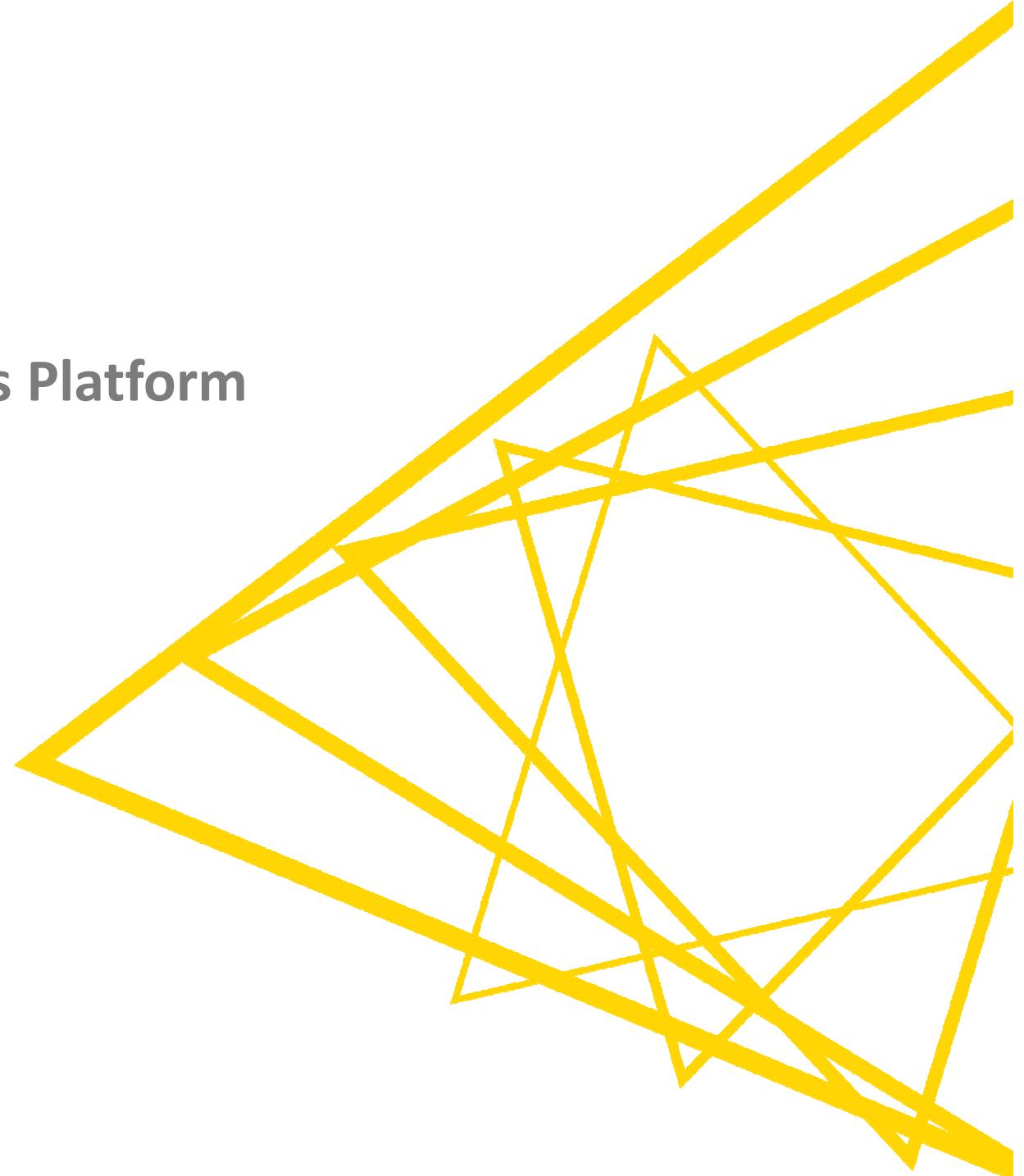




# The Open Analytics Platform

Bernd Wiswedel  
KNIME.com AG



# Agenda

---

- KNIME.com AG
- The KNIME Platform
  - Recognition
  - Small Sales Pitch
- KNIME and R – the best of two worlds
- KNIME (Node) Development

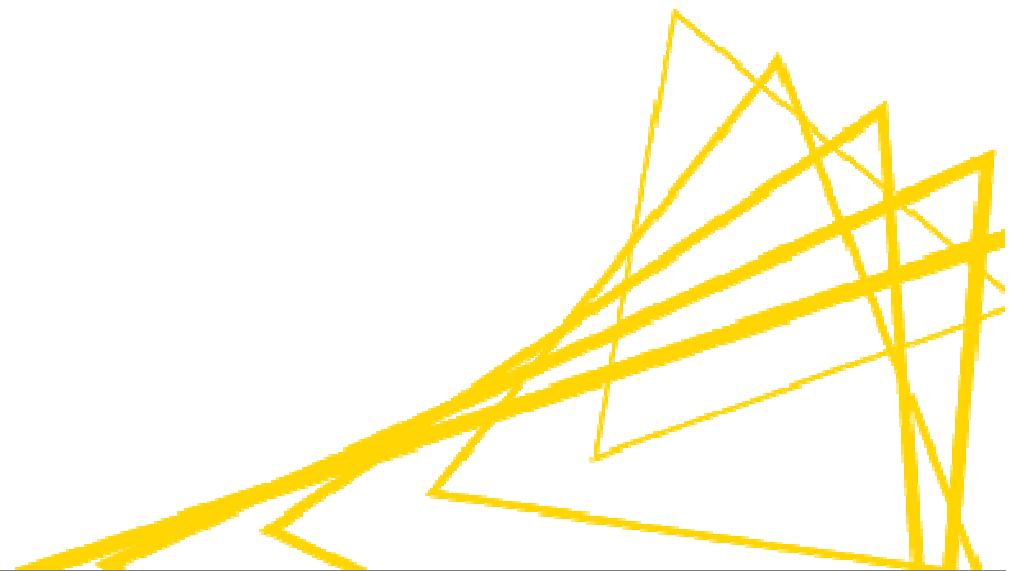
# A Brief History of KNIME

---

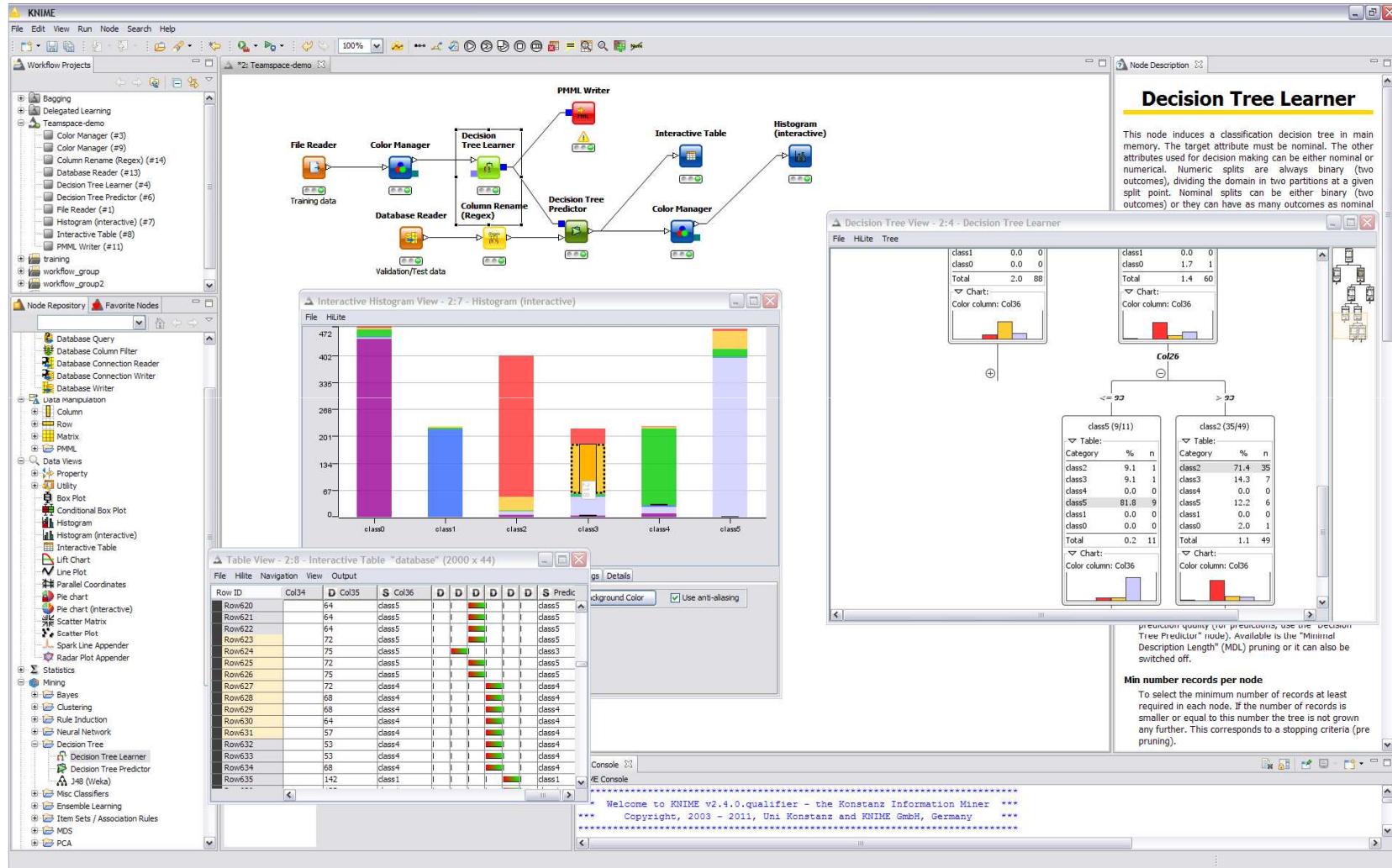
- 2004: KNIME development commences
- 2006: KNIME v1 released
- 2006: Spin-off in Konstanz, Germany
- 2008: KNIME moves to Zurich
- 2010: Enterprise products released
- 2011: KNIME.com AG founded
- 2013: KNIME opens San Francisco office
- 2014: KNIME opens Berlin office

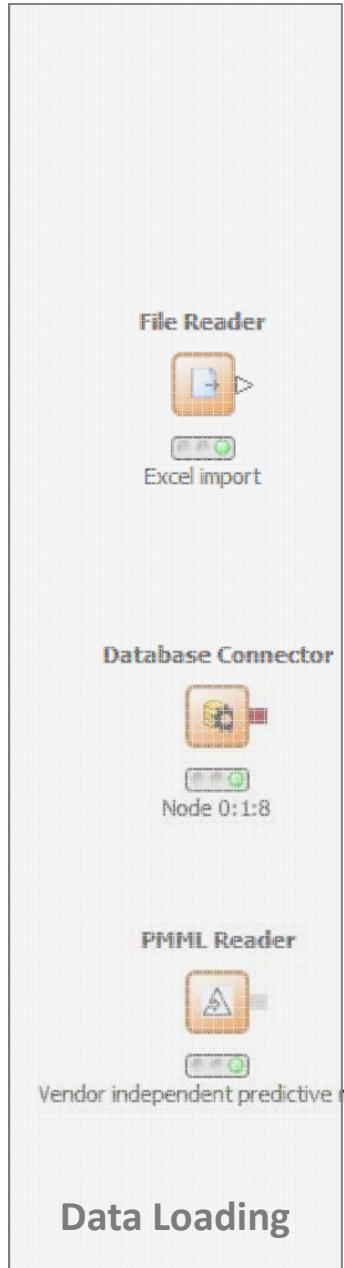
„KNIME saved my  
life in a world of scripts  
that I do not want to learn!

# The KNIME Platform



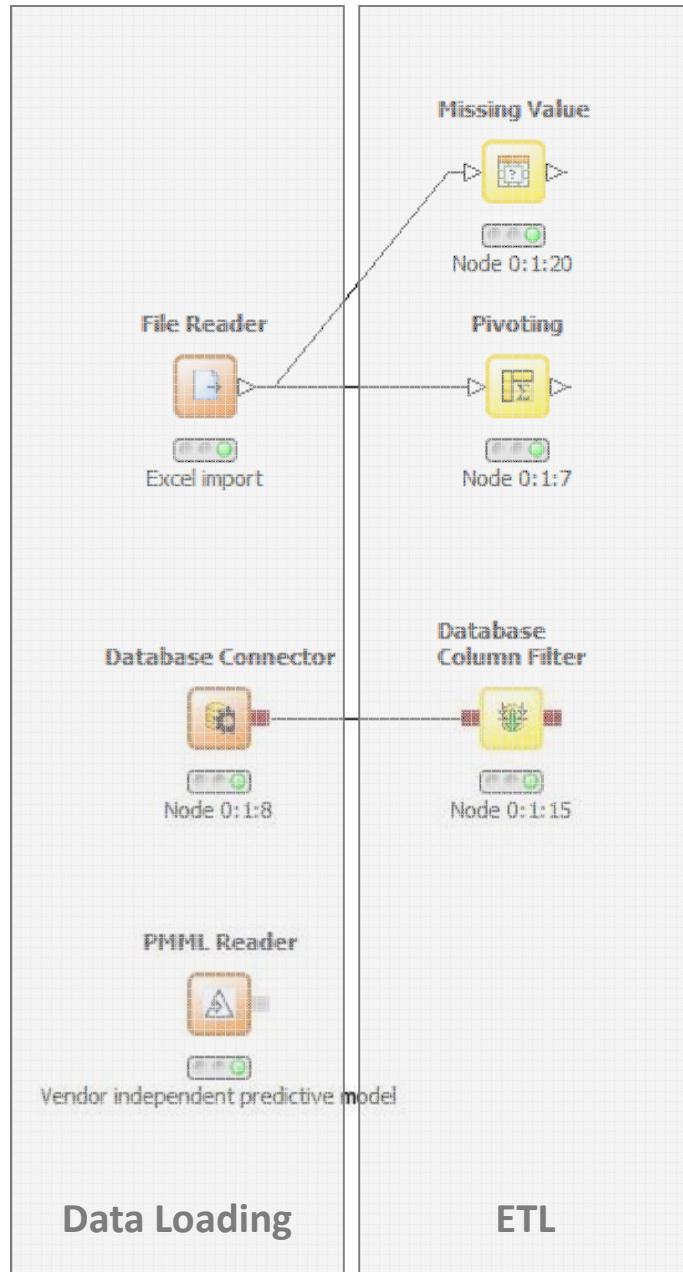
# The KNIME Platform





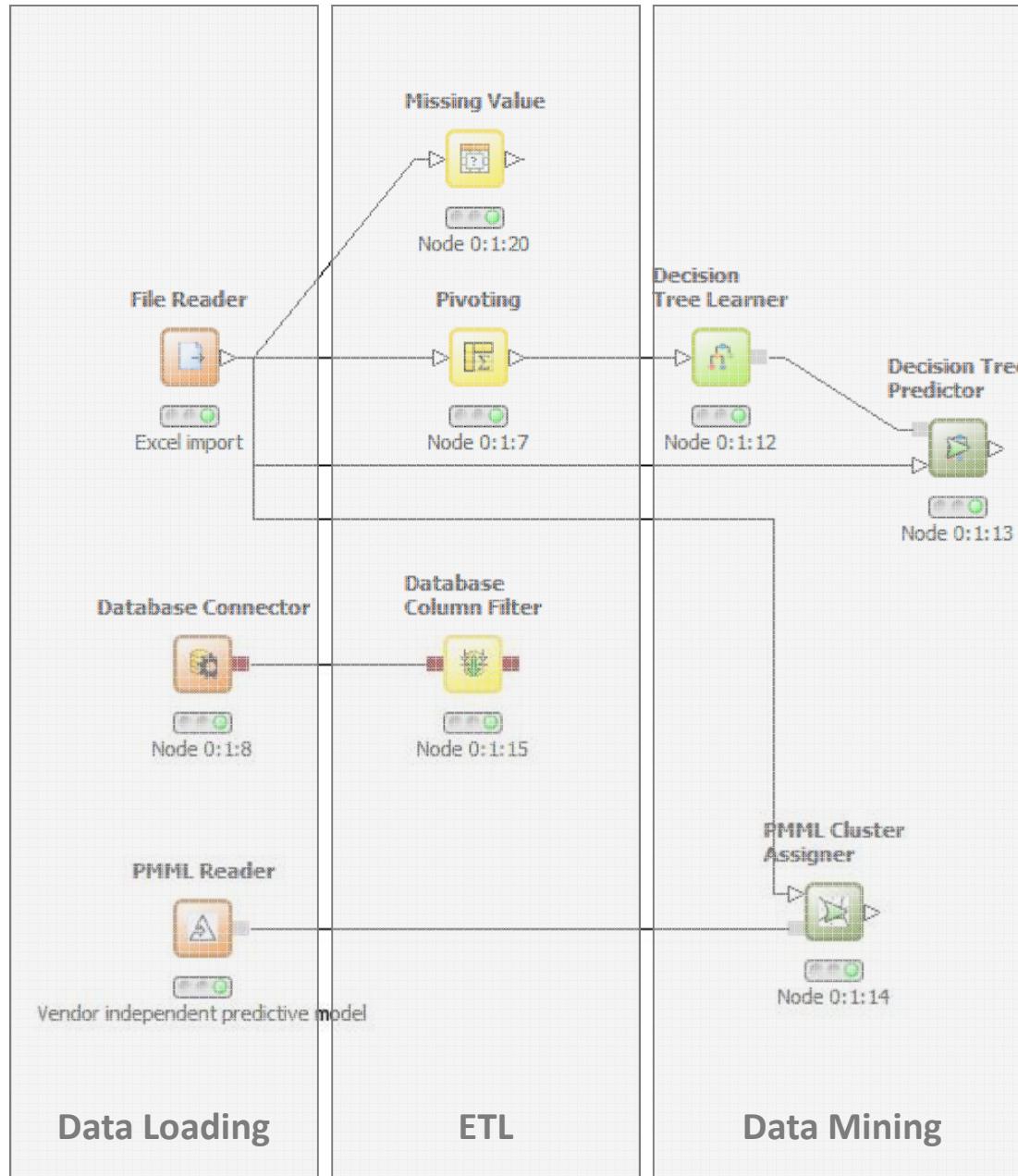
KNIME loads and integrates data from diverse data sources:

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

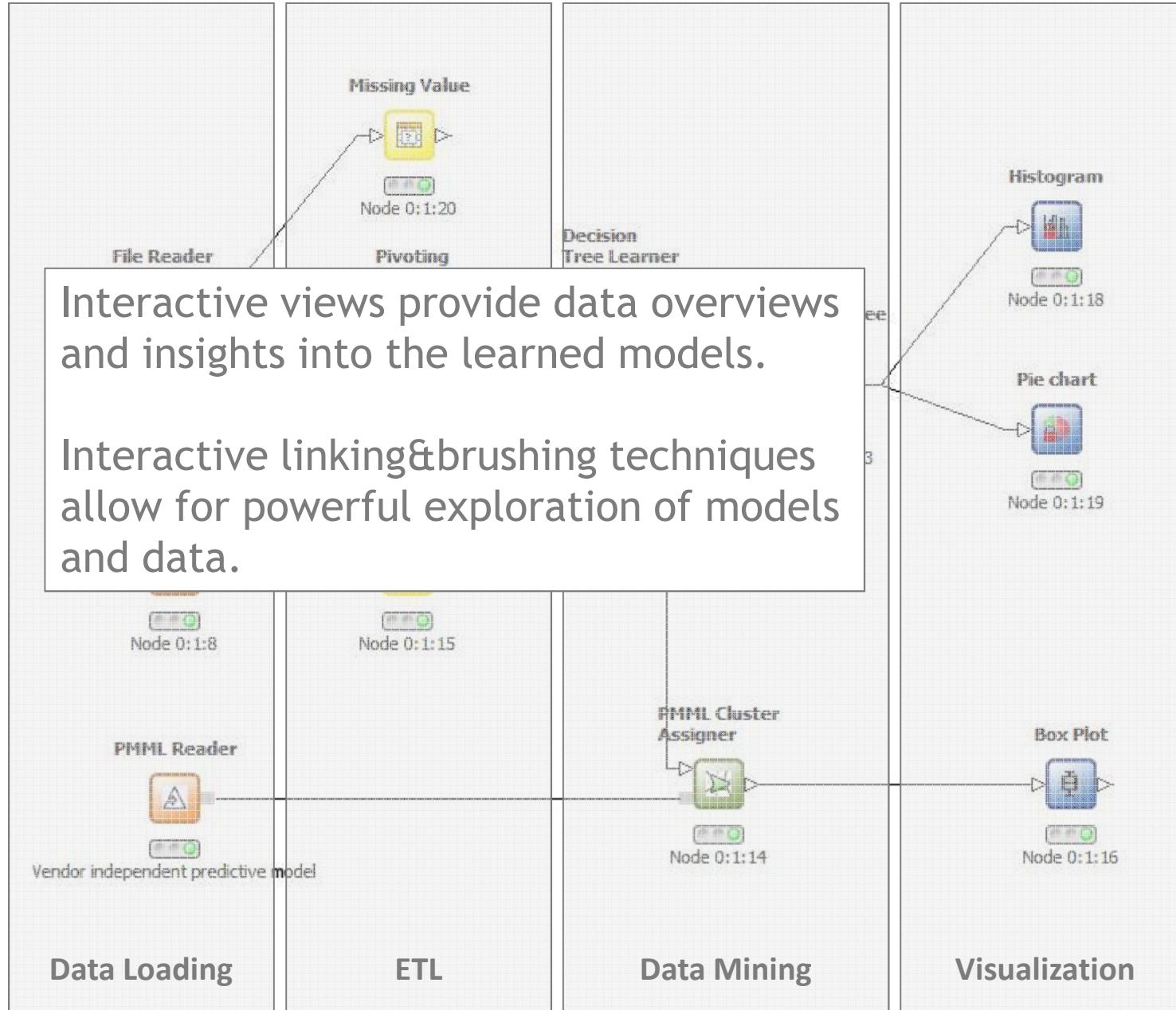


KNIME provides huge repository of modules for easy-to-use, modular

- Data preprocessing
- Data fusion
- Data transformation



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



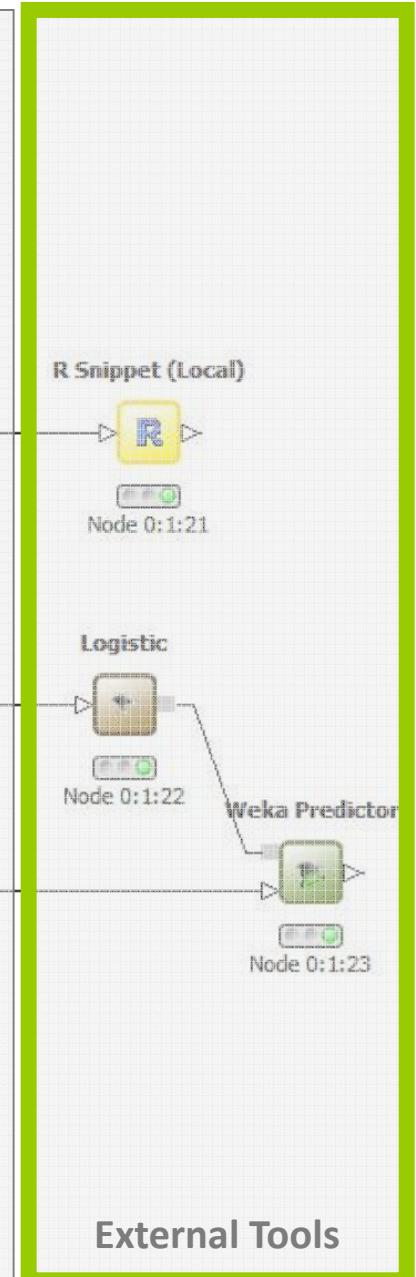
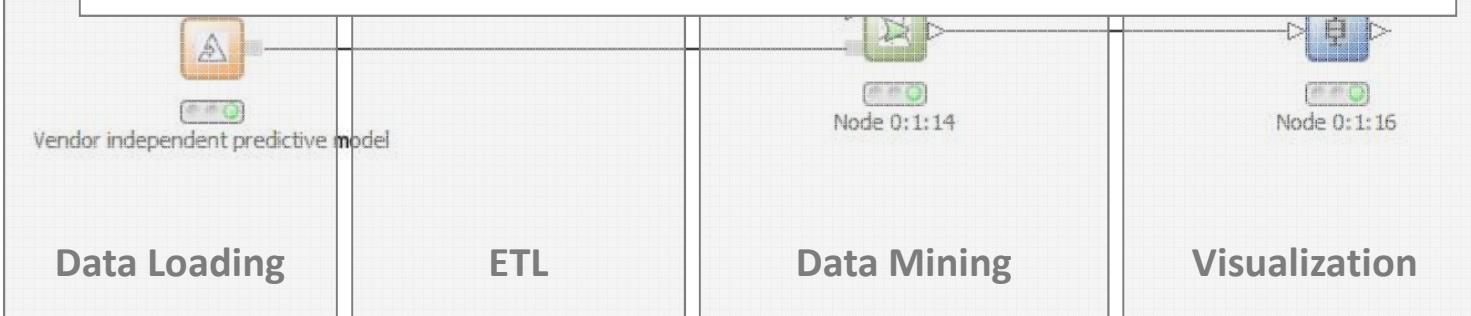
Due to its open API and “node-in-a-sandbox”-approach additional (also external) tools are easily integrated,

e.g.

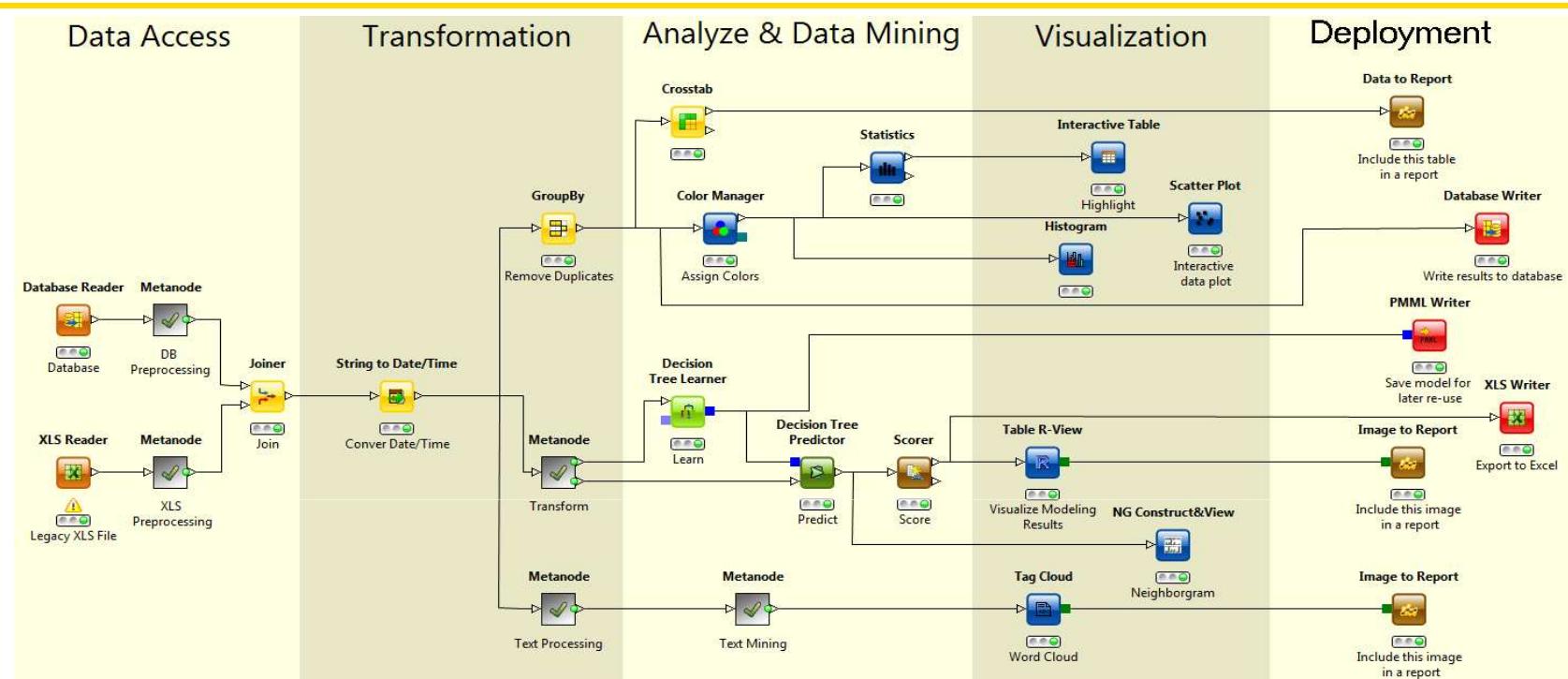
- Access to the R Project (statistical analysis/visualizations)
- Complete integration of the machine learning library WEKA
- Application area specific integration, e.g. CDK (Chemical Development Kit), RDKit, ImageJ, ...



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



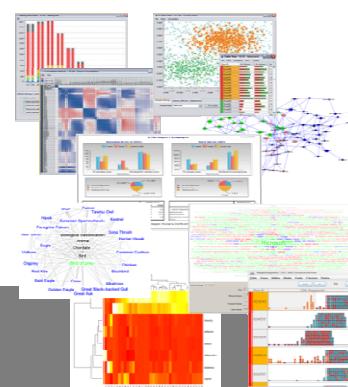
# Over 1000 native and embedded nodes included:



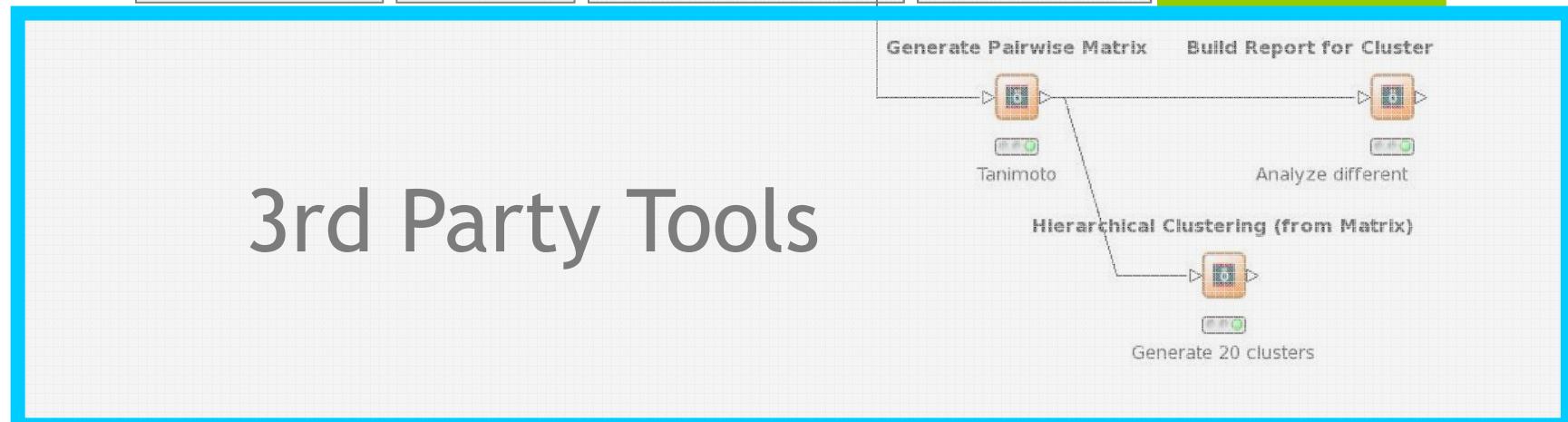
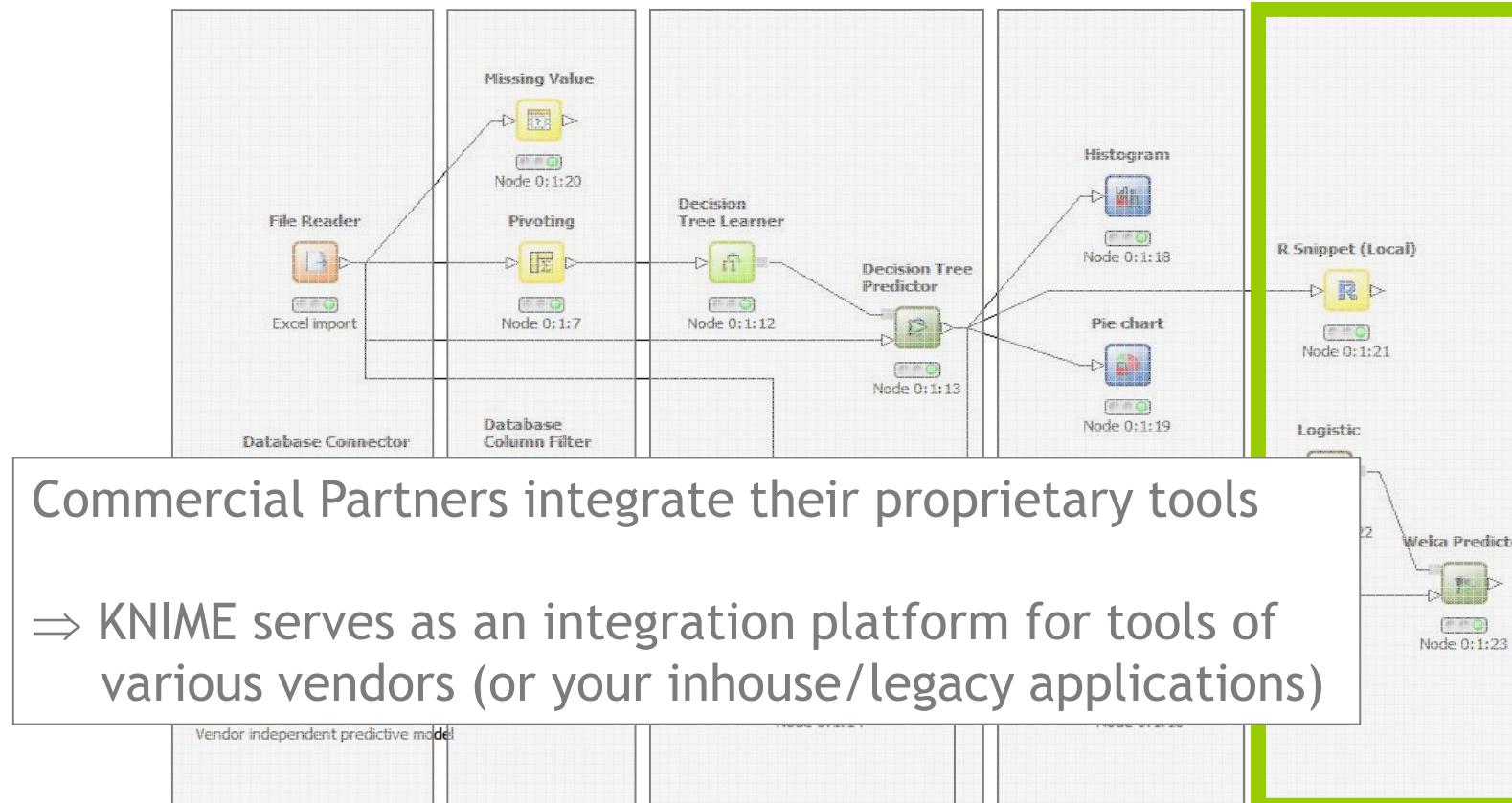
MySQL, Oracle, etc.  
SAS, SPSS, etc.  
Excel, Flat, etc.  
Hive etc.  
XML, PMML  
Text, Doc, Image  
Web Crawlers  
Industry Specific  
Community / 3rd

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

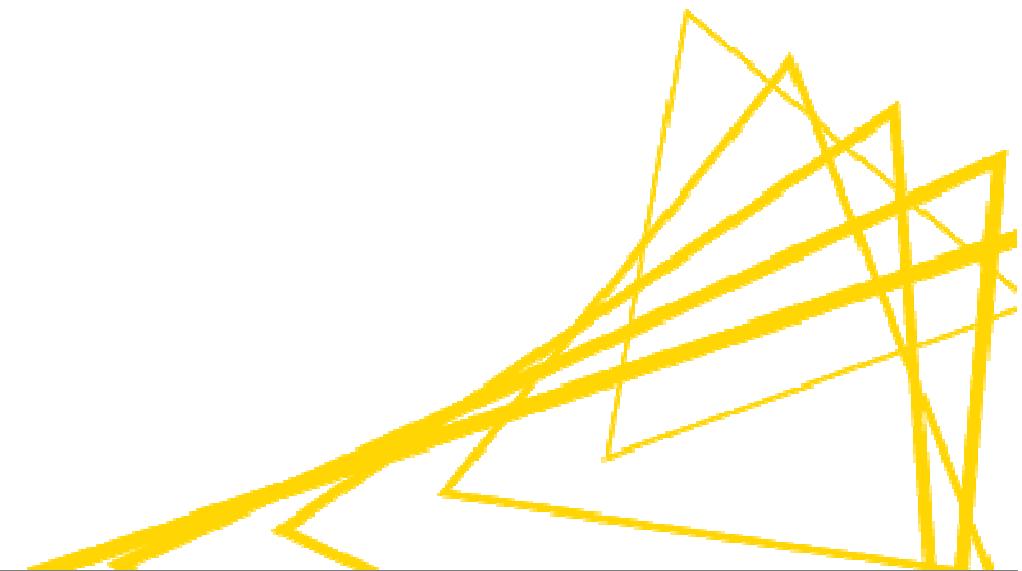
Statistics  
Data Mining  
Machine Learning  
Web Analytics  
Text Mining  
Network Analysis  
Social Media Analysis  
WEKA  
R  
Community / 3rd



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



# Small KNIME Demo



# Who's Using KNIME?

---

>25,000

**Individuals using KNIME**

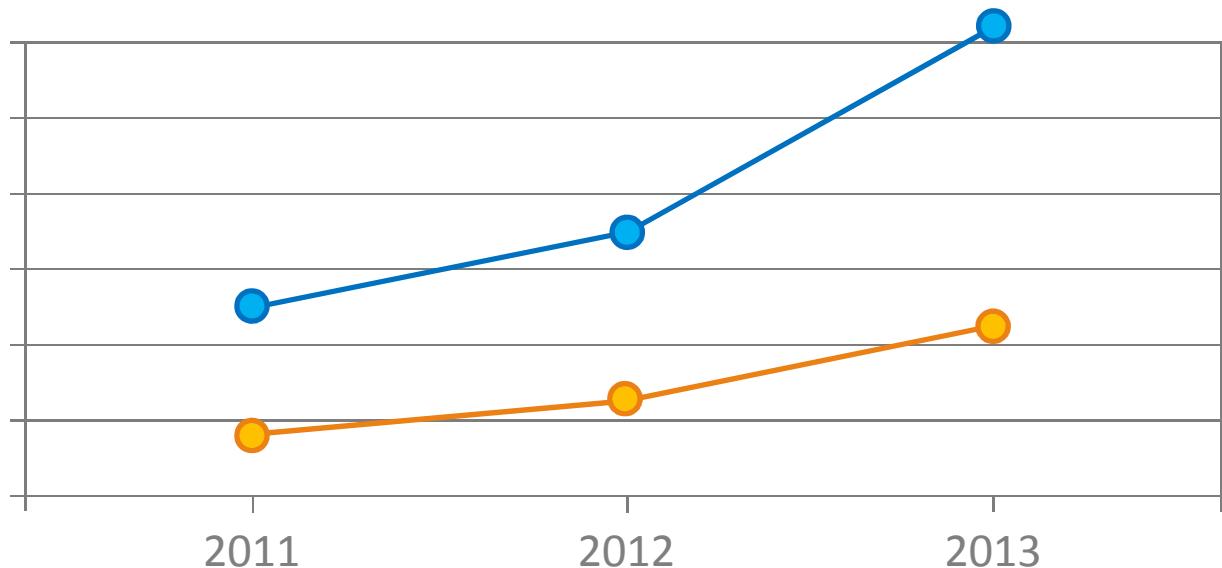
>3,000

**Organizations using KNIME**

>300

**Customers paying for KNIME**

as of January 2014



# Broad Range of KNIME Application Areas

---



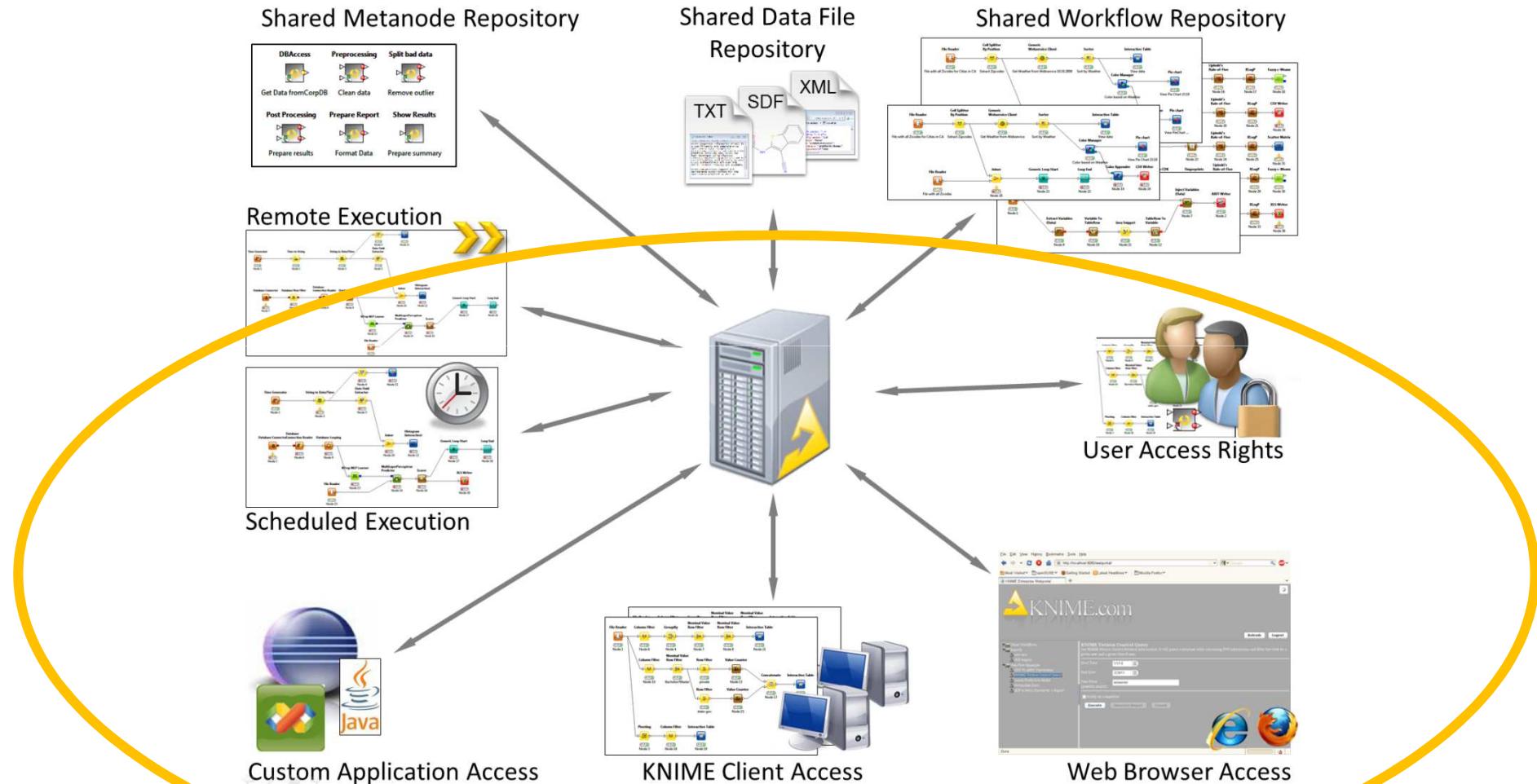
# Top in User Satisfaction

---

	Overall	IBM SPSS Statistics	IBM SPSS Modeler	KNIME	R	Rapid Miner	SAS	SAS Enterprise Miner	STATISTICA (StatSoft)	Weka
Software or platform Type:		Proprietary	Proprietary	Open Source	Open Source	Open Source	Proprietary	Proprietary	Proprietary	Open Source
Quality and accuracy of model performance	4,34	3,93	4,32	4,32	4,43	4,37	4,34	4,26	4,69	4,21
Dependability/Stability of software	4,25	4,19	4,05	4,43	4,34	4,22	4,32	4,44	4,56	3,73
Variety of available algorithms	4,20	3,69	4,30	4,48	4,72	4,54	4,01	4,00	4,63	4,33
Ease of use	4,19	4,28	4,60	4,76	3,58	4,47	3,69	4,00	4,49	4,06
Ability to automate repetitive tasks	4,17	3,75	3,96	4,39	4,39	4,40	4,31	4,00	4,45	3,71
Data manipulation capabilities	4,15	4,00	4,32	4,53	4,10	4,27	4,45	3,82	4,41	3,52
Quality of output / Ease of interpretation	4,10	3,91	4,04	4,39	4,04	4,36	3,69	4,00	4,53	3,66
Good metrics of model quality	4,10	3,85	3,96	4,05	4,13	4,28	4,08	4,18	4,50	3,85
Good variable discovery, profiling and selection	4,03	3,70	4,06	4,17	3,98	4,33	3,81	4,35	4,44	3,69
Quality of user interface	4,03	4,11	4,53	4,62	3,36	4,45	3,58	3,91	4,49	3,59
Ease of model deployment (scoring models to other data sets)	4,03	3,61	4,13	4,43	3,82	4,20	3,90	4,21	4,46	3,77
Speed	4,02	3,84	4,13	4,12	3,58	3,90	4,08	3,97	4,48	3,53
Data quality assessment and data preparation capabilities	4,00	3,94	4,22	4,33	3,76	4,20	4,05	3,68	4,38	3,53
Ability to handle very large data sets	3,99	3,82	4,21	4,35	2,95	3,74	4,41	4,44	4,58	3,03
Ability to modify algorithm options to fine-tune analyses	3,95	3,17	3,59	3,98	4,33	4,23	3,97	3,97	4,33	3,88
Enables mining within one's database	3,94	3,54	4,26	4,12	3,75	4,10	3,92	4,00	4,19	3,61
Ability to easily incorporate data at different levels of granularity (e.g. transaction data and customer data)	3,90	3,56	4,06	4,24*	3,77	3,99	4,14	3,94	4,24*	3,29
Useful help menu, demos and tutorials	3,87	3,83	3,99	3,93	3,68	3,90	3,76	3,79	4,35	3,61
Strong graphical visualization of models	3,83	3,24	3,68	3,88	4,14	4,28	3,02	3,88	4,62	3,28
Cost of software	3,79	3,16	3,00	4,93	4,90	4,82	2,33	2,74	3,90	4,88
Overall	4,04	3,76	4,07	4,32	3,99	4,25	3,89	3,98	4,44	3,74

2012 & 2013 Rexter Analytics Survey

# Sales Pitch: The KNIME Server at Work

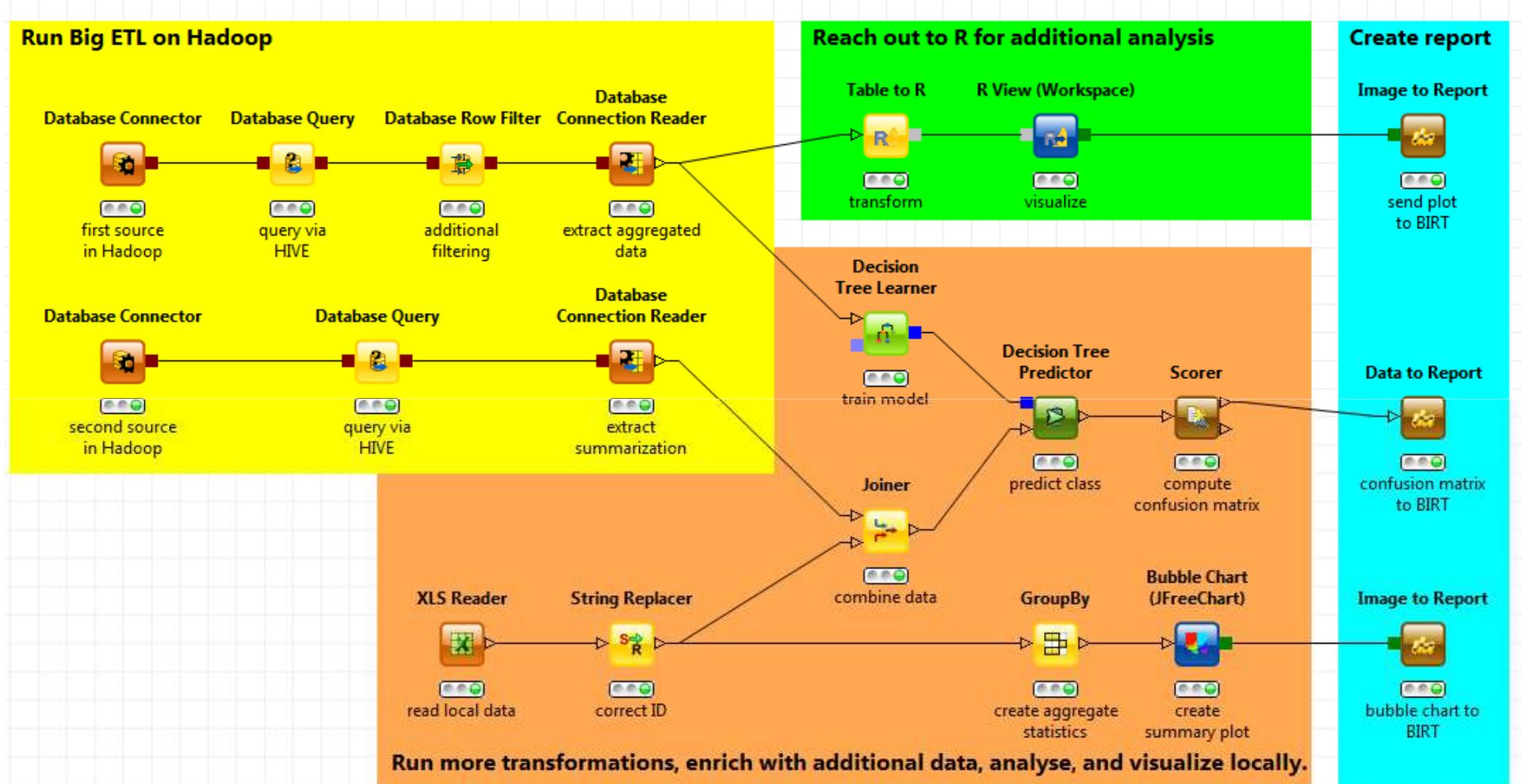


# KNIME in Action: Big Data

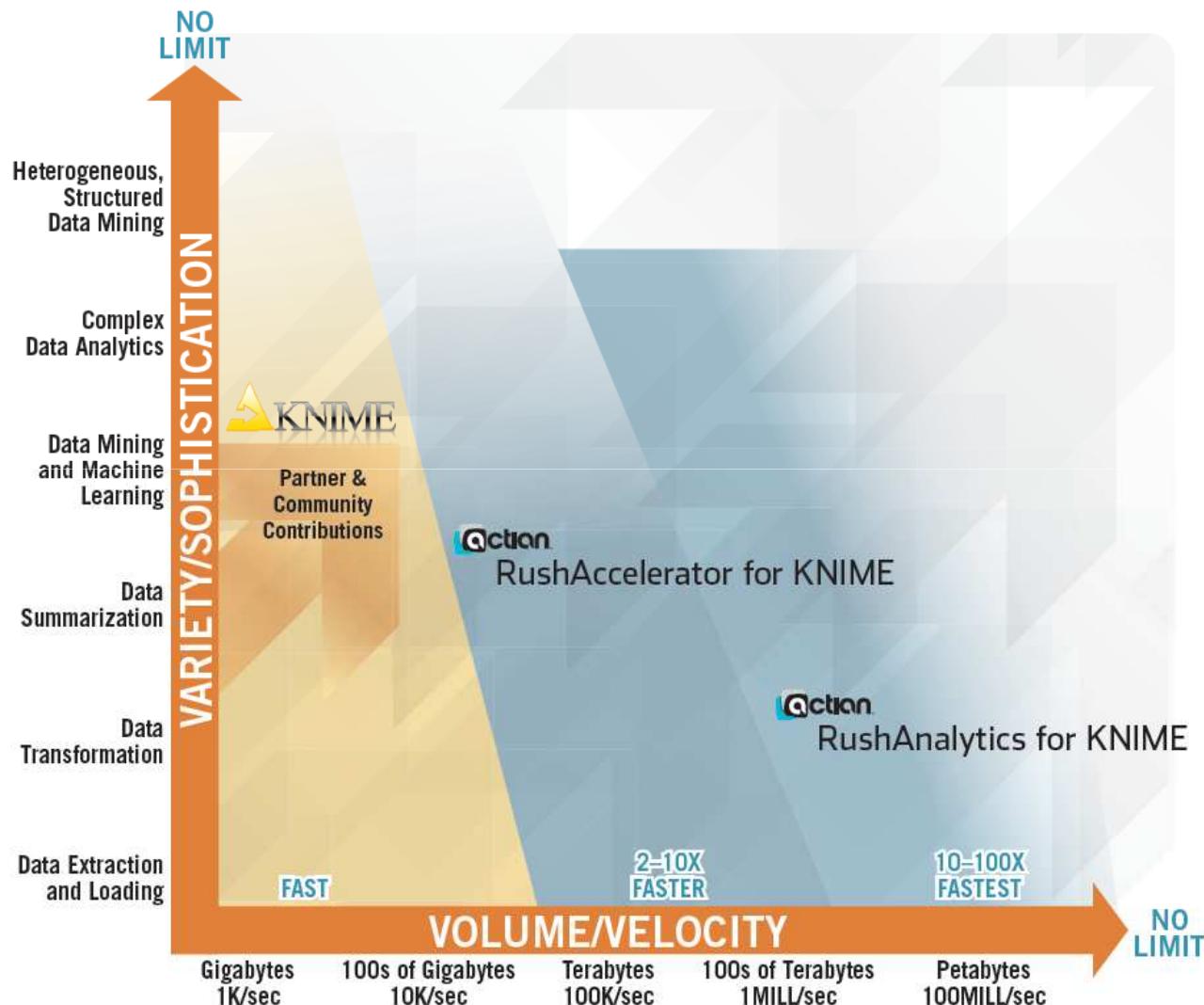
---

“As long as your machine can handle it, KNIME will play along.”

# KNIME and Big Data

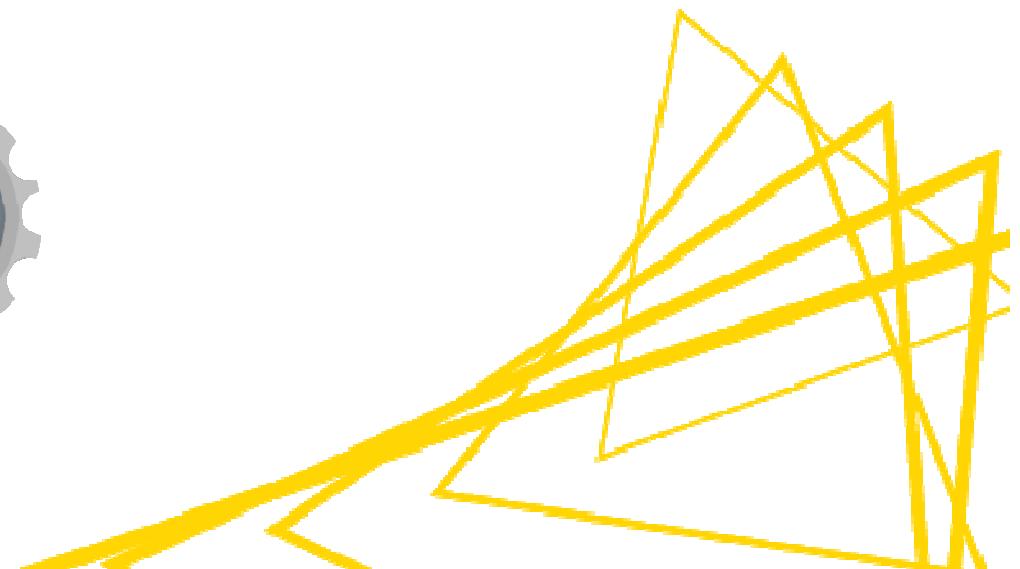
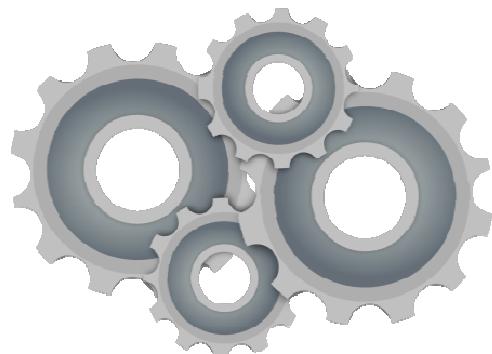


# KNIME and Big Data



# KNIME and R

## The best of two worlds



# Why use KNIME and R?

---

## R

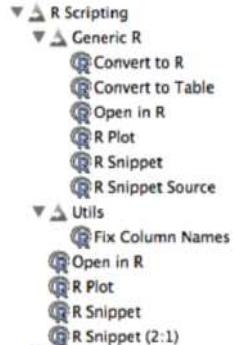
- Powerful statistics
- (b)Leading edge algorithms
- Powerful/flexible graphics
- Widely accepted language

## KNIME

- Powerful GUI
- Good Extract/Transform/Load
- Integrates diverse tools
- Enterprise grade solutions

# Two Integrations

---

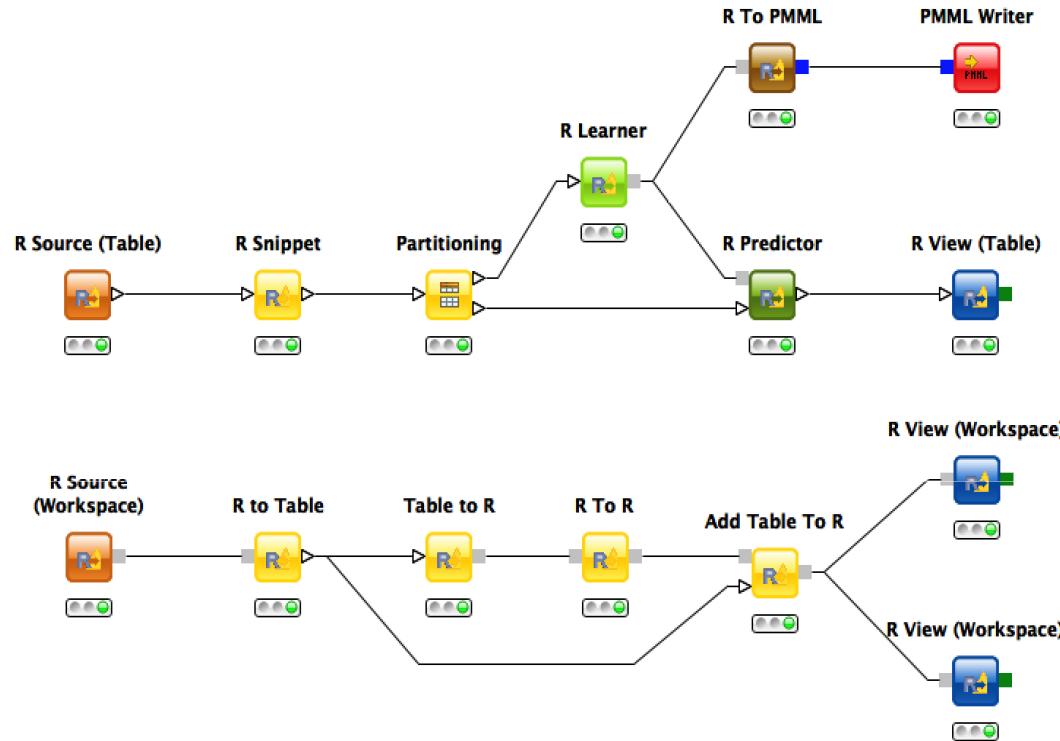


- Community (RServe Integration)



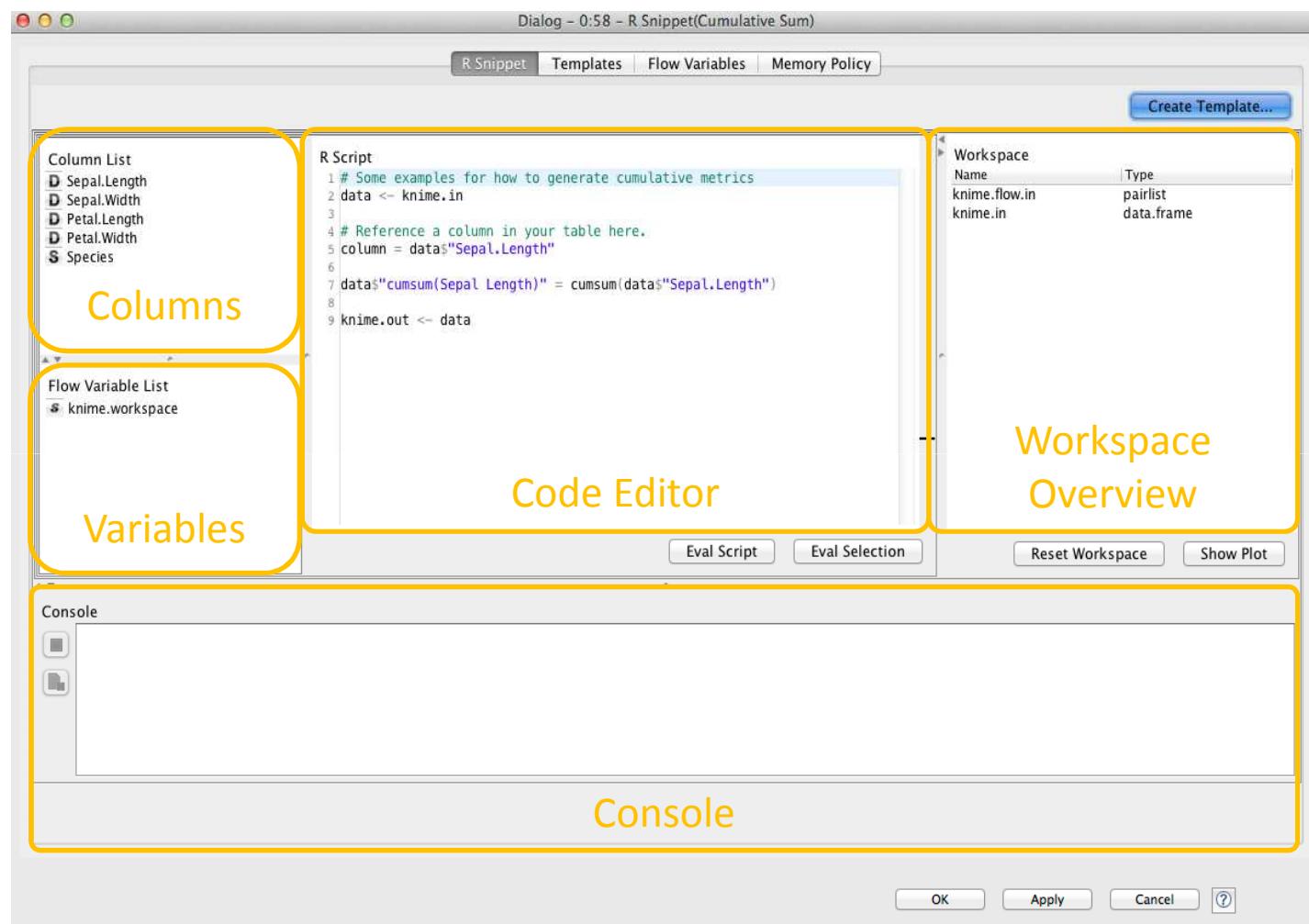
- R Interactive (Today's topic)

# Overview of new nodes

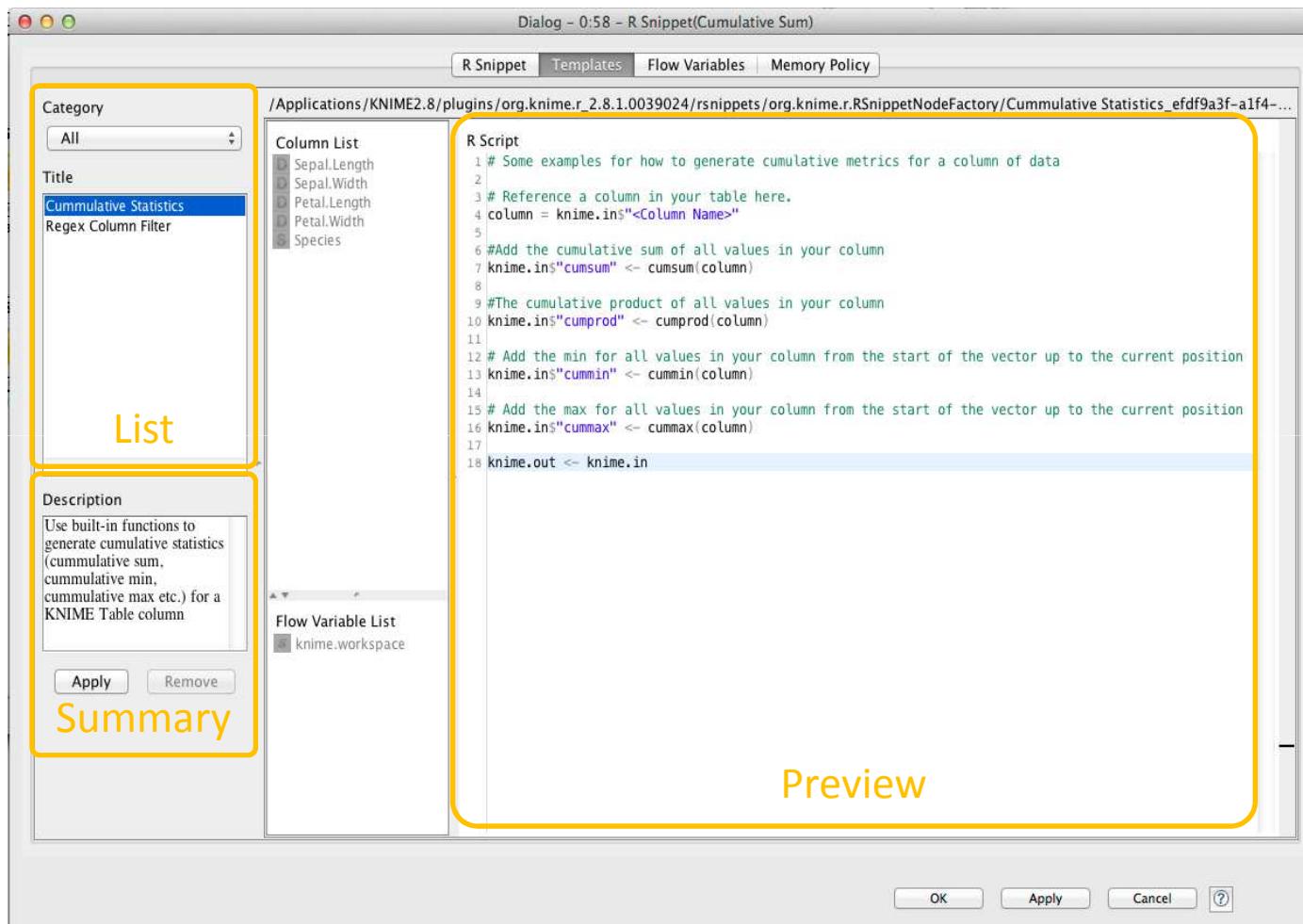


- Different input and output options
- Grey ports enable workspace branching

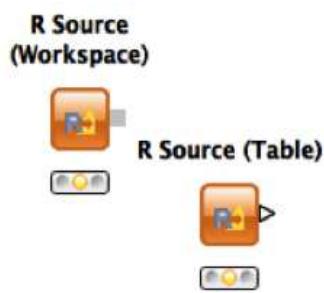
# The Interactive Editor



# Templates



# R Source nodes



```
R Script
1 # The foreign library provides access to many 3rd party data formats.
2 # Just a few examples are listed below, many others exist.
3 # More details cran.r-project.org/web/packages/foreign/foreign.pdf
4
5 library(foreign)
6
7 # map filepath from a flow variable here.
8 path = "/Users/knime/Desktop/iris.stx"
9
10 # Read SAS XPORT
11 data = read.xport(path)
12
13 knime.out <- data
```

A screenshot of the KNIME interface showing the output of the R Source (Table) node. The window title is "Data from R - 0:62 - R Source (Table)". The table has 149 rows and 5 columns. The columns are labeled "Row ID", "S:SETOSA", "D:X\_5\_1", "D:X\_3\_5", "D:X\_1\_4", and "D:X\_0\_". The data shows the first 10 rows of the Iris dataset.

Row ID	S:SETOSA	D:X_5_1	D:X_3_5	D:X_1_4	D:X_0_
1	setosa	4.9	3	1.4	0.2
2	setosa	4.7	3.2	1.3	0.2
3	setosa	4.6	3.1	1.5	0.2
4	setosa	5	3.6	1.4	0.2
5	setosa	5.4	3.9	1.7	0.4
6	setosa	4.6	3.4	1.4	0.3
7	setosa	5	3.4	1.5	0.2
8	setosa	4.4	2.9	1.4	0.2
9	setosa	4.9	3.1	1.5	0.1
10	setosa	5.4	3.7	1.5	0.2

- Get data from an R data frame
- Assign output to a data frame named knime.out
- Use with foreign, RCurl, or ...

# R Snippet nodes

The screenshot shows the KNIME interface. On the left, there is an R Snippet node with the following R script:

```
R Snippet
R Script
1 # Reference a column in your table here.
2 column = knime.in$"Sepal.Length"
3
4 data = knime.in
5 #Add the cumulative sum of all values in your column
6 data$cumsum(Sepal.Length) <- cumsum(column)
7
8 knime.out <- data
```

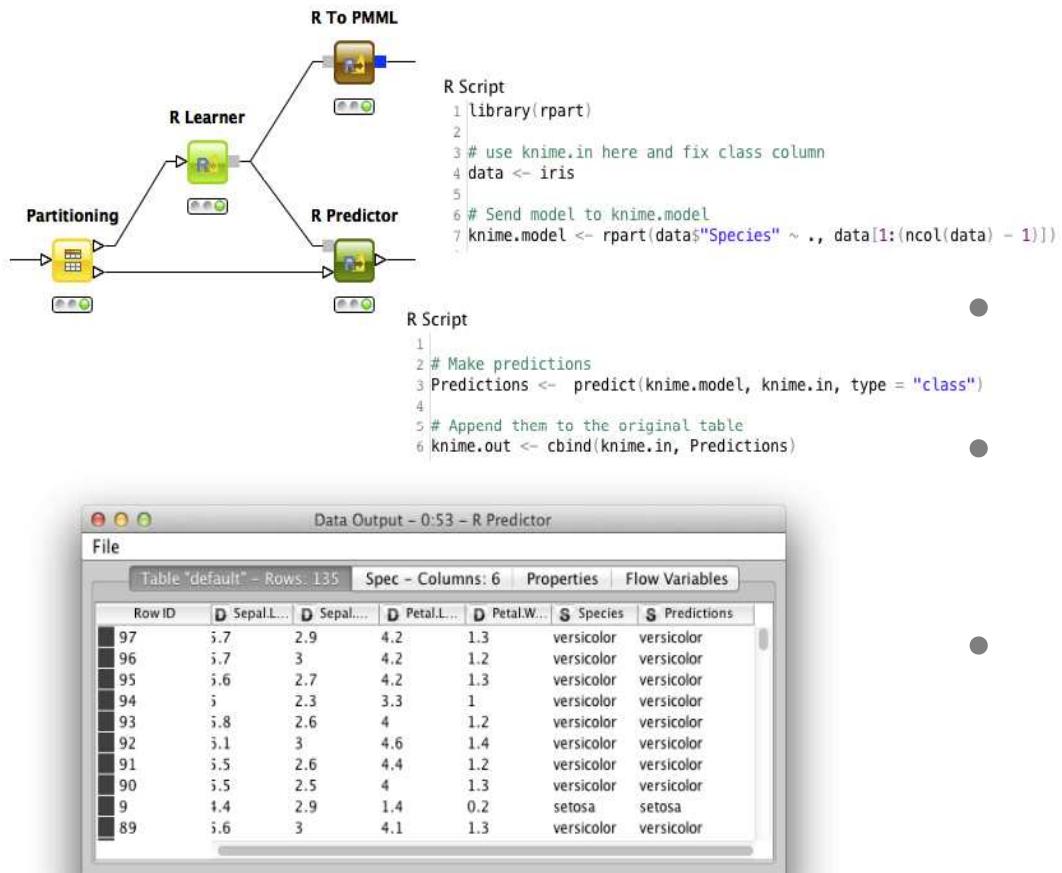
On the right, a Data Output node displays the resulting table "default" with 150 rows. The table has two columns: "Row ID" and "cumsum(Sepal.Length)". The data for the first 10 rows is as follows:

Row ID	cumsum(Sepal.Length)
2	10
3	14.7
4	19.3
5	24.3
6	29.7
7	34.3
8	39.3
9	43.7
10	48.6

- Generic data manipulation
- Edit tables or workspaces
- Derive knime.out from knime.in
- Use for cumulative stats, plyr, or

...

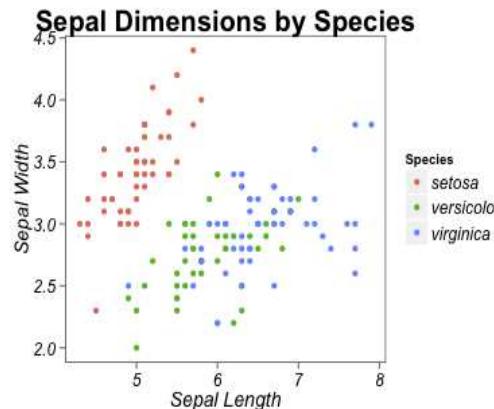
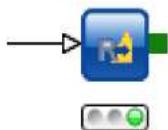
# R Data Mining nodes



- Use R models in KNIME
- Learner (knime.model) & Predictor motif
- R to PMML support for model portability

# R View nodes

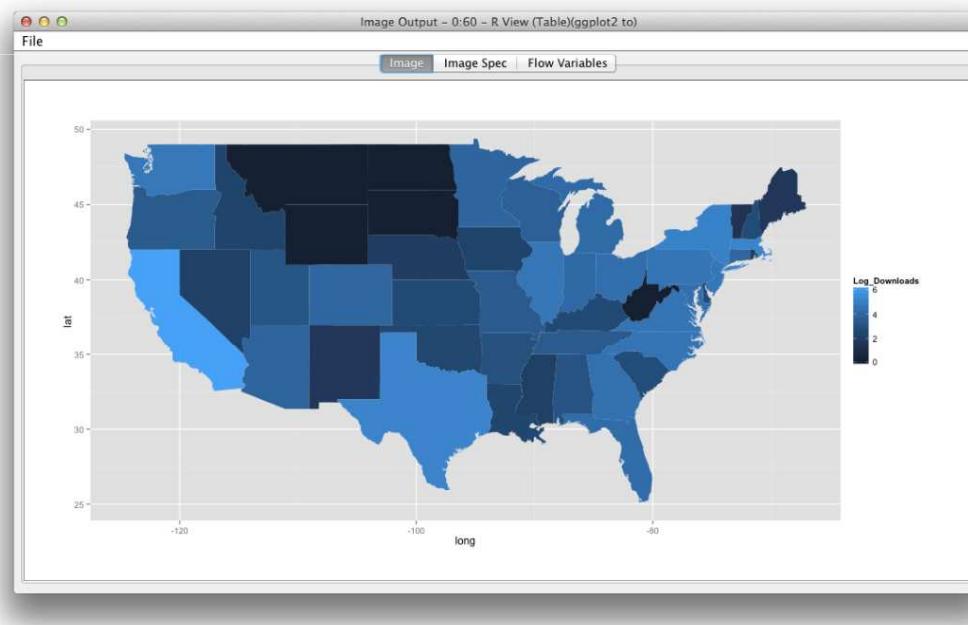
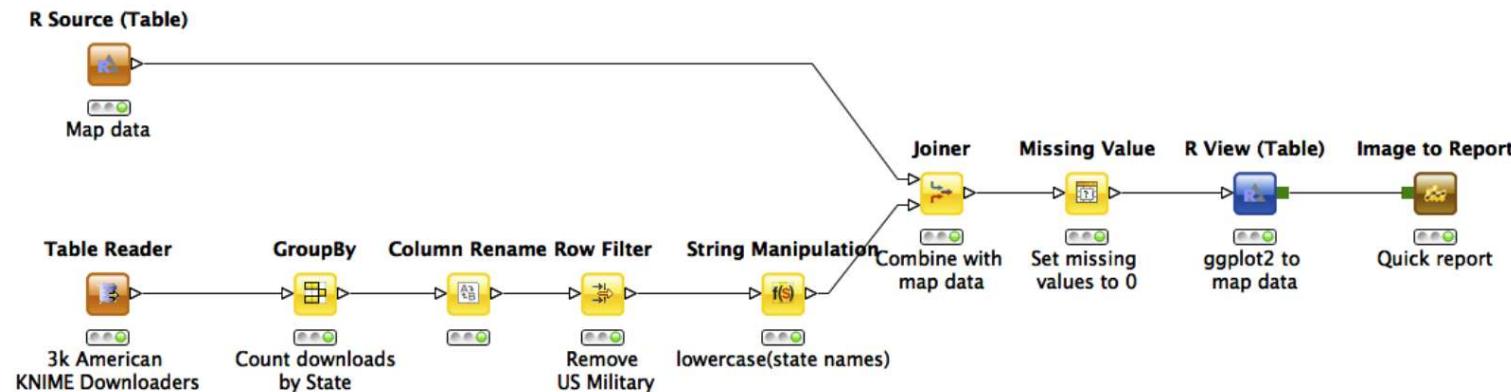
## R View (Table)



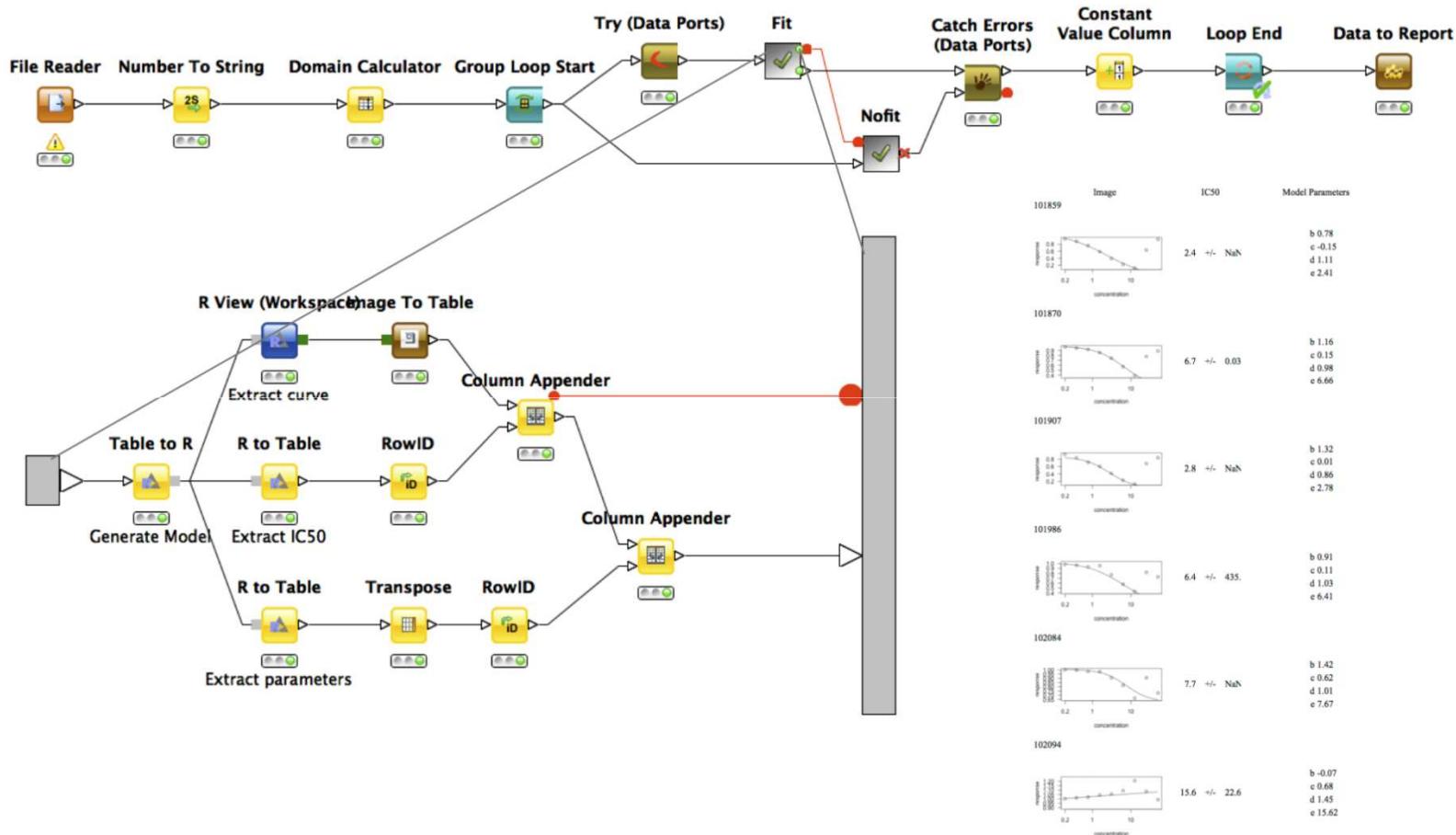
```
R Script
1 require(ggplot2)
2 require(grid)
3
4 # Insert data references here,
5 data = iris
6
7 x = data$Sepal.Length
8 x_label = "Sepal Length"
9
10 y = data$Sepal.Width
11 y_label = "Sepal Width"
12
13 #Column for coloring:
14 class = data$Species
15 legend_title = "Species"
16
17 # A plot title
18 title = "Sepal Dimensions by Species"
19
20 # Define a plot theme: http://docs.ggplot2.org/0.9.2.1/theme.html for more options
21 clean_theme = theme(panel.background = element_blank(),
22   plot.title = element_text(size=20, face="bold", colour = "black"),
23   panel.border = element_rect(color = "black", linetype = "solid", fill = "transparent"),
24   axis.title.x = element_text(size=14, face="italic", colour = "black"),
25   axis.title.y = element_text(size=14, face="italic", colour = "black"),
26   axis.text = element_text(size=12, face="italic", colour = "black"),
27   legend.text = element_text(size=12, face="italic", colour = "black"),
28   panel.grid = element_blank())
29 )
30
31 #Define some labels
32 labels = labelslist(title = title, x = x_label, y = y_label, color = legend_title)
33
34 # Generate a plot
35 plot = qplot(x, y, color = class, main = title)
36
37 # Apply theme and labels
38 plot + labels + clean_theme
```

- Generic R plots
- Plot(knime.in)
- Use with many packages including ggplot2

# R in Action: Choropleth Generation



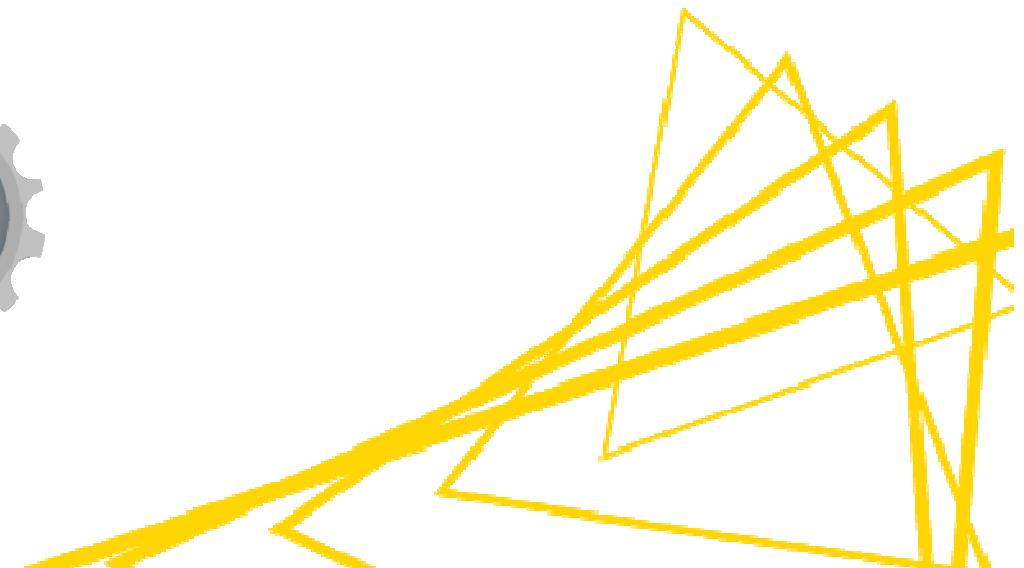
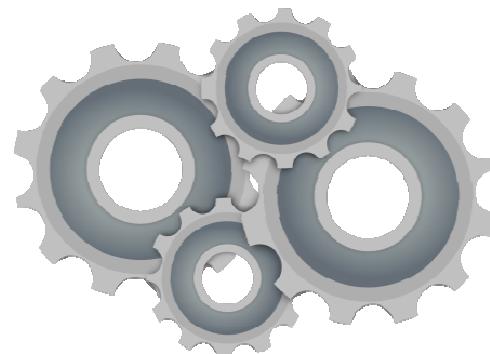
# R in Action: Dose Response modeling



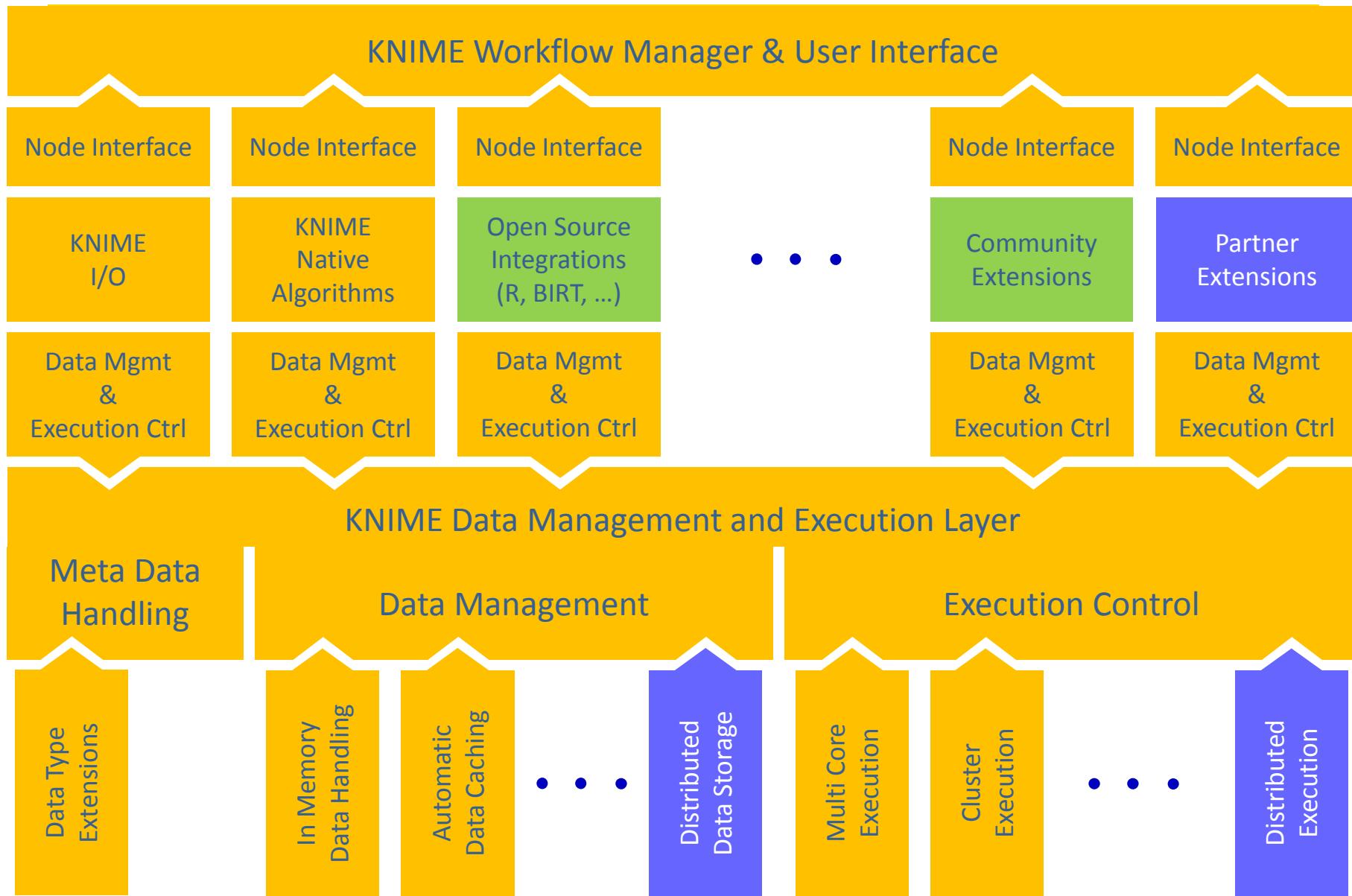
Created with KNIME Report Designer. Provided by KNIME.com AG, Zurich, Switzerland



# A Peak under the Hood: KNIME (Node) Development

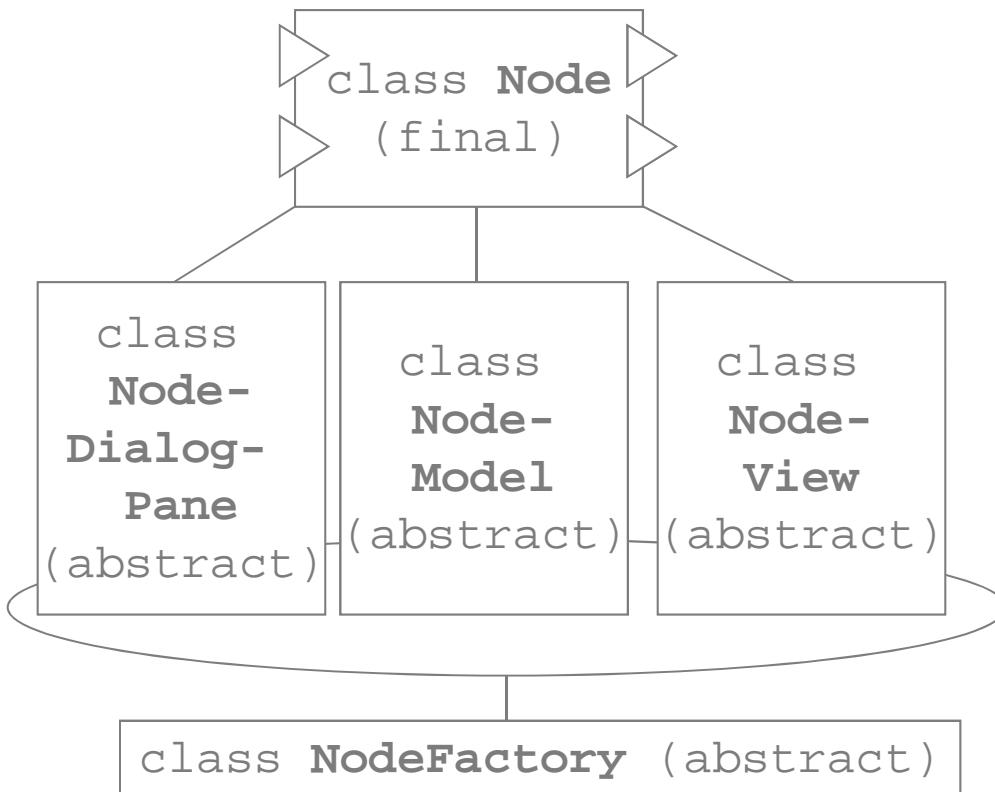


# KNIME Analytics Platform: Technology Overview



# Node Architecture

---



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

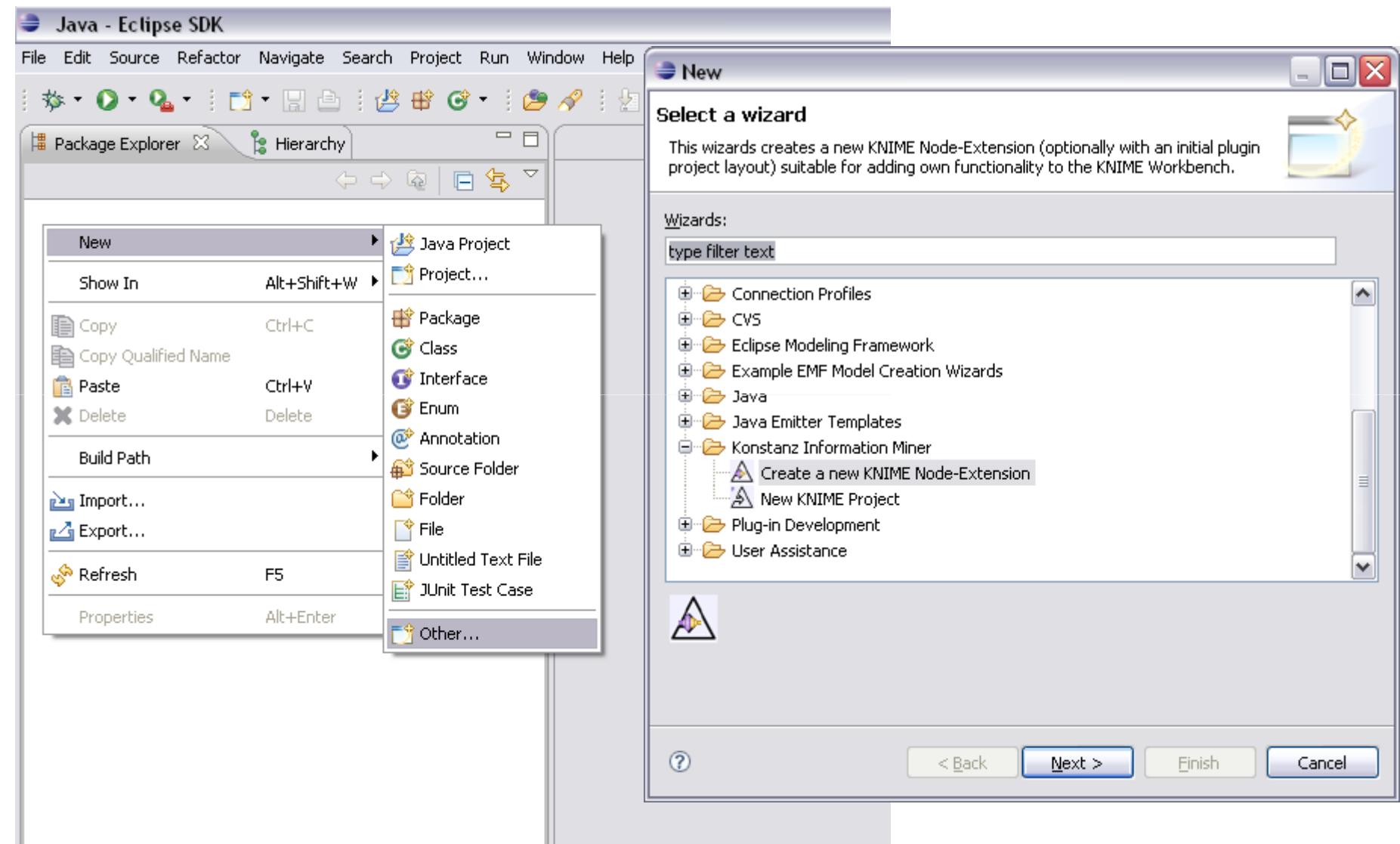
# Node Extension Wizard

---

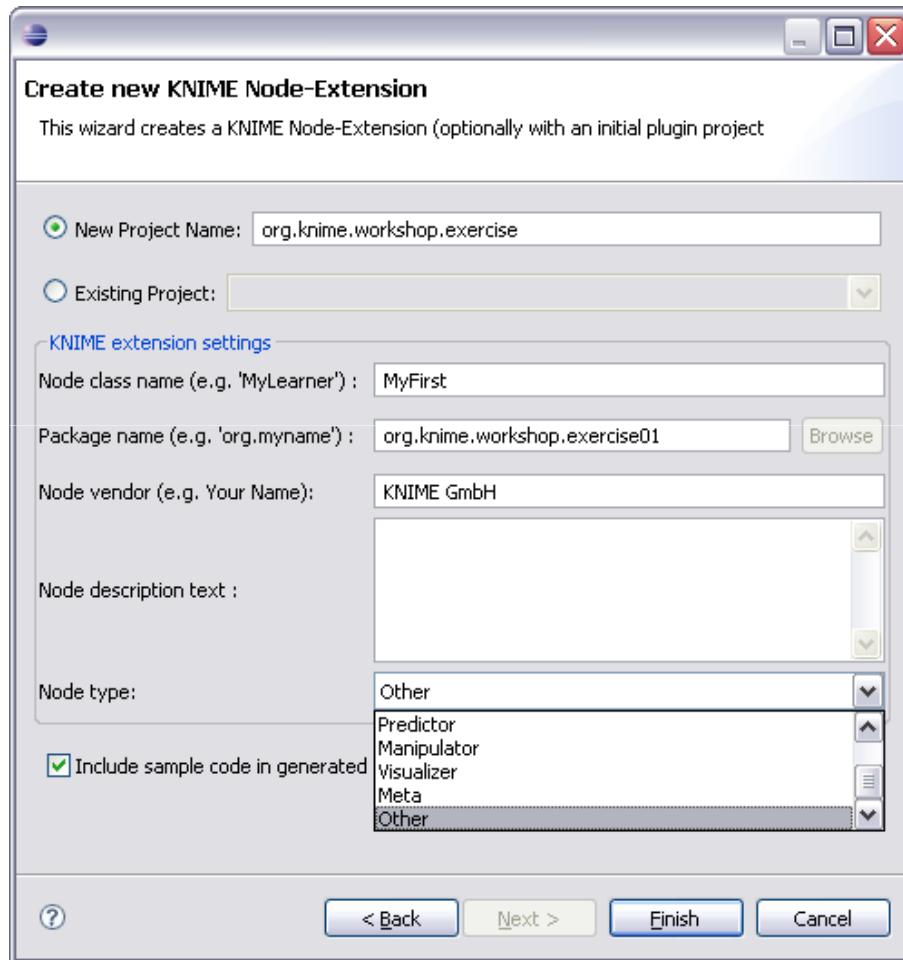


- Included in the KNIME Developer Version
- Allows creation of plugin projects including functioning KNIME nodes (with sample code)
- Helpful to easily create all node classes
  - Generates all Java classes
  - Node is registered with the plugin project
  - Launch KNIME and enjoy the new node working!

# Node Extension Wizard



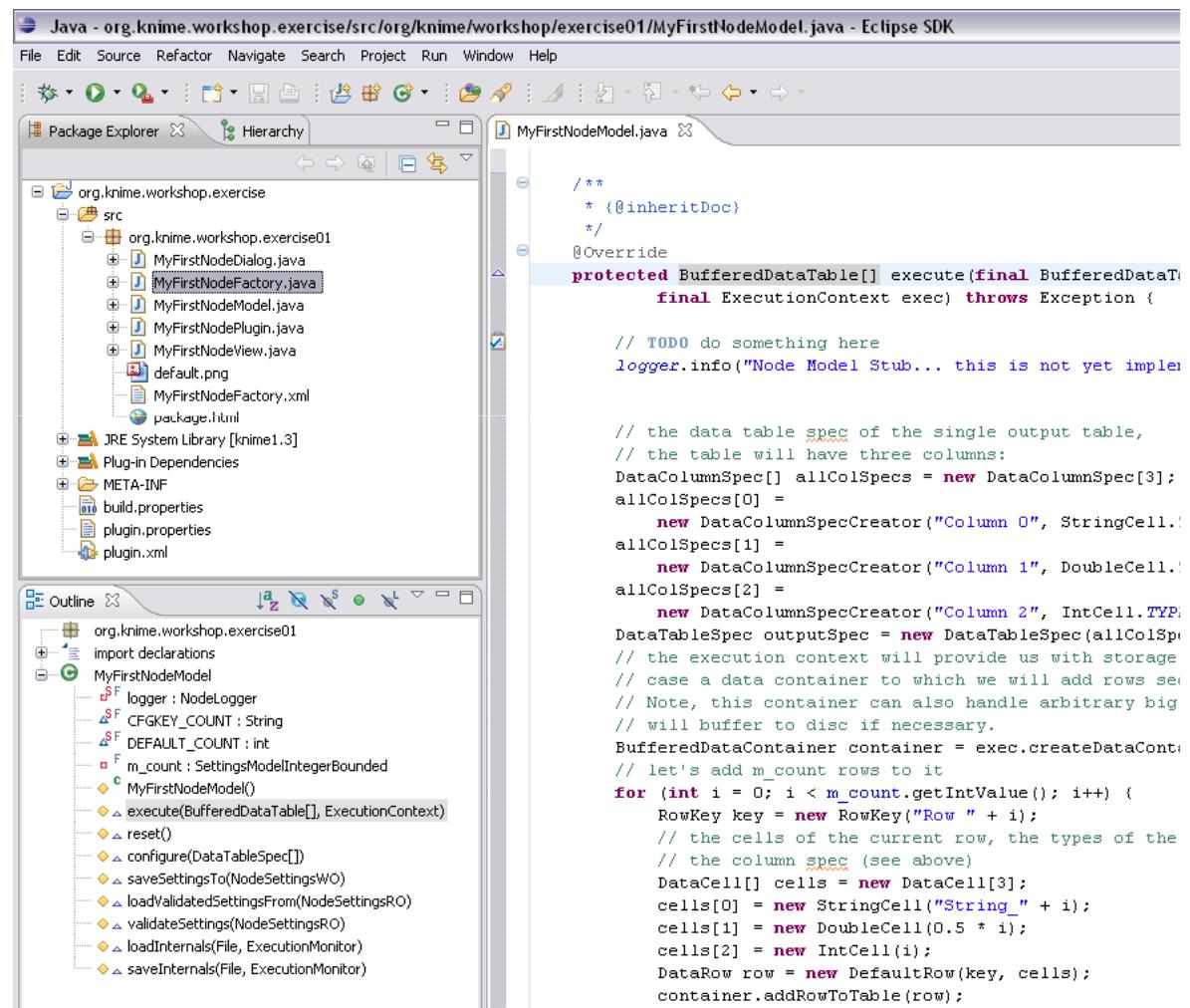
# Node Extension Wizard



- 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`



The screenshot shows the Eclipse IDE interface with the following details:

- Java Editor:** Displays the file `MyFirstNodeModel.java`. The code includes annotations like `@Override` and `@inheritDoc`, and handles `ExecutionContext` and `BufferedDataTable`.
- Package Explorer:** Shows the project structure under `org.knime.workshop.exercise`. It includes subfolders `src` and `META-INF`, and files like `MyFirstNodeDialog.java`, `MyFirstNodeFactory.java`, `MyFirstNodeModel.java`, `MyFirstNodePlugin.java`, `MyFirstNodeView.java`, `default.png`, `MyFirstNodeFactory.xml`, `package.html`, `JRE System Library [knime1.3]`, `Plug-in Dependencies`, `build.properties`, `plugin.properties`, and `plugin.xml`.
- Outline View:** Shows the class hierarchy and member declarations for `MyFirstNodeModel`.

# Resources

---

- KNIME pages ([www.knime.org](http://www.knime.org))
  - APPLICATIONS for example workflows
  - LEARNING HUB under RESOURCES  
[www.knime.org/learning-hub](http://www.knime.org/learning-hub)
- KNIME Tech pages ([tech.knime.org](http://tech.knime.org))
  - FORUM for questions and answers
  - DOCUMENTATION for documentation, FAQ, changelogs, ...
  - LABS where to find new experimental nodes
  - COMMUNITY CONTRIBUTIONS for development instructions and third party nodes
- KNIME TV channel on 

## Learning Hub

If you would like to learn more about KNIME, please take a look at the learning materials available on a variety of different media. There are many books, there are also lots of video materials available online and of course the KNIME courses and webinars.

[Basic KNIME](#)  
[Advanced KNIME](#)  
[Reporting](#)  
[KNIME Server](#)  
[Data Mining](#)  
[Web and Social Media Analysis](#)  
[Chemistry](#)  
[R and SAS](#)



# Thank you

