Open for Innovation®
KNIME

KNIME User Training

KNIME.com AG

Copyright © 2017 KNIME.com AG

Overview
KNIME Analytics Platform

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ®
KNIME

What is KNIME Analytics Platform?

- A tool for data analysis, manipulation, visualization, and reporting
- Based on the graphical programming paradigm
- Provides a diverse array of extensions:
  - Text Mining
  - Network Mining
  - Cheminformatics
  - Weka machine learning
  - Many integrations, such as Java, R, Python, etc.

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Additional Resources

KNIME pages (www.knime.org)

- **SOLUTIONS** for example workflows
- RESOURCES/**LEARNING HUB** www.knime.org/learning-hub
- RESOURCES/**NODE GUIDE** https://www.knime.org/nodeguide

KNIME Tech pages (tech.knime.org)

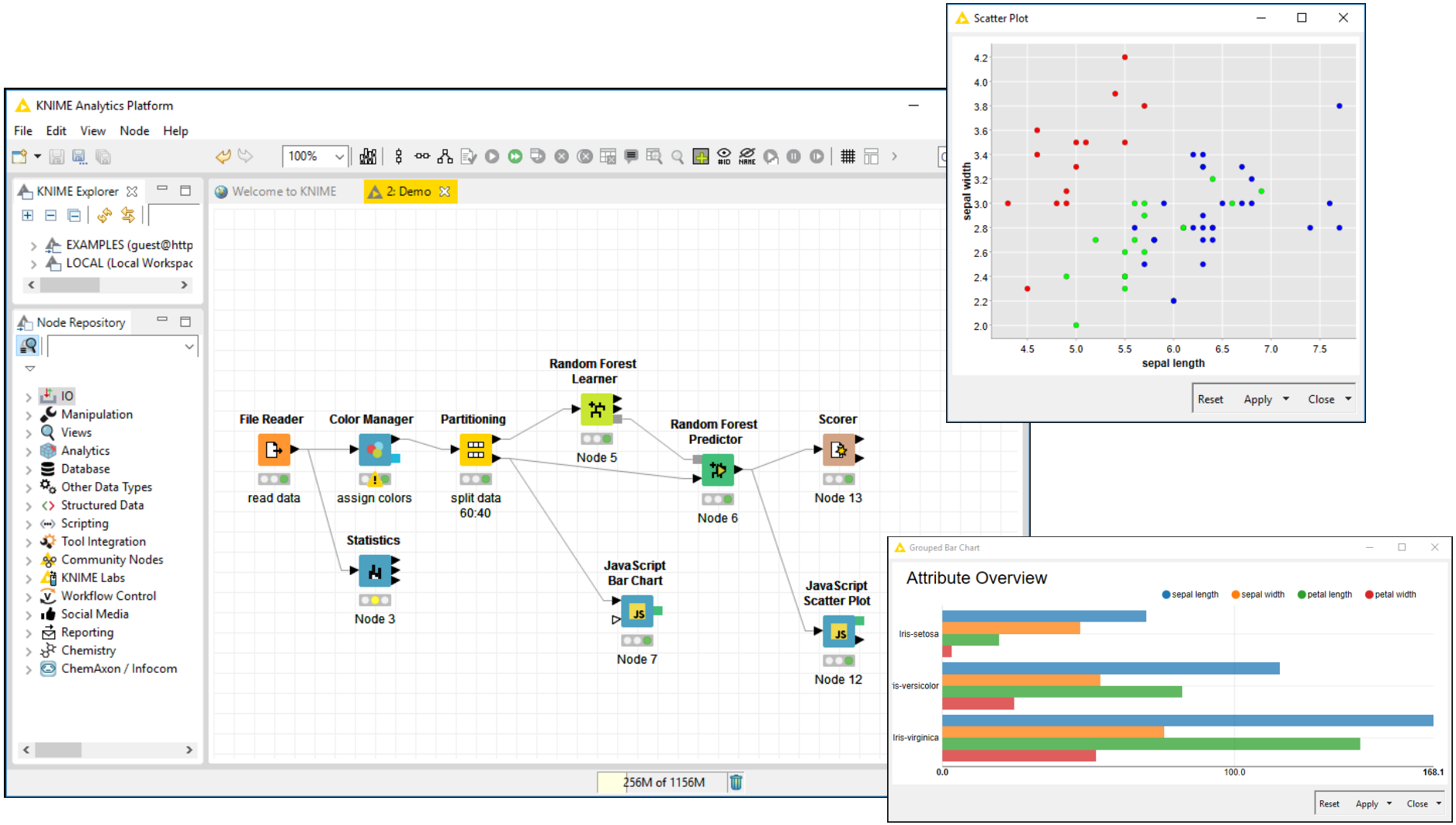
- **FORUM** for questions and answers
- **DOCUMENTATION** for docs, FAQ, changelogs, ...
- **COMMUNITY CONTRIBUTIONS** for dev instructions and third party nodes

KNIME TV on **YouTube** https://www.youtube.com/user/KNIMETV

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

The KNIME® Analytics Platform



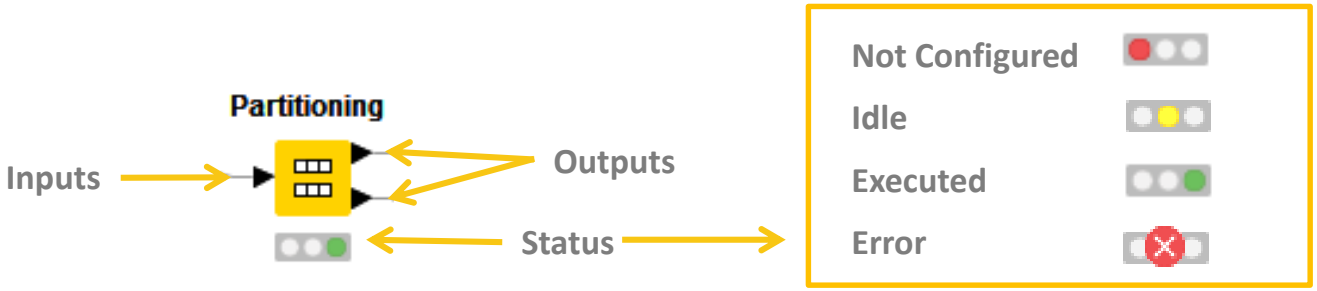
Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

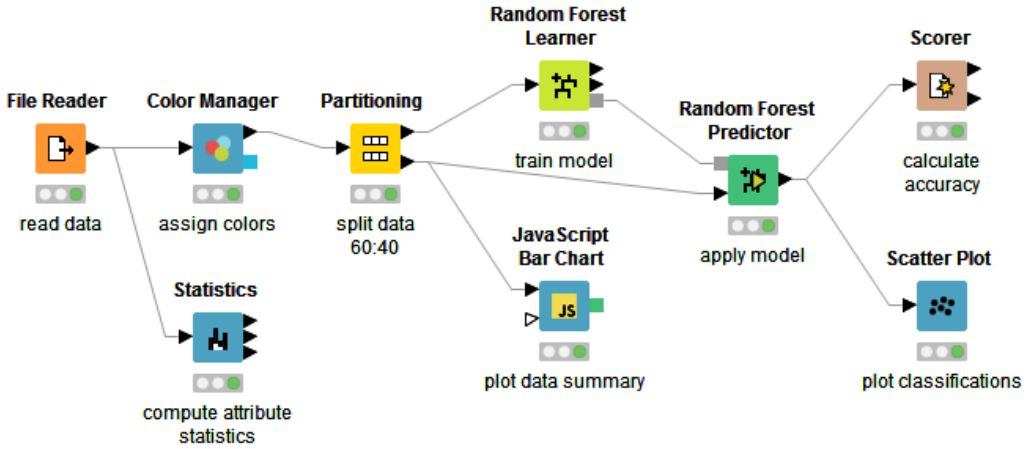
Open for Innovation ® KNIME

Visual KNIME Workflows

NODES perform tasks on data



Nodes are combined to create **WORKFLOWS**

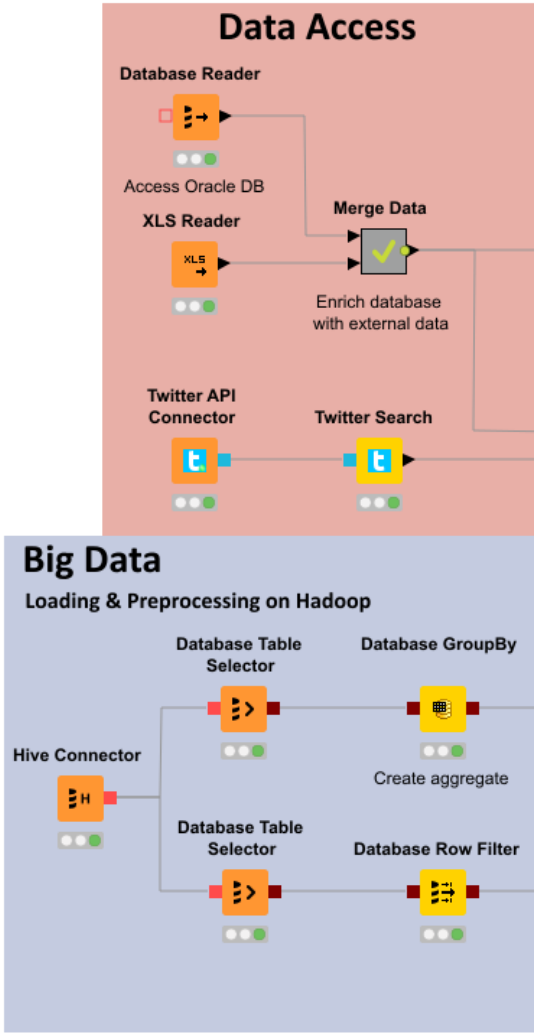


Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

Data Access

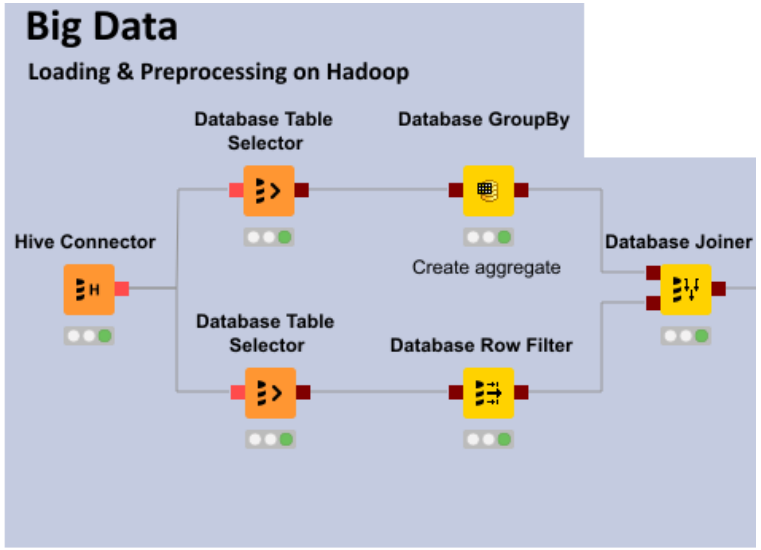


- Databases
  - MySQL, PostgreSQL
  - any JDBC (Oracle, DB2, MS SQL Server)
- Files
  - Csv, txt
  - Excel, Word, PDF
  - SAS, SPSS
  - XML
  - PMML
  - Images, texts, networks, chem
- Web, Cloud
  - REST, Web services
  - Twitter, Google

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

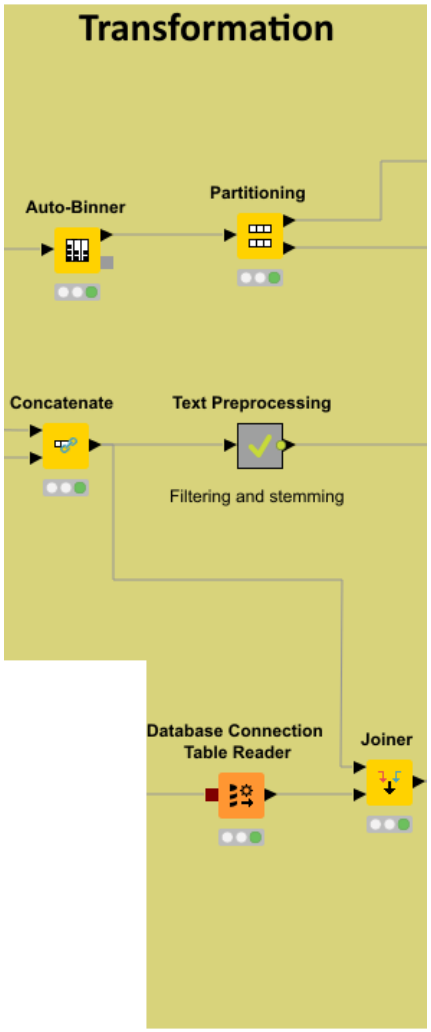
Big Data



- Spark
- HDFS support
- Hive
- Impala
- HP Vertica
- In-database processing

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Transformation

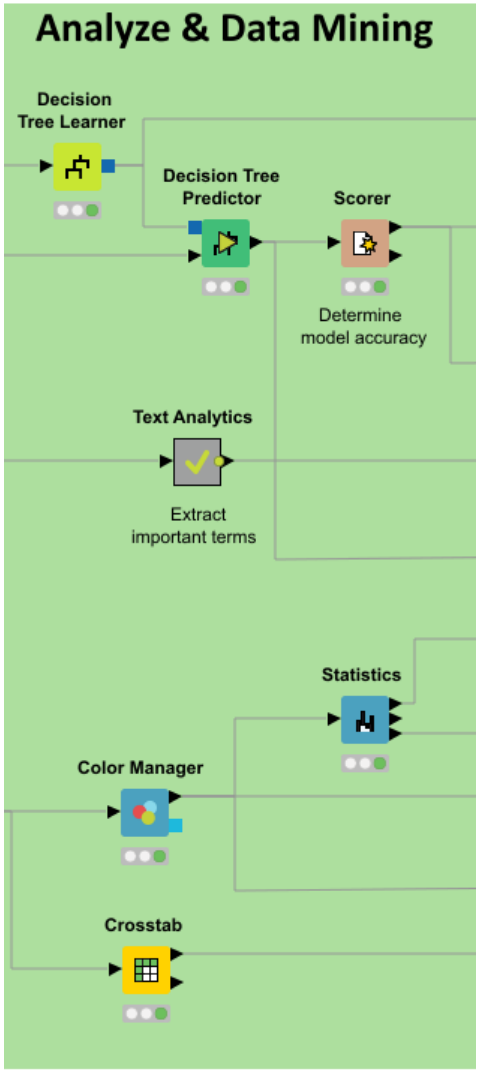


- Preprocessing
  - Row, column, matrix based
- Data blending
  - Join, concatenate, append
- Aggregation
  - Grouping, pivoting, binning
- Feature Creation and Selection

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Analyze & Data Mining

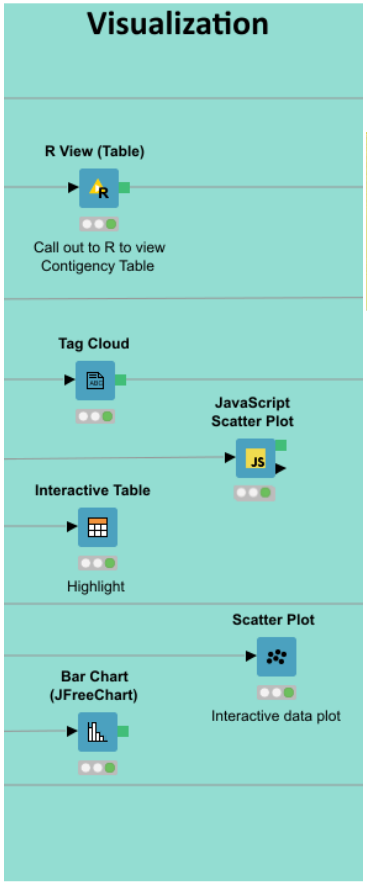


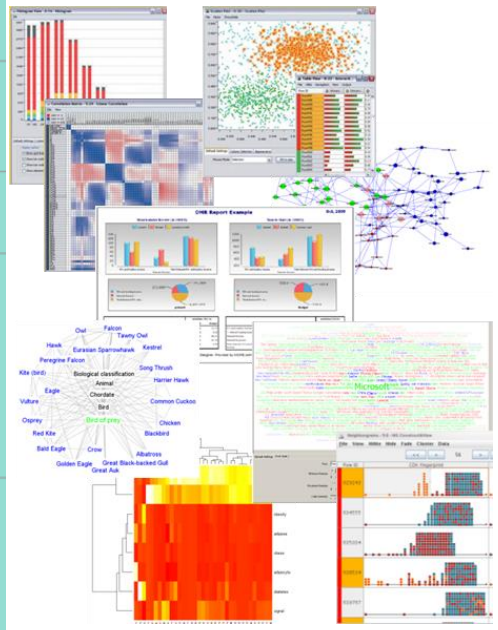
- Regression
  - Linear, logistic
- Classification
  - Decision tree, ensembles, SVM, MLP, Naïve Bayes
- Clustering
  - k-means, DBSCAN, hierarchical
- Validation
  - Cross-validation, scoring, ROC
- Misc
  - PCA, MDS, item set mining
- External
  - R, Weka

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Visualization



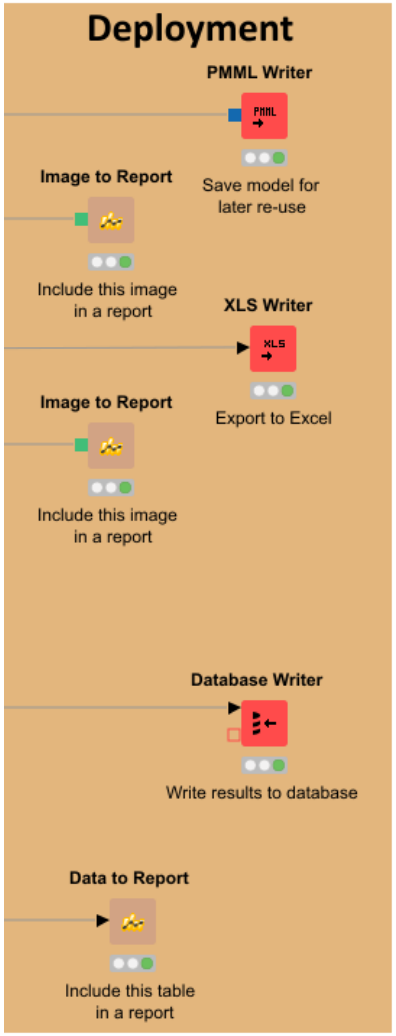


- Interactive
  - Scatter plot, histogram, pie charts, box plot
  - Highlighting (brushing)
- JFreeChart
- JavaScript
- Misc
  - Tag cloud, open street map, networks, molecules
- External
  - R

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Deployment

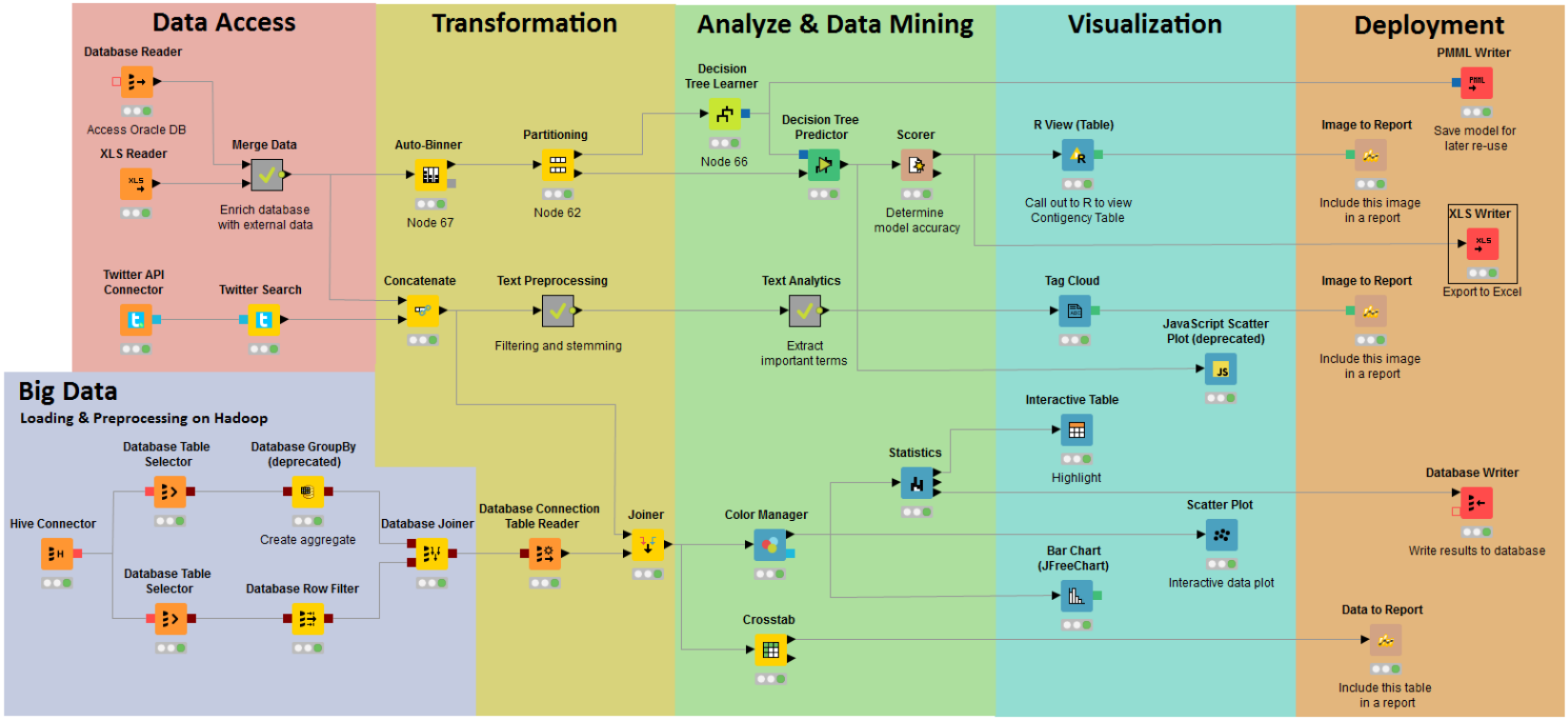


- Database
- Files
  - Excel, csv, txt
  - XML
  - PMML
  - to: local, KNIME Server, SSH-, FTP-Server
- BIRT Reporting

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Over 1500 native and embedded nodes included:



Data Access
MySQL, Oracle, ...
SAS, SPSS, ...
Excel, Flat, ...
Hive, Impala, ...
XML, JSON, PMML
Text, Doc, Image, ...
Web Crawlers
Industry Specific
Community / 3rd

Transformation
Row,
Column
Matrix
Text, Image
Time Series
Java
Python
Community / 3rd

Analysis & Mining
Statistics
Data Mining
Machine Learning
Web Analytics
Text Mining
Network Analysis
Social Media Analysis
R, Weka, Python
Community / 3rd

Visualization
R
JFreeChart
JavaScript
Community / 3rd

Deployment
via BIRT
PMML
XML, JSON
Databases
Excel, Flat, etc.
Text, Doc, Image
Industry Specific
Community / 3rd

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Overview

- Installing KNIME Analytics Platform
- The KNIME Workspace
- The KNIME File Extensions
- The KNIME Workbench
  - Workflow editor
  - Explorer
  - Node repository
  - Node description
- Installing new features

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Install KNIME Analytics Platform

- Select the KNIME version for your computer:
  - Mac, Win, or Linux and 32 / 64bit
- Note different downloads (minimal or **full**)
- Download archive and extract the file, or download installer package and run it

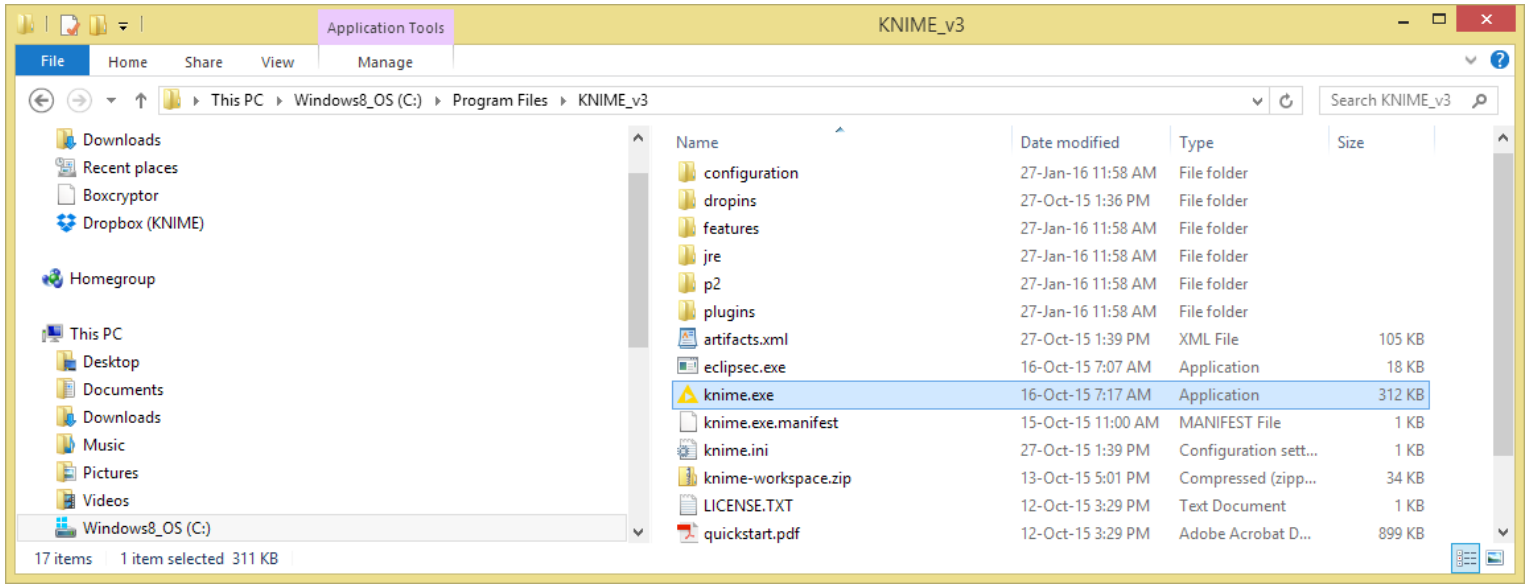
Windows	
KNIME Analytics Platform for Windows (installer) *The installer adds an icon to the desktop and suggests suitable memory settings*	32 bit *(301,82 MB)* 64 bit *(305 MB)*
KNIME Analytics Platform + all free extensions for Windows (installer) *The installer adds an icon to the desktop and suggests suitable memory settings*	32 bit *(1,74 GB)* 64 bit *(1,92 GB)*
KNIME Analytics Platform for Windows (self-extracting archive) *The self-extracting archive only creates a folder holding the KNIME installation*	32 bit *(304,18 MB)* 64 bit *(306,45 MB)*
KNIME Analytics Platform for Windows (zip archive)	32 bit *(346,61 MB)* 64 bit *(349,93 MB)*
Linux	
KNIME Analytics Platform for Linux	32 bit *(363,02 MB)* 64 bit *(360,1 MB)*
KNIME Analytics Platform + all free extensions for Linux	32 bit *(1,66 GB)* 64 bit *(2,07 GB)*
Mac OSX	
KNIME Analytics Platform for Mac OSX (10.7 and above)	64 bit *(329,41 MB)*
KNIME Analytics Platform + all free extensions for Mac OSX (10.7 and above)	64 bit *(1,99 GB)*

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Start KNIME Analytics Platform

- Go to the installation directory and launch **KNIME**, or use the shortcut created on your Desktop.

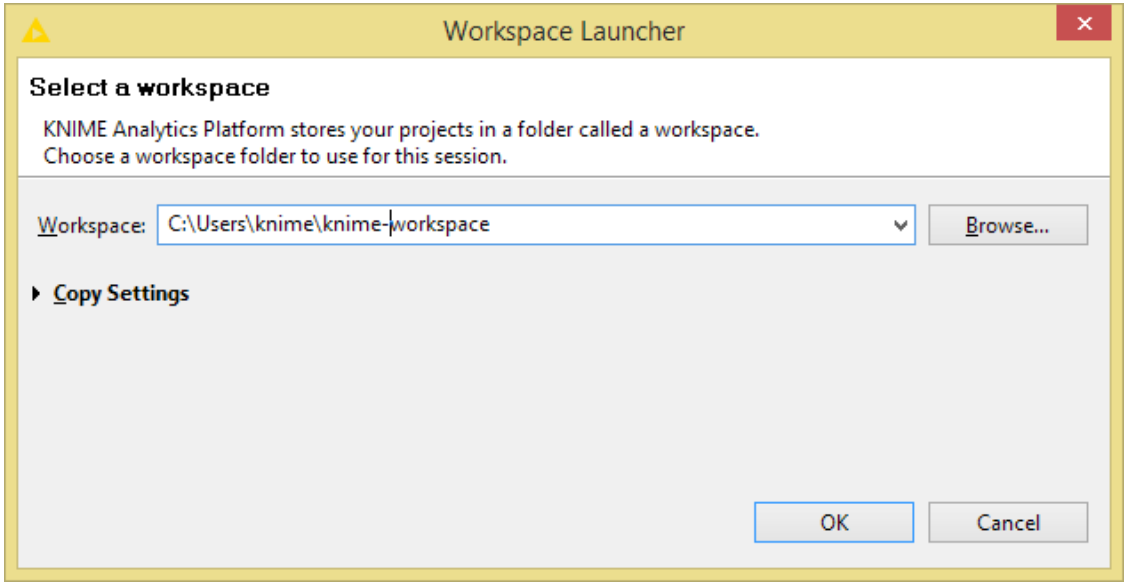


Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

The KNIME Workspace

- The workspace is the **folder/directory** in which workflows (and potentially data files) are stored for the current KNIME session.
- Workspaces are portable (just like KNIME)



Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Welcome Page

Open for Innovation KNIME

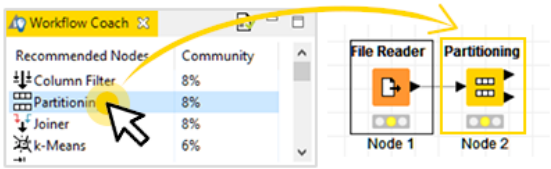
Welcome to KNIME Analytics Platform!

New to KNIME? Looking for resources to get started?

- Register for emails with introductory tips here.
- Explore our Quickstart Guide.
- Check out 7 things to do after installing KNIME Analytics Platform
- Find more hints and how-tos in the Learning Hub.
- And register for our release and event emails right here.

This page will be displayed upon startup but you can customize the content using the checkboxes at the bottom.

NEW since 3.2: Workflow Coach recommends matching nodes.



Updates for the following components are available:

- **DYMATRIX Uplift Modeling Extensions**
- **Palladian for KNIME**

Click here in order to install updates.

Where to go from here

- Create new workflow
- Learning Hub
- Browse example workflows
- Get additional nodes
- Go to my workflows
- Mount KNIME Cloud Server

Most recently used workflows

- ModelSelection_WebPortal_Part1
- ModelSelection_WebPortal_Part1
- ModelSelection_BasicWorkflow
- DataCleaning_WebPortal_v2.0
- KNIME_project2
- Sexy ETL_v2.0

Tips & Tricks

Specialist Nodes

Did you know there are a whole variety of specialist nodes available from KNIME Labs and the Community around Scripting, Image Processing, Text Processing, Internet Mining, Network Mining, Cell Biology and Genetics, and Chemistry. To access them, go to Help Menu and choose Install New< Software.

- [x] Show intro text at next start
- [x] Show update notifications at next start
- [x] Show links and most recently used workflows at next start

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation KNIME

The KNIME Workbench

KNIME Analytics Platform - /Users/rb/knime-workspace

KNIME Explorer

EXAMPLES (knime-guest@http:
LOCAL (Local Workspace)
BigData Training
KNIMEUserTraining
data
exercises
solutions
01. Importing Data - s
02. Data Manipulation
03. Visualization - solu
04. Data Mining - solut
05. Flow Variables - sc
06. Exporting Data - s

Servers and Workflows

0: 04. Data Mining - solution

- Evaluate the quality of a model with the Scorer node.
- What was the overall accuracy of your model?
- Optional: Evaluate Receiver Operation Characteristics with the ROC Curve node.

Workflow Editor

Fully Joined Data
Partitioning
Decision Tree Learner
Decision Tree Predictor
Scorer
JavaScript ROC Curve
Node 278
ROC Curve

Workflow Coach

Recommended Nodes
Decision Tree Predictor
Decision Tree Learner
Naive Bayes Predictor
Naive Bayes Learner
Regression Predictor

Node Recommendations

Node Repository

IO
Manipulation
Views
Analytics
Database
Other Data Types
Structured Data
Scripting
Tool Integration
Community Nodes
KNIME Labs

Node Repository

Node Description

Partitioning

The input table is split into two partitions (i.e. row-wise), e.g. train and test data. The two partitions are available at the two output ports. The following options are available in the dialog:

Dialog Options

Absolute

Specify the absolute number of rows in the first partition. If there are less rows than specified here, all rows are entered into the first table, while the second table contains no rows.

Relative

The percentage of the number of rows in the input table that are in the first partition. It must be between 0 and 100, inclusively.

Take from top

This mode puts the top-most rows into the first output table and the remainder in the second table.

Linear sampling

Node Description

Outline

Outline

Console

KNIME Console

```
*****************************************************************************
***        Welcome to the KNIME Analytics Platform v3.3.1.v201612192115
***                  Copyright by KNIME GmbH, Konstanz, Germany
*****************************************************************************
Log file is located at: /Users/rb/knime-workspace/.metadata/knime/knime.l
WARN  LicenseStore                    License for 'KNIME Spark Executor'
WARN  LicenseStore                    License for 'KNIME Big Data Connect
```

Console

347M of 812M

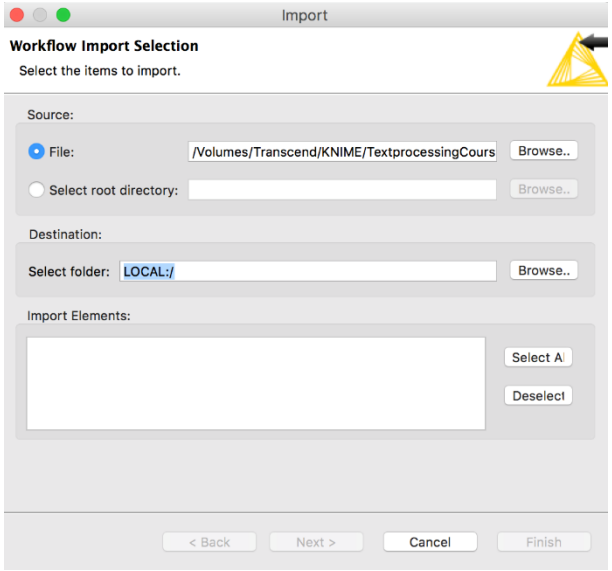
Copyright © 2017 KNIME.com AG

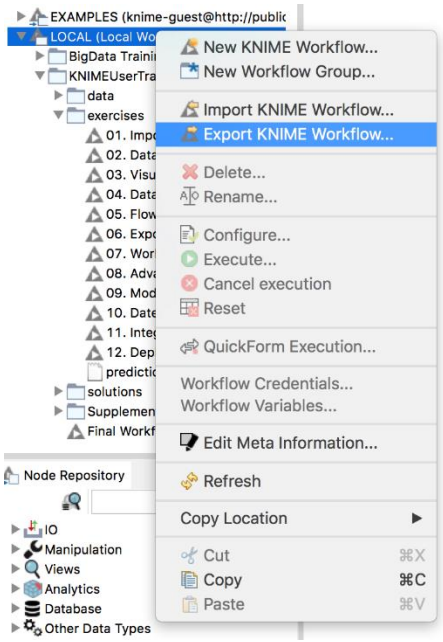
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

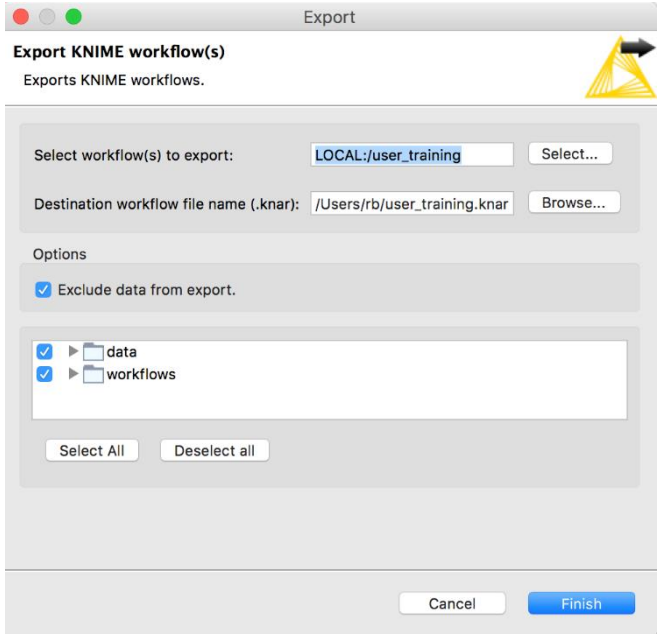
Open for Innovation ® KNIME

Creating New Workflows, Importing and Exporting

- Right-click Workspace in KNIME Explorer to create new workflow or workflow group or to import workflow
- Right-click on workflow or workflow group to export







Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

KNIME File Extensions

- Dedicated file extensions for Workflows and Workflow groups associated with KNIME Analytics Platform

- ***.knwf** for KNIME Workflow Files

Workflow_1.knwf

- ***.knar** for KNIME Archive Files

Group_wf_1.knar

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

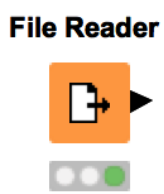
Open for Innovation ® KNIME

More on Nodes…

A node can have 3 states:

Idle:
The node is not yet configured and cannot be executed with its current settings.

Configured:
The node has been set up correctly, and may be executed at any time



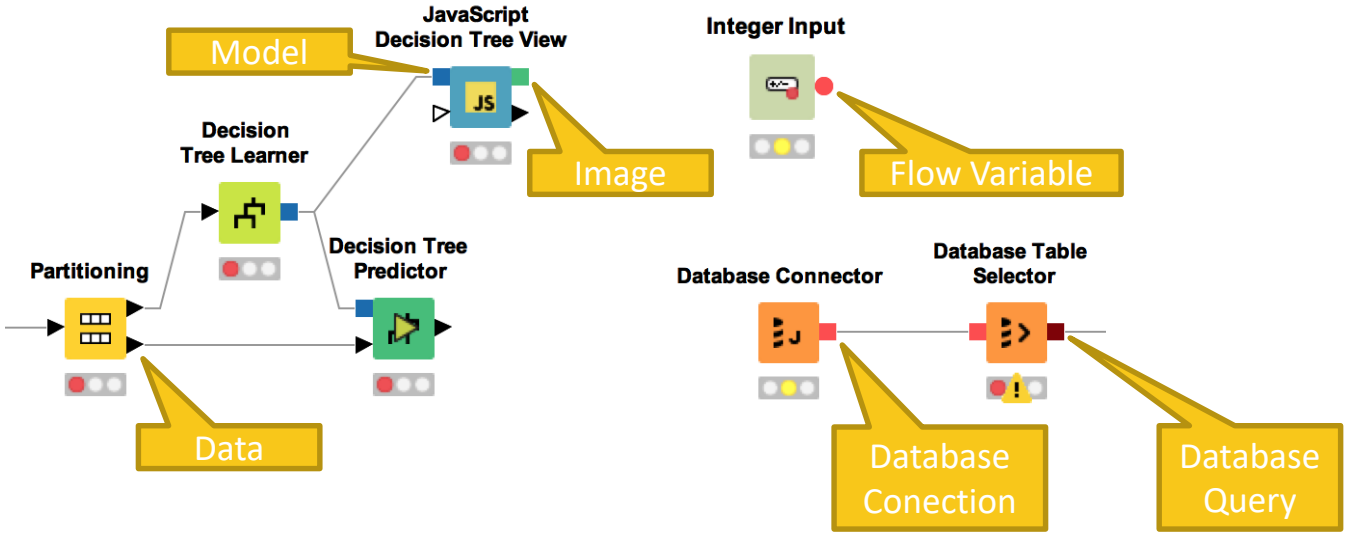
Executed:
The node has been successfully executed. Results may be viewed and used in downstream nodes.

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Inserting and Connecting Nodes

- Insert nodes into workspace by dragging them from Node Repository or by double-clicking in Node Repository
- Connect nodes by left-clicking output port of Node A and dragging the cursor to (matching) input port of Node B
- Common port types:

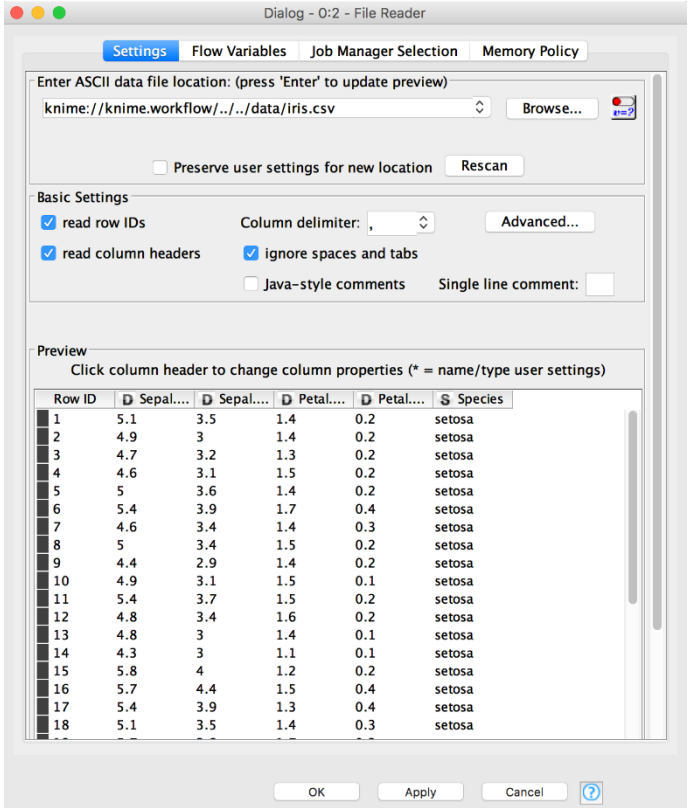


Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Node Configuration

- Most nodes require configuration
- To access a node configuration window:
  - Double-click the node
  - Right-click > Configure

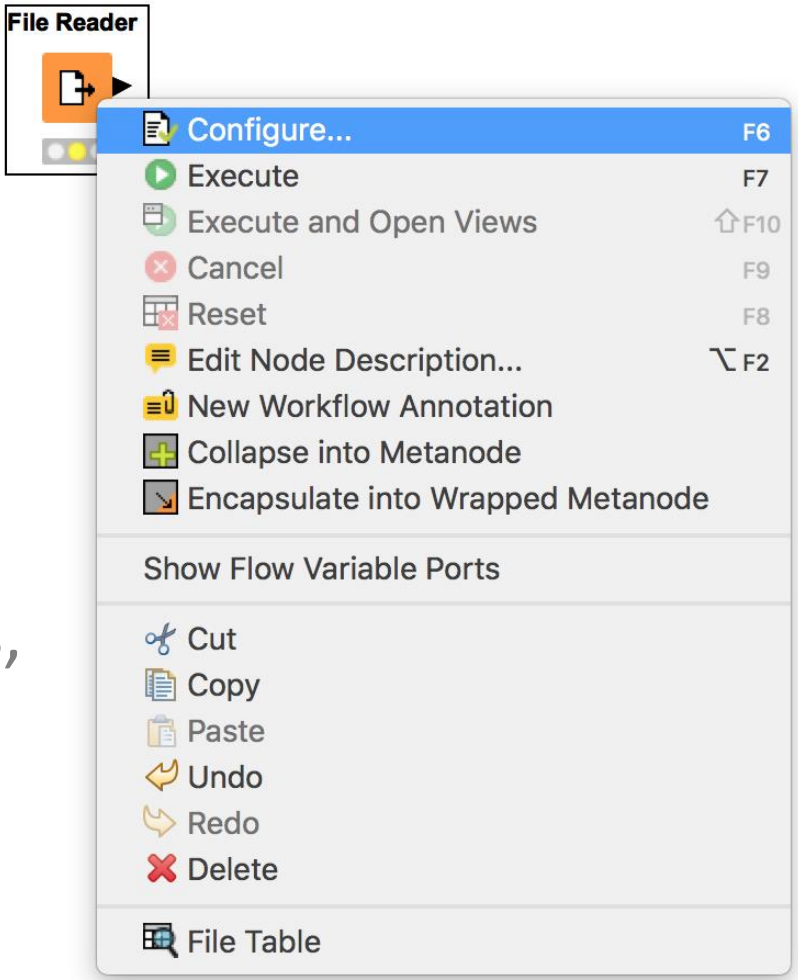


Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Node Execution

- Right-click node
- Select Execute in context menu
- If execution is successful, status shows green light
- If execution encounters errors, status shows red light

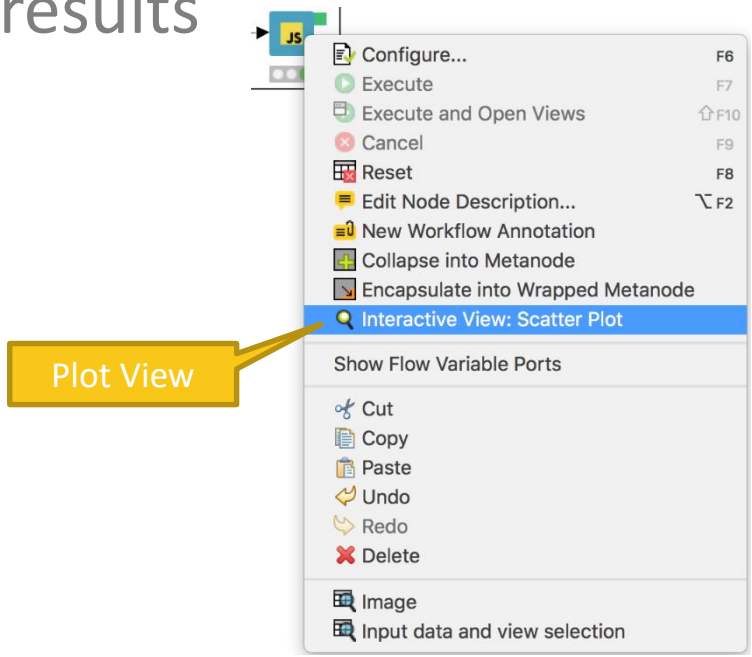


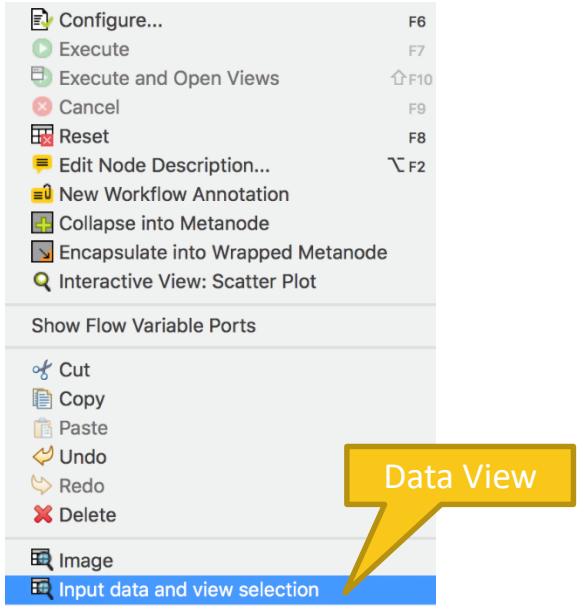
Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Node Views

- Right-click node
- Select Views in context menu
- Select output port to inspect execution results



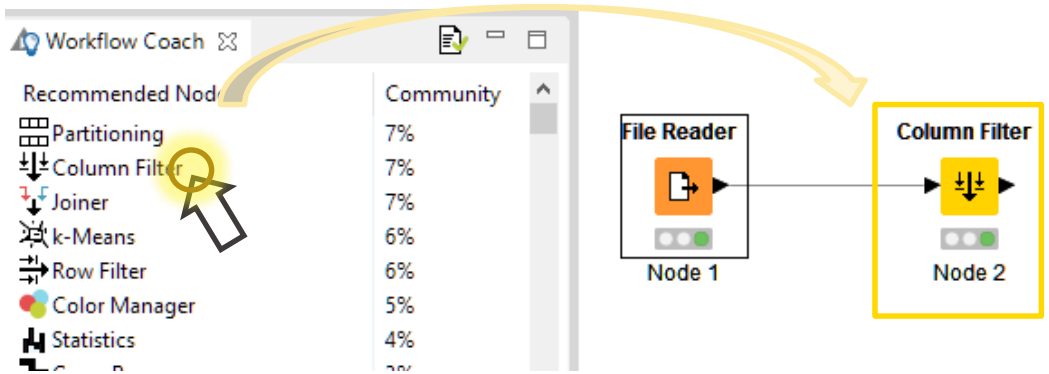


Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Workflow Coach

- Recommendation engine
  - It gives hints about which node use next in the workflow
  - Based on KNIME communities' usage statistics
  - Usage statistics available also with Personal Productivity Extension and KNIME Server products (these products require a purchased license)

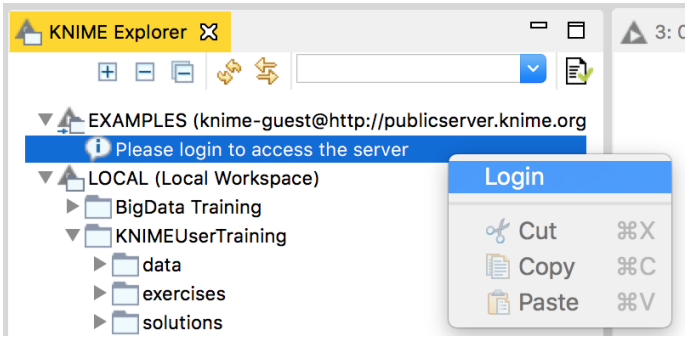


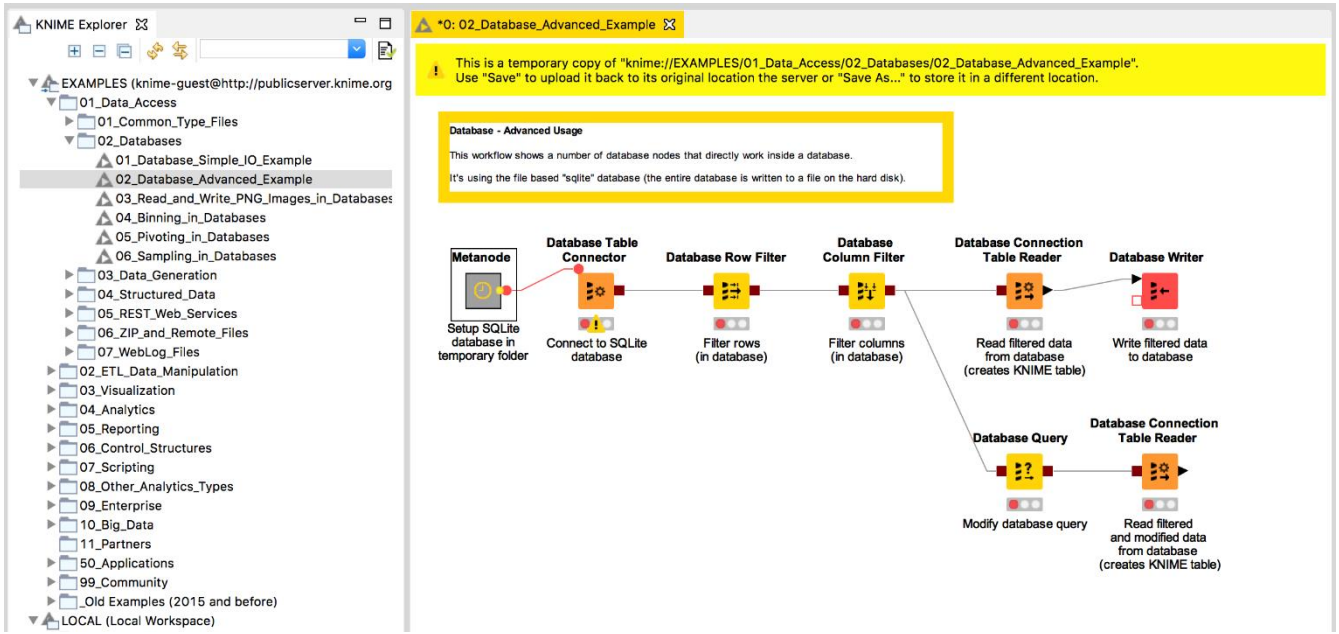
Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Getting Started: KNIME Example Server

- Public repository with large selection of example workflows for many, many applications
- Connect via KNIME Explorer





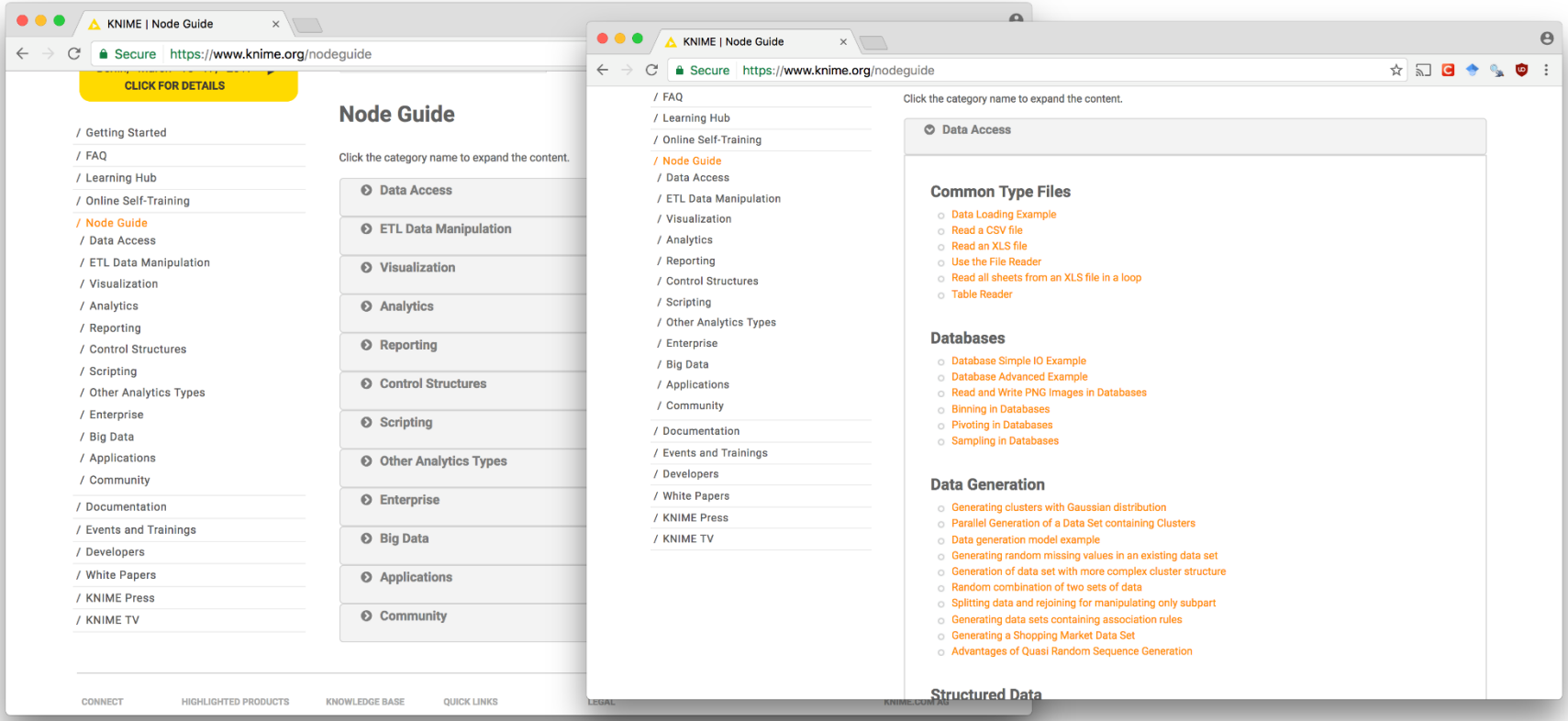
Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

Online Node Guide

- Workflows from Example Server also available online
  - https://www.knime.org/nodeguide



Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Hot Keys (for future reference)

Task	Hot key	Description
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
	F9	cancels selected running nodes
	Shift + F9	cancels all running nodes
Move Nodes and Annotations	Ctrl + Shift + Arrow	moves the selected node in the arrow direction
	Ctrl + Shift + PgUp/PgDown	moves the selected annotation in the front or in the back of all overlapping annotations
Workflow Operations	F8	resets selected nodes
	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

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Today's Example: Next Best Offer (NBO)

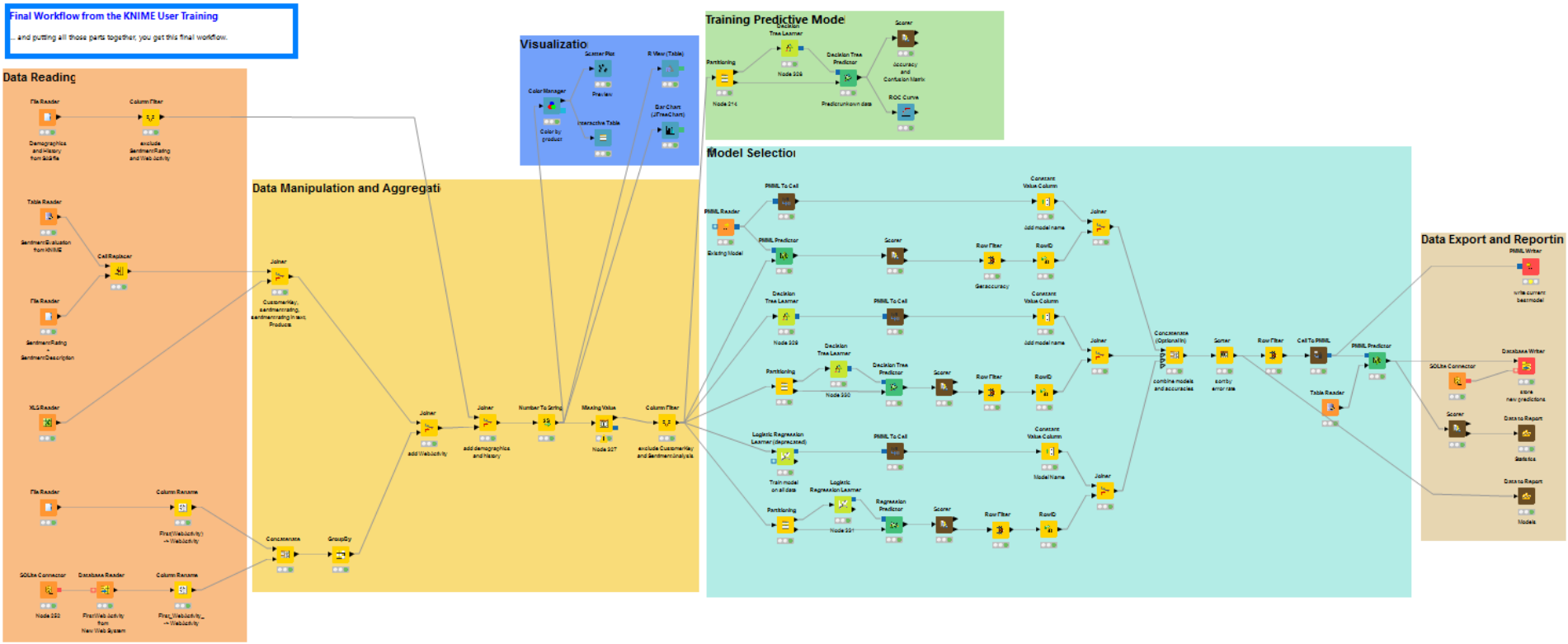
- Traditional Direct Marketing advertises a single product to a specific audience. The Next Best Offer (NBO) approach focuses on taking existing customers (and their data) and using upsell models to find interesting new products for them.

- Today we construct a workflow that joins diverse data sources into a set of complete customer records. Using this, we will build and deploy a predictive model to find people who might be interested in a newly available product.

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Today's Example: Next Best Offer (NBO)



Explore the final workflow

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

Importing Data

Accessing files and databases

Copyright © 2017 KNIME.com AG

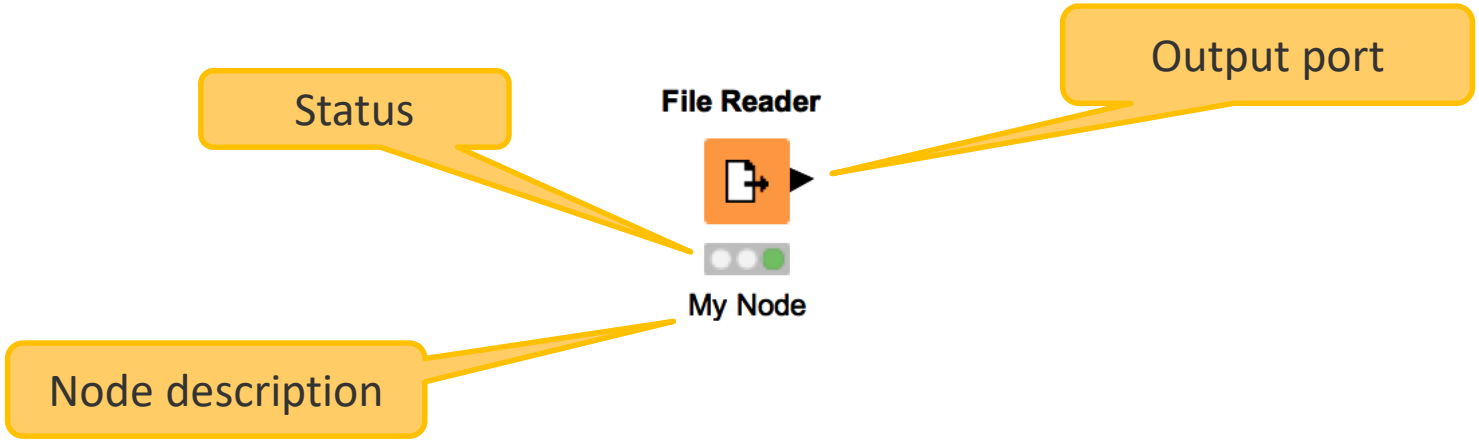
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ®
KNIME

Data Source Nodes

Typically characterized by:

- Orange color
- No input ports, 1-2 output ports



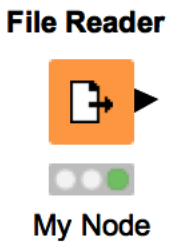
Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

New Node: File Reader

Workhorse of the KNIME Source nodes

- Reads text based files
- Many advanced features allow it to read most ‘weird’ files
  - Short lines, inline comments, headers and special encoding

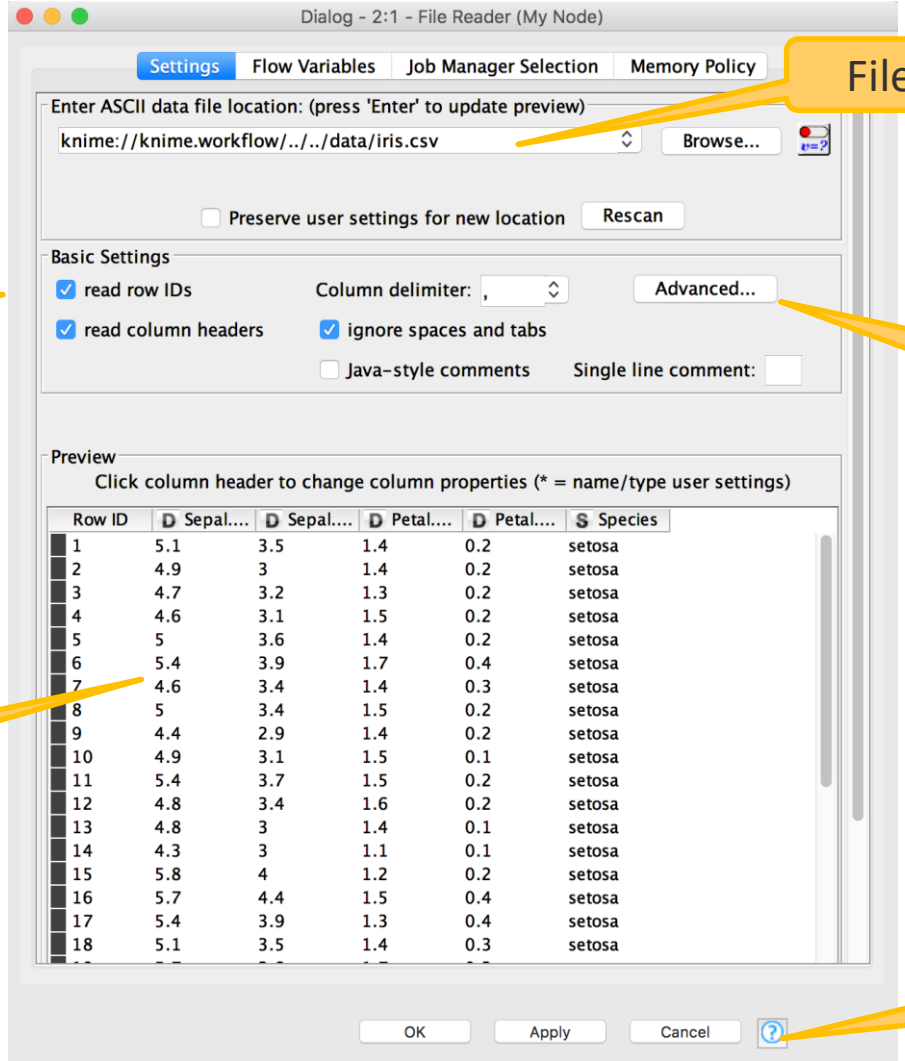


YouTube KNIME TV Channel video:
https://youtu.be/flaHQw-Qhlg

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

File Reader Configuration

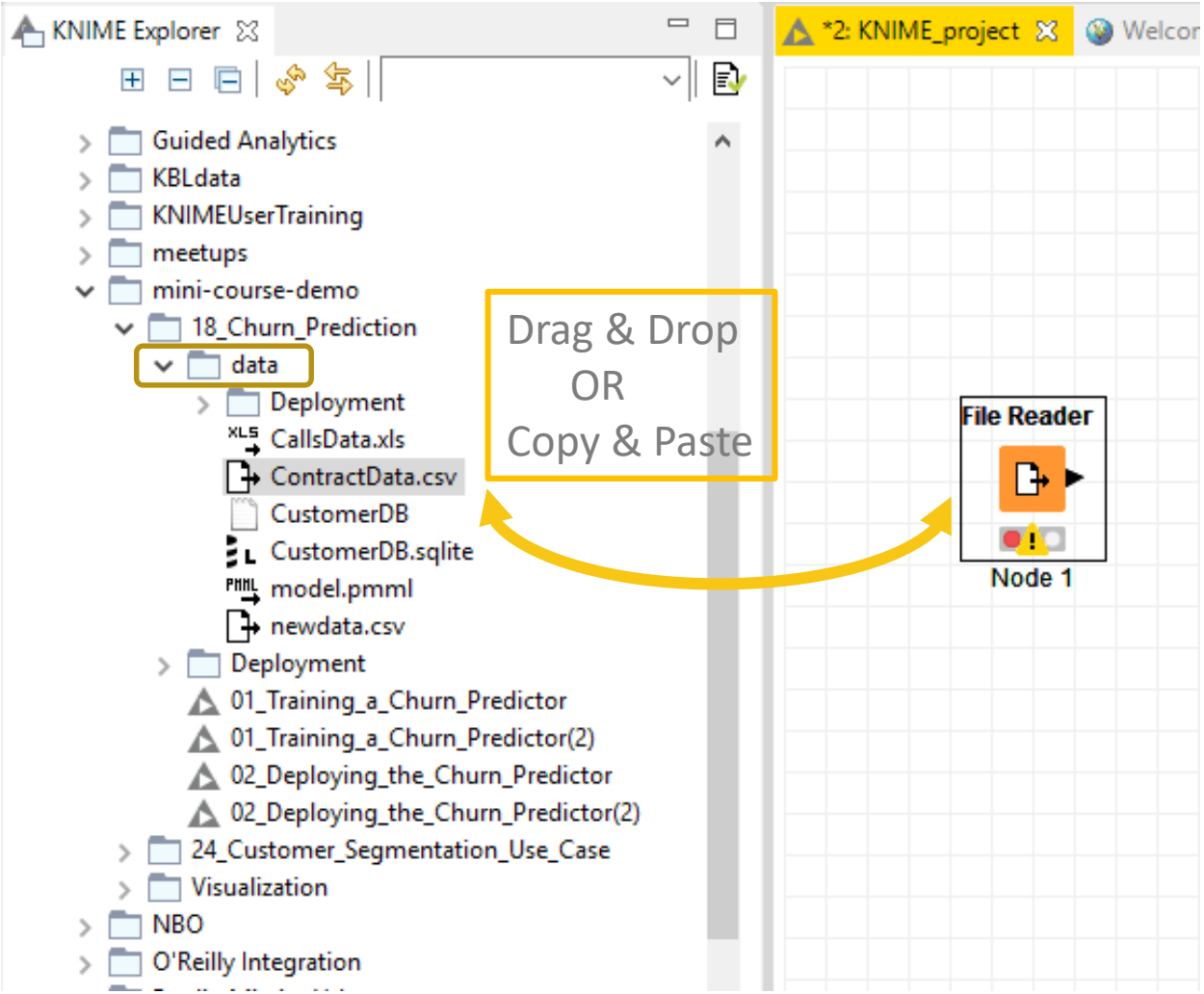


Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

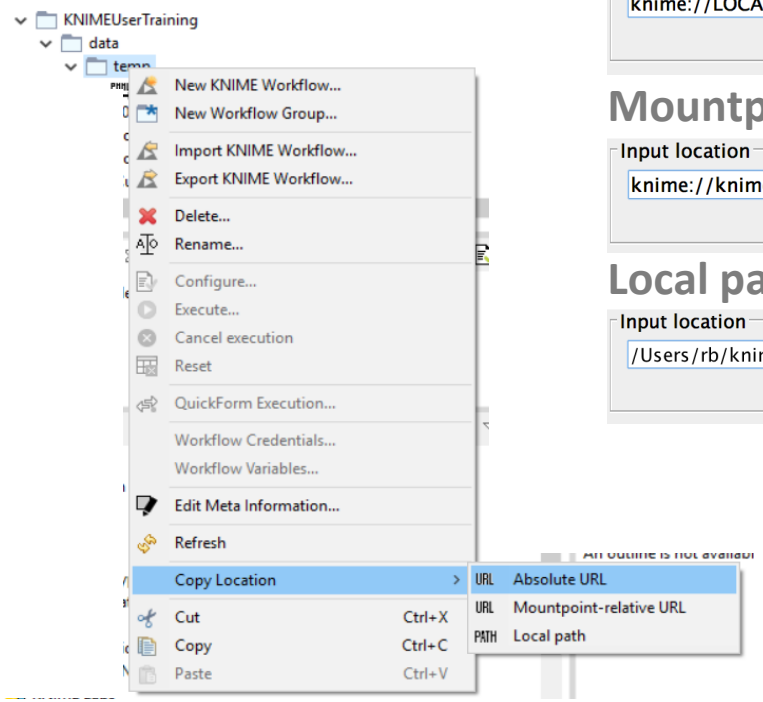
Open for Innovation ® KNIME

Alternative Faster Way …

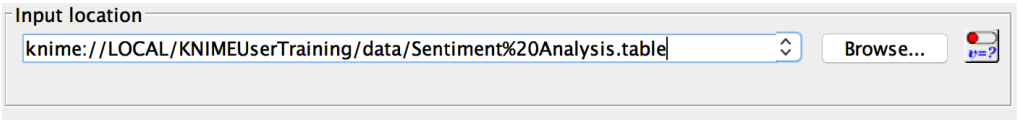


Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Filenames and the knime:// protocol



Absolute URL



Mountpoint-relative URL

Local path

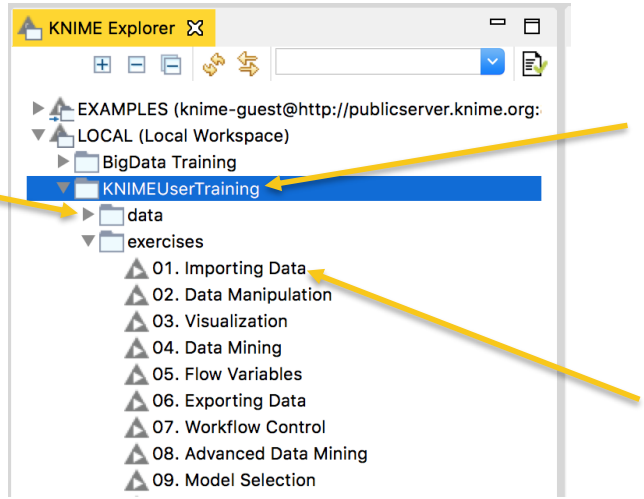
Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

Workflow Relative File Paths

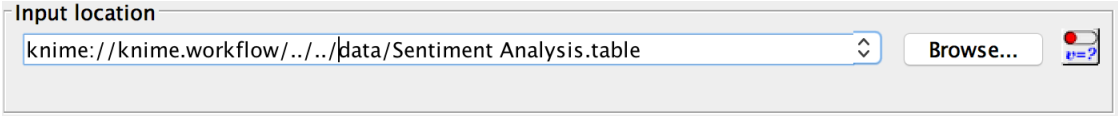
- Best choice if workflows are to be shared
- Requires matching folder structure within workflow group
  - Independent of environment outside of workflow group



Example: Path to „Sentiment Analysis.table“

- Local path:

  C:\Users\rb\knime-workspace\KNIMEUserTraining\data\Sentiment Analysis.table
- Workflow relative:



YouTube KNIME TV Channel:
https://youtu.be/U9sP4g4yGwY

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

New Node: Excel Reader (XLS)

- Reads .xls and .xlsx file from Microsoft Excel
  - Supports reading from multiple sheets

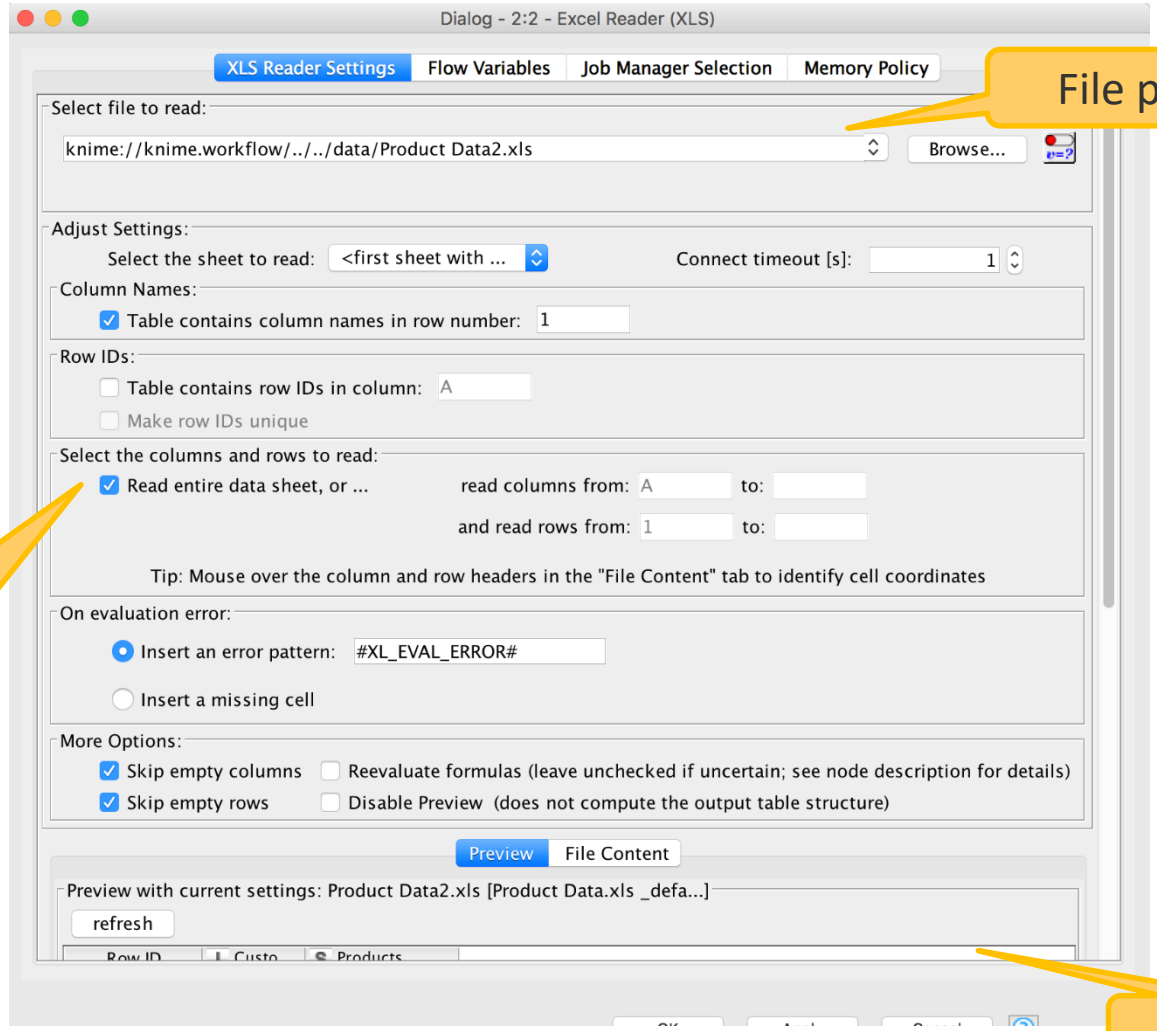
Excel Reader (XLS)

Read Excel Sheet Names (XLS)

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Excel Reader Configuration



File path

Sheet specific settings

Preview

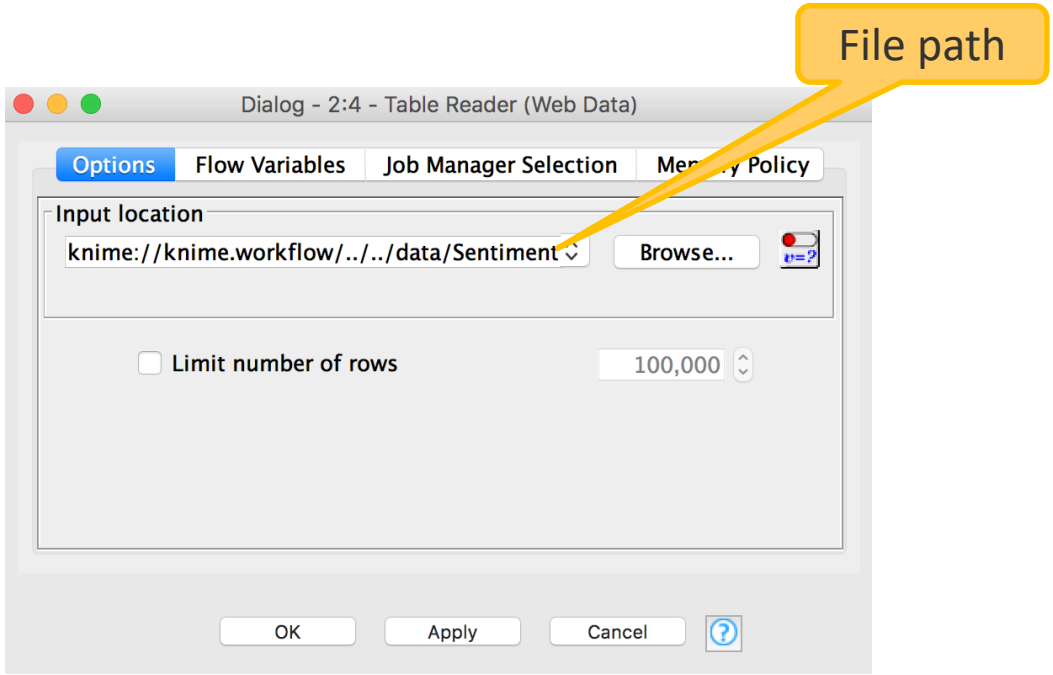
Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

New Node: Table Reader

- Reads tables from the native KNIME Format.
  - Maximum performance, minimum configuration



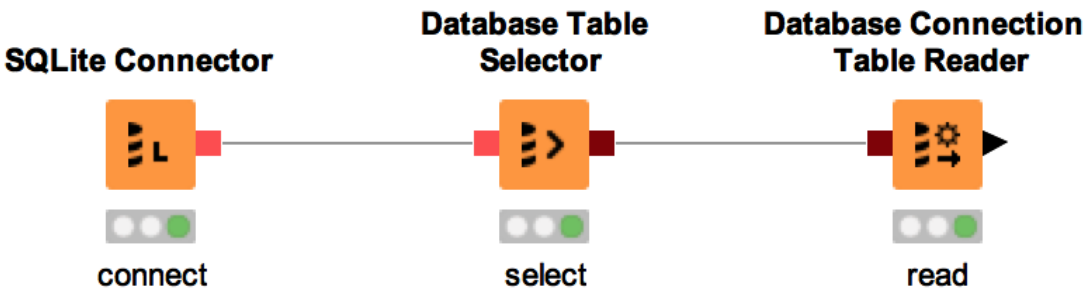
YouTube KNIME TV channel video:
https://youtu.be/tid1qi2HAOo

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Database Connectivity

- Read data from any JDBC enabled database
- Write your own SQL or model it using dedicated nodes

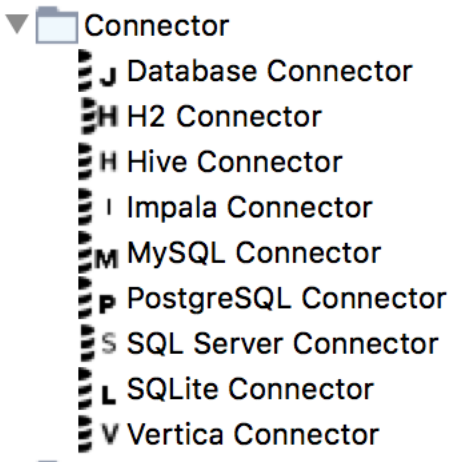


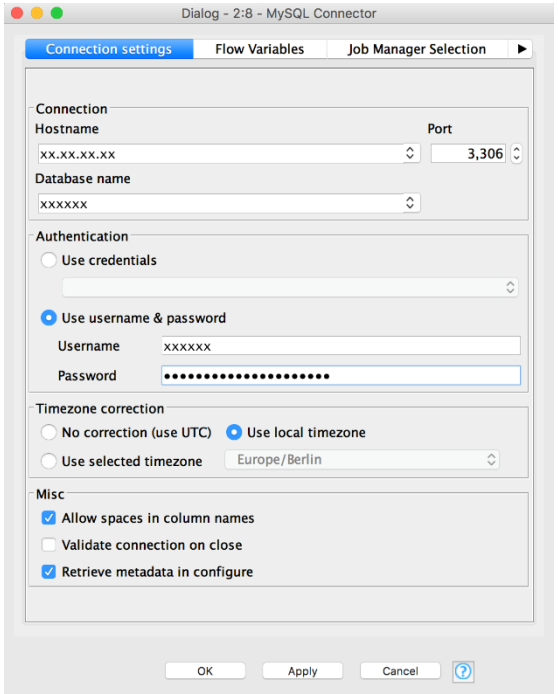
Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

New Nodes: Database Connectors

- Native: Postgres, MySQL, SQLite, MSSQL (SQL Server)
- Database Connector (e.g. Oracle, DB2, HANA).
- Commercial extensions: HIVE and Impala





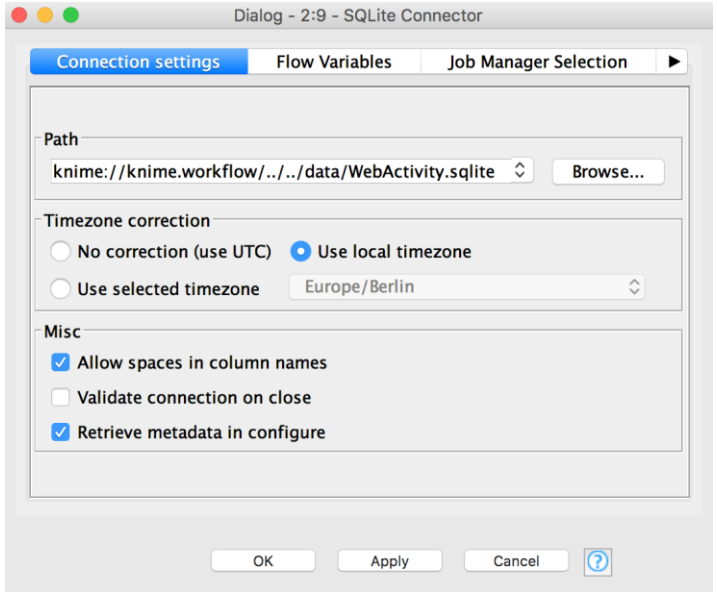
Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

New Nodes: SQLite

- Propagate connection information to other DB nodes
- File-based database
- Useful for prototyping (switch to real connector on deployment)

SQLite Connector

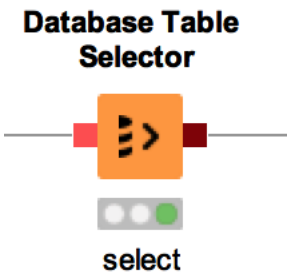


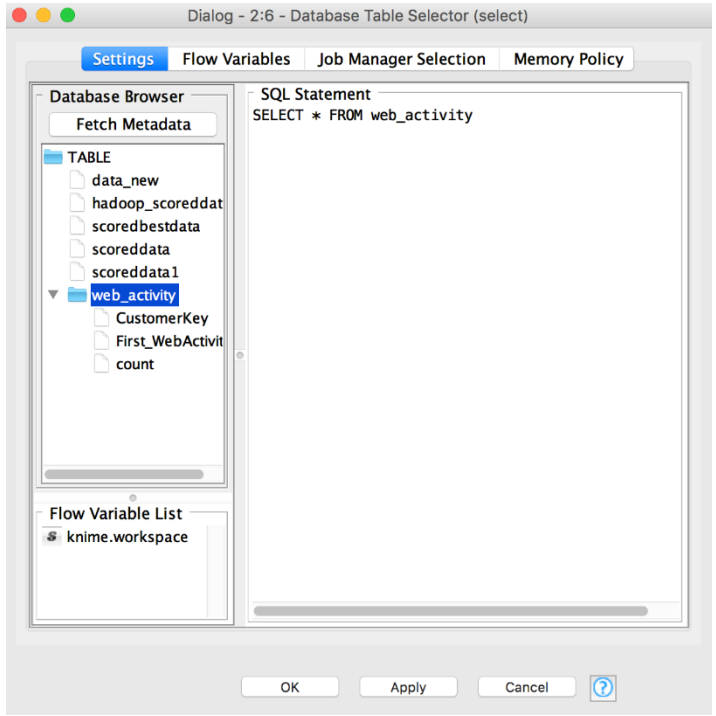
Copyright © 2017 KNIME.com AG

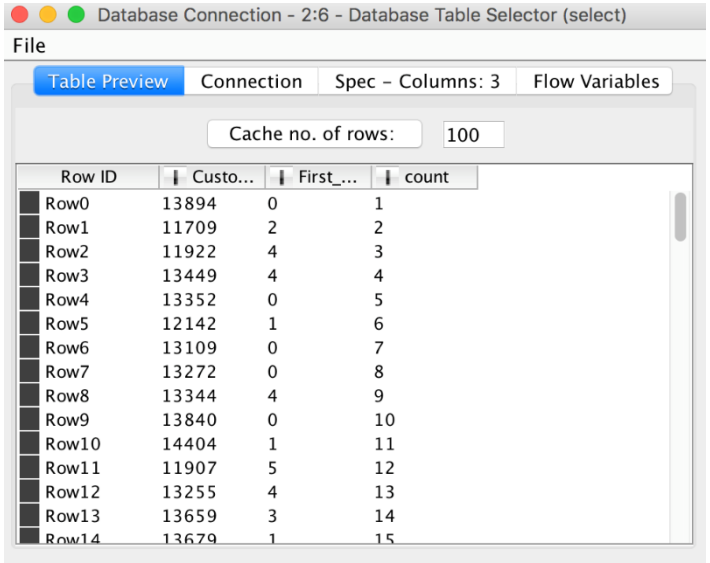
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

New Node: Database Table Selector

- Take connection information and construct a query
- Explore DB metadata
- Output is an SQL query





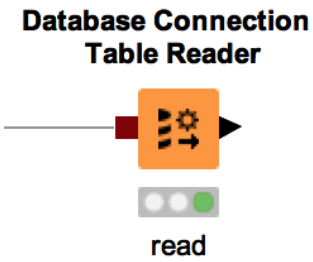


Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

New Node: Database Connection Table Reader

- Executes SQL Query
- Reads results into a KNIME table



Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Other Useful Data Sources

- PMML Reader – reads standard predictive models
- XML Reader with XPATH support
- Python/R Source nodes
- SAS7BDAT (Labs)
- REST/SOAP for web services, and many more

XML Reader

with XPath

SDF Reader

chemistry

CSV Reader

less magic

R Source (Table)

foreign library

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

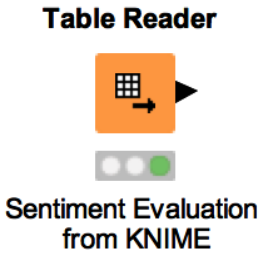
Open for Innovation ® KNIME

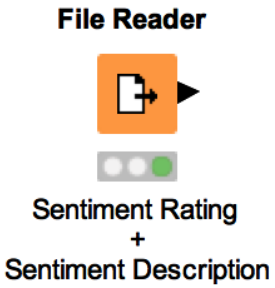
Importing Data Exercise

Starting with exercise: *Importing Data*

Read the following files

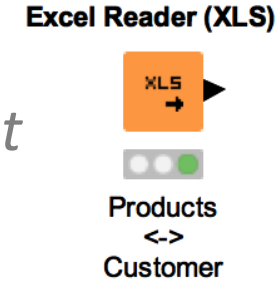
- Sentiment Analysis.table
- Sentiment Rating.csv
- Product Data2.xls





Optional: Read table *web_activity* from the database *WebActivity.sqlite*

(hint: drag and drop the file to your workflow to get started)



Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

Data Manipulation

Clean, join, aggregate

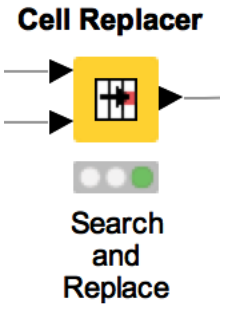
Copyright © 2017 KNIME.com AG

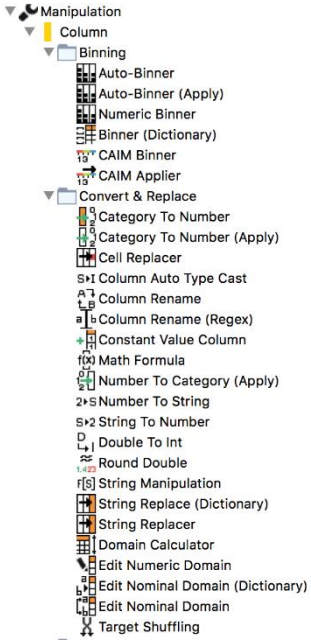
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

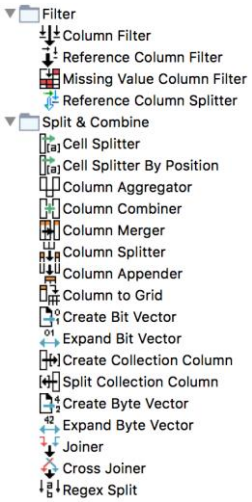
Open for Innovation ®
KNIME

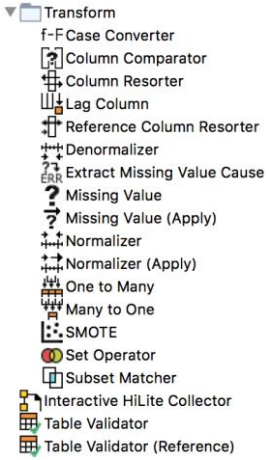
Data Manipulation Nodes

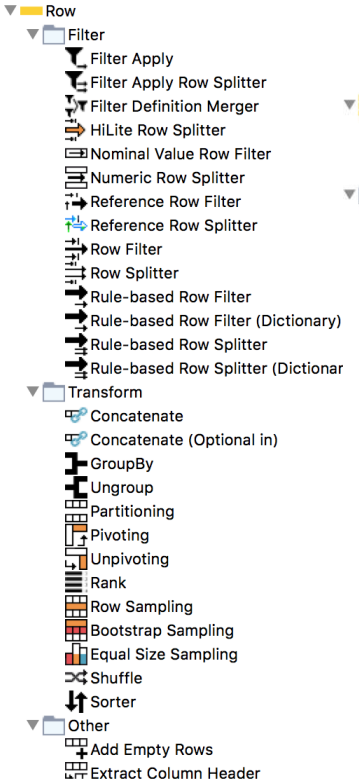
- Yellow color with a variety of input and output ports
- Apply a transformation to input data
- Many, many nodes!

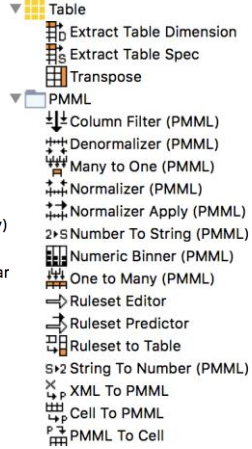












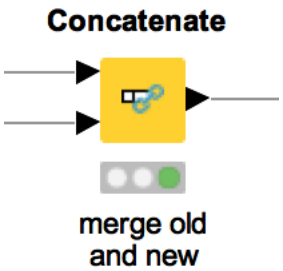
Copyright © 2017 KNIME.com AG

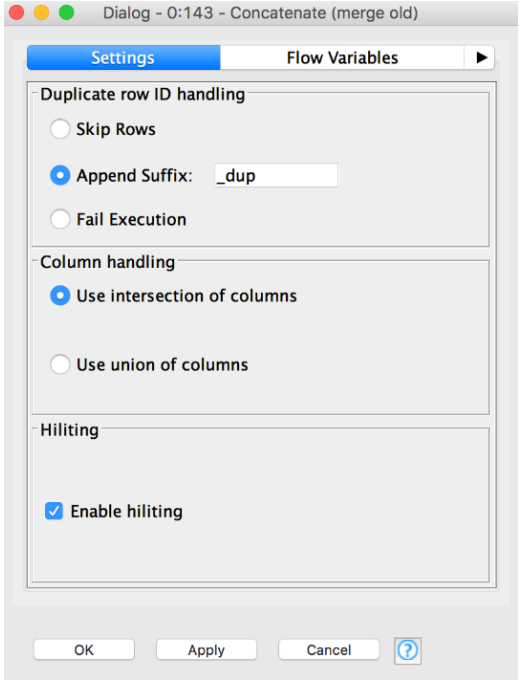
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

New Node: Concatenate

Combine rows from 2 tables with shared columns

- Handles duplicate row keys gracefully
- Take the union or intersection of columns





Copyright © 2017 KNIME.com AG

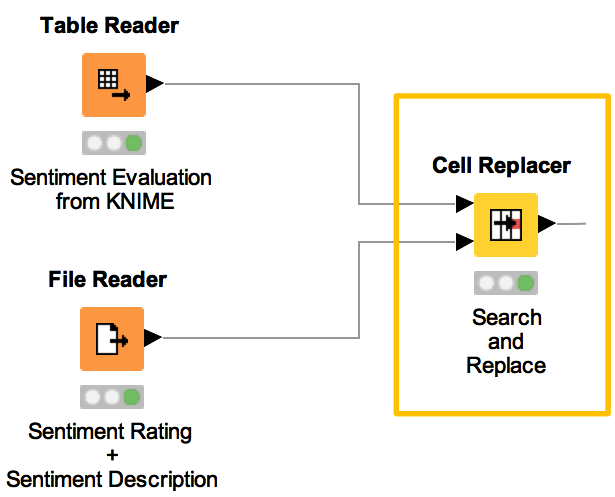
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

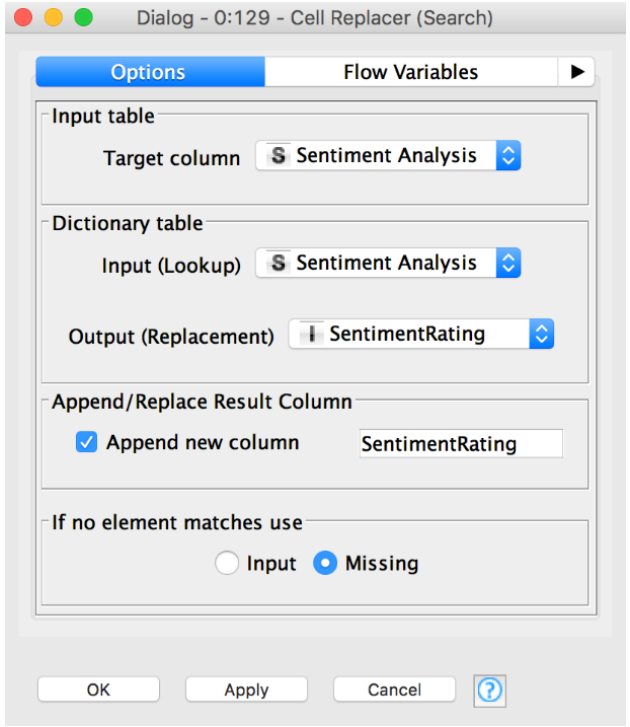
Open for Innovation ® KNIME

New Node: Cell Replacer

Replaces the content of a column based on a lookup

- Top port references the table to be searched
- Bottom port holds the lookup table (search keys and replacement values)





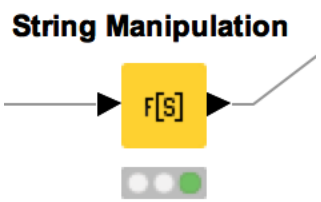
Copyright © 2017 KNIME.com AG

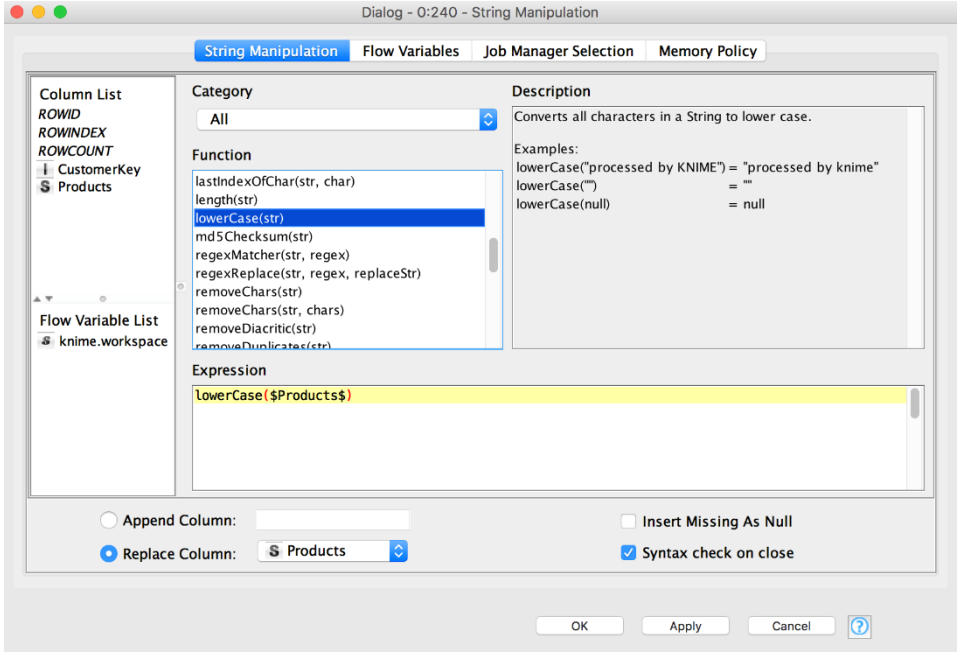
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation KNIME

New Node: String Manipulation

Create and edit values in String columns

- Clean up capitalization (eg. Lowercase)
- Modify existing strings or create new columns





Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

Data Manipulation Exercise, Activity I

Starting with exercise: *Data Manipulation, Activity I*

- Concatenate web activity data from old and new systems
- Replace sentiment evaluation (strings) with corresponding numeric values
- Use String Manipulation to ensure that all entries of the Products column are lower case from the product data spreadsheet.

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Joining Columns of Data

Left Table

Manually created table - 2:1 - Table Creator(left table)

File

Table "default" - Rows: 4 | Spec - Columns: 4 | Properties | Flow Variables

Row ID	ID	age	Income	class
Row1	1	23	<=50K	F1
Row2	3	25	<=50K	F3
Row3	6	22	>50K	A4
Row4	8	21	<=50K	C3

Join by ID

Right Table

Manually created table - 2:2 - Table Creator(right ta...

File

Table "default" - Rows: 6 | Spec - Columns: 4 | Properties | Flow Variables

Row ID	ID	age	Income	sex
Row1	1	23	<=50K	M
Row2	2	25	<=50K	F
Row3	4	23	>50K	M
Row4	5	21	<=50K	F
Row5	6	25	>50K	M
Row6	7	24	<=50K	M

Inner Join

Joined table - 2:3 - Joiner

File

Table "default" - Rows: 2 | Spec - Columns: 7 | Properties | Flow Variables

Row ID	ID	age	Income	class	age(*)	Income(*)	sex
Row1_Row1	1	23	<=50K	F1	23	<=50K	M
Row3_Row5	6	22	>50K	A4	25	>50K	M

Left Outer Join

Missing values in the right table.

Joined table - 2:3 - Joiner

File

Table "default" - Rows: 4 | Spec - Columns: 7 | Properties | Flow Variables

Row ID	ID	age	Income	class	age(*)	Income(*)	sex
Row1_Row1	1	23	<=50K	F1	23	<=50K	M
Row3_Row5	6	22	>50K	A4	25	>50K	M
Row2_?	3	25	<=50K	F3	?	?	?
Row4_?	8	21	<=50K	C3	?	?	?

Right Outer Join

Missing values in the left table.

Spec - Columns: 7 | Properties | Flow Variables

	age	Income	class	ID	age(*)	Income(*)	sex
Row1_	23	<=50K	F1	1	23	<=50K	M
Row3_R	22	>50K	A4	6	25	>50K	M
?_Row2	?	?	?	2	25	<=50K	F
?_Row3	?	?	?	4	23	>50K	M
?_Row4	?	?	?	5	21	<=50K	F
?_Row6	?	?	?	7	24	<=50K	M

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

Joining Columns of Data

Left Table

Manually created table - 2:1 - Table Creator(left table)

File

Table "default" - Rows: 4 | Spec - Columns: 4 | Properties | Flow Variables

Row ID	ID	age	Income	class
Row1	1	23	<=50K	F1
Row2	3	25	<=50K	F3
Row3	6	22	>50K	A4
Row4	8	21	<=50K	C3

Join by ID

Right Table

Manually created table - 2:2 - Table Creator(right ta...

File

Table "default" - Rows: 6 | Spec - Columns: 4 | Properties | Flow Variables

Row ID	ID	age	Income	sex
Row1	1	23	<=50K	M
Row2	2	25	<=50K	F
Row3	4	23	>50K	M
Row4	5	21	<=50K	F
Row5	6	25	>50K	M
Row6	7	24	<=50K	M

Full Outer Join

Joined table - 2:3 - Joiner

File

Table "default" - Rows: 8 | Spec - Columns: 8 | Properties | Flow Variables

Row ID	ID	age	Income	class	ID(*)	age(*)	Income(*)	sex
Row1_Row1	1	23	<=50K	F1	1	23	<=50K	M
Row3_Row5	6	22	>50K	A4	6	25	>50K	M
Row2_?	3	25	<=50K	F3	?	?	?	?
Row4_?	8	21	<=50K	C3	?	?	?	?
?_Row2	?	?	?	?	2	25	<=50K	F
?_Row3	?	?	?	?	4	23	>50K	M
?_Row4	?	?	?	?	5	21	<=50K	F
?_Row6	?	?	?	?	7	24	<=50K	M

Missing values in the left table.

Missing values in the right table.

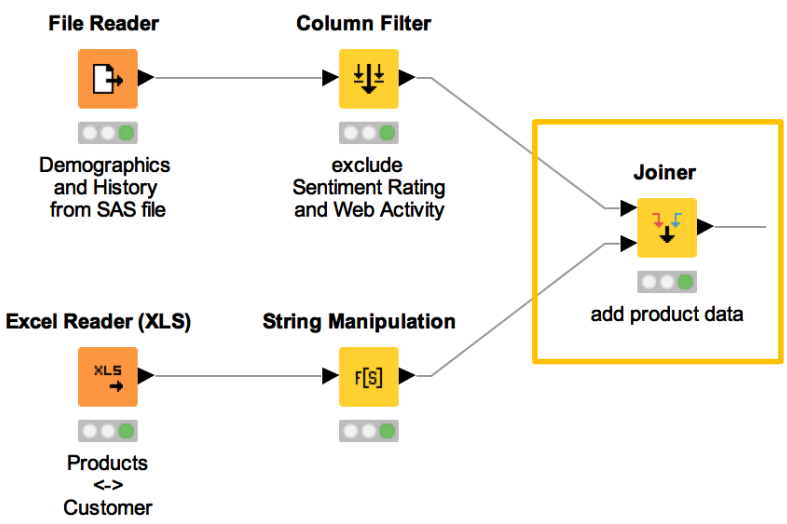
Copyright © 2017 KNIME.com AG

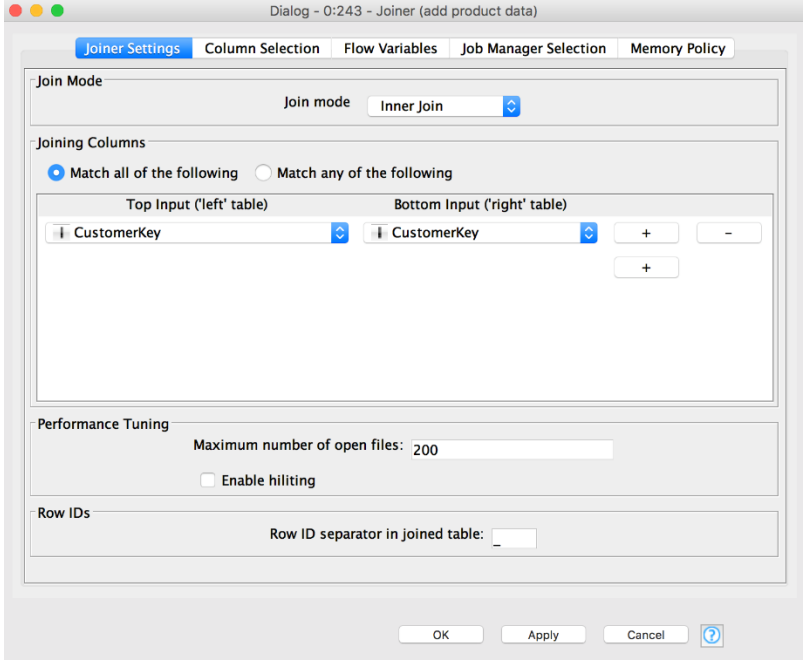
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

New Node: Joiner

- Combines columns from 2 different tables
- Top port contains “Left” data table
- Bottom port contains the “Right” data table

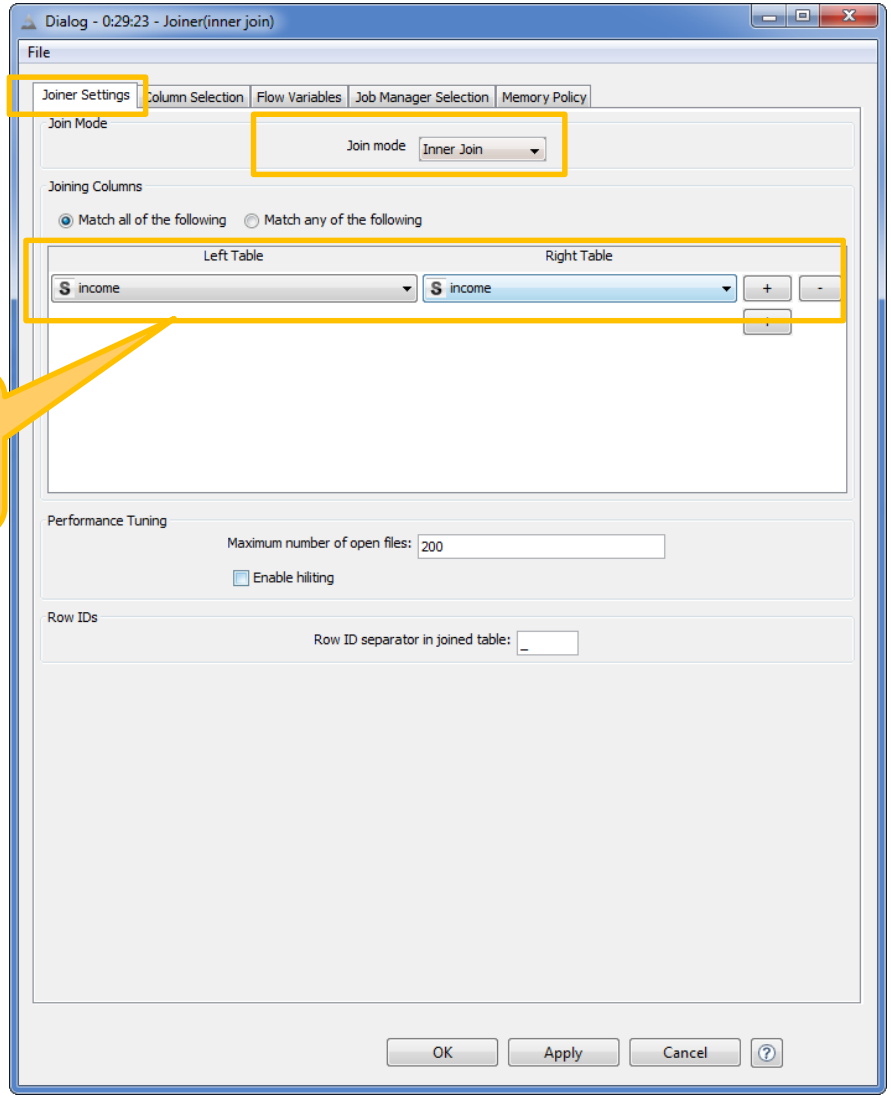




Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

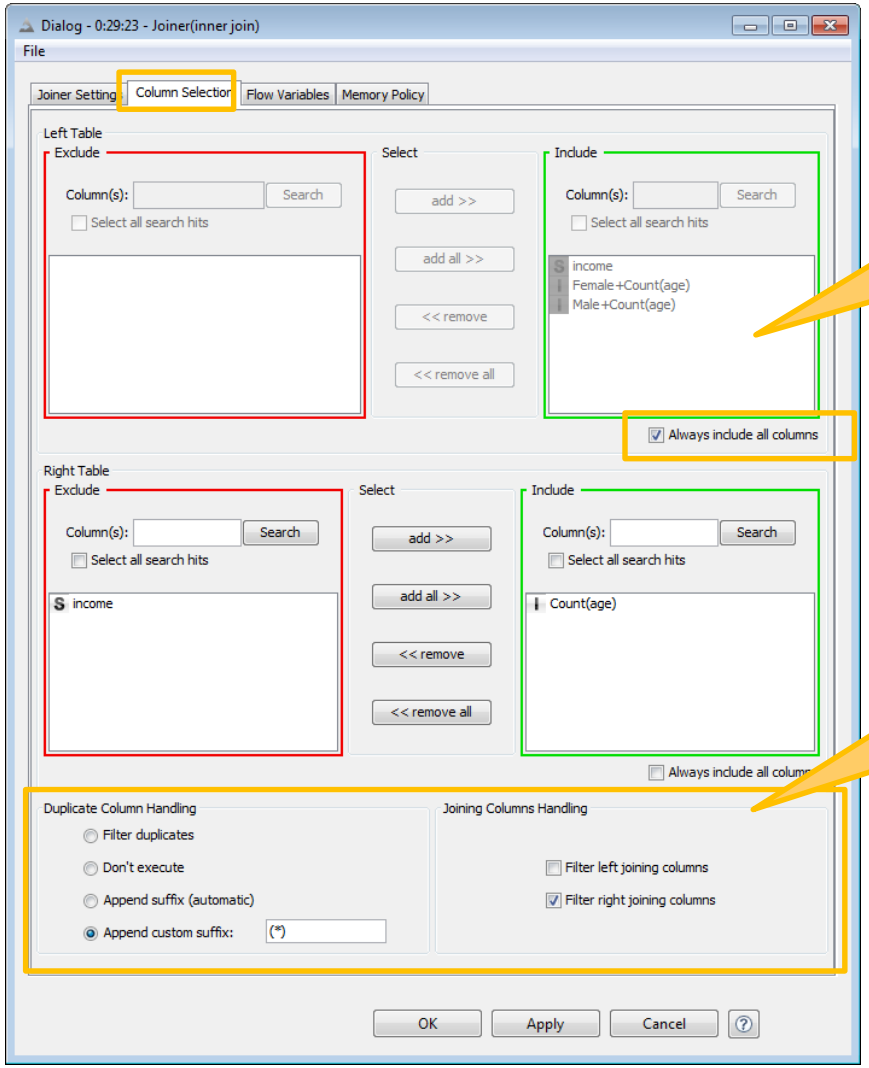
Joiner Configuration – Linking Rows



Values to join on. Multiple joining columns are allowed.

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Joiner Configuration – Column Selection



Columns from left table to output table

Columns from right table to output table

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Data Aggregation

rowid	group	value
r1	m	2
r2	f	3
r3	m	1
r4	f	5
r5	f	7
r6	m	5

rowid	group	sum(value)
r1+r3+r6	m	8
r2+r4+r5	f	15

aggregated on “group” by method: sum(“value”)

Copyright © 2017 KNIME.com AG

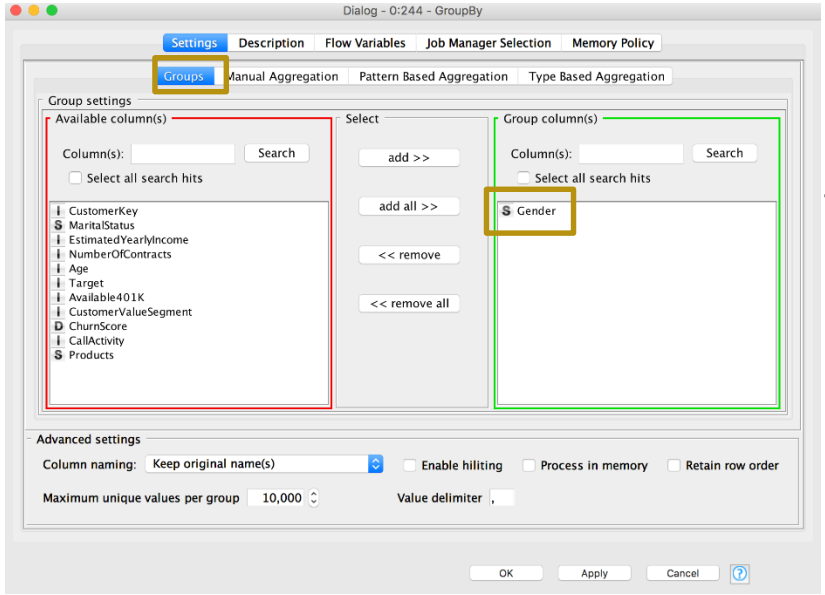
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

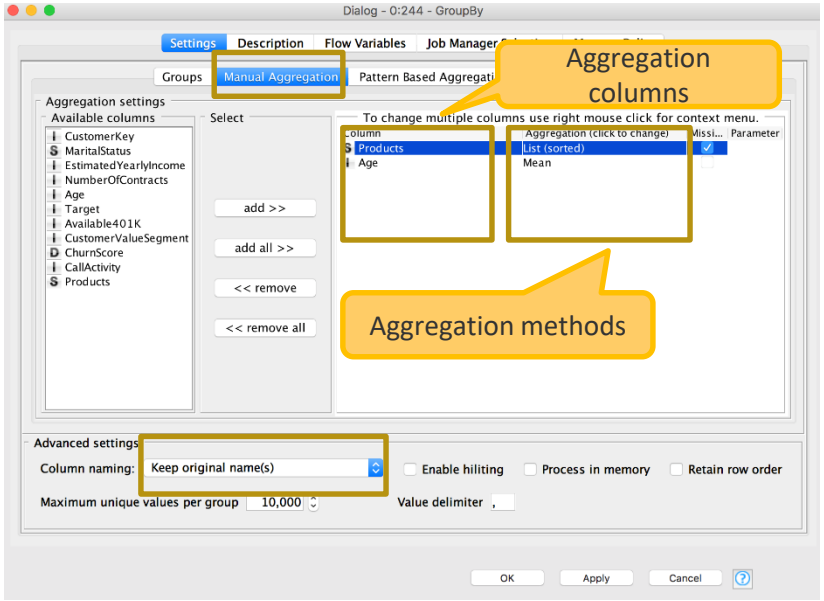
New Node: GroupBy

Aggregate to remove duplicates or summarize data

- First tab provides grouping options
- Second tab provides control over aggregation details

YouTube KNIME TV video:
https://youtu.be/bDwF-TOMtWw





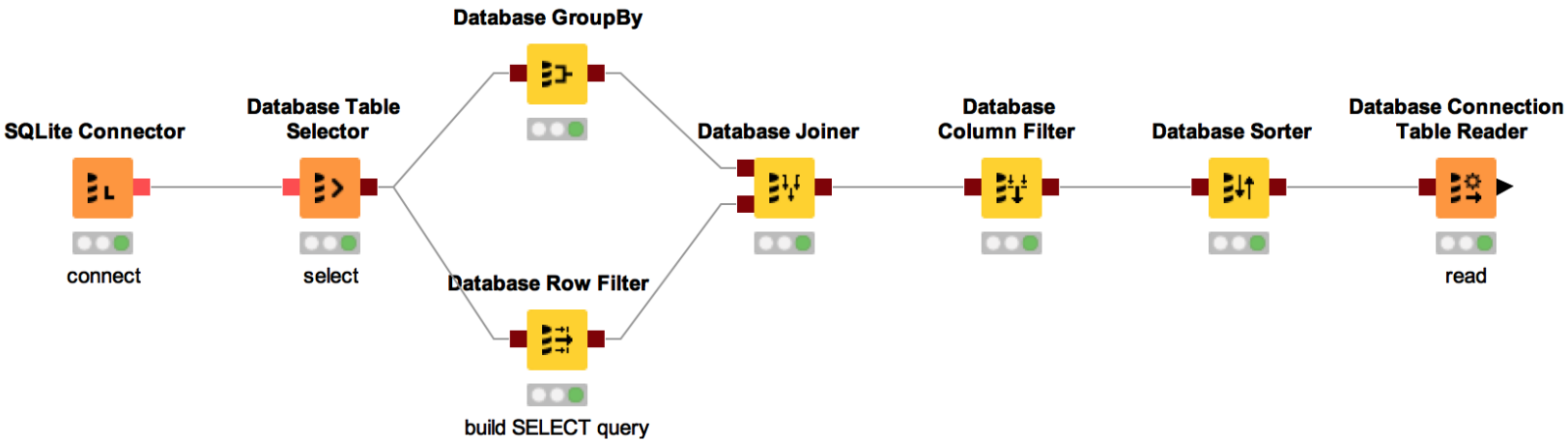
Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

In-database Data Manipulation

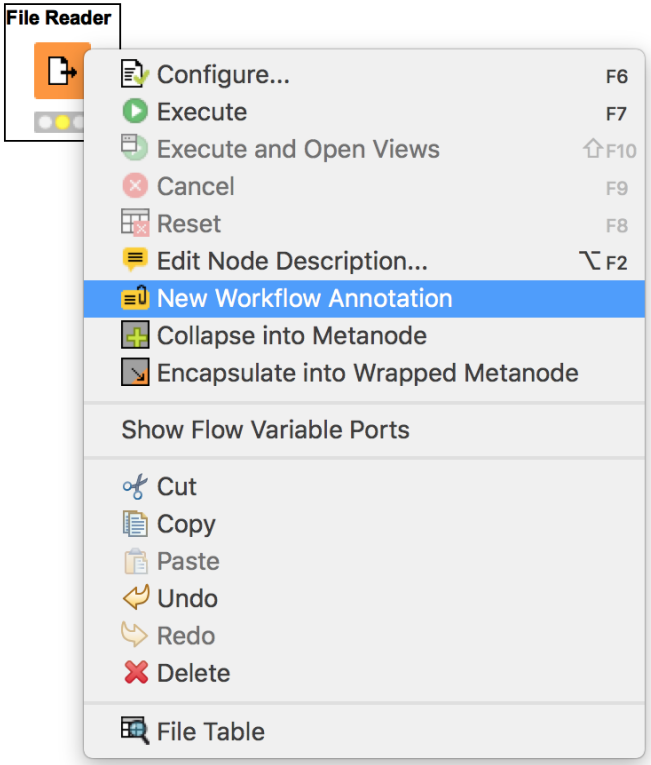
- Model SQL query using nodes
- DB versions of GroupBy, Joiner, Row Filter, Sorter, etc.

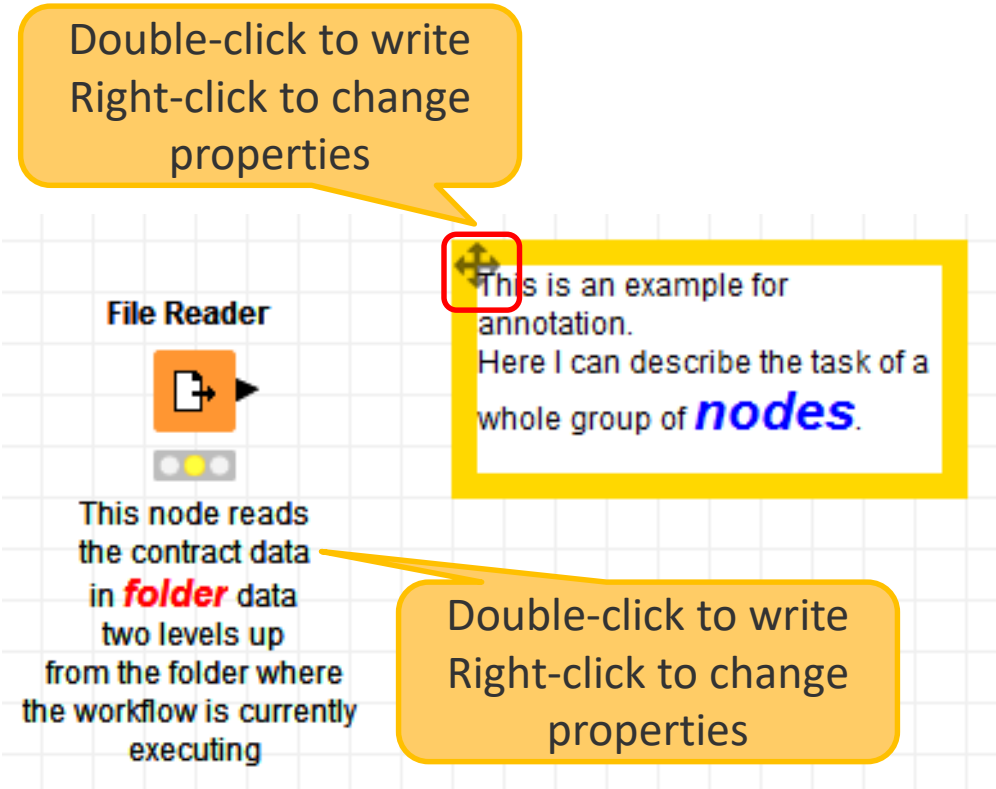


Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Comments & Annotations





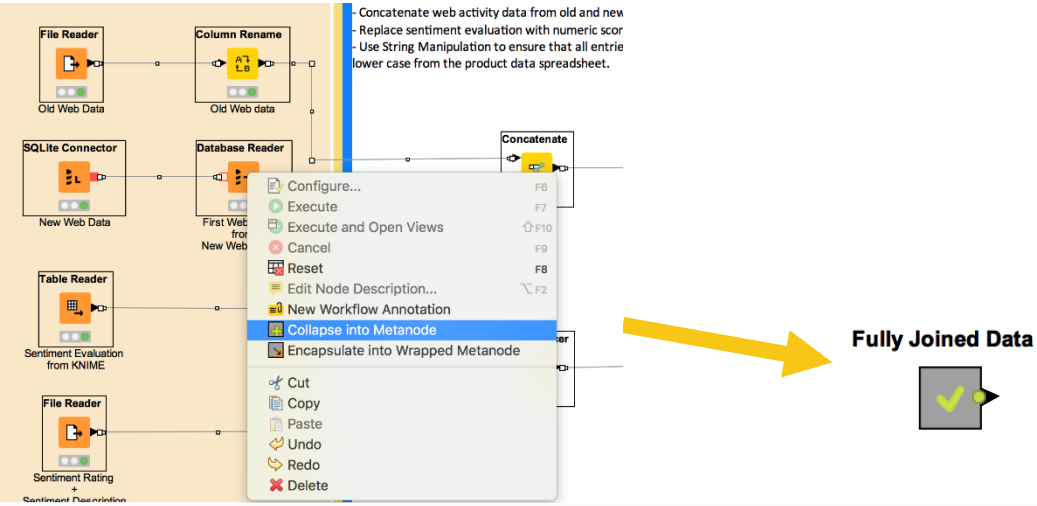
YouTube KNIME TV Channel:
https://youtu.be/AHURYB_O8sA

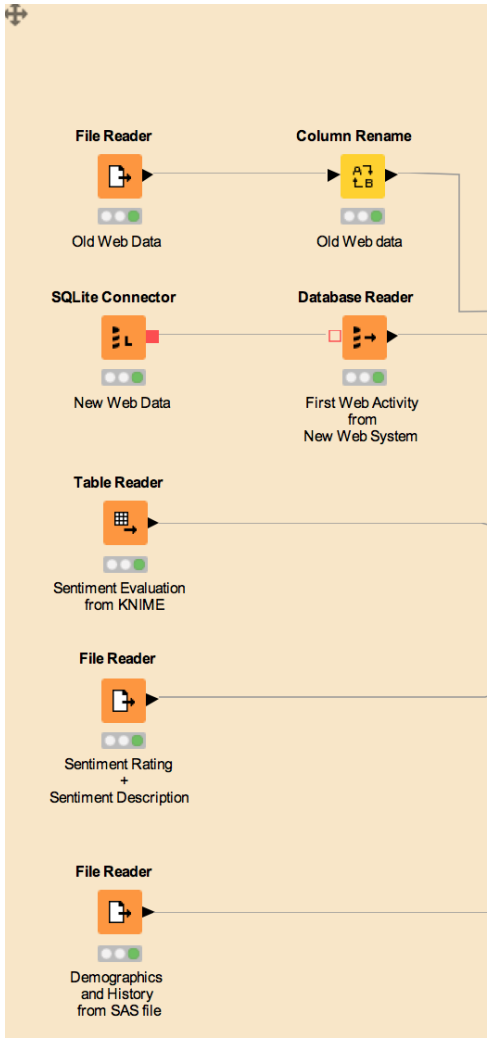
Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Workflow Organisation – Good Practices

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





Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

Data Manipulation Exercise, Activity II

Starting with exercise *Data Manipulation, Activity II*

- Join all data together using a series of joiner nodes and the "Customer Key" field
- Resolve duplicates in the joined dataset (hint: GroupBy node)
- Clean up and document your workflow using annotations, node labels and metanodes

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Data Visualization

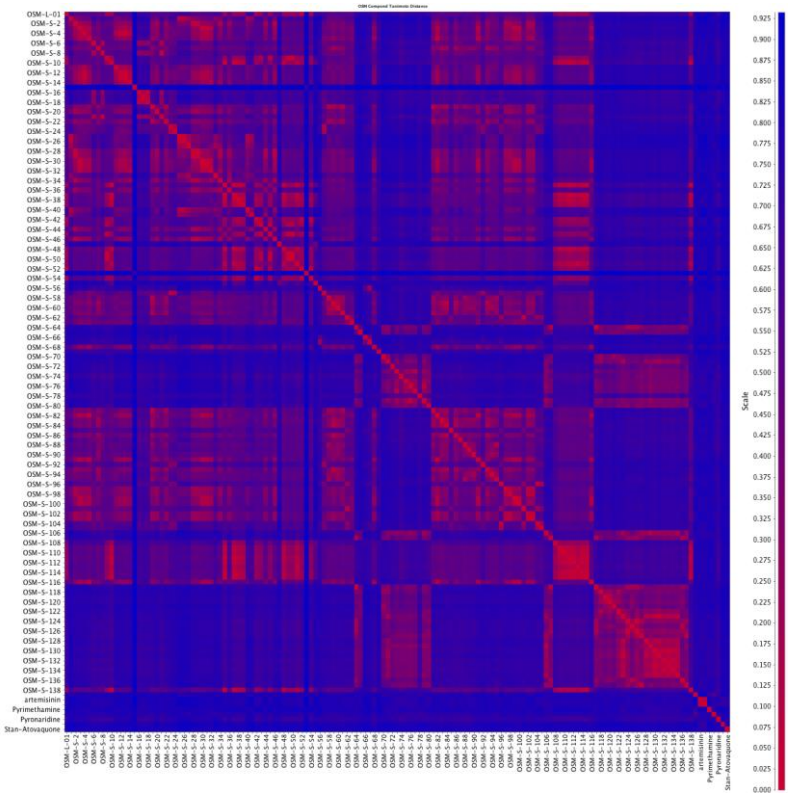
Charts and tables

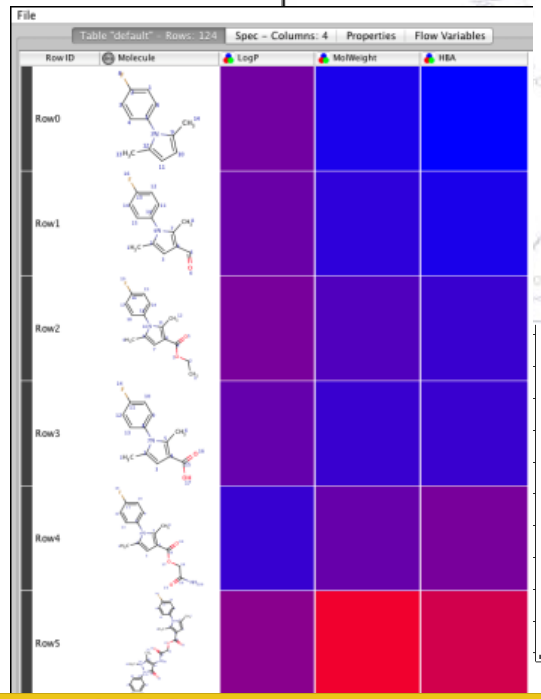
Copyright © 2017 KNIME.com AG

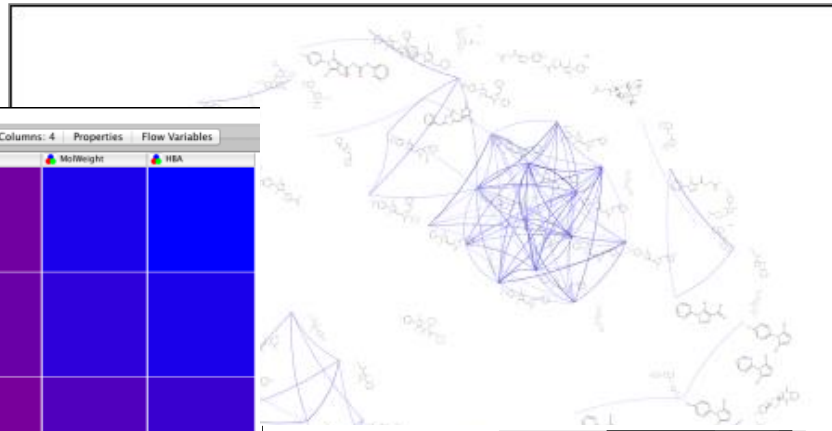
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

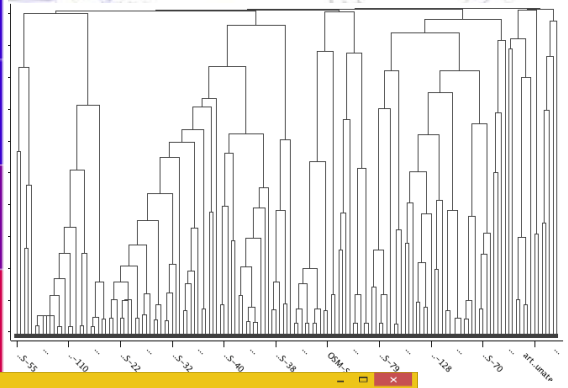
Open for Innovation ®
KNIME

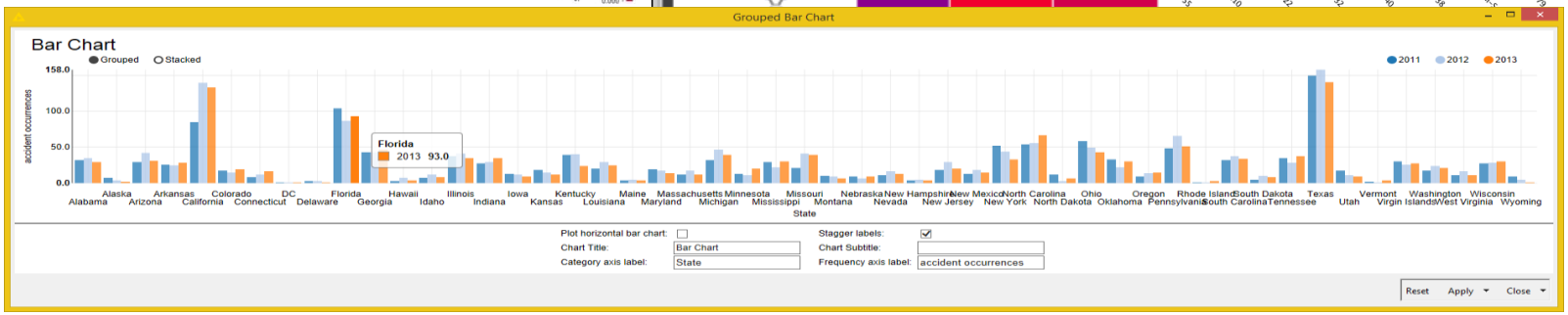
Data Visualization











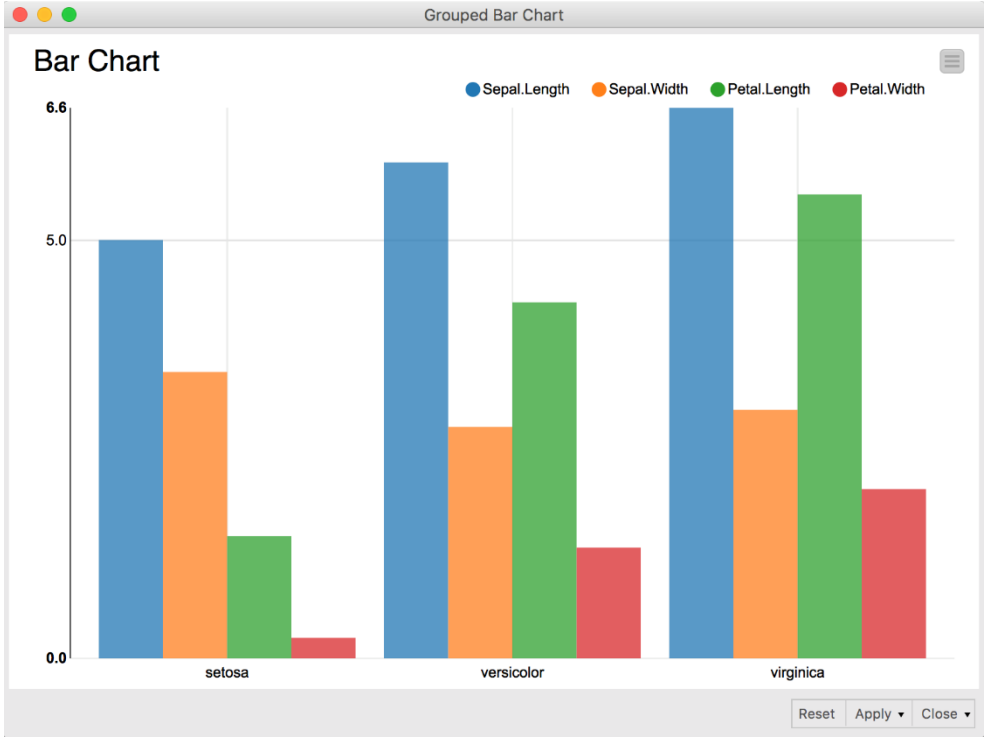
Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

Data Visualization

- Interactive plots and tables (with Highlighting)
- JavaScript integration for interactive views
- R View nodes for building advanced graphics in KNIME



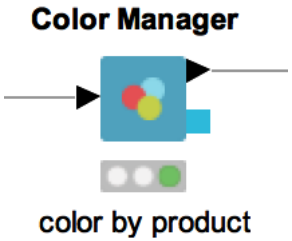
Copyright © 2017 KNIME.com AG

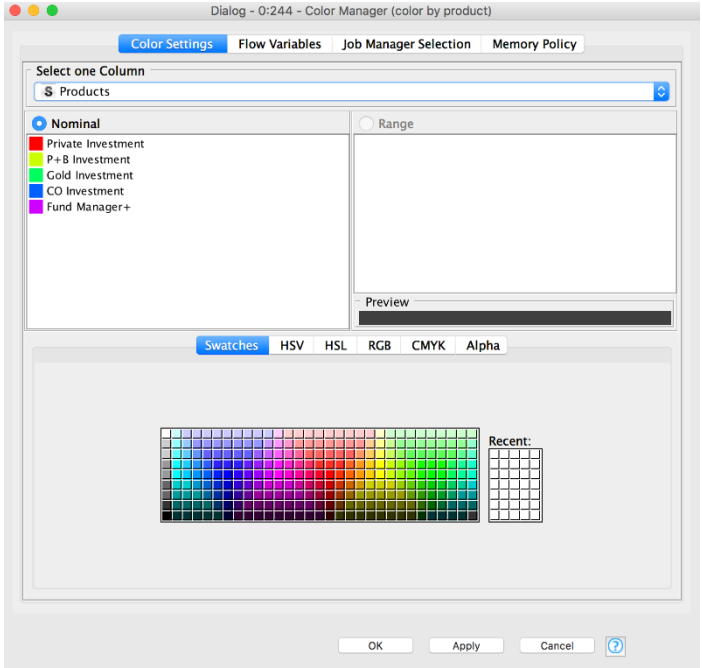
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

New Node: Color Manager

One of several visual property managers (e.g. size, shape)

- Color by nominal or continuous values
- Sync colors between views using the color model port





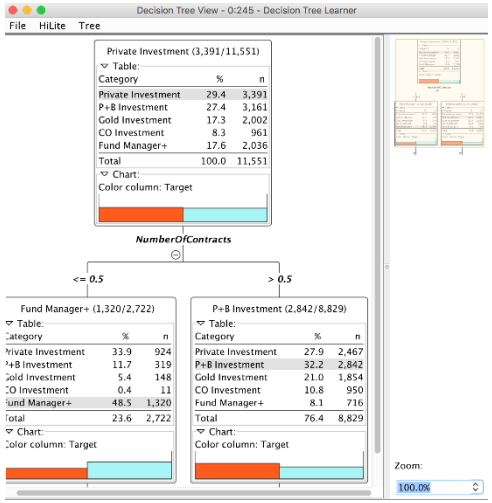
Copyright © 2017 KNIME.com AG

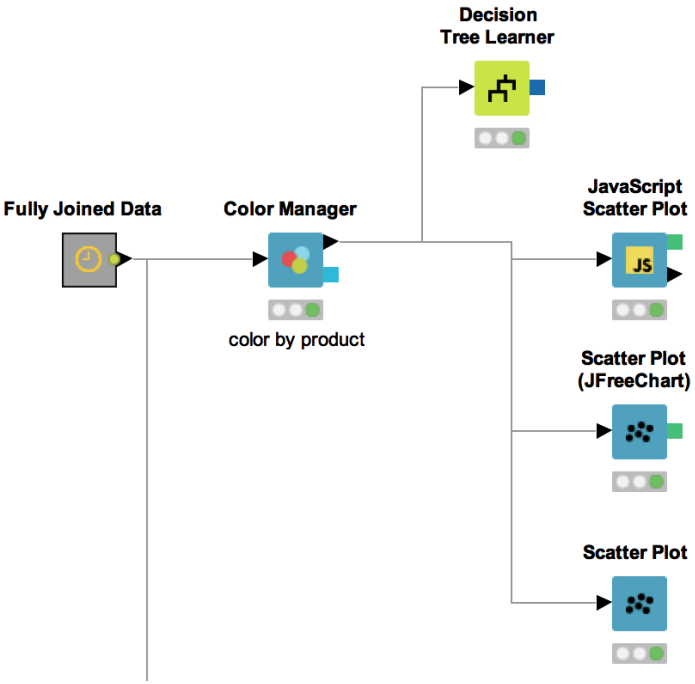
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

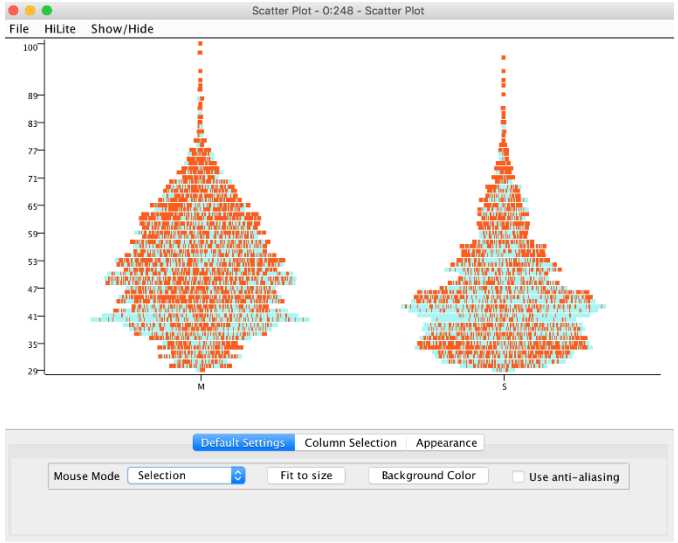
Open for Innovation ® KNIME

Hiliting

- **Hilited** data is visible across all views
- Keep multiple views open to explore complex data



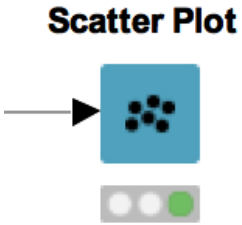


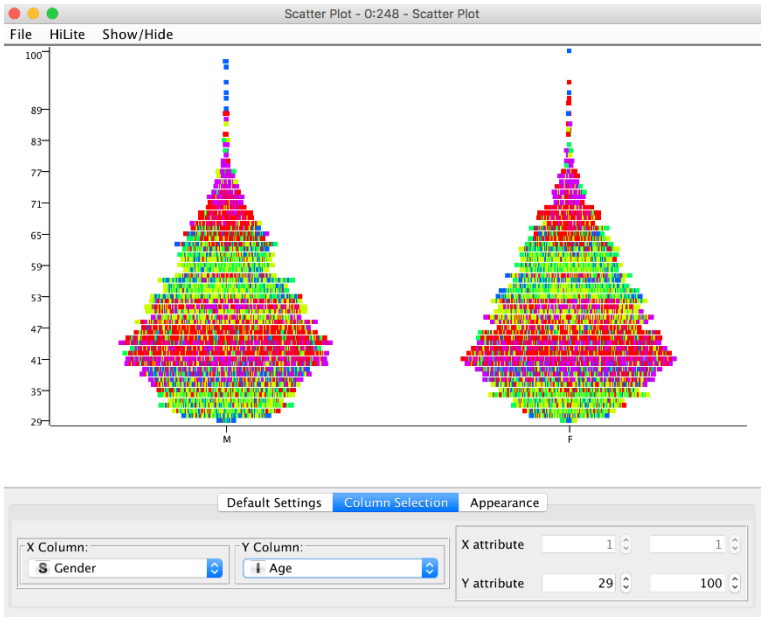


Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

New Node: Scatter Plot

- Plot different columns on X and Y
- Displays data including pre calculated visual properties (size, shape, color)
- Supports highlighting
- Produces a view, no image





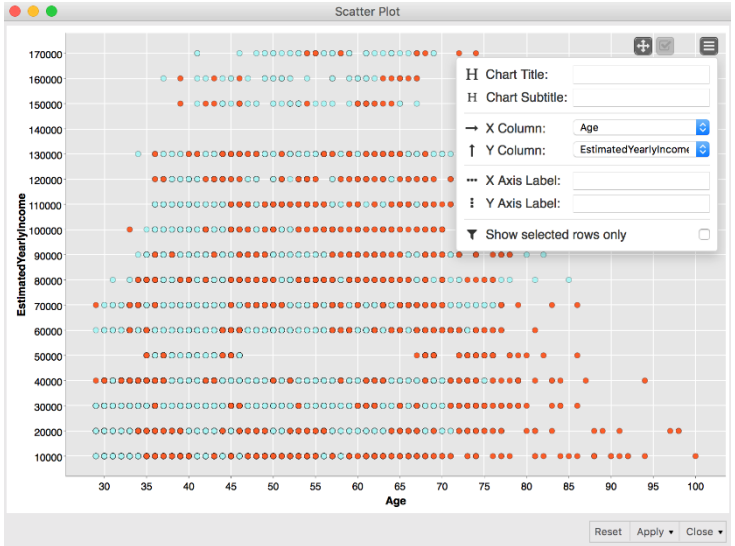
Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

New Node: JavaScript Scatter Plot

- Plot different columns on X and Y
- Displays data including pre calculated visual properties (size, shape, color)
- Does not support highlighting
- Produces an interactive View and an Image



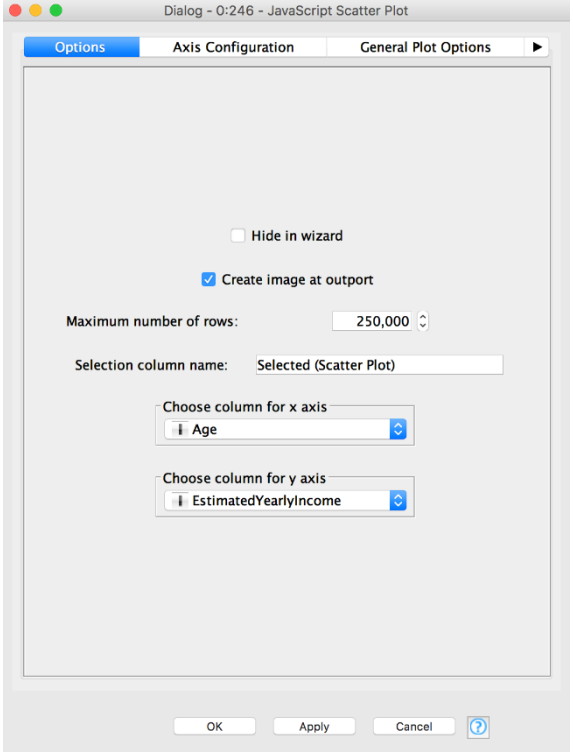
Copyright © 2017 KNIME.com AG

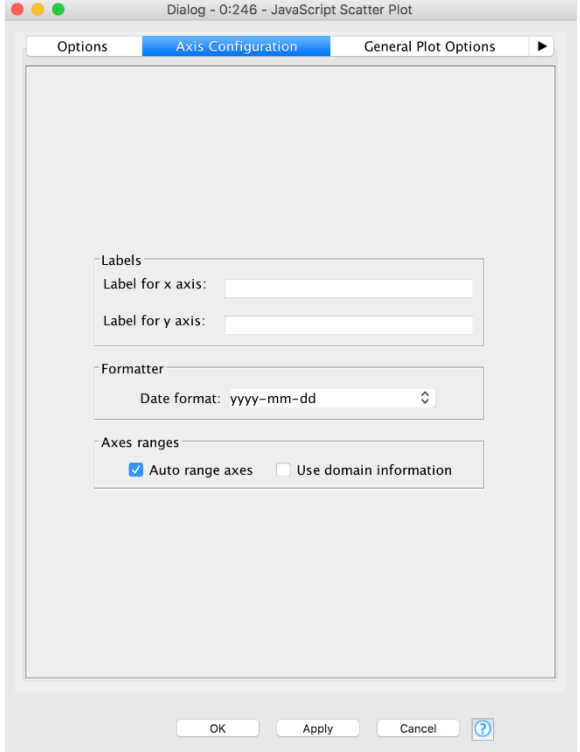
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

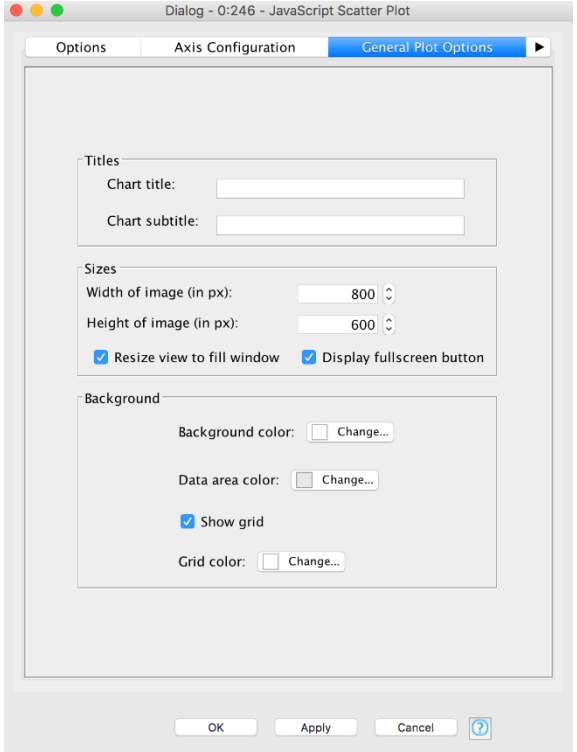
Open for Innovation ® KNIME

New Node: JavaScript Scatter Plot

- 3 configuration tabs





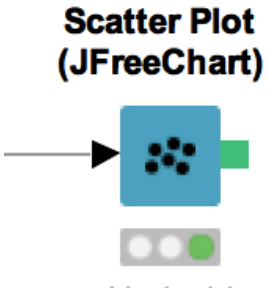


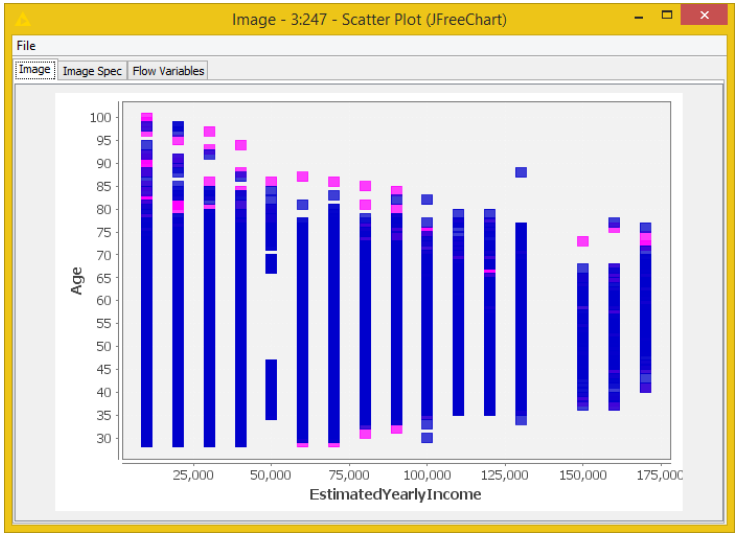
Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

New Node: Scatter Plot (JFreeChart)

- Plot different columns on X and Y
- Displays data including pre-calculated visual properties (size, shape, color)
- Does not support highlighting
- Produces a static view and an image





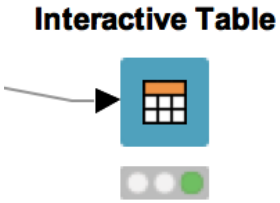
Copyright © 2017 KNIME.com AG

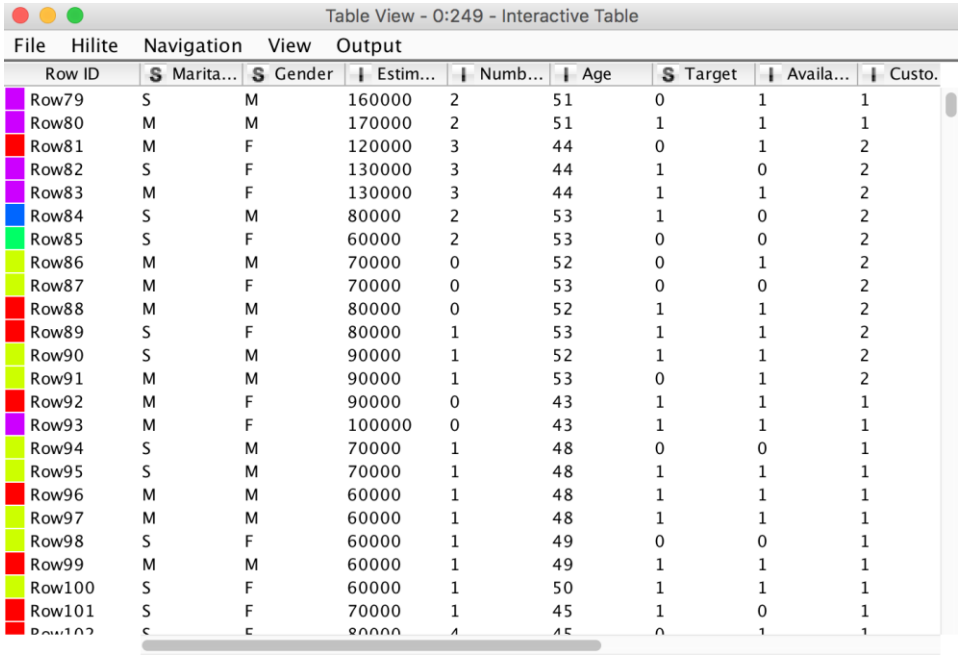
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ®
KNIME

New Node: Interactive Table

- No graphics
- Supports highlighting





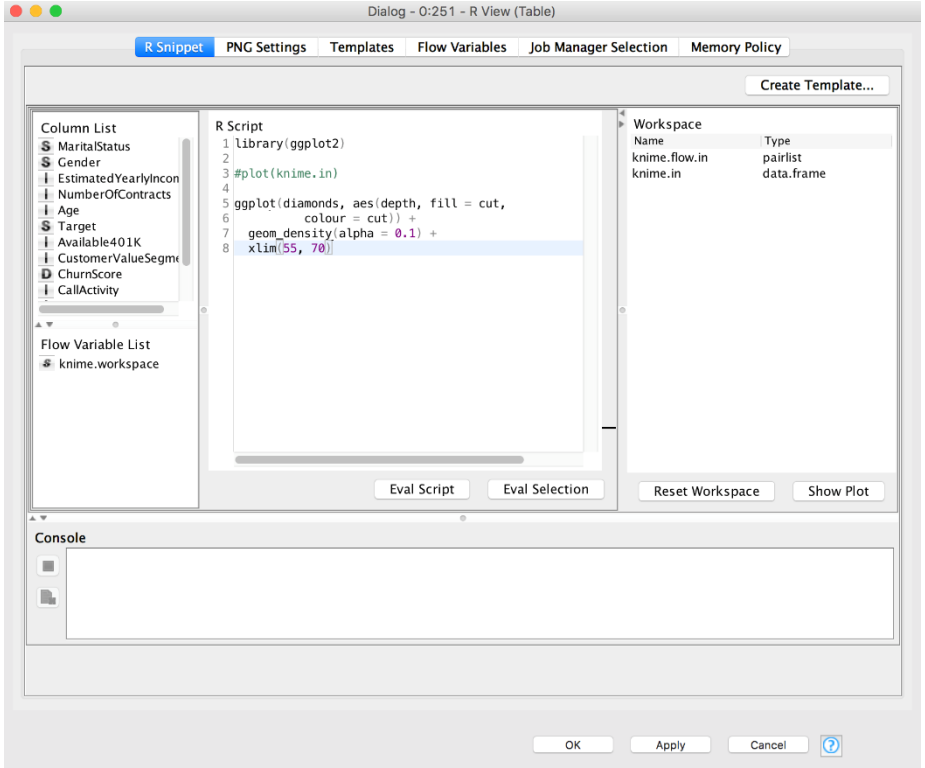
Row ID	S Marita...	S Gender	I Estim...	I Numb...	I Age	S Target	I Availa...	I Custo.
Row79	S	M	160000	2	51	0	1	1
Row80	M	M	170000	2	51	1	1	1
Row81	M	F	120000	3	44	0	1	2
Row82	S	F	130000	3	44	1	0	2
Row83	M	F	130000	3	44	1	1	2
Row84	S	M	80000	2	53	1	0	2
Row85	S	F	60000	2	53	0	0	2
Row86	M	M	70000	0	52	0	1	2
Row87	M	F	70000	0	53	0	0	2
Row88	M	M	80000	0	52	1	1	2
Row89	S	F	80000	1	53	1	1	2
Row90	S	M	90000	1	52	1	1	2
Row91	M	M	90000	1	53	0	1	2
Row92	M	F	90000	0	43	1	1	1
Row93	M	F	100000	0	43	1	1	1
Row94	S	M	70000	1	48	0	0	1
Row95	S	M	70000	1	48	1	1	1
Row96	M	M	60000	1	48	1	1	1
Row97	M	M	60000	1	48	1	1	1
Row98	S	F	60000	1	49	0	0	1
Row99	M	M	60000	1	49	1	1	1
Row100	S	F	60000	1	50	1	1	1
Row101	S	F	70000	1	45	1	0	1
Row102	S	F	80000	4	45	0	1	1

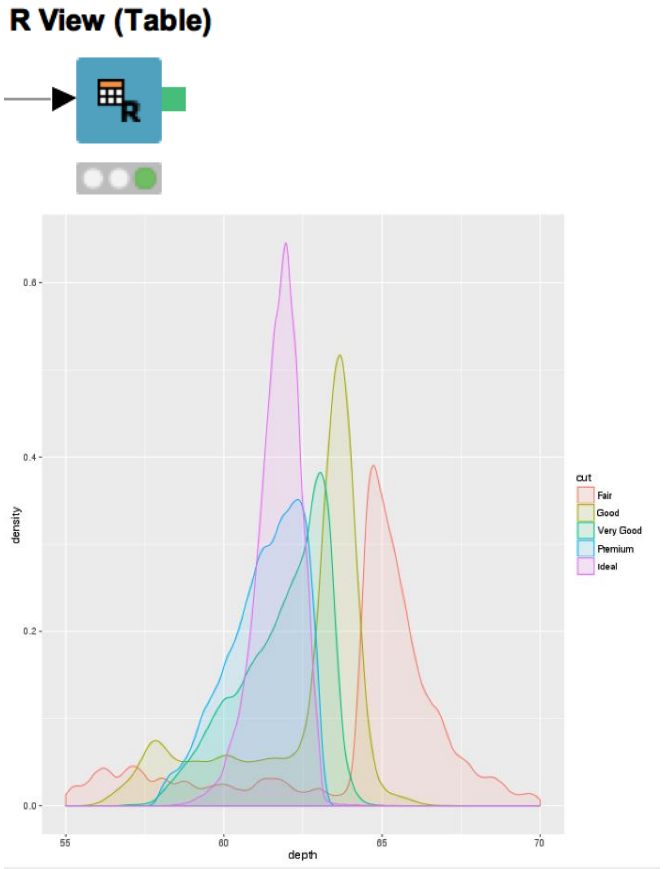
Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Other Nodes: R View

- R View nodes for maximum customizibility





Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Visualization Exercise

Start with exercise: *Visualization*

- Color data by Product
- Produce a scatter plot of Age vs. Estimated Yearly Income
- In the “Scatter Plot” node highlight some data points and view only highlighted points using an Interactive Table node
- Optional: Visualize data with the R (View node)
  - (start with a script like: *plot(knime.in)*

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Data Mining

Partition, learn, predict, score

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ®
KNIME

Data Mining Strategies

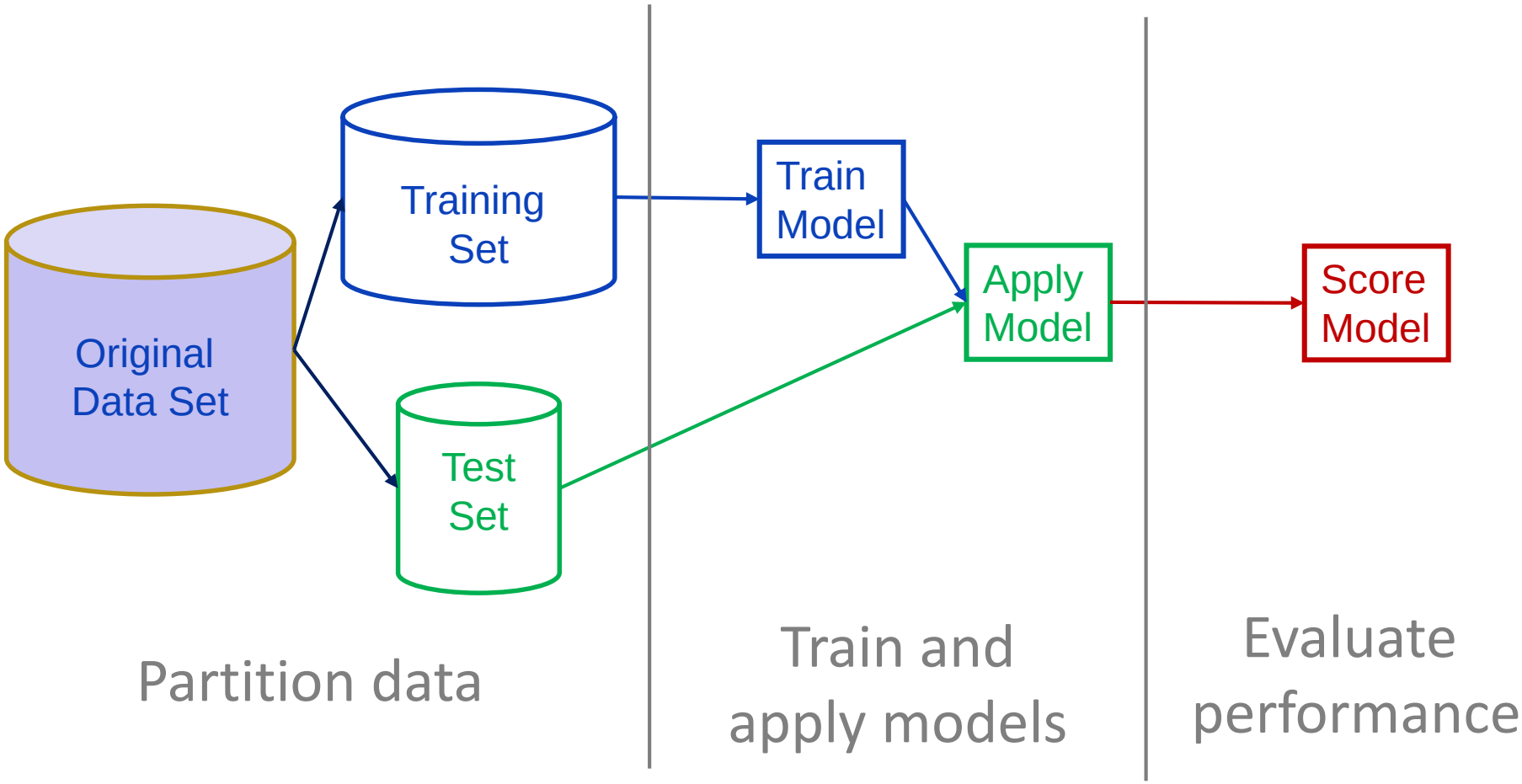
Example applications:

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

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Data Mining: Process Overview

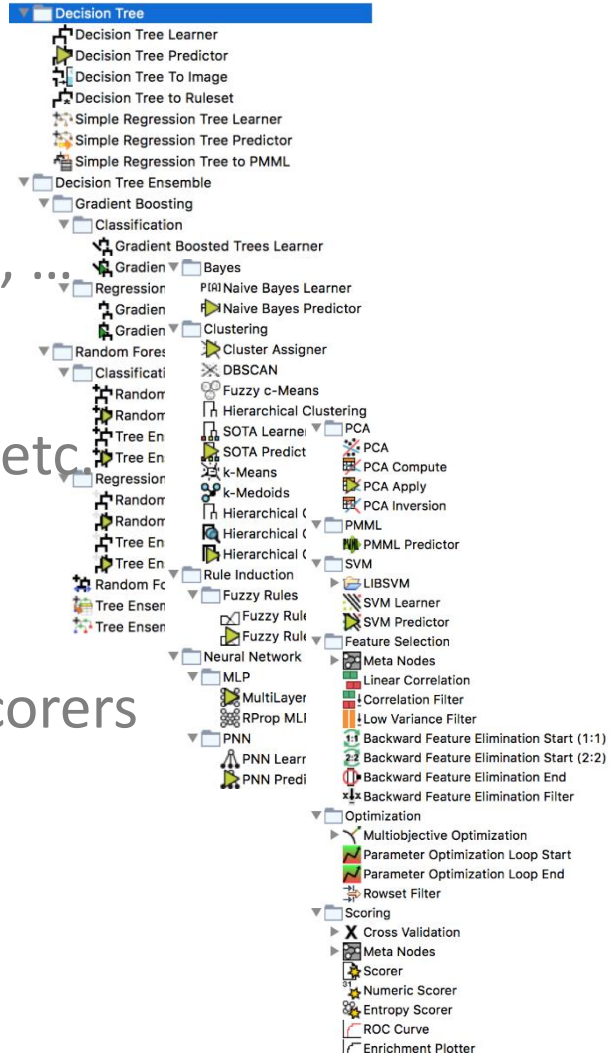


Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Data Mining in KNIME

- KNIME has many modeling tools!
  - Decision tree, random forest, SVM, regression, neural networks, clustering, …
  - and integrations with other libraries: WEKA, libSVM, R, Python (scikit-learn) etc.
- And many model evaluation nodes
  - ROC, standard, numeric and entropy scorers
  - Feature elimination
  - Cross validation

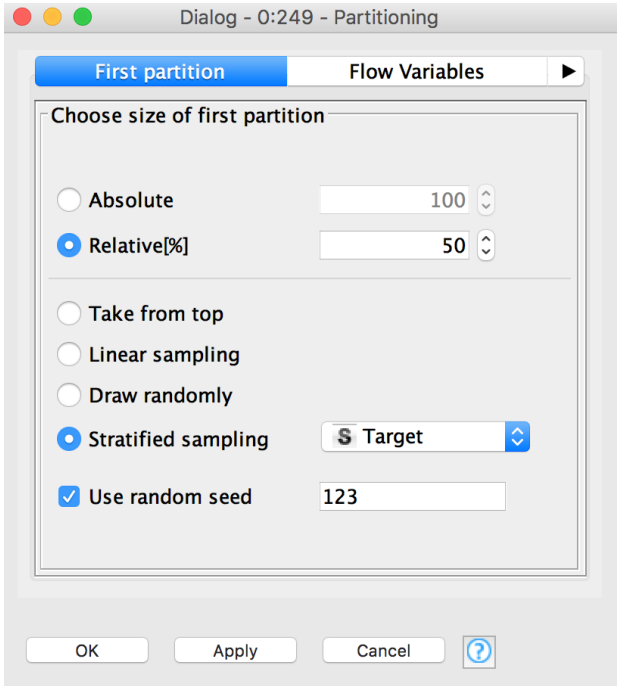


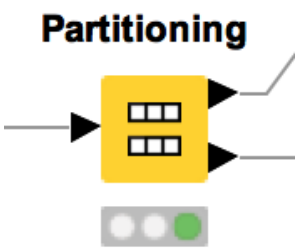
Copyright © 2017 KNIME.com AG

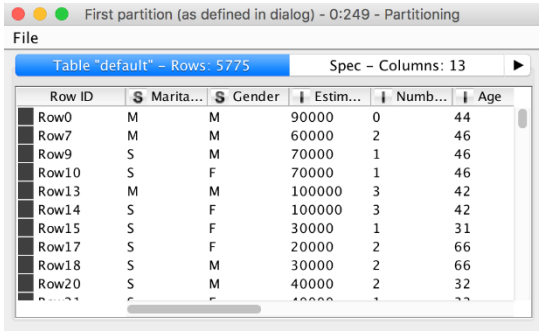
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

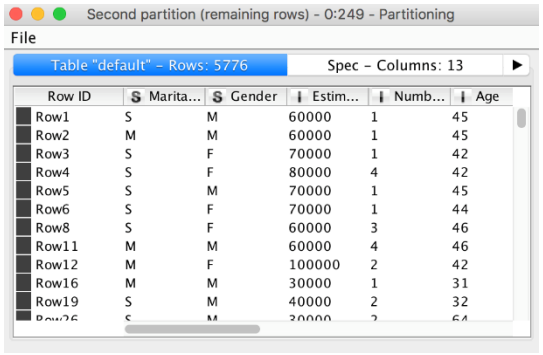
New Node: Partitioning

- Use to split data into training and evaluation sets
  - Partition by count (e.g. 10 rows) or fraction (e.g. 10%)
  - Sample by a variety of methods; random, linear, stratified







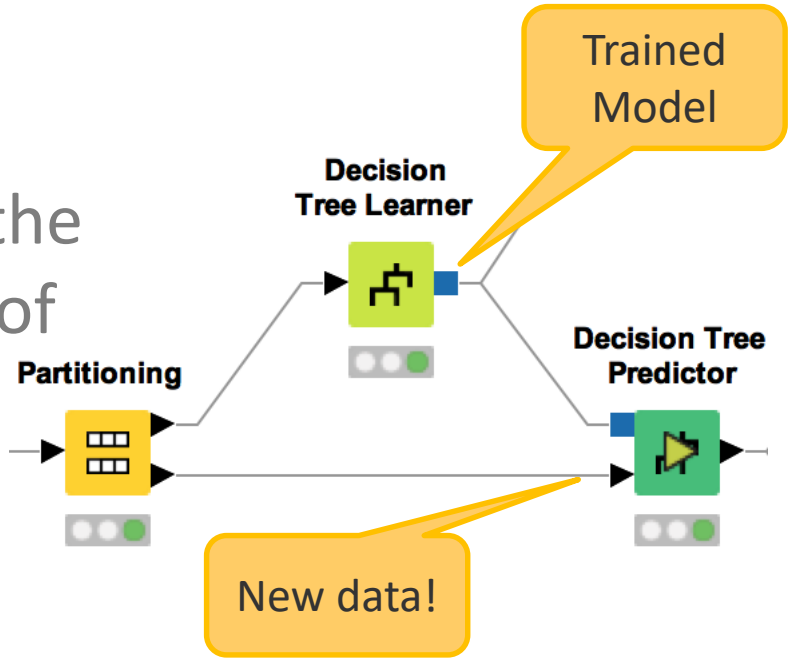


Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Learner-predictor Motif

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



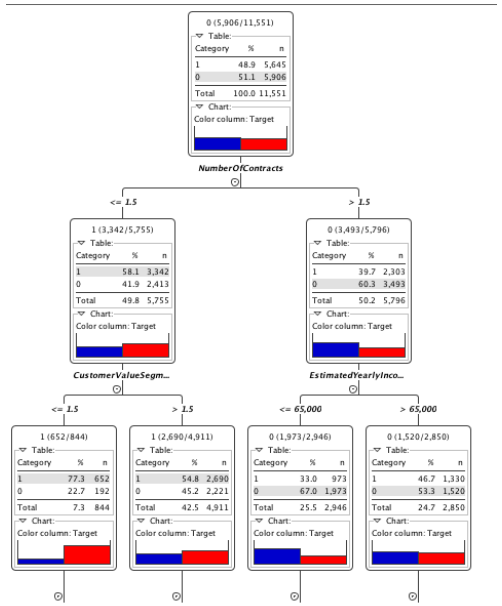
Copyright © 2017 KNIME.com AG

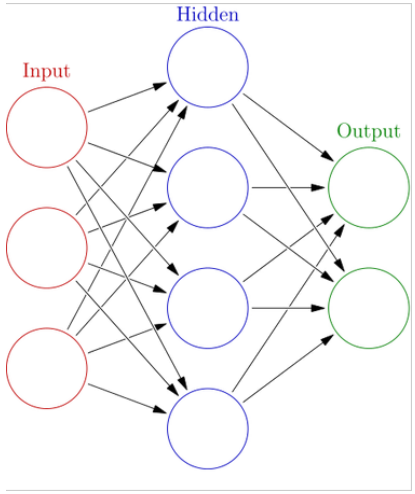
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Classification

Predict *nominal* outcomes on existing data (supervised)

- Applications
  - Churn analysis (yes/no)
  - Chemical activity (active/inactive)
  - Spam detection (spam/not spam)
  - Optical character recognition (A-Z)
- Methods
  - Decision Trees
  - Neural Networks
  - Naïve Bayes
  - Logistic Regression





Class counts for Target

Class:	0	1
Count:	5906	5645

Total count: 11551

Threshold to used for zero probabilities: 0.0

Gaussian distribution for Age per class value

	0	1
Count:	5906	5645
Mean:	49.68557	46.82604
Std. Deviation:	12.27388	10.16363
Rate:	51%	49%

Gaussian distribution for Available401K per class value

	0	1
Count:	5906	5645
Mean:	0.68134	0.68485
Std. Deviation:	0.466	0.46462
Rate:	51%	49%

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

KNIME's Decision Tree

J.R. Quinlan, "C4.5 Programs for machine learning"

J. Shafer, R. Agrawal, M. Mehta, "SPRINT: A Scalable Parallel Classifier for Data Mining"

C4.5 builds a tree from a set of training data using the concept of information entropy.

At each node of the tree, the attribute of the data with the highest **normalized information gain** (difference in entropy) is chosen to split the data.

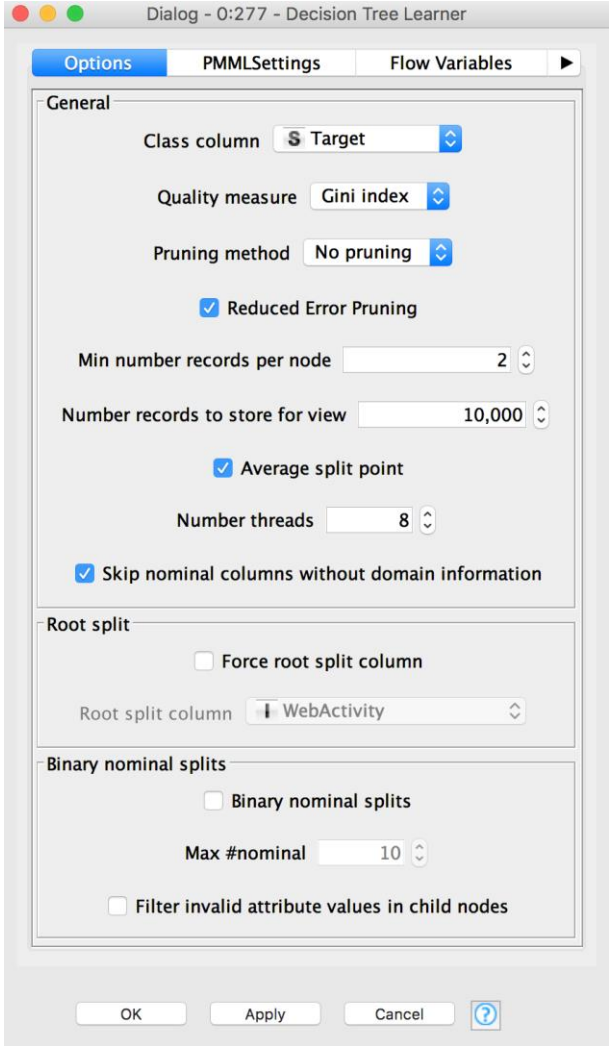
The C4.5 algorithm then recurses on the smaller sub lists.

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

New Node: Decision Tree Learner

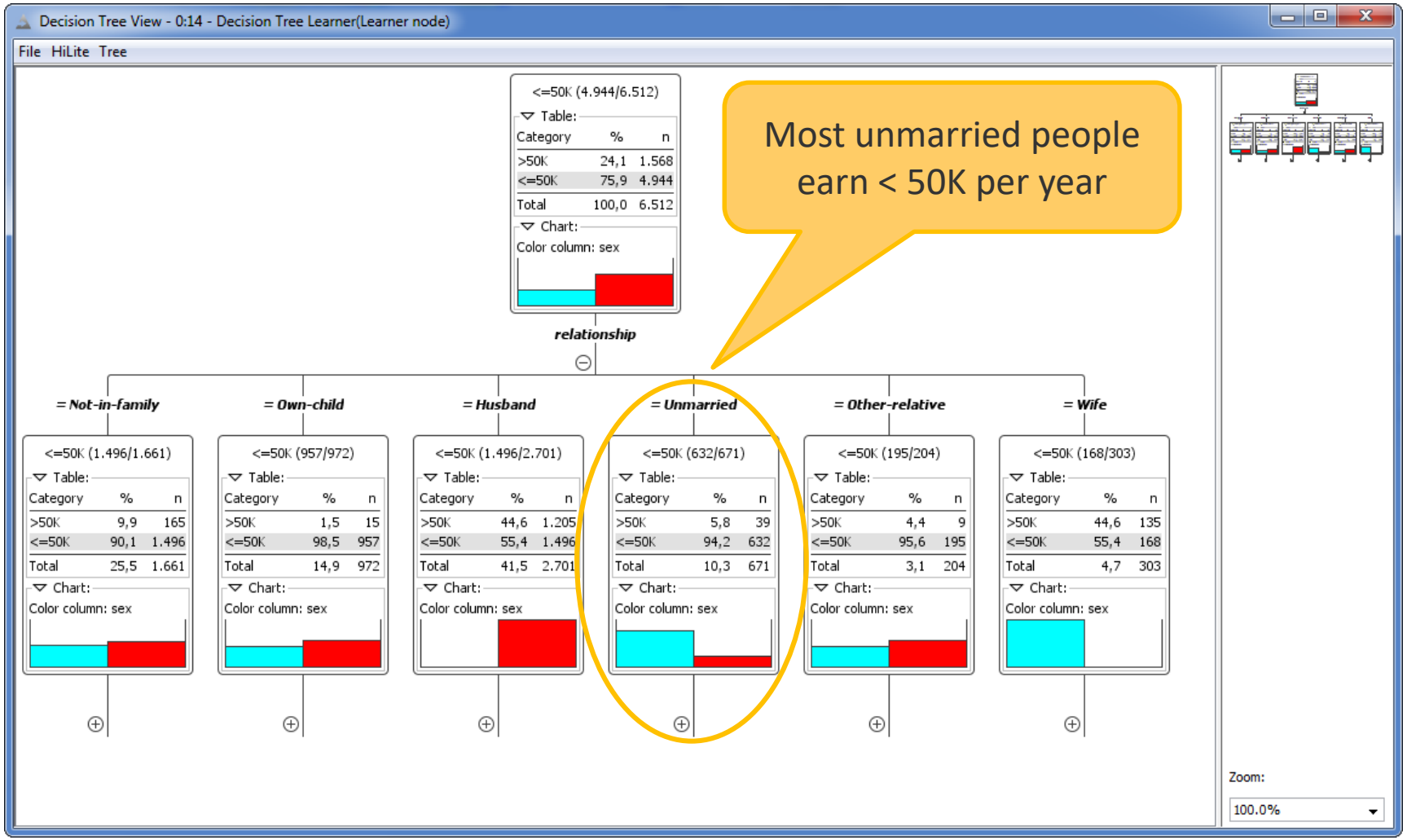


Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ®
KNIME

Decision Tree View



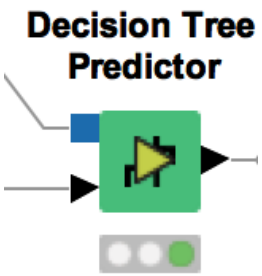
Copyright © 2017 KNIME.com AG

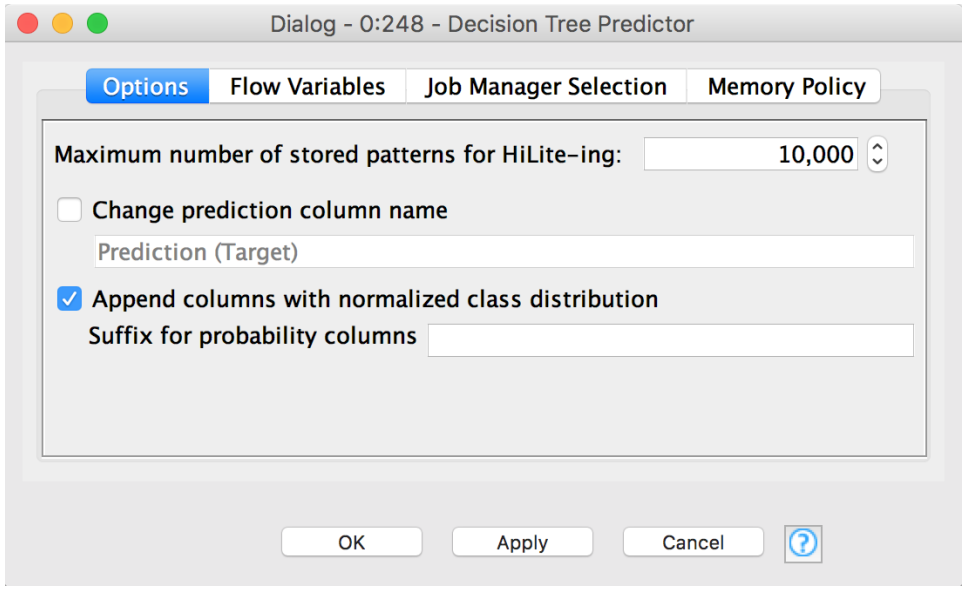
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

New Node: Decision Tree Predictor

- Takes a decision tree model & applies it to new data
- Check the box to append class probabilities



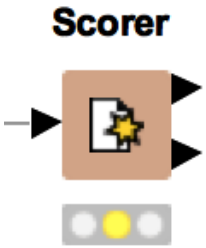


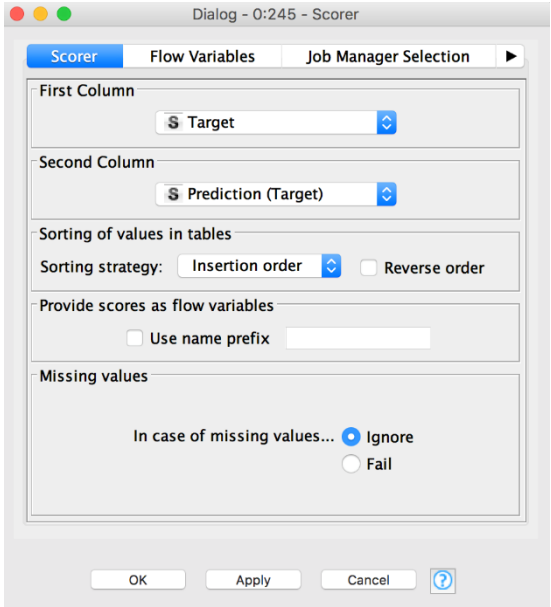
Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

New Node: Scorer

- Compare predicted results to known truth in order to evaluate model quality
  - Confusion matrix shows the distribution of model errors
  - An accuracy statistics table provides a detailed analysis of model quality.





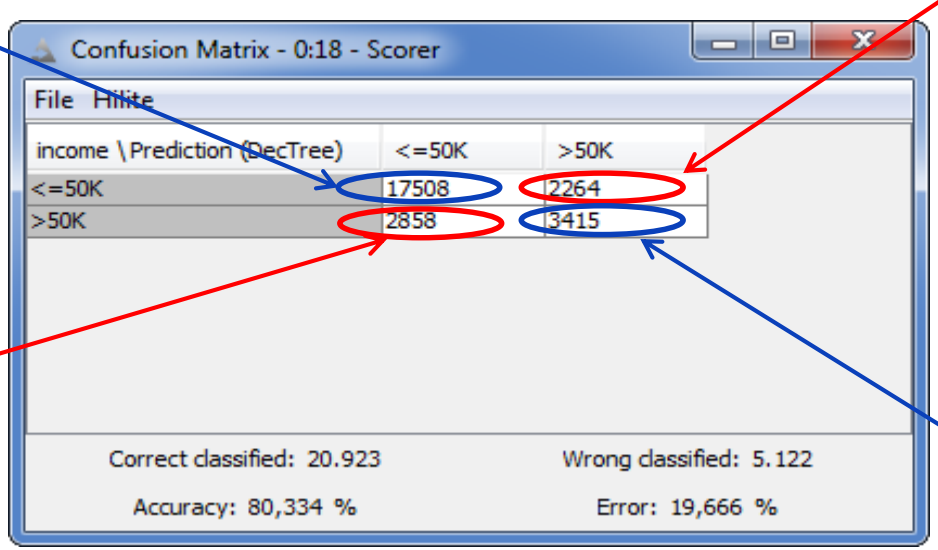
Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

New Node: Scorer

True Positives
Income = “<=50K”
Predicted = “<=50K”

False Positives
Income = “<=50K”
Predicted = “>50K”



False Negatives

True Negatives

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Scorer: Accuracy Measures

Accuracy statistics - 3:246 - Scorer (Accuracy and)

File

Table "default" - Rows: 3 | Spec - Columns: 11 | Properties | Flow Variables

Row ID	TruePositives	FalsePositives	TrueNegatives	FalseNegatives	Recall	Precision	Sensitivity	Specifity	F-meas...	Accuracy	Cohen's kappa
M	8330	1062	8071	1021	0.891	0.887	0.891	0.884	0.889	?	?
F	8071	1021	8330	1062	0.884	0.888	0.884	0.891	0.886	?	?
Overall	?	?	?	?	?	?	?	?	?	0.887	0.775

Copyright © 2017 KNIME.com AG

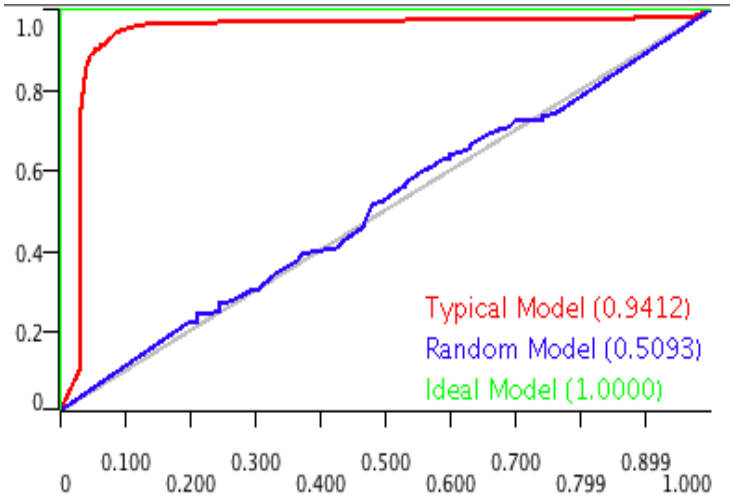
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

Receiver Operating Characteristics

- Sort by confidence in target class
- Plot true positive rate vs false positive rate
- Ideal models achieve 100% TPR with 0% FPR
- Area under the curve indicates model quality

(1=ideal model, 0.5 = random outcome)

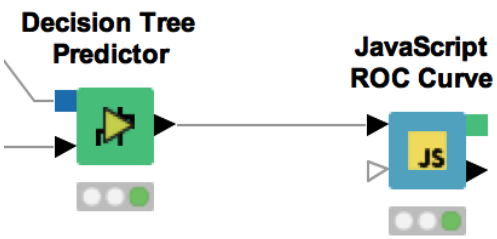


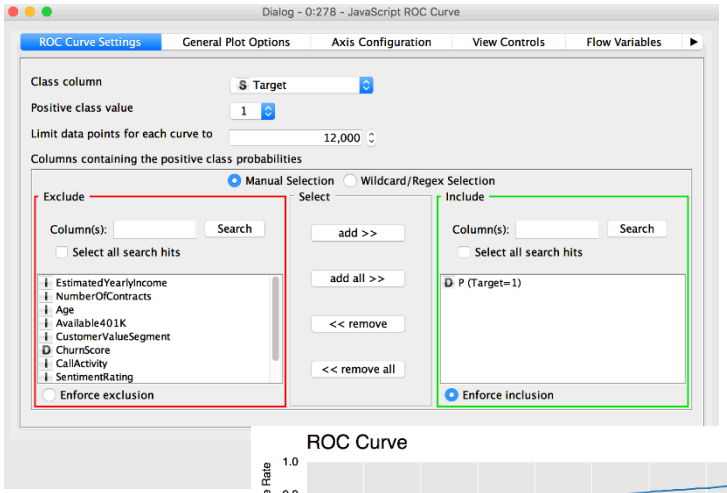
Copyright © 2017 KNIME.com AG

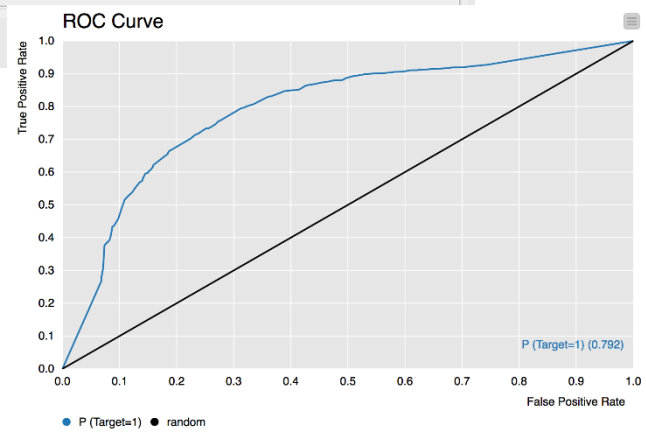
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

New Node: ROC Curve

- Requires individual class probabilities from a preceding predictor
- User must define:
  1. Original class column
  2. Positive class value
  3. Probability for that class from 1 or more models
- See also the JavaScript ROC Curve node







Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

Data Mining Exercise, Activity I

Starting with exercise: *Data Mining, Activity I*:

- Partition the fully joined data
  - 50%, Stratified Sampling
- Train a decision tree on the training data
  - (Learn against "Target" column)
- Use the model to predict the upsell potential for remaining records.
- Evaluate the quality of a model with a Scorer.
- Optional: Find AUC for the model using ROC curve.

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Regression

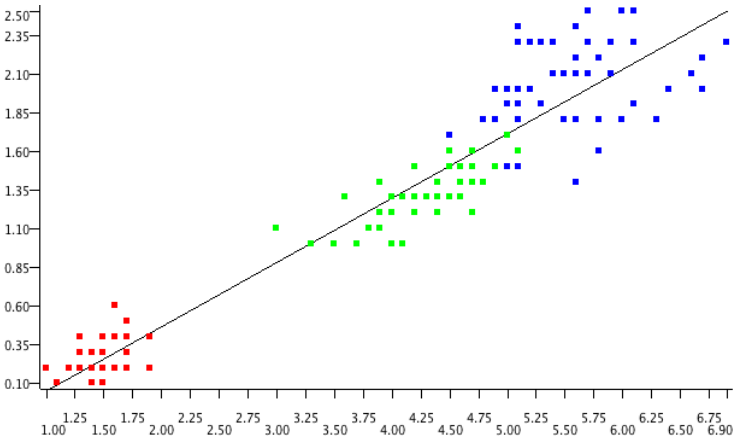
Predict *numeric* outcomes on existing data (supervised)

Applications

- Forecasting
- Quantitative Analysis

Methods

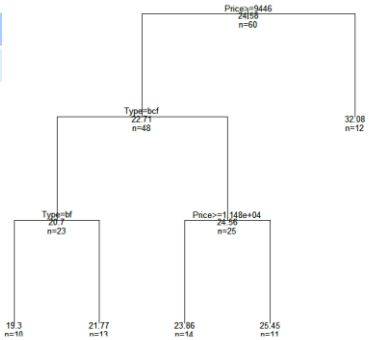
- Linear
- Polynomial
- Regression Trees
- Partial Least Squares



Statistics on Linear Regression

Variable	Coeff.	Std. Err.	t-value	P>|t|
Petal.Length	0.4158	0.0096	43.3872	0.0
Intercept	-0.3631	0.0398	-9.1312	4.44E-16

Multiple R-Squared: 0.9271
Adjusted R-Squared: 0.9266



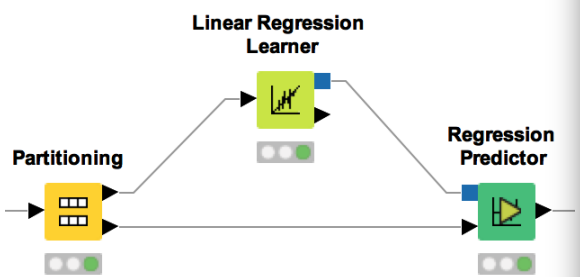
Copyright © 2017 KNIME.com AG

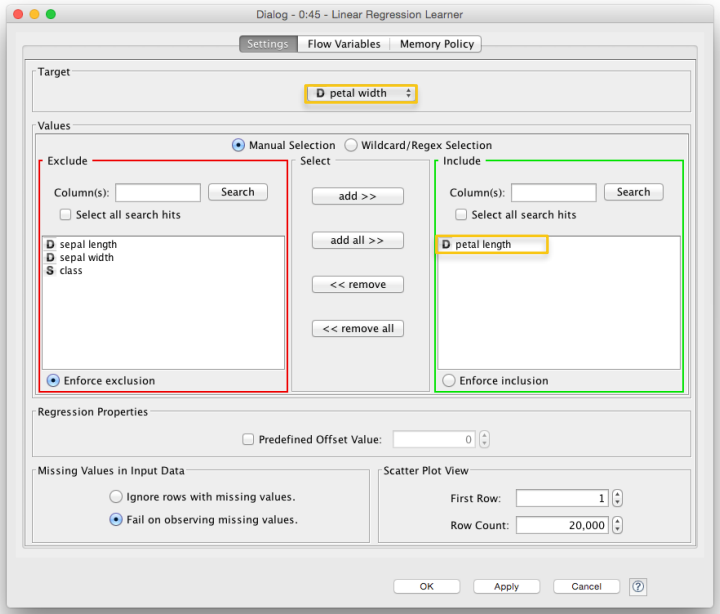
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

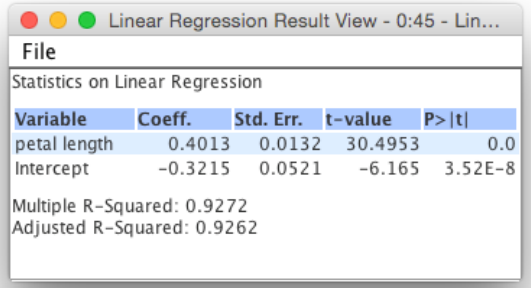
Open for Innovation ® KNIME

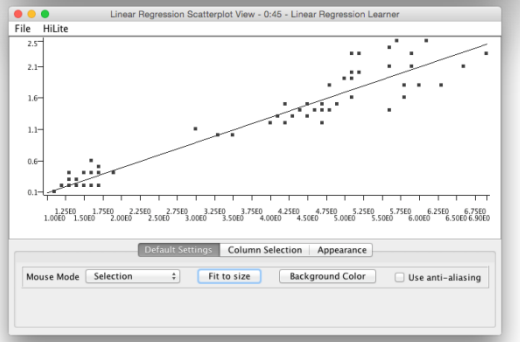
New Nodes: Linear Regression Learner & Regression Predictor

- A linear model relating a dependent variable to 1 or more independent variables.
  - Model coefficients provided in 2nd output port
  - Also available: Polynomial and Tree Ensemble Regression nodes









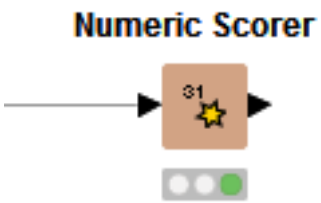
Copyright © 2017 KNIME.com AG

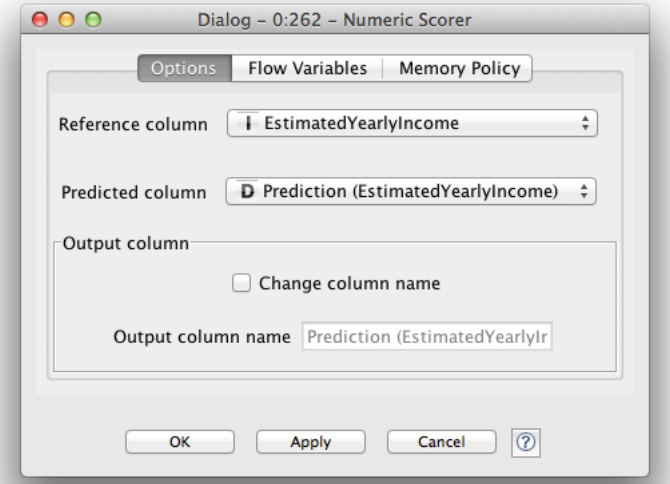
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

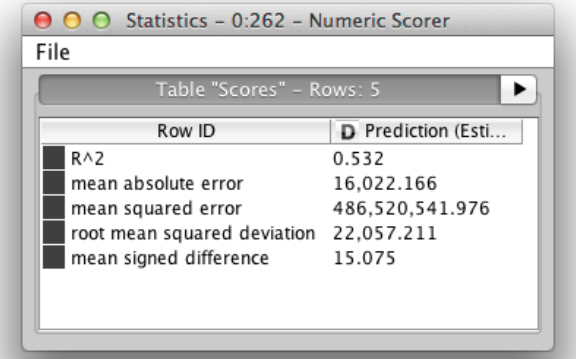
New Node: Numeric Scorer

Similar to scorer node, but for nodes with *numeric* predictions (e.g. linear/polynomial regression)

- Compare dependent variable values to predicted values to evaluate goodness of fit.
- Report R^2, RMSD, SEM etc.







Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

Data Mining Exercise, Activity II

Starting with exercise: *Data Mining, Activity II*:

- Partition the fully joined data
  - 50%, Stratified Sampling
- Train a linear regression model that predicts age as a function of some other parameters in the data set
- Use the model to predict the age of the remaining users
- Evaluate the quality of a model with a Numeric Scorer.
- Is this model useful for predicting customer age?

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Clustering

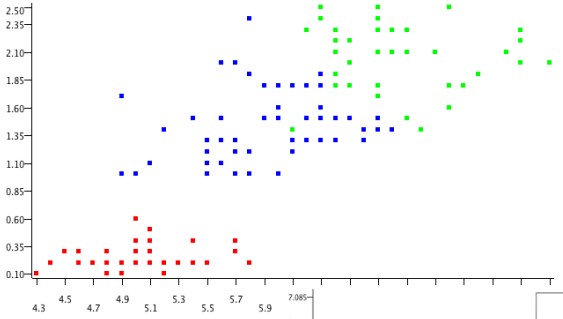
Discover hidden structure in unlabeled data (unsupervised)

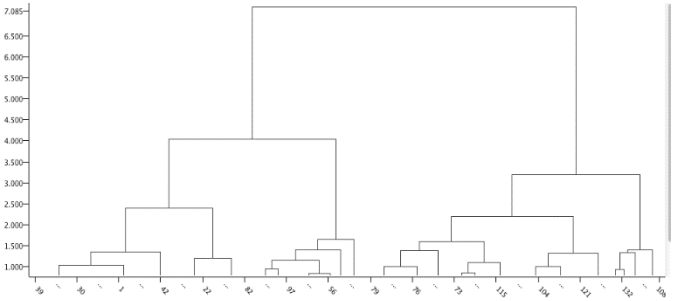
Applications

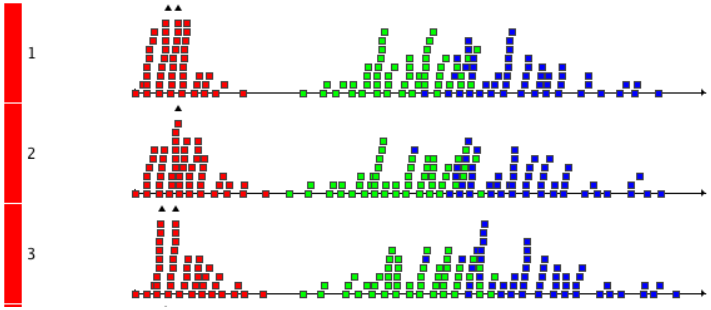
- Market Segmentation
- Diversity picking

Methods

- K-means/medoids
- Hierarchical
- DBScan
- Neighbourgrams







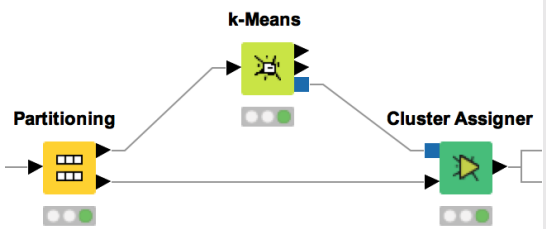
Copyright © 2017 KNIME.com AG

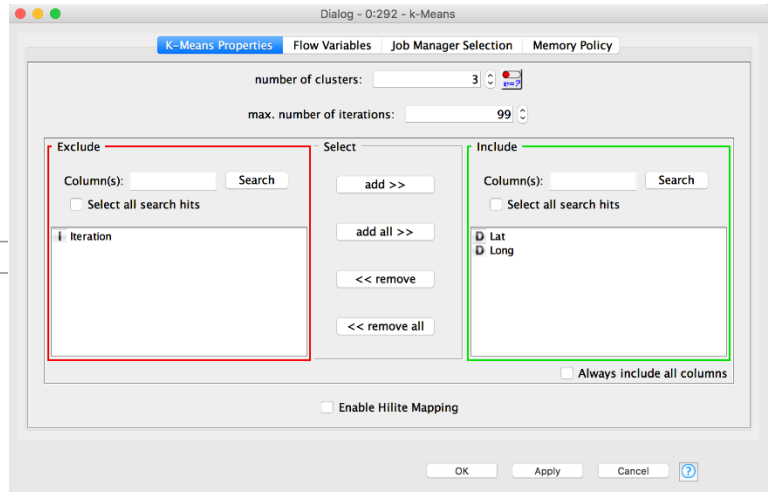
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

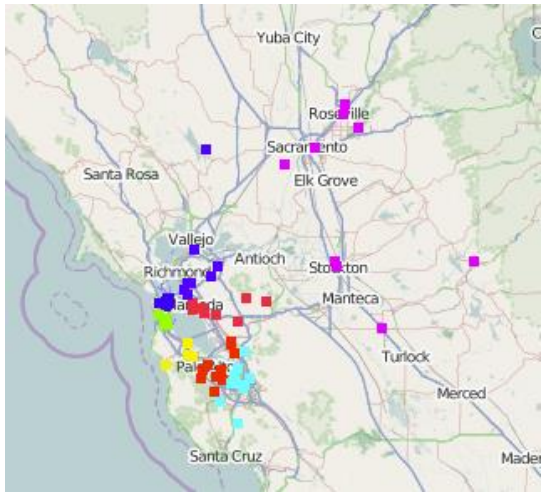
Open for Innovation ® KNIME

New Nodes: k-Means Clustering

- Looks at n observations to define the means for k clusters.
- Each observation is then assigned to its closest cluster center.
- You must provide k.







Copyright © 2017 KNIME.com AG

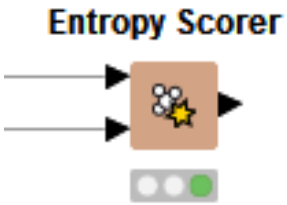
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

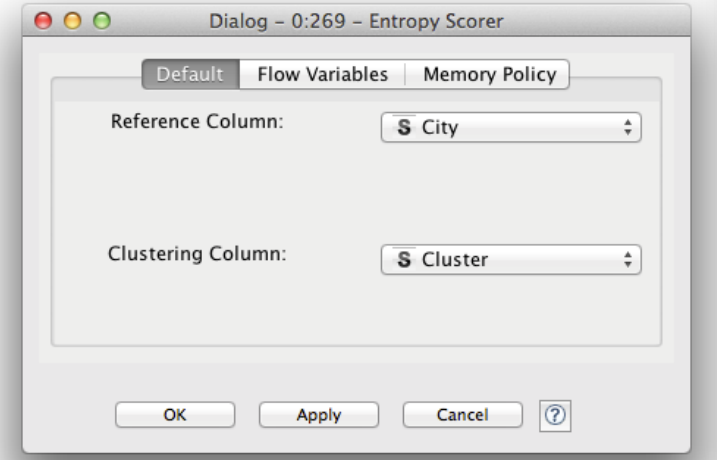
Open for Innovation ® KNIME

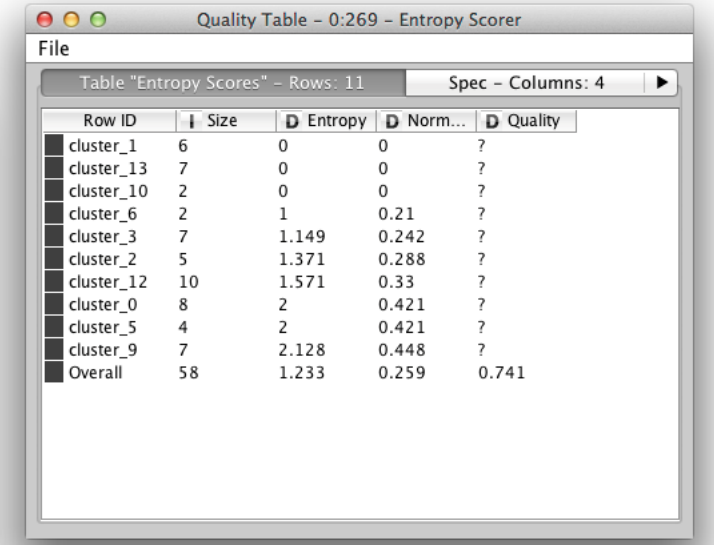
New Node: Entropy Scorer

- Similar to scorer node, but used with unsupervised learning (no target to predict)
  - Cluster labels and reference clusters do not need to be in the same domain (e.g. Match “Cluster 1” to “iris setosa”)
  - Reports entropy based statistics which indicate model quality

    (low entropy, high quality is the aim)







Row ID	I Size	D Entropy	D Norm...	D Quality
cluster_1	6	0	0	?
cluster_13	7	0	0	?
cluster_10	2	0	0	?
cluster_6	2	1	0.21	?
cluster_3	7	1.149	0.242	?
cluster_2	5	1.371	0.288	?
cluster_12	10	1.571	0.33	?
cluster_0	8	2	0.421	?
cluster_5	4	2	0.421	?
cluster_9	7	2.128	0.448	?
Overall	58	1.233	0.259	0.741

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Data Mining Exercise, Activity III

Start with exercise: *Data Mining, Activity III*

- Read the location_data.table file
- Filter to entries from California (region_code = CA)
- Train a k-means model with k=3. Use only position data for clustering (latitude and longitude)
- Evaluate with Entropy Scorer. Compare cluster labels to the "City" column. Hint: use the k-means output for both ports in the input to the scorer.
- Optional: Plot latitude and longitude in a view (OSM Map or Scatter Plot) and use that to help you visually optimize k.

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Integrating External Tools

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ®
KNIME

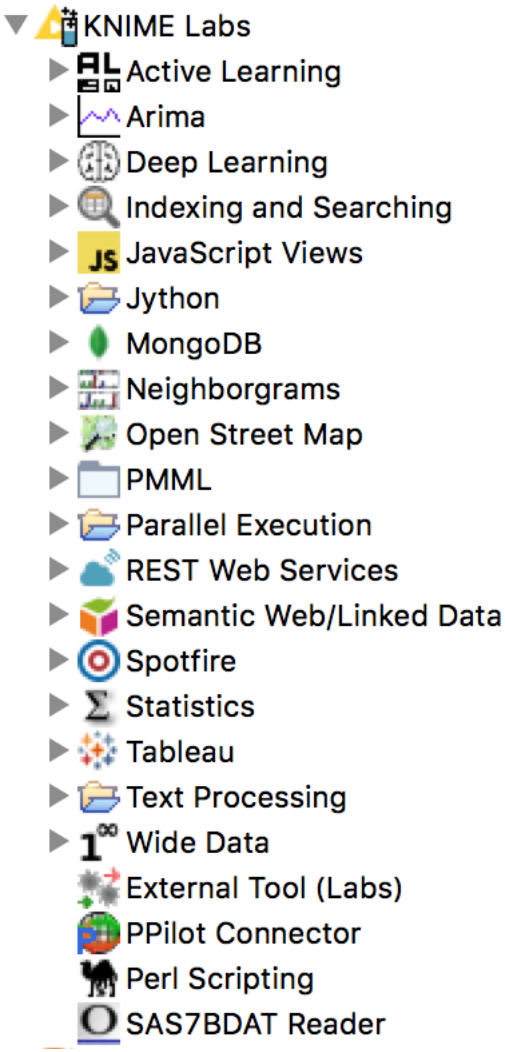
Goal of This Session

This session gives a quick overview of the external tools that can be called within KNIME, e.g.: Java, R, any external tool of choice, web service, and so on.

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

KNIME Labs



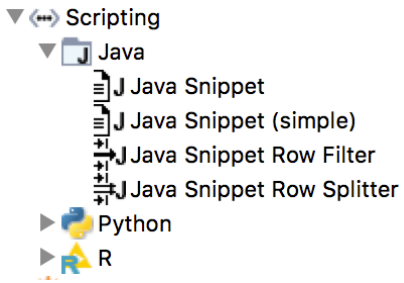
- KNIME Labs enable you to preview new KNIME features and plug-ins that are still under development.
- The nodes provided in KNIME Labs (KNIME Tech) are not (yet) part of the official KNIME version because the functionality and/or API may not be finalized.
- You can get these plug-ins by installing the KNIME Labs extension package. Most of the plug-ins come already with a detailed description and some workflow examples.

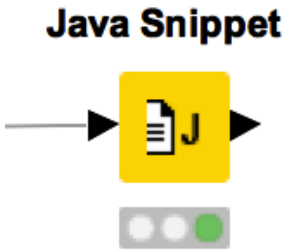
Copyright © 2017 KNIME.com AG

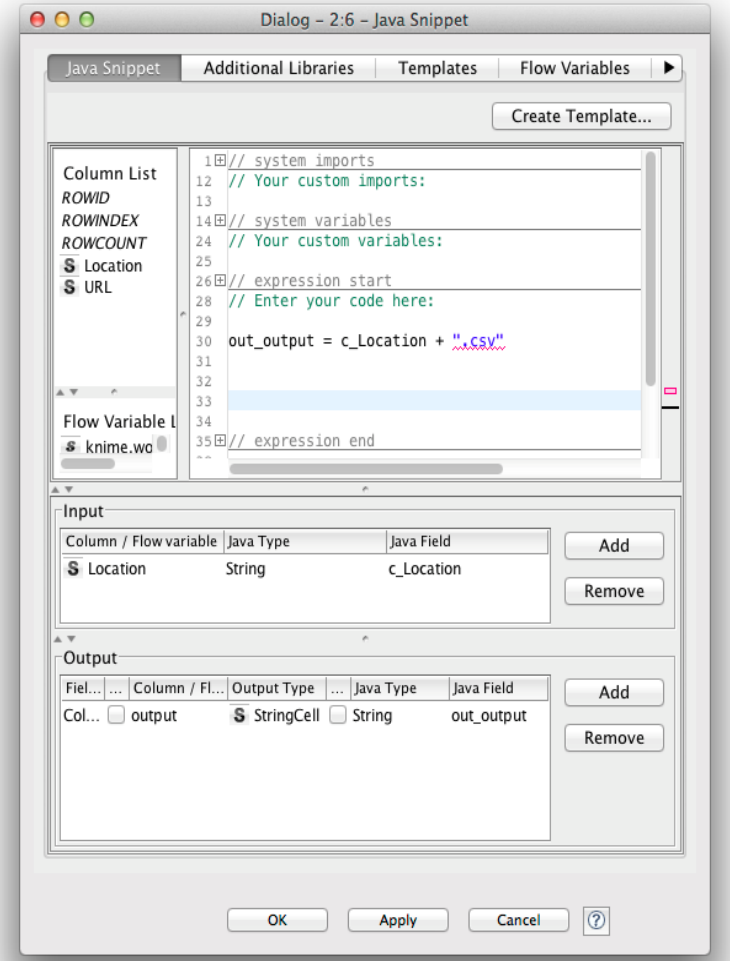
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Java Snippet

- Fastest running scripting node in KNIME
- Syntax highlighting, auto completion, error checking
- Templates allow you to save scripts for later re-use
- Import custom libraries





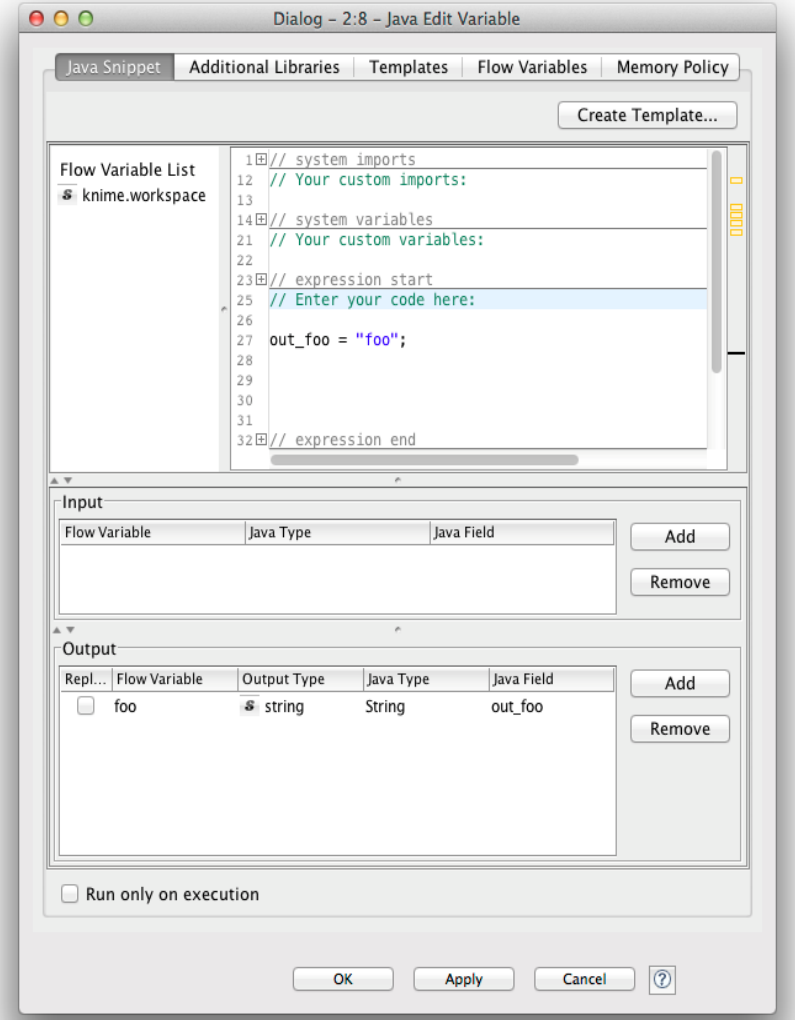


Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Java Edit Variable

Same as Java snippet, but without table ports

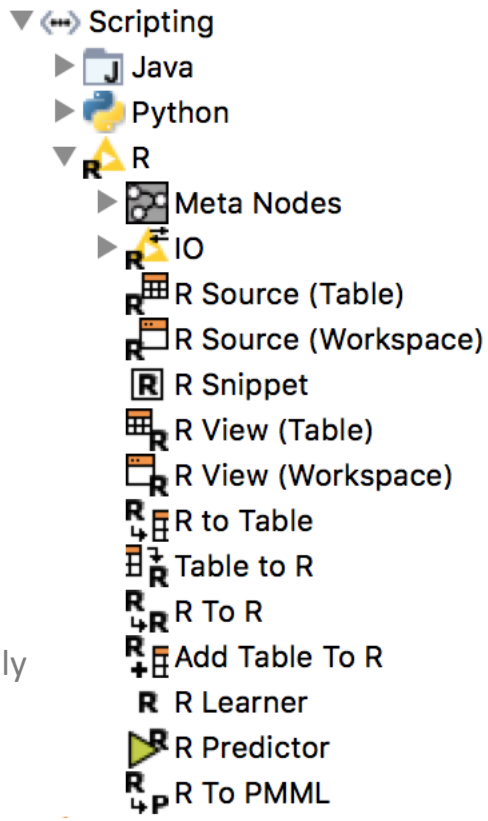


Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

R Integration

- Run R inside KNIME.
- Works with existing R installations.
- Nodes for many tasks, but all with similiar interfaces…
- First run: install.packages(‘Rserve‘) and install.packages(‘Cairo‘)*

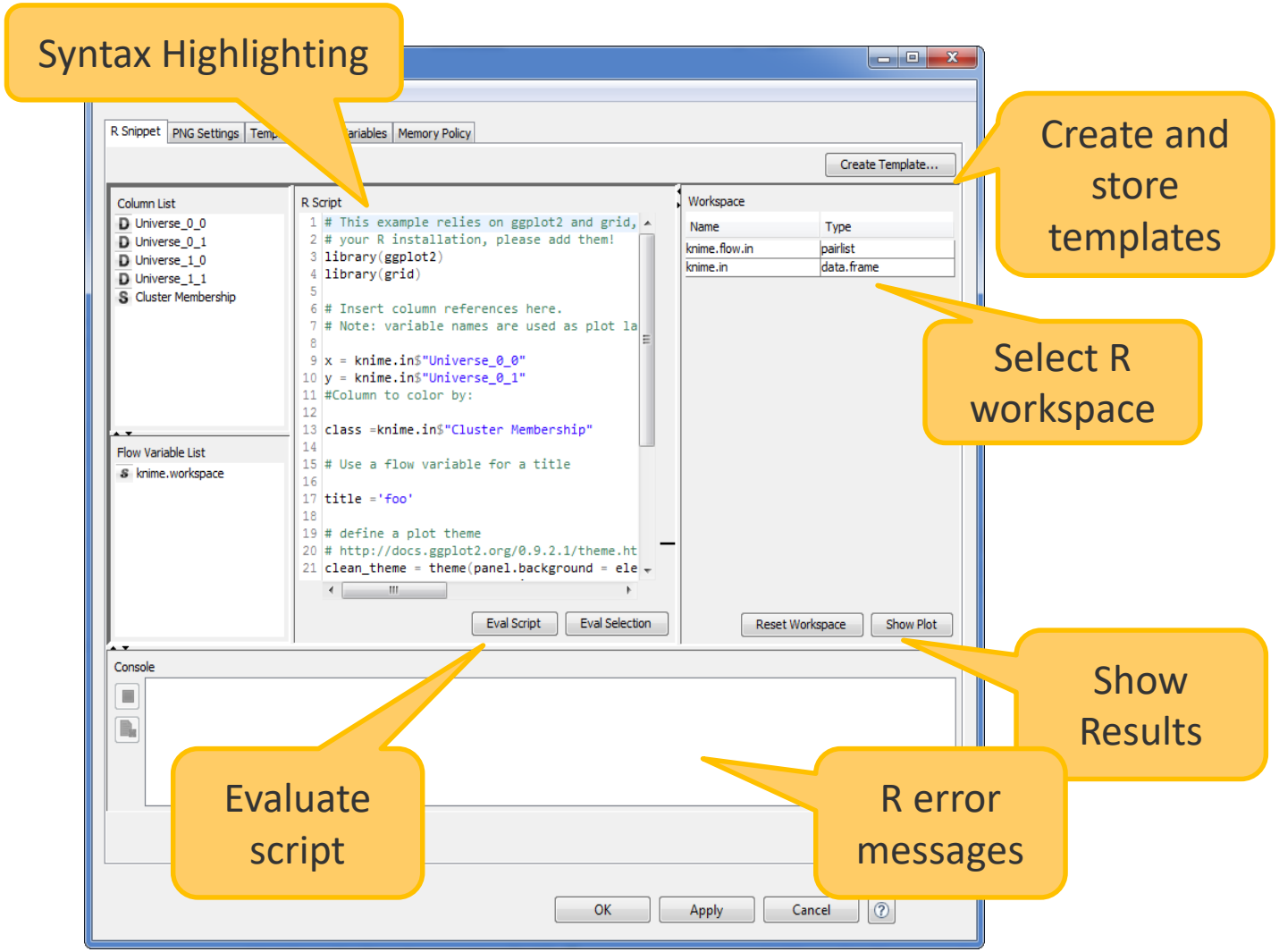
*mac only



Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

R Integration



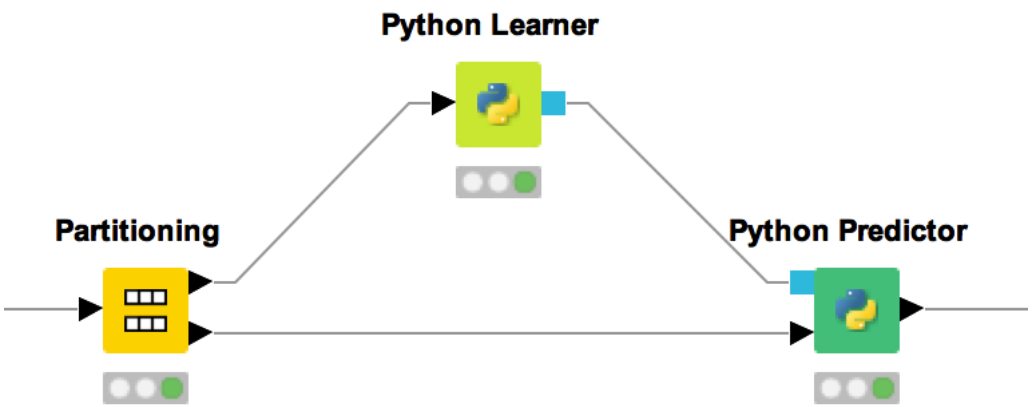
Copyright © 2017 KNIME.com AG

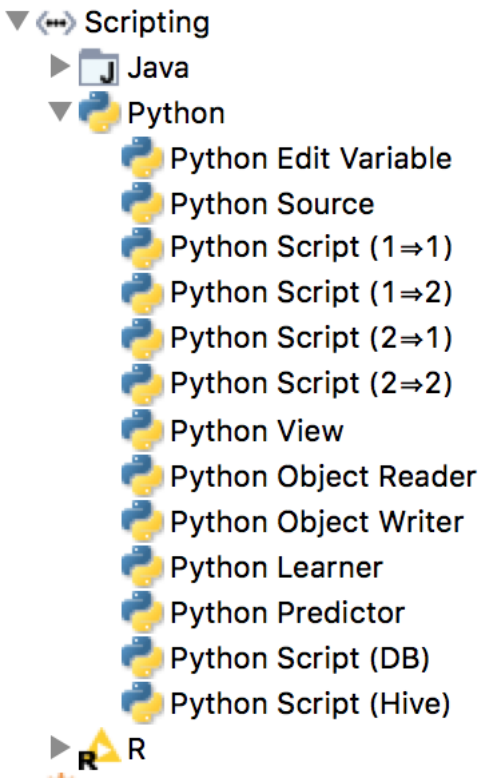
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

Python Integration

- Run Python inside KNIME
- Works with existing installations
- UI modeled after R integration

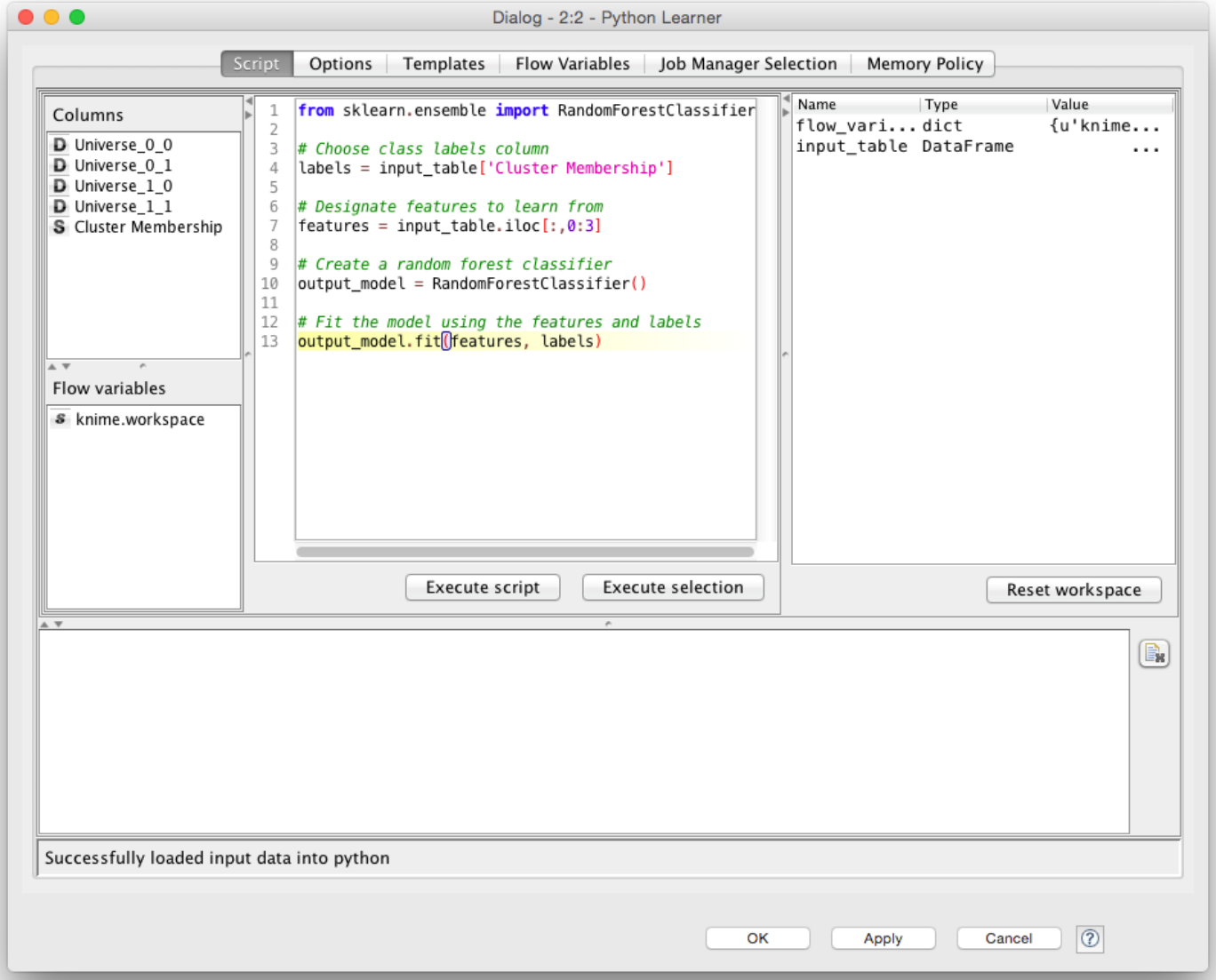




Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Python Scripting UI



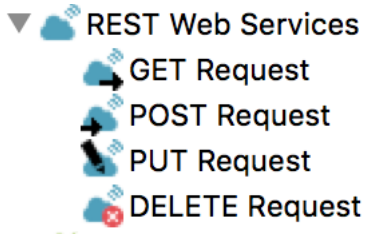
Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

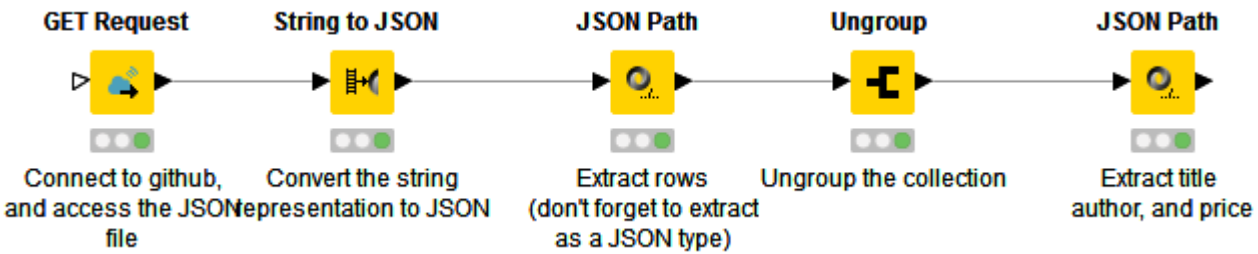
Open for Innovation ® KNIME

RESTful Web Services

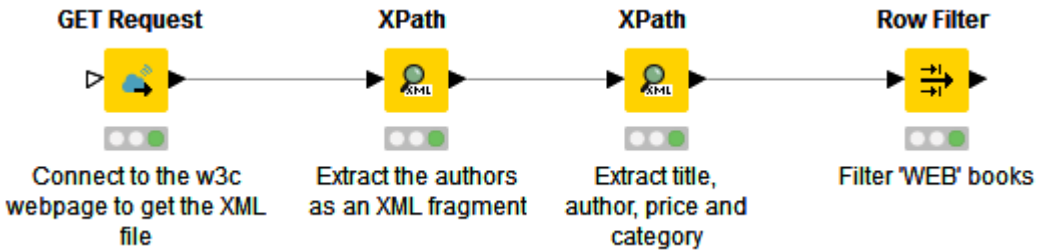
In KNIME Labs:



JSON Response:



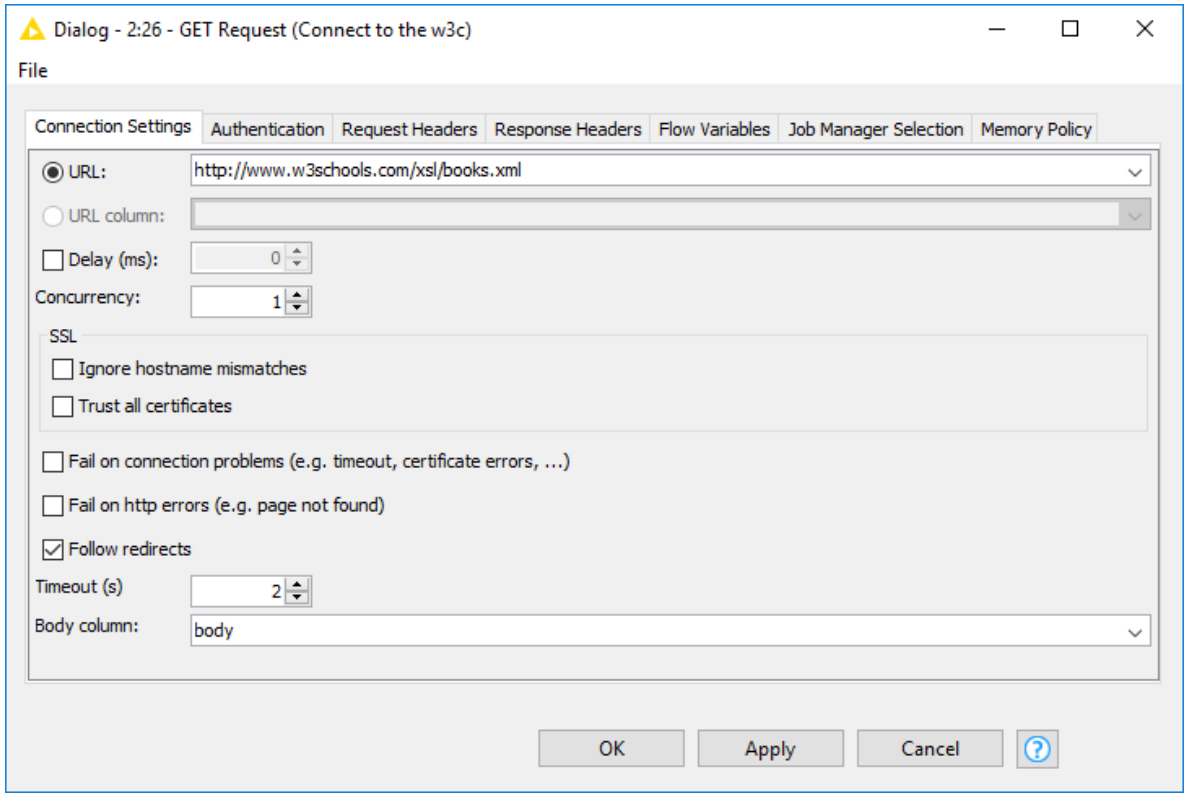
XML Response:



Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

RESTful Web Services



https://www.knime.org/blog/IBM-Watson-meets-Google-API

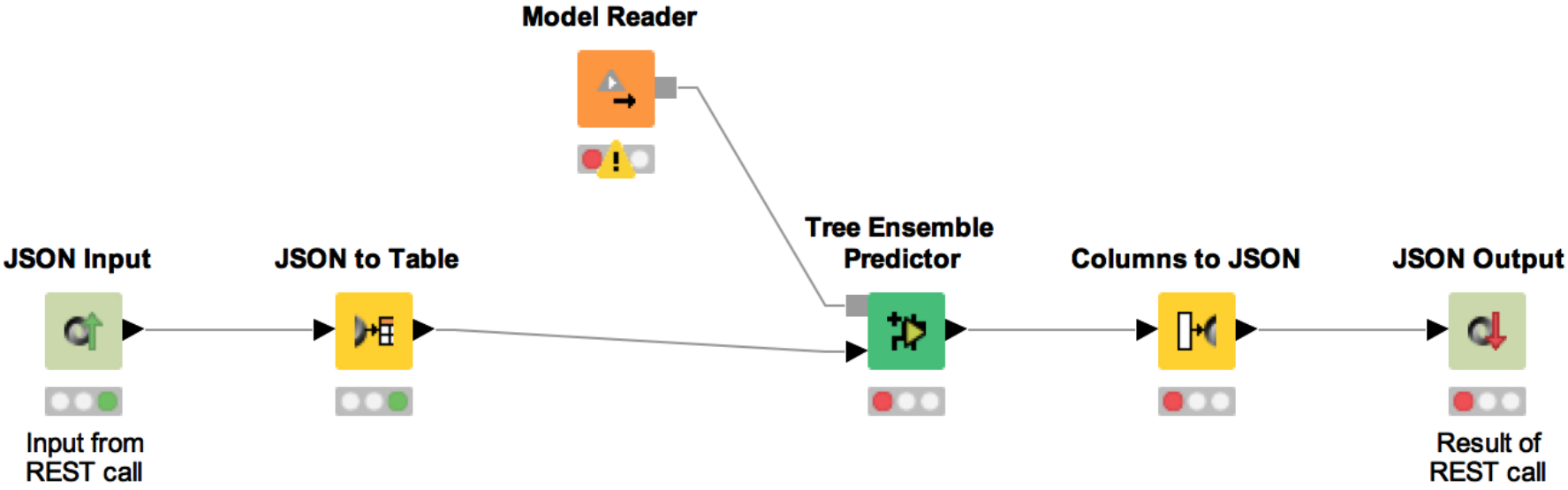
https://www.knime.org/blog/OSM-meets-CSV-file-and-Google-API

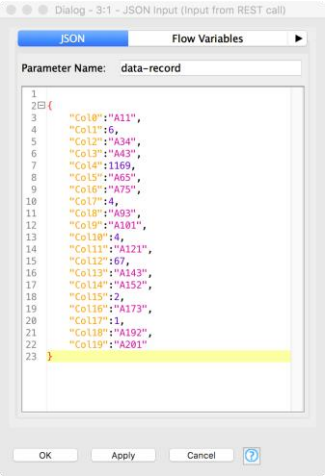
Copyright © 2017 KNIME.com AG

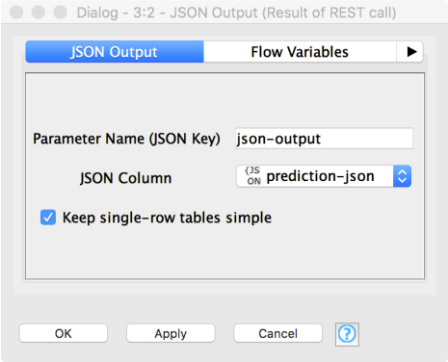
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

KNIME Server as a REST resource







https://www.knime.org/blog/giving-the-knime-server-a-rest

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

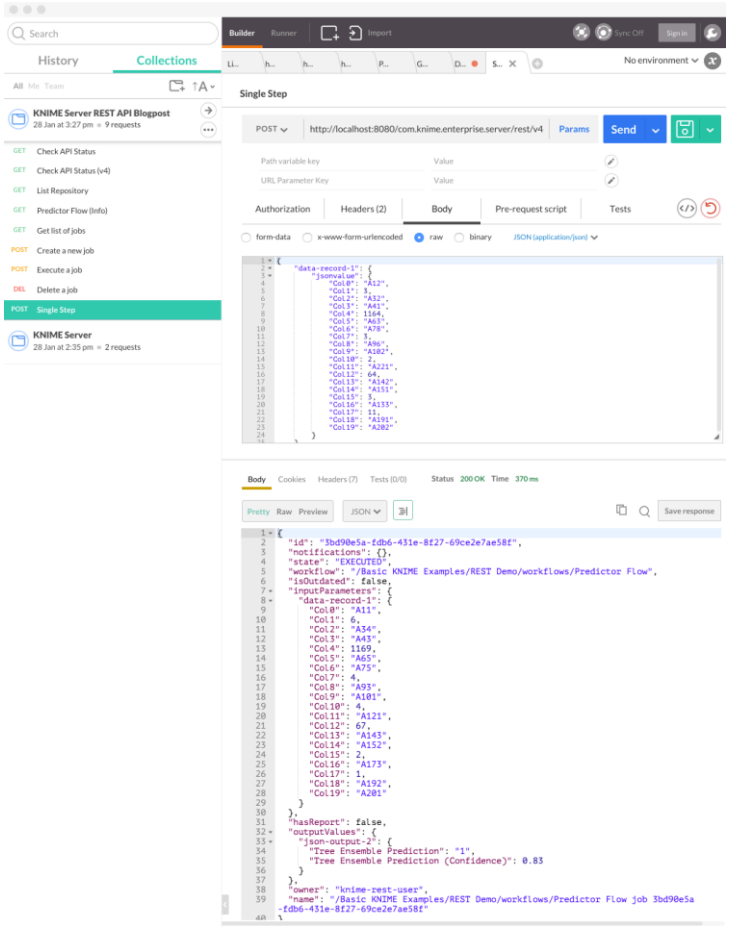
Open for Innovation ® KNIME

KNIME Server as a REST resource

- Use cURL, SOAPUI or Chrome extension Postman to explore the REST API

```
jon — -bash — 80×41
Jons-MacBook-Pro:~ jon$ curl -u 'knime-rest-user:knockknock' -i http://localhost
:8080/com.knime.enterprise.server/rest
HTTP/1.1 200 OK
Server: Apache-Coyote/1.1
Cache-Control: private
Expires: Thu, 01 Jan 1970 00:00:00 UTC
Cache-Control: no-transform,max-age=86400
Date: Thu, 28 Jan 2016 13:27:45 GMT
KNIME-Class: com.knime.enterprise.server.rest.api.ent.ServerVersion
Link: <http://localhost:8080/com.knime.enterprise.server/rest/_profile/knime-ser
ver-doc.xml>;rel="profile"
Content-Type: application/vnd.mason+json;charset=UTF-8
Content-Length: 481

{
  "@controls" : {
    "self" : {
      "href" : "http://localhost:8080/com.knime.enterprise.server/rest/",
      "method" : "GET"
    },
    "knime:v4" : {
      "href" : "http://localhost:8080/com.knime.enterprise.server/rest/v4/",
      "title" : "KNIME Server API v4",
      "method" : "GET"
    }
  },
  "version" : {
    "major" : 4,
    "minor" : 2,
    "revision" : 2
  },
  "@namespaces" : {
    "knime" : {
      "name" : "http://www.knime.com/server/rels#"
    }
  }
}
Jons-MacBook-Pro:~ jon$
```



Copyright © 2017 KNIME.com AG

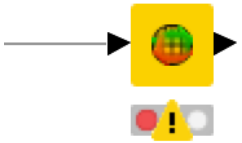
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

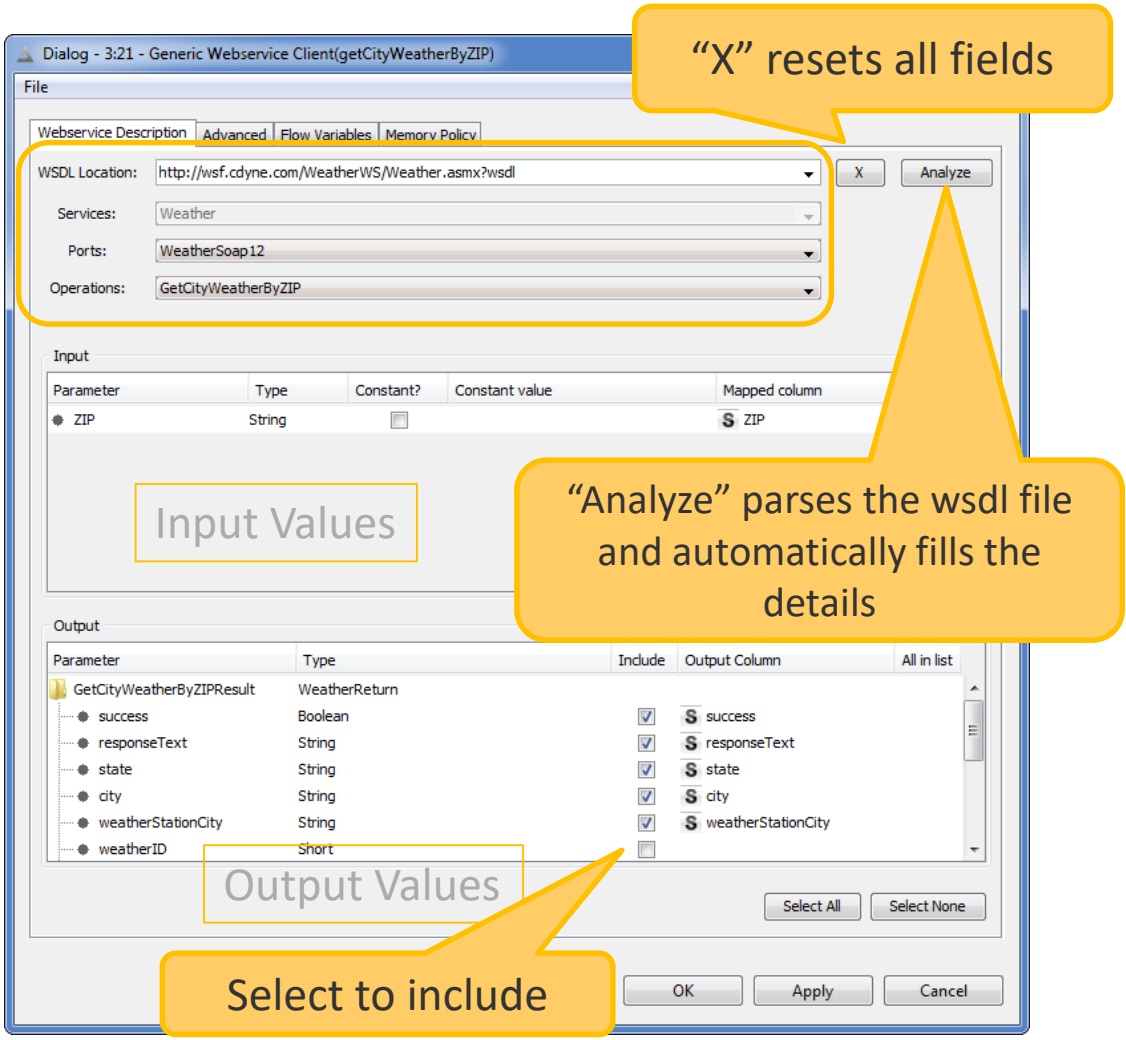
Open for Innovation ® KNIME

Generic Web Service Client (SOAP)

- Tool Integration
  - Spark
  - External SSH Tool
  - External Tool
  - Generic Web Service Client

Generic Web Service Client



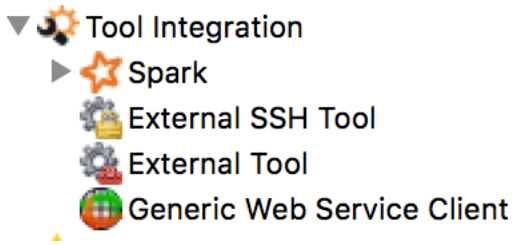


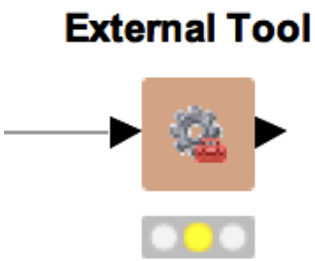
Copyright © 2017 KNIME.com AG

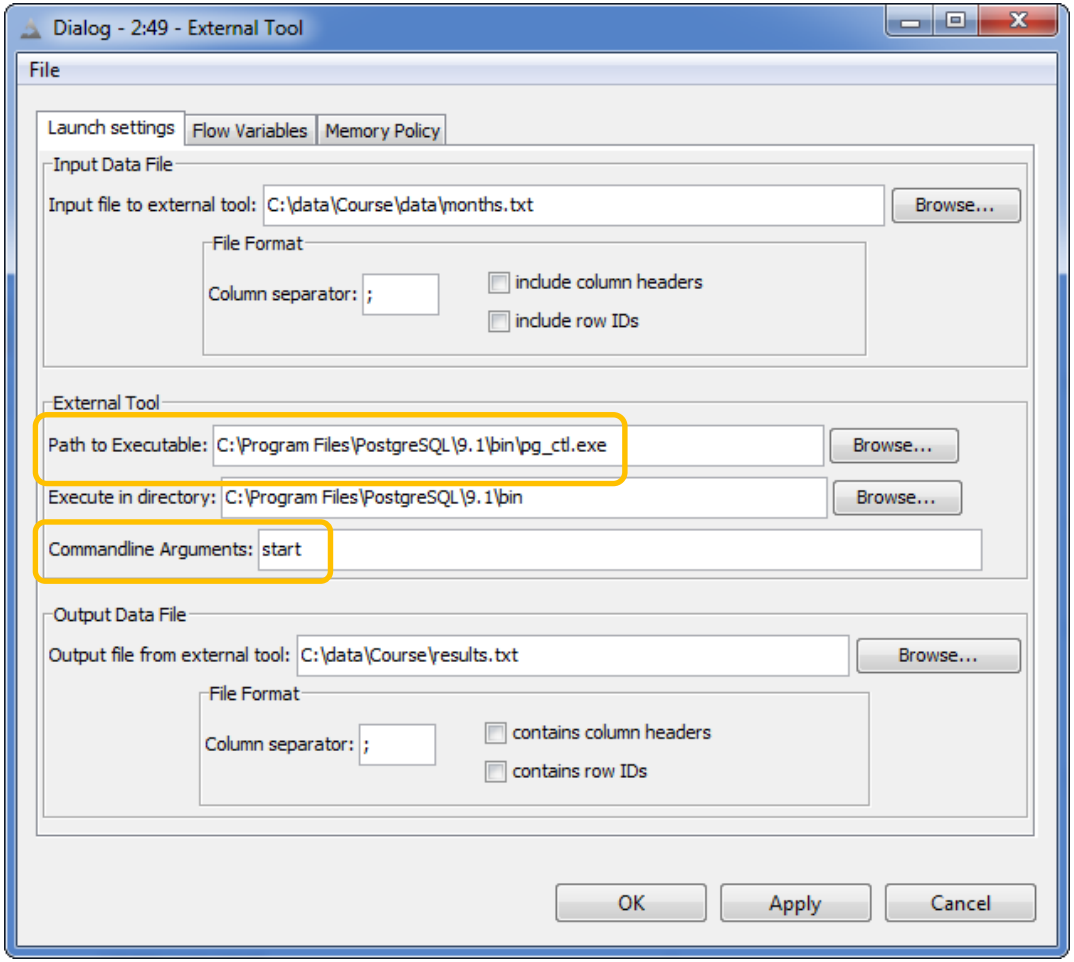
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

External Tool Node







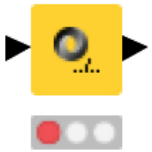
Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

JSON and JSONPath

JSON Reader

JSON Path



- Use the JSON Reader (or the GET Resource) nodes to get an JSON cell
- Use JSONPath nodes to ‘query’ the JSON and extract certain parameters
- Editor window simplifies construction of JSON queries by auto-generating them (click on properties)

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

JSONPath

Dialog - 2:6 - JSON Path (Extract rows)

Settings | Flow Variables | Job Manager Selection | Memory Policy

Input

Representation

Remove source column

Outputs

Output column

rows

Add single query

JSON-Cell Preview

```
1 [ {
2   "id" : "978-0641723445",
3   "cat" : [ "book", "hardcover" ],
4   "name" : "The Lightning Thief",
5   "author" : "Rick Riordan",
6   "series_t" : "Per
7   "sequence_i" : 1,
8   "genre_s" : "fant
9   "inStock" : true,
```

Undo
Can't Redo
Cut
Copy
Paste
Delete
Select All
Folding
Add JSONPath

Edit

JSONPath: $[0]['author']

Column name: author

Result type: String

Result is list

Return the paths instead of values

OK | Cancel

OK | Apply | Cancel

No destination location provided! Please enter

1083M of 1334M

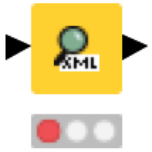
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

XML and XPath

XML Reader

XPath



- Use the XML Reader (or the GET Resource) nodes to get an XML cell
- Use XPath nodes to ‘query’ the XML and extract certain parameters
- Editor window simplifies construction of XPath queries by auto-generating them (click on XML elements)

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

XPath

Dialog - 13:46 - XPath

Settings | Namespace | Flow Variables

XML column: XML XML

Remove source column.

XPath summary

Column name	XPath query	Type
author	/book/author	String(SingleCell)

Selected XPath: /book/author

Add XPath | Edit XPath | Remove XPath

XML-Cell Preview

```
1  <?xml version="1.0" encoding="UTF-8"?>
2  <book year="2001">
3      <author>Vincent Bugliosi</author>
4      <title>The Betrayal of America: How the Supreme Court Ur
5      <category>About Elections</category>
6      <isbn>1-56025-355-X</isbn>
7  </book>
8
```

OK | Apply | Cancel

XPath Query Settings

Column name:

New column name: author

XPath query for column name: (relative to value query) name

XPath value query

/book/author

Return type

XPath data type: String cell

Options:

Return missing cell on empty string.

Multiple tag options

Single Cell | Collection Cell

Multiple Columns | Multiple Rows

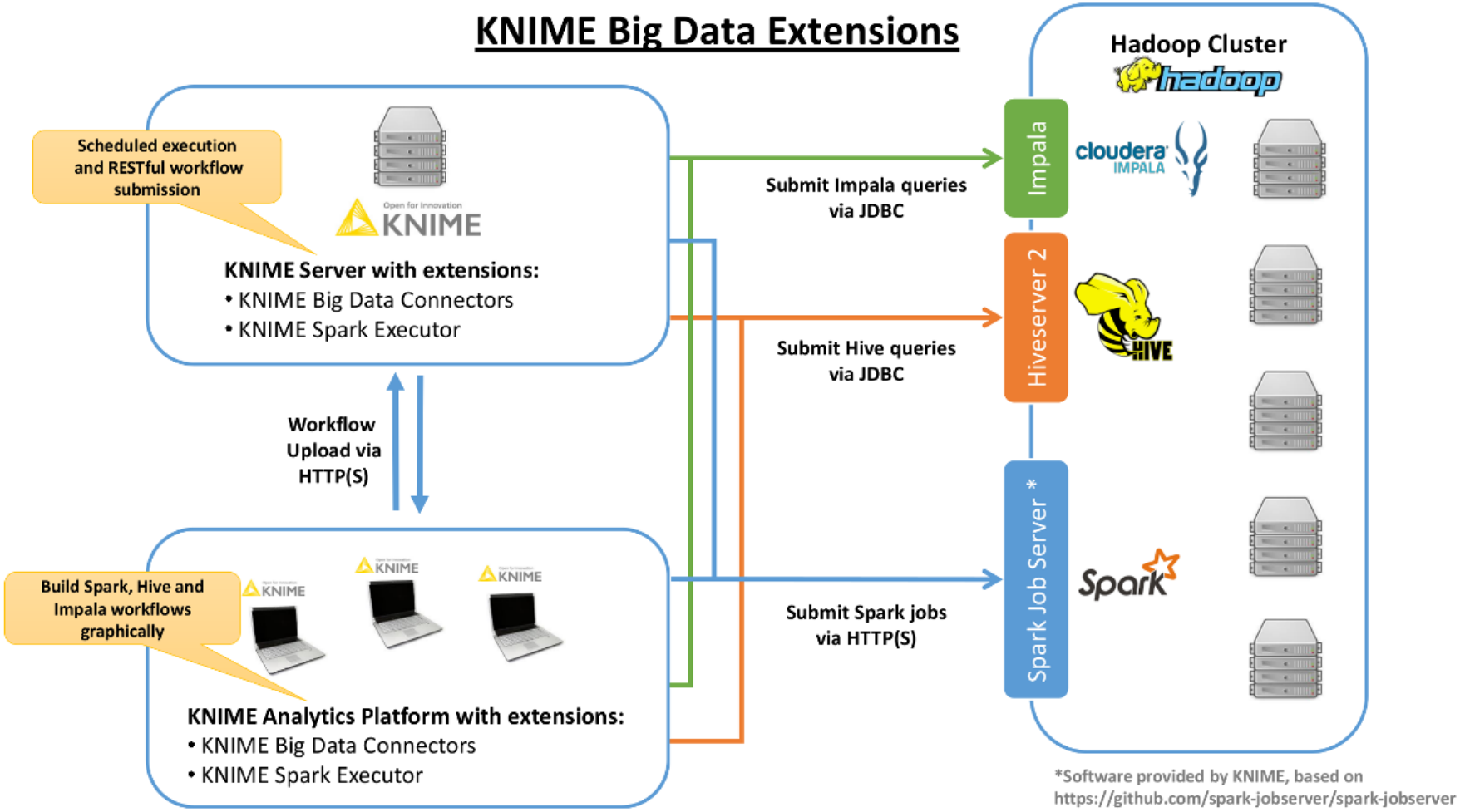
Ok | Cancel

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

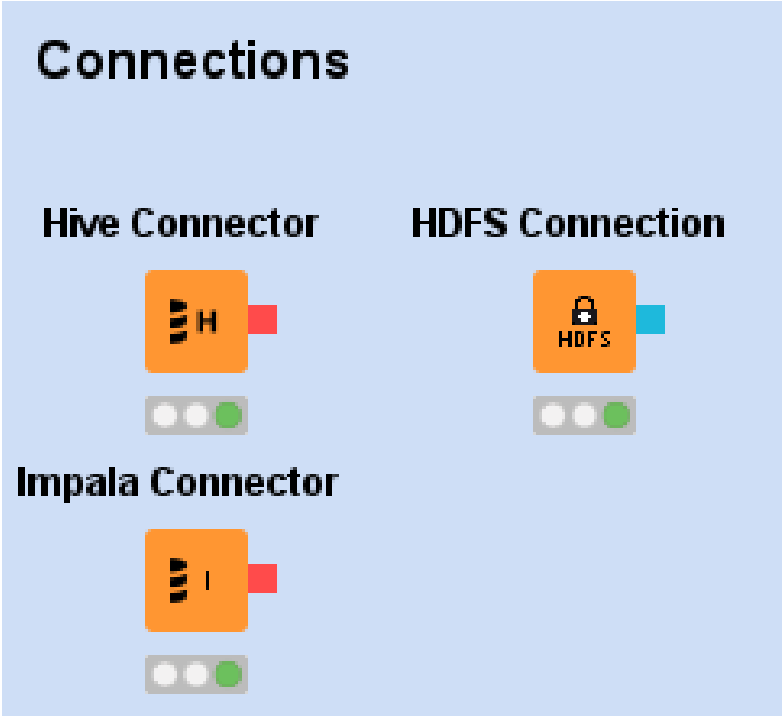
Open for Innovation KNIME

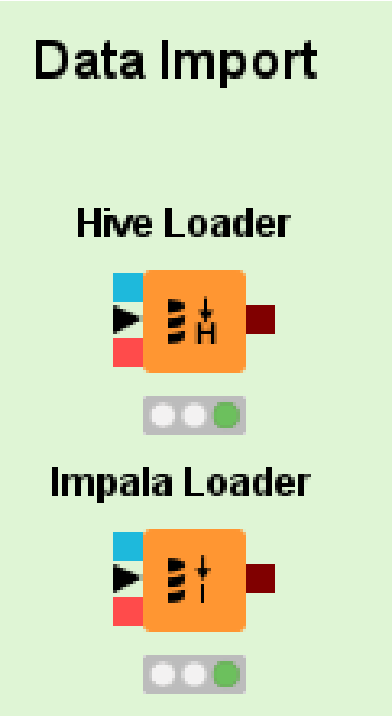
Hive, HDFS, Spark Architecture

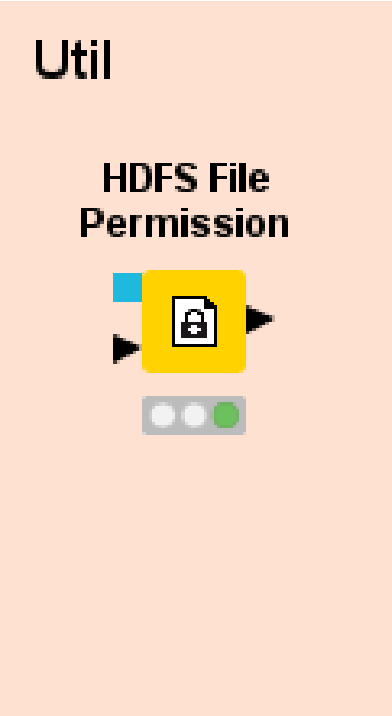


Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Big Data Connector Extensions





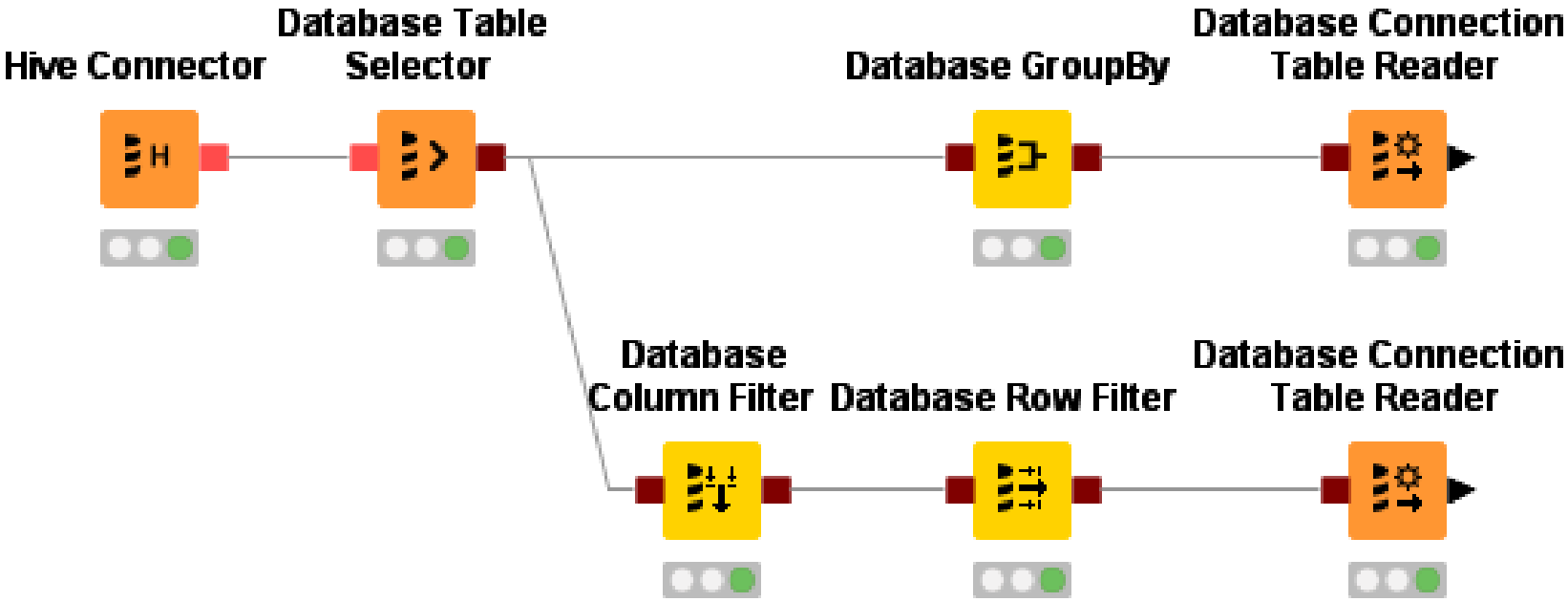


Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

Query Big Data Platform with Hive

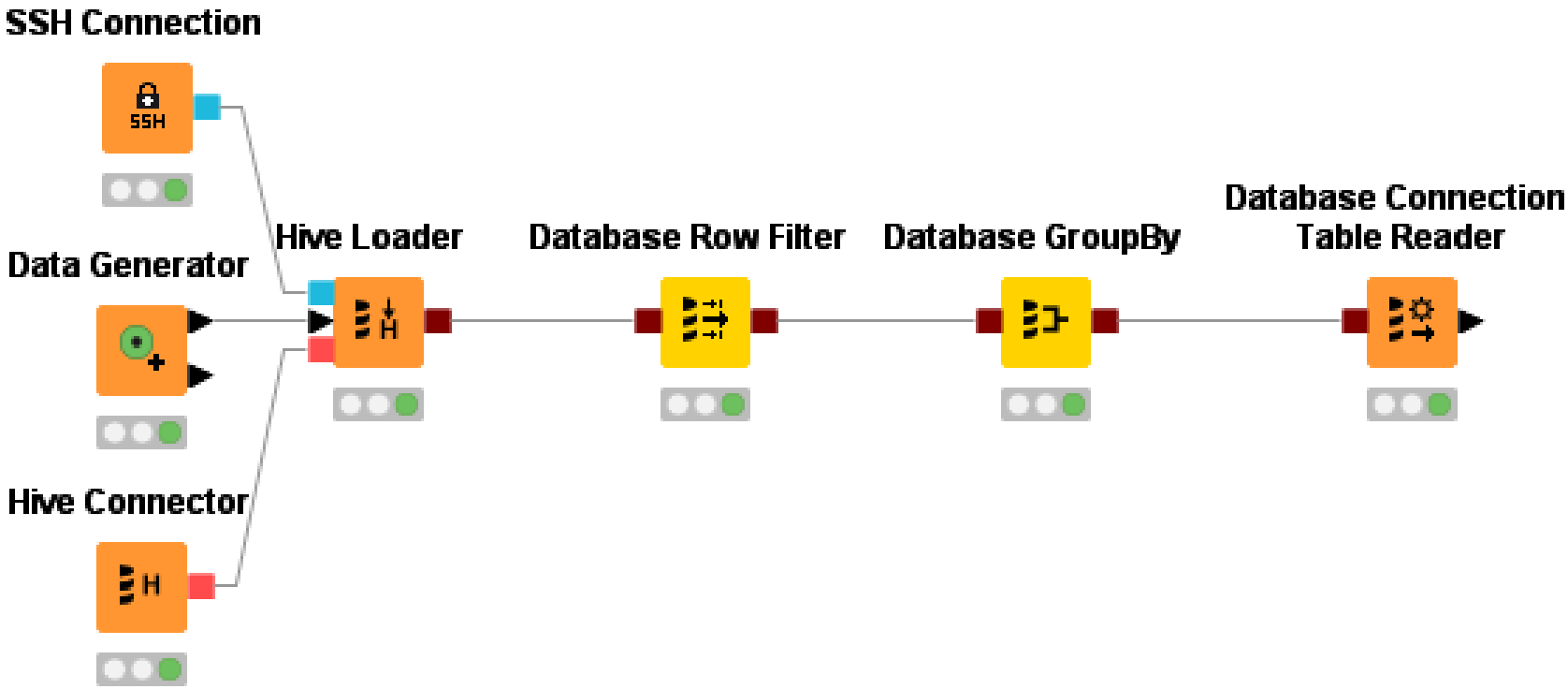


Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

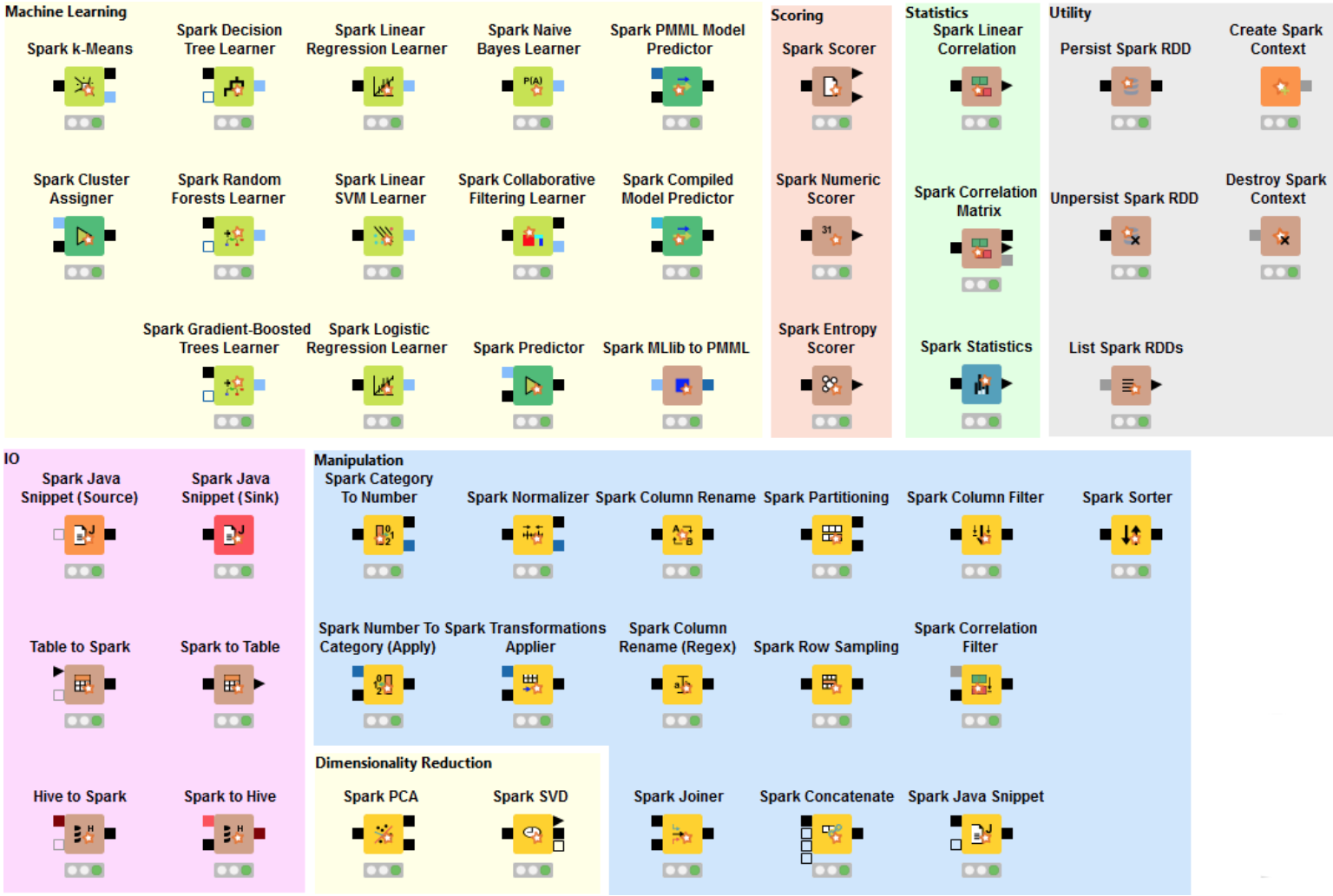
Loading Data to Hive



Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Spark Functionality

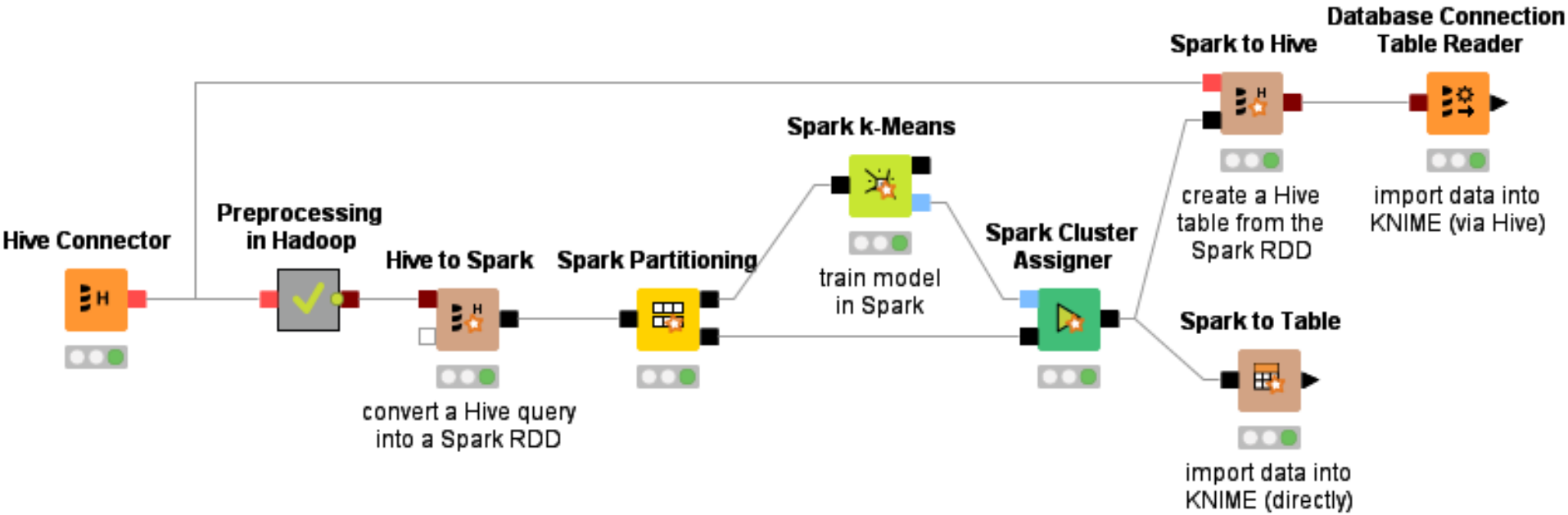


Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

Machine Learning in Spark



Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

Exercises

- Use the REST nodes to call an external web service
- Choose either books.json or books.xml and use the appropriate tools to extract the book name, author and title
- Programmers: choose your favorite from Java, Python and R
  - Take the KNIME table and multiply all values in one column by 10.
  - Python/R: Build a simple decision tree learner
  - Alternatively use the Generic JavaScript node and D3.js to create a custom JavaScript visualization that can be viewed in KNIME Analytics Platform, and KNIME WebPortal.

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Exporting Data & Deployment

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ®
KNIME

Exporting Data

After an analysis is completed, what next?

- Write results to a file
- Create/update a database
- Save the model for use elsewhere
- Generate a rich report
- Deploy via KNIME WebPortal
- Deploy via workflow as RESTful web service

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Input/Output in Deployment

Input

- File (CSV, Table, XLS, …)
- Database
- JSON for REST API

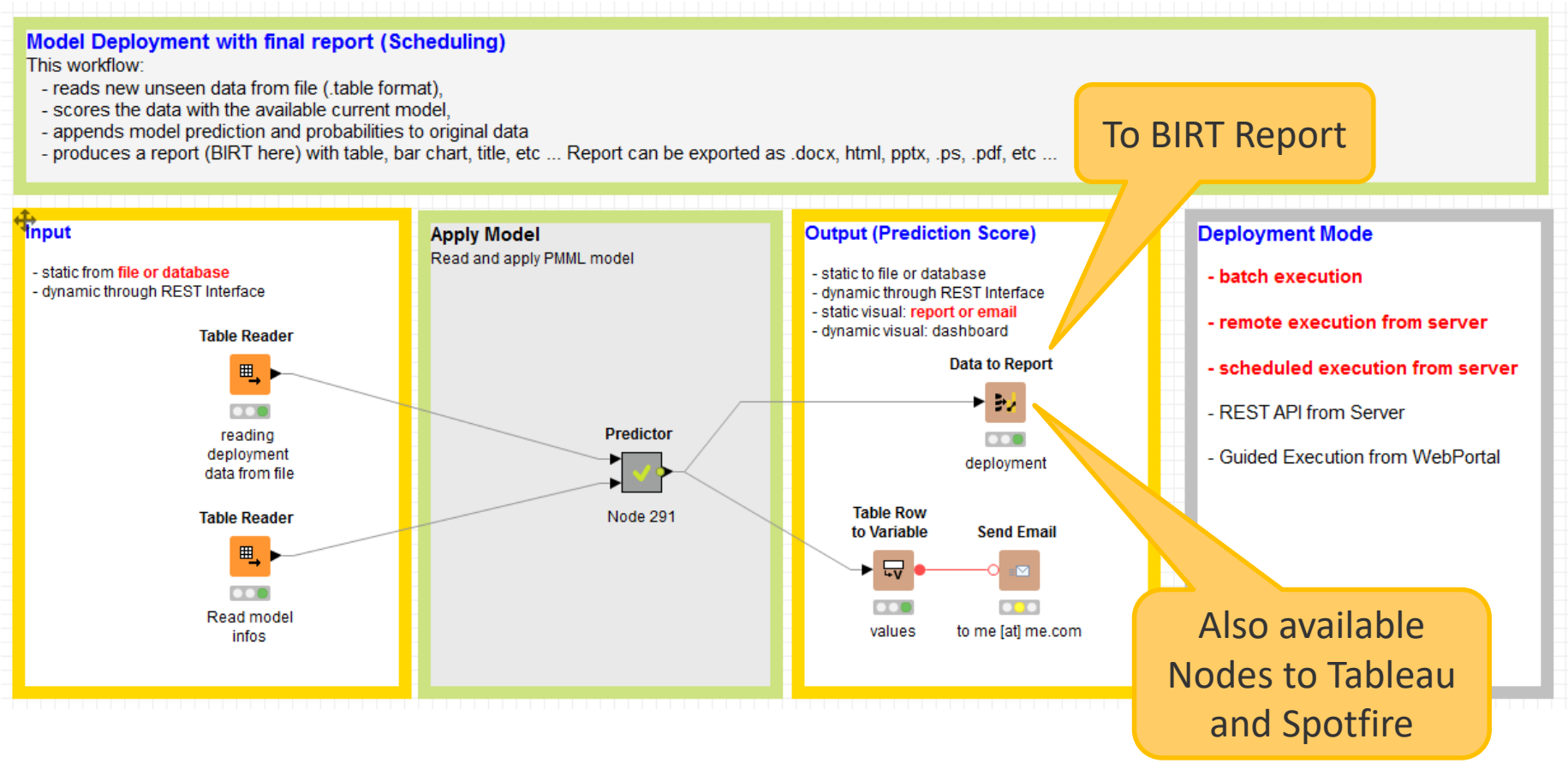
Output

- Report (BIRT, Tableau, Spotfire)
- Email
- File (CSV, table, XLS, …)
- Dashboard on WebPortal

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

To Report / Email

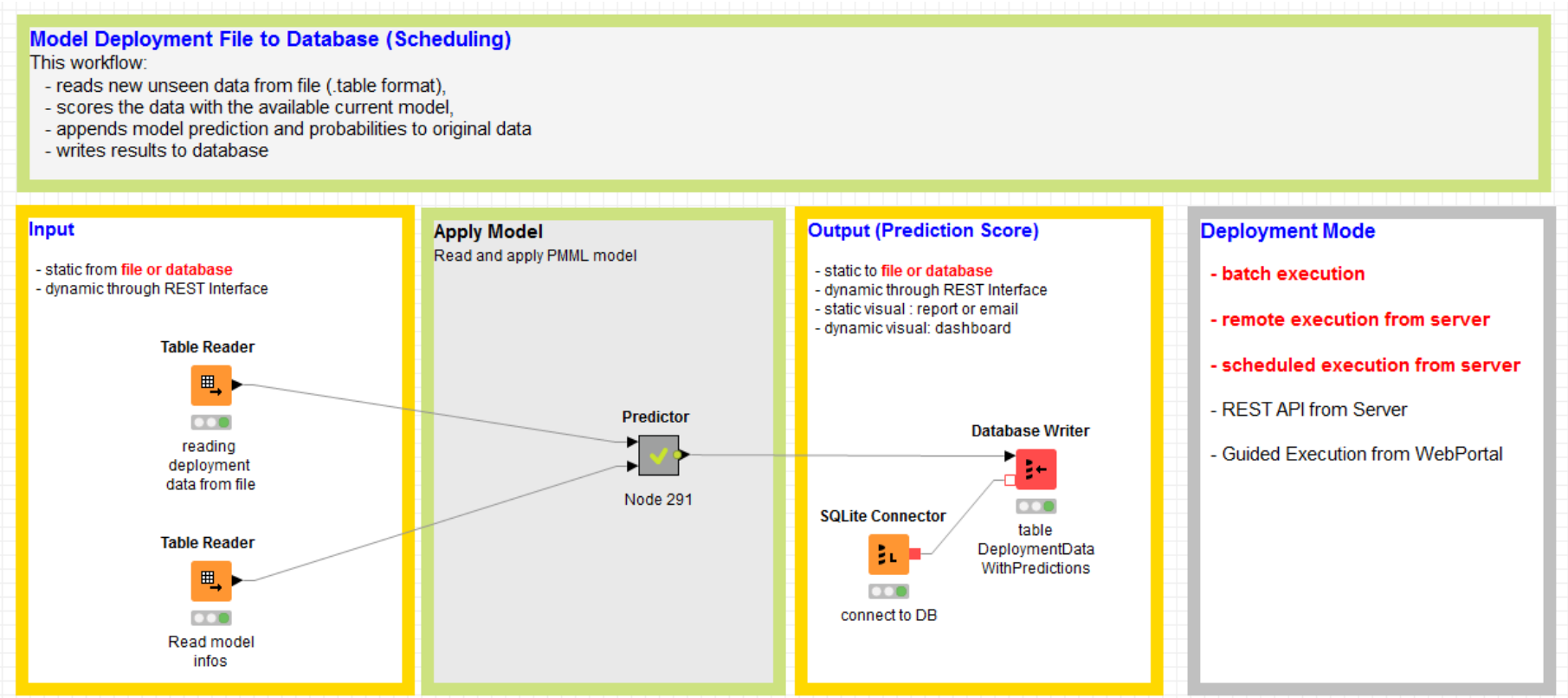


Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

To File / Database

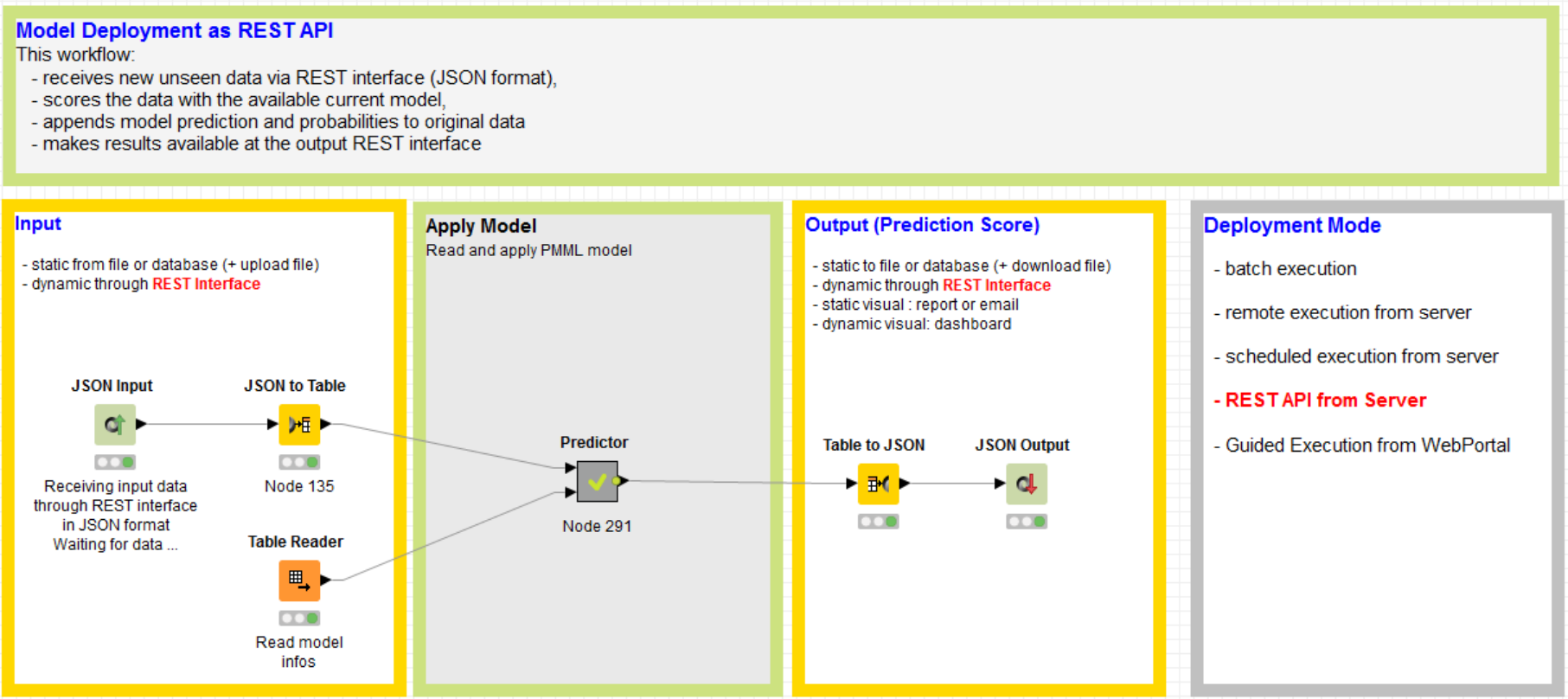


Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

REST API on Server

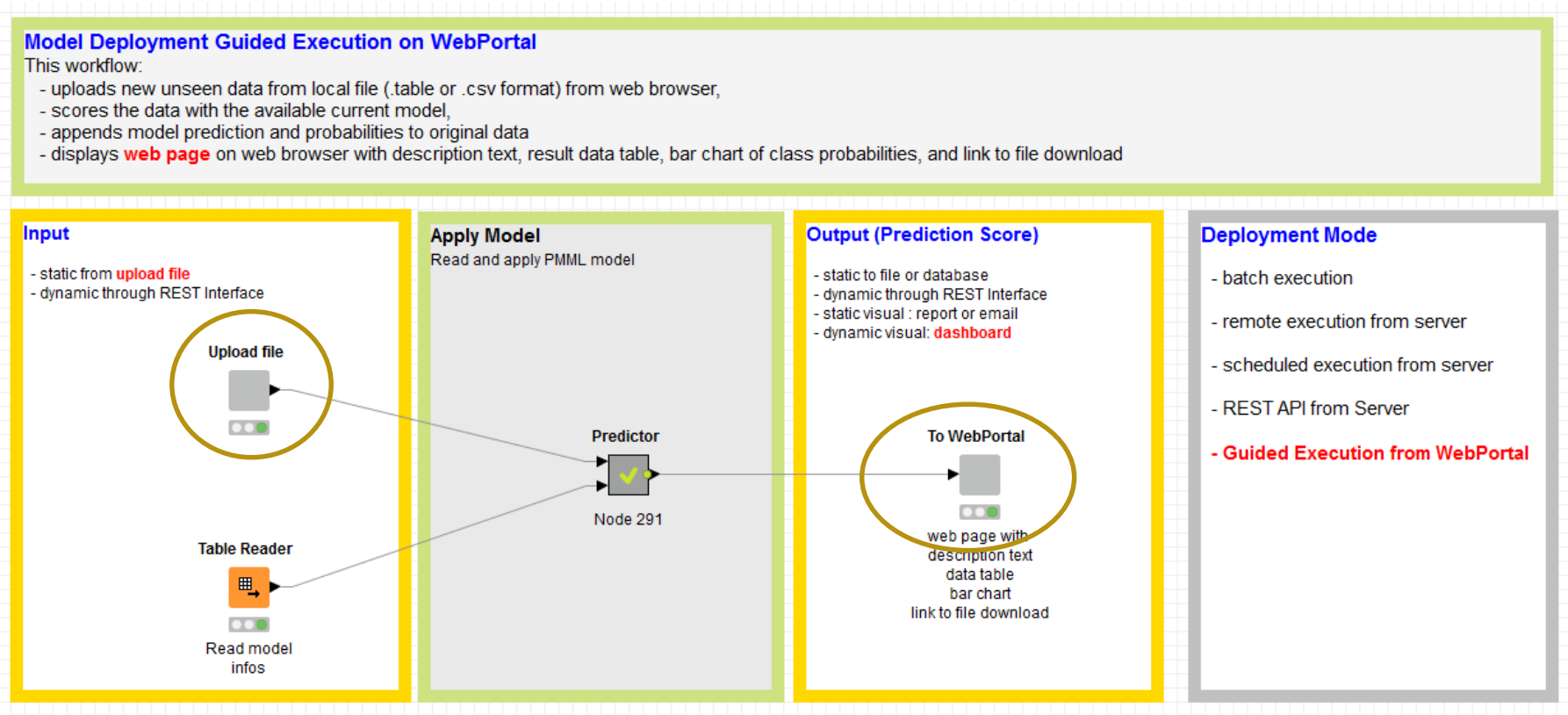


Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

To Dashboard on WebPortal



Step 1
Upload File

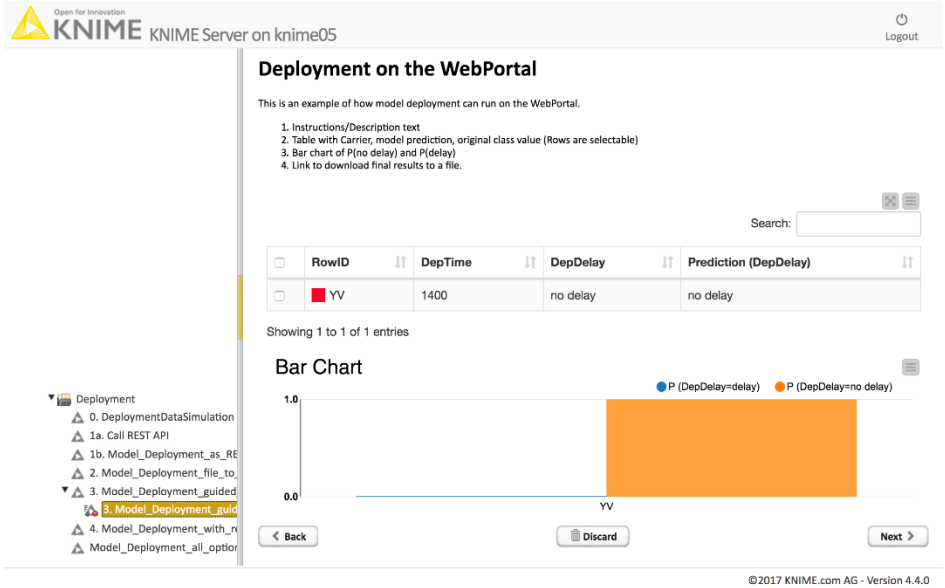
Step 2
Dashboard

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

Workflow on KNIME WebPortal



Step 1
Upload File

Step 2
Dashboard

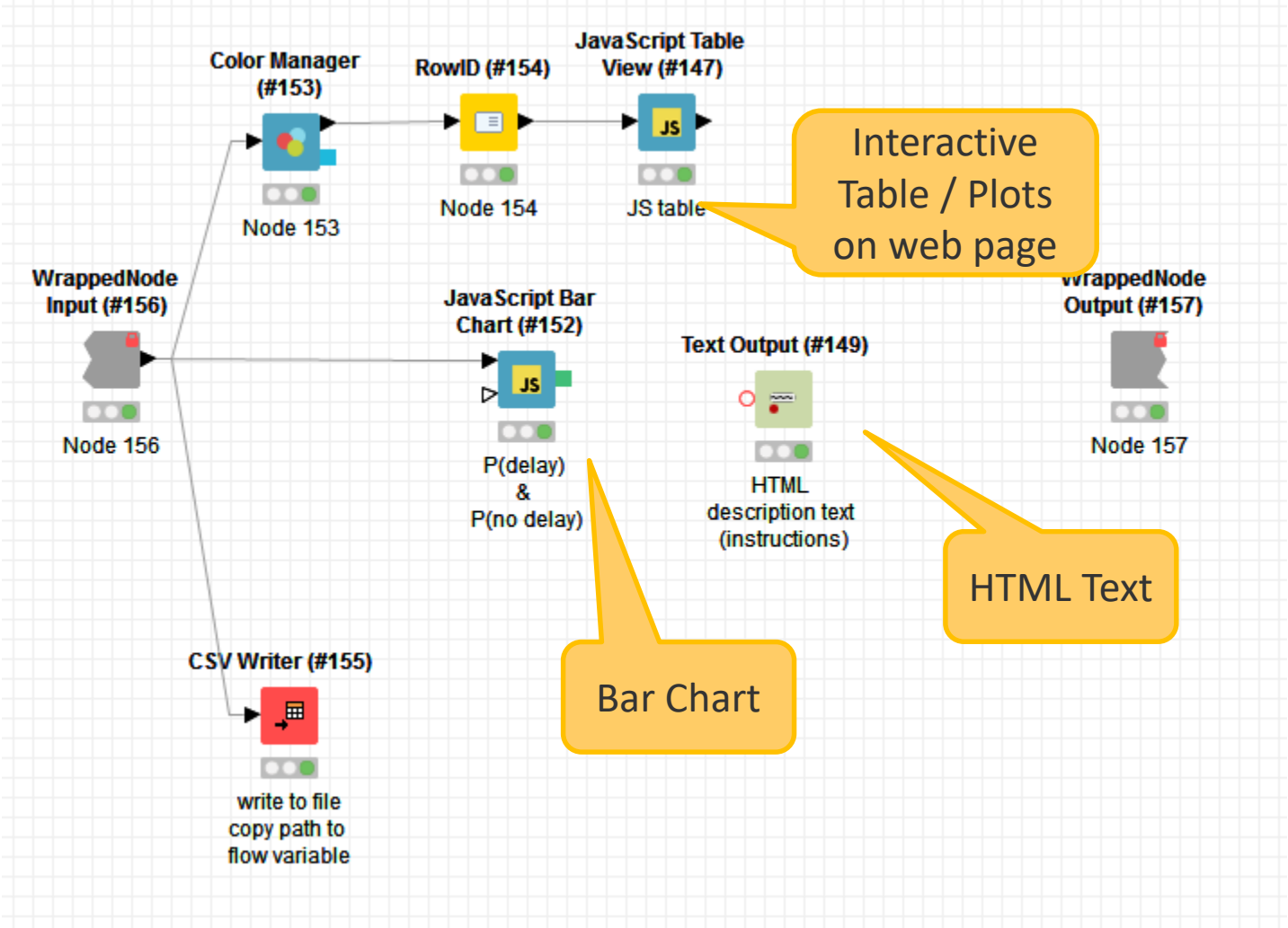
Available in KNIME Server

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

Wrapped Node to produce Dashboard on Web Page



Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

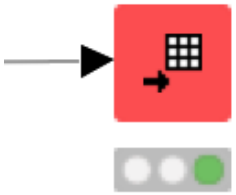
Open for Innovation ® KNIME

Data Export Nodes

Typically characterized by:

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

Table Writer



- IO
  - Write
    - CSV Writer
    - ARFF Writer
    - Table Writer
    - PMML Writer
    - Model Writer
    - Image Port Writer
    - Image Column Writer
    - Excel Writer (XLS)
- Analytics
  - Mining
    - Weka
      - Weka (3.7)
        - IO
          - Weka Classifier Writer (3.7)
          - Weka Clustering Writer (3.7)
  - Distance Calculation
    - Distance Matrix
      - Distance Matrix Writer
- Database
  - Read/Write
    - Database Connection Table Writer
    - Database Writer
- Other Data Types
  - Network
    - IO
      - File
        - Network Writer

- KNIME Labs
  - Deep Learning
    - I/O
      - DL4J Model Writer
  - MongoDB
    - MongoDB Writer
  - Semantic Web/Linked Data
    - IO
      - Triple File Writer
      - SPARQL File Writer
  - Spotfire
    - TIBCO File Writer
  - Tableau
    - Tableau Writer (TDE)
- Chemistry
  - I/O
    - Mol2 Writer
    - Molfile Writer
    - SDF Writer
    - Smiles Directory Writer

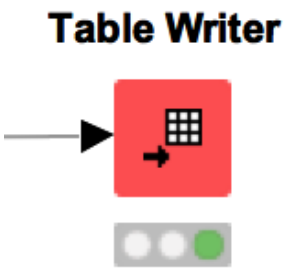
- Structured Data
  - JSON
    - JSON Writer
    - JSON Row Combiner and Writer
  - XML
    - XML Writer
- Scripting
  - Python
    - Python Object Writer
  - R
    - IO
      - R Model Writer
- Community Nodes
  - KNIME Image Processing
    - IO
      - Image Writer
  - Erlwood Nodes
    - IO
      - XLS Coloured Writer
  - NGS
    - IO
      - BEDGraphWriter
      - FQCWriter
      - FastQWriter
  - Palladian
    - Text Classifier
      - TextClassifierModelWriter
  - RDKit
    - Fingerprints
      - RDKit Fingerprint Writer

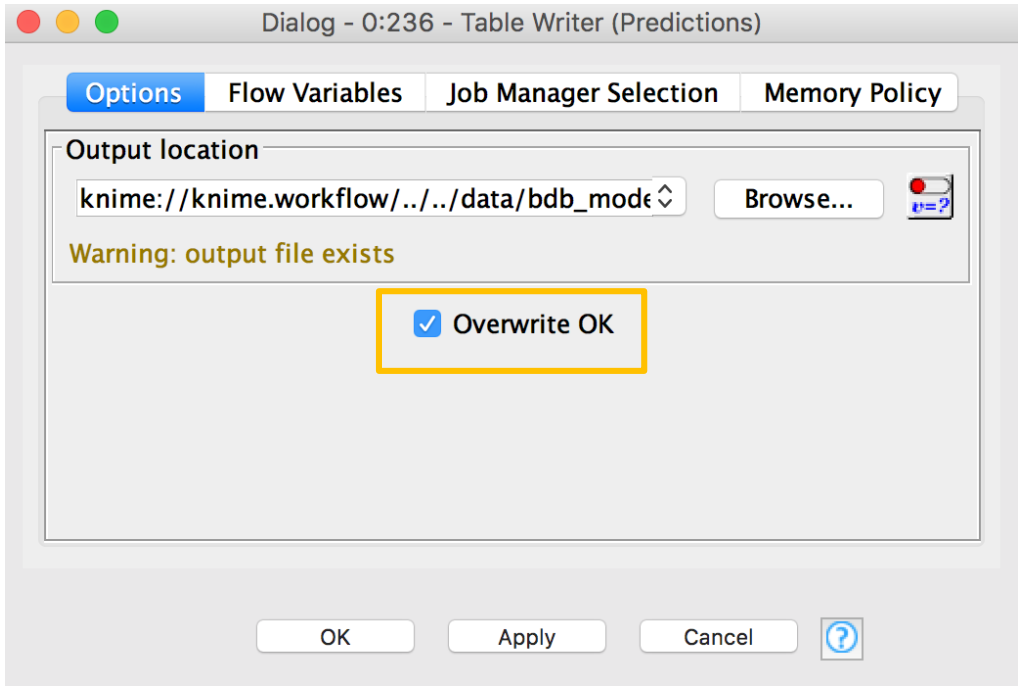
Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

New Node: Table Writer

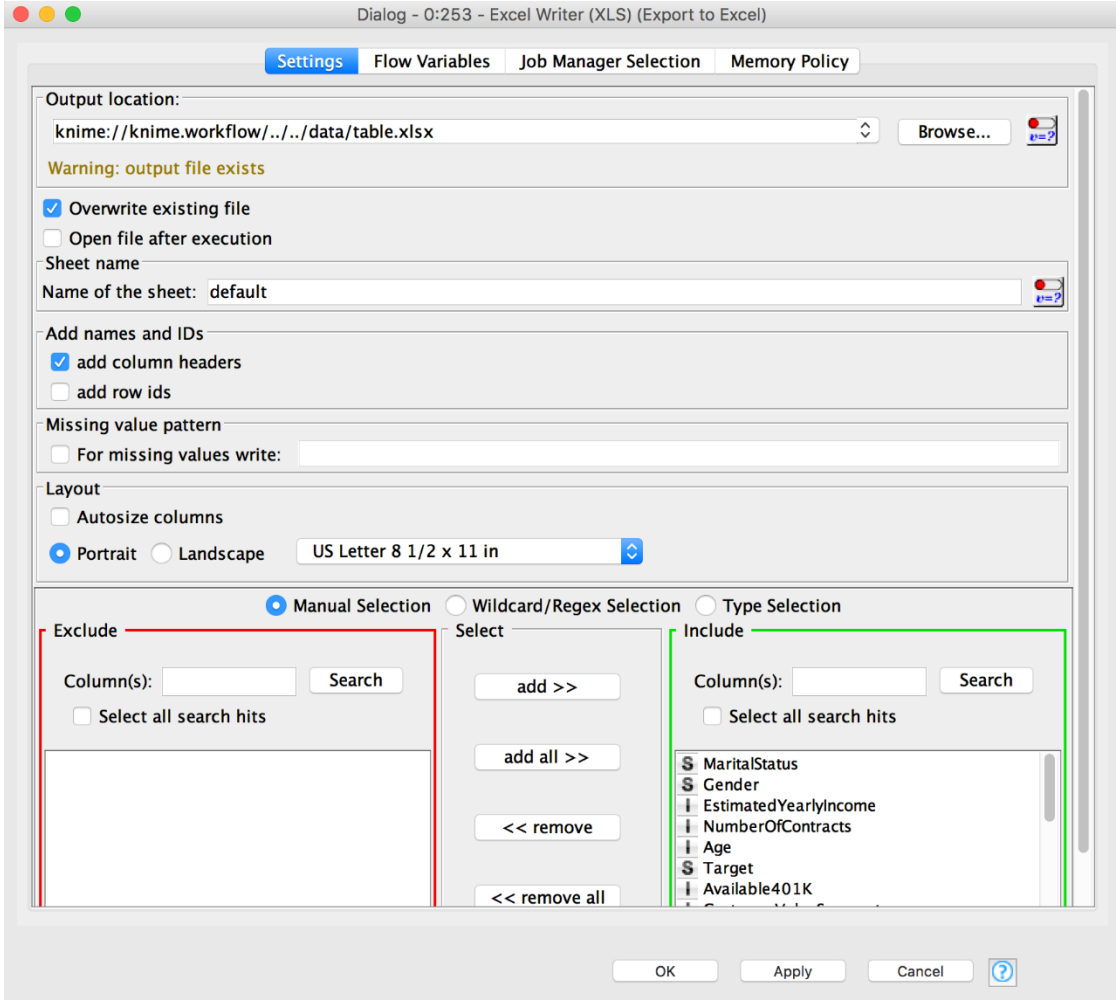




Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

New Node: XLS Writer



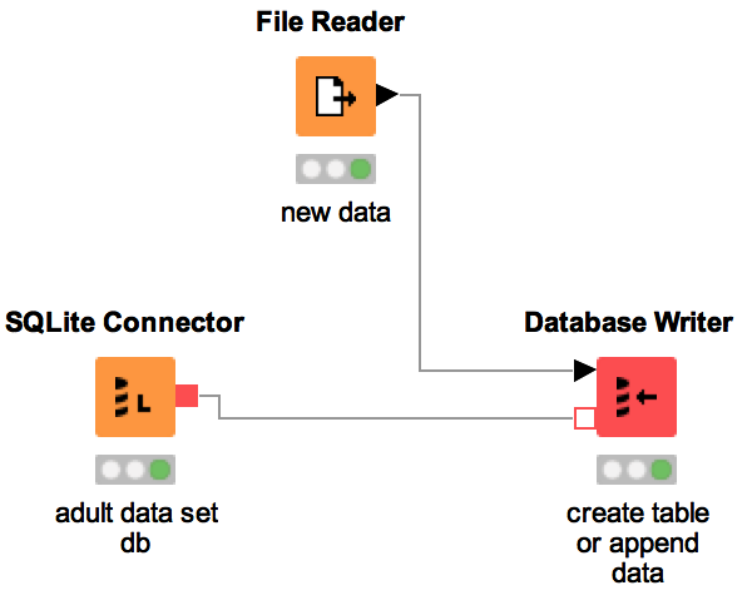
Copyright © 2017 KNIME.com AG

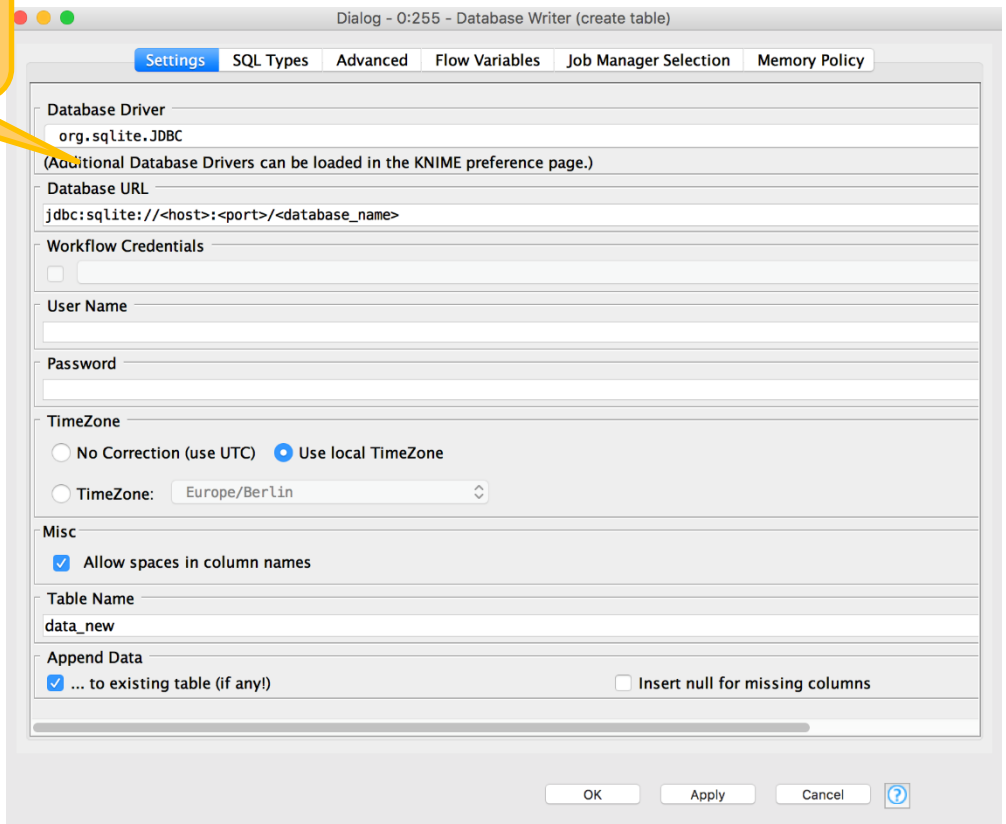
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

New Node: Database Writer

Only if no Connector node



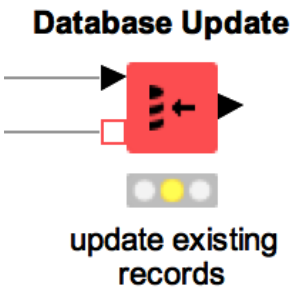


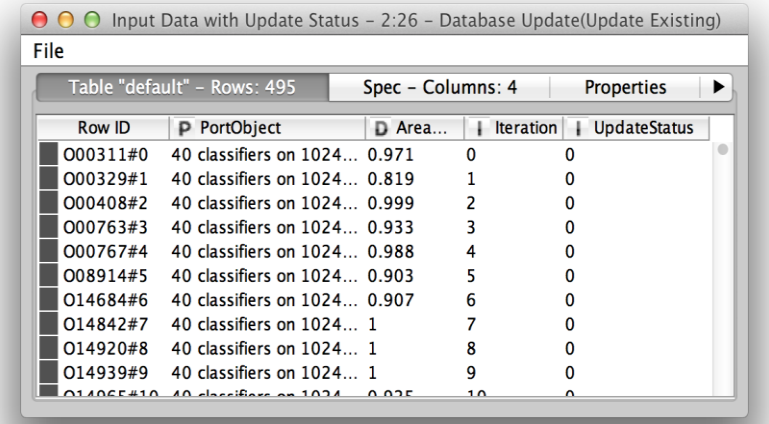
Copyright © 2017 KNIME.com AG

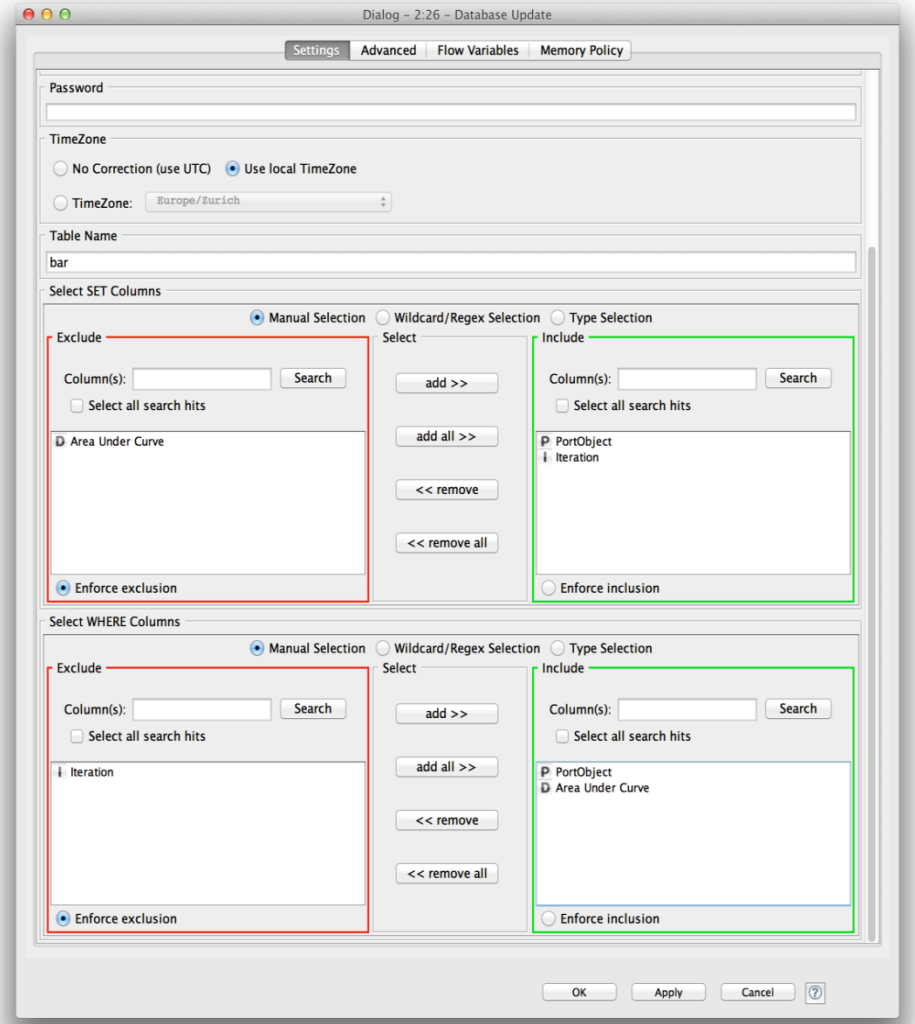
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

New Node: Database Update







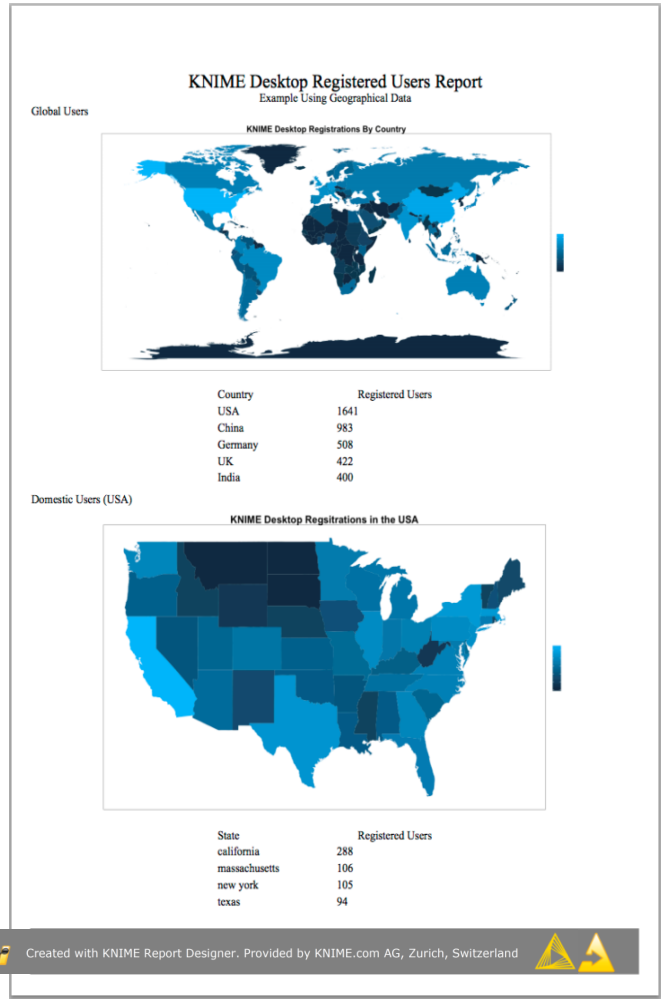
Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

Reporting in KNIME

- Reporting in KNIME is done via a 3rd party application named BIRT (Business Intelligence Reporting Tool)
- Data is sent to BIRT from KNIME using special nodes.
- Reports in BIRT are constructed from report items, which may include images, tables, charts and labels.
- Reports may be generated in a variety of formats (html, pdf, pptx, xlsx, docx)



KNIME Desktop Registered Users Report

Example Using Geographical Data

Global Users

KNIME Desktop Registrations By Country

Country	Registered Users
USA	1641
China	983
Germany	508
UK	422
India	400

Domestic Users (USA)

KNIME Desktop Regsitrations in the USA

State	Registered Users
california	288
massachusetts	106
new york	105
texas	94

Created with KNIME Report Designer. Provided by KNIME.com AG, Zurich, Switzerland

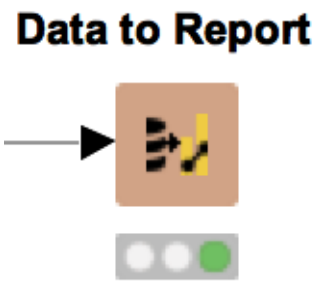
Copyright © 2017 KNIME.com AG

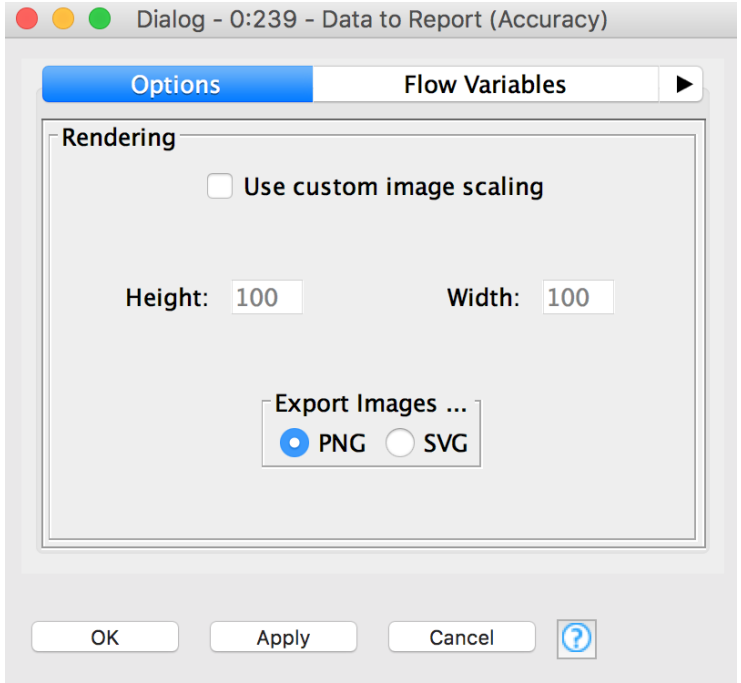
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

New Node: Data to Report

Send a data table to BIRT





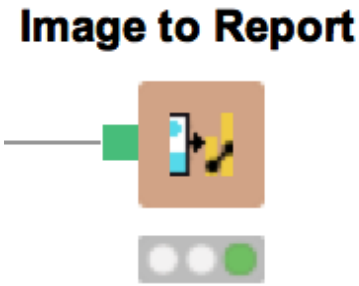
Copyright © 2017 KNIME.com AG

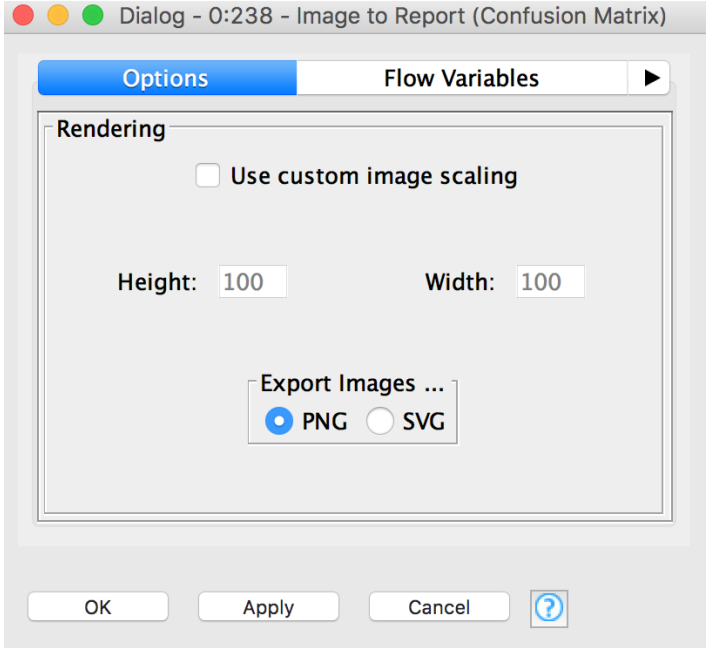
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

New Node: Image to Report

Send an image to BIRT

- PNG and SVG are supported formats (see node description for details)



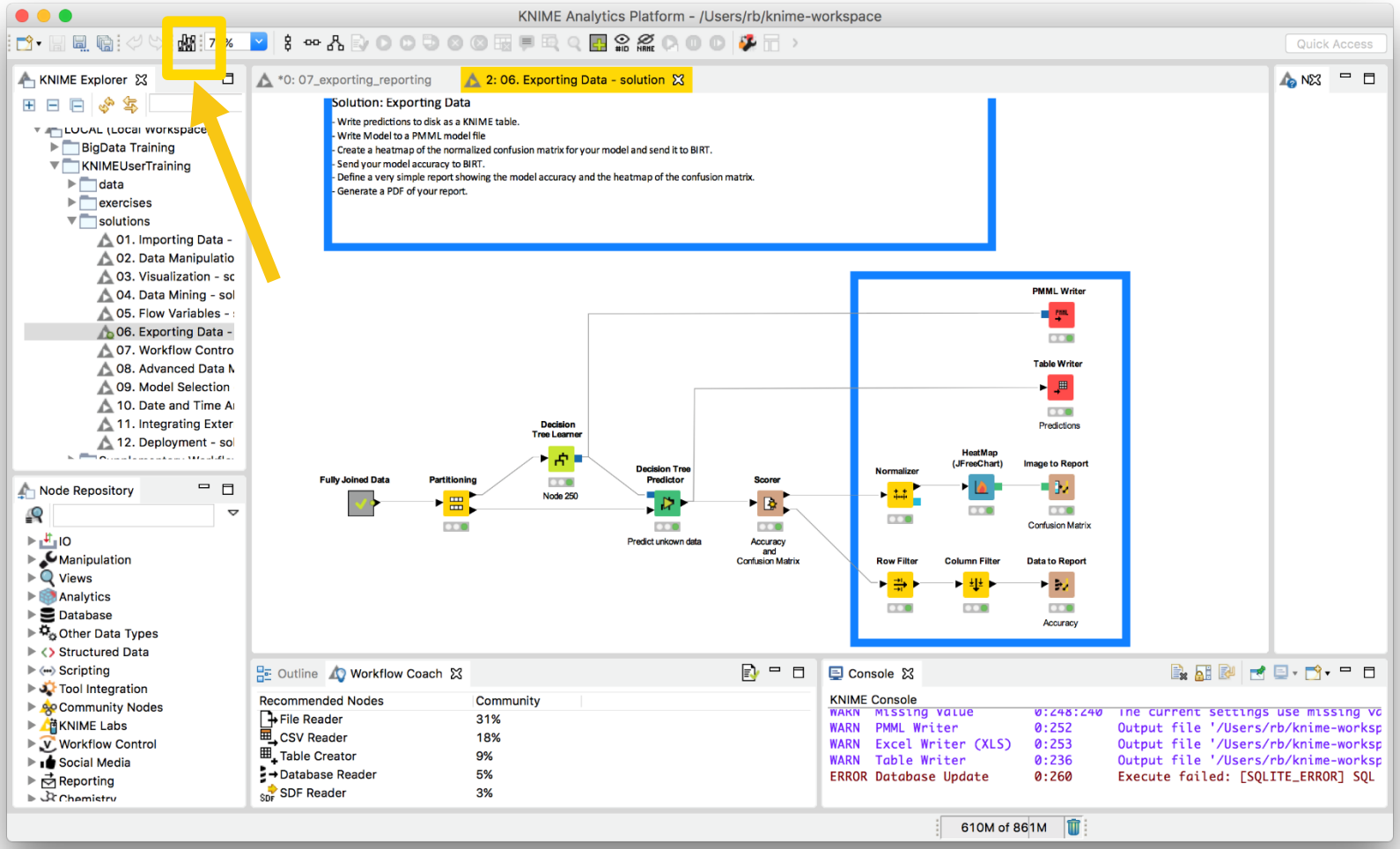


Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Edit the Report

Open the workflow > Click the Report Editor button in the tool bar



Copyright © 2017 KNIME.com AG

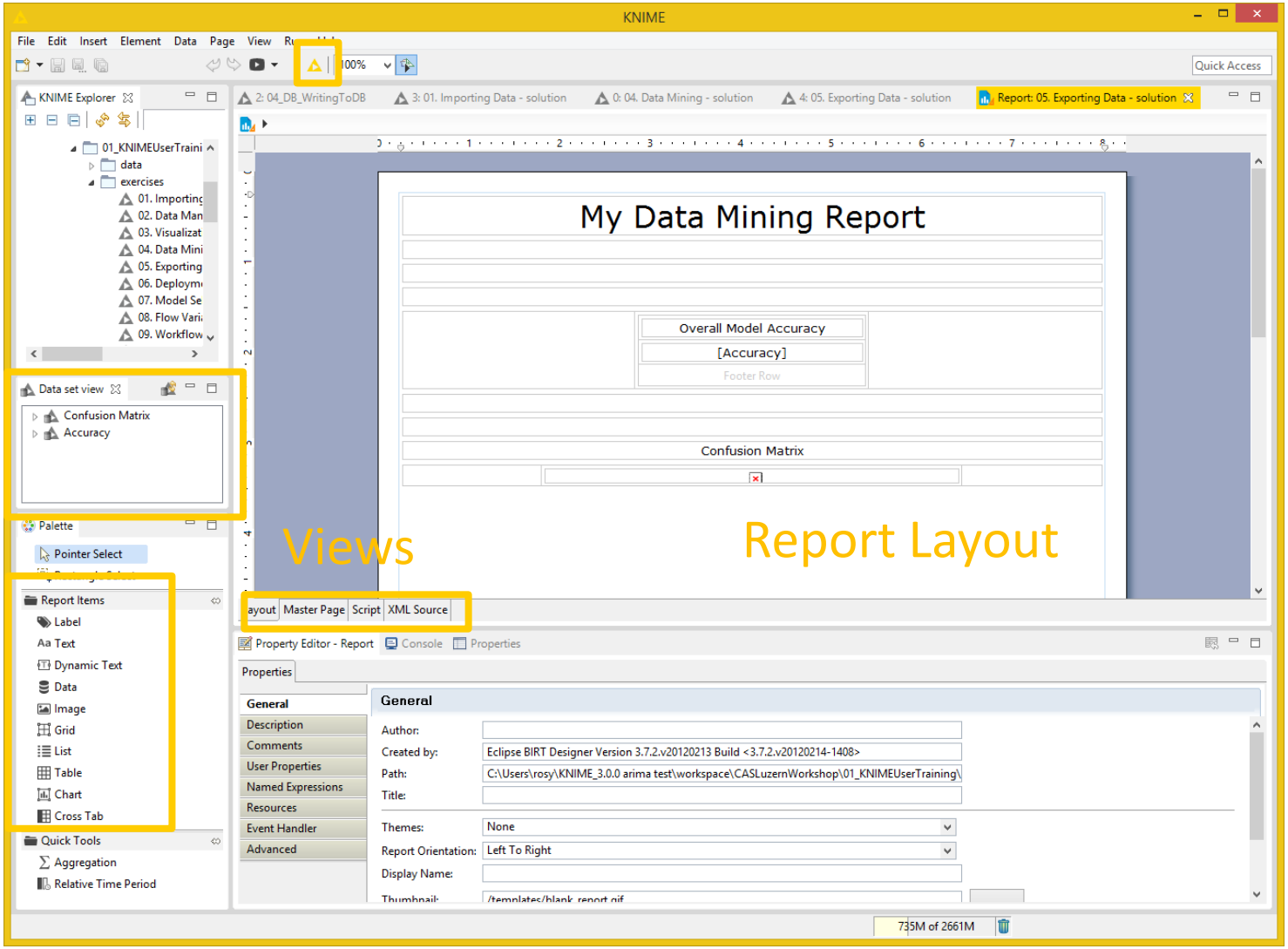
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

Reporting Perspective

Data from KNIME

Report Items



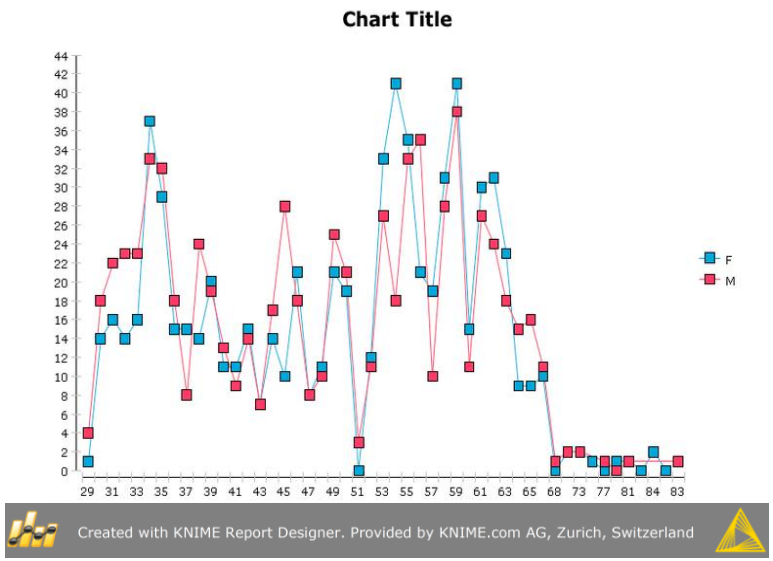
Copyright © 2017 KNIME.com AG

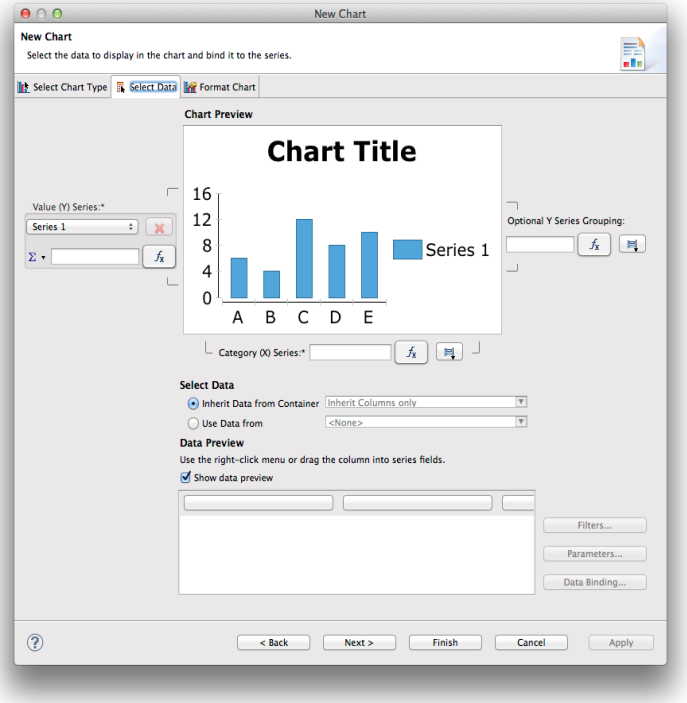
Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/

Open for Innovation ® KNIME

Charting in BIRT

- Many chart types
- Fine control of plot appearance
- Familiar ‘Excel Like’ interface
- Supports interactivity





Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Exporting Data Exercise

Starting with exercise: *Exporting Data*

- Write your predictions to disk as a KNIME table
- Create a heatmap of the normalized confusion matrix and send it to BIRT
- Send your model accuracy table to BIRT
- Define a very simple report showing the accuracy of your model
- Generate a PDF of your report

Copyright © 2017 KNIME.com AG

Licensed under a Creative Commons Attribution-Noncommercial-Share Alike license
https://creativecommons.org/licenses/by-nc-sa/4.0/
Open for Innovation ® KNIME

Open for Innovation®
KNIME

The End

education@knime.com

Copyright © 2017 KNIME.com AG