



University of Minho
School of Engineering



Machine Learning and Decision-Making

ADI @ LEI/3º, MiEI/4º - 2º Semestre
Filipe Gonçalves, Inês Alves, Cesar Analide

Part VII – March 2022

Contents

2

The Elbow Method

Quality Measures

Hands On

- Clustering
 - Clustering Pipeline
 - The Elbow Method
- Quality measures
- Hands On

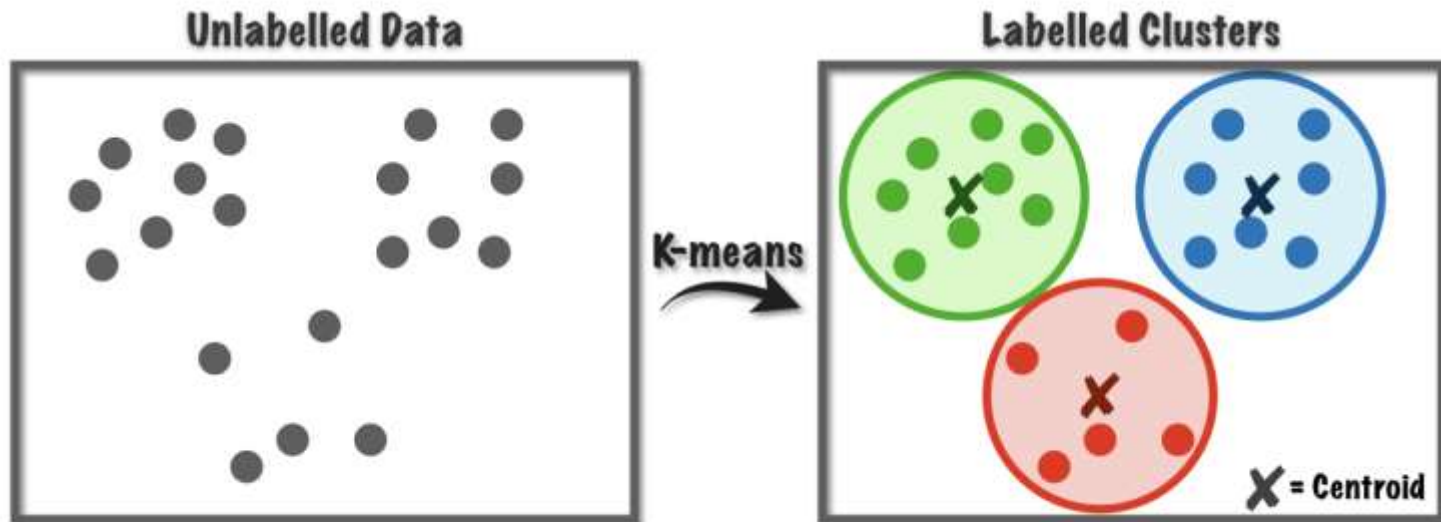
Clustering

3

THE ELBOW METHOD

Quality Measures

Hands On



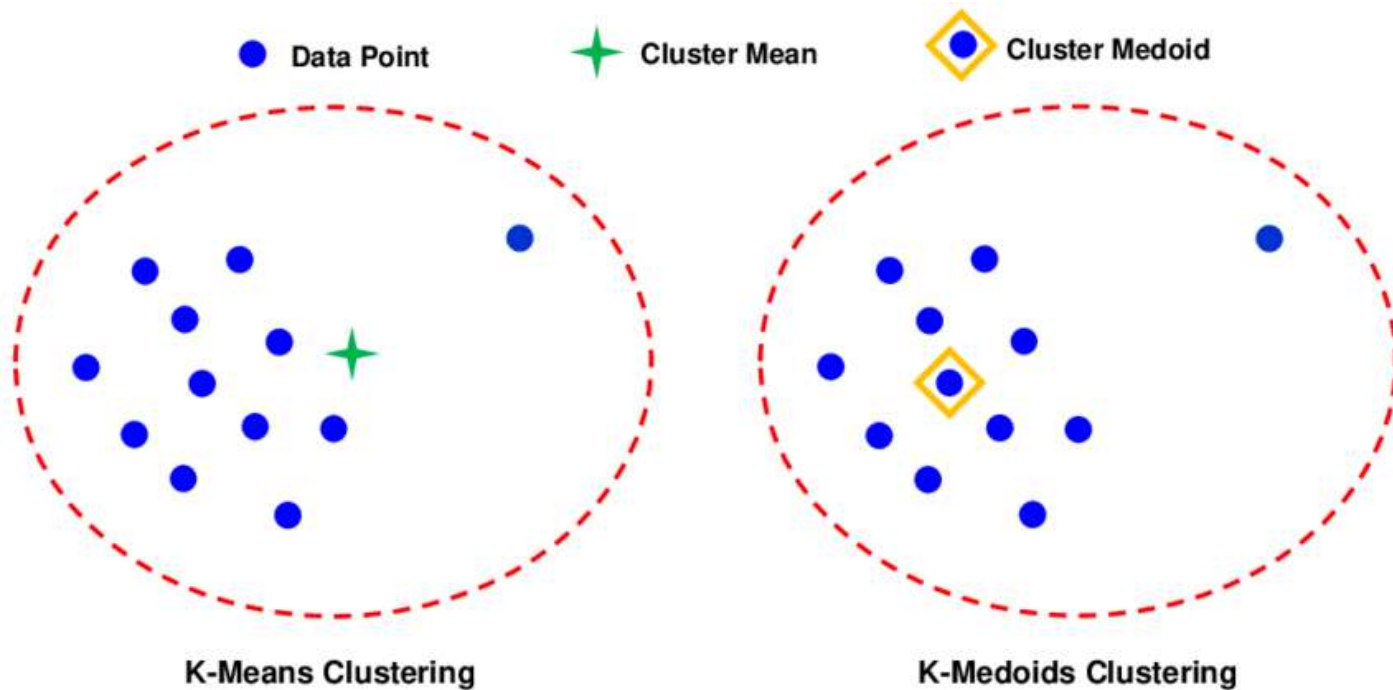
Clustering

4

THE ELBOW METHOD

Quality Measures

Hands On



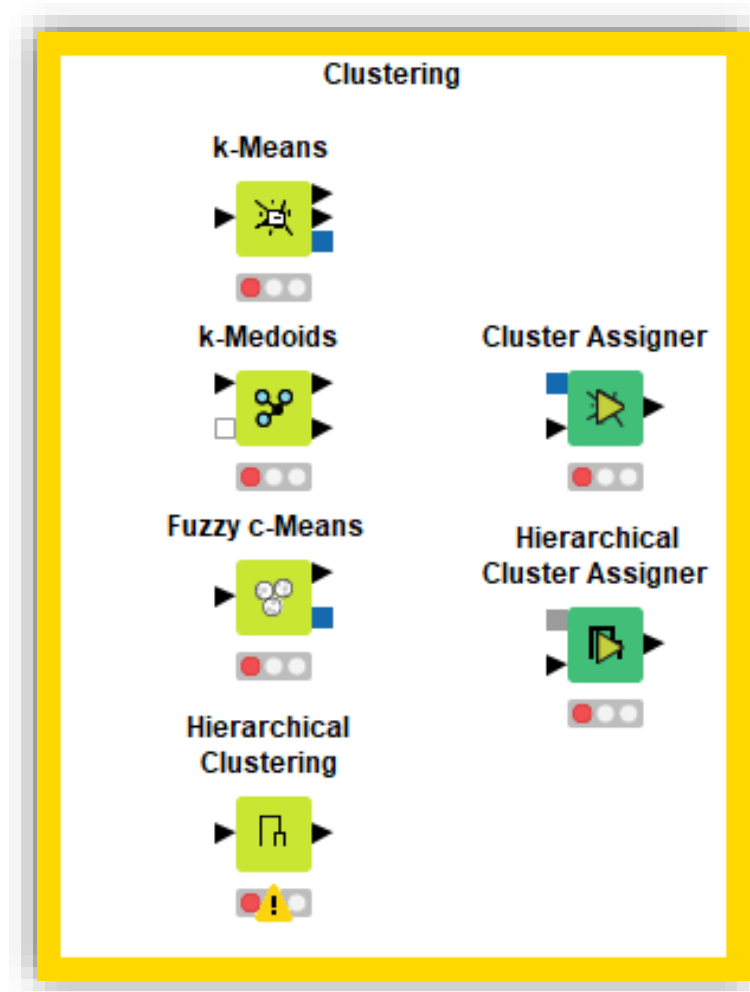
Clustering

5

THE ELBOW METHOD

Quality Measures

Hands On



Clustering

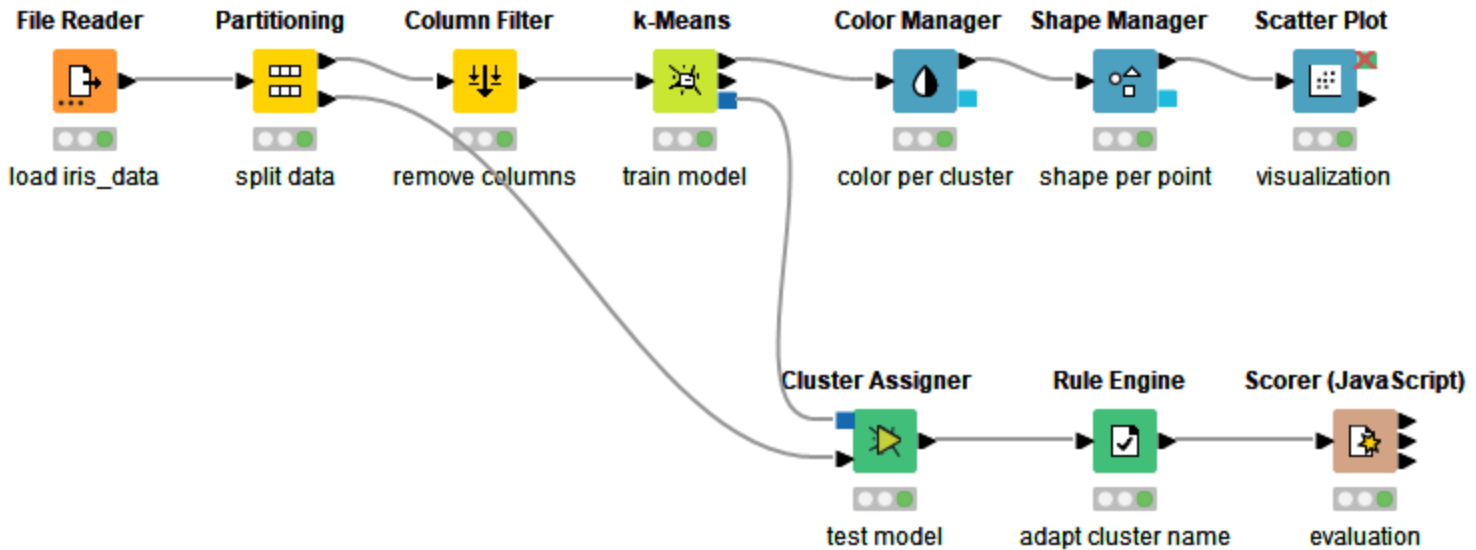
6

THE ELBOW METHOD

Quality Measures

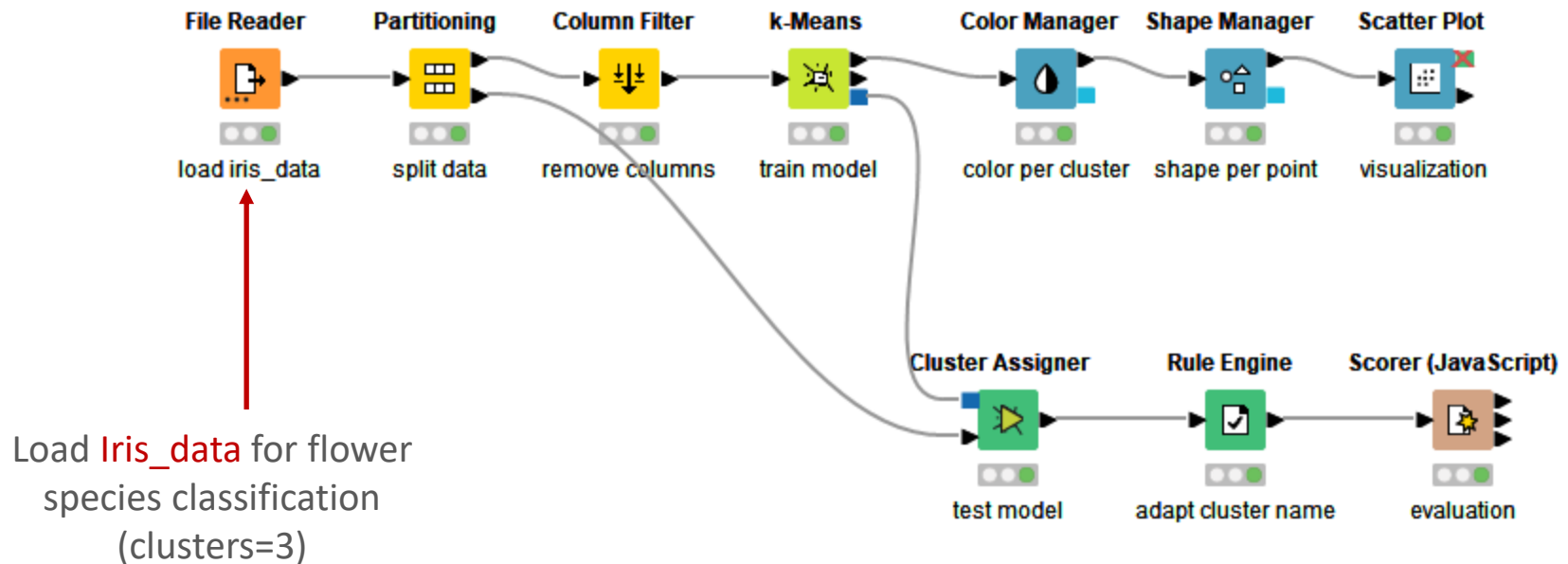
Hands On

Clustering Pipeline



Clustering

Clustering Pipeline



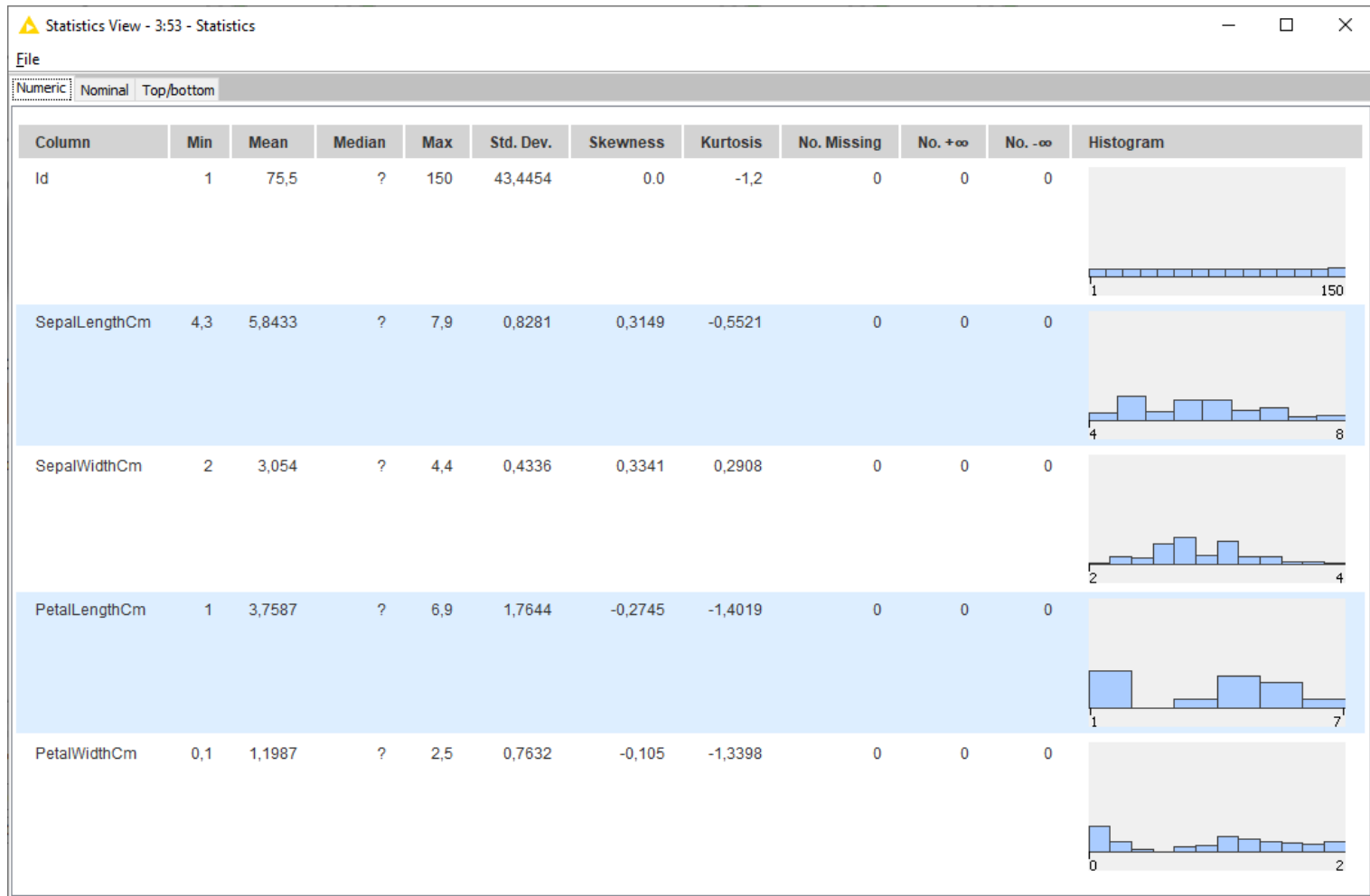
Clustering

8

THE ELBOW METHOD

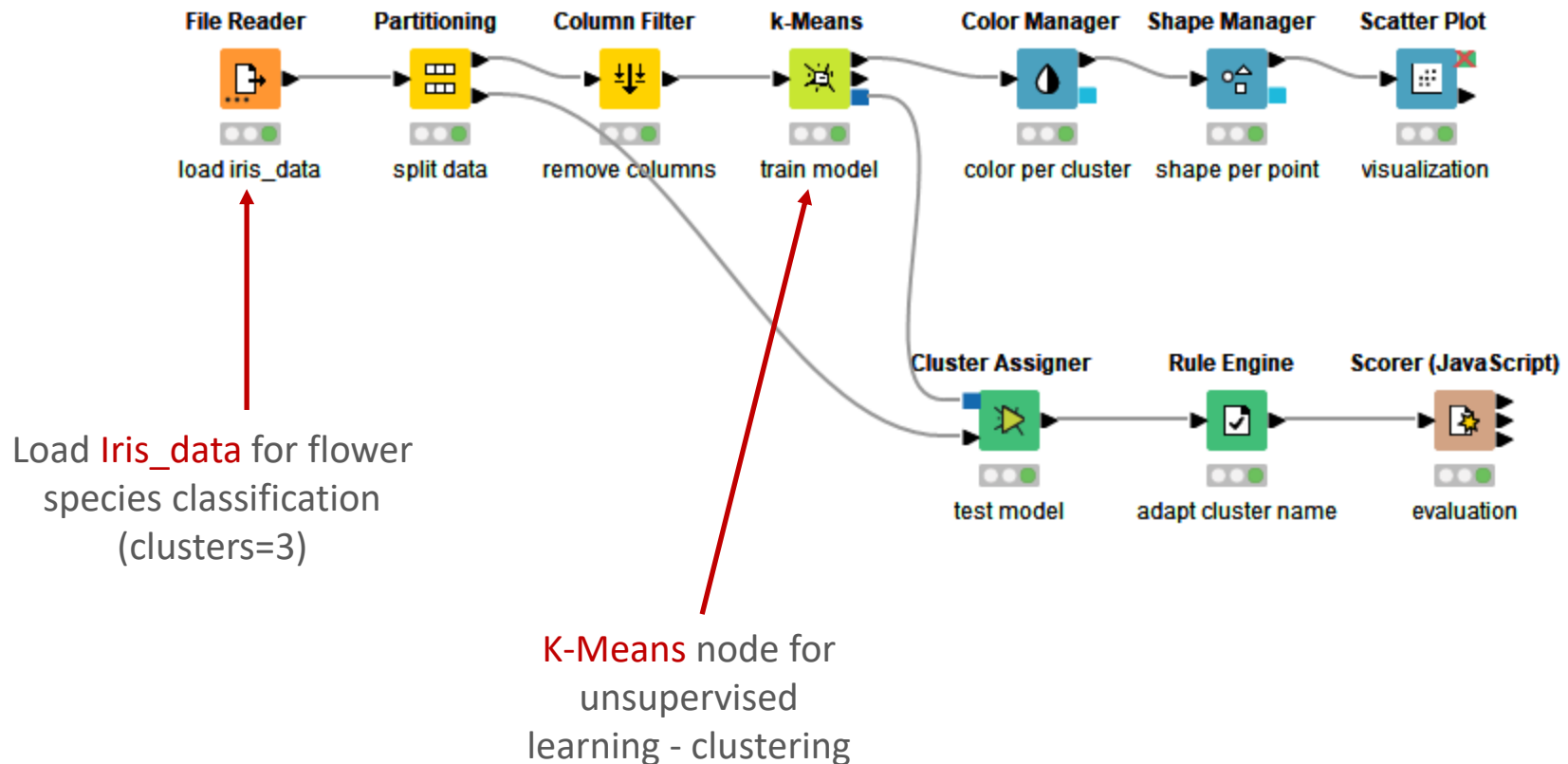
Quality Measures

Hands On



Clustering

Clustering Pipeline



K-Means Settings

10

THE ELBOW METHOD

Quality Measures

Hands On

The screenshot shows the 'Dialog - 0:21 - k-Means' window. The 'K-Means Properties' tab is active. The 'number of clusters' is set to 3 and is circled in red. The 'max. number of iterations' is set to 99. Below these, there are 'Exclude' and 'Include' sections, each with a 'Filter' input and a list area. The 'Exclude' section is outlined in red, and the 'Include' section is outlined in green. Between these sections are navigation buttons: '>', '>>', '<<', and '<'. At the bottom, there is a checkbox for 'Always include all columns' and another for 'Enable Hilite Mapping'. The 'OK', 'Apply', 'Cancel', and a help icon are at the bottom right.

Node Description

k-Means

This node outputs the cluster centers for a predefined number of clusters (no dynamic number of clusters). K-means performs a crisp clustering that assigns a data vector to exactly one cluster. The algorithm terminates when the cluster assignments do not change anymore.

The clustering algorithm uses the Euclidean distance on the selected attributes. The data is not normalized by the node (if required, you should consider to use the "Normalizer" as a preprocessing step).

Dialog Options

number of clusters

The number of clusters (cluster centers) to be created.

max number of iterations

The number of iterations after which the algorithm terminates, independent of the accuracy improvement of the cluster centers.

Enable Hilite Mapping

If enabled, the hiliting of a cluster row (2nd output) will hilite all rows of this cluster in the input table and the 1st output table. Depending on the number of rows, enabling this feature might consume a lot of memory.

Ports

Input Ports	
0	Input to clustering. All numerical values and only these are considered for clustering.

Clustering

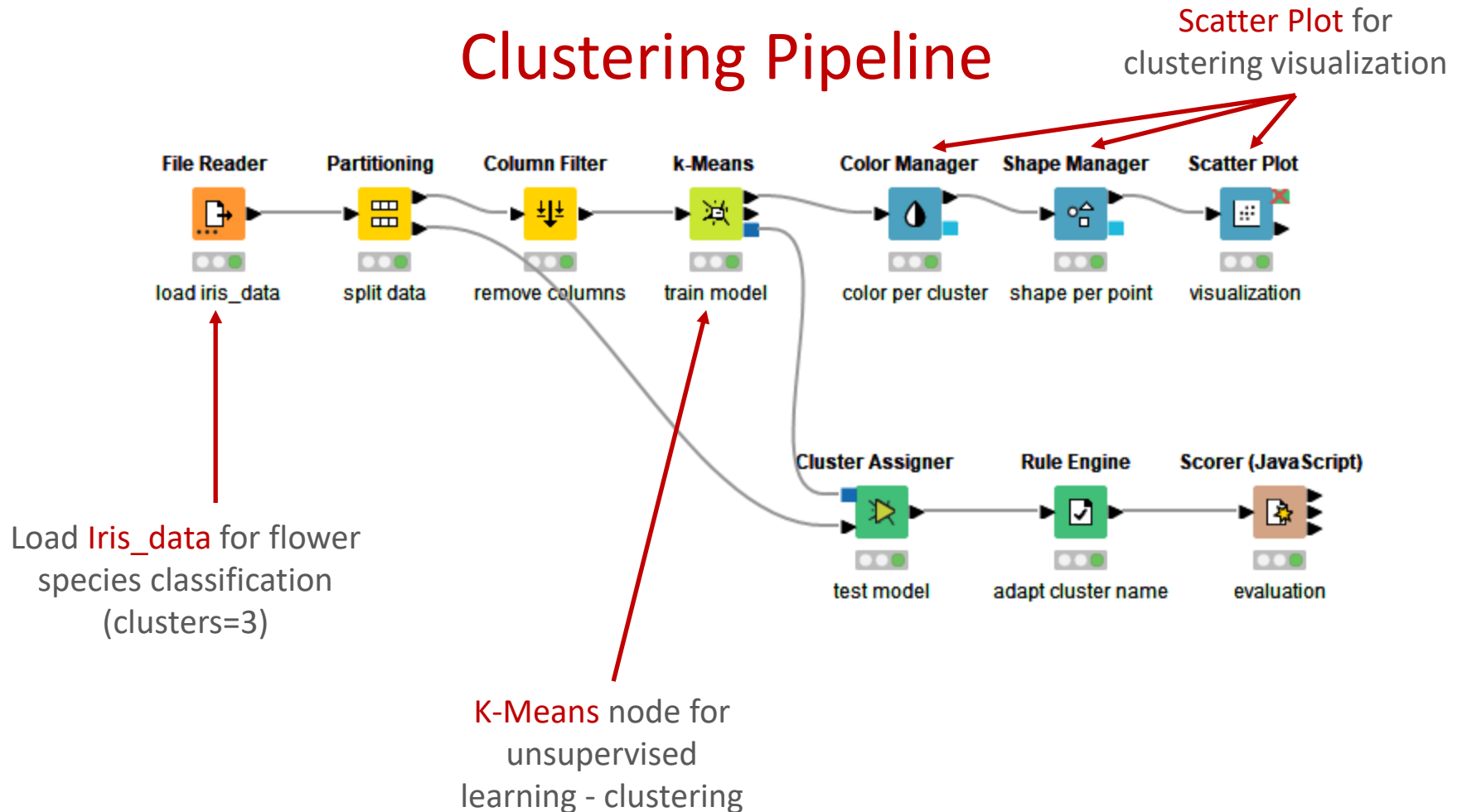
11

THE ELBOW METHOD

Quality Measures

Hands On

Clustering Pipeline



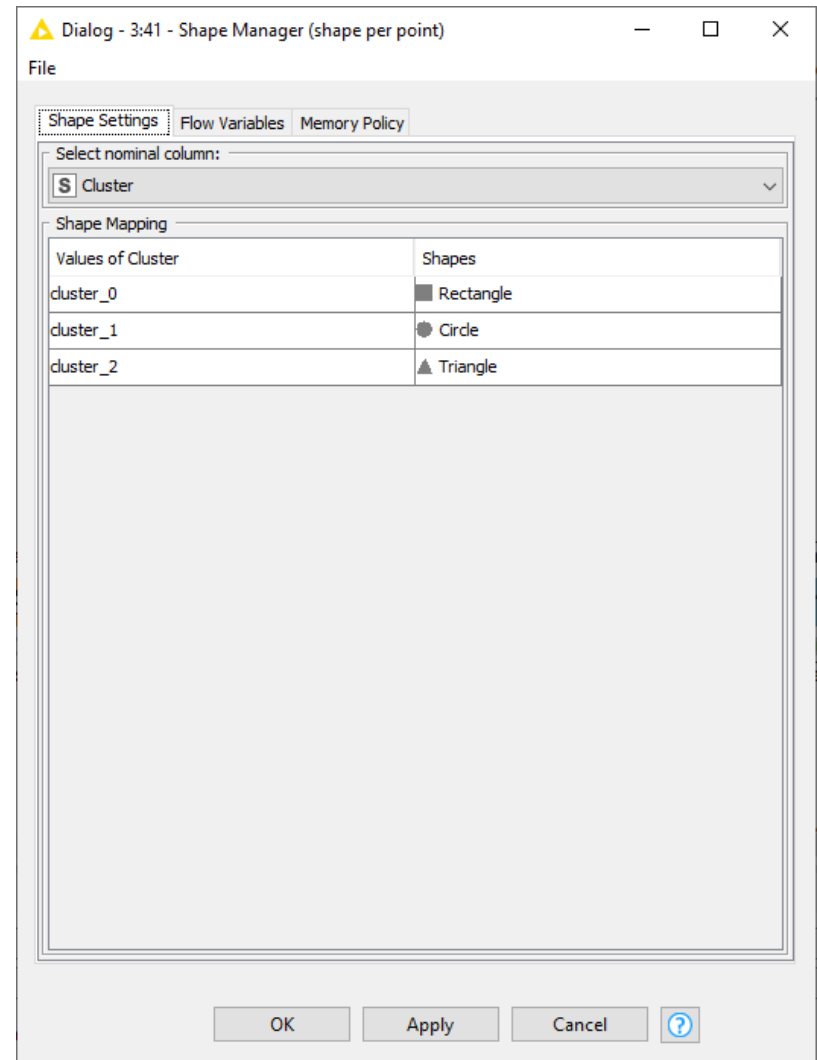
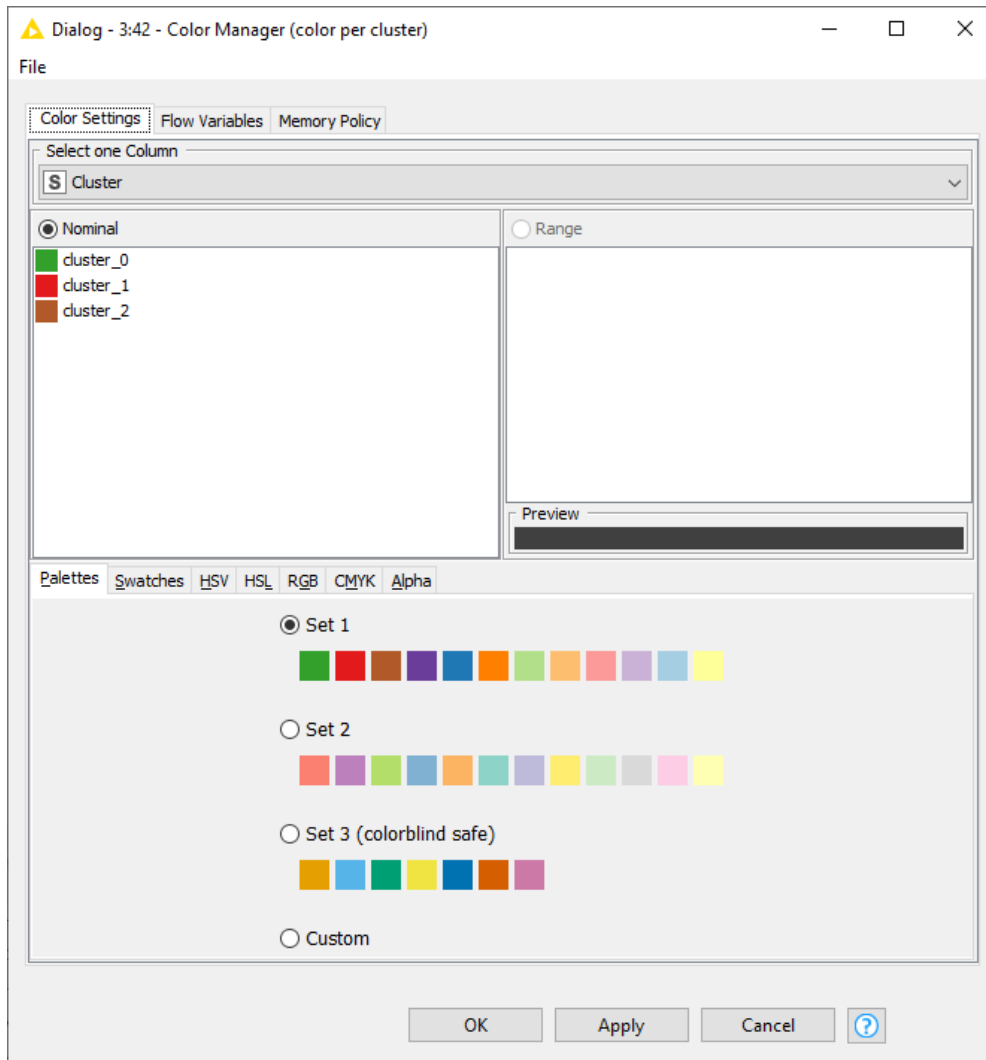
Clustering

12

THE ELBOW METHOD

Quality Measures

Hands On



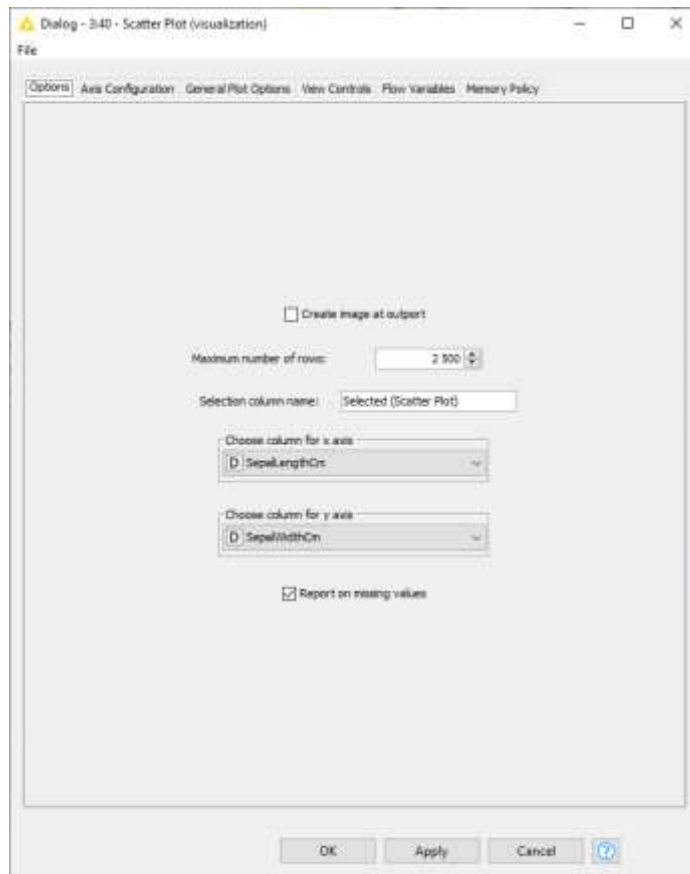
Clustering

13

THE ELBOW METHOD

Quality Measures

Hands On



Clustering

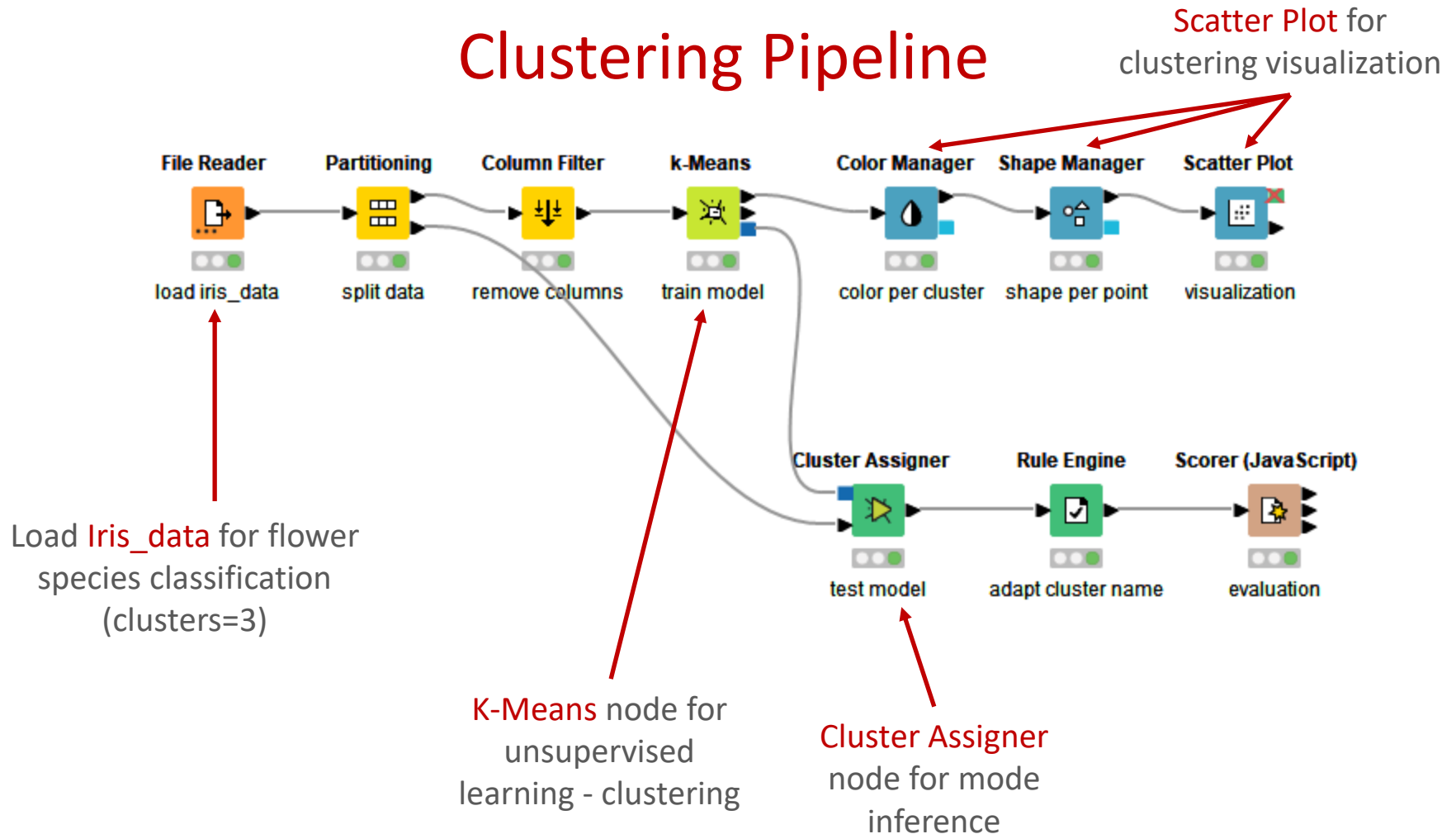
14

THE ELBOW METHOD

Quality Measures

Hands On

Clustering Pipeline



Clustering

15

THE ELBOW METHOD

Quality Measures

Hands On

Assigned Data - 3:43 - Cluster Assigner (test model)

File Edit Hilite Navigation View

Table "default" - Rows: 30 Spec - Columns: 7 Properties Flow Variables

Row ID	I Id	D SepalLe...	D SepalW...	D PetalLe...	D PetalWi...	S Species	S Cluster
Row9	10	4.9	3.1	1.5	0.1	Iris-setosa	cluster_1
Row20	21	5.4	3.4	1.7	0.2	Iris-setosa	cluster_1
Row28	29	5.2	3.4	1.4	0.2	Iris-setosa	cluster_1
Row29	30	4.7	3.2	1.6	0.2	Iris-setosa	cluster_1
Row31	32	5.4	3.4	1.5	0.4	Iris-setosa	cluster_1
Row33	34	5.5	4.2	1.4	0.2	Iris-setosa	cluster_1
Row39	40	5.1	3.4	1.5	0.2	Iris-setosa	cluster_1
Row42	43	4.4	3.2	1.3	0.2	Iris-setosa	cluster_1
Row43	44	5	3.5	1.6	0.6	Iris-setosa	cluster_1
Row45	46	4.8	3	1.4	0.3	Iris-setosa	cluster_1
Row73	74	6.1	2.8	4.7	1.2	Iris-versicolor	cluster_0
Row76	77	6.8	2.8	4.8	1.4	Iris-versicolor	cluster_0
Row80	81	5.5	2.4	3.8	1.1	Iris-versicolor	cluster_0
Row85	86	6	3.4	4.5	1.6	Iris-versicolor	cluster_0
Row86	87	6.7	3.1	4.7	1.5	Iris-versicolor	cluster_0
Row88	89	5.6	3	4.1	1.3	Iris-versicolor	cluster_0
Row91	92	6.1	3	4.6	1.4	Iris-versicolor	cluster_0
Row92	93	5.8	2.6	4	1.2	Iris-versicolor	cluster_0
Row94	95	5.6	2.7	4.2	1.3	Iris-versicolor	cluster_0
Row96	97	5.7	2.9	4.2	1.3	Iris-versicolor	cluster_0
Row102	103	7.1	3	5.9	2.1	Iris-virginica	cluster_2
Row105	106	7.6	3	6.6	2.1	Iris-virginica	cluster_2
Row110	111	6.5	3.2	5.1	2	Iris-virginica	cluster_2
Row114	115	5.8	2.8	5.1	2.4	Iris-virginica	cluster_0
Row121	122	5.6	2.8	4.9	2	Iris-virginica	cluster_0
Row132	133	6.4	2.8	5.6	2.2	Iris-virginica	cluster_2
Row134	135	6.1	2.6	5.6	1.4	Iris-virginica	cluster_2
Row140	141	6.7	3.1	5.6	2.4	Iris-virginica	cluster_2
Row144	145	6.7	3.3	5.7	2.5	Iris-virginica	cluster_2
Row148	149	6.2	3.4	5.4	2.3	Iris-virginica	cluster_2

Clustering

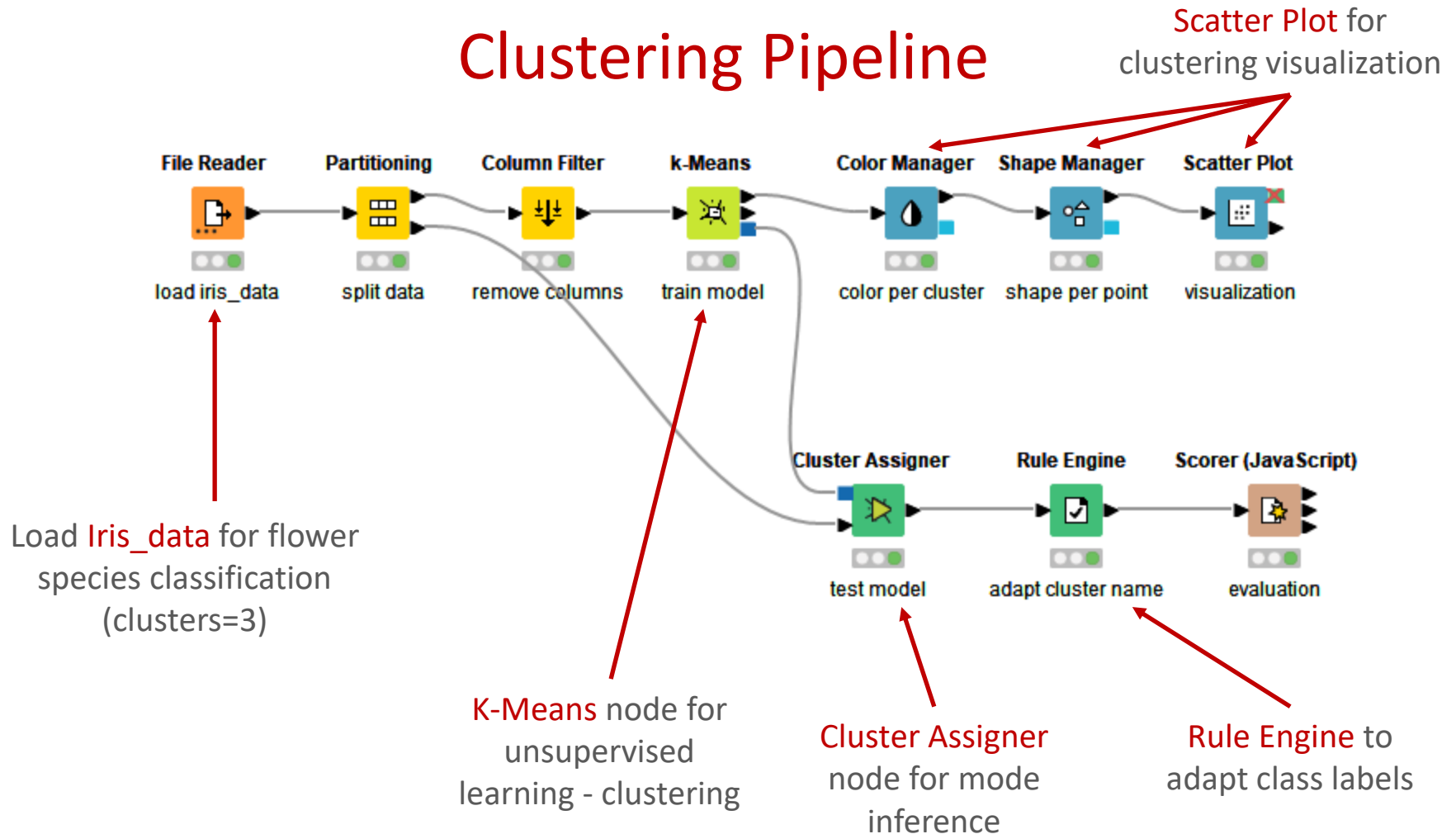
16

THE ELBOW METHOD

Quality Measures

Hands On

Clustering Pipeline



Clustering

17

THE ELBOW METHOD

Quality Measures

Hands On

Dialog - 3:45 - Rule Engine (adapt cluster name)

File

Rule Editor | Flow Variables | Memory Policy

Column List

- ROWID
- ROWINDEX
- ROWCOUNT
- I Id
- D Sepal.LengthCm
- D Sepal.WidthCm
- D Petal.LengthCm
- D Petal.WidthCm
- S Species
- S Cluster

Flow Variable List

- s knime.workspace

Category

All

Function

- ? < ?
- ? <= ?
- ? = ?
- ? > ?
- ? >= ?
- ? AND ?
- ? IN ?
- ? LIKE ?
- ? MATCHES ?
- ? OR ?
- ? XOR ?
- FALSE
- MISSING ?
- NOT ?

Description

Expression

```
// enter ordered set of rules, e.g.:  
// $double column name$ > 5.0 => "large"  
// $string column name$ LIKE "*blue*" => "small and blue"  
// TRUE => "default outcome"  
$cluster$ = "cluster_0" => "Iris-versicolor"  
$cluster$ = "cluster_1" => "Iris-setosa"  
$cluster$ = "cluster_2" => "Iris-virginica"
```

Append Column: Continent

Replace Column: Cluster

OK Apply Cancel ?

Clustering

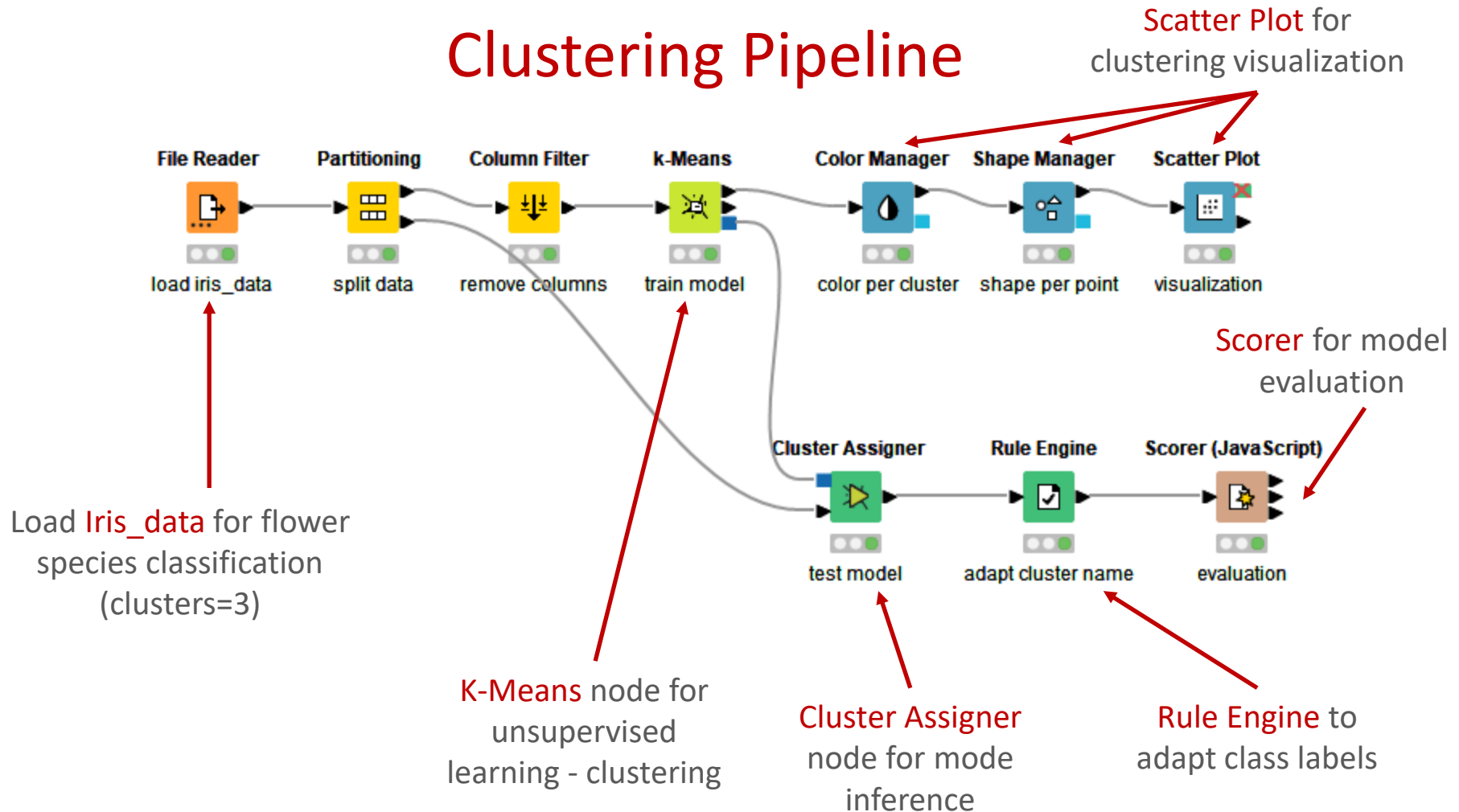
18

THE ELBOW METHOD

Quality Measures

Hands On

Clustering Pipeline



Clustering

19

THE ELBOW METHOD

Quality Measures

Hands On

Confusion Matrix				
Scorer View				
Confusion Matrix				
	Iris-setosa (Predicted)	Iris-versicolor (Predic...	Iris-virginica (Predict...	
Iris-setosa (Actual)	10	0	0	100.00%
Iris-versicolor (Actual)	0	10	0	100.00%
Iris-virginica (Actual)	0	2	8	80.00%
	100.00%	83.33%	100.00%	
Overall Statistics				
Overall Accuracy	Overall Error	Cohen's kappa (κ)	Correctly Classified	Incorrectly Classified
93.33%	6.67%	0.900	28	2

When not defined.. how many clusters to use?

20

THE ELBOW METHOD

Quality Measures

Hands On



The Elbow Method

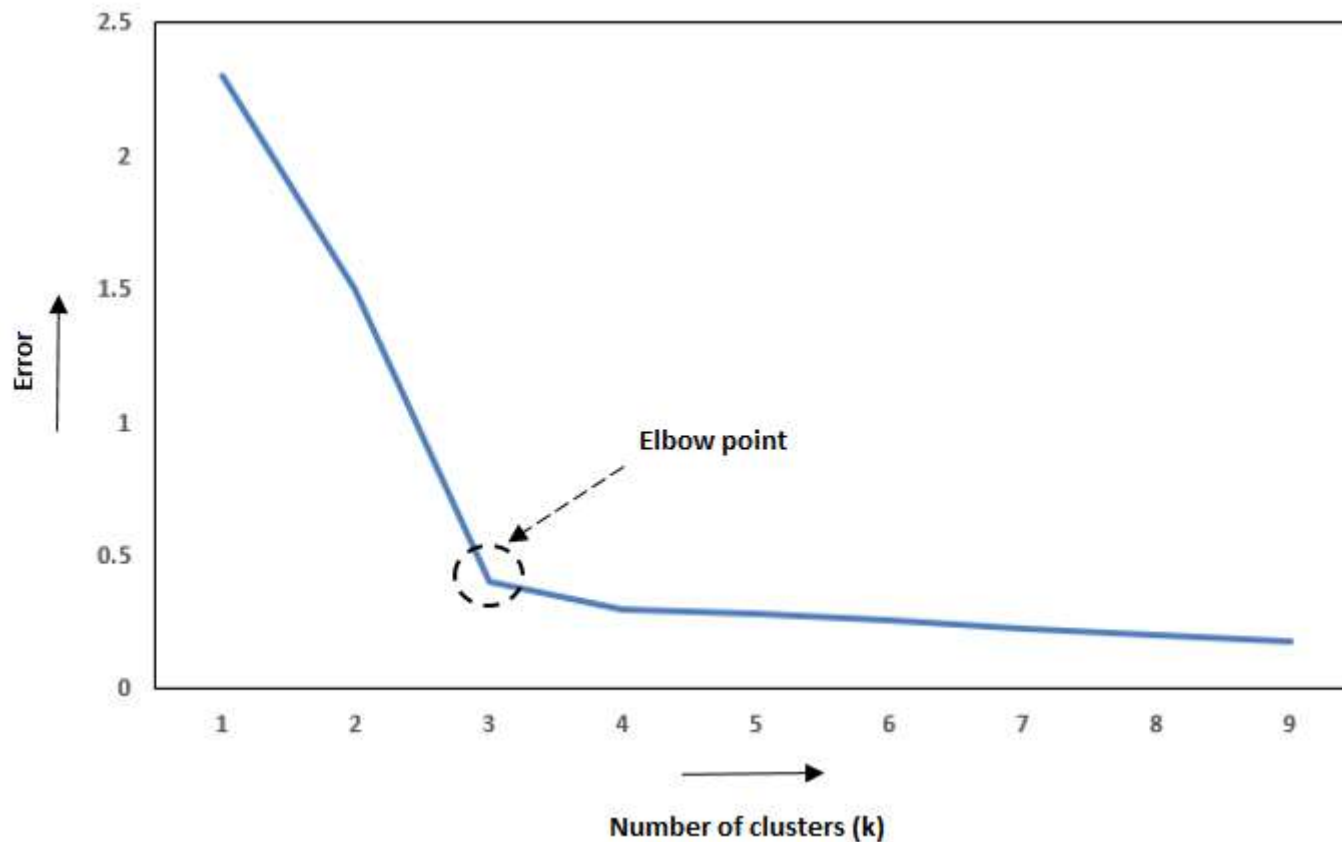
The Elbow Method

21

THE ELBOW METHOD

Quality Measures

Hands On



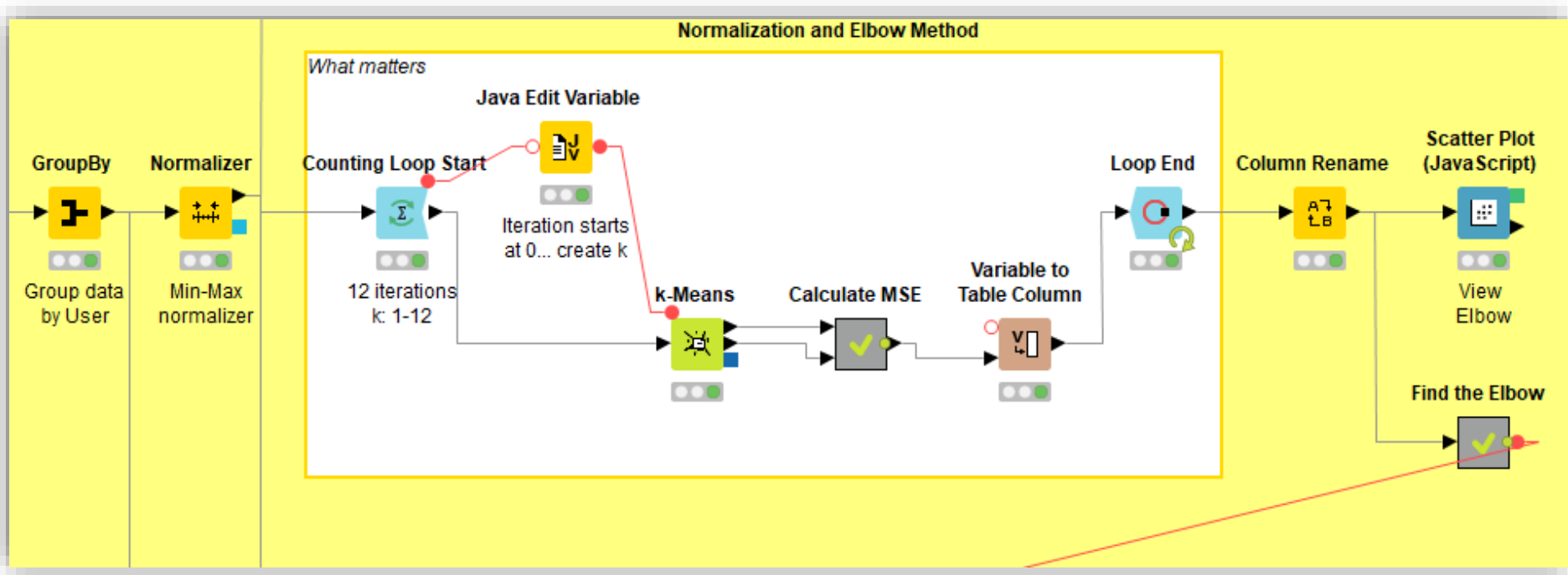
The Elbow Method

22

THE ELBOW METHOD

Quality Measures

Hands On



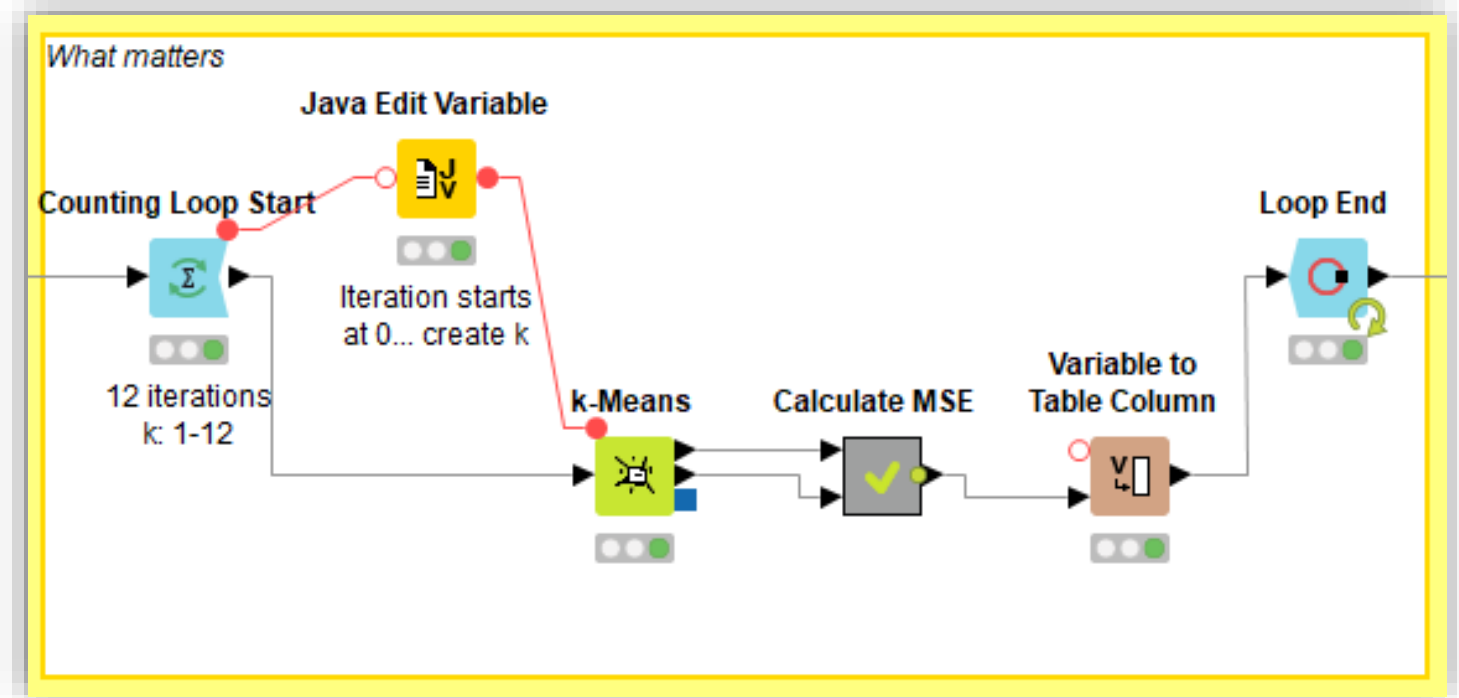
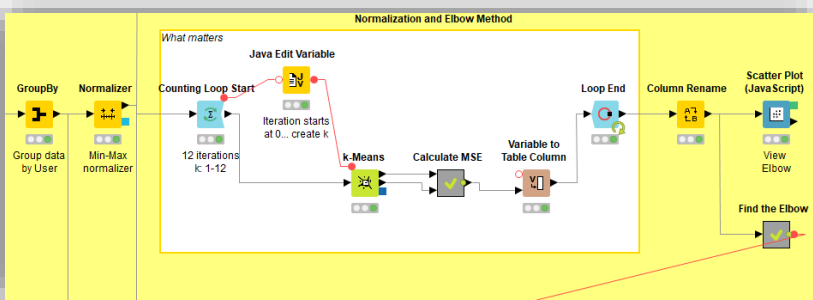
The Elbow Method

23

THE ELBOW METHOD

Quality Measures

Hands On



The Elbow Method

24

THE ELBOW METHOD

Quality Measures

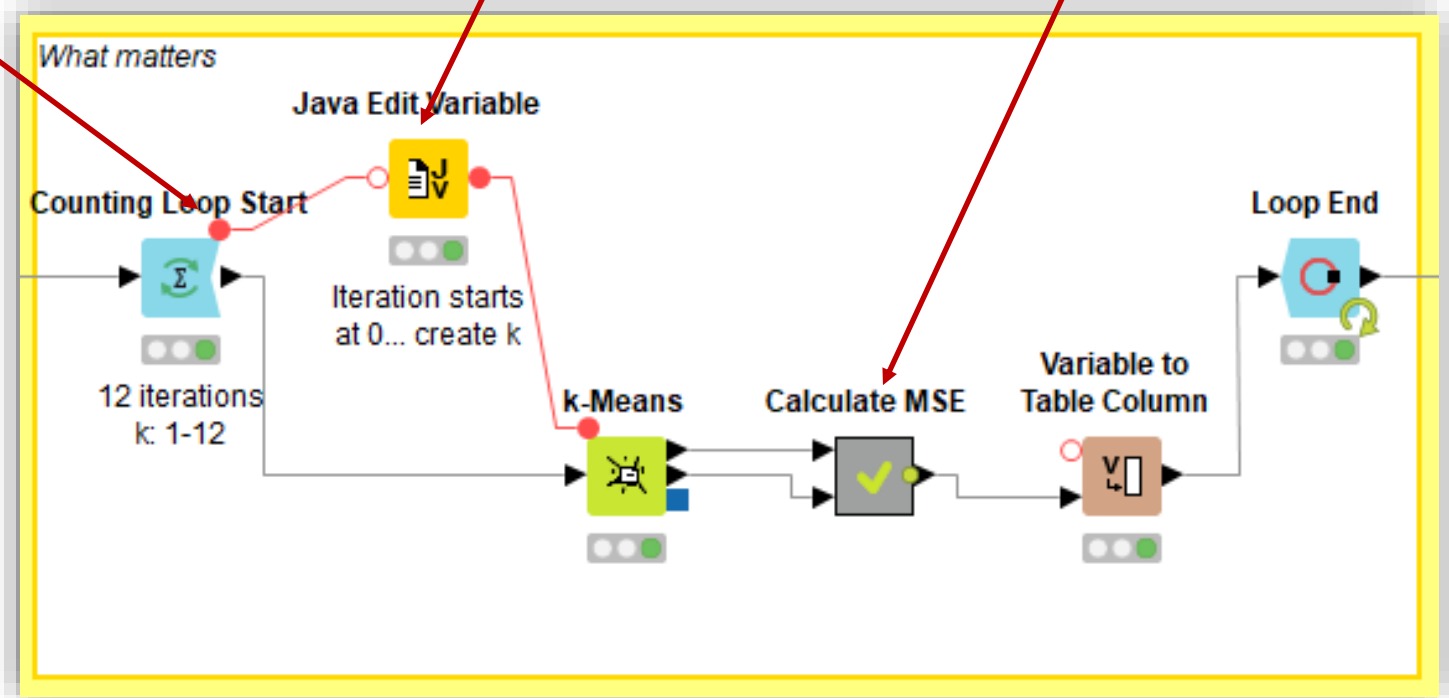
Hands On

Flow Variable made available by the Counting Loop Start:

currentIteration

Makes use of *currentIteration* to define the k of k-Means. Adds one because it starts at 0!

Calculates a **error metric** (MSE) to quantify k!



The Elbow Method

25

THE ELBOW METHOD

Quality Measures

Hands On

The screenshot displays the Orange3 data mining interface. A workflow titled "Normalization and Elbow Method" is visible, featuring components like GroupBy, Normalizer, Counting Loop Start (highlighted with a red box), Java Edit Variable, Loop End, Column Rename, and Scatter Plot (JavaScript). A "k-Means" widget is also present, with a "Calculate MSE" widget connected to it. A "Variable to Table Column" widget is also visible.

In the foreground, two dialog boxes are open. The top one is titled "Unaltered input - 2:142:0:122 - Counting Loop Start (12 iterations)" and shows a table of flow variables. The bottom one is titled "Dialog - 2:142:0:122 - Counting Loop Start (...)" and shows the "Standard settings" tab with a "Number of loops" spinner set to 12.

Index	Owner ID	Name	Value
	0 2:142:0:122	i* maxIterations	12
	0 2:142:0:122	i* currentIteration	11
	0 2:142:0:122	i* Loop-Execute	
	0 2:142:0:122	i* Loop (0)	
	0 2:142	i* Component Context	
	0	s* knime.workspace	C:\Users\bruno\knime-workspace

Standard settings	Flow Variables	Memory Policy
Number of loops <input type="text" value="12"/>		

The Elbow Method

26

THE ELBOW METHOD

Quality Measures

Hands On

The screenshot displays a KNIME workflow on the left and a 'Dialog - 2:142:0:123 - Java Edit Variable (iteration starts)' window on the right.

Workflow (Left):

- GroupBy:** Group data by User.
- Normalizer:** Min-Max normalizer.
- Counting Loop Start:** 12 iterations, k: 1-12.
- Java Edit Variable:** A node highlighted with a red box, containing the text 'iteration starts at 0... create k'.
- k-Means:** A clustering node.

Dialog - 2:142:0:123 - Java Edit Variable (iteration starts):

File

Java Snippet Additional Libraries Additional Bundles Templates Flow Variables Memory Policy

Create Template...

Flow Variable List

- currentIteration
- knime.workspace
- maxIterations

```
1 // system imports
13 // Your custom imports:
14
15 // system variables
25 // Your custom variables:
26
27 // expression start
29 // Enter your code here:
30
31
32 out_k = v_currentIteration+1;
33
34
35 // expression end
38
```

Input

Flow Variable	Java Type	Java Field
currentIteration	Integer	v_currentIteration

Output

Replace	Flow Variable	KNIME Type	Java Type	Java Field
<input type="checkbox"/>	k	int	Integer	out_k

OK Apply Cancel ?

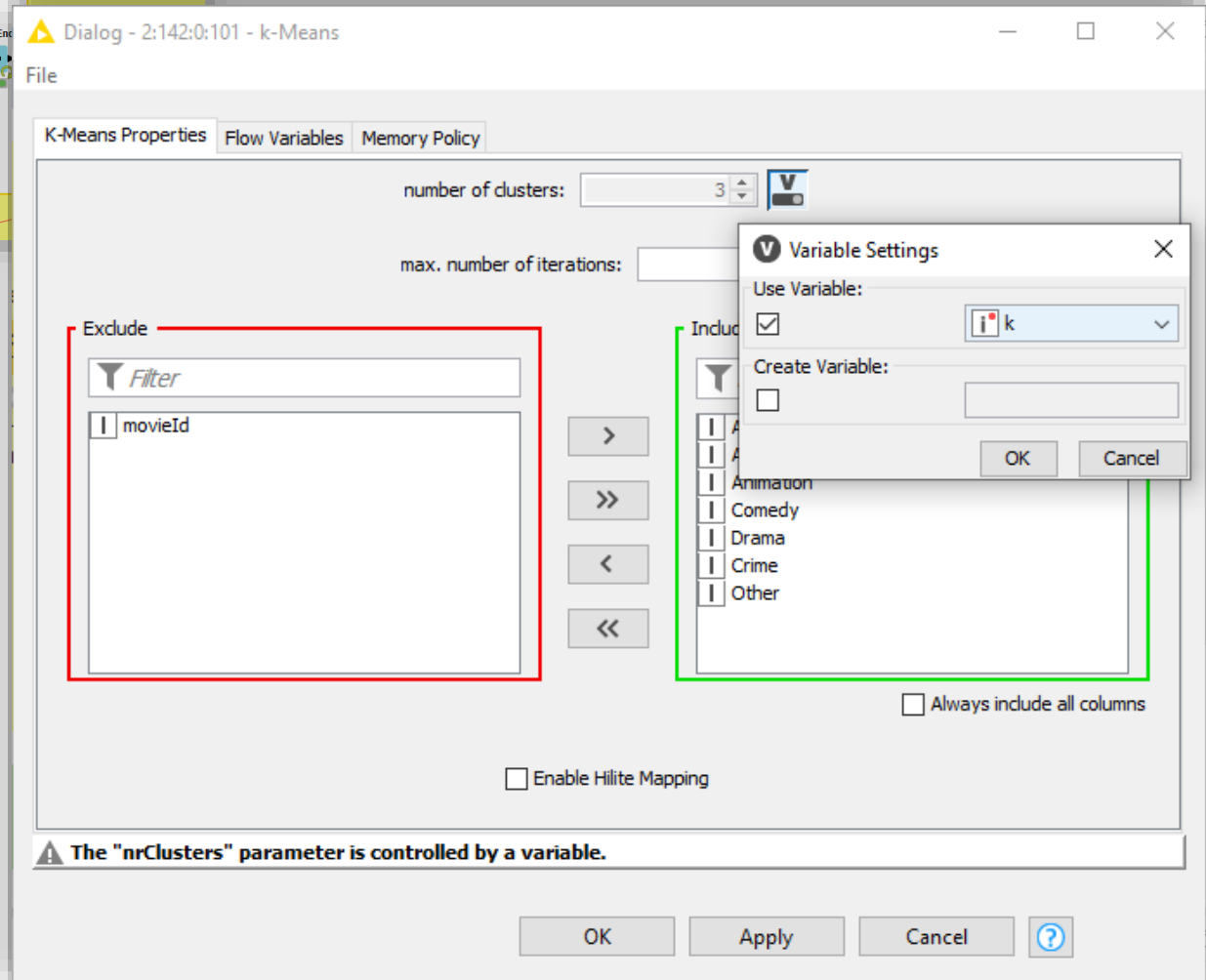
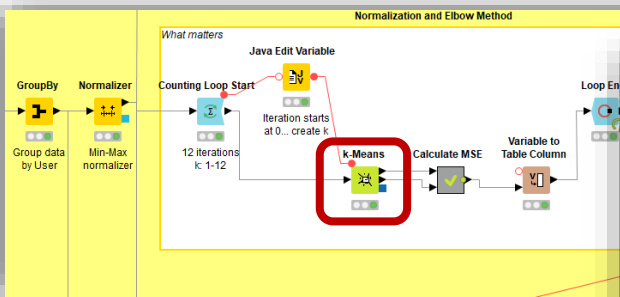
The Elbow Method

27

THE ELBOW METHOD

Quality Measures

Hands On



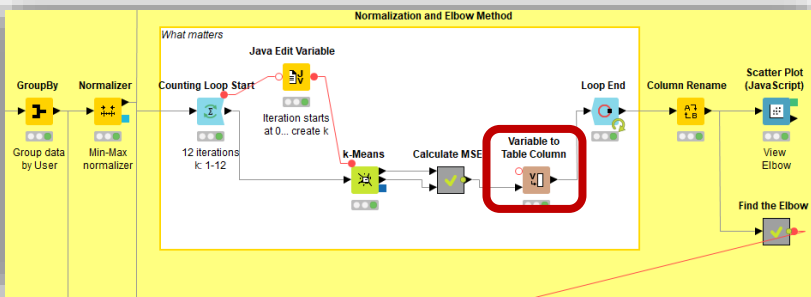
The Elbow Method

28

THE ELBOW METHOD

Quality Measures

Hands On



Input table with additional columns -...

File Hilite Navigation View

Table "default" - Rows: 1 Spec - Columns: 2 Properties Flow Variables

Row ID	D Mean(SquaredError)	I k
Row0	0.053	12

Dialog - 2:142:0:127 - Variable to Table Column

File

Variable Selection Flow Variables Memory Policy

☒ Manual Selection ☐ Wildcard/Regex Selection

Exclude

Filter

maxIterations
currentIteration
knome.workspace

☐ Enforce exclusion

Include

Filter

k

☒ Enforce inclusion

OK

Apply

Cancel

?

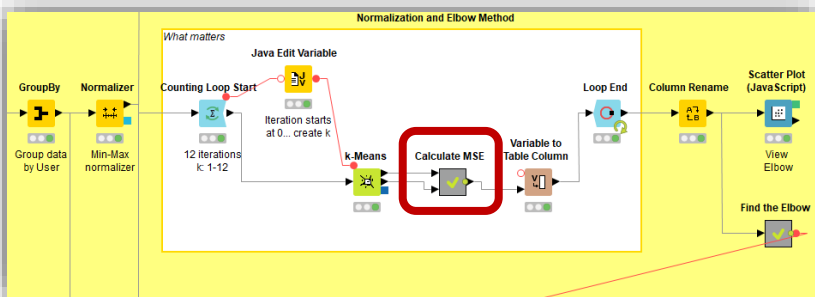
The Elbow Method

29

THE ELBOW METHOD

Quality Measures

Hands On



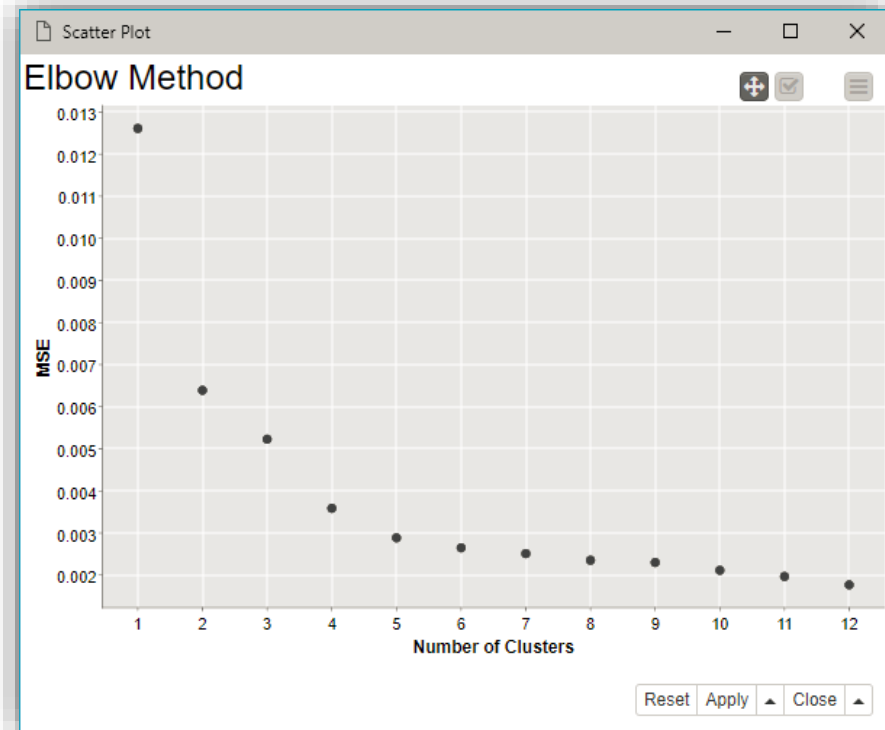
Collected results - ...

File Hilite Navigation View

Properties Table "default" - Rows: 12

Flow Variables Spec - Columns: 2

Row ID	Mean(S...	k
Row0#0	0.013	1
Row0#1	0.006	2
Row0#2	0.005	3
Row0#3	0.004	4
Row0#4	0.003	5
Row0#5	0.003	6
Row0#6	0.003	7
Row0#7	0.002	8
Row0#8	0.002	9
Row0#9	0.002	10
Row0#10	0.002	11
Row0#11	0.002	12



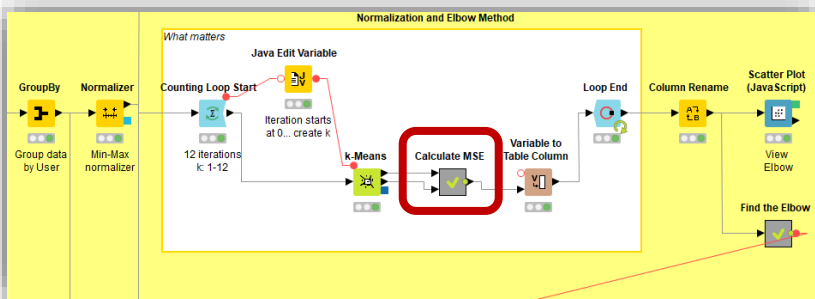
The Elbow Method

30

THE ELBOW METHOD

Quality Measures

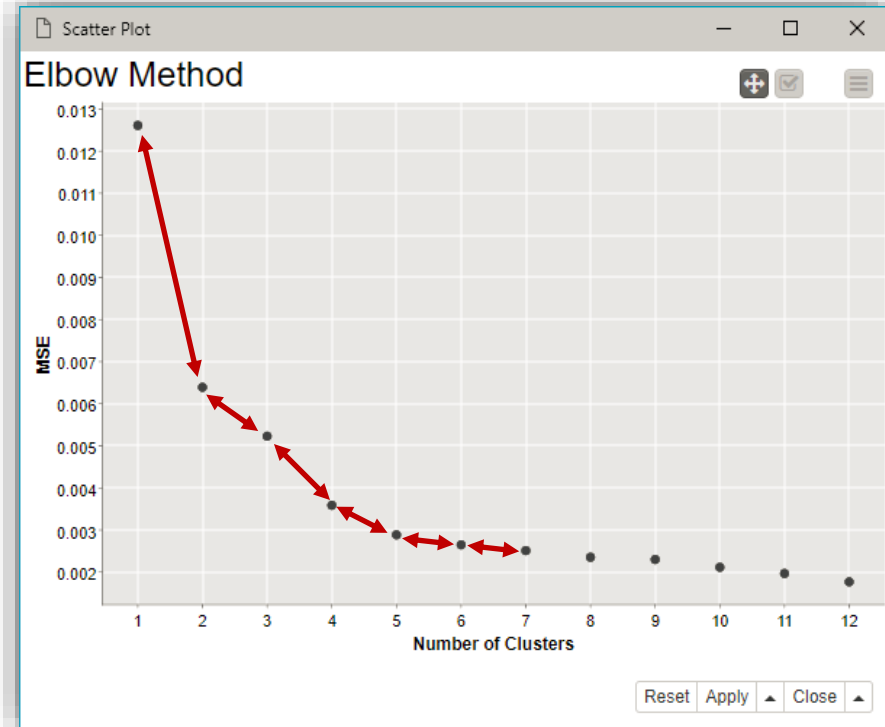
Hands On



Collected results - ...

File Hilite Navigation View

Properties		Flow Variables	
Table "default" - Rows: 12		Spec - Columns: 2	
Row ID	Mean(S...	k	
Row0#0	0.013	1	
Row0#1	0.006	2	
Row0#2	0.005	3	
Row0#3	0.004	4	
Row0#4	0.003	5	
Row0#5	0.003	6	
Row0#6	0.003	7	
Row0#7	0.002	8	
Row0#8	0.002	9	
Row0#9	0.002	10	
Row0#10	0.002	11	
Row0#11	0.002	12	



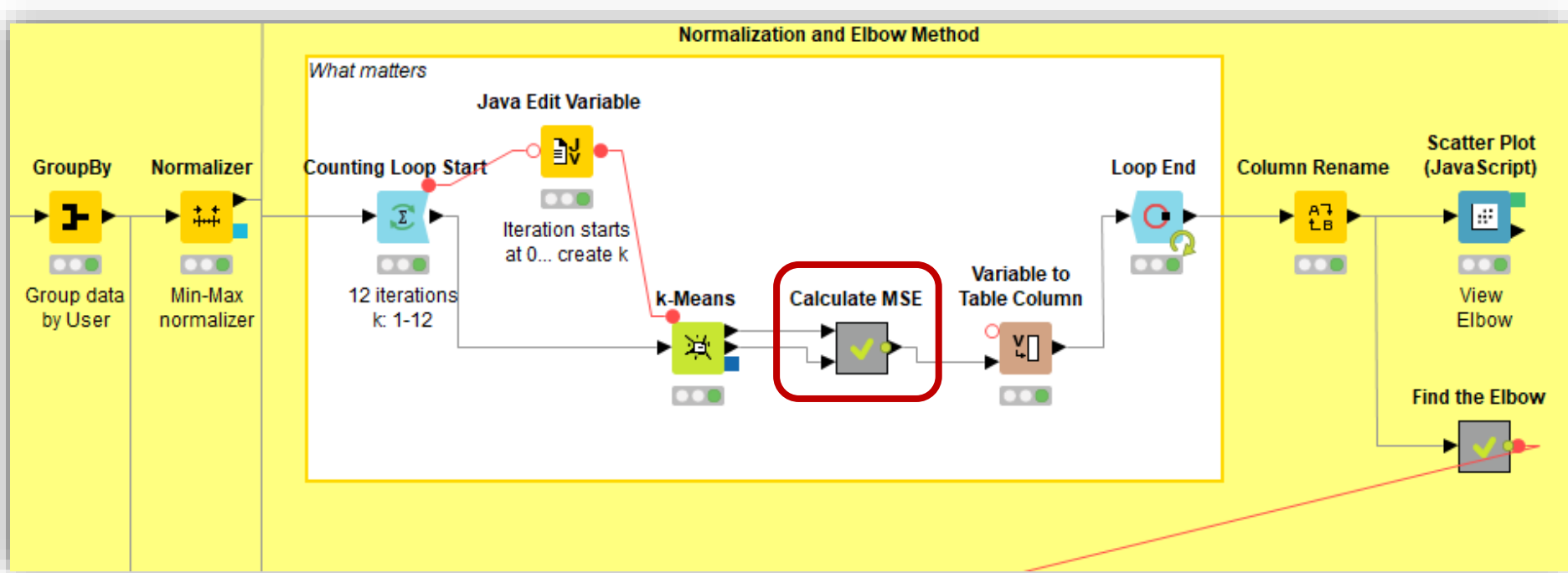
Quality Measures

31

The Elbow Method

QUALITY MEASURES

Hands On



Quality Measures

MAE, MSE and RMSE

32

The Elbow Method

QUALITY MEASURES

Hands On

MAE

Mean Absolute Error measures the average magnitude of the errors in a set of predictions, without considering their direction.

$$MAE = \frac{1}{n} \sum_{j=1}^n |y_j - \hat{y}_j|$$

MSE

Mean Squared Error consists of the average of squared differences between the prediction and the actual observation, without considering their direction

$$MSE = \frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2$$

RMSE

Root Mean Squared Error consists of the square root of the average of squared differences between the prediction and the actual observation, without considering their direction

$$RMSE = \sqrt{\frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2}$$

Where n is the number of observations, and y_j and \hat{y}_j are the actual observation and the predicted value, respectively.

Quality Measures for Clustering

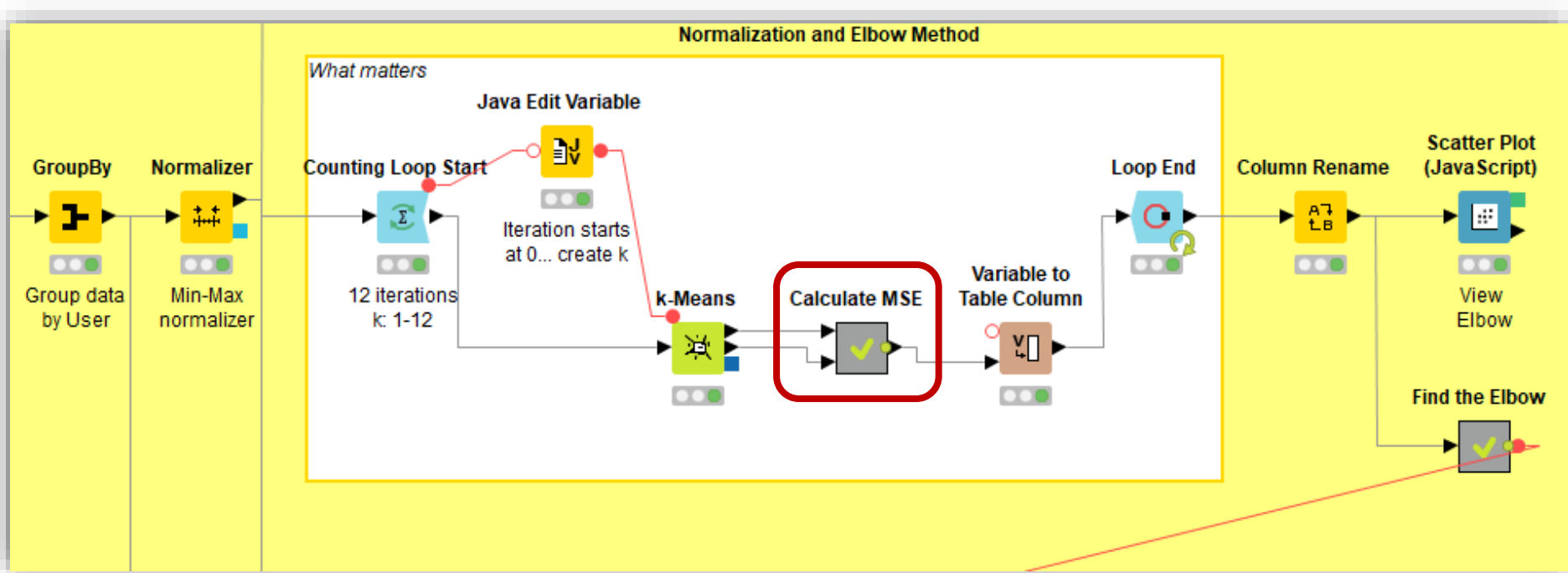
The Elbow Method

35

The Elbow Method

QUALITY MEASURES

Hands On



Distance from each point to the centroid of the cluster it belongs to!

Quality Measures for Clustering

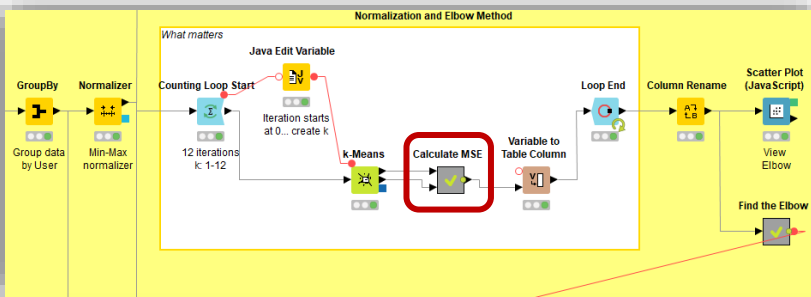
The Elbow Method

36

The Elbow Method

QUALITY MEASURES

Hands On



The **two outputs** of the **k-Means node**:

- (1) Input data labeled with the cluster;
- (2) The created clusters and centroids.

Labeled input - 0:144:0:141 - k-Means (1)

File Hilite Navigation View

Table "default" - Rows: 610 Spec - Columns: 12 Properties Flow Variables

Row ID	D Comedy	D Crime	D Drama	D Other	S Cluster
Row0	0.077	0.107	0.051	0.101	cluster_3
Row1	0.006	0.024	0.012	0.005	cluster_1
Row2	0.008	0.005	0.011	0.012	cluster_2
Row3	0.096	0.064	0.091	0.069	cluster_10
Row4	0.014	0.029	0.018	0.013	cluster_7
Row5	0.118	0.084	0.106	0.116	cluster_4
Row6	0.045	0.062	0.043	0.067	cluster_4
Row7	0.022	0.021	0.014	0.016	cluster_7
Row8	0.014	0.017	0.015	0.01	cluster_8
Row9	0.073	0.031	0.054	0.058	cluster_8
Row10	0.011	0.031	0.018	0.023	cluster_1
Row11	0.017	0.002	0.011	0.007	cluster_11
Row12	0.01	0.007	0.011	0.007	cluster_7
Row13	0.016	0.017	0.018	0.014	cluster_7
Row14	0.026	0.045	0.041	0.06	cluster_4
Row15	0.023	0.057	0.036	0.039	cluster_10
Row16	0.018	0.076	0.038	0.037	cluster_11
Row17	0.137	0.329	0.166	0.202	cluster_9
Row18	0.35	0.172	0.115	0.308	cluster_6
Row19	0.097	0.06	0.054	0.107	cluster_4
Row20	0.186	0.174	0.071	0.189	cluster_6

Clusters - 0:144:0:141 - k-Means (2)

File Hilite Navigation View

Table "default" - Rows: 12 Spec - Columns: 8 Properties Flow Variables

Row ID	D Animation	D Comedy	D Crime	D Drama	D Other
cluster_0	0.747	0.767	0.76	0.672	0.786
cluster_1	0.017	0.015	0.021	0.016	0.014
cluster_2	0.026	0.02	0.024	0.018	0.022
cluster_3	0.113	0.107	0.127	0.098	0.119
cluster_4	0.153	0.085	0.073	0.058	0.085
cluster_5	0.397	0.386	0.392	0.315	0.408
cluster_6	0.477	0.257	0.154	0.14	0.229
cluster_7	0.02	0.019	0.021	0.016	0.017
cluster_8	0.025	0.023	0.025	0.021	0.021
cluster_9	0.167	0.186	0.237	0.17	0.223
cluster_10	0.049	0.051	0.066	0.055	0.059
cluster_11	0.026	0.014	0.03	0.021	0.019

Quality Measures for Clustering

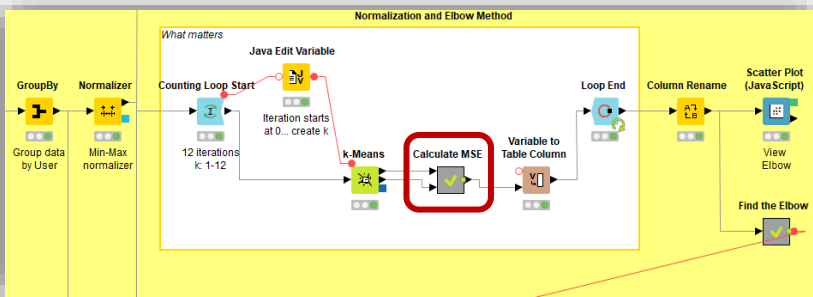
The Elbow Method

37

The Elbow Method

QUALITY MEASURES

Hands On



Observation with id Row0 was assigned to cluster 3. Its value for comedy is 0.077. How far is it from the centroid's center of cluster 3 (0.107)? And for the other genres?

How far is this observation from the centroid of the cluster?

Clusters - 0:144:0:141 - k-Means

File Hilite Navigation View

Table "default" - Rows: 12 Spec - Columns: 8 Properties Flow Variables

Row ID	D Animation	D Comedy	D Crime	D Drama	D Other
cluster_0	0.747	0.767	0.76	0.672	0.786
cluster_1	0.017	0.015	0.021	0.016	0.014
cluster_2	0.026	0.02	0.024	0.018	0.022
cluster_3	0.113	0.107	0.127	0.098	0.119
cluster_4	0.153	0.085	0.073	0.058	0.085
cluster_5	0.397	0.386	0.392	0.315	0.408
cluster_6	0.477	0.257	0.154	0.14	0.229
cluster_7	0.02	0.019	0.021	0.016	0.017
cluster_8	0.025	0.023	0.025	0.021	0.021
cluster_9	0.167	0.186	0.237	0.17	0.223
cluster_10	0.049	0.051	0.066	0.055	0.059
cluster_11	0.026	0.014	0.03	0.021	0.019

Labeled input - 0:144:0:141 - k-Means

File Hilite Navigation View

Table "default" - Rows: 610 Spec - Columns: 12 Properties Flow Variables

Row ID	D Comedy	D Crime	D Drama	D Other	S Cluster
Row0	0.077	0.107	0.051	0.101	cluster_3
Row1	0.006	0.024	0.012	0.005	cluster_1
Row2	0.008	0.005	0.011	0.012	cluster_2
Row3	0.096	0.064	0.091	0.069	cluster_10
Row4	0.014	0.029	0.018	0.013	cluster_7
Row5	0.118	0.084	0.106	0.116	cluster_4
Row6	0.045	0.062	0.043	0.067	cluster_4
Row7	0.022	0.021	0.014	0.016	cluster_7
Row8	0.014	0.017	0.015	0.01	cluster_8
Row9	0.073	0.031	0.054	0.058	cluster_8
Row10	0.011	0.031	0.018	0.023	cluster_1
Row11	0.017	0.002	0.011	0.007	cluster_11
Row12	0.01	0.007	0.011	0.007	cluster_7
Row13	0.016	0.017	0.018	0.014	cluster_7
Row14	0.026	0.045	0.041	0.06	cluster_4
Row15	0.023	0.057	0.036	0.039	cluster_10
Row16	0.018	0.076	0.038	0.037	cluster_11
Row17	0.137	0.329	0.166	0.202	cluster_9
Row18	0.35	0.172	0.115	0.308	cluster_6
Row19	0.097	0.06	0.054	0.107	cluster_4
Row20	0.186	0.174	0.071	0.189	cluster_6

Quality Measures for Clustering

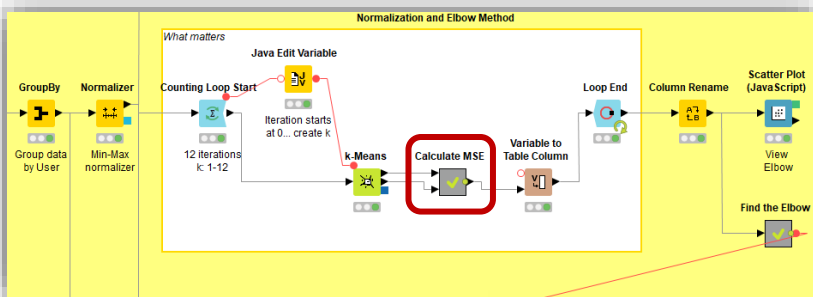
The Elbow Method

38

The Elbow Method

QUALITY MEASURES

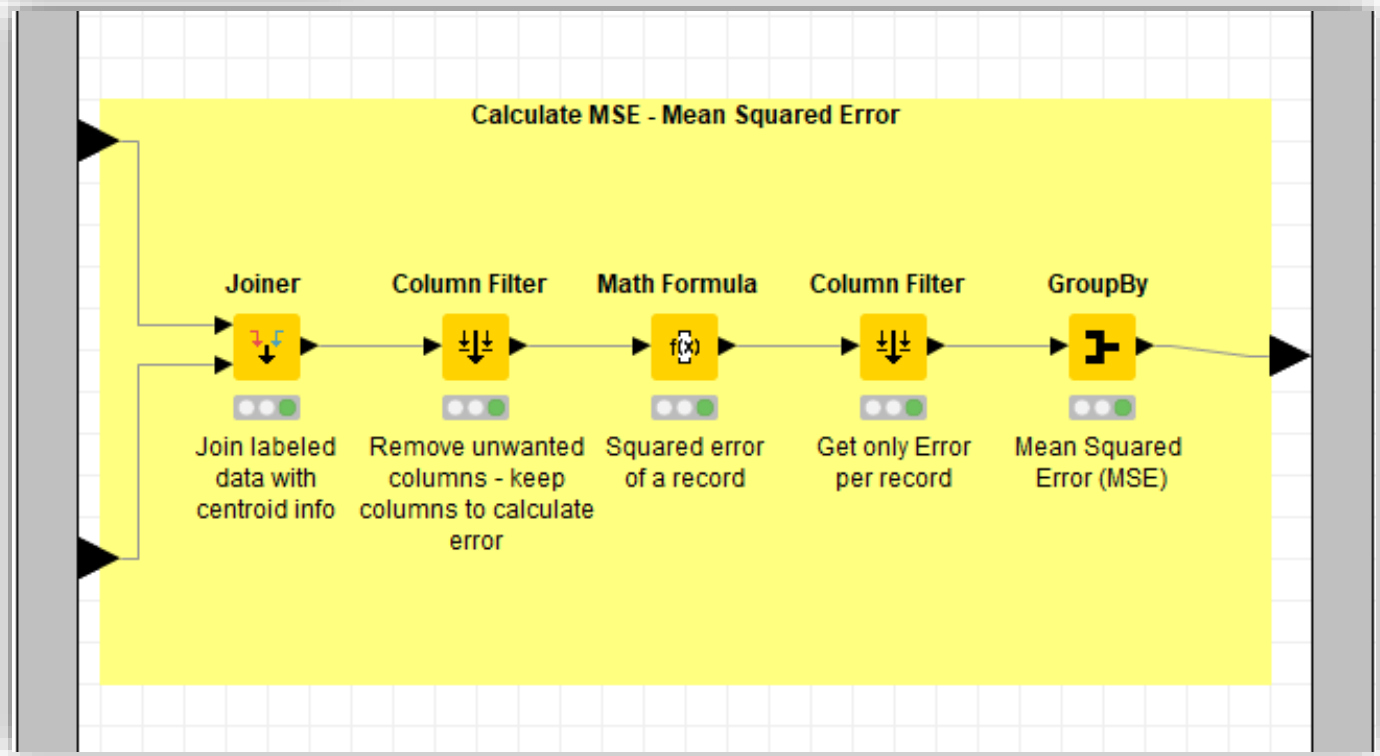
Hands On



We may use MSE, MAE or RMSE to compute this error metric, i.e., **how far are records from the centroid's of their cluster.**

Input: The input data labeled with the cluster they belong.

Input: The created clusters and centroids.



Quality Measures for Clustering

The Elbow Method

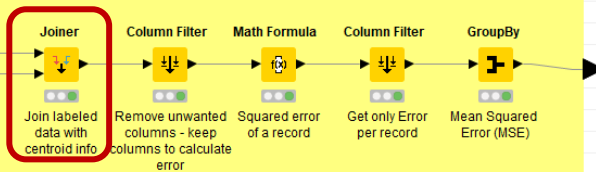
39

The Elbow Method

QUALITY MEASURES

Hands On

Calculate MSE - Mean Squared Error



Joined table - 2:142:0:126:130 - Joiner (Join labeled)

File Hilite Navigation View

Table "default" - Rows: 9742 Spec - Columns: 17 Properties Flow Variables

Row ID	Animation	Comedy	Drama	Crime	Other	Cluster	Advent...	Action (...)	Animati...	Comed...	
Row0_cluster_0	1	0	0	0	1	cluster_0	0.572	0.044	0.783	0.629	0
Row1_cluster_7	0	0	0	0	1	cluster_7	1	0	0.055	0	0
Row2_cluster_6	1	0	0	0	1	cluster_6	0	0	0	1	0
Row3_cluster...	1	1	0	0	1	cluster_10	0.027	0.012	0.019	1	1
Row4_cluster_4	1	0	0	0	0	cluster_4	0.033	0.078	0.028	0.99	0
Row5_cluster_5	0	0	1	1	1	cluster_5	0	1	0.053	0	0
Row6_cluster_6	1	0	0	0	1	cluster_6	0	0	0	1	0
Row7_cluster_7	0	0	0	0	1	cluster_7	1	0	0.055	0	0
Row8_cluster_8	0	0	0	0	0	cluster_8	0.31	1	0.087	0.142	0
Row9_cluster_9	0	0	0	0	1	cluster_9	1	1	0.073	0.136	0
Row10_duste...	1	1	0	1	1	cluster_10	0.027	0.012	0.019	1	1
Row11_duste...	1	0	0	0	1	cluster_6	0	0	0	1	0
Row12_duste...	0	0	0	0	1	cluster_0	0.572	0.044	0.783	0.629	0
Row13_duste...	0	1	0	0	0	cluster_3	0.018	0	0.014	0.15	1
Row14_duste...	0	0	0	0	1	cluster_9	1	1	0.073	0.136	0
Row15_duste...	0	1	1	0	0	cluster_3	0.018	0	0.014	0.15	1
Row16_duste...	0	1	0	1	1	cluster_3	0.018	0	0.014	0.15	1
Row17_duste...	1	0	0	0	0	cluster_4	0.033	0.078	0.028	0.99	0

Quality Measures for Clustering

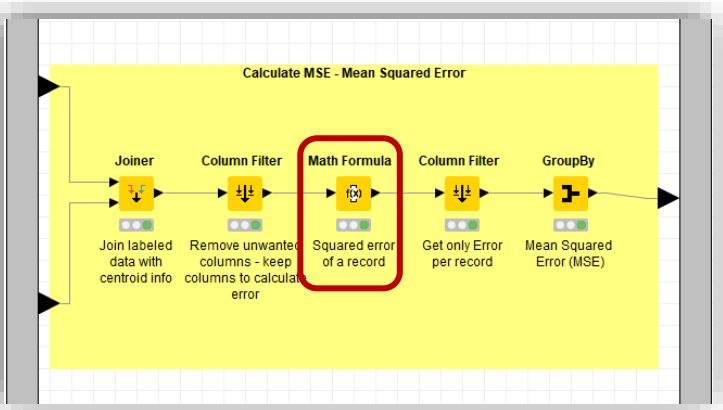
The Elbow Method

40

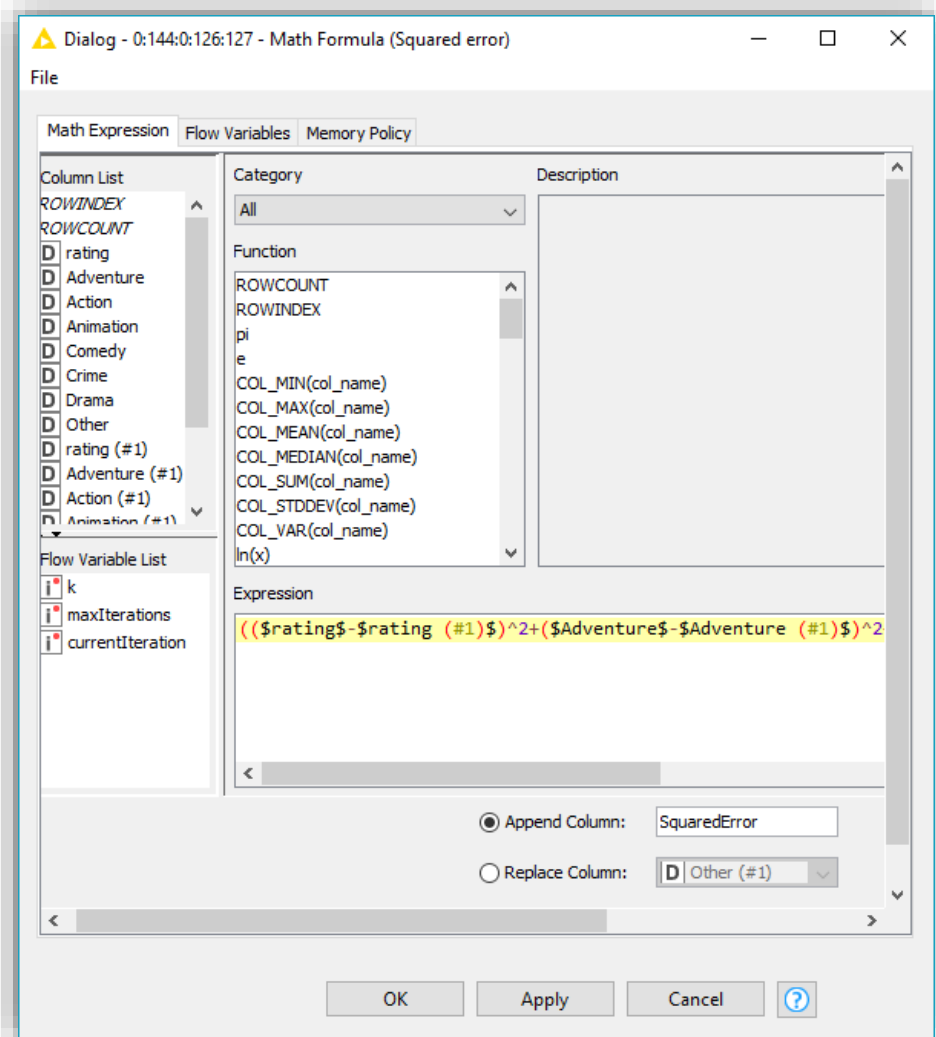
The Elbow Method

QUALITY MEASURES

Hands On



$$MSE = \frac{1}{n} \sum_{j=1}^n (y_j - \hat{y}_j)^2$$



Quality Measures for Clustering

The Elbow Method

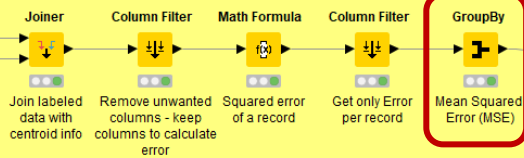
41

The Elbow Method

QUALITY MEASURES

Hands On

Calculate MSE - Mean Squared Error



Group table - 2:142:0:126:129 - GroupBy (Mean Squared)

File Hilite Navigation View

Table "default" - Rows: 1 Spec - Column: 1 Properties Flow Variables

Row ID	D Mean(SquaredError)
Row0	0.053

Quality Measures for Clustering

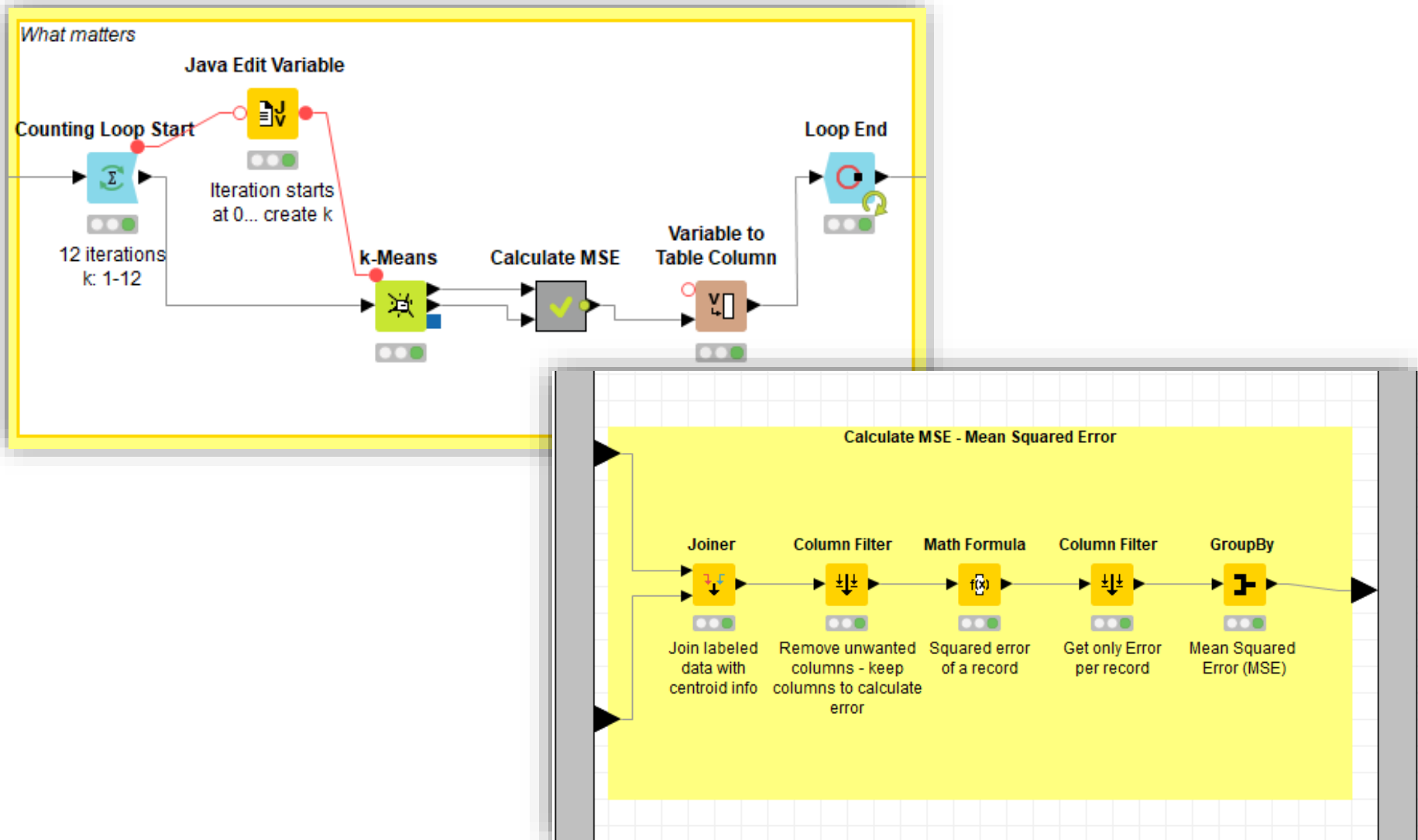
The Elbow Method Metanode

42

The Elbow Method

QUALITY MEASURES

Hands On



Quality Measures for Clustering

The Elbow Method Metanode

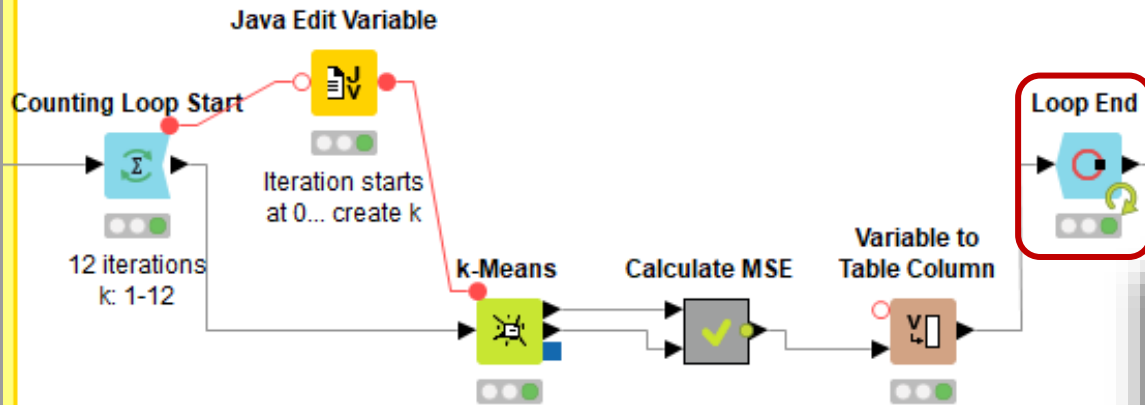
43

The Elbow Method

QUALITY MEASURES

Hands On

What matters



Collected results - ...

File Hilite Navigation View

Properties		Flow Variables	
Table "default" - Rows: 12		Spec - Columns: 2	
Row ID	Mean(S...	k	
Row0#0	0.013	1	
Row0#1	0.006	2	
Row0#2	0.005	3	
Row0#3	0.004	4	
Row0#4	0.003	5	
Row0#5	0.003	6	
Row0#6	0.003	7	
Row0#7	0.002	8	
Row0#8	0.002	9	
Row0#9	0.002	10	
Row0#10	0.002	11	
Row0#11	0.002	12	

Quality Measures for Clustering

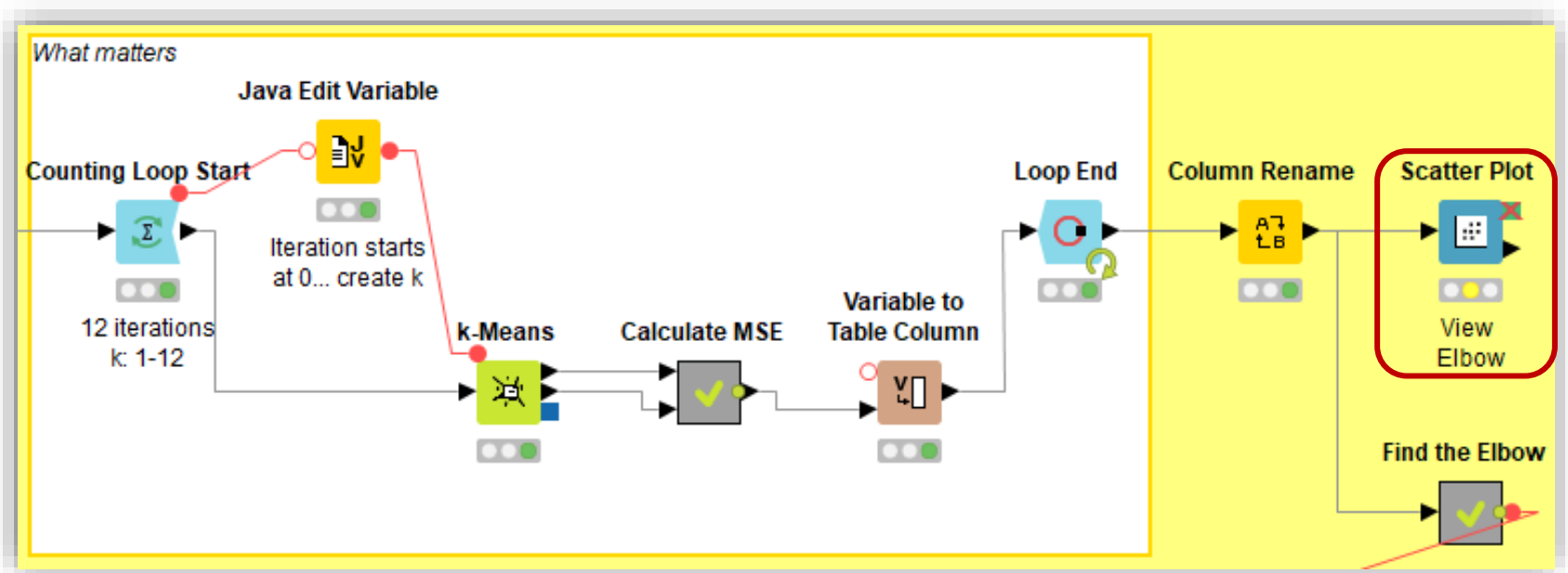
View the Elbow

44

The Elbow Method

QUALITY MEASURES

Hands On



Quality Measures for Clustering

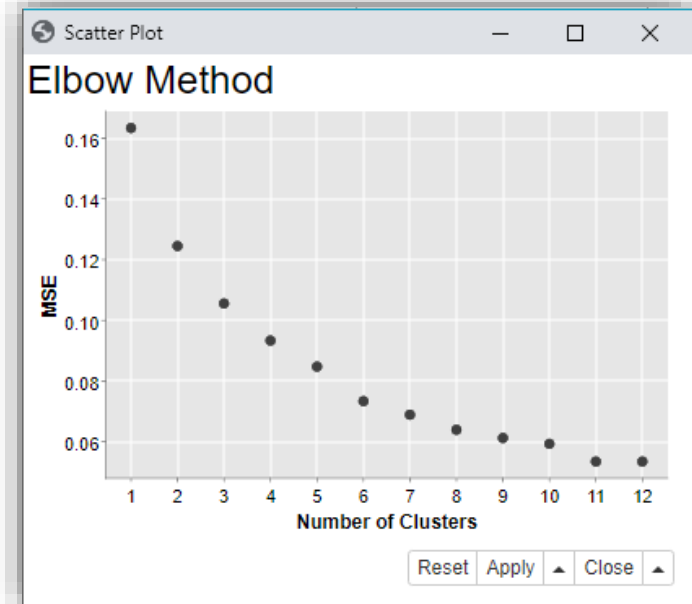
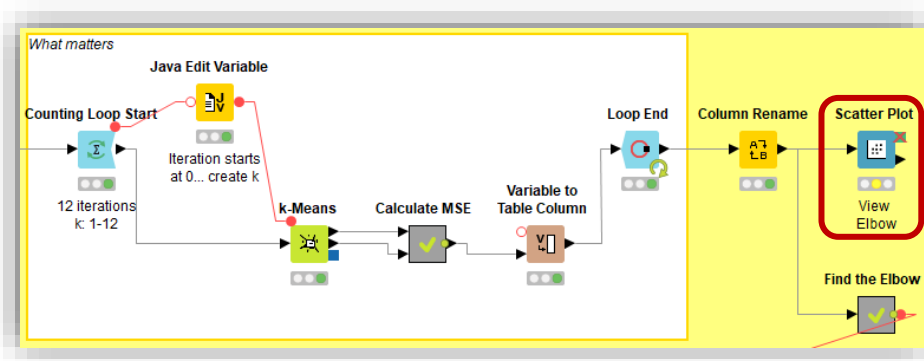
View the Elbow

45

The Elbow Method

QUALITY MEASURES

Hands On



Quality Measures for Clustering

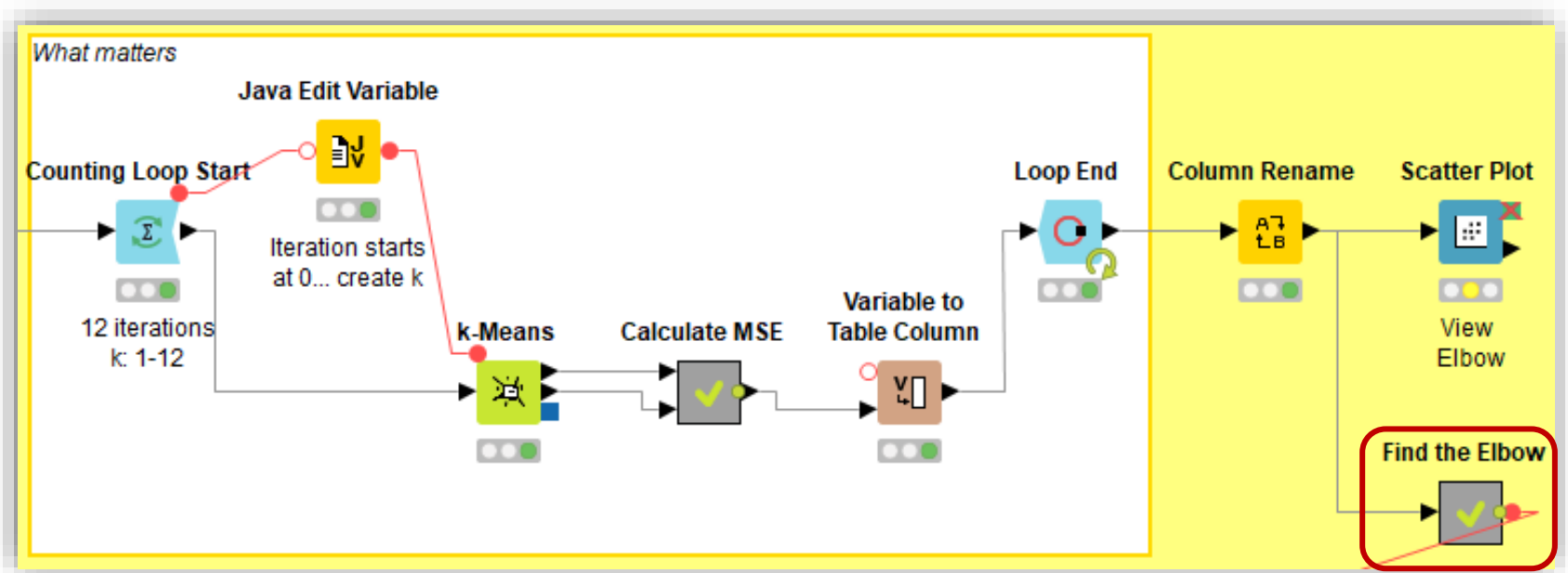
Finding the Elbow ... Automatically

46

The Elbow Method

QUALITY MEASURES

Hands On



Quality Measures for Clustering

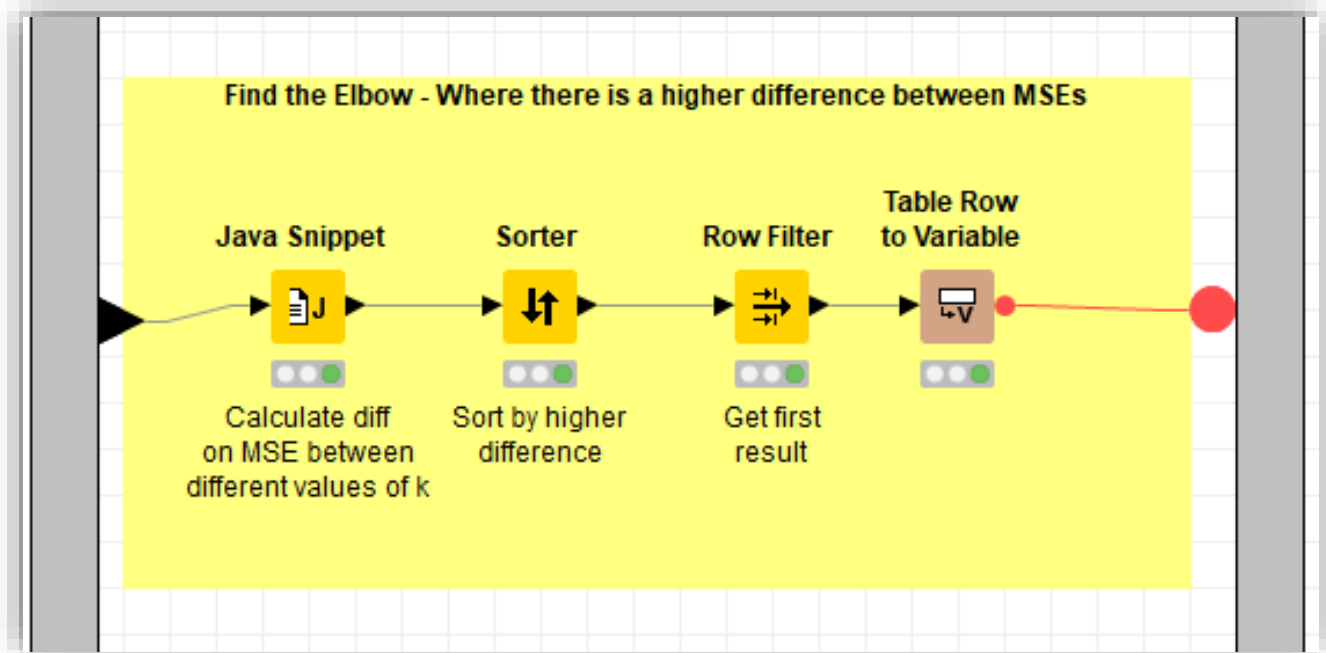
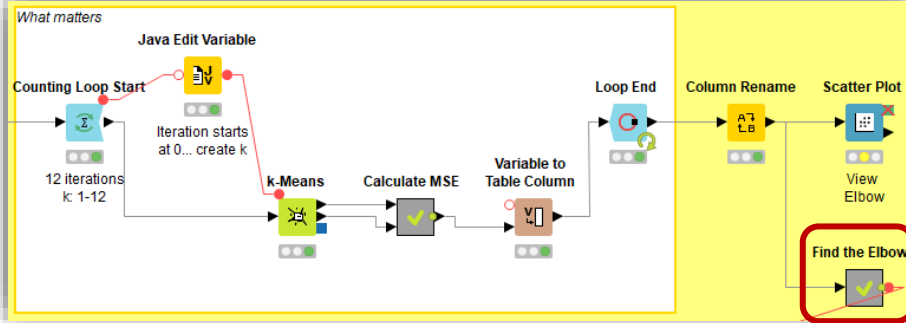
Finding the Elbow ... Automatically

47

The Elbow Method

QUALITY MEASURES

Hands On



Quality Measures for Clustering

Finding the Elbow ... Automatically

48

The Elbow Method

QUALITY MEASURES

Hands On

Find the Elbow - Where there is a higher difference between MSEs

Java Snippet

Calculate diff
on MSE between
different values of k

Sorter

Sort by higher
difference

Row Filter

Get first
result

Table Row
to Variable

Appended table - 0:142:0:137:132 - Jav...

File Hilite Navigation View

Table "default" - Rows: 12 Spec - Columns: 3 Properties Flow Variables

Row ID	D MSE	I k	D MseDiff
Row0#0	0.163	1	0
Row0#1	0.124	2	0.039
Row0#2	0.106	3	0.019
Row0#3	0.093	4	0.012
Row0#4	0.085	5	0.009
Row0#5	0.073	6	0.011
Row0#6	0.069	7	0.004
Row0#7	0.064	8	0.005
Row0#8	0.061	9	0.003
Row0#9	0.059	10	0.002
Row0#10	0.053	11	0.006
Row0#11	0.053	12	0

Sorted Table - 0:142:0:137:133 - Sorter (...)

File Hilite Navigation View

Table "default" - Rows: 12 Spec - Columns: 3 Properties Flow Variables

Row ID	D MSE	I k	D MseDiff
Row0#1	0.124	2	0.039
Row0#2	0.106	3	0.019
Row0#3	0.093	4	0.012
Row0#5	0.073	6	0.011
Row0#4	0.085	5	0.009
Row0#10	0.053	11	0.006
Row0#7	0.064	8	0.005
Row0#6	0.069	7	0.004
Row0#8	0.061	9	0.003
Row0#9	0.059	10	0.002
Row0#0	0.163	1	0
Row0#11	0.053	12	0

Hands On

49

The Elbow Method

Quality Measures

HANDS ON

HANDS ON

