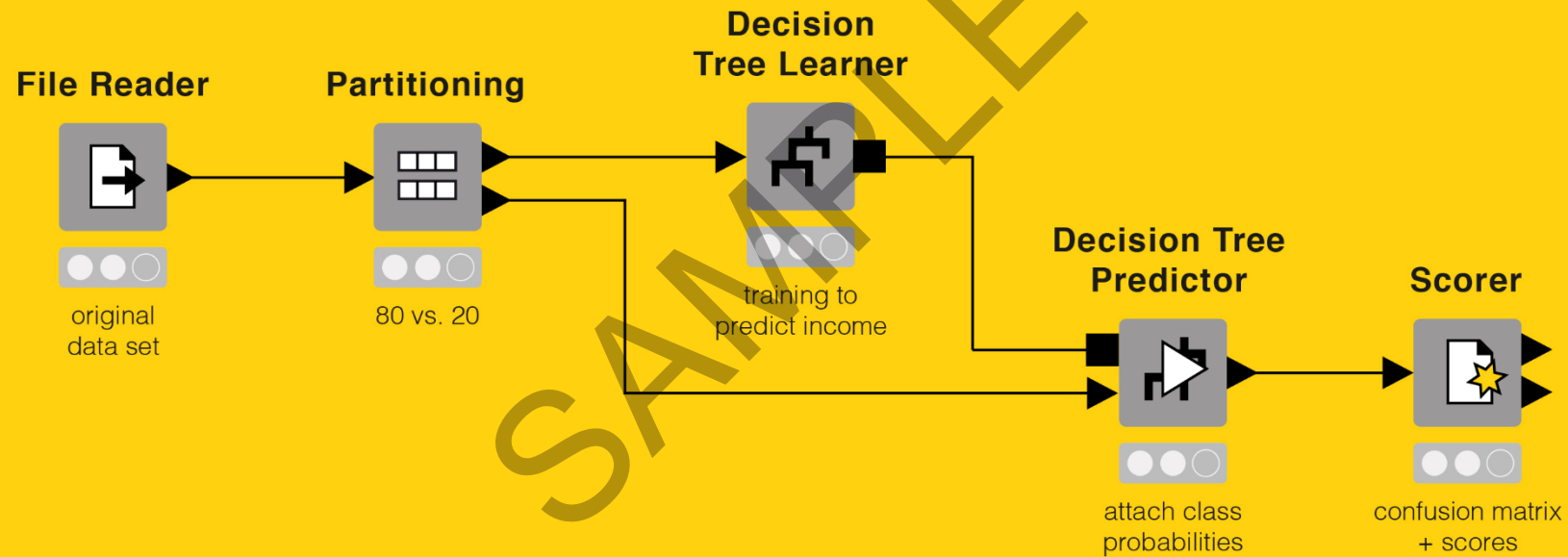


KNIME® BEGINNER'S LUCK



A Guide to KNIME Analytics Platform for Beginners

Author: Rosaria Silipo

SAMPLE

Copyright© 2019 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 4.0**.

For information regarding permissions and sales, write to:

KNIME Press
Technoparkstr. 1
8005 Zurich
Switzerland

knimepress@knime.com

ISBN: 978-3-033-02850-0

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	20
The “Workspace Launcher”	21
1.5. KNIME workflow	21
What is a workflow	22
What is a node	23
1.6. <i>.knwf</i> and <i>.knar</i> file extensions	23
1.7. KNIME workbench	24
The KNIME Workbench	25
Top menu	26
Tool Bar	29
Hotkeys	30
Node Repository	31
Search box	31
KNIME Explorer	31

EXAMPLES Server.....	32
Mounting Servers in KNIME Explorer	33
Workflow Editor	34
Customizing the Workflow Editor.....	35
Workflow Annotations	35
Other Workbench Customizations	36
Node Monitor View	36
1.8. Download the KNIME Extensions	37
Installing KNIME Extensions	37
1.9. Data and workflows for this book.....	38
1.10. Exercises	39
Exercise 1	39
Exercise 2	39
Exercise 3	40
Chapter 2. My first workflow.....	44
2.1. Workflow operations.....	44
Create a new Workflow Group.....	45
Create a new workflow.....	46
Save a workflow.....	47
Delete a workflow.....	47
2.2. Node operations	48
Create a new node.....	48
Configure a node	49
Execute a node	49
Node Text	50

Node Description	50
View the processed data	51
2.3. Read data from a file	51
Create a “File Reader” node	52
Configure the “File Reader” node.....	53
Customizing Column Properties	54
Advanced Reading Options.....	55
The <i>knime://</i> protocol.....	56
2.4. KNIME data structure and data types.....	57
KNIME data structure	59
2.5. Filter Data Columns	59
Create a “Column Filter” node	60
Configure the “Column Filter” node.....	61
2.6. Filter Data Rows.....	62
Create a “Row Filter” node.....	63
Configure the “Row Filter” node	63
Row filter criteria	65
2.7. Write Data to a File.....	67
Create a “CSV Writer” node.....	67
Configure the “CSV Writer” node	68
2.8. Exercises	69
Exercise 1	69
Exercise 2.....	72
Chapter 3. My first data exploration	75
3.1. Introduction	75

3.2. Replace Values in Columns	76
Column Rename	77
Rule Engine	79
3.3. String Splitting	81
Cell Splitter by Position	82
Cell Splitter [by Delimiter]	83
RegEx Split (= Cell Splitter by RegEx)	84
3.4. String Manipulation	85
String Manipulation	85
Case Converter	87
String Replacer	88
Column Combiner	89
Column Resorter	90
3.5. Type Conversions	91
Number To String	91
String To Number	92
Double To Int	93
3.6. Database Operations	93
SQLite Connector	95
MySQL Connector	96
Workflow Credentials	97
Master Key (deprecated)	98
DB Writer	99
Import a JDBC Database Driver	99
DB Table Selector	102

DB Reader	103
3.7. Aggregations and Binning	103
Numeric Binner	105
GroupBy: "Groups" tab	106
GroupBy: Aggregation tabs	107
Pivoting	108
3.8. Nodes for Data Visualization	110
3.9. Scatter Plot	110
Scatter Plot: Interactive View	112
3.10. Graphical Properties	113
Color Manager	114
3.11. Line Plots and Parallel Coordinates	116
Line Plot	116
Parallel Coordinates	118
3.12. Bar Charts and Histograms	119
Bar Chart	120
Table View	123
3.13. Exercises	126
Exercise 1	126
Exercise 2	128
Exercise 3	128
Chapter 4. My First Model	132
4.1. Introduction	132
4.2. Split and Combine Data Sets	133
Row Sampling	133

Partitioning	134
Shuffle.....	135
Concatenate	136
4.3. Transform Columns	137
PMML	138
Missing Value.....	139
Normalizer	140
Normalization Methods.....	141
Normalizer (Apply).....	141
4.4. Machine Learning Models	142
Naïve Bayes Model	143
Naïve Bayes Learner	144
Naïve Bayes Predictor	144
Scorer (Javascript).....	146
Decision Tree	150
Decision Tree Learner: Options Tab.....	151
Decision Tree Learner: PMML Settings Tab	152
Decision Tree Predictor	153
Decision Tree View	158
ROC Curve.....	159
Artificial Neural Network	161
RProp MLP Learner	161
Multilayer Perceptron Predictor	163
Write/Read Models to/from file.....	164
PMML Writer	164

PMML Reader	166
Statistics.....	167
Regression	169
Linear Regression Learner	170
Regression Predictor.....	171
Clustering.....	171
k-Means	172
Cluster Assigner	173
Hypothesis Testing.....	173
4.5. Exercises	174
Exercise 1	174
Exercise 2	176
Exercise 3	176
Chapter 5. The Workflow for my First Report	178
5.1. Introduction	178
5.2. Installing the Report Designer Extension.....	179
5.3. Transform Rows.....	179
RowID	182
Unpivoting	183
Sorter	185
5.4. Joining Columns.....	185
Joiner	187
Joiner node: the „Joiner Settings” tab	188
Joiner node: the “Column Selection” tab	189
Join mode	190

5.5. Misc Nodes	191
Java Snippet (simple)	192
Java Snippet	193
Math Formula	194
Math Formula (Multi Column)	195
5.6. Marking Data for the Reporting Tool	196
Data to Report	196
5.7. Cleaning Up the Final Workflow	197
Create a Meta-node from scratch	197
Collapse pre-existing nodes into a Meta-node	199
Expand and Reconfigure a Meta-node	199
5.8. Exercises	201
Exercise 1	201
Exercise 2	202
Exercise 3	203
Chapter 6. My First Report	206
6.1. Switching from KNIME to BIRT and back	206
6.2. The BIRT Environment	207
6.3. Master Page	208
6.4. Data Sets	210
6.5. Title	211
6.6. Grid	212
6.7. Tables	214
Toggle Breadcrumb	218
6.8. Style Sheets	218

Create a new Style Sheet	219
Apply a Style Sheet	220
6.9. Maps	222
6.10. Highlights	223
6.11. Page Break	225
6.12. Charts.....	225
Select Chart Type	226
Select Data.....	227
Format Chart.....	229
How to change the chart properties.....	237
6.13. Generate the final document	237
6.14. Exercises	238
Exercise 1	238
Exercise 1a	239
Exercise 2	240
Exercise 3	241
References	244
Node and Topic Index.....	245

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)

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 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 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](#)".

The 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

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 4.0. 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 for KNIME users.

Useful Web Pages	
http://www.knime.org	The root page in the KNIME web site.
https://www.knime.com/knime-software	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-introductory-course	“Introductory Course to Data Science”. This is the landing page to learn more about the specific KNIME functionalities. It covers the whole data science cycle from data access and data exploration to machine learning and control structures.
http://www.knime.org/learning-hub	This is a collection of very basic learning material - as web sites, videos, webinars, courses, and more. It is organized by topic, like text mining or chemistry, or basic KNIME nodes, etc...
https://forum.knime.com/	In the www.knime.org site you can find a number of resources. What I find particularly useful is the KNIME Forum. Here you can ask questions about how to use KNIME or about how to extend KNIME with new nodes. Someone from the KNIME community answers always and quickly.

Courses, Events, and Videos

Course for KNIME Analytics Platform	KNIME periodically offers onsite two-day courses for KNIME Analytics Platform. This includes basic and advanced elements. To check for the next available date/place and to register, just go to the KNIME Course web site https://www.knime.com/courses
KNIME Webinars	A number of webinars are also available since May 2013 on specific topics, like chemistry nodes, text mining, integration with other analytics tools, automated machine learning, 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
KNIME Meetups and KNIME Summits	KNIME Meetups and KNIME Summits are held periodically all over the world. These are always good chances to learn more about KNIME, 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 and https://www.knime.com/summits)
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

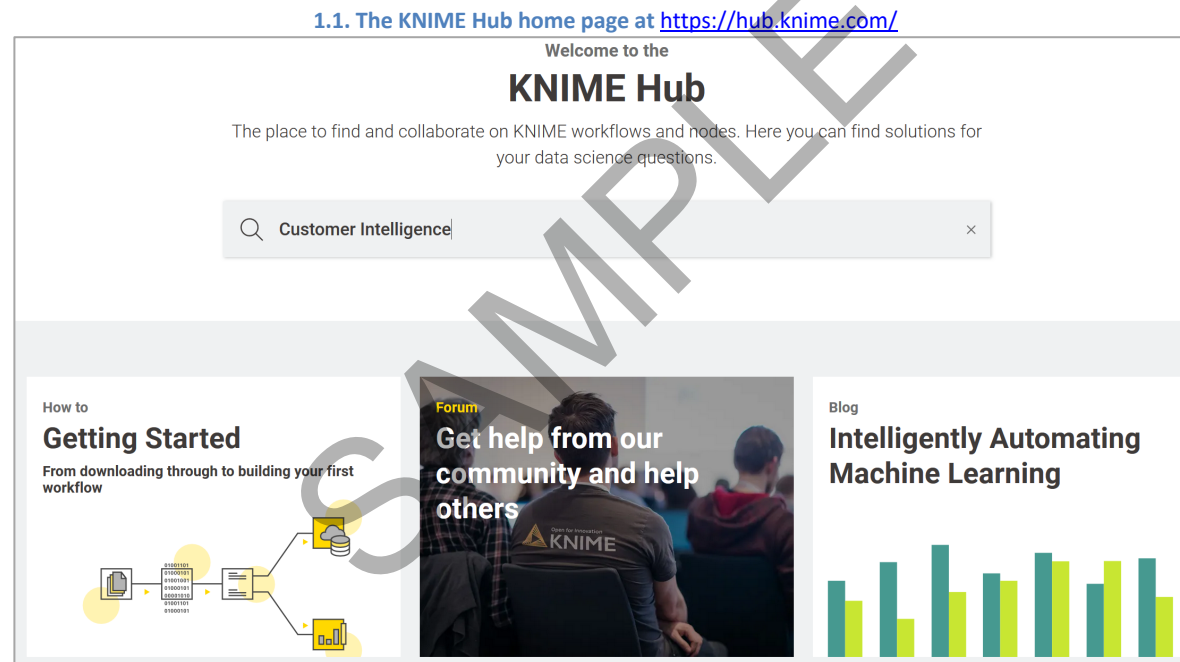
KNIME Platform	For the advanced use: Rosaria Silipo, Jeanette Prinz, "KNIME Advanced Luck" (https://www.knime.com/knimepress/knime-advanced-luck)
Reporting Suite	The KNIME Reporting Suite is based on BIRT, another open source tool for reporting. Here is a basic guide on how to use BIRT: <i>D. Peh, N. Hague, J. Tatchell, "BIRT. A field Guide to Reporting.", Addison-Wesley, 2008</i>
Data Science and KNIME	For an overview of data science, data mining, and data analytics, please check: <i>Berthold M.R., Borgelt C., Höppner F., Klawonn F., "Guide to intelligent data analysis", Springer 2010.</i>

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 search for applications, nodes, and components uploaded by KNIME users in this shared place of the KNIME community.



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](#).

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

The screenshot shows a search interface on the KNIME Hub. At the top, a search bar contains the text "Customer Intelligence" with a magnifying glass icon on the left and a close button (X) on the right. Below the search bar, the text "488 results" is displayed. A navigation bar below the results shows "All", "Nodes", and "Workflows" (which is selected and underlined). The main content area displays a list of four workflow cards. Each card has a small icon on the left, a title, a description, a list of tags, a breadcrumb trail, and a profile picture or icon on the right.

Workflow Title	Description	Tags	Breadcrumb	Author
B2B Customer Intelligence Use Case	Showcasing tools and methods available for the Citizen Data Scientist to improve and predict B2B customer behaviour.		Users > knime > Examples > 50_Applications > 42_Customer_Intelligence	A
Customer Segmentation	This workflow performs customer segmentation by means of clustering k-Means node. The second part of the workflow implements an interactive wizard on the WebPortal to visualize and label (or write not...	clustering, k-Means, customer segmentation, WebPortal, visualization, labelling, interactive visualization, Customer Intelligence, CI	Users > knime > Examples > 50_Applications > 24_Customer_Segmentation_Use_Case	[Profile Picture]
Basic Customer Segmentation	This workflow implements a basic customer segmentation through a clustering procedure. No input is required from the business analyst.	clustering, k-Means, customer segmentation, Customer Intelligence, CI	Users > knime > Examples > 50_Applications > 24_Customer_Segmentation_Use_Case	[Profile Picture]
Training a Churn Predictor	This workflow is an example of how to build a basic PMML model for a churn prediction using a Decision Tree algorithm.	Customer Intelligence, CI, churn	Users > knime > Examples > 50_Applications > 18_Churn_Prediction	R

Clicking one of applications in the list opens its web page (Fig. 1.3), with a nice explanatory picture of the implementation. The button on the right “Open workflow” then will let you open the application within your current KNIME Analytics Platform installation.

3.3. The page dedicated to the application named "Customer Segmentation" on the KNIME Hub

Customer Segmentation

This workflow performs:

1. clustering (k-Means)
2. visualization and labelling of clusters
3. summary of cluster stats

Data Reading

- Contract Data
- Operational Data

Parameter Selection

- No. of Clusters
- Input Columns

Clustering

- k-Means

Cluster Labeling

- Visualize Cluster in Scatter Plot & Table of Cluster Centers
- Label Cluster Loop End (2 ports)
- collect all cluster centers with new labels
- visualize cluster centers and cluster stats

On WebPortal

- New Labelling of Clusters
- Cluster Visualization
- Write data to File with new cluster labels

Display Cluster Result

- PCA Scatter Plot
- Data Scatter Plot
- Cluster Centers Scatter Plot

Display Labeled Clusters

- OutputFile.txt

This workflow performs customer segmentation by means of clustering k-Means node. The second part of the workflow implements an interactive wizard on the WebPortal to visualize and label (or write notes) about the single clusters.

1.3. Download and install KNIME Analytics Platform

There are two available KNIME products:

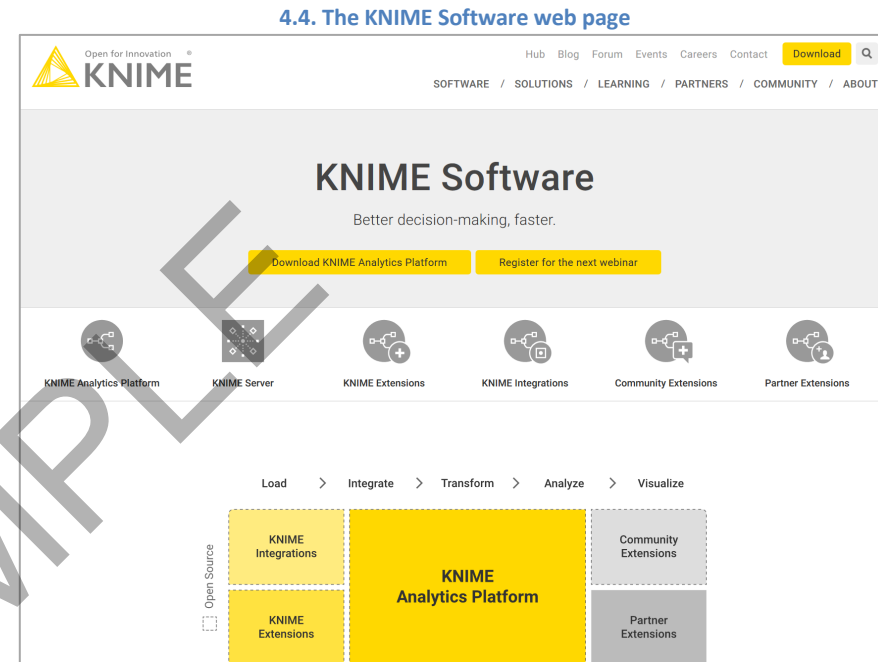
- the open source [KNIME Analytics Platform](https://www.knime.com/knime-analytics-platform), which can be downloaded free of charge at <https://www.knime.com/knime-software> under the GPL version 3 license
- the [KNIME server](https://www.knime.com/knime-server), which is described at <https://www.knime.com/knime-server>

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 development and management, data warehousing, integration, and scalability for the data science lab.

In this book, however, we work with KNIME Analytics Platform (open source) version 4.0. To start playing with KNIME Analytics Platform, first, you need to download it to your machine.

Download KNIME Analytics Platform

- Go to www.knime.org
- In the lower part of the first screen of the main page, click “KNIME Software”
- In the “KNIME Software” page, click the button “Download KNIME Analytics Platform”.
- 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. 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.

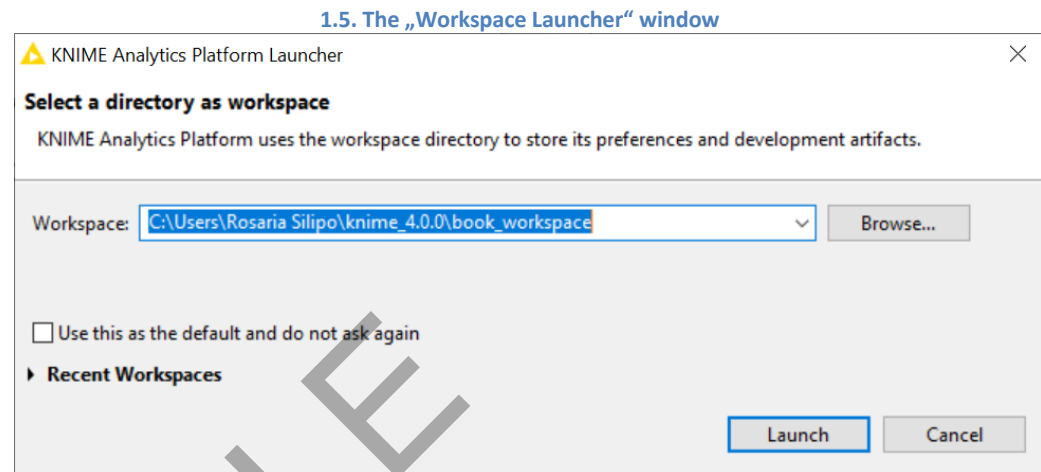
After the splash screen, the “Workspace Launcher” window requires you to enter the path of the workspace.

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.



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.x.y\book_workspace”.

1.5. KNIME workflow

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

Small little boxes, 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

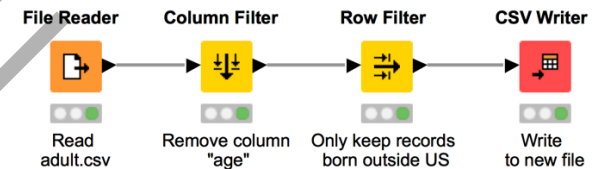
KNIME Analytics Platform implements its workflows **graphically**. Each step of the data analysis is implemented and executed through a little box, called **node**. A sequence of nodes makes a workflow.

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



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.

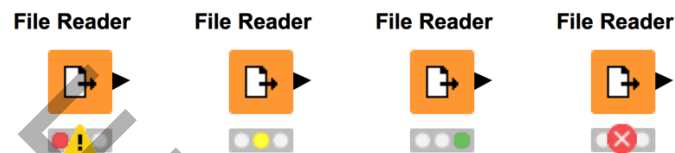
Every node in KNIME has 4 states:

- 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



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 platform.

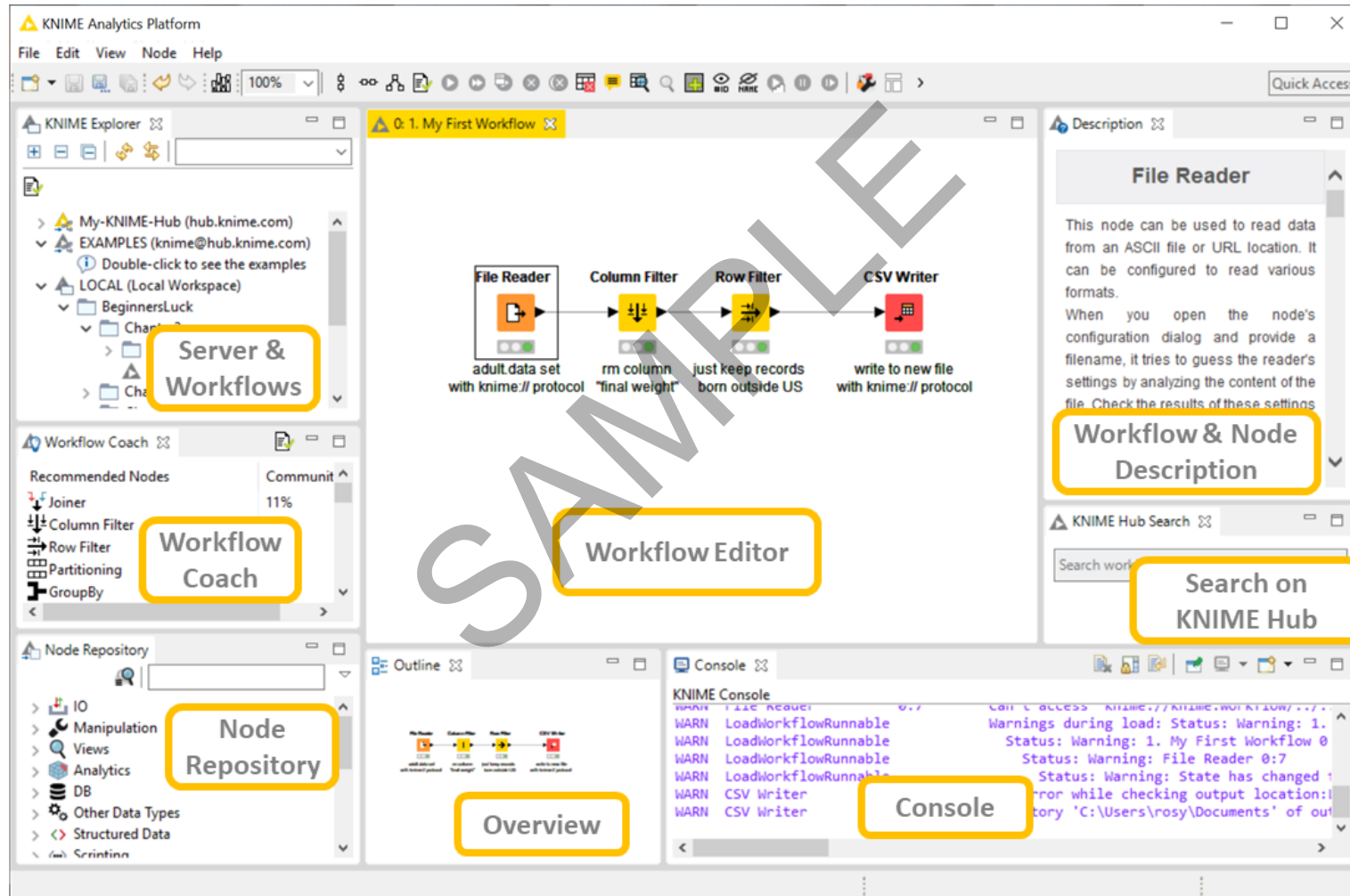
1.8. *.knwf* and *.knar* files are associated with KNIME Analytics Platform. A double-click opens the workflow(s) directly inside the platform

▲ 01_From_Strings_to_Documents.knwf	10/4/2017 9:45 AM	KNIME Workflow ...	18,619 KB
▲ 04_Interaction_Graph.knwf	9/29/2017 8:20 AM	KNIME Workflow ...	9,465 KB
▲ 06_REST_Examples_Google_Geocode.knwf	7/29/2017 7:09 PM	KNIME Workflow ...	62 KB
▲ 06_Semantic_Web_updated.knar	11/3/2016 2:24 PM	KNIME Archive File	178 KB
▲ AzureDemoWorkflowArchive.knar	5/5/2017 11:24 AM	KNIME Archive File	24,104 KB
▲ Building a Simple Classifier_.knwf	2/18/2017 5:46 PM	KNIME Workflow ...	43 KB
▲ Cookbook_Ch5.knar	11/24/2017 10:03 ...	KNIME Archive File	477 KB
▲ Cookbook_Ch6.knar	11/24/2017 10:26 ...	KNIME Archive File	155 KB
▲ Corsair.knwf	7/10/2017 4:20 PM	KNIME Workflow ...	106 KB

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.

1.9. The KNIME workbench



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.

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.

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

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.

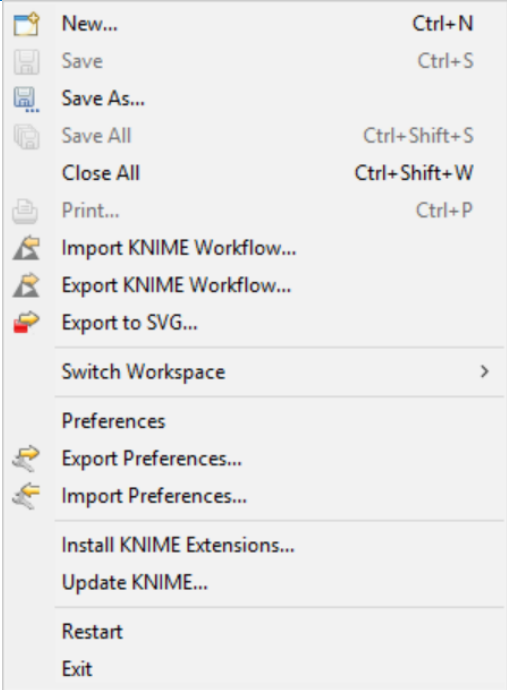
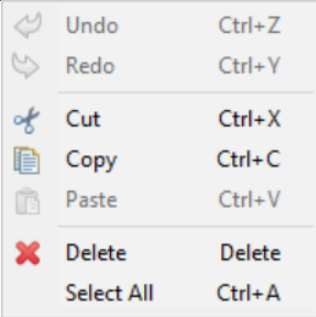
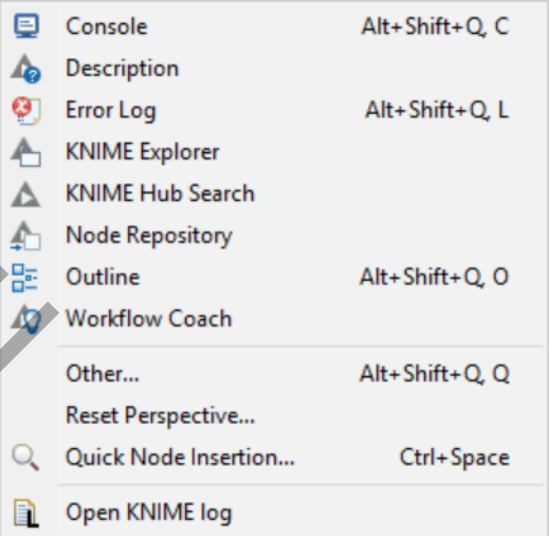
Console

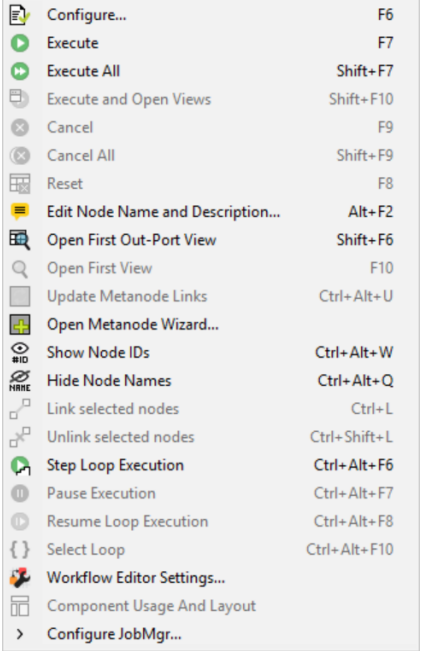
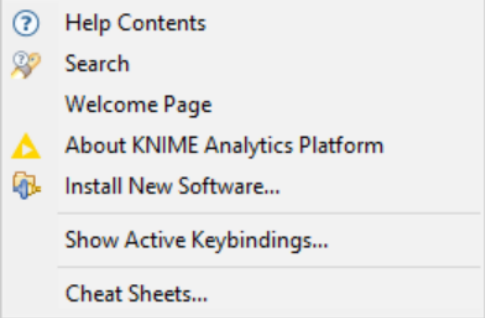
The "Console" panel displays error and warning messages to the user.

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.

Top menu

File	Edit	View
		
<p>File includes the traditional File commands, like “New” and “Save”, in addition to some KNIME specific commands, like:</p> <ul style="list-style-type: none"> - Import/Export KNIME workflow... - Export to SVG - Switch Workspace - Preferences with Export/Import Preferences - Install KNIME Extensions - Update KNIME 	<p>Edit contains edit commands.</p> <p>Undo and Redo refer to the last performed actions.</p> <p>Cut, Copy, Paste, and Delete refer to selected nodes in the workflow.</p> <p>Select All selects all the nodes of the workflow in the workflow editor.</p>	<p>View contains the list of all panels that can be opened in the KNIME workbench.</p> <p>A closed panel can be re-opened here.</p> <p>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.</p> <p>Option “Other” opens additional views useful to customize the workbench.</p>

Node	Help
	
<p>Node refers to all possible operations that can be performed on a node. A node can be:</p> <ul style="list-style-type: none"> - 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) <p>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.</p> <p>Option “Open Meta Node Wizard” starts the wizard to create a new meta node in the workflow editor.</p>	<p>Help Contents provides general Help about the Workbench, BIRT, and KNIME.</p> <p>Search opens a panel on the right of the “Node Description” panel to search for specific Help topics or nodes.</p> <p>Welcome Page (re-)opens the Welcome Page</p> <p>Install New Software is the door to install KNIME Extensions from the KNIME Update sites.</p> <p>Show Active Keybindings summarizes all keyboard commands for the workflow editor.</p> <p>Cheat Sheets offer tutorials on specific topics: the reporting tool, cvs, Plug-ins.</p>

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 resetted workflows without the data produced by each node. This generates considerably smaller export files.

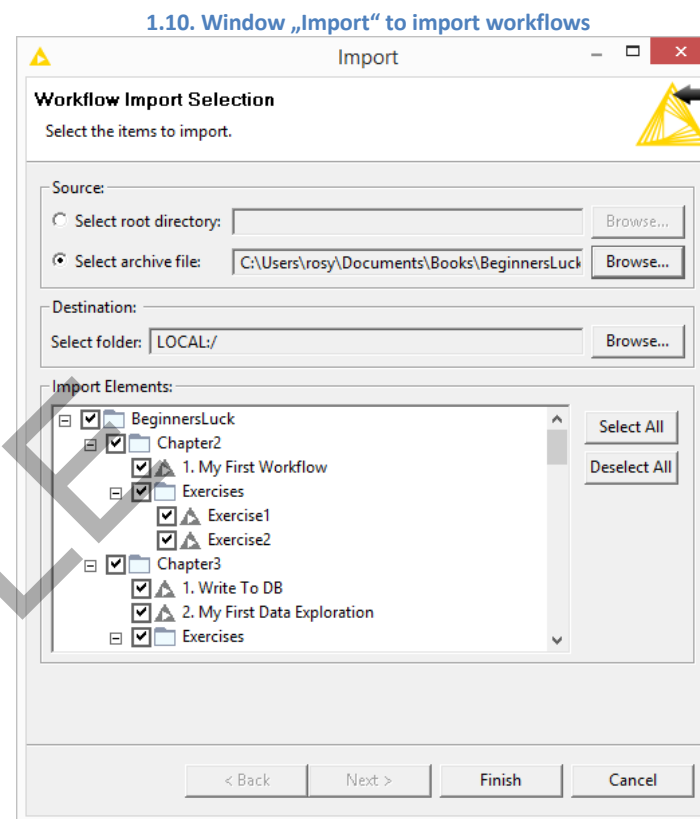
Simply copying a workflow from one folder to another can create a number of problems related to internal KNIME updates. Copying workflows by using the option **"Import KNIME workflow"** or by double-click is definitely safer.

"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).

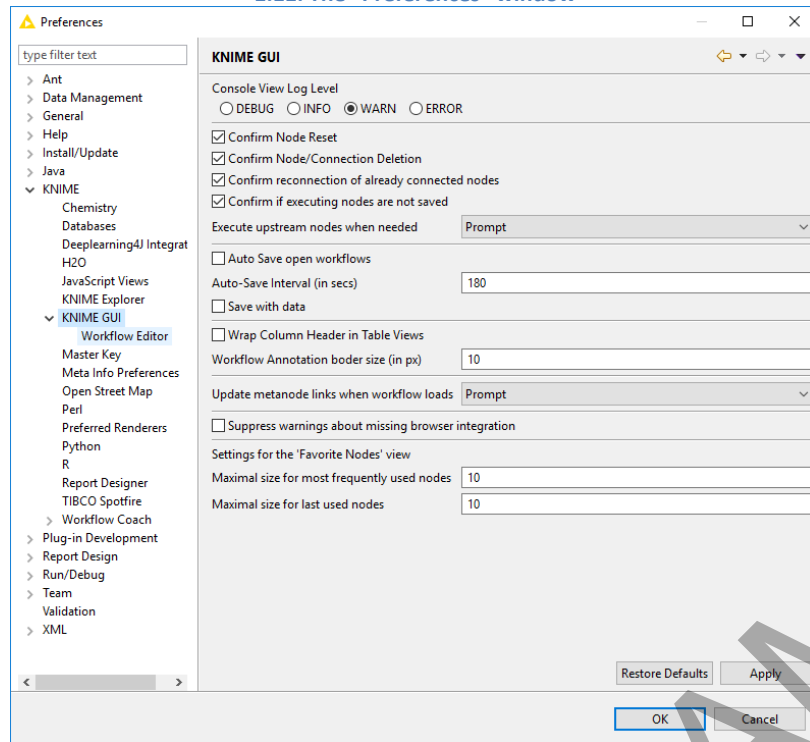
"File" -> "Switch Workspace" changes the current workspace with a new one.

"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.



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 right, 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,

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 menu 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

Node Repository

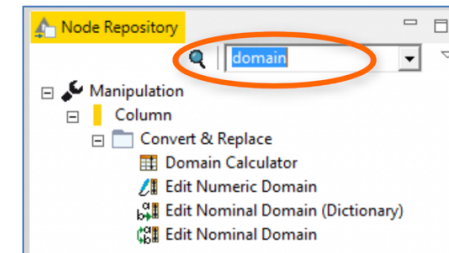
In the lower left corner we find the Node Repository, containing all installed nodes organized in categories and subcategories. KNIME Analytics Platform has accumulated by now more than 1500 nodes. It has become hard to remember the location of each node in the Node Repository. To solve this problem, 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

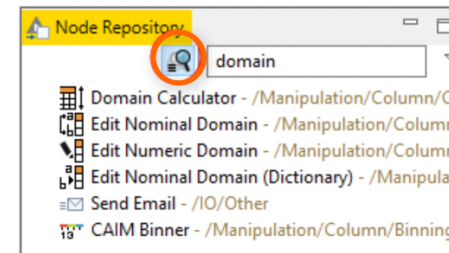
At the top of the “Node Repository” panel there is a **search box**. If you type a keyword in the search box and hit “Enter”, you obtain the list of nodes containing an exact match of that keyword. Press the “Esc” key to see all nodes again.

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.

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



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



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.

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:

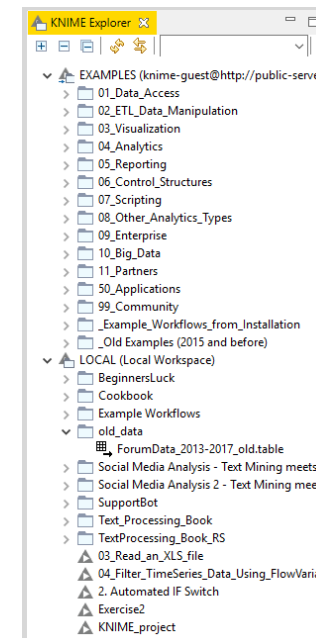
- right click “EXAMPLES” in the “KNIME Explorer” panel
- select “Login”

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.

1.15. 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.

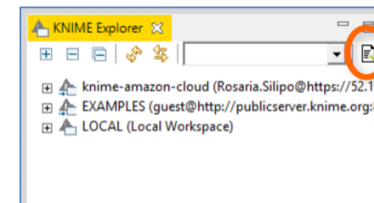
Mounting Servers in KNIME Explorer

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

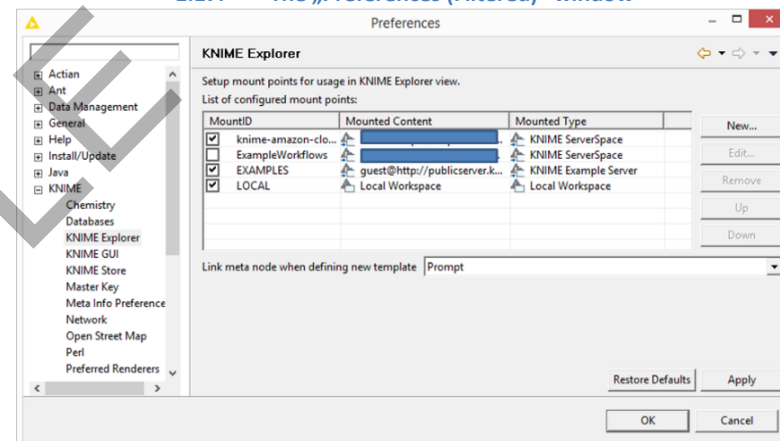
- Select the “KNIME Explorer” panel
- Click the “Configure Explorer View” button
- The “Preferences (Filtered)” window opens on the “KNIME Explorer” page and lists all KNIME Servers already mounted in this KNIME instance. The two KNIME servers available by default on every KNIME instance are the local workspace “LOCAL” and the KNIME public Server “EXAMPLES”.
- 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. 1.18)
- Use the “Test Connection” button to automatically retrieve the default mountpoint for the selected server.

The same KNIME Explorer “Preferences” page in figure 1.17 can be reached via “File” in the top menu -> “Preferences” -> “KNIME Explorer”.

1.16. The „Configure KNIME Explorer“ button



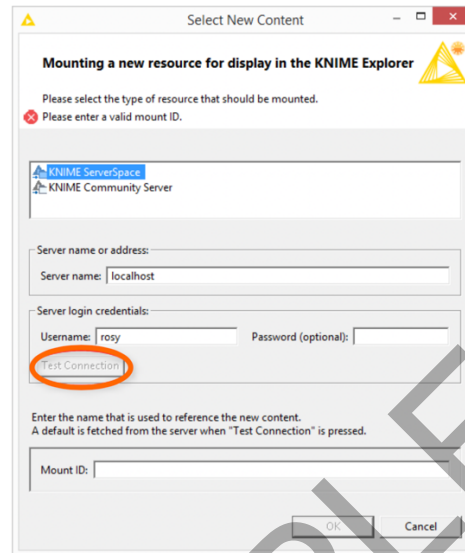
1.17. The „Preferences (Filtered)“ window



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

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

1.18. The "Select New Content" window



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.

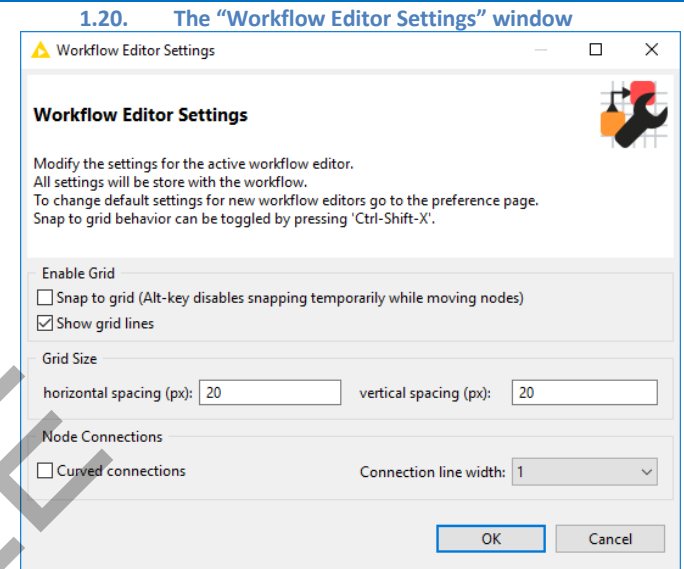
1.19. "Change Workflow Editor Settings" Button in Tool Bar



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.



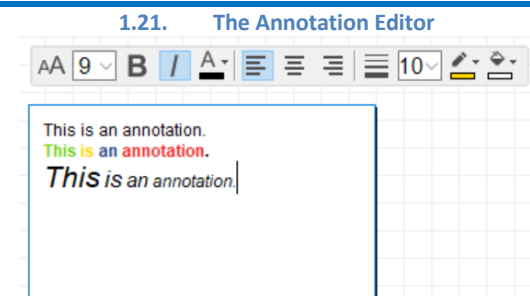
Adding annotations to the canvas

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, thickness).
- To reopen an annotation, just double-click at the top left corner, where the pencil icon is.



Other Workbench Customizations

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

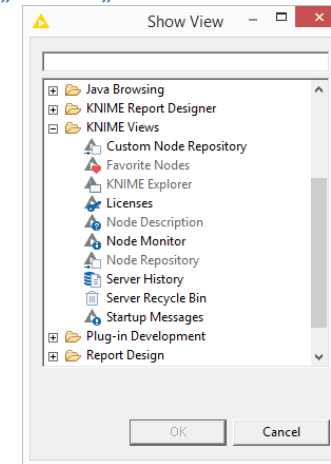
Popular views, for example, are the “Node Monitor”, the “Custom Node Repository”, and the “Licenses” and “Server” views, if you have a connected server. All these extra views can be found in the Top menu under “View” -> “Other” -> “KNIME Views”.

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”

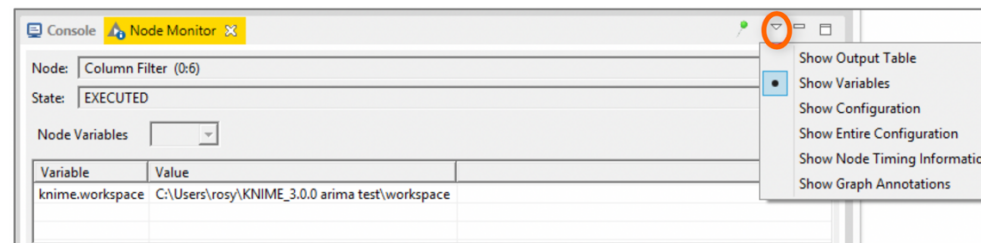


Node Monitor View

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

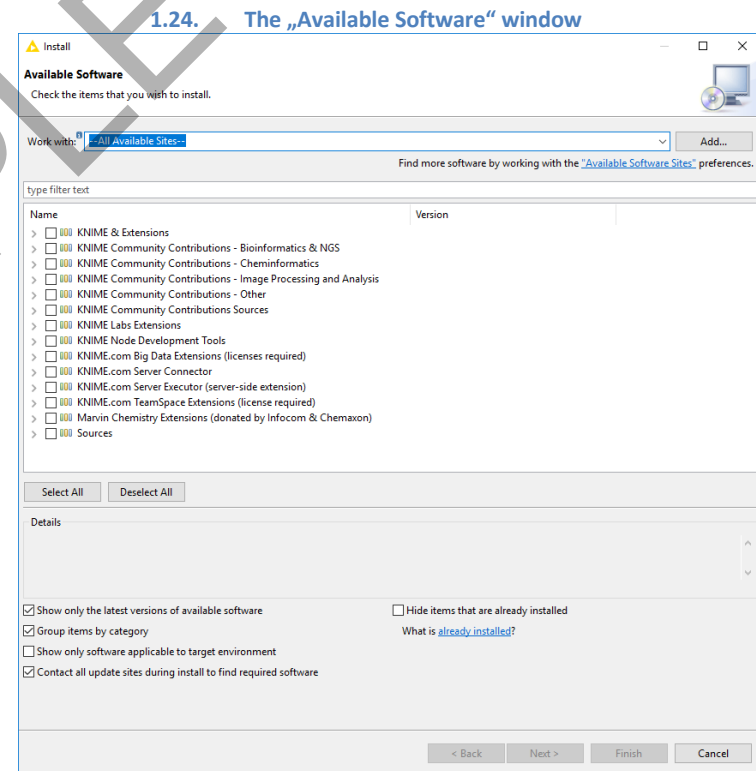
To install a new KNIME extension from within KNIME Analytics Platform, there are two options.

1. From the Top Menu, select **“File”** -> **“Install KNIME Extensions”**, select the desired extension, click the **“Next”** button and follow the wizard instructions.

OR

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.

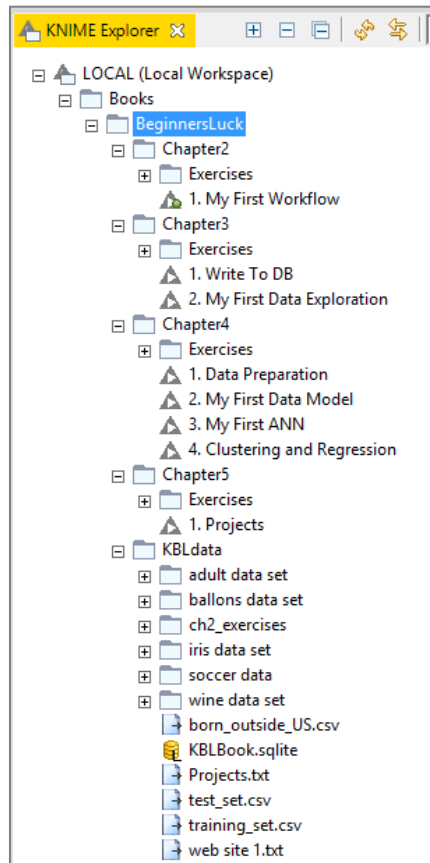
Once the selected KNIME extension(s) has/have been installed and KNIME has been restarted, you should see the new category, corresponding to the installed extension, in the **“Node Repository”**.



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

1.25. Workflows and data imported from the Download Zone .knar file



When you purchased this book, in the same email with the link to this pdf file, you should also have found a link to the Download Zone file. The Download Zone file is a .knar file that contains the data and workflows used and implemented throughout this book.

- Download the Download Zone .knar file onto your machine. Then:
 - Double click it
- OR
- import it into the KNIME Explorer via Select File -> Import KNIME Workflow ...

At the end of the import operation, in the KNIME Explorer panel you should find a BeginnersLuck folder containing Chapter2, Chapter3, Chapter4 and Chapter5 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 “Download Zone” 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

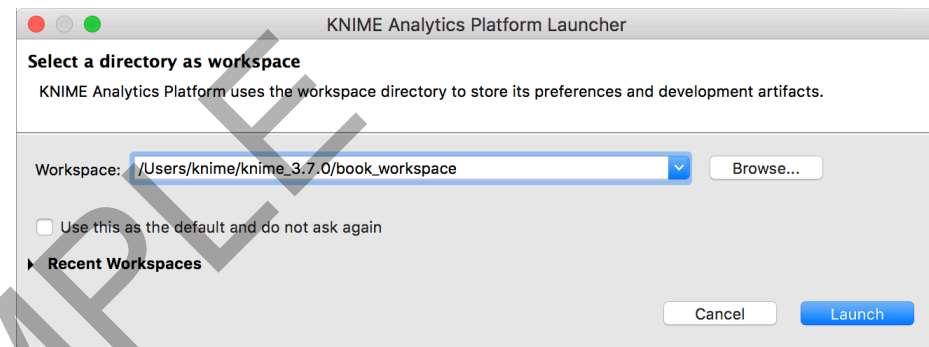
Create your own workspace and name it “book_workspace”. You will use this workspace for the next workflows and exercises.

Solution to Exercise 1

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

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

1.26. Exercise 1: Create workspace "book_workspace"



Exercise 2

Install the following extensions:

- KNIME Database
- KNIME Javascript Views
- KNIME Report Designer

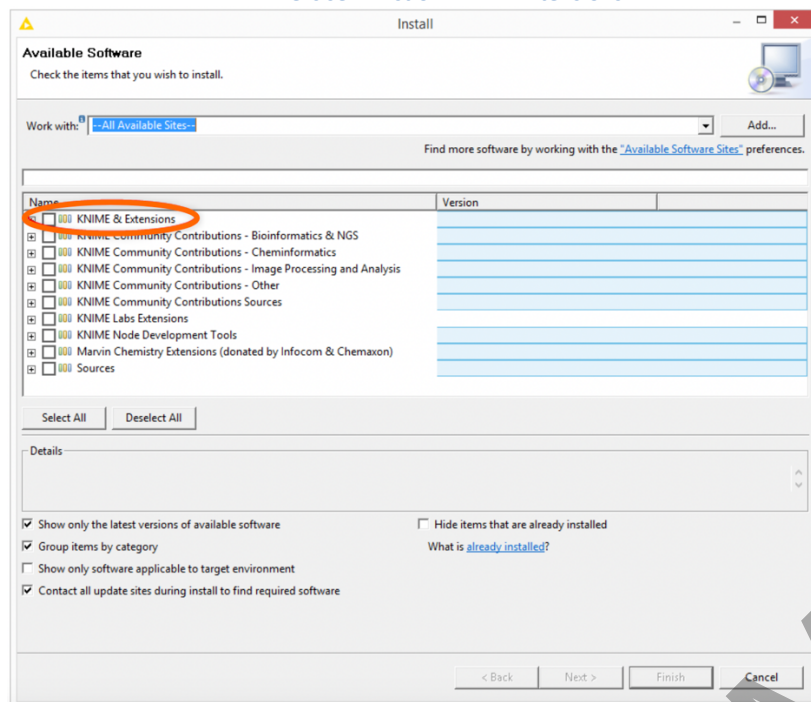
Solution to Exercise 2

From the Top Menu, select “File” -> “Install KNIME Extensions”

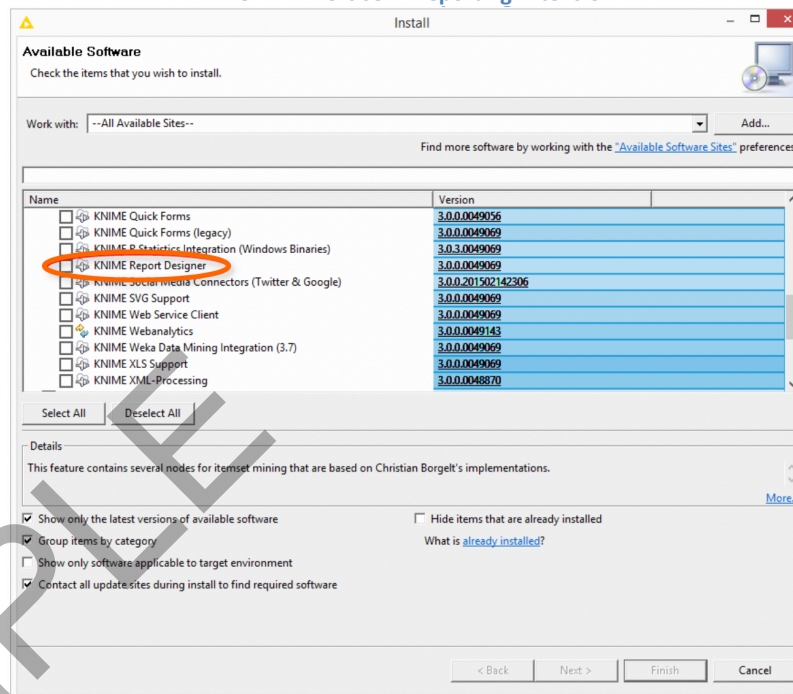
Select the required Extensions

Click “Next” and follow instructions

1.27 . Exercise 2: List of KNIME Extensions



1.28. Exercise 2: Reporting Extension



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

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 top port, is taken as the reference table; the second input table, connected to the bottom 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	Tinkerbelle	4
4	Mother	4
5	Augusta	3
6	The Seven Seas	3

Resulting Table

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

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