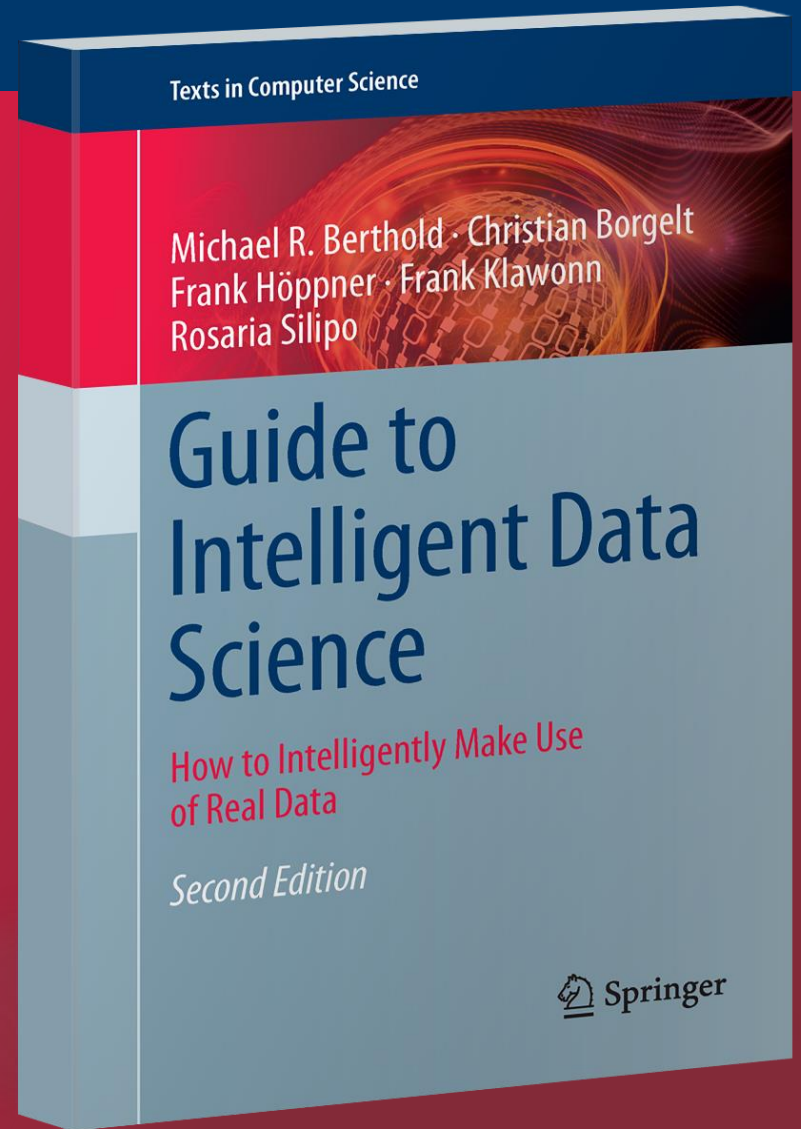


KNIME Analytics Platform



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

-Abraham Maslow

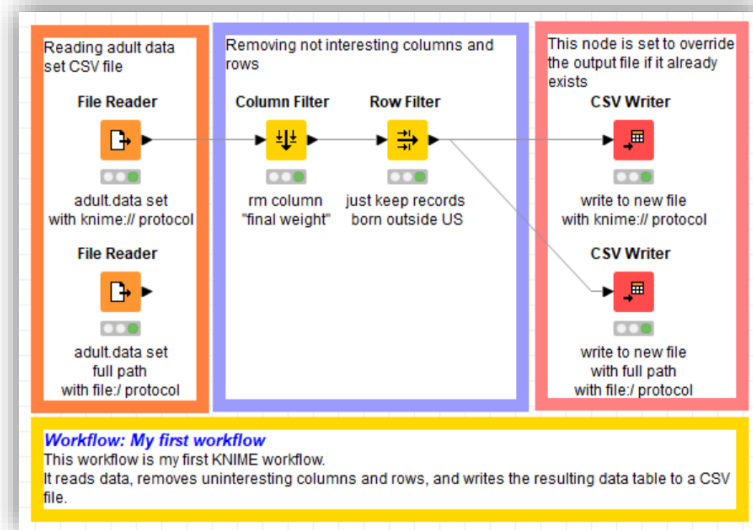
First the tool or the knowledge?

**This lesson refers to Appendix B of the GIDS book*

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

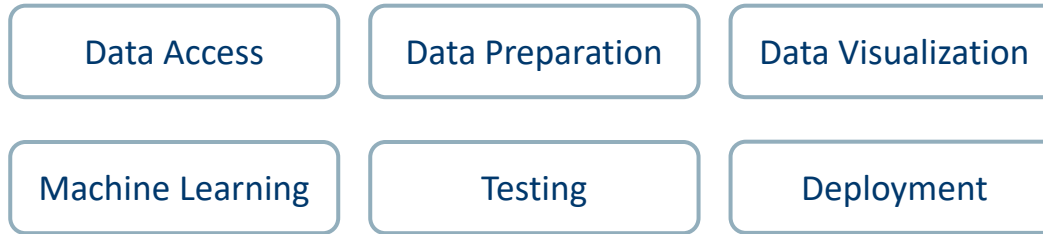
Datasets

- Dataset used : adult dataset
- Example Workflows:
 - „My First Workflow“ <https://kni.me/w/kYeZOLeAJXo9Mvol>
 - 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

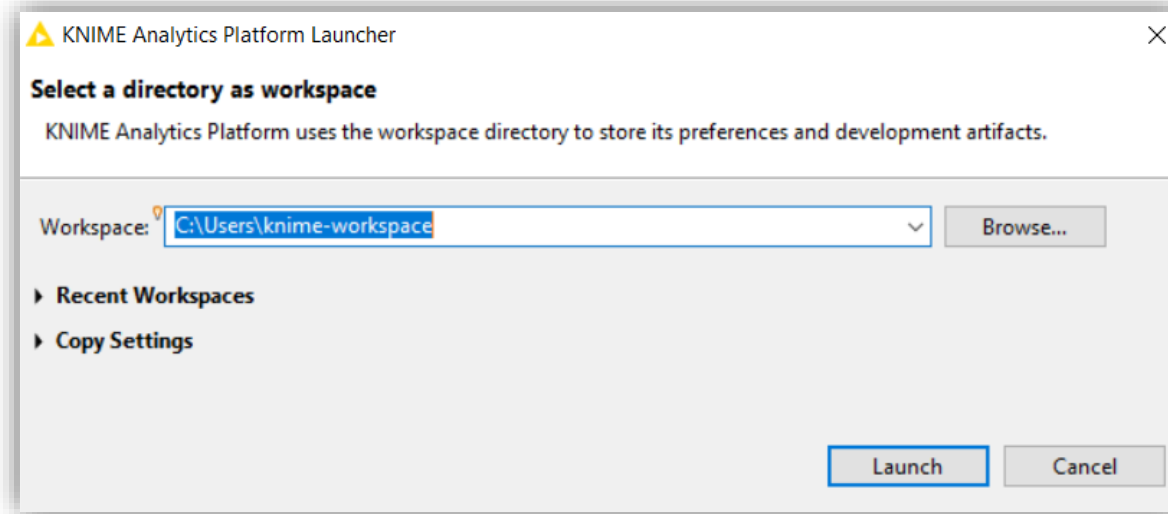
- To integrate the solutions into the IT environment
 - Scheduling
 - MLOps
 - Easy deployment
 - REST architecture
 - Auditing tools
- Closed source
- Yearly license

<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



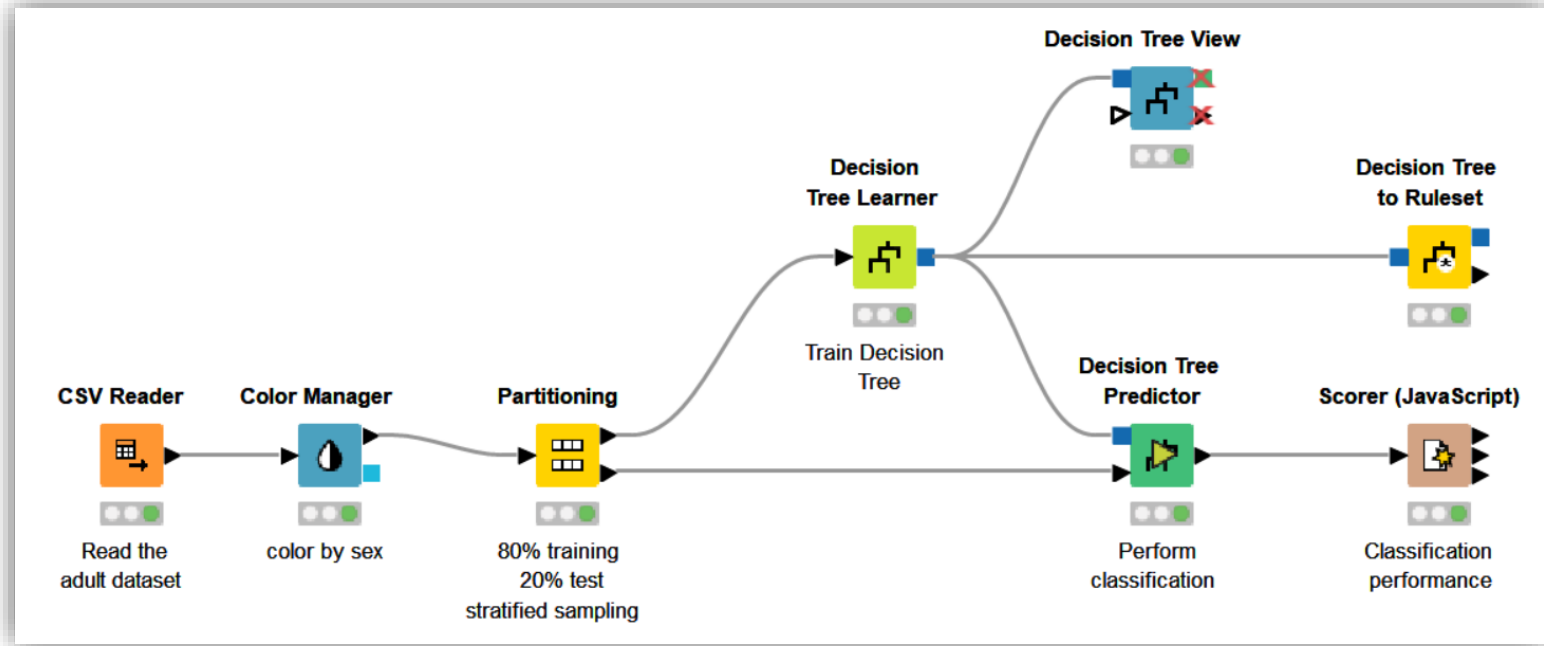
- The workspace is the **folder/directory** in which workflows (and potentially data files) are stored for the current session.
- Workspaces are portable (just like KNIME Analytics Platform)

The KNIME Workbench

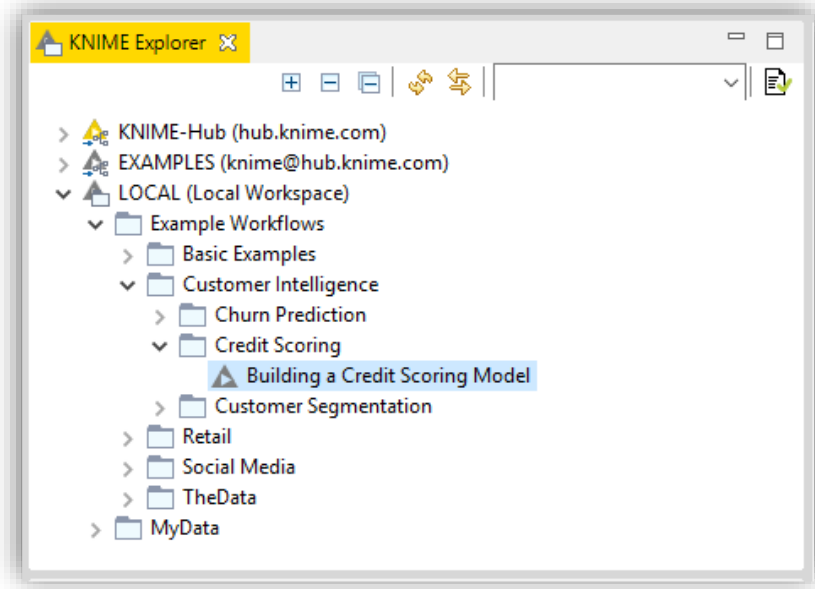
The screenshot displays the KNIME Analytics Platform interface. The main window, titled "My First Workflow", shows a workflow diagram with the following nodes: File Reader (labeled "Read sales data"), Column Filter (labeled "Include country, date and amount"), Row Filter (labeled "Exclude rows with country 'unknown'"), Stacked Area Chart (labeled "Sales over time"), and Pie/Donut Chart (labeled "Sales per country").



Yellow callout boxes identify several key components:

- KNIME Explorer:** Located in the top-left pane, showing a project tree with folders like "My_First_Project" and "data".
- Workflow Coach:** Located in the middle-left pane, displaying a list of "Recommended Nodes" such as Joiner, Column Filter, Row Filter, Partitioning, GroupBy, Missing Value, and Statistics.
- Node Repository:** Located in the bottom-left pane, showing categories like IO, Manipulation, Views, Analytics, DB, and Other Data Types.
- Workflow Editor:** The central workspace where the workflow is built.
- Description:** A panel on the right showing the description for the "File Reader" node.
- KNIME Hub Search:** A search bar at the bottom right for finding workflows.
- Outline:** A small thumbnail of the workflow in the bottom-left corner of the main editor.
- Console:** A panel at the bottom right showing the KNIME Console output, which includes a welcome message: "Welcome to KNIME Analytics Platform v4.0.0.v20190... Copyright by KNIME AG, Zurich, Switzerland".

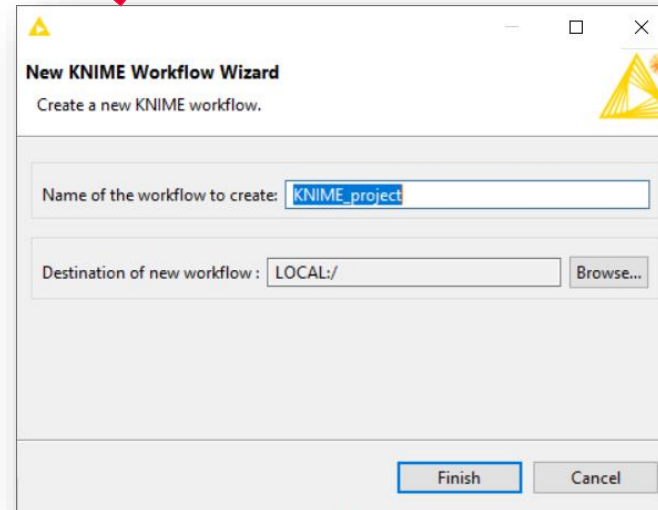
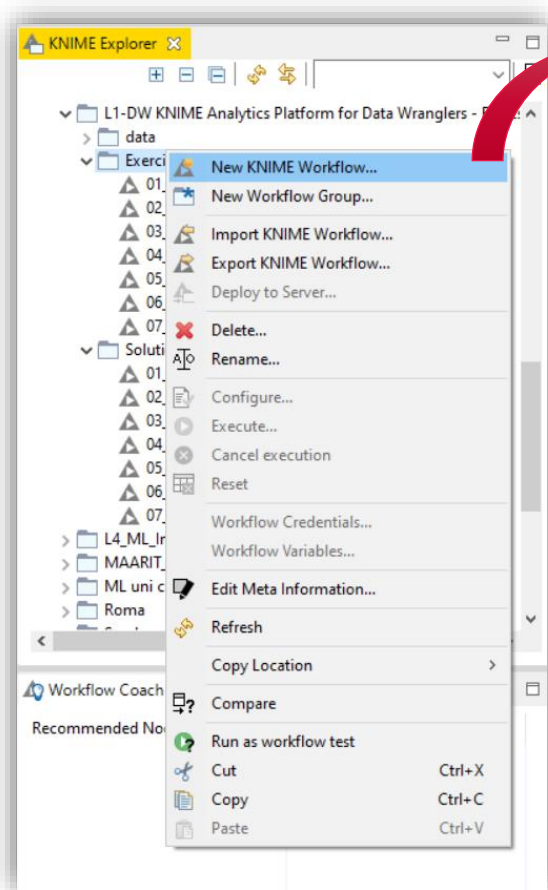


A workflow is a pipeline of nodes, each configurable to perform a specific task.
The data flow through nodes from left to right



- 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

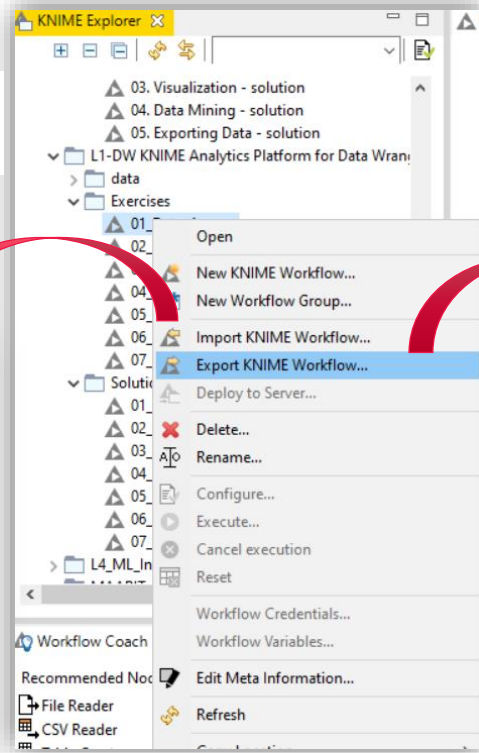
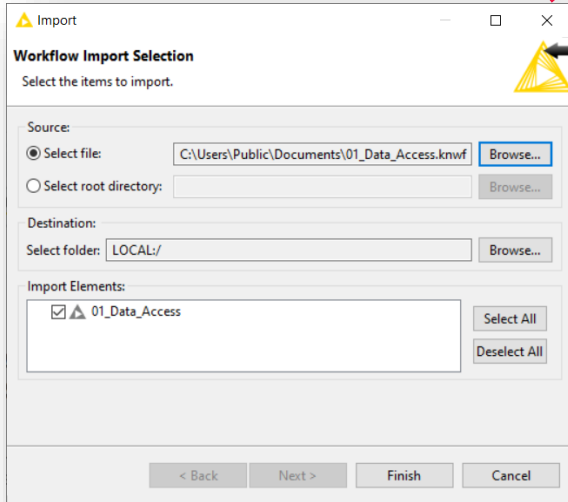
Creating a new workflow



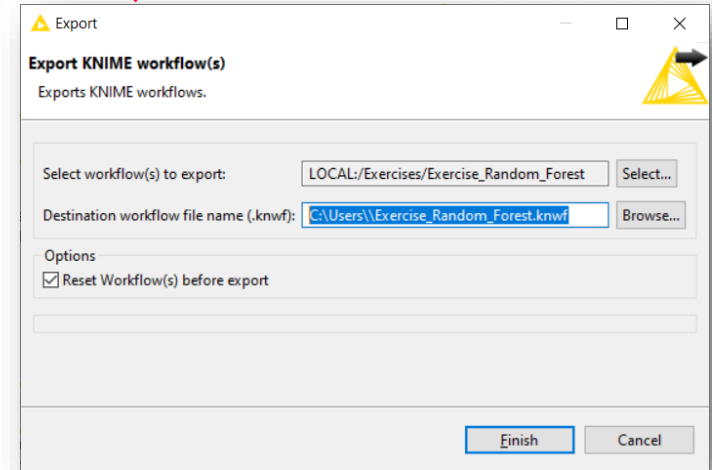
Click anywhere on the KNIME Explorer to create a new workflow or workflow group

Importing and Exporting Workflows

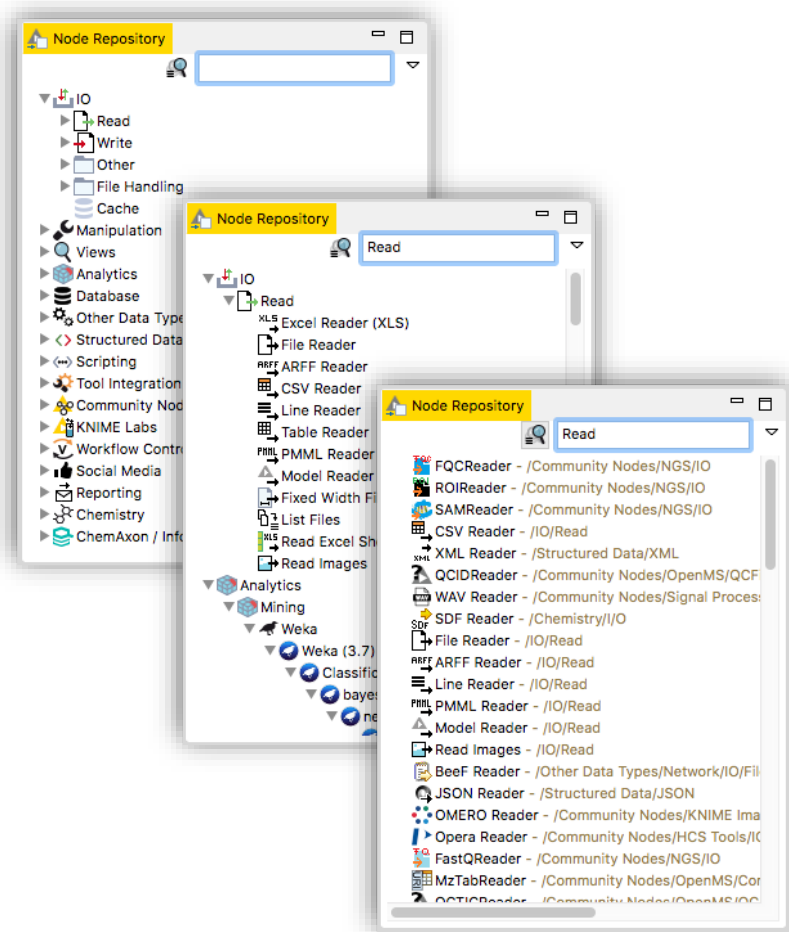
Right-click anywhere in KNIME Explorer to import a workflow





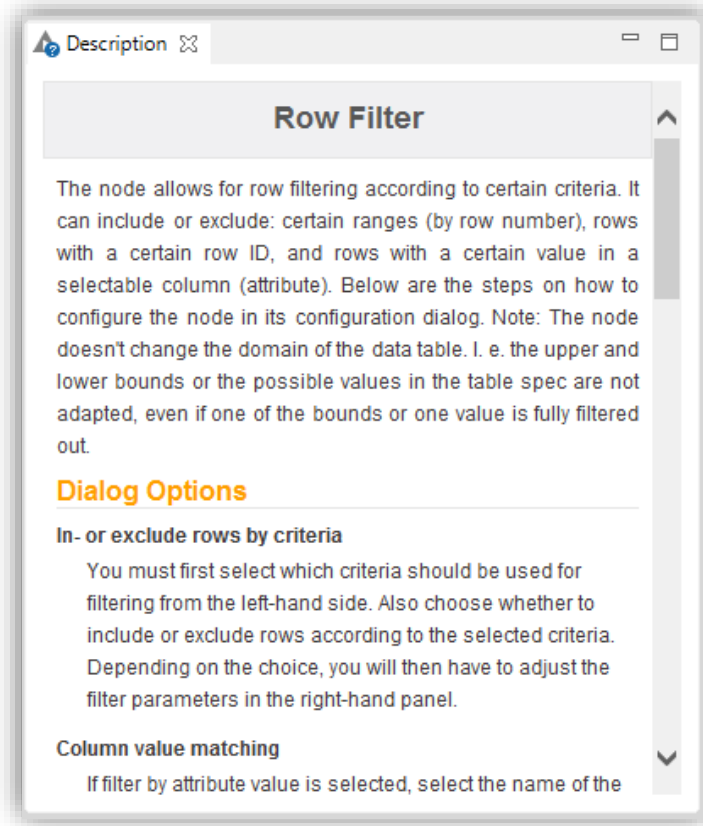
Right-click on a workflow or workflow group to export the selected workflow



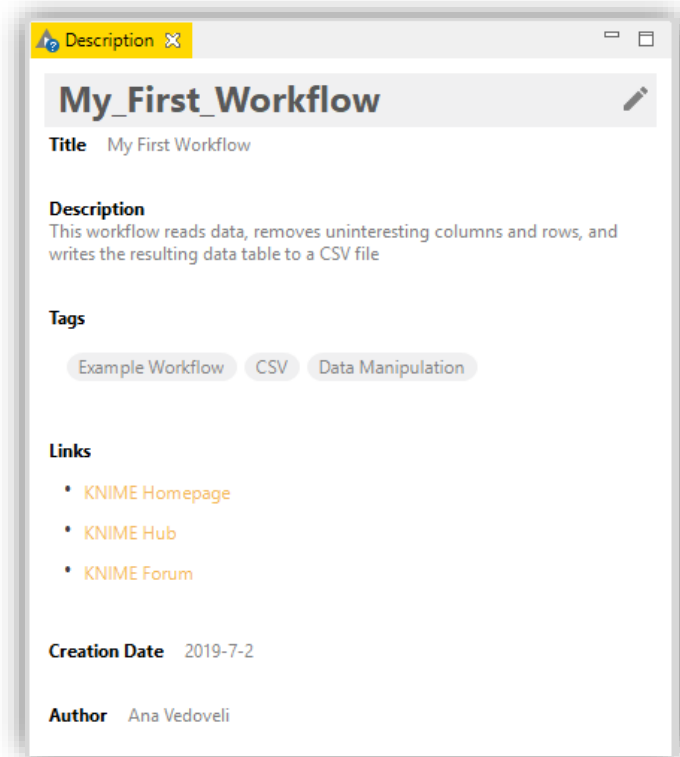
Node Repository



- 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

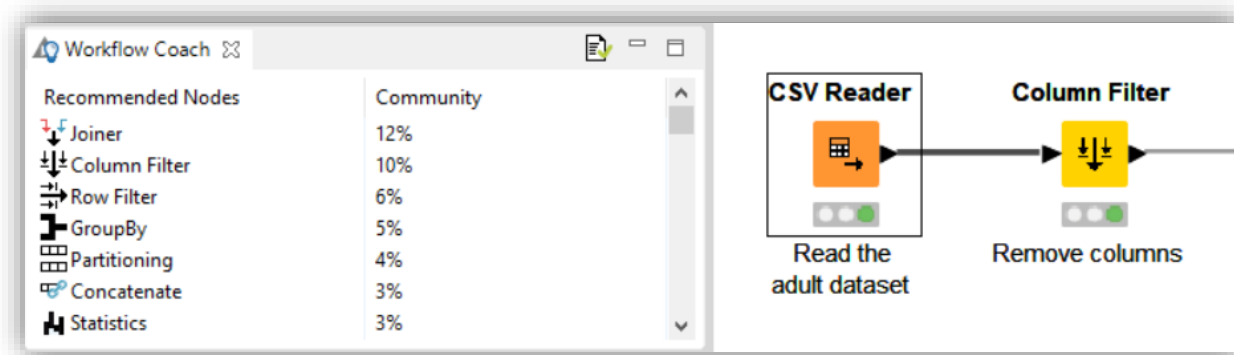


- The Description window gives information about:
 - Node Functionality
 - Input & Output
 - Node Settings
 - Ports
 - References to literature

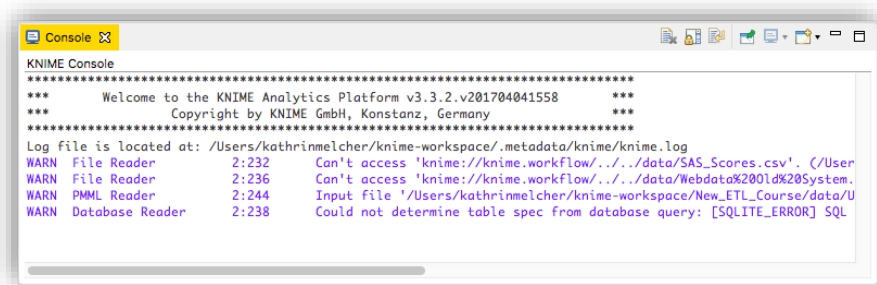


- When selecting the workflow, the Description window gives information about the workflow's:
 - Title
 - Description
 - Associated Tags and Links
 - Creation Date
 - Author

- 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

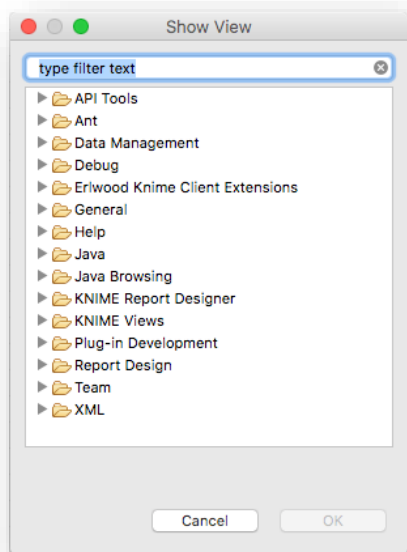
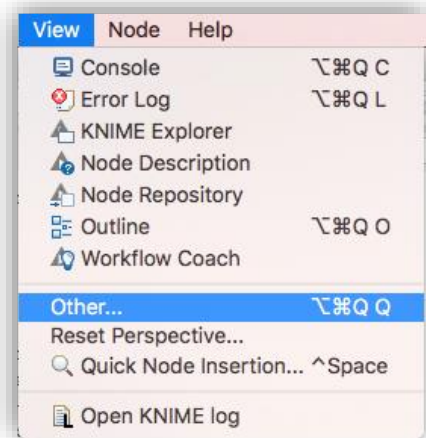


Console and Other views



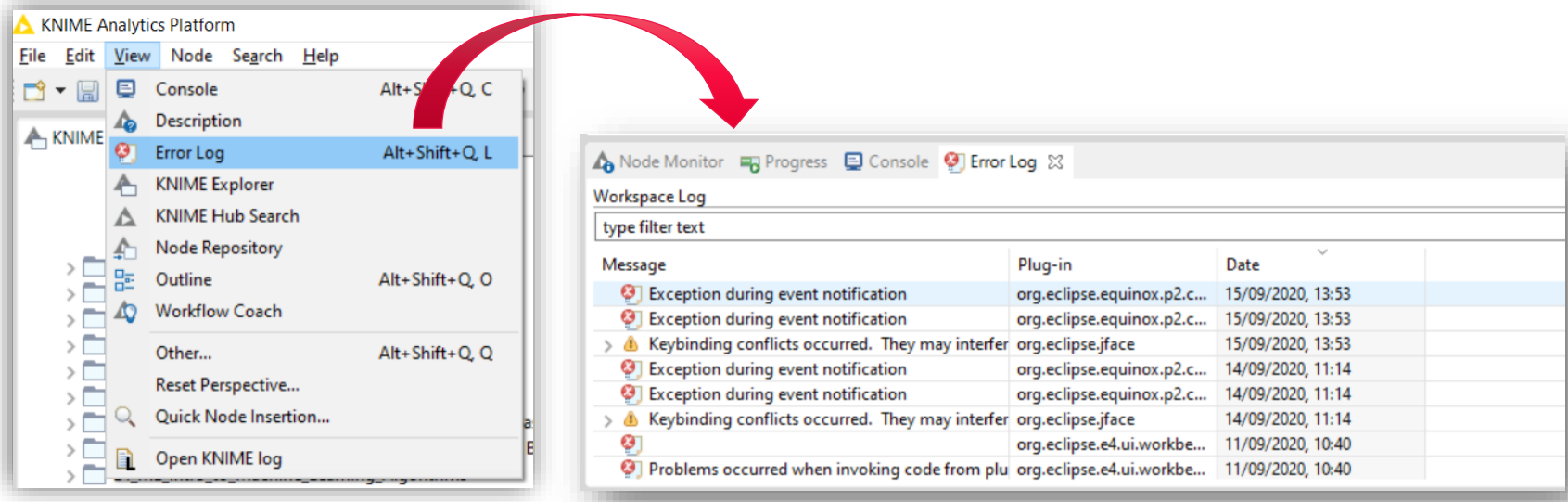
```
KNIME Console
*****
*** Welcome to the KNIME Analytics Platform v3.3.2.v201704041558 ***
*** Copyright by KNIME GmbH, Konstanz, Germany ***
*****
Log file is located at: /Users/kathrinmelcher/knime-workspace/.metadata/knime/knime.log
WARN File Reader 2:232 Can't access 'knime://knime.workflow/./././data/SAS_Scores.csv'. C:/User
WARN File Reader 2:236 Can't access 'knime://knime.workflow/./././data/Webdata%201d%20System.
WARN PMML Reader 2:244 Input file '/Users/kathrinmelcher/knime-workspace/New_ETL_Course/data/U
WARN Database Reader 2:238 Could not determine table spec from database query: [SQLITE_ERROR] SQL
```

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



- Click on View and select *Other...* to add additional views

Error Log View

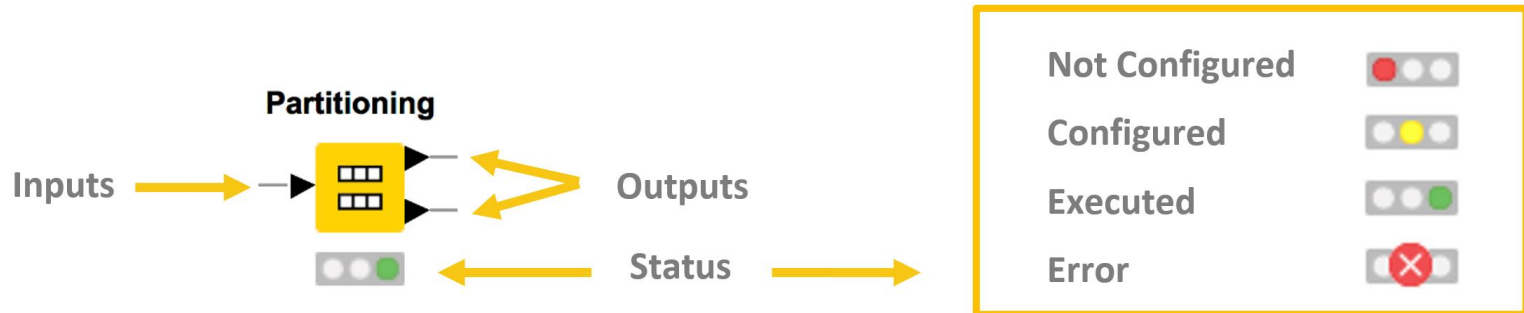


Tip: enabling and checking the Error Log view can help while debugging your project

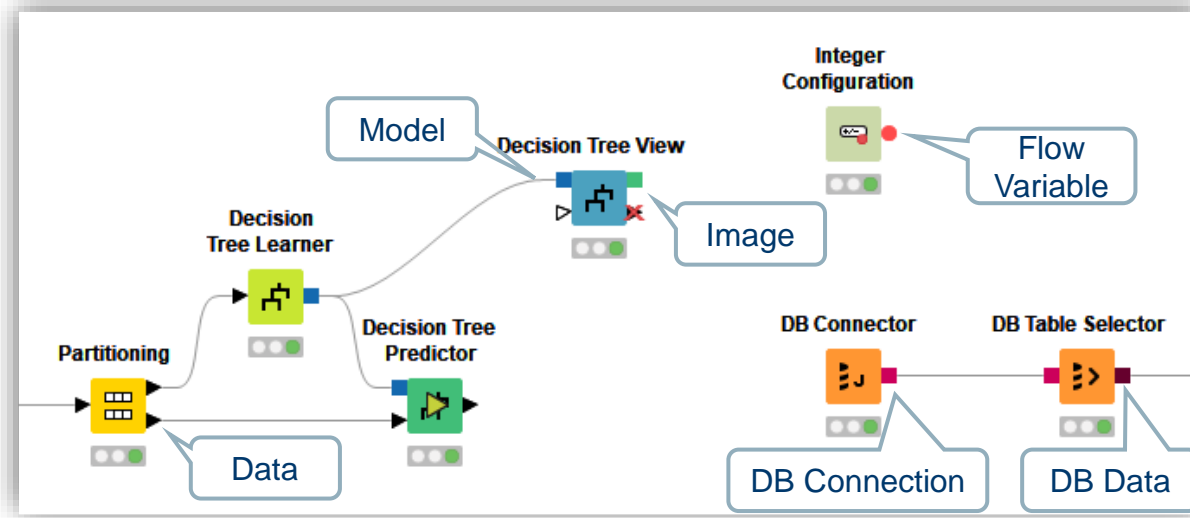
More on Nodes

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

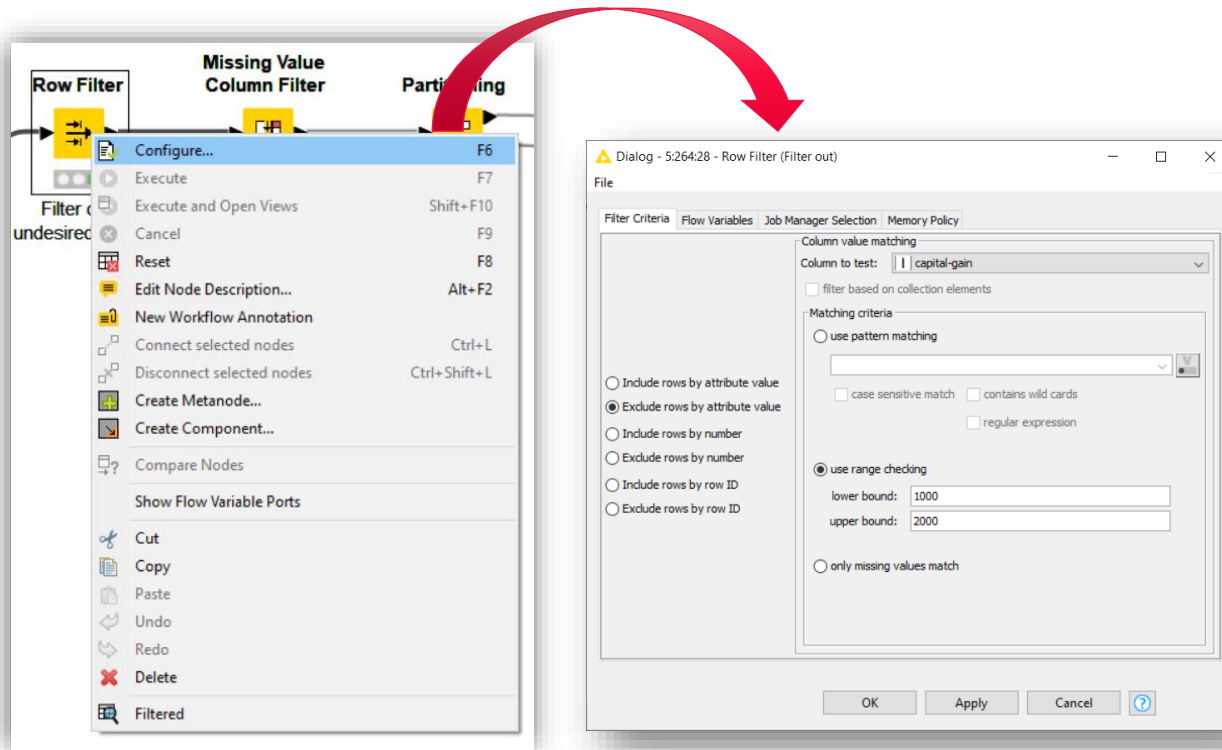


Data Port Types



- 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

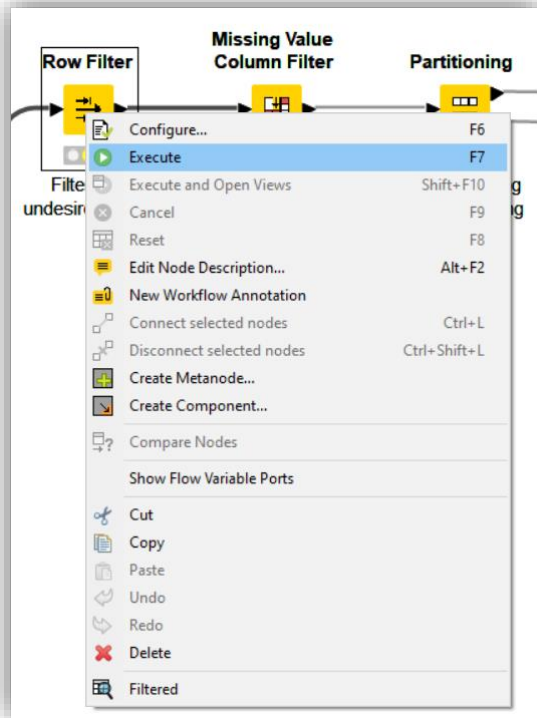
Node Configuration



- 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



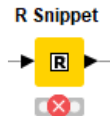
If execution is successful, status shows green light



If execution produces warnings, status show yellow triangle



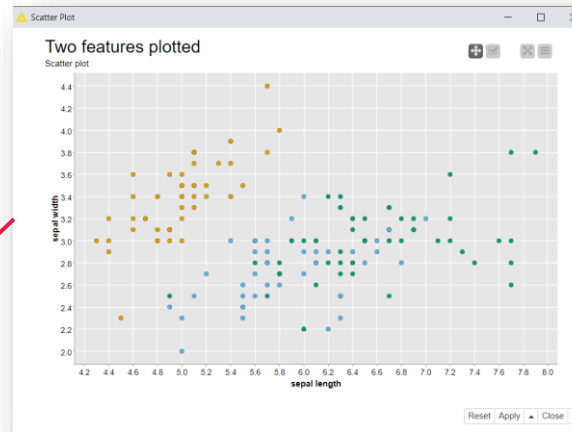
If execution encounters errors, status shows a red X



Node Views

Scatter Plot

- Configure... F6
- Execute F7
- Execute and Open Views Shift+F10
- Cancel F9
- Reset F8
- Edit Node Description... Alt+F2
- New Workflow Annotation
- Connect selected nodes Ctrl+L
- Disconnect selected nodes Ctrl+Shift+L
- Create Metanode...
- Create Component...
- Select Loop
- Interactive View: Scatter Plot**
- Compare Nodes
- Show Flow Variable Ports
- Cut
- Copy
- Paste
- Undo
- Redo
- Delete
- Image
- Input data and view selection**



Interactive View

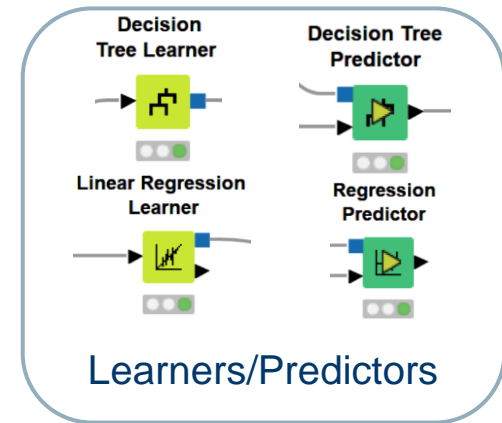
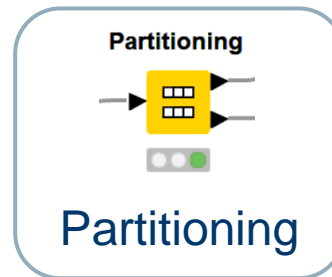
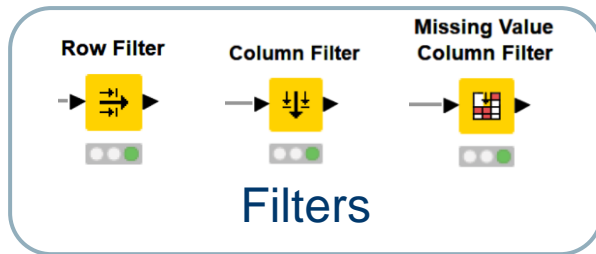
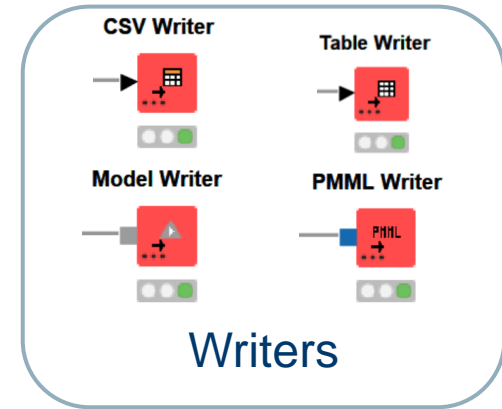
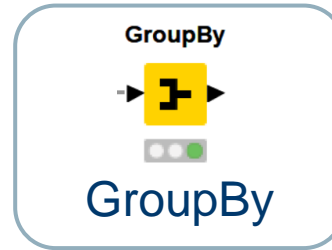
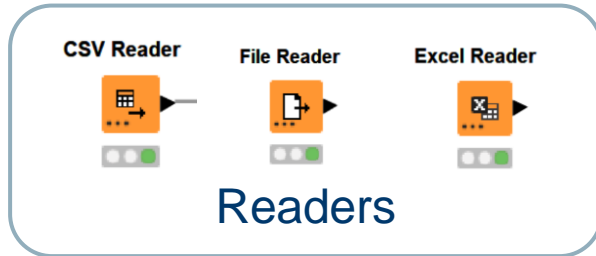
Input data and view selection - 1083:0:10 - Scatter Plot (Two features)

Table "default" - Rows: 148 Spec - Columns: 6 Properties Flow Variables

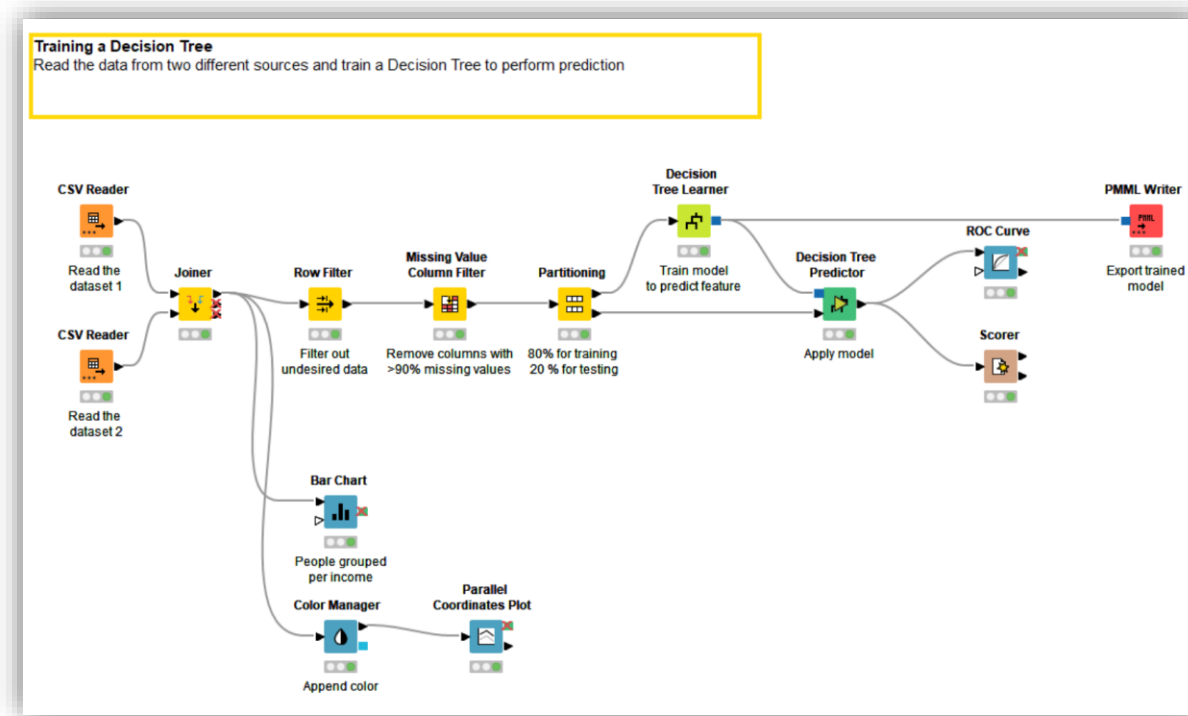
Row ID	D sepal le...	D sepal w...	D petal le...	D petal w...	S class n...	B Selecte...
Row27_Row0	5	3.5	1.6	0.6	Iris-setosa	false
Row28_Row0	4.8	3	1.4	0.3	Iris-setosa	false
Row29_Row0	4.6	3.2	1.4	0.2	Iris-setosa	false
Row30_Row0	5	3.3	1.4	0.2	Iris-setosa	false
Row31_Row1	6.4	3.2	4.5	1.5	Iris-versicolor	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	6.4	3.0	4.2	1.2	Iris-versicolor	false

Data View

Frequently Used Nodes



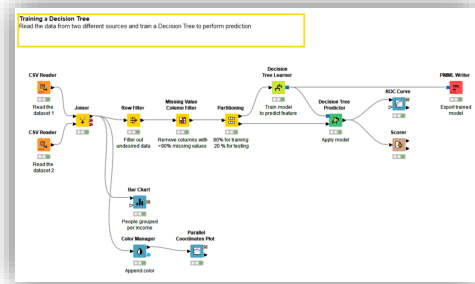
Tidy up workflows



- Workflow can easily become complex and difficult to understand

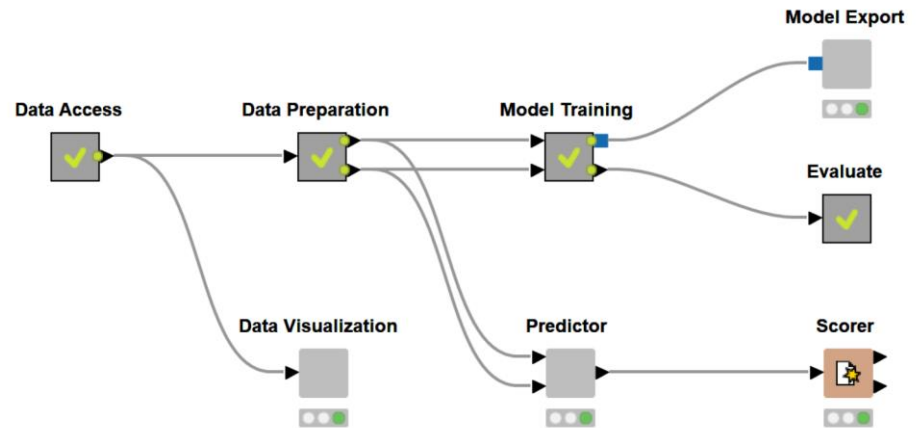
Metanodes and Components

Tidy up workflows



Training a Decision Tree

Read the data from two different sources and train a Decision Tree to perform prediction



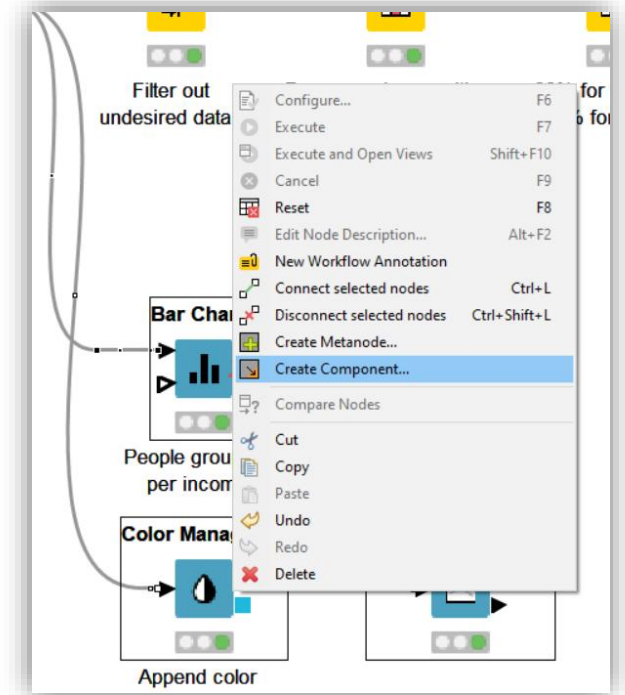
- Metanodes and components can help tidying up, encapsulating nodes performing common operations

Steps to build a component or a metanode

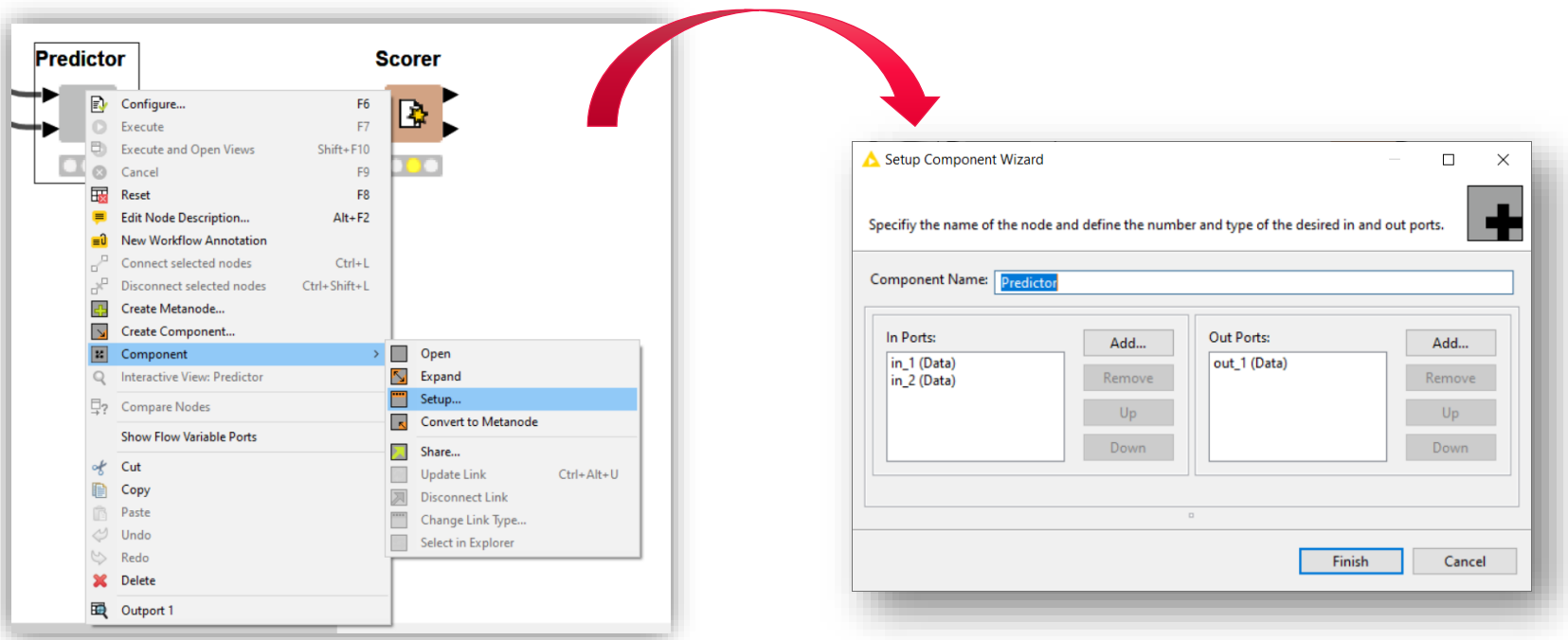
- Select related nodes that you want to group
- Right click
- Select *Create component...* or *Create Metanode...*
- Give it a name

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

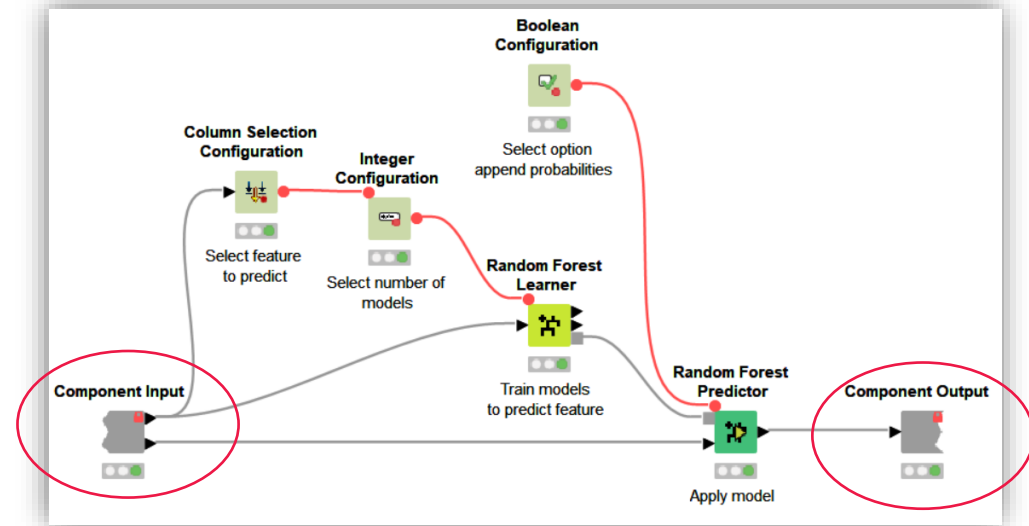
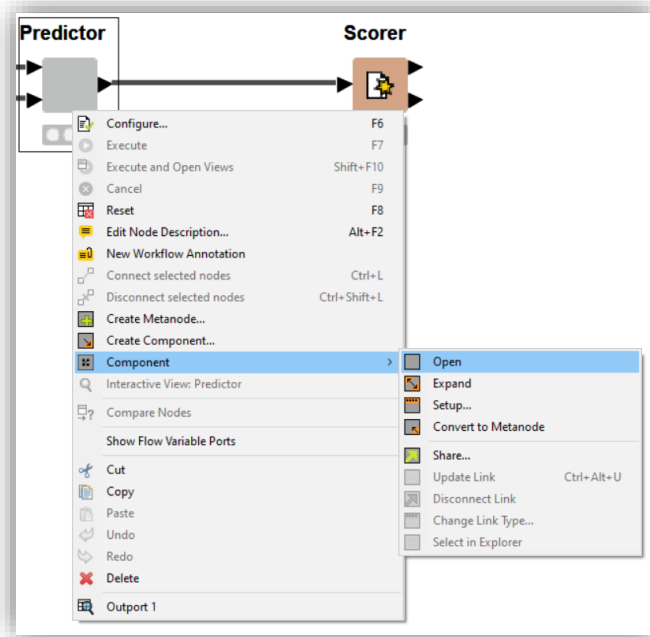


Submenu Component



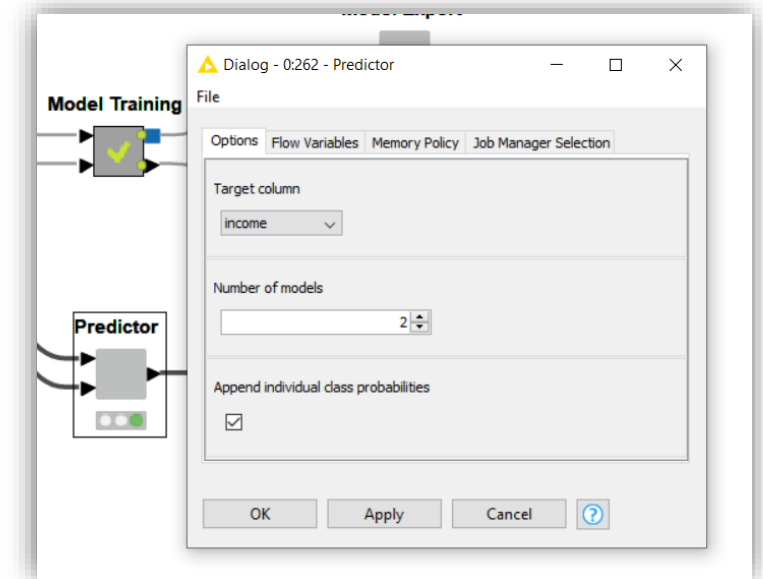
- Right click on the Component and select Setup... from the submenu Component to access further customization settings, such as the component name and the ports

Inside a component

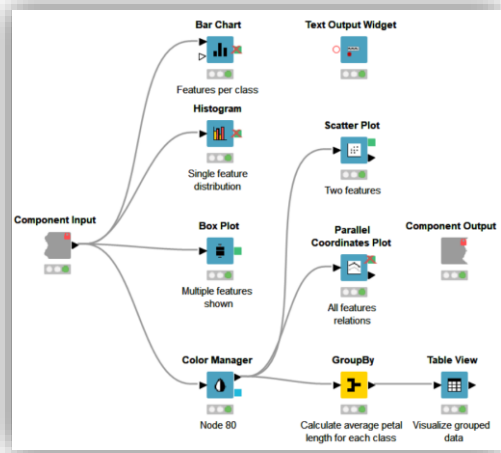


Shortcut:
Ctrl + double click on
component to open its
content

- Components 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

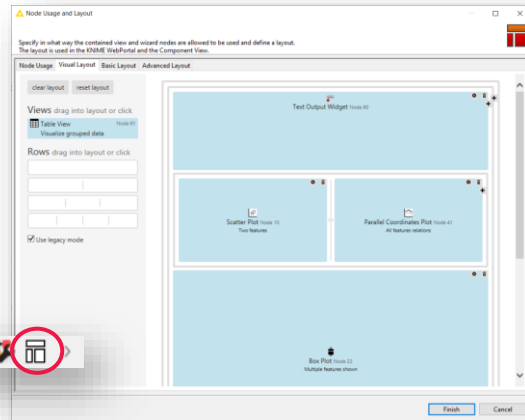


Components Composite View

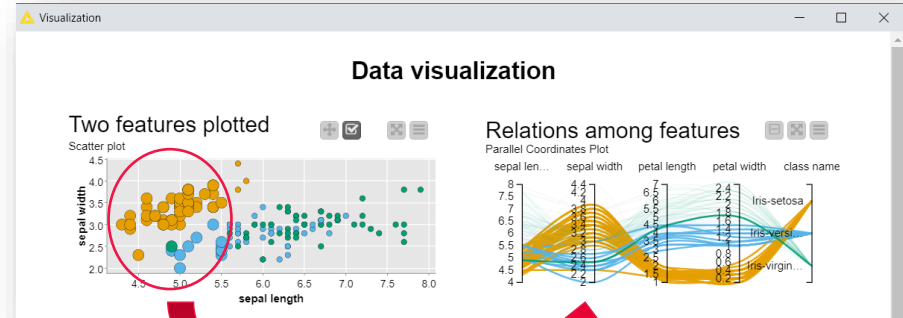
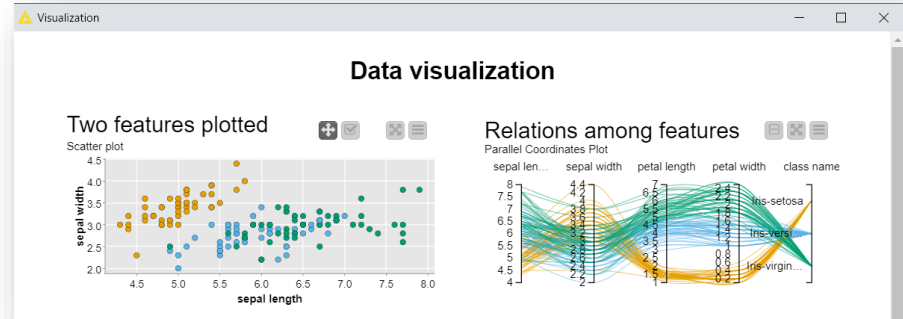
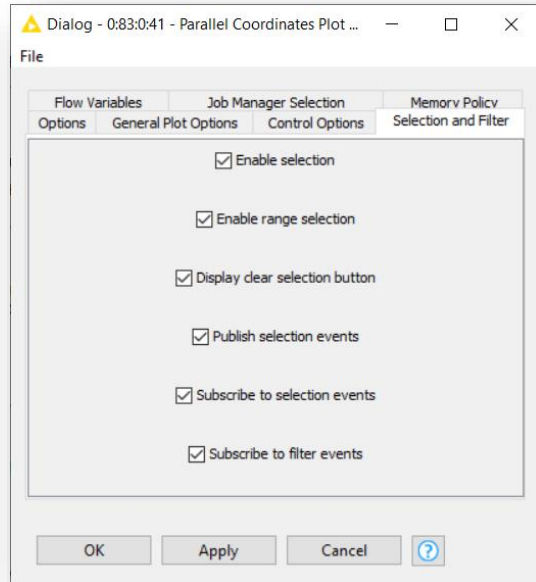


The visualization nodes within the component can be organized to build an interactive composite view

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

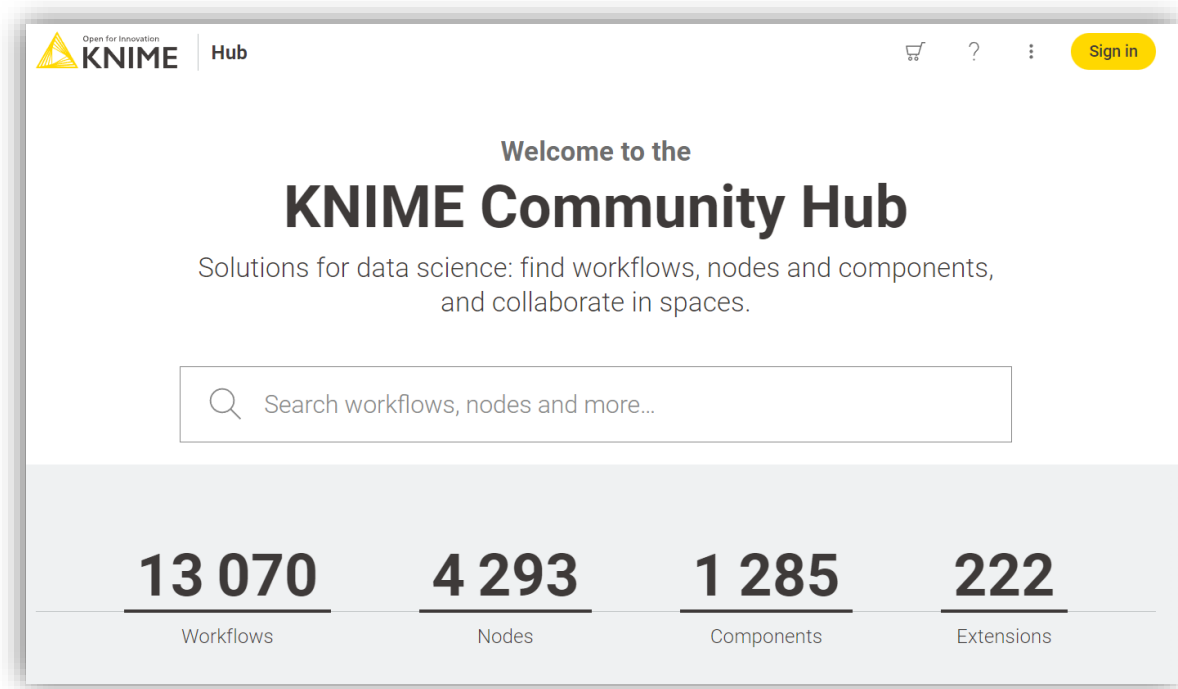


Composite views interactivity



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



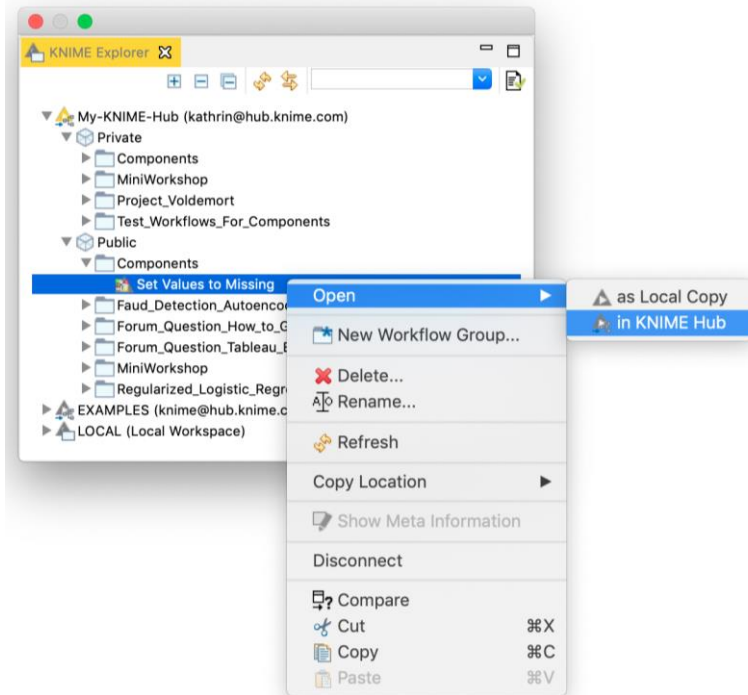
A place to share knowledge about Workflows and Nodes
<https://hub.knime.com>

The screenshot shows the KNIME Community Hub interface. At the top, there is a search bar and a 'Sign in' button. Below the navigation bar, the breadcrumb trail reads 'KNIME Hub > rs > Spaces > Data Science Guide > Workflows > Chapter8 > 01_DecisionTree'. The main content area features a workflow titled 'Decision Tree' with a 'Thumbnail' button and a user profile picture. Below the title, it says 'Last Update: 20 Feb 2020'. The workflow diagram includes nodes for 'CSV Reader', 'Color Manager', 'Partitioning', 'Train Decision Tree', 'Decision Tree Predictor', and 'Scorer (JavaScript)'. A 'Short link' is provided: <https://knt.me/rs/PV3W2ZpUrMML>. A text box below the diagram reads: '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

Nodes, Shared Components and Extensions

The screenshot shows the 'Node / Visualizer' interface for a 'Scatter Plot' node. The title is 'Scatter Plot'. Below the title, there is a description: '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.' The configuration section includes: '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.' Below the configuration, there is a note: '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 follows: 'Additionally a static SVG image can be rendered, which is then made available at the first output port.' A final note states: 'Note, this node is currently under development. Future versions of the node might have more or changed functionality.' On the right side, there is an 'Extension' section: 'KNIME JavaScript Views' (Version 4.2.1) with a 'Short link' <https://knt.me/rs/HuJl14WMyvH12x>.



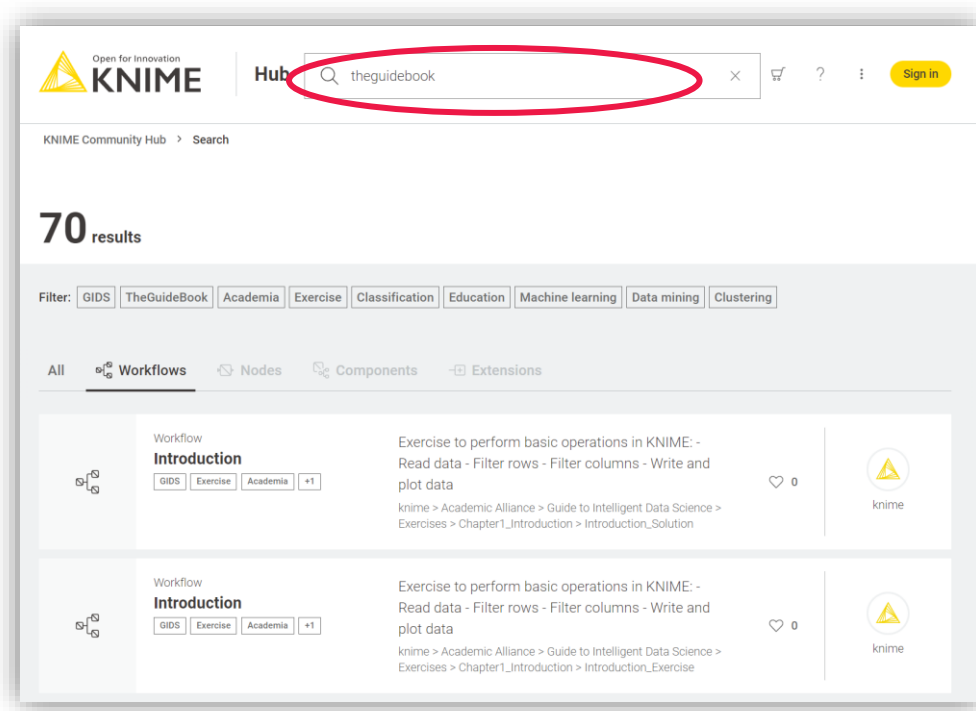
– Private Space

- 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



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

TheOutlookBook | deployment | dashboard | GDS

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

Import

Workflow Import Selection
Select the items to import.

Source:
 File: C:\Users\Public\Documents\01_Deploy_on_Dash Browse...
 Select root directory: Browse...

Destination:
Select folder: LOCAL:/ Browse...

Import Elements:
 01_Deploy_on_Dashboard Select All Deselect All

< Back Next > Finish Cancel

Downloading and importing a workflow from the KNIME Community Hub

Workflow
Deployment to a Dashboard

TheGuideBook | deployment | dashboard | GDS

Last Update: 24 Feb 2020

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


Model Reader
Load trained model

Decision Tree Predictor
Predict churn

Data upload

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 Analytics Platform

File Edit View Node Search Help

KNIME Explorer

- > My-KNIME-Hub (hub.knime.com)
- > EXAMPLES (knime@hub.knime.com)
- > LOCAL (Local Workspace)

Cheat Sheet: Machine Learning with KNIME Analytics Platform

SUPERVISED LEARNING

- CLASSIFICATION**
 - Decision Tree Classifier**: A simple model that splits data into two or more classes based on a single feature.
 - Naive Bayes Classifier**: A probabilistic model based on Bayes' theorem with the assumption of independence between features.
 - Logistic Regression**: A linear model for binary classification.
 - SVM**: Finds the optimal hyperplane that separates the classes.
 - Random Forest Classifier**: An ensemble method that combines multiple decision trees.
 - AdaBoost**: A meta-algorithm that combines weak classifiers to create a strong classifier.
 - Gradient Boosting Classifier**: An ensemble method that builds trees sequentially, minimizing a loss function.
- NUMERIC REGRESSION**
 - Linear Regression**: A simple model that fits a straight line to the data.
 - Polynomial Regression**: A model that fits a polynomial curve to the data.
 - Support Vector Regression**: A regression method that uses support vectors to define the regression line.
 - Decision Tree Regressor**: A regression model that uses decision trees.
 - Random Forest Regressor**: An ensemble regression model.
 - Gradient Boosting Regressor**: An ensemble regression model.

UNSUPERVISED LEARNING

- CLUSTERING**
 - K-Means**: A partitioning method that divides data into K clusters.
 - Hierarchical Clustering**: A clustering method that builds a hierarchy of clusters.
 - DBSCAN**: A density-based clustering method that finds core points of high density and expands them into clusters.
 - Gaussian Mixture Model**: A probabilistic model that represents the data as a mixture of Gaussian distributions.

IMPUTABLE LEARNING

- EM ALGORITHM**: An iterative algorithm for finding maximum likelihood estimates of parameters in a statistical model.

TRAINING

- Model Training**: The process of fitting a model to a dataset.
- Model Evaluation**: The process of assessing the performance of a model on a new dataset.
- Model Deployment**: The process of using a trained model to make predictions on new data.

Cheat Sheet: Control and Orchestration with KNIME Analytics Platform

CONTROL

- Flow Variables**: Used to pass data between nodes and store information.
- Widget and Configuration nodes**: Used to create and configure widgets.
- Column Selection Wizard**: Used to select columns for a node.
- Column Selection Filter**: Used to filter columns based on a condition.
- Column Selection Wizard**: Used to select columns for a node.
- Column Selection Filter**: Used to filter columns based on a condition.
- Column Selection Wizard**: Used to select columns for a node.
- Column Selection Filter**: Used to filter columns based on a condition.

OPERATION

- CSV Reader**: Reads CSV files.
- Excel Reader**: Reads Excel files.
- Table Reader**: Reads tables from a database.
- Google Sheets Reader**: Reads data from Google Sheets.
- Table Filter**: Filters rows based on a condition.
- Table Joiner**: Joins two tables.
- Table Splitter**: Splits a table into multiple tables.
- Table Sorter**: Sorts a table.
- Table Joiner**: Joins two tables.
- Table Splitter**: Splits a table into multiple tables.
- Table Sorter**: Sorts a table.

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

Cheat Sheet: Building a KNIME Workflow for Beginners

Getting started with KNIME Analytics Platform

- Check out the 7 things you should do after installing KNIME Analytics Platform.
- Take the Learning Course at www.knime.com/learning.
- Understanding the basic light system.
- Not configured: Nodes that are not configured and cannot be executed with its current settings.
- Configured: Nodes that have been correctly configured and results can be executed and sent to downstream nodes.
- Dynamic: Nodes that have been successfully executed and results can be executed and sent to downstream nodes.
- Dynamic: Nodes that have been successfully executed and results can be executed and sent to downstream nodes.

EXPLORE

- Scatter Plot**: Represents data points in two dimensions.
- Bar Chart**: Displays categorical data.
- Stacked Area Chart**: Plots multiple numerical data series over time.
- Color Manager**: Assigns a color property to each data point.
- Box Plot**: Visualizes statistical data.
- Data Explorer**: Provides an interactive view to explore data.

ANALYZE

- Decision Tree**: A model that splits data into two or more classes.
- K-Means**: A clustering method that divides data into K clusters.
- Logistic Regression**: A linear model for binary classification.
- Score**: Calculates a number of performance measures such as accuracy, F1 score, and Cohen's Kappa.
- Names Score**: Calculates a number of performance measures such as accuracy, F1 score, and Cohen's Kappa.
- RUC Curve**: Displays the quality of a Receiver Operating Characteristic (ROC) curve.

OPERATION

- CSV Reader**: Reads CSV files.
- Excel Reader**: Reads Excel files.
- Table Reader**: Reads tables from a database.
- Google Sheets Reader**: Reads data from Google Sheets.
- Table Filter**: Filters rows based on a condition.
- Table Joiner**: Joins two tables.
- Table Splitter**: Splits a table into multiple tables.
- Table Sorter**: Sorts a table.
- Table Joiner**: Joins two tables.
- Table Splitter**: Splits a table into multiple tables.
- Table Sorter**: Sorts a table.

EXPLORE

- Model Reader**: Reads machine learning models.
- Table Reader**: Reads tables from a database.
- Table Filter**: Filters rows based on a condition.
- Table Joiner**: Joins two tables.
- Table Splitter**: Splits a table into multiple tables.
- Table Sorter**: Sorts a table.

TRANSFORM

- Math Formula**: Applies mathematical operations to data.
- Joiner**: Joins two tables.
- Sorter**: Sorts a table.
- Concatenator**: Concatenates two tables.
- Missing Value**: Replaces missing values.
- String Manipulation**: Performs operations on strings.

OUTPUT

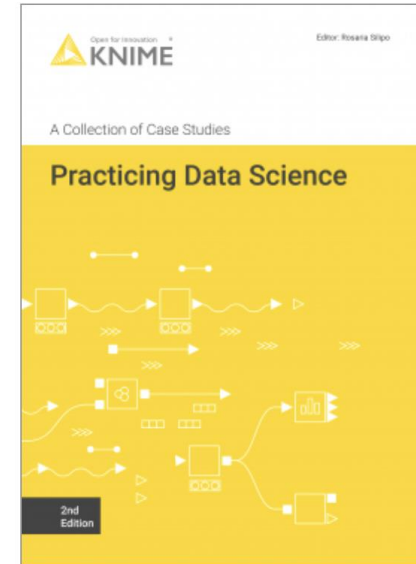
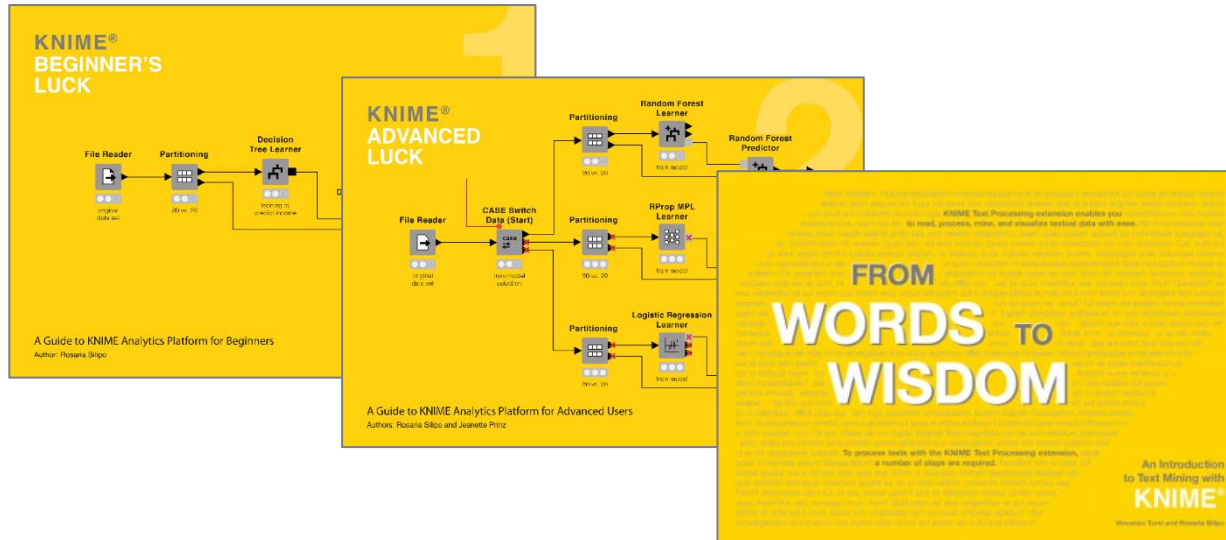
- CSV Writer**: Writes data to a CSV file.
- Table Writer**: Writes data to a table.
- Table Reader**: Reads tables from a database.
- Table Filter**: Filters rows based on a condition.
- Table Joiner**: Joins two tables.
- Table Splitter**: Splits a table into multiple tables.
- Table Sorter**: Sorts a table.

Resources

- KNIME Forum**: A community for sharing tips and tricks.
- KNIME Books**: A collection of books on KNIME.
- KNIME Wiki**: A collection of articles on KNIME.
- KNIME YouTube**: A collection of videos on KNIME.
- KNIME LinkedIn**: A collection of posts on KNIME.
- KNIME Twitter**: A collection of tweets on KNIME.
- KNIME Facebook**: A collection of posts on KNIME.
- KNIME Instagram**: A collection of posts on KNIME.
- KNIME Snapchat**: A collection of posts on KNIME.
- KNIME TikTok**: A collection of posts on KNIME.

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

with code: **<Promotion-Code>**



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

Open for Innovation
KNIME

Community Hub Blog Forum Educators Events Solutions Careers Contact

SOFTWARE / PRICING / COMMUNITY / LEARNING / PARTNERS / ABOUT

Download

KNIME Self-Paced Courses

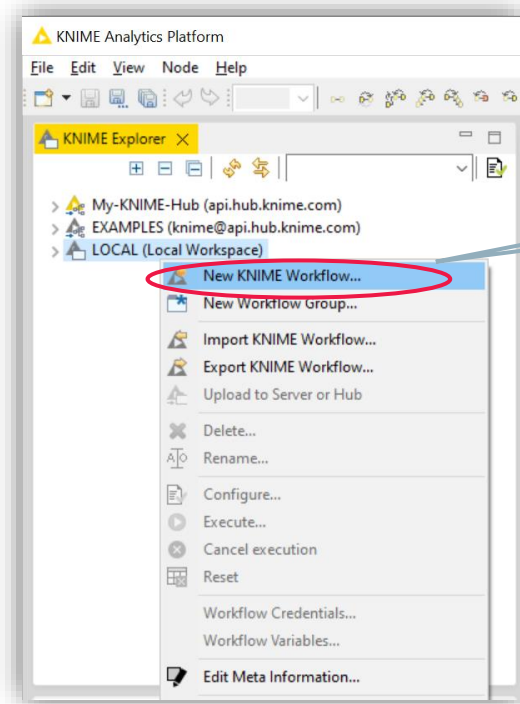
Courses are organized by level: L1 basic, L2 advanced, L3 deployment, L4 specialized.

In each course, go through the lessons with ~5 minutes videos, hands-on exercises, and knowledge-check questions.

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

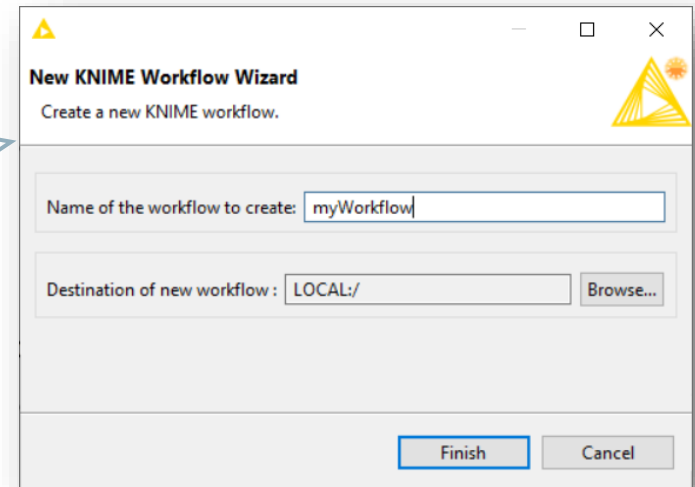
Build your first Hello Workflow

Create your first workflow

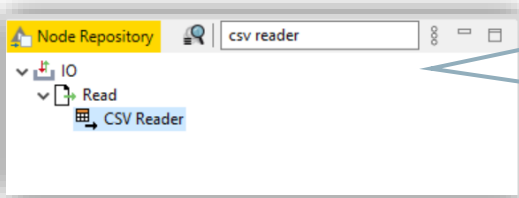
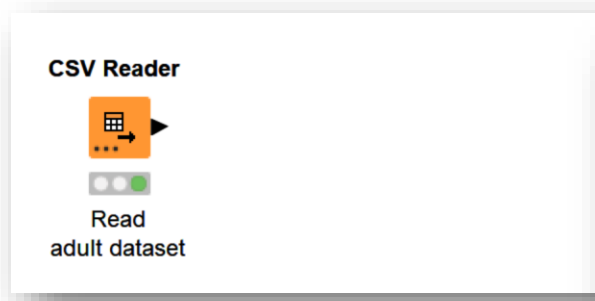


Right Click on the LOCAL folder in the KNIME Explorer and select *New KNIME Workflow*

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

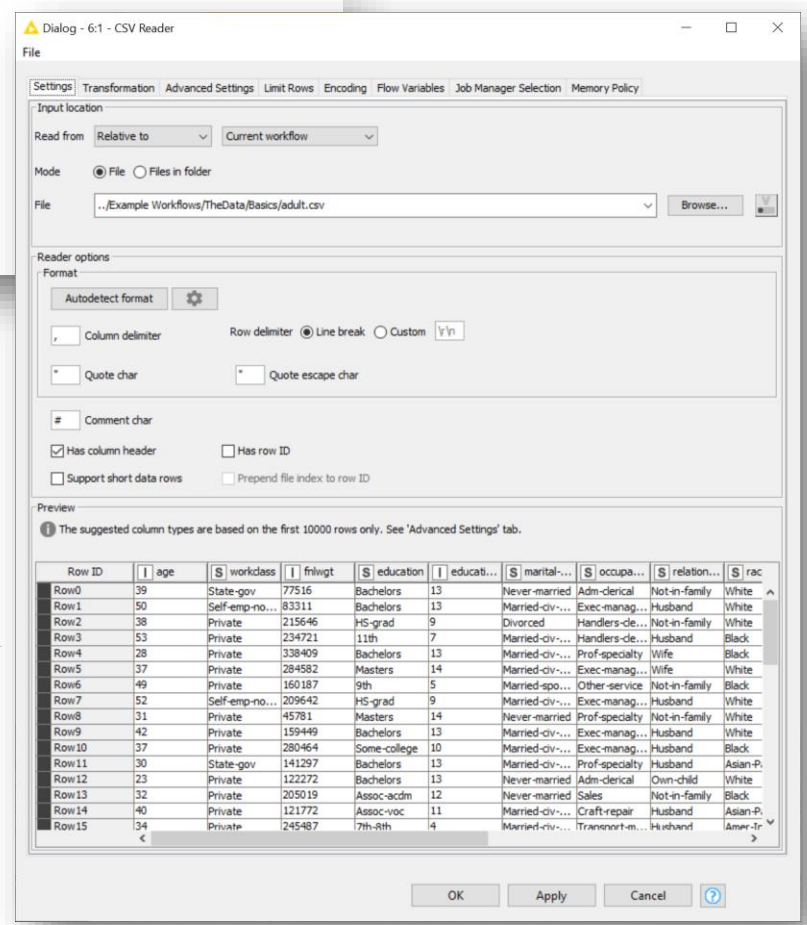


Read the dataset

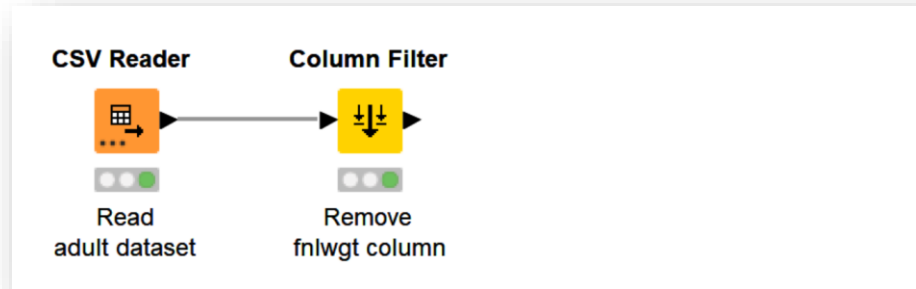


Drag and drop the **CSV Reader** node from the Node Repository to add it to the workflow

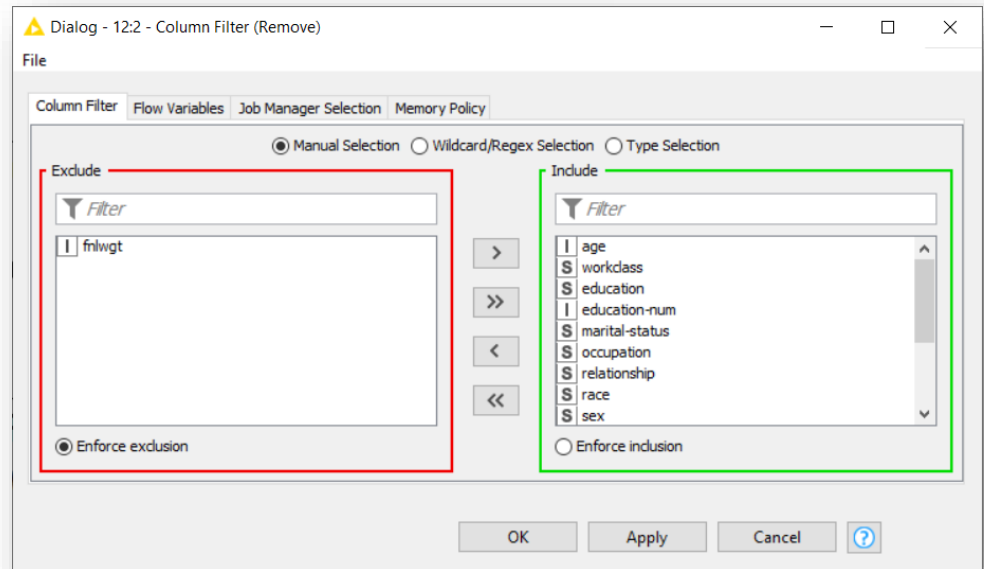
Open the configuration window (double click) and select the file on your machine containing the adult dataset



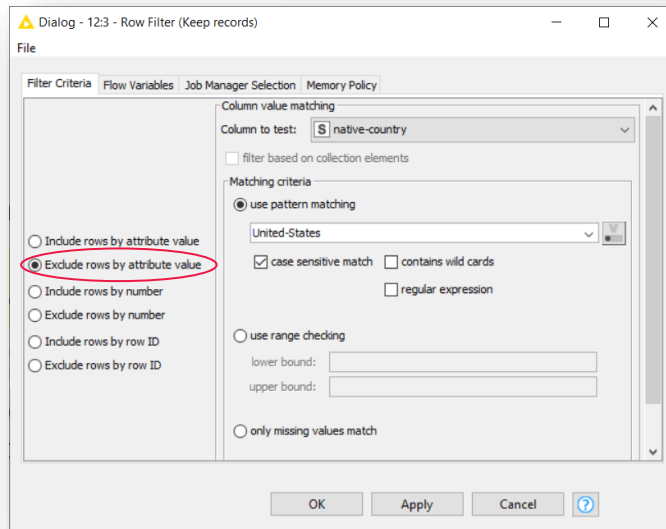
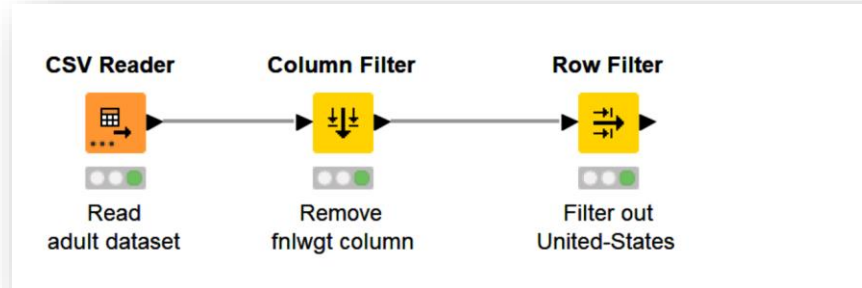
Remove columns



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

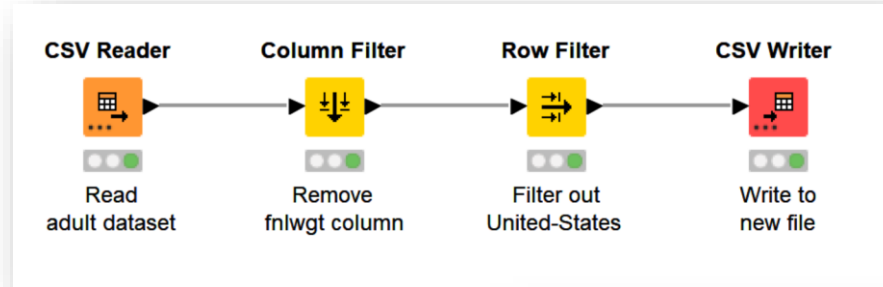


Remove Rows



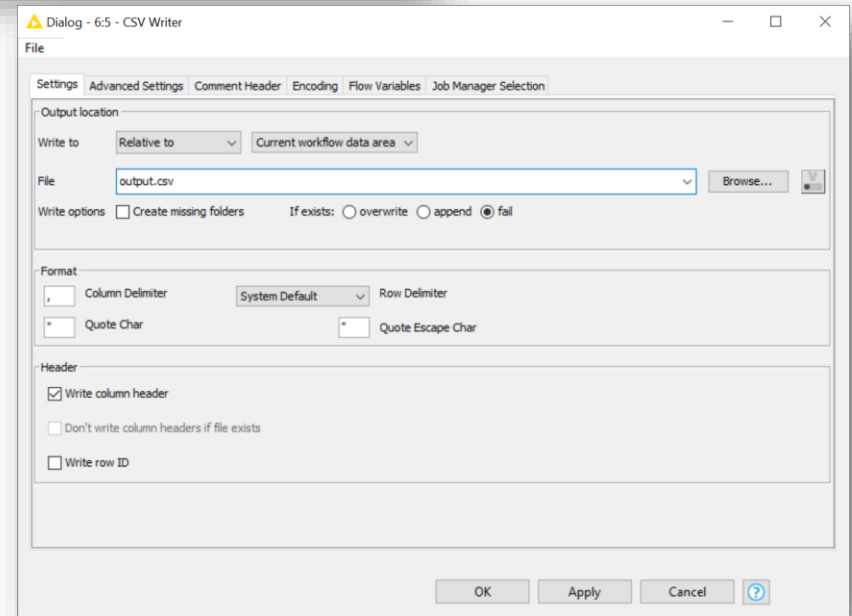
Add a Row Filter node and configure it to only keep entries whose “native-country” value is not “United-States”

Write to new file



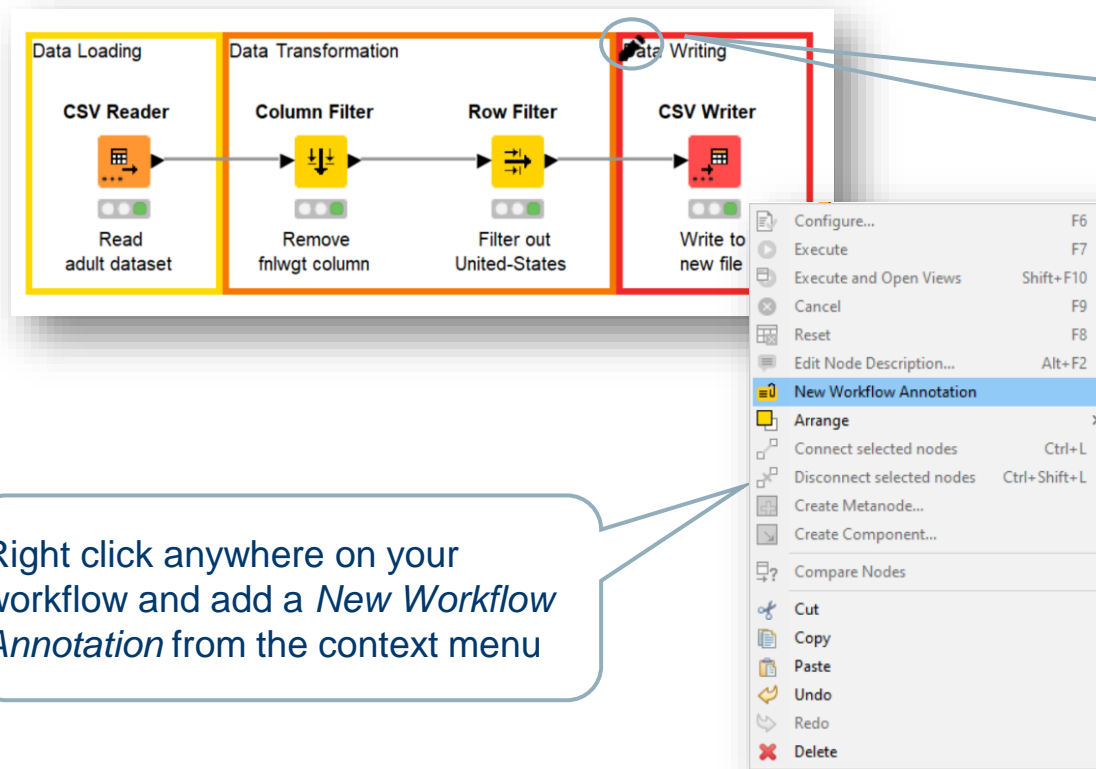
Finally add a CSV Writer node to the pipeline.

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



Click on the upper left corner icon to customize text and appearance of an annotation

Right click anywhere on your workflow and add a *New Workflow Annotation* from the context menu

Thank you

Guide to Intelligent Data Science Second Edition, 2020