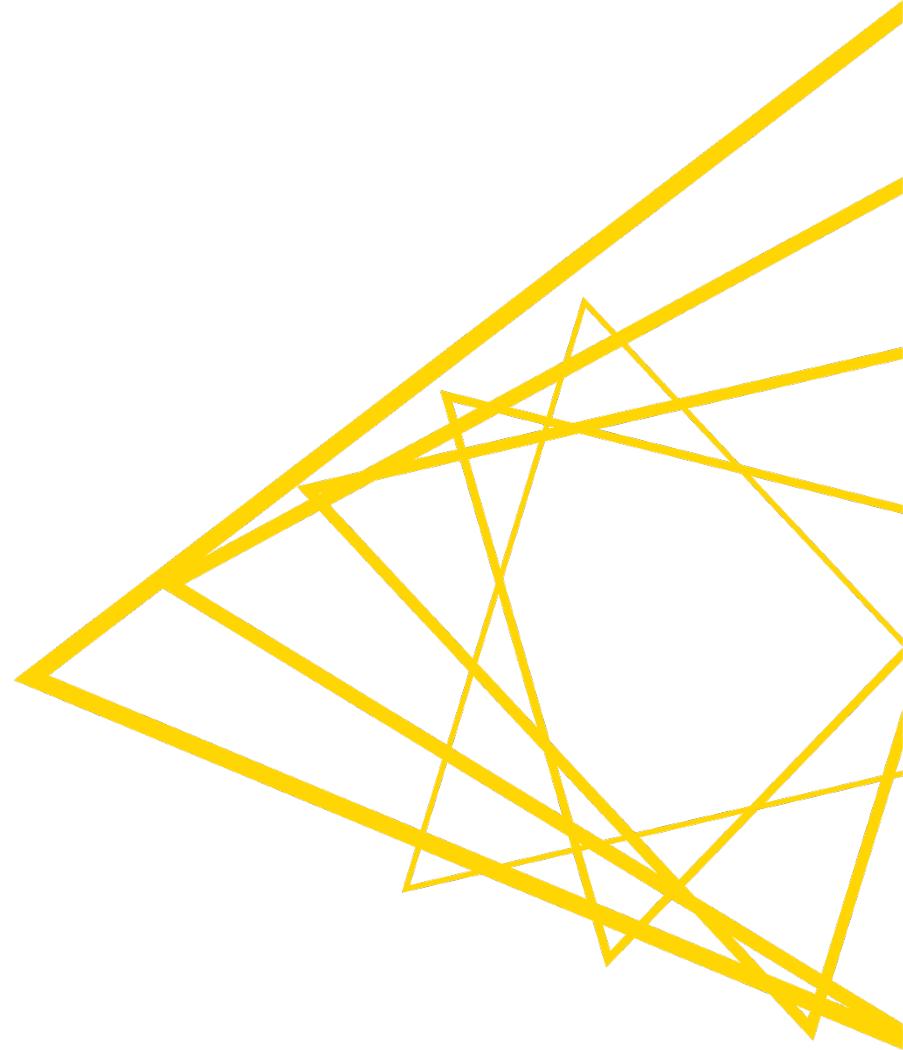
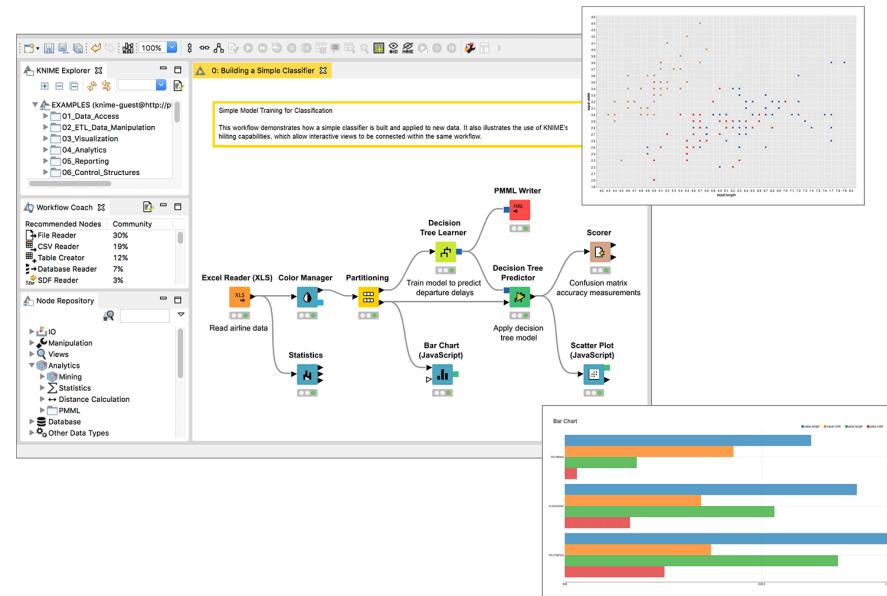


# Overview KNIME Analytics Platform



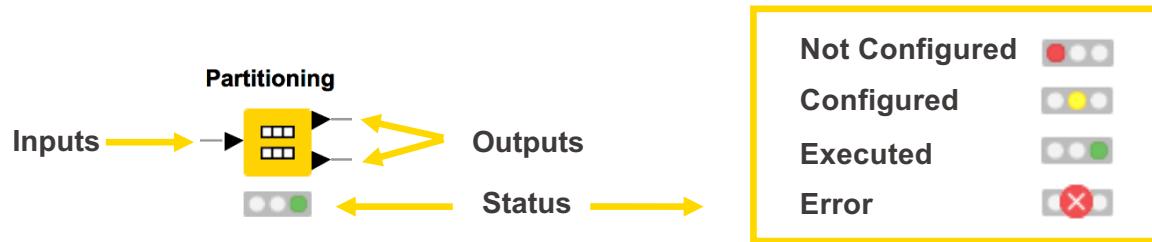
# What is KNIME Analytics Platform?

- 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.

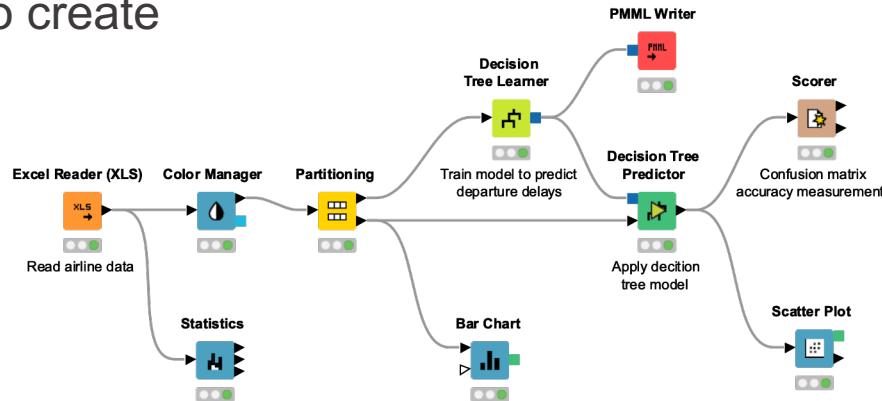


# Visual KNIME Workflows

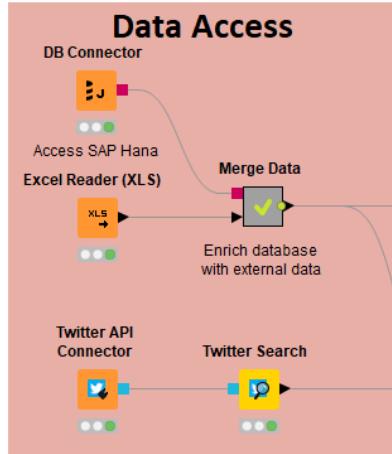
**NODES** perform tasks on data



Nodes are combined to create  
**WORKFLOWS**



# Data Access



## Databases

- MySQL, PostgreSQL, Oracle
- Theobald
- any JDBC (DB2, MS SQL Server)
- Amazon DynamoDB

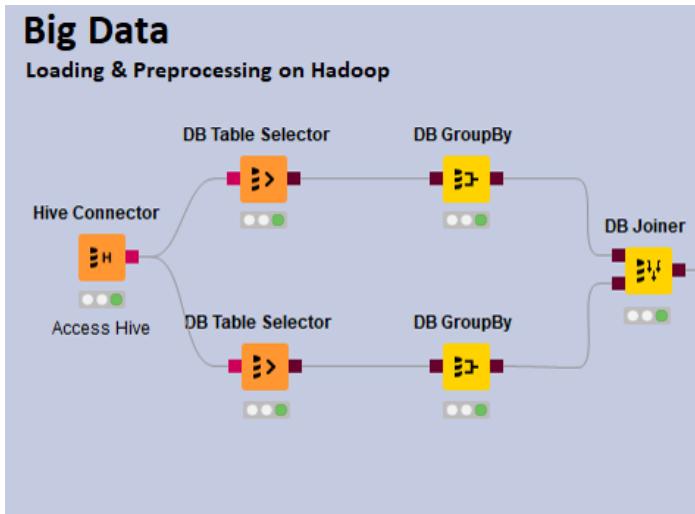
## Files

- 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

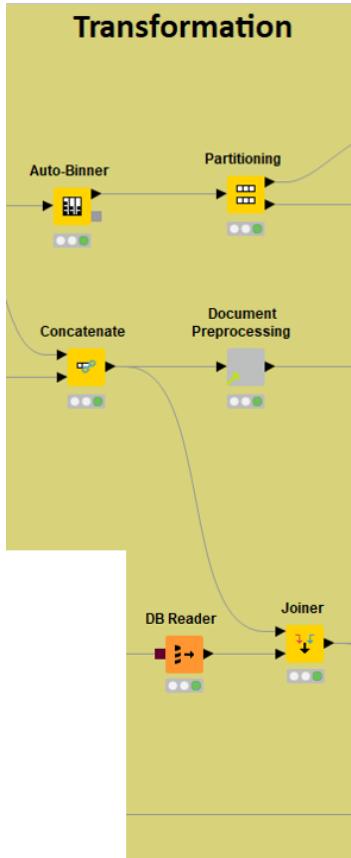
# Big Data



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

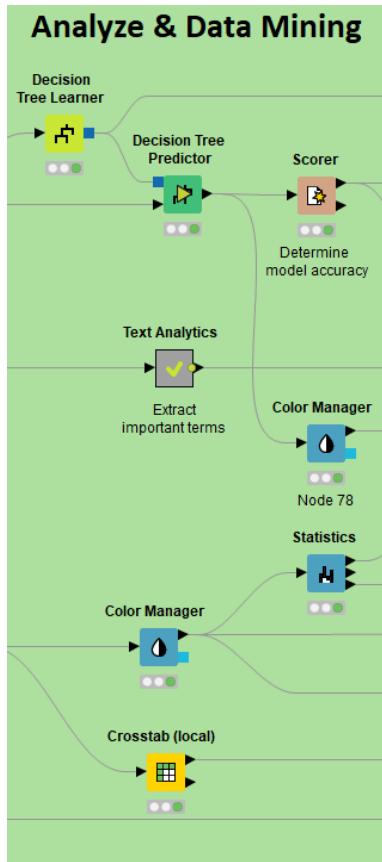


# Transformation



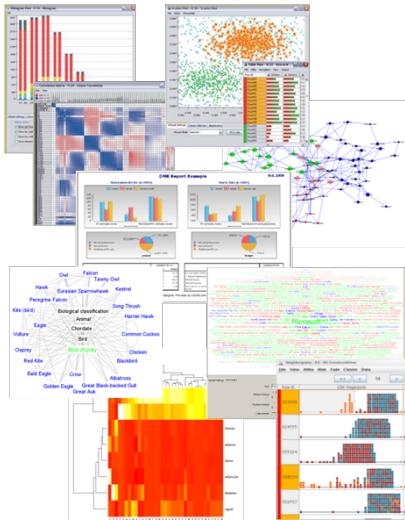
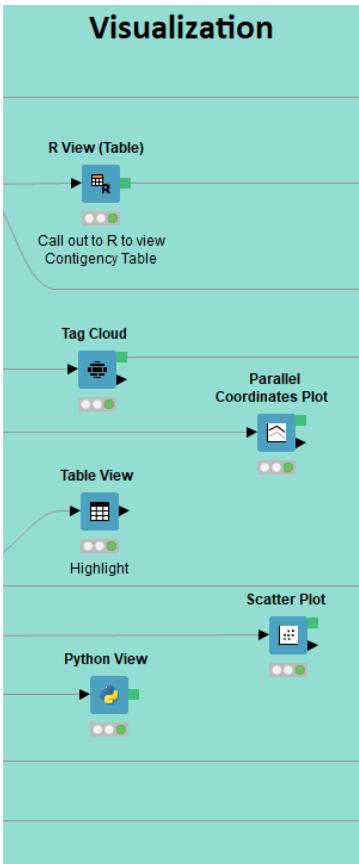
- Preprocessing
  - Row, column, matrix based
- Data blending
  - Join, concatenate, append
- Aggregation
  - Grouping, pivoting, binning
- Feature Creation and Selection

# Analysis & Data Mining



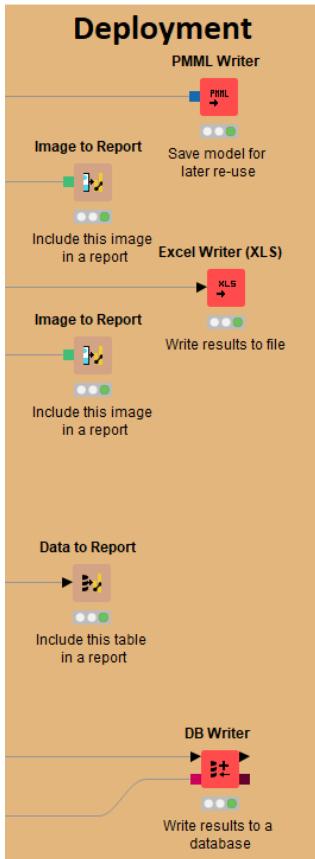
- Regression
  - Linear, logistic
- Classification
  - Decision tree, ensembles, SVM, MLP, Naïve Bayes
- Clustering
  - k-means, DBSCAN, hierarchical
- Validation
  - Cross-validation, scoring, ROC
- Deep Learning
  - Keras, DL4J
- External
  - R, Python, Weka, H2O, Keras

# Visualization



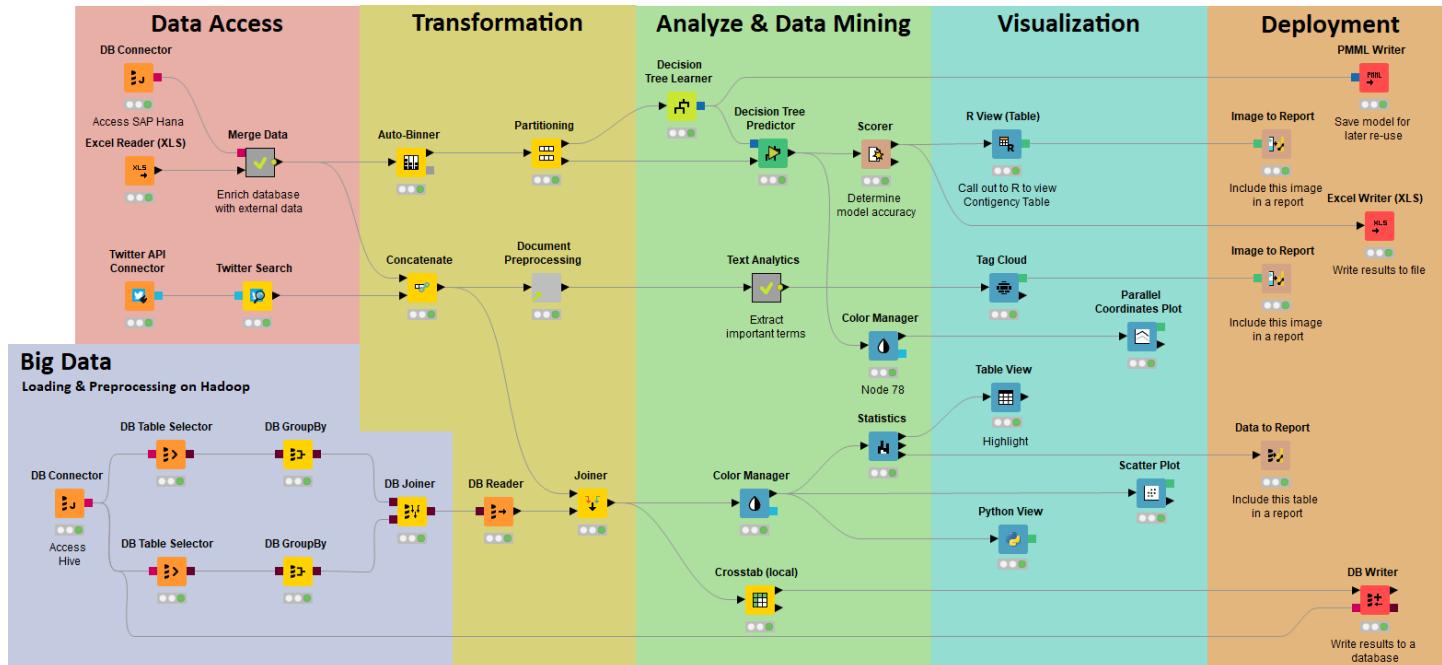
- Interactive Visualizations
- JavaScript-based nodes
  - Scatter Plot, Box Plot, Line Plot
  - Networks, ROC Curve, Decision Tree
  - Plotly Integration
  - Adding more with each release!
- Misc
  - Tag cloud, open street map, molecules
- Script-based visualizations
  - R, Python

# Deployment



- Database
- Files
  - Excel, CSV, txt
  - XML
  - PMML
  - to: local, KNIME Server, Amazon S3, Azure Blob Store
- BIRT Reporting

# Over 2000 Native and Embedded Nodes Included:



## Data Access

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

## Transformation

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

## Analysis & Mining

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

## Visualization

R  
JFreeChart  
JavaScript  
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 archive and extract the file, or download installer package and run it

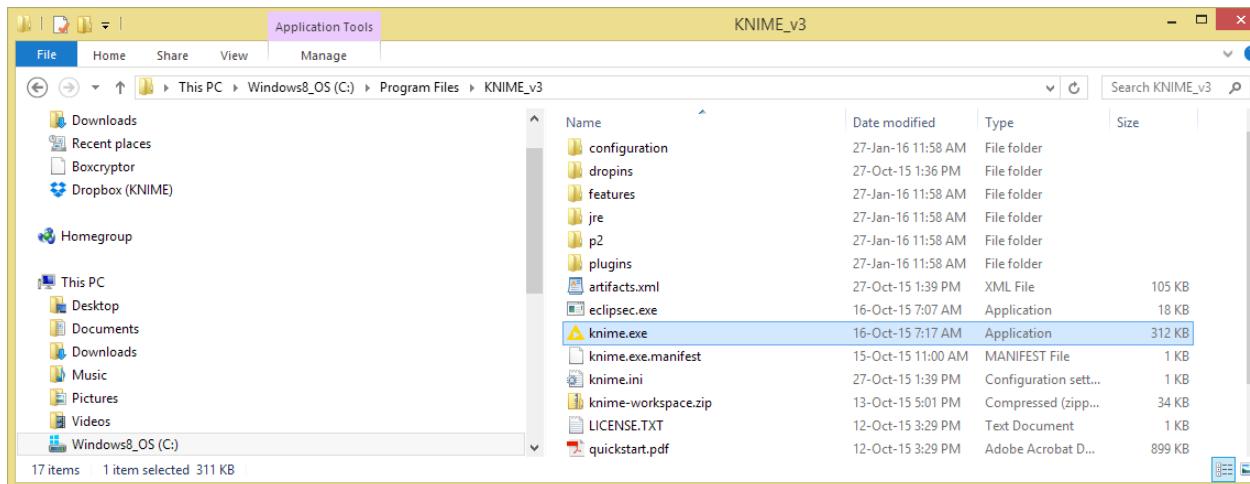
Windows	
KNIME Analytics Platform for Windows (installer) <i>The installer adds an icon to the desktop and suggests suitable memory settings</i>	<a href="#">32 Bit</a> (393.38 MB) <a href="#">64 Bit</a> (396.38 MB)
KNIME Analytics Platform for Windows (self-extracting archive) <i>The self-extracting archive only creates a folder holding the KNIME installation</i>	<a href="#">32 Bit</a> (396.87 MB) <a href="#">64 Bit</a> (400.72 MB)
KNIME Analytics Platform for Windows (zip archive)	<a href="#">32 Bit</a> (466.11 MB) <a href="#">64 Bit</a> (470.07 MB)

Linux	
KNIME Analytics Platform for Linux	<a href="#">64 Bit</a> (417.21 MB)

Mac	
KNIME Analytics Platform for Mac OSX (10.11 and above)	<a href="#">64 Bit</a> (388.44 MB)

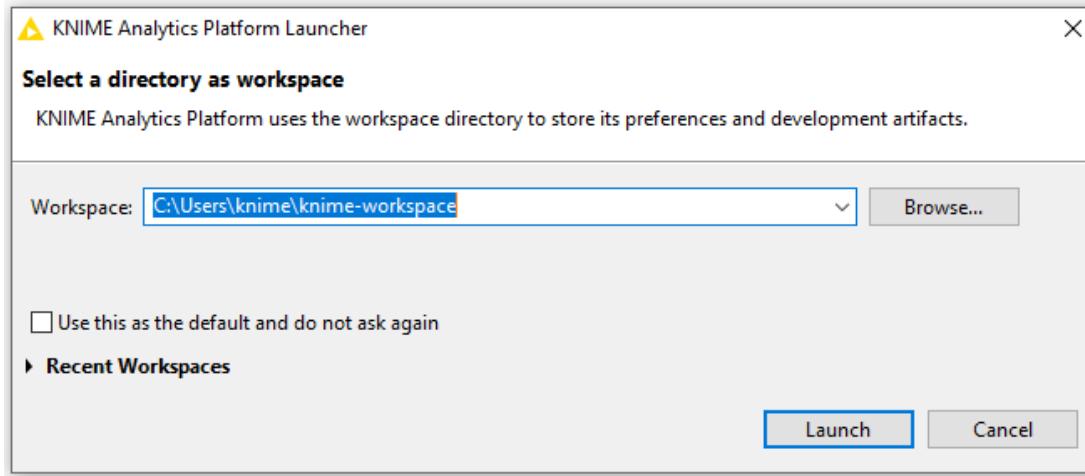
# Start KNIME Analytics Platform

- 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)

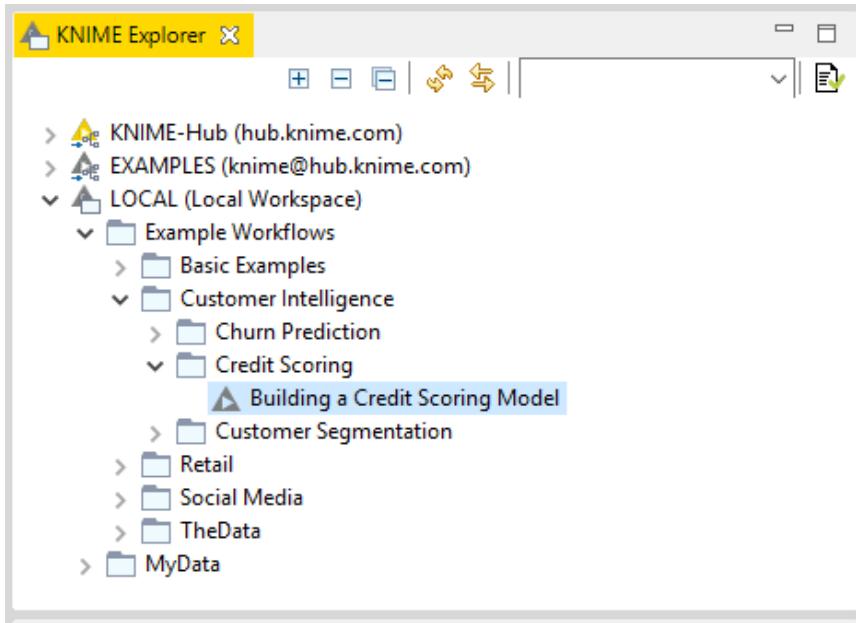


# The KNIME Analytics Platform Workbench

The screenshot displays the KNIME Analytics Platform Workbench interface with several panels:

- KNIME Explorer**: Shows the project structure with sections like My-KNIME-Hub, EXAMPLES, LOCAL (Local Workspace), and Example Workflows. A yellow box highlights the "KNIME Explorer" section.
- Workflow Coach**: Lists recommended nodes by category (Community, Recommended Nodes) such as GroupBy, Joiner, Column Filter, Concatenate, CSV Writer, Reference Row Filter, String Manipulation, and Math Formula. A yellow box highlights the "Workflow Coach" section.
- Node Repository**: Lists categories like IO, Manipulation, Views, Analytics, DB, Other Data Types, Structured Data, Scripting, Tools & Services, and Community Nodes. A yellow box highlights the "Node Repository" section.
- Workflow Editor**: Displays a workflow titled "My first Workflow" with nodes: File Reader (read adult.csv), Row Filter (keep only records born in the US), Column Filter (remove gender), and Table Writer (Write table). A yellow box highlights the "Workflow Editor" section.
- Outline**: Shows the outline of the current workflow.
- Console & Node Monitor**: Monitors the execution status of nodes. It shows the node "Row Filter (3:2)" is EXECUTED. The "Port Output" table displays data from the "adult.csv" file. A yellow box highlights the "Console & Node Monitor" section.
- Description**: Provides a detailed description of the "Row Filter" node, explaining its function and configuration options. A yellow box highlights the "Node Description" section.
- KNIME Hub Search**: Allows searching for workflows, nodes, and more.

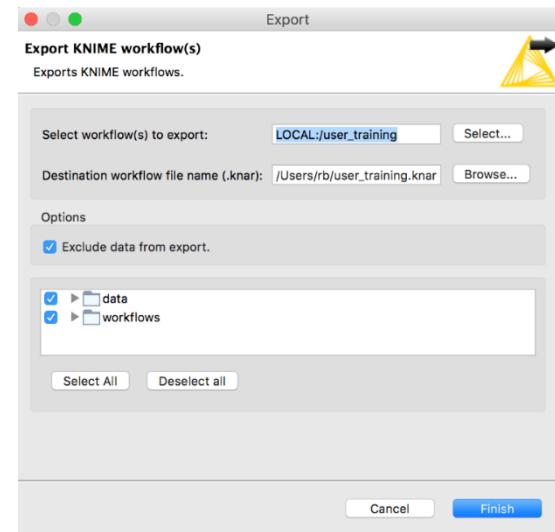
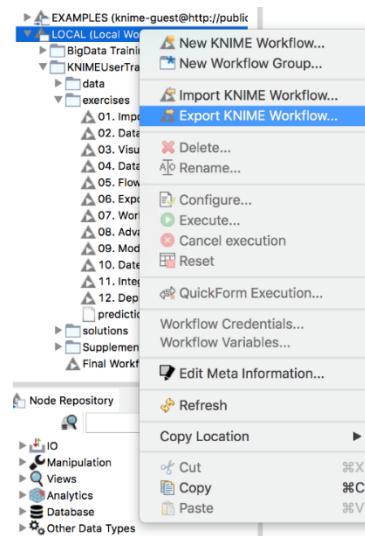
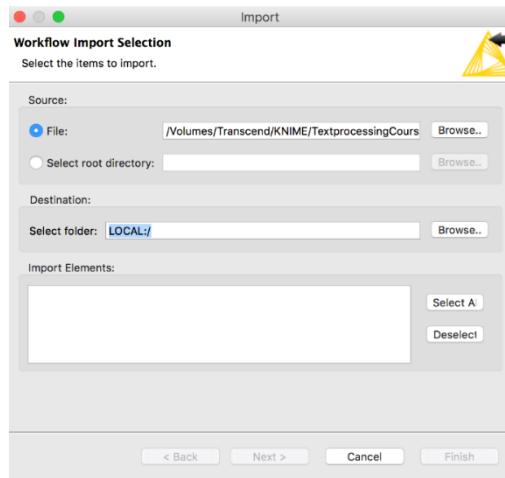
# KNIME Explorer



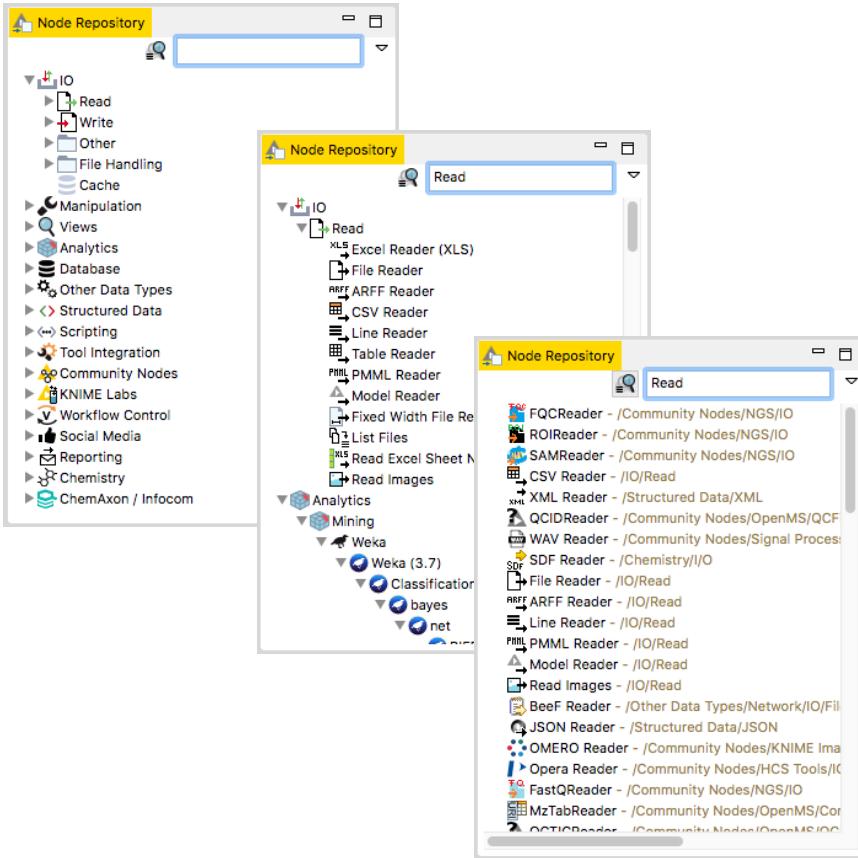
- In LOCAL you can access your own workflow projects.
- The Explorer toolbar on the top has a search box and buttons to
  - ➡ select the workflow displayed in the active editor
  - ⟳ refresh the view
- The KNIME Explorer can contain 4 types of content:
  - Workflows
  - Workflow groups
  - Data files
  - 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

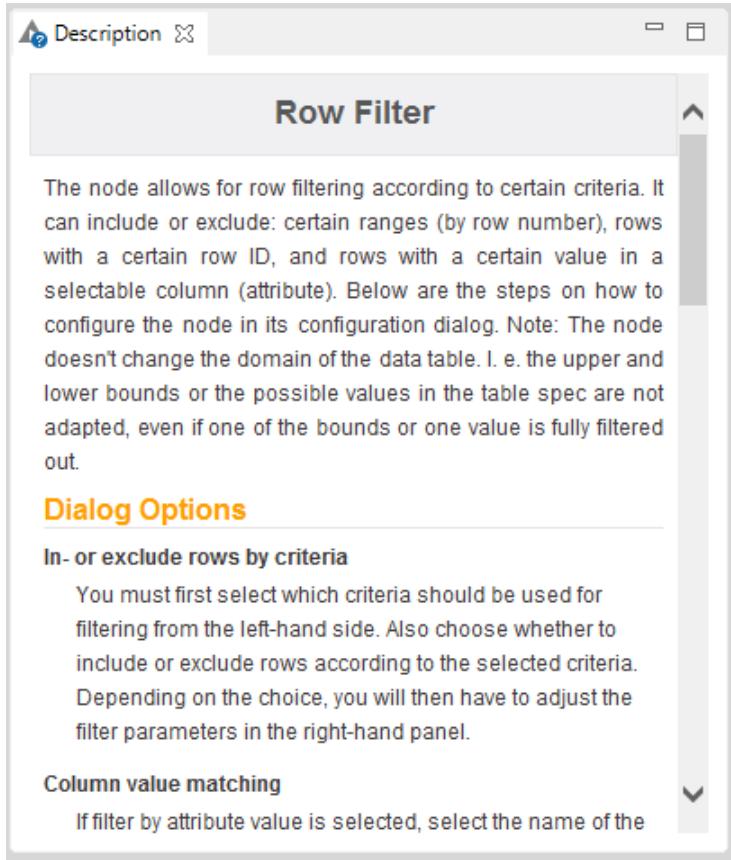


# Node Repository



- 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 screenshot shows the 'Description' window for the 'Row Filter' node. The title bar says 'Description' with a help icon. The main content area has a header 'Row Filter'. Below it, a text block explains the node's functionality: it allows for row filtering according to certain criteria, such as ranges by row number, specific row IDs, or values in a column. It notes that the node does not change the domain of the data table. A section titled 'Dialog Options' follows, with a heading 'In- or exclude rows by criteria'. The text here instructs users to select filtering criteria and adjust parameters. Another section, 'Column value matching', is partially visible at the bottom.

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

# Workflow Description

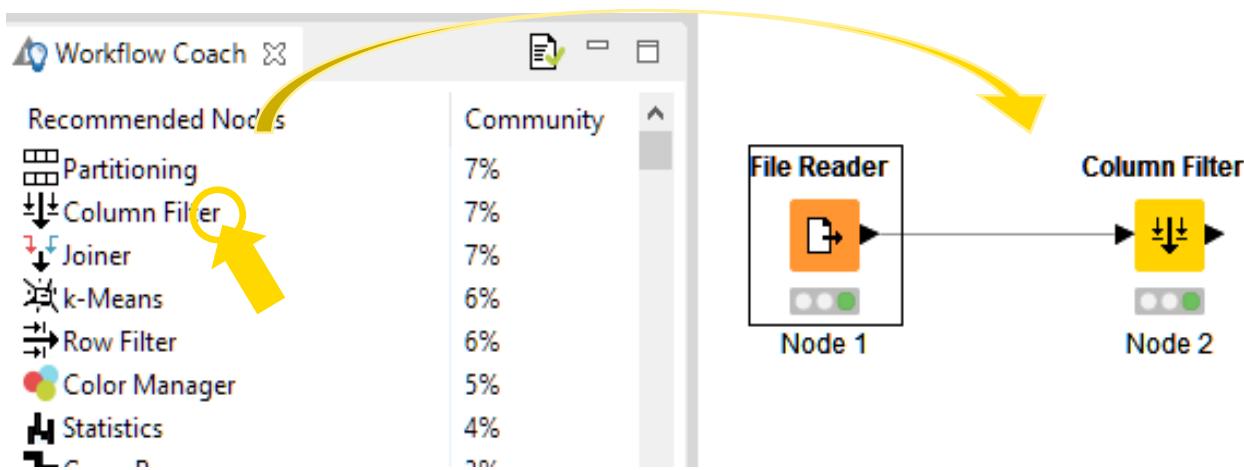
The screenshot shows the KNIME Description window for a workflow titled "My\_First\_Workflow". The window has a yellow header bar with a question mark icon, the title "Description", and a close button. The main content area contains the following information:

- Title:** My First Workflow
- Description:** This workflow reads data, removes uninteresting columns and rows, and writes the resulting data table to a CSV file.
- Tags:** Example Workflow, CSV, Data Manipulation
- Links:**
  - KNIME Homepage
  - KNIME Hub
  - KNIME Forum
- Creation Date:** 2019-7-2
- Author:** Ana Vedoveli

- 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



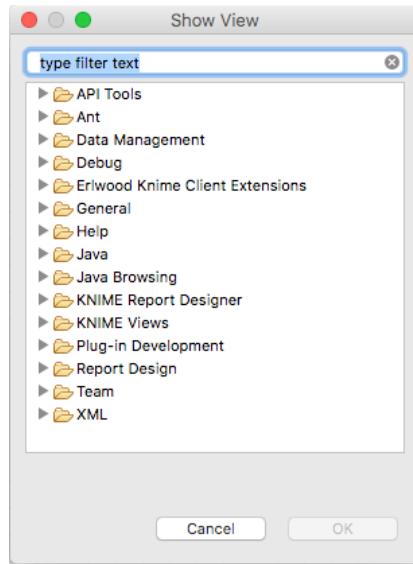
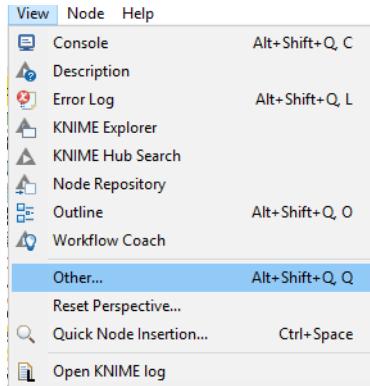
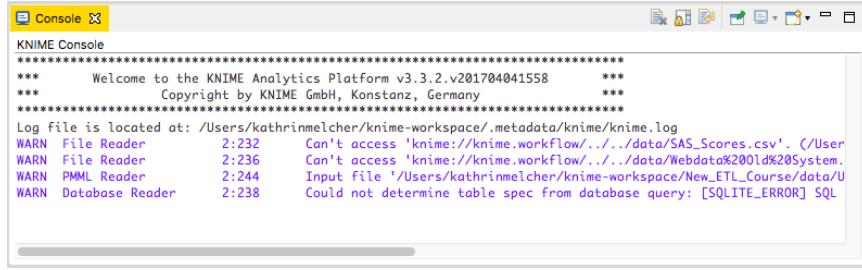
# Node Monitor

- 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 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 following options: 'Show Output Table' (which is checked), 'Show Variables', 'Show Configuration', 'Show Entire Configuration', 'Show Node Timing Information', and 'Show Graph Annotations'. The 'Products' column in the table lists categories like 'Private Investn', 'P+B Investmer', 'Gold Investme', 'P+B Investmer', and 'Fund Manager'.

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

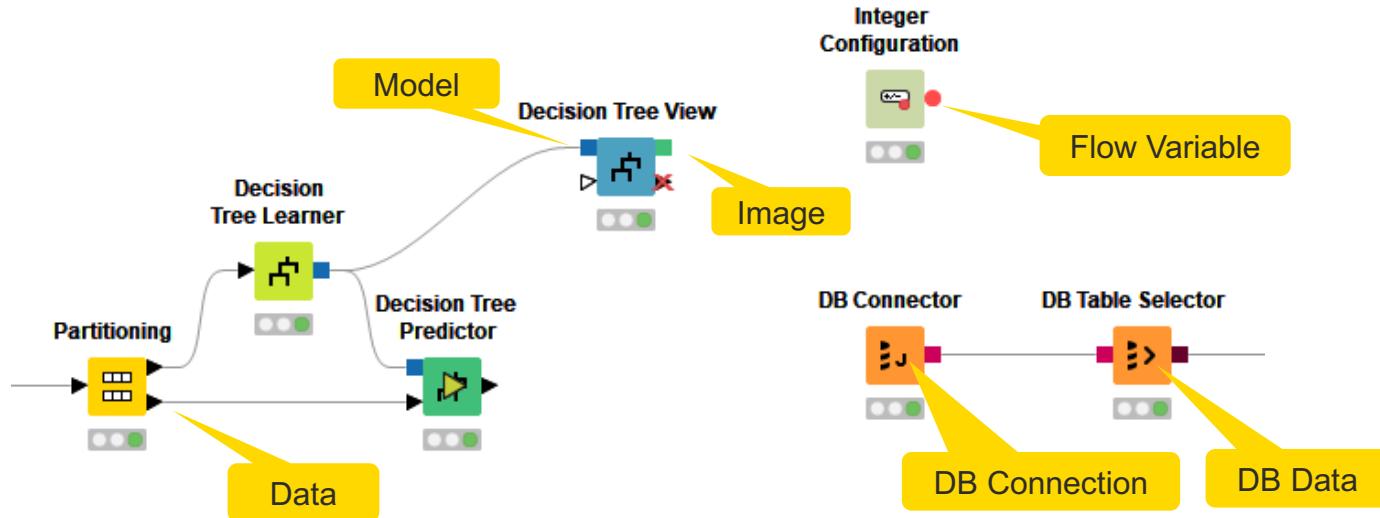
# Console and Other Views



- Console view prints out error and warning messages about what is going on under the hood
  
- Click 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:

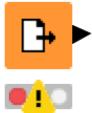


# More on Nodes...

---

- 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

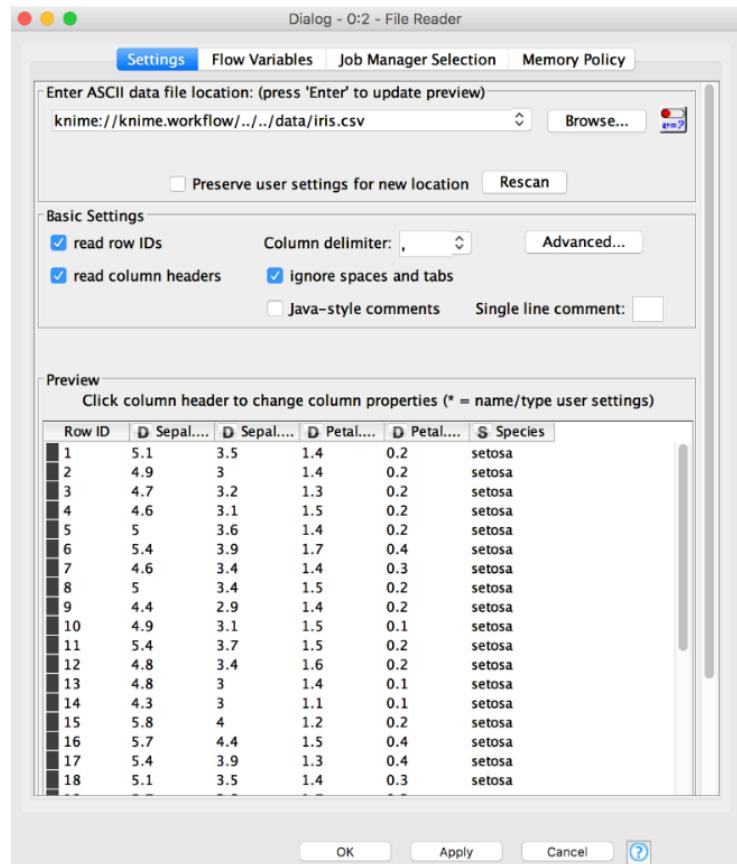


## Error:

The node has encountered an error during execution.

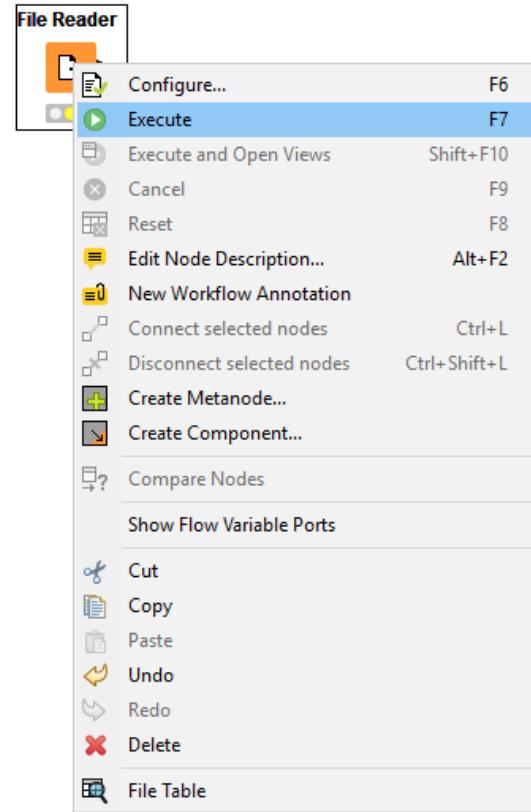
# Node Configuration

- 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 the context menu
- If execution is successful, status shows green light
- If execution encounters errors, status shows red light



## Tool Bar

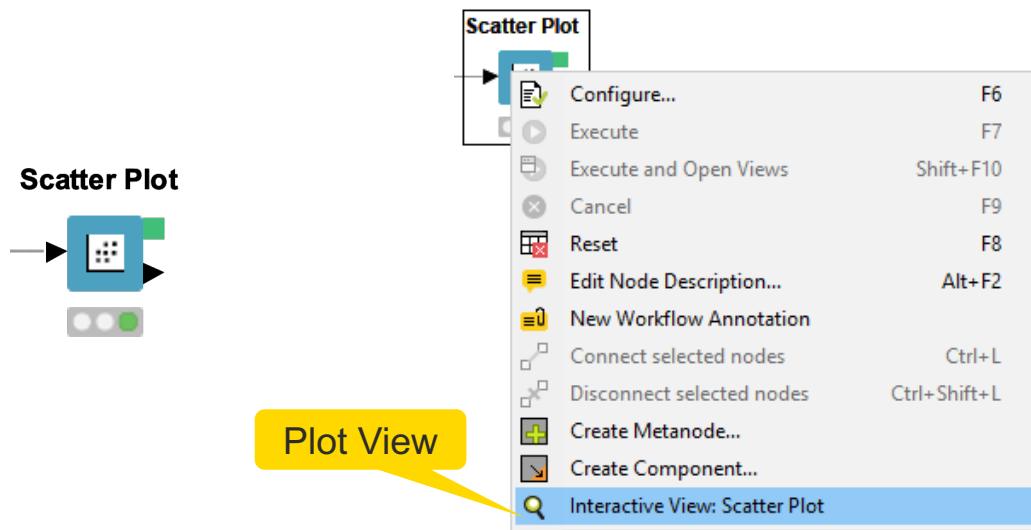


The buttons in the toolbar can be used for the active workflow. The most important buttons are:

-  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 Hub

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

The screenshot shows the KNIME Hub search results for the query "Sentiment Analysis". The results page displays 350 items. At the top, there are filters for All, Nodes, Components, Workflows, and Extensions. Below the filters, two workflow cards are visible:

- 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 example is located in the KNIME Examples > 04\_Analytics > 14\_Deep\_Learning > 02\_Keras > 08\_Sentiment\_Analysis\_with\_Deep\_Learning\_KNIME\_nodes folder.
- 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 example is located in the KNIME Examples > 04\_Analytics > 14\_Deep\_Learning > 02\_Keras > 07\_Sentiment\_Analysis\_with\_Deep\_Learning folder.

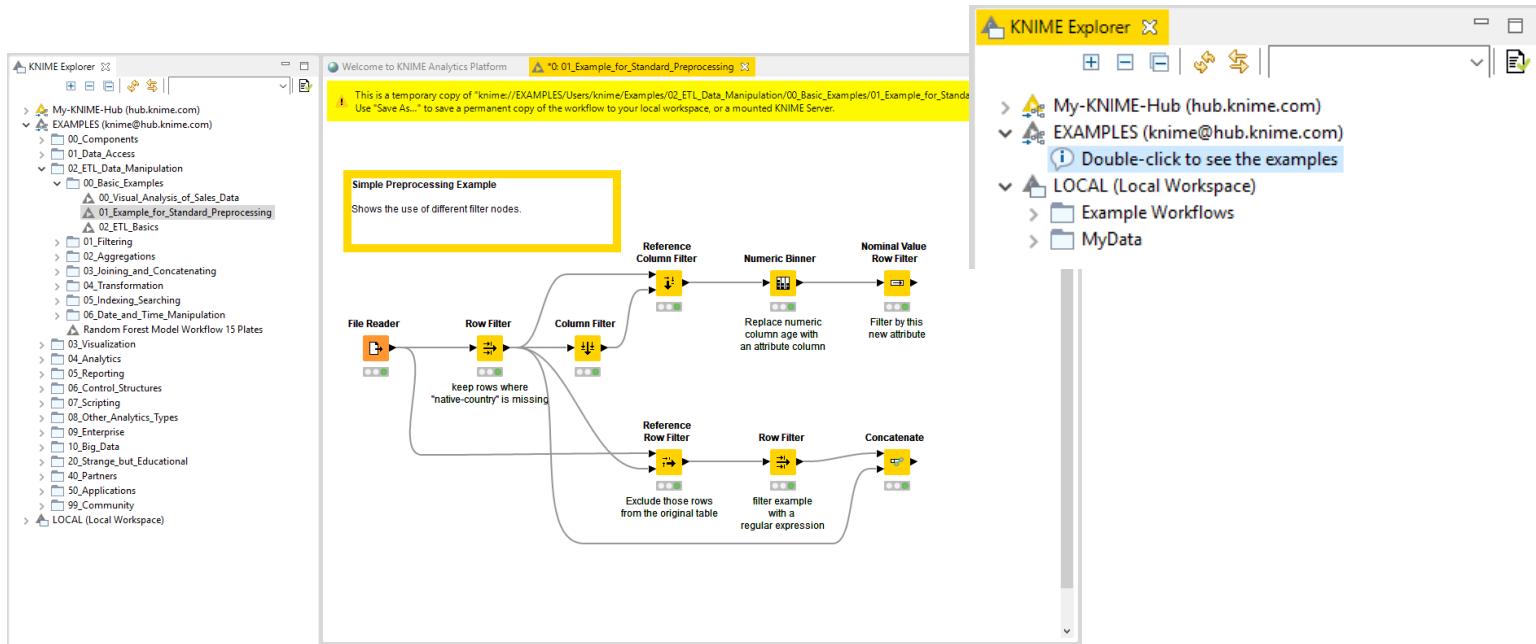
At the bottom, there is a third card for "Sentiment Analysis (Classification) of Documents".

The screenshot shows the main homepage of the KNIME Hub. It features a search bar at the top with the placeholder "Search workflows, nodes and more...". Below the search bar, the text "Welcome to the KNIME Hub" is displayed, followed by the subtext "The place to find and collaborate on KNIME workflows and nodes. Here you can find solutions for your data science questions." To the right of the text, there are four large numerical statistics: 3 993 Nodes, 265 Components, 2 541 Workflows, and 211 Extensions. Below these statistics are three call-to-action boxes: "How to Getting started From downloading through to building your first workflow", "KNIME Forum Get help from our community and help others", and a thumbnail image of a person working on a laptop.

<https://hub.knime.com>

# Getting Started: KNIME Example Server

- 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