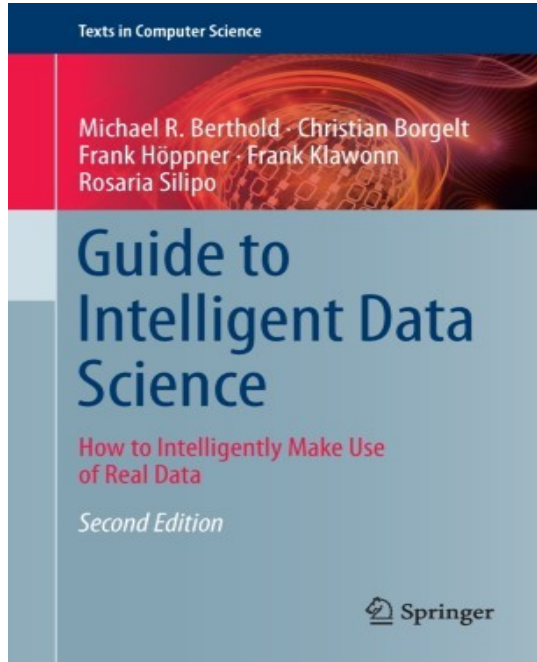




Big Data ACUA6002 Tutorial 1

KNIME- Friend or Foe ?

„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?

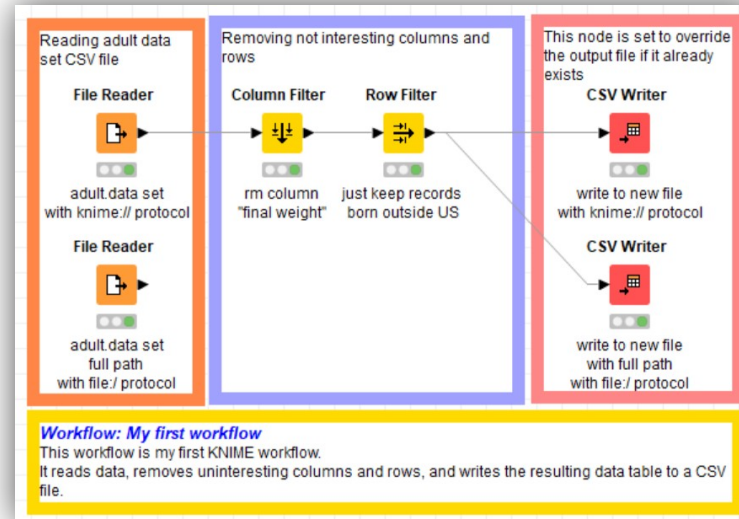
**This lesson refers to Appendix B of the GIDS book*

Content of this lesson

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

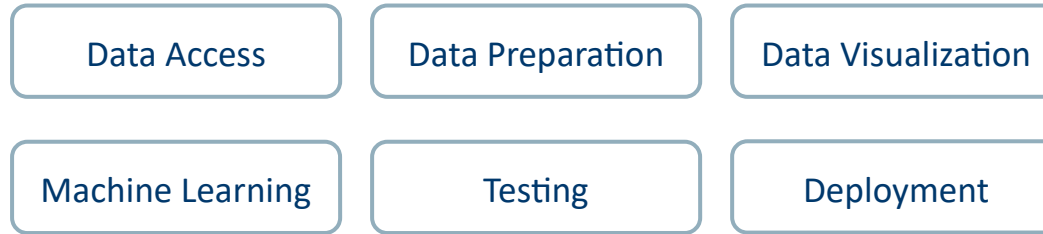
Datasets

- 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 Server

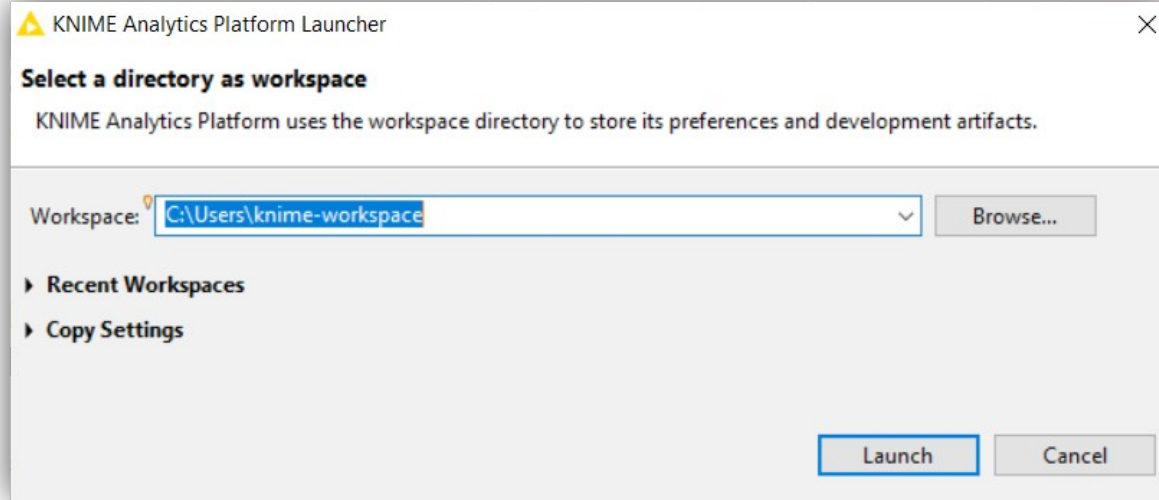
- To integrate the solutions into the IT environment
 - Scheduling
 - MLOps
 - Easy deployment
 - REST architecture
 - Auditing tools
- Closed source
- Yearly license

<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 Workbench interface. The main window is titled "My First Workflow" and contains a workflow diagram. The workflow starts with a "File Reader" node (labeled "Read sales data"), followed by a "Column Filter" node (labeled "Include country, date and amount"), and then a "Row Filter" node (labeled "Exclude rows with country 'unknown'"). The output of the Row Filter node is split into two paths: one leading to a "Stacked Area Chart" (labeled "Sales over time") and the other leading to a "Pie/Donut Chart" (labeled "Sales per country").

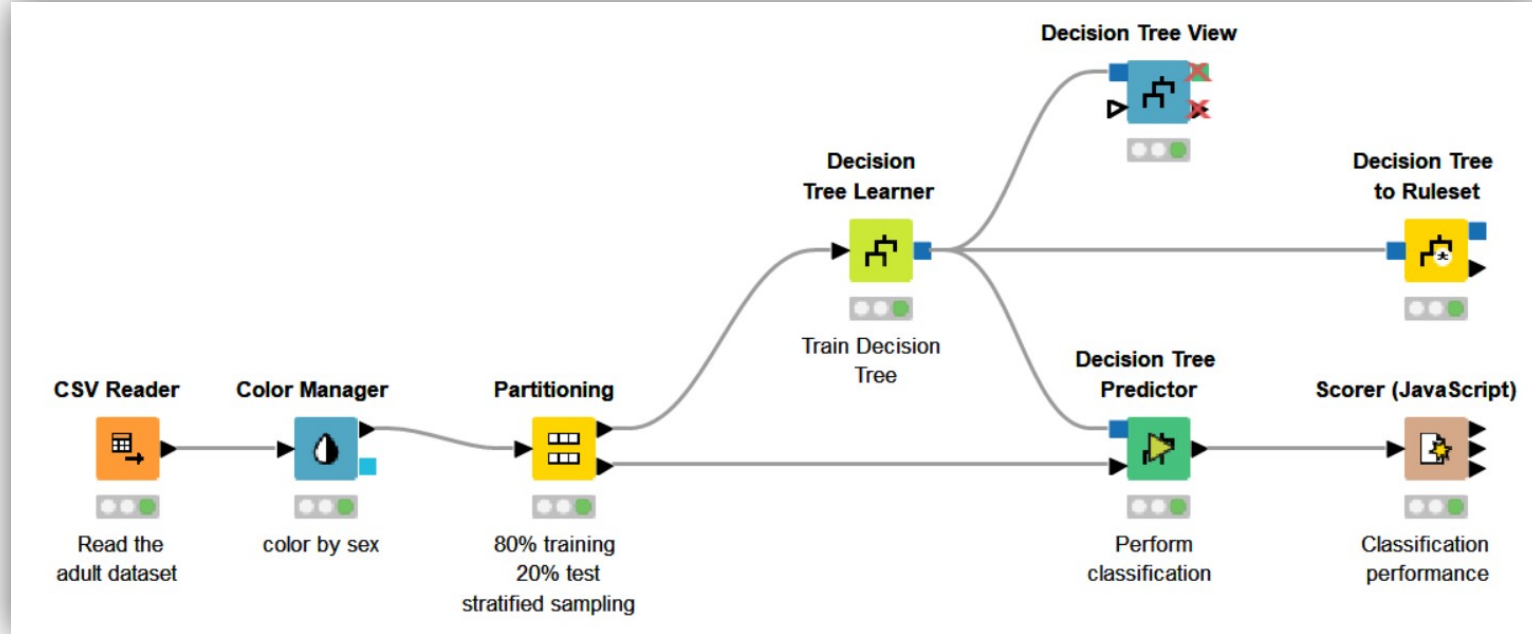
Surrounding the main workflow editor are several panels:

- KNIME Explorer:** Located on the top left, it shows a tree view of the project structure, including "My-KNIME-Hub", "EXAMPLES", "LOCAL (Local Workspace)", "Example Workflows", "My_First_Project", "data", and "My_First_Workflow".
- Workflow Coach:** Located on the middle left, it lists "Recommended Nodes" such as "Joiner", "Column Filter", "Row Filter", "Partitioning", "GroupBy", "Missing Value", and "Statistics".
- Node Repository:** Located on the bottom left, it provides a search bar and a list of node categories: "IO", "Manipulation", "Views", "Analytics", "DB", and "Other Data Types".
- Description:** Located on the top right, it provides detailed information about the selected "File Reader" node, explaining its purpose and configuration options.
- KNIME Hub Search:** Located on the bottom right, it provides a search bar for finding workflows and nodes.
- Outline:** Located at the bottom center, it shows a hierarchical view of the workflow structure.
- Console:** Located at the bottom right, it displays the KNIME Console output, which includes a welcome message and copyright information.

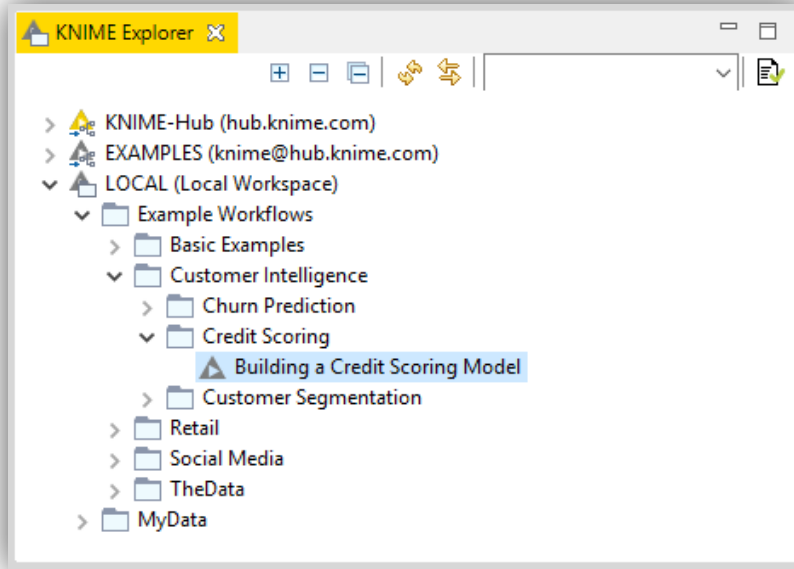
Yellow callout boxes with labels identify the following components:

- KNIME Explorer**
- Workflow Editor**
- Workflow Coach**
- Node Repository**
- Description**
- KNIME Hub Search**
- Outline**
- Console**

Workflow

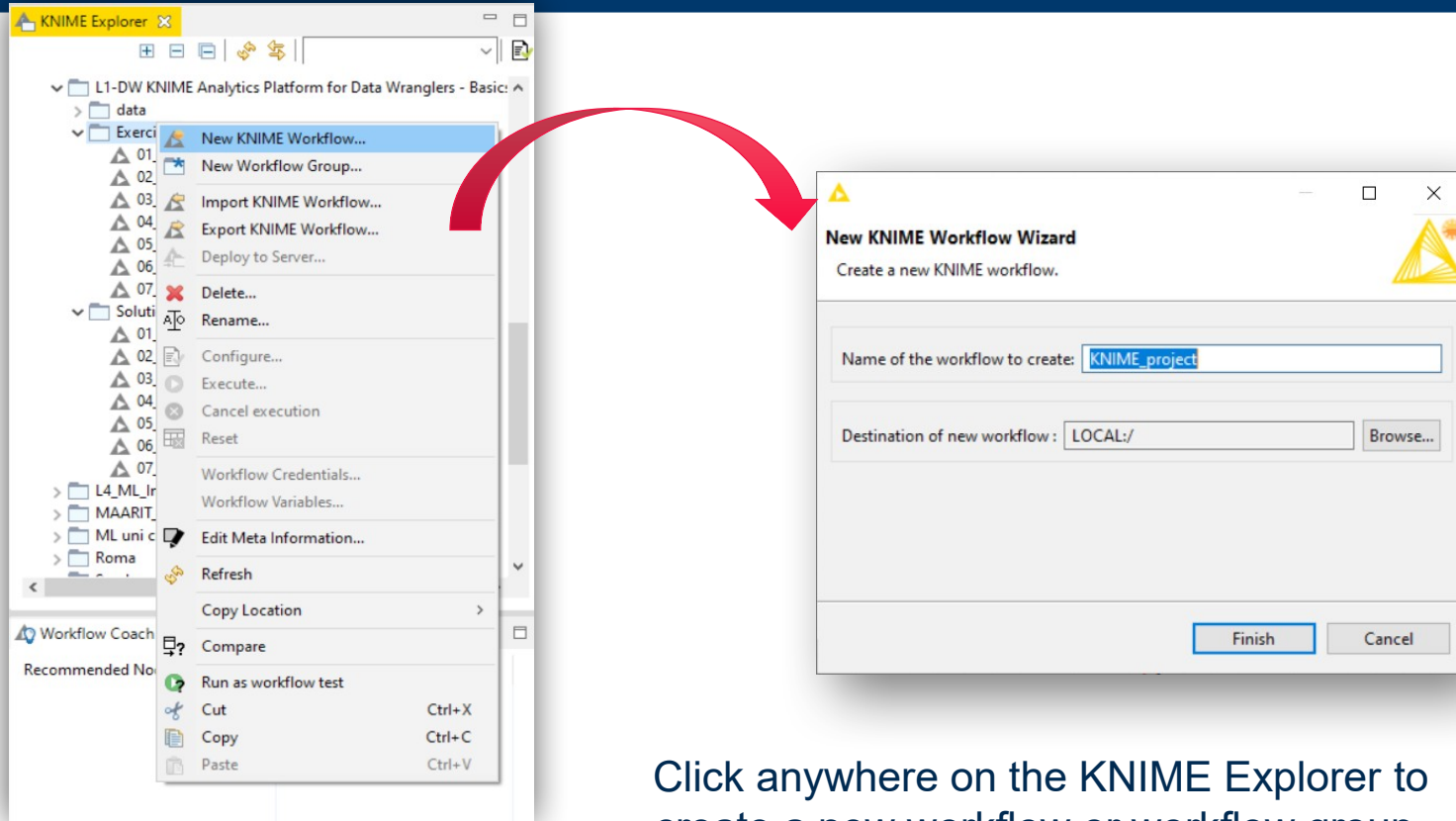


A workflow is a pipeline of nodes, each configurable to perform a specific task.
The data flow through nodes from left to right



- 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

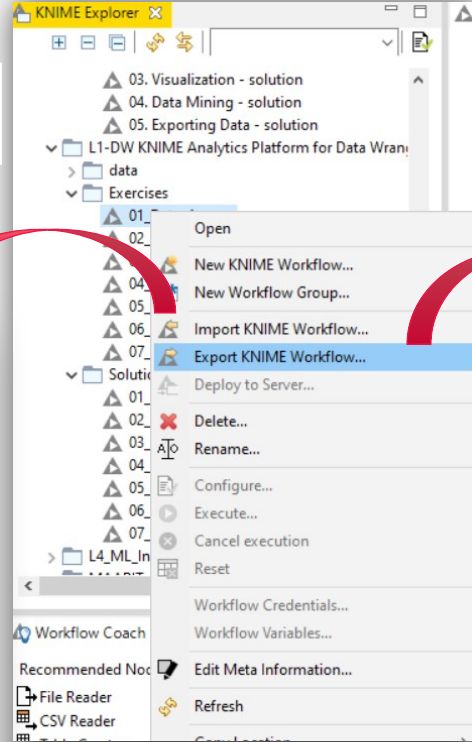
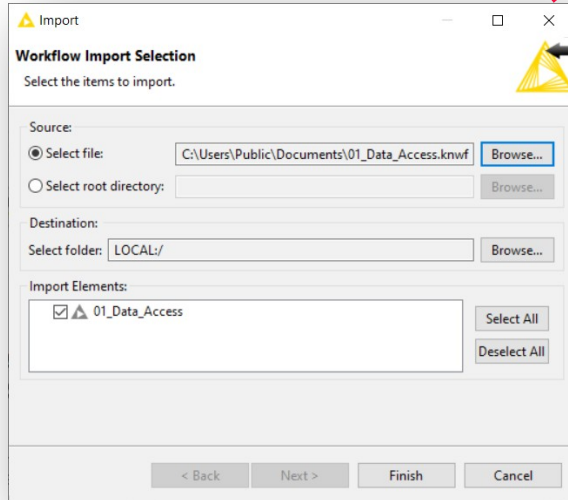
Creating a new workflow



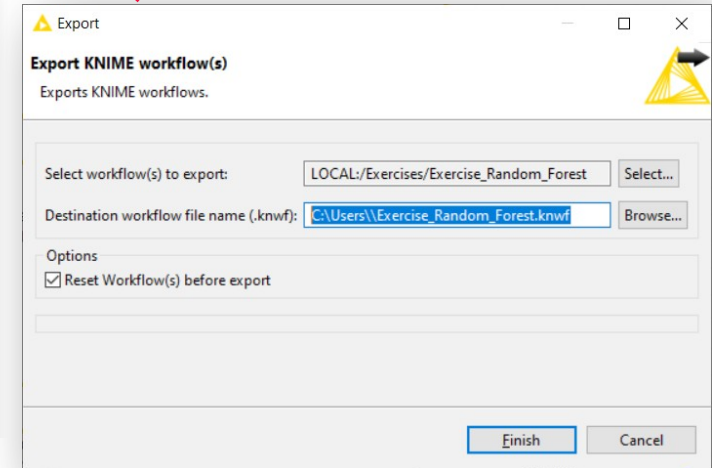
Click anywhere on the KNIME Explorer to create a new workflow or workflow group

Importing and Exporting Workflows

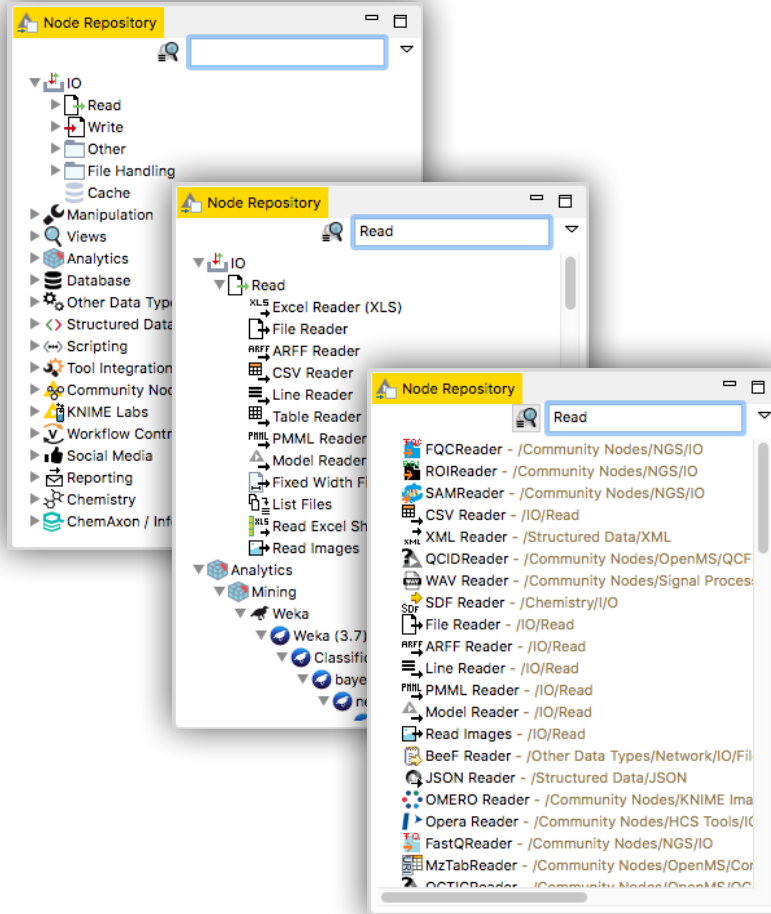
Right-click anywhere in KNIME Explorer to import a workflow



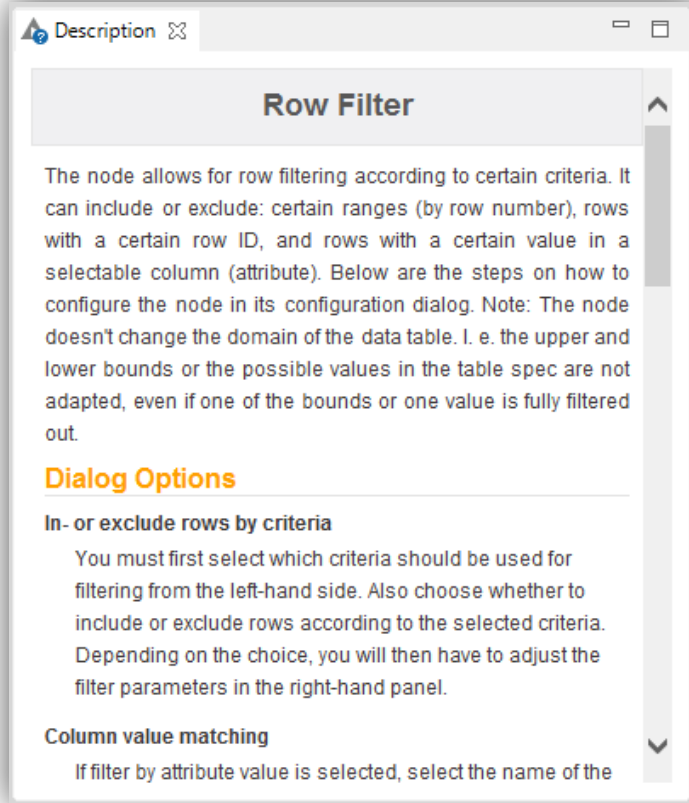
Right-click on a workflow or workflow group to export the selected workflow



Node Repository

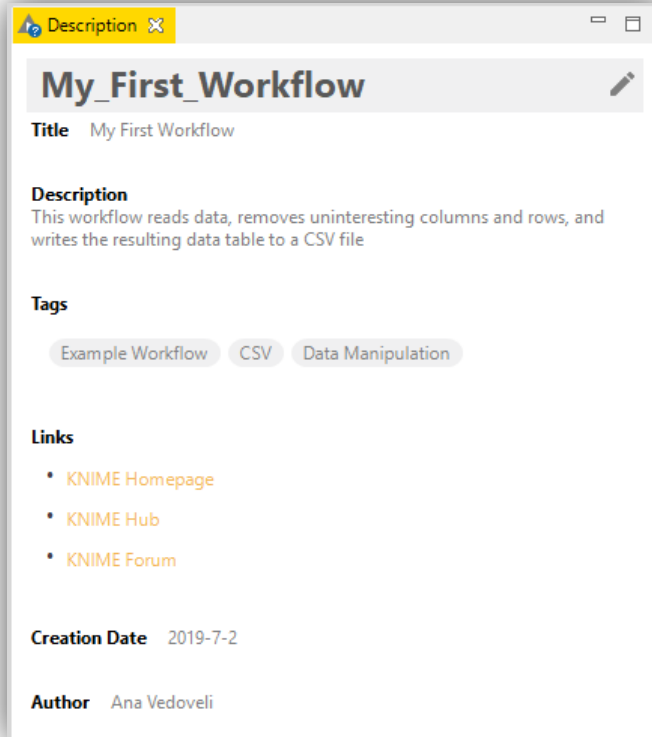


- The Node Repository contains all KNIME nodes - ordered by category with further subcategories.
- Extensions installation can sensibly increase the number of nodes
- Two search methods:
 -  Crisp Search
 -  Fuzzy Search
- Nodes can be added by drag and drop from the Node Repository to the Workflow Editor



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

Workflow Description

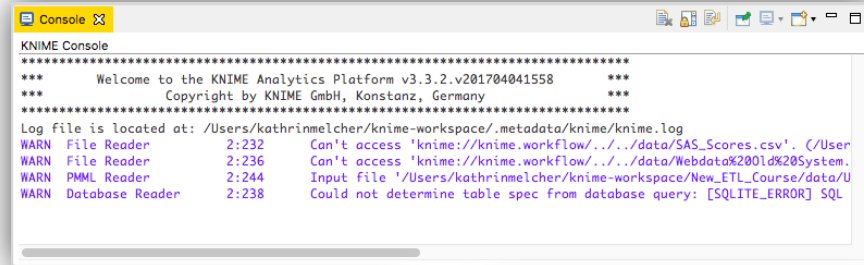


- When selecting the workflow, the Description window gives information about the workflow's:
 - Title
 - Description
 - Associated Tags and Links
 - Creation Date
 - Author

- 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

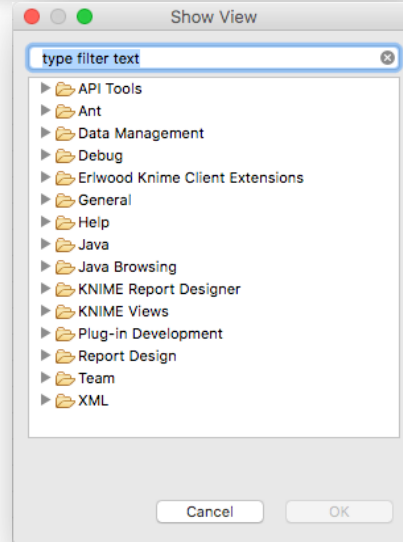
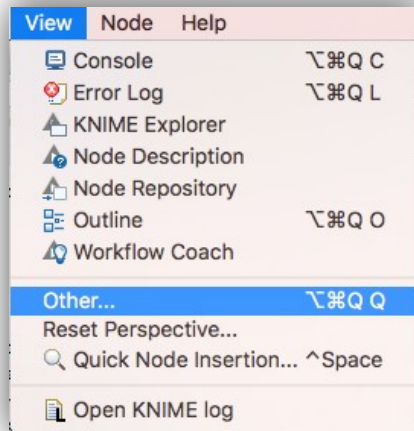


Console and Other views



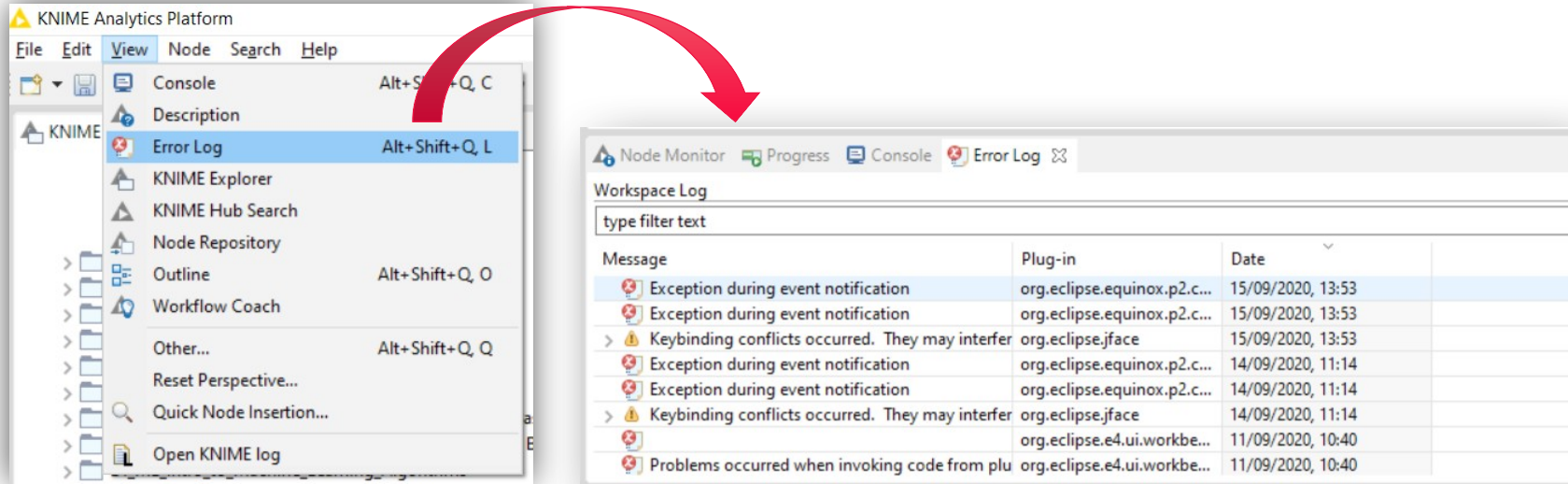
```
KNIME Console
*****
*** Welcome to the KNIME Analytics Platform v3.3.2.v201704041558 ***
*** Copyright by KNIME GmbH, Konstanz, Germany ***
*****
Log file is located at: /Users/kathrinmelcher/knime-workspace/.metadata/knime/knime.log
WARN File Reader 2:232 Can't access 'knime://knime.workflow/.../data/SAS_Scores.csv'.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

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



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

Error Log View

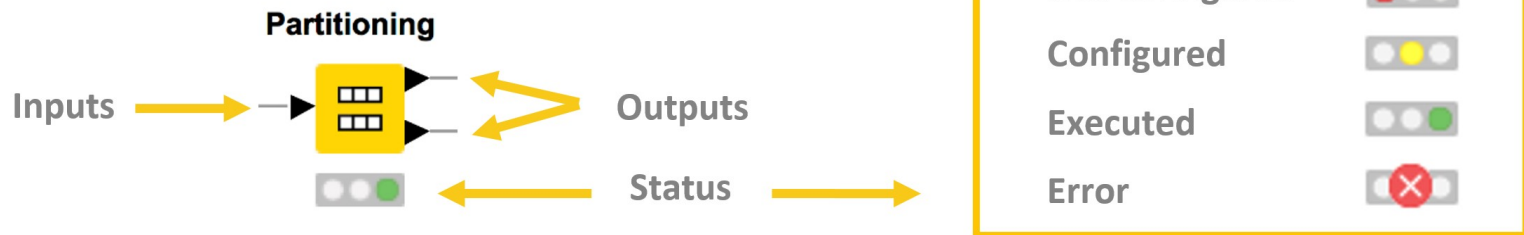


Tip: enabling and checking the Error Log view can help while debugging your project

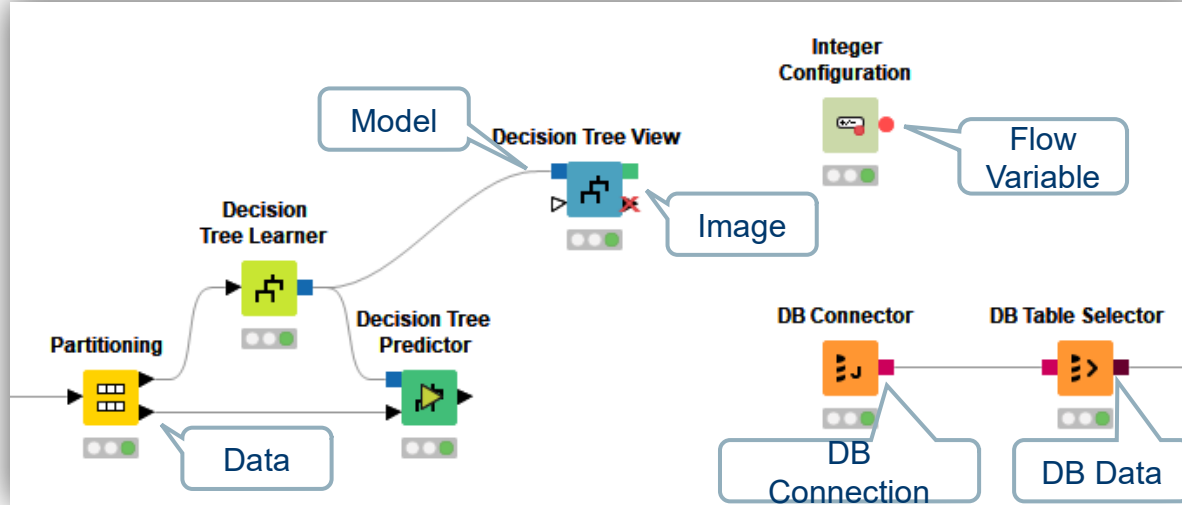
More on Nodes

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

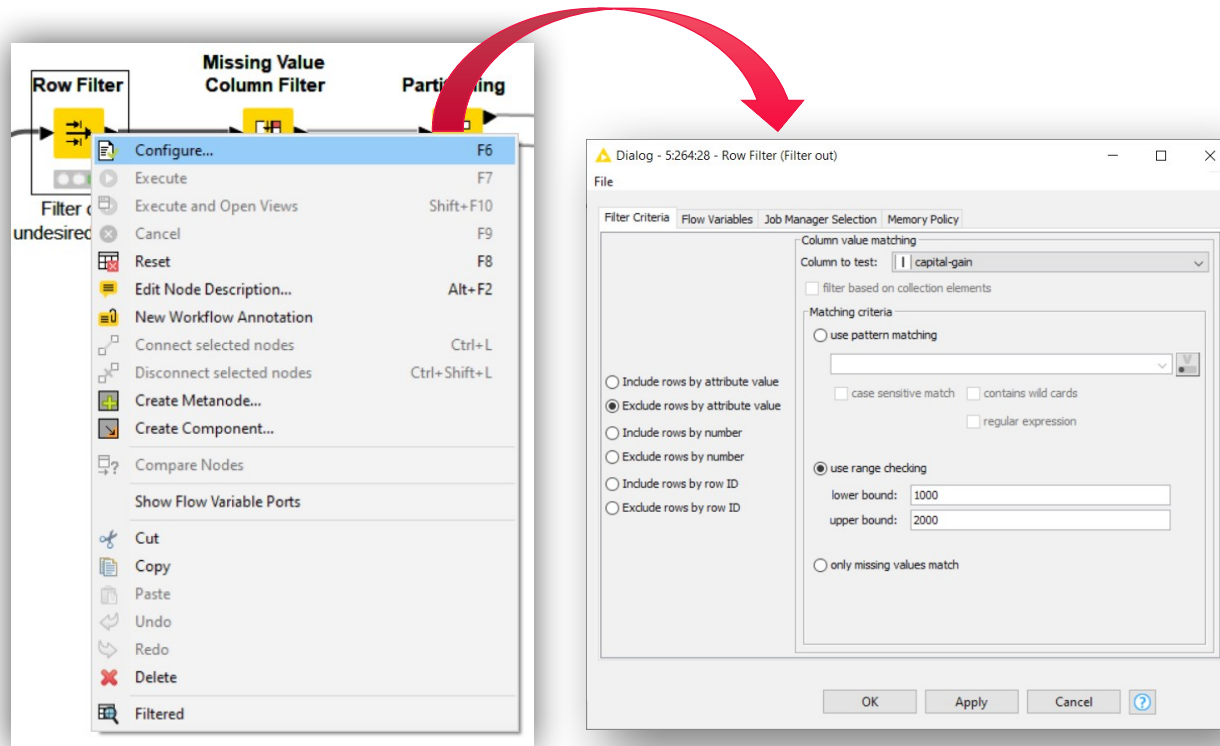


Data Port Types



- 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

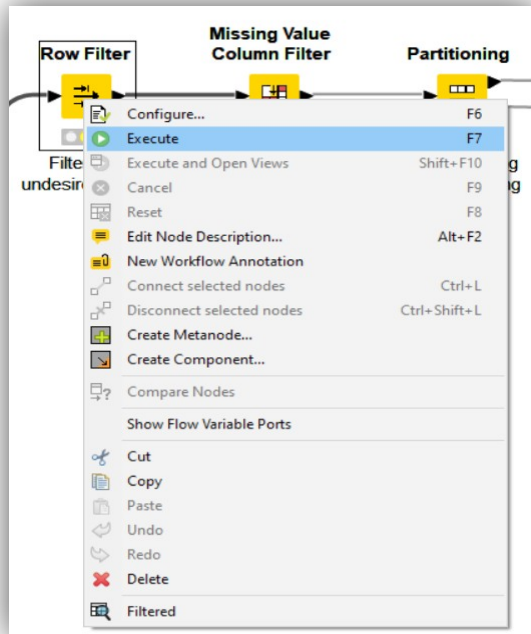
Node Configuration



- 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



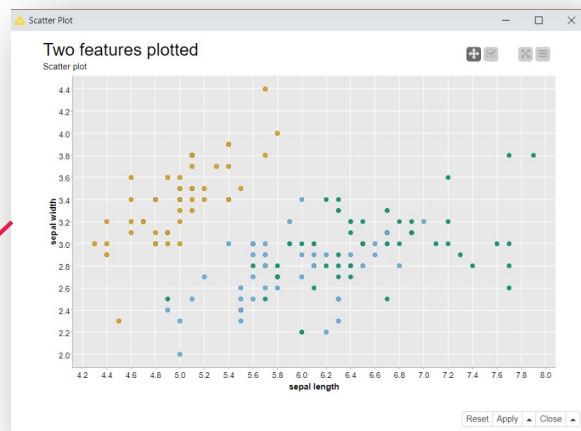
Node Views

Scatter Plot

Two features plotted

Two feature visualization

- 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 - 10:83:0:10 - Scatter Plot (Two features)

File Edit Hilite Navigation View

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	4.3	1.2	Iris-versicolor	false

Data View

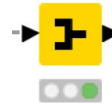
Frequently Used Nodes

CSV Reader File Reader Excel Reader (XLS)



Readers

GroupBy



GroupBy

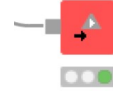
CSV Writer



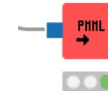
Table Writer



Model Writer

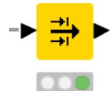


PMML Writer

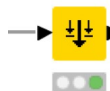


Writers

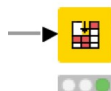
Row Filter



Column Filter

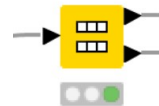


Missing Value
Column Filter



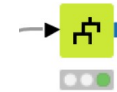
Filters

Partitioning

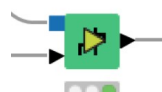


Partitioning

Decision
Tree Learner



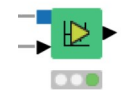
Decision Tree
Predictor



Linear Regression
Learner

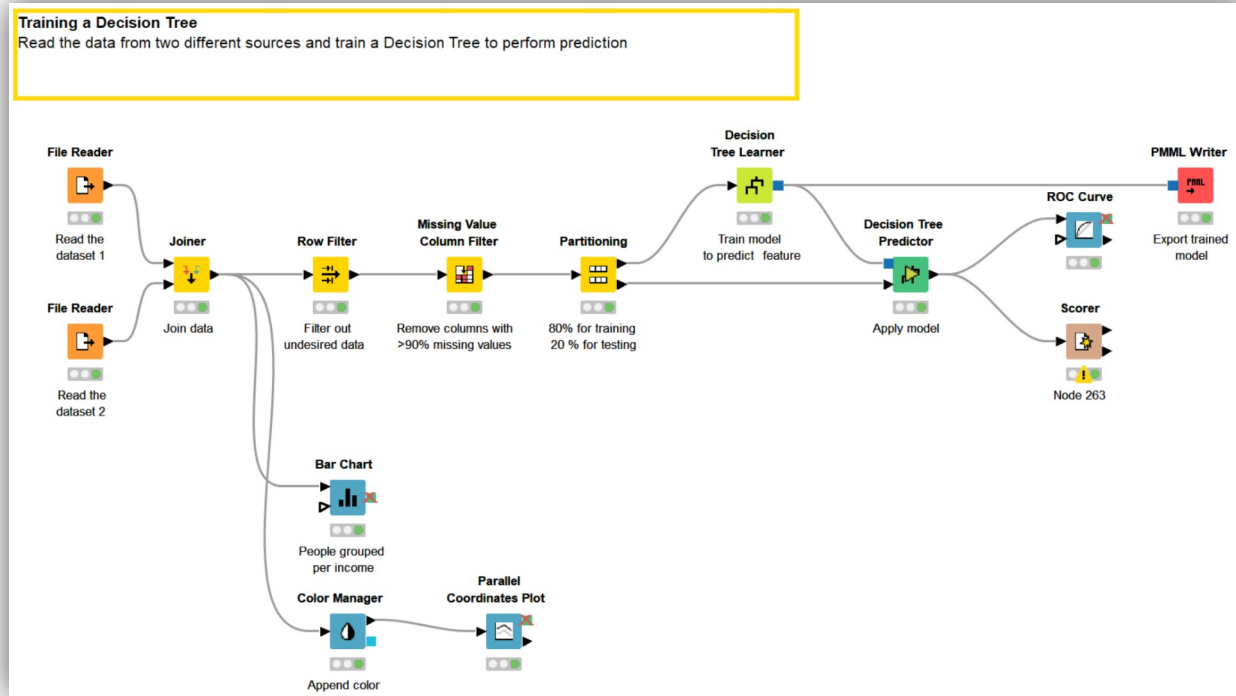


Regression
Predictor



Learners/Predictors

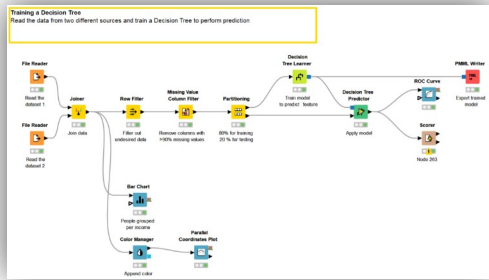
Tidy up workflows



- Workflow can easily become complex and difficult to understand

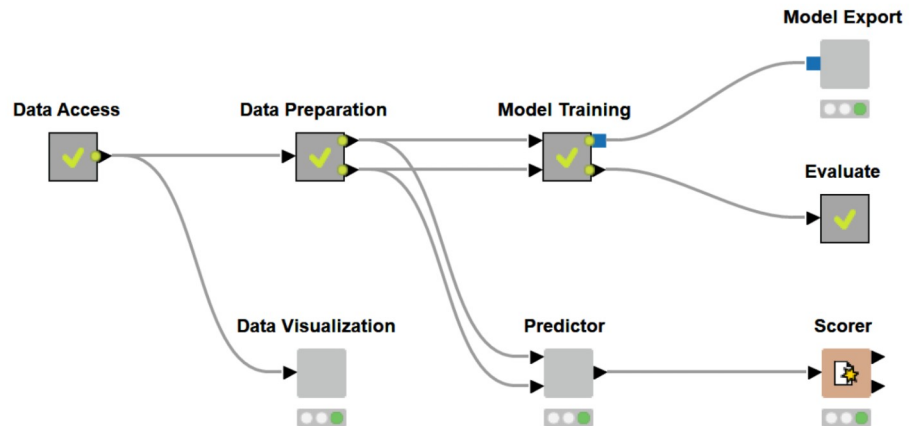
Metanodes and Components

Tidy up workflows



Training a Decision Tree

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



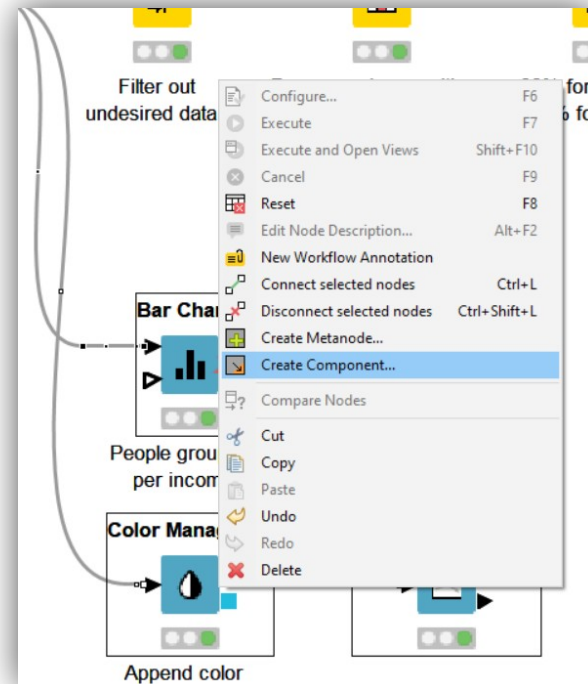
- Metanodes and components can help tidying up, encapsulating nodes performing common operations

Steps to build a component or a metanode

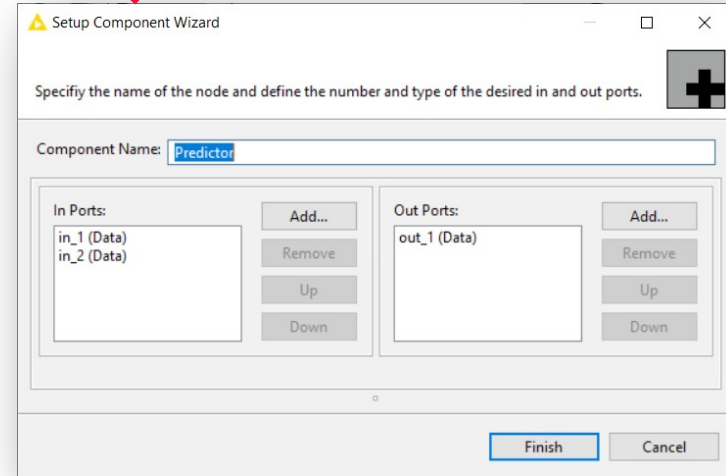
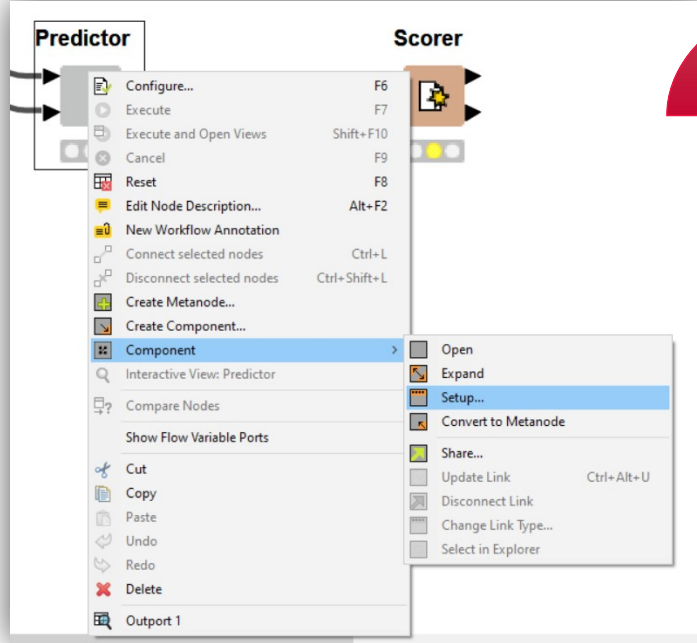
- Select related nodes that you want to group
- Right click
- Select *Create component...* or *Create Metanode...*
- Give it a name

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

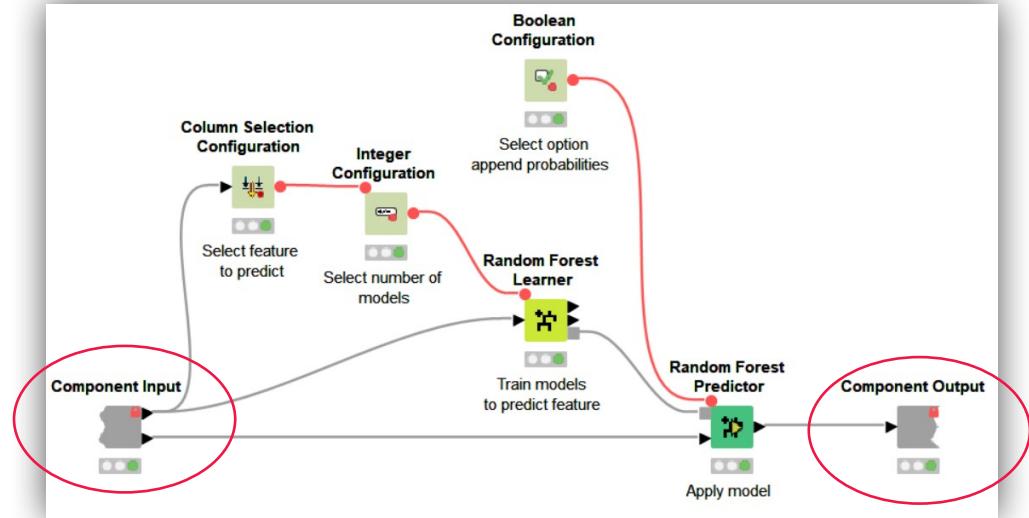
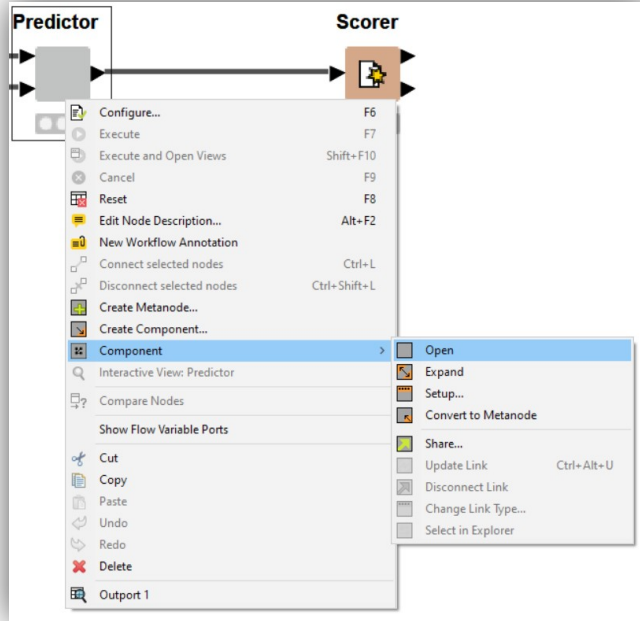


Submenu Component



- 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

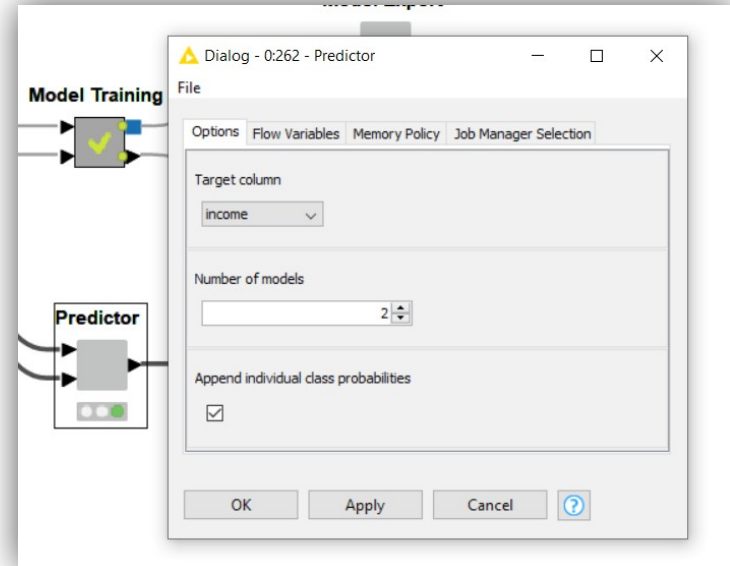
Inside a component



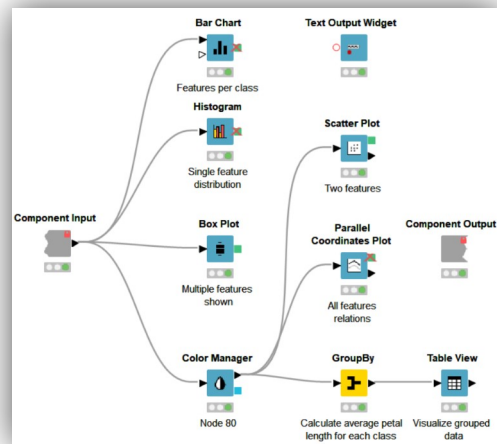
Shortcut:
Ctrl + double click on
component to open its
content

Components 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

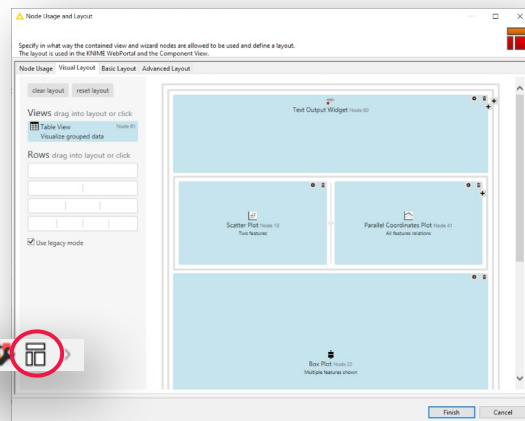


Components Composite View

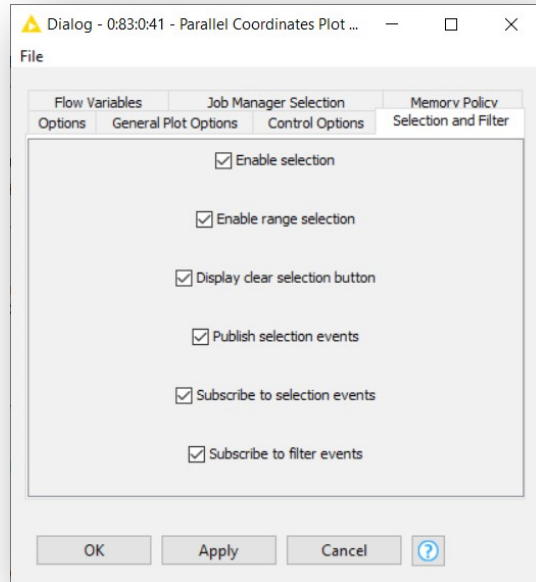


The visualization nodes within the component can be organized to build an interactive composite view

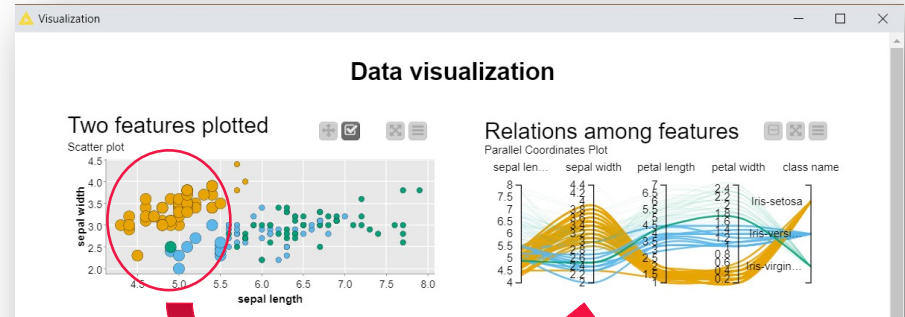
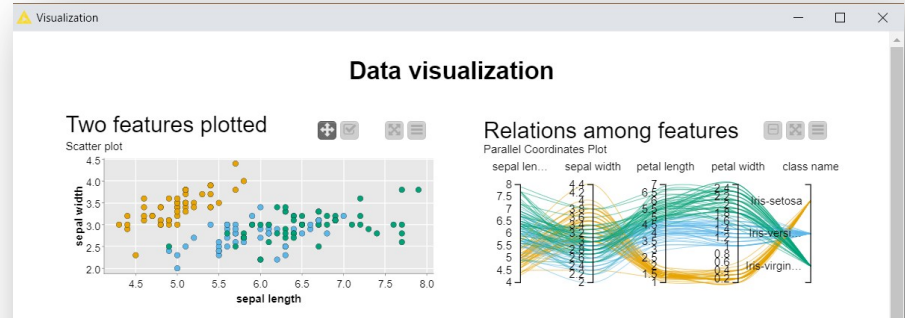
You can organize and reshape the node views from the Visual Layout window (from inside the component, last icon on the toolbar)



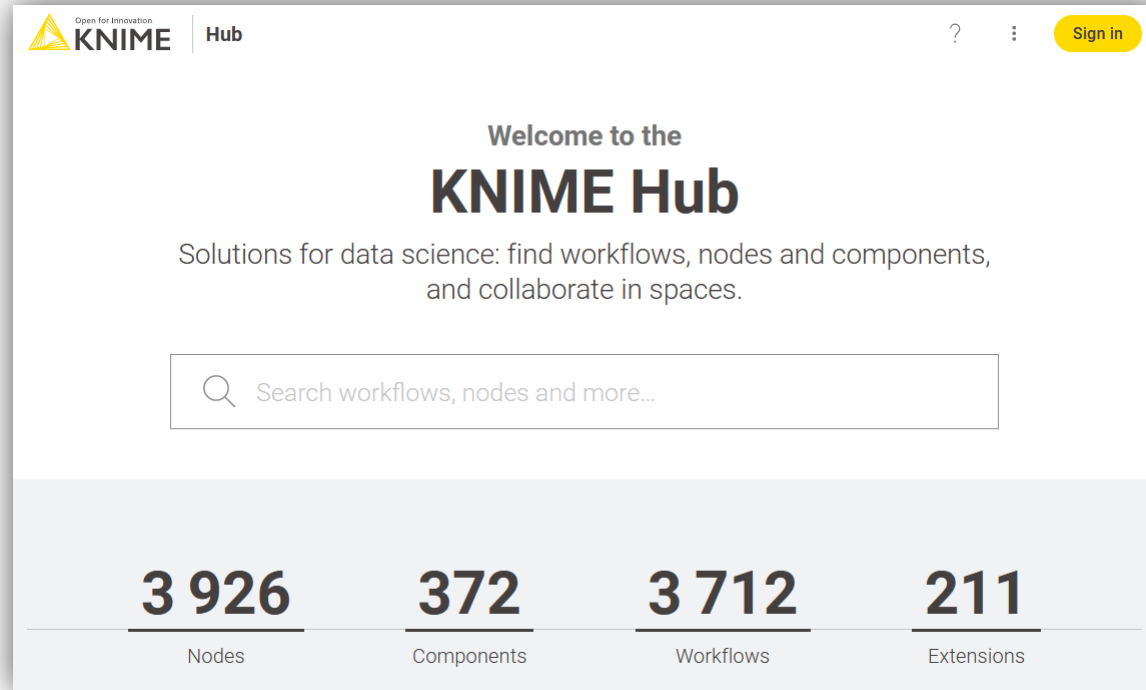
Composite views interactivity



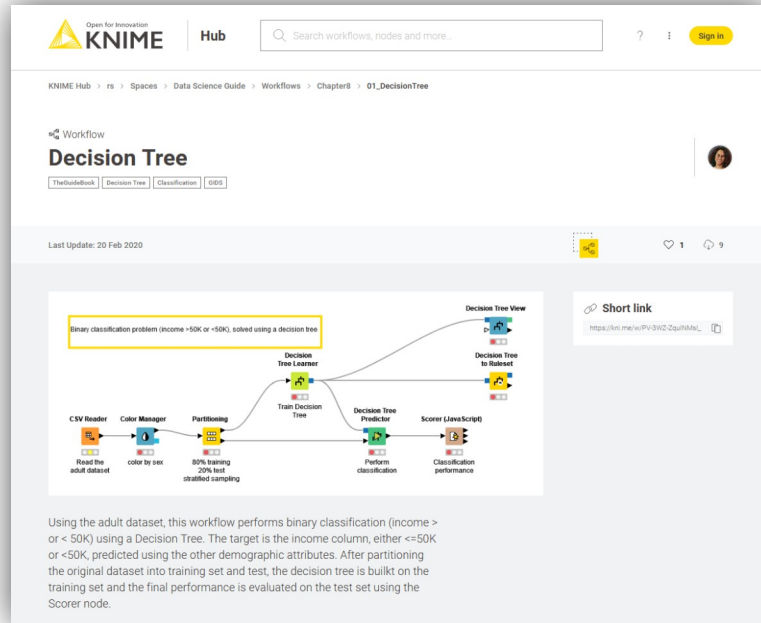
Enable publication and subscription to selection events to make the composite view interactive: data selected in one view are highlighted in the others



KNIME Hub



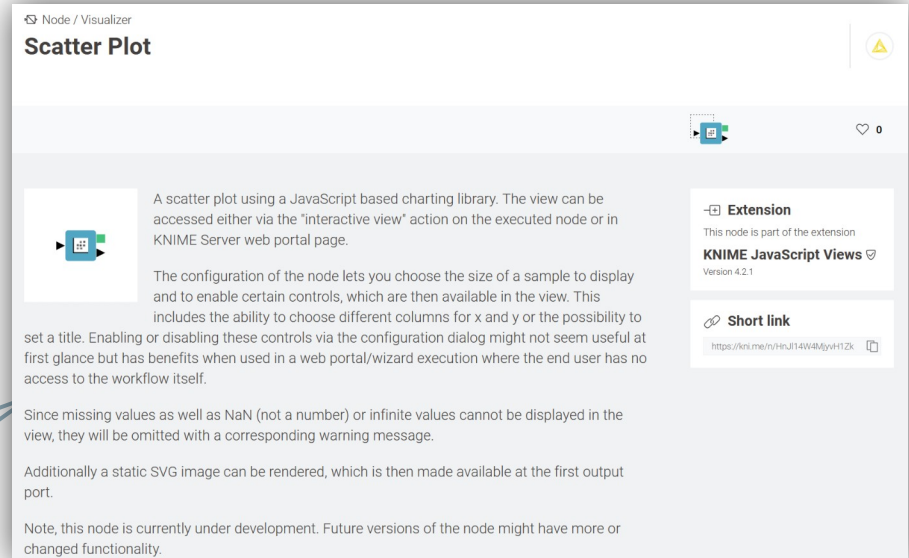
A place to share knowledge about Workflows and Nodes
<https://hub.knime.com>



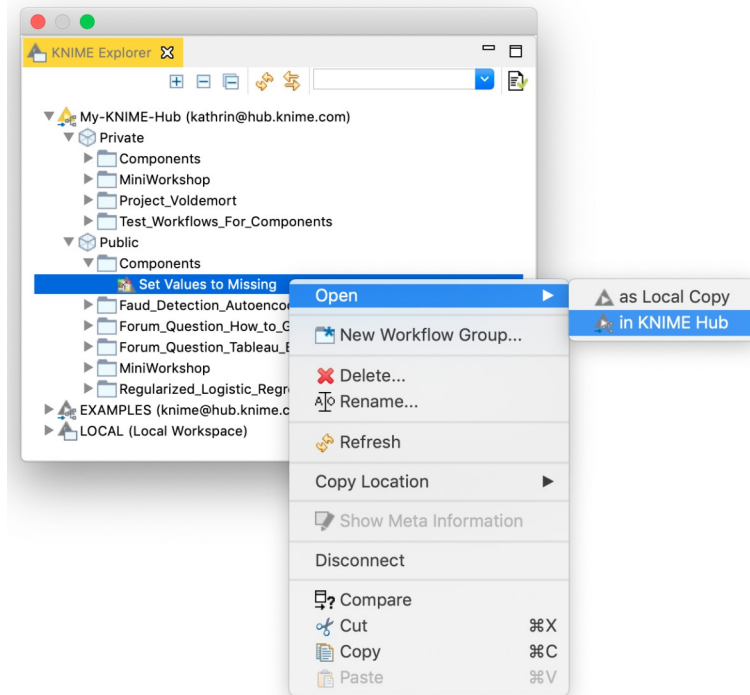
The screenshot shows the KNIME Hub interface. At the top, there's a search bar and a 'Sign in' button. Below the navigation bar, the breadcrumb trail reads: KNIME Hub > rs > Spaces > Data Science Guide > Workflows > Chapter8 > 01_DecisionTree. The main content area displays a workflow titled 'Decision Tree' with tabs for 'The Notebook', 'Decision Tree', 'Classification', and 'GBS'. The workflow diagram shows a sequence of nodes: CSV Reader (labeled 'Read the adult dataset'), Color Manager (labeled 'color by sex'), Partitioning (labeled '80% training, 20% test, stratified sampling'), Train Decision Tree (labeled 'Decision Tree Learner'), Decision Tree Predictor (labeled 'Decision Tree to Infer'), and Scorer (JavaScript) (labeled 'Perform classification', 'Classification performance'). A 'Short link' is provided: <https://knime.me/vPv3W2ZqurMML>. Below the diagram, a text block explains: '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 the 'Scatter Plot' node. The title bar reads 'Node / Visualizer' and 'Scatter Plot'. The main content area features a small icon of the node and 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.' Below this, it states: '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.' A note at the bottom says: '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. Additionally a static SVG image can be rendered, which is then made available at the first output port. Note, this node is currently under development. Future versions of the node might have more or changed functionality.' On the right side, there's an 'Extension' section: 'KNIME JavaScript Views' (Version 4.2.1) and a 'Short link' <https://knime.me/rvHrU14W48yH1Zk>.



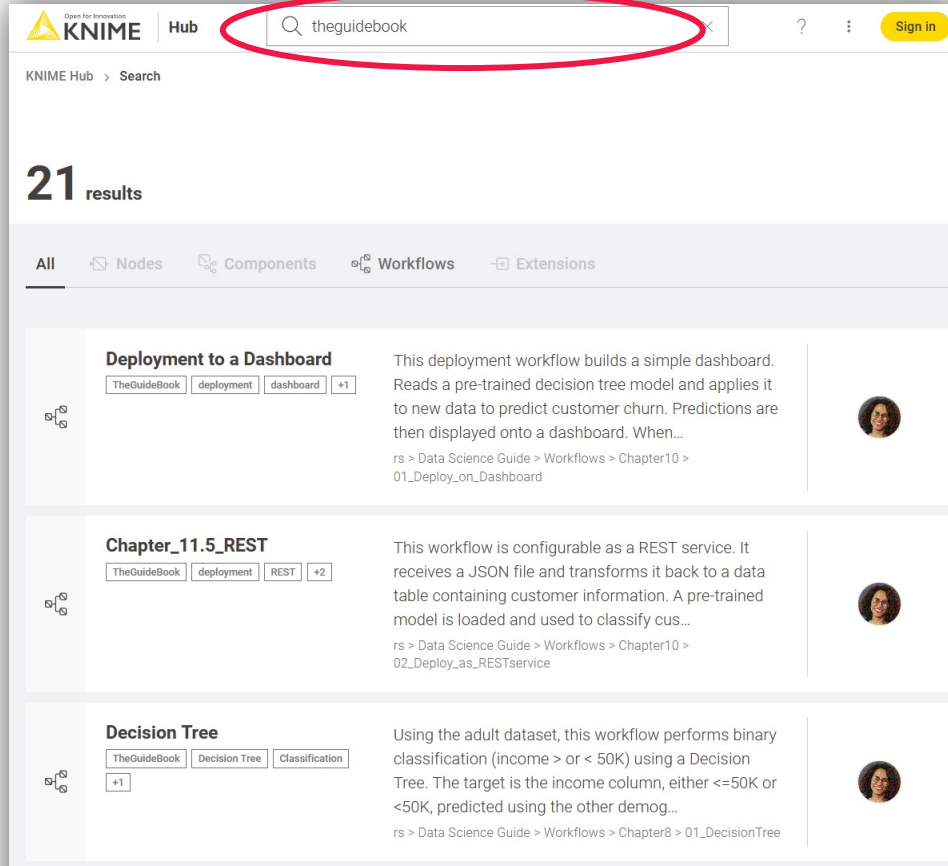
– Private Space

- 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 Hub

Downloading and importing a workflow from the KNIME Hub



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

Downloading and importing a workflow from the KNIME 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

Short link

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 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

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 Analytics Platform

File Edit View Node Search Help

KNIME Explorer

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

KNIME Cheat Sheets

Cheat Sheet: Machine Learning with KNIME Analytics Platform

SUPERVISED LEARNING

CLASSIFICATION

NUMERIC PREDICTION

TEXT CLASSIFICATION

UNSUPERVISED LEARNING

CLUSTERING

ASSOCIATION

SEQUENCE ANALYSIS

TIME SERIES ANALYSIS

ENSEMBLE LEARNING

ENGINING

ROUTING

TRAINING

DEPLOY

Cheat Sheet: Control and Orchestration with KNIME Analytics Platform

CONTROL

EXPLORE

ANALYZE

KNIME

Cheat Sheet: Building a KNIME Workflow for Beginners

READ

EXPLORE

TRANSFORM

DEPLOY

KNIME

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

Cheat Sheet: Building a KNIME Workflow for Beginners

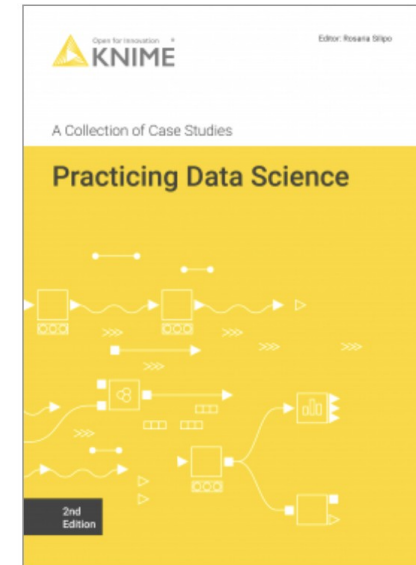
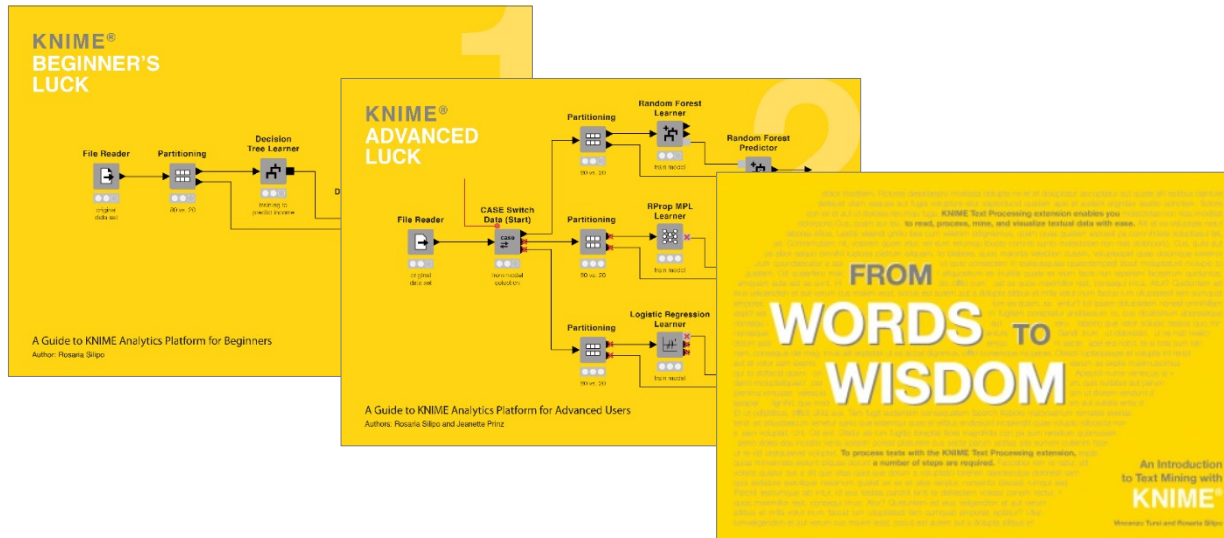
EXPLORE

ANALYZE

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

Hub Blog Forum Events Use Cases Careers Contact [Download](#) 

SOFTWARE / PRICING / COMMUNITY / LEARNING / PARTNERS / ABOUT

Home > Learning > KNIME Self-Paced Courses

[/ Getting Started](#)
[/ FAQ](#)
[/ Learning Hub](#)
[/ KNIME Self-Paced Courses](#)
[/ KNIME Server E-Learning Course](#)
[/ Documentation](#)
[/ Events and Courses](#)
[/ Developers](#)
[/ KNIME Cheat Sheets](#)
[/ White Papers](#)
[/ KNIME Press](#)
[/ KNIME Certification Program](#)
[/ KNIME TV](#)
[/ Changelogs](#)

KNIME Self-Paced Courses

Start here to learn more about data science, data wrangling, text processing, big data, and collaboration and deployment at your own pace and in your own schedule!

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

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

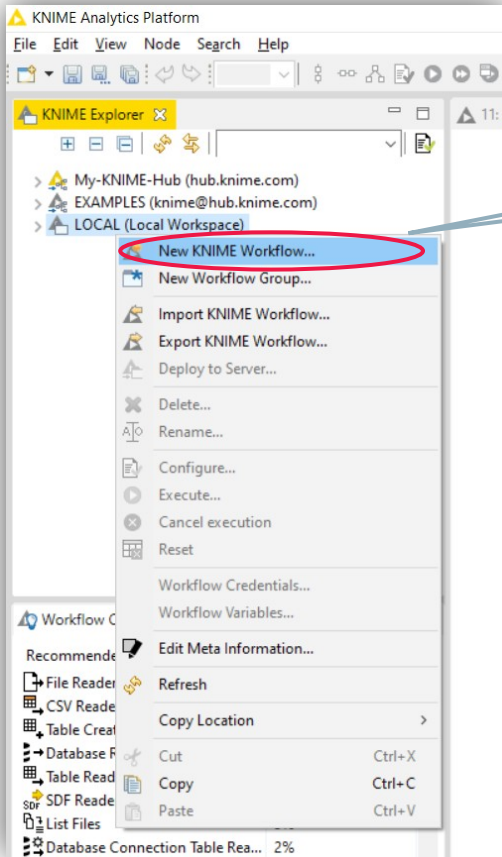
L4
L4-TS Introduction to Time Series Analysis
L4-ML Introduction to Machine Learning Algorithms
L4-TP Introduction to Text Processing
L4-BD Introduction to Big Data with KNIME Analytics Platform
L4-CH Introduction to Working with Chemical Data

Courses

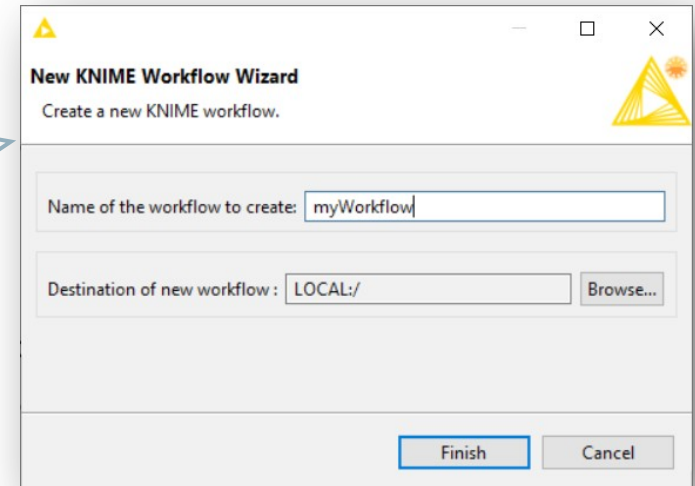
- > [L1-DS] - KNIME Analytics Platform for Data Scientists: Basics
- > [L1-DW] - KNIME Analytics Platform for Data Wranglers: Basics
- > [L2-DS] - KNIME Analytics Platform for Data Scientists: Advanced
- > [L2-DW] - KNIME Analytics Platform for Data Wranglers: Advanced

Build your first Hello Workflow

Create your first workflow



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

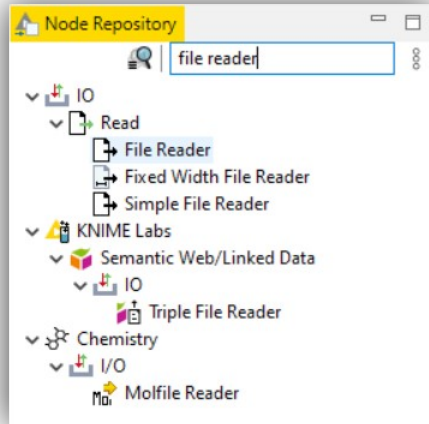


Read the dataset

File Reader

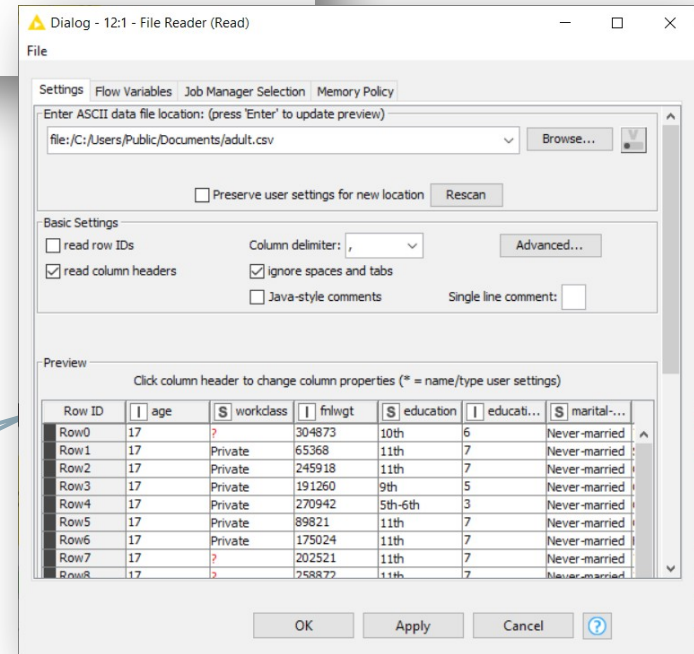


Read
adult dataset

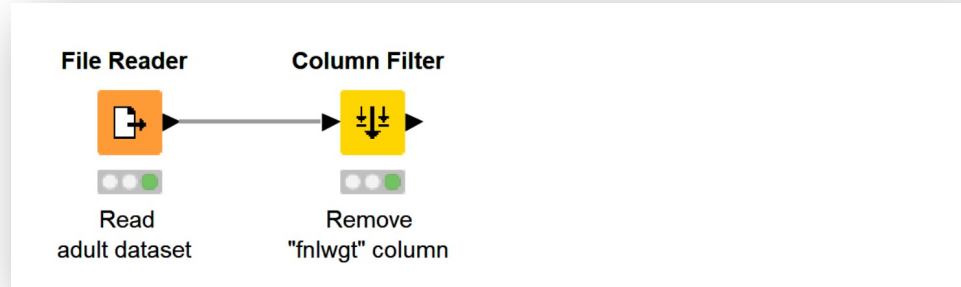


Drag and drop the **File Reader** node from the Node Repository to add it to the workflow

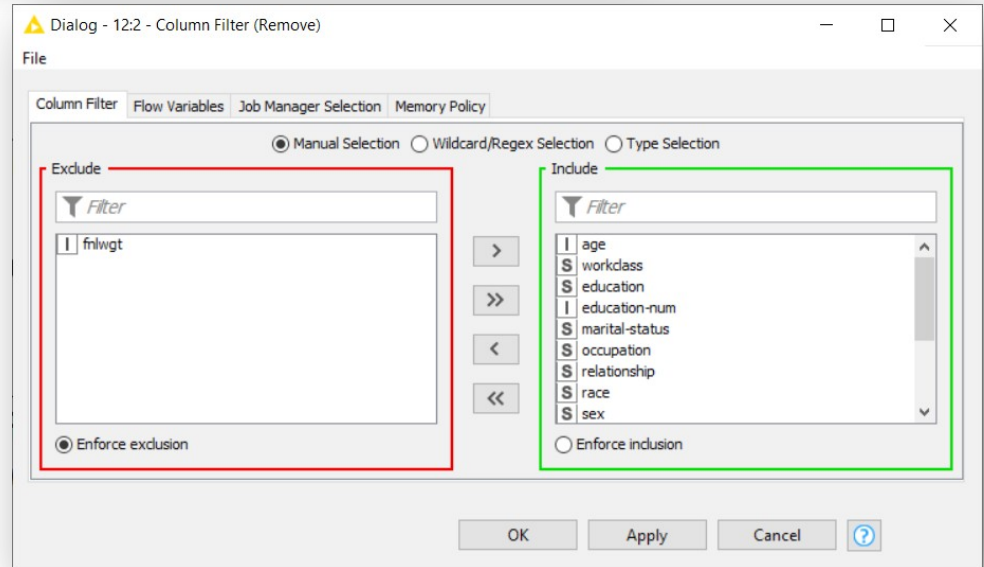
Open the configuration window (double click) and select the file on your machine containing the adult dataset



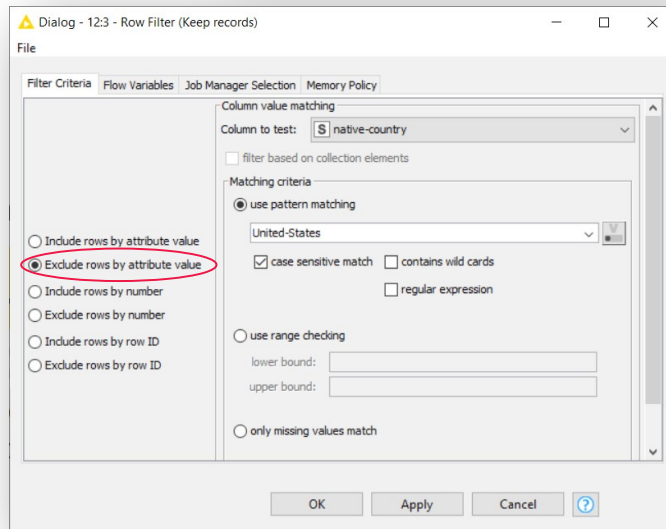
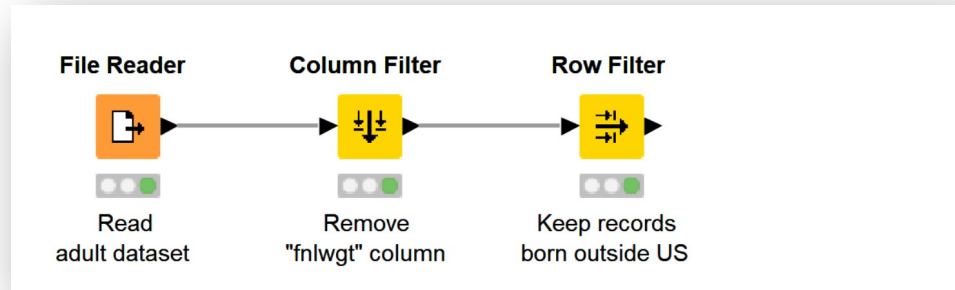
Remove columns



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

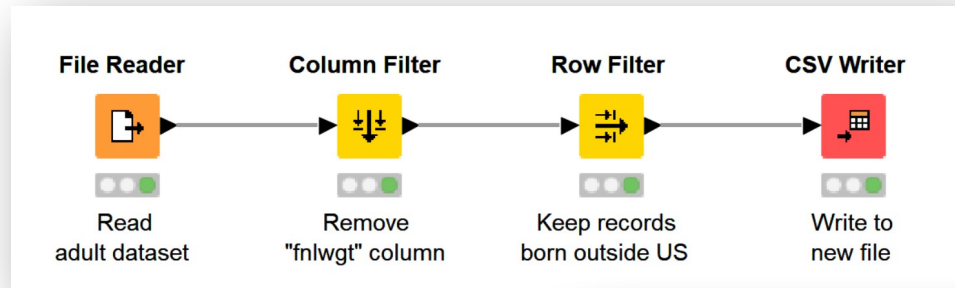


Remove Rows



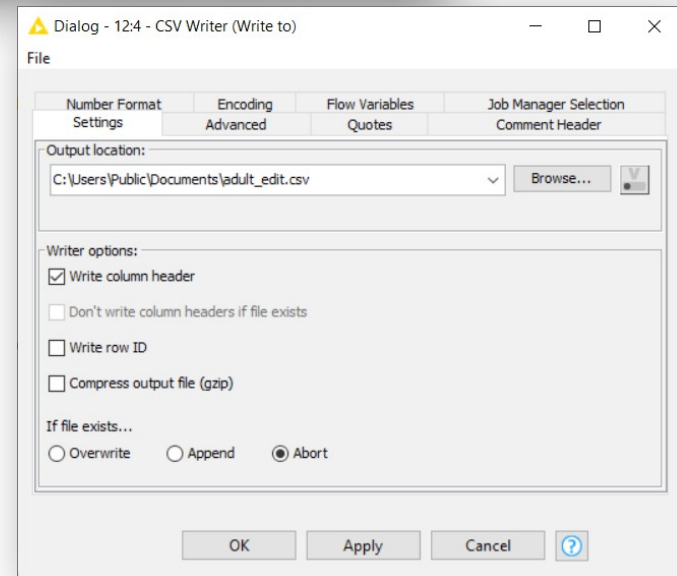
Add a Row Filter node and configure it to only keep entries whose "native-country" value is not "United-States"

Write to new file



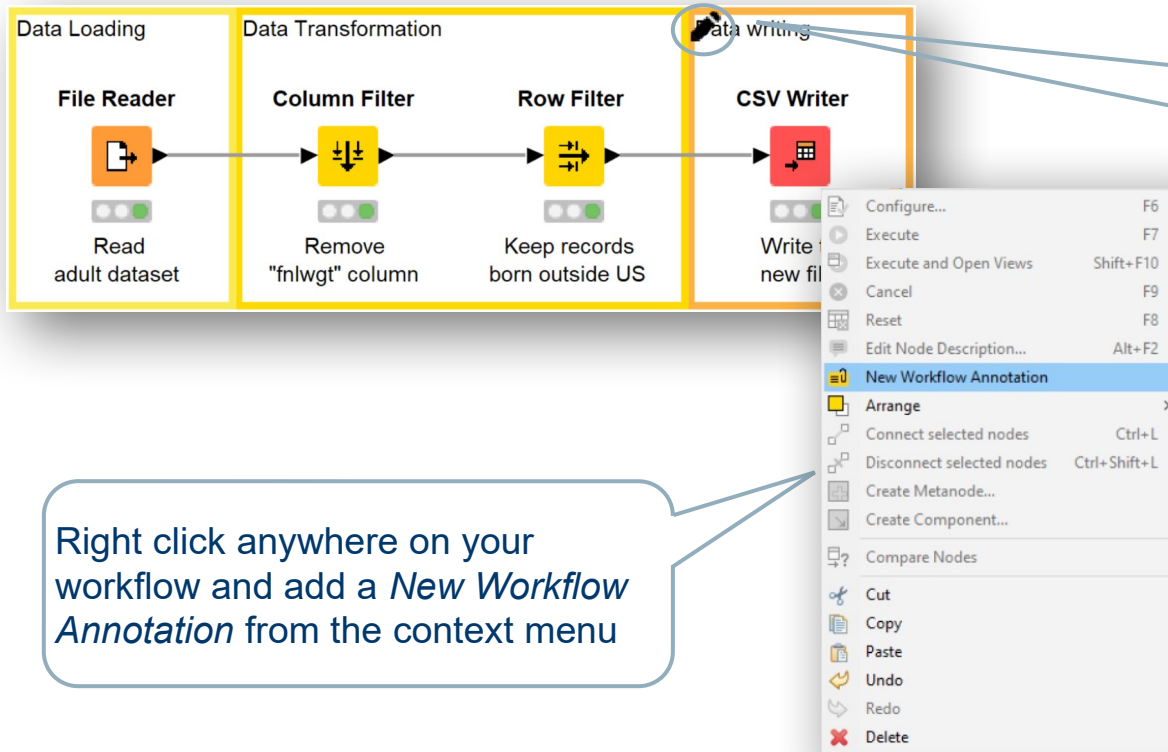
Finally add a CSV Writer node to the pipeline.

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

Tutorial 1: Completed