

Open for Innovation

KNIME

Visual Workflows for Pushing Data Literacy

Dr. Stefan Helfrich (KNIME GmbH)

January 27, 2021



KNIME in a Nutshell

- **KNIME, the Company**

- Offices in Europe and USA
- Close ties to Universities and Research
- Global Partners Network
- Long-term funding



- **KNIME Software**

- Used by 96% of EUROSTOXX 50 and DOW JONES Indu companies

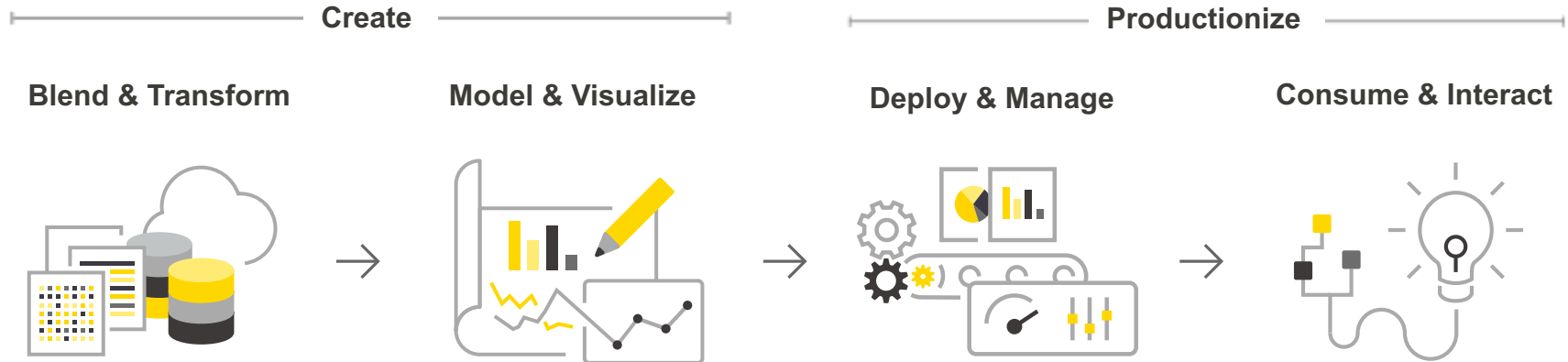
- **KNIME Analytics Platform**

- Gather, Wrangle, Model, Visualize
- Open source and free

- **KNIME Server**

- Deploy, Manage, Consume, Optimize
- Commercial complement

KNIME Software – One Ecosystem



KNIME Analytics Platform

KNIME Extensions

KNIME Integrations

Community Extensions

Partner Extensions

KNIME Server

Team Collaboration

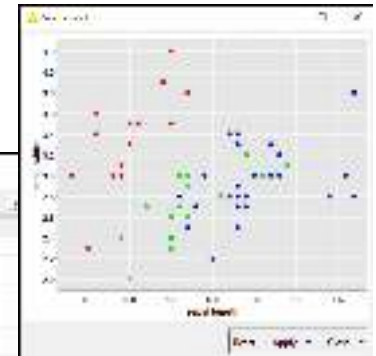
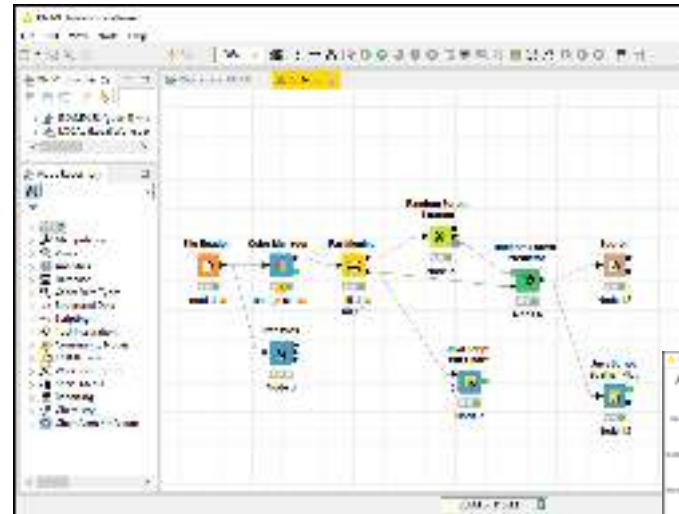
End User Applications

API Services

Managed Execution

What is KNIME Analytics Platform?

- A platform for data analysis/mining, manipulation, visualization, ML and AI
- Build around **visual workflows** (what you see is what you run)
- Open source and free to use for everyone (no limitations)
- Provides a diverse array of extensions and integrations:
 - Text Mining
 - Network Mining
 - Cheminformatics
 - Image Mining
 - Other Tools: Weka, Keras, H2O, ...
 - Other Languages: R, Python,



Visual KNIME Workflows

Nodes

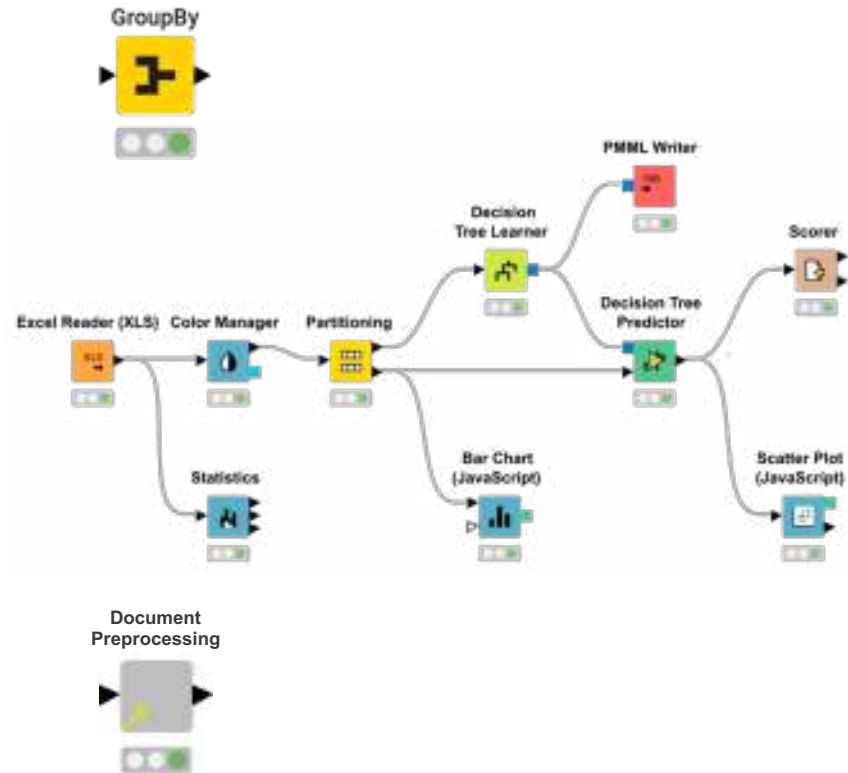
perform tasks on data

Workflows

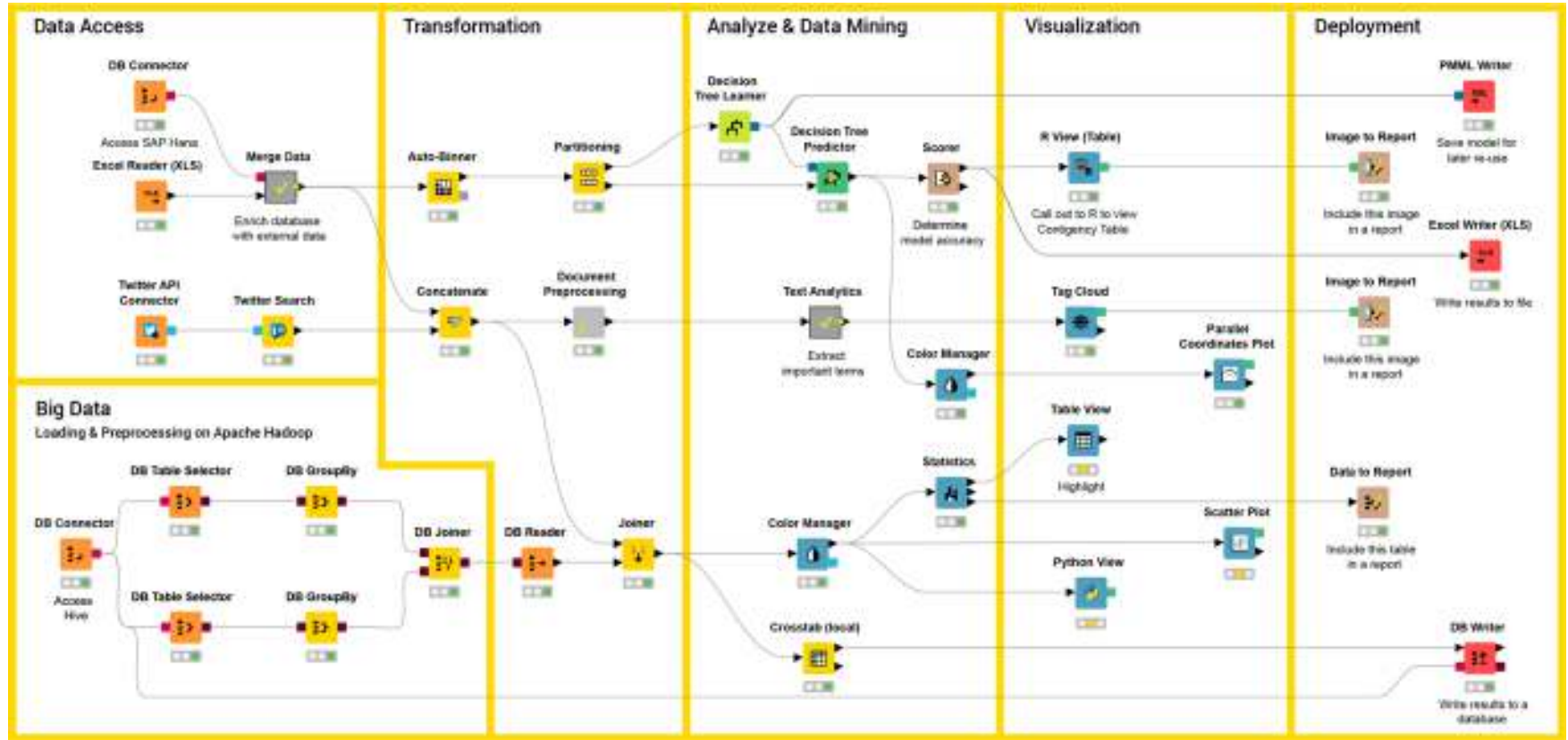
combine nodes to model data flow

Components

encapsulate complexity & expertise



Covering all Stages of the Data Science Life Cycle



KNIME in Research

A KNIME Workflow for Automated Structure Verification

[JA Lumley, G Sharman, T Wilkin...](#) - ... the Science of Drug ..., 2020 - journals.sagepub.com

Adequate characterization of chemical entities made for biological screening in the drug discovery context is critical. Incorrectly characterized structures lead to mistakes in the interpretation of structure–activity relationships and confuse an already multidimensional ...

☆ Cited by 4 Related articles All 3 versions

Data Analytics and Visualization Techniques of Corona Impact

[A Pankh, S Shah, V Bhatt](#) - researchgate.net

... We have taken around 1.5 million points to plot the graphs over the data analytic tools based on python and **Knime**. Keywords - Data Analytic, Python, **Knime**, Coronavirus visualization, descriptive analytic, machine learning prediction, Linear regression 1. Introduction ...

☆ All 3 versions

Five years of the **KNIME** vernalis cheminformatics community contribution

[SD Roughley](#) - Current medicinal chemistry, 2020 - ingentaconnect.com

Since the official release as a **KNIME** Community Contribution in June 2013, the Vernalis **KNIME** nodes have increased from a single node (the 'PDB Connector' node) to around 126 nodes (November 2017; Version 1.12. 0); furthermore, a number of nodes have been ...

☆ Cited by 6 Related articles All 3 versions

Hitachi Materials Informatics Analytics Platform Assisting Rapid Development.

[Y Osakabe, A Asahara, H Morita](#) - AAAI Spring Symposium: Combining ..., 2020 - ceur-ws.org

... Second, upload them to the MIAP database. MIAP automatically converts files with different formats into a predetermined format using **KNIME**, the open source Figure 2: Screen capture of checking learning results software (**KNIME** 2019) ...

☆ Related articles

Integration of the ImageJ Ecosystem in **KNIME** Analytics Platform

[C Dietz, CT Rueden, S Helfrich, ETA Dobson...](#) - Frontiers in Computer ..., 2020 - frontiersin.org

Open-source software tools are often used for analysis of scientific image data due to their flexibility and transparency in dealing with rapidly evolving imaging technologies. The complex nature of image analysis problems frequently requires many tools to be used in ...

☆ Cited by 3 Related articles All 2 versions

Speech Signal Analysis and Classification of Dominant Parameter for Pathological Voices

[W Christina Subiksha, A Nandhini, KP Bharath...](#) - researchgate.net

... The secondary objective is the classification of the voice signal into normal and abnormal voice samples using the machine learning software Konstanz Information Miner (**KNIME**) ... For classification of voice signal to be normal or abnormal, the **Knime** software is used ...

☆

Data Analysis and Application of Retail Enterprises Based on **Knime**

[X Pu, N Qi, J Huang](#) - MS&E, 2020 - iopscience.iop.org

With the rapid development of information technology, the application of cross-discipline has shown explosive growth. Data mining, big data and other technologies are rapidly entering all walks of life. Research on data mining technology and its application in data analysis of ...

☆ Related articles All 2 versions

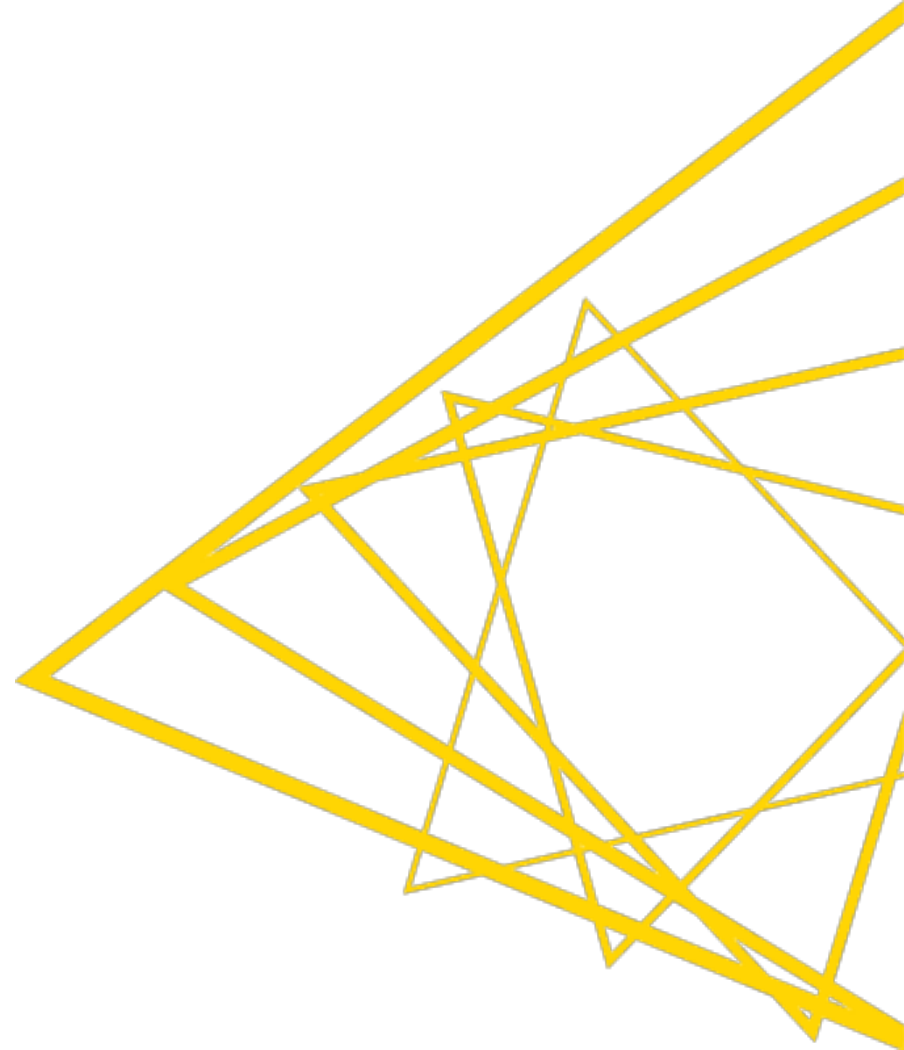
AI supported Topic Modeling using **KNIME**-Workflows.

[J Al Qundus, S Peikert, A Paschke](#) - Qurator, 2020 - pdfs.semanticscholar.org

Topic modeling algorithms traditionally model topics as list of weighted terms. These topic models can be used effectively to classify texts or to support text mining tasks such as text summarization or fact extraction. The general procedure relies on statistical analysis of term ...

☆ Cited by 1 Related articles All 4 versions

Let's Take a Look at KNIME (Demo)



The KNIME Analytics Platform Workbench

The screenshot displays the KNIME Analytics Platform Workbench interface. The main workspace shows a workflow titled "My first Workflow" with four nodes: File Reader (read adult.csv), Row Filter (keep only records born in the US), Column Filter (remove gender), and Table Writer (Write table). The Row Filter node is highlighted with a yellow box and a callout. The interface is divided into several panels:

- KNIME Explorer:** Located in the top-left, showing a file tree with folders like "My-KNIME-Hub", "EXAMPLES", and "LOCAL (Local Workspace)".
- Workflow Coach:** Located in the middle-left, displaying "Recommended Nodes" with a list including GroupBy, Joiner, Column Filter, Concatenate, CSV Writer, Reference Row Filter, String Manipulation, and Math Formula, along with their respective community percentages.
- Node Repository:** Located in the bottom-left, showing a hierarchical list of node categories such as IO, Manipulation, Views, Analytics, DB, and Other Data Types.
- Workflow Editor:** The central workspace where the workflow is built and executed.
- Node Description:** Located in the top-right, providing details for the selected "Row Filter" node, including its purpose and configuration options.
- Outline:** Located in the bottom-middle, showing a small overview of the workflow structure.
- Console & Node Monitor:** Located in the bottom-right, displaying the execution status of the selected node and a table of output data.

ID	age	workclass	lnthgt	education	education-num	occupation	relationshp	
Row1	50	Self-emp-not-inc	83311			Exec-managerial	Husband	
Row2	38	Private	21564			Handlers-cleaners	Not-in-fa	
Row3	53	Private	23472			Handlers-cleaners	Husband	
w5	37	Private	28452	Masters	14	Married-civ-spouse	Exec-managerial	Wife
w7	52	Self-emp-not-inc	209642	HS-grad	9	Married-civ-spouse	Exec-managerial	Husband

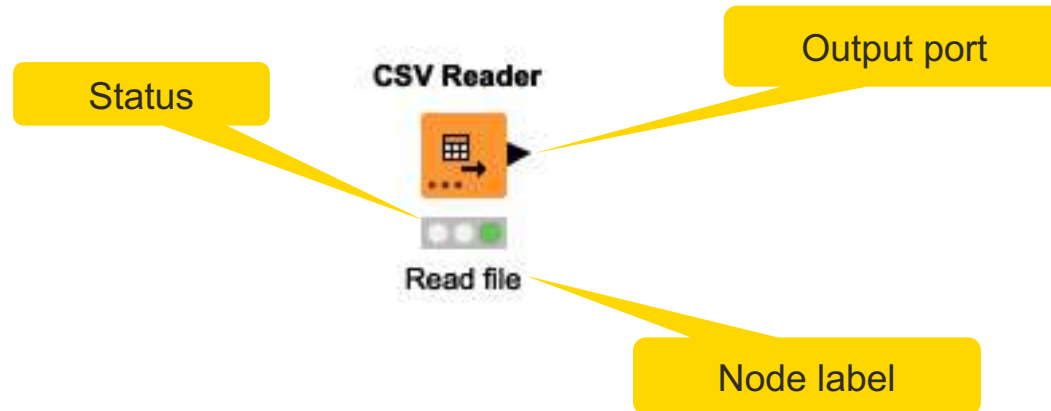
Importing Data and File Handling



Data Source Nodes

Typically characterized by:

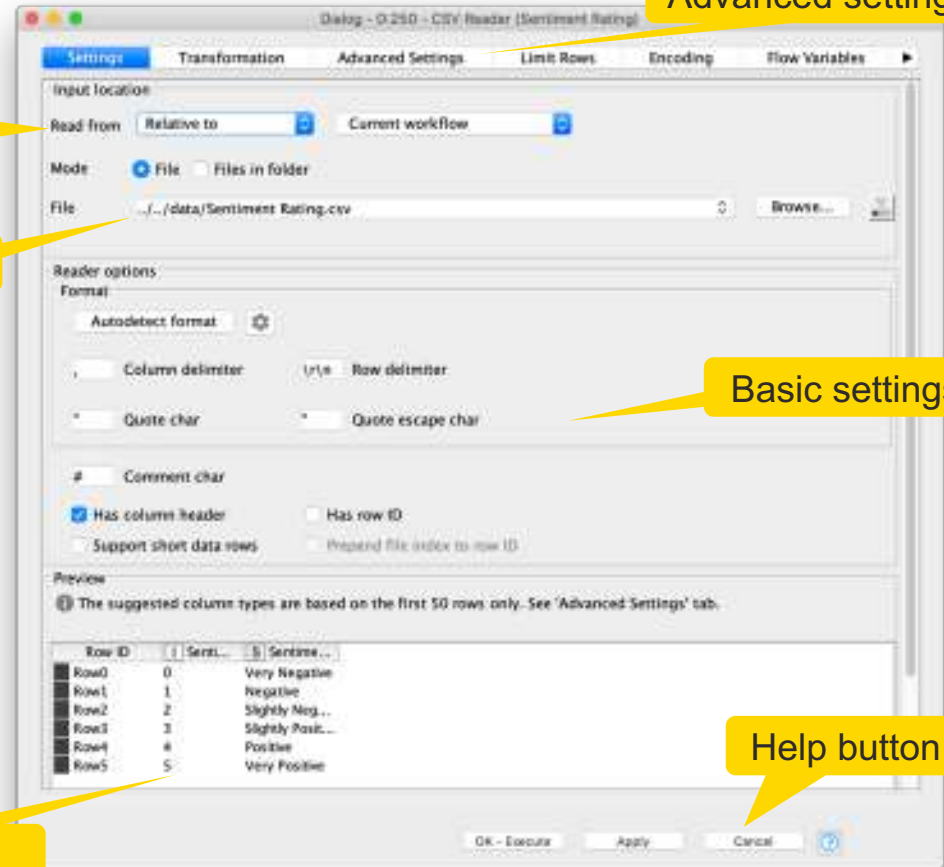
- Orange color
- By default no input ports, 1-2 output ports
- Consistent UX across all nodes and files systems (ongoing work)
- Able to manage various file systems within the same workflow



CSV Reader

- Reads either one or multiple .csv and .txt files
- Further tabs to
 - limit the rows
 - select encoding

CSV Reader



Default File Systems

- Local File System



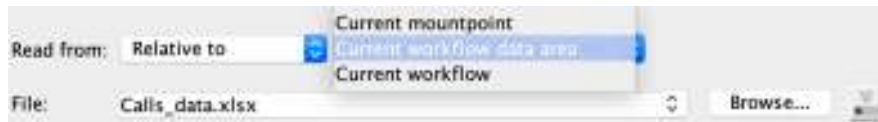
Input location

Read from: Local File System

Mode: File Files in folder

File: /Users/kathrinmelcher/Desktop/course_data.csv

- Relative to ...



Read from: Relative to

File: Calls_data.xlsx

Current mountpoint
Current workflow data area
Current workflow

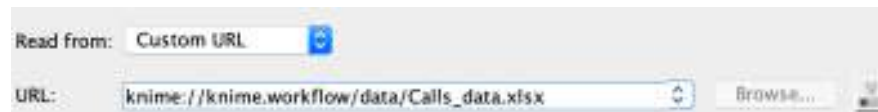
- Mountpoint



Read from: Mountpoint LOCAL

File: /Example Workflows/TheData/Customers/CallsData.xls

- Custom URL



Read from: Custom URL

URL: knime://knime.workflow/data/Calls_data.xlsx

Connecting to other File Systems

- Add file system connection port to connect to another file system

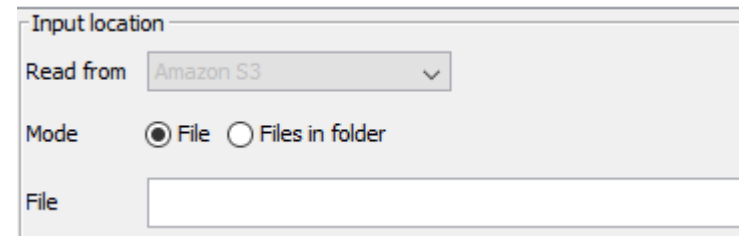
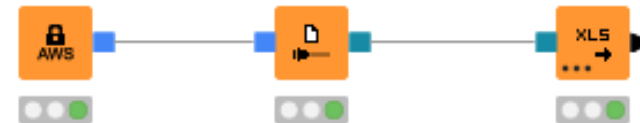
- Click on the three dots on the lower left to add or remove a dynamic port.



- Supported file systems

- Microsoft Azure
- Google
- Amazon
- Databricks
- BigData file systems (hdfs, httpFS, ...)
- On-premise (e.g. ssh, ftp, ...)

Amazon Authentication Amazon S3 Connector Excel Reader (XLS)

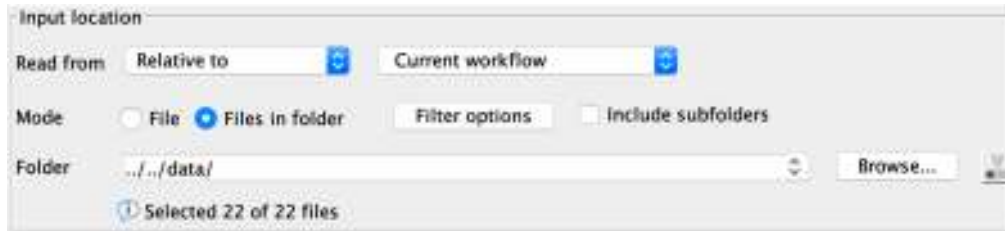
A screenshot of the configuration panel for the 'Excel Reader' widget. It has a title 'Input location'. Under 'Read from', there is a dropdown menu set to 'Amazon S3'. Under 'Mode', there are two radio buttons: 'File' (selected) and 'Files in folder'. At the bottom, there is a text input field labeled 'File'.

Read Single or Multiple Files

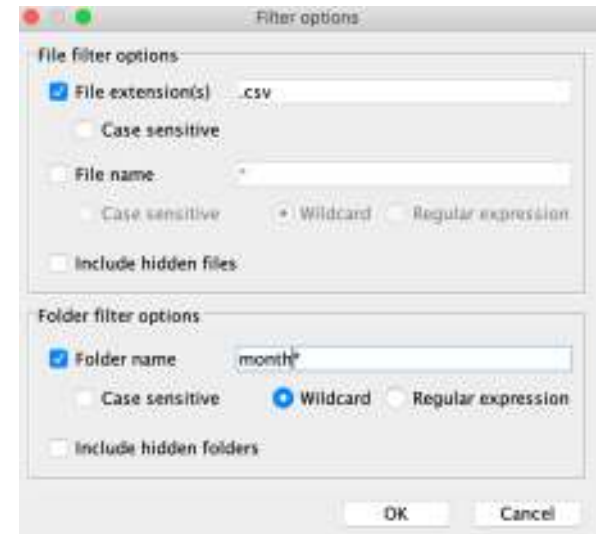
- Single file



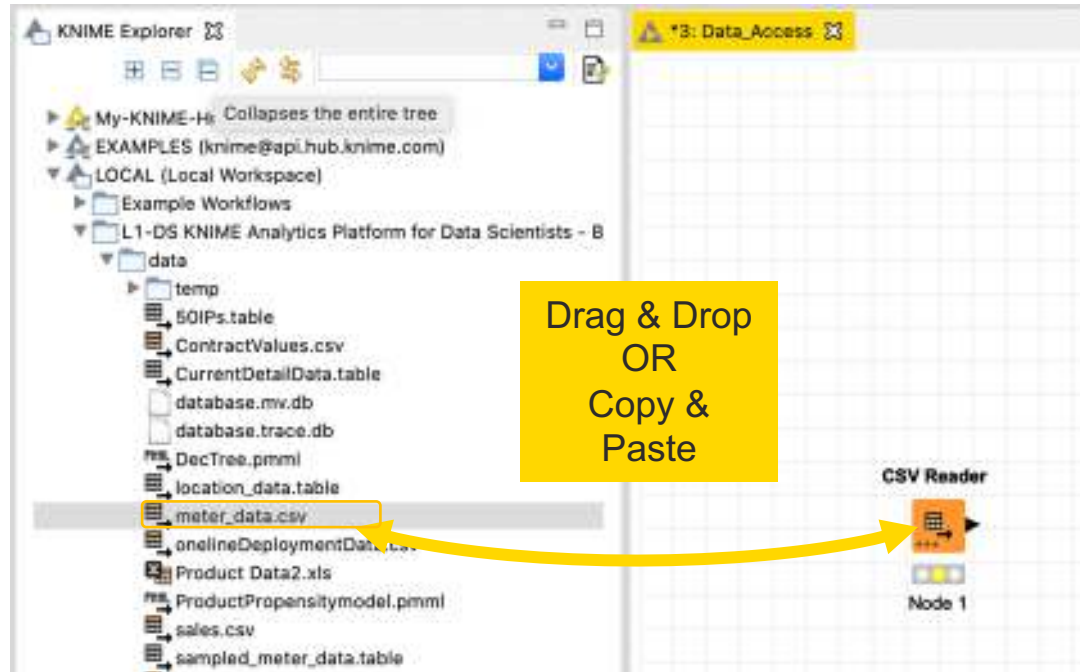
- Files in a folder



- Option to include subfolder
- Option to define filter criterions

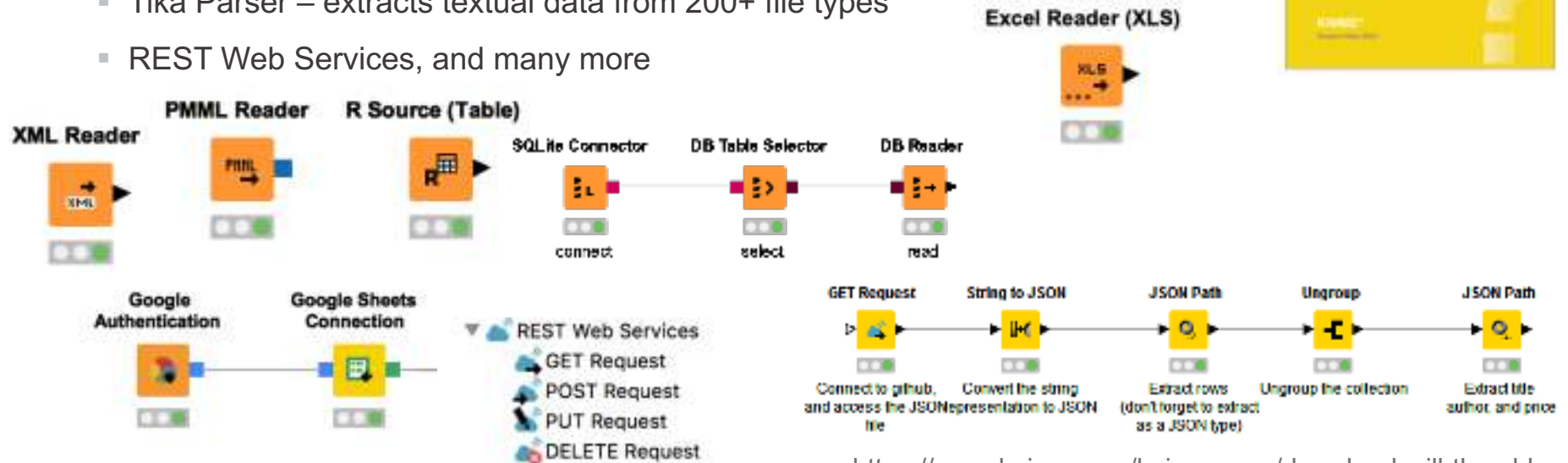


Alternative Faster Way ...



Other Useful Data Sources

- KNIME Analytics Platform provides many more options to access data:
 - PMML Reader – reads standard predictive models
 - XML Reader with XPATH support
 - Python/R Source nodes
 - Tika Parser – extracts textual data from 200+ file types
 - REST Web Services, and many more



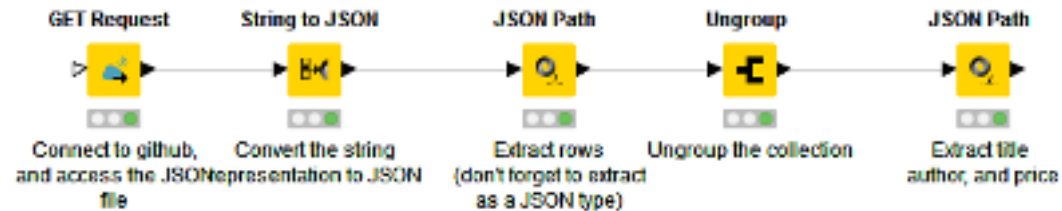
<https://www.knime.com/knimepress/download-will-they-blend>

RESTful Web Services

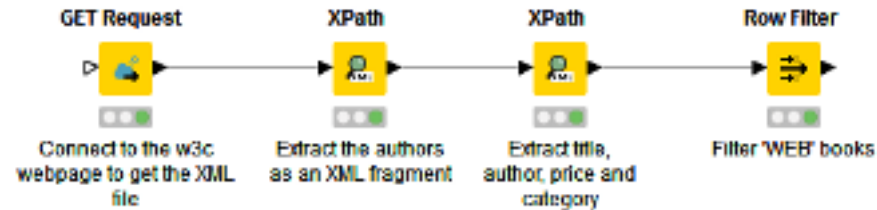
- Use KNIME nodes to interact with RESTful web services
- Send requests using standard HTTP methods



JSON Response:



XML Response:



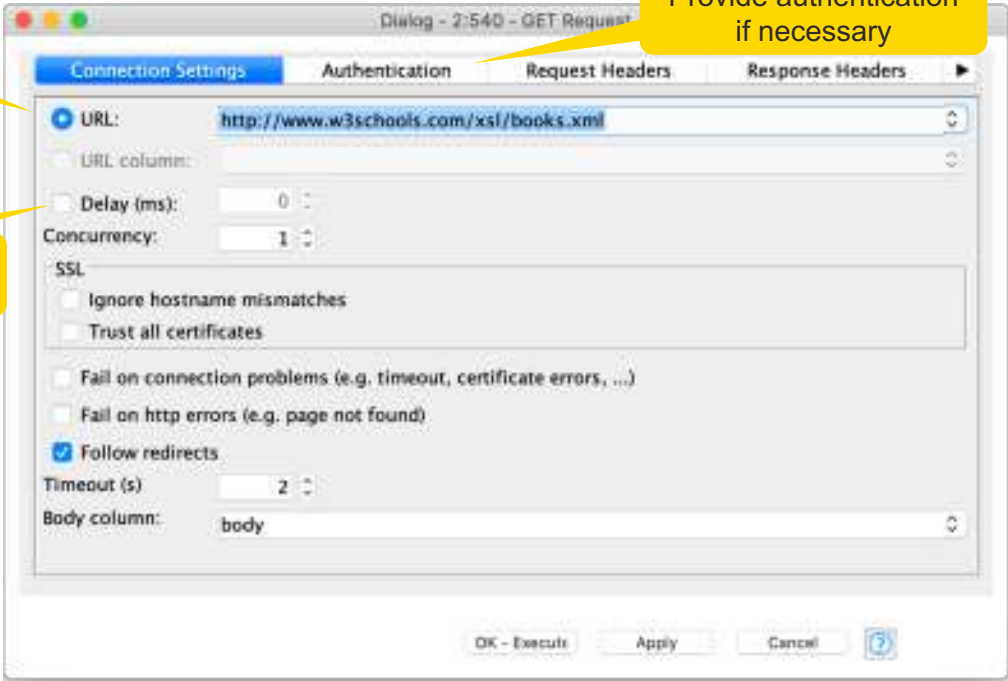
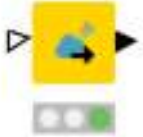
RESTful Web Services

Enter URL, or use from column

Provide authentication if necessary

Add delay between individual requests

GET Request

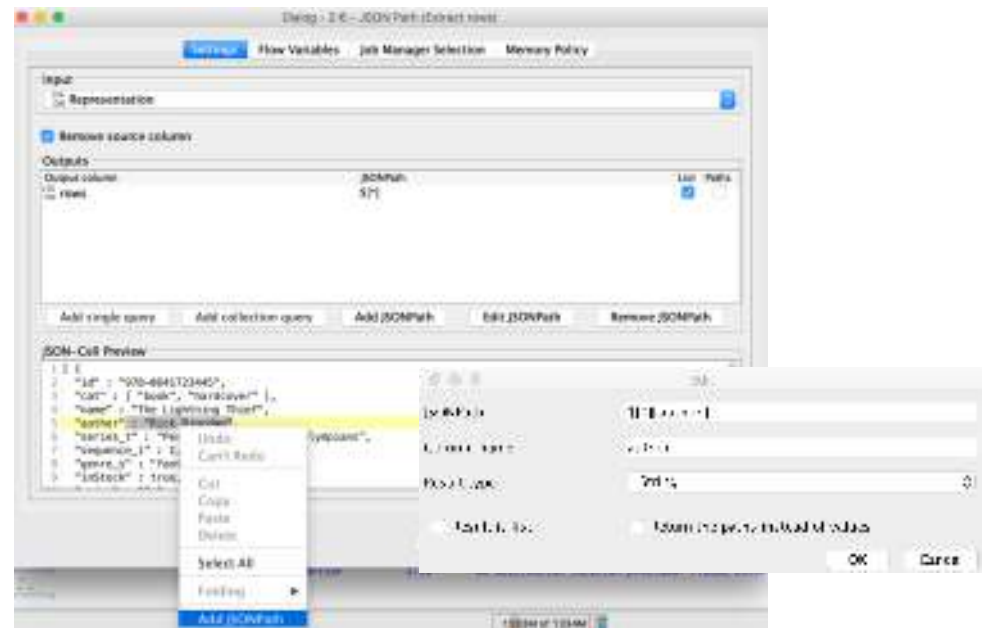


<https://www.knime.com/blog/a-restful-way-to-find-and-retrieve-data>

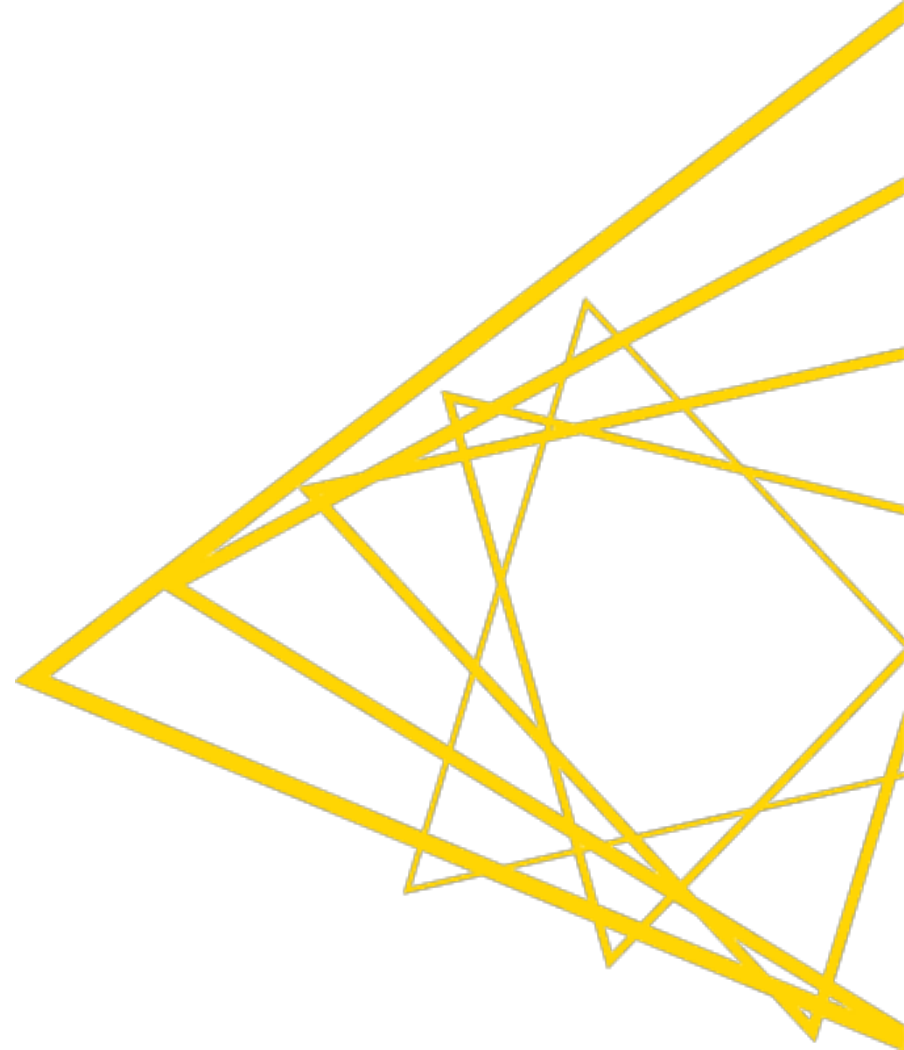
<https://www.knime.com/blog/OSM-meets-CSV-file-and-Google-API>

JSON Reader and JSON Path nodes

- Use the JSON Reader (or GET Request) node to get a JSON cell
- Use the JSON Path node to query the JSON file and extract parameters
- Editor window simplifies construction of JSON queries by auto-generating them (click on properties)

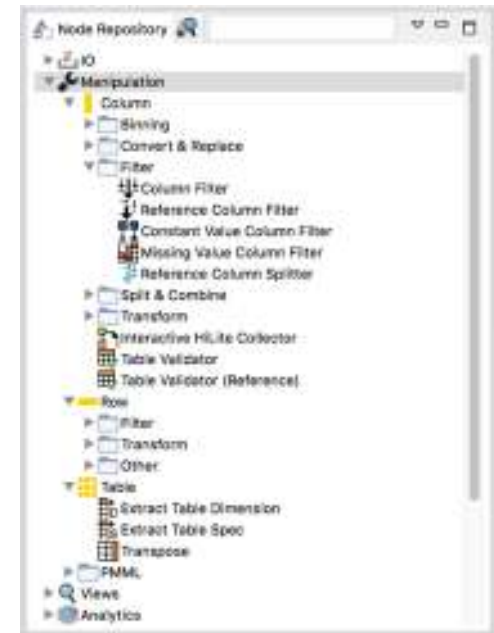
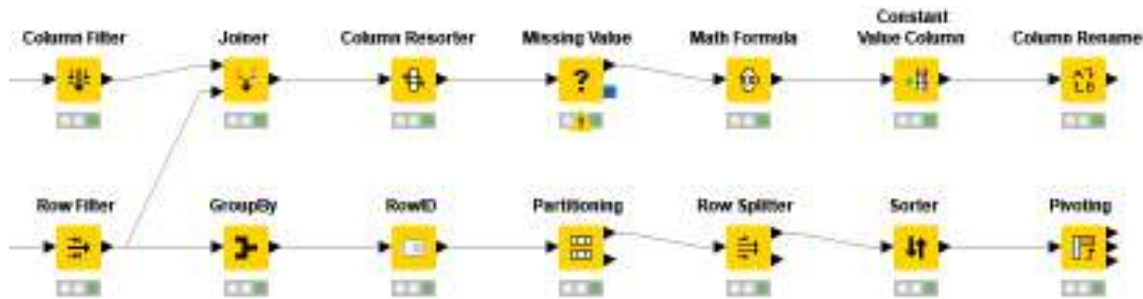


Data Transformation



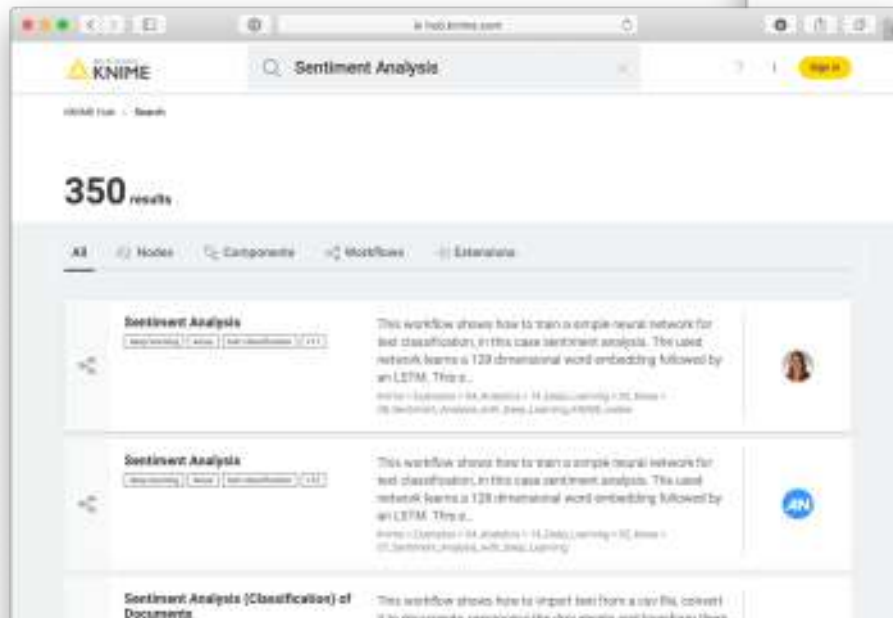
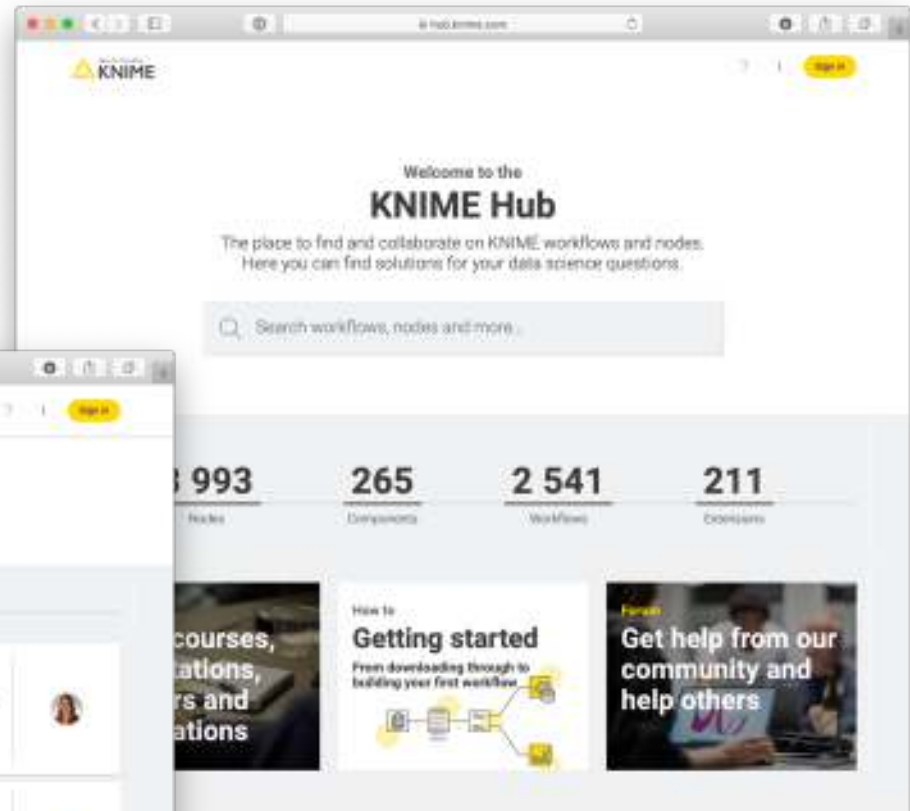
Data Transformation Nodes

- Yellow color with a variety of input and output ports
- Apply a transformation to input data
- Many, many nodes!



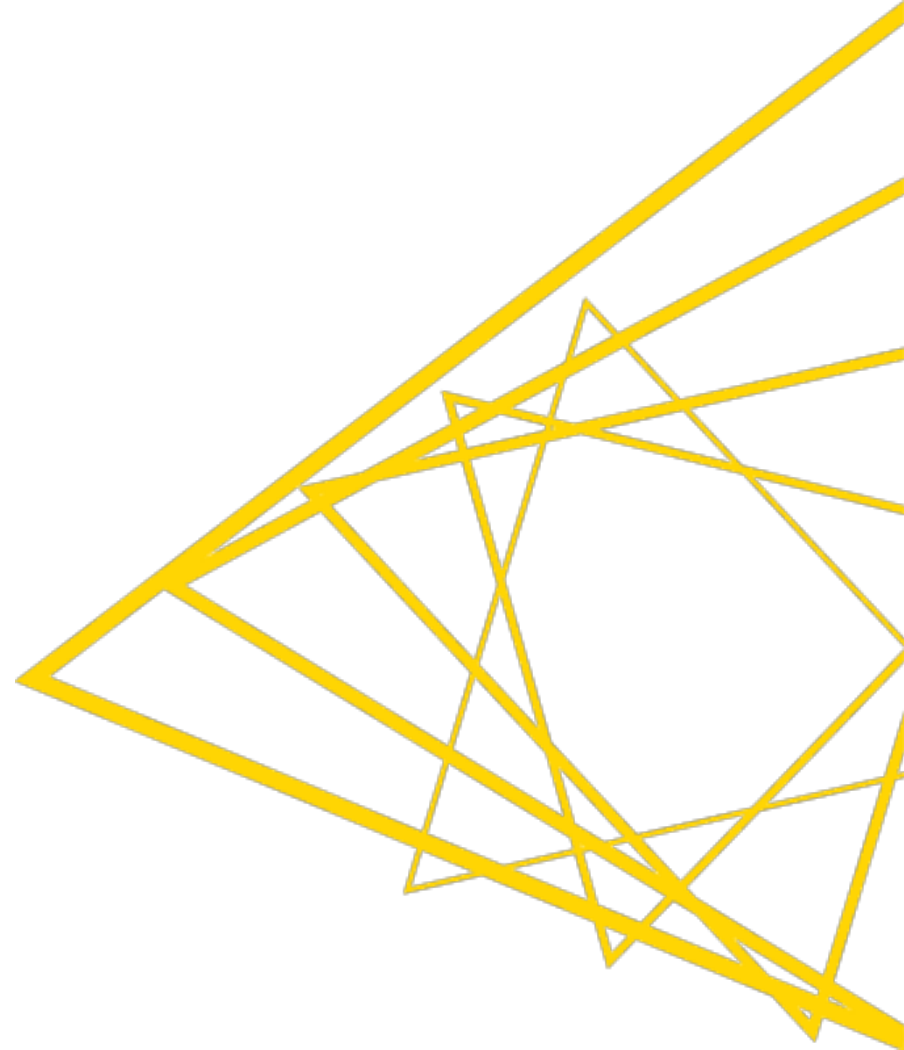
Finding Your Way: KNIME Hub

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



<https://hub.knime.com>

Data Visualization



Data Visualization

- Large selection of easy to use visualization nodes
 - Web-based and interactive
 - Dedicated nodes,
 - no scripting required
- Plotly nodes
 - Similar but integrated from an external library
- R and Python View nodes for highly customizable graphics
 - Require scripting

Views

- JS JavaScript
 - JS Generic JavaScript View
 - Bar Chart
 - Box Plot
 - Conditional Box Plot
 - Decision Tree View
 - Heatmap
 - Histogram
 - Lift Chart
 - Line Plot
 - Parallel Coordinates Plot
 - Pie/Donut Chart
 - ROC Curve
 - Scatter Plot
 - Stacked Area Chart
 - Sunburst Chart
 - Table Editor
 - Table View
 - Tag Cloud
 - Tile View

JS JavaScript Views (Labs)

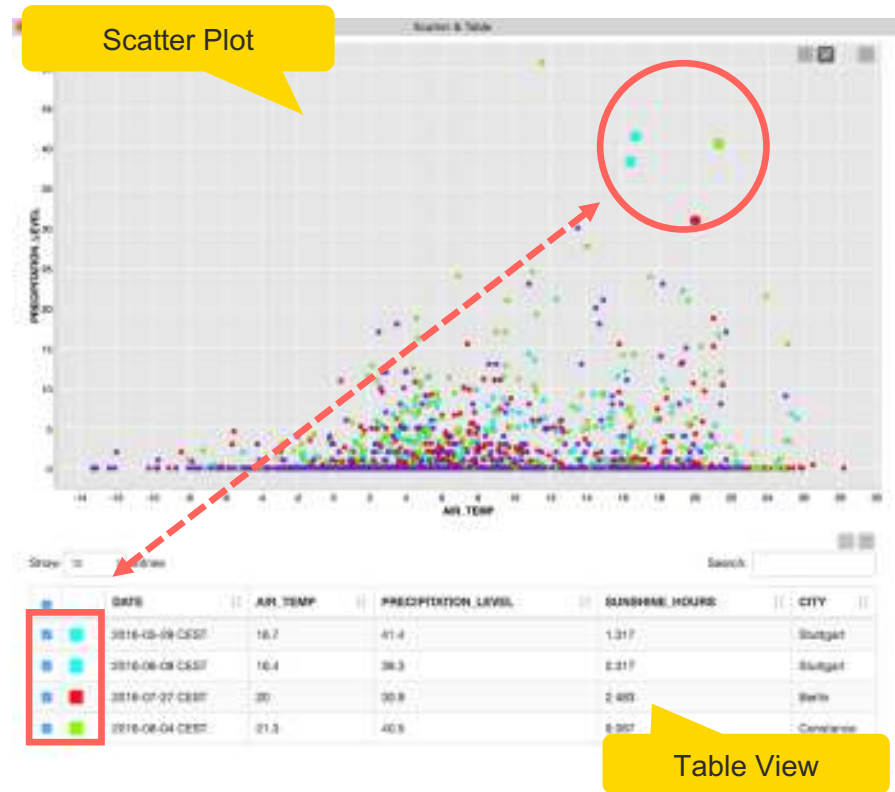
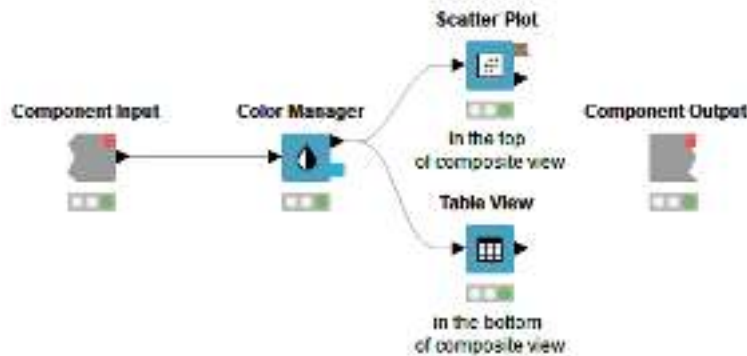
- JS Plotly
 - JS Bubble Chart (Plotly)
 - JS Continuous Error Plot (Plotly)
 - JS Contour Plot (Plotly)
 - JS 2D Density Plot (Plotly)
 - JS Error Bars Plot (Plotly)
 - JS Line Plot (Plotly)
 - JS Radar Plot (Plotly)
 - JS Scatter Plot (Plotly)
 - JS 3D Scatter Plot (Plotly)
 - JS Stacked Area Chart (Plotly)
 - JS Surface Plot (Plotly)
 - JS Violin Plot (Plotly)

Scripting

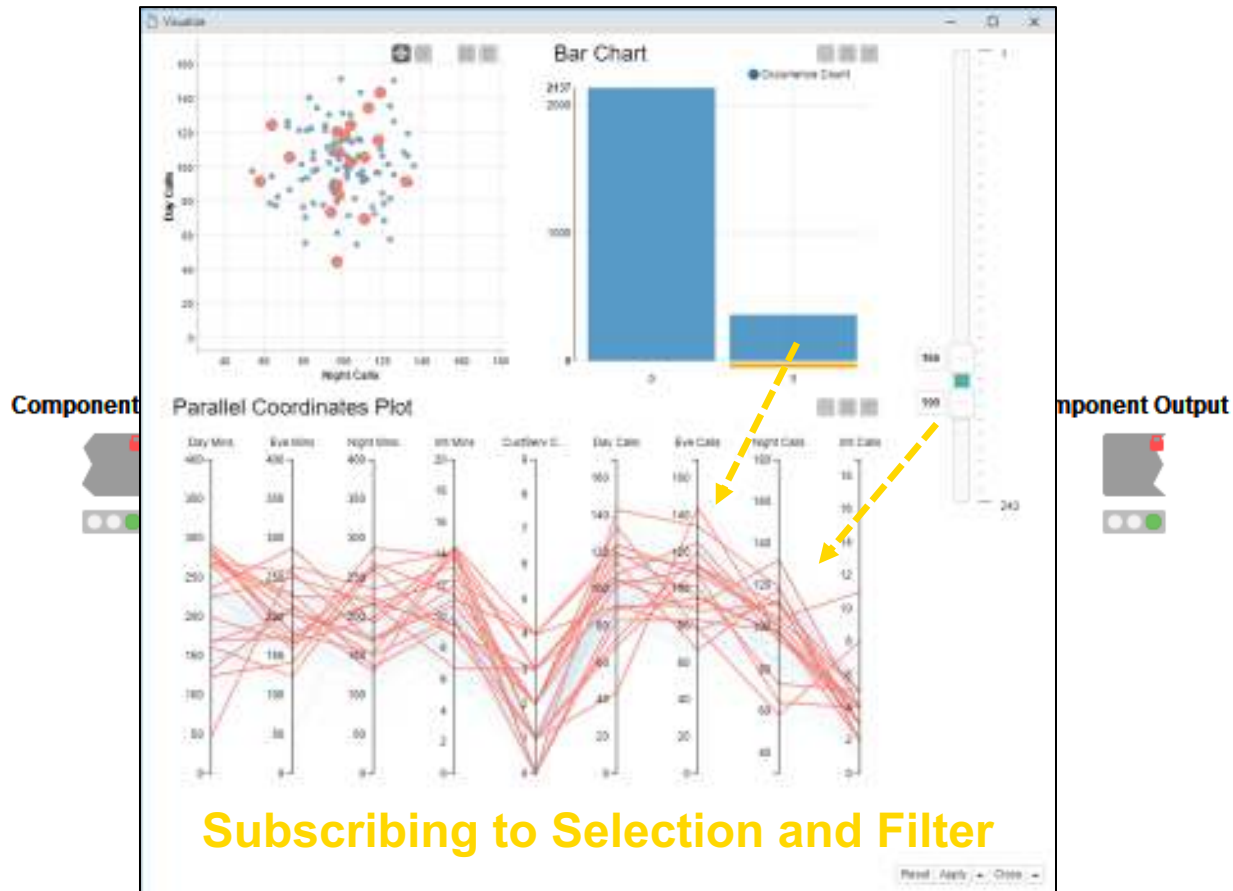
- Python
 - Python View
- R
 - R View (Table)
 - R View (Workspace)

Components – Combined Views

- Multiple JavaScript View nodes can be combined in Components
- Selections are transmitted to all other views
- Also for use on the KNIME



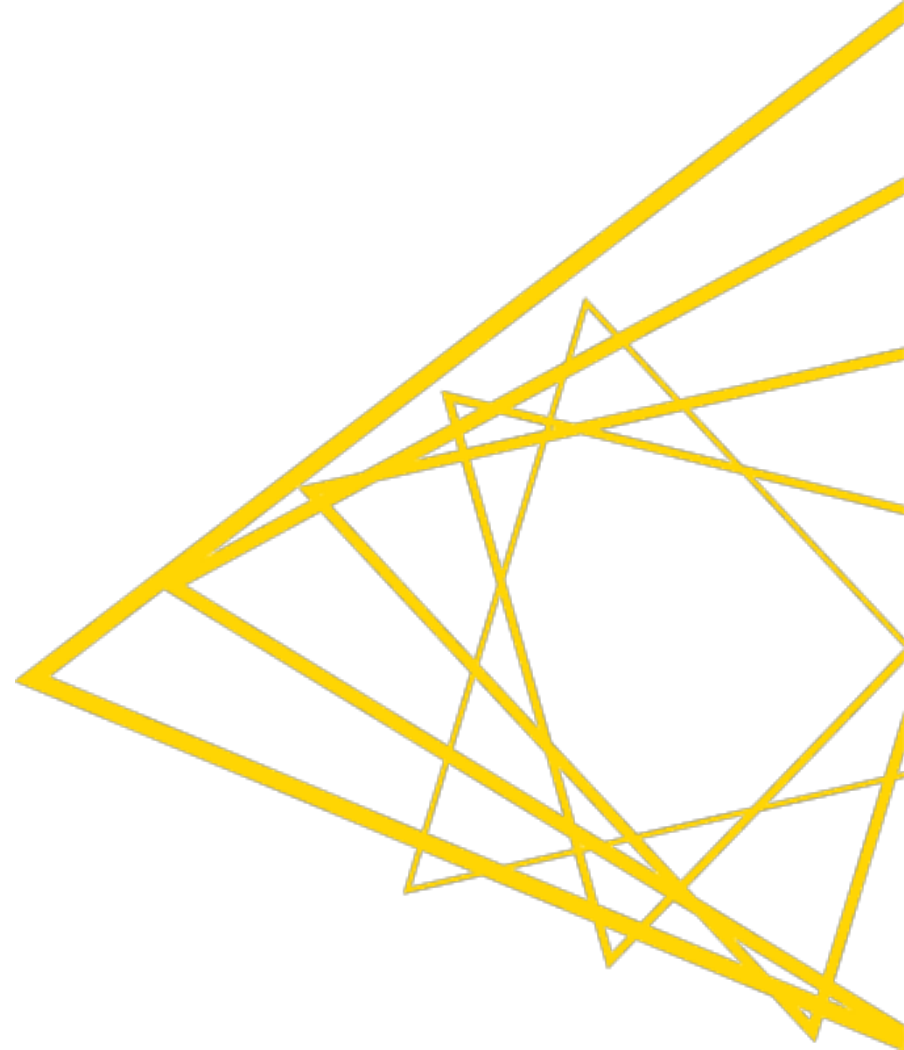
Interactivity across Charts: Selection and Filter Events



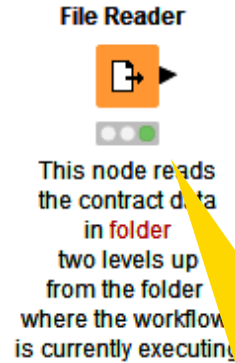
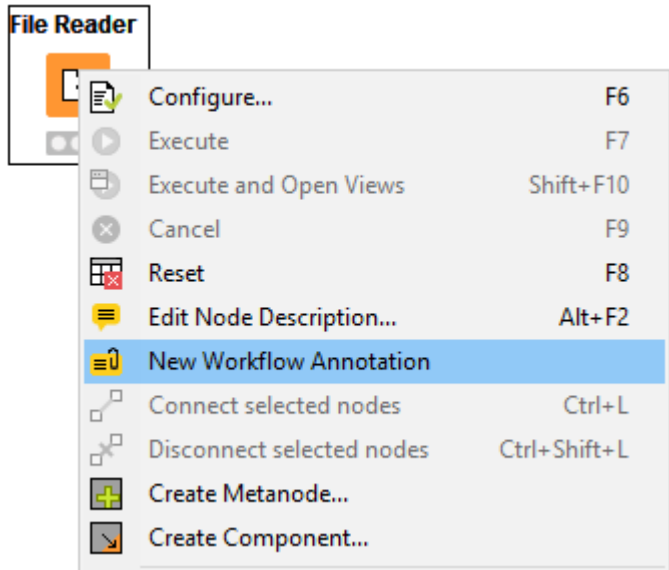
Interactivity across Charts: Selection and Filter Events



Workflow Organization and Documentation



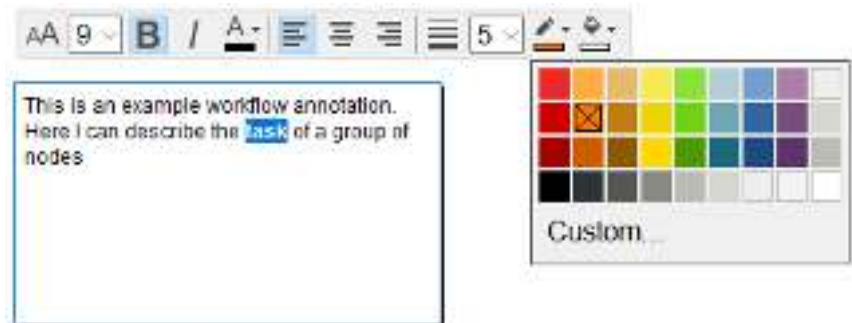
Comments & Annotations



This is an example workflow annotation. Here I can describe the **task** of a group of nodes

Double-click to write
Use the panel to
change properties

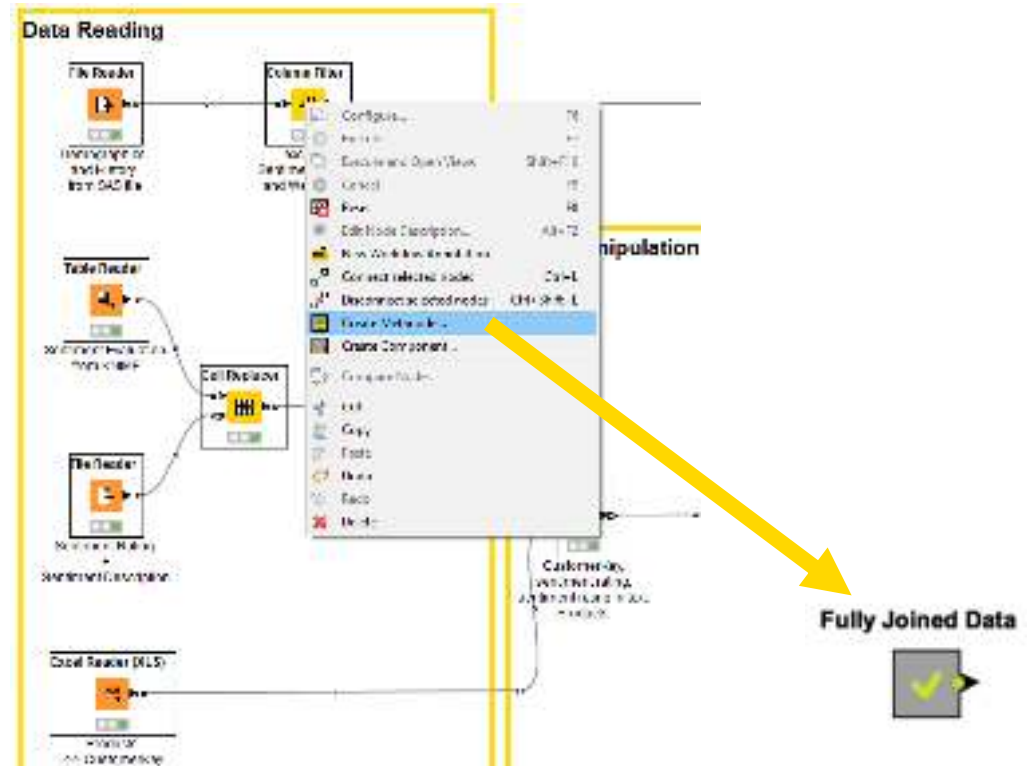
Double-click to write
Use the panel to
change properties



YouTube KNIME TV Channel:
https://youtu.be/AHURYB_O8sA

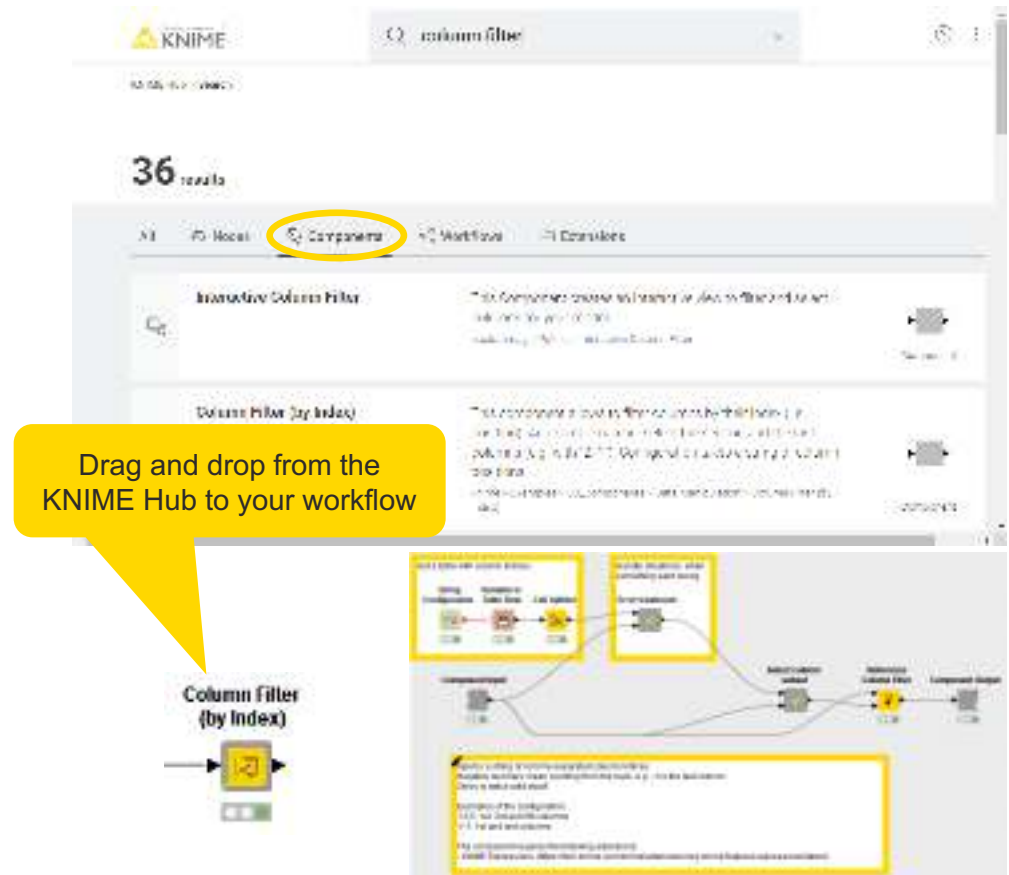
Workflow Organization – Good Practices

- Workflow annotations
- Node labels
- Metanodes
 - Right click -> Create Metanode...
 - Organize workflow by task
 - Hide complexity & improve readability



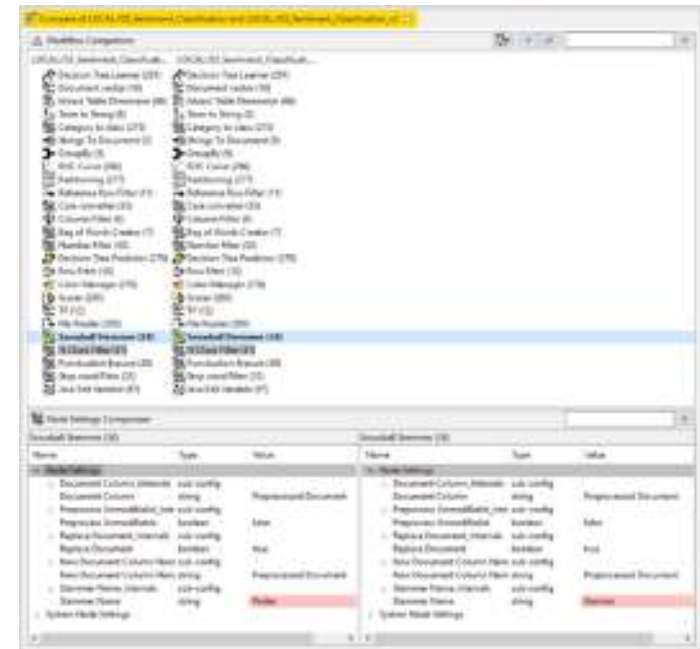
Workflow Organization – Components

- Component encapsulates a reusable functionality as a KNIME workflow
- Components can be configured as any KNIME nodes
- Access and share components on the KNIME Hub

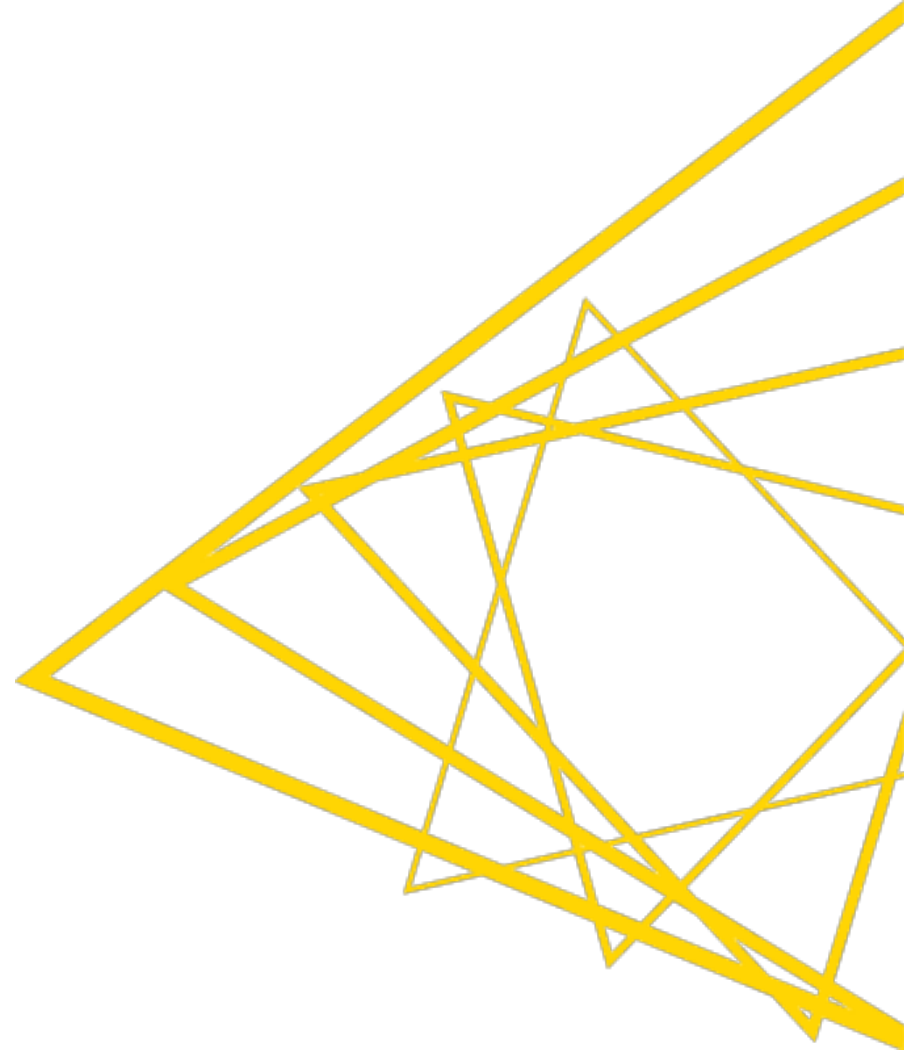


KNIME WorkflowDiff

- Automates identification and comparison of nodes in a workflow, metanodes, and two different workflows
- Identifies insertions, deletions, substitutions, and parameter changes



Data Mining

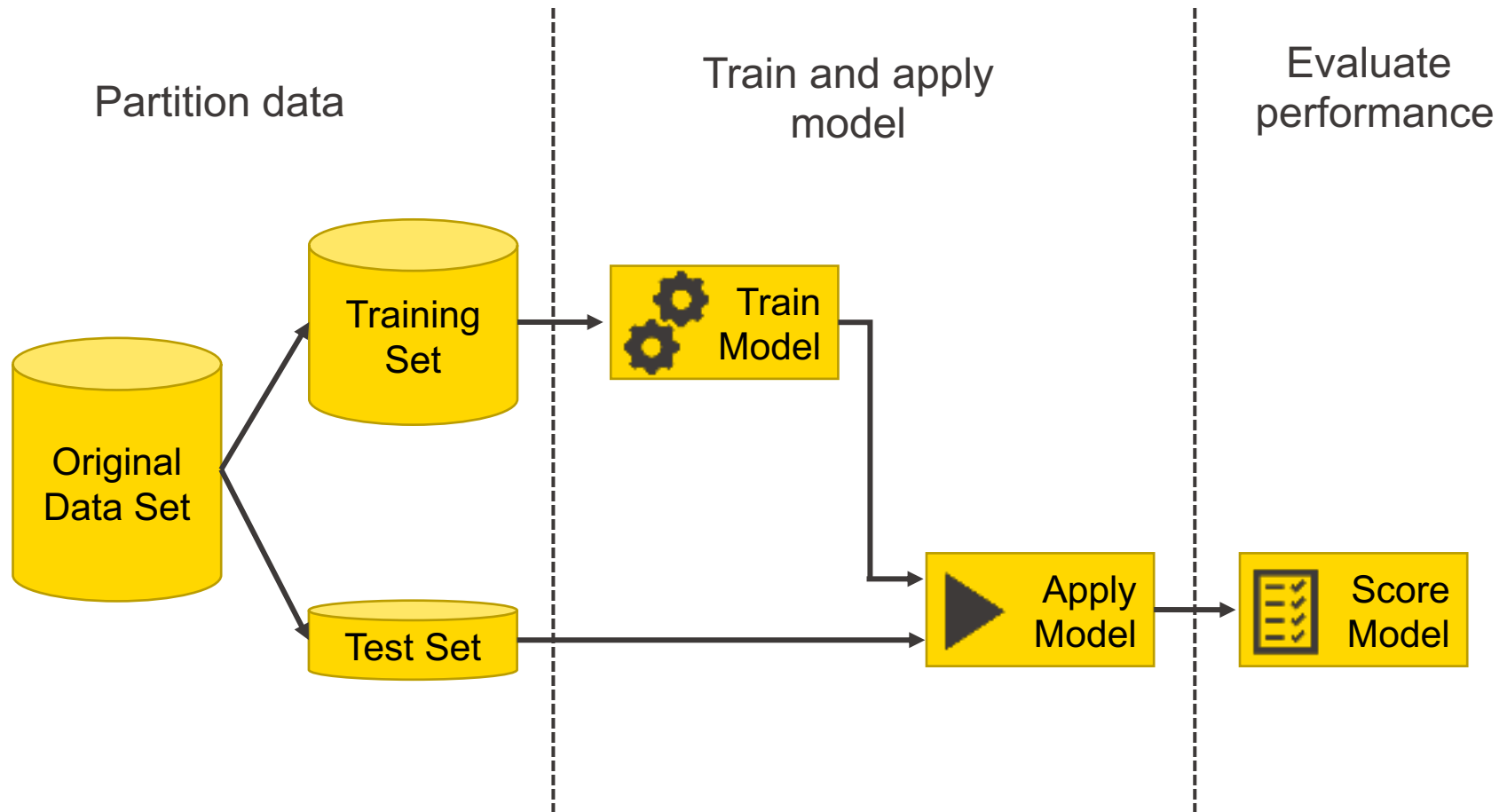


Data Mining Strategies

Example Applications:

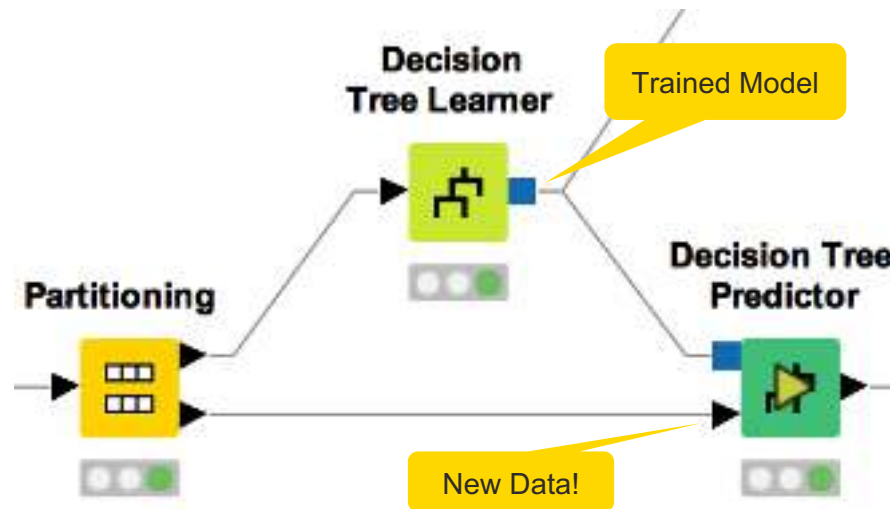
- Anomaly Detection (fraud, predictive maintenance)
- Association Rule Learning (market basket analysis)
- Clustering (customer / market segmentation)
- Classification (next best offer, churn preventions)
- Regression (trend estimation)

Data Mining: Process Overview

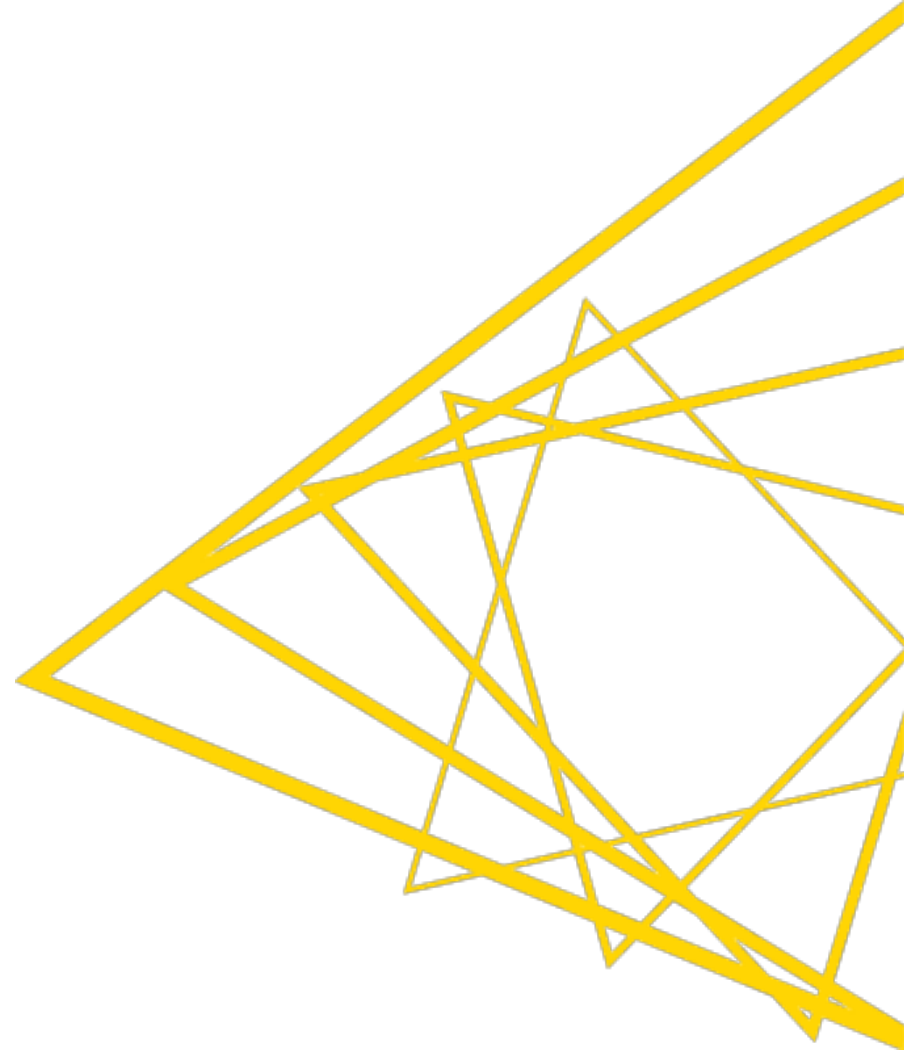


Learner-Predictor Motif

- Most data mining approaches in KNIME use a Learner-predictor motif.
- The Learner node trains the model with its input data.
- The Predictor node applies the model to a different subset of data.



Deployment

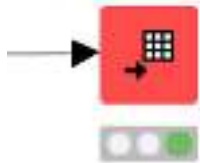


Data Export Nodes

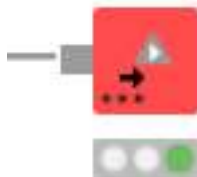
Typically characterized by:

- Magenta color
- 1 input port, no output ports
- Create file on file system or write to database

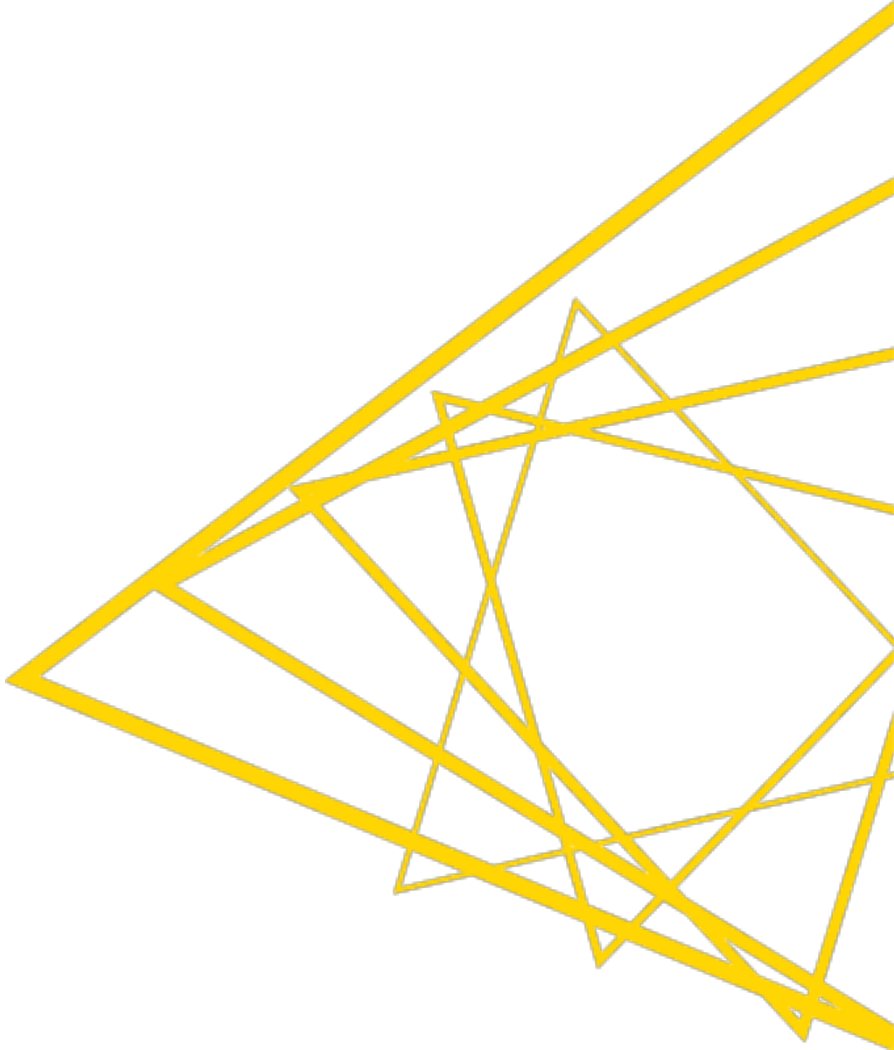
Table Writer



Model Writer



Some Use-Cases



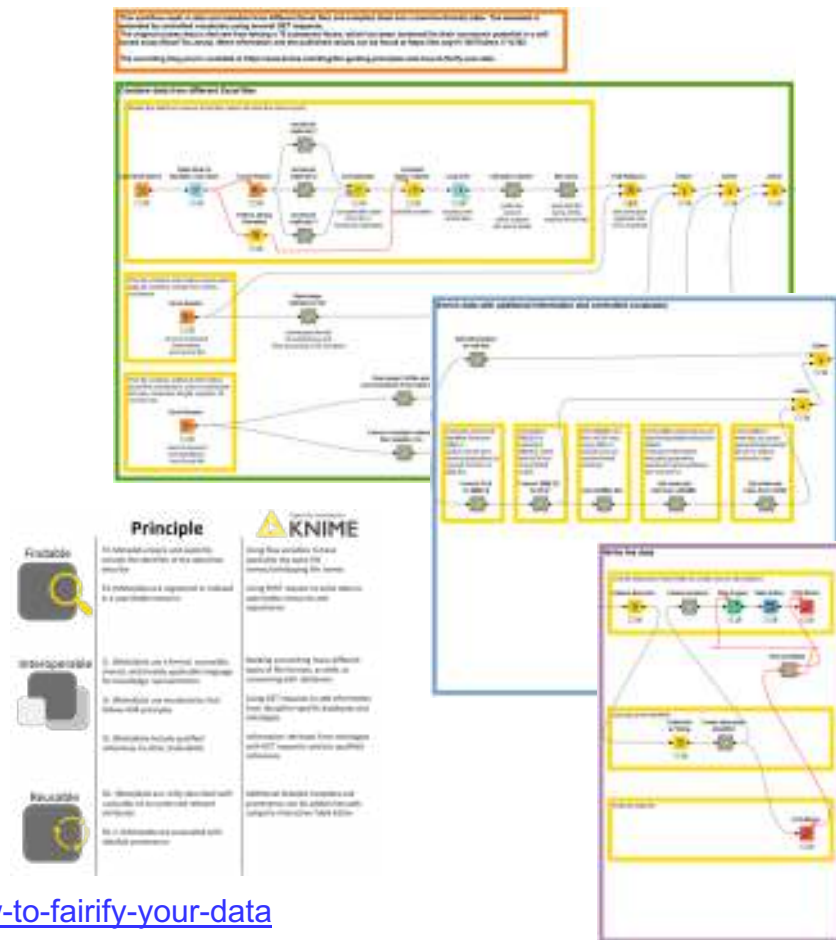
How to use KNIME to make data FAIRer

- Restructure and combine previously generated bioassay data to comply with the FAIR principles
- Extend the metadata with domain-specific controlled vocabulary using GET requests
- Add user-defined information using the Interactive Table Editor

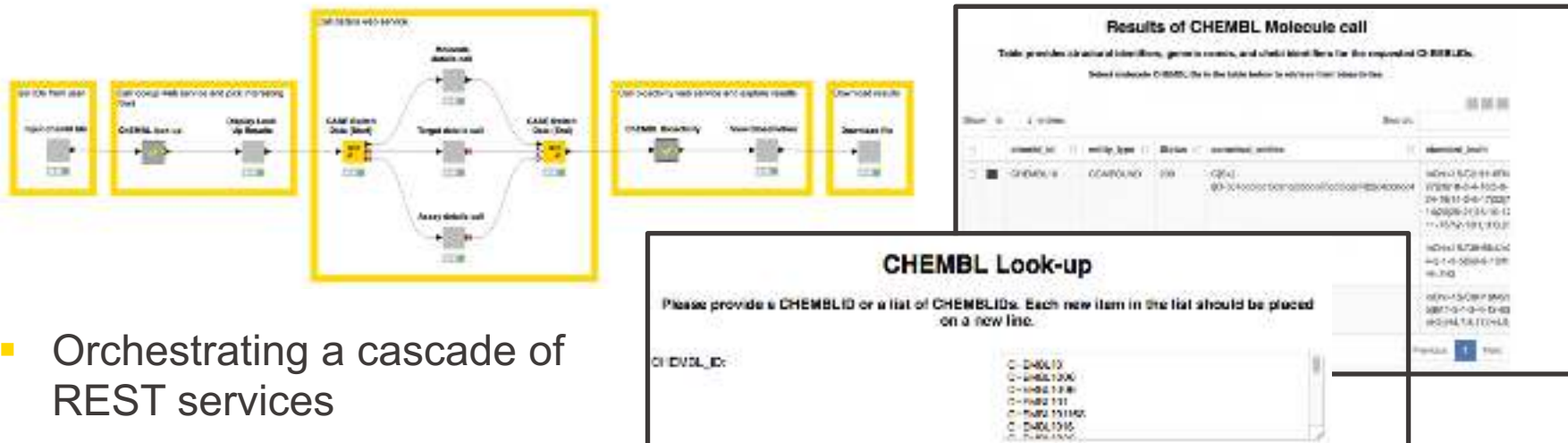


Workflow on the Hub: <https://kni.me/w/Py9va0SQesbauPwS>

Blog Post: <https://www.knime.com/blog/fair-guiding-principles-and-how-to-fairify-your-data>

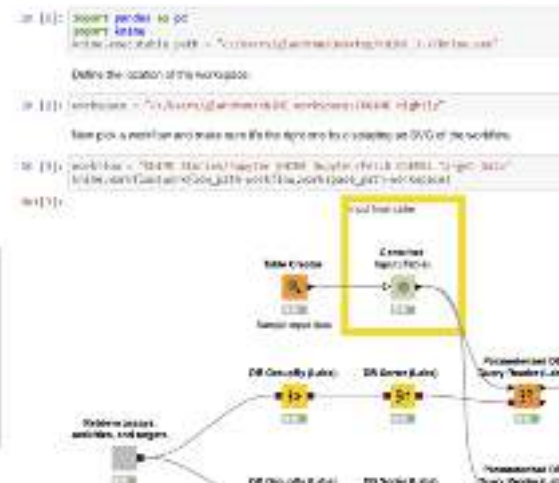
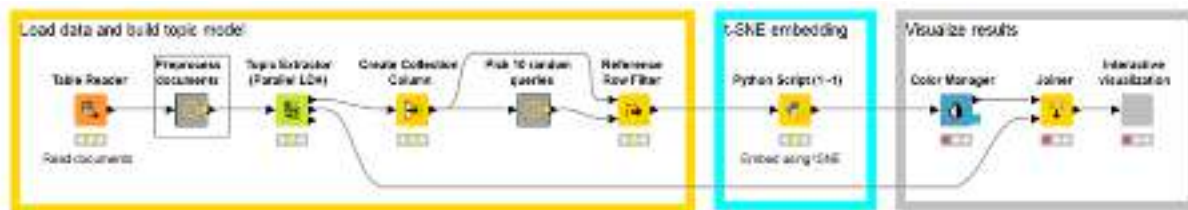


Calling ChEMBL Web Services



Jupyter – KNIME - Jupyter

- Using Python code in Jupyter notebooks from KNIME
- Calling KNIME workflows from Jupyter notebooks



Workflows on KNIME EXAMPLES Server:

07_Scripting/03_Python/03_Fetch_ChEMBL_Target_Data

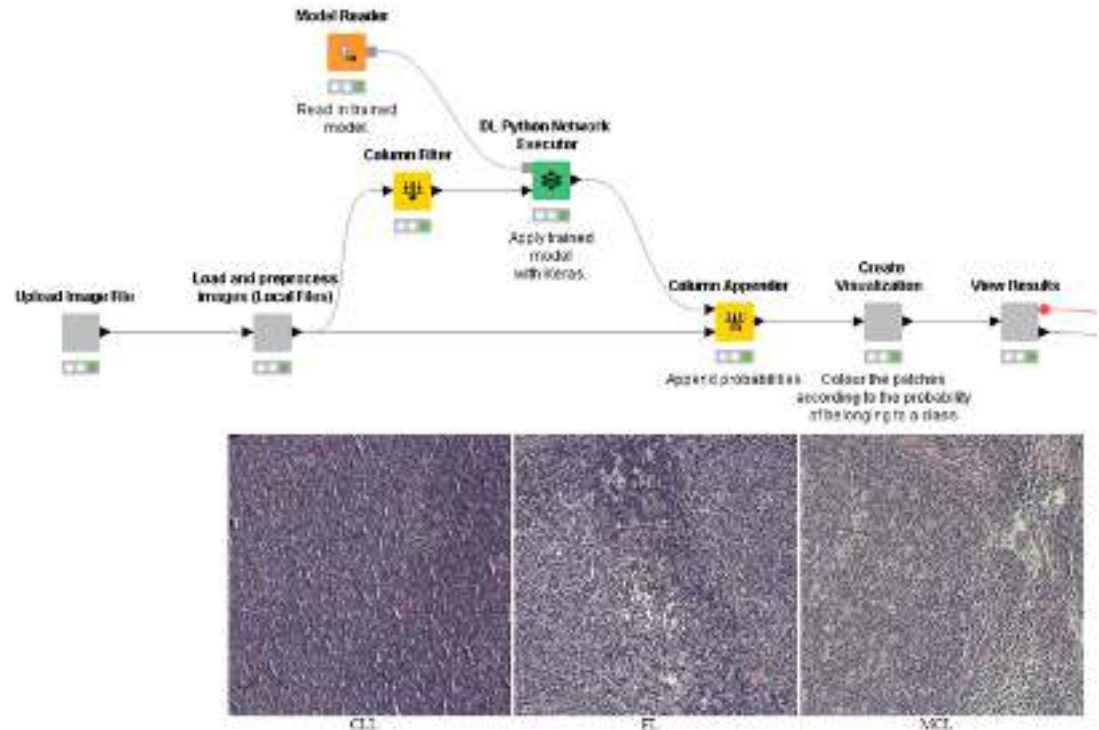
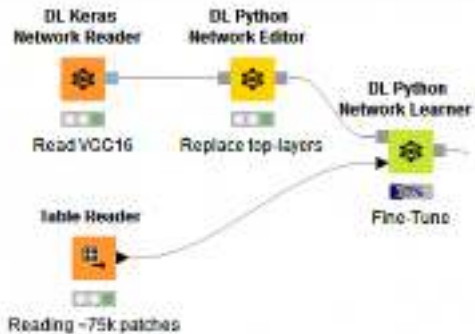
07_Scripting/03_Python/04_Using_Jupyter_from_KNIME_to_embed_documents

Blog post: <https://www.knime.com/blog/knime-and-jupyter>



Deep Learning and Histopathology

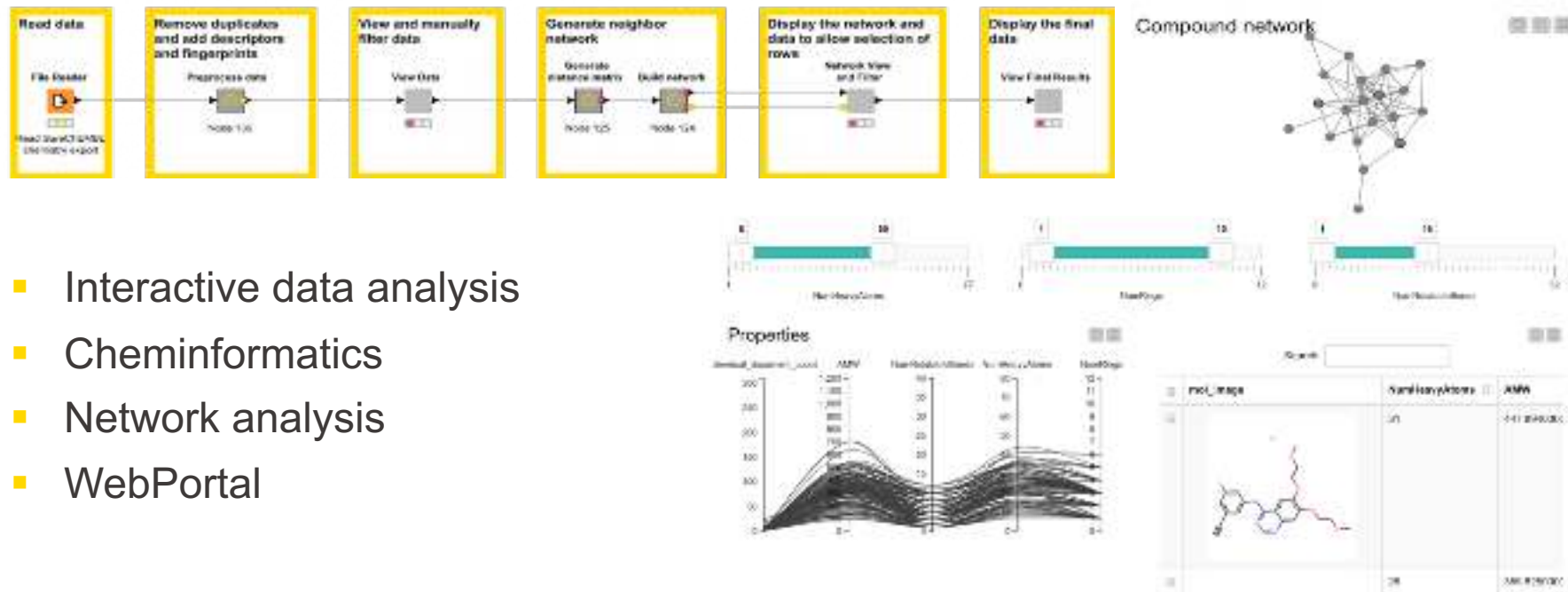
- Image analysis
- Deep learning
- Keras
- Custom visualization



Workflow on KNIME EXAMPLES server: [50_Applications/31_Histopathology_Blog_Post](https://www.knime.com/blog/using-the-new-knime-deep-learning-keras-integration-to-predict-cancer-type-from-histopathology)

Blog post: <https://www.knime.com/blog/using-the-new-knime-deep-learning-keras-integration-to-predict-cancer-type-from-histopathology>

Finding the key compound in a patent



Video from ChemAxon UGM: <https://www.youtube.com/watch?v=MQmmm7tzmB0>

Workflow: https://workflows.knime.com/knime/hub/workflows/50_Applications%3A29_Patent_Network_Analysis%3A01_Tarceva_neighbor_network_from_SureChEMBL_knime://EXAMPLES/50_Applications/29_Patent_Network_Analysis

Some Final Remarks

- Visual workflows allow you build documented and reproducible analyses
- KNIME Analytics Platform is open-source and free to use (no limitations)
- KNIME covers the entire data (science) life cycle

