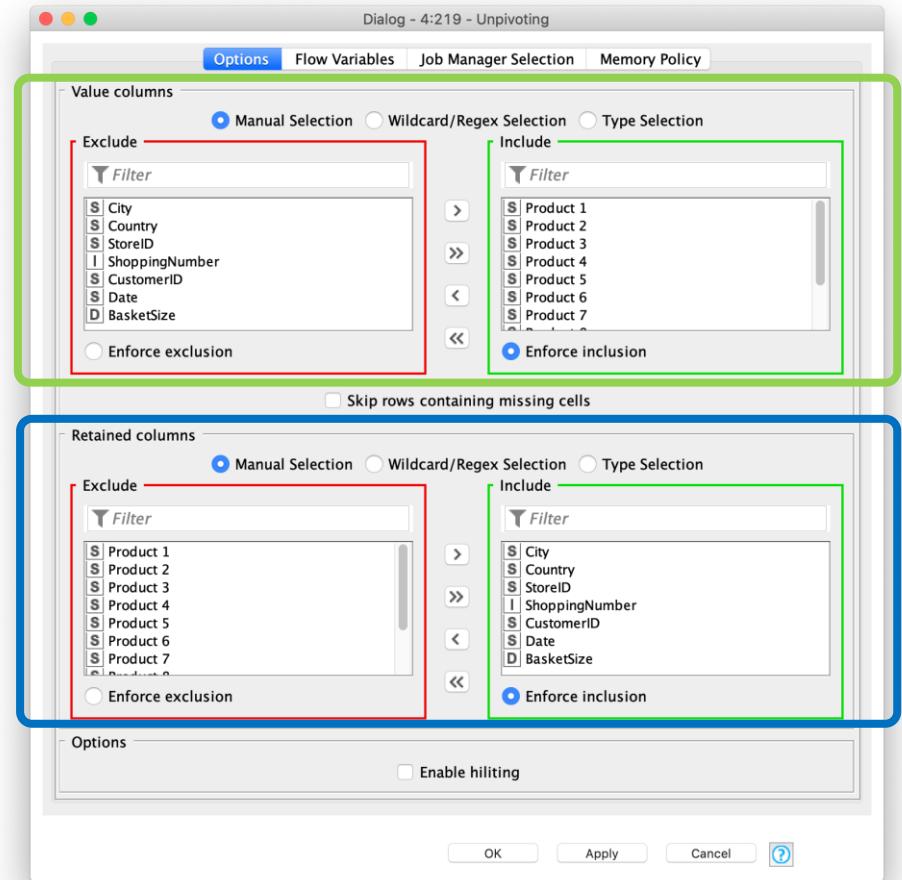
Value Column

Retaining Columns

Solution: Unpivoting Node

# Unpivoting

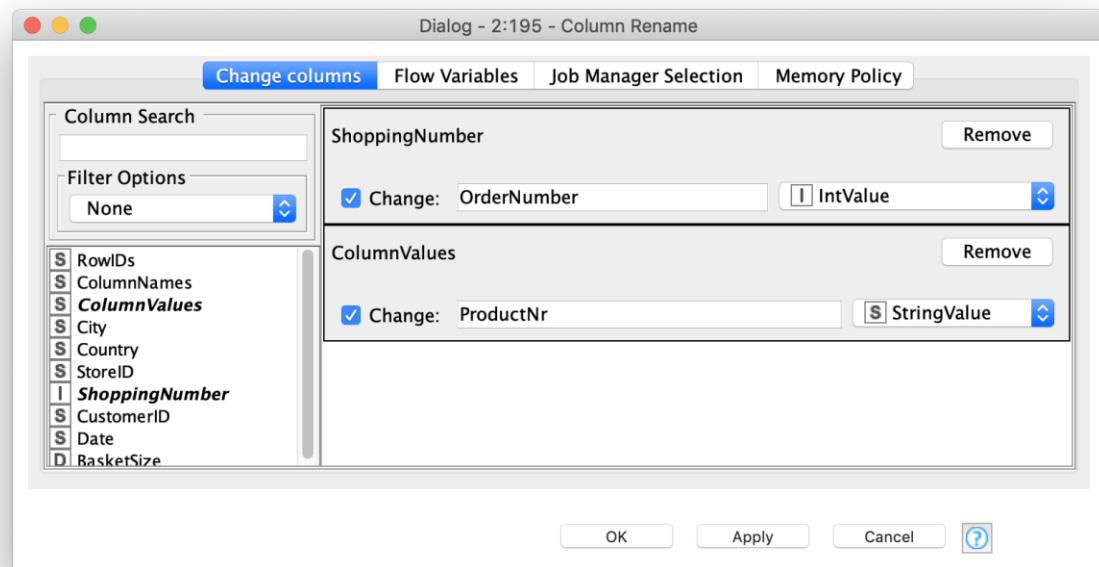
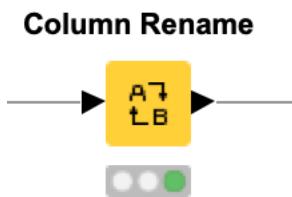
- Rotates the value columns to rows
- Duplicates the remaining columns and appends them to each corresponding row



Value Column  
Retaining Columns

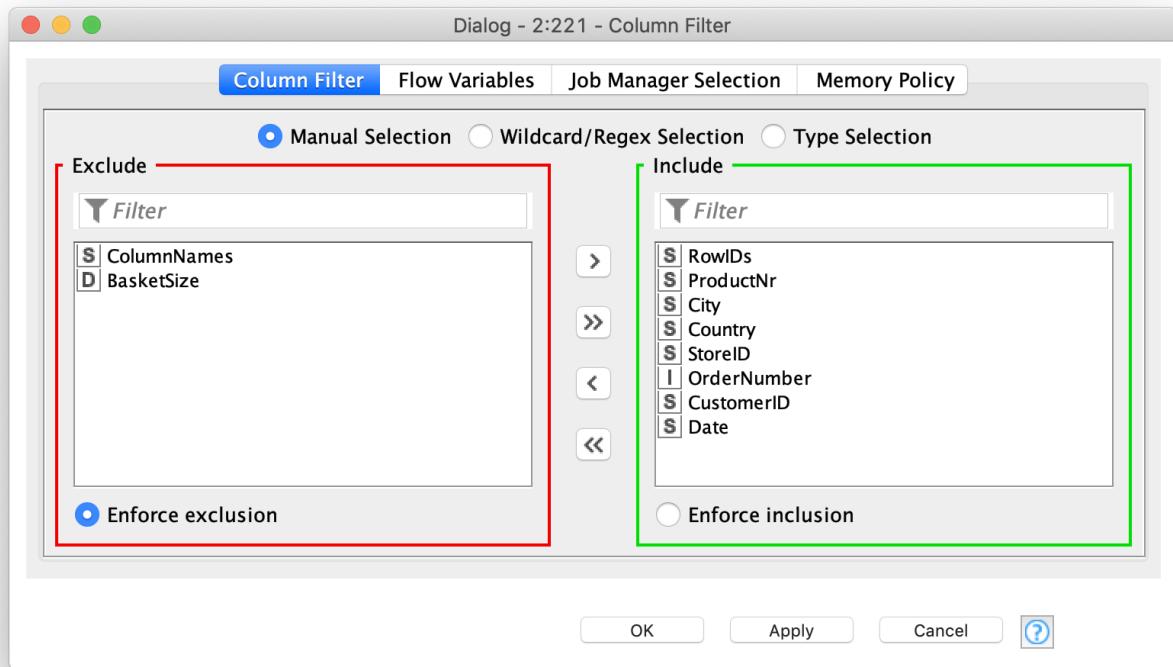
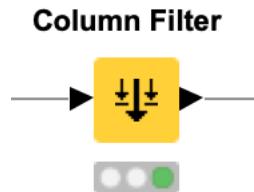
# Column Rename

- Renames column names or changes their types



# Column Filter

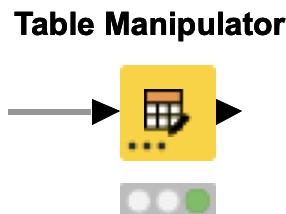
- Excludes columns from the table by moving them to the Exclude list



# Table Manipulator

Allows for

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



Dialog - 0.3 - Table Manipulator

File

Settings Flow Variables Memory Policy

Row ID handling

Use existing row ID  Prepend table index to row ID

Transformations

Reset actions    Enforce types Take columns from:  Union  Intersection

	Column	New name	Type
⋮	<input checked="" type="checkbox"/> \$ City		\$ String
⋮	<input checked="" type="checkbox"/> \$ Country		\$ String
⋮	<input checked="" type="checkbox"/> \$ CustomerID	ID	\$ String
⋮	<input checked="" type="checkbox"/> \$ FirstName		⌚ String → Local Time
⋮	<input checked="" type="checkbox"/> \$ LastName		⌚ String → Number (double)
⋮	<input checked="" type="checkbox"/> \$ Birthday		⌚ String → Number (integer)
⋮	<input checked="" type="checkbox"/> \$ Age		⌚ String → Number (long)
⋮	<input checked="" type="checkbox"/> \$ Email		⌚ String → PXML
⋮	<input checked="" type="checkbox"/> \$ Newsletter		⌚ String → Period
⋮	<input checked="" type="checkbox"/> ? <any unknown new column>		⌚ String → SVG Image
⋮			\$ String

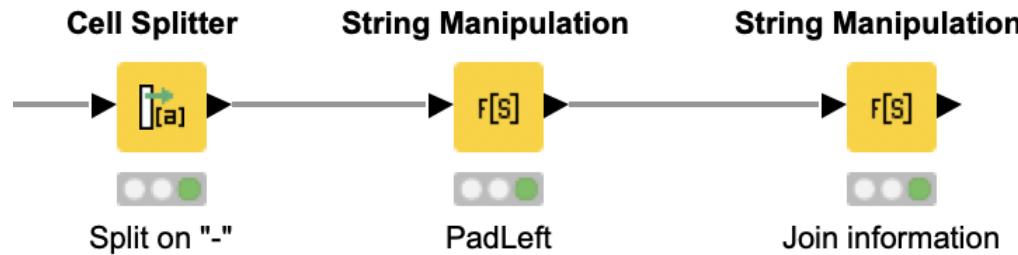
Preview

Data analysis successfully completed.

Row ID	\$ City	\$ Country	\$ ID	\$ FirstName	\$ LastName	\$ Birthday	\$ Age	\$ Email	\$ Newsletter
Row0	Glasgow	United Kingdom	17-171-832-104	Alois	Berger	23.9.1972	47	Alois.Berger@mcr.com	0
Row1	Szczecin	Poland	37-370-580-177	Michaela	Schultz	9.6.1998	21	Michaela.Schultz@mcr.com	0
Row2	Sheffield	United Kingdom	27-270-743-182	Rotraut	GrÄ¼nwald	20.4.1975	44	Rotraut.GrÄ¼nwald@mcr.com	0

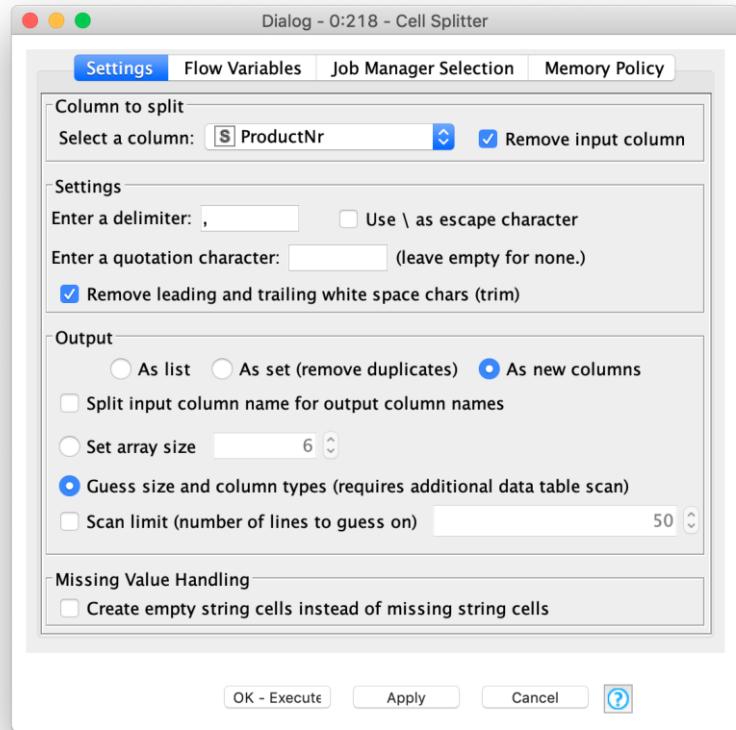
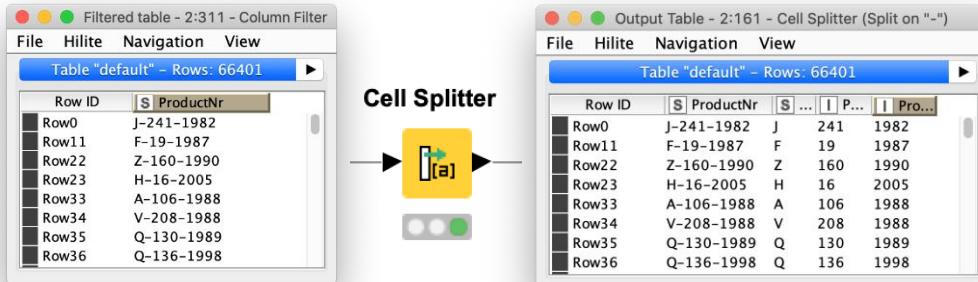
OK Apply Cancel ?

# Standardization of Product Numbers



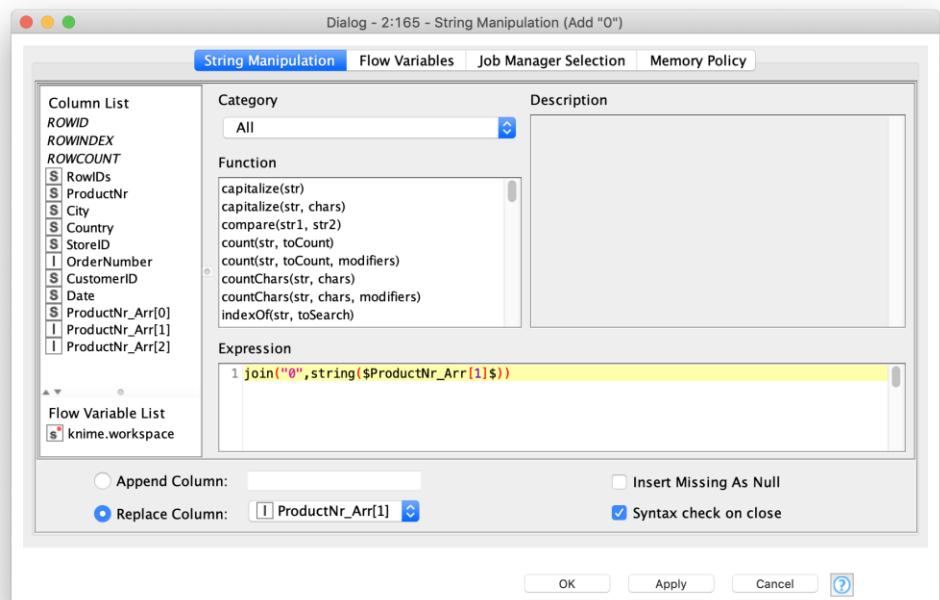
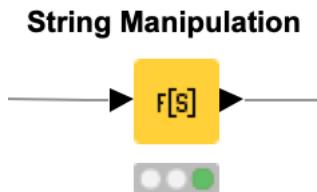
# Cell Splitter

Splits the content of one column into many columns based on a delimiter



# String Manipulation

- Modifies existing strings or creates new columns
  - Cleans up capitalization
  - Joins string values
  - Pads strings, e.g. padLeft
  - Replaces string values



# Exercise: 03\_Data Cleaning (Part 3 & 4)

---

- If the age of a customer is missing, replace the birthday with a missing value  
Hint: Use the expression NOT MISSING \$Age\$=> \$Birthday\$
- Impute the missing values in the age column with the column mean
- Remove rows for duplicate CustomerIDs
- Correct the spelling mistakes in the “Country” column (Optional)
  - Extract the values with spelling mistakes
  - Manually define the correct spelling for the lookup table
  - Create the lookup table automatically, using a similarity search

# Exercise: 04\_Data\_Transformation

---

- Change the structure of the table with the onsite purchases so that each purchased product is in a separate row and not the whole purchase event
  - Unpivot the columns that show the products ordered in one purchase event. Retain other columns in the table.
  - Remove rows that have missing values
  - Rename the "ColumnValues" column to "ProductNr" and "ShoppingNumber" to "OrderNumber" and remove unnecessary columns
- Optional:
  - Standardize the Product Numbers

# Exercise: 05\_Data\_Manipulation (optional)

---

- Add the price to the onsite product purchase data
- Add transaction types to each product purchase
  - "Store - no CC" if the customer ID is not available in the onsite transaction
  - "Store - CC" if the customer ID is available in the onsite transaction
  - "OnlineStore" for the orders coming from the online store
- Concatenate the data of online and onsite purchases
- Add the customer information to each transaction

# Data Aggregation

# Data Aggregation

RowID	Group	Value	
r1	m	2	
r2	f	3	
r3	m	1	
r4	f	5	
r5	f	7	
r6	m	5	



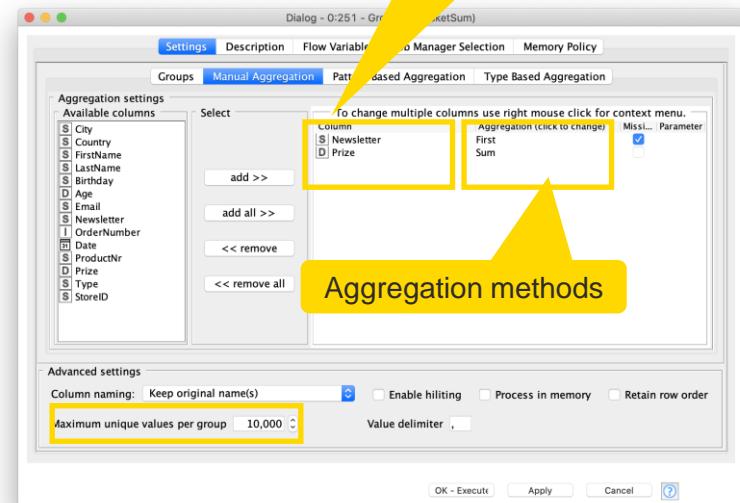
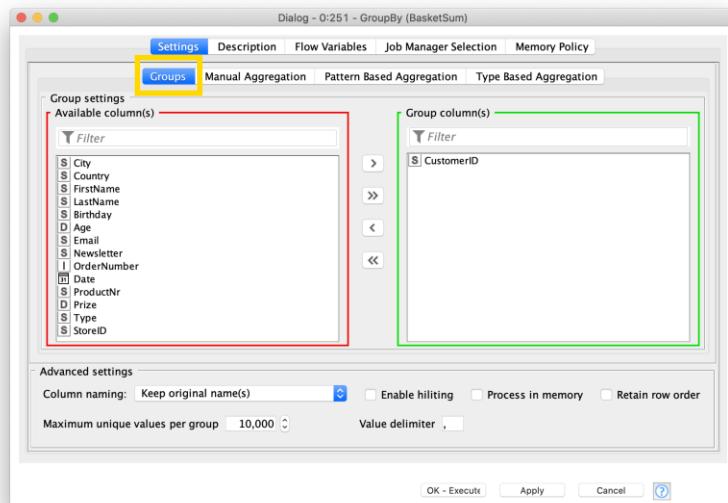
RowID	Group	Sum(Value)
r1+r3+r6	m	8
r2+r4+r5	f	15

aggregated on “group”  
by method: sum(“value”)

# GroupBy

Aggregate rows 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

Gender	Hair	Age
f	blond	31
m	red	22
f	blond	53
m	brown	16
f	brown	47
f	black	22
m	blond	13
m	red	55

Aggregation: Count

Gender	blond	brown	black	red
f	2	1	1	0
m	1	1	0	2

Aggregation: Mean(Age)

Gender	blond	brown	black	red
f	42	47	22	0
m	13	16	0	38,5

Solution: Pivoting Node

# Data Aggregation

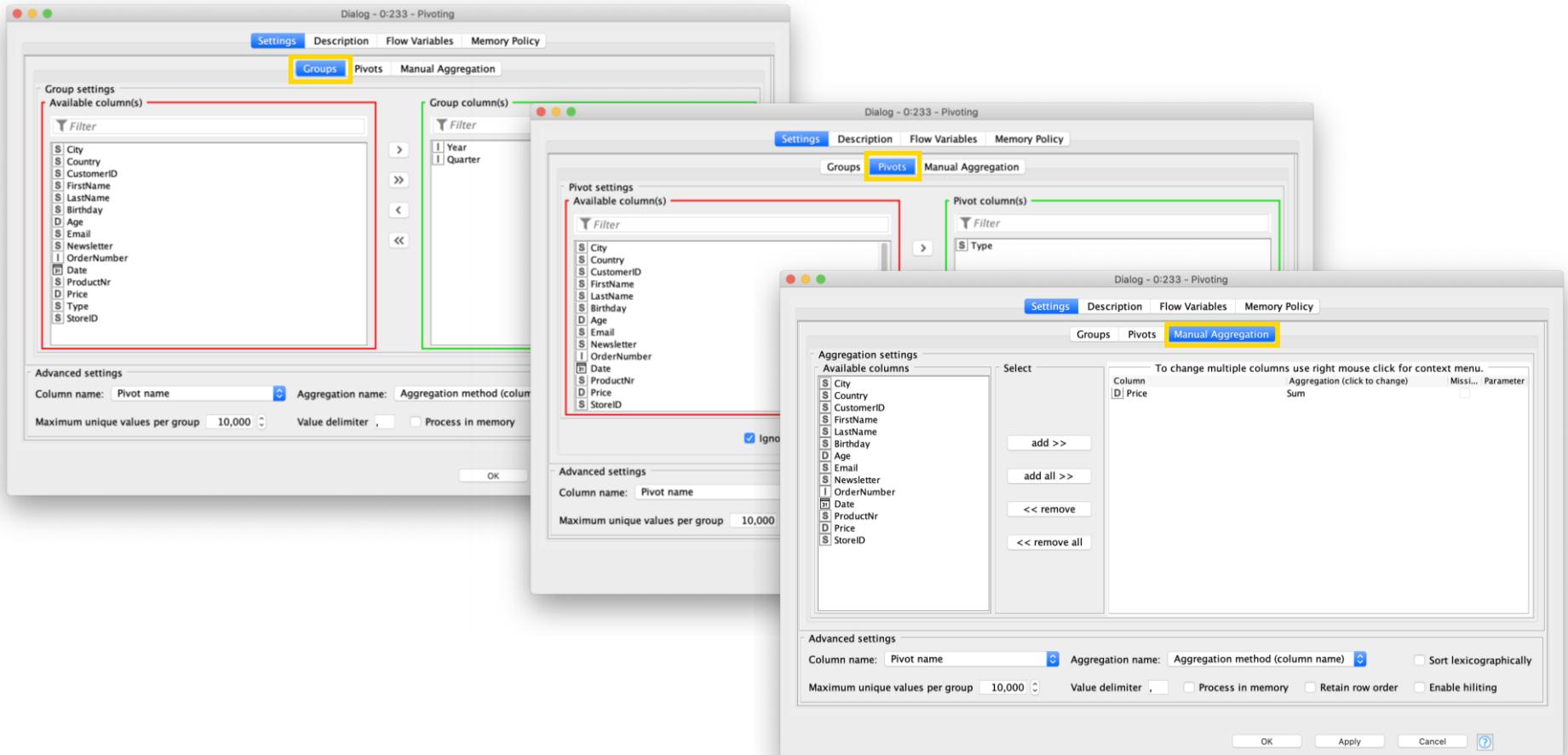
The diagram illustrates the process of data aggregation. On the left, a flat table contains three columns: Gender, Hair, and Age. The Gender column has values f and m. The Hair column has values blond, red, brown, and black. The Age column has values 31, 22, 53, 16, 47, 22, 13, and 55. A yellow arrow points from this table to the right, where the results of the aggregation are shown. The results are presented in two tables. The first table, titled "Aggregation: Mean(Age)", shows the mean age for each combination of Gender and Hair color. The second table provides the raw data used for these calculations.

Gender	blond	brown	black	red
f	42	53	22	0
m	13	16	0	38,5

Gender	Hair	Age
f	blond	31
m	red	22
f	blond	53
m	brown	16
f	brown	47
f	black	22
m	blond	13
m	red	55

Pivoting Node: Group - Pivot - Aggregate

# Pivoting

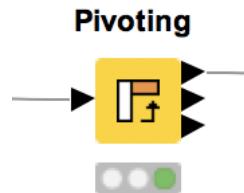


# Pivoting

---

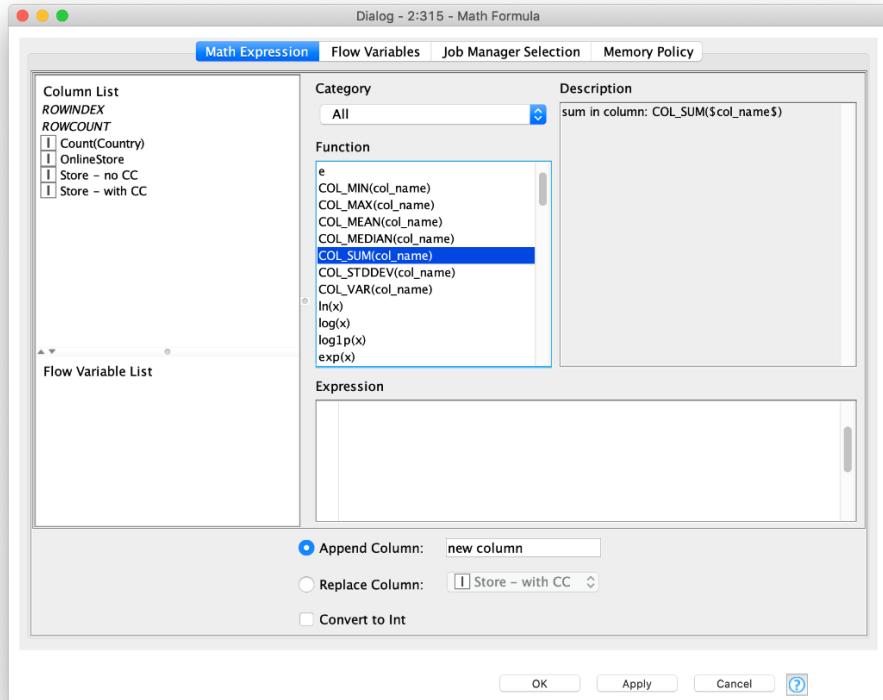
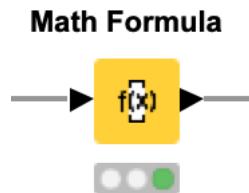
Performs pivoting on selected columns for grouping and pivoting

- Values of group columns become unique rows
- Values of the pivot columns become unique columns for each set of column combinations together with each aggregation
- Many aggregation methods are provided



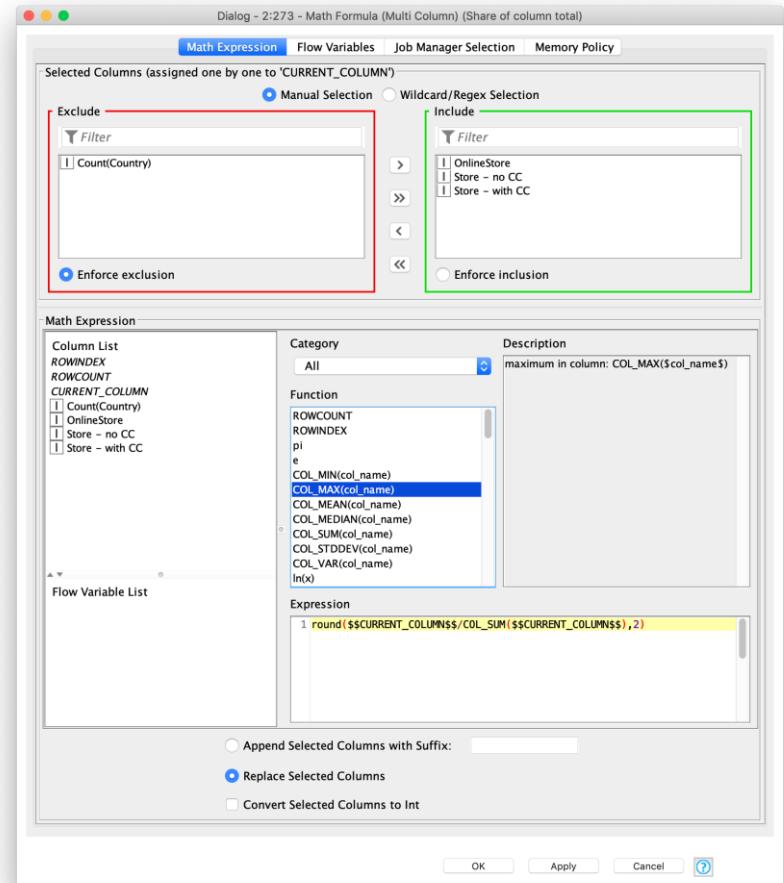
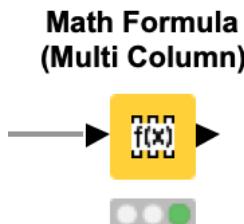
# Math Formula

- Row-wise calculations
- Some col-wise statistics
- Many mathematical functions
- Double-click function, then select col by click



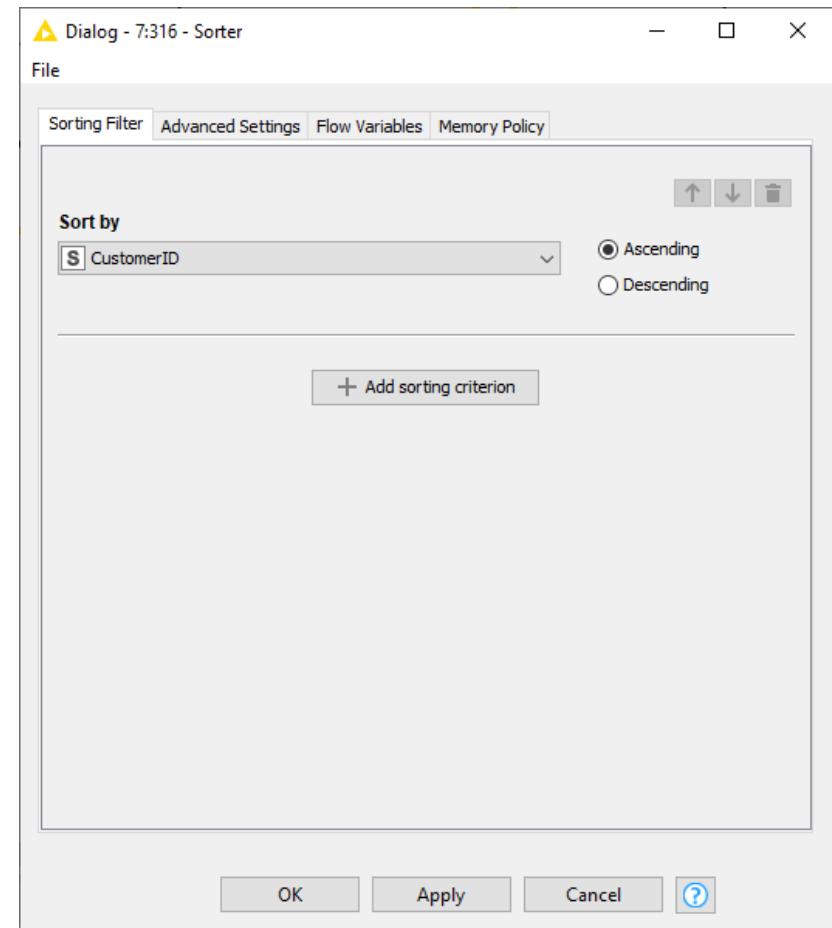
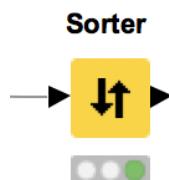
# Math Formula (Multi Column)

- Useful if you want to make the same calculations on multiple columns.
- The selected columns from the upper part are called **CURRENT\_COLUMN** in the Column List and Expression dialog.



# Sorter

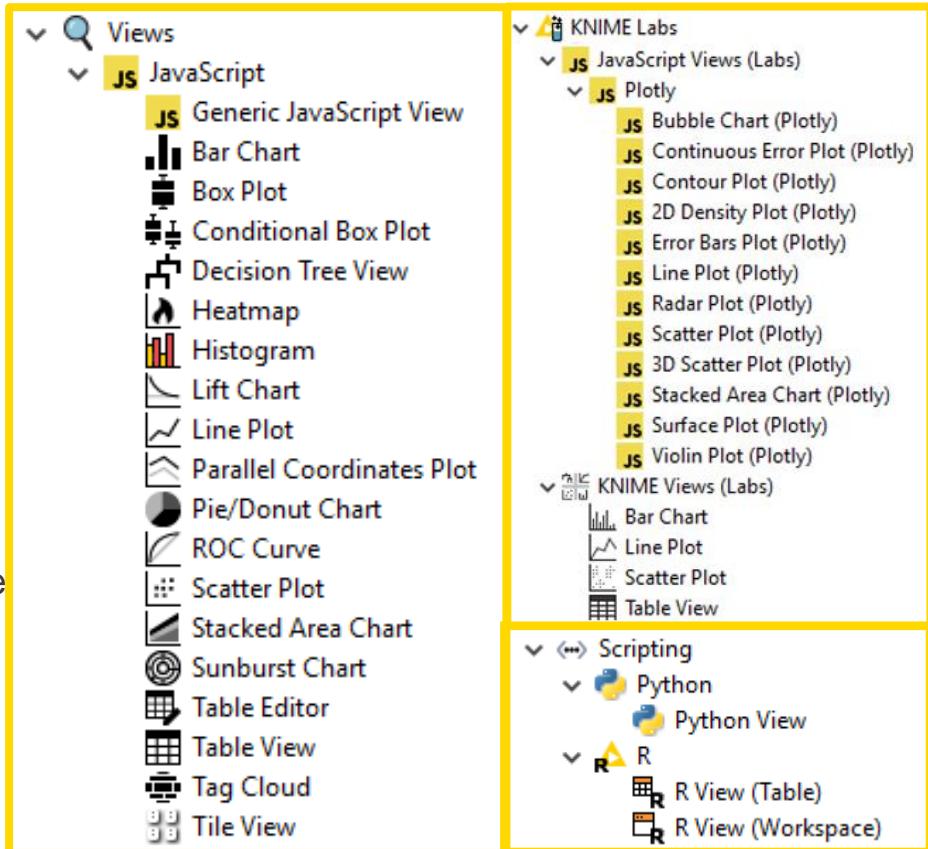
- Sorts the rows based on the values of the selected column(s), either
  - ascending or
  - descending



# Data Visualization

# 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
- New Visualization Nodes in Labs
  - A live preview of the visualization next to the configuration dialog
- R and Python View nodes for highly customizable graphics
  - Require scripting

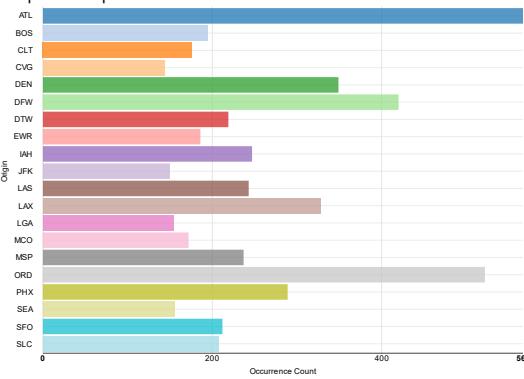


# Visualizations Using One Column

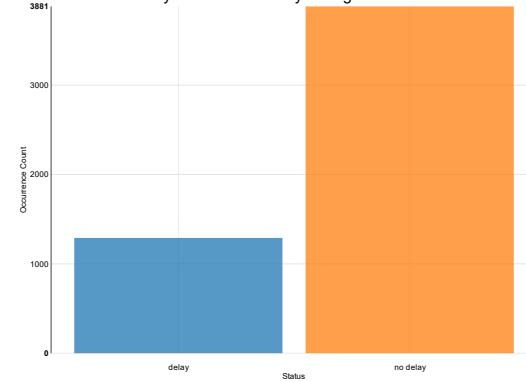
Number of Flights by Date



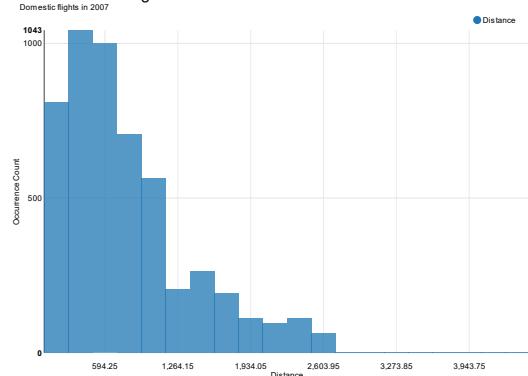
Departure Airports



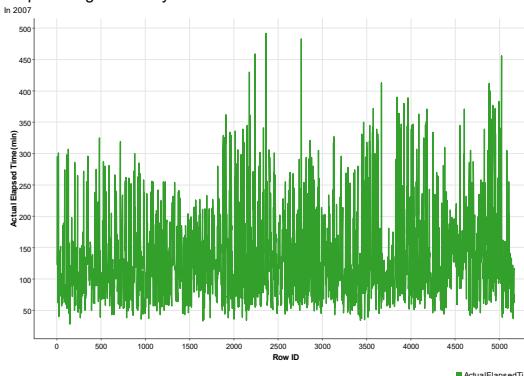
Distribution of Delayed and Non-Delayed Flights



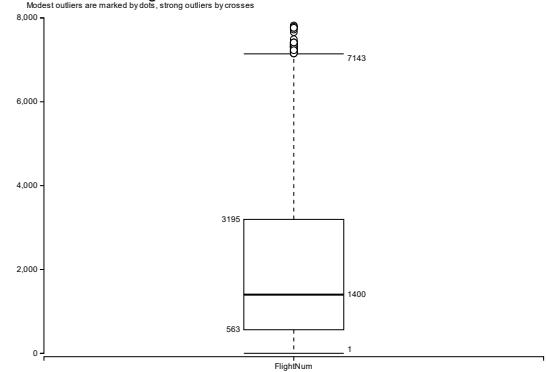
Distribution of Flight Distances



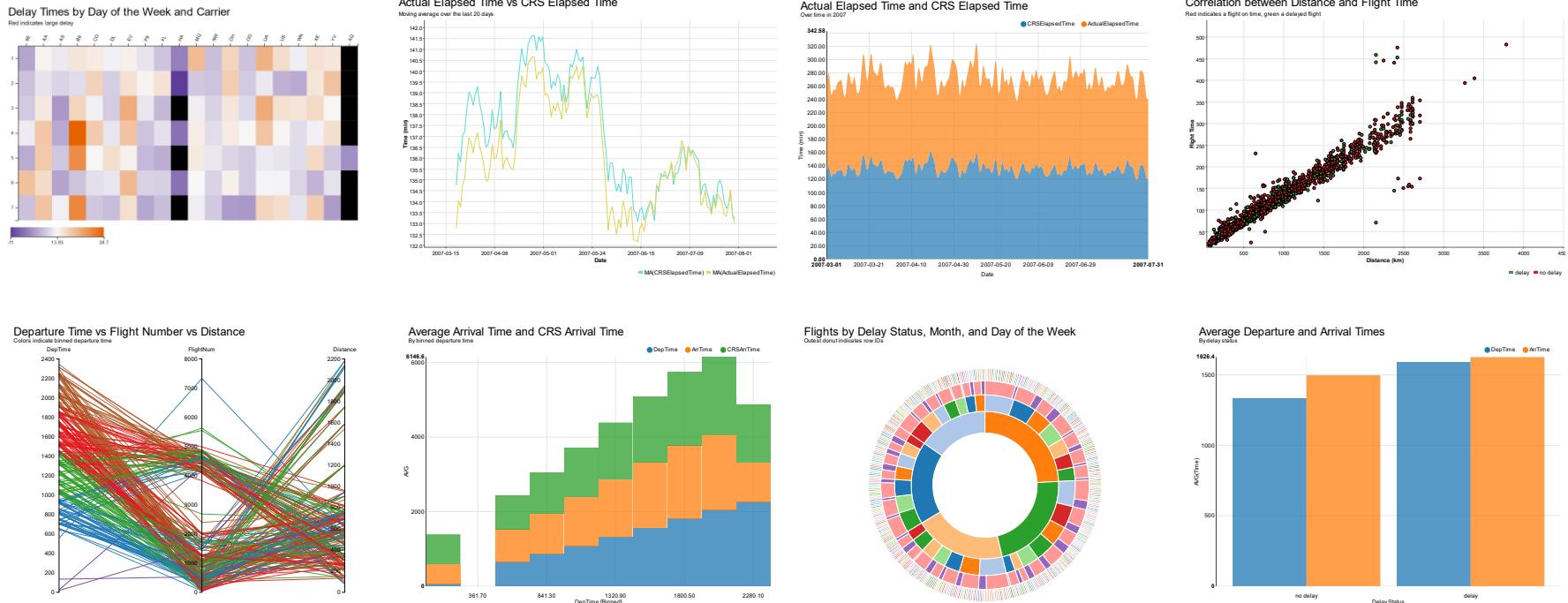
Elapsed Flight Time by Row ID



Distribution of Flight Numbers

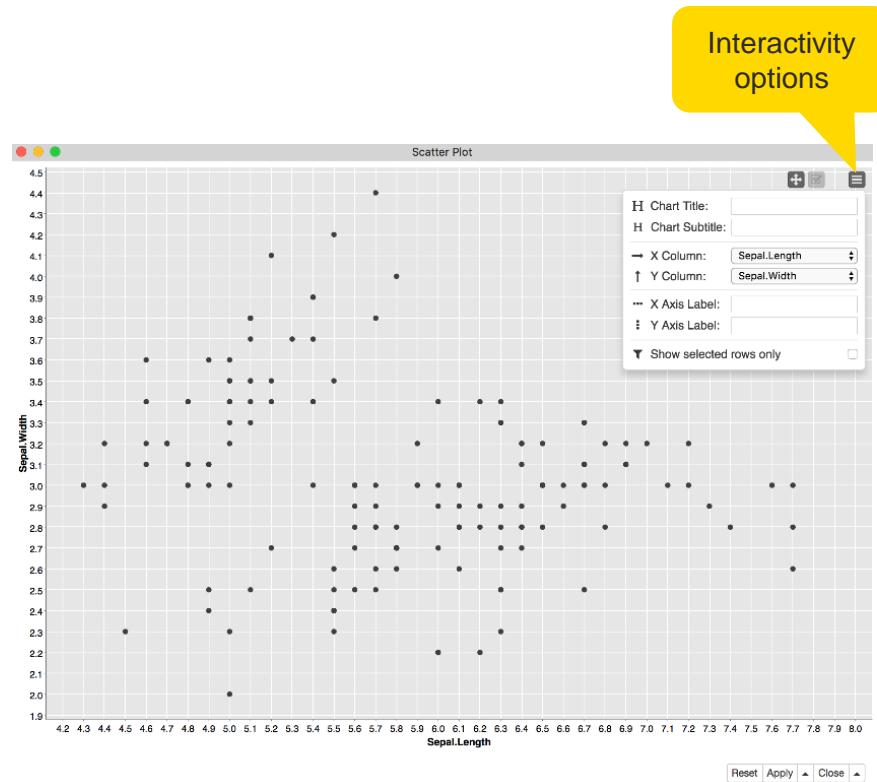
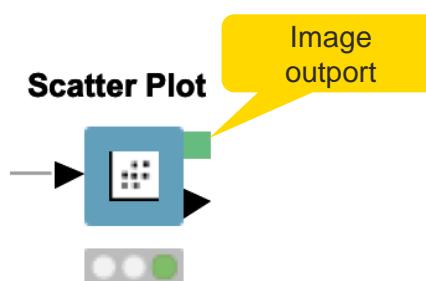


# Visualizations Using Three Columns



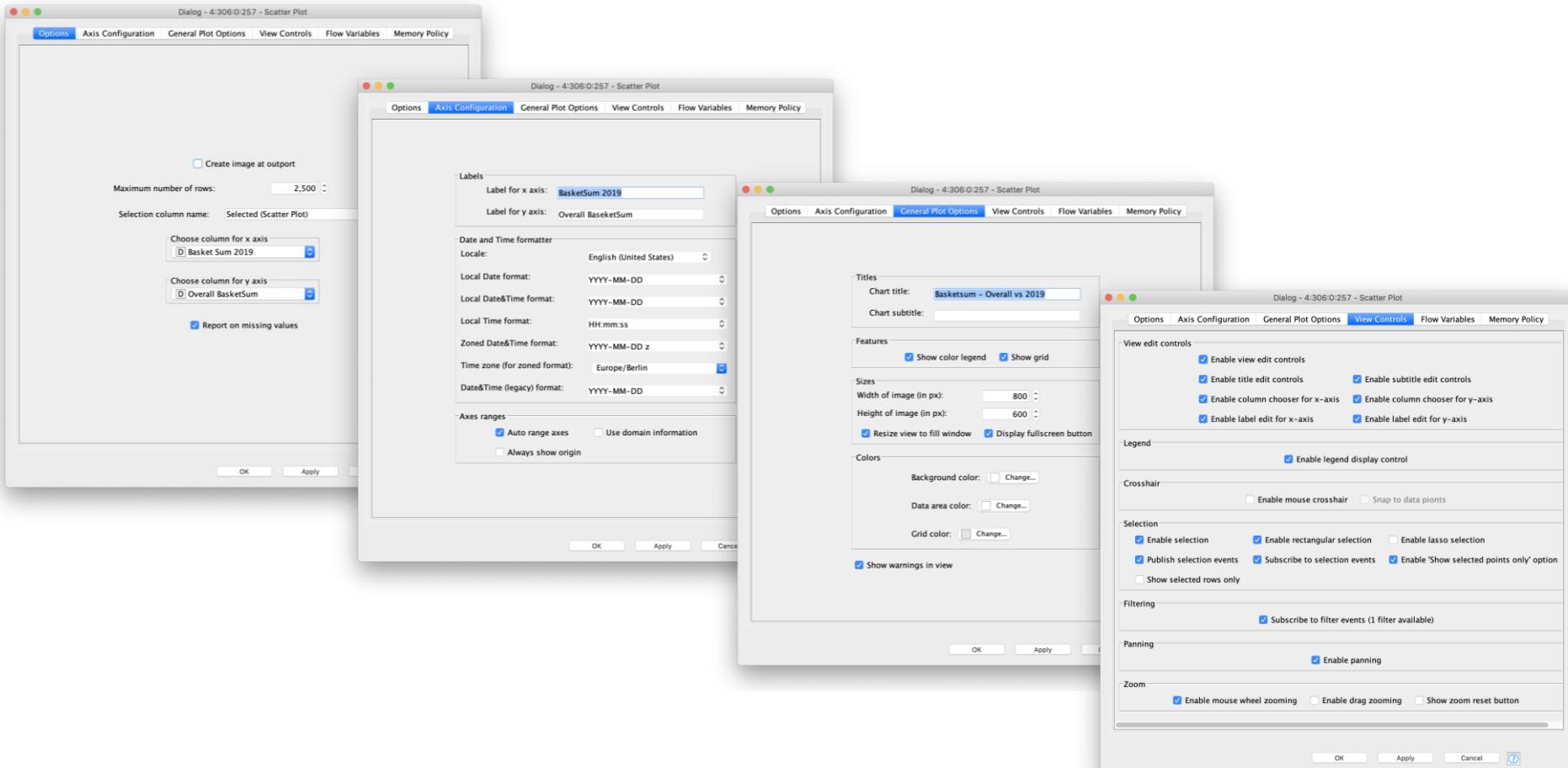
# Scatter Plot

- Plots different columns on X and Y
- Displays data including color information
- Produces an interactive view and an image



Interactivity options

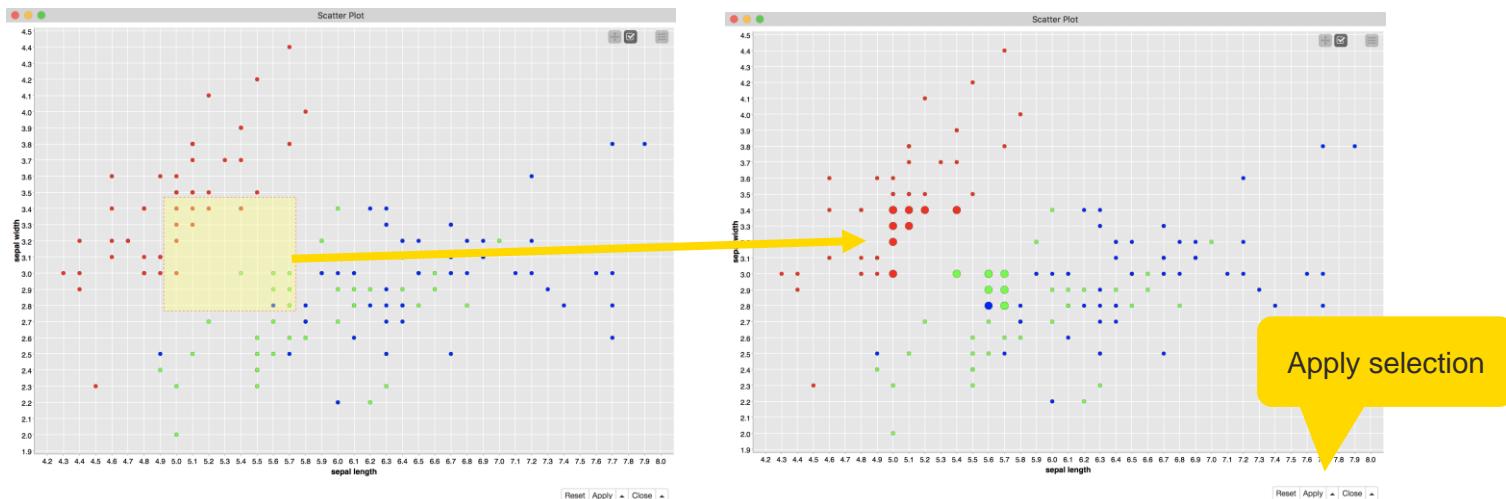
# Scatter Plot



# Selection and Filtering in JavaScript Views

Interactivity allows you to select data points in views

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



# Color Manager

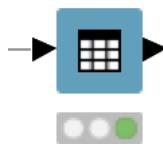
- Colors by nominal or continuous values
- Syncs colors between views using the color model port and Color Appender node

The image illustrates the KNIME Color Manager feature. On the left, the 'Dialog - 2:7 - Color Manager' window is shown. It has tabs for 'Color Settings', 'Flow Variables', 'Job Manager Selection', and 'Memory Policy'. Under 'Color Settings', a dropdown menu 'Select one Column' is set to 'D | petal\_length'. Below it, there are two radio button options: 'Nominal' and 'Range'. The 'Range' option is selected, with 'min=1.0' and 'max=6.9' displayed. A 'Preview' section shows a horizontal color bar. A yellow callout bubble points to this preview bar with the text 'Color range for numerical values'. At the bottom of the dialog are buttons for 'OK', 'Apply', 'Cancel', and a help icon. To the right of the dialog is a blue rectangular node icon labeled 'Color Manager' with a small water droplet icon. An arrow points from this icon to a scatter plot on the right. The scatter plot is titled 'Scatter Plot' and shows data points colored according to their petal length values. The x-axis is labeled 'Sepal.Length' and ranges from 4.2 to 8.0. The y-axis is labeled 'Sepal.Width' and ranges from 1.9 to 4.5. Data points are colored using a gradient corresponding to the 'Range' settings defined in the dialog.

# Table View

- Displays data in an HTML table view
- The view offers several interactive features, as well as the possibility to select rows

**Table View**



JavaScript Table View

	RowID	age	workclass	fnlwgt	education	education-num
<input checked="" type="checkbox"/>	Row0	39	State-gov	77516	Bachelors	13
<input type="checkbox"/>	Row1	50	Self-emp-not-inc	83311	Bachelors	13
<input type="checkbox"/>	Row9	42	Private	159449	Bachelors	13
<input type="checkbox"/>	Row12	23	Private	122272	Bachelors	13
<input type="checkbox"/>	Row25	56	Local-gov	216851	Bachelors	13
<input type="checkbox"/>	Row32	45	Private	386940	Bachelors	13
<input type="checkbox"/>	Row41	53	Self-emp-not-inc	88506	Bachelors	13
<input type="checkbox"/>	Row42	24	Private	172987	Bachelors	13
<input type="checkbox"/>	Row45	57	Federal-gov	337895	Bachelors	13
<input type="checkbox"/>	Row53	50	Federal-gov	251585	Bachelors	13

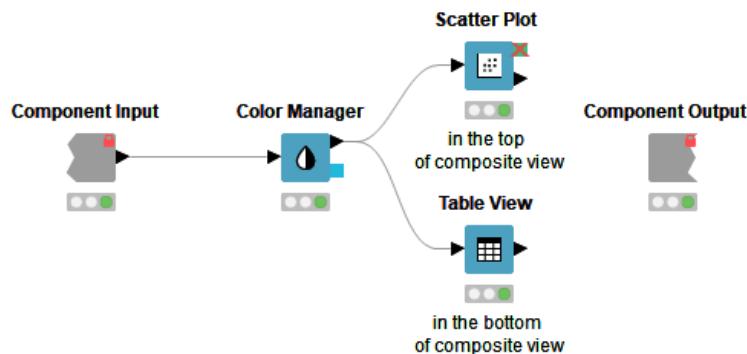
Show 10 entries Search:

Search age Search workclass Search fnlwgt  Search education-

Loading data (28710 of 29170 records) - Displaying 1 to 10 of 29170 entries.

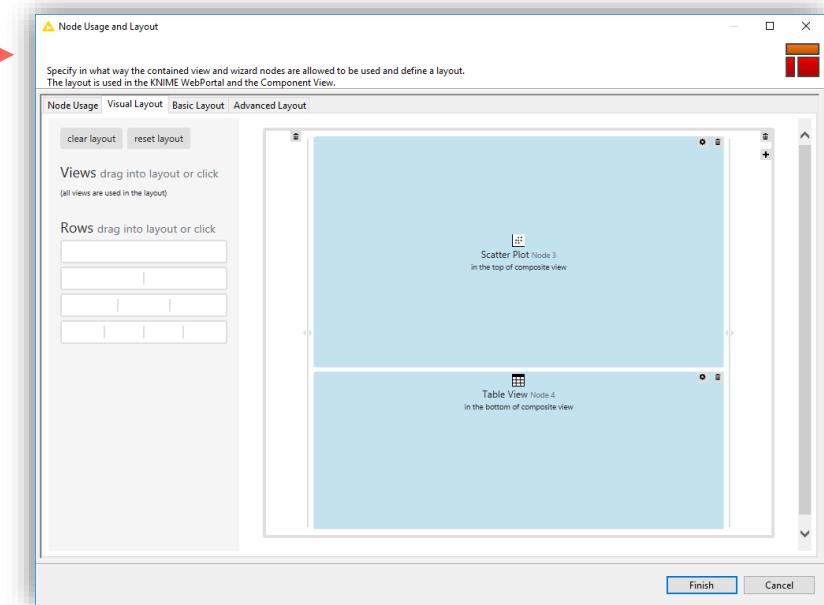
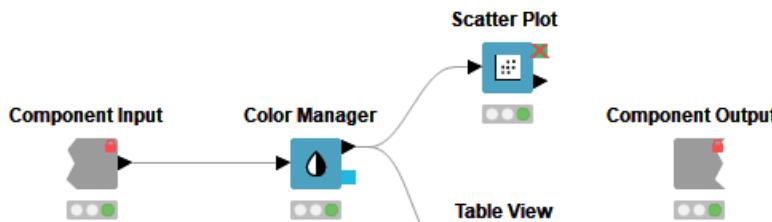
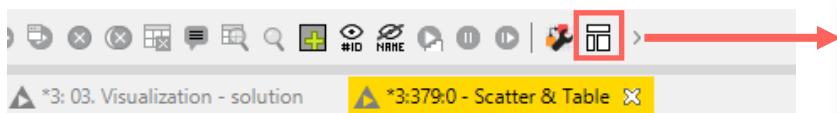
# Components – Combined Views

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



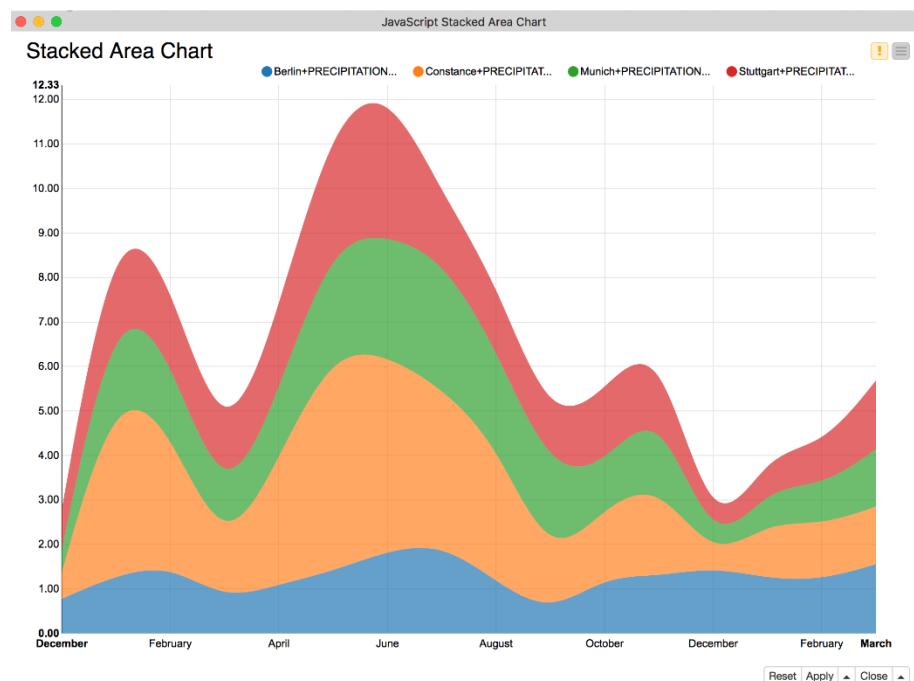
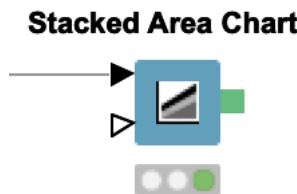
# Configure Content and Views Layout

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



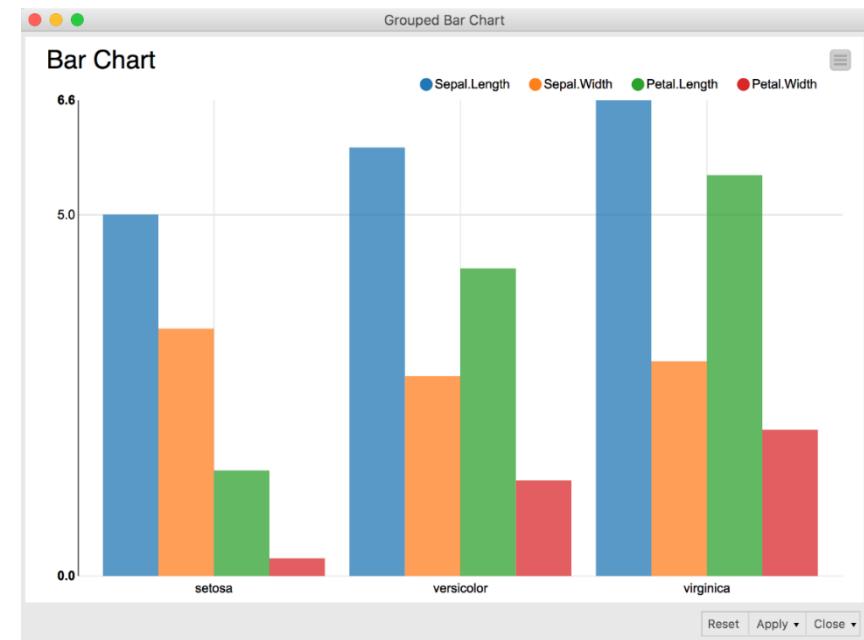
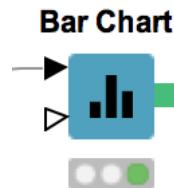
# Stacked Area Chart

- Visualizes numerical values from multiple columns as stacked areas
- Great for plotting distributions over time



# Bar Chart

- Shows numerical values across categories
- Vertical or horizontal bars
- Bars can be grouped or stacked



# The Optional Color Input Port

- Many of the visualization nodes have an optional port to change the colors
- Expects table with column headers of first table in the first column with assigned colors

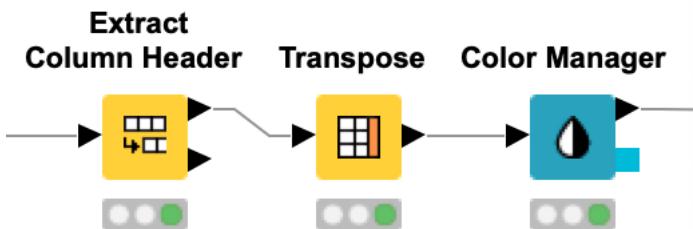
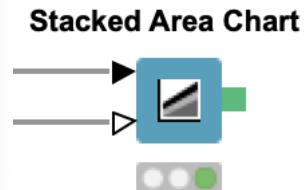


Table with Colors - 3:306:0:265 - Color M...

File Hilite Navigation View

Table "default" – Rows: 5

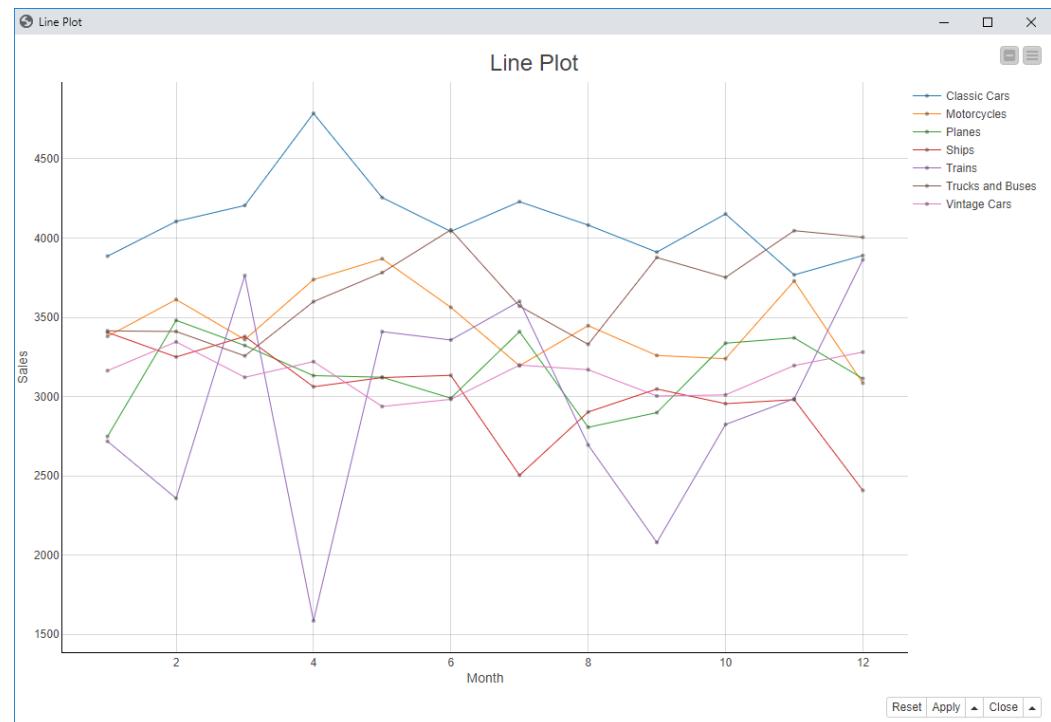
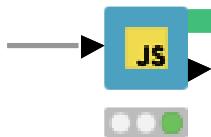
Row ID	Column Header
Column 0	Year
Column 1	Quarter
Column 2	Store – no CC
Column 3	Store – with CC
Column 4	OnlineStore



# Line Plot

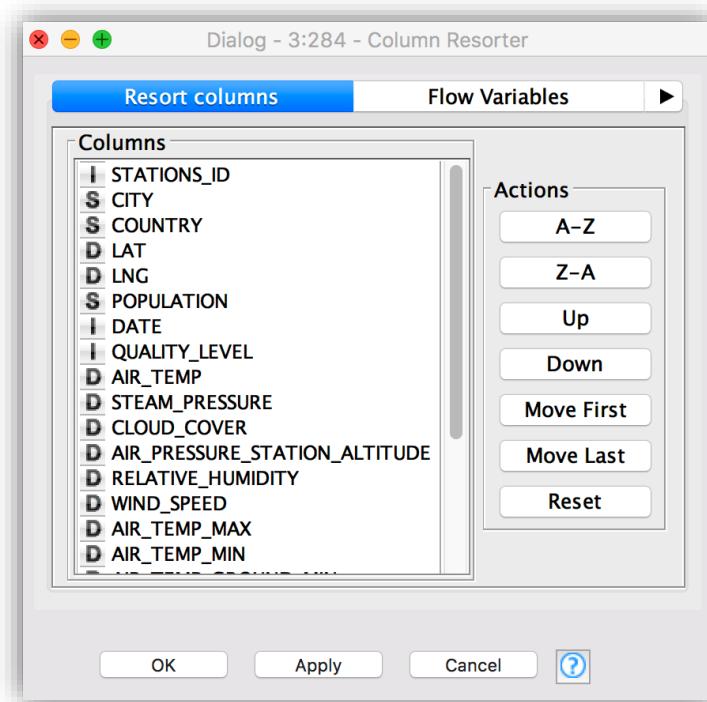
- Plots sequence of values, e.g. over time
- Useful to identify trends, also between groups

Line Plot (Plotly)

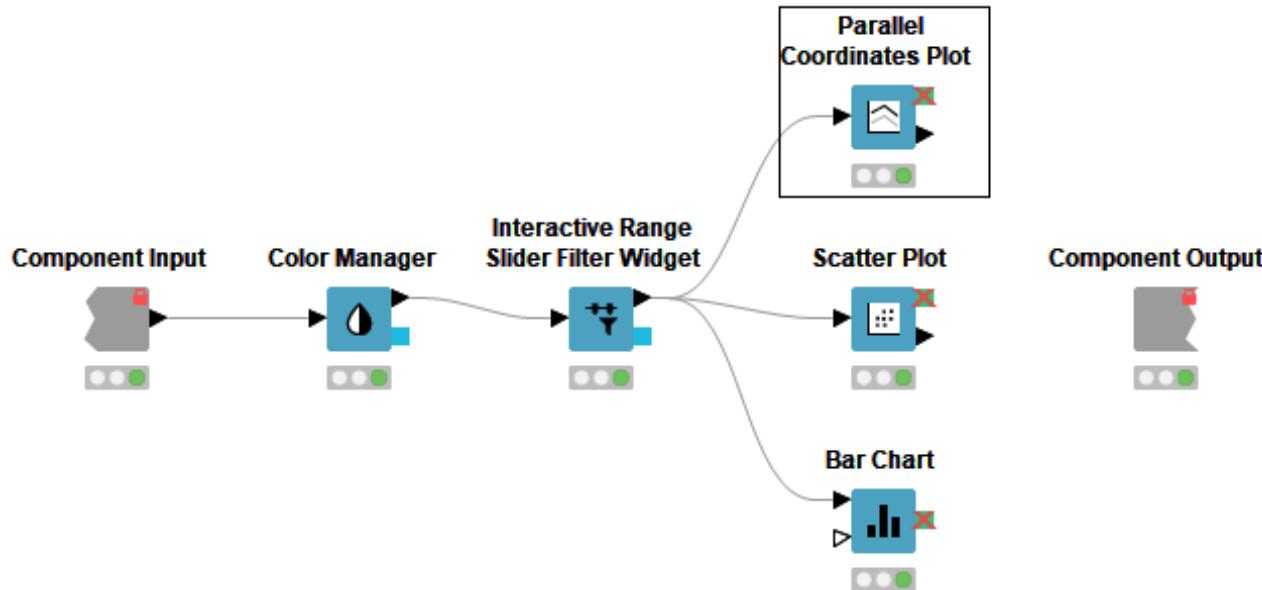


# Column Resorter

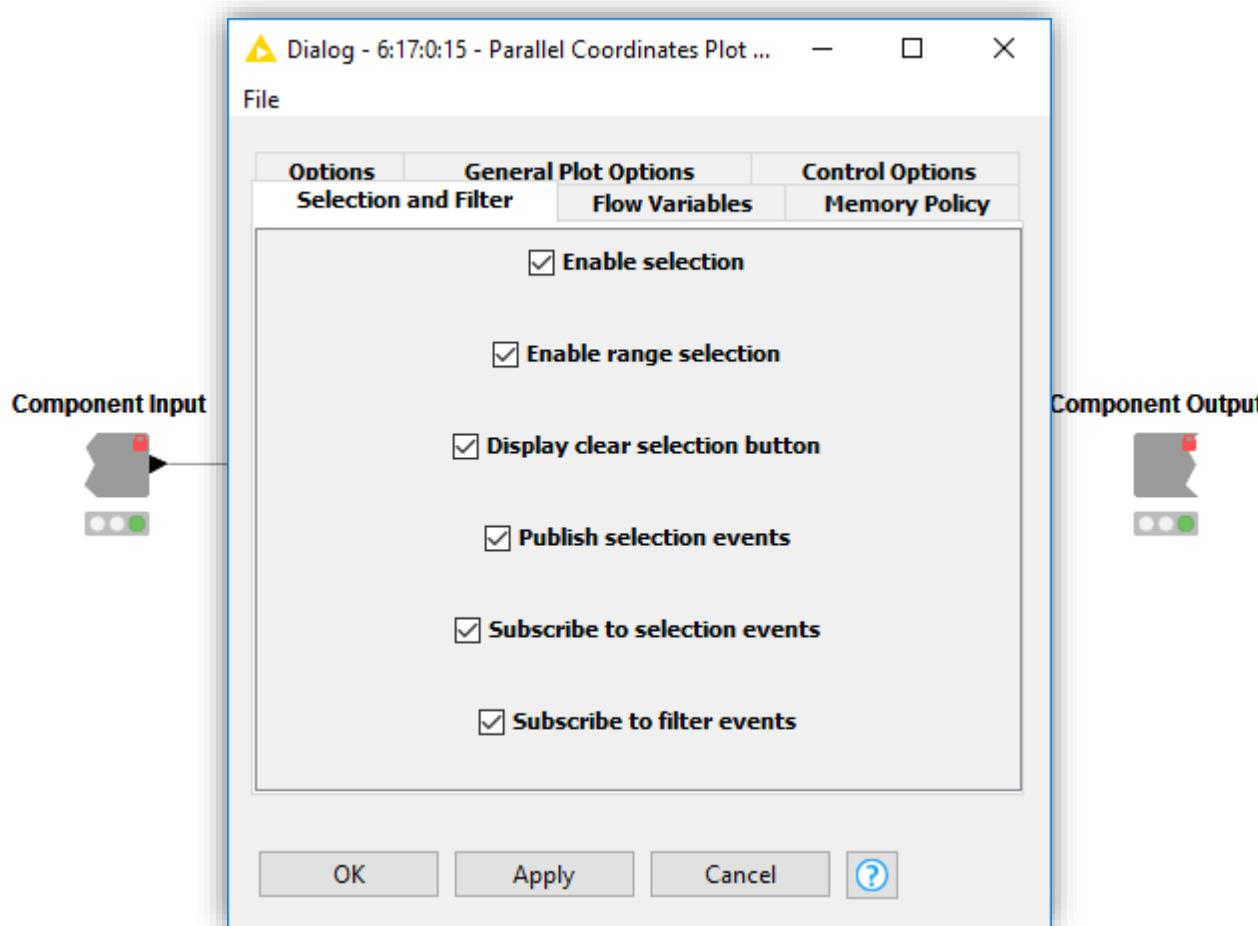
- Changes the order of the input column based on user defined settings
- Options:
  - Sort alphabetical (A-Z or Z-A)
  - Move the selected columns one step (Up or Down)
  - Move the selected columns to top or end (Move First / Last)



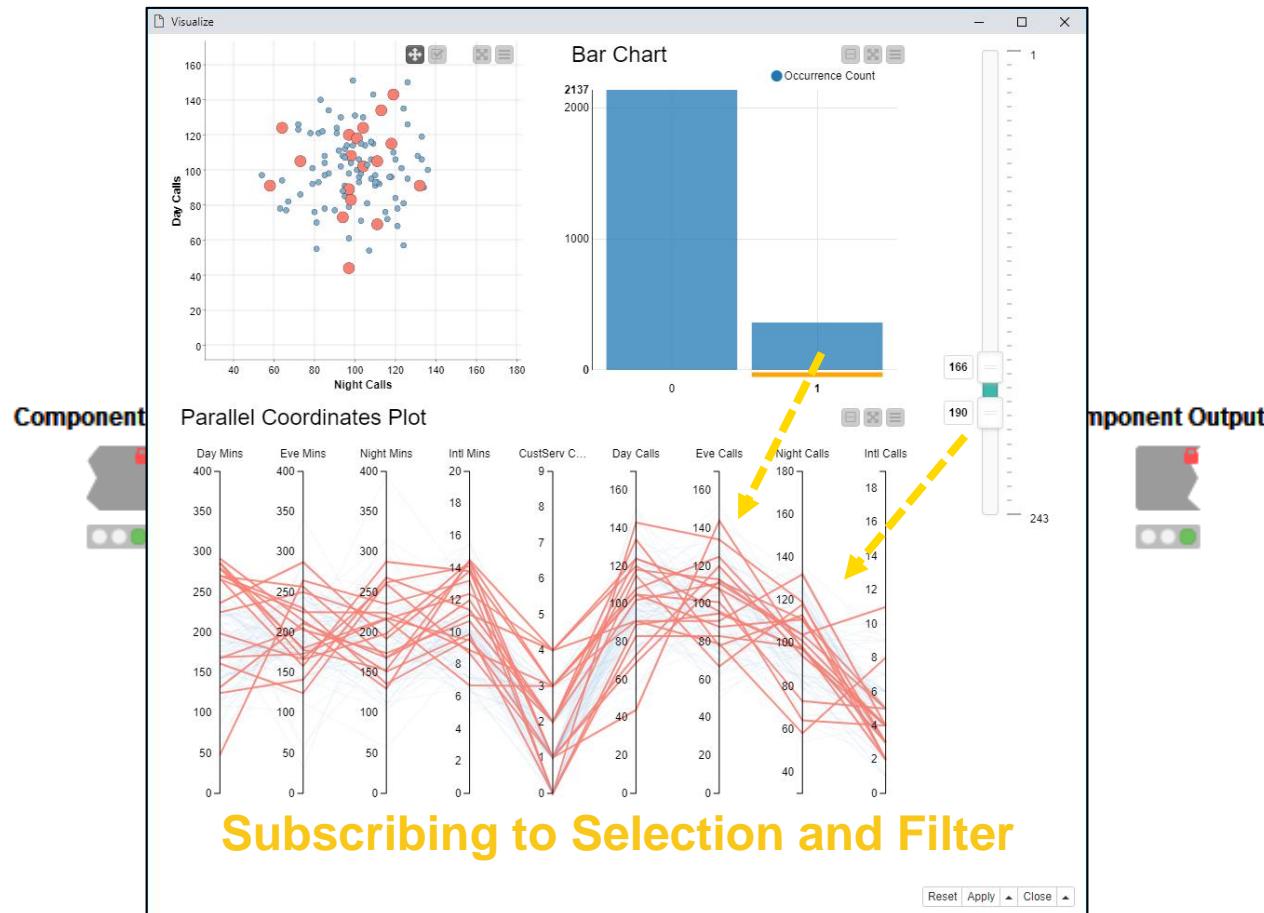
# Interactivity across Charts: Selection and Filter Events



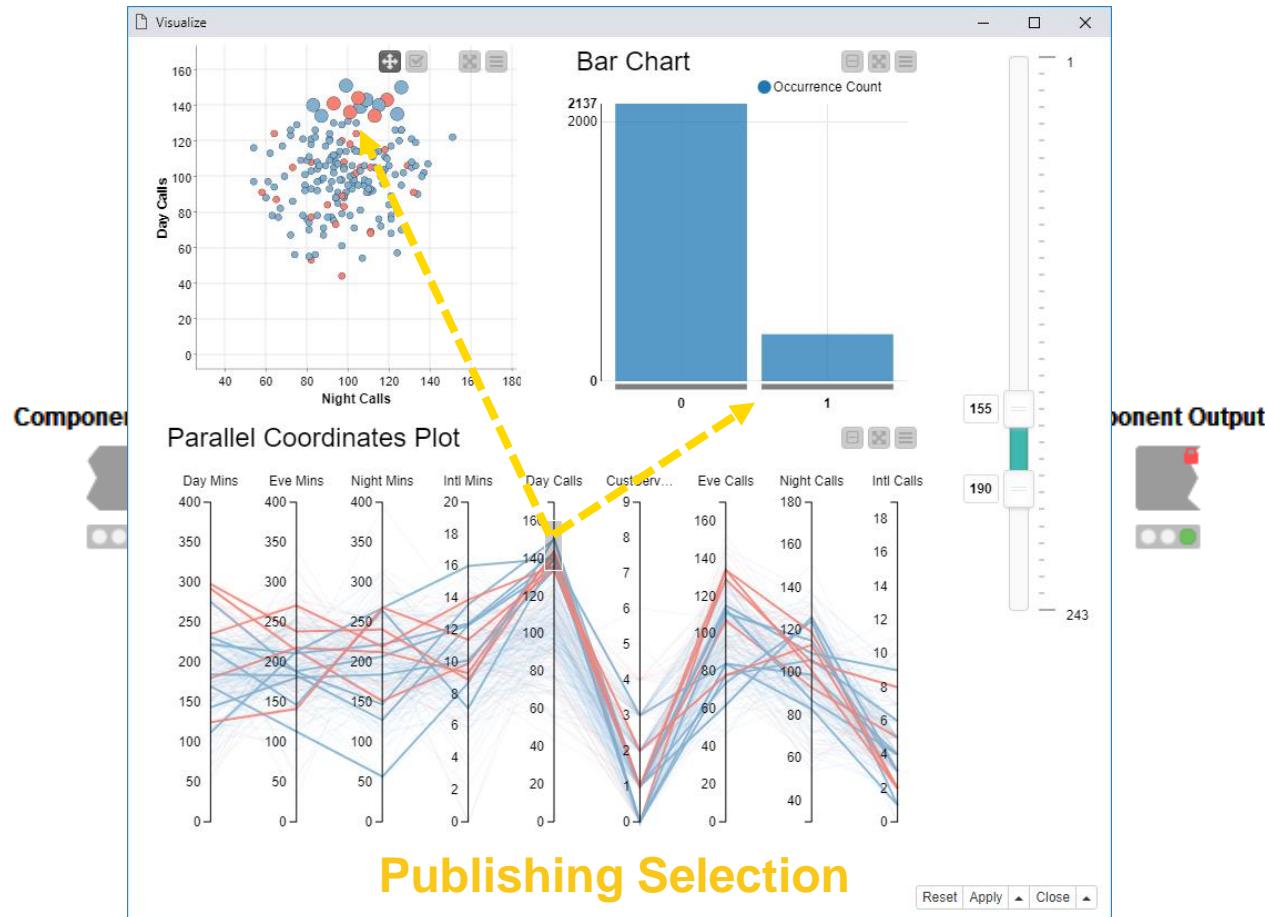
# Interactivity across Charts: Selection and Filter Events



# Interactivity across Charts: Selection and Filter Events

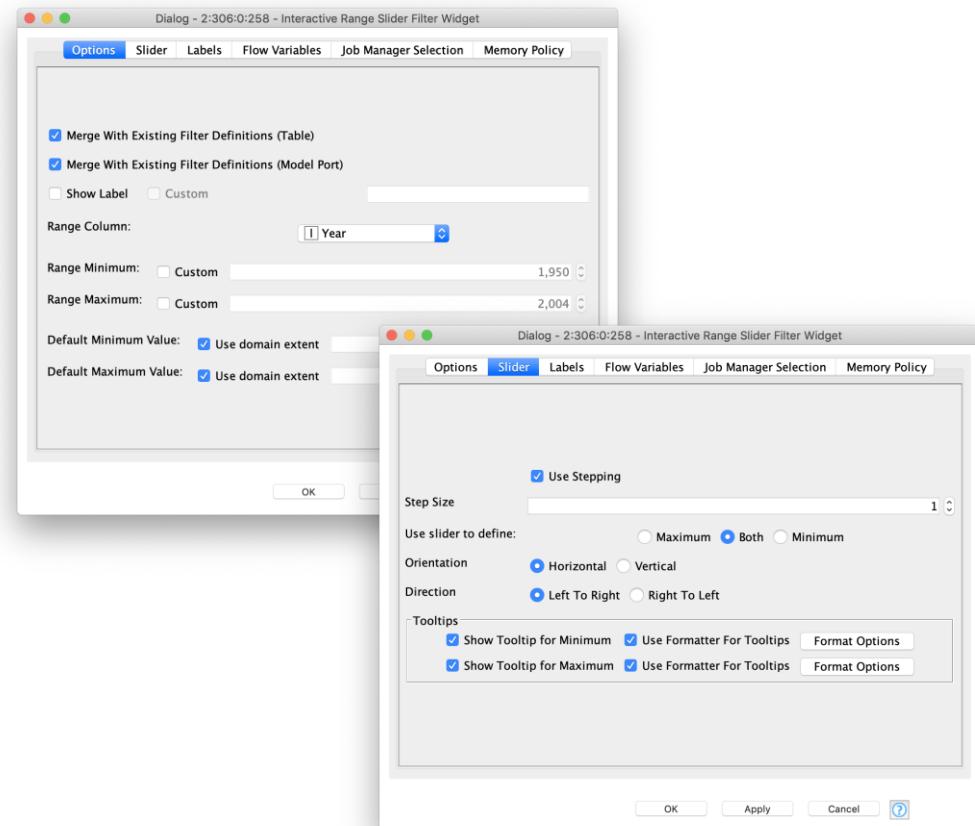
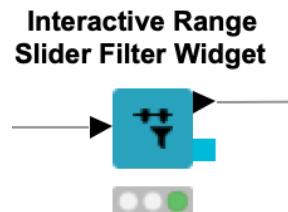


# Interactivity across Charts: Selection and Filter Events



# Interactive Range Slider Filter Widget

- Slider which can be used to trigger interactive filter events in the view of a component



# New Visualization Nodes in KNIME (Labs)

- Brand new configuration dialog (available with KNIME 4.6)
  - Explore the visualization as you change the configuration settings

The screenshot illustrates the new configuration dialog for KNIME Labs, specifically for a Bar Chart. On the left, a preview window shows a bar chart titled "Total Sales by Product as of August 3, 2022". The chart displays sales values for five categories: Fund Manager+, P+B Investment, Private Investment, Gold Investment, and CO Investment. A tooltip for the "Fund Manager+" bar indicates a value of 1,123,031. A yellow callout bubble points to this bar with the text: "The view updates as you change the configuration settings >>". On the right, the configuration dialog is open, showing various settings sections:

- Data**: Category dimension set to "Products".
  - Aggregate**: Sum (radio button selected).
  - Frequency dimensions**: Excludes "CustomerKey" and "BasketSize"; Includes "BasketValue".
- Plot**: Title set to "Total Sales by Product as of August 3, 2022".
  - Orientation**: Vertical (radio button selected).
  - Arrange bars**: Grouped (radio button selected).
  - Interactivity**: Enable image download (checkbox checked), Show tooltip (checkbox checked).
- OK** and **Cancel** buttons at the bottom right.

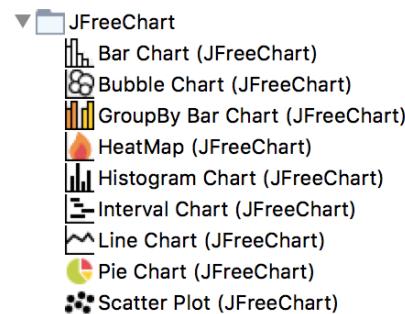
A second yellow callout bubble points to the "Title" field in the Plot section with the text: "Settings controlled by variables".

At the top right, a sidebar shows the KNIME Labs navigation tree: KNIME Labs (selected) and KNIME Views (Labs) expanded to show Bar Chart, Line Plot, Scatter Plot, and Table View.

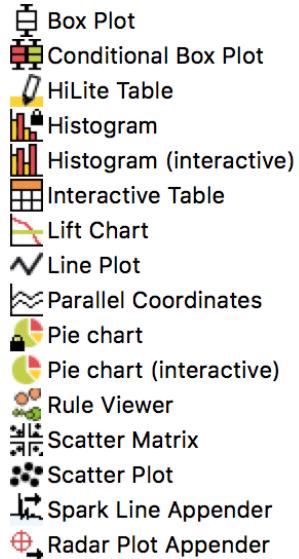
Category	Value
Fund Manager+	1,123,031
P+B Investment	~850,000
Private Investment	~2,200,000
Gold Investment	~1,500,000
CO Investment	~200,000

# Legacy View Nodes: JFreeChart & KNIME Views

- KNIME provides three types of visualizations
  - **JavaScript Views**
  - JFreeChart
  - KNIME Views
- Active development only for JavaScript Views -> use those!
- JFreeChart and KNIME Views still useful until all plot types are implemented in JS (we're on it)



JFreeChart



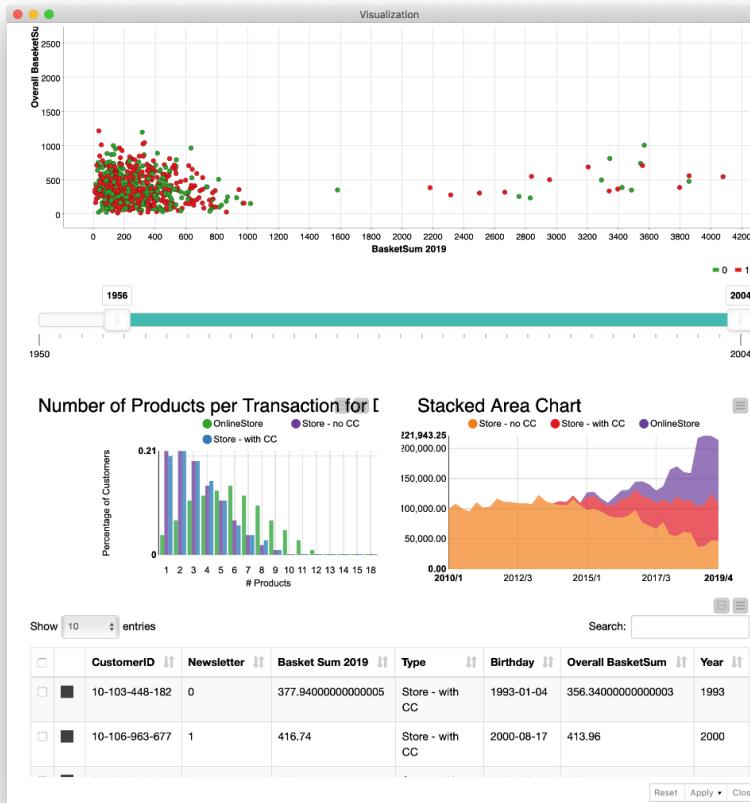
KNIME Views

# Exercise: 06\_Data\_Aggregation

---

- Calculate the total purchase amount by a customer ID both in 2019 and earlier
- Calculate the total purchase amount by quarter and transaction type
- Calculate the numbers of orders by basket size and transaction type (optional)
- Convert the dates of births of the customers to Date&Time and extract the birth year into a separate column (optional)

# Exercise: 07\_Visualization – Goal

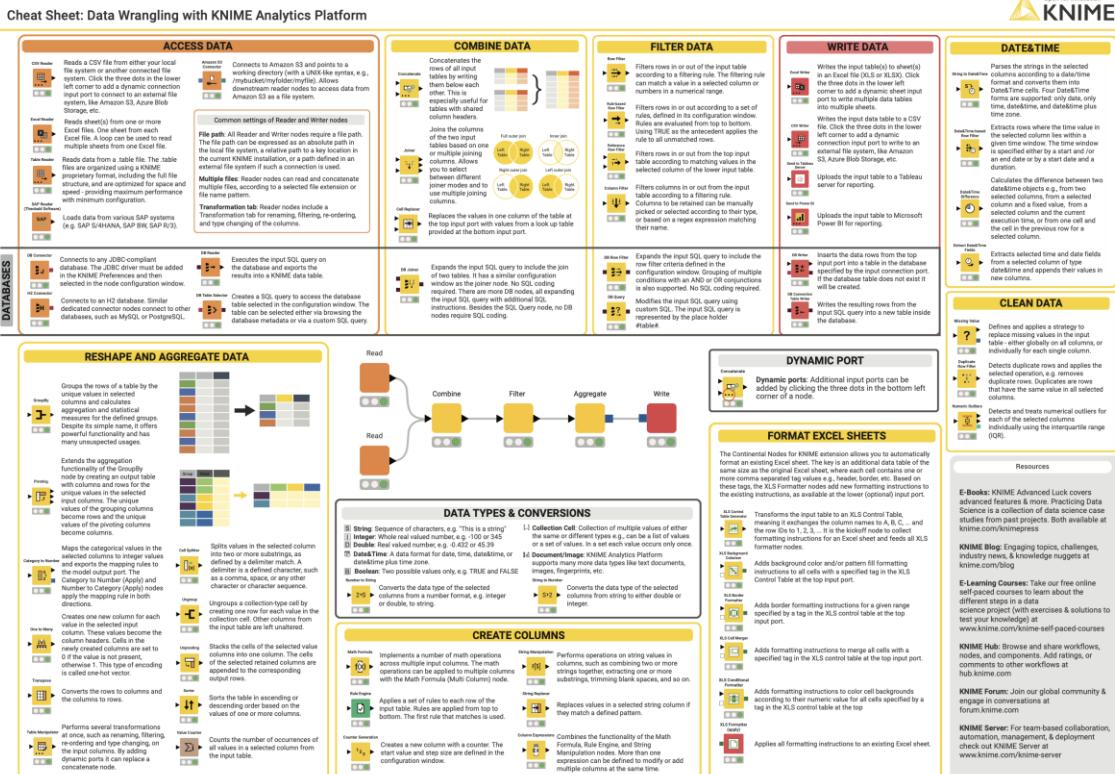


# Exercise: 07\_Visualization

---

- Create a scatter plot to show the relationship between the total purchase amount in the year 2019 and in the years before
- Visualize the customer data in an interactive table
- Create a stacked area chart to show the development of the total purchase amount over time for each transaction type
- Create a composite view and define the layout
  
- Optional tasks:
  - Add a range slider to filter the scatter plot by age (optional)
  - Build a bar chart to show the number of products per order for the different transaction types (optional)
  - Change the color for the different transaction types (optional)

# Data Wrangler Cheat Sheet



<https://www.knime.com/sites/default/files/2021-07/cheat-sheet-data-wrangling.pdf>

# **Attachment: How to use a local update site to install extensions**

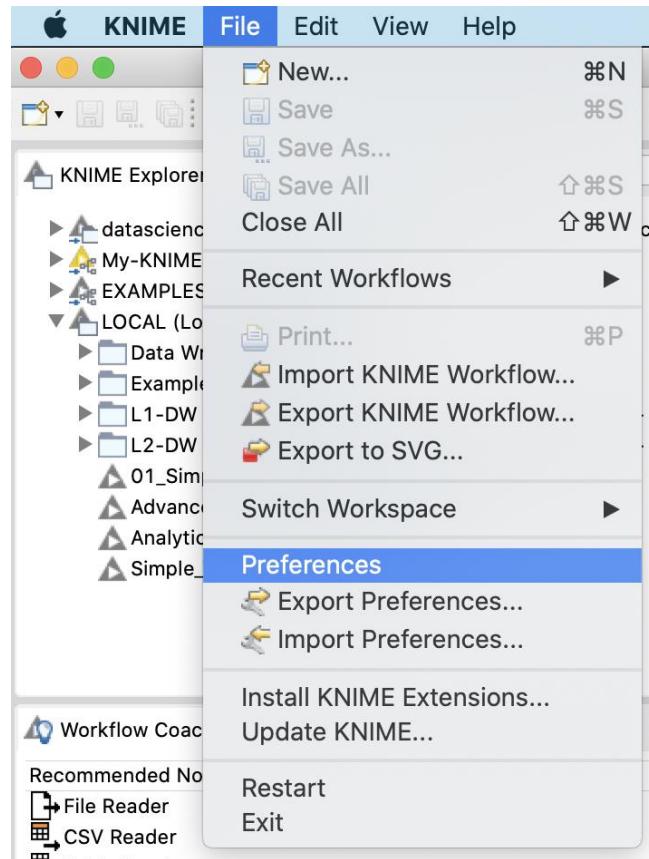
# Adding a Local Update Site

---

- Download the update site as zip
  - [KNIME update](#) site as zip
  - [Previous versions](#) of the KNIME update site as zip
  - [Community update](#) sites as zip

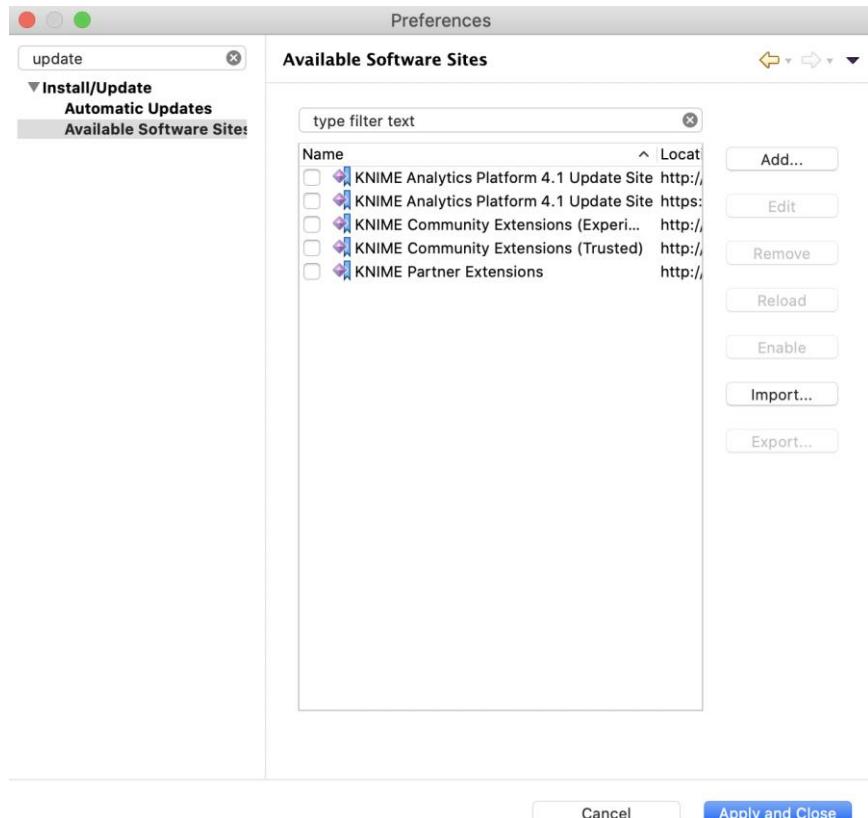
# Adding a Local Update Site

- Open KNIME Analytics Platform and go to the preference page by clicking on
- File -> Preferences



# Adding a Local Update Site

1. Search for update (upper left search bar) and go to Available Software sites.
2. Uncheck all existing software sites.
3. Click on Add.. on the upper right.



# Adding a Local Update Site

1. Define a name
2. Click on Archive and select the folder you've just downloaded
3. Click OK
4. Click Apply and Close

