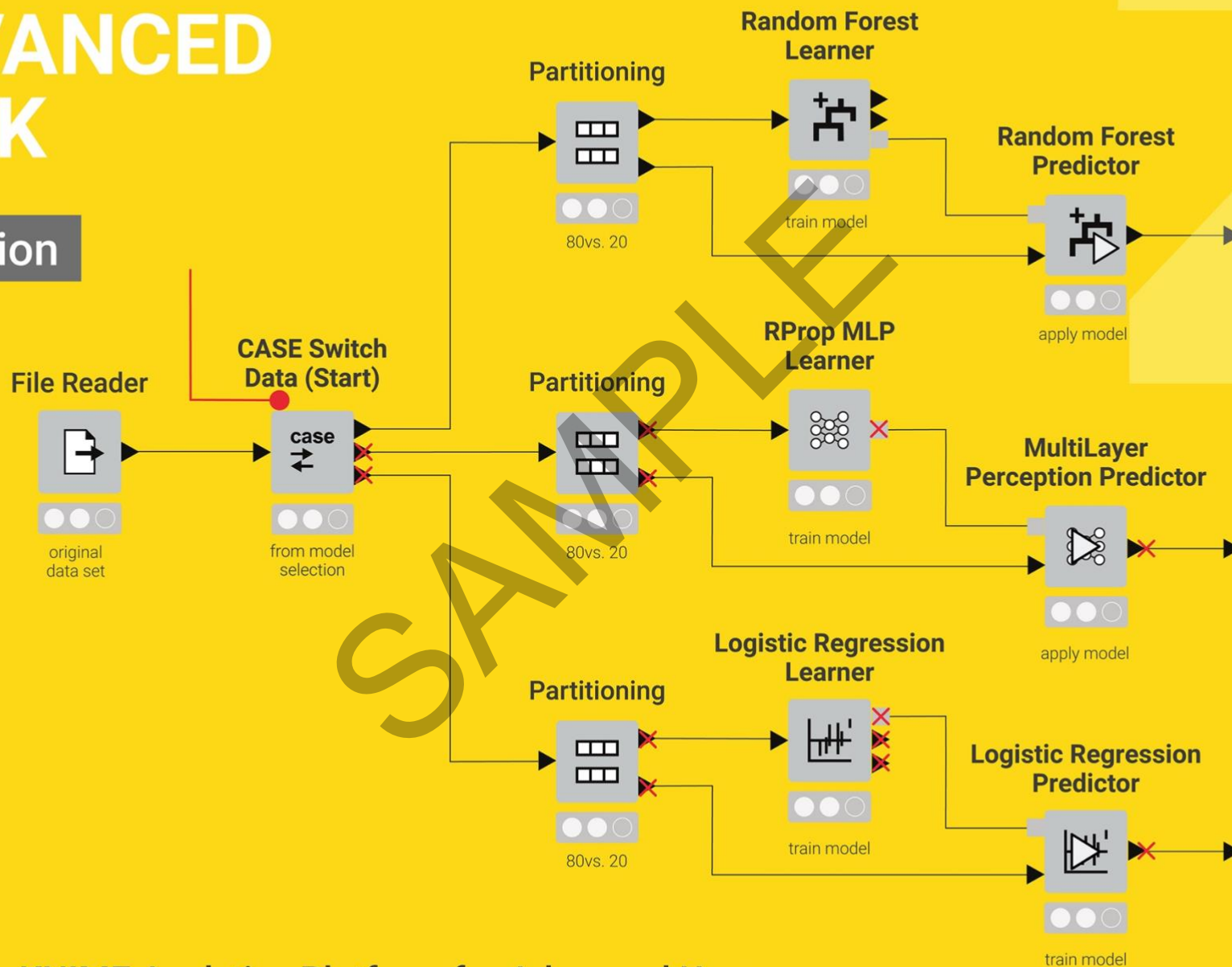


KNIME[®] ADVANCED LUCK

2nd Edition



A Guide to KNIME Analytics Platform for Advanced Users

Authors: Rosaria Silipo and Victor Palacios

Copyright©2022 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 4.5**.

For information regarding permissions and sales, write to:

KNIME Press
Hardturmstr. 66
8005 Zurich
Switzerland

knimepress@knime.com

ISBN: 978-3-9523926-0-7

Table of Contents

Acknowledgements	11
Chapter 1. Introduction.....	12
1.1. Purpose and Structure of this Book	12
1.2. Data and Workflows for this Book.....	13
1.3. Memory Usage in KNIME Analytics Platform.....	16
Chapter 2. Database Operations	19
2.1. Database Nodes	19
2.2. Connect to a Database: DB Connector Nodes	20
DB Connector	21
Register your own JDBC driver	22
Workflow Credentials.....	24
SQLite Connector	25
2.3. Select the Table to work on: the DB Table Selector Node.....	26
DB Table Selector.....	26
2.4. In-Database Processing.....	28
DB Row Filter.....	29
DB Column Filter	31
DB Query	32
2.5. Utility Nodes for Databases.....	33
DB SQL Executor	34
DB Query Injector	34
DB Query Extractor.....	35
2.6. Reading Data resulting from a SQL Query.....	35

DB Reader	36
Table Creator	37
Parameterized DB Query Reader	38
DB Query Reader	39
2.7. Writing Data resulting from a SQL Query	40
DB Writer	41
DB Connection Table Writer	42
2.8. Database UPDATE and DELETE Commands	42
DB Delete	43
DB Update	44
2.9. Database Type Mapping	46
2.10. Big Data Platforms and MongoDB	47
2.11. Exercises	49
Exercise 1	49
Exercise 2	50
Exercise 3	51
Chapter 3. Accessing Information from the Web	53
3.1. Accessing Google Sheets	53
Google Authentication	54
Google Sheets Connection	55
Google Sheets Reader	56
Google Sheets Appender	58
Google Sheets Updater	59
Google Sheets Writer	60

3.2. Accessing REST Services	60
GET Request: "Configuration Settings" Tab	63
GET Request: the other Tabs	64
JSON Path.....	65
JSON to Table	66
POST Request: "Configuration Settings" Tab	68
POST Request: "Request Body" Tab	69
3.3. Exercises.....	69
Exercise 1	69
Chapter 4. Date&Time Manipulation	71
4.1. The Date&Time Type	71
4.2. How to produce a Date&Time Column.....	72
String to Date&Time.....	73
Date&Time to String.....	75
Create Date&Time Range.....	76
4.3. Refine Date&Time Values	77
Modify Time	78
Date&Time Shift	79
4.4. Row Filtering based on Date&Time Criteria	80
Date&Time-based Row Filter	81
Extract Date&Time Fields	82
Date&Time Difference.....	84
4.5. Moving Average and Aggregation	85
Moving Average	88

Moving Aggregation.....	89
4.6. Time Series Analysis	91
Lag Column.....	92
4.7. Exercises.....	96
Exercise 1	96
Exercise 2.....	97
Chapter 5. Flow Variables.....	99
5.1. What is a Flow Variable?	99
5.2. Creating a Flow Variable for all Nodes in the Workflow.....	100
5.3. Flow Variable Values as Node Settings	102
The “Flow Variable” Button	103
The “Flow Variables” Tab in the Configuration Window.....	104
5.4. Creating a Flow Variable from within a Workflow.....	105
Transform a Data Value into a Flow Variable.....	106
Table Row to Variable.....	107
Transform a Configuration Setting into a Flow Variable.....	108
Configuration Nodes to Create Flow Variables.....	111
Integer Configuration	112
5.5. Inject a Flow Variable through the Flow Variable Ports	113
Flow Variable Injection into the Workflow	114
Merge Variables.....	115
5.6. Configurations, Widgets, Components, and KNIME WebPortal.....	115
5.7. Transform a Flow Variable into a Data Value.....	119
Variable to Table Row.....	119

5.8. Modifying Flow Variable Values	120
5.9. More Configuration Nodes and Widget Nodes	123
Value Selection Widget	124
Local File Browser Configuration.....	125
5.10. Composite View in Components	127
Interactive Range Slider Filter Widget.....	130
5.11. Components are for Sharing.....	132
5.12. Exercises.....	134
Exercise 1	134
Exercise 2.....	134
Exercise 3.....	137
Exercise 4.....	140
Chapter 6. Advanced Dashboards with Composite Views.....	142
6.1. A few examples of Advanced Dashboards	142
6.2. Interactively selecting one or more attributes to represent in a chart	147
Column Filter Widget	148
6.3. Dynamically update the dashboard	150
Refresh Button Widget.....	150
6.4. Interactively selecting rows by column values.....	154
Nominal Row Filter Widget	155
6.5. Text autocompletion.....	160
Autocomplete Text Widget	161
6.6. Custom filtering	163
Single Selection Widget	164

6.7. Animating a vanilla bar chart with community components	168
Animated Bar Chart.....	169
6.8. Exercises.....	172
Exercise 1.....	172
Exercise 2.....	175
Chapter 7. Loops	177
7.1. What is a Loop.....	177
7.2. Loop with a pre-defined number of iterations	179
Data Generator.....	180
Counting Loop Start.....	183
Loop End	183
7.3. Dedicated Commands for Loop Execution	186
7.4. Appending Columns to the Output Data Table.....	188
Loop End (Column Append).....	189
7.5. Loop on a List of Columns	192
Column List Loop Start	193
7.6. Loop on a List of Values.....	197
Table Row To Variable Loop Start.....	198
Cache	200
7.7. Loop on Data Groups and Data Chunks	201
Group Loop Start.....	202
Chunk Loop Start	204
Breakpoint.....	206
7.8. Keep Looping till a Condition is verified.....	206

Generic Loop Start	207
Variable Condition Loop End	207
7.9. Recursive Loop	209
Recursive Loop Start.....	210
Recursive Loop End	211
7.10. Exercises.....	212
Exercise 1.....	212
Exercise 2.....	214
Exercise 3.....	215
Exercise 4.....	217
Chapter 8. Switches.....	220
8.1. Introduction to Switches.....	220
8.2. The “IF Switch”- “END IF” switch block.....	221
IF Switch.....	222
End IF	223
Auto-Binner	225
8.3. The “Java IF (Table)” node	226
Java IF (Table)	227
8.4. The CASE Switch Block	228
CASE Switch Start.....	229
CASE Switch.....	230
8.5. Transforming an Empty Data Table Result into an Inactive Branch	231
Empty Table Switch	232
8.6. Exercises.....	233

Exercise 1.....	233
Exercise 2.....	235
References.....	238
Node and Topic Index.....	239

SAMPLE

Acknowledgements

We would like to thank a number of people for their help and encouragement in writing this book.

In particular, we would like to thank Bernd Wiswedel for answering our endless questions about calling external REST services from inside a workflow, Iris Adae for explaining the most advanced features of the Date&Time nodes, Andisa Dewi for significantly contributing to the Database Operations chapter, and Lada Rudnitckaia for updating the whole book to the latest version of KNIME Analytics Platform.

Special thanks go to Peter Ohl for reviewing the book contents and making sure that they comply with KNIME intended usage and to Heather Fyson for reviewing the book's English written style.

Finally, we would like to thank the whole KNIME Team for their support in publishing and advertising this book.

SAMPLE

Chapter 1. Introduction

1.1. Purpose and Structure of this Book

KNIME Analytics Platform is a powerful tool for data analytics and data visualization. It provides a complete environment for data analysis which is fairly simple and intuitive to use. This, coupled with the fact that KNIME Analytics Platform is open source, has led a large number of professionals to use it. In addition, third-party software vendors develop KNIME extensions in order to integrate their tools into it. KNIME nodes are now available that reach beyond customer relationship management and business intelligence, extending into the field of finance, life sciences, biotechnology, pharmaceutical, and chemical industries. Thus, the archetypal KNIME user is no longer necessarily a data science expert, although his/her goal is still the same: to understand data and to extract useful information.

This book was written with the intention of building upon the reader's first experience with KNIME software. It expands on the topics that were covered in the first KNIME user guide ("[KNIME Beginner's Luck](#)" [1]) and introduces more advanced functionalities. In the first guide [1], we described the basic principles of [KNIME Analytics Platform](#) and showed how to use it. We demonstrated how to build a basic workflow to manipulate, visualize, and model data, and how to build reports. Here, we complete these descriptions by introducing the reader to more advanced concepts. A summary of the chapters provides you with a short overview of the contents to follow.

Chapter 2 describes the nodes needed to connect to a database, import data, build an appropriate SQL query to select a subset of the data or for some required processing, and finally to write data back into the database. Accessing a database, importing data, and building SQL queries are the basic operations necessary for any, even very simple, data warehousing strategy.

Of course, the largest source of data is nowadays the Internet. Chapter 3 is dedicated to alternative ways of getting data besides files and databases, i.e. web data sources. Chapter 3 starts with the connectors to Google Sheets and continues with access to REST services. Those are definitely powerful tools to search for data elsewhere.

Chapter 4 introduces the Date&Time object and the nodes to turn a String column into a Date&Time column, to format it, to extract a time difference, and in general to perform date/time based operations. The Date&Time object provides the basis for working with time series. The last section of chapter 4 briefly describes a few nodes dedicated to time series analysis.

A very important concept for the KNIME workflows is the concept of "flow variables". Flow variables enable external parameters to be introduced into a workflow to control its execution. Chapter 5 describes what a flow variable is, how to create it, and how to edit it inside

the workflow, if needed. Capitalizing on the concept of flow variables, we introduce the components and explain how to make them configurable via Configuration nodes and Widget nodes.

Widget nodes especially can be helpful in the creation of dashboards, reports, and data apps. Chapter 6 shows how to build a rich dashboard through the composite view of an advanced component.

Most data operations in KNIME Analytics Platform are executed on a data matrix, named data table. This means that an operation is executed on all data rows. This is a big advantage in terms of speed and programming compactness. However, from time to time, a workflow also needs to run its rows, one after the other, through an operation. That is, sometimes it needs a real loop. Chapter 7 introduces a few nodes that implement loops: from a simple “for” cycle to more complex loops, such as looping on a list of values or feeding the current iteration results into the next iteration.

Chapter 8 illustrates the use of logical switches to change the workflow path upon compliance with some predefined condition.

In this introductory chapter, we list the data and the example workflows that have been built for this book and note the KNIME Extensions required to run some of the example workflows.

1.2. Data and Workflows for this Book

This book builds a few examples and provides solutions to the exercises. Solutions and exercises are contained in folder “AdvancedLuck” downloadable from the [KNIME Hub space of one of the authors](#) of this book. To download material from the KNIME Hub, you need to create a KNIME account, the same as for the [KNIME Forum](#). After entering the KNIME Hub, in order to download the workflows, just click on the cloud icon.

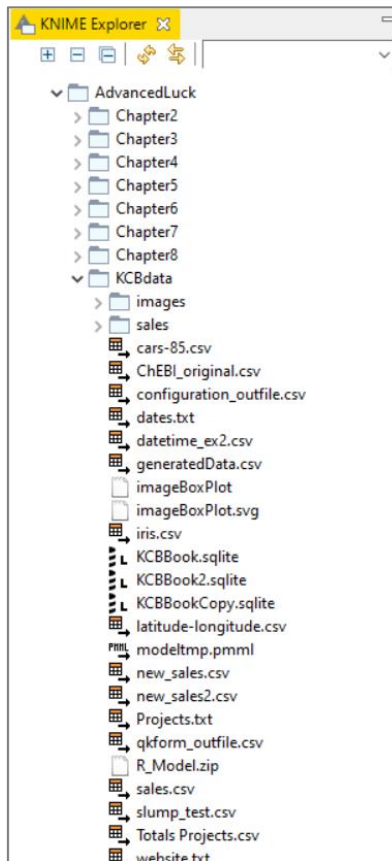
- Download the whole folder “AdvancedLuck” onto your machine, which will result in a .knar file. Then:
- Double click it OR import it into the KNIME Explorer via Select File -> Import KNIME Workflow ...

1.25. Workflows and data for this book on the KNIME Hub (hub.knime.com/rs1/spaces/Public/latest/KNIMEPress%2FKNIME_Advanced_Luck_4.5_2022011)

The screenshot shows a public space on the KNIME Hub. At the top left, it says "Public space" with a cube icon. Below that is the title "Public" in a large, bold font. To the right of the title are icons for editing (a pencil in a circle) and a user profile picture. Below the title, it says "Last edited: 26 Jun 2020". To the right of this text are icons for a heart (likes), the number "0", a link icon, and a vertical ellipsis (more options). Below this is a breadcrumb trail: "Home > KNIMEPress > KNIME_Advanced_Luck_4.5_2022011". Below the breadcrumb trail is a list of items. The first item is a folder icon, a left-pointing arrow, and the text "AdvancedLuck". To the right of the folder icon is a share icon (a cloud with an arrow pointing up).

SAMPLE

1.1. Workflows and data used in this book, as imported from the “AdvancedLuck” folder on the KNIME Hub



At the end of the import operation, in your “KNIME Explorer” panel you should find a folder named “AdvancedLuck” and containing Chapter2, Chapter3, Chapter4, etc ... subfolders, each one with workflows and exercises to be implemented in the corresponding chapters of this book. In addition, under the main folder “Advanced Luck”, you should find a KCBdata subfolder containing all necessary data.

The data used for the exercises and for the demonstrative workflows of this book were either generated by the authors or downloaded from the UCI Machine Learning Repository [2], 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 to download it. Data generated by the author, that is not public data, are located only in the KCBdata folder.

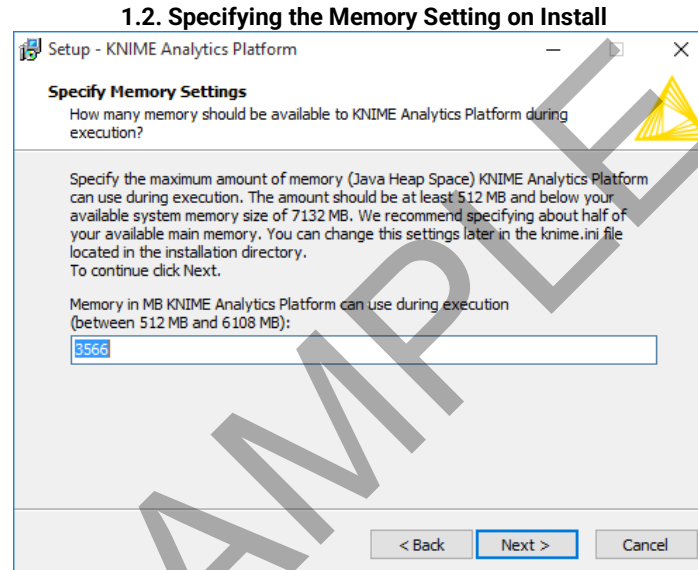
Data sets from the UCI Machine Learning Repository [2]:

- Automobile: <http://archive.ics.uci.edu/ml/datasets/Automobile>
- Slump_test: <http://archive.ics.uci.edu/ml/datasets/Concrete+Slump+Test>

This book is not meant as an exhaustive reference for KNIME Analytics Platform, although many useful workflows and aspects of it are demonstrated through worked examples. This text is intended to give you the confidence to use the advanced functions in KNIME Analytics Platform to manage and analyze your own data.

1.3. Memory Usage in KNIME Analytics Platform

The maximum amount of memory to use is set at installation time, if you install it using the Windows Installer. However, in all other installation procedures, or if you want to change the set number of MB later, you will need to set yourself the maximum amount of memory available to KNIME Analytics Platform.



The amount of memory available to KNIME Analytics Platform is stored in the knime.ini file. The knime.ini file is located in the directory in which KNIME Analytics Platform has been installed, together with the knime.exe file. The knime.ini file contains a number of required settings.

-Xmx<size> is the setting that defines the maximum heap size available to run workflows. You can define this value by editing the knime.ini file or at installation time. If you run into memory problems, you probably need to manually increase the heap space (-Xmx option) directly in the knime.ini file to a size compatible with the memory you have on your machine (like 4G for 4 Gigabytes).

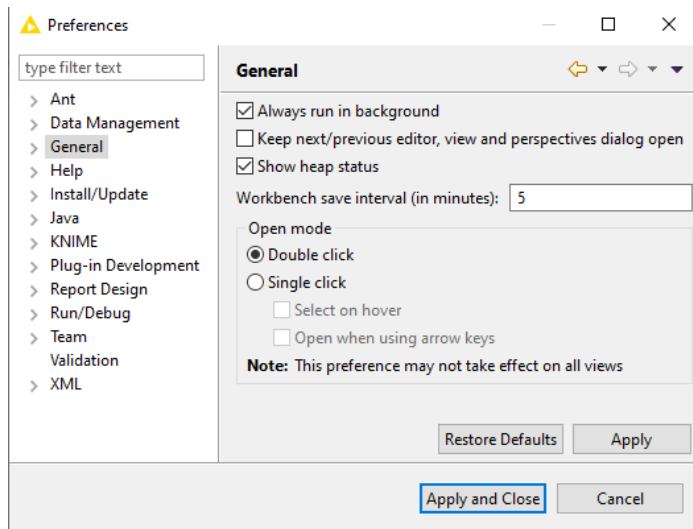
1.3. The "knime.ini" file

```
-startup
plugins/org.eclipse.equinox.launcher_1.3.100.v20150511-1540.jar
--launcher.library
plugins/org.eclipse.equinox.launcher.win32.win32.x86_64_1.1.300.v20150602-1417
-vmargs
-server
-Dsun.java2d.d3d=false
-Dosgi.classloader.lock=classname
-XX:+UnlockDiagnosticVMOptions
-XX:+UnsyncloadClass
-Dknime.enable.fastload=true
-XX:CompileCommand=exclude,javax/swing/text/GlyphView,getBreakSpot
-Xmx3566m
-Dorg.eclipse.swt.browser.IEVersion=10001
-Dsun.awt.noerasebackground=true
```

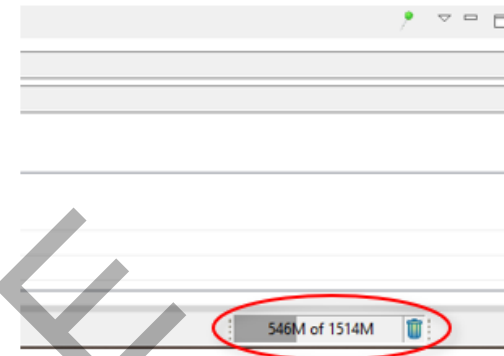
There is also an easy way to monitor how much heap space is being used by a workflow and if this reaches the maximum limit assigned by the `-Xmx` option. In the Top Menu in the KNIME workbench:

- Click "File"
- Select "Preferences"
- "Preferences" window opens
- In "Preferences" window
 - o Select "General"
 - o In the frame on the right named "General", enable the option "Show heap status"
 - o Click "OK"
 - o Now, in the lower right corner you can see a small number showing the heap status.

1.4. The "Preferences" window with the "Show heap status" option



1.5. The bottom right corner shows the heap status



To run the example workflows and the exercises provided in this book, you will need to install the whole "KNIME & Extensions" group.

In order to install a KNIME Extension:

- In the top menu of the KNIME workbench, select "File" -> "Install KNIME Extensions ..."
- In the "Install" window:
 - Open the group containing your extension, like for example "KNIME & Extensions" group
 - If you do not know where your extension package is located, just run a search by inserting a few related keywords in the top textbox
 - Select your extension
 - Click "Next" and follow installation instructions