KNIME Analytics Platform

Michael R. Berthold · Christian Borgelt
Frank Höppner · Frank Klawonn
Rosaria Silipo

Guide to Intelligent Data Science

How to Intelligently Make Use of Real Data

Second Edition

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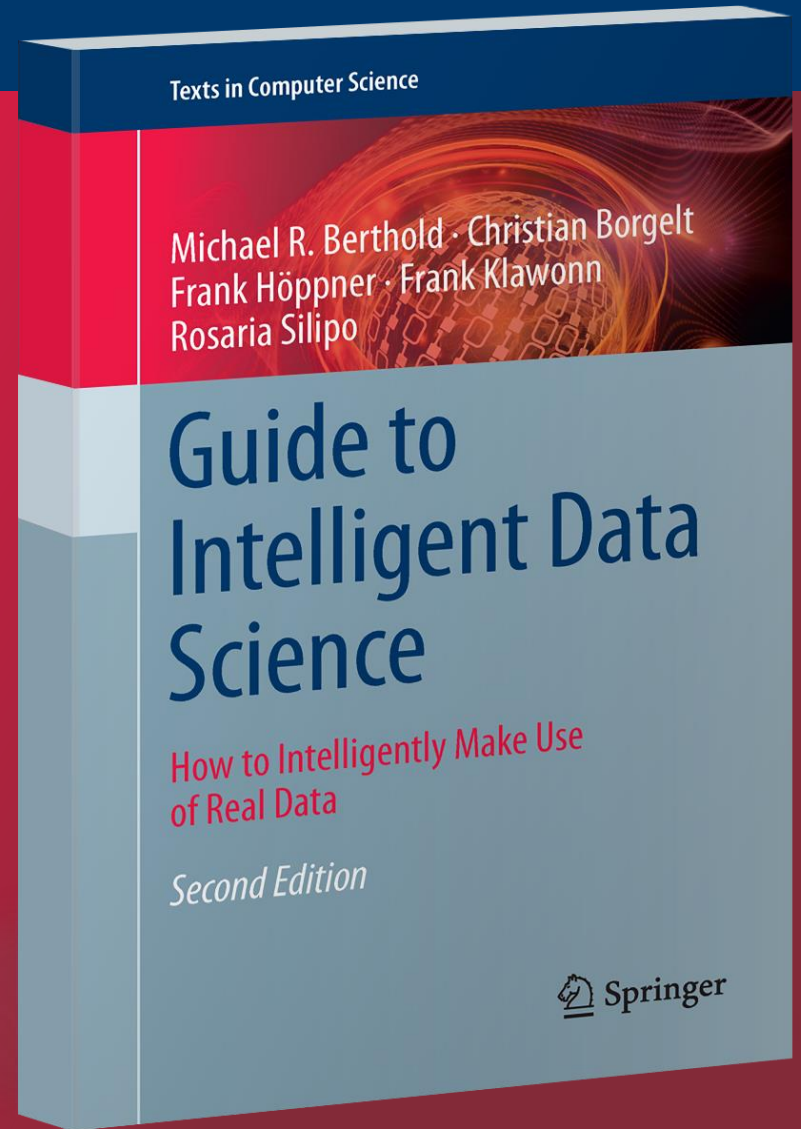
„If the only tool you have is a hammer, you tend to see every problem as a nail"

-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

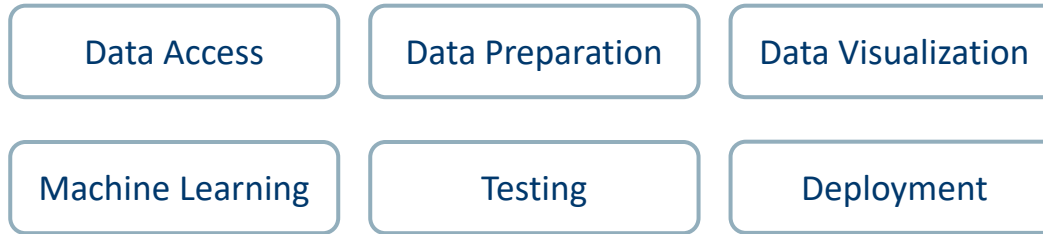- KNIME Community Hub

- 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
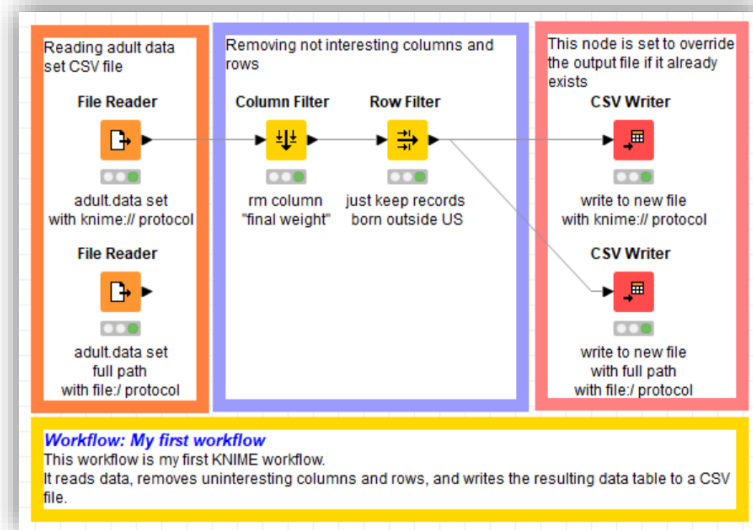
Download and Install

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

Data Access	Data Preparation	Data Visualization
Machine Learning	Testing	Deployment

Open for Innovation
KNIME

- 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
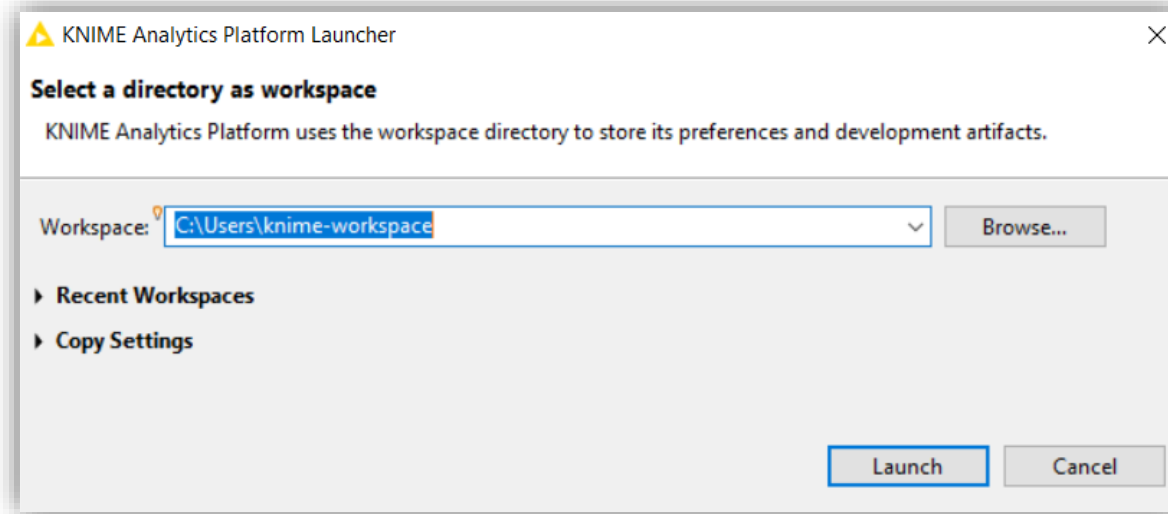
- The workspace is the **folder/directory** in which workflows (and potentially data files) are stored for the current session.

- Workspaces are portable (just like KNIME Analytics Platform)

The KNIME Workbench

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
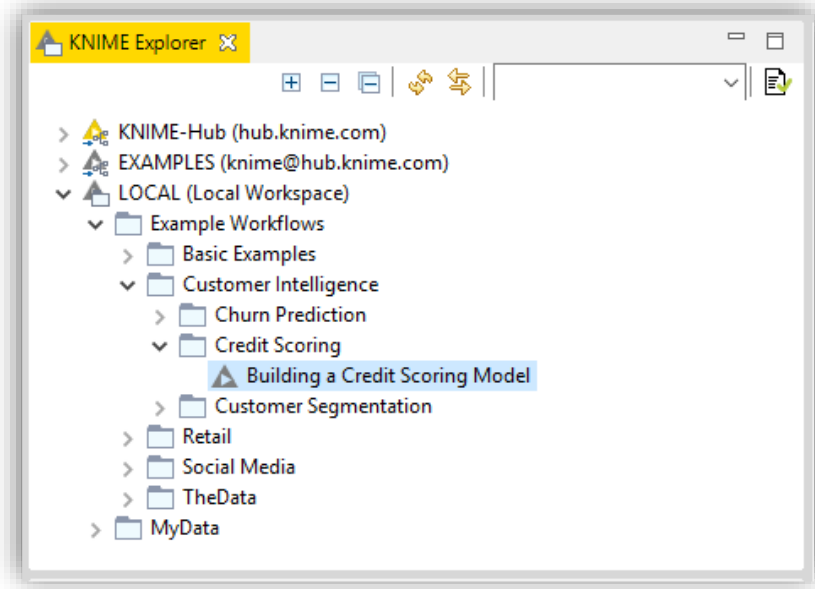
– This panel displays all the workflows in the selected workspace

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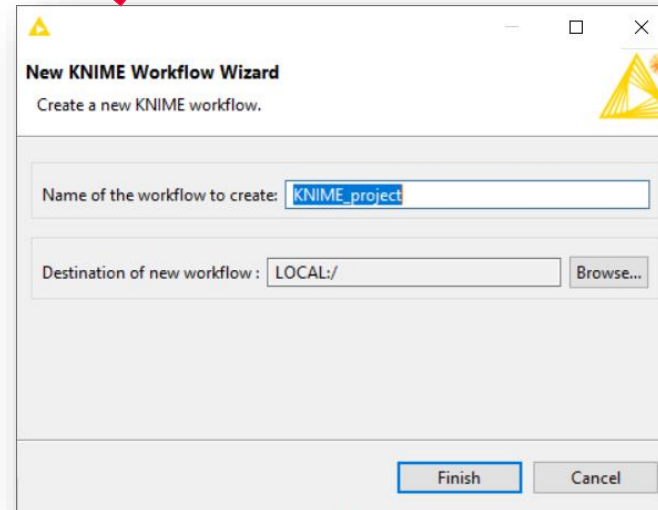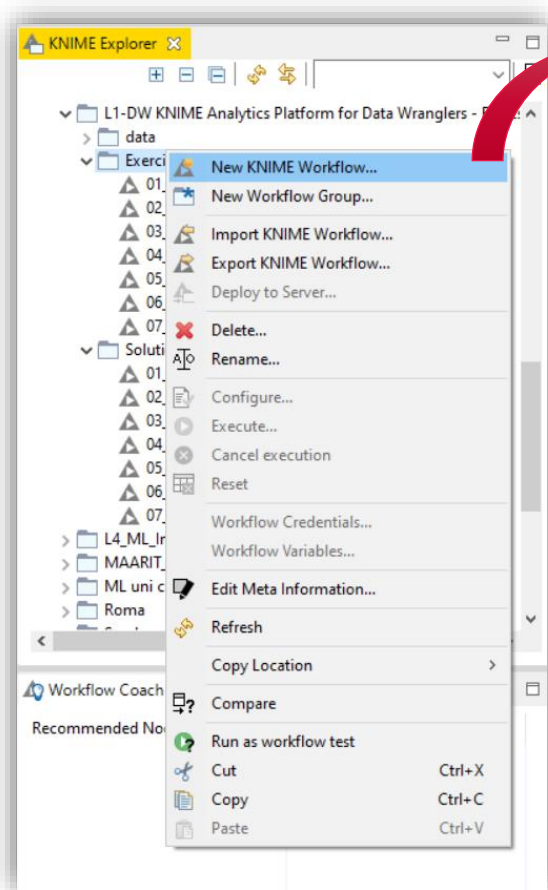
Click anywhere on the KNIME Explorer to create a new workflow or workflow group

Importing and Exporting Workflows

Right-click anywhere in KNIME Explorer to import a workflow
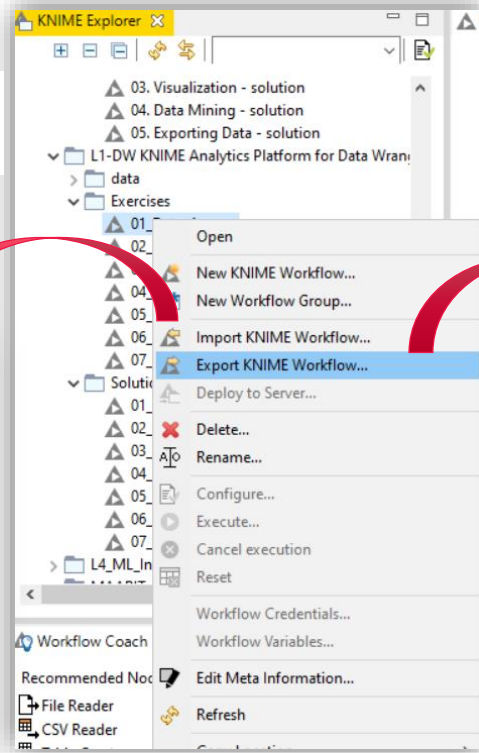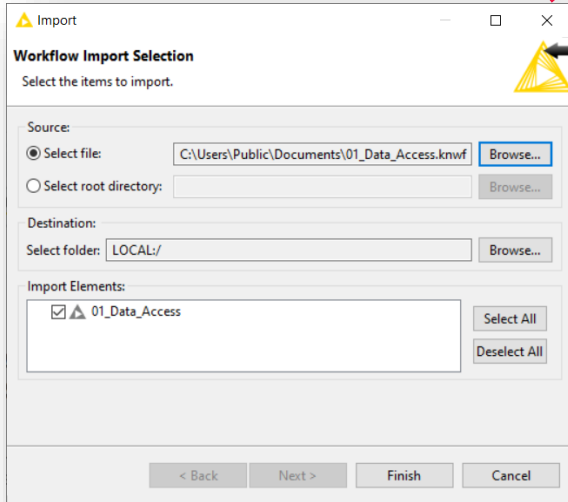
Right-click on a workflow or workflow group to export the selected workflow
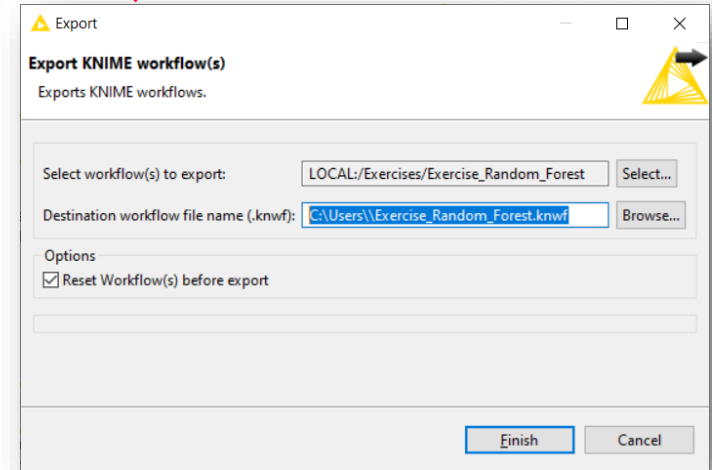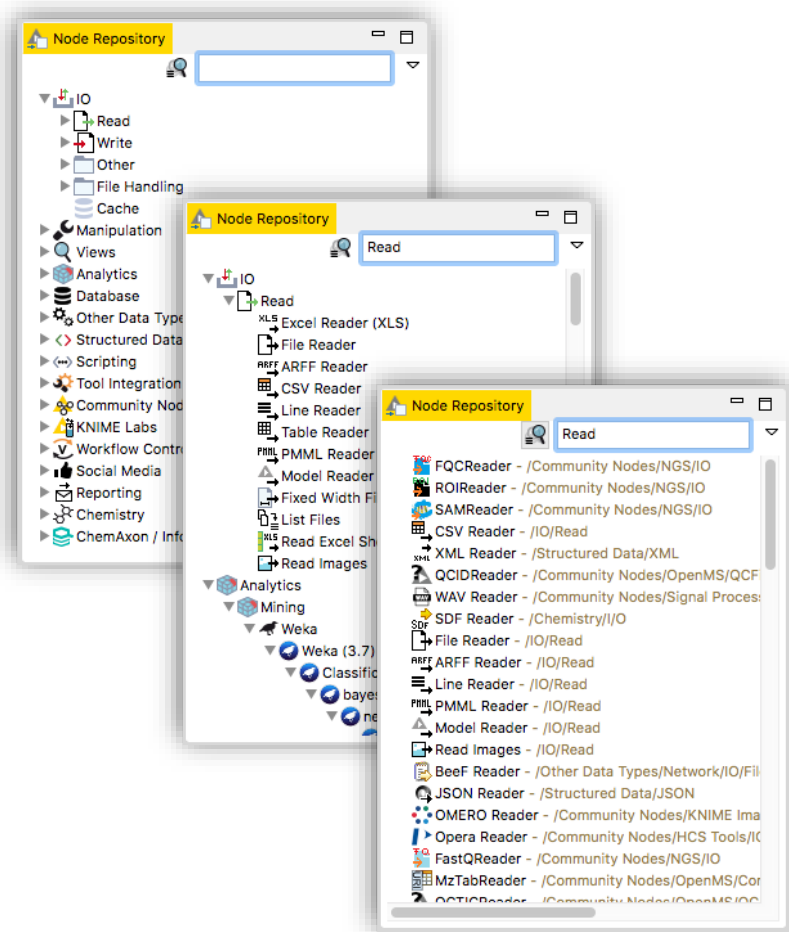
- The Node Repository contains all KNIME nodes - ordered by category with further subcategories.

- 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
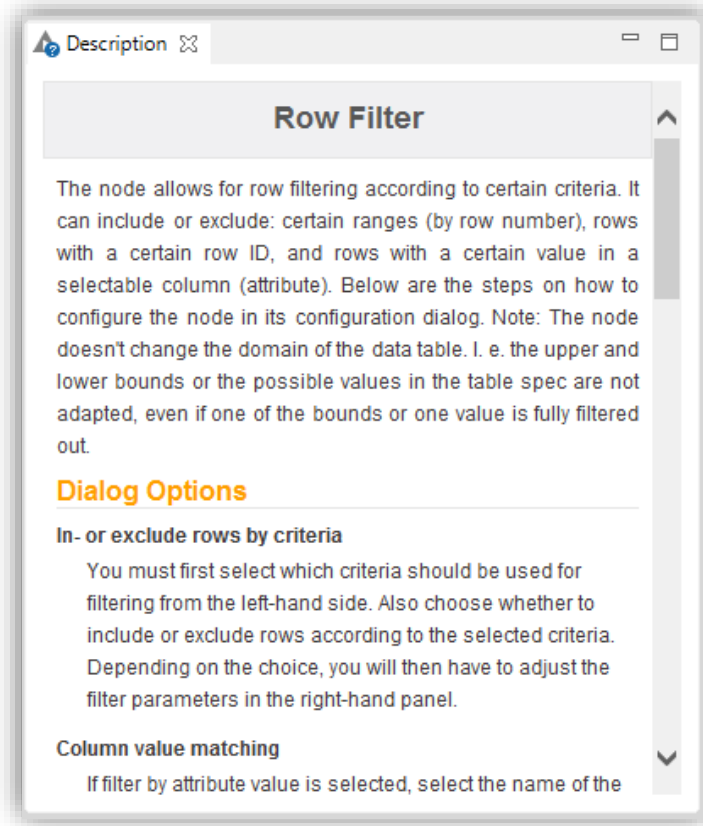
– The Description window gives information about:

- Node Functionality
- Input & Output
- Node Settings
- Ports
- References to literature
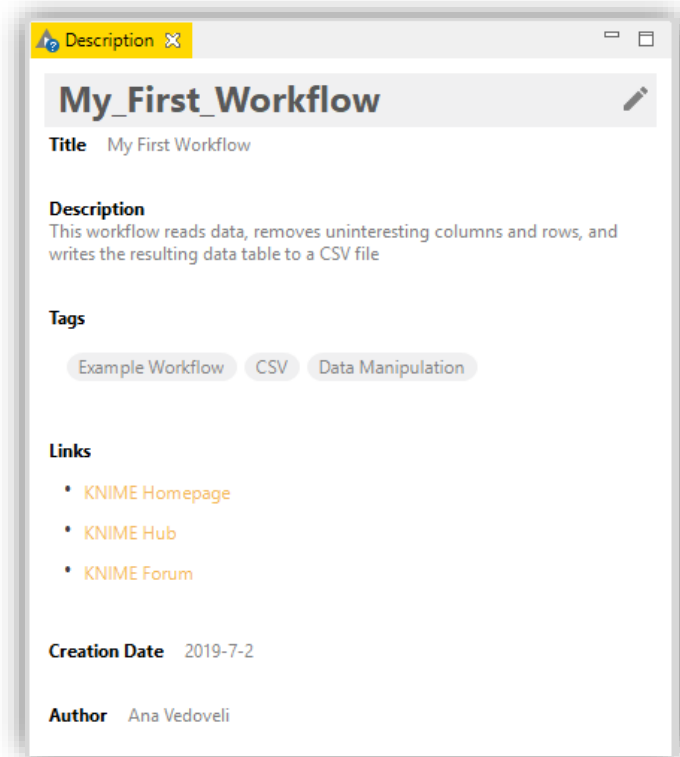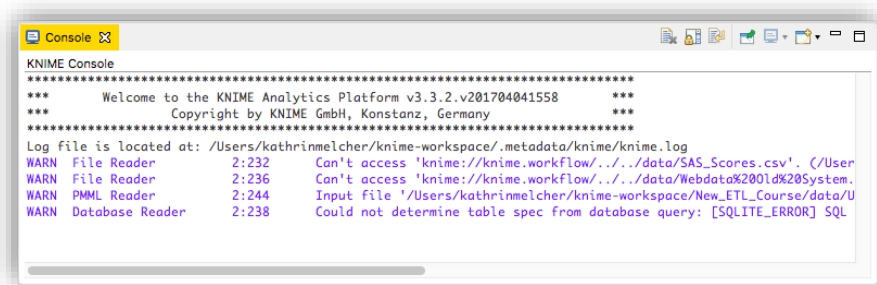
– When selecting the workflow, the Description window gives information about the workflow's:

  – Title

  – Description

  – Associated Tags and Links

  – Creation Date

  – Author

Workflow Coach

– 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
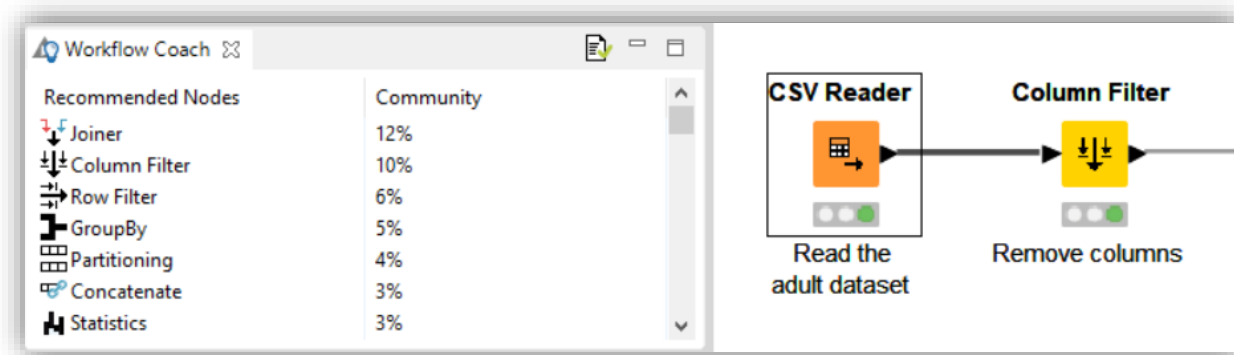
Console and Other views

- Console view prints out error and warning messages about what is going on under the hood
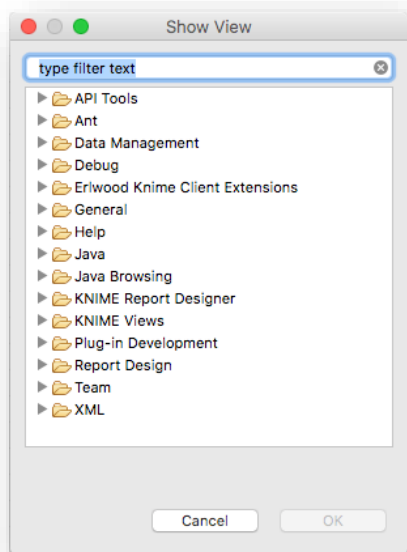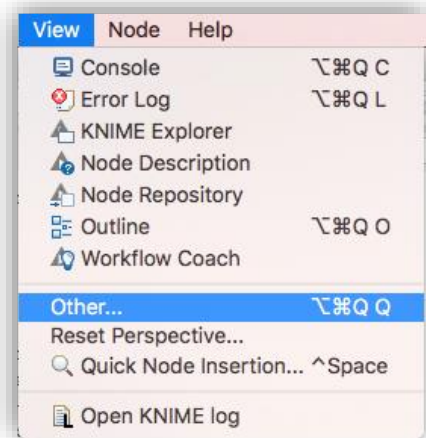
- Click on View and select *Other…* to add additional views

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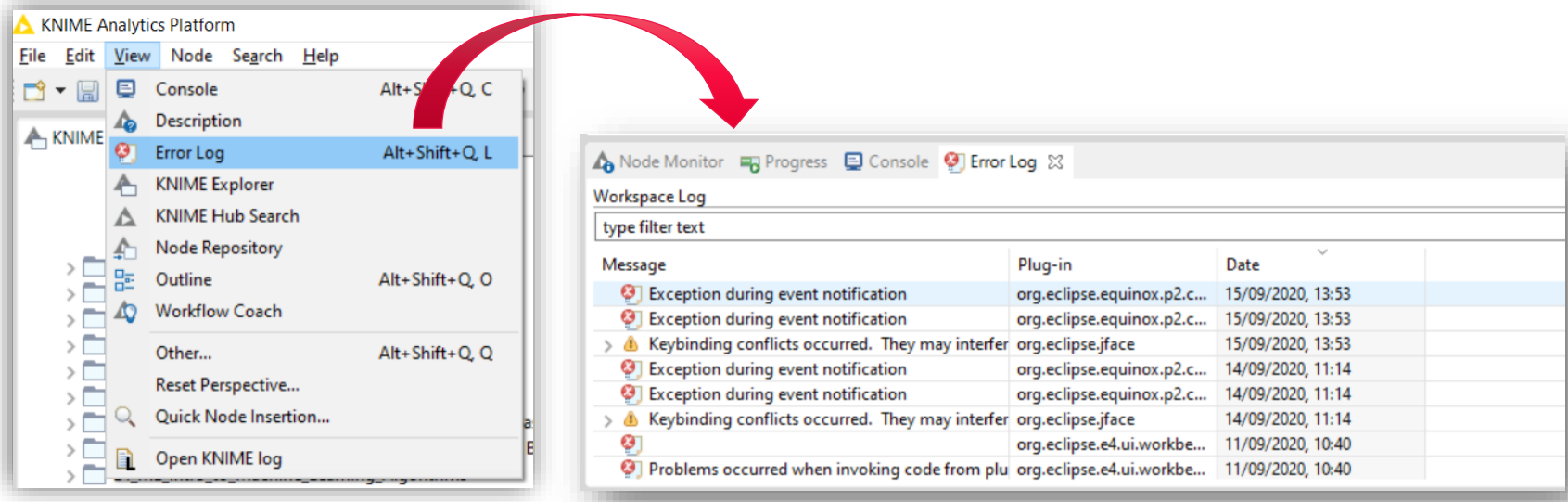
Tip: enabling and checking the Error Log view can help while debugging your project

More on Nodes

- 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
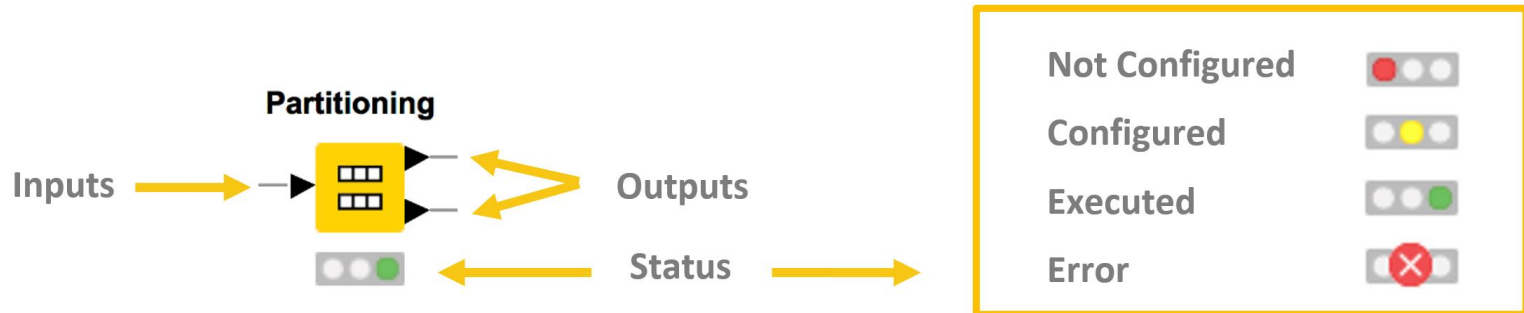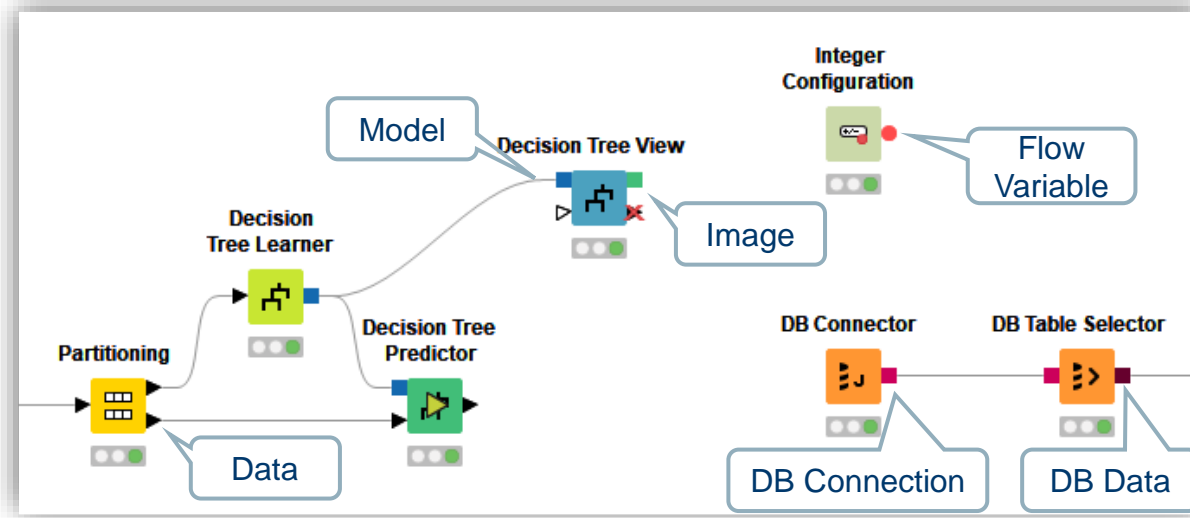
- A pipeline of such nodes makes a **workflow**

- 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

- 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
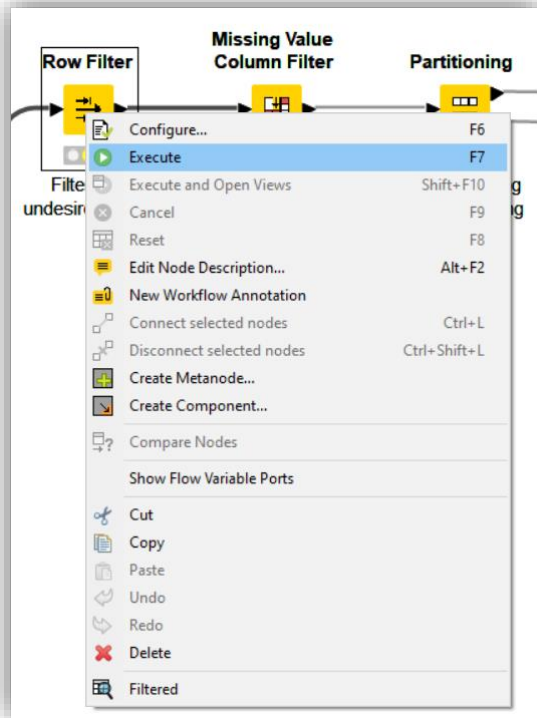
If execution is successful, status shows green light

If execution produces warnings, status show yellow triangle

If execution encounters errors, status shows a red X

Interactive View

Data View

Readers

GroupBy

Writers

Filters

Partitioning

Learners/Predictors

– Workflow can easily become complex and difficult to understand

Metanodes and Components

Tidy up workflows
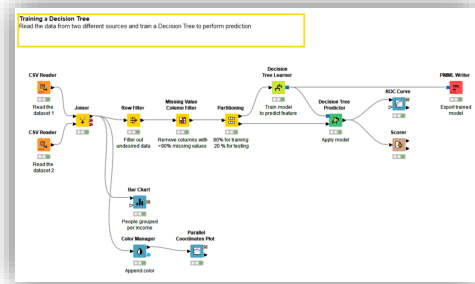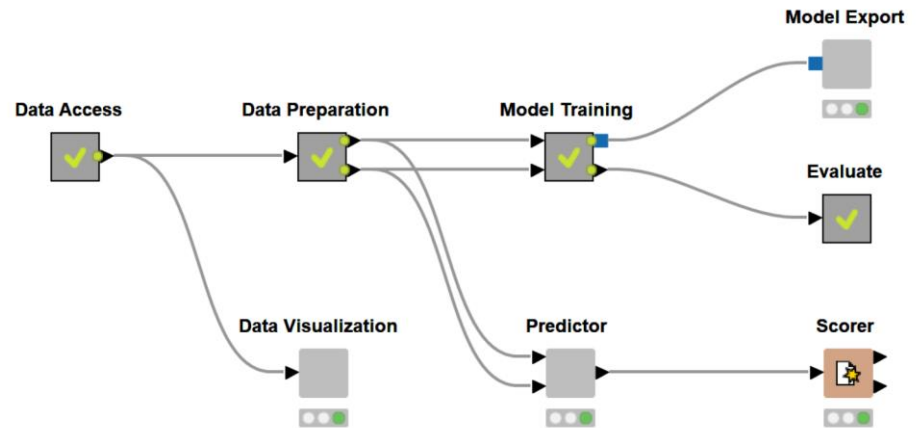
– Metanodes and components can help tidying up, encapsulating nodes performing common operations

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
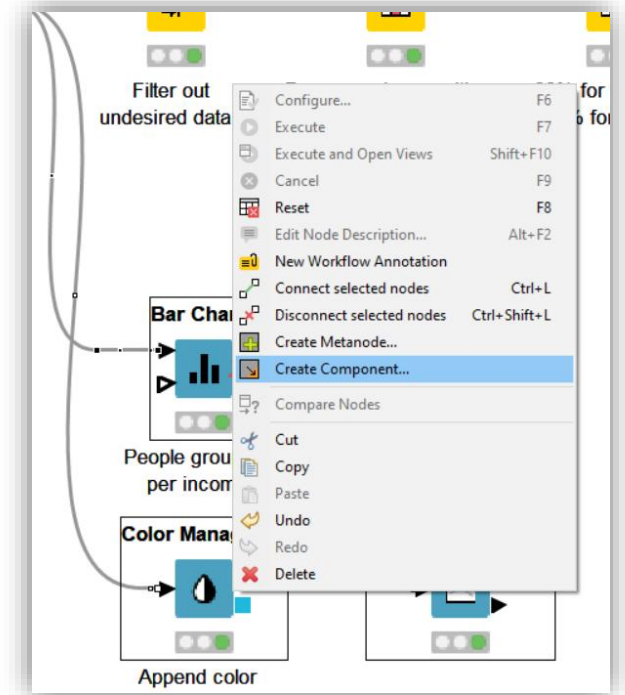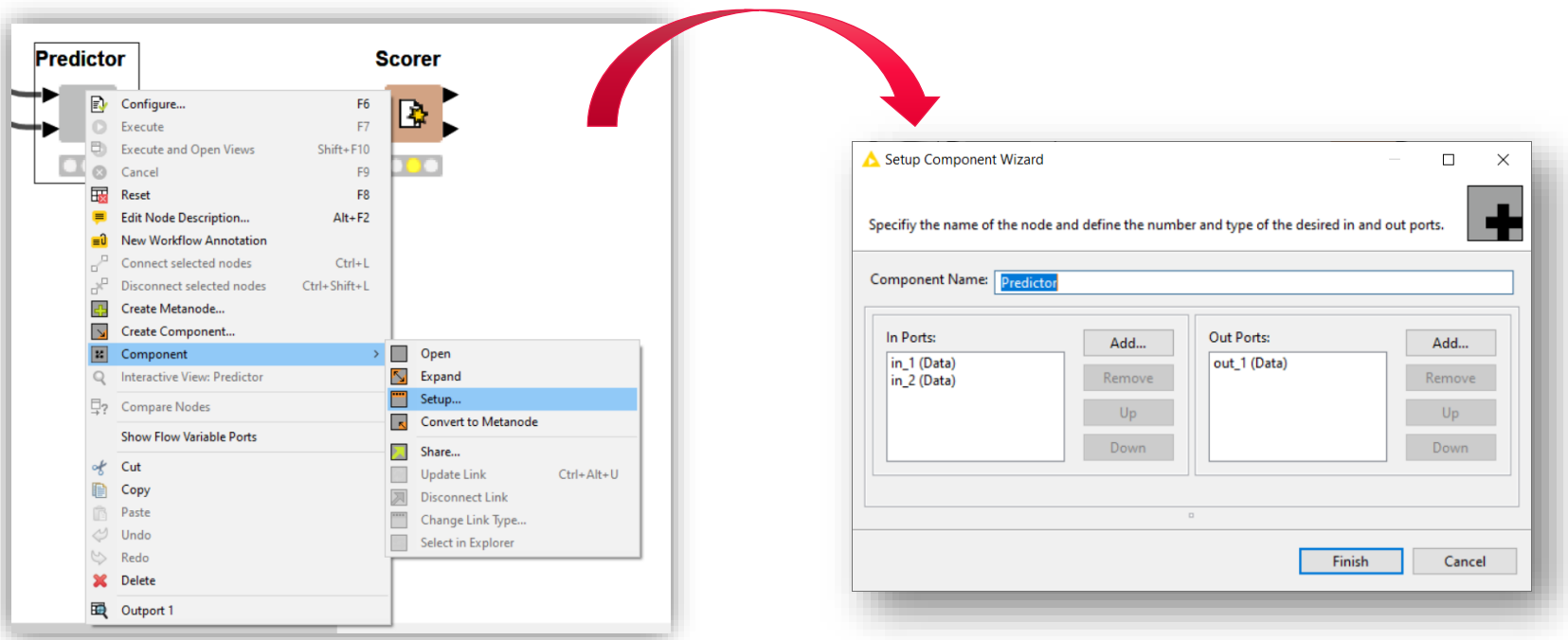
– Right click on the Component and select Setup… from the submenu Component to access further customization settings, such as the component name and the ports
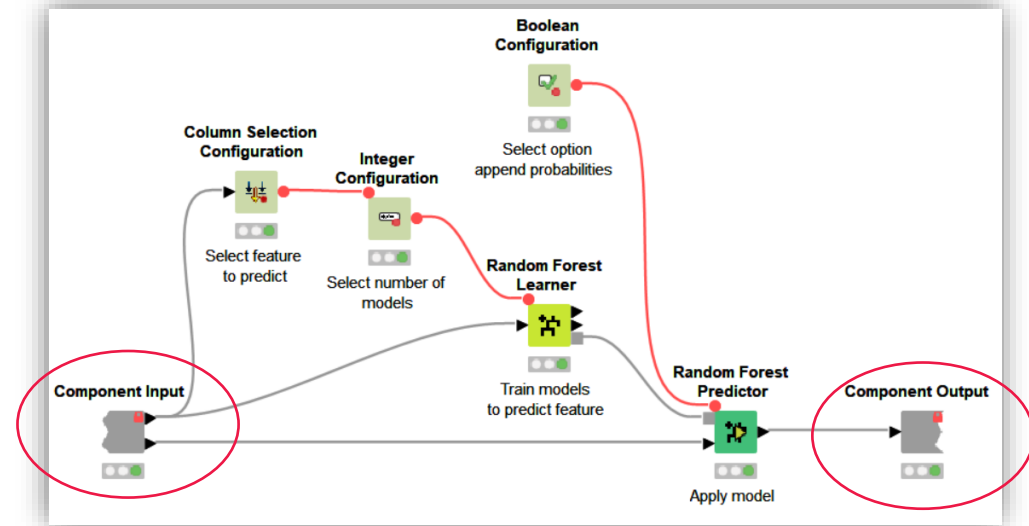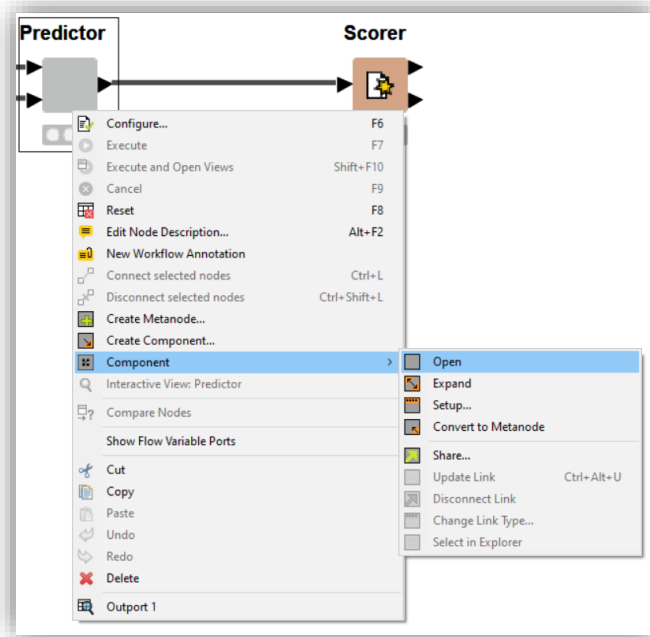
Shortcut:

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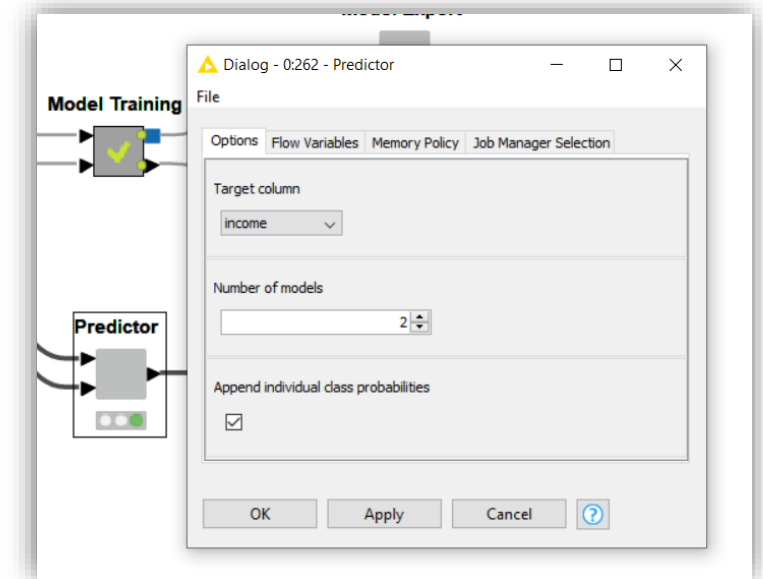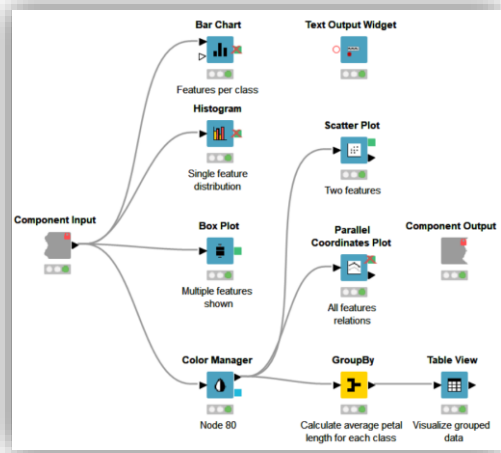
The visualization nodes within the component can be organized to build an interactive composite view

You can organize and reshape the node views from the Visual Layout window (from inside the component, last icon on the toolbar)
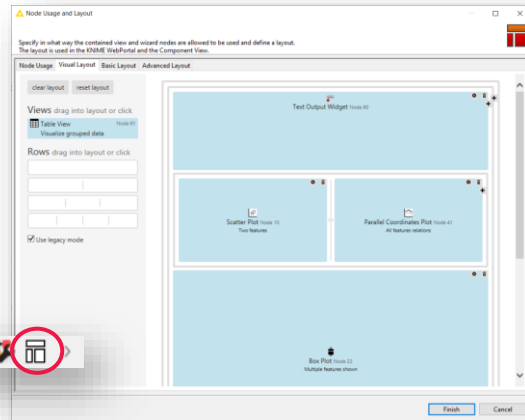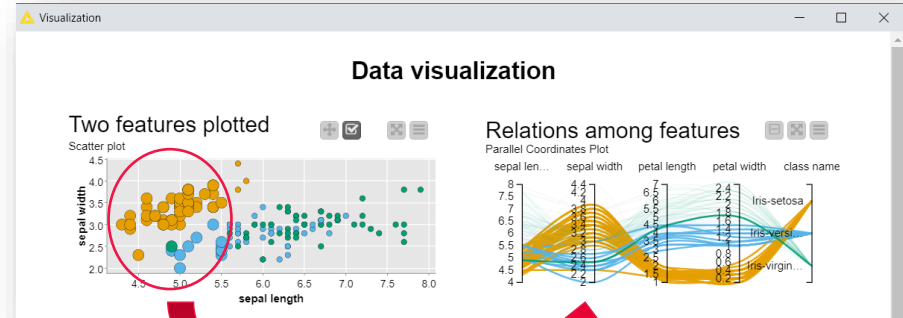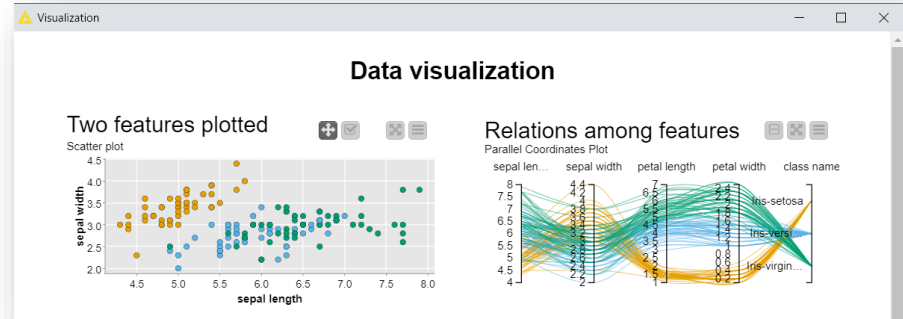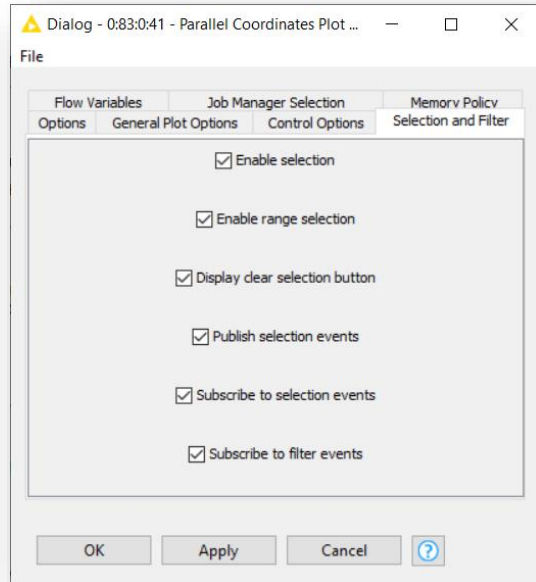
Enable publication and subscription to selection events to make the composite view interactive: data selected in one view are highlighted in the others

KNIME Community Hub
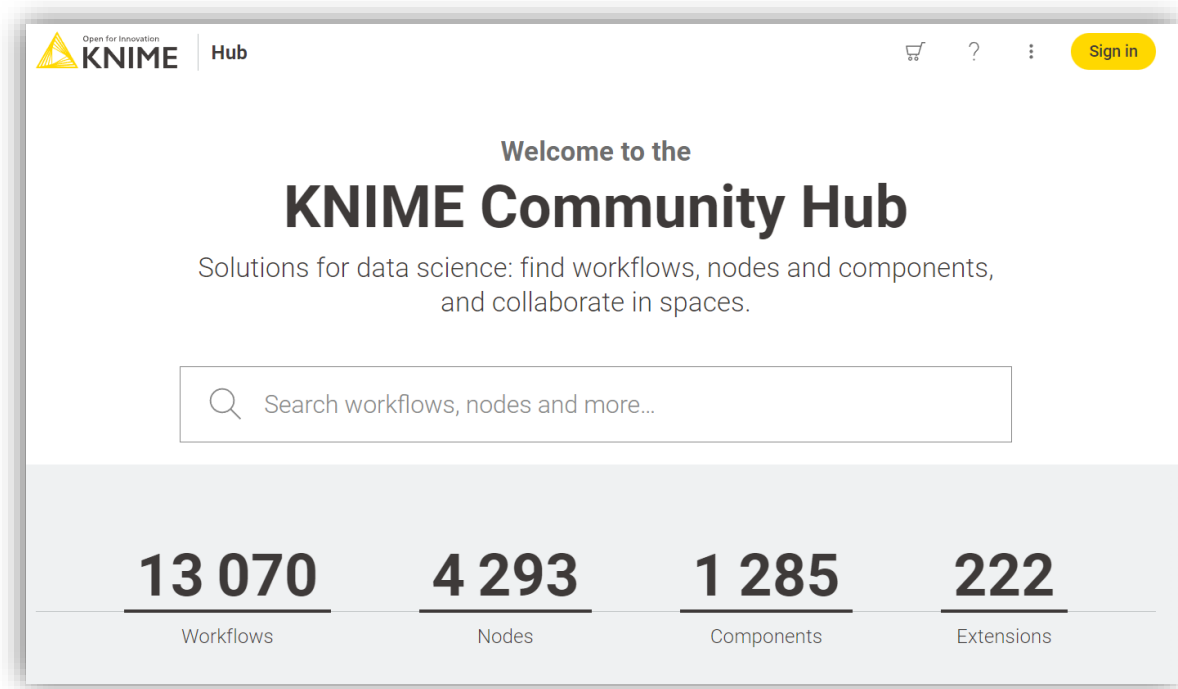
A place to share knowledge about Workflows and Nodes
https://hub.knime.com

KNIME Community Hub

Workflows

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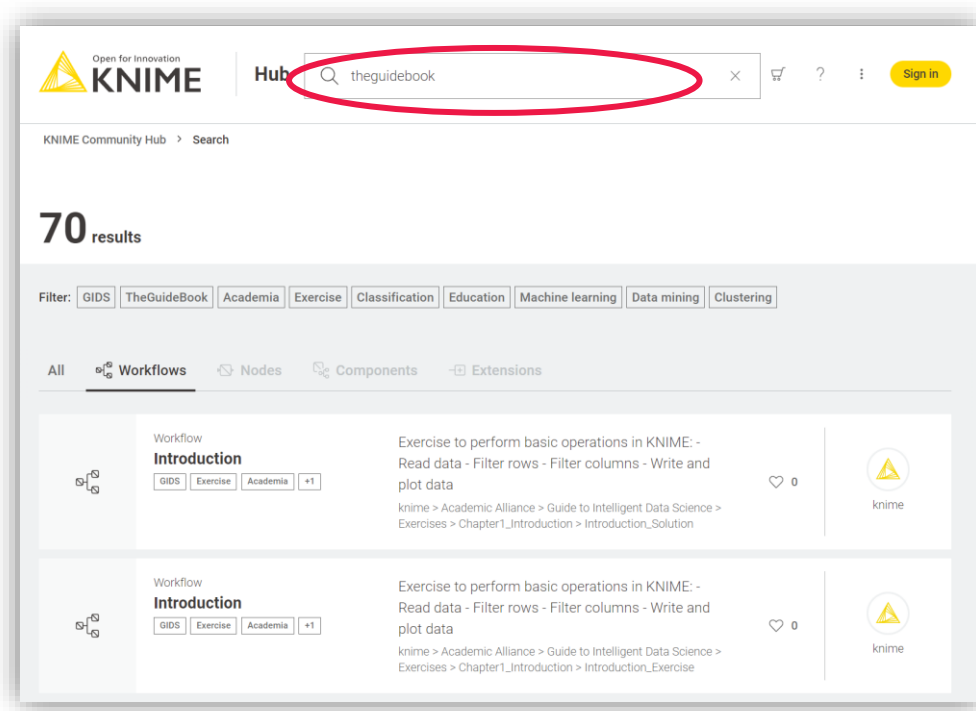
Searching for the Tag "***theguidebook***" will show you all the workflows related to this book

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

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

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

Build your first Hello Workflow

Create your first workflow
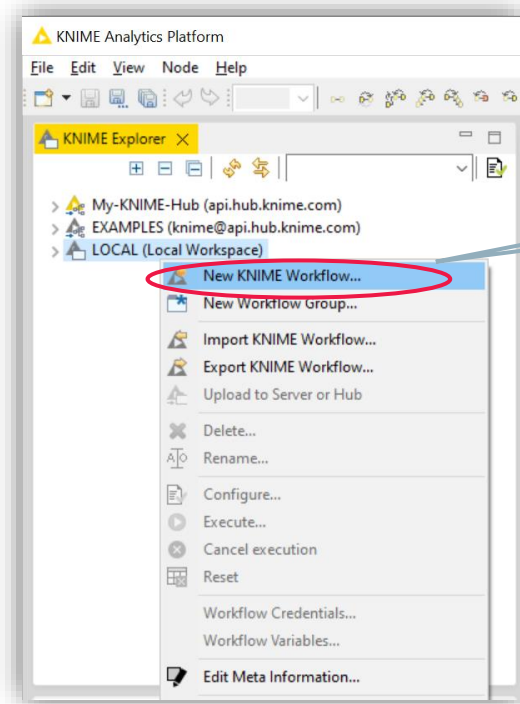
Right Click on the LOCAL folder in the KNIME Explorer and select *New KNIME Workflow*

From the pop up window, insert the name of your first workflow
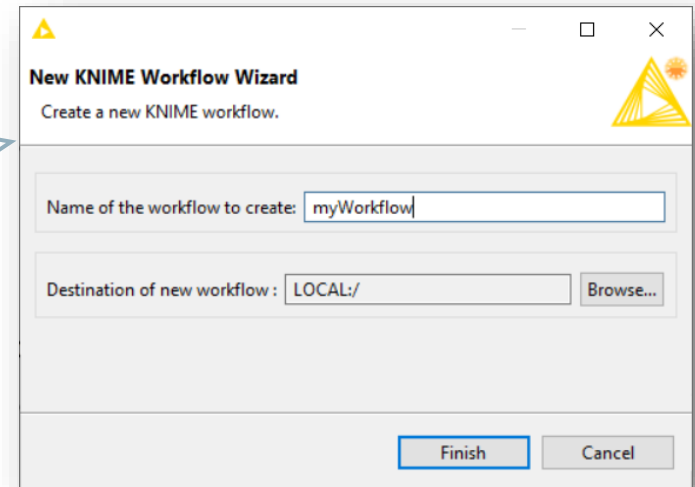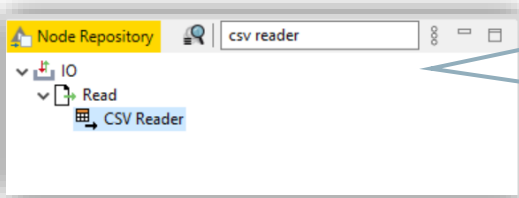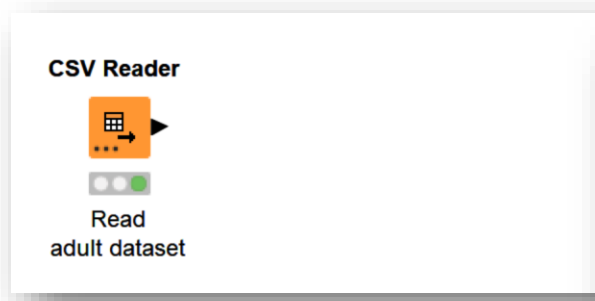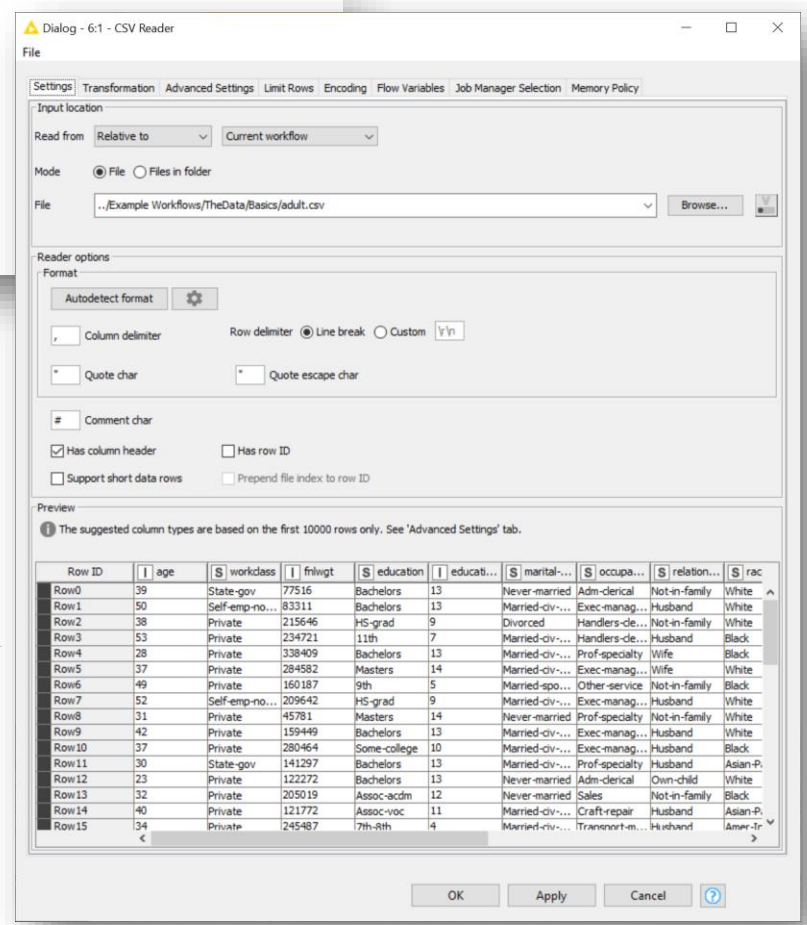
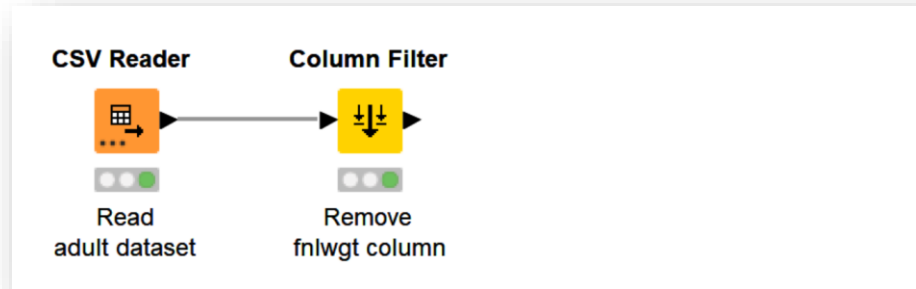Drag and drop the **CSV Reader** node from the Node Repository to add it to the workflow
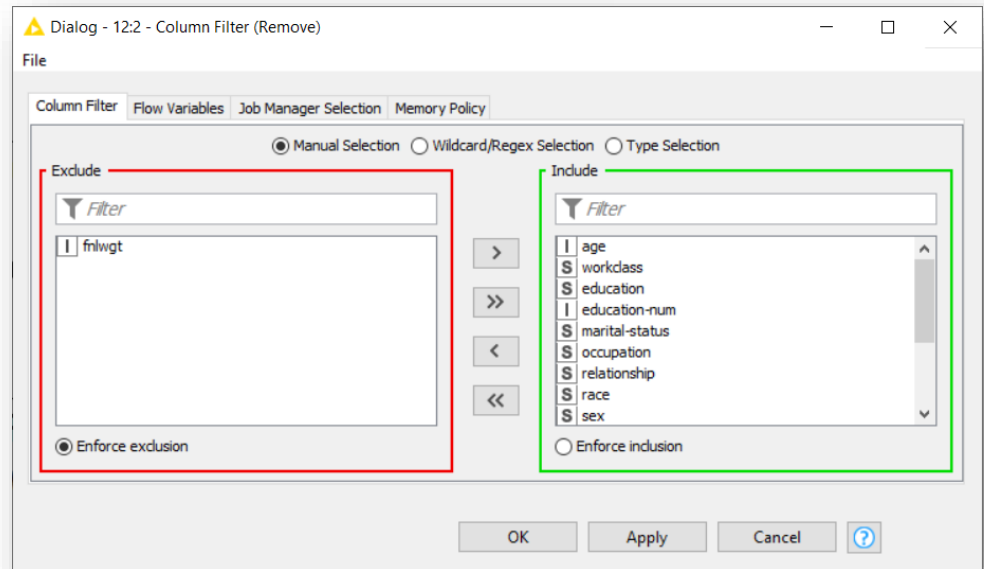
Open the configuration window (double click) and select the file on your machine containing the adult dataset
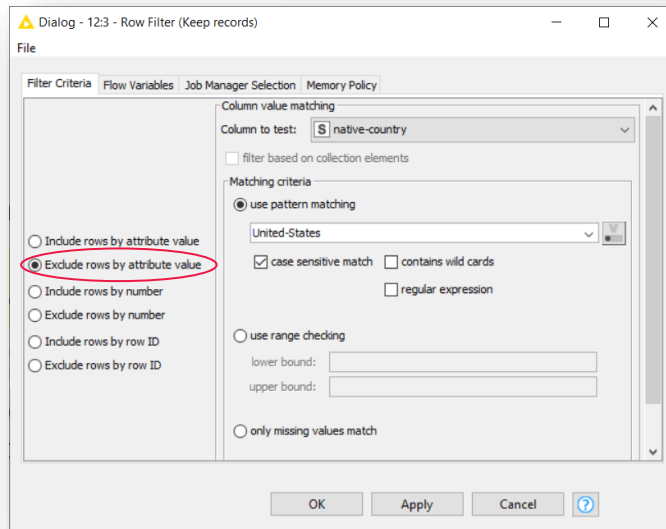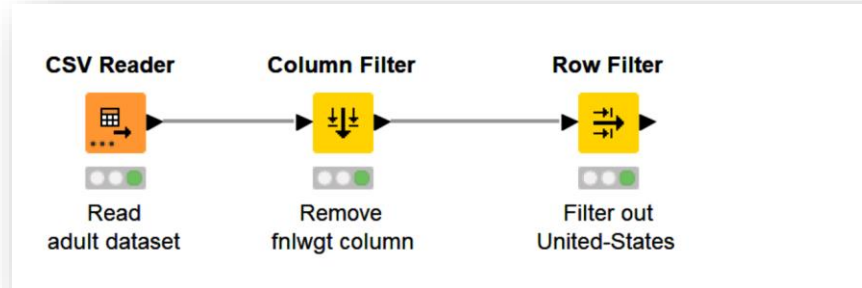
Remove columns

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

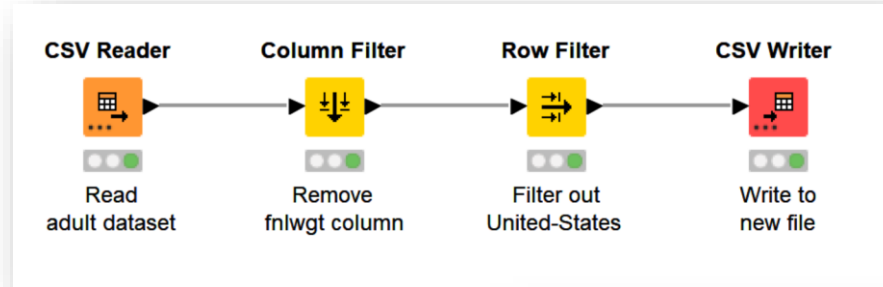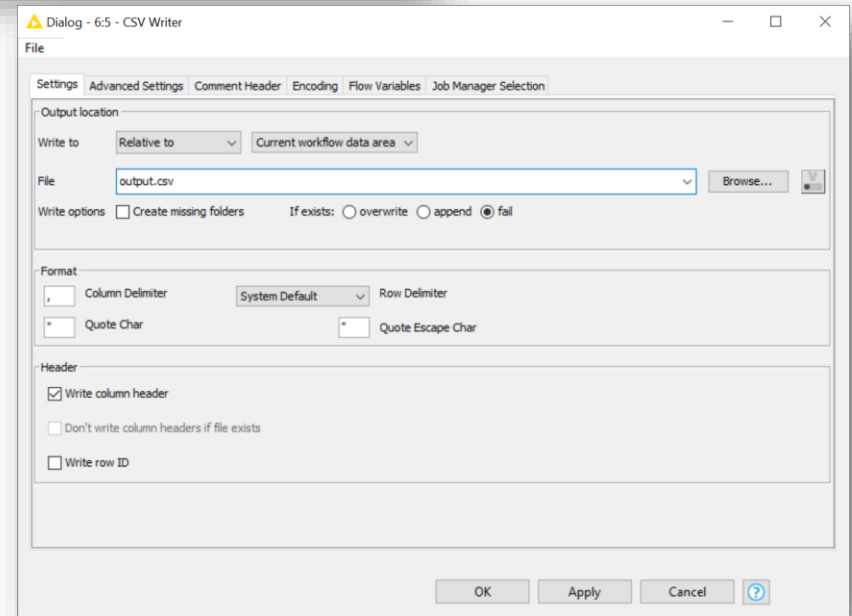Add a Row Filter node and configure it to only keep entries whose "native-country" value is not "United-States"
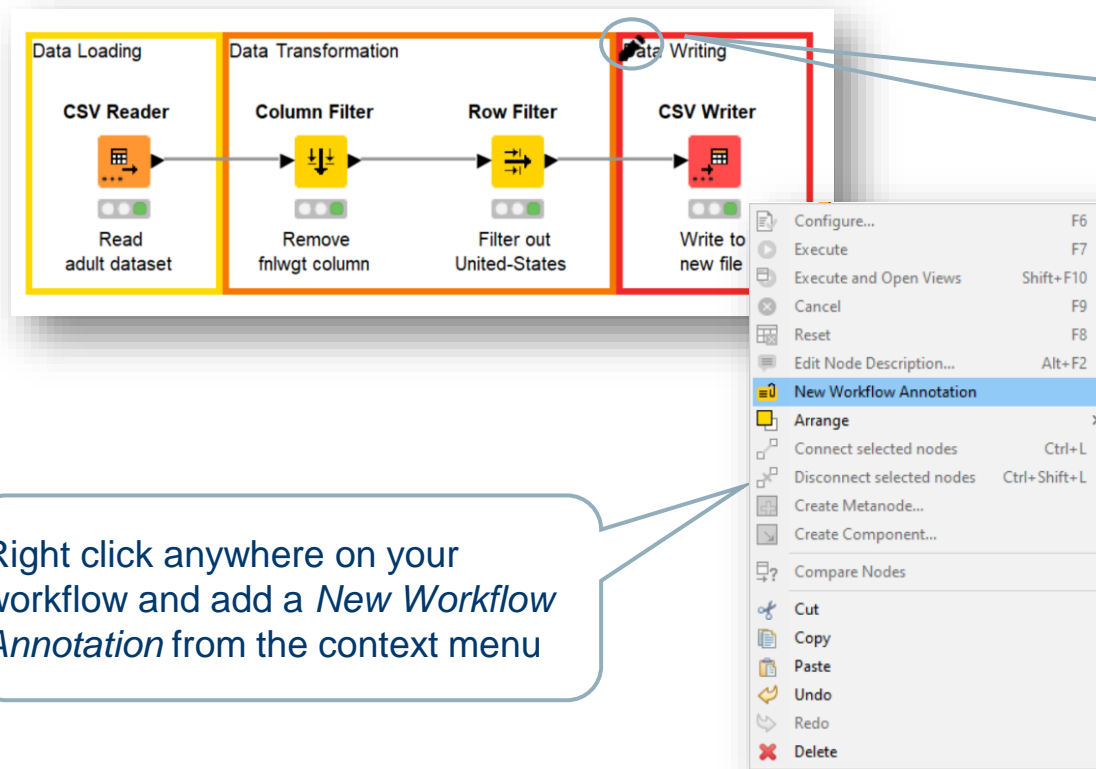
Write to new file

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