

# **KNIME**

# **Analytics**

# **Platform**



**Dr. José Ramón Iglesias**

DSP-ASIC BUILDER GROUP

Director Semillero TRIAC

Ingenieria Electronica

Universidad Popular del Cesar

## Summary of this lesson

*„If the only tool you have is a hammer, you tend to see every problem as a nail”*

*-Abraham Maslow*

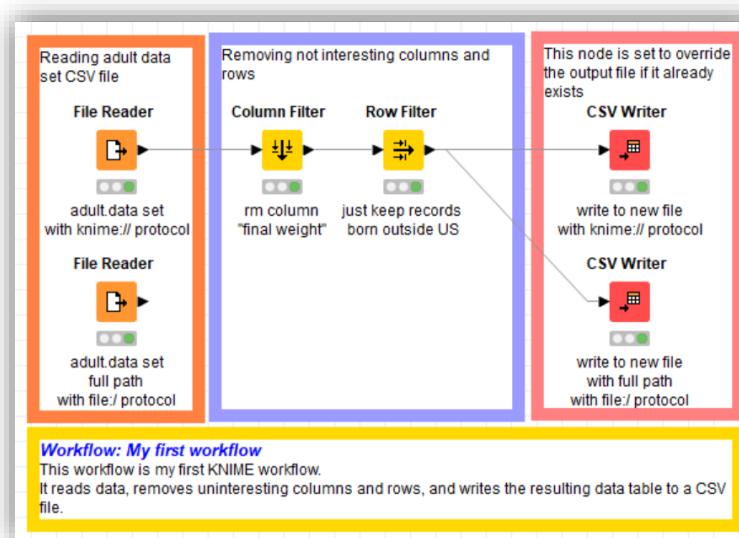
First the tool or the knowledge?

## Content of this lesson

- Download and Install
- The Workbench
- More on Nodes
- Metanodes and Components
- KNIME Community Hub
- Build your first „Hello“ Workflow

# Datasets

- Dataset used : adult dataset
- Example Workflows:
  - „My First Workflow“ <https://kni.me/w/kYeZOLeAJXo9MvoI>
    - Read from CSV file, Excel file and SQLite.
    - Filter rows and columns
    - Write to CSV file



# Download and Install

- Open and opensource modular Data Science platform
- Covers all the data science needs:



- Based on the visual programming paradigm
- Provides a diverse array of extensions:
  - Text Mining
  - Network Mining
  - Cheminformatics
  - Deep Learning
  - Many integrations, such as Java, R, Python, Weka, Keras, Plotly, H2O, etc
  - ... And more

## KNIME Analytics Platform

- To develop data science solutions
  - Structured data
  - Unstructured data
  - Machine Learning
  - Statistics
- Open source
- Free

## KNIME Business Hub

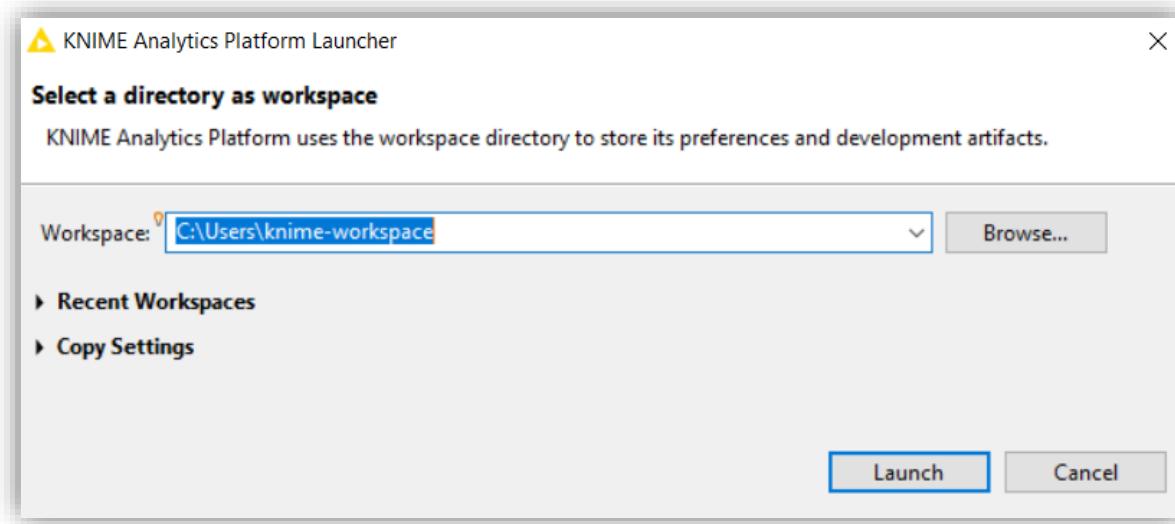
- To integrate the solutions into the IT environment
  - Scheduling
  - MLOps
  - Easy deployment
  - REST architecture
  - Auditing tools
- Closed source
- Yearly license

<https://www.knime.com/downloads>

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

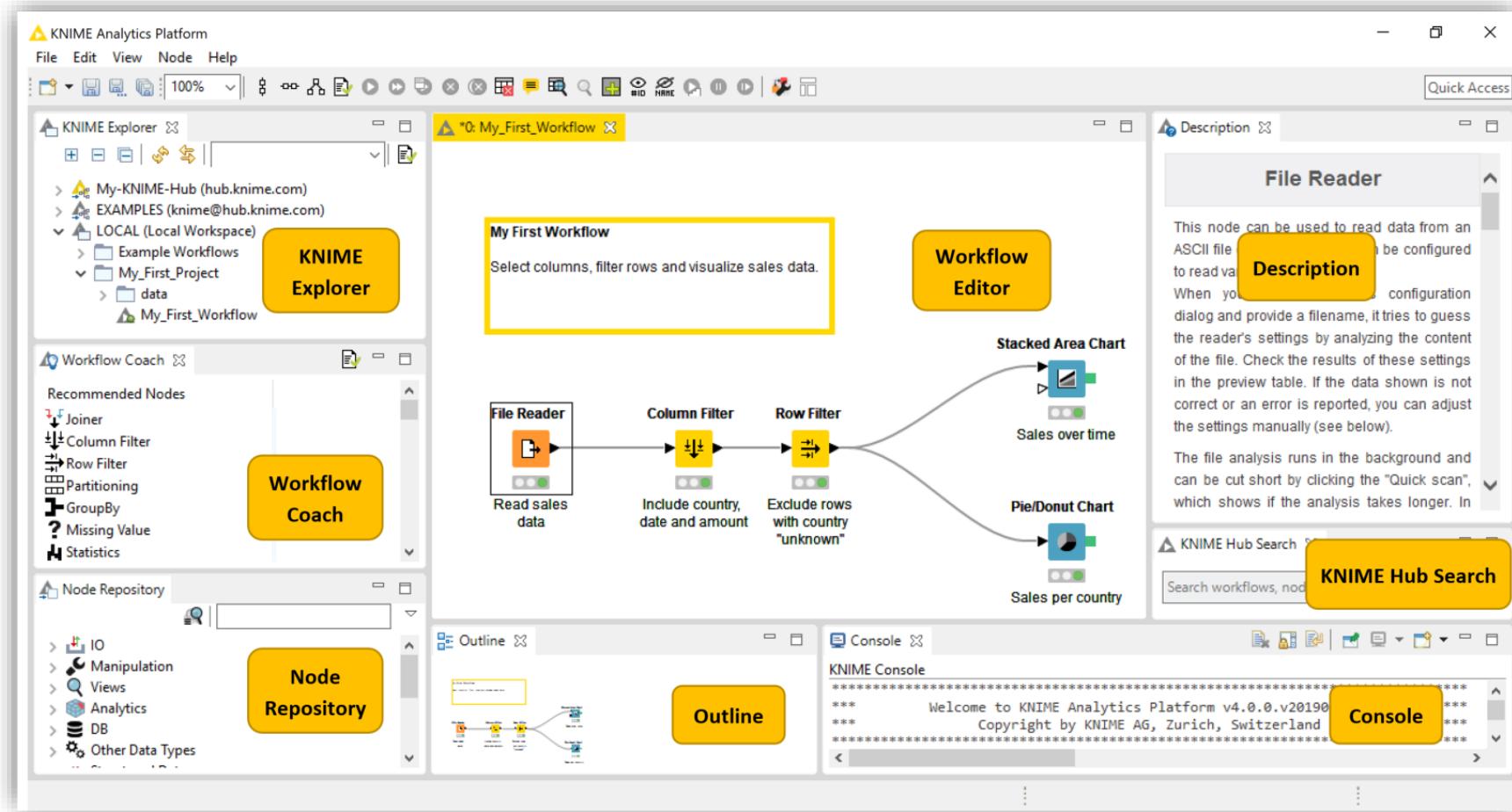
# The Workbench

# The KNIME Workspace

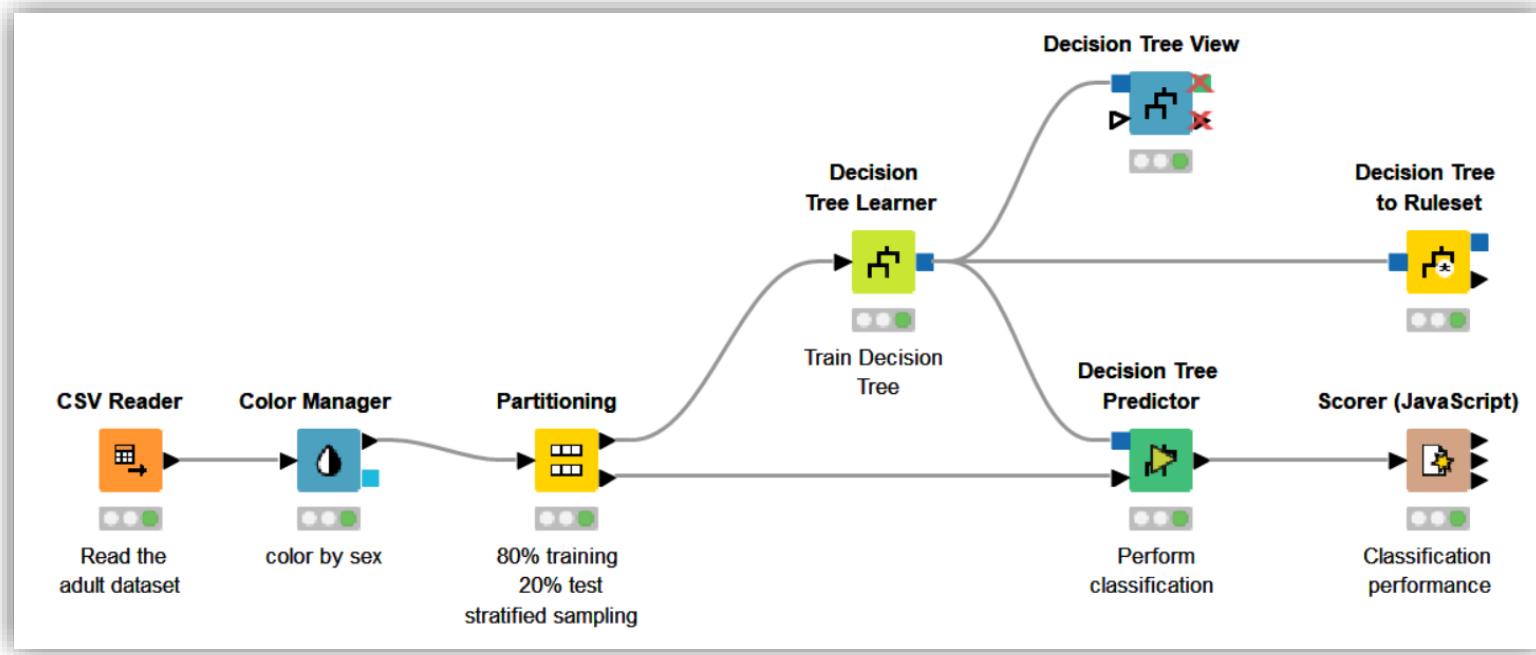


- The workspace is the **folder/directory** in which workflows (and potentially data files) are stored for the current session.
- Workspaces are portable (just like KNIME Analytics Platform)

# The KNIME Workbench

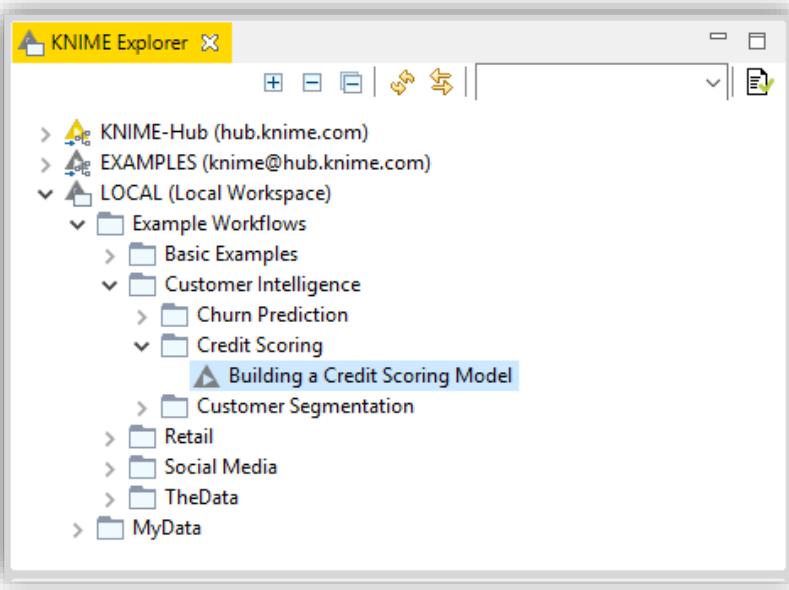


# Workflow



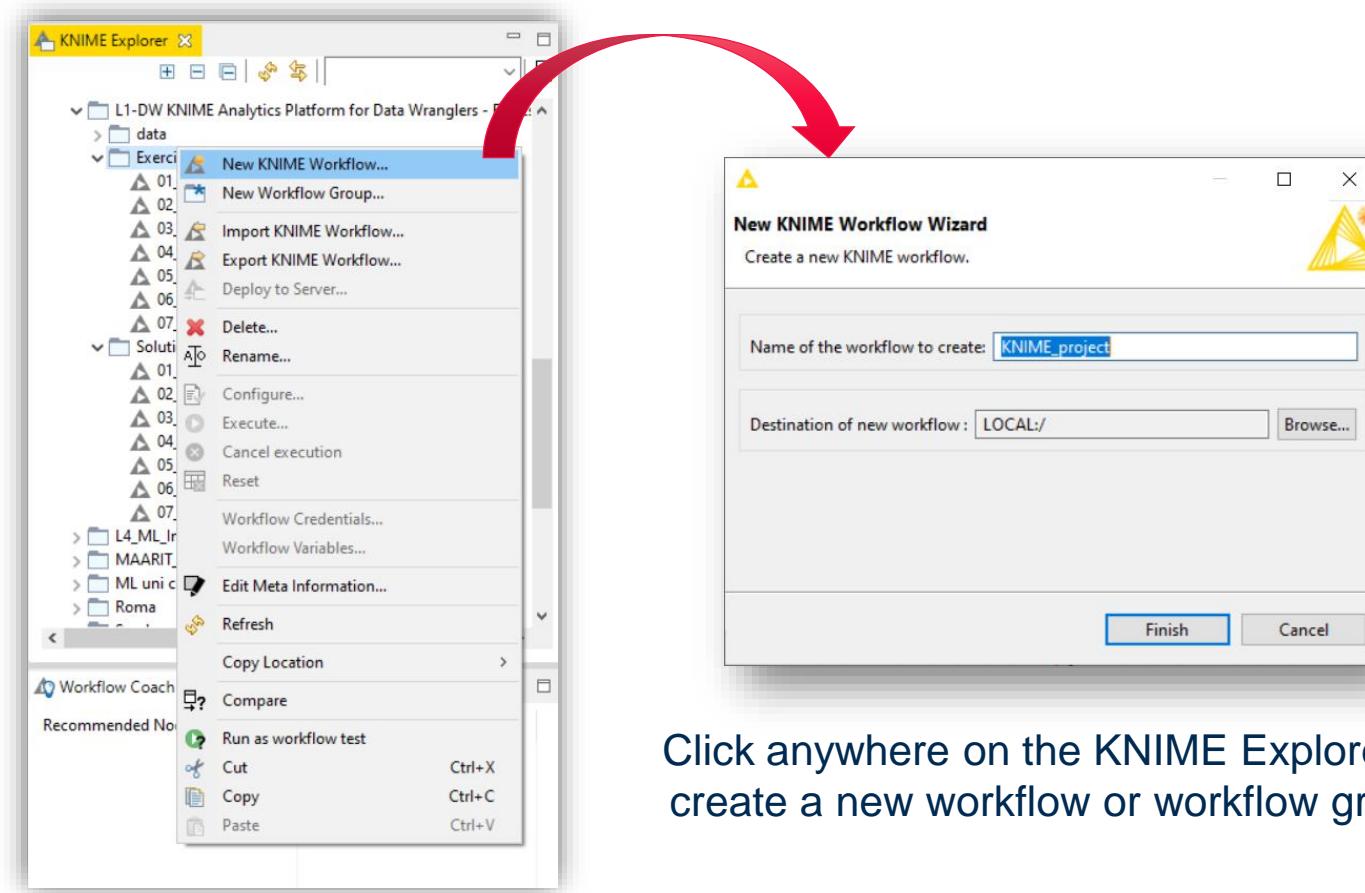
A workflow is a pipeline of nodes, each configurable to perform a specific task.  
The data flow through nodes from left to right

# KNIME Explorer



- This panel displays all the workflows in the selected workspace
  - LOCAL: projects saved on your own machine
  - EXAMPLES: hundreds of read-only example workflows
  - My-KNIME-Hub: additional space where you can share your workflows with the community or just park your work for yourself
- Provides a search box and buttons to
  - Refresh the view
  - Select the currently displayed workflow
- Can display 4 types of content
  - Workflows
  - Workflow groups
  - Data files
  - Shared Components

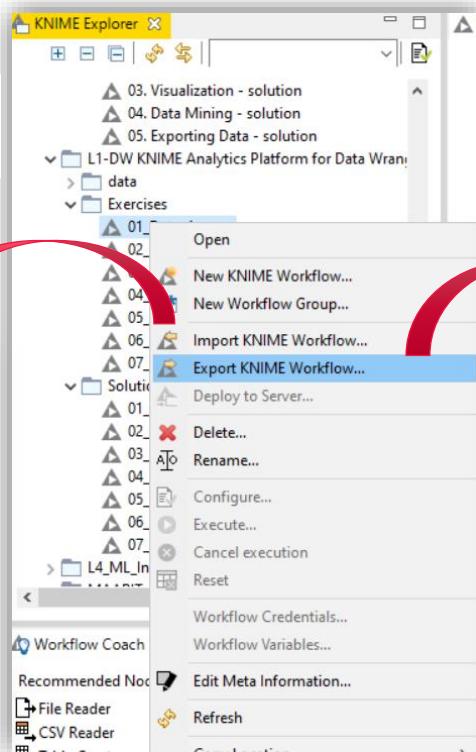
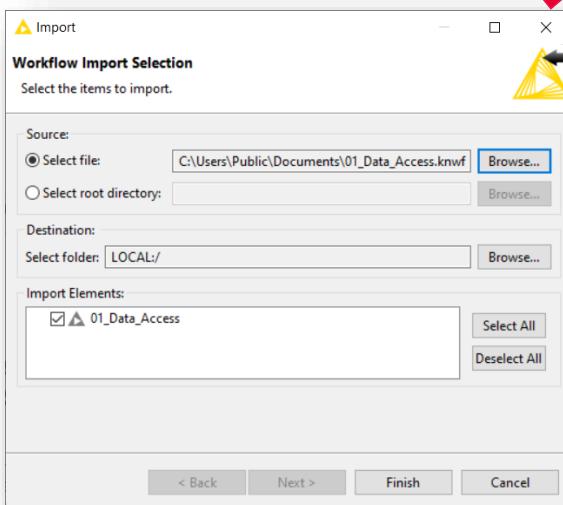
# Creating a new workflow



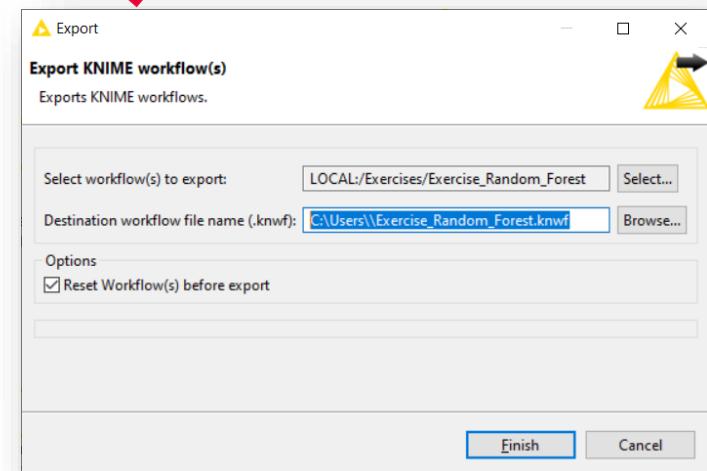
Click anywhere on the KNIME Explorer to create a new workflow or workflow group

# Importing and Exporting Workflows

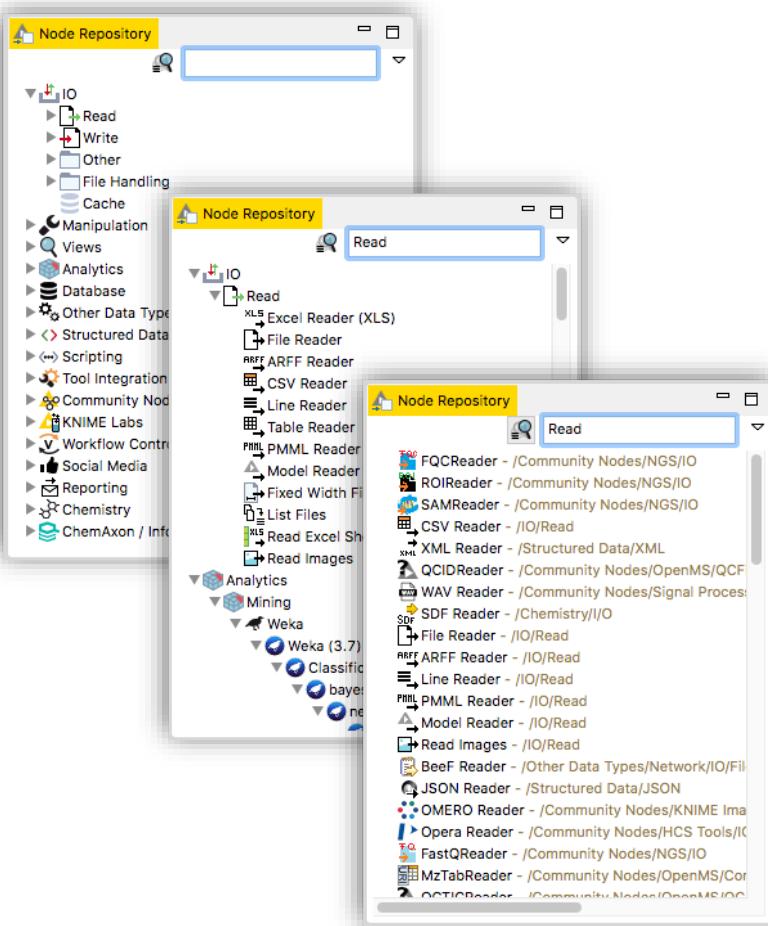
Right-click anywhere in KNIME Explorer to import a workflow



Right-click on a workflow or workflow group to export the selected workflow

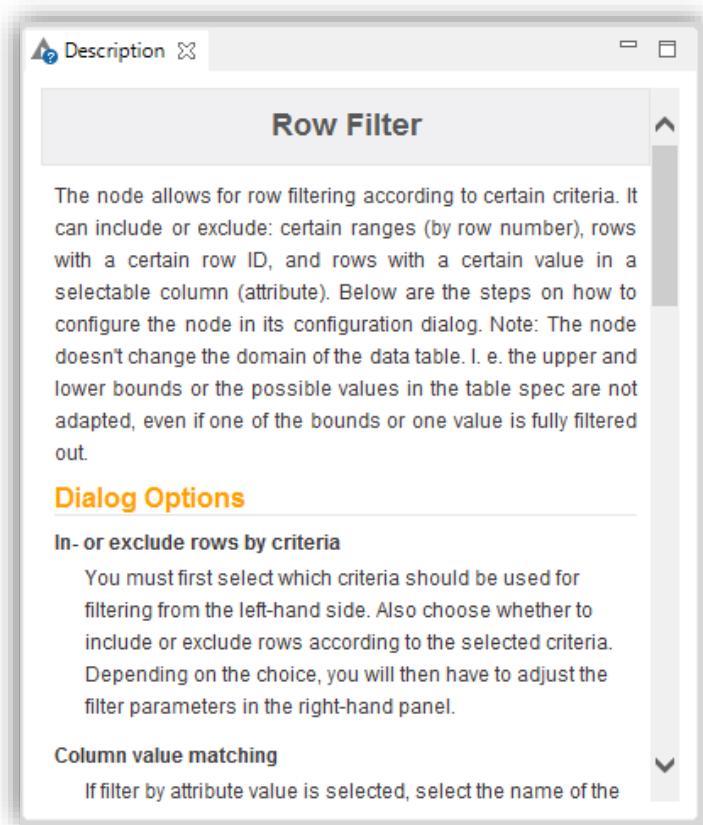


# Node Repository



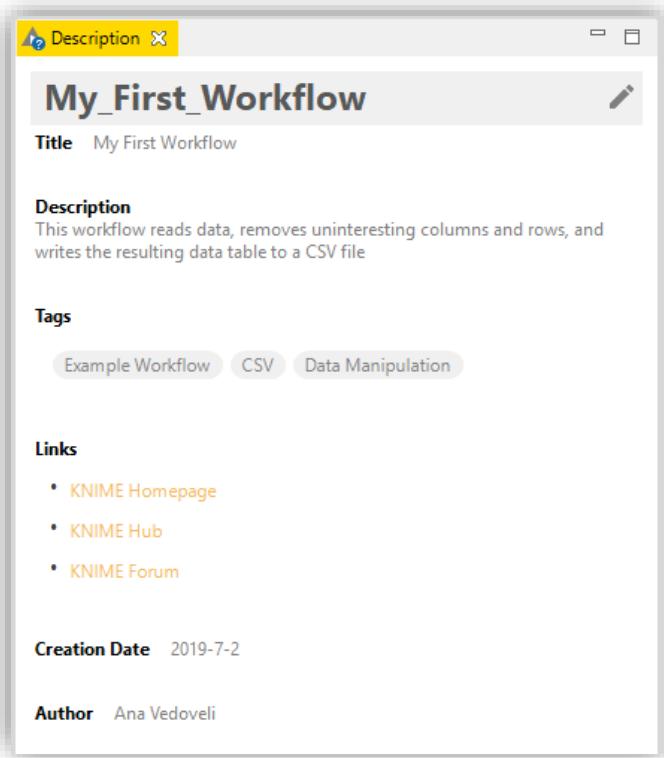
- The Node Repository contains all KNIME nodes - ordered by category with further subcategories.
- Extensions installation can sensibly increase the number of nodes
- Two search methods:
  - Crisp Search
  - Fuzzy Search
- Nodes can be added by drag and drop from the Node Repository to the Workflow Editor

# Node Description



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

# Workflow Description



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

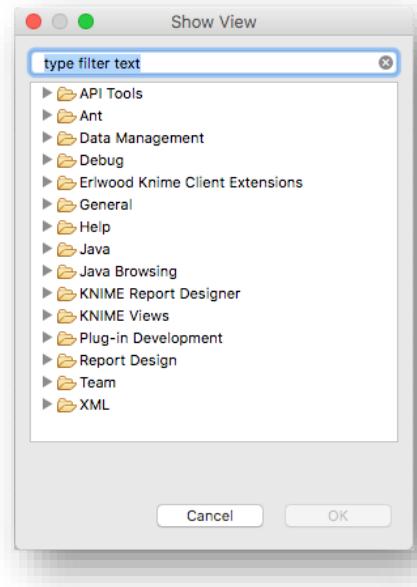
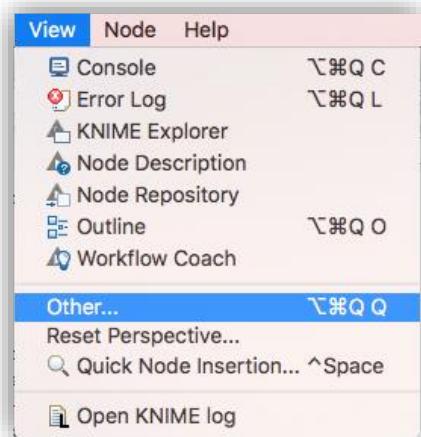
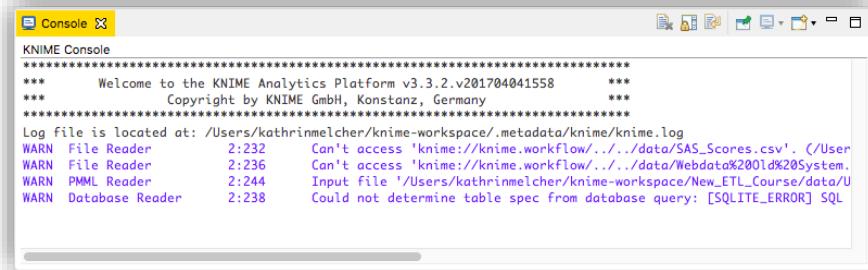
- Node Recommendation engine
- It gives hints about which node to use next in the workflow
- It is based on world-wide KNIME community usage statistics
- It can also be set to use personal and local group usage statistics

The screenshot shows the KNIME Workflow Coach interface. On the left, a sidebar titled "Recommended Nodes" lists several nodes with their usage percentages in the "Community" column:

Node	Community
Joiner	12%
Column Filter	10%
Row Filter	6%
GroupBy	5%
Partitioning	4%
Concatenate	3%
Statistics	3%

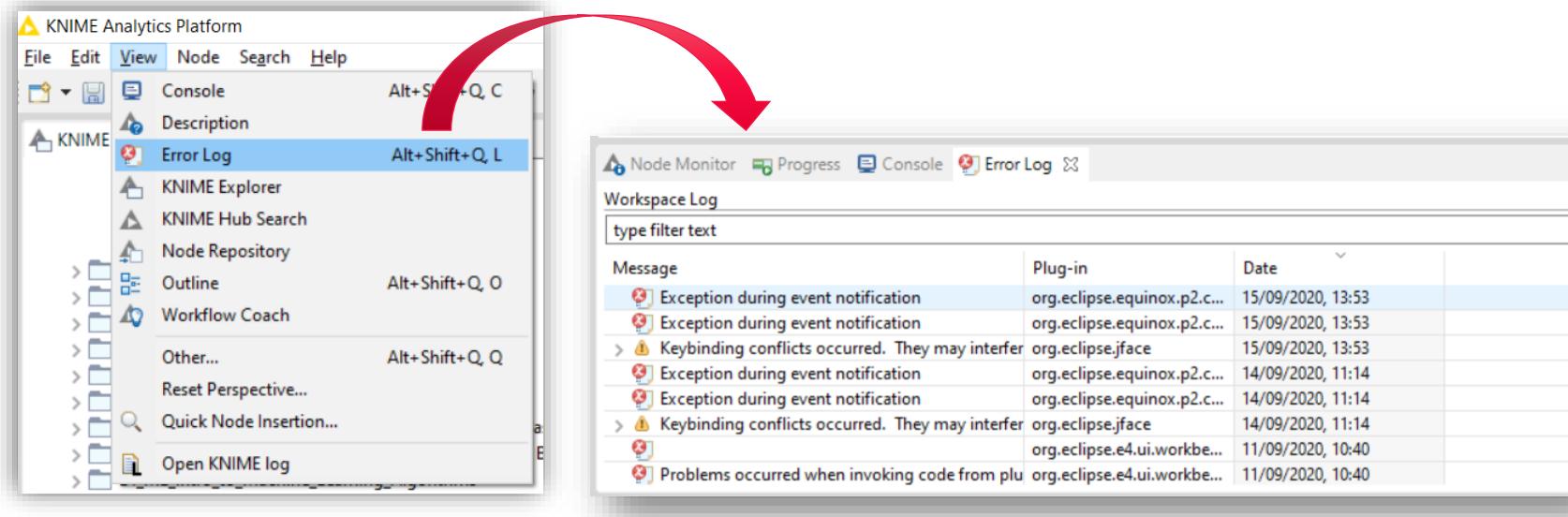
On the right, a workflow diagram is displayed. It starts with a "CSV Reader" node, followed by a "Column Filter" node. A tooltip for the "Column Filter" node reads "Remove columns". Below the diagram, the text "Read the adult dataset" is visible.

## 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 additional views

## Error Log View

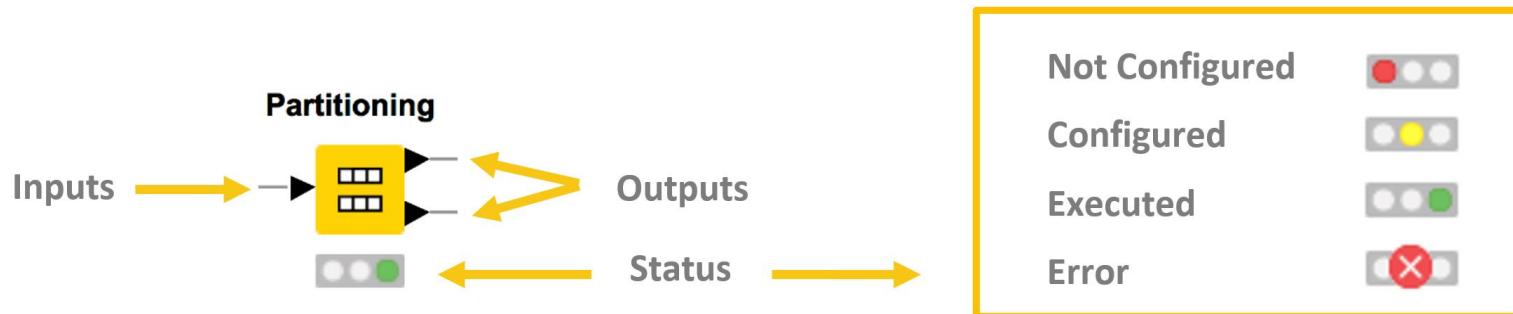


Tip: enabling and checking the Error Log view can help while debugging your project

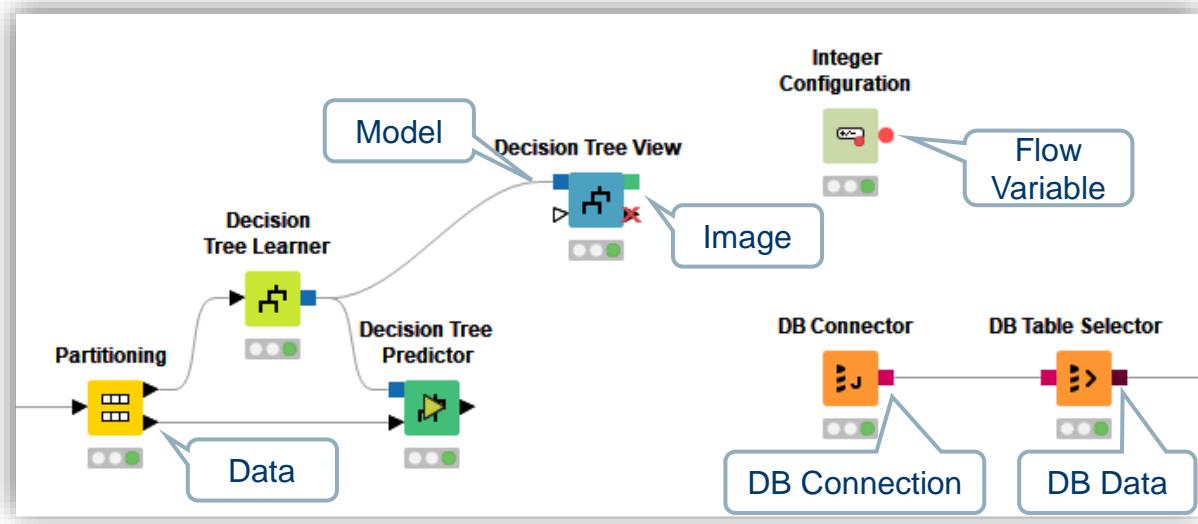
# More on Nodes

## More on Nodes...

- Nodes are the basic processing units of a workflow
- Each node has a number of input and/or output ports
- Data is transferred over a connection from an out-port to the in-port(s) of other nodes
- Under each node, a light shows its status

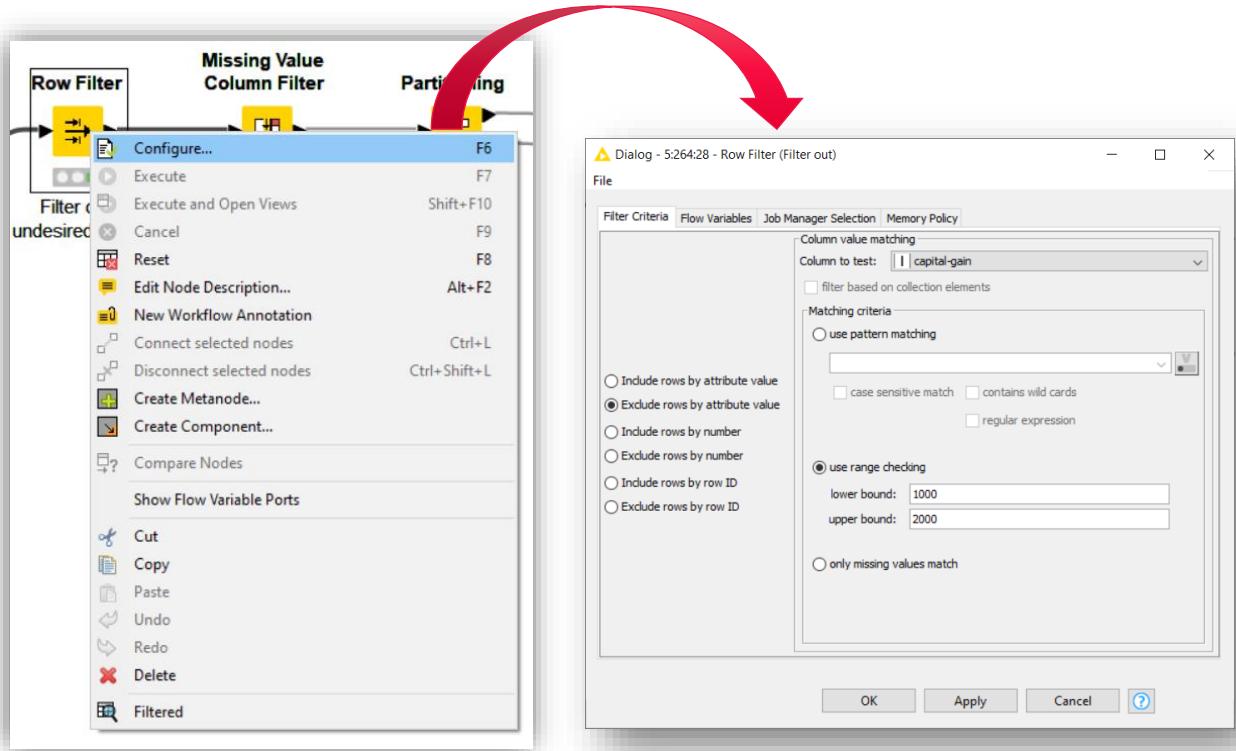


# Data Port Types



- A pipeline of such nodes makes a **workflow**
- The result of the node's operation on the data is provided at the out-port to successor nodes
- Only port of the same type can be connected

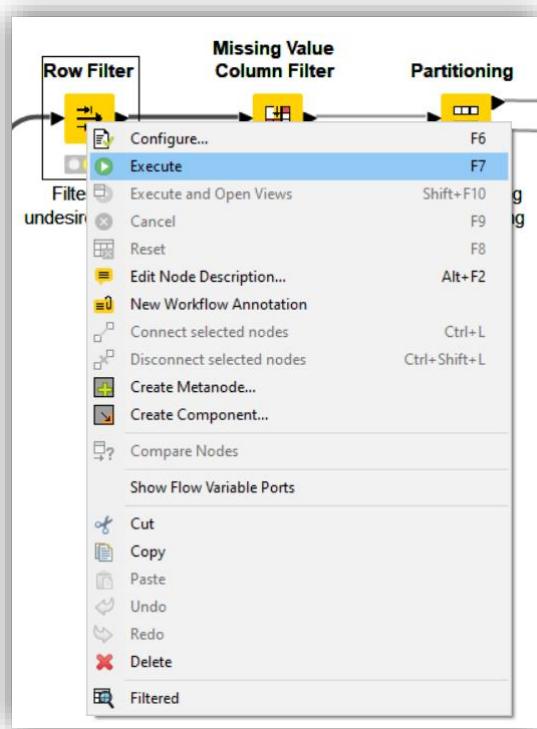
# Node Configuration



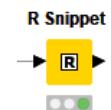
- Most nodes require configuration
- To access a node configuration window:
  - Double-click the node
  - OR
  - Right-click > Configure

# Node Execution

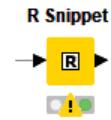
- Right-click node
- Select Execute in context menu



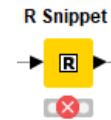
If execution is successful, status shows green light



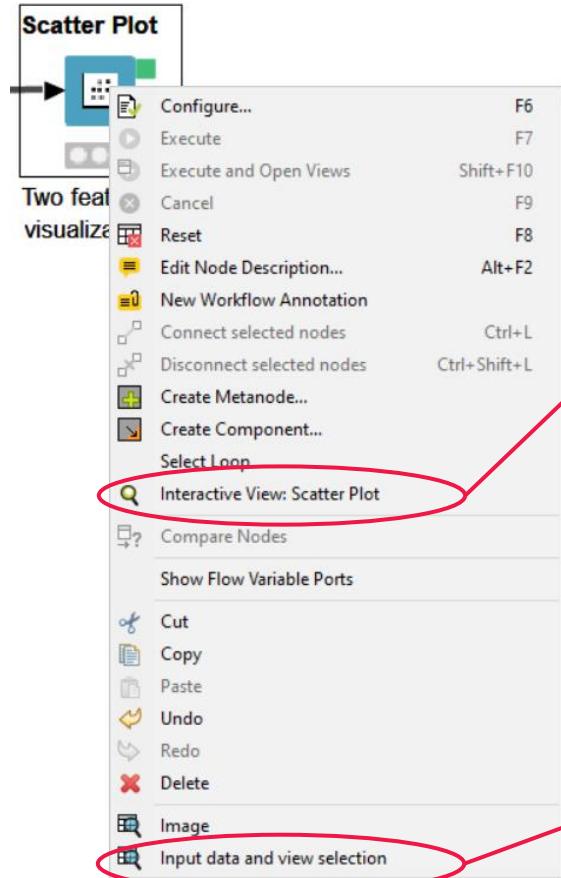
If execution produces warnings, status show yellow triangle



If execution encounters errors, status shows a red X



# Node Views



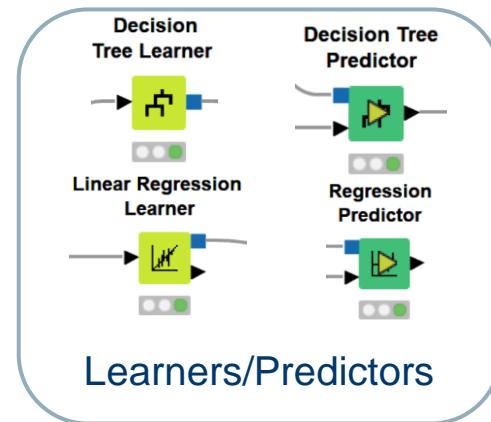
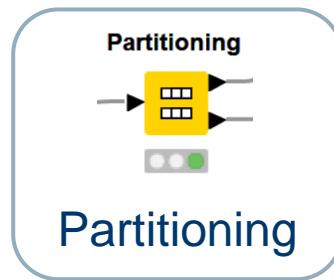
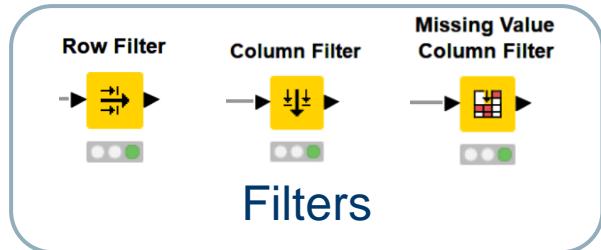
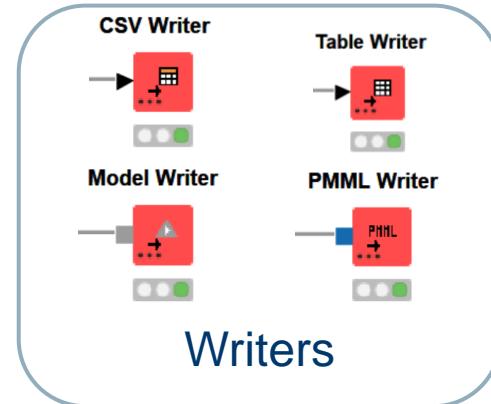
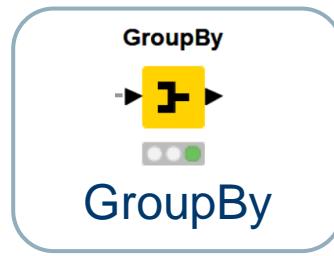
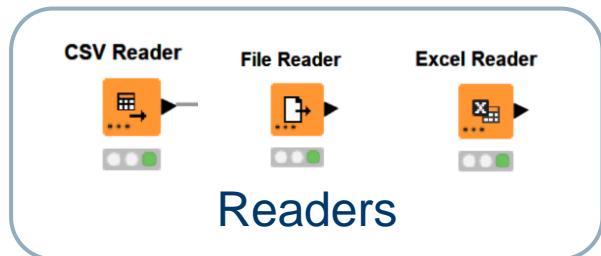
Interactive View

A table titled 'Input data and view selection - 1083:0:10 - Scatter Plot (Two features)' showing 148 rows of data. The columns are: Row ID, Sepal length, Sepal width, Petal length, Petal width, Class name, and Selected. The data includes rows like Row27\_Row0, Row28\_Row0, etc., with values such as 5.0, 3.5, 1.6, 0.6, Iris-setosa, false.

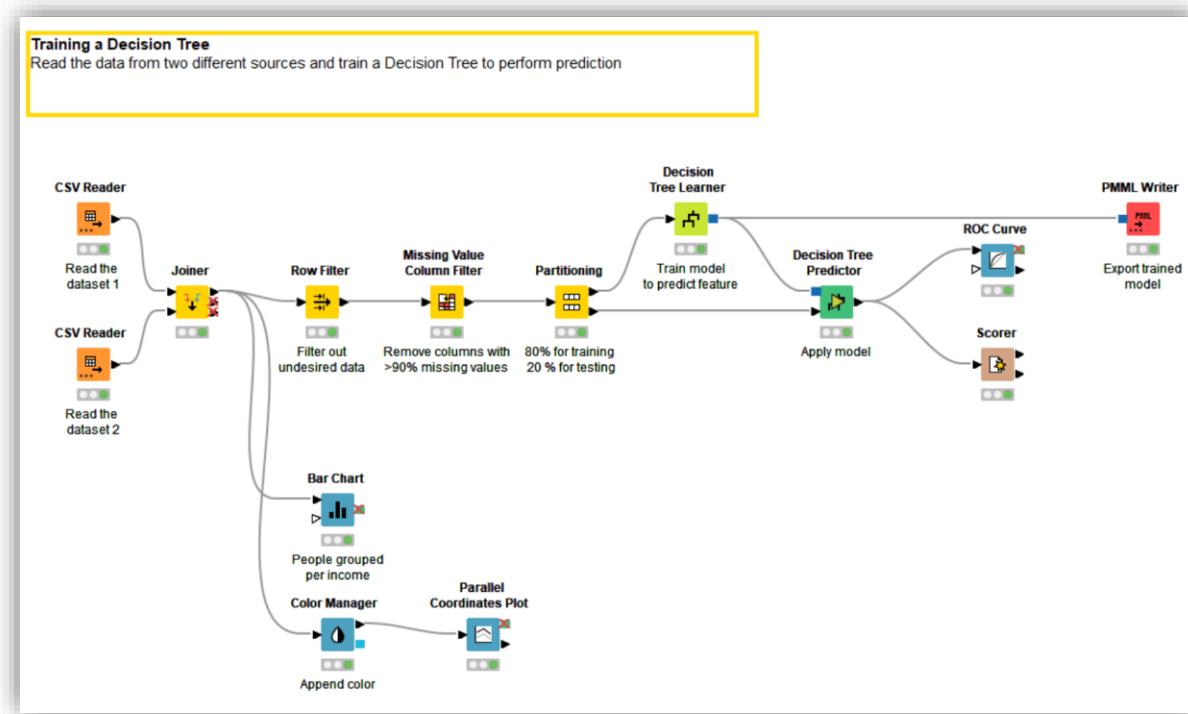
Row ID	D sepal le...	D sepal w...	D petal le...	D petal w...	S class n...	B Select...
Row27_Row0	5	3.5	1.6	0.6	Iris-setosa	false
Row28_Row0	4.8	3	1.4	0.3	Iris-setosa	false
Row29_Row0	4.6	3.2	1.4	0.2	Iris-setosa	false
Row30_Row0	5	3.3	1.4	0.2	Iris-setosa	false
Row31_Row1	6.4	3.2	4.5	1.5	Iris-versicolor	false
Row32_Row1	5.5	2.3	4	1.3	Iris-versicolor	false
Row33_Row1	6.5	2.8	4.6	1.5	Iris-versicolor	false
Row34_Row1	5.7	2.8	4.5	1.3	Iris-versicolor	false
Row35_Row1	4.9	2.4	3.3	1	Iris-versicolor	false
Row36_Row1	6.6	2.9	4.6	1.3	Iris-versicolor	false
Row37_Row1	5	2	3.5	1	Iris-versicolor	false
Row38_Row1	5.9	3	4.2	1.5	Iris-versicolor	false
Row39_Row1	6	2.2	4	1	Iris-versicolor	false
Row40_Row1	5.6	2.9	3.6	1.3	Iris-versicolor	false
Row41_Row1	6.7	3.1	4.4	1.4	Iris-versicolor	false
Row42_Row1	5.8	2.7	4.1	1	Iris-versicolor	false
Row43_Row1	6.2	2.2	4.5	1.5	Iris-versicolor	false
Row44_Row1	5.6	2.5	3.9	1.1	Iris-versicolor	false
Row45_Row1	6.1	2.8	4	1.3	Iris-versicolor	false
Row46_Row1	4.4	2.0	4.2	1.2	Iris-versicolor	false

Data View

# Frequently Used Nodes



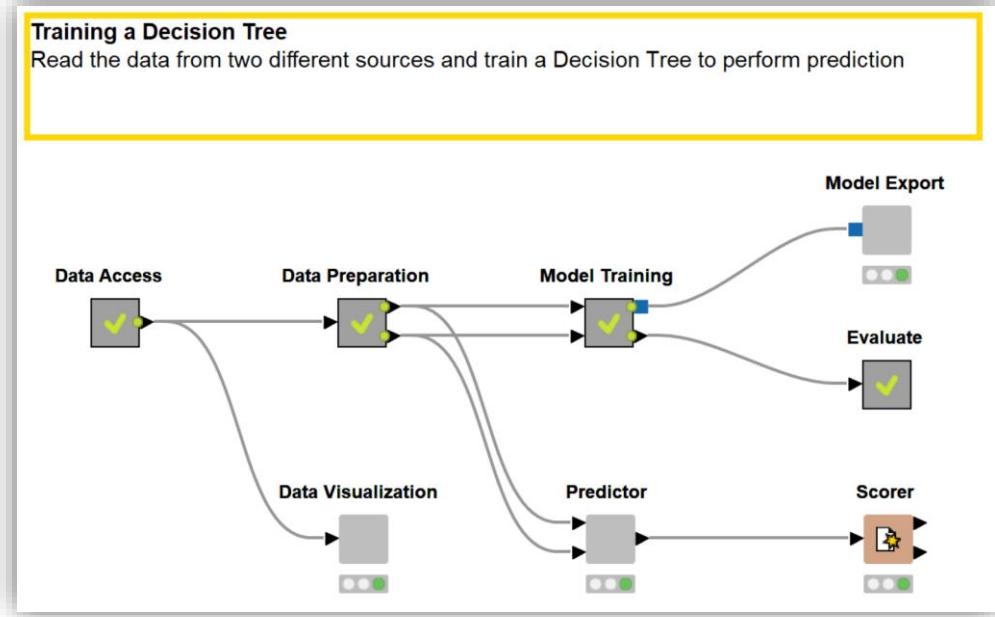
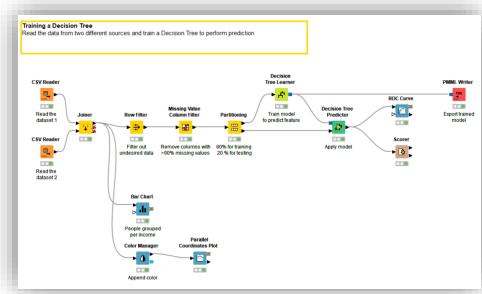
# Tidy up workflows



- Workflow can easily become complex and difficult to understand

# Metanodes and Components

# Tidy up workflows



- Metanodes and components can help tidying up, encapsulating nodes performing common operations

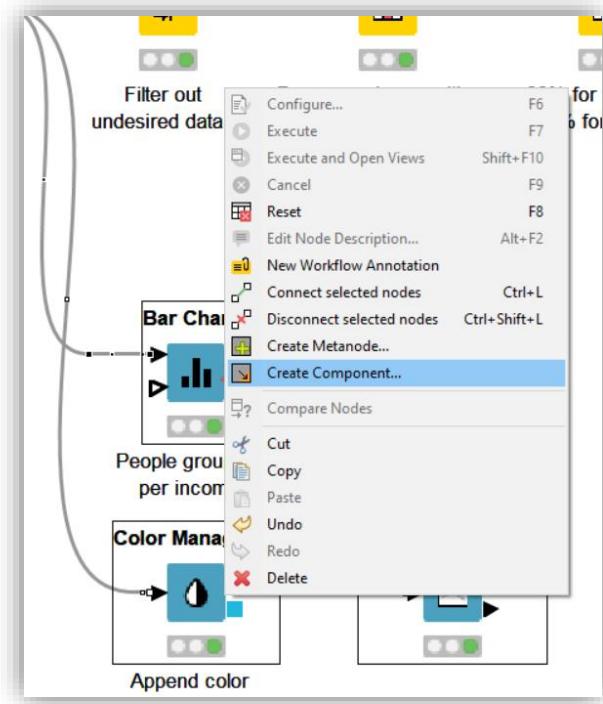
# Components

## Steps to build a component or a metanode

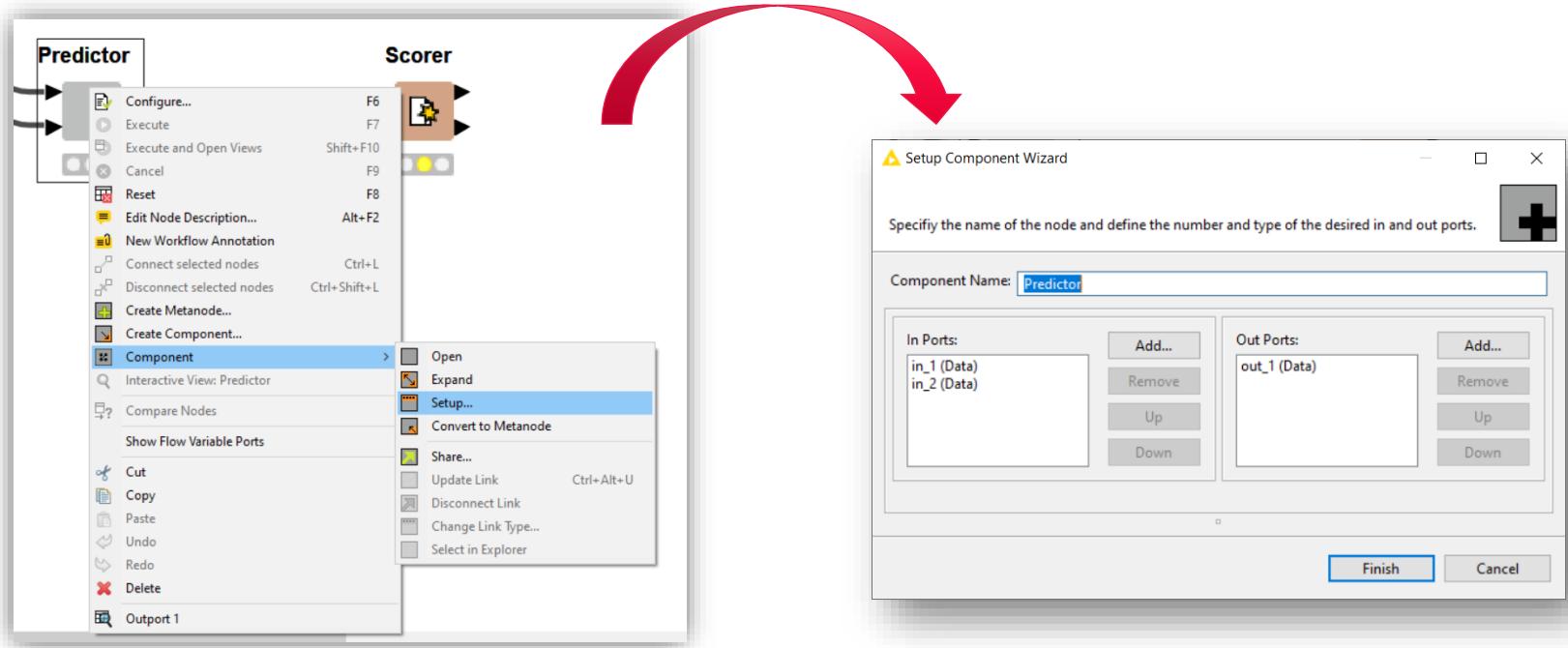
- Select related nodes that you want to group
- Right click
- Select *Create component...* or *Create Metanode...*
- Give it a name

## Components have more sophisticated features:

- Encapsulate flow variables, i.e. the parameters only live inside the component
- Provide a **configuration window**: variables and parameters within the component can be edited by Right Click -> Configure...
- Build a **composite view**: Visualization inside the component can be grouped in a dashboard

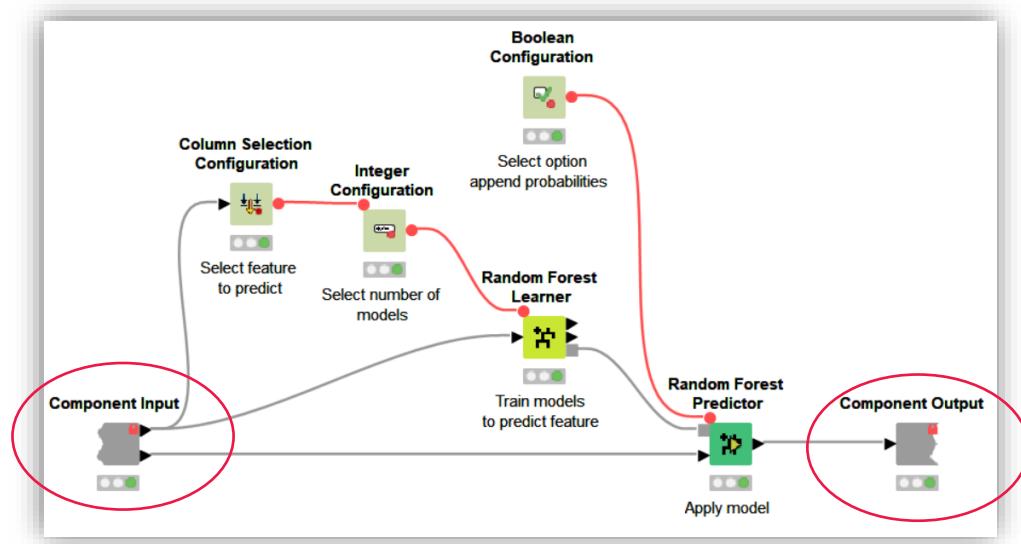
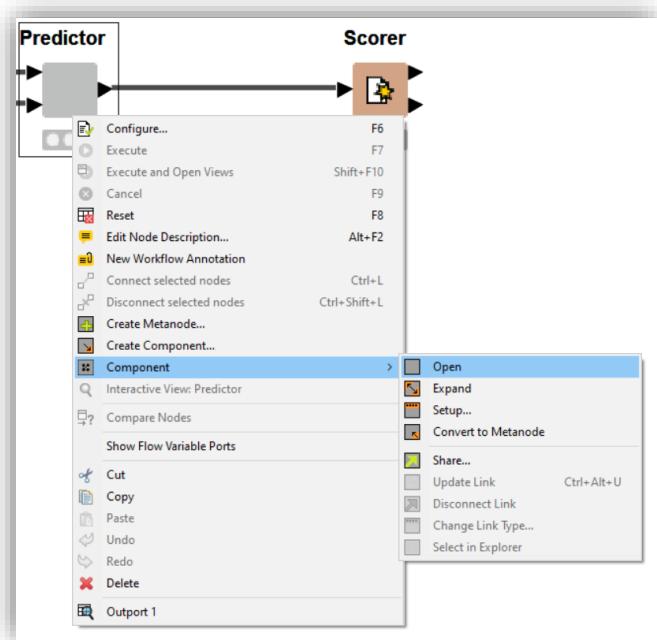


## Submenu Component



- Right click on the Component and select Setup... from the submenu Component to access further customization settings, such as the component name and the ports

# Inside a component

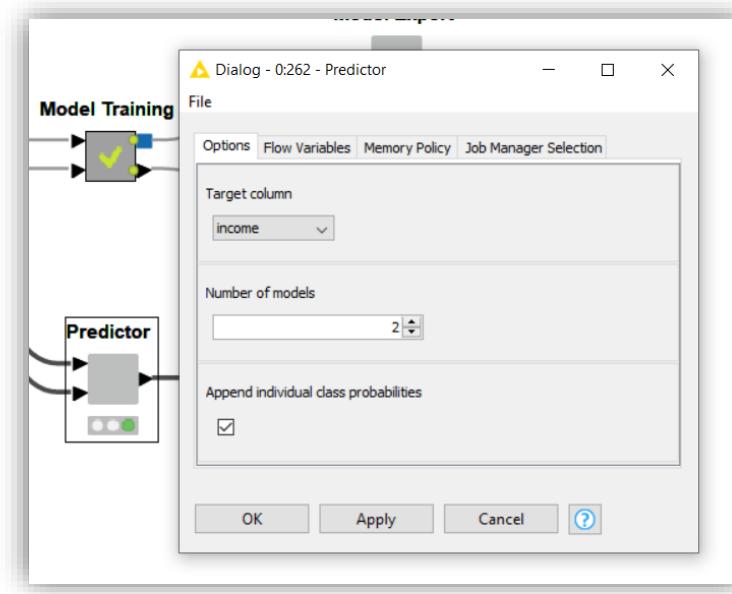


Shortcut:

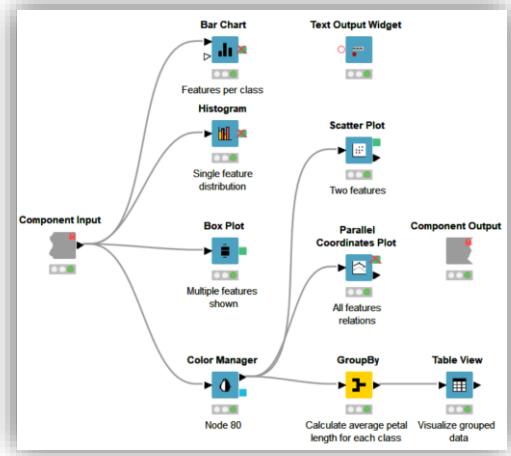
**Ctrl + double click** on component to open its content

## Components Configuration Window

- Components can be configurable
- From the configuration window (Right click -> Configure...) the user can enter some parameters
- The entered parameters change the behaviour of the nodes inside the component

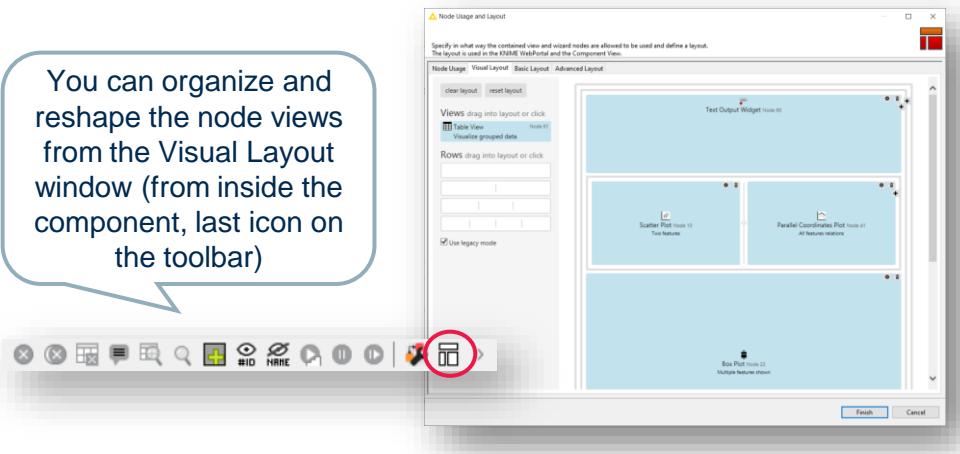


# Components Composite View

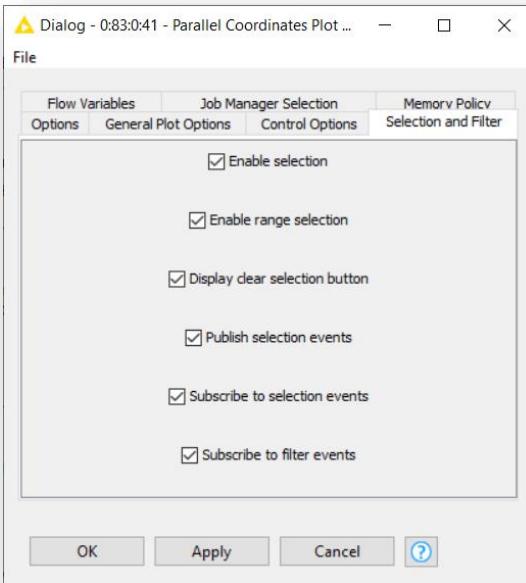


The visualization nodes within the component can be organized to build an interactive composite view

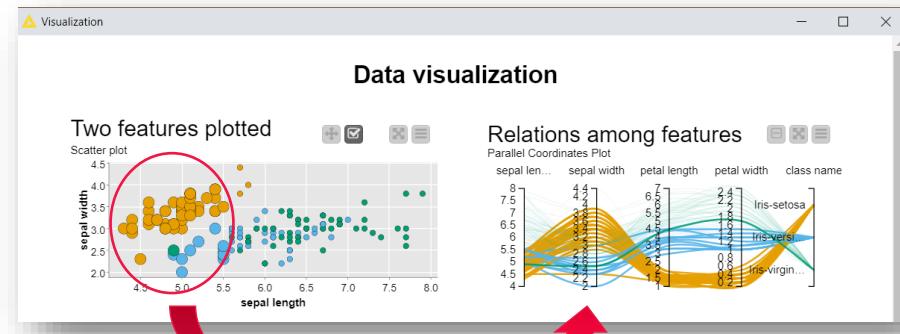
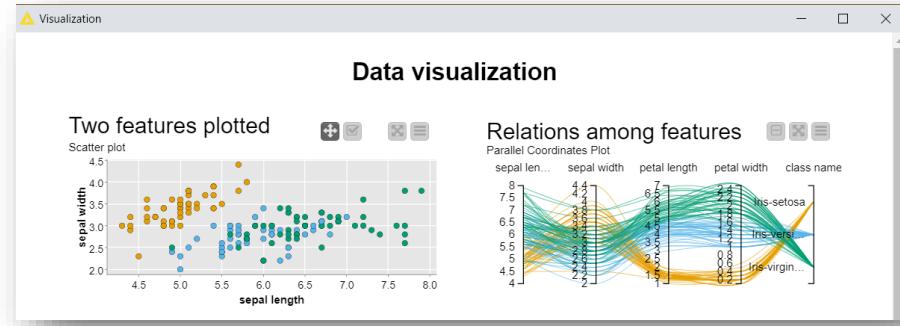
You can organize and reshape the node views from the Visual Layout window (from inside the component, last icon on the toolbar)



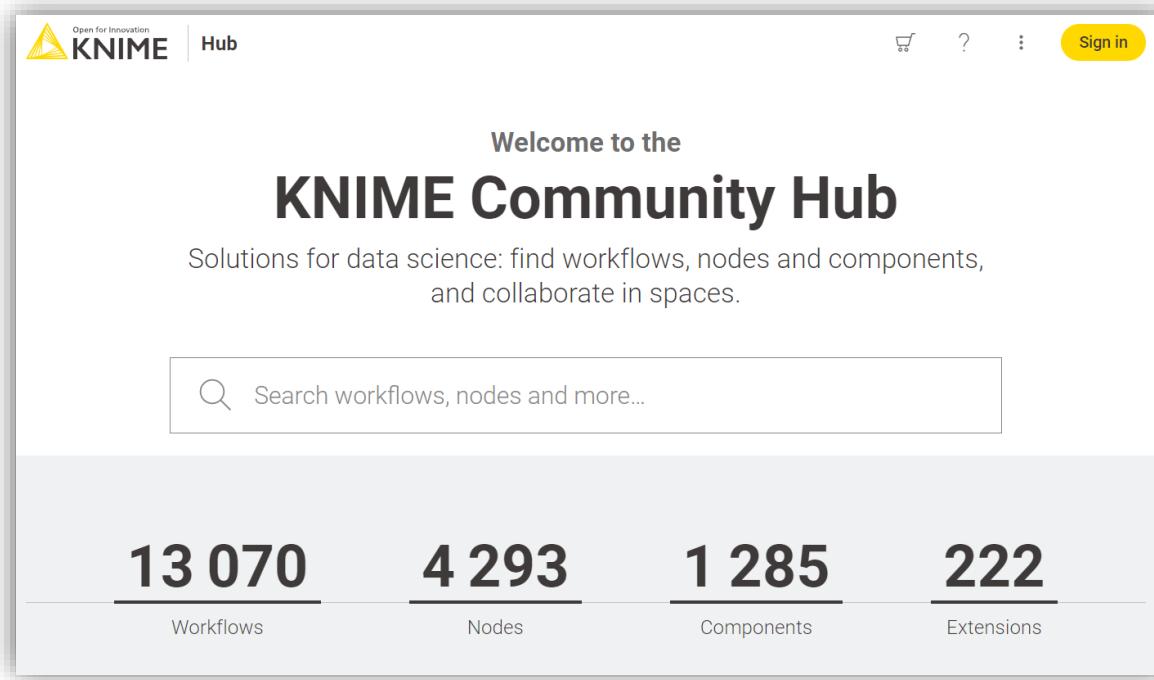
# Composite views interactivity



Enable publication and subscription to selection events to make the composite view interactive: data selected in one view are highlighted in the others



# KNIME Community Hub



The screenshot shows the KNIME Community Hub homepage. At the top left is the KNIME logo with the tagline "Open for Innovation". To the right are navigation links for "Hub", "Sign in", and other user icons. The main title "Welcome to the KNIME Community Hub" is centered above a subtitle "Solutions for data science: find workflows, nodes and components, and collaborate in spaces." Below this is a search bar with the placeholder "Search workflows, nodes and more...". At the bottom, there's a summary of statistics: 13 070 Workflows, 4 293 Nodes, 1 285 Components, and 222 Extensions.

Workflows	Nodes	Components	Extensions
13 070	4 293	1 285	222

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

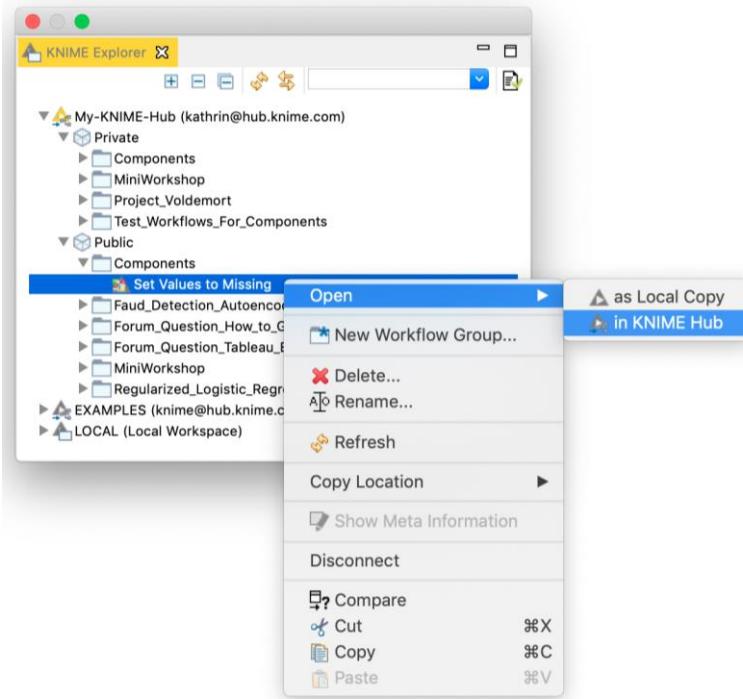
# KNIME Community Hub

The screenshot shows the KNIME Community Hub interface. At the top, there's a search bar with placeholder text "Search workflows, nodes and more..." and a "Sign In" button. Below the header, the URL "KNIME Hub > rs > Spaces > Data Science Guide > Workflows > Chapter8 > 01\_DcisionTree" is visible. The main content area displays a workflow titled "Decision Tree". The workflow starts with a "CSV Reader" node, followed by a "Color Manager" node, and then a "Partitioning" node. The "Partitioning" node has two outputs: one for the training set and one for the test set. The training set is processed by a "Decision Tree Learner" node, which then connects to a "Decision Tree View" node. The test set is processed by a "Decision Tree Predictor" node, which then connects to a "Scorer (JavaScript)" node. Finally, the "Scorer (JavaScript)" node connects to a "Classification performance" node. A yellow callout box highlights the "Decision Tree View" node. Below the workflow, a descriptive text explains the purpose of the workflow: "Using the adult dataset, this workflow performs binary classification (income > or < 50K) using a Decision Tree. The target is the income column, either <=50K or >50K, predicted using the other demographic attributes. After partitioning the original dataset into training set and test, the decision tree is built on the training set and the final performance is evaluated on the test set using the Scorer node."

## Workflows

The screenshot shows a node detail page for the "Scatter Plot" node. At the top, there's a "Node / Visualizer" icon and the node name "Scatter Plot". Below the title, there's a "Short link" button with the URL "https://knime.me/w/Pv9nGZquuMh...". The main content area contains a large image of a scatter plot. To the right of the image, a text block describes the node: "A scatter plot using a JavaScript based charting library. The view can be accessed either via the 'interactive view' action on the executed node or in KNIME Server web portal page." Below this, another text block explains the configuration: "The configuration of the node lets you choose the size of a sample to display and to enable certain controls, which are then available in the view. This includes the ability to choose different columns for x and y or the possibility to set a title. Enabling or disabling these controls via the configuration dialog might not seem useful at first glance but has benefits when used in a web portal/wizard execution where the end user has no access to the workflow itself." Further down, a note states: "Since missing values as well as NaN (not a number) or infinite values cannot be displayed in the view, they will be omitted with a corresponding warning message." Another note mentions: "Additionally a static SVG image can be rendered, which is then made available at the first output port." At the bottom, a note indicates: "Note, this node is currently under development. Future versions of the node might have more or changed functionality." On the right side of the page, there's a section for "Extension" with the text "This node is part of the extension KNIME JavaScript Views Version 4.2.1" and a "Short link" button with the URL "https://knime.me/n/HnJ14W8MjyH12k".

## Nodes, Shared Components and Extensions



- **Private Space**
- Your personal space. Upload here your workflows and components (max 1GB) to have them always available in a central place

- **Public Space**
- Shared with the KNIME community. Everyone can find and download them from the KNIME Community Hub

# Downloading and importing a workflow from the KNIME Community Hub

The screenshot shows the KNIME Community Hub search interface. A red oval highlights the search bar at the top, which contains the text "theguidebook". Below the search bar, the text "70 results" is displayed. Under the search bar, there are filter options: GIDS, TheGuideBook, Academia, Exercise, Classification, Education, Machine learning, Data mining, and Clustering. The "Workflows" tab is selected. Two workflow cards are visible:

- Workflow Introduction**:  
GIDS | Exercise | Academia | +1  
Exercise to perform basic operations in KNIME: - Read data - Filter rows - Filter columns - Write and plot data  
knime > Academic Alliance > Guide to Intelligent Data Science > Exercises > Chapter1\_Introduction > Introduction\_Solution
- Workflow Introduction**:  
GIDS | Exercise | Academia | +1  
Exercise to perform basic operations in KNIME: - Read data - Filter rows - Filter columns - Write and plot data  
knime > Academic Alliance > Guide to Intelligent Data Science > Exercises > Chapter1\_Introduction > Introduction\_Exercise

Searching for the Tag  
***theguidebook*** will show you all  
the workflows related to this book

# Downloading and importing a workflow from the KNIME Community Hub

Workflow  
Deployment to a Dashboard

Last Update: 24 Feb 2020

Prediction dashboard. The visualization component shows predictor accuracy and data insights

Model Reader  
Load trained model

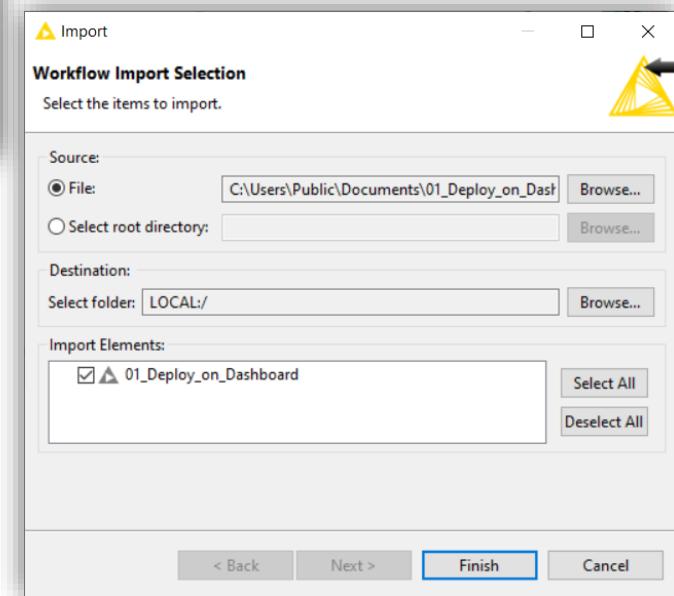
Data upload

Decision Tree Predictor  
Predict churn

Visualization

This deployment workflow builds a simple dashboard. Reads a pre-trained decision tree model and applies it to new data to predict customer churn. Predictions are then displayed onto a dashboard.  
When run from a KNIME WebPortal, it can be called and executed step by step from a web browser. First page asks for the data upload. Second page shows the interactive dashboard where the user can drill in into the data and the predictions..

**Method 1**  
Download the workflow, locate it into your machine and import it as seen before



# Downloading and importing a workflow from the KNIME Community Hub

Workflow  
**Deployment to a Dashboard**

TheGuidedBook deployment dashboard GIDS

Last Update: 24 Feb 2020

Prediction dashboard. The visualization component shows predictor accuracy and data insights

**Model Reader**  
Load trained model

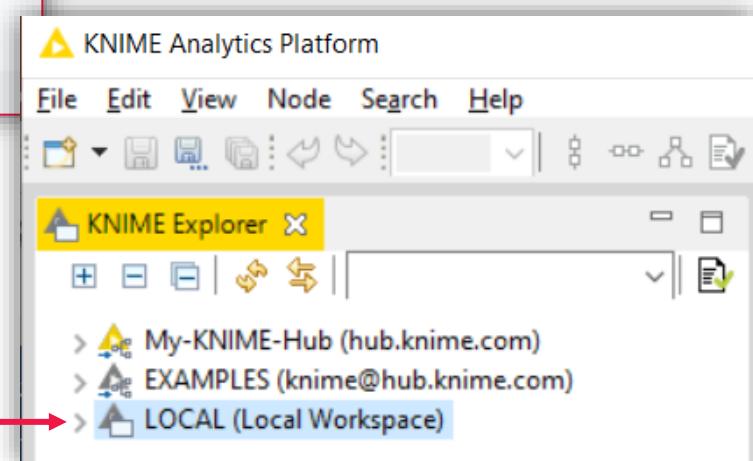
**Data upload**

Decision Tree Predictor  
Predict churn

Visualization

This deployment workflow builds a simple dashboard. Reads a pre-trained decision tree model and applies it to new data to predict customer churn. Predictions are then displayed onto a dashboard. When run from a KNIME WebPortal, it can be called and executed step by step from a web browser. First page asks for the data upload. Second page shows the interactive dashboard where the user can drill in into the data and the predictions..

**Method 2**  
Drag and drop the  icon directly into the KNIME Explorer at the desired location



# KNIME Cheat Sheets

<https://www.knime.com/cheat-sheets>

This cheat sheet provides an overview of machine learning components in KNIME. It is divided into three main sections: CLASSIFICATION, REGRESSION, and CLUSTERING. Each section contains several nodes with their descriptions and icons.

- CLASSIFICATION:** Includes nodes like Logistic Regression, Decision Tree, Random Forest, Naive Bayes, and K-Means.
- REGRESSION:** Includes nodes like Linear Regression, Ridge Regression, Lasso Regression, and Support Vector Regression.
- CLUSTERING:** Includes nodes like K-Means, DBSCAN, and Hierarchical Clustering.

This cheat sheet covers control and orchestration in KNIME. It includes sections on TRAINING and CONTROL.

- TRAINING:** Shows nodes for creating training datasets from various sources like CSV, databases, and APIs.
- CONTROL:** Shows nodes for managing loops and switches in workflows.

This cheat sheet is designed for beginners. It shows how to build a workflow for traffic light management. It includes sections on ORCHESTRATION and READ.

- ORCHESTRATION:** Shows nodes for managing workspaces and services.
- READ:** Shows nodes for reading data from various sources like CSV, MySQL, and REST services.

This cheat sheet focuses on data transformation. It includes sections on TRANSFORM and PUBLISH.

- TRANSFORM:** Shows nodes for manipulating data structures like Joiner, Splitter, and Filter.
- PUBLISH:** Shows nodes for publishing transformed data to various formats like Excel, CSV, and Tableau.

This section continues the workflow for traffic light management, showing how to use loops and switches. It also includes sections on EXPLORE and ANALYZE.

- EXPLORE:** Shows nodes for creating charts like Scatter Plot, Box Plot, and Bar Chart.
- ANALYZE:** Shows nodes for machine learning like Learner Node, Predictor Node, and Logistic Regression.

This section continues the data transformation process, showing how to use various mathematical and string operators. It includes sections on READ, EXPLORE, ANALYZE, and PUBLISH.

- READ:** Shows nodes for reading data from Google Sheets and CSV files.
- EXPLORE:** Shows nodes for exploring data with Learner and Predictor nodes.
- ANALYZE:** Shows nodes for performing logistic regression and other statistical analyses.
- PUBLISH:** Shows nodes for publishing results to various platforms like Tableau and Python.

This section continues the data transformation process, showing how to use various mathematical and string operators. It includes sections on TRANSFORM and PUBLISH.

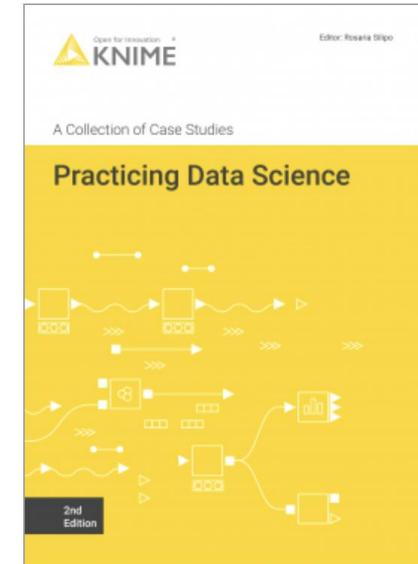
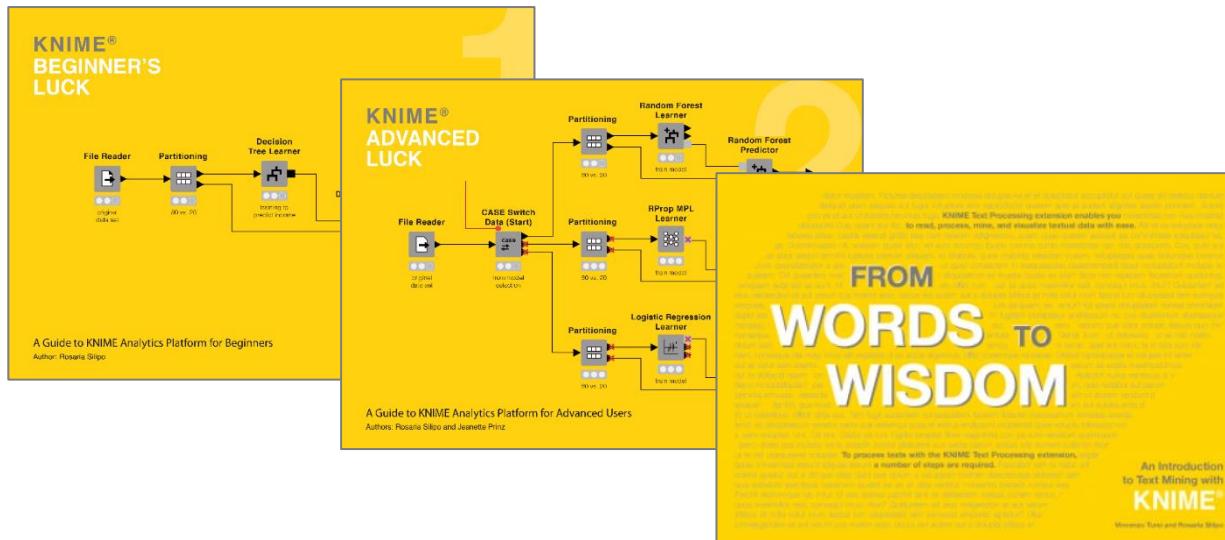
- TRANSFORM:** Shows nodes for complex transformations like Math Formula and Column Rename.
- PUBLISH:** Shows nodes for publishing results to various platforms like Excel, CSV, and Tableau.

This section concludes the data transformation process, showing how to use various mathematical and string operators. It includes sections on PUBLISH and Resources.

- PUBLISH:** Shows nodes for publishing results to various platforms like Excel, CSV, and Tableau.
- Resources:** Provides links to the KNIME community forum, blog, and documentation.

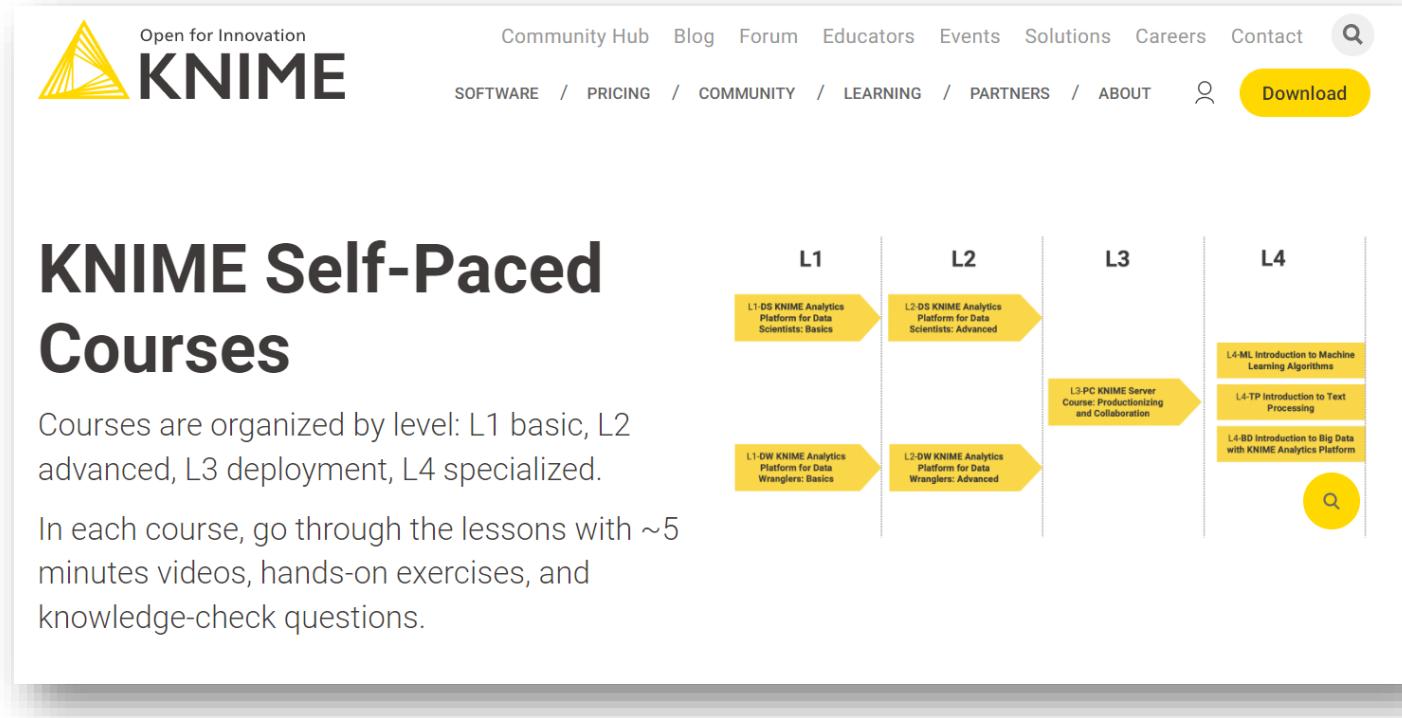
e-book downloads from **KNIME Press**  
<https://www.knime.com/knimepress>

with code: <Promotion-Code>



# Free e-Learning Courses about KNIME Analytics Platform

<https://www.knime.com/knime-self-paced-courses>



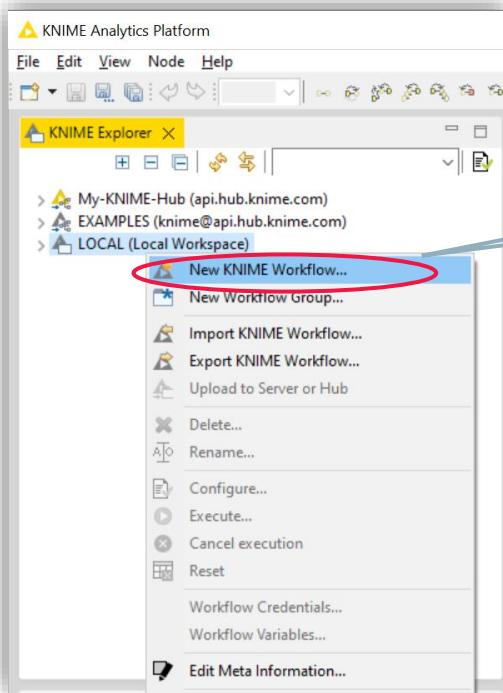
The screenshot shows the KNIME website's self-paced courses section. At the top, the KNIME logo is displayed with the tagline "Open for Innovation". The navigation bar includes links for Community Hub, Blog, Forum, Educators, Events, Solutions, Careers, Contact, a search icon, and a "Download" button. Below the navigation, the main heading "KNIME Self-Paced Courses" is prominently displayed. A descriptive text explains that courses are organized by level: L1 basic, L2 advanced, L3 deployment, and L4 specialized. To the right, a diagram illustrates the course structure across four levels:

- L1:** L1-DS KNIME Analytics Platform for Data Scientists: Basics; L1-DW KNIME Analytics Platform for Data Wranglers: Basics
- L2:** L2-DS KNIME Analytics Platform for Data Scientists: Advanced; L2-DW KNIME Analytics Platform for Data Wranglers: Advanced
- L3:** L3-PC KNIME Server Course: Productionizing and Collaboration
- L4:** L4-ML Introduction to Machine Learning Algorithms; L4-TP Introduction to Text Processing; L4-BD Introduction to Big Data with KNIME Analytics Platform

A magnifying glass icon is located at the bottom right of the diagram area.

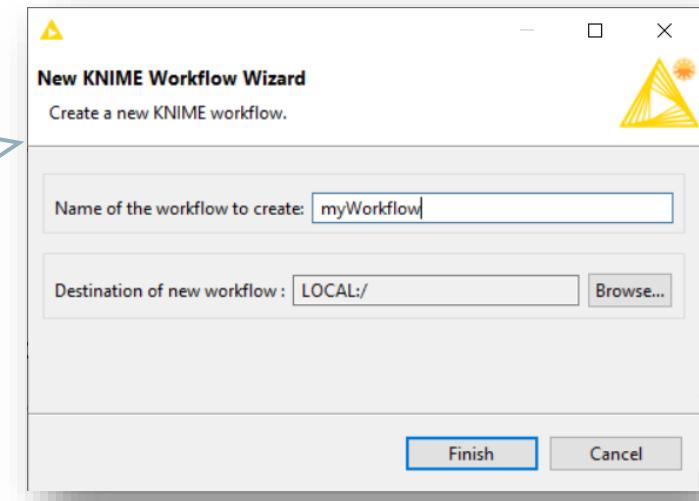
# Build your first Hello Workflow

# Create your first workflow



Right Click on the LOCAL folder in the KNIME Explorer and select New KNIME Workflow

From the pop up window, insert the name of your first workflow



# Read the dataset

The screenshot shows the KNIME interface. On the left, the Node Repository is open, displaying a 'csv reader' node under the 'Read' category. A callout bubble points from this node to the workflow area, instructing to 'Drag and drop the CSV Reader node from the Node Repository to add it to the workflow'. In the center, a 'CSV Reader' node is placed in the workflow, with a callout bubble pointing to its configuration window. The configuration window title is 'Dialog - 6:1 - CSV Reader'. It shows the 'Input location' tab selected, with 'Read from' set to 'Relative to' and 'Current workflow'. The 'File' field contains the path '.../Example Workflows/TheData/Basics/adult.csv'. Below this, the 'Reader options' section includes fields for 'Format' (Autodetect format), 'Column delimiter' (set to comma), 'Row delimiter' (set to line break), 'Quote char' (empty), 'Comment char' (empty), and checkboxes for 'Has column header' (checked) and 'Support short data rows' (unchecked). The 'Preview' section shows the first 15 rows of the 'adult.csv' dataset. The preview table has 15 columns and 15 rows, with the first few rows visible:

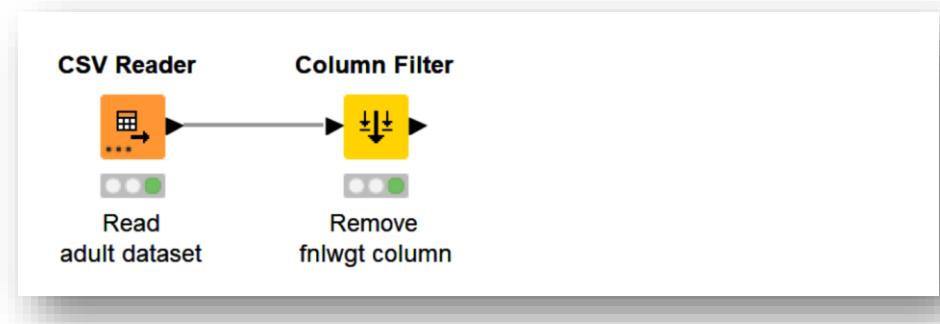
Row ID	age	workclass	fnwght	education	educatn	marital	occupa	relation	rac
Row0	39	State-gov	77516	Bachelors	13	Never-married	Adm-clerical	Not-in-family	White
Row1	50	Self-emp-no... Self-emp-no...	83311	Bachelors	13	Married-civ...	Exec-manag...	Husband	White
Row2	38	Private	215646	HS-grad	9	Divorced	Handlers-de...	Not-in-family	White
Row3	53	Private	234721	11th	7	Married-civ...	Handlers-de...	Husband	Black
Row4	28	Private	338409	Bachelors	13	Married-civ...	Prof-specialty	Wife	Black
Row5	37	Private	284582	Masters	14	Married-civ...	Exec-manag...	Wife	White
Row6	49	Private	160187	9th	5	Married-spo...	Other-service	Not-in-family	Black
Row7	52	Self-emp-no... Self-emp-no...	209642	HS-grad	9	Married-civ...	Exec-manag...	Husband	White
Row8	31	Private	45781	Masters	14	Never-married	Prof-specialty	Not-in-family	White
Row9	42	Private	159449	Bachelors	13	Married-civ...	Exec-manag...	Husband	White
Row10	37	Private	280464	Some-college	10	Married-civ...	Exec-manag...	Husband	Black
Row11	30	State-gov	141297	Bachelors	13	Married-civ...	Prof-specialty	Husband	Asian-P.
Row12	23	Private	122272	Bachelors	13	Never-married	Adm-clerical	Own-child	White
Row13	32	Private	205019	Assoc-acdm	12	Never-married	Sales	Not-in-family	Black
Row14	40	Private	121772	Assoc-voc	11	Married-civ...	Craft-repair	Husband	Asian-P.
Row15	34	Private	245487	7th-8th	4	Married-civ...	Transmitt-m...	Husband	Amer-Isr

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

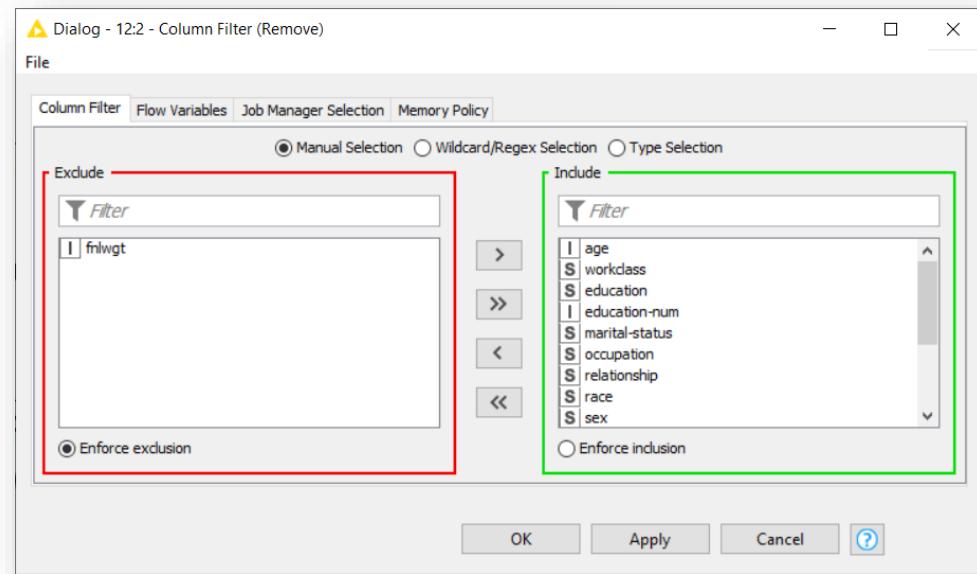
Drag and drop the **CSV Reader** node from the Node Repository to add it to the workflow

Open the configuration window (double click) and select the file on your machine containing the adult dataset

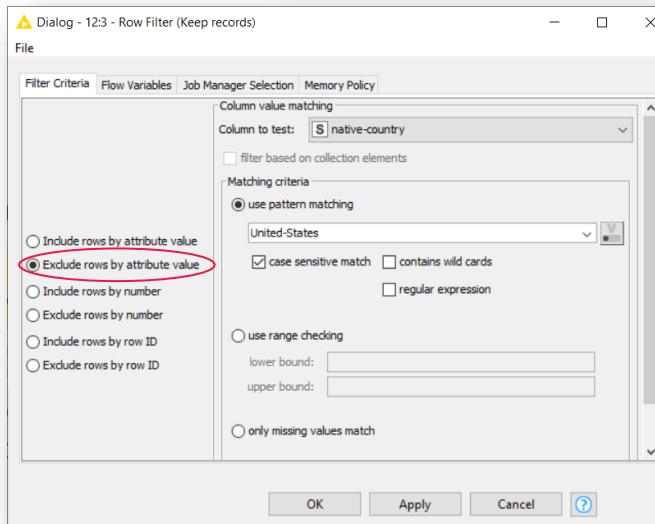
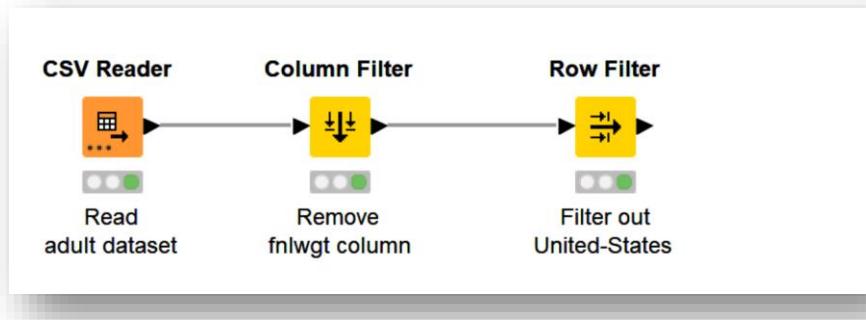
# Remove columns



Some columns have unnecessary information. Remove them with a Column Filter node

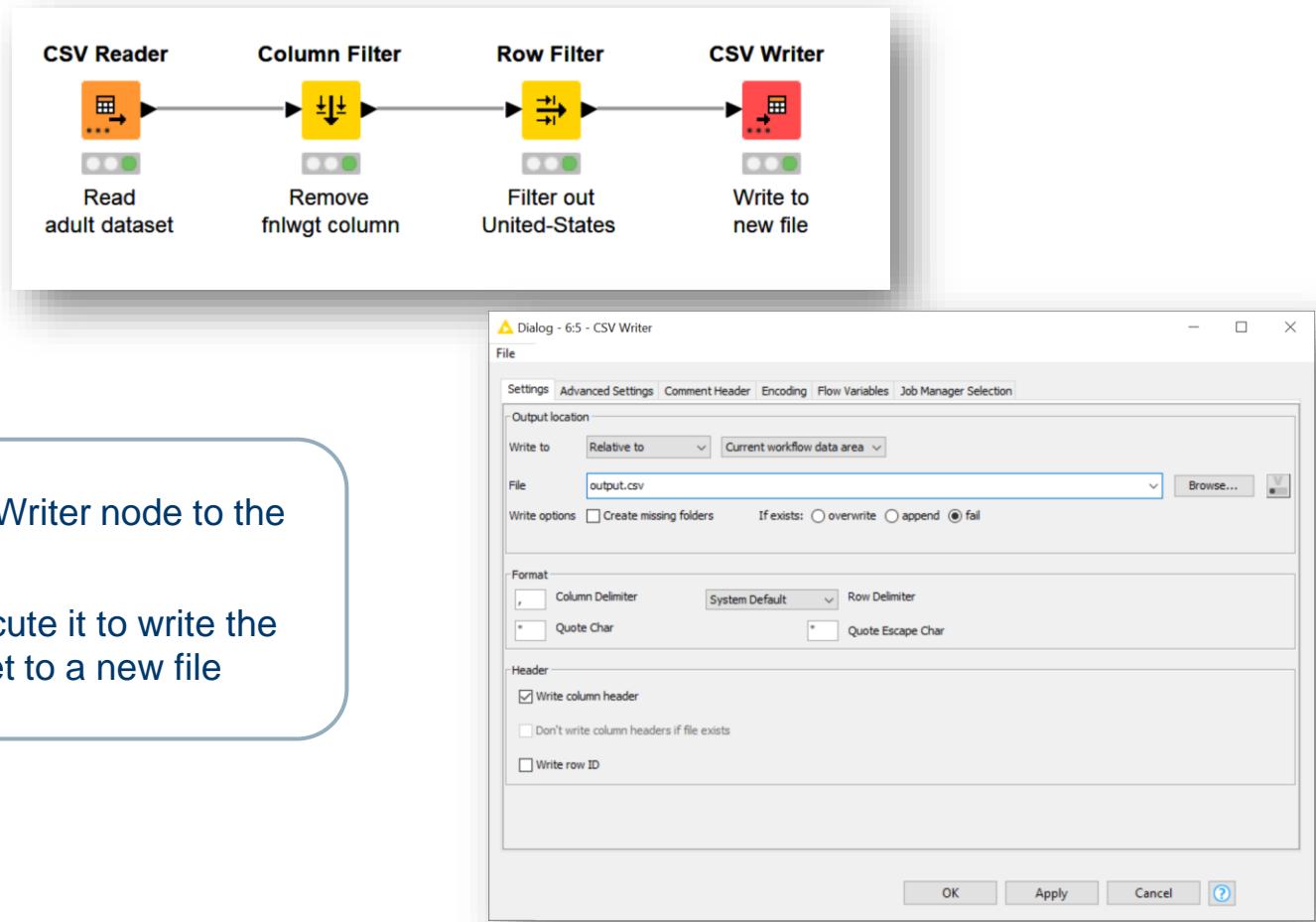


## Remove Rows



Add a Row Filter node and configure it to only keep entries whose “native-country” value is not “United-States”

## Write to new file

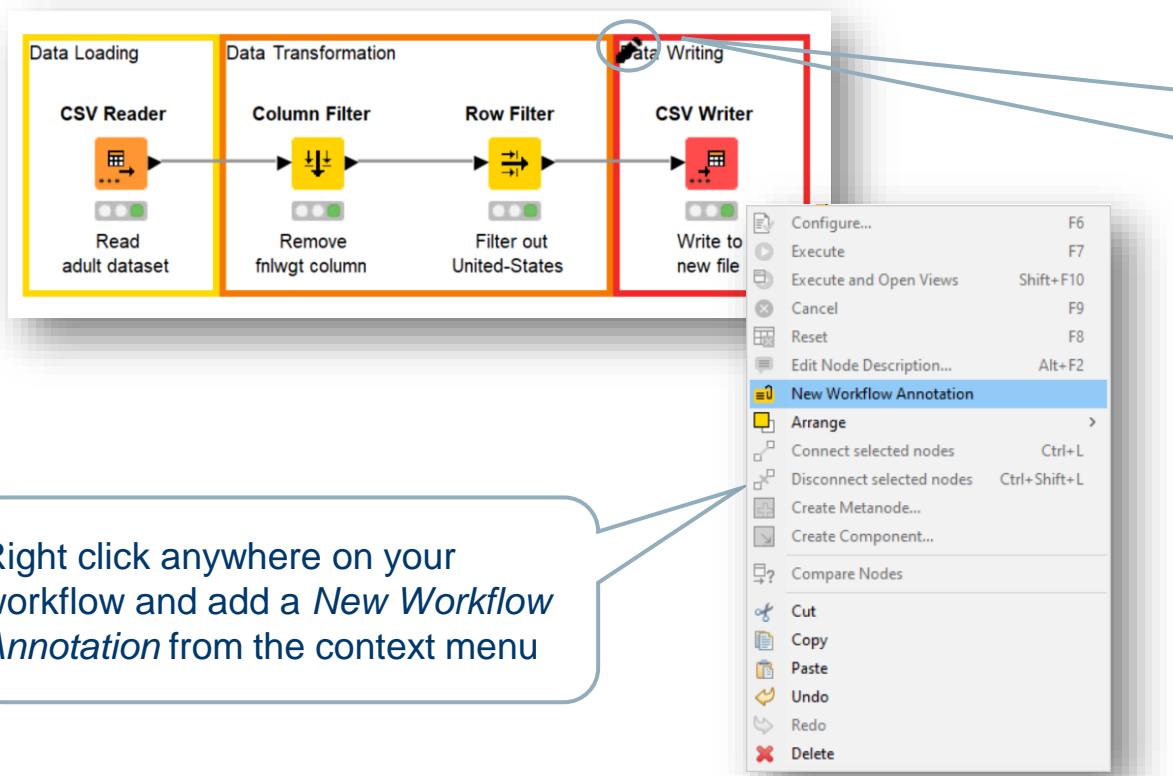


Finally add a CSV Writer node to the pipeline.

Configure and execute it to write the transformed dataset to a new file

## Annotations

- Annotations are coloured editable boxes that you can add to your workflow
- They help you making it more readable and visually pleasant



# Thank you