

# Machine Learning with WEKA

Eibe Frank

Department of Computer Science,  
University of Waikato, New Zealand

- WEKA: A Machine Learning Toolkit
- The Explorer
  - Classification and Regression
  - Clustering
  - Association Rules
  - Attribute Selection
  - Data Visualization
- The Experimenter
- The Knowledge Flow GUI
- Conclusions

# WEKA: the bird



*Copyright: Martin Kramer (mkramer@wxs.nl)*

# WEKA: the software

- Machine learning/data mining software written in Java (distributed under the GNU Public License)
- Used for research, education, and applications
- Complements “Data Mining” by Witten & Frank
- Main features:
  - ◆ Comprehensive set of data pre-processing tools, learning algorithms and evaluation methods
  - ◆ Graphical user interfaces (incl. data visualization)
  - ◆ Environment for comparing learning algorithms

# WEKA: versions

- There are several versions of WEKA:
  - ◆ WEKA 3.0: “book version” compatible with description in data mining book
  - ◆ WEKA 3.2: “GUI version” adds graphical user interfaces (book version is command-line only)
  - ◆ WEKA 3.3: “development version” with lots of improvements
- This talk is based on the latest snapshot of WEKA 3.3 (soon to be WEKA 3.4)

# WEKA only deals with “flat” files

```
@relation heart-disease-simplified
```

```
@attribute age numeric  
@attribute sex { female, male}  
@attribute chest_pain_type { typ_angina, asympt, non_anginal, atyp_angina}  
@attribute cholesterol numeric  
@attribute exercise_induced_angina { no, yes}  
@attribute class { present, not_present}
```

```
@data  
63,male,typ_angina,233,no,not_present  
67,male,asympt,286,yes,present  
67,male,asympt,229,yes,present  
38,female,non_anginal,?,no,not_present  
...
```

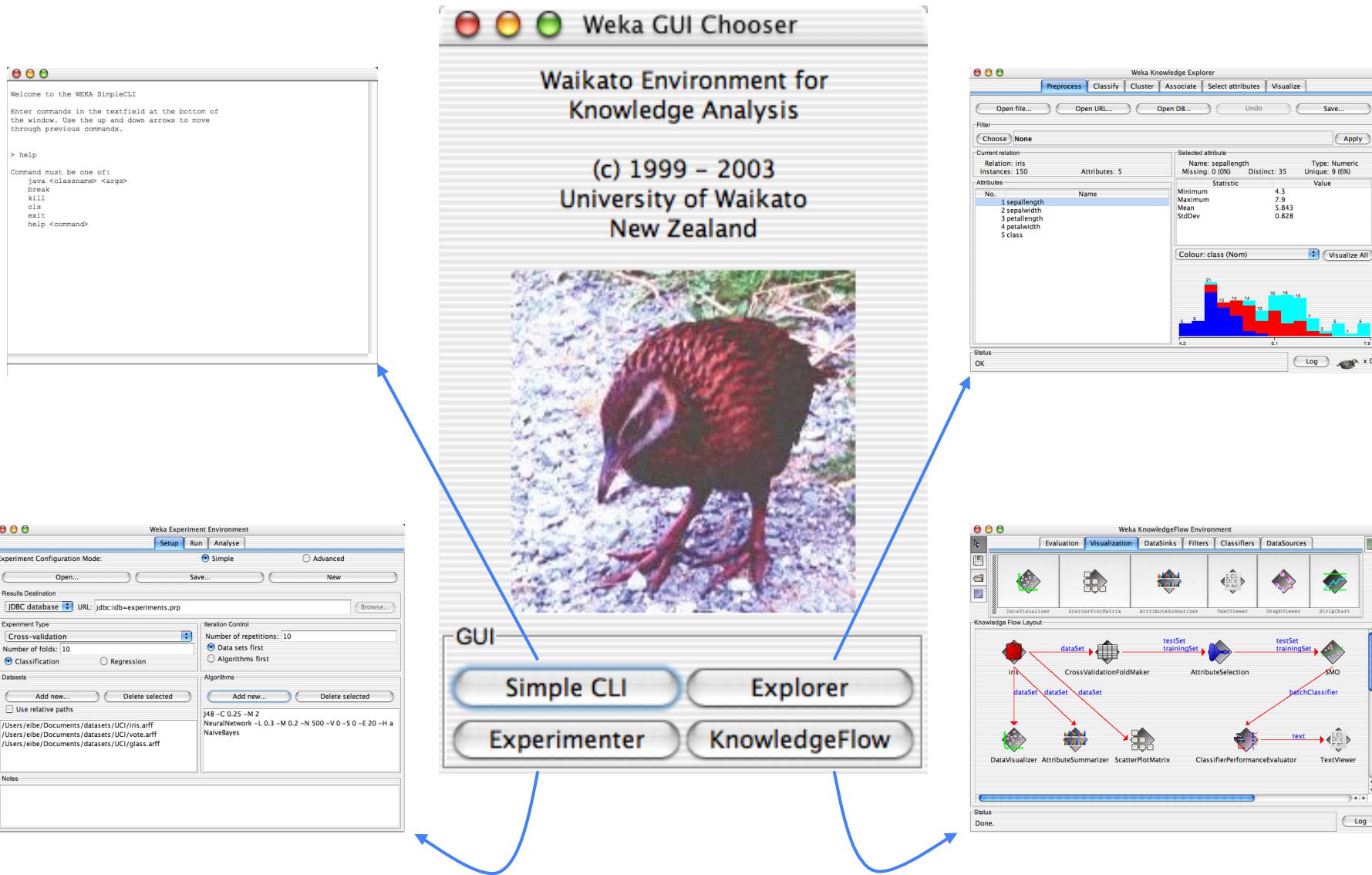


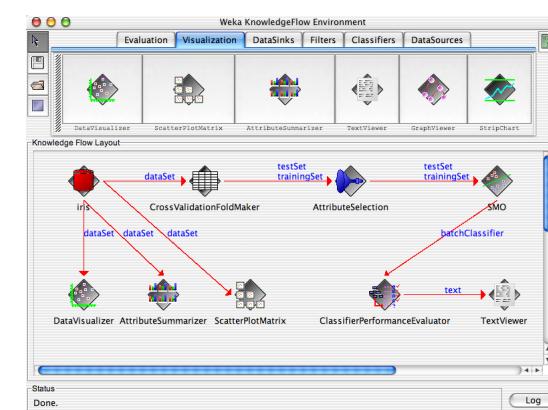
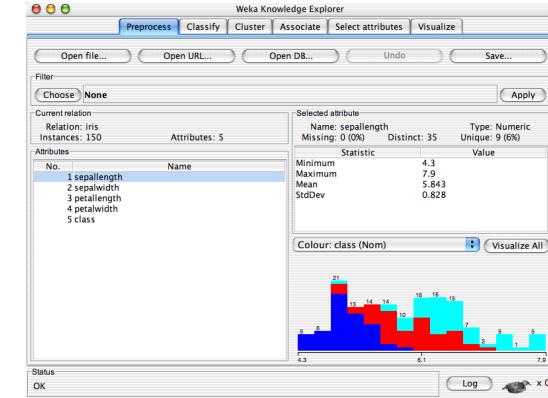
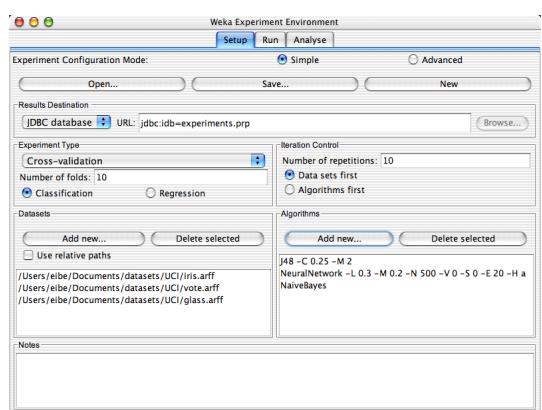
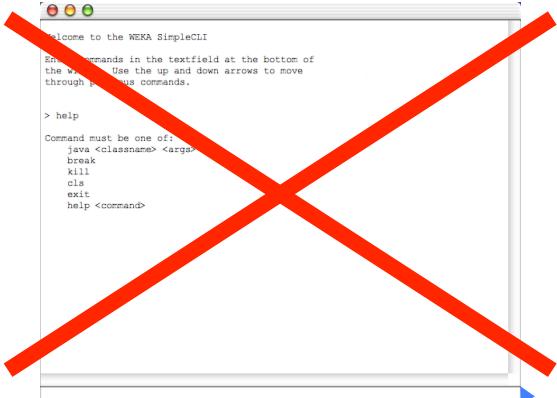
Flat file in  
ARFF format

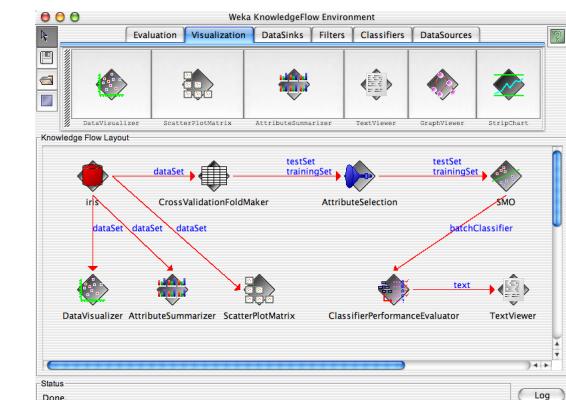
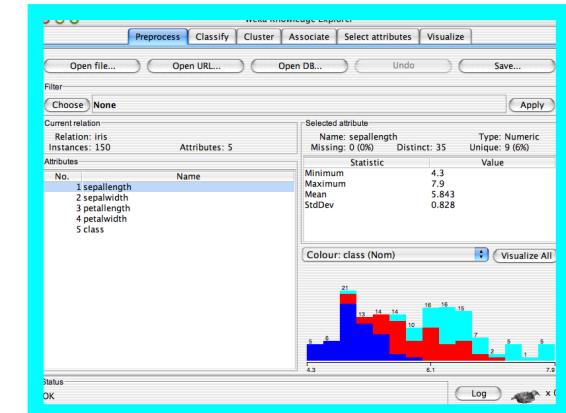
# WEKA only deals with “flat” files

```
@relation heart-disease-simplified  
  
@attribute age numeric ← numeric attribute  
@attribute sex { female, male} ← nominal attribute  
@attribute chest_pain_type { typ_angina, asympt, non_anginal, atyp_angina}  
@attribute cholesterol numeric  
@attribute exercise_induced_angina { no, yes}  
@attribute class { present, not_present}
```

```
@data  
63,male,typ_angina,233,no,not_present  
67,male,asympt,286,yes,present  
67,male,asympt,229,yes,present  
38,female,non_anginal,?,no,not_present  
...
```







# Explorer: pre-processing the data

- Data can be imported from a file in various formats: ARFF, CSV, C4.5, binary
- Data can also be read from a URL or from an SQL database (using JDBC)
- Pre-processing tools in WEKA are called “filters”
- WEKA contains filters for:
  - ◆ Discretization, normalization, resampling, attribute selection, transforming and combining attributes, ...

**Weka Knowledge Explorer**

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Undo Save...

Filter

Choose **None** Apply

Current relation

Relation: None Instances: None Attributes: None

Selected attribute

Name: None Missing: None Distinct: None Type: None Unique: None

Attributes

Visualize All

Status

Welcome to the Weka Knowledge Explorer

Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Undo Save...

Filter

Choose None Apply

Current relation

Relation: None Instances: None Attributes: None

Selected attribute

Name: None Missing: None Distinct: None Type: None Unique: None

Attributes

Visualize All

Status

Welcome to the Weka Knowledge Explorer

Log x 0

A red arrow points from the 'Open file...' button in the top toolbar to the 'Choose' button in the 'Filter' section.

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Undo Save...

Filter

Choose None Apply

Current relation

Relation: iris Instances: 150 Attributes: 5

Attributes

No.	Name
1	sepallength
2	sepalwidth
3	petallength
4	petalwidth
5	class

Selected attribute

Name: sepallength Type: Numeric  
Missing: 0 (0%) Distinct: 35 Unique: 9 (6%)

Statistic	Value
Minimum	4.3
Maximum	7.9
Mean	5.843
StdDev	0.828

Colour: class (Nom) Visualize All

4.3 6.1 7.9

5 6 21 13 14 14 10 16 16 15 7 2 5 1 5

Status

OK Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Undo Save...

Filter

Choose None Apply

Current relation

Relation: iris Instances: 150 Attributes: 5

Attributes

No.	Name
1	sepallength
2	sepalwidth
3	petallength
4	petalwidth
5	class

Selected attribute

Name: sepallength Type: Numeric  
Missing: 0 (0%) Distinct: 35 Unique: 9 (6%)

Statistic	Value
Minimum	4.3
Maximum	7.9
Mean	5.843
StdDev	0.828

Colour: class (Nom) Visualize All

Status

OK Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Undo Save...

Filter

Choose None Apply

Current relation

Relation: iris Instances: 150 Attributes: 5

Attributes

No.	Name
1	sepallength
2	sepalwidth
3	petallength
4	petalwidth
5	class

Selected attribute

Name: class Type: Nominal  
Missing: 0 (0%) Distinct: 3 Unique: 0 (0%)

Label	Count
Iris-setosa	50
Iris-versicolor	50
Iris-virginica	50

Colour: class (Nom) Visualize All

50 50 50

Status

OK Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Undo Save...

Filter

Choose None Apply

Current relation

Relation: iris Instances: 150 Attributes: 5

Attributes

No.	Name
1	sepallength
2	sepalwidth
3	petallength
4	petalwidth
5	class

Selected attribute

Name: class Type: Nominal  
Missing: 0 (0%) Distinct: 3 Unique: 0 (0%)

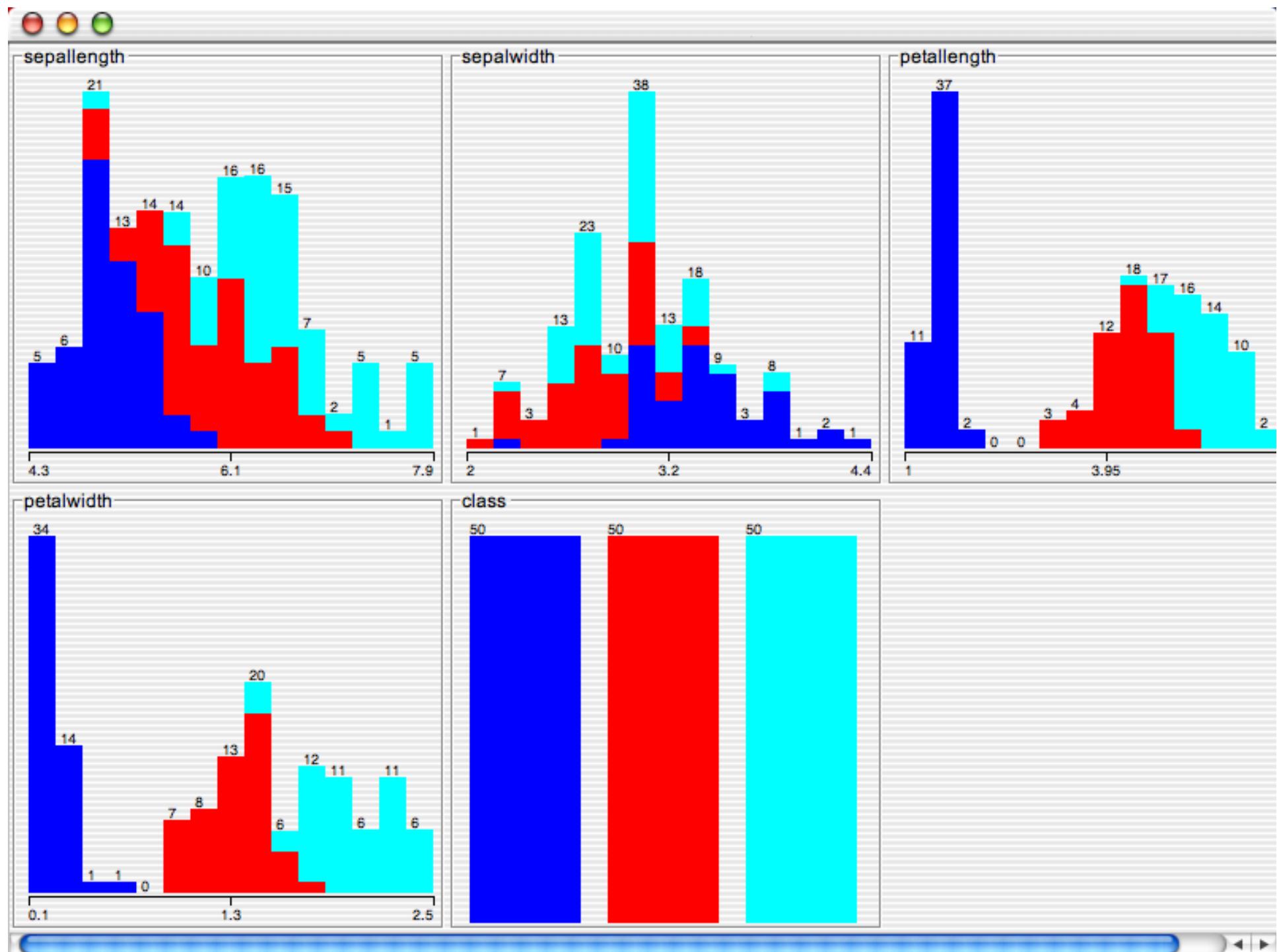
Label	Count
Iris-setosa	50
Iris-versicolor	50
Iris-virginica	50

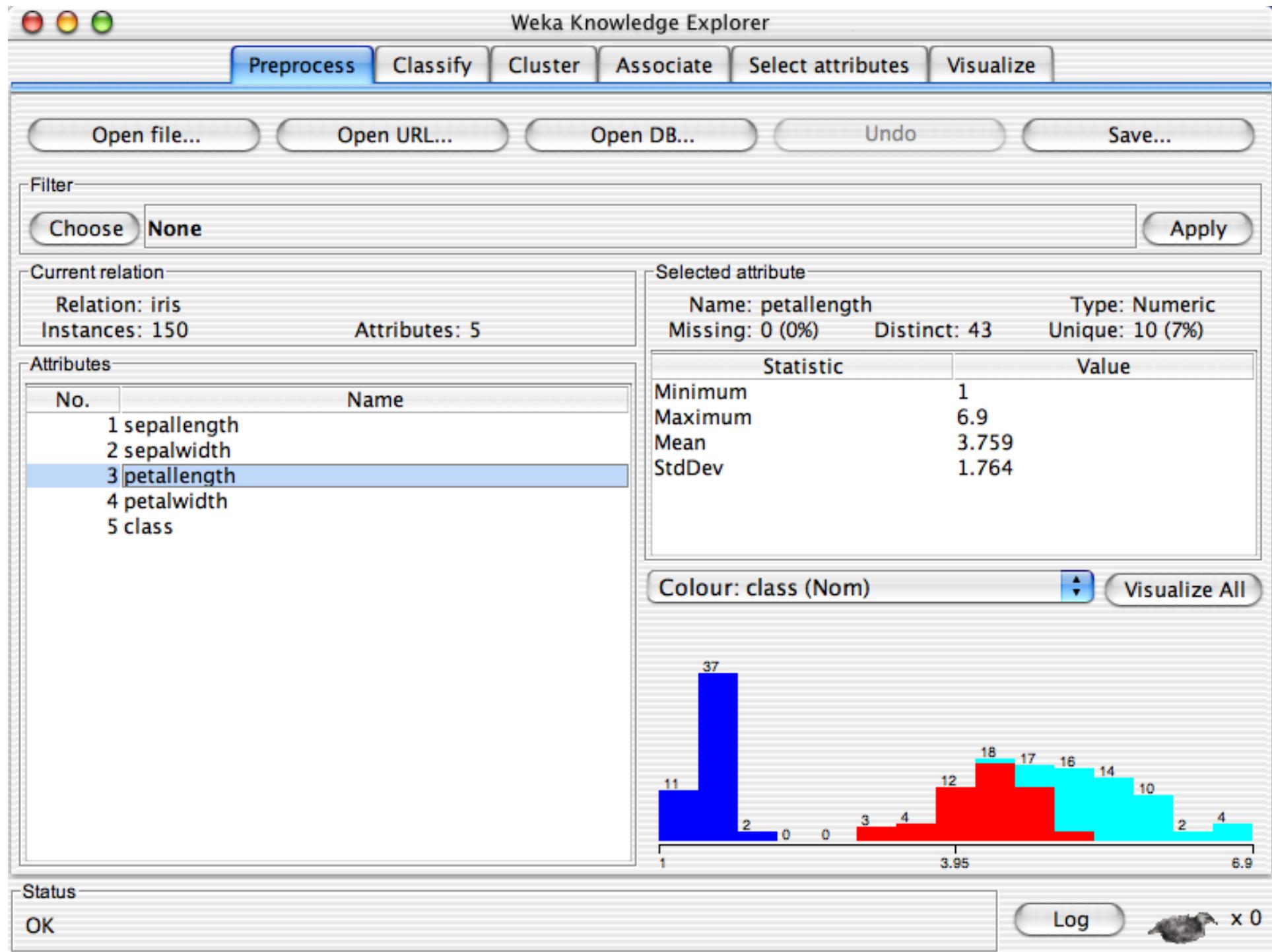
Colour: class (Nom) Visualize All

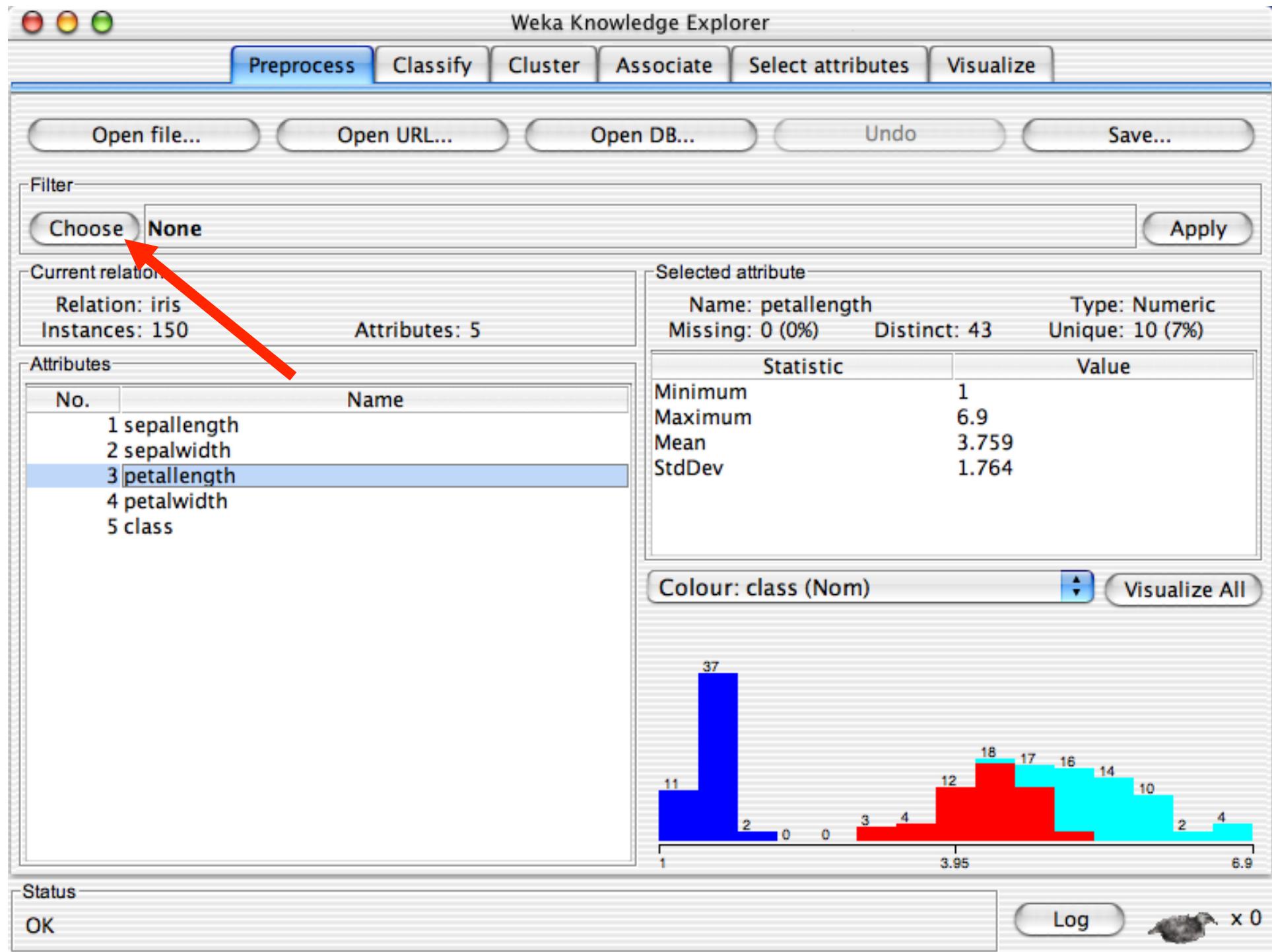
50 50 50

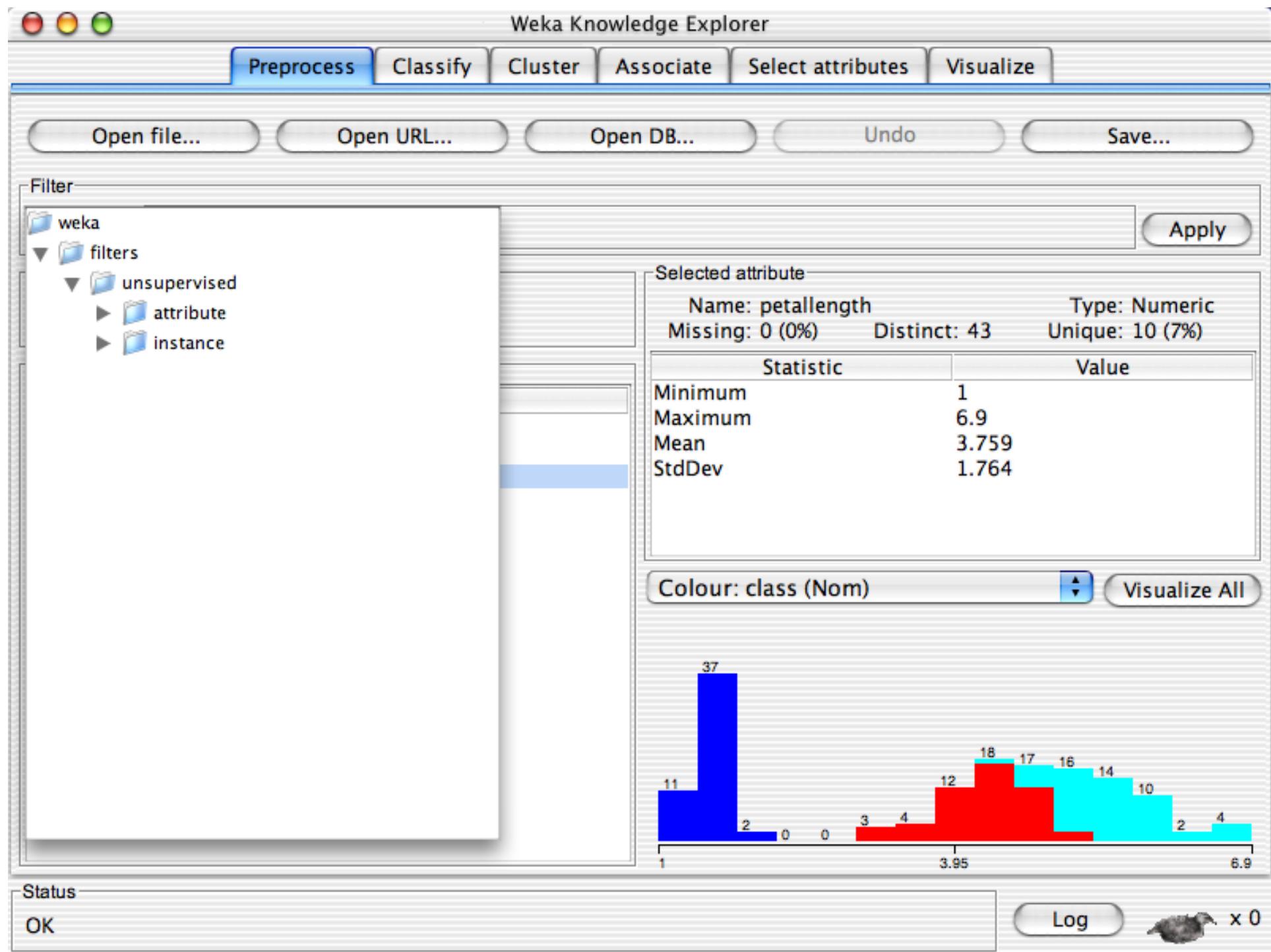
Status

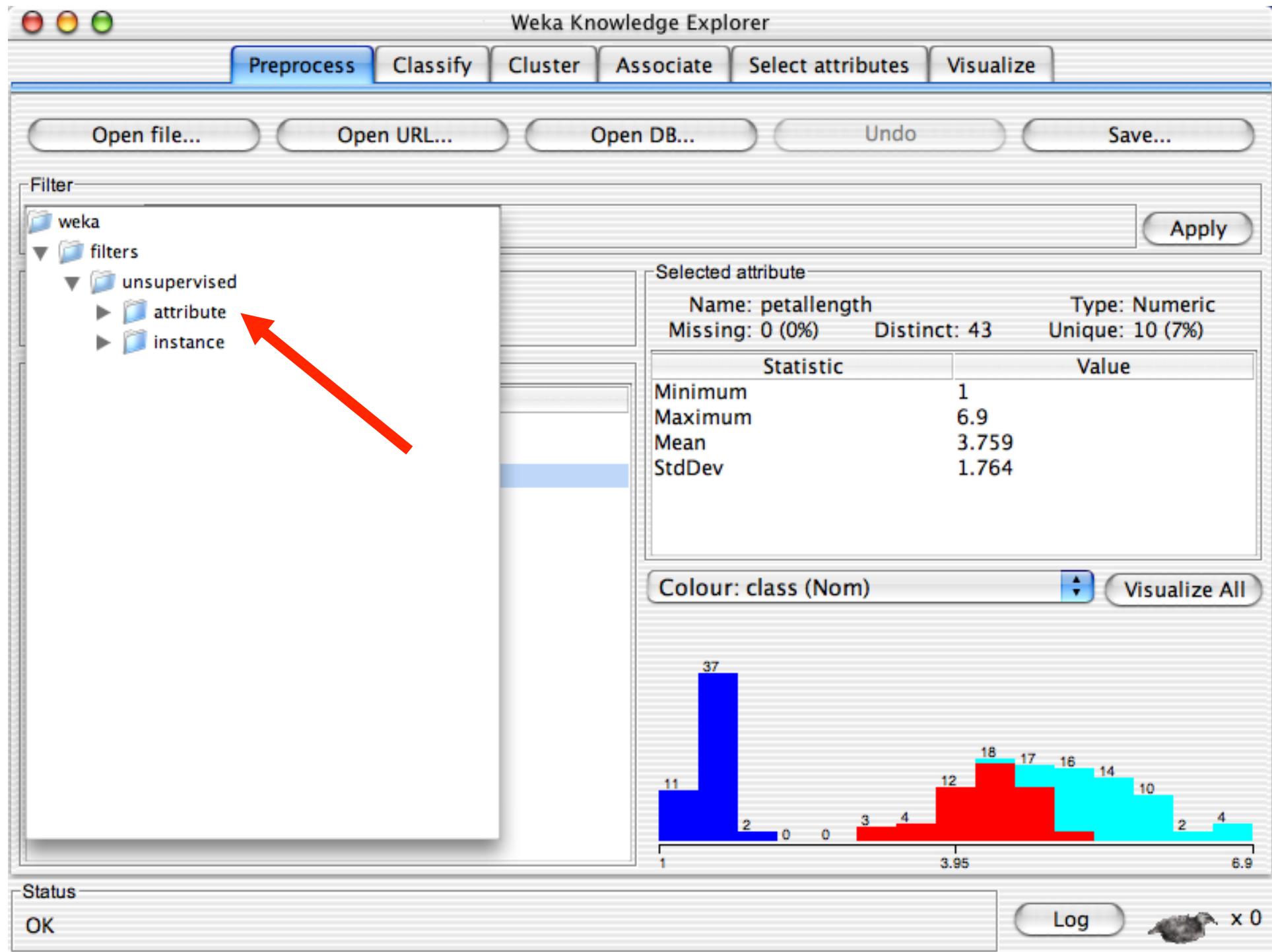
OK Log x 0











Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Undo Save...

Filter

weka

filters

  unsupervised

    attribute

- Add
- AddCluster
- AddExpression
- AddNoise
- Copy
- Discretize
- FirstOrder
- MakeIndicator
- MergeTwoValues
- NominalToBinary
- Normalize
- NumericToBinary
- NumericTransform
- Obfuscate
- PKIDiscretize
- Remove
- RemoveType

Selected attribute

Name: petallength      Type: Numeric  
Missing: 0 (0%)      Distinct: 43      Unique: 10 (7%)

Statistic	Value
Minimum	1
Maximum	6.9
Mean	3.759
StdDev	1.764

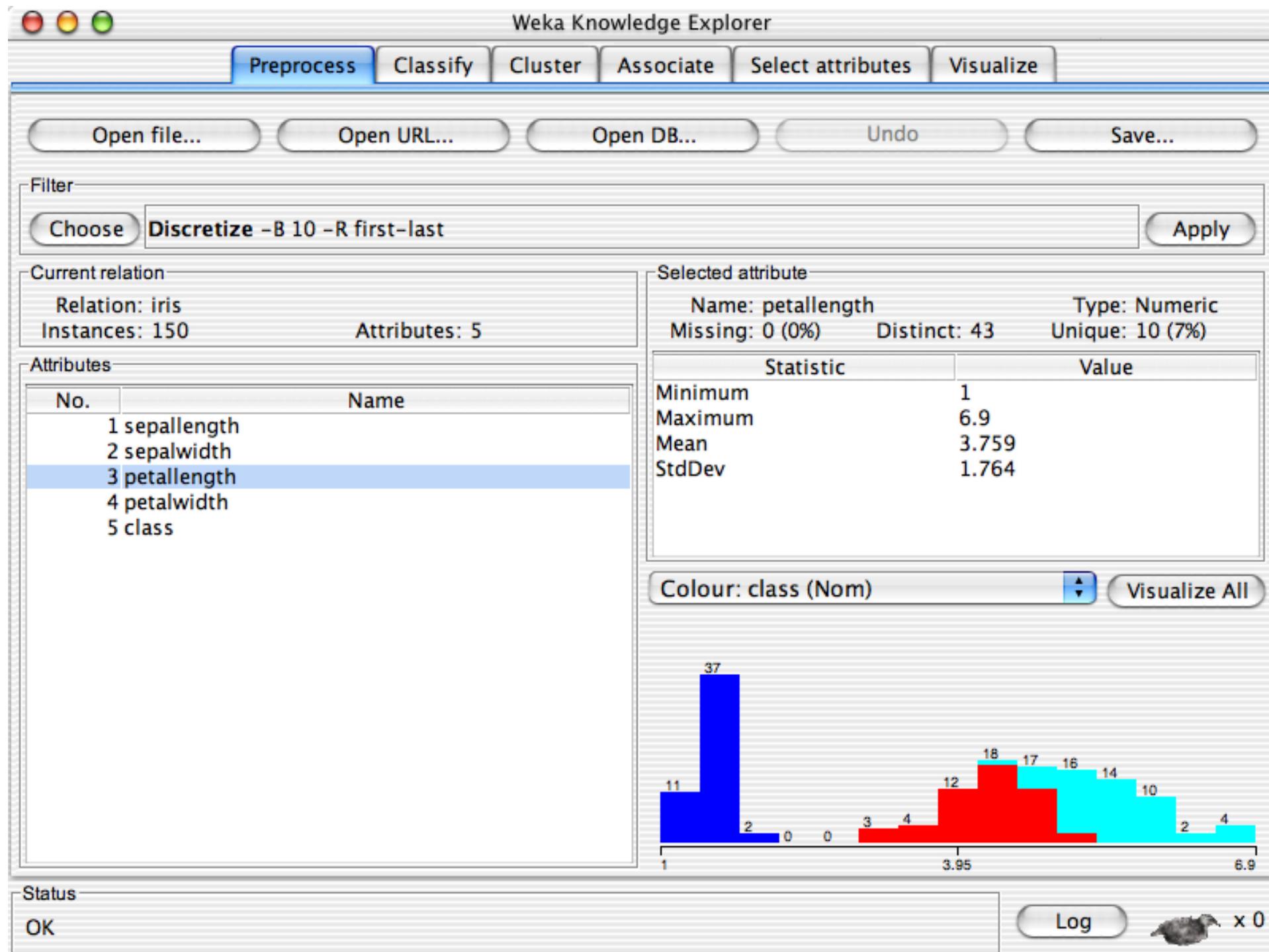
Colour: class (Nom)      Visualize All

A histogram visualizing the distribution of the 'petallength' attribute. The x-axis represents petallength values from 1 to 6.9, with major ticks at 1, 3.95, and 6.9. The y-axis represents frequency. The distribution is right-skewed, with the highest frequency bin (1-1.5) containing 37 instances. Other visible frequencies include 11, 2, 0, 0, 3, 4, 12, 18, 17, 16, 14, 10, 2, and 4.

Status

OK

Log x 0



Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Undo Save...

Filter

Choose Discretize -B 10 -R first-last **Apply**

Current relation  
Relation: iris Instances: 150 Attributes: 5

Attributes

No.	Name
1	sepallength
2	sepalwidth
3	<b>petallength</b>
4	petalwidth
5	class

Selected attribute  
Name: petallength Type: Numeric  
Missing: 0 (0%) Distinct: 43 Unique: 10 (7%)

Statistic	Value
Minimum	1
Maximum	6.9
Mean	3.759
StdDev	1.764

Colour: class (Nom) **Visualize All**

Status OK Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Undo Save...

Filter

Choose Discretize -B 10 -R first-last weka.gui.GenericObjectEditor Apply

Current relation  
Relation: iris Instances: 150 Attributes: 4

About  
An instance filter that discretizes a range of numeric attributes in the dataset into nominal attributes. More

Attributes

No.	Name
1	sepallength
2	sepalwidth
3	petallength
4	petalwidth
5	class

attributeIndices first-last  
bins 10  
findNumBins False  
invertSelection False  
makeBinary False  
useEqualFrequency False

Visualize All

Open... Save... OK Cancel

Status OK Log x 0

Detailed description: A histogram visualization showing the distribution of the 'petallength' attribute. The x-axis represents the petallength values from 1 to 6.9. The y-axis represents the frequency of each bin. The distribution is highly skewed to the right, with the vast majority of values falling between 1 and 3. The highest frequency is in the first bin (1-1.5), which is colored blue and has a value of 11. Subsequent bins show a sharp decline in frequency, with values of 2, 0, 0, 3, 4, 12, 10, 2, and 4 respectively. The last bin (6.5-6.9) has a value of 0.

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Undo Save...

Filter

Choose Discretize -B 10 -R first-last weka.gui.GenericObjectEditor Apply

Current relation  
Relation: iris Instances: 150 Attributes: 4

About  
An instance filter that discretizes a range of numeric attributes in the dataset into nominal attributes. More

Attributes

No.	Name
1	sepallength
2	sepalwidth
3	petallength
4	petalwidth
5	class

attributeIndices first-last  
bins 10  
findNumBins False  
invertSelection False  
makeBinary False  
useEqualFrequency False

Visualize All

Open... Save... OK Cancel

11 2 0 0 3 4 12 10 2 4 6.9 3.95 1

Status OK Log x 0

The screenshot shows the Weka Knowledge Explorer interface with the 'Preprocess' tab selected. A 'Discretize' filter is applied to the 'iris' dataset, which has 150 instances and 4 attributes. The 'About' section describes the filter as an instance filter that discretizes numeric attributes into nominal ones. The 'attributeIndices' is set to 'first-last', 'bins' to 10, and 'useEqualFrequency' is set to False. A red arrow points to the 'useEqualFrequency' checkbox. Below the controls is a histogram for the 'petallength' attribute, showing frequencies for bins 1 through 6.9. The status bar at the bottom indicates 'OK'.

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Undo Save...

Filter

Choose Discretize -B 10 -R first-last weka.gui.GenericObjectEditor Apply

Current relation  
Relation: iris Instances: 150 Attributes: 4

About  
An instance filter that discretizes a range of numeric attributes in the dataset into nominal attributes. More

Attributes

No.	Name
1	sepallength
2	sepalwidth
3	petallength
4	petalwidth
5	class

attributeIndices first-last  
bins 10  
findNumBins False  
invertSelection False  
makeBinary False  
useEqualFrequency True

Visualize All

Open... Save... OK Cancel

Status OK Log x 0

Detailed description: A histogram visualization showing the distribution of the 'petallength' attribute. The x-axis represents the petallength values, ranging from 1 to 6.9. The y-axis represents the frequency of each bin. The distribution is highly skewed to the right, with the vast majority of values falling between 1 and 3. The histogram has 10 bins. The first bin (1-1.4) has a frequency of 11. The second bin (1.4-1.7) has a frequency of 2. The third bin (1.7-2.0) has a frequency of 0. The fourth bin (2.0-2.3) has a frequency of 0. The fifth bin (2.3-2.6) has a frequency of 3. The sixth bin (2.6-2.9) has a frequency of 4. The seventh bin (2.9-3.2) has a frequency of 12. The eighth bin (3.2-3.5) has a frequency of 10. The ninth bin (3.5-3.8) has a frequency of 2. The tenth bin (3.8-4.1) has a frequency of 4.

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Undo Save...

Filter

Choose Discretize -B 10 -R first-last weka.gui.GenericObjectEditor Apply

Current relation  
Relation: iris Instances: 150 Attributes: 4

About  
An instance filter that discretizes a range of numeric attributes in the dataset into nominal attributes. More

Attributes

No.	Name
1	sepallength
2	sepalwidth
3	petallength
4	petalwidth
5	class

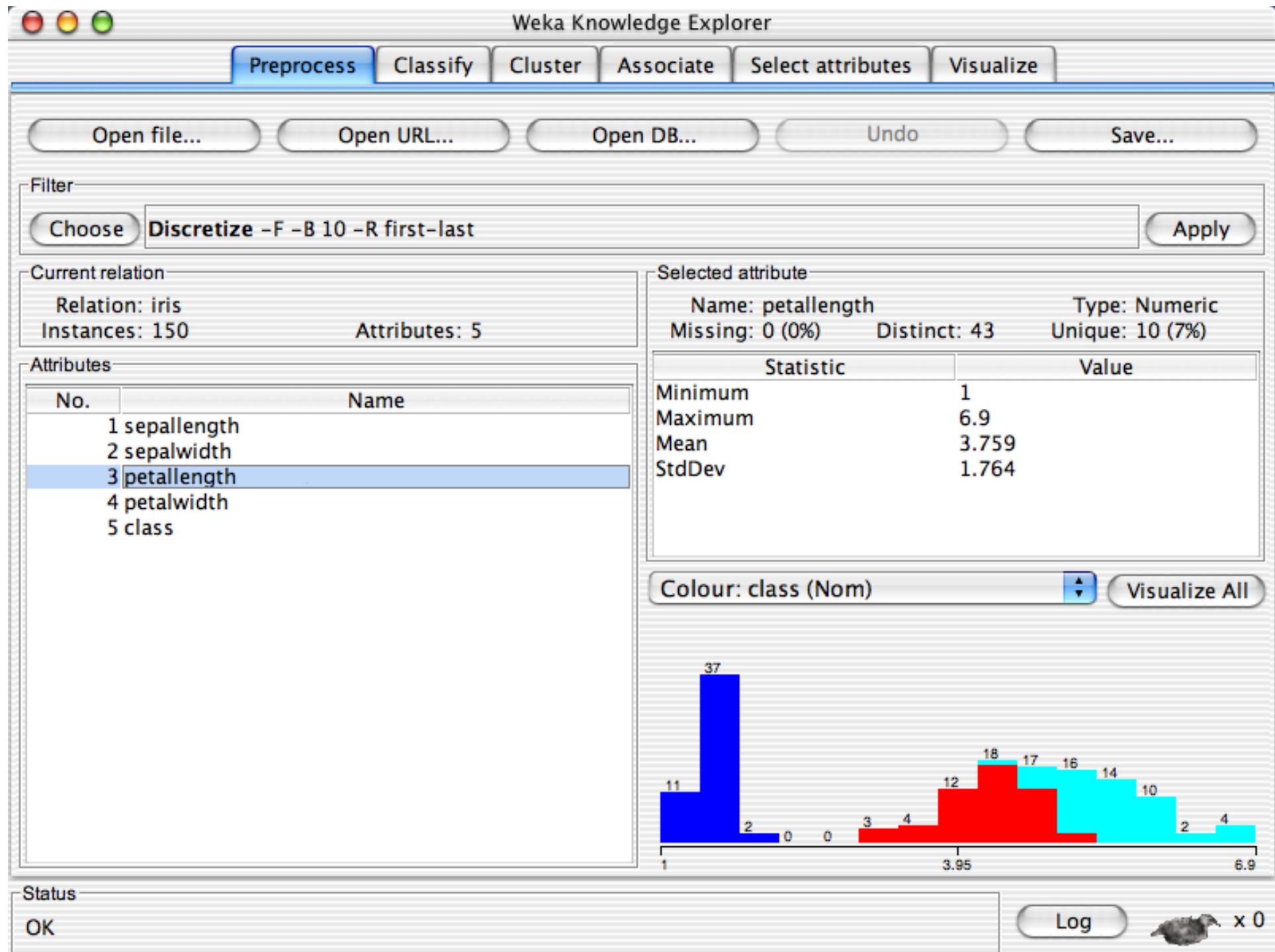
attributeIndices first-last  
bins 10  
findNumBins False  
invertSelection False  
makeBinary False  
useEqualFrequency True

Visualize All

Open... Save... OK Cancel

Status OK Log x 0

A histogram showing the distribution of petallength for the Iris dataset. The x-axis ranges from 1 to 6.9, and the y-axis shows frequency. The distribution is bimodal, with one peak around 1.7 (blue bars) and another around 4.3 (red bars).



Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Undo Save...

Filter

Choose Discretize -F -B 10 -R first-last **Apply**

Current relation  
Relation: iris Instances: 150 Attributes: 5

Attributes

No.	Name
1	sepallength
2	sepalwidth
3	<b>petallength</b>
4	petalwidth
5	class

Selected attribute  
Name: petallength Type: Numeric  
Missing: 0 (0%) Distinct: 43 Unique: 10 (7%)

Statistic	Value
Minimum	1
Maximum	6.9
Mean	3.759
StdDev	1.764

Colour: class (Nom) **Visualize All**

Frequency distribution of petallength:

Bin Range	Frequency
[1.0, 1.5)	11
[1.5, 2.0)	37
[2.0, 2.5)	2
[2.5, 3.0)	0
[3.0, 3.5)	0
[3.5, 4.0)	3
[4.0, 4.5)	4
[4.5, 5.0)	12
[5.0, 5.5)	18
[5.5, 6.0)	17
[6.0, 6.5)	16
[6.5, 7.0)	14
[7.0, 7.5)	10
[7.5, 8.0)	2
[8.0, 8.5)	4

Status OK Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Open file... Open URL... Open DB... Undo Save...

Filter

Choose Discretize -F -B 10 -R first-last Apply

Current relation  
Relation: iris-weka.filters.unsupervised.attribute.Disc...  
Instances: 150 Attributes: 5

Attributes

No.	Name
1	sepallength
2	sepalwidth
3	<b>petallength</b>
4	petalwidth
5	class

Selected attribute  
Name: petallength Type: Nominal  
Missing: 0 (0%) Distinct: 10 Unique: 0 (0%)

Label	Count
'(-inf-1.45]'	23
'(1.45-1.55]'	14
'(1.55-1.8]'	11
'(1.8-3.95]'	13
'(3.95-4.35]'	14
'(4.35-4.65]'	15
'(4.65-5.05]'	18

Colour: class (Nom) Visualize All

Status Log x 0

# Explorer: building “classifiers”

- Classifiers in WEKA are models for predicting nominal or numeric quantities
- Implemented learning schemes include:
  - ◆ Decision trees and lists, instance-based classifiers, support vector machines, multi-layer perceptrons, logistic regression, Bayes’ nets, ...
- “Meta”-classifiers include:
  - ◆ Bagging, boosting, stacking, error-correcting output codes, locally weighted learning, ...

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose ZeroR

Test options

Use training set

Supplied test set Set...

Cross-validation Folds 10

Percentage split % 66

More options...

(Nom) class

Start Stop

Result list (right-click for options)

Status

OK Log x 0

The screenshot shows the Weka Knowledge Explorer interface. The 'Classify' tab is selected. In the 'Classifier' section, 'ZeroR' is chosen. Under 'Test options', 'Cross-validation' is selected with 10 folds. The 'Result list' panel is empty. The 'Status' bar at the bottom shows 'OK' and a log icon with 'x 0'.

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose ZeroR

Test options

Use training set

Supplied test set Set...

Cross-validation Folds 10

Percentage split % 66

More options...

(Nom) class

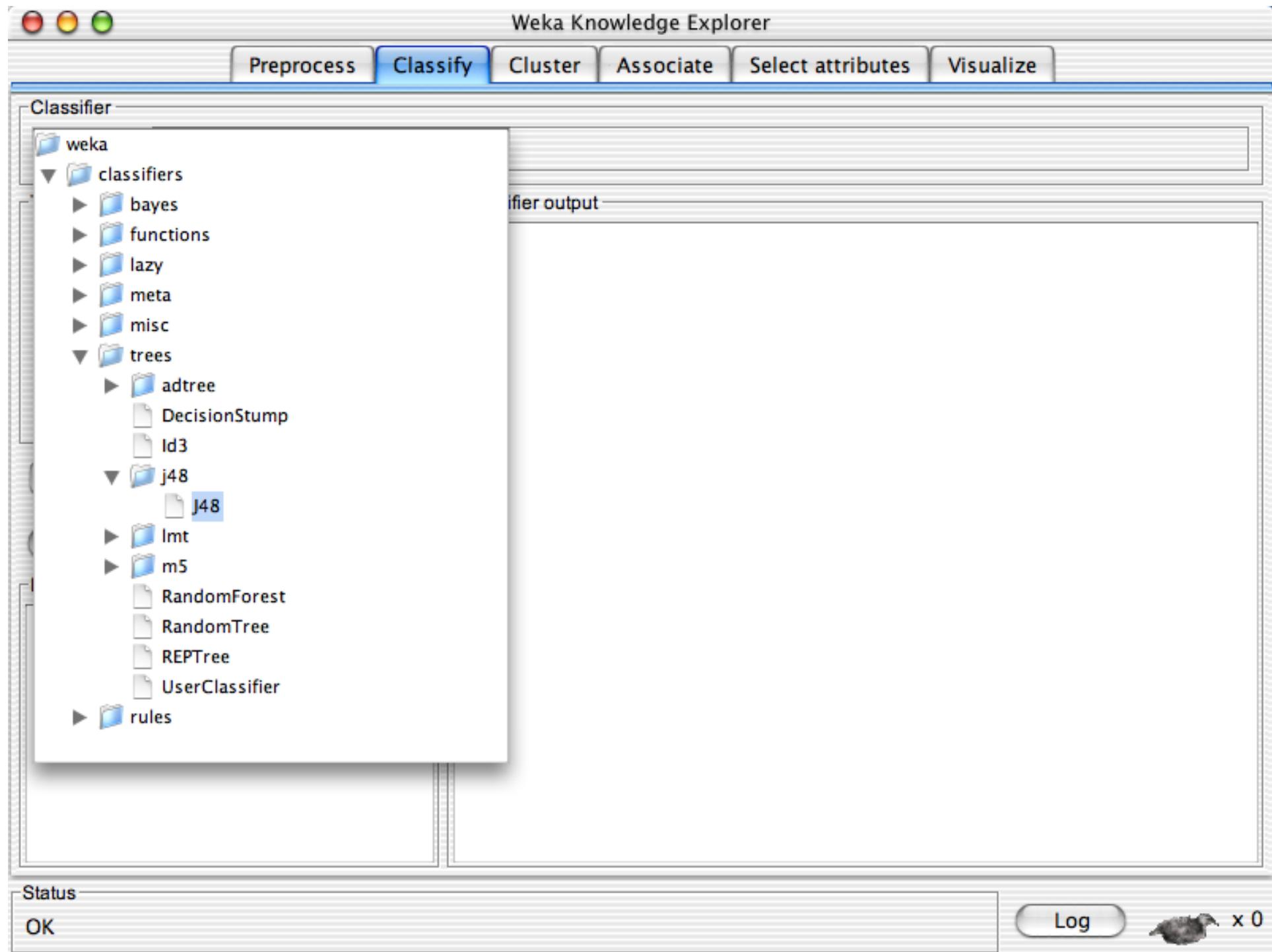
Start Stop

Result list (right-click for options)

Status

OK Log x 0

The image shows the Weka Knowledge Explorer application window. The 'Classify' tab is active. In the 'Classifier' section, 'ZeroR' is selected from the dropdown menu. A red arrow points from the 'Choose' button to the 'ZeroR' text. Below this, the 'Test options' section shows 'Cross-validation' is selected with 10 folds. The 'Result list' section is empty. At the bottom, the status bar displays 'OK'.



Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 - C 0.25 - M 2

Test options

Use training set  
 Supplied test set Set...  
 Cross-validation Folds 10  
 Percentage split % 66  
More options...

(Nom) class

Start Stop

Result list (right-click for options)

Status

OK Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 - C 0.25 - M 2

Test options

Use training set

Supplied test set Set...

Cross-validation Folds 10

Percentage split % 66

More options...

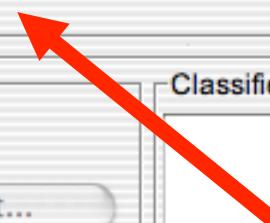
(Nom) class

Start Stop

Result list (right-click for options)

Status

OK Log x 0



Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 - C 0.25 - M 2 weka.gui.GenericObjectEditor weka.classifiers.trees.j48.J48

Test options

Use training set  
 Supplied test set Set...  
 Cross-validation Folds 10  
 Percentage split % 66  
More options...

(Nom) class

Start Stop

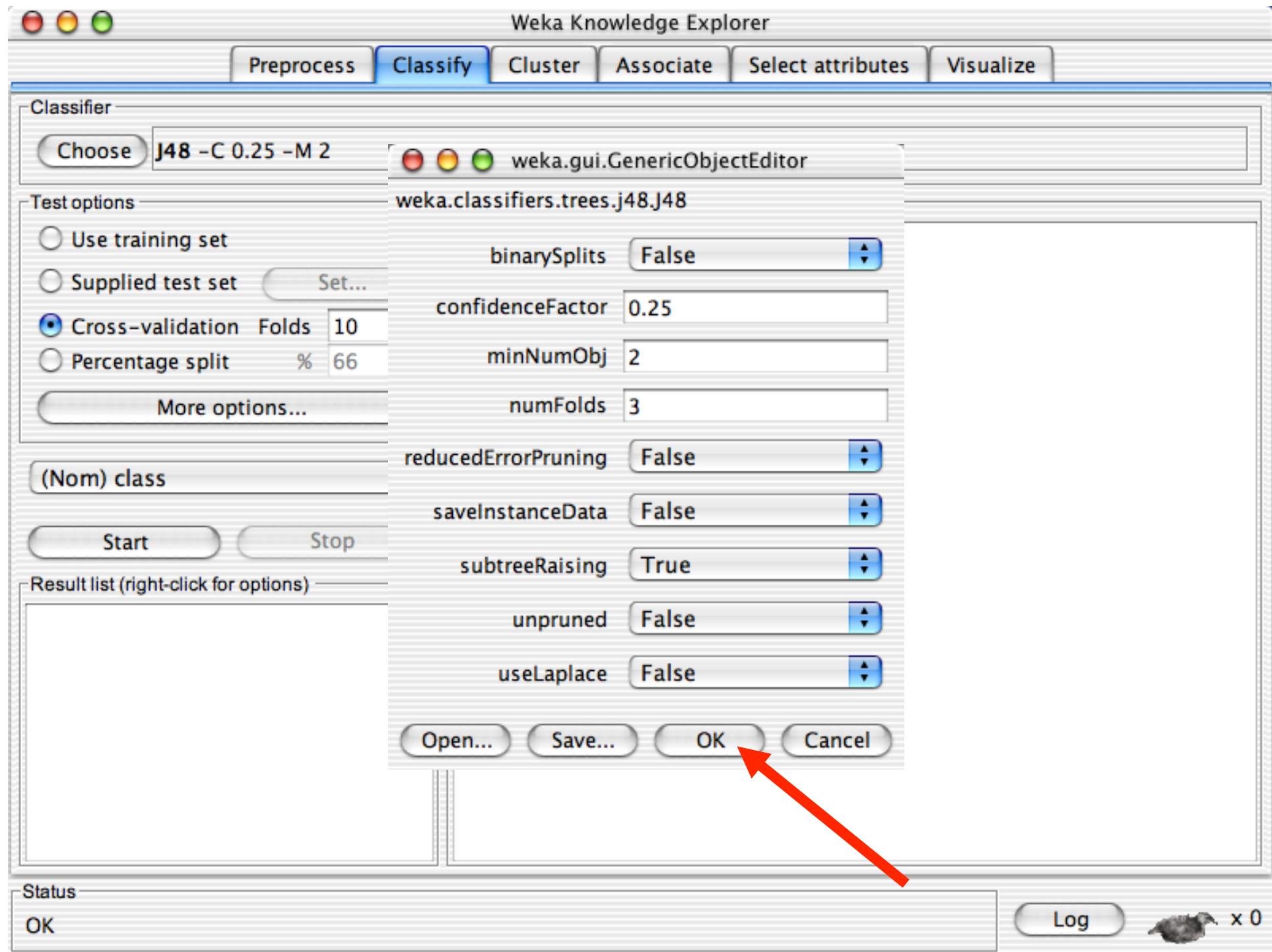
Result list (right-click for options)

binarySplits False  
confidenceFactor 0.25  
minNumObj 2  
numFolds 3  
reducedErrorPruning False  
saveInstanceData False  
subtreeRaising True  
unpruned False  
useLaplace False

Open... Save... OK Cancel

Status

OK Log x 0



Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 - C 0.25 - M 2

Test options

Use training set  
 Supplied test set Set...  
 Cross-validation Folds 10  
 Percentage split % 66  
More options...

(Nom) class

Start Stop

Result list (right-click for options)

Status

OK Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 - C 0.25 - M 2

Test options

Use training set  
 Supplied test set Set...  
 Cross-validation Folds 10  
 Percentage split % 66  
More options...

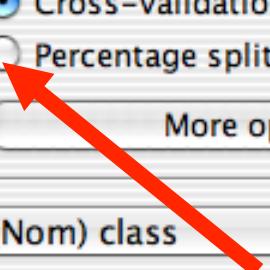
(Nom) class

Start Stop

Result list (right-click for options)

Status

OK Log x 0



Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 - C 0.25 - M 2

Test options

Use training set

Supplied test set Set...

Cross-validation Folds 10

Percentage split % 66

More options...

(Nom) class

Start Stop

Result list (right-click for options)

Status

OK Log x 0

The screenshot shows the Weka Knowledge Explorer interface. The 'Classify' tab is selected. In the 'Classifier' section, 'J48 - C 0.25 - M 2' is chosen. Under 'Test options', 'Percentage split' is selected with 66% chosen. A 'More options...' button is available. A dropdown menu shows '(Nom) class'. Below are 'Start' and 'Stop' buttons. A 'Result list' panel is present but empty. At the bottom, the status is 'OK' and there is a 'Log' button with a count of 0.

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 - C 0.25 - M 2

Test options

Use training set  
 Supplied test set Set...  
 Cross-validation Folds 10  
 Percentage split % 66

More options...

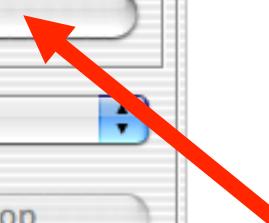
(Nom) class

Start Stop

Result list (right-click for options)

Status

OK Log x 0



Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 - C 0.25 - M 2

Test options

Use training set  
 Supplied test set Set...  
 Cross-validation Folds 10  
 Percentage split % 66  
More options...

(Nom) class

Start Stop

Result list (right-click for options)

Classifier output

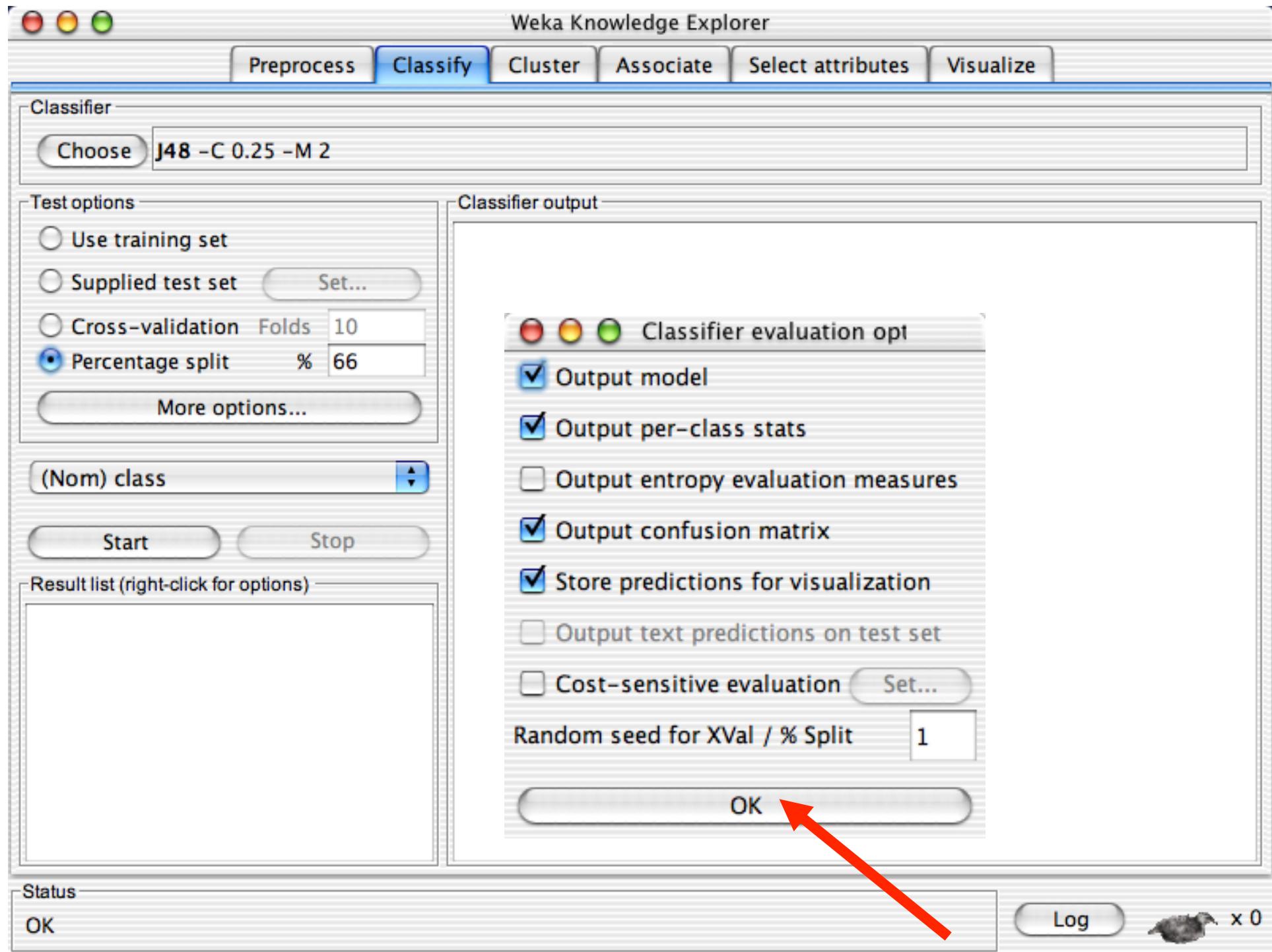
Classifier evaluation opt

Output model  
 Output per-class stats  
 Output entropy evaluation measures  
 Output confusion matrix  
 Store predictions for visualization  
 Output text predictions on test set  
 Cost-sensitive evaluation Set...  
Random seed for XVal / % Split 1

OK

Status

OK Log x 0



Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 - C 0.25 - M 2

Test options

Use training set

Supplied test set Set...

Cross-validation Folds 10

Percentage split % 66

More options...

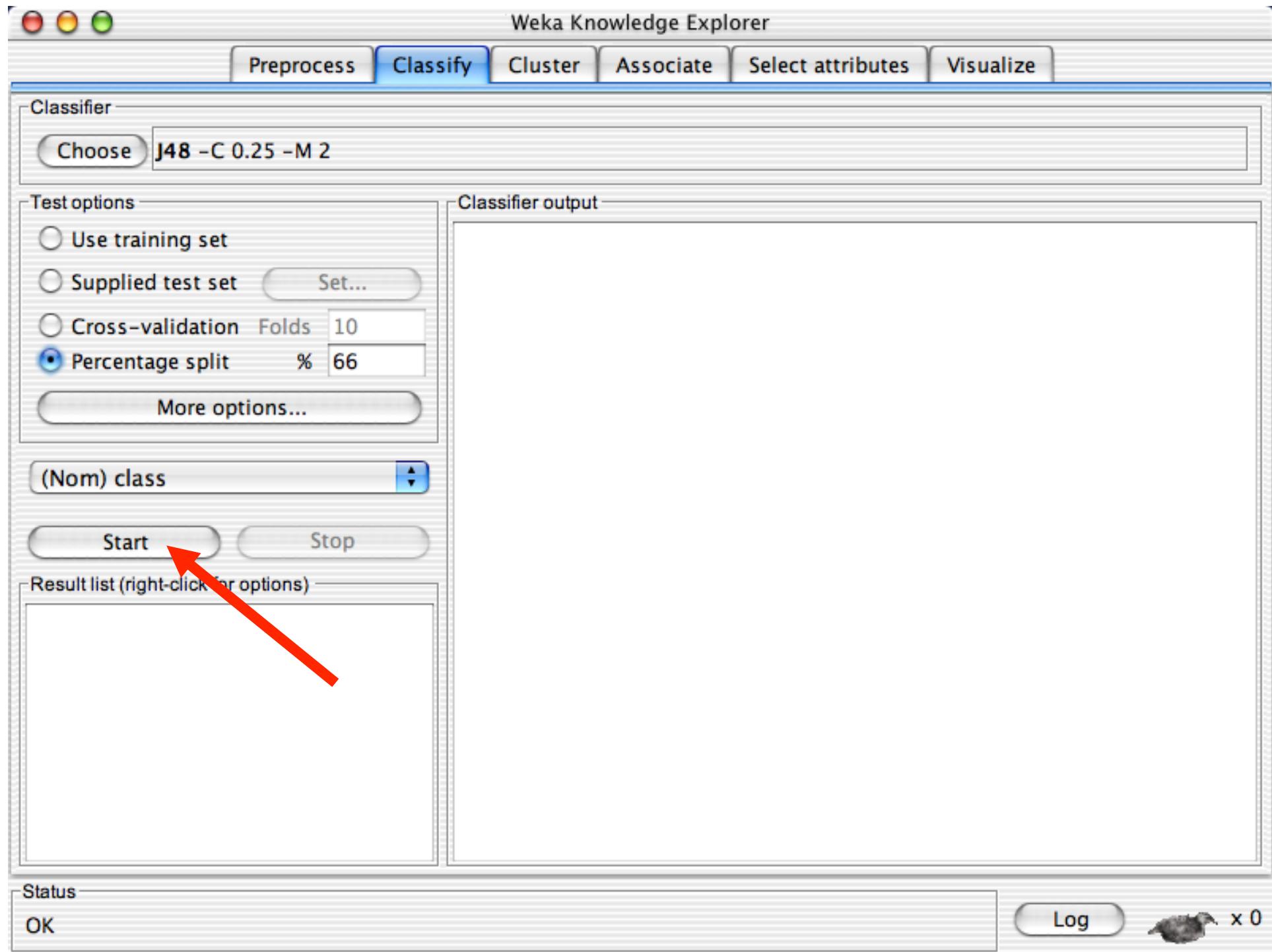
(Nom) class

Start Stop

Result list (right-click for options)

Status

OK Log x 0



Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 -C 0.25 -M 2

Test options

Use training set  
 Supplied test set Set...  
 Cross-validation Folds 10  
 Percentage split % 66  
More options...

(Nom) class

Start Stop

Result list (right-click for options)

11:49:05 – trees.j48.J48

Classifier output

```
==== Run information ====
Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: iris
Instances: 150
Attributes: 5
           sepallength
           sepalwidth
           petallength
           petalwidth
           class
Test mode: split 66% train, remainder test

==== Classifier model (full training set) ====
J48 pruned tree
-----
petalwidth <= 0.6: Iris-setosa (50.0)
petalwidth > 0.6
|   petalwidth <= 1.7
|   |   petallength <= 4.9: Iris-versicolor (48.0/1.0)
|   |   petallength > 4.9
|   |   |   petalwidth <= 1.5: Iris-virginica (3.0)
|   |   |   petalwidth > 1.5: Iris-versicolor (3.0/1.0)
|   petalwidth > 1.7: Iris-virginica (46.0/1.0)

Number of Leaves : 5
```

Status

OK Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 -C 0.25 -M 2

Test options

Use training set  
 Supplied test set Set...  
 Cross-validation Folds 10  
 Percentage split % 66  
More options...

(Nom) class

Start Stop

Result list (right-click for options)

11:49:05 – trees.j48.J48

Classifier output

==== Run information ====  
Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2  
Relation: iris  
Instances: 150  
Attributes: 5  
sepallength  
sepalwidth  
petallength  
petalwidth  
class  
Test mode: split 66% train, remainder test

==== Classifier model (full training set) ====  
J48 pruned tree  
-----  
petalwidth <= 0.6: Iris-setosa (50.0)  
petalwidth > 0.6  
| petalwidth <= 1.7  
| | petallength <= 4.9: Iris-versicolor (48.0/1.0)  
| | petallength > 4.9  
| | | petalwidth <= 1.5: Iris-virginica (3.0)  
| | | petalwidth > 1.5: Iris-versicolor (3.0/1.0)  
| petalwidth > 1.7: Iris-virginica (46.0/1.0)  
Number of Leaves : 5

Status

OK Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 - C 0.25 - M 2

Test options

Use training set  
 Supplied test set Set...  
 Cross-validation Folds 10  
 Percentage split % 66  
More options...

(Nom) class

Start Stop

Result list (right-click for options)

11:49:05 - trees.j48.J48

Classifier output

Time taken to build model: 0.24 seconds

==== Evaluation on test split ====  
==== Summary ====  

Correctly Classified Instances	49	96.0784 %
Incorrectly Classified Instances	2	3.9216 %
Kappa statistic	0.9408	
Mean absolute error	0.0396	
Root mean squared error	0.1579	
Relative absolute error	8.8979 %	
Root relative squared error	33.4091 %	
Total Number of Instances	51	

==== Detailed Accuracy By Class ====  

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
1	0	1	1	1	Iris-setosa
1	0.063	0.905	1	0.95	Iris-versicolor
0.882	0	1	0.882	0.938	Iris-virginica

==== Confusion Matrix ====  

a	b	c	<-- classified as
15	0	0	a = Iris-setosa
0	19	0	b = Iris-versicolor
0	2	15	c = Iris-virginica

Status

OK Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 - C 0.25 - M 2

Test options

Use training set  
 Supplied test set Set...  
 Cross-validation Folds 10  
 Percentage split % 66  
More options...

(Nom) class

Start Stop

Result list (right-click for options)

11:49:05 - trees.j48.J48

Classifier output

Time taken to build model: 0.24 seconds

==== Evaluation on test split ====  
==== Summary ====  

Correctly Classified Instances	49	96.0784 %
Incorrectly Classified Instances	2	3.9216 %
Kappa statistic	0.9408	
Mean absolute error	0.0396	
Root mean squared error	0.1579	
Relative absolute error	8.8979 %	
Root relative squared error	33.4091 %	
Total Number of Instances	51	

==== Detailed Accuracy By Class ====  

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
1	0	1	1	1	Iris-setosa
1	0.063	0.905	1	0.95	Iris-versicolor
0.882	0	1	0.882	0.938	Iris-virginica

==== Confusion Matrix ====  

a	b	c	<-- classified as
15	0	0	a = Iris-setosa
0	19	0	b = Iris-versicolor
0	2	15	c = Iris-virginica

Status

OK

Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 - C 0.25 - M 2

Test options

Use training set  
 Supplied test set Set...  
 Cross-validation Folds 10  
 Percentage split % 66  
More options...

(Nom) class

Start Stop

Result list (right-click for options)  
11:49:05 – trees.j48.J48

View in main window  
View in separate window  
Save result buffer  
Load model  
Save model  
Re-evaluate model on current test set  
Visualize classifier errors  
Visualize tree  
Visualize margin curve  
Visualize threshold curve  
Visualize cost curve

Classifier output

Time taken to build model: 0.24 seconds

==== Evaluation on test split ====  
==== Summary ====  

Correctly Classified Instances	49	96.0784 %
Incorrectly Classified Instances	2	3.9216 %
Kappa statistic	0.9408	
Mean absolute error	0.0396	
Root mean squared error	0.1579	
Relative absolute error	8.8979 %	
Root relative squared error	33.4091 %	
Total Number of Instances	51	

==== Detailed Accuracy By Class ====  

	Recall	F-Measure	Class
Iris-setosa	1	1	Iris-setosa
Iris-versicolor	1	0.95	Iris-versicolor
Iris-virginica	0.882	0.938	Iris-virginica

Status  
OK

Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 - C 0.25 - M 2 Weka Classifier Tree Visualizer: 11:49:05 – trees.j48.J48 (iris)

Test options

- Use training set
- Supplied test set
- Cross-validation
- Percentage split

More options

(Nom) class

Start

Result list (right-click for details)

11:49:05 – trees.j48.J48

Weka Classifier Tree Visualizer: 11:49:05 – trees.j48.J48 (iris)

Tree View

```
graph TD; Root[petalwidth] --<= 0.6--> Setosa[Iris-setosa (50.0)]; Root --> Virginica[petalwidth]; Virginica --<= 1.7--> Versicolor1[Iris-versicolor (48.0/1.0)]; Virginica --> Virginica1[petalwidth]; Virginica1 --<= 1.5--> Virginica2[Iris-virginica (3.0)]; Virginica1 --> Versicolor2[Iris-versicolor (3.0/1.0)];
```

96.0784 %  
3.9216 %

ass  
is-setosa  
is-versicolor  
is-virginica

Status

OK

Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 - C 0.25 - M 2

Test options

Use training set  
 Supplied test set Set...  
 Cross-validation Folds 10  
 Percentage split % 66  
More options...

(Nom) class

Start Stop

Result list (right-click for options)

11:49:05 – trees.j48.J48

View in main window  
View in separate window  
Save result buffer  
Load model  
Save model  
Re-evaluate model on current test set  
Visualize classifier errors  
Visualize tree  
Visualize margin curve  
Visualize threshold curve  
Visualize cost curve

Classifier output

Time taken to build model: 0.24 seconds

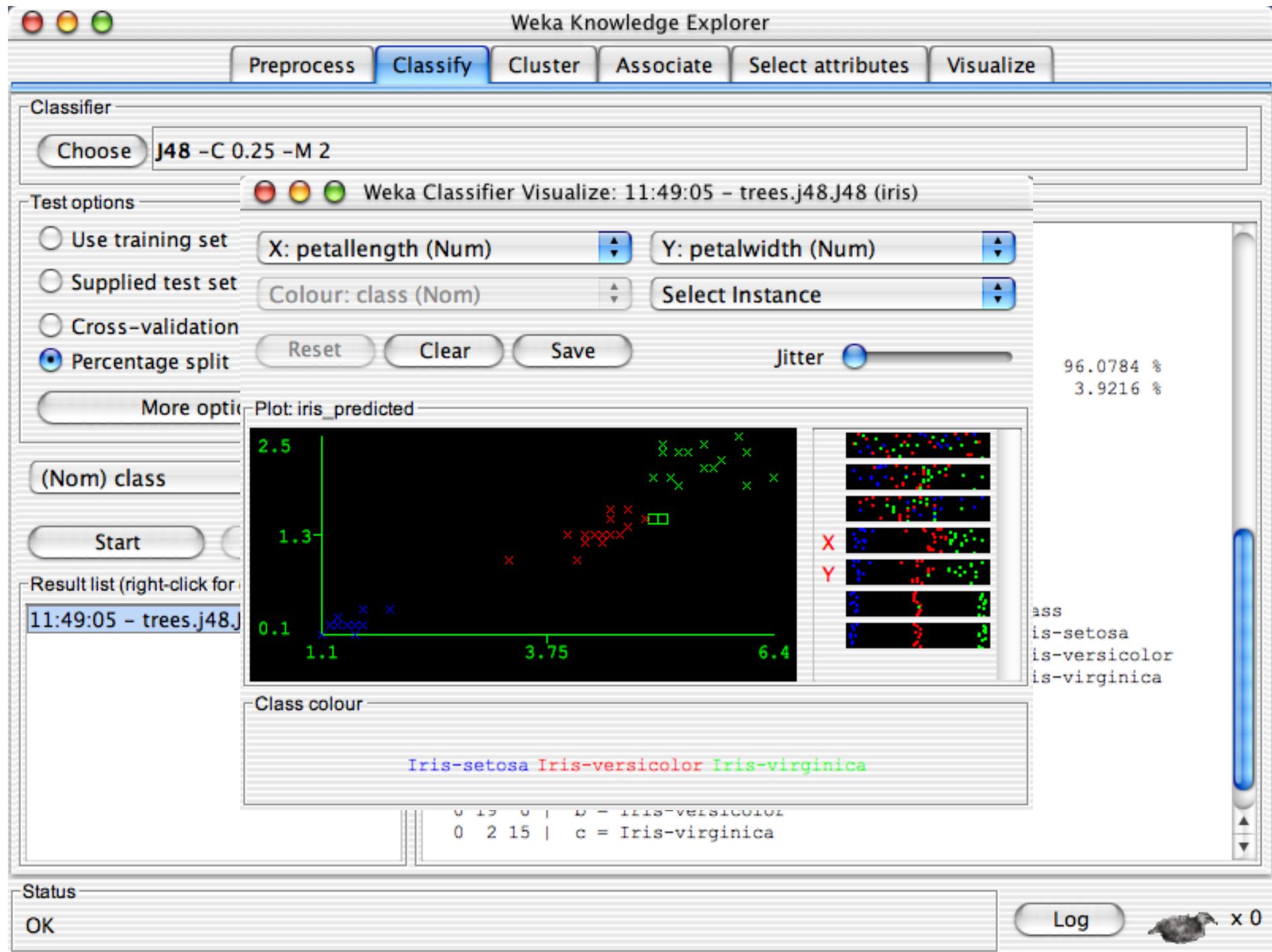
==== Evaluation on test split ====  
==== Summary ====  

Correctly Classified Instances	49	96.0784 %
Incorrectly Classified Instances	2	3.9216 %
Kappa statistic	0.9408	
Mean absolute error	0.0396	
Root mean squared error	0.1579	
Relative absolute error	8.8979 %	
Root relative squared error	33.4091 %	
Total Number of Instances	51	

==== Detailed Accuracy By Class ====  

	Recall	F-Measure	Class
Iris-setosa	1	1	Iris-setosa
Iris-versicolor	1	0.95	Iris-versicolor
Iris-virginica	0.882	0.938	Iris-virginica

Status OK Log x 0



Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 - C 0.25 - M 2

Test options

Use training set  
 Supplied test set Set...  
 Cross-validation Folds 10  
 Percentage split % 66  
More options...

(Nom) class

Start Stop

Result list (right-click for options)

11:49:05 - trees.j48.J48

Classifier output

Time taken to build model: 0.24 seconds

==== Evaluation on test split ====  
==== Summary ====  

Correctly Classified Instances	49	96.0784 %
Incorrectly Classified Instances	2	3.9216 %
Kappa statistic	0.9408	
Mean absolute error	0.0396	
Root mean squared error	0.1579	
Relative absolute error	8.8979 %	
Root relative squared error	33.4091 %	
Total Number of Instances	51	

==== Detailed Accuracy By Class ====  

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
1	0	1	1	1	Iris-setosa
1	0.063	0.905	1	0.95	Iris-versicolor
0.882	0	1	0.882	0.938	Iris-virginica

==== Confusion Matrix ====  

a	b	c	<-- classified as
15	0	0	a = Iris-setosa
0	19	0	b = Iris-versicolor
0	2	15	c = Iris-virginica

Status

OK Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose J48 - C 0.25 - M 2

Test options

Use training set  
 Supplied test set Set...  
 Cross-validation Folds 10  
 Percentage split % 66  
More options...

(Nom) class

Start Stop

Result list (right-click for options)

11:49:05 - trees.j48.J48

Classifier output

Time taken to build model: 0.24 seconds

==== Evaluation on test split ====  
==== Summary ====  

Correctly Classified Instances	49	96.0784 %
Incorrectly Classified Instances	2	3.9216 %
Kappa statistic	0.9408	
Mean absolute error	0.0396	
Root mean squared error	0.1579	
Relative absolute error	8.8979 %	
Root relative squared error	33.4091 %	
Total Number of Instances	51	

==== Detailed Accuracy By Class ====  

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
1	0	1	1	1	Iris-setosa
1	0.063	0.905	1	0.95	Iris-versicolor
0.882	0	1	0.882	0.938	Iris-virginica

==== Confusion Matrix ====  

a	b	c	<-- classified as
15	0	0	a = Iris-setosa
0	19	0	b = Iris-versicolor
0	2	15	c = Iris-virginica

Status

OK

Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

**Classifier**

- weka
- └ classifiers
  - bayes
  - functions
    - LeastMedSq
    - LinearRegression
    - Logistic
  - neural
    - NeuralNetwork
  - pace
  - supportVector
  - SimpleLinearRegression
  - SimpleLogistic
  - VotedPerceptron
  - Winnow
- lazy
- meta
- misc
- trees
- rules

**output**

```
Time taken to build model: 0.24 seconds

Evaluation on test split ===

Model ==>

    correctly Classified Instances      49           96.0784 %
    incorrectly Classified Instances     2            3.9216 %
    statistic                           0.9408
    absolute error                      0.0396
    mean squared error                  0.1579
    mean absolute error                 8.8979 %
    relative squared error              33.4091 %
    Number of Instances                 51

Detailed Accuracy By Class ===

          :   FP Rate    Precision    Recall    F-Measure    Class
          :   0          1           1          1           Iris-setosa
          :  0.063       0.905       1          0.95        Iris-versicolor
          :   0          1           0.882      0.938        Iris-virginica

Confusion Matrix ===

      a   b   c   <-- classified as
15  0  0 |  a = Iris-setosa
 0 19  0 |  b = Iris-versicolor
 0  2 15 |  c = Iris-virginica
```

**Status**

OK

Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose NeuralNetwork -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

Test options

Use training set  
 Supplied test set Set...  
 Cross-validation Folds 10  
 Percentage split % 66  
More options...

(Nom) class

Start Stop

Result list (right-click for options)

11:49:05 - trees.j48.J48

Classifier output

```
==== Evaluation on test split ====
==== Summary ====
Correctly Classified Instances          49           96.0784 %
Incorrectly Classified Instances        2            3.9216 %
Kappa statistic                         0.9408
Mean absolute error                     0.0396
Root mean squared error                 0.1579
Relative absolute error                  8.8979 %
Root relative squared error            33.4091 %
Total Number of Instances               51
```

```
==== Detailed Accuracy By Class ====
TP Rate    FP Rate    Precision    Recall    F-Measure    Class
 1          0          1            1          1          Iris-setosa
 1          0.063       0.905        1          0.95       Iris-versicolor
 0.882      0          1            0.882      0.938       Iris-virginica
```

```
==== Confusion Matrix ====
a  b  c  <-- classified as
15  0  0  |  a = Iris-setosa
 0 19  0  |  b = Iris-versicolor
 0  2 15  |  c = Iris-virginica
```

Status

OK

Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose NeuralNetwork -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

Test options

Use training set  
 Supplied test set Set...  
 Cross-validation Folds 10  
 Percentage split % 66  
More options...

(Nom) class

Start Stop

Result list (right-click for options)

11:49:05 - trees.j48.J48

Classifier output

```
==== Evaluation on test split ====
==== Summary ====
Correctly Classified Instances          49           96.0784 %
Incorrectly Classified Instances        2            3.9216 %
Kappa statistic                         0.9408
Mean absolute error                     0.0396
Root mean squared error                 0.1579
Relative absolute error                  8.8979 %
Root relative squared error            33.4091 %
Total Number of Instances               51

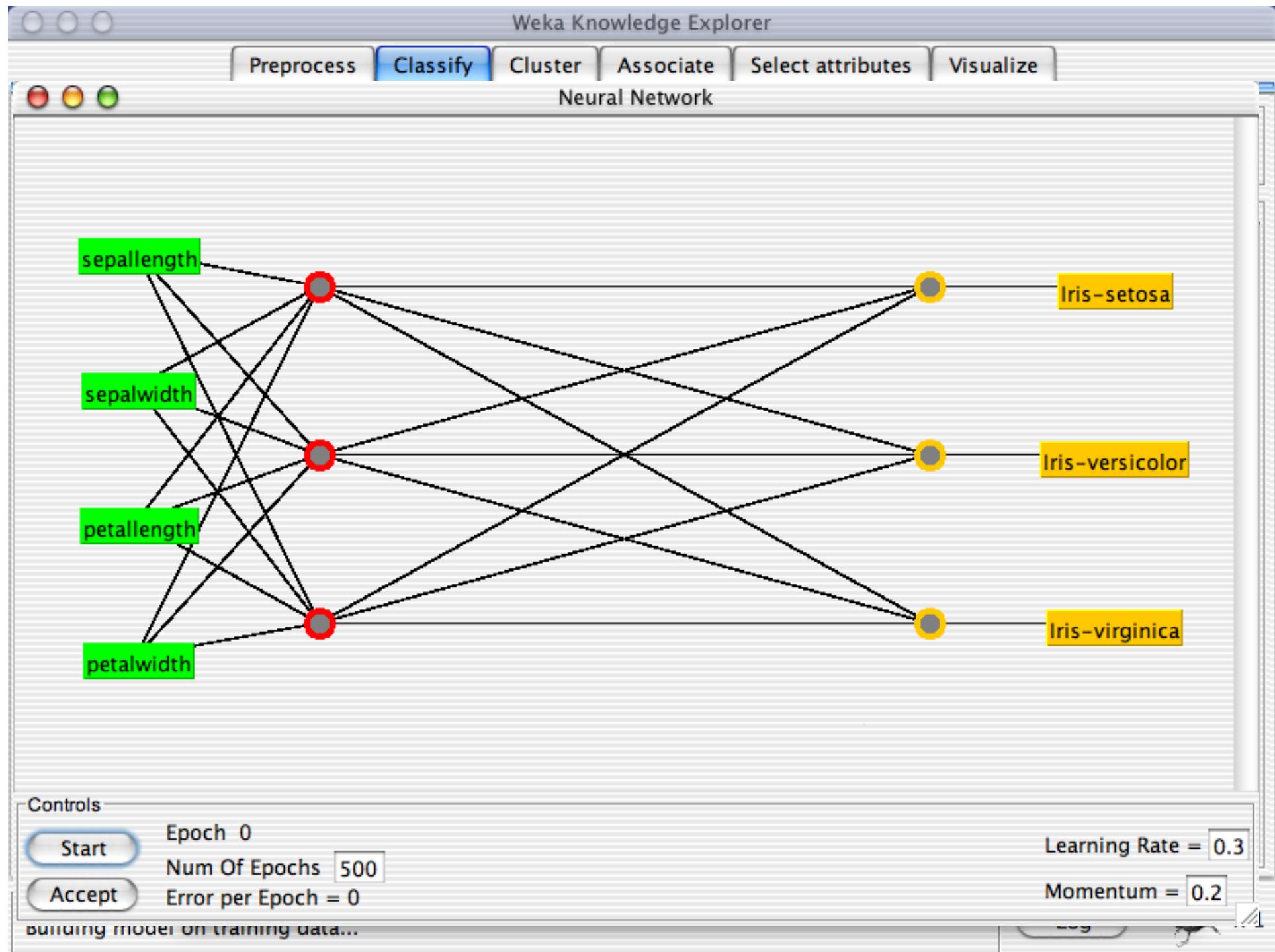
==== Detailed Accuracy By Class ====
TP Rate      FP Rate      Precision      Recall      F-Measure      Class
 1           0           1             1           1
 1           0.063        0.905         1           0.95        Iris-versicolor
 0.882       0           1             0.882       0.938        Iris-virginica

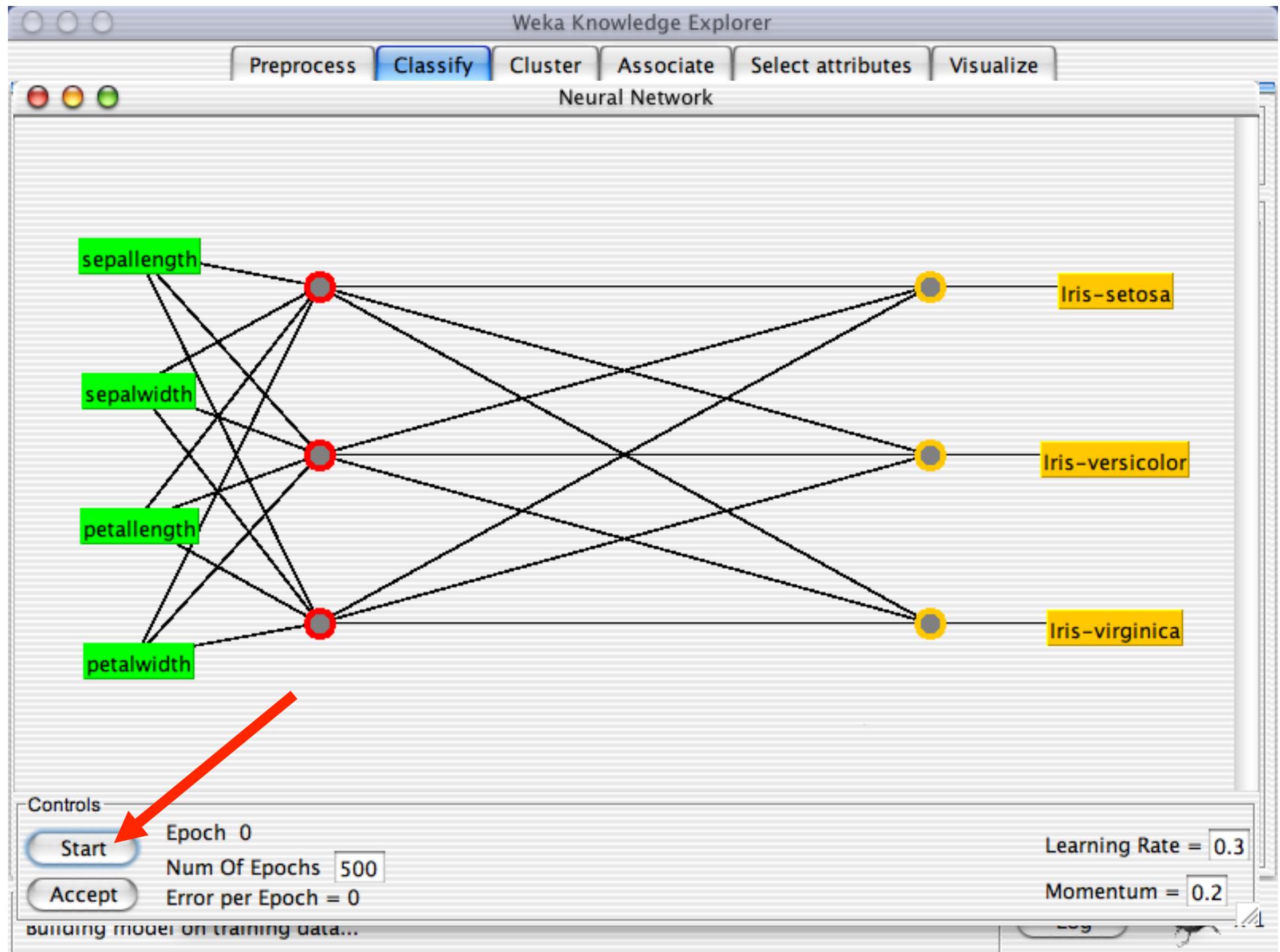
==== Confusion Matrix ====
a  b  c  <-- classified as
15  0  0 |  a = Iris-setosa
 0 19  0 |  b = Iris-versicolor
 0  2 15 |  c = Iris-virginica
```

Status

OK

Log x 0





Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose NeuralNetwork -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a -G -R

Test options

Use training set  
 Supplied test set Set...  
 Cross-validation Folds 10  
 Percentage split % 66  
More options...

(Nom) class

Start Stop

Result list (right-click for options)

11:49:05 – trees.j48.J48  
14:34:28 – functions.neural.NeuralNetwork

Classifier output

==== Evaluation on test split ====  
==== Summary ====  

Correctly Classified Instances	50	98.0392 %
Incorrectly Classified Instances	1	1.9608 %
Kappa statistic	0.9704	
Mean absolute error	0.0239	
Root mean squared error	0.1101	
Relative absolute error	5.3594 %	
Root relative squared error	23.2952 %	
Total Number of Instances	51	

  
==== Detailed Accuracy By Class ====  

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
1	0	1	1	1	Iris-setosa
1	0.031	0.95	1	0.974	Iris-versicolor
0.941	0	1	0.941	0.97	Iris-virginica

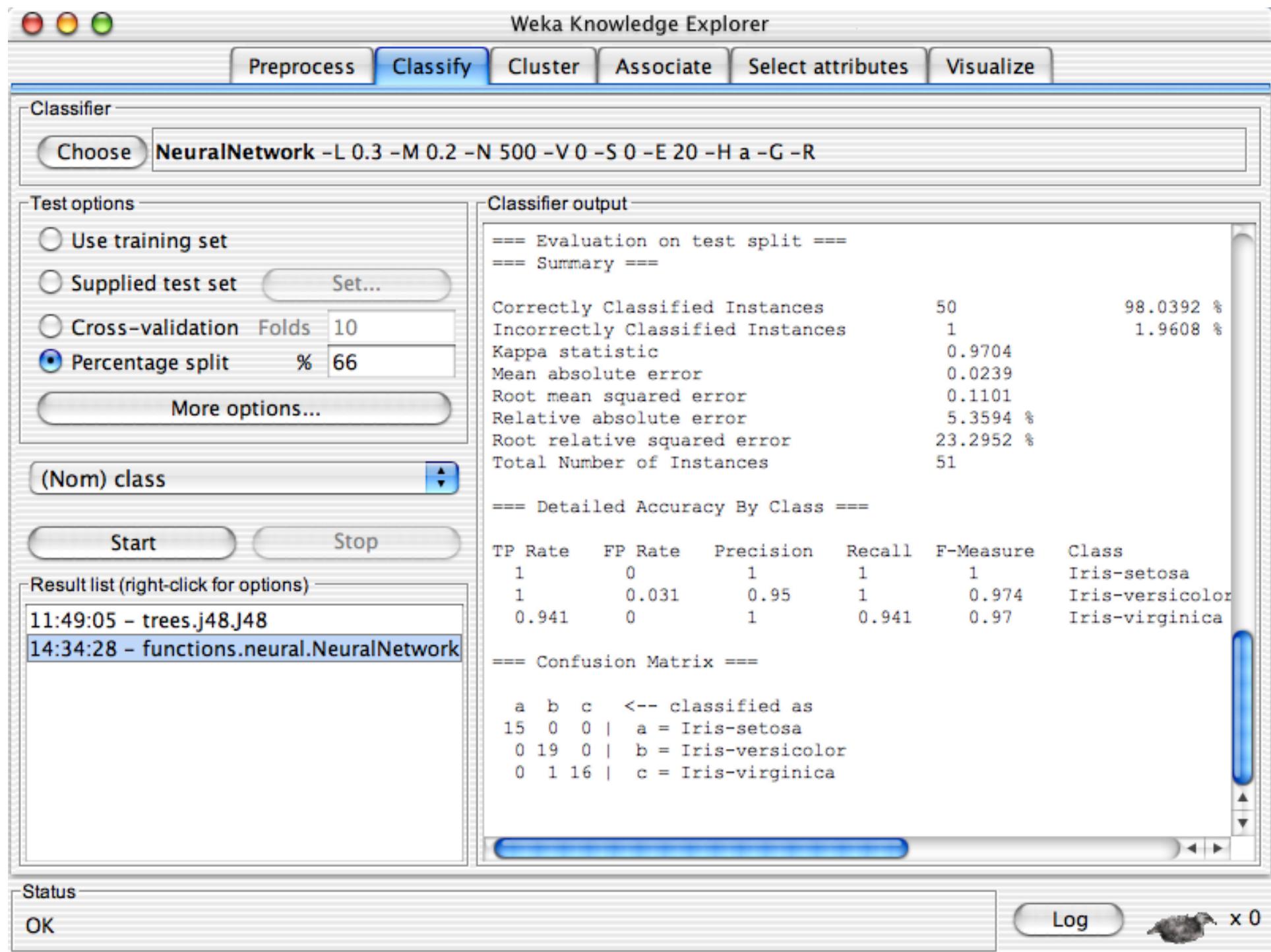
  
==== Confusion Matrix ====  

a	b	c	<-- classified as
15	0	0	a = Iris-setosa
0	19	0	b = Iris-versicolor
0	1	16	c = Iris-virginica

Status

OK

Log x 0



Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose NeuralNetwork -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a -G -R

Test options

Use training set

Supplied test set Set...

Cross-validation Folds 10

Percentage split % 66

More options...

(Nom) class

Start Stop

Result list (right-click for options)

11:49:05 – trees.j48.J48

14:34:28 – functions.neural.NeuralNetwork

Classifier output

==== Evaluation on test split ===  
==== Summary ===

Correctly Classified Instances	50	98.0392 %
Incorrectly Classified Instances	1	1.9608 %
Kappa statistic	0.9704	
Mean absolute error	0.0239	
Root mean squared error	0.1101	
Relative absolute error	5.3594 %	
Root relative squared error	23.2952 %	
Total Number of Instances	51	

==== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
1	0	1	1	1	Iris-setosa
1	0.031	0.95	1	0.974	Iris-versicolor
0.941	0	1	0.941	0.97	Iris-virginica

==== Confusion Matrix ===

a	b	c	<-- classified as
15	0	0	a = Iris-setosa
0	19	0	b = Iris-versicolor
0	1	16	c = Iris-virginica

Status

OK

Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

- weka
- ▼ classifiers
  - bayes
    - AODE
    - BayesNetK2
    - BayesNetB
    - NaiveBayes
    - NaiveBayesMultinomial
    - NaiveBayesSimple
    - NaiveBayesUpdateable
  - functions
  - lazy
  - meta
  - misc
  - trees
  - rules

Classifier output

```
== Evaluation on test split ===
== Summary ===

    correctly Classified Instances           50          98.0392 %
    incorrectly Classified Instances         1           1.9608 %
    kappa statistic                         0.9704
    mean absolute error                   0.0239
    root mean squared error               0.1101
    relative absolute error              5.3594 %
    root relative squared error          23.2952 %
    total Number of Instances            51

== Detailed Accuracy By Class ===

      P  Rate   FP Rate   Precision   Recall   F-Measure   Class
      1       0          1          1          1          1   Iris-setosa
      1       0.031        0.95        1          1          0.974  Iris-versicolor
      0.941   0          1          0.941        0.941        0.97  Iris-virginica

== Confusion Matrix ===

  a  b  c  <-- classified as
15  0  0  |  a = Iris-setosa
  0 19  0  |  b = Iris-versicolor
  0  1 16  |  c = Iris-virginica
```

Status

OK

Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose NaiveBayes

Test options

Use training set  
 Supplied test set Set...  
 Cross-validation Folds 10  
 Percentage split % 66  
More options...

(Nom) class

Start Stop

Result list (right-click for options)

11:49:05 - trees.j48.J48  
14:34:28 - functions.neural.NeuralNetwork

Classifier output

==== Evaluation on test split ====  
==== Summary ====  

Correctly Classified Instances	50	98.0392 %
Incorrectly Classified Instances	1	1.9608 %
Kappa statistic	0.9704	
Mean absolute error	0.0239	
Root mean squared error	0.1101	
Relative absolute error	5.3594 %	
Root relative squared error	23.2952 %	
Total Number of Instances	51	

  
==== Detailed Accuracy By Class ====  

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
1	0	1	1	1	Iris-setosa
1	0.031	0.95	1	0.974	Iris-versicolor
0.941	0	1	0.941	0.97	Iris-virginica

  
==== Confusion Matrix ====  

a	b	c	<- classified as
15	0	0	a = Iris-setosa
0	19	0	b = Iris-versicolor
0	1	16	c = Iris-virginica

Status

OK Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose NaiveBayes

Test options

Use training set  
 Supplied test set Set...  
 Cross-validation Folds 10  
 Percentage split % 66  
More options...

(Nom) class

Start Stop

Result list (right-click for options)

11:49:05 - trees.j48.J48  
14:34:28 - functions.neural.neuralNetwork

Classifier output

==== Evaluation on test split ====  
==== Summary ====  
Correctly Classified Instances 50 98.0392 %  
Incorrectly Classified Instances 1 1.9608 %  
Kappa statistic 0.9704  
Mean absolute error 0.0239  
Root mean squared error 0.1101  
Relative absolute error 5.3594 %  
Root relative squared error 23.2952 %  
Total Number of Instances 51

==== Detailed Accuracy By Class ====  
TP Rate FP Rate Precision Recall F-Measure Class  
1 0 1 1 1 Iris-setosa  
1 0.031 0.95 1 0.974 Iris-versicolor  
0.941 0 1 0.941 0.97 Iris-virginica

==== Confusion Matrix ====  
a b c <-- classified as  
15 0 0 | a = Iris-setosa  
0 19 0 | b = Iris-versicolor  
0 1 16 | c = Iris-virginica

Status OK Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose NaiveBayes

Test options

Use training set  
 Supplied test set Set...  
 Cross-validation Folds 10  
 Percentage split % 66  
More options...

(Nom) class

Start Stop

Result list (right-click for options)

11:49:05 - trees.j48.J48  
14:34:28 - functions.neural.NeuralNetwork  
14:48:05 - bayes.NaiveBayes

Classifier output

==== Evaluation on test split ====  
==== Summary ====  

Correctly Classified Instances	48	94.1176 %
Incorrectly Classified Instances	3	5.8824 %
Kappa statistic	0.9113	
Mean absolute error	0.0447	
Root mean squared error	0.1722	
Relative absolute error	10.0365 %	
Root relative squared error	36.4196 %	
Total Number of Instances	51	

  
==== Detailed Accuracy By Class ====  

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
1	0	1	1	1	Iris-setosa
0.947	0.063	0.9	0.947	0.923	Iris-versicolor
0.882	0.029	0.938	0.882	0.909	Iris-virginica

  
==== Confusion Matrix ====  

a	b	c	<- classified as
15	0	0	a = Iris-setosa
0	18	1	b = Iris-versicolor
0	2	15	c = Iris-virginica

Status

OK

Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose NaiveBayes

Test options

Use training set  
 Supplied test set Set...  
 Cross-validation Folds 10  
 Percentage split % 66  
More options...

(Nom) class

Start Stop

Result list (right-click for options)

11:49:05 - trees.j48.J48  
14:34:28 - functions.neural.NeuralNetwork  
14:48:05 - bayes.NaiveBayes

Classifier output

==== Evaluation on test split ====  
==== Summary ====  
Correctly Classified Instances 48 94.1176 %  
Incorrectly Classified Instances 3 5.8824 %  
Kappa statistic 0.9113  
Mean absolute error 0.0447  
Root mean squared error 0.1722  
Relative absolute error 10.0365 %  
Root relative squared error 36.4196 %  
Total Number of Instances 51

==== Detailed Accuracy By Class ====  
TP Rate FP Rate Precision Recall F-Measure Class  
1 0 1 1 1 Iris-setosa  
0.947 0.063 0.9 0.947 0.923 Iris-versicolor  
0.882 0.029 0.938 0.882 0.909 Iris-virginica

==== Confusion Matrix ====  
a b c <-- classified as  
15 0 0 | a = Iris-setosa  
0 18 1 | b = Iris-versicolor  
0 2 15 | c = Iris-virginica

Status

OK Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose NaiveBayes

Test options

Use training set  
 Supplied test set Set...  
 Cross-validation Folds 10  
 Percentage split % 66  
More options...

(Nom) class

Start

Result list (right-click for context menu):  
11:49:05 - trees.j48.J  
14:34:28 - functions.  
14:48:05 - bayes.Nai

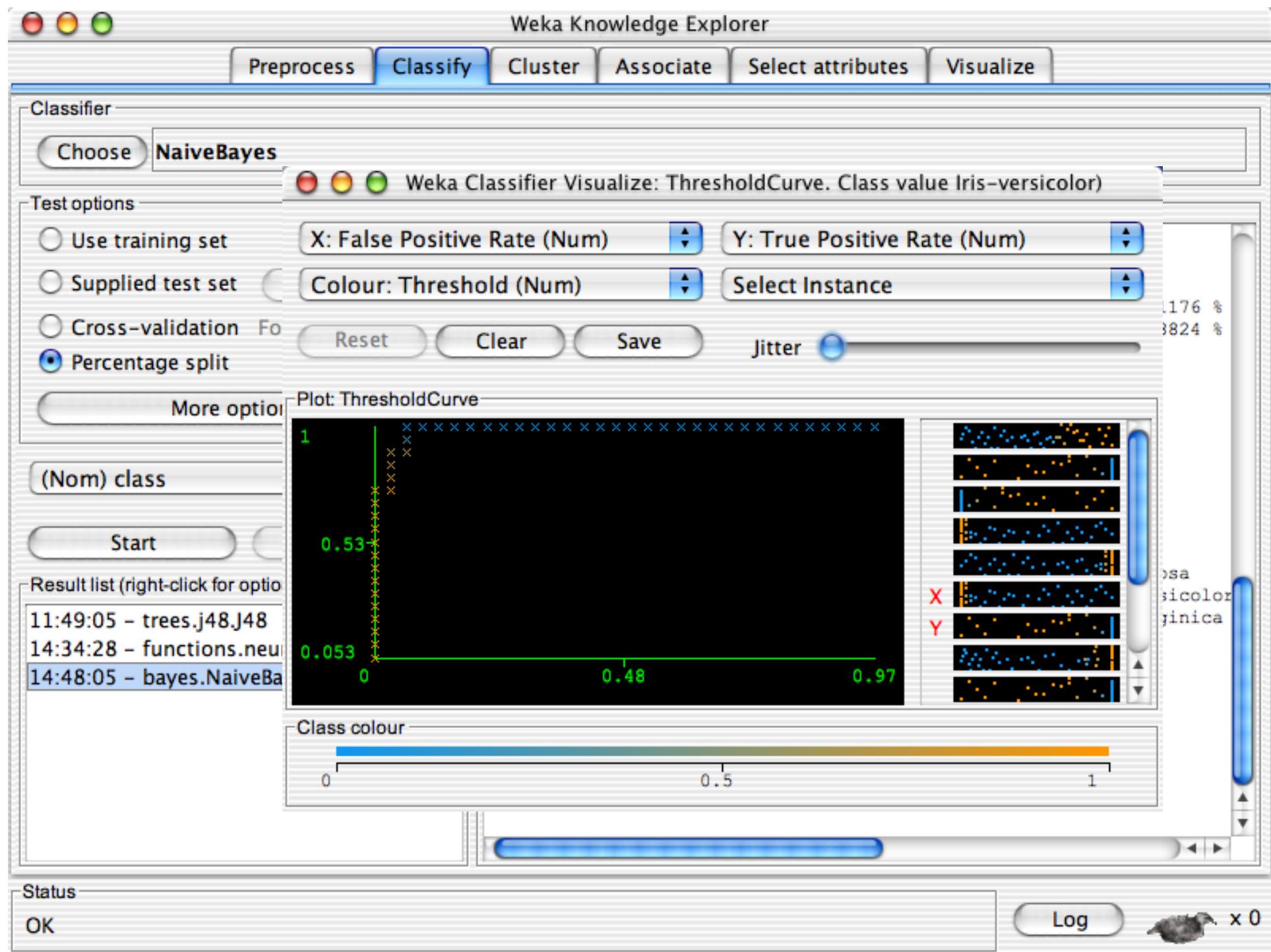
View in main window  
View in separate window  
Save result buffer  
Load model  
Save model  
Re-evaluate model on current test set  
Visualize classifier errors  
Visualize tree  
Visualize margin curve  
Visualize threshold curve  
Visualize cost curve

Classifier output

==== Evaluation on test split ====  
==== Summary ====  
Correctly Classified Instances 48 94.1176 %  
Incorrectly Classified Instances 3 5.8824 %  
Kappa statistic 0.9113  
Mean absolute error 0.0447  
Root mean squared error 0.1722  
Relative absolute error 10.0365 %  
Root relative squared error 36.4196 %  
Total Number of Instances 51

==== Detailed Accuracy By Class ====  
Precision Recall F-Measure Class  
1 1 1 Iris-setosa  
0.9 0.947 0.923 Iris-versicolor  
0.938 0.882 0.909 Iris-virginica

Status OK Log x 0



Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose NaiveBayes

Test options

Use training set  
 Supplied test set Set...  
 Cross-validation Folds 10  
 Percentage split % 66  
More options...

(Nom) class

Start Stop

Result list (right-click for options)

11:49:05 - trees.j48.J48  
14:34:28 - functions.neural.NeuralNetwork  
14:48:05 - bayes.NaiveBayes

Classifier output

==== Evaluation on test split ====  
==== Summary ====  

Correctly Classified Instances	48	94.1176 %
Incorrectly Classified Instances	3	5.8824 %
Kappa statistic	0.9113	
Mean absolute error	0.0447	
Root mean squared error	0.1722	
Relative absolute error	10.0365 %	
Root relative squared error	36.4196 %	
Total Number of Instances	51	

  
==== Detailed Accuracy By Class ====  

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
1	0	1	1	1	Iris-setosa
0.947	0.063	0.9	0.947	0.923	Iris-versicolor
0.882	0.029	0.938	0.882	0.909	Iris-virginica

  
==== Confusion Matrix ====  

a	b	c	<- classified as
15	0	0	a = Iris-setosa
0	18	1	b = Iris-versicolor
0	2	15	c = Iris-virginica

Status

OK

Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose **NaiveBayes**

Test options

Use training set

Supplied test set Set...

Cross-validation Folds 10

Percentage split % 66

More options...

(Nom) class

Start Stop

Result list (right-click for options)

11:49:05 - trees.j48.J48

14:34:28 - functions.neural.NeuralNetwork

14:48:05 - bayes.NaiveBayes

Classifier output

==== Evaluation on test split ====  
==== Summary ====  

Correctly Classified Instances	48	94.1176 %
Incorrectly Classified Instances	3	5.8824 %
Kappa statistic	0.9113	
Mean absolute error	0.0447	
Root mean squared error	0.1722	
Relative absolute error	10.0365 %	
Root relative squared error	36.4196 %	
Total Number of Instances	51	

  
==== Detailed Accuracy By Class ====  

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
1	0	1	1	1	Iris-setosa
0.947	0.063	0.9	0.947	0.923	Iris-versicolor
0.882	0.029	0.938	0.882	0.909	Iris-virginica

  
==== Confusion Matrix ====  

a	b	c	<- classified as
15	0	0	a = Iris-setosa
0	18	1	b = Iris-versicolor
0	2	15	c = Iris-virginica

Status

OK

Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

- weka
- └ classifiers
  - bayes
  - functions
  - lazy
  - meta
  - misc
  - trees
    - adtree
    - DecisionStump
    - Id3
    - j48
    - lmt
    - m5
    - RandomForest
    - RandomTree
    - REPTree
    - UserClassifier
  - rules

Classifier output

```
== Evaluation on test split ===  
== Summary ===  
    correctly Classified Instances      48          94.1176 %  
    incorrectly Classified Instances     3           5.8824 %  
    kappa statistic                   0.9113  
    mean absolute error               0.0447  
    root mean squared error           0.1722  
    relative absolute error           10.0365 %  
    root relative squared error      36.4196 %  
    total Number of Instances        51  
  
== Detailed Accuracy By Class ===  


| P     | Rate  | FP Rate | Precision | Recall | F-Measure | Class           |
|-------|-------|---------|-----------|--------|-----------|-----------------|
| 1     | 0     | 0       | 1         | 1      | 1         | Iris-setosa     |
| 0.947 | 0.063 | 0.063   | 0.9       | 0.947  | 0.923     | Iris-versicolor |
| 0.882 | 0.029 | 0.029   | 0.938     | 0.882  | 0.909     | Iris-virginica  |

  
== Confusion Matrix ===  


| a  | b  | c  | <-- classified as   |
|----|----|----|---------------------|
| 15 | 0  | 0  | a = Iris-setosa     |
| 0  | 18 | 1  | b = Iris-versicolor |
| 0  | 2  | 15 | c = Iris-virginica  |



Status



OK



Log x 0


```

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose UserClassifier

Test options

Use training set  
 Supplied test set Set...  
 Cross-validation Folds 10  
 Percentage split % 66  
More options...

(Nom) class

Start Stop

Result list (right-click for options)

11:49:05 - trees.J48.J48  
14:34:28 - functions.neural.NeuralNetwork  
14:48:05 - bayes.NaiveBayes

Classifier output

==== Evaluation on test split ====  
==== Summary ====  

Correctly Classified Instances	48	94.1176 %
Incorrectly Classified Instances	3	5.8824 %
Kappa statistic	0.9113	
Mean absolute error	0.0447	
Root mean squared error	0.1722	
Relative absolute error	10.0365 %	
Root relative squared error	36.4196 %	
Total Number of Instances	51	

  
==== Detailed Accuracy By Class ====  

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
1	0	1	1	1	Iris-setosa
0.947	0.063	0.9	0.947	0.923	Iris-versicolor
0.882	0.029	0.938	0.882	0.909	Iris-virginica

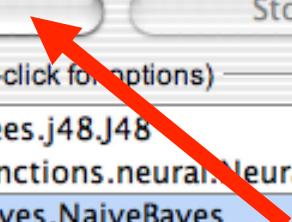
  
==== Confusion Matrix ====  

a	b	c	<- classified as
15	0	0	a = Iris-setosa
0	18	1	b = Iris-versicolor
0	2	15	c = Iris-virginica

Status

OK

Log x 0



Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose UserClassifier

Test options

Use training set

Supplied test set

Cross-validation

Percentage split

More options

(Nom) class

Start

Result list (right-click for options)

11:49:05 – trees.J48.J48  
14:34:28 – functions.neural  
14:48:05 – bayes.NaiveBayes  
15:26:57 – trees.UserClass

Tree View

Tree Visualizer Data Visualizer

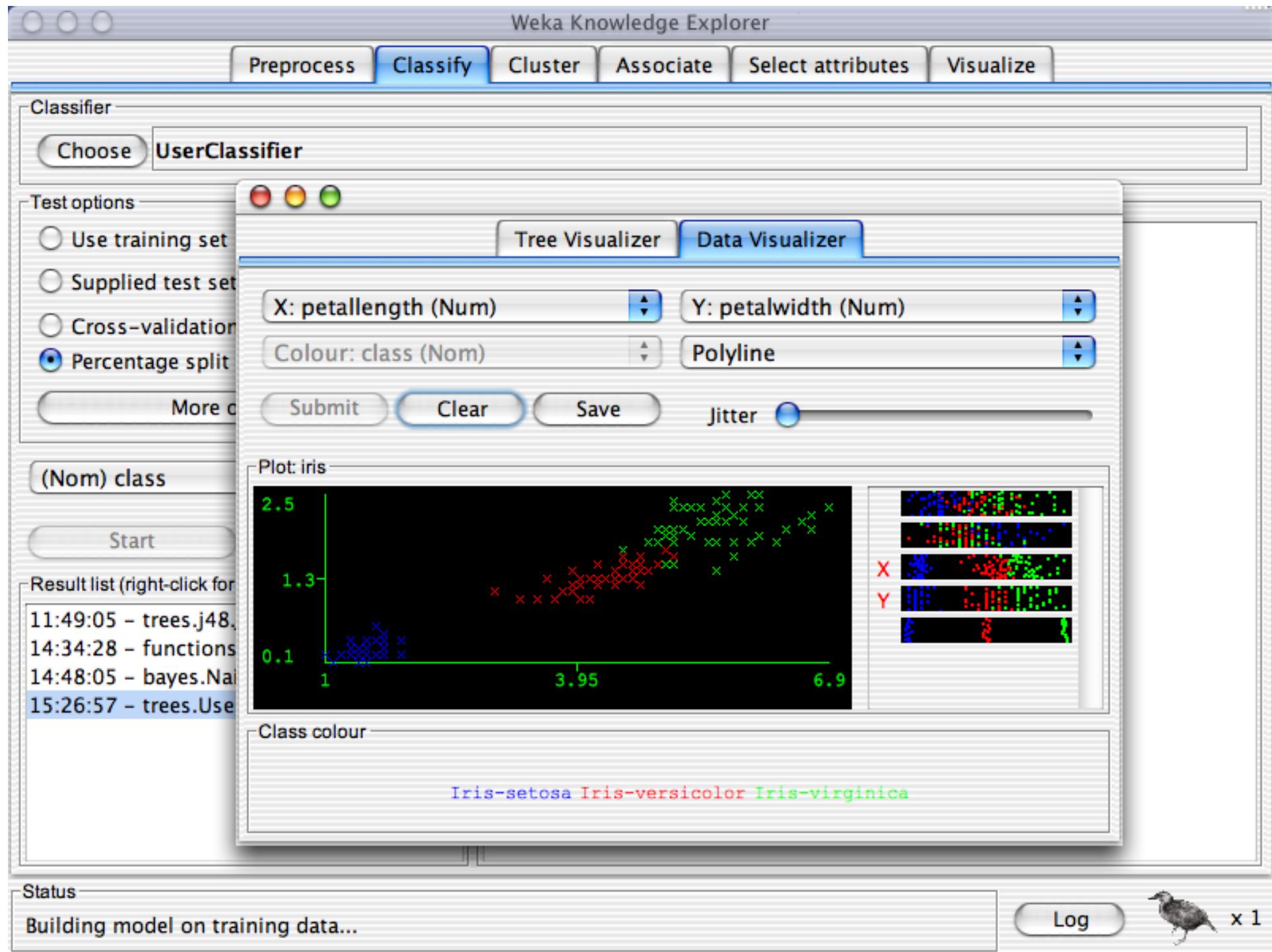
[Iris-setosa, 50.0]  
[Iris-versicolor, 50.0]  
[Iris-virginica, 50.0]

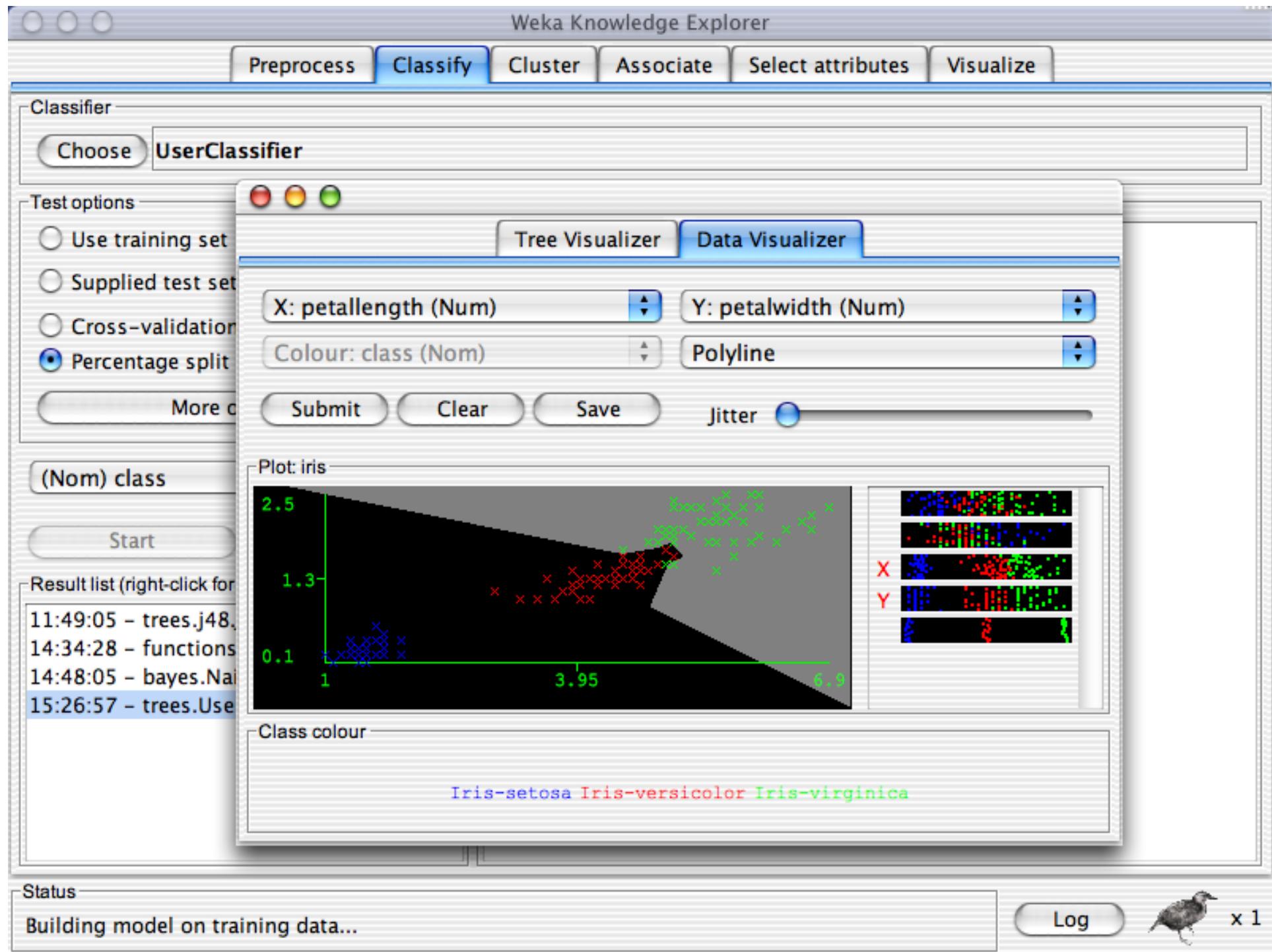
Status

Building model on training data...

Log

x 1





Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose UserClassifier

Test options

Use training set  
 Supplied test set  
 Cross-validation  
 Percentage split  
More options

(Nom) class

Start

Result list (right-click for context menu)

11:49:05 - trees.j48.J48  
14:34:28 - functions.  
14:48:05 - bayes.NaiveBayes  
15:26:57 - trees.User

Status

Building model on training data...

Tree Visualizer Data Visualizer

Tree View

Split on petallength AND petalwidth

```
graph TD; Root[Split on petallength AND petalwidth] -- True --> L1L[Iris-versicolor, 1.0]; Root -- False --> L1R[Iris-setosa, 50.0]; L1L --> L2L1[Iris-virginica, 48.0]; L1R --> L2R1[Iris-versicolor, 49.0]; L1R --> L2R2[Iris-virginica, 2.0]
```

Log x 1

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose UserClassifier

Test options

Use training set  
 Supplied test set Set...  
 Cross-validation Folds 10  
 Percentage split % 66  
More options...

(Nom) class

Start Stop

Result list (right-click for options)

11:49:05 - trees.j48.J48  
14:34:28 - functions.neural.NeuralNetwork  
14:48:05 - bayes.NaiveBayes  
15:44:32 - trees.UserClassifier

Classifier output

==== Evaluation on test split ====  
==== Summary ====  

Correctly Classified Instances	49	96.0784 %
Incorrectly Classified Instances	2	3.9216 %
Kappa statistic	0.9408	
Mean absolute error	0.0319	
Root mean squared error	0.1622	
Relative absolute error	7.1634 %	
Root relative squared error	34.312 %	
Total Number of Instances	51	

  
==== Detailed Accuracy By Class ====  

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
1	0	1	1	1	Iris-setosa
1	0.063	0.905	1	0.95	Iris-versicolor
0.882	0	1	0.882	0.938	Iris-virginica

  
==== Confusion Matrix ====  

a	b	c	<- classified as
15	0	0	a = Iris-setosa
0	19	0	b = Iris-versicolor
0	2	15	c = Iris-virginica

Status

OK

Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

**Classifier**

Choose UserClassifier

**Test options**

Use training set  
 Supplied test set Set...  
 Cross-validation Folds 10  
 Percentage split % 66

**(Nom) class**

Start Stop

**Result list (right-click for options)**

11:49:05 – trees.j48.J48  
14:34:28 – functions.neural.NeuralNetwork  
14:48:05 – bayes.NaiveBayes  
15:44:32 – trees.UserClassifier

**Classifier output**

==== Evaluation on test split ====  
==== Summary ====  

	Value	%
Correctly Classified Instances	49	96.0784 %
Incorrectly Classified Instances	2	3.9216 %
Kappa statistic	0.9408	
Mean absolute error	0.0319	
Root mean squared error	0.1622	
Relative absolute error	7.1634 %	
Root relative squared error	34.312 %	
Total Number of Instances	51	

  
==== Detailed Accuracy By Class ====  

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
1	0	1	1	1	Iris-setosa
1	0.063	0.905	1	0.95	Iris-versicolor
0.882	0	1	0.882	0.938	Iris-virginica

  
==== Confusion Matrix ====  

	a	b	c	<-- classified as
15	0	0		a = Iris-setosa
0	19	0		b = Iris-versicolor
0	2	15		c = Iris-virginica

Status OK Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

**Classifier**

Choose UserClassifier

**Test options**

Use training set  
 Supplied test set Set...  
 Cross-validation Folds 10  
 Percentage split % 66

(Num) sepallength  
(Num) sepalwidth  
**✓ (Num) petallength**  
(Num) petalwidth  
(Nom) class

**Result list (right-click for options)**

11:49:05 – trees.j48.J48  
14:34:28 – functions.neural.NeuralNetwork  
14:48:05 – bayes.NaiveBayes  
**15:44:32 – trees.UserClassifier**

**Classifier output**

```
==== Evaluation on test split ====
==== Summary ====
Correctly Classified Instances           49          96.0784 %
Incorrectly Classified Instances        2           3.9216 %
Kappa statistic                         0.9408
Mean absolute error                     0.0319
Root mean squared error                 0.1622
Relative absolute error                  7.1634 %
Root relative squared error            34.312 %
Total Number of Instances               51

==== Detailed Accuracy By Class ====
TP Rate    FP Rate    Precision    Recall    F-Measure    Class
1          0          1            1          1
1          0.063      0.905       1          0.95
0.882     0          1            0.882     0.938
Iris-setosa
Iris-versicolor
Iris-virginica

==== Confusion Matrix ====
a  b  c  <-- classified as
15  0  0  |  a = Iris-setosa
 0  19  0  |  b = Iris-versicolor
 0  2  15 |  c = Iris-virginica
```

Status

OK

Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose UserClassifier

Test options

Use training set

Supplied test set Set...

Cross-validation Folds 10

Percentage split % 66

More options...

(Nom) class

Start Stop

Result list (right-click for options)

11:49:05 – trees.j48.J48

14:34:28 – functions.neural.NeuralNetwork

14:48:05 – bayes.NaiveBayes

15:44:32 – trees.UserClassifier

Classifier output

==== Evaluation on test split ===  
==== Summary ===

Correctly Classified Instances	49	96.0784 %
Incorrectly Classified Instances	2	3.9216 %
Kappa statistic	0.9408	
Mean absolute error	0.0319	
Root mean squared error	0.1622	
Relative absolute error	7.1634 %	
Root relative squared error	34.312 %	
Total Number of Instances	51	

==== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	Class
1	0	1	1	1	Iris-setosa
1	0.063	0.905	1	0.95	Iris-versicolor
0.882	0	1	0.882	0.938	Iris-virginica

==== Confusion Matrix ===

a	b	c	<-- classified as
15	0	0	a = Iris-setosa
0	19	0	b = Iris-versicolor
0	2	15	c = Iris-virginica

Status

OK

Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

weka  
└ classifiers  
 ├── bayes  
 ├── functions  
 ├── lazy  
 ├── meta  
 ├── misc  
 └── trees  
 ├── adtree  
 ├── DecisionStump  
 ├── Id3  
 ├── j48  
 ├── lmt  
 └── m5  
 ├── MSP  
 ├── RandomForest  
 ├── RandomTree  
 ├── REPTree  
 └── UserClassifier  
 └── rules

Classifier output

```
== Evaluation on test split ===  
== Summary ===  
    correctly Classified Instances      49          96.0784 %  
    incorrectly Classified Instances   2           3.9216 %  
    kappa statistic                  0.9408  
    mean absolute error              0.0319  
    root mean squared error          0.1622  
    relative absolute error          7.1634 %  
    root relative squared error     34.312  %  
    total Number of Instances       51  
  
== Detailed Accuracy By Class ===  


|                 | P Rate | FP Rate | Precision | Recall | F-Measure | Class           |
|-----------------|--------|---------|-----------|--------|-----------|-----------------|
| Iris-setosa     | 1      | 0       | 1         | 1      | 1         | Iris-setosa     |
| Iris-versicolor | 1      | 0.063   | 0.905     | 1      | 0.95      | Iris-versicolor |
| Iris-virginica  | 0.882  | 0       | 1         | 0.882  | 0.938     | Iris-virginica  |

  
== Confusion Matrix ===  


|                 | a  | b  | c  | <-- classified as   |
|-----------------|----|----|----|---------------------|
| Iris-setosa     | 15 | 0  | 0  | a = Iris-setosa     |
| Iris-versicolor | 0  | 19 | 0  | b = Iris-versicolor |
| Iris-virginica  | 0  | 2  | 15 | c = Iris-virginica  |



Status



OK



Log x 0


```

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose M5P - M 4.0

Test options

Use training set  
 Supplied test set Set...  
 Cross-validation Folds 10  
 Percentage split % 66  
More options...

(Num) petallength

Start Stop

Result list (right-click for options)

11:49:05 – trees.j48.J48  
14:34:28 – functions.neural.NeuralNetwork  
14:48:05 – bayes.NaiveBayes  
15:44:32 – trees.UserClassifier  
15:49:03 – trees.m5.M5P

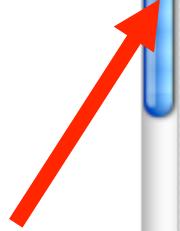
Classifier output

==== Run information ====  
Scheme: weka.classifiers.trees.m5.M5P -M 4.0  
Relation: iris  
Instances: 150  
Attributes: 5  
sepallength  
sepalwidth  
petallength  
petalwidth  
class  
Test mode: split 66% train, remainder test

==== Classifier model (full training set) ====  
M5 pruned model tree:  
(using smoothed predictions)  
petalwidth <= 0.8 : LM1 (50/10.469%)  
petalwidth > 0.8 :  
| class=Iris-virginica <= 0.5 : LM2 (50/14.325%)  
| class=Iris-virginica > 0.5 : LM3 (50/17.598%)  
LM num: 1  
Linear Regression Model  
petallength =

Status

OK Log x 0



Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose M5P - M 4.0

Test options

Use training set  
 Supplied test set Set...  
 Cross-validation Folds 10  
 Percentage split % 66  
More options...

(Num) petallength

Start Stop

Result list (right-click for options)

11:49:05 – trees.j48.J48  
14:34:28 – functions.neural.NeuralNetwork  
14:48:05 – bayes.NaiveBayes  
15:44:32 – trees.UserClassifier  
15:49:03 – trees.m5.M5P

Classifier output

```
| CLASS=IRIS-virginica > 0.5 : LM3 (50/17.598%)  
LM num: 1  
Linear Regression Model  
petallength =  
0.4957 * petalwidth +  
1.343  
LM num: 2  
Linear Regression Model  
petallength =  
0.4208 * sepallength +  
1.2692 * petalwidth +  
0.0795  
LM num: 3  
Linear Regression Model  
petallength =  
0.7501 * sepallength +  
0.6105  
Number of Rules : 3
```

Status

OK

Log x 0



Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose M5P - M 4.0

Test options

Use training set  
 Supplied test set Set...  
 Cross-validation Folds 10  
 Percentage split % 66  
More options...

(Num) petallength

Start Stop

Result list (right-click for options)

11:49:05 - trees.j48.J48  
14:34:28 - functions.neural.NeuralNetwork  
14:48:05 - bayes.NaiveBayes  
15:44:32 - trees.UserClassifier  
15:49:03 - trees.m5.M5P

Classifier output

```
0.4208 * sepallength +
1.2692 * petalwidth +
0.0795

LM num: 3
Linear Regression Model

petallength =

0.7501 * sepallength +
0.6105

Number of Rules : 3

Time taken to build model: 1.31 seconds

==== Evaluation on test split ====
==== Summary ===

Correlation coefficient 0.9889
Mean absolute error 0.1861
Root mean squared error 0.255
Relative absolute error 11.9578 %
Root relative squared error 14.9153 %
Total Number of Instances 51
```

Status

OK Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose M5P - M 4.0

Test options

Use training set  
 Supplied test set Set...  
 Cross-validation Folds 10  
 Percentage split % 66  
More options...

(Num) petallength

Start Stop

Result list (right-click for options)

11:49:05 - trees.j48.J48  
14:34:28 - functions.neural.NeuralNetwork  
14:48:05 - bayes.NaiveBayes  
15:44:32 - trees.UserClassifier  
15:49:03 - trees.m5.M5P

Classifier output

```
0.4208 * sepallength +
1.2692 * petalwidth +
0.0795

LM num: 3
Linear Regression Model

petallength =

0.7501 * sepallength +
0.6105

Number of Rules : 3

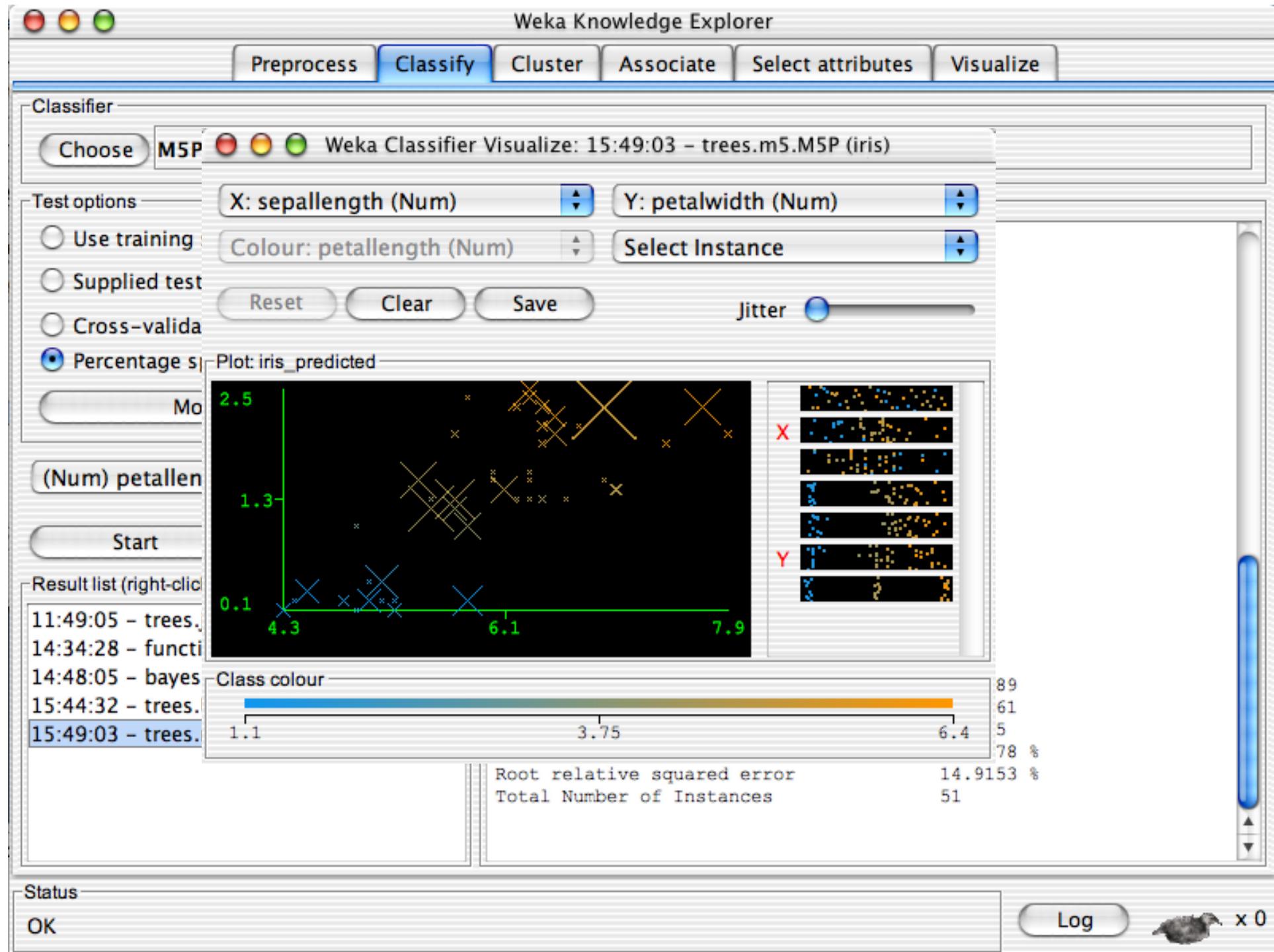
Time taken to build model: 1.31 seconds

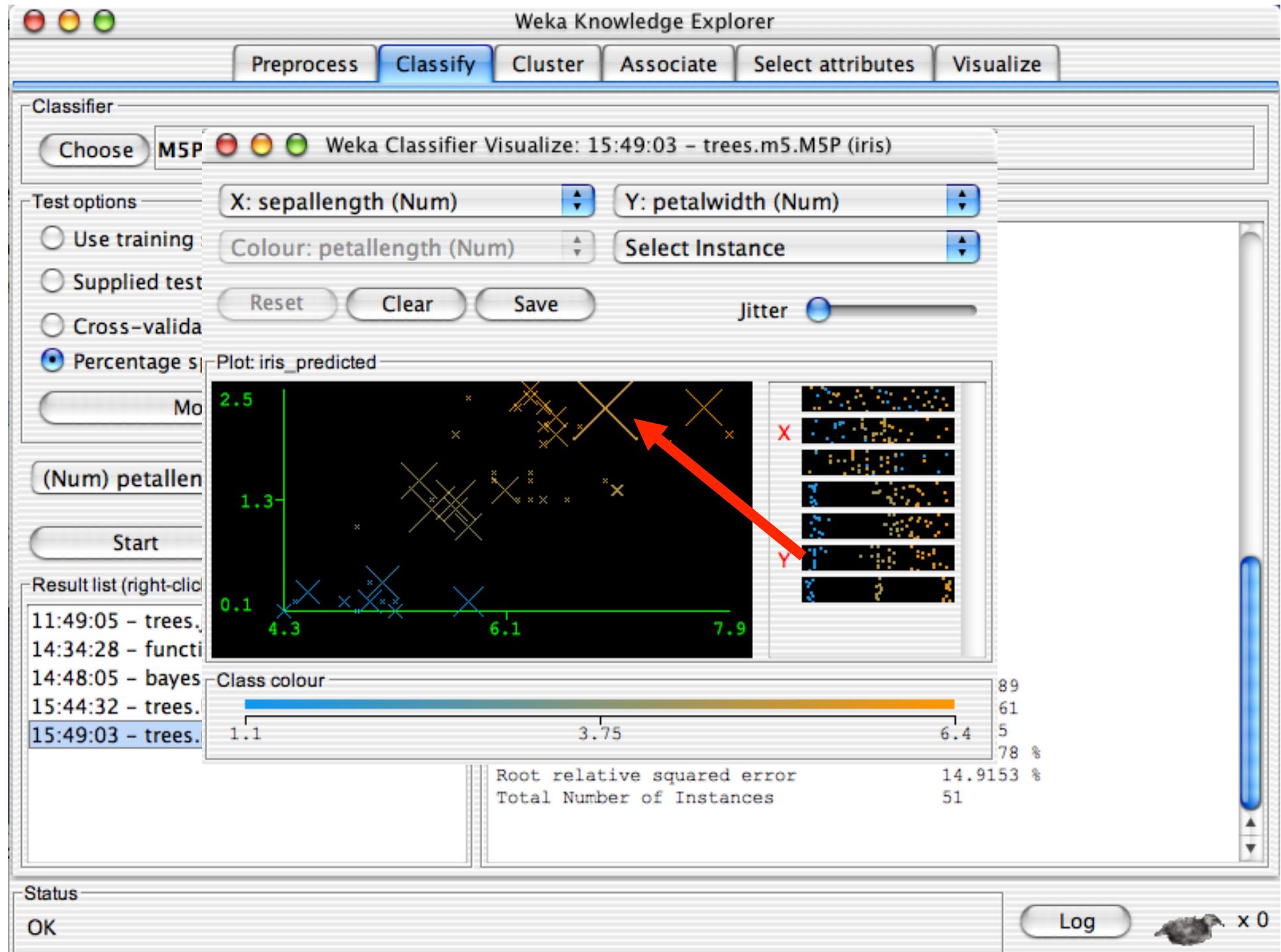
==== Evaluation on test split ====
==== Summary ===

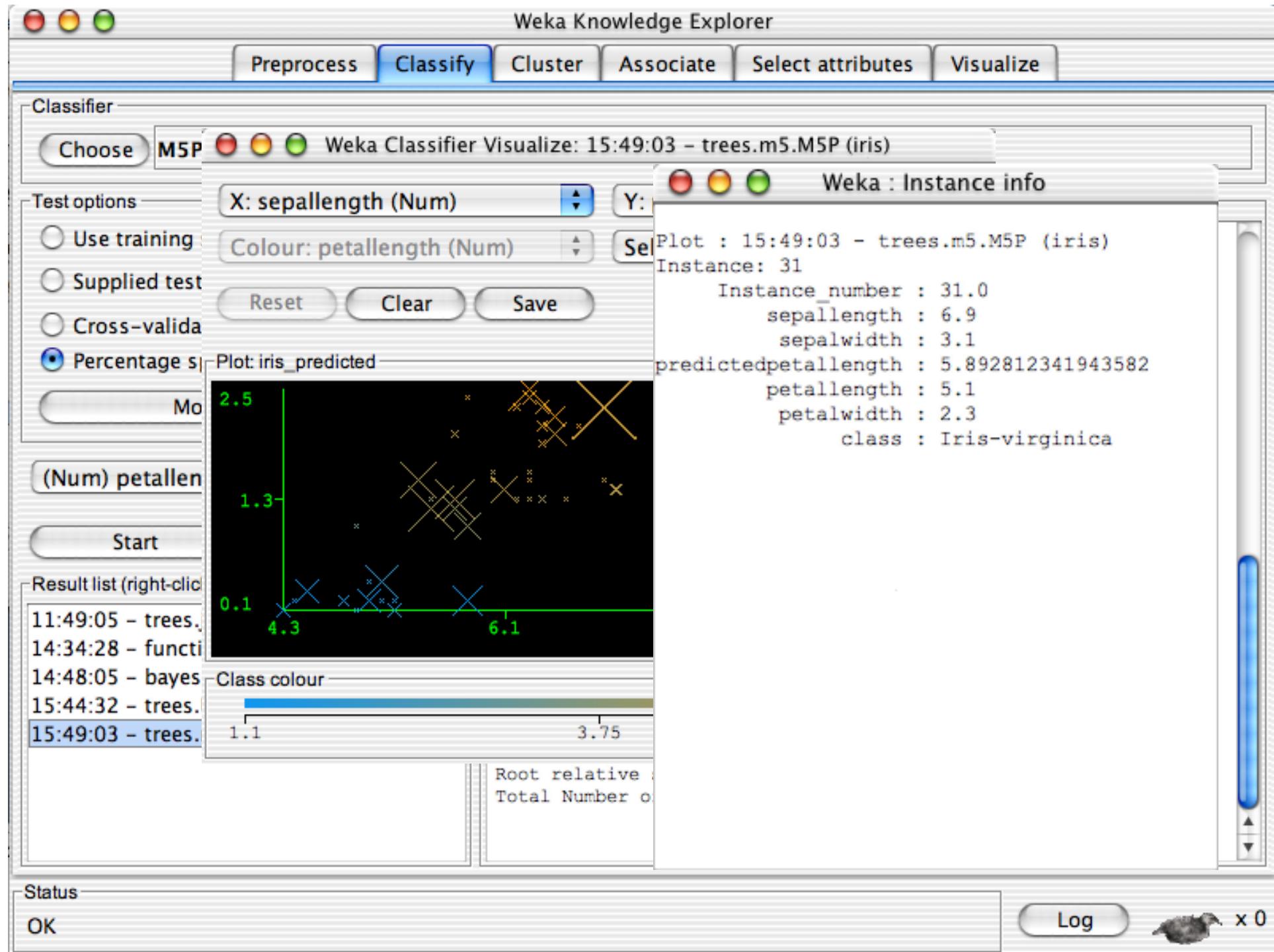
Correlation coefficient 0.9889
Mean absolute error 0.1861
Root mean squared error 0.255
Relative absolute error 11.9578 %
Root relative squared error 14.9153 %
Total Number of Instances 51
```

Status

OK Log x 0







# Explorer: clustering data

- WEKA contains “clusterers” for finding groups of similar instances in a dataset
- Implemented schemes are:
  - ◆  $k$ -Means, EM, Cobweb,  $X$ -means, FarthestFirst
- Clusters can be visualized and compared to “true” clusters (if given)
- Evaluation based on loglikelihood if clustering scheme produces a probability distribution

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Clusterer

Choose EM -I 100 -N -1 -S 100 -M 1.0E-6

Cluster mode

Use training set

Supplied test set Set...

Percentage split % 66

Classes to clusters evaluation (Nom) class

Store clusters for visualization

Ignore attributes

Start Stop

Result list (right-click for options)

Status

OK Log x 0

The screenshot shows the Weka Knowledge Explorer interface with the 'Cluster' tab selected. In the 'Clusterer' section, 'EM' is chosen with parameters '-I 100', '-N -1', '-S 100', and '-M 1.0E-6'. The 'Cluster mode' panel contains several options: 'Use training set' (selected), 'Supplied test set' (unchecked), 'Percentage split' (set to 66%), 'Classes to clusters evaluation' (unchecked), and a dropdown for '(Nom) class'. A checked checkbox for 'Store clusters for visualization' is also present. Below these are 'Ignore attributes', 'Start', and 'Stop' buttons. A large 'Result list' area is empty. The 'Status' bar at the bottom shows 'OK' and a log icon with 'x 0'.

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Clusterer

Choose EM -I 100 -N -1 -S 100 -M 1.0E-6

Cluster mode

Use training set

Supplied test set Set...

Percentage split % 66

Classes to clusters evaluation

(Nom) class ▾

Store clusters for visualization

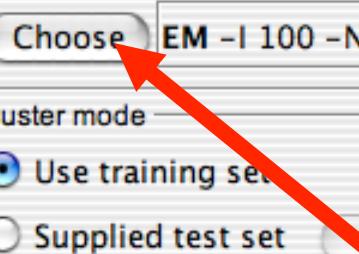
Ignore attributes

Start Stop

Result list (right-click for options)

Status

OK Log x 0



Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Clusterer

weka

clusterers

- EM
- SimpleKMeans
- Cobweb
- FarthestFirst
- XMeans

77387815

Clusterer output

Status

OK

Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Clusterer

Choose Cobweb -A 1.0 -C 0.0028209479177387815

Cluster mode

Use training set

Supplied test set Set...

Percentage split % 66

Classes to clusters evaluation (Nom) class

Store clusters for visualization

Ignore attributes

Start Stop

Result list (right-click for options)

Status

OK Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Clusterer

Choose Cobweb -A 1.0 -C 0.0028209479177387815

Cluster mode

Use training set

Supplied test set Set...

Percentage split % 66

Classes to clusters evaluation (nom) class

Store clusters for visualization

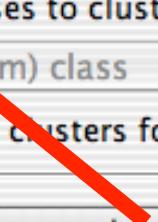
Ignore attributes

Start Stop

Result list (right-click for options)

Status

OK Log x 0



Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Clusterer

Choose Cobweb -A 1.0 -C 0.0028209479177387815

Cluster mode

Use training set

Supplied test set Set...

Percentage split % 66

Classes to clusters evaluation  
(Nom) class

Store clusters for visualization

Ignore attributes

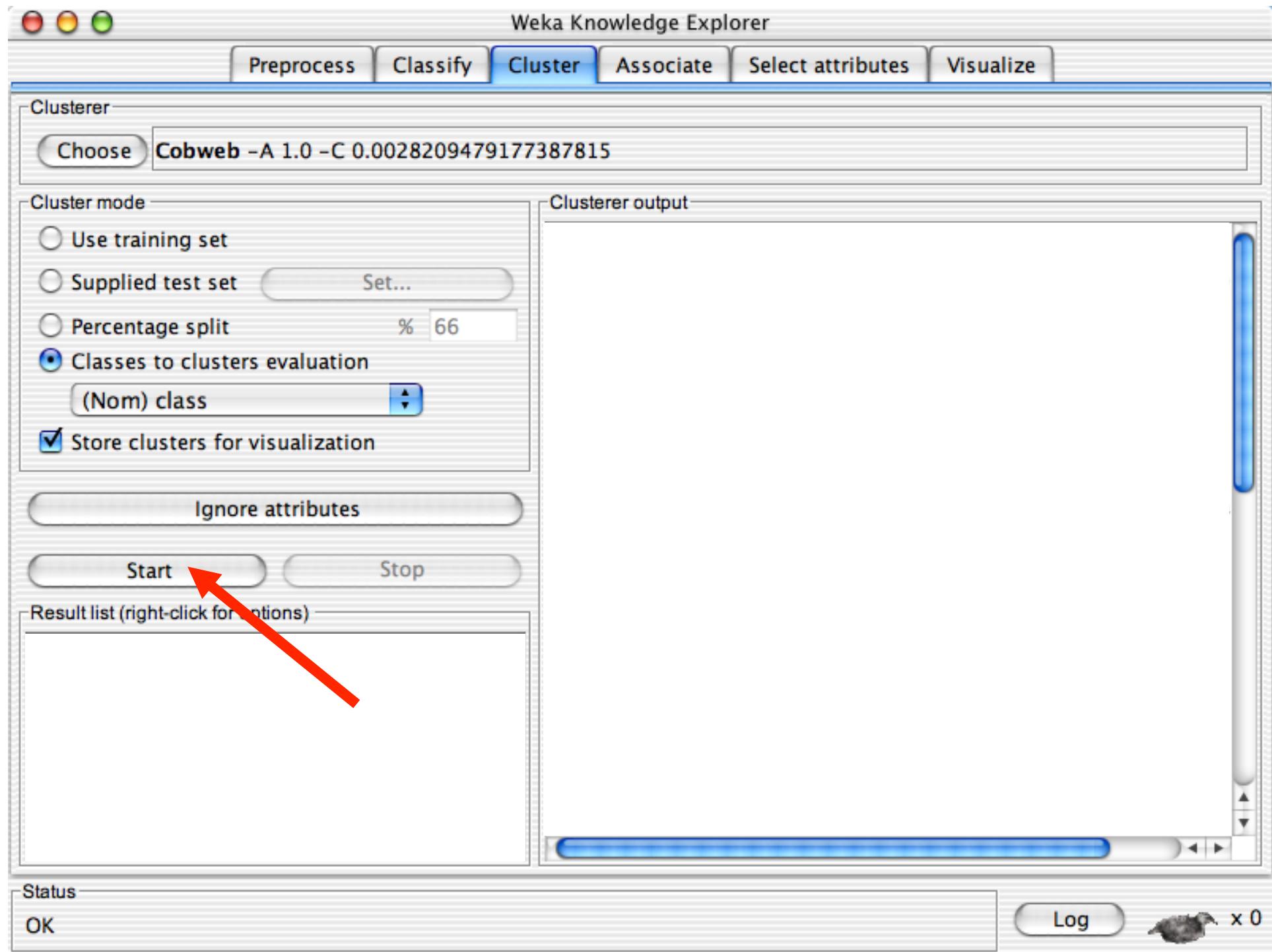
Start Stop

Result list (right-click for options)

Status

OK Log x 0

Clusterer output



Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Clusterer

Choose Cobweb -A 1.0 -C 0.0028209479177387815

Cluster mode

Use training set

Supplied test set Set...

Percentage split % 66

Classes to clusters evaluation  
(Nom) class

Store clusters for visualization

Ignore attributes

Start Stop

Result list (right-click for options)

16:05:58 – Cobweb

Clusterer output

```
==== Run information ====
Scheme: weka.clusterers.Cobweb -A 1.0 -C 0.002820947917
Relation: iris
Instances: 150
Attributes: 5
           sepallength
           sepalwidth
           petallength
           petalwidth
Ignored:
           class
Test mode: Classes to clusters evaluation on training data

==== Clustering model (full training set) ====
Number of merges: 0
Number of splits: 0
Number of clusters: 3

node 0 [150]
|   leaf 1 [ 96]
node 0 [150]
|   leaf 2 [ 54]

==== Evaluation on training set ====

```

Status

OK

Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Clusterer

Choose Cobweb -A 1.0 -C 0.0028209479177387815

Cluster mode

Use training set

Supplied test set Set...

Percentage split % 66

Classes to clusters evaluation  
(Nom) class

Store clusters for visualization

Ignore attributes

Start Stop

Result list (right-click for options)

16:05:58 – Cobweb

Clusterer output

```
==== Run information ====
Scheme: weka.clusterers.Cobweb -A 1.0 -C 0.002820947917
Relation: iris
Instances: 150
Attributes: 5
           sepallength
           sepalwidth
           petallength
           petalwidth
Ignored:
           class
Test mode: Classes to clusters evaluation on training data

==== Clustering model (full training set) ====
Number of merges: 0
Number of splits: 0
Number of clusters: 3

node 0 [150]
|   leaf 1 [ 96]
node 0 [150]
|   leaf 2 [ 54]

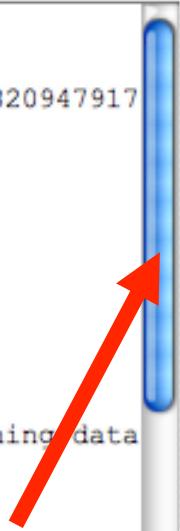
==== Evaluation on training set ====

```

Status

OK

Log x 0



Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Clusterer

Choose Cobweb -A 1.0 -C 0.0028209479177387815

Cluster mode

Use training set

Supplied test set Set...

Percentage split % 66

Classes to clusters evaluation (Nom) class

Store clusters for visualization

Ignore attributes

Start Stop

Result list (right-click for options)

16:05:58 - Cobweb

Status

OK

Clusterer output

Number of clusters: 3

node 0 [ 150]  
| leaf 1 [ 96]  
node 0 [ 150]  
| leaf 2 [ 54]

Clustered Instances

1 100 ( 67%)  
2 50 ( 33%)

Class attribute: class  
Classes to Clusters:

1 2 <-- assigned to cluster  
0 50 | Iris-setosa  
50 0 | Iris-versicolor  
50 0 | Iris-virginica

Cluster 1 <-- Iris-versicolor  
Cluster 2 <-- Iris-setosa

Incorrectly clustered instances : 50.0 33.3333 %

Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Clusterer

Choose Cobweb -A 1.0 -C 0.0028209479177387815

Cluster mode

Use training set

Supplied test set Set...

Percentage split % 66

Classes to clusters evaluation (Nom) class

Store clusters for visualization

Ignore attributes

Start Stop

Result list (right-click for options)

16:05:58 – Cobweb

Clusterer output

Number of clusters: 3

node 0 [ 150]  
| leaf 1 [ 96]  
node 0 [ 150]  
| leaf 2 [ 54]

Clustered Instances

1 100 ( 67%)  
2 50 ( 33%)

Class attribute: class  
Classes to Clusters:

1 2 <-- assigned to cluster  
0 50 | Iris-setosa  
50 0 | Iris-versicolor  
50 0 | Iris-virginica

Cluster 1 <-- Iris-versicolor  
Cluster 2 <-- Iris-setosa

Incorrectly clustered instances : 50.0 33.3333 %

Status OK Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Clusterer

Choose Cobweb -A 1.0 -C 0.0028209479177387815

Cluster mode

Use training set

Supplied test set Set...

Percentage split % 66

Classes to clusters evaluation  
(Nom) class

Store clusters for visualization

Ignore attributes

Start Stop

Result list (right-click for options)

16:05:58 - Cobweb

- View in main window
- View in separate window
- Save result buffer
- Load model
- Save model
- Re-evaluate model on current test set

Visualize cluster assignments

Visualize tree

Clusterer output

```
==== Run information ====
Scheme: weka.clusterers.Cobweb -A 1.0 -C 0.002820947917
Relation: iris
Instances: 150
Attributes: 5
           sepallength
           sepalwidth
           petallength
           petalwidth
Ignored:
           class
Test mode: Classes to clusters evaluation on training data
==== Clustering model (full training set) ====
Number of merges: 0
Number of splits: 0
Number of clusters: 3
```

Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Clusterer

Choose Cobweb -A 1.0 -C 0.0028209479177387815

Cluster mode

- Use training set
- Supplied test se
- Percentage split
- Classes to cluster  
(Nom) class
- Store clusters fo

Tree View

Weka Classifier Tree Visualizer: 16:05:58 – Cobweb (iris)

node 0 (150)

leaf 1 (96)

leaf 2 (54)

Result list (right-click for context menu)

16:05:58 – Cobweb

Status

OK

Log x 0

on on training data

==

Weka Knowledge Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Clusterer

Choose Cobweb -A 1.0 -C 0.0028209479177387815

Cluster mode

Use training set

Supplied test set Set...

Percentage split % 66

Classes to clusters evaluation  
(Nom) class

Store clusters for visualization

Ignore attributes

Start Stop

Result list (right-click for options)

16:05:58 – Cobweb

- [View in main window](#)
- [View in separate window](#)
- [Save result buffer](#)
- [Load model](#)
- [Save model](#)
- [Re-evaluate model on current test set](#)

Clusterer output

```
==== Run information ====
Scheme: weka.clusterers.Cobweb -A 1.0 -C 0.002820947917
Relation: iris
Instances: 150
Attributes: 5
           sepallength
           sepalwidth
           petallength
           petalwidth
Ignored:
           class
Test mode: Classes to clusters evaluation on training data
==== Clustering model (full training set) ====
Number of merges: 0
Number of splits: 0
Number of clusters: 3
==== Evaluation on training set ====

```

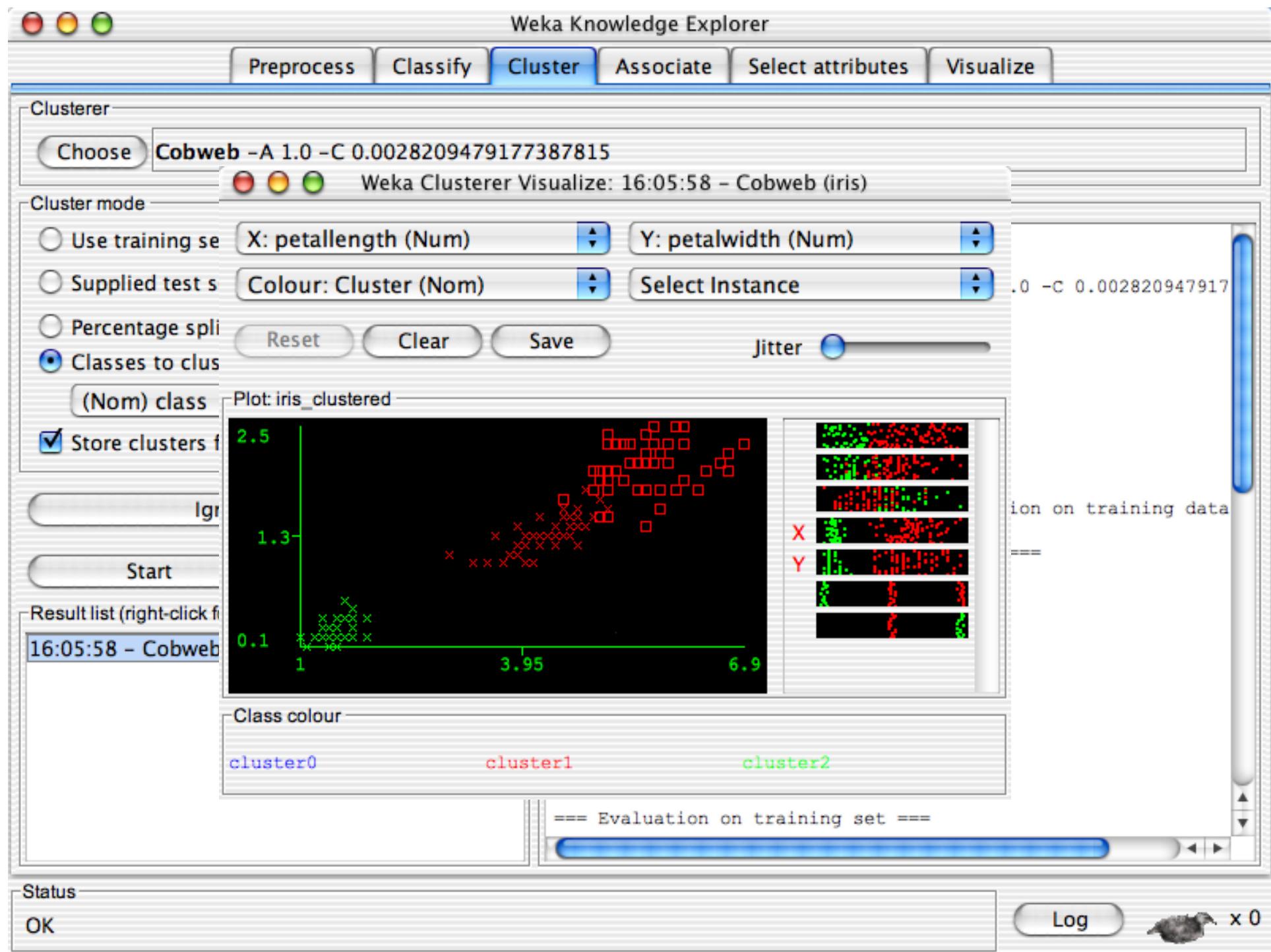
Status

OK

Visualize cluster assignments

Visualize tree

Log x 0



# Explorer: finding associations

- WEKA contains an implementation of the Apriori algorithm for learning association rules
  - ◆ Works only with discrete data
- Can identify statistical dependencies between groups of attributes:
  - ◆ milk, butter  $\Rightarrow$  bread, eggs (with confidence 0.9 and support 2000)
- Apriori can compute all rules that have a given minimum support and exceed a given confidence

Weka Knowledge Explorer

Preprocess Classify Cluster Associate **Select attributes** Visualize

Associator

Choose **Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0**

Start Stop

Associator output

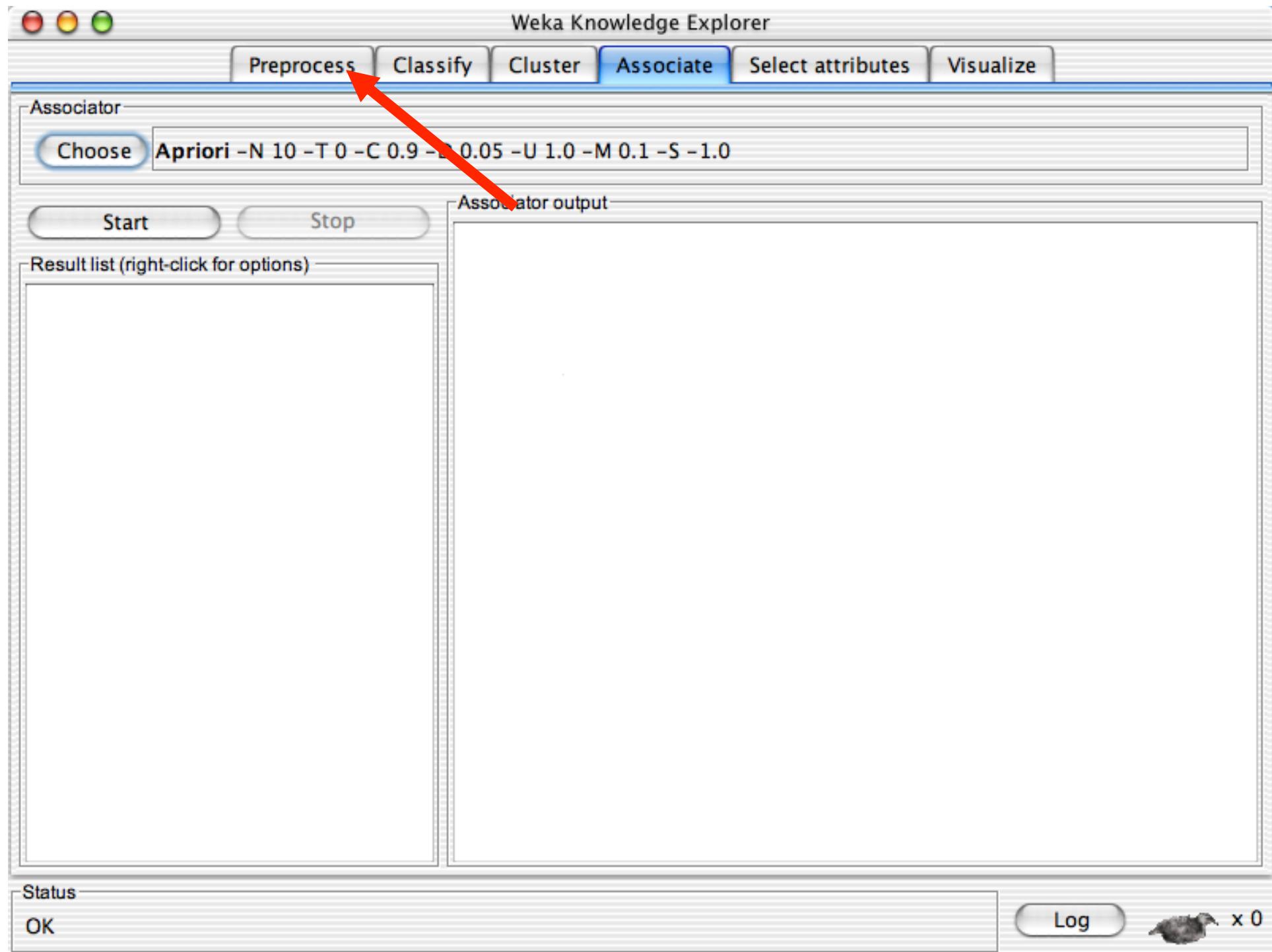
Result list (right-click for options)

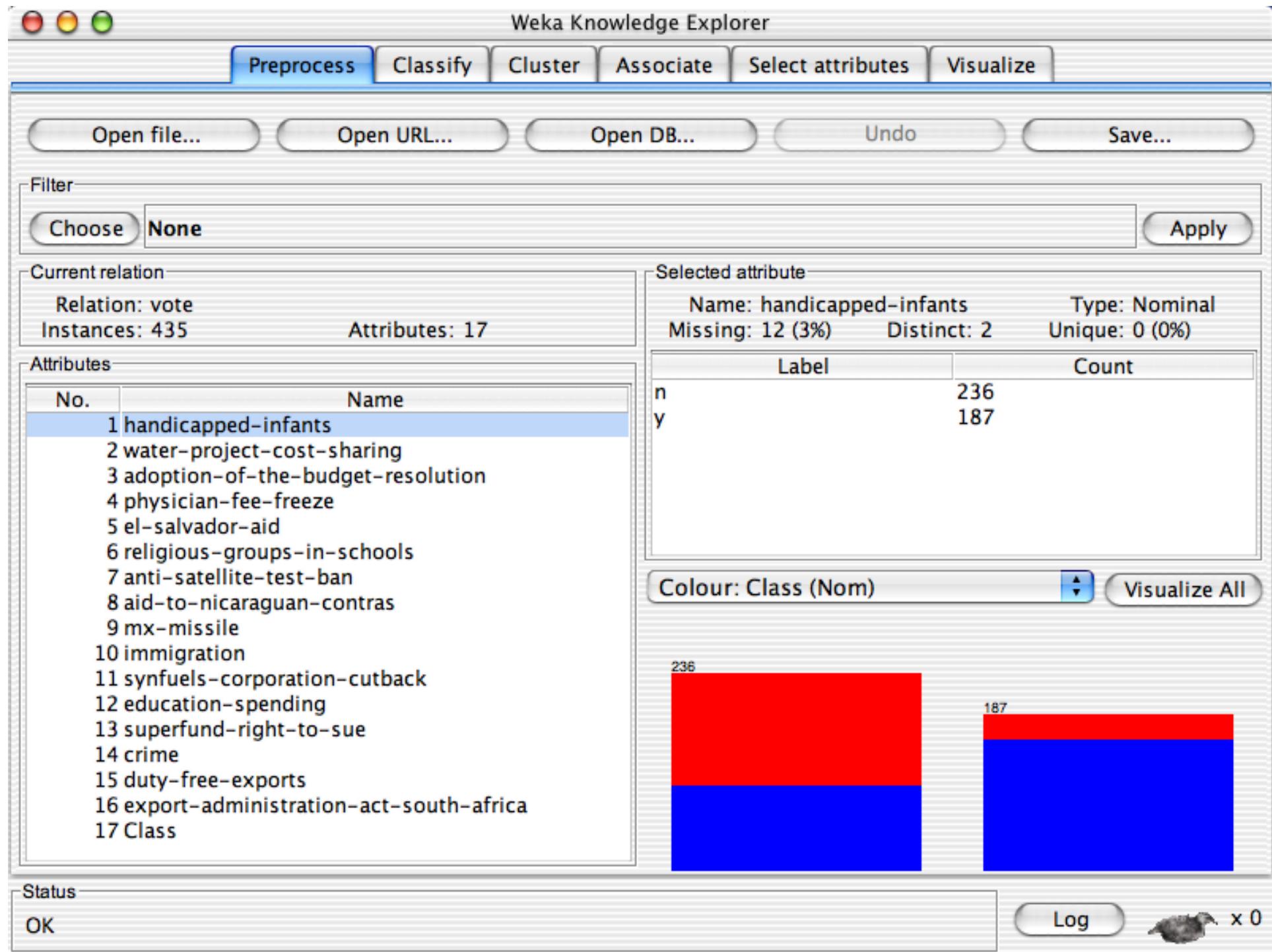
Status

OK

Log x 0

The screenshot displays the Weka Knowledge Explorer interface, specifically the Associate tab. The main window title is "Weka Knowledge Explorer". The top navigation bar includes tabs for Preprocess, Classify, Cluster, Associate (which is highlighted in blue), Select attributes, and Visualize. Below the tabs, a section titled "Associator" contains a "Choose" button followed by the command "Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0". There are "Start" and "Stop" buttons below this. To the right, a large rectangular area is labeled "Associator output" and is currently empty. On the left, a vertical panel is labeled "Result list (right-click for options)" and also appears empty. At the bottom, a "Status" field shows "OK" and a "Log" button with a small icon and the text "x 0".





Weka Knowledge Explorer

Preprocess Classify Cluster Associate **Select attributes** Visualize

Open file... Open URL... Open DB... Undo Save...

Filter

Choose **None** Apply

Current relation

Relation: vote Instances: 435 Attributes: 17

Attributes

No.	Name
1	handicapped-infants
2	water-project-cost-sharing
3	adoption-of-the-budget-resolution
4	physician-fee-freeze
5	el-salvador-aid
6	religious-groups-in-schools
7	anti-satellite-test-ban
8	aid-to-nicaraguan-contras
9	mx-missile
10	immigration
11	synfuels-corporation-cutback
12	education-spending
13	superfund-right-to-sue
14	crime
15	duty-free-exports
16	export-administration-act-south-africa
17	Class

Selected attribute

Name: handicapped-infants Type: Nominal  
Missing: 12 (3%) Distinct: 2 Unique: 0 (0%)

Label	Count
n	236
y	187

Colour: Class (Nom) Visualize All

Status

OK Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate **Select attributes** Visualize

Associator

Choose **Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0**

Start Stop

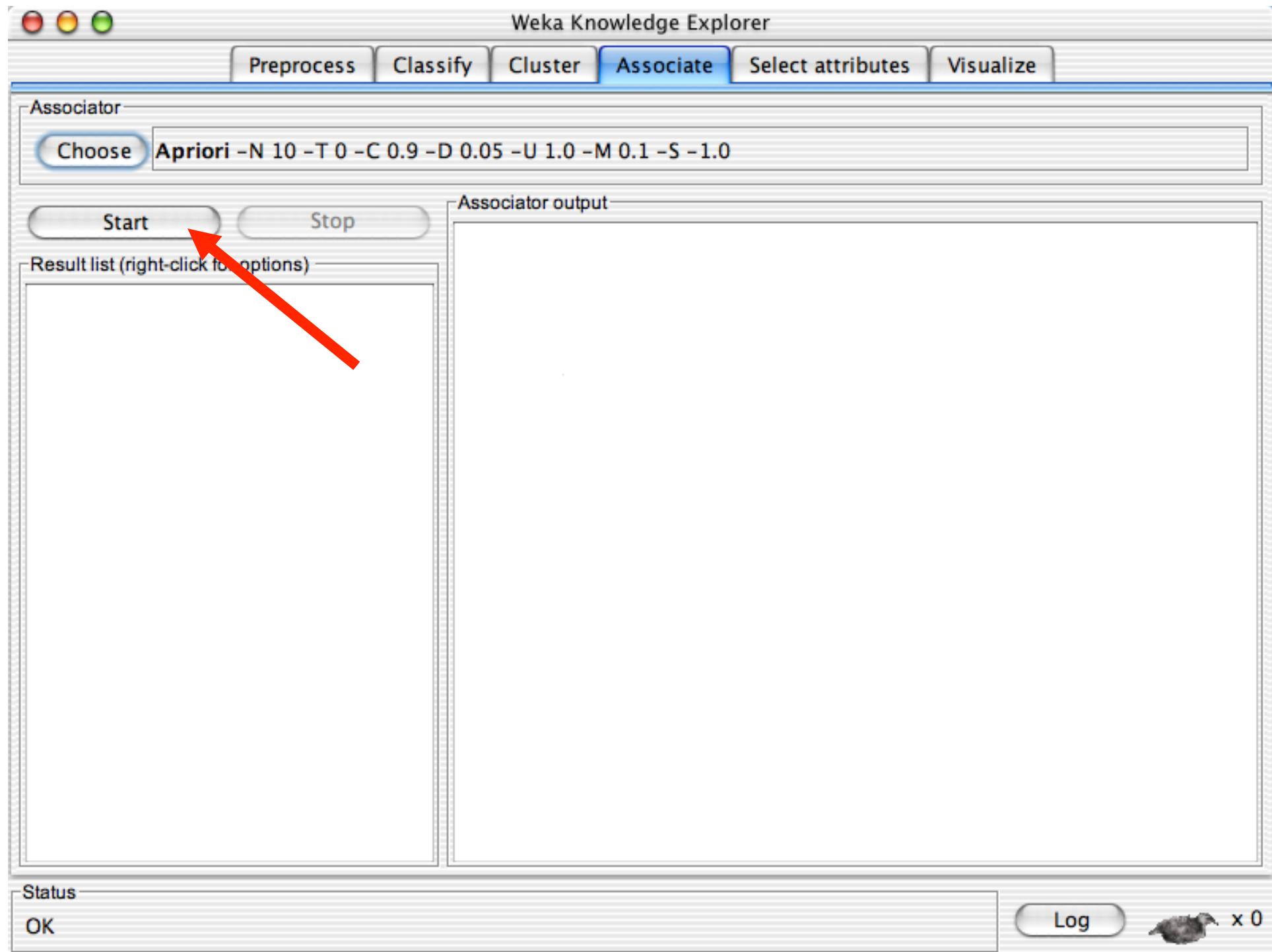
Associator output

Result list (right-click for options)

Status

OK

Log x 0



Weka Knowledge Explorer

Preprocess Classify Cluster Associate **Select attributes** Visualize

Associator

Choose **Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0**

**Start** **Stop**

Result list (right-click for options)

16:29:37 – Apriori

Associator output

```
Minimum metric <confidence>: 0.9
Number of cycles performed: 11

Generated sets of large itemsets:

Size of set of large itemsets L(1): 20
Size of set of large itemsets L(2): 17
Size of set of large itemsets L(3): 6
Size of set of large itemsets L(4): 1

Best rules found:

1. adoption-of-the-budget-resolution=y physician-fee-freeze=n 219 => Class=democrat
2. adoption-of-the-budget-resolution=y physician-fee-freeze=n aid-to-nicaraguan-contras=y 210 => Class=democrat 210
3. physician-fee-freeze=n aid-to-nicaraguan-contras=y 211 => Class=democrat 210
4. physician-fee-freeze=n education-spending=n 202 => Class=democrat 201 conf:(0.99)
5. physician-fee-freeze=n 247 => Class=democrat 245 conf:(0.99)
6. el-salvador-aid=n Class=democrat 200 => aid-to-nicaraguan-contras=y 197 conf:(0.98)
7. el-salvador-aid=n 208 => aid-to-nicaraguan-contras=y 204 conf:(0.98)
8. adoption-of-the-budget-resolution=y aid-to-nicaraguan-contras=y Class=democrat 200
9. el-salvador-aid=n aid-to-nicaraguan-contras=y 204 => Class=democrat 197 conf:(0.98)
10. aid-to-nicaraguan-contras=y Class=democrat 218 => physician-fee-freeze=n 210
```

Status

OK

Log x 0

# Explorer: attribute selection

- Panel that can be used to investigate which (subsets of) attributes are the most predictive ones
- Attribute selection methods contain two parts:
  - ◆ A search method: best-first, forward selection, random, exhaustive, genetic algorithm, ranking
  - ◆ An evaluation method: correlation-based, wrapper, information gain, chi-squared, ...
- Very flexible: WEKA allows (almost) arbitrary combinations of these two

Weka Knowledge Explorer

Preprocess Classify Cluster Associate **Select attributes** Visualize

Attribute Evaluator

Choose **CfsSubsetEval**

Search Method

Choose **BestFirst - D 1 - N 5**

Attribute Selection Mode

Use full training set  
 Cross-validation      Folds 10  
                            Seed 1

Attribute selection output

(Nom) Class

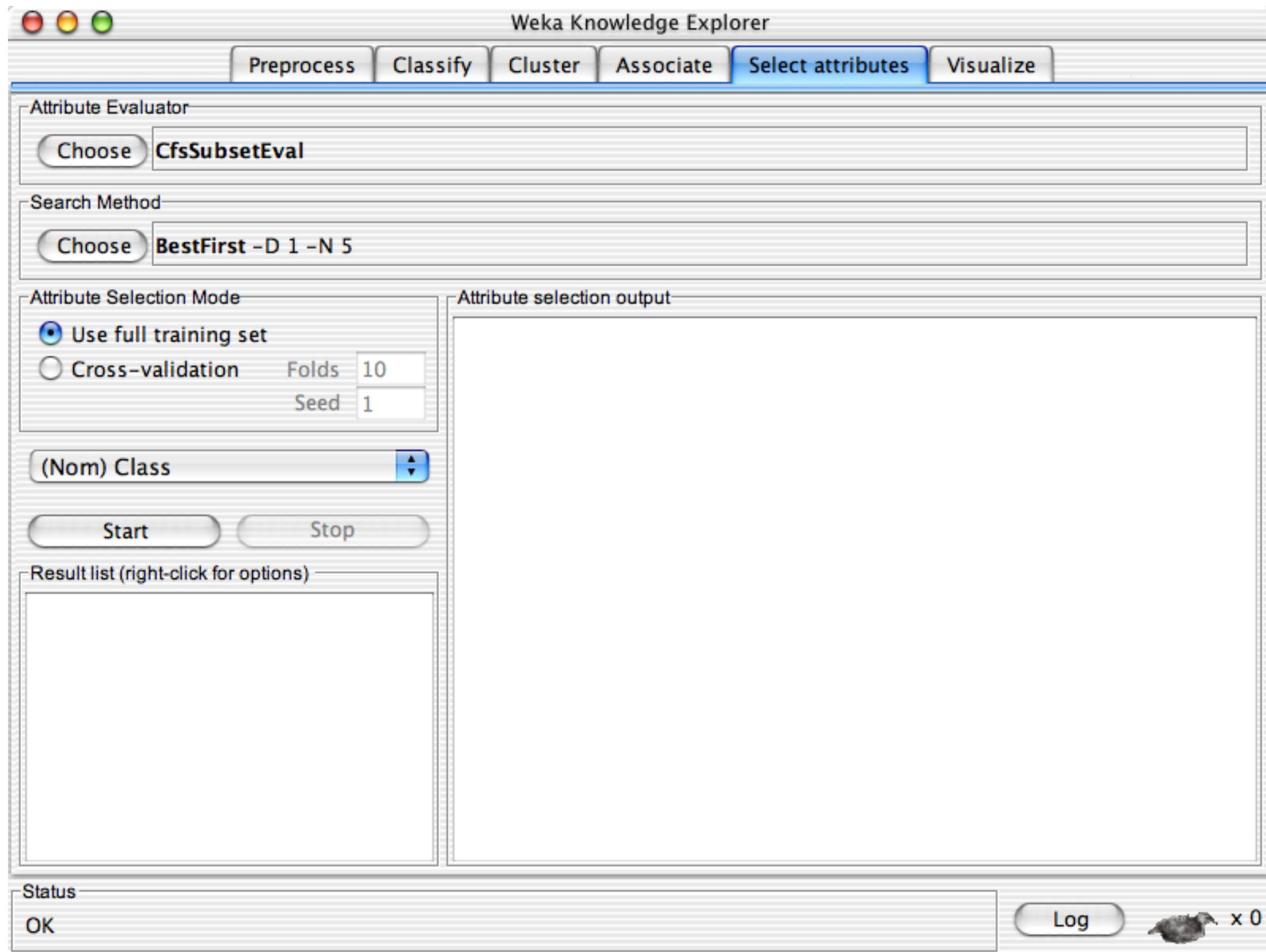
Start Stop

Result list (right-click for options)

Status

OK

Log x 0



Weka Knowledge Explorer

Preprocess Classify Cluster Associate **Select attributes** Visualize

Attribute Evaluator

Choose **CfsSubsetEval**

Search Method

Choose **BestFirst - D 1 - N 5**

Attribute Selection Mode

Use full training set  
 Cross-validation Folds 10  
Seed 1

(Nom) Class

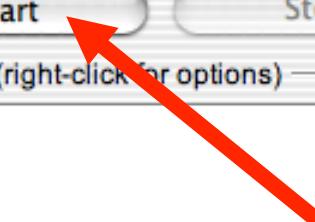
Start Stop

Result list (right-click for options)

Status

OK

Log x 0



Weka Knowledge Explorer

Preprocess Classify Cluster Associate **Select attributes** Visualize

Attribute Evaluator

Choose **CfsSubsetEval**

Search Method

Choose **BestFirst - D 1 - N 5**

Attribute Selection Mode

Use full training set  
 Cross-validation      Folds 10  
                            Seed 1

(Nom) Class

Start Stop

Result list (right-click for options)

16:39:40 – BestFirst + CfsSubsetEval

Attribute selection output

```
duty-free-exports
export-administration-act-south-africa
Class
Evaluation mode: evaluate on all training data

==== Attribute Selection on all input data ====

Search Method:
    Best first.
    Start set: no attributes
    Search direction: forward
    Stale search after 5 node expansions
    Total number of subsets evaluated: 83
    Merit of best subset found: 0.729

Attribute Subset Evaluator (supervised, Class (nominal): 17 Class):
    CFS Subset Evaluator

Selected attributes: 4 : 1
    physician-fee-freeze
```

Status

OK

Log x 0

Weka Knowledge Explorer

Preprocess Classify Cluster Associate **Select attributes** Visualize

Attribute Evaluator

Choose **CfsSubsetEval**

Search Method

Choose **BestFirst - D 1 - N 5**

Attribute Selection Mode

Use full training set  
 Cross-validation Folds 10 Seed 1

(Nom) Class

Start Stop

Result list (right-click for options)

16:39:40 – BestFirst + CfsSubsetEval

Attribute selection output

```
duty-free-exports
export-administration-act-south-africa
Class
Evaluation mode: evaluate on all training data

==== Attribute Selection on all input data ====

Search Method:
  Best first.
  Start set: no attributes
  Search direction: forward
  Stale search after 5 node expansions
  Total number of subsets evaluated: 83
  Merit of best subset found: 0.729

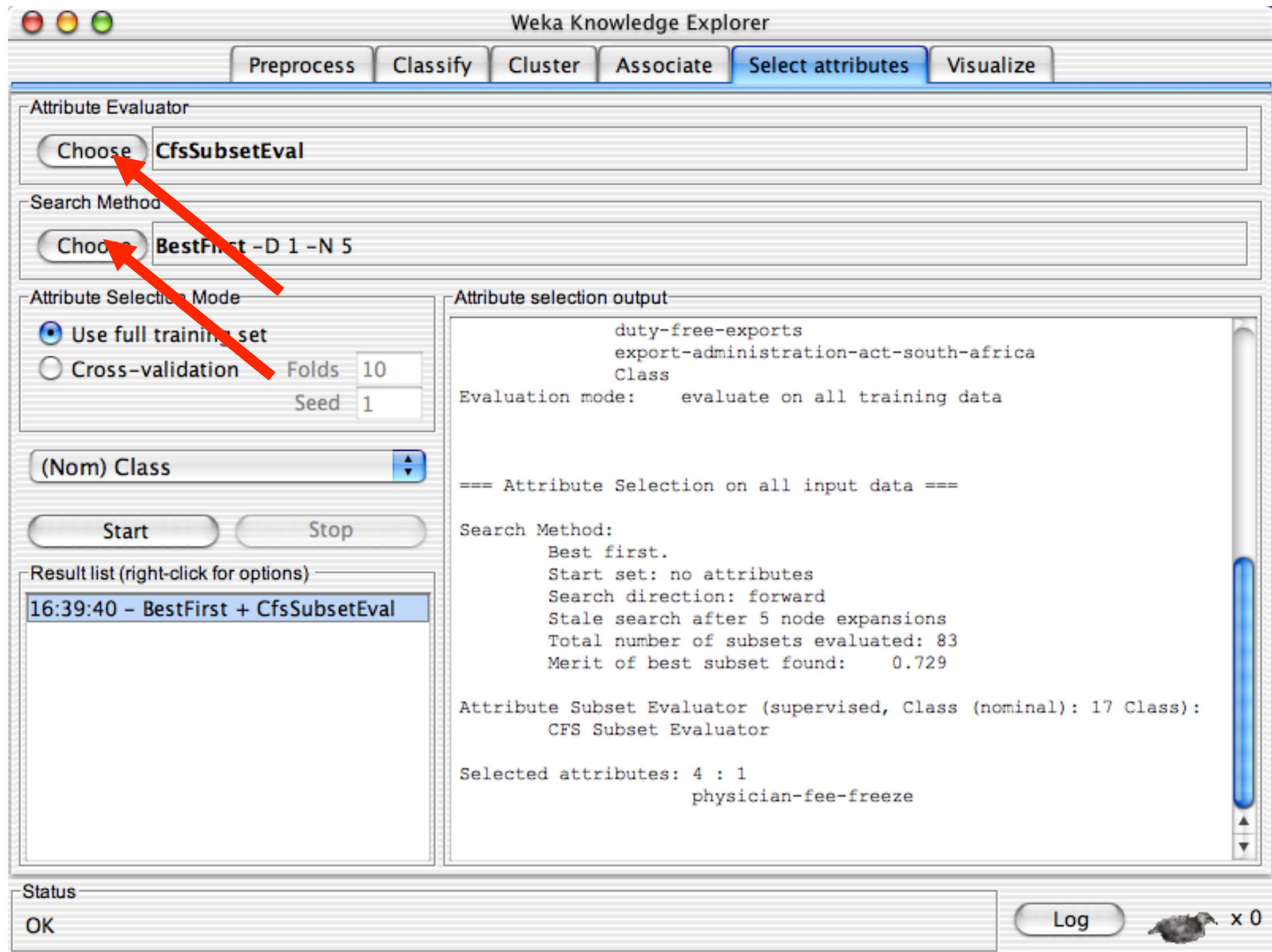
Attribute Subset Evaluator (supervised, Class (nominal): 17 Class):
  CFS Subset Evaluator

Selected attributes: 4 : 1
  physician-fee-freeze
```

Status

OK

Log x 0



# Weka Knowledge Explorer

[Preprocess](#)[Classify](#)[Cluster](#)[Associate](#)[Select attributes](#)[Visualize](#)

## Attribute Evaluator

- weka
- attributeSelection
  - CfsSubsetEval
  - ClassifierSubsetEval
  - WrapperSubsetEval
  - ConsistencySubsetEval
  - ReliefFAttributeEval
  - InfoGainAttributeEval**
  - GainRatioAttributeEval
  - SymmetricalUncertAttributeEval
  - OneRAttributeEval
  - ChiSquaredAttributeEval
  - PrincipalComponents
  - SVMAttributeEval

### Attribute selection output

```
duty-free-exports  
export-administration-act-south-africa  
Class
```

```
uation mode: evaluate on all training data
```

```
Attribute Selection on all input data ===
```

```
ch Method:  
Best first.  
Start set: no attributes  
Search direction: forward  
Stale search after 5 node expansions  
Total number of subsets evaluated: 83  
Merit of best subset found: 0.729
```

```
lolute Subset Evaluator (supervised, Class (nominal): 17 Class):  
CFS Subset Evaluator
```

```
Selected attributes: 4 : 1  
physician-fee-freeze
```

Status

OK

Log



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

## Attribute Evaluator

Choose

InfoGainAttributeEval

## Search Method

- weka
- attributeSelection
  - BestFirst
  - ForwardSelection
  - RaceSearch
  - GeneticSearch
  - RandomSearch
  - ExhaustiveSearch
  - Ranker
  - RankSearch

E308 - N - 1

Attribute selection output

```
duty-free-exports
export-administration-act-south-africa
Class
uation mode: evaluate on all training data
```

Attribute Selection on all input data ===

```
ch Method:
    Best first.
Start set: no attributes
Search direction: forward
Stale search after 5 node expansions
Total number of subsets evaluated: 83
Merit of best subset found: 0.729
```

```
ibute Subset Evaluator (supervised, Class (nominal): 17 Class):
    CFS Subset Evaluator
```

```
ected attributes: 4 : 1
                    physician-fee-freeze
```

## Status

OK

Log



Weka Knowledge Explorer

Preprocess Classify Cluster Associate **Select attributes** Visualize

Attribute Evaluator

Choose **InfoGainAttributeEval**

Search Method

Choose **Ranker -T -1.7976931348623157E308 -N -1**

Attribute Selection Mode

Use full training set

Cross-validation Folds 10  
Seed 1

(Nom) Class

Start Stop

Result list (right-click for options)

16:39:40 – BestFirst + CrsSubsetEval

Status

OK

Log x 0

A red arrow points from the 'Start' button to the 'Result list' area, indicating where the user should click to begin the attribute selection process.

Weka Knowledge Explorer

Preprocess Classify Cluster Associate **Select attributes** Visualize

Attribute Evaluator

Choose **InfoGainAttributeEval**

Search Method

Choose **Ranker -T -1.7976931348623157E308 -N -1**

Attribute Selection Mode

Use full training set  
 Cross-validation Folds 10 Seed 1

(Nom) Class

Start Stop

Result list (right-click for options)

16:39:40 – BestFirst + CfsSubsetEval  
16:43:05 – Ranker + InfoGainAttributeEval

Attribute selection output

Information Gain Ranking Filter

Ranked attributes:

0.7078541	4 physician-fee-freeze
0.4185726	3 adoption-of-the-budget-resolution
0.4028397	5 el-salvador-aid
0.34036	12 education-spending
0.3123121	14 crime
0.3095576	8 aid-to-nicaraguan-contras
0.2856444	9 mx-missile
0.2121705	13 superfund-right-to-sue
0.2013666	15 duty-free-exports
0.1902427	7 anti-satellite-test-ban
0.1404643	6 religious-groups-in-schools
0.1211834	1 handicapped-infants
0.1007458	11 synfuels-corporation-cutback
0.0529956	16 export-administration-act-south-africa
0.0049097	10 immigration
0.0000117	2 water-project-cost-sharing

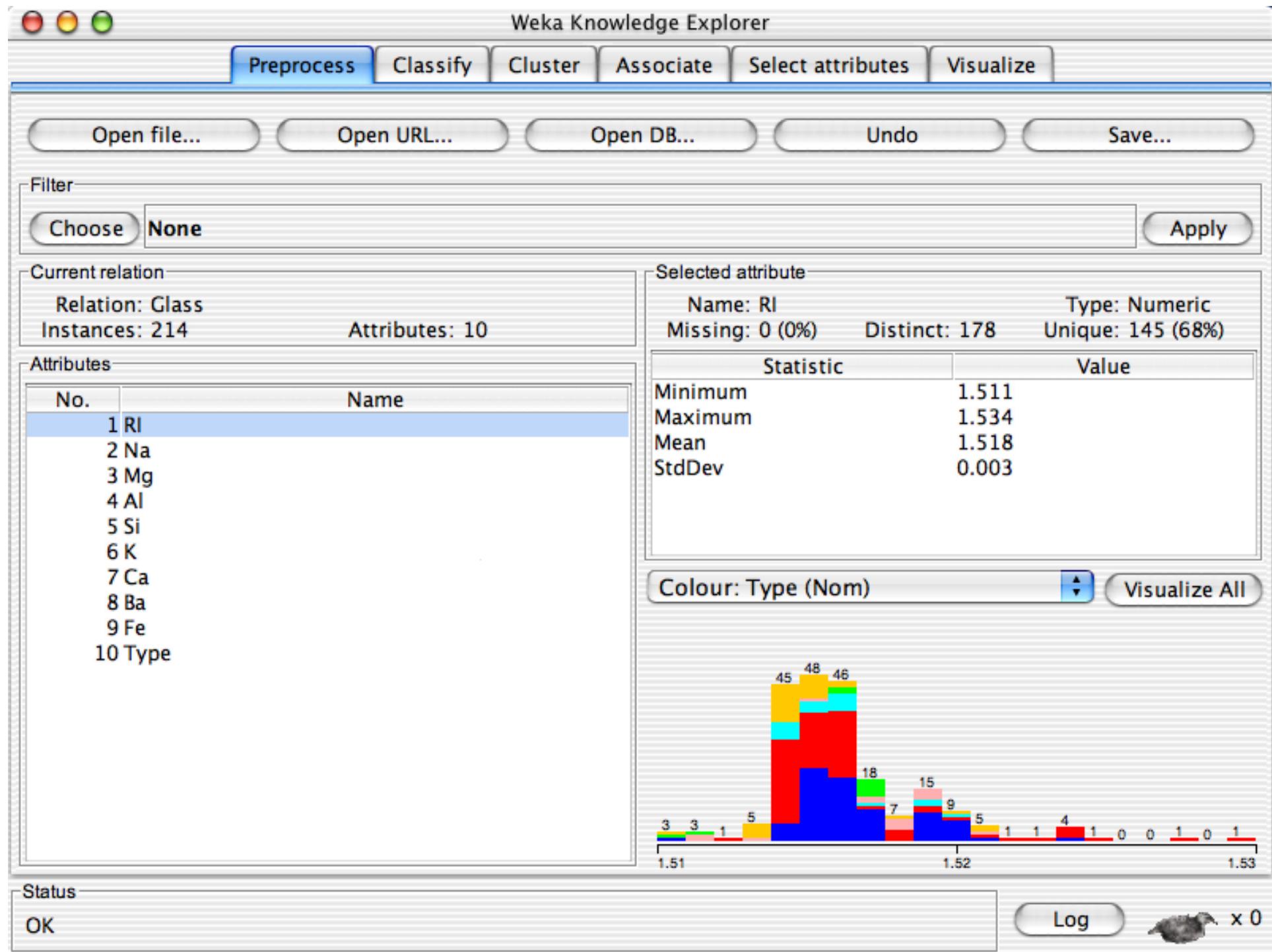
Selected attributes: 4,3,5,12,14,8,9,13,15,7,6,1,11,16,10,2 : 16

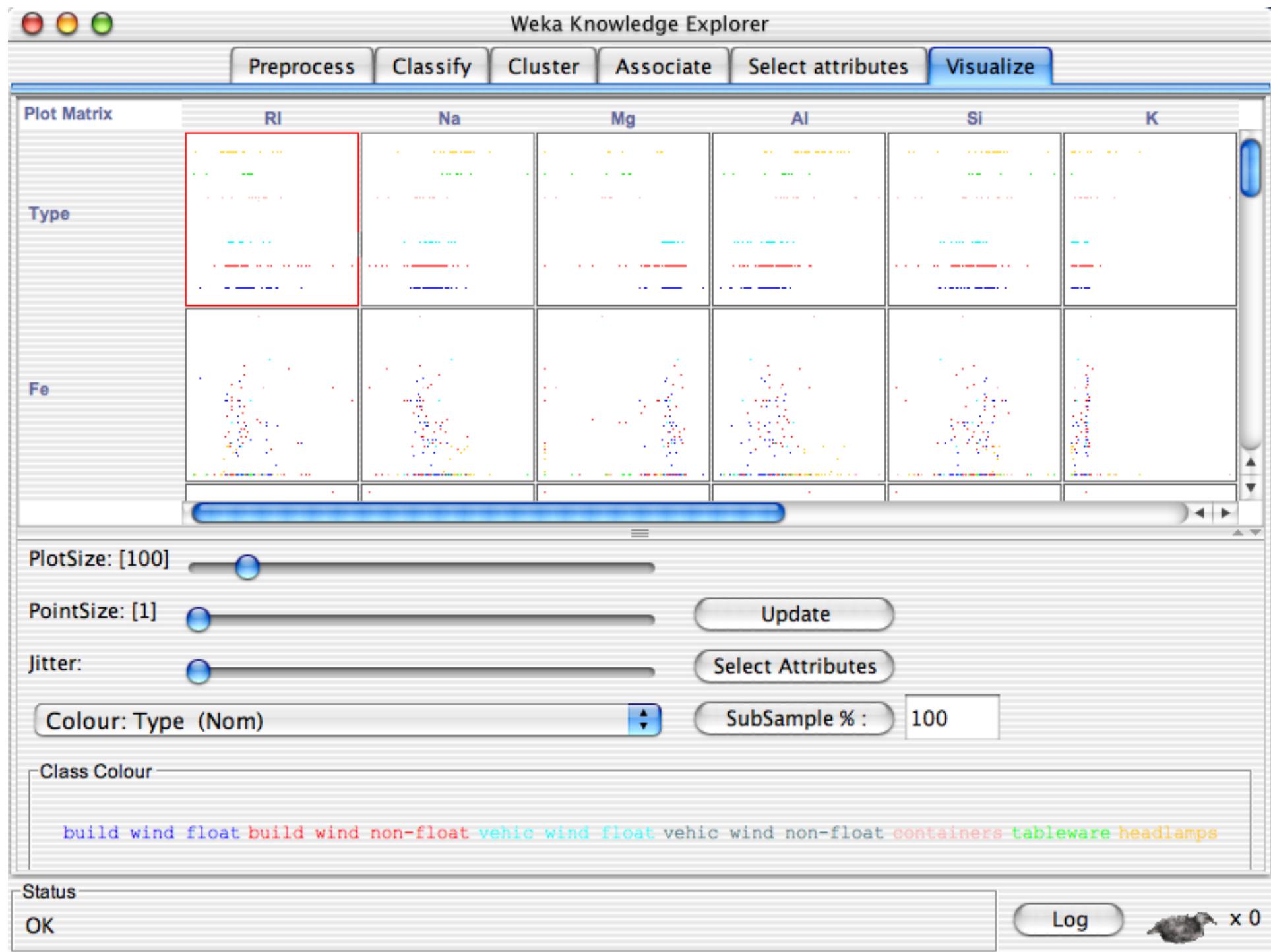
Status

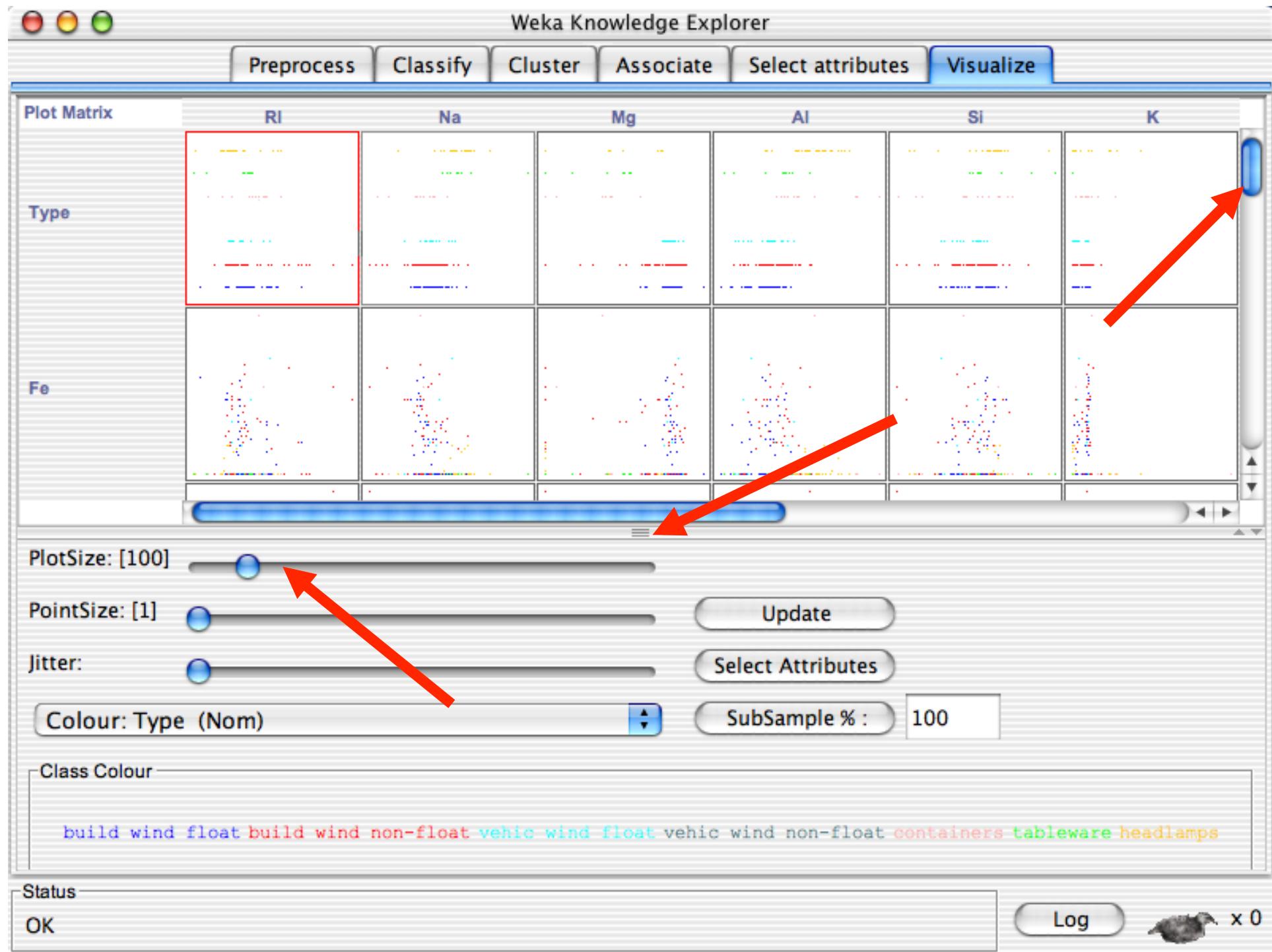
OK Log x 0

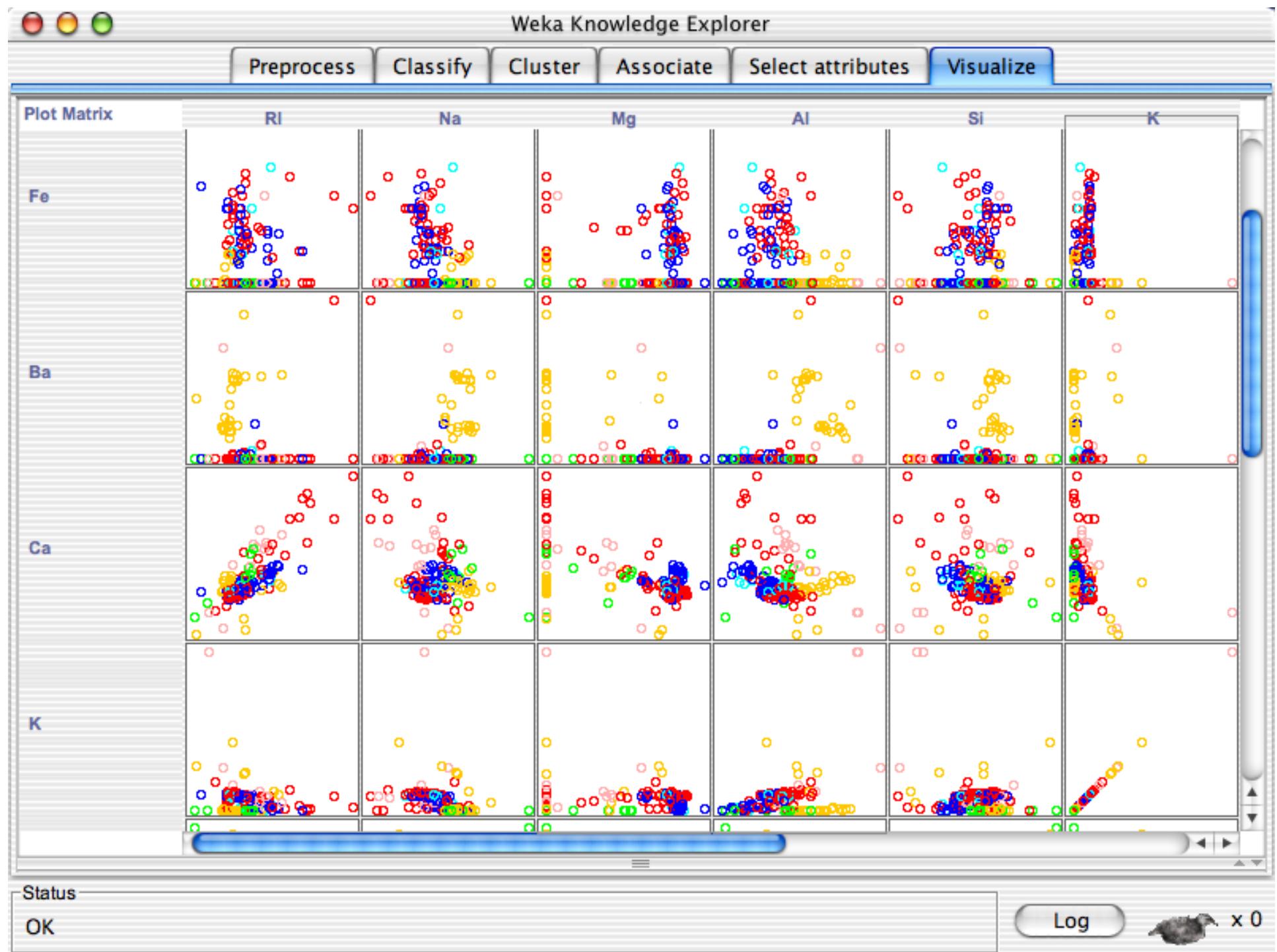
# Explorer: data visualization

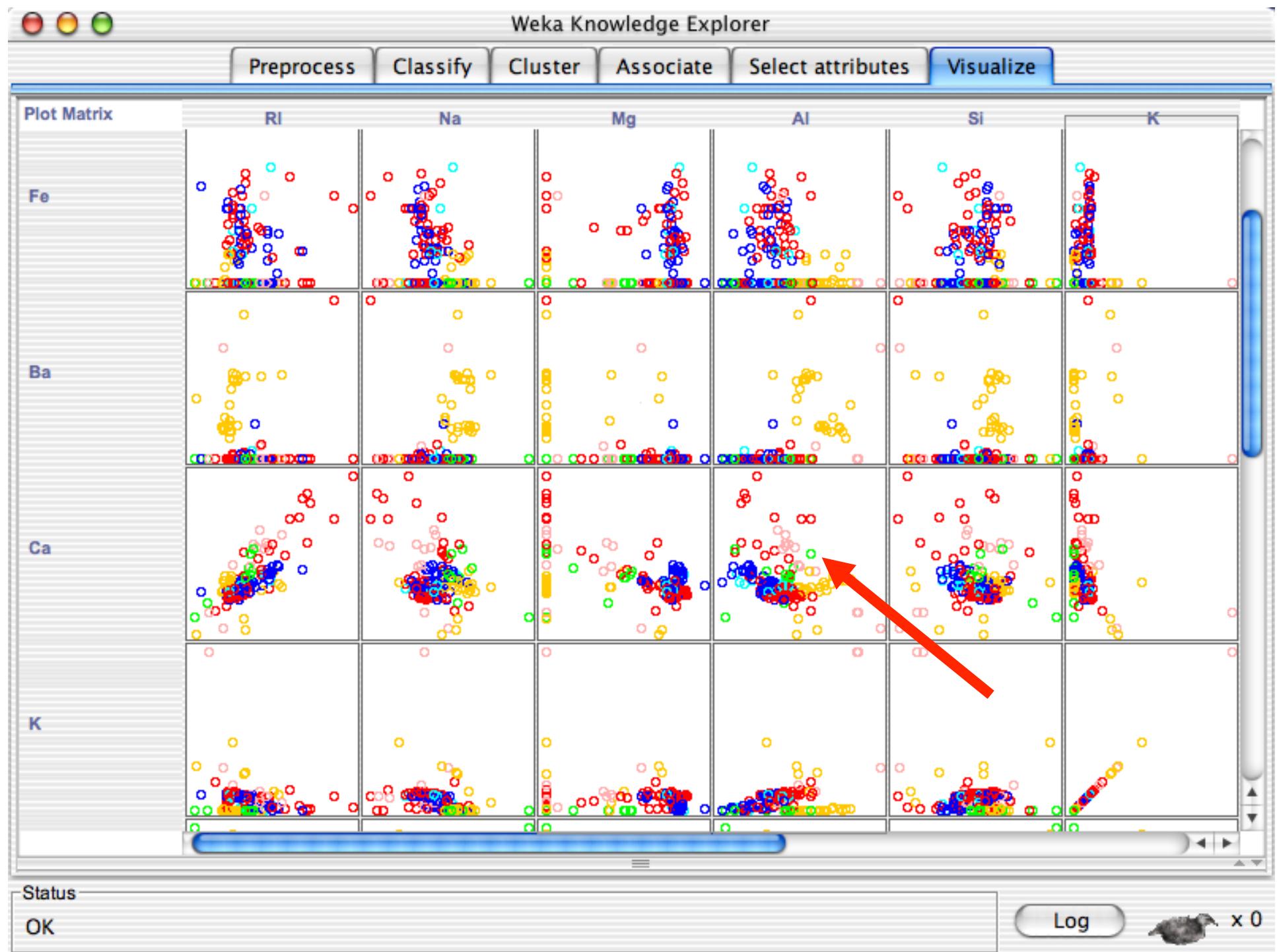
- Visualization very useful in practice: e.g. helps to determine difficulty of the learning problem
- WEKA can visualize single attributes (1-d) and pairs of attributes (2-d)
  - ◆ To do: rotating 3-d visualizations (Xgobi-style)
- Color-coded class values
- “Jitter” option to deal with nominal attributes (and to detect “hidden” data points)
- “Zoom-in” function

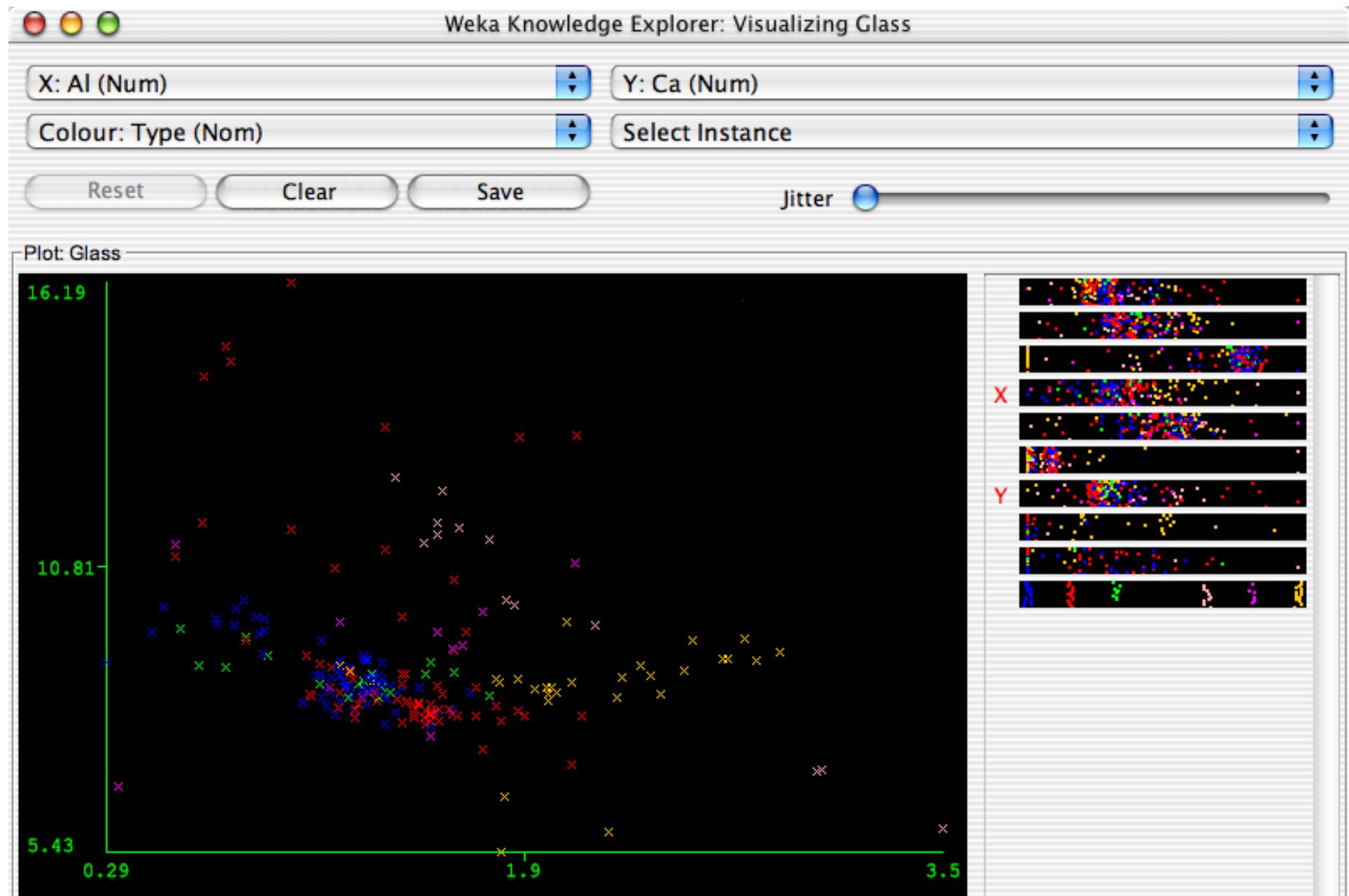




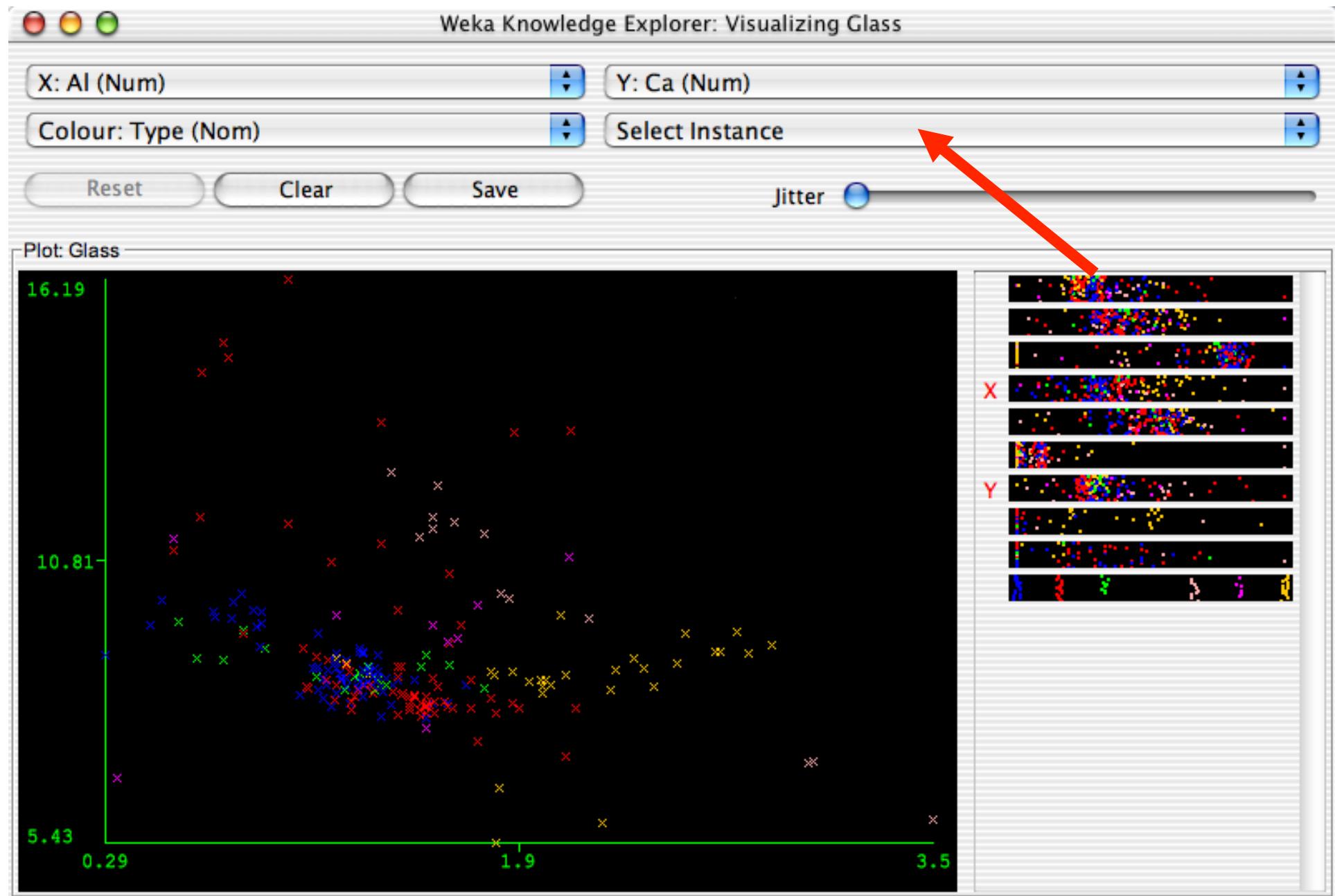








build	wind	float
vehic	wind	float
vehic	wind	non-float
	non-float	
tableware		



# Weka Knowledge Explorer: Visualizing Glass

X: Al (Num)

Y: Ca (Num)

Colour: Type (Nom)

Rectangle

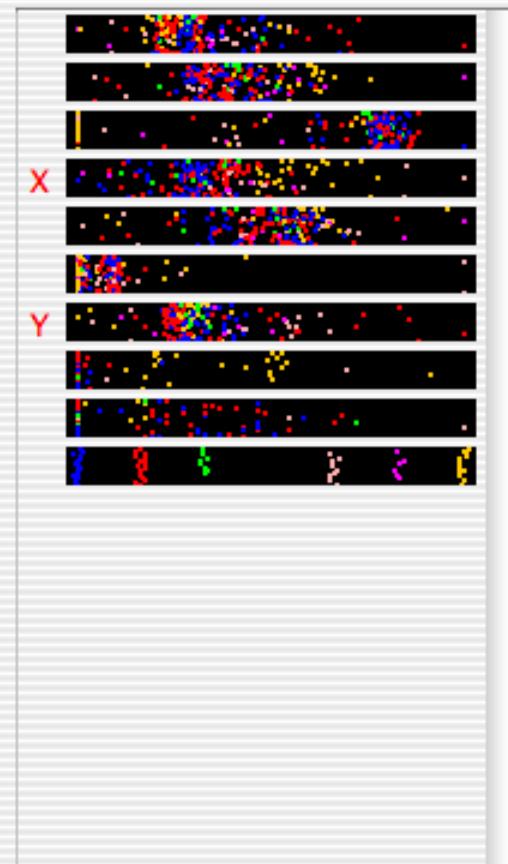
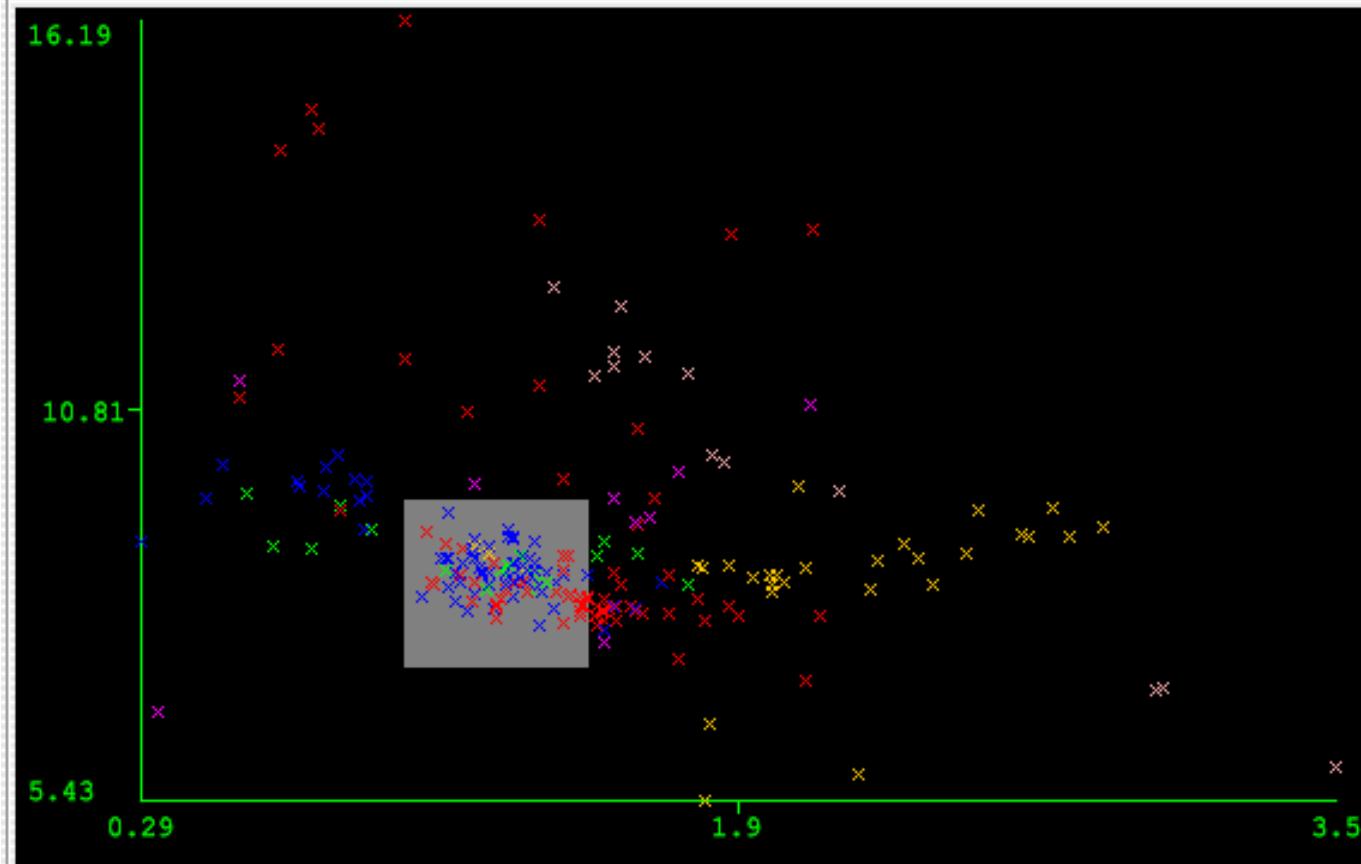
Submit

Clear

Save

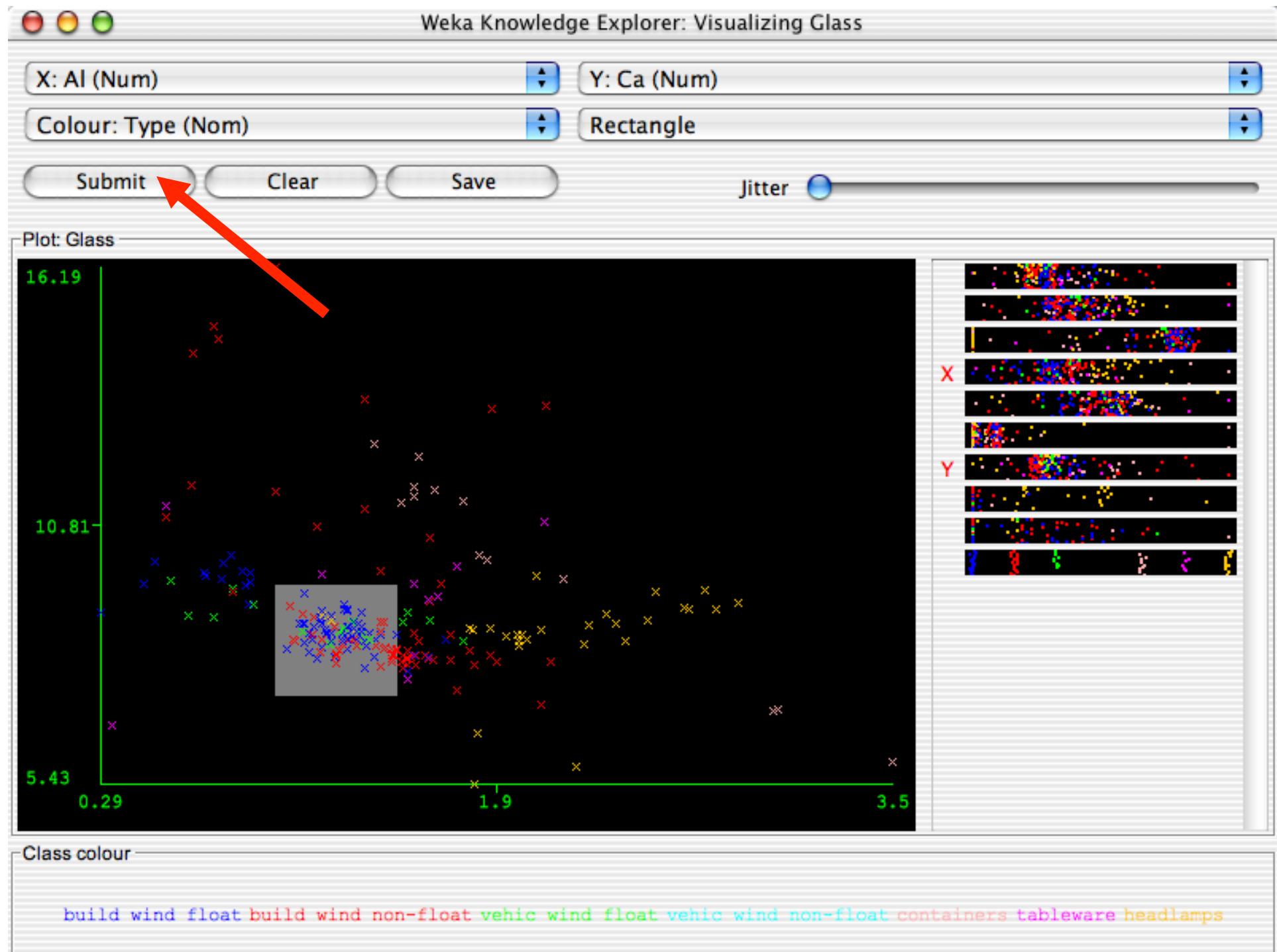
Jitter

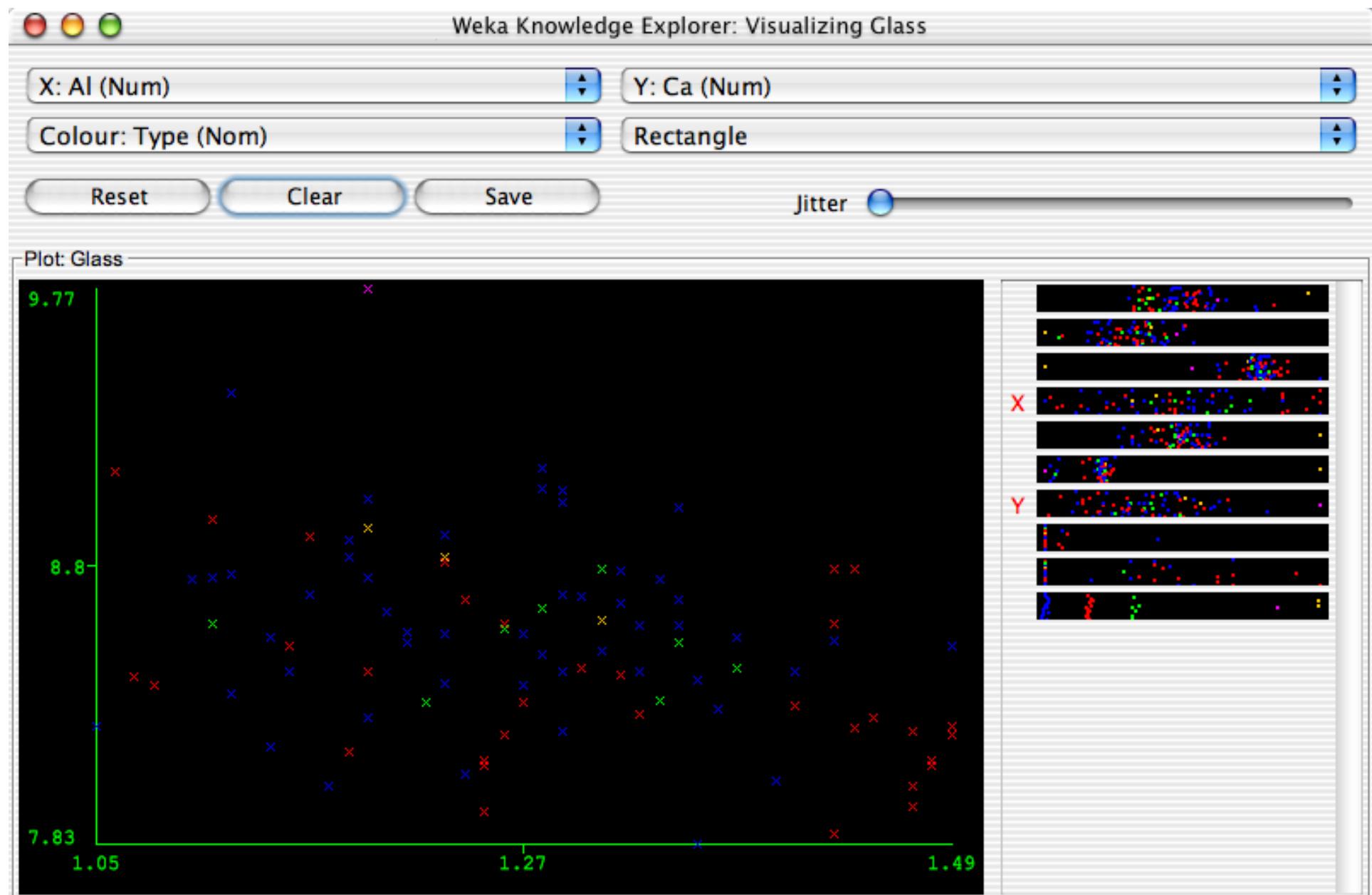
Plot: Glass



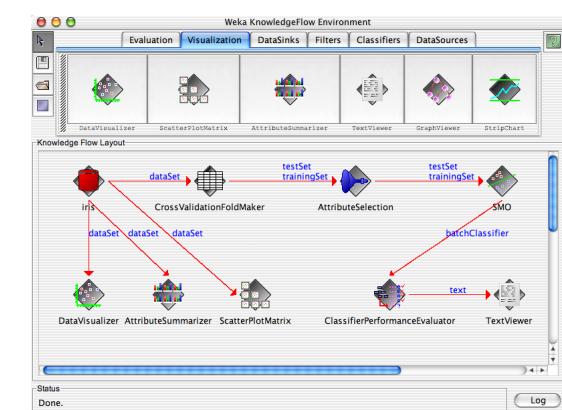
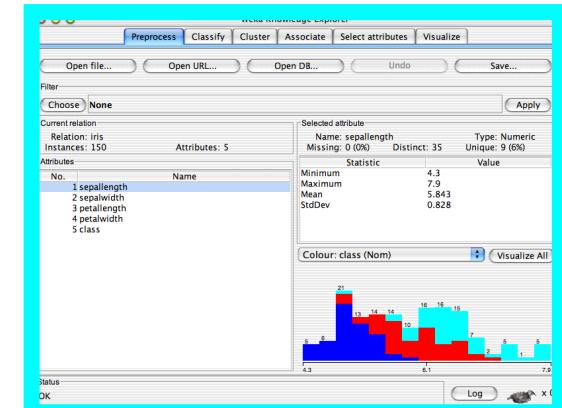
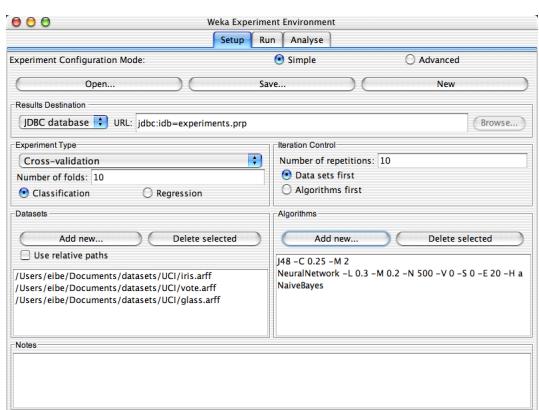
Class colour

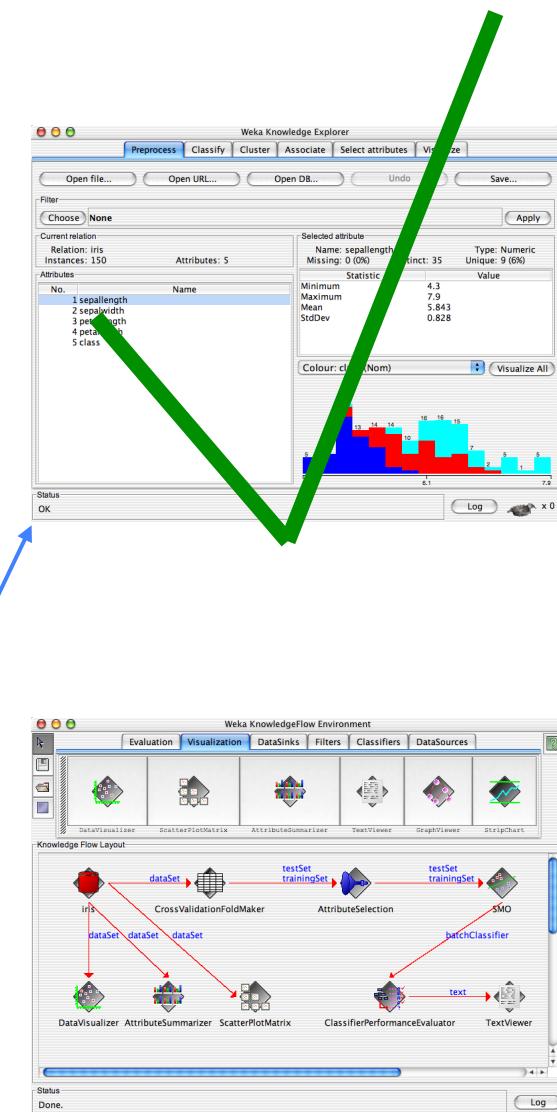
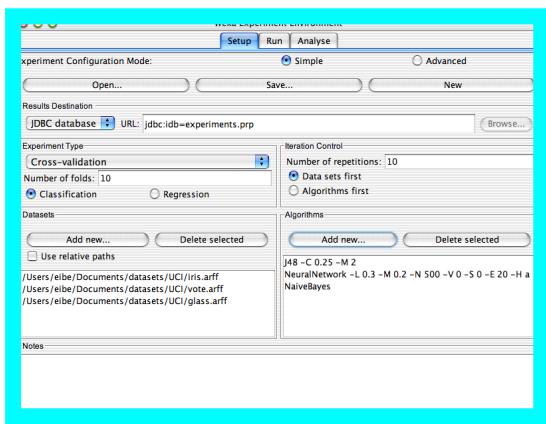
build wind float build wind non-float vehic wind float vehic wind non-float containers tableware headlamps





build wind float	build wind non-float	vehic wind float
vehic wind non-float	containers	tableware
headlamps		





# Performing experiments

- Experimenter makes it easy to compare the performance of different learning schemes
- For classification and regression problems
- Results can be written into file or database
- Evaluation options: cross-validation, learning curve, hold-out
- Can also iterate over different parameter settings
- Significance-testing built in!

Weka Experiment Environment

Setup Run Analyse

Experiment Configuration Mode:  Simple  Advanced

**Open...** **Save...** **New**

Results Destination  
JDBC database  **Filename:**

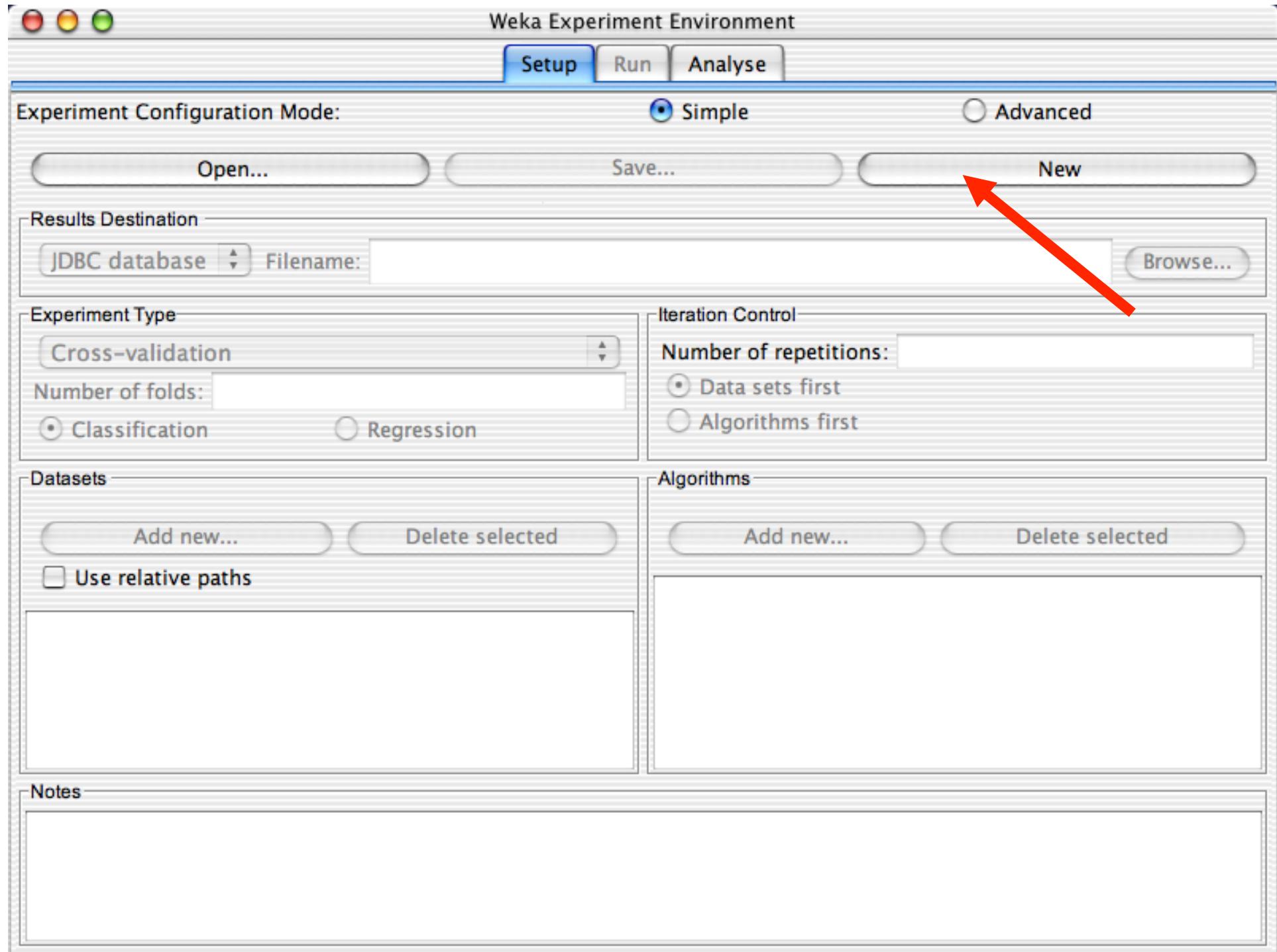
Experiment Type  
Cross-validation  
Number of folds:  
 Classification  Regression

Iteration Control  
Number of repetitions:  
 Data sets first  Algorithms first

Datasets  
Add new... Delete selected  
 Use relative paths

Algorithms  
Add new... Delete selected

Notes



Weka Experiment Environment

Setup Run Analyse

Experiment Configuration Mode:  Simple  Advanced

Open... Save... New

Results Destination

JDBC database URL: jdbc:idb=experiments.prp

Experiment Type

Cross-validation

Number of folds: 10

Classification  Regression

Iteration Control

Number of repetitions: 10

Data sets first  Algorithms first

Datasets

Add new... Delete selected

Use relative paths

/Users/eibe/Documents/datasets/UCI/iris.arff  
/Users/eibe/Documents/datasets/UCI/vote.arff  
/Users/eibe/Documents/datasets/UCI/glass.arff

Algorithms

Add new... Delete selected

J48 -C 0.25 -M 2  
NeuralNetwork -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a  
NaiveBayes

Notes

Weka Experiment Environment

Setup Run **Analyse**

Experiment Configuration Mode:  Simple  Advanced

Open... Save... New

Results Destination

JDBC database URL: jdbc:idb=experiments.prp

Experiment Type

Cross-validation

Number of folds: 10

Classification  Regression

Iteration Control

Number of repetitions: 10

Data sets first  Algorithms first

Datasets

Add new... Delete selected

Use relative paths

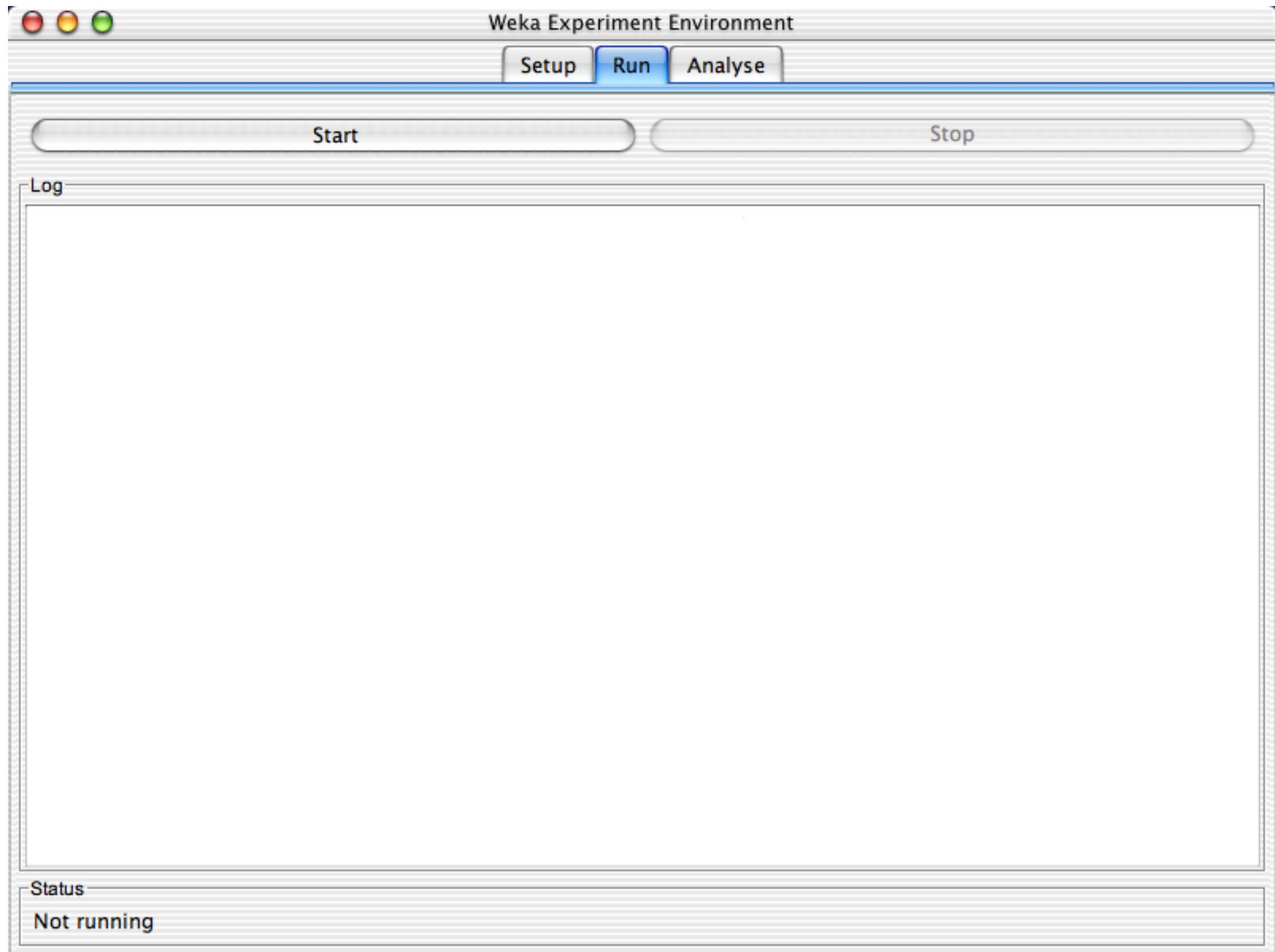
/Users/eibe/Documents/datasets/UCI/iris.arff  
/Users/eibe/Documents/datasets/UCI/vote.arff  
/Users/eibe/Documents/datasets/UCI/glass.arff

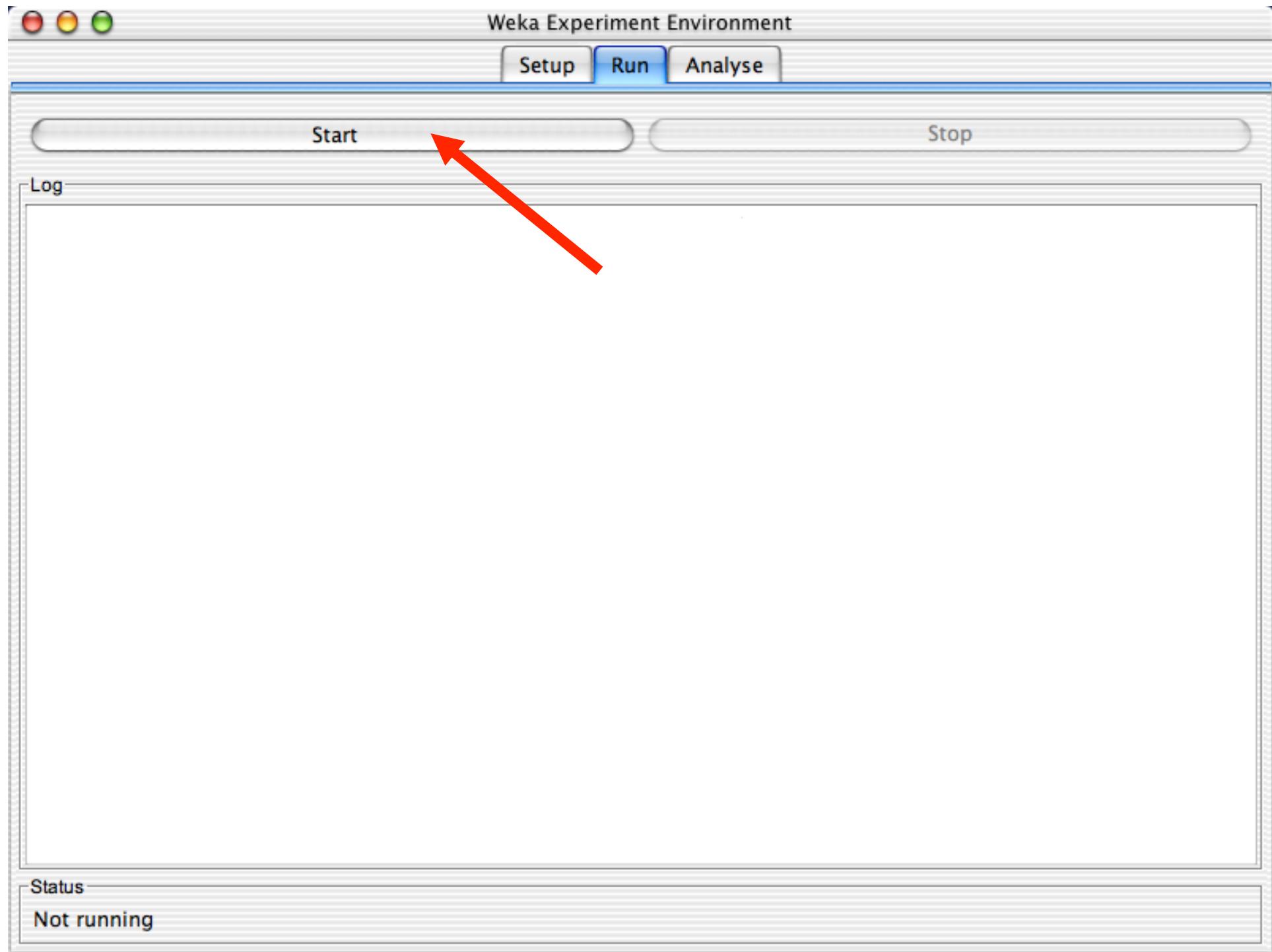
Algorithms

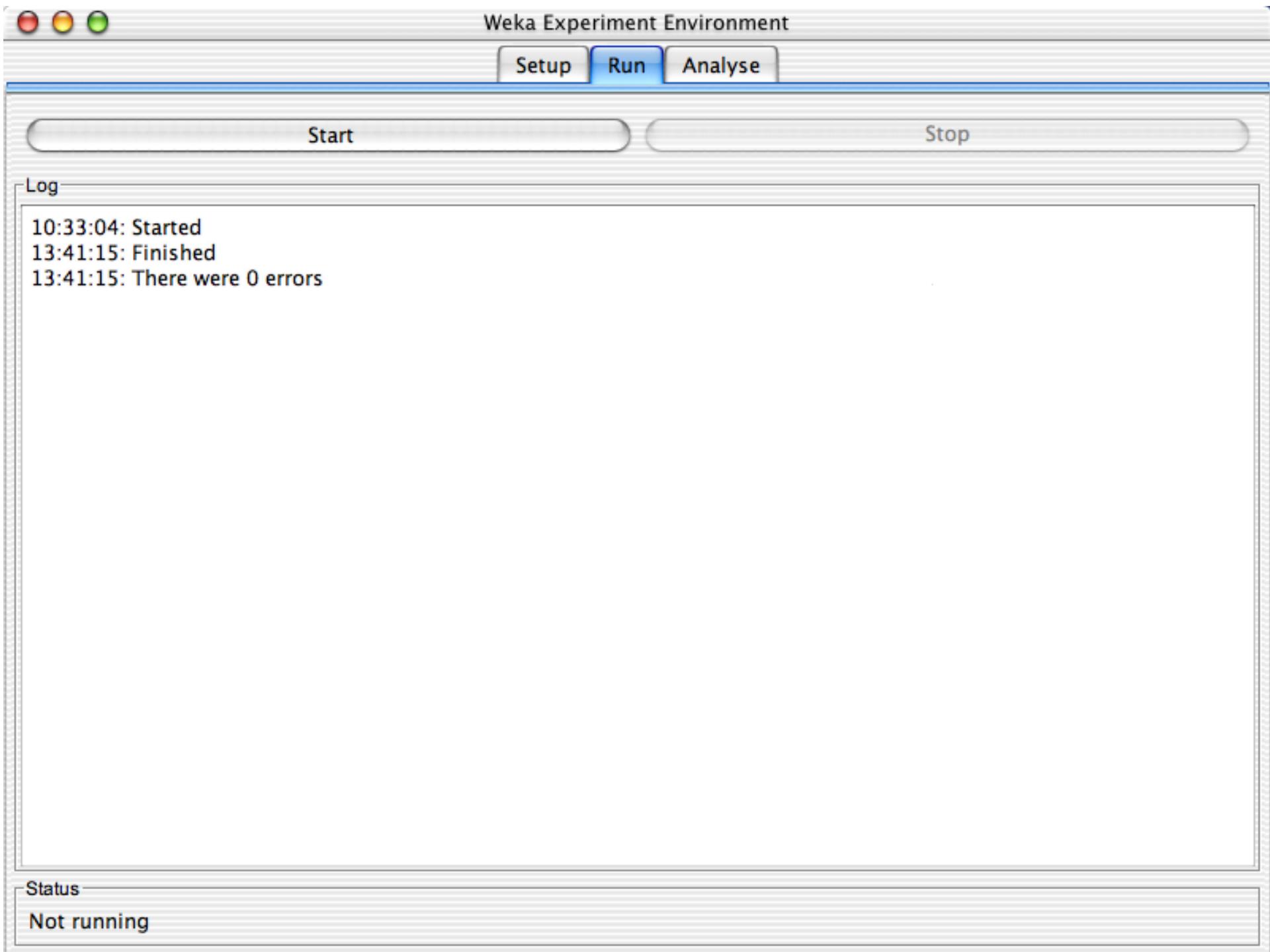
Add new... Delete selected

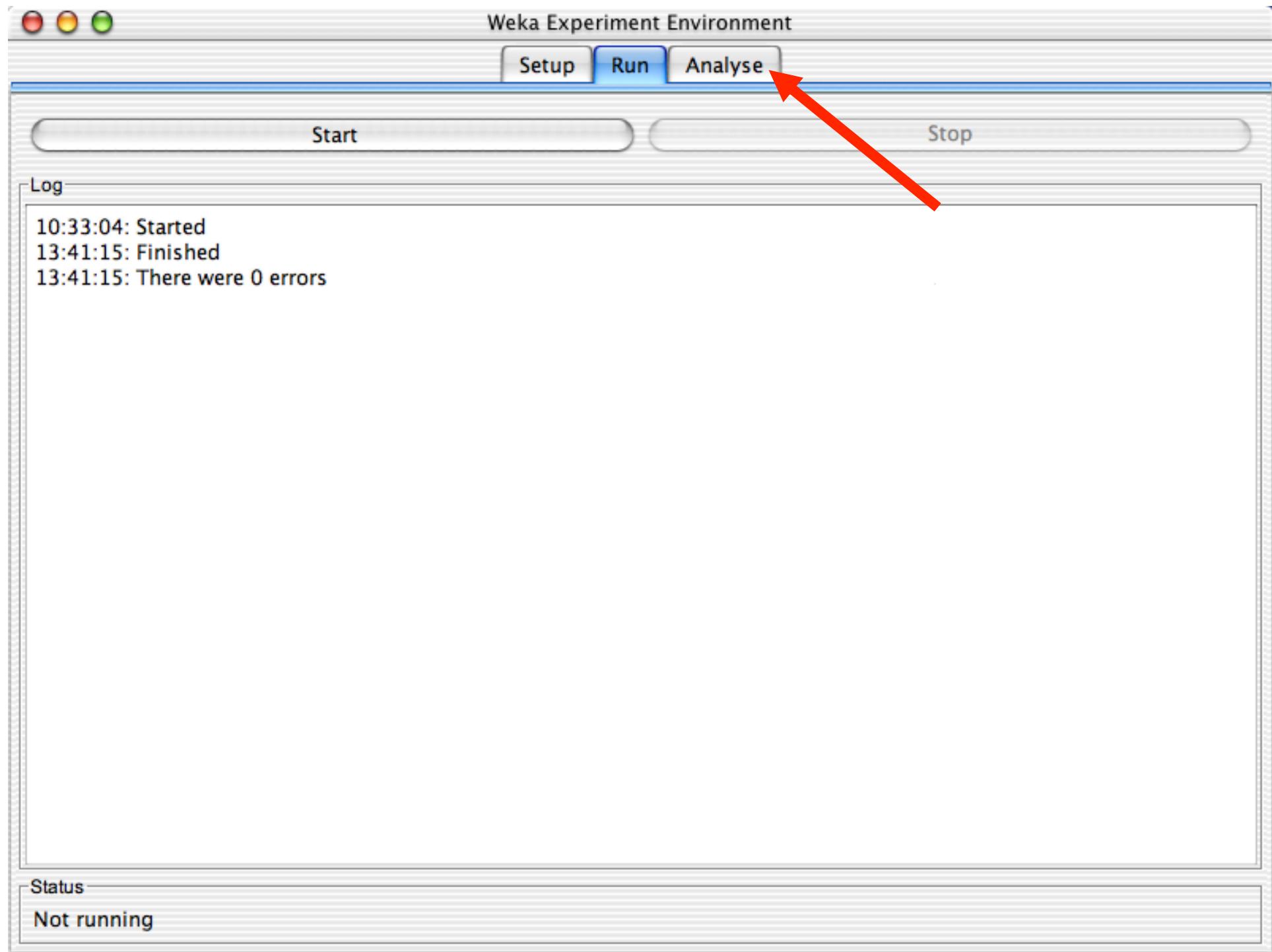
J48 -C 0.25 -M 2  
NeuralNetwork -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a  
NaiveBayes

Notes









Weka Experiment Environment

Setup Run Analyse

Source

No source File... Database... Experiment

Configure test Test output

Row key fields Select keys...

Run field

Column key fields Select keys...

Comparison field

Significance 0.05

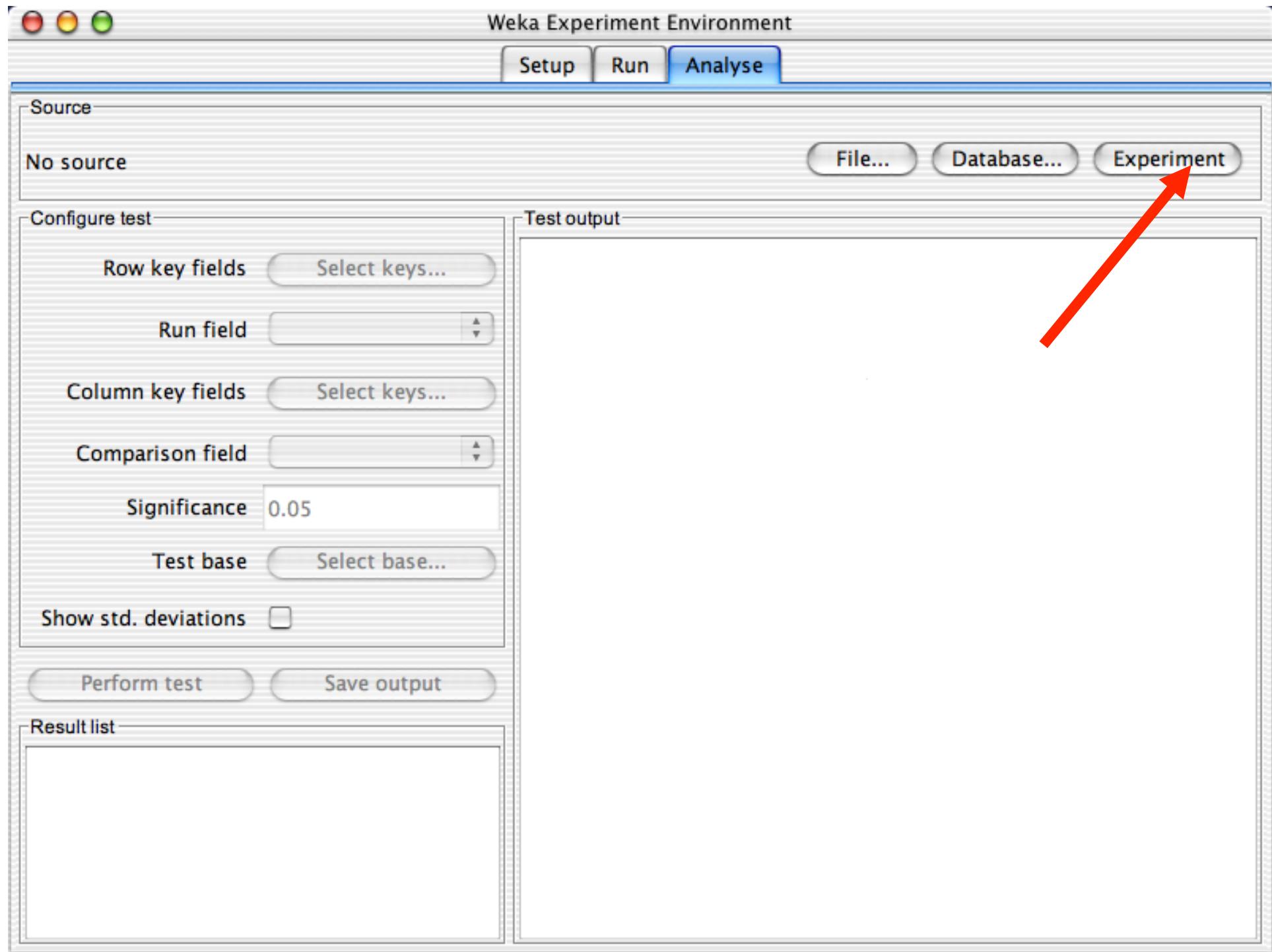
Test base Select base...

Show std. deviations

Perform test Save output

Result list

The screenshot shows the Weka Experiment Environment window. At the top, there are three colored window control buttons (red, yellow, green). Below them is the title "Weka Experiment Environment". A horizontal menu bar contains three tabs: "Setup" (disabled), "Run" (disabled), and "Analyse" (selected). On the left side, a vertical panel titled "Source" displays the message "No source" and includes three buttons: "File...", "Database...", and "Experiment". Below this is a section titled "Configure test" which contains several configuration fields: "Row key fields" with a "Select keys..." button, "Run field" with a dropdown menu, "Column key fields" with a "Select keys..." button, "Comparison field" with a dropdown menu, "Significance" set to "0.05", "Test base" with a "Select base..." button, and a checkbox for "Show std. deviations" which is unchecked. At the bottom of this panel are two buttons: "Perform test" and "Save output". To the right of the configuration panel is a large, empty rectangular area titled "Test output". At the very bottom left, there is a small section titled "Result list" with a corresponding empty rectangular area.



Weka Experiment Environment

Setup Run Analyse

Source

Got 900 results

File... Database... Experiment

Configure test

Row key fields

Run field

Column key fields

Comparison field

Significance 0.05

Test base

Show std. deviations

Perform test

Result list

13:44:17 - Available resultsets

13:44:55 - Percent\_correct - trees.j48.J48 '-C 0.25 -M 2' -2177331683936444444

Test output

Analysing: Percent\_correct  
Datasets: 3  
Resultsets: 3  
Confidence: 0.05 (two tailed)  
Date: 9/9/03 1:44 PM

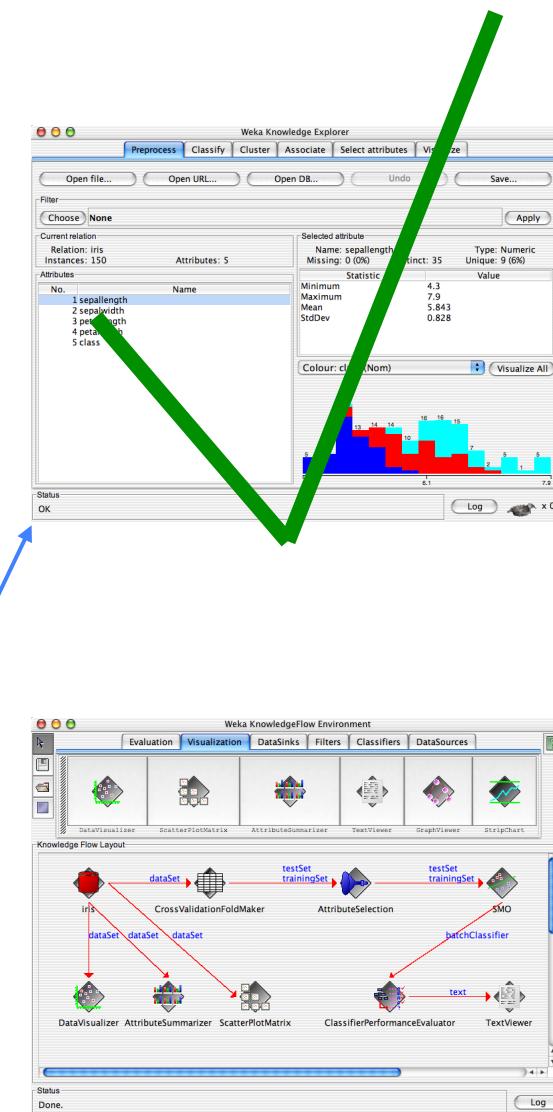
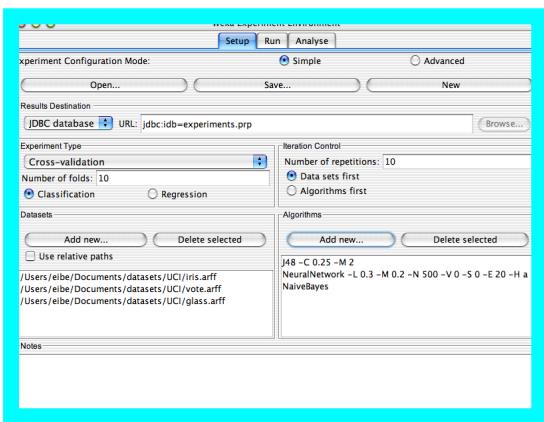
Dataset	(1) trees.j4		(2) funct	(3) bayes	
iris	(100)	94.73		96.4	95.53
vote	(100)	96.57		94.71 *	90.02 *
Glass	(100)	67.63		66.78	49.45 *

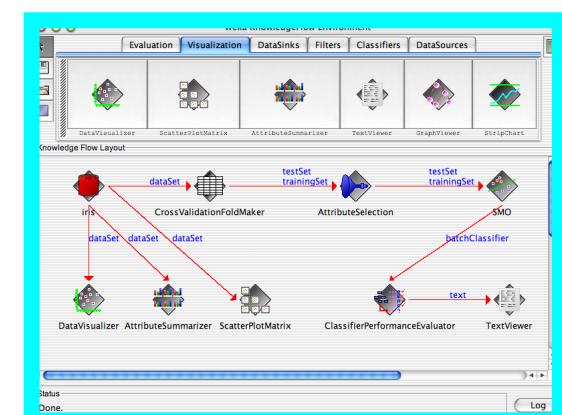
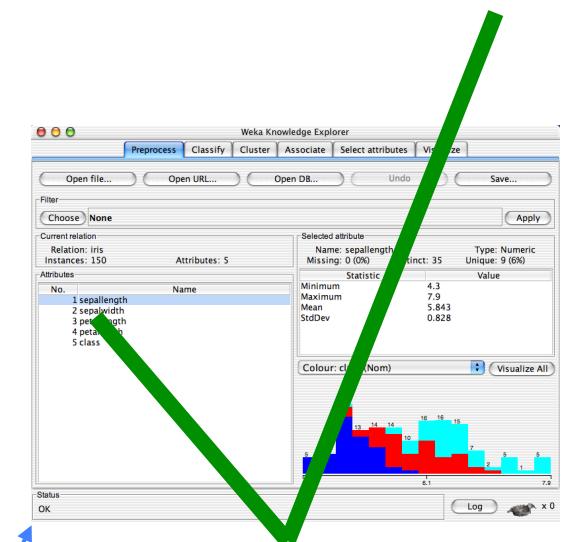
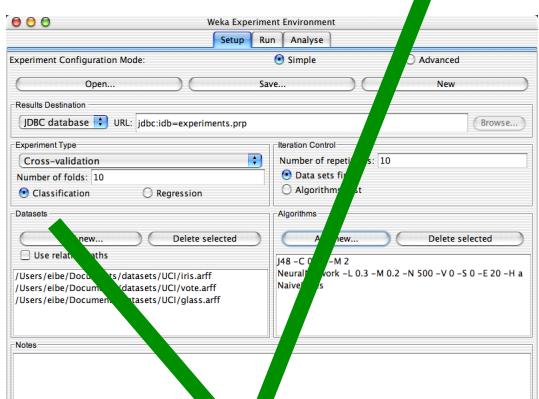
(v/ /\*) | (0/2/1) (0/1/2)

Skipped:

Key:

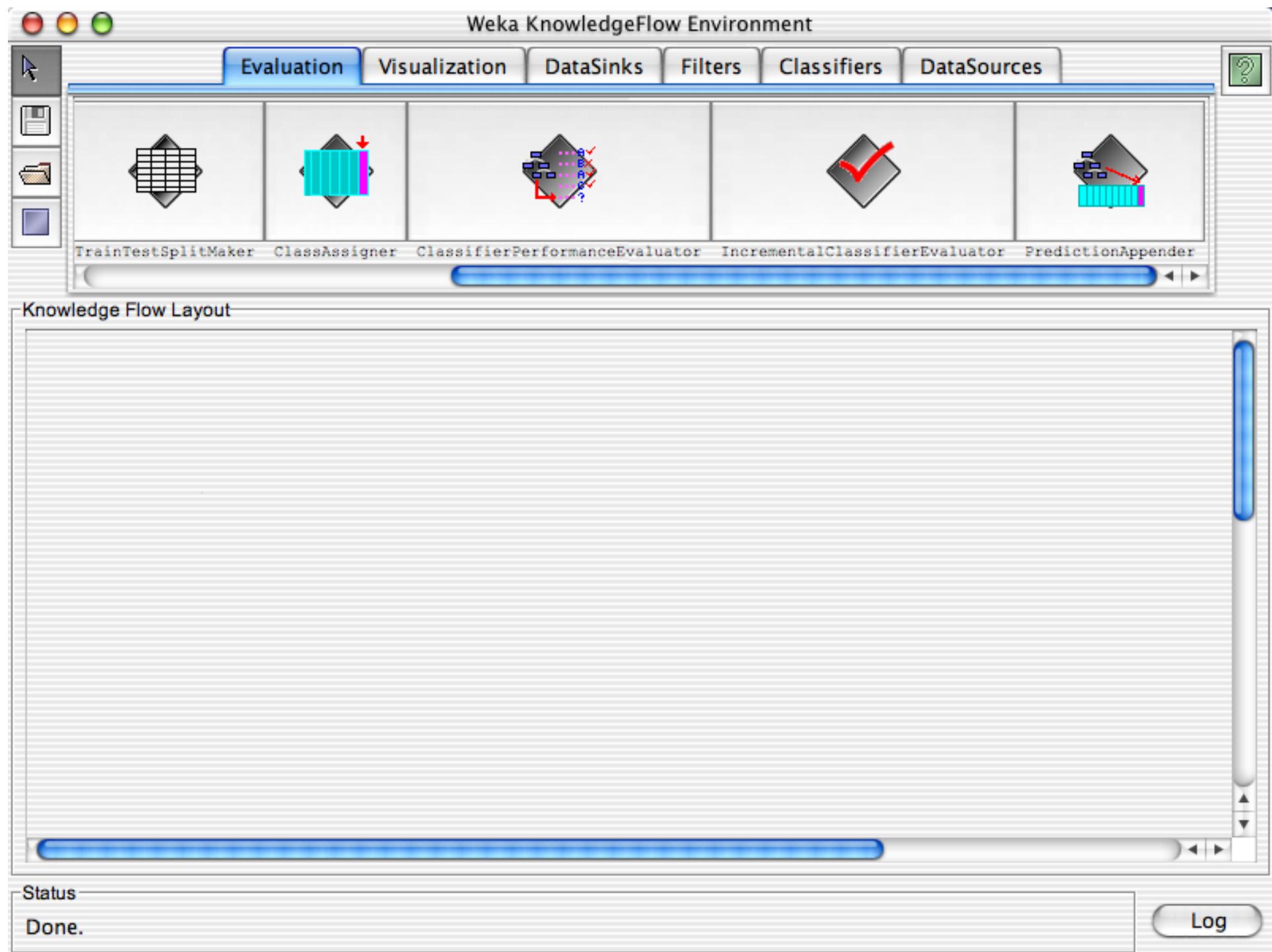
(1) trees.j48.J48 '-C 0.25 -M 2' -2177331683936444444  
(2) functions.neural.NeuralNetwork '-L 0.3 -M 0.2 -N 500 -V 0 -S 0  
(3) bayes.NaiveBayes '' 2029074699749330519

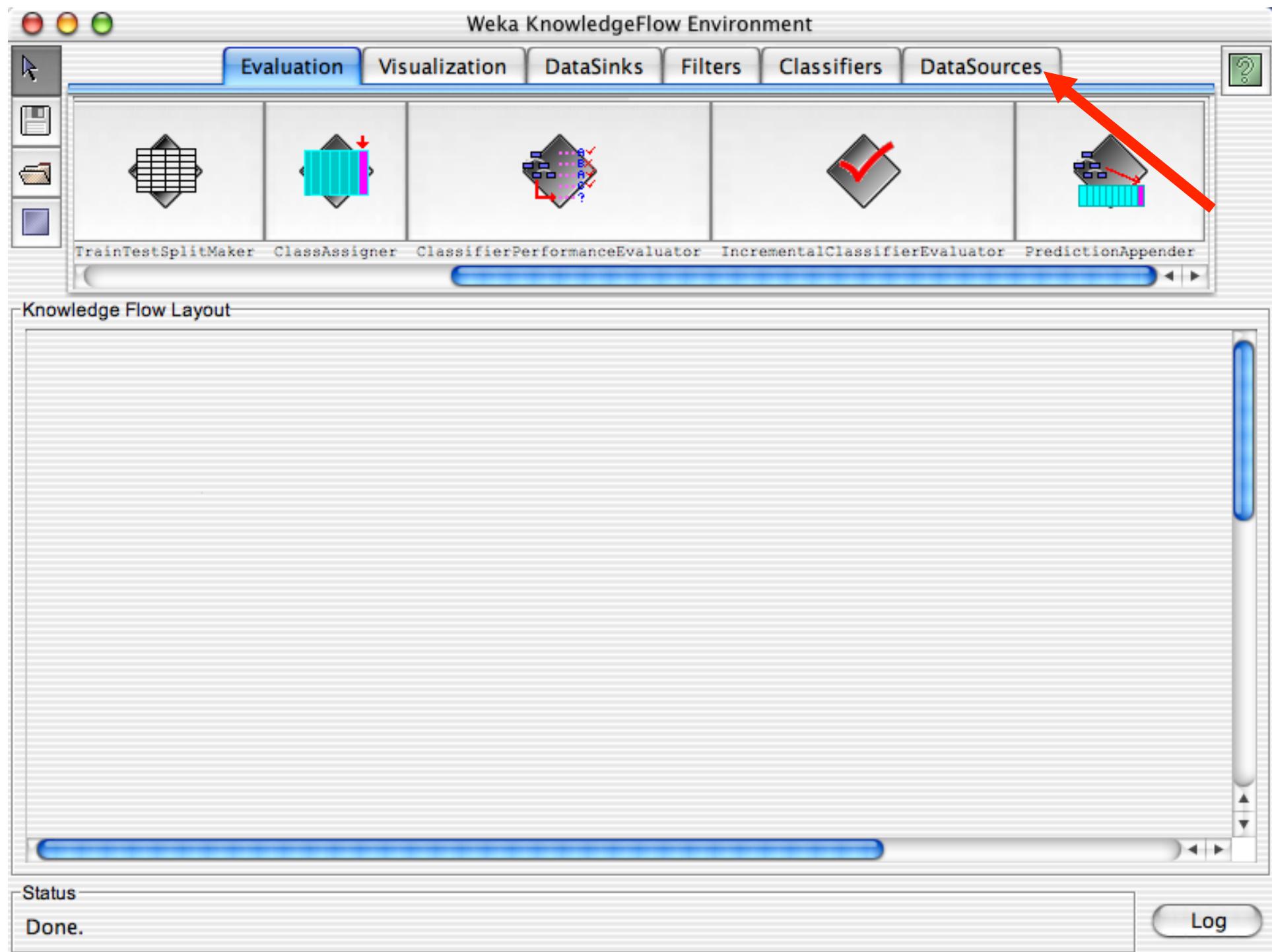




# The Knowledge Flow GUI

- New graphical user interface for WEKA
- Java-Beans-based interface for setting up and running machine learning experiments
- Data sources, classifiers, etc. are beans and can be connected graphically
- Data “flows” through components: e.g., “data source” -> “filter” -> “classifier” -> “evaluator”
- Layouts can be saved and loaded again later





Weka KnowledgeFlow Environment

Evaluation   Visualization   DataSinks   Filters   Classifiers   **DataSources**

ArffLoader   CSVLoader   C45Loader   SerializedInstancesLoader

Knowledge Flow Layout

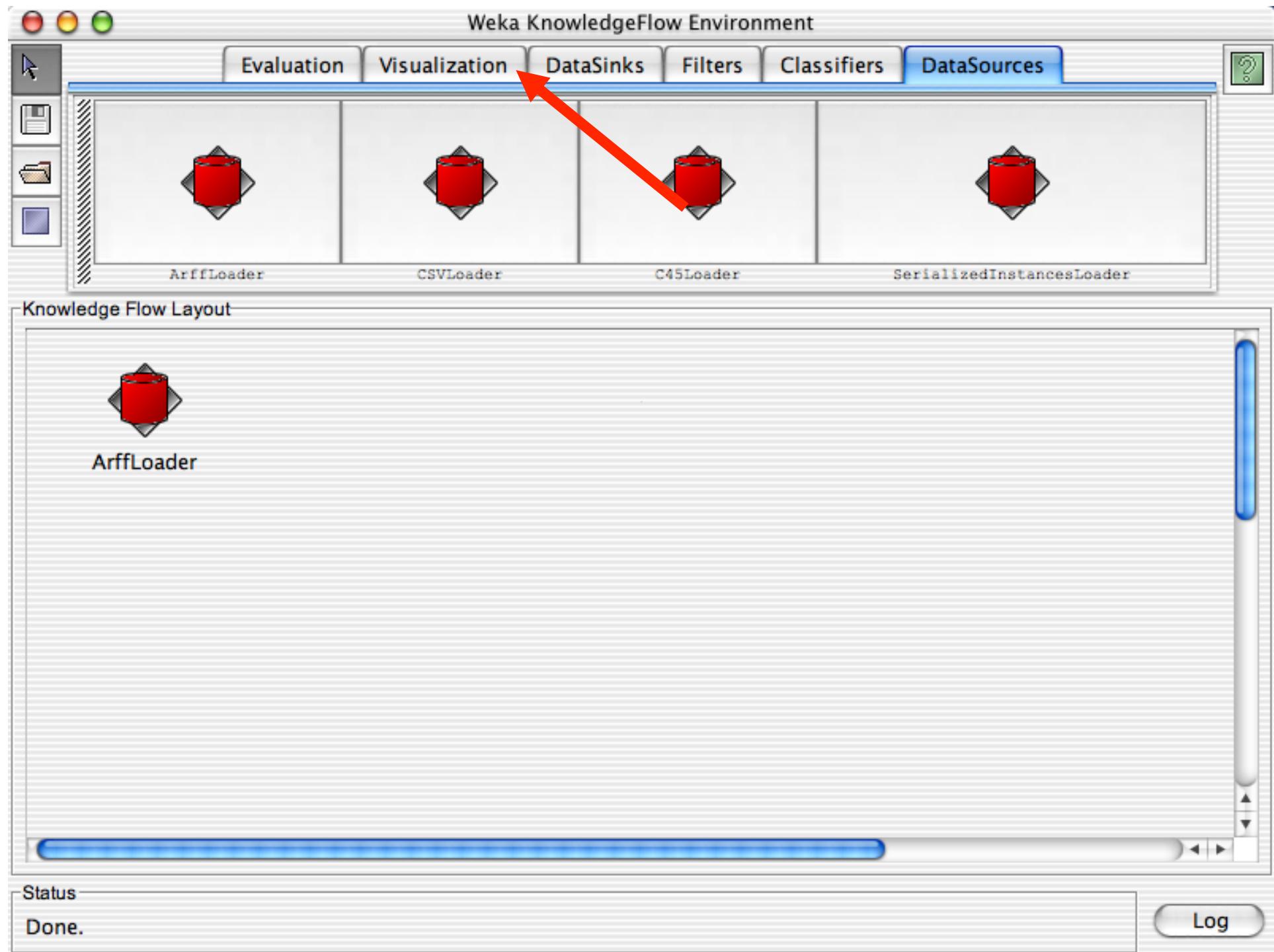
ArffLoader

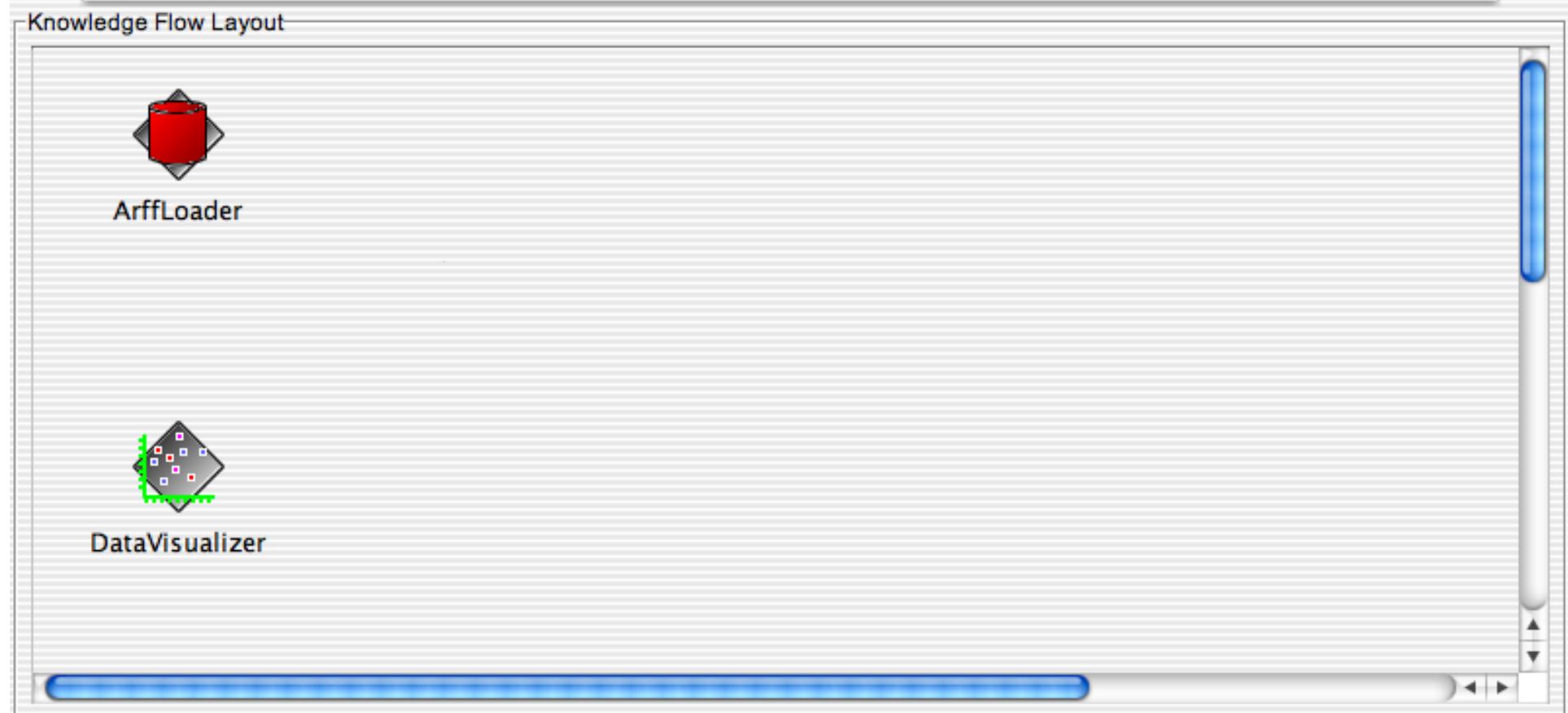
Status

Done.

Log

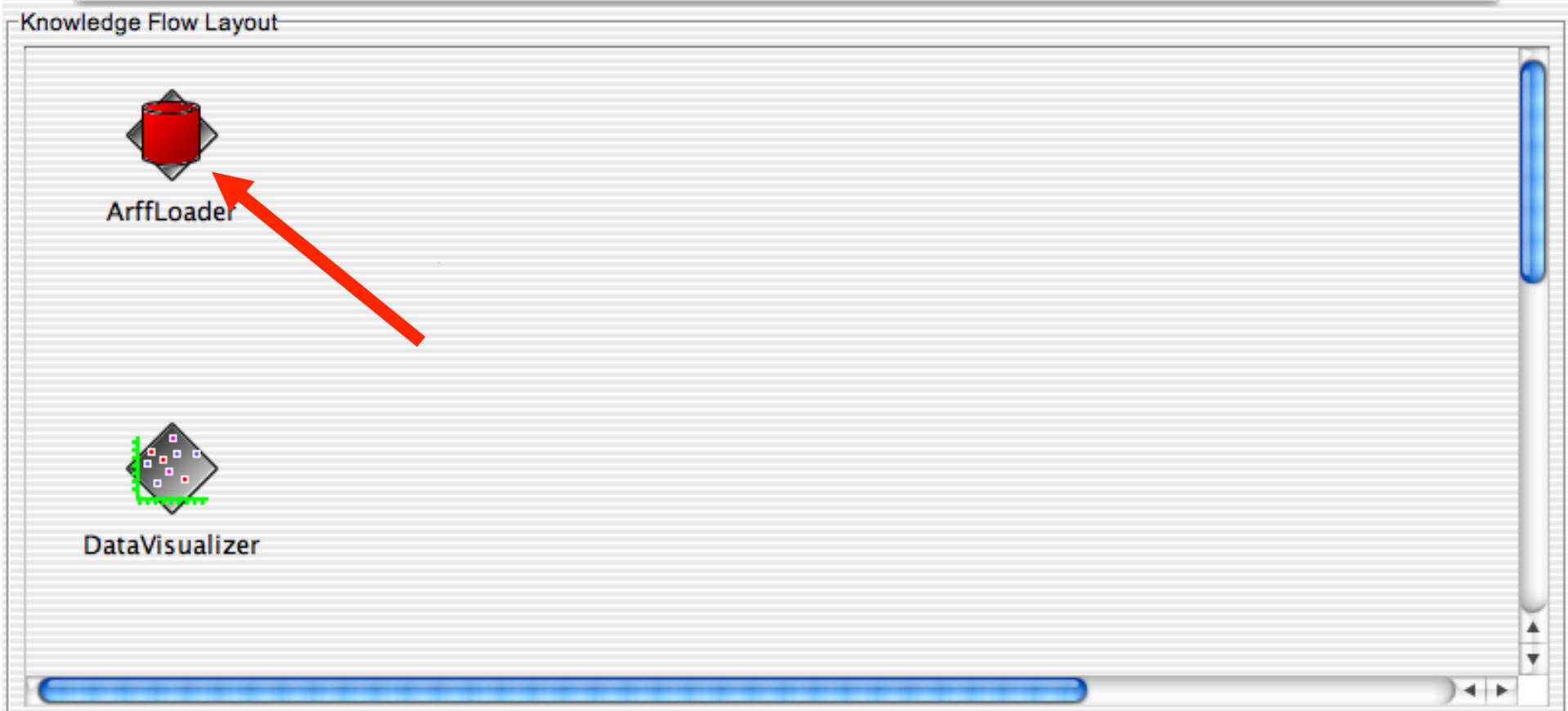
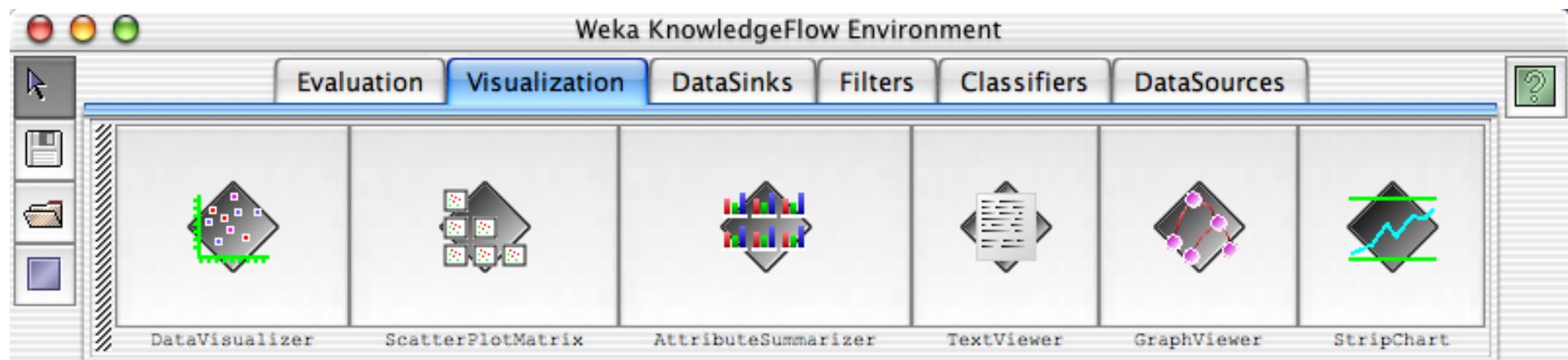
The screenshot shows the Weka KnowledgeFlow Environment interface. At the top, there are three colored window control buttons (red, yellow, green) and a title bar with the text "Weka KnowledgeFlow Environment". Below the title bar is a menu bar with tabs: Evaluation, Visualization, DataSinks, Filters, Classifiers, and DataSources. The DataSources tab is currently selected, indicated by a blue border around it. To the right of the tabs is a question mark icon. On the left side of the main workspace, there is a vertical toolbar with icons for a cursor, a floppy disk, a folder, and a square. Below the toolbar, four red cylindrical components with black diamond bases are arranged horizontally, each labeled with its name: ArffLoader, CSVLoader, C45Loader, and SerializedInstancesLoader. In the main workspace below, there is a large rectangular area titled "Knowledge Flow Layout" containing a single instance of the ArffLoader component. A vertical scroll bar is located on the right side of this layout area. At the bottom of the screen, there is a status bar with the text "Status" and "Done." followed by a "Log" button. A horizontal progress bar is also visible at the bottom.





Status: Done.

Log



Status  
Done.

Log



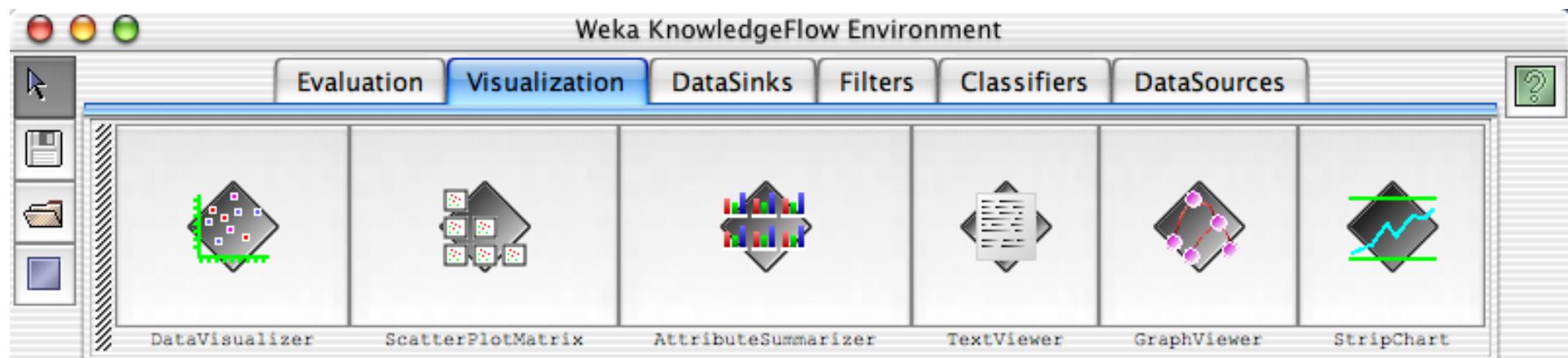
Knowledge Flow Layout



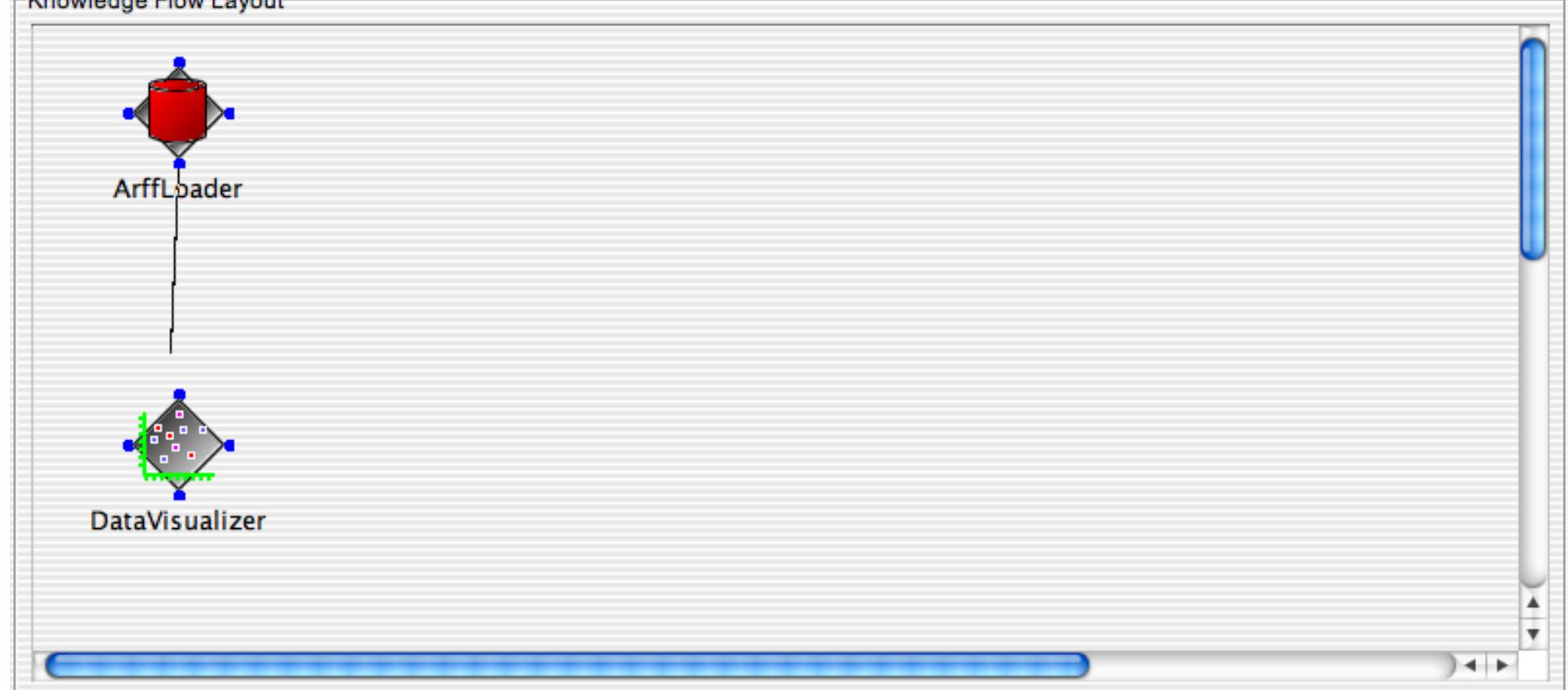
Status

Done.

Log



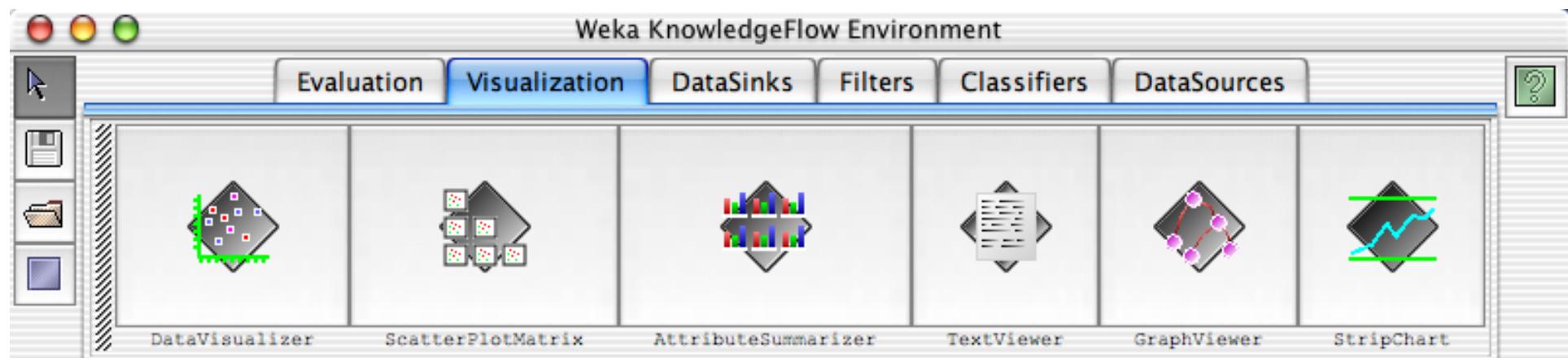
Knowledge Flow Layout



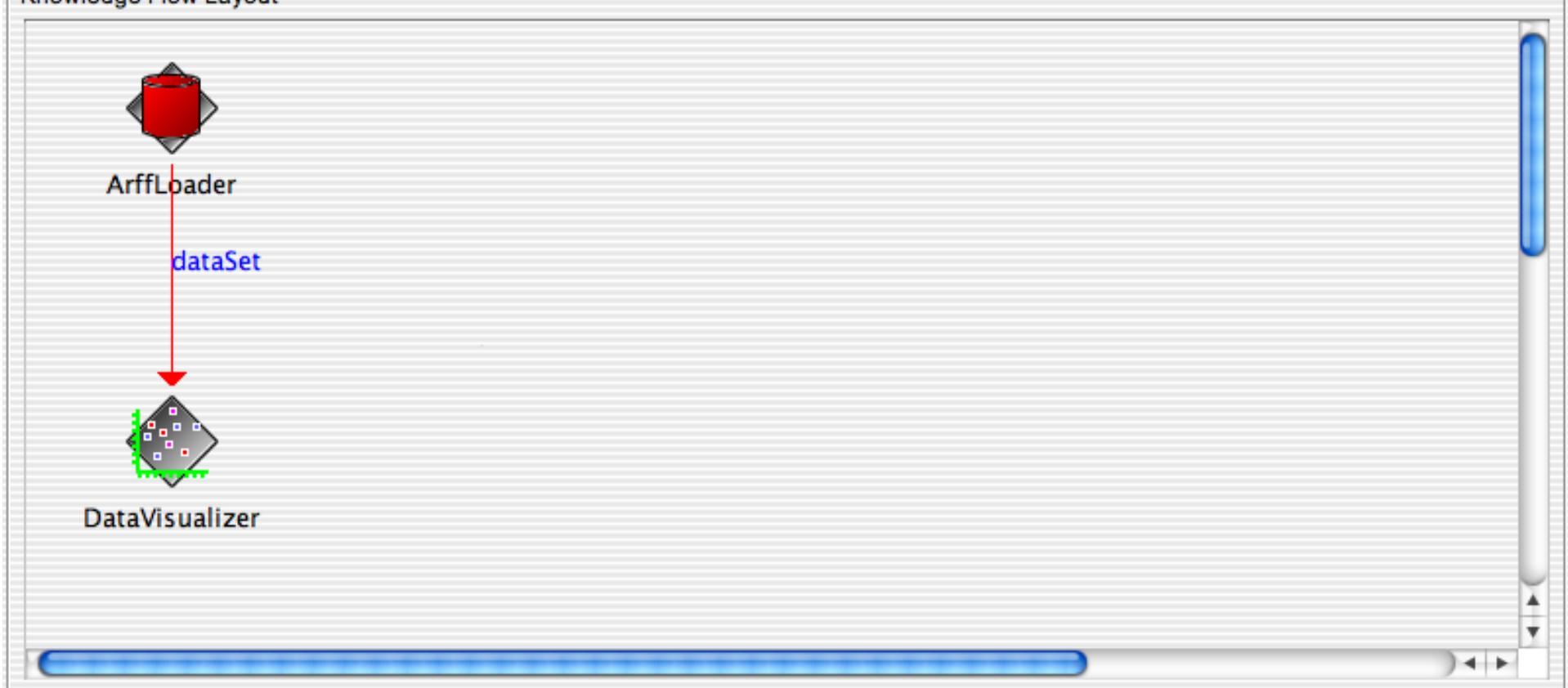
Status

Done.

Log



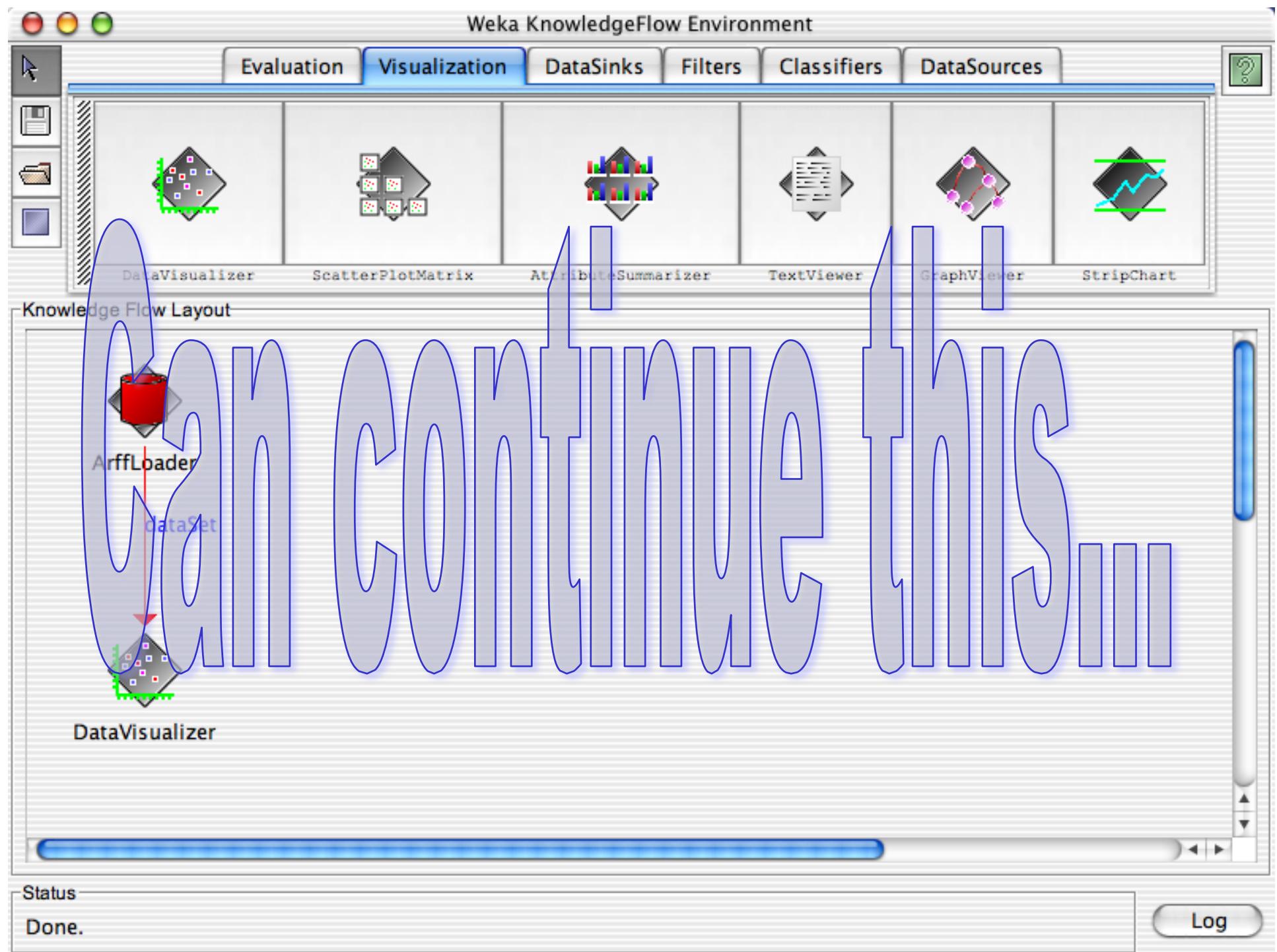
Knowledge Flow Layout

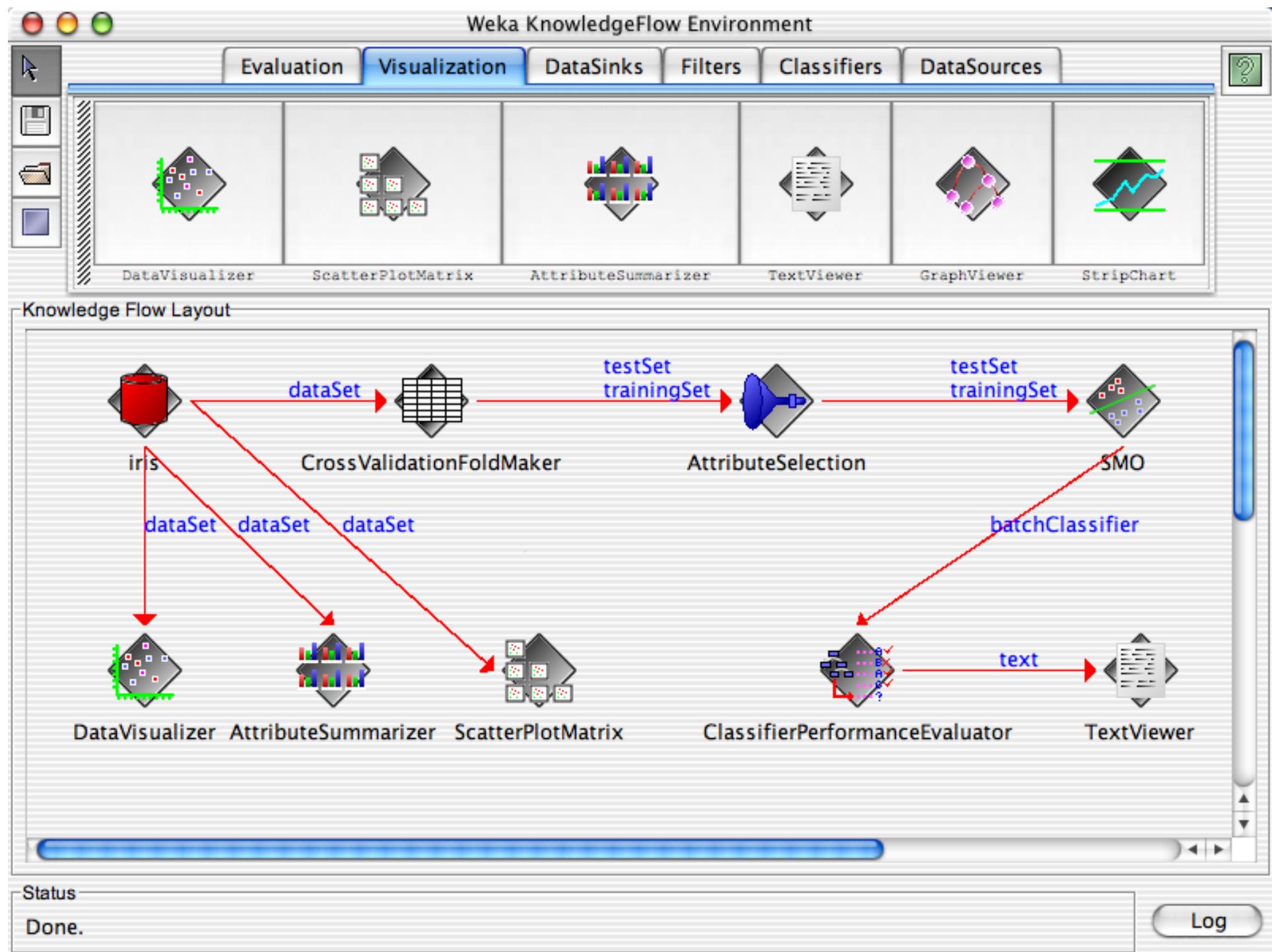


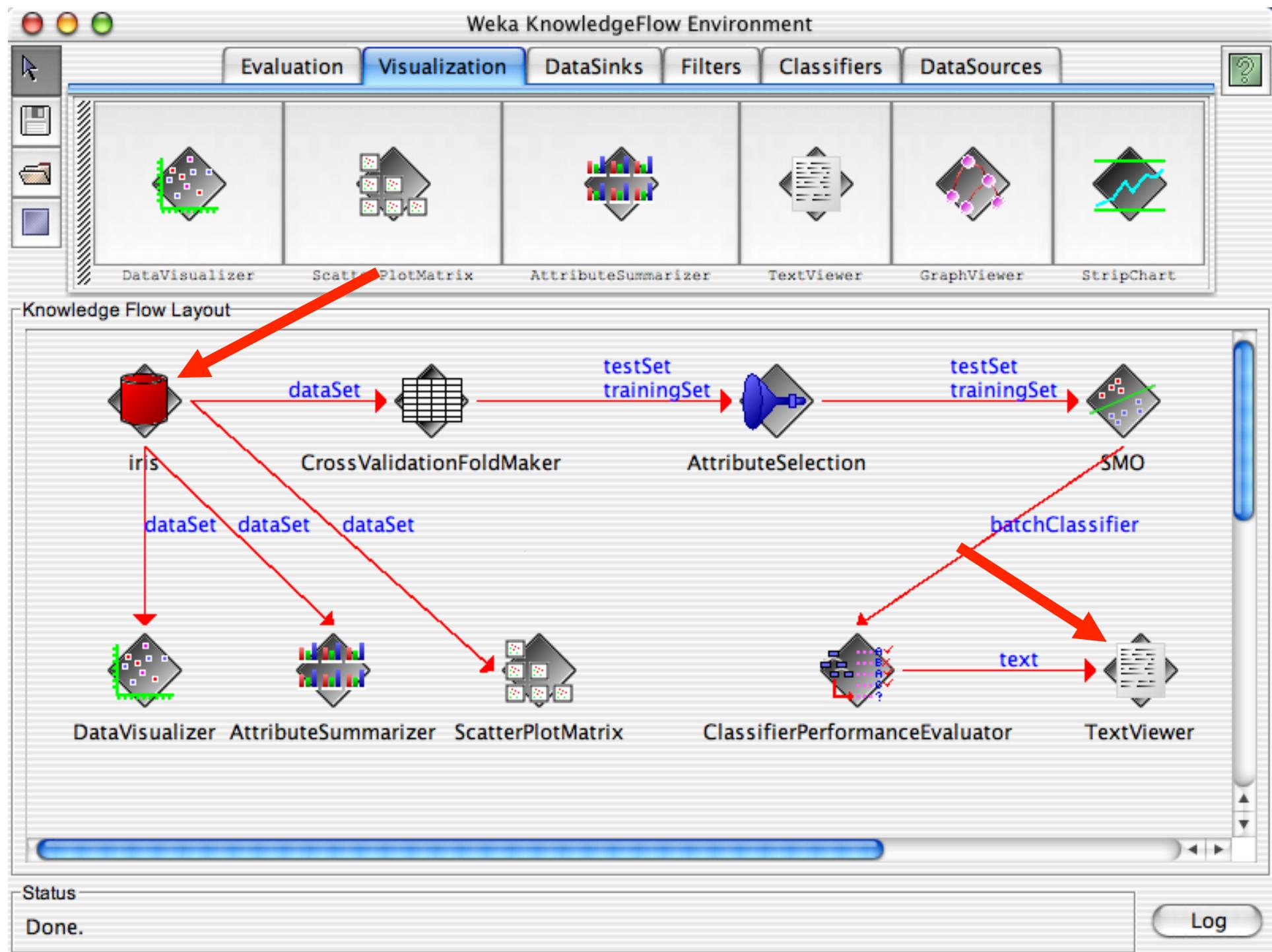
Status

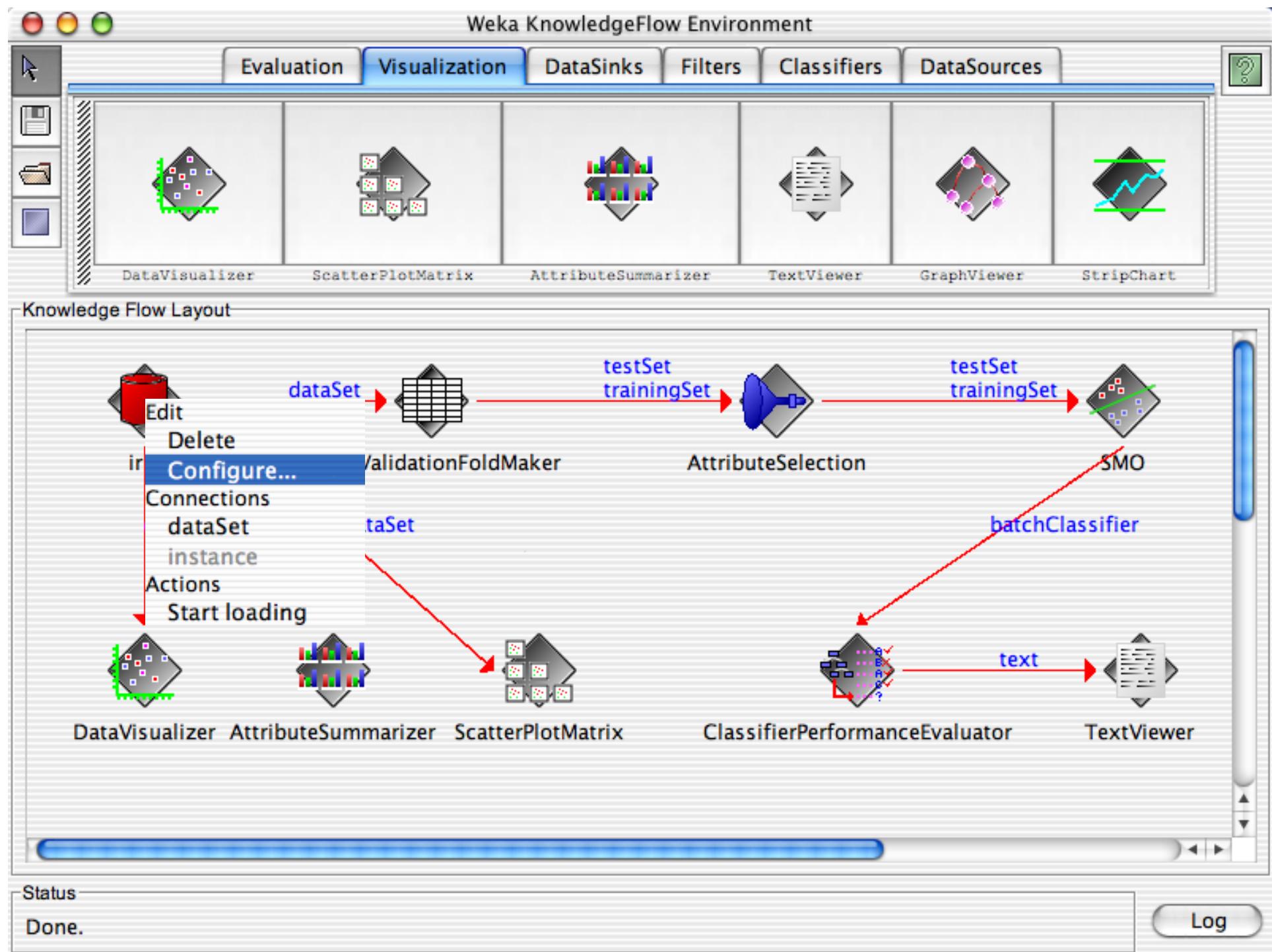
Done.

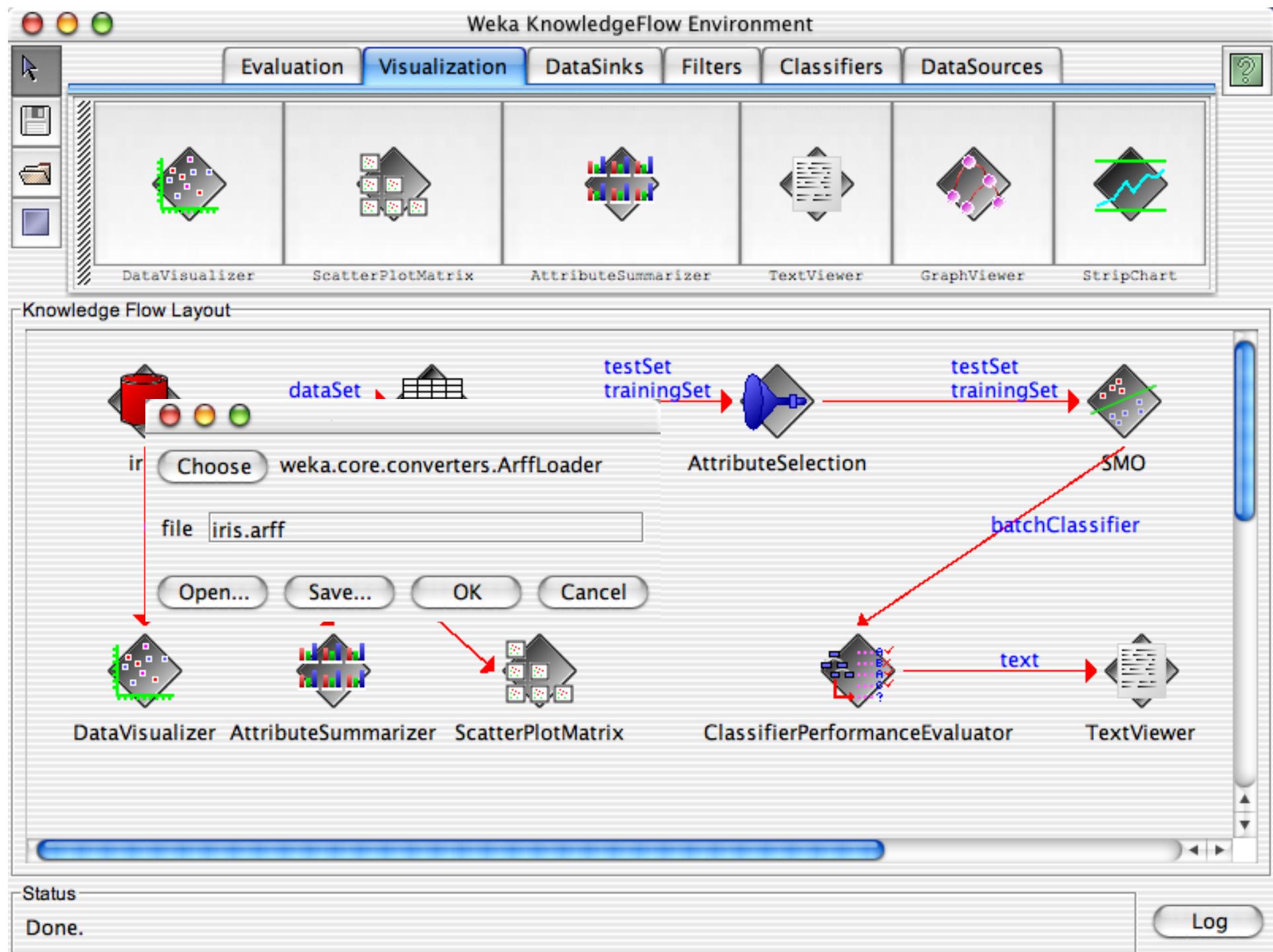
Log

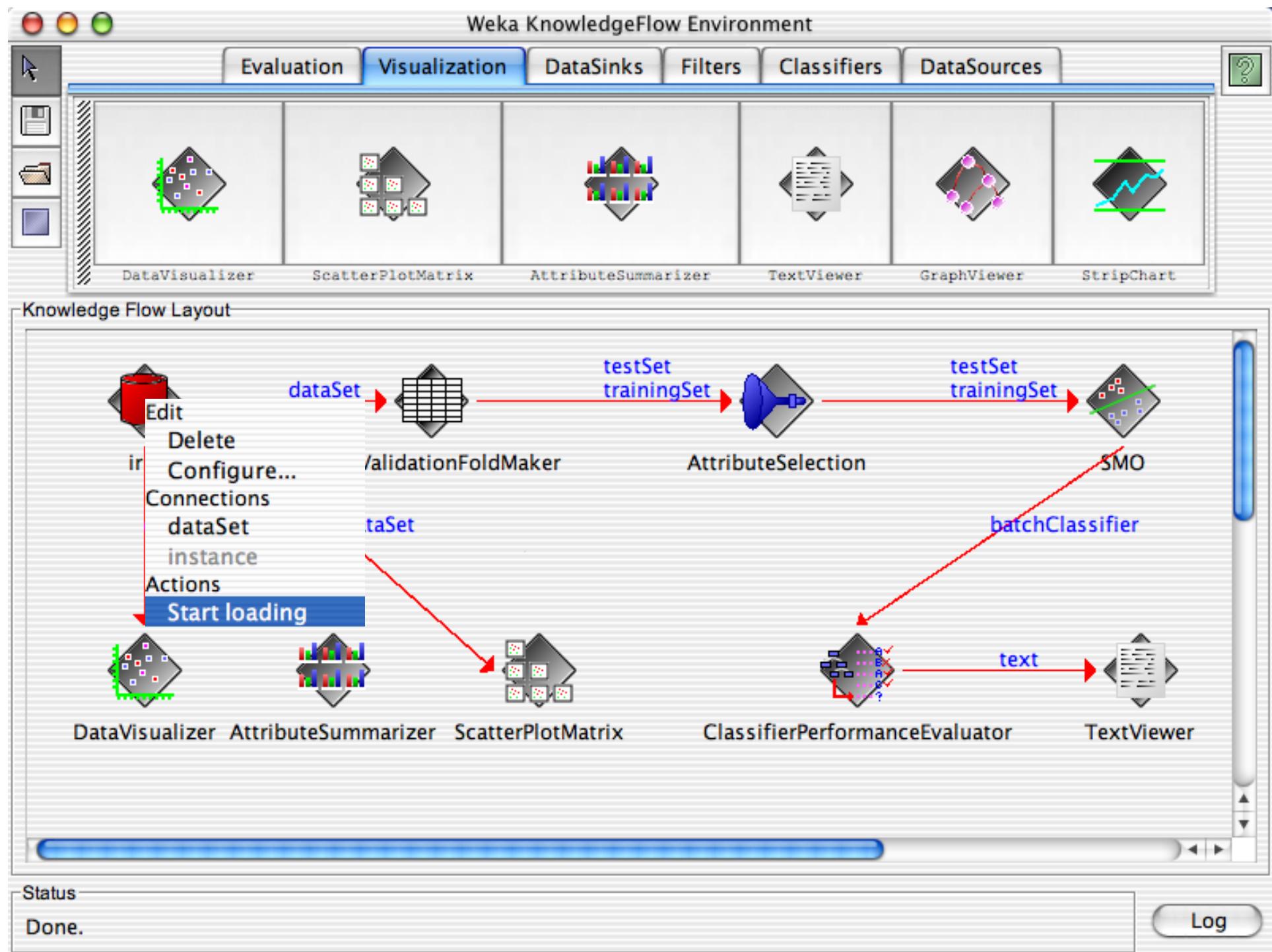


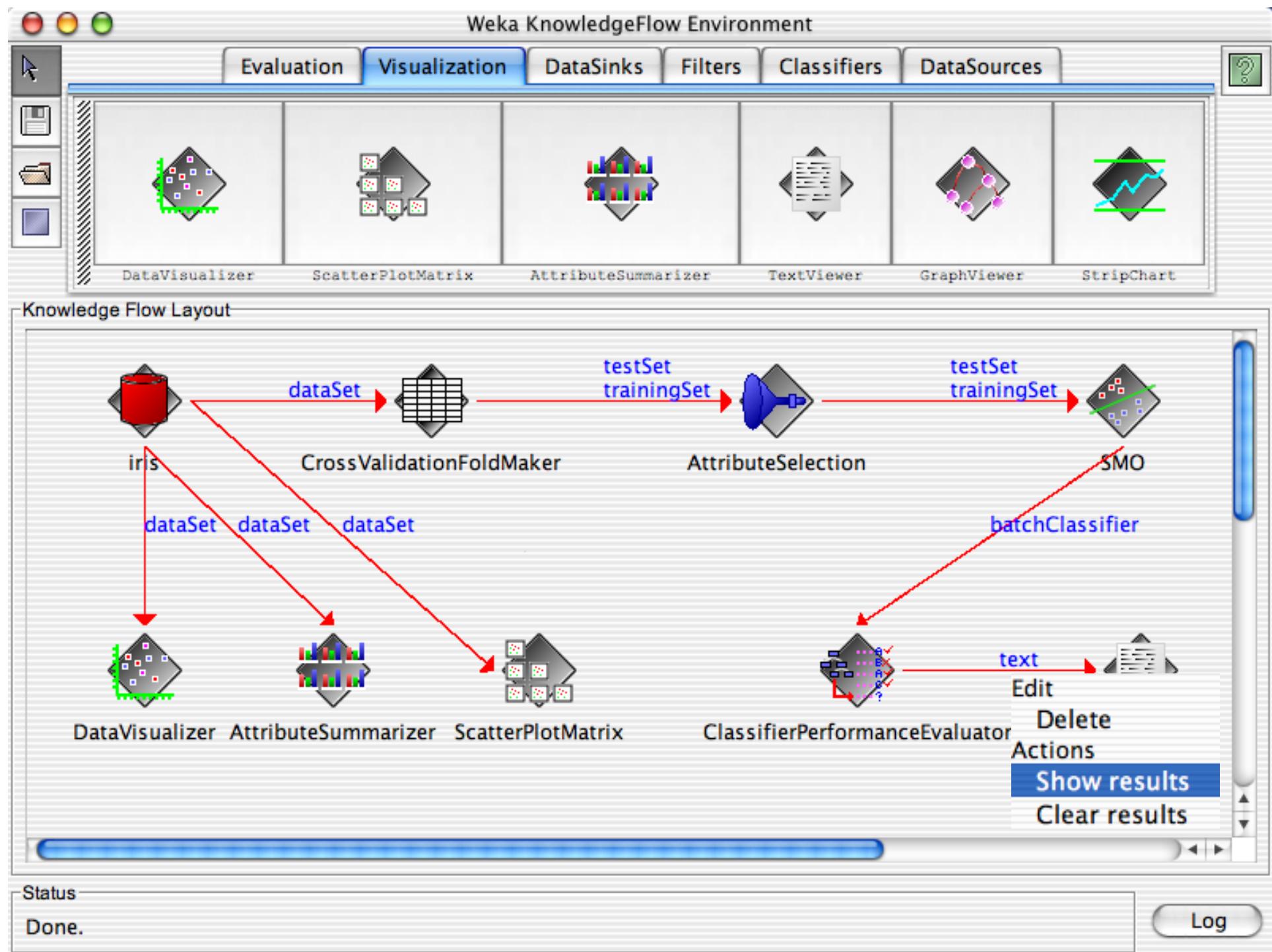














Knowledge Flow Layout

Result list

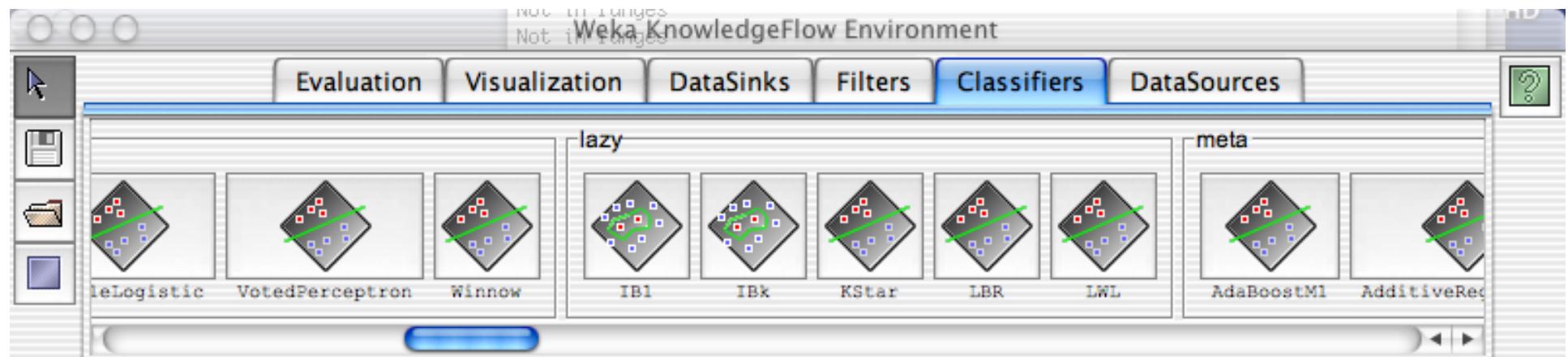
Text

09:59:02 - SMO

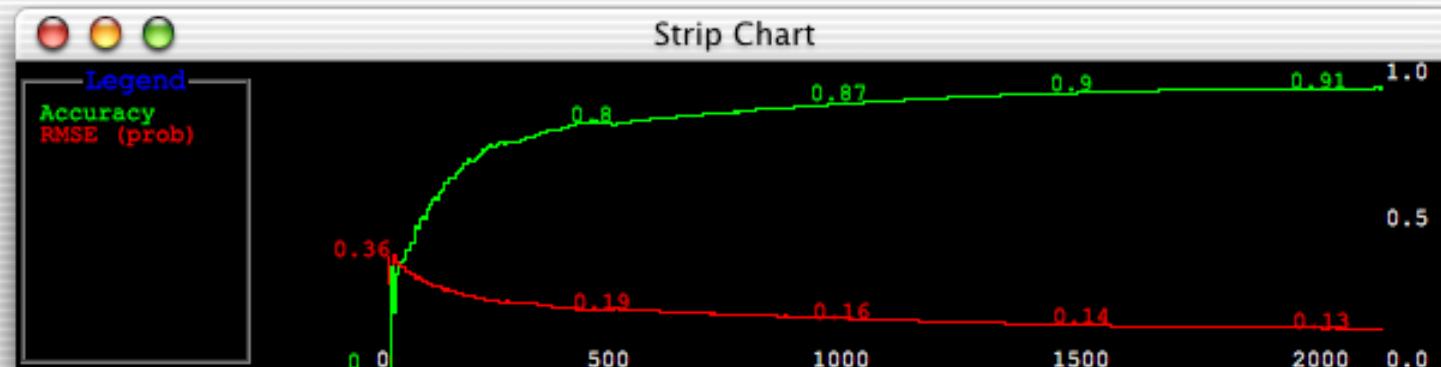
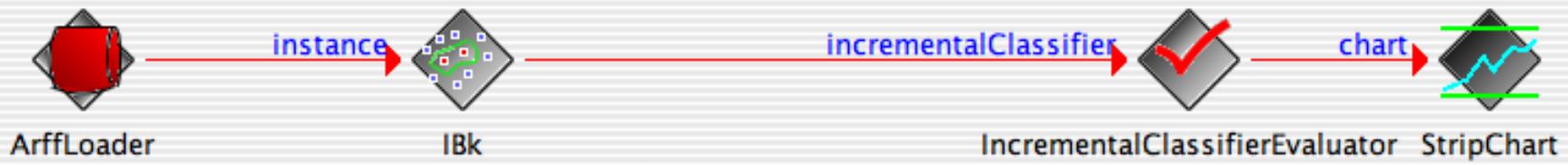
Correctly Classified Instances	144	96	%	
Incorrectly Classified Instances	6	4	%	
Kappa statistic	0.94			
Mean absolute error	0.2311			
Root mean squared error	0.288			
Relative absolute error	52	%		
Root relative squared error	58.704	%		
Total Number of Instances	150			

Data





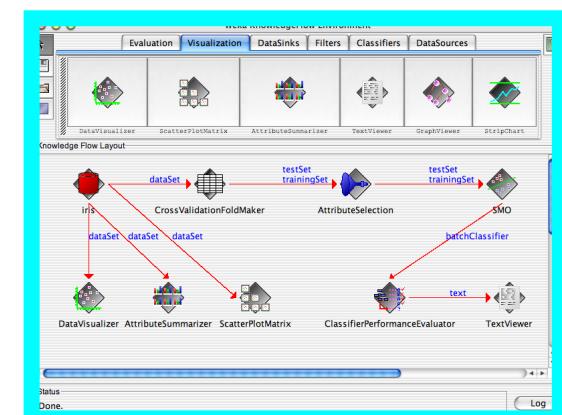
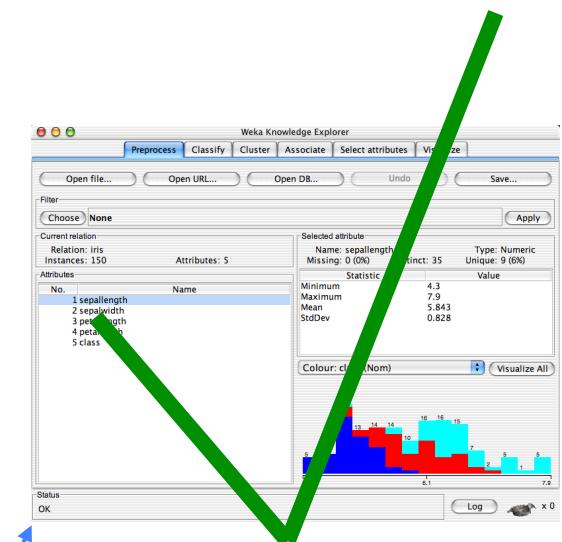
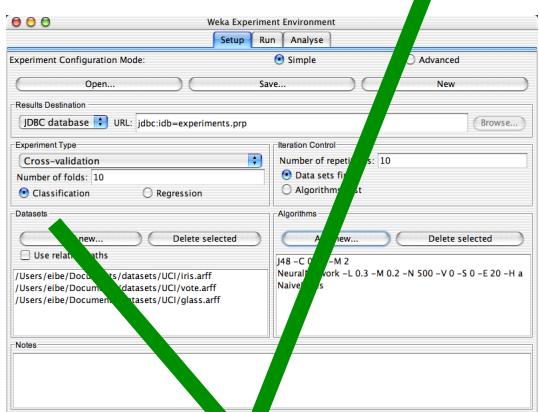
Knowledge Flow Layout

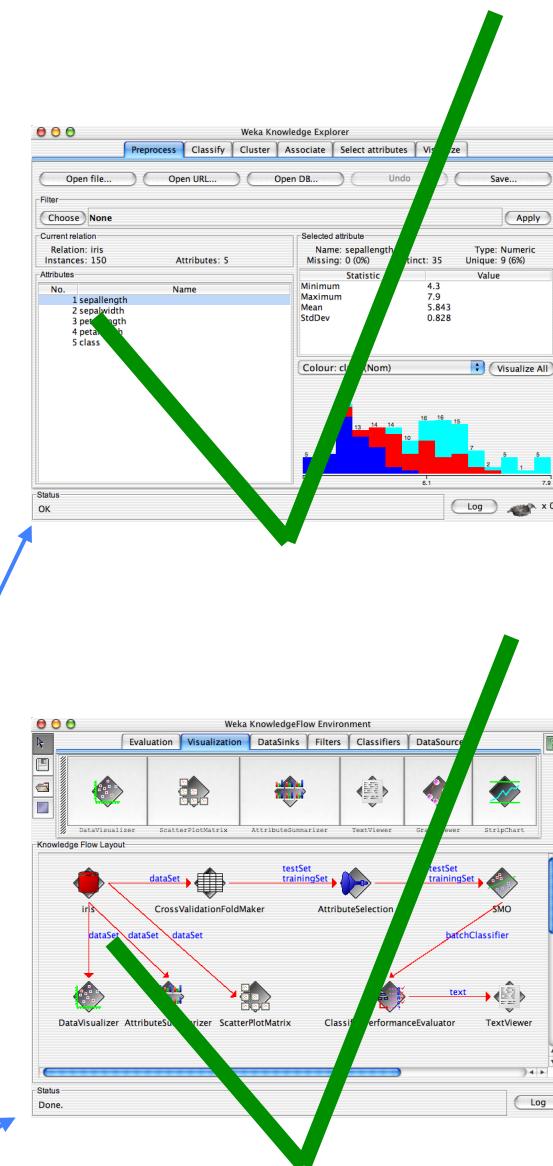
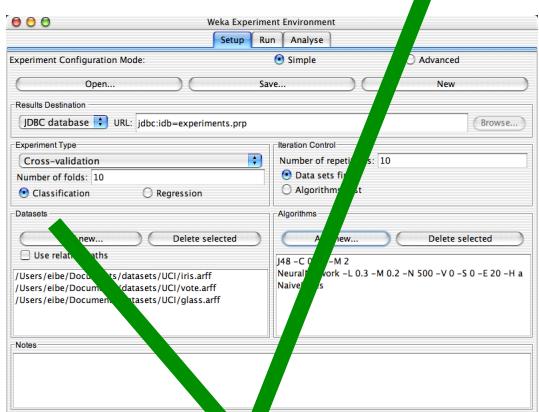


Status

Done.

Log





# Conclusion: try it yourself!

- WEKA is available at  
<http://www.cs.waikato.ac.nz/ml/weka>
- Also has a list of projects based on WEKA
- WEKA contributors:

Abdelaziz Mahoui, Alexander K. Seewald, Ashraf M. Kibriya, Bernhard Pfahringer , Brent Martin, Peter Flach, Eibe Frank ,Gabi Schmidberger ,Ian H. Witten , J. Lindgren, Janice Boughton, Jason Wells, Len Trigg, Lucio de Souza Coelho, Malcolm Ware, Mark Hall ,Remco Bouckaert , Richard Kirkby, Shane Butler, Shane Legg, Stuart Inglis, Sylvain Roy, Tony Voyle, Xin Xu, Yong Wang, Zhihai Wang