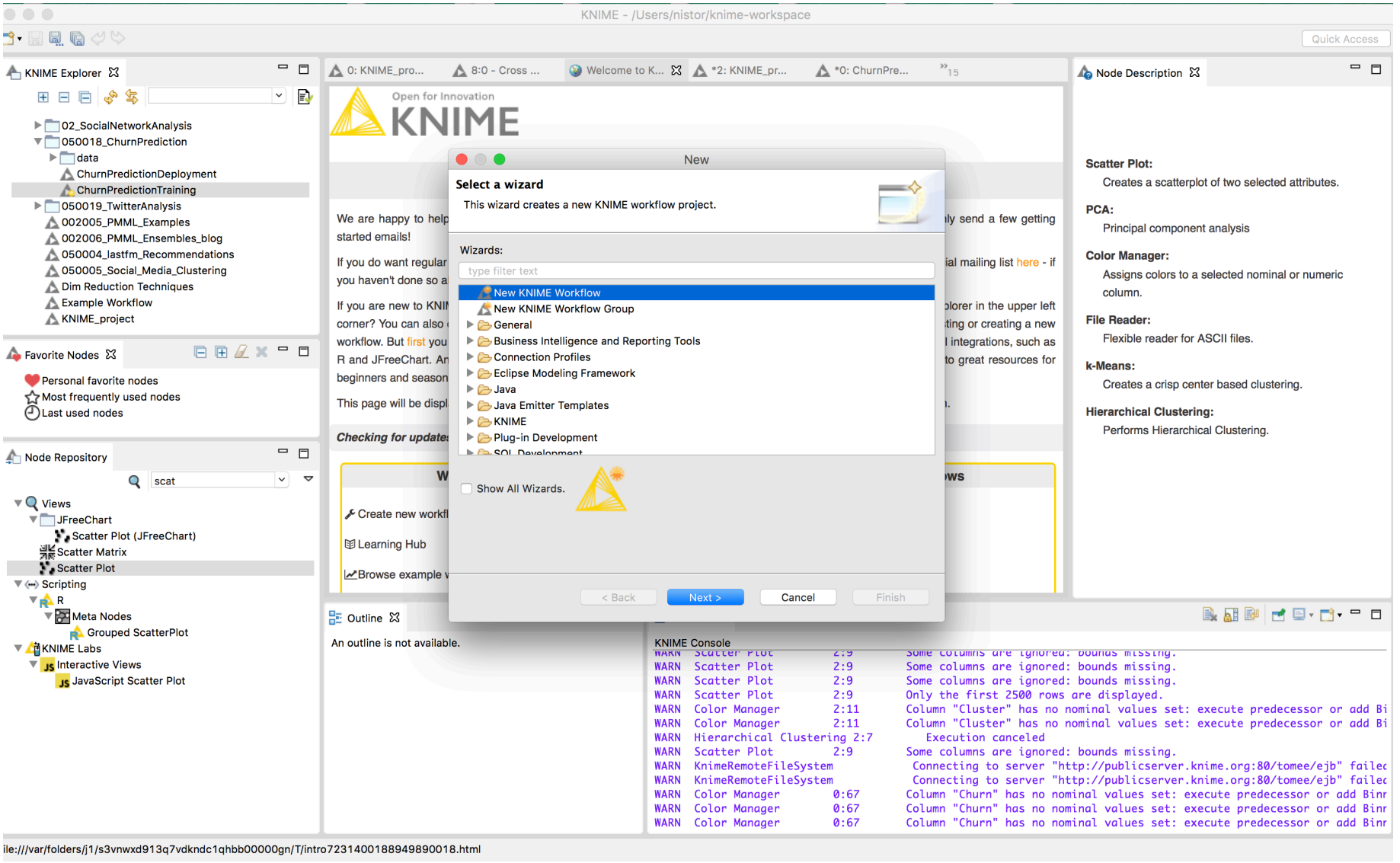
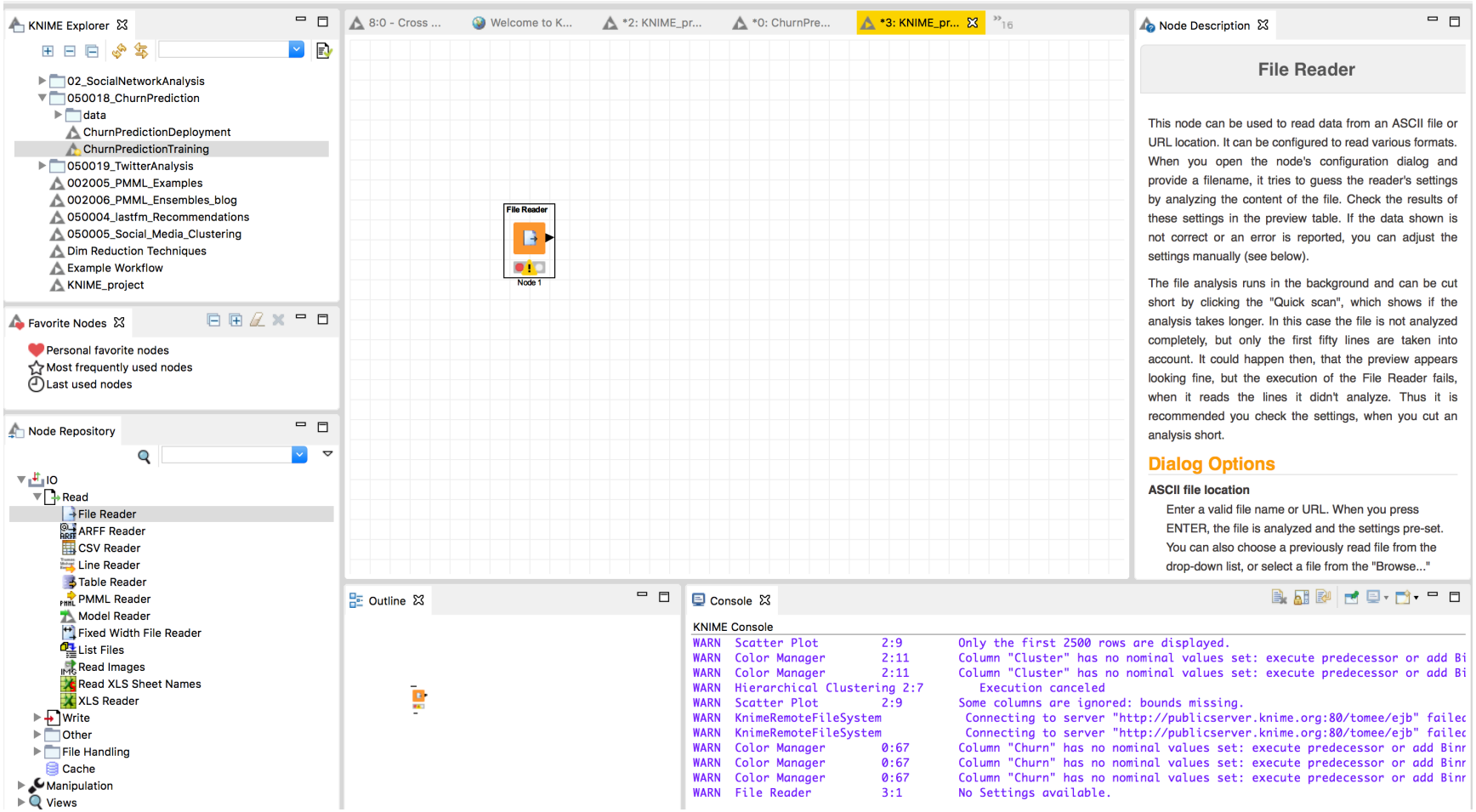
Data Mining, BI, Predictive Analysis, Visualization

with KNIME

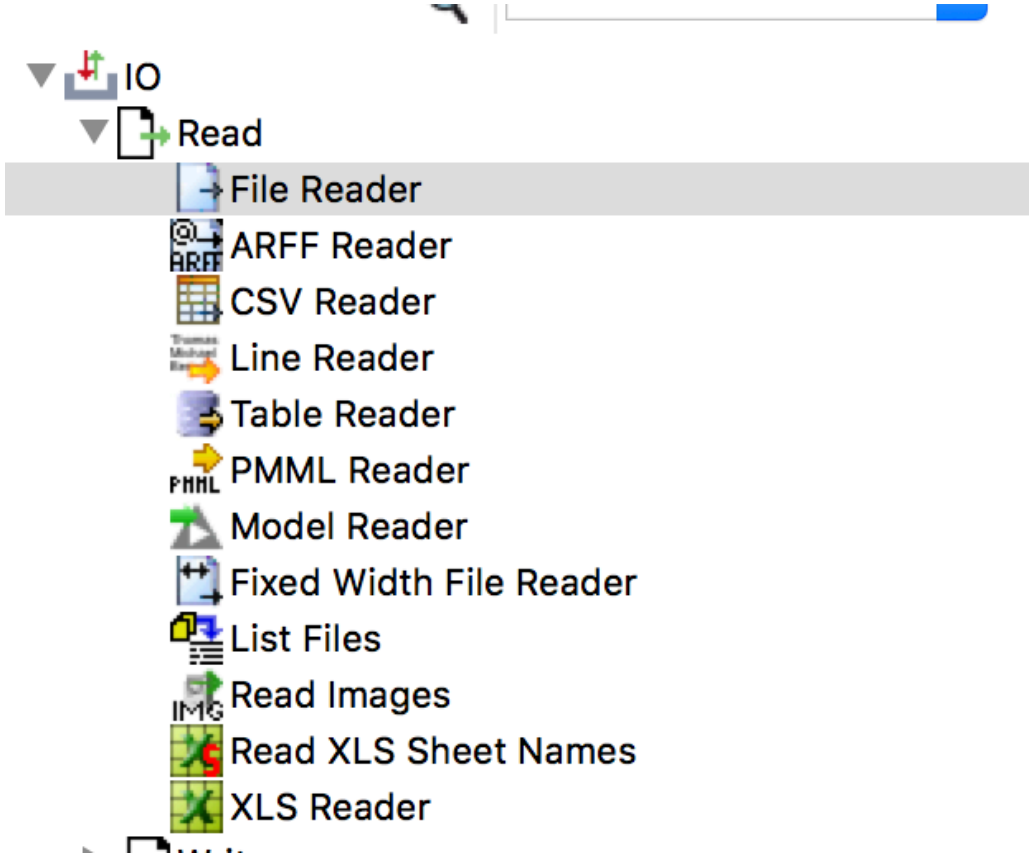
KNIME : create workflow



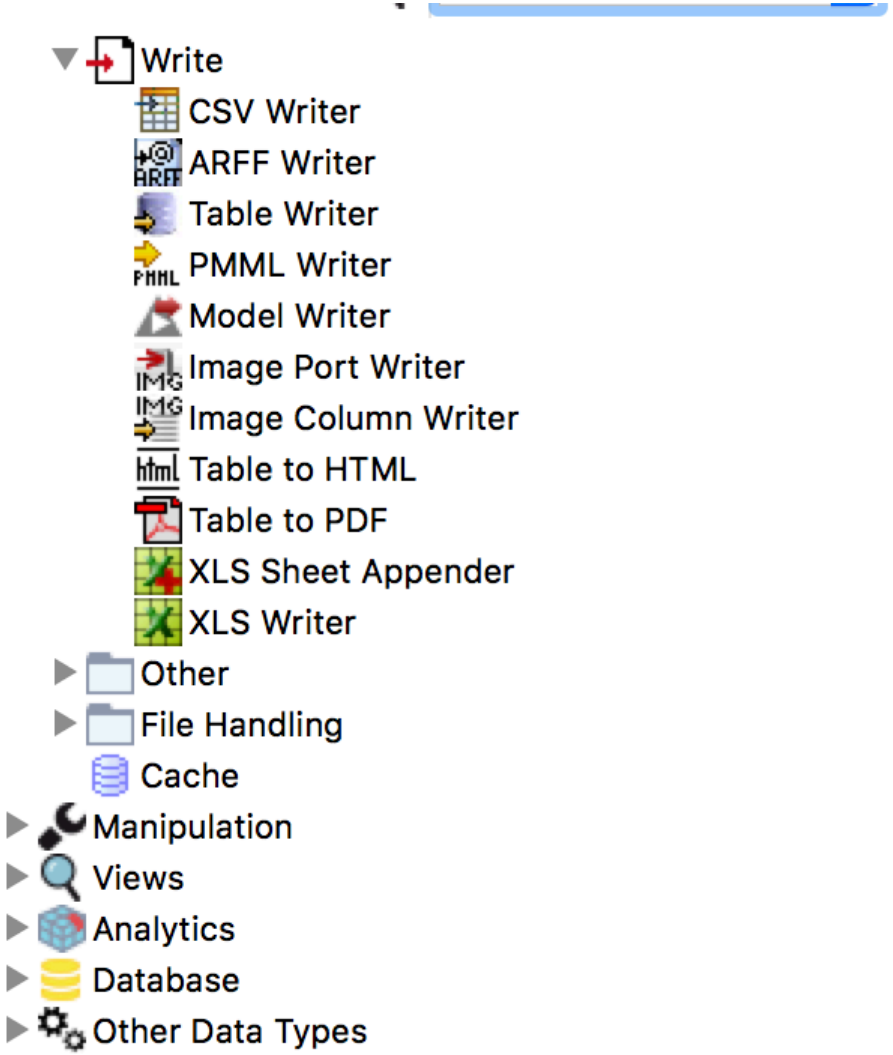
Read data


KNIME Explorer
02_SocialNetworkAnalysis
050018_ChurnPrediction
data
ChurnPredictionDeployment
ChurnPredictionTraining
050019_TwitterAnalysis
002005_PMML_Examples
002006_PMML_Ensembles_blog
050004_lastfm_Recommendations
050005_Social_Media_Clustering
Dim Reduction Techniques
Example Workflow
KNIME_project
Favorite Nodes
Personal favorite nodes
Most frequently used nodes
Last used nodes
Node Repository
IO
Read
File Reader
ARFF Reader
CSV Reader
Line Reader
Table Reader
PMML Reader
Model Reader
Fixed Width File Reader
List Files
Read Images
Read XLS Sheet Names
XLS Reader
Write
Other
File Handling
Cache
Manipulation
Views
8:0 - Cross ...
Welcome to K...
*2: KNIME_pr...
*0: ChurnPre...
*3: KNIME_pr...
File Reader
Node 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
ASCII file location
Enter a valid file name or URL. When you press ENTER, the file is analyzed and the settings pre-set. You can also choose a previously read file from the drop-down list, or select a file from the "Browse..."
Console
KNIME Console
WARN Scatter Plot 2:9 Only the first 2500 rows are displayed.
WARN Color Manager 2:11 Column "Cluster" has no nominal values set: execute predecessor or add Bi
WARN Color Manager 2:11 Column "Cluster" has no nominal values set: execute predecessor or add Bi
WARN Hierarchical Clustering 2:7 Execution canceled
WARN Scatter Plot 2:9 Some columns are ignored: bounds missing.
WARN KnimeRemoteFileSystem Connecting to server "http://publicserver.knime.org:80/tomee/ejb" failed
WARN KnimeRemoteFileSystem Connecting to server "http://publicserver.knime.org:80/tomee/ejb" failed
WARN Color Manager 0:67 Column "Churn" has no nominal values set: execute predecessor or add Binr
WARN Color Manager 0:67 Column "Churn" has no nominal values set: execute predecessor or add Binr
WARN Color Manager 0:67 Column "Churn" has no nominal values set: execute predecessor or add Binr
WARN File Reader 3:1 No Settings available.

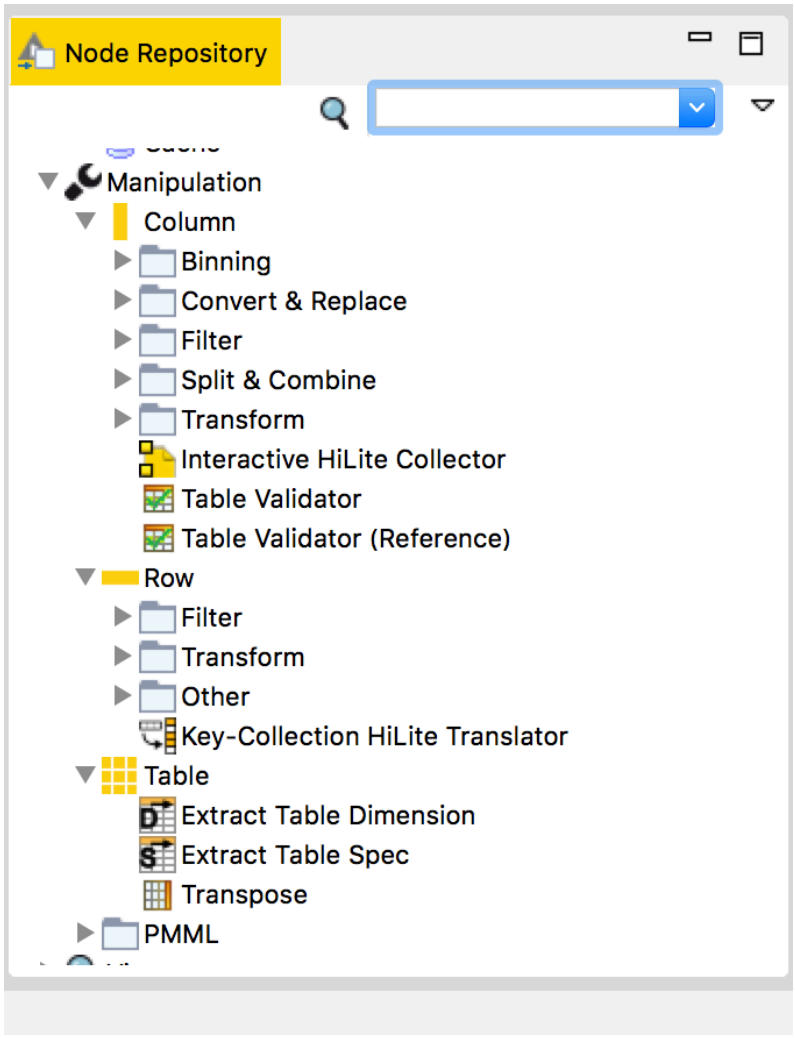
Several types of importing



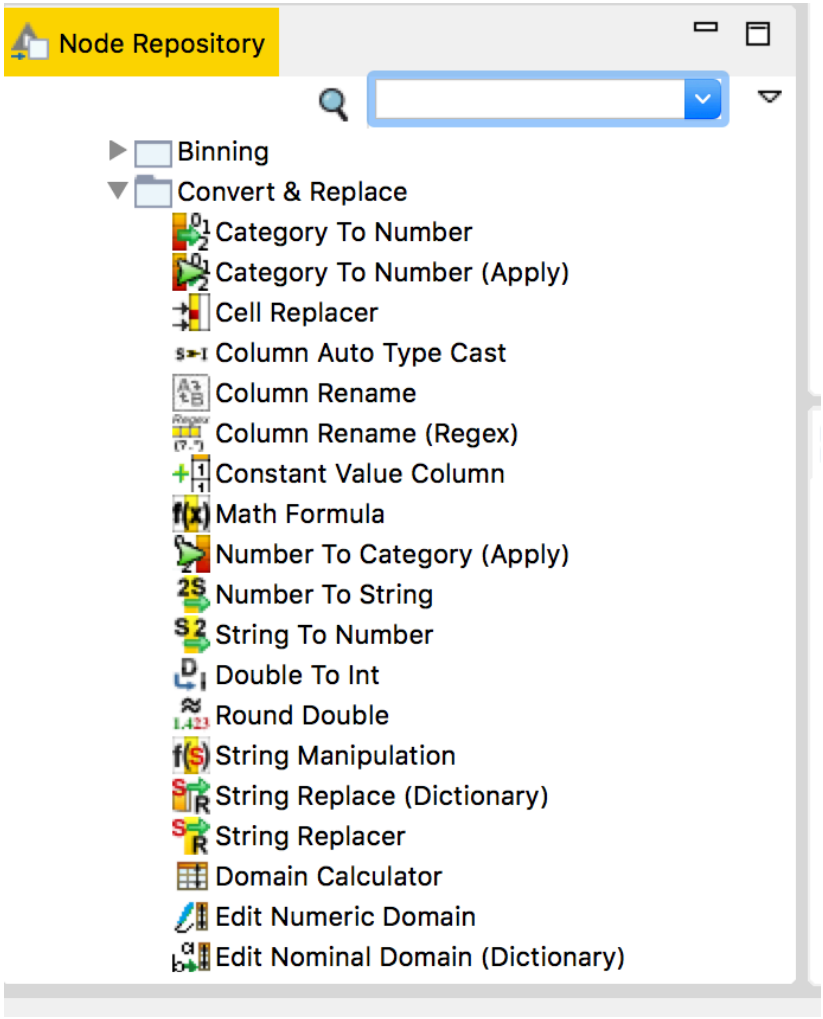
Write data & results



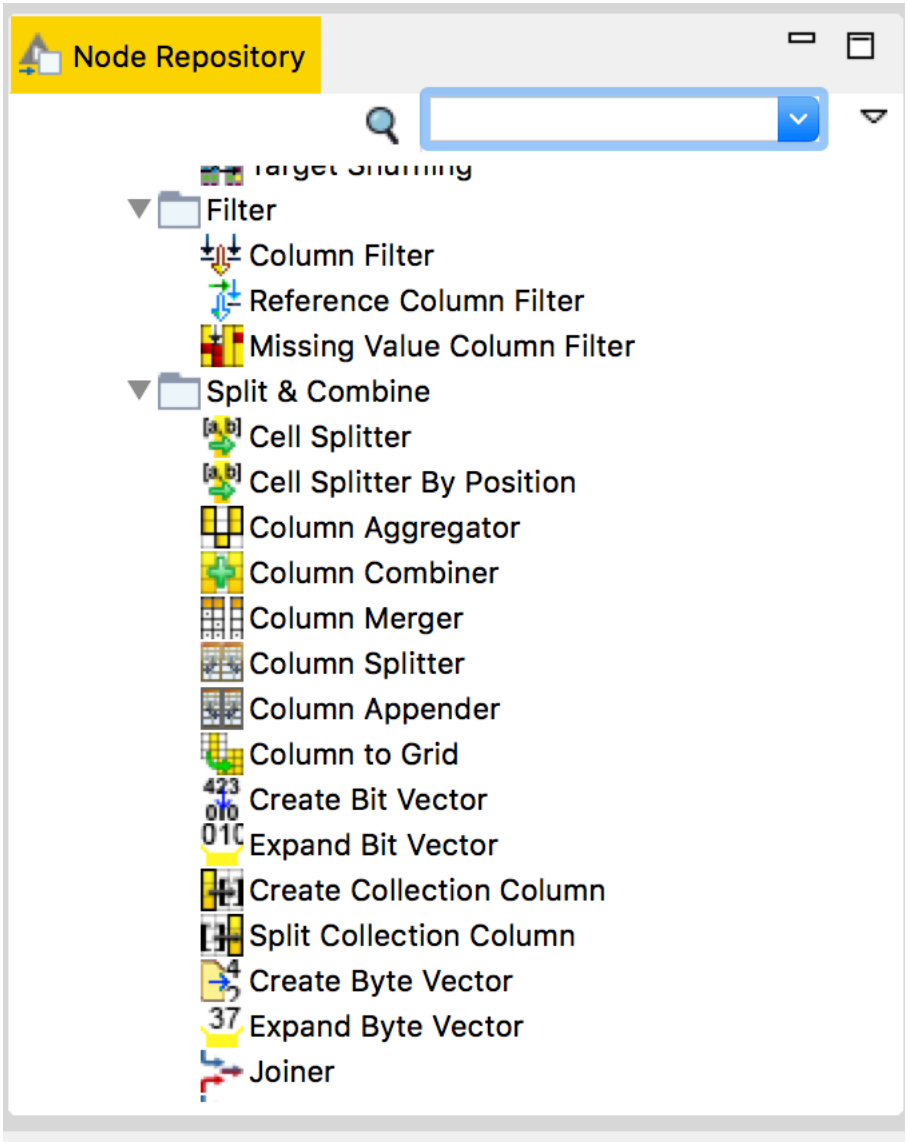
Data manipulation



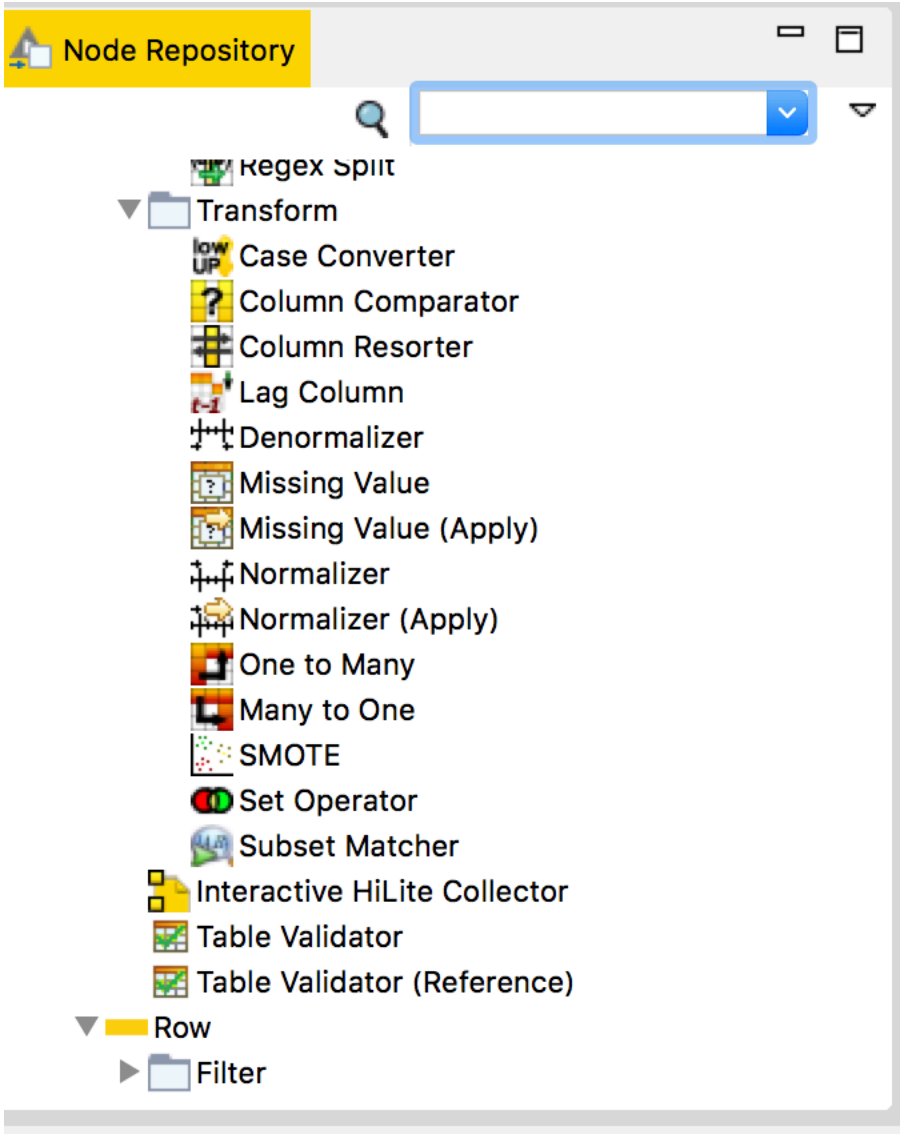
Data : convert & replace



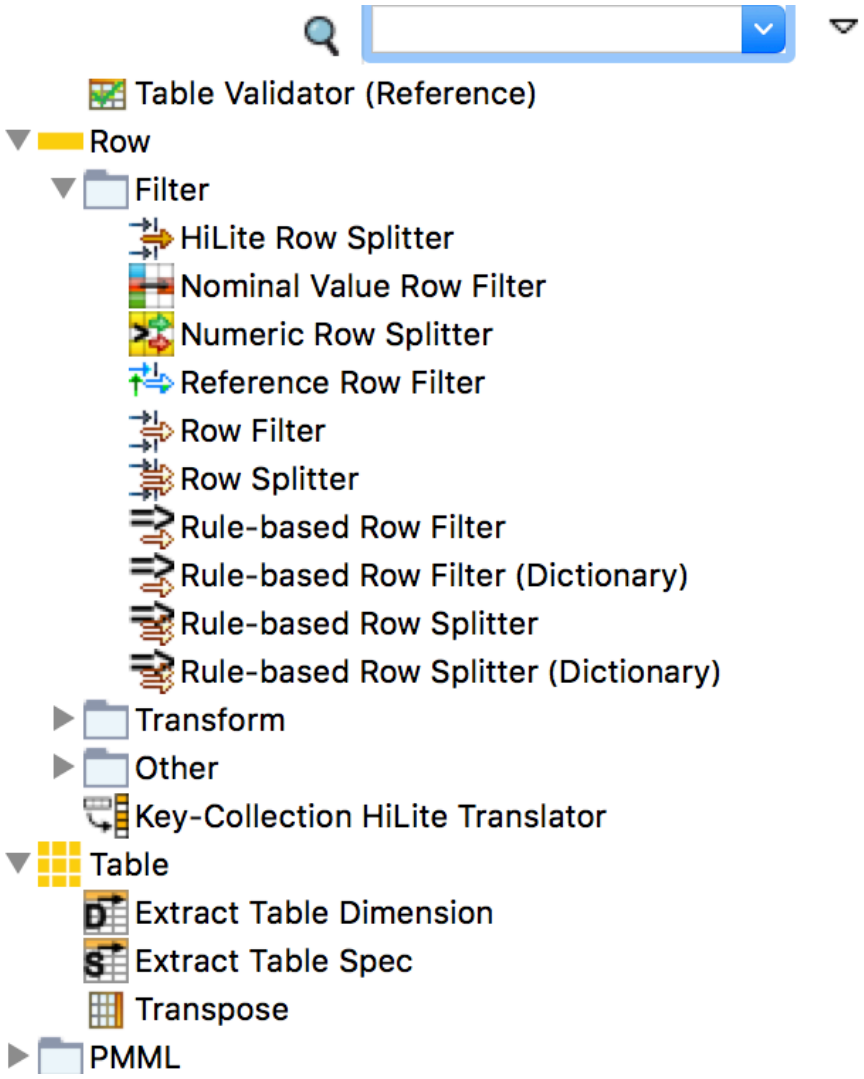
Split & Combine data



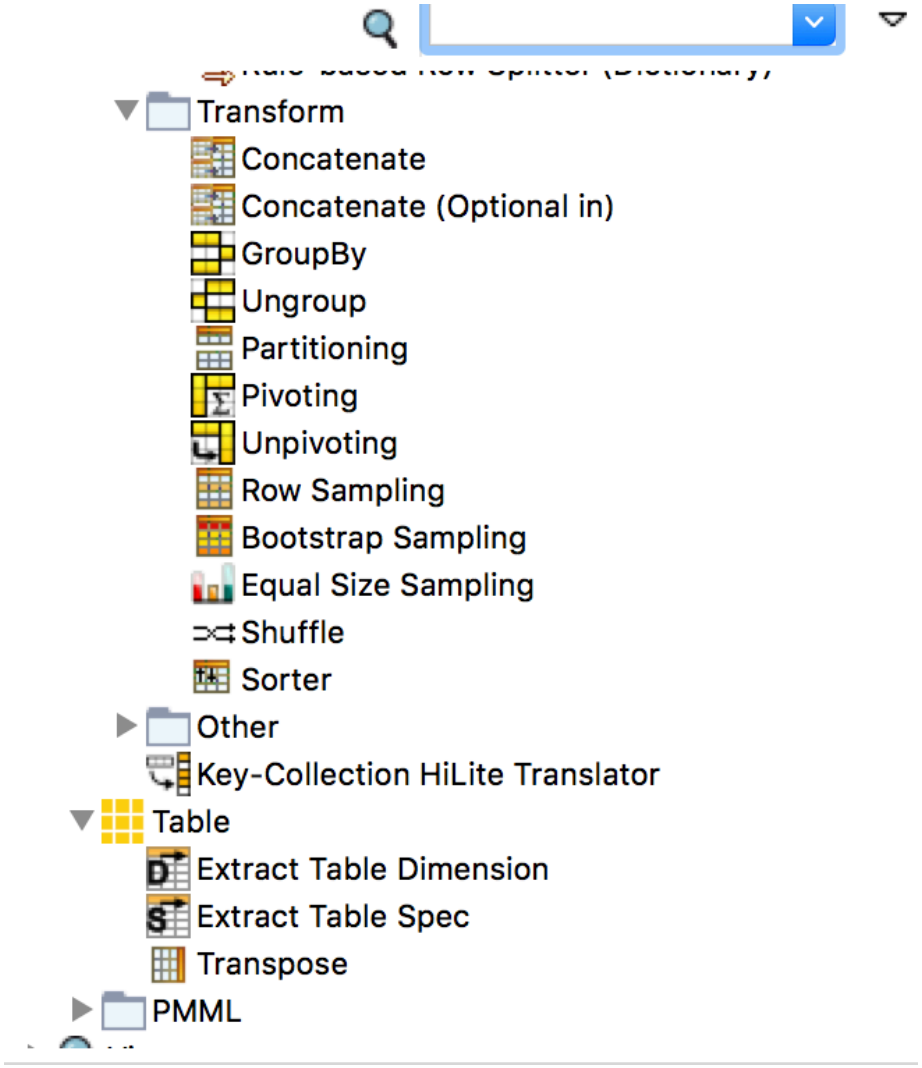
Data transformation



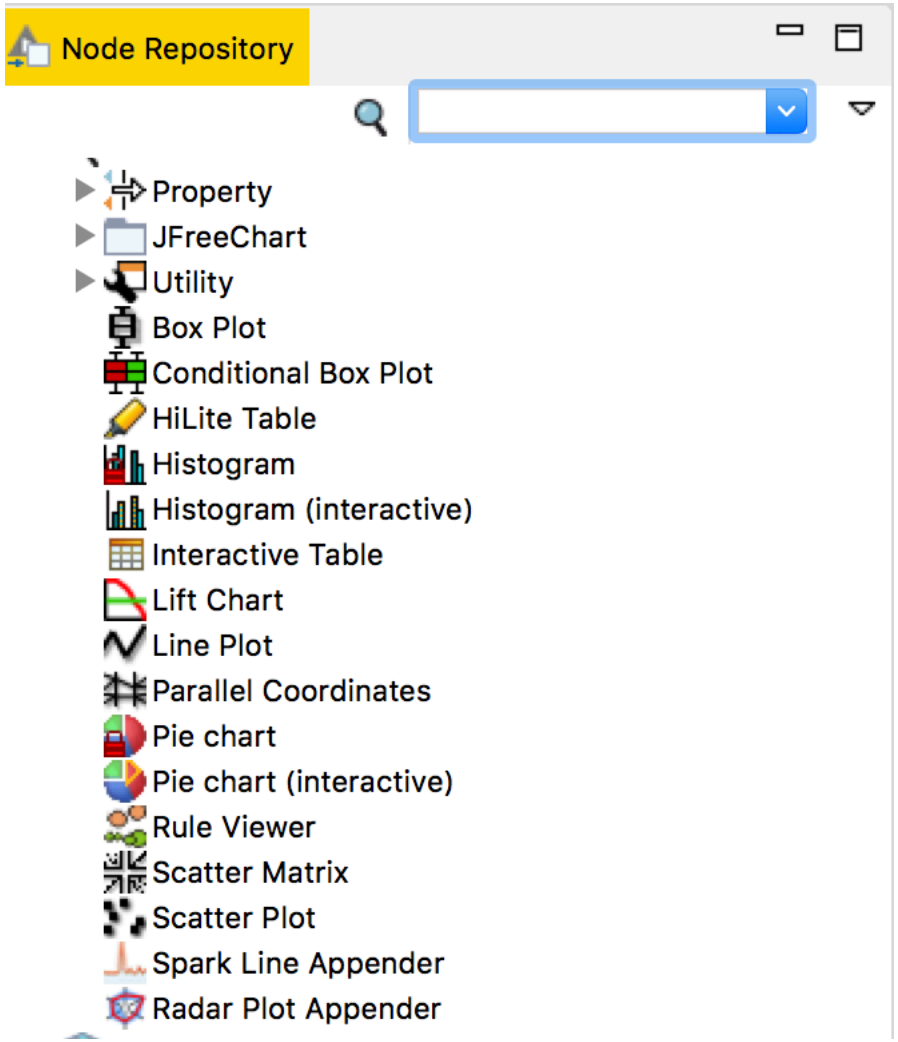
Transform data : by Row



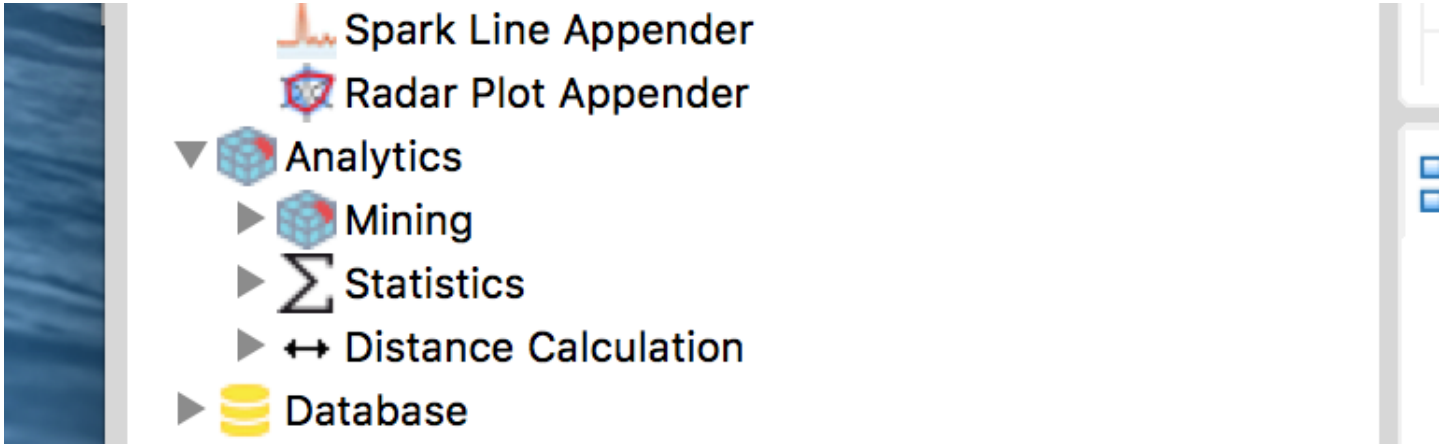
Transform data: by Column



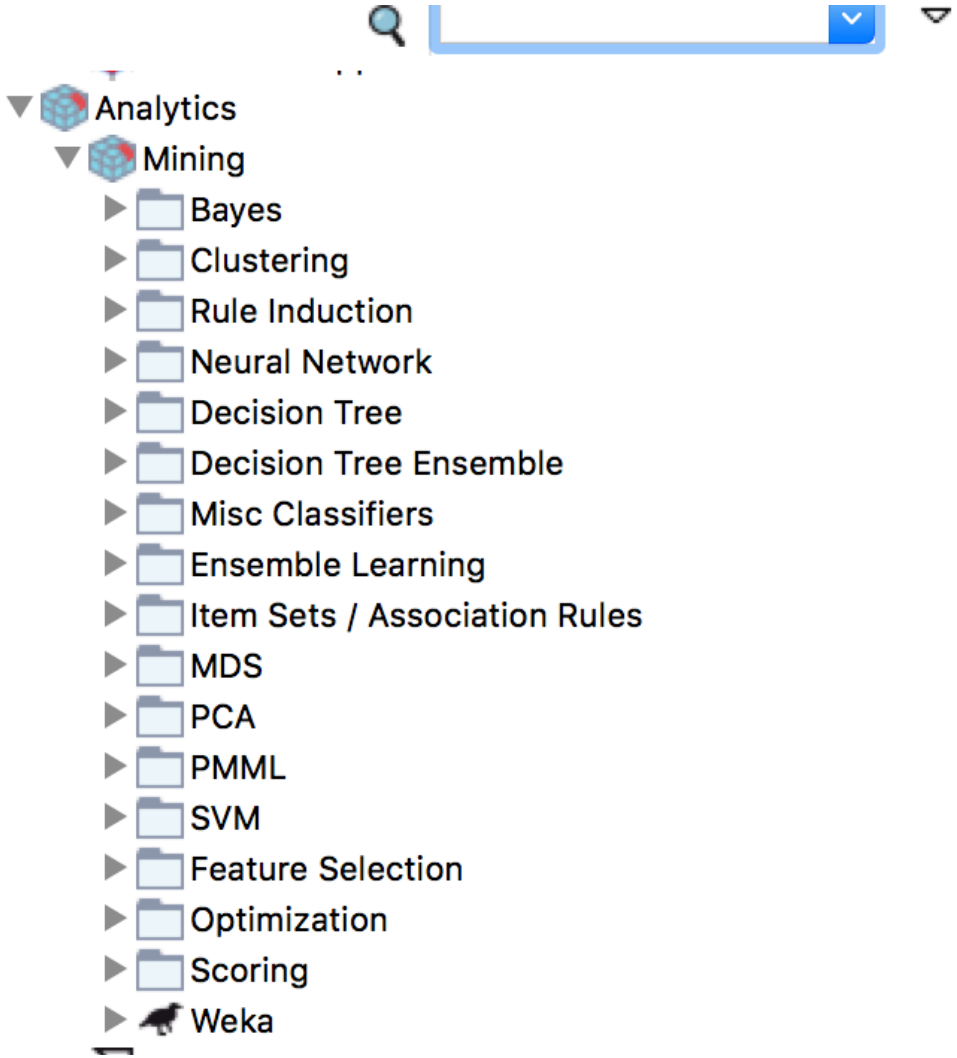
Data visualization



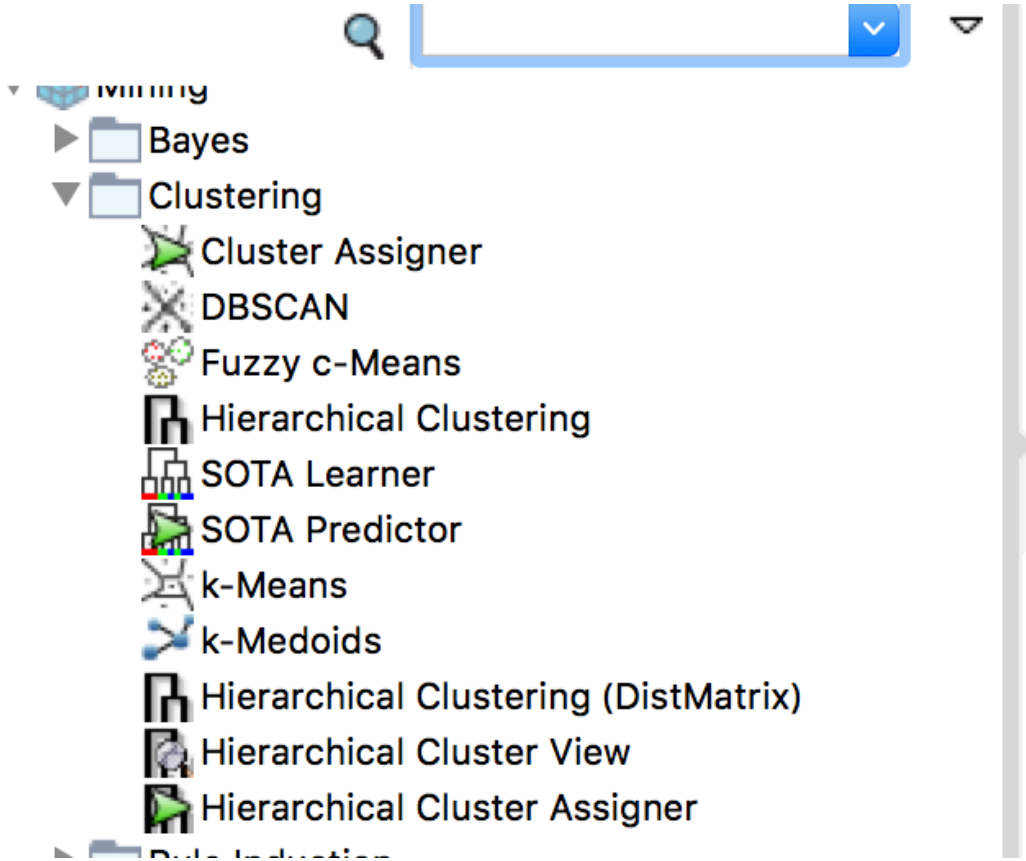
Other functions



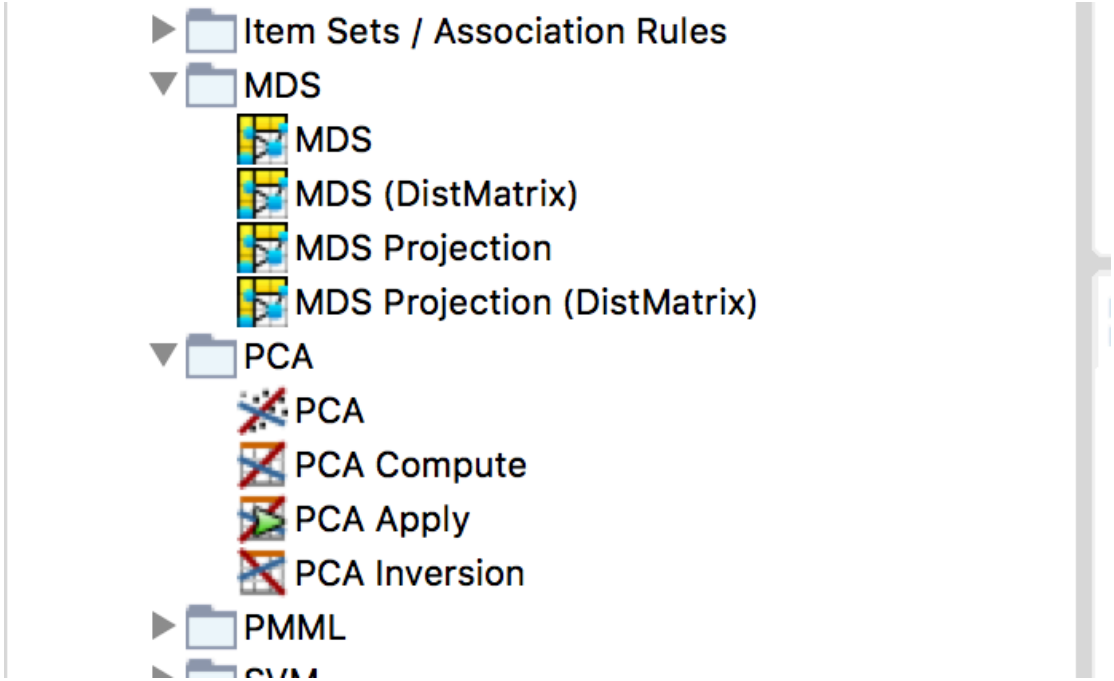
Knime: Analytics & Data Mining



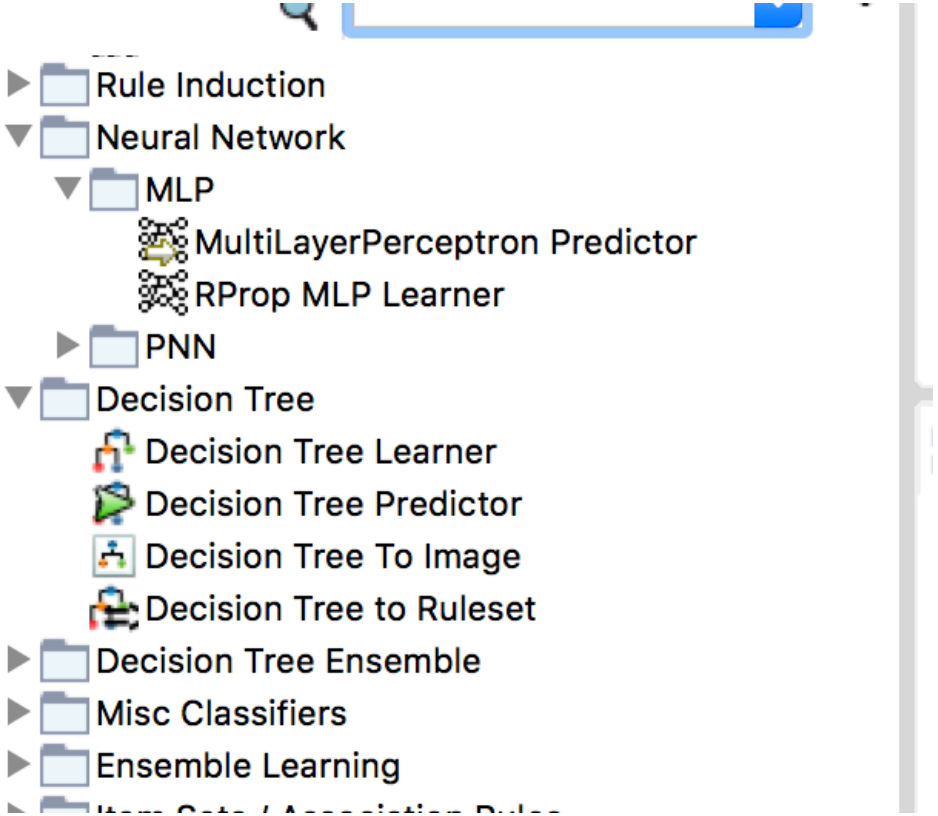
Clustering data



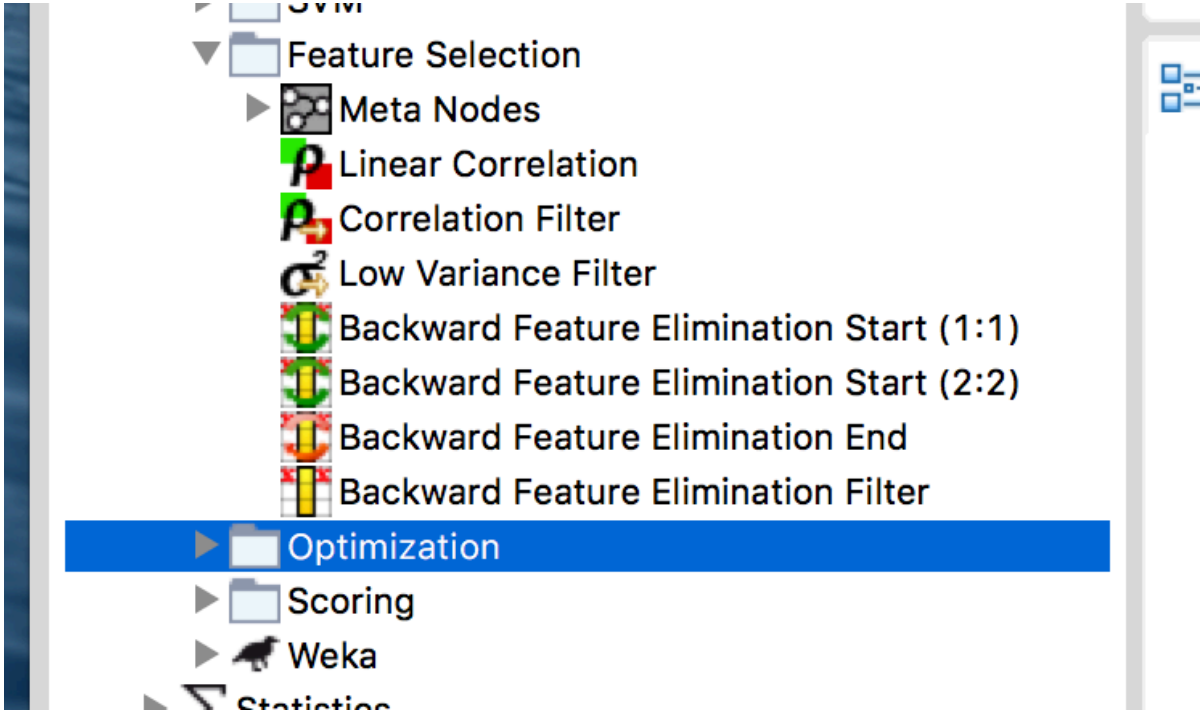
Data projections methods



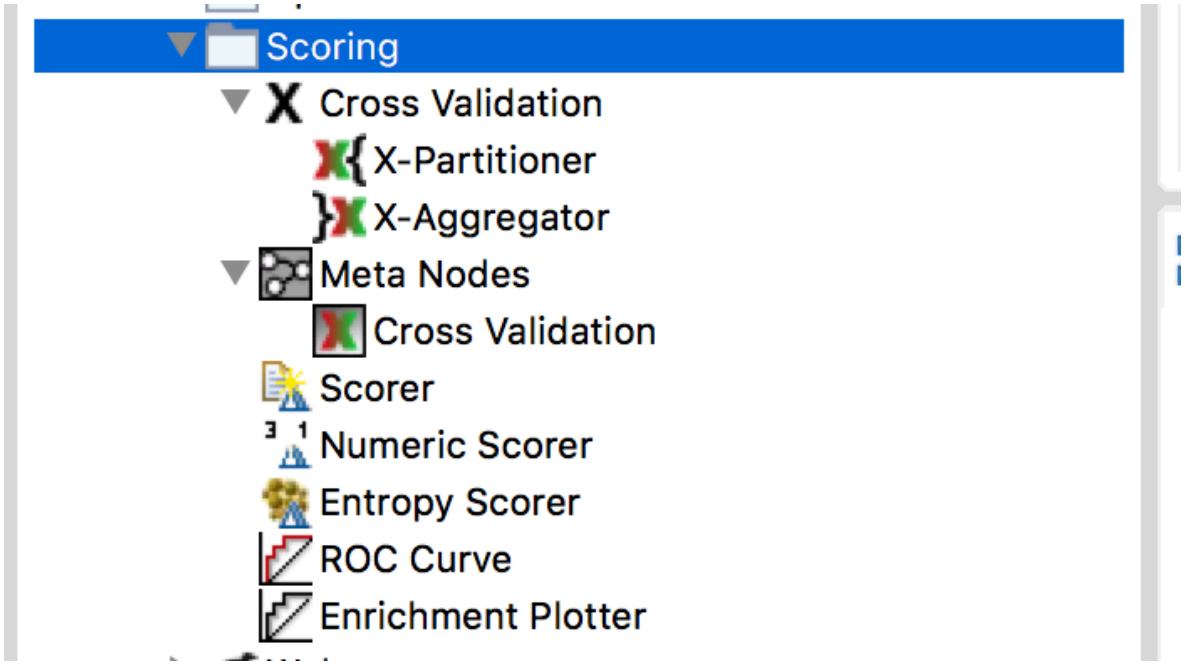
Predictions models

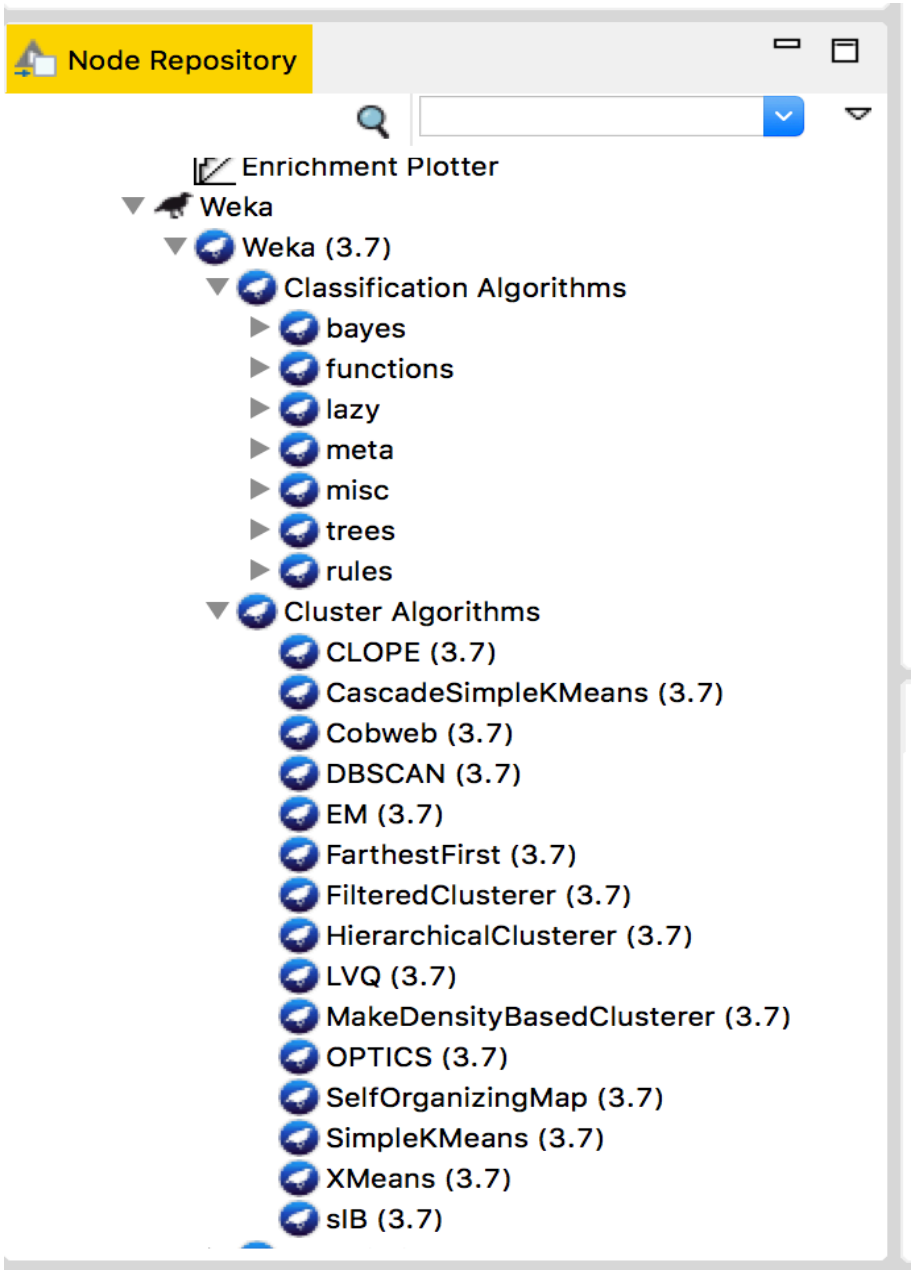


Features selections approaches

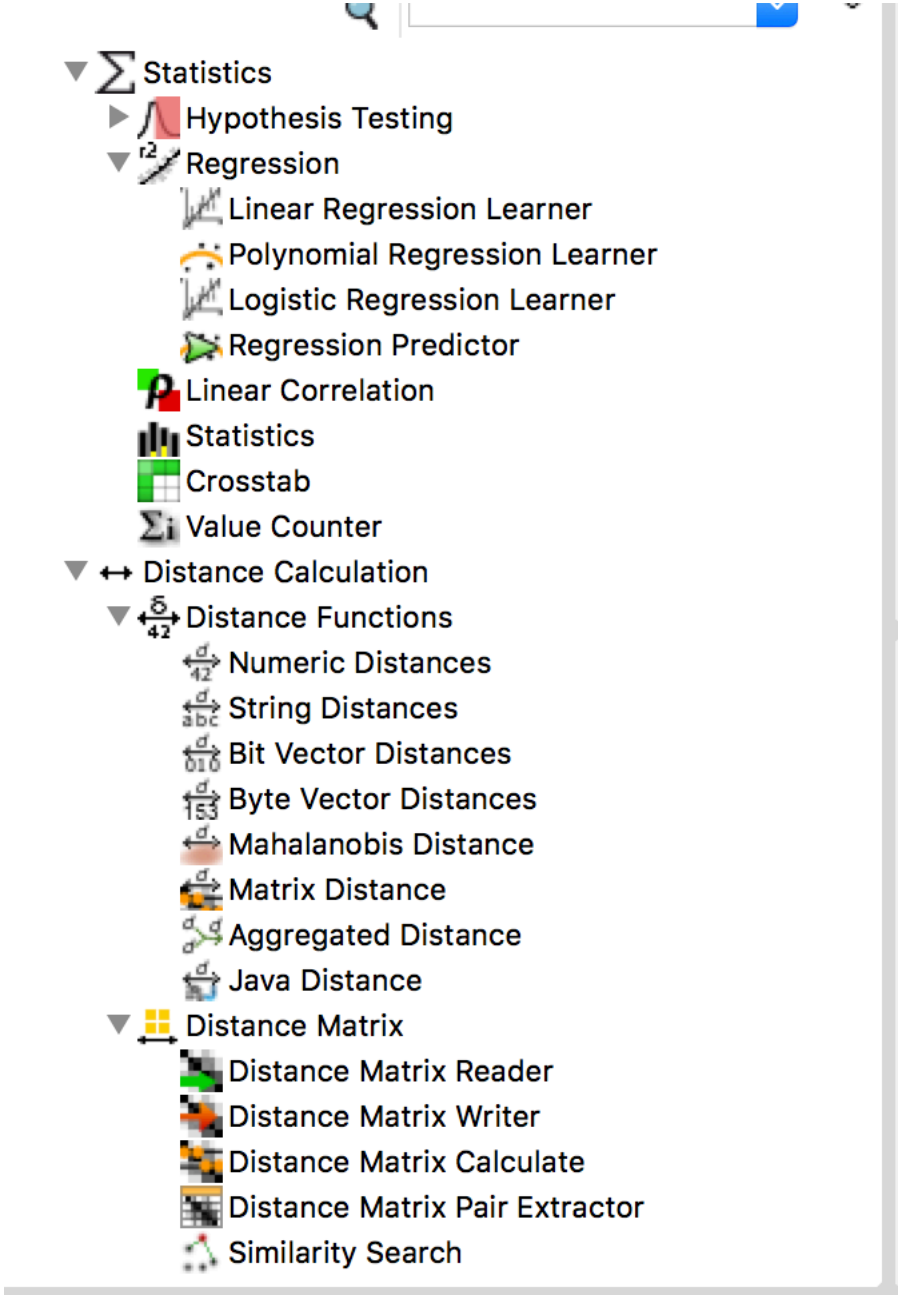


Validation & Scoring




Node Repository
Enrichment Plotter
Weka
Weka (3.7)
Classification Algorithms
bayes
functions
lazy
meta
misc
trees
rules
Cluster Algorithms
CLOPE (3.7)
CascadeSimpleKMeans (3.7)
Cobweb (3.7)
DBSCAN (3.7)
EM (3.7)
FarthestFirst (3.7)
FilteredClusterer (3.7)
HierarchicalClusterer (3.7)
LVQ (3.7)
MakeDensityBasedClusterer (3.7)
OPTICS (3.7)
SelfOrganizingMap (3.7)
SimpleKMeans (3.7)
XMeans (3.7)
sIB (3.7)

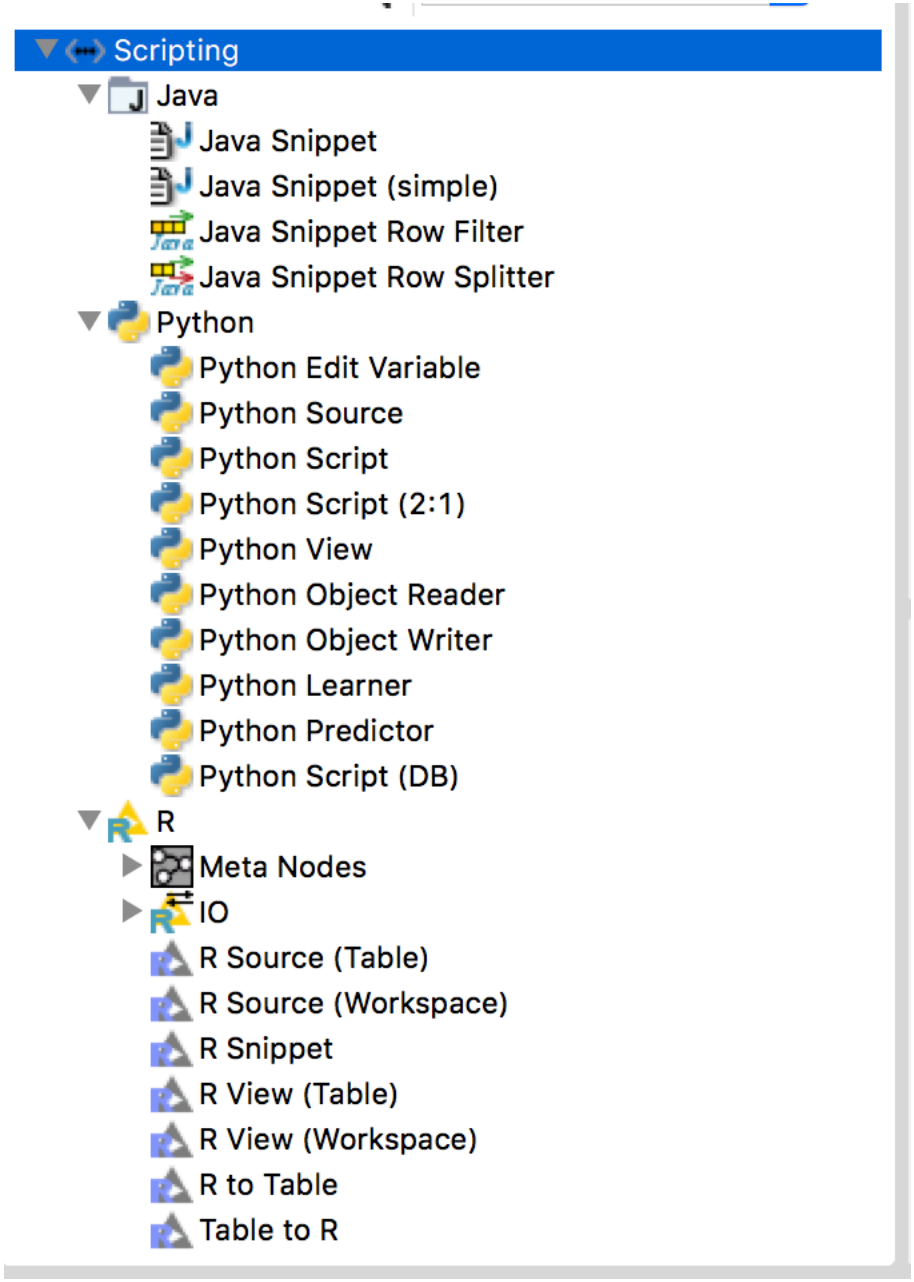
Weka



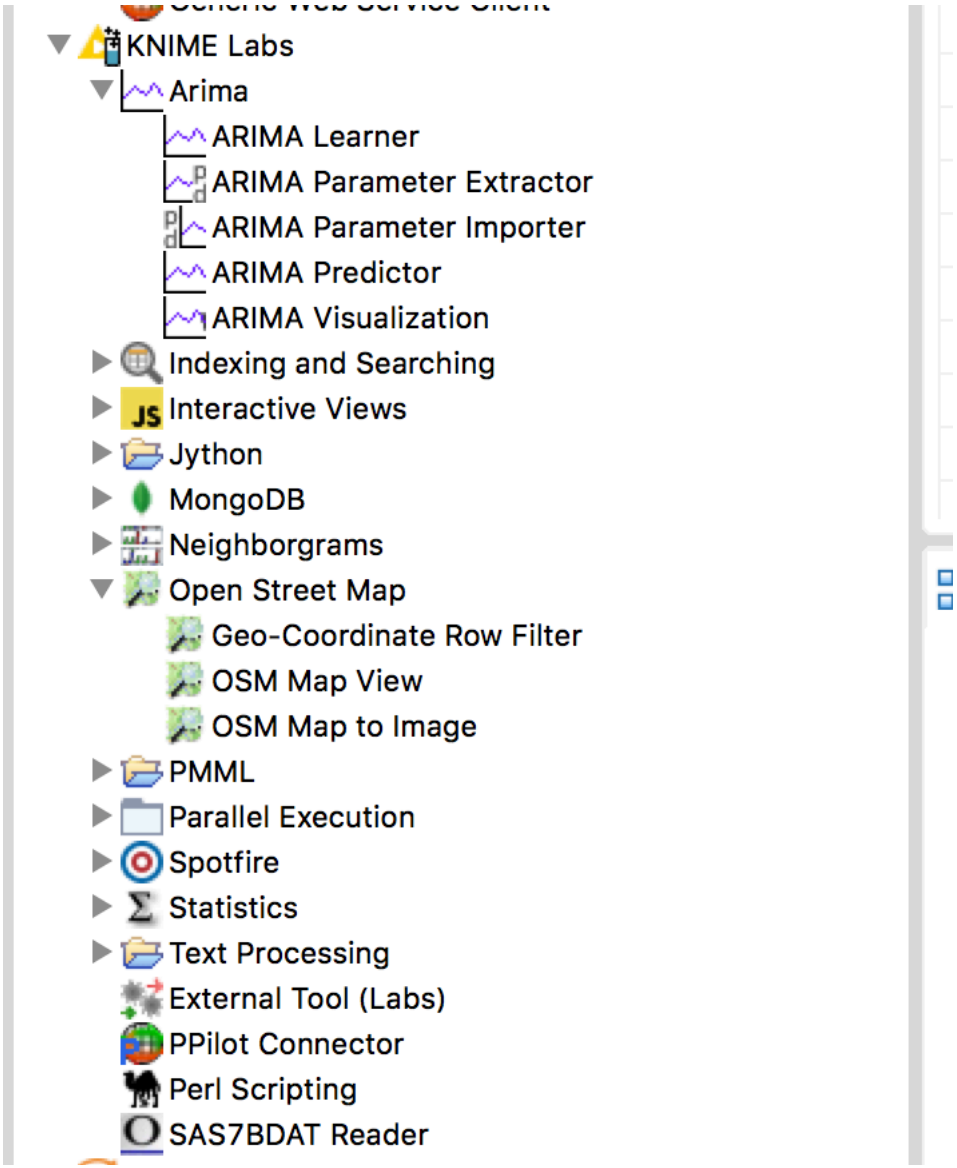
Statistics

Distance computing

Distance matrix

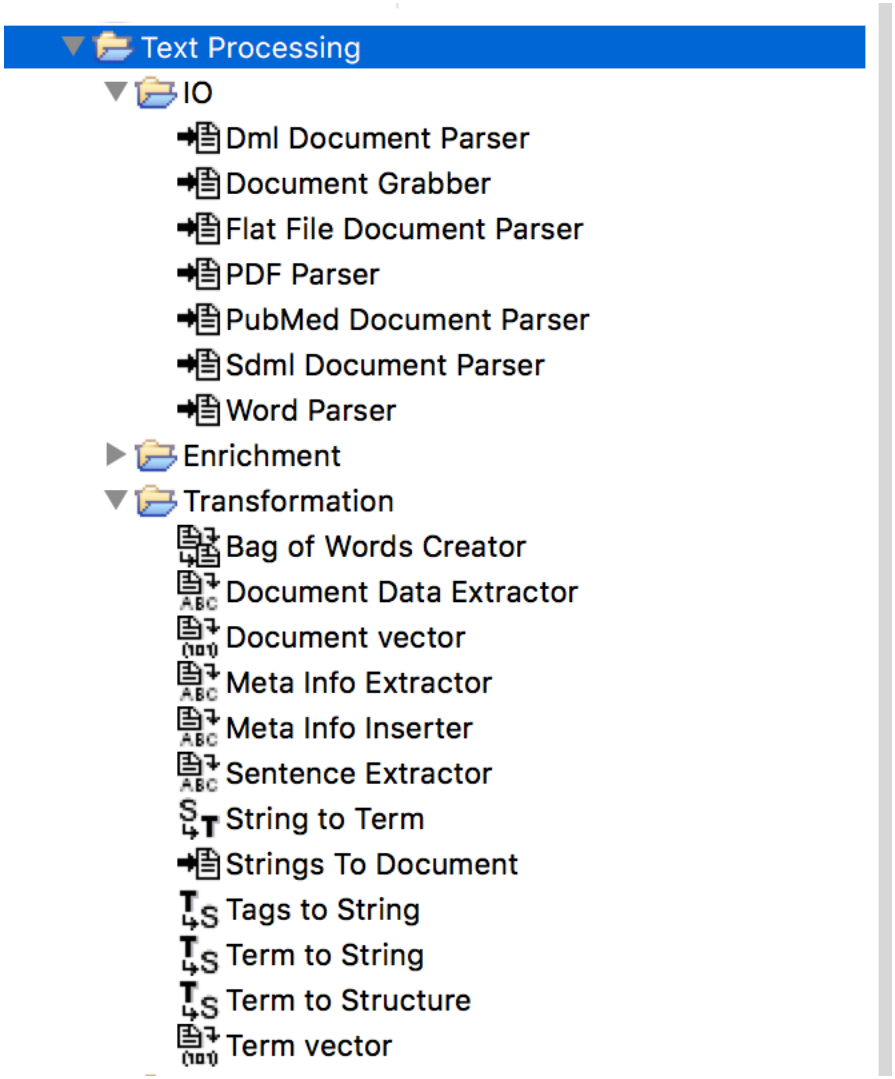

Scripting
Java
Java Snippet
Java Snippet (simple)
Java Snippet Row Filter
Java Snippet Row Splitter
Python
Python Edit Variable
Python Source
Python Script
Python Script (2:1)
Python View
Python Object Reader
Python Object Writer
Python Learner
Python Predictor
Python Script (DB)
R
Meta Nodes
IO
R Source (Table)
R Source (Workspace)
R Snippet
R View (Table)
R View (Workspace)
R to Table
Table to R

Scripting

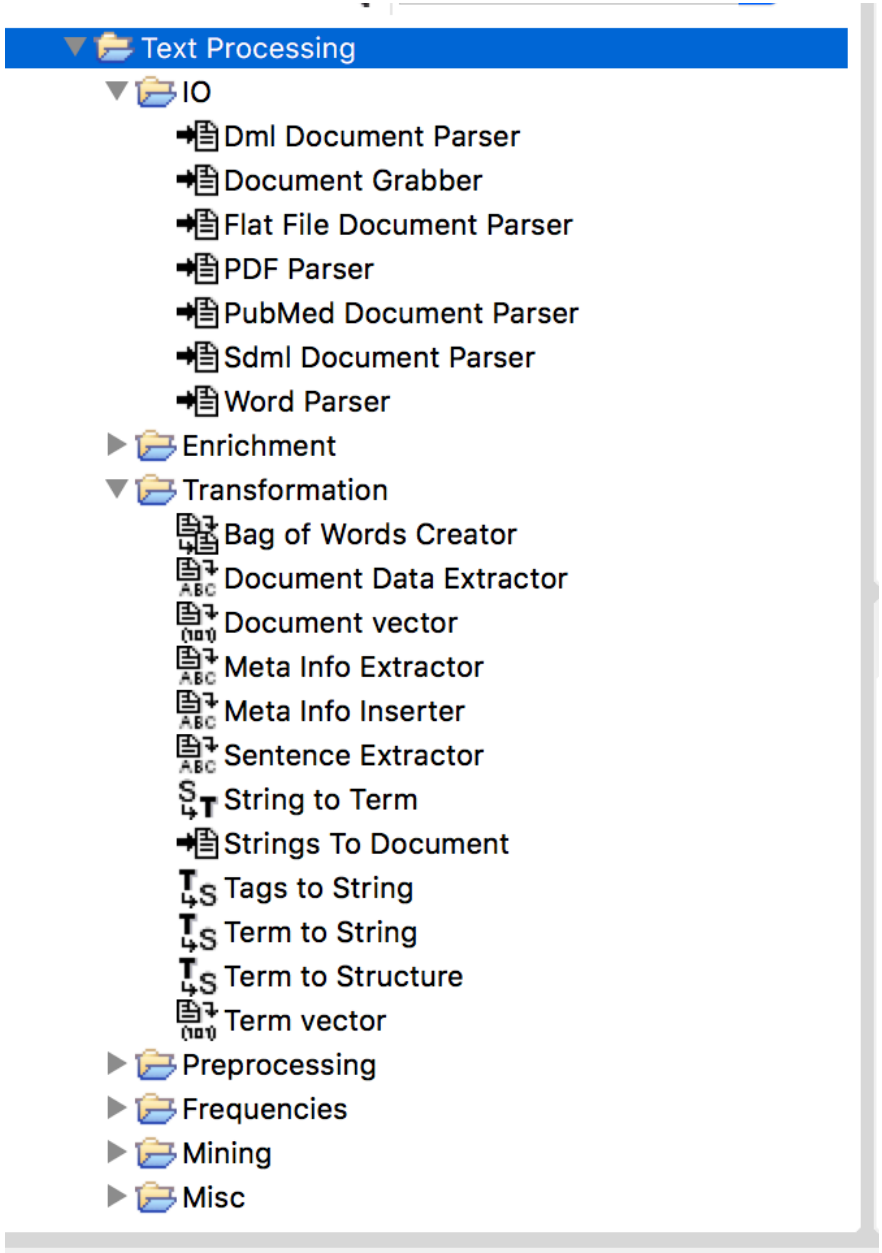

KNIME Labs
Arima
ARIMA Learner
ARIMA Parameter Extractor
ARIMA Parameter Importer
ARIMA Predictor
ARIMA Visualization
Indexing and Searching
Interactive Views
Jython
MongoDB
Neighborgrams
Open Street Map
Geo-Coordinate Row Filter
OSM Map View
OSM Map to Image
PMML
Parallel Execution
Spotfire
Statistics
Text Processing
External Tool (Labs)
PPilot Connector
Perl Scripting
SAS7BDAT Reader

Knime Labs

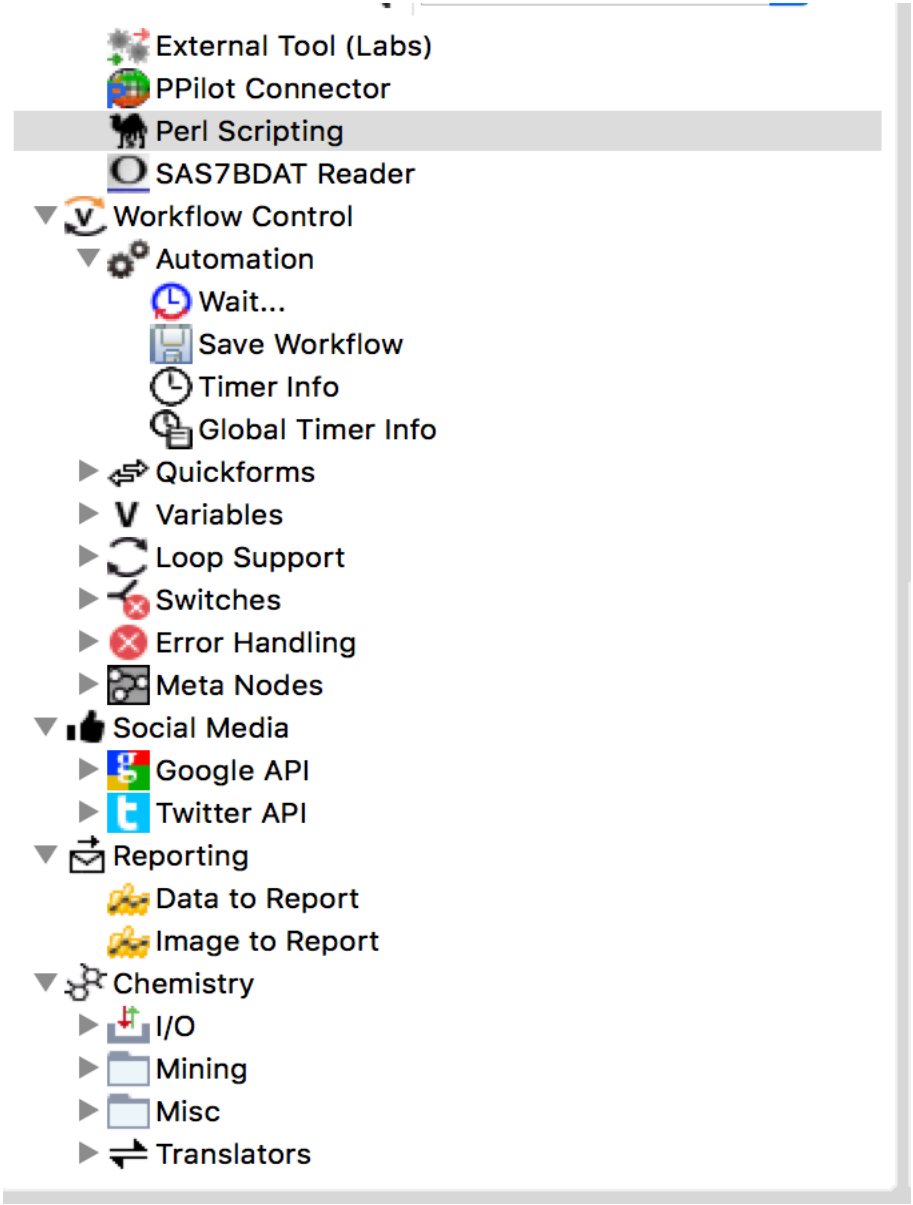
Text Processing (1)



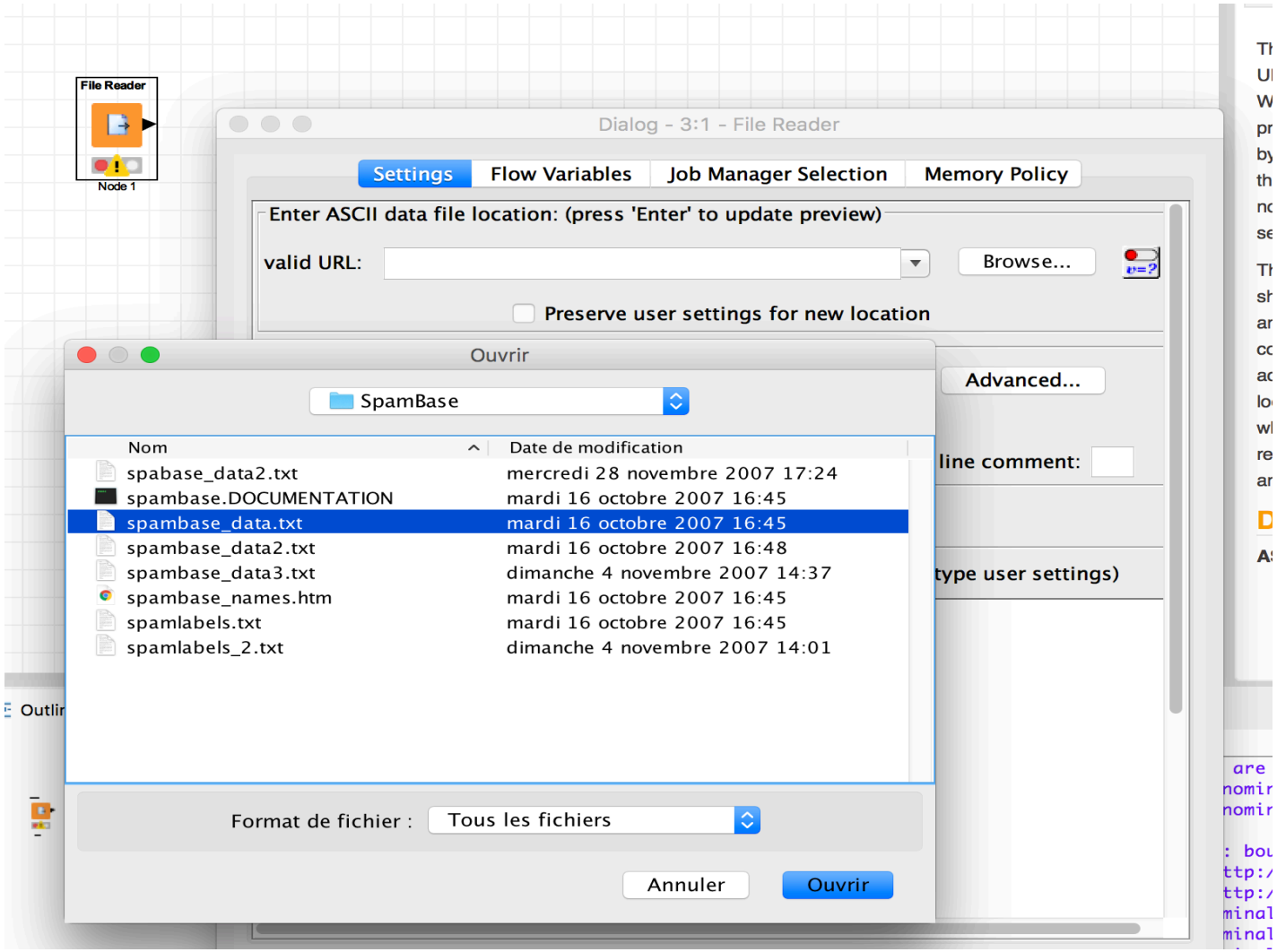
Text Processing (2)



Other utilities



File Reader


File Reader
Node 1
Dialog - 3:1 - File Reader
Settings
Flow Variables
Job Manager Selection
Memory Policy
Enter ASCII data file location: (press 'Enter' to update preview)
valid URL:
Browse...
Preserve user settings for new location
Advanced...
line comment:
type user settings)
Ouvrir
SpamBase
Nom
Date de modification
spabase_data2.txt
mercredi 28 novembre 2007 17:24
spambase.DOCUMENTATION
mardi 16 octobre 2007 16:45
spambase_data.txt
mardi 16 octobre 2007 16:45
spambase_data2.txt
mardi 16 octobre 2007 16:48
spambase_data3.txt
dimanche 4 novembre 2007 14:37
spambase_names.htm
mardi 16 octobre 2007 16:45
spamlabels.txt
mardi 16 octobre 2007 16:45
spamlabels_2.txt
dimanche 4 novembre 2007 14:01
Format de fichier :
Tous les fichiers
Annuler
Ouvrir

File Reader configuration

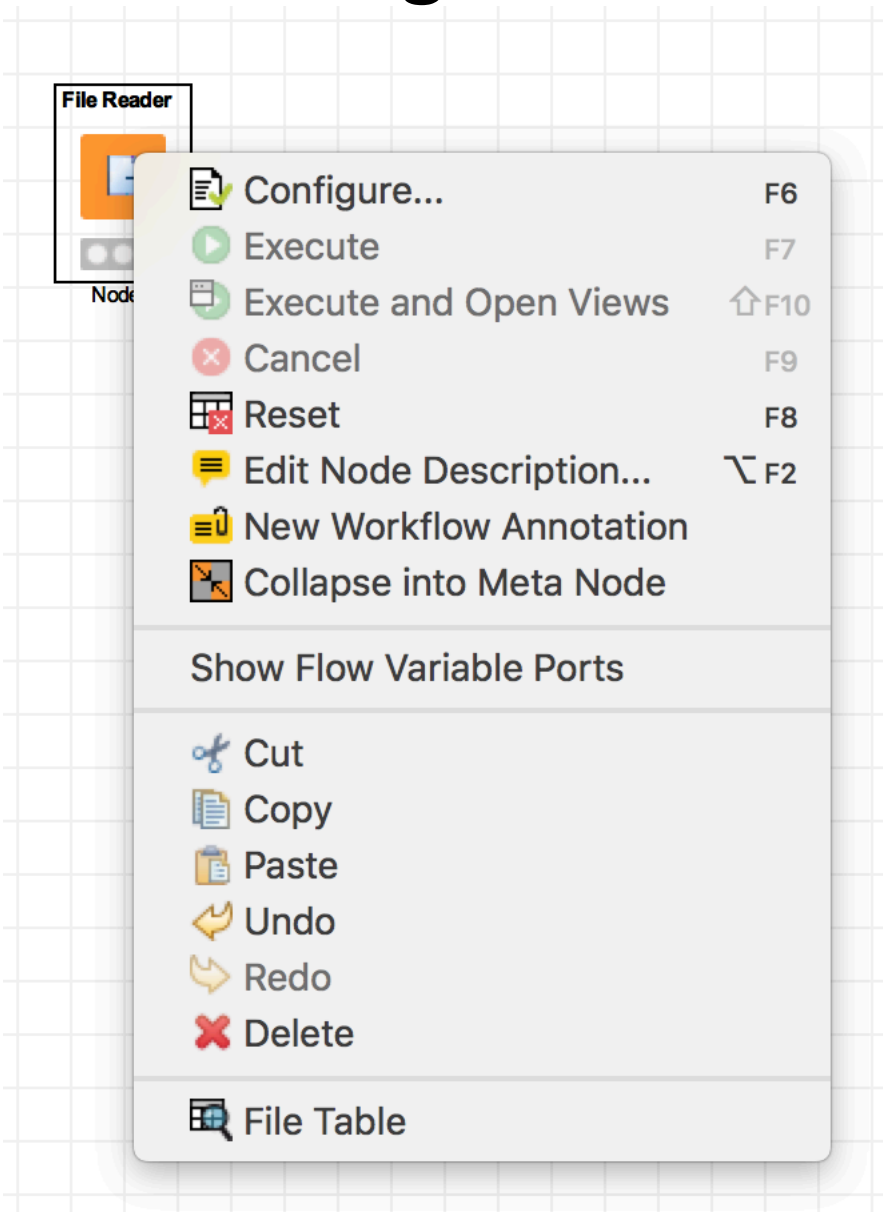
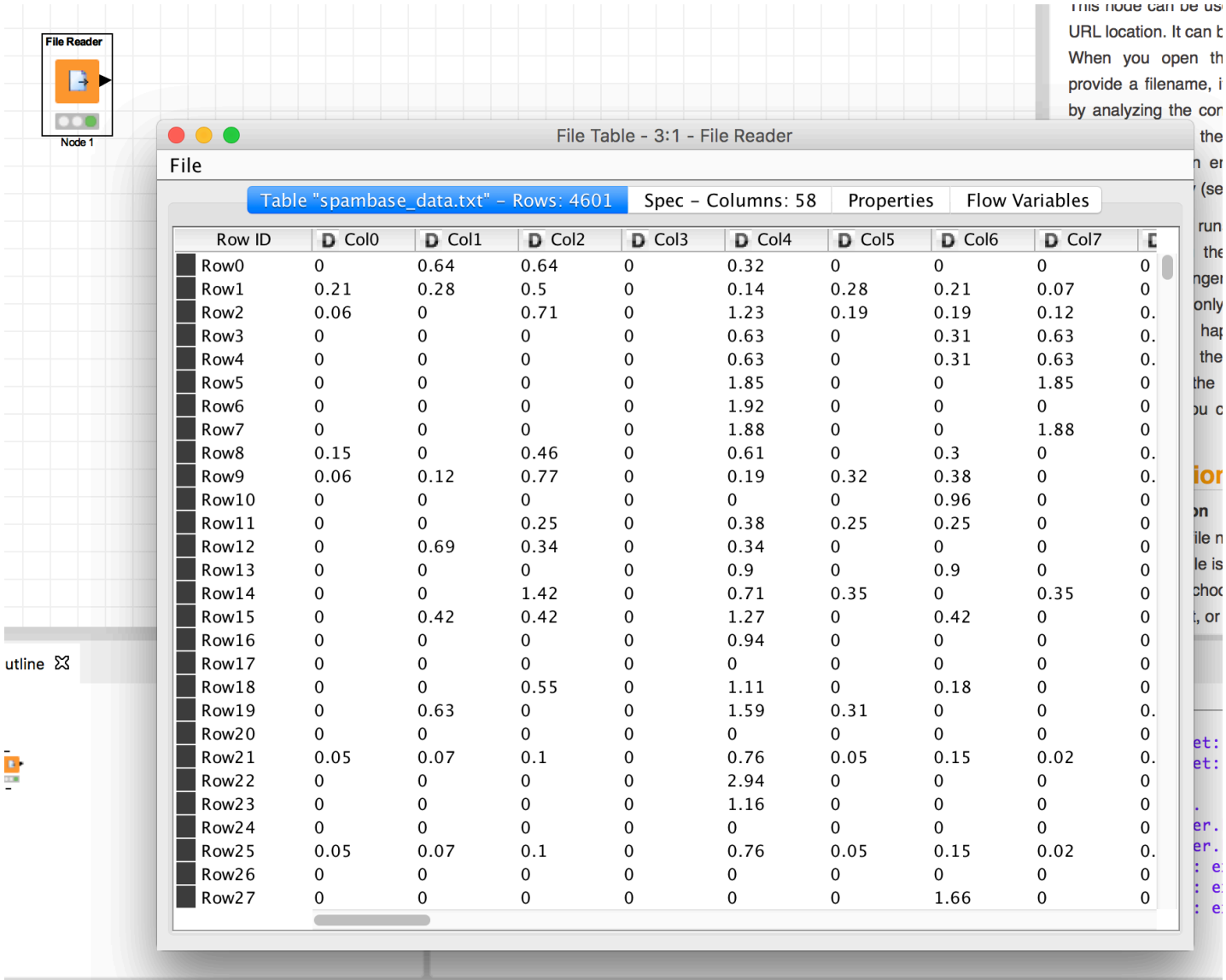
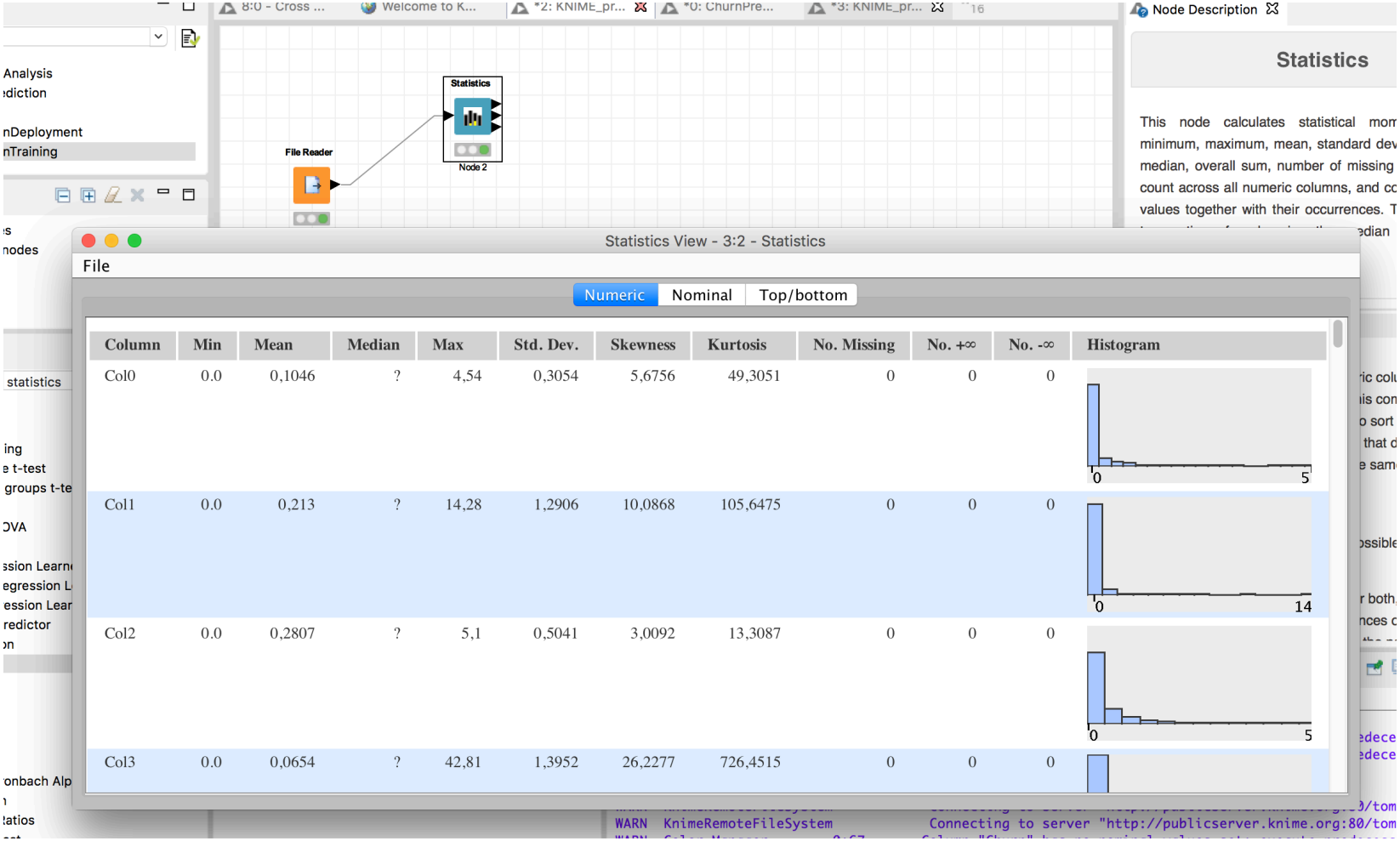


Table in File Reader

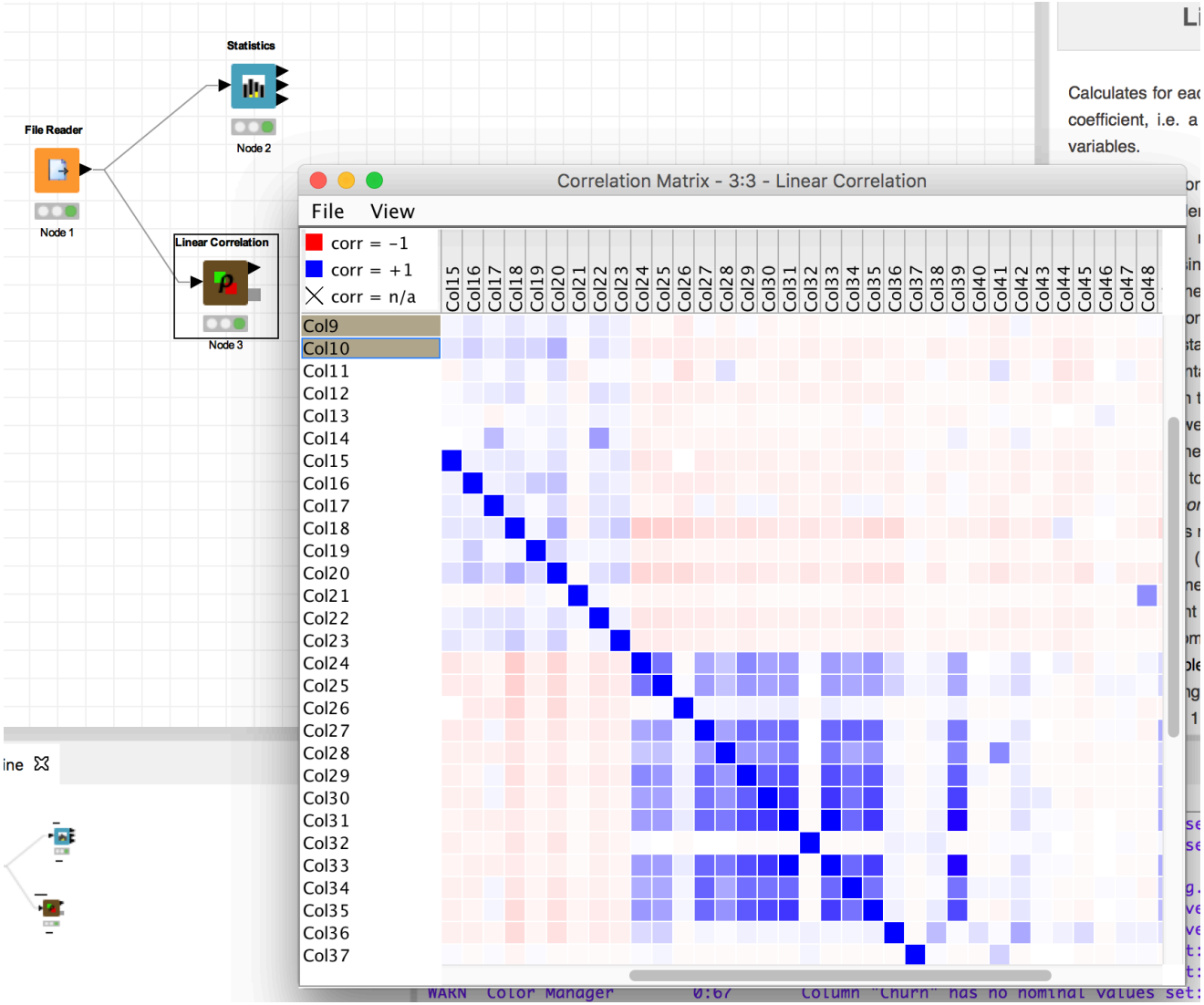


Row ID	Col0	Col1	Col2	Col3	Col4	Col5	Col6	Col7	
Row0	0	0.64	0.64	0	0.32	0	0	0	0
Row1	0.21	0.28	0.5	0	0.14	0.28	0.21	0.07	0
Row2	0.06	0	0.71	0	1.23	0.19	0.19	0.12	0.
Row3	0	0	0	0	0.63	0	0.31	0.63	0.
Row4	0	0	0	0	0.63	0	0.31	0.63	0.
Row5	0	0	0	0	1.85	0	0	1.85	0
Row6	0	0	0	0	1.92	0	0	0	0
Row7	0	0	0	0	1.88	0	0	1.88	0
Row8	0.15	0	0.46	0	0.61	0	0.3	0	0.
Row9	0.06	0.12	0.77	0	0.19	0.32	0.38	0	0.
Row10	0	0	0	0	0	0	0.96	0	0
Row11	0	0	0.25	0	0.38	0.25	0.25	0	0
Row12	0	0.69	0.34	0	0.34	0	0	0	0
Row13	0	0	0	0	0.9	0	0.9	0	0
Row14	0	0	1.42	0	0.71	0.35	0	0.35	0
Row15	0	0.42	0.42	0	1.27	0	0.42	0	0
Row16	0	0	0	0	0.94	0	0	0	0
Row17	0	0	0	0	0	0	0	0	0
Row18	0	0	0.55	0	1.11	0	0.18	0	0
Row19	0	0.63	0	0	1.59	0.31	0	0	0.
Row20	0	0	0	0	0	0	0	0	0
Row21	0.05	0.07	0.1	0	0.76	0.05	0.15	0.02	0.
Row22	0	0	0	0	2.94	0	0	0	0
Row23	0	0	0	0	1.16	0	0	0	0
Row24	0	0	0	0	0	0	0	0	0
Row25	0.05	0.07	0.1	0	0.76	0.05	0.15	0.02	0.
Row26	0	0	0	0	0	0	0	0	0
Row27	0	0	0	0	0	0	1.66	0	0

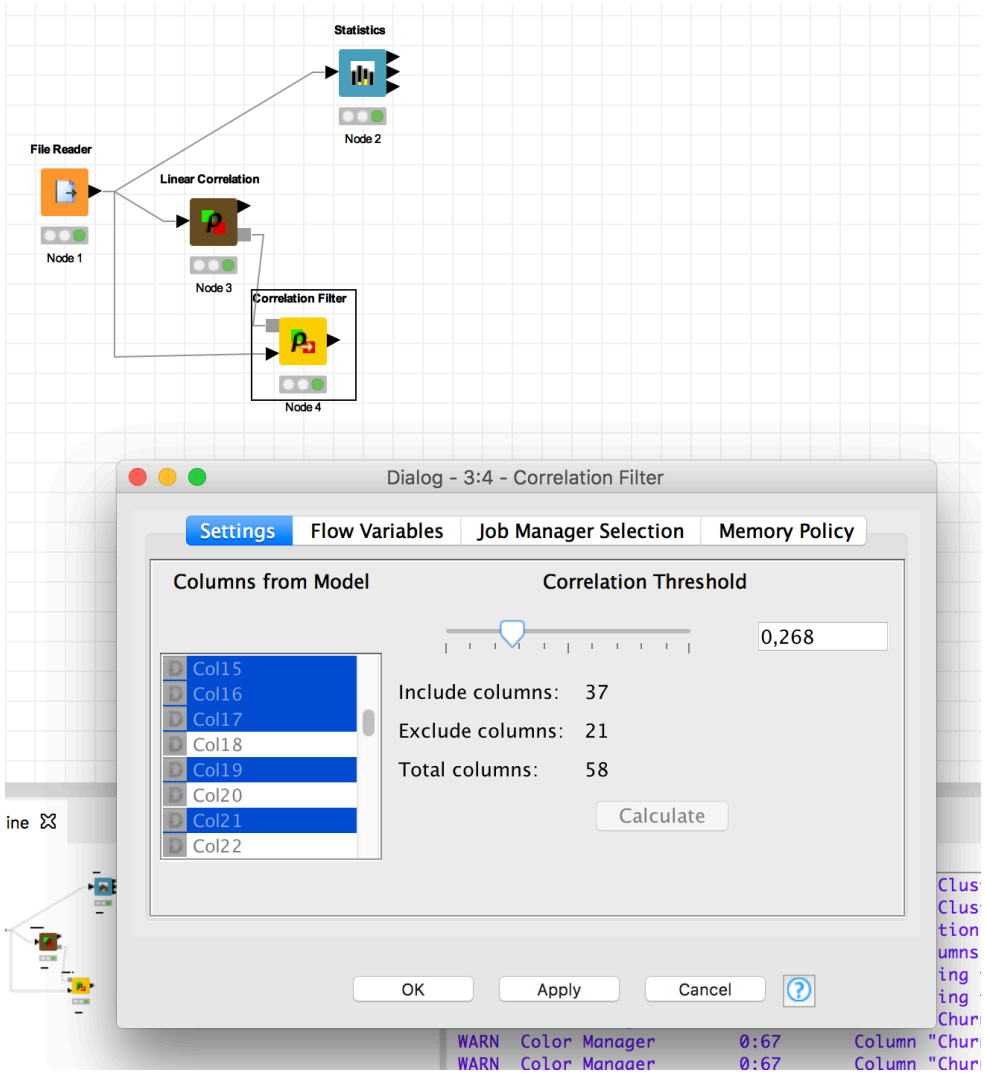
Statistics view



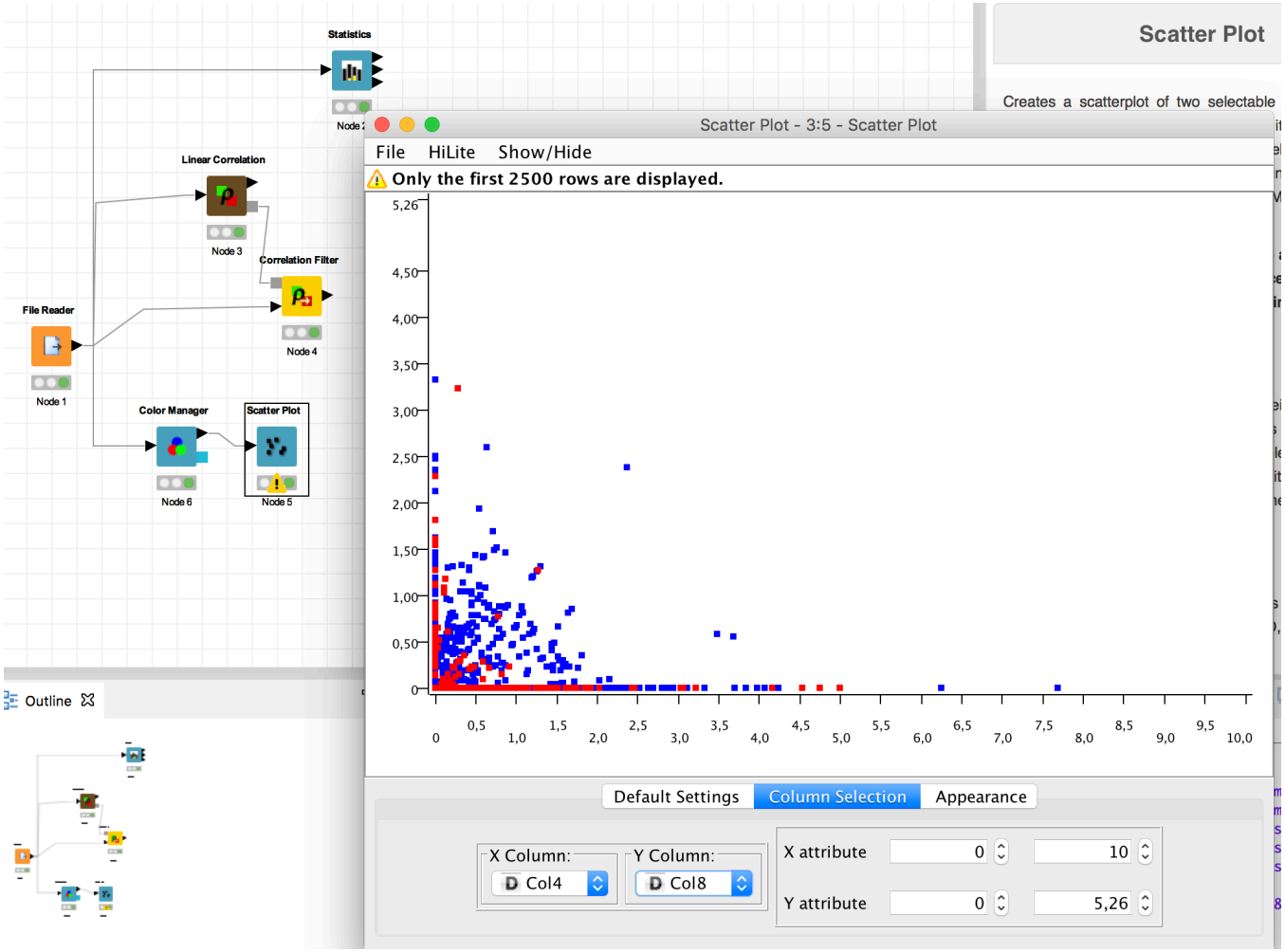
Correlation view



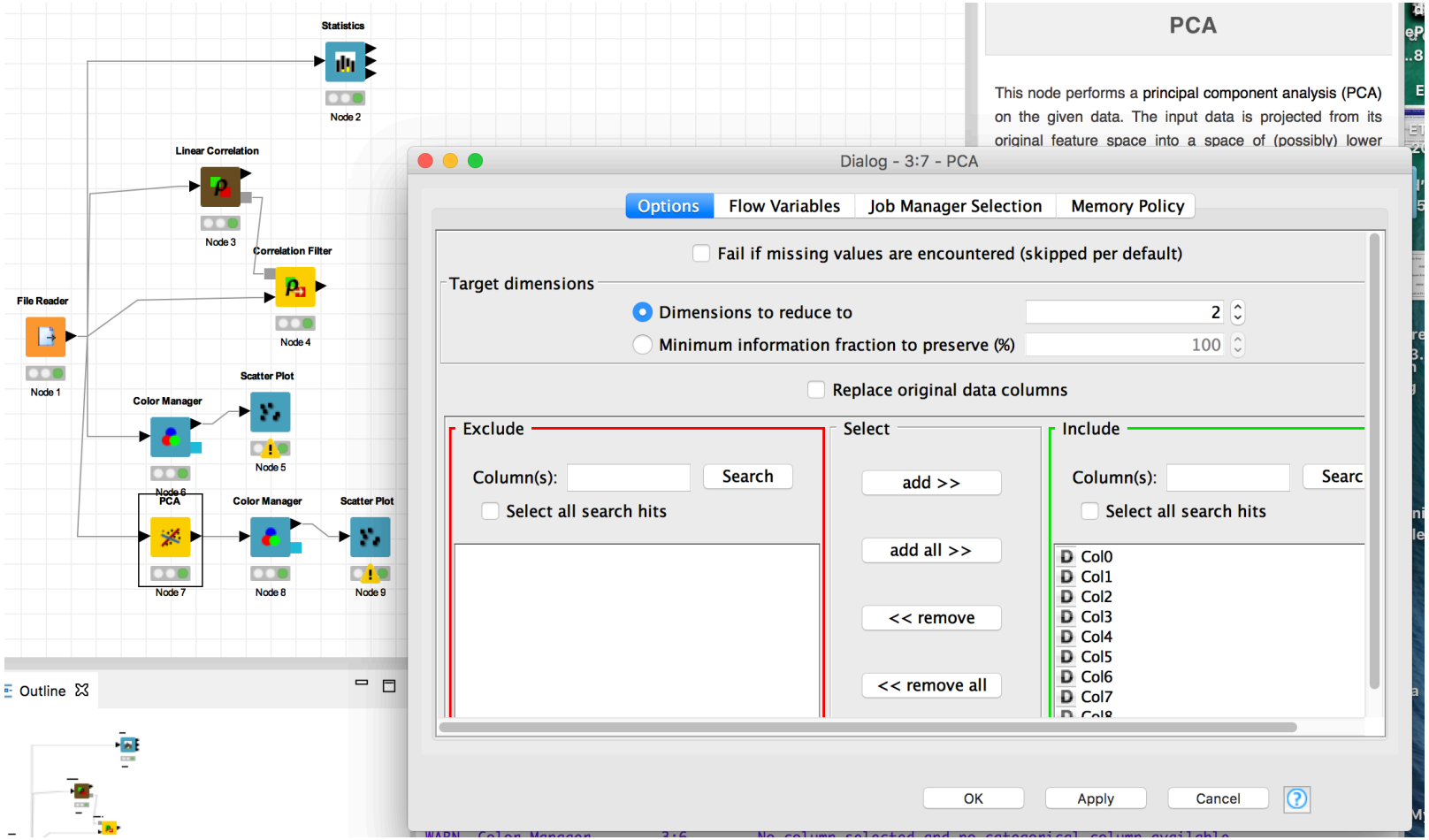
Correlation Filter



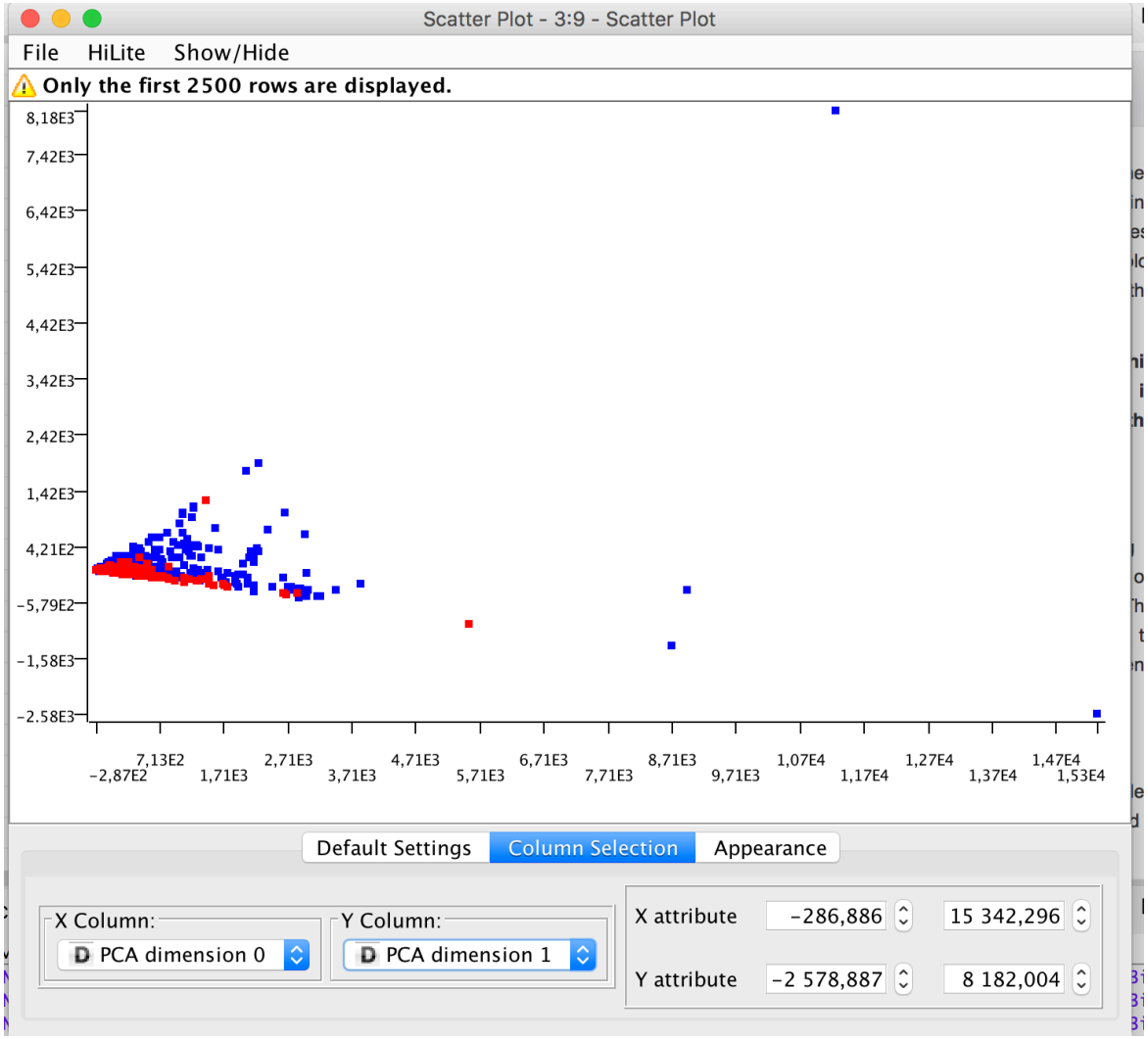
Visualization: Scatter Plot



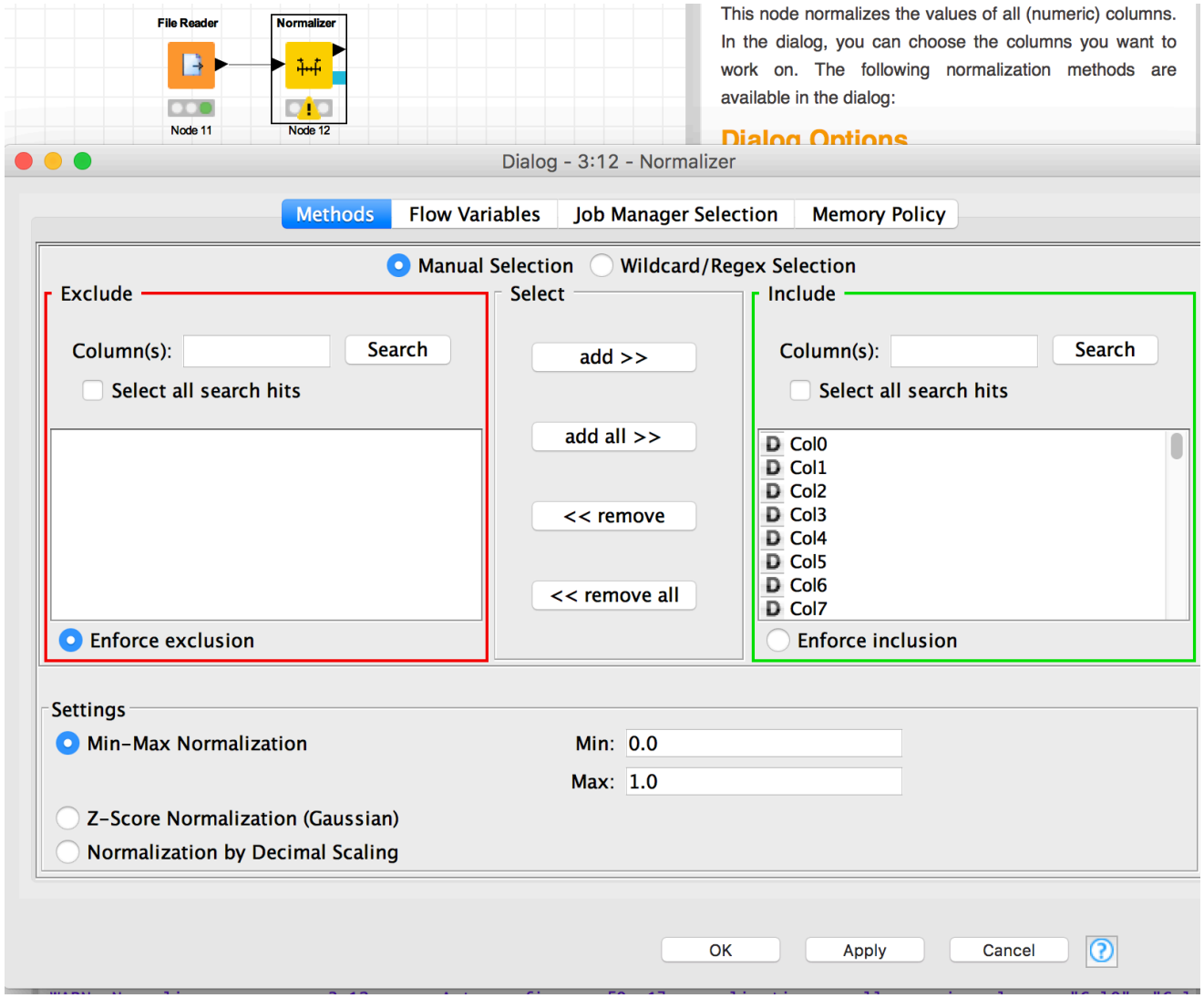
Principal Component Analysis (PCA)



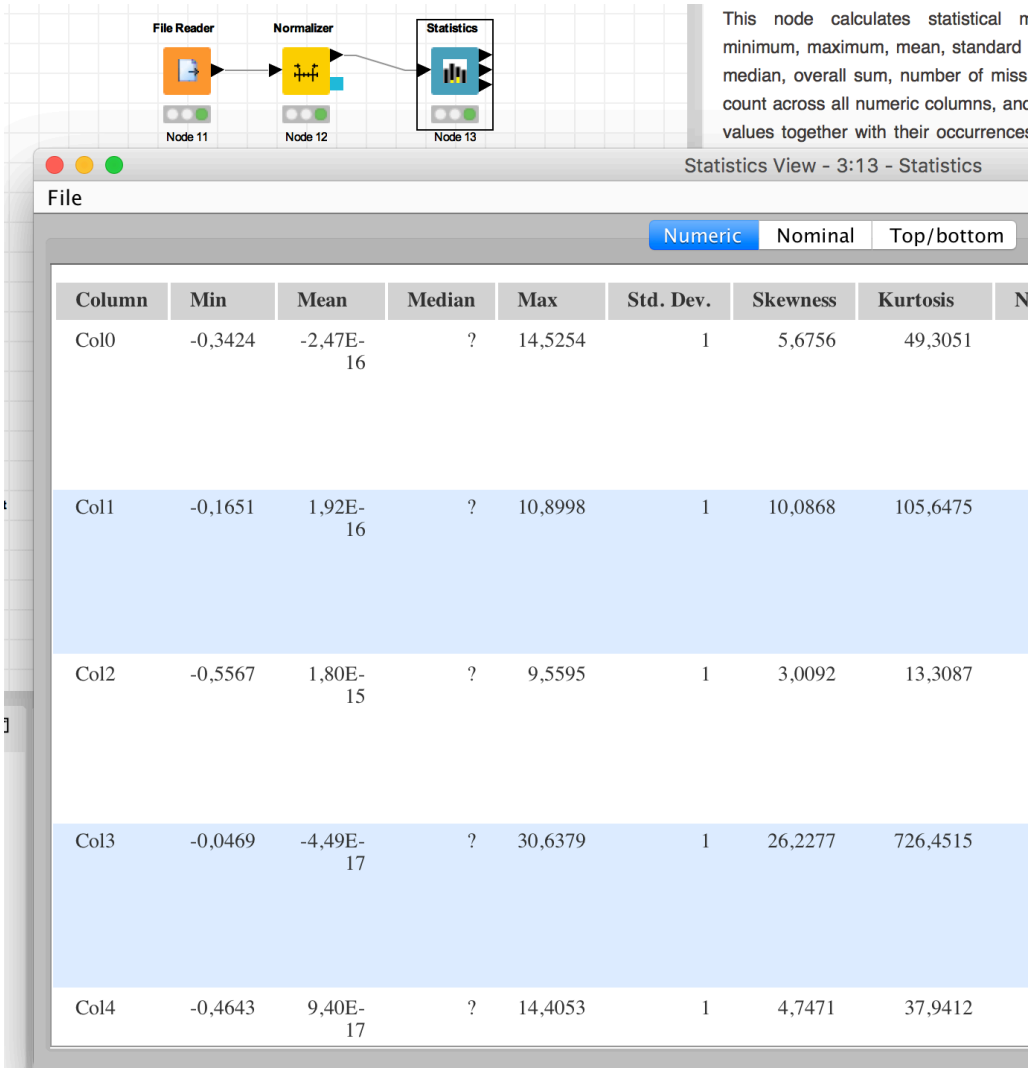
Project the data using PCA



Data normalization

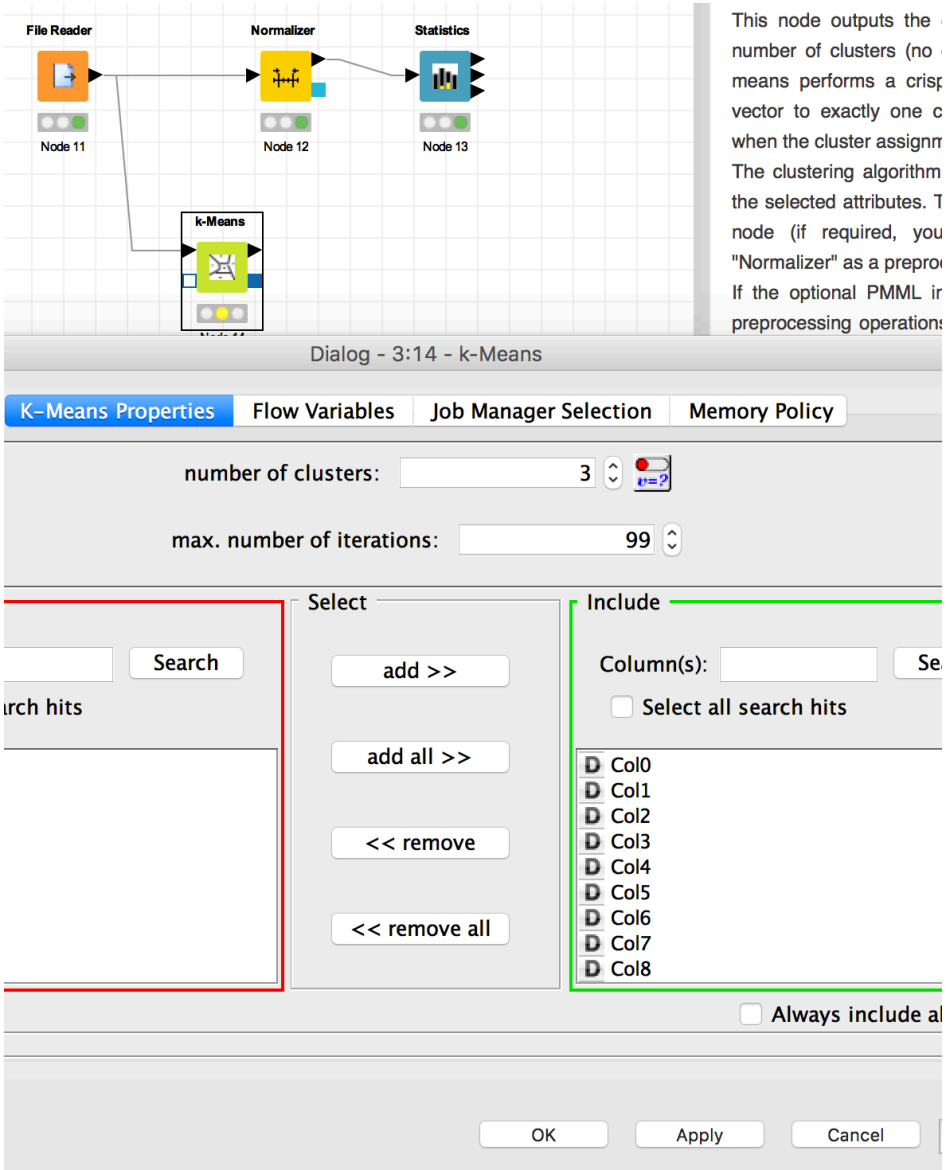


Data normalization

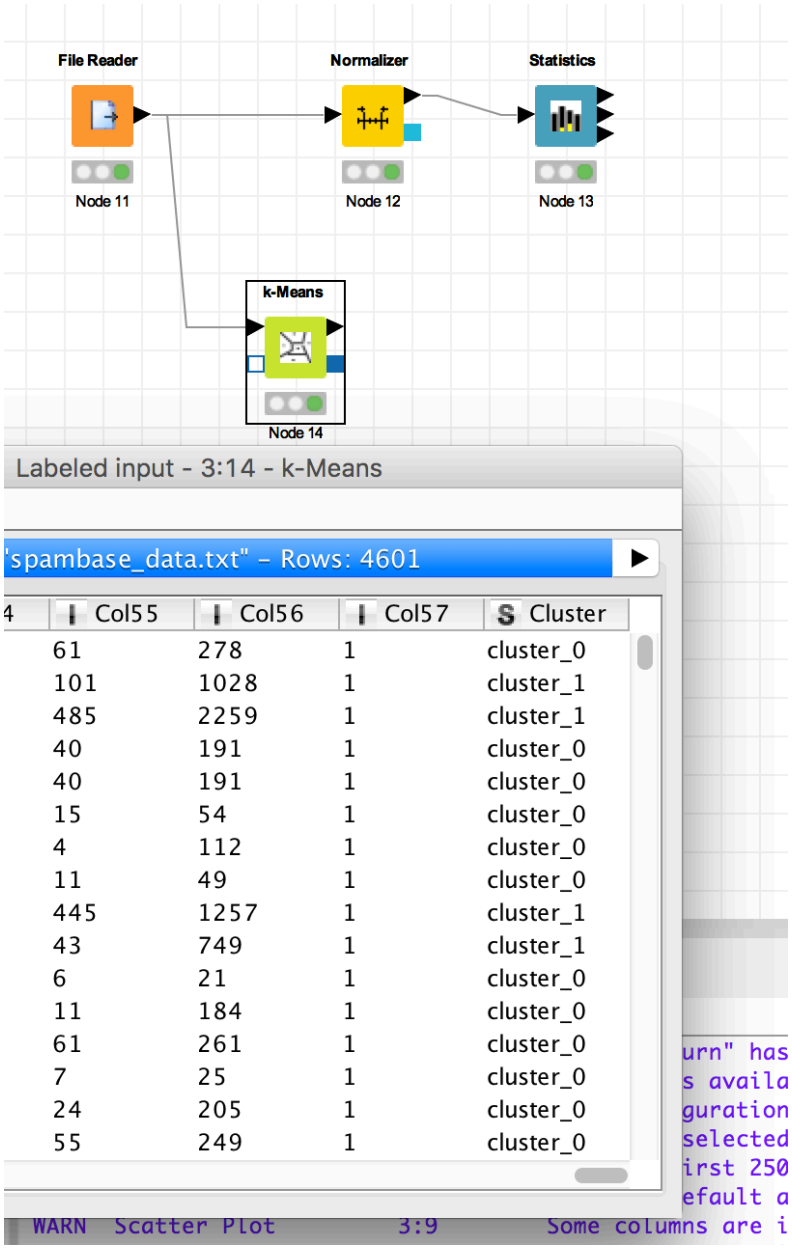


Column	Min	Mean	Median	Max	Std. Dev.	Skewness	Kurtosis	N
Col0	-0,3424	-2,47E-16	?	14,5254	1	5,6756	49,3051	
Col1	-0,1651	1,92E-16	?	10,8998	1	10,0868	105,6475	
Col2	-0,5567	1,80E-15	?	9,5595	1	3,0092	13,3087	
Col3	-0,0469	-4,49E-17	?	30,6379	1	26,2277	726,4515	
Col4	-0,4643	9,40E-17	?	14,4053	1	4,7471	37,9412	

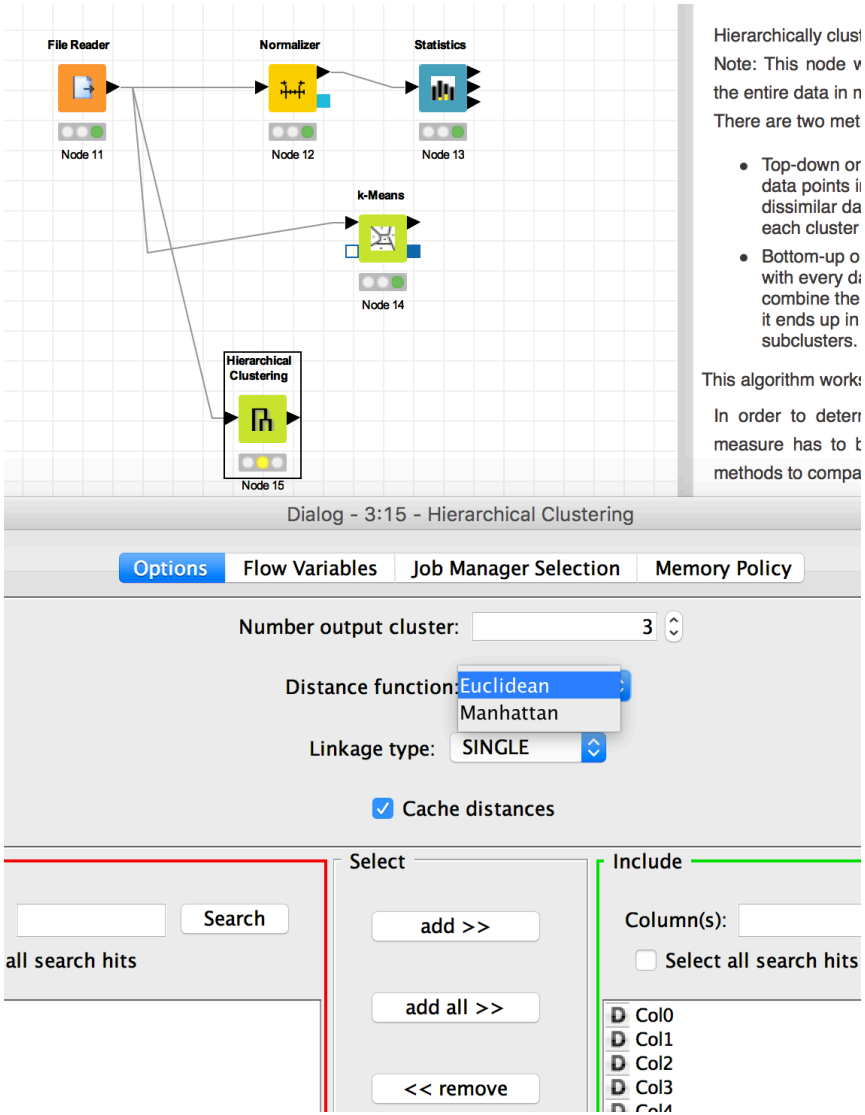
Clustering


File Reader
Normalizer
Statistics
Node 11
Node 12
Node 13
k-Means
This node outputs the
number of clusters (no
means performs a crisp
vector to exactly one c
when the cluster assignm
The clustering algorithm
the selected attributes. T
node (if required, you
"Normalizer" as a prepro
If the optional PMML ir
preprocessing operations
Dialog - 3:14 - k-Means
K-Means Properties
Flow Variables
Job Manager Selection
Memory Policy
number of clusters:
3
max. number of iterations:
99
Select
Include
Search
add >>
add all >>
<< remove
<< remove all
Column(s):
Select all search hits
Col0
Col1
Col2
Col3
Col4
Col5
Col6
Col7
Col8
OK
Apply
Cancel

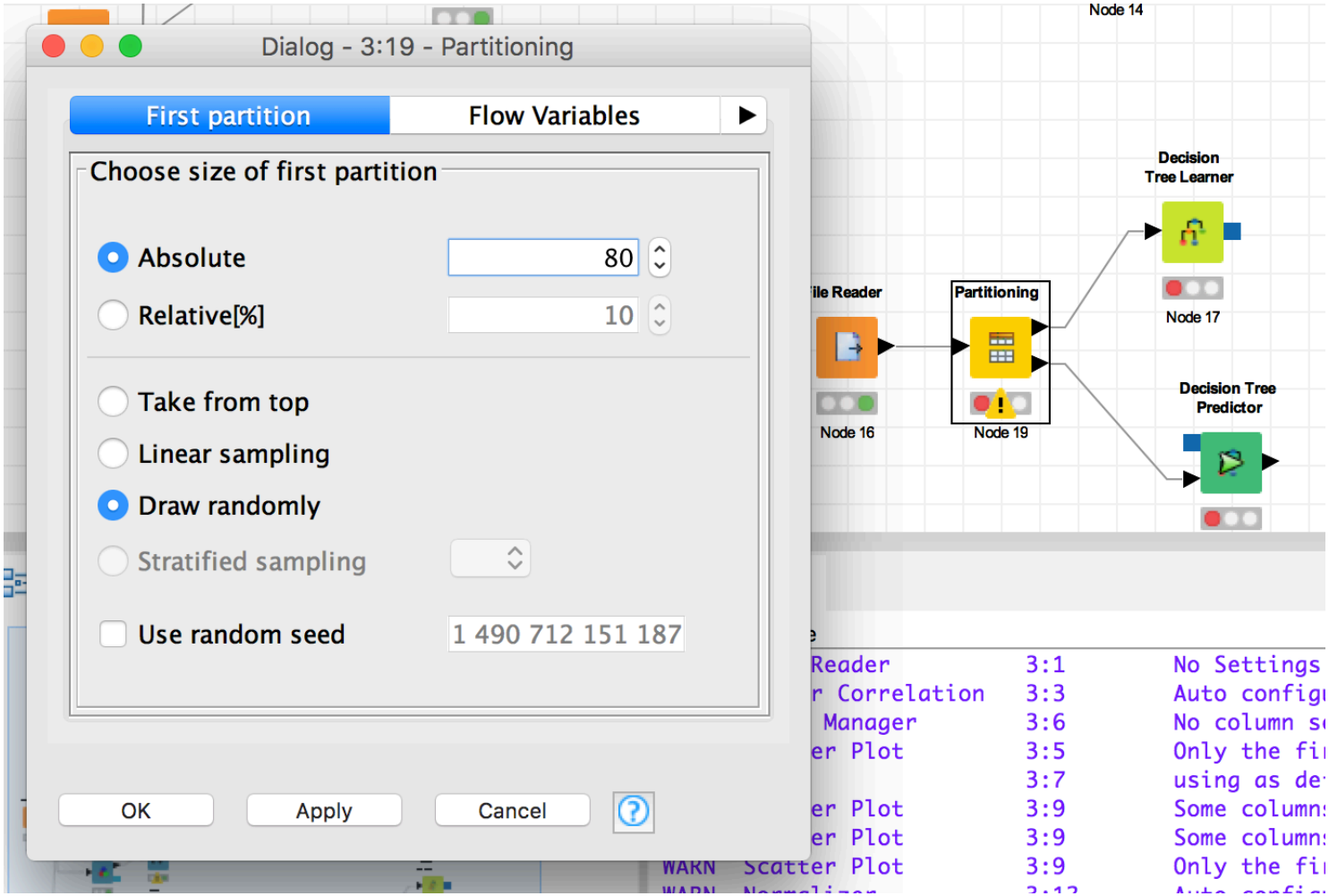
Clustering


File Reader
Normalizer
Statistics
Node 11
Node 12
Node 13
k-Means
Node 14
Labeled input - 3:14 - k-Means
"spambase_data.txt" - Rows: 4601
Col55 Col56 Col57 Cluster
61 278 1 cluster_0
101 1028 1 cluster_1
485 2259 1 cluster_1
40 191 1 cluster_0
40 191 1 cluster_0
15 54 1 cluster_0
4 112 1 cluster_0
11 49 1 cluster_0
445 1257 1 cluster_1
43 749 1 cluster_1
6 21 1 cluster_0
11 184 1 cluster_0
61 261 1 cluster_0
7 25 1 cluster_0
24 205 1 cluster_0
55 249 1 cluster_0
WARN Scatter Plot 3:9 Some columns are i

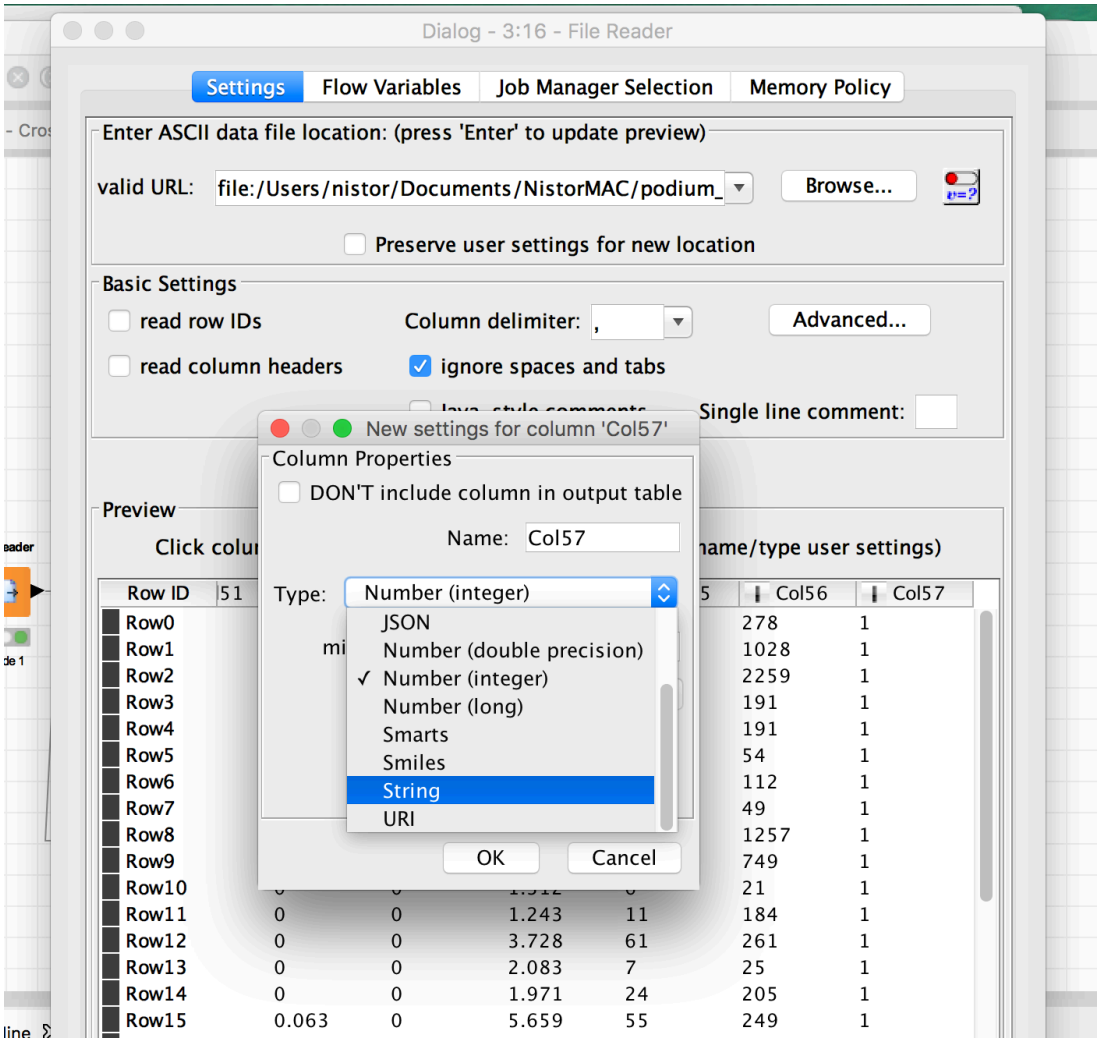
Clustering with Hierarchical clustering



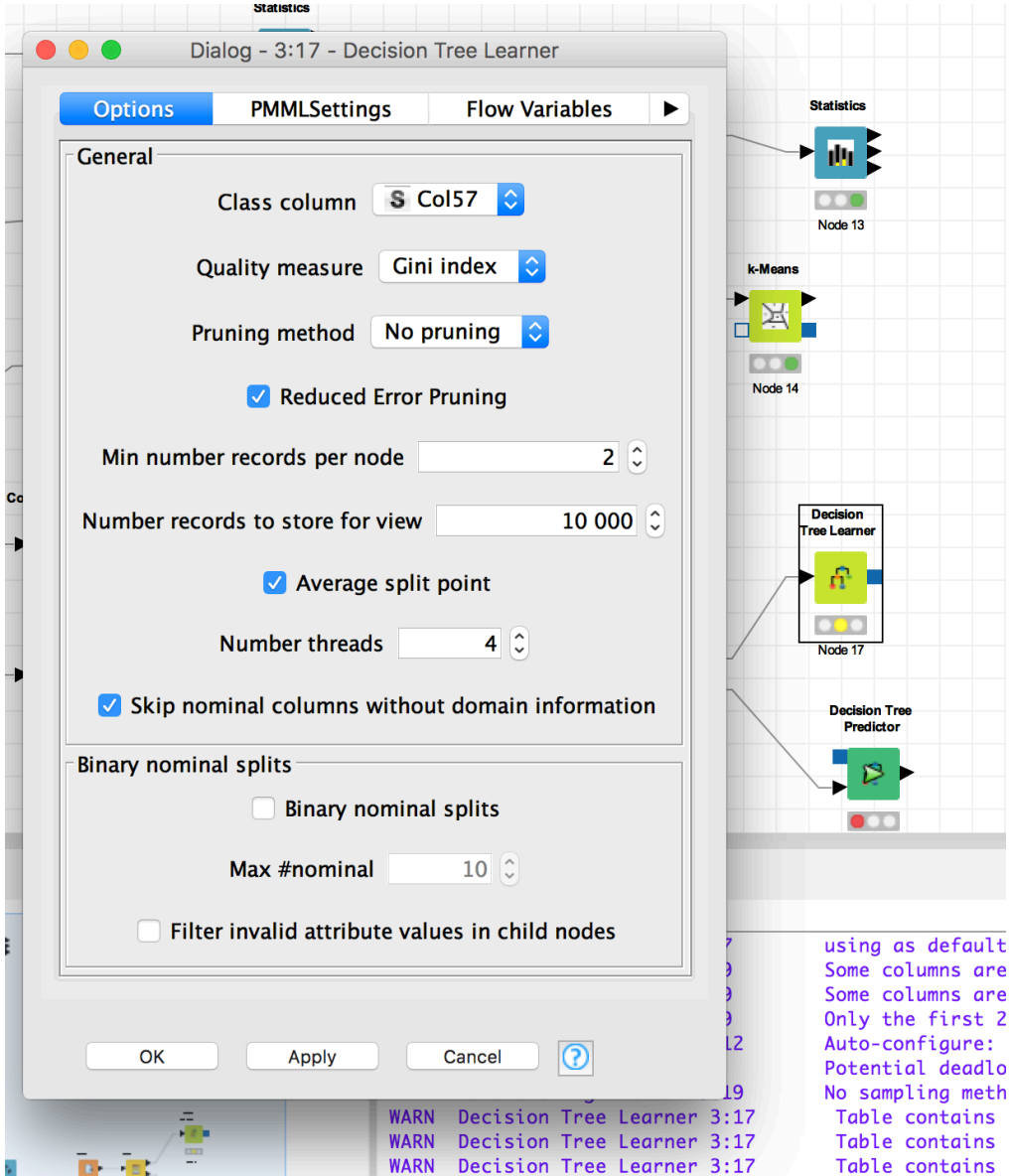
Machine Learning (1)



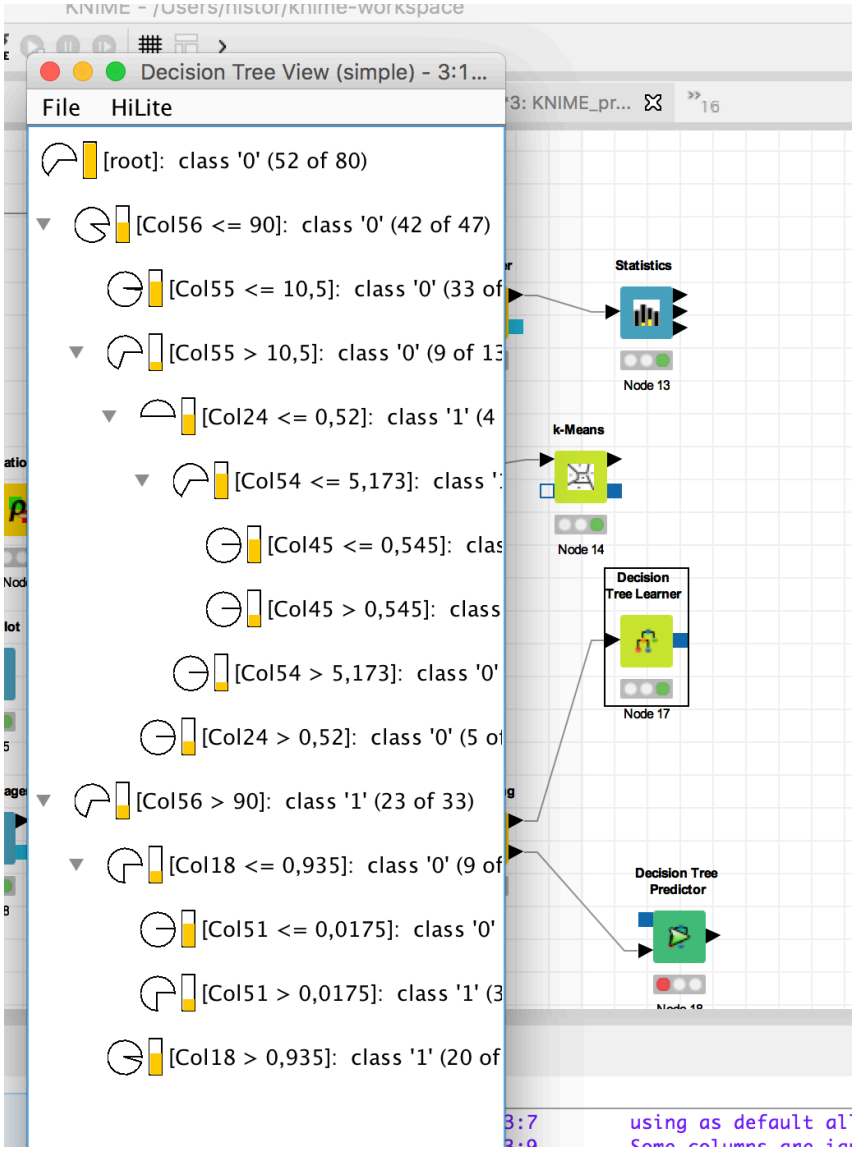
Machine Learning (2)



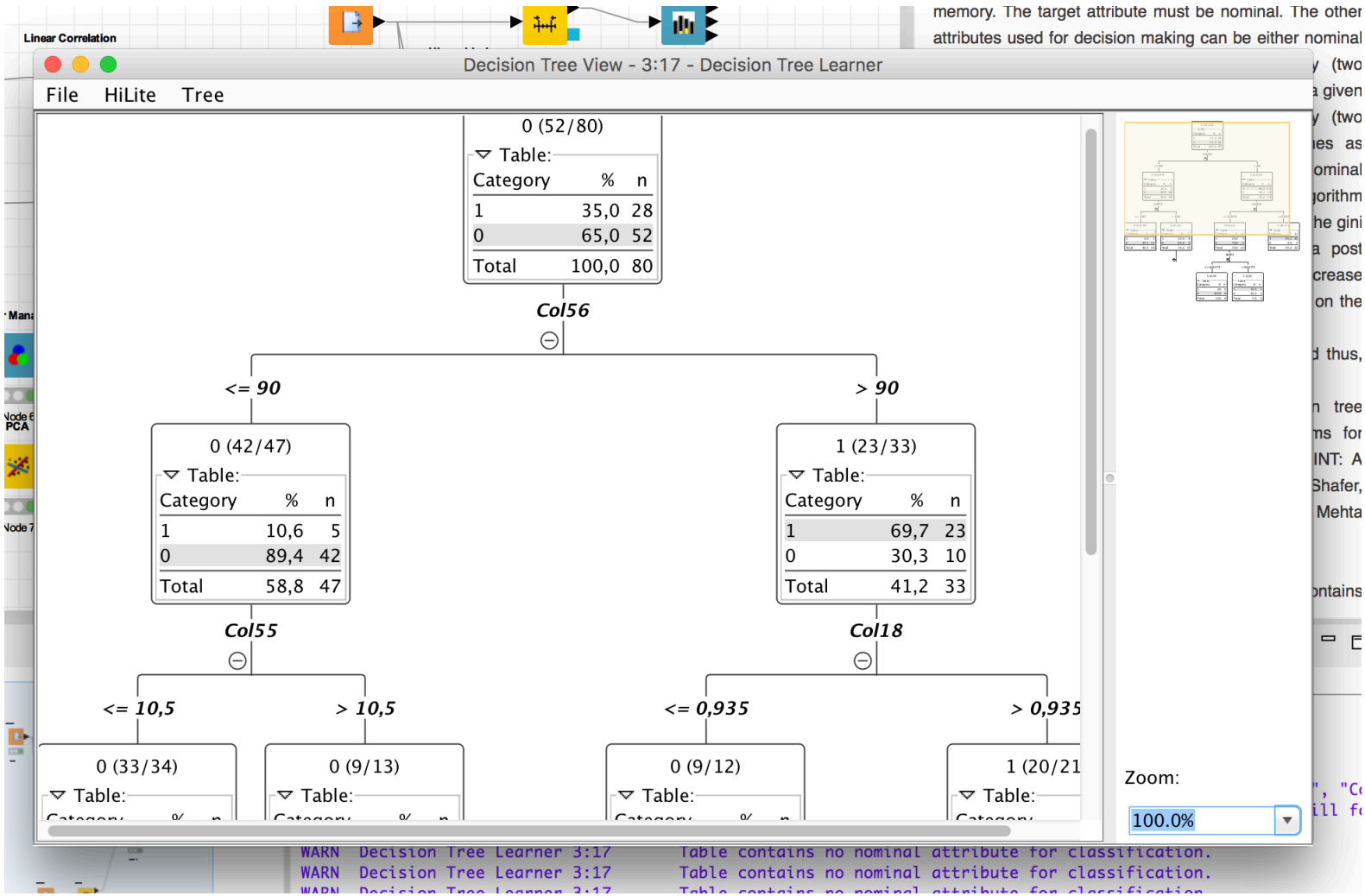
Machine Learning (3)



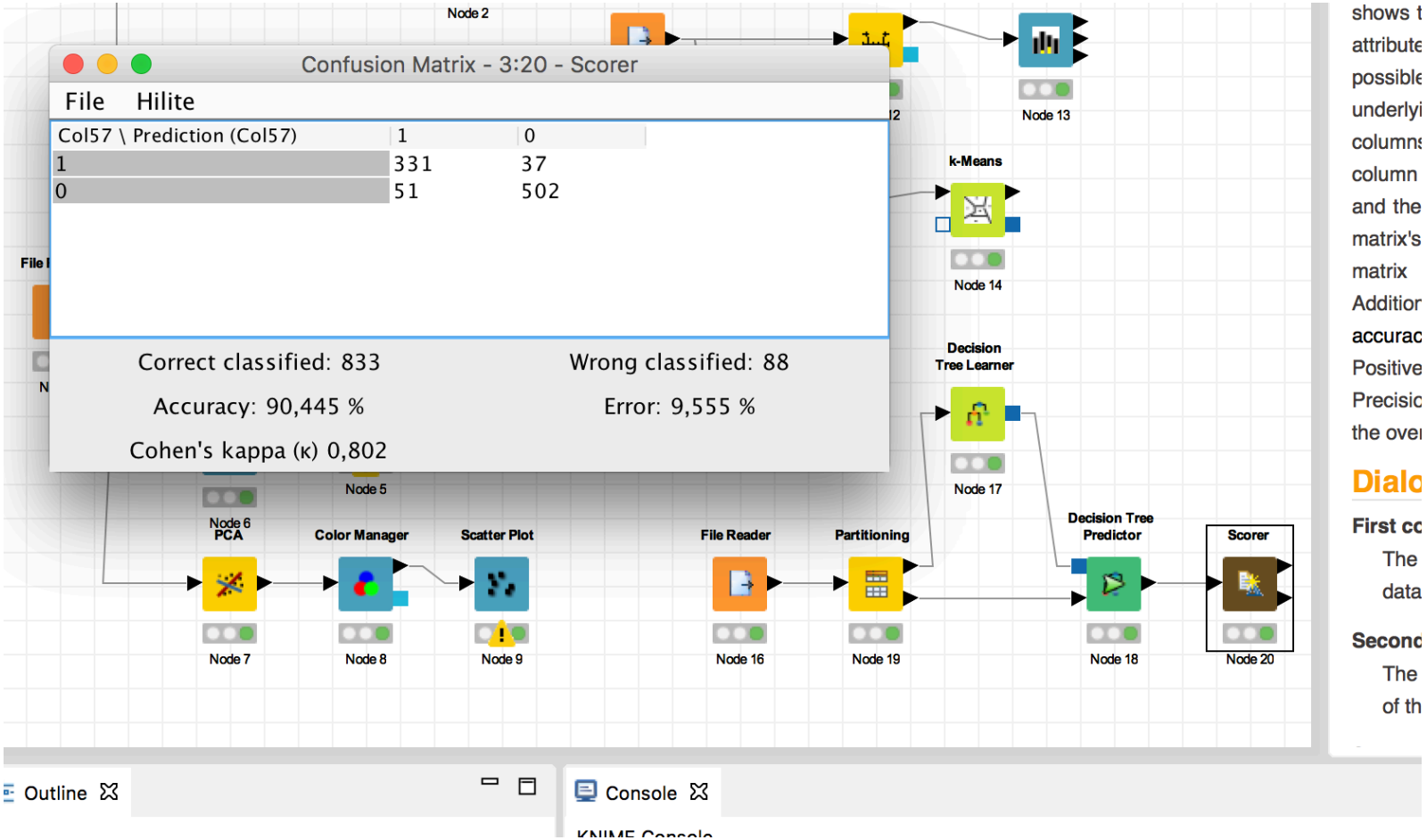
Machine Learning (4)



Machine Learning (5)



Machine Learning (6)



Machine Learning (7)