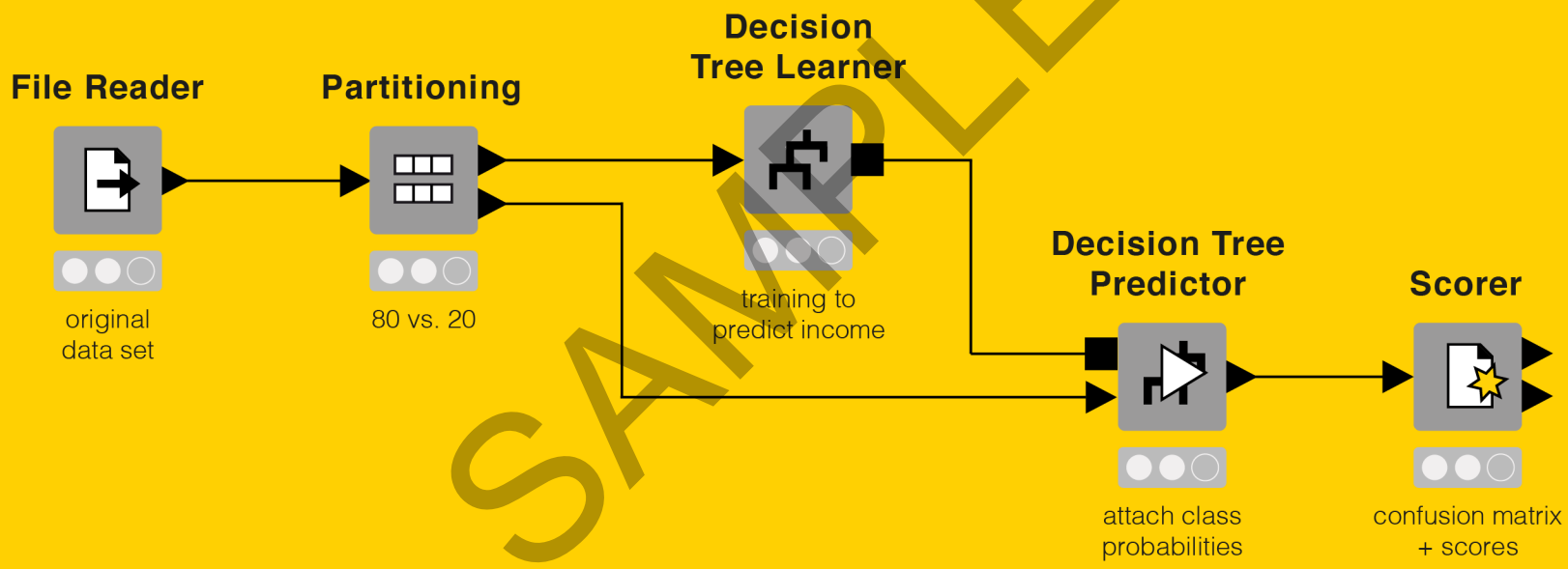
KNIME®
BEGINNER'S LUCK



SAMPLE

A Guide to KNIME Analytics Platform for Beginners

Authors: Satoru Hayasaka and Rosaria Silipo

SAMPLE

Copyright© 2021 by KNIME Press

All Rights Reserved. This publication is protected by copyright, and permission must be obtained from the publisher prior to any prohibited reproduction, storage in a retrieval system, or transmission in any form or by any means, electronic, mechanical, photocopying, recording or likewise.

This book has been updated for **KNIME 4.3**.

For information regarding permissions and sales, write to:

KNIME Press
Technoparkstr. 1
8005 Zurich
Switzerland

knimepress@knime.com

ISBN: 978-3-033-02850-0

SAMPLE

Table of Contents

Foreword 12
Acknowledgements 13
Chapter 1. Introduction 14
1.1. Purpose and structure of this book 14
1.2. KNIME community 15
Useful Web Pages 15
Courses, Events, and Videos 16
Books 16
KNIME Hub 17
1.3. Download and install KNIME Analytics Platform 19
1.4. Workspace 21
The “Workspace Launcher” 21
1.5. KNIME workflow 22
What is a workflow 22
What is a node 23
1.6. *.knwf* and *.knar* file extensions 23
1.7. KNIME workbench 24
The KNIME Workbench 26
Top menu 27
Tool Bar 30
Hotkeys 31
Node Repository 32
Search box 32
KNIME Explorer 32

SAMPLE

My-KNIME-Hub 33
EXAMPLES Server 34
Mounting Servers in KNIME Explorer 35
Workflow Editor 36
Customizing the Workflow Editor 37
Workflow Annotations 37
Other Workbench Customizations 38
Node Monitor View 38
1.8. Download the KNIME Extensions 39
Installing KNIME Extensions 39
1.9. Data and workflows for this book 40
1.10. Exercises 41
Exercise 1 41
Exercise 2 42
Exercise 3 43
Chapter 2. My first workflow 46
2.1. Workflow operations 46
Create a new Workflow Group 47
Create a new workflow 48
Save a workflow 49
Delete a workflow 49
2.2. Node operations 50
Create a new node 50
Configure a node 51
Execute a node 51

SAMPLE

Node Text 52
Node Description 52
View the processed data 53
2.3. Read data from a file 53
Create a “CSV Reader” node 54
Configure the “CSV Readernode 55
Customizing Column Properties 56
The *knime://* protocol 58
2.4. KNIME data structure and data types 59
KNIME data structure 61
2.5. Filter Data Columns 61
Create a “Column Filter” node 62
Configure the “Column Filter” node 64
2.6. Filter Data Rows 65
Create a “Row Filter” node 66
Configure the “Row Filter” node 66
Row filter criteria 68
2.7. Write Data to a File 70
Create a “CSV Writer” node 70
Configure the “CSV Writer” node 71
2.8. Exercises 72
Exercise 1 72
Exercise 2 75
Chapter 3. My first data exploration 79
3.1. Introduction 79

SAMPLE

3.2. Replace Values in Columns 80
Column Rename 81
Rule Engine 83
3.3. String Splitting 85
Cell Splitter by Position 86
Cell Splitter [by Delimiter] 87
RegEx Split (= Cell Splitter by RegEx) 88
3.4. String Manipulation 89
String Manipulation 89
Case Converter 91
String Replacer 92
Column Combiner 93
Column Resorter 94
3.5. Type Conversions 95
Number To String 95
String To Number 96
Double To Int 97
3.6. Database Operations 97
SQLite Connector 99
MySQL Connector 100
Workflow Credentials 101
DB Writer 102
Import a JDBC Database Driver 103
DB Table Selector 105
DB Reader 106

SAMPLE

3.7. Aggregations and Binning 106
Numeric Binner 108
GroupBy: “Groups” tab 109
GroupBy: Aggregation tabs 110
Pivoting 111
3.8. Nodes for Data Visualization 113
3.9. Scatter Plot 113
Scatter Plot: Interactive View 115
3.10. Graphical Properties 116
Color Manager 117
3.11. Line Plots and Parallel Coordinates 119
Line Plot 119
Parallel Coordinates 121
3.12. Bar Charts and Histograms 122
Bar Chart 123
Table View 126
3.13. Exercises 128
Exercise 1 128
Exercise 2 130
Exercise 3 130
Chapter 4. My First Model 134
4.1. Introduction 134
4.2. Split and Combine Data Sets 135
Row Sampling 135
Partitioning 136

SAMPLE

Shuffle 137
Concatenate 138
4.3. Transform Columns 139
PMML 140
Missing Value 141
Normalizer 142
Normalization Methods 143
Normalizer (Apply) 143
4.4. Machine Learning Models 144
Naïve Bayes Model 145
Naïve Bayes Learner 146
Naïve Bayes Predictor 146
Scorer (Javascript) 148
Decision Tree 152
Decision Tree Learner: Options Tab 153
Decision Tree Learner: PMML Settings Tab 154
Decision Tree Predictor 155
Decision Tree View 160
ROC Curve 161
Artificial Neural Network 163
RProp MLP Learner 163
Multilayer Perceptron Predictor 165
Write/Read Models to/from file 166
Model Writer 166
Model Reader 168

SAMPLE

Statistics 169
Regression 171
Linear Regression Learner 172
Regression Predictor 173
Clustering 173
k-Means 174
Cluster Assigner 175
Hypothesis Testing 175
4.5. Exercises 176
Exercise 1 176
Exercise 2 178
Exercise 3 178
Chapter 5. Preparing Data for Reporting 180
5.1. Introduction 180
5.2. Transform Rows 180
RowID 183
Unpivoting 184
Sorter 186
5.4. Joining Columns 186
Joiner 188
Joiner node: the „Joiner Settings" tab 189
Joiner node: the "Column Selection" tab 190
Join mode 191
5.5. Misc Nodes 192
Java Snippet (simple) 193

SAMPLE

Java Snippet 194
Math Formula 195
Math Formula (Multi Column) 196
5.7. Cleaning Up the Final Workflow 198
Collapse pre-existing nodes into a Meta-node 198
Create a Meta-node from scratch 199
Expand and Reconfigure a Meta-node 201
5.8. Exercises 202
Exercise 1 202
Exercise 2 203
Exercise 3 204
Chapter 6. Reporting with BIRT 207
6.1. Reporting with KNIME Analytics Platform 207
6.2. Installing the Report Designer Extension (BIRT) 209
6.3. Marking Data in the Workflow 209
Data to Report 210
6.4. From KNIME to BIRT and Back 210
6.5. The BIRT Environment 211
6.6. The Layout 215
6.7. The Tables 218
Toggle Breadcrumb 222
6.8. Page Break 227
6.9. The Charts 227
Select Chart Type 228
Select Data 229

SAMPLE

Format Chart 231
6.10. Style Sheets 239
Create a new Style Sheet 240
Apply a Style Sheet 241
6.11. Generate the final document 242
6.12. Dynamic Text 242
6.13. Exercises 245
Exercise 1 245
References 247
Node and Topic Index 248

SAMPLE

Foreword

This is the first book I wrote in 2010 for the KNIME Press on how to use KNIME Analytics Platform. Since we are getting close to the 10-year anniversary of this book, we (the KNIME Press Team and I) thought that it might be in need of a new Foreword text. It does not actually need an overall update, as ever since its birth it has been updated twice a year every year, following each new release of KNIME Analytics Platform; not immediately after - but close enough.

That is right! KNIME Beginner's Luck, like all other e-books from KNIME Press, is a live e-book, constantly changing to fit the newest version of the software. This liveness of the e-book is also the reason why it has only rarely been printed. Updating printed pages is undoubtedly harder than updating a pdf file!

As this is the first book, it is inevitably about the basics: the basics of KNIME Analytics Platform of course and also the basics of a data science project. This book guides you through the most important access functions, data transformation operations, and of course machine learning nodes available in KNIME Analytics Platform. Supplemented with many example workflows, exercises, and screenshots, it will quickly familiarize you with the basic functions of the software. If you are looking for more advanced topics, you won't find them here, instead….

If you want to learn more about advanced machine learning algorithms, flow variables, or loops, check the sequel to this book: "KNIME Advanced Luck". If you want to learn more about text processing, have a look at the book, "From Words To Wisdom". If you come from that school of thoughts where reading manuals or instructions is overrated, you can start directly with reading about solutions to case studies in various application fields in our collection "Practicing Data Science". If your job is more about integrating and blending different data sources and data types, then the book for you is the "Will they blend?" collection. More useful booklets are available on the KNIME Press page, if you are transitioning from SAS, Excel, or Alteryx.

All this is to say that the KNIME Press team and I have been working hard to provide you with the learning material, books, and tutorials, to become progressively more and more productive with KNIME Software and data science concepts.

Rosaria Silipo (Author of a number of KNIME Press Books, PhD)

SAMPLE

Acknowledgements

First of all, I would like to thank the whole KNIME Team for their patience in dealing with me and my infinite questions.

Among all others in the KNIME Team I would like to specifically thank Peter Ohl for having reviewed this book in order to find any possible aspects that were not compatible with KNIME best practice.

I would also like to thank Casiana Rimbu for the help in providing the most beautiful, clear, and artistic screenshots I could ever imagine and Meta Brown for encouraging me in the first steps of developing the embryonic idea of writing this book.

Many thanks finally go to Heather Fyson for reviewing the book's English.

SAMPLE

Chapter 1. Introduction

1.1. Purpose and structure of this book

We live in the age of data! Every purchase we make is dutifully recorded; every money transaction is carefully registered; every web click ends up in a web click archive. Nowadays everything carries an RFID chip and can record data. We have data available like never before. What can we do with all these data? Can we make some sense out of it? Can we use it to learn something useful and profitable? We need a tool, a surgical knife that can empower us to cut deeper and deeper into our data, to look at it from many different perspectives, to represent its underlying structure.

Let's suppose then that we have this huge amount of data already available, waiting to be dissected. What are the options for a professional to enter the world of Business Intelligence (BI) and Data Science (DS)? The options available are of course multiple and growing rapidly. If our professional does not control an excessive budget, he or she could turn to the world of open-source software. Open-source software, however, is more than a money driven choice. In many cases it represents a software philosophy for resource sharing and control that many professionals support.

Inside the open-source software world, we can find a few Data Science and BI tools. KNIME Analytics Platform represents an easy choice for the non-initiated professional. It does not require learning a specific script and it offers a Graphical User Interface (GUI) to implement and document analysis procedures. In addition - and this is not a secondary advantage - KNIME Analytics Platform can work as an integration platform into which many other BI and Data Science tools can be plugged. It is then not only possible but even easy to analyze data with KNIME Analytics Platform and then to build dashboards on the same processed data with a different BI tool.

Even though KNIME Analytics Platform is very simple and intuitive to use, any beginner would profit from an accelerated orientation through all of the nodes, categories, and settings. This book represents the beginner's luck, because it is aimed to help any beginner to gear up his/her learning process. This book is not meant to be an exhaustive guide to the whole KNIME software. It does not cover implementations under the KNIME Server, which is not open-source, or topics which are considered advanced. Flow Variables, for example, and implementations of database SQL queries are discussed in the sequel book "KNIME Advanced Luck".

This book is divided into six chapters. The first chapter covers the basic concepts of KNIME Analytics Platform, while chapter two takes the reader by the hand into the implementation of the very first KNIME application. From the third chapter, we start the exploration of data science concepts in a more in-depth manner. The third chapter indeed explains how to perform some basic data exploration and visualization, in terms of nodes and processing flow. Chapter four is dedicated to data modeling. It covers a few demonstrative approaches to machine learning, Naïve Bayes, decision trees, and artificial neural networks. Finally, chapters five and six are dedicated to reporting. Usually, the results of an investigation based on data visualization

SAMPLE

or, in a later phase, on data modeling must be shown at some point to colleagues, management, directors, customers, or external workers. Thus, reporting is a very important phase at the end of the data analysis process. Chapter five shows how to prepare the data to export into a report, while chapter six shows how to build the report itself.

Each chapter guides the reader through an ETL or a machine learning (ML) process step by step. Each step is explained in detail and offers some explanations about alternative employments of the current nodes. At the end of each chapter several exercises are proposed to the reader to test and perfect what he/she has learned so far.

Examples and exercises in this book have been implemented using KNIME 4.3. They should also work under subsequent KNIME versions, although there might be slight differences in their appearance.

1.2. KNIME community

Being an open-source software, KNIME Analytics Platform benefits of a number of forums and groups of KNIME users all around the world. This is a good safety net for advises, hints, and learning material. We report below the most popular sites and groups.

Useful Web Pages	
http://www.knime.com	The **root page in the KNIME web site**.
https://www.knime.com/software-overview	The first place to look for an overview of all **KNIME products**. The open source KNIME Analytics Platform can be downloaded here.
https://www.knime.com/knime-self-paced-courses	This is the root page of the **KNIME self-paced courses** (e-learning) where you can learn more about the specific KNIME functionalities. It covers the whole data science creation cycle from data access and data exploration to machine learning and control structures.
https://hub.knime.com/	The **KNIME Hub** is a public repository for KNIME material: workflows, examples, extensions, component templates, and nodes. If you want to start with practical examples, here you can find many according to the search term you type in.
https://forum.knime.com/	What I find particularly useful is the **KNIME Forum**. Here you can ask questions about how to use KNIME Analytics Platform or about how to extend it with new nodes. Someone from the KNIME community answers always and quickly.

SAMPLE

Courses, Events, and Videos

Courses for KNIME Analytics Platform	KNIME periodically offers onsite and online courses for KNIME software. This includes basic and advanced elements. To check for the next available date and to register, just go to the KNIME Events web page https://www.knime.com/learning/events and select the tab Online Courses.
KNIME Webinars	A number of webinars are also frequently available on specific topics, like chemistry nodes, text mining, integration with other analytics tools, automated machine learning, deep learning, time series analysis, best practices, and so on. To know about the next scheduled webinars, check the KNIME Events web page at https://www.knime.com/learning/events and select the tab Webinars.
KNIME Meetups and KNIME Summits	KNIME Meetups and KNIME Data Talks are held periodically all over the world. These are always good chances to learn more about KNIME software, to get inspired about new data science projects, and to get to know other people from the KNIME Community (https://www.knime.com/learning/events)
KNIME TV Channel on YouTube	KNIME has its own video channel on YouTube, named KNIMETV. There, a number of videos are available to learn more about many different topics and specially to get updated about the new features in the new KNIME releases (http://www.youtube.com/user/KNIMETV)

Books

Advanced Features in KNIME Analytics Platform	For the advanced use: Rosaria Silipo, Jeanette Prinz, "KNIME Advanced Luck" (https://www.knime.com/knimepress/advanced-luck)
Reporting Suite	The KNIME Reporting Suite is based on BIRT, another open source tool for reporting. Here is a basic guide on how to use BIRT: *D. Peh, N. Hague, J. Tatchell, "BIRT. A field Guide to Reporting.", Addison-Wesley, 2008*
Data Science and KNIME	For an overview of data science, data mining, and data analytics, please check: *Berthold M.R., Borgelt C., Höppner F., Klawonn F., Silipo R., "Guide to intelligent data science", Springer 2020.*

SAMPLE

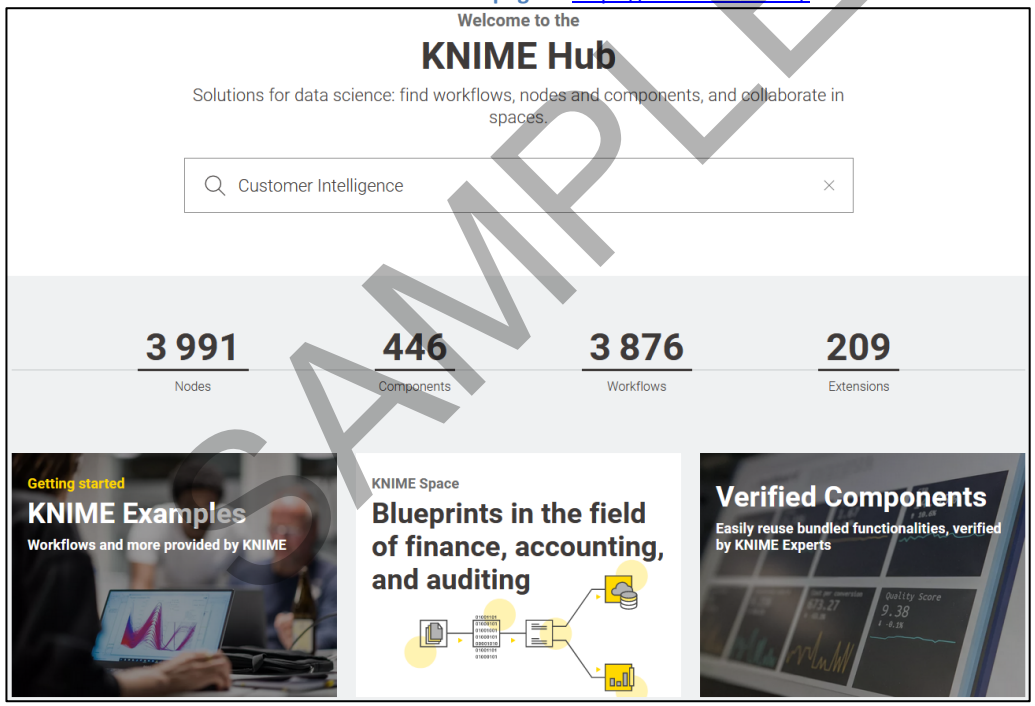
KNIME Hub

However, there is a privileged place where to find information about KNIME nodes and example workflows for your next projects: the KNIME Hub (https://hub.knime.com/).

The KNIME Hub is a repository of applications, components, and nodes to recycle, reuse, and assemble on KNIME Analytics Platform. Or as it says on the home page: The KNIME Hub is "the place to find and collaborate on KNIME workflows and nodes. Here you can find solutions for your data science questions."

When you access the KNIME Hub the first time, you end up with the page in figure 1.1. This page offers a few links to the starting guide documentation, the KNIME Forum, and the KNIME blog. Most importantly at the top it offers a search box to find applications, nodes, and components uploaded by KNIME users in this shared place of the KNIME community.

1.1. The KNIME Hub home page at https://hub.knime.com/

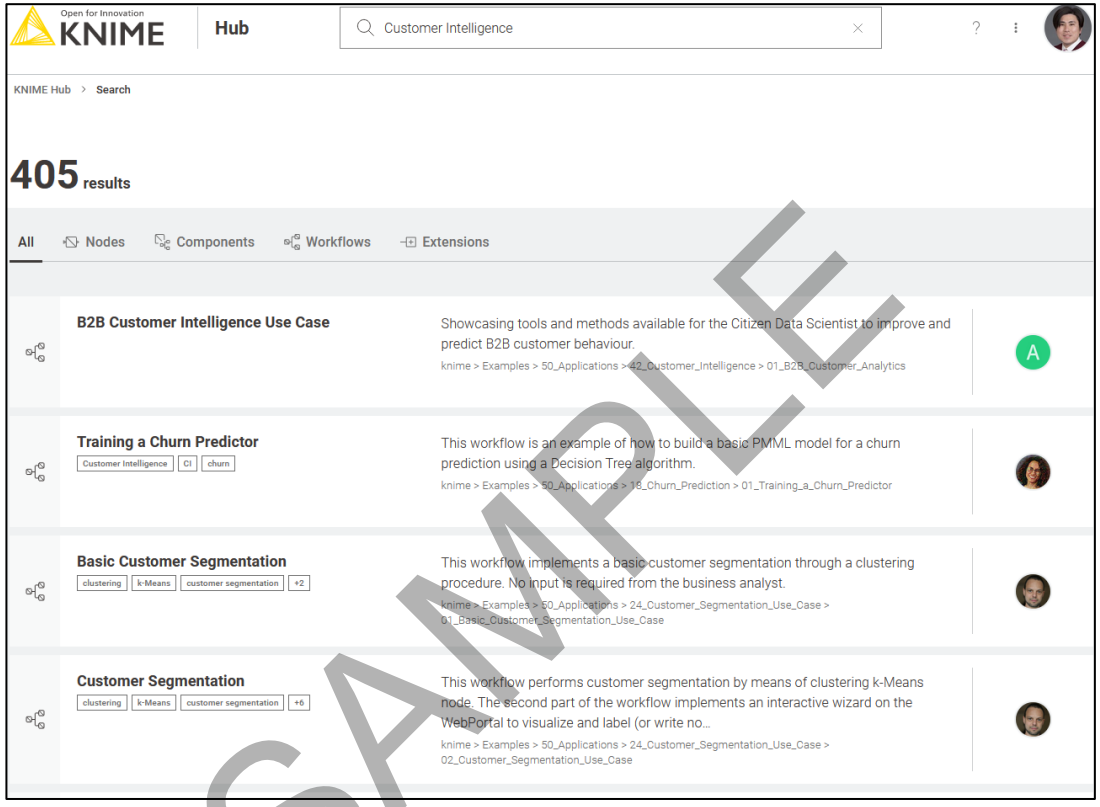


SAMPLE

If we type "Customer Intelligence" in the search box, we end up with a list of nodes and workflows related with Customer Intelligence. Let's select just "Workflows" in the top menu. Then below in figure 1.2 you can see the list of applications (workflows) implementing some aspects of Customer

Intelligence - and appropriately tagged - as uploaded by users of the KNIME community. Indeed, you can upload your own developed applications on the KNIME Hub. All you need is an account with the KNIME Forum.

1.2. The list of applications (workflows) related (and tagged) with Customer Intelligence and available on the KNIME Hub



SAMPLE

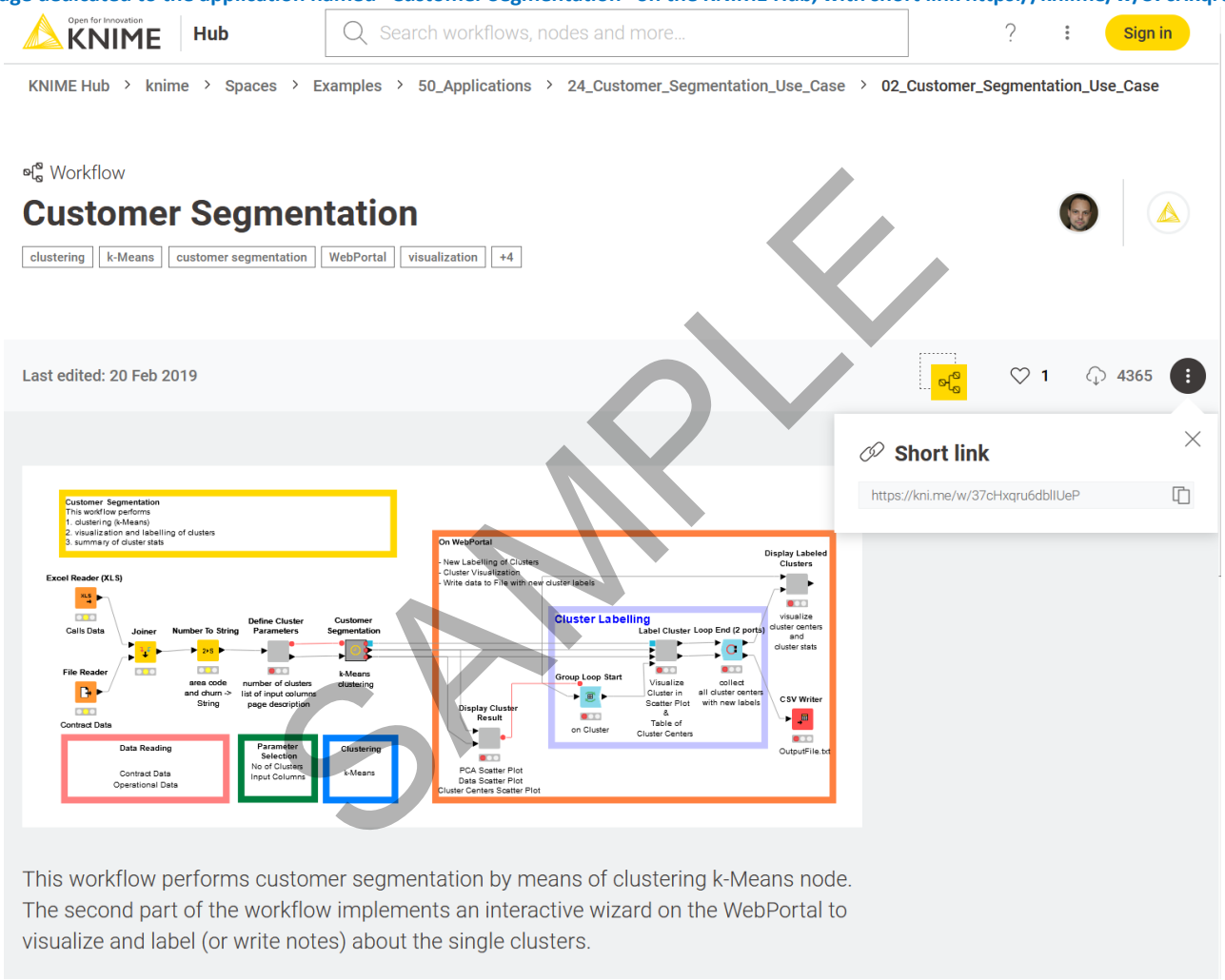
Clicking one of the applications in the list opens the corresponding web page (Fig. 1.3), with a nice explanatory picture of the implemented workflow.

On the top right corner, you can see the button to log in with the KNIME Forum account. Being logged in allows you to upload, download, comment, like, and update the spaces for which you have permissions. Below that, you can find the author picture and below that a number of utility buttons: to download the workflow, to like it, to drag & drop it onto your KNIME Analytics Platform installation, and the short permanent link to this workflow to share.

If you hover over the author image, and if you have editing permissions for this Hub space, a pen will appear. Clicking on it will allow you to allow other KNIME users to upload and change this space.

Notice that the KNIME Hub is a repository for workflows, but also for nodes, components, and extensions.

1.3. The page dedicated to the application named "Customer Segmentation" on the KNIME Hub, with short link https://kni.me/w/37cHxqru6dbllUeP



SAMPLE

1.3. Download and install KNIME Analytics Platform

There are two available KNIME products:

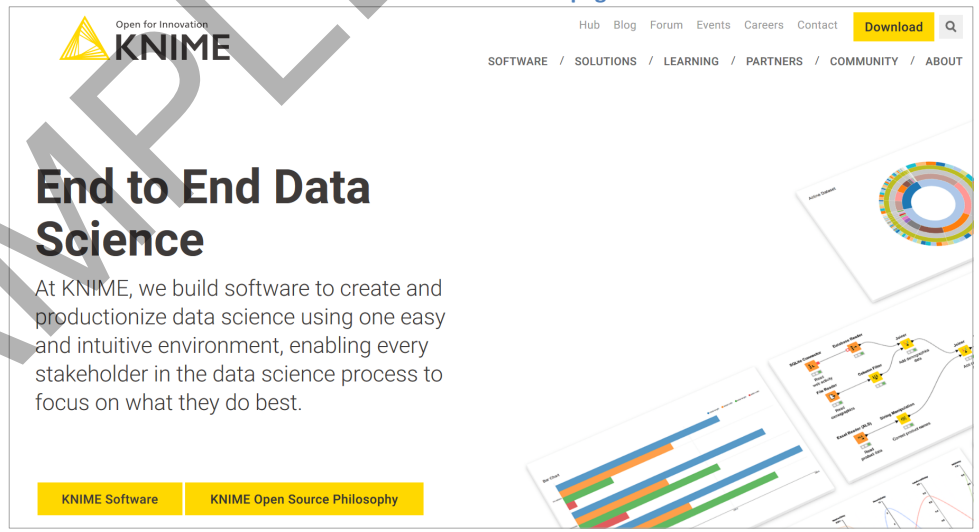
- the open source KNIME Analytics Platform, which can be downloaded free of charge at https://www.knime.com/software-overview under the GPL version 3 license
- the KNIME server, which is described at https://www.knime.com/knime-server

Analytically speaking, the functionalities of the two products are the same. The KNIME Server in addition includes a number of useful IT features for team collaboration, enterprise workflow deployment and management, data warehousing, integration, and scalability for the data science lab. In this book, however, we will work with KNIME Analytics Platform (open-source). To start playing with KNIME Analytics Platform, first, you need to download it to your machine.

Download KNIME Analytics Platform

- Go to www.knime.com
- In the upper right corner of the main page, click "Download"
- Provide a little information about yourself (that is appreciated), then proceed to step 2 "Download KNIME"
- Choose the version that suits your environment (Windows/Mac/Linux, 32 bit/64 bit, with or without Installer for Windows) optionally including all free extensions
- Accept the terms and conditions
- Start downloading. You will end up with a zipped (*.zip), a self-extracting archive file (*.exe), or an Installer application
- For .zip and .exe files, just unpack it in the destination folder. If you selected the installer version, just run it and follow the installer instructions.

1.4. The KNIME web page



SAMPLE

1.4. Workspace

To start KNIME Analytics Platform, open the folder where KNIME has been installed and run knime.exe (or knime on a Linux/Mac machine). If you have installed KNIME using the Installer, then you can just click the icon on your desktop or on your Windows main menu.

If you are starting KNIME Analytics Platform for the first time, you will be asked if you want to share your usage statistics with KNIME. These numbers will be used to fuel the best practice recommendation engine provided within KNIME Analytics Platform workbench: the Workflow Coach. No personal information will ever reach KNIME and your anonymous statistics will never be shared with anybody.

After the splash screen, the "Workspace Launcher" window requires you to enter the path of the workspace.

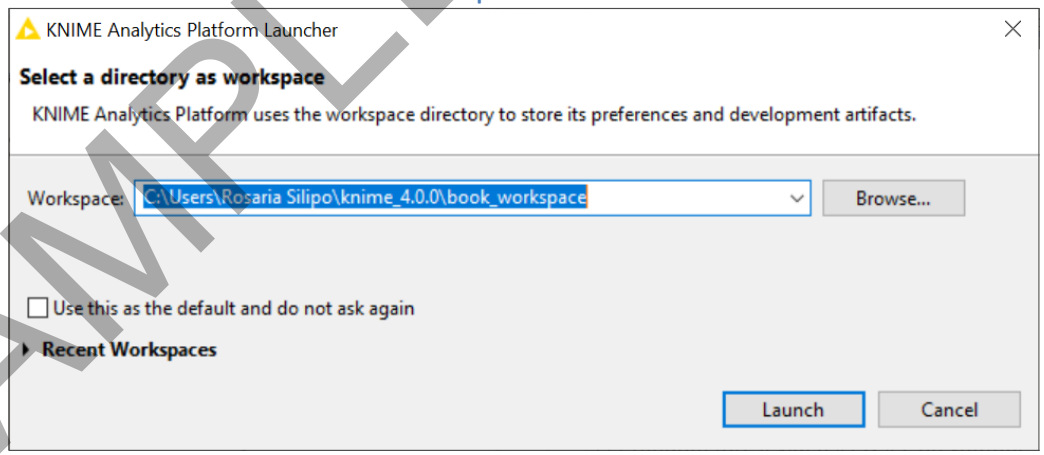
The "Workspace Launcher"

The **workspace** is a folder where all preferences and applications (workflows), both developed and currently under development, are saved for the next KNIME session.

The workspace folder can be located anywhere on the hard-disk.

By default, the workspace folder is "..\knime-workspace". However, you can easily change that, by changing the path proposed in the "Workspace Launcher" window, before starting the KNIME working session.

1.5. The "Workspace Launcher" window



SAMPLE

Once KNIME Analytics Platform has been opened, from within the KNIME workbench you can switch to another workspace folder, by selecting "File" in the top menu and then "Switch Workspace". After selecting the new workspace, KNIME Analytics Platform restarts, showing the workflow list from the newly selected workspace. Notice that if the workspace folder does not exist, it will be automatically created.

If I have a large number of customers for example, I can use a different workspace for each one of them. This keeps my work space clean and tidy and protects me from mixing up information by mistake. For this project I used the workspace "KNIME_4.3.1\book_workspace".

1.5. KNIME workflow

KNIME Analytics Platform does not work with scripts, it works with graphical workflows.

Small little icons, called nodes, are dedicated each to implement and execute a given task. A sequence of nodes makes a workflow to process the data to reach the desired result.

What is a workflow

A workflow is an **analysis flow**, i.e. the **sequence of analysis steps** necessary to reach a given result. It is the pipeline of the analysis process, something like:

Step 1. Read data
Step 2. Clean data
Step 3. Filter data
Step 4. Train a model

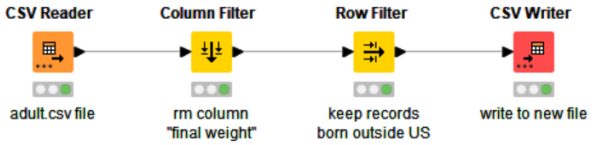
KNIME Analytics Platform implements its workflows **graphically**. Each step of the data analysis is implemented and executed through a little box, called **node**. A sequence of nodes makes a workflow.

In the KNIME whitepaper [1] a workflow is defined as follows: *"Workflows in KNIME are graphs connecting nodes, or more formally, direct acyclic graphs (DAG)."* (http://www.kdd2006.com/docs/KDD06_Demo_13_Knime.pdf)

Below is an example of a KNIME workflow, with:

- a node to read data from a file
- a node to exclude some data columns
- a node to filter out some data rows
- a node to write the processed data into a file

1.6. Example of a KNIME workflow



SAMPLE

Note. A workflow is a data analysis sequence, which in a traditional programming language would be implemented by a series of instructions and calls to functions. KNIME Analytics Platform implements it graphically. This graphical representation is more intuitive to use, lets you keep an overview of the analysis process, and makes for the documentation as well.

What is a node

A node is the **single processing unit** of a workflow.

A node takes a data set as input, processes it, and makes it available at its output port. The "processing" action of a node ranges from modeling - like an Artificial Neural Network Learner node - to data manipulation - like transposing the input data matrix - from graphical tools - like a scatter plot, to reading/writing operations.

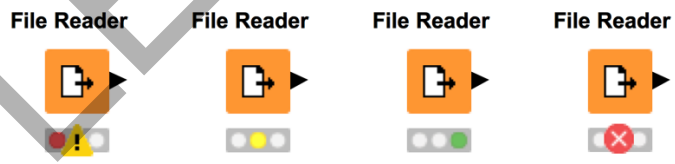
Every node in KNIME has 4 states:

- Inactive and not yet configured → **red** light
- Configured but not yet executed → **yellow** light
- Executed successfully → **green** light
- Executed with errors → **red with cross** light

Nodes containing other nodes are called **metanodes** or **components**.

Below are four examples of the same node (a File Reader node) in each one of the four states.

1.7. File Reader node with different states

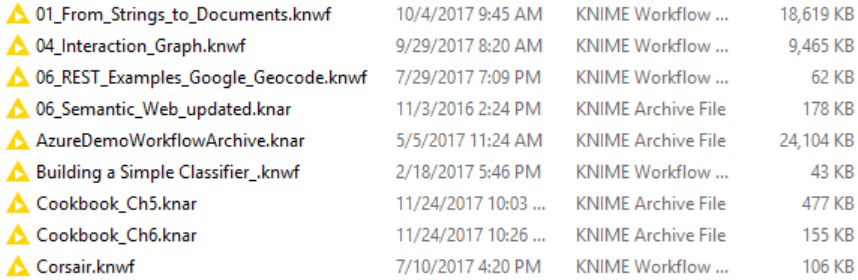


SAMPLE

1.6. *.knwf* and *.knar* file extensions

KNIME workflows can be packaged and exported in *.knwf* or *.knar* files. A *.knwf* file contains only one workflow, while a *.knar* file contains a group of workflows. Such extensions are associated with KNIME Analytics Platform. A double-click opens the workflow inside KNIME Analytics Platform.

1.8. *.knwf* and *.knar* files are associated with KNIME Analytics Platform. A double-click opens the workflow(s) directly inside the platform

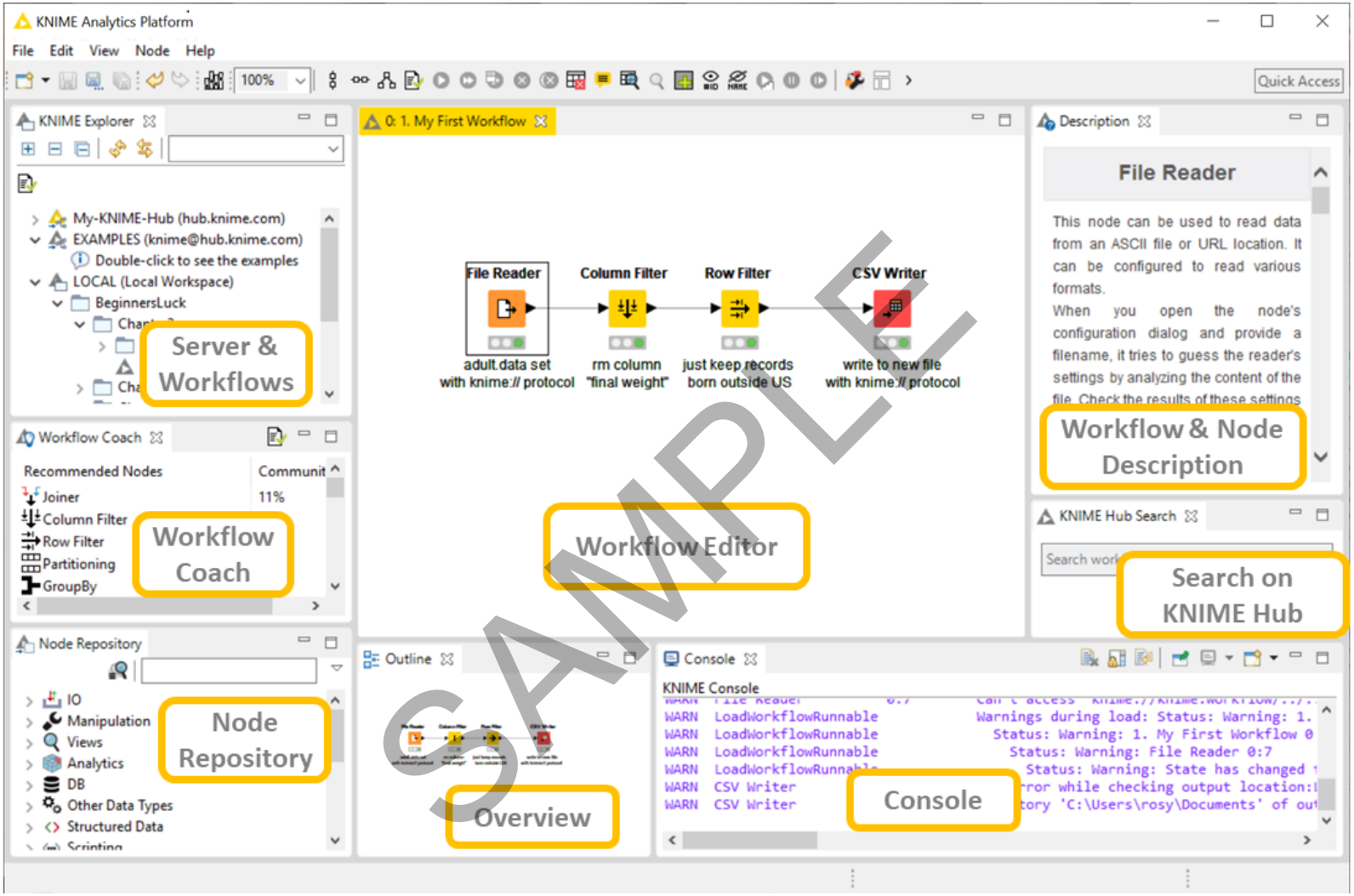


1.7. KNIME workbench

After accepting the workspace path, the KNIME workbench opens on a “Welcome to KNIME” page. This page provides a few links to get started, such as for example to the KNIME Hub, to some basic documentation, to the current courses and events, to available updates, and so on. The “KNIME Workbench” consists of a top menu, a tool bar, and a few panels. Panels can be closed, re-opened, and moved around.

SAMPLE

1.9. The KNIME workbench



SAMPLE

The KNIME Workbench

Top Menu: File, Edit, View, Node, Help

Tool Bar: New, Save (Save As, Save All), Undo/Redo, Open Report (if reporting was installed), zoom (in %), Align selected nodes vertically/horizontally, Auto layout, Configure, Execute options, Cancel execution options, Reset, Edit node name and description, Open node's first out port table, Open node's first view, Open the "Add Meta node" Wizard, Append IDs to node names, Hide all node names, Loop execution options, Change Workflow Editor Settings, Edit Layout in Components, configure job manager.

KNIME Explorer

This panel shows the list of workflow projects available in the selected workspace (LOCAL), on the EXAMPLES server, on the My-KNIME-Hub (your own space on the KNIME Hub), or on other connected KNIME servers.

Workflow Coach

This is a node recommendation engine. It will provide the list of the top most likely nodes to follow the currently selected node.

Workflow Editor

The central area consists of the "Workflow Editor" itself.

A node can be selected from the "Node Repository" panel and dragged and dropped here, in the "Workflow Editor" panel.

Nodes can be connected by clicking the output port of one node and releasing the mouse either at the input port of the next node or at the next node itself.

Node Description

If a node or a workflow is selected, this panel displays a summary description of the node's functionalities or the workflow's meta information.

Search box for KNIME Hub

To search for material on the KNIME Hub

Node Repository

This panel contains all the nodes that are available in your KNIME installation. It is something similar to a palette of tools when working in a report or with a web designer software. There we use graphical tools, while in KNIME we use data analytics tools.

Outline

The "Outline" panel contains a small overview of the contents of the "Workflow Editor". The "Outline" panel might not be of so much interest for small workflows. However, as soon as the workflows reach a considerable size, all the workflow's nodes may no longer be visible in the "Workflow Editor" without scrolling. The "Outline" panel, for example, can help you locate newly created nodes.

Console

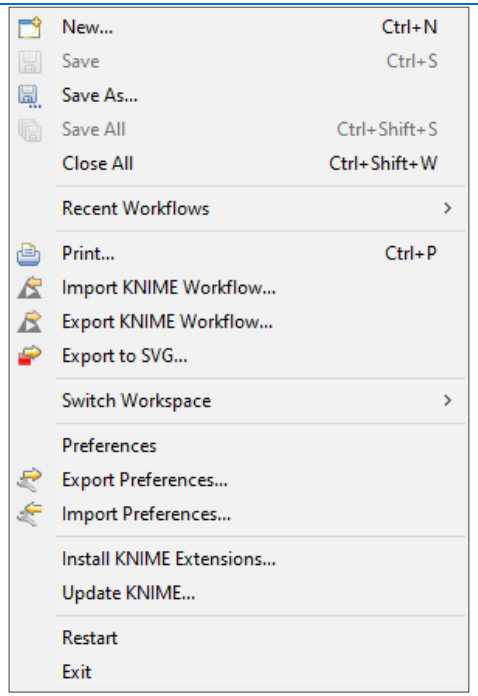
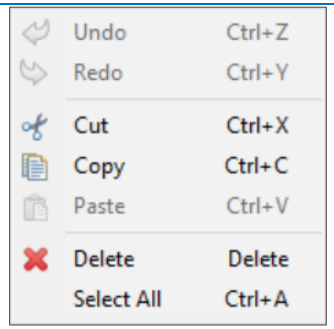
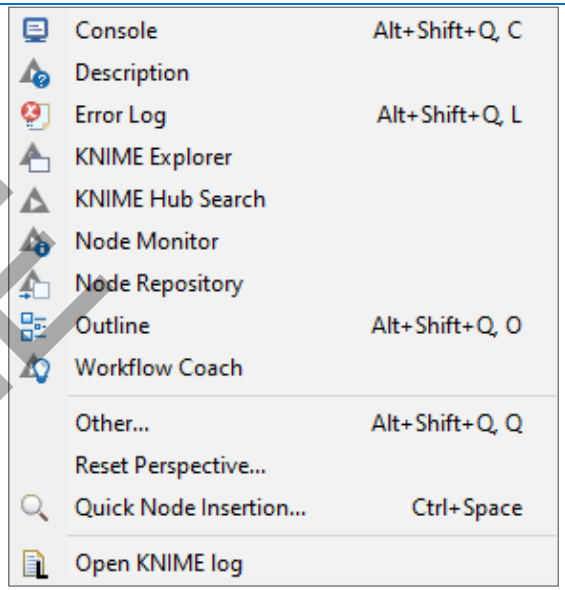
The "Console" panel displays error and warning messages to the user.

This panel also shows the location of the log file, which might be of interest when the console does not show all messages.

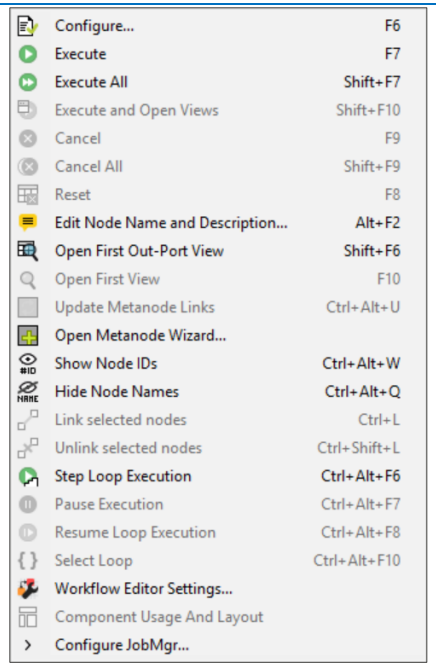
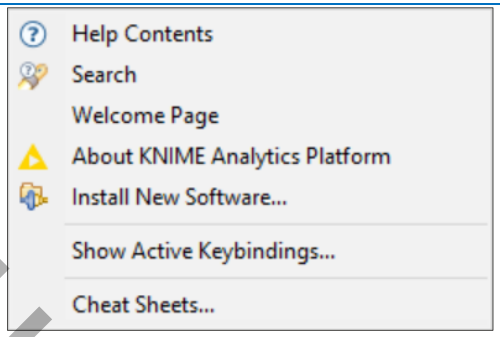
There is a button in the tool bar as well to show the log file associated with this KNIME instance.

SAMPLE

Top menu

File	Edit	View
		
File includes the traditional File commands, like "New" and "Save", in addition to some KNIME specific commands, like: - Import/Export KNIME workflow… - Export to SVG - Switch Workspace - Preferences with Export/Import Preferences - Install KNIME Extensions - Update KNIME	**Edit** contains edit commands. **Undo** and **Redo** refer to the last performed actions. **Cut, Copy, Paste,** and **Delete** refer to selected nodes in the workflow. **Select All** selects all the nodes of the workflow in the workflow editor.	**View** contains the list of all panels that can be opened in the KNIME workbench. A closed panel can be re-opened here. Also, when the panel disposition is messed up, the option "Reset Perspective" re-creates the original panel layout when the workbench was started for the first time. Option "Other" opens additional views useful to customize the workbench.

SAMPLE

Node	Help
	
Node refers to all possible operations that can be performed on a node. A node can be: - Configured - Executed - Cancelled (stopped during execution) - Reset (resets the results of the last “Execute” operation) - Given a name and description - Set to show its View (if any) Options are only active if they are possible. For example, an already successfully executed node cannot be re-executed unless it is first reset or its configuration has been changed. The “Cancel” and “Execute” options are then inactive. Option “Open Meta Node Wizard” starts the wizard to create a new meta node in the workflow editor.	**Help Contents** provides general Help about the Workbench, BIRT, and KNIME. **Search** opens a panel on the right of the “Node Description” panel to search for specific Help topics or nodes. **Welcome Page** (re-)opens the Welcome Page **Install New Software** is the door to install KNIME Extensions from the KNIME Update sites. **Show Active Keybindings** summarizes all keyboard commands for the workflow editor. **Cheat Sheets** offer tutorials on specific topics: the reporting tool, csv, Plug-ins.

SAMPLE

Let's now go through the most frequently used items in the Top Menu.

"File" → "Import KNIME workflow" reads and copies workflows into the current workspace.

Option "Select root directory" copies the workflow directly from a folder into the current workspace (LOCAL).
Option "Select archive file" reads a workflow from a .knwf or .knar file into the current workspace (LOCAL). .knwf /.knar files can be created through the option "File"→ "Export KNIME workflow".

"File" → "Export KNIME workflow" exports the one selected workflow to a .knwf or the many selected workflows to a .knar file.

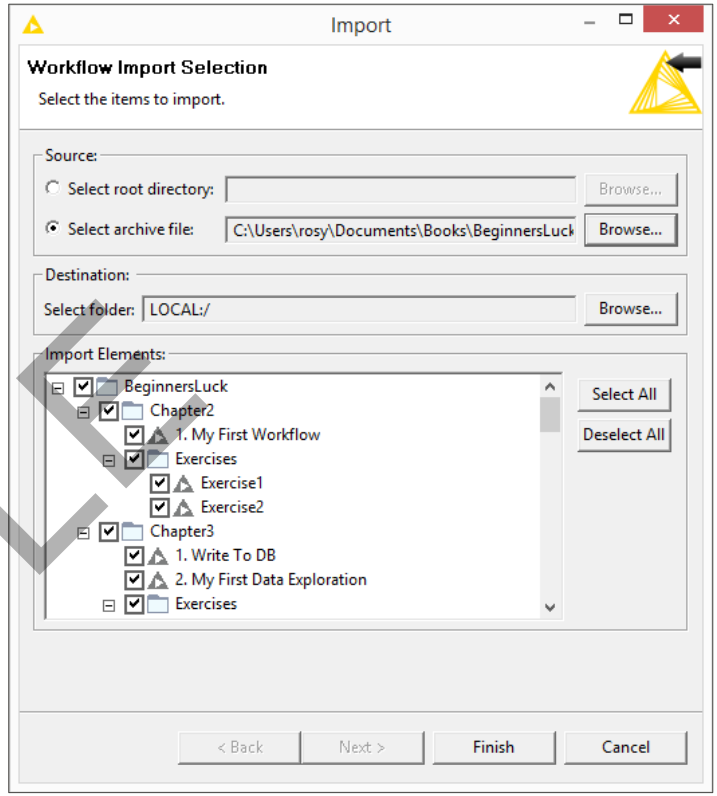
Option "Reset Workflow(s) before export" exports fully reset workflows without the data produced by each node. This generates considerably smaller export files.

Simply copying a workflow from one folder to another can create a number of problems related to internal KNIME updates. Copying workflows by using the option "Import KNIME workflow" or by double-click is definitely safer.

"File" → "Install KNIME Extensions" and "**Help" → "Install New Software"** both link to the dialog window for the installation of KNIME Extensions from the KNIME Update sites (see next sections).

"File" → "Switch Workspace" changes the current workspace with a new one.

1.10. Window "Import" to import workflows

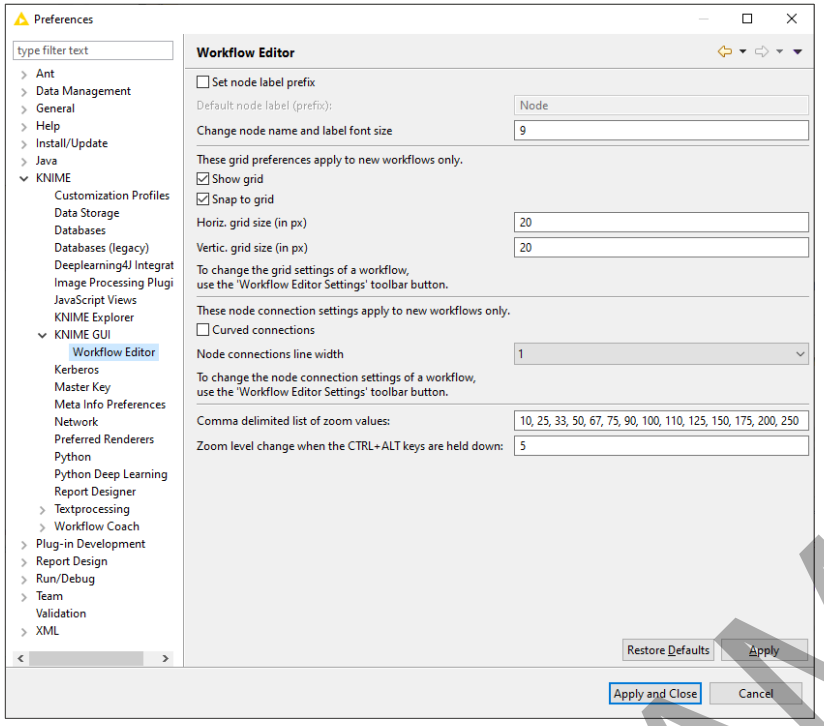


"File" → "Preferences" brings you to the window where all KNIME settings can be customized. They can be found under item "KNIME". Let's check them.

- **Chemistry** has settings related to the KNIME Renderers in the chemistry packages.
- **Databases** specifies the location of specific database drivers, not already available within KNIME. Indeed, the most common and most recent database drivers are already available in the driver menu of Database nodes. However, if you need some specific driver file, you can set its path here.

SAMPLE

1.11. The "Preferences" window



- **KNIME Explorer** contains the list of the shared repositories via KNIME Server.
- **KNIME GUI** allows the customization of the KNIME workbench options and layout via a number of settings.
- **Master Key** contains the master key to be used in nodes with an encryption option, like database connection nodes. Since KNIME 2.3 database passwords are passed via the “Credentials” workflow variables and the Master Key preference has been deprecated. You can still find it in the Preferences menu for backward compatibility.
- In **Meta Info Preferences** you can upload meta-info template for nodes and workflows.
- Here you can also find the preference settings for the **external packages**, like: H2O, R, Report Designer, Perl, Perl, Open Street Map, and others if you have them installed. In particular, for the external scripts, this page offers the option to set the path to the reference script installation.
- Finally, **Workflow Coach** contains the dataset to be used for the node recommendation engine: the community, a server workspace, or your own local workspace.

Export Preferences and **Import Preferences** in the “File” menu respectively exports and imports the “Preferences” settings into and from a *.epf file. These two commands come in handy when, for example, a new version of KNIME is installed and we want to import the old preferences settings.

Tool Bar

The tool bar is another important piece of the KNIME workbench.

From the left, we find the icon to create a new workflow, save the selected workflow, save as the selected workflow in another location, save all open workflows, undo and redo, switch to the reporting environment, zoom (in %), align selected nodes vertically, align selected nodes horizontally, auto-layout, configure the selected node, execute the selected node, execute all executable nodes, execute selected nodes and open the first data view, cancel selected running nodes, cancel all running nodes, reset selected nodes, edit description of selected node, open first data view of selected nodes, open views of selected nodes, open the Add Metanode Wizard, append IDs to node names, hide node names,

SAMPLE

do one loop step, pause loop execution, resume loop execution, change workflow editor settings, open layout editor for components, configure job manager for all selected nodes. We will see all these options along the course of this book.

For now, I just want to describe the "**Auto Layout**" button. The auto-layout button automatically adjusts the position of the nodes in the workflow to produce a clean, ordered, and easy to explore workflow. This auto-layout operation becomes particularly useful when, for example after a long development session, the workflow overview has become difficult.

1.12. The "Auto Layout" button in the tool bar

For all keyboard lovers, most KNIME commands can also run via **hotkeys**. All hotkeys are listed in the KNIME menus on the side of the corresponding commands or in the tooltip messages of the icons in the Tool Bar under the Top Menu. Here are the most frequently used hotkeys.

Hotkeys

Node Configuration

- **F6** opens the configuration window of the selected node

Node Execution

- **F7** executes selected configured nodes
- **Shift + F7** executes all configured nodes
- **Shift + F10** executes all configured nodes and opens all views

Stop Node Execution

- **F9** cancels selected running nodes
- **Shift + F9** cancels all running nodes

To move nodes

- **Ctrl + Shift + Arrow** moves the selected node in the arrow direction

Node Resetting

- **F8** resets selected nodes

Save Workflows

- **Ctrl + S** saves the workflow
- **Ctrl + Shift + S** saves all open workflows
- **Ctrl + Shift + W** closes all open workflows

Meta-Node

- **Shift + F12** opens Meta Node Wizard

To move Annotations

- **Ctrl + Shift + PgUp/PgDown** moves the selected annotation in the front or in the back of all the overlapping annotations

SAMPLE

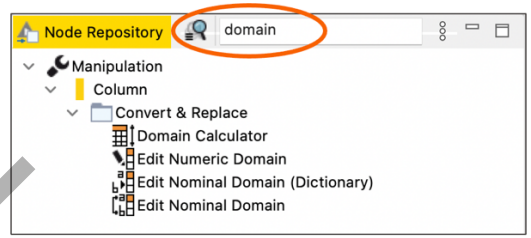
Node Repository

In the lower left corner we find the Node Repository, containing all installed nodes organized in categories and subcategories. KNIME Analytics Platform has accumulated by now more than 1500 nodes. It has become hard to remember the location of each node in the Node Repository. To solve this problem, two search options are available: by exact match and by fuzzy match, both in the search box placed at the top of the Node Repository panel.

Search box

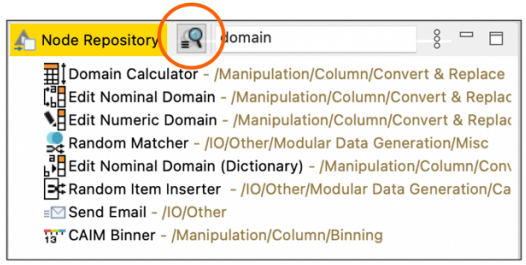
At the top of the "Node Repository" panel there is a **search box**. If you type a keyword in the search box and hit "Enter", you obtain the list of nodes containing an exact match of that keyword. Press the "Esc" key to see all nodes again.

1.13. Word search in the Node Repository panel: exact match mode



Clicking the lens on the left of the search box runs a fuzzy search algorithm leading to a wider matching result list than what found in the previous figure.

1.14. Word Search in the Node Repository panel: fuzzy search mode



SAMPLE

KNIME Explorer

In the top left corner of the KNIME workbench, we find the KNIME Explorer panel. This panel contains:

- Under LOCAL the workflows that have been developed in the selected workspace
- The mount points to a number of KNIME Servers
- The workflows contained in the reference workspace of such servers
- The access to the My-KNIME-Hub, that is to your space on the KNIME Hub. Remember that you need an account with the KNIME Forum to access this space.

At the beginning, the KNIME Explorer panel only contains LOCAL, My-KNIME-Hub, and EXAMPLES. As we already stated, LOCAL shows the content of the selected workspace. EXAMPLES points to a read-only public server, accessible via anonymous login. This server hosts a number of example workflows that you can use to jump start a new project. My-KNIME-Hub allows to access your space on the KNIME Hub.

When you open KNIME Analytics Platform for the first time, you will find a folder named “Example Workflows” containing the solutions to a few common data science use cases, comprehensive of data.

Folders in “KNIME Explorer”, containing workflows, are also called “Workflow Groups”.

Note. KNIME Explorer panel can also host data. Just create a folder under the workspace folder, fill it with data files on the machine, and select “Refresh” in the context-menu (right-click) of the “KNIME Explorer” panel.

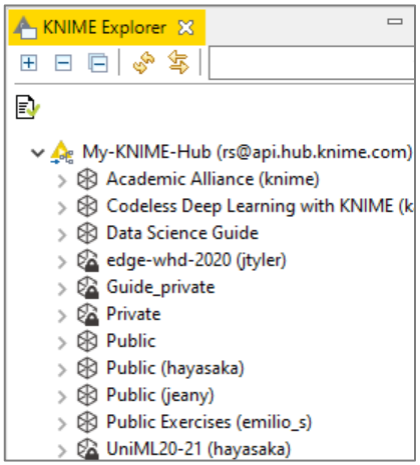
My-KNIME-Hub

From the KNIME Explorer panel, you can access your spaces on the KNIME Hub and upload and update new or existing content in there.

By default an authenticated KNIME user has a public space, for material to share publicly, and a private space to park his/her own material for further usage. However, new private or public spaces can be created with a right-click on My-KNIME-Hub in the KNIME Explorer panel and then a selection of the option “Create new Space…”.

By default, you are the only owner of your own spaces. However, when accessing this space form a web browser, after hovering on your image in the top right corner, a pen appears. This will allow to add colleagues and teammates as contributors to the space.

1.15. KNIME Explorer panel. At the top the content of the EXAMPLES server; below the content of the LOCAL workspace



SAMPLE

EXAMPLES Server

A link to the KNIME Public Server (EXAMPLES) is available in the “KNIME Explorer” panel. This is a server provided by KNIME to all users for tutorials and demos. There you can find a number of useful examples on how to implement specific tasks with KNIME. To connect to the EXAMPLES Server:

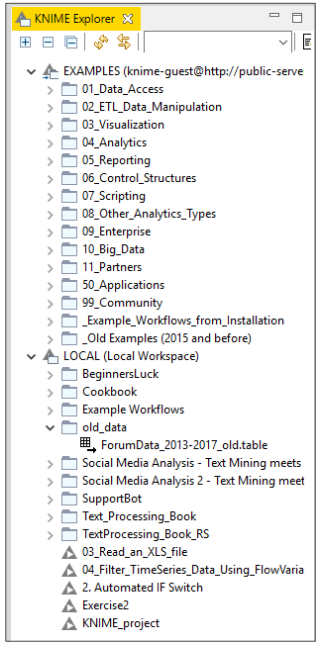
- double click “EXAMPLES” in the “KNIME Explorer” panel
- double click “Double click to connect…”

You should be automatically logged in as a guest.

To transfer example workflows from the EXAMPLES Server to your LOCAL workspace, just drag and drop or copy and paste (Ctrl-C, Ctrl-V in Windows) them from “EXAMPLES” to “LOCAL”.

You can also open the EXAMPLES workflows in the workflow editor, however only temporarily and in read-only mode. A yellow warning box on top warns that this workflow copy will not be saved.

1.16. KNIME Explorer panel. At the top the content of the EXAMPLES server; below the content of the LOCAL workspace



The KNIME Explorer panel can of course host more than one KNIME Server. It is enough to add server mount points to the list of the available KNIME servers.

SAMPLE

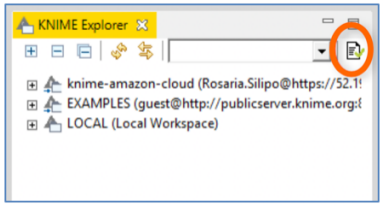
Mounting Servers in KNIME Explorer

To add KNIME servers to the “KNIME Explorer” panel:

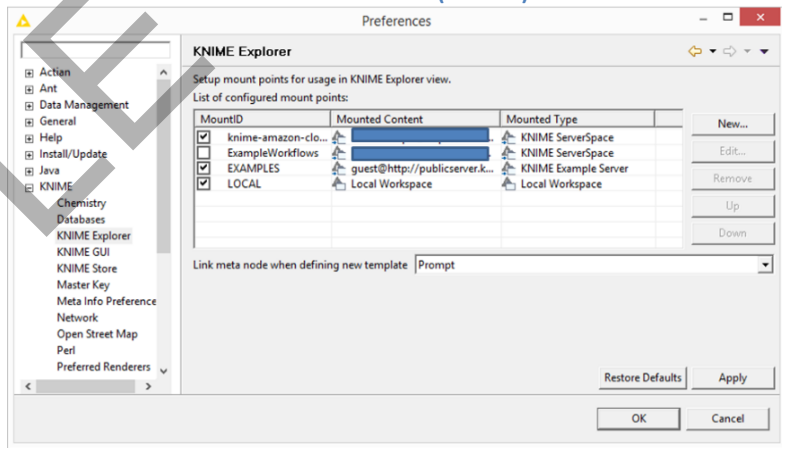
- Select the “KNIME Explorer” panel
- Click the “Configure Explorer View” button
- The “Preferences (Filtered)” window opens on the “KNIME Explorer” page and lists all KNIME spaces already mounted in this KNIME instance. The three KNIME spaces available by default on every KNIME instance are the local workspace “LOCAL”, the KNIME public Server “EXAMPLES”, and the My-KNIME-Hub located on the KNIME Hub Server (hub.knime.com).
- Use the “New” and the “Remove” button to add /remove connections to remote servers.
- After clicking the ”New“ button, fill in the required information about the server in the “Select New Content” window (Fig. 1.19)
- Use the “Test Connection” button to automatically retrieve the default mountpoint for the selected server.

The same KNIME Explorer “Preferences” page can be reached via “File” in the top menu → “Preferences” → “KNIME Explorer”.

1.17. The ”Configure KNIME Explorer“ button



1.18. The “Preferences (Filtered)“ window

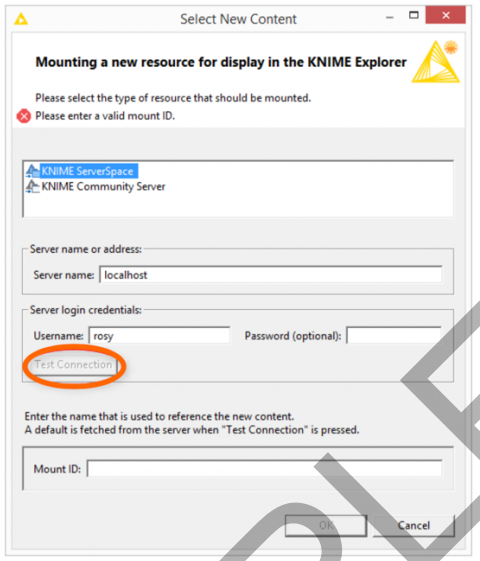


To login into any of the available servers in the “KNIME Explorer” panel:

- right-click or double-click the server name
- provide the credentials

SAMPLE

1.19. The “Select New Content”” window



SAMPLE

Workflow Editor

The central piece of the KNIME workbench consists of the workflow editor itself. This is the place where a workflow is built by adding one node after the other. Nodes are inserted in the workflow editor by drag and drop or double-click from the Node Repository or the Workflow Coach. The workflow building process will be described widely in the next sections of this book. Here, we will describe how to customize and probably improve the canvas role of the workflow editor space. We will describe two options:

- change the canvas appearance with grids and different visualizations for the connections;
- introducing annotations to comment the work.

Adding a grid to the canvas and curved connections to the workflows

Almost towards the end, on the right of the tool bar, you can see the “Change Workflow Editor Settings” button. If you click it, the “Workflow Editor Settings” window opens.

1.19. "Change Workflow Editor Settings" Button in Tool Bar

Customizing the Workflow Editor

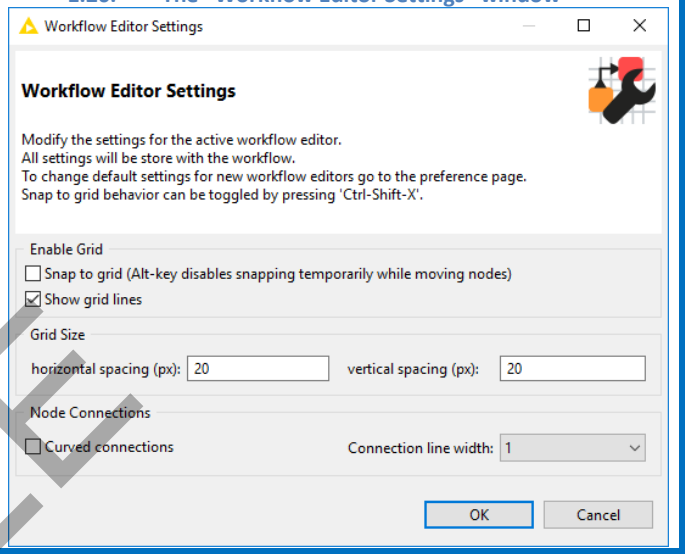
The grid feature contains a few options:

1. “Show grid lines”. This shows grid lines in the workflow editor and allows to better align nodes and annotations manually.

2. “Snap to grid”. This option attaches nodes and annotations to the closest available corner of the grid. It gives you less manual freedom, but the result is cleaner and more ordered in shorter time.

3. “Curved Connections”. Here you can enable node connections to follow a curve rather than a straight line. This might lead to more appealing workflow graphics.

1.20. The “Workflow Editor Settings” window



Adding annotations to the canvas

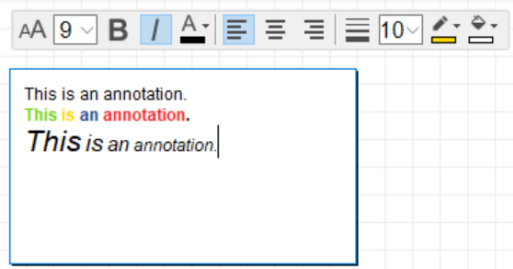
It is also possible to include **annotations** in the workflow editor. Annotations can help to explain the task of the workflow and the function of each node or group of nodes. The result is an improved documentation-like overview of the workflow general task and of the single sub-tasks.

Workflow Annotations

To insert a new annotation:

- right-click anywhere in the workflow editor and select “New Workflow Annotation”
- a pale-yellow small frame appears; this is the default annotation frame
- double-click the frame to edit its content
- Notice the tool bar appearing at the top to edit text style, font color, background color, text alignment, and border properties (color, thinkness).
- To reopen an annotation, just double-click at the top left corner, where the pencil icon is.

1.21. The Annotation Editor



SAMPLE

Other Workbench Customizations

Another possibility for customization consists of adding views. Available views are found in the "View" item in the Top menu.

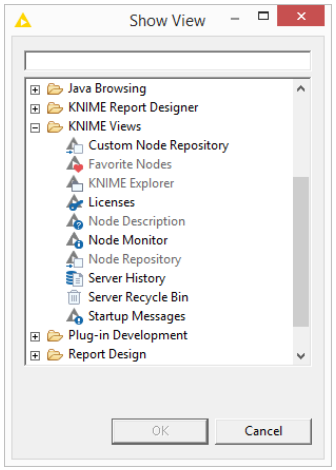
Popular views, for example, are the "Node Monitor", the "Custom Node Repository", and the "Licenses" and "Server" views, if you have a connected server. All these extra views can be found in the Top menu under "View" → "Other" → "KNIME Views".

The "Node Monitor" view helps, especially during the development phase, to monitor and debug the workflow execution.

The "Custom Node Repository allows for a customized "Node Repository" with only a subset of nodes.

"Licenses" allows to monitor your license situation, if you have any.

1.22. Additional Views from "View" → "Other" → "KNIME Views"



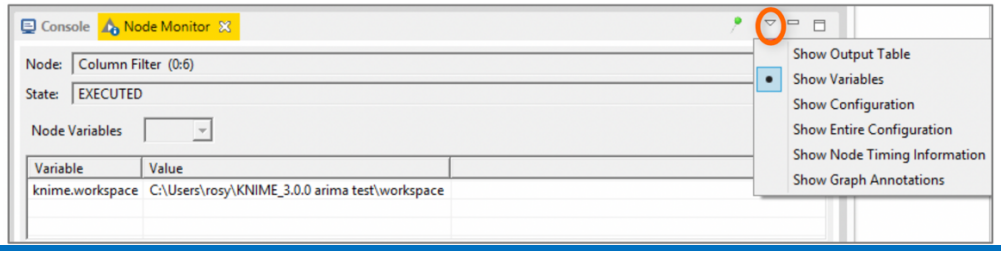
SAMPLE

Node Monitor View

To insert the "Node Monitor" panel in the workbench:

- Select "View"→ "Other…" in the top menu
- In the "Show View" window, expand the "KNIME Views" item and double-click "Node Monitor"; a panel, named "Node Monitor", appears on the side of the "Console" panel; the panel shows the values for the output flow variables, the output data, or the configuration settings of the selected node in the workflow editor.
- There you can decide what to show (data, configuration, variables), via the menu in the top right corner.

1.23. The Node Monitor View



1.8. Download the KNIME Extensions

KNIME Analytics Platform is an open source product. As every open source product, it benefits from the feedback and the functionalities that the community develops. A number of extensions are available for KNIME Analytics Platform. If you have downloaded and installed KNIME Analytics Platform including all its free extensions, you will see the corresponding categories in the Node Repository panel, such as KNIME Labs, Text Processing, R Integration, and many others. However, if at installation time, you have chosen to install the bare KNIME Analytics Platform without the free extensions, you might need to install them separately at some point on a running instance.

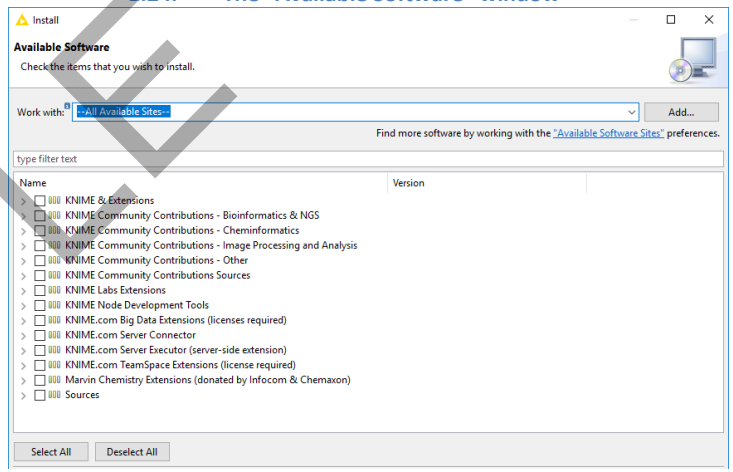
Installing KNIME Extensions

To install a new KNIME extension from within KNIME Analytics Platform, there are three options.

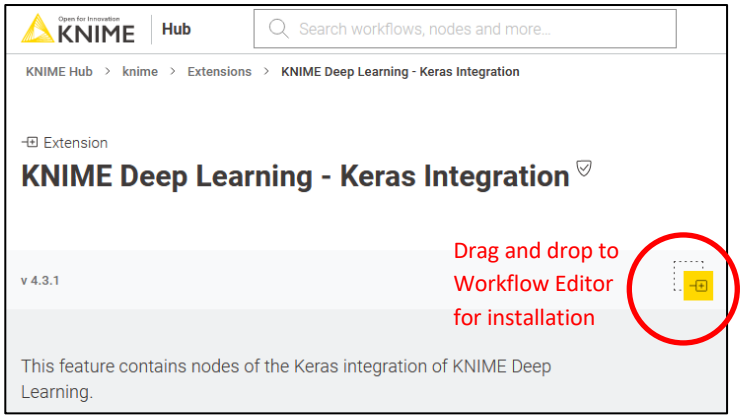
1. From the Top Menu, select "**File**" → "**Install KNIME Extensions**", select the desired extension, click the "**Next**" button and follow the wizard instructions.
2. From the Top Menu, select "**Help**" → "**Install New Software**". In the "Available Software" window, in the "Work with" textbox, select the URL with the KNIME update site (usually named "KNIME Analytics Platform 4.x Update Site" - http://update.knime.com/analytics-platform/4.x). Then select the extension, click the "**Next**" button and follow the wizard instructions.
3. Search the KNIME Hub, on a web browser or from the KNIME Hub panel. When the desired extension is found, drag and drop the extension icon from the browser to the Workflow Editor.

Once the selected KNIME extension(s) has/have been installed and KNIME has been restarted, you should see the new category, corresponding to the installed extension, in the "Node Repository".

1.24. The "Available Software" window



1.25. An extension from the KNIME Hub



SAMPLE

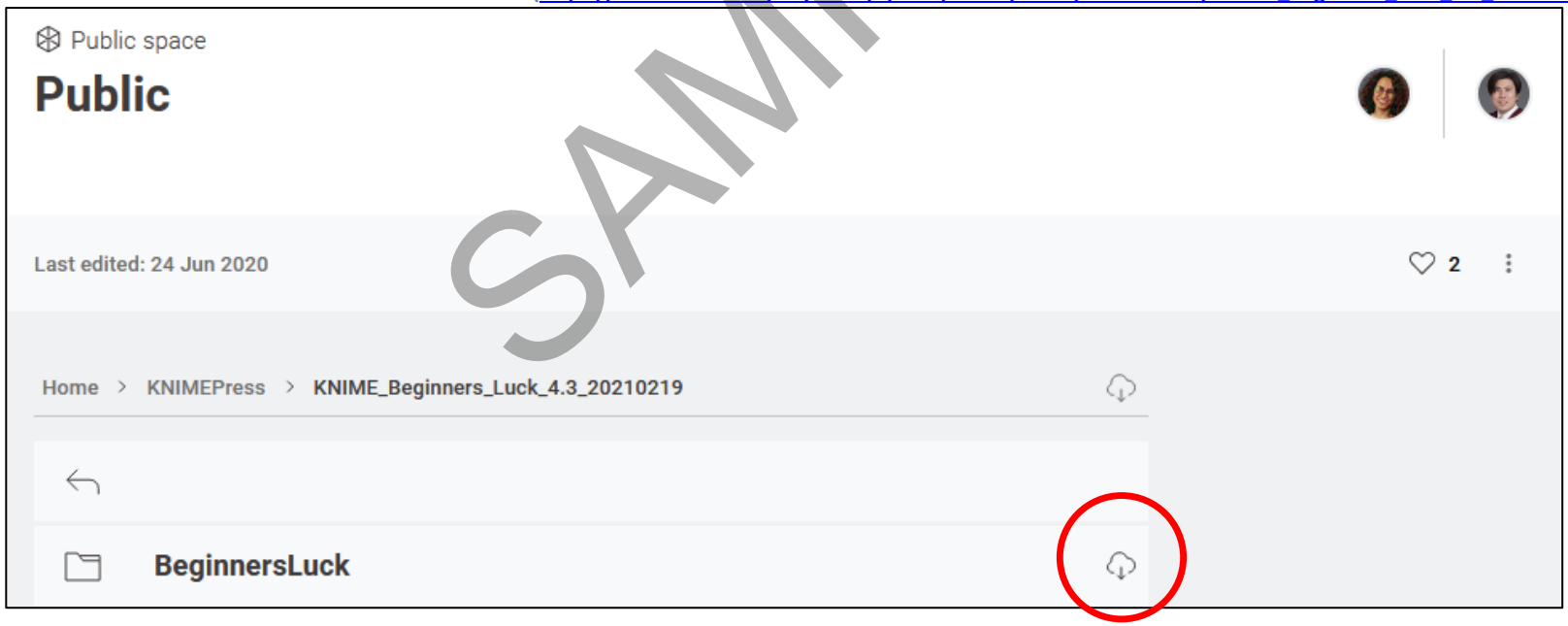
In the “Available Software” window you can find some extension groups: KNIME & Extensions, KNIME Labs Extensions, KNIME Node Development Tools, Sources, and more. “KNIME & Extensions” contains all extensions provided for the current release; “KNIME Labs Extensions” contains a number of extensions ready to use, but not yet of x.1 release quality; “KNIME Node Development Tools” contains packages with some useful tools for Java programmers to develop nodes; “Sources” contains the KNIME source code. Specific packages donated by third parties or community entities might also be available in the list of extensions. These are usually grouped under “Community” categories. My advice is to install all extensions, even the cheminformatics ones. Many of them contain several useful nodes not necessarily restricted to a particular domain.

1.9. Data and workflows for this book

This book builds a few examples and provides the solutions to the exercises. They are contained in folder “BeginnersLuck” downloadable from the KNIME Hub space of one of the authors of this book. To access the KNIME Hub, you need to create an account with the KNIME Forum. After entering the KNIME Hub, in order to download the workflows, just click on the cloud icon.

- Download the whole folder onto your machine, which will result in a .knar file. Then:
- Double click it OR import it into the KNIME Explorer via Select File → Import KNIME Workflow …

1.26. Workflows and data for this book on the KNIME Hub (https://hub.knime.com/hayasaka/spaces/Public/latest/KNIMEPress/KNIME_Beginners_Luck_4.3_20210219/)



SAMPLE

At the end of the import operation, in the KNIME Explorer panel you should find a BeginnersLuck folder containing Chapter2, Chapter3, Chapter4, Chapter5, and Chapter6 subfolders, each one with workflows and exercises to be implemented in the next chapters. You should also find a KBLdata folder containing the required data.

The data used for the exercises and for the demonstrative workflows of this book were either generated by the author or downloaded from the UCI Machine Learning Repository, a public data repository (http://archive.ics.uci.edu/ml/datasets). If the data set belongs to the UCI Repository, a full link is provided here for download. Data generated by the author, that is not public data, are located in the KBLData folder.

Data from the UCI Machine Learning Repository:

- Adult.data: http://archive.ics.uci.edu/ml/datasets/Adult
- Iris data: http://archive.ics.uci.edu/ml/datasets/Iris
- Yellow-small.data (Balloons) http://archive.ics.uci.edu/ml/datasets/Balloons
- Wine data: http://archive.ics.uci.edu/ml/datasets/Wine

1.10. Exercises

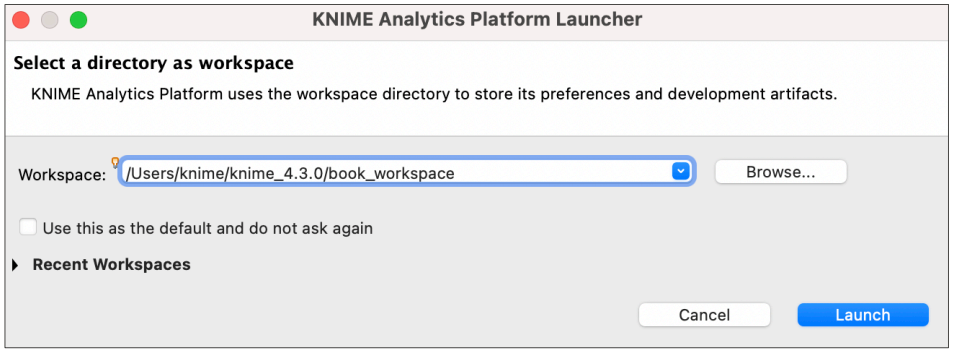
Exercise 1

Create your own workspace and name it “book_workspace”. You will use this workspace for the next workflows and exercises.

Solution to Exercise 1

- Launch KNIME
- In Workspace Launcher window, click “Browse”
- Select the path for your new workspace
- Click “OK”

1.27. Exercise 1: Create workspace "book_workspace"



To keep this as your default workspace, enable the option on the lower left corner.

SAMPLE

Exercise 2

Install the following extensions:

- KNIME Database
- KNIME Javascript Views
- KNIME Report Designer

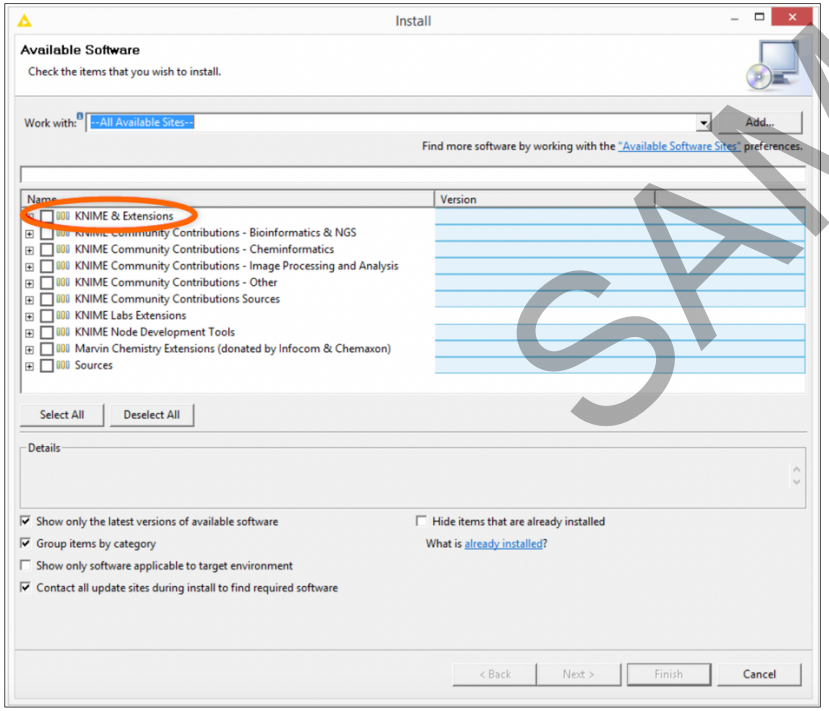
Solution to Exercise 2

From the Top Menu, select **"File" → "Install KNIME Extensions"**

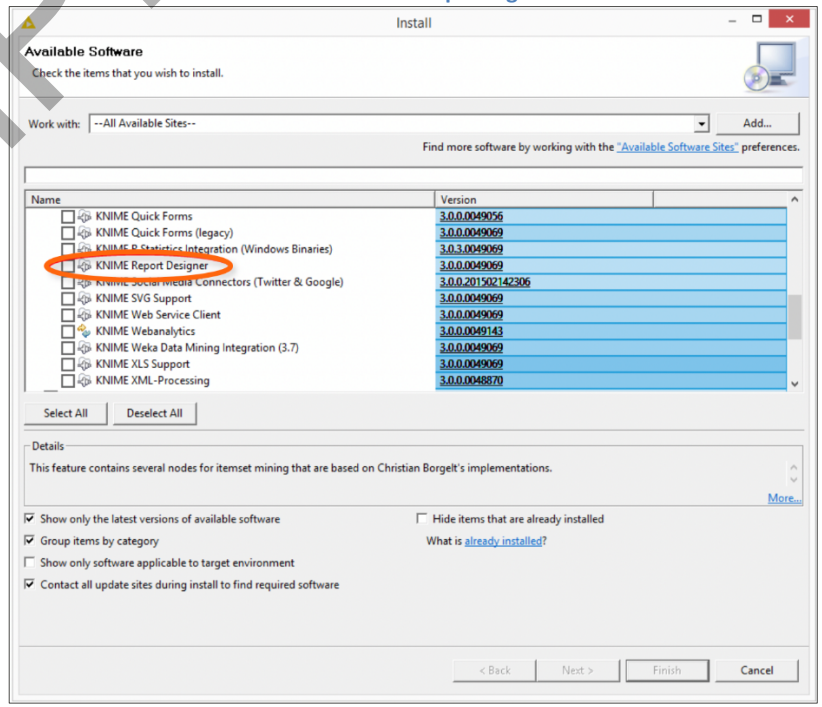
Select the required Extensions

Click **"Next"** and follow instructions

1.28. Exercise 2: List of KNIME Extensions



1.29. Exercise 2: Reporting Extension



SAMPLE

Exercise 3

Search all “Row Filter” nodes in the Node Repository.

From the “Node Description” panel, can you explain what the difference is between a “Row Filter”, a “Reference Row Filter”, and a “Nominal Value Row Filter”?

Show the node effects by using the following data tables:

Original Table

Position	name	team
1	The Black Rose	4
2	Cynthia	4
3	Tinkerbell	4
4	Mother	4
5	Augusta	3
6	The Seven Seas	3

Reference Table

Ranking	scores
1	22
3	14
4	10

SAMPLE

Solution to Exercise 3

Row Filter

The node allows for row filtering according to certain criteria. It can include or exclude: certain ranges (by row number), rows with a certain row ID, and rows with a certain value in a selectable column (attribute). In the example below we used the following filter criterion: `team > 3`

Original table

Position	name	team
1	The Black Rose	4
2	Cynthia	4
3	Tinkerbell	4
4	Mother	4
5	Augusta	3
6	The Seven Seas	3

Filtered table

Position	name	team
1	The Black Rose	4
2	Cynthia	4
3	Tinkerbell	4
4	Mother	4

Reference Row Filter

This node has two input tables. The first input table, connected to the bottom port, is taken as the reference table; the second input table, connected to the top port, is the table to be filtered. You have to choose the reference column in the reference table and the filtering column in the second table. All rows with a value in the filtering column that also exists in the reference column are kept, if the option "include" is selected; they are removed if the option "exclude" is selected.

Reference Table

Ranking	scores
1	22
3	14
4	10

Filtering Table

Position	name	team
1	The Black Rose	4
2	Cynthia	4
3	Tinkerbell	4
4	Mother	4
5	Augusta	3
6	The Seven Seas	3

Resulting Table

Position	name	team
1	The Black Rose	4
3	Tinkerbell	4
4	Mother	4

SAMPLE

In the example above, we use "Ranking" as the reference column in the reference table and "Position" as the filtering column in the filtering table. We have chosen to include the common rows.

Nominal Value Row Filter

Filters the rows based on the selected value of a nominal attribute. A nominal column and one or more nominal values of this attribute can be selected as the filter criterion. Rows that have these nominal values in the selected column are included in the output data. Basically it is a Row Filter applied to a column with nominal values. Nominal columns are string columns and nominal values are the values in it.

In the example below, we use "name" as the nominal column and "`name = Cynthia`" as the filtering criterion.

Original table

Position	name	team
1	The Black Rose	4
2	Cynthia	4
3	Tinkerbell	4
4	Mother	4
5	Augusta	3
6	The Seven Seas	3

Filtered table

Position	name	team
2	Cynthia	4

SAMPLE