

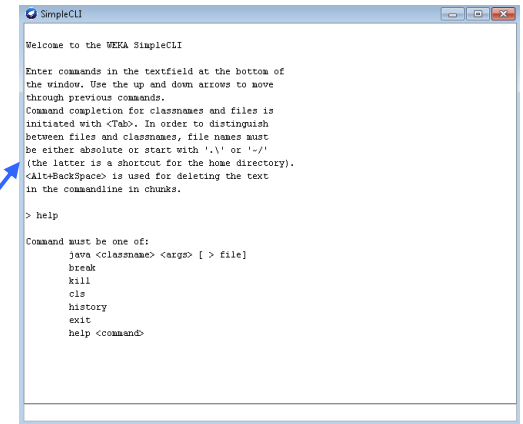
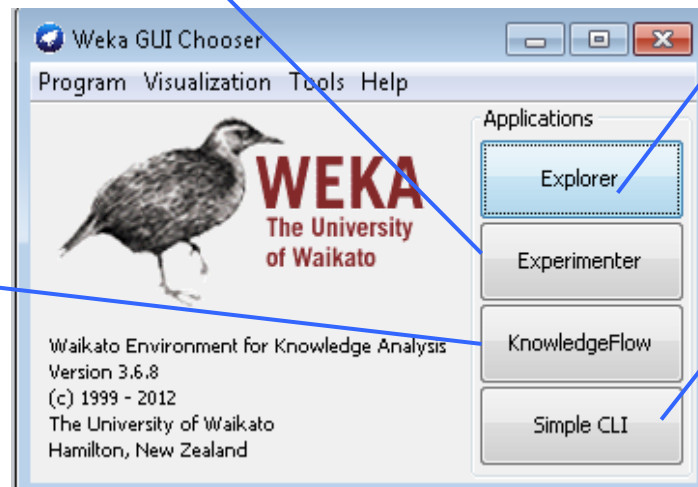
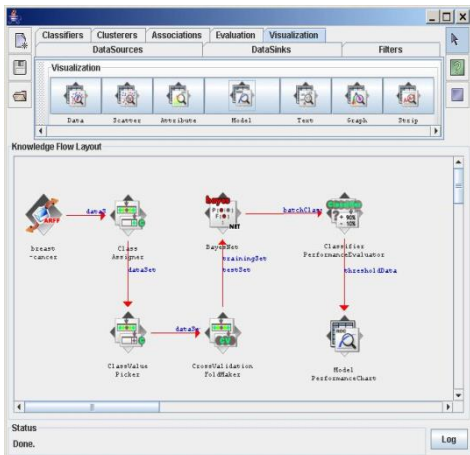
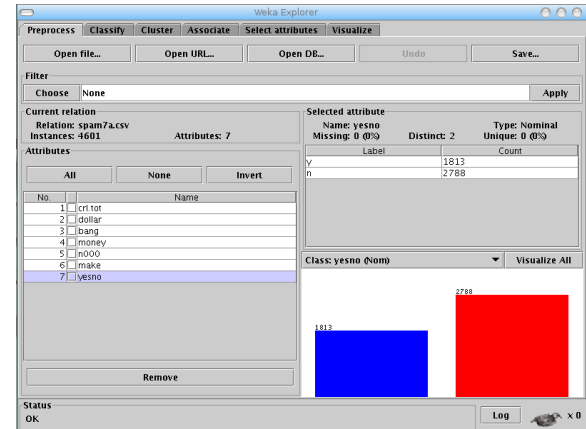
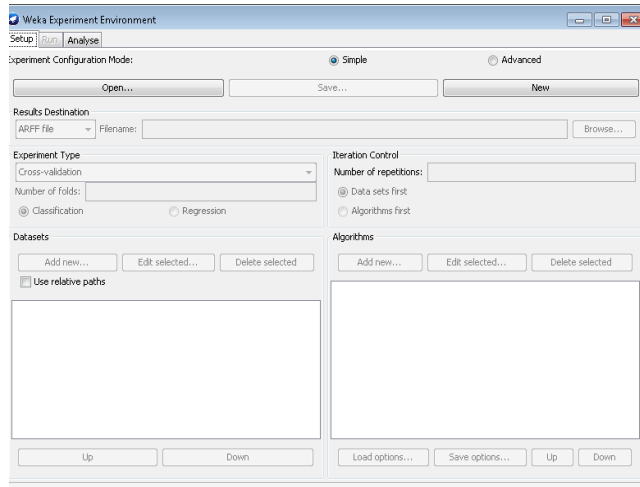
Τεχνητή Νοημοσύνη, Μηχανική Μάθηση και Εφαρμογές

WEKA

Εργαλείο WEKA

- Το WEKA (Waikato Environment for Knowledge Analysis) είναι μια σουίτα λογισμικού μηχανικής μάθησης γραμμένο σε Java που αναπτύχθηκε στο Πανεπιστήμιο του Waikato της Νέα Ζηλανδία.
- Ανήκει στην κατηγορία του λεγόμενου “ελεύθερου λογισμικού” (freeware).
- Είναι ένα από τα πιο διαδεδομένα λογισμικά Εξόρυξης Δεδομένων.

Το Περιβάλλον του WEKA



Explorer

Weka Explorer

Preprocess | **Classify** | Cluster | Associate | Select attributes | Visualize

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

Filter: Choose Apply

Current relation
Relation: spam7a.csv
Instances: 4601 Attributes: 7

Attributes

All | None | Invert

No.	Name
1	<input type="checkbox"/> crl.tot
2	<input type="checkbox"/> dollar
3	<input type="checkbox"/> bang
4	<input type="checkbox"/> money
5	<input type="checkbox"/> n000
6	<input type="checkbox"/> make
7	<input checked="" type="checkbox"/> yesno

Remove

Selected attribute
Name: yesno Type: Nominal
Missing: 0 (0%) Distinct: 2 Unique: 0 (0%)

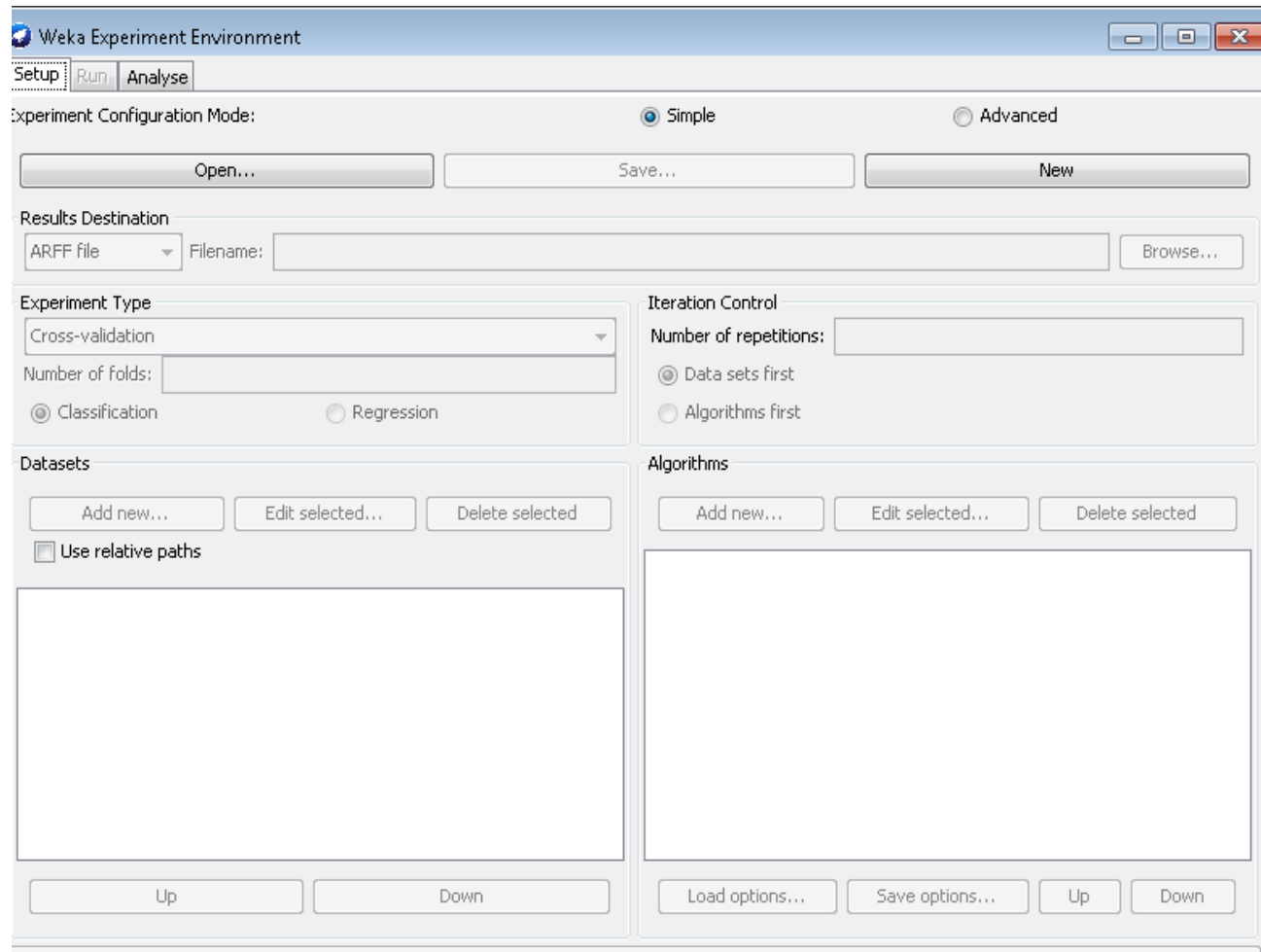
Label	Count
y	1813
n	2788

Class: yesno (Nom) Visualize All

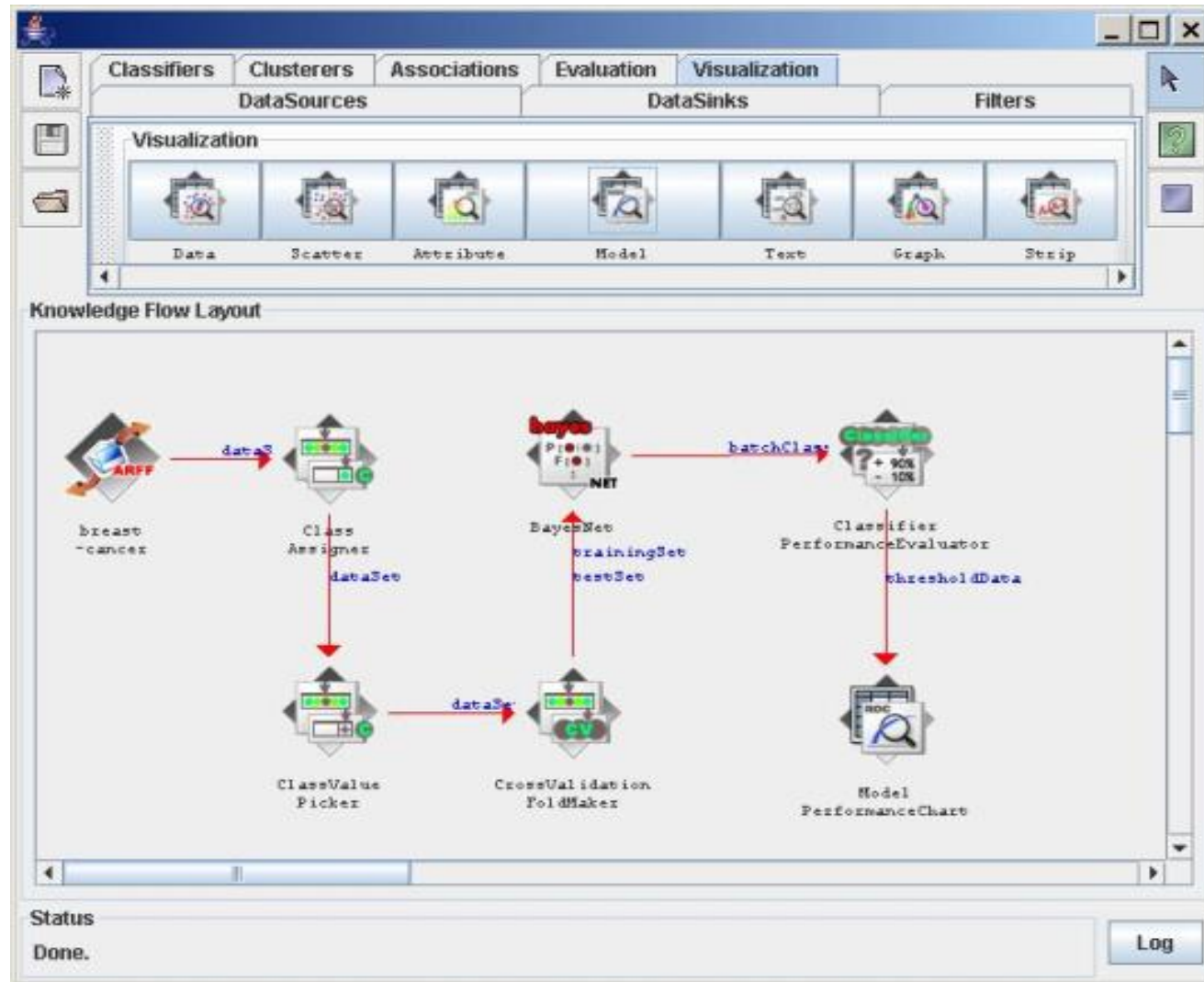
Label	Count
y	1813
n	2788

Status: OK Log x 0

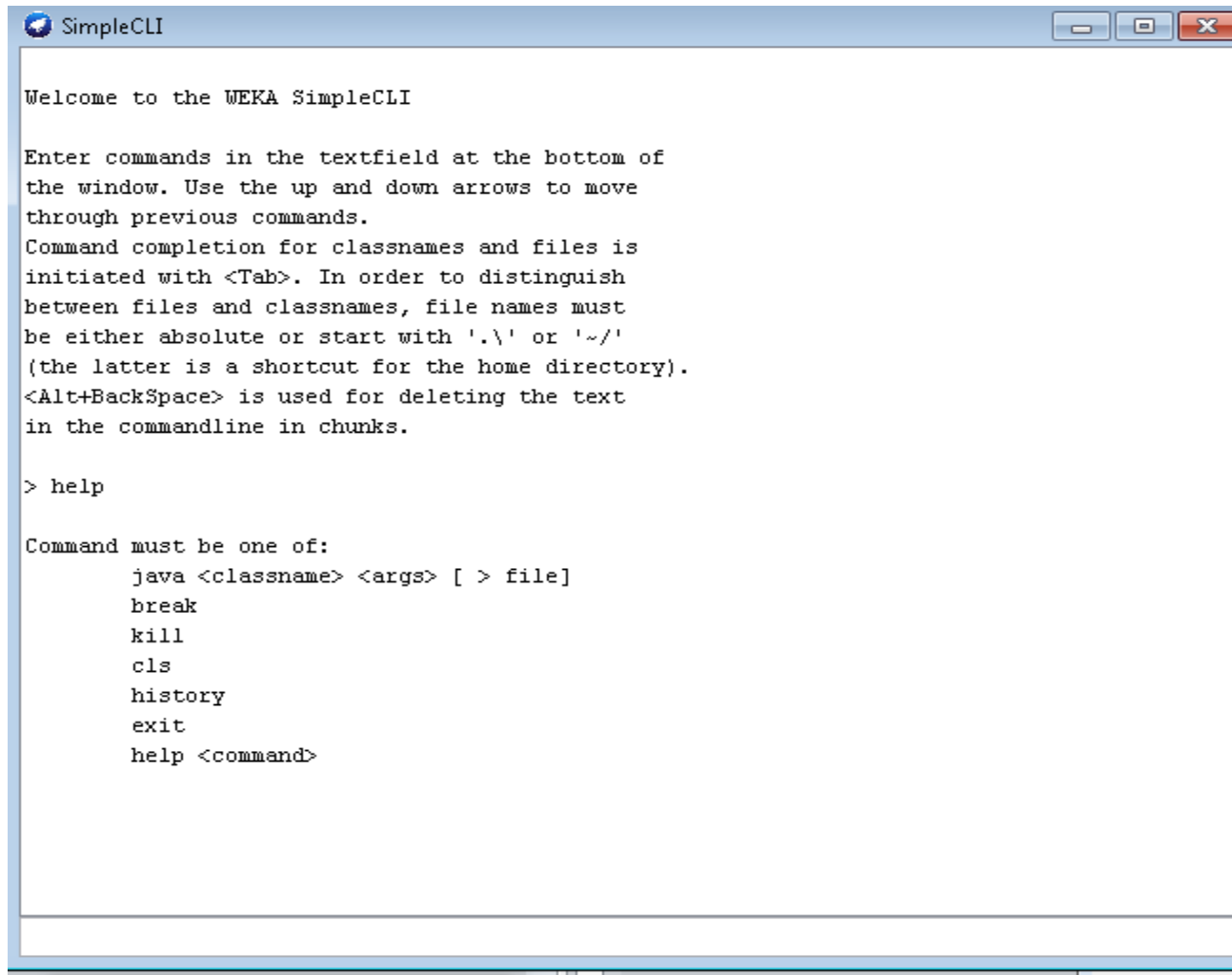
Experiment



KnowledgeFlow



Simple CLI



```
SimpleCLI

Welcome to the WEKA SimpleCLI

Enter commands in the textfield at the bottom of
the window. Use the up and down arrows to move
through previous commands.
Command completion for classnames and files is
initiated with <Tab>. In order to distinguish
between files and classnames, file names must
be either absolute or start with './' or '~/ '
(the latter is a shortcut for the home directory).
<Alt+BackSpace> is used for deleting the text
in the commandline in chunks.

> help

Command must be one of:
  java <classname> <args> [ > file]
  break
  kill
  cls
  history
  exit
  help <command>
```

Το Περιβάλλον του WEKA

- Βασικές Διεπαφές:
 - **“The Explorer”**: Ο χρήσης μπορεί να εκτελέσει όλες τις κύριες εργασίες εξόρυξης Δεδομένων, όπως κατηγοριοποίηση, παλινδρόμηση, ανάλυση συστάδων, ανακάλυψη κανόνων συσχέτισης, προ-επεξεργασία των δεδομένων και οπτικοποίηση.
 - **“Experimenter”** είναι ένα περιβάλλον για διεξαγωγή πειραμάτων, όπου αξιολογούνται μέθοδοι κατηγοριοποίησης και παλινδρόμησης. Σύγκριση της επίδοσης διαφορετικών μοντέλων και παρουσιάζει τα αποτελέσματα σε μορφή πίνακα.
 - **“Knowledge Flow”** είναι ένα περιβάλλον που επιτρέπει τη διεξαγωγή των ιδίων εργασιών με τον Explorer, διαθέτει όμως διαφορετική διεπαφή (interface). Στο περιβάλλον αυτό χρησιμοποιούνται components, τα οποία συνδέονται μεταξύ τους με γραφικό τρόπο, ο οποίος ορίζει τη ροή εργασίας

Explorer: Προ-επεξεργασία Δεδομένων

- Τα δεδομένα μπορούν να εισαχθούν από ένα αρχείο σε διάφορες μορφές: ARFF, CSV, C4.5, Binary
- Τα δεδομένα μπορούν επίσης να εισαχθούν από ένα URL ή από μια βάση δεδομένων SQL (χρησιμοποιώντας το JDBC)
- Τα εργαλεία προ-επεξεργασίας στο WEKA ονομάζονται "filters"

Το WEKA περιέχει filters για:

Διακριτοποίηση, κανονικοποίηση, επαναδειγματοληψία, επιλογή χαρακτηριστικών, μετασχηματισμός και συνδυασμός χαρακτηριστικών, κ.α

Explorer: Προ-επεξεργασία Δεδομένων

- Ο πιο συνηθισμένος όμως τρόπος φόρτωσης δεδομένων είναι μέσω ενός αρχείου ARFF.
- Τα αρχεία ARFF είναι απλά αρχεία κειμένου, όπου οι τιμές διαχωρίζονται με κόμμα (Coma Separated Values(CSV)).
- ARFF αρχεία:
 - Στην αρχή αναφέρεται η λέξη "@relation" και ακολουθεί το όνομα του πίνακα δεδομένων.
- Στη συνέχεια γίνεται η δήλωση των πεδίων. Για κάθε πεδίο χρειάζεται μια γραμμή, στην αρχή της οποίας υπάρχει η λέξη "@attribute", ακολουθεί το όνομα του πεδίου (πχ Turnover), και κατόπιν δηλώνεται ο τύπος του πεδίου. Αν το πεδίο είναι αριθμητικό, χρησιμοποιείται η λέξη "numeric". Αν το πεδίο είναι ονομαστικό, δηλώνονται οι δυνατές τιμές μέσα σε αγκύλες

Παράδειγμα αρχείου ARFF WEKA

```

% Summary Statistics:
%
%           Min  Max   Mean   SD   Class Correlation
%  sepal length: 4.3  7.9   5.84  0.83    0.7826
%  sepal width:  2.0  4.4   3.05  0.43   -0.4194
%  petal length:  1.0  6.9   3.76  1.76    0.9490 (high!)
%  petal width:  0.1  2.5   1.20  0.76    0.9565 (high!)
%
% 9. Class Distribution: 33.3% for each of 3 classes.

@RELATION iris

@ATTRIBUTE sepallength REAL
@ATTRIBUTE sepalwidth  REAL
@ATTRIBUTE petallength REAL
@ATTRIBUTE petalwidth  REAL
@ATTRIBUTE class       {Iris-setosa,Iris-versicolor,Iris-virginica}

@DATA
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
4.6,3.1,1.5,0.2,Iris-setosa
5.0,3.6,1.4,0.2,Iris-setosa
5.4,3.9,1.7,0.4,Iris-setosa
4.6,3.4,1.4,0.3,Iris-setosa
5.0,3.4,1.5,0.2,Iris-setosa

```

Ο «δημιουργός γνώσης» WEKA

- Είναι ένα φιλικό γραφικό περιβάλλον εργασίας
- Όλα τα εργαλεία του WEKA: φίλτρα, κατηγοριοποιητές, ταξινομητές, συσχετιστές και επιλογείς χαρακτηριστικών μπορούν να παρουσιασθούν γραφικά με δισδιάστατα γραφήματα
- Με βάση τις γραφικές αναπαραστάσεις των επιλεγμένων δεδομένων δοκιμάζονται και προβλέπεται η απόδοση των εναλλακτικών μοντέλων που θα επιλεγούν να χρησιμοποιηθούν τελικά για την διαμόρφωση της «θαμμένης» γνώσης.
- Είναι επιλέξιμα μέσα από αναδυόμενα μενού

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
```

```
...
```



Δομή
ARFF
αρχείου

WEKA only deals with “flat” files

@relation heart-disease-simplified

Αριθμητική
τιμή

@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

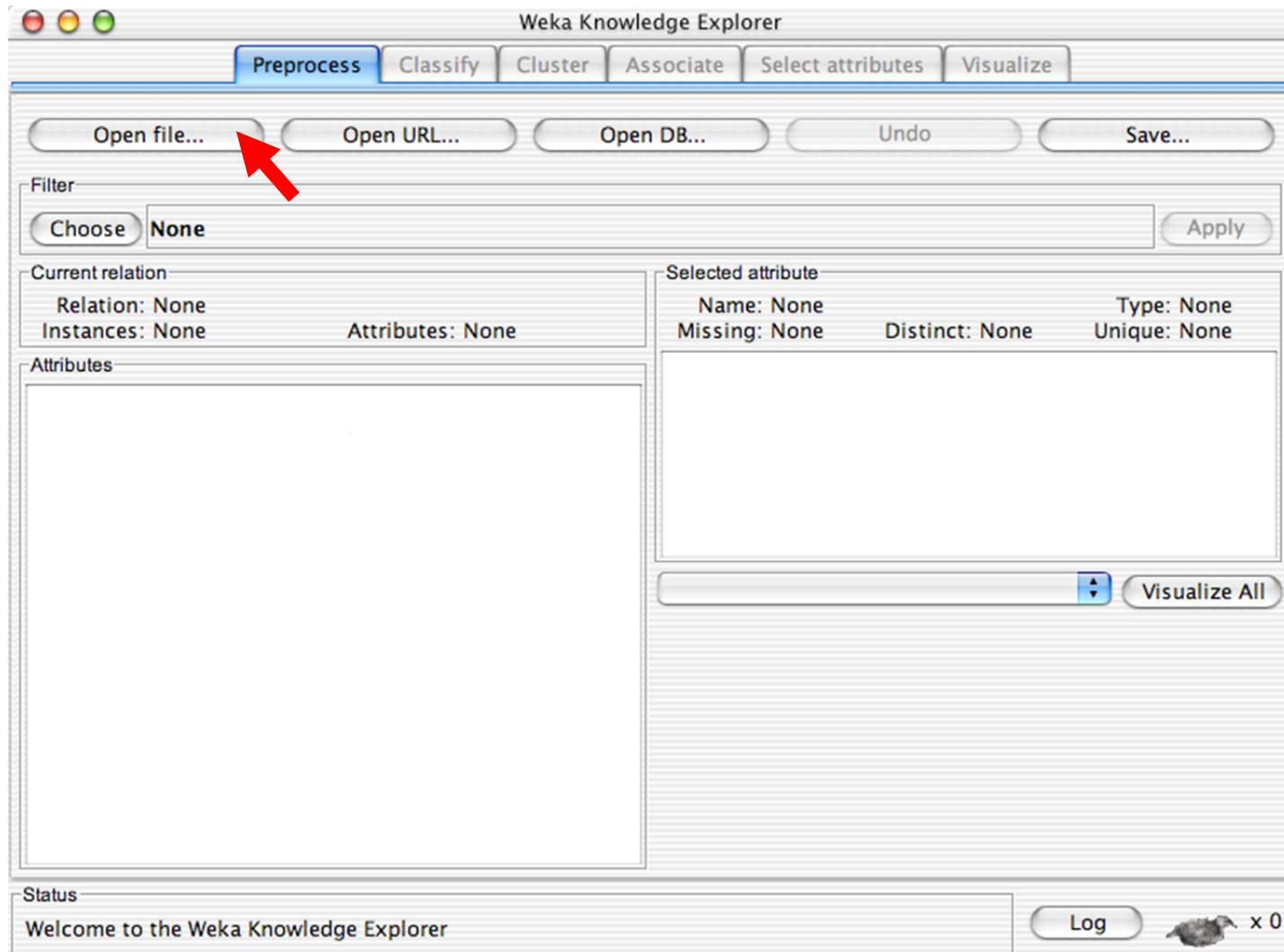
38,female,non_anginal,?,no,not_present

...

WEKA Explorer

- **Preprocess:** Επιλογή και τροποποίηση των δεδομένων.
- **Classify:** Εκπαίδευση και έλεγχος των learning schemes που κάνουν classification ή κάνουν regression.
- **Cluster:** Εφαρμογή clustering στα δεδομένα.
- **Associate:** Δημιουργεί κανόνες συσχέτισης για τα δεδομένα.
- **Select attributes:** Επιλογή των πιο σχετικών χαρακτηριστικών γνωρισμάτων των δεδομένων.
- **Visualize:** Προβολή διαδραστικών 2D διαγραμμάτων των δεδομένων.

WEKA Διεπαφή



Προεπεξεργασία Δεδομένων

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	sepalength
2	sepalwidth
3	petallength
4	petalwidth
5	class

Selected attribute
Name: sepalength
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

A histogram showing the distribution of sepalength values for three classes. The x-axis represents sepalength from 4.3 to 7.9, and the y-axis represents the number of instances. The bars are colored by class: blue for class 1, red for class 2, and cyan for class 3. The counts for each bin are: 5 (blue), 6 (blue), 21 (blue), 13 (red), 14 (red), 14 (red), 10 (cyan), 16 (cyan), 16 (cyan), 15 (cyan), 7 (cyan), 2 (red), 5 (cyan), 1 (cyan), 5 (cyan).

Status: OK

Log x 0

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

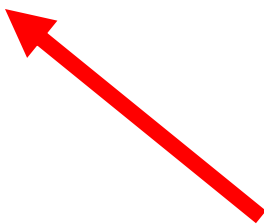
Attributes: 5

Selected attribute

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

Attributes

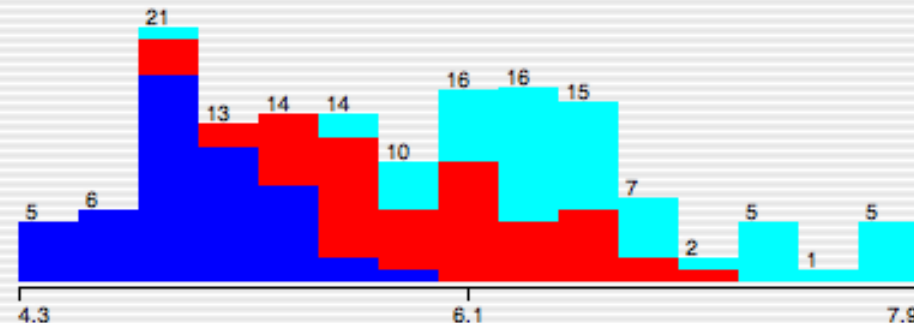
No.	Name
1	sepallength
2	sepalwidth
3	petallength
4	petalwidth
5	class



Statistic	Value
Minimum	4.3
Maximum	7.9
Mean	5.843
StdDev	0.828

Colour: class (Nom)

Visualize All



Status

OK

Log



Προεπεξεργασία Δεδομένων

The screenshot shows the Weka Knowledge Explorer interface. The 'Preprocess' tab is active. The 'Current relation' is 'iris' with 150 instances and 5 attributes. The 'Selected attribute' is 'class', which is a nominal attribute with 3 distinct values and 0 missing values. The 'Attributes' list shows 5 attributes: 1 sepallength, 2 sepalwidth, 3 petallength, 4 petalwidth, and 5 class. The 'Selected attribute' table shows the distribution of the 'class' attribute: Iris-setosa (50), Iris-versicolor (50), and Iris-virginica (50). The 'Visualize All' button is visible, and the resulting visualization shows three colored bars (blue, red, cyan) representing the counts for each class.

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

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

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

Attributes

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

Colour: class (Nom) Visualize All

50 50 50

Status
OK Log x 0

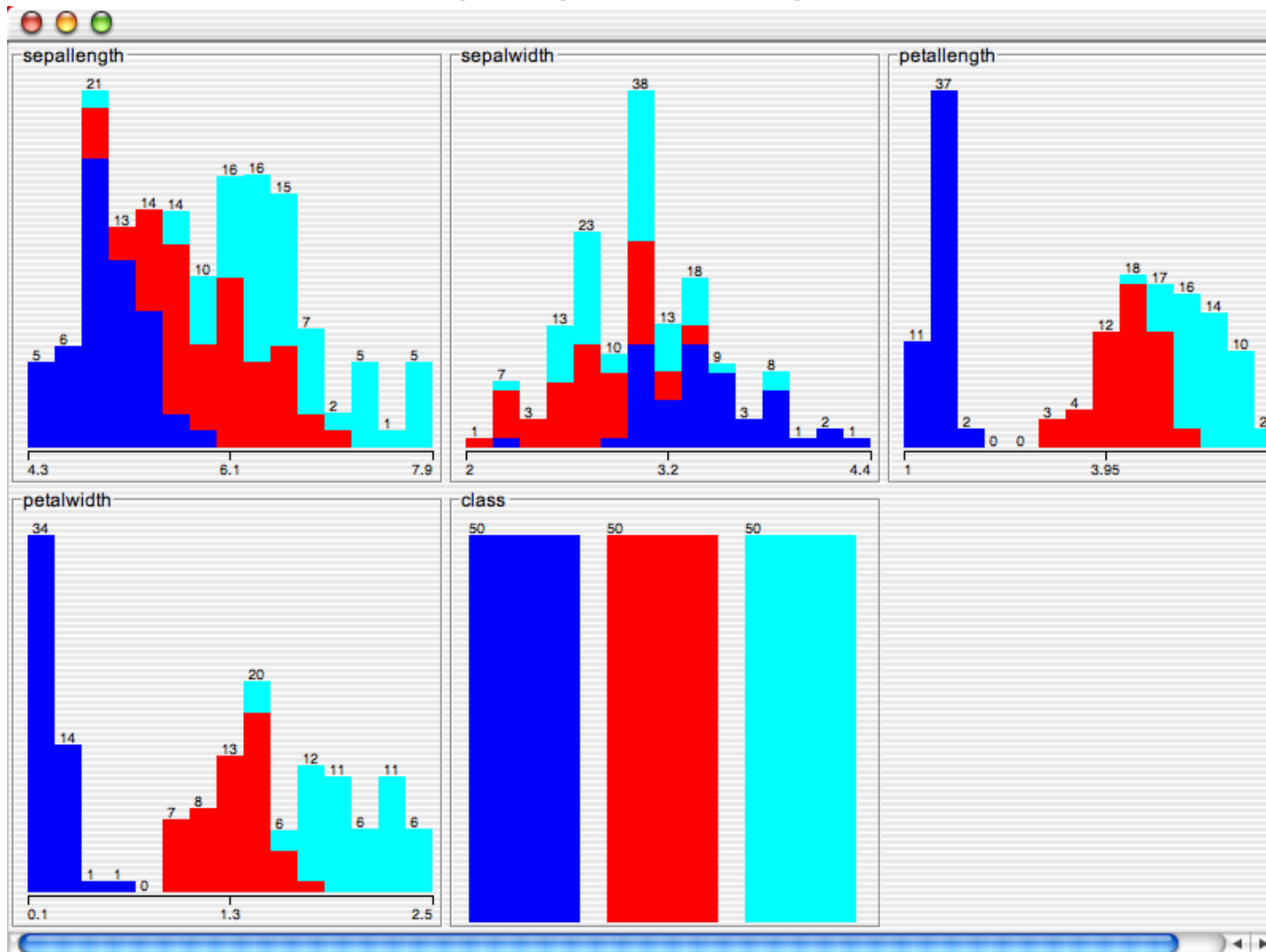
Προ-επεξεργασία Δεδομένων

The screenshot shows the Weka Knowledge Explorer interface. The 'Preprocess' tab is active. The 'Current relation' is 'iris' with 150 instances and 5 attributes. The 'Attributes' list includes 'class' (highlighted). The 'Selected attribute' section shows 'class' with 3 distinct values and 0 missing values. A table displays the distribution of the 'class' attribute:

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

A red arrow points to the 'Iris-versicolor' bar in the bar chart below. The bar chart shows three bars, each with a count of 50, representing the distribution of the 'class' attribute. The bars are colored blue, red, and cyan. The status bar at the bottom shows 'OK' and a 'Log' button.

Οπτικοποίηση δεδομένων



Προεπεξεργασία δεδομένων

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

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

Statistic	Value
Minimum	1
Maximum	6.9
Mean	3.759
StdDev	1.764

Attributes

No.	Name
1	sepalength
2	sepalwidth
3	petalength
4	petalwidth
5	class

Colour: class (Nom) Visualize All

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

Status: OK

Log x 0

Προεπεξεργασία δεδομένων

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	sepalength
2	sepalwidth
3	petallength
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

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	3
3.5 - 4.0	4
4.0 - 4.5	12
4.5 - 5.0	18
5.0 - 5.5	17
5.5 - 6.0	16
6.0 - 6.5	14
6.5 - 7.0	10
7.0 - 7.5	2
7.5 - 8.0	4

Status: OK

Log x 0

Προεπεξεργασία δεδομένων

The screenshot displays the Weka Knowledge Explorer interface with the 'Preprocess' tab selected. The 'Discretize' filter is applied to the 'iris' dataset, which has 150 instances and 5 attributes. The 'petal length' attribute is selected for discretization. The filter configuration is as follows:

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

A histogram at the bottom of the dialog shows the distribution of the 'petal length' attribute across 10 bins. The x-axis ranges from 1 to 6.9, with a major tick at 3.95. The y-axis represents the number of instances in each bin. The distribution is as follows:

Bin	Count
1	11
2	2
3	0
4	0
5	3
6	4
7	12
8	14
9	10
10	2
11	4

The status bar at the bottom shows 'OK' and a 'Log' button.

Προεπεξεργασία δεδομένων

The screenshot shows the Weka Knowledge Explorer interface with the 'Preprocess' tab selected. A 'weka.gui.GenericObjectEditor' window is open, displaying the configuration for the 'weka.filters.unsupervised.attribute.Discretize' filter. The filter is applied to the 'iris' relation, which has 150 instances. The 'Attributes' list shows 'petallength' selected. The configuration parameters are as follows:

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

A red arrow points to the 'makeBinary' parameter. The 'About' section describes the filter as one that discretizes a range of numeric attributes into nominal attributes. The status bar at the bottom shows 'OK' and a 'Log' button.

Προεπεξεργασία δεδομένων

The screenshot displays the Weka Knowledge Explorer interface with the 'Preprocess' tab selected. A dialog box titled 'weka.gui.GenericObjectEditor' is open, showing the configuration for the 'weka.filters.unsupervised.attribute.Discretize' filter. The 'iris' dataset is currently selected, with 150 instances and 5 attributes. The 'petalength' attribute is highlighted in the attributes list.

weka.filters.unsupervised.attribute.Discretize

About
An instance filter that discretizes a range of numeric attributes in the dataset into nominal attributes. [More](#)

attributeIndices: first-last

bins: 10

findNumBins: False

invertSelection: False

makeBinary: False

useEqualFrequency: True

Buttons: Open... Save... OK Cancel

Progress bar: 1 2 0 0 3 4 3.95 6.9

Status: OK Log x 0

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

Προεπεξεργασία δεδομένων

The screenshot displays the Weka Knowledge Explorer interface. The 'Preprocess' tab is active, and the 'Discretize -B 10 -R first-last' filter is applied to the 'iris' dataset. The 'Attributes' list shows 'petallength' selected. The 'weka.gui.GenericObjectEditor' window shows the filter configuration: 'attributeIndices' is 'first-last', 'bins' is '10', 'findNumBins' is 'False', 'invertSelection' is 'False', 'makeBinary' is 'False', and 'useEqualFrequency' is 'True'. A histogram at the bottom shows the distribution of 'petallength' values, with a red arrow pointing to the 'Save...' button.

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**

Current relation: Relation: iris, Instances: 150, Attributes: 5

Attributes:

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

weka.gui.GenericObjectEditor

weka.filters.unsupervised.attribute.Discretize

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

attributeIndices: first-last

bins: 10

findNumBins: False

invertSelection: False

makeBinary: False

useEqualFrequency: True

Open... | Save... | OK | Cancel

Status: OK

Log x 0

Προεπεξεργασία δεδομένων

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	sepalength
2	sepalwidth
3	petallength
4	petalwidth
5	class

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

Statistic	Value
Minimum	1
Maximum	6.9
Mean	3.759
StdDev	1.764

Colour: class (Nom) Visualize All

Bin Range	Count	Class
1.0 - 1.5	11	1
1.5 - 2.0	37	1
2.0 - 2.5	2	1
2.5 - 3.0	0	1
3.0 - 3.5	0	1
3.5 - 4.0	3	2
4.0 - 4.5	4	2
4.5 - 5.0	12	2
5.0 - 5.5	18	2
5.5 - 6.0	17	3
6.0 - 6.5	16	3
6.5 - 7.0	14	3
7.0 - 7.5	10	3
7.5 - 8.0	2	3
8.0 - 8.5	4	4

Status: OK

Log x 0

Προεπεξεργασία δεδομένων

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	sepalwidth
2	sepalwidth
3	petalwidth
4	petalwidth
5	class

Selected attribute
Name: petalwidth
Missing: 0 (0%)
Distinct: 43
Type: Numeric
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

Προεπεξεργασία δεδομένων

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	sepalength
2	sepalwidth
3	petallength
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

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

Status: OK

Log x 0

Κατηγοριοποίηση Δεδομένων

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

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)

Classifier output

Status

OK

Log

 x 0

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose **ZeroR**

Test options

Use training set

Supplied test set

Cross-validation Folds

Percentage split %

(Nom) class

Result list (right-click for options)

Classifier output

Status

OK



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

- weka
 - classifiers
 - bayes
 - functions
 - lazy
 - meta
 - misc
 - trees
 - adtree
 - DecisionStump
 - Id3
 - j48
 - J48**
 - lmt
 - m5
 - RandomForest
 - RandomTree
 - REPTree
 - UserClassifier
 - rules

Classifier output

Status

OK

Log



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

Status

OK

Log

 x 0

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose **J48 -C 0.25 -M 2**

Test options

Use training set

Supplied test set

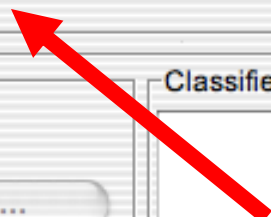
Cross-validation Folds

Percentage split %

(Nom) class

Result list (right-click for options)

Classifier output



Status

OK

x 0

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)

weka.gui.GenericObjectEditor

weka.classifiers.trees.j48.J48

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

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)

weka.gui.GenericObjectEditor

weka.classifiers.trees.j48.J48

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

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

Status

OK

Log

 x 0

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

Status

OK

Log

 x 0

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

Status

OK

Log

 x 0

Classifier
Choose **J48 -C 0.25 -M 2**

Test options

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

(Nom) class

Result list (right-click for options)

Classifier output

Status
OK

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

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

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

Status

OK

Log



x 0

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

Status

OK

Log

 x 0

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.J48 -C 0.25 -M 2

Relation: iris

Instances: 150

Attributes: 5

sepalength

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

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.J48 -C 0.25 -M 2

Relation: iris

Instances: 150

Attributes: 5

sepalength

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

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

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose J48 -C 0.25 -M 2

Test options

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

(Nom) class

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

x 0

Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Classifier

Choose J48 -C 0.25 -M 2

Test options

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

(Nom) class

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
1	1	Iris-setosa
1	0.95	Iris-versicolor
0.882	0.938	Iris-virginica

Status

OK

x 0

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

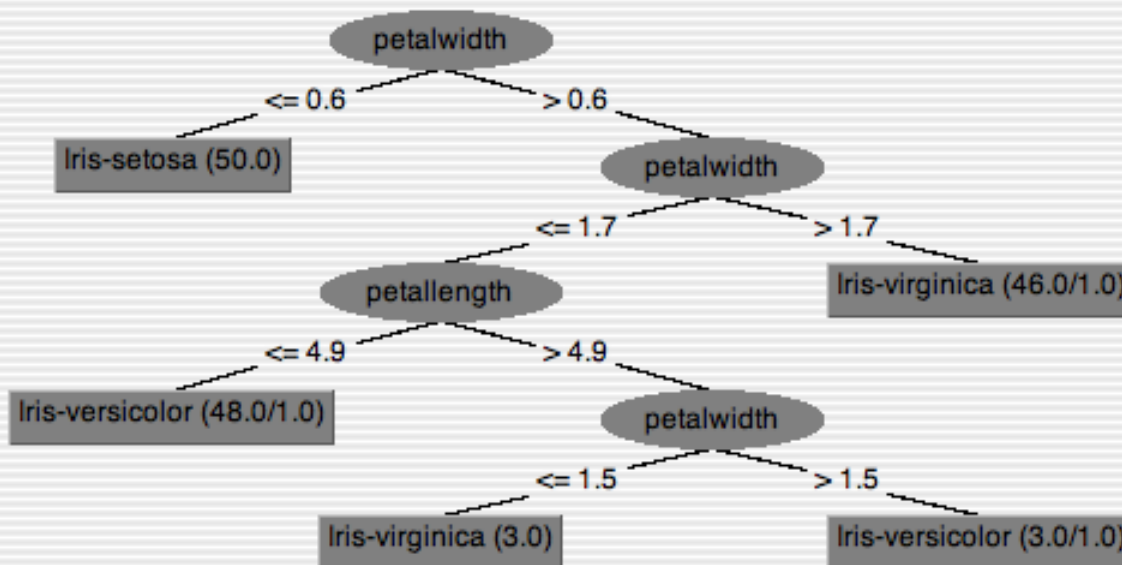
(Nom) class

Start

Result list (right-click for)

11:49:05 - trees.j48.J

Tree View



96.0784 %
3.9216 %

class
is-setosa
is-versicolor
is-virginica

```

10 0 0 | a = Iris-setosa
0 19 0 | b = Iris-versicolor
0 2 15 | c = Iris-virginica
  
```

Status

OK

Log



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

Συσταδοποίηση δεδομένων

- Αφού έχει επιλεγεί ένα σύνολο δεδομένων είναι δυνατόν να γίνει συσταδοποίηση (εύρεση ομάδων 'όμοιων' δεδομένων).
- Από την καρτέλα cluster μπορεί να επιλεγεί ένας αλγόριθμος με βάση τον οποίο θα γίνει συσταδοποίηση και με το κουμπί start να αρχίσει η εκτέλεση του αλγορίθμου.

Συσταδοποίηση δεδομένων

The image shows two overlapping windows from the Weka software. The background window is the 'Clusterer' interface, and the foreground window is the 'Clusterer Visualize' window.

Clusterer Window:

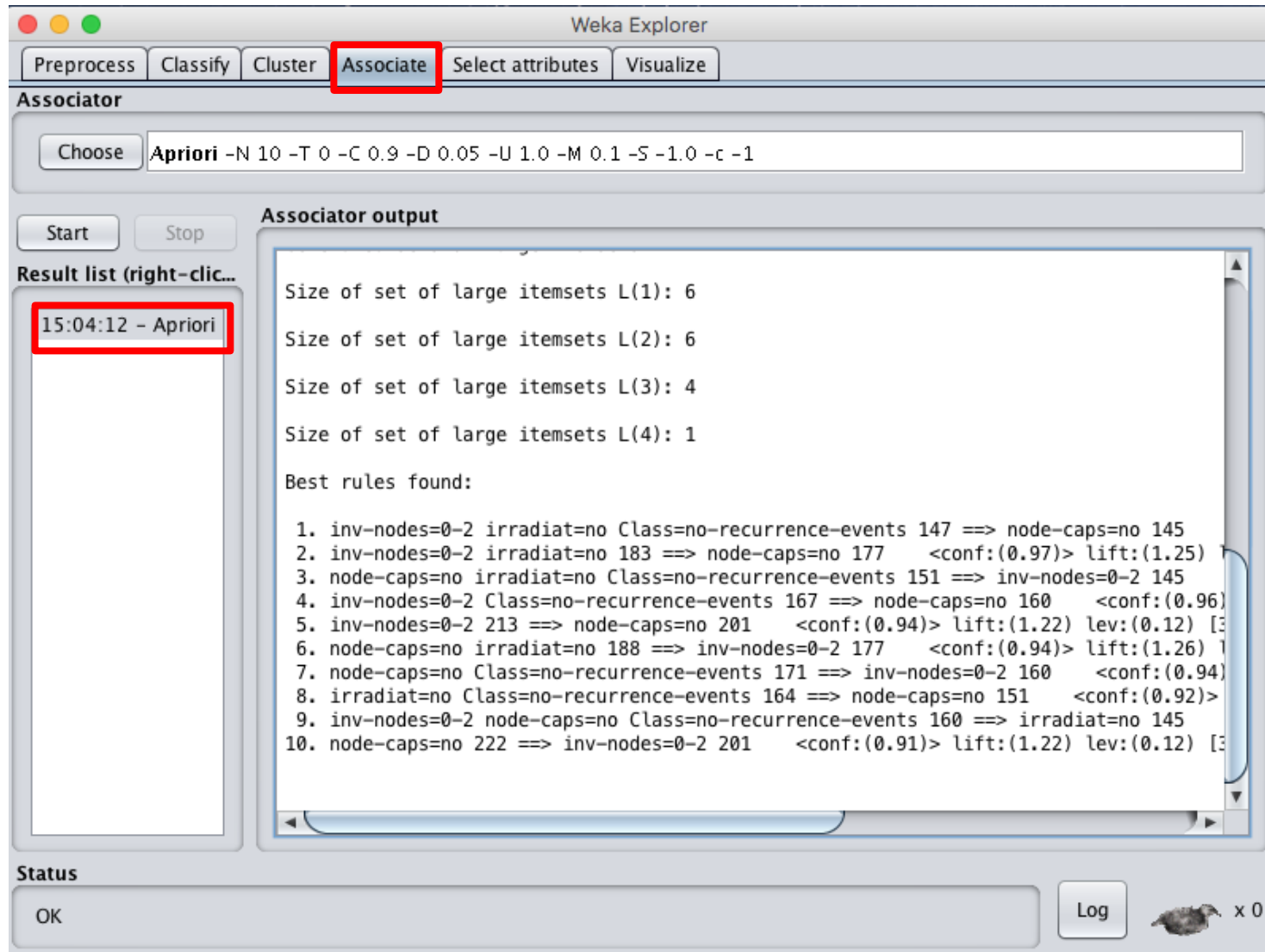
- Buttons: Preprocess, Classify, Cluster
- Clusterer: Choose EM -I 100 -N -1 -S 100 -M 1.0E-6
- Cluster mode:
 - Use training set
 - Supplied test set (Set...)
 - Percentage split (% 66)
 - Classes to clusters evaluation (Nom) class
- Store clusters for visualization
- Buttons: Ignore attributes, Start, Stop
- Result list (right-click for options): 12:24:46 - EM
- Status: OK

Clusterer Visualize Window:

- Title: Weka Clusterer Visualize: 12:24:46 - EM (iris)
- X: petallength (Num), Y: petalwidth (Num)
- Colour: Cluster (Nom), Select Instance
- Buttons: Reset, Clear, Save
- Jitter: [Slider]
- Plot: iris_clustered (Scatter plot with axes and data points)
- Class colour: cluster0 (blue), cluster1 (red), cluster2 (green), cluster3 (cyan)
- Normal Distribution. Mean = 1.031 StdDev = 0.0464
- Clustered Instances:

0	50	(33%)
1	36	(24%)
2	54	(36%)
3	10	(7%)
- Log likelihood: -1.80561
- Buttons: Log, x 0

Κανόνες Συσχέτισης



The screenshot shows the Weka Explorer interface with the 'Associate' tab selected. The 'Associator' section shows the 'Apriori' algorithm chosen with the following command: `Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1`. The 'Associator output' window displays the results of the algorithm, including the size of the set of large itemsets for each level (L(1) to L(4)) and a list of the best rules found. The 'Result list' on the left shows a single entry: '15:04:12 - Apriori'. The 'Status' bar at the bottom indicates 'OK'.

Weka 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 -c -1`

Start Stop

Result list (right-click...)

15:04:12 - Apriori


Associator output

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

Best rules found:

1. `inv-nodes=0-2 irradiat=no Class=no-recurrence-events 147 ==> node-caps=no 145`
2. `inv-nodes=0-2 irradiat=no 183 ==> node-caps=no 177 <conf:(0.97)> lift:(1.25)`
3. `node-caps=no irradiat=no Class=no-recurrence-events 151 ==> inv-nodes=0-2 145`
4. `inv-nodes=0-2 Class=no-recurrence-events 167 ==> node-caps=no 160 <conf:(0.96)`
5. `inv-nodes=0-2 213 ==> node-caps=no 201 <conf:(0.94)> lift:(1.22) lev:(0.12) [3`
6. `node-caps=no irradiat=no 188 ==> inv-nodes=0-2 177 <conf:(0.94)> lift:(1.26) 1`
7. `node-caps=no Class=no-recurrence-events 171 ==> inv-nodes=0-2 160 <conf:(0.94)`
8. `irradiat=no Class=no-recurrence-events 164 ==> node-caps=no 151 <conf:(0.92)>`
9. `inv-nodes=0-2 node-caps=no Class=no-recurrence-events 160 ==> irradiat=no 145`
10. `node-caps=no 222 ==> inv-nodes=0-2 201 <conf:(0.91)> lift:(1.22) lev:(0.12) [3`

Status

OK Log  x 0

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)

Associator output

Status

OK

Log



x 0

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)

Associator output

Status

OK

Log



x 0

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 219
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:(0.98)
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

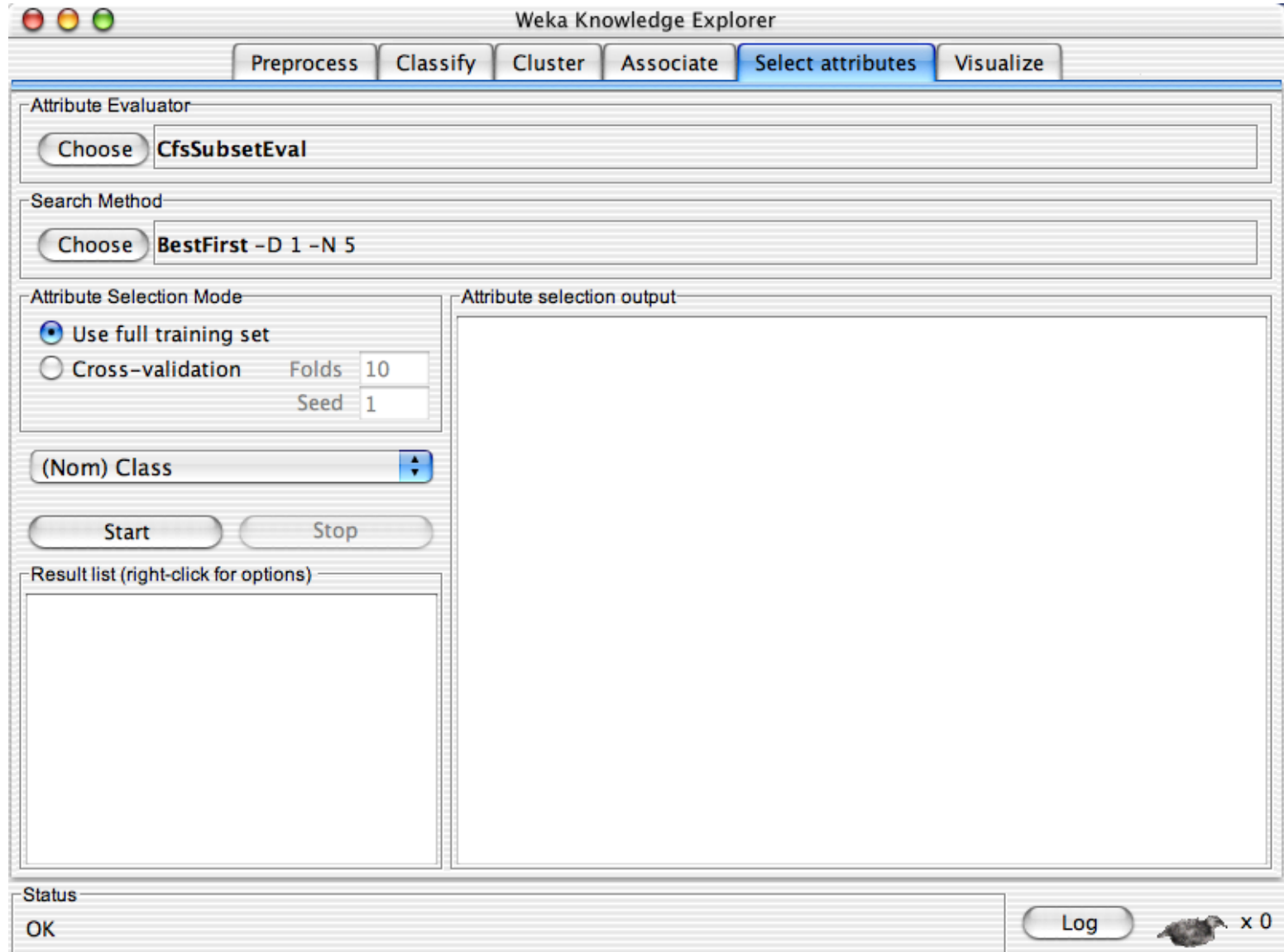
Status

OK

Log



Επιλογή χαρακτηριστικών



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)

Attribute selection output

Status

OK

Log



x 0

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)

Attribute selection output

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



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

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
    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

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
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

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

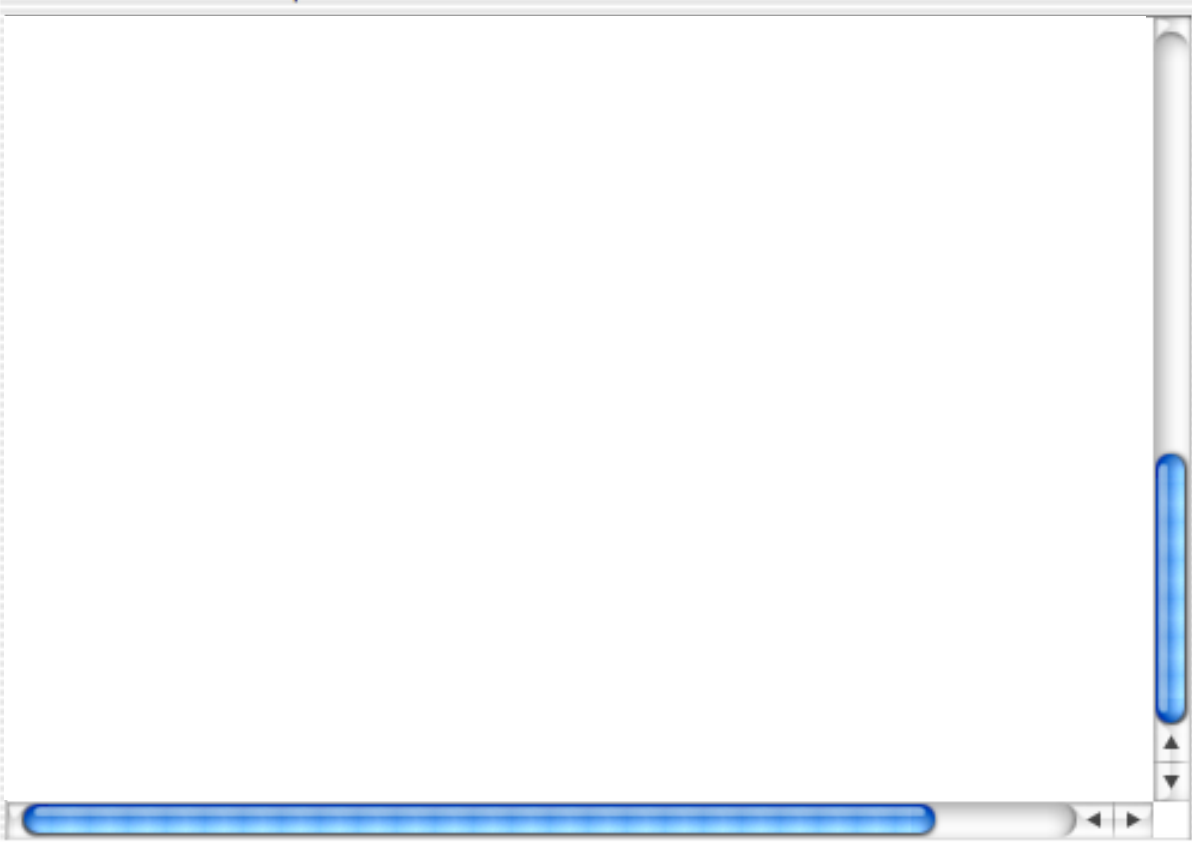
Start

Stop

Result list (right-click for options)

16:39:40 - BestFirst + CrossSubsetEval

Attribute selection output



Status

OK

Log



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

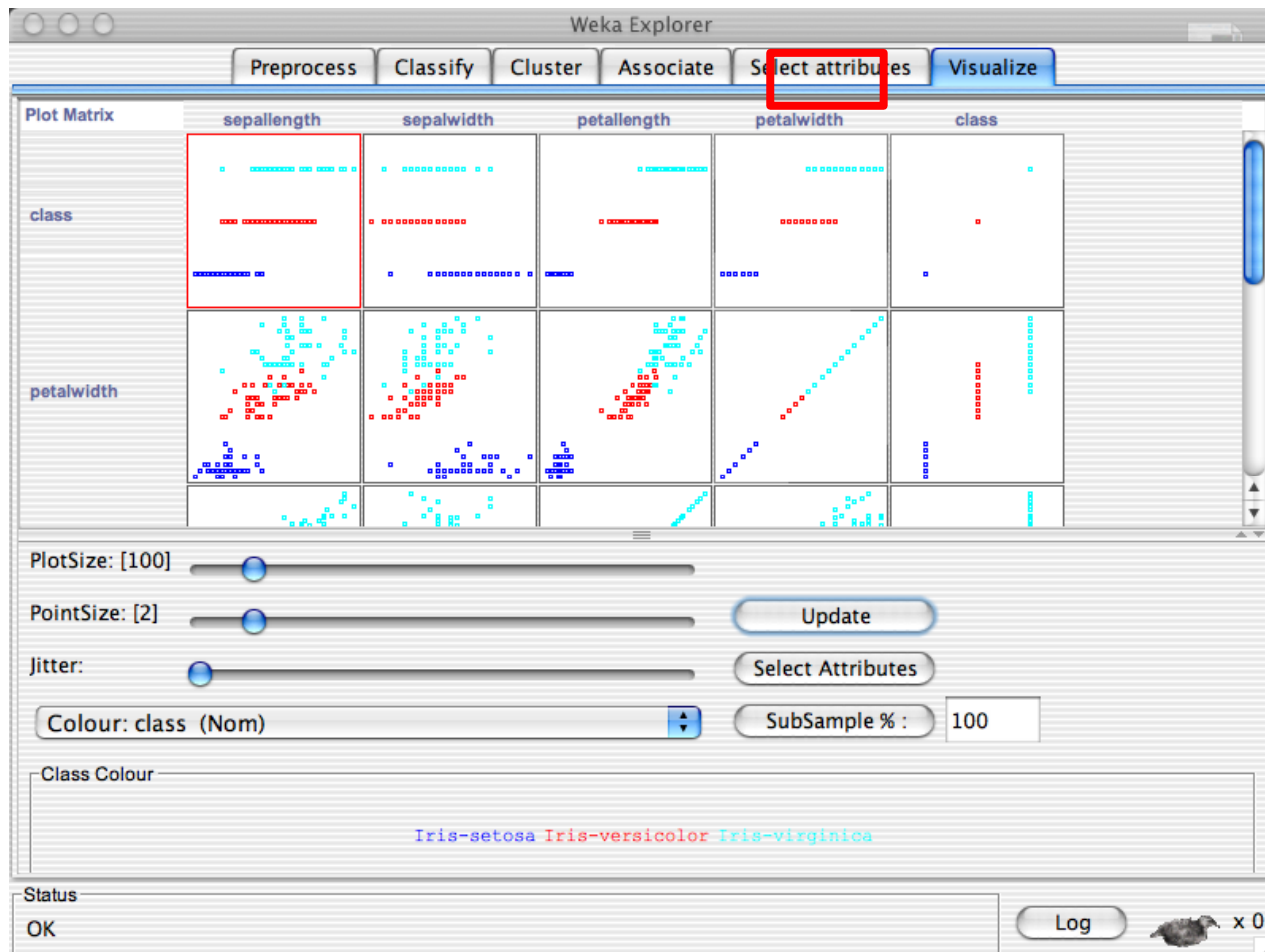
OK

Log

x 0

Οπτικοποίηση

- Δυσδιάστατα την κατανομή των χαρακτηριστικών.



Preprocess

Classify

Cluster

Associate

Select attributes

Visualize

Plot Matrix

RI

Na

Mg

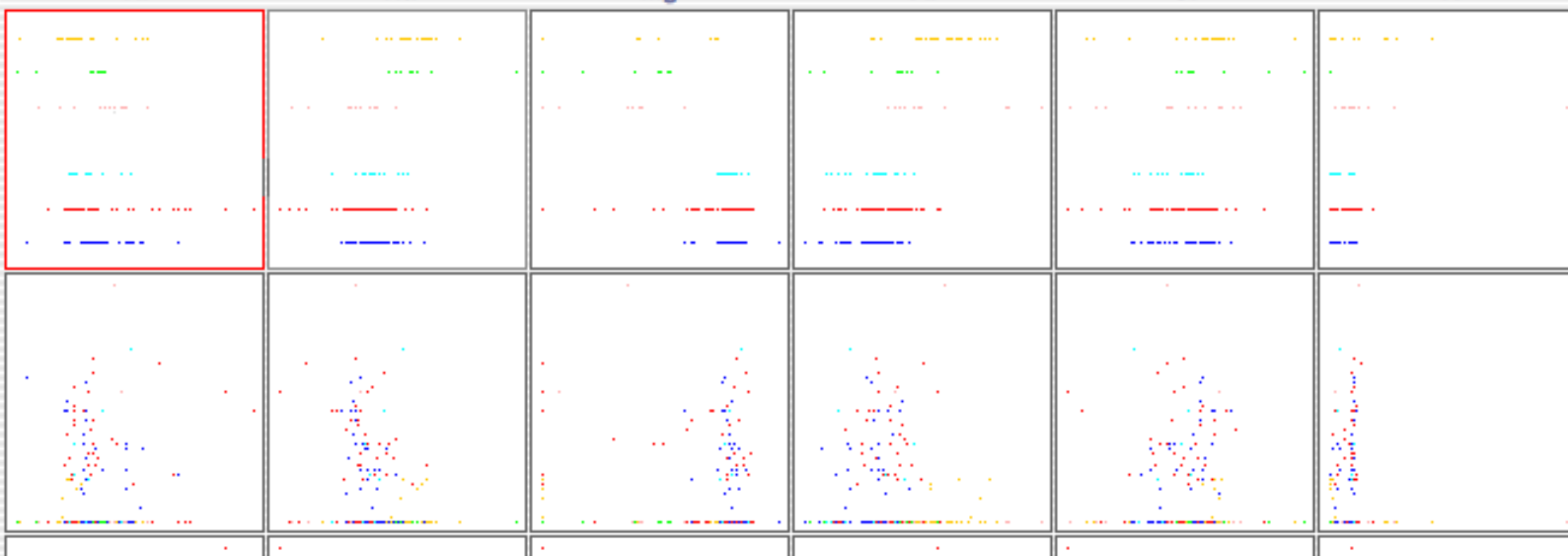
Al

Si

K

Type

Fe



PlotSize: [100]



PointSize: [1]



Jitter:



Update

Select Attributes

Colour: Type (Nom)



SubSample % :

100

Class Colour

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

Status

OK

Log



Preprocess

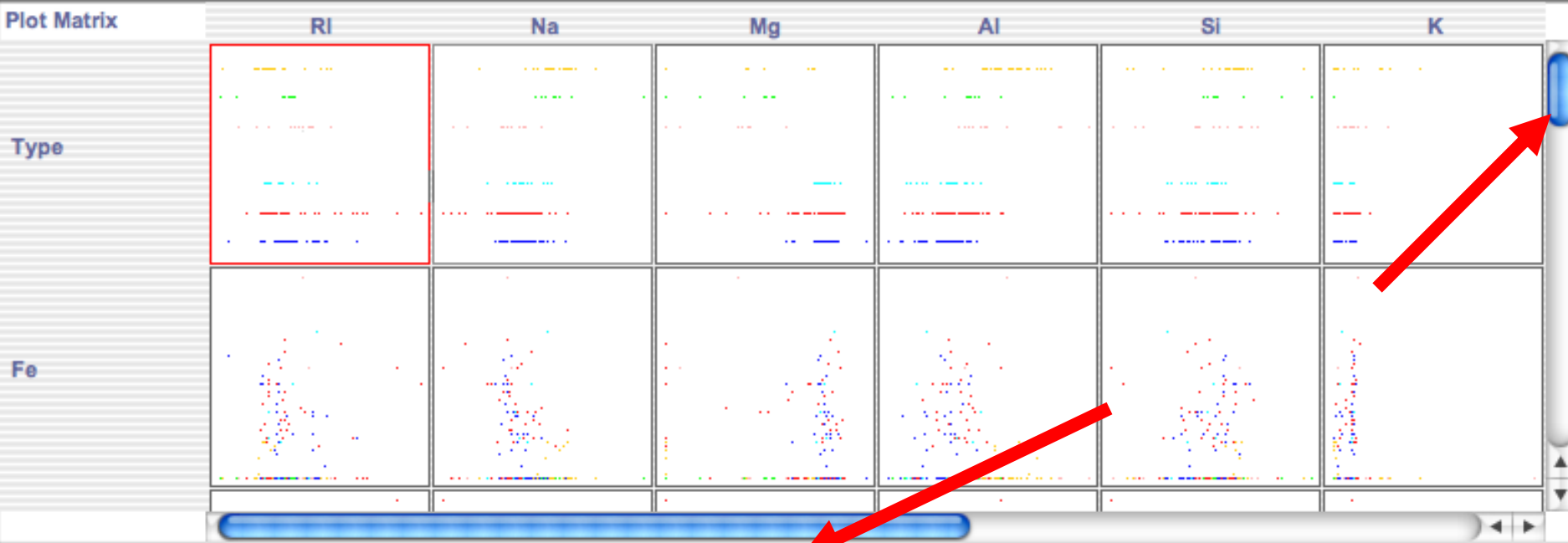
Classify

Cluster

Associate

Select attributes

Visualize



PlotSize: [100]

PointSize: [1]

Jitter:

Update

Select Attributes

Colour: Type (Nom)

SubSample % : 100

Class Colour

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

Status
OK

Log



Preprocess

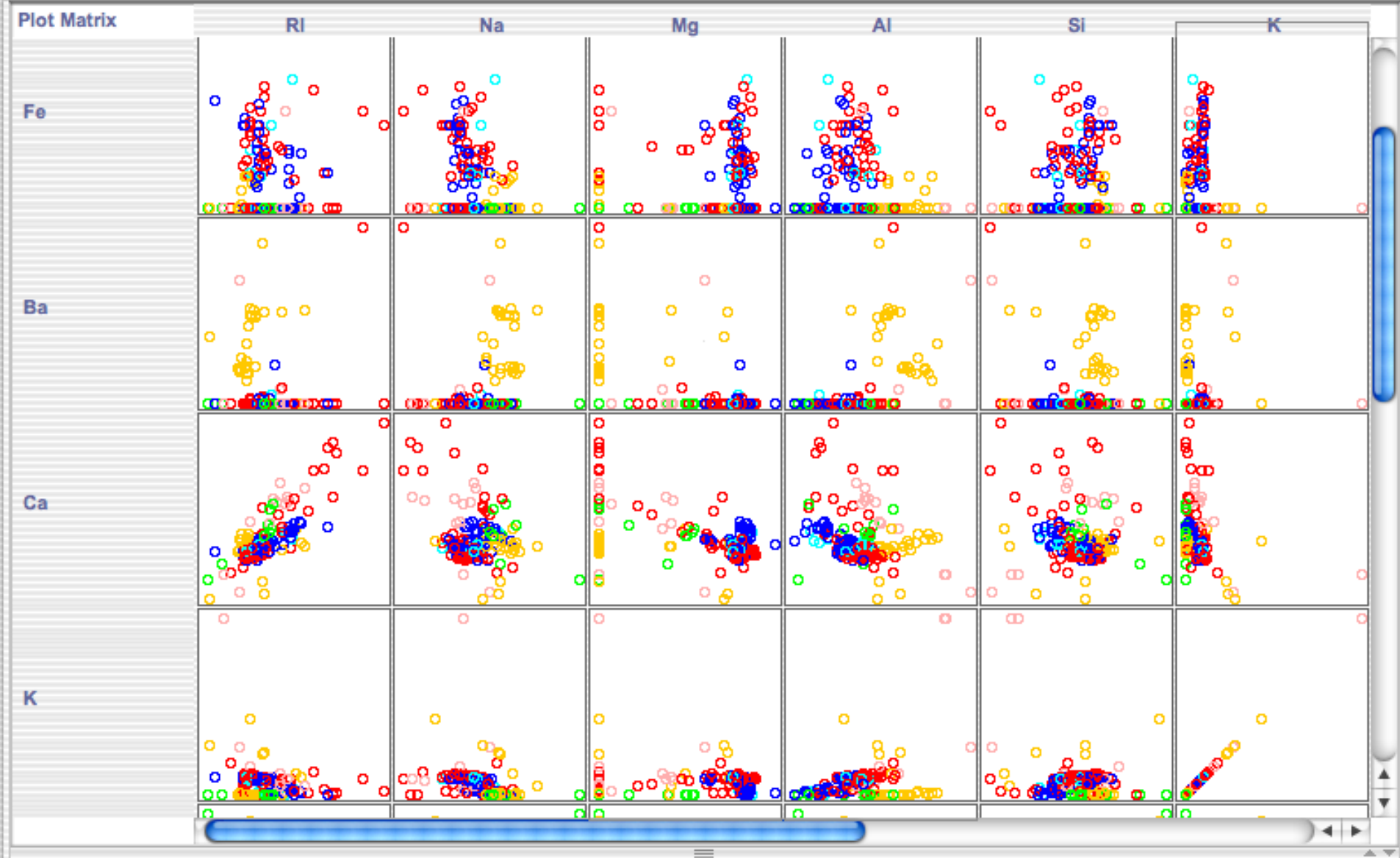
Classify

Cluster

Associate

Select attributes

Visualize



Status

OK

Log

x 0

Preprocess

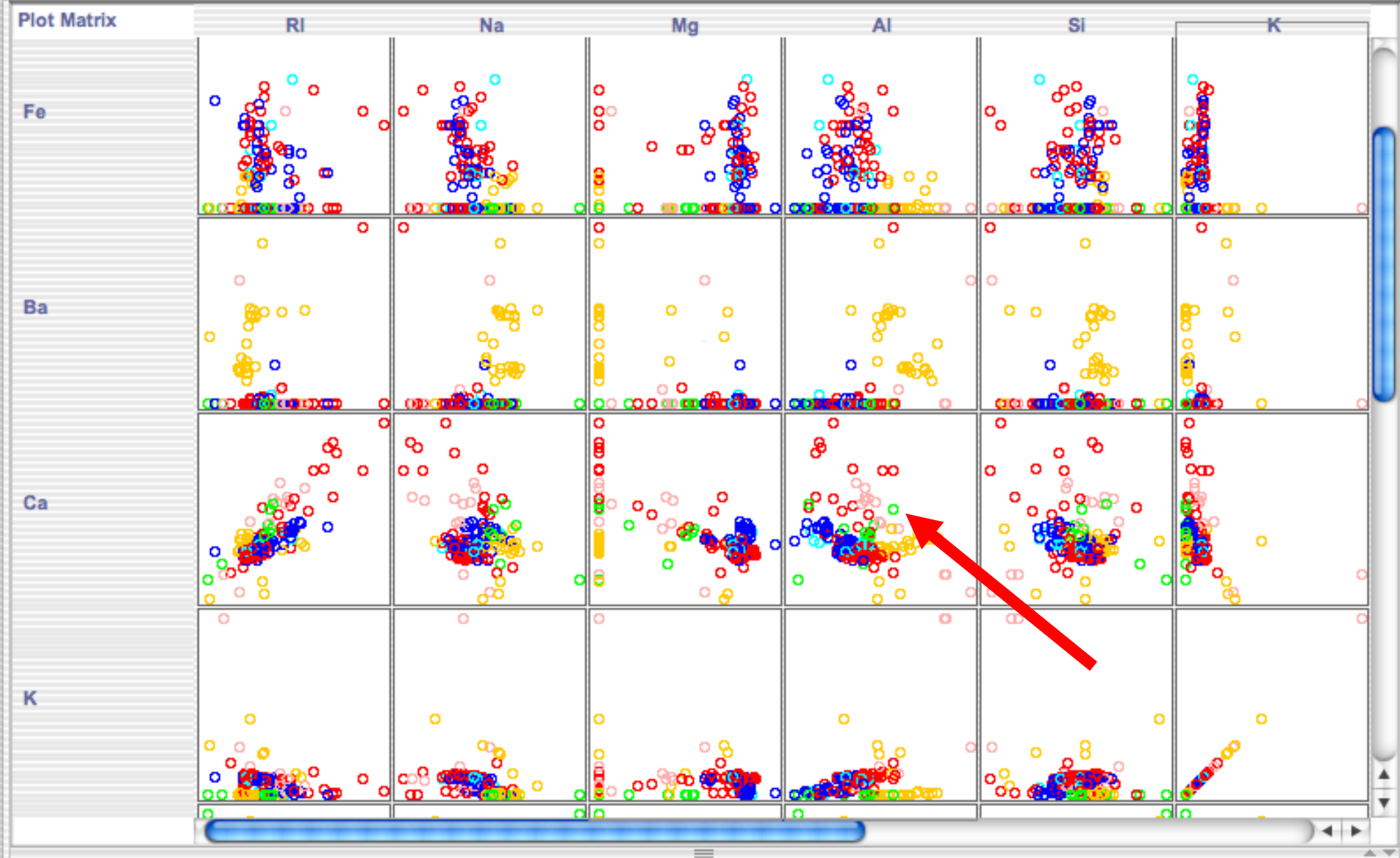
Classify

Cluster

Associate

Select attributes

Visualize



Status

OK

Log

x 0

X: Al (Num)

Y: Ca (Num)

Colour: Type (Nom)

Select Instance

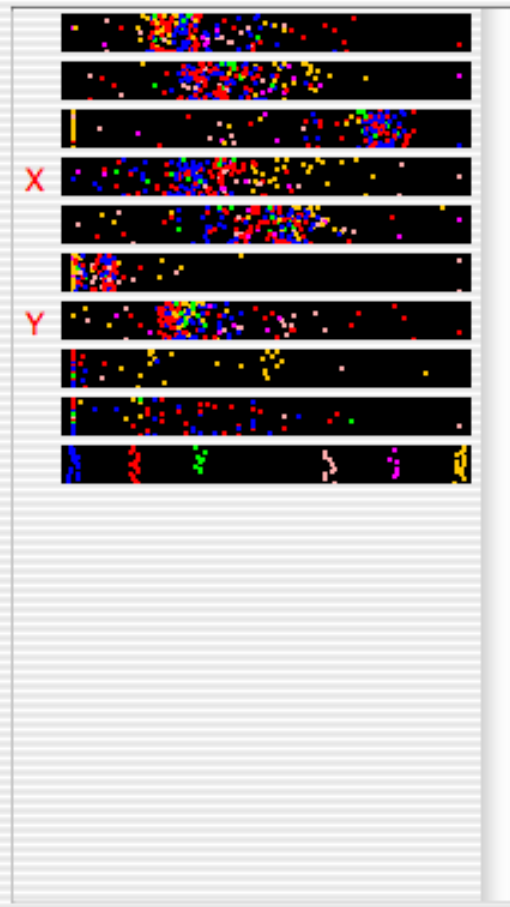
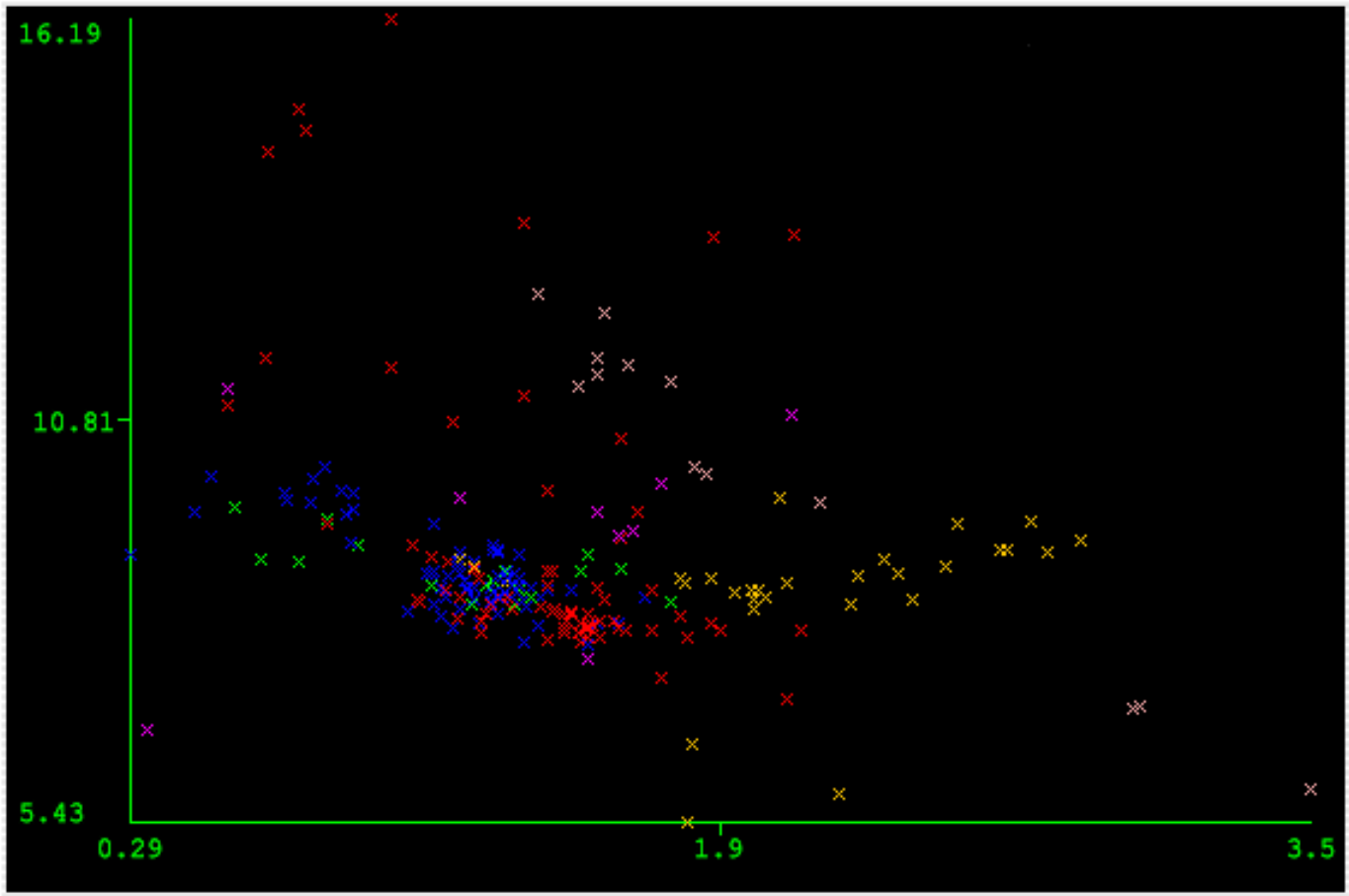
Reset

Clear

Save

Jitter

Plot: Glass



Class colour

build wind float

build wind non-float

vehic wind float

vehic wind non-float

containers

tableware

headlamps

X: Al (Num)

Y: Ca (Num)

Colour: Type (Nom)

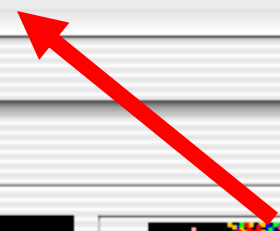
Select Instance

Reset

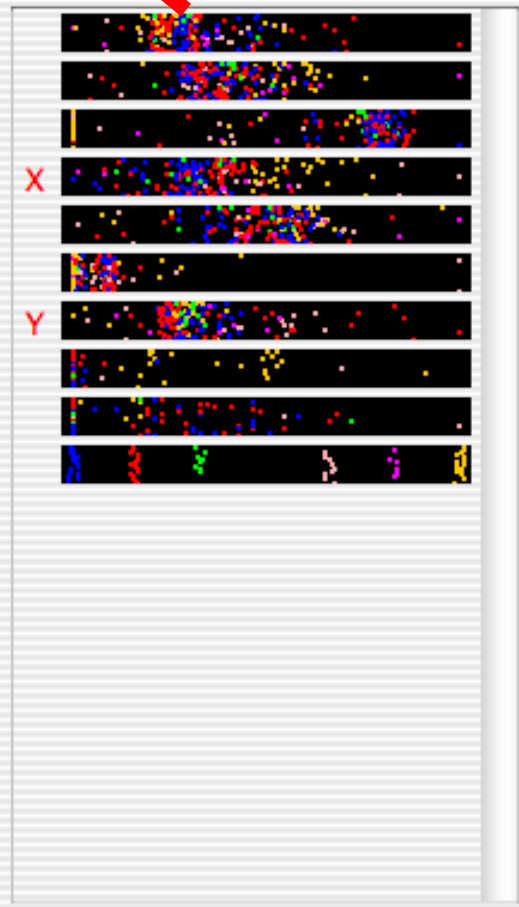
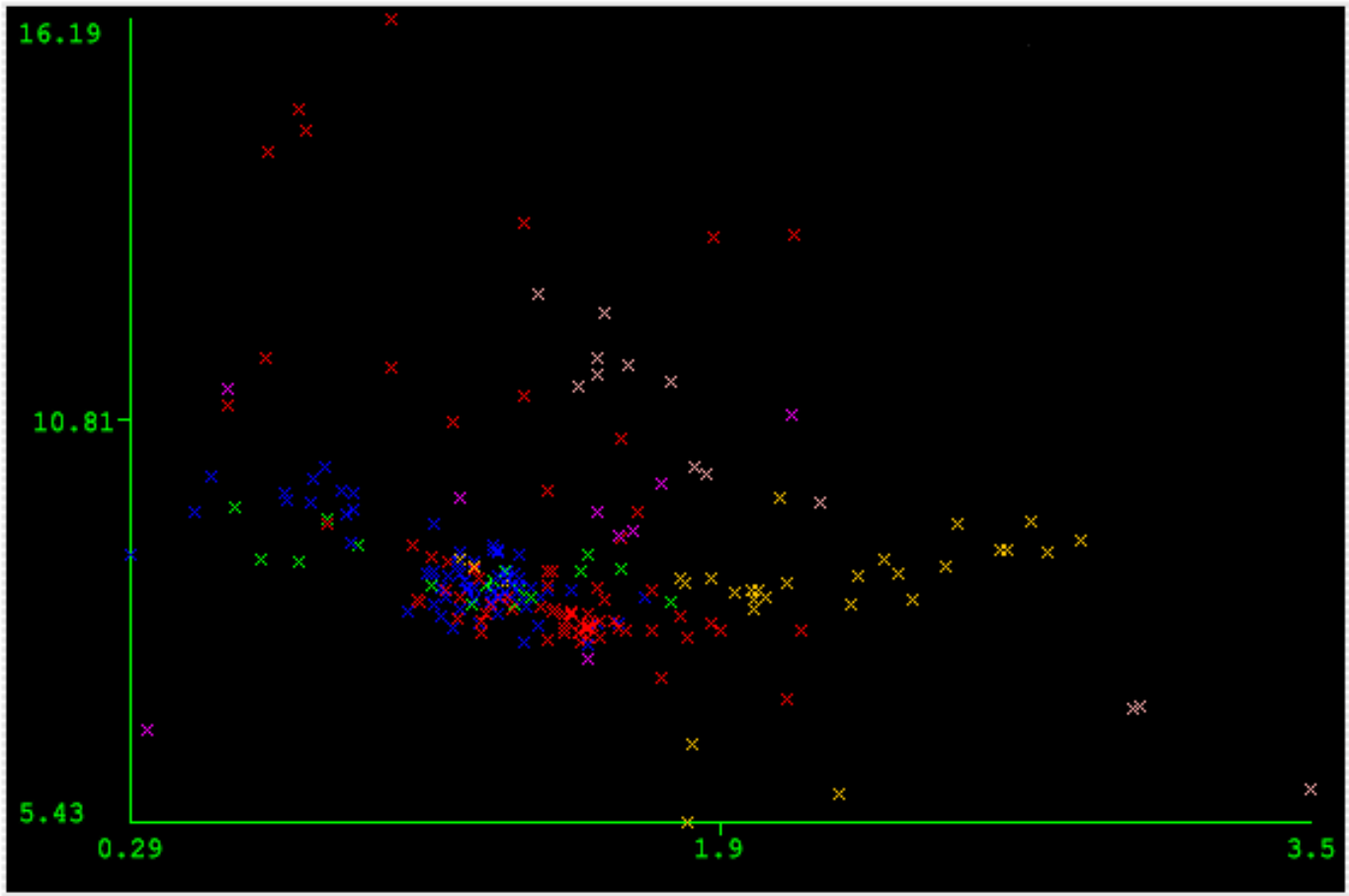
Clear

Save

Jitter



Plot: Glass



Class colour

build wind float

build wind non-float

vehic wind float

vehic wind non-float

containers

tableware

headlamps

X: Al (Num)

Y: Ca (Num)

Colour: Type (Nom)

Rectangle

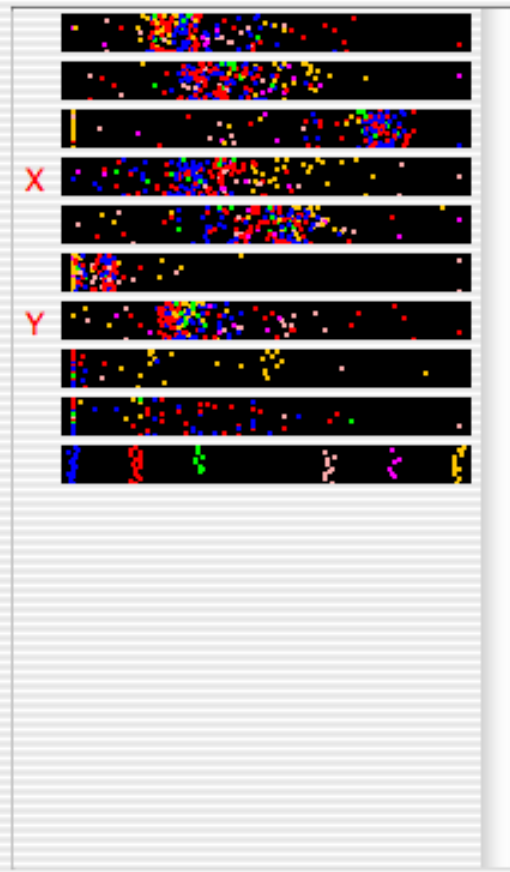
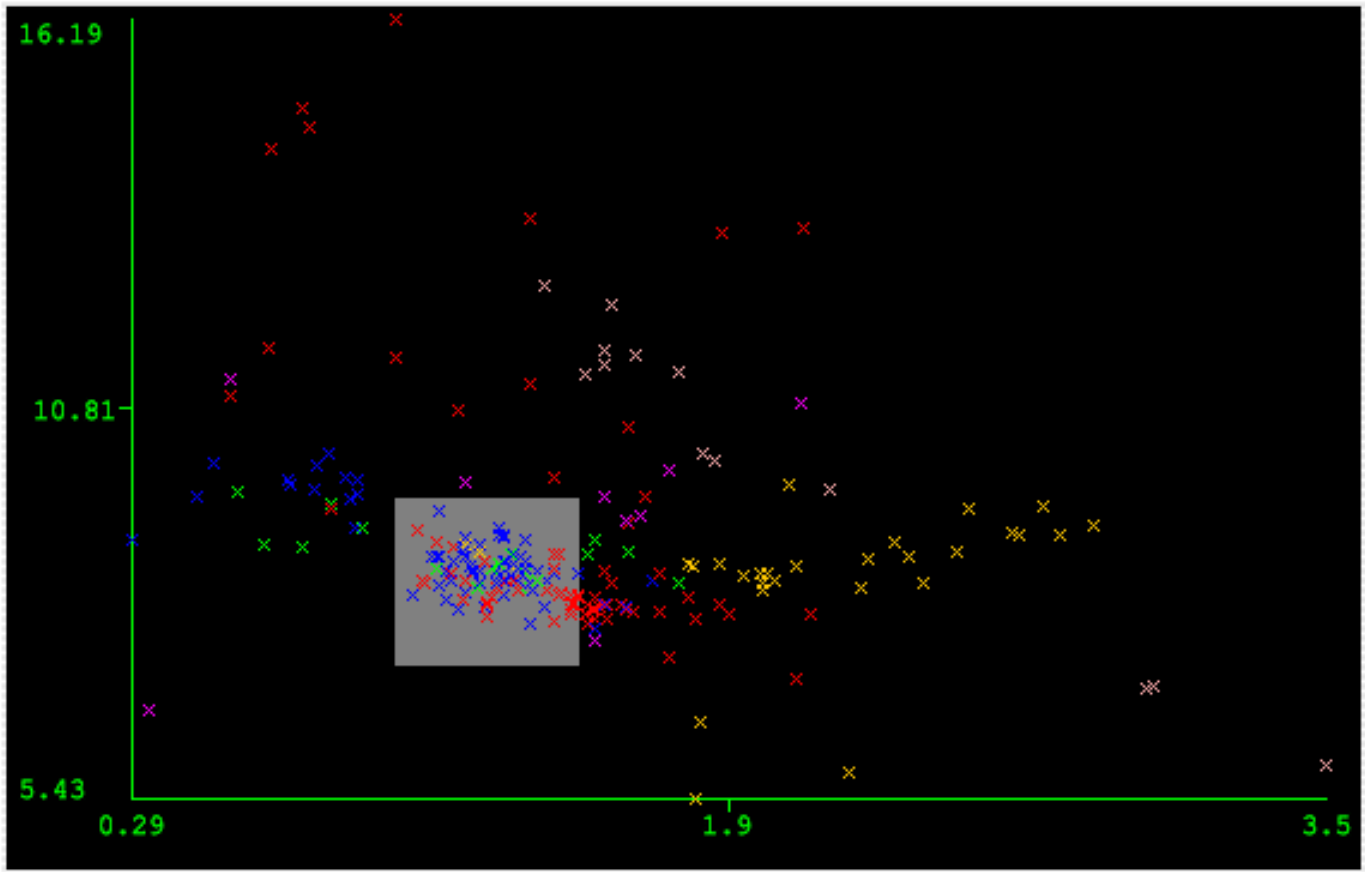
Submit

Clear

Save

Jitter

Plot: Glass



Class colour

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

X: Al (Num)

Y: Ca (Num)

Colour: Type (Nom)

Rectangle

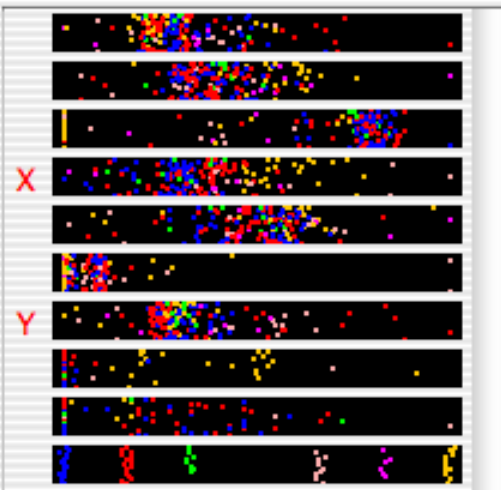
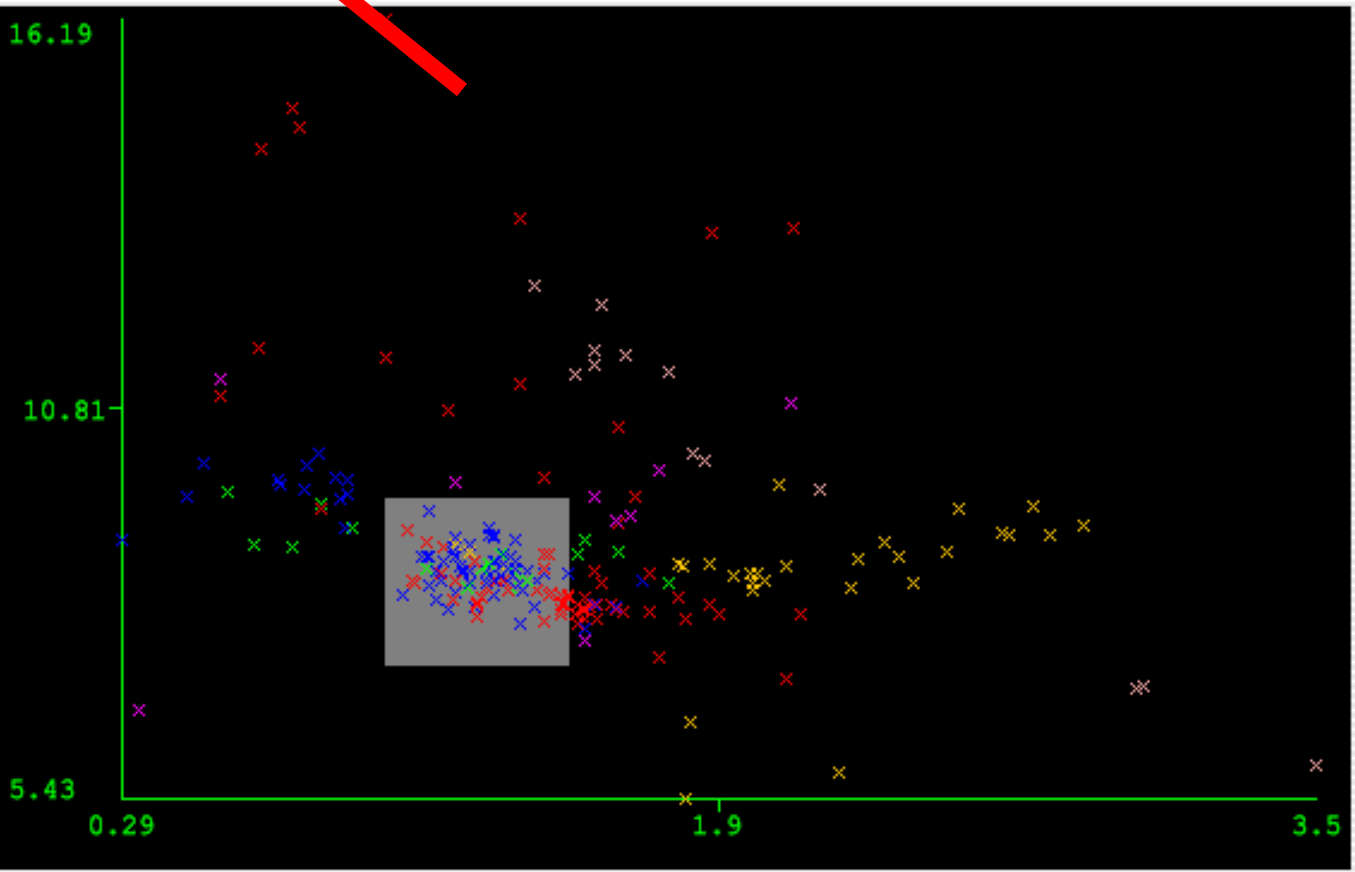
Submit

Clear

Save

Jitter

Plot: Glass



Class colour

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

X: Al (Num)

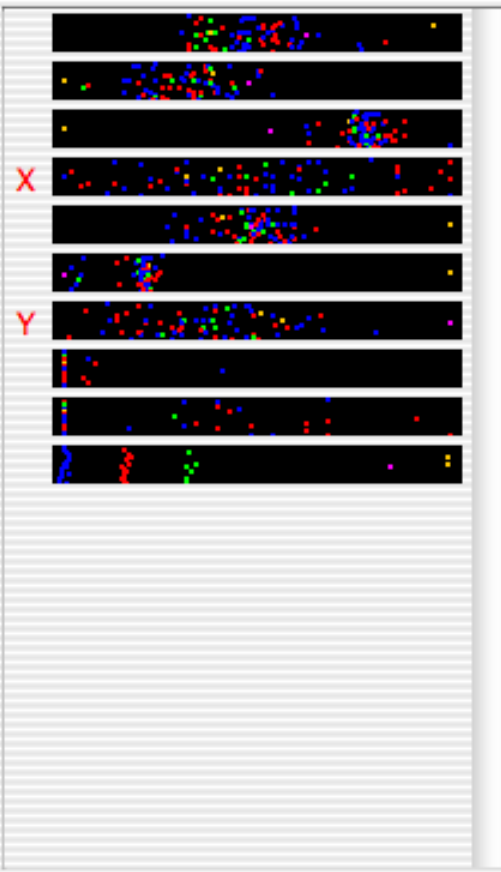
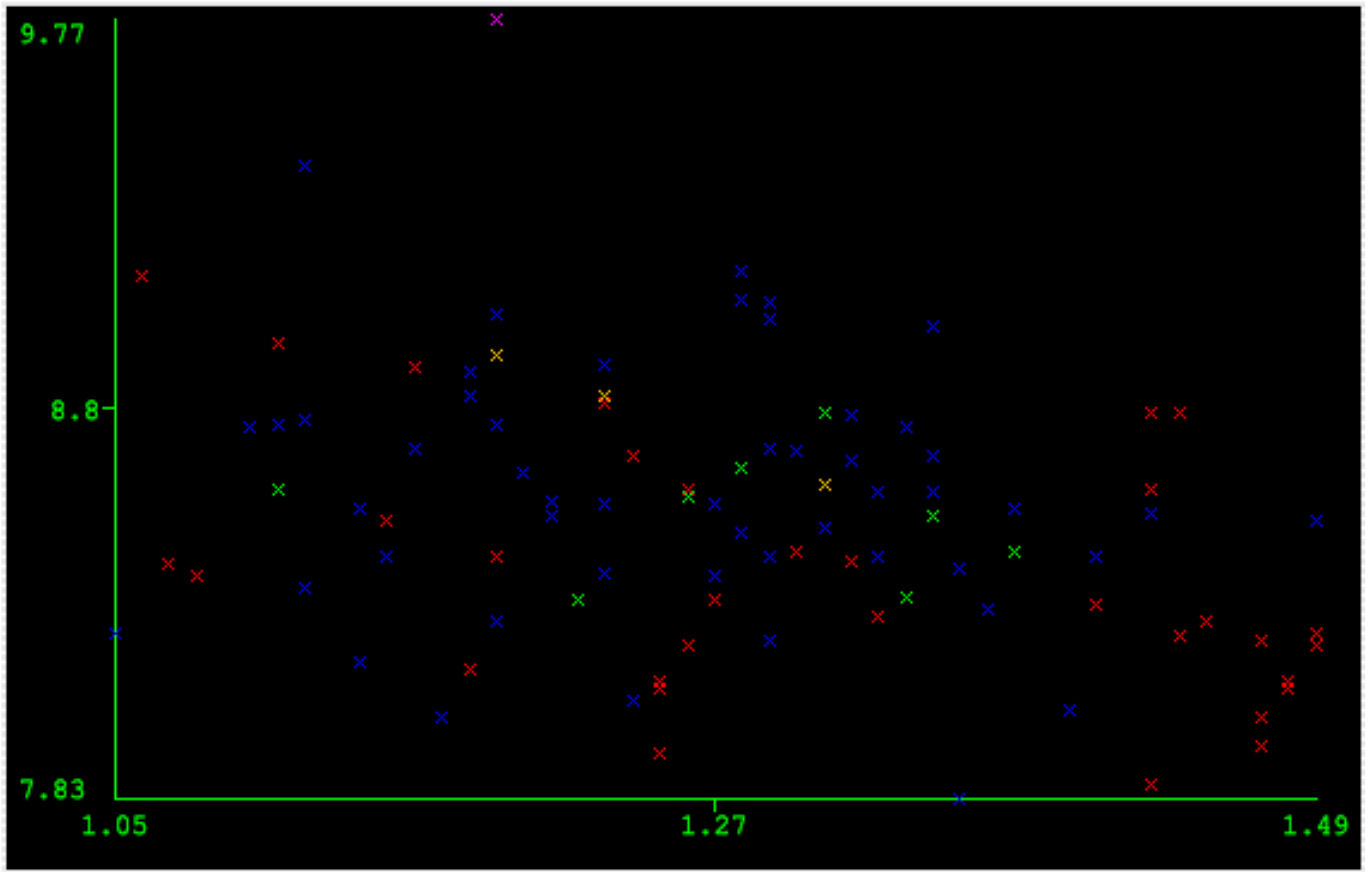
Y: Ca (Num)

Colour: Type (Nom)

Rectangle

Jitter

Plot: Glass



Class colour

build wind float

build wind non-float

vehic wind float

vehic wind non-float

containers

tableware

headlamps

Οπτικοποίηση

- Επιτρέπει την δισδιάστατη γραφική παρουσίαση της τρέχουσας ομάδας δεδομένων.
- **Scatter plot matrix:**
 - “Visualize panel” παρουσιάζεται το scatter plot matrix για όλα τα χαρακτηριστικά (attributes), έχοντας το καθένα το δικό του χρώμα, σε σχέση πάντα, με την επιλεγμένη κλάση.
 - Δίνεται η δυνατότητα να αλλαχθεί το μέγεθος κάθε δισδιάστατου - 2D - plot , όπως και το μέγεθος των σημείων. Επίσης, δίνεται η δυνατότητα, τυχαία να μετακινηθούν τα δεδομένα (jitter – ανακαλύπτοντας έτσι κρυμμένα σημεία).
- **Επιλέγοντας ξεχωριστά ένα-ένα το δισδιάστατο - 2D scatter plot**
 - Κάνοντας κλικ σε ένα κελί (cell) μέσα στο scatter plot matrix, επιστρέφεται ένα ξεχωριστό παράθυρο οπτικοποίησης (visualization) του scatter plot που επιλέχθηκε

Οπτικοποίηση

Selecting Instances

- Υπάρχουν περιπτώσεις, που είναι χρήσιμο να επιλεγεί ένα υποσύνολο δεδομένων χρησιμοποιώντας το εργαλείο οπτικοποίησης (visualization).

4 Τρόποι Επιλογής Σημείων Δεδομένων πάνω στο plot:

- **Select Instance.** Κάνοντας κλικ πάνω σε ένα data point, ανοίγει ένα παράθυρο που καταγράφει σε μια λίστα όλα τα χαρακτηριστικά αυτού του σημείου (attributes).
- **Rectangle.** Μπορείτε να δημιουργήσετε ένα τετράγωνο, by dragging, μέσα στο οποίο θα επιλέγετε να βάλετε τα σημεία-δεδομένα που θέλετε (από αυτά που υπάρχουν στο plot).
- **Polygon.** Μπορείτε να δημιουργήσετε ένα πολύγωνο, επιλέγοντας τα σημεία που θέλετε να βρίσκονται μέσα σε αυτό. Χρησιμοποιείστε, Αριστερό-Κλικ πάνω σε ένα σημείο για να το προσθέσετε μέσα στο πολύγωνο και Δεξί-Κλικ για να φτιαχτεί, ολοκληρωθεί το πολύγωνο.
- **Polyline.** Μπορείτε να δημιουργήσετε μία ή περισσότερες γραμμές που θα τοποθετούν τα σημεία μεταξύ των δύο πλευρών που θα διαχωρίζει η μία ή οι περισσότερες γραμμές. Χρησιμοποιείστε, Αριστερό-Κλικ για να προσθέσετε ένα σημείο και Δεξί-Κλικ για να φτιαχτεί, ολοκληρωθεί το Polyline.

Δέντρα Απόφασης (Interactive decision tree construction)

