Introduction to KNIME

The Konstanz Information Miner

Tobias Kötter

University of Konstanz, Germany

Tobias.Koetter@uni-konstanz.de

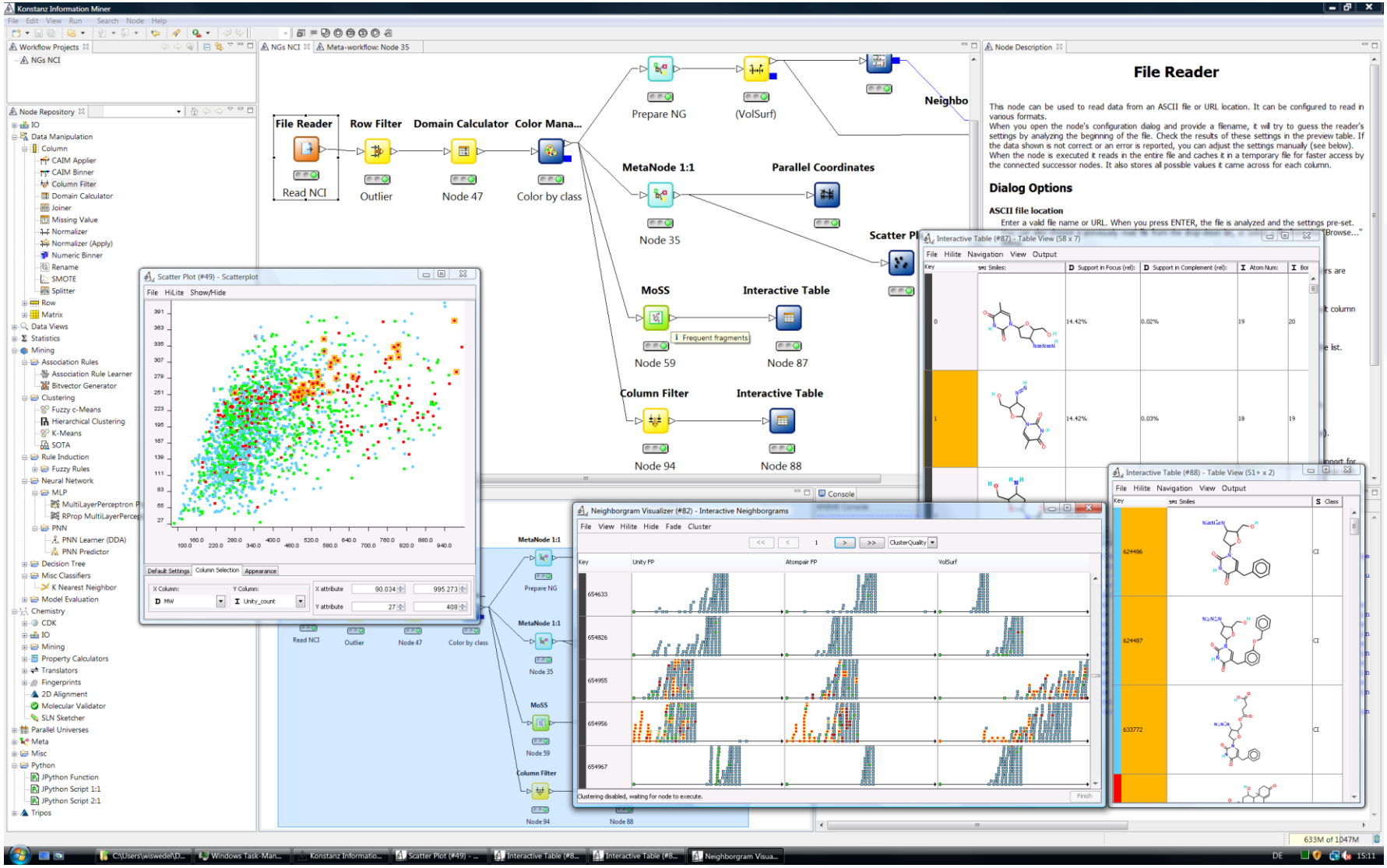
Agenda

- Introduction
- Installation
- Usage
- Example Workflows
- Development
- Summary

What is KNIME

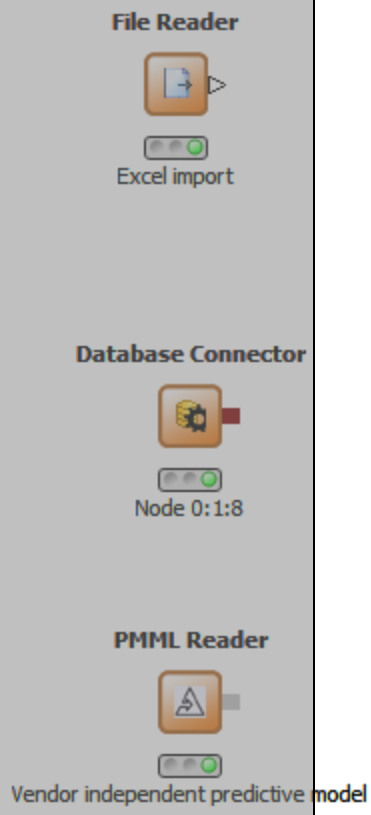
- Data exploration platform
  - Data flows (data pipelining)
  - Operations are encapsulated in nodes
  - More than 100 nodes for processing and analyzing your data
- Based on the Eclipse platform and Java
  - Open Source
  - Development started in 2004
  - Developed by KNIME.com and the Nycomed chair
- Public release of version 1.0 in 2006
- More than 5500 active users

KNIME


Konstanz Information Miner
Workflow Projects
NGs NCI
Node Repository
Meta-workflow: Node 35
File Reader
Row Filter
Domain Calculator
Color Mana...
Read NCI
Outlier
Node 47
Color by class
Prepare NG
(VolSurf)
Neighbo
MetaNode 1:1
Parallel Coordinates
Node 35
Scatter Pl
MoSS
Interactive Table
Frequent fragments
Node 59
Node 87
Column Filter
Interactive Table
Node 94
Node 88
Node Description
File Reader
This node can be used to read data from an ASCII file or URL location. It can be configured to read in various formats.
When you open the node's configuration dialog and provide a filename, it will try to guess the reader's settings by analyzing the beginning of the file. Check the results of these settings in the preview table. If the data shown is not correct or an error is reported, you can adjust the settings manually (see below).
When the node is executed it reads in the entire file and caches it in a temporary file for faster access by the connected successor nodes. It also stores all possible values it came across for each column.
Dialog Options
ASCII file location
Enter a valid file name or URL. When you press ENTER, the file is analyzed and the settings pre-set.
Scatter Plot (#49) - Scatterplot
Interactive Table (#87) - Table View (58 x 7)
Interactive Table (#88) - Table View (51+ x 2)
Neighborgram Visualizer (#82) - Interactive Neighborgrams
Clustering disabled, waiting for node to execute.
Console
633M of 1047M

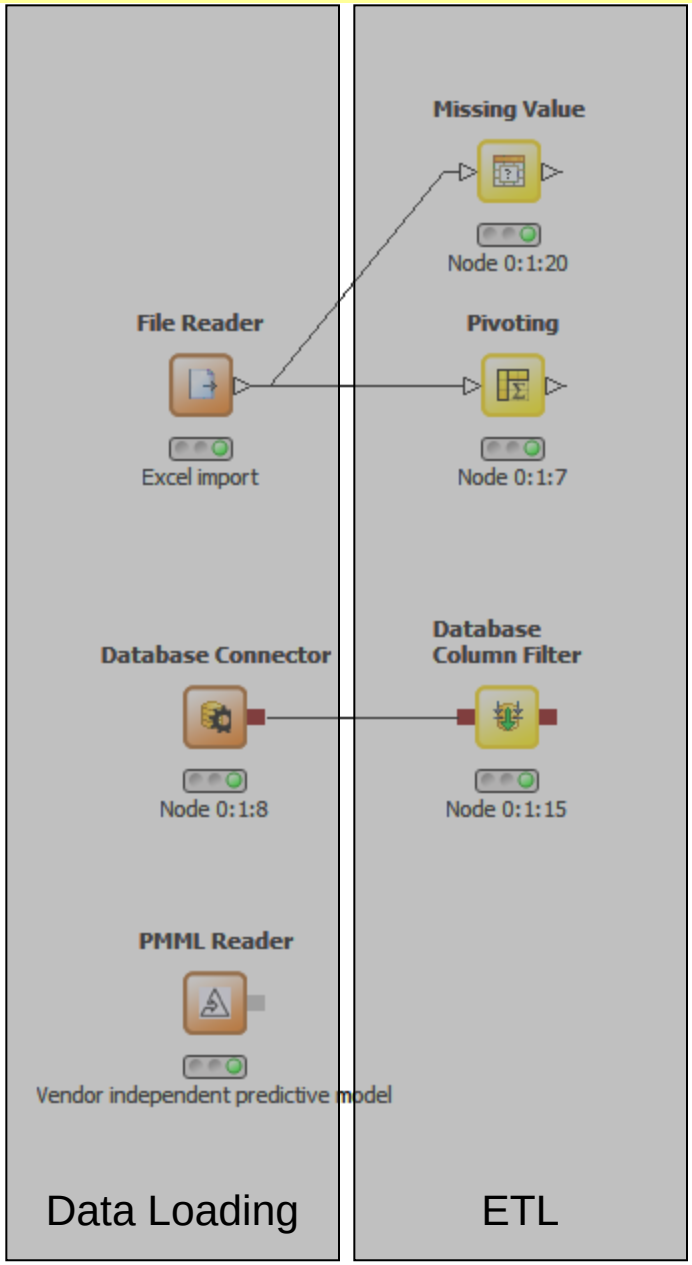
KNIME loads and integrates data from diverse data sources:

- Different data bases
- Various file formats (CSV, XML, SDF, etc.)



Data Loading

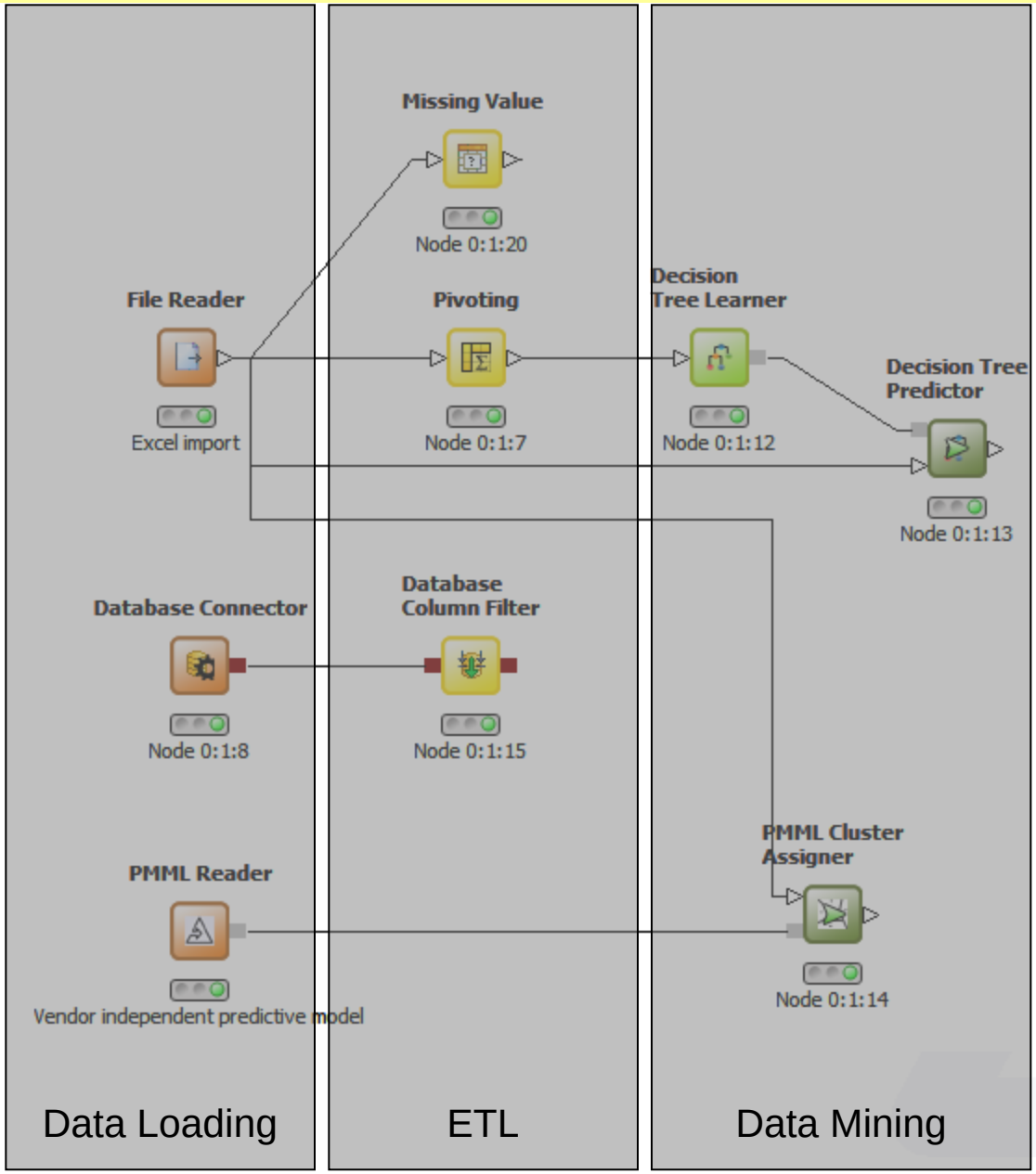
KNIME



KNIME provides huge repository of modules for easy-to-use, modular

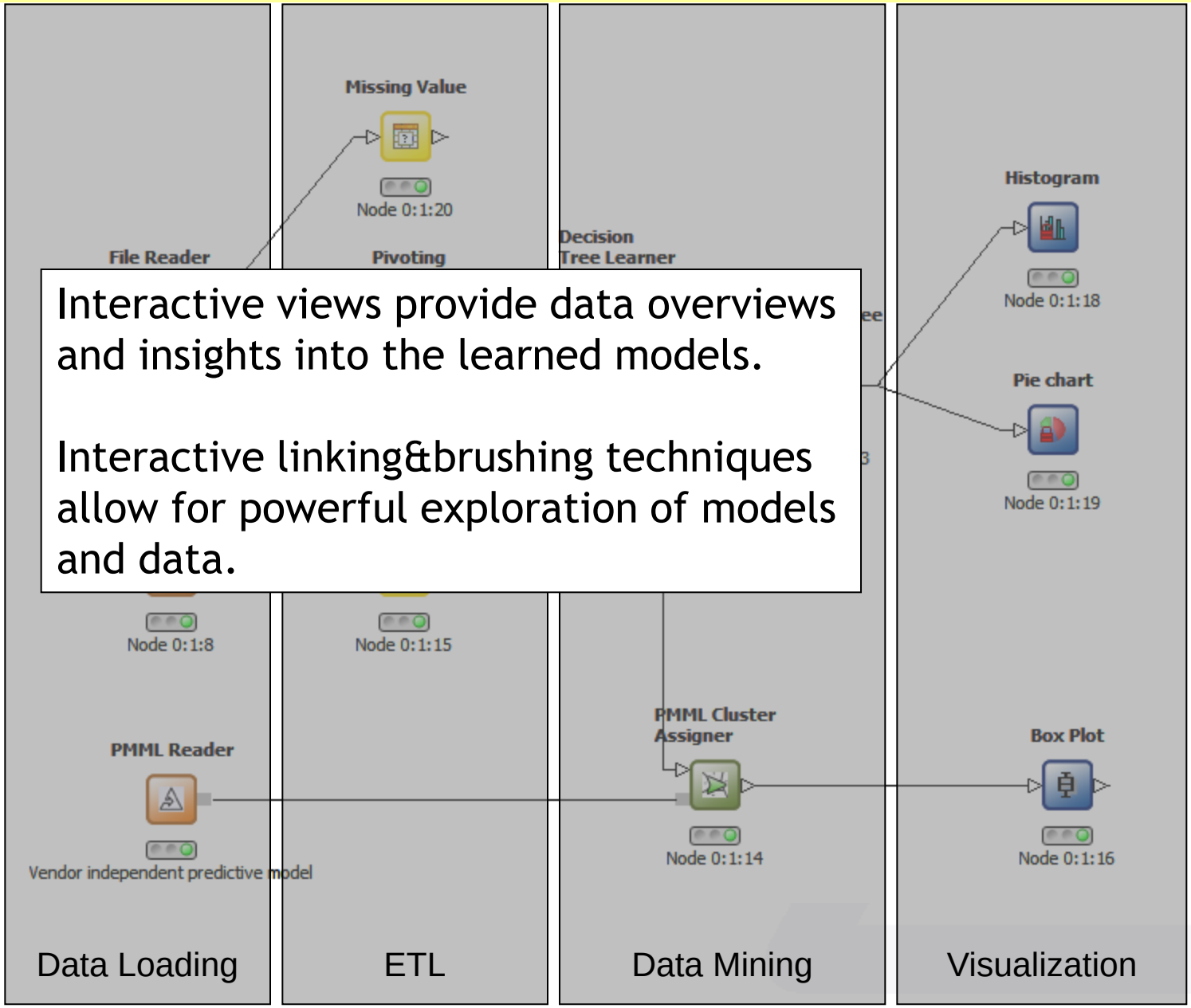
- Data preprocessing
- Data fusion
- Data transformation

KNIME



In addition to standard data mining techniques, KNIME adds cutting edge data analysis algorithms. (…thanks to its academic roots)

KNIME



Interactive views provide data overviews and insights into the learned models.

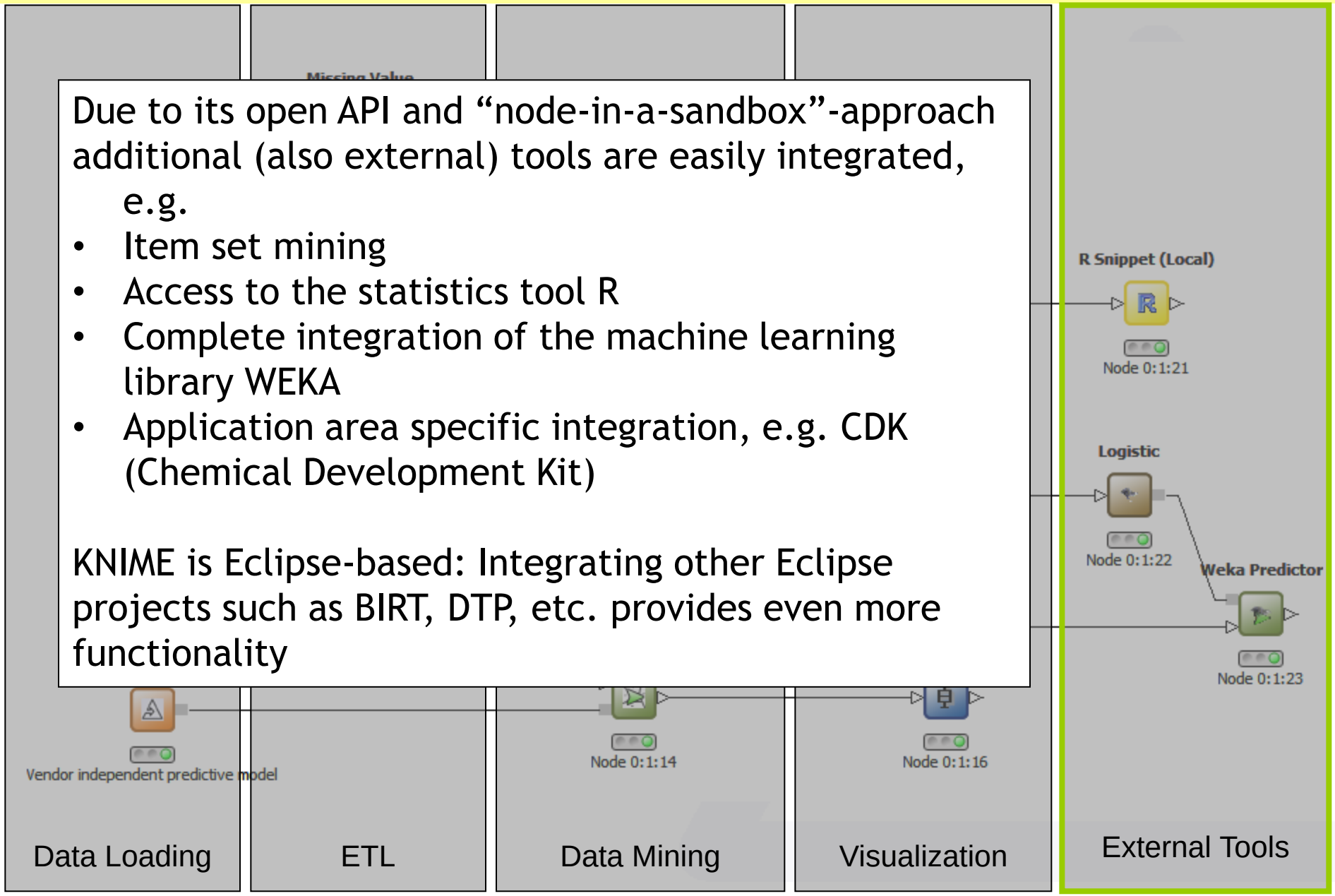
Interactive linking&brushing techniques allow for powerful exploration of models and data.

KNIME

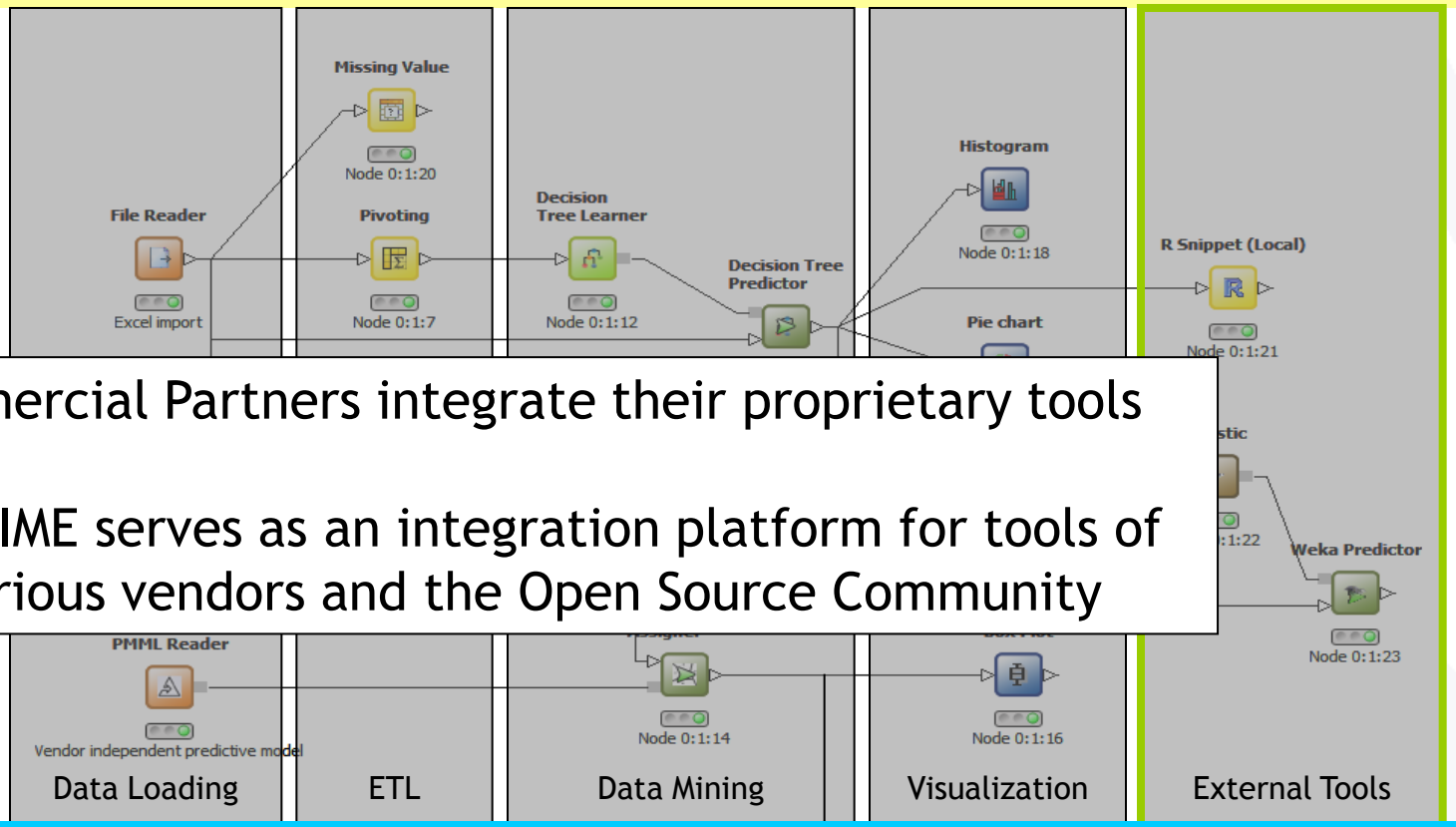
Due to its open API and “node-in-a-sandbox”-approach additional (also external) tools are easily integrated, e.g.

- Item set mining
- Access to the statistics tool R
- Complete integration of the machine learning library WEKA
- Application area specific integration, e.g. CDK (Chemical Development Kit)

KNIME is Eclipse-based: Integrating other Eclipse projects such as BIRT, DTP, etc. provides even more functionality

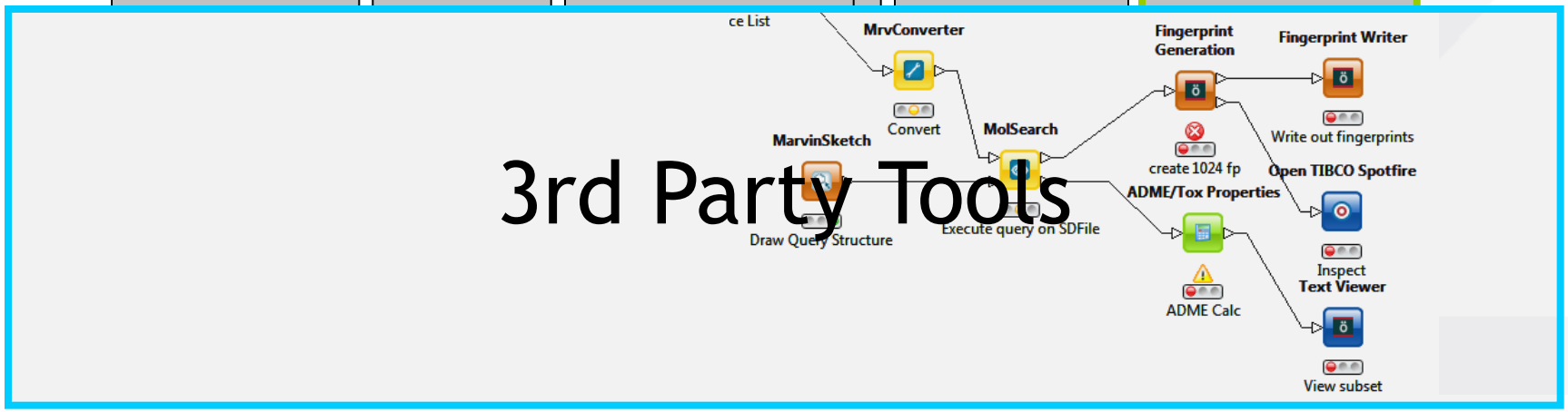


KNIME



Commercial Partners integrate their proprietary tools

⇒ KNIME serves as an integration platform for tools of various vendors and the Open Source Community



3rd Party Tools

INSTALLATION

KNIME Installation

- KNIME: http://knime.org/download-desktop
  - KNIME Desktop Version
  - KNIME SDK Developer Version

- Available for:
  - Windows 64/32bit
  - Linux 64/32bit
  - Mac OS X 64bit

- No installation needed simply unpack the downloaded archives

KNIME Standard Nodes

- IO
  - Read
  - Write
  - Other
  - Cache
- Database
- Data Manipulation
  - Column
  - Row
  - Matrix
  - PMML
- Data Views
- Statistics
  - Regression
  - Linear Correlation
  - Correlation Filter
  - Statistics
  - Crosstab
  - Value Counter
- Mining
- Meta
  - Feature Elimination
  - Iterate List of Files
  - Loop x-times
  - Variables Loop (Data)
  - Variables Loop (Database)
  - X-Validation
- Flow Control
  - Loop Support
  - Switches
  - Variables
- Misc
- Time Series

- Database
  - Database Reader
  - Database Connector
  - Database Looping
  - Database Row Filter
  - Database Query
  - Database Column Filter
  - Database Connection Reader
  - Database Connection Writer
  - Database Writer

- Mining
  - Bayes
  - Clustering
  - Rule Induction
  - Neural Network
  - Decision Tree
  - Misc Classifiers
  - Item Sets / Association Rules
  - MDS
  - PCA
  - SVM
  - Scoring
  - Meta

- Data Views
  - Property
  - Utility
  - Box Plot
  - Conditional Box Plot
  - Histogram
  - Histogram (interactive)
  - Interactive Table
  - Lift Chart
  - Line Plot
  - Parallel Coordinates
  - Pie chart
  - Pie chart (interactive)
  - Scatter Matrix
  - Scatter Plot

- Time Series
  - Date Field Extractor
  - Time Field Extractor
  - Extract Time Window
  - Mask Date/Time
  - Moving Average
  - Preset Date/Time
  - String to Date/Time
  - Time Difference
  - Time Generator
  - Time to String

KNIME Extensions

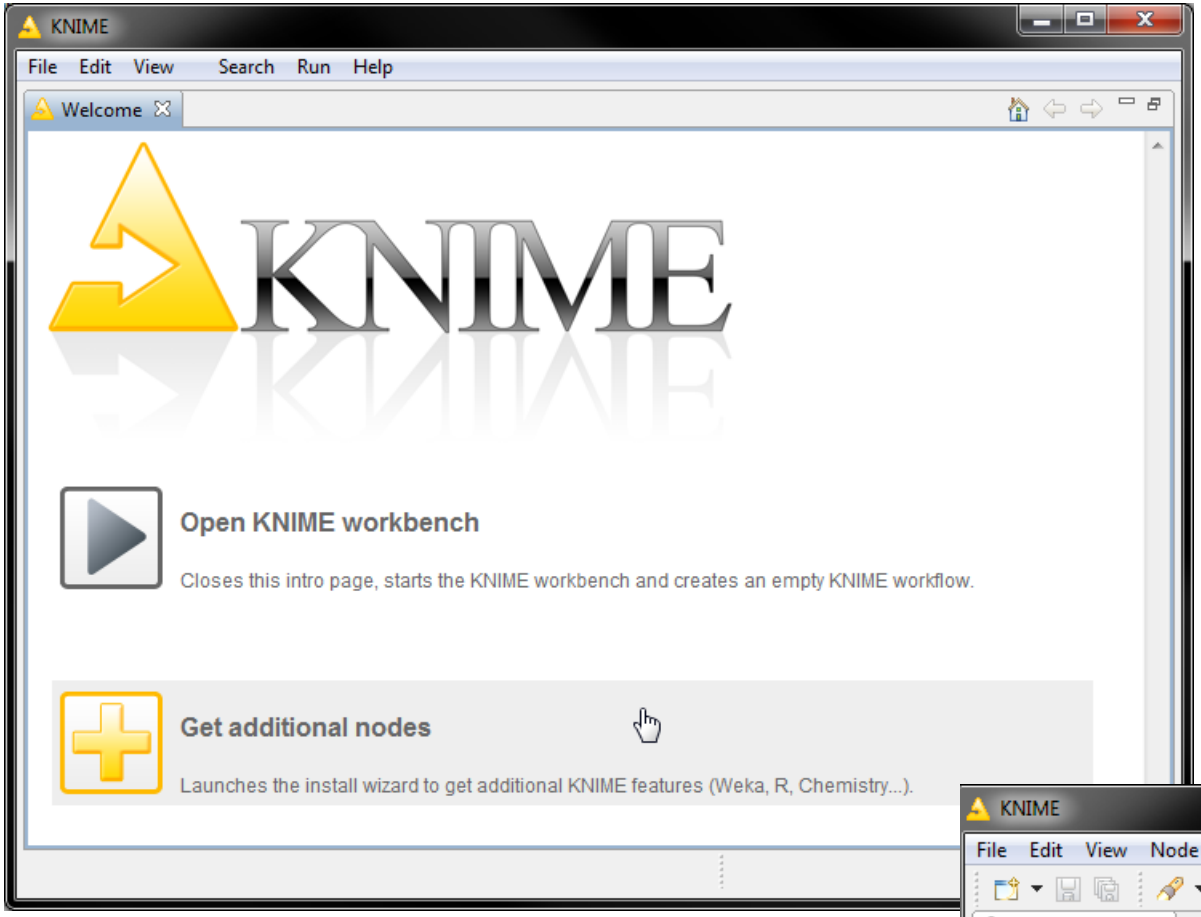
- Chemistry types and features
- Distance Matrix
- Ensemble Learning
- External Tool Support
- HTML/PDF Writer
- Item Set Mining
- R Statistics Integration
- Report Designer
- Webservice Client
- Weka Data Mining Integration
- XLS Support
- XML Processing
- …

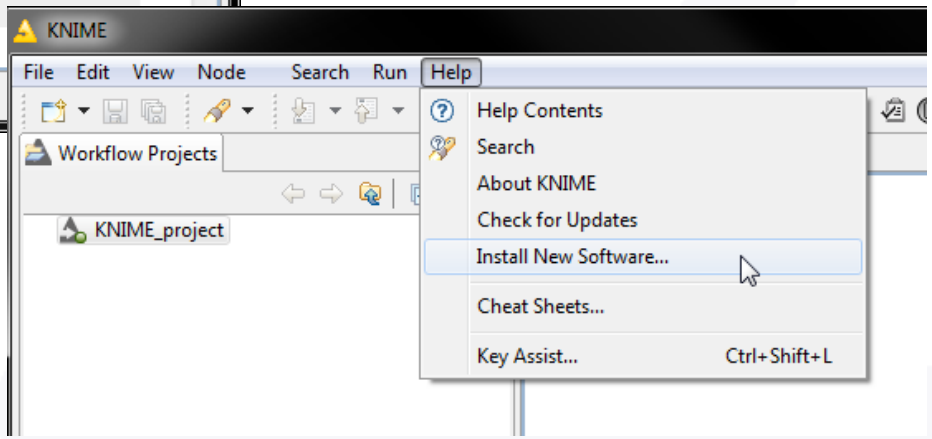
Additional Extensions

- Experimental Extensions: http://tech.knime.org/knime-labs
  - Modular Data Generators
  - Network Mining
  - Perl Scripting
  - Text Processing
  - …

- Community Contributions: http://tech.knime.org/community
  - Chemoinformatics
  - High Content Screening
  - Image Processing
  - Next Generation Sequencing
  - R/Groovy/Matlab/Python Scripting
  - STARK
  - …

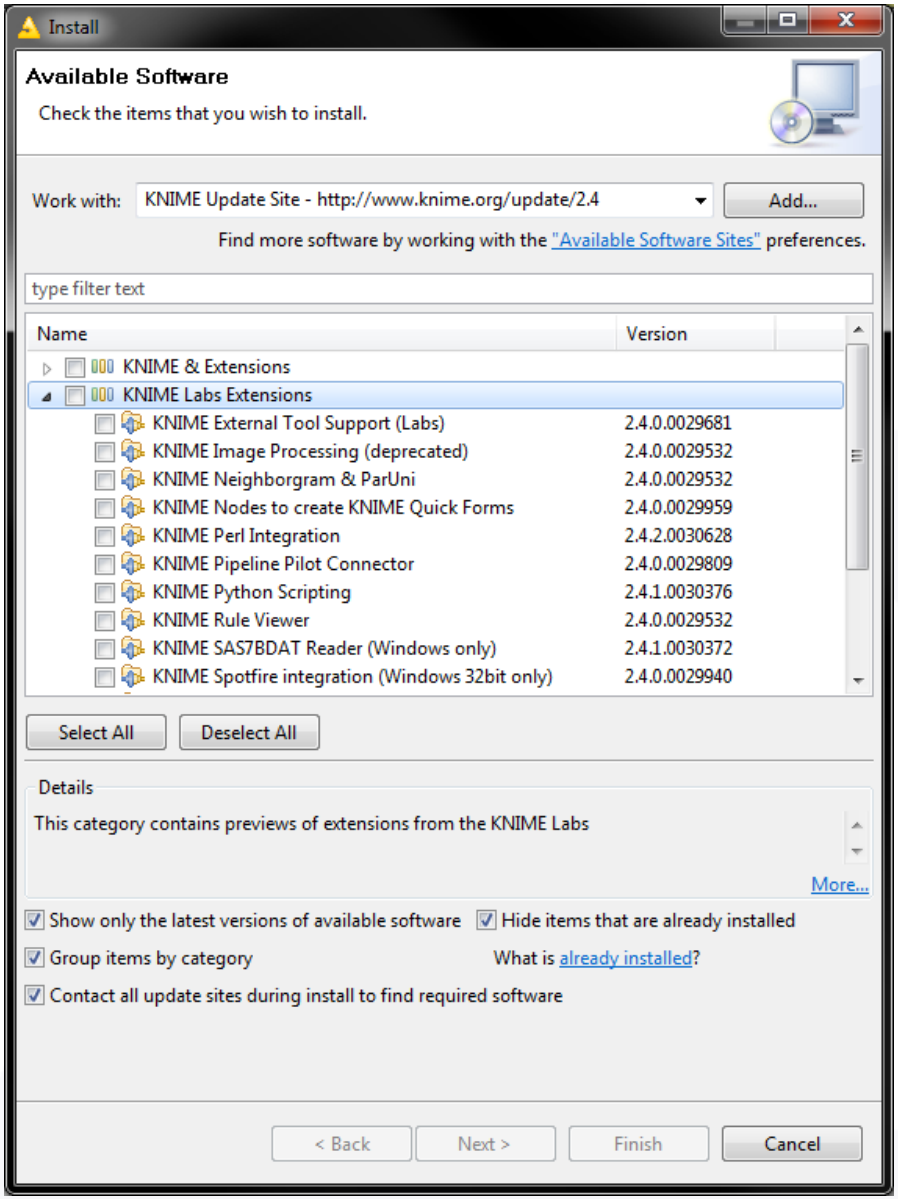
Extension Installation





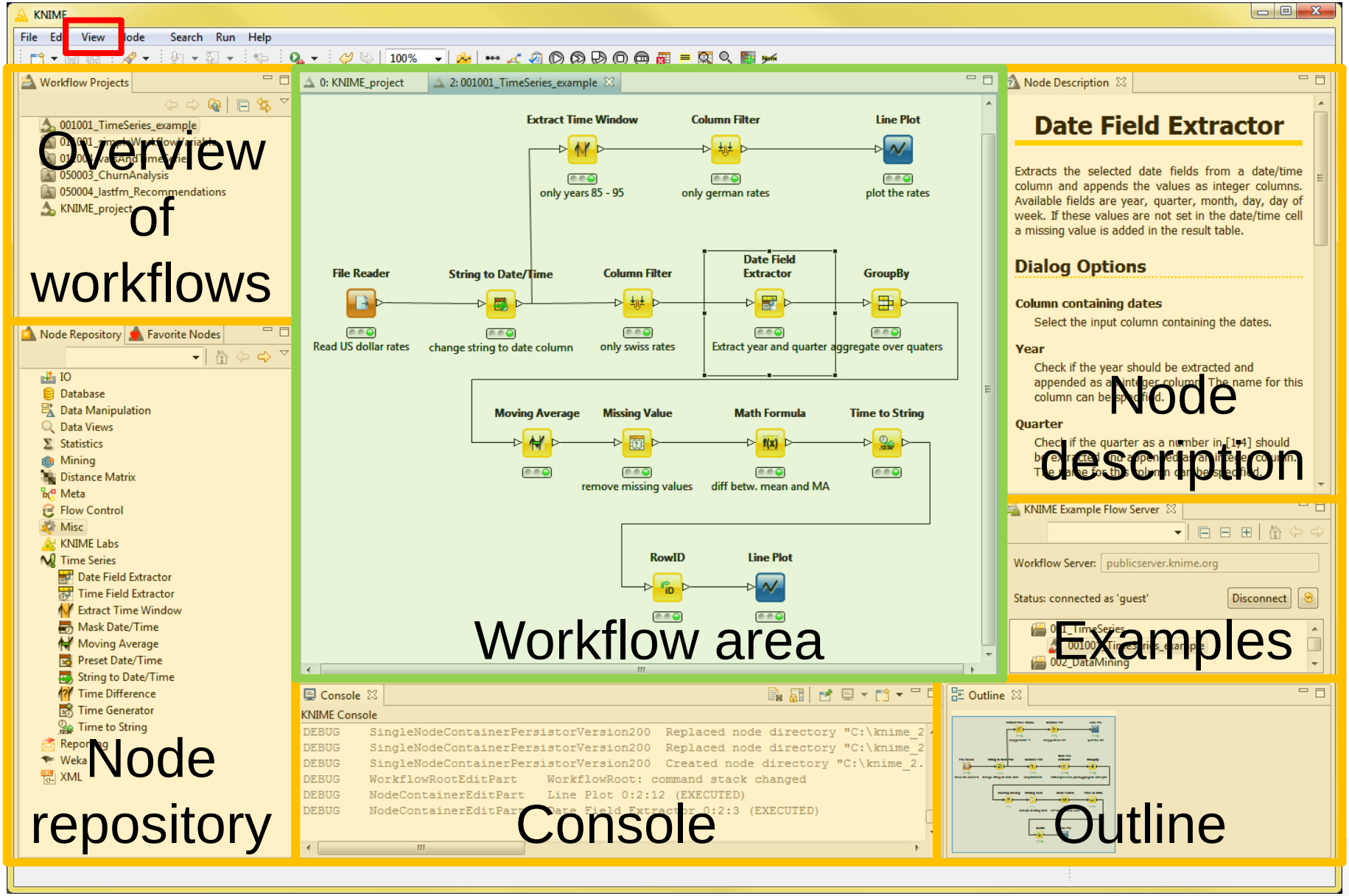
Extension Installation

- Installation via Eclipse update manager
- KNIME and Labs extensions at
  - http://www.knime.org/update/2.4
- Community Contributions at
  - http://tech.knime.org/update/community-contributions/release
  - http://tech.knime.org/update/community-contributions/nightly



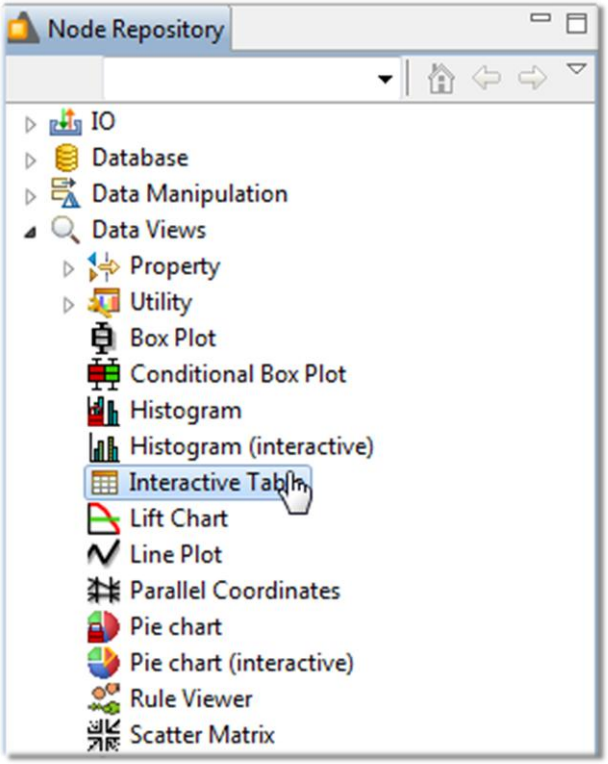
USAGE

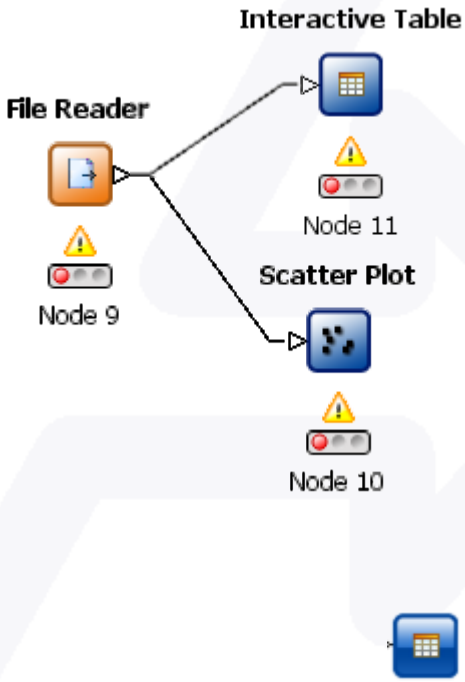
The KNIME Workbench



Creating a Simple Flow

- Drag and drop nodes from the repository to the workflow
- Connect the nodes by “drawing” connections between output and input ports





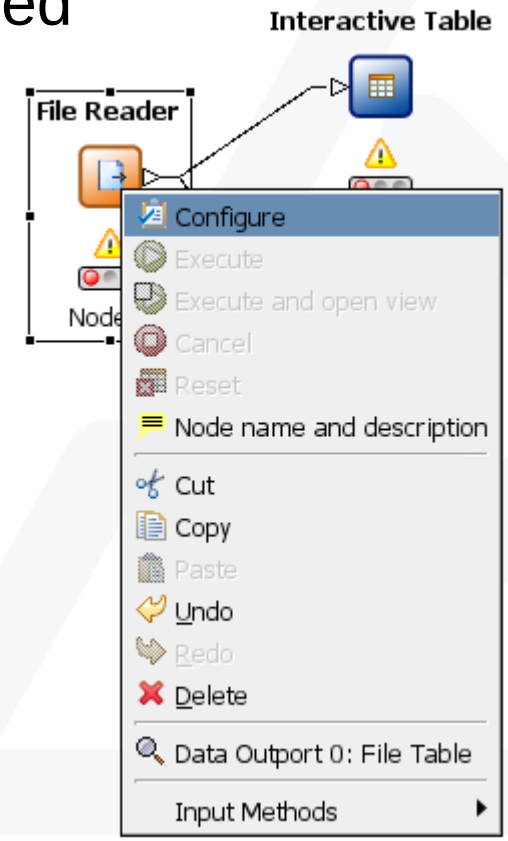
Configuring Nodes

- Right-click on node to get context menu
- *Configure* opens a dialog where parameters can be set
- Not all nodes need to be configured manually
- Node states


Node is not configured properly or not fully connected


Node is ready to be executed

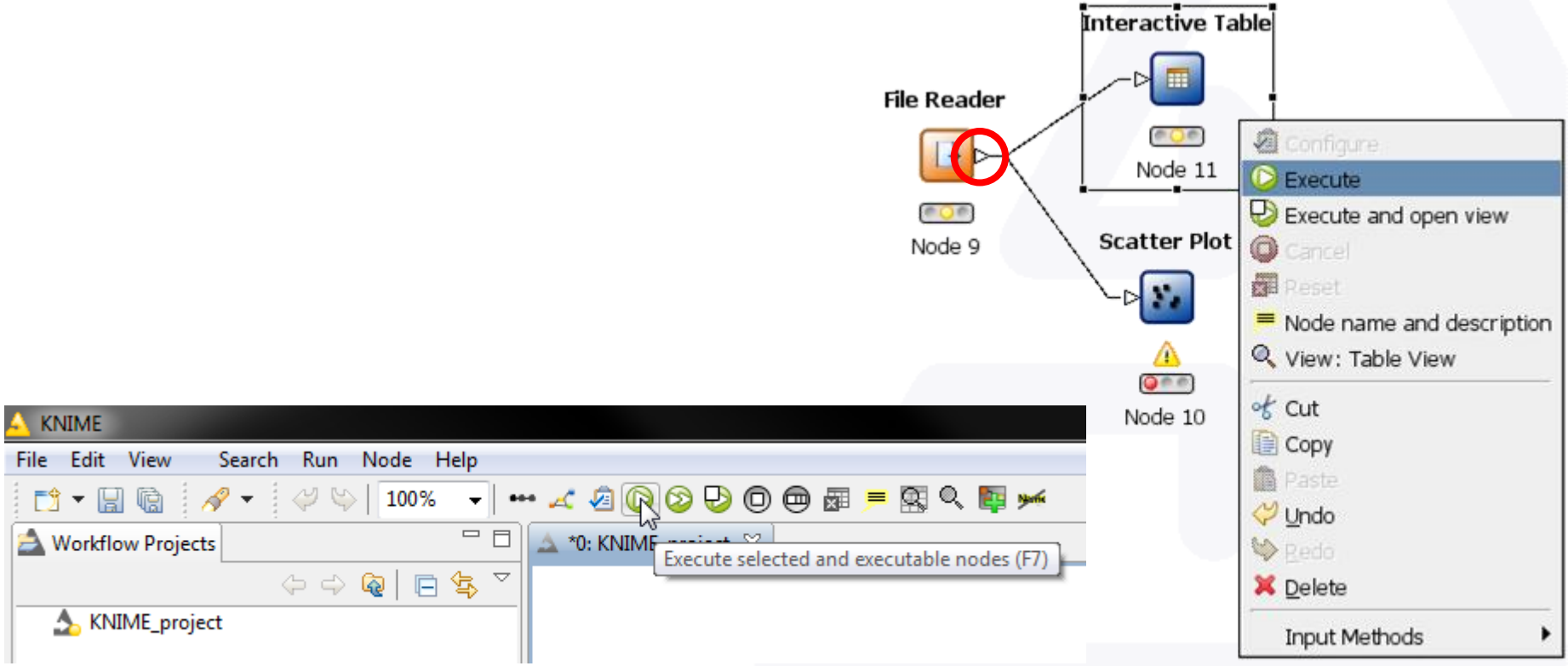

Node is executed and has data at its output port(s)



Executing Nodes

- Click *Execute* in the context menu or the toolbar button
  - only “yellow” nodes can be executed

- All predecessing nodes are executed automatically



Data Table

- Contains meta information (spec)
  - data types
  - domains
  - # of rows / cols
- Large tables are buffered on disc
- Blob cell support for large data cells e.g. images

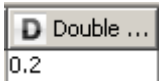
Bit vector data - 4:8 - Bitvector Generator

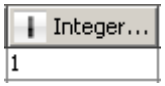
File

Table "default" - Rows: 150 | Spec - Columns: 5 | Properties

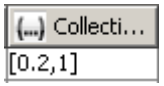
Row ID	D Double ...	S String Col	I Integer...	(...) Collecti...	BitVectors
Row0	0.2	Iris-setosa	1	[0.2,1]	10
Row1	0.2	Iris-setosa	1	[0.2,1]	10
Row2	0.2	Iris-setosa	1	[0.2,1]	10
Row3	0.2	Iris-setosa	1	[0.2,1]	10
Row4	0.2	Iris-setosa	1	[0.2,1]	10
Row5	0.4	Iris-setosa	1	[0.4,1]	10
Row6	0.3	Iris-setosa	1	[0.3,1]	10
Row7	0.2	Iris-setosa	1	[0.2,1]	10
Row8	0.2	Iris-setosa	1	[0.2,1]	10
Row9	0.1	Iris-setosa	1	[0.1,1]	10
Row10	0.2	Iris-setosa	1	[0.2,1]	10
Row11	0.2	Iris-setosa	1	[0.2,1]	10
Row12	0.1	Iris-setosa	1	[0.1,1]	10
Row13	0.1	Iris-setosa	1	[0.1,1]	10
Row14	0.2	Iris-setosa	1	[0.2,1]	10
Row15	0.4	Iris-setosa	1	[0.4,1]	10
Row16	0.4	Iris-setosa	1	[0.4,1]	10
Row17	0.3	Iris-setosa	1	[0.3,1]	10
Row18	0.3	Iris-setosa	1	[0.3,1]	10
Row19	0.3	Iris-setosa	1	[0.3,1]	10
Row20	0.2	Iris-setosa	1	[0.2,1]	10

Data Types

- Common data types
  - Double Value 

  - Int Value 

  - String Value 

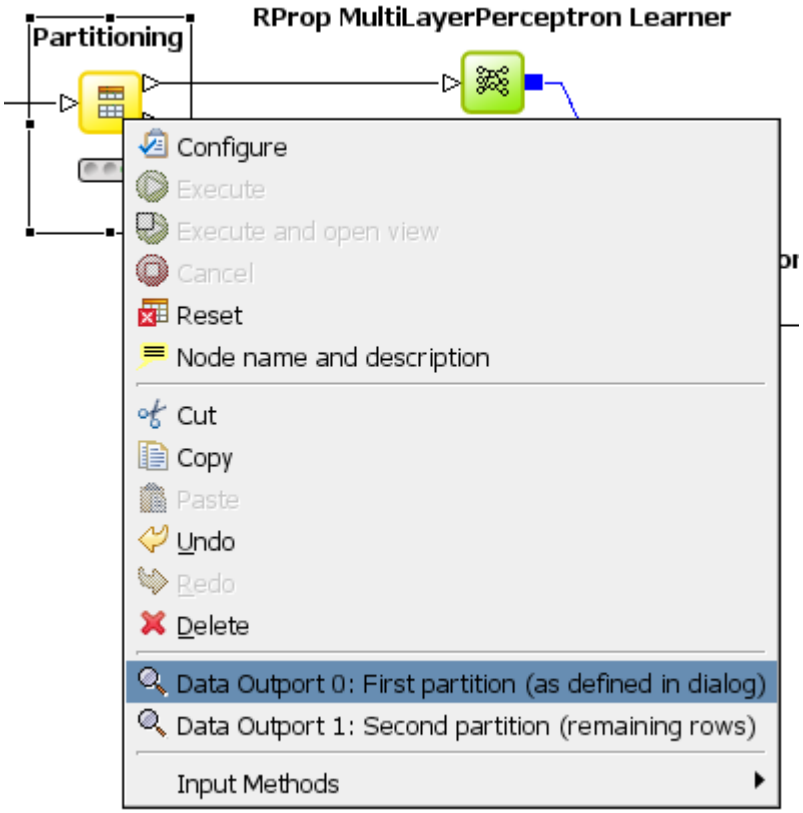
  - Collections 

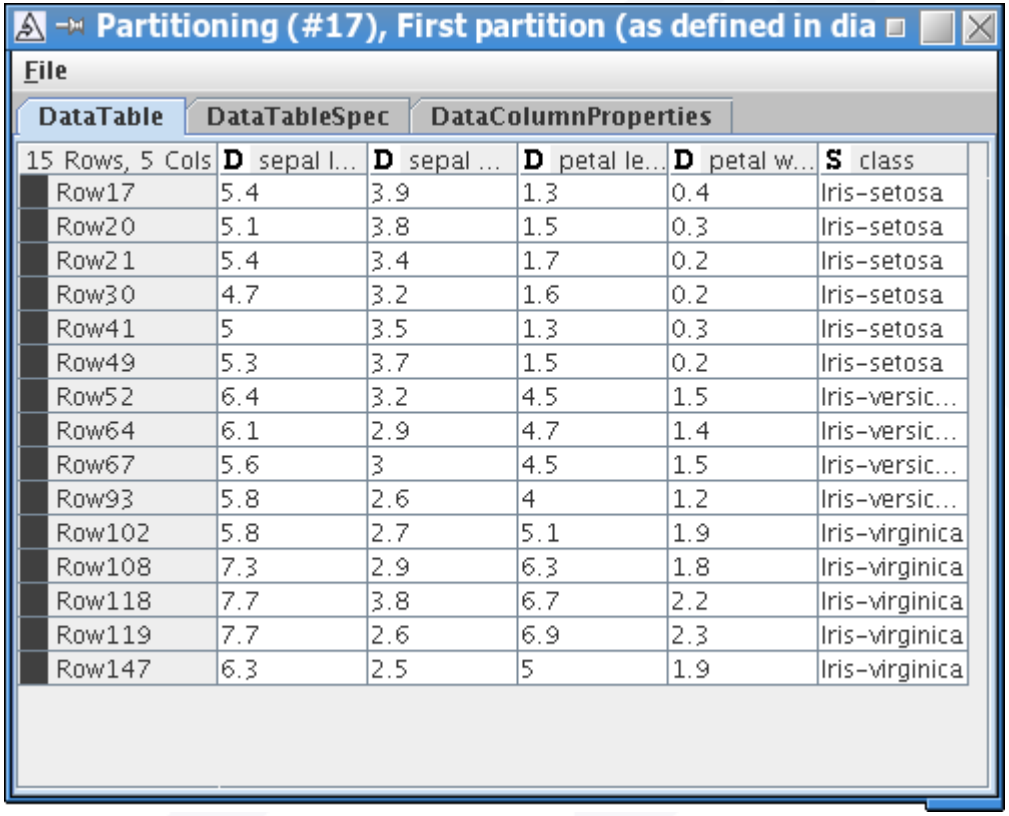
    - Sets
    - Lists
  - Bit vectors 

- Additional data types
  - Terms and Documents
  - Image
  - Network
  - Chemical types
    - Molecules i.e. CDK, Smiles, SDF, …
  - Distance Matrix
  - Custom data types

Intermediate Results

- Quick view on the output tables at each executed node





15 Rows, 5 Cols	D sepal l...	D sepal ...	D petal le...	D petal w...	S class
Row17	5.4	3.9	1.3	0.4	Iris-setosa
Row20	5.1	3.8	1.5	0.3	Iris-setosa
Row21	5.4	3.4	1.7	0.2	Iris-setosa
Row30	4.7	3.2	1.6	0.2	Iris-setosa
Row41	5	3.5	1.3	0.3	Iris-setosa
Row49	5.3	3.7	1.5	0.2	Iris-setosa
Row52	6.4	3.2	4.5	1.5	Iris-versic...
Row64	6.1	2.9	4.7	1.4	Iris-versic...
Row67	5.6	3	4.5	1.5	Iris-versic...
Row93	5.8	2.6	4	1.2	Iris-versic...
Row102	5.8	2.7	5.1	1.9	Iris-virginica
Row108	7.3	2.9	6.3	1.8	Iris-virginica
Row118	7.7	3.8	6.7	2.2	Iris-virginica
Row119	7.7	2.6	6.9	2.3	Iris-virginica
Row147	6.3	2.5	5	1.9	Iris-virginica

Learners

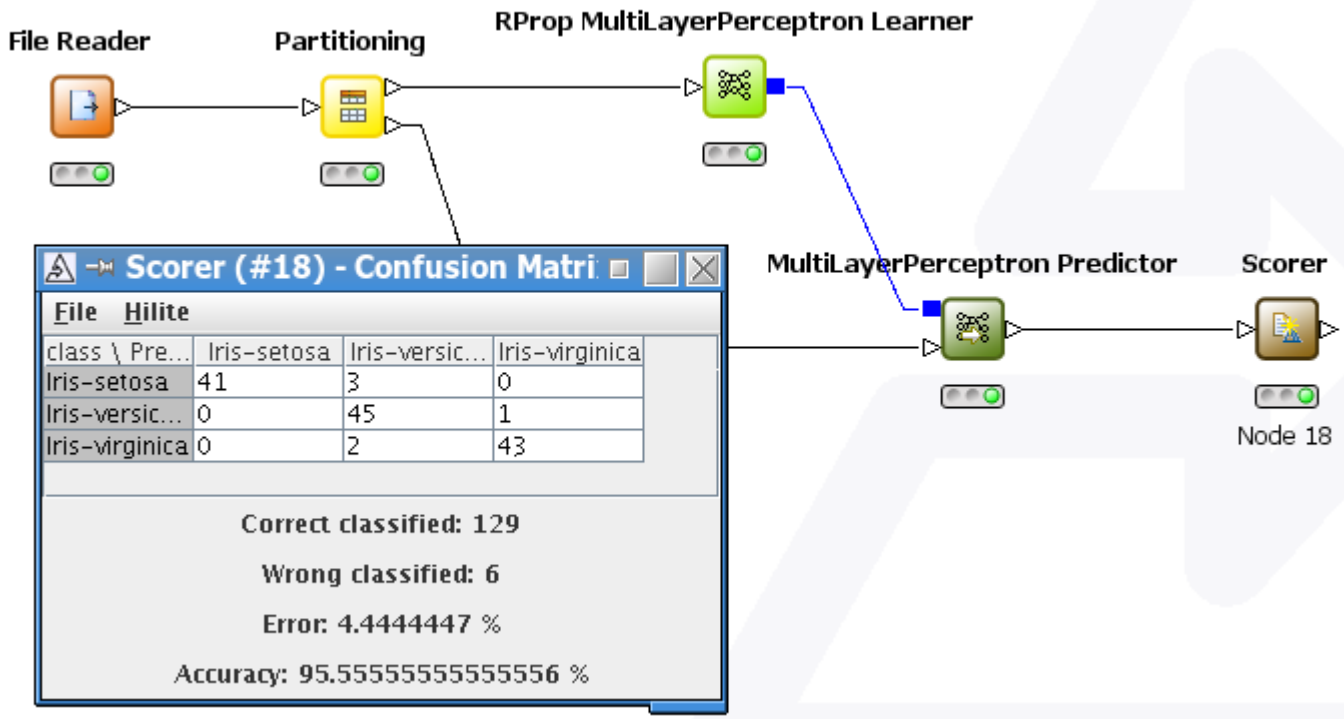
- Learners build “models” based on the input data

- Many different algorithms
  - Decision Trees
  - Neural Networks
  - Regression
  - All of Weka
  - …

- Model output port

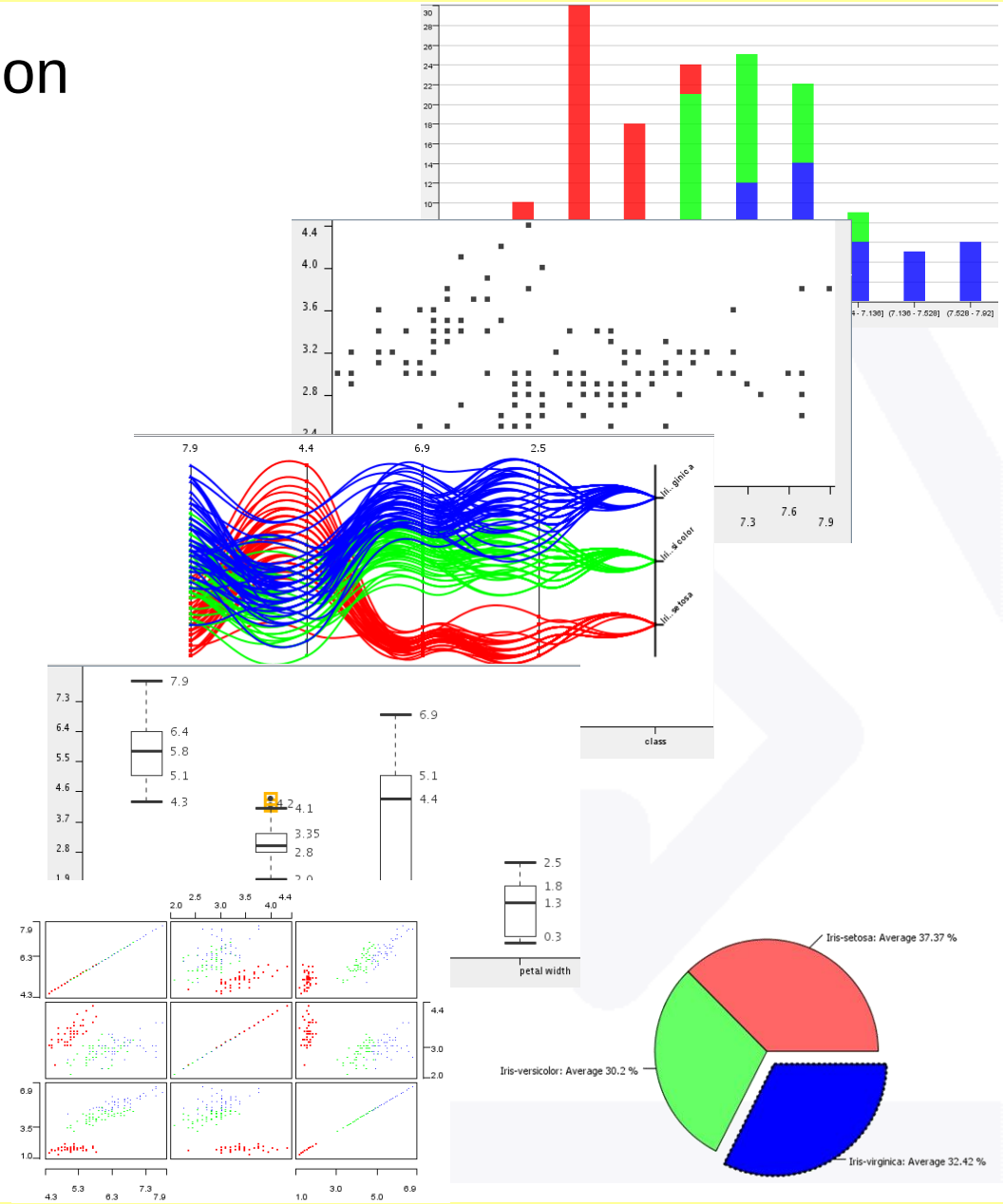
Predictors

- Predictors predict class values
- Take a model from a learner and suitable data



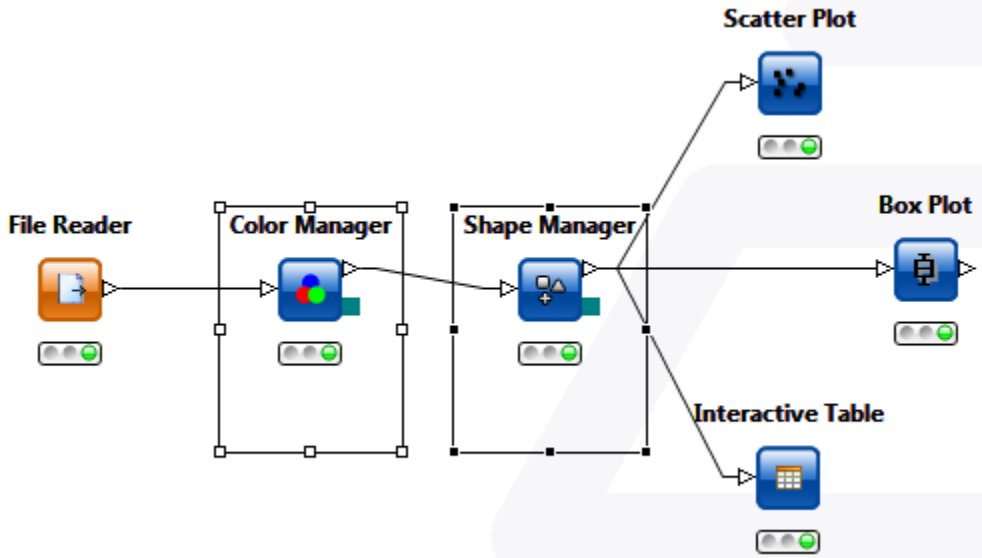
Views

- Many nodes offers views on the input data
  - Scatter Plot, Scatter Matrix
  - Histograms
  - Interactive Table
  - …
- Click *Open View* in the context menu
- Some nodes have more than one view

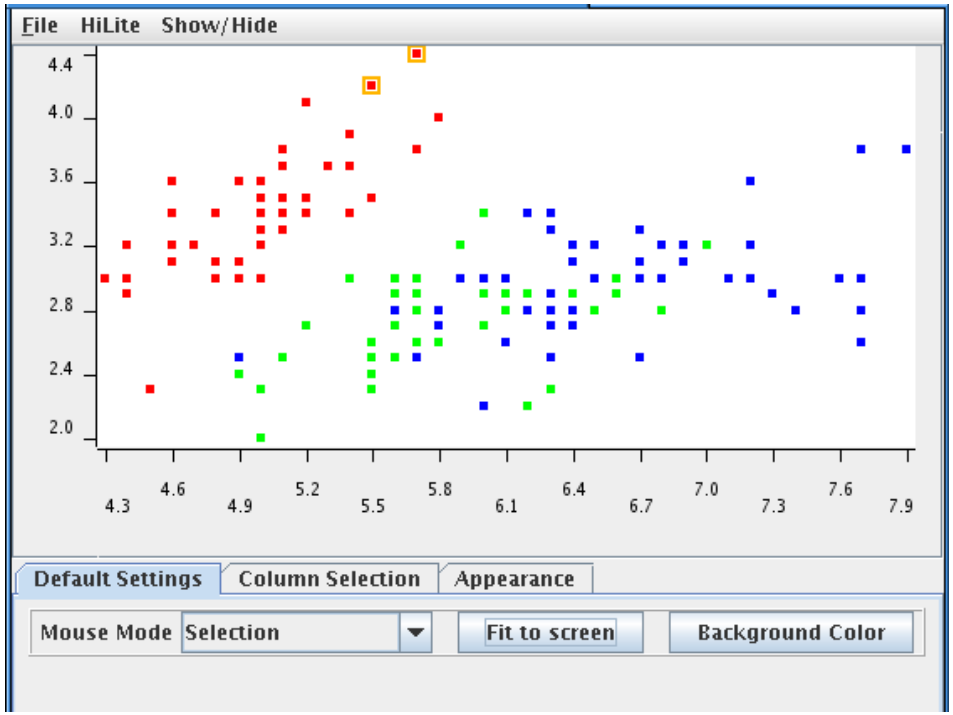


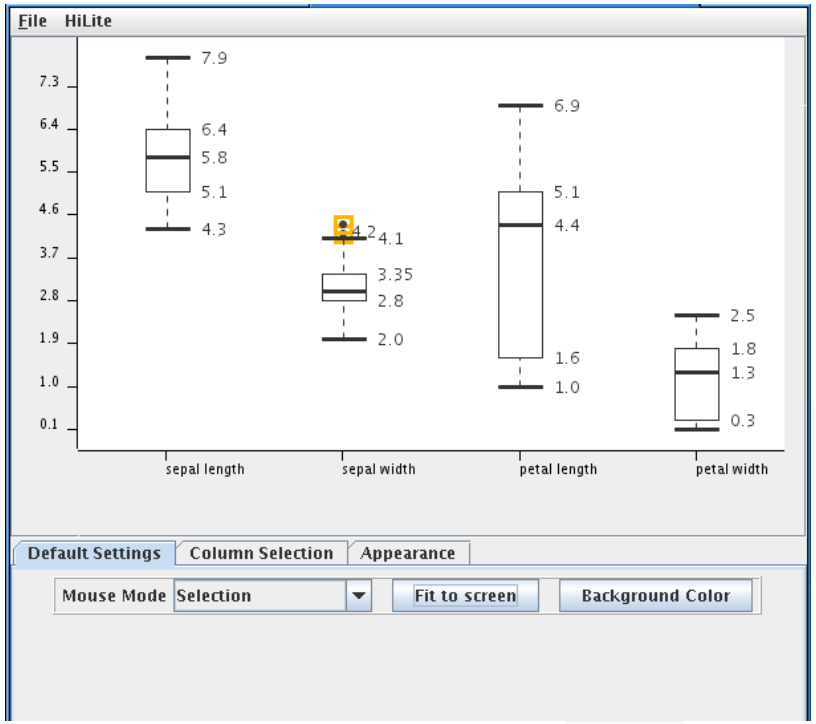
Adding Color and Shape

- Data rows can be assigned colors and shapes
- Based on a certain column
- Special *Color Manager* and *Shape Manager* nodes
- Accessible in all following nodes

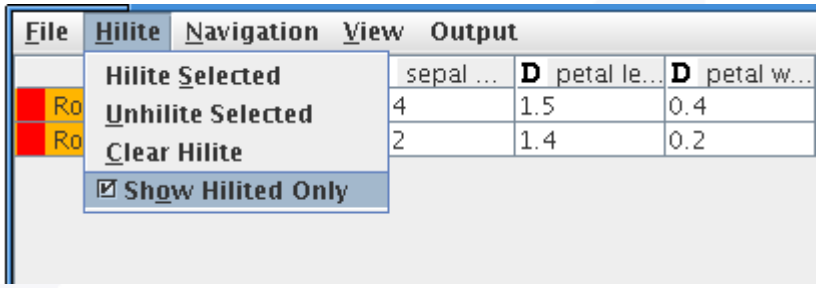


Linked Brushing



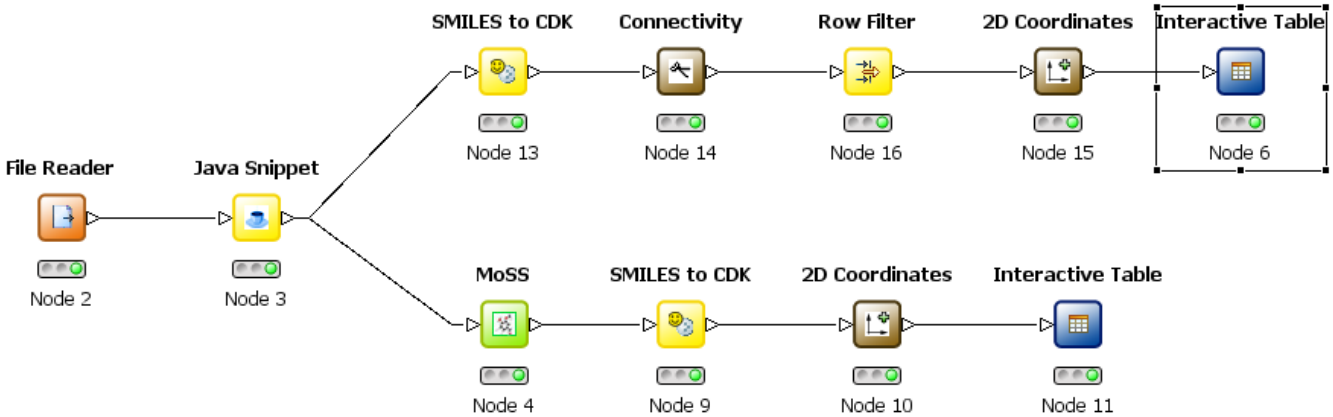


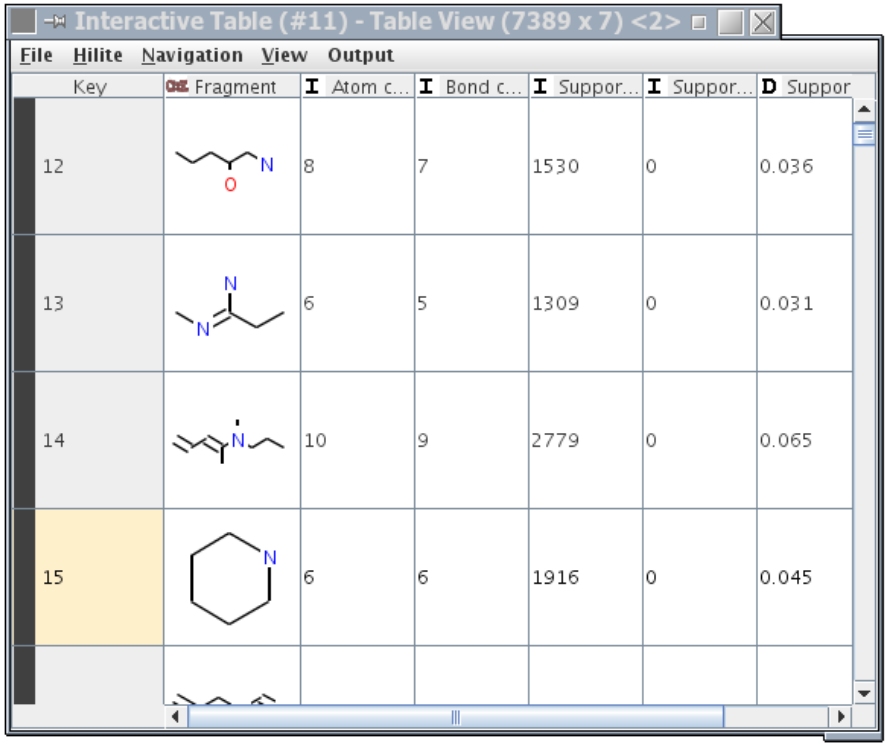
- Hiliting can be changed from every view in the flow that supports it

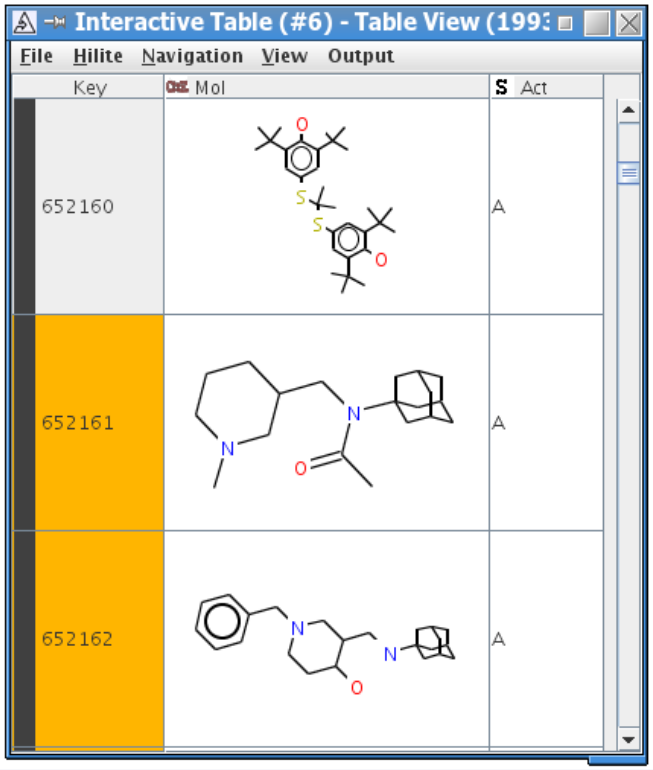


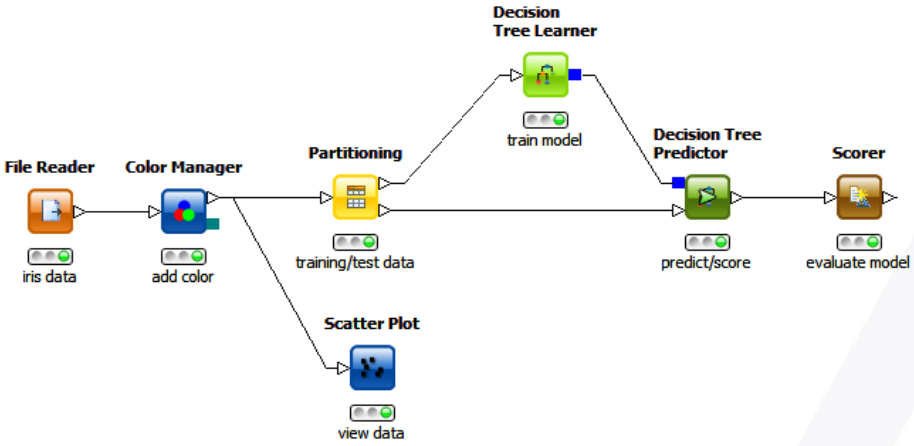
- Views can show only hilited rows

1:n Hiliting


File Reader
Node 2
Java Snippet
Node 3
SMILES to CDK
Node 13
Connectivity
Node 14
Row Filter
Node 16
2D Coordinates
Node 15
Interactive Table
Node 6
MoSS
Node 4
SMILES to CDK
Node 9
2D Coordinates
Node 10
Interactive Table
Node 11


Interactive Table (#11) - Table View (7389 x 7) <2>
File Hilite Navigation View Output
Key
Fragment
Atom c...
Bond c...
Suppor...
Suppor...
Suppor
12 8 7 1530 0 0.036
13 6 5 1309 0 0.031
14 10 9 2779 0 0.065
15 6 6 1916 0 0.045


Interactive Table (#6) - Table View (199
File Hilite Navigation View Output
Key
Mol
Act
652160 A
652161 A
652162 A



EXAMPLE WORKFLOWS

Example Workflow (step by step)

KNIME

File Edit View Node Search Run Help

Workflow Projects

- 001001_TimeSeries_example
- 011001_simpleWorkflowVariable
- 011004_varsAndTimeSeries
- 050003_ChurnAnalysis
- 050004_lastfm_Recommendations
- KNIME_project

Node Repository | Favorite Nodes

- IO
- Database
- Data Manipulation
- Data Views
- Statistics
- Mining
- Distance Matrix
- Meta
- Flow Control
- Misc
- KNIME Labs
- Time Series
  - Date Field Extractor
  - Time Field Extractor
  - Extract Time Window
  - Mask Date/Time
  - Moving Average
  - Preset Date/Time
  - String to Date/Time
  - Time Difference
  - Time Generator
  - Time to String
- Reporting
- Weka
- XML

*0: KNIME_project | *2: 001001_TimeSeries_example

File Reader

Read US dollar rates

Console

KNIME Console

```
DEBUG    KnimeResourceNavigator  Workflow Moving Average 0:2:8 removed
DEBUG    WorkflowEditor  Workflow event triggered: WorkflowEvent [type=NODE_REMOVED;
DEBUG    WorkflowRootEditPart    WorkflowRoot: workflow changed, refreshing children
DEBUG    NodeContainerEditPart   Missing Value 0:2:10 (IDLE)
DEBUG    NodeContainerEditPart   Math Formula 0:2:15 (IDLE)
```

Outline

Node Description

File Reader

This node can be used to read data from an ASCII file or URL location. It can be configured to read various formats.
When you open the node's configuration dialog and provide a filename, it tries to guess the reader's settings by analyzing the content of the file. Check the results of these settings in the preview table. If the data shown is not correct or an error is reported, you can adjust the settings manually (see below).

The file analysis runs in the background and can be cut short by clicking the "Quick scan", which shows if the analysis takes longer. In this case the file is not analyzed completely, but only the first fifty lines are taken into account. It could happen then, that the preview appears looking fine, but the execution of the File Reader fails, when it reads the lines it didn't analyze. Thus it is recommended you check the settings, when you cut an analysis short.

Dialog Options

KNIME Example Flow Server

Workflow Server: publicserver.knime.org

Status: connected as 'guest' Disconnect

- 001_TimeSeries
  - 001001_TimeSeries_example
- 002_DataMining

Example Workflow (step by step)

KNIME

File Edit View Node Search Run Help

Workflow Projects

- 001001_TimeSeries_example
- 011001_simpleWorkflowVariable
- 011004_varsAndTimeSeries
- 050003_ChurnAnalysis
- 050004_lastfm_Recommendations
- KNIME_project

Node Repository | Favorite Nodes

- IO
  - Read
    - File Reader
    - ARFF Reader
    - CSV Reader
    - Line Reader
    - Table Reader
    - PMML Reader
    - XLS Reader
    - Model Reader
  - Write
  - Other
  - Cache
- Database
- Data Manipulation
- Data Views
- Statistics
- Mining
- Distance Matrix
- Meta
- Flow Control
- Misc
- KNIME Labs
- Time Series
- Reporting
- Weka
- XML

*0: KNIME_project | *2: 001001_TimeSeries_example

File Reader

Read US dollar rates

- Configure...
- Execute
- Execute and Open Views
- Cancel
- Reset
- Edit Node Name and Description...
- New Workflow Annotation
- Collapse into Meta Node
- Expand Meta Node
- Show Flow Variable Ports
- Cut
- Copy
- Paste
- Undo
- Redo
- Delete
- 0 File Table

Node Description

File Reader

This node can be used to read data from an ASCII file or URL location. It can be configured to read various formats.
When you open the node's configuration dialog and provide a filename, it tries to guess the reader's settings by analyzing the content of the file. Check the results of these settings in the preview table. If the data shown is not correct or an error is reported, you can adjust the settings manually (see below).

The file analysis runs in the background and can be cut short by clicking the "Quick scan", which shows if the analysis takes longer. In this case the file is not analyzed completely, but only the first fifty lines are taken into account. It could happen then, that the preview appears looking fine, but the execution of the File Reader fails, when it reads the lines it didn't analyze. Thus it is recommended you check the settings, when you cut an analysis short.

Dialog Options

KNIME Example Flow Server

Workflow Server: publicserver.knime.org

Status: connected as 'guest' Disconnect

- 001_TimeSeries
  - 001001_TimeSeries_example
- 002_DataMining

Console

KNIME Console

```
DEBUG    NodeContainer    File Reader 0:2:1 has new state: CONFIGURED
DEBUG    KnimeResourceNavigator  state changed to CONFIGURED
DEBUG    NodeContainer    001001_TimeSeries_example 0:2 has new state: CONFIGURED
DEBUG    NodeContainer    001001_TimeSeries_example 0:2 has new state: CONFIGURED
DEBUG    NodeContainer    Workflow Manager 0 has new state: IDLE
```

Outline

Example Workflow (step by step)

KNIME

File Edit View Node Search Run Help

Workflow Projects

- 001001_TimeSeries_example
- 011001_simpleWorkflowVariable
- 011004_varsAndTimeSeries
- 050003_ChurnAnalysis
- 050004_lastfm_Recommendations
- KNIME_project

Node Repository | Favorite Nodes

- IO
  - Read
    - File Reader
    - ARFF Reader
    - CSV Reader
    - Line Reader
    - Table Reader
    - PMML Reader
    - XLS Reader
    - Model Reader
  - Write
  - Other
  - Cache
- Database
- Data Manipulation
- Data Views
- Statistics
- Mining
- Distance Matrix
- Meta
- Flow Control
- Misc
- KNIME Labs
- Time Series
- Reporting
- Weka
- XML

*0: KNIME_project | *2: 001001_TimeSeries_example

File Reader

Read US dollar rates

Console

KNIME Console

```
DEBUG    NodeContainer    001001_TimeSeries_example 0:2 has new state: EXECUTED
DEBUG    NodeContainer    Workflow Manager 0 has new state: IDLE
DEBUG    NodeContainer    001001_TimeSeries_example 0:2 has new state: EXECUTED
DEBUG    OpenPortViewAction      Open Port View File Reader (#1)
DEBUG    OpenPortViewAction      Open Port View File Reader (#1)
```

Outline

Node Description

File Reader

This node can be used to read data from an ASCII file or URL location. It can be configured to read various formats.
When you open the node's configuration dialog and provide a filename, it tries to guess the reader's settings by analyzing the content of the file. Check the results of these settings in the preview table. If the data shown is not correct or an error is reported, you can adjust the settings manually (see below).

The file analysis runs in the background and can be cut short by clicking the "Quick scan", which shows if the analysis takes longer. In this case the file is not analyzed completely, but only the first fifty lines are taken into account. It could happen then, that the preview appears looking fine, but the execution of the File Reader fails, when it reads the lines it didn't analyze. Thus it is recommended you check the settings, when you cut an analysis short.

Dialog Options

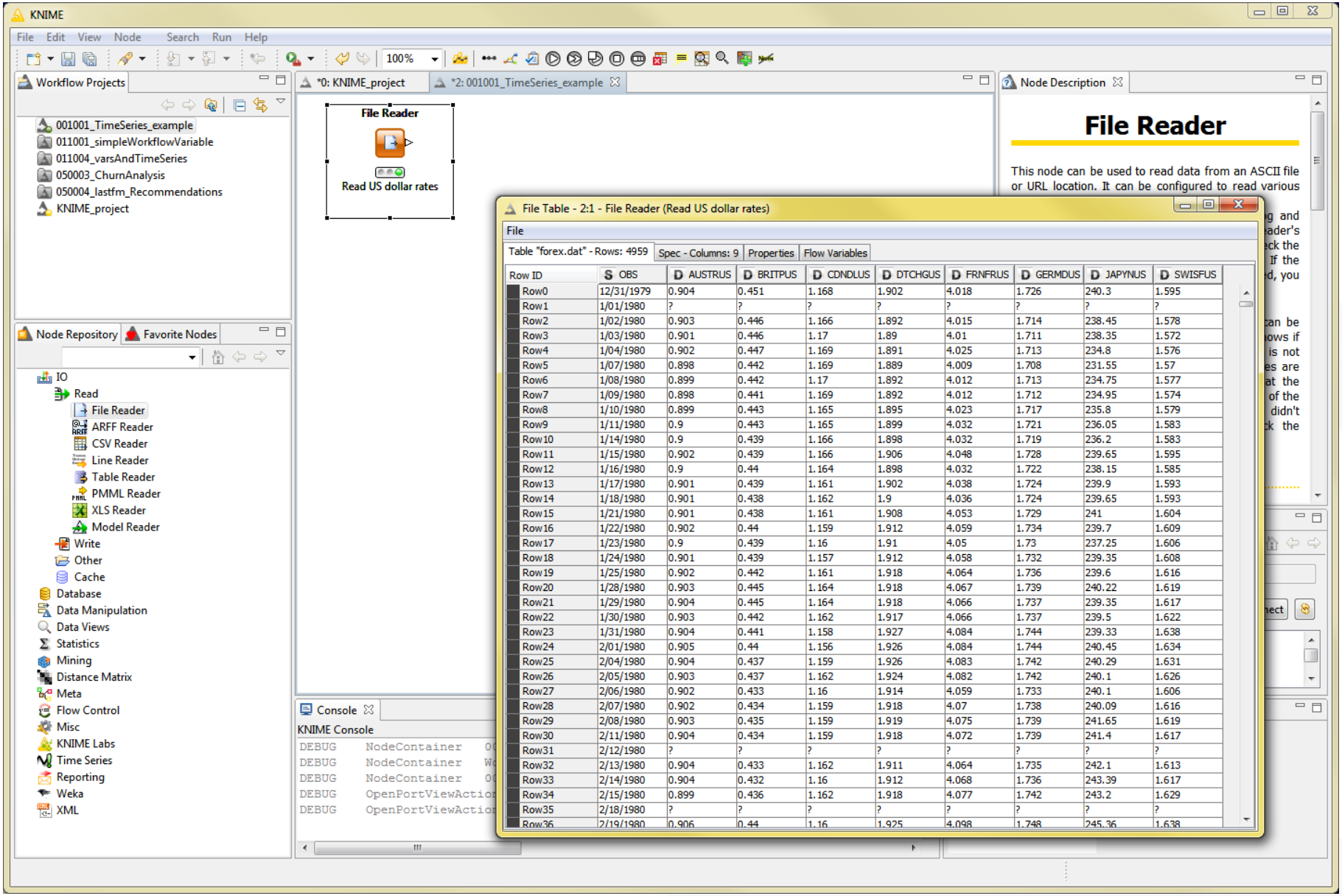
KNIME Example Flow Server

Workflow Server: publicserver.knime.org

Status: connected as 'guest' Disconnect

- 001_TimeSeries
  - 001001_TimeSeries_example
- 002_DataMining

Example Workflow (step by step)



Row ID	S OBS	D AUSTRUS	D BRITPUS	D CDNDLUS	D DTCHGUS	D FRNFRUS	D GERMDUS	D JAPYNUS	D SWISFUS
Row0	12/31/1979	0.904	0.451	1.168	1.902	4.018	1.726	240.3	1.595
Row1	1/01/1980	?	?	?	?	?	?	?	?
Row2	1/02/1980	0.903	0.446	1.166	1.892	4.015	1.714	238.45	1.578
Row3	1/03/1980	0.901	0.446	1.17	1.89	4.01	1.711	238.35	1.572
Row4	1/04/1980	0.902	0.447	1.169	1.891	4.025	1.713	234.8	1.576
Row5	1/07/1980	0.898	0.442	1.169	1.889	4.009	1.708	231.55	1.57
Row6	1/08/1980	0.899	0.442	1.17	1.892	4.012	1.713	234.75	1.577
Row7	1/09/1980	0.898	0.441	1.169	1.892	4.012	1.712	234.95	1.574
Row8	1/10/1980	0.899	0.443	1.165	1.895	4.023	1.717	235.8	1.579
Row9	1/11/1980	0.9	0.443	1.165	1.899	4.032	1.721	236.05	1.583
Row10	1/14/1980	0.9	0.439	1.166	1.898	4.032	1.719	236.2	1.583
Row11	1/15/1980	0.902	0.439	1.166	1.906	4.048	1.728	239.65	1.595
Row12	1/16/1980	0.9	0.44	1.164	1.898	4.032	1.722	238.15	1.585
Row13	1/17/1980	0.901	0.439	1.161	1.902	4.038	1.724	239.9	1.593
Row14	1/18/1980	0.901	0.438	1.162	1.9	4.036	1.724	239.65	1.593
Row15	1/21/1980	0.901	0.438	1.161	1.908	4.053	1.729	241	1.604
Row16	1/22/1980	0.902	0.44	1.159	1.912	4.059	1.734	239.7	1.609
Row17	1/23/1980	0.9	0.439	1.16	1.91	4.05	1.73	237.25	1.606
Row18	1/24/1980	0.901	0.439	1.157	1.912	4.058	1.732	239.35	1.608
Row19	1/25/1980	0.902	0.442	1.161	1.918	4.064	1.736	239.6	1.616
Row20	1/28/1980	0.903	0.445	1.164	1.918	4.067	1.739	240.22	1.619
Row21	1/29/1980	0.904	0.445	1.164	1.918	4.066	1.737	239.35	1.617
Row22	1/30/1980	0.903	0.442	1.162	1.917	4.066	1.737	239.5	1.622
Row23	1/31/1980	0.904	0.441	1.158	1.927	4.084	1.744	239.33	1.638
Row24	2/01/1980	0.905	0.44	1.156	1.926	4.084	1.744	240.45	1.634
Row25	2/04/1980	0.904	0.437	1.159	1.926	4.083	1.742	240.29	1.631
Row26	2/05/1980	0.903	0.437	1.162	1.924	4.082	1.742	240.1	1.626
Row27	2/06/1980	0.902	0.433	1.16	1.914	4.059	1.733	240.1	1.606
Row28	2/07/1980	0.902	0.434	1.159	1.918	4.07	1.738	240.09	1.616
Row29	2/08/1980	0.903	0.435	1.159	1.919	4.075	1.739	241.65	1.619
Row30	2/11/1980	0.904	0.434	1.159	1.918	4.072	1.739	241.4	1.617
Row31	2/12/1980	?	?	?	?	?	?	?	?
Row32	2/13/1980	0.904	0.433	1.162	1.911	4.064	1.735	242.1	1.613
Row33	2/14/1980	0.904	0.432	1.16	1.912	4.068	1.736	243.39	1.617
Row34	2/15/1980	0.899	0.436	1.162	1.918	4.077	1.742	243.2	1.629
Row35	2/18/1980	?	?	?	?	?	?	?	?
Row36	2/19/1980	0.906	0.44	1.16	1.925	4.098	1.748	245.36	1.638

Example Workflow (step by step)

KNIME

File Edit View Node Search Run Help

Workflow Projects

- 001001_TimeSeries_example
- 011001_simpleWorkflowVariable
- 011004_varsAndTimeSeries
- 050003_ChurnAnalysis
- 050004_lastfm_Recommendations
- KNIME_project

Node Repository | Favorite Nodes

- IO
- Database
- Data Manipulation
- Data Views
- Statistics
- Mining
- Distance Matrix
- Meta
- Flow Control
- Misc
- KNIME Labs
- Time Series
  - Date Field Extractor
  - Time Field Extractor
  - Extract Time Window
  - Mask Date/Time
  - Moving Average
  - Preset Date/Time
  - String to Date/Time
  - Time Difference
  - Time Generator
  - Time to String
- Reporting
- Weka
- XML

*0: KNIME_project | *2: 001001_TimeSeries_example

File Reader → String to Date/Time

Read US dollar rates | change string to date column

Dialog - 2:2 - String to Date/Time (change string to dat...

File

Options | Flow Variables | Memory Policy

Select column: S OBS

☑ Replace selected column

New column name: OBS_time

Date format: MM/dd/yyyy

Abort execution

☑ Abort execution... ...after this number of unresolved rows: 100

OK | Apply | Cancel

Node Description

String to Date/Time

Parses the strings in the selected String column and converts them into date/time cells by using the entered format pattern.

Dialog Options

Select column

Select the input column containing the timestamp string.

Replace selected column

Check if the string column is not needed anymore and should be replaced by the parsed and converted date/time values.

New column name

If the string column should not be replaced enter here the name for the new column containing the parsed and converted date/times.

Date format

KNIME Example Flow Server

Workflow Server: publicserver.knime.org

Status: connected as 'guest' | Disconnect

- 001_TimeSeries
  - 001001_TimeSeries_example
- 002_DataMining

Console

KNIME Console

```
DEBUG    KnimeResourceNavigator  Workflow Extract Time Window 0:2:5 removed
DEBUG    WorkflowEditor  Workflow event triggered: WorkflowEvent [type=NODE_REMOVED;
DEBUG    WorkflowRootEditPart    WorkflowRoot: workflow changed, refreshing children
DEBUG    NodeContainerEditPart   String to Date/Time 0:2:2 (CONFIGURED)
DEBUG    OpenDialogAction        Opening node dialog...
```

Outline

Example Workflow (step by step)

KNIME

File Edit View Node Search Run Help

Workflow Projects

- 001001_TimeSeries_example
- 011001_simpleWorkflowVariable
- 011004_varsAndTimeSeries
- 050003_ChurnAnalysis
- 050004_lastfm_Recommendations
- KNIME_project

*0: KNIME_project | *2: 001001_TimeSeries_example

File Reader → String to Date/Time

Read US dollar rates | change string to date column

Node Repository | Favorite Nodes

- IO
- Database
- Data Manipulation
- Data Views
- Statistics
- Mining
- Distance Matrix
- Meta
- Flow Control
- Misc
- KNIME Labs
- Time Series
  - Date Field Extractor
  - Time Field Extractor
  - Extract Time Window
  - Mask Date/Time
  - Moving Average
  - Preset Date/Time
  - String to Date/Time
  - Time Difference
  - Time Generator
  - Time to String
- Reporting
- Weka
- XML

Console

KNIME Console

```
DEBUG    KnimeResourceNavigator  stat
DEBUG    NodeContainer    001001_TimeS
DEBUG    NodeContainer    001001_TimeS
DEBUG    NodeContainer    Workflow Manager 0 has new state: IDLE
DEBUG    OpenPortViewAction      Open Port View String to Date/Time (#1)
```

Node Description

String to Date/Time

Parses the strings in the selected String column and converts them into date/time cells by using the entered format pattern.

Parsed time - 2:2 - String to Date/Time (change string to date column)

File

Table "default" - Rows: 4959 | Spec - Columns: 9 | Properties | Flow Variables

Row ID	OBS	AUSTRUS	BRITPUS	CDNDLUS	DTCHGUS	FRNFRUS	D
Row30	11.Feb.1980	0.904	0.434	1.159	1.918	4.072	1.73
Row31	12.Feb.1980	?	?	?	?	?	?
Row32	13.Feb.1980	0.904	0.433	1.162	1.911	4.064	1.73
Row33	14.Feb.1980	0.904	0.432	1.16	1.912	4.068	1.73
Row34	15.Feb.1980	0.899	0.436	1.162	1.918	4.077	1.74
Row35	18.Feb.1980	?	?	?	?	?	?
Row36	19.Feb.1980	0.906	0.44	1.16	1.925	4.098	1.74
Row37	20.Feb.1980	0.907	0.439	1.155	1.926	4.095	1.74
Row38	21.Feb.1980	0.907	0.437	1.151	1.925	4.096	1.74
Row39	22.Feb.1980	0.908	0.44	1.153	1.938	4.127	1.75
Row40	25.Feb.1980	0.909	0.439	1.151	1.941	4.132	1.76
Row41	26.Feb.1980	0.91	0.439	1.15	1.945	4.133	1.76
Row42	27.Feb.1980	0.91	0.437	1.146	1.939	4.125	1.75
Row43	28.Feb.1980	0.91	0.438	1.146	1.94	4.13	1.76
Row44	29.Feb.1980	0.912	0.44	1.145	1.954	4.16	1.77
Row45	03.Mar.1980	0.912	0.446	1.143	1.96	4.182	1.78
Row46	04.Mar.1980	0.911	0.447	1.153	1.968	4.192	1.79
Row47	05.Mar.1980	0.912	0.446	1.147	1.963	4.19	1.78
Row48	06.Mar.1980	0.913	0.448	1.153	1.971	4.202	1.79
Row49	07.Mar.1980	0.913	0.449	1.158	1.972	4.202	1.79
Row50	10.Mar.1980	0.914	0.451	1.164	1.983	4.23	1.81
Row51	11.Mar.1980	0.912	0.449	1.165	1.988	4.242	1.81
Row52	12.Mar.1980	0.912	0.448	1.167	1.987	4.228	1.80
Row53	13.Mar.1980	0.914	0.45	1.17	1.999	4.25	1.82
Row54	14.Mar.1980	0.915	0.452	1.179	2.011	4.276	1.83
Row55	17.Mar.1980	0.918	0.457	1.185	2.059	4.376	1.87
Row56	18.Mar.1980	0.92	0.457	1.18	2.058	4.385	1.87
Row57	19.Mar.1980	0.921	0.457	1.177	2.056	4.371	1.87
Row58	20.Mar.1980	0.92	0.456	1.179	2.047	4.349	1.86
Row59	21.Mar.1980	0.921	0.458	1.183	2.056	4.365	1.87

Example Workflow (step by step)

KNIME

File Edit View Node Search Run Help

Workflow Projects

- 001001_TimeSeries_example
- 011001_simpleWorkflowVariable
- 011004_varsAndTimeSeries
- 050003_ChurnAnalysis
- 050004_lastfm_Recommendations
- KNIME_project

Node Repository | Favorite Nodes

- IO
- Database
- Data Manipulation
- Data Views
- Statistics
- Mining
- Distance Matrix
- Meta
- Flow Control
- Misc
- KNIME Labs
- Time Series
  - Date Field Extractor
  - Time Field Extractor
  - Extract Time Window
  - Mask Date/Time
  - Moving Average
  - Preset Date/Time
  - String to Date/Time
  - Time Difference
  - Time Generator
  - Time to String
- Reporting
- Weka
- XML

*0: KNIME_project | *2: 001001_TimeSeries_example

File Reader → String to Date/Time → Extract Time Window

Read US dollar rates | change string to date column | only years 85 - 95

Dialog - 2:5 - Extract Time Window (on...

File

Options | Flow Variables | Memory Policy

Columns containing Timestamp: OBS

Select starting point:

Use date

Date: Year: 1985 Month: 1 Day: 1

Use Time

Time: Hour: 0 Minute: 0 Second: 0 Milli: 0

Select end point:

Use date

Date: Year: 1995 Month: 12 Day: 31

Use Time

Time: Hour: 0 Minute: 0 Second: 0 Milli: 0

OK | Apply | Cancel

Console

KNIME Console

```
DEBUG    ConnectionContainerEditPart      modelli
DEBUG    WorkflowRootEditPart     part: NodeConta
DEBUG    ConnectionContainerEditPart      refresh
DEBUG    ConnectionContainerEditPart      modelli
DEBUG    OpenDialogAction         Opening node di
```

Node Description

Extract Time Window

Extracts all rows where the time value of the selected column lies within a given time window from the input. The time window is specified by a start and end date (and time).

Dialog Options

Columns containing date/time
Select the input column containing date/time stamps.

Starting point
Start date/time of the time window.

End point:
End date/time of the time window.

Ports

Input Ports

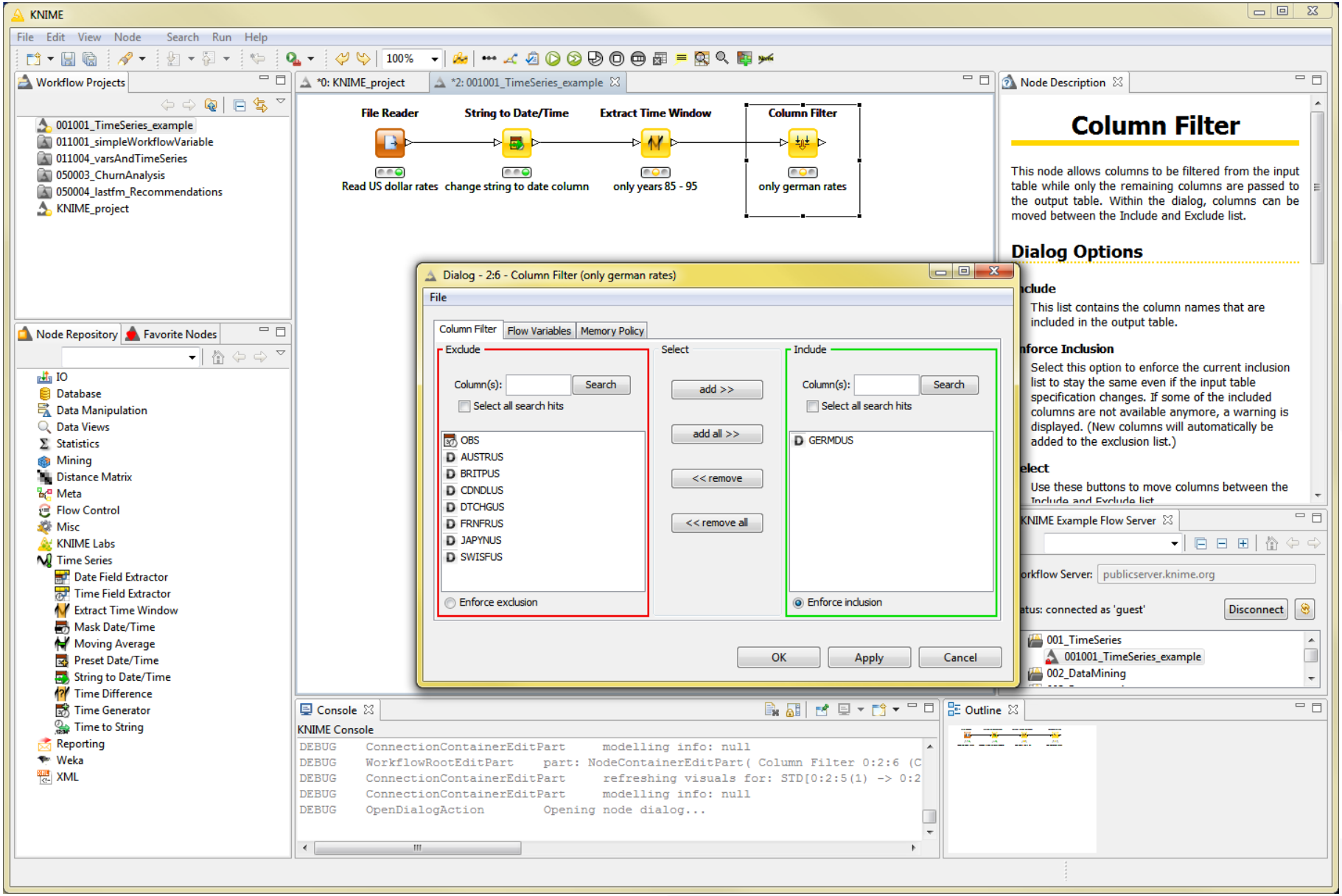
KNIME Example Flow Server

Workflow Server: publicserver.knime.org

Status: connected as 'guest' | Disconnect

- 001_TimeSeries
  - 001001_TimeSeries_example
- 002_DataMining

Example Workflow (step by step)



Example Workflow (step by step)

KNIME

File Edit View Node Search Run Help

Workflow Projects

- 001001_TimeSeries_example
- 011001_simpleWorkflowVariable
- 011004_varsAndTimeSeries
- 050003_ChurnAnalysis
- 050004_lastfm_Recommendations
- KNIME_project

Node Repository | Favorite Nodes

- IO
- Database
- Data Manipulation
- Data Views
- Statistics
- Mining
- Distance Matrix
- Meta
- Flow Control
- Misc
- KNIME Labs
- Time Series
  - Date Field Extractor
  - Time Field Extractor
  - Extract Time Window
  - Mask Date/Time
  - Moving Average
  - Preset Date/Time
  - String to Date/Time
  - Time Difference
  - Time Generator
  - Time to String
- Reporting
- Weka
- XML

*0: KNIME_project | *2: 001001_TimeSeries_example

File Reader → String to Date/Time → Extract Time Window → Column Filter → Line Plot

Read US dollar rates | change string to date column | only years 85 - 95 | only german rates | plot the rates

Node Description

KNIME Example Flow Server

Workflow Server: publicserver.knime.org

Status: connected as 'guest' Disconnect

- 001_TimeSeries
  - 001001_TimeSeries_example
- 002_DataMining

Console

KNIME Console

```
DEBUG    ConnectionContainerEditPart     modelling info: null
DEBUG    WorkflowRootEditPart    part: NodeContainerEditPart( Line Plot 0:2:7 (CONFI
DEBUG    ConnectionContainerEditPart     refreshing visuals for: STD[0:2:6(1) -> 0:2
DEBUG    ConnectionContainerEditPart     modelling info: null
DEBUG    NodeContainerEditPart   Line Plot 0:2:7 (CONFIGURED)
```

Outline

Example Workflow (step by step)

KNIME

File Edit View Node Search Run Help

Workflow Projects

- 001001_TimeSeries_example
- 011001_simpleWorkflowVariable
- 011004_varsAndTimeSeries
- 050003_ChurnAnalysis
- 050004_lastfm_Recommendations
- KNIME_project

Node Repository | Favorite Nodes

- IO
- Database
- Data Manipulation
- Data Views
- Statistics
- Mining
- Distance Matrix
- Meta
- Flow Control
- Misc
- KNIME Labs
- Time Series
  - Date Field Extractor
  - Time Field Extractor
  - Extract Time Window
  - Mask Date/Time
  - Moving Average
  - Preset Date/Time
  - String to Date/Time
  - Time Difference
  - Time Generator
  - Time to String
- Reporting
- Weka
- XML

*0: KNIME_project | *2: 001001_TimeSeries_example

Console

KNIME Console

```
DEBUG    NodeContainer    001001_TimeSeries_example 0:2 has new state: EXECUTED
DEBUG    NodeContainer    001001_TimeSeries_example 0:2 has new state: EXECUTED
DEBUG    NodeContainer    Workflow Manager 0 has new state: IDLE
DEBUG    OpenViewAction   Open Node View Line Plot (#0)
DEBUG    DefaultVisualizationNodeModel   Registering view at model (total count 1)
```

Outline

Node Description

Line Plot

Plots the numeric columns of the input table as lines. All values are mapped to a single y coordinate. This may distort the visualization if the difference of the values in the columns is large.

Only columns with a valid domain are available in this view. Make sure that the predecessor node is executed or set the domain with the DomainCalculator node!

HiLiting:

The data points may be selected by either dragging a rectangle with the mouse over the points or by clicking on points (if the dots are not displayed you won't see the selection or hilite since only data points can be selected or hilited). Hold control pressed for multiple selections. The selected datapoints may be hilited by either right-click to get the context menu or via the hilite menu in the menu bar.

Tooltip

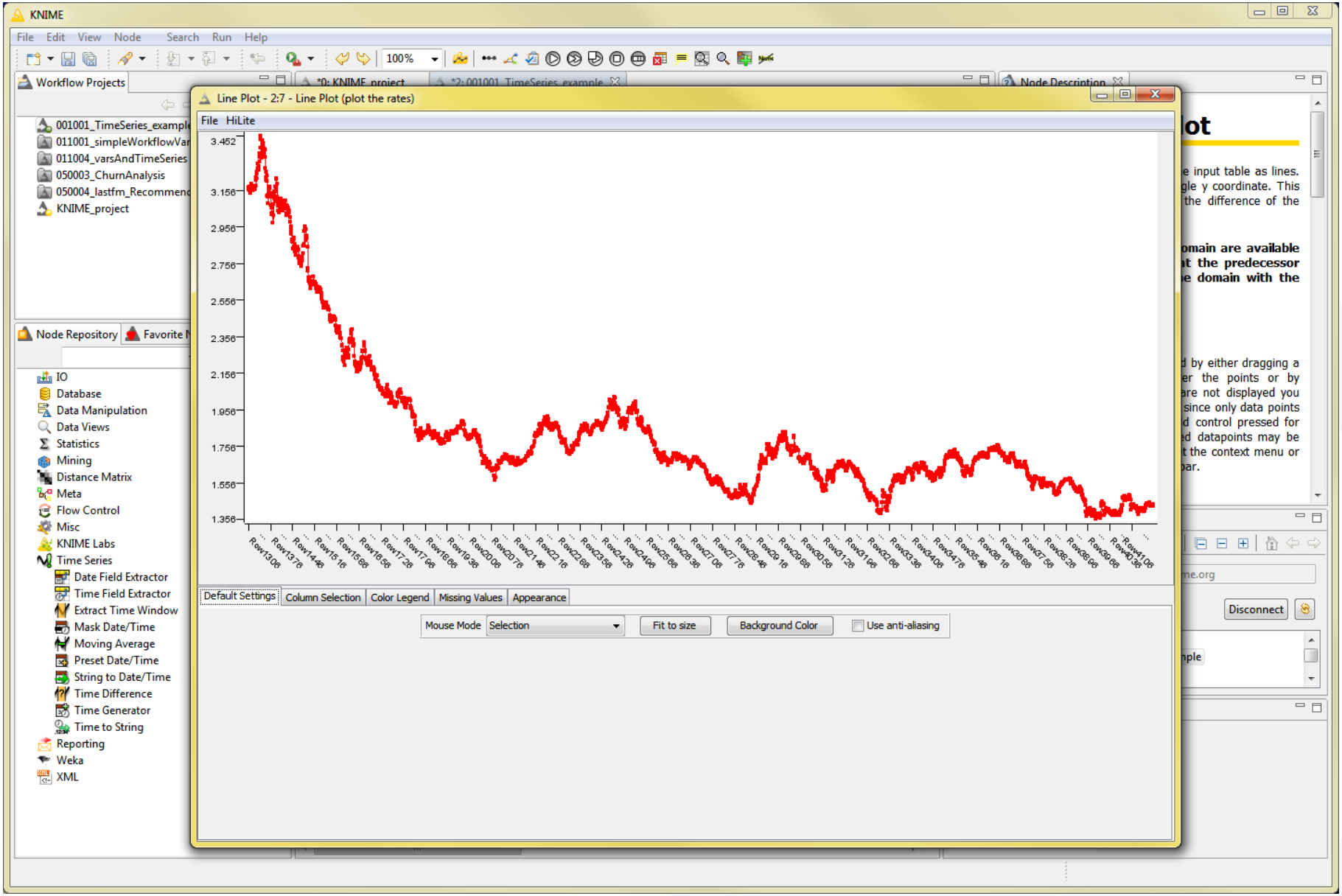
KNIME Example Flow Server

Workflow Server: publicserver.knime.org

Status: connected as 'guest' Disconnect

- 001_TimeSeries
  - 001001_TimeSeries_example
- 002_DataMining

Example Workflow (step by step)



Example Workflow (step by step)

KNIME

File Edit View Node Search Run Help

Workflow Projects

- 001001_TimeSeries_example
- 011001_simpleWorkflowVariable
- 011004_varsAndTimeSeries
- 050003_ChurnAnalysis
- 050004_lastfm_Recommendations
- KNIME_project

Node Repository | Favorite Nodes

- IO
- Database
- Data Manipulation
- Data Views
- Statistics
- Mining
- Distance Matrix
- Meta
- Flow Control
- Misc
- KNIME Labs
- Time Series
  - Date Field Extractor
  - Time Field Extractor
  - Extract Time Window
  - Mask Date/Time
  - Moving Average
  - Preset Date/Time
  - String to Date/Time
  - Time Difference
  - Time Generator
  - Time to String
- Reporting
- Weka
- XML

*0: KNIME_project | *2: 001001_TimeSeries_example

File Reader → String to Date/Time → Extract Time Window → Column Filter → Line Plot

Read US dollar rates | change string to date column | only years 85 - 95 | only german rates | plot the rates

Column Filter: only swiss rates

Node Description

Column Filter

This node allows columns to be filtered from the input table while only the remaining columns are passed to the output table. Within the dialog, columns can be moved between the Include and Exclude list.

Dialog Options

Include

This list contains the column names that are included in the output table.

Enforce Inclusion

Select this option to enforce the current inclusion list to stay the same even if the input table specification changes. If some of the included columns are not available anymore, a warning is displayed. (New columns will automatically be added to the exclusion list.)

KNIME Example Flow Server

Workflow Server: publicserver.knime.org

Status: connected as 'guest' Disconnect

- 001_TimeSeries
  - 001001_TimeSeries_example

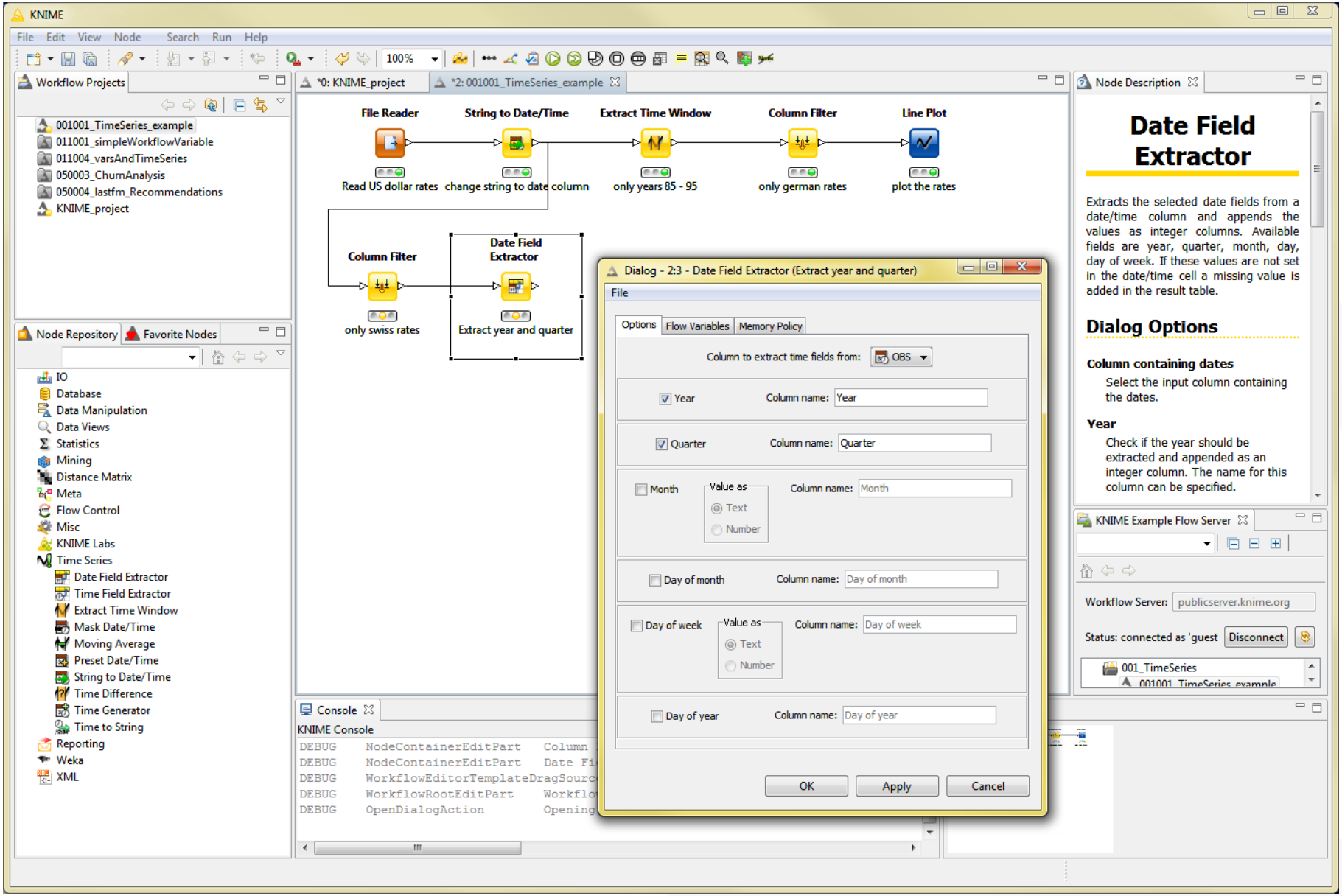
Console

KNIME Console

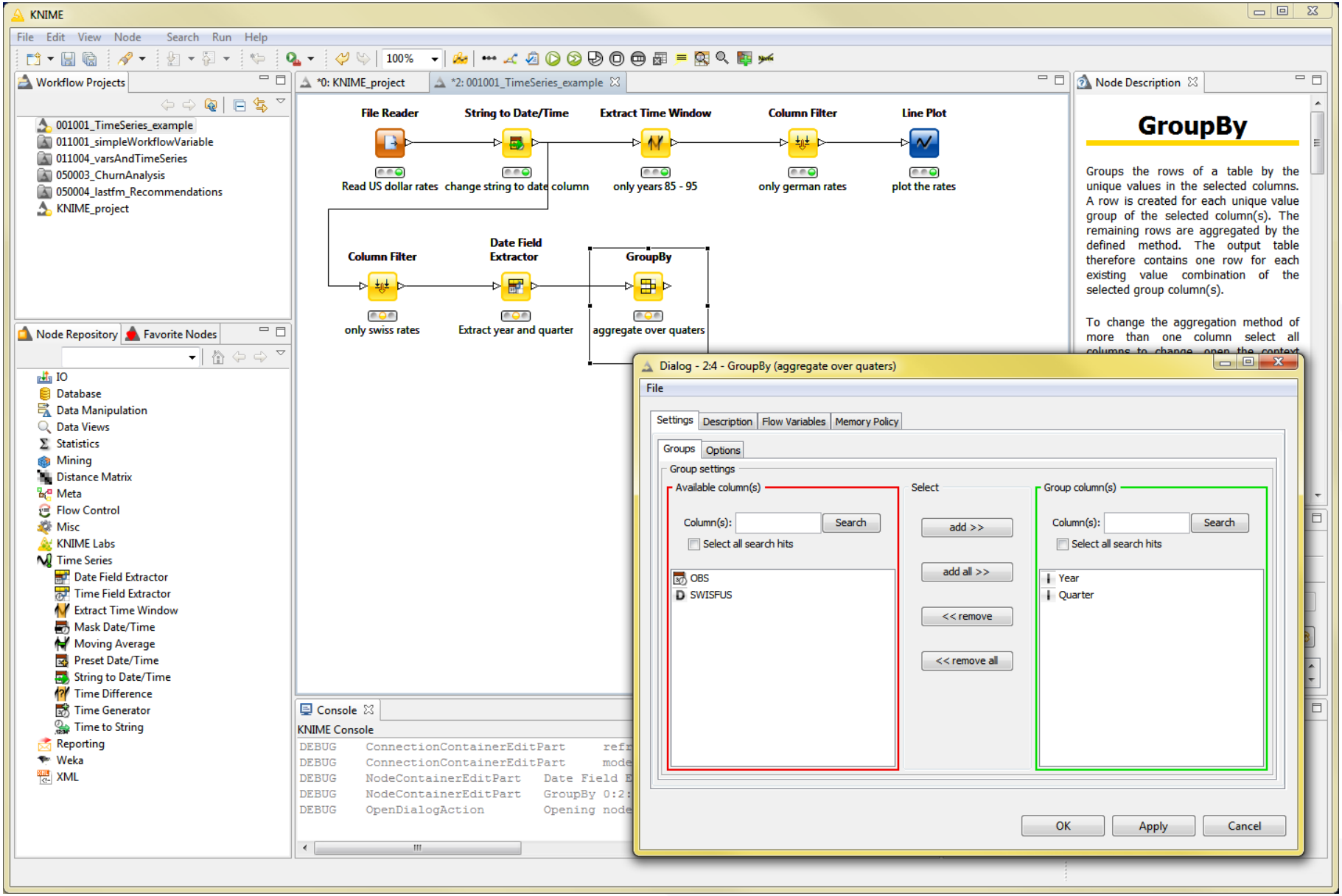
```
DEBUG    KnimeResourceNavigator  ROOT's workflow has changed NODE_REMOVED
DEBUG    KnimeResourceNavigator  Workflow Date Field Extractor 0:2:3 removed
DEBUG    WorkflowEditor  Workflow event triggered: WorkflowEvent [type=NODE_REMOVED;
DEBUG    WorkflowRootEditPart    WorkflowRoot: workflow changed, refreshing children
DEBUG    NodeContainerEditPart   Column Filter 0:2:9 (CONFIGURED)
```

Outline

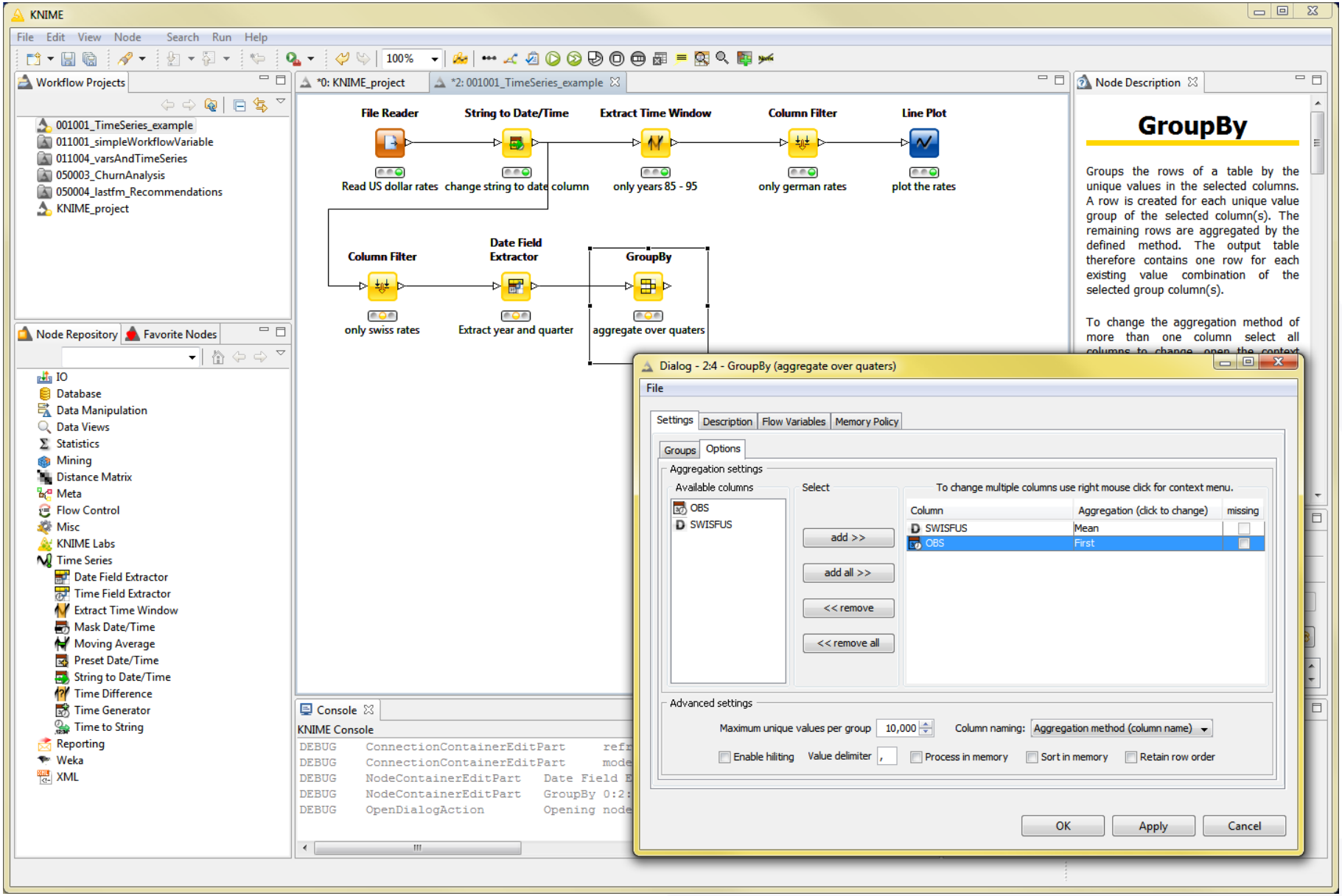
Example Workflow (step by step)



Example Workflow (step by step)



Example Workflow (step by step)



Example Workflow (step by step)

KNIME

File Edit View Node Search Run Help

Workflow Projects

- 001001_TimeSeries_example
- 011001_simpleWorkflowVariable
- 011004_varsAndTimeSeries
- 050003_ChurnAnalysis
- 050004_lastfm_Recommendations
- KNIME_project

*0: KNIME_project | *2: 001001_TimeSeries_example

File Reader → String to Date/Time → Extract Time Window → Column Filter → Line Plot

Read US dollar rates | change string to date column | only years 85 - 95 | only german rates | plot the rates

Column Filter → Date Field Extractor → GroupBy

only swiss rates | Extract year and quarter | aggregate over quaters

Node Repository | Favorite Nodes

- IO
- Database
- Data Manipulation
- Data Views
- Statistics
- Mining
- Distance Matrix
- Meta
- Flow Control
- Misc
- KNIME Labs
- Time Series
  - Date Field Extractor
  - Time Field Extractor
  - Extract Time Window
  - Mask Date/Time
  - Moving Average
  - Preset Date/Time
  - String to Date/Time
  - Time Difference
  - Time Generator
  - Time to String
- Reporting
- Weka
- XML

Console

KNIME Console

```
DEBUG    KnimeResourceNavigator  state changed to EXECUTED
DEBUG    NodeContainer   001001_TimeSeries_example 0:2 has new s
DEBUG    NodeContainer   001001_TimeSeries_example 0:2 has new s
DEBUG    NodeContainer   Workflow Manager 0 has new state: IDLE
DEBUG    OpenPortViewAction      Open Port View GroupBy (#1)
```

Node Description

GroupBy

Groups the rows of a table by the unique values in the selected columns. A row is created for each unique value group of the selected column(s). The remaining rows are aggregated by the defined method. The output table therefore contains one row for each existing value combination of the selected group column(s).

To change the aggregation method of

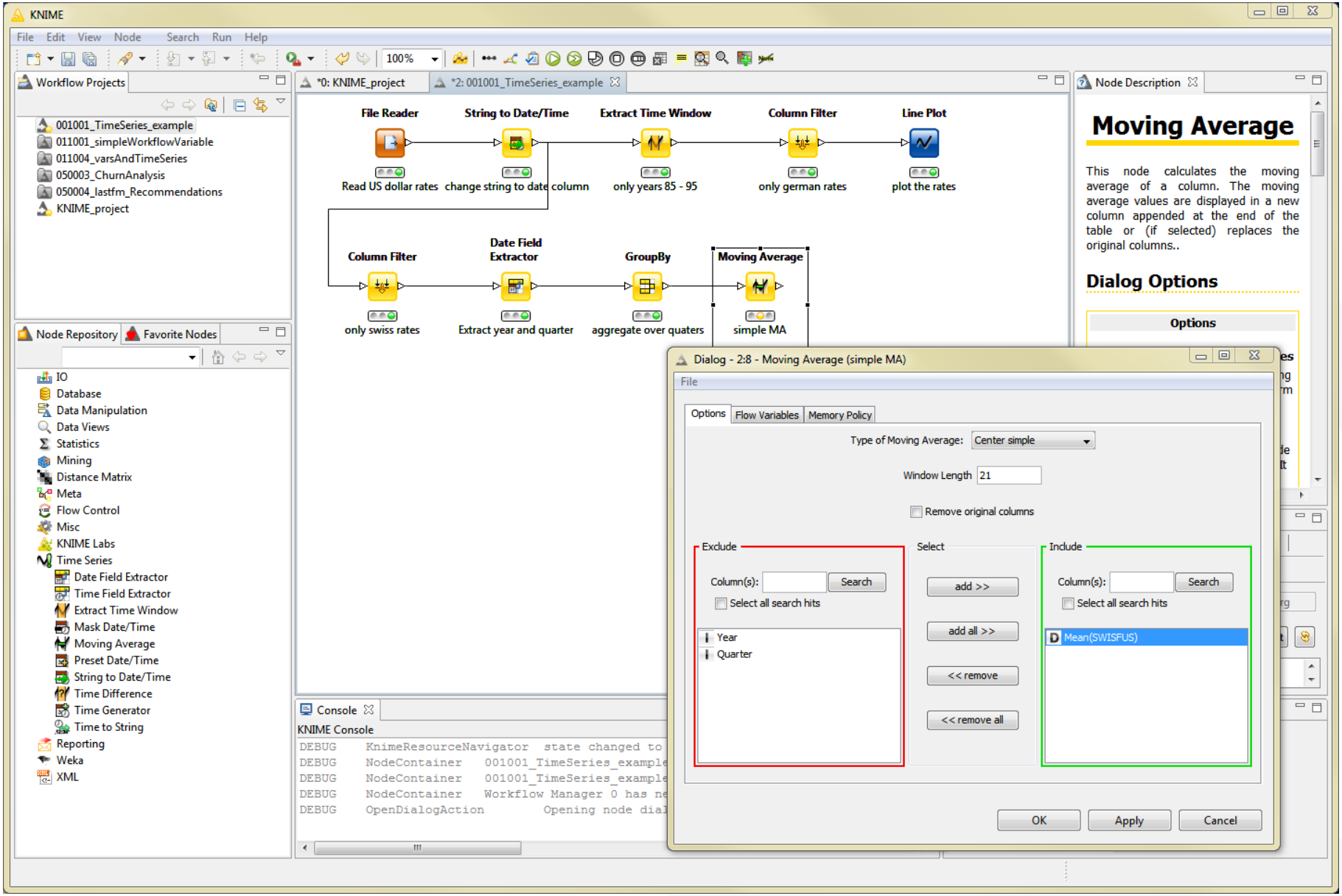
Group table - 2:4 - GroupBy (aggregate over quaters)

File

Table "default" - Rows: 77 | Spec - Columns: 4 | Properties | Flow Variables

Row ID	Year	Quarter	Mean(SWISFUS)	First(OBS)
Row23	1985	3	2.346	01.Jul.1985
Row24	1985	4	2.135	01.Oct.1985
Row25	1986	1	1.979	01.Jan.1986
Row26	1986	2	1.866	01.Apr.1986
Row27	1986	3	1.688	01.Jul.1986
Row28	1986	4	1.663	01.Oct.1986
Row29	1987	1	1.547	01.Jan.1987
Row30	1987	2	1.492	01.Apr.1987
Row31	1987	3	1.526	01.Jul.1987
Row32	1987	4	1.402	01.Oct.1987
Row33	1988	1	1.376	01.Jan.1988
Row34	1988	2	1.419	01.Apr.1988
Row35	1988	3	1.566	01.Jul.1988
Row36	1988	4	1.495	03.Oct.1988
Row37	1989	1	1.584	02.Jan.1989
Row38	1989	2	1.696	03.Apr.1989
Row39	1989	3	1.658	03.Jul.1989
Row40	1989	4	1.607	02.Oct.1989
Row41	1990	1	1.507	01.Jan.1990
Row42	1990	2	1.443	02.Apr.1990
Row43	1990	3	1.336	02.Jul.1990
Row44	1990	4	1.274	01.Oct.1990
Row45	1991	1	1.312	01.Jan.1991
Row46	1991	2	1.474	01.Apr.1991
Row47	1991	3	1.517	01.Jul.1991
Row48	1991	4	1.433	01.Oct.1991
Row49	1992	1	1.457	01.Jan.1992

Example Workflow (step by step)



Example Workflow (step by step)

KNIME

File Edit View Node Search Run Help

Workflow Projects

- 001001_TimeSeries_example
- 011001_simpleWorkflowVariable
- 011004_varsAndTimeSeries
- 050003_ChurnAnalysis
- 050004_lastfm_Recommendations
- KNIME_project

*0: KNIME_project | *2: 001001_TimeSeries_example

File Reader → String to Date/Time → Extract Time Window → Column Filter → Line Plot

Read US dollar rates | change string to date column | only years 85 - 95 | only german rates | plot the rates

Column Filter → Date Field Extractor → GroupBy → Moving Average → Math Formula → Line Plot

only swiss rates | Extract year and quarter | aggregate over quaters | simple MA | diff betw. mean and MA

Node Description

Math Formula

This node evaluates a (free-form) mathematical expression based on the values in a row. The computed results can be either appended as new column or be used to replace an input column. Available variables are the values in the corresponding row of the table (left list in the dialog). Commonly used functions are shown in the list "Mathematical Functions". There are also some constants available, such as *pi* (the ratio of the circumference of a circle to its diameter), *e* (the base of the natural logarithms), the total number of rows in the table, and some other column based constants (right list).

This node uses *JEP*, the Java Math Expression Parser.

Dialog Options

Node Repository | Favorite Nodes

Dialog - 2:15 - Math Formula (diff betw. mean and MA)

File

Math Expression | Flow Variables | Memory Policy

Column List: Year, Quarter, Mean(SWISFUS), MA(Mean(SWISFUS))

Mathematical Function: ln(x), log(x), exp(x), abs(x), sqrt(x), rand(), mod(x, y), if(x, y, z), round(x), round(x, y), ceil(x)

Constants: ROWCOUNT, ROWINDEX, pi, e, COL_MIN(col_name), COL_MAX(col_name), COL_MEAN(col_name), COL_STDDEV(col_name), COL_VAR(col_name)

Expression: abs($MA(Mean(SWISFUS))$-$Mean(SWISFUS)$)

Append Column: trend

Replace Column: MA(Mean(SWISFUS))

OK | Apply | Cancel

KNIME Example Flow Server

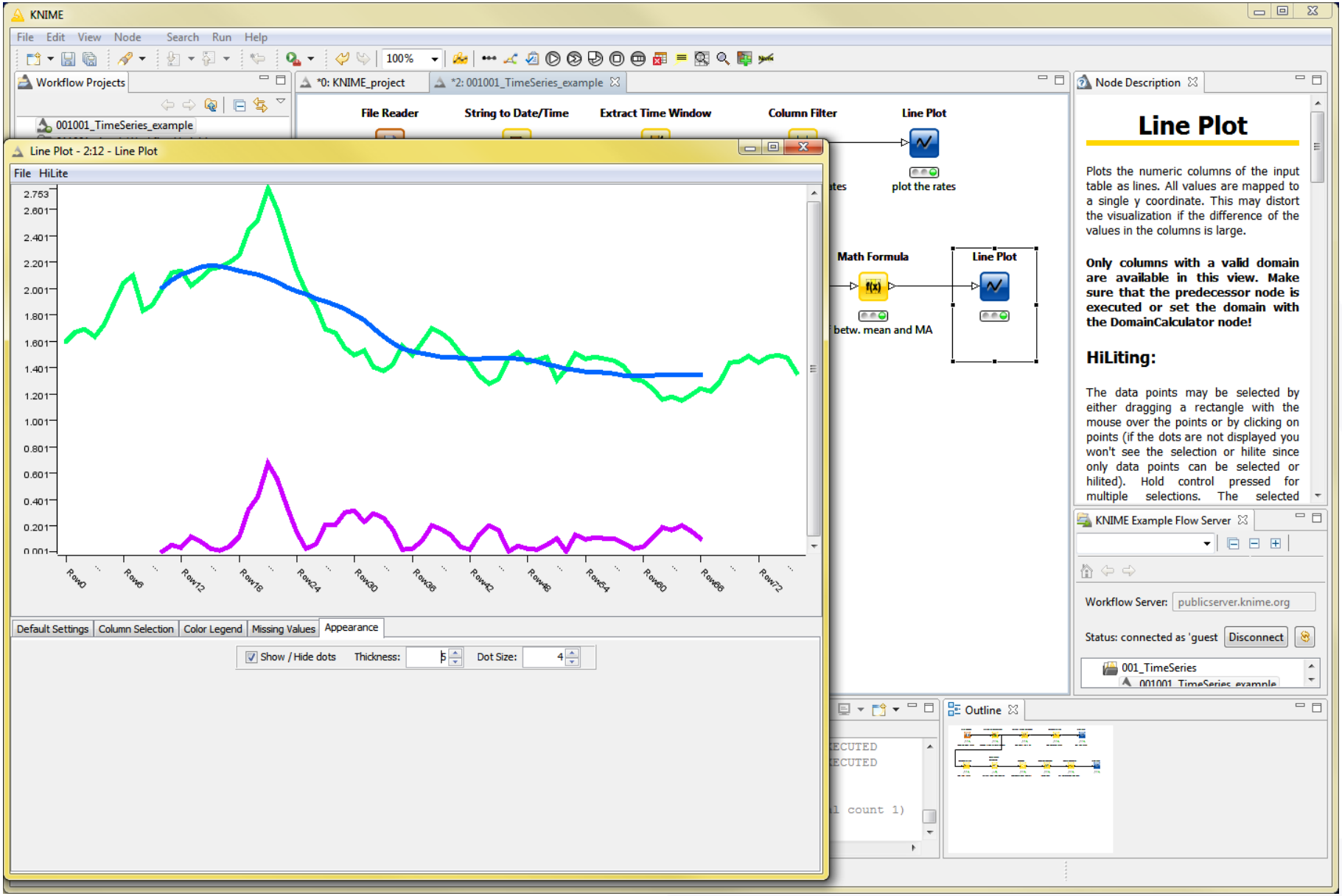
Workflow Server: publicserver.knime.org

Status: connected as 'guest' Disconnect

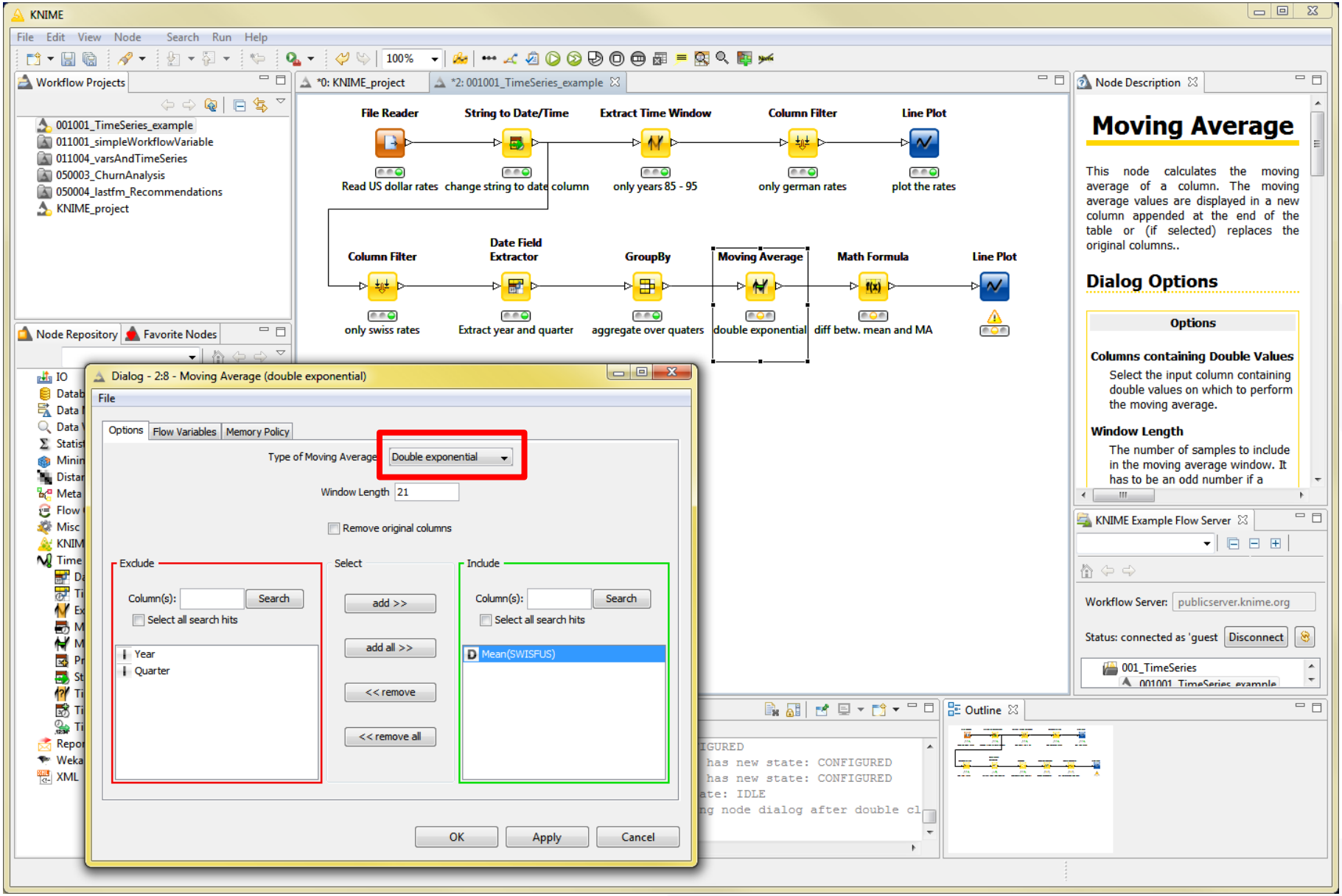
001_TimeSeries

Outline

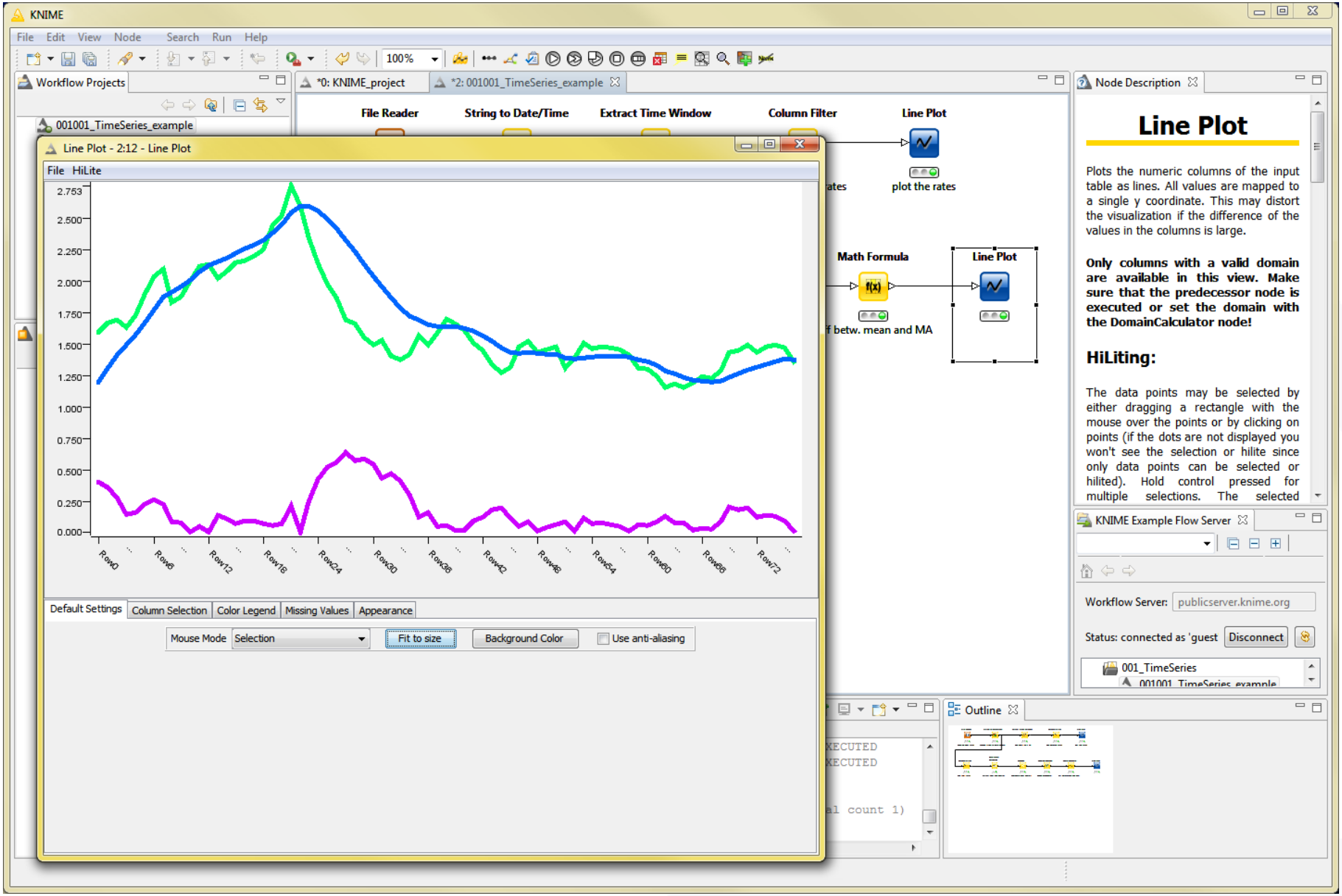
Example Workflow (step by step)



Example Workflow (step by step)

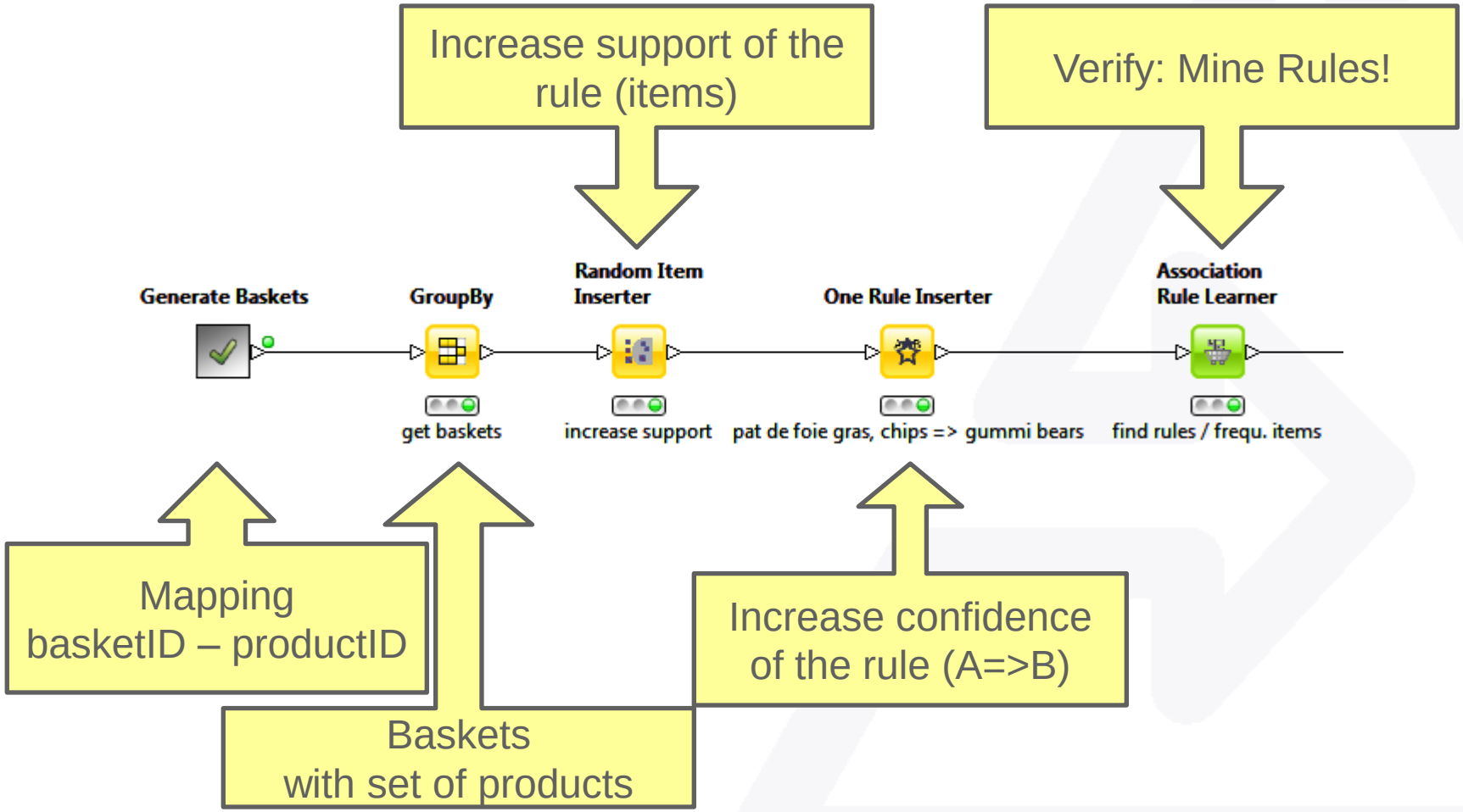


Example Workflow (step by step)

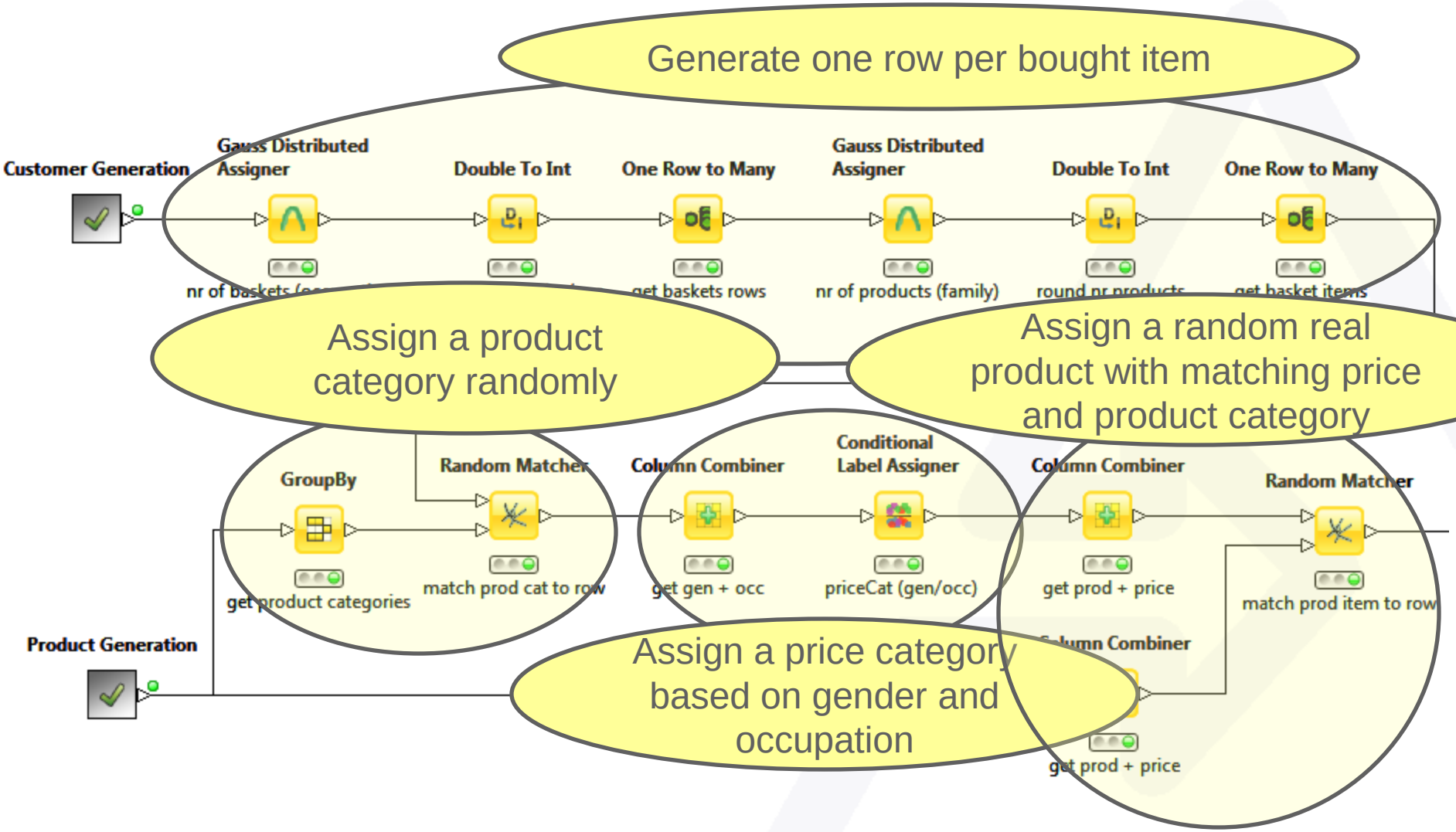


Association Rules

- Using the Modular Data Generation from KNIME Labs

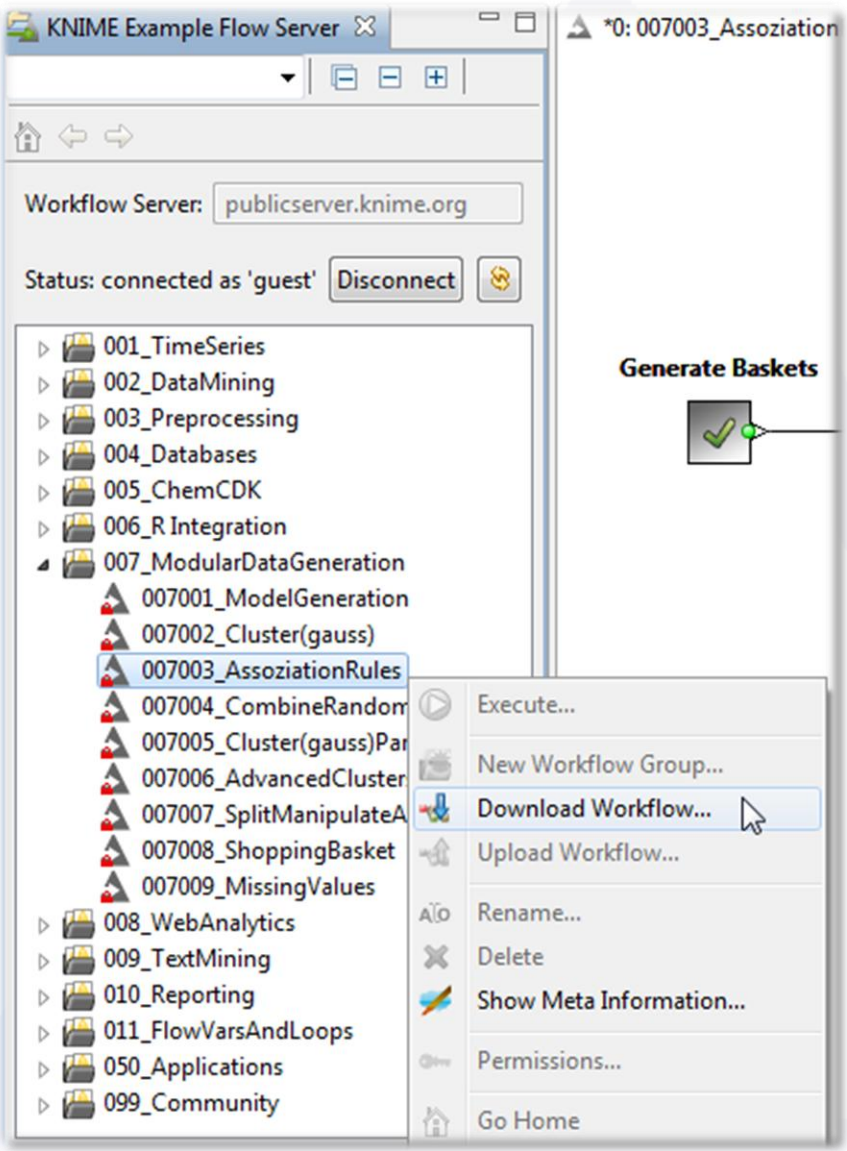


Generate Customer Data



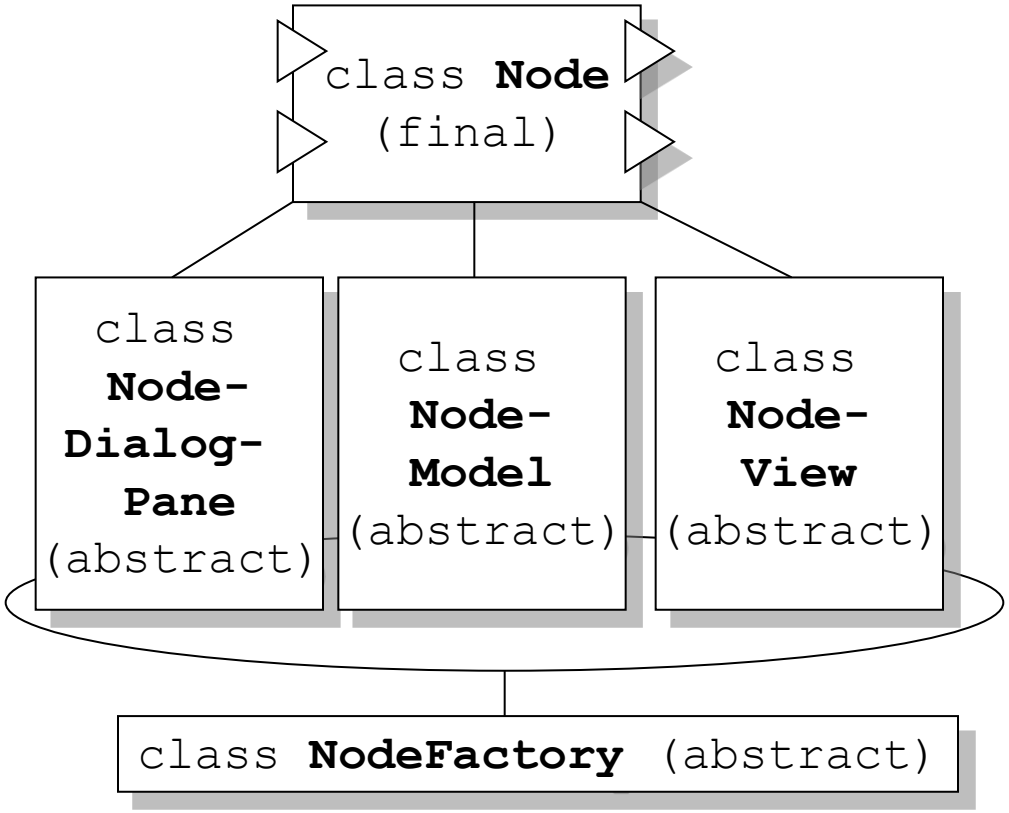
Download Example Workflows

- KNIME Example Flow Server
- Available via the view menu in KNIME
- Public workflow repository
- Demonstrated workflows:
  - 001001_TimeSeries_example
  - 007003_AssoziationRules
  - 007008_ShoppingBasket



DEVELOPMENT

Node Architecture

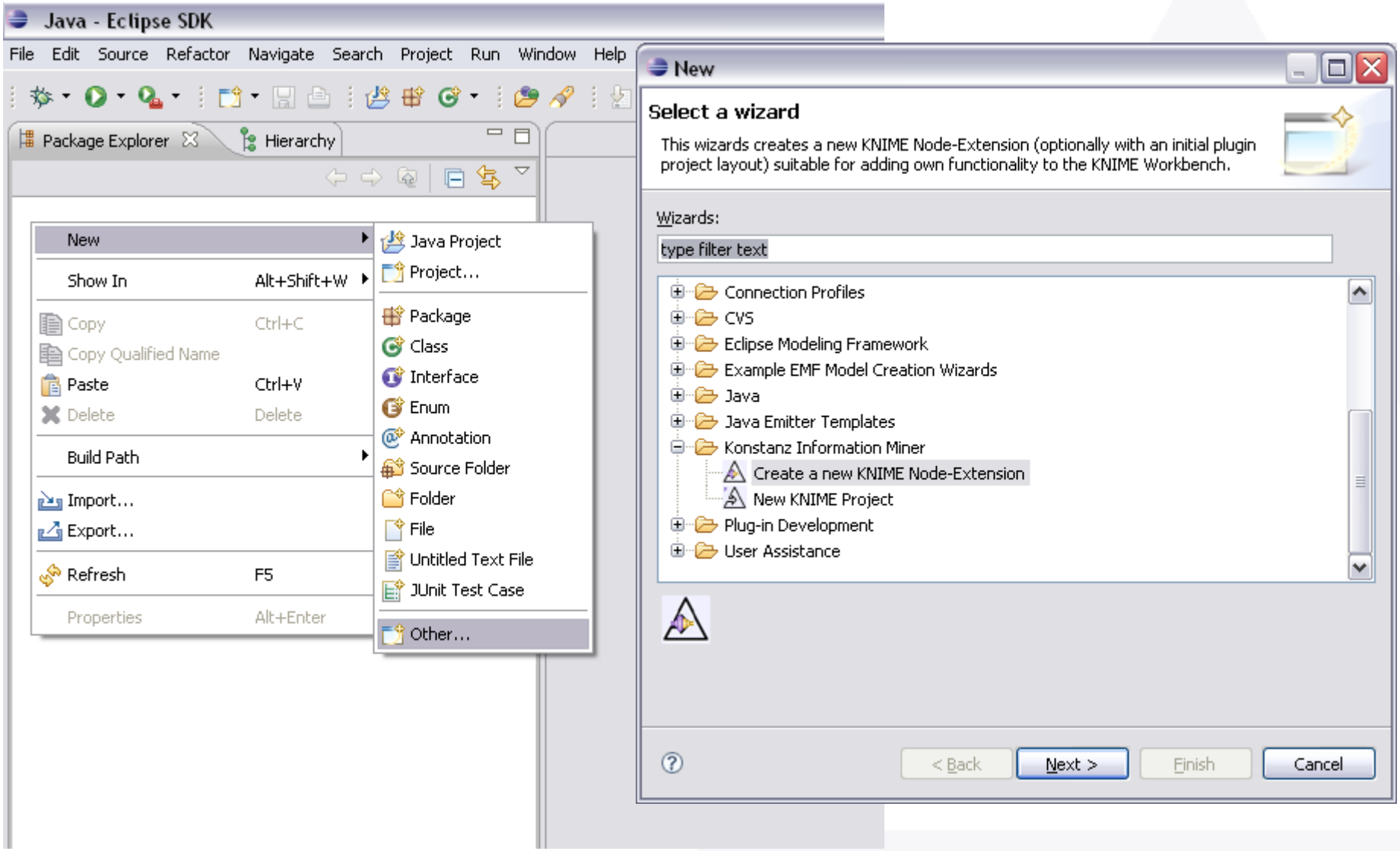


- KNIME interacts only with a **Node**
- **Node** takes care of embedding the node in the infrastructure
- New nodes implement Model/View/Dialog

Node Extension Wizard

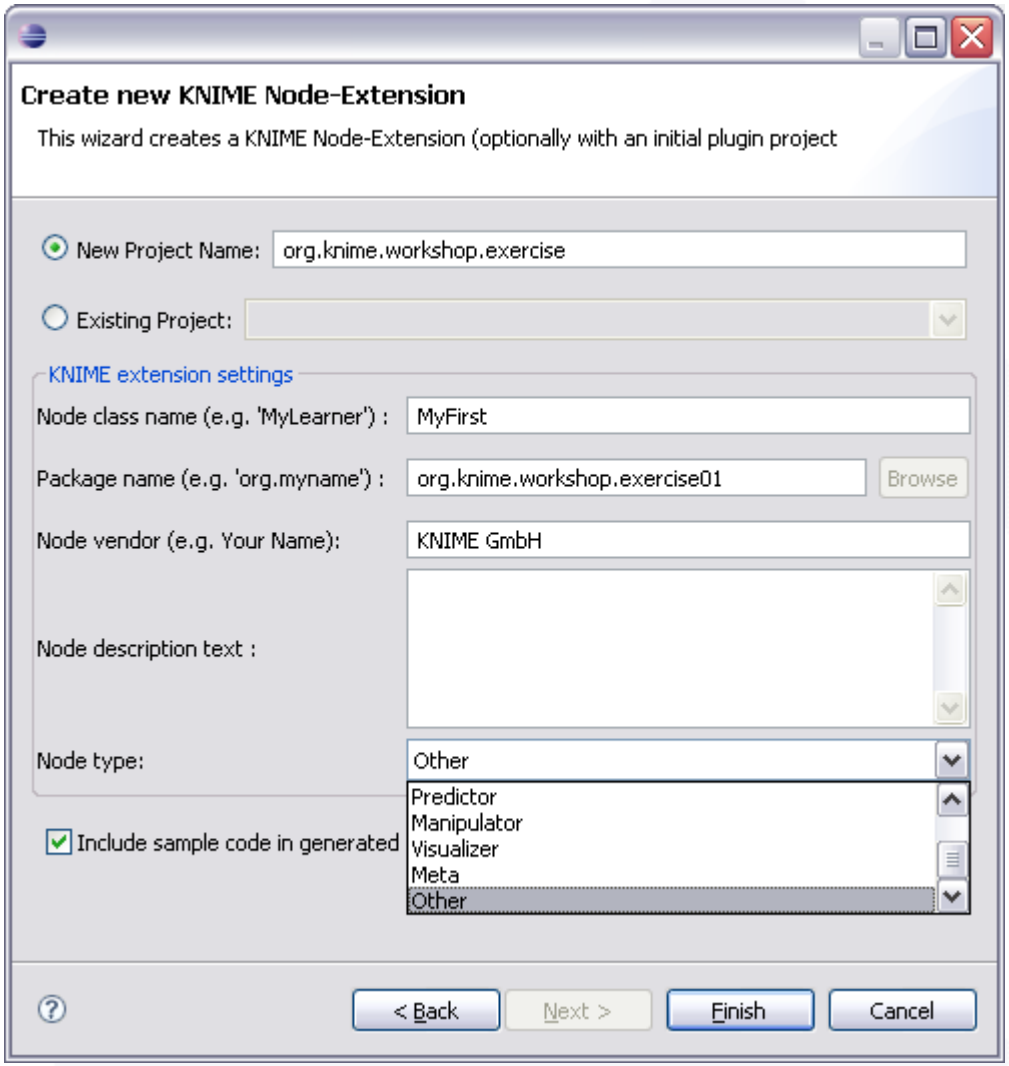
- Included in the KNIME Developer Version
- Allows creation of plugin projects including functioning KNIME nodes (with sample code)
- Helpful to easily create all node classes
  - Generates all Java classes
  - Node is registered with the plugin project
  - Launch KNIME and enjoy the new node working!

Node Extension Wizard



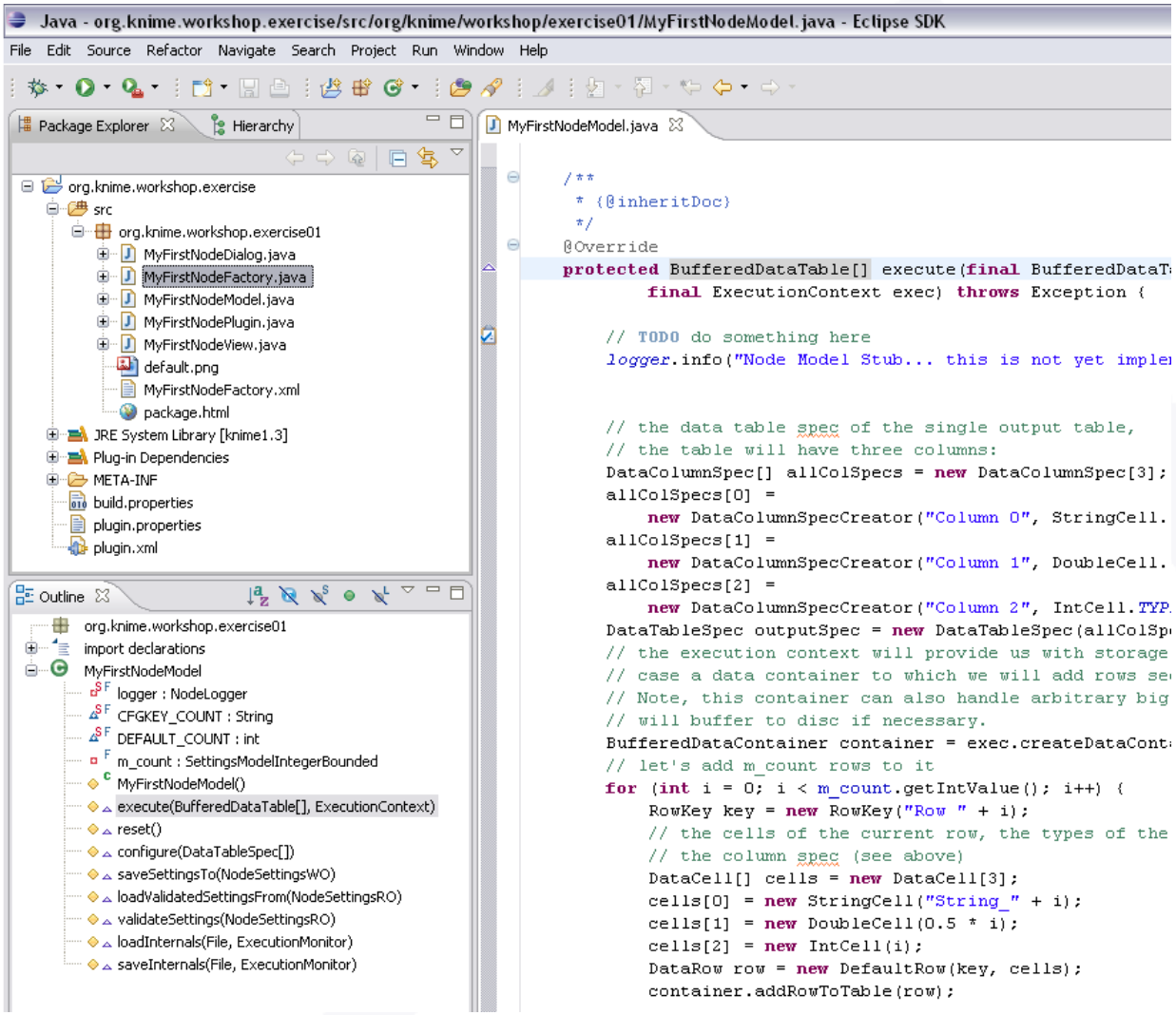
Node Extension Wizard

- Specify all settings to create a new KNIME node
  - In a completely new plugin project, or
  - Into an existing project
- Node type: Sink, Source, Learner, Predictor, Manipulator, Visualizer, Meta, or Other
- Include sample code or not



Node Extension Wizard

- Contains all Java classes (including sample code)
- Node is registered in the **`plugin.xml`**
- **`NodeDialog`** and **`NodeView`** class are also created and registered to the **`NodeFactory`**



Summary

- Nodes encapsulate operations
- Nodes provide intermediate results
- Self explaining workflows
- Handles large data sets
- Easy extensible