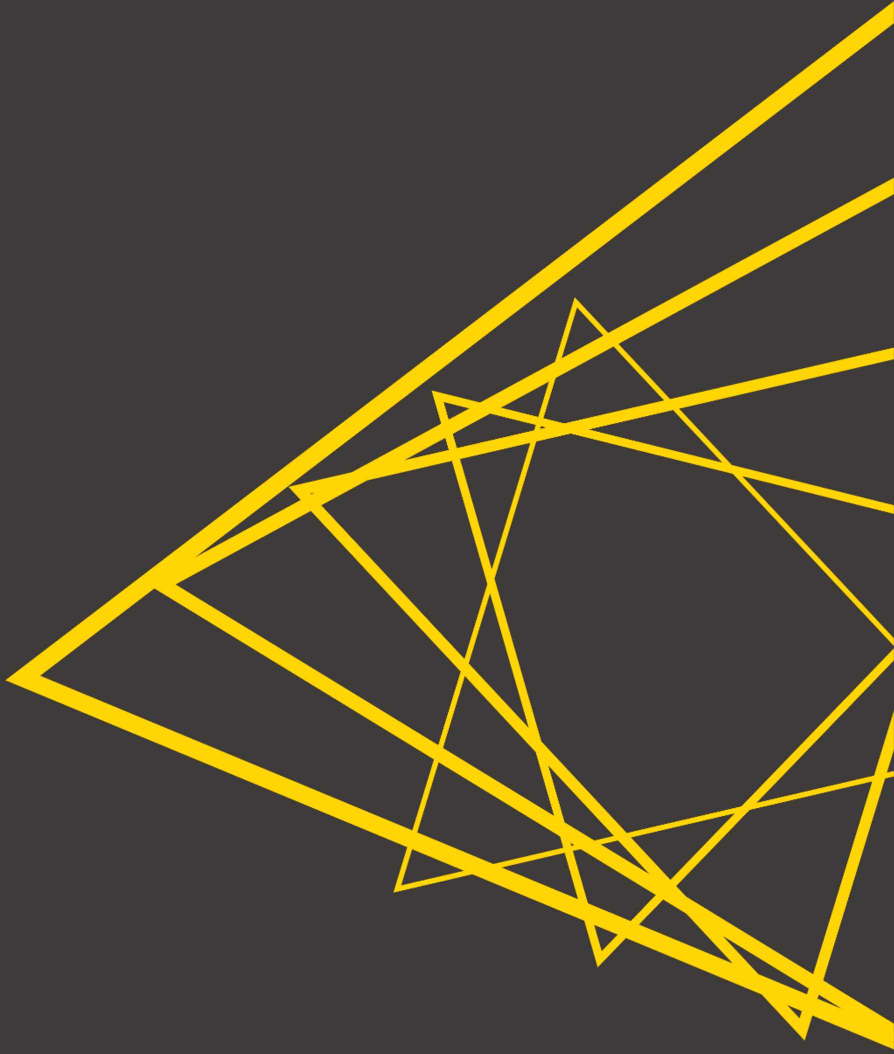
Boston Workshop

Monday, June 12 2023

Dr. Alice Krebs



Before we get started…

… some questions for you:

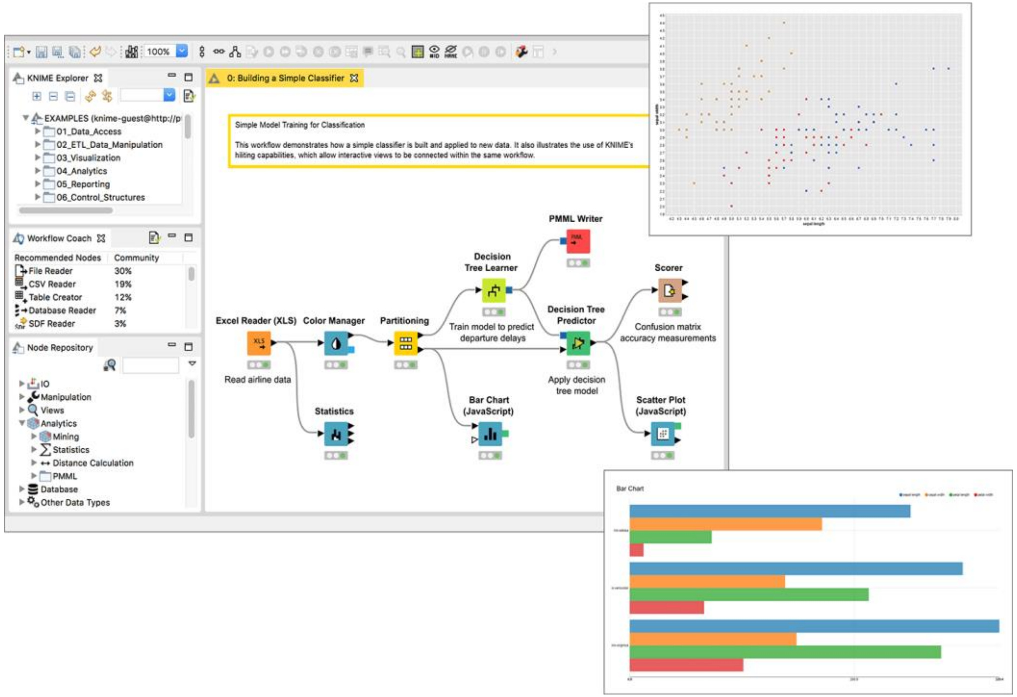
- What tools are you using for data analysis?
- What do you expect/hope to learn today?
- Did you manage to open the “00_Setup” workflow already?

© 2023 KNIME AG. All rights reserved.

Introduction to KNIME Analytics Platform

What is KNIME Analytics Platform?

- A tool for data analysis, manipulation, visualization, and reporting
- Based on the graphical programming paradigm
- Provides a diverse array of extensions:
    - Text Mining
    - Network Mining
    - Cheminformatics
    - Many integrations, such as Java, R, Python, Weka, Keras, Plotly, H2O, etc.



Recommended Nodes	Community
File Reader	30%
CSV Reader	19%
Table Creator	12%
Database Reader	7%
SDF Reader	3%

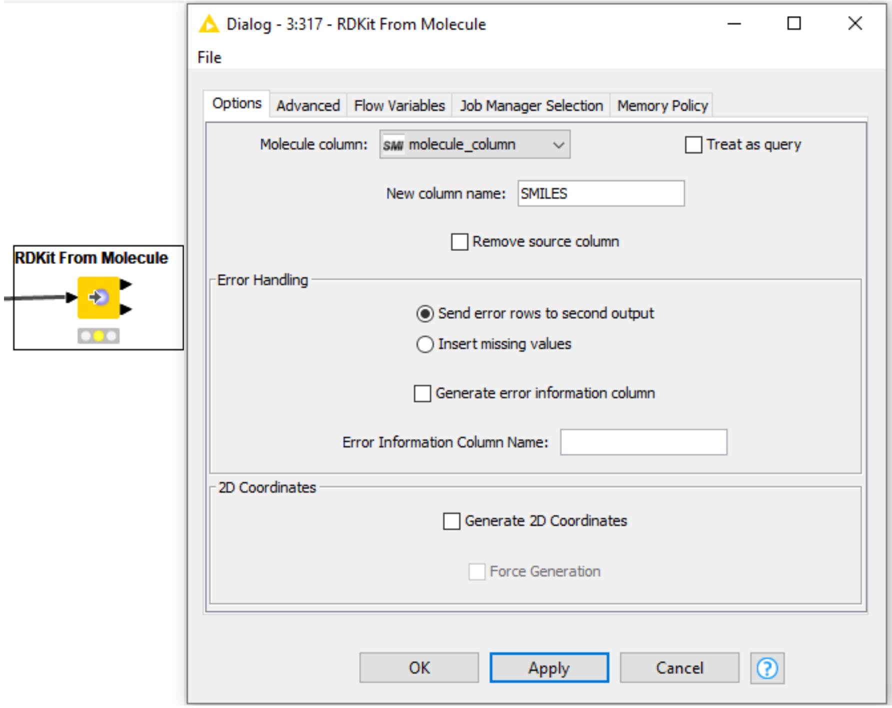
Simple Model Training for Classification

This workflow demonstrates how a simple classifier is built and applied to new data. It also illustrates the use of KNIME's hiliting capabilities, which allow interactive views to be connected within the same workflow.

© 2023 KNIME AG. All rights reserved.

Graphical programming paradigm?

- Lets users create programs by manipulating program elements graphically rather than by specifying them textually



RDKit From Molecule

Dialog - 3:317 - RDKit From Molecule

Molecule column: SMI molecule_column

☐ Treat as query

New column name: SMILES

☐ Remove source column

Error Handling

● Send error rows to second output

○ Insert missing values

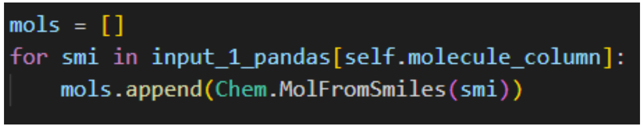
☐ Generate error information column

Error Information Column Name:

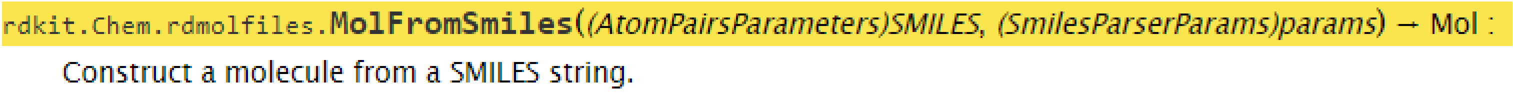
2D Coordinates

☐ Generate 2D Coordinates

☐ Force Generation



```
mols = []
for smi in input_1_pandas[self.molecule_column]:
    mols.append(Chem.MolFromSmiles(smi))
```



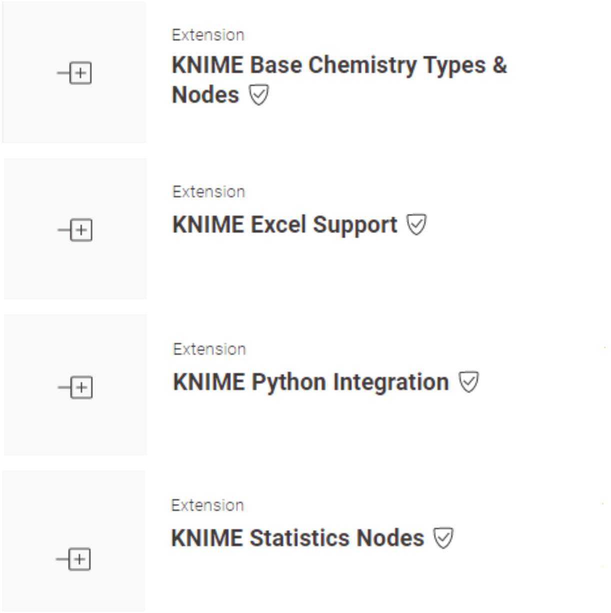
`rdkit.Chem.rdmolfiles.`**MolFromSmiles**(*(AtomPairsParameters)SMILES*, *(SmilesParserParams)params*) → Mol :

Construct a molecule from a SMILES string.

Extensions?

- KNIME AP comes by default only with basic functionality

- Extensions (and integrations) add specific functionality
- Not default, need to be added



Extension

KNIME Base Chemistry Types & Nodes

Extension

KNIME Excel Support

Extension

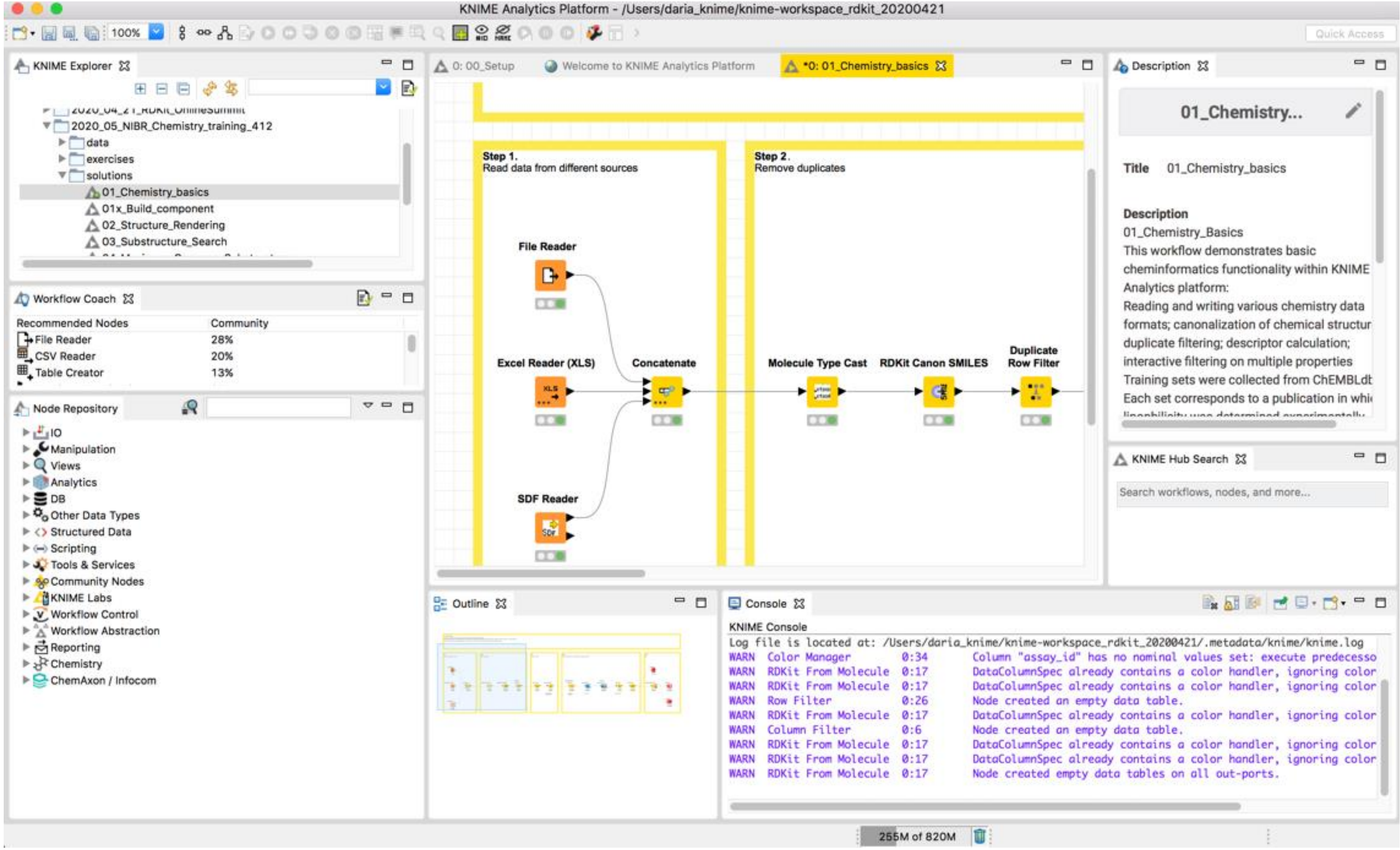
KNIME Python Integration

Extension

KNIME Statistics Nodes

© 2023 KNIME AG. All rights reserved.

The KNIME Workbench



Recommended Nodes	Community
File Reader	28%
CSV Reader	20%
Table Creator	13%

01_Chemistry...

Title 01_Chemistry_basics

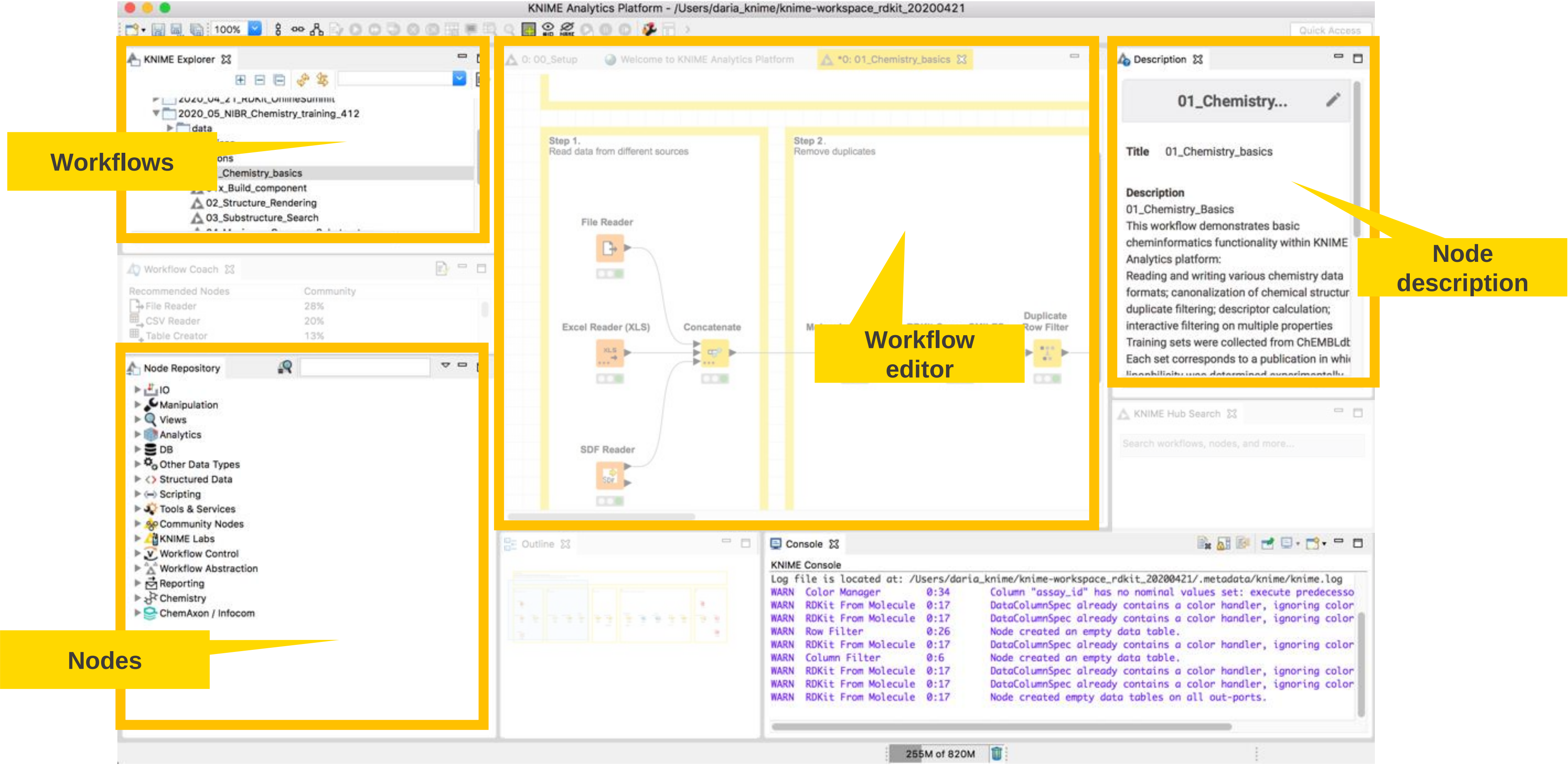
Description
01_Chemistry_Basics
This workflow demonstrates basic
cheminformatics functionality within KNIME
Analytics platform:
Reading and writing various chemistry data
formats; canonalization of chemical structur
duplicate filtering; descriptor calculation;
interactive filtering on multiple properties
Training sets were collected from ChEMBLdt
Each set corresponds to a publication in whi

KNIME Console

```
Log file is located at: /Users/daria_knime/knime-workspace_rdkit_20200421/.metadata/knime/knime.log
WARN  Color Manager         0:34      Column "assay_id" has no nominal values set: execute predecesso
WARN  RDKit From Molecule   0:17      DataColumnSpec already contains a color handler, ignoring color
WARN  RDKit From Molecule   0:17      DataColumnSpec already contains a color handler, ignoring color
WARN  Row Filter            0:26      Node created an empty data table.
WARN  RDKit From Molecule   0:17      DataColumnSpec already contains a color handler, ignoring color
WARN  Column Filter         0:6       Node created an empty data table.
WARN  RDKit From Molecule   0:17      DataColumnSpec already contains a color handler, ignoring color
WARN  RDKit From Molecule   0:17      DataColumnSpec already contains a color handler, ignoring color
WARN  RDKit From Molecule   0:17      Node created empty data tables on all out-ports.
```

© 2023 KNIME AG. All rights reserved.

The KNIME Workbench – most important windows



Workflows

Workflow editor

Node description

Nodes

01_Chemistry...

Title 01_Chemistry_basics

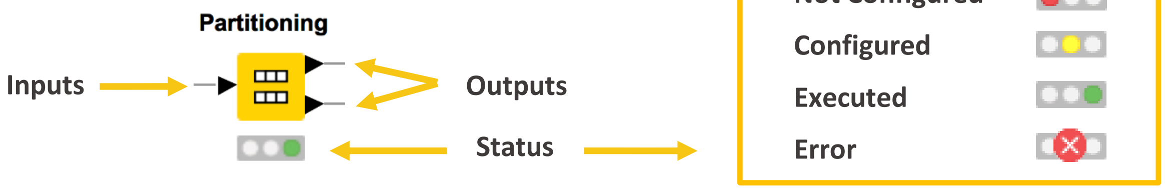
Description
01_Chemistry_Basics
This workflow demonstrates basic cheminformatics functionality within KNIME Analytics platform:
Reading and writing various chemistry data formats; canonalization of chemical structur
duplicate filtering; descriptor calculation;
interactive filtering on multiple properties
Training sets were collected from ChEMBLdt
Each set corresponds to a publication in whi

KNIME Console

```
Log file is located at: /Users/daria_knime/knime-workspace_rdkit_20200421/.metadata/knime/knime.log
WARN  Color Manager        0:34     Column "assay_id" has no nominal values set: execute predecesso
WARN  RDKit From Molecule  0:17     DataColumnSpec already contains a color handler, ignoring color
WARN  RDKit From Molecule  0:17     DataColumnSpec already contains a color handler, ignoring color
WARN  Row Filter           0:26     Node created an empty data table.
WARN  RDKit From Molecule  0:17     DataColumnSpec already contains a color handler, ignoring color
WARN  Column Filter        0:6      Node created an empty data table.
WARN  RDKit From Molecule  0:17     DataColumnSpec already contains a color handler, ignoring color
WARN  RDKit From Molecule  0:17     DataColumnSpec already contains a color handler, ignoring color
WARN  RDKit From Molecule  0:17     Node created empty data tables on all out-ports.
```

Visual KNIME Workflows

NODES perform tasks on data



Partitioning

Inputs

Outputs

Status

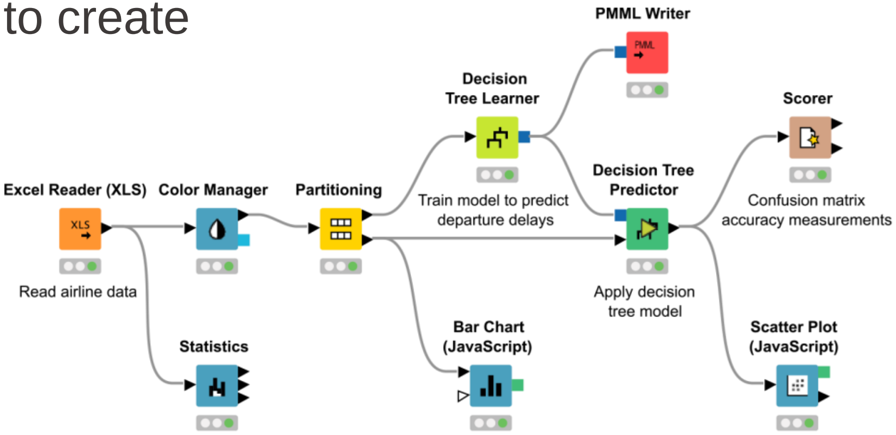
Not Configured

Configured

Executed

Error

Nodes are combined to create **WORKFLOWS**



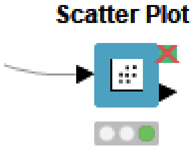
© 2023 KNIME AG. All rights reserved.

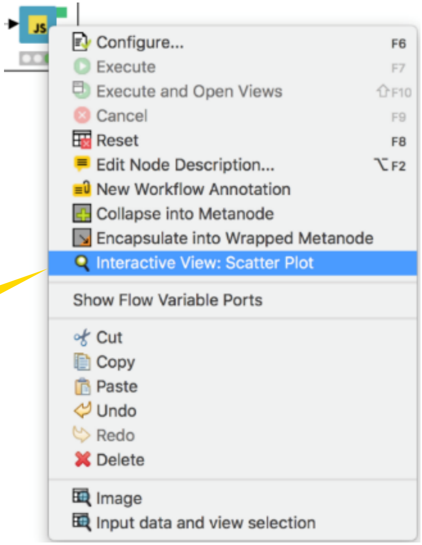
Node Outputs and Views

- Right-click executed node
- Select View option in context menu

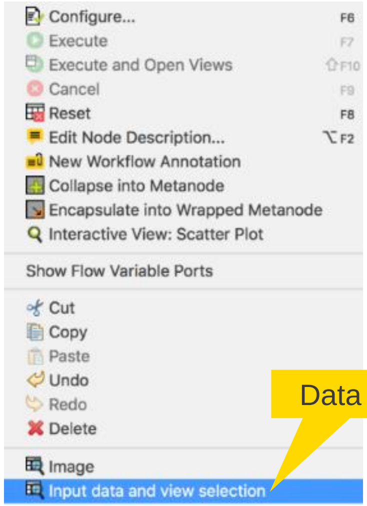
OR

- Select output port (last item) to inspect execution results




JS
Configure... F6
Execute F7
Execute and Open Views ⇧F10
Cancel F9
Reset F8
Edit Node Description... ⌥F2
New Workflow Annotation
Collapse into Metanode
Encapsulate into Wrapped Metanode
Interactive View: Scatter Plot
Show Flow Variable Ports
Cut
Copy
Paste
Undo
Redo
Delete
Image
Input data and view selection

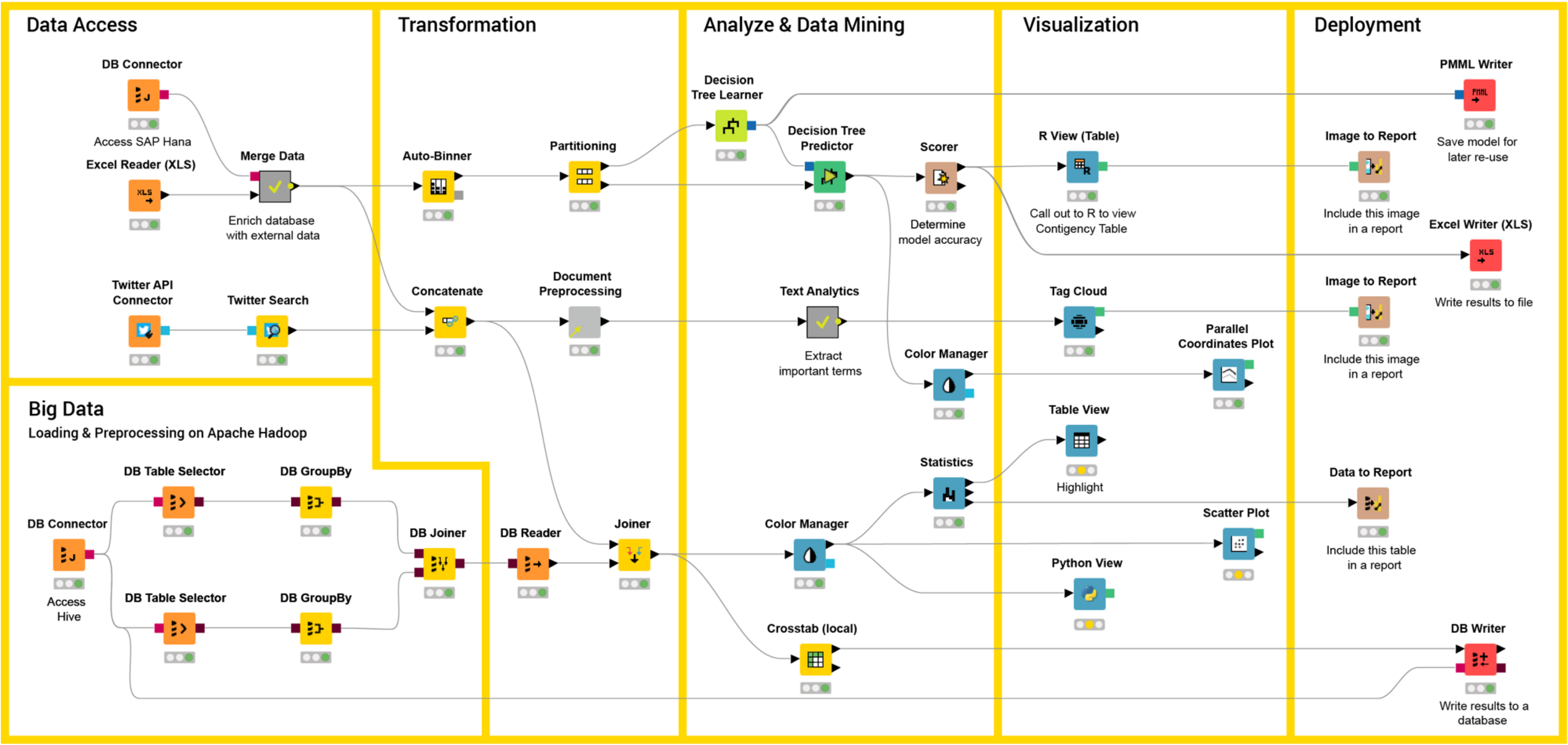
Plot View


Configure... F6
Execute F7
Execute and Open Views ⇧F10
Cancel F9
Reset F8
Edit Node Description... ⌥F2
New Workflow Annotation
Collapse into Metanode
Encapsulate into Wrapped Metanode
Interactive View: Scatter Plot
Show Flow Variable Ports
Cut
Copy
Paste
Undo
Redo
Delete
Image
Input data and view selection

Data View

© 2023 KNIME AG. All rights reserved.

4000+ Nodes for all Steps of End-To-End Data Science



Data Access

DB Connector
Access SAP Hana

Excel Reader (XLS)

Merge Data
Enrich database with external data

Twitter API Connector

Twitter Search

Big Data

Loading & Preprocessing on Apache Hadoop

DB Connector
Access Hive

DB Table Selector

DB GroupBy

DB Table Selector

DB GroupBy

DB Joiner

Transformation

Auto-Binner

Partitioning

Concatenate

Document Preprocessing

DB Reader

Joiner

Analyze & Data Mining

Decision Tree Learner

Decision Tree Predictor

Scorer
Determine model accuracy

Text Analytics
Extract important terms

Color Manager

Statistics

Color Manager

Crosstab (local)

Visualization

R View (Table)
Call out to R to view Contigency Table

Tag Cloud

Parallel Coordinates Plot

Table View
Highlight

Scatter Plot

Python View

Deployment

PMML Writer
Save model for later re-use

Image to Report
Include this image in a report

Excel Writer (XLS)
Write results to file

Image to Report
Include this image in a report

Data to Report
Include this table in a report

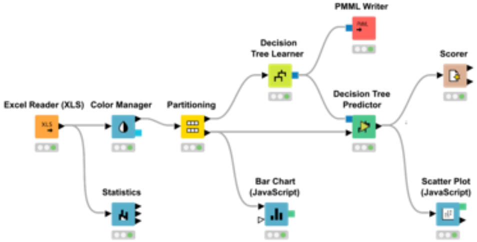
DB Writer
Write results to a database

© 2023 KNIME AG. All rights reserved.

Mix & Match Data Sources, Technologies, and Execution

sas mongoDB DynamoDB
Google Drive salesforce SAP
REST

External Data Sources



TensorFlow H2O.ai ONNX
R python
dmlc XGBoost K

AI / ML / DL Libraries

Spotfire
Power BI plotly
PDF Adobe
tableau

Reporting & Visualization

Scripting Languages

R python

databricks
Google Cloud
cloudera IMPALA
HIVE Spark Amazon Athena hadoop
aws REDSHIFT Azure

Distributed / Cloud Execution

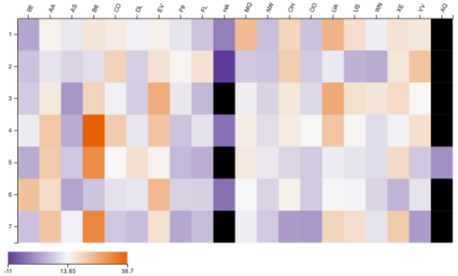
© 2023 KNIME AG. All rights reserved.

Open for Innovation KNIME

Data Visualization

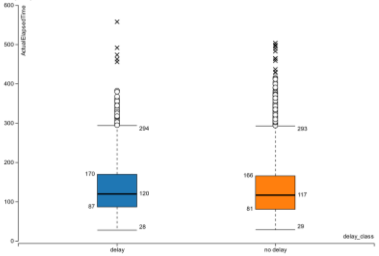
Delay Times by Day of the Week and Carrier

Red indicates large delay



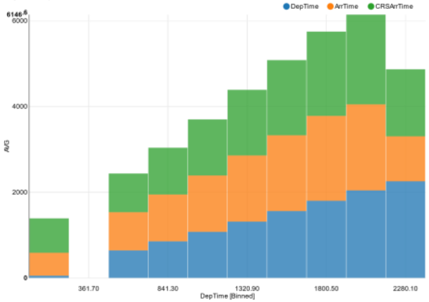
Distribution of Elapsed Time

By delay class



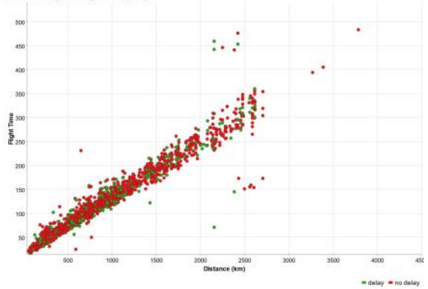
Average Arrival Time and CRS Arrival Time

By binned departure time



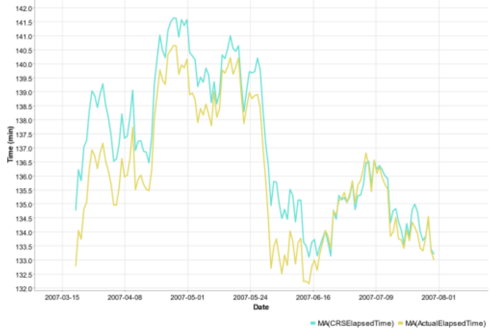
Correlation between Distance and Flight Time

Red indicates a flight on time, green a delayed flight

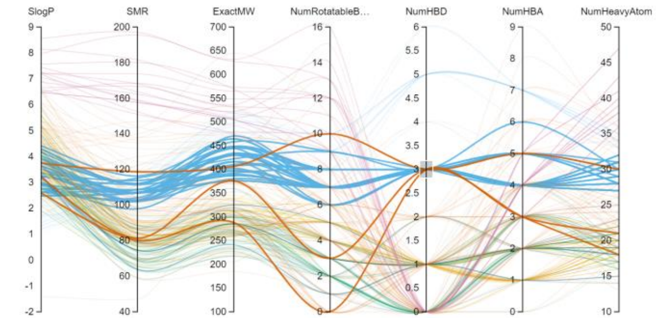


Actual Elapsed Time vs CRS Elapsed Time

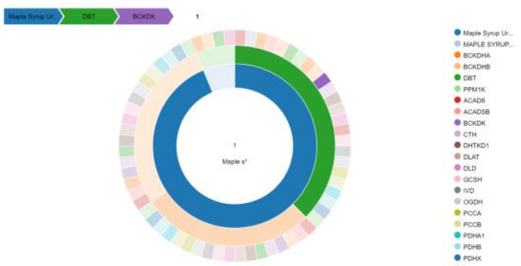
Moving average over the last 20 days



PhysChem Properties of Compounds



Interactions



© 2023 KNIME AG. All rights reserved.

Why KNIME?

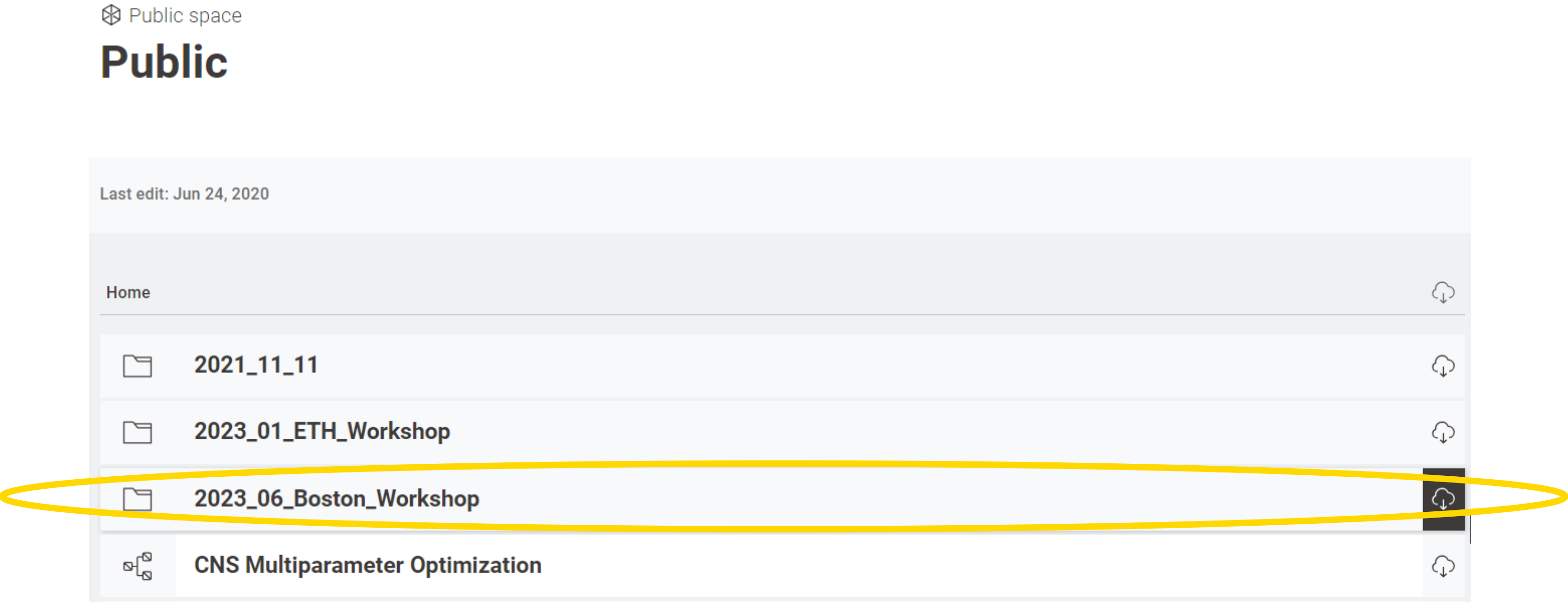
- Self-documenting
  - Every workflow is a track of what you did to your data
- Reproducible
  - Once configured, it will run the same way every time and have the same results
- Data is available at every manipulation step
  - Trace and control how the data table is changing throughout the workflow
- Interactive data visualization
- Chemistry capabilities

© 2023 KNIME AG. All rights reserved.

Setup KNIME AP for today

Set up KNIME Analytics Platform

- Download workflows from the KNIME Community Hub
  - https://kni.me/s/5Goi-yKrwMLzAU4o



Public space

Public

Last edit: Jun 24, 2020

Home

2021_11_11

2023_01_ETH_Workshop

2023_06_Boston_Workshop

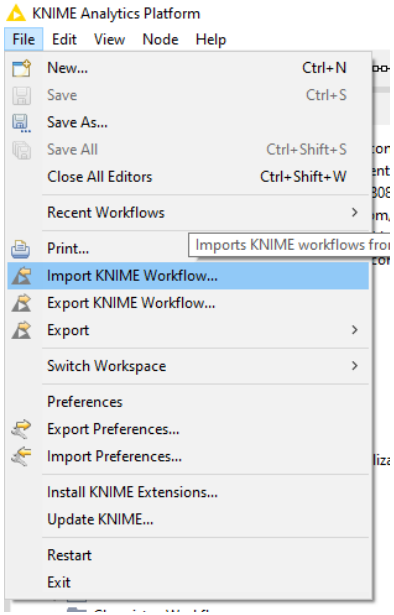
CNS Multiparameter Optimization

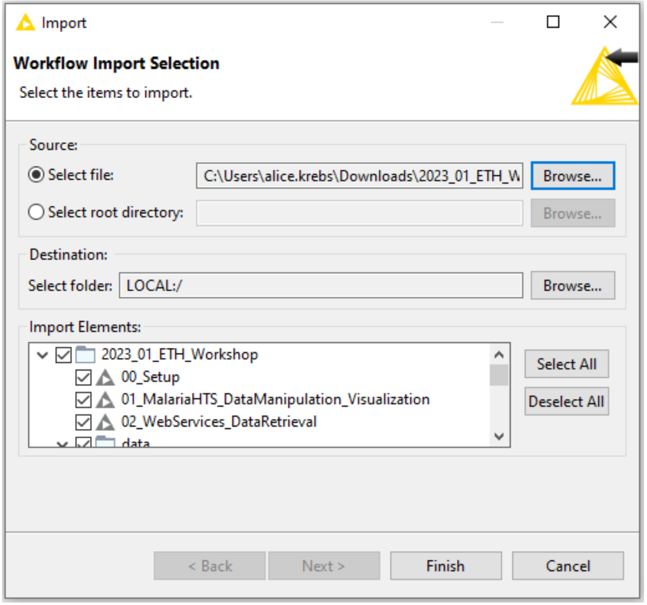
© 2023 KNIME AG. All rights reserved.

Import the workflow group

- Navigate to your download folder and choose the “2023_06_Boston_Workshop.knar” file

.knwf → single workflow
.knar → group of workflows





● Select file: C:\Users\alice.krebs\Downloads\2023_01_ETH_W

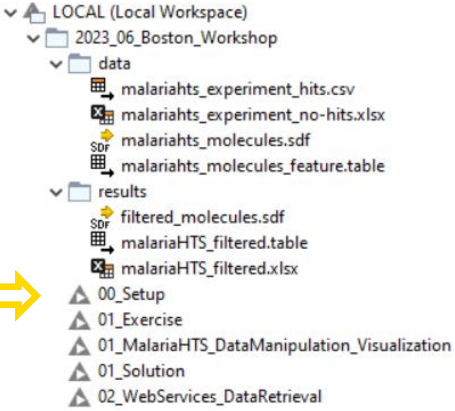
○ Select root directory:

Select folder: LOCAL:/

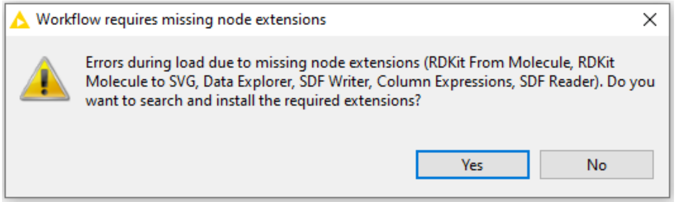
- ☑ 2023_01_ETH_Workshop
  - ☑ 00_Setup
  - ☑ 01_MalariaHTS_DataManipulation_Visualization
  - ☑ 02_WebServices_DataRetrieval
  - ☑ data

Getting started

- Open the “00_Setup” workflow to install all the extensions you need today

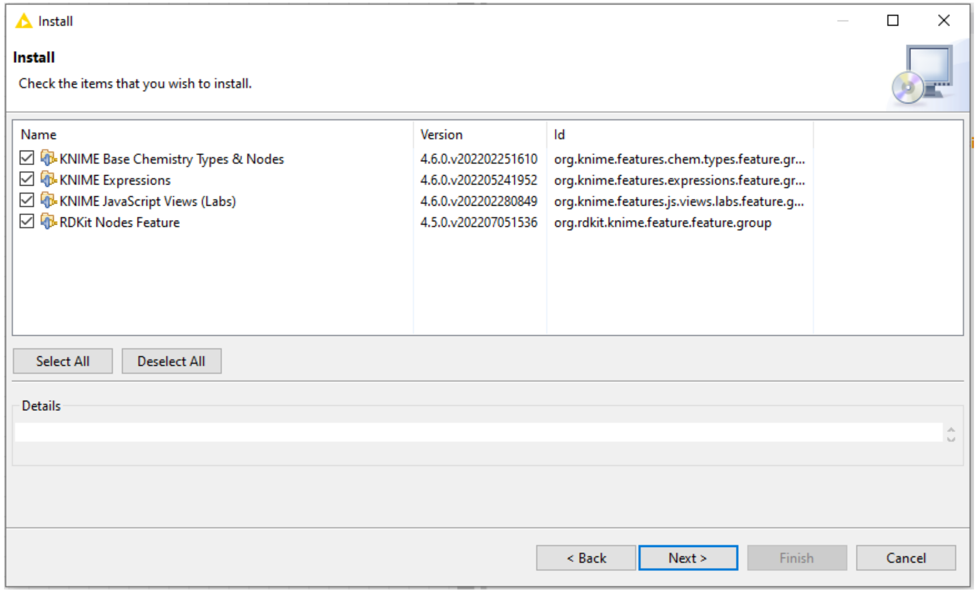


- LOCAL (Local Workspace)
  - 2023_06_Boston_Workshop
    - data
      - malariahts_experiment_hits.csv
      - malariahts_experiment_no-hits.xlsx
      - malariahts_molecules.sdf
      - malariahts_molecules_feature.table
    - results
      - filtered_molecules.sdf
      - malariaHTS_filtered.table
      - malariaHTS_filtered.xlsx
    - 00_Setup
    - 01_Exercise
    - 01_MalariaHTS_DataManipulation_Visualization
    - 01_Solution
    - 02_WebServices_DataRetrieval



Workflow requires missing node extensions

Errors during load due to missing node extensions (RDKit From Molecule, RDKit Molecule to SVG, Data Explorer, SDF Writer, Column Expressions, SDF Reader). Do you want to search and install the required extensions?



Install

Install

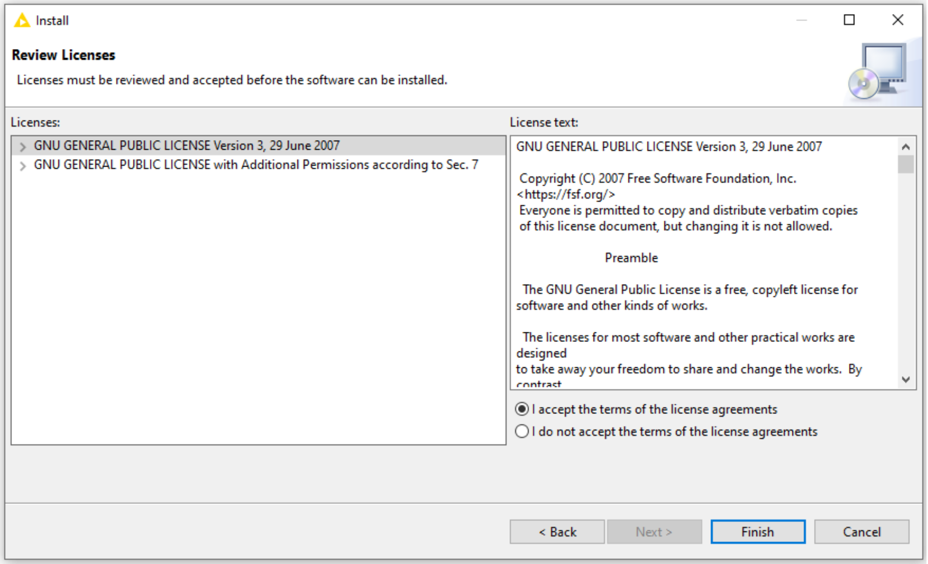
Check the items that you wish to install.

Name	Version	Id
☑ KNIME Base Chemistry Types & Nodes	4.6.0.v202202251610	org.knime.features.chem.types.feature.gr...
☑ KNIME Expressions	4.6.0.v202205241952	org.knime.features.expressions.feature.gr...
☑ KNIME JavaScript Views (Labs)	4.6.0.v202202280849	org.knime.features.js.views.labs.feature.g...
☑ RDKit Nodes Feature	4.5.0.v202207051536	org.rdkit.knime.feature.feature.group

© 2023 KNIME AG. All rights reserved.

Getting started

- Open the “00_Setup” workflow to install all the extensions you need today



Review Licenses

Licenses must be reviewed and accepted before the software can be installed.

Licenses:

- GNU GENERAL PUBLIC LICENSE Version 3, 29 June 2007
- GNU GENERAL PUBLIC LICENSE with Additional Permissions according to Sec. 7

License text:

GNU GENERAL PUBLIC LICENSE Version 3, 29 June 2007

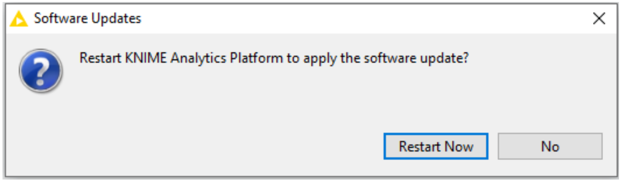
Copyright (C) 2007 Free Software Foundation, Inc.
<https://fsf.org/>
Everyone is permitted to copy and distribute verbatim copies
of this license document, but changing it is not allowed.

Preamble

The GNU General Public License is a free, copyleft license for software and other kinds of works.

The licenses for most software and other practical works are designed
to take away your freedom to share and change the works. By contrast

● I accept the terms of the license agreements
○ I do not accept the terms of the license agreements

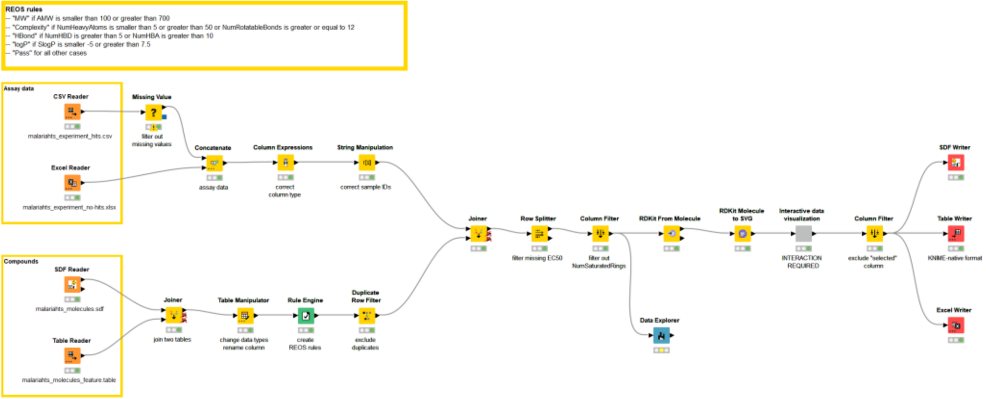


Restart KNIME Analytics Platform to apply the software update?

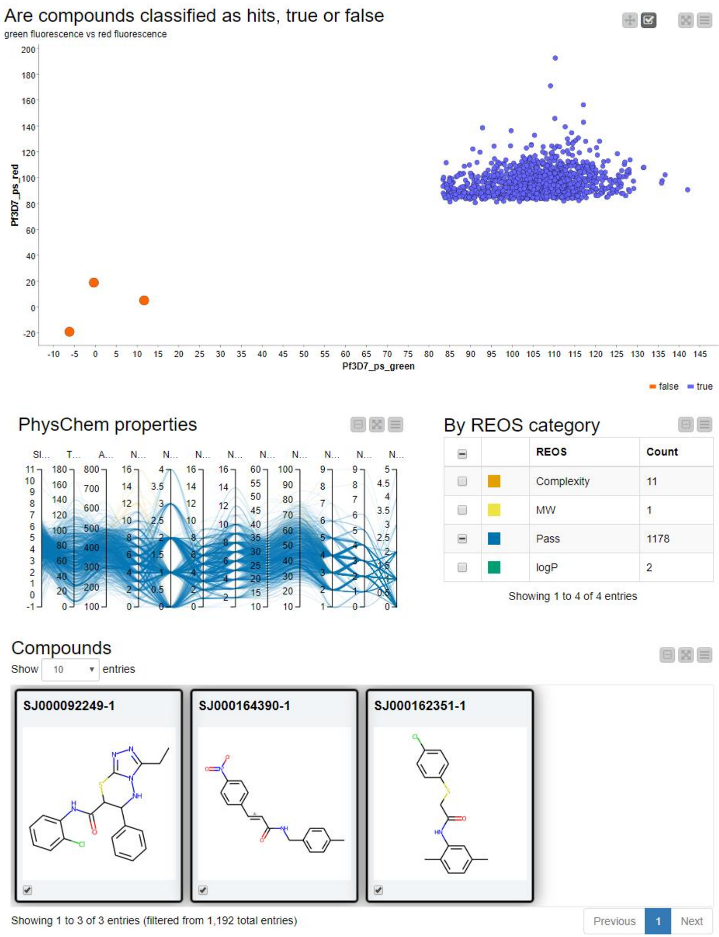
Ready to go!

Today’s workflows I

- 01_MalariaHTS_DataManipulation_Visualization



REOS rules
– "MW" if AMW is smaller than 100 or greater than 700
– "Complexity" if NumHeavyAtoms is smaller than 5 or greater than 50 or NumRotatableBonds is greater or equal to 12
– "HBond" if NumHBD is greater than 5 or NumHBA is greater than 10
– "logP" if SlogP is smaller -5 or greater than 7.5
– "Pass" for all other cases



By REOS category

⊟		REOS	Count
☐	🟧	Complexity	11
☐	🟨	MW	1
⊟	🟦	Pass	1178
☐	🟩	logP	2

Showing 1 to 4 of 4 entries

Compounds

Show 10 entries

SJ000092249-1 ☑

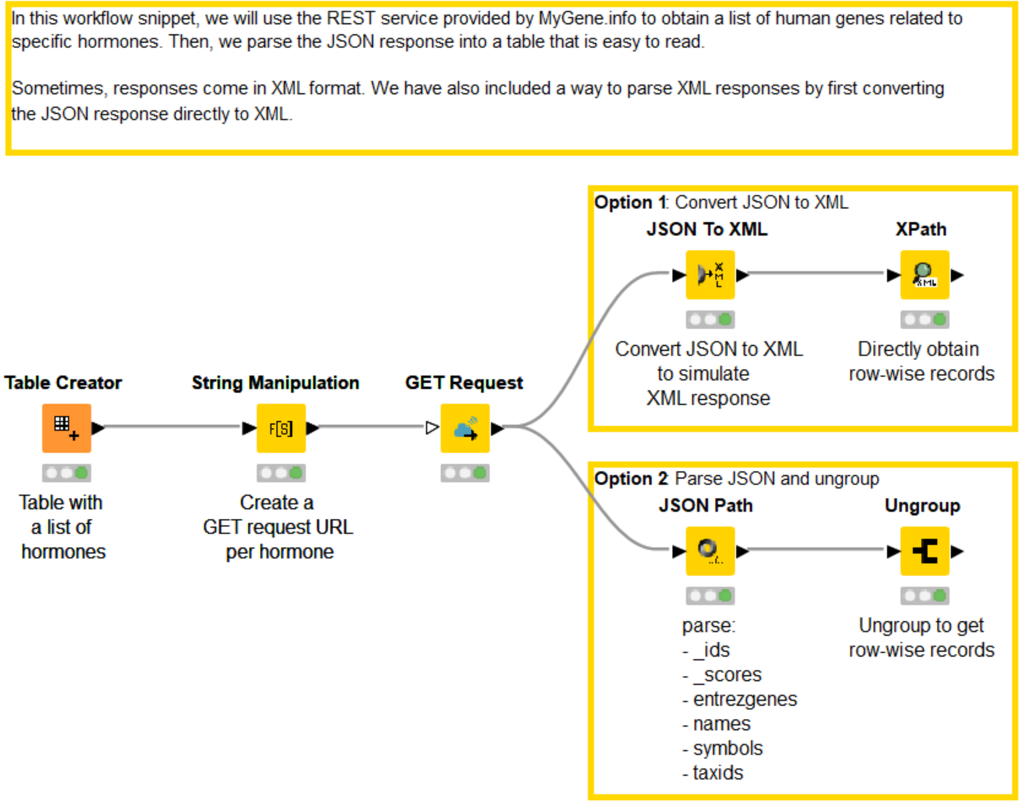
SJ000164390-1 ☑

SJ000162351-1 ☑

Showing 1 to 3 of 3 entries (filtered from 1,192 total entries)

Today’s workflows II

- 02_WebServices_DataRetrieval



In this workflow snippet, we will use the REST service provided by MyGene.info to obtain a list of human genes related to specific hormones. Then, we parse the JSON response into a table that is easy to read.

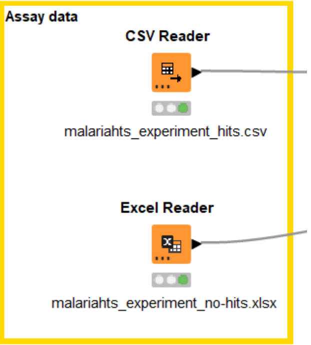
Sometimes, responses come in XML format. We have also included a way to parse XML responses by first converting the JSON response directly to XML.

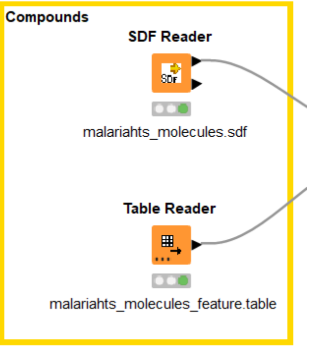
Table Creator
Table with a list of hormones
String Manipulation
Create a GET request URL per hormone
GET Request
Option 1: Convert JSON to XML
JSON To XML
Convert JSON to XML to simulate XML response
XPath
Directly obtain row-wise records
Option 2: Parse JSON and ungroup
JSON Path
parse:
- _ids
- _scores
- entrezgenes
- names
- symbols
- taxids
Ungroup
Ungroup to get row-wise records

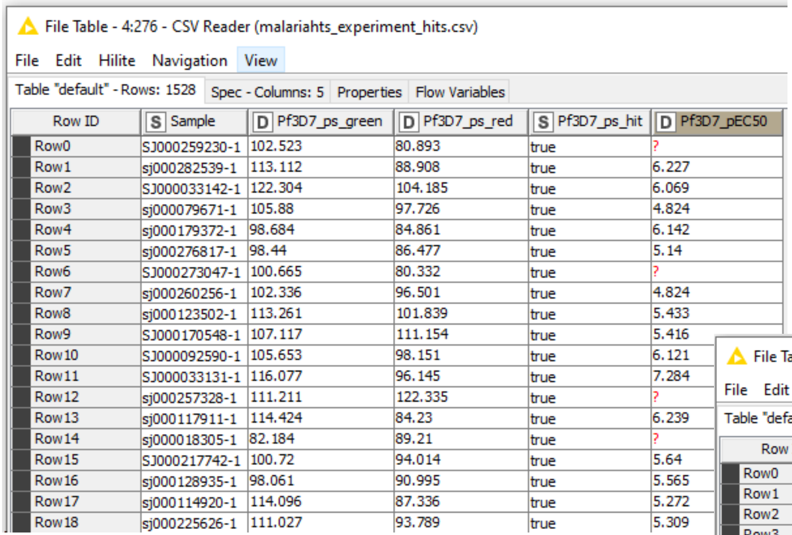
© 2023 KNIME AG. All rights reserved.

01_MalariaHTS_DataManipulation_Visualization

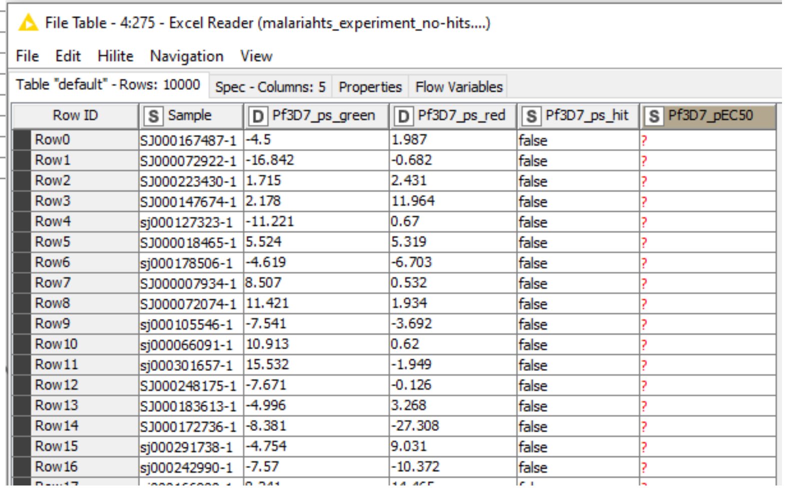
- **Step 1:** import data from 4 different sources


Assay data
CSV Reader
malariahts_experiment_hits.csv
Excel Reader
malariahts_experiment_no-hits.xlsx




File Table - 4:276 - CSV Reader (malariahts_experiment_hits.csv)
File
Edit
Hilite
Navigation
View
Table "default" - Rows: 1528
Spec - Columns: 5
Properties
Flow Variables

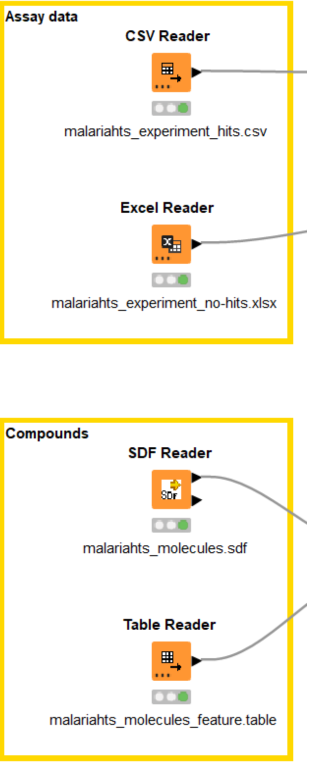
Row ID	S Sample	D Pf3D7_ps_green	D Pf3D7_ps_red	S Pf3D7_ps_hit	D Pf3D7_pEC50
Row0	SJ000259230-1	102.523	80.893	true	?
Row1	sj000282539-1	113.112	88.908	true	6.227
Row2	SJ000033142-1	122.304	104.185	true	6.069
Row3	sj000079671-1	105.88	97.726	true	4.824
Row4	sj000179372-1	98.684	84.861	true	6.142
Row5	sj000276817-1	98.44	86.477	true	5.14
Row6	SJ000273047-1	100.665	80.332	true	?
Row7	sj000260256-1	102.336	96.501	true	4.824
Row8	sj000123502-1	113.261	101.839	true	5.433
Row9	SJ000170548-1	107.117	111.154	true	5.416
Row10	SJ000092590-1	105.653	98.151	true	6.121
Row11	SJ000033131-1	116.077	96.145	true	7.284
Row12	sj000257328-1	111.211	122.335	true	?
Row13	sj000117911-1	114.424	84.23	true	6.239
Row14	sj000018305-1	82.184	89.21	true	?
Row15	SJ000217742-1	100.72	94.014	true	5.64
Row16	sj000128935-1	98.061	90.995	true	5.565
Row17	sj000114920-1	114.096	87.336	true	5.272
Row18	sj000225626-1	111.027	93.789	true	5.309

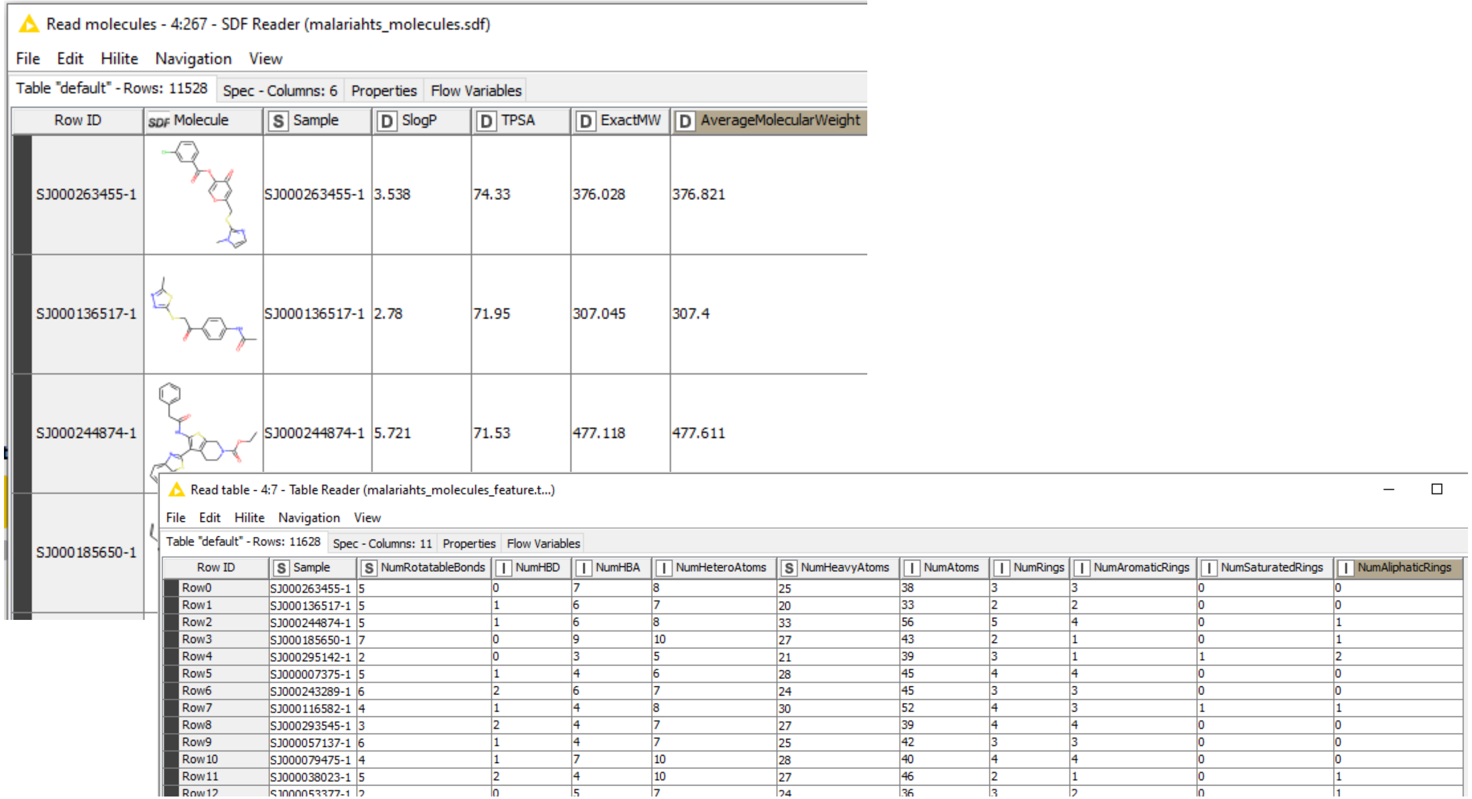

File Table - 4:275 - Excel Reader (malariahts_experiment_no-hits....)
File
Edit
Hilite
Navigation
View
Table "default" - Rows: 10000
Spec - Columns: 5
Properties
Flow Variables

Row ID	S Sample	D Pf3D7_ps_green	D Pf3D7_ps_red	S Pf3D7_ps_hit	S Pf3D7_pEC50
Row0	SJ000167487-1	-4.5	1.987	false	?
Row1	SJ000072922-1	-16.842	-0.682	false	?
Row2	SJ000223430-1	1.715	2.431	false	?
Row3	SJ000147674-1	2.178	11.964	false	?
Row4	sj000127323-1	-11.221	0.67	false	?
Row5	SJ000018465-1	5.524	5.319	false	?
Row6	sj000178506-1	-4.619	-6.703	false	?
Row7	SJ000007934-1	8.507	0.532	false	?
Row8	SJ000072074-1	11.421	1.934	false	?
Row9	sj000105546-1	-7.541	-3.692	false	?
Row10	sj000066091-1	10.913	0.62	false	?
Row11	sj000301657-1	15.532	-1.949	false	?
Row12	SJ000248175-1	-7.671	-0.126	false	?
Row13	SJ000183613-1	-4.996	3.268	false	?
Row14	SJ000172736-1	-8.381	-27.308	false	?
Row15	sj000291738-1	-4.754	9.031	false	?
Row16	sj000242990-1	-7.57	-10.372	false	?

01_MalariaHTS_DataManipulation_Visualization

- **Step 1:** import data from 4 different sources


Assay data
CSV Reader
malariahts_experiment_hits.csv
Excel Reader
malariahts_experiment_no-hits.xlsx
Compounds
SDF Reader
malariahts_molecules.sdf
Table Reader
malariahts_molecules_feature.table



Read molecules - 4:267 - SDF Reader (malariahts_molecules.sdf)

Table "default" - Rows: 11528

Row ID	SDF Molecule	S Sample	D SlogP	D TPSA	D ExactMW	D AverageMolecularWeight
SJ000263455-1		SJ000263455-1	3.538	74.33	376.028	376.821
SJ000136517-1		SJ000136517-1	2.78	71.95	307.045	307.4
SJ000244874-1		SJ000244874-1	5.721	71.53	477.118	477.611
SJ000185650-1						

Read table - 4:7 - Table Reader (malariahts_molecules_feature.t...)

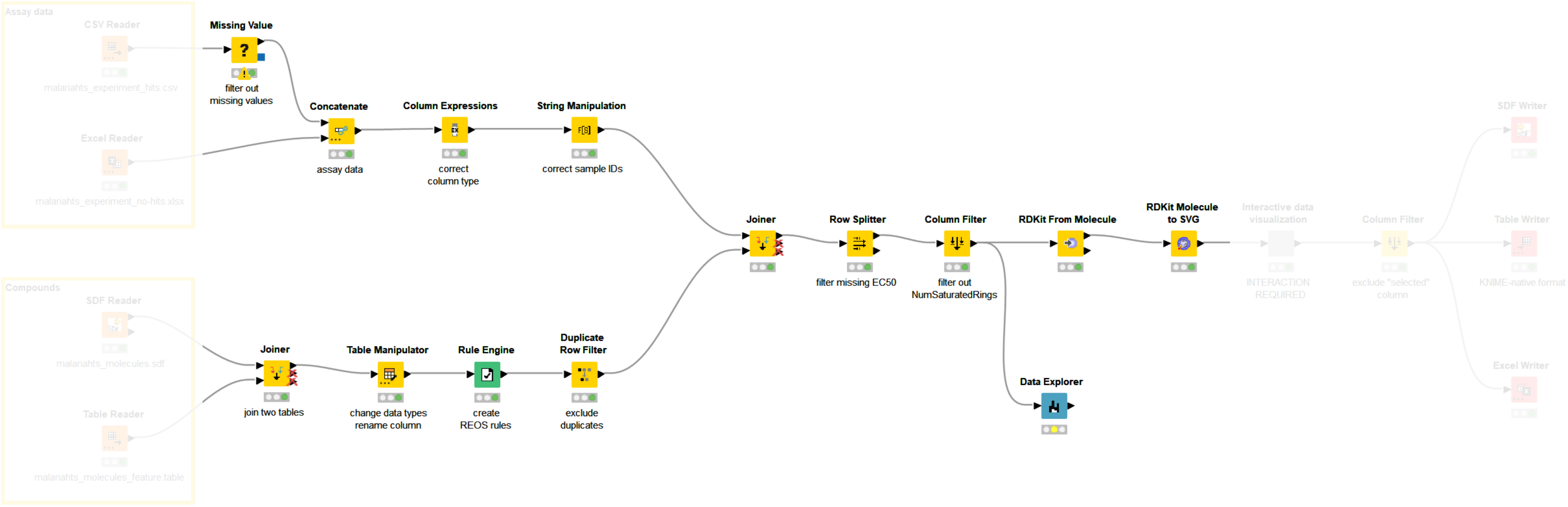
Table "default" - Rows: 11628

Row ID	S Sample	S NumRotatableBonds	I NumHBD	I NumHBA	I NumHeteroAtoms	S NumHeavyAtoms	I NumAtoms	I NumRings	I NumAromaticRings	I NumSaturatedRings	I NumAliphaticRings
Row0	SJ000263455-1	5	0	7	8	25	38	3	3	0	0
Row1	SJ000136517-1	5	1	6	7	20	33	2	2	0	0
Row2	SJ000244874-1	5	1	6	8	33	56	5	4	0	1
Row3	SJ000185650-1	7	0	9	10	27	43	2	1	0	1
Row4	SJ000295142-1	2	0	3	5	21	39	3	1	1	2
Row5	SJ000007375-1	5	1	4	6	28	45	4	4	0	0
Row6	SJ000243289-1	6	2	6	7	24	45	3	3	0	0
Row7	SJ000116582-1	4	1	4	8	30	52	4	3	1	1
Row8	SJ000293545-1	3	2	4	7	27	39	4	4	0	0
Row9	SJ000057137-1	6	1	4	7	25	42	3	3	0	0
Row10	SJ000079475-1	4	1	7	10	28	40	4	4	0	0
Row11	SJ000038023-1	5	2	4	10	27	46	2	1	0	1
Row12	SJ000053377-1	2	0	5	7	24	36	3	2	0	1

© 2023 KNIME AG. All rights reserved.

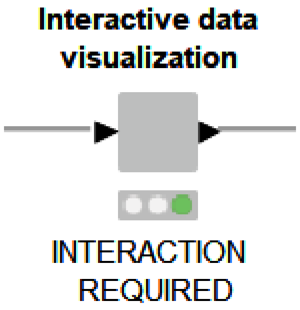
01_MalariaHTS_DataManipulation_Visualization

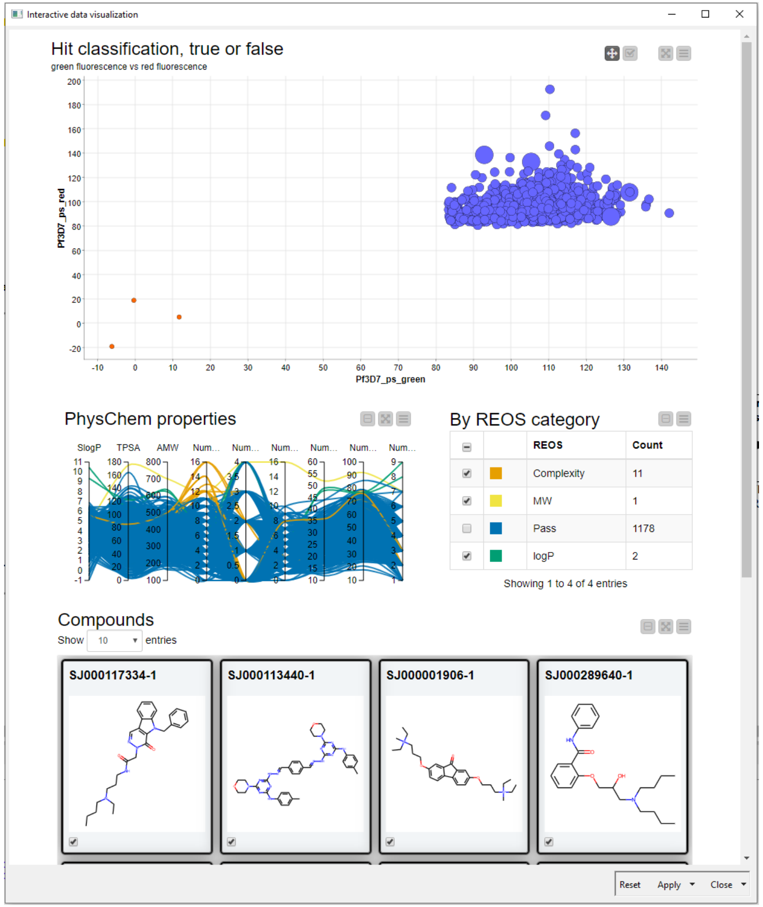
- **Step 2:** data cleaning, merging datasets, manipulate data, create classification



01_MalariaHTS_DataManipulation_Visualization

- **Step 3:** visualize and interactively select data
  - build a component to combine different plots in one view





Hit classification, true or false

green fluorescence vs red fluorescence

PhysChem properties

By REOS category

⊟		REOS	Count
☑	🟧	Complexity	11
☑	🟨	MW	1
☐	🟦	Pass	1178
☑	🟩	logP	2

Showing 1 to 4 of 4 entries

Compounds

Show 10 entries

SJ000117334-1

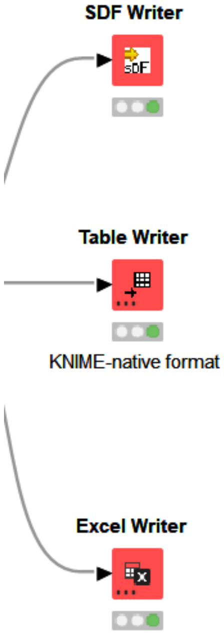
SJ000113440-1

SJ000001906-1

SJ000289640-1

01_MalariaHTS_DataManipulation_Visualization

- **Step 4:** Export the data
  - SD file
  - KNIME-native table (only re-usable in KNIME)
  - Excel



© 2023 KNIME AG. All rights reserved.

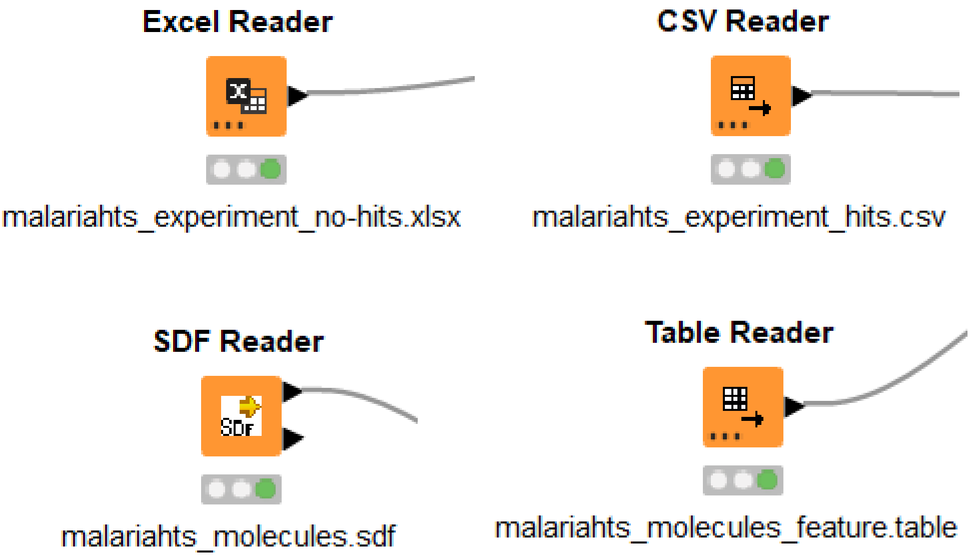
Content, theory, background on…

- Import data
  - Local file system
  - Workflow relative path
- Merging data sets
  - Concatenate
  - Join
- Exclude data
  - Filter
  - Splitter
- Handle missing values and duplicates
- Change data
  - Change assigned data type
  - Capitalize the letters in the sample ID
  - Rename columns
  - Create classification of REOS rules
- Visualize data in components

© 2023 KNIME AG. All rights reserved.

Import data

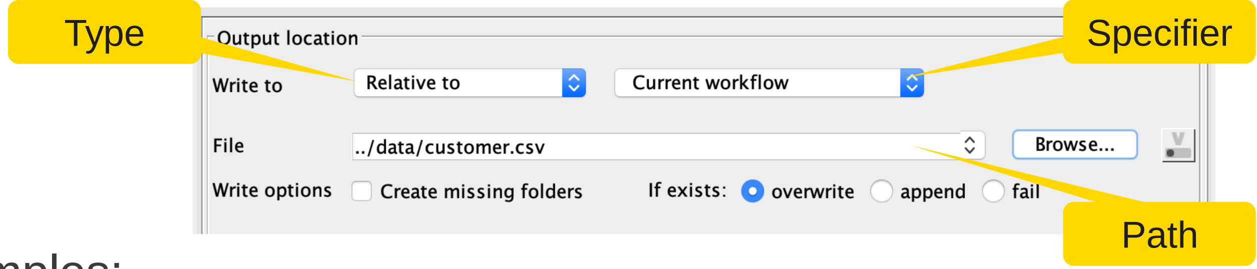
- Different Reader nodes for different file formats
- “Table” is a KNIME native format
- Drag and drop the files from the data folder in the KNIME Explorer


Excel Reader
CSV Reader
malariahts_experiment_no-hits.xlsx
malariahts_experiment_hits.csv
SDF Reader
Table Reader
malariahts_molecules.sdf
malariahts_molecules_feature.table

© 2023 KNIME AG. All rights reserved.

Common Settings: File Path

- A path consists of three parts:
  - **Type**: Specifies the file system type e.g. local, relative, mountpoint, custome_url or connected.
  - **Specifier**: Optional string with additional file system specific information e.g. relative to which location (knime.workflow)
  - **Path**: Specifies the location within the file system



Write to: Relative to | Current workflow

File: ../data/customer.csv

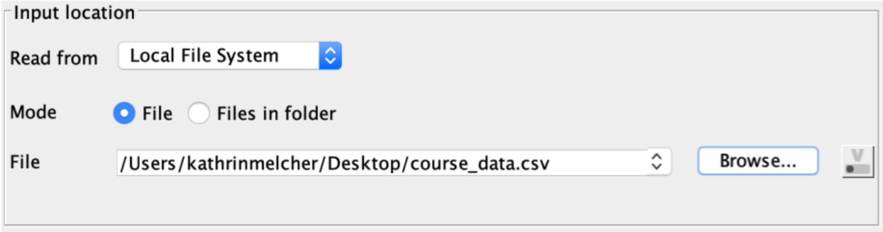
Write options: ☐ Create missing folders   If exists: ● overwrite ○ append ○ fail

- Examples:
  - (LOCAL, , C:\Users\username\Desktop)
  - (RELATIVE, knime.workflow, file1.csv)
  - (MOUNTPOINT, MOUNTPOINT_NAME, /path/to/file1.csv)
  - (CONNECTED, amazon-s3:eu-west-1, /mybucket/file1.csv)

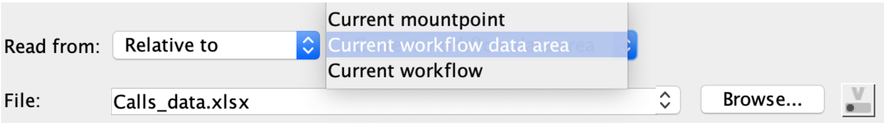
© 2023 KNIME AG. All rights reserved.

Common Settings: 4 Default File Systems

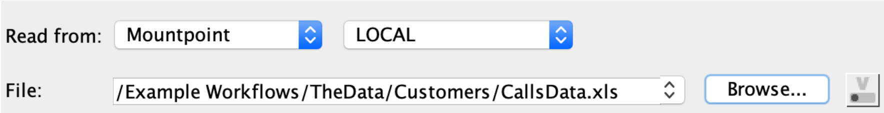
- ## Local File System


Input location
Read from: Local File System
Mode: ● File ○ Files in folder
File: /Users/kathrinmelcher/Desktop/course_data.csv

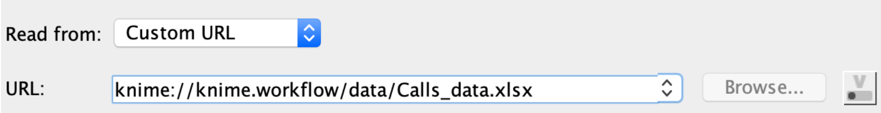
- ## Relative to …


Read from: Relative to
Current mountpoint
Current workflow data area
Current workflow
File: Calls_data.xlsx

- ## Mountpoint


Read from: Mountpoint LOCAL
File: /Example Workflows/TheData/Customers/CallsData.xls

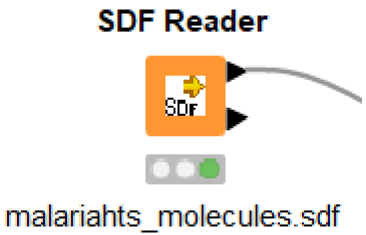
- ## Custom URL

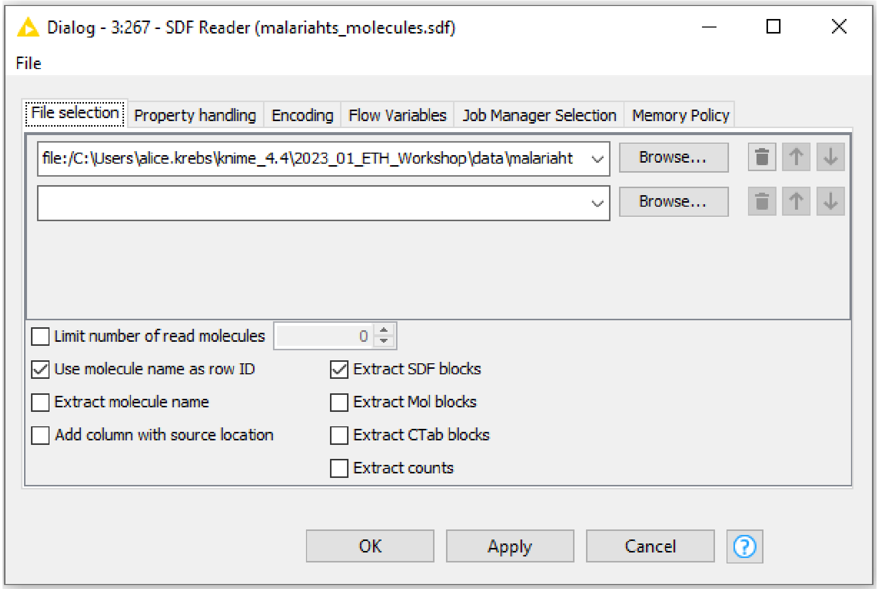

Read from: Custom URL
URL: knime://knime.workflow/data/Calls_data.xlsx

© 2023 KNIME AG. All rights reserved.

Special case – SDF reader node

- No new file handling
- Extract properties to get all data





file:/C:\Users\alice.krebs\knime_4.4\2023_01_ETH_Workshop\data\malariaht

☐ Limit number of read molecules 0

☑ Use molecule name as row ID

☐ Extract molecule name

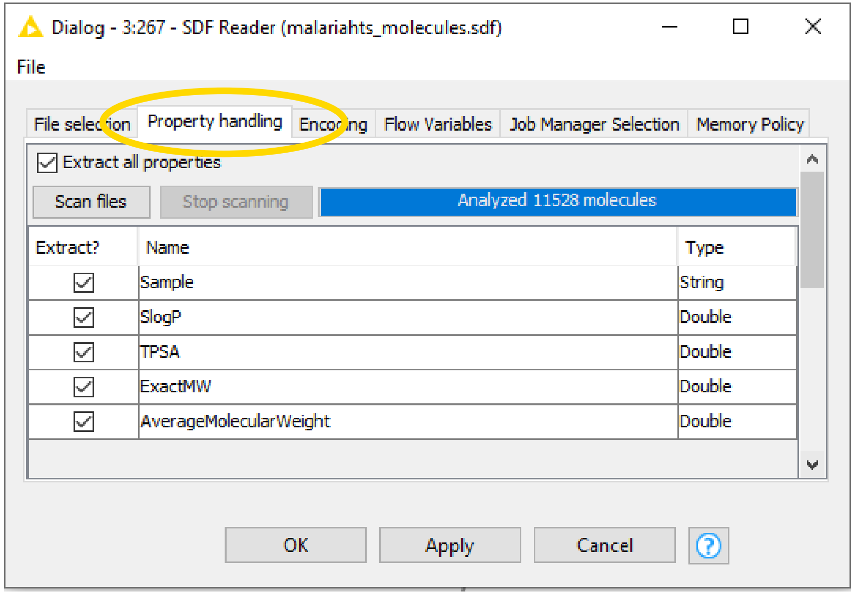
☐ Add column with source location

☑ Extract SDF blocks

☐ Extract Mol blocks

☐ Extract CTab blocks

☐ Extract counts



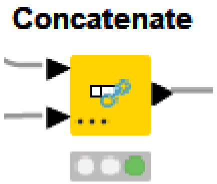
☑ Extract all properties

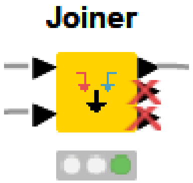
Analyzed 11528 molecules

Extract?	Name	Type
☑	Sample	String
☑	SlogP	Double
☑	TPSA	Double
☑	ExactMW	Double
☑	AverageMolecularWeight	Double

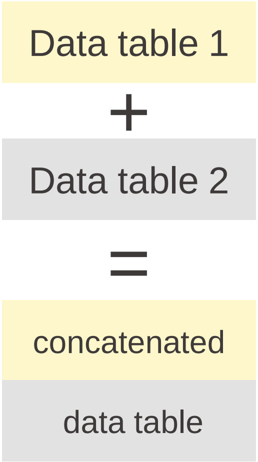
© 2023 KNIME AG. All rights reserved.

Concatenate vs. Join

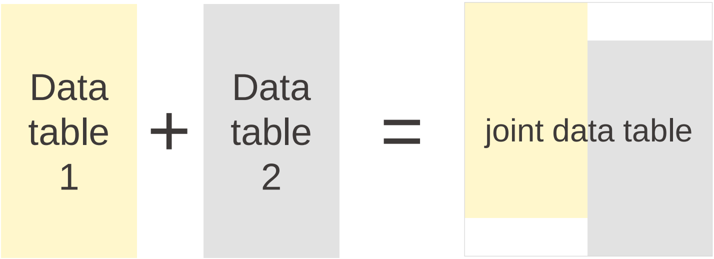




- Simply links to tables, appends them underneath each other, like in a chain



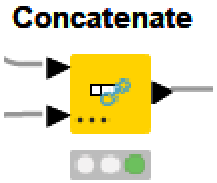
- Combines two tables row wise

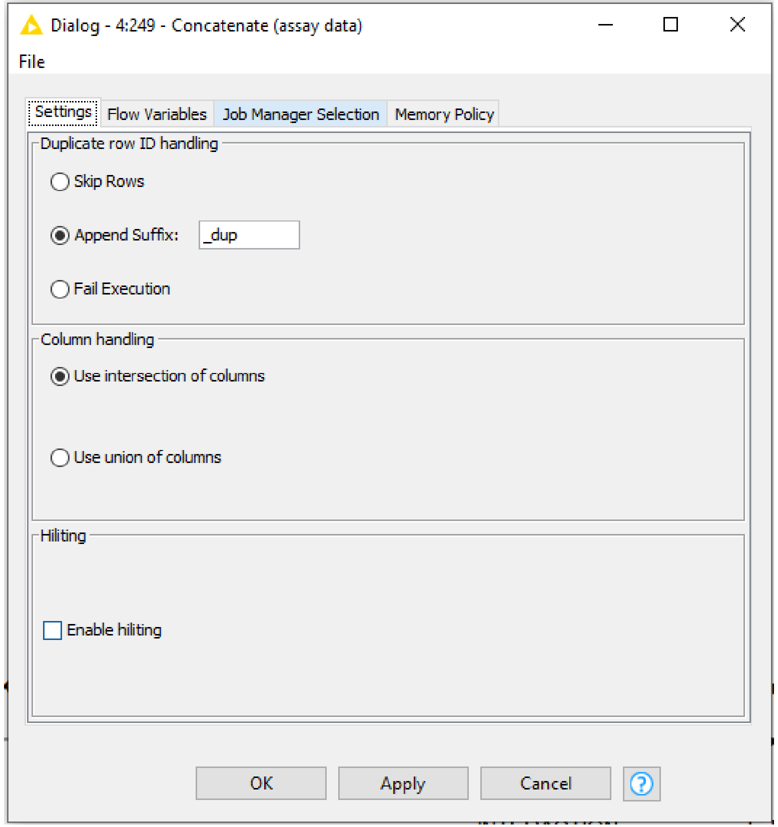


Concatenate node

- Intersection
  - use only the columns that appear in both input tables

- Union
  - use all columns available in the input tables




Dialog - 4:249 - Concatenate (assay data)

File

Settings | Flow Variables | Job Manager Selection | Memory Policy

Duplicate row ID handling

○ Skip Rows

● Append Suffix: _dup

○ Fail Execution

Column handling

● Use intersection of columns

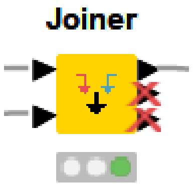
○ Use union of columns

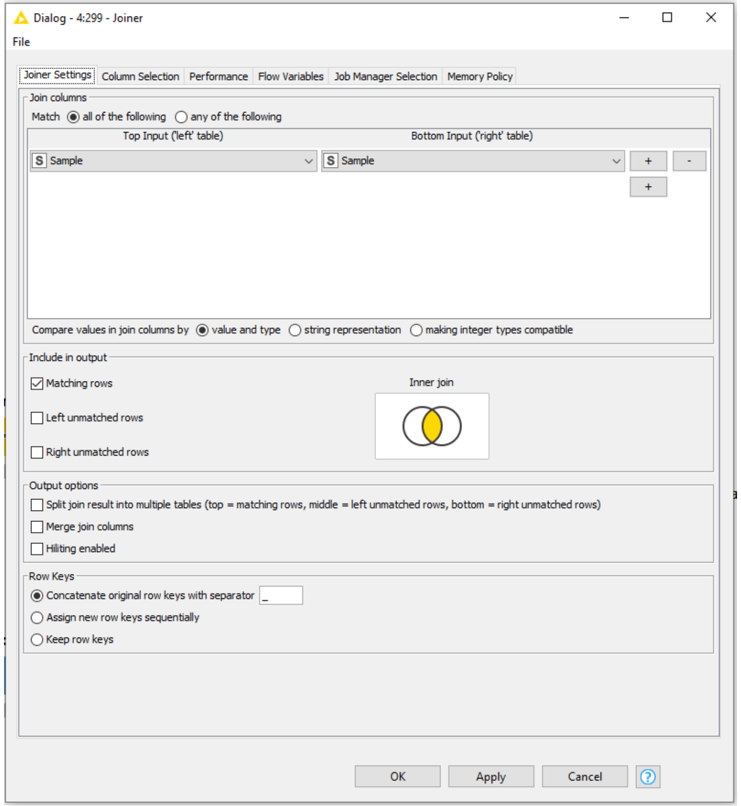
Hiliting

☐ Enable hiliting

© 2023 KNIME AG. All rights reserved.

Joiner node





Join columns

Match ● all of the following ○ any of the following

Top Input ('left' table)	Bottom Input ('right' table)
S Sample	S Sample

Compare values in join columns by ● value and type ○ string representation ○ making integer types compatible

Include in output

☑ Matching rows

☐ Left unmatched rows

☐ Right unmatched rows

Inner join

Output options

☐ Split join result into multiple tables (top = matching rows, middle = left unmatched rows, bottom = right unmatched rows)

☐ Merge join columns

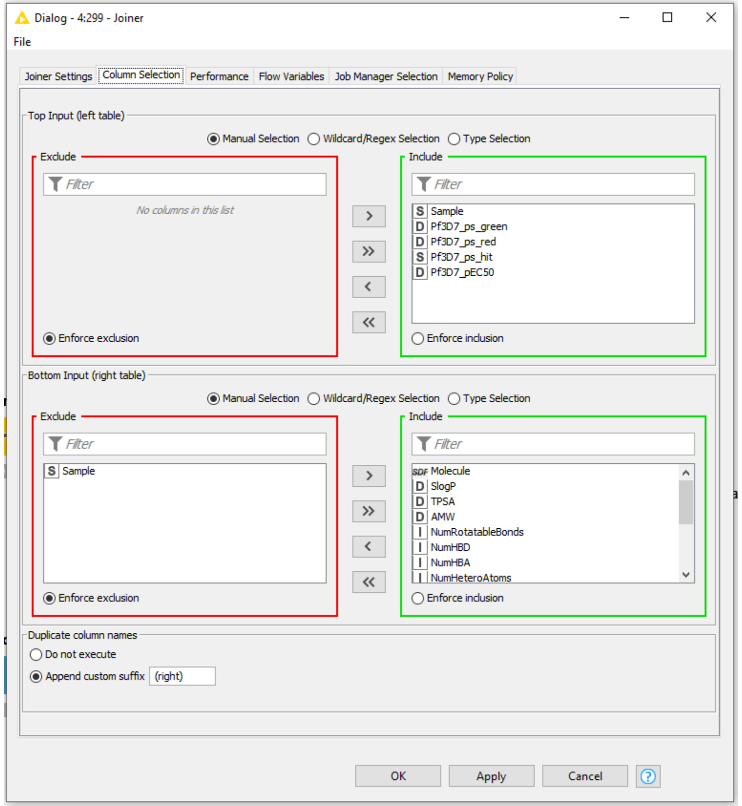
☐ Hiliting enabled

Row Keys

● Concatenate original row keys with separator _

○ Assign new row keys sequentially

○ Keep row keys



Top Input (left table)

● Manual Selection ○ Wildcard/Regex Selection ○ Type Selection

Exclude	Include
No columns in this list	S Sample D Pf3D7_ps_green D Pf3D7_ps_red S Pf3D7_ps_hit D Pf3D7_pEC50
● Enforce exclusion	○ Enforce inclusion

Bottom Input (right table)

● Manual Selection ○ Wildcard/Regex Selection ○ Type Selection

Exclude	Include
S Sample	SDF Molecule D SlogP D TPSA D AMW I NumRotatableBonds I NumHBD I NumHBA I NumHeteroAtoms
● Enforce exclusion	○ Enforce inclusion

Duplicate column names

○ Do not execute

● Append custom suffix (right)

© 2023 KNIME AG. All rights reserved.

Joining Columns of Data – Inner Join

Left Table

molregno	chembl_id	SMILES
22	CHEMBL1794855	CCCN(CCC)
24	CHEMBL278751	CCN(C)
15	CHEMBL103772	CCCN1CC
10	CHEMBL328107	C1CN(CCN1)

Right Table

molregno	Ki_value	Ki_relation	Ki_unit
17	76.0	=	nM
65	6.56	=	nM
35	100	>	nM
15	8	=	nM
10	95.8	=	nM

Inner Join

molregno	chembl_id	SMILES	Ki_value	Ki_relation	Ki_unit
15	CHEMBL103772	CCCN1CC	8	=	nM
10	CHEMBL328107	C1CN(CCN1)	95.8	=	nM

© 2023 KNIME AG. All rights reserved.

Joining Columns of Data – Left Outer Join

Left Table

molregno	chembl_id	SMILES
22	CHEMBL1794855	CCCN(CCC)
24	CHEMBL278751	CCN(C)
15	CHEMBL103772	CCCN1CC
10	CHEMBL328107	C1CN(CCN1)

Right Table

molregno	Ki_value	Ki_relation	Ki_unit
17	76.0	=	nM
65	6.56	=	nM
35	100	>	nM
15	8	=	nM
10	95.8	=	nM

Left Outer Join

molregno	chembl_id	SMILES	Ki_value	Ki_relation	Ki_unit
22	CHEMBL1794855	CCCN(CCC)	?	?	?
24	CHEMBL278751	CCN(C)	?	?	?
15	CHEMBL103772	CCCN1CC	8	=	nM
10	CHEMBL328107	C1CN(CCN1)	95.8	=	nM

© 2023 KNIME AG. All rights reserved.

Joining Columns of Data – Right Outer Join

Left Table

molregno	chembl_id	SMILES
22	CHEMBL1794855	CCCN(CCC)
24	CHEMBL278751	CCN(C)
15	CHEMBL103772	CCCN1CC
10	CHEMBL328107	C1CN(CCN1)

Right Table

molregno	Ki_value	Ki_relation	Ki_unit
17	76.0	=	nM
65	6.56	=	nM
35	100	>	nM
15	8	=	nM
10	95.8	=	nM

Right Outer Join

molregno	chembl_id	SMILES	Ki_value	Ki_relation	Ki_unit
17	?	?	76.0	=	nM
65	?	?	6.56	=	nM
35	?	?	100	>	nM
15	CHEMBL103772	CCCN1CC	8	=	nM
10	CHEMBL328107	C1CN(CCN1)	95.8	=	nM

© 2023 KNIME AG. All rights reserved.

Joining Columns of Data – Full Outer Join

Left Table

molregno	chembl_id	SMILES
22	CHEMBL1794855	CCCN(CCC)
24	CHEMBL278751	CCN(C)
15	CHEMBL103772	CCCN1CC
10	CHEMBL328107	C1CN(CCN1)

Right Table

molregno	Ki_value	Ki_relation	Ki_unit
17	76.0	=	nM
65	6.56	=	nM
35	100	>	nM
15	8	=	nM
10	95.8	=	nM

Full Outer Join

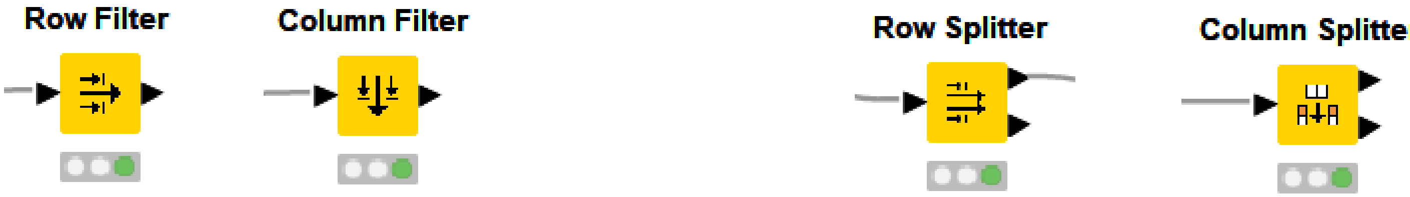
Values missing in the left table

molregno	chembl_id	SMILES	Ki_value	Ki_relation	Ki_unit
17	?	?	76.0	=	nM
65	?	?	6.56	=	nM
35	?	?	100	>	nM
15	CHEMBL103772	CCCN1CC	8	=	nM
10	CHEMBL328107	C1CN(CCN1)	95.8	=	nM
22	CHEMBL1794855	CCCN(CCC)	?	?	?
24	CHEMBL278751	CCN(C)	?	?	?

Values missing in the right table

© 2023 KNIME AG. All rights reserved.

Filter vs. Splitter



- User-defined criteria
- Only one output port
- Excluded rows or columns are not available for downstream processing

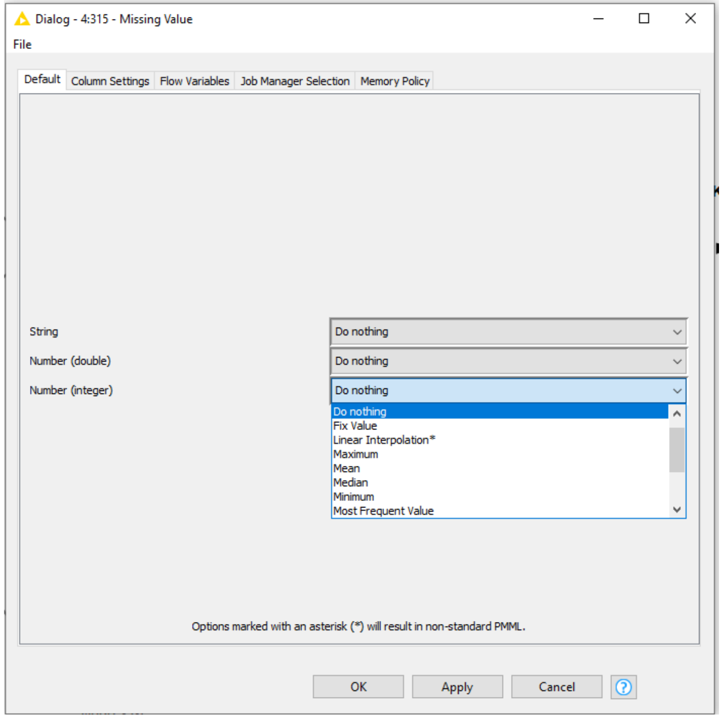
- User-defined criteria
- Two output ports
- Splits a dataset into two
- The ‘excluded’ data is available at the lower output port for downstream processing

© 2023 KNIME AG. All rights reserved.

Missing value handling

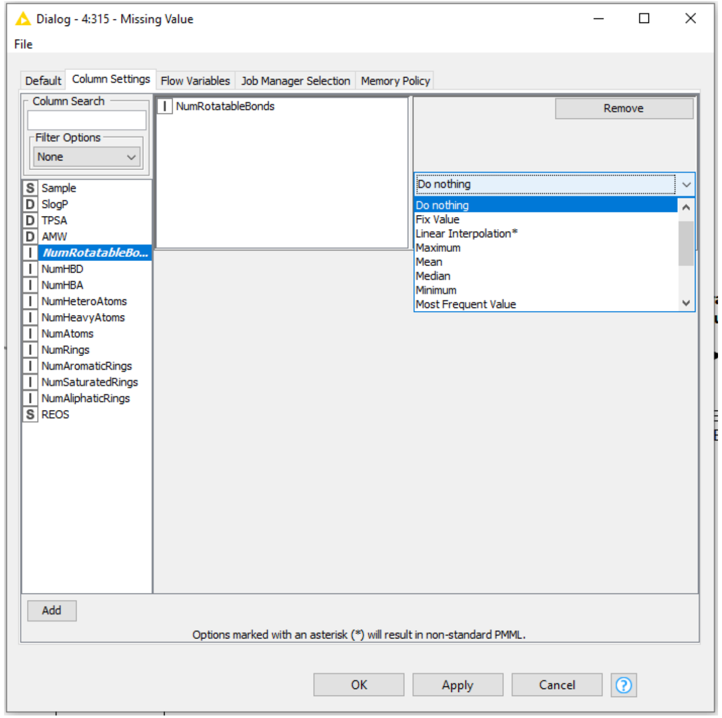
- First tab: define table-wide action depending on data type
- Second tab: define action for individual columns

Missing Value



String: Do nothing
Number (double): Do nothing
Number (integer): Do nothing

Options marked with an asterisk (*) will result in non-standard PMML.



I NumRotatableBonds: Do nothing

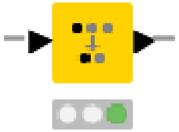
Options marked with an asterisk (*) will result in non-standard PMML.

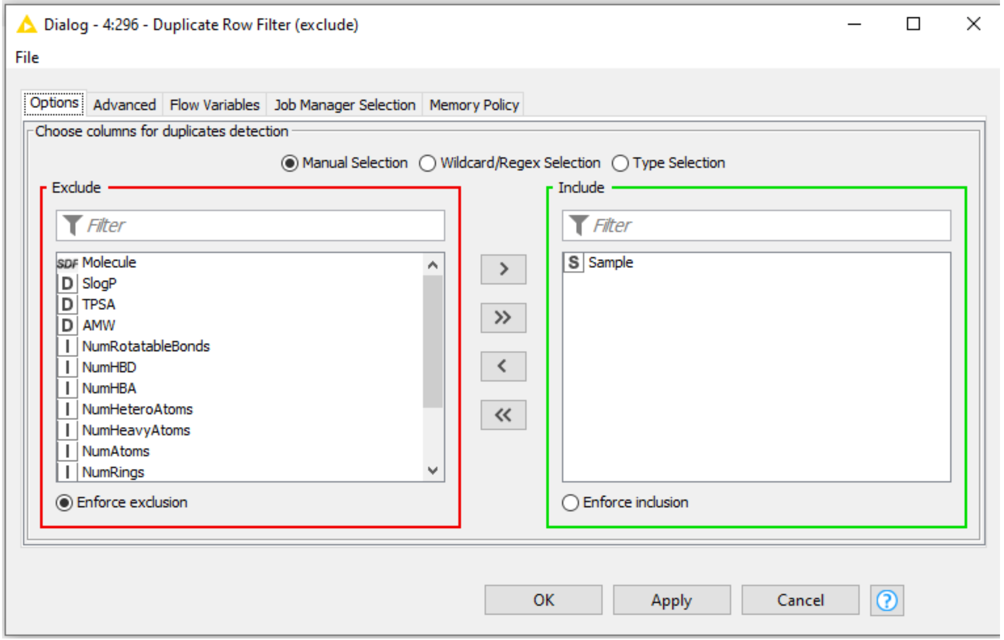
© 2023 KNIME AG. All rights reserved.

Exclude duplicates

- identifies duplicate rows
- duplicate rows have identical values in certain columns
- user defines the column(s) for duplicate detection

Duplicate Row Filter





Choose columns for duplicates detection

● Manual Selection ○ Wildcard/Regex Selection ○ Type Selection

Exclude

- SDF Molecule
- D SlogP
- D TPSA
- D AMW
- I NumRotatableBonds
- I NumHBD
- I NumHBA
- I NumHeteroAtoms
- I NumHeavyAtoms
- I NumAtoms
- I NumRings

● Enforce exclusion

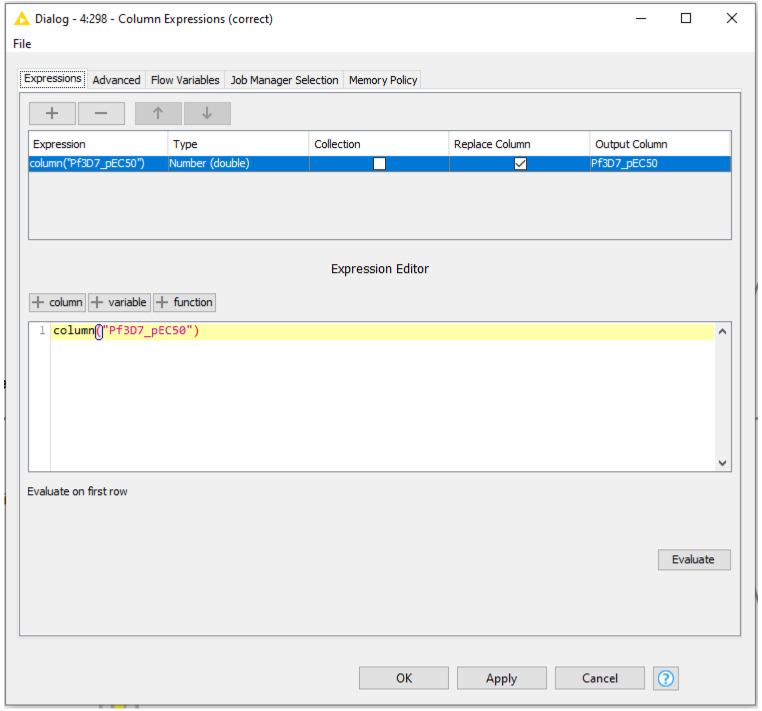
Include

- S Sample

○ Enforce inclusion

Change assigned data type

- modify existing columns using expressions

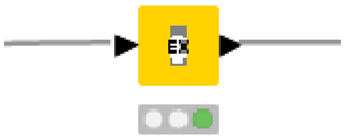


Dialog - 4:298 - Column Expressions (correct)

Expression	Type	Collection	Replace Column	Output Column
column("Pf3D7_pEC50")	Number (double)	☐	☑	Pf3D7_pEC50

Expression Editor

```
column("Pf3D7_pEC50")
```



Column Expressions

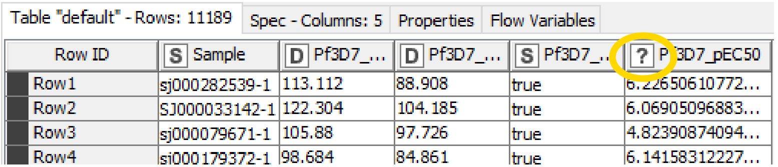


Table "default" - Rows: 11189
Spec - Columns: 5
Properties
Flow Variables

Row ID	S Sample	D Pf3D7_...	D Pf3D7_...	S Pf3D7_...	? Pf3D7_pEC50
Row1	sj000282539-1	113.112	88.908	true	6.22650610772...
Row2	SJ000033142-1	122.304	104.185	true	6.06905096883...
Row3	sj000079671-1	105.88	97.726	true	4.82390874094...
Row4	sj000179372-1	98.684	84.861	true	6.14158312227...

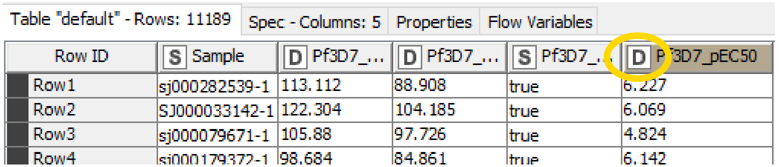


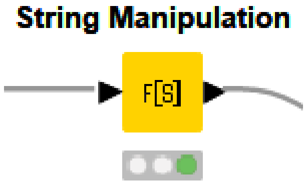
Table "default" - Rows: 11189
Spec - Columns: 5
Properties
Flow Variables

Row ID	S Sample	D Pf3D7_...	D Pf3D7_...	S Pf3D7_...	D Pf3D7_pEC50
Row1	sj000282539-1	113.112	88.908	true	6.227
Row2	SJ000033142-1	122.304	104.185	true	6.069
Row3	sj000079671-1	105.88	97.726	true	4.824
Row4	sj000179372-1	98.684	84.861	true	6.142

© 2023 KNIME AG. All rights reserved.

Capitalize the letters in the sample ID

- Manipulates strings
- Many different functions available
- Can also be used for type converting



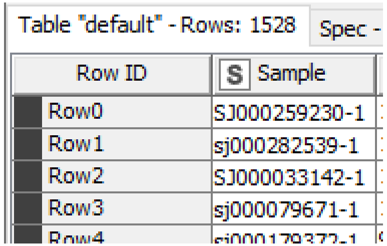
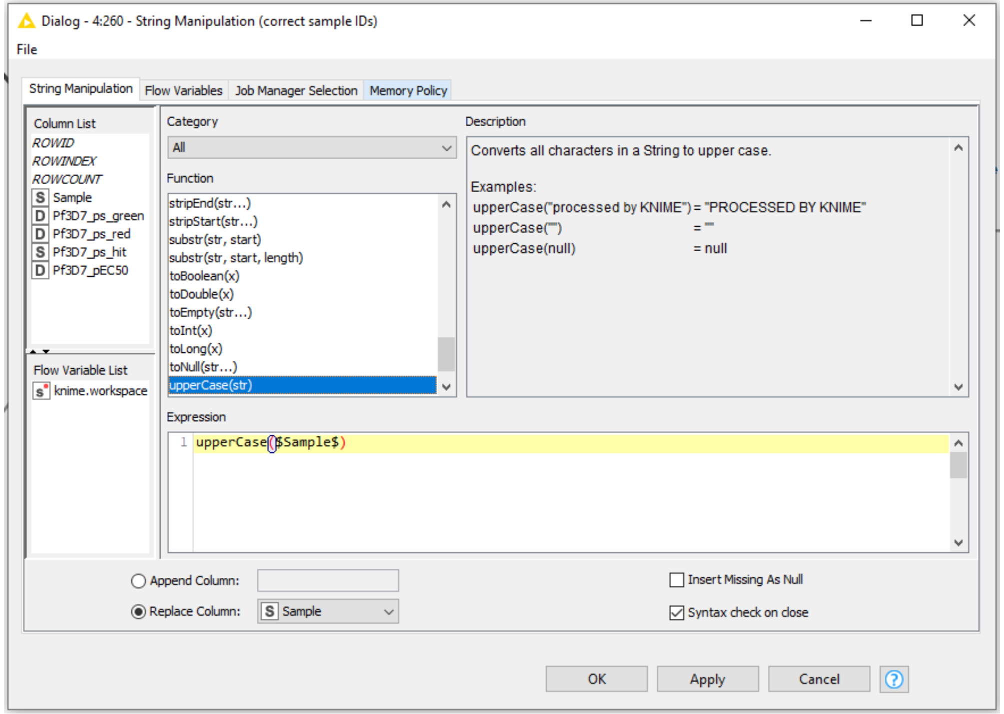


Table "default" - Rows: 1528 | Spec -

Row ID	S Sample
Row0	SJ000259230-1
Row1	sj000282539-1
Row2	SJ000033142-1
Row3	sj000079671-1
Row4	sj000179372-1



Dialog - 4:260 - String Manipulation (correct sample IDs)

Column List
- *ROWID*
- *ROWINDEX*
- *ROWCOUNT*
- S Sample
- D Pf3D7_ps_green
- D Pf3D7_ps_red
- S Pf3D7_ps_hit
- D Pf3D7_pEC50

Flow Variable List
- s knime.workspace

Category: All

Function
- stripEnd(str...)
- stripStart(str...)
- substr(str, start)
- substr(str, start, length)
- toBoolean(x)
- toDouble(x)
- toEmpty(str...)
- toInt(x)
- toLong(x)
- toNull(str...)
- upperCase(str)

Description

Converts all characters in a String to upper case.

Examples:

```
upperCase("processed by KNIME") = "PROCESSED BY KNIME"
upperCase("")                   = ""
upperCase(null)                 = null
```

Expression

```
upperCase($Sample$)
```

○ Append Column:

● Replace Column: S Sample

☐ Insert Missing As Null

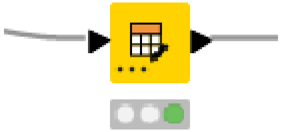
☑ Syntax check on close

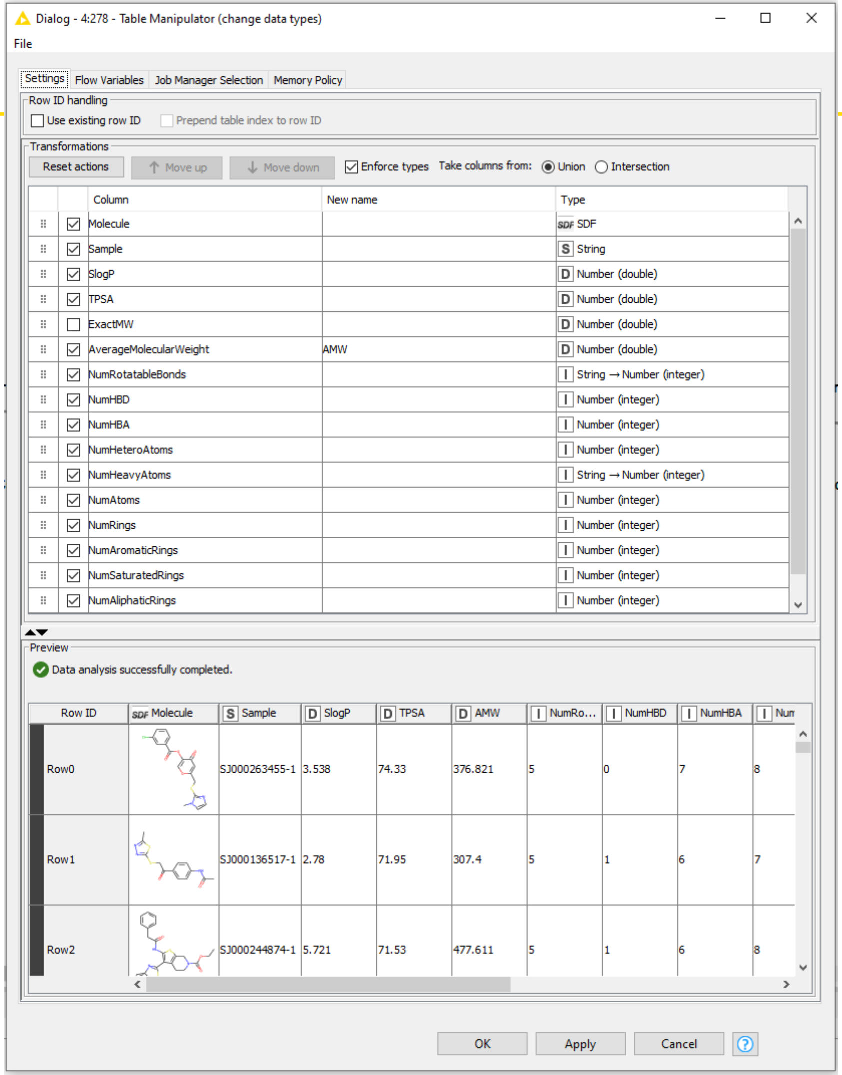
© 2023 KNIME AG. All rights reserved.

Rename column

- Transformations of input columns
  - Renaming
  - Filtering
  - Re-ordering
  - Type changing
- Check data manipulation in the preview

Table Manipulator





Dialog - 4:278 - Table Manipulator (change data types)

Row ID handling

☐ Use existing row ID ☐ Prepend table index to row ID

Transformations

☑ Enforce types Take columns from: ● Union ○ Intersection

	Column	New name	Type
☑	Molecule		SDF SDF
☑	Sample		S String
☑	SlogP		D Number (double)
☑	TPSA		D Number (double)
☐	ExactMW		D Number (double)
☑	AverageMolecularWeight	AMW	D Number (double)
☑	NumRotatableBonds		I String → Number (integer)
☑	NumHBD		I Number (integer)
☑	NumHBA		I Number (integer)
☑	NumHeteroAtoms		I Number (integer)
☑	NumHeavyAtoms		I String → Number (integer)
☑	NumAtoms		I Number (integer)
☑	NumRings		I Number (integer)
☑	NumAromaticRings		I Number (integer)
☑	NumSaturatedRings		I Number (integer)
☑	NumAliphaticRings		I Number (integer)

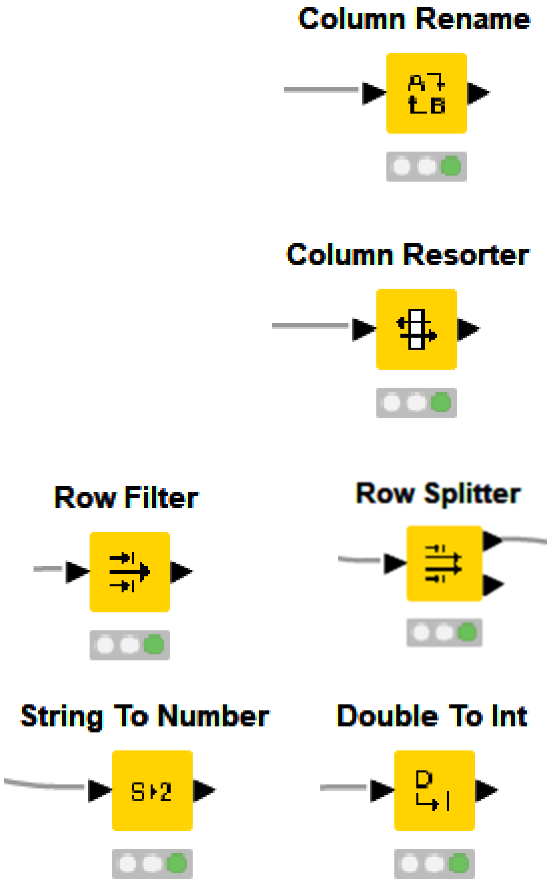
Preview

Data analysis successfully completed.

Row ID	SDF Molecule	S Sample	D SlogP	D TPSA	D AMW	I NumRo...	I NumHBD	I NumHBA	I Nu
Row0		SJ000263455-1	3.538	74.33	376.821	5	0	7	8
Row1		SJ000136517-1	2.78	71.95	307.4	5	1	6	7
Row2		SJ000244874-1	5.721	71.53	477.611	5	1	6	8

© 2023 KNIME AG. All rights reserved.

There isn't just one way of doing it…

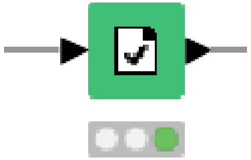


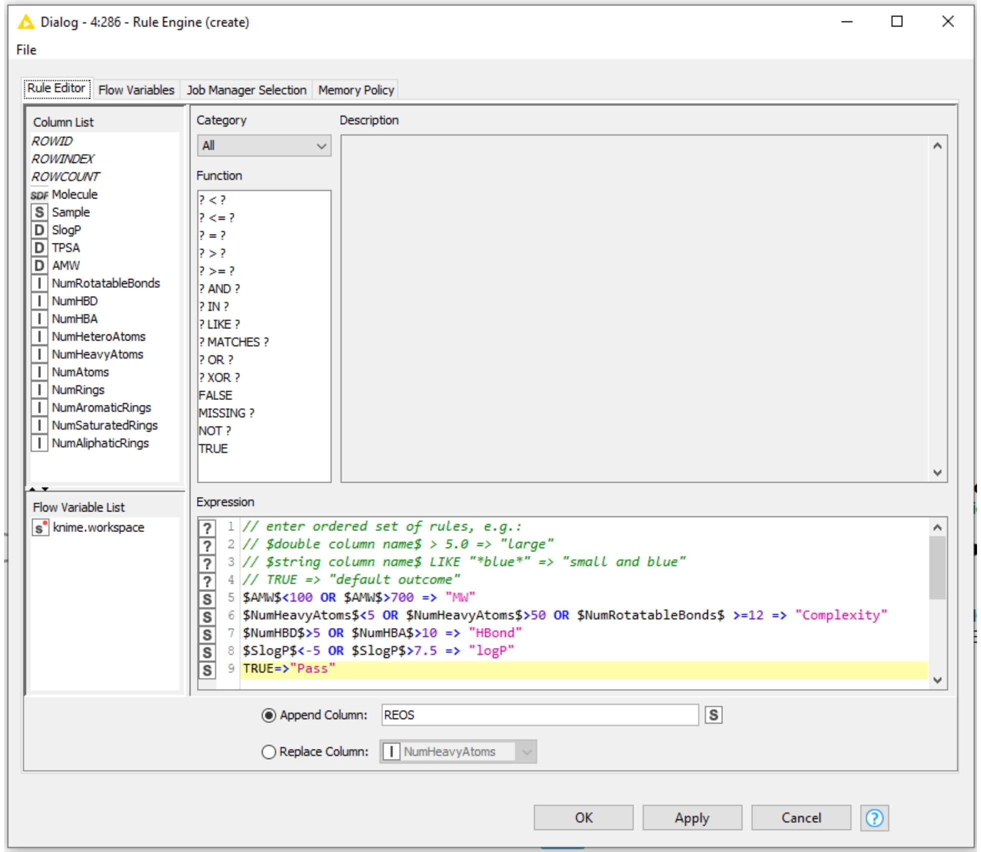
- Rename columns
- Change the column order
- Exclude missing values
- Convert data types

© 2023 KNIME AG. All rights reserved.

Create classification REOS rules

Rule Engine





```
// enter ordered set of rules, e.g.:
// $double column name$ > 5.0 => "large"
// $string column name$ LIKE "*blue*" => "small and blue"
// TRUE => "default outcome"
$AMW$<100 OR $AMW$>700 => "MW"
$NumHeavyAtoms$<5 OR $NumHeavyAtoms$>50 OR $NumRotatableBonds$ >=12 => "Complexity"
$NumHBD$>5 OR $NumHBA$>10 => "HBond"
$SlogP$<-5 OR $SlogP$>7.5 => "logP"
TRUE=>"Pass"
```

● Append Column: REOS

○ Replace Column: NumHeavyAtoms

Define Condition => *Outcome if condition is met*

If not do this

IF… THEN… ELSE

GroupBy node

- Aggregate rows to summarize data
- Different aggregation methods available

Product ID	**REOS**	**AMW**
P001	Pass	462.55
P002	Pass	397.92
P003	Complexity	431.28
P004	Pass	431.28
P005	MW	371.46
P006	logP	389.20

Group	**Count*(AMW)**
Pass	3
Complexity	1
MW	1
logP	1

© 2023 KNIME AG. All rights reserved.

GroupBy node

- Aggregate rows to summarize data
- Different aggregation methods available

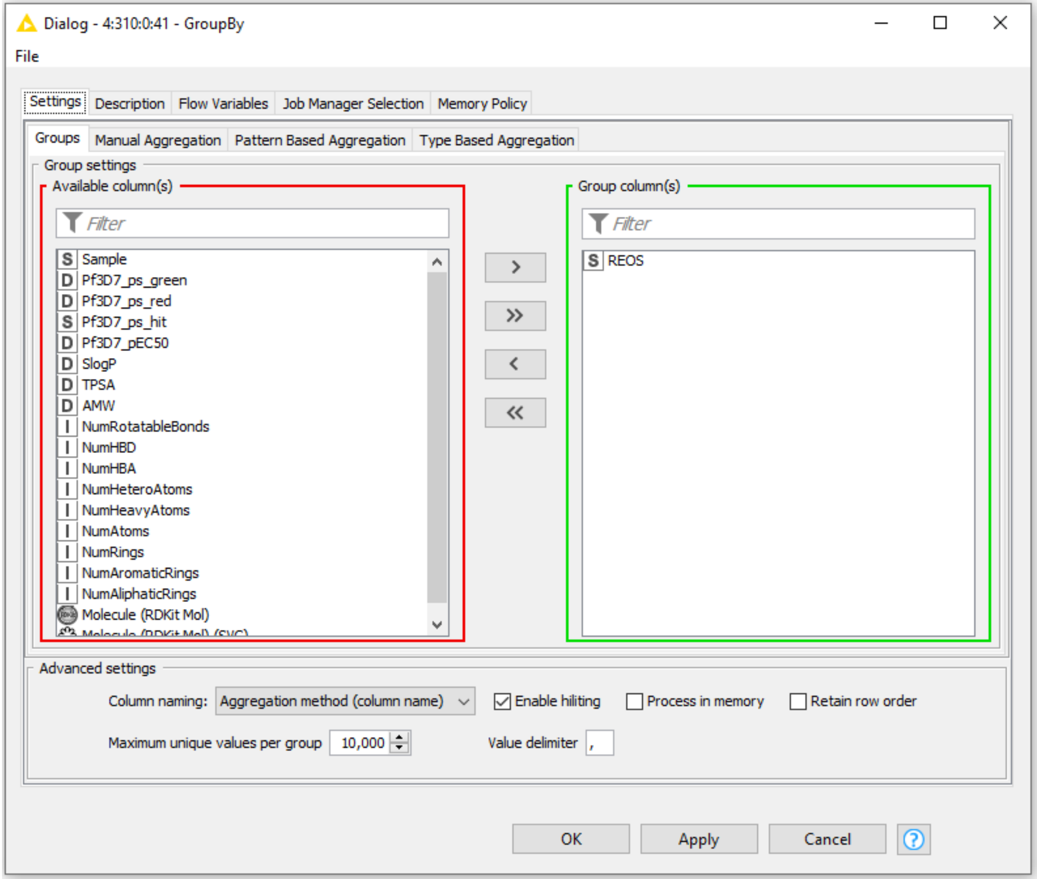
Product ID	REOS	AMW
P001	Pass	462.55
P002	Pass	397.92
P003	Complexity	431.28
P004	Pass	431.28
P005	MW	371.46
P006	logP	389.20

Group	Mean(AMW)
Pass	430.58
Complexity	431.28
MW	371.46
logP	389.20

© 2023 KNIME AG. All rights reserved.

Groupby node configuration

- Define the column to group on in the first tab



Group settings

Available column(s)

- S Sample
- D Pf3D7_ps_green
- D Pf3D7_ps_red
- S Pf3D7_ps_hit
- D Pf3D7_pEC50
- D SlogP
- D TPSA
- D AMW
- I NumRotatableBonds
- I NumHBD
- I NumHBA
- I NumHeteroAtoms
- I NumHeavyAtoms
- I NumAtoms
- I NumRings
- I NumAromaticRings
- I NumAliphaticRings
- Molecule (RDKit Mol)

Group column(s)

- S REOS

Advanced settings

Column naming: Aggregation method (column name)

☑ Enable hiliting ☐ Process in memory ☐ Retain row order

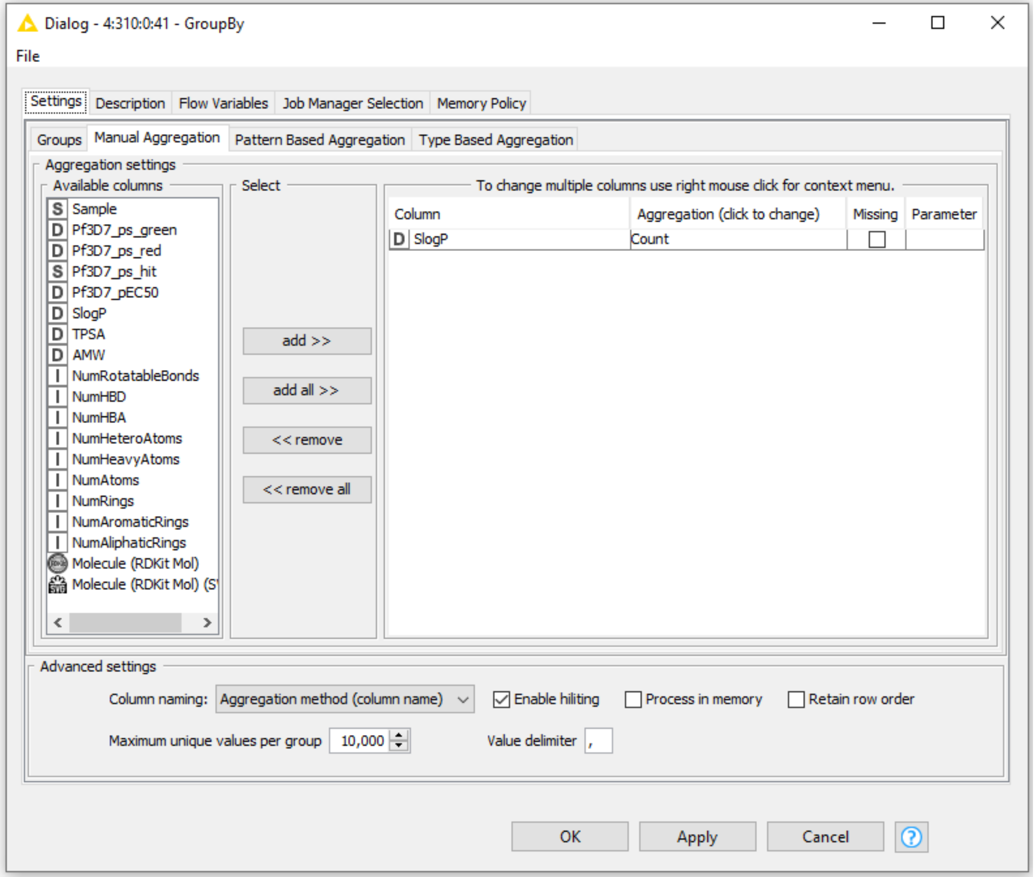
Maximum unique values per group 10,000

Value delimiter ,

Groupby configuration

- Define the aggregation method in the second tab

- Information about the methods is provided in the Description tab



Available columns:

- S Sample
- D Pf3D7_ps_green
- D Pf3D7_ps_red
- S Pf3D7_ps_hit
- D Pf3D7_pEC50
- D SlogP
- D TPSA
- D AMW
- I NumRotatableBonds
- I NumHBD
- I NumHBA
- I NumHeteroAtoms
- I NumHeavyAtoms
- I NumAtoms
- I NumRings
- I NumAromaticRings
- I NumAliphaticRings
- Molecule (RDKit Mol)
- Molecule (RDKit Mol) (S

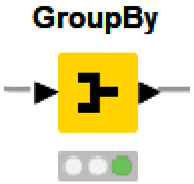
To change multiple columns use right mouse click for context menu.

Column	Aggregation (click to change)	Missing	Parameter
D SlogP	Count	☐	

Advanced settings

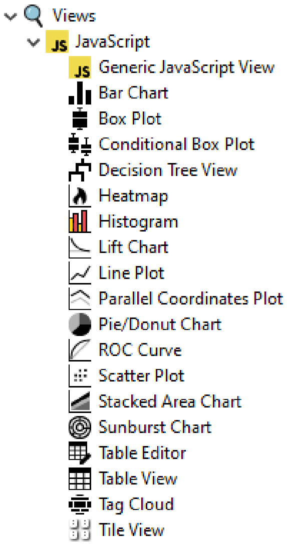
Column naming: Aggregation method (column name) ☑ Enable hiliting ☐ Process in memory ☐ Retain row order

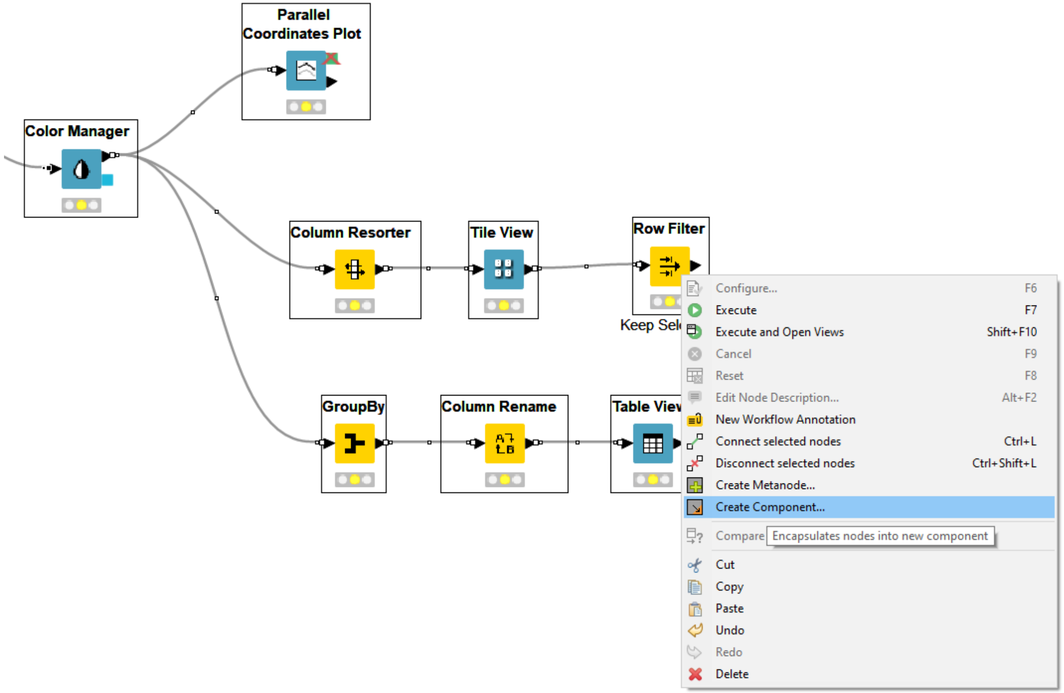
Maximum unique values per group 10,000 Value delimiter ,



Visualization in components

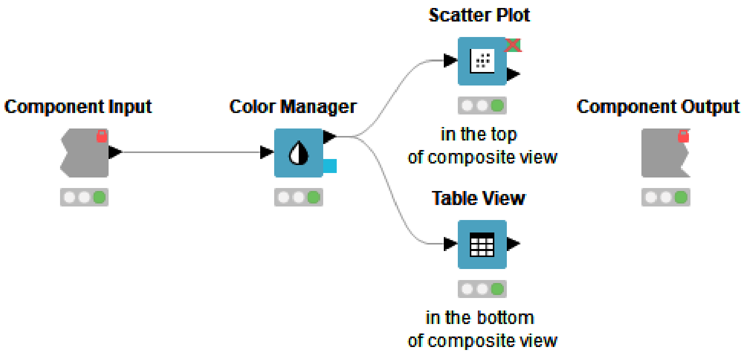
- Mark the view nodes you want to combine
- Right-click > create component
- Use the Javascript-based nodes





Components – Combined Views

- Multiple JavaScript View nodes can be combined in components
- Selections are transmitted to all other views
- Also for use on the KNIME WebPortal (commercial product)



Scatter Plot

in the top
of composite view

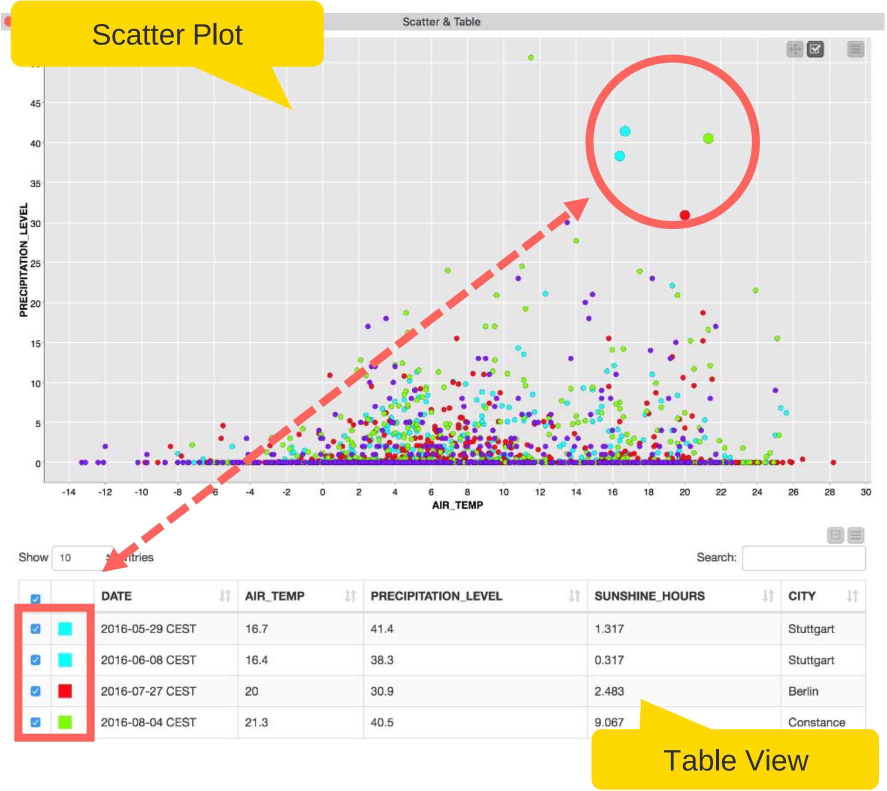
Component Input

Color Manager

Component Output

Table View

in the bottom
of composite view



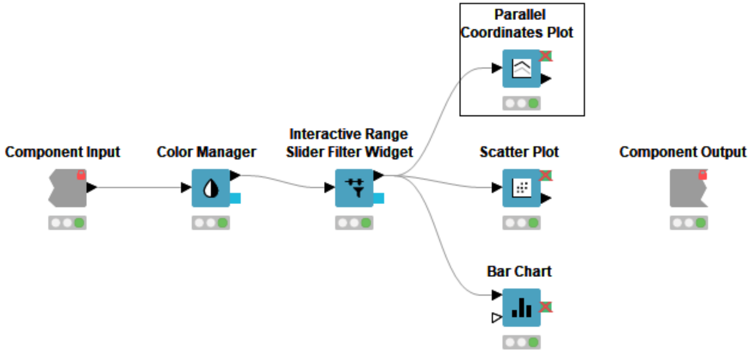
Show 10 entries

Search:

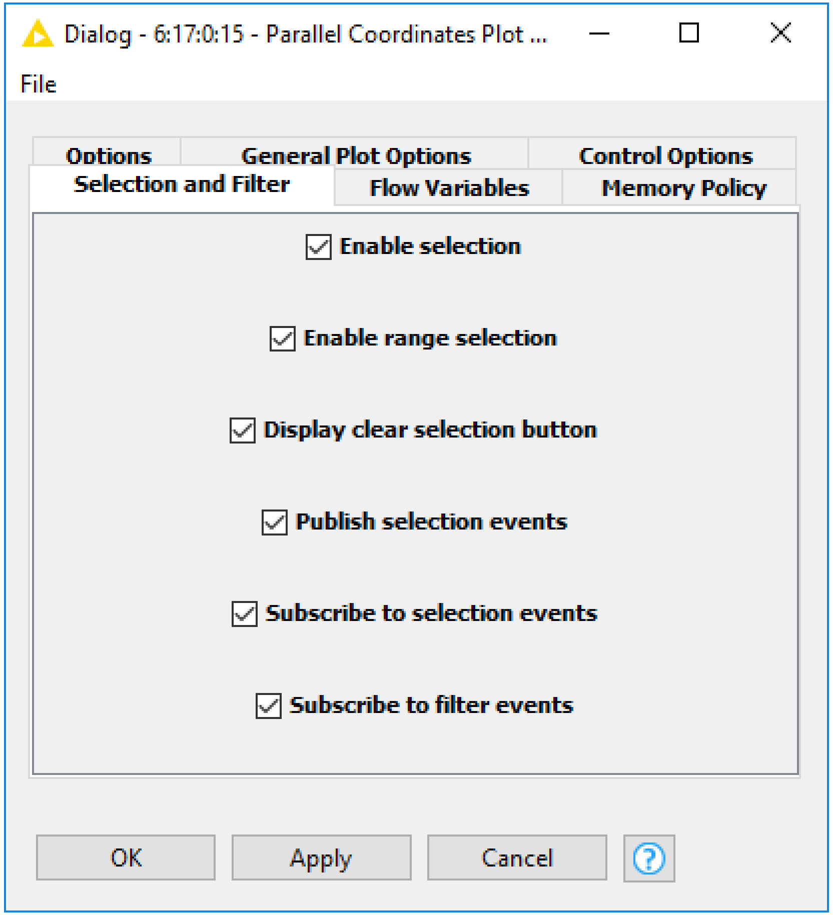
☑		DATE	AIR_TEMP	PRECIPITATION_LEVEL	SUNSHINE_HOURS	CITY
☑	🟦	2016-05-29 CEST	16.7	41.4	1.317	Stuttgart
☑	🟦	2016-06-08 CEST	16.4	38.3	0.317	Stuttgart
☑	🟥	2016-07-27 CEST	20	30.9	2.483	Berlin
☑	🟩	2016-08-04 CEST	21.3	40.5	9.067	Constance

© 2023 KNIME AG. All rights reserved.

JS-based nodes: Selection and Filter Events



- Selection and Filter tab in many Javascript-based view nodes



☑ Enable selection

☑ Enable range selection

☑ Display clear selection button

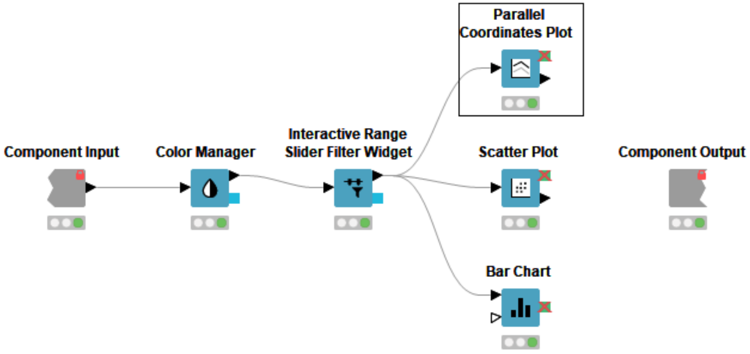
☑ Publish selection events

☑ Subscribe to selection events

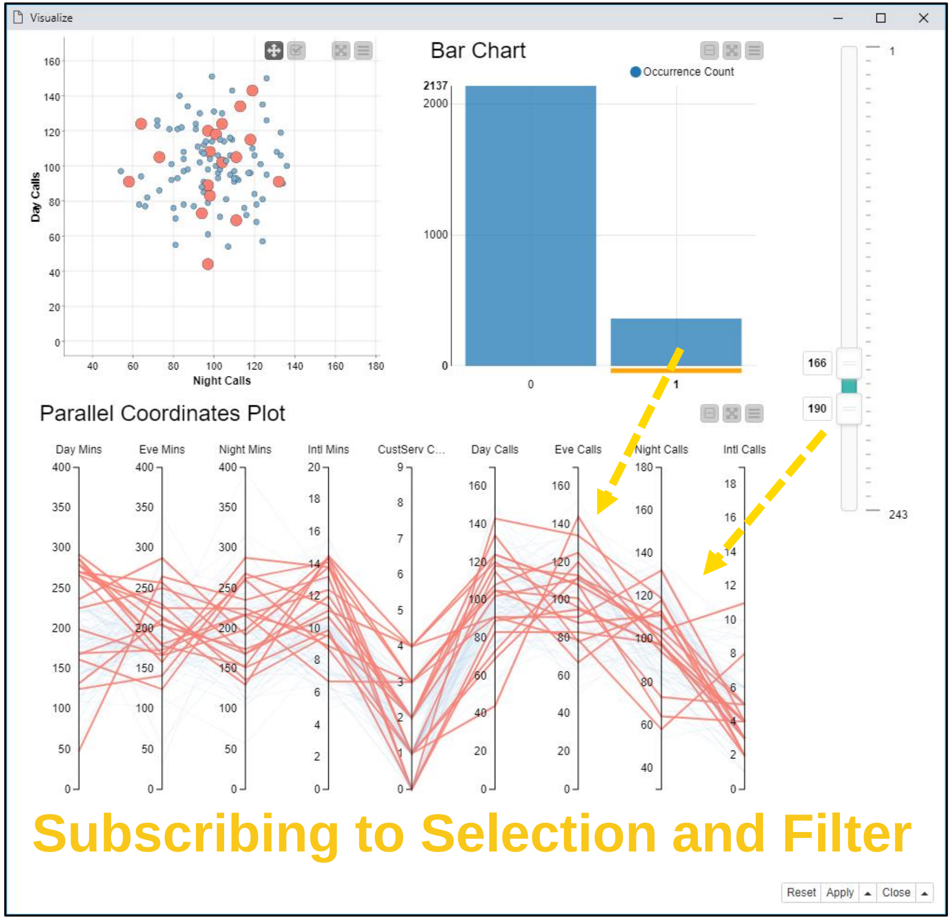
☑ Subscribe to filter events

© 2023 KNIME AG. All rights reserved.

JS-based nodes: Selection and Filter Events

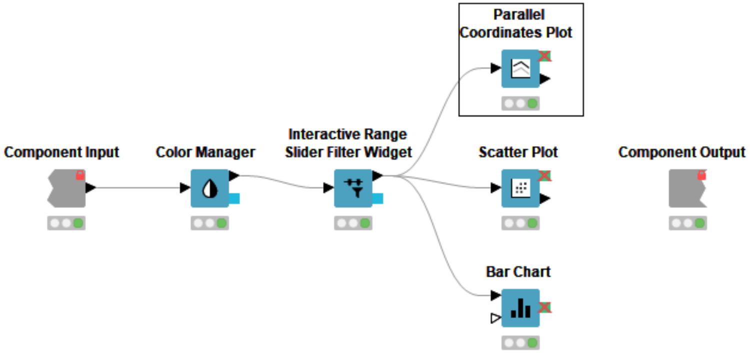


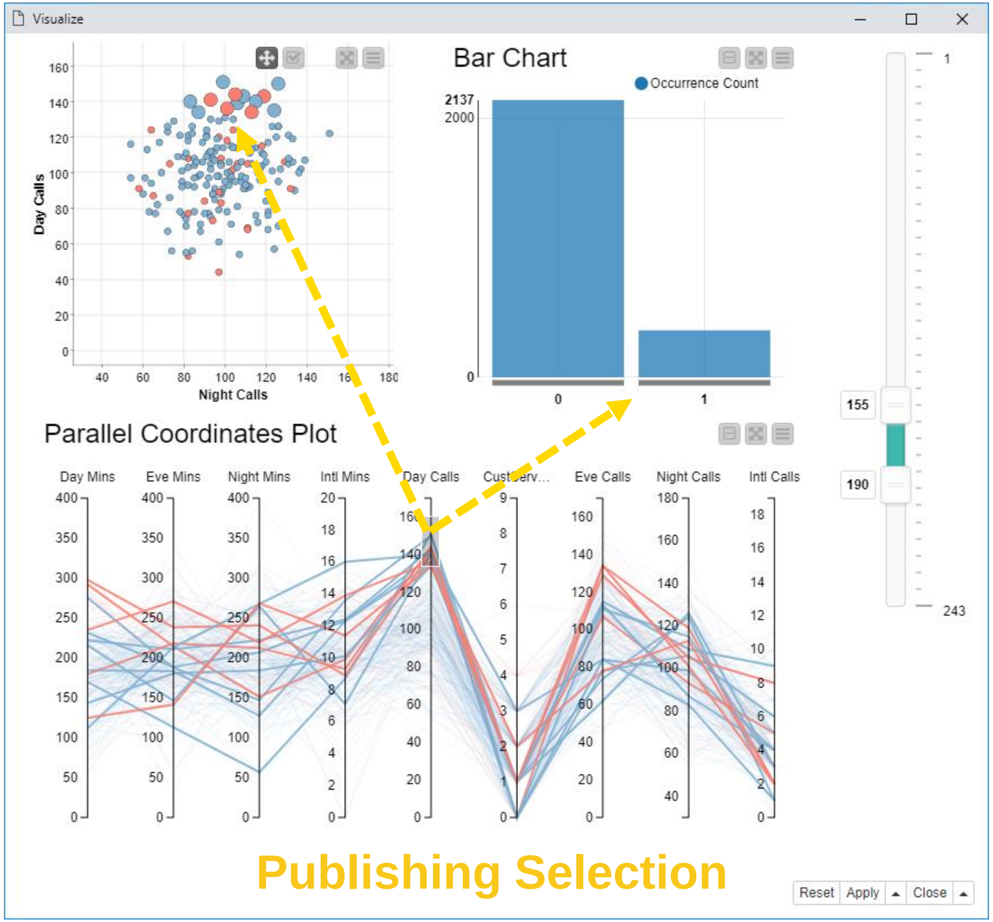
- Listens to the selection made in other view



Subscribing to Selection and Filter

JS-based nodes: Selection and Filter Events





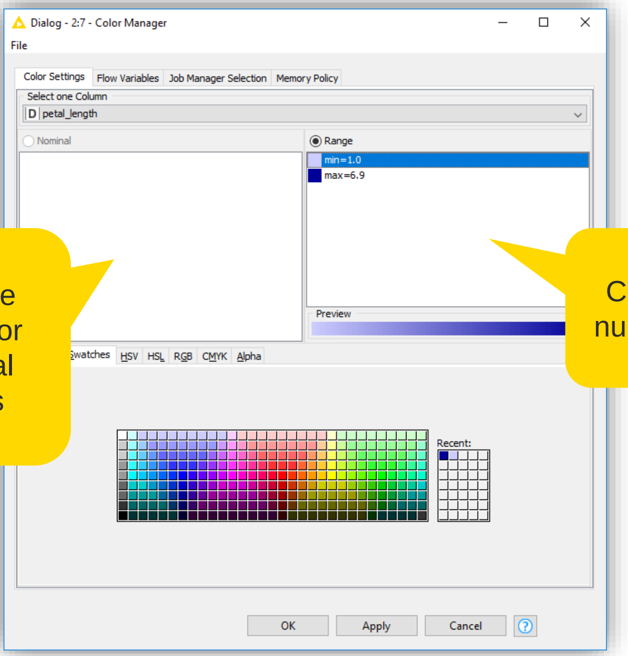
Publishing Selection

- Makes the other views listen to the selection made in this view

© 2023 KNIME AG. All rights reserved.

Color Manager

- Colors by nominal or continuous values
- Syncs colors between views using the color model port and Color Appender node



Select one Column: D petal_length

○ Nominal

● Range

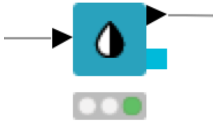
min=1.0

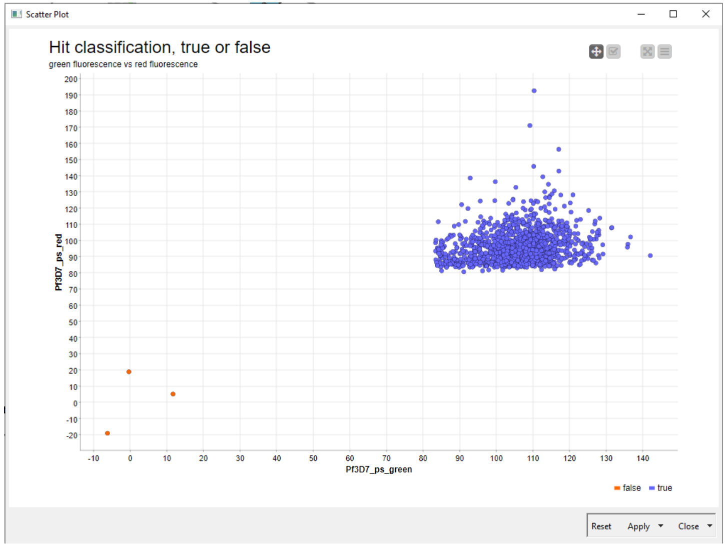
max=6.9

Discrete colors for nominal values

Color range for numerical values

Color Manager



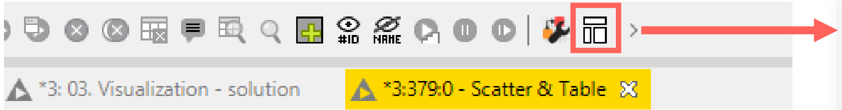


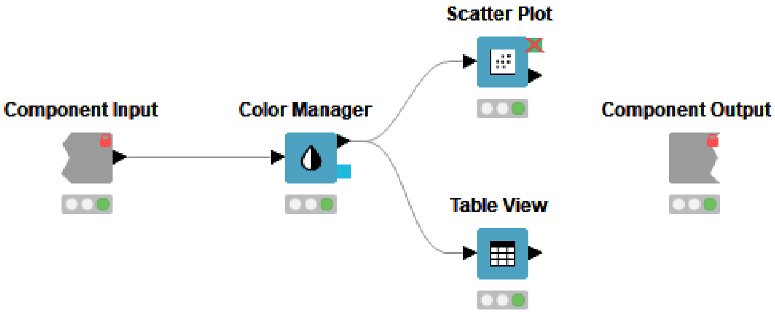
© 2023 KNIME AG. All rights reserved.

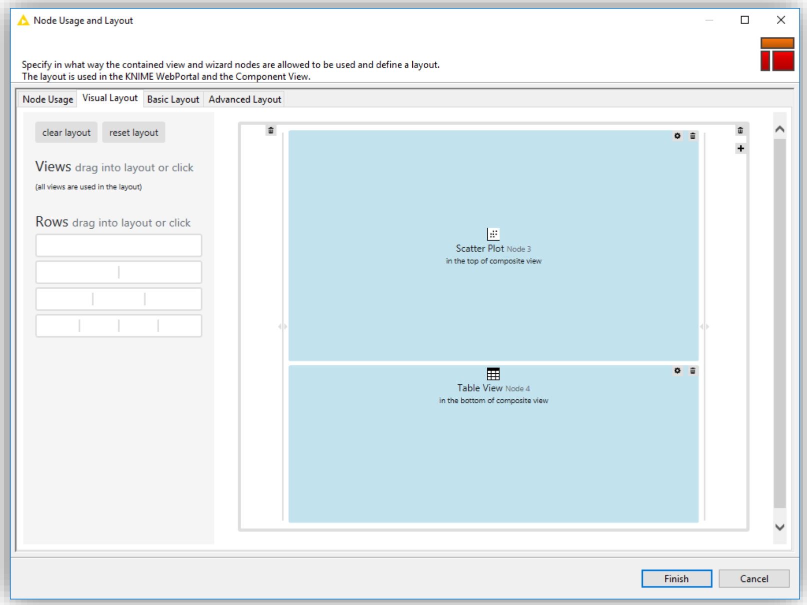
Configure Content and Views Layout

- Click layout button when inside Component to assign views to rows and columns

- Add views and rows via drag&drop
- Add columns using **+** buttons







Specify in what way the contained view and wizard nodes are allowed to be used and define a layout.
The layout is used in the KNIME WebPortal and the Component View.

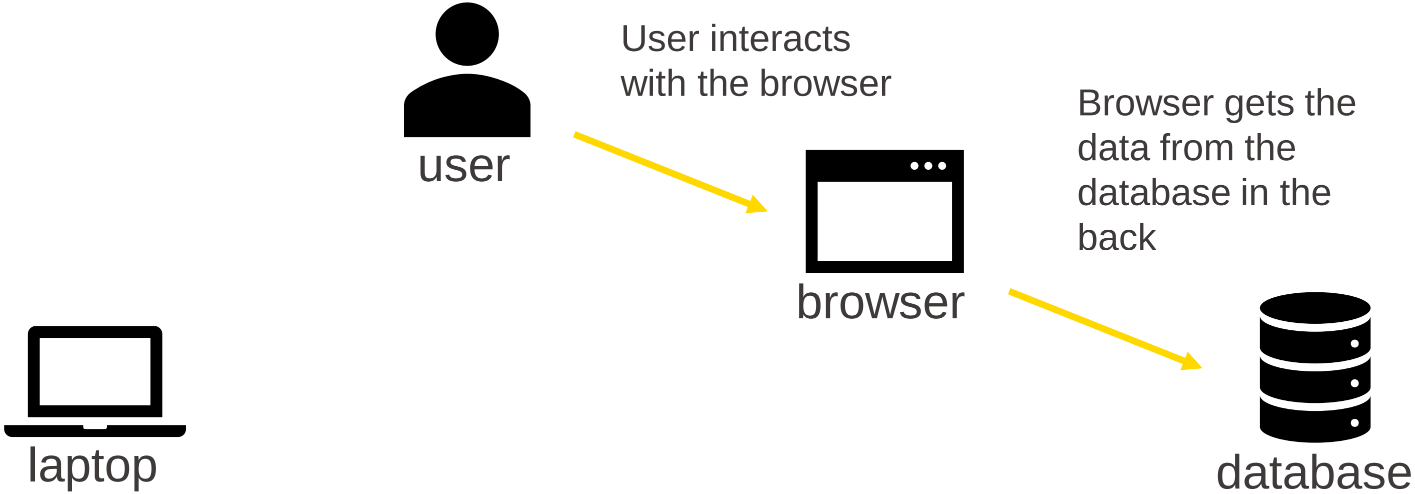
Node Usage
Visual Layout
Basic Layout
Advanced Layout
clear layout
reset layout
Views drag into layout or click
(all views are used in the layout)
Rows drag into layout or click
Scatter Plot Node 3
in the top of composite view
Table View Node 4
in the bottom of composite view
Finish
Cancel

That’s it for workflow I…

Workflow II – web service data retrieval

- Instead of retrieving data manually, I can do it in an automated way
- Especially useful for large amounts of data
- Many data sources offer the 'endpoint' to do so (→ REST API, web service)

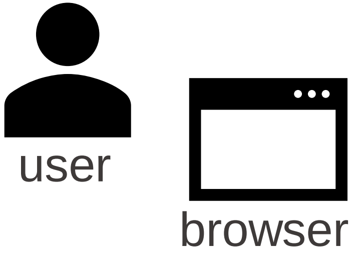
MANUALLY

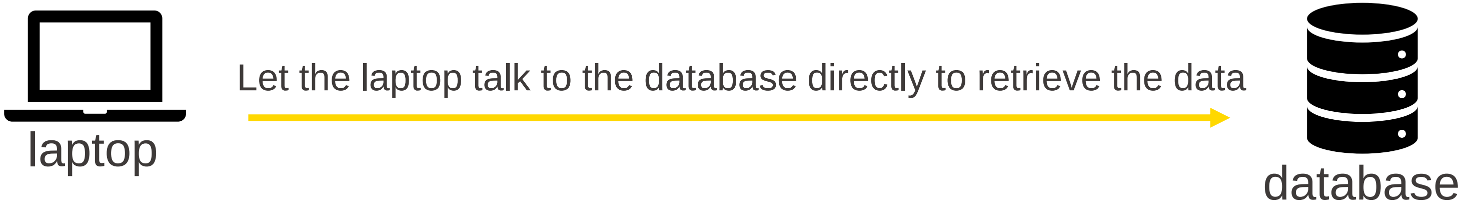


Workflow II – web service data retrieval

- Instead of retrieving data manually, I can do it in an automated way
- Especially useful for large amounts of data
- Many data sources offer the ‘endpoint’ to do so (→ REST API, web service)

AUTOMATICALLY



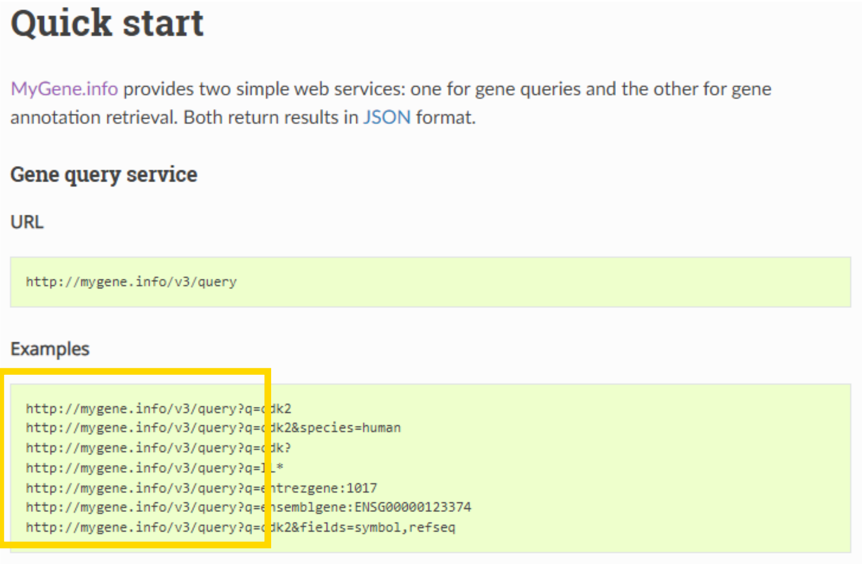

laptop
Let the laptop talk to the database directly to retrieve the data
database

© 2023 KNIME AG. All rights reserved.

Workflow II – web service data retrieval

- Examples of data sources with an API:
  - https://docs.mygene.info/en/latest/index.html
  - https://chembl.gitbook.io/chembl-interface-documentation/web-services/chembl-data-web-services
- Construct the according URL to retrieve data for according query

Steady part



Quick start

MyGene.info provides two simple web services: one for gene queries and the other for gene annotation retrieval. Both return results in JSON format.

Gene query service

URL

```
http://mygene.info/v3/query
```

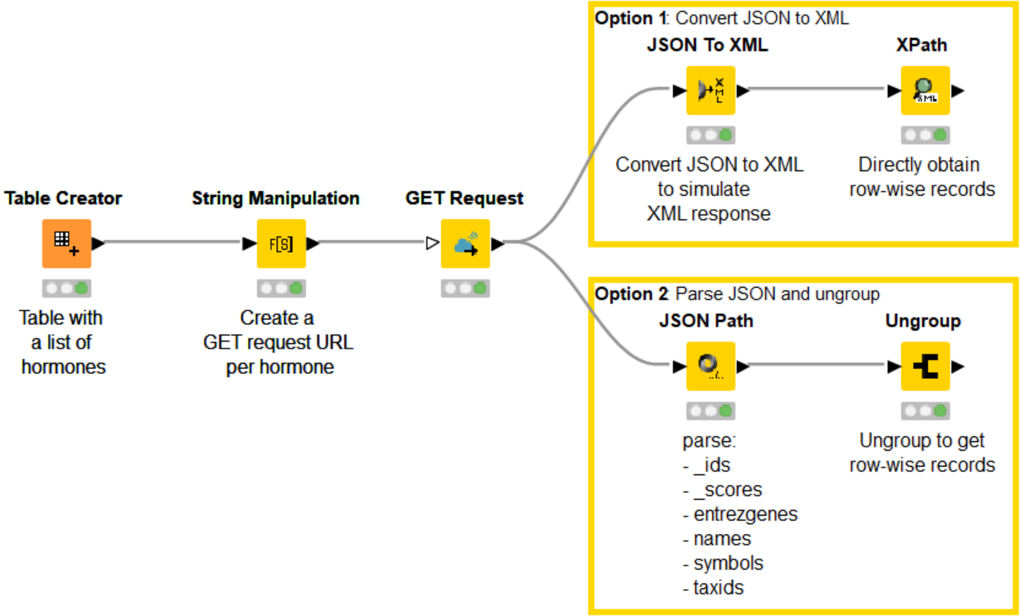
Examples

```
http://mygene.info/v3/query?q=cdk2
http://mygene.info/v3/query?q=cdk2&species=human
http://mygene.info/v3/query?q=cdk?
http://mygene.info/v3/query?q=IL*
http://mygene.info/v3/query?q=entrezgene:1017
http://mygene.info/v3/query?q=ensemblgene:ENSG00000123374
http://mygene.info/v3/query?q=cdk2&fields=symbol,refseq
```

© 2023 KNIME AG. All rights reserved.

Workflow II – web service data retrieval

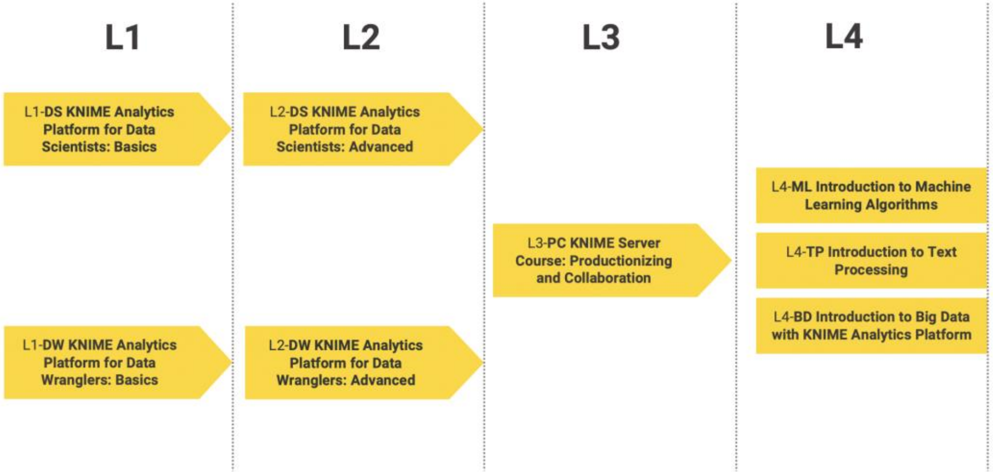
- Join the steady part of the URL with your query using the String Manipulation
- The GET Request node ‘talks’ to the DB and retrieves the data
- The result is often not very human-friendly (JSON or XML data format)
- Parse the results to get a
  neat table


Option 1: Convert JSON to XML
JSON To XML
XPath
Convert JSON to XML to simulate XML response
Directly obtain row-wise records
Table Creator
String Manipulation
GET Request
Table with a list of hormones
Create a GET request URL per hormone
Option 2 Parse JSON and ungroup
JSON Path
Ungroup
parse:
- _ids
- _scores
- entrezgenes
- names
- symbols
- taxids
Ungroup to get row-wise records

That’s it for workflow II…

KNIME courses

- **Self-paced courses**
  - Courses are organized by level L1 (basic) - L4 (specialized)
  - lessons with ~5 minutes videos, hands-on exercises, and knowledge-check questions
- **Instructor-lead online courses**
  - One week, 1h15 per day
  - Possibility to ask questions
  - Discount available for academics



L1

L1-DS KNIME Analytics Platform for Data Scientists: Basics

L1-DW KNIME Analytics Platform for Data Wranglers: Basics

L2

L2-DS KNIME Analytics Platform for Data Scientists: Advanced

L2-DW KNIME Analytics Platform for Data Wranglers: Advanced

L3

L3-PC KNIME Server Course: Productionizing and Collaboration

L4

L4-ML Introduction to Machine Learning Algorithms

L4-TP Introduction to Text Processing

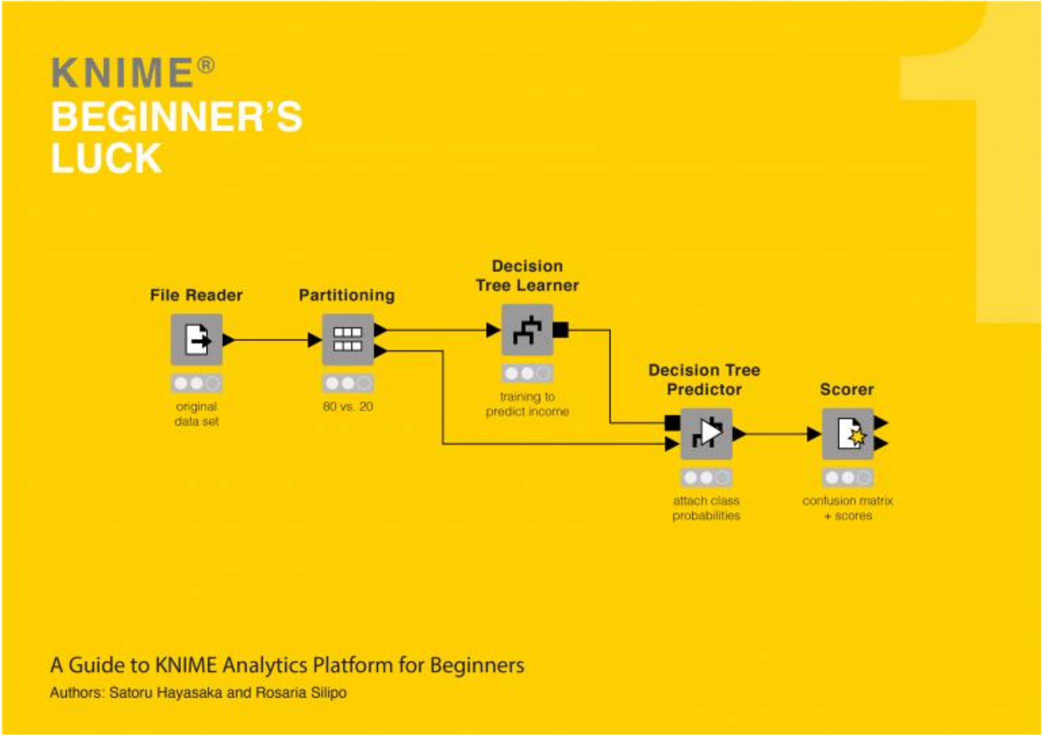
L4-BD Introduction to Big Data with KNIME Analytics Platform

https://www.knime.com/knime-courses

© 2023 KNIME AG. All rights reserved.

KNIME book

- KNIME Beginners Luck



KNIME®
BEGINNER'S
LUCK

A Guide to KNIME Analytics Platform for Beginners

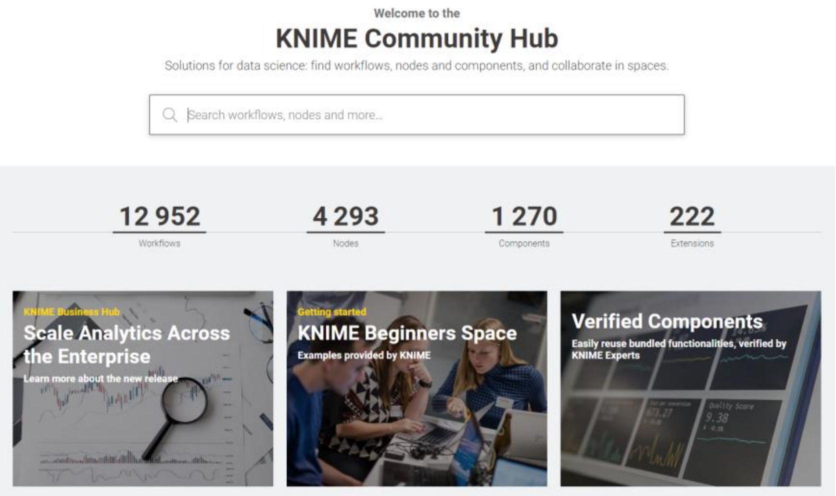
Authors: Satoru Hayasaka and Rosaria Silipo

https://www.knime.com/knimepress/beginners-luck

© 2023 KNIME AG. All rights reserved.

More resources

- **KNIME Community Hub**
  - Find example workflows you can adapt
  - Find out e.g. how others use a certain nodes



Welcome to the

KNIME Community Hub

Solutions for data science: find workflows, nodes and components, and collaborate in spaces.

Search workflows, nodes and more...

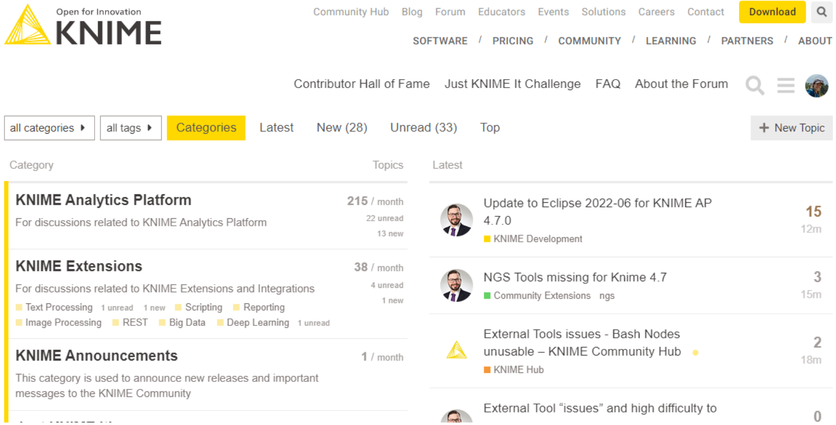
12 952	4 293	1 270	222
Workflows	Nodes	Components	Extensions

KNIME Business Hub
Scale Analytics Across the Enterprise
Learn more about the new release

Getting started
KNIME Beginners Space
Examples provided by KNIME

Verified Components
Easily reuse bundled functionalities, verified by KNIME Experts

- **KNIME Forum**
  - Get and find answers to your questions



Open for Innovation
KNIME

Community Hub Blog Forum Educators Events Solutions Careers Contact Download

SOFTWARE / PRICING / COMMUNITY / LEARNING / PARTNERS / ABOUT

Contributor Hall of Fame Just KNIME It Challenge FAQ About the Forum

all categories ▸ all tags ▸ Categories Latest New (28) Unread (33) Top + New Topic

Category	Topics
KNIME Analytics Platform For discussions related to KNIME Analytics Platform	215 / month 22 unread 13 new
KNIME Extensions For discussions related to KNIME Extensions and Integrations Text Processing 1 unread 1 new · Scripting · Reporting · Image Processing · REST · Big Data · Deep Learning 1 unread	38 / month 4 unread 1 new
KNIME Announcements This category is used to announce new releases and important messages to the KNIME Community	1 / month

Latest	
Update to Eclipse 2022-06 for KNIME AP 4.7.0 KNIME Development	15 12m
NGS Tools missing for Knime 4.7 Community Extensions ngs	3 15m
External Tools issues - Bash Nodes unusable – KNIME Community Hub KNIME Hub	2 18m
External Tool "issues" and high difficulty to	0

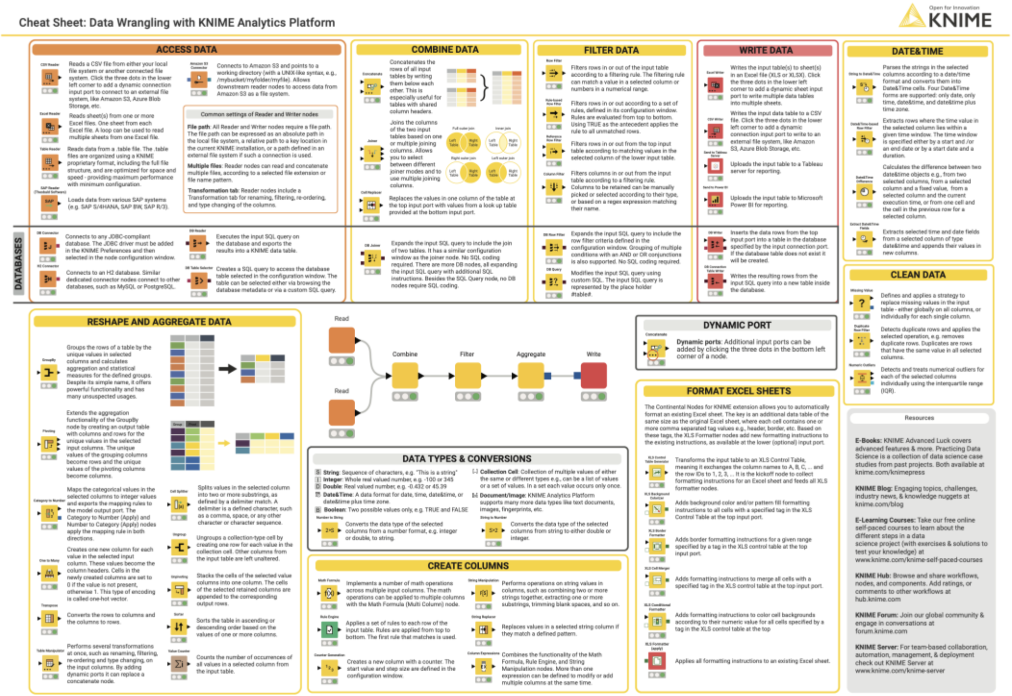
… and **Youtube**
https://www.youtube.com/knimetv

© 2023 KNIME AG. All rights reserved.

Cheat Sheets

- Building a KNIME Workflow for Beginners
- Building Components for your Team or the KNIME Community
- Control and Orchestration with KNIME AP
- Data Wrangling with KNIME AP
- Connectors with KNIME AP
- Machine Learning with KNIME AP

https://www.knime.com/cheat-sheets



Cheat Sheet: Data Wrangling with KNIME Analytics Platform

ACCESS DATA

CSV Reader	Reads a CSV file from either your local file system or another connected file system. Click the three dots in the lower left corner to add a dynamic connection input port to connect to an external file system, like Amazon S3, Azure Blob Storage, etc.
Excel Reader	Reads sheet(s) from one or more Excel files. One sheet from each Excel file. A loop can be used to read multiple sheets from one Excel file.
Table Reader	Reads data from a .table file. The .table files are organized using a KNIME proprietary format, including the full file structure, and are optimized for space and speed - providing maximum performance with minimum configuration.
SAP Reader (Theobald Software)	Loads data from various SAP systems (e.g. SAP S/4HANA, SAP BW, SAP R/3).
Amazon S3 Connector	Connects to Amazon S3 and points to a working directory (with a UNIX-like syntax, e.g., /mybucket/myfolder/myfile). Allows downstream reader nodes to access data from Amazon S3 as a file system.

Common settings of Reader and Writer nodes

File path: All Reader and Writer nodes require a file path. The file path can be expressed as an absolute path in the local file system, a relative path to a key location in the current KNIME installation, or a path defined in an external file system if such a connection is used.

Multiple files: Reader nodes can read and concatenate multiple files, according to a selected file extension or file name pattern.

Transformation tab: Reader nodes include a Transformation tab for renaming, filtering, re-ordering, and type changing of the columns.

COMBINE DATA

Concatenate	Concatenates the rows of all input tables by writing them below each other. This is especially useful for tables with shared column headers.
Joiner	Joins the columns of the two input tables based on one or multiple joining columns. Allows you to select between different joiner modes and to use multiple joining columns.
Cell Replacer	Replaces the values in one column of the table at the top input port with values from a look up table provided at the bottom input port.

FILTER DATA

Row Filter	Filters rows in or out of the input table according to a filtering rule. The filtering rule can match a value in a selected column or numbers in a numerical range.
Rule-based Row Filter	Filters rows in or out according to a set of rules, defined in its configuration window. Rules are evaluated from top to bottom. Using TRUE as the antecedent applies the rule to all unmatched rows.
Reference Row Filter	Filters rows in or out from the top input table according to matching values in the selected column of the lower input table.
Column Filter	Filters columns in or out from the input table according to a filtering rule. Columns to be retained can be manually picked or selected according to their type, or based on a regex expression matching their name.

WRITE DATA

Excel Writer	Writes the input table(s) to sheet(s) in an Excel file (XLS or XLSX). Click the three dots in the lower left corner to add a dynamic sheet input port to write multiple data tables into multiple sheets.
CSV Writer	Writes the input data table to a CSV file. Click the three dots in the lower left corner to add a dynamic connection input port to write to an external file system, like Amazon S3, Azure Blob Storage, etc.
Send to Tableau Server	Uploads the input table to a Tableau server for reporting.
Send to Power BI	Uploads the input table to Microsoft Power BI for reporting.

DATABASES

DB Connector	Connects to any JDBC-compliant database. The JDBC driver must be added in the KNIME Preferences and then selected in the node configuration window.
H2 Connector	Connects to an H2 database. Similar dedicated connector nodes connect to other databases, such as MySQL or PostgreSQL.
DB Reader	Executes the input SQL query on the database and exports the results into a KNIME data table.
DB Table Selector	Creates a SQL query to access the database table selected in the configuration window. The table can be selected either via browsing the database metadata or via a custom SQL query.
DB Joiner	Expands the input SQL query to include the join of two tables. It has a similar configuration window as the joiner node. No SQL coding required. There are more DB nodes, all expanding the input SQL query with additional SQL instructions. Besides the SQL Query node, no DB nodes require SQL coding.
DB Row Filter	Expands the input SQL query to include the row filter criteria defined in the configuration window. Grouping of multiple conditions with an AND or OR conjunctions is also supported. No SQL coding required.
DB Query	Modifies the input SQL query using custom SQL. The input SQL query is represented by the place holder #table#.
DB Writer	Inserts the data rows from the top input port into a table in the database specified by the input connection port. If the database table does not exist it will be created.
DB Connection Table Writer	Writes the resulting rows from the input SQL query into a new table inside the database.

DATE&TIME

String to Date&Time	Parses the strings in the selected columns according to a date/time format and converts them into Date&Time cells. Four Date&Time forms are supported: only date, only time, date&time, and date&time plus time zone.
Date&Time-based Row Filter	Extracts rows where the time value in the selected column lies within a given time window. The time window is specified either by a start and /or an end date or by a start date and a duration.
Date&Time Difference	Calculates the difference between two date&time objects e.g., from two selected columns, from a selected column and a fixed value, from a selected column and the current execution time, or from one cell and the cell in the previous row for a selected column.
Extract Date&Time Fields	Extracts selected time and date fields from a selected column of type date&time and appends their values in new columns.

CLEAN DATA

Missing Value	Defines and applies a strategy to replace missing values in the input table - either globally on all columns, or individually for each single column.
Duplicate Row Filter	Detects duplicate rows and applies the selected operation, e.g. removes duplicate rows. Duplicates are rows that have the same value in all selected columns.
Numeric Outliers	Detects and treats numerical outliers for each of the selected columns individually using the interquartile range (IQR).

RESHAPE AND AGGREGATE DATA

GroupBy	Groups the rows of a table by the unique values in selected columns and calculates aggregation and statistical measures for the defined groups. Despite its simple name, it offers powerful functionality and has many unsuspected usages.
Pivoting	Extends the aggregation functionality of the GroupBy node by creating an output table with columns and rows for the unique values in the selected input columns. The unique values of the grouping columns become rows and the unique values of the pivoting columns become columns.
Category to Number	Maps the categorical values in the selected columns to integer values and exports the mapping rules to the model output port. The Category to Number (Apply) and Number to Category (Apply) nodes apply the mapping rule in both directions.
One to Many	Creates one new column for each value in the selected input column. These values become the column headers. Cells in the newly created columns are set to 0 if the value is not present, otherwise 1. This type of encoding is called one-hot vector.
Transpose	Converts the rows to columns and the columns to rows.
Table Manipulator	Performs several transformations at once, such as renaming, filtering, re-ordering and type changing, on the input columns. By adding dynamic ports it can replace a concatenate node.
Cell Splitter	Splits values in the selected column into two or more substrings, as defined by a delimiter match. A delimiter is a defined character, such as a comma, space, or any other character or character sequence.
Ungroup	Ungroups a collection-type cell by creating one row for each value in the collection cell. Other columns from the input table are left unaltered.
Unpivoting	Stacks the cells of the selected value columns into one column. The cells of the selected retained columns are appended to the corresponding output rows.
Sorter	Sorts the table in ascending or descending order based on the values of one or more columns.
Value Counter	Counts the number of occurrences of all values in a selected column from the input table.

DYNAMIC PORT

Dynamic ports: Additional input ports can be added by clicking the three dots in the bottom left corner of a node.

DATA TYPES & CONVERSIONS

- S **String**: Sequence of characters, e.g. "This is a string"
- I **Integer**: Whole real valued number, e.g. -100 or 345
- D **Double**: Real valued number, e.g. -0.432 or 45.39
- **Date&Time**: A data format for date, time, date&time, or date&time plus time zone.
- B **Boolean**: Two possible values only, e.g. TRUE and FALSE
- [..] **Collection Cell**: Collection of multiple values of either the same or different types e.g., can be a list of values or a set of values. In a set each value occurs only once.
- **Document/Image**: KNIME Analytics Platform supports many more data types like text documents, images, fingerprints, etc.

Number to String	Converts the data type of the selected columns from a number format, e.g. integer or double, to string.
String to Number	Converts the data type of the selected columns from string to either double or integer.

CREATE COLUMNS

Math Formula	Implements a number of math operations across multiple input columns. The math operations can be applied to multiple columns with the Math Formula (Multi Column) node.
Rule Engine	Applies a set of rules to each row of the input table. Rules are applied from top to bottom. The first rule that matches is used.
Counter Generation	Creates a new column with a counter. The start value and step size are defined in the configuration window.
String Manipulation	Performs operations on string values in columns, such as combining two or more strings together, extracting one or more substrings, trimming blank spaces, and so on.
String Replacer	Replaces values in a selected string column if they match a defined pattern.
Column Expressions	Combines the functionality of the Math Formula, Rule Engine, and String Manipulation nodes. More than one expression can be defined to modify or add multiple columns at the same time.

FORMAT EXCEL SHEETS

The Continental Nodes for KNIME extension allows you to automatically format an existing Excel sheet. The key is an additional data table of the same size as the original Excel sheet, where each cell contains one or more comma separated tag values e.g., header, border, etc. Based on these tags, the XLS Formatter nodes add new formatting instructions to the existing instructions, as available at the lower (optional) input port.

XLS Control Table Generator	Transforms the input table to an XLS Control Table, meaning it exchanges the column names to A, B, C, ... and the row IDs to 1, 2, 3, ... It is the kickoff node to collect formatting instructions for an Excel sheet and feeds all XLS formatter nodes.
XLS Background Colorizer	Adds background color and/or pattern fill formatting instructions to all cells with a specified tag in the XLS Control Table at the top input port.
XLS Border Formatter	Adds border formatting instructions for a given range specified by a tag in the XLS control table at the top input port.
XLS Cell Merger	Adds formatting instructions to merge all cells with a specified tag in the XLS control table at the top input port.
XLS Conditional Formatter	Adds formatting instructions to color cell backgrounds according to their numeric value for all cells specified by a tag in the XLS control table at the top
XLS Formatter (apply)	Applies all formatting instructions to an existing Excel sheet.

Resources

E-Books: KNIME Advanced Luck covers advanced features & more. Practicing Data Science is a collection of data science case studies from past projects. Both available at knime.com/knimepress

KNIME Blog: Engaging topics, challenges, industry news, & knowledge nuggets at knime.com/blog

E-Learning Courses: Take our free online self-paced courses to learn about the different steps in a data science project (with exercises & solutions to test your knowledge) at www.knime.com/knime-self-paced-courses

KNIME Hub: Browse and share workflows, nodes, and components. Add ratings, or comments to other workflows at hub.knime.com

KNIME Forum: Join our global community & engage in conversations at forum.knime.com

KNIME Server: For team-based collaboration, automation, management, & deployment check out KNIME Server at www.knime.com/knime-server

Thank You!

Questions? Please reach out

alice.krebs@knime.com
zachary.durso@knime.com