

KNIME Quickstart Guide

KNIME AG, Zurich, Switzerland Version 4.4 (last updated on 2020-09-07)



Table of Contents

Introduction	1
Start KNIME Analytics Platform	2
Nodes and Workflows	4
Build Your First Workflow	4
Example Workflows	4
Accessing example workflows on the KNIME Hub	4
Accessing example workflows from within KNIME Analytics Platform:	5
Extensions and Integrations	7
Tips & Tricks	20
More Learning Material	24

Introduction

KNIME Analytics Platform is open source software for creating data science applications and services. Intuitive, open, and continuously integrating new developments, KNIME makes understanding data and designing data science workflows and reusable components accessible to everyone.

With KNIME Analytics Platform, you can create visual workflows with an intuitive, drag and drop style graphical interface, without the need for coding.

In this quickstart guide we'll take you through the KNIME Workbench and show you how you can build your first workflow. Most of your questions will probably arise as soon as you start with a real project. In this situation, you'll find a lot of answers in the KNIME Workbench Guide, and in the E-Learning Course on our website. But don't get stuck in the guides. Feel free to contact us and the wide community of KNIME Analytics Platform users, too, at the KNIME Forum. Another way of getting answers to your data science questions is to explore the nodes and workflows available on the KNIME Hub. We are happy to help you there!

Start KNIME Analytics Platform

If you haven't yet installed KNIME Analytics Platform, you can do that on this download page. For a step by step introduction, follow this Installation Guide.

Start KNIME Analytics Platform and when the KNIME Analytics Platform Launcher window appears, define the KNIME workspace here as shown in Figure 1.

A KNIME Analytics Platform Launcher	×
Select a directory as workspace	
KNIME Analytics Platform uses the workspace directory to store its preferences and development artifacts.	
Workspace: C:\Users\KNIME\knime-workspace Browse	
Use this as the default and do not ask again	
Recent Workspaces	
Launch Cancel	

Figure 1. KNIME Analytics Platform Launcher

The KNIME workspace is a folder on your local computer to store your KNIME workflows, node settings, and data produced by the workflow. The workflows and data stored in your workspace are available through the KNIME Explorer in the upper left corner of the KNIME Workbench.

After selecting a folder as the KNIME workspace for your project, click *Launch*. When in use, the KNIME Analytics Platform user interface - the KNIME Workbench - looks like the screenshot shown in Figure 2.

KNIME Analytics Platform File Edit View Node Help □ ▼ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □	•• & D 0 0 0	o o 🗃 🗮 🗮 🔍 🛃	s 2 0 0 0 🖗 🗔		-	Ø X Quick Access
A KNIME Explorer 🐹	 	A "0: My_First_Workflow X			A Description 23	
 My-KNIME-Hub (hub.kn) EXAMPLES (knime@hub.kn) LOCAL (Local Workspace Example Workflows My-First, Project data My-First, Workflow 	KNIME Explorer	My First Workflow Select columns, filter	rows and visualize sales data.	Workflow Editor	This node c ASCII file or to read varios When you open the node's con dialog and provide a filename, it thiss	figuration to guess
Workflow Coach 23 Recommended Nodes Joiner Joiner Workflow Filter Postitioning PorcupBy Missing Value	Workflow Coach	File Reader	Column Filter Row Fil Column Filter Row Fil Column Country Include country date and amount with cou	Sales over time Soles over time PielDonut Chart onty	the reader's settings by analyzing th of the file. Check the results of these in the preview table. If the data sho correct or an error is reported, you or the settings manually (see below). The file analysis runs in the backgr can be out short by clicking the "Qui which shows if the analysis takes I	e settings wn is not an adjust ound and ck scan", onger. In
Node Repository			"unimo	wn"		AE Hub arch
> 💾 IO > 🖋 Manipulation	Node Repository	E Outine 12	Outline		s Platform v4.0.0.v20190620	nsole
				1		

Figure 2. KNIME Workbench

The KNIME Workbench is made up of the following components:

- **KNIME Explorer**: Overview of the available workflows and workflow groups in the active KNIME workspaces, i.e. your local workspace, KNIME Servers, and your personal KNIME Hub space.
- Workflow Coach: Lists node recommendations based on the workflows built by the wide community of KNIME users. It is inactive if you don't allow KNIME to collect your usage statistics.
- **Node Repository**: All nodes available in core KNIME Analytics Platform and in the extensions you have installed are listed here. The nodes are organized by categories but you can also use the search box on the top of the node repository to find nodes.
- Workflow Editor: Canvas for editing the currently active workflow.
- **Description**: Description of the currently active workflow, or a selected node (in the Workflow Editor or Node Repository).
- Outline: Overview of the currently active workflow.
- **Console**: Shows execution messages indicating what is going on under the hood.

i

A more detailed explanation of the different views in the KNIME Workbench is provided in the KNIME Workbench Guide, as well as on the introductory video *The KNIME Workbench*.

Nodes and Workflows

In KNIME Analytics Platform, individual tasks are represented by nodes. Each node is displayed as a colored box with input and output ports, as well as a status, as shown in Figure 3. The input(s) are the data that the node processes, and the output(s) are the resulting datasets. Each node has specific settings, which we can adjust in a configuration dialog. When we do, the node status changes, shown by a traffic light below each node. Nodes can perform all sorts of tasks, including reading/writing files, transforming data, training models, creating visualizations, and so on.

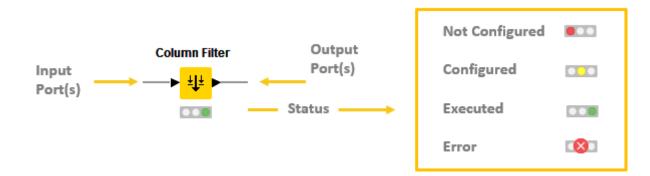


Figure 3. Node ports and node status

A collection of interconnected nodes constitutes a workflow, and usually represents some part - or perhaps all - of a particular data analysis project.

Build Your First Workflow

Let's now start building an example workflow, where we analyze some sales data. When we're finished it will look like the workflow shown in Figure 4. Don't worry if you get stuck along the way, the finished workflow is also available on the KNIME Hub here.

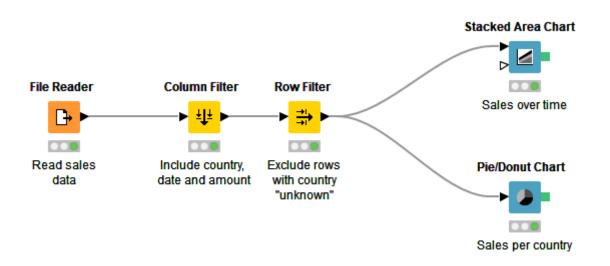


Figure 4. Example workflow

The example workflow in Figure 4 reads data from a CSV file, filters a subset of the columns, filters out some rows, and visualizes the data in two graphs: a stacked area chart and a pie chart, which you can see in Figure 5: one showing the development of sales over time, and the other showing the share of different countries on total sales.

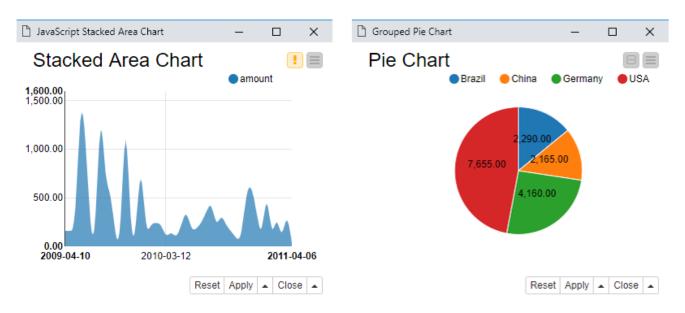


Figure 5. Output views of the example workflow

To get started, first download the CSV file that contains the data that we are going to use in the workflow. You can find it here. Next, create a new, empty workflow by:

- Clicking New in the toolbar panel at the top of the KNIME Workbench
- Or by right clicking a folder of your local workspace in the KNIME Explorer, as shown in Figure 6

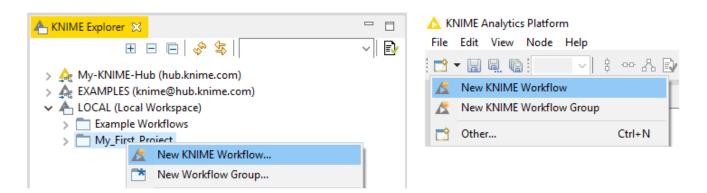


Figure 6. Creating a new, empty workflow

The first node you need is the File Reader node, which you'll find in the node repository. You can either navigate to $IO \rightarrow Read \rightarrow File Reader$, or type a part of the name in the search box in the node repository panel.

To use the node in your workflow you can either:

- Drag and drop it from the node repository to the workflow editor
- Or double click the node in the node repository. It automatically appears in the workflow editor.

Let's now define the settings for this node:

• Open the configuration dialog either by double clicking the File Reader node, or by right clicking it and selecting *Configure...* as shown in Figure 7.

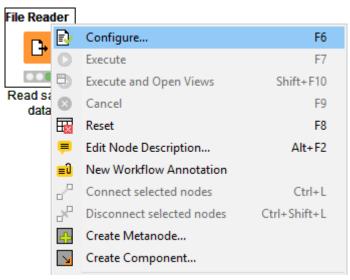


Figure 7. Configuring a node

• In the configuration dialog, define the file path by clicking the *Browse* button, then check other available settings, and preview the data as shown in Figure 8.

🛕 Dialog - 0:1 - File Reader (Rea	id sales)		×
file	a suresy		~
iie.			
Settings Flow Variables Job Ma	nager Selection Memory Policy		
Enter ASCII data file location: (pr	ess 'Enter' to update preview)		^
knime://knime.workflow/data/sa	les_data.csv	V Browse	
Pr	reserve user settings for new location	Rescan	
Basic Settings			
read row IDs	Column delimiter: , \sim	Advanced	
read column headers	✓ ignore spaces and tabs		
	Java-style comments	Single line comment:	
. .			
Preview Click column head	der to change column properties (* = name	e/type user settings)	
			¥
	OK Apply	Cancel	

Figure 8. Configuration dialog of the File Reader node

You may now want to inspect the output table to see if the data file was read as you intended. To inspect the output table:

- Execute the File Reader node by right clicking the node and selecting Execute
- Open the output table by right clicking the executed node and selecting the last option in the menu: *File Table*

If the data was read in correctly, add the Column Filter node to the workflow editor and connect it to the File Reader node:

- Click the output port of the File Reader node, holding the mouse button and releasing it at the input port of the Column Filter node
- Alternatively, select the File Reader node by clicking it once in your workflow, and then double clicking the Column Filter node in the node repository. This method automatically connects the Column Filter node to the File Reader node.

Before you proceed, you must configure the Column filter node:

· Move the columns "country", "date", and "amount" into the green-framed Include field

either by double clicking them, or using the buttons between the *Exclude* and *Include* fields in the configuration dialog shown in Figure 9

∧ Dialog - 0:2 - Column Filter (Include country,) File	-		×
FIIE			
Column Filter Flow Variables Job Manager Selection Memory Policy Manual Selection Wildcard/Regex Selection Trype Selection Exclude Filter guantity S card Cust_ID Cust_ID Enforce exclusion Image: Column Filter Column Filter Image: Column Filter Image: Column Filter Image: Column Filter Image: Column Filter Image: Column Filter Image: Column Filter Image: Column Filter Image: Column Filter Image: Column Filter Image: Column Filter Image: Column Filter Image: Column Filter Image: Column Filter Image: Column Filter Image: Column Filter Image: Column Filter Image: Column Filter Image: Column Filter Image: Column Filter Image: Column Filter Image: Column Filter Image: Column Filter Image: Column Filter Image: Column Filter Image: Column Filter Image: Column Filter Image: Column Filter Image: Column Filter Image: Column Filter Image: Column Filter Image: Column Filter Image: Column Filter Image: Column Filter Image: Column Filter Image: Column Filter <			
OK Apply Cancel	6	2	

Figure 9. Configuring the Column Filter node

• Finish the configuration by clicking OK

Continue building the workflow:

- Adding the Row Filter node to the workflow editor and connecting it to the Column Filter node
- Open the configuration dialog of the Row Filter node and exclude rows from the input table where the column "country" has the value "unknown" as shown in Figure 10

A Dialog - 0:3 - Row Filter (Exclude File Filter Criteria Flow Variables Job M	
 Include rows by attribute value Exclude rows by attribute value Include rows by number Exclude rows by number Include rows by row ID Exclude rows by row ID 	Column to test: S country
	OK Apply Cancel 🕐

Figure 10. Configuring the Row Filter node

Now that the data has been filtered, let's move on to data visualization:

- Search for the nodes Stacked Area Chart and Pie/Donut Chart in the node repository, and add them to the workflow editor, both connected to the Row Filter node
- Open the configuration dialog of the Stacked Area Chart node. Select the column "date" as the x-axis column as shown in Figure 11.

▲ Dialog - 0:4 - Stacked A File	rea Chart (JavaScript)	(Sales over time)			-	-	×
Options General Plot Option	ns Axis Configuration	Control Options	Flow Variables	Job Manager Selection	Memory Policy		
- General Settings		Ge	enerate image				
	м	aximum number of	rows	2,500 🜩			
		Choose column for	x-axis Sdate	2 ~			
		Sort f	for x-axis column	1			
	۲	Manual Selection	○ Wildcard/Re	gex Selection			_
Filter			T	<i>Filter</i>			
No c	olumns in this list			amount			
			>>>				
			<				
			«				
Enforce exclusion			0	Enforce inclusion			
			OK	Apply	Cancel	0	

Figure 11. Configuring the Stacked Area Chart node

• Now open the configuration dialog of the Pie/Donut Chart node and select "country" as category column, "Sum" as aggregation method, and "amount" as frequency column for the pie chart. Both of these configuration options are shown in Figure 12.

▲ Dialog - 0:16 - Pie/Donut Chart (JavaScript) — □	×
File	
Flow Variables Job Manager Selection Memory F	olicy
Options General Plot Options Control Options Intera	activity
General Settings	
Generate image	
Category Column S date 🗸	
Aggregation Method	
Occurence Count ○ Sum ○ Average	
Report on missing values	
Include 'Missing values' category	
Frequency Column	
Process table in memory	
OK Apply Cancel ?	

Figure 12. Configuring the Pie/Donut Chart node

The workflow is finished, and the next step is to execute it and view the output. You do this either by clicking the "Execute all executable nodes" button in the toolbar shown in Figure 13...



Figure 13. Executing all executable nodes from the toolbar

..., or by selecting the last nodes of the different branches of the workflow, right clicking the selection, and then clicking *Execute* in the menu.

To inspect the interactive output view of a JavaScript based node:

Choose the *Execute and Open Views* option for an unexecuted node as shown in Figure
14

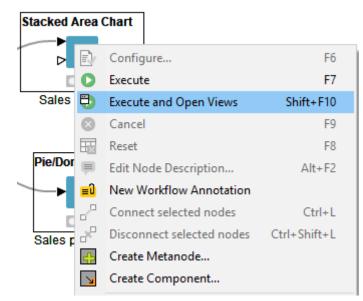


Figure 14. Executing and opening interactive view

• Or, once a node is executed, right click the node and select *Interactive View:* ... instead, as shown in Figure 15

Stacked A	rea (Chart	
	Ē	Configure	F6
	C	Execute	F7
Sales	8	Execute and Open Views	Shift+F10
	8	Cancel	F9
	Ξx	Reset	F8
Pie/Don	F	Edit Node Description	Alt+F2
	≡Ĵ	New Workflow Annotation	
	~	Connect selected nodes	Ctrl+L
Sales p	28	Disconnect selected nodes	Ctrl+Shift+L
	4	Create Metanode	
	4	Create Component	
	Q	Interactive View: JavaScript Stacked Area Chart	

Figure 15. Opening interactive view of an executed node

Currently, the pie chart uses default colors for different countries in the data. With the Color Manager node, you can assign the countries other colors than the default ones you see in Figure 5. The colors have to be assigned before building the graph, so you'll have to add the Color Manager node in the middle of the workflow.

Add the Color Manager node:

• By dragging the node from the node repository and releasing it on its place between the Row Filter node and the Pie Donut Chart node in the workflow when the connection has

turned red, as shown in Figure 16. The red connection means that it's ready to accept the new node when you release the mouse.

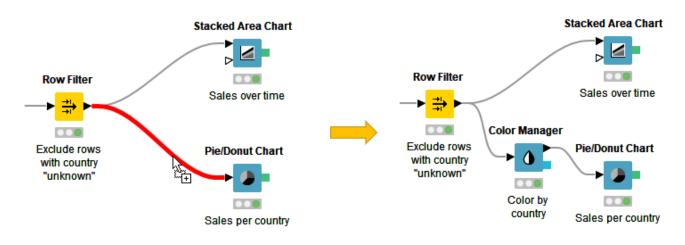


Figure 16. Insert a node between two nodes in a workflow

Example Workflows

Jump start your analysis with the example workflows on the KNIME Hub. You can either browse the example workflows directly on hub.knime.com or via the EXAMPLES mountpoint in the KNIME Explorer. You can copy the workflows to your local workspace, modify them according to your use case, or copy parts of them to your workflow. The workflows on the KNIME Hub are also a useful resource to learn about different use cases in KNIME Analytics Platform.

Accessing example workflows on the KNIME Hub

- Go to the KNIME Hub
- Next, enter a search term in the search field as shown in Figure 17

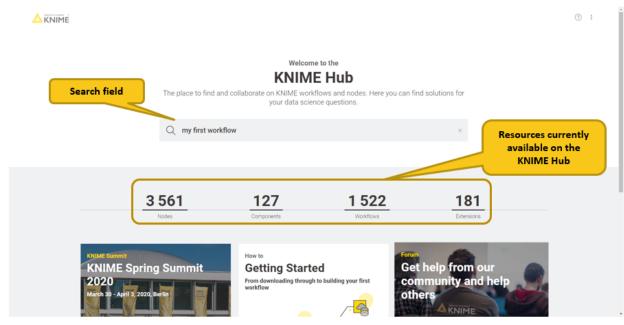


Figure 17. Searching for workflows on the KNIME Hub

- · Navigate through the workflows in the search results
- Select one workflow, and click the "Open workflow" button on the right as shown in Figure 18

	Q Search workflows, nodes and more	⑦ :
KNIME Hub , knime , Spaces , Exar ତ୍ୱରି Visual Analys	Inples > 02_ETL_Data_Manipulation > 00_Basic_Examples > 00_Visual_Analysis_of_Sales_Data	
My First Workflow - Visualizati	on of Sales Data	Open in KNIME Analytics Platform
- Filter data to columns that cont - Filter data to rows that you wan - Visualize the data using a Staci	I to include in your analysis ed Area Chart and a Pie/Donut Chart Data Visualization	Open workflow or download workflow
Data Access - Provide file path File Reader	- Show sales by time and country Data Preprocessing - Filter columns and rows Column Filter Row Filter	By downloading the workflow, you agree to our terms and conditions.
Read sales_2008-2011.csv	Sales by time	D Short link https://mi.me/w/88_0uf-mtF0acBt-

Figure 18. Opening a workflow from the KNIME Hub

You can also install extensions and add nodes or components to your workflow just by dragging and dropping them from the KNIME Hub page to the KNIME Workbench.

Accessing example workflows from within KNIME Analytics Platform:

- Expand the EXAMPLES mountpoint in the KNIME Explorer
- Next, double click to see the example workflows ordered by categories, as shown in Figure 19. No credentials are necessary.

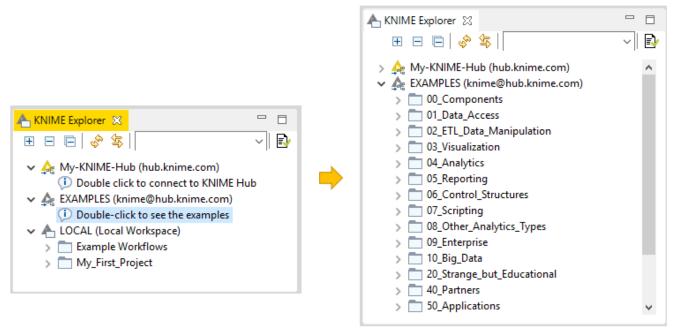


Figure 19. Logging in to the EXAMPLES mountpoint

- Inside these categories, some workflow groups are named after single operations, e.g. filtering
- Other workflow groups have names that refer to broader topics, e.g. time series analysis
- The "50_Applications" workflow group contains workflows that cover entire use cases like churn prediction or fraud detection

To download an example workflow:

- Drag and drop
- Or, copy and paste

the workflow into your LOCAL workspace. Double click the downloaded copy of the example workflow to open and edit it like any other workflow.

i

The video The EXAMPLES Server provides a more detailed introduction to the EXAMPLES server.

Extensions and Integrations

If you want to add capabilities to KNIME Analytics Platform, you can install extensions and integrations. The available extensions range from free open source extensions and integrations provided by KNIME to free extensions contributed by the community and commercial extensions including novel technology nodes provided by our partners.

The KNIME extensions and integrations developed and maintained by KNIME contain deep learning algorithms provided by Keras, high performance machine learning provided by H2O, big data processing provided by Apache Spark, and scripting provided by Python and R, just to mention a few.

Install extensions by:

- Clicking *File* on the menu bar and then *Install KNIME Extensions*.... The dialog shown in Figure 20 opens.
- Selecting the extensions you want to install
- · Clicking Next and following the instructions
- Restarting KNIME Analytics Platform

▲ Install			
Available Software			
Check the items that you wish to install.			
Name	Version		^
	version		
> 000 KNIME & Extensions			
> 000 KNIME Big Data Extensions			
> 000 KNIME Community Extensions - Bioinformatics & NGS			
> 100 KNIME Community Extensions - Cheminformatics			
> 100 KNIME Community Extensions - Image Processing and Analysis			
> IIII KNIME Community Extensions - Other			
> 000 KNIME Community Extensions Sources			
> 000 KNIME Labs Extensions			~
Select All Deselect All			
Details			
			E.4
Show only the latest versions of available software	Hide items that are already installed		
Group items by category	What is <u>already installed</u> ?		
Show only software applicable to target environment			
	< Back Next >	Finish	Cancel

Figure 20. Installing Extensions and Integrations

The KNIME extensions and trusted community extensions are available per default via an URL to their update sites. Other extensions can be installed by first adding their update sites.

To add an update site:

- Navigate to File \rightarrow Preferences \rightarrow Install/Update \rightarrow Available Software Sites
- Click Add...

i

- And either add a new update site by providing a URL via the Location field
- Or, by providing a file path to a zip file that contains a local update site, via Archive...
- Finally, give the update site some meaningful name and click OK

After this is done, the extensions can be installed as described further above.



information about extensions can be found in the Extensions and Integrations Guide.

Update to the latest KNIME version by:

- Clicking *File* and then *Update KNIME*... to make sure that you use the latest version of the KNIME Software and the installed extensions
- In the window that opens, select the updates, accept the terms and conditions, wait until the update is finished, and restart KNIME Analytics Platform

Tips & Tricks

• Get Help and Discuss at the KNIME Forum

Log in to our KNIME Community Forum, and join the discussions under different categories ranging from KNIME Analytics Platform, to extensions and integrations, special interest groups, and KNIME development. The forum is a lively community, where KNIME staff, along with other experienced KNIME users, are available to answer your questions.

Import and Export Workflows

To import a workflow or a workflow group, right click anywhere in the local workspace in the KNIME Explorer and select *Import (Export) KNIME Workflow...*, as shown in Figure 21

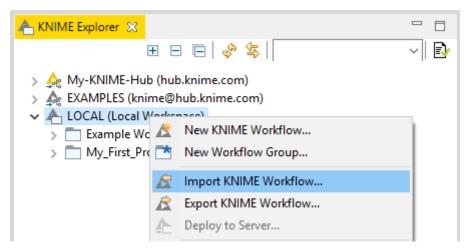


Figure 21. Importing and exporting workflows and workflow groups

Then, follow the steps explained below and shown in Figure 22:

- To export a workflow or a workflow group first select the workflow (or group) you want to export
- Next, write the path to the destination folder and the file name. If you export a workflow group, you can select the elements you want to export from inside the folder.

				_	~
🛆 Import — 🗆	X 🔥 Expor	t			×
Workflow Import Selection	Export K	NIME workflow(s)			
Select the items to import.	Exports #	NIME workflows.			
Source:					
Select file: C:\Users\Public\My_First_Pr Browner	/se Select w	orkflow(s) to export:	LOCAL:/My_Fit	Select	t
O Select root directory: Bro	/se Destinat	ion workflow file name (.kna	r): C:\Users\Public	Brow	se
Destination:	Options				
Select folder: LOCAL:/My_First_Project Bro	/se Rese	t Workflow(s) before export			
Import Elements:					
V My_First_Project					
V 🗹 🗖 data		sales_data.csv			
Sales_data.csv Desel	t All	My_First_Workflow			
My_First_Workflow	Select	All Deselect all			
< Back Next > Finish Ca	cel		Finish	Cance	:1

Figure 22. Defining path to a file to import or export

· Import Data by Dragging and Dropping a Data File

You can import a data file from the KNIME workspace or any location on your system by dragging and dropping it from the KNIME Explorer, Desktop or File Explorer to the workflow editor as shown in Figure 23. This method automatically creates the right node to read the file type, and it preconfigures the node by populating the file path setting with a file path URL relative to the KNIME Explorer location.

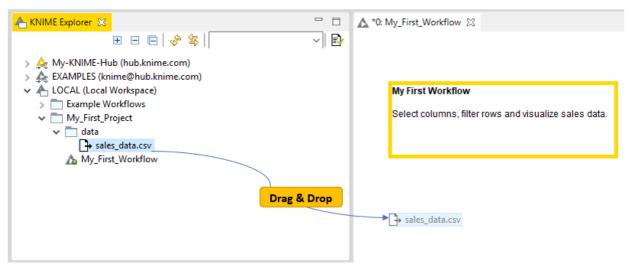


Figure 23. Reading data files by drag and drop

• Replace a Node in a Workflow

You can replace a node in your workflow by dragging a node from the repository and dropping it on top of an existing node as soon as a white arrow and boxes appear

inside it as shown in Figure 24.

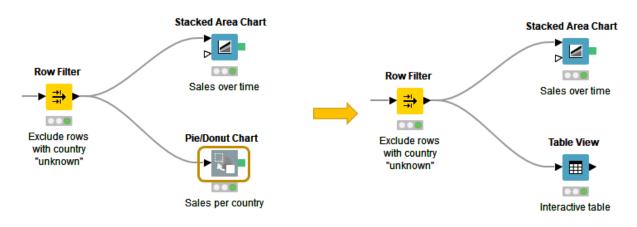


Figure 24. Replacing a node in a workflow

Expand Your Node Search: Fuzzy Search and Crisp Search

If you are not sure of the name of the node you're searching for, switch to fuzzy search mode in the node repository by clicking the icon next to the search field as shown in Figure 25. Your search results will now include any nodes related to the search term. In the crisp search mode, the search text must exactly match the node name itself. With more practice building workflows, you'll remember more and more node names. After some time you'll probably switch back to the crisp search mode to find the node you're looking for faster.

Crisp Search	Fuzzy Search						
🖍 Node Repository 🙀 🛛 grouping 🔍 🖓 🗖	🖌 Node Repository 🛒 grouping 🗸 🖓 🗖						
No matches found.	 Ungroup - /Manipulation/Row/Transform GroupBy - /Manipulation/Row/Transform Pivoting - /Manipulation/Row/Transform Spark GroupBy - /Tools & Services/Apache Spark/Row Group Loop Start - /Workflow Control/Loop Support Unpivoting - /Manipulation/Row/Transform Database GroupBy - /Database/Manipulation Independent groups t-test - /Analytics/Statistics/Hypothe DB GroupBy (Labs) - /KNIME Labs/DB (Labs)/Query Row Sampling - /Manipulation/Row/Transform Group Mutual Information - /Community Nodes/HCS Tool Grouped ScatterPlot - /Scripting/R/Meta Nodes Grading (3.7) - /Analytics/Mining/Weka/Weka (3.7)/Classi Profit Engine - /Scientific Strategy/Market Simulation Bagging - /Analytics/Mining/Ensemble Learning/Meta Notes 						

Figure 25. Crisp and fuzzy search mode

Monitor the State of a Node

If you want to see the intermediate output tables in your workflow, you can add a Node Monitor panel to your KNIME Workbench:

 \circ Click View → Other → KNIME Views → Node Monitor as shown in Figure 26

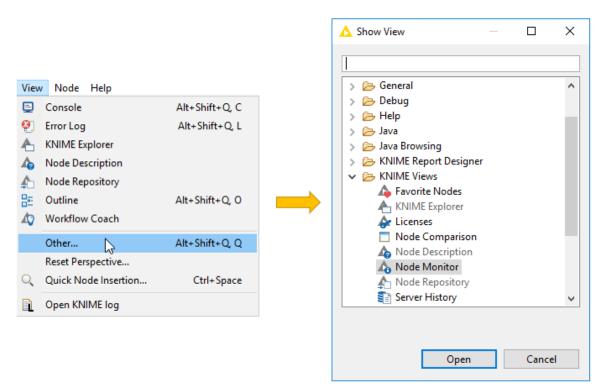


Figure 26. Adding a Node Monitor to the KNIME Workbench

 Next select Show Output Table in the View Menu in the upper right corner of the Node Monitor as shown in Figure 27. This panel now shows the output of the node you have selected in the workflow.

	/ ▽ □ □		Cons	ole 🧥 No File Reade		/ ▽□ Ε						
•	Show Output Table Show Variables Show Configuration Show Entire Configuration		State:	EXECUTED								
		Þ	Port O	Coutput Port 0 ~ Load data								
	Show Node Timing Information		Row0	prod_4	unknown	2008-12-12	1	3	?	Cust_8		
	Show Graph Annotations		Row1	prod_3	China	2009-04-10	2	160	Ν	Cust_2		
			Row2	prod_3	China	2009-04-10	2	160	Υ	Cust_5		
			Row3	prod_3	China	2009-05-10	2	160	?	Cust_2	~	

Figure 27. Showing output table in the node monitor

Besides the output table, you can also have the panel show the execution time of the node. You can also pin the output of one node. This means that the output of the currently selected node is shown independently of the selection in the workflow editor.

More Learning Material

- Installation Guide
- KNIME Workbench Guide
- Cheat Sheet(s)
- E-Learning Course
- Navigate the Learning Options at KNIME
- 7 things to do after installing KNIME Analytics Platform





KNIME AG Talacker 50 8001 Zurich, Switzerland www.knime.com info@knime.com

The KNIME® trademark and logo and OPEN FOR INNOVATION® trademark are used by KNIME AG under license from KNIME GmbH, and are registered in the United States. KNIME® is also registered in Germany.